ENCYCLOPEDIA OF ANALYTICAL CHEMISTRY

Applications, Theory and Instrumentation

Edited by R. A. Meyers

Biomedical Spectroscopy • Biomolecules Analysis • Carbohydrate Analysis
• Chemical Weapons Chemical Analysis

WILEY

Biomedical Spectroscopy: Introduction

Henry H. Mantsch
National Research Council, Winnipeg, Canada

1 Introduction 1
2 Overview of Contributions 1
3 Apologia and Outlook 3
Abbreviations and Acronyms 3

1 INTRODUCTION

Biomedical Spectroscopy, fortuitously the first section in this comprehensive encyclopedia, is one of the youngest branches of analytical chemistry. Throughout much of their history, medicine and spectroscopy have evolved quite separately, yet in the new millennium the two solitudes are closing the gap that has separated them for so long as they cross-pollinate each other increasingly. Although spectroscopy, by its very nature, has always had an interdisciplinary focus, the recent marriage between spectroscopy and medicine is only now beginning to bear fruit.

It is unfortunate that the word ''chemical'' has acquired such a bad reputation in the lay press and yet the same public is increasingly captivated by ''natural biochemicals''. So what are natural biochemicals, and what is a natural substance? To answer this question we have to go back almost 20 billion years to when the universe arose with a cataclysmic explosion that hurled hot, energy-rich subatomic particles throughout space. Gradually, as the universe evolved, the various chemical elements were created, including those in all living organisms on earth. Hence, we humans are literally made of stardust. The unique molecules from which living organisms are constructed are called biomolecules and were selected during the course of evolution for their fitness to perform specific functions. It is therefore quite legitimate to ask what the purpose or the specific biological function of a given biomolecule in a living organism may be. When examined separately these biomolecules conform to all the physical and chemical laws that describe the behavior of inanimate matter, and yet living things possess unique properties not displayed by collections of inanimate molecules, thus presenting a distinctive challenge to the analytical chemist and to the biospectroscopist alike.

Medicine, known as ''the art of science of healing disease'', is more art than science to some, where the patient is treated as a whole and not as the site in which a particular disease has chosen to manifest itself. However, to many others medicine is a life science that rests firmly upon a foundation of biological sciences, which in turn, trace their origin to the basic physical sciences of chemistry, physics and mathematics. Sadly, in spite of our success to combat disease and illness, there are still health care problems, for which there is little we can do. Even if the best of contemporary medicine were universally available, regardless of financial barriers, cancer would continue to kill, rheumatoid arthritis would still cripple, and Alzheimer's disease would continue to render many elderly people helpless. Although modern medicine appears powerless to reduce the impact of these diseases, largely due to our inability to treat them successfully, the paucity of effective diagnostic techniques plays a significant role in the continued occurrence of these diseases. Early diagnosis is an important component of preventive medicine, and spectroscopic methods play an increasingly important role in biodiagnostics. Medical biospectroscopy uses the entire arsenal of electromagnetic radiation, from the high-energy end (gamma rays) down to the low-energy end (radio waves), to probe individual biomolecules and indeed the whole human body.

2 OVERVIEW OF CONTRIBUTIONS

The *Biomedical Spectroscopy* section consists of 15 articles, among which are five dedicated to optical, infrared (IR), and magnetic resonance (MR) spectroscopy. From the most ancient times, medical practitioners performed physical examinations of their patients using their eyes as optical spectroscopes. Indeed, analytical methods based on optical spectroscopy were used early in medical diagnosis, and even today many tests performed in the clinical chemistry laboratory still rely on visible spectroscopy to monitor chemical or enzymatic color reactions. Bioanalytical methods based on IR spectroscopy or on MR spectroscopy are of more recent vintage, as are several optical approaches such as optical coherence tomography or photodynamic therapy.

Biodiagnostic methods fall into two large categories: (a) biomedical spectroscopy and (b) biomedical imaging. The former is based on the interaction of selected electromagnetic waves with individual or collections of biomolecules. The resulting ''molecular spectra'', represented as plots of intensity versus electromagnetic energy, provide answers to the questions ''what?'' (qualitative bioanalytical chemistry) and ''how much?'' (quantitative bioanalytical chemistry). The second category,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

biomedical imaging, goes on to answer the question ''where is what?'', by localizing and mapping the spectroscopic information. Both biomedical spectroscopy and biomolecular imaging can be performed either ex vivo, on extracted biofluids or excised tissue, when the biomaterial is brought to the spectrometer, or in vivo, in which case the electromagnetic radiation is delivered to the target area of the patient via optical fibers, endoscopes, catheters or even through space (e.g. radio waves in MR imaging). In an ideal world, all diagnostic procedures would be noninvasive, but we live within the constraints of the real world.

The individual articles in the *Biomedical Spectroscopy* section are largely self-contained, each covering a particular area of expertise of the contributing author(s). A first group of contributions deals with optical spectroscopy. The term ''optical spectroscopy'' is not synonymous with visible spectroscopy, which at times can be confusing. Optical spectroscopy involves transitions between electronic energy levels, and thus extends beyond the violet into the ultraviolet and beyond the red into the near-IR. Ramanujam (**Fluorescence Spectroscopy In Vivo**) provides a comprehensive account of optical fluorescence and absorption spectroscopy, introducing the technique, the types of chromophores and fluorophores, and the spectrometers and fluorimeters used in clinical settings. Both absorption and fluorescence spectroscopies have been explored extensively as diagnostic tools, in particular for cancer (precancer) screening in epithelial surface layers of various organ sites (cervix, bladder, gastrointestinal tract, trachelial tube, and oral cavity). The contribution by Ramjiawan et al. (**Fluorescence Imaging**) is dedicated to optical fluorescence imaging. Fluorescence, a zero-background technique, has a much higher sensitivity and specificity than absorption spectroscopy or imaging. In particular immunofluorescence imaging, a technique based on the interaction of labeled antibodies with specific antigens, shows great diagnostic potential, even if it is not yet a common sight in hospitals. Heise (**Glucose, In Vivo Assay of**) tackles glucose, the Holy Grail of all analytes assessed in biological fluids. It is the dream of many diabetics, and of their attending physicians, to do away with the daily finger pricking and to depend on a simple, noninvasive optical wand for determining their blood glucose levels. As the reader will find out, great progress has been made towards this goal, but we are not quite there yet. Fercher (**Optical Coherence Tomography**) provides an up-to-date synopsis of the novel field of optical coherence spectroscopy and tomography, addressing a number of applications in medicine that range from ex vivo biopsy studies in dermatology, urology, and gynecology to in vivo imaging in ophthalmology, dentistry, and gastroenterology. Another type of optical spectroscopy suitable for medical applications, photodynamic therapy, is reviewed by Röder (**Photodynamic Therapy**). Photodynamic therapy, known as PDT to its practitioners, uses photosensitizers that are nontoxic in the dark but become toxic after photoactivation by light. The evolution of photodynamic treatment over three generations of photosensitizers is illustrated by a number of clinical applications with particular emphasis on skin diseases and cancer.

A second group of contributions deals with IR spectroscopy, a more recent tool in medical research and practice. Both IR spectroscopy and the complementary technique of Raman spectroscopy, derive information from the vibrations of chemical bonds in the biomolecules of interest and are therefore referred to as vibrational spectroscopies. Jackson and Mantsch (**Infrared Spectroscopy, Ex Vivo Tissue Analysis by**) introduce the reader to the basics of ex vivo tissue analysis by mid-IR spectroscopy, focusing on two facets of such an analysis. First, there is an experimental aspect related to potential pitfalls with spectroscopic measurements on samples as complex as human tissue. The second aspect concerns the interpretation of spectra, meaning the proper translation of spectroscopic information into diagnostic, medically relevant information using such tools as chemometrics and nonsubjective multivariate statistical classification methods. Every type of human tissue, although structurally highly complex, has a unique vibrational pattern (fingerprint) in the mid-IR region, which is different in healthy and in diseased tissue. This has led to the creation of a new field, IR histopathology. In vivo tissue analysis by IR spectroscopy is addressed by Sowa et al. (**Near-infrared Spectroscopy, In Vivo Tissue Analysis by**). Compared to both visible light and mid-IR radiation, near-IR light can traverse a greater distance into tissue, therefore the spectral range of near-IR is ideally suited for in vivo tissue spectroscopy and imaging. The penetration depth of near-IR light is of the order of several centimeters, whereas only the top 10–20 μm of tissue can be explored by mid-IR light. The clinically relevant tissue chromophores oxy- and deoxyhemoglobin, oxy- and deoxymyoglobin, oxidized and reduced cytochrome *c*, as well as water, provide vital information related to oxygen delivery, storage, utilization and tissue hydration/dehydration. The article by Shaw and Mantsch (**Infrared Spectroscopy in Clinical and Diagnostic Analysis**) offers an overview of the clinical analyses that have been carried out by IR spectroscopic methods on such common biological fluids as serum, whole blood and urine, as well as on less-common body fluids such as amniotic fluid, synovial fluid, cerebrospinal fluid, and saliva. The term ''IR clinical chemistry'' was coined as an analytical technique that does not require chemical or biochemical reagents for the quantitative determination of analytes; instead, the analysis relies on chemometric algorithms. A

digression into the realm of microbiology, utilizing mid-IR spectroscopy as the investigative tool, is recounted by Naumann (**Infrared Spectroscopy in Microbiology**). This methodology offers an alternative analytical tool for the detection, enumeration, classification and identification of pathogenic bacteria in a clinical setting. Finally, Ozaki and Noda (**Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences**) introduce the reader to a new modality for extracting additional information from IR spectra. Two-dimensional correlation spectroscopy already has revolutionized MR spectroscopy and is expected to benefit IR biomedical spectroscopy as well.

A third group of contributions explores nuclear magnetic resonance (NMR) in medicine. There is now a tendency to drop the ''N'' in NMR, because the word ''nuclear'' is unpopular. Even though optical and IR spectroscopy have been around much longer, MR spectroscopy enjoys remarkable success today. Winter and Bansal (**Magnetic Resonance, General Medical**) provide an authoritative overview of multinuclear NMR in medicine. As MR spectroscopy involves the absorption of certain radiofrequencies by individual nuclei in a magnetic field, it must be performed in a magnet. The higher the magnetic field, the greater the detection sensitivity, a fact which has fuelled the race for ever higher-field magnets. Smith and Blandford (**Magnetic Resonance in Medicine, High Resolution Ex Vivo**) introduce the reader to the intricacies of ex vivo high-resolution MR spectroscopy in medicine. The applications they highlight range from the study of inborn errors of metabolism to the diagnosis of neurological disorders and many types of cancer. The last three contributions describe various MR imaging modalities. Richards (**Multinuclear Magnetic Resonance Spectroscopic Imaging**) discusses MR spectroscopic imaging, a procedure for generating spatially resolved maps and images of selected biomolecules in tissue (ex vivo), or in the body (in vivo). This molecular-level imaging, also known as chemical shift imaging, is similar to chemical group imaging by IR spectroscopy. Both imaging modalities allow the noninvasive mapping of chemicals in the body, although IR imaging is limited to external body parts such as skin, whereas the whole body is accessible to MR imaging. Richter (**Magnetic Resonance Imaging, Functional**) familiarizes the reader with functional magnetic resonance imaging (fMRI), perhaps the most powerful imaging technique available today. The stellar performance of fMRI in the field of neuroscience has filled in many blanks on the human brain map, and fMRI is now a premier method for the study of brain function. Finally, Yuan (**Magnetic Resonance Angiography**) exploits the emerging use of MR in angiography, an area until recently reserved for X-ray radiologists. The name ''angiography'' applies to any imaging modality that can visualize blood vessels and blood flow. Potential clinical applications are discussed, the challenge being coronary MR angiography because the three-dimensional structure of the coronary tree changes shape during each heart contraction. A concern sometimes voiced with in vivo imaging is the potentially harmful electromagnetic radiation. Although this may apply to X-ray imaging, molecular spectroscopic imaging uses low-energy electromagnetic fields and very few adverse effects have been recorded.

3 APOLOGIA AND OUTLOOK

While I want to thank all the contributing authors for their effort and dedication, as editor of the *Biomedical Spectroscopy* section I also apologize to my colleagues in the biospectroscopic community whose work could not be included. I had to be selective and no doubt this selectivity reflects my own bias and preoccupations with certain aspects of the exploding field of biomedical spectroscopy. As to the future of biomedical spectroscopy, I am optimistic that in the process of growing and maturing it will embrace other traditional areas of analytical chemistry. When pressed to be more specific, I would like to refer the reader to a statement by Alan Kay, one of the founders of Silicon Valley in California, ''the best way to predict the future is to invent it''. So, good luck and ''happy hunting grounds'' to future generations of bioanalytical chemists.

ABBREVIATIONS AND ACRONYMS

fMRI	Functional Magnetic Resonance Imaging
IR	Infrared
MR	Magnetic Resonance
NMR	Nuclear Magnetic Resonance

Fluorescence Imaging

Bram Ramjiawan, Michael Jackson, and Henry Mantsch
Institute for Biodiagnostics, Winnipeg, Canada

1 Introduction **1**
2 Chromophores **2**
2.1 Endogenous Chromophores 2
2.2 Exogenous Chromophores 3
3 Immunofluorescence Imaging **5**
3.1 Choice of Antibody 5
3.2 Choice of Fluorophore 6
4 Choice of Detection Systems **7**
4.1 Charge-coupled Device Cameras 7
4.2 Filter Selection 7
4.3 Typical Experimental Set-up 7
5 Immunofluorescence Imaging of Tumors **8**
5.1 Qualitative Analysis of Images 8
5.2 Quantitative Analysis of Images 9
5.3 Statistical Analysis of Images 11
6 Summary and Future Prospects **11**
Abbreviations and Acronyms **13**
Related Articles **13**
References **14**

A number of fluorescence imaging techniques show diagnostic promise. Imaging endogenous fluorescence has been proposed as a method for cancer diagnosis. Unfortunately, tissue autofluorescence is relatively weak and poor contrast between malignant and normal tissue is seen. Contrast may be enhanced with the addition of fluorescent materials that are selectively accumulated by malignant cells, such as fluoroscein or porphyrin derivatives. The limited penetration of light at the emission maxima of these materials restricts the use of fluorescence techniques utilizing these chromophores to superficial phenomena. However, many potential applications still exist. For example, monitoring fluorescence during surgery may allow resection margins to be clearly delineated.

Other exogenous chromophores that may have may have diagnostic utility include indocyanine green (ICG). Techniques based upon visualization of the distribution of ICG fluorescence (i.e. choroidal angiography) are already prominent in ophthalmology. ICG fluorescence imaging may also find a useful niche in monitoring of burns and transplant tissue. In addition, monitoring of vascular parameters during cardiac surgery presents exciting opportunities. For example, low oxygen levels (e.g. during bypass surgery) in the heart can result in alterations in microvascular permeability. As ICG is largely bound to serum albumin, it should not be seen in extravascular spaces in normal hearts. However, increased permeability will allow albumin to diffuse into the extravascular spaces, and diffuse fluorescence across the surface of the heart will be seen. In principle, the infusion of polymers (e.g. dextrans) of various molecular weights labeled with dyes that fluoresce at various wavelengths will allow the assessment of the porosity of capillary beds in such systems.

Immunofluorescence techniques have the potential to provide unmatched sensitivity and specificity. The unique nature of antibody–antigen interactions ensures specific delivery of the fluorophore to the site of interest. The specific interaction of labeled antibodies with antigens means that the fluorophore persists in the body for a prolonged period of time (days). Following a single injection of labeled antibody, repeated measurements on the same site over the course of hours or days allow kinetic information to be readily obtained. In principle, this means that the effect of therapeutic intervention, i.e. radiation therapy, chemotherapy, etc., can be monitored.

1 INTRODUCTION

Owing to their high sensitivity, fluorescence techniques have found a valuable place in biology. The sensitivity of fluorescence methods is due in large part to the fact that they are ''zero background'' techniques. In other words, in the absence of the chromophore of interest, no signal (other than random noise) can be detected. Any signal that is detected must therefore arise from the chromophore of interest.

Fluorescence methods with potential for use in medicine utilize both intrinsic and extrinsic chromophores. The number of intrinsic chromophores (i.e. chromophores that occur naturally in tissues) is limited; the most important are listed in Table 1. A number of literature reports suggest that tissue nicotinamide adenine dinucleotide (reduced form) (NADH) fluorescence may prove useful in the detection of malignancies, the rationale being that changes in the metabolic status of malignantly transformed cells are reflected in changes in fluorescence signals arising from NADH.[1,2] Unfortunately, intrinsic fluorescence is typically low in most tissues and applications to date are limited.[3]

The range of extrinsic chromophores of use (i.e. chromophores which are added to samples in some

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Selected chromophores and their excitation and emission properties

Chromophore	Excitation maximum (nm)	Emission maximum (nm)
Tryptophan	275	350
Collagen	340	395
Elastin	460	520
NADH	350	460
Porphyrins	400	610, 675
ICG	790	815
Cy3	550	565
Cy5	690	710
Cy7	750	777

manner) is obviously much wider, and fluorescent chromophores are available covering a wide spectral range [from the visible to the near-infrared (NIR)].[4,5] Typically, extrinsic chromophores are introduced into samples either as free agents or attached to carrier materials such as antibodies to specific materials present in the sample.[6–8]

Detection of fluorescent materials in samples may be achieved using either spectroscopic or imaging technologies. Spectroscopic techniques may be used to analyze small regions of tissue and provide information relating to the average concentration of fluorescent materials within the sample. Fluorescence spectroscopy may be performed in vitro, e.g. using homogenized tissues, or in vivo, e.g. with the use of fiber-optic bundles.[1,9] The advantage to the spectroscopic approach is that fluorescence across the entire spectral range of interest is acquired, allowing information on multiple chromophores to be obtained with one measurement. However, the obvious disadvantage to this approach is that only a small volume of tissue can be analyzed, and the resulting signal is an average signal from that tissue volume. Spatial information is not obtained.

Spatial information may be obtained by fluorescence imaging, in which the distribution of fluorescence intensity is measured as a function of position within the sample. Fluorescence imaging is typically achieved using imaging arrays such as charge-coupled device (CCD) cameras equipped with appropriate filters. While spatial information may be obtained in this manner, fluorescence imaging has typically been performed at a single frequency. Thus, spectral information is lost. Multiple imaging frequencies (to detect multiple chromophores) require changing between multiple filters, a tedious procedure. The development of the liquid-crystal tunable filter (LCTF) has removed this limitation. The transmission characteristics of LCTFs may be varied under computer control, allowing the user to vary the imaging wavelength without changing filters. With such filters, fluorescence spectroscopic imaging becomes feasible.

2 CHROMOPHORES

2.1 Endogenous Chromophores

Although a number of endogenous fluorescent chromophores exist in the body (see Table 1), their use in in vivo fluorescence studies can be problematic. The most common problems are as follows:

1. The low level of fluorescence typically seen in tissues.
2. Overlap of excitation and emission wavelengths for absorption. Many endogenous chromophores may be excited at similar wavelengths. For example, collagen fluorescence may be stimulated by illumination at 340 nm, very close to the excitation wavelength of NADH (350 nm). Overlap of excitation and emission wavelengths can be even more problematic. For instance, whereas tryptophan and NADH fluorescence is characterized by well separated excitation wavelengths (275 and 350 nm, respectively), the emission maximum for tryptophan fluorescence is seen at 350 nm. Hence, tryptophan fluorescence at 350 nm may be masked owing to absorption of the emitted radiation by NADH.
3. Ultraviolet (UV) or visible excitation and emission maxima. As discussed above, chromophores that exhibit excitation and emission maxima in the UV or visible spectral regions limit studies to the characterization of superficial structures, owing to limited penetration of light through tissues at these wavelengths. In addition, illumination of tissues with UV light is obviously undesirable.

Despite these limitations, imaging fluorescence arising from endogenous chromophores has been demonstrated to have clinical potential, particularly in the diagnosis of malignancies. For example, it has been demonstrated that malignant tissue shows enhanced accumulation of endogenous porphyrins, which exhibit fluorescence in the red part of the visible spectrum.[7,10] It has therefore been suggested that changes in fluorescence arising from endogenous porphyrins may be useful in tumor detection.

Differences in NADH fluorescence are also apparent between normal and malignant cells and tissues. NADH exhibits relatively strong fluorescence at 350 nm, whereas its oxidized counterpart, NAD^+, exhibits only weak fluorescence. The ratio of NADH to NAD^+ is decreased in malignant tissues, thus leading to a reduced fluorescence signal in malignant tissues. It should be stressed that factors that result in metabolic disturbances may also potentially affect NADH/NAD^+ ratios and produce altered tissue fluorescence signatures. The specificity of NADH fluorescence methods for malignancies is at present unclear.

2.2 Exogenous Chromophores

Exogenous chromophores are those introduced into samples by some mechanism either to allow the detection of specific materials (e.g. when attached to antibodies), to probe environmental parameter (e.g. pH) or to provide contrast (e.g. angiography). As discussed above, the number of potential exogenous fluorescent chromophores is large, allowing greater flexibility in the choice of the spectral range to be employed for excitation and emission monitoring. In fact, in many instances chromophores can be chemically tailored to meet excitation and emission criteria.

The chromophore of choice depends upon a number of factors, including the nature of the experiment, excitation and emission wavelengths and chemical structure. Chemical structure is important if the fluorophore is to be attached to a carrier material: the fluorophore must contain a reactive group or be capable of derivatization or produce an active group which will allow chemical linkage to the carrier material. In many respects, the limitations in chromophore selection imposed by the nature of the experiment and excitation and emission wavelengths are more severe. The applications of fluorescent dyes that show excitation and emission wavelength maxima in the UV or visible regions of the spectrum are restricted owing to the limited penetration of UV and visible light into tissues. Thus only fluorescence signals from superficial chromophores can be recorded in these spectral regions. Dyes which fluoresce in the far-red or NIR regions have greater potential diagnostic use, owing to the enhanced penetration of light through tissues in these spectral regions.

2.2.1 Ultraviolet/Visible Fluorophores: Porphyrin Derivatives

Any fluorescent material that is selectively accumulated by tumor cells can in principle be used as a diagnostic marker. In this respect, many therapeutic agents have potential as diagnostic agents. Porphyrin derivatives used in photodynamic therapy are particularly well suited in this respect.[6,10–15] Already used as therapeutic agents (with toxicity data, etc., available) these agents are selectively accumulated by malignant cells and exhibit fluorescence. Illumination with light at around 400 nm allows a fluorescence signal to be imaged in tissues and cells at 600–650 nm, depending upon the nature of the porphyrin derivative. The utility of this approach has been demonstrated by imaging the fluorescence arising from a hematoporphyrin derivative accumulated in colon adenocarcinoma cells implanted in the leg of a rat. Illumination at 337 nm 24 h following administration of a low dose of the porphyrins resulted in detectable fluorescence and visualization of the tumor.

Accumulation of porphyrins, including hematoporphyrin derivative and protoporphyrin IX, have been used to diagnose a number of forms of cancer in humans. In vivo fluorescence imaging of skin showed higher fluorescence in basal cell carcinomas lesions than the surrounding normal tissue. Clear demarcation of lesion borders was seen in images. Malignant tumors of the breast, head and neck region and urinary bladder have been visualized using a fluorescence imaging system using 100-ns long optical pulses at 390 nm delivered via an endoscope system which was also coupled to the imaging system. Tumor detection was achieved based on differential fluorescence between normal and malignant tissue, related to the selective uptake of tumor-marking agents such as hematoporphyrin derivative and levulinic acid, and natural chromophore differences between various tissues. A clear demarcation from normal surrounding tissue was found in measurements of superficial bladder carcinoma and in vitro investigations of resected breast cancer.[16]

While the use of porphyrins as chromophores in fluorescence imaging is appealing, it is not without drawbacks.[6,7,10,14] The most important drawbacks are the time required to achieve significant differential accumulation of the porphyrin in malignant tissue and potential side effects such as photosensitization.

2.2.2 Visible Fluorophores: Fluorescein

In contrast to porphyrin derivatives, toxic side effects associated with fluorescein use are minimal. Fluorescein has been used for many years in studies of the vascular beds within the eye. Illumination of the retina with blue light produces green fluorescence in the vessels of the eye following injection of a bolus of fluorescein. Acquisition of images (typically with the use of photographic film) of the distribution of fluorescence produces angiograms, images of the vessels within the eye. While fluorescein angiography is slowly being superseded by ICG angiography, other applications of fluorescein fluorescence are being pursued.

Malignant tissues show an enhanced accumulation of fluorescein compared with normal tissues. At neutral pH fluorescein exists as a mixture of charged and uncharged species. Accumulation of fluorescein by cells results from passive diffusion of the neutral form of the dye across the cell membrane. In malignant tissues, the extracellular pH is often lower than that seen in normal tissue. This lower pH shifts the dissociation curve of fluorescein, resulting in a higher concentration of the neutral form of the dye in the extracellular fluid. This in turn results in an increased passive diffusion of the dye across the cell membrane. As tumor cells are generally able to maintain a neutral intracellular pH, the equilibrium within the cell is shifted

in favor of dissociation, reducing the concentration of the neutral form of the dye and reducing outward diffusion. The dye therefore accumulates within malignant cells.[17]

Preferential uptake of fluorescein by malignant tissues has been known for more than 50 years, However, the dye exhibits green fluorescence when excited with blue light. In other words, the excitation and emission maxima are both in the visible portion of the spectrum. Penetration of visible light through tissues is poor and fluorescein spectroscopy and imaging are limited to accessible surfaces. Fluorescein imaging is therefore most useful if endoscopic techniques are used to study organs such as the lungs or the gastrointestinal tract. Endoscopic detection of fluorescein fluorescence has been demonstrated for gastric cancer. Even more promising is the finding that dysplasia could be detected based upon fluorescein fluorescence in 22 of 23 specimens of cheek pouch from carcinogen-treated Syrian hamsters.[9]

Of course it should be noted that conditions other than malignancies might induce a fall in extracellular pH, e.g. inflammatory conditions. If such conditions are also associated with maintenance of intracellular pH, i.e. the formation of a pH gradient across the cell, then they may show fluorescein accumulation and enhanced fluorescence may result. The specificity for fluorescein for malignant tissues has yet to be determined.

2.2.3 Near-infrared Fluorophores: Indocyanine Green

If information from deeper, non-surface structures is required, fluorophores that exhibit excitation and emission maxima in the far-red or NIR spectral regions, in which tissues are relatively transparent to light, are required.

The most widely used NIR fluorescent dye is ICG (also known as Cardio-Green or Fox Green). ICG exhibits absorption and emission maxima at 780 and 810 nm, respectively, wavelengths at which light can readily traverse several millimeters of tissues. Originally used to monitor cardiac and hepatic function (based upon the transit time of a bolus of the fluorescent dye through the circulatory system), ICG has recently become more widely utilized as a contrast agent in studies of the vasculature.[18–21]

The most common application of ICG fluorescence in vascular studies is without doubt in choroidal angiography, i.e. imaging the choroidal vessels of the eye, vessels that provide 70% of the nutrient flow to the retina. A discussion of this active field of clinical use is beyond the scope of this article and the interested reader is referred to a number of papers on this subject.[18,22–24] It is sufficient to say here that by coupling a CCD camera to the fundus camera normally used to view the retina in a clinical setting, fluorescence images of the choroid vasculature can be obtained. By studying the time course of fluorescence changes in images, hemodynamic information can be obtained. Static images and hemodynamic information can be used to evaluate vision-related problems such as age-related macular degeneration.[18,23]

Monitoring tissue vascularity is important in many other clinical settings, and ICG fluorescence imaging may play an important role in these areas. One obvious application of ICG fluorescence imaging is monitoring vascularization of tumors. Typically tumors have high metabolic rates, and so have requirements for high rates of blood flow. As the tumor develops, this high blood flow requirement is met by neovascularization, that is, the formation of new blood vessels. Thus many tumors are extremely well vascularized. Imaging the distribution of blood vessels within tumors is therefore of potential diagnostic utility. This technique has been demonstrated to be effective for studying vascular bends in melanoma of the retina. However, it should be stressed that only a limited subset of tumors will be amenable to study in this manner. An absolute requirement for such a technique is a clear differentiation between the vascular structures of the tumor and surrounding tissues. If this clear differentiation is not present, then diffuse fluorescence from ICG in the vasculature of the surrounding tissue will mask fluorescence of ICG in vessels within the tumor. In other words, the zero background advantage of fluorescence techniques will be lost, reducing sensitivity and specificity.

In addition to providing contrast for studying vasculature, ICG has also been used in the investigation of burns.[25] In contrast to the above studies, studies involving burns have focused upon the degree of ICG fluorescence recorded at various times after a burn was induced was correlated with clinical outcome of burns of various depths. Burns were produced in anesthetized pigs by placing a heated brass block on the skin. Burn depth was varied by varying the time of contact between the block and the skin. Following creation of the burn, a bolus of ICG was injected into the circulatory system at various time points, the skin illuminated at 780 nm and fluorescence images acquired using a CCD camera. Comparison of fluorescence in burns and adjacent normal tissue revealed interesting differences that could be related not only to the stage of the burn (i.e. time after the burn was created) but also to burn depth. Fluorescence images from deep burns (i.e. those produced by prolonged contact between heated metal block and the skin) showed a decreased fluorescence intensity at the burn site compared with superficial burns. This difference in fluorescence intensity was related to a decreased ICG content in deeper vessels in the deep burn. In superficial burns, such vessels remain largely intact, allowing ICG flow and a strong ICG fluorescence signal. In deep burns

these vessels are damaged and become occluded, reducing blood and ICG flow and reducing the fluorescence signal.

In addition to discriminating between burns of different thickness, ICG fluorescence imaging was also able to discriminate between fresh and older (24 h) burns. A significant (twofold) elevation of fluorescence intensity was seen in images of fresh burns compared with images of burns recorded after 24 h. This difference was attributed to an increase in capillary permeability in fresh burns, leading to an increased efflux of ICG from capillaries and accumulation in the extravascular space.

Finally, and perhaps most importantly, ICG fluorescence imaging was able to discriminate between burns which healed within 21 days and those which did not heal within 24 h. At all time points, ICG fluorescence images showed a higher fluorescence intensity in burns which healed within 21 days compared with those which did not heal. Presumably this difference may be explained at least in part by a greater blood flow (leading to a greater ICG fluorescence signal) in burns which would heal. This greater blood flow in burns that would heal would have two beneficial effects, namely delivery of an adequate supply of nutrients to the regenerating tissues and removal of toxic waste products.

3 IMMUNOFLUORESCENCE IMAGING

The examples discussed above share a common drawback: a lack of specificity. In addition, toxicity may pose difficulties. These problems can be alleviated with the use of more effective targeting strategies that do not require the use of potentially toxic compounds. The most promising of these alternative approaches is immunofluorescence imaging. Immunofluorescence techniques combine the sensitivity of fluorescence measurement methods with the specificity of immunochemistry, allowing highly specific detection of low concentrations of materials. Specificity is achieved with the use of an antibody specific to the material that the investigator wishes to detect, e.g. a cell surface antigen expressed uniquely by cancer cells. Exposure of cells to such an antibody labeled with a fluorescent dye results in accumulation of the labeled antibody only on the surface of cancer cells and not on the surface of normal cells. Detection of a fluorescence signal from samples therefore confirms the presence of malignant cells.[26–30] Specificity is provided by the unique nature of antibody–antigen interactions, while low limits of detection are possible owing to the inherent sensitivity of fluorescence techniques.

Immunofluorescence techniques are of course related to radioimmunoassays.[31–37] However, radiolabeled materials pose a human and environmental risk before, during and after use. Specialized precautions, equipment and training are required for their use. Immunofluorescence techniques eliminate the need for radionuclides and the associated specialized equipment, expertise and health and safety risks while maintaining sensitivity.

Success in immunofluorescence imaging depends upon the correct choice of antibody, fluorophore and detection system.

3.1 Choice of Antibody

3.1.1 General Properties of Antibodies

To be detectable by immunofluorescence techniques, a substance must be immunogenic when introduced into a host animal, i.e. capable of inducing an immune response, and this immune response must result in the production of antibodies. Such immunogenic substances are termed antigens. Compounds that are immunogenic have certain general characteristics that include the following: they are foreign to the immunized organism, generally of high molecular weight and chemically complex. Proteins are good examples of immunogenic compounds. When a protein is used as an immunogen, it induces an immune response that results in the production of antibodies that exhibit a remarkable specificity towards that protein.

Although the ideal condition for eliciting an immune response involves having foreign substances of high molecular weight that are chemically complex, situations exist in which an immune response can be mounted against simple compounds of low molecular weight. In these situations, the small compound is rendered immunogenic by chemical linkage to a high molecular weight substance such as a protein.

Following incubation of the antigen in a host animal, antibodies to the antigen may be isolated. Antibodies are globular glycoproteins produced by B-lymphocytes in response to the presence of foreign substances. Figure 1 shows the generalized structure of an antibody molecule. At the molecular level, antibody molecules are made up of four polypeptide chains, two identical light chains (25 000 Da) and two identical heavy chains (50 000 Da). The structure of the antibody molecule is stabilized by a number of disulfide bridges. There are two antigen-binding sites on each antibody molecule, each having both a constant and a variable region. This variable portion can adopt an apparently infinite variety of subtly different forms that allow it to bind specifically to a vast variety of antigens.[38] Interaction of an antibody with an antigen is governed by noncovalent forces, including hydrophobic, electrostatic and van der Waals forces and hydrogen bonding.

The nature of the heavy chain components of the constant region of the antibody molecule determines the

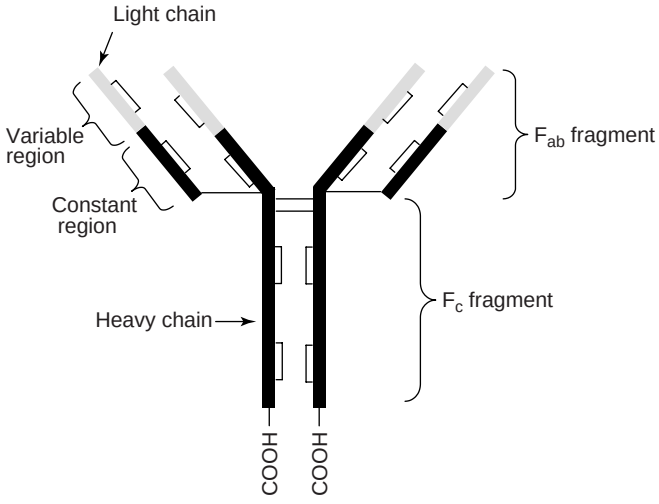

Figure 1 Generalized structure of an antibody molecule.

physical properties of the antibody, and antibodies are grouped into five classes based upon these properties. The five functional classes of immunoglobulin (Ig) are IgA, IgM, IgG, IgE and IgD. The biological properties of each of these classes are unique. For example, IgE is the major class of Ig involved in allergy and binds with high affinity to mast cells; IgG is the only class of Ig that crosses the placenta, providing immunity to the fetus; IgA is the major antibody that is found in saliva and tears; IgM can activate other components of the immune system to rupture bacteria and other cells.[39]

3.1.2 Monoclonal Versus Polyclonal Antibodies

Normal B-lymphocytes are each capable of producing an antibody to a specific antigen determinant (the region of the antigen recognized by the antibody) when activated. As antigens usually have multiple antigenic determinants, a mixture of antibodies is therefore produced in the serum of inoculated animals. Each activated B-lymphocyte forms a clonal population of cells in the spleen, which produce the same antibody as the parent cell. Thus, a polyclonal population of cells is found in the spleen, secreting a wide variety of antibodies. Even relatively simple antigens normally lead the generation of a mixture of antibodies with different specificities and affinities.[39] This problem of a heterogeneous antibody population can be eliminated with the production of monoclonal antibodies. By fusing the clonal B-lymphocytes from the spleen with immortal myeloma cells, a group of immortal, antibody-producing hybrids are produced. If individual hybrid cells are then harvested and grown in culture, each individual cell gives rise to a colony of clonal cells, and each cell in the colony secretes the same (monoclonal) antibody. The hybrid colonies are then screened to determine which colony produces the monoclonal antibody with the desired properties. This colony is then used to produce an essentially limitless supply of monoclonal antibody with the desired properties.

Molecular homogeneity and an abundant supply have revolutionized immunoassays. Monoclonal antibodies are not without disadvantages. Lower affinity and the special techniques required for production are the major disadvantages of monoclonal antibodies compared with polyclonal antibodies. However, low affinities can be overcome by the careful selection of high-affinity antibody-producing hybrids, and currently many monoclonal antibodies have affinities in the range 10^{-10}–10^{-12} L mol^{-1}.[40]

3.1.3 Antibody Fragments

In addition to the use of entire antibody molecules, fragments of antibodies may also be used in immunofluorescence experiments. Obviously, this must involve the use of the F_{ab} fragment (i.e. the variable, high-affinity binding region) rather than the F_c fragment (i.e. the constant, non-binding region) of the antibody. The main advantage in the use of antibody fragments rather than the entire antibody lies in the potential for increased delivery of antibody fragments to tumor sites in in vivo studies.

In vivo immunofluorescence studies are limited by the rate of antibody delivery to the binding site. In part, the rate of delivery is determined by the molecular weight of the antibody. The use of a relatively small fragment of an antibody (MW 30 000–40 000) rather than the entire antibody increases the rate of diffusion across capillary walls and into the body of the tumor.[41] This increased rate of delivery of labelled material to the site of interest obviously improves the changes of the fluorescent label being detected.

3.2 Choice of Fluorophore

The following characteristics should be taken into consideration in selecting the appropriate fluorophore:

1. *Coupling properties*. Coupling of a fluorophore label to immunological reagents (antibodies or their fragments) is an absolute prerequisite for immunofluorescence imaging. The fluorophore must therefore contain a reactive chemical group or be capable of derivatization to introduce a reactive group such as an isothiocyanate or ester group. Importantly, coupling should not affect the affinity or specificity of labeled antibodies or result in diminished signals.
2. *Sensitivity*. The sensitivity of immunoassays is highly dependent upon the properties of the fluorophore used. To be useful a fluorescent dye must have a high

quantum yield, i.e. the efficiency of conversion of light from the absorption wavelength to the emission wavelength must be high.

3. *Stability*. When linked to antibodies, fluorophores should be nonreactive to materials within the sample of interest. Labeled materials should also be stable for extended periods of time. Stable signal levels both from day to day and from experiment to experiment are required. Stable fluorophores decrease the need for frequent standardization. It should be noted that many fluorophores are sensitive to factors such as temperature and pH.
4. *Availability*. In order for a label to be accepted in research and in the clinical setting, a label should be readily available, conditions for labeling the antibody should be mild and easily optimized and labeling should be highly reproducible.
5. *Excitation and emission maxima*. The excitation and emission maxima of the fluorophore must also be considered. For example, absorption maxima should not be in spectral regions in which endogenous materials absorb strongly and emission maxima should not be in spectral regions where endogenous chromophores exhibit significant emission or absorption. In addition, the choice of absorption and emission maxima, and so fluorophore, depend upon whether the experiment is conducted in vitro of in vivo. For example, dyes such as fluorescein isothiocyanate (FITC) are commonly used in immunofluorescence studies in vitro. However, such dyes are excited by UV or visible light and fluoresce in the visible range of the electromagnetic spectrum. Unfortunately, visible light has limited penetration through tissues. This limited penetration obviously reduces the utility of the dyes for in vivo studies. In contrast, cyanine fluorochromes have emission maxima in the far-red spectral region. Blood and tissue are relatively transparent at such wavelengths, leading to enhanced transmission of light through tissues. Cyanine dyes such as Cy5 are therefore more suited to in vivo applications.

4 CHOICE OF DETECTION SYSTEMS

For in vivo detection of fluorescence using visible or NIR fluorophores, CCD camera systems equipped with the appropriate filters are the detector systems of choice.

4.1 Charge-coupled Device Cameras

The detection system of choice is the CCD camera. CCD cameras are available with two-dimensional arrays of detectors that allow two-dimensional images to be acquired at wavelengths in the visible and NIR regions. The advantages of such cameras include high quantum efficiency, low noise characteristics, flexibility and ruggedness. CCD cameras are available in many forms, with many types of sensing element to choose from. However, a two-dimensional array of silicon detectors will provide adequate response over the spectral range 200–1000 nm and allow the detection of most common chromophores.

4.2 Filter Selection

If the detection of fluorescence at a particular wavelength is required, then all other wavelengths must be prevented from reaching the imaging array. This wavelength selection can be achieved either with the use of fixed-wavelength filters such as band-pass or interference filters or by using LCTFs. Band-pass and interference filters allow light at predetermined wavelengths to be transmitted to the sensing element. If multiple wavelengths are of interest (i.e. if multiple antibodies labeled with different fluorophores are used), then a filter is required for each wavelength. The requirement for multiple filters is removed if LCTFs are used. As their name implies, LCTFs utilize liquid-crystal technology and have tunable transmission characteristics. The transmission window of the filter is set and changed electronically. Thus, images can be acquired at multiple wavelengths with no hardware adjustments. In addition to the time saving, LCTFs have the additional advantage that as no moving parts are involved in wavelength selection, registration of images at different wavelengths is excellent.

4.3 Typical Experimental Set-up

A typical set-up for an immunofluorescence imaging experiment is shown in Figure 2. A CCD camera equipped

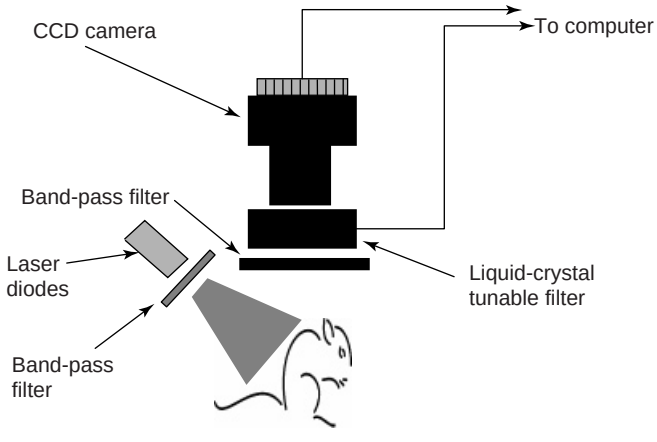

Figure 2 Typical pictorial representation of experimental set-up.

with a zoom lens for focusing is mounted above the sample area. A computer-controlled LCTF for wavelength selection is attached to the zoom lens. An illumination source is positioned so as to provide even illumination over the sample area The source may be a broad-band (white light) source or may be monochromatic radiation provided by a laser diode. If laser diodes are the source of choice, then precautions must be taken owing to the intense nature of the radiation produced. Although laser diodes are essentially monochromatic, side lobes (i.e. lower intensity emission on either side of the main emission line) may be evident. Although such side lobes are much weaker than emission at the main emission wavelength, the intensity may still be sufficient to result in transmission of some light through the LCTF. To eliminate this possibility, a band-pass filter should be inserted between the source and sample to remove wavelengths which are not of interest.

As an additional precaution when high-intensity sources such as laser diodes are used, a band-pass filter should be placed in front of the LCTF to eliminated scattered light at the frequency of the laser diode.

5 IMMUNOFLUORESCENCE IMAGING OF TUMORS

5.1 Qualitative Analysis of Images

The experimental set-up described above was used to study the interaction of a fragment of an antitumor antibody labeled with Cy5 with tumor cells implanted in athymic mice.[42] The antibody fragment was a fragment of a human monoclonal antibody to a cell surface antigen expressed by a wide variety of human cancer cells. The athymic mouse has two major advantages in such experiments. First, the athymic mouse is hairless. This alleviates potential problems due to scattering of light by hair, which would not only reduce the amount of excitation light reaching the area of interest but would also result in blurring at the fluorescence site. Second, the athymic mouse is immune compromised. As such it provides an excellent host for human tumor cells, which will not be rejected following implantation. The ability to monitor binding of the antibody fragment to human tumor cells is important if the antibody is of human origin. This raises an important, if obvious, issue in immunofluorescence experiments: the choice of model system requires thought. It is crucial that the antigen to which the antibody is raised is expressed in the system under investigation. For example the use of a carcinogen-treated animal as a model system may be inappropriate when utilizing human antibodies. The antigens expressed in tumors induced by carcinogens in animal models may not be the same as those expressed in human tumors. Use of a model system that hosts human tumor cells is preferable.

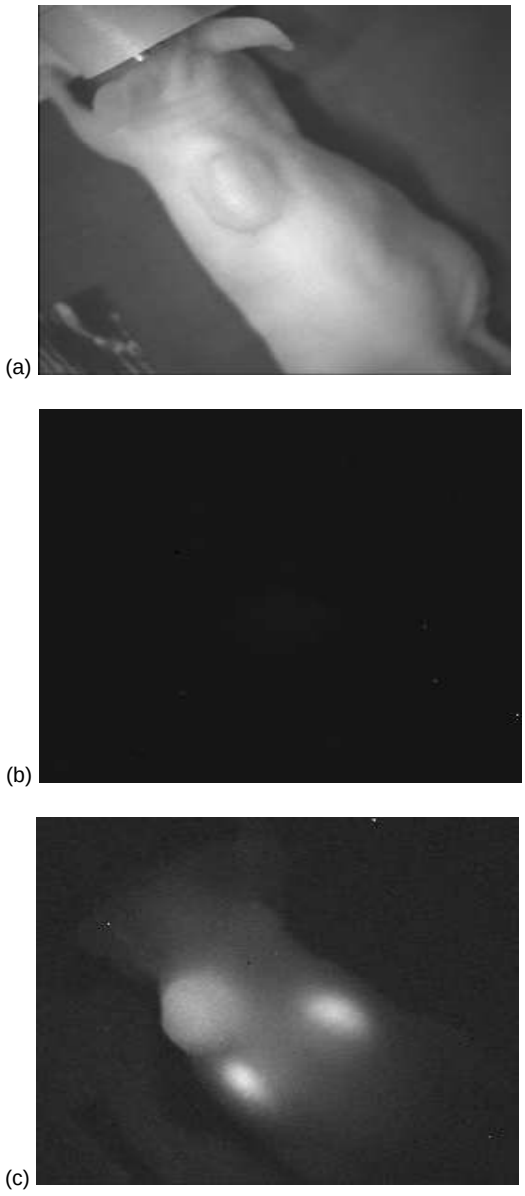

Figure 3 (a) Visible, (b) 670-nm and (c) 720-nm images of a tumor-bearing nude mouse 2 h postinjection of an antibody–Cy5.5.18 complex.

With the correct choice of imaging system (including illumination source and filters), antibody, dye and model, imaging can be successfully performed.[42] A visible image and images acquired at 670 nm (excitation maximum) and 720 nm (near the emission maximum) from a tumor-bearing mouse injected with the antibody–Cy5 complex are shown in Figure 3. The tumor can be clearly seen in the visible image (Figure 3a). Images acquired at 670 nm 2 h postinjection of the antibody–dye complex show no discernible features, as expected. However, images

acquired at 720 nm clearly show fluorescence at the site of the tumor and also in the region of the kidneys (Figure 3c), indicating uptake of the antibody–dye complex by the tumor and kidneys. No fluorescence is seen at the tumor site following injection of free dye or following injection of a labeled irrelevant (i.e. nonbinding) fragment. However, injection of both free dye and the labeled irrelevant fragment resulted in fluorescence in the region of the kidneys (not shown). Fluorescence was also noted in the bladder following injection of free dye, antibody–dye and irrelevant antibody–dye, due to accumulation of the dye prior to excretion in the urine. Fluorescence was not detected in the bladder 24 h postinjection.

The kinetics of antibody binding can be qualitatively monitored by examining images of a tumor-bearing mouse taken at a number of time points. Figure 4(a–g) shows visible and fluorescence images acquired at $t = 0$ (preinjection), 2, 6, 12, 24 and 48 h postinjection of antibody–dye. As expected, no fluorescence is seen at the site of the tumor or in the kidneys preinjection. However, fluorescence is clearly seen in both the tumor and kidneys 2 h postinjection. Fluorescence in the tumor peaks 24 h postinjection in this animal, whereas fluorescence in the kidneys peaks 6 h postinjection. At 48 h postinjection, fluorescence in the kidneys has largely disappeared, but fluorescence is still apparent at the tumor site. Peak fluorescence in the kidneys is much greater than the peak fluorescence in the tumor.

5.2 Quantitative Analysis of Images

Visual inspection of the fluorescence signal arising from the tumors and kidneys presents problems. First, it is difficult to detect subtle differences in fluorescence intensity by eye. Second, the perceived intensity of fluorescence in images depends upon scaling of the images. Variations

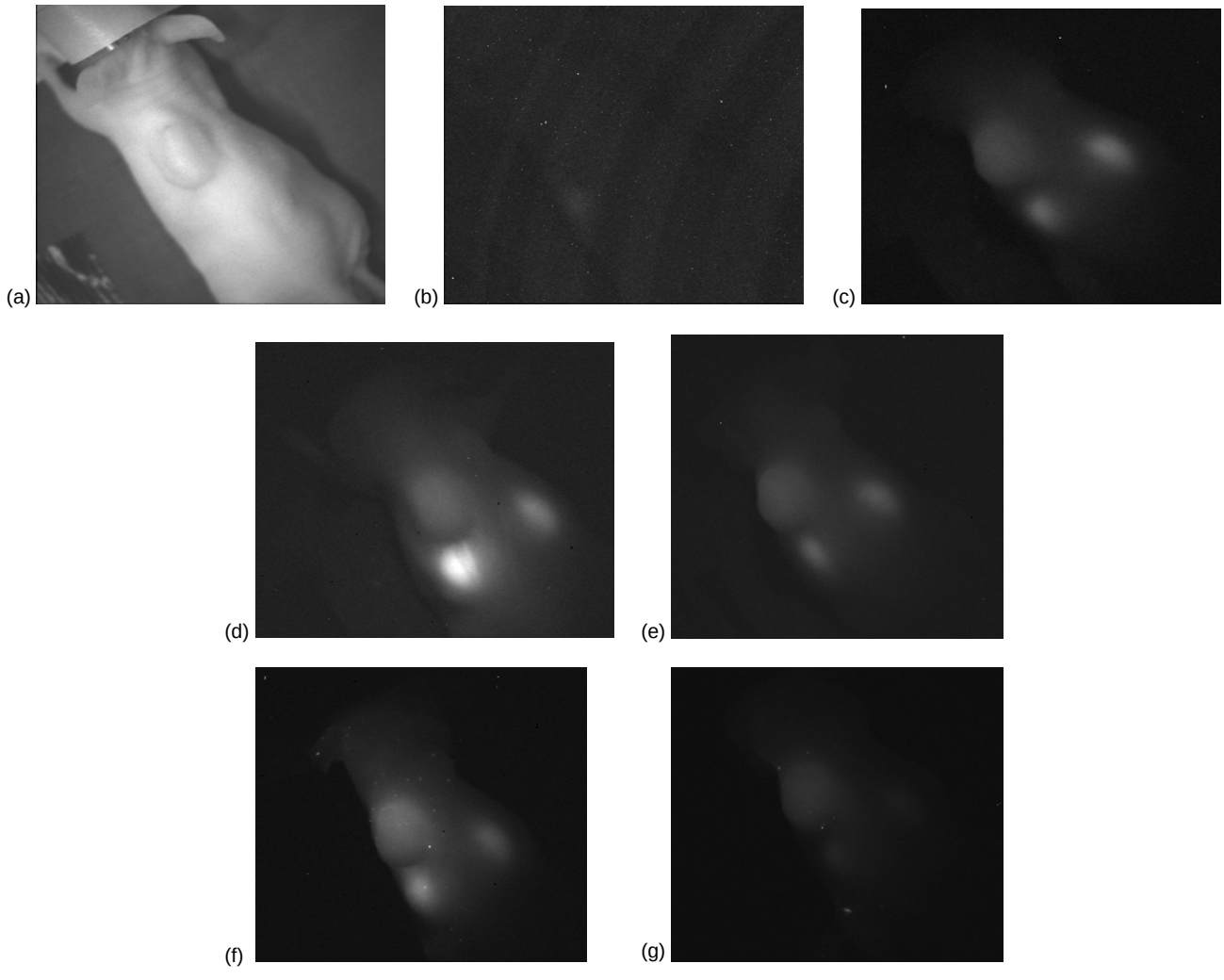

Figure 4 (a) Visible and 720-nm images of a tumor-bearing nude mouse at (b) 0, (c) 2, (d) 6, (e) 12, (f) 24 and (g) 48 h postinjection of an antibody–Cy5.5.18 complex.

in scaling between images can therefore lead to artefactual changes in intensity that may mask true changes in fluorescence intensity. A more rigorous analysis requires a quantitative assessment of fluorescence intensity. To obtain qualitative information relating to fluorescence intensity in each image (composed of 256×256 pixels), the intensity at each pixel through the region of the image showing the maximum fluorescence intensity was extracted.[42] This is illustrated in Figure 5(a–f). The fluorescence intensity at each pixel along the line shown in each image is plotted as a function of position for each image. Pixel 0 corresponds to the pixel at the top of the image and pixel 256 corresponds to the pixel at the bottom of the image. This sequence of images clearly shows accumulation of the antibody in the both the tumor and kidney beginning at 2 h postinjection. The second, sharp peak observed in Figure 5(b) arises from fluorescence in the left kidney, part of which is traversed by the line of pixels extracted. The intensity of fluorescence from the tumor increased over time, peaking 24 h postinjection. The fluorescence then rapidly declines, falling to about 5% of the peak fluorescence 48 h postinjection. Fluorescence intensity persists at about 5% of peak intensity 72 h postinjection.

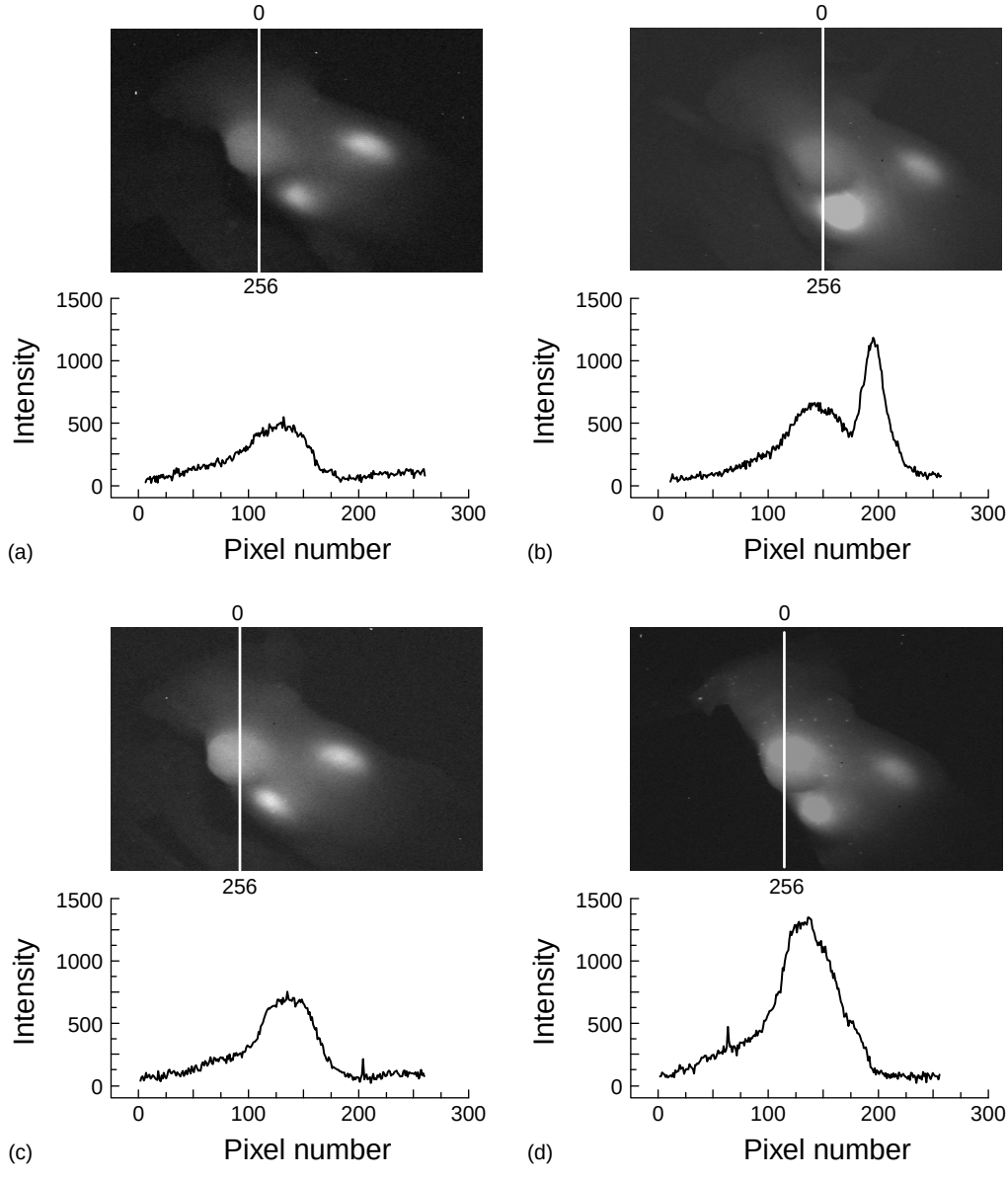

Figure 5 Images acquired at 720 nm from a tumor-bearing mouse (a) 2, (b) 6, (c) 12, (d) 24, (e) 48 and (f) 72 h following injection of an antibody–Cy5.5.18 complex. The fluorescence intensity at each pixel on the line through the region of greatest fluorescence intensity for each image (shown in white) is plotted below each image.

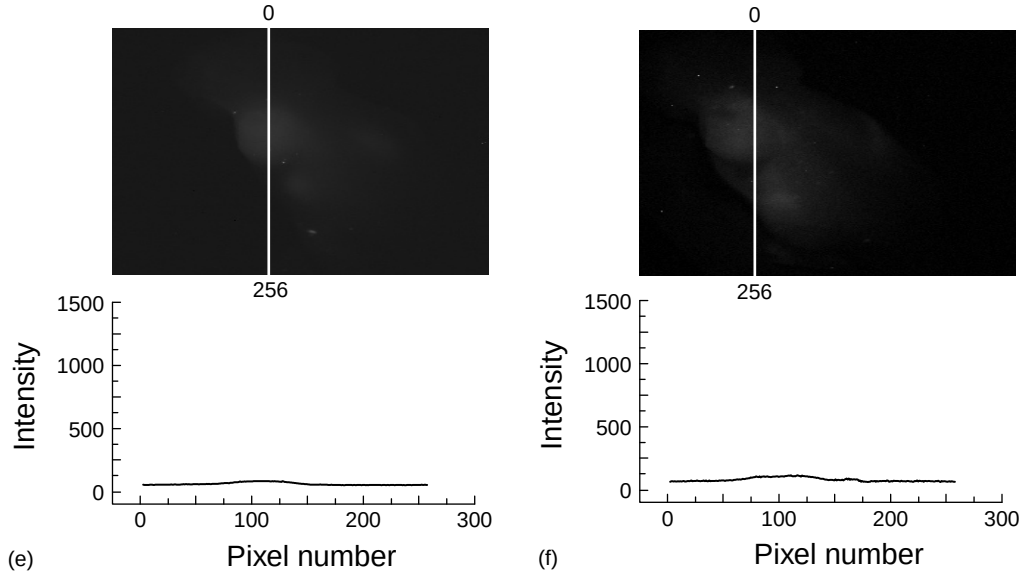

Figure 5 (*Continued*)

5.3 Statistical Analysis of Images

The method discussed above allows the quantitation of fluorescence as a function of time in individual animals but does not allow data from different animals to be pooled for statistical analysis. To allow comparison between animals, data must be normalized in some manner. Quantitative data from tumors may be normalized by scaling all data from each animal with respect to data obtained at a predetermined time point. For example, data may be normalized to images obtained from each animal 2 h postinjection. In other words, all data at 6 h postinjection are expressed as a percentage of the fluorescence intensity seen at 2 h postinjection in the same mouse. The mean difference in fluorescence intensity compared with 2 h postinjection may then be calculated for each time point. The mean and standard deviation difference in fluorescence relative to 2 h postinjection may then be calculated for each time point and analyzed using straightforward tests such as the paired difference test to assess statistically significant differences.

For the experiments described above the mean decrease in fluorescence seen 6 h postinjection (40%) was found to be significant ($p < 0.05$). Fluorescence continued to decrease over time, reaching a minimum 72 h postinjection, with a mean decrease in fluorescence of 75%.

Figure 6(a–f) shows a series of images obtained from a control mouse, demonstrating uptake of the antibody–Cy5 complex by the kidneys as a function of time. In this example, the left kidney is masked and only the right kidney is monitored. Fluorescence from the complex can be seen at all time points measured. Uptake of the complex peaked at 6–12 h (Figure 6b, c) and then diminished with time, reaching a minimum at 72 h postinjection.

Data from kidneys were also normalized and analyzed using the paired difference test. Peak fluorescence occurred 6 h postinjection in the majority of animals and so data were normalized with respect to measurements made 6 h postinjection. The mean difference in fluorescence 2 h postinjection compared with 6 h postinjection (10%) was not significant. A nonsignificant decrease of only 24% ($p > 0.05$) was seen 12 h postinjection. However, fluorescence significantly decreased 24, 48 and 72 h postinjection ($p < 0.05$).

6 SUMMARY AND FUTURE PROSPECTS

A number of fluorescence imaging techniques show diagnostic promise. Imaging endogenous fluorescence has been proposed as a method for cancer diagnosis.[7] Typically such studies rely on the decrease in the NADH/NAD^+ ratio observed in malignant tissues, which results in a decreased fluorescence signal in malignant tissue. Unfortunately, tissue autofluorescence (whether from NADH or other endogenous chromophores) is relatively weak and poor contrast between malignant and normal tissue is seen. Contrast may be enhanced with the addition of fluorescent materials that are selectively accumulated by malignant cells, such as fluorescein or porphyrin derivatives.[9,16] Although fluorescence from these chromophores occurs in the visible region of the spectrum, the limited penetration of light at these wavelengths restricts the use of fluorescence techniques

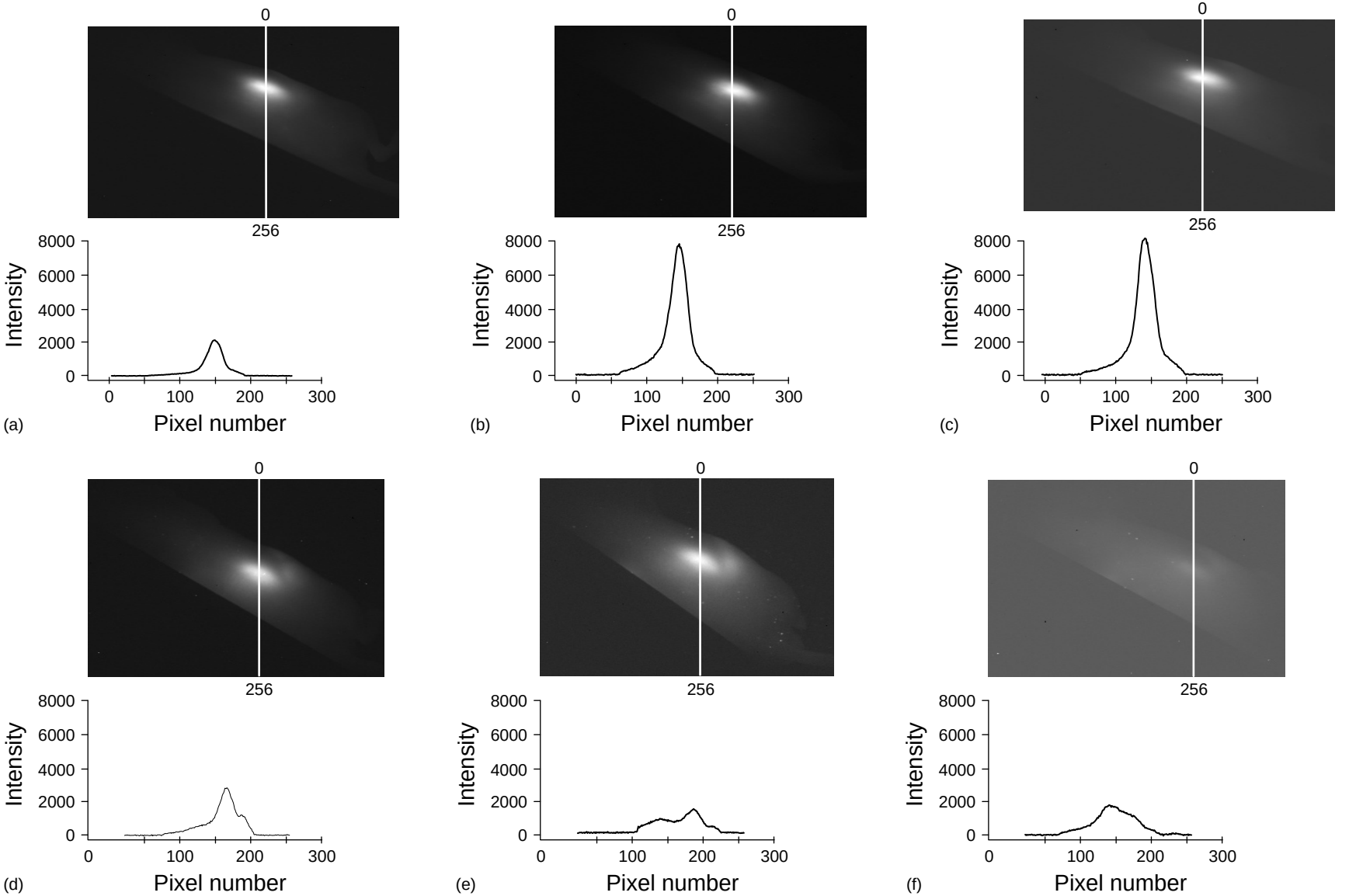

Figure 6 Images acquired at 720 nm from a control mouse (a) 2, (b) 6, (c) 12, (d) 24, (e) 48 and (f) 72 h following injection of an antibody–Cy5.5.18 complex. The fluorescence intensity at each pixel on the line through the region of greatest fluorescence intensity for each image (shown in white) is plotted below each image.

utilizing these chromophores to superficial phenomena. However, many potential applications still exist. For example, monitoring fluorescein fluorescence during surgery may allow resection margins to be clearly delineated.

Other exogenous chromophores that may have diagnostic utility include ICG. Techniques based upon visualization of the distribution of ICG fluorescence (i.e. choroidal angiography) are already prominent in ophthalmology.[18,22–24] As discussed above, ICG fluorescence imaging may also find a useful niche in monitoring of burns and transplant tissue. Two features are common to each of these applications: distribution of fluorescence at an accessible surface is monitored and the main phenomena under investigations related to microvascular circulation. Obviously the requirement for an accessible surface limits non-invasive diagnostic applications. However, monitoring of vascular parameters during cardiac surgery presents exciting opportunities. For example, low oxygen levels (e.g. during bypass surgery) in the heart can result in alterations in microvascular permeability. As ICG is largely bound to serum albumin, it should not be seen in extravascular spaces in normal hearts and ICG fluorescence will be seen only in blood vessels. However, increased permeability will allow albumin to diffuse into the extravascular spaces, and diffuse fluorescence across the surface of the heart will be seen. In principle, the infusion of polymers (e.g. dextrans) of various molecular weights labeled with dyes that fluoresce at different wavelengths will allow the assessment of the porosity of capillary beds in such systems.

Imaging ICG fluorescence may also be useful in detecting artherosclerotic plaques in superficial vessels such as the carotid artery in the neck, a major site of plaque formation. Regions of plaque formation should in principle appear as areas of vessels with a narrow fluorescence profile.

ICG is largely cleared from the blood by the liver. However, it has been demonstrated that areas of the liver affected by hepatocellular carcinoma show decreased uptake of ICG.[43] Clearly this reduced uptake of ICG has diagnostic implications, and again may prove useful for determining tumor boundaries during surgery.

Immunofluorescence techniques have the potential to provide unmatched sensitivity and specificity.[26,27,29,42] The unique nature of antibody–antigen interactions ensures specific delivery of the fluorophore to the site of interest. Most studies to date have been limited, but the potential should be apparent. The specific interaction of labeled antibodies with antigens means that the fluorophore persists in the body for a prolonged period of time (days). Following a single injection of labeled antibody, repeated measurements on the same site over the course of hours or days allow kinetic information to be readily obtained. In principle this means that the effect of therapeutic intervention, i.e. radiation therapy, chemotherapy, etc., can be monitored. This would be achieved by comparing an initial fluorescence image acquired immediately prior to treatment with images obtained as treatment progressed.

The clinical potential of fluorescence imaging techniques is only now becoming apparent. Advances in instrumentation and continued development of new chromophores are opening up new possibilities and it is surely only a matter of time before fluorescence imaging techniques become established clinical tools.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
FITC	Fluorescein Isothiocyanate
ICG	Indocyanine Green
Ig	Immunoglobulin
LCTF	Liquid-crystal Tunable Filter
NADH	Nicotinamide Adenine Dinucleotide (reduced form)
NIR	Near-infrared
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo • Magnetic Resonance Imaging, Functional • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors • Infrared Spectroscopy of Biological Applications

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis

REFERENCES

1. A.M. Kluftinger, N.L. Davis, N.F. Quenville, S. Lam, J. Hung, B. Palcic, 'Detection of Squamous Cell Cancer and Pre-cancerous Lesions by Imaging of Tissue Autofluorescence in the Hamster Cheek Pouch Model', *Surg. Oncol.*, **1**, 183–188 (1992).
2. S.C. Lam, J. MacAulay, J. Hung, J. LeRiche, A.E. Profio, B. Palcic, 'Detection of Dysplasia and Carcinoma In Situ with a Lung Imaging Fluorescence Endoscope Device', *J. Thorac. Cardiovasc. Surg.*, **105**, 1035–1040 (1993).
3. I. Pathak, N.L. Davis, Y.N. Hsiang, N.F. Quenville, B. Palcic, 'Detection of Squamous Neoplasia by Fluorescence Imaging Comparing Porfimer Sodium Fluorescence to Tissue Autofluorescence in the Hamster Cheek-pouch Model', *Am. J. Surg.*, **170**, 423–426 (1995).
4. A. Pelegrin, S. Folli, F. Buchegger, J.P. Mach, G. Wagnieres, H. van-den-Bergh, 'Antibody–Fluorescein Conjugates for Photoimmunodiagnosis of Human Colon Carcinoma in Nude Mice', *Cancer*, **67**, 2529–2537 (1991).
5. M.W. Wessendorf, T.C. Brelje, 'Which Fluorophore is Brightest? A Comparison of the Staining Obtained Using Fluorescein, Tetramethylrhodamine, Lissamine Rhodamine, Texas Red, and Cyanine 3.18', *Histochemistry*, **98**, 81–85 (1992).
6. F.N. Jiang, D.J. Liu, H. Neyndorff, M. Chester, S.Y. Jiang, J.G. Levy, 'Photodynamic Killing of Human Squamous Cell Carcinoma Cells Using a Monoclonal Antibody–Photosensitizer Conjugate', *J. Natl. Cancer. Inst.*, **83**, 1218–1225 (1991).
7. F.N. Jiang, A.M. Richter, A.K. Jain, J.G. Levy, C. Smits, 'Biodistribution of a Benzoporphyrin Derivative-monoclonal Antibody Conjugate in A549-tumor-bearing Nude Mice', *Biotechnol. Ther.*, **4**, 43–61 (1993).
8. R.B. Mujumdar, L.A. Ernst, S.R. Mujumdar, C.J. Lewis, A.S. Waggoner, 'Cyanine Dye Labeling Reagents: Sulfoindocyanine Succinimidyl Esters', *Bioconjug. Chem.*, **4**, 105–111 (1993).
9. V.V. Lazarev, R.A. Roth, Y. Kanzakevich, J. Hang, 'Detection of Premalignant Oral Lesions in Hamsters with an Endoscopic Fluorescence Imaging System', *Cancer*, **85**, 1421–1429 (1999).
10. A.W. Hemming, N.L. Davis, B. Dubois, N.F. Quenville, R.J. Finley, 'Photodynamic Therapy of Squamous Cell Carcinoma. An Evaluation of a New Photosensitizing Agent, Benzoporphyrin Derivative and New Photoimmunoconjugate', *Surg. Oncol.*, **2**, 187–196 (1993).
11. E.S. Andersson, J. Ankerst, J. Johansson, K. Svanberg, S. Svanberg, 'Laser-induced Fluorescence in Malignant and Normal Tissue of Rats Injected with Benzoporphyrin Derivative', *Photochem. Photobiol.*, **57**, 978–983 (1993).
12. L. Baert, R. Berg, B. Van-Damme, M.A. D'Hallewin, J. Johansson, K. Svanberg, S. Svanberg, 'Clinical Fluorescence Diagnosis of Human Bladder Carcinoma Following Low-dose Photofrin Injection', *Urology*, **41**, 322–330 (1993).
13. T.J. Dougherty, 'Photodynamic Therapy', *Photochem. Photobiol.*, **58**, 895–900 (1993).
14. D. Mew, C.K. Wat, G.H. Towers, J.G. Levy, 'Photoimmunotherapy: Treatment of Animal Tumors with Tumor-specific Monoclonal Antibody–Hematoporphyrin Conjugates', *J. Immunol.*, **130**, 1473–1477 (1983).
15. H.W. Pogrebniak, W. Matthews, C. Black, A. Russo, J.B. Mitchell, P. Smith, J.A. Roth, H.I. Pass, 'Targetted Phototherapy with Sensitizer–Monoclonal Antibody Conjugate and Light', *Surg. Oncol.*, **2**, 31–42 (1993).
16. M. Korbelik, G. Krosl, 'Photofrin Accumulation in Malignant and Host Cell Populations of Various Tumours', *Br. J. Cancer*, **73**, 506–513 (1996).
17. O.V. Braginskaja, V.V. Lazarev, I.N. Pershina, K.V. Petrov, L.B. Rubin, O.V. Tikhonova, 'Sodium Fluorescein Accumulation in Cultured Cells', *Gen. Physiol. Biophys.*, **12**, 453–464 (1993).
18. H. Kuck, W. Inhoffen, U. Schneider, I. Kreissig, 'Diagnosis of Occult Subretinal Neovascularization in Age-related Macular Degeneration by Infrared Scanning Laser Videoangiography', *Retina*, **13**, 36–39 (1993).
19. J.P. Lund, 'The Dye Dilution Method for Measurement of Cardiac Output', *Eur. Heart J.*, **11**(Suppl.I), 6–12 (1990).
20. G. Moneta, M. Brulisauer, K. Jager, A. Bollinger, 'Infrared Fluorescence Videomicroscopy of Skin Capillaries with Indocyanine Green', *Int. J. Microcirc. Clin. Exp.*, **6**, 25–34 (1987).
21. H. Schad, M. Haider, H. Brechtelsbauer, 'Determination of Plasma Volume with Indocyanine Green', *Anaesthesist*, **36**, 608–614 (1987).
22. M. Brulisauer, G. Moneta, K. Jager, A. Bollinger, 'Infrared Fluorescence Videomicroscopy with Indocyanine Green (Cardiogreen)', *Adv. Exp. Med. Biol.*, **220**, 219–221 (1987).
23. L.A. Yannuzzi, J.A. Sorenson, D.R. Guyer, J.S. Lakter, B. Chang, D. Orlock, 'Indocyanine Green Videoangiography: Current Status', *Eur. J. Ophthalmol.*, **4**, 69–81 (1994).
24. J. Lim, R.W. Flower, 'Indocyanine Green Angiography', *Int. Ophthalmol. Clin.*, **35**, 59–70 (1995).
25. M.R. Jerath, K.T. Schomacker, R.L. Sheridan, N.S. Ishioka, 'Burn Wound Assessment in Porcine Skin Using Indocyanine Green Fluorescence', *J. Trauma Injury Infect. Crit. Care.*, **46**, 1085–1088 (1999).
26. B. Ballou, G.W. Fisher, A.S. Waggoner, D.L. Farkas, J.M. Reiland, R. Jaffe, R.B. Mujumdar, S.R. Mujumdar, T.R. Hakala, 'Tumor Labeling In Vivo Using Cyanine-conjugated Monoclonal Antibodies', *Cancer Immunol. Immunother.*, **41**, 257–263 (1995).
27. B. Ballou, J.M. Reiland, G. Levine, R.J. Taylor, W.C. Shen, H.J. Ryser, D. Solter, T.R. Hakala, 'Tumor Location and Drug Targeting Using a Monoclonal Antibody

(Anti-SSEA-1) and Antigen-binding Fragments', *J. Surg. Oncol.*, **31**, 1–12 (1986).
28. E.S. Andersson, J. Johansson, K. Svanberg, S. Svanberg, 'Fluorescence Imaging and Point Measurements of Tissue: Applications to the Demarcation of Malignant Tumors and Atherosclerotic Lesions from Normal Tissue', *Photochem. Photobiol.*, **53**, 807–814 (1999).
29. S. Folli, P. Westermann, D. Braichotte, A. Pelegrin, G. Wagnieres, H. van-den-Bergh, J.P. Mach, 'Antibody–Indocyanin Conjugates for Immunophotodetection of Human Squamous Cell Carcinoma in Nude Mice', *Cancer Res.*, **54**, 2643–2649 (1994).
30. J. Little, C.M. Fenoglio-Preiser, 'Tumor Immunodiagnosis', in *Comprehensive Textbook of Oncology*, eds. A.R. Moossa, S.C. Schimpff, C.M. Robson, Williams & Wilkins, Baltimore, 247–260, 1991.
31. M.W. Arnold, S. Schneebaum, A. Berens, C. Mojzisik, G. Hinkle, E.W. Martin, Jr, 'Radioimmunoguided Surgery Challenges Traditional Decision Making in Patients with Primary Colorectal Cancer', *Surgery*, **112**, 624–629 (1992).
32. B. Ballou, R. Jaffe, R.J. Taylor, D. Solter, T.R. Hakala, 'Tumor Radioimmunolocation: Differential Antibody Retention by Antigenic Normal Tissue and Tumor', *J. Immunol.*, **132**, 2111–2116 (1984).
33. D.M. Goldenberg, S.M. Larson, R.A. Reisfeld, J. Schlom, 'Targeting Cancer with Radiolabeled Antibodies', *Immunol. Today*, **16**, 261–264 (1995).
34. D.J. Buchsbaum, 'Experimental Approaches to Increase Radiolabeled Antibody Localization in Tumors', *Cancer Res.*, **55**, 5729s–5732s (1995).
35. D.J. Buchsbaum, 'Experimental Tumor Targeting with Radiolabeled Ligands', *Cancer*, **80**, 2371–2377 (1997).
36. L.M. Cobb, J.L. Humm, 'Radioimmunotherapy of Malignancy Using Antibody Targeted Radionuclides', *Br. J. Cancer*, **54**, 863–869 (1999).
37. B.D. Mann, M.B. Cohen, R.E. Saxton, D.L. Morton, W.F. Benedict, E.L. Korn, L. Spolter, L.S. Graham, C.C. Chang, M.W. Burk, 'Imaging of Human Tumor Xenografts in Nude Mice with Radiolabeled Monoclonal Antibodies. Limitations of Specificity Due to Nonspecific Uptake of Antibody', *Cancer*, **54**, 1318–1327 (1984).
38. C.A. Janeway, P. Travers, *Immunobiology*, Garland Publishing, New York, 1:13–1:15, 1997.
39. E. Benjamini, S. Leskowitz, *Immunology. A Short Course*, Alan R. Liss, New York, 5–7, 1998.
40. I.A. Hemmila, 'Antibodies as Analytical Reagents', in *Applications of Fluorescence in Immunoassays*, eds. J.D. Winefordner, I.M. Kolthoff, John Wiley & Sons, New York, 1–343, 1991.
41. R.K. Jain, 'Physiological Barriers to Delivery of Monoclonal Antibodies and Other Macromolecules in Tumors', *Cancer Res.*, **50**, 814s–819s (1990).
42. B. Ramjiawan, P. Maiti, A. Aftanas, H. Kaplan, D. Fast, H.H. Mantsch, M. Jackson, 'Non-invasive Localization of Tumors by Immunofluorescence Imaging Using a Single Chain Fv Fragment of a Human Monoclonal Antibody with Broad Cancer Specificity', *Cancer*, in press.
43. Y. Kimura, T. Higashi, N. Kuwahara, K. Nouso, S. Ohguchi, N. Hino, M. Tanimizu, H. Nakatsukasa, K. Tobe, T. Tsuji, 'Usefulness of Indocyanine Green Injection During Ultrasound-guided Liver Biopsy for the Diagnosis of Small Hepatocellular Carcinoma', *Acta Med. Okayama*, **50**, 255–259 (1996).

Fluorescence Spectroscopy In Vivo

Nirmala Ramanujam
University of Wisconsin, Madison, USA

1 Fluorescence Spectroscopy 1
1.1 Introduction 1
1.2 Principles and Definitions 2
2 Fluorophores 4
2.1 Endogenous Fluorophores 4
2.2 Exogenous Fluorophores 6
3 Fluorescence Spectroscopy of Turbid Media 7
3.1 Fluorescence Spectroscopy of an Optically Dilute, Homogeneous Medium 7
3.2 Fluorescence Spectroscopy of Turbid Media such as Tissue 7
3.3 Deconvolution of Absorption and Scattering from Tissue Fluorescence Emission Spectra 9
3.4 Turbid Tissue-simulating Phantoms for Fluorescence Spectroscopy of Tissue 11
4 Instrumentation 11
4.1 Light Sources 11
4.2 Illumination and Collection of Light 12
4.3 Monochromators and Spectrographs 13
4.4 Detectors 14
4.5 Signal-to-noise Ratio Analysis of an Instrument Used for Fluorescence Spectroscopy of Tissue 15
4.6 Calculation of Tissue Fluorescence Efficiency 15
5 Clinical Applications 16
5.1 Neoplasia 17
5.2 Atherosclerosis 19
6 Clinical Instruments 19
6.1 Single-pixel, Three-excitation-wavelength Fluorimeter 19
6.2 Single-pixel, Excitation–Emission Matrix System 21
6.3 Fluorescence Imaging Systems 23
7 Methods of Analysis 23
7.1 Statistically Based Model 24
7.2 Physically Based Models 25
7.3 Effect of Excitation and Emission Geometry on Fluorescence Emission Spectra of Turbid Media 29
8 Future Perspectives 30
Acknowledgments 30
Abbreviations and Acronyms 31
Related Articles 31
References 31

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Diagnostic techniques based on optical spectroscopy have the potential to link the biochemical and morphological properties of tissues to individual patient care. In particular, these techniques are fast, noninvasive and quantitative. Furthermore, they can be used to elucidate key tissue features, such as the cellular metabolic rate, vascularity, intravascular oxygenation and alterations in tissue morphology. These tissue features can be interpreted to shed light on a variety of clinical problems, such as precancerous and cancerous growth and atherosclerosis. The goal of this report is to review the development and application of optical spectroscopy in the ultraviolet (UV) and visible (VIS) spectral regions, as a diagnostic tool in clinical applications. A particular emphasis is placed on steady-state, UV/VIS fluorescence spectroscopy for the detection of precancers and cancers, in vivo.

1 FLUORESCENCE SPECTROSCOPY

1.1 Introduction

Diagnostic techniques based on optical spectroscopy have the potential to link the biochemical and morphological properties of tissues to individual patient care. In particular, these techniques are fast, noninvasive and quantitative. Furthermore, they can be used to elucidate key tissue features, such as the cellular metabolic rate, vascularity, intravascular oxygenation and alterations in tissue morphology. These tissue features can be interpreted to shed light on a variety of clinical problems, such as precancerous and cancerous growth and atherosclerosis.[1] If applied successfully, optical spectroscopy has the potential to represent an important step forward toward advances in diagnostic and therapeutic medical applications.

Spectroscopy is the study of the interaction of electromagnetic radiation with matter. There are three aspects to a spectroscopic measurement: irradiation of a sample with electromagnetic radiation; measurement of the absorption, spontaneous emission (fluorescence, phosphorescence) and/or scattering (Rayleigh elastic

scattering, Raman inelastic scattering) from the sample; and analysis and interpretation of these measurements. Detailed study of absorption, spontaneous emission and scattering provides information that can be classified broadly as analytical, structural, dynamic, and energetic.[2]

Optical spectroscopy deals with interactions of electromagnetic radiation with matter that occur at the UV, VIS, near-infrared (NIR) and infrared (IR) wavelengths. In the UV/VIS spectral regions (<700 nm), light can penetrate only superficial tissue volumes (a few hundred microns to a millimeter in depth) due to the fact that this biological medium is highly absorbing.[3] However, in the NIR spectral region (700–900 nm) tissue is generally less absorbing and, furthermore, the number of elastic scattering events of light in tissue is approximately two orders of magnitude greater than the number of absorption events.[4] This enables the light to propagate through tissue volumes that are up to several centimeters in depth.

In the UV/VIS spectral regions, absorption and fluorescence spectroscopy have been explored extensively as diagnostic tools for precancer and cancer detection in the surface epithelia of various organ sites (colon, cervix, bronchus, lung, bladder, brain, esophagus, head and neck, skin, bile duct, breast and stomach tissues)[5–12] and for the characterization of atherosclerosis in the surface of artery vessel walls.[5,6,11] Recently, fluorescence spectroscopy was also used to detect Alzheimer's disease of the brain in vitro.[13] Elastic scattering spectroscopy in this spectral region also has been evaluated for the detection of precancers and cancers in the surface epithelia of several organ sites, but to a much lesser extent.[14]

NIR absorption,[4] fluorescence and phosphorescence[15] spectroscopies have been used to interrogate larger tissue volumes. In particular, NIR absorption has been used to detect brain bleeds,[16] brain oxygenation,[17] functional activity of the brain,[18,19] bioenergetics of skeletal muscle (e.g. in the case of genetic disease of mitochondrial function)[20] and breast tumors.[21] NIR correlation spectroscopy, which monitors fluctuations in the elastic scatter intensity, is emerging as a potential diagnostic tool to detect blood flow deep within thick tissue.[22]

To date, most clinical applications have concentrated on absorption, fluorescence and elastic scattering spectroscopies, because these measurements can be obtained with a good signal-to-noise ratio in reasonably short integration times. However, with advances in illumination and detection technologies, NIR Raman scattering, which is a relatively weak phenomenon, is emerging as a potential diagnostic tool for precancer and cancer detection in the surface epithelia of various organ sites,[23] for the characterization of atherosclerosis on the surface of artery vessel walls[24] and for glucose monitoring.[25]

The goal of this article is to review the development and application of UV/VIS optical spectroscopy as a diagnostic tool in clinical applications. Particular emphasis is placed on steady-state UV/VIS fluorescence spectroscopy for the detection of precancers and cancers in vivo. In this section the principles and definitions that are related to the phenomenon of fluorescence are summarized. In the next section the endogenous fluorescence properties of molecules present in cells and tissues and the fluorescence properties of exogenous molecules, which can be used as contrast agents, are presented. This is followed in section 3 by a discussion of the effect of absorption and scattering on fluorescence spectroscopy of turbid media such as tissue, and methods for deconvolving this effect. Next, the instrumentation requirements for fluorescence spectroscopy of tissues are reviewed in section 4. In the last several sections, a summary of the clinical applications of fluorescence spectroscopy, particularly precancer and cancer detection (section 5), a review of representative instruments (section 6) and methods of analysis used (section 7) are presented. With respect to methods of analysis, statistical models, which are used solely for the purpose of discriminating diseased from nondiseased tissues based on the spectral information, as well as physical models, which have the potential to elucidate the biochemical/morphological basis for the spectral differences observed, are presented. In the final part of section 7, the effect of the illumination and collection geometry on fluorescence spectroscopy of tissues is discussed and approaches to resolve it are presented. This article concludes with a brief discussion of future perspectives in section 8.

1.2 Principles and Definitions

1.2.1 Probing Energy Levels with Electromagnetic Radiation

Optical spectroscopy probes the energy levels of a molecule.[2] The energy level of a molecule is defined as its characteristic state, which is related to the molecular structure of the molecule and to the energetics and dynamics of any chemical processes that the molecule may undergo. The ground state of a molecule is defined as the state of lowest energy. States of higher energy are called excited states. A molecule possesses several distinct reservoirs of energy levels, including electronic, vibrational, rotational, translational and those associated with nuclear and electron spin.[2] In the optical regime, the energy levels of interest are those that are associated with vibrational and electronic transitions. The separation between vibrational energy levels is determined by the mass of the atoms and the flexibility of the chemical bonds joining them. The separation between electronic

energy levels, which is greater, occurs when electrons are displaced from one region of a molecule to another.

At any finite temperature, the molecules will be distributed among the energy levels available to them because of thermal agitation.[2] The exact distribution will depend on the temperature (T) and on the separation between the energy levels (ΔE) in the energy ladder. At a given temperature, the number of molecules in an upper level (n_{upper}) relative to that in a lower level (n_{lower}) is given by the Boltzmann distribution, as shown in Equation (1):

$$\frac{n_{\text{upper}}}{n_{\text{lower}}} = \exp\left(\frac{-\Delta E}{kT}\right) \tag{1}$$

where k is the Boltzmann constant ($1.38 \times 10^{-23}\,\text{J K}^{-1}$). When electromagnetic radiation is applied to a molecule, it is just as likely to cause transitions from a higher to a lower energy level as it is to cause transitions from a lower to a higher energy level. Consequently, net absorption or transition to a higher energy level can occur only if the difference between the populations of the energy levels concerned is significant, with the lower one being significantly higher. Calculation of the population of energy levels at room temperature[2] has shown that for vibrational energy level spacings the ratio $n_{\text{upper}}/n_{\text{lower}}$ is $\sim 10^{-3}$ and for electronic energy level spacings it is 10^{-21}.

1.2.2 The Fluorescence Phenomenon

Figure 1 displays an energy level diagram with ground (S_0) and excited (S_1) electronic states as well as vibrational energy levels within each electronic state of a molecule.[26] When a molecule is illuminated at an excitation wavelength that lies within the absorption spectrum of that molecule, it will absorb the energy and be activated from its ground state (S_0) to an excited singlet state (S_1), with the electron in the same spin as its ground state. The molecule can then relax back from the excited state to the ground state by generating energy either nonradiatively or radiatively, depending on the local environment. In a nonradiative transition, relaxation occurs by thermal generation (dashed arrows). In a radiative transition, relaxation occurs via fluorescence at specific emission wavelengths (solid arrow). Fluorescence generation occurs in three steps: thermal equilibrium is achieved rapidly as the electron makes a nonradiative transition to the lowest vibrational level of the first excited state; the electron then makes a radiative transition to a vibrational level of the ground state; and finally the electron makes a nonradiative transition to the lowest vibrational level of the ground state. When there is inter-system crossing, in which the spin of the electron is flipped in the excited state, the time for radiative transition from the excited state to ground state is longer because the transition must occur with a spin change. This excited state is termed the triplet state (not shown). Radiative transition from the excited triplet state is termed phosphorescence. Tissue absorption, fluorescence and phosphorescence monitor changes in electronic energy levels, to provide biochemical information from biological molecules.

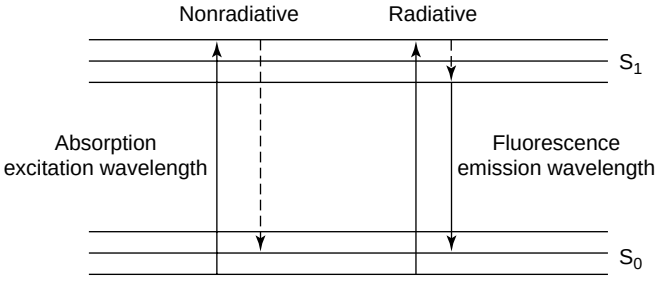

Figure 1 Energy level diagram illustrating the phenomena of absorption and fluorescence of a molecule.

The phenomenon of fluorescence displays several general characteristics for a particular biological molecule.[26] First, due to the losses in energy between absorption and emission that occur as a result of nonradiative transitions, fluorescence occurs at emission wavelengths that are always red-shifted relative to the excitation wavelength. Second, the emission wavelengths are independent of the excitation wavelength. Third, the fluorescence spectrum of a biological molecule is generally a mirror image of its absorption spectrum.

The fluorescence of a biological molecule is characterized by its quantum yield and its lifetime.[26] The quantum yield is simply the ratio of the number of photons emitted to the number absorbed. The lifetime is defined as the average time the biological molecule spends in the excited state prior to return to the ground state. The fluorescence quantum yield and lifetime are modified by a number of factors that can increase or decrease the energy losses. For example, a molecule may be nonfluorescent as a result of a large rate of nonradiative decay (thermal generation) or a slow rate of radiative decay (fluorescence generation).

Fluorescence spectroscopy is the measurement and analysis of various features that are related to the fluorescence quantum yield and/or lifetime of a biological molecule. The fluorescence intensity of a biological molecule is a function of its concentration, its extinction coefficient (absorbing power) at the excitation wavelength, and its quantum yield at the emission wavelength.[2] A fluorescence emission spectrum represents the fluorescence intensity measured over a range of emission wavelengths at a fixed excitation wavelength. On the other hand, a fluorescence excitation spectrum is a plot of the fluorescence intensity at a particular emission wavelength for a range of excitation wavelengths.

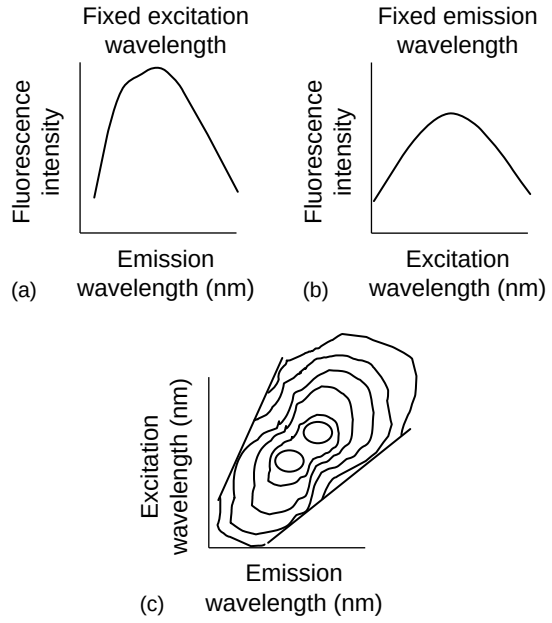

Figure 2 Illustration of: (a) a fluorescence emission spectrum; (b) a fluorescence excitation spectrum; (c) an EEM of a molecule.

A fluorescence excitation–emission matrix (EEM) is a two-dimensional contour plot that displays the fluorescence intensities as a function of a range of excitation and emission wavelengths. Each contour represents points of equal fluorescence intensity. Figure 2 illustrates a fluorescence emission spectrum (Figure 2a), a fluorescence excitation spectrum (Figure 2b) and a fluorescence EEM (Figure 2c).

2 FLUOROPHORES

2.1 Endogenous Fluorophores

Table 1 lists the biological molecules that exhibit endogenous fluorescence, along with their excitation and emission maxima.[11,27] These endogenous fluorophores include amino acids, structural proteins, enzymes and co-enzymes, vitamins, lipids and porphyrins. Their excitation maxima lie in the range 250–450 nm (spanning the UV/VIS spectral range), whereas their emission maxima lie in the range 280–700 nm (spanning the UV/VIS/NIR spectral range). Details of the molar extinction coefficients, fluorescence quantum yields and lifetimes are provided elsewhere.[11] The endogenous fluorophores that are speculated to play a role in transformations that occur with carcinogenesis are the amino acids tryptophan and tyrosine,[28,29] the structural proteins collagen and elastin,[30–36] the coenzymes NADH[28,37] and FAD,[37,38] and porphyrins.[39]

Table 1 Excitation and emission maxima of biological molecules that exhibit endogenous fluorescence

Endogenous fluorophores	Excitation maxima (nm)	Emission maxima (nm)
Amino acids		
Tryptophan	280	350
Tyrosine	275	300
Phenylalanine	260	280
Structural proteins		
Collagen	325	400, 405
Elastin	290, 325	340, 400
Enzymes and coenzymes		
FAD, flavins	450	535
NADH	290, 351	440, 460
NADPH	336	464
Vitamins		
Vitamin A	327	510
Vitamin K	335	480
Vitamin D	390	480
Vitamin B_6 compounds		
Pyridoxine	332, 340	400
Pyridoxamine	335	400
Pyridoxal	330	385
Pyridoxic acid	315	425
Pyridoxal 5′-phosphate	330	400
Vitamin B_{12}	275	305
Lipids		
Phospholipids	436	540, 560
Lipofuscin	340–395	540, 430–460
Ceroid	340–395	430–460, 540
Porphyrins	400–450	630, 690

FAD, flavin adenine dinucleotide; NADH, reduced nicotinamide adenine dinucleotide; AND(P)H, reduced nicotinamide adenine dinucleotide phosphate.

2.1.1 Amino Acids: Tryptophan and Tyrosine

Amino acids are the basic structural units of a protein[40] and proteins play crucial roles in virtually all of the biological processes.[40] Three amino acids with aromatic side chains are fluorescent. At excitation wavelengths above 295 nm only tryptophan is fluorescent.[26] From 280 to 295 nm, both tyrosine and tryptophan are fluorescent; however, energy transfer from tryptophan to tyrosine is common.[26] Below 280 nm, all three amino acids can be excited, albeit the quantum yield of phenylalanine is relatively low compared with that of tryptophan and tyrosine.[26]

2.1.2 Structural Proteins: Collagen and Elastin

Collagen, a family of fibrous proteins, is distinctive in forming insoluble fibers that have a high tensile strength.[40] It is the major extracellular matrix component that is present to some extent in nearly all organs

and serves to hold cells together in discrete units. Elastin is the major component of elastic fibers and is found in most connective tissues in conjunction with collagen and polysaccharides.[40]

Fluorescence has been noted from collagen[41] and elastin.[42] Collagen fluorescence in load-bearing tissues is associated with cross-links, hydroxylysyl pyridoline and lysyl pyridinoline.[43] Collagen fluorescence has an excitation maximum at 325 nm and an emission maximum at 400 nm, as displayed in Figure 3.[41] The fluorescent material in elastin is a tricarboxylic triamino pyridinium derivative, which is very similar in spectral properties to the fluorophore in collagen.[44] The excitation and emission maxima of the fluorescence of this elastin cross-link occur at 320 and 405 nm, respectively, as shown in Figure 4.[42]

2.1.3 Metabolic Coenzymes: Pyridine Nucleotides and Flavoproteins

Living organisms require a continual input of free energy through cellular metabolism for the performance of mechanical work in muscle contraction and other cellular movements, the active transport of molecules and ions, and the synthesis of macromolecules and other biomolecules from simple precursors.[40] In most processes, the carrier of free energy is adenosine triphosphate (ATP), which is derived from the oxidation of fuel molecules such as carbohydrates and fatty acids. In aerobic organisms, the ultimate oxidizing agent or electron acceptor is molecular oxygen. However, electrons are not transferred directly from fuel molecules and their breakdown products to molecular oxygen. Instead, these substrates transfer electrons to special

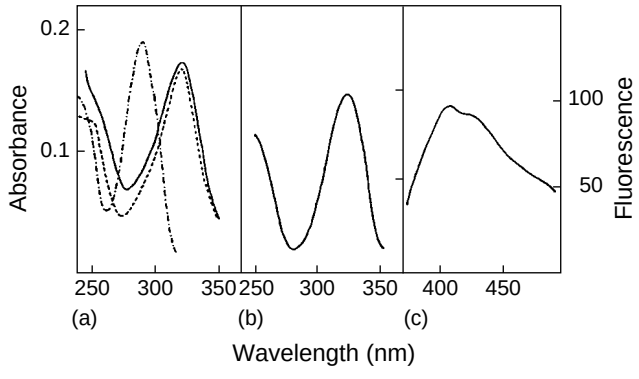

Figure 3 Collagen fluorescence: (a) UV absorption spectra in 0.1 N HCl (·····), in 0.1 M potassium phosphate buffer, pH 7.4 (- - -), and in 0.1 N NaOH (——); (b) fluorescence excitation spectrum in 0.02 M potassium phosphate buffer, pH 7.4, with emission fixed at 400 nm; (c) fluorescence emission spectrum in 0.02 M potassium phosphate buffer, pH 7.4, with excitation fixed at 325 nm. (Reproduced by permission of Academic Press, Orlando, Florida, from Fujimoto.[41])

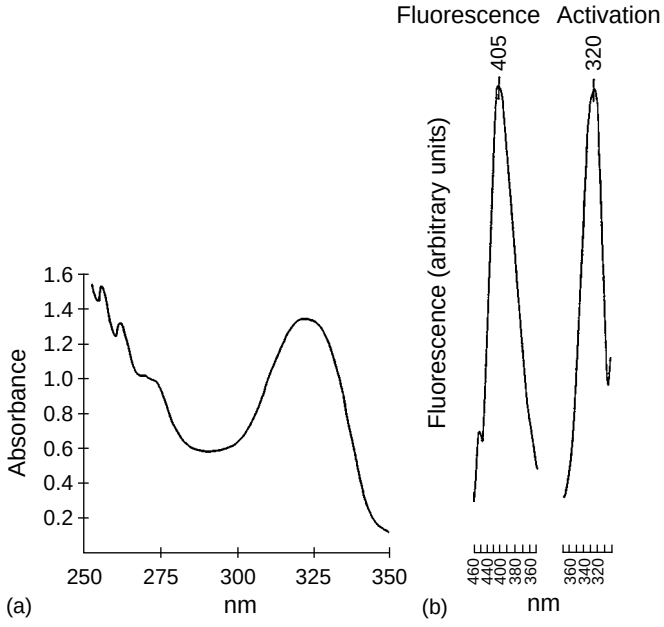

Figure 4 Elastin fluorescence: (a) UV absorption spectrum in 0.1 M potassium phosphate buffer; (b) fluorescence emission and excitation (activation) spectrum in 0.1 M HCl. (Reprinted from Deyl et al.,[42] with permission from Elsevier Science.)

carriers called pyridine nucleotide (PN) and flavoprotein (Fp). The oxidized form of PN – nicotinamide adenine dinucleotide (NAD^+) – and the oxidized form of Fp/FAD – are the major electron acceptors in the oxidation of fuel molecules. After accepting the electrons, these carriers become reduced. The now reduced electron carriers – NADH and reduced flavin adenine dinucleotide ($FADH_2$) – then transfer their electrons to molecular oxygen by means of the electron transport chain in the inner membrane of the mitochondria within the cell. As a result, molecular oxygen oxidizes them to NAD^+ and FAD. The ultimate flow of these electrons to molecular oxygen drives the synthesis of ATP.

The fluorescent forms of these electron carriers are the reduced form of PN and the oxidized form of Fp, which emit fluorescence when excited with UV and blue light respectively. Figure 5 displays the fluorescence excitation and emission spectra of Fp and PN from isolated pigeon heart mitochondria measured at −196 °C.[45] Figure 5(a) and (b) displays the oxidized (Ox) and reduced (Red) Fp excitation and emission spectra, respectively, and Figure 5(c) and (d) shows the oxidized (Ox) and reduced (Red) PN excitation and emission spectra, respectively. Clearly, the excitation and emission of Fp is maximal when it is oxidized and minimal when it is reduced. The converse is true for PN. The ratio of the concentration of oxidized and reduced electron carriers (or, synonymously, the ratio of the fluorescence of Fp and PN) gives a measure of the cellular metabolic rate.[45] In particular, a reduced

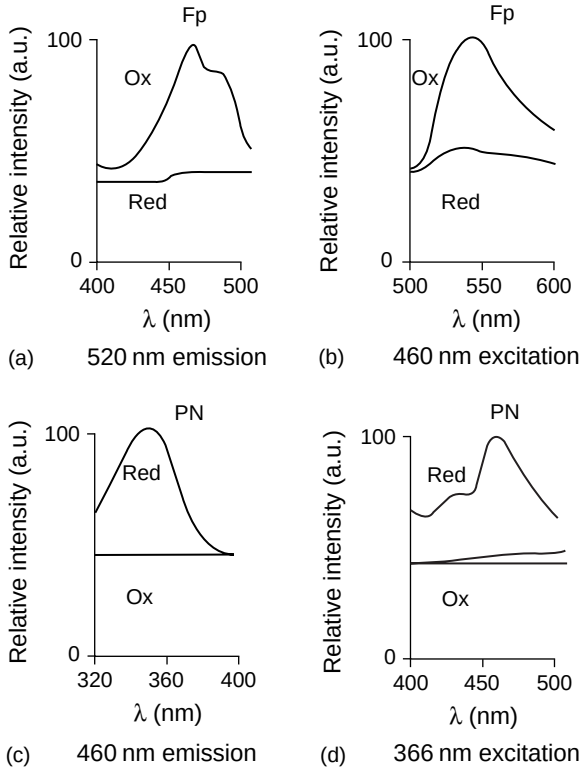

Figure 5 (a, b) The oxidized (Ox) and reduced (Red) Fp excitation and emission spectra, respectively. (c, d) The oxidized (Ox) and reduced (red) PN excitation and emission spectra, respectively. The measurements were made from pigeon heart mitochondria at −196 °C. (Reproduced by permission from Chance et al.[45])

ratio is indicative of a high phosphate potential (state 4), whereas an oxidized ratio indicates a low phosphate potential (states 2 and 3).

2.1.4 Porphyrins

Porphyrin is a precursor in heme biosynthesis.[40] It leads to the synthesis of protoporphyrin IX (PpIX), which acquires an atom to form heme. δ-Aminolevulinate synthase, the enzyme catalyzing the committed step in this pathway, is feedback inhibited by heme. Porphyrins produce a red fluorescence, with peaks at 630 and 690 nm, when excited in the blue spectral region between 400 and 450 nm.[46]

2.1.5 Fluorescence Microscopy and Spectroscopy of Microstructures in Cultured Cells and Tissue Sections

Fluorescence microscopy and spectroscopy of fluorescent micro-structures in cultured cells and optically thin, unstained, frozen tissue sections have indicated that differences in the fluorescence emission spectra of normal, precancerous and cancerous tissues may be attributed to differences associated with several endogenous fluorophores within the various sublayers of these tissues. Pradhan et al.[28] suggested that there is an increase in NADH and tryptophan, whereas Anidjar et al.[38] reported that there is probably a decrease in FAD as cells progress from a normal to cancerous state. Romer et al.[30] and Fairman et al.[32] observed a decreased fluorescence intensity from the collagens in precancerous colon tissue sections relative to that in normal colon tissue sections. Bottiroli et al.[34] observed a red fluorescence in some parts of a tissue section that was from a cancerous colon. The red fluorescence that was observed by Bottiroli et al.[34] has been observed previously by several groups.[37,47–51]

2.2 Exogenous Fluorophores

Most exogenous fluorophores currently being evaluated as contrast agents for precancer and cancer detection include photosensitizers, developed originally for photodynamic therapy (PDT).[12] These include hematoporphyrin derivative (HpD), pheophorbide-a, meso-tetra-(hydroxyphenyl)-chlorin (MTHPC), benzoporphyrin derivative (BPD), tin etiopurpurin (SnET2), hypercin and pthalocyanine.[12] These compounds have strong absorption characteristics in the blue spectral region or up to the highest Q-band at 635 nm, resulting in fluorescence between 625 and 675 nm.[12] These agents are generally administered systemically and the difference in accumulation of these exogenous fluorophores in tumor tissue relative to their normal counterpart appears to correlate with the differences in the vascularity of the two tissue types.[12] Lutetium texaphyrin is another photosensitizing agent that has absorption and fluorescence characteristics in the NIR spectral region.[12] This agent seems to be selectively phototoxic to tumors via apoptosis rather than necrosis (which results from vascular stasis), suggesting that this agent localizes in the cell.[12]

During the past few years an alternative concept has been introduced, which is based on the initial observations of Ghadially et al.,[51] who suggested using the photosensitizing precursor δ-aminolevulinic acid (5-ALA) to induce PpIX fluorescence in tumors. PpIX is characterized by an absorption maximum at 405 nm and fluorescence maxima at 630 and 700 nm.[52] The compound 5-ALA is a natural precursor of PpIX in the biosynthetic pathway for heme.[52] Normally, the synthesis of heme regulates the synthesis of PpIX through feedback control. However, the administration of 5-ALA bypasses this feedback and induces the accumulation of PpIX in tumor tissue. It has been suggested that deficiency in ferrochelatase (the enzyme required for conversion of PpIX to heme) in tumors results in accumulation of PpIX

in these tissues relative to normal tissue.[12] There are several advantages associated with using 5-ALA as a contrast agent. Both 5-ALA and PpIX are substances naturally present in the body, making the toxicity issue less critical. Furthermore, the drug can be administered conveniently both orally and topically.[12]

The use of photosensitizers as purely contrast agents has raised potential safety and toxicity concerns. Hence, recent work has examined nonphotosensitizing agents with similar absorption and fluorescence characteristics to photosensitizers for their potential to serve as contrast agents. Several exogenous fluorophores, including Nile blue and its derivatives and the caretenoporphyrins, have been shown to fulfil this criterion.[12] Nonphototoxic or weakly phototoxic exogenous fluorophores that are in routine clinical use currently are fluorescin (fluorescin angiography) and indocyanine green (cardiac output measurements and liver assessment), which have absorption and fluorescence characteristics in the VIS and NIR regions, respectively.[11] In clinical use, these drugs are injected into the vasculature, through which they are ultimately cleared.

3 FLUORESCENCE SPECTROSCOPY OF TURBID MEDIA

3.1 Fluorescence Spectroscopy of an Optically Dilute, Homogeneous Medium

Using the Beer–Lambert law,[2] one can express the fluorescence intensity of an optically dilute, homogeneous medium (sum of the optical densities at the excitation and emission wavelengths. Everywhere is less than unity)[26] as a linear function of the concentration of the fluorophores in that medium. Specifically, the fluorescence intensity, $F(\lambda_x, \lambda_m)$, at particular excitation (λ_x) and emission (λ_m) wavelengths due to k fluorophores is defined according to Equation (2):

$$F(\lambda_x, \lambda_m) = P_0(\lambda_x) \int_0^L dz\, \Omega \sum_{k=1}^{N} \mu_{a_k}(\lambda_x)\phi_k(\lambda_m) \quad (2)$$

and Equation (3) gives the definition of $\mu_a(\lambda_x)$, the absorption coefficient of the fluorophore at the excitation wavelength:

$$\mu_a(\lambda_x) = 2.303\varepsilon(\lambda_x)C \quad (3)$$

In Equations (2) and (3) $\phi(\lambda_m)$ is the quantum yield of the fluorophore at the emission wavelength, $P_0(\lambda_x)$ is the power of the intensity at the excitation wavelength, L is the path length and Ω is the detector collection efficiency. The absorption coefficient $\mu_a(\lambda_x)$ is a linear function of $\varepsilon(\lambda_x)$ and C, which are the extinction coefficients at excitation wavelength and concentration, respectively, of the fluorophore. Although fluorescence spectroscopy of optically dilute, homogeneous media is well understood, fluorescence spectroscopy of turbid media, such as human tissue, is complicated by its highly absorbing and scattering properties.[53]

3.2 Fluorescence Spectroscopy of Turbid Media such as Tissue

Fluorescence spectroscopy of turbid media such as tissue depends on one or more of the following. Specifically, it depends on the concentration and distribution of fluorophore(s) present in the tissue as well as the biochemical/biophysical environment, which may alter the quantum yield and lifetime of the fluorophore(s). For example, epithelial tissues generally have two primary sublayers – a surface epithelium and an underlying stroma or submucosa; the specific fluorophores, as well as their concentration and distribution, can vary significantly between these two tissue layers. Fluorescence spectroscopy of turbid media such as tissue also depends on the absorption and scattering that result from the concentration and distribution of nonfluorescent absorbers and scatterers, respectively, within the different sublayers of the tissue.

The effect of the aforementioned variables on fluorescence spectroscopy of tissue is wavelength dependent. First, the endogenous fluorophores that have absorption bands that lie in the same wavelength range as the excitation light will be excited and hence will emit fluorescence. The absorption and scattering properties of the tissue will affect light at both of these excitation and emission wavelengths. Therefore only those fluorophores contained in the tissue layers to which the excitation light penetrates and from which the emitted light can escape the tissue surface will produce measurable fluorescence. Elastic scattering events in tissue are caused by random spatial variations in the density, refractive index and dielectric constants of extracellular, cellular and subcellular components.[54] Tissue scattering generally decreases monotonically with increasing wavelength over the UV, VIS and NIR spectral regions.[11] Tissues are generally forward scattering (mean anisotropy factor is between 0.97 and 0.98), with scattering coefficients ranging from 10 to 1000 cm^{-1} from the NIR to the UV spectral range.[11] Absorption in tissue in the UV, VIS and NIR regions is primarily attributed to hemoglobin.[55] Although absorption in tissue is strongly wavelength dependent, it tends to decrease generally with increasing wavelengths.[11] Typically, tissue absorption coefficients range from 0.1 to 10 000 cm^{-1} from the NIR to the UV spectral range. Consequently, the penetration depth of light, which is primarily a function of the tissue absorption properties,

decreases from several centimeters to a few hundred microns from the NIR to the UV.[11] For example, in the UV spectral region, the penetration depth of light in tissue is approximately 225 μm at 337 nm.[3]

Hemoglobin, which is contained in red blood cells, serves as the oxygen carrier in blood and also plays a vital role in the transport of carbon dioxide and hydrogen ions.[40] The capacity of hemoglobin to bind oxygen depends on the presence of a nonpolypeptide unit, namely a heme group. The heme consists of an organic part, a protoporphyrin ring and an iron atom. Figure 6 displays the absorption spectra of oxygenated and deoxygenated hemoglobin over the UV/VIS (Figure 6a) and NIR (Figure 6b) spectral regions.[55] In Figure 6(a), the absorption spectra are characterized by the Soret band at 400–450 nm, the bands at 540 (α band) and 569 nm (β band) of oxygenated hemoglobin and the 557 nm band of deoxygenated hemoglobin. In Figure 6(b), the absorption spectra are characterized by the decreasing absorption of deoxygenated hemoglobin and the increasing absorption of oxygenated hemoglobin as a function of increasing wavelength.

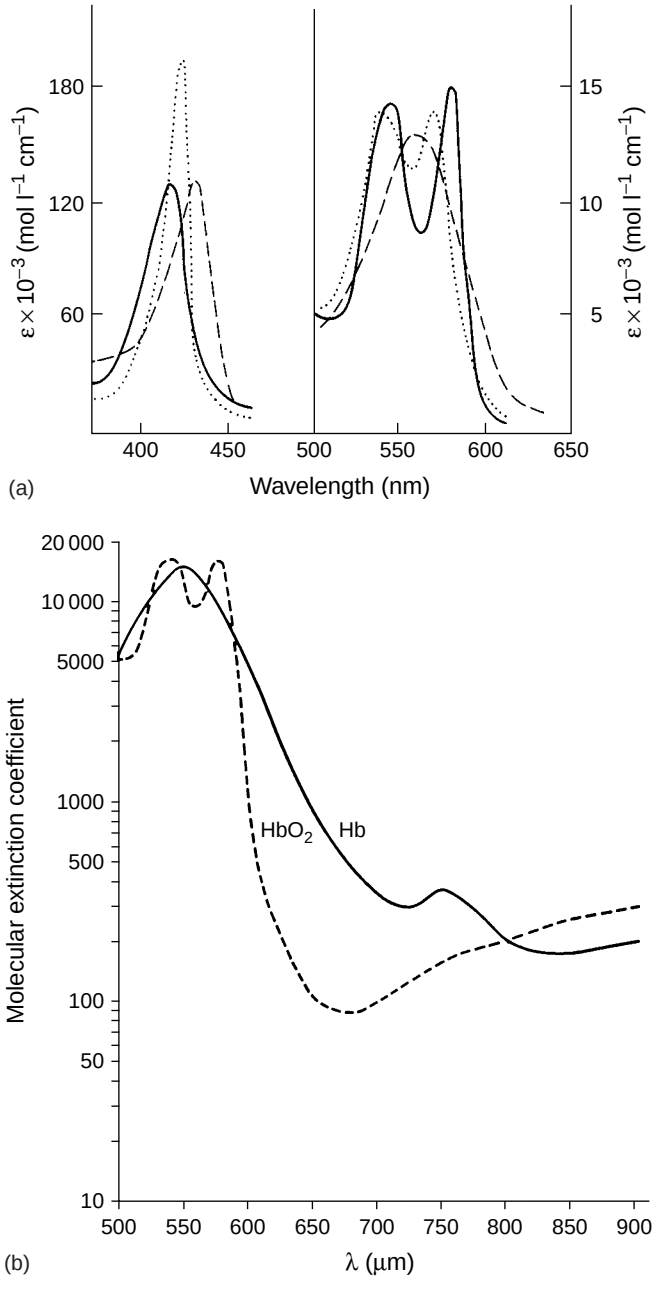

Figure 6 Absorption spectra of oxygenated and deoxygenated hemoglobin over: (a) UV/VIS spectral region (solid line, HbO_2; dashed line, Hb; dotted line, Soret band); (b) NIR spectral region. (Reproduced by permission of Academic Press, Orlando, Florida, from Brown.[55])

The effect of hemoglobin absorption is noticeable when comparing fluorescence emission spectra measured from an optically thick arterial tissue sample (250 μm thickness) to that measured from a corresponding, optically thin tissue section (4 μm thickness) using the same illumination and collection geometry, as shown in Figure 7.[56] Evaluation of Figure 7 indicates that the fluorescence intensity of the optically thick tissue sample is significantly reduced and, furthermore, its line shape is characterized by a valley at 420 nm, which corresponds to the Soret absorption band of hemoglobin.[55] The fluorescence emission spectrum of the optically thin tissue section may be attributed to collagen fluorescence.

The illumination and collection geometry of the excitation and the emitted light, respectively, can also affect fluorescence measurements from tissue with respect to both the intensity and line shape.[54,57,58] This may be attributed to the fact that although fluorescence is generated isotropically from the fluorophores within the

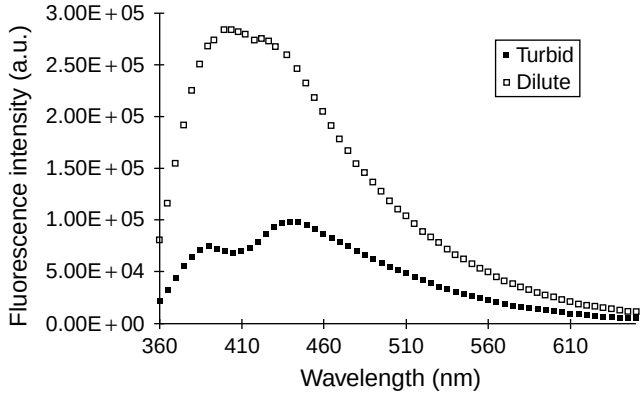

Figure 7 Fluorescence emission spectra at 340 nm excitation, measured from an optically thick, arterial tissue sample (250 μm thickness) to that measured from a corresponding, optically thin, tissue section (4 μm thickness) using the same excitation and collection geometry. (Reproduced by permission of the Optical Society of America, from Durkin et al.[56])

biological medium, the fluorescence emitted from the surface of the medium may range from isotropic to anisotropic, depending on whether the medium is highly absorbing, dilute or turbid.[54] Monte-Carlo simulations have been used extensively to simulate light distribution in turbid media to explore the effect of absorption and scattering on the fluorescence emitted from the surface of tissues using finite excitation beam profiles and complex excitation and emission geometries.[57] The results of these simulations indicate several important findings.

The profile of the excitation beam greatly affects the distribution of the excitation light in the tissue and thus is an important factor. For a Gaussian beam profile, the fluorescence distribution at a specific emission wavelength on the surface of the tissue is peaked and narrow, whereas for a uniform profile it is wider and less peaked. Furthermore, the effect of absorption, particularly due to hemoglobin, on the measured fluorescence increases as the collection fiber is moved further away from the illumination fiber on the surface of the tissue, resulting in decreased fluorescence intensity. In addition, the effect of this absorption is wavelength dependent, suggesting that it will affect the line shape of the fluorescence as well. This is evident when comparing the fluorescence emitted at two wavelengths at which hemoglobin has different absorption characteristics. As the distance of the collection fiber increases relative to the illumination fiber, the ratio of the fluorescence intensity at the two emission wavelengths deviates from unity. This suggests that collecting the emitted light from the tissue site directly illuminated by the excitation light will minimize the distortion due to hemoglobin absorption, albeit this will reduce the sampling depth for the fluorescence measurements.

The effect of the illumination and collection geometry on fluorescence measurements from tissue is illustrated in Figure 8, which displays the fluorescence emission spectra at 476 nm excitation of normal human cadaver aorta (N) and of one with atheromatous plaque (P).[58] In Figure 8(a) fluorescence was collected from the tissue site directly illuminated by the excitation light, and in Figure 8(b) the fluorescence was collected from a circular area around the directly illuminated area. Both the intensity and line shape of the fluorescence are affected by the probe geometry. For the normal aorta (Figure 8a) the ratio of the fluorescence intensity at 600 and 580 nm is 1.27, whereas it is 1.72 in Figure 8(b). For the atheromatous tissue, the ratio is 0.91 for both collection areas.

3.3 Deconvolution of Absorption and Scattering from Tissue Fluorescence Emission Spectra

Fluorescence excitation and emission from turbid media such as tissues consist of three components: the distribution of the excitation light in the tissue, which is a function of the absorption and scattering coefficients of the medium at that wavelength; the fluorescence of isotropically radiating point sources located at different depths within the tissue, which is determined by the fluorescence quantum yield of the fluorophore and the excitation intensity at that depth; and the total fluorescence escaping the surface of the tissue, which is a function of the absorption and scattering properties of the medium at that wavelength. Quantification of the concentration of fluorophores within tissue in principle involves deconvoluting the absorption and scattering properties of the tissue from the measured fluorescence emission spectrum and quantifying the identity and distribution of the fluorophores that contribute to the deconvolved spectrum. The former requires the development of a transfer function based on the measurements of tissue optical properties, i.e. the absorption and scattering coefficients and the anisotropy parameter, whereas the latter requires knowledge of the identity and distribution of the fluorophores within the tissue.

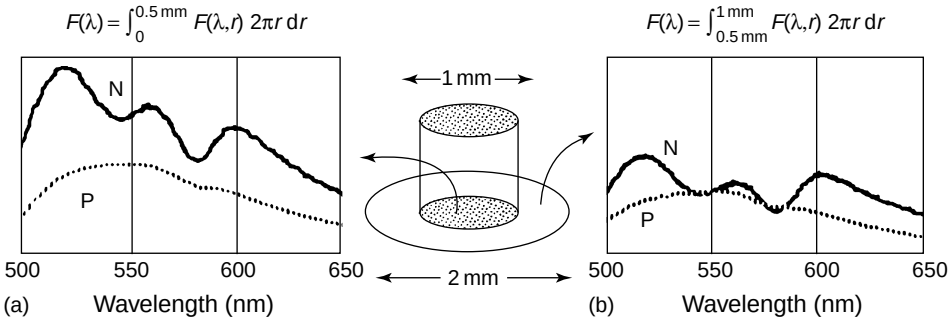

Figure 8 Fluorescence emission spectra at 476 nm excitation of normal human cadaver aorta (N) and one with atheromatous plaque (P): (a) the fluorescence was measured from the tissue site directly illuminated by the excitation light; (b) the fluorescence was measured from a circular area around the directly illuminated area. (Reproduced by permission of the Optical Society of America, from Keijzer et al.[58])

For a turbid medium such as tissue that contains k fluorophores, the fluorescence intensity can be rewritten to include a transfer function (TF) that describes the attenuation of the excited and emitted light due to the absorption and scattering properties of the tissue at these wavelengths.[54] This is defined in Equation (4):

$$F(\lambda_x, \lambda_m) = P_0(\lambda_x) \int_0^\infty dz \sum_{k=1}^{N} \mu_{a_k}(\lambda_x, z)\phi_k(\lambda_m, z) \times TF(\lambda_x, \lambda_m, z)\Omega \quad (4)$$

where z refers to the depth within the medium. Here, the transfer function for a particular depth can be determined from the tissue absorption coefficient, the scattering coefficient and the anisotropy factor. The wavelength-dependent absorption coefficient (μ_a) denotes the probability of photon absorption per unit path length, whereas the scattering coefficient (μ_s) denotes the probability of photon scattering per unit path length. The anisotropy factor (g) denotes the cosine of the average scattering angle. The reduced or isotropic scattering coefficient (μ_s') is the product of the scattering coefficient (μ_s) and one minus the anisotropy parameter ($1 - g$). Models of light transport[59–65] can be used to deduce the absorption and scattering coefficients and the anisotropy parameter needed to compute the transfer function in Equation (4).

3.3.1 Models of Light Transport

Most models of light transport used to quantify tissue optical properties to date have been either numerical or approximate analytical solutions of the neutral particle transport equation.[59,60] Because of the inhomogeneity of tissues, the solution of Maxwell's equations, which might otherwise be used to model light propagation in tissues, is generally not feasible. However, if polarization and diffraction effects are ignored, the transport of photons through random media may be modeled as neutral particle transport. The neutral particle transport theory is heuristic and is based on a statistical approximation of photon transport in a multiple- scattering medium. Within this framework, the propagation of light is described in terms of the transport of discrete photons, which may be either scattered or absorbed. The transport equation (Equation 5) is defined as:

$$\frac{1}{c}\frac{\partial\phi(\vec{r}, \hat{\Omega})}{\partial t} + \hat{\Omega} \cdot \nabla\phi(\vec{r}, \hat{\Omega}) + [\mu_a(\vec{r}) + \mu_s(\vec{r})]\phi(\vec{r}, \hat{\Omega}) = \mu_s(\vec{r}) \int_{4\pi} d\hat{\Omega}' f(\hat{\Omega}' \to \hat{\Omega})\phi(\vec{r}, \hat{\Omega}') + s(\vec{r}, \hat{\Omega}) \quad (5)$$

where $\phi(\vec{r}, \hat{\Omega})$ is the local angular flux of photons at position $\vec{r}$, angle $\hat{\Omega}$ and time t in a scattering medium. The absorption coefficient $\mu_a(\vec{r})$ describes photon absorption at position $\vec{r}$, and the scattering coefficient $\mu_s(\vec{r})$ describes photon scattering at position $\vec{r}$. The term $f(\hat{\Omega}' \to \hat{\Omega})$ is the scattering phase function, which describes the probability of a photon scattered from direction $\hat{\Omega}'$ into direction $\hat{\Omega}$. Term $S(\vec{r}, \hat{\Omega})$ is a source term for photons at position $\vec{r}$ and angle $\hat{\Omega}$. Equation (5) can be solved numerically or approximated analytically to determine the position- and angle-dependent flux of photons in a medium described by an arbitrary absorption, scattering and phase function distribution.

Approximate analytical models include the following approximations. The simplest analytical model is the Beer–Lambert law,[2] which assumes that the incident light on the medium attenuates exponentially in the direction of propagation within that medium. This is valid for a medium that does not scatter light, but is sufficiently absorbing that it can no longer be dilute, which is generally not true for tissues. The Kubelka–Munk approximation[61] describes the time-independent, diffuse reflectance and transmittance for light incident on a tissue slab in terms of its spatially uniform absorption and isotropic scattering coefficients. The radiative transport equation also can be solved using the adding–doubling method[62] for appropriate boundary conditions, and its solution can be compared with measurements of time-independent, diffusely reflected and transmitted light to estimate the absorption and scattering, as well as the anisotropy factors. Models based on Kubelka–Munk and the adding–doubling method are appropriate for the determination of tissue absorption and scattering coefficients in a wavelength region in which neither absorption nor scattering in tissue is dominant (UV/VIS spectral regions).

Within the diffusion approximation,[63] which is valid far from sources and when the scattering coefficient is several orders of magnitude greater than the absorption coefficient, the spatial and temporal distribution of the diffusely reflected or transmitted light from tissue is related to the absorption and scattering coefficients and anisotropy factor. Diffusion theory is valid for cases where the absorption is low and scattering is high in tissues, such that the light is multiply scattered and travels through various optical path lengths before being detected (as in the NIR spectral region). When the probability of absorption in tissue becomes significant (as in the UV/VIS spectral regions), re-emitted light is minimally or moderately scattered and diffusion theory is no longer valid.

Numerical solutions to the transport equation include the method of discrete ordinates[64] and the Monte-Carlo technique.[65] These techniques can be used to model light transport in tissue over the entire UV/VIS/NIR spectral regions. The Monte-Carlo technique, which has been the

more popular of these numerical models, tracks individual photon trajectories, which are computer-simulated to calculate absorption and scattering coefficients (inverse calculation) or space irradiance distributions (forward calculation) for any tissue, tissue geometry and wavelength region. However, because tens of millions of photon trajectories must be tracked for statistical simulation of light transport in tissue, computations are time intensive.

The aforementioned numerical techniques can be used to compute the absorption and scattering coefficients and anisotropy parameter from measurements of diffusely reflected and transmitted light from tissue. Additionally, if the optical properties are known at the specific excitation and emission wavelengths and the fluorophore(s) quantum yield and distributions are provided, fluorescence excitation, emission and escape from the different depths within the tissue can be computed. Hence, these numerical techniques can be used to compute numerically the transfer function in Equation (4).

3.4 Turbid Tissue-simulating Phantoms for Fluorescence Spectroscopy of Tissue

In order to evaluate the effect of absorption and scattering and the illumination and collection geometry experimentally on fluorescence measurements from tissue, it is useful to employ tissue phantoms that simulate the optical properties and fluorescent properties of these biological media.

Durkin et al.[66] developed a liquid tissue phantom that simulates the optical and fluorescence properties of tissues over the range 350–650 nm, with absorption coefficients in the range 25–5 cm^{-1}, scattering coefficients, in the range 400–200 cm^{-1} and fluorophores in the micromolar concentration range. Phantom fluorophores included FAD and rhodamine B. Absorption was controlled by adjusting the hemoglobin concentration. Polystyrene spheres, which have a smoothly varying scattering coefficient as a function of wavelength, were used as the scatterers. Sample inhomogeneities were simulated by preparing the phantom in a gelatin substrate. These phantoms could be made dilute, absorbing and/or turbid.

Wagnieres et al.[67] also developed tissue phantoms that have the optical characteristics of biological tissues in the wavelength range 400–650 nm, with absorption coefficients in the range 21–0.7 cm^{-1} and scattering coefficients in the range 680–220 cm^{-1}. The phantoms are made up of agarose dissolved in water to provide a transparent matrix, which is then loaded with silicon dioxide, Intralipid, ink, blood, azide, penicillin, bovine serum and fluorophores. The silicon dioxide and Intralipid particles are responsible for scattering, whereas the ink and blood are absorbers. The penicillin and azide are used to ensure conservation of the phantoms at 4 °C. The serum, and fluorophores such as Coumarin 30, produce fluorescence. The mechanical properties of these gelatinous phantoms render them easily moldable so that complex structures and shapes simulating layered or other inhomogeneous structures containing various amounts of absorber, scatterer and fluorophore can be developed.

4 INSTRUMENTATION

In order to perform quantitative fluorescence spectroscopy of tissue, the fluorescence emission spectra and the reflectance and/or the diffuse reflectance and transmittance need to be measured. The latter are needed if tissue optical properties are to be quantified. However, because intact issues are semi-infinite media, only fluorescence emission and reflectance spectra are measured from these biological systems.

A schematic of the basic components of an instrument used for fluorescence and/or reflectance spectroscopy of tissues is shown in Figure 9. It consists of a light source, a flexible, conduit that contains optical fibers for the illumination and collection of light, a dispersing element that separates the emitted light into its respective wavelengths, and a detector that measures the intensity at these wavelength(s). In this scenario, fluorescence emission and reflectance are measured in a re-emission geometry in which the illumination and collection are performed on the same surface of the tissue. The various types of instruments employed for these measurements essentially have the same basic components.

4.1 Light Sources

Lasers, which have a very narrow spectral output, are generally used as monochromatic excitation light sources for fluorescence spectroscopy, whereas lamps with a broad spectral output are more appropriate for

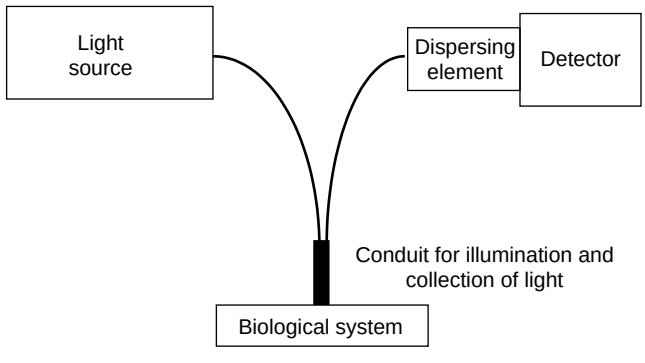

Figure 9 Schematic drawing of a fiber-optic-based instrument for fluorescence and/or reflectance spectroscopy.

reflectance spectroscopy. Lamps also can be used as quasi-monochromatic excitation light sources for fluorescence spectroscopy when coupled with a monochromator or narrow-bandpass filter (see section 4.3). The need for portable light sources for clinical applications has limited the use of monochromatic light sources in the UV/VIS spectral range to the nitrogen pumped-dye laser (fundamental wavelength is 337 nm) and the helium–cadmium laser (wavelengths are 325 and 442 nm). Additionally, a dye module coupled to the nitrogen laser enables the generation of VIS laser wavelengths through the use of appropriate chemical dyes in a resonant cavity. Other lasers that have been used for fluorescence spectroscopy include the argon ion (UV and blue), helium neon (red) and krypton ion (blue) lasers. Portable polychromatic sources include the mercury and xenon arc lamps. The primary criteria for the selection of light sources, in addition to portability, are output power, bandwidth, wavelength tunability, coupling efficiency (into optical fibers) and the need for pulsed versus continuous wave light sources.

The power from a portable nitrogen laser at 337 nm is ~300 μJ per pulse. At a maximal repetition rate of 30 Hz its average power is ~9 mW. Because the nominal conversion efficiency of dyes used to generate VIS wavelengths is approximately 10%, the average power in this case would be ~1 mW. The output power from the helium–cadmium laser ranges from 5 mW at 325 nm to 15 mW at 442 nm. Commercially available mercury and xenon arc lamps can be operated over a range of 50–1000 W. Owing to excess generation of heat in the IR, clinical applications employing these lamps have been restricted to the use of power supplies between 75 and 300 W, with a liquid filter for IR rejection. Note that the output powers available for tissue illumination are significantly lower than the output powers specified, due to coupling losses from the light source into the optical fibers and/or coupling optics.

Although the UV/VIS lasers afford a very narrow spectral band (1–2 nm), the bandwidth of illumination from a lamp is determined by the use of narrow-bandpass filters or a monochromator (see section 4.3), which is used for wavelength dispersion. In general a 10–20 nm bandpass for the light source is sufficient, because tissue fluorescence emission spectra are broad with a 40–60 nm full width at half-maximum (fwhm).

The advantages and disadvantages of using a lamp versus a laser for fluorescence spectroscopy should be considered. An advantage of using a mercury or xenon lamp over a laser is that it provides wavelength tunability over the UV/VIS spectral range, as shown in Figure 10.[68] However, the disadvantage of using a lamp over a laser is that the coupling efficiency into optical fibers can be reduced by a factor of 20 compared to laser coupling,[69] due to the relatively large focal spot diameter of arc lamps. Finally, lasers with very short pulse duration (of the order of nanoseconds) are necessary when the tissue needs to be illuminated with pulsed excitation light for gated detection (which provides effective rejection of ambient light during fluorescence measurements) and for fluorescence lifetime measurements.

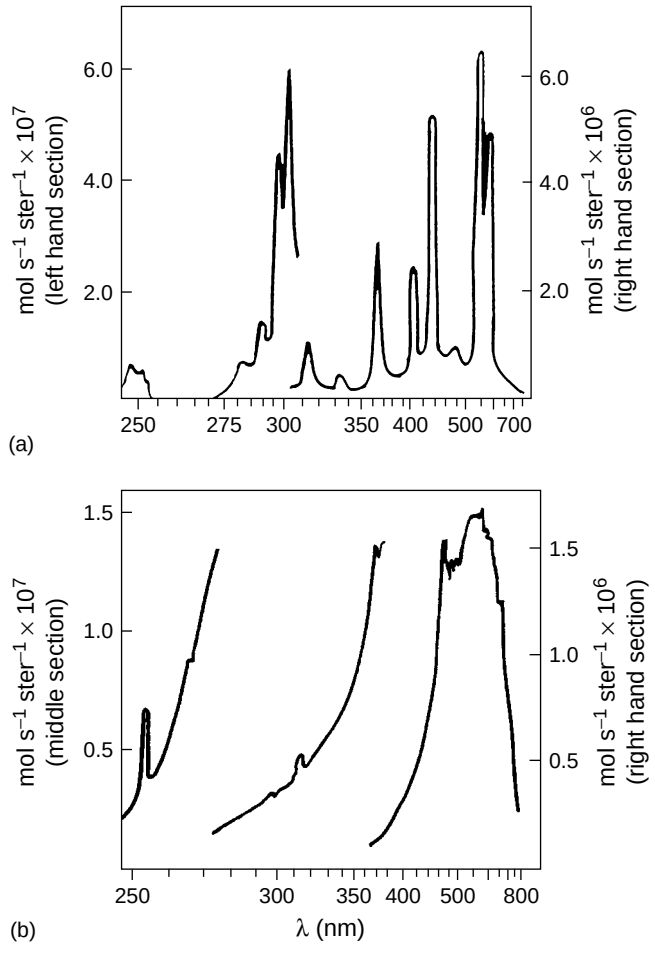

Figure 10 Typical radiant intensity and shape of the emission spectra of: (a) mercury lamp; (b) xenon arc lamp. (Reproduced by permission of Academic Press, Orlando, Florida, from Kohen et al.[68])

4.2 Illumination and Collection of Light

With respect to illumination and collection of light from tissue, two different approaches may be considered: the contact approach, where fiber-optic probes are placed directly in contact with the tissue surface; and the noncontact approach, where a series of lenses are used to project the light onto the tissue surface and collect it in a similar manner. With the contact approach, variable pressure on the tissue may distort the signal. However, with the noncontact approach the signal strength will vary

with the source–tissue and tissue–detector distances. In general, the contact approach is used for steady-state and time-resolved fluorescence and reflectance spectroscopy from small tissue volumes, whereas the noncontact approach is more suited for imaging larger tissue areas.

For the fiber-optic-based approach, typically multimode, step-index fibers with a core diameter ranging from 100 μm to 1 mm and *f* numbers (see section 4.3) of 2–4 are used for UV/VIS, fluorescence and reflectance spectroscopy. Fused-silica optical fibers are used for UV-excited fluorescence spectroscopy because standard silica fibers generate fluorescence with UV excitation.

The fluorescence emission spectra of turbid media such as tissue are extremely dependent on the illumination and collection geometry due to the interaction of absorption and fluorescence.[57] As shown elsewhere,[57,58] design of a fiber-optic probe configuration that reduces the distortion of the fluorescence emission spectrum due to tissue absorption consists of a geometry in which the tissue fluorescence is collected only from that surface directly illuminated by the excitation light. Richards-Kortum et al.[57] first designed optical fiber probes with illumination and collection geometries that enable maximal overlap between excitation and emission areas on the tissue surface.

In general, reflectance spectroscopy refers to the detection of both the diffuse (multiply scattered) and specular (surface) components of the reflectance. When measuring the diffuse reflectance from tissues, the photons that are detected generally undergo multiple scattering events and the photon path lengths are much greater than the geometric separation between the source and detector. Consequently, through this measurement, the probe is able to sample features in the tissue volume. Because specular reflection is due to refractive index mismatch and is generally a 2% reflection of the incident light in the case of an air/tissue interface, it needs to be minimized in attempting to make diffuse reflectance measurements. In order to minimize specular reflection, geometries have been employed in which the source and detector fiber are placed adjacent to each other, so that the specularly reflected light is not coupled to the detector fiber.[9] However, this geometry precludes the ability to collect diffuse reflectance from the same site that is illuminated. This presents two problems: there is attenuation of the signal; and if fluorescence measurements are made from the same site of illumination there will be a lack of congruence between the fluorescence and reflectance measurements. One way to minimize specular reflection in a case where the diffuse reflectance is collected from the same site that is illuminated is to have the distal tip of the collection fiber angled at 17° from the direction normal to the optical fiber axis, such that the specular reflection is not coupled but the diffuse reflectance is.[69]

4.3 Monochromators and Spectrographs

Light can be dispersed spectrally using a monochromator or a spectrograph, which are both dispersing components. A monochromator presents one wavelength or bandpass at a time of the input light from its exit slit, whereas a spectrograph presents a range of wavelengths simultaneously at the exit focal plane. Monochromators can be used as filters in conjunction with arc lamps to produce a series of monochromatic outputs for sample illumination (if only a few wavelengths are needed, narrow-bandpass filters may be more appropriate) or can be used to disperse the emitted light into its respective wavelengths, each of which can be detected serially using a single-channel detector. Spectrographs can used to disperse the emitted light into its respective wavelengths simultaneously for multichannel detection.

The key components of a monochromator/spectrograph are: an entrance slit; a collimating lens; a grating (dispersing unit) for wavelength selection; a focusing lens; and an exit slit (see Figure 11). Grooves in diffraction gratings are manufactured either classically with the use of a ruling engine (ruled gratings) or holographically with the use of interference fringes generated at the intersection of two laser beams (holographic gratings). Groove densities generally range from 50 to 3600 grooves mm^{-1}. Holographic gratings have a higher groove density than ruled gratings.

Parameters that are relevant to the selection of a monochromator/spectrograph are the reciprocal linear dispersion ($mm\,nm^{-1}$), transmission efficiency, *f* number and stray light rejection. The reciprocal linear dispersion (which is inversely related to the grating groove density, diffraction order and focal length of the focusing element), when multiplied by the entrance slit width, provides the spectral resolution. For fluorescence and

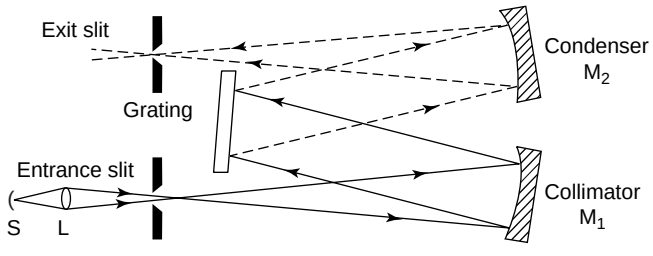

Figure 11 Schematic drawing of a monochromator that shows the arrangement of the light source (S), the condensing lens (L), the entrance slit, the mirror optics (M_1 and M_2), the grating and the exit slit. (Reproduced by permission of Academic Press, Orlando, Florida, from Kohen et al.[68])

reflectance spectroscopy, generally a spectral resolution of 10 nm is sufficient because these spectra are broad band and hence lack structure within this bandpass. Transmission efficiencies of both holographic and ruled gratings are comparable (~40%) in the VIS spectral region, and the efficiency of a particular grating is maximal at the wavelength at which it is blazed. Collection efficiency is reflected by the f number, which is defined as the ratio of the focal length and diameter of the collimating element. Lower f numbers are usually associated with a higher light gathering power or throughput. However, higher f-number spectrographs with longer focal-length-collimating elements are generally used to project a wide, flat field ideal for multichannel imaging detectors. For the applications indicated here, generally f numbers that match those of standard optical fibers (~2–4) are selected in order to maximize coupling efficiency. Stray light in monochromators/spectrographs is due to diffusion of light by optical components, including gratings and mirrors. Stray light is reduced by using holographic gratings, which usually give an order of magnitude better stray light rejection than classically ruled gratings. Typical stray light rejection for UV/VIS, fluorescence and reflectance spectroscopy should be specified to five orders of magnitude lower than the signal. Otherwise, long-pass filters with an optical density of at least 5.0 at the wavelength of excitation have to be employed to block the excitation light from the detector.

4.4 Detectors

The important considerations in choosing a detector are the type of measurements being made, i.e. single wavelength versus multiwavelength and single pixel (small region of the tissue surface) versus multipixel (relatively larger tissue area). Fluorescence and reflectance measurements from single pixels can be made either using a single-channel or multichannel detector. If measurements at only one or several wavelengths are being made, a single-channel photoemissive tube – called a photomultiplier tube (PMT) – or a semiconductor-based avalanche photodiode (APD) with a bandpass filter can be used. For fluorescence or reflectance spectroscopy, a spectrograph coupled to a multichannel, linear photodiode array (PDA) is appropriate. Fluorescence and reflectance spectroscopy also can be performed using a monochromator coupled to a PMT, albeit this substantially increases the measurement time. In the case of measurements from multiple pixels on the tissue surface, a two-dimensional charge-coupled device (CCD) camera may be employed. In order to reduce or eliminate the detection of ambient light during fluorescence spectroscopy, a detector with an intensifier for fast gating (several nanoseconds) must be used in conjunction with a pulsed excitation light source.

With respect to general detector performance, other considerations include quantum efficiency and sources of noise. For the detector, the quantum efficiency is defined as the ratio of induced current to the induced flux (often measured in electrons per photon). The quantum efficiency depends on the wavelength of light used, the material type and shape and other physical parameters. Figure 12 displays the quantum efficiency of a PMT and PDA/CCD.[70]

The noise components that we must consider when detecting light are as follows:

- Shot noise, which is due to random statistical fluctuations of the incident light. Shot noise increases with the square root of the signal.
- Dark signal, which comes from the random generation of electrons. Dark signal is temperature dependent and thus can be reduced by cooling the detector.
- Readout noise, which comes from the electronic process of reading the signal from the detector.

The dark flux for a PMT lies in the range 2–50 $e^- s^{-1}$. The readout noise is essentially zero because the pulse-height discrimination effectively avoids quantization noise. Although the un-intensified PDAs have been applied where the light levels are high, as in the case of reflectance spectroscopy, intensified PDAs are more suitable for fluorescence spectroscopy in which light levels are several orders of magnitude lower. In an intensified PDA, a microchannel plate intensifier is used to amplify the photoelectrons by a factor of 3000. Typical dark (at −20 °C) and readout noises are 40 counts s^{-1} and 3 e^-, respectively (manufacturer specification is 1 count per e^-). The dark current of the CCD cameras range from 12 e^- $pixel^{-1}$ s^{-1} at −40 °C to 0 e^- $pixel^{-1}$ s^{-1} when cooled to −110 °C. Their readout noise ranges from 12 to 4 e^-.

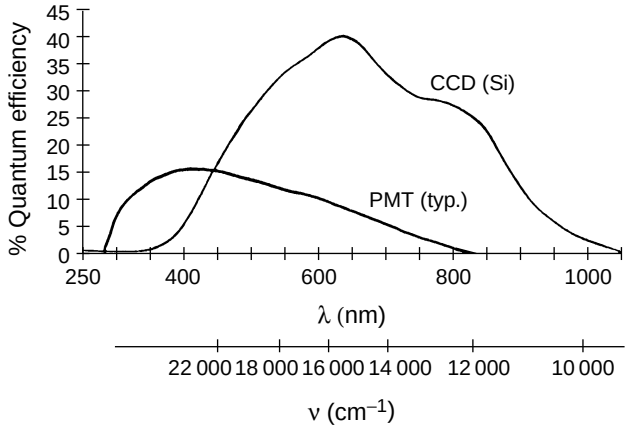

Figure 12 Quantum efficiency of a PMT and PDA/CCD.

4.5 Signal-to-noise Ratio Analysis of an Instrument Used for Fluorescence Spectroscopy of Tissue

To establish the performance of an instrument for fluorescence spectroscopy of tissue (fluorescence rather than reflectance measurements are limiting in signal to noise), Zangaro et al.[69] did a signal-to-noise ratio analysis using typical components. They used a nitrogen pumped-dye laser coupled to a fiber-optic probe for illumination and collection, and a spectrograph coupled to an intensified PDA for detection. Their calculation is as follows. The output pulse energy of the nitrogen laser is ~300 μJ. If the dye module is used, the nominal conversion efficiency is around 10%, but with surface reflection losses it becomes 25 μJ per pulse. They calculated losses of 40% in coupling this light to a 200-μm core diameter fiber, resulting in an output energy of 16 μJ per pulse. Assuming a tissue fluorescence efficiency in the 400-nm region to be 0.01%, the average fluorescence energy that they calculated is ~1.6 nJ. Given a collection efficiency of 1% of the fluorescence escaping from the tissue surface to the detector fiber, the fluorescence energy coupled to the detector fiber is ~16 pJ. Accounting for coupling losses and transmission efficiency through the spectrograph, the total energy pulse that reaches the PDA is two orders of magnitude lower, i.e. 400 fJ.

The signal-to-noise ratio obtainable by the PDA is determined by S, which is the total number of photoelectrons that reach the detector, and N, which is the number of photoelectrons associated with noise. The total noise generated by the various sources is defined according to Equation (6):

$$N = (S + N_R^2 + N_D^2)^{1/2} \tag{6}$$

where the square root of S is the shot noise from the signal photoelectrons collected by the detector, N_R is the readout noise inherent in the amplifier and electronics and N_D is the noise due to the dark current and is dependent on the exposure time and temperature. Because the detector is Peltier cooled to 5 °C during operation, the dark current is negligible. The signal that reaches the photodetector is estimated to be 400 fJ. Given a quantum efficiency in the 400-nm region of 50% (1×1018 photoelectrons J^{-1}), approximately 400 000 photoelectrons reach the PDA. The shot noise is calculated to be 632 electrons. The readout noise for the PDA is 3000 electrons. Using these quantities, the signal-to-noise ratio obtainable with these parameters is approximately 130. Trujillo et al. have also done a systematic evaluation of the signal-to-noise ratio of a typical instrument used for fluorescence spectroscopy of tissue.[71] In addition, they have developed an analytical expression to quantify the fluorescence efficiency of the tissue, which they have incorporated into the signal-to-noise ratio analysis of their fluorimeter.

4.6 Calculation of Tissue Fluorescence Efficiency

In order to calculate the expected signal-to-noise ratio of a fluorimeter, two quantities are required: the throughput and inherent noise of the fluorimeter; and the fluorescence efficiency of the tissue volume probed. The throughput and inherent noise of the fluorimeter is a function of the delivery of the excitation light to the tissue surface, the collection of the fluorescence from the tissue surface, and finally conversion of the optical signal to an electronic signal. The calculation of these three quantities has been described in section 4.5. The fluorescence efficiency is the quantitative relationship between the excitation and emission energy. Trujillo et al.[71] have developed an analytical expression to calculate the fluorescence efficiency of tissue from fluorescence emission spectra of tissues measured in vivo.

The fluorescence efficiency (FE) of a turbid sample is defined as the ratio of the total emitted photons to the excitation photons and is described as shown in Equation (7):

$$FE = \int_0^\infty dz \int_{\lambda_m}^\infty d\lambda_m H_{in}(\lambda_x, z) \sum_k [\mu_{a_k}(\lambda_x, z) \times \phi_k(\lambda_m, z)] H_{out}(\lambda_m, z) \tag{7}$$

where λ_x is the excitation wavelength, λ_m is the emission wavelength, z is the depth within the sample, $\mu_a(\lambda_x, z)$ represents the absorption coefficient at the excitation wavelength of the fluorophore at depth z, $\phi(\lambda_m, z)$ is the fraction of the absorbed energy converted to fluorescence at the emission wavelength of the fluorophore at depth z, and $H_{in}(\lambda_x, z)$ and $H_{out}(\lambda_m, z)$ represent the attenuation of the excitation and emission light, respectively, at depth z. The fluorescence efficiency is independent of the power of the intensity at the excitation wavelength and the illumination/collection geometry. The fluorescence efficiency of an unknown turbid sample can be calculated from fluorescence measurements of the unknown sample and fluorescence measurements of an optically dilute homogeneous standard with a known absorption coefficient (μ_a) and quantum efficiency using the same apparatus. Given the fluorescence emission spectrum of the turbid sample and that of a homogeneous standard with a known μ_a and quantum efficiency, the fluorescence efficiency of the turbid sample can be expressed by Equation (8):

$$FE_t(\lambda_x, \lambda_m) = \frac{\int_{\lambda_m}^\infty P_t(\lambda_x, \lambda_m)\, d\lambda_m}{\int_l^{\lambda_m} P_{std}(\lambda_x, \lambda_m)\, d\lambda_m} \times \frac{\int_0^l dz \int_{\lambda_m}^\infty d\lambda_m H_{in-std}(\lambda_x, z) \mu_{a-std} \times (\lambda_x) \phi_{std}(\lambda_m) H_{out-std}(\lambda_m, z) \Omega_t}{\Omega_{std}} \tag{8}$$

where the subscript "t" corresponds to that of the turbid sample, the subscript "std" corresponds to that of the homogeneous standard, P_t is the total integrated fluorescence power of the turbid sample, P_{std} is the total integrated fluorescence power of the homogeneous standard (both of which can be measured), and Ω_t and Ω_{std} represent the collection efficiency of the turbid sample and homogeneous standard respectively. The first term in Equation (8) can be calculated from the ratio of the total integrated fluorescence power of the turbid sample and the homogeneous standard. The second term requires knowledge of the optical properties as well as the collection geometry of the homogeneous standard. The authors used a rhodamine solution in a cuvette as the homogeneous standard in these calculations. Because the rhodamine standard used is an optically dilute homogeneous solution, and the numerical aperture of the excitation and collection is small (0.22), the attenuation of the excitation light and fluorescent light in the rhodamine standard can be described with the simple Beer–Lambert Law.[2] Hence, the wavelength-dependent μ_a can be calculated simply from the measured rhodamine absorbance. The quantum efficiency of rhodamine is unity.[72] The double integral in Equation (8) can be evaluated by using the measured values of $\mu_a(\lambda_x)$, $\mu_a(\lambda_m)$ and $\phi(\lambda_m)$ of the homogeneous standard. If $\phi(\lambda_m)$ is not known but the quantum efficiency is, then an alternative method can be used to calculate the fluorescence efficiency, which requires changing $\phi(\lambda_m)$ into the quantum efficiency QE (defined by Equation 9):

$$QE = \int_{\lambda_x}^{\infty} d\lambda_m \phi(\lambda_m) \tag{9}$$

Figure 13 displays the average fluorescence efficiency calculated from the fluorescence emission spectra of a total of 374 human cervical tissues at three excitation wavelengths in vivo. In Figure 13, the average fluorescence efficiency values are plotted for normal squamous tissue, normal columnar tissues, low-grade squamous intra-epithelial lesions (SIL) and high-grade SIL at excitation wavelengths 337, 380 and 460 nm. Standard deviations are shown as error bars. For squamous cervical tissues, the fluorescence efficiency decreases as tissue progresses from a nondiseased normal squamous to a diseased state (low-grade and high-grade SIL). Furthermore, as the excitation wavelength increases, the fluorescence efficiency decreases. The fluorescence efficiencies shown here are instrument independent and can be used to determine the signal-to-noise ratio needed to measure fluorescence emission spectra from tissues for different instrument configurations and excitation–emission collection geometries.

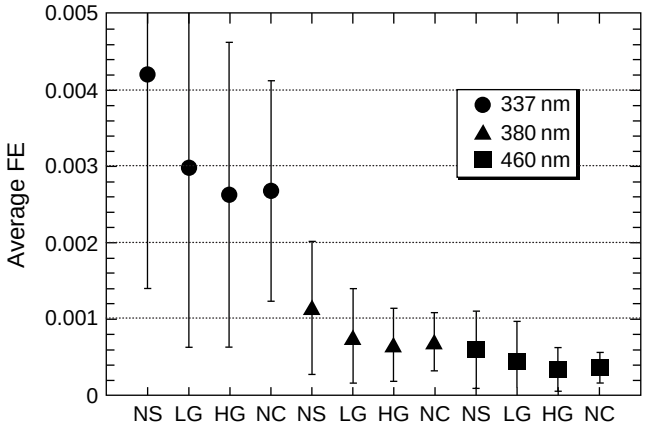

Figure 13 Average fluorescence efficiency (FE) (±standard deviation) for human cervical normal squamous tissues (NS), normal columnar tissues (NC), low-grade SIL (LG) and high-grade SIL (HG) calculated from their fluorescence spectra at the excitation wavelengths 337, 380 and 460 nm. (Reproduced by permission of the Society for Applied Spectroscopy, from McCreery.[70])

5 CLINICAL APPLICATIONS

Currently, one of the most widely explored applications of fluorescence spectroscopy is the detection of endoscopically invisible, early neoplastic growth in epithelial tissue sites. Early neoplastic growth refers to precancerous changes such as dysplasia and carcinoma in situ, which precede invasive cancer or carcinoma. Currently, there are no effective diagnostic techniques for these early tissue transformations. Fluorescence spectroscopy is ideally suited for this application because of its ability to examine tissue surfaces rather than tissue volumes, and its adaptability to an endoscopic device. If fluorescence spectroscopy can be applied successfully as a diagnostic technique in this clinical context, it may increase the potential for curative treatment and thus reduce complications. In addition to the potential for improved patient outcome, the fast and noninvasive nature of this diagnostic technique may also reduce health care costs.

Steady-state fluorescence measurements from small tissue regions (less than a few millimeters in diameter) as well as steady-state fluorescence imaging of relatively large tissue fields (a few centimeters in diameter) have been performed.[11,12] To a much lesser extent, time-resolved fluorescence measurements are being explored currently.[11,12] Furthermore, sources of both intrinsic (endogenous fluorophores) and extrinsic (exogenous fluorophores) fluorescence have been considered.[11,12] The advantage of using exogenous fluorophores is that the photophysical and pharmacokinetic properties can

be selected and are known. Furthermore, exogenous fluorophores are more highly fluorescent than endogenous fluorophores. On the other hand, the disadvantage of using exogenous fluorophores is that issues relating to safety and toxicity of the drug being used have to be addressed. Furthermore, the selection of the optimal time delay after administration of the drug is nontrivial.

Several authors have reviewed the progress in this field over the last decade. Andersson-Engels et al.[5,6] discussed the possibilities of using fluorescence spectroscopy for tissue diagnosis. The tissue types discussed were malignant tumors and atherosclerotic lesions. Studies of endogenous fluorescence as well as exogenous fluorescence with contrast agents were also presented. In a subsequent review, Andersson-Engels et al.[7] reviewed the scientific, technical and practical issues related to the use of fluorescence spectroscopy for tissue diagnosis in clinical oncological applications. Papazoglou[8] presented a review on the diagnosis of malignancies and atherosclerotic plaques using fluorescence spectroscopy. Special emphasis was given to problems that were raised during clinical trials and recent experimental studies, such as the identification of the fluorescent chemical species and the determination of the illumination/collection geometry for fluorescence spectroscopy of tissue. Bigio et al.[9] reviewed fluorescence spectroscopy and elastic scattering spectroscopy of small regions of tissue, whereas Andersson-Engels et al.[10] focused on fluorescence spectroscopy of relatively large tissue fields. Richards-Kortum et al.[11] and Wagnieres et al.[12] have written more comprehensive reviews. The former review focuses on the quantitative aspects of fluorescence spectroscopy. Specifically, this review describes optical interactions pursued for biomedical applications, provides a descriptive framework for light interactions in tissue and, finally, reviews the important endogenous and exogenous molecules and how they are exploited for tissue diagnosis. The latter review, which focuses more on clinical aspects, presents a critical status report on the diagnosis of neoplastic tissues in vivo using fluorescence spectroscopy.

5.1 Neoplasia

5.1.1 Endogenous Fluorescence

The phenomenon of fluorescence was first observed by Stokes.[73] Much later, Stubel recognized the diagnostic potential of tissue fluorescence.[74] Policard[47] observed red fluorescence when examining necrotic tumors under illumination with UV and VIS light. Ronchese[48] demonstrated in 1954 that ulceration is essential for the production of red fluorescence in human cutaneous, squamous cell carcinoma. The observed fluorescence was attributed to endogenous porphyrins in the tissue. Ghadially et al.[39,49–51] also reported that ulcerated squamous carcinomas exhibit a red fluorescence when exposed to UV light. They concluded that this red fluorescence may be due in part to the action of bacteria on a protoporphyrin precursor. In 1965, Lycette et al.[75] suggested that fluorescence spectroscopy could be used to discriminate between normal tissues and malignant tumors. Fluorescence emission spectra were recorded from excised normal tissue and malignant tumors of the esophagus, stomach, breast and thyroid at 330 nm excitation. All tissues fluoresced in the range 360–600 nm. It was found that the fluorescence intensities of malignant tumors were less than that of normal tissue from the same patient. Subsequently, the groups of Profio,[76,77] Alfano,[78] Lohmann[79] and Yang[80] did pioneering studies on in vitro and in vivo fluorescence spectroscopy of neoplastic and nonneoplastic animal and human tissues.

Fluorescence spectroscopy has been used to evaluate neoplastic and nonneoplastic tissues in vitro and in vivo. Generally, the results of in vitro studies have been used to guide the design of experimental parameters for in vivo studies.[81] However, in vitro studies can introduce significant artifacts due to hemoglobin reabsorption and oxidation of certain fluorophores,[82] questioning the relevance of in vitro studies to in vivo studies. Richards-Kortum et al.[11] review in vitro and in vivo studies on fluorescence spectroscopy of neoplasia and in vitro studies on fluorescence spectroscopy of atherosclerosis. This review covers only in vivo studies, in order to demonstrate the diagnostic potential of fluorescence spectroscopy in clinical applications.

UV/VIS fluorescence spectroscopy can be developed and employed to differentiate diseased from nondiseased tissues in vivo. The altered biochemical and morphological state that occurs as tissue progresses from a nondiseased to a diseased state is reflected in the spectral characteristics of the measured fluorescence. This spectral information can be compared to tissue histology – the current gold standard – which indicates the absence or presence and grade of disease. Mathematical algorithms then can be developed and optimized to classify tissues into their respective histological category based on their spectral features. These mathematical algorithms can be implemented in software, thereby potentially enabling fast, noninvasive, automated screening and diagnosis in a clinical setting. Although the signal-to-noise ratio of fluorescence spectroscopy is several orders of magnitude lower than that associated with absorption spectroscopy over the same wavelength range, this technique affords higher sensitivity and specificity for biochemical characterization.

Table 2[83–100] summarizes the key aspects of a representative set of clinical applications of fluorescence spectroscopy (the most recent investigation per group and investigations from groups who reported the performance

Table 2 Fluorescence spectroscopy of neoplastic and nonneoplastic tissues in vivo obtained from a representative set of clinical applications[a]

Tissue	Excitation wavelength (nm)	Measurement type	Dimension reduction: emission wavelengths (nm)	Classification: decision scheme	Sample size ND; D	SE; SP (%)
Colon[83]	337	Spectra	MVLR (spectra)	Binary	121; 56	80; 92
Colon[84]	337	Spectra	MVLR (spectra)	Binary	182; 32	100; 84
Colon[85]	370	Spectra	I (460, 600, 680)	Bayes theorem	67; 32	94; 91[b]
Colon[86]	360	Images	I (>400)	Binary	12; N/A	83; N/A
Colon[87]	337	Transient decay	τ (550)	Binary	13; 11	85; 87
Cervix[88]	337, 380, 460	Spectra	PCA (spectra)	Bayes theorem	122; 59	82; 68[b]
Cervix[88,89]	337, 380, 460	Spectra	PCA of 15 $I(\lambda_x, \lambda_m)$	Bayes theorem	122; 59	84; 65[b]
			15 $I(\lambda_x, \lambda_m)$	Neural network		91; 67[b]
Bronchus[90]	442	Images	I (red, green)	Binary	558; 142	67; N/A[c]
Bladder[91]	337	Spectra	I (385, 455)	Binary	85; 29	97; 98[d]
Bladder[92]	308	Spectra	I (360, 440)	Binary	35; 31	100; 100[d]
Esophagus[93]	410	Spectra	I (480, 680)	Binary	252; 56	40; 96
Oral cavity[94]	370	Spectra	I (490)	Binary		100; 87.5
	410		I (640)			
Oral cavity[95]	337	Spectra	I (peak)	Binary	17; 11	94; 100
	410		I (635, 490)			
Head and neck[96]	442	Images	I (red, green)	Binary	16; 16	100; 87.5[d]
Head and neck[97]	375–440	Images, spectra	I (>515)	Binary	36; N/A	43.3; N/A[d]
					36; N/A	94.4; N/A[d]
Larynx[98]	442	Images	I (red, green)	Binary	328 total	72.5; 94[c,d]
Larynx[99]	442	Images	I (red, green)	Binary	87; 28	85; 87[c,d]
Skin[100]	336	Images	I (475)	Binary	318; 90[d]	83; 79

[a] Acronyms: I = intensity; τ = decay time; MVLR = multivariate linear regression; PCA = principal component analysis; λ_x = excitation wavelength; λ_m = emission wavelength; ND = nondiseased; D = diseased; SE = sensitivity; SP = specificity; N/A = not applicable.
[b] Sensitivity and specificity were evaluated prospectively.
[c] Sensitivity and specificity were evaluated for a combination of fluorescence and conventional endoscopic techniques.
[d] Sensitivity and specificity were calculated for the discrimination between cancers and noncancers (other investigations calculated the sensitivity and specificity for discriminating precancers and cancers from normal tissues).

of their technique). In these applications, fluorescence spectroscopy in the UV/VIS spectral regions was used for the detection of neoplastic tissues in vivo. In particular, Table 2 provides the organ sites studied, the excitation wavelength(s) used, the type of measurements that were made, the dimensionally reduced spectral variables and the corresponding classification scheme used in the mathematical algorithm, the sample size for the diseased and nondiseased populations and the corresponding sensitivity and specificity. Note that in all these clinical applications only the endogenous fluorescence, absorption and scattering properties of the tissue were exploited.

The neoplastic tissues that were evaluated spectroscopically are from the colon,[83–87] cervix,[88,89] bronchus,[90] bladder,[91,92] brain[101] (preliminary studies only), esophagus,[93] oral cavity,[94,95] head and neck,[96,97] larynx,[98,99] skin,[100] bile duct[102] (preliminary studies only), stomach[103] (in vitro studies only) and breast tissues[104–107] (in vitro studies only). The excitation wavelengths that were selected correspond to those used to excite fluorophores in the UV/VIS spectral regions (see Table 1). Most groups measured fluorescence emission spectra from tissue sites that are 1–2 mm in diameter. Only a few groups measured fluorescence images from tissue regions that are a few centimeters in diameter, and only one group measured the transient fluorescence decay from colon tissues in vivo.

There are generally two steps involved in the development of a mathematical algorithm that is based on fluorescence spectroscopy. The first part is to reduce dimensionally the measured spectral variables, and the second step is to develop a classification scheme for the discrimination of these useful spectral parameters into relevant histological/histopathological categories. The development of current mathematical algorithms based on fluorescence spectroscopy can be classified broadly into three categories: algorithms based on qualitatively selected spectral variables (fluorescence intensities at several emission wavelengths); algorithms based on statistically selected spectral parameters (a more robust use of all the measured spectral information); and algorithms based on parameters that reflect the biochemical and/or morphological features of the tissue. Classification schemes have been primarily based on a binary discrimination or a probability-based classification scheme.

Mostly, algorithms have been based on qualitatively or statistically selected spectral variables with binary classification schemes as indicated in Table 2.

The sensitivity is defined as the fraction of diseased samples correctly classified, and the specificity is defined as the fraction of nondiseased tissues correctly classified. In several clinical studies, the sensitivity and specificity were evaluated prospectively rather than retrospectively to obtain an unbiased estimate of the performance of this technique.[85,88,89] In some cases the sensitivity and specificity were evaluated for a combination of fluorescence spectroscopy and conventional endoscopy.[90,98,99] Finally, in the case of the bladder, head and neck and larynx, the sensitivity and specificity are reported for the discrimination between cancers and noncancers.[91,92,96–99] In other clinical applications, the sensitivity and specificity are reported for the discrimination of precancers and cancers from normal tissues. In the majority of clinical studies performed the sensitivity and specificity are greater than 80%, reflecting the high classification accuracy of fluorescence spectroscopy for the detection of neoplastic tissues in vivo. The sensitivity and specificity reported here are similar or superior to current clinical modalities that are used routinely.

5.1.2 Exogenous Fluorescence

There are a number of groups that have explored the diagnostic potential of fluorescence spectroscopy of exogenous fluorophores (particularly photosensitizing agents developed for PDT) in tissues in vivo.[12] Auler et al.[108] and Figgie[109] first observed red fluorescence from animal tumors after administration of exogenous porphyrins. The first use of fluorescin to improve the detection and identification of brain tumors in vivo was reported by Moore et al.[110] The accumulation of hematoporphyrin in various types of cancers was discovered and exploited during the 1950s.[111–113] HpDs were evaluated subsequently for localization in cancers of various organs, including the esophagus, tracheobronchial tree and cervix,[114–120] and for characterizing suspicious head and neck lesions.[121] The first cystoscopic observations of HpD fluorescence in urinary bladder tumors were reported by Kelly et al.[122] These and other initial clinical and animal model studies of HpD-mediated PDT led to fluorescence diagnostic applications by many other groups using this exogenous fluorophore. Subsequently, Kreigmair et al.[123] proposed the use of 5-ALA-induced PpIX fluorescence for the detection of bladder neoplasms.

Photosensitizing agents, particularly HpD and 5-ALA, have been used as contrast agents for fluorescence spectroscopy of neoplastic tissues in a wide variety of tissue sites, including the skin, bladder, bronchus, colon, esophagus, head and neck, and breast.[12] However, because most of these exogenous fluorophores rely on the differences in the vasculature between diseased tissues (which generally exhibit leaky vessels) relative to their nondiseased counterparts, it remains to be seen whether the characteristics that determine localization of the exogenous fluorophore in cancers are also present in precancerous tissues.

5.2 Atherosclerosis

Edholm et al.[124] suggested that quantitative optical techniques could be used to improve the detection of atherosclerosis in vivo; they measured the reflectance of the aorta at 500 and 550 nm in 15 patients undergoing aortography. Kitrell et al.[125] demonstrated that fluorescence spectroscopy could be used to discriminate between normal aorta and fibrous plaque, and suggested that optical diagnosis (using low-power illumination) and therapy (using high-power illumination) of atherosclerosis could be combined in a single fiber-optic device. Subsequently, a number of groups investigated the utility of fluorescence spectroscopy for the diagnosis of atherosclerotic plaque, many with the goal of developing a guidance system for laser angiosurgery catheters.[8,11]

The fluorescence emission spectra of normal artery tissue and atherosclerotic plaques have been measured over the entire range of UV/VIS excitation wavelengths.[11] However, most of these investigational studies to demonstrate the diagnostic potential of fluorescence spectroscopy have been performed in vitro on excised tissues and hence have not addressed the problems of making these measurements in vivo on the intact artery wall.

6 CLINICAL INSTRUMENTS

In this section representative examples of instruments used in clinical applications of fluorescence spectroscopy are described. In particular, details of the instruments and typical data are presented to provide the reader with an understanding of the main characteristics of instruments for UV/VIS fluorescence spectroscopy.

6.1 Single-pixel, Three-excitation-wavelength Fluorimeter

Although fluorescence emission spectra of normal tissue, dysplasia and invasive carcinoma have been measured previously from several organ sites in vivo at single excitation wavelengths (see Table 2), Ramanujam et al.[88] were among the first to address the necessity to measure fluorescence emission spectra at several excitation

wavelengths in order to exploit more fully the tissue biochemistry and morphology. They developed a portable fluorimeter that can be used to measure fluorescence emission spectra from 1-mm-diameter human cervical tissue sites in vivo at three excitation wavelengths in the UV/VIS spectral regions. These wavelengths were selected primarily because they coincide with the absorption bands of biologically relevant fluorophores, namely the metabolic coenzymes and structural proteins and chromophores, which include oxygenated and deoxygenated hemoglobin. A schematic of the fluorimeter is shown in Figure 14.

Two nitrogen pumped-dye lasers (5 ns pulse duration, 30 Hz repetition rate) are used to provide illumination at three different excitation wavelengths: one laser generates light at 337 nm (fundamental) and has a dye module in a resonant cavity that can be used to generate light at 380 nm using the fluorescent dye BBQ. The dye module in the resonant cavity of the second laser is used to generate light at 460 nm, using the fluorescent dye Coumarin 460. Laser illumination at each excitation wavelength of 337, 380 and 460 nm is coupled to each of three excitation fibers in a fiber-optic probe. Note that two 10-nm bandpass filters – one centered at 380 nm and the other centered at 460 nm – are placed between the excitation fiber and the two dye modules, to prevent leakage from the fundamental at 337 nm.

The fiber-optic probe consists of a central fiber surrounded by a circular array of six fibers; all seven fibers have the same characteristics (0.22 numerical aperture, 200 μm core diameter, 245 μm diameter with cladding). Three fibers along the diameter of the distal end of the probe are used for excitation light delivery. The purpose of the remaining four fibers is to collect the emitted light from the tissue area directly illuminated by the excitation light.[57,58] A quartz shield (3 mm in diameter and 2 mm thick) at the tip of the distal end of the probe, which is in direct tissue contact, provides a fixed distance between the fibers and the tissue surface so that the fluorescence intensities can be measured in calibrated units. A tissue area that is 1 mm in diameter is illuminated by each excitation delivery fiber. The average energies per pulse on the tissue surface at 337, 380 and 460 nm excitation were 15.2, 11.5 and 18 $\mu J\,mm^{-2}$, respectively, in this case.

The proximal ends of the four emission collection fibers are arranged in a circular array and imaged at the 500-μm-wide entrance slit of an f 3.8 spectrograph equipped with a 300 $grooves\,mm^{-1}$ grating coupled to a 1024-intensified PDA controlled by a multichannel analyzer. The collection optics between the proximal end of the four fibers and the spectrograph are two, fused-silica, planoconvex lenses. Between these lenses is a filter wheel assembly containing long-pass filters with 50% transmission at 360, 400 and 475 nm, which are used to reject backscattered excitation light at 337, 380 and 460 nm excitation, respectively, from the detector. The nitrogen pumped-dye lasers are used for external triggering of a pulser, which serves to synchronize the 200-ns collection gate of the detector to the leading edge of the laser pulse. The use of gated detection in conjunction with pulsed excitation eliminates the effects of ambient light during fluorescence measurements. The total time required to record fluorescence emission spectra at all three excitation wavelengths from one tissue site was approximately 5 s. Spectra were collected in the VIS spectral region, with a resolution of 10 nm (fwhm)

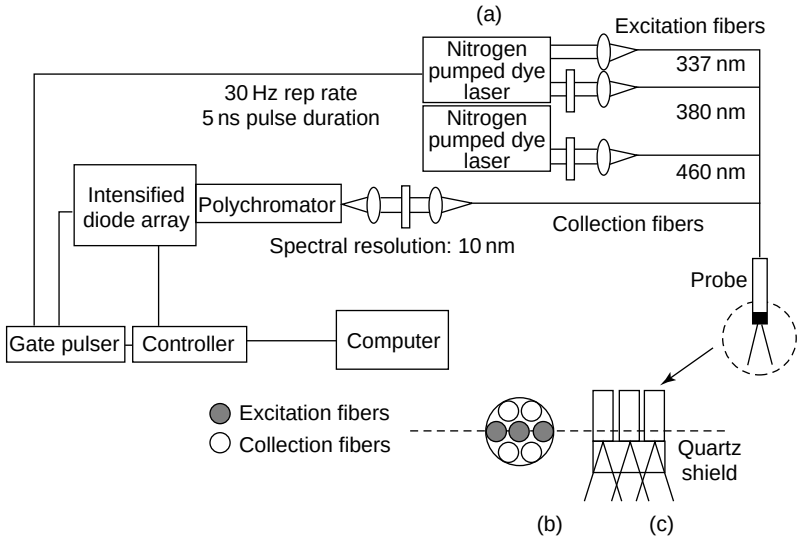

Figure 14 Schematic of a three-excitation-wavelength system. (Reproduced by permission of the American Society for Photobiology, from Ramanujam et al.[88])

and a signal-to-noise ratio of 100 : 1 at the fluorescence maximum at each excitation wavelength. Data acquisition and storage are achieved using a computer.

All spectra were corrected for the nonuniform spectral response of the detection system using correction factors obtained by recording the spectrum of a National Institute of Standards and Technology (NIST) traceable, calibrated, tungsten ribbon filament lamp. Spectra from each cervical tissue site at each excitation wavelength were averaged to obtain a single spectrum per site. The average tissue spectra were then normalized to the average peak fluorescence intensity of the rhodamine 610 calibration standard at the corresponding excitation wavelength for that patient; absolute fluorescence intensities are reported in these calibrated units.

Figure 15 illustrates average fluorescence emission spectra per site acquired from cervical tissue sites at 337, 380 and 460 nm excitation from a typical patient. All fluorescence intensities are reported in the same set of calibrated units. Evaluation of the tissue spectra at 337 nm excitation indicates that the fluorescence intensity of the precancerous SIL is less than that of the corresponding normal squamous tissue and greater than that of the corresponding normal columnar tissue. Evaluation of the spectra at 380 nm excitation indicates that the fluorescence intensity of SIL is less than that of the corresponding normal squamous tissue, with the low-grade SIL exhibiting the weakest fluorescence intensity. Note that the fluorescence intensity of the normal columnar tissue is indistinguishable from that of the high-grade SIL. Finally, evaluation of spectra at 460 nm excitation indicates that the fluorescence intensity of SIL is less than that of the corresponding normal squamous tissue and greater than that of the corresponding normal columnar tissue. The spectra shown here demonstrate that it is possible to obtain fluorescence emission spectra from tissues in vivo at several excitation wavelengths, with a high signal-to-noise ratio and in a fast and noninvasive manner.

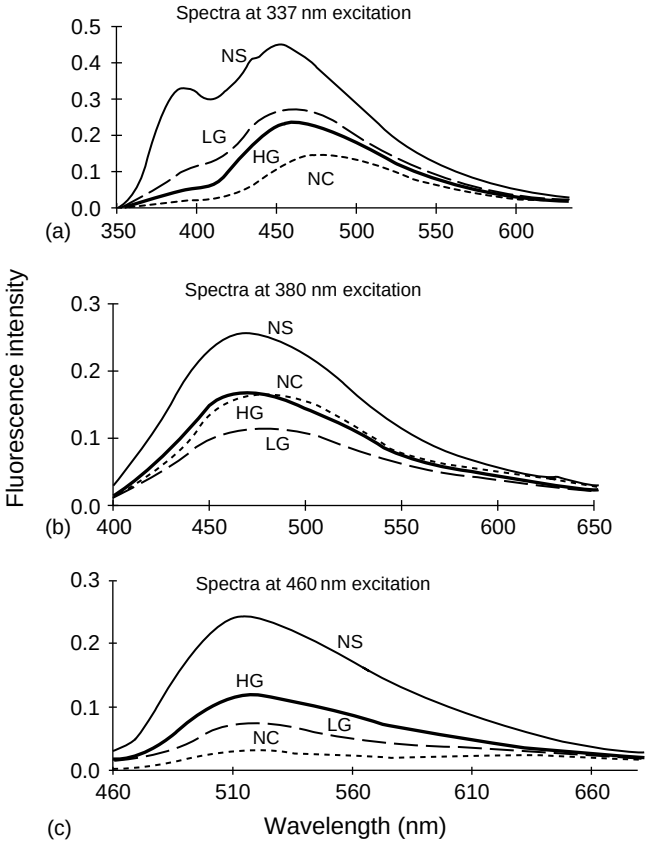

Figure 15 Fluorescence spectra measured at 337, 380 and 460 nm excitation from the cervix of a typical patient using the three-excitation-wavelength fluorimeter. (Reproduced by permission of the American Society for Photobiology, from Ramanujam et al.[88])

6.2 Single-pixel, Excitation–Emission Matrix System

It is apparent that fluorescence emission spectra at multiple excitation wavelengths spanning the UV/VIS spectral regions are needed to characterize properly the patho-physiologically relevant, endogenous fluorophores in tissue. Furthermore, reflectance measurements at these wavelengths are also needed to probe tissue absorption and scattering properties and enable deconvolution between the fluorescence, absorption and scattering.[126,127] Zeng et al.[128] developed an instrument to measure the fluorescence emission spectra and diffuse reflectance spectra from the human skin, simultaneously. However, this instrument was capable of measuring fluorescence emission spectra at only a single excitation wavelength of 380 nm.

Zangaro et al.[69] integrated and improved upon the designs of Ramanujam et al.[88] and Zeng et al.[128] by constructing a fast EEM system that can measure fluorescence emission spectra at multiple excitation wavelengths and diffuse reflectance spectra from 1- to 2-mm-diameter tissue sites in vivo quickly and noninvasively. Preliminary considerations for the design of this system were as follows: rapid wavelength tunability and real-time data acquisition for fluorescence spectroscopy at multiple excitation wavelengths; the ability to measure diffuse reflectance in addition to fluorescence; gated detection of spectral measurement with minimal interference from ambient light; and capability for optical fiber delivery and collection. The system, which was designed with the aforementioned considerations in mind, is shown in Figure 16.

A nitrogen pumped-dye laser is used as the excitation light source. Delivery of light to tissue and collection of returned spectral information is accomplished by

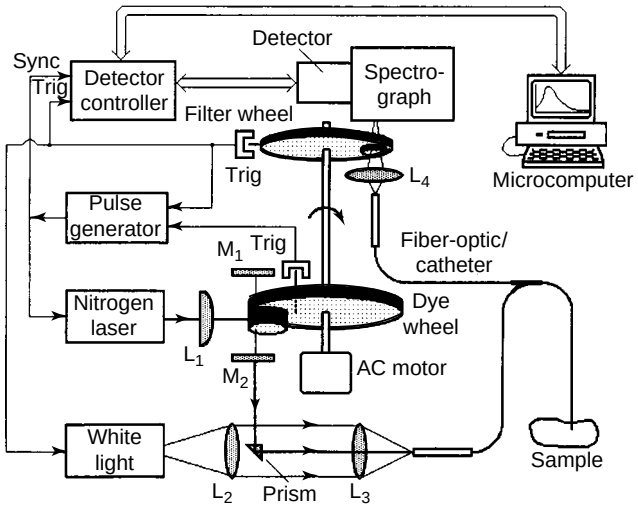

Figure 16 EEM system designed to measure tissue fluorescence spectra at multiple excitation wavelengths and spectrally resolved broad band diffuse reflectance spectra in vivo. (Reproduced by permission of the Optical Society of America, from Zangaro et al.[69])

means of an optical fiber probe. A filter wheel placed in front of the spectrograph rejects backscattered excitation light. The fluorescence is dispersed in the spectrograph, detected by an intensified PDA, coupled to an optical multichannel analyzer and stored in a computer. The novel aspects of this system, which lends itself to multi-excitation fluorescence spectroscopy and diffuse reflectance spectroscopy of tissues in the UV/VIS spectral regions, are described below.

Dye cuvettes (with appropriate dyes) mounted on a rotating wheel that rapidly traverses a single resonant cavity are capable of generating multiple wavelengths. Also, a corresponding filter wheel placed in front of the spectrograph allows rejection of backscattered excitation light at each wavelength when synchronized with the dye wheel. Furthermore, a 10-μs pulsed xenon lamp is also incorporated such that diffuse reflectance measurements can be made over the UV/VIS spectral regions. Fiber-optic delivery and collection have been designed to provide for the remote operation of the system and for the measurement of fluorescence and diffuse reflectance directly from the tissue area that is illuminated, in order to remove the effects of the illumination/collection geometry.[57,58] Furthermore, the fiber-optic probe was designed such that the diffuse reflectance and fluorescence can be made from the same tissue site. The acquisition time for both fluorescence and reflectance measurements is less than 1 s.

The performance of this system has been validated clinically by passing the fiber-optic probe through one of the accessory channels of a colonoscope and bringing

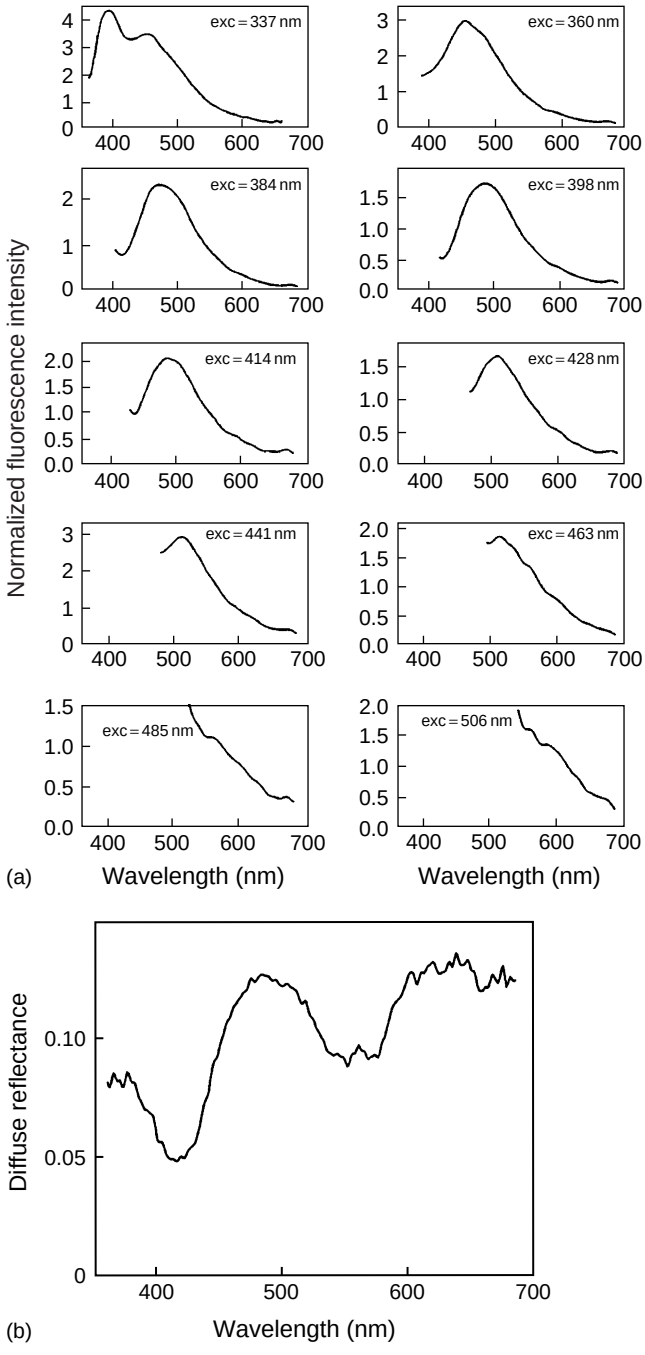

Figure 17 (a) Fluorescence spectra at 10 excitation wavelengths and (b) diffuse reflectance spectrum over the UV/VIS spectral regions measured from a colon tissue site in vivo. (Reproduced by permission of the Optical Society of America, from Zangaro et al.[69])

it into flush contact with the tissue surface. Figure 17(a) displays fluorescence emission spectra at multiple excitation wavelengths and Figure 17(b) displays a diffuse reflectance spectrum from colon tissue. Figure 17(a) shows spectra measured at 10 excitation wavelengths in

the UV/VIS spectral regions. The spectra at the different excitation wavelengths appear to correspond with spectra of biologically relevant fluorophores, including collagen (337 nm excitation), reduced PN (360 nm excitation) and oxidized Fp (428 nm excitation). Note that the effects of hemoglobin absorption on the fluorescence emission spectra are observed at 337 nm excitation. Figure 17(b) shows the diffuse reflectance spectrum obtained from the same tissue site. The spectrum exhibits the absorption bands of hemoglobin at 420 nm (Soret band) and at 540 and 580 nm (α and β bands).

Zuluaga et al.[129] have also developed a fast EEM system that measures fluorescence emission spectra at multiple excitation wavelengths and diffuse reflectance spectra rapidly (4 min total measurement time) and noninvasively from tissues in vivo. Specifically, the system measures the fluorescence EEM at 330–500 nm excitation and 380–700 nm emission. The diffuse reflectance spectra are measured at 380–950 nm. The primary difference between this system and that developed by Zangaro et al.[69] is that it employs a white light source with a series of bandpass filters instead of a nitrogen pumped-dye laser for generating multiple excitation wavelengths for fluorescence spectroscopy. Furthermore, Zuluaga et al.[129] present a method based on autocorrelation vectors to identify the excitation and emission wavelengths where the spectra of diseased and nondiseased tissues differ the most.

6.3 Fluorescence Imaging Systems

Several groups have developed endoscopic-compatible[86,97,130–134] and nonendoscopic-based[100,135,136] fluorescence imaging systems. The most notable is that originally developed by Palcic et al. for fluorescence bronchoscopy,[130] which has now led to a commercial light-induced fluorescence endoscopy (LIFE) device (Xillix Technologies Corporation, Richmond, BC, Canada) that is used for fluorescence imaging of relatively large tissue fields (a few centimeters in diameter). The LIFE device consists of a white light source and a color CCD for the acquisition of white light images, a helium–cadmium laser and two filtered, high-sensitivity CCD cameras for the acquisition of green and red fluorescence images, an endoscope-compatible fiber-optic bundle and a computer with a monitor. The laser is used to provide excitation light at 442 nm. The fluorescence emitted is collected and coupled via an imaging, fiber-optic bundle to two filtered, image-intensified CCD cameras (one for the green wavelength band at 480–520 nm and one for the red wavelength band at 630 nm and longer). The fluorescence image is digitized using an imaging board in the computer. Using a mathematical transformation (a nonlinear discriminant function combination of the red and green intensity images), a pseudo-image of the observed tissue site is obtained and displayed on the monitor in real-time. Normal tissues appear green and neoplastic tissues appear brownish-red.

The LIFE device has been tested in a multicenter clinical trial to evaluate if fluorescence bronchoscopy, when used as an adjunct to white light bronchoscopy, can improve the bronchoscopist's ability to locate areas suspicious of dysplasia for biopsy and histological examination as compared to white light bronchoscopy alone.[90] The LIFE device also has been used successfully to evaluate neoplastic lesions in other organ sites, including the head and neck,[96] larynx,[98,99] bile duct,[102] and gastrointestinal tract.[137,138]

Recently, Zeng et al.[134] elaborated on the fluorescence imaging system for the gastrointestinal tract. The device consists of a mercury arc lamp, two intensified CCD cameras, a fiber-optic endoscope and a computer-based console. The system is capable of working in three different modalities: conventional white light imaging mode; light-induced fluorescence imaging mode, based on fluorescence imaging of two wavelength bands in the red and green; and light-induced fluorescence and reflectance imaging mode, based on the combination of red and green fluorescence images and a red-NIR reflectance image.

7 METHODS OF ANALYSIS

Most screening and diagnostic algorithms developed from fluorescence spectroscopy of tissues incorporate qualitatively or statistically selected spectral variables, which are evaluated using a binary or probability-based classification scheme. For example, Panjehpour et al.[93] have developed an algorithm that uses qualitatively selected, fluorescence intensities at several emission wavelengths in a binary classification scheme for the detection of Barrett's esophagus in vivo. On the other hand, Ramanujam et al.[88] have developed an algorithm that uses statistically selected spectral variables and probability-based classification for cervical precancer detection in vivo. This multivariate statistical algorithm employs PCA[139] to reduce dimensionally the preprocessed tissue fluorescence emission spectra into orthogonal principal components with minimal information loss. Bayes theorem[140] is then used to develop probability-based classification using the diagnostically relevant principal components. The advantage of using statistical rather than qualitative analysis of the tissue fluorescence emission spectra is that the entire spectral information content can be exploited. Furthermore, the benefit of using a probability, rather than a binary, classification scheme is that the likelihood of the classification being correct is also provided.

An example of an algorithm that uses biochemical and/or morphological features that are related to the tissue fluorescence emission spectra, coupled with a probability-based classification, is that developed by Richards-Kortum et al.[141] This algorithm discriminates between normal coronary arteries and noncalcified and calcified atherosclerotic plaque in vitro. The contribution of the biochemical and/or morphological features is extracted from the tissue fluorescence emission spectra via an analytical model of tissue fluorescence, based on the exponential attenuation of light in an absorption-dominant medium. Bayes' theorem is then used for probability classification.

The advantage of using a physically based model over a statistically based model is that the former method provides insight into the biochemical and morphological basis. However, there are notable advantages to using a statistical model such as PCA. Although it is a linear method of analysis, PCA can still be used to model effectively the nonlinear turbid tissue fluorescence. Because it is less restrictive, PCA can permit a better fit to the fluorescence emission spectra than simple physical models can, through the use of a linear combination of orthogonal principal components.[142] In addition, the principal components are not correlated with each other and therefore can be used in a variety of classification algorithms that generally require un-correlated variables. Furthermore, each principal component can be correlated to the spectral variables of the original tissue fluorescence emission spectra, thus providing insight into the spectral features that contribute to the classification.

Most algorithms to date use qualitatively and, to a lesser extent, statistically selected variables with binary or probability-based classification schemes as indicated in Table 2. The development of a physically based model that uses biochemical and morphological features that are related to the measured tissue fluorescence emission spectra has been hampered by the fact that fluorescence spectroscopy of human tissue is greatly affected by the absorption and scattering of the excitation light and the emitted light, making interpretation of the measured spectral information challenging.[53]

Given the high classification accuracy that can be achieved using statistically based algorithms,[88] and the difficulties associated with the development of physically based algorithms,[141] it is perhaps worthwhile considering the development of a hybrid algorithm that incorporates the key features of both. For example, the statistical model could be related to the physical model in order to realize the biochemical and/or morphological basis of the statistically selected spectral variables that are used for classification purposes.

Section 7.1 provides a detailed description of the algorithm developed by Ramanujam et al.,[88] which uses statistically selected spectral parameters in conjunction with a probability based classification process. Section 7.2 discusses physically based models to analyze tissue fluorescence emission spectra and their potential to elucidate the biochemical and morphological features that contribute to the tissue fluorescence emission spectra. Section 7.3 presents various approaches to account for the illumination/collection geometry-related distortion of the tissue fluorescence emission spectra measured in vivo.

7.1 Statistically Based Model

Ramanujam et al.[88] developed a formal analytical process for the development of screening and diagnostic algorithms for the detection of human cervical precancer or SIL in vivo. The formal analytical process is displayed in Figure 18, where the text in the dashed-line boxes represents the mathematical steps implemented on the spectral data, and the text in the solid-line boxes represents the output after each mathematical process. There are four primary steps involved in the multivariate statistical analysis of tissue fluorescence emission spectra. The first step is to preprocess spectral data to reduce interpatient and intrapatient variation within a tissue type; the preprocessed spectra are then dimensionally reduced to an informative set of principal components that describe most of the variance of the original spectral data set using PCA.[139] Next, the principal components that contain diagnostically relevant information are selected using an unpaired, one-sided Student's *t*-test, and finally a classification algorithm based on Bayes theorem[140] is developed using these diagnostically relevant principal components.

In summary, three constituent algorithms were developed using multivariate statistical analysis: constituent algorithm 1 discriminates between SIL and normal squamous tissues; constituent algorithm 2 discriminates between SIL and normal columnar tissues; and algorithm 3 differentiates high-grade SIL from low-grade SIL. The three constituent algorithms were then combined to develop two composite algorithms: constituent algorithms 1 and 2 were combined to develop a composite screening algorithm that discriminates between SIL and non-SIL. All three constituent algorithms were then combined to develop a composite diagnostic algorithm that differentiates high-grade SIL from non-high grade SIL.

Inputs into the multivariate statistical algorithm included the preprocessed fluorescence spectra at all three excitation wavelengths (full-parameter) and fluorescence intensities at a reduced number of excitation–emission wavelength pairs (15 reduced parameters) selected from the component loadings calculated from PCA.[139] The

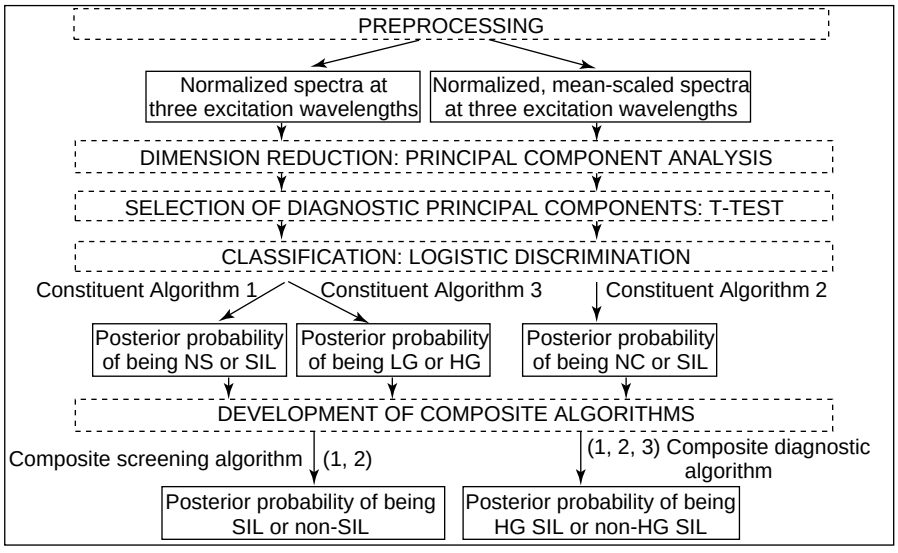

Figure 18 Formal analytical process for the development of screening and diagnostic algorithms for the differential detection of cervical pre-cancer or SILs. NS = normal squamous; NC = normal columnar; LG = low grade; HG = high grade. (Reproduced by permission of the American Society for Photobiology, from Ramanujam et al.[88])

Table 3 Comparison of accuracy of composite screening and diagnostic (full and reduced-parameter) algorithms to that of Pap smear screening and colposcopy in expert hands

Classification	SIL *v.* Non-SIL		HG SIL *v.* Non-HG SIL	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Pap smear screening	62 ± 23	68 ± 21	N/A	N/A
Colposcopy in expert hands	94 ± 6	48 ± 23	79 ± 23	76 ± 13
Full-parameter composite algorithm	82 ± 1.4	68 ± 0.0	79 ± 2	78 ± 6
Reduced-parameter composite algorithm	84 ± 1.5	65 ± 2	78 ± 0.7	74 ± 2

HG = high grade; N/A = not applicable.

algorithm was developed on a calibration set and tested on a prediction set of approximately equal numbers of samples in each tissue category.

Table 3 compares the sensitivity and specificity of the composite screening and diagnostic algorithms with that of Pap smear screening[143] and colposcopy in expert hands.[144] Table 3 indicates that the composite screening algorithm has a similar specificity and a significantly improved sensitivity relative to Pap smear screening. A comparison of the composite screening algorithm to that of colposcopy in expert hands for differentiating SIL from non-SIL indicates that this algorithm demonstrates a 10% decrease in sensitivity but a 20% improvement in specificity. The composite diagnostic algorithm and colposcopy in expert hands both discriminate high-grade SIL from non-high-grade SIL with a similar sensitivity and specificity. Finally, the sensitivity and specificity of the reduced-parameter algorithms that use an order of magnitude fewer variables is within 5% of the sensitivity and specificity reported for the full-parameter algorithms. This has important implications in using fluorescence spectroscopy to identify fluorescence intensities at a reduced number of optimal excitation–emission wavelength pairs for the implementation of fluorescence imaging.

7.2 Physically Based Models

Although it has been demonstrated that fluorescence spectroscopy can be used to differentiate diseased from nondiseased tissues with high sensitivity and specificity, the underlying biochemical and morphological basis for the spectral differences is poorly understood. This has been hampered by the fact that fluorescence spectroscopy of a turbid medium such as tissue is complicated by its absorption and scattering properties.[53]

A fluorescence emission spectrum measured from tissue over the UV/VIS spectral region is primarily attributed to the superposition of the fluorescence of a variety of biological molecules that contain naturally occurring fluorophores. Fluorescence emission spectra

measured from tissues also contain information about the absorbing and scattering properties of that medium. Quantification of the concentration and distribution of fluorophores within tissue in principle involves: deconvoluting the absorption and scattering properties of the tissue from the measured fluorescence emission spectrum; and quantifying the identity and distribution of the fluorophores that contribute to the deconvolved spectrum. The former step requires the development of a transfer function based on the measurements of tissue optical properties, whereas the latter step requires knowledge of the identity and distribution of the fluorophores within the tissue.

7.2.1 Transfer Function

Deconvolution of the absorption and scattering properties of the tissue from the measured fluorescence emission spectrum requires knowledge of the tissue optical properties, i.e. the absorption coefficient, the scattering coefficient and the anisotropy parameter. Analytical methods based on Kubelka–Munk theory[61] and the adding–doubling method[62] have been developed to calculate tissue optical properties in the UV/VIS spectral range. However, these models require the measurement of diffuse transmittance and reflectance from tissues. Diffuse transmittance measurements cannot be made from tissues in vivo, thus limiting the number of parameters needed to determine tissue optical properties using these simple analytical techniques.

Several groups have developed simple analytical models based on measurements of fluorescence and diffuse reflectance only (no transmittance measurements are needed) to deconvolve absorption and scattering properties of the tissue from the measured fluorescence emission spectrum.[126,127] These analytical techniques, although simplistic, do provide a first step towards quantifying the biochemical and morphological characteristics of tissue fluorescence emission spectra.

In particular, Gardner et al.[127] have developed an analytical expression for recovering the intrinsic fluorescence coefficient (which is defined as the product of the fluorophore absorption coefficient and the fluorescence quantum yield) of a homogeneous turbid medium from a surface measurement of fluorescence and diffuse reflectance. The intrinsic fluorescence coefficient $\beta(\lambda_x, \lambda_m)$ is defined simply by Equation (10):

$$\beta(\lambda_x, \lambda_m) = \mu_a(\lambda_x)\phi(\lambda_m) \tag{10}$$

where λ_x is the excitation wavelength, λ_m is the emission wavelength, $\mu_a(\lambda_x)$ is the absorption coefficient of the fluorophore at the excitation wavelength and $\phi(\lambda_m)$ is its fluorescence quantum yield at the emission wavelength.

Using the analytical method developed by Gardner et al., one can recover the intrinsic fluorescence coefficient from the measured fluorescence using Equation (11):

$$\beta(\lambda_x, \lambda_m) = \frac{F(\lambda_x, \lambda_m)}{P_0(\lambda_x)[(\Delta\Omega/\pi)\cos\theta]D(\lambda_m)X_{1D}(\lambda_x, \lambda_m)} \tag{11}$$

where $F(\lambda_x, \lambda_m)$ is the measured fluorescence intensity, $P_0(\lambda_x)$ is the power of the incident light at the excitation wavelength, D is the detector's wavelength-dependent response function, $[(\Delta\Omega/\pi)\cos\theta]$ is the detector collection efficiency for a distant detector geometry and a tissue surface with Lambertian intensity distribution, and X_{1D} is the one-dimensional path length factor. The term $P_0(\lambda_x)[(\Delta\Omega/\pi)\cos\theta]D(\lambda_m)$ can be quantified by calibration of the fluorescence measurement for source and detector constants. Specifically, the fluorescence calibration can be performed with an optically dilute fluorophore solution with a predetermined intrinsic fluorescence coefficient. The path length factor X_{1D} requires knowledge of the tissue optical properties at the excitation and emission wavelengths. This function is based upon the exponential attenuation of light away from the source, as are the Beer–Lambert and diffusion theories, with coefficients that are empirically developed from extensive Monte-Carlo simulations. Factor X_{1D} can be calculated from noninvasive measurements of the diffuse reflectance (R_d) and effective penetration depth (δ). Because two optical properties are needed, it is possible to make two, unique, diffuse reflectance measurements with two illumination–collection fiber distances in order to specify R_d and δ.

The recovery method presented successfully identifies the intrinsic fluorescence coefficient (both spectral line shape and intensity) of turbid tissue phantoms that incorporate various concentrations of fluorophore (rhodamine 6G), absorber (adult hemoglobin) and scatterer (polystyrene spheres) with absorption and scattering coefficients that coincide with the range found in soft tissues. The results are shown in Figure 19. Furthermore, fluorophore concentrations of the turbid tissue phantoms are predicted to within 15% of the true concentration (not shown).

Although this analytical model is simple and can recover the intrinsic fluorescence coefficient from turbid tissue phantom fluorescence emission spectra, there are several limitations associated with it. The optical properties of the medium need to be uniform. Furthermore, the size of the medium should be large enough such that only the surface boundary affects the distribution of light. Finally, the one-dimensional fluence rate expression is valid for collimated incident light with a beam diameter that is at least a factor of four larger than the penetration depth of the light.

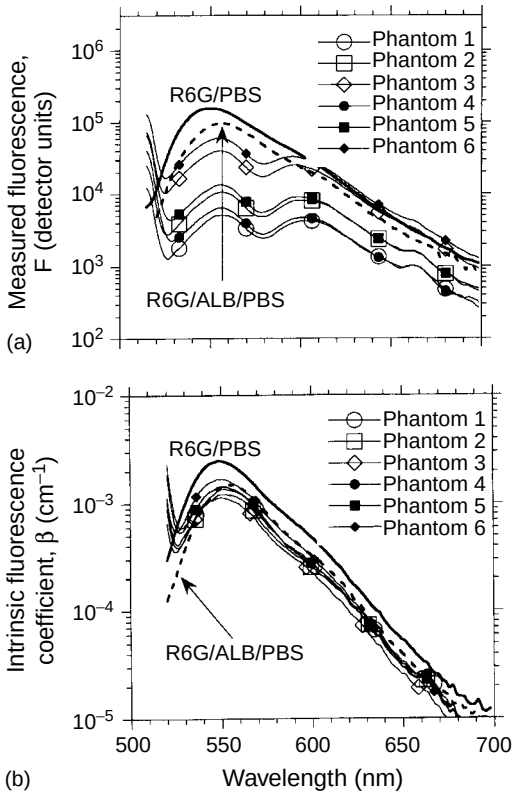

Figure 19 Summary of intrinsic fluorescence recovery from six turbid tissue phantoms: (a) measured uncorrected fluorescence spectra; (b) corrected fluorescence spectra, displayed as the intrinsic fluorescence coefficient β. Note that R6G/PBS corresponds to the fluorescence of rhodamine 6G in phosphate-buffered saline and that R6G/ALB/PBS corresponds to the fluorescence of rhodamine 6G in 2.5% bovine serum albumin–phosphate-buffered saline. (Reproduced by permission of the Optical Society of America, from Gardner et al.[127])

Wu et al.[126] have achieved a similar goal as Gardner et al.[127] using a photon migration approach to model fluorescence from a homogeneous turbid medium such as tissue. The model provides an analytical relationship between the bulk fluorescence emission spectrum F and the diffuse reflectance spectrum R for arbitrary geometries and boundary conditions, which represents an advantage over the previous analytical model.[126] Wu et al. demonstrate that the distortion can be removed by measuring R from a turbid medium over the same wavelength range and with the same geometry as is used for measuring F. The validity of this approach has been demonstrated from tissue experiments using human aortic media and with Monte-Carlo simulations. This analytical model does provide accurate fluorescence line shape information, but unlike the previous analytical model[126] the absolute intensity is still coupled to the absorption coefficient. This implies that the fluorescence measurements from two tissues that have the same amount of fluorophore but different amounts of absorption will have the same line shape but different fluorescence intensities.

Unlike Gardner et al.[127] and Wu et al.,[126] Panou-Diamandi et al.[145] modeled tissue fluorescence using electromagnetic theory. In this mathematical approach, the inelastic property of fluorescence was expressed via the polarization vector of the medium. The fluorescence scattering spectral function was independent of the excitation and emission geometry, expressing the energy transfer from the excitation frequency to all the emission frequencies. The model assumed that the tissue is a single, homogeneous, infinitely thick dielectric layer under plane wave illumination. Experimental measurements were carried out on optically turbid collagen gels that contained fluorescent dyes, in order to validate the mathematical model. Comparison between experimental and theoretically expected fluorescence emission spectra gave satisfactory results.

Another technique that has been employed to deconvolve the absorption and scattering properties from the fluorescence emission spectra of turbid media is partial least squares (PLS) regression.[146] This method is not based on analytical approximations of the transport equations[126,127] or Maxwell's equations,[145] but rather describes the fluorescence emission spectra measured from turbid medium as a linear combination of basis vectors that are representative of those to be predicted. The PLS method involves the regression between a fluorescence spectral matrix $X(n \times m)$ and a concentration matrix $Y(n \times p)$. PLS seeks a calibration matrix $B(m \times p)$ such that Equation (12) holds:

$$Y = XB \tag{12}$$

The PLS method was compared to models of light transport based on Beer's law[147] and Kubelka–Munk theory[56] to determine which is most effective in extracting the concentration of fluorophores in a set of turbid tissue phantoms containing absorbers and scatterers. The model based on Beer's law significantly underestimated the fluorophore concentrations, whereas the model based on Kubelka–Munk theory significantly overestimated the concentrations. However, this method overestimated the concentrations by a constant scaling factor that potentially can be corrected for. Finally, using the PLS method, which was trained on a set of turbid tissue phantom spectra that have optical properties similar to those to be analyzed, fluorophore concentrations were predicted to within 5% of the true concentration. It should be noted that the success of the PLS method relies on a training set that reflects the chemical and optical

complexity of the turbid media to be investigated. Several important obstacles remain before PLS can be applied to analyze fluorescence emission spectra from tissues. The main obstacle is the question of how to construct reliably a training set in which the concentrations of fluorophores and absorbers of biological interest can be quantified.

7.2.2 Identity and Distribution of Fluorophores that Contribute to Tissue Fluorescence Emission Spectra

It was mentioned earlier that quantification of the concentration of fluorophores within tissue in principle involves two steps. The first step requires deconvolution of the absorption and scattering properties of the tissue from the bulk tissue fluorescence emission spectrum. The second step requires knowledge of the identity and distribution of the fluorophores within the tissue. Although the previously described analytical models[126,127,145,146] address the first step, they do not address the second. The following example addresses the second step specifically through the measurement and quantification of the biochemical and morphological basis of colonic tissue fluorescence emission spectra measured in vivo.[52]

Zonios et al.[52] measured and quantified the contribution of the fluorescent microstructures in frozen, unstained tissue sections to the tissue fluorescence emission spectra measured from colonic tissues in vivo. In order to achieve this, first fluorescence emission spectra were measured at 370 nm excitation from normal colon tissues and colonic adenomas (precancer) in vivo. Deconvolution of absorption and scattering from fluorescence emission spectra measured from the tissues in vivo was achieved using optical properties obtained from Kubelka–Munk[61] analysis of diffuse transmittance and reflectance measurements of excised human colon tissues in vitro. Fluorescence, microscopy and microspectrofluorimetry at 363 nm excitation of thin, frozen, unstained tissue sections were used to characterize the spectral line shapes and the distribution of the fluorescent microstructures that contribute to the measured fluorescence emission spectra of colonic tissues in vivo. A model based on Monte-Carlo simulations[65] was used to relate the spectral line shape and distribution of the fluorescent microstructures to the colonic tissue spectra measured in vivo. This is perhaps one of the first comprehensive attempts to measure and quantify the biochemical and morphological basis of tissue fluorescence emission spectra measured in vivo.

The model developed by Zonios et al.[52] describes the tissue fluorescence emission spectrum $F(\lambda_x, \lambda_m)$, as shown by Equation (13):

$$F(\lambda_x, \lambda_m) = \kappa \sum_i \phi_i(\lambda_m) D_i(z) T(\lambda_x, \lambda_m, z)\, dz \qquad (13)$$

where λ_x represents the excitation wavelength, λ_m represents the emission wavelength, z represents the depth within the tissue sample, ϕ_i represents the intrinsic fluorescence of each individual microstructure with index i, $D_i(z)$ represents the fluorescence intensity spatial distribution of an individual microstructure with index i, $T(\lambda_x, \lambda_m, z)$ is a transfer function that incorporates the tissue optical properties and is calculated using Monte-Carlo simulations of light propagation in tissue, and κ represents a scaling factor.

First, a database of fluorescence line shapes and distributions of fluorescent microstructures within colon tissue were obtained. A microspectrofluorimeter incorporating a fluorescence microscope, an argon laser source (363.8 nm), a spectrograph and an intensified PDA, coupled to an optical multichannel analyzer was used to record the fluorescence emission spectra of microstructures from unstained, frozen tissue sections cut from tissue biopsies. Quantitative fluorescence imaging studies to quantify the distribution of the fluorescent microstructures within the tissue section were performed using a 75-W short-arc filtered xenon lamp (380 nm) coupled to a microscope and a thermoelectrically cooled CCD camera.

Finally, the transmission and reflection spectra of thick samples of normal mucosa and adenomatous polyps were measured in the range 300–700 nm using a spectrophotometer equipped with an integrating sphere. The tissue absorption and scattering properties were determined by calculating the Kubelka–Munk coefficients[61] and transforming these into transport theory absorption and reduced scattering coefficients.[148] The anisotropy parameter was obtained from the literature for tissues with histology similar to the colon.[53] The transfer function was determined using Monte-Carlo simulations.[65] The transfer function accounts for the scattering and absorption of the excitation light and emitted light at wavelengths λ_x and λ_m, respectively. It also accounts for the specific light illumination/collection geometry used to make the fluorescence measurements from tissues in vivo.

Figure 20 shows the computed fluorescence emission spectra of normal and adenomatous colon tissue versus the corresponding average fluorescence emission spectra measured from tissues in vivo. The peak intensity of the in vivo spectra has been scaled for comparison. The computed spectra contain all the characteristic spectral features observed in the in vivo spectra. The following interpretations may be made from the intrinsic fluorescent microstructures, the fluorescence density function and the transfer function used to calculate the computed fluorescence emission spectra shown in Figure 20:

- Although the intrinsic fluorescence of collagen (the dominant fluorophore) peaks at 420 nm, the observed peak in both normal tissues and adenoma is at 460 nm

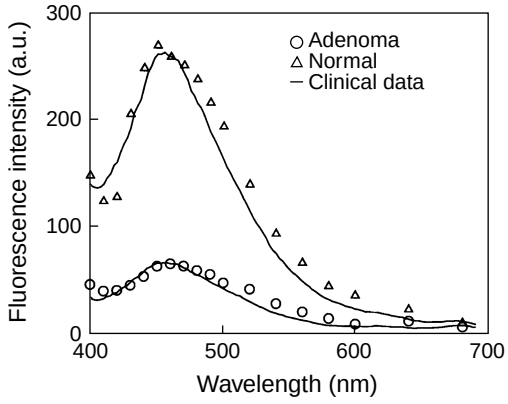

Figure 20 Computed fluorescence emission spectra from normal colon tissues and adenomas versus the average fluorescence spectra measured in vivo. (Reproduced by permission of IEEE, Piscataway, NJ (© 1996 IEEE), from Zonios et al.[82])

due to the effect of large hemoglobin absorption at 420 nm.

- The fluorescence intensity of adenoma is smaller than that of normal colon. This is due to three factors: the mucosal collagen fluorescence is decreased in adenoma due to the enlargement of crypts, which displace the lamina propria; the submucosa contributes to the fluorescence in normal tissue but not in adenoma, because of the increased thickness of a polyp; and adenoma exhibits increased absorption due to hemoglobin content.
- Red fluorescence is increased in adenoma. This additional red fluorescence is primarily associated with the intrinsic fluorescence of the dysplastic crypt cells.

In conclusion, the model developed by Zonios et al.[52] quantifies the contribution of the fluorescent microstructures that are responsible for differences observed in the fluorescence emission spectra of normal and adenomatous colon tissues. Their findings indicate that both the biochemical and morphological features of the tissue contribute to the measured fluorescence emission spectrum.

Zeng et al.[149] have used a similar approach to reconstruct in vivo skin fluorescence emission spectra. Specifically, they measured the fluorescence emission and diffuse reflectance spectra of skin in vivo and the fluorescence of the microstructures in the skin in vitro. They reconstructed the spectra measured from skin in vivo using Monte-Carlo simulations in which they used the microscopic fluorophore distribution and spectra measured from excised skin tissue sections in vitro. The optical properties of skin, however, were obtained from published values in the literature. They did not report similar data for any skin lesions such as naevi or melanoma.

7.3 Effect of Excitation and Emission Geometry on Fluorescence Emission Spectra of Turbid Media

Several groups have evaluated the effect of the excitation and emission geometry on the measurement of tissue fluorescence emission spectra from turbid media and tissue.[58,150–153] Keijzer et al.[58] used Monte-Carlo simulations and experimental measurements to show that the illumination/collection geometry affects both the intensity and line shape of fluorescence emission measured from tissue. Specifically, they demonstrated that as the separation between the illumination and collection fiber is increased, the measured fluorescence intensity decreases as a result of hemoglobin absorption. Furthermore, the attenuation due to hemoglobin absorption is wavelength dependent, thus affecting the fluorescence line shape as well. Richards-Kortum et al.[57] suggested using a fiber-optic probe, which collects the emitted light from an area, that is directly illuminated by the excitation light in order to minimize the effect of hemoglobin absorption on the fluorescence measurement. Qu et al.[150] also used Monte-Carlo simulations and experimental measurements to demonstrate that the distortion to the fluorescence emission spectrum is a function of the diameter of the area viewed by the illumination and collection fiber as well. Using bronchial tissue optical properties (for the excitation and emitted light) in their Monte-Carlo simulations, they found that the distortions to the fluorescence emission spectrum were minimized when the diameter of the tissue area viewed by the excitation and emission fiber was less than 1 mm. This was verified experimentally.

Pogue et al.[151] went a step further and demonstrated that if fluorescence is measured from a sample volume that is smaller than the average mean free scattering path (inverse of the reduced scattering coefficient) within that turbid medium, then the effects of absorption will be diminished. To achieve this, they used a confocal detection scheme to measure fluorescence as a function of fluorophore concentration from a turbid medium. They used a pinhole detector with a diameter (10 μm) that was an order of magnitude smaller than the mean free scattering path in the medium. Using this method, they were able to obtain a linear relationship between the fluorescence intensity and concentration of the fluorophore in the turbid medium for a wide range of fluorophore concentrations. Although this relationship was independent of the absorption coefficient, the fluorescence intensity varied linearly with the scattering coefficient for a fixed fluorophore concentration. Fortuitously, this is not a serious limitation because most tissues do not have a large variation in their scattering coefficient. It was also observed

for a fixed detector size that, as the mean free scattering path decreases, the number of nonscattered fluorescent photons relative to the number of scattered fluorescent photons that are detected also decreases. This suggests that the size of the detector has to be reduced as the mean free scattering path length of the turbid medium is reduced. This has direct implications on the strength of the fluorescence intensities detected and the size of the sampled volume. In a subsequent report, Pogue et al.[152] showed that they could overcome the problem of poor signal-to-noise ratio by developing and employing a fiber-optic probe that has multiple fibers (100 μm in diameter), each of which serves as an independent source and detector. Using this probe, the fluorescence intensity measured by each fiber was obtained from a sample volume smaller than the mean free scattering path for that medium, thus minimizing the effect of absorption. By integrating the fluorescence intensity from a total of 30 fibers, the signal-to-noise ratio of the fluorescence measurements was enhanced significantly. The main limitation of this technique is the very small sampled volume.

Avrillier et al.[153] used an alternative approach to resolve the effect of the illumination/collection geometry on the fluorescence emission spectra of tissue. They used Monte-Carlo simulations and optical properties of tissues measured in vitro to predict the effect of absorption and scattering on the tissue fluorescence emission spectra measured in vivo. Specifically, they predicted the correction factors as a function of increased separation between the excitation and emission fibers. These predictions were used to correct the distortion in the fluorescence emission spectra measured from the tissues. However, because the optical properties of tissues measured in vitro are fraught with error associated with extraction, this approach is considered to be, at best, approximate unless the tissue optical properties used in the Monte-Carlo simulations are measured in vivo.

8 FUTURE PERSPECTIVES

Diagnostic tools based on optical spectroscopy in the UV/VIS spectral regions have the potential to link the biochemical and morphological properties of tissues to individual patient care. In particular, these techniques are fast, noninvasive and quantitative. Furthermore, the accuracy and efficacy of the technology have been demonstrated clinically to be comparable or superior to current clinical modalities.

However, the biochemical and morphological basis for the diagnostic capability of fluorescence spectroscopy is not completely understood. This is limited by the lack of: sophisticated mathematical models that can be used to quantify tissue optical properties in order to account for the effects of absorption and scattering in the UV/VIS spectral region; and a comprehensive understanding of the biochemical constituents that contribute to the measured fluorescence. Hence, today's challenge for biomedical optics is to elucidate tissue biochemistry and morphology in greater detail for the specific disease process so that appropriate therapeutic interventions may have the greatest impact. In order to achieve this, both experimental and mathematical techniques need to developed that can enable a greater understanding of the chromophores and fluorophores that are modulated by the disease process and their contribution to the measured spectrum.

Animal models are useful with regard to developing methodologies that can be used to elucidate the biochemical and morphological properties of tissue and optimizing the instruments to exploit these features. The advantages that they offer are: they are well characterized; spectral measurements can be made per tissue site for a variety of experimental and biological conditions; and data can be obtained from a statistically significant number of animals for the purpose of validation without the need for expensive clinical trials. Although animal model studies have been performed extensively in relation to the use of photosensitizing agents for PDT, only a handful of studies on measuring and quantifying tissue autofluorescence have been performed.[154–159]

With respect to clinical applications, the diagnostic potential of fluorescence spectroscopy is defined by histology. In order to evaluate if fluorescence spectroscopy is sensitive to changes that precede the morphological manifestations seen by histological evaluation, it is valuable to correlate the measured spectral information to other biochemical and/or genetic markers of increased cancer risk. Furthermore, although fluorescence spectroscopy has been most widely explored for the detection of precancer or cancer in a screening and diagnostic setting, its diagnostic potential for guiding surgery has been evaluated only in a limited number of clinical studies.[101] Because this technique can discriminate between diseased and nondiseased tissues in a fast and noninvasive manner, it could be used to define margins of resection for various surgical procedures, such as that of the breast and brain. Additionally, this technique also could be used to monitor the response of tissues to various therapeutic interventions.

ACKNOWLEDGMENTS

The author gratefully acknowledges Ms Mary Leonard for her extensive help with the preparation of this manuscript.

ABBREVIATIONS AND ACRONYMS

APD	Avalanche Photodiode
ATP	Adenosine Triphosphate
BPD	Benzoporphyrin Derivative
CCD	Charge-coupled Device
EEM	Excitation–Emission Matrix
FAD	Flavin Adenine Dinucleotide
$FADH_2$	Reduced Flavin Adenine Dinucleotide
Fp	Flavoprotein
fwhm	Full Width at Half-maximum
HpD	Hematoporphyrin Derivative
IR	Infrared
LIFE	Light-induced Fluorescence Endoscopy
MTHPC	Meso-tetra-(hydroxyphenyl)-chlorin
NAD^+	Nicotinamide Adenine Dinucleotide
NADH	Reduced Nicotinamide Adenine Dinucleotide
NIR	Near-infrared
NIST	National Institute of Standards and Technology
PCA	Principal Component Analysis
PDA	Photodiode Array
PDT	Photodynamic Therapy
PLS	Partial Least Squares
PMT	Photomultiplier Tube
PN	Pyridine Nucleotide
PpIX	Protoporphyrin IX
SIL	Squamous Intra-epithelial Lesions
SnET2	Tin Etiopurpurin
UV	Ultraviolet
VIS	Visible
5-ALA	δ-Aminolevulinic Acid

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by • Photodynamic Therapy

REFERENCES

1. R.S. Cotran, V. Kumar, S.L. Robbins, *Robbins Pathologic Basis of Disease*, WB Saunders Co., Philadelphia, PA, 1989.
2. I.D. Campbell, R.A. Dwek, *Biological Spectroscopy*, Benjamin/Cummings Publishing Co., Menlo Park, CA, 1984.
3. F. Koenig, F.J. McGovern, H. Enquist, R. Larne, T.F. Deutsch, K.T. Schomacker, 'Autofluorescence Guided Biopsy for the Early Diagnosis of Bladder Carcinoma', *J. Urol.*, **159**(6), 1871–1875 (1998).
4. B. Chance, M. Cope, B. Dugan, N. Ramanujam, B. Tromberg, 'Phase Measurements of Absorbers and Scatterers in Tissue: Review', *Rev. Sci. Instrum.*, **69**(10), 3457–3481 (1998).
5. S. Andersson-Engels, J. Johansson, U. Stenram, K. Svanberg, S. Svanberg, 'Malignant Tumor and Atherosclerotic Plaque Diagnosis Using Laser Induced Fluorescence', *IEEE J. Quant. Electron.*, **26**(12), 2207–2217 (1990).
6. S. Andersson-Engels, J. Johansson, K. Svanberg, S. Svanberg, 'Fluorescence Imaging and Point Measurements of Tissue: Applications to the Demarcation of Malignant Tumors and Atherosclerotic Lesions from Normal Tissue', *Photochem. Photobiol.*, **53**(6), 807–814 (1991).
7. S. Andersson-Engels, B.C. Wilson, 'In Vivo Fluorescence in Clinical Oncology: Fundamentals and Practical Issues', *J. Cell. Pharmacol.*, **3**, 66–79 (1992).
8. T.G. Papazoglou, 'Malignancies and Atherosclerotic Plaque Diagnosis – Is Laser Induced Fluorescence Spectroscopy the Ultimate Solution?', *J. Photochem. Photobiol. B: Biol.*, **28**, 3–11 (1995).
9. I.J. Bigio, J.R. Mourant, 'Ultraviolet and Visible Spectroscopies for Tissue Diagnostics: Fluorescence Spectroscopy and Elastic-scattering Spectroscopy', *Phys. Med. Biol.*, **42**, 803–814 (1996).
10. S. Andersson-Engels, C.A. Klinteberg, K. Svanberg, S. Svanberg, 'In Vivo Fluorescence Imaging for Tissue Diagnostics', *Phys. Med. Biol.*, **42**, 815–824 (1997).
11. R. Richards-Kortum, E. Sevick-Muraca, 'Quantitative Optical Spectroscopy for Tissue Diagnosis', *Annu. Rev. Phys. Chem.*, **47**, 555–606 (1996).
12. G.A. Wagnieres, W.M. Star, B.C. Wilson, 'In Vivo Fluorescence Spectroscopy and Imaging for Oncological Applications', *Photochem. Photobiol.*, **68**(5), 603–632 (1998).
13. E.B. Hanlon, I. Itzkan, R.R. Dasari, M.S. Feld, R.J. Ferrante, A.C. McKee, D. Lathi, N.W. Kowall, 'Near-infrared Fluorescence Spectroscopy Detects Alzheimer's Disease In Vitro', *Photochem. Photobiol.*, **70**(2), 236–242 (1999).
14. L.T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J.M. Crawford, M.S. Feld, 'Observation of Periodic Fine Structure in Reflectance from Biological Tissue: a New Technique for Measuring Nuclear Size Distribution', *Phys. Rev. Lett.*, **80**(3), 627–630 (1998).
15. B.B. Das, F. Liu, R.R. Alfano, 'Time-resolved Fluorescence and Photon Migration Studies in Biomedical

and Random Media', *Rep. Prog. Phys.*, **60**(2), 227–292 (1997).
16. S.P. Gopinath, C.S. Robertson, R.G. Grossman, B. Chance, 'Near-infrared Spectroscopic Localization of Intracranial Hematomas', *J. Neurosurg.*, **79**, 43–47 (1993).
17. W.J. Levy, S. Levin, B. Chance, 'Near-infrared Measurement of Cerebral Oxygenation: Correlation with Electroencephalographic Ischemia During Ventricular Fibrillation', *Anesthesiology*, **83**(4), 738–746 (1995).
18. S. Nioka, Q. Luo, B. Chance, 'Human Brain Functional Imaging with Reflectance CWS', *Adv. Exp. Med. Biol.*, **428**, 237–242 (1997).
19. A. Villringer, B. Chance, 'Non-invasive Optical Spectroscopy and Imaging of Human Brain Function', *Trends Neurosci.*, **20**(10), 435–442 (1997).
20. W. Bank, B. Chance, 'Diagnosis of Defects in Oxidative Muscle Metabolism by Non-invasive Tissue Oximetry', *Mol. Cell. Biochem.*, **174**(1/2), 7–10 (1997).
21. S. Nioka, M. Miwa, S. Orel, M. Shnall, M. Haida, S. Zhao, B. Chance, 'Optical Imaging of Human Breast Cancer', *Adv. Exp. Med. Biol.*, **361**, 171–179 (1994).
22. D.A. Boas, A.G. Yodh, 'Spatially Varying Dynamical Properties of Turbid Media Probed with Diffusing Temporal Light Correlation', *J. Opt. Soc. Am. A*, **14**, 192–215 (1997).
23. A. Mahadevan-Jansen, R. Richards-Kortum, 'Raman Spectroscopy of Pre-cancers and Cancers', *J. Biomed. Opt.*, **1**(1), 31–70 (1996).
24. J.F. Brennan III, M.E. Beattie, Y. Wang, M.J. Cantella, B.Y. Tsaur, R.R. Dasari, M.S. Feld, 'PdSi Focal-plane Array Detectors for Short-wave Infrared Raman Spectroscopy of Biological Tissue: a Feasibility Study', *Appl. Opt.*, **35**(28), 5736–5739 (1996).
25. A.J. Berger, T.W. Koo, I. Itzkan, G. Horowitz, M.S. Feld, 'Multicomponent Blood Analysis by Near-infrared Raman Spectroscopy', *Appl. Opt.*, **38**(13), 2916–2926 (1999).
26. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum Press, New York, 1983.
27. M.B. Silberberg, H.E. Savage, G.C. Tang, P.G. Sacks, R.R. Alfano, S.P. Schantz, 'Detection of Retinoic Acid Induced Biochemical Alterations in Squamous Cell Carcinoma Using Intrinsic Fluorescence Spectroscopy', *Laryngoscopy*, **104**(3), 278–282 (1994).
28. A. Pradhan, P. Pal, G. Durocher, L. Villeneuve, A. Balassy, F. Babai, L. Gaboury, L. Blanchard, 'Steady State and Time Resolved Fluorescence Properties of Metastatic and Non-metastatic Malignant Cells from Different Species', *J. Photochem. Photobiol. B: Biol.*, **3**(3), 101–112 (1995).
29. S. Ganesan, P.G. Sacks, Y. Yang, A. Katz, M. Al-Rawi, H.E. Savage, S.P. Schantz, R.R. Alfano, 'Native Fluorescence Spectroscopy of Normal and Malignant Epithelial Cells', *Cancer Biochem. Biophys.*, **16**, 365–373 (1998).
30. T.J. Romer, M. Fitzmaurice, R.M. Cothren, R. Richards-Kortum, M.V. Sivak, Jr, J.R. Kramer, Jr, 'Laser-induced Fluorescence Microscopy of Normal Colon and Dysplasia in Colonic Adenomas: Implications for Spectroscopic Diagnosis', *Am. J. Gastroenterol.*, **90**(1), 81–87 (1995).
31. H. Wang, J. Willis, M.J.F. Canto, M.V. Sivak, Jr, J.A. Izatt, 'Quantitative Laser Scanning Confocal Autofluorescence Microscopy of Normal, Premalignant, and Malignant Colonic Tissues', *IEEE Trans. Biomed. Eng.*, **46**(10), 1246–1252 (1999).
32. G.S. Fairman, M.H. Nathanson, A.B. West, L.I. Deckelbaum, L. Kelly, C.R. Kapadia, 'Differences in Laser-induced Autofluorescence Between Adenomatous and Hyperplastic Polyps and Normal Colonic Mucosa by Confocal Microscopy', *Digest. Dis. Sci.*, **40**(6), 1261–1268 (1995).
33. K. Izuishi, H. Tajiri, T. Fujii, N. Boku, A. Ohtsu, T. Ohnishi, M. Ryi, T. Kinoshita, S. Yoshida, 'The Histological Basis of Detection of Adenoma and Cancer in the Colon by Autofluorescence Endoscopic Imaging', *Endoscopy*, **31**(7), 511–516 (1999).
34. G. Bottiroli, A.C. Croce, D. Locatelli, R. Marchesini, E. Pignoli, S. Tomatis, C. Cuzzoni, S. Di Palma, M. Dalfante, P. Spinellu, 'Natural Fluorescence of Normal and Neoplastic Human Colon: a Comprehensive ''ex vivo'' Study', *Lasers Surg. Med.*, **16**(1), 48–60 (1995).
35. G.I. Zonios, R.M. Cothren, J.T. Arendt, J. Wu, J. Van Dam, J.M. Crawford, R. Manoharan, M.S. Feld, 'Morphological Model of Human Colon Tissue Fluorescence', *IEEE Trans. Biomed. Eng.*, **43**(2), 113–122 (1996).
36. R.S. DaCosta, L. Ligle, J. Kost et al., 'Confocal Microscopy/Macroscopy and Microspectrofluorimetry Analysis of Human Colorectal Tissue', *J. Anal. Morphol.*, **4**, 192–194 (1997).
37. W.S. Glassman, M. Steinberg, R.R. Alfano, 'Time Resolved and Steady State Fluorescence Spectroscopy from Normal and Malignant Cultured Human Breast Cell Lines', *Lasers Life Sci.*, **6**(2), 91–98 (1994).
38. M. Anidjar, O. Cussenot, J. Blais, O. Bourdon, S. Avrillier, D. Ettori, J.M. Villter, J. Fiet, P. Teillac, A. Le Duc, 'Argon Laser Induced Autofluorescence may Distinguish Between Normal and Tumor Human Urothelial Cells: a Microspectrofluorimetric Study', *J. Urol.*, **155**(5), 1771–1774 (1996).
39. F.N. Ghadially, W.J.P. Neish, H.C. Dawkins, 'Mechanisms Involved in the Production of Red Fluorescence of Human and Experimental Tumors', *J. Pathol. Bacteriol.*, **85**, 77–92 (1963).
40. L. Stryer, *Biochemistry*, WH Freeman and Company, New York, 1988.
41. D. Fujimoto, 'The Structure of Pyridinoline, a Collagen Crosslink', *Biochem. Biophys. Res. Commun.*, **76**(4), 1124–1129 (1977).
42. Z. Deyl, K. Macek, M. Adam, O. VanCikova, 'Studies on the Chemical Nature of Elastin Fluorescence', *Biochim. Biophys. Acta*, **625**, 248–254 (1980).

43. D.R. Eyre, M.A. Paz, 'Cross-linking in Collagen and Elastin', *Annu. Rev. Biochem.*, **53**, 717–748 (1984).
44. D.P. Thornhill, 'Separation of a Series of Chromophores and Fluorophores Present in Elastin', *Biochem. J.*, **147**, 215–219 (1975).
45. B. Chance, B. Schoener, R. Oshino, F. Itshak, Y. Nakase, 'Oxidation–Reduction Ratio Studies of Mitochondria in Freeze-trapped Samples', *J. Biol. Chem.*, **254**(11), 4764–4771 (1979).
46. A.M. Kluftinger, N.L. Davis, N.F. Quenville, S. Lam, J. Hung, B. Palcic, 'Detection of Squamous Cell Cancer and Pre-cancerous Lesions by Imaging Tissue Autofluorescence in the Hamster Cheek Pouch Model', *Surg. Oncol.*, **1**, 183–188 (1992).
47. A. Policard, 'Etudes sur les Aspects Offerts par des Tumeurs Experimentales Examinees a la Lumiere de Woods', *CR Soc. Biol.*, **91**, 1423–1425 (1924).
48. F. Ronchese, 'The Fluorescence of Cancer Under the Wood's Light', *Oral Surg. Oral Med. Pathol.*, **7**, 967–971 (1954).
49. F.N. Ghadially, 'Red Fluorescence of Experimentally Induced and Human Tumors', *J. Pathol. Bacteriol.*, **80**, 345–361 (1960).
50. F.N. Ghadially, W.J.P. Neish, 'Porphyrin Fluorescence of Experimentally Induced Squamous Cell Carcinoma', *Nature*, **188**, 1124 (1960).
51. F.N. Ghadially, W.J.P. Neish, H.C. Dawkins, 'Mechanisms Involved in the Production of Red Fluorescence of Human and Experimental Tumors', *J. Pathol. Bacteriol.*, **85**, 77–92 (1963).
52. J.C. Kennedy, R.H. Pottier, D.C. Pross, 'Photodynamic Therapy with Endogenous Protoporphyrin IX: Basic Principles and Present Clinical Experience', *J. Photochem. Photobiol. B*, **6**, 143–148 (1990).
53. W.F. Cheong, A.J. Welch, 'A Review of the Optical Properties of Tissues', *IEEE J. Quantum Electron.*, **26**, 2166–2185 (1990).
54. R. Richards-Kortum, 'Fluorescence Spectroscopy of Turbid Media', in *Optical–Thermal Response of Laser Irradiated Tissue*, eds. A.J. Welch, M.J.C. van Gemert, Plenum Press, New York, 667–706, 1995.
55. S.B. Brown, *An Introduction to Spectroscopy for Biochemists*, Academic Press, London, 1980.
56. A.J. Durkin, S. Jaikumar, N. Ramanujam, R. Richards-Kortum, 'Relation Between Fluorescence Spectra of Dilute and Turbid Samples', *Appl. Opt.*, **33**(3), 414 (1994).
57. R. Richards-Kortum, A. Mehta, G. Hayes, R. Cothren, T. Kolubayev, C. Kitrell, N.B. Ratliff, J.R. Kramer, M.S. Feld, 'Spectral Diagnosis of Atherosclerosis Using an Optical Fiber Laser Catheter', *Am. Heart J.*, **118**(2), 381 (1989).
58. M. Keijzer, R. Richards-Kortum, S.L. Jacques, M.S. Feld, 'Fluorescence Spectroscopy of Turbid Media: Autofluorescence of the Human Aorta', *Appl. Opt.*, **28**, 4286–4292 (1989).
59. S. Chandrasekhar, *Radiative Transfer*, Oxford University Press, Oxford, 1950.
60. A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Academic Press, New York, 1987.
61. P. Kubelka, 'New Contributions to the Optics of Intensely Light Scattering Materials. Part II: Nonhomogeneous Layers', *J. Opt. Soc. Am.*, **344**, 330–335 (1954).
62. S. Prahl, M.J.C. van Gemert, A.J. Welch, 'Determining the Optical Properties of Turbid Media by Using the Adding–Doubling Method', *Appl. Opt.*, **32**, 559–568 (1993).
63. M.S. Patterson, B. Chance, B.C. Wilson, 'Time Resolved Reflectance and Transmittance for the Non-invasive Measurement of Tissue Optical Properties', *Appl. Opt.*, **28**, 2331–2336 (1989).
64. J.J. Duderstadt, L.J. Hamilton, *Nuclear Reactor Analysis*, Wiley, New York, 103–144, 1976.
65. S.L. Jacques, 'Monte-Carlo Methods', in *Optical–Thermal Response of Laser-irradiated Tissue*, eds. A.J. Welch, M.J.C. van Gemert, Plenum Press, New York, 73–99, 1995.
66. A.J. Durkin, S. Jaikumar, R. Richards-Kortum, 'Optically Dilute, Absorbing and Turbid Phantoms for Fluorescence Spectroscopy of Homogeneous and Inhomogeneous Samples', *Appl. Spectrosc.*, **47**, 2114–2121 (1993).
67. G. Wagnieres, S. Cheng, M. Zellweger, N. Utke, D. Braichotte, J.-P. Ballini, H. van den Bergh, 'An Optical Phantom with Tissue-like Properties in the Visible for Use in PDT and Fluorescence Spectroscopy', *Phys. Med. Biol.*, **42**(7), 1415–1426 (1997).
68. E. Kohen, R. Santus, J.G. Hirschberg, *Photobiology*, Academic Press, San Diego, CA, 1995.
69. R.A. Zangaro, L. Silveira, R. Manoharan, G. Zonios, I. Itzkan, R.R. Dasari, J. Van Dam, M.S. Feld, 'Rapid Multi-excitation Fluorescence Spectroscopy System for In Vivo Tissue Diagnosis', *Appl. Opt.*, **35**(25), 5211–5219 (1996).
70. R.L. McCreery, 'CCD Array Detectors for Multichannel Raman Spectroscopy', in *Charged Transfer Devices in Spectroscopy*, eds. J.V. Sweedler, K.L. Ratzlaff, M.B. Denton, VCH Publishers, New York, 227–279, 1994.
71. E.V. Trujillo, D.R. Sandison, U. Utzinger, N. Ramanujam, M. Follen Mitchell, R. Richards-Kortum, 'Method to Determine Tissue Fluorescence Efficiency In Vivo and Predict Signal-to-noise Ratio for Spectrometers', *Appl. Spectrosc.*, **52**(7), 943–951 (1998).
72. D.G. Taylor, J.N. Demas, *Anal. Chem.*, **51**, 712 (1979).
73. G.G. Stokes, 'Uber Die Anderung der Brechbarkeit des Lichtes', *Philos. Trans. R. Soc.*, **107**, 11 (1852).
74. H. Stubel, 'Die Fluoreszenz Tierischer Gewebe in Ultravioletten Licht', *Pflugers Arch. Physiol.*, **142**, 1 (1911).
75. R.M. Lycette, R.B. Leslie, *Lancet*, **2**, 436 (1965).

76. A.E. Profio Profio, D.R. Doiron, E.G. King, 'Laser Fluorescence Bronchoscope for Localization of Occult Lung Tumors', *Med. Phys.*, **6**(6), 523–525 (1979).
77. A.E. Profio, D.R. Doiron, 'A Feasibility Study of the Use of Fluorescence Bronchoscopy for the Localization of Small Lung Tumors', *Phys. Med. Biol.*, **22**, 949–957 (1977).
78. R.R. Alfano, D.B. Tata, J. Cordero, P. Tomashefsky, F.W. Longo, M.A. Alfano, 'Laser Induced Fluorescence Spectroscopy from Native Cancerous and Normal Tissue', *IEEE J. Quant. Electron.*, **20**, 1507–1511 (1984).
79. W. Lohmann, 'In Situ Detection of Melanomas by Fluorescence Measurements', *Maturwissenschaften*, **75**, 201–202 (1988).
80. Y. Yang, Y. Ye, F. Li, Y. Li, P. Ma, 'Characteristic Autofluorescence for Cancer Diagnosis and its Origin', *Lasers Surg. Med.*, **7**, 528–532 (1987).
81. A. Mahadevan, M.F. Mitchell, E. Silva, S. Thomsen, R. Richards-Kortum, 'A Study of the Fluorescence Properties of Normal and Neoplastic Human Cervical Tissue', *Lasers Surg. Med.*, **13**, 647–655 (1994).
82. G.I. Zonios, R.M. Cothren, J.T. Arendt, J. Wu, J. Van Dam, J.M. Crawford, R. Manoharan, M.S. Feld, 'Morphological Model of Human Colon Tissue Fluorescence', *IEEE Trans. Biomed. Eng.*, **43**, 113–122 (1996).
83. K.T. Schomacker, J.K. Frisoli, C.C. Compton, T.J. Flotte, J.M. Richter, N.S. Nishioka, T.F. Deutsch, 'Ultraviolet Laser-induced Fluorescence of Colonic Tissue: Basic Biology and Diagnostic Potential', *Lasers Surg. Med.*, **12**(1), 63–78 (1992).
84. C. Eker, S. Montan, E. Jaramillo, K. Koizumi, C. Rubio, S. Andersson-Engels, K. Svanberg, S. Svanberg, P. Slezak, 'Clinical Spectral Characterization of Colonic Mucosal Lesions Using Autofluorescence and Delta Amniolevulinic Acid Sensitization', *Gut*, **44**(4), 511–518 (1999).
85. R.M. Cothren, M.V. Sivak, J. Van Dam, R.E. Petras, M. Fitzmaurice, J.M. Crawford, J. Wu, J.F. Brennan, R.P. Rava, R. Manoharan, M.S. Feld, 'Detection of Dysplasia at Colonoscopy Using Laser Induced Fluorescence: a Blinded Study', *Gastrointest. Endosc.*, **44**(2), 168–176 (1996).
86. T.D. Wang, J.M. Crawford, M.S. Feld, Y. Wang, I. Itzkan, J. Van Dam, 'In Vivo Identification of Colonic Dysplasia Using Fluorescence Endoscopic Imaging', *Gastrointest. Endosc.*, **49**(4), 447–455 (1999).
87. M.A. Mycek, K.T. Schomacker, N.S. Nishioka, 'Colonic Polyp Differentiation Using Time-resolved Autofluorescence Spectroscopy', *Gastrointest. Endosc.*, **48**(4), 390–394 (1998).
88. N. Ramanujam, M.F. Mitchell, A. Mahadevan-Jansen, S. Thomsen, G. Staerkel, A. Malpica, T. Wright, A. Atkinson, R. Richards-Kortum, 'Cervical Pre-cancer Detection Using a Multivariate Statistical Algorithm Based on Laser Induced Fluorescence Spectra at Multiple Excitation Wavelengths', *Photochem. Photobiol.*, **64**, 720–735 (1996).
89. K. Tumer, N. Ramanujam, J. Ghosh, R. Richards-Kortum, 'Ensembles of Radial Basis Function Networks for Spectroscopic Detection of Cervical Precancer', *IEEE Trans. Biomed. Eng.*, **45**(8), 953–961 (1998).
90. S. Lam, T. Kennedy, M. Unger, Y.E. Miller, D. Gelmont, V. Rusch, B. Gipe, D. Howard, J. LeRiche, A. Coldman, A. Gazdar, 'Localization of Bronchial Intraepithelial Neoplastic Lesions by Fluorescence Bronchoscopy', *Chest*, **113**(3), 696–702 (1998).
91. F. Koenig, F.J. McGovern, A.F. Althausen, T.F. Deutsch, K.T. Schomacker, 'Laser Induced Autofluorescence Diagnosis of Bladder Cancer', *J. Urol.*, **156**(5), 1597–1601 (1996).
92. M. Anidjar, D. Ettori, O. Cussenot, P. Meria, F. Desgrandchamps, A. Cortesse, P. Teillac, A. Le Duc, S. Avrillier, 'Laser Induced Autofluorescence of Bladder Tumors: Dependence on Excitation Wavelength', *J. Urol.*, **156**(5), 1590–1596 (1996).
93. M. Panjehpour, B.F. Overholt, T. Vo-Dinh, R.C. Haggit, D.H. Edwards, F.P. Buckely III, 'Endoscopic Fluorescence Detection of High Grade Dysplasia in Barrett's Esophagus', *Gastroenterology*, **111**, 93–101 (1996).
94. J.K. Dhingra, D.F. Perrault, K. McMillan, E.E. Rebeiz, S. Kabani, R. Manoharan, I. Itzkan, M.S. Feld, S.M. Shapshay, 'Early Diagnosis of Upper Aerodigestive Tract Cancer by Autofluorescence', *Arch. Otolaryngol. – Head Neck Surg.*, **122**(11), 1181–1186 (1996).
95. A. Gillenwater, R. Jacob, R. Ganeshappa, B. Kemp, A.K. El-Naggar, J.L. Palmer, G. Clayman, M.F. Mitchell, R. Richards-Kortum, 'Noninvasive Diagnosis of Oral Neoplasia Based on Fluorescence Spectroscopy and Native Tissue Autofluorescence', *Arch. Otolaryngol. – Head Neck Surg.*, **124**(11), 1251–1258 (1998).
96. B. Kulapaditharom, V. Boonkitticharoan, 'Laser Induced Fluorescence Imaging in Localization of Head and Neck Cancers', *Ann. Otol. Rhinol. Laryngol.*, **107**(3), 241–246 (1998).
97. C.S. Betz, M. Mehlmann, K. Rick, H. Stepp, G. Grevers, R. Baumgartner, A. Leunig, 'Autofluorescence Imaging and Spectroscopy of Normal and Malignant Mucosa in Patients with Head and Neck Cancer', *Lasers Surg. Med.*, **25**, 323–334 (1999).
98. M.L. Harries, S. Lam, C. MacAulay, J.A. Qu, B. Palcic, 'Diagnostic Imaging of the Larynx: Autofluorescence of Laryngeal Tumors Using the Helium Cadmium Laser', *J. Laryngol. Otol.*, **109**, 108–110 (1995).
99. M. Zargi, L. Smid, I. Fajdiga, B. Bubnic, J. Lenarcic, P. Oblak, 'Laser Induced Fluorescence in Diagnostics of Laryngeal Cancer', *Acta Oto-Laryngol. Suppl.*, **527**, 125–127 (1997).
100. B.W. Chwirot, S. Chwirot, J. Redzinski, Z. Michiniewicz, 'Detection of Melanomas by Digital Imaging of

Spectrally Resolved Ultraviolet Light Induced Autofluorescence of Human', *Eur. J. Cancer*, **34**(11), 1730–1734 (1998).
101. G. Bottiroli, A.C. Croce, D. Locatelli, R. Nano, E. Giombelli, A. Messina, E. Benericetti, 'Brain Tissue Autofluorescence: an Aid for Intra-operative Delineation of Tumor Resection Margins', *Cancer Detect. Prevent.*, **22**(4), 330–339 (1998).
102. K. Izuishi, H. Tajiri, M. Ryu, J. Furuse, Y. Maru, K. Inoue, M. Konishi, T. Kinoshita, 'Detection of Bile Duct Cancer by Autofluorescence Cholangioscopy: a Pilot Study', *Hepato-Gastroenterology*, **46**(26), 804–807 (1999).
103. B.W. Chwirot, S. Chwirot, W. Jedrzejczyk, M. Jackowski, A.M. Raczynska, J. Winczakiewicz, J. Dobber, 'Ultraviolet Laser Induced Fluorescence of Human Stomach Tissues: Detection of Cancer Tissues by Imaging Techniques', *Lasers Surg. Med.*, **21**(2), 149–158 (1997).
104. R.R. Alfano, G.C. Tang, A. Pradhan, W. Lam, D.S.J. Choy, E. Opher, 'Fluorescence Spectra from Cancerous and Normal Breast and Lung Tissue', *IEEE J. Quant. Electron.*, **23**, 1806–1811 (1987).
105. Y. Yang, A. Katz, E.J. Celmer, M. Zurawska-Szczepaniak, R.R. Alfano, 'Optical Spectroscopy of Benign and Malignant Tissues', *Lasers Life Sci.*, **7**(2), 115–127 (1996).
106. Y. Yang, A. Katz, E.J. Celmer, M. Zurawska-Szczepaniak, R.R. Alfano, 'Excitation Spectrum of Malignant and Benign Breast Tissues: a Potential Optical Biopsy Approach', *Lasers Life Sci.*, **7**(4), 249–265 (1997).
107. P.K. Gupta, S.K. Majumder, A. Uppal, 'Breast Cancer Diagnosis Using Nitrogen Laser Excited Autofluorescence Spectroscopy', *Lasers Surg. Med.*, **21**(5), 417–422 (1997).
108. H. Auler, G. Banzer, 'Untersuchungen uber die Rolle der Porphine bei Geschwulstkranken Menschen und Tieren', *Krebforschung*, **53**, 1423–1425 (1942).
109. R.H. Figge, 'Near Ultraviolet Rays and Fluorescence Phenomena as Aids to Discovery and Diagnosis in Medicine', *Univ. Md. Med. Bull.*, **26**, 165–168 (1942).
110. G.E. Moore, W.T. Peyton, L.A. French, W.W. Walker, 'The Clinical Use of Fluorescin in Neurosurgery', *J. Neurosurg.*, **5**, 392–398 (1948).
111. G.C. Peck, H.P. Mack, H.F. Figge, 'Cancer Detection and Therapy III. Affinity of Lymphatic Tissues for Hematoporphyrin', *Bull. School Med. Univ. Md.*, **38**, 124–127 (1953).
112. G.C. Peck, H.P. Mack, W.A. Holbrook, H.F. Figge, 'Use of Hematoporphyrin Fluorescence in Biliary and Cancer Surgery', *Bull. School Med. Univ. Md.*, **38**, 124–127 (1955).
113. D.S. Rassmussen-Taxdal, G.R. Ward, F.H. Figge, 'Fluorescence of Human Lymphatic and Cancer Tissue Following High Doses of Intravenous Hematoporphyrin', *Cancer*, **8**, 78–81 (1958).
114. R.L. Lipson, E.J. Baldes, A.M. Olsen, 'Haematoporphyrin Derivative: a New Aid for Endoscopic Detection of Malignant Disease', *J. Thorac. Cardiovasc. Surg.*, **42**, 623–629 (1961).
115. R.L. Lipson, E.J. Baldes, A.M. Olsen, 'The Use of a Derivative of Haematoporphyrin in Tumor Detection', *J. Natl. Cancer Inst.*, **26**, 1–11 (1961).
116. R.L. Lipson, E.J. Baldes, A.M. Olsen, 'A Further Evaluation of the Use of Haematoporphyrin Derivative as a New Aid for Endoscopic Detection of Malignant Disease', *Dis. Chest*, **46**, 676–679 (1964).
117. R.L. Lipson, J.H. Pratt, E.J. Baldes, M.B. Dockerty, 'Haematoporphyrin Derivative for Detection of Cervical Cancer', *Obstet. Gynecol.*, **24**, 78–84 (1964).
118. R.L. Lipson, E.J. Baldes, M.J. Gray, 'Hematoporphyrin Derivative for Detection and Management of Cancer', *Cancer*, **20**, 2255–2257 (1967).
119. M.J. Gray, R.L. Lipson, J.V.S. Maeck, L. Parker, D. Romeyn, 'Use of Hematoporphyrin Derivative in Detection and Management of Cervical Cancer: a Preliminary Report', *Am. J. Obstet. Gynecol.*, **99**, 766–771 (1967).
120. H.B. Gregorie, E.O. Horger, J.L. Ward, J.R. Green, T. Richards, H.C. Robertson, T.B. Stevenson, 'Hematoporphyrin Derivative Fluorescence in Malignant Neoplasms', *Ann. Surg.*, **167**, 820–828 (1968).
121. T.B. Leonard, W.L. Beck, 'Hematoporphyrin Fluorescence: an Aid in Diagnosis of Malignant Neoplasms', *Laryngoscopy*, **81**, 365–377 (1971).
122. J.F. Kelly, J.F. Snell, 'Hematoporphyrin Derivative: a Possible Aid in the Diagnosis and Therapy of Carcinoma of the Bladder', *J. Urol.*, **115**, 150–151 (1976).
123. M. Kreigmair, R. Baumgartner, W. Lumper, R. Reisenberg, S. Stocker, A. Hofstetter, 'Fluorescence Cystoscopy Following Intravesical Instillation of 5-Aminolevulinic Acid (ALA)', *J. Urol.*, **149**, 240A (1993).
124. P. Edholm, B. Jacobson, 'Detection of Aortic Atheromatosis In Vivo by Reflection Spectrophotometry', *J. Atheroscler. Res.*, **5**, 592–595 (1965).
125. C. Kitrell, R.L. Willett, C. de los Santos-Pacheco, N.B. Ratliff, J.R. Kramer, E.G. Malk, M.S. Feld, 'Diagnosis of Fibrous Arterial Atherosclerosis Using Fluorescence', *Appl. Opt.*, **24**, 2280–2281 (1985).
126. J. Wu, M.S. Feld, R.P. Rava, 'Analytical Model for Extracting Intrinsic Fluorescence from a Turbid Media', *Appl. Opt.*, **32**(19), 3585–3595 (1993).
127. C.M. Gardner, S.L. Jacques, A.J. Welch, 'Fluorescence Spectroscopy of Tissue: Recovery of Intrinsic Fluorescence from Measured Fluorescence', *Appl. Opt.*, **35**(10), 1780–1792 (1996).
128. H. Zeng, C. MacAulay, B. Palcic, D. McLean, 'A Computerized Autofluorescence and Diffuse Reflectance Spectroanalyzer for In Vivo Skin Studies', *Phys. Med. Biol.*, **38**, 231–240 (1993).

129. A.F. Zuluaga, U. Utzinger, A.J. Durkin, F. Holger, A. Gillenwater, R. Jacob, B. Kemp, J. Fan, R. Richards-Kortum, 'Fluorescence Excitation and Emission Matrices of Human Tissues: a System for In Vivo Measurement and Method of Data Analysis', *Appl. Spectrosc.*, **53**(3), 302–311 (1999).
130. B. Palcic, S. Lam, J. Hung et al., 'Detection and Localization of Early Lung Cancer by Imaging Techniques', *Chest*, **99**, 742–743 (1991).
131. P.S. Anderson, S. Montan, T. Persson, S. Svanberg, S. Tapper, S.-E. Karlsson, 'Fluorescence Endoscopy Instrumentation for Improved Tissue Characterization', *Med. Phys.*, **14**(4), 633–636 (1987).
132. G.A. Wagnieres, A.P. Studzinski, H.E. van den Bergh, 'An Endoscopic Fluorescence Imaging System for Simultaneous Visual Examination and Photodetection of Cancers', *Rev. Sci. Instrum.*, **68**(1), 203–212 (1997).
133. T. Mckechnie, A. Jahan, I. Tait, A. Cuschieri, W. Sibbett, M. Padgett, 'An Endoscopic System for the Early Detection of Cancers of the Gastrointestinal Tract', *Rev. Sci. Instrum.*, **69**(6), 2521–2523 (1998).
134. H. Zeng, A. Weiss, R. Cline, C.E. MacAulay, 'Real-time Endoscopic Imaging for Early Cancer Detection in the Gastrointestinal Tract', *Bioimaging*, **6**(4), 151–165 (1998).
135. P.S. Andersson, S. Montan, S. Svanberg, 'Multispectral System for Medical Fluorescence Imaging', *IEEE J. Quant. Electron.*, **QE-23**(10), 1798–1805 (1987).
136. S. Andersson-Engels, J. Johansson, S. Svanberg, 'Medical Diagnostic System Based on Simultaneous Multispectral Fluorescence Imaging', *Appl. Opt.*, **34**(1), 8022–8029 (1994).
137. G.A. Duvall, J. Kost et al., 'Laser Induced Fluorescence Endoscopy (LIFE): a Pilot Study of a Real Time Autofluorescence Imaging System for Early Detection of Dysplasia and Carcinoma in the GI Tract', *Endoscopy*, **28**, S45 (1996).
138. G.A. DuVall, B.C. Wilson, N. Marcon, 'Tissue Autofluorescence', *Ann. Gastrointest. Endosc.*, **10**, 25–30 (1997).
139. R.W. Dillon, M. Goldstein, *Multivariate Analysis: Methods and Applications*, Wiley, New York, 1984.
140. A. Albert, E.K. Harris, *Multivariate Interpretation of Clinical Laboratory Data*, Marcel Dekker, New York, 1987.
141. R. Richards-Kortum, R.P. Rava, M. Fitzmaurice, J.R. Kramer, M.S. Feld, '476 nm Excited Laser Induced Fluorescence Spectroscopy of Human Coronary Artery: Applications in Cardiology', *Am. Heart J.*, **122**(4), 1141–1150 (1991).
142. D.M. Haaland, 'Multivariate Calibration Methods Applied to Quantitative Analysis of Infrared Spectra', in *Computer-enhanced Analytical Spectroscopy*, ed. P.C. Jurs, Plenum Press, New York, 1–30, Vol. 3, 1992.
143. M.T. Fahey, L. Irwig, P. Macaskill, 'Meta-analysis of Pap Test Accuracy', *Am. J. Epidemiol.*, **141**(7), 680–689 (1995).
144. M.F. Mitchell, 'Accuracy of Colposcopy', *Consult. Obstet. Gynecol.*, **6**(1), 70–73 (1994).
145. G. Panou-Diamandi, N.K. Uzunoglu, G. Zacharakis, G. Filippidis, T. Papazoglou, D. Koutsouris, 'A One Layer Tissue Fluorescence Model Based on Electromagnetic Theory', *J. Electromag. Waves Appl.*, **12**(8), 1101–1121 (1998).
146. A.J. Durkin, R. Richards-Kortum, 'Comparison of Methods to Determine Chromophore Concentrations from Fluorescence Spectra of Turbid Samples', *Lasers Surg. Med.*, **19**, 75–89 (1996).
147. R. Richards-Kortum, R.P. Rava, M. Fitzmaurice, L.L. Tong, N.B. Ratliff, J.R. Kramer, M.S. Feld, 'A One-layer Model of Laser-induced Fluorescence for Diagnosis of Disease in Human Tissue: Applications to Atherosclerosis', *IEEE Trans. Biomed. Eng.*, **36**(12), 1222–1232 (1989).
148. M.J.C. Van Gemert, W.M. Star, 'Relations Between the Kubelka–Munk and the Transport Equation Models for Anisotropic Scattering', *Lasers Life Sci.*, **1**, 287–298 (1987).
149. H. Zeng, C. MacAulay, D.I. McLean, B. Palcic, 'Reconstruction of In Vivo Skin Autofluorescence Spectrum from Microscopic Properties by Monte Carlo Simulations', *J. Photochem. Photobiol. B: Biol.*, **38**(2/3), 234–240 (1997).
150. J. Qu, C. MacAulay, S. Lam, 'Laser Induced Fluorescence Spectroscopy at Endoscopy: Tissue Optics, Monte Carlo Modeling, and In Vivo Measurements', *Opt. Eng.*, **34**, 3334–3343 (1995).
151. S. Avrillier, E. Tinet, D. Ettori, J.M. Tualle, B. Gelebart, 'Influence of the Emission-reception Geometry on Laser Induced Fluorescence Spectra from Turbid Media', *Appl. Opt.*, **37**(13), 2781–2787 (1998).
152. B.W. Pogue, T. Hasan, 'Fluoropore Quantitation in Tissue Simulating Media with Confocal Detection', *IEEE J. Select. Topics Quant. Electron.*, **12**(4), 959–964 (1996).
153. B.W. Pogue, G. Burke, 'Fiber-optic Bundle Design for Quantitative Fluorescence Measurements from Tissue', *Appl. Opt.*, **37**(31), 7429–7436 (1998).
154. J.K. Dhingra, X. Zhang, K. McMillan, S. Kabani, R. Manoharan, I. Itzkan, M.S. Feld, S.M. Shapshay, 'Diagnosis of Head and Neck Precancerous Lesions in an Animal Model Using Fluorescence Spectroscopy', *Laryngoscopy*, **108**(4), 471–475 (1998).
155. E.W.J. Van der Breggen, A.I. Rem, M.M. Christian, C.J. Yang, K.H. Calhoun, H.J.C.M. Sterenborg, M. Motamedi, 'Spectroscopic Detection of Oral and Skin Tissue Transformation in a Model for Squamous Cell Carcinoma: Autofluorescence versus Aminolevulinic Acid-induced Fluorescence', *IEEE J. Select Topics*, **QE2**, 997–1007 (1997).
156. A.M. Kluftinger, N.L. Davis, N.F. Quenville, S. Lam, J. Hung, B. Palcic, 'Detection of Squamous Cell Cancer and Pre-cancerous Lesions by Imaging Tissue

Autofluorescence in the Hamster Cheek Pouch Model', *Surg. Oncol.*, **1**, 183–188 (1992).

157. I. Pathak, N. Davis, Y.N. Hsiang, N.F. Quenville, B. Palcic, 'Detection of Squamous Neoplasia by Fluorescence Imaging Comparing Porfimer Sodium Fluorescence to Tissue Autofluorescence in the Hamster Cheek Pouch Model', *Am. J. Surg.*, **170**(5), 423–426 (1995).
158. J.M. Nauta, O.C. Speelman, H.L. van Leengoed, P.G. Nikkels, J.L. Roodenburg, W.M. Star, M.J. Witjes, A. Vermey, 'In Vivo Photo-detection of Chemically Induced Pre-malignant Lesions and Squamous Cell Carcinoma of the Rat Palatal Mucosa', *J. Photochem. Photobiol., B: Biol.*, **39**(2), 156–166 (1997).
159. H.J.C.M. Sterenborg, S. Thomsen, S.L. Jacques, M. Motamedi, 'In Vivo Autofluorescence of an Unpigmented Melanoma in Mice. Correlation of Spectroscopic Properties to Microscopic Structure', *Melanoma Res.*, **5**, 211–216 (1995).

Glucose, In Vivo Assay of

H.M. Heise
Institut für Spektrochemie und Angewandte Spektroskopie an der Universität, Dortmund, Germany

1 Introduction **1**
2 Medical and Analytical Aspects **2**
2.1 Carbohydrate Metabolism 2
2.2 Glucose in Blood and Tissue 4
2.3 Conventional Assays Based on Invasive Technology 6
3 Optical Spectroscopy for In Vivo Measurements **7**
3.1 Optical Glucose Sensing 7
3.2 Near-infrared Spectroscopy 8
3.3 Tissue Spectroscopy 15
3.4 In Vivo Assay Using Diffuse Reflectance Spectroscopy of Lip Mucosa 18
4 Review on Current Alternative Technologies **23**
4.1 Fluorescence Spectroscopy 23
4.2 Polarimetry 23
4.3 Scattering Approaches 23
4.4 Raman Spectroscopy 24
Concluding Remarks **24**
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **25**
References **25**

In vivo glucose sensing and regulation is necessary for patients with carbohydrate metabolism disorders, particularly those caused by diabetes mellitus. Self-monitoring of blood glucose is part of the daily routine for such patients, and the measurement of capillary blood glucose, using different enzymatic assays in combination with photometric or electrochemical detection, still remains the standard method. Research activities concentrate on developing minimally invasive and noninvasive methodologies, the latter being based on spectroscopic techniques exploiting optical glucose specific characteristics, i.e. wavelength-dependent absorptivities and refractive indices, mainly in the near-infrared (NIR), or the rotation of linearly polarized radiation in the visible range. Owing to the complexity of the integrally probed tissue, and the presence of many interfering compounds, only multivariate spectroscopic measurement strategies, using several wavelengths, can be used for a noninvasive in vivo assay of glucose. Further complications arise from the heterogeneous distribution of glucose in the intravascular, interstitial and intracellular space and differences in their dynamics. In contrast to the established reference methodology which uses capillary blood, especially for the hypoglycemic concentration range, the imprecision observed so far within several experimental optical approaches is regarded as unacceptable for patient self-monitoring. Further research is needed to establish an in vivo glucose assay for diabetic patients, or for direct monitoring in intensive care units and operating theatres, based on optical spectroscopy.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

All vital processes in cells or larger organisms require a source of energy and the supply or removal of essential substances. In this context, aldohexose glucose is a substance of great metabolic importance, because it can be considered as the main energy carrier for the human organism, and is the most abundant monosaccharide in physiological samples. D-Glucose can be found in two different stereoisomers, i.e. the α- and the β-anomeric form. In aqueous solution each form slowly changes, by means of the free aldehyde, into an equilibrium mixture of approximately one-third of α-D- and two-thirds of β-D-glucose **(1)**, a process which is called mutarotation. Due to their chirality, both isomers show different optical activity and may exhibit dissimilar properties in enzymatic reactions.

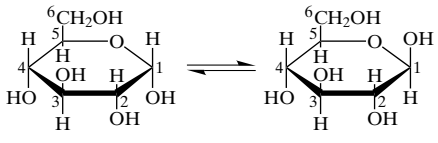

(1)

In the fasting human body at rest, much of the utilized glucose is consumed by the brain and the nervous system, while erythrocytes account for much of the remainder. To avoid large fluctuations of blood glucose, an endocrine regulation exists, a mechanism for which the term homeostasis was coined. The hormone insulin, which is secreted by the B cells of the islets of Langerhans in the pancreas, plays a particularly important part in lowering the glucose concentration. Other effects will also be discussed.

Owing to the importance of D-glucose, most analyses in a clinical laboratory are concerned with this sugar. Continuous measurements are carried out, for example, in critical care units when bedside monitoring of severely ill patients is performed. Furthermore, measurements of glucose are essential for patients without adequate glycemic control, which leads to enormous and adverse variations in blood glucose concentrations. Disorders within carbohydrate metabolism are frequently found, especially diabetes mellitus. About 6% of the adult population in developed countries is affected by this metabolic disease, and there are estimates that as many as 120 million people worldwide suffer from diabetes. According to the World Health Organization there are different categories of diabetes, of which the dominant primary one exists in two different forms, Type I and Type II diabetes.

Type I diabetes, also known as insulin-dependent diabetes mellitus (IDDM), often starts during childhood, and accounts for about 5–10% of the diabetic population. It is caused by an autoimmune disease and is characterized by a complete insulin deficiency due to the destruction of the beta cells producing the hormone.

In Type II diabetes, enough insulin may be available but, due to insulin resistance of the target organs, homeostasis is perturbed. A second mechanism, manifested by progressively diminishing insulin release, can also be responsible for the increase in plasma glucose concentration. This type is often called noninsulin-dependent diabetes mellitus (NIDDM), and occurs mainly in people over 40 years of age.

For all diabetic patients the medical therapy aims to regulate glucose levels to those of metabolically healthy individuals. The results from intensive studies within the Diabetes Control and Complications Trial[1] or the UK Prospective Diabetes Study[2] recommended treatment programs designed to achieve normal or near-normal glycemia (intensified insulin therapy for insulin-dependent diabetic patients), which demands frequent testing of blood glucose. Without proper management, this disease can lead to severe long-term health complications, including perturbations in the microcirculation of blood, with associated blindness, kidney and heart diseases, nerve damage and finally amputation of extremities.

Patient self-monitoring certainly represents the greatest advancement in diabetes therapy since the discovery of insulin. In current practice, diabetics prick a finger with, for example, a lancet, take a drop of blood, and place it on a dry chemical test strip, which is either visually or instrumentally evaluated (also called color reflectance technology). Recent biosensor technology has provided miniaturized alternative equipment that utilizes electrochemical measurements (amperometry). Further development of blood glucose meters aims at using only a few microliters of blood, providing the analysis result within a minute and faster. For a better understanding of the requirements for the development of in vivo glucose assays, some fundamental aspects will be discussed first. Since spectroscopic assays are presented, substances containing carbohydrate moieties showing spectral similarity to the free glucose are worth mentioning.

2 MEDICAL AND ANALYTICAL ASPECTS

2.1 Carbohydrate Metabolism

The principal product of carbohydrate digestion is glucose, which can be found in peripheral venous blood at a concentration of 70–110 mg dL^{-1} (3.9–6.1 mmol L^{-1}). In arterial blood, the plasma glucose level is between 15 and 30 mg dL^{-1} higher than in venous blood. The arteriovenous level difference usually depends on nutritional and metabolic conditions. Glucose, upon entering the cells, becomes phosphorylated to glucose-6-phosphate, after which it may be catabolized. Phosphorylation is actually the initiating reaction for glycolysis, where the citric acid cycle is the final pathway for the oxidation of carbohydrates to CO_2 and H_2O. The metabolism of glucose leads to the intermediate reaction product of pyruvate, which gives lactate following reduction. The latter reactions are carried out under anaerobic conditions, i.e. in the absence of O_2. The production of glucose (gluconeogenesis) from amino acids and glycerol in the liver is also possible.

Another pathway for the depletion of glucose is by the synthesis of macromolecules. In particular, glycogen, a storage form of glucose which is mainly found in the liver and skeletal muscles, is of great importance. The blood also contains glycogen at an average concentration of 27.5 mg L^{-1}, mainly in the granulocytes (45%) and in the platelets (45%). A pathway within skin tissue allows the synthesis of mucopolysaccharides. In principle, the medical physiology and biochemistry concerning glucose is rather complex and is beyond the scope of this article. For more information, see Ganong.[3] In Figure 1 the significant factors that influence plasma glucose levels are sketched. Renal excretion usually takes place when the venous plasma glucose concentration is above 180 mg dL^{-1}, which explains why glucose in urine remains an index component of metabolic disorders in patients. Utilization of glucose by the brain, muscles and other tissues, its metabolic activities in the liver, or its transformation into neutral fat, stored in adipose cells as a long-term energy reserve of the body, must also be mentioned. Modeling the homeostasis of blood glucose

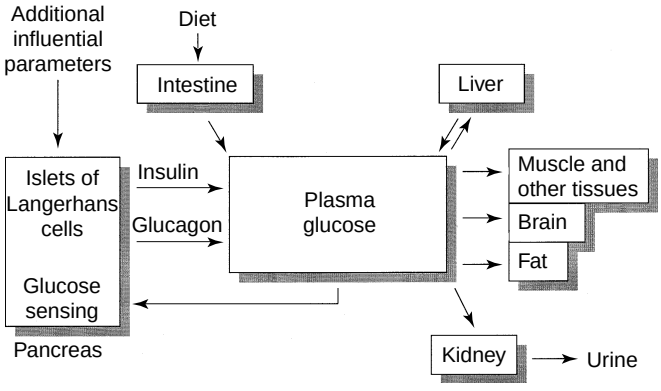

Figure 1 Scheme for blood glucose regulation and principal factors influencing the plasma glucose level.

concentration is indeed complicated, and problems have been discussed by other authors.[4,5]

The principal determinants for glucose entering and leaving the bloodstream are dietary intake, rates of entry into different tissues and the glucostatic activity of the liver. An increase in insulin plasma concentration causes more glucose to be stored, whereas glucagon, another peptide secreted by the A-cells of the islets of Langerhans, shows reciprocal hormonal activity, mobilizing glucose. Excess insulin causes hypoglycemia (lower than normal blood glucose concentration), which can lead as a first complication to unconsciousness. Chronic hyperglycemia (higher than normal glucose concentration) can cause coma and death due to so-called acidosis and dehydration of the body. In the longer term, it leads to nonenzymatic glycation of proteins inducing cellular dysfunction and modifications of the extracellular matrix. After glycation and further transformation, advanced glycation endproducts (AGE) are formed.

A fraction of the blood hemoglobin also becomes glycated. Normally, hemoglobin in adults is found as two variants, HbA1 and HbA2 (the A indicates adult). Hemoglobin consists of 97.5% HbA1 (molecular weight of 64.450). Glycation commences with the nonenzymatic formation of an aldimine (Schiff's base), which is labile. It follows an irreversible rearrangement to the stable ketoamine form. At normal blood glucose concentration, the glycated hemoglobins account for about 6–8% of the total hemoglobin. Owing to the life time of the erythrocytes, an indicator for the mean blood glucose concentration over the previous 10–12 weeks is provided by the glycohemoglobin fraction HbA1c, which is the major glycohemoglobin. It is biochemically characterized as the stable adduct of glucose to the N-terminal valine within the two polypeptide β-chains [*N*-(1-deoxyfructosyl)hemoglobin] with a reference range of 4.5–6.3% of total hemoglobin for healthy individuals. For diabetic persons, the HbA1c fraction is usually elevated to 8% and higher, with values sometimes exceeding 10%. An increase of 4% in this value gives an equivalent concentration of bound glucose of about $4\,\mathrm{mg\,dL^{-1}}$, compared with the glycated hemoglobin fraction normally found. Albumin is also glycated, but is metabolized more rapidly, which allows it to be used as an indicator of more recent average glucose levels.

It must be remembered that sugar moieties of mammalian glycoproteins show significant variations in their structure and relative occurrence during growth, development and differentiation due to their presence in many biological functions. This must be kept in mind when assaying body fluids and tissues by spectroscopic methods. For blood, analytical results for total carbohydrates, bound to proteins, were reported as $2.7\,\mathrm{g\,L^{-1}}$, with hexoses alone providing a level of $1.2\,\mathrm{g\,L^{-1}}$.[6]

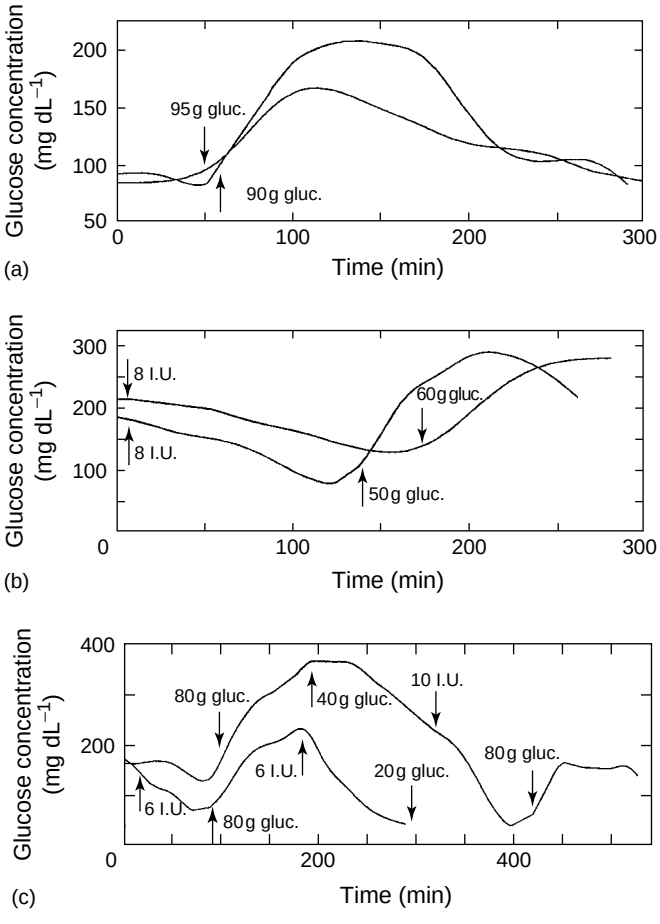

Figure 2 Blood glucose profiles obtained during nonstandard oral glucose tolerance testing (times for carbohydrate ingestion and insulin injections are marked by arrows). (a) Glucose concentrations for two nondiabetic subjects (trace with larger maximum glucose concentration is for a woman; second trace is for a healthy man). (b) Glucose profiles for two male diabetic subjects; and (c) glucose profiles for a third male diabetic subject with experiments on two different days.

The dynamics of changes in blood glucose concentration are of special interest for in vivo assays, since they dictate the necessity for assay speed and timing. In Figure 2 some experiments are presented with deliberately induced changes. Under such conditions it is also usual for calibration experiments for in vivo assays to be carried out. For the clinical diagnosis of diabetes, the response to a standard test dose of glucose is evaluated, a procedure which is called the oral glucose tolerance test (OGTT). Usually, the adult being tested is given a sugar solution equivalent to 75 g of glucose. In healthy individuals the fasting plasma glucose level is less than 115 mg dL^{-1}, and the 2-h value less than 140 mg dL^{-1} with no further rise above 200 mg dL^{-1}. Values between 140 and 200 mg dL^{-1} during the first 2 h of the test indicate an impaired glucose tolerance. Such tests are shown in Figure 2(a) with the results from a healthy person and a woman who shows the possibility of impaired glucose tolerance (the glucose load was larger than in standard oral glucose tolerance testing). The results shown in Figure 2(b) and (c) below were obtained from diabetic patients. Their 2-h blood glucose values are all higher than 200 mg dL^{-1}. It is noteworthy that for some periods very sharp changes in glucose concentration occur (about 40 mg dL^{-1} per 10 min), which exemplifies the need for a rapidly responding sensor system.

2.2 Glucose in Blood and Tissue

A noninvasive and painless measurement assumes that blood sampling is unnecessary. Obtaining quantitative analytical information concerning metabolites by using transcutaneous in vivo spectroscopy of body tissue poses many practical problems. Most information, as discussed above, is gained from sampled blood, which represents only a relatively small fraction of the tissue volume under investigation. This fraction is not well known and may vary considerably depending on the location and the physiological state of the microvasculature. Generally, about 70% of the total blood volume is related to the venous vasculature, 20% is arterial and about 5% is capillary blood. The arterio-venous glucose difference also plays a role in integral tissue probing within spectroscopic assays. Under physiological conditions, the difference can vary considerably as a result of nutritional and metabolic activities. After glucose intake during oral glucose tolerance testing a maximum difference of 45 mg dL^{-1} (2.5 mmol L^{-1}) has been found.[7] Additionally, the blood flow in response to thermoregulatory stimuli can vary between 1 and 150 mL per 100 g of skin tissue,[3] which greatly affects the arterio-venous difference.

For transcutaneous in vivo measurements, the most suitable tissue has to be selected on the basis of possessing a high fraction of blood within its composition. For human skin at different measuring sites, at a given skin temperature, the tip of the index finger showed the highest hemoglobin concentrations which could be related to the tissue blood volume.[8] Our own measurements showed that the human lip was an even better test location for the same reasons (see Figure 3). In Figure 3(a) the diffuse reflectance spectra of different skin tissues are presented, which are dominated by the absorption features of oxygenated hemoglobin (hemoglobin oxygenation saturation, $HbSO_2$, is usually high at around 90%). Figure 3(b) shows the effect of hyperemization of the inner lip tissue by applying a fiber sensor head thermostated at 42 °C, leading to an increase in blood volume and hemoglobin oxygenation. Figure 3(c) illustrates the effect of external pressure applied to the skin tissue, pushing blood out of the probed tissue.

In addition to the intravascular fluid, the interstitial fluid, bathing the different cells, belongs to the extracellular body water compartment. It provides nutrients and oxygen to the tissue and takes up metabolic waste products. The constituents of the interstitial fluid are constantly replenished by the blood. Some fluid from the interstitial space is steadily removed by the lymphatic system. Interstitial fluid is a passive medium with virtually no flow.

The percentage of the intravascular, interstitial and intracellular compartments is tissue dependent. The volume of interstitial fluid, compared to the whole body, is about three times greater than the plasma volume, but the amount of interstitial fluid in a tissue depends on the pressure in the capillaries, the oncotic pressure,[3] the capillary filtration coefficient, the number of active capillaries, the lymph flow, and the total extracellular fluid. However, no significant changes in the concentration of low molecular weight compounds in plasma were found when the plasma volume was decreased by orthostasis. For spectroscopic limb measurements, fluid values of 63% for the intracellular and 37% for the extracellular space were reported, where 27% related to the interstitium and 10% to blood plasma. A model calculation for glucose tissue average was presented for a blood glucose value of 100 mg dL^{-1}, leading to a tissue average of 38 mg dL^{-1}, with more than half of the glucose related to the interstitial fluid (58%), less than a third to blood (26%), and 16% due to the intracellular fluid.[9]

For various reasons it is apparent that the composition of the interstitial fluid is similar to plasma, but some temporal relationships for the concentration profiles of different metabolites can be expected. The physiological balance differs in each of the compartments, and response times to drastic changes in the vascular compartment

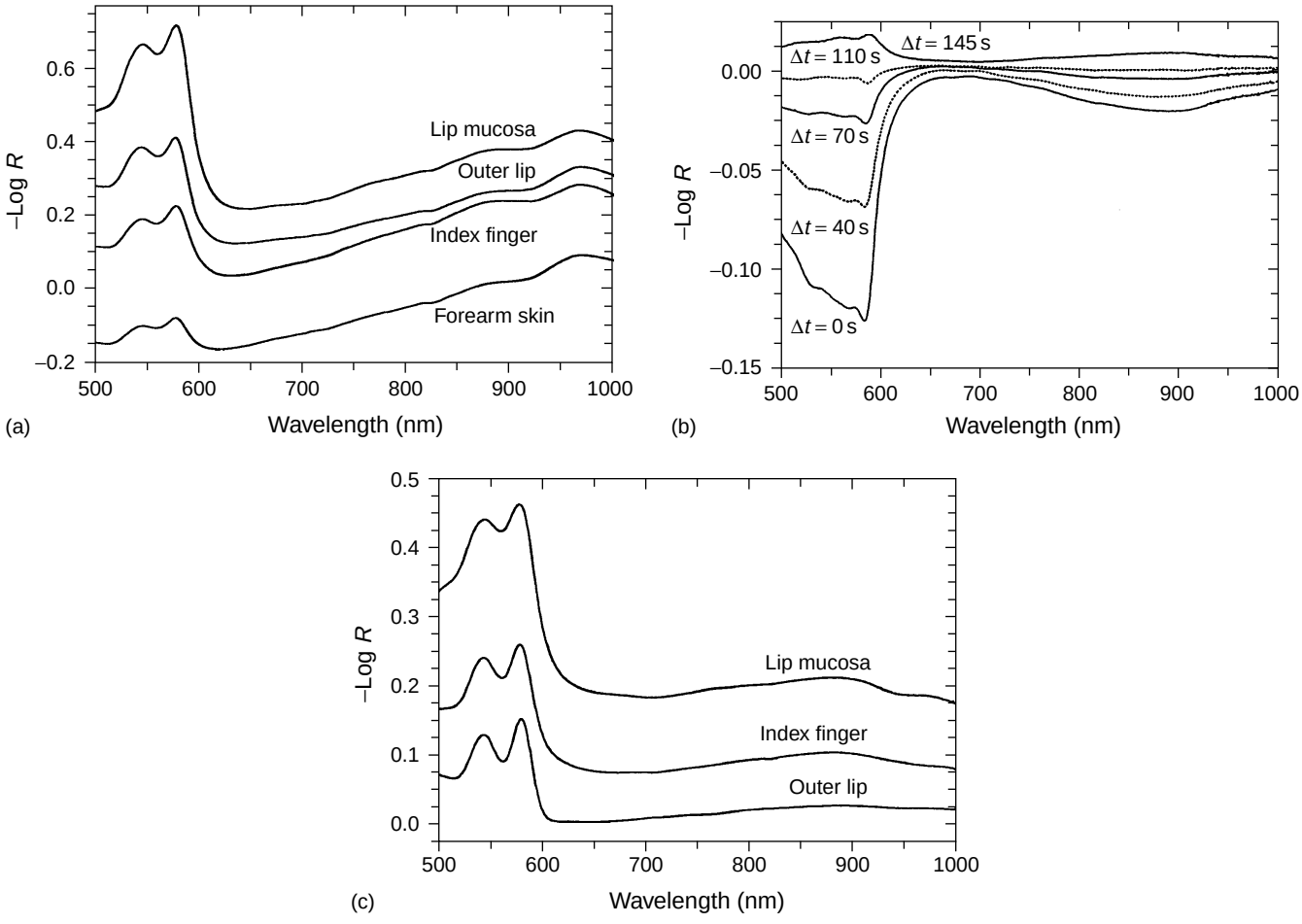

Figure 3 (a) Reflectance visible/NIR spectra of different skin tissues recorded at normal skin temperature by using a bifurcated fiber optic probe (for clarity, the individual spectra are offset). (b) Difference spectra showing the effect of hyperemization of the inner lip tissue by application of a fiber sensor head thermostated at 42 °C (first contact at $\Delta t = 0$ s, difference spectra were calculated versus a lip spectrum taken after 120 s). (c) Difference tissue spectra calculated from spectra under low and high contact pressure applied by the fiber sensor head illustrating the corresponding blood volume changes. (Reproduced in part by permission of Walter De Gruyter from H.M. Heise, A. Bittner, R. Marbach, 'Near-infrared Reflectance Spectroscopy for Non-invasive Monitoring of Metabolites', *Clin. Chem. Lab. Med.*, **38**, 137–145 (2000).)

may vary due to the physiological condition, such as blood circulation. For glucose, concentration differences between blood and the subcutaneous (s.c.) or dermis interstitial fluid also exist.

There is still controversy concerning the absolute glucose concentration in the interstitial fluid, when compared to that of blood. Some investigators, using microdialysis probing, claim the s.c. interstitial glucose concentration, under steady state conditions, to be practically identical to that in blood, whereas other authors have found values of only about 75% of the blood glucose concentration, which supports the theory of glucose extraction by the peripheral tissue. In a recent review, the results from many different investigations were compiled and experimental effects from different analytical probes were discussed.[9] According to theory, the physiological diffusion delay is caused by a change in the diffusion equilibrium as obtained under steady-state conditions. This change can be triggered, for example, by a sharp rise in blood glucose concentration under the same cellular glucose utilization rate. Thus, the physical transfer of glucose from blood, through the interstitial space to the cells, is maintained. The modeling of stationary glucose diffusion processes from a central capillary to the neighbouring tissue under constant cellular glucose utilization, the so-called Krogh's cylinder model, in addition to the time-dependent diffusion processes, have been described previously.[10] Key parameters are the diffusion constant, vessel diameter and spatial capillary density, allowing the calculation of average tissue concentrations.

We have been modeling the transport processes by using moving average filters or impulse-invariant designed Butterworth filters of the first order.[11] The

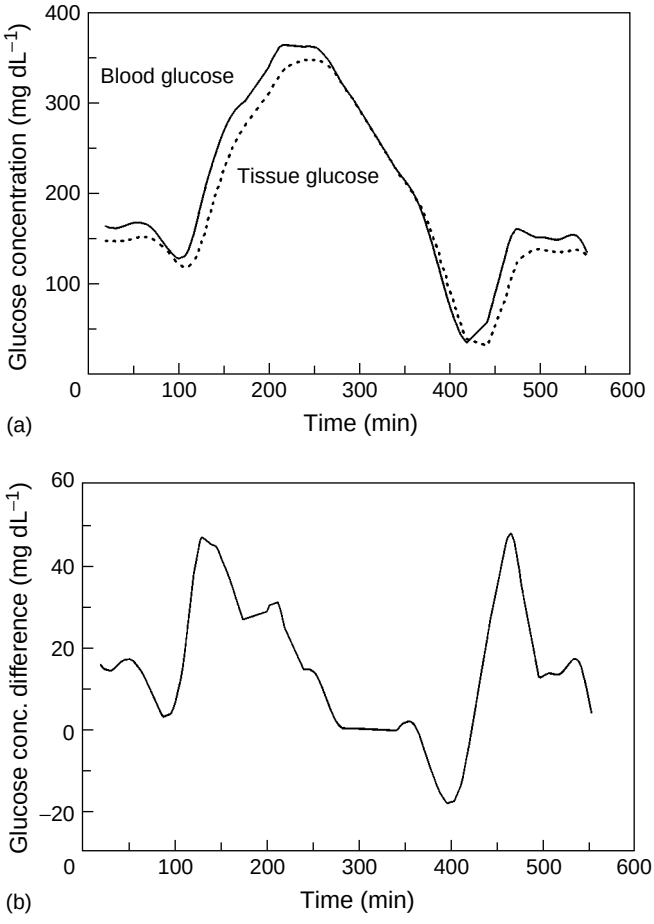

Figure 4 (a) Blood glucose profile of a diabetic patient and a model calculation for tissue glucose concentrations taking into account the delay of glucose diffusion into the interstitial fluid space and a constant cellular glucose utilization rate. (b) Corresponding difference values between blood and mean tissue glucose concentrations.

delays seen by taking the diffusion processes and the cellular glucose utilization into account were dependent on the rate of blood glucose changes and glucose extraction, which may explain some of the large variations seen within lag-time studies. A model calculation is given in Figure 4(a), where the time-dependent traces of blood glucose and that of an average tissue measurement are shown. Figure 4(b) provides absolute differences, particularly in the hypoglycemic range, where systematic errors are relatively large for estimating a blood glucose level from an integral tissue measurement. For changes in the concentration profiles, delays between 10 and 20 min in relation to glycemia were observed. In addition, the impoverishment of glucose in a tissue can precede hypoglycemia, which can be used as an early marker.[12,13] It was possible to model such effects by taking standard glucose diffusion and different cellular glucose uptakes into account.

Blood flow cannot be neglected as a variable in interstitial glucose studies (see literature in Roe and Smoller[9]), although the reduction of gradients within the tissue, e.g. along the blood capillaries, is certainly a factor which accounts for an increase in integral tissue glucose, particularly in skin tissue. Skin is very sensitive to circulatory shunting effects, as seen in Figure 3(b), and the rise in blood flow parallels vasodilation, equivalent to an increase in blood volume.

The permeability of the capillary walls is crucial to the rate of metabolite exchange between blood and interstitial fluid. Diabetes is known to affect the vasculature, and vast changes in the microcirculation can occur with disease progression.

2.3 Conventional Assays Based on Invasive Technology

Despite the abundance of diverse test strip devices for blood glucose measurement, the support of a clinical chemistry laboratory is still required. In the past, it had been the exclusive task of such a laboratory to analyse blood or other body fluids. Since blood serves as the primary metabolic transport system within the body, medical interest focuses on monitoring its metabolite concentrations as a guide to the metabolic state. Other fluids, such as urine, saliva, sweat and tears, have also been analysed for glucose, but no useful correlation with the concentrations in blood could be established, although glucose in urine is used as an index component for metabolic disorders.

Before the advent of enzymatic and spectroscopic assays, a variety of chemical methods[14] had been developed for glucose analysis, whereas enzymatic procedures are generally used nowadays in clinical and food laboratories.[15]

Enzymatic methods are widely used for reference analysis, necessary for calibrating spectroscopic assays. For glucose determination three enzyme systems are commercially available. The hexokinase (HK) method has been internationally recommended as a reference method. Here, a two-step reaction is necessary: in the first reaction HK (EC 2.7.1.1) is used, followed by an indicator reaction catalyzed by glucose-6-phosphate dehydrogenase (G-6-P-DH; EC 1.1.1.49), which provides the specificity of the method. The second common method uses glucose dehydrogenase (G-DH; EC 1.1.1.47), which requires the addition of mutarotase, accelerating the conversion of the anomeric α- into the β-form of D-glucose. A third method uses the enzyme glucose oxidase (GOD; EC 1.1.3.4), which catalyzes the oxidation of glucose to gluconic acid providing H_2O_2. An indicator reaction is usually necessary for quantification by spectrophotometry. Other detection methods follow oxygen consumption (e.g. by using the Clark electrode) or the formation of hydrogen

peroxide amperometrically. Assay kits are commercially available, and various automated analyzers have been developed.

Many electrochemical biosensors are based on these measurement principles allowing in vivo or ex vivo glucose measurements. Similarly, solid-phase reagent systems for self-monitoring of blood glucose rely on enzymatic methodology. These systems have the advantage that the chemical reaction products can be evaluated visually, or with greater precision by means of a reflectance photometer. The availability of reagent strips and pocket-sized reflectance meters has had considerable impact on the management of diabetes, allowing a more specific therapy. This technology is now challenged by biosensory methods, which require sample volumes of only a few microliters.

The usual invasive methods of sampling disrupt the skin barrier and puncture the capillary bed in order to draw a blood sample. The current trend is certainly to enable minimally invasive measurement of physiological analytes. An emerging technology uses reverse iontophoresis to electroosmotically extract body fluid with glucose. Thermal microporation and suction have both been proposed for interstitial fluid harvesting, but these competing techniques still face various problems for in vivo assays. For reviews see Roe and Smoller[9] and Klonoff.[16]

At present, intermittent invasive testing is widely practised, although continuous monitoring offers many advantages, since a glucose sensor could be coupled to an insulin pump, thus completing the artificial pancreas. Therefore, much effort is currently under way to develop sensors for continuous monitoring. However, invasive in vivo sensors face the problem of biocompatibility, which usually compromises the performance of the sensing device. In addition there is the risk of infection, and the implantation into the vascular space, as an option for long-term monitoring, is not advisable due to the formation of thrombi, as this could present a major health risk to the patient. Another approach may be the implantation of an electrochemical biosensor into the subcutaneous tissue. An alternative to intravascular measurement is the analysis of the interstitial fluid, but complications have been reported for such in vivo sensing. However, ex vivo sensor applications avoid these problems by using microdialysis probes, although this is at the cost of a dilution of the fluid under investigation. Other systems proposed for on-line continuous ex vivo monitoring rely on open tissue perfusion. In the longer term, noninvasive technology, based, for example, on NIR spectroscopy, will be available for intermittent and continuous monitoring of blood glucose concentration. The benefits of this are obvious, because such a transcutaneous measurement is painless, and does not require a blood sample to be taken.

3 OPTICAL SPECTROSCOPY FOR IN VIVO MEASUREMENTS

3.1 Optical Glucose Sensing

Many medical instruments, including those for magnetic resonance imaging and spectroscopy, rely on the interaction of electromagnetic radiation with body fluids and tissues. The same is true of optical techniques, which find numerous applications in the medical field and offer the potential for the development of small, rugged and moderately priced instruments. The use of infrared (IR) spectroscopy for biomedical applications has increased tremendously in recent years. One objective is to develop IR spectroscopic methods for medical diagnostics (IR pathology) in order to investigate the transformation from healthy to diseased cells and tissues. Quantitative analysis of body fluid constituents is another major field in clinical chemistry where optical analytical methods may find a use.

Generally, spectroscopic methods exploit specific optical characteristics of the analyte, e.g. wavelength-dependent absorptivities or refractive indices, mainly in the NIR or, as for optically active glucose, the rotation of linearly polarized radiation. For in vitro and in vivo assays, IR, Raman spectroscopy and other optical techniques such as polarization, scattering and photoacoustic approaches have recently been proposed for measuring glucose concentrations in blood, interstitial fluid, tissues or the aqueous humor of the eye. Many of these methods have worked well for carefully controlled in vitro samples, e.g. blood plasma or other body fluids. However, in vivo measurements are confounded by large variations in tissue physiology, including a greater sample complexity and microheterogeneity as well as spectral artifacts caused by the optical properties of tissues.

The Raman effect involves inelastic radiation scattering, which results in a characteristic spectrum. The photoacoustic NIR measurement technique has also been used and promising results were recently published by MacKenzie et al.[17] A further technique currently under investigation is fluorescence spectroscopy. Usually, the specimens studied are multicomponent systems, with individual components contributing many overlapping signals. Such a situation requires the application of multivariate techniques, such as measurement at several wavelengths and chemometric methods to determine concentration. Until now, the NIR absorption approach has been shown to be the most effective method of proving in vivo blood glucose monitoring capabilities, and this method will be presented in more detail. Finally, results from novel techniques using pulse spectrometry (NIR photoplethysmography) are discussed. In the final section, some optical techniques are reviewed and their

impact on the development of in vivo assays of glucose is discussed.

3.2 Near-infrared Spectroscopy

For quantitative analysis, the mid-IR and NIR spectral ranges are of special importance. In the mid-IR (2.5 μm–25 μm; 4000–400 cm^{-1}) significant absorptions caused by excitation of molecular vibrations assigned to fundamental and combination bands occur. Such absorptions are characteristic of a molecule and are useful for identification. In the NIR range (780–2500 nm; 12 800–4000 cm^{-1}) molecules absorb radiation for excitation of overtone and combination vibrations. The most intense bands are mainly from vibrations of molecular moieties that involve hydrogen atoms, e.g. O–H, C–H and N–H. The short-wave near-infrared (SWNIR) carries higher overtone bands with even smaller absorptivities, so that the optical sample pathlength is usually increased, and aqueous solutions can be measured with cuvettes of 1-cm pathlength. The visible and the SWNIR range exhibit significant absorption bands due to electronic transitions.

In Figure 5, some NIR spectra of substances measured mainly as crystalline powders are shown. These are important in the analysis of blood, and the individual bands can be assigned to certain vibrations of molecular substructures.[18] The intensities of the higher overtone vibration bands decrease compared to lower

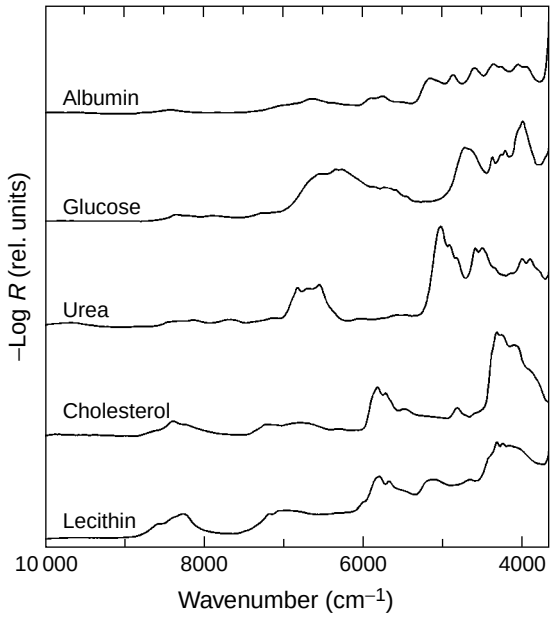

Figure 5 NIR spectra of several compounds important in clinical blood analysis, apart from lecithin, which was measured as a thin layer backed by a diffuse reflecting gold-coated substrate (transflectance technique); the other diffuse reflectance spectra were measured from powders.

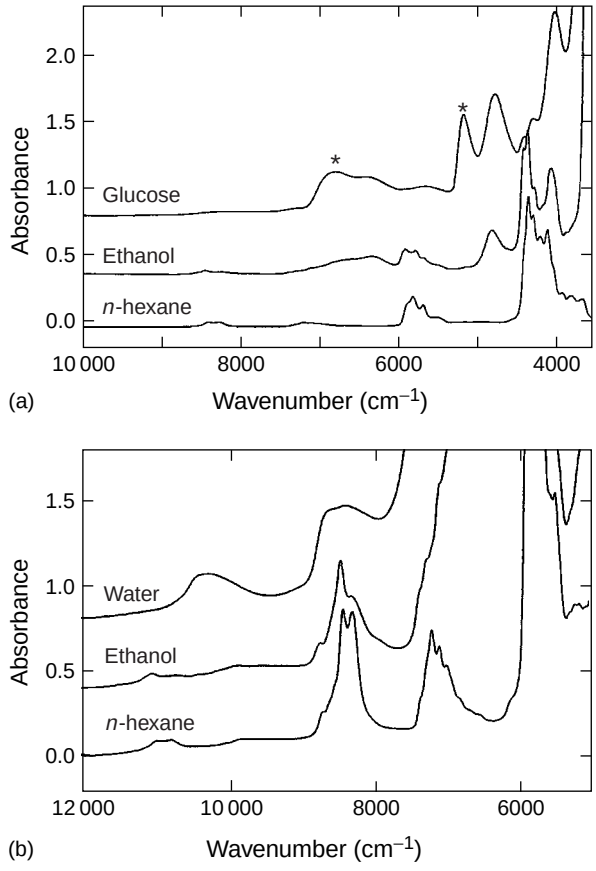

Figure 6 NIR spectra of model compounds. (a) Glucose spectrum recorded from a glass-like sample prepared as a layer on a diffuse reflecting gold-coated substrate (bands indicated by an asterisk are assignable to water residue). The spectra of ethanol and hexane were recorded using a 0.5-mm pathlength quartz-cell. (b) SWNIR absorbance spectra recorded using a 10-mm pathlength quartz cell. (All measurements were carried out at 30 °C. For clarity some spectra are offset.)

overtone bands. This becomes clearer when the spectra of small, simply structured molecules are interpreted. In Figure 6(a) a comparison between glass-like glucose, which gives a similar spectrum to that in aqueous solution, ethanol and hexane is given, and the assignment of bands involving hydrocarbon subunits and C–OH groups is evident through the corresponding band frequencies. In the glucose spectrum, the band indicated with an asterisk at around 5200 cm^{-1} belongs to some water residue. SWNIR spectra recorded at 1-cm pathlength contain similar structural information (Figure 6b).

For different classes of compounds, NIR spectra show significant differences. Frequently, substances with similar structures, which could easily be quantified by using enzyme methods, have to be determined, but the similarity of their NIR spectra can lead to a failure of the spectroscopic assay. Of particular importance are the monosaccharides found in biosystems. For example,

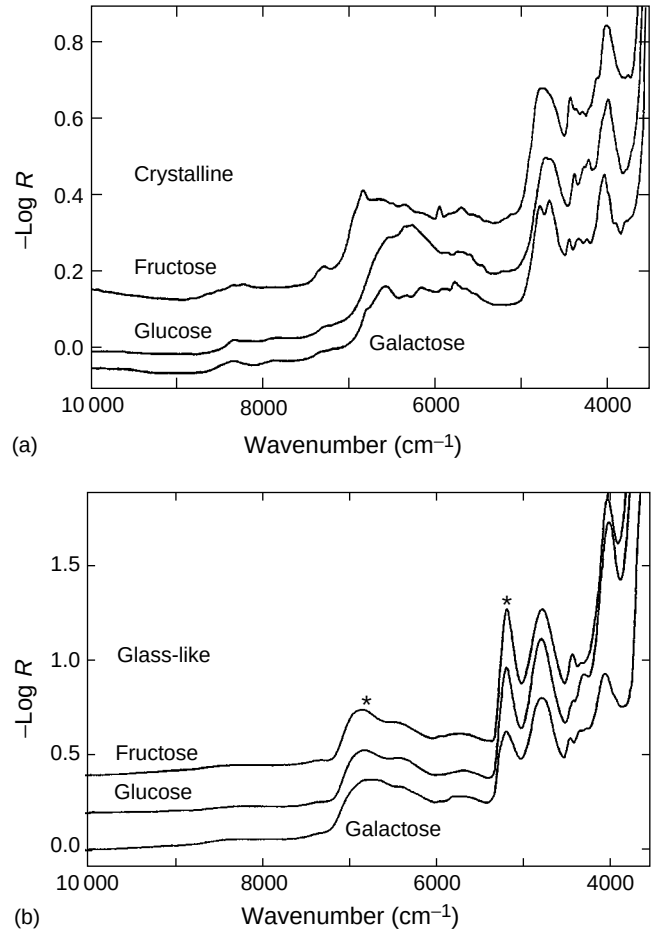

Figure 7 Comparison of pure carbohydrate NIR spectra. (a) Diffuse reflectance spectra of different crystalline monosaccharides. (b) Spectra from glass-like sugar samples obtained from syrup preparations and measured in transflectance. (Reproduced by permission of American Institute of Physics from H.M. Heise, A. Bittner, *Fourier Transform Spectroscopy: 11th International Conference*, ed. J.A. de Haseth, American Institute of Physics, New York, AIP Conf. Proc., **430**, 274–277 (1998).)

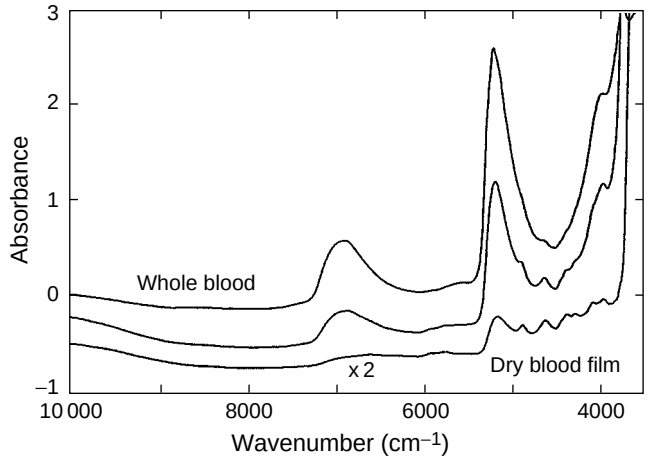

Figure 8 Diffuse reflectance spectrum of a whole-blood film on a gold-coated diffuse reflecting substrate; further spectra were recorded after different stages of water evaporation (both lower spectra were scaled up by a factor of 2).

the spectra of crystalline glucose, fructose and galactose (Figure 7a) show impressive differences which can be used advantageously for substance identification. Quite different features exist when glass-like samples from syrup preparations and samples after the evaporation of water are studied (Figure 7b). These spectra are similar to those obtained from difference spectroscopy of aqueous solutions to be discussed later. The largest spectral differences are found at a spectral range around 4000 cm^{-1}. This is the reason why many researchers favor this spectral range, as it can be used for aqueous samples by utilizing cuvettes of about 0.5-mm optical pathlength.

The biomedical samples considered here contain water as the main constituent, and its optical properties, in particular the characteristics of its radiation absorption in terms of the absorption coefficient, dominate the IR spectra. An interesting experiment is the follow-up of the evaporation of water in whole blood by diffuse reflectance, the spectra of which are shown in Figure 8. The disappearance of the water absorption bands is evident and finally, spectral features from proteins such as hemoglobin and albumin dominate the dry film spectrum.

The quantitative investigation of biofluids by multivariate IR spectrometry can be carried out successfully, and novel assays for reagent-free multicomponent analysis have been presented in the literature. Whole blood is certainly the most frequently studied fluid, but other liquids such as aqueous humor, urine, saliva and synovial fluid have been analyzed by spectroscopic methods. For literature reviews, see Heise et al.[19] and Khalil.[20]

As water plays an important role in spectroscopic bioassays, some further characteristics are discussed in order to better understand the challenges encountered when devising in vivo assays. Figure 9(a) shows the water spectra recorded when two cells of different optical pathlength were used. The spectra illustrate the intense absorptions caused by H–O–H overtone and combination vibrations. In addition, the hydrogen bonding network of the water molecules in the liquid phase must be considered, as it leads to a tremendous broadening of the absorption bands due to variations in the molecular force field and structure that result from the intermolecular interactions. The association equilibrium between water molecules is very sensitive to temperature change, which can be estimated from spectral changes seen in Figure 9(b) and (c). This is the reason why spectroscopic in vitro experiments are usually carried out under cell thermostatting. On the other hand, the temperature dependencies of the

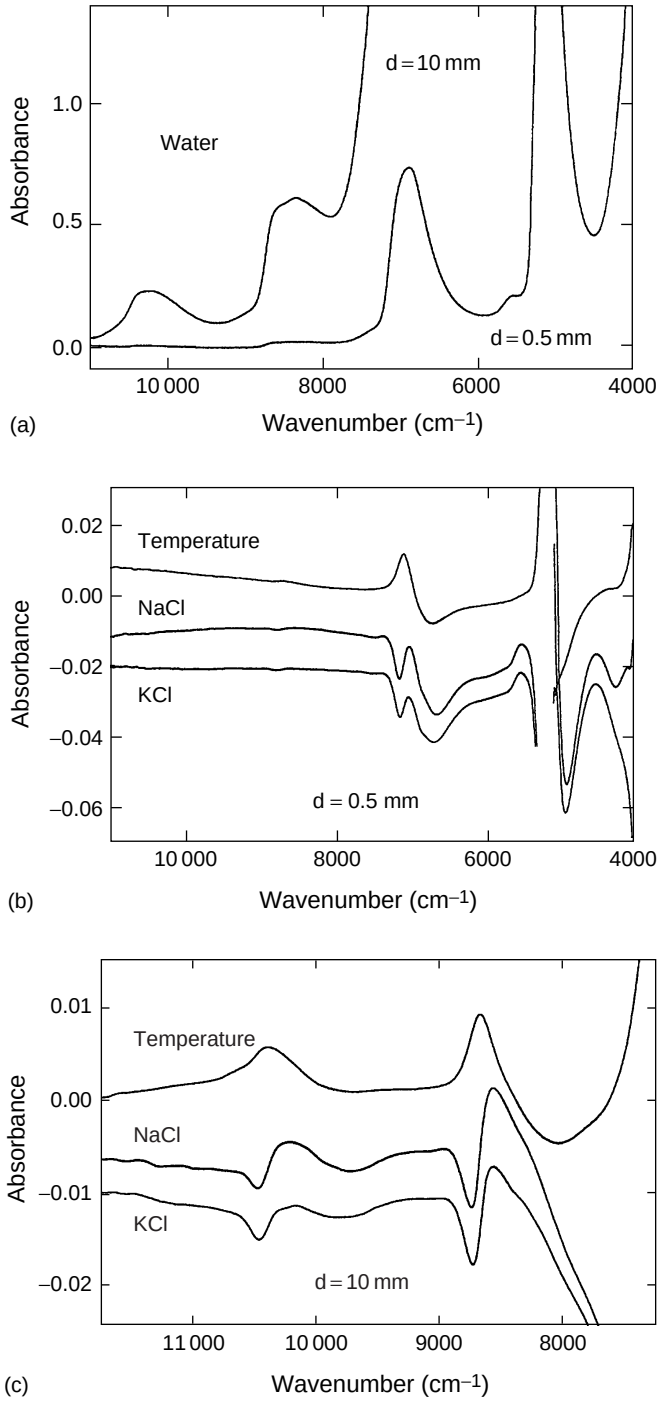

Figure 9 (a) NIR spectra of water recorded at 25 °C and different cell pathlengths. (b) Difference spectra recorded with a 0.5-mm pathlength quartz cell: water at 32.5 °C, an aqueous solution of NaCl and of KCl (5% each) versus a water spectrum recorded at 30 °C. (c) SWNIR spectra recorded using a 10-mm pathlength cell; for further conditions, see legend to (b). (Reproduced by permission of American Institute of Physics, from H.M. Heise, A. Bittner, *Fourier Transform Spectroscopy: 11th International Conference*, ed. J.A. de Haseth, American Institute of Physics, New York, AIP Conf. Proc., **430**, 274–277 (1998).)

NIR water spectrum can be exploited for temperature determination.

Additionally, the hydrogen bonding between the water molecules is influenced by electrolytes and, to a rough approximation, the change in the IR spectra of aqueous solutions is similar to that caused by a change of temperature of pure water (Figure 9b and c). The linear dependencies exhibited by the water spectrum can be used under special conditions for determining NaCl concentration. By using the NIR spectra of aqueous solutions the measurement of pH is also possible, and titrations have been followed by NIR spectroscopy.[21]

Similarly, strong interactions between water and biopolymers can occur, which also affect the hydrogen-bonded water clusters, leading to some significant shifting of water absorption bands. Proteins are particularly prone to such interactions. For example, collagen, in the form of gelatine or cartilage, and polysaccharides, such as dextrin and hyaluronate, can easily be hydrated. Recently, the water in skin has been studied in detail with respect to such effects.[22] This clearly demonstrates the complexity of water spectra and the numerous factors which must be considered when using NIR spectroscopy for the analysis of biofluids.

The spectra of several compounds in aqueous solution, of particular interest for the in vivo glucose assay, are presented in Figure 10. Here, again, absorbance spectra within the SWNIR and the conventional range are highlighted using transmission cells of appropriate pathlength. The difference spectra were compensated for the absorbance of water by using a spectrum of pure water, but due to the interactions between solute and solvent the compensation is incomplete (note the sharp spectral dips at the position of the water bands). As discussed, apart from differences above 4000 cm^{-1}, there are large spectral similarities. An interesting compound is hydroxyethyl starch, which is water soluble and used medically as a blood plasma substitute for reducing blood viscosity or compensating for blood loss. The spectral features of the polysaccharide are due to glucose subunits, and resemble those of glycogen. Another compound, ethanol, whose spectrum is given in Figure 6 can be found in blood after consumption of alcoholic beverages. There are certain conditions where the a priori knowledge of the absence of such substances is critical as their presence could lead to large perturbations in assay performance. See, for example, Berentsen.[23]

Owing to the large water absorptivities in the mid-IR, penetration of IR radiation is not sufficient to establish transcutaneous measurements of metabolites in tissue, in particular when the so-called attenuated total reflectance technique is applied for in vivo measurements.[24] This problem may be overcome by using NIR spectroscopy for

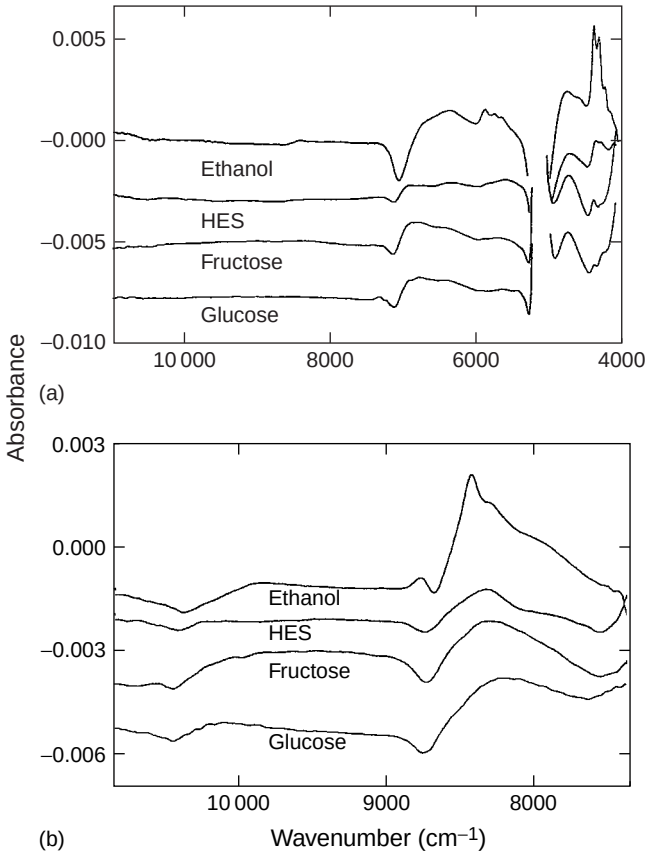

Figure 10 Difference spectra of some diluted aqueous solutions of compounds important for biotic assays with partial water absorbance compensation. (a) NIR data from solutions of 1.5% concentration recorded at 30 °C with 0.5-mm pathlength (HES, hydroxyethyl starch). (b) SWNIR spectral range recorded from solutions of 0.5% concentration each recorded at 30 °C with a 10-mm pathlength cell. (Reproduced by permission of American Institute of Physics, from H.M. Heise, A. Bittner, *Fourier Transform Spectroscopy: 11th International Conference*, ed. J.A. de Haseth, American Institute of Physics, New York, AIP Conf. Proc., **430**, 274–277 (1998).)

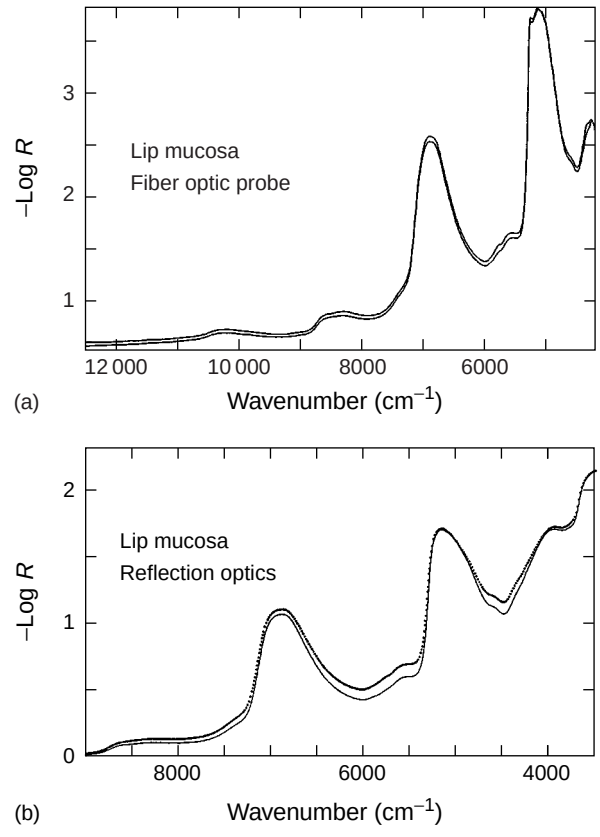

Figure 11 Diffuse reflectance spectra of lip mucosa recorded by using different accessories. (a) Lip spectra of two different men measured against a white reflection standard by using a bifurcated fiber-optic probe. (b) Mean spectra from a group of 390 spectra from 133 patients (upper dotted trace) and from 219 spectra of one male (lower solid trace) obtained by using a diffuse reflectance accessory; the latter spectra were offset to give the same $-\log R$ value at 9000 cm^{-1}, deliberately shifted to zero (for the respective accessories, see Figure 12).

which the water absorptivities are much smaller. Due to the optical constants of tissue, i.e. the spectral absorption and scattering coefficients, a wavelength-dependent penetration depth for such radiation exists. For wavelengths between 600 and 1300 nm (16 700–7700 cm^{-1}, the so-called therapeutic window) transmission measurements of body tissues are possible.[25] The spectral range above these wavelengths, in particular the interval between 1600 and 1850 nm (6250–5400 cm^{-1}), is dominated by two water absorption bands, and is accessible by diffuse reflectance measurements,[26] due to the tissue scattering characteristics.

Figure 11 shows several spectra of lip mucosa, which was chosen as an example of soft tissue, obtained using different instrumentation (more details on tissue optics and spectroscopy are given in later sections). The accessories used were an optical fiber probe and a mirror based device. It can be seen that tissue optics play an important role in recording the in vivo spectra. In particular, the type and collection geometry of the accessory and source–detector separation are critical factors in the optimization of in vivo assays.

3.2.1 Instrumentation and Novel Developments

Different types of spectrometer are used for NIR spectrometry, and there is a growing trend towards miniaturization. In the past, scanning grating spectrometers were the most widespread. Originally, NIR spectroscopy was carried out by extending the range of ultraviolet/visible (UV/VIS) instrumentation. Nowadays, many instruments are based on polychromators with diode arrays to measure the whole spectrum simultaneously. Other NIR measurement technology uses acousto-optical tunable filters. An acoustic wave is

generated from a transducer, bonded to one face of a birefringent crystal, to produce a refractive index modulation. By tuning a range of radio frequencies it is possible to change the diffraction within the crystal, allowing a spectrum to be scanned at a special band pass. Other potent technology uses Fourier transform (FT) spectrometers, which are widely distributed and based mainly on a Michelson interferometer. Their primary information is coded in the interferogram which is Fourier-transformed to obtain the spectrum.[27]

The detection electronics are an important part of the instrumentation. Usually for the NIR range, photodiodes of different semiconductor materials are used. Detectors manufactured by silicon technology can be photodiodes or extremely sensitive charge-coupled devices (CCD). Unfortunately, their cutoff is above 1050 nm, so that other materials for photon detection are preferred (e.g. InGaAs, Ge or InSb). The diodes from the latter material are generally used at −196 °C. Nowadays, linear InGaAs arrays with up to 1024 pixels are available, often in combination with thermoelectric cooling by a Peltier device, allowing temperatures down to −30 °C and lower.

As thermal sources, tungsten–halogen lamps are used to provide the broad emission of electromagnetic radiation according to Planck's law. Another option for radiation sources are light-emitting diodes (LEDs) and tunable lasers. With the limited tunability of the latter, the spectral range necessary for the applications discussed here can only be covered so far by a large number of different laser diodes. However, semiconductor diodes operating at about room temperature or below by thermoelectric cooling will be more accessible in the future. With the availability of such devices, other measurement techniques, as used in laser photo-acoustic spectroscopy, are possible with special probes for radiation delivery and detection.[28]

Miniaturization and integration of optical and optoelectronic instruments is necessary to achieve portable, personal and hand-held instrumentation with minimal power consumption and high reliability, while maintaining low production costs. The question is whether low-level integration as seen, for example, with the availability of a four-channel filter based IR detector assembly is sufficient for the ambitious requirements of multi-wavelength in vivo measurements.[29] For greater flexibility, mini- and microspectrometers are currently under development. Potent instrumentation is based on a concept which uses a LED array spectrometer.[30] A linear LED array serves as the radiation source, coupled to a fixed grating monochromator for wavelength resolution and stability. The exit beam, with up to 32 different possible wavelengths in the SWNIR spectral range, delivers spectral pulses through an optical fiber, but extension to longer wavelengths has been promised. High integration of an IR spectrometer based on silicon was also achieved, and the developments reached in microsystems for spectroscopic applications have recently been described.[31]

Modern spectrometers often use optical fibers for the guidance of radiation, either employing single fibers or fiber bundles, sometimes with the aim of adapting the bundle cross-section to the spectrometer needs (a circular shape is optimal for focusing radiation into the fiber; the end could be shaped to suit the spectrometer entry slit). One affordable material useful in the NIR is an especially low-OH grade quartz with high transmission and the necessary range to reach into the mid-IR range, although other materials have also been in use.[32] Additionally, measurement accessories for routine and process analysis can be constructed from such fibers. In Figure 12(a), a bifurcated, Y-shaped fiber-optic accessory is sketched. The fiber bundles, one for sample illumination and the other for radiation collection, usually contain fibers of about 0.2 mm diameter, and at the common end, all fibers can be arranged at random. Other configurations, for

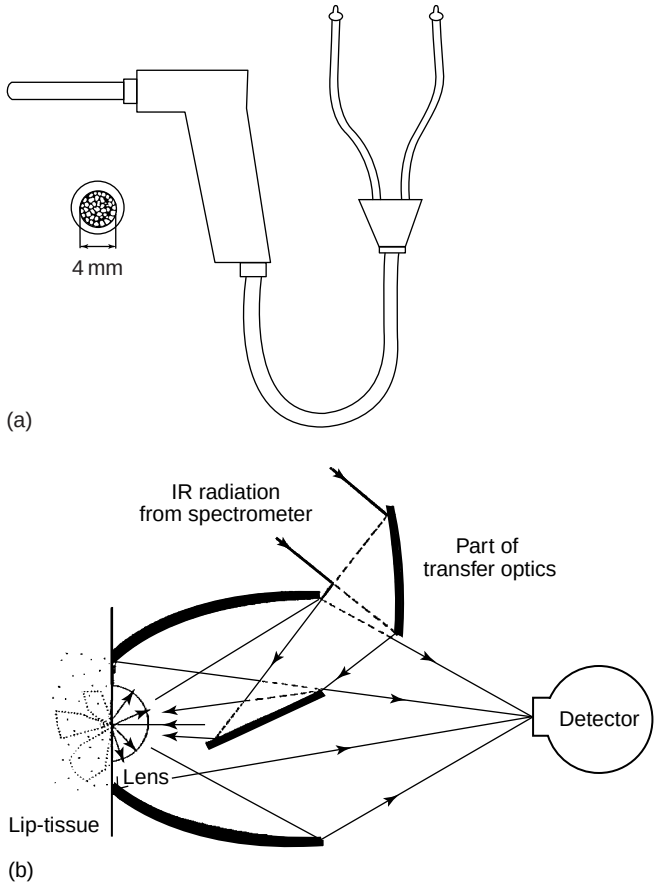

Figure 12 Scheme of different accessories for a diffuse reflectance measurement. (a) Bifurcated fiber-optic probe. (b) Accessory based on a rotational ellipsoidal mirror (reflection optics).

example concentric arrangements with or without a gap to select a particular optical pathlength and depth for tissue photons, are possible.[33]

The optical properties of the fiber core and its cladding are critical for the radiation throughput of the fiber-optic cable. The so-called numerical aperture (NA) represents its radiation-gathering capability. The difference in the refractive indices of core and cladding defines the largest angle (Θ = acceptance half angle) for a cone of meridional rays, at which total reflection at the interface of both still takes place (n is the refractive index of the medium containing the cone vertex)

$$\mathrm{NA} = n \sin \Theta = (n_{\mathrm{core}}^2 - n_{\mathrm{clad}}^2)^{1/2}$$

For conventional measurements in transmittance or diffuse reflectance, accessories are also based on reflection optics.[26,34] The arrangement favored by us consists of a rotational ellipsoidal mirror with one focus provided for the iluminated sample, whereas the second one is reserved for the detector (Figure 12b). The collection efficiency for diffusely reflected radiation is more than 50% for a Lambertian scatterer. The spectral signal-to-noise ratio has to be chosen according to the demands of sensing glucose absorption signatures. For diffuse reflectance measurements with glucose absorption bands around a wavelength of 1.6 μm, the noise level in the skin spectra for the configuration described should be lower than 10^{-5} AU (absorbance units) to tackle the normal physiological glucose concentration level.[35]

3.2.2 *Chemometrics for Quantitative Spectrum Evaluation*

For quantitative measurements the absorbance spectra of the sample are usually evaluated. Absorbance is the negative (decadic) logarithm of the transmittance spectrum (T), which provides the ratio of the transmitted against nonattenuated spectral radiant power. The absorbance values (A) are proportional to sample thickness and compound concentration ($A = \varepsilon cb$, where ε is the absorptivity, c is the concentration, and b is the sample pathlength of the absorbing substance), as stated by the Lambert–Beer law. Whilst transmission spectroscopy has been used as a standard method to solve analytical problems since the beginning of IR spectroscopy, other techniques such as measuring diffuse reflectance (R) have become routine for the analysis of, for example, pharmaceutical substances, or the in vivo determination of the degree of oxygen saturation of hemoglobin in arterial blood in skin or other tissues. For linearization, as shown empirically by several authors, the use of $-\log R$ values can provide a linear relationship between reflectance and concentration. A different, more rigorous approach is based on the theory derived by Kubelka and Munk (a description of this theory is given by Loyalka and Riggs[36]), which is usually considered for quantitative measurements carried out in the visible spectral range, as well as in the mid-IR.

Generally, linear modeling of the spectra is the basis for their quantitative evaluation. A multicomponent analysis requires at least the same number of wavelengths as components. The classical multivariate approach for evaluating a sample spectrum containing the contributions from several compounds is to model this with all component reference spectra, which can be pure compound spectra or, in the case of substance interactions leading to spectral changes, the component spectra can be derived from mixture spectra of similar composition to that of the sample to be analyzed. The component concentration values for the sample are estimated by least-squares for equation systems, overdetermined in wavelengths. Spectral residuals can assist in outlier detection or elucidating modeling deficits.

For assays of biosamples, it is nearly impossible to have quantitative information on all components contributing to the spectrum. However, regressing concentrations against spectral data, the so-called soft modeling approach, works well. For statistical calibrations, it is essential that the calibration data span the range of variations that can influence the spectra of future unknown samples. There are a number of different multivariate calibration algorithms available. Often, calibration systems are ill-conditioned due to linear dependencies in the spectral data (so-called collinearity problem). For that reason, factor methods such as partial least-squares (PLS) are applied.[37] A different approach, which is able to model nonlinearities explicitly, is the application of neural networks.[33,38] As in other multivariate methods, calibration by ''learning'' is essential, but such an approach tends to be more susceptible to overfitting, and more difficult to handle and interpret than the linear approaches discussed before.

Special spectrum preprocessing has been presented, for example, to reach a temperature insensitive algorithm for in vitro glucose assays by using adapted Fourier filtering. The advantage gained from filtering is that an optimal PLS calibration model can be built from a smaller number of factors due to the elimination of spectral variance during the preprocessing step, but with nearly unchanged performance. In practice, it has been shown that a preselection of spectral intervals can significantly enhance the performance of calibration models. A first choice can be based on inspection of the so-called property correlation spectrum with the compound response spectrum, which is necessary a priori knowledge obtained from the calibration data.[39] Correlations arising from other components can be manifested by calculating such a spectrum. This is

proportional to the first derived PLS factor spectrum, also defining its first regression vector, which is modified by considering further consecutive PLS factors as obtained in the iterative PLS factorization. Extension of the spectral interval limits can be carried out to find a minimum for the cross-validated prediction error.

For calibration testing, the data are usually divided into two subsets, the first for calibration modeling and the second for prediction validation with independent data. This approach has the advantage that it does not rely on any statistical assumptions derived from the calibration stage. Another strategy is to have a modeling subset of samples (size $n - m$) and a validation subset of size m, but putting the data after validation back for repeat calibrations and selecting another subset of samples previously not implemented in the validation stage, until all samples have been used in the validation process. The root mean square error of prediction (RMSEP) is calculated from the prediction residual sum of squares. When the validation subset size is 1, this is called the ''leave-one-out'' strategy, which is a reasonable approach when only a limited number of samples is available for model training:

$$\mathrm{RMSEP} = \left[\frac{\sum [c_i - c_{i,\mathrm{pred}}(R)]^2}{M}\right]^{1/2}$$

where c_i and $c_{i,\mathrm{pred}}$ are the sample reference concentration values and the independently predicted values, respectively; M is the total number of samples; and R is, for example, the number of PLS factors used for the estimation of regression variables. Usually, the calibration model with the lowest RMSEP value is considered for future application, thus avoiding model overfitting.

The number of wavelengths optimally required for spectroscopic calibration is still a subject of intense debate. Especially for NIR spectrometry with filter instruments, the optimal design has been frequently discussed. In fact, once one uses as many variables as there are independent spectral constituents, the addition of further wavelengths should serve to reduce effects from noise. On the other hand, as more wavelengths are used, the probability of encountering additional spectral interferences increases. This progression eventually leads to a situation where the use of more variables starts to degrade the accuracy of the result. Special selection strategies have been followed in multiple linear regression (MLR) methods, which generally rely on a reduced number of spectral variables, see for example Draper and Smith.[40] Step-wise regression is well known in statistical applications, and in a popular form, the forward or backward step-wise procedure, the algorithm calculates the linear regression between two sets of variables using least-squares, and selection is based on statistical confidence levels.

Recently, considerable effort has been made to evaluate procedures that identify spectral variables carrying useful information for the set-up of robust calibration models. A selection of papers dealing with these aspects, which consider search strategies such as genetic algorithms and simulated annealing or artificial neural network approaches, are given by McShane et al.[41] We developed a rapid and reliable variable selection for statistical calibrations based on PLS regression vector choices, which has been tested for various calibration scenarios.[42–44] As the optimum regression vector obtained from statistical calibrations contains the weights for the spectral variables needed for concentration prediction, spectral variables were chosen pairwise, providing the minima and neighbouring maxima of the regression vector, but in a ranking order decided by their coefficient weight size. The analytical performance of the calibration models, based on a consecutively increasing number of spectral variable pairs, is tested by cross-validation. The minimum RMSEP was taken for the selection of the optimum calibration model with a reduced number of variables under this scheme. For mid-IR studies with clinical calibrations for glucose in blood plasma or whole blood, only 10 variables were needed to reach the equivalent analytical performance obtained with broad spectrum evaluation.[42] In contrast, the NIR studies with the same clinical blood plasma population showed that more than 20 variables were needed to reach the clinical acceptance level.[43] The same procedure was successfully applied to data sets obtained within noninvasive assay development[44] which will be discussed later.

Problems with statistical calibrations can occur if correlations exist between two compounds within the set of calibration samples, because this covariance can result in systematic prediction errors when the calibration model is applied to samples outside the calibration sample population. Therefore, the design of the calibration experiments is of critical importance. Extremely important for statistical calibration testing is the model validation, i.e. checking the performance for prediction using calibration-independent samples. As mentioned already, the number of samples needed for calibration, in particular for assays on natural specimens, can be large in order to model the range of variance factors to be expected, although strategies have been developed recently to reduce the calibration workload, including synthetically prepared samples of less complex composition than found for the natural specimens.[45]

An appropriate experimental design is also necessary for sensitive in vivo calibrations when the analytical signals are comparable to the prevailing noise level or signal drifting. It is well known that the pickup of spurious drift

effects can significantly influence the results from multivariate statistical calibrations, especially when data were recorded under continuous monitoring conditions. In the past, our strategy for handling complex experiments was able to avoid temporal chance correlations. Surprisingly, positive results from calibrations were reached for single OGTTs using mid-IR spectra of the test person's inner lip, obtained with the attenuated total reflection technique.[24] However, by combining the calibration data from 2 days with several peaks and troughs in the blood glucose level, it became clear that the 1-day evaluations were dominated by spurious drift effects. If a priori knowledge on the compound signal features exists, this can be used for studying individual PLS factors and the regression vector shape to verify the physical basis of the calibration.

3.3 Tissue Spectroscopy

There are certain spectroscopic aspects to be discussed for the quantitative analysis of biomedical samples. Scattering in soft tissues is mainly due to discontinuities in the refractive index on the subcellular level. The mechanism of (elastic) radiation scattering from biological cells has recently been investigated by Mourant et al.,[46] who studied the scattering behavior of mammalian cells for wavelengths between 500 and 800 nm. Most of the scattering was caused by structures within the cells, but for larger wavelengths further investigations are needed.

Knowledge of the spatial distribution of photon-visit within tissue is important for quantitative diagnostics, since it defines the effective tissue volume which is probed by in vivo spectroscopy. For such a purpose, modeling of radiation transport in tissues is required. Estimation of the wavelength-dependent optical pathlength through tissue is required in order to quantify absolute analyte concentration values.[47] For biological tissues we find anisotropic scattering, which is highly forward-directed. This is the reason why simple theories such as that derived by Kubelka and Munk are not valid, so that more sophisticated modeling is required, although under special conditions such a theory may be applicable. A detailed description of the theory can be found in a paper by Loyalka and Riggs.[36]

The radiative transfer theory commonly preferred allows a treatment of radiation–tissue interaction, as only parameters of statistical relevance are used for optical tissue characterization; these are the average refractive index n, the absorption coefficient μ_a, the scattering coefficient μ_s (both in units of cm^{-1}), and the average scattering phase function $S(\Theta)$. The absorption coefficient μ_a equals the absorbance per unit length, $2.303\,\varepsilon_m C$, where ε_m is the molar absorptivity and C is the molar concentration. The reduced scattering coefficient is defined as $\mu_s' = (1 - g)\mu_s$, where the anisotropy parameter g is the mean cosine of the scattering angle Θ. More complex cases than found for isotropic media have been considered, for example, such as layered structures to model skin. A comprehensive summary of relevant terms in tissue optics was given by Patterson et al.[48] In order to explain the NIR spectra and for a better understanding of radiation transport, e.g. by means of maximum penetration depths and integral photon pathlengths, Monte Carlo simulations of the "photon random walk" through tissue can be carried out.[26,34] This refers to numerical calculations based on random sampling from appropriate probability distributions. With these tools, it is also possible to model the effects of physiological factors and different analytes for certain measurement scenarios, provided that their effects on the absorption and scattering properties of the tissue are known.[49]

The in vivo spectroscopy can be described from a practical point of view. In Figure 13, the diffuse reflectance spectra of lean muscle tissue are presented, recorded at different layer thicknesses by using a mirror optics based accessory. The intensities at lower wavenumbers are displayed, which are more favorable than found for a fiber-optic probe. The backscattered radiation intensity is low, so it is necessary to have extremely accurate and reproducible measurements, in addition to a high spectral signal-to-noise ratio. For SWNIR applications, also based on Monte Carlo simulations, a modified Lambert–Beer law can be expressed as $A = (\varepsilon cDb) + G$, where D is the differential pathlength factor due to photon diffusion, and G is an additive term due to scattering losses. This interpretation is valid for the reflectance data from Figure 13, when transformed into absorbance equivalent values. Additional losses from scattering, which are nearly

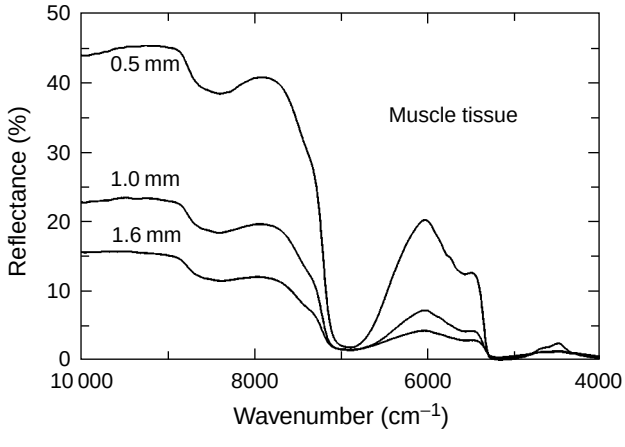

Figure 13 Reflectance spectra of different layers of lean muscle tissue backed by a brass plate (from bottom: layer thicknesses 1.6, 1.0, and 0.5 mm) and measured by means of the accessory based on reflection optics.

constant over a wide spectral range, are accompanied by an increase in sample layer depth.

3.3.1 Skin Structure and Chemistry

Skin is composed of three layers: the epidermis, the dermis and the subcutaneous-tissue layer, which includes adipose fat cells. It contains other elements such as hair and different glands and structures like blood vessels, lymphatics and nerves. Epidermal thickness can vary between 60 and 100 μm, but it is remarkably constant over the body, apart from special areas with a thick stratum corneum layer. The thickness of the dermis, which is connective tissue composed of fibrous proteins (predominantly collagen and elastin fibers) and proteoglycans, can vary tremendously, depending on the body location. The heterogenity of body tissues is illustrated in Figure 14, which shows a cross-section through the upper part of inner lip skin. The cornified, thin keratinized surface layer of normal skin epidermis, the stratum corneum, is missing here. A thorough review of the anatomy of the skin is beyond the scope of this article, but a few details relevant to in vivo spectroscopy are given.

The skin can be considered as the largest body organ, and it has different functions, such as providing a protective barrier. For a greater part, this is achieved by the stratum corneum, a horny layer which evolves from a dense population of actively dividing epidermal cells. For mucous tissues such as oral mucosa, the stratum corneum does not exist, which can be advantageous in diffuse reflectance measurements. Other skin functions include the regulation of body temperature and water content. A considerable amount of body water is accumulated in the skin, and a large proportion is found interstitially. The ability of skin to act as a reservoir is observed under various pathological conditions, with prominent uptake, for example, in kidney insufficiency and obesity, whereas dehydration is seen, for example, in diabetes insipidus. Surgical skin specimens of fat-free skin show a typical average water content of 72%. The epidermal water content is lower due to the horny layer. These figures vary because of substantial variations in water content over the surface and between different types of skin.

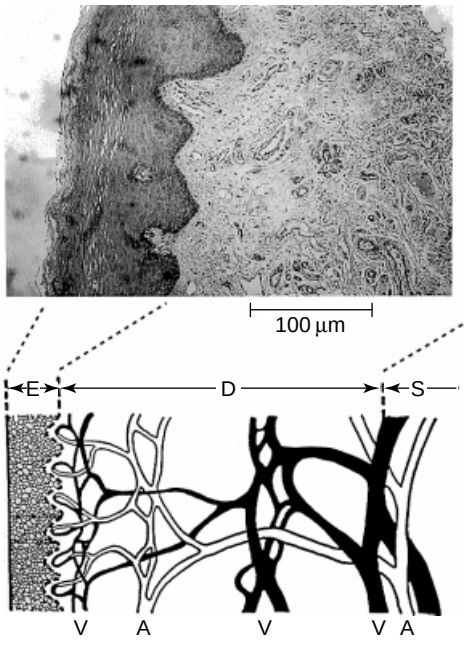

Figure 14 Microscopic cross-section of a tissue sample from the mucosal inner lip epidermis and schematics of the skin vasculature (E, epidermis; D, dermis; and S, subcutaneous tissue; V, venous and A, arterial blood vessels). (Reproduced in part by permission of Thieme-Stratton from R. Huch, A. Huch, D.W. Lübbers, *Transcutaneous PO_2*, Thieme-Stratton Inc., New York, 1981.)

In addition to collagen, hyaluronic acid is an important component. The actual location of water and the amounts associated with various components such as collagen and other proteins is not clear, but water has been seen as an integral component of collagen structure, being essential for maintaining the fibrillar collagen structure. Swelling of the polymer network by water is observable. The quantity and nature of proteins and mucopolysaccharides very profoundly affect the extent of water uptake. Variations in water content with age are well-known. Differences due to gender and diet have also been reported.

Apart from the structural variability, changes in vasodilation and vascular permeability, the concentration of hemoglobin in blood and its oxygenation status, as well as pigmentation from melanin, carotenoids, etc. influence the optical characteristics of cutaneous tissue, especially in the visible and SWNIR spectral ranges. As water is the main constituent in soft tissue, the NIR absorption spectrum is dominated by this compound and its variations. Additionally, scattering changes are dependent on the tissue water content, because the differences in the refractive index discontinuities at the subcellular level are affected.

The networks of blood vessels within the skin are shown schematically in Figure 14(b). Capillary loops extend almost to the surface of the skin. Vascular patterns in skin and their variations have been discussed recently.[3] By changing blood flow, thermoregulation can be controlled. There is certainly a temperature gradient along the skin depth, affected by the blood flow within the vasculature, which influences the spectroscopic variations seen during in vivo monitoring.[50]

Other variations are a result of skin irritation, photobiological processes and/or skin diseases, which also affect the structural and chemical characteristics. Exposure to a diversity of chemicals in soaps, moisturizers and

detergents certainly affects the skin and may even cause inflammation. These variations can also be followed by skin spectroscopy.

3.3.2 Skin Spectroscopy

The effects from different accessory geometries on skin spectra are discussed in this section. In Figure 11, spectra from the inner lip obtained with a fiber probe and a diffuse reflection (DR) mirror-based accessory are shown. For these cases, a reasonable match in refractive indices is given for the interface from tissue and the contacting optical material, which fortunately leads to negligible Fresnel reflection, which would otherwise cause unwanted perturbations if larger refractive index differences between the layers were present. This would be especially undesirable as this radiation fraction carries no analytical information. When we look at the spectral features below 6000 cm^{-1}, subtle differences can be detected. In the spectrum of skin from a man, recorded with the DR accessory, a broad single band attributed to water is seen. The top spectrum in Figure 11(b), which is an average spectrum from a large population of men and women, indicates a weak doublet, whereas the spectral features obtained with the fiber-optic accessory clearly indicate the probing of subcutaneous fatty tissue.

In Figure 15, the diffuse reflectance spectra of muscle and fatty tissue are shown, and these support this view. For the phospholipid lecithin, a special lipid representative, a spectrum is shown in Figure 5. For completeness, further lipid classes are mentioned: fatty acids and their derivatives, triglycerides and sterols. In Figure 16 the variability in lip spectroscopy is illustrated. In Figure 16(a), the spectral differences between consecutively measured

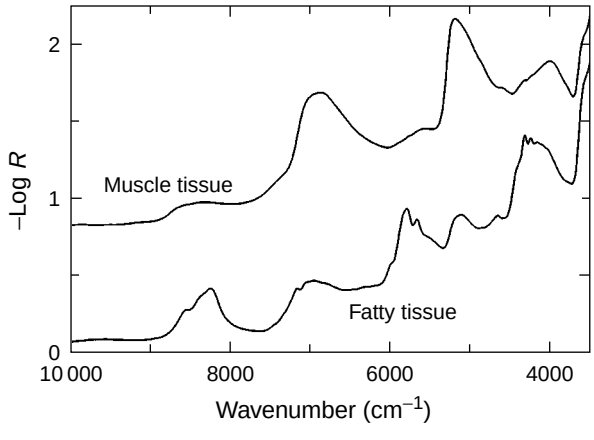

Figure 15 Comparison of diffuse reflectance NIR spectra of lean muscle and fatty tissue measured by means of the accessory based on reflection optics (the latter spectrum is offset for clarity).

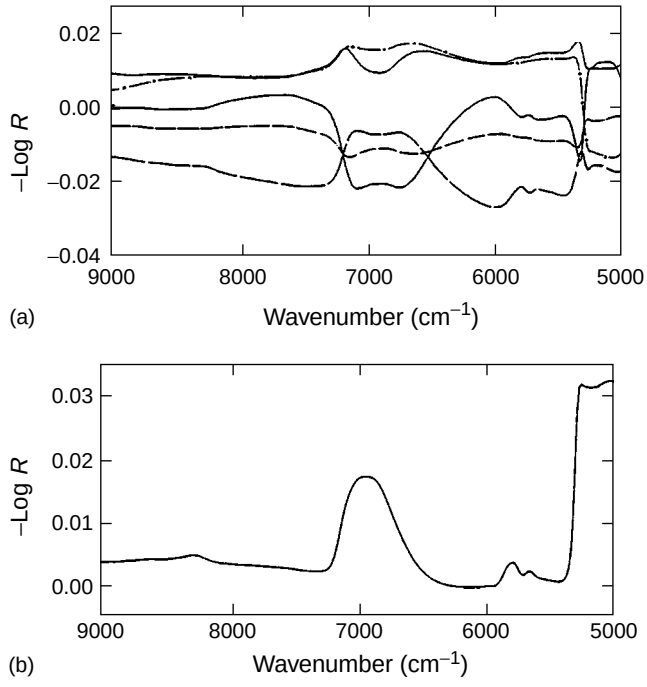

Figure 16 Difference spectra illustrating the variability of lip tissue measurements within the spectral range considered for blood glucose prediction. (a) Spectral variations calculated from the difference of two subsequent lip measurements under constant blood glucose. (b) Difference spectrum between the mean spectra calculated from lips of 182 female and 199 male volunteers.

lip spectra of a man are shown, also indicating changes in the tissue water. Figure 16(b) gives the difference spectrum between the average inner lip spectra obtained from women and men, respectively. In addition to the difference in water content, a larger contribution to the spectrum from fatty tissue of the oral mucosa in women can be seen.

The effect of different contact pressures on the lip spectrum is illustrated in Figure 17. The spectra were again recorded by using the DR accessory, and by applying low and heavy pressure during measurement. The resulting difference spectrum shows spectral features similar to a cuvette spectrum of blood plasma.[19]

In Figure 18 the influence of the stratum corneum on the diffuse reflectance skin spectra is shown. In Figure 18(a), the spectra of a finger tip, recorded by a fiber probe and a mirror-based DR accessory (see Figure 12), are given. Whereas the scattering top layer has little effect on the skin spectra (Figure 18b), there is a tremendous difference seen compared with the oral mucosa spectrum (cf. Figure 17). With a thicker stratum corneum layer, as seen for thumb skin compared with that of the outer lip, the spectral features of keratin become evident below 6000 cm^{-1}. These effects are even more intense for spectra of the heel or thumb tissue when measured through the

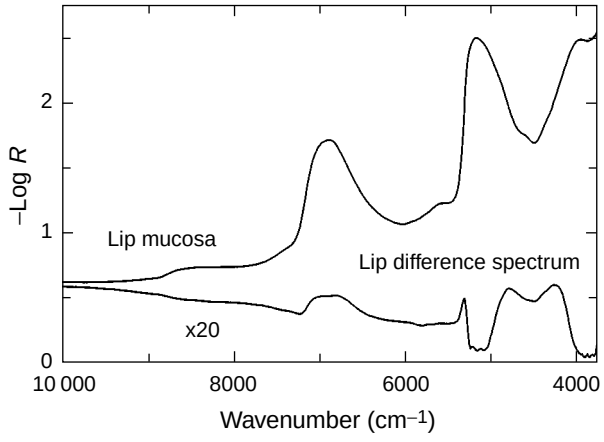

Figure 17 Diffuse reflectance spectrum of lip mucosa recorded by means of the accessory based on reflection optics and difference lip spectrum obtained by application of low and high contact pressure of the probe (lower trace, enlarged).

finger nail. For completeness the keratin spectrum of hair is also shown in Figure 18(c).

The variability within skin tissue spectroscopy is summarized in Figure 19. Figure 19(a) shows the spectral standard deviation, as obtained by using the DR accessory with repeated repositioning of the probe to the inner lip, for a multiperson and a single-person experiment (the corresponding mean lip spectra are shown in Figure 11b), whereas in Figure 19(b) the standard deviation data are presented for a fiber-optic probe, measuring inner lip and thumb spectra with and without repositioning of the probe (for the corresponding tissue spectra, see Figure 18b). There are significant reductions in spectral variance with a fixed spectroscopic probe.

3.4 In Vivo Assay Using Diffuse Reflectance Spectroscopy of Lip Mucosa

The information content of different NIR spectral ranges versus spectral perturbations (noise, component spectrum

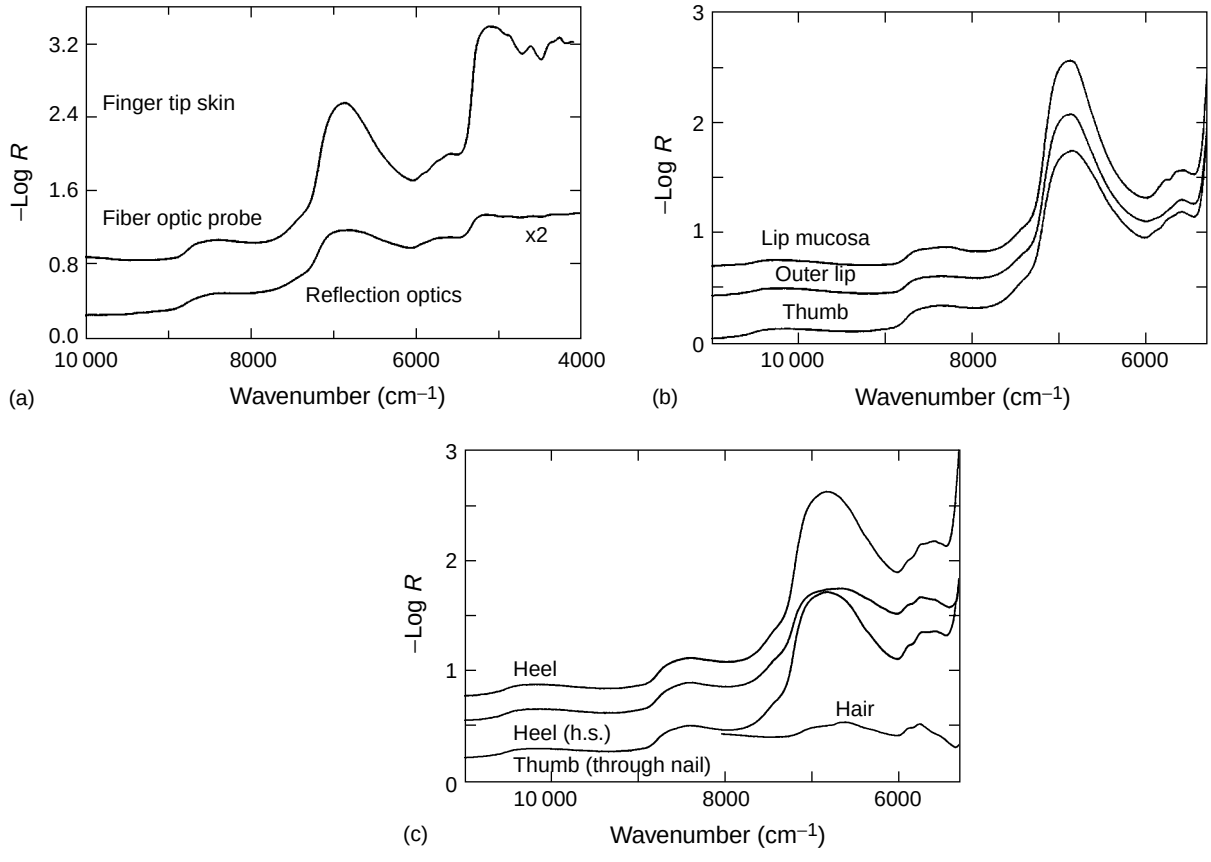

Figure 18 In vivo reflectance spectra of different skin tissues. (a) Comparison of single-person spectra of finger tip skin as measured with a bifurcated fiber-optic probe and an ellipsoidal mirror-based accessory. (b) Spectra measured from lip mucosa, outer lip, and thumb skin using a bifurcated fiber-optic probe. (c) Spectra showing significant contributions from keratin: heel with a thick layer of stratum corneum, as for the trace above, but highly scattering (h.s.); thumb measured through finger nail; and hair. (Reproduced in part by permission of Springer-Verlag from A. Bittner, S. Thomaßen, H.M. Heise, 'In Vivo Measurements of Skin Tissue by Near-infrared Diffuse Reflectance Spectroscopy', *Mikrochim. Acta*, **14**(Suppl.), 429–432 (1997).)

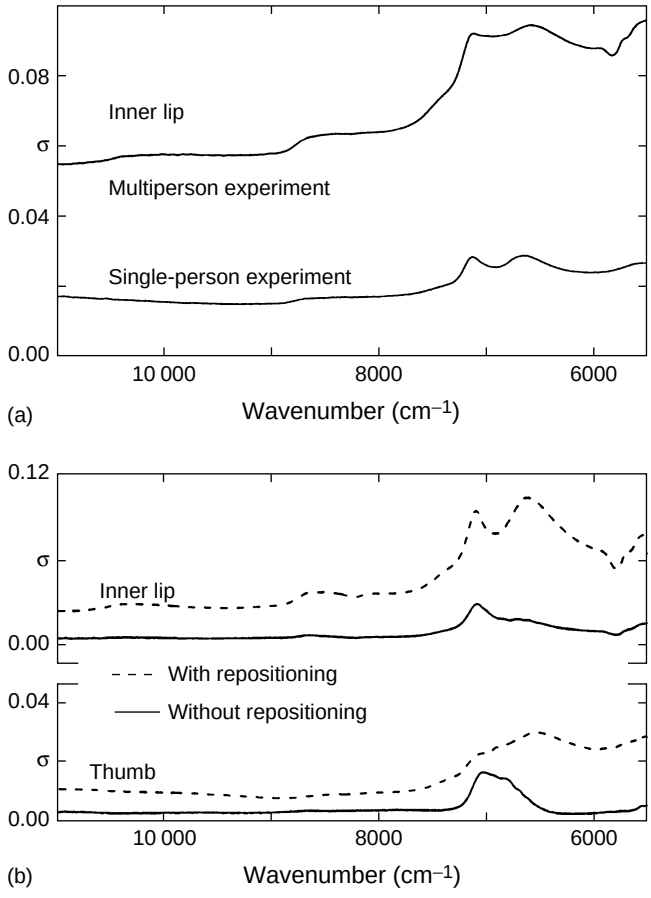

Figure 19 Reproducibility in the measurement of NIR skin tissue spectra. (a) Standard deviations of inner lip spectra obtained within a multiperson experiment (133 persons, 390 spectra) and from a single-person 2-week experiment (219 spectra) as measured by a reflection optics-based accessory. (b) Standard deviations of inner lip and thumb spectra, respectively, as measured with and without repositioning of a fiber-optic probe. (Reproduced in part by permission of Springer-Verlag from A. Bittner, S. Thomaßen, H.M. Heise, 'In Vivo Measurements of Skin Tissue by Near-infrared Diffuse Reflectance Spectroscopy', *Mikrochim. Acta*, **14**(Suppl.), 429–432 (1997).)

overlap) with respect to the assay performance for glucose has been investigated recently.[44,50] For such a purpose, a large population of human plasma samples from a hospital was analyzed by clinical enzymatic methods. The spectra were recorded by using either a 10-mm or 1-mm quartz transmission cell. Respective spectral data for calibration ranged from 11 000 to 7600 cm^{-1} (SWNIR) or from intervals below 6800 cm^{-1}. By using the optimum calibration model, the NIR spectra recorded with the 1-mm cell produced a relative standard deviation of 7.8% (based on the mean glucose concentration of the samples), whereas significantly worse performance was achieved using SWNIR data with a respective value of 23%. Handicaps of the SWNIR region are the weaker absorption coefficients of the plasma components and the much broader absorption bands, leading to a poorer selectivity for the calibration compared to other spectral ranges. For the long-wave NIR with data below 6800 cm^{-1}, exclusion of spectral intervals with the largest variances from water was necessary to reach the calibration results mentioned above.

Generally, the measurement conditions for the in vitro analysis of aqueous biofluids can be kept much more constant with regard to temperature, sample homogeneity and optical pathlength than for in vivo measurements, where many physiological variables are to be considered for skin tissue: peripheral blood perfusion with shunting or other circulatory variations, an inhomogeneous distribution superimposed on an unknown coarse glucose gradient, and a temperature profile with additional fluctuations in mean body temperature. In addition, the thickness and anatomic structure of the skin can vary tremendously. Despite these variations, the spectrometric assay performance for in vitro samples allows some extrapolation for in vivo measurements.

3.4.1 Integral Tissue Measurements

The most promising results have been obtained for individual patients where the measurement conditions have been investigated in greater detail. In vivo experiments can be performed on a subject, e.g. under fasting conditions, oral glucose or meal tolerance tests or glucose clamp experiments with constant blood glucose levels. Parallel to this, reliable reference data for the calibration must also be obtained from invasive blood tests, either by puncture of the skin or by using, for example, syringes for taking venous blood. It is advisable that the data gathering on the time-dependent blood glucose profile are dense enough to allow data interpolation to reconstruct the temporal dependency needed for mean tissue glucose modeling. There is certainly a great deficit for modeling when only single blood glucose measurements with large time gaps are available, which do not permit any estimates on mean tissue glucose.

As discussed above, the blood fraction in skin is quite variable, so skin hyperemization is advisable to reach optimum conditions with a high blood content and largest blood flow for reducing gradients in the vascular system. The blood volume could be monitored by diffuse reflectance spectroscopy in the visible spectral range by determining the hemoglobin concentration. When oral mucosa is the preferred tissue for in vivo monitoring, cleaning of the oral cavity is important in order to avoid spurious results caused by the presence of food residues.

It is a prerequisite for a working calibration that the concentration variance in the calibration data set is large enough to allow for the calculation of satisfying

calibration models. The signal-to-noise ratio must be sufficiently high to enable extraction of meaningful information from the spectral data.

Various calibration experiments with an individual person, as well as with 133 patients, were carried out by us (see also Heise et al.[44] and references cited therein). NIR spectra of the human inner lip were measured using the specially optimized accessory for diffuse reflectance measurements, attached to an FT spectrometer, to allow skin spectra with a noise level of about 10^{-5} AU.[26] Measurement time was about 1 min to accumulate 1200 single-sided interferograms equivalent to a spectral resolution of 32 cm^{-1} after Fourier transformation. This is the maximum measurement time usually tolerated by patients. PLS multivariate calibration with spectral variable selection was applied for regression of the spectral data between 9000 and 5500 cm^{-1}. In Figure 20 the design of two different calibration experiments is sketched with data gathering over 2 days and over 2 weeks, respectively. For compressed presentation the time gaps between the different daily sessions were deliberately reduced.

In the 2-day calibration experiment, a nonstandard OGTT was used and a time delay of the glucose concentration profile of 10 min for calculating the reference concentration data was applied (optimized by variation testing). The optimum calibration models were obtained with ''leave-one-out'' cross-validation. In Figure 21(a) the RMSEP results from different PLS models based on 26 selected spectral variables are shown, and an optimum calibration model from 132 calibration samples, logarithmized lip reflectance spectra with corresponding reference blood glucose concentrations, is obtained with 16 PLS factors. On the other hand, a least-squares solution of the regression coefficients gives about the same performance. The calibration model can predict glucose concentrations with an RMSEP of 36 mg dL^{-1}. The robustness of the calibration model is also underlined by the fact that cross-validation with packages of five standards shows only an insignificant deterioration compared to the ''leave-one-out'' strategy. The prediction results are plotted in a scatter diagram for acceptance testing. Routinely, the Clarke error grid has been used for data presentation,[51] but we prefer a straightforward presentation similar to that proposed by Koschinsky et al.[52] Such a presentation using logarithmic scales is given in Figure 21(b), where the confidence interval for a 30% relative standard deviation is marked by dashed lines (for self-monitoring devices a value of 15% has been claimed, but this is not always reached).

The 2-week experiment led to an optimum RMSEP value of 47 mg dL^{-1}, which is significantly higher than the results from the 2-day calibration data due to the larger variability in the spectral data compared to the 2-day

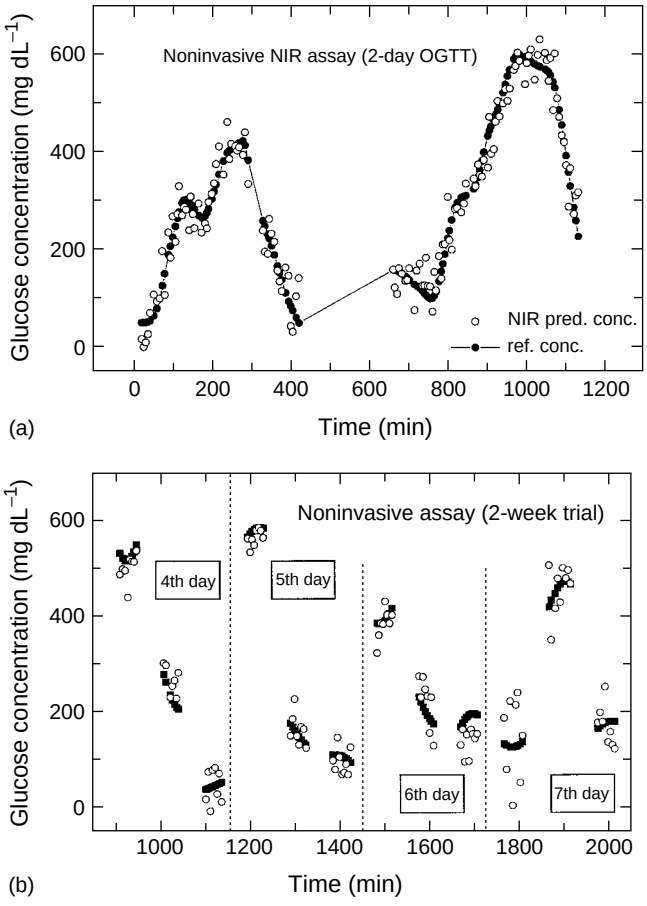

Figure 20 Comparison of capillary blood glucose concentration values (calibration reference data) and prediction results based on PLS calibration using diffuse reflectance lip spectra obtained in experiments with a single diabetic person. (a) Results for the 2-day nonstandard OGTT. (b) Representative section of results obtained for 4 days within a 2-week testing experiment. Three sessions with different blood glucose levels were chosen during the whole day (for compressed presentation the time gap between the sessions is deliberately reduced). (Reproduced by permission of Walter De Gruyter from H.M. Heise, A. Bittner, R. Marbach, 'Near-infrared Reflectance Spectroscopy for Non-invasive Monitoring of Metabolites', *Clin. Chem. Lab. Med.*, **38**, 137–145 (2000).)

test. However, such a testing scenario is more realistic for diabetic patients performing everyday blood glucose monitoring. The performance is illustrated for a sequence of days in Figure 20(b). There are some situations where the scatter of prediction results is unacceptably large. Further investigations are necessary to study the major factors affecting the prediction performance of such a spectroscopic assay. In this context, perturbations from other compounds, such as from medications, must also be investigated in the future.

A better understanding of NIR in vivo spectroscopy of integral body tissue, in particular the skin, is still

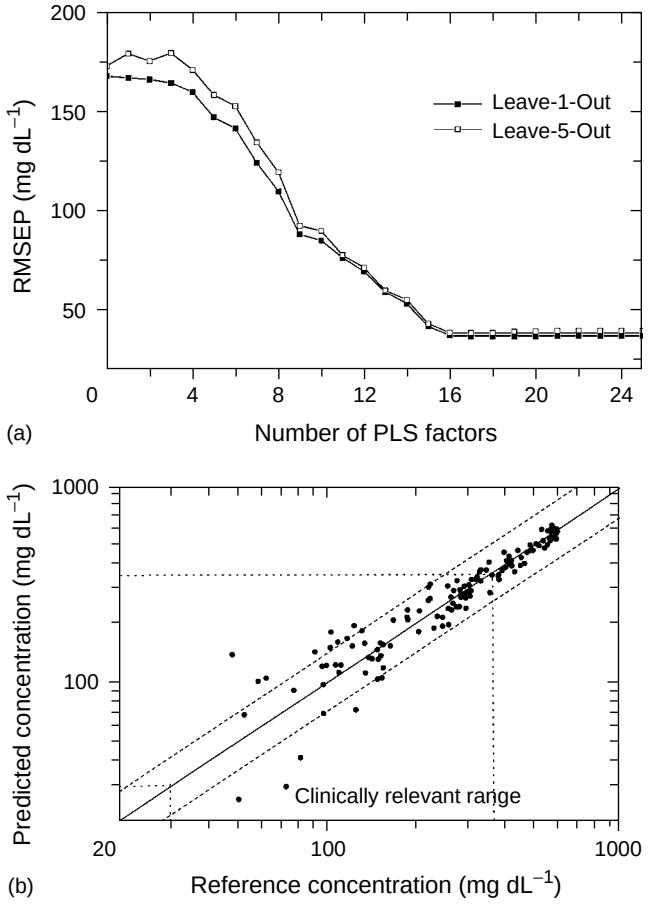

Figure 21 Calibration results for blood glucose using 132 diffuse reflectance lip spectra obtained from a single-person experiment with nonstandard OGTT (see also Figure 20a). (a) RMSEP from calibration models based on 26 specially selected spectral variables versus number of PLS factors. (b) Predicted concentrations versus reference values based on the optimum calibration calculated from 16 PLS factors (the confidence interval for a 30% relative standard deviation is marked by dashed lines). (Reproduced in part by permission of Walter De Gruyter from H.M. Heise, A. Bittner, R. Marbach, 'Near-infrared Reflectance Spectroscopy for Non-invasive Monitoring of Metabolites', *Clin. Chem. Lab. Med.*, **38**, 137–145 (2000).)

needed to learn more about biosystem complexity, but the same applies to other body samples, such as the aqueous humor. It also means that standardization of the measurement is vital, but an even more sophisticated data reduction scheme, including outlier detection, will be necessary rather than the simple calibration approaches chosen so far. With the increasing capabilities of NIR spectrometers, and computers handling more data with refined algorithms, the time when a noninvasive approach for metabolite determination could be utilized for the patient at home or in the hospital is not out of reach.

3.4.2 Pulsatile Spectroscopy

Integral tissue probing faces many unknown physiological variables. Further development of noninvasive assay technology can be based on fast subsecond NIR spectroscopic measurements, allowing the probing of parts of the intravascular fluid space, since its arterial volume is modulated by the heart beat. It is the situation as found for pulse oximetry, by which the degree of $HbSO_2$ can be determined.[53,54] Spectrum changes can be monitored synchronously to the cardiac cycle. During systole the capillary blood volume is increased, which leads to a higher absorption of IR radiation due to water, hemoglobin and other blood constituents; the situation is opposite during diastole. However, since the arterial blood constitutes a small and unknown fraction of the tissue, the alternating signal is only a minor fraction compared to the total tissue water, which also varies with the physiological state of the microvasculature and the tissue texture. Owing to limitations in signal-to-noise ratio, such a measurement principle, although proposed in patents, has not yet been applied in practice for metabolite measurements.

Preliminary results of time-resolved measurements on human oral mucosa were recently presented using diffuse reflectance spectroscopy.[44] The first individual lip spectra obtained with fast measurements are shown in Figure 22(a). These are difference spectra smoothed for noise reduction, as calculated versus the first measured spectrum of the data set, after application of a polynomial baseline fitted to predefined spectral intervals which are located in the spectrum minima. It clearly illustrates the intensity fluctuations caused by changes in the arterial blood compartment associated with the cardiac cycle.

Each Fourier analysis of the time-dependent logarithmized intensities, assigned to the individual spectral variables, provided the spectral Fourier amplitudes for each frequency component. Considering a three-dimensional diagram with spectral wavenumber variables, frequency components in hertz and pulsatile amplitudes as ordinates, the pulsatile spectrum evolves along the heart beat frequency. This has no static component, originating from the aqueous unmodulated compartments of the skin tissue. The pulsatile spectrum, showing only component changes with pulse, is a type of difference spectrum (see Figure 22b), and is similar to the water absorbance spectrum recorded with a transmission cell of 0.5-mm pathlength. The water absorbance alterations due to the cardiac blood pressure changes are about 20 mAU for the water band at 6900 cm^{-1}. This is equivalent to a water layer of 15-μm thickness, which is about a factor of 50 smaller than obtained for the integrating measurements discussed above. It is noteworthy that the ratio of the maximum amplitudes of the water combination band at 5200 cm^{-1} and of the overtone band at 6900 cm^{-1} is much smaller for the pulsatile spectrum than for the

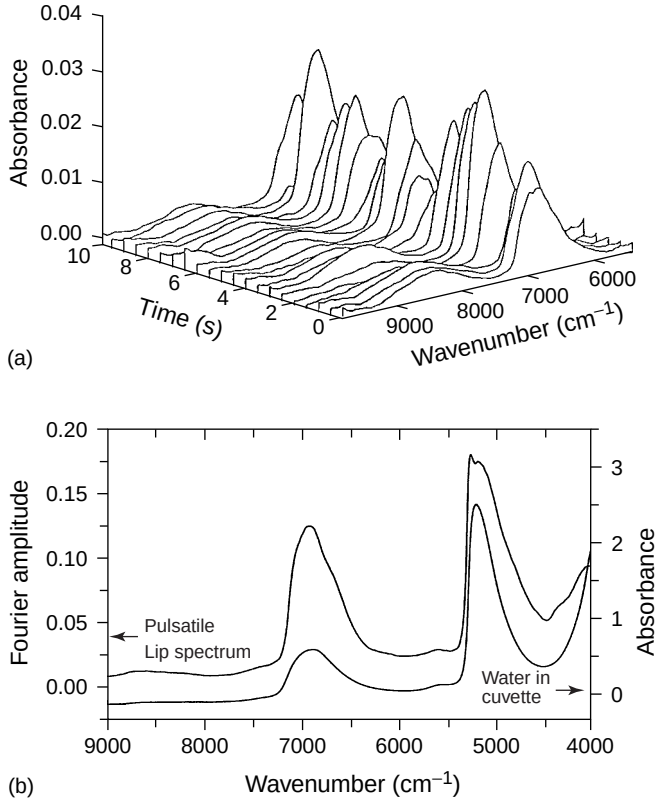

Figure 22 (a) Fast measurements of inner lip spectra of a single person shown as differences versus the first measured lip spectrum after spectral smoothing. (b) Fourier amplitudes (upper trace) illustrate the relative pulsatile spectral components in the near-IR diffuse reflectance spectra of human oral mucosa due to cardiac-modulated blood volume variations (for each spectral variable the Fourier amplitude coefficients were averaged around the heart beat frequency within a frequency interval of 0.06 Hz); for comparison the absorbance spectrum of water as measured in a 0.5-mm cell is given below (lower trace, right ordinate scale). (Reproduced by permission of Walter De Gruyter from H.M. Heise, A. Bittner, R. Marbach, 'Near-infrared Reflectance Spectroscopy for Non-invasive Monitoring of Metabolites', *Clin. Chem. Lab. Med.*, **38**, 137–145 (2000).)

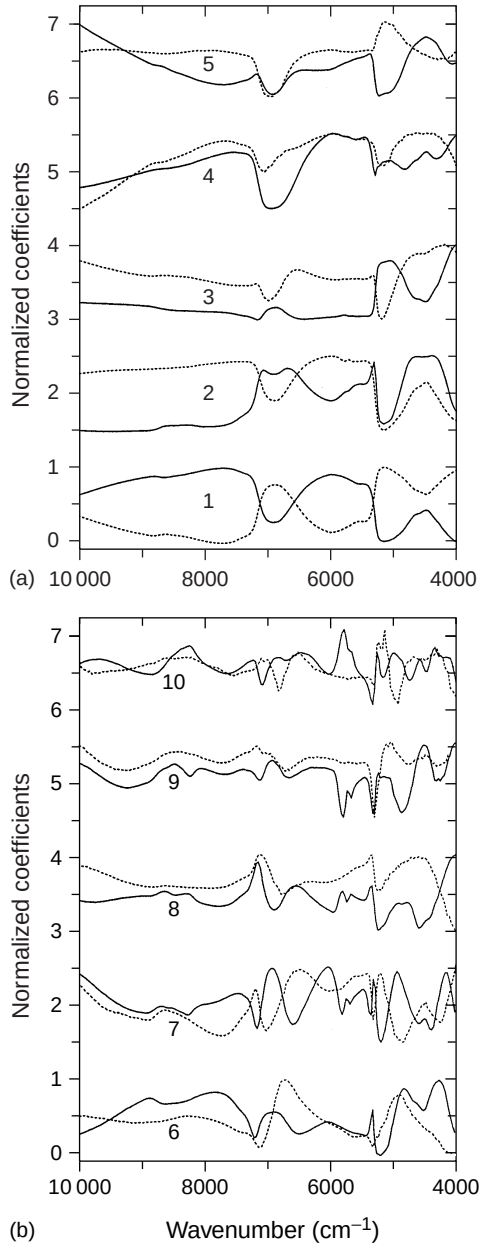

Figure 23 Factor spectra from a PCA of the logarithmized inner lip spectra of a single person recorded during a 2-week testing experiment (solid lines) and of logarithmized inner lip spectra from a different person obtained during 1 min (sampling frequency: two spectra recorded per second; dashed lines). (a) The minimum–maximum normalized loading spectra of the first five principal components. (b) The additional five factor spectra related to further decreasing singular values of the original calibration spectra matrix.

water absorbance spectrum recorded in transmission. This can be explained by the significantly different penetration depths for the NIR radiation obtained for those wavelengths.

The complexity of lip spectra can best be evaluated by using a principal component analysis (PCA). For this, the matrix of calibration spectra is decomposed to give factor spectra which provide a system of orthogonal basis vectors to span the spectral space from the calibration data. The vectors can be derived from a singular value decomposition of the matrix, defined by maximizing the variance subsequently in the remaining subspace. Such an analysis was carried out for the individual lip spectra measured within 350 ms and recorded for the lapse of 2 min, as well as for the lip spectra obtained within the one-person, 2-week, calibration experiment. The first ten factor spectra, normalized to the same minimum–maximum distance of their vector components, are presented in Figure 23. As the first

five vectors (without regard to their sign) show great similarity due to the dominating variance contributions from tissue water, the other factor spectra from the 1-min integrating inner lip measurements possibly demonstrate that different sampling tissue sites due to repositioning over 2 weeks increase the complexity of the data. As a consequence, calibration modeling with a greater number of factors is required for efficient glucose assays using this approach than would be necessary using pulse spectrometry. Additional complexity is given for any of the optical approaches probing a certain skin volume integrally, because it is important to know the relationship between glucose concentrations in blood and interstitial and intracellular compartments.

4 REVIEW ON CURRENT ALTERNATIVE TECHNOLOGIES

4.1 Fluorescence Spectroscopy

In the past, fluorescence had been investigated as an implantable glucose sensing approach.[55] A similar strategy was followed by Lakowicz and Maliwal who investigated lifetime-based sensing using fluorescence resonance energy transfer based on the reversible association of concanavalin A with dextran.[56] A slightly variant method was previously presented based on the luminescence decay time of a long lifetime metal–ligand complex.[57] An attempt has been made to develop a transdermal methodology using sensing patches implanted below the skin which show a fluorescence change correlated to the tissue glucose concentration (''quasi-noninvasive approach'').[58] A NIR assay based on long-wavelength dyes has recently been evaluated, although this technological advance has not yet been applied to physiological fluids.[59]

4.2 Polarimetry

The monosaccharide D-glucose exists in two stereoisomers (α-D-glucose and β-D-glucose). Owing to their chirality, these substances are optically active, i.e. they give rise to an opposite optical rotation of linearly polarized electromagnetic radiation. In aqueous solution, this effect, which is proportional to concentration and sample pathlength, can be conveniently measured, because the concentrations of the α- and the β-form of glucose are not equally balanced. Quantitative measurements have been carried out by polarimetry for a long time, since optical activity was first observed in the beginning of the last century by Arago, Biot and Seebeck. However, the interaction of physiological glucose with electromagnetic radiation allows the determination of such low concentrations by using microdegree polarimetry.

This technique was successfully applied for cell culture media.[60] As the skin shows significant radiation scattering, such a measurement is extremely difficult due to depolarization of the probing light. Therefore, this methodology has been suggested for measuring the glucose concentration of the aqueous humor within the anterior chamber of the eye. Glucose can be measured when a beam of polarized light passes laterally through the eye. The polarimeter used a Faraday modulator to generate the polarization modulation of the incident laser beam before it reached the sample. A digital closed-loop approach using a Faraday compensator and an analyzer for nulling the optical rotation due to glucose increased the robustness of the measurement apparatus. A recent paper gives results for a multispectral in vitro method for compensating effects from albumin and preliminary single wavelength in vivo results supporting the use of such a technique to be employed for a noninvasive glucose assay.[61] A device for coupling light through the anterior chamber has also been presented. Chou et al. used an optical heterodyne polarimetric approach for in vitro samples and illustrated its use for glucose detection by using a rabbit's eye.[62] Problems reported so far concern corneal birefringence and eye motion. There was a 30-min time delay between the concentration profiles of the humor and blood glucose. Another topic of interest is the patient's acceptance of such an eye measurement device, apart from the fact that such a time delay results in a loss in glucose regulation within a closed-loop feedback system.

4.3 Scattering Approaches

Apart from absorption measurements, a different approach was evaluated by several groups. There are four different effects from dissolving glucose in an aqueous scattering biosystem: two effects influence the absorption coefficient due to an increase in intrinsic glucose absorption and to water displacement, and the latter is not selective enough for in vivo monitoring. Two further effects are noticeable: changes in the refractive index and the scattering coefficient, which has been mentioned earlier. Utilization of the latter effect led to the development of portable instrumentation with two optical sensor heads, each containing four LEDs of different wavelengths in the visible and SWNIR range, and six photodetectors were located at different millimeter distances. The monitoring of tissue glucose concentration was carried out during several glucose clamp experiments.[63] The general variability, i.e. glucose-independent changes in scattering of skin, was also investigated. Specificity is the greatest concern, because many other effects unrelated to glucose can influence the scattering signal. Some significant influencing factors have been discussed in section 3.3.1. On the

other hand, the instrumentation is attractive as it is simple and affordable, and allows continuous monitoring.

4.4 Raman Spectroscopy

A complementary technique to IR is Raman spectroscopy: a transition between vibrational energy levels is also induced, but this involved radiation scattering by the molecules. Unlike IR spectroscopy, however, monochromatic laser excitation is used. One great advantage is that the Raman spectrum of water is weak, which is certainly in contrast to mid-IR spectroscopy. Much effort has recently been placed on a multicomponent assay using NIR Raman spectroscopy,[64] emphasizing the splendid potential of vibrational spectroscopy for its suitability in clinical chemistry assays due to its selectivity based on fingerprint-like fundamental vibration spectra. In contrast to skin, analysis of components in the aqueous ocular humor may well be achieved by using Raman spectroscopy, although laser excitation power is a primary concern. For avoiding a large background fluorescence, NIR laser excitation is usually considered for the Raman spectrum recording,[65] although a tremendous tradeoff is seen due to longer wavelength excitation, compared to excitation with lasers in the visible range. Stimulated Raman spectroscopy using a pump and a probe laser for the same biofluid has also been proposed, and theoretical studies on the sensitivity have been presented.[66] More details on Raman spectroscopic approaches can be found in a review on noninvasive glucose measurements.[20] So far, in vivo experiments still present many problems.

CONCLUDING REMARKS

There are further spectroscopic techniques, the potential of which has still to be assessed. Photoacoustic spectroscopy within the NIR range has been mentioned in section 3.1. Another challenging technique is emission spectroscopy in the mid-IR produced by utilizing temperature gradients in the skin, and this is currently under intensive testing.[67] American companies such as Biocontrol Technology Inc. or Futrex Inc., using NIR absorption spectroscopy, have frequently been in the headlines, but their photonics technology has still to achieve levels needed to meet the expectations in noninvasive blood glucose testing. Their technology has yet to demonstrate the reliability required by the United States Federal Drug Agency (USFDA) to market such devices in the United States. The fluctuations in the industrial diagnostic instrument area are large, where many different systems are currently under design and testing, driven by the plethora of patents issued in this field of noninvasive testing. A snapshot of current company activities can be found in a recent article by Diller,[68] which is incomplete, as her research did not cover the European and Japanese activities!

One necessary trend is that the analytical instruments have to move closer to the patient, which certainly culminates in the development of miniature noninvasive measurement devices for blood glucose self-monitoring. The optical approach, in particular in vivo absorption NIR spectroscopy using diffuse reflectance of skin tissue, is promising for the future. In the meantime, implanted biosensors or microdialysis probes, in combination with glucose biosensors, are competing techniques, but face different problems within in vivo assays. These can be avoided by optical techniques, ranked highly owing to their noninvasive analytical methodology.

ACKNOWLEDGMENTS

The author is indebted to Mrs M. Hillig, Dr R. Marbach, Dipl.-Ing. S. Thomaßen and Dr A. Bittner for fine collaboration in the past. Professor Dr med. Th. Koschinsky, Professor Dr med. H. Reinauer and Dr med. C. Niederau from the Diabetes-Forschungsinstituts an der Heinrich Heine-Universität Düsseldorf are thanked for providing support for in vivo experiments and clinical reference data. Professor Dr med. D.W. Lübbers from the Max Planck-Institut für Molekulare Physiologie, Dortmund is thanked for continued discussions in the field of skin physiology. Financial support by the Deutsche Forschungsgemeinschaft, the Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen and the Bundesministerium für Bildung und Forschung is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

AGE	Advanced Glycation Endproducts
AU	Absorbance Units
CCD	Charge-coupled Devices
DR	Diffuse Reflection
FT	Fourier Transform
G-DH	Glucose Dehydrogenase
GOD	Glucose Oxidase
G-6-P-DH	Glucose-6-phosphate Dehydrogenase
$HbSO_2$	Hemoglobin Oxygenation Saturation
HK	Hexokinase
IDDM	Insulin-dependent Diabetes Mellitus
IR	Infrared
LED	Light-emitting Diode
MLR	Multiple Linear Regression
NA	Numerical Aperture

NIDDM	Noninsulin-dependent Diabetes Mellitus
NIR	Near-infrared
OGTT	Oral Glucose Tolerance Test
PCA	Principal Component Analysis
PLS	Partial Least-squares
RMSEP	Root Mean Square Error of Prediction
SWNIR	Short-wave Near-infrared
USFDA	United States Federal Drug Agency
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Fluorescence Spectroscopy In Vivo • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Monosaccharides and Sugar Alcohol Analysis

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Glucose Measurement • Infrared Spectroscopy in Clinical Chemistry • Point-of-care Testing • Urinalysis and Other Bodily Fluids

Food **(Volume 5)**
Enzyme Analysis and Bioassays in Food Analysis

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Multivariate Calibration of Analytical Data • Soft Modeling of Analytical Data

Electronic Absorption and Luminescence **(Volume 12)**
Near-infrared Absorption/Luminescence Measurements

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Spectral Data, Modern Classification Methods for • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. The Diabetes Control and Complications Trial Research Group, 'The Effect of Intensive Treatment of Diabetes on the Development of Diabetes on the Development and Progression of Long-term Complications in Insulin-dependent Diabetes Mellitus', *N. Engl. J. Med.*, **329**, 977–986 (1993).
2. UK Prospective Diabetes Study Group, 'Intensive Blood-glucose Control with Sulphonylureas or Insulin Compared with Conventional Treatment and Risk of Complications in Patients with Type 2 Diabetes (UKPDS 33)', *Lancet*, **352**, 837–853 (1998).
3. W.F. Ganong, *Review of Medical Physiology*, 17th edition, Appleton & Lange, Norwalk, 1995.
4. J.T. Sorensen, C.K. Colton, R.S. Hillman, J.S. Soeldner, 'Use of a Physiologic Pharmacokinetic Model of Glucose Homeostasis for Assessment of Performance Requirements for Improved Insulin Therapies', *Diabetes Care*, **5**(3), 148–157 (1982).
5. C. Cobelli, A. Mari, 'Validation of Mathematical Models of Complex Endocrine-metabolic Systems, a Case Study on a Model of Glucose Regulation', *Med. Biol. Eng. Comp.*, **21**, 390–399 (1983).
6. C. Lentner (ed.), 'Wissenschaftliche Tabellen Geigy', *Teilband Phys. Chem.*, 8th edition, Blut, Humangenetik, Ciba-Geigy, Basel, 66–195, 1979.
7. H. Keller, *Klinisch-chemische Labordiagnostik*, Georg Thieme, Stuttgart, 1991.
8. L. Caspary, J. Thum, A. Creutzig, D.W. Lübbers, K. Alexander, 'Quantitative Reflection Spectrophotometry: Spatial and Temporal Variation of Hb Oxygenation in Human Skin', *Int. J. Microcirc.*, **15**, 131–136 (1995).
9. J.N. Roe, B.R. Smoller, 'Bloodless Glucose Measurements', *Crit. Rev. Therapeutic Carrier Syst.*, **15**, 199–241 (1998).
10. G. Thews, 'Diffusion und Permeation', in *D-Glucose und verwandte Verbindungen in Medizin und Biologie*, eds. H. Barthelheimer, W. Heyde, W. Thorn, Ferdinand Enke Verlag, Stuttgart, 250–276, 1966.
11. R. Marbach, Th. Koschinsky, F.A. Gries, H.M. Heise, 'Noninvasive Blood Glucose Assay by Near-infrared Diffuse Reflectance Spectroscopy of the Human Inner Lip', *Appl. Spectrosc.*, **47**, 875–881 (1993).
12. F. Sternberg, C. Meyerhoff, F.J. Mennel, H. Mayer, F. Bischof, E.F. Pfeiffer, 'Does Fall in Tissue Glucose Precede Fall in Blood Glucose?', *Diabetologia*, **39**, 609–612 (1996).
13. V. Thomé-Duret, G. Reach, M.N. Gangnerau, F. Lemonnier, J.C. Klein, Y. Zhang, Y. Hu, G.S. Wilson, 'Use of a Subcutaneous Glucose Sensor to Detect Decreases in Glucose Concentration Prior to Observation in Blood', *Anal. Chem.*, **68**, 3822–3826 (1996).

14. J.M. Burrin, C.P. Price, 'Measurement of Blood Glucose', *Ann. Clin. Biochem.*, **22**, 327–342 (1985).
15. H.U. Bergmeyer (ed.), *Methods of Enzymatic Analysis*, VCH, Weinheim, 1983–1988.
16. D.C. Klonoff, 'Noninvasive Blood Glucose Monitoring', *Diabetes Care*, **20**, 433–437 (1997).
17. H.A. MacKenzie, H.S. Ashton, Y.C. Shen, J. Lindberg, P. Rae, K.M. Quan, S. Spiers, 'Blood Glucose Measurements by Photoacoustics', *OSA Trends Opt. Photon. Ser.*, **22**, 156–159 (1998).
18. J.J. Workman, 'Interpretive Spectroscopy for Near Infrared', *Appl. Spectrosc. Rev.*, **31**, 251–320 (1996).
19. H.M. Heise, R. Marbach, A. Bittner, Th. Koschinsky, 'Clinical Chemistry and Near-infrared Spectroscopy: Multicomponent Assay for Human Plasma and its Evaluation for the Determination of Blood Substrates', *J. Near Infrared Spectrosc.*, **6**, 361–374 (1998).
20. O.S. Khalil, 'Spectroscopic and Clinical Aspects of Noninvasive Glucose Measurements', *Clin. Chem.*, **45**, 165–177 (1999).
21. K. Molt, A. Niemöller, Y.J. Cho, 'Analysis of Aqueous Solutions by Near-infrared Spectrometry (NIRS) II. Titrations of Weak and Very Weak Acids with Strong Bases', *J. Mol. Struct.*, **410/411**, 565–572 (1997).
22. K. Martin, 'In Vivo Measurements of Water in Skin by Near-infrared Reflectance', *Appl. Spectrosc.*, **52**, 1001–1007 (1998).
23. S. Berentsen, T. Stolz, K. Molt, 'Analysis of Aqueous Solutions by Near-infrared Spectrometry (NIRS) IV. One- and Two-component Systems of Organic Compounds in Water', *J. Mol. Struct.*, **410/411**, 581–585 (1997).
24. H.M. Heise, R. Marbach, 'Human Oral Mucosa Studies with Varying Blood Glucose Concentration by Noninvasive ATR-FT-IR-Spectroscopy', *Cell. Mol. Biol.*, **44**(6), 899–912 (1998).
25. B.C. Wilson, S.L. Jacques, 'Optical Reflectance and Transmittance of Tissue: Principles and Applications', *IEEE J. Quant. Electron.*, **26**, 2186–2199 (1990).
26. R. Marbach, H.M. Heise, 'Optical Diffuse Reflectance Accessory for Measurements of Skin Tissue by Near-infrared Spectroscopy', *Appl. Opt.*, **34**, 610–621 (1995).
27. J. Coates, 'Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy', *Appl. Spectrosc. Rev.*, **33**, 267–425 (1998).
28. G. Spanner, R. Nießner, 'New Concepts for the Noninvasive Determination of Physiological Glucose Concentrations Using Modulated Laser Diodes', *Fresenius J. Anal. Chem.*, **354**, 306–310 (1996).
29. A.K. Amerov, K.J. Jeon, Y.-J. Kim, G. Yoon, 'Method and Device for Non-invasive Blood Glucose Measurement', *Proc. SPIE*, **3599**, 33–42 (1999).
30. J. Malinen, M. Känsäkoski, 'LED-based Spectrometer Modules for Hand-held Sensors and On-line Process Monitoring', *Proc. SPIE*, **3537**, 88–95 (1999).
31. H. Kopola, P. Karioja, O. Rusanen, A. Lehto, J. Lammasniemi, 'Hybridization, Assembling, and Testing of Miniaturized Optoelectronic Modules for Sensors and Microsystems', *Proc. SPIE*, **3276**, 56–64 (1998).
32. H.M. Heise, L. Küpper, L.N. Butvina, 'Attenuated Total Reflection Mid-infrared Spectroscopy for Clinical Chemistry Applications Using Silver Halide Fibers', *Sensors Actuators*, **B51**, 84–91 (1998).
33. K.-U. Jagemann, C. Fischbacher, K. Danzer, U.A. Müller, B. Mertes, 'Application of Near-infrared Spectroscopy for Non-invasive Determination of Blood/Tissue Glucose Using Neural Networks', *Z. Phys. Chem.*, **191**, 179–190 (1995).
34. A. Bittner, S. Thomaßen, H.M. Heise, 'In-vivo Measurements of Skin Tissue by Near-infrared Diffuse Reflectance Spectroscopy', *Mikrochim. Acta*, **14**(Suppl.), 429–432 (1997).
35. H.M. Heise, 'Medical Applications of Infrared Spectroscopy', *Mikrochim. Acta*, **14**(Suppl.), 67–77 (1997).
36. S.K. Loyalka, C.A. Riggs, 'Inverse Problem in Diffuse Reflectance Spectroscopy: Accuracy of the Kubelky–Munk Equations', *Appl. Spectrosc.*, **49**, 1107–1110 (1995).
37. H. Martens, T. Naes, *Multivariate Calibration*, John Wiley & Sons, New York, 1989.
38. P. Bhandare, Y. Mendelson, R.A. Peura, G. Janatsch, J.D. Kruse-Jarres, R. Marbach, H.M. Heise, 'Multivariate Determination of Glucose in Whole Blood Using Partial Least-squares and Artificial Neural Networks Based on Mid-infrared Spectroscopy', *Appl. Spectrosc.*, **47**, 1214–1221 (1993).
39. H.M. Heise, R. Marbach, G. Janatsch, J.D. Kruse-Jarres, 'Multivariate Determination of Glucose in Whole Blood by Attenuated Total Reflection Infrared Spectroscopy', *Anal. Chem.*, **61**, 2009–2015 (1989).
40. N.R. Draper, H. Smith, *Applied Regression Analysis*, 2nd edition, John Wiley & Sons, New York, 1981.
41. M.J. McShane, B.D. Cameron, G.L. Coté, C.H. Spiegelman, 'Peak-hopping Stepwise Wavelength Selection Algorithm', *Proc. SPIE*, **3599**, 101–109 (1999).
42. H.M. Heise, A. Bittner, 'Rapid and Reliable Spectral Variable Selection for Statistical Calibrations Based on PLS-regression Vector Choices', *Fresenius' J. Anal. Chem.*, **359**, 93–99 (1997).
43. H.M. Heise, A. Bittner, 'Multivariate Calibration for Near-infrared Spectroscopic Assays of Blood Substrates in Human Plasma Based on Variable Selection using PLS-Regression Vector Choices', *Fresenius' J. Anal. Chem.*, **362**, 141–147 (1998).
44. H.M. Heise, A. Bittner, R. Marbach, 'Clinical Chemistry and Near-infrared Spectroscopy: Technology for Noninvasive Glucose Monitoring', *J. Near Infrared Spectrosc.*, **6**, 349–359 (1998).
45. M.R. Riley, M.A. Arnold, D.W. Murhammer, 'Matrix-enhanced Calibration Procedure for Multivariate Calibration Models with Near-infrared Spectra', *Appl. Spectrosc.*, **52**, 1339–1347 (1998).

46. J.R. Mourant, J.P. Freyer, A.H. Hielscher, A.A. Eick, D. Shen, T.M. Johnson, 'Mechanisms of Light Scattering from Biological Cells Relevant to Noninvasive Optical-tissue Diagnostics', *Appl. Opt.*, **37**, 3586–3593 (1998).
47. S.J. Matcher, M. Cope, D.T. Delpy, 'Use of the Water Absorption Spectrum to Quantify Tissue Chromophore Concentration Changes in Near Infrared Spectroscopy', *Phys. Med. Biol.*, **39**, 177–196 (1994).
48. M.S. Patterson, B.C. Wilson, D.R. Wyman, 'The Propagation of Optical Radiation in Tissue. I. Models of Radiation Transport and Their Application', *Lasers Med. Sci.*, **6**, 155–168 (1991).
49. J. Qu, B.C. Wilson, 'Monte Carlo Modeling Studies of the Effect of Physiological Factors and Other Analytes on the Determination of Glucose Concentration In Vivo by Near Infrared Optical Absorption and Scattering Measurements', *J. Biomed. Opt.*, **2**, 319–325 (1997).
50. H.M. Heise, 'Near-infrared Spectrometry for In Vivo Glucose Sensing', in *Biosensors in the Body – Continuous In Vivo Monitoring*, ed. D.M. Fraser, John Wiley, Chichester, 79–116, 1997.
51. W. Clarke, D. Cox, L. Gonder-Frederick, W. Carter, S. Pohl, 'Evaluating Clinical Accuracy of Systems for Self-monitoring of Blood Glucose', *Diabetes Care*, **10**, 622–628 (1987).
52. Th. Koschinsky, K. Dannehl, F.A. Gries, 'New Approach to Technical and Clinical Evaluation of Devices for Self-monitoring of Blood Glucose', *Diabetes Care*, **11**, 619–629 (1988).
53. J.P. de Kock, L. Tarassenko, 'Pulse Oximetry: Theoretical and Experimental Models', *Med. Biol. Eng. Comp.*, **31**, 291–300 (1993).
54. M.J. Hayes, P.R. Smith, 'Artifact Reduction in Photoplethysmography', *Appl. Optics*, **37**, 7437–7446 (1998).
55. S. Mansouri, J.S. Schultz, 'A Miniature Optical Glucose Sensor Based on Affinity Binding', *Biotechnology*, **2**, 885–890 (1984).
56. J.R. Lakowicz, B. Maliwal, 'Optical Sensing of Glucose Using Phase-modulation Fluorimetry', *Anal. Chim. Acta*, **271**, 155–164 (1992).
57. L. Tolosa, H. Szmacinski, G. Rao, J.R. Lakowicz, 'Lifetime-based Sensing of Glucose Using Energy Transfer with a Long Lifetime Donor', *Anal. Biochem.*, **250**, 102–108 (1997).
58. M.J. McShane, S. Rastegar, G.L. Coté, 'Probe Design for Implantable Fluorescence-based Sensors', *Proc. SPIE*, **3599**, 93–100 (1999).
59. O.J. Rolinski, D.J.S. Birch, L.J. McCartney, J.C. Pickup, 'Near Infrared Assay for Glucose Determination', *Proc. SPIE*, **3602**, 6–14 (1999).
60. G.L. Coté, B.D. Cameron, 'Noninvasive Polarimetric Measurement of Glucose in Cell Culture Media', *J. Biomed. Opt.*, **2**, 275–281 (1997).
61. B.D. Cameron, H. Gorde, G.L. Coté, 'Development of an Optical Polarimeter for In Vivo Glucose Monitoring', *Proc. SPIE*, **3599**, 43–49 (1999).
62. C. Chou, C.Y. Han, W.C. Kuo, Y.C. Huang, C.M. Feng, J.C. Shyu, 'Noninvasive Glucose Monitoring In Vivo with an Optical Heterodyne Polarimeter', *Appl. Opt.*, **37**, 3553–3557 (1998).
63. L. Heinemann, G. Schmelzeisen-Redeker et al., 'Noninvasive Continuous Glucose Monitoring in Type I Diabetic Patients with Optical Glucose Sensors', *Diabetologia*, **41**, 848–854 (1998).
64. A.J. Berger, T.-W. Koo, I. Itzkan, G. Horowitz, M.S. Feld, 'Multicomponent Blood Analysis by Near-infrared Raman Spectroscopy', *Appl. Opt.*, **38**, 2916–2926 (1999).
65. J. Lambert, M. Storrie-Lombardi, M. Borchert, 'Measurement of Physiologic Glucose Levels Using Raman Spectroscopy in a Rabbit Aqueous Humor Model', *LEOS Newsl.*, **12**(2), 19–22 (1998).
66. R.V. Tarr, P.G. Steffes, 'The Non-invasive Measure of D-glucose in the Ocular Aqueous Humor Using Stimulated Raman Spectroscopy', *LEOS Newsl.*, **12**(2), 22–27 (1998).
67. D.C. Klonoff, J. Braig, B. Sterling, C. Kramer, D. Goldberger, R. Trebino, 'Mid-infrared Spectroscopy for Noninvasive Blood Glucose Monitoring', *LEOS Newsl.*, **12**(2), 13–14 (1998).
68. W. Diller, 'Just Around the Corner: Pursuing the Elusive Diabetes Market', *Start-up*, **Dec.**, 27–38 (1998).

Infrared Spectroscopy in Clinical and Diagnostic Analysis

R. Anthony Shaw and Henry H. Mantsch
Institute for Biodiagnostics, National Research Council of Canada, Winnipeg, Canada

1 Introduction 1
2 Infrared Spectroscopy of Biological Fluids 2
2.1 Mid-infrared Attenuated Total Reflectance Spectroscopy 3
2.2 Mid-infrared Spectroscopy of Dried Films 3
2.3 Near-infrared Spectroscopy 4
3 Calibration Methods 4
3.1 Multiple-wavelength Linear Regression 5
3.2 Principal Component Regression and Partial Least Squares 6
3.3 Spectral Preprocessing 6
4 Serum Analysis 6
4.1 Infrared Spectroscopy of Serum 6
4.2 Serum Analysis using Near-infrared Spectroscopy 7
4.3 Serum Analysis using Mid-infrared Spectroscopy 10
5 Serum and Blood Glucose 14
6 Fetal Lung Maturity Determined by Infrared Spectroscopy 14
7 Other Fluids 15
7.1 Urine Analysis 15
7.2 Saliva 16
8 Disease Diagnosis Based on Infrared Spectral Pattern Recognition 17
8.1 Arthritis Diagnosis from Infrared Spectroscopy of Synovial Fluid 17
8.2 Disease Pattern Recognition in Mid-infrared Spectra of Serum 18
9 Summary 18
Acknowledgments 18
Abbreviations and Acronyms 18
Related Articles 18
References 19

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The infrared spectrum of a mixture serves as the basis to quantitate its constituents, and a number of common clinical chemistry tests have proven to be feasible using this approach. This article reviews the infrared spectroscopy-based analytical methods that have been developed for consideration as clinical assays, including serum analysis, urine analysis, amniotic fluid assays for the estimation of fetal lung maturity, and others. Because of the widespread interest in the potential for in vivo measurement of blood glucose using near-infrared spectroscopy, a separate section is devoted to the analysis of glucose in whole blood.

A related technique uses the infrared spectrum of biomedical specimens directly as a diagnostic tool. For example, the spectra of serum and of synovial fluid have proven to be useful in the diagnosis of metabolic disorders and arthritis, respectively, without explicitly recovering their chemical composition from the spectra. Rather, characteristic spectral features and patterns have been identified as the basis to distinguish spectra corresponding to healthy patients from those corresponding to diseased patients. These applications are reviewed here.

Issues such as ease of use, speed, reliability, sample size, and calibration stability all play important roles in governing the practical acceptability of infrared spectroscopy-based analytical methods. To provide a framework to illustrate these issues, descriptions are included for the various procedures that have been explored to wed successfully infrared spectroscopy to clinical chemistry.

1 INTRODUCTION

Infrared (IR) spectroscopy has emerged in recent years as the analytical method of choice in an enormous variety of applications. What brought about this revolution? The clearest advantage is that no specific reagents are required. Automated, repetitive analyses can therefore be carried out at very low cost. The appeal of these factors has spurred the development of a new generation of analytical IR spectrometers that combine high acquisition speed with superb spectral sensitivity. Powerful chemometric algorithms and software packages have emerged in parallel with the new hardware, and new applications emerge continually.

Rather than relying upon reagents to promote color reactions, IR-based analysis is founded upon the spectrum of IR colors characteristic of the analyte itself. If a particular component provides an IR absorption spectrum, and its concentration is high enough that the spectrum contributes meaningfully to the IR absorption profile, then it may, in principle, be quantified by using IR spectroscopy. Although the requirement that the

component exhibits an IR absorption spectrum rules out the quantitation of simple ions, a number of very common clinical analytical tests may, in principle, be carried out using IR spectroscopy.

This article begins by comparing and contrasting mid-infrared (MIR) and near-infrared (NIR) spectroscopy in the context of analytical applications. The second section describes the general approach to generating an IR-based quantitation method. Although Beer's law generally holds true for common analytes in biological fluids, it is very unusual to find a single absorption that can be used as the basis to quantify any single component in real-life samples. Analytical methods that are based upon IR spectroscopy must nearly always be calibrated by reference to accepted clinical analyses, using multiple-wavelength linear regression or other full-spectrum methods.

The function of the clinical chemistry laboratory is to perform quantitative and qualitative analyses on body fluids such as serum, blood, urine, and spinal fluid, as well as other materials such as tissue, calculi, and feces. The main body of this article describes IR-based methods to carry out some of the most common clinical analytical tests, specifically those involving serum, whole blood, and urine. Fluids that are less commonly assayed (e.g. saliva and amniotic fluid) are also discussed separately. NIR spectroscopy has achieved some notoriety in the clinical chemistry arena because of the early promise that it might serve as the basis for a noninvasive blood glucose test. Some relevant in vitro studies are surveyed briefly here. The article closes with a discussion of novel approaches to derive diagnosis directly, without explicit quantitative analysis, from the spectra of biological fluids.

2 INFRARED SPECTROSCOPY OF BIOLOGICAL FLUIDS

The IR spectral region ranges from the red end of the visible spectrum at 780 nm (12 820 cm^{-1}) to the onset of the microwave region at a wavelength of 1 mm (10 cm^{-1}). Traditionally, this range is further subdivided into the near-infrared (NIR), mid-infrared (MIR), and far-infrared (FIR). The MIR region covers the range 400–4000 cm^{-1}, and is the region most familiar to the organic chemist as providing a "fingerprint" characteristic of molecular species. It is this region that includes the rich spectrum of absorptions corresponding to fundamental vibrations of the species being probed.

Although MIR absorption positions are almost universally reported in units of wavenumbers (cm^{-1}; the inverse of the wavelength in centimeters), it remains the norm for NIR spectra to be reported in wavelength units, generally in nanometers. The NIR spans the range 780–2500 nm, encompassing weak transitions that correspond to combinations and overtones of the vibrational modes observed in the MIR. Because NIR absorptions are generally broad and therefore strongly overlapping, it is difficult or impossible to arrive at specific assignments for individual absorptions in the NIR. Partly for that reason, these transitions were long ignored by the spectroscopic communities (indeed by all communities!), although their potential for use in analytical work was noted as early as the mid-1950s.[1] By the end of the following decade, the technique had caught the attention of the agricultural community as a possible means to determine protein, oil, and moisture content of agricultural commodities. In this application and many others developed since, the inherently weak absorptions proved advantageous, permitting the convenience of longer optical path lengths than are feasible in MIR work, and hence relatively easy sample handling. Since the seminal work of Norris,[2] Williams,[3] and others, NIR spectroscopy has largely matured, and now finds acceptance in an enormous variety of analytical applications. The majority of NIR spectrometers manufactured today are customized for analytical applications, including appropriate software and simplified user interfaces for routine operation.

In considering the use of IR spectroscopy for clinical analyses, we are confronted with the fact that the most abundant species found in all biological fluids is water, and the IR spectra reflect this fact. To illustrate the dominance of water in the IR spectra, Figures 1 and 2 depict the absorption profiles for native serum in the MIR and NIR spectral regions. Although some of the stronger solute absorptions do emerge in the MIR spectra, water clearly dominates the overall appearance. The NIR spectra are

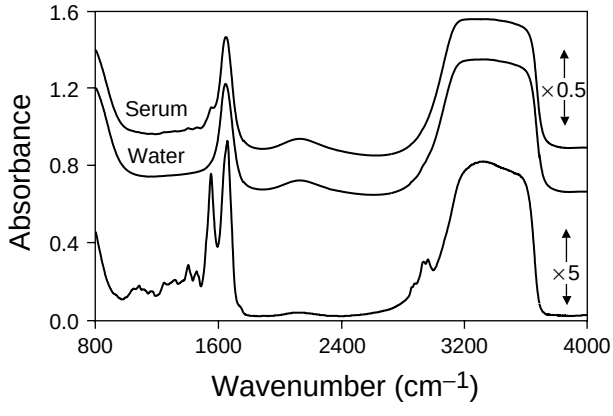

Figure 1 MIR absorption spectra of serum and of water, collected with an optical path length of 6 μm. The lower trace is a difference spectrum, with the spectrum of water subtracted from that of serum. Note the tenfold difference in the absorbance scales.

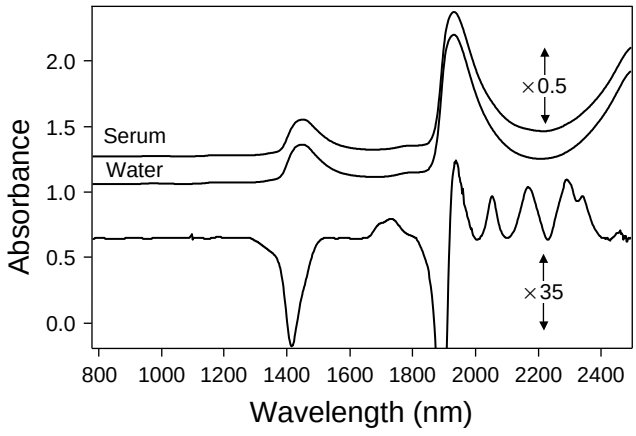

Figure 2 NIR absorption spectra of serum and of water, collected with an optical path length of 0.5 mm. The lower trace is a difference spectrum, with the spectrum of water subtracted from that of serum. Note the 70-fold difference in the absorbance scales.

apparently devoid of any absorptions other than those of water.

MIR and NIR spectroscopies in fact offer quite different, complementary, approaches to analysis. The richness of the MIR spectrum makes it instinctively appealing as the method of choice for analytical work, however NIR has practical benefits such as convenience in sample handling and the fact that the sample cells do not require specialized materials. Whereas MIR spectroscopy of aqueous specimens typically requires optical path lengths of the order of microns, NIR transmission spectra are generally collected using path lengths of 0.5 mm or greater. The question of whether to use NIR or MIR spectroscopy for analytical purposes then translates to the question of whether the additional effort generally required to acquire MIR spectra is compensated by other possible benefits such as greater analytical accuracy or smaller sample volume.

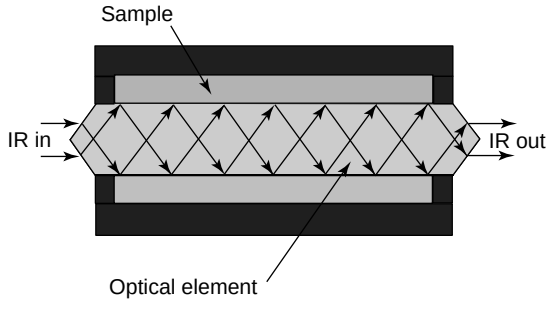

Figure 3 Apparatus to measure the ATR spectrum for a liquid specimen. The ATR spectrum is derived by ratioing the single-beam spectrum measured with the specimen in place against a single-beam spectrum for the clean optical element.

2.1 Mid-infrared Attenuated Total Reflectance Spectroscopy

Attenuated total reflectance (ATR) spectroscopy provides an alternative means to measure absorption spectra by using the experimental arrangement illustrated in Figure 3. The clearest advantage of this method is that it provides a means to measure MIR spectra for strongly absorbing aqueous solutions, without the inconvenience and imprecision involved in working at very short path lengths that are required for transmission spectroscopy. Rather than transmitting IR radiation through the specimen, the liquid sample is placed in contact with the ATR optical element.[4,5] The refractive index of the element (typically zinc selenide) is high enough that the IR beam propagating through it undergoes several internal reflections as it travels through the crystal. A background spectrum is first measured with no sample in the cell. The sample is then placed in contact with the crystal. The internally reflected beam effectively penetrates the sample to depths of 0.5–2 μm and hence is attenuated at wavelengths corresponding to sample absorptions. Ratioing the resulting single-beam spectrum against the background spectrum results in a spectrum that is nearly identical to the absorption spectrum, differing only by virtue of the wavelength dependence of the penetration depth.

2.2 Mid-infrared Spectroscopy of Dried Films

This approach finesses the difficulties associated with strong water absorptions by simply eliminating water from the specimen. Typically 5–50 μL of liquid is spread on a suitable substrate and allowed to dry, and a transmission spectrum is acquired for the resulting film. In addition to eliminating the spectral interference of water,

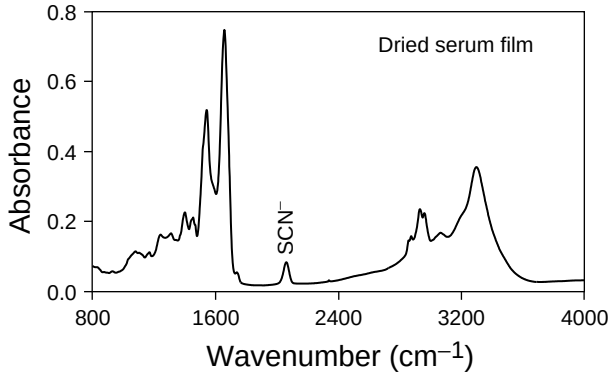

Figure 4 Absorption (transmission) spectrum for a serum film dried onto a barium fluoride window. The specimen was first diluted twofold in aqueous 4 g L^{-1} potassium thiocyanate (KSCN) solution. The absorption of SCN^- at 2060 cm^{-1} was used for subsequent normalization of the spectra as part of the development of quantitation models (see Shaw et al.[23]).

this approach can provide inherently better spectral resolution by virtue of eliminating the water/solute interactions. A representative spectrum of a dry serum film is illustrated in Figure 4.

2.3 Near-infrared Spectroscopy

Although the NIR is defined as encompassing the 780–2500 nm spectral range, it is convenient to subdivide further this span into natural subregions. The 2000–2500 nm range includes the most intense absorptions and thus is the region most commonly exploited for analytical purposes. Absorptions in this region correspond to ''combination bands'', combining X–H (where X = C, N, O) stretches with other fundamental vibrations, whereas practically all of the higher energy transitions correspond to vibrational first (1400–1800 nm), second (950–1250 nm), and third overtones (Table 1).

The diversity of transitions in the NIR region has interesting practical consequences. For aqueous solutions, the 2000–2500 nm region is best explored by using a transmission cell with an optical path length of 0.5–2.5 mm. The optimal path length to observe the first overtone transitions is longer – of the order of 5–10 mm – whereas observation of the second overtones requires a path length of several centimeters. Where sample volume is a consideration, a relatively short path length is a necessity and the combination region is therefore preferred. Another outgrowth of this trend is that tissue is relatively transparent at shorter NIR wavelengths. A key consideration in the search for in vivo analytical methods (e.g. blood glucose) is therefore to arrange the experiment such that the effective optical path length is optimized for the appropriate analyte NIR absorption features. A proposed in vivo method based upon the combination bands will require a short effective path length, whereas a method that monitors second overtone absorptions would require a much longer one.

Table 1 NIR vibrational transitions

NIR spectral range (nm)	Nature of vibrational transitions
2200–2450	C–H stretch combinations
2000–2200	N–H, O–H stretch combinations
1650–1800	C–H stretch, 1st overtones
1400–1500	N–H, O–H stretch, 1st overtones
1100–1225	C–H stretch, 2nd overtones
950–1100	N–H, O–H stretch, 2nd overtones
850–950	C–H stretch, 3rd overtones
775–850	N–H stretch, 3rd overtones

3 CALIBRATION METHODS

In the vast majority of cases, IR-based analytical methods are developed via calibration to accepted reference analyses. The term ''calibration'' therefore describes the derivation of a model with which to recover quantitative analytical information from the IR spectra. Although this step is obviously a trivial one for very simple one- or two-component systems, more complex mixtures require a more sophisticated approach.

The general procedure is the same regardless of the details of the process. The first stage is to accumulate both IR spectra and reference assays for a set of appropriate clinical specimens. Ideally, this set of calibration samples should span the range of concentrations expected both for the analyte of interest and for any interfering species (i.e. any IR absorber other than the target compound). Separate calibration models are then developed for each of the target analytes. Finally, each of the calibration models is validated by comparing IR-predicted levels to the reference levels determined for an independent set of test specimens. An outline of the model development process is presented in Table 2.

This section introduces three of the more common techniques: multiple-wavelength linear regression (MLR),

Table 2 Development of an IR-based clinical analytical method

Preparation

- Collect clinical specimens
- Carry out reference analyses for species of interest
- Measure corresponding IR spectra
- Designate two-thirds of the total number of spectra as the calibration set and the remaining one-third of spectra as the validation set
- Preprocess spectra

Calibration

- Choose modeling method (e.g. MLR, PLS, PCR)
- Generate models ranging in complexity from a very few variables (wavelength terms for MLR; factors for PLS, PCR) to many variables
- Predict concentrations using all models and compare to reference analyses
- Identify outliers and correct or remove as appropriate
- Recalibrate models with outliers removed
- Evaluate standard errors of calibration and cross-validation for each model and plot as a function of the number of variables in the model

Validation

- Predict concentrations using all models and compare to reference analyses
- Evaluate standard error of prediction for each model and plot as a function of the number of variables in the model

Where IR methods are sought for more than one species, the calibration/validation procedure is carried out independently for each analyte.

principal component regression (PCR), and partial least squares (PLS).

3.1 Multiple-wavelength Linear Regression

This approach is simply an extension of Beer's law to include multiple wavelengths. The need for several wavelengths is dictated by the inherent richness of the IR spectra – it is generally difficult or impossible to find a single absorption corresponding to a particular analyte that is not overlapped by the absorptions of at least one other constituent. The Beer's law relationship shown in Equation (1):

$$A = \varepsilon c L \tag{1}$$

is therefore replaced by the more general form shown in Equation (2):

$$A = \sum \varepsilon_i c_i L \tag{2}$$

where A is the absorbance, ε_i is the molar absorptivity of the ith constituent, c_i is the concentration of the ith constituent, and L is the optical path length. The expression relating the concentration to IR absorption intensities then takes the form of Equation (3):

$$c_i = K_{0i} + K_{1i}A(\lambda_1) + K_{2i}A(\lambda_2) + \cdots + K_{Ni}A(\lambda_N) \tag{3}$$

where K_{Ni} are the calibration coefficients for the ith constituent, and λ_N are the corresponding analytical wavelengths.

This approach is most readily applied when the spectra show dominant absorptions corresponding to the analyte of interest. This proved to be the case in the NIR analysis of urine urea.[6] Figure 5 compares the spectra of five representative urine specimens in the NIR region of 2050–2500 nm, together with the spectra of five aqueous urea solutions spanning the concentration range 30–230 mmol L^{-1}. In this instance the spectral features due to urea clearly dominate the urine spectra, and it is not surprising that a single-wavelength Beer's law relationship provided quite good accuracy in extracting the urea concentrations. In particular, the intensity of the feature at 2152 nm proved sufficient to recover urea concentrations with a standard error of 20 mmol L^{-1} over the physiological concentration range of 100–500 mmol L^{-1}. The accuracy was improved, however, by including additional terms as shown in Equation (4):

$$C(\text{urea}) = -10 + 68\ \frac{A(2152\,\text{nm})}{A(1194\,\text{nm})} + 1.3 \times 10^5 A(1724\,\text{nm}) \tag{4}$$

There are two new wavelengths in this model. The first, at 1194 nm, corresponds to a weak water absorption. The most common rationale for including such a term as a divisor of the primary wavelength is to correct for fluctuations in the effective optical path length, generally caused by light scattering due to particulate matter in the sample. The second new term, at 1724 nm, corresponds to a weak protein absorption. This term, particularly in those samples with unusually high protein levels, may serve to correct for the contribution of protein absorptions to the intensity at 2152 nm.

For analytes that do not yield prominent absorptions in the spectra of the target specimens, the simple single-term Beer's law relationship fails completely. One solution in this case is again to assume a solution of the type represented by Equation (3), whereupon the task becomes to determine how many and which wavelengths/frequencies should be included in the analysis. One approach is to regress the set of spectral intensities, for each wavelength, against the analyte concentrations for the calibration specimens. The single wavelength that provides the best correlation with concentration is then taken as the ''primary'' wavelength, and further regressions are carried out holding the primary wavelength fixed to determine additional terms to complement the single-term model. The same process may be used to determine divisor terms (see the first term of Equation 4).

The stepwise regression approach to determining MLR terms is not guaranteed to find the optimal set of wavelengths, particularly for complex specimens where many terms may be required. The general problem is illustrated by the fact that in a set of calibration spectra, each comprising 2000 absorbance values, there are 2.5×10^{26} possible eight-term wavelength combinations. Brute-force evaluation of all possible eight-term MLR models is clearly out of the question, and there is an

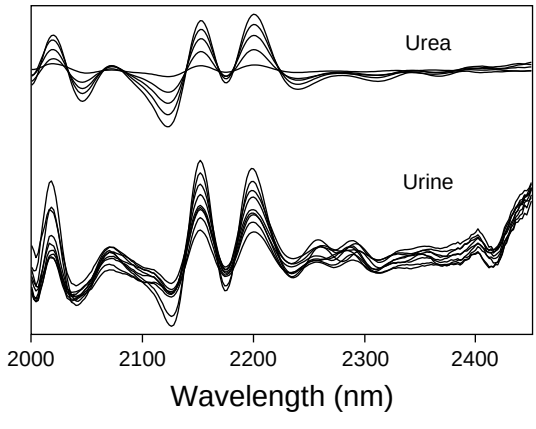

Figure 5 NIR spectra (second derivatives, inverted) for a representative set of nine urine specimens (lower traces) and for five aqueous urea samples ranging from 30 to 230 mmol L^{-1} (upper traces).

ongoing search for more efficient methods.[7] Recent developments include genetic algorithms to identify the optimal spectral regions. For example, an algorithm originally developed to identify diagnostic patterns in magnetic resonance spectra[8] has been modified recently to seek out optimal spectral subregions for MLR.

3.2 Principal Component Regression and Partial Least Squares

The feature common to both of these approaches is that each spectrum is reduced to a sum of pseudospectra, or ''loading vectors''. Each spectrum is newly represented by a unique set of ''scores'' – the set of coefficients required to reconstruct the original spectrum from the set of loading vectors. Typically, each of the spectra can be reconstructed to within the noise limits by a combination of typically 5–15 loading vectors, as compared to the hundreds or thousands of intensity values in the original spectra. The scores then provide the basis for quantitation.

The essential relationship in both the PCR and PLS models takes the form of Equation (5):

$$\mathbf{A} = \mathbf{TB} + \mathbf{E_A} \tag{5}$$

With m spectra in the calibration set, each having n absorbance values, $\mathbf{A}$ is the $m \times n$ matrix of the calibration spectra. The spectra are reconstructed as a product of $\mathbf{B}$ ($h \times n$), the new basis set of loading vectors, and $\mathbf{T}$ ($m \times h$), the scores. To reiterate, the key to the process is that each spectrum is reduced from a vector of length n (a row in $\mathbf{A}$) to a new vector of length h (the corresponding row in $\mathbf{T}$), where h is typically between 5 and 15.

The column matrix of concentrations $\mathbf{c}$ is also related to the loading vectors $\mathbf{T}$, according to Equation (6):

$$\mathbf{c} = \mathbf{Tv} + \mathbf{e_c} \tag{6}$$

Here, $\mathbf{v}$ is the matrix of coefficients that relates the scores to the concentrations.

The reader is referred to several works in the literature[9–14] for fuller explanations of PLS and PCR methods. For the sake of the present discussion, we note the following features common to the two methods:

- The main challenge in developing a method is to decide how many (and, in the case of PCR, which) loading vectors to include.
- The overall performance of either method may be improved by eliminating superfluous spectral regions from $\mathbf{A}$.
- The modeling of the spectra provides a means to detect outliers (those spectra with extraordinarily large spectral residuals $\mathbf{E_A}$).

3.3 Spectral Preprocessing

It is almost always necessary, or at least desirable, to preprocess the absorption spectra in some fashion; the aim is to enhance the spectral features that carry information regarding the analyte of interest, and effectively to suppress or eliminate superfluous features. The simplest form of ''preprocessing'' is the selection of appropriate wavelengths in MLR model development; the analogy in PLS and PCR is the selection of a limited spectral region (or regions).

The most common procedures are mean centering, variance scaling, and derivation. Mean centering simply subtracts the average of the calibration spectra from each of the individual spectra. Variance scaling involves first evaluating the standard deviation among spectra for the intensity at each wavelength. All spectra are then divided by the pseudospectrum of standard deviations, and hence scaled so that the variance is unity at all wavelengths. This operation effectively enhances the prominence of features due to species of relatively low concentration, while suppressing the intensities of strong (and variable) absorptions. The procedure is therefore most appropriate for the analysis of minor components. Derivation is commonly used to eliminate random fluctuations in the baseline (first derivative) and slope (second derivative) of the absorption spectra. Another benefit is the effective narrowing of spectral features, which may enhance specificity in the analytical method. Note that the features in the second-derivative spectrum are inverted relative to the absorption spectra. Although the second-derivative spectra plotted in Figure 5 have been inverted to yield peaks rather than valleys at positions corresponding to absorption peaks, this convention is not followed universally.

4 SERUM ANALYSIS

These analyses play a critical role in diagnosing and monitoring a wide variety of disorders (see Table 3), and a typical central hospital laboratory typically carries out many thousands of such tests every month. In order for a new testing procedure to be accepted clinically it must meet well-defined accuracy and precision standards. Although practical considerations such as the degree of automation also play a role in the acceptability of novel methods, these issues lie outside the scope of this article. In this section we present the current state of the art in the IR-based analysis of serum.

4.1 Infrared Spectroscopy of Serum

Among the most common clinical serum tests are those for the most abundant organic species. For at least

Table 3 Selected serum analytes that may be determined using IR spectroscopy

Analyte	Reference intervals[a]		Associated conditions[a]
Total protein	60–83 g L^{-1} (adult)	↑	Hypovolemic states
		↓	Nutritional deficiency
			Liver disease
			Renal disease
			Fever
			Inflammation
Albumin[b]	32–48 g L^{-1} (adult)	↑	Dehydration
		↓	Pregnancy
Urea	7–18 mg dL^{-1} (adult)	↑	Impaired kidney function
	(2.5–6.4 mmol L^{-1})		Congestive heart failure
			Stress
		↓	Severe liver damage
			Low protein diet
			Nephrotic syndrome
Glucose	65–105 mg dL^{-1}	↑	Diabetes mellitus
	(3.6–5.8 mmol L^{-1})		Acute pancreatitis
			Stress/shock
		↓	Pancreatic disorders
			Hepatic disease
			Extrapancreatic tumors
Cholesterol	150–235 mg dL^{-1} (male)[c]	↑	Idiopathic hypercholesterolemia
	(3.9–6.1 mmol L^{-1})		Biliary obstruction
			Pregnancy
	141–219 mg dL^{-1} (female)[c]		Hypothyroidism
	(3.6–5.7 mmol L^{-1})	↓	Severe liver damage
			Malnutrition
			Hyperthyroidism
Triglycerides	48–189 mg dL^{-1} (male)[d]	↑	Liver diseases
	(0.5–2.1 mmol L^{-1})		Familial hyperlipidemia
			Alcoholism
	40–117 mg dL^{-1} (female)[d]		Gout
	(0.45–1.3 mmol L^{-1})	↓	Malnutrition

[a] From Wallach[15] and Tietz:[16] ↑ indicates conditions associated with levels above the reference range; ↓ indicates conditions associated with levels below the reference range.
[b] Serum albumin levels generally parallel to total protein levels.
[c] Desirable range (5th percentile to 75th percentile) for 40-year-old individuals. For men, the upper limit of the desirable range rises by approximately 1 mg dL^{-1} for every year after 40; for women, the increase is ~3 mg dL^{-1} for every year after 40.
[d] Desirable range (reference interval is somewhat wider).

six of these, the MIR spectra are distinctive enough and the concentrations are high enough that they may be determined from the MIR spectra of serum. These include glucose, total protein, albumin, triglycerides, urea, and cholesterol. The basis for detecting and discriminating among the six analytes is illustrated by the spectra of the pure compounds shown in Figure 6. The NIR spectra also permit quantitation of the same six species.

Four comprehensive feasibility studies have been published, all of which differ in significant ways. Two were based upon MIR spectroscopy, and two on NIR spectroscopy. One MIR investigation used ATR spectroscopy, and another used dried serum films; the two NIR studies differed in more subtle, yet substantial, details.

4.2 Serum Analysis using Near-infrared Spectroscopy

Two major systematic investigations have been carried out. The first of these, reported in a pair of publications by Hall and Pollard,[17,18] was based upon a rapid-scanning NIR spectrometer. The authors reported analytical methods for urea, triglycerides, total protein, and albumin.

Because the absorptions of protein overwhelm those of other dissolved species, it proved possible to use a simple MLR model to quantitate serum total protein. The absorption spectra of albumin further proved to be clearly distinguishable from those of the remaining proteins (primarily globulins; see Figure 7), so that a second two-term MLR model was sufficient to determine albumin levels. Based upon the second derivatives of the absorption spectra, the two models were as shown in

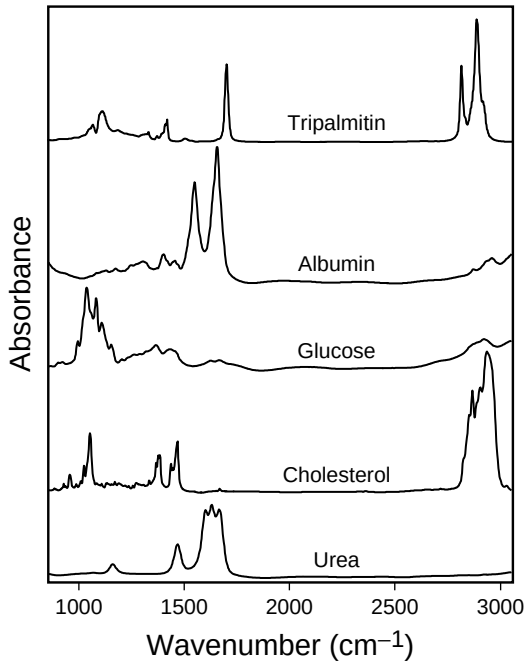

Figure 6 MIR absorption spectra for selected serum constituents. The spectra for urea, glucose, and albumin were acquired for aqueous solutions using an optical path length of 6 μm (the spectrum of water has been subtracted). Those for cholesterol and tripalmitin (tripalmitidoylglycerol) were measured for solutions in carbon tetrachloride using an optical path length of 0.5 mm.

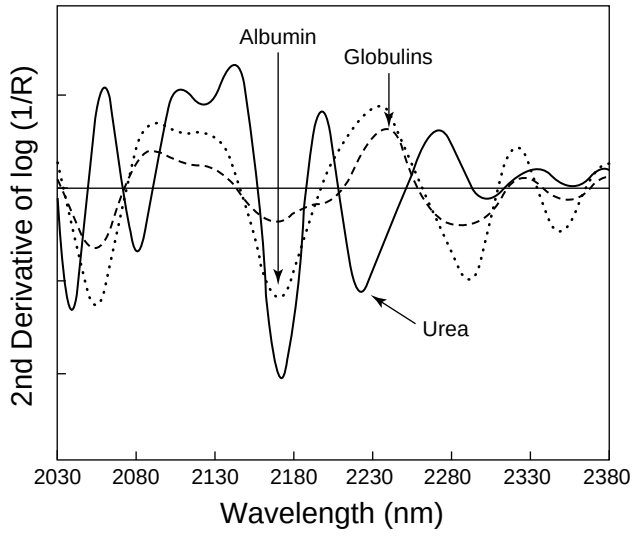

Figure 7 NIR reflectance spectra (second derivatives) for albumin, globulins, and urea. Total serum protein may be quantified by the intensity of the serum absorption at 2064 nm, corresponding to minima (absorption maxima) in the albumin and globulin second-derivative spectra. (Adapted by permission of Elsevier Science from J.W. Hall, A. Pollard, 'Near-infrared Spectroscopic Determination of Serum Total Proteins, Albumin, Globulins, and Urea', *Clinical Biochemistry*, 483–490, Vol. 26, © 1993 by the Canadian Society of Clinical Chemists.)

Equations (7) and (8):

$$C_{\text{albumin}} = 15 - 4419A(2178\,\text{nm}) + 3655A(2206\,\text{nm}) \quad (7)$$

$$C_{\text{total protein}} = 65 - 7821A(2064\,\text{nm}) - 2373A(1440\,\text{nm}) \quad (8)$$

The protein levels predicted by the NIR model are compared to the reference analytical results in Figure 8.

The models for urea and triglyceride quantitation, also based upon the second-derivative spectra, required eight and eleven PLS factors, respectively. The spectral regions employed as the basis for these models differed slightly: optimal for urea quantitation (Figure 8) was a combination of the ranges 1324–1800 and 2304–2370 nm, whereas triglyceride levels were optimally predicted by combining the ranges 1635–1800

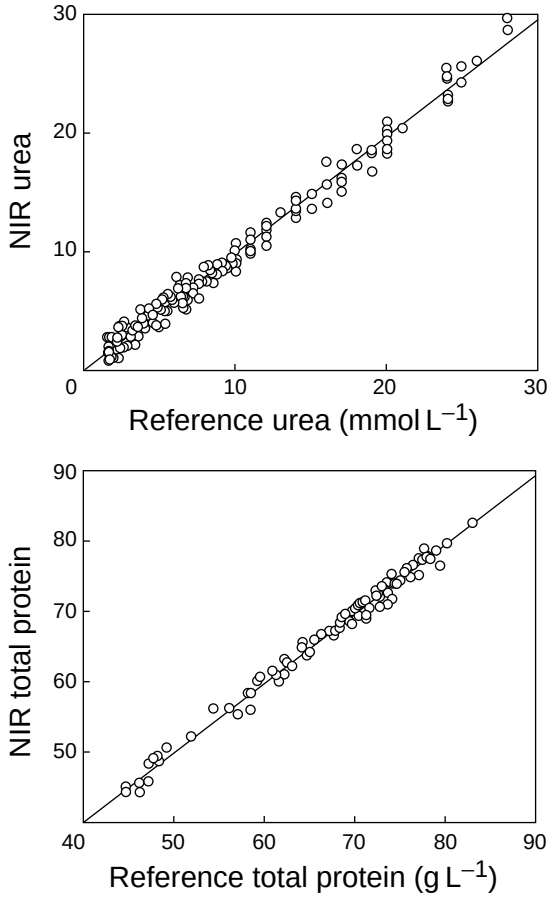

Figure 8 NIR-predicted serum urea and serum protein levels compared to reference analytical results (see also "NIR A" in Table 4). The line of identity is included. (Adapted by permission of Elsevier Science from J.W. Hall, A. Pollard, 'Near-infrared Spectroscopic Determination of Serum Total Proteins, Albumin, Globulins, and Urea', *Clinical Biochemistry*, 483–490, Vol. 26, © 1993 by the Canadian Society of Clinical Chemists.)

and 2035–2375 nm. The 1850–2025 nm range, spanning a very intense water absorption, was explicitly excluded from both models. This provides an example of how PLS modeling benefits from the exclusion of spectral segments that carry no relevant analytical information. These NIR analytical methods for serum total protein, albumin, urea, and triglycerides are summarized as part of the synopsis of methods presented in Table 4.

A more recent NIR investigation was identical in spirit to the inaugural studies but incorporated one substantial change in the experimental protocol.[19] Although the early work was carried out using an optical path length of 0.5 mm, the more recent work used a path length of 2.5 mm. It is quite counter-intuitive to expect improved accuracy at this relatively long path length, because the region richest in solute absorptions (2050–2450 nm) is bordered by two strong water absorptions (see Figure 2). The accessible range within this window is substantially reduced as the optical path length is increased, by virtue of the further encroachment of the shoulders of the two water absorptions. It emerged, however, that this effect was more than compensated for by the enhanced signal-to-noise for solute absorptions in the spectral window that remained accessible.

Figure 9 demonstrates the accuracy of the second NIR study in assays for triglycerides, urea, and cholesterol. Although the analytical methods for total protein and albumin also proved successful, the attempt to quantitate serum lactate proved to be fruitless. The poor results for lactate are largely due to the relatively low serum concentration. Another contributing factor may be that the NIR spectrum is not rich enough to differentiate lactate from other dissolved species (the only NIR bands arise from the methyl group). The analytical methods for the other six analytes are summarized in Table 4. As indicated in Table 4, all methods were based upon PLS models and all made use of the same 2062–2353 nm spectral region.

The NIR quantitation of glucose is of extraordinary interest. This interest stems from the early promise of NIR spectroscopy as a means of monitoring blood glucose levels noninvasively. Indeed, one of the primary aims of the in vitro study[19] was to delineate better the ability of NIR spectroscopy to quantitate serum glucose under ideal experimental conditions. The success of this endeavor is summarized in Figure 10, which superimposes a Clarke error grid[20] on the scatterplot comparing NIR to reference glucose levels. The error grid serves as a template indicating regions corresponding to acceptable

Table 4 Serum analyses using MIR and NIR spectroscopy

Analyte	Method[a]	PLS spectral region(s)	No. of PLS factors	SEP ($mmol\,L^{-1}$)[b]
Glucose	ATR	950–1200 cm^{-1}	9	0.58
	Film MIR	925–1250 cm^{-1}	10	0.41
	NIR B	2062–2353 nm	13	1.3
Triglycerides	ATR	1100–1500, 1700–1800 cm^{-1}	13	0.11
	Film MIR	900–1500, 1700–1800, 2800–3000 cm^{-1}	7	0.23
	NIR A	1635–1800, 2035–2375 nm	8	0.19
	NIR B	2062–2353 nm	13	0.11
Total protein	ATR	1350–1700 cm^{-1}	3	1.4
	Film MIR	900–1800 cm^{-1}	13	2.8
	NIR A	2064, 1440 nm (MLR model)	(2)[c]	1.7
	NIR B	2062–2353 nm	10	2.3
Albumin	Film MIR	1100–1800 cm^{-1}	12	2.2
	NIR A	2178, 2206 nm (MLR model)	(2)[c]	1.1
	NIR B	2062–2353 nm	7	2.0
Cholesterol	ATR	2800–3000 cm^{-1}	8	0.22
	Film MIR	1100–1300, 1700–1800, 2800–3000 cm^{-1}	11	0.28
	NIR B	2062–2353 nm	13	0.32
Urea	ATR	1130–1800 cm^{-1}	20	0.48
	Film MIR	1400–1800 cm^{-1}	13	1.1
	NIR A	1324–1800, 2304–2370 nm	11	0.8
	NIR B	2062–2353 nm	12	0.46

[a] "ATR" = MIR ATR spectroscopy of native serum;[21] "Film MIR" = MIR spectroscopy of dried serum films;[23] "NIR A" = NIR spectroscopy of native serum at 0.5-mm path length;[17,18] "NIR B" = NIR spectroscopy of native serum at 2.5-mm path length.[19]

[b] Standard error of prediction (SEP) for independent test sets except for "ATR" study, where the standard error of cross-validation for the calibration set is given.

[c] Two wavelength terms were required for the albumin and total protein MLR calibration models.

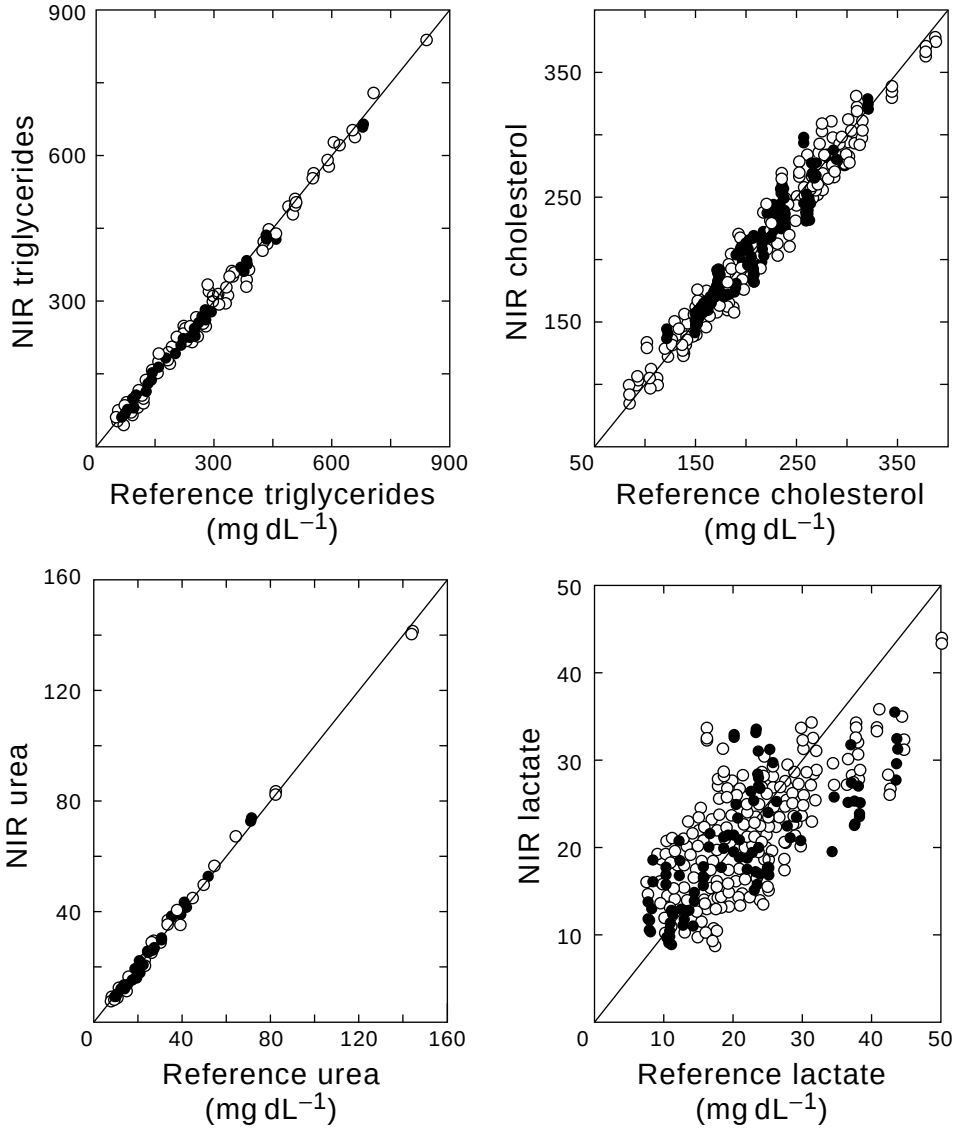

Figure 9 Comparison of NIR-predicted serum analyte levels to reference analytical results (see also ''NIR B'' in Table 4). Open circles correspond to the calibration (training) set, solid circles to the validation (test) set and the solid line is the line of identity. (Adapted from K.H. Hazen, M.A. Arnold, G.W. Small, 'Measurement of Glucose and Other Analytes in Undiluted Human Serum with Near-infrared Transmission Spectroscopy', *Analytica Chimica Acta*, 255–267, Vol. 371, © 1998, with permission from Elsevier Science.)

analytical errors (A, B) and regions corresponding to errors that would lead to dangerous or fatal clinical decisions (C, D, E). As the authors point out, this analytical method is not accurate enough to meet clinical demands but it is accurate enough to suggest that further investigation is warranted.

4.3 Serum Analysis using Mid-infrared Spectroscopy

The two comprehensive feasibility studies have followed two different paths to avoid the problems associated with transmission spectroscopy of the native serum. One of these made use of ATR spectroscopy of the liquid, and the second was based upon transmittance spectroscopy of dried serum films.

4.3.1 Attenuated Total Reflectance Spectroscopy of Native Serum

In this work the investigators sought to quantify glucose, total protein, cholesterol, triglycerides, urea, and uric acid on the basis of the MIR ATR spectra collected using a CIRCLE ATR cell (Spectra-Tech Inc., Shelton, CT, USA).[21] It had been concluded on the basis of

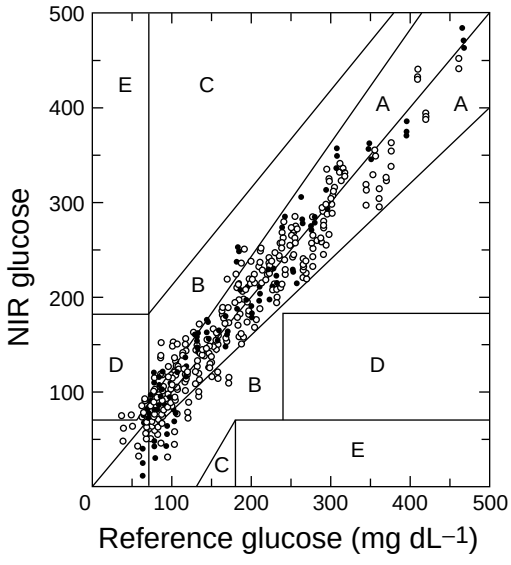

Figure 10 NIR-predicted serum glucose levels vs reference assays (see also "NIR B" in Table 4). Open circles correspond to the calibration (training) set, solid circles to the validation (test) set, and the solid line is the line of identity. A Clarke error grid[30] is superimposed, distinguishing regions corresponding to clinically safe analytical errors (regions A, B) from analytical errors that would result in dangerously inappropriate clinical decisions (C, D, E). (Adapted from K.H. Hazen, M.A. Arnold, G.W. Small, 'Measurement of Glucose and Other Analytes in Undiluted Human Serum with Near-infrared Transmission Spectroscopy', *Analytica Chimica Acta*, 255–267, Vol. 371, © 1998, with permission from Elsevier Science.)

an earlier study[22] that the most critical factor in this procedure is careful cleaning of the zinc selenide ATR optical element between spectral acquisitions. Therefore, the more recent investigation included a cleaning cycle using first a detergent solution, then distilled water, and finally ethanol. The element was then dried by pressurized nitrogen before admitting the next sample. Finally, the specimen was allowed 30 s to reach thermal equilibrium (the cell temperature was kept at $37 \pm 0.02\,^{\circ}C$) before the spectrum was acquired.

The accuracy of this approach is illustrated by the scatterplots in Figure 11. All of the IR-predicted concentrations were based upon PLS models, summarized as part of Table 4. The MIR ATR approach is substantially more accurate than NIR for the quantitation of serum glucose. Why is this the case? The answer relates to the nature of the glucose structure. A rich set of strong absorptions appears in the MIR region of 950–1250 cm^{-1}, corresponding to skeletal C–O stretching vibrations, whereas the NIR spectrum shows only a single absorption in the CH combination region (see Table 1) centered at 2270 nm. The combination bands involving glucose OH groups are very diffuse and are overlapped with the water absorption at 1935 nm to such an extent that they are essentially valueless.

4.3.2 Transmittance Spectroscopy of Dried Serum Films

When a small volume of serum is spread evenly on an IR-transparent window and allowed to dry, the resulting film may be used as the basis to quantitate at least six analytes. A large study based upon this approach used 200 specimens as the basis to develop PLS calibration models for glucose, triglycerides, total protein, albumin, cholesterol, and urea, and an additional 100 specimens to test the accuracy of these models.[23]

A representative spectrum is shown in Figure 4. This spectrum is for a film dried from a 50 : 50 mixture of serum and aqueous potassium thiocyanate solution ($4\,g\,L^{-1}$); the prominent absorption at 2060 cm^{-1} originates with the SCN^- ion. All serum specimens were diluted in this fashion prior to measurement, with the objective of using the absorption intensity at 2060 cm^{-1} to normalize the spectra and hence compensate for possible imprecision in preparation of the films. The specimens were prepared for IR spectroscopy by spreading 7 μL of the diluted serum evenly on the surface of a 13-mm-diameter BaF_2 window. Duplicate samples were prepared in each case, and the corresponding spectra were averaged for PLS analysis.

The PLS trials were preceded by a normalization stage (all spectra were normalized to a common integrated intensity in the SCN^- absorption at 2060 cm^{-1}). The second derivatives of the normalized spectra then served as the basis for the analyses. Scatterplots comparing the reference analytical levels to the IR-predicted albumin, total protein, glucose, cholesterol, triglycerides, and urea are shown in Figure 12. The corresponding PLS models and their analytical accuracies are compiled in Table 4. Attempts to quantitate uric acid and creatinine proved unsuccessful due to their relatively low serum concentrations.

We use the example of glucose to illustrate the procedure that is used to gauge the appropriate number of PLS factors to include in the final model. Recalling that the pool of spectra is divided into a calibration set of 200 samples and a validation set of 100 samples, the aim was to arrive at a final model that optimally extracted the analytical information latent in the calibration spectra. To guard against the possibility of overfitting, the IR-predicted analytical levels were typically compared to reference values for models with 1–15 factors. The trends illustrated in Figure 13 are typical; the standard error of calibration (SEC) in the calibration set specimens decreases rapidly as the initial factors are added, and then tends to plateau as all of the analytically relevant factors were extracted. Additional factors provide rapidly

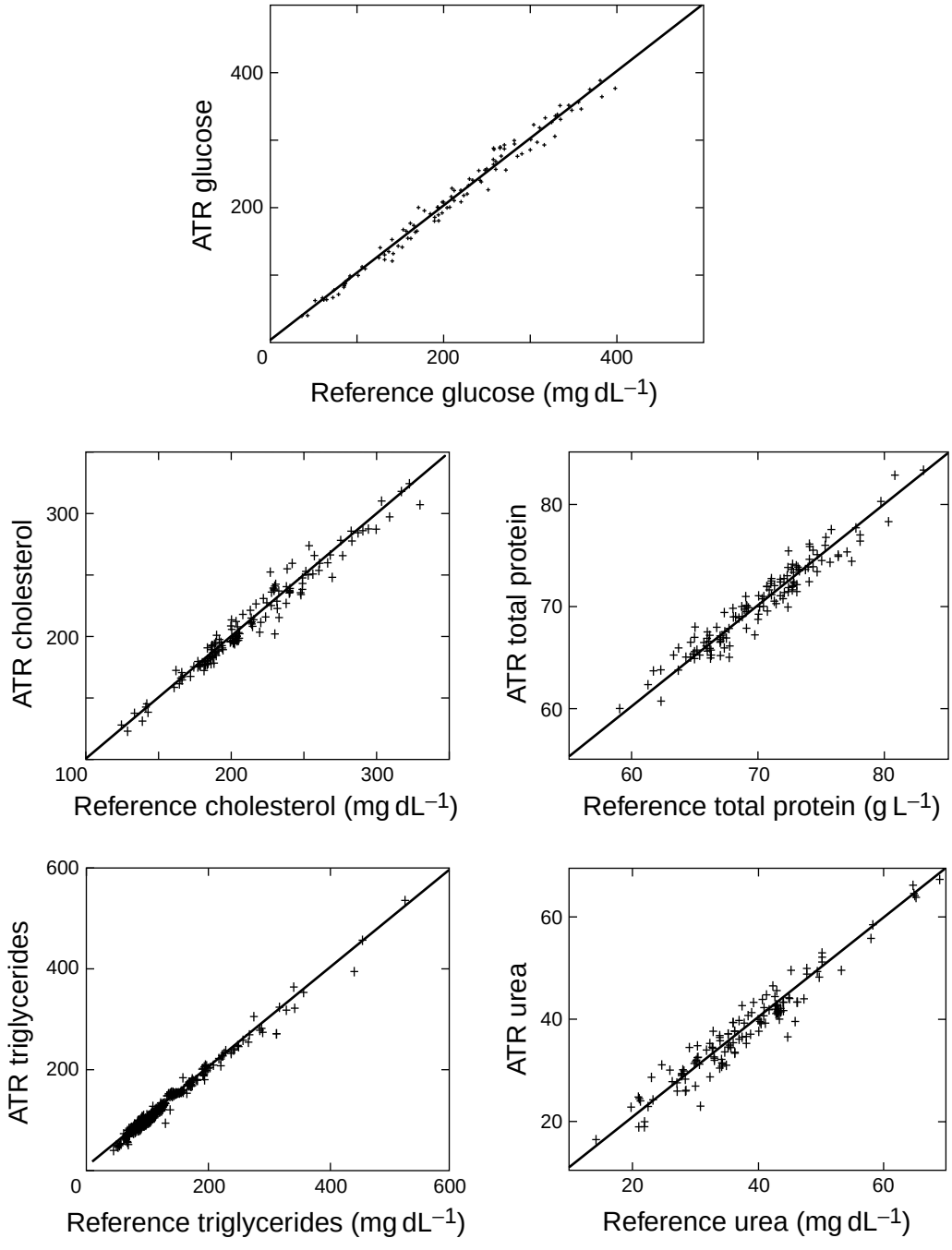

Figure 11 Serum ATR MIR-predicted glucose, total protein, cholesterol, urea, and triglycerides plotted against the gold standard clinical assays, with the line of identity included for reference (see also ''ATR'' in Table 4). (Adapted from Heise et al.[21] by permission of the Society for Applied Spectroscopy.)

diminishing returns, and generally model spectral features that are unique to the set of calibration spectra (i.e. noise). This may be inferred by examining the corresponding trend in the standard error of prediction (SEP) for the validation set. Although the SEC and SEP are essentially identical for all models up to and including 10 factors, this is no longer the case for models including 11 factors or more – although the errors continue to diminish for the calibration set, the opposite trend takes hold for the validation set. The appropriate number of PLS factors corresponds to the point at which the standard error in the validation set begins to increase; this model, corresponding to 10 PLS factors, is equally accurate for the samples in the calibration and validation sets

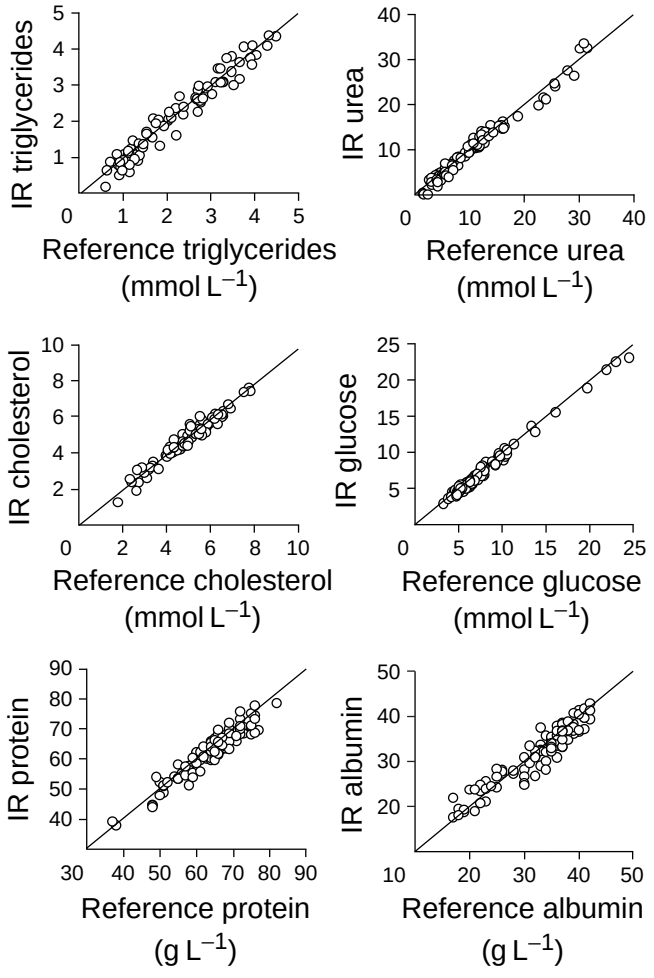

Figure 12 Scatterplots comparing MIR-predicted serum analyte levels to reference analytical results (see also ''Film MIR'' in Table 4). The line of identity is also plotted. The spectra were acquired for dried serum films (see Shaw et al.[23]).

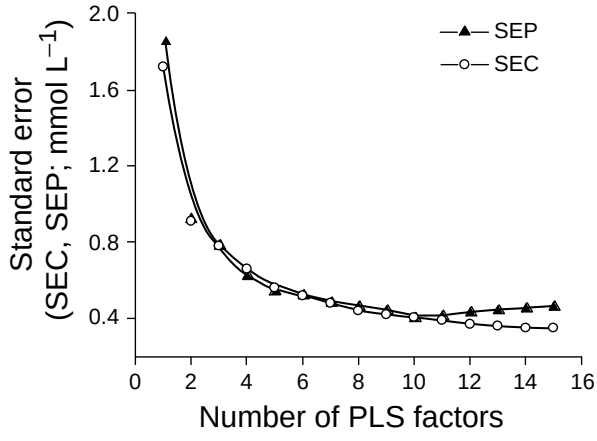

Figure 13 Trends in the SEC and SEP with increasing number of PLS factors for a glucose quantitation model. The 10-factor model was chosen as optimal, based upon these trends (see Shaw et al.[23]).

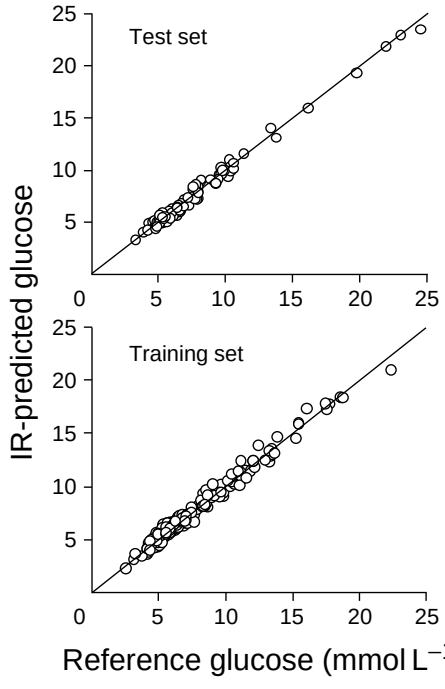

Figure 14 Comparison of the scatterplots comparing MIR-predicted serum glucose levels to reference analytical results for the training (calibration) and test (validation) sets. Results are based upon the spectra of dried serum films using the 10-factor PLS model (see Figure 13 and ''Film MIR'' in Table 4), and the line of identity is superimposed on each plot.

(Figure 14). The serum PLS models determined in this fashion for the dried serum films varied from a 7-factor model for triglycerides to 13 factors for urea and total protein (see Table 4).

Finally, the example of serum glucose provides a good demonstration of a useful feature of PLS modeling. The first PLS weight vector is a least-squares estimate of the spectrum of the analyte of interest; it is a weighted sum of all the calibration spectra, with the weights being the reference concentrations.[12] If this estimate shows similarity to the spectrum of the pure compound, it may be inferred that the PLS model is soundly based in that it is incorporating spectroscopic patterns that originate with that species. To illustrate this, we have evaluated a PLS model for glucose based upon the absorption spectra (not their derivatives) of the dried films. The first PLS weight vector is plotted in Figure 15 together with the absorption spectrum for an aqueous glucose solution. The striking similarities between these two traces provide a good illustration of how this PLS weight vector can be used to support the validity of the PLS model as a whole. At the same time, it should be emphasized that the first weight vector does not always show such a strong resemblance to the IR spectrum of the target analyte. Particularly

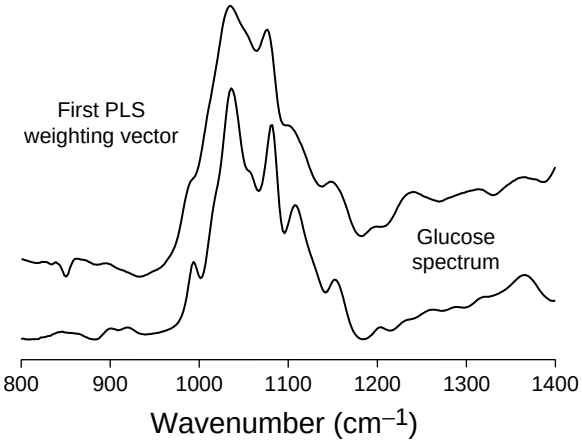

Figure 15 Comparison of the spectrum of an aqueous glucose solution to the first weighting vector for a serum glucose PLS calibration model. Comparisons of this type can provide confirmation, as in this case, that PLS modeling is soundly based upon genuine spectroscopic features of the analyte of interest rather than chance correlations.

for weakly absorbing species of low concentration, the first weight vector – even for a soundly based PLS model – may incorporate compensating features from other components of the mixture that are as large as or larger than features ascribed to the species of interest.

5 SERUM AND BLOOD GLUCOSE

The analysis for glucose is probably the most common blood/serum test. Much of the demand arises from the requirement for frequent self-testing in the diabetic population, and the majority of attempts to quantify glucose using IR spectroscopy have been motivated by the prospect of a noninvasive in vivo test. Because NIR radiation penetrates tissue to depths of millimeters or more, and because an absorption spectrum may be measured for living tissue by using fiber optics or other arrangements, hope has been held out for a NIR method to quantitate blood glucose in vivo.[24,25] Although this subject matter is reviewed elsewhere in this encyclopedia, the various in vitro NIR laboratory studies provide some interesting insights regarding the prospects for in vivo measurement.

In principle, serum or blood glucose may be quantified either by using MIR spectroscopy or by exploiting any of three sets of NIR absorptions, namely those corresponding to vibrational combination bands (2000–2500 nm), the first overtone absorptions (1400–1800 nm), or the second overtone bands (950–1250 nm). All of these have been explored in attempting to quantitate serum glucose,[24] but there are surprisingly few published studies attempting to quantitate glucose in whole blood. Three MIR ATR investigations[26–28] yielded methods with standard errors of 0.8–1.1 mM, whereas the lone report addressing the use of NIR spectroscopy for the measurement of glucose in whole blood reported a SEP of 2.2 mM.[29] The PLS models are summarized in Table 5.

Table 5 IR spectroscopic determinations of glucose in whole blood

Study	Optical path length	Spectral range	No. of PLS factors	SEP (mM)
MIR 1[26]	ATR	750–1500 cm^{-1}	16	1.1
MIR 2[27]	ATR	750–1500 cm^{-1}	ANN[a]	0.9
MIR 3[28]	ATR	950–1200 cm^{-1}	11	0.8
NIR[29]	1 mm	1515–1818, 2062–2353 nm	8	2.1

[a] Artificial neural network model.

How accurate must the glucose analysis be in order to be acceptable clinically? One set of guidelines for clinical testing that has been adopted widely is that of Barnett.[30] Under those guidelines, a new analytical method for blood or serum glucose method should agree with established methods with a maximum standard deviation of 0.28 mM. A more detailed examination of the clinical consequences of inaccurate glucose testing in diabetics has provided the ''Clarke grid''[20] (see Figure 10). Although the scatterplot superimposed on this grid represents a serum NIR analytical method with a standard deviation (SEP) of 1.3 mM relative to reference analyses (substantially larger than the allowable error limit suggested by Barnett), the method clearly approaches the criteria for acceptability set out by Clarke et al. As mentioned earlier in this article, the regions labelled A and B in Figure 10 correspond to clinically acceptable errors. The most serious deficiency of the NIR method is in the analyses for specimens with lower glucose concentrations.

To summarize the present state of the art: the most accurate NIR analysis of serum glucose, carried out using the spectral window and transmission path length optimal for NIR detection, approaches the level of accuracy required for clinical use; and the NIR detection of glucose in blood is less accurate, mainly due to the confounding influence of light scattering by the blood cells.

6 FETAL LUNG MATURITY DETERMINED BY INFRARED SPECTROSCOPY

Among the most common concerns with problematic pregnancies is the possibility that the baby, if born

prematurely, will suffer from respiratory distress syndrome. Failing to produce pulmonary surfactant properly, these infants have traditionally been at high risk of severe respiratory problems or even death. It was recognized in the late 1970s that the fetal lung maturity could be estimated by analyzing for lung surfactants in the amniotic fluid. These tests have been used as the basis for the clinical decision as to whether and when to induce labor, balancing the risk to the mother in continuing the pregnancy with the benefit to the fetus of further lung development within the womb.

The test that has gained widest acceptance is the determination of the amniotic fluid lecithin/sphingomyelin ratio, using thin-layer chromatography (TLC). By its nature, this is a time-consuming and labor-intensive test, and an alternative based upon fluorescence depolarization has been proposed and widely adopted. This procedure measures the ratio of surfactant to protein in amniotic fluid.

Both the lipid and protein constituents provide clear absorptions in the MIR spectra (Figure 16), and both the lecithin/sphingomyelin ratio[31] and the surfactant/protein ratio[32] may be determined from the IR spectra of dry amniotic fluid films. For the lecithin/sphingomyelin ratio determination, the values predicted from the IR spectra (a 14-factor PLS model incorporating the spectral region 2800–3200 cm^{-1}) showed a very good correlation ($r = 0.90$) with the TLC values. Similarly, when surfactant/protein ratios (determined using the Abbott TDx analyzer) were used to calibrate the PLS model, the resulting IR-based analytical method closely reproduced the reference TDx assays (Figure 17).

By their nature, both the TLC and TDx reference methods are inherently less precise than, for example, the common serum assays. For that reason, the scatterplots illustrated in Figure 17 inevitably reflect imprecision in the reference analyses. To confirm that the PLS model

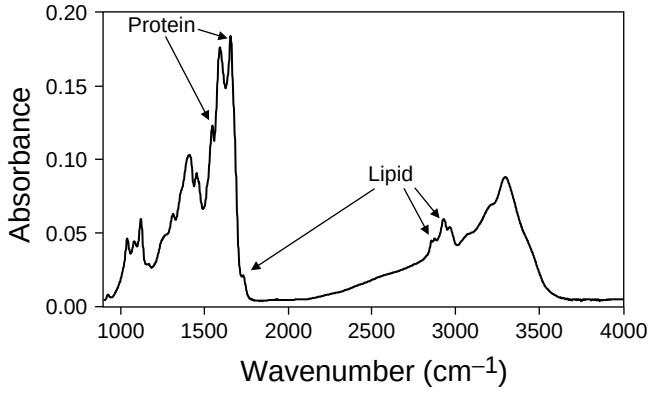

Figure 16 Representative MIR absorption spectrum of a dried amniotic fluid film.

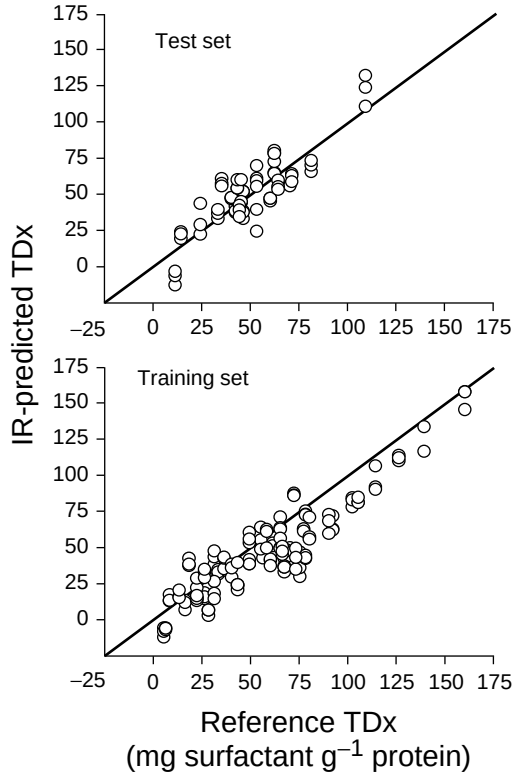

Figure 17 Scatterplots summarizing a PLS model for the determination of amniotic fluid surfactant/protein ratio from the MIR spectra of dried films.[32] The sloping line in each plot is the line of identity.

is based upon the spectral features of surfactant constituents, and is not built upon an accidental correlation with reference analyses, Figure 18 compares the first PLS weight vector to an experimental spectrum. The region plotted in Figure 18 corresponds to CH stretching vibrations of the amniotic fluid constituents, and therefore includes the characteristic vibrations of the long lipid methylene chains as well as absorptions from the proteins. The experimental trace is a difference spectrum, obtained by subtracting the average of all IR spectra corresponding to surfactant/protein ratios less than 55 mg g^{-1} from the average of those with surfactant/protein ratios above that value. The similarity between these two traces confirms that the PLS model is founded upon genuine spectral features originating with the lipid and protein constituents of amniotic fluid.

7 OTHER FLUIDS

7.1 Urine Analysis

Two of the most common analytical tests are for urine creatinine and protein, both of which are key indicators

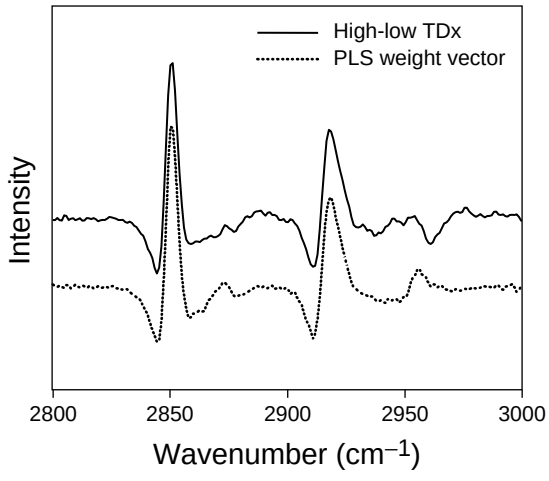

Figure 18 Validation of a PLS model to quantitate the amniotic fluid surfactant/protein ratio from MIR spectra of dried amniotic fluid films. The lower trace is the first PLS weight vector for a MIR quantitation model calibrated with reference to laboratory TDx measurements of the surfactant/protein ratio. The upper trace is a difference spectrum, obtained by subtracting the average of all IR spectra corresponding to surfactant/protein ratios of less than $55\,mg\,g^{-1}$ from the average of those with surfactant/protein ratios above that value. The $2800-3000\,cm^{-1}$ region proved nearly optimal for PLS calibration, with only marginal improvements gained by the addition of segments in the $900-1800\,cm^{-1}$ region (see Figure 16).

Table 6 Selected urine analytes that may be determined using IR spectroscopy

Analyte	Reference intervals[a]
Urea	$12-20\,g\,24\,h^{-1}$ ($428-714\,mmol\,24\,h^{-1}$)
Creatinine	$1-1.6\,g\,24\,h^{-1}$ ($8.8-14.2\,mmol\,24\,h^{-1}$)
Total protein	$<0.1\,g\,24\,h^{-1}$

[a] From Wallach[15] and Tietz.[16] Diagnostic information is generally inferred from totals passed in urine over a 24-h period. Concentrations for the NIR study reported in Shaw et al.[6] were grouped in the ranges $100-400\,mmol\,L^{-1}$ (urea), $2.5-12.5\,mmol\,L^{-1}$ (creatinine) and $0-3\,g\,L^{-1}$ (protein).

of kidney function (Table 6). Urea is also measured for the same reason, although less frequently. Because the normal physiological concentration range is quite high for urea, intermediate for creatinine, and very low in the case of protein, each of these three analytes posed unique challenges in developing analytical methods based upon the NIR spectra of the native fluid.[6]

As discussed earlier, a simple MLR model sufficed to quantitate urea across the physiological range. A 9-factor PLS model was required to determine creatinine with a SEP of $0.79\,mmol\,L^{-1}$. The best that could be achieved for protein was an 8-factor PLS model with a SEP of $0.23\,g\,L^{-1}$. This model is clearly insufficient to provide the information that is sought most often clinically, because a protein level above $\sim 0.1\,g\,L^{-1}$ is considered to be a warning sign of possible kidney malfunction. In an effort to improve the accuracy of the IR-based method for samples with low protein concentrations, a second PLS model was derived using only those specimens with reference protein concentrations of less than $1\,g\,L^{-1}$. Although the result was an improved accuracy in quantitating protein in the range $0-1\,g\,L^{-1}$, the SEP of $0.12\,g\,L^{-1}$ remained unacceptably high. With a detection limit of $0.36\,g\,L^{-1}$ (taken as three times the standard error), many urine specimens have protein levels below the threshold that is detectable by NIR spectroscopy.

7.2 Saliva

Because it is so readily available, saliva has often been considered as a potential source of diagnostic information.[33] The MIR spectrum of the dried film (Figure 19) reveals not only the protein constituents but also thiocyanate (SCN^-). Although it is somewhat surprising to the layman to learn that this ion is present in appreciable amounts in human saliva, it plays a functional role; enzymatic conversion yields salivary hypothiocyanate ($OSCN^-$), which is a highly effective antibacterial agent.

The SCN^- ion shows an absorption in a spectral region that is typically devoid of any other bands. As a result, it proved possible to quantitate this ion in saliva through a simple Beer's law relationship according to

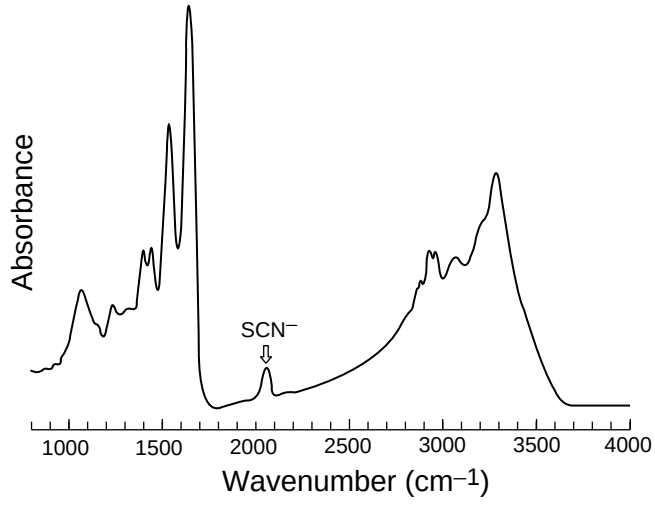

Figure 19 Representative MIR spectrum of a dried saliva film. The absorption at $2060\,cm^{-1}$ is from endogenous thiocyanate (SCN^-).[34]

Equation (9):[34]

$$C_{SCN^-} = -0.001 + 0.65A(2058\,cm^{-1}) \quad (9)$$

where $A(2058\,cm^{-1})$ denotes the integrated intensity for absorption at $2058\,cm^{-1}$ for a 5-μL saliva aliquot dried to a spot of diameter 4 mm. The same study indicated further that saliva thiocyanate concentration levels may follow a circadian rhythm, with maximum levels early in the day.

8 DISEASE DIAGNOSIS BASED ON INFRARED SPECTRAL PATTERN RECOGNITION

Although the analysis of biological fluids has a long tradition in providing information to suggest or corroborate diagnosis, a complementary technique is emerging for the interpretation of the IR spectra. Rather than deriving analyte levels explicitly from them, the spectra may be viewed as fingerprints that correlate directly with the presence or absence of disease. Because the spectra are complex, patterns characteristic of specific diseases are rarely (if ever) discernable from visual examination of the spectra. However, multivariate analytical methods may identify subtle patterns distinguishing the spectra corresponding to ''normal'' specimens from those corresponding to diseased patients.

The general procedure for developing this diagnostic test has much in common with the techniques employed to develop IR-based analytical methods. The first step is to acquire appropriate specimens from two sets of donors. One set of normal or control samples is required, whereas the second set corresponds to patients who have been diagnosed by traditional methods as having the disease of interest. The corresponding IR spectra are collected and subjected to two interlinked procedures: feature extraction and classification. The feature extraction procedure identifies characteristics that distinguish the normal spectra from diseased, whereas the aim of the classification stage is to separate optimally the two groups of spectra based upon those characteristics. Finally, the general applicability of the optimal classifier is tested by predicting the class assignments (normal or diseased) for a separate group of test spectra and comparing these a posteriori to the ''gold standard'' diagnoses.

8.1 Arthritis Diagnosis from Infrared Spectroscopy of Synovial Fluid

Synovial fluid is an ultrafiltrate of blood plasma that serves to transport nutrients to cartilage as well as to lubricate the joints. It is readily available for diagnostic testing (the fluid is commonly drained from joints that are inflamed, either as a result of disease or physical trauma), however there is no synovial fluid analytical test or combination of tests to reliably diagnose arthritis or to distinguish arthritic conditions from one another.

Two studies have suggested that the IR spectra of synovial fluid specimens provide the basis to diagnose arthritis and to differentiate among its variants.[35,36] A NIR study demonstrated that osteoarthritis, rheumatoid arthritis, and spondyloarthropathy could be distinguished on the basis of the synovial fluid absorption patterns in the range 2000–2400 nm.[35] In that case, the pool of synovial fluid spectra was subject to principal component analysis, and eight principal component scores for each spectrum were employed as the basis for linear discriminant analysis (LDA). On that basis, the optimal LDA classifier matched 105 of the 109 spectra to the correct clinical designation (see Table 7).

An investigation based upon the MIR spectra of dried synovial fluid films showed similar success in distinguishing spondyloarthropathy, rheumatoid arthritis, and osteoarthritis from one another and from control specimens (generally synovial fluid aspirates from individuals with injuries rather than diseased joints).[36] This study is of interest in that it made use of a genetic region selection algorithm[8] to identify a set of 15 discrete spectral subregions differentiating the 4 classes of spectra. With each spectrum represented by a set of 15 regional intensities, LDA provided the basis for successful classification.

Although applications of this type are not analytical in the traditional sense, they may provide analytical information indirectly. The successful classification of IR spectra according to disease type implies that the composition of the specimen is altered in a characteristic fashion with the onset of disease – in this case the synovial fluid make-up reflects the presence and type of arthritis. This is an intriguing finding, particularly as it is very unlikely that the IR spectra are detecting any particular constituent that cannot be (and has not been) assayed

Table 7 Classification table: clinical versus IR-based arthritis diagnoses

		IR-based diagnoses[a]		
		SA	RA	OA
Clinical diagnoses	SA	12	3	0
	RA	0	65	0
	OA	0	2	27

[a] The IR-based diagnoses are from principal component analysis and LDA of synovial fluid NIR spectra (see text): SA = spondyloarthropathy; RA = rheumatoid arthritis; OA = osteoarthritis. The table indicates, for example, that of the 29 spectra corresponding to patients with osteoarthritis, 27 were classified correctly but two were misclassified as rheumatoid arthritis. (see Shaw et al.[35]).

previously. The strength of this approach therefore lies not in the ability of IR spectroscopy to identify novel disease markers, but rather in the ability of multivariate pattern recognition and classification methods to perceive characteristic changes in the balance among the IR-detectable components.

8.2 Disease Pattern Recognition in Mid-infrared Spectra of Serum

The premise underlying this research is that the combination of serum MIR spectroscopy with pattern recognition methods can distinguish healthy subjects from those with specific disease types. One report has demonstrated success rates of better than 90% in distinguishing diabetics from healthy subjects, type I from type II diabetics, and patients with rheumatoid arthritis from healthy subjects.[37] Again, this work is intriguing because it is very unlikely that the IR spectra reveal any fundamentally new serum constituents as ''disease markers''. Although the successful classifications indicate that the presence of disease is linked to specific relationships among specific spectral features, the challenge remains to interpret the physiological significance of those features. As research progresses in this area, we may anticipate that these characteristic relationships among spectral features will be interpreted to reveal characteristic relationships among serum metabolite levels distinguishing healthy from diseased donors.

9 SUMMARY

This article has provided an overview of the common clinical analyses that may, in principle, be carried out by using IR spectroscopy, as well as pointing out some novel diagnostic tests that have emerged by combining IR spectroscopy with multivariate pattern recognition methods. In the case of analysis, it is clear that IR spectroscopy can meet the standards of accuracy that are required for a number of standard clinical serum and urine tests. More specialized tests such as amniotic fluid assays to assess fetal lung maturity may also be carried out, and other less common analyses, such as cerebrospinal fluid and interstitial fluid tests, can certainly be envisaged as being suited for IR spectroscopy.

IR spectroscopic analysis offers a number of practical advantages that make it a natural fit for high-volume applications of the type epitomized by serum and urine analysis, as well as for labor-intensive tests such as the amniotic fluid lecithin/sphingomyelin ratio. No reagents are required, there is generally no need to dilute very concentrated specimens (as may be required for certain other analytical methods), several analyses are available simultaneously from a single IR spectrum, and – at least in the method using dried films – very little sample (microliters) is required. Time will tell whether these advantages will be exploited through the development of dedicated, IR-based clinical analyzers.

ACKNOWLEDGMENTS

Ms Sarah Low Ying is gratefully acknowledged for her assistance in the preparation of this article.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflectance
FIR	Far-infrared
IR	Infrared
LDA	Linear Discriminant Analysis
MIR	Mid-infrared
MLR	Multiple-wavelength Linear Regression
NIR	Near-infrared
PCR	Principal Component Regression
PLS	Partial Least Squares
SEC	Standard Error of Calibration
SEP	Standard Error of Prediction
TLC	Thin-layer Chromatography

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Glucose, In Vivo Assay of

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis • Infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics

Infrared Spectroscopy **(Volume 12)**
Quantitative Analysis, Infrared

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. W. Kaye, 'Near-infrared Spectroscopy I. Spectral Identification and Analytical Applications', *Spectrochim. Acta*, **6**, 257–287 (1954).
2. K.H. Norris, 'History of NIR', *J. Near Infrared Spectrosc.*, **4**, 31–37 (1996).
3. P.C. Williams, 'Application of Near Infrared Reflectance Spectroscopy to Analysis of Cereal Grains and Oilseeds', *Cereal Chem.*, **52**, 561–576 (1975).
4. N.J. Harrick, *Internal Reflection Spectroscopy*, Interscience Publishers, New York, 1967.
5. F.M. Mirabella, *Internal Reflection Spectroscopy: Review and Supplement*, ed. N.J. Harrick, Harrick Scientific, Ossining, NY, 1985.
6. R.A. Shaw, S. Kotowich, H.H. Mantsch, M. Leroux, 'Quantitation of Protein, Creatinine, and Urea in Urine by Near-infrared Spectroscopy', *Clin. Biochem.*, **29**, 11–19 (1996).
7. D.E. Honigs, J.M. Freelin, G.M. Hieftje, T.B. Hirschfeld, 'Near-infrared Reflectance Analysis by Gauss–Jordan Linear Algebra', *Appl. Spectrosc.*, **37**, 491–497 (1983).
8. A. Nikulin, B. Dolenko, T. Bezebah, R.L. Somorjai, 'Near-optimal Region Selection for Feature Space Reduction: Novel Preprocessing Methods for Classifying MR Spectra', *NMR Biomed.*, **11**, 209–216 (1998).
9. P.M. Fredericks, J.B. Lee, P.R. Osborn, D.A.J. Swinkels, 'Materials Characterization using Factor Analysis of FT-IR Spectra. Part 1: Results', *Appl. Spectrosc.*, **39**, 303–310 (1985).
10. P.M. Fredericks, J.B. Lee, P.R. Osborn, D.A.J. Swinkels, 'Materials Characterization using Factor Analysis of FT-IR Spectra. Part 2: Mathematical and Statistical Considerations', *Appl. Spectrosc.*, **39**, 311–316 (1985).
11. K.R. Beebe, B.R. Kowalski, 'An Introduction to Multivariate Calibration and Analysis', *Anal. Chem.*, **59**, 1007A–1017A (1987).
12. D.A. Haaland, E.V. Thomas, 'Partial Least-squares Methods for Spectral Analysis. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information', *Anal. Chem.*, **60**, 1193–1203 (1988).
13. D.A. Haaland, E.V. Thomas, 'Partial Least-squares Methods for Spectral Analysis. 2. Application to Simulated and Glass Spectral Data', *Anal. Chem.*, **60**, 1203–1209 (1988).
14. H. Martens, T. Næs, *Multivariate Calibration*, John Wiley & Sons, Chichester, 1991.
15. J. Wallach, *Interpretation of Diagnostic Tests. A Synopsis of Laboratory Medicine*, 5th edition, Little, Brown and Company, London, 1992.
16. C.A. Burtis, E.R. Ashwood (eds.), 'Appendix' in *Tietz. Fundamentals of Clinical Chemistry*, WB Saunders, Philadelphia, 1996.
17. J.W. Hall, A. Pollard, 'Near-infrared Spectrophotometry: a New Dimension in Clinical Chemistry', *Clin. Chem.*, **38**, 1623–1631 (1992).
18. J.W. Hall, A. Pollard, 'Near-infrared Spectroscopic Determination of Serum Total Proteins, Albumin, Globulins, and Urea', *Clin. Biochem.*, **26**, 483–490 (1993).
19. K.H. Hazen, M.A. Arnold, G.W. Small, 'Measurement of Glucose and Other Analytes in Undiluted Human Serum with Near-infrared Transmission Spectroscopy', *Anal. Chim. Acta*, **371**, 255–267 (1998).
20. W.L. Clarke, D. Cox, L.A. Gonder-Frederick, W. Carter, S.L. Pohl, 'Evaluating Clinical Accuracy of Systems for Self-monitoring of Blood Glucose', *Diabetes Care*, **10**, 622–628 (1987).
21. H.M. Heise, R. Marbach, Th. Koschinsky, F.A. Gries, 'Multicomponent Assay for Blood Substrates in Human Plasma by Mid-infrared Spectroscopy and its Evaluation for Clinical Analysis', *Appl. Spectrosc.*, **48**, 85–95 (1994).
22. G. Janatsch, J.D. Kruse-Jarres, R. Marbach, H.M. Heise, 'Multivariate Calibration for Assays in Clinical Chemistry using Attenuated Total Reflection Infrared Spectra of Human Blood Plasma', *Anal. Chem.*, **61**, 2016–2023 (1989).
23. R.A. Shaw, S. Kotowich, M. Leroux, H.H. Mantsch, 'Multianalyte Serum Analysis using Mid-infrared Spectroscopy', *Ann. Clin. Biochem.*, **35**, 624–632 (1998).
24. M.A. Arnold, 'Non-invasive Glucose Monitoring', *Curr. Opin. Biotechnol.*, **7**, 46–49 (1996).
25. O.S. Khalil, 'Spectroscopic and Clinical Aspects of Noninvasive Glucose Measurements', *Clin. Chem.*, **45**, 165–177 (1999).
26. H. Heise, R. Marbach, G. Janatsch, J.D. Kruse-Jarres, 'Multivariate Determination of Glucose in Whole Blood by Attenuated Total Reflection Infrared Spectroscopy', *Anal. Chem.*, **61**, 2009–2015 (1989).
27. P. Bhandare, Y. Mendelson, R.A. Peura, G. Janatsch, J. Kruse-Jarres, R. Marbach, H.M. Heise, 'Multivariate Determination of Glucose in Whole Blood using Partial-least-squares and Artificial Neural Networks Based on Mid-infrared Spectroscopy', *Appl. Spectrosc.*, **8**, 1214–1221 (1993).
28. K.J. Ward, D.M. Haaland, M.R. Robinson, R.P. Eaton, 'Post-prandial Blood Glucose Determination by Quantitative Mid-infrared Spectroscopy', *Appl. Spectrosc.*, **46**, 959–965 (1992).
29. D.M. Haaland, M.R. Robinson, G.W. Koepp, E.V. Thomas, R.P. Eaton, 'Reagentless Near-infrared Determination of Glucose in Whole Blood using Multivariate Calibration', *Appl. Spectrosc.*, **46**, 1575–1578 (1992).
30. R.N. Barnett, 'Medical Significance of Laboratory Results', *Am. J. Clin. Pathol.*, **50**, 671–676 (1968).
31. K.Z. Liu, T.C. Dembinski, H.H. Mantsch, 'Prediction of RDS from Amniotic Fluid Analysis: a Comparison of the Prognostic Value of TLC and Infrared Spectroscopy', *Prenatal Diagn.*, **18**, 1267–1275 (1998).
32. K.Z. Liu, R.A. Shaw, T.C. Dembinski, G.J. Reid, S. Low Ying, H.H. Mantsch, 'A Comparison of the Accuracy of the Infrared Spectroscopy and TDx-FLM Assays in the Estimation of Fetal Lung Maturity', (in press).

33. D. Malamud, L. Tabak, 'Saliva as a Diagnostic Fluid', *Ann. NY Acad. Sci.*, **694**, 1–348 (1993).
34. C.P. Schultz, M.K. Ahmed, C. Dawes, H.H. Mantsch, 'Thiocyanate Levels in Human Saliva: Quantitation by Fourier Transform Infrared Spectroscopy', *Anal. Biochem.*, **240**, 7–12 (1996).
35. R.A. Shaw, S. Kotowich, H.H. Eysel, M. Jackson, G.T.D. Thomson, H.H. Mantsch, 'Arthritis Diagnosis Based upon the Near-infrared Spectrum of Synovial Fluid', *Rheumatol. Int.*, **15**, 159–165 (1995).
36. H.H. Eysel, M. Jackson, A. Nikulin, R.L. Somorjai, G.T.D. Thomson, H.H. Mantsch, 'A Novel Diagnostic Test for Arthritis: Multivariate Analysis of Infrared Spectra of Synovial Fluid', *Biospectroscopy*, **3**, 161–167 (1997).
37. G.H. Werner, J. Früh, F. Keller, H. Greger, R. Somorjai, B. Dolenko, M. Otto, D. Böcker, 'Mid Infrared Spectroscopy as a Tool for Disease Pattern Recognition from Human Blood', *Proc. SPIE*, **3257**, 35–41 (1998).

Infrared Spectroscopy in Microbiology

Dieter Naumann
Robert Koch-Institute, Berlin, Germany

1 Introduction and History 1
1.1 Objectives 2
1.2 Advantages and Disadvantages of the Method 2
1.3 Infrared Spectroscopy 3
1.4 Fourier Transform Infrared Spectroscopy Technology: the Fourier Transform Infrared Spectrometer 3
2 Chemical Structures and Composition of Microbial Cells 5
2.1 Bacterial Cells 5
2.2 Yeast Cells 6
3 Main Biological Building Blocks and Assignment of Infrared Bands 7
4 Experimental Methodologies 10
4.1 Infrared Measurement Techniques for Microbial Samples 10
4.2 Sampling of Microbial Cells and Data Acquisition 12
4.3 Variability of the Microbial Specimen and the Problem of Reproducibility 13
5 Data Treatment and Evaluation Techniques 13
5.1 Factor Analysis 14
5.2 Multivariate Clustering 15
5.3 Artificial Neural Networks 17
6 Applications 17
6.1 Characterization of Particular Cell Compounds and Structures 17
6.2 Differentiation, Classification, and Identification of Microorganisms by Infrared Spectroscopy 19
6.3 Sensitivity Testing by Infrared Spectroscopy 23
6.4 Detection, Enumeration, and Identification of Microcolonies 25
7 Perspectives 27
Acknowledgments 27
Abbreviations and Acronyms 27
Related Articles 28
References 28

Infrared (IR) spectra of intact microbial cells are highly specific fingerprint-like signatures which are used to differentiate, classify, and identify diverse microbial species and strains. Microbial IR spectra are also useful to (1) detect in situ intracellular compounds or structures such as inclusion bodies, storage materials, and endospores, (2) monitor and quantify metabolically released CO_2 in response to various different substrates, and (3) characterize growth-dependent phenomena and cell–drug interactions. The characteristic information, useful for microbial characterizations, is generally distributed over the entire IR region of the electromagnetic spectrum, i.e. over the near-infrared (NIR), mid-infrared (MIR), and far-infrared (FIR). The spectral traits can be systematically extracted from the typically broad and complex spectral contours applying resolution enhancement techniques, difference spectroscopy, and pattern recognition methods such as factor-analysis and cluster-analysis, and artificial neural networks (ANNs). Additional applications arise by means of a light microscope coupled to the IR spectrometer. IR spectra of microcolonies containing less than 10^3 cells are obtained from colony replica by a stamping technique that transfers microcolonies growing on culture plates to a special IR-sample holder. Using a computer-controlled x,y-stage together with mapping and video techniques, the fundamental tasks of microbiological analysis, namely detection, enumeration, and differentiation of microorganisms, are perspectively integrated in one single apparatus.

1 INTRODUCTION AND HISTORY

The 1980s and 1990s have witnessed the emergence of sensitive, rapid and increasingly precise physical techniques for microbiological analysis. These new techniques range from mass spectroscopy (MS), molecular spectroscopy (including fluorescence, Fourier transform infrared (FTIR) and Raman spectroscopy), the application of laser technologies, and flow cytometry, to different separation techniques like gas chromatography and high-performance liquid chromatography.[1] This development is paralleled by the dynamic advance of molecular genetic techniques in microbiology, which presumably will develop into most sensitive and specific tools for microbial characterizations in the future.[2]

In 1911 W.W. Coblentz was probably the first scientist to suggest that biological materials can profitably be analyzed by means of IR spectroscopy. The use of IR spectroscopy as a means of differentiating between and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

identifying bacteria was extensively reported as early as in the 1950s and 1960s.[3,4] A critical review on this subject published in 1959 summarized that, although bacteria definitely exhibit IR spectra that are unique for individual strains, the identification of bacteria via IR techniques cannot be regarded as a practical scheme because it is too time-consuming and impractical a procedure.[3] Because of the limited specifications of IR spectrometers at that time (sensitivity, time, and reproducibility), reports on IR applications to microorganisms became less frequent in the 1960s and had virtually ceased by the mid-1970s. The development of modern interferometric IR spectroscopy, the availability of low-cost minicomputers and powerful new algorithms of multivariate statistical analysis (MSA) and pattern recognition methodologies such as factor analysis and ANNs have contributed to the revival of IR spectroscopy as a means of characterizing microbial samples.

1.1 Objectives

This article describes details of recent technological developments, data elaboration methodologies, and representative examples of applications in this field. Only results that are concerned with samples of intact cells will be presented. No attempt will be made to comment on the use of IR spectroscopy for the analysis of isolated cell fragments, compounds, macromolecules, or cell metabolites.

1.2 Advantages and Disadvantages of the Method

Among the most important aspects of IR spectroscopy for investigating intact cells of microorganisms the following are noteworthy:

- Advantages:
  1. The method is uniformly applicable to virtually all microorganisms which can be grown in culture. Biomass requirements can be scaled down to single colonies taken directly from the culture plates. With the use of IR microscopes even microcolonies as small as 20 μm in diameter corresponding to a few hundred cells can be analyzed. Results are available within minutes after obtaining adequate samples of the pure culture.
  2. Detection, enumeration, classification, and identification can be integrated in one single instrument when using the IR microscope. Results are available within one working day in a clinical setting, including isolation, cultivation and identification.
  3. IR spectroscopy can classify microorganisms at very different levels of taxonomic discrimination without any preselection of strains by other taxonomic criteria. In contrast to most other techniques, IR spectroscopy is useful at the strain, species and genus level. The specificity of the method is generally extremely high allowing differentiations at the strain and/or serogroup/serotype level.
  4. These advantages make the method applicable to (i) very rapid identification of life-threatening pathogens; (ii) epidemiological investigations, conductance of case studies, screening of pathogens, hygiene control, elucidation of infection chains, therapy control, and detection of recurrent infections; (iii) characterization and screening of microorganisms from the environment, biotechnological process control, microbiological quality control in the food and pharmaceutical industries, and maintenance of strain collections.
  5. In a number of cases, in situ detection of specific cell components is possible (e.g. storage materials, spore formation, encapsulation of microorganisms) and drug resistance and cell–drug interaction can be monitored and characterized.
- Notable disadvantages:
  1. Only microorganisms that can be grown in culture and are available as pure cultures can be analyzed. Mixed cultures can only be investigated with the help of the IR microscope provided single colony growth is obtained.
  2. Classification and identification are based on the analysis of spectral fingerprints. Specific information on one or only a few specific compounds present in the cells is generally not available.
  3. Since the technique provides a spectral fingerprint of all cell constituents, stable results can only be obtained provided the microbiological parameters (culture medium, cultivation time, and temperature) can be controlled rigidly. Spectral databases that are used in different laboratories require the use of appropriate culture media available on the market.
  4. Acceptance of the IR technique by the microbiological community presupposes that spectral information can be transformed to classification schemes that are at least similar to those obtained by chemotaxonomic or by

molecular genetic techniques. This task can only be fulfilled by trained people in expert laboratories.

5. While the strength of the technique is its ability to differentiate microorganisms very rapidly below the species level, classifications at the genus level cannot be expected to be taxonomically relevant in all cases.

1.3 Infrared Spectroscopy

The IR region of the spectrum extends from the visible region until it overlaps the microwave, or very short radar, range at wavelengths of some millimeters (see Figure 1). The basic characteristic of this region is that the radiation originates from thermal emission from a hot source. Per convention throughout the IR region the ''wavenumber'', that is the number of waves per centimeter, is used to characterize the radiation. The wavenumber unit is the reciprocal centimeter (cm^{-1}) and, hence, one is dealing with the convenient values of $10\,000\,cm^{-1}$ to $10\,cm^{-1}$ throughout the IR region. Also per convention, the IR region of the electromagnetic spectrum is subdivided into the NIR, MIR, and FIR regions (Figure 1). An IR spectrum of a sample is obtained by scanning the intensity of IR radiation before and after passage of the IR beam through the sample. The IR spectrum is displayed by plotting the quantity $T = I_S/I_R$ as a function of wavenumbers, where T is the transmittance, I_S the intensity of the IR beam after and I_R before passing through the sample. In most cases the absorbance A is used ($A = -\log T$), since the absorbance at a given wavelength is directly proportional to the concentration of a sample according to Beer's law. The IR spectra of most materials consist of a large number of absorption bands. These bands originate from the interaction (energy exchange) between discrete light quanta and mechanical motions (vibrational and rotational modes) of the molecules which are excited by the absorption of IR radiation. Since the

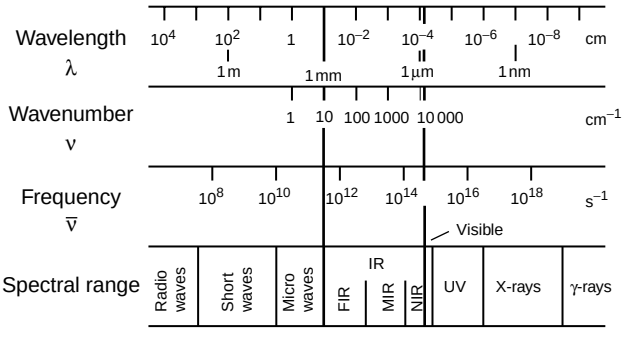

Figure 1 The electromagnetic spectrum. (Adapted from Naumann et al.[6])

constituents of a typical biological sample are present in a condensed phase (solids, liquids or solutions), only vibrational modes are observed. Consequently, IR spectra of biological specimens are only vibrational spectra.[5]

1.4 Fourier Transform Infrared Spectroscopy Technology: the Fourier Transform Infrared Spectrometer

With the advent of FTIR spectroscopy, major drawbacks of ''classical'' dispersive IR spectroscopy could be circumvented. Briefly, FTIR no longer measures one wavelength after the other and is not like dispersive spectrometers which are equipped with prism- and/or grating-monochromators where the light energy emanating from the IR source is strongly limited by several entrance and exit slits. FTIR, instead, applies interferometric modulation of radiation (see Figure 2). As in dispersive instruments, a hot white-light source (globar) is used in the MIR (block scheme, Figure 2a). A widespread type of interferometer used in FTIR is the so-called Michelson interferometer mounted with a KBr beam splitter coated with a germanium film to split the collimated IR beam into two parts (Figure 2b, upper panel). For rapid-scanning interferometers liquid nitrogen cooled mercury cadmium telluride (MCT) detectors are used. For slower scanning types of interferometer, pyroelectric detectors (e.g. a deuterated triglycine sulfate (DTGS) detector element) can be used. The moving mirror of the interferometer (Figure 2b, upper panel) yields a sinusoidal signal at the detector for each modulated frequency (see Figure 2b, central panel). For a white-light source all modulated signals superimpose to an interferogram as shown in Figure 2b, lower panel. This interferogram is amplified and digitized by the A/D converter (Figure 2a), computer stored, and finally transformed to a spectrum by fast Fourier transform (FT) techniques. Thus, the auxiliary computer in case of FTIR is a virtual necessity.[7] The interferometer can be thought of as a means of encoding the initial frequencies into a special form which the detector can observe. The most important feature of the interferogram is that every individual data point of this signal contains information over the entire IR region. In essence, the detector is always observing all frequencies at the same time. The Fourier transformation is simply a mathematical means of sorting out the individual frequencies for the final representation of an IR spectrum. This, along with some other features of the interferometer, leads to several distinct advantages:[7]

1. The Jaquinot Advantage. There are no longer slits in the interferometers that limit the amount of

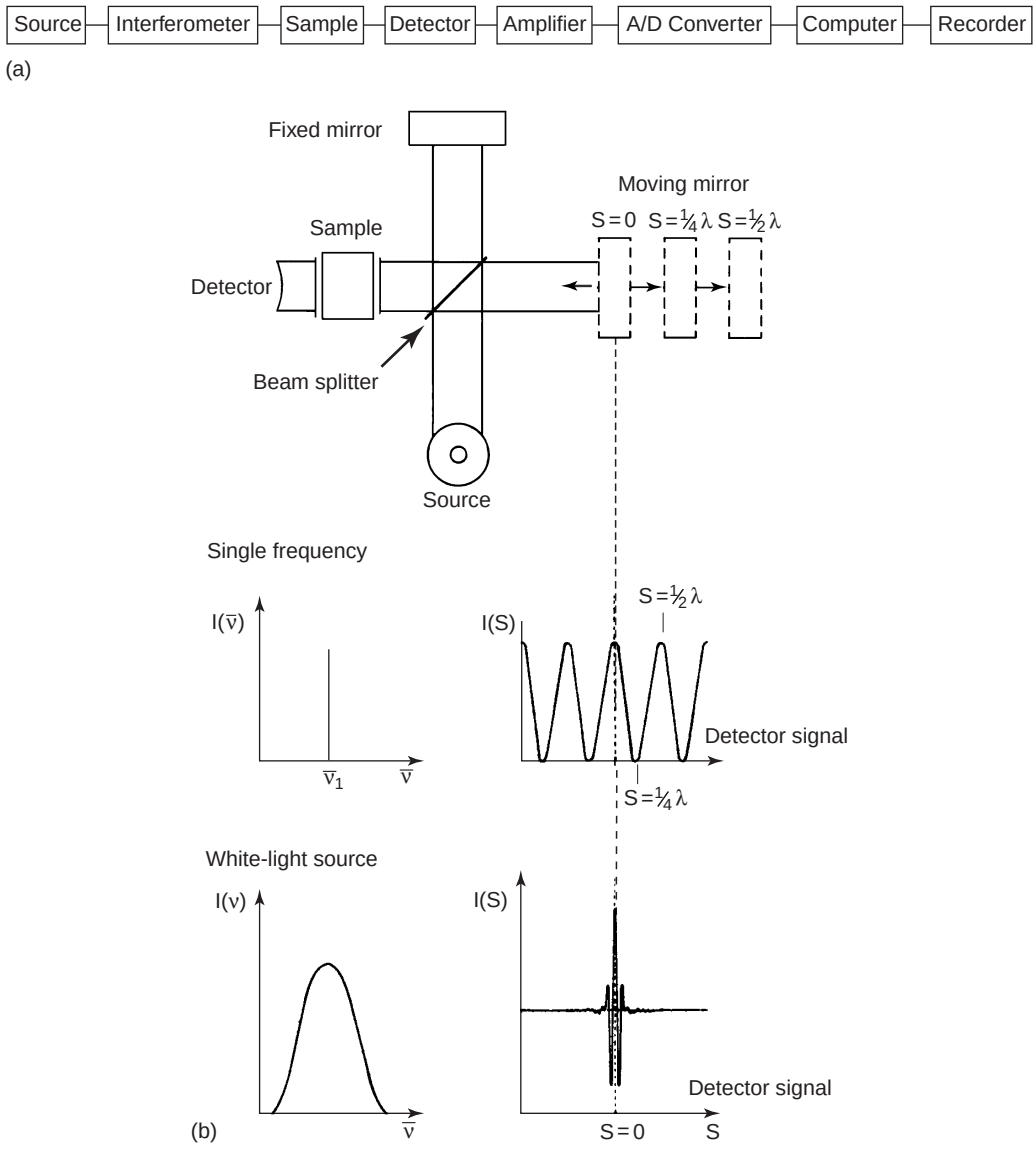

Figure 2 FTIR spectrometer. (a) Block scheme of the basic components of an FTIR spectrometer. (b) Working principle of a Michelson interferometer consisting of a light source, beam splitter, fixed mirror, moving mirror, detector and a sample (upper panel). A single frequency light source (central panel, left) is modulated to a sinusoidal signal observed by the detector (central panel, right). A white-light source (e.g. emitted from a globar) is transformed to the interferogram (lower panel).

energy getting to the detector. The resolution of the FTIR spectrum is solely defined by the pathlength of the moving mirror away from the zeropoint (see Figure 2b, central panel). This advantage is particularly useful in cases where a small sample limits the optical throughput.

2. The Felgett Advantage. Because the detector is observing all frequencies simultaneously, the signal is said to be ''multiplexed''. One complete scan of the moving mirror of the interferometer takes only a fraction of the time that a ''dispersive scan'' would need. Repetitive scans can be co-added and signal-averaged to reduce the random noise. Thus, very tiny signals may be extracted by decreasing the noise to a satisfactory level.
3. The Connes Advantage. The means by which the moving mirror position is monitored involves the use of an internal HeNe reference laser. This laser is also used as an internal wavelength calibration standard. Therefore, the wavelength accuracy of most of the FTIR spectrometers is specified to lie within $0.01\,cm^{-1}$, which is particularly useful when absorbance subtraction and resolution enhancement techniques are used and when stored spectra, for

data comparison, have to be transferred to different loci.

4. The Stray Light Advantage. Because the interferometer modulates each IR frequency differently, there is no equivalence to stray light in FTIR. This means that the absorbance values remain very linear even beyond three absorbance units.
5. The Data Handling Advantage. Since FTIR is based on highly efficient minicomputers controlling the interfermometer, collecting the digitized interferograms, and calculating the IR spectra, spectral data can now be easily stored on electronic devices and can be manipulated (resolution-enhanced, baseline-corrected, etc.).

2 CHEMICAL STRUCTURES AND COMPOSITION OF MICROBIAL CELLS

Understanding of IR spectra of microbial cells requires a general perception of its composition, major cell types, and chemical structures present, and knowledge of the differentiation of cells and tissues. At the simplest level all biological systems are composed of water, lipids, proteins, and carbohydrates. The gross composition of bacterial (prokaryotic), yeast, and mammalian (eukaryotic) cells is given in Table 1.

2.1 Bacterial Cells

In contrast to animal, plant, and yeast or fungal cells (so-called eukaryotic organisms) bacteria (also called prokaryotes) exist in only a limited number of morphological forms (e.g. rods, cocci, chains and spirals). Their chemical composition and structures, however, vary considerably. The cytoplasmic structures of bacteria are less organized (compartmentalized) and are simpler than those of yeasts and fungi, but have complex and diverse molecular and supramolecular structures outside the plasmic membrane. These include the cell wall, outer membrane, capsules and sometimes specific layers, e.g. the S-layers. Some bacteria are capable of sporulation or storage material production. Many structural differences providing the possibility of differentiation between bacteria by IR spectroscopy reside in the cell envelope, which is generally defined as the cytoplasmic membrane plus the cell wall.[8]

Table 1 Gross composition of pro- and eukaryotic cells[a]

	Bacteria[b]	Yeasts[b]	Animal cells[b]
Radius (μm)	~1	~10	10–100
Volume (μm^3)	~1	~1000	>10 000
Surface/volume (μm^{-1})	~1	~0.1	Tissue
Generation time (h)	0.2–10	2–10	~20
DNA base-pairs	$\sim 4 \times 10^6$	$\sim 20 \times 10^6$	$500–5000 \times 10^6$
Number of "genes"	~4000	~20 000	>50 000
Size of ribosome	70 S	80 S	80 S
RNA % (w/w)[c]	5–15	3–10	10–20
DNA % (w/w)	2–4	1–3	~5
Proteins % (w/w)	40–60	40–50	~60
Lipids % (w/w)	10–15	5–20	10–15
Carbohydrates % (w/w)	10–20	10–25	6–8

[a] Values have been taken from different sources.
[b] All numbers are rough estimations which depend on organism/organ, growth conditions, cell division cycle, and other factors.
[c] Content in % of cell dry weight.

Most cell envelopes fall in two categories: the so-called Gram-positive bacteria consisting only of the cytoplasm, the cytoplasmic membrane, and the cell wall; and the more complex Gram-negative which contain the so-called outer membrane in addition to the cell envelope. Some bacterial species, the mycoplasms, lack any cell wall at all, but express a rather rigid plasma membrane.

The bacterial cell wall is a rigid high-molecular network made up primarily of the peptidoglycan, which has a shape-giving function and protects the cells from osmotic disruption. Its primary structure consists basically of disaccharide–pentapeptide subunits with unusual features such as the occurrence of alternating D- and L-amino acids and a γ-bonded D-glutamic acid residue. Its structural variants are found to be different for various groups of bacteria. Additionally, a lipoprotein has been found to be covalently bound to the peptidoglycan of Gram-negative organisms. Many Gram-positive bacteria have an additional polymer, covalently bound to the peptidoglycan, the teichoic and teichuronic acids. The teichoic acids are ribitol or glycerol-containing macromolecules, built up by a phosphate-carrying backbone with side chains of variable composition. Teichuronic acids or neutral polysaccharides are also sometimes found in the Gram-positive cell wall. Gram-negative bacteria exhibit an additional membrane, the so-called outer membrane.

The outer membrane is an asymmetric membrane, the inner leaflet of which contains only phospholipids with nearly the same composition as found in the cytoplasmic membrane, while the outer leaflet contains exclusively one particular type of amphiphilic molecule, the lipopolysaccharides (LPS) and the various pore-forming proteins, the porins. The structure of LPS contains three basic regions: the so-called *O*-specific side chain (a hetero-oligosaccharide, responsible for *O*-antigenicity), the inner and outer core regions and a lipid anchor called lipid A. This outer membrane is the major permeability barrier that protects these cells against bile salts, degradation by digesting enzymes and

which prevents the cells from hydrophobic drugs and antibiotics diffusing through this particular layer. The mycobacteria, nocardia, corynebacteria, and some related groups have very unusual cell envelopes which form thick wax-like layers around the outside of the cell wall. Major compounds present in this impermeable and rigid layer are complex, long-chain fatty acids, the mycolic acids.[9]

Some bacteria form capsules (sometimes referred to also as ''slime layers'') surrounding the cell envelope. These are not essential structures and are frequently built up of negatively charged polysaccharide compounds. Some bacilli exhibit capsules composed of negatively charged homo-oligopolypeptides such as poly-D-glutamic acids.

Many bacilli and clostridia may form endospores which are modified cell structures that can survive under unfavorable environmental conditions. These endospores have two membrane-like layers. Between these two layers, a spore-specific peptidoglycan (the so-called cortex) is found which differs in primary and three-dimensional structure from the peptidoglycan of vegetative cells in that the muramic acid of peptidoglycan is modified to a lactam derivative and is less cross-linked. A keratin coat is located on the cell exterior and it has been established that large quantities of Ca^{2+}-dipicolinate are related to heat resistance of endospores.[8]

2.2 Yeast Cells

The proteins, lipids and polysaccharides that make up the membrane, cell wall, and capsules of capsulated yeasts have a significant impact on the systematics and phylogeny of yeasts. Only a small number of species of yeasts has been investigated in depth and even fewer studies have focused on structural details of these constituents.

The storage compounds in yeasts have been reviewed.[10] The principal, readily mobilizable reserve polysaccharide in yeasts is glycogen (see also Figure 3). Glycogen occurs in yeast cells both in the cytoplasm and nucleoplasm either in soluble form or as aggregates of spherical particles having a diameter of 40–50 nm. The glycogen content of yeast cells is highly dependent on the physiological state and may reach up to 20% of the dry weight of the cells. Glycogen may have a molecular weight up to 10^7; it contains three types of α-(1 → 4) linked chains of glycosyl units: short A-chains (side chains) attached to B-chains (main chains) by α-(1 → 6) glycosidic linkages. The main chains (carrying one or more side chains) are attached by α-(1 → 6)-glycosidic links.

The cell walls of yeasts are composed of complex polysaccharides and glycoproteins. The principal low-molecular building blocks of these polysaccharides are glucose and mannose followed by galactose, xylose, *N*-acetyl-D-glucosamine, and uronic acids. While the qualitative composition of these compounds is a taxonomic marker, the quantitative composition varies with cultivation conditions. Major structures of yeast cell walls are β-glucans in which the glycosyl units are mutually linked by β-(1 → 3), β-(1 → 6), and possibly β-(1 → 2) glycosidic bonds. These structures are found with different molecular weights and branching and may form microfibrillar structures of crystalline nature.[11] Chitin, a linear β-(1 → 4) polymer of *N*-acetylglucosamine is a typical constituent of primary septa and budding yeast.

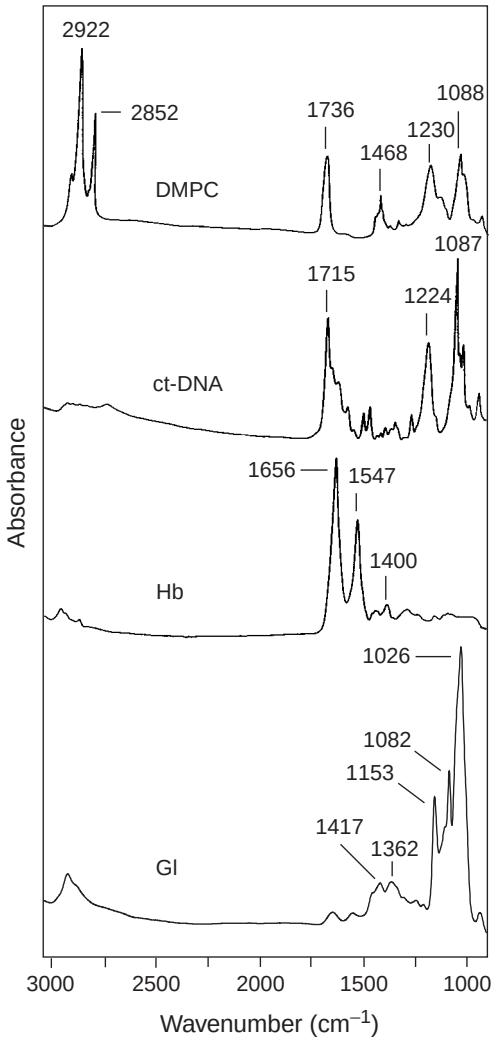

Figure 3 MIR-spectra of the main biological building blocks. DMPC; dimyristoylphosphatidylcholin; ct-DNA; calf thymus deoxyribonucleic acid; Hb; human hemoglobin; Gl; the carbohydrate glycogen. Spectra have been measured in H_2O and corrected for H_2O/buffer technique: absorbance/transmission (A/T); nominal physical resolution: 4 cm^{-1}; apodization function: Blackman-Harris 3-term; number of scans: 128; detector: DTGS; spectrometer: IFS 66 (Bruker Optics, Germany).

Chitosan, a β-(1 $\rightarrow$ 4)-linked polymer of D-glucosamine may be considered to be a minor yeast polysaccharide but in some dimorphic fungi belonging to the group of Zygometes it may represent one of the principal wall components. Mannans exist in yeasts as covalent complexes with proteins. The yeast mannoproteins are large molecules (molecular weight up to 500 000) consisting of a covalently linked carbohydrate and protein. The polysaccharide portion contains up to 150 mannosyl units, being connected via *N*-glycosidically linked polymannose units. A second group of carbohydrate of yeast mannans are short manno-oligosaccharides, *O*-glycosidically linked to serine and/or threonine residues of the polypeptide (protein).[11]

In some yeasts, particularly of the genus *Cryptococcus*, the polymers containing D-glucoronic acid residues are important constituents of extracellular capsules. For example, the capsule of *Cryptococcus laurentii* is made up of a α-(1 $\rightarrow$ 3)-linked mannose backbone with xylosyl and glucuronosyl residues as side groups.[12]

Yeast membranes contain a number of lipids and pigments that are not present in prokaryotic cells (bacteria).[8,13] These are sterols, sphingolipids, ergosterins, melanins, and some glycolipids. Culture conditions have a marked influence on the total lipid content and lipid composition of yeasts. Factors controlling lipid content and composition are pH of the medium, temperature, and time of growth, and the ratio of N- and C-sources. Sterols occur both in free form and as esters with long-chain fatty acids. Both forms are interconvertible. Free sterols are associated with membrane functions; sterol esters may fulfil a storage or ''pool'' function. Common sterol molecules of yeasts are ergosterol, lanosterol, episterol, zymosterol, and fecosterol. Major structures of sphingolipids found in yeasts are the sphingosines, cerebrins (ceramides), sphingomyelines, and cerebrosides. A typical membrane lipid in yeast is ergosterin. Its structure is similar to cholesterin and it belongs to the group of sterines. Further compounds frequently found in the membranes of yeasts are melanins which are black pigments built up from tyrosine derivatives.[8]

3 MAIN BIOLOGICAL BUILDING BLOCKS AND ASSIGNMENT OF INFRARED BANDS

IR absorption bands observed in the MIR region between approximately 800 and 4000 cm^{-1} mainly arise from the fundamental vibrational modes and in the group frequency notion, can often be assigned to particular functional groups. At wavenumbers lower than 1400 cm^{-1}, IR bands tend no longer to arise from localized vibrational modes but rather from skeletal and strongly coupled modes, which are difficult to describe. For practical purposes, rough band assignments can be obtained from group frequency charts published in several bibliographies that may be helpful to obtain correlations between partial structures and band frequencies. Several spectra descriptions and excellent structure–spectra correlations for the most important biological macromolecules can be obtained from the literature.[14,15] Efforts to interpret the IR spectra of biological molecules are mainly based on the analysis of known structures, normal coordinate analysis and isotope exchange experiments. Nucleic acids, proteins, lipids, and carbohydrates are constantly present in different amounts and in a wide diversity in microbial cells. Figure 3 gives the MIR-spectra of a typical phospholipid, DNA, protein, and carbohydrate structure. Some of the IR bands are numbered and tentative assignments can be taken from Table 2.

Figure 4 shows IR spectra of fully hydrated samples of intact cells of *Escherichia coli* that span the entire MIR and NIR range from 1000 to 10 000 wavenumbers. The MIR spectrum of Figure 4(a) was obtained on a hydrated film sample with a thickness of about 8 μm. Figure 4(b) shows the NIR spectrum recorded on a sample with a film thickness of about 250 μm. In both cases CaF_2 was used as the optical cuvette material. The insets of Figure 4(a) and (b) give the second derivatives as calculated from the original absorbance spectra. Both the MIR and the NIR regions provide a large number of spectral traits that may potentially be useful to characterize complex biological materials such as microorganisms. Some rough assignments in respect to OH, NH and CH functional groups are given at the top of Figure 4.

Figure 5 (curve a) shows a MIR spectrum of a typical microbial sample dehydrated to a thin film layer on a ZnSe optical plate. While exact assignment to specific structures is still too complex a task, many spectral features of intact cells can be visualized by applying resolution enhancement techniques. In general 50–70 spectral features are resolved by derivative or Fourier self-deconvolution techniques (see Figure 5, curve b), and are the basis of assignments to functional groups, known chemical compounds, partial structures or different conformational states of particular molecules present. These spectral features constitute a spectral fingerprint useful for differentiating between different types or states of cells.

IR spectra of microbial specimens provide not only a number of absorption bands that describe molecular composition of the cells. Many of these bands are also sensitive to structural changes, various intra- and intermolecular interactions including H-bonding pattern, membrane constitution, lipid–protein interaction, and conformational states like different secondary structures of proteins. The physical state of the sample such as hydration or aggregation state, interaction with ions

Table 2 Tentative assignment of some bands frequently found in microbial IR spectra (peak frequencies have been obtained from the second derivative spectra)[a]

Band numbering (see Figure 4)	Frequency (cm^{-1})	Assignment
	~3500	O−H str of hydroxyl groups
	~3200	N−H str (amide A) of proteins
1	2959	C−H str (asym) of $-CH_3$
	2934	C−H str (asym) of $>CH_2$
2	2921	C−H str (asym) of $>CH_2$ in fatty acids
	2898	C−H str of $\rightarrow$C−H methine
3	2872	C−H str (sym) of $-CH_3$
4	2852	C−H str (sym) of $>CH_2$ in fatty acids
5	1741	$>$C=O str of esters
	1715	$>$C=O str of esters, RNA/DNA, $-C(=O)OH$
	~1695	Amide I band components
	~1685	resulting from antiparallel
	~1675	pleated sheets and β-turns of proteins
6	~1655	Amide I of α-helical structures
7	~1637	Amide I of β-pleated sheet structures
8	1548	Amide II
9	1515	"Tyrosine" band
10	1468	C−H def of $>CH_2$
11	~1400	C=O str (sym) of COO^-
	1310–1240	Amide III band components of proteins
12	1250–1220	P=O str (asym) of $>PO_2^-$ phosphodiesters
13	1200–900	C−O−C, C−O dominated by ring vibrations of carbohydrates
		C−O−P, P−O−P
	1085	P=O str (sym) of $>PO_2^-$
	720	C−H rocking of $>CH_2$
	900–600	"Fingerprint region"

[a] Adapted from Naumann et al.[6]
asym = asymmetric; sym = symmetric; str = stretching; def = deformation.

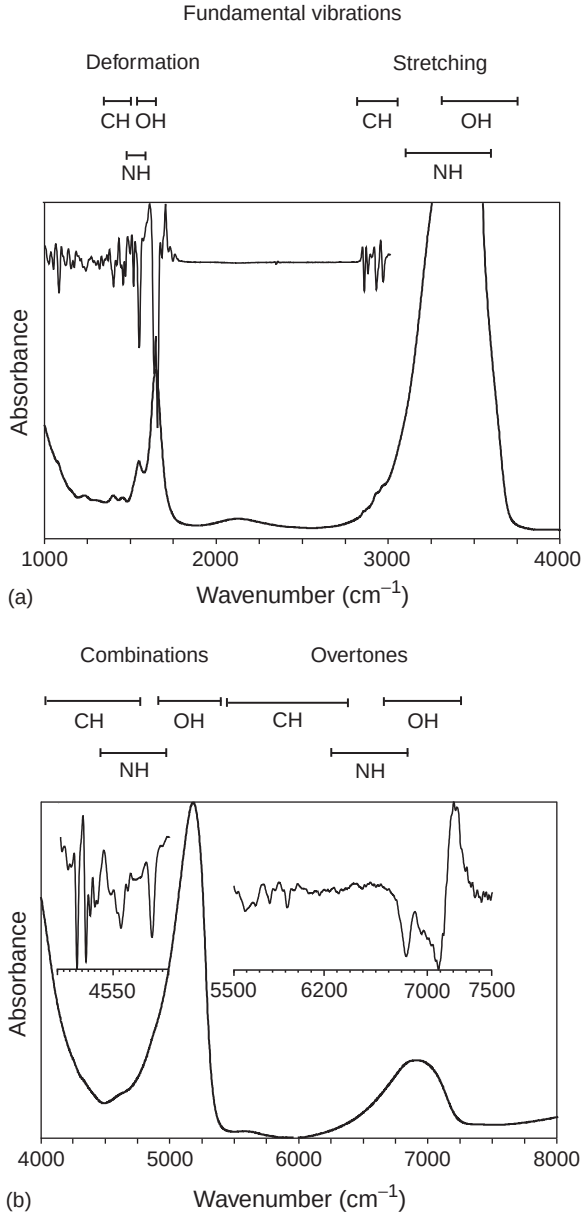

Figure 4 IR absorbance spectra of a fully hydrated sample of *Escherichia coli* (strain RKI/A 139) in the spectral range between 1000 and 10 000 cm^{-1} (a) MIR; (b) NIR. The insets show the second derivatives as calculated from the original absorbance spectra. Physical resolution applied was 6 cm^{-1} for the MIR and 16 cm^{-1} for the NIR measurements, respectively. Technique: A/T; cuvettes: CaF_2 windows with 8 µm (MIR) and 250 µm (NIR), respectively; number of scans: 64 (MIR, DTGS-detector), 512 (NIR, InSb-detector); apodization function: Blackman-Harris, 3-term; spectrometer: IFS 66 (Bruker Optics, Germany).

and so on has a strong influence on results. These facts necessitate the rigorous standardization of sampling, preparation, and data acquisition procedures.[6,16–18]

Tentative band assignments based on the systematic comparison of resolution-enhanced microbial IR spectra with those of the known building blocks constantly present in intact cells can be derived from Table 2 (see also spectra of Figure 3 for comparison):

1. The region between 4000 and 3100 cm^{-1} is dominated by rather broad spectral features resulting from −OH (~3400 cm^{-1}) and N−H stretching modes (amide A ~ 3300 cm^{-1} and amide B ~ 3030 cm^{-1}).

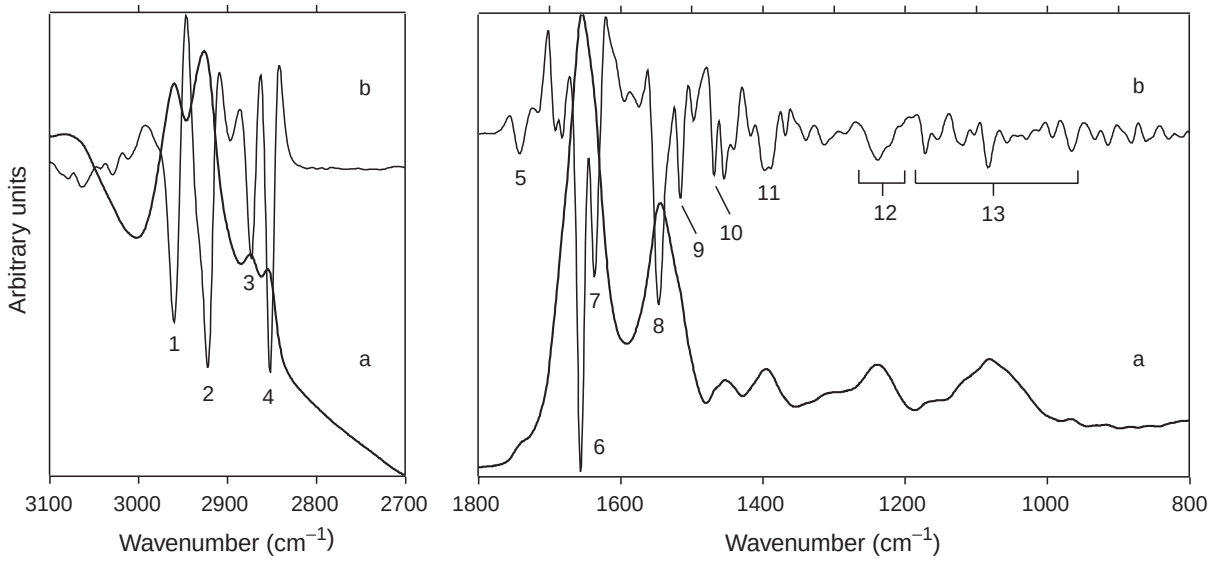

Figure 5 Tentative band assignment of some bands frequently observed in bacterial MIR spectra (*Staphylococcus aureus* (strain SG 511)). (a) Original absorbance spectrum; (b) second derivative. (1) $-CH_3$ asymmetric stretching; (2) $-CH_2$ asymmetric stretching; (3) $-CH_3$ symmetric stretching; (4) $-CH_2$ symmetric stretching; (5) $>C=O$ stretching; (6) amide I of α-helical structures; (7) amide I of β-sheet structures; (8) amide II; (9) tyrosine ring vibration band; (10) $>CH_2$ bending; (11) $-COO^-$ symmetric stretching; (12) $>PO_2^-$ asymmetric stretching; (13) spectral range dominated by complex ring vibrations of carbohydrates, P−O−P stretching, C−O−P stretching, $>PO_2^-$ symmetric stretching. Technique: A/T; number of scans: 64; nominal physical resolution: 6 cm^{-1}; apodization function: Blackman-Harris, 3-term; IR cuvette used, see Figure 6(a); spectrometer: IFS 28/B (Bruker Optics, Germany).

2. The region between 3100 and 2800 cm^{-1} exhibits the C−H stretching vibrations of $-CH_3$ and $>CH_2$ functional groups and, hence, is generally dominated by the spectral characteristics of fatty acid chains of the various membrane amphiphiles (e.g. phospholipids) and by some amino acid side-chain vibrations. Complementary information can be deduced from the region between 1470 and 1350 cm^{-1}, where the various deformation modes of these functional groups are found. In rare cases, a weak band near 3015 cm^{-1} is also observed, resulting from =C−H double bond stretching modes of unsaturated fatty acid chains.

3. The region between 1800 and 1500 cm^{-1} is dominated by the conformation-sensitive amide I and amide II bands, which are the most intensive bands in the spectra of nearly all bacterial samples so far tested. Since IR spectroscopy is an averaging technique, the amide I and amide II bands cannot provide structure information on a single protein; they rather indicate the predominance of α- or β-structures present. Useful information can also be obtained from bands near 1740 cm^{-1}, resulting from $>C=O$ stretching vibrations of the ester functional groups in lipids. Absorptions of nucleic acids can also be expected in this spectral domain due to $>C=O$, $>C=N$, $>C=C<$ stretching of the DNA or RNA heterocyclic base structures.[14,15] A band near 1715 cm^{-1}, which is assigned to a $>C=O$ stretching vibration, is routinely observed in the spectra of hydrated microbial cells and tissue material, and is known as a sensitive probe of base pairing in nucleic acids. Weak features of nucleic acids between 1600 and 1700 cm^{-1} are often overlapped by the much stronger protein amide I bands, since the relative amount of DNA or RNA per cell mass generally does not exceed 10% w/w. Weak bands, which can be assigned to amino acid side-chain vibrations, occur near 1498 cm^{-1} (phenylalanine), 1516 cm^{-1} (tyrosine) and between 1585 and 1570 cm^{-1} (aspartate and glutamate carboxylate stretching).

4. Complex absorption profiles are observed between 1300 and 1500 cm^{-1} arising predominantely from $>CH_2$ and $>CH_3$ bending modes of lipids and proteins. A characteristic, but weak, feature is often observed around 1400 cm^{-1}, which may be attributed to the symmetric stretching vibrations of $-COO^-$ functional groups of amino acid side chains or free fatty acids.

5. Around 1230 cm^{-1} superimposed bands typical of different $>P=O$ double bond asymmetric stretching vibrations of phosphodiester, free phosphate and monoester phosphate functional groups are observed. In most cases, three to four different weakly pronounced features can be discriminated by resolution enhancement, with the band near 1220 cm^{-1} being

most probably due to the phosphodiester functional groups of DNA/RNA polysaccharide backbone structures. The other >P=O double-bond stretching frequencies are due to head group vibrations of phospholipids or, for example, the phosphorus-containing carbohydrates such as ''teichoic acids'' and ''lipoteichoic acids'' (charged polymers, which may be present in substantial amounts in Gram-positive bacteria).

6. The spectral region between 1200 and 900 cm^{-1} is generally dominated by the symmetric stretching vibration of PO_2^- groups in nucleic acids and a complex sequence of peaks mainly due to C–O–C and C–O–P stretching vibrations of various oligo- and polysaccharides.
7. The region between 900 and 600 cm^{-1} exhibits a variety of weak, but extremely characteristic, features superimposed on an underlying broad spectral contour. This region may contain weakly expressed bands arising from aromatic ring vibrations of phenylalanine, tyrosine, tryptophan and the various nucleotides. With the exception of only a few peaks (e.g. a band near 720 cm^{-1}, resulting from the $>CH_2$ rocking modes of the fatty acid chains), valid assignments can hardly be achieved. Therefore, we refer to this spectral domain as the ''true fingerprint region''.

4 EXPERIMENTAL METHODOLOGIES

A major advantage of IR spectroscopy is that almost any kind of material can be measured and that it is not limited to the physical state of the sample (samples may be solutions, viscous liquids, suspensions, inhomogeneous solids or powders).[5] Additionally, there are no principal restrictions to recording IR spectra of a given sample under very different physicochemical conditions with respect to temperature, pressure, state of dispersion, pH and so on. This is of advantage for biomedical analyses, since it is imperative to test biological specimens under conditions that leave the sample's structures ''as they are'', preferentially hydrated, unperturbed and nondisintegrated.

4.1 Infrared Measurement Techniques for Microbial Samples

In general, biological samples of microorganisms do not behave ideally. Different biological specimens should be compared under conditions where (1) IR absorbances of the samples to be compared are not too different, (2) IR bands are not too intensive in order to avoid detector nonlinearities, (3) Beer's law is at least approximately obeyed, (4) signal-to-noise ratio is sufficiently high, and (5) varying baseline shifts caused by diffuse scattering at the sample surface and/or by inhomogeneity within the sample itself is minimized. These requirements are best fulfilled by using the traditional A/T, absorbance/reflectance (A/R), and the attenuated total reflection techniques.[5–7]

4.1.1 Absorbance/Transmission Measurements

A/T spectra can be obtained from liquid solutions, dispersions or suspensions, from viscous or solid films cast on suitable IR transparent plates. Since water is the medium ubiquitously present in all biological samples, water-insoluble and IR transparent optical materials have to be used. These are CaF_2, BaF_2, ZnSe, ZnS, KRS-5™, or germanium which differ in refractive index, spectral transparencies, and water solubility. Table 3 surveys relevant optical properties of these materials. Different cuvette systems have been described in the literature for A/T–IR measurements

Table 3 Optical properties of the most important water-insoluble optical materials

Material (thickness)	ZnSe ($d = 3$ mm)	ZnS ($d = 2$ mm)	Ge ($d = 1.5$ mm)	BaF_2 ($d = 10$ mm)	KRS-5™ ($d = 1$ mm)	CaF_2 ($d = 1$ mm)	AgCl ($d = 1$ mm)
Transmission range (μm)	0.5–20	0.4–14	1.8–23	0.18–12	0.45–45	0.12–12	0.4–30
Maximal transmission (%)	70	70	45	85	75	>90	80
Loss of reflection (for two surfaces) (%)	30 (10.6 μm)	25 (10.6 μm)	53 (10.6 μm)	7.7 (0.6 μm)	28 (10 μm)	12 (4 μm)	19.5 (10 μm)
Solubility in water at 300 K (g/100 g H_2O)	1×10^{-3}	6.9×10^{-4}	Insoluble	0.17	5×10^{-2}	1.3×10^{-3}	1.5×10^{-4}
Refractive index at							
0.5 μm	2.66	2.42		1.48	2.73	1.44	2.10
5.0 μm		2.25	4.01	1.45		1.40	2.00
10.0 μm	2.41	2.20	4.003	1.40	2.37		1.98
20.0 μm	2.30				2.34		1.95

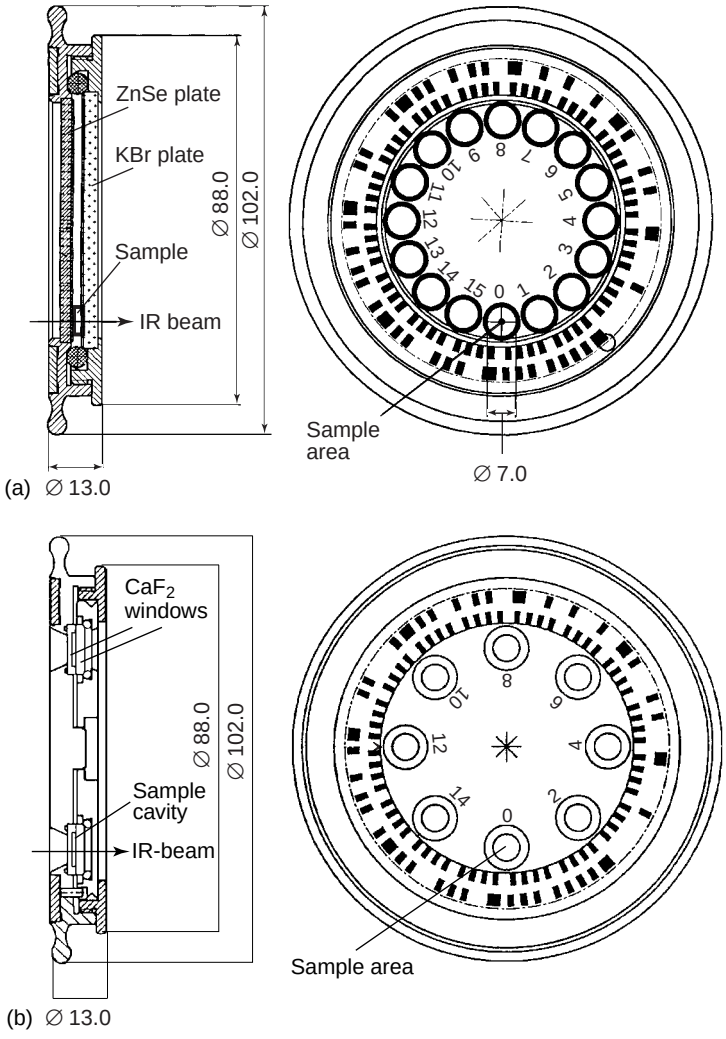

Figure 6 Technical drawing of two computer-controlled multisample cuvettes useful for measuring dried (a) or fully hydrated biological samples (b) in A/T. The cuvette systems are used in the dedicated FTIR spectrometer IFS 28/B (Bruker Optics, Germany).

of microbial samples. Two versatile technical solutions for the consecutive measurement of microbial samples (dried film samples or aqueous solutions, suspensions or gel discs) are given in Figure 6.[6,19]

4.1.2 Absorbance/Reflectance Measurements

A/R measurements (sometimes also called ''external reflectance'') can be performed on the same samples used for A/T. The samples are placed on a light-reflecting instead of transmitting medium (e.g. silver, gold or polished stainless steel) which leads to a double-passage or, for special A/R devices, even a multipassage of IR radiation through the sample. At a grazing incidence of light, molecular monolayers, e.g. from lipids, may be measured as well.[5,7]

4.1.3 Attenuated Total Reflectance Measurements

For an attenuated total reflectance (ATR) experiment, the sample is no longer placed in the path of the propagating IR beam as it is for A/T or A/R experiments, but is brought into contact with the surface of an internal reflection plate or prism, where it can interact with the IR radiation evanescing from the optically denser (ATR plate) to the rarer (sample) medium.[20] The penetration depth of the electromagnetic wave into the rarer medium is defined by the wavelength dependent ratio n_2/n_1 of the refractive indices of the denser (n_2) and the rarer (n_1) media and by the angle of incidence α and is in the order of a few micrometers. For a defined ATR element (n_1) and angle of incidence, the effective optical pathlength is constant or similar for different samples that have comparable refractive indices and provided the optical contact between the ATR

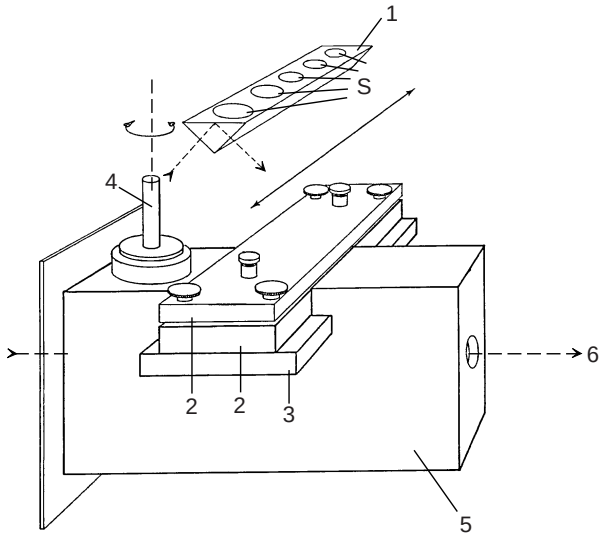

Figure 7 Schematic drawing of a multisample ATR cuvette: (1) ZnSe prism with five marked sample areas S; (2) demountable, hermetically sealed cartridge (two parts) carrying the ATR prism (1); (3) sliding carriage of the cartridge by which the ZnSe prism (sample areas S) may be moved through the IR beam for consecutive measurements; (4) cartridge drive; (5) housing with reflecting optics; (6) IR beam. (Adapted from Naumann et al.[6])

element and the samples is the same. These experimental prerequisites can be fulfilled at least approximately for large numbers of biological samples. A specially designed multisample ATR cuvette suitable for the measurement of hydrated and dried microbial samples is shown in Figure 7.[6]

4.1.4 Diffuse Reflectance Measurements

Diffuse reflectance (DR) measurements of highly scattering samples like freeze-dried biological specimens, surface-rich materials like catalytic or chromatographic carriers, and finely ground or powdered materials can be performed using standard accessories supplied by most manufacturers of FTIR spectrometers.[7] A thin-layer chromatography accessory useful for FTIR measurements of dried microbial samples in DR has been described.[21]

4.2 Sampling of Microbial Cells and Data Acquisition

In order to obtain reproducible results, sampling of biological species, sample treatment procedures, IR measurement techniques, and IR acquisition parameters have to be controlled and standardized rigidly. There is no simple and uniform answer to all these requirements. For microbial analyses standardized experimental protocols including data acquisition and evaluation procedures have already been published.[6,17,18] These standardization efforts have been stimulated by the necessity to exchange spectral data between different laboratories and to construct validated reference databases for routine identification of microorganisms isolated in different laboratories.

IR spectra may be obtained with sufficient reproducibility from microorganisms provided the microbiological parameters influencing cell growth (composition of growth media, incubation time, temperature of growth etc.) can be controlled and standardized. Compared with these requirements sample collection, sample preparation, and spectroscopic data acquisition parameters (spectral resolution, scanning time etc.) are of secondary importance. Microbial samples suitable for IR measurements can be obtained either from liquid cultures or directly from solid agar plates. These samples can be measured as hydrated pellets or dried films applying either the A/T, A/R or ATR techniques.

Using the A/T technique a typical simple protocol runs as follows. Subcultured microbial strains are cultivated on appropriate solid nutrient agar plates. Time and temperature of growth depends on the type of microorganism tested (e.g. 24 h at 37 °C for many human pathogens). Small amounts of microorganisms are carefully removed from the agar plate with a standard calibrated (e.g. 1 mm in diameter) platinum loop and are suspended in 80 μL of distilled water. Subsequently, 30 μL of the suspensions are transferred to the water-insoluble IR transparent optical plate as small drops covering predefined sample areas of the ZnSe plate of the multicuvette system shown in Figure 6. The drop of microbial suspension is then dehydrated in a desiccator over a drying agent (P_4O_{10} Sicapent from Merck or Silicagel) applying a moderate vacuum (≈25 Torr) to form transparent film discs suitable for A/T–IR measurements. The optical plate is then sealed in a gas-tight cuvette-cartridge to control humidity and to prevent the instrument from contamination and is transferred to the automatic cuvette holder of the instrument. The physical parameters have to be kept constant for all measurements to be compared. A suitable set of physical parameters is as follows: 6 cm^{-1} nominal physical resolution, Blackman–Harris three-term apodization function, and a sufficient number of scans to reach a signal-to-noise ratio better than 3000 : 1. It is advisable to take the single beam reference spectrum through an empty place of the multisample cuvette system directly before the single beam sample spectrum is obtained. This eliminates virtually all contributions from impurities on the optical materials and minimizes problems arising from water vapor and CO_2 bands caused by possible instabilities of the instrument and of dry air purging of the system.[6,17–19]

4.3 Variability of the Microbial Specimen and the Problem of Reproducibility

The enormous diversity of microbial species and strains makes it an absolute necessity to perform measurements on statistically significant numbers of samples on hundreds if not thousands of different species and strains of microorganisms. The most important factors influencing data quality are repeatability and reproducibility of measurements. First, different levels of reproducibility that define and limit the discriminative power of the IR technique have to be considered. For quantification purposes it is useful to define different reproducibility levels (RL_i) of repetitive IR measurements on microbiologically identical strains:[6,17,18]

- (RL_1) that describes repeatability of independent IR measurements on samples prepared, e.g. from aliquots of one and the same aqueous suspension of microorganisms
- (RL_2) that defines repeatability of independent IR measurements on samples prepared from aliquots of microbial suspensions with the microorganisms grown on the same agar medium produced from different batches over a sufficient period of time (e.g. one year).

To calculate quantitative numbers for these RL, an objective measure for description and comparison of independent IR measurements is necessary. One possibility is the crosswise calculation of correlation coefficients r_{y1y2} (Pearson's product–moment correlation coefficient) between pairs of measured spectra according to Equation (1).

$$r_{y1y2} = \frac{\sum_{i=1}^{n} y_{1i}y_{2i} - n\overline{y_1}\,\overline{y_2}}{\sqrt{\sum_{i=1}^{n} y_{1i}^2 - n\overline{y_1^2}}\;\sqrt{\sum_{i=1}^{n} y_{2i}^2 - n\overline{y_2^2}}} \qquad (1)$$

y_{1i} and y_{2i} are the individual absorbance values of the two spectra to be compared; n is the number of data points in the given frequency range; $\overline{y_1}$ and $\overline{y_2}$ are the arithmetic mean values of y_1 and y_2.

From the correlation coefficient r_{y1y2}, a so-called differentiation index D_{y1y2} may be defined according to Equation (2):

$$D_{y1y2} = (1 - r_{y1y2})1000 \qquad (2)$$

where r_{y1y2} is the correlation coefficient. D may adopt values between zero and 2000, with zero for identical spectra (or spectral ranges), 1000 for completely noncorrelated and 2000 for completely negatively noncorrelated spectra.[6,17,18]

Mean D values between 0.4 and 0.8 for RL_1 and between 7 and 10 with Poisson-like distributions of calculated individual D values for RL_2 are obtained in practice when analyzing the first derivatives in the spectral range between 900 and 1200 cm^{-1} of dried microbial film samples with the cuvette system shown in Figure 6.[6,17,18] RL_1 is generally one order of magnitude lower than RL_2. Hence the microbiological parameters (quality of growth medium, cultivation, and sampling of bacterial biomass, e.g. from the surface of solid agar plates) define the repeatability of IR measurements on microorganisms. The comparison of strains from different species or genera may yield D values higher than 300.[6,17–19] Obviously IR spectroscopy provides considerable spectral variance that can be utilized to discriminate between different microorganisms.

5 DATA TREATMENT AND EVALUATION TECHNIQUES

Because thousands of measurements on hundreds of different species and strains of microorganisms have to be evaluated and compared for spectral similarity/dissimilarity, the use of multivariate data compression and pattern recognition techniques is a necessity. While univariate statistical analysis considers only a single property of a given object (e.g. a single intensity at a given wavenumber), multivariate statistics evaluate several properties of the objects at the same time. In this way the interrelationships between the properties of the objects can also be taken into account. The arsenal of multivariate statistical techniques provides ample methodologies for the pretreatment, evaluation, and representation of huge and complex data structures. Of the large number of pattern recognition techniques available, three are of particular interest when IR spectra of microorganisms have to be analyzed. These are factor analysis, hierarchical clustering, and ANNs.[22–26]

Factor analysis is primarily used to achieve data reduction. Hierarchical clustering, a so-called unsupervised classification method, attempts to find intrinsic group structures within the data set without the need of any class assignment nor partitioning of the data into a training and test data set. ANN analysis on the other hand is a supervised classifier by which the class assignment of each individual object is needed from the beginning. Partitioning of the whole data into a training and a test data set is needed to ensure reliability of results. Robustness of classification is further enhanced by cross-validation using the stringent ''leave-one-out'' method.

5.1 Factor Analysis

Briefly, the goal of factor analysis is to extract essential information from mixed and large data sets and to attain the classification of pattern.[22,23] As in any MSA technique, the objects (here IR spectra) are first represented as points in hyperdimensional space with as many dimensions as there are properties (in our case digitized absorbance values at discrete wavenumber positions). The procedure of how data compression is achieved by factor analysis is shown schematically in Figure 8.

Beginning with a primary database, the spectral range of interest is selected, eventually filtered (e.g. by calculating the first or second derivative, or by applying a band pass filter or Fourier self-deconvolution); finally the spectra are normalized (e.g. by peak or vector normalization). The aim of factor analysis is now to calculate by a specific eigenvector/eigenvalue analysis a new variance-weighted coordinate system which is adapted to the characteristic structure of the data cloud. The orthogonal vectors of this new coordinate system, the so-called eigenvectors (in our case ''factor spectra''), then represent characteristic structure information of the complex data cloud. In this way the objects (IR spectra) can be represented in the new, variance-weighted coordinate system as a function of specific properties of the whole data set rather than as a function of absorbances or frequencies. Typically, the first 2–24 eigenvectors of the new coordinate system represent, to a decreasing amount, the bulk of interspectra variance. The higher indexed eigenvectors (factor spectra) soon represent predominantly merely noise. Since the original IR spectra can be represented in this new coordinate system by only 2–12 numbers (factorial coordinates or factor loadings) compared to the 2000–4000 absorbance values of an MIR spectrum, a considerable data reduction is achieved, which allows the structures within the data cloud to be recognized. The mathematics of factor analysis are generally described in the matrix notation. This notation allows the use of general metrics (e.g. Euclidean metrics for principal component analysis (PCA) or chi-squared metrics for correspondence analysis). For more details, the reader is referred to the literature.[22,23]

Figure 9 shows a so-called PCA map calculated from a data set of approximately 140 spectra collected from independent measurements on 13 different species and strains comprising Gram-positive and Gram-negative

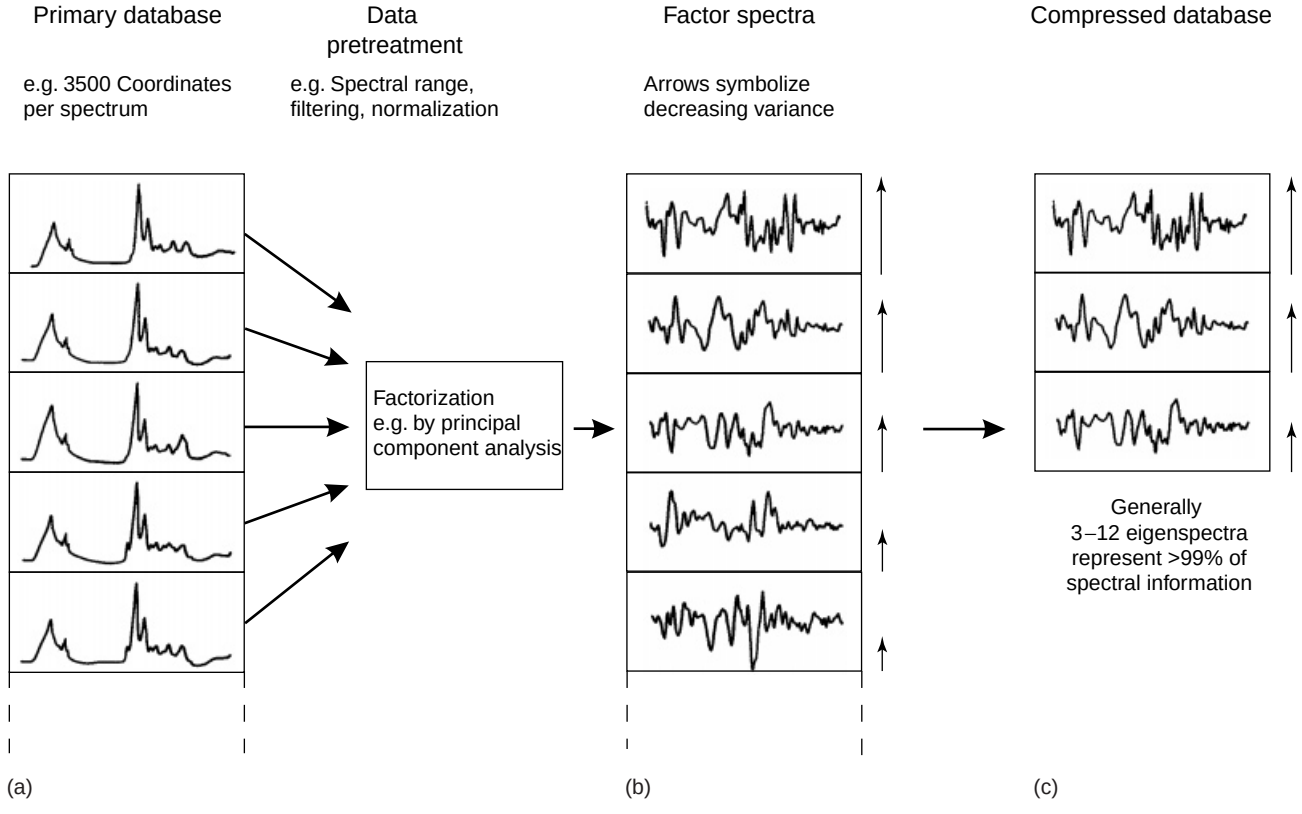

Figure 8 Schematic representation of data compression by factor analysis. (a) The primary database of IR spectra. These spectra are filtered, for example by calculating the first derivatives, and the spectral range(s) of interest are selected and normalized. (b) Variance-weighted factor spectra are obtained as a consequence of factor analysis. Arrows indicate decreasing variance. (c) Data compression is obtained since the first 2–24 eigenspectra generally describe more than 99% of variance within the data set.

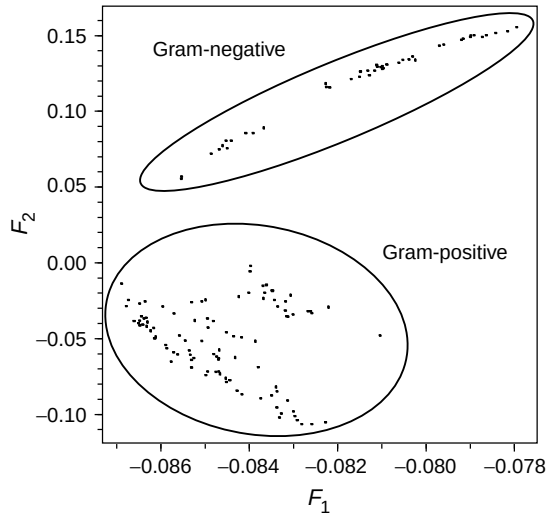

Figure 9 PCA map as calculated from a data set of 143 IR spectra obtained from Gram-positive and Gram-negative bacteria comprising repetitive measurements on 13 different species and strains. For projection of data, the factorial coordinates (factor loadings) F_1 and F_2 are used. As input data, the normalized first derivatives in the spectral ranges of 2800–3000 cm^{-1} and 1400–1500 cm^{-1} have been used. Derivation, normalization, and PCA (factor analysis) were performed using the software OPUS Vers. 3.0 (Bruker Optics, Germany running, under OS/2 Warp Vers. 3.0 (IBM, USA)).

strains of microorganisms. The map has been constructed by two-dimensional projection of eigenvectors (factor spectra) 2 and 3 such that the intrinsic group or class structure of the whole data set can be inspected by eye. Each point in the map represents a spectrum, the factorial coordinates (factor loadings) 2 and 3 are used for data representation. PCA was applied using the normalized first derivatives in the spectral ranges between 2800–3000 and 1400–1500 wavenumbers as input data. Two diffuse clouds can be distinguished with the upper and lower clouds representing exclusively the Gram-negative and the Gram-positive bacteria, respectively. Since the spectral ranges used are known to represent the spectral characteristics of the C–H stretching and deformation vibrations of the $-CH_3$, $>CH_2$, and $\equiv$C–H functional groups, respectively, this finding is caused because these Gram-positive and Gram-negative organisms differ in carrying (Gram-negative) or not carrying (Gram-positive) a membrane structure additional to the cytoplasmic membrane, the outer membrane, which, in turn, leads to specific differences in the spectral range where the various C–H stretching bands of the membrane's lipids are expressed.[6,17]

5.2 Multivariate Clustering

Cluster analysis aims to classify objects, i.e. the description of the structure and property interrelationships intrinsic to a given set of objects each defined by a multiplicity of properties.[24] The various classification algorithms use different distance measures to calculate similarity between the objects (e.g. correlation coefficients, Euclidean distances, Mahalanobis distances). Starting from the calculated distance matrix, an attempt is made to partition the objects into groups, classes or clusters, such that all objects belonging to the same cluster are as similar as possible and all objects belonging to different clusters are as different as possible, which means that intraclass variance should be as small as possible, while the interclass variance is a maximum. Often it is sufficient to find classifications that are only locally optimal. So far all hierarchical cluster analysis techniques are of approximative character. All cluster analysis techniques can be divided into hierarchical and nonhierarchical procedures. In most cases the hierarchical techniques are favored, since it is assumed that they describe best the inherent hierarchical structure of class-division within the given data set. The various different hierarchical cluster analysis techniques are generally distinguished between divisive and agglomerative methods. Divisive techniques are thought to be more potent, but much more computer intensive. Thus, the agglomerative techniques are used more frequently. All agglomerative clustering procedures work according to the same hierarchical ascendant principles: (1) in the beginning there are n objects that have to be classified, (2) then the two objects lying closest have to be found and are collected in one cluster, (3) the distances between this new cluster and all other objects are calculated, (4) in a next step, a search is made for the next nearest objects or clusters which are collected in one cluster. This procedure is repeated until all objects are collected in one single cluster. The history of hierarchical classification analysis is represented by a minimal spanning tree, also called a dendrogram, in which the merging process of classes can be followed visually. These classification schemes cannot provide an objective criterion of best partitioning. In one way or other, the number of classes has to be predetermined by the user, who needs at least some a priori knowledge about the inherent class structure of the data cloud. It is possible to provide the user with additional information that facilitates the interpretation (e.g. a curve showing the number of classes in a hierarchical cluster analysis calculation versus intraclass variance), but the final decision if the partition is useful or not is subjective. Hierarchical cluster analysis can be applied directly to IR spectra (the objects) represented in spectra space as points in hyperdimensional space with as many dimensions as there are encoded absorbance values, or to the IR spectra represented in a data compressed form in factor space with as many dimensions as there are factorial coordinates (factor loadings) selected for data representation (see Figures 10a and b). Adequate distance measures used

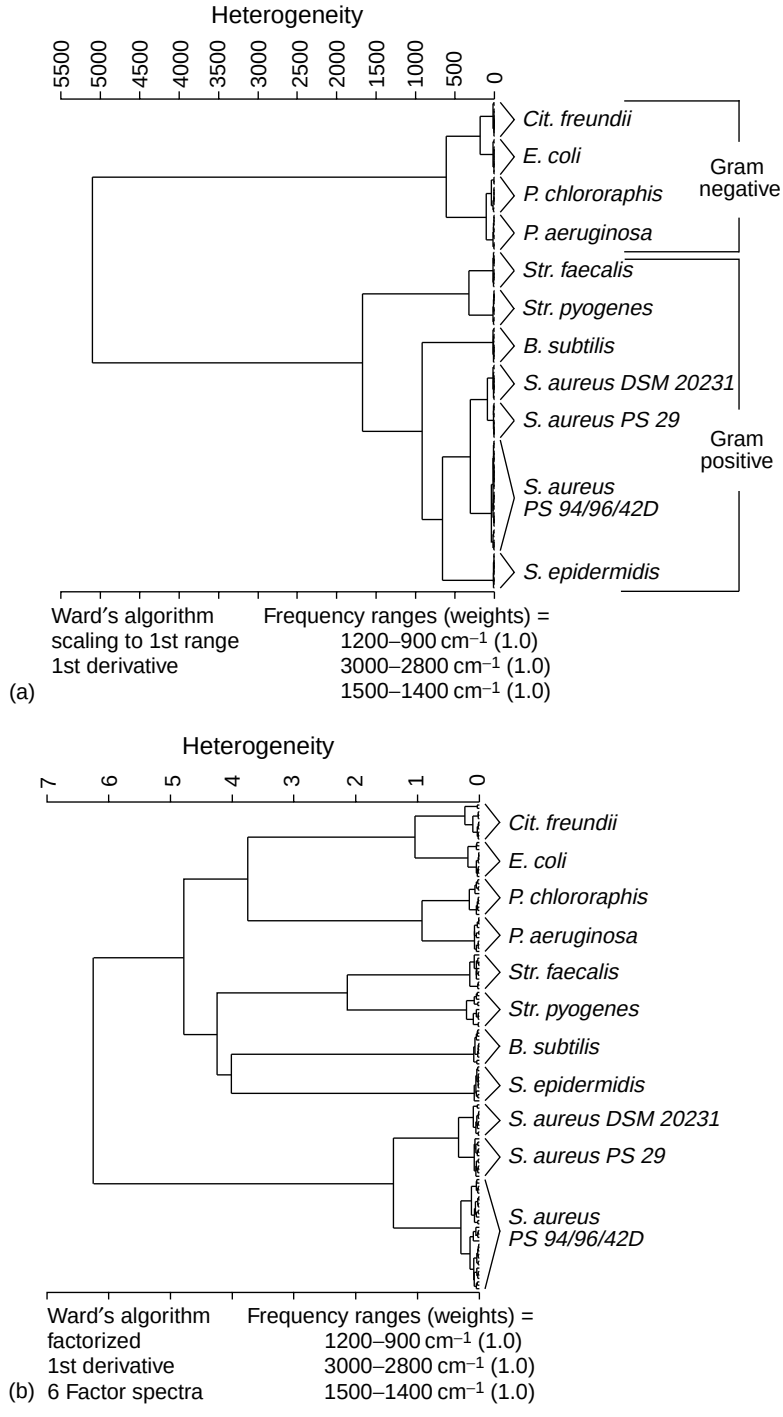

Figure 10 Minimal spanning trees, also called dendrograms, that graphically show hierarchical cluster analysis results on the same 143 spectra of Gram-positive and Gram-negative microorganisms shown in Figure 9. (a) Dendrogram obtained when the interspectral distance matrix is calculated in spectra space using D value calculation of spectral distances as defined by Equation (2). (b) Dendrogram obtained when interspectral distances (Euclidean) are calculated in factor space after data compression using the first six factors. Cluster and factor analysis are based on the first derivatives in three spectral ranges (2800–3000, 1400–1500, and 900–1200 cm^{-1}) each separately vector-normalized. For cluster analysis the Ward's algorithm was used. Factor analysis, hierarchical clustering, derivation, and normalization were performed using the software OPUS Vers. 3.0 (Bruker Optics, Germany) running under OS/2 Warp Vers. 3.0 (IBM, USA).

to calculate the spectral distances between the objects in spectra or factor space can be, for example, the Euclidean distance or the Pearson's product momentum correlation coefficient. The decision which is the best algorithm for cluster analysis and the optimal distance measure for calculating the distance matrix has to be found by trial and error. Satisfactory results on IR spectra of complex biological molecules can be expected from the Ward's and the so-called average linkage algorithms using the Pearson's product momentum correlation coefficient as a distance measure.[17,18]

5.3 Artificial Neural Networks

In contrast to cluster analysis, ANN analysis is a supervised classification approach that needs a priori knowledge of the group or class structure of the data.[25,26] ANNs are self-adaptive, parallel machine-learning systems made up of simple processing units, which have a natural propensity for storing experimental knowledge and making it available for practical use. ANNs are primarily used for pattern recognition purposes. They resemble the connectivity in brain in two respects. The first is that knowledge is acquired by the network from a learning process and in the second the interneuron connection strengths, known as synaptic weight, store the acquired knowledge. Several models of information processing of biological systems are used as an analogy. In nerve cells (neurons) the following abstract elements can be defined: (1) synapses which have chemical signal substances, (2) dendrites, where the incoming signals are received and processed, (3) the cell nucleus (kernel) which controls this process, and (4) axons, which propagate the impulse irritation forward. The incoming weights are calculated via different input layers of adjacent cell connections. After summation, an impulse is induced according to the strength of the weighted signals. The propagation function transports the impulse to a new cell, or directly to the output layer. The mathematical functions involved in this process are: (1) an activation function, (2) a description of the activation state of the neuron and (3) an output function. The activation function describes how a new activation state is developed from a preceding state of the neuron. The activation state defines the degree of activation and the propagation function has to weigh and combine all incoming signals. Therefore, a summation is often used as an algorithm. The value produced by the propagation function is used as input by the activation function. The activation function then produces the output. Binary, linear, sigmoid, and logarithmic functions are used as activation functions. The connecting network defines the manner in which the architecture of the network operates. For feed-forward ANNs, the information propagates from the input layer through the hidden layer(s) to the output layer. During training the reaction value is compared with the expected value at the output of the net. As long as deviations occur, the extent of the deviation is used as a measure of, how strong and where the weights are to be changed to minimize the output error. The second group of algorithms used are the feedback nets which can feed back the outcoming information from the neurons to a preceding layer. Examples are the so-called Hopfield, Elman, and Jordan nets. A supervised ANN requires a desired output for each input vector (spectrum) which is then compared to the actual output generated at a certain stage of analysis. A couple of learning strategies and learning rules are used to obtain incremental changes in the weights in order to optimize an error criterion. The learning strategy attempts to minimize a global error function for the given set of training data. The process of computing local errors for each processing element for optimization of the weights is continued interactively until the error is minimized to an acceptable value. The advantage of ANNs is that they can be used as self-learning parallel working machines. ANNs can be trained using a maximum of repetitive measurements on each of the objects to be compared. The redundant information is then ignored, only the significant information is stored. This saves computer storage capacity. The more that data are used for training, the more precise will be the assignment of an unknown object to one of the predefined patterns. This self-learning capacity increases the flexibility significantly, and diffuse, possibly small but characteristic variance in the spectra can be recognized correctly, and systematic errors in measurements such as noise, bad baselines and so on are trained and will be recognized during the evaluation process, and may thus help to avoid misinterpretations. Unlike cluster analysis, ANN needs the data set to be divided into a training and a test data set. The training should be performed on a representative random sampling of the data population. In order to estimate how effectively the ANN works, the stringent ''leave-one-out'' cross-validation method may be used. Here, validation is performed on the basis of $n - 1$ of the cases (spectra) investigated and the one case left out is tested (reclassified).

6 APPLICATIONS

6.1 Characterization of Particular Cell Compounds and Structures

This section describes the characterization of particular cell components identified and analyzed by specific vibrational bands. In this context resolution enhancement techniques and/or difference spectroscopy are of particular help in identifying compound-specific vibrational

features. IR spectra from a number of microorganisms exhibit characteristic IR bands of variable intensities that cannot be considered to be spectral variations because of experimental conditions or changes in microbiological parameters. A detailed analysis of these spectral features revealed the presence of particular cell constituents such as intracellularly accumulated storage materials, endospores, and cell surface structures like proteineacous and polysaccharidal capsules, or the detection of metabolically released CO_2 in cells and cell cultures.[19,27–29]

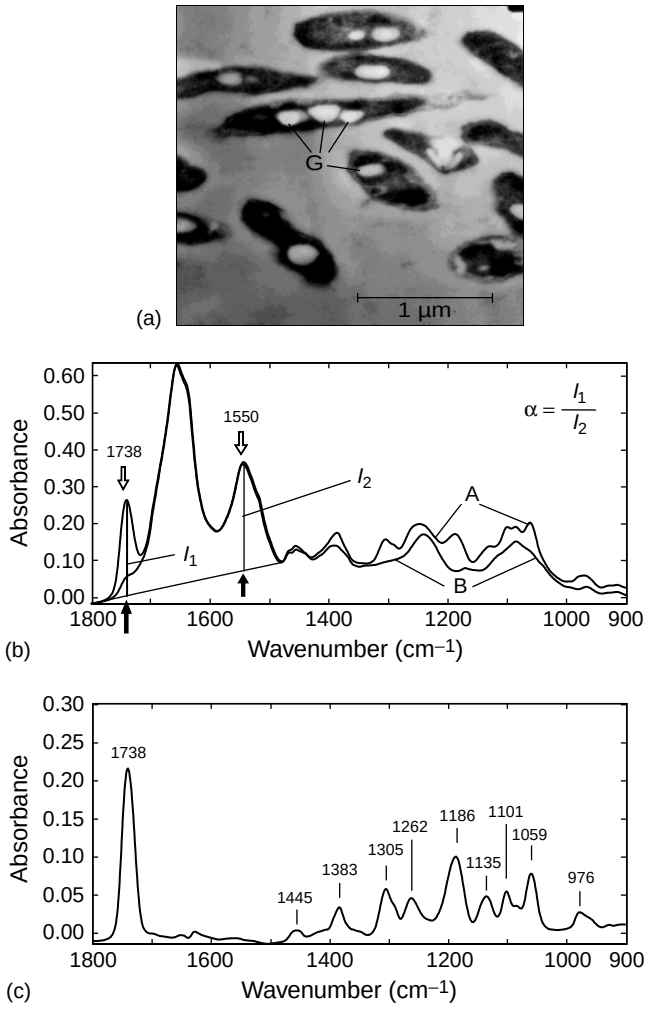

Figure 11 IR spectra of a *Legionella pneumophila* strain (isolate RKI II 8) that produces different amounts of the intracellular storage material poly-β-hydroxybutyric (PHB) acid as a function of growth time. (a) Electron micrograph showing granular PHB inclusion bodies (G); (b) spectra obtained after 48 h (A), and after 120 h (B); (c) difference spectrum (A minus B). Some absorption bands diagnostic for PHB are annotated. The *L. pneumophila* strain was grown on CYE agar plates at 37 °C. The intensity of the ester carbonyl band at 1738 cm^{-1} and the amide II band near 1550 cm^{-1} can be used to determine the relative amount of PHB present in the cells. For data acquisition parameters, see legend to Figure 5.

6.1.1 Cell Storage Materials

Poly-β-hydroxy fatty acids (PHFs), to give an example, are energy and carbon reserve compounds found in many prokaryotes. Generally PHFs are accumulated under the limitation of nutrients when supply of energy and carbon is in excess. Under conditions of starvation, PHFs can be utilized and degraded by the microorganisms helping the cells to survive under severe starvation conditions. It is known that the survival rate is related to the amount of PHFs, which are intracellularly accumulated as small granules. These granules can easily be detected by a light microscope or by electron microscopy (marked by G in Figure 11a). Poly-β-hydroxybutyrate (PHB), to give an example, is frequently found in bacteria (e.g. various strains of *Bacilli, Clostridia, Acetobacter, Legionella*, and *Pseudomonas*). In most microbial IR spectra the ester carbonyl band at 1738 cm^{-1} is only a small weakly expressed shoulder. This band is caused predominantly by the >C=O stretching vibration of ester-bound fatty acids, of the membrane-forming ester-bound lipids. Gram-negative bacteria generally show a stronger ester carbonyl band than the Gram-positive organisms owing to the presence of an additional membrane layer, the outer membrane. In spectra obtained from *Legionella, Pseudomonas* or *Bacillus* strains, for instance, a rather prominent ester >C=O stretching band, accompanied by a number of additional bands between 900 and 1500 cm^{-1}, is occasionally observed which is not permanently present throughout the cell cycle. For some *Legionella pneumophila* strains this band reaches a maximum of intensity after 48 h, while at 120 h of growth practically no additional >C=O ester band can be detected. Figure 11(b) shows the overlaid FTIR spectra of *L. pneumophila* strain RKI/II8 grown for 48 (curve A) and 120 h (curve B), respectively. Inspection of the two spectra reveals characteristic differences, most prominent in the ester carbonyl stretching region around 1740 cm^{-1}. Figure 11(c) gives the difference spectrum as calculated from A and B. This difference spectrum closely resembles the FTIR spectrum recorded from isolated and purified PHB. At least ten bands can be identified and are assigned to a typical polyester compound. The quantitative determination of PHB is achieved by calculation of the ratio $\alpha = I_1/I$ (see Figure 11b), where I_1 is the intensity of the ester carbonyl peak at 1738 cm^{-1} used as a measure marker band of PHB content, and I_2 is the intensity of the amide II peak at 1550 cm^{-1} used as an internal standard which measures approximate total cell mass.

6.1.2 Endospore Formation in Cells

Spore formation in bacteria serves as a strategy for survival under unfavorable conditions. Endospores are resistant to heat, radiation and chemicals, and can

survive starvation. In contrast, vegetative cells spores are not killed by standard sterilization techniques and can cause severe medical problems. The multistage transformation process that occurs within the mother cell has been investigated thoroughly by a variety of molecular genetic, analytical and structural techniques. FTIR spectroscopy can be used to monitor the multiphase process of sporulation, avoiding or complementing more time-consuming procedures like staining, microscopic enumeration and isolation.

The IR spectra of some strains of bacilli and clostridia exhibited some relatively weak extra bands around 1279 cm^{-1}, 767 cm^{-1}, 728 cm^{-1}, 703 cm^{-1}, and 660 cm^{-1} (see Figure 12). Calculation of the difference spectra reveals prominent bands between 1650 cm^{-1} and 1250 cm^{-1}. Light microscopic examination of the same suspensions used for FTIR spectroscopy reveals up to 50% free and mature endospores. Five of the ten most intensive bands of the difference spectrum depicted in Figure 12 are diagnostic for dipicolinic acid (DPA), a compound involved in sporulation. The strong >C=O stretching band of COOH of DPA (expected to be around 1705 cm^{-1}) cannot be discovered in the difference spectra, since chelate binding of Ca^{2+} ions in vivo results in the two stretching vibration bands of the COO^- group at ~1605 cm^{-1} and ~1405 cm^{-1}, respectively. A large proportion of the spore-coat is composed of polypeptides. It is very interesting that the amide I and amide II bands are not found to be increased in spectra from sporulating cells. This is probably because the total amount of protein roughly remains unchanged. It is anticipated that the effect of antibacterial agents on sporulating cells can be accessed quantitatively by calculation of appropriate band ratios, for example the ratio of 1279 cm^{-1} (marker band) and 1545 cm^{-1} (amide II band).[27]

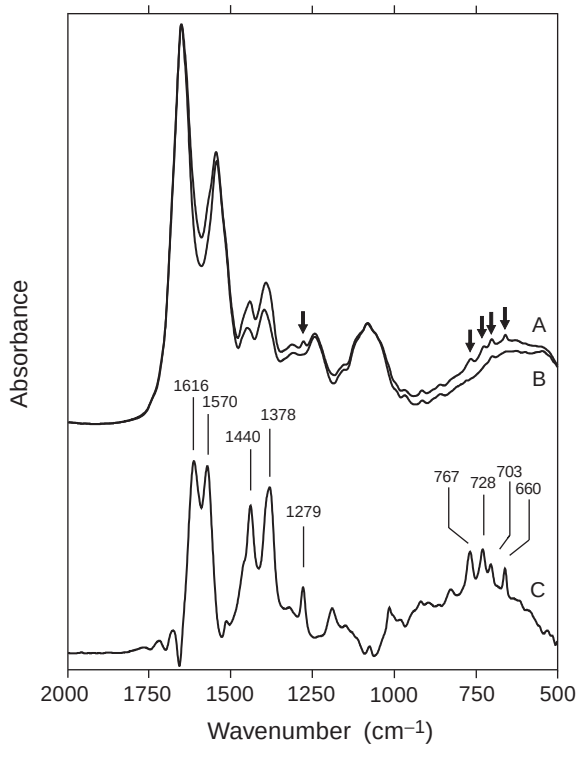

Figure 12 IR spectra of *Clostridium sordelli* (strain DSMZ (Deutsche Sammlung van Mikroorganism und Zellkulturen) 44) which may produce different amounts of endospore as a function of growth time. Spectra were obtained from two populations grown for 48 h (A) and 24 h (B) on Columbia blood agar plates at 37 °C, respectively. The difference spectrum (A) minus (B) is shown by curve (C). Data acquisition parameters, see legend to Figure 5. (This figure has been adapted from Naumann et al.[28])

6.1.3 Metabolically Released CO_2

The detection and quantitation of metabolically released CO_2 in bacteria and yeasts is an additionally interesting aspect of the FTIR technique. Carbon dioxide provided intracellularly or elsewhere, which is usually detected as CO_2 hydrates in water, can be determined extremely sensitively, since the CO_2 band near 2343 cm^{-1} is present in a spectral region where signal-to-noise ratio is optimal and which is usually devoid of overlapping spectral features of the biological material.[29] Figure 13(a) shows the spectra of a fully hydrated sample of a *Candida dubliniensis* isolate (1), of pure water (2), and the difference spectrum 1 minus 2 (3). In the spectral region between 2400 and 2300 cm^{-1}, the asymmetric stretching band of CO_2 hydrates can be detected near 2343 cm^{-1}. Apparently, the CO_2 band monitors distinct levels of metabolic activity in the cells or within the culture. Figure 13(b) shows two series of IR spectra obtained from yeast cultures of *Issatchenkia occidentalis* differing in supplementation, with glucose as a substrate for CO_2 production. Spectra are collected as a function of time in intervals of 8 min after addition of glucose to the yeast culture. A control culture without glucose is measured identically. Using the integral intensity of the peak at 2343 cm^{-1} to quantify CO_2 evolution, the kinetics of CO_2 release after addition of various different substrates to a yeast culture can be sensitively monitored (Figure 13c). Provided that ^{13}C-labeled substrates are available, the simultaneous detection of the kinetics of CO_2 production from two different substrates can be determined as well, since the $^{13}CO_2$-peak is detected near 2277 cm^{-1}, quite separate from the $^{12}CO_2$ stretching band near 2343 cm^{-1}.[19,29]

6.2 Differentiation, Classification, and Identification of Microorganisms by Infrared Spectroscopy

More than 3000 bacterial species have been named systematically with each species comprising 5–1000

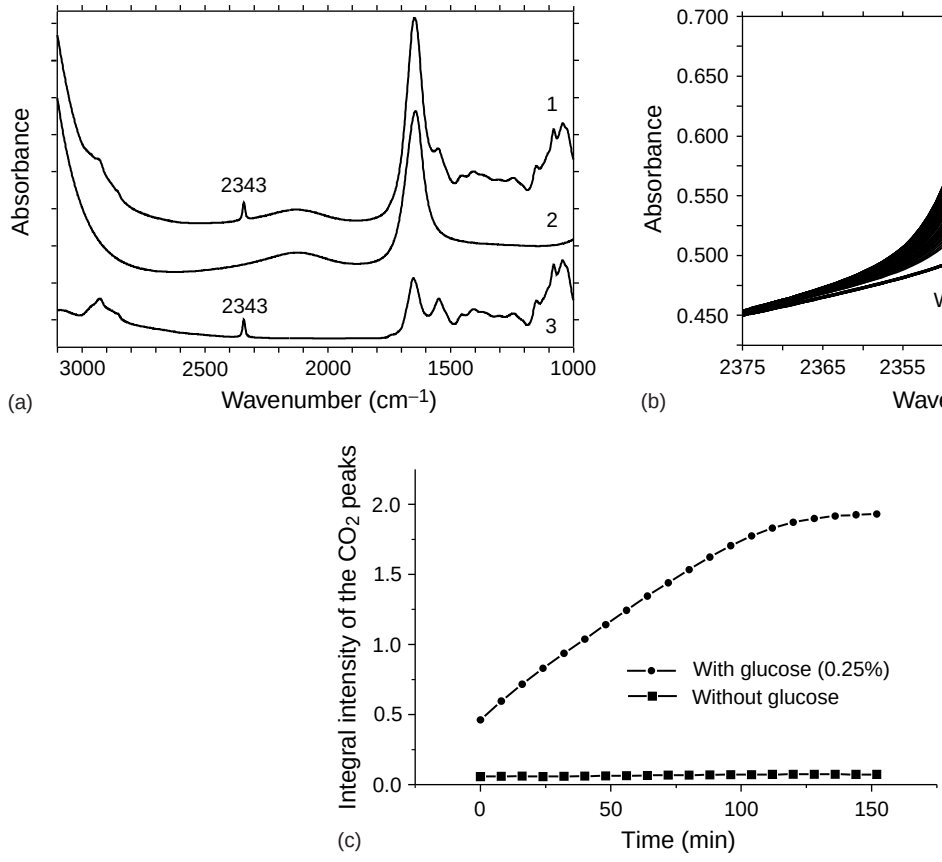

Figure 13 CO_2 production by yeast cells. (a) IR spectra of fully hydrated samples of a *Candida dubliniensis* clinical isolate (strain M467II/97) (1), pure water (2), and (3) the difference spectrum (1) minus (2). (b) Time evolution of CO_2 production after adding glucose to a yeast culture of *Issatchenkia occidentalis*. (c) Kinetic plots of CO_2 production in a yeast culture of *I. occidentalis*. Spectra of (a) were measured in A/T using a cuvette equipped with CaF_2 windows and a pathlength of 8 µm. Kinetic spectra of (b) and (c) were obtained with a 250-µm pathlength using CaF_2 optical windows. Physical resolution: 4 cm^{-1}; number of scans: 64; apodization function: Happ-Genzel; detector: DTGS; spectrometer: IFS 66 (Bruker Optics, Germany).

strains also described by strain numbers (corresponding numbers for yeasts and fungi are not yet available). Many have been described as human, plant or animal pathogenic which emphasizes that most efforts devoted to the isolation and characterization of microorganisms to date are related to the problem of host–cell interaction. Probably there are a million or even more different bacterial species on earth which have not yet been recognized. Taxa, such as families, genera, species, and strains are defined as groups of related microorganisms comprising different levels of discrimination. Modern methods used in microbial taxonomy in addition to classical microscopy, staining techniques, biochemical or enzyme reaction patterns are molecular genetic techniques.[2]

Microbial diversity is always structural and biochemical diversity. IR spectroscopy on intact microorganisms provides information on the structure and composition of the whole cell. Figure 14 shows three typical MIR spectra of microbial samples, a Gram-positive (*Staphylococcus* aureus, strain ATCC 6538), Gram-negative (*Pseudomonas aeruginosa*, strain ATCC 27853), and yeast (*Candida albicans*, strain ATCC 10231) isolate, each dehydrated to a thin film disc on a ZnSe optical plate and measured with the multicuvette system shown in Figure 6(a). While spectrum 3 differs significantly from 1 and 2, the latter two are seemingly similar. However, a more detailed analysis of these two spectra by focusing on selected spectral regions and calculating second derivatives proves the presence of distinct spectral differences between these microorganisms as well (see Figure 14).

Because microbial IR spectra are complex spectroscopic signals encoding the superposition of hundreds or even thousands of bands that cannot be resolved by any means, pattern recognition techniques have to be used which consider the spectra as fingerprints rather than a combination of discrete band intensities, frequencies and bandwidths. Figure 15 shows two dendrograms obtained by hierarchical cluster analysis performed on IR spectra

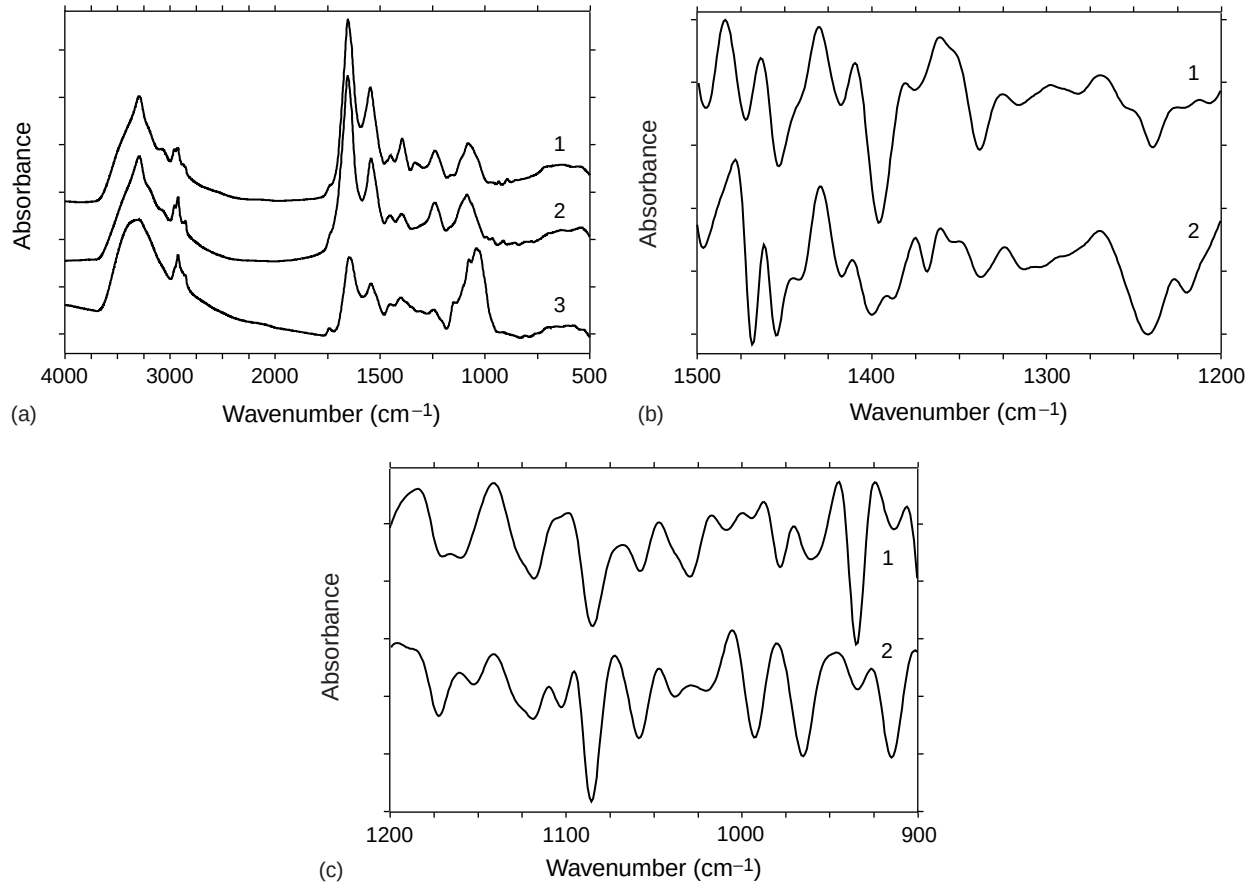

Figure 14 (a) Typical MIR spectra of pathogenic microorganisms: *Staphylococcus aureus* (strain ATCC 6538 (American type culture collection)) (1), *Pseudomonas aeruginosa* (strain ATCC 27853) (2), and *Candida albicans* (strain ATCC 10231) (3). Second derivatives are shown in (b) and (c), respectively. Data acquisition parameters, see legend to Figure 5.

obtained from very diverse microorganisms. It is particularly interesting that the IR technique is not restricted to the analysis of bacteria. Yeasts and fungi can also be analyzed. Figure 15(a) gives the dendrogram of a hierarchical classification trial based on approximately 240 microbial IR spectra obtained from independent measurements on 13 different strains of Gram-positive and Gram-negative bacteria and of 87 spectra from 29 strains of six different *Candida* species.[17] Three main clusters are recognized, which discriminate perfectly between the Gram-positive and Gram-negative bacteria on the one hand, and the yeast strains on the other hand. A closer inspection of the fine structure of the dendrogram shown in Figure 15(a) reveals a perfect discrimination at the species level, as established by molecular genetic techniques. Figure 15(b) shows the result of a separate cluster analysis performed on the different species and strains of *Candida* only using a different set of spectral regions as input for cluster analysis. This dendrogram proves that *Candida* species can be perfectly classified at species level by IR spectroscopy.[30]

As all cell components depend on the expression of smaller or larger parts of the genome, the FTIR spectra of microorganisms display specifically a complete phenetic and a genetic fingerprint of the cells under study. This is why the specificity of the technique is very high, allowing differentiations at quite different taxonomic levels, even down to the subspecies, strain and/or serogroup/serotype level.[6,16–19,28,30] The latter is demonstrated by the FTIR-based classification of different isolates of *Escherichia coli* and some strains of *S. aureus*. Figure 16(a) gives the dendrogram of a classification trial on a selection of serologically different *E. coli* strains. The purpose of this analysis was to group the spectra of the isolates according to their *O*-antigenic structure. *O*-antigenicity is usually determined by serological and sodium dodecyl sulfate–gel electrophoretic techniques. The *O*-antigenic epitopes are structurally determined by the *O*-specific side chains of the LPS which are complex heterooligosaccharide compounds. Therefore, these structures are primarily observed in the spectral region between 700 and 1200 cm^{-1}, where the carbohydrates dominate the

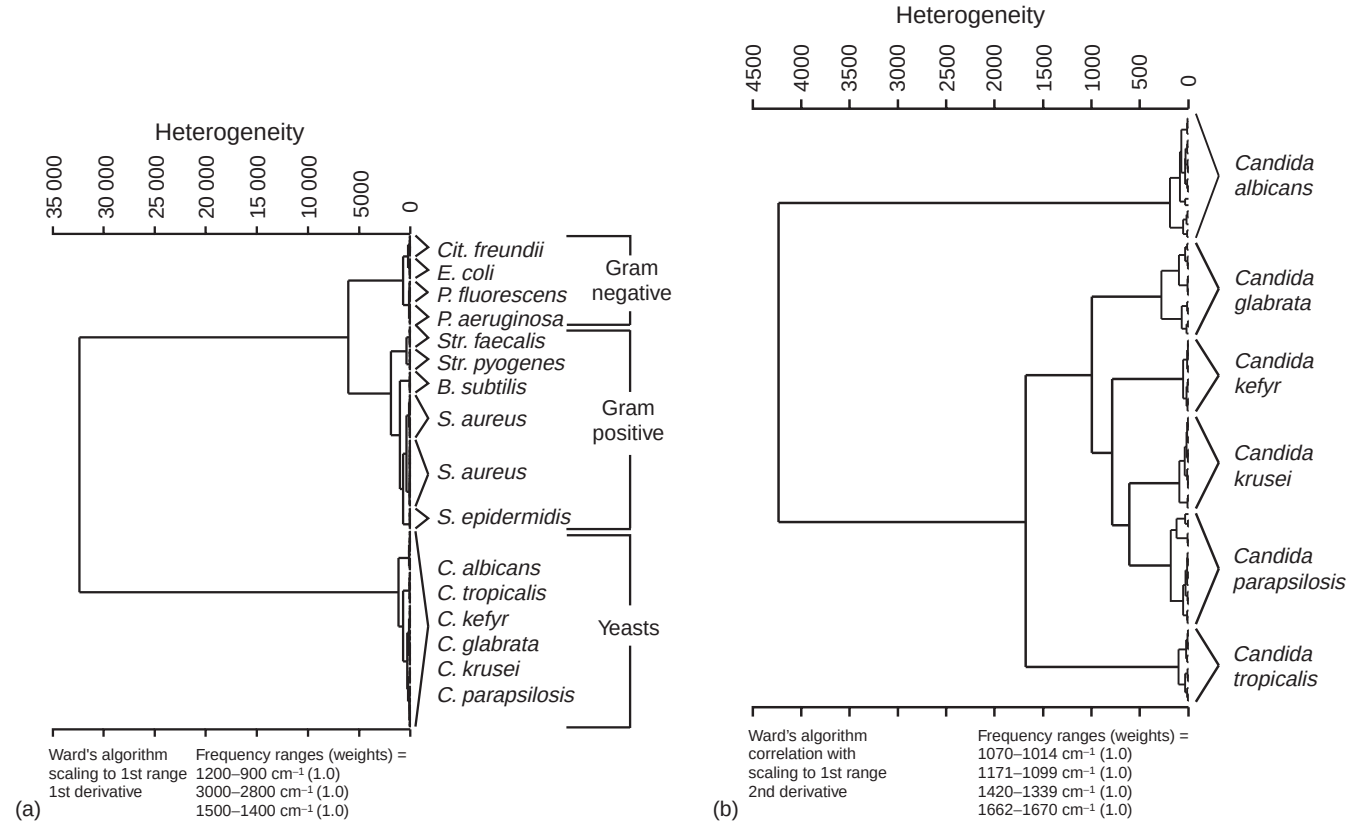

Figure 15 Dendrogram of a hierarchical cluster analysis performed on 240 spectra of different strains of Gram-positive and Gram-negative bacteria, and of yeasts belonging to the genus *Candida* (a). Dendrogram obtained when cluster analysis is performed on the yeast spectra only (b). Cluster analysis was performed using the first derivatives, considering the spectral ranges 2800–3000, 1400–1500, and 900–1200 cm^{-1} (a), and 1070–1014, 1171–1099, 1420–1339, 1662–1670 cm^{-1} (b). All spectral ranges were equally weighted and Ward's algorithm was applied.[17,18] The Pearson's correlation coefficient (*D* values, see Equation 2) was used to calculate the distance matrix. Further experimental and data evaluation parameters, see legends to Figures 5 and 10.

spectral features observed. Using Ward's algorithm for cluster analysis, three main clusters are obtained which perfectly correspond to the O 18, O 25, and O 114 group scheme defined by *O*-serology. Figure 16(b) gives results on hierarchical clustering of IR spectra from different strains of *S. aureus* obtained by manipulating genes related to methicillin resistance of these organisms. Results of cluster analysis corroborate that differentiations can be obtained even at the strain level.

Different algorithms can be used to identify unknown microbial strains on the basis of a reference database. It is pertinent that these reference data sets contain representative numbers of spectra covering all relevant spectral types to be identified. Unknown microbial samples will only be correctly identified with validated reference databases of microbial IR spectra. Identification is then achieved by comparing the IR spectrum of an unknown microorganism with all entries of the reference database. An already tested algorithm is based on the calculation of differentiation indices (*D* values).[18] The spectrum is first subdivided into several spectral windows, selected such that they contain the most discriminative spectral information. The combination of these spectral windows is then used in a stepwise correlation procedure to determine the most similar spectrum contained in the database. Finally a list of most probable hits is reported by the program.[6,17,18,31,32] In this way, IR reference databases for routine identifications are already commercially available from an FTIR spectrometer producing company. These libraries contain thousands of spectra of different species and strains of *Staphylococci, Enterobacteriaceae, Pseudomonas, Bacilli, Clostridia* and yeasts. IR identification data sets are also available for a spectroscopic Gram test and for the identification of waterborne microorganisms.

Identification libraries based on ANN analysis in combination with feature extraction methodologies are used to identify unknown microorganisms.[33] The elaboration of such ANN libraries is divided into several steps including data pretreatment, normalization of data, spectral feature extraction, elaboration of adequate network architectures, training functions, and learning parameters.

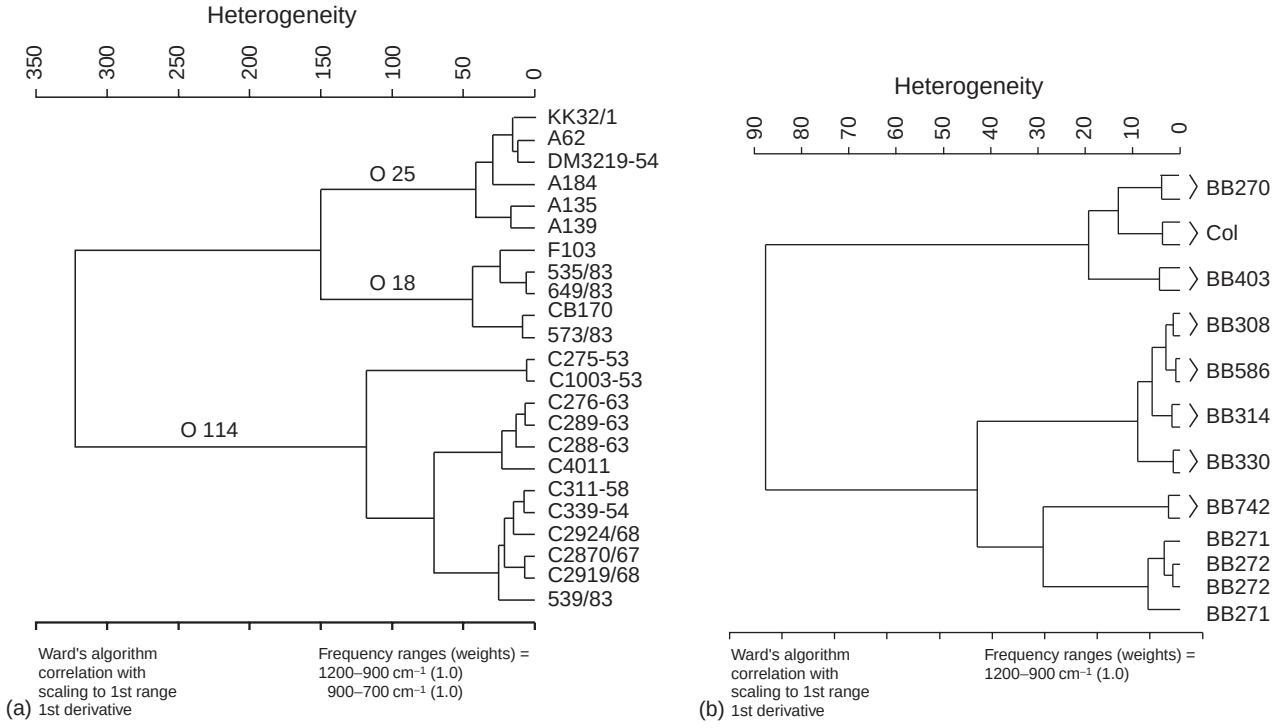

Figure 16 Spectral typing of closely related microorganisms. (a) Clinical isolates of *E. coli* (numbers in right column) belonging to different serogroups: O 25, O 18, and O 114 according to their *O*-antigenic structure. (b) Isogenic strains of *S. aureus* (numbers in right column) differing only in respect of a number of genetic changes influencing methicillin resistance. Hierarchical clustering was achieved applying the Ward's algorithm and using the Pearson's correlation coefficient as the distance measure (*D* values according to Equation 2).[17,18] The first derivatives of the spectra considering the spectral ranges 700–900 and 900–1200 cm^{-1} (a) and 900–1200 cm^{-1} (b) were used, respectively. For further experimental parameters, see legends to Figures 5 and 10. (This figure has been adapted from Helm et al.[17])

The most important idea in this approach is first to establish optimized ANN libraries for subgroups of the whole reference library and then to connect them in the form of a multilayered neural architecture. The user working with such libraries can solve complicated identification problems by running through all layers (neural networks of subgroups) of these reference databases in a controlled multistack procedure. The information flux is controlled by one heading network whose single task is to separate, for example, Gram-positive and Gram-negative bacteria, yeasts, and a number of different genera. In this way a microbial IR identification library is available that comprises six different bacterial genera each containing a representative number of species and strains within the genera *Pseudomonas, Bacillus, Staphylococcus, Streptococcus, Aeromonas, Mycobacteria* and two subnets for the identification of different *Candida* species and for differentiation between fluconazole-sensitive and fluconazole-resistant *C. albicans* strains, respectively.[33]

6.3 Sensitivity Testing by Infrared Spectroscopy

One important additional item of clinical microbiology is antibiotic susceptibility testing. The procedures that have been elaborated for practice are generally based on photometric determination of bacterial growth by light transmission/absorption (optical density) or light scattering (nephelometry). Some promising physical techniques such as radiometry, microcalorimetry, bioluminescence and electrical impedance have not yet reached the stage of general application. IR spectroscopy is also useful in this field, since quantitation of cell mass as a function of antibiotic treatment, as well as the detection of antibiotic-induced structural changes in microbial cells, is well within reach of sensitivity and specificity of IR spectroscopy.[6,34,35] A simple and rapid sampling technique has been elaborated for the detection of time-dependent variations within bacterial cells grown in liquid cultures.[6] A typical protocol of this technique includes the following steps: aliquots of the control and the treated liquid cultures are centrifuged, the cells are washed once or twice, and are again pelleted. The pellet is then quantitatively transferred to an IR transparent plate suitable for A/T-measurement (see Figure 6a).

The principles of this technique can be derived from Figure 17 which compares the spectral information obtained from measurements on a control and on a chloramphenicol-treated culture. Most of the bands

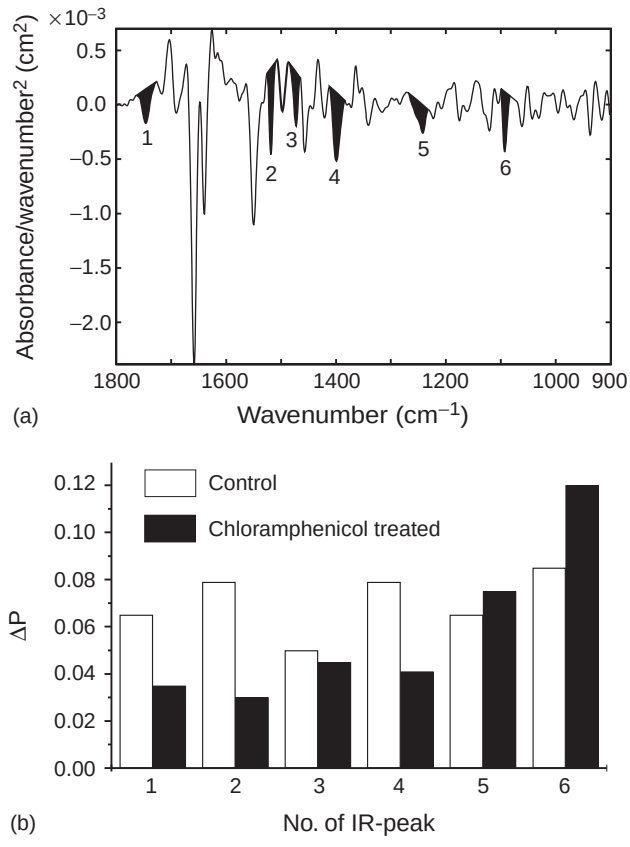

Figure 17 IR bands useful for antibiotic susceptibility testing. (a) Assignment of six IR peaks selected from the second derivative spectrum of *S. aureus* (strain SG 511) sampled from the logarithmic phase of growth: (1) $>C=O$ ester stretching near 1740 cm^{-1}; (2) tyrosine band near 1516 cm^{-1}; (3) $>CH_2$ scissoring band of fatty acids near 1467 cm^{-1}; (4) $-COO^-$ symmetric stretching near 1400 cm^{-1}; (5) P=O asymmetric stretching frequency, for example of $>PO_2^-$ phosphodiester functional groups, near 1245 cm^{-1}; (6) most intensive band of the sugar ring vibrations near 1085 cm^{-1}. (b) Mean change ΔP of peak heights (in arbitrary units) of the six selected IR bands shown in (a) after 30 min of growth of a control culture of *S. aureus* (strain SG 511) (open bars) in comparison to the corresponding ΔP values of a culture treated with 20 $\mu g\ mL^{-1}$ of the protein biosynthesis inhibitor chloramphenicol (black bars). (Adapted from Naumann.[6])

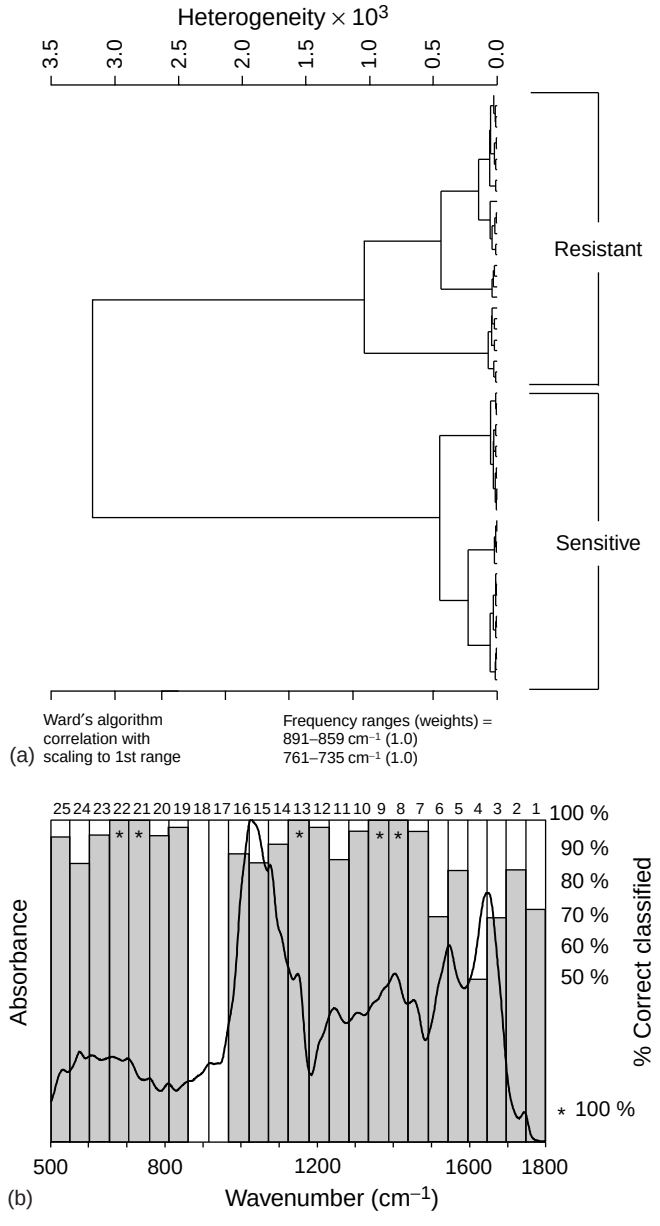

Figure 18 Differentiation between fluconazole-sensitive and fluconazole-resistant strains of *C. albicans* (clinical isolates). (a) Dendrogram of a hierarchical cluster analysis of measurements on seven highly resistant and seven sensitive strains (56 spectra in all). (b) Schematic representation of classification results obtained by linear discrimination analysis on the same number of isolates (see dendrogram of Figure 18a). For linear discriminant analysis the spectrum was subdivided into 25 subregions between 500 and 1800 cm^{-1}. Each subregion was used to test classification by linear discriminant analysis. Regions marked by an asterix yielded 100% correct classification.

shown in Figure 17(a) respond sensitively to antibiotic-induced changes and can thus be used to test drug–cell interaction. Some of them reveal considerably reduced relative peak intensities compared to the control (see Figure 17b). Peaks 5 and 6, assigned to components of the cell wall, in contrast, show increased intensities. This is explained by the reaction mechanism of chloramphenicol (a protein biosynthesis inhibitor), which at this early stage of cell–drug interaction inhibits cell division but not yet cell-wall synthesis. Thus, the analysis of drug-induced changes in bacterial cells is feasible using specific IR marker bands (in case of a protein synthesis inhibitor, for example, the various amide bands). The time of growth, after application of the antibiotic, needed to detect specific changes in the culture can be as little as 30 min.

In some cases, testing for resistance or sensitivity against antimicrobial drugs by IR spectroscopy is possible on the basis of the spectral information alone, i.e. without any application of the antibiotic. This is feasible where the causes and mechanisms of resistance reside somewhere in the cell wall's or in the membrane's altered structure. Figure 18(a) gives an example on seven *C. albicans* strains (clinical isolates) that were known by other techniques to be sensitive and seven strains that were highly resistant to the antimycotic azole derivative fluconazole. The cluster analysis performed on these strains provides a classification scheme that separates the strains into two distinct groups quite well according to what is known a priori. This successful classification could only be obtained by using a few rather narrow spectral ranges that were identified, e.g. by trial and error or by using linear discriminant analysis (see Figure 18b).

6.4 Detection, Enumeration, and Identification of Microcolonies

The light microscope has long been a standard analytical instrument in almost all microbiological research and routine laboratories. While light microscopes provide information on shape, color, and contrast of a given sample, IR spectroscopy may give information about structure and identity of complex samples at the molecular level. Thus, the combination of light microscopy with the sensitivity and specificity of IR spectroscopy provides considerable additional information, in particular the possibility of obtaining structure information on ''what you see''. With the development of modern FTIR spectrometers and increased sensitivities of small detector elements, the coupling of an IR spectrometer to a light microscope was much improved and high quality ''FTIR microscopes'' appeared on the market. This combination pushed the detection limit down to the subnanogram level and opened up the field of spatial resolution and imaging to IR spectroscopy. Because the conventional light microscope uses condensers and objective materials that have limited transparency for IR radiation, the FTIR microscope generally requires all-reflecting optical devices, so-called cassegrainian optics. An assortment of apertures placed in the plane of the real intermediate image of the sample is used to mask the sample areas of interest down to the 10-μm spot of spatial resolution, close to the value restricted by the wavelength of the IR light used.[36]

The microbiological characterization of unknown clinical, food, water, or airborne microbial specimens includes the fundamental steps of detection, enumeration, and differentiation of cells. Different techniques are used to detect, count, and differentiate microorganisms. These techniques include counting colony-forming units, measurement of optical density, use of cell counters or cell sorters and application of a whole arsenal of techniques by which microbial cells can be differentiated including molecular genetic methodologies.

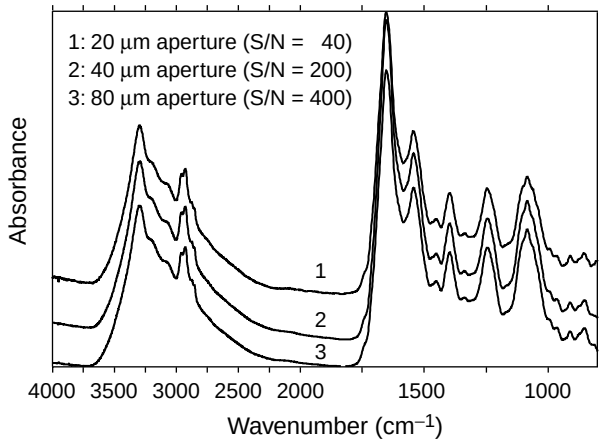

Figure 19 Sensitivity test for a standard FTIR microscope. Spectra shown represent the IR signals from microcolonies containing approximately $1-5 \times 10^4$ cells (spectrum 1), $1-5 \times 10^3$ cells (spectrum 2), and $1-5 \times 10^2$ cells (spectrum 3) using 80, 40, and 20-μm apertures, respectively. Dried colony spots of *E. coli* (strain RKI A139) deposited on a round BaF_2 plate (Ø 25 mm) were measured. The spots were obtained by the stamping technique described in the text. Technique: A/T; number of scans applied: 512; nominal physical resolution: 6 cm^{-1}; apodization function: Blackman-Harris, three-term; FTIR microscope: IRscope I coupled to an IFS 28/B spectrometer (Bruker Optics, Germany); note that signal-to-noise ratios were calculated for the first derivatives with the signal determined at 1600–1700 cm^{-1} (amide I band), and the noise (peak-to-peak) between 2000 and 2100 cm^{-1}.

Some fundamental efforts have been made to integrate detection, enumeration, and identification in one single instrument using IR microspectrometry.[6] When applying IR microscopy to microbial samples, three questions have to be addressed: (1) what are the real detection limits for microorganisms? (2) how can nanogram amounts of microorganisms be sampled and reproducibly measured by the spectrometer? and (3) can detection, enumeration, and differentiation of microorganisms be achieved by combining optical, spectroscopic and computer imaging techniques? The sensitivity of a standard IR microscopic instrumentation is demonstrated in Figure 19 which shows IR spectra obtained from three different microbial spots of a given uniform bacterial biofilm dried down to a BaF_2 optical plate. From the diameter and the film thickness of the spots, the number of bacterial cells detected by the IR beam can be estimated to be of the order of 10^4 (spectrum 1, 80-μm spot), 10^3 (spectrum 2, 40-μm spot) and 10^2 (spectrum 3, 20-μm spot), respectively.

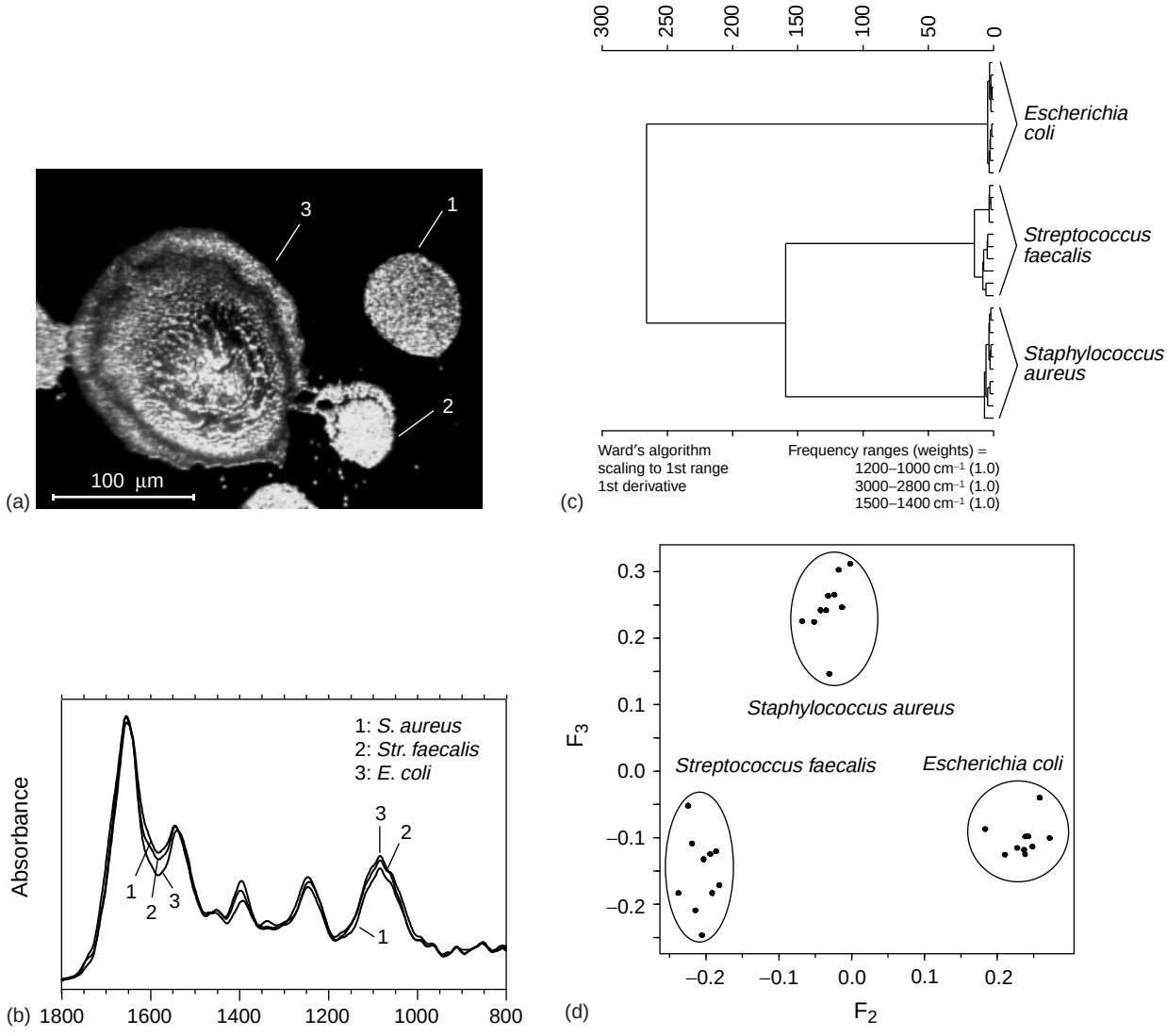

Figure 20 Detection, differentiation, and classification of different microbial microcolonies. (a) Micrograph (magnification approximately 200×) of three different colony spots deposited on BaF_2 windows by the stamping technique described. (b) FTIR species of the microcolonies shown in (a). (c) Hierarchical cluster analysis and (d) factor analysis performed on IR measurements of approximately 30 different colony spots. The groupings suggested by hierarchical clustering are: C_1 = *S. aureus* (strain RKI/WG PS42D); C_2 = *Streptococcus faecalis* (strain DSMZ 20371); C_3 = *E.coli* (strain RKI A139). For instrumental and data acquisition parameters, see legend to Figure 19.

To sample microbial cells for FTIR microscopic measurements, the following experimental protocol has been established.[6] Aliquots of the microbial cell suspension, sufficiently diluted to guarantee single colony growth on solid agar plates, are plated and incubated over a period of 6–8 h. After these growing times, colony-forming units are generally not yet visible to the naked eye. Round IR transparent plates made from BaF_2, CaF_2 or ZnSe are then gently pressed to the agar surface. This imprinting technique transfers spatially accurate small amounts of the microcolonies (two to three microbial layers) on to the plate and provides replica of dried and round microbial colony spots that can be measured by the FTIR microscope. With the aid of a computer-controlled *x*,*y*-stage these spots are measured automatically or operator-controlled using video techniques. Additionally, the number of colony spots can be counted and size classification of these spots is also possible.

Figure 20 gives an example of IR microscope measurements performed on a mixed culture of three different microorganisms, namely *S. aureus, Streptococcus faecalis* and *E. coli*. Figure 20(a), upper panel, shows the micrograph from a selected area of the colony replica, which shows three different microcolonies that can

already be morphologically differentiated by eyes. Typical spectra obtained from these three colony spots are shown in the lower panel of Figure 20(a). These spectra immediately suggest that the three microcolony imprints have indeed been derived from three different microorganisms. The next steps in analysis include (1) representative sampling of FTIR microscopic spectra of bacterial film spots, which (2) are then subjected to MSA. Figure 20(b) and (c) show typical results when using hierarchical clustering (Figure 2b) and factor analysis (Figure 2c). Both evaluation techniques unequivocally prove that the spectra obtained from the microcolonies are indeed characteristic of three different microorganisms present in a mixed culture.

7 PERSPECTIVES

The main advantages of IR spectroscopy which make it attractive are extreme rapidity compared to conventional techniques, uniform applicability to diverse microorganisms, and a high specificity which allows differentiation even down to subspecies levels. The IR methodology requires only low amounts of consumables, it is computer compatible, and may thus promote exchange of results and databases via data nets. The strength of the IR technique is its ability to conduct epidemiological case studies and large screening experiments very quickly. Additional fields of application are the detection of infection chains and the control of therapy, the maintenance of strain collections, and the differentiation of microorganisms from the environment for which established systems are not yet available. In the food, water, and pharmaceutical industries IR spectroscopy may contribute to improve microbiological quality control. For the control of biotechnological processes it is an alternative or additional technique to already existing analytical tools. Drawing upon the knowledge obtained to date, the serial type of a dedicated instrument for IR characterizations of microorganisms is now available from an IR spectrometer producing company. The prospects for the use of FTIR microscopy for microbiological characterization are very promising. This new technology may help to scale down the number of cells needed to a few hundred, to analyze mixed cultures, and to detect light-microscopic and spectroscopic features of microorganisms simultaneously. The development of a fully automated IR spectroscopic system of microbiological analysis that combines detection, enumeration, and identification of microorganisms is already being addressed. The sampling techniques, spectroscopic procedures, and data evaluation strategies elaborated in the context of microbiological IR spectroscopy can easily be carried over to characterize other microorganisms such as amoebae and viruses, and even plant or mammalian cells and tissues. Information density can be increased by combining the spectral traits accessible from the NIR, MIR, and FIR regions of the electromagnetic spectrum and by using IR and Raman spectroscopy in tandem. IR and FT-NIR Raman spectra of nearly identical samples of intact microorganisms can now be obtained with excellent reproducibilities.[37] This makes it possible for the first time to establish a combined use of IR and Raman spectroscopy of complex biological samples, which constantly and notoriously give Raman spectra that are overwhelmed by the fluorescence background using laser excitation in the visible region. The exploitation of the complementarity of IR and Raman spectroscopy particularly may open up new avenues for biomedical applications in the future. A particularly intriguing challenge is to elaborate computer-based and user-friendly pattern-recognition software which gives data reduction, effective feature extraction, and optimal classification results useful for practical purposes collected in the NIR, MIR, and FIR, or by combining the two vibrational spectroscopic techniques of IR and Raman spectroscopy.

ACKNOWLEDGMENTS

The excellent technical assistance of Angelika Brauer in preparing and proof-reading the manuscript is gratefully acknowledged. People who have contributed to this work in the laboratory of the author are acknowledged in the original literature. Particularly acknowledged are Dieter Helm, Christian Schultz, Vesna Fijala, and Siedy Sällström-Baum. The assistance of Maren Stämmler in many sample preparations, measurements and evaluations is much appreciated.

ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Network
A/R	Absorbance/Reflectance
A/T	Absorbance/Transmission
ATCC	American Type Culture Collection
ATR	Attenuated Total Reflectance
DPA	Dipicolinic Acid
DR	Diffuse Reflectance
DTGS	Deuterated Triglycine Sulfate
FIR	Far-infrared
FT	Fourier Transform
FTIR	Fourier Transform Infrared
IR	Infrared
LPS	Lipopolysaccharides
MCT	Mercury Cadmium Telluride

MIR	Mid-infrared
MS	Mass Spectroscopy
MSA	Multivariate Statistical Analysis
NIR	Near-infrared
PCA	Principal Component Analysis
PHB	Poly-β-hydroxybutyrate
PHF	Poly-β-hydroxy Fatty Acid
RL	Reproducibility Levels

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Chemometrics **(Volume 11)**
Clustering and Classification of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Fourier Transform Raman Instrumentation

REFERENCES

1. W.H. Nelson, *Modern Techniques for Rapid Microbiological Analysis*, VCH, New York, 1991.
2. P.R. Murray, E.J. Baron, M.A. Pfaller, F.C. Tenover, R.H. Yolken (eds.), 'Manual of Clinical Microbiology', *Section III: Diagnostic Technologies in Clinical Microbiology*, 6th and 7th Edition, American Society of Microbiology, Washington, 1995, 1999.
3. K.P. Norris, 'Infrared Spectroscopy and its Application to Microbiology', *J. Hyg.*, **57**, 326–345 (1959).
4. J.W. Riddle, P.W. Kabler, B.A. Kenner, R.H. Bordner, S.W. Rockwood, H.J.R. Stevenson, 'Bacterial Identification by Infrared Spectrophotometry', *J. Bacteriol.*, **72**, 593–603 (1956).
5. N.B. Colthup, L.H. Daly, S.E. Wiberley, 'Introduction to Infrared and Raman Spectroscopy', Academic Press, New York, 1975.
6. D. Naumann, H. Labischinski, P. Giesbrecht, 'The Characterization of Microorganisms by Fourier Transform Infrared Spectroscopy (FT/IR)', in *Modern Techniques for Rapid Microbiological Analysis*, ed. W.H. Nelson, VCH, New York, 43–96, 1991.
7. P.R. Griffiths, J.A. de Haseth, 'Fourier Transform Infrared Spectrometry', Wiley (Interscience), New York, 1986.
8. H.J. Rogers, H.R. Perkins, J.B. Ward, *Microbial Cell Walls and Membranes*, Chapman and Hall, London, 1980.
9. P.J. Brennan, H. Nikaido, 'The Envelope of Mycobacteria', *Annu. Rev. Biochem.*, **64**, 29–63 (1995).
10. V. Farkas, 'Polysaccharide Metabolism', in *The Yeasts*, eds. A.H. Rose, J.S. Harrison, Academic Press, New York, 317–366, Vol. 3, 1989.
11. F.M. Klis, 'Review: Cell Wall Assembly in Yeast', *Yeast*, **10**(7), 851–869 (1994).
12. H.J. Phaff, 'Chemotaxonomy on the Polysaccharide Composition of Cell Walls and Capsules', in *The Yeasts. A Taxonomic Study*, eds. C.P. Kurtzman, J.W. Fell, Elsevier Science, Amsterdam, 45–47, 1998.
13. C. Ratledge, C.T. Evans, 'Lipids and Their Metabolism', in *The Yeasts*, eds. A.H. Rose, J.S. Harrison, Academic Press, 367–455, Vol. 3, 1989.
14. D. Li-Vien, N.B. Colthup, W.G. Fateley, J.G. Grasselli, *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press, Boston, 1991.
15. F.S. Parker, *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*, Plenum Press, New York, 1983.
16. D. Naumann, D. Helm, H. Labischinski, 'Microbiological Characterizations by FT/IR Spectroscopy', *Nature*, **351**, 81–82 (1991).
17. D. Helm, H. Labischinski, G. Schallehn, D. Naumann, 'Classification and Identification of Bacteria by Fourier Transform Infrared Spectroscopy', *J. Gen. Microbiol.*, **137**, 69–79 (1991).
18. D. Helm, H. Labischinski, D. Naumann, 'Elaboration of a Procedure for Identification of Bacteria Using Fourier Transform Infrared Spectral Libraries: A Stepwise Correlation Approach', *J. Microbiol. Methods*, **14**, 127–142 (1991).
19. D. Naumann, 'Infrared and NIR Raman Spectroscopy in Medical Microbiology', in *Infrared Spectroscopy: New Tool in Medicine, Proceedings of SPIE*, eds. H.H. Mantsch, M. Jackson, Bellingham, Washington, 245–257, Vol. 3257, 1998.
20. N.J. Harrick, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, Ossining, New York, 1979.
21. M.K. Winson, R. Goodacre, É.M. Timmins, A. Jones, B.J. Alsberg, A.M. Woodward, J.J. Rowland, D.B. Kell, 'Diffuse Reflectance Absorbance Spectroscopy Taking

in Chemometrics (DRASTIC). A Hyperspectral FT/IR-Based Approach to Rapid Screening for Metabolite Overproduction', *Anal. Chim. Acta*, **348**, 273–282 (1997).
22. B.S. Everitt, *Statistical Methods for Medical Investigations*, John Wiley, Toronto, 1994.
23. B.F.J. Manly, *Multivariate Statistical Methods*, a Primer', Chapman & Hall, New York, 1986.
24. B.S Everitt, *Cluster Analysis*, John Wiley, Toronto, 1993.
25. J.E. Dayhoff, *Neural Network Architectures*, Nostrand Reinhold, New York, 1990.
26. J. Zupan, J. Gasteiger, *Neural Networks for Chemists*, VCH, Weinheim, 1993.
27. D. Helm, D. Naumann, 'Identification of Some Bacterial Cell Components by FT/IR Spectroscopy', *FEMS Microbiol. Lett.*, **126**, 75–80 (1995).
28. D. Naumann, C.P. Schultz, D. Helm, 'What can Infrared Spectroscopy Tell us about the Structure and Composition of Intact Bacterial Cells?', in *Infrared Spectroscopy of Biomolecules*, eds. H.H. Mantsch, D. Chapman, Wiley-Liss, New York, 279–310, 1996.
29. C.P. Schultz, H.H. Eysel, H.H. Mantsch, M. Jackson, 'Carbon Dioxide in Tissues, Cells, and Biological Fluids Detected by FTIR Spectroscopy', *J. Phys. Chem.*, **100**, 6845–6848 (1996).
30. É.M. Timmins, S.A. Howell, B.K. Alsberg, W.C. Noble, R. Goodacre, 'Rapid Differentiation of Closely Related *Candida* Species and Strains by Pyrolysis-Mass Spectrometry and Fourier Transform Infrared Spectroscopy', *J. Clin. Microbiol.*, **36**, 367–374 (1998).
31. M. Kümmerle, S. Scherer, H. Seiler, 'Rapid and Reliable Identification of Food-borne Yeasts by Fourier Transform Infrared Spectroscopy', *Appl. Environ. Microbiol.*, **64**, 2207–2214 (1998).
32. M.C. Curk, F. Peladan, J.C. Hubert, 'Fourier Transform Infrared (FTIR) Spectroscopy for Identifying *Lactobacillus* Species', *FEMS Microbiol. Lett.*, **123**, 241–248 (1994).
33. J. Schmitt, T. Udelhoven, D. Naumann, H.-C. Flemming, 'Stacked Spectral Data Processing and Artificial Neural Networks Applied to FT/IR and FT-Raman Spectra in Biomedical Applications', in *Infrared Spectroscopy: New Tool in Medicine, Proceedings of SPIE*, eds. H.H. Mantsch, M. Jackson, Bellingham, Washington, 236–244, Vol. 3257, 1998.
34. W. Bouhedja, G.D. Sockalingum, P. Pina, P. Allouch, C. Bloy, R. Labia, J.M. Millot, M. Manfait, 'ATR-FTIR Spectroscopic Investigation of *E. coli* Transconjugants β-Lactams-Resistance Phenotype', *FEBS Letters*, **412**, 39–42 (1997).
35. G.D. Sockalingum, W. Bouhedja, P. Pina, P. Allouch, C. Mandray, R. Labia, J.M. Millot, M. Manfait, 'ATR-FTIR Spectroscopic Investigation of Imipenem-susceptible and -resistant *Pseudomonas aeruginosa* Isogenic Strains', *Biochem. Biophys. Res. Comm.*, **232**, 240–246 (1997).
36. R.G. Messerschmidt, M.A. Harthcock (eds.), *Infrared Microspectroscopy*, Marcel Dekker, New York, 1988.
37. D. Naumann, S. Keller, D. Helm, Ch. Schultz, B. Schrader, 'FT/IR Spectroscopy and FT-Raman Spectroscopy are Powerful Analytical Tools for the Non-invasive Characterization of Intact Microbial Cells', *J. Mol. Struct.*, **347**, 399–406 (1995).

Infrared Spectroscopy, Ex Vivo Tissue Analysis by

Michael Jackson and Henry H. Mantsch
Institute for Biodiagnostics, National Research Council, Winnipeg, Canada

1 Introduction **1**
2 Experimental Aspects **2**
2.1 Sampling Techniques 2
2.2 Sample Stability: Fixation of Samples 5
2.3 Sample Preparation 6
3 Assignment of Major Absorptions **8**
4 Interpretation of Spectra: Estimation of Sample Histology **10**
5 Nonsubjective Analysis of Spectra **11**
6 Analysis of Microspectroscopic Maps **13**
6.1 Functional Group Mapping 13
6.2 Nonsubjective Analysis of Microscopic Maps 20
7 Summary **24**
Abbreviations and Acronyms **25**
Related Articles **25**
References **25**

Infrared (IR) spectroscopy provides information relating to the vibration of covalent bonds within molecules. The wavelength of light absorbed by a vibrating bond depends upon the atoms in the bond, the type of bond, the type of vibration and inter- and intramolecular interactions. For complex samples such as human tissues an IR spectrum therefore provides a direct indication of sample biochemistry. With the correct choice of sampling methodology (usually an IR microscope) information on the biochemical nature of disease states can be obtained from tissue samples, which can often be useful diagnostically. Variations in spectral signatures arising from nucleic acids, proteins and lipids can provide important information in a number of disease states, including Alzheimer's disease (AD), breast cancer and skin cancer.

The information obtained by IR spectroscopy is difficult to obtain with many other instrumental techniques. For example the signal-to-noise ratio obtained by IR microscopy is far superior to that seen with Raman methods, allowing more sophisticated data processing and so more information to be extracted. Furthermore, many species of interest cannot be studied in situ with other techniques. Nucleic acids are a case in point. Powerful techniques such as nuclear magnetic resonance spectroscopy provide no information concerning this material, while Raman techniques only provide information relating to individual nucleotides. In contrast, DNA and RNA give rise to IR signatures that provide information relating to nucleic acid content/structure.

To appreciate fully the information contained in the complex spectra obtained from human tissues and cells, a unique combination of expertise in spectroscopy, biochemistry and anatomy/histology is required. This combination allows the investigator to avoid potential artefacts due to incorrect sampling and spatial variations in sample composition and to attribute the major absorptions present in spectra to individual biochemical species. However, spectral interpretation is often a highly subjective process, a fact that is made worse when one considers that many thousands of spectra are often acquired from a single tissue section. The application of pattern recognition techniques to IR data removes this subjectivity and allows realistic processing of these large data sets. In addition, many new methods are being developed which allow presentation of these complex data sets in a form readily interpreted by the nonexpert.

1 INTRODUCTION

IR spectroscopy provides information relating to the vibration of covalent bonds within molecules. The wavelength of light absorbed by a vibrating bond depends upon the atoms in the bond, the type of bond, the type of vibration and any inter- and intramolecular interactions, while the intensity of light absorption is directly related (by the Beer–Lambert law) to the concentration of molecules containing the bond. The IR spectrum of a material therefore provides information relating to the quantity, composition, structure and environment of materials in a sample. Furthermore, IR spectroscopy provides this information for all IR active materials present in a sample simultaneously. For complex samples such as human tissues an IR spectrum therefore provides a direct indication of sample biochemistry.[1–3]

Any method that is able to provide information relating to tissue biochemistry can also be used to study and diagnose disease states. This arises from the fact that all disease states are accompanied/caused by alterations in tissue biochemistry. These alterations in tissue biochemistry will result in changes in the IR spectrum of the tissue. A comparison of the spectra

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of healthy and diseased tissues should therefore provide information that can be used to determine the presence of disease and also to understand the disease at the molecular level.[4–6]

IR spectroscopy has a number of advantages over other instrumental methods for the analysis of tissues including speed, sensitivity, flexibility and low cost. Sampling techniques exist to ensure that spectra can be obtained from any type of biological sample, and high-quality spectra may be obtained within minutes on compact low-cost instruments. In addition, no endogenous chromophores (which may potentially perturb the system under investigation) are required. One obvious drawback is the requirement for tissue to be removed from subjects before analysis. However, as biopsies are routinely taken for many pathological conditions, obtaining samples is generally not problematic.

2 EXPERIMENTAL ASPECTS

With all instrumental techniques, determining the most appropriate manner in which to make measurements is one of the most critical steps. Inadequate experimental design may lead to no results, or worse, to misleading results. For example, for many years biophysical studies of proteins and peptides were carried out using organic solvents such as trifluoroethanol (TFE) or dimethylsulfoxide (DMSO).[7,8] The rationale for the use of these solvents was that their organic nature allowed them to serve as membrane mimetic agents, thus allowing studies of peptides in a ''membrane-like'' environment without the membrane (a necessity for nuclear magnetic resonance studies). Typically such studies found that peptides and proteins dissolved in TFE adopted a helical configuration, a finding taken to indicate that in a cell membrane such peptides would adopt a helical configuration. In contrast, peptides dissolved in DMSO were found to be unstructured. However, IR spectroscopic studies have since demonstrated that all peptides and proteins, regardless of native structure, adopt helical configurations in TFE and are unstructured in DMSO. Conclusions drawn concerning native structures of peptides dissolved in DMSO and TFE reflect the experimental design rather than inherent properties of the peptides and proteins and must therefore be treated with caution.

Clearly the complex nature of human tissues requires great care to be taken in designing experiments for the ex vivo analysis of tissues and body fluids by IR spectroscopy. In determining the most appropriate manner in which to make measurements one must consider the range of sampling techniques available, the form of the sample, the nature of the sample, the stability of the sample and the nature of any possible interferents.

2.1 Sampling Techniques

IR spectroscopy is not limited by the physical state of the material under investigation. Methodologies exist which allow spectra to be obtained from solids (e.g. bone, teeth), liquids (e.g. amniotic fluid), slurries (e.g. feces) and particulate suspensions (e.g. blood). However, it should be stressed that care is required in determining the optimal sampling method to be used, perhaps more so in the case of cells, biofluids and tissues than for other types of samples. The complexity of cells, biofluids and tissues present unique challenges to the IR spectroscopist, and these complexities often dictate the choice of sampling technique.

2.1.1 Sampling Techniques for Liquids

Many body fluids are important sources of diagnostic information in medicine, including amniotic fluid, synovial fluid, sweat, urine and blood. Fluid samples are most readily analyzed using transmission techniques. A small volume (typically 5–10 μL) of sample is placed between IR-transparent substrates separated by a thin spacer to provide a reproducible pathlength. It should be stressed that the strength of the absorption bands of water which are seen in the mid-IR region of the spectrum is such that with even moderately thick samples, total absorption of the IR light may occur at a broad range of wavelengths.[9] In general, for biological fluids the pathlength should be kept to 10 μm or less for measurements in the mid-IR. For measurements in the near-IR region of the spectrum, the pathlength may be much greater due to the reduced intensity of the overtone and combination absorptions from water found in this spectral region. In fact extended pathlengths (0.1–1 mm) are not only possible but a necessity in near-IR studies, owing to the reduced intensity of the overtone and combination absorptions of the analytes of interest in the near-IR compared to the fundamentals observed in mid-IR spectra.

In addition to allowing longer pathlengths to be used, near-IR spectroscopy has the additional advantage that spectra are generally simpler than mid-IR spectra. Typically only overtones and combination absorptions arising from vibrations of C–H, N–H and O–H groups are seen in near-IR spectra. This may facilitate identification of species of interest in the absence of absorptions from confounding species. However, it should be borne in mind that this reduced complexity of near-IR spectra also means a reduced information content. For example near-IR spectra do not provide information relating to phosphate or carbonyl groups. When deciding whether to acquire mid- or near-IR spectra, the nature of the information sought should therefore guide the choice.

The analysis of biological fluids by recording transmission spectra from films prepared by drying the fluid on to an IR-transparent substrate is gaining popularity.[9]

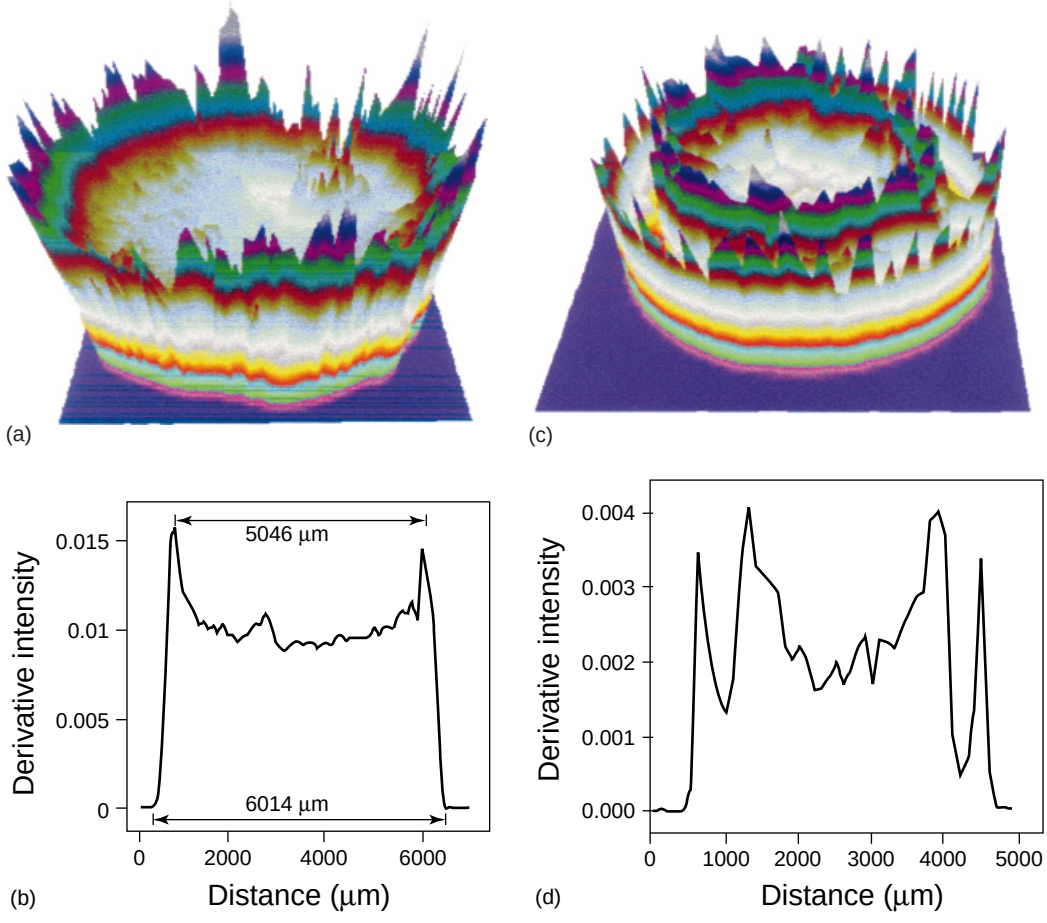

Figure 1 Functional group maps showing the distribution of protein (integrated amide I intensity) within two dry films of amniotic fluid (a) and (c) and cross-sections showing the intensity through the center of the films (b) and (d).

This method has the advantages that the strong interfering absorptions from water are removed and the sample becomes more stable when dried. It should be stressed that any volatile materials present in the sample will be lost upon drying. In addition any information relating to perturbation of the structure of water by analytes in the sample and information relating to hydration is lost.

Investigators should be aware that artifacts associated with heterogeneity of the film upon drying may cause significant problems. For instance, it is possible that upon drying preferential interactions occur between some solutes such that they co-deposit. Alternatively, formation of microdomains of relatively apolar materials may exclude polar material from these areas of the film. In addition, each of the materials present will have a different solubility in the bulk fluid, and so may be deposited at different rates as the fluid dries. IR microscopic studies of films prepared by drying amniotic fluid on to calcium fluoride windows show significant heterogeneity in solute deposition. Functional group mapping (see section 6.1) allows the distribution of protein in dry films to be monitored. The distribution of protein in two amniotic fluid films is shown in Figure 1(a) and (c). Clearly, deposition of protein is not uniform but is concentrated at the edges of the films, most likely as a result of capillary flow during drying. In some films the situation was complicated by the formation of two concentric rings. The reason for the appearance of this second ring is not clear, but it may be related to the rate of drying or surface properties of the window. Additional rings were also seen in the region of air bubbles that formed during drying.

This sample heterogeneity has important repercussions. Variations in the distribution of photons within the IR beam of the spectrometer can potentially have significant effects upon experimental results, due to variations in the area of the film sampled. Similarly, positioning of the sample within the sample compartment will be crucial. Such factors are of particular significance when working with more than one spectrometer and will have a significant impact upon data transferability. To avoid large variations in the distribution of solutes when preparing

films, substrates should be cleaned thoroughly (preferably with a plasma cleaner) before deposition of the film, the fluid should be degassed to prevent formation of bubbles and the film should be dried rapidly under high vacuum. While these precautions will not prevent capillary flow and the formation of a ring of solute at the edge of the film, it will prevent formation of a second, concentric ring or smaller rings formed by air bubbles. It should also be stressed that the fluid viscosity will affect capillary flow and the extent of solute deposition at the edges of films may vary with different types of fluid.

2.1.2 Sampling Techniques for Solids

2.1.2.1 Macroscopic Sampling Techniques for obtaining spectra from macroscopic solid materials include standard transmission, attenuated total reflectance (ATR) and photoacoustic spectroscopy. Each technique has advantages and limitations. Transmission techniques are the simplest available for acquiring spectra of solid materials. Traditional methods of spectral acquisition involve either dispersion of the sample in KBr that is then compressed to form a disk or mounting of the sample directly on an IR-transparent substrate. While dispersion of the sample in KBr is straightforward, it does require the sample to be in powder form. This requires that the sample is physically robust. Obviously, most tissues cannot be analyzed in this fashion. Samples that may be analyzed in this way include calcified tissues such as bone and tooth enamel. However, it should be borne in mind that the process of grinding the sample to form a powder may result in alterations in crystallinity of the mineral matrix in such samples.

Mounting samples on or between IR-transparent substrates is more generally applicable to studies of tissues. However, the strength of the absorption bands seen in the mid-IR region of the spectrum is such that with even relatively thin samples, total absorption of the IR light may occur. In general, for samples containing significant quantities of water, sample thickness should be kept to 10 μm or less. In practice this may be achieved in two ways. For tissues, the simplest method of preparing samples to minimize the pathlength is to compress a small piece of tissue between IR-transparent windows. The windows are generally composed of CaF_2 or BaF_2. To ensure that reproducible pathlengths are obtained, one window may be etched with a 5–10-μm depression. Unfortunately, this simple procedure cannot be used for all samples. Some tissues, such as muscle tissue, cannot be compressed easily between windows. For such samples, compression between diamond windows is required.

Rigid samples such as muscle tissue may also be analyzed by specialized ATR accessories such as the Split Pea developed by Harrick. This consists of a small hemispherical ATR element (typically a 1–2-mm diameter germanium element) on to which the sample is placed. The sample is then compressed against the element with an anvil of some type under reproducible pressure. IR light is then directed through the germanium element, penetrating the sample once and being reflected back into the element and directed towards the detector. While this approach allows samples to be analyzed rapidly, it suffers from the reduced signal-to-noise ratio common to ATR methods. In particular, the fact that the crystal allows only a single reflection into the sample contributes to a reduced signal-to-noise ratio.

The major drawback of these simple approaches to sampling is that the physical integrity of the sample is lost. It then becomes impossible to correlate spectroscopic information with the histology of the sample, a fundamental issue in the analysis of tissues. To maintain the physical integrity of the sample, sections of tissue can be produced with the aid of a cryotome. Thin sections (7–15 μm) are cut and allowed to air dry (which takes only a few seconds) on an IR-transparent substrate. Spectra are then acquired from the dry tissue. This approach has three immediate advantages. First, the interfering absorptions from tissue water are removed. Second, removal of water effectively fixes the sample and prevents degradation of tissue components. Finally, and perhaps most importantly, the sample can be stained following data acquisition. Microscopic analysis of the stained tissue then allows a direct correlation to be drawn between sample histology/composition and spectra.

Occasionally, samples cannot be conveniently prepared by microtoming. For example, bone requires either specialized equipment to prepare sections or must be demineralized. Demineralization is obviously to be avoided in IR spectroscopic studies of bone, as usually the mineral matrix is the component of interest. Similar problems exist when studying teeth. Photoacoustic spectroscopy may be used as an alternative sampling technique. Photoacoustic spectroscopy makes use of the fact that when samples absorb IR light, they undergo a small thermal expansion. In a sealed chamber, this expansion produces a pressure wave that can be detected by a sensitive microphone, and the signal from the microphone can be Fourier transformed to generate a spectrum of the sample. This technique is ideal for irregular samples such as teeth. Furthermore, with a step scan interferometer, depth profiling can be achieved by varying the modulation of the IR light.[10]

2.1.2.2 Microscopic Sampling The complexity of tissues limits the applicability of the measurements discussed above. Tissue heterogeneity is particularly problematic. An IR spectrum represents the sum of the spectra of all of the IR active materials present in the

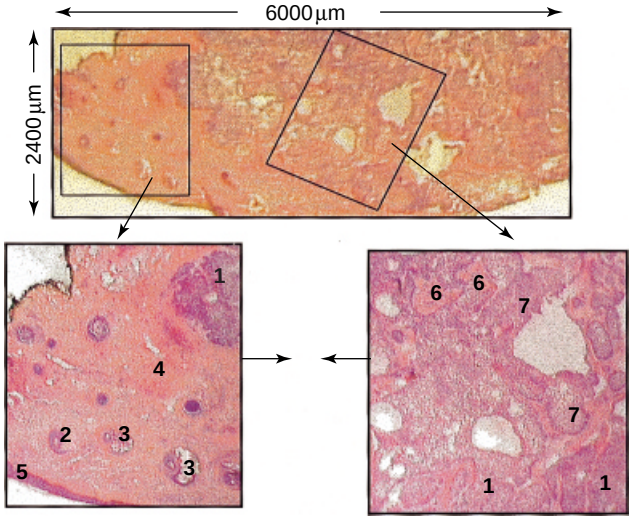

Figure 2 A photomicrograph of a stained section of a skin biopsy containing a basal cell carcinoma. The individual tissue components are labeled, highlighting the complexity of the tissue. 1, Nodular tumor; 2, follicle sheath; 3, sebaceous glands; 4, dermis; 5, epidermis; 6, stroma; 7, infiltrative tumor.

region of tissue through which the IR light passes. While this ''averaging'' is undoubtedly an advantage (discrete probes for individual materials are not required) it may lead to significant problems for the investigator who does not appreciate the heterogeneous nature of tissues. The composition of all tissues varies both spatially and temporally. Temporal variation may arise due to cyclical variations in tissue composition (for example due to hormonal effects) or as a consequence of aging. Spatial variability arises due to the heterogeneous distribution of tissue components within a sample. This spatial variability is illustrated clearly in Figure 2, which shows a photomicrograph of a stained section of a skin tumor. The sample is clearly composed of a number of distinct tissues, including the epidermis, dermis, hair follicles, sebaceous glands and tumor. A spectrum of the entire section (a macroscopic spectrum) will therefore contain contributions from all of these tissues, weighted with respect to their relative proportions in the sample. Thus, spectral features arising from one or more of the other tissue components may mask spectral features characteristic of the tumor.

To obtain information relating to the spatial distribution of materials within sections and to obtain spectra of pure tissue components the technique of IR microscopy is used. In its most basic form, an IR microscope is a beam condenser that reduces the diameter of the IR beam, focusing the IR light on to a small area of sectioned tissue.[11] This allows high-quality spectra to be obtained from extremely small samples ($30 \times 30\,\mu m$). In fact, with the use of highly focused synchrotron radiation, spectra may be acquired from samples as small as $3 \times 3\,\mu m$. By coupling an IR microscope to a high-precision computer-controlled stage the automated collection of IR spectra at each position in a two-dimensional grid defined over the tissue section becomes possible. Information collected at each point (pixel) is combined to produce an IR spectroscopic map of the tissue section. More recently, IR focal plane array detectors have been coupled to step-scan interferometers equipped with IR microscopes.[12–14] In this case, multiple detector elements enable spectra at each pixel to be collected simultaneously. The focal plane array detector method is much faster than the traditional mapping technique, but can suffer from poor signal-to-noise ratios below $1400\,cm^{-1}$, which to date has limited applicability. With both techniques the end result is the same; large spectral maps are collected containing thousands of spectra per map with enormous amounts of both spatial and spectral information.

While a wide variety of sampling methodologies exist, it is becoming increasingly clear that IR microscopy is the sampling technique of choice. The ability to obtain spectra nondestructively with high spatial resolution and precision allows extremely detailed studies of tissues to be undertaken. Again, the ability to stain samples subsequently and compare histological information and spectroscopic information provides a significant advantage. Given these advantages and the almost universal acceptance of IR microscopy as the measurement technique of choice, the following sections will focus predominantly on acquisition, processing and interpretation of microspectroscopic data.

2.2 Sample Stability: Fixation of Samples

In all studies involving biological materials, questions of sample degradation over time arise. Of particular importance in this respect is the degradation of tissues and fluids in the brief period immediately following removal from the body. In this period, cells continue to metabolize but metabolic fuels are not replaced. Cellular disturbances during this period may result in spectral changes that may be confused with alterations in biochemistry caused by disease processes. With cell suspensions and tissues, it is possible to prevent this degradation by fixation with preservatives such as formalin (an aqueous solution of formaldehyde) or 70% ethanol in water. However, common fixatives such as ethanol/water and formalin might be expected to exhibit strong IR absorptions, and thus themselves be potential sources of artifacts in spectra. Furthermore, these fixatives preserve tissue by inactivation of degradative enzymes, which may also introduce artifacts into spectra.

The effect of ethanol/water on spectra of cultured human melanoma cells dried 30 min after suspension in

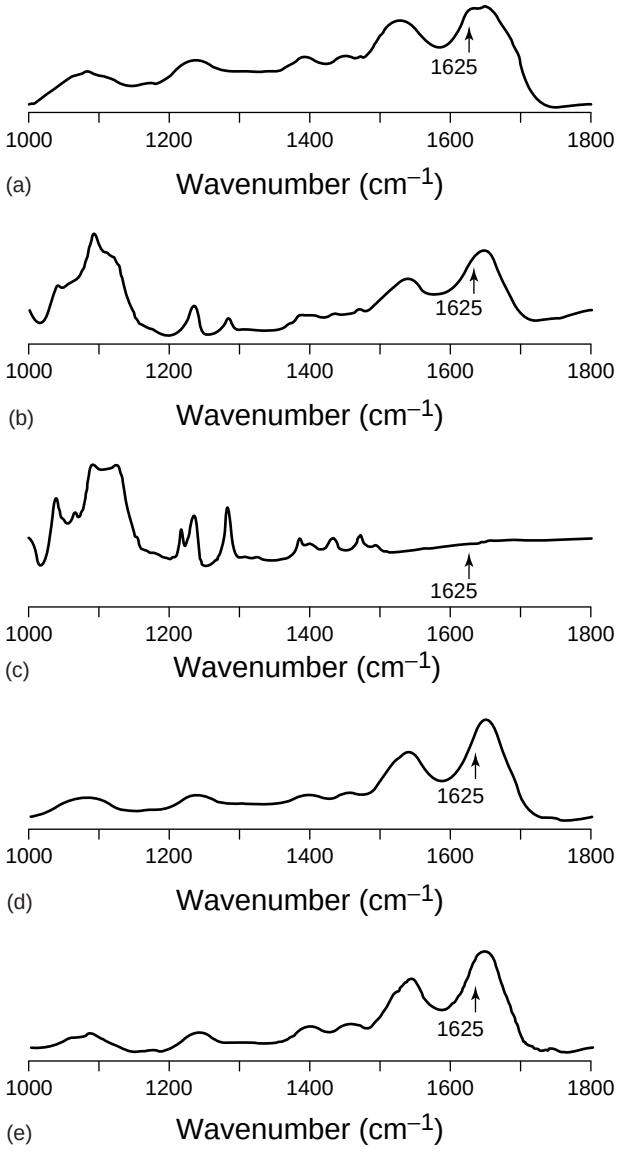

Figure 3 IR spectra of (a) cultured human melanoma cells dried from 70% ethanol in water, (b) cultured human melanoma cells dried from formalin solution, (c) a film prepared by drying formalin, (d) cultured human melanoma cells suspended in formalin, washed in saline and dried and (e) cultured human melanoma cells dried from saline.

70% ethanol is shown in Figure 3(a). It is immediately apparent that suspension of the cells in ethanol significantly alters the spectrum compared to suspension in isotonic saline (Figure 3e). Most noticeably, a pronounced shoulder is seen at 1625 cm^{-1}, with a second shoulder apparent at 1680 cm^{-1}. The appearance of such spectroscopic features is highly indicative of the formation of aggregates of protein molecules stabilized by intermolecular hydrogen bonding, and is typical of spectra of proteins in aqueous alcohol mixtures.[7,8,15] This aggregation is a result of the disruption of intramolecular hydrogen bonds within protein molecules, resulting in large-scale protein denaturation. Presumably this protein denaturation and aggregation forms the basis for the fixative properties of alcohol.

The intensity of the absorptions at 1625 and 1680 cm^{-1} is time dependent, increasing with prolonged suspension in the alcohol due to continued penetration of the alcohol into the cells. Fixation with ethanol can also be seen to reduce the intensity of the ester C=O stretching vibration at 1740 cm^{-1}, suggesting a decreased lipid content in ethanol-fixed cells. This reduction in lipid content reflects solubilization of membrane lipids by ethanol. Large-scale protein denaturation and delipidation obviously make ethanol unsuitable as a fixative for IR spectroscopic studies.

The spectrum of melanoma cells dried from neutral buffered formalin is shown in Figure 3(b). The characteristic absorptions associated with hydrogen bond-stabilized aggregates are not present in this spectrum, indicating that formalin fixation does not induce unfolding of proteins. This is consistent with the known mechanism of action of formalin, which is the promotion of methylene bridge formation between lysine side chains, leaving the structure of the proteins essentially intact. While drying of cells from formalin does not appear to have significant affects upon the major protein absorptions, it does result in the appearance of a series of new absorptions between 1000–1500 cm^{-1}. Analysis of the spectrum of a film prepared by drying neutral buffered formalin (Figure 3c) suggests that these absorptions arise from formaldehyde. It may be thought surprising that formaldehyde, a gas, is retained upon drying of the formalin. However, it should be borne in mind that water, as water of hydration, can be trapped in the salt crystals that form during drying, and it is likely that the salt crystals will also contain traces of the formaldehyde which was dissolved in this water. These distinct absorptions are removed upon washing the cell suspension with isotonic saline before drying (Figure 3d), resulting in spectra essentially identical to those obtained from cells suspended in saline (Figure 3e). Thus, if fixation of samples is desired, then formalin should be the fixative of choice for the vibrational spectroscopist and samples should be rinsed thoroughly with isotonic saline prior to analysis.

2.3 Sample Preparation

2.3.1 Preparation of Cell Samples

Preparation of cell samples for IR spectroscopic analysis is relatively straightforward. In general, cells are usually studied as dry smears. As discussed above this has the advantage of halting all biochemical reactions within the cells, effectively fixing the sample. If chemical fixation is

required, as we have seen, formalin should be the fixative of choice and samples should be carefully washed with isotonic saline before analysis.

Preparation of the cell smear is the most crucial aspect of sample preparation, and care should be taken to avoid deposition of cells in thick multilayers, which results in a long pathlength for spectroscopic measurements. Such long pathlengths will result in strong absorption of IR light and introduce artifacts into spectra caused by detector nonlinearity. The simplest method of avoiding formation of dense regions of cells is to dilute the cell suspension before preparing the smear. Alternatively, some investigators deposit monolayers of cells on the surface of the substrate by centrifugation using specialized rotors. Whichever technique is employed, this step is essential to ensure reproducible artifact-free spectra.

In some circumstances it may be of interest to study cells in an aqueous environment. In such circumstances it is important to remember to keep cells thermostated at 37 °C. Not only does this ensure that physiologically relevant information is obtained, but thermostating the sample holder when acquiring both the sample and buffer spectra allows correct compensation/subtraction of the aqueous buffer. Samples may then be analyzed in aqueous buffers containing labeled substrates or pharmacologically active substances and spectral changes induced by these substances may be monitored.[16]

2.3.2 *Preparation of Tissue Samples*

Samples of tissue are prepared for IR microscopy in much the same way as for light microscopy, with some important exceptions. Samples being prepared for light microscopy are normally embedded in a liquid hydrocarbon (such as paraffin wax or products such as optimal cutting tool, or OCT), cooled to between −10 and −20 °C and sectioned using a cryotome. Typically, sections are 10–20-μm in thickness. The section is then mounted on a glass slide, stained and protected with a glass cover slip. The purpose of the hydrocarbon matrix is simply to maintain the physical integrity of the sample during sectioning (cooling the sample causes the matrix to solidify). However, such materials will produce strong IR absorption bands in spectra and should not be allowed to contaminate tissues prepared for IR microscopy. Rather than completely embedding the tissue, a small amount of embedding material should be carefully placed on one end of the tissue block. This should then be used to attach the tissue block to the cooled chuck of the cryotome. Solidification of the embedding material upon contact with the cooled chuck will effectively mount the tissue for sectioning. Sectioning in this manner may result in some loss of physical integrity (predominantly due to shrinkage of tissue during drying on the substrate and manifest as small holes in the tissue section) but not usually sufficient to impair analysis.

In general, once the tissue is mounted, sections should be cut to a thickness of no more than 10 μm. If the section is too thick, the intensity of resultant absorbance bands in some spectral regions (particularly the amide I region of the spectrum) will be such that problems due to detector nonlinearity will result. For most tissues, we find 7-μm sections result in acceptable spectra with adequate signal-to-noise ratios and with absorbance values in the linear range of the detector. However, sample thickness obviously depends upon the material under investigation. A good rule of thumb is to try and keep absorbance values in the range 0.4–0.6 absorbance units.

Once sectioned, light microscopists typically mount tissue on glass slides. While glass is obviously transparent to visible light, most (all wavelengths below 2500 cm^{-1}) mid-IR light is absorbed by glass. Thus, glass should be used as a substrate only if the investigator is solely interested in near-IR spectra or the CH_2, CH_3 and NH stretching absorptions in the mid-IR. In all other cases, the standard materials used for transmission spectroscopy (CaF_2, BaF_2) should be used. Obviously, water-soluble materials such as NaCl are to be avoided. A few investigators mount tissue sections on gold-coated slides and acquire reflectance–absorption spectra. However, this approach may lead to the introduction of artifacts into spectra and as adequate spectra are usually achieved with standard transmission microscopy, we suggest that this approach be avoided.

Having sectioned and mounted the tissue, light microscopists next employ stains to allow visualization of tissue components. Staining is itself a complex multistep procedure which involves not only the addition of stains to samples but also immersion of the sample in a variety of organic solvents, including various concentrations of ethanol in water and xylene. The final result is a tissue section that shows differential uptake of the stain used, allowing discrimination between tissue components. It is tempting to take this approach in IR microscopic studies, as it allows complete identification of tissues in the sample before spectra are acquired.[17] However, it should be stressed strongly that under no circumstances should tissues that have been washed in xylene and/or ethanol/water be used for IR studies, because of the potential for artifacts due to delipidation and protein denaturation.

The introduction of stains into the sample may also introduce artifacts, in three possible ways. First, the stain itself may exhibit strong spectral features that confound the analysis. Second, even if the stain has no discernible IR absorptions, the interaction of the stain with species within the sample may result in spectroscopic changes. Finally, it is also possible that very weak effects not

discernible to the naked eye (either due to the stain itself or indirectly due to interactions with materials within the tissue) may bias advanced statistical analyses such as multivariate pattern recognition analysis. Such methods are capable of detecting subtle differences between tissues (see below) and thus may be affected by differential accumulation of stains within different tissue components.

The message then is simple: keep sample preparation simple and to a minimum.

3 ASSIGNMENT OF MAJOR ABSORPTIONS

Assuming the correct sampling technique has been chosen and the sample has been processed correctly, generation of large quantities of data becomes straightforward. However, interpretation of this data is not always straightforward, and relies heavily upon an understanding of the spectroscopic properties of the individual constituents of the sample. Of necessity, this assignment of IR absorptions to specific chromophores arising from individual chemical species in human tissues is largely empirical, relying upon extrapolation of IR spectra of the major components of human tissue (lipids, proteins and polynucleotides). The IR spectrum of most tissues should be closely approximated by summation of these three spectral types, weighted according to relative concentrations, and so the major spectral features characteristic of each of these building blocks and the information provided by these absorptions will be discussed. Spectra of a typical lipid, protein and a polynucleotide are shown in Figure 4 and the major peaks assigned below. It should be noted that these spectra were acquired from dry films of material to enable full-range spectra to be shown (i.e. with no interference from water). As such, the frequencies reported are those observed in the dry state, and may vary with hydration.

The spectrum of a typical phospholipid (dimyristoylphosphatidylcholine, DMPC) is shown in Figure 4(a). The dominant absorptions in this spectrum, as in the spectra of all lipids, are found in the region 2800–3000 cm^{-1} and are assigned, by analogy with IR spectra of alkanes, predominantly to the asymmetric and symmetric stretching vibrations of CH_3 (2956 and 2874 cm^{-1}) and CH_2 (2922 and 2852 cm^{-1}) groups of the acyl chains.[18] As expected given the greater number of CH_2 groups, the intensity of the CH_2 absorptions is some 10–20 times that of the corresponding CH_3 absorption. The frequency of the CH_2 stretching absorptions provides a useful probe of lipid bilayer order, low frequencies being associated with a higher degree of conformational order.

In spectra of nucleic acids (Figure 4b) and proteins (Figure 4c) the intensity of absorptions in this spectral

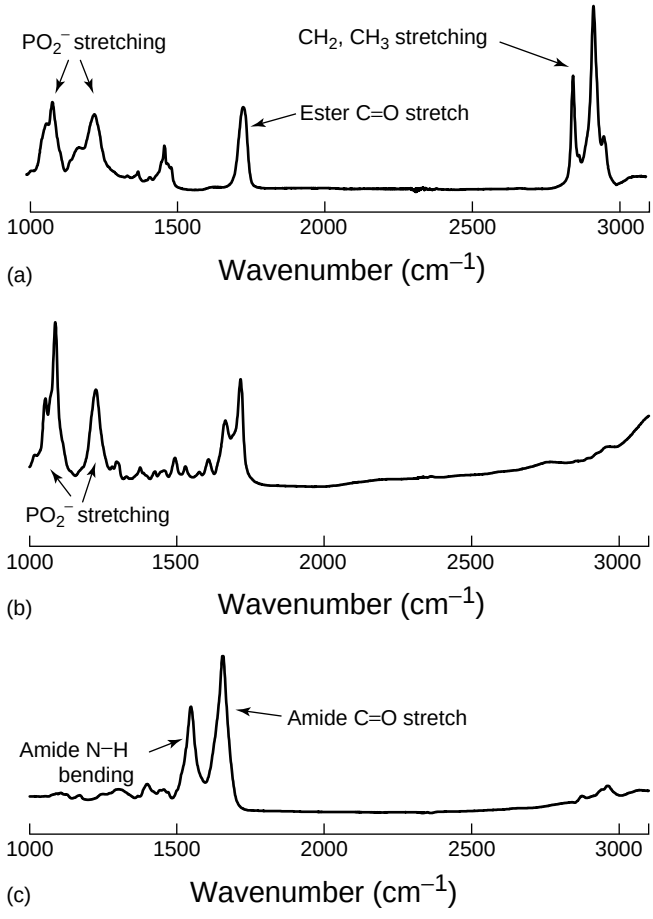

Figure 4 IR spectra of dry films of (a) dimyristoylphosphatidylcholine (DMPC), (b) a polynucleotide and (c) hemoglobin.

region is much reduced. No CH_3 stretching absorptions are seen in spectra of polynucleotides and only weak absorptions due to the CH_2 stretching vibrations of the carbohydrate residues and the C–H stretching vibrations of the bases of the nucleotides are seen. Weak features arising from the CH_2 and CH_3 stretching vibrations of the amino acid side chains dominate the spectrum of proteins in this region. The ratio of the relative intensities of the CH_2 and CH_3 stretching absorptions is much reduced compared with that seen for DMPC and reflects the almost equal proportion of CH_2 and CH_3 groups which occur in protein side chains.

The region between 1750 and 2800 cm^{-1} is generally free from absorptions in biological materials, the major exception being the combination band of water in wet tissue and fluids and a sharp absorption from CO_2 produced by cellular metabolism which is seen in cell suspensions and biological fluids (for the purposes of this discussion we will assume that the spectrometer is adequately purged and no residual water vapor or atmospheric CO_2 is present). Between 1600 and

1800 cm^{-1}, absorptions attributed to C=O stretching vibrations are typically found. In phospholipids such as DMPC a strong absorption at 1736 cm^{-1} arises from the stretching vibration of the ester C=O groups of the lipid. The frequency of this absorption is strongly affected by hydration (hydrogen bonding to water). In nucleic acids two major absorptions are seen which may be attributed to C=O stretching vibration of the purine and pyrimidine bases (1717 and 1666 cm^{-1}). The absorption at 1717 cm^{-1} is a sensitive probe of base pairing in nucleic acids. In proteins the major absorption in this region is the so-called amide I band, which arises predominantly from the C=O stretching vibration of the amide C=O group. It has been repeatedly demonstrated that this absorption is conformationally sensitive and can be used to predict protein secondary structure in model studies.[13]

The amide II (predominantly an N–H bending vibration coupled to C–N stretching, seen between 1500 and 1560 cm^{-1}) and the amide III (attributed to a complex vibration involving C–N stretching and N–H in plane bending, often with significant contributions from CH_2 wagging vibrations, which occurs at 1220–1350 cm^{-1}) absorption bands are also used to assess protein conformation. However, as they arise from complex vibrations involving multiple functional groups they are less useful for protein structure prediction than the amide I.

The only other significant absorptions in the spectra of most proteins are attributed to side chain vibrations such as the COO^- asymmetric and symmetric stretching vibration of the acidic amino acids aspartate and glutamate (1580 and 1400 cm^{-1}). Collagen is a noticeable exception: in addition to a prominent series of absorptions arising from the amide III/CH_2 wagging vibrations, two distinct absorptions at 1030 and 1080 cm^{-1} are seen which are attributed to C–O stretching vibrations of the carbohydrate moieties attached to this important structural glycoprotein. In fact, the series of absorptions at 1030, 1080, 1204, 1240, 1280 and 1330 cm^{-1} provide a highly characteristic fingerprint that can be used to monitor the connective tissue matrix in tissues.[19,20]

In nucleic acids the region between 1000 and 1500 cm^{-1} is populated by a number of sharp but weak absorptions. The major absorptions arise from vibrational modes of the phosphate groups of the phosphodiester linkages of the polynucleotide chain and are assigned to PO_2^- asymmetric (1224 cm^{-1}) and symmetric (1087 cm^{-1}) stretching vibrations.

Phospholipids exhibit similar phosphate absorptions to those shown by nucleic acids. The frequency of absorptions arising from the PO_2^- groups of phospholipids can provide important information concerning headgroup hydration. In addition to phosphate absorptions, nucleic acids and phospholipids show a relatively strong absorption in this region arising from C–O–C stretching vibrations at 1050–70 cm^{-1}. The only other major absorption in spectra of phospholipids arises from the CH_2 scissoring vibration of the acyl chains (1469 cm^{-1}). A series of weak absorptions between 1250 and 1400 cm^{-1} are assigned to the wagging vibrations of the CH_2 groups of the acyl chains.

In addition to lipids, proteins and nucleic acids, carbohydrates are present in all cells. The major absorptions from carbohydrates are found in the 1000–1200 cm^{-1} region of the spectrum and are attributed to C–O stretching vibrations. However, significant carbohydrate absorptions are only expected in those tissues with large carbohydrate stores (e.g. liver, cervix) or with a high glycoprotein (e.g. collagen) content.

Assignments of the major absorptions in biological tissues are summarized in Table 1. However, it should be stressed that these frequencies are offered as guidelines only, and that the recipe for assignment of IR absorption bands to specific chromophores involves an equal proportion of histological and spectroscopic understanding, liberally seasoned with common sense. The next section will highlight how knowledge of histology and spectroscopy can be combined successfully to aid in the interpretation of tissue spectra.

Table 1 Representative frequencies of the major IR absorption bands in tissues and cells

Absorption peak (cm^{-1})	Assignment
3290	Amide A (N–H stretch); protein
3050	Amide B (N–H bending 1st overtone); protein
3010	Olefinic C–H stretch; lipids
2956	CH_3 asymmetric stretch; lipids, proteins
2925	CH_2 asymmetric stretch; lipids, proteins
2873	CH_3 symmetric stretch; lipids, proteins
2854	CH_2 symmetric stretch; lipids, proteins
1735	Ester C=O stretch; lipids
1655	Amide I (protein C=O stretch), α-helices
1636	Amide I (protein C=O stretch), β-sheet
1545	Amide II (protein N–H bending)
1515	Tyrosine ring breathing vibration (C–C/C=C stretching)
1467	CH_2 bending; lipids, protein
1400	COO^- symmetric stretching; proteins, fatty acids
1380	CH_3 bending; lipids, proteins
1280	Amide III of collagen
1220–1240	PO_2^- asymmetric stretch; nucleic acids, lipids
	Amide III of collagen
1204	Amide III of collagen
1000–1200	C–O stretching; carbohydrates
1080	PO_2^- symmetric stretch; nucleic acids, lipids

4 INTERPRETATION OF SPECTRA: ESTIMATION OF SAMPLE HISTOLOGY

Using the knowledge gained from studies of isolated biological materials, assignments of the major absorptions in IR spectra of tissues can be made. While these assignments are of course empirical, they can be made with some confidence when coupled with knowledge of potential sample histology and tissue function. Spectra of breast tissue and liver tissue will be used as illustrative examples (Figure 5).

Clear differences can be seen not only between the spectra of breast[21] and liver tissue,[22] but also between different samples of the same tissue. These striking differences must be due to differences in composition. By attributing the major absorptions in each spectrum to specific chemical species in accordance with the assignments discussed above, the nature of these compositional differences can be suggested. Knowledge of the histology of the tissue under investigation can then be used to confirm or reject these proposed compositional differences. For example, the spectrum of breast tissue in Figure 5(a) does not exhibit the characteristic amide I absorption at around 1650 cm^{-1} typical of protein-rich tissues. Breast tissue is comprised predominantly of epithelial cells, connective tissue (predominantly type I collagen) and adipose tissue. As epithelial cells and connective tissue both contain significant quantities of protein, the lack of an amide I absorption in Figure 5(a) clearly suggests that the tissue cannot contain either epithelial cells or connective tissue. Logically the tissue must therefore consist predominantly of adipose tissue. The presence of strong absorptions arising from CH_2 stretching (not shown), C=O stretching, CH_2 bending and C–O–C stretching vibrations clearly point to the presence of a high proportion of lipid-like species, supporting the suggestion that this sample is composed predominantly of adipose tissue. In contrast, the breast tissue giving rise to the spectrum in Figure 5(b) is composed predominantly of protein, as indicated by the intense amide I absorption. More specifically, the series of characteristic absorptions at 1030, 1080, 1204, 1240, 1280 and 1330 cm^{-1} (highlighted on the figure) is indicative of a high collagen content, suggesting that this sample is composed predominantly of connective tissue. The third section of breast tissue (Figure 5c) exhibits a strong protein content, but no absorptions attributable to either collagen or adipose tissue. It can therefore be reliably deduced that this section of tissue is comprised predominantly of epithelial cells.

The spectrum of healthy and hypercholesteremic rabbit liver tissue differ considerably from all three spectra of breast tissue (Figure 5d and e). In particular, a strong series of absorptions is present between 1000 and

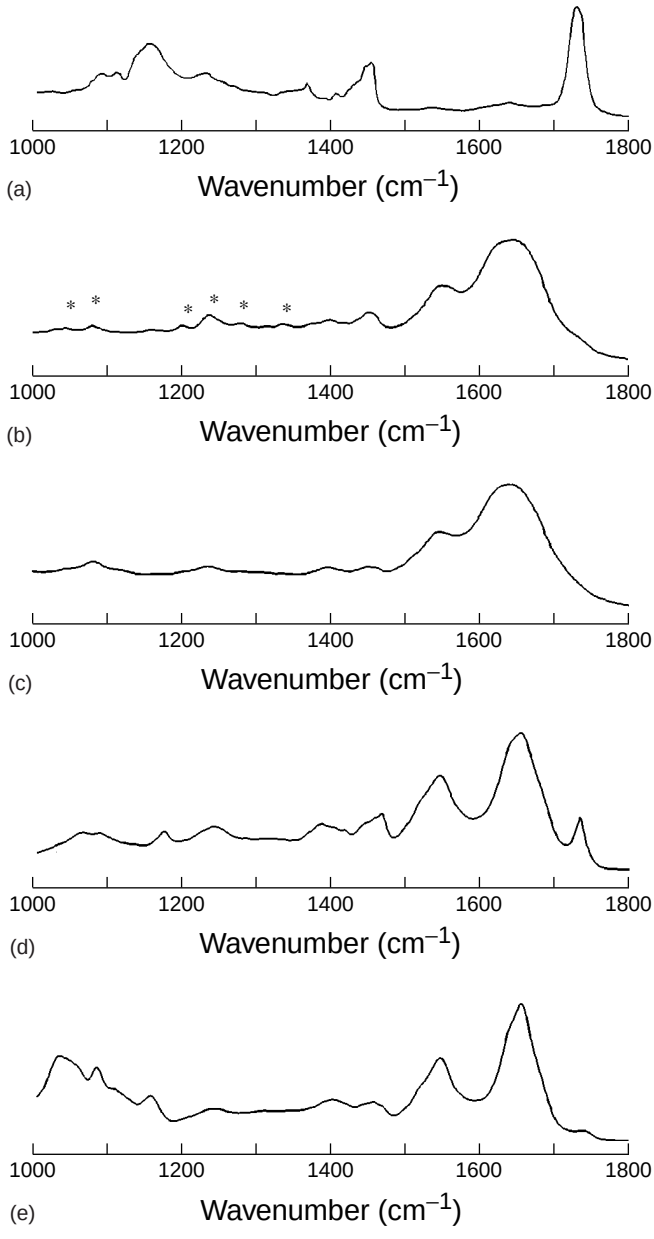

Figure 5 IR spectra of human breast tissue (a–c), and liver tissue from hypercholesteremic (d) and (e) normal rabbit liver tissue.

1200 cm^{-1} in healthy liver tissue. As one of the main functions of the liver is the storage of carbohydrate in the form of glycogen, it is reasonable to speculate that this series of absorptions arises from C–O stretching vibrations of glycogen present in the liver. Comparison of spectra of liver tissue with spectra of glycogen confirms this assignment. A high cholesterol diet results in significant changes in the spectrum of rabbit liver (Figure 5d). Close examination of the spectrum reveals the presence of a number of absorption bands typically found in lipid-rich tissues (e.g. a strong ester C=O

stretching absorption band), suggesting accumulation of lipid-like species in the liver. Given the fact that this rabbit was ingesting a high cholesterol diet, it seems reasonable to speculate that these lipid-like absorptions arise from deposition of cholesterol esters within the liver. Comparison with spectra of low-density lipoprotein (rich in cholesterol esters) shows a high degree of similarity, supporting this hypothesis.[22]

Spectra of both normal and hypercholesteremic liver differ significantly from spectra of protein-rich breast samples in the region of the amide I and II absorption bands. This difference arises as a consequence of the two measurement techniques used. Breast spectra were acquired by placing thin sections of tissue between CaF_2 windows. As such, the sections were still fully hydrated. On the other hand, spectra of liver tissue were recorded from an air dried section of tissue using an IR microscope. The difference in the amide I and II profiles of protein-rich breast and liver samples therefore reflects the presence of the strong, broad O–H bending absorption at around 1640 cm^{-1} in the breast tissue, which distorts the amide I profile.

5 NONSUBJECTIVE ANALYSIS OF SPECTRA

Clearly, interpretation of spectra involves a detailed understanding of both spectroscopy and histology and entails a significant degree of subjectivity. It is possible to remove this subjectivity by the application of statistical techniques. Statistical treatments of spectroscopic data to date have usually involved multivariate pattern recognition. Multivariate pattern recognition methods fall into two basic categories: unsupervised and supervised.

Unsupervised pattern recognition methods are used to identify spectra that have similar (within some predetermined limit) characteristics. As such, unsupervised methods are driven by gross spectral similarities and no knowledge of the nature of the sample is required. Typical examples include cluster analysis techniques such as hierarchical cluster analysis, fuzzy C-means (FCM) cluster analysis, k-means cluster analysis and ISODAT (interactive self-organized data analysis techniques) clustering. The earliest applications of such techniques to IR spectra of cells and tissues were those of Naumann et al., who applied hierarchical cluster analysis to spectra of bacteria.[23] These highly successful studies demonstrated that bacteria could be grouped by strain using this approach, and methods for the rapid identification of bacteria based upon IR spectra are now being developed.

In contrast, supervised pattern recognition techniques such as linear discriminant analysis (LDA) make use of the fact that the investigator often has a substantial amount of information available (either biochemical or clinical) regarding spectroscopic data.[24] For example, the investigator may know that spectra arise from well-defined tissue types. This information may then be used to train an LDA algorithm to recognize the particular combinations of variables (peak frequencies, bandwidths, relative intensities etc.) in these spectra that are characteristic of these tissue types. The trained LDA algorithm can then be applied to unknown spectra, and the unknown spectra partitioned into one of these classes based upon the spectral pattern found.

Supervised pattern recognition techniques, including artificial neural networks and LDA have been applied with great success to the classification of IR spectra of cells, tissues and fluid.[21,25,26] However, a few points should be borne in mind when applying supervised pattern recognition techniques.

Preprocessing of data is almost always a prerequisite to obtaining good classification. The simplest forms of preprocessing are normalization and derivatization. Normalization is required to ensure that classification is not based upon variations in sample thickness, while derivation may be required to remove baseline offsets which can affect classification. (It should be noted that these preprocessing steps should also be considered when employing clustering techniques).

More sophisticated preprocessing is aimed at data reduction, with the aim of removing superfluous data while simultaneously reducing computation time.[21,27] One method of achieving this is to identify the spectral regions which contain the most diagnostic information, a process termed *region selection*. In practice, region selection is achieved by selecting a window of only a few data points (typically $n = 10–12$) within the spectrum. Beginning at one end of the spectrum, only this window of points should be used in the training and test sets. Repeat this procedure with 25 randomly generated training sets and record the average accuracy with which the test sets are classified. Next, move the window by $n/2$ points and repeat the process. This process is repeated until the entire spectrum has been traversed. The spectral window that produced the highest classification accuracy is then determined. If this spectral window allows classification with a predetermined accuracy (say 85–90%) then only this region of the spectrum is required for further computations. If insufficient accuracy is obtained, then all combinations of the best six to eight subregions should be used in the training set to determine the combination of these subregions that results in optimal classification accuracy.

While preprocessing is important, it is crucial to any analysis that the training set contains a sufficient number of spectra representative of the population as a

whole to allow robust training. Unfortunately, often only small data sets are available. One method of improving robustness classification with small sample sizes is to apply the *leave-one-out* method. This is best illustrated by an example. If we have 100 spectra in our entire database, we need to separate spectra into the training and test sets. Typically, the training set contains two-thirds of the spectra, so we have a training set containing 66 spectra. The LDA is then trained and the remaining spectra are each then classified. However, with such a small number of spectra used in the training set, the training set may not be representative of the population at large, in which case poor classification will result. An alternative approach, termed leave-one-out, makes use of the fact that each spectrum in the test set is classified independently of the other spectra in the test set. The basis of this approach is to remove one spectrum (leave-one-out) and train the LDA algorithm using the remaining 99 spectra. This trained algorithm is then applied to the single spectrum and a classification for this spectrum determined. The spectrum is then returned to the training set and a second spectrum removed. The LDA is then trained on this new training set of 99 spectra and applied to the new single test spectrum. This process is repeated, essentially using 100 discrete training sets of 99 spectra to classify the 100 spectra. This approach has been used to classify macroscopic spectra of dry films of synovial fluid as arising from either control joints or joints affected by one of three forms of arthritis with an accuracy of greater than 95%.[9]

Another effective method for developing a robust training set is known as bootstrapping.[21] Using this approach the data are split randomly into training and

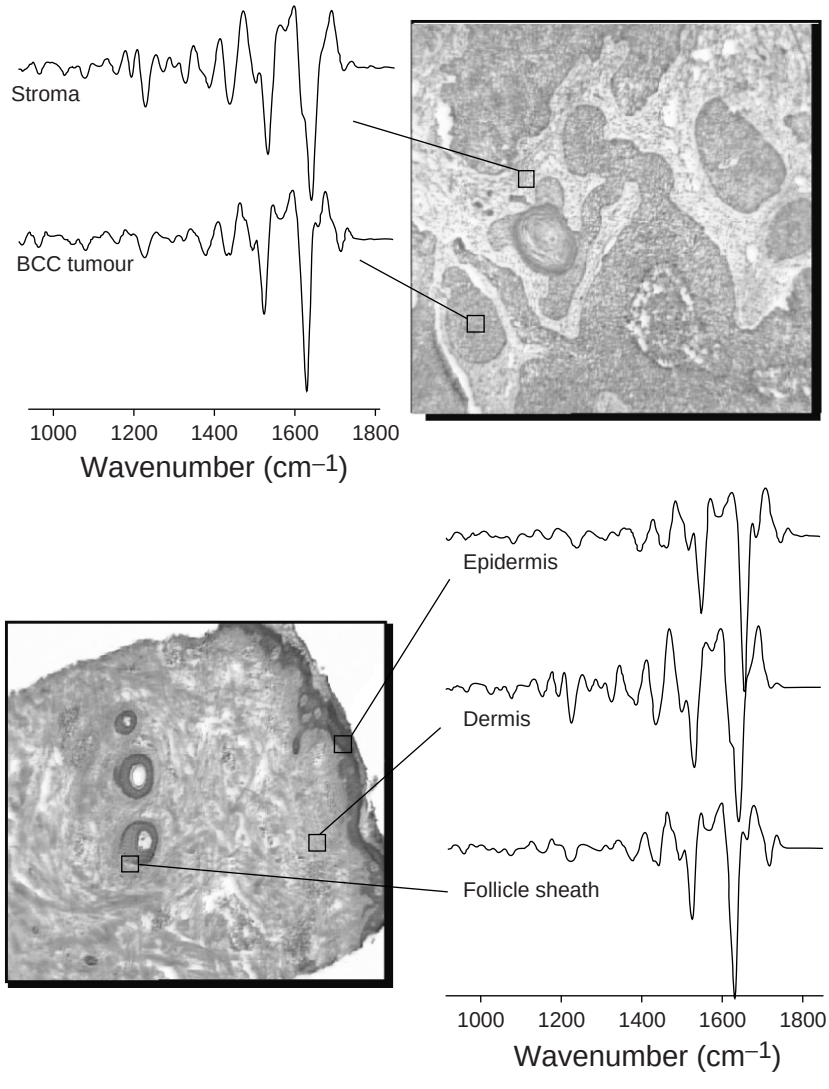

Figure 6 Photomicrographs of skin biopsies showing regions of tissue used to produce spectra of ''pure'' tissue components. Derivative spectra from each region are shown.

test sets with two-thirds of the spectra in the training set. The classifier is then trained, and a new training set of 66 spectra is randomly selected. This process is repeated, typically 250 times, and the average classifier from these 250 iterations computed. This average classifier is then applied to all 100 spectra.

Both the bootstrapping and the leave-one-out methods enhance the robustness of LDA by reducing the risk of ''overtraining''. With a test set that is unrepresentative of the population at large one runs the risk of reaching a suboptimal solution, that is a local minimum rather than a global minimum. The use of large training sets which adequately represent the population at large or the average classifier resulting from multiple training sets reduces the likelihood of this occurring.

The robustness of spectral analysis by supervised classification techniques is also affected by the requirement for assignment of spectra to a particular class of tissue. Uncertainty about the histology of the sample can result in uncertainty or ''fuzziness'' of classification. For example, in a recent study of breast tissue a novel classification strategy was required to classify spectra of breast tumor by grade, as a traditional approach to LDA proved to be unsatisfactory.[21] To a large extent, this difficulty arose because spectra were obtained by macroscopic transmission methods. As discussed above, such spectra represent an average of all tissues present in the sample, which in this instance may be normal epithelial cells, epithelial cells exhibiting varying different degrees of malignancy, adipose tissue and connective tissue. Obviously such a complex composition makes assignment of macroscopic spectra to a particular class difficult.

This problem may be alleviated by the acquisition of spectra with the use of an IR microscope. The ability to generate highly accurate classifiers using spectra of essentially pure tissue components has been demonstrated in studies of skin biopsies. The ability to acquire spectra from 30×30-μm regions of tissue ensured that spectra of tumor, epidermis, dermis, stroma and hair follicles could be obtained without contamination by the other tissues (Figure 6). Notable differences may be seen between the derivative spectra of the five tissue components.[26] For example, spectra from the dermis and the stroma running throughout the tumor body exhibit the characteristic signature of collagen. In contrast, spectra obtained from the follicle sheath and sebaceous glands exhibit strong lipid absorptions arising from sebum, a complex mixture of lipids, acylglycerides, waxes, squalene, cholesterol and cholesterol esters of unknown function. Finally, the epidermis and tumor could be differentiated by a reduced protein : nucleic acid ratio in regions of tumor.

It can readily be seen that IR microscopic examination of skin allows individual tissue components to be identified. Acquisition of spectra from a number of skin biopsies in this manner allowed a large representative database of spectra of the pure tissue components to be produced. Application of LDA to this data set allowed classification of spectra of each of the tissue components with extremely high accuracy (98%).

Clearly the application of pattern recognition techniques to IR spectra of cells and tissues allows nonsubjective classification of spectra with an impressive degree of accuracy. However, it should be stressed that no classification strategy can account for poor data, and it is vital that care be taken to ensure that all spectra are characterized by linear detector responses, high signal-to-noise ratios and good water vapor compensation.

6 ANALYSIS OF MICROSPECTROSCOPIC MAPS

In standard transmission methods, a single spectrum is obtained which describes the gross spectroscopic characteristics of the sample. An estimate of gross sample histology can be obtained relatively easily by analyzing the spectrum using the approach describe above. With the acquisition of spectra from single points by IR microscopy, spectra of pure tissue components may be obtained and data correlated with sample histology, allowing characterization of tissues with much higher precision in a relatively straightforward manner. However, with IR microscopic mapping this single spectrum is replaced by many thousands of spectra, and data analysis and presentation present unique challenges. While the interpretation of individual spectra from a mapping experiment can be achieved as described above, this approach is obviously impractical when analyzing a map containing 10 000 spectra. A number of more efficient methods for analysis of spectroscopic maps have recently been developed, many of which are borrowed from, or have parallels in, image processing.

6.1 Functional Group Mapping

The most straightforward and most common method of analyzing and presenting data from IR microscopic mapping experiments is so-called functional group mapping. In functional group mapping, parameters obtained from a functional group of interest are plotted as a function of position within the map. The parameters that may be plotted depend upon the information sought. Examples of parameters that may be plotted include the frequency of an absorption maximum, the peak intensity of an absorption band, the integrated intensity of an absorption band and the ratio of peak or integrated intensities.

6.1.1 Intensity Maps

As the intensity of a functional group is directly proportional to the concentration of the material giving rise to the absorption, plotting the intensity (either peak intensity of integrated intensity) of a functional group at each pixel within a map provides a simple method for assessing the distribution of the material throughout the sample. A number of approaches are available for calculation of peak intensities, and care must be taken to choose the correct method and avoid the introduction of artifacts.

Calculation of intensity values is most straightforward if performed using the original absorption spectra, although baseline fluctuations may be problematic. For example, spectra of tissues often show a considerable baseline slope and offset due to scattering effects. In a map containing thousands of spectra, this offset and slope are likely to vary considerably. Simply calculating the peak height above a zero baseline will therefore lead to artifactual changes in intensity that are related to changes in this slope and offset rather than absorption intensity. It is therefore important that investigators include some type of baseline compensation in intensity calculations to avoid introduction of erroneous variations in intensity. The position is similar when considering integrated intensities, the correct calculation being the area enclosed by the spectrum and this baseline.

Unfortunately, in many instances it is not feasible to analyze absorption spectra because of overlap of the absorption band of interest with other absorption bands. Calculation of second derivative spectra may result in sufficient separation of absorption bands to allow calculations on such highly overlapped absorption bands and has the additional advantage that baseline variations are removed. However, calculations on derivative spectra require considerable care, for a number of reasons. First, second derivative spectra exhibit negative peaks. The presence of negative values results in presentational problems, as the functional group maps will be inverted. Second, most derivative peaks have both positive and negative contributions, which may have a significant impact upon the calculation of both peak and integrated intensities. For example, reading of the value at the minimum of an absorption band in derivative spectra significantly underestimates the height of the derivative peak, as the contribution of the positive components of the peak are ignored. Finally, broad, weak absorption bands may produce derivative peaks that have intensity values that vary around zero, which complicates interpretation.

Inversion of spectra and the application of an offset correction to ensure that all calculated values are positive remove these problems. Following derivation, inversion and offset correction, calculation of reliable peak or integrated intensities is possible. However, again a baseline should be included to define the absorption band of interest, and either the height above this baseline or the area bounded by the baseline and the absorption band calculated.

With adequate precautions, the distribution of materials within tissues can now be monitored. Absorptions of particular interest are those arising from lipid (e.g. the symmetric and asymmetric CH_2 stretching and ester C=O stretching absorption bands), DNA (C=O stretching absorption band, phosphate stretching absorption bands), protein (amide I absorption) and collagen (amide II absorption at 1204 cm^{-1}). Plotting the intensity of these absorption bands as a function of position allows the distribution of lipids, protein and DNA to be assessed, which provides an overview of the gross biochemical characteristics of tissues.

This utility of intensity-based functional group mapping for characterizing tissue is illustrated in Figure 7, which shows functional group maps obtained from a section of skin containing a basal cell carcinoma. To aid interpretation of the functional group maps, skin morphology will be briefly discussed. The outermost layer of the skin is the epidermis (see Figure 7a). The epidermis consists of densely packed keratinized epithelial cells that grow in layers, the deepest cell layer resting on the basement membrane that separates the epidermis and dermis. The dermis or connective tissue layer is significantly thicker than the epidermis and is largely acellular, consisting mainly of a connective tissue matrix comprised of collagen fibrils. It supports sebaceous glands and hair follicles, which are surrounded by a layer of epithelial cells called the follicle sheath. These anatomical structures are highlighted in Figure 7(a). Two regions in which the tissue has become folded during deposition on to the window are also highlighted, as these regions of tissue are potential sources of artifacts.

The amide I intensity, which may be used to monitor protein distribution, is plotted in Figure 7(b) as a function of position within the map. The highest concentration of protein is seen in the regions of folded tissue. This is expected, as the pathlength in this region of the tissue is obviously greater than in the surrounding tissue. Clearly this is artifactual and does not provide useful information relating to the composition of the tissue in these regions. Such regions of the tissue should be ignored when interpreting data, and whenever possible tissue sections containing folds should not be analyzed. In the remaining tissue, the highest protein concentration is seen in the outer layers of the skin, corresponding to the epidermis. The protein concentration is generally reduced in the dermis and is lowest in the region of the tissue containing the tumor. However, small islands of high protein concentration can be seen, and these islands of protein clearly correspond to hair follicles. This

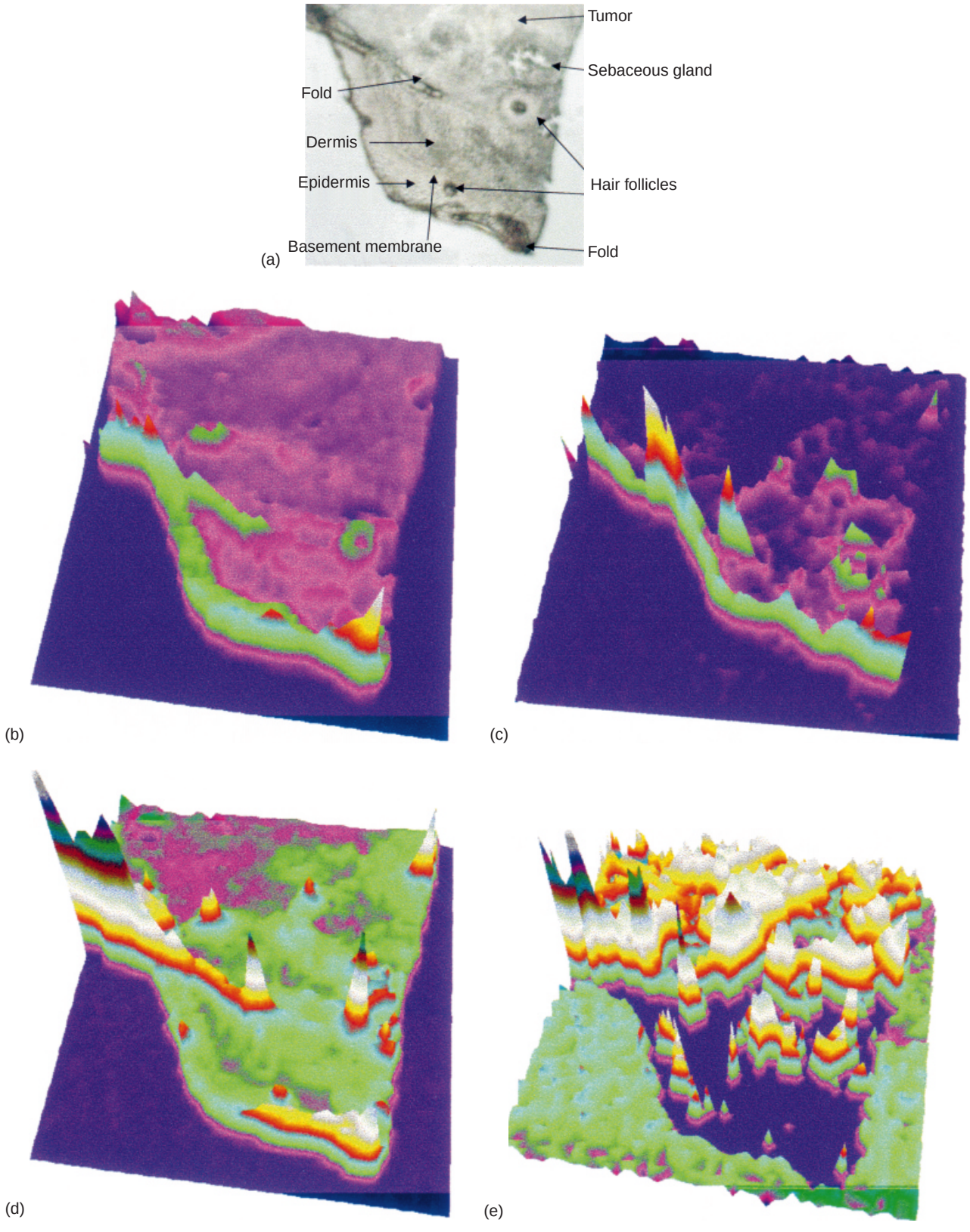

Figure 7 (a) Photomicrograph of a skin biopsy. Functional group maps showing distribution of (b) protein (amide I peak height), (c) collagen (peak height at 1204 cm^{-1}), (d) lipids (peak height at 1740 cm^{-1}) and (e) nucleic acids (peak height at 1718 cm^{-1}).

distribution results from the differences in composition of the three tissues. The epidermis consists of tightly packed cells, many of which have significant deposits of keratin, which results in a high protein content of this tissue. Hair follicles are surrounded by a layer of cells that have the same origin as the epidermal cells, and thus exhibit the same high protein content. It is interesting to note however the low protein content at the center of the hair follicle, which arises because hairs are hollow.

The dermis is largely acellular and consists predominantly of collagen fibers interspersed throughout a glycosaminoglycan ground substance. The protein content is thus reduced compared to the epidermis. The region of the tumor exhibits the lowest protein content. This is most likely to be a consequence of nuclear enlargement which is seen in many tumor cells and which reduces the amount of cytoplasm, and so protein, inside the tumor cells.

The distribution of intensity at $1204\,cm^{-1}$, which corresponds to collagen distribution, is notably different from the distribution of the amide I intensity (Figure 7c). No intensity is seen at $1204\,cm^{-1}$ in the region of the epidermis, but strong contributions are seen in the region of the basement membrane, hair follicles and sebaceous glands. Such a distribution of intensity is expected given that the basement membrane is composed predominantly of collagen, while hair follicles and sebaceous glands are both surrounded by a protective collagen sheath. A more diffuse distribution of collagen, corresponding to the collagen fibers interspersed throughout the ground substance, is apparent in the dermis.

As seen when monitoring protein distribution, artificially high lipid concentrations are seen in regions of folded tissue (Figure 7d). Again, these regions of tissue should be ignored. Strong contributions from lipid can clearly be seen in the region of hair follicles and sebaceous glands, but little lipid-like material is seen in other regions of the sample. The appearance of lipid-like material in sebaceous glands and hair follicles is the result of the presence of sebum. Interestingly, it can be seen that the low concentration of protein seen at the center of hair follicles corresponds to peaks in sebum content, suggesting that at the level at which the hair follicles have been sectioned the center contains a plug of sebum.

Finally, the distribution of DNA, as indicated by the distribution of intensity of the DNA absorption at $1718\,cm^{-1}$, is markedly different from the distribution of other materials, with the highest intensity appearing in the region of tumor (Figure 7e).

From this brief discussion, it should be apparent that discrete tissue types within a sample can be identified based upon the distribution of the major biochemical species present.

6.1.2 Frequency Maps

While plotting intensity information provides information concerning the amount of materials at each position within the map, plotting the frequency of absorption maxima as a function of position provides information relating to structural properties. For example, plotting the frequency of the amide I absorption maximum (or the center of gravity of the amide I absorption band) provides information relating to variations in protein structure throughout the tissue. Such information may be useful in the characterization of proteinaceous deposits associated with pathological conditions. A recent study illustrated the effectiveness of this approach in tissues affected by AD.[28]

AD is characterized by progressive degeneration of the gray matter of the central nervous system. It is believed that deposition and self-association (aggregation) of a peptide known as βA4 amyloid lead to the formation of so-called neuritic plaques. Neuritic plaque formation is followed by neuronal death producing severe cortical atrophy and loss of language skills, memory and intellectual function and disorientation. The accumulation of βA4 is one of the major pathological hallmarks of AD, and the structure of the peptide in situ is of considerable interest.

To locate deposits of βA4 in situ, intensity-based functional group mapping may be used. Deposition of the peptide should lead to a significant increase in amide I intensity in the region of the neuritic plaque. This is illustrated in the amide I functional group map produced from maps of the tissue shown in the photomicrograph in Figure 8(a), which contains an amorphous deposit which may be a neuritic plaque. Analysis of the amide I functional group map shows that there is indeed a dramatic increase in the integrated intensity of the amide I band in the center of the area mapped, corresponding exactly to the amorphous deposit of interest (Figure 8b). This observation implies that the amorphous deposit is proteinaceous in composition. However, it does not allow any conclusions concerning the structure of the protein in the deposit to be drawn. To obtain this information, a plot of the center of gravity of the amide I absorption as a function of position is required.

Frequency-based functional group maps are usually presented as two-dimensional contour plots. The contour plots are color coded according to the center of gravity of the amide I. In the representation shown in Figure 8(c), orange/red coloration corresponds to an amide I center of gravity at $1650–56\,cm^{-1}$ while blue coloration corresponds to an amide I center of gravity at $1630\,cm^{-1}$. As the frequency of the amide I is indicative of protein structure, regions of the tissue containing proteins with different secondary structures are colored differently. Thus, regions of the tissue containing aggregated proteins are shown

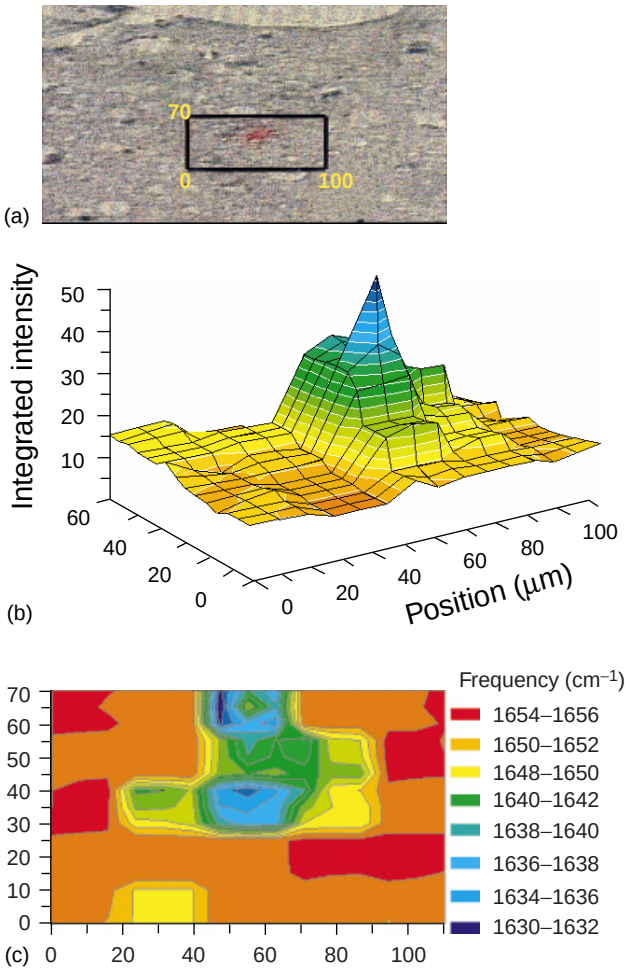

Figure 8 Photomicrograph of Alzhiemer's tissue showing the region mapped (a) and intensity-based (b) and frequency-based (c) functional group maps derived from the amide I absorption.

in blue, while those containing predominantly unordered and/or helical proteins are shown in orange/red.

The area corresponding to the proteinaceous deposit can clearly be seen to exhibit a low amide I frequency (1630 cm^{-1}) indicative of aggregated protein. This strongly suggests that the protein deposit identified by intensity-based functional group mapping is a neuritic plaque. Staining of the tissue with Congo red, a stain that specifically colors amyloid deposits pink/red, revealed that the mass was indeed a neuritic plaque, confirming the spectroscopic assessment (Figure 8a).

Frequency-based functional group mapping may also be used to monitor variations in more subtle structural parameters such as membrane fluidity. The frequency of the absorptions attributed to CH_2 symmetric and asymmetric stretching vibrations of lipids are sensitive to the degree of conformational order within membranes. Increased membrane disorder (i.e. a more fluid environment) is associated with an increase in the frequency of the CH_2 stretching absorption bands. Thus, a plot of the frequency of the CH_2 stretching absorption band frequency may be used to probe membrane dynamics in tissue sections and to assess variations in membrane dynamics. Although such studies have been few to date, potential applications include investigation of neurological disorders, particularly those involving lipid-rich white matter such as multiple sclerosis.

While frequency-based functional group mapping is potentially a powerful method for assessing structural properties of tissues, care should be taken in interpreting results. Particular care should be taken when the chromophore of interest arises from more than one biochemical species. For example, the CH_2 stretching absorption bands in tissues may arise predominantly from lipids but often contain significant contributions from proteins. As the frequency of CH_2 stretching absorption bands differs in lipids and proteins, the frequency of this composite absorption is determined by the frequencies of the CH_2 stretching absorption bands of the lipid and protein components and the relative proportions of lipid and protein. A change in the frequency of this composite absorption band may therefore result from a change in the frequency of the lipid CH_2 stretching absorption bands, a change in the frequency of the protein CH_2 stretching absorption band or a change in the relative amounts of lipid and protein.

To distinguish between these three possibilities requires other functional group maps to be examined. For example, examination of the amide I frequency-based functional group map will allow the effect of variations in protein structure to be assessed, while examination of the amide I and lipid C=O intensity-based functional group maps allow the role of variations in lipid and protein content to be determined. If these parameters remain constant across the area of the map of interest then a change in membrane fluidity may then be inferred.

6.1.3 Ratio Maps

While intensity- and frequency-based functional group maps provide much information, this information typically relates only to a single chromophore. From the brief discussion above it can be seen that there are often instances when the ratio of materials (for example lipid and protein) within tissue is of interest. In such cases, functional group maps that represent the ratio of a parameter measured for two distinct chromophores may be of use. For example, a ratio map showing the variation in the lipid : protein ratio across a tissue section can be calculated in a number of ways. The most obvious methods involve measuring the intensity of the amide A, amide I, or amide II absorption band from proteins and

the intensity of the CH_2 asymmetric or symmetric stretching or CH_2 bending absorption band of lipids. The ratio of the intensity of any of the lipid absorption bands to any of the protein absorption bands will provide an indication of the variation of the lipid : protein ratio across the tissue.

Two important, if obvious, factors should be stressed. First, the lipid : protein ratio obtained will depend upon the chromophores used. Second, the lipid : protein ratio obtained will be different if the investigator uses absorption or derivative spectra, even if the same chromophores are used to calculate the ratio, as relative intensities are not retained in derivative spectra. Obviously then this methodology shows only trends in the lipid : protein ratio across the tissue, and cannot be used to calculate absolute values. However, as interest is usually in changes in ratios within the tissue, this does not usually represent a problem.

A second method of estimating changes in the lipid : protein ratio across the surface of a tissue section involves calculating the ratio of the CH_2 and CH_3 stretching absorptions. Both lipids and proteins give rise to CH_2 and CH_3 stretching absorptions. However, the ratio of the intensity of the CH_2 and CH_3 absorption bands is significantly different in lipids and proteins. Lipids typically contain 14–18 CH_2 groups and only one CH_3 group per acyl chain, while proteins contain almost equal proportions of CH_2 and CH_3 groups in their side chains. Thus, the $CH_2 : CH_3$ intensity ratio for lipid-rich and protein-rich tissues differs substantially. Plotting the $CH_2 : CH_3$ intensity ratio as a function of position therefore provides an alternative method of monitoring variations in the lipid : protein ratio throughout tissues.

6.1.4 Gray Scale Imaging

While visualization of the distribution of tissue components by functional group mapping is a powerful method of analysis, it requires a degree of familiarity to generate maps with the correct orientation or numbers of contours to portray information usefully. This is particularly so for two-dimensional contour plots. This can be avoided by converting the maps to gray scale images, in which each pixel is assigned a value between 0 (white, no absorption) and 255 (black, total absorption) depending upon the intensity of the absorption band under investigation.[29] Gray scale functional group maps calculated from absorption bands at 2924, 1650, 1204 cm^{-1} and the ratio at 1240/1204 cm^{-1} for the tissue section shown in Figure 2 are shown in Figure 9. The CH_2 stretching absorption band gray scale map (2924 cm^{-1}, Figure 9a) indicates a high lipid content (areas shaded black) around hair follicles and in sebaceous glands. Additionally, there is a high lipid CH_2 content in the epidermis, probably owing to membrane phospholipids in the keratinocytes of this skin layer. The dermis and stroma have low lipid content indicating the relatively low amount of cellular material in these skin areas.

Protein deposition within the section closely follows lipid deposition (Figure 9b), with the highest protein content seen in the epidermis and hair follicles and the lowest protein content seen in the tumor. As discussed above this low protein content arises from nuclear enlargement, decreasing the volume of cytoplasm within tumor cells.

Collagen (1204 cm^{-1}) distribution throughout the section (Figure 9c) is clearly different from total protein distribution. Collagen content is very high in the dermis of the skin, especially in dermis close to the tumor area. As expected, the tumor region contains little if any collagen.

Monitoring nucleic acid distribution in tissue sections can be difficult. In Figure 7 the intensity of the DNA absorption at 1717 cm^{-1} was used as a marker for DNA content. However, in many tissues this weak absorption is difficult to detect. In such tissues the only option is to analyse the PO_2^- stretching absorption bands at 1240 and 1080 cm^{-1}. However, each of these absorption bands contains contributions from phosphate groups of phospholipids and the amide III absorption band of collagen. Analyzing absorptions at 1240 and 1080 cm^{-1} is therefore inadvisable as a method of monitoring nucleic acid distribution.

We have found that the most reliable method for tracking nucleic acid content in the absence of a clear absorption band at 1717 cm^{-1} is to monitor the ratio of absorbance at 1240/1204 cm^{-1}. Collagen exhibits absorption bands at both frequencies, while nucleic acids exhibit an absorption band only at 1240 cm^{-1}. Thus, this ratio will provide a useful indication of the relative proportions of nucleic acids and collagen. A high ratio indicates a high nucleic acid content, a lower ratio indicates significant collagen contributions. For practical purposes, the contributions from lipid phosphate groups can be ignored in tissues without a significant adipose tissue component. This can be verified by examination of the ester C=O stretching band at 1740 cm^{-1}. The lipid phosphate and ester C=O absorptions exhibit approximately equal intensity in isolated materials. As the lipid C=O absorption is typically weak in most tissues, then it can be assumed that the lipid phosphate groups contribute minimally to the absorption at 1240 cm^{-1} in most tissues.

The distribution of nucleic acids (1240/1204 cm^{-1}) in the tissue section is shown in Figure 9(d). The epidermis, follicle sheath and tumor area all have a high content of nucleic acids. The dermis and stroma exhibit the lowest

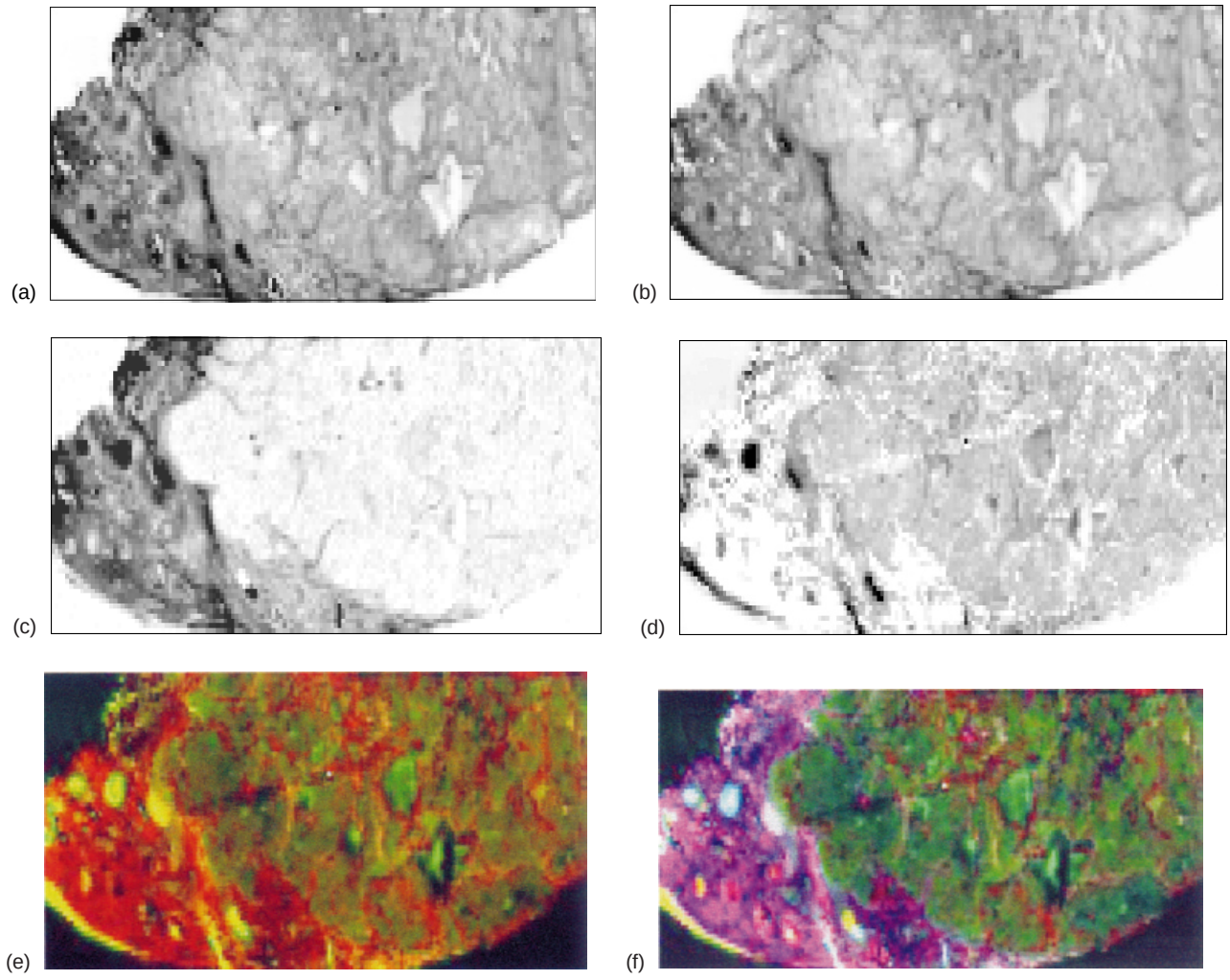

Figure 9 Gray scale images showing distribution of lipid (peak height at 2924 cm^{-1}, a), protein (amide I peak height, b) collagen (peak height at 1204 cm^{-1}, c) and nucleic acids (ratio of absorbance at 1240/1204 cm^{-1}, d) in the section of skin shown in Figure 2. Digitally stained maps produced by combing color-scaled maps are shown for (e) protein (red) and nucleic acids (green) and (f) phospholipid (red), nucleic acids (green) and collagen (blue). See text for more detail.

nucleic acid content. Interestingly, this presentation allows the most fine structure to be visualized within the tumor, clearly allowing detection of the stroma running throughout the body of the tumor.

It is apparent that functional group mapping, either in the form of two-dimensional contour plots or three-dimensional intensity plots or gray scale images, readily identifies the major tissue types in skin due to their widely different chemical compositions (i.e. collagen in dermis, lipid in adipose tissue, etc.). However, such representations of data allow only the distribution of one or two (in the case of ratio maps) components to be assessed in a given map. It would be of more practical use if many components could be simultaneously displayed in one map. This is achieved with a technique known as digital staining.

6.1.5 Digital Staining

As its name implies, digital staining[29] mimics the process of staining tissues, allowing discrimination of tissue components based upon color. In digital staining, 8-bit gray scale images showing the distribution of three tissue components are converted to 8-bit red, green and blue scale images. These 8-bit images are then combined to produce 16- or 24-bit images. By superimposing the chemical information contained in the separate functional group maps, digital staining provides a means of representing relative proportions of chemical species visually.

Figure 9(e) shows the digitally stained map obtained by combining the functional group maps for protein and nucleic acids shown in Figure 9(b) and (d). The nucleic acid gray scale map was converted to a green scale map while the protein gray scale map was converted to red

scale. Thus, regions of tissue with a high proportion of nucleic acids appear green while areas of tissue rich in protein appear red. Note that colors appear as the eye would see them if light of these colors were shone into the eye, not as if reflected from the page. Thus areas of equal proportions of nucleic acid and protein appear yellow and not purple. This 16-bit presentation of the data clearly shows discrimination between various types of tissue. The body of the tumor can clearly be seen colored green, indicating a relatively high nucleic acid content and a relatively low protein content. Areas of stroma can also be seen running throughout the body of the tumor and colored red, indicative of a relatively high protein content and a relatively low nucleic acid content. The epidermis and follicle sheath appear yellow in this representation, indicating equal spectral contributions from nucleic acids and proteins.

Figure 9(f) shows a 24-bit representation of the distribution of phospholipids (red), nucleic acids (green), and collagen (blue). From this three-color digitally stained image it is clear that the dermis contains a high proportion of collagen with some lipid (blue/purple), that the tumor is comprised mainly of nucleic acids (green) and that the epidermis contains high proportions of both nucleic acids and phospholipids (yellow). In addition, it is apparent that the dermis exhibits considerable structure. Histologically, the dermis is comprised of two distinct layers, the reticular dermis and the papillary dermis. The reticular layer is the furthest from the epidermis (closest to the tumor) and forms the bulk of the dermis. It has prominent broad bands of dense collagen, with intervening long fibers of elastin. In contrast, the upper papillary layer contains less collagen and elastin, but more glycosaminoglycan matrix and more cellular elements. The two layers of the dermis can readily be seen as dark purple (papillary layer, high collagen content, low lipid content) and light purple (reticular layer, lower collagen content and higher lipid content from cell membranes). Thus the 24-bit digitally stained image separates these two dermal layers well, demonstrating the discriminatory power of this mode of presentation.

6.2 Nonsubjective Analysis of Microscopic Maps

Functional group mapping, either as three-dimensional intensity plots, two-dimensional contour plots, gray scale images or digitally stained images, is clearly a very powerful analytical tool. However, in order to understand the significance of such plots this method requires absorption bands to be assigned to particular functional groups, which may not only be difficult but also highly subjective, even for the trained spectroscopist. This inherent subjectivity reduces the usefulness of the technique in a clinical setting. To remove or reduce this subjectivity a number of unsupervised multivariate pattern recognition/classification techniques, particularly cluster analysis techniques, have been applied to spectroscopic maps.

6.2.1 Unsupervised Methods: Cluster Analysis

Unsupervised pattern recognition techniques classify data based upon gross spectral features. Classification is therefore based upon inherent properties of spectra, and no knowledge of the nature of the sample is required. For example, cluster analysis techniques may be used to group spectra based upon some measure of similarity. This will be illustrated by a description of a technique known as ISODAT clustering, one of the most straightforward methods of cluster analysis.[30] ISODAT clustering requires an estimate of the number of clusters expected together with the minimum size of each of the clusters (i.e. the minimum number of spectra in each cluster). In addition, an estimate of the maximum standard deviation that is acceptable within a cluster before the cluster is split into two smaller clusters and the minimum difference between the means of different cluster which is acceptable before the clusters are merged are required.

ISODAT clustering partitions the data set randomly into the estimated number of clusters, n. The mean and standard deviation are then calculated for each cluster. The means of all of the clusters are then compared. If the difference between the means of any two clusters is less than the threshold value set, then it may be deemed that spectra within these two clusters exhibit a high degree of similarity and the clusters are merged. If on the other hand the standard deviation of any cluster is above the standard deviation threshold set then it may be inferred that spectra in this cluster show a high degree of dissimilarity. This cluster is then split to reduce the standard deviation of the two new clusters below the threshold value. The cluster mean and standard deviation thresholds therefore determine the sensitivity of the analysis. This process is iterated to produce the optimal number of clusters for the input parameters.

ISODAT cluster analysis was applied to a section of hypercholesteremic rabbit liver.[22] The area outlined in Figure 10(a) was mapped, spectra were truncated to the region 3100–3690 cm^{-1} (N–H stretching vibrations, to decrease computation time) and the maximum allowed standard deviation set to 12%. Each shaded block in Figure 10(b) corresponds to spectra acquired from a 40×40-μm pixel of tissue, with steps of 51 and 48 μm in the x and y directions. Pixels of the same shade belong to the same cluster. Clearly with the parameters used in this analysis five distinct clusters can be seen, implying the presence of five discrete tissue types with different

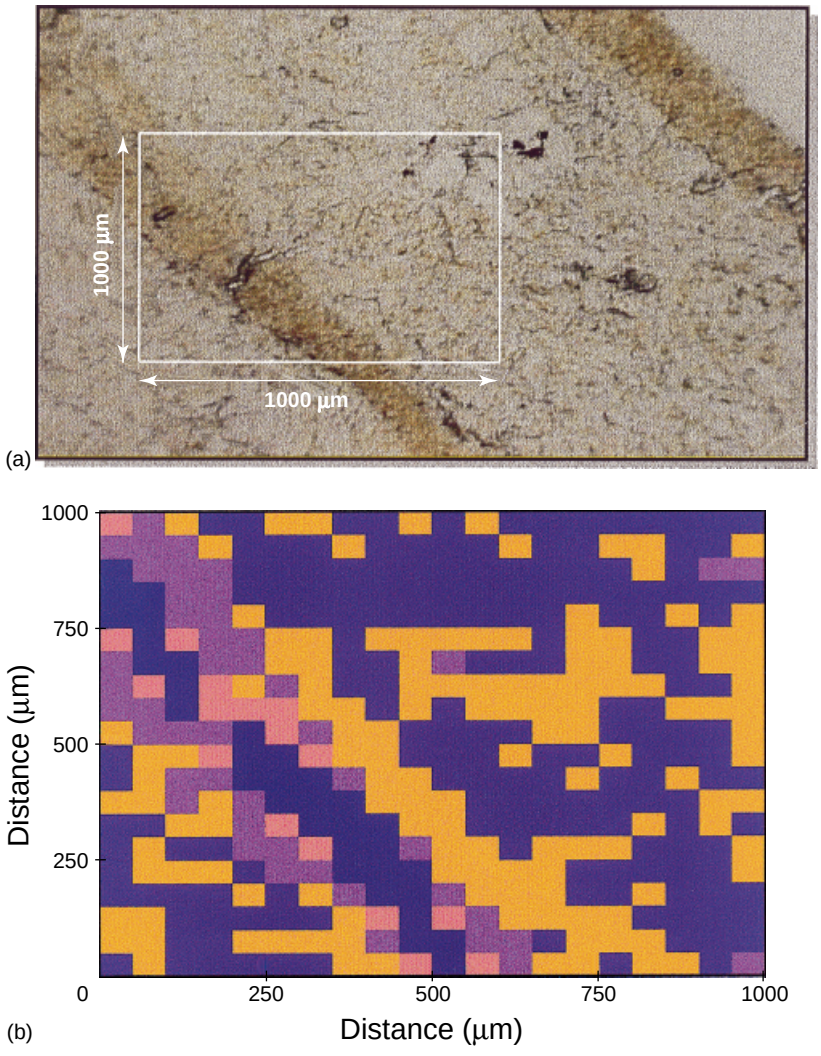

Figure 10 Photomicrograph of hypercholesteremic liver tissue showing the area mapped (a, white box) and the results of ISODAT cluster analysis of spectra (b). Pixels shaded the same color belong to the same cluster.

populations of N–H groups. It can be seen that two of these populations are found almost exclusively in the region of the tissue corresponding to the yellow streak caused by bilirubin deposition in the tissue.

FCM cluster analysis uses similar principles, partitioning spectra such that the differences in the intracluster spectral responses are minimized, while simultaneously maximizing the intercluster differences between spectral responses.[31] Application of FCM cluster analysis to IR microspectroscopic maps of skin tissue is illustrated in Figure 11.[29] Five discrete clusters of spectra were identified. Comparison of the cluster map (Figure 11b) with histological data (Figure 11a) reveals that the clusters correspond to distinct histological types. Pixels colored turquoise correspond to adipose tissue and sebaceous glands in the histological section. Spectra in this cluster exhibited the highest lipid intensities, the highest ester C=O stretching vibration and the lowest amide I and II intensities. Pixels shown in blue are found almost exclusively in the dermis. Spectra in this cluster exhibited the series of absorptions assigned to collagen vibrations. The clustering routine did not differentiate between the two layers of the dermis. Pixels shown in yellow are interspersed throughout the body of the tumor. Spectra in this cluster exhibited weak collagen absorptions, and are most likely to arise from regions of stroma within the tumor. This stromal component is not the same as that seen in the dermis, as it occurs in a separate cluster. This may be due to spectral contributions from malignant epithelial cells in the stroma interspersed throughout the tumor. Alternatively, it may indicate a difference in the chemistry of the stroma in the two locations (i.e. matrix modifications induced by the tumor cells). Green pixels correspond to the majority of tumor

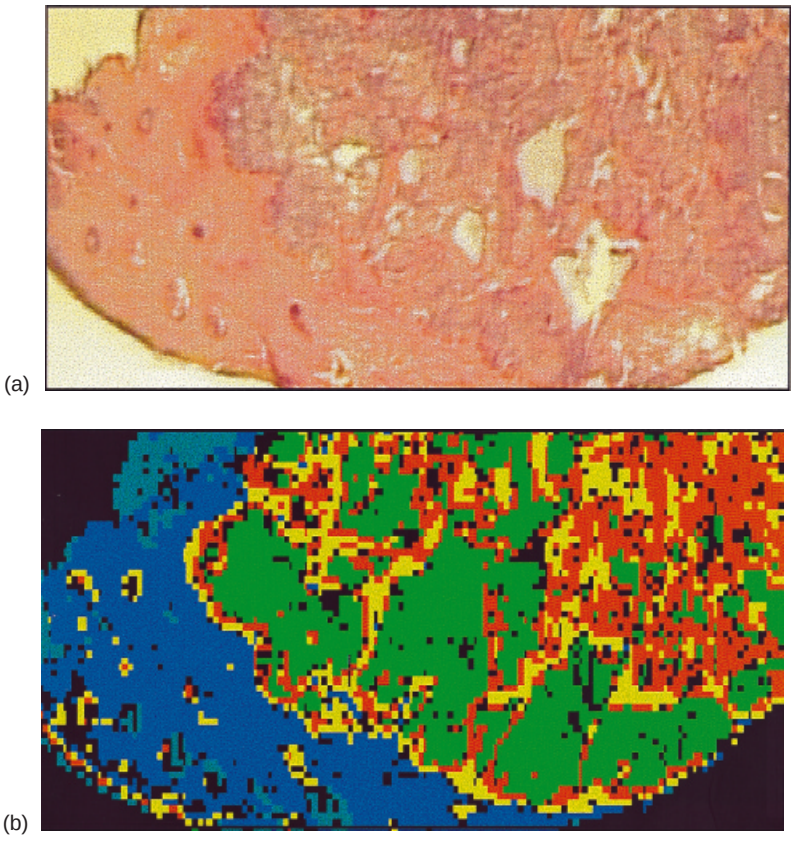

Figure 11 Results of FCM cluster analysis applied to a spectroscopic map of a skin biopsy. Pixels shaded the same color belong to the same cluster.

cells (nodular-type basal cell carcinoma). Spectra in this cluster exhibited weak lipid absorptions and minimal collagen absorptions. An additional cluster (shown in red) is interspersed throughout the tumor. Spectra in this cluster were similar to those in the larger, green cluster, but correspond histologically to infiltrative tumor regions.

Clearly, cluster analysis techniques can be used to distinguish between tissue compartments having similar spectral, and so biochemical, features. However, as discussed above, cluster analysis techniques group spectra based solely upon gross spectral similarities. Clusters of spectra that exhibit only small spectral differences may therefore be grouped together. In addition, the biochemical/histological basis for the appearance ofused these groupings must still be determined by visual assessment and interpretation of spectra from each cluster. Techniques that are better able to separate tissue compartments showing subtle spectral differences and which require no visual interpretation to allow attribution to specific histological groups are therefore required. Supervised pattern recognition methods are potentially better suited to the development of such data analysis methods.

6.2.2 Supervised Methods: Linear Discriminant Analysis

LDA has been very successful when applied to single spectra obtained from multiple tissue samples.[9,21,26] An obvious extension is to use LDA to develop ''search'' methods that can be used to locate individual tissue components within a spectroscopic map of tumor-bearing skin.[32]

As discussed above, we have used spectra acquired from pure components of skin biopsies to train a linear discriminant algorithm to recognize spectra of tumor, epidermis, dermis and follicle sheath.[26] This analysis was extended to allow LDA to be performed on the skin biopsy discussed in section 6.1. Briefly, 914 spectra acquired from a range of skin biopsies were used as the training set for an LDA. To reduce computation time the combination of spectral subregions that allowed the most efficient discrimination between spectral classes were identified. The 10 subregions so identified were then used to train the LDA algorithm to partition spectra into the five classes (epidermis, dermis, follicle sheath/sebaceous glands, stroma or tumor). The trained LDA was then applied to the spectroscopic map. The implementation of LDA used in this study returned a membership value for each spectrum in the spectroscopic map. For each

spectrum, a value ranging between 0 (not belonging) and 1 (belonging) is returned which indicates membership of each class, with the sum of values for all classes being unity. The values returned therefore provide an indication of the likelihood of the spectrum at a particular pixel belonging to each of the five classes. Thus, for a spectrum that arises from tumor, the ideal result would be a returned value of 1 for the tumor class and 0 for all other classes.

To simplify data interpretation, a visual representation of the results of the LDA can be produced by plotting the membership values of spectra at each pixel for each of the five classes, with the data displayed as a gray scale image. In this representation, a likelihood of zero for a spectrum belonging to a particular class results in pixels being shaded black, while a likelihood of one results in pixels being shaded white. Figure 12 shows the gray scale images of the membership values for tumor, dermis and stroma. All membership likelihood maps in Figure 12 correspond well with histological detail. For example in Figure 12(a) areas shaded white (high likelihood of membership of the tumor class) correspond exactly with the tumor area on the histological section (cf. Figure 2). Similarly, the light regions in the membership maps shown in Figure 12(b) and (c) correspond to stroma and dermis regions in the stained tissue section.

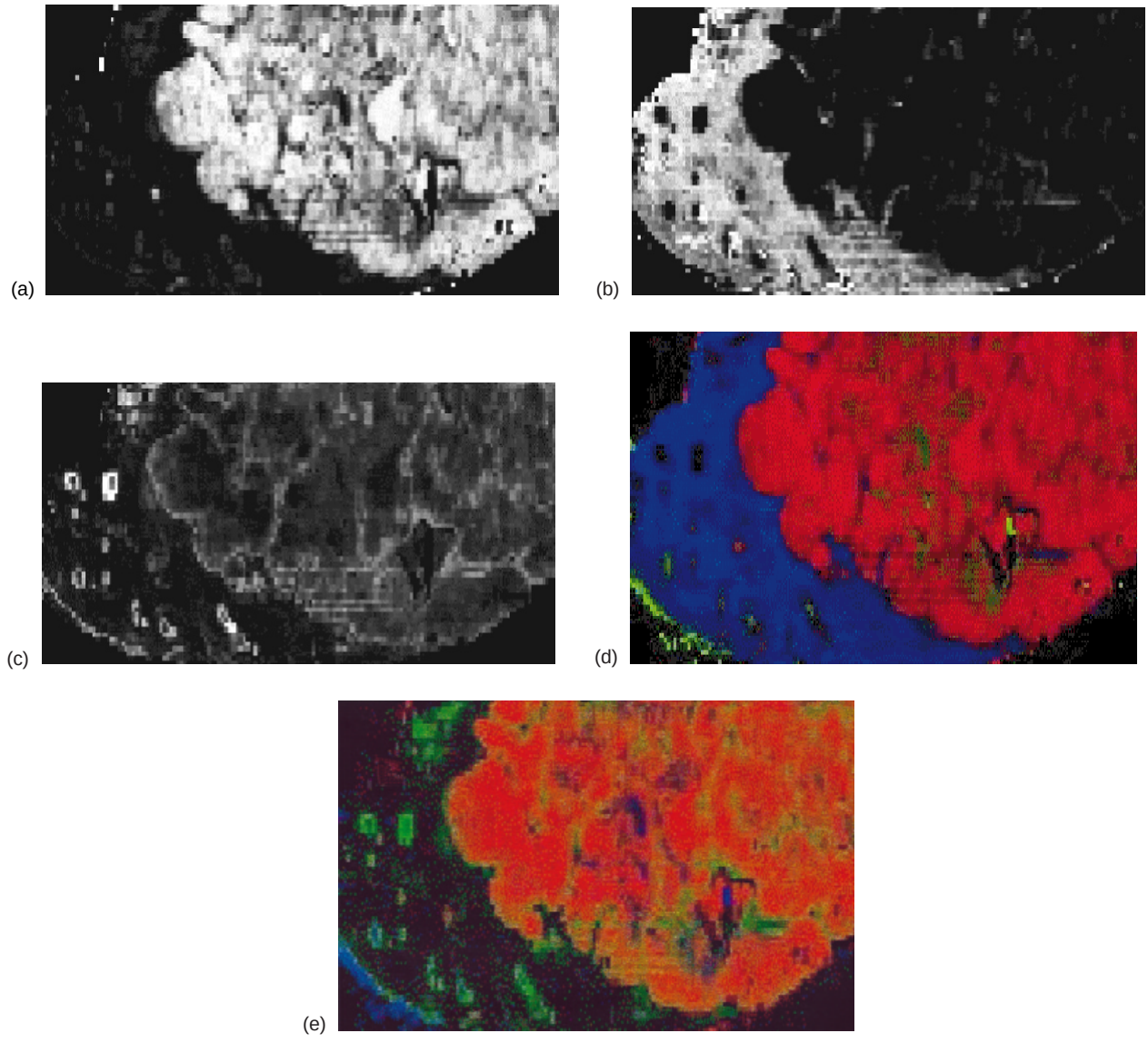

Figure 12 Gray scale maps indicating the likelihood of tissue belonging to (a) tumor, (b) dermis or (c) stroma as predicted by LDA. Digital staining was then applied to produce color-coded membership likelihood maps for (d) tumor (red), dermis (blue) and epidermis (green) and (e) tumor (red), follicle sheath (green) and epidermis (blue). See text for more detail.

Considerable fine structure is seen for some membership likelihood maps. For example, the stroma membership likelihood maps clearly show regions of stroma running through the body of the tumor. While the likelihood of these pixels containing stroma is high, it is significantly below 1 and the stromal fibers running through the tumor appear as gray pixels. Presumably this results from the fact that even with relatively small aperture sizes, spectra from these pixels contain contributions from both the fibrous stroma running through the tumor and also from tumor cells, reducing the stroma membership likelihood for these pixels. Interestingly, stromal regions are also identified within the dermis (bright areas in Figure 12c). The most likely origin of this ''stromal signature'' is the collagen sheath that surrounds hair follicles and sebaceous glands. This suggestion is supported by the concordance between the position of these bright regions and follicle and sebaceous glands within the dermis in Figure 2.

To display the likelihood of membership of more than one class simultaneously, the digital staining approach may be used. In this case, three membership likelihood gray scale maps are converted to red, green and blue scale maps and the images combined as described above. Superimposing three maps such that the color displayed for each pixel is the combined RGB value from the three individual maps then creates a composite image which provides information about the likelihood of membership of a pixel in three classes simultaneously. In essence, this produces a color-coded membership likelihood map.

Examples of color-coded membership likelihood maps produced by digital staining of the LDA results are illustrated in Figure 12(d) and (e). Figure 12(d) combines the likelihood maps for tumor (red), epidermis (green) and dermis (blue). Thus, a high likelihood of a pixel falling into the tumor class results in the appearance of a red pixel, in the epidermis class a green pixel and in the dermis class a blue pixel. The striking feature of the digitally stained LDA results is the excellent discrimination between the three classes, and how closely the classes correspond to histological tissue compartments. There can be no question that spectra of basal cell carcinoma tumor cells are distinct from any of the other tissue components in this skin section (green and blue). The demarcation between tumor and dermis can clearly be seen, providing excellent tumor marginalization.

Figure 12(e) combines the likelihood maps for tumor (red), epidermis (blue) and follicle sheath/sebaceous glands (green). Again, excellent discrimination is seen between the three tissue types. This is particularly gratifying given the fact that the epidermis, tumor and follicle sheath share a common cellular origin (epidermal cells). Despite this similarity in origin a clear differentiation between these three tissue components is seen, highlighting the power of this analytical approach. Interestingly, there is currently some debate as to the origin of basal cell carcinoma. While many investigators feel that the tumors are epidermal in origin, others feel that the follicle sheath is a more likely point of origin. Whatever the origin of the tumor, it is clearly a distinct spectroscopic entity.[33]

7 SUMMARY

Long before the instrumental advances that revolutionized IR spectroscopy, Sutherland and Thompson commented that ''the infrared spectrum of a chemical compound is probably the most characteristic property of that compound''.[34] Half a century later technical developments have firmly established IR spectroscopy as the most widely applicable analytical tool. This highly versatile technique has now found yet another new application: ex vivo characterization of tissues, cells and biofluids. While this application is new, the general principles that guide design and implementation of all good spectroscopic experiments hold true. Obtaining adequate spectra requires the correct choice of sampling technique, appropriate sample preparation and common sense. With these precautions observed, a wealth of information can be extracted from spectra. However, to make use of this information a unique combination of spectroscopic and histological knowledge is required.

The complexity of the information obtained from spectroscopic studies of tissues, particularly those involving IR microscopy, provides a daunting challenge to the IR spectroscopist, but one that can be overcome. Statistical methods play an important role in analysis of large IR spectroscopic data sets, although this field is still in its infancy. Equally important, effective presentation of the large volume of data which is routinely generated is crucial, and methods such as gray scale imaging and digital staining described here will be important factors in determining the degree to which IR spectroscopy becomes generally accepted as a tool for tissue characterization.

In summary, IR spectroscopy has become a powerful tool for the analysis of tissues, cells and biofluids. With a few simple precautions and a multidisciplinary approach, information relating to normal biochemistry and pathological processes within tissues can be reliably obtained. This information may be used to further the understanding of disease processes or may be used to develop novel methods for diagnosis, staging, prognosis and therapeutic monitoring.

ABBREVIATIONS AND ACRONYMS

AD	Alzheimer's Disease
ATR	Attenuated Total Reflectance
DMPC	Dimyristoylphosphatidylcholine
DMSO	Dimethylsulfoxide
FCM	Fuzzy C-means
IR	Infrared
ISODAT	Interactive Self-organized Data Analysis Techniques
LDA	Linear Discriminant Analysis
OCT	Optimal Cutting Tool
TFE	Trifluoroethanol

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Microbiology • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Chemometrics **(Volume 11)**
Clustering and Classification of Analytical Data

REFERENCES

1. L.-P. Choo, M. Jackson, W.C. Halliday, H.H. Mantsch, 'Infrared Spectroscopic Characterisation of Multiple Sclerosis Plaques in the Human Central Nervous System', *Biochim. Biophys. Acta*, **1182**, 333–337 (1993).
2. H. Fabian, M. Jackson, L. Murphy, P.H. Watson, I. Fichtner, H.H. Mantsch, 'A Comparitive Infrared Spectroscopic Study of Human Breast Tumors and Breast Tumor Cell Xenografts', *Biospectroscopy*, **1**, 37–46 (1995).
3. P. Lasch, W. Wasche, W.J. McCarthy, G. Muller, D. Naumann, 'Imaging of Human Colon Carcinoma Thin Sections by FTIR Microspectroscopy', *Proc. SPIE*, **3257**, 187–198 (1998).
4. M. Jackson, H.H. Mantsch, 'The Medical Challenge to Infrared Spectroscopy', *J. Mol. Struct.*, **408/409**, 105–111 (1997).
5. M. Jackson, H.H. Mantsch, 'Biomedical Infrared Spectroscopy', in *Infrared Spectroscopy of Biomolecules*, eds. H.H. Mantsch, D. Chapman, Wiley-Liss, Toronto, 311–340, 1996.
6. M. Diem, S. Boydston-White, L. Chiriboga, 'Infrared Spectroscopy of Cells and Tissues: Shining New Light on a Novel Subject', *Appl. Spectrosc.*, **53**, 148A–161A (1999).
7. M. Jackson, H.H. Mantsch, 'Beware of Proteins in DMSO', *Biochim. Biophys. Acta*, **1078**, 231 (1991).
8. M. Jackson, H.H. Mantsch, 'Halogenated Alcohols as Solvents for Proteins: FTIR Spectroscopic Studies', *Biochim. Biophys. Acta*, **1118**, 139 (1991).
9. H.H. Eysel, M. Jackson, A. Nikhulin, R.L. Somorjai, G.T.D. Thomson, H.H. Mantsch, 'A Novel Diagnostic Test for Arthritis: Multivariate Analysis of Infrared Spectra of Synovial Fluid', *Biospectroscopy*, **3**, 161–167 (1997).
10. M.G. Sowa, H.H. Mantsch, 'FTIR Photoacoustic Depth Profiling Spectroscopy of Enamel', *Calcif. Tissue Int.*, **54**, 481–485 (1994).
11. J.E. Katon, 'Infrared Microscopy. A Review of Fundamentals and Applications', *Micron*, **27**, 303–314 (1996).
12. E.N. Lewis, P.J. Treado, R.C. Reeder, G.M. Story, A.E. Dowrey, C. Marcott, I.W. Levin, 'Fourier Transform Spectroscopic Imaging Using an Infrared Focal-Plane Array Detector', *Anal. Chem.*, **67**, 3377–3381 (1995).
13. P. Colarusso, L.H. Kidder, I.W. Levin, J.C. Fraser, J.F. Arens, E.N. Lewis, 'Infrared Spectroscopic Imaging: From Planetary to Cellular Systems', *Appl. Spectr.*, **52**, 106A–120A (1998).
14. L.H. Kidder, I.W. Levin, E.N. Lewis, V.D. Kleinman, E.J. Heilweil, 'Mercury Cadmium Telluride Focal-Plane Array Detection for Mid-infrared Fourier Transform Spectroscopic Imaging', *Opt. Lett.*, **22**, 742–744 (1997).
15. M. Jackson, H.H. Mantsch, 'The Use and Misuse of FTIR Spectroscopy in the Determination of Protein Structure', *Crit. Rev. Biochem. Mol. Biol.*, **30**, 95–120 (1995).
16. C.P. Schultz, H.H. Eysel, H.H. Mantsch, M. Jackson, 'Carbon Dioxide in Tissue, Cells and Biological Fluids Detected by FTIR Spectroscopy', *J. Phys. Chem.*, **100**, 6845–6848 (1996).
17. R.K. Dukor, M.N. Liebman, B.L. Johnson, 'A New, Non-destructive Method for Analysis of Clinical Samples with FTIR Microspectroscopy. Breast Cancer Tissue As An Example', *Cell. Mol. Biol.*, **44**, 211–217 (1998).
18. M. Jackson, H.H. Mantsch, 'Biomembrane Structure from FTIR Spectroscopy', *Spectrochim. Acta Rev.*, **15**, 53–69 (1993).
19. M. Jackson, L.-P. Choo, P.H. Watson, W.C. Halliday, H.H. Mantsch, 'Beware of Connective Tissue Proteins: Assignment and Implications of Collagen Absorptions in Infrared Spectra of Human Tissues', *Biochim. Biophys. Acta*, **1270**, 1–6 (1995).

20. K.Z. Liu, M. Jackson, M.G. Sowa, H. Ju, I.M.C. Dixon, H.H. Mantsch, 'Modification of the Extracellular Matrix Following Myocardial Infarction Monitored by FTIR Spectroscopy', *Biochim. Biophys. Acta*, **1315**, 73–77 (1996).
21. M. Jackson, J.R. Mansfield, B. Dolenko, R.L. Somorjai, H.H. Mantsch, P.H. Watson, 'Prediction of Breast Tumor Grade and Streoid Receptor Status by Pattern Recognition Analysis of Infrared Spectra', *Cancer Detect. Preven.*, **23**, 245–253 (1999).
22. M. Jackson, B. Ranmjiawan, H.H. Mantsch, 'Functional Group Imaging of the Jaundiced Rabbit Liver', *Cell. Mol. Biol.*, **44**, 89–98 (1998).
23. H.C. van-der-Mei, D. Naumann, H.J. Busscher, 'Grouping of Oral Streptococcal Species using Fourier Transform Infrared Spectroscopy in Comparison with Classical Microbiological Identification', *Arch. Oral Biol.*, **38**, 1013–1019 (1993).
24. G.J. McLachlan, 'Discriminant Analysis and Statistical Pattern Recognition', John Wiley, New York, 1992.
25. N. Pizzi, L.-P. Choo, J.R. Mansfield, M. Jackson, W.C. Halliday, H.H. Mantsch, R.L. Somorjai, 'Neural Network Classification of Infrared Spectra of Control and Alzheimer's Diseased Tissue', *Artificial Intelligence Med.*, **7**, 67–79 (1995).
26. L.M. McIntosh, M. Jackson, H.H. Mantsch, M.F. Stranc, D. Pilavdzic, A.N. Crowson, 'Infrared Spectra of Basal Cell Carcinomas are Distinct from Non-tumor Bearing Skin Components', *J. Invest. Dermatol.*, **112**, 951–956 (1999).
27. A.S. Bangalore, R.E. Shaffer, G.W. Small, M.A. Arnold, 'Genetic Algorithm-based Method for Selecting Wavelengths and Model Size for Use with Partial Least-squares Regression: Application to Near-infrared Spectroscopy', *Anal. Chem.*, **68**, 4200–4212 (1996).
28. L.-P. Choo, D.L. Wetzel, W.C. Halliday, M. Jackson, S.M. LeVine, H.H. Mantsch, 'In-Situ Characterization of β-Amyloid in Alzheimer's Diseased Tissue by Synchrotron FTIR Microspectroscopy', *Biophys. J.*, **71**, 1672–1679 (1996).
29. L.M. Mcintosh, J.R. Mansfield, A.N. Crowson, H.H. Mantsch, M. Jackson, 'Analysis and Interpretation of Infrared Microscopic Maps: Visualisation and Classification of Skin Components by Digital Staining and Multivariate Analysis', *Biospectroscopy*, **5**, 265–275 (1999).
30. J.T. Tou, R.C. Gonzalez, 'Pattern Recognition Principles', in *Applied Mathematics and Computation*, ed. R. Kalaba, Addison-Wesley, Reading, 97–104, Vol. 7, 1974.
31. R.L. Somorjai, M. Jarmasz, Richard Baumgartner, 'Exploratory Data Analysis of Functional MRI Data Comparing the Influence of Different Preprocessing Techniques and Distance Metrics on Fuzzy Clustering Analysis', NeuroImage, in press.
32. J.R. Mansfield, L.M. McIntosh, A.N. Crowson, H.H. Mantsch, M. Jackson, 'A LDA-Guided Search Engine for the Non-subjective Analysis of Infrared Microscopic Maps', *Appl. Spectrosc.*, **53**, 1323–1330 (1990).
33. F. Ponten, Z. Ren, M. Nister, B. Westermark, J. Ponten, 'Epithelial–Stromal Interactions in Basal Cell Cancer: The PDGF System', *J. Invest. Dermatol.*, **102**, 304–309 (1994).
34. G.B.B.M. Sutherland, H.W. Thompson, 'Developments in the Technique of Infrared Spectroscopy', *Trans. Faraday Soc.*, **41**, 174–179 (1945).

Magnetic Resonance Angiography

Chun Yuan
University of Washington, Seattle, USA

1 Introduction 1
2 Flow Phenomena in Magnetic Resonance Imaging 1
2.1 Time-of-flight Phenomenon 2
2.2 Flow Induced Phase Shift Phenomena 4
3 Basic Magnetic Resonance Angiography Imaging Techniques 5
3.1 Two- and Three-dimensional Time-of-flight Techniques 6
3.2 Two- and Three-dimensional Phase Contrast Techniques 6
3.3 Cardiac-gated Two-dimensional Phase Contrast Angiography 7
4 Post Processing Techniques and Blood Vessel Visualization 7
5 Recent Developments 7
5.1 Contrast-enhanced Magnetic Resonance Angiography 7
5.2 Coronary Magnetic Resonance Angiography 9
6 Comparison with Other Angiography Techniques 9
Acknowledgments 10
Abbreviations and Acronyms 10
Related Articles 10
References 10

Angiography generally refers to any imaging method that can visualize blood vessels. Magnetic resonance angiography (MRA) is a class of magnetic resonance (MR) imaging techniques that create blood vessel images. During the past decade, there has been a tremendous amount of development and refinement in MRA techniques, making MRA a powerful tool in the diagnosis of cardiovascular disease. In most of the cases, MRA techniques are applied noninvasively, without the use of ionizing radiation. The inherent advantages of MR imaging (noninvasive, flow sensitive, high soft tissue contrast, and flexible imaging approaches) make it ideally suited to the study of the vasculature in almost any part of the human body. MRA presents a wide spectrum of techniques for the study of flowing blood: different methods for high spatial resolution, high temporal resolution, velocity sensitization, flow quantification, and visualizing pulsatility. This article addresses the basic principles of MRA and describes its applications.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Blood flow phenomena and the basics of flow measurement with MR were recognized long before the invention of MR imaging.[1,2] Singer proposed to determine blood flow rates using MR relaxation time measurements in mice. However, flow phenomena in actual MR images of humans were not reported until 1977, by Hinshaw.[3] He observed that in single-slice transverse images of the wrist, vessels appeared as high intensity structures. Developments in medical applications of MR imaging have caused renewed interest in flow for two reasons: first, flow effects are a significant source of image artifacts in many clinical applications.[4,5] Second, and perhaps most importantly, flow imaging techniques can be devised which exploit these effects, showing clear pictures of blood vessels and allowing for the study of blood flow phenomena (Figure 1). Since 1985, when MRA was first shown to be clinically feasible,[6] the imaging techniques and hardware used for MRA have greatly improved. As little as fifteen years ago, MRA was rarely utilized clinically, whereas it now has become commonplace in the standard care of patients.[7–10]

2 FLOW PHENOMENA IN MAGNETIC RESONANCE IMAGING

All MR blood flow phenomena arise from one of the two basic properties of moving nuclei (proton or spin) in a magnetic field. The first property is simply the change in position of spins (for example: blood) as they move during the scanning procedure. In this case, the MR signals measured are primarily related to the longitudinal components of the magnetization of spins. The second property is a change in the phase of the transverse magnetization of spins. This occurs when spins move in the direction of a magnetic field gradient applied during a scan. The MR signal detected by a scanner is the transverse component of the magnetization, modulated by phase and longitudinal components. The first property is usually referred to as the TOF phenomenon and the second as flow-induced phase shift phenomenon.

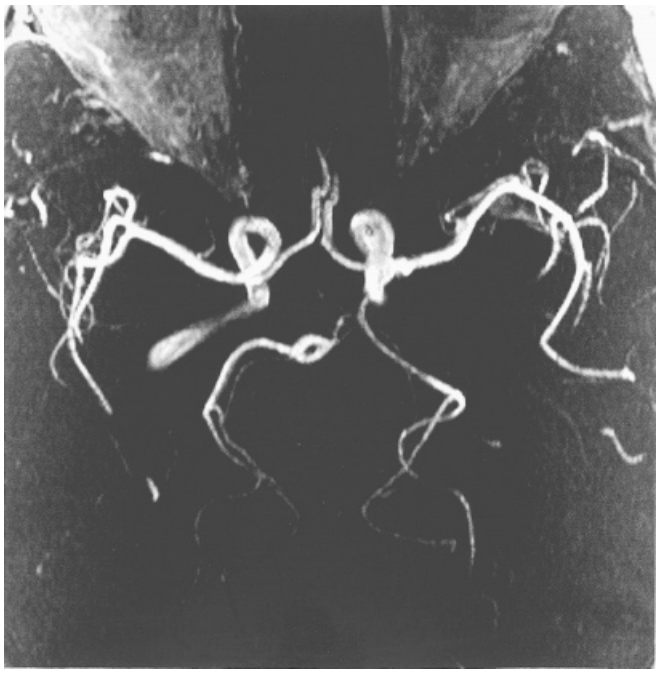

Figure 1 A three-dimensional (3D) time-of-flight (TOF) image of a normal human brain MRA. This is a top view of a region centered around the Circle of Willis, assembled from a stack of cross-sectional images. A phased-array coil was used to acquire this set of data for optimal signal-to-noise performance. (Courtesy of Maravilla KRM, University of Washington, Seattle, WA.)

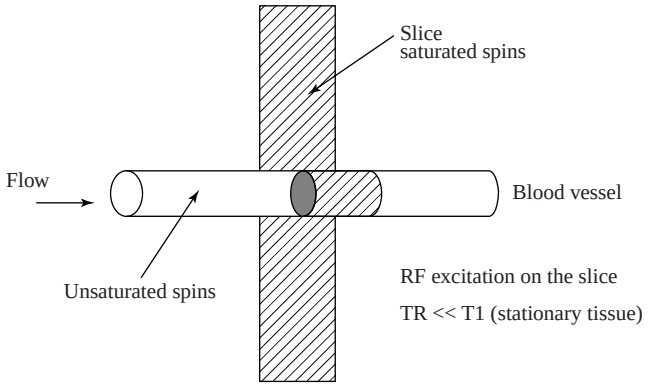

Figure 2 Illustration of inflow enhancement. With the application of RF pulses at every TR (TR $\ll$ T1 of stationary tissue), the longitudinal magnetization, M_Z, of spins in the imaging slice is reduced due to partial saturation. Blood flowing into the imaging slice, however, causes moving spins to enter the slice with full magnetization. These spins contribute a higher signal intensity than the surrounding, stationary tissue. This enhancement is dependent on the flow velocity, and also on the orientation of the vessel relative to the imaging plane.

2.1 Time-of-flight Phenomenon

It is well known that the excitation and detection of a MR signal are not simultaneous. Bulk motion of flow during image acquisition can give rise to TOF effects. TOF effects are changes in the MR signal due to spins flowing into, or out of, an imaging volume during the application of an imaging sequence. For example, if a vessel containing flowing blood passes through a slab of tissue being imaged, the bolus of blood which experiences the 90° radiofrequency (RF) pulse may be outside the slab when the refocusing 180° RF pulse occurs. In this case, the magnetization of the blood is not refocused, and no signal is observed within the vessel. This is usually referred to as ''flow void''.[11] Based on this phenomenon and other flow signal suppression schemes, one class of MRA techniques, named ''black blood MRA'', have developed.[12] However, another TOF effect, referred to as ''inflow enhancement'' or ''bright blood MRA'', is used more widely in clinical application.

The cause of inflow enhancement is physical displacement of the spins between successive RF excitation. This enhancement is especially true for gradient echo based imaging sequences, since a single RF pulse (as opposed to a 90°–180° train in spin echo sequence) is applied for each data acquisition. The basic idea is illustrated in Figure 2. Let us assume the blood flow to be perpendicular to the imaging slice. For an image sequence, repetition time (TR) is shorter with respect to the stationary tissue's longitudinal relaxation time (T1) within the slice, where the magnetization becomes saturated by the RF excitations and gives low signal intensities. Blood flow in the vessel moves spins from outside the slice, which are in an unsaturated or fully relaxed state, into the imaging slice. These spins, with full equilibrium magnetization, produce a much stronger signal than stationary spins. The stronger signal creates the desired contrast between flowing blood and stationary tissues. This enhancement would be difficult to achieve with a spin echo sequence due to the flow void effect.

The TOF effect is flow velocity dependent. Thus, with the same imaging plane, regions with fast flow will appear brighter than the regions with relatively slow flow. In general, TOF MRA techniques are designed to generate positive contrast between flowing blood and background stationary tissues. There are three essential and interdependent factors for optimal use of TOF effect:

1. suppression of background tissue signal;
2. signal enhancement of moving spins with accurate spatial depiction;
3. sufficient MR image quality to properly represent both flow and surrounding structures.

There are many TOF MRA imaging sequences.[13–22] Generally, these techniques are designed to balance among the three factors for specific applications. Design

considerations are on imaging parameters, such as TR, echo time (TE), strength of the RF excitation (as specified by the flip angle), slice thickness, and flow dynamics at the region of interest. Additional design considerations include: blood vessel geometry, orientation, and surrounding environment.

The wide range of physiological hemodynamic effects results in many significant technical challenges for TOF MRA techniques. Ideal TOF effects result from constant flow and speed across a vessel perpendicular to the imaging plane. Blood, however, travels at velocities that range from near zero to about $1\,\mathrm{m\,s^{-1}}$, and constantly changes both velocity and direction. The time duration for peak flow can be as short as a few milliseconds, then the flow may slow down and even reverse direction. With each cardiac cycle, this whole process is repeated and modulated by respiration and body motion. Blood may pass through tortuous vessels and focal narrowings that reduce the cross-section by >95%. These morphological features as well as surface irregularities, such as ulceration and atherosclerotic plaque, disturb the smooth laminar flow of blood through the vessel, and therefore complicate TOF techniques.

Another major image quality concern is the extra phase accumulated when spins move during the application of gradient magnetic fields. The phase accumulation can be described by a simple integral over time (Equation 1):

$$\Phi = \gamma \int \vec{G} \cdot \vec{x}\, \mathrm{d}t \tag{1}$$

where γ is the gyromagnetic ratio, $\vec{x}$ is the physical location of spins expressed as a vector and $\vec{G}$ is the magnetic gradient field applied as a function of time expressed as a vector for gradient field applied in all three axes. As a simple example, when spins are moving with a constant velocity v along x direction during the application of a constant gradient G_x along x direction for a period T, the phase is given by Equation (2):

$$\Phi = \gamma G_x \left(x_0 + \frac{vT^2}{2} \right. \tag{2}$$

where x_0 is the spin location at the start of the gradient application. A zero or constant phase shift will not affect the final magnitude data (MR image). For flow with variable velocity, however, the extra phase accumulated may cause image artifacts to appear as a streak of ghosting (image artifacts) along one of the data acquisition axes, as shown in Figure 3. In addition, phase distribution within a pixel may cause intra-pixel dephasing that can lead to signal cancellation.

An effective technique to deal with the effect of extra phase accumulation is gradient moment nulling.[23–25] Based on Equation (2), when $v = 0$, a bipolar shape gradient (two lobes, one positive and one negative) can be easily used to set the phase to zero. If $v \neq 0$, it is possible to design a polynomial gradient shape to set the phase to zero. This approach is called gradient moment nulling. In most cases, gradient moment nulling is applied to set the phase to zero for constant flow, and can be applied to all three directions. It has been very effective even for

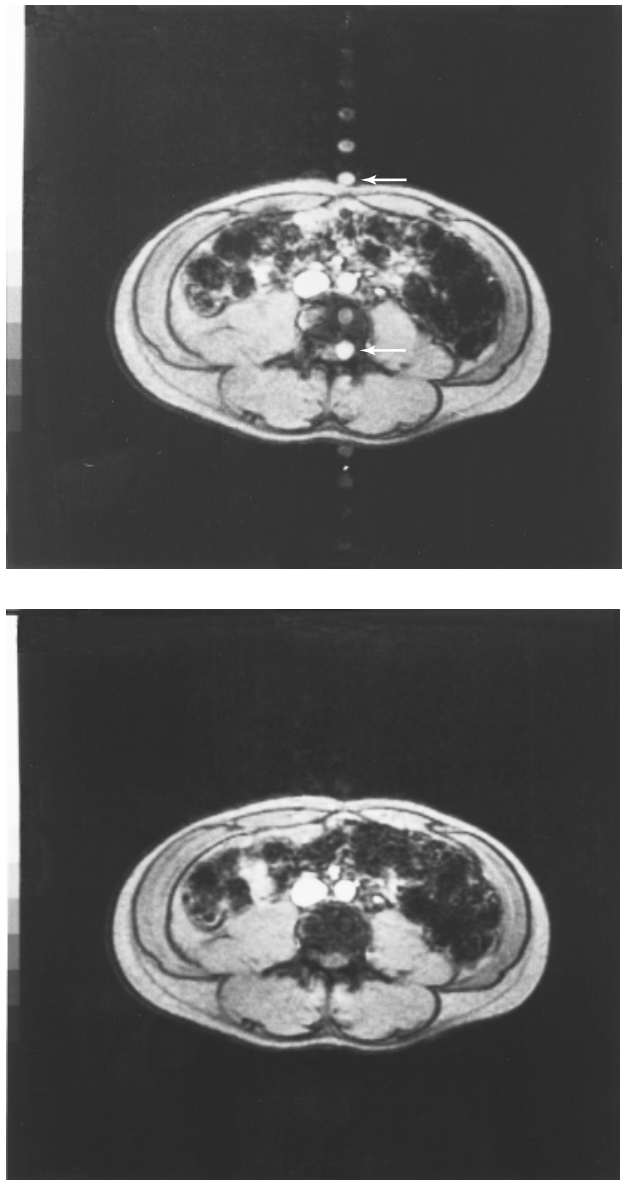

Figure 3 Illustration of flow related image artifacts on cross-sectional TOF images of the abdominal region of a normal volunteer. (a) is taken with no flow compensation along the cross-sectional direction and (b) is taken with flow compensation in that direction. Notice the ghosts (arrows), which are centered on the aorta and spread periodically on a line parallel to one direction (phase encoding) of the image, and also the removal of these ghosts with the application of gradient nulling based flow compensation.

situations where flow velocity is a variable (see Figure 3). Generally, gradient moment nulling prolongs the TE.

Image contrast in TOF images depends on enhancing signals from flowing blood, and suppressing signals from stationary tissues. Suppression of background tissue signal can be accomplished through the use of rapid TRs (short TR), and relatively high RF power levels (high flip angle). However, the same saturation process that reduces stationary signal can also reduce signal from flowing spins. Flow enhancement, on the other hand, can be achieved by placing the imaging slice perpendicular to the direction of flow. For example, cross-sectional (axial) TOF images of the carotid artery tend to have very good flow enhancement because of the orientation of the artery and the imaging plane. For certain arteries, however, it may not be feasible to place a slice perpendicular to the flow direction.

2.2 Flow Induced Phase Shift Phenomena

The second class of flow effects arises from changes in the phase of transverse magnetization induced (of spins) as the blood moves along a magnetic field gradient.[26,27] The phase accumulation is described in Equation (1). During an imaging data acquisition, magnetic field gradients are applied for short periods of time for spatial encoding. During the short period of a gradient pulse, nuclei at different locations are at slightly different frequencies. This causes the phase of transverse magnetization to change. Typically, bipolar gradients are applied so that any dephasing is compensated by a gradient of the opposite sign, so that the phase shift accumulated equals zero.

From Equations (1) and (2), it is clear that the position dependent phase shifts generated by a gradient pulse are proportional to blood flow velocity and the amplitude of the gradient pulse applied at the direction of flow. This indicates that gradient pulses can be shaped so that a desired phase shift can be achieved from moving spins while stationary tissues still maintain a zero phase shift after the gradient trains. A gradient train that satisfies this condition is usually referred to as a flow encoding gradient.

There are many different ways to design the shape of gradients for flow encoding. A typical scheme[28] is shown in Figure 4. This scheme has become known as phase-contrast (PC) imaging. With the PC method, two sets of data are acquired under identical conditions, with the exception of the polarity of the flow encoding gradient pulses. Assuming constant flow v, from Equation (1), the phase Φ can be expressed in two terms:

$$\Phi_{\text{velocity}} = \Phi_{\text{stationary}} + \gamma v T A$$

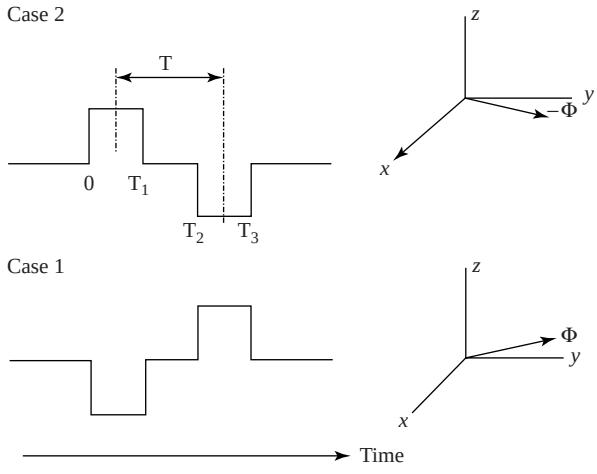

Figure 4 Illustration of phase modulation of the MR signal during the application of flow encoding gradient fields. In each data acquisition, a bipolar gradient is applied which in case 1 generates a phase shift Φ due to motion. In the second data acquisition, the gradient field is of the same duration and amplitude, only their polarities are different. Thus, the phase shift from case 2 is $-\Phi$. The two acquired data sets are identical except for the phase shift arising from flow.

where A is the area of each gradient lobe and T is center-to-center time interval between the two gradient lobes. With bipolar gradient, $\Phi_{\text{stationary}} = 0$, thus, Φ_{velocity} is proportional to spin velocity v. For subsequent acquisition, this gradient pulse inverts the polarity of the bipolar flow-encoding gradients. The phase Φ_{velocity} is $-\gamma v T A$. The only procedural difference in these two acquisitions is the negation of the bipolar flow-encoding gradients. This time related phase accumulation may cause intra-pixel dephasing and signal loss. Each flow encoding gradient, therefore, must have sufficient amplitude to cause a detectable phase shift in the MR signal due to flow, and yet must be small enough to minimize phase cancellation caused by velocity distribution within the imaging voxel.

There are multiple ways of extracting flow information from the imaging system. A simple example is to subtract the magnitude of the first image data set from the second. In this case, the only remaining data are from the moving spins. Alternatively, the phase of the MR signal can be extracted by complex arithmetic on the acquired data sets.

2.2.1 Velocity Encoding and Aliasing

With PC MRA, signal intensities at each pixel are proportional to the flow related phase shift, and therefore also proportional to velocity. If the phase shift is too large, however, it becomes indistinguishable from a smaller phase shift due to the 2π range limit. This cyclical

nature of phase shift is usually referred to as aliasing. Aliasing occurs when higher flow velocities are incorrectly represented in the speed image as lower velocities due to the 2π phase shift range. To provide quantitative information regarding velocity in PC MRA, the velocity encoding (VENC) should be selected to encompass the highest velocities that are likely to be encountered within the vessel of interest. When a VENC is selected, the amplitude of the bipolar flow-encoding gradients is adjusted so that all velocities up to the selected value can be imaged without aliasing.[29]

Flow encoding gradient pulses induce a phase shift in transverse spin magnetization that is proportional only to the component of velocity parallel to the direction of the applied magnetic field gradient. Consequently, an angiogram containing information for all components of velocity must be constructed from data acquired in three orthogonal flow-sensitive directions. Modern phase-sensitive pulses acquire 3D flow information using a four-excitation scheme.[28,30,31]

PC technique is distinctly different from TOF technique, and has several features that distinguish it from other angiographic techniques as well.

1. Multiple data acquisitions with different flow encoding gradients are needed for 3D flow detection. Thus the overall scan time is longer than the TOF MRA for a similar range of coverage and spatial resolution.
2. Multiple types of images can be generated. A magnitude image can be generated that shows the anatomy, which usually has standard gradient echo image contrast. This magnitude image is generated using a complex difference method that is discussed later. A phase image shows the flow map with signals from stationary tissue removed (Figure 5). Based on this phase map, flow speed in a certain direction can be measured. In addition, image contrast is a direct reflection of flow speed and direction.
3. With PC MRA, images of high velocity blood travelling in large vessels within a large field-of-view can be acquired as easily as can slowly moving blood in small vessels. This is different from TOF MRA where image contrast reflects different levels of spin saturation. Thus, TOF MRA is sensitive to parameters such as vessel geometry and blood velocity that influence the inflow enhancements.
4. Excellent suppression of stationary tissue is possible with PC MRA because the main source of image contrast is phase shift or motion.
5. PC MRA can be used to perform quantitative velocity and flow measurements. Methods currently in use include thin-slice methods in which vessels are imaged in cross-section and the intensity of each pixel of the image is made proportional to the flow-induced phase shift.[32–35]
6. Another characteristic of PC MRA is that the quality of the angiogram is heavily dependent on the quality of the instrument. In particular, the phase stability of the instrument is critical because motion is detected and quantified by measuring phase shifts in the MR signals. Eddy currents (occurring during fast switching on and off of a gradient field) are a common source of phase instability in a MR system and can lead to poor image quality.

(a) (b)

Figure 5 Phase and magnitude image of a normal volunteer's carotid artery using a gated two-dimensional (2D) phase contrast imaging. (a) is the phase map and (b) is the magnitude image. In (a), stationary tissue is almost invisible due to a near zero phase. The carotid artery appears dark and the jugular vein appears bright. The darkness of the carotid artery indicates the flow direction in that artery is from inferior to superior and vice versa. In (b), artery, vein, and background stationary tissue are visualized. Note the relative positioning of the blood vessel to the background. The flow directional information, however, is not presented in this image. (Courtesy of GE Medical Systems, Waukesha, WI.)

3 BASIC MAGNETIC RESONANCE ANGIOGRAPHY IMAGING TECHNIQUES

There are three MRA imaging techniques: 2D/3D TOF, 2D/3D phase contrast, and cardiac gated phase contrast (these are the names used on scanners made by General

Electric Medical Systems, different manufacturers tend to use different terminology to describe the same techniques).[29]

3.1 Two- and Three-dimensional Time-of-flight Techniques

TOF MRA is used extensively in the diagnosis of lumen narrowing of carotid bifurcations and of intracranial aneurysms. It is used to assess the pathologic thoracic and abdominal aorta as well as for lower extremity runoff studies.[36–43] These techniques are most effective when ''inflow enhancement'' is highly pronounced. A typical example is the cross-sectional (axial) imaging of the carotid artery, where blood flow is almost completely perpendicular to the imaging plane. As mentioned in the previous section, the amount of inflow enhancement depends on several factors, including tissue specific parameters such as T1, and sequence specific parameters such as flip angle, slice thickness or blood flow velocity. The 2D and 3D techniques are illustrated in Figure 6. In Figure 6(a), a sequential 2D technique is shown that provides multiple thin sections of the vessels. Alternatively, a 3D technique can be applied as demonstrated in Figure 6(b). The whole volume is excited simultaneously, then subdivided into thin partitions or slices by using an additional phase encoding scheme in the slice select direction. Unlike 2D imaging, in which the slice resolution is defined by the excitation profile of the RF pulse, here the slice resolution is defined by the spatially encoding gradients and can be less than 1 mm thick. In 3D imaging, complete compensation for the flow-induced phase shifts is also necessary to avoid flow artifacts in the form of signal losses and ghosting. Table 1 shows a comparison of the advantages and disadvantages of 2D and 3D TOF angiography.

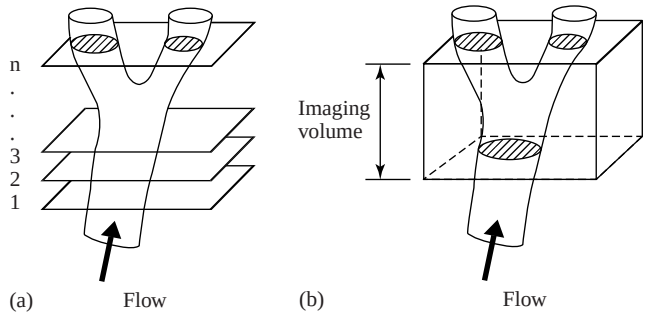

Figure 6 Comparison of 2D and 3D TOF MRA techniques. A 2D technique (a) is based on the acquisition of multiple thin slices at different locations in the vessel at different times. In general, images at different locations are acquired sequentially such as based on the image numbering system. This technique takes advantage of in-flow enhancement at every imaging location. In 3D technique (b), the whole volume is excited simultaneously, and spatial resolution along the slab direction is achieved, similar to the in-plane resolution by the application of gradients. Thus, the slice thickness can be thinner in 3D acquisition. But the magnetization of flowing spins traversing the imaging volume may be saturated to result in reduced in-flow enhancement, especially at the distal end of the volume.

3.2 Two- and Three-dimensional Phase Contrast Techniques

PC techniques can also be divided into 2D and 3D acquisitions. 2D PC angiograms can be acquired very quickly. This makes the 2D technique useful as a localizer to determine the locations of blood vessels for detailed studies. 3D PC technique, similar to 3D TOF technique, enjoys higher spatial resolution. For the purposes of post processing, where an isotropic voxel (same size along all cubic directions) is desired, 3D PC technique has definitive advantages. This technique, however, can require a long data acquisition time. For example, when flow encoding is required along all three directions, the scan time using 3D PC can be four times longer than a 3D TOF technique that covers a comparable region. The benefits are the excellent background signal suppression, the clear depiction of vessels with fast

Table 1 Comparison of 2D TOF and 3D TOF techniques

	2D TOF	3D TOF
Cross plane flow	Sensitive to both slow and fast flow	Slow flow may be saturated while traversing the imaging volume
In-plane flow	Insensitive	Sensitive
Imaging time	Relatively short	Relatively short
Image signal-to-noise ratio	Low	High
Spatial resolution	Low (especially in z direction)	High

Table 2 Comparison of 2D PC and 3D PC techniques[29]

	2D PC	3D PC
Variable velocities	Sensitive	Sensitive
Saturation effects	Minimized (to 2D TOF)	Minimized (to 3D TOF)
Imaging time	Relatively short	Relatively long
Image signal-to-noise ratio	Low	High
Spatial resolution	Low (especially in z direction)	High
Magnitude and phase images	Yes	Yes

and slow flow, and the information on flow direction (Table 2).

3.3 Cardiac-gated Two-dimensional Phase Contrast Angiography

Cardiac-gated 2D phase contrast MRA is based on cine (or time-resolved) MR acquisition methods in which TR remains constant, and cardiac gating is used to trigger the detection of images at multiple points in the cardiac cycle.[44,45] The cine gradient echo pulse sequence is modified to include bipolar gradients for positive and negative flow encoding, in a fashion analogous to multi-slice 2D phase contrast. VENC is also similar to non-gated 2D PC. With this approach, data at multiple points (for example 16) in the cardiac cycle are sorted retrospectively from the scan data. Magnitude and phase images are then generated for each time point in the cardiac cycle. On the phase images, signal intensity is proportional to blood flow velocity for each phase of the cardiac cycle. The phase images also display the direction of blood flow in the same direction as the flow-encoding gradient, a dark signal indicates flow in the opposite direction. The Cine PC technique is useful for imaging pulsatile blood flow in peripheral arteries. Imaging in which thin slices are excited and the velocity-induced phase of each pixel is displayed to provide a quantitative measure of blood velocity has become widely used in the past few years. Cardiac gated PC methods have been applied in head,[46] aorta,[47] renal,[48] mesenteric,[49] and peripheral vessels.[50]

4 POST PROCESSING TECHNIQUES AND BLOOD VESSEL VISUALIZATION

Post processing of MRA image data is almost a necessity in clinical applications for demonstrating spatial positioning, connectivity of vessels, and correlative MR imaging. In addition, with many MRA techniques it is possible to accumulate upwards of 100 images. There is a need to integrate the information provided by individual images and display them in the familiar form of an angiogram. This is most frequently accomplished with a maximum intensity projection (MIP) algorithm.[51] The MIP algorithm, performed on a computer, finds the maximum intensity pixel along a specified projection line in all slices of a MRA data set. These maximum intensity pixels are then copied onto the projection plane, forming the projection image. With high-contrast bright blood images, the result is an angiogram. Unlike X-ray angiography, the MIP pixel signal intensities are not additive. The algorithm will only detect the maximum intensity pixel and only that pixel will be projected. Projection angles can be arbitrary.

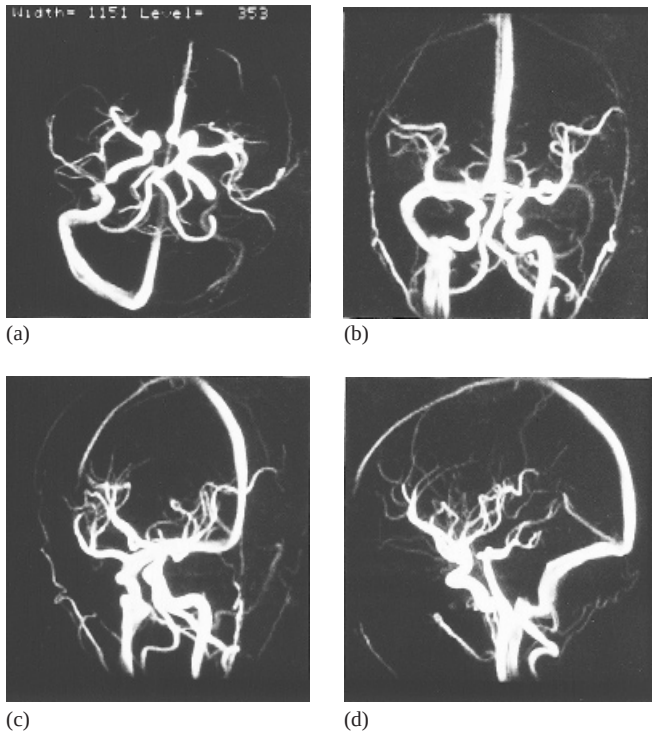

Figure 7 Maximum pixel projection of the intracranial arteries obtained from 3D phase contrast MRA technique. Four projections are presented (a) top to bottom, (b) front to back, (c) side to side, and (d) left to right. The middle cerebral artery, anterior cerebral artery, and the sagittal sinus are well visualized. Note that depth perception is not presented in these images. (Courtesy of GE Medical Systems, Waukesha, WI.)

Hence the 3D MRA data set can be viewed from multiple angles for viewing purposes (Figure 7). However, some problems arise with this method. One is that the overlying vessels may complicate the view angle. Another is the lack of depth information on the projection image. There are techniques introduced to enhance the MIP algorithm and to explore other methods of processing and display.[52–55]

5 RECENT DEVELOPMENTS

5.1 Contrast-enhanced Magnetic Resonance Angiography

Both TOF and PC MRA techniques are essentially non-invasive and rely on in-flow enhancement or phase shift to generate contrast between flowing blood and stationary tissues. There are several disadvantages associated with these techniques. One is inplane saturation, which can be a problem with slow flowing or tortuous arteries. A second is turbulence-induced signal loss in a region of narrowing. Imaging times also tend to be long, especially for covering a large vasculature. Recent advances

in the implementation and understanding of 3D contrast enhanced MRA are making vascular imaging of the body safer and more accurate.[56,57] This is accomplished by using paramagnetic contrast media injected intravenously. Gadolinium chelate is a paramagnetic metal ion that decreases both the spin-lattice (T1) and spin-spin (T2) relaxation times.[58] These agents have been widely used to generate contrast enhanced MR images, and are also now used for angiographic images. These extracellular agents diffuse from the intravascular compartment into the interstitial space in a matter of minutes. Due to the strong enhancement effect of paramagnetic contrast media, a small dose injected as an intravenous bolus is sufficient to briefly enhance the entire arterial vascular tree. However, imaging time must be short, and timing is critical to get the desired enhancement.[57,58] By using a 3D gradient echo pulse sequence on scanners with high performance gradient systems, high resolution 3D volumes of image data can be acquired in a single breath-hold. The 3D acquisition allows imaging with a large field-of-view that encompasses an extensive region of vascular anatomy. Newly deployed hardware and software on commercial scanners has vastly improved image quality of 3D contrast MRA exams. This is particularly true in the chest and abdomen, where breathing has always been a problem for imaging. Subsequent post-processing, in the form of edited maximum pixel projection, allows an angiographic display of image data in any desired obliquity. Figure 8 is an example of contrast enhanced MRA of human carotid arteries. Figure 9 is an example of contrast enhanced MRA of human abdominal region.[59–62] Clinical applications of contrast enhanced MRA can be found in many different parts of the human body.[63–68]

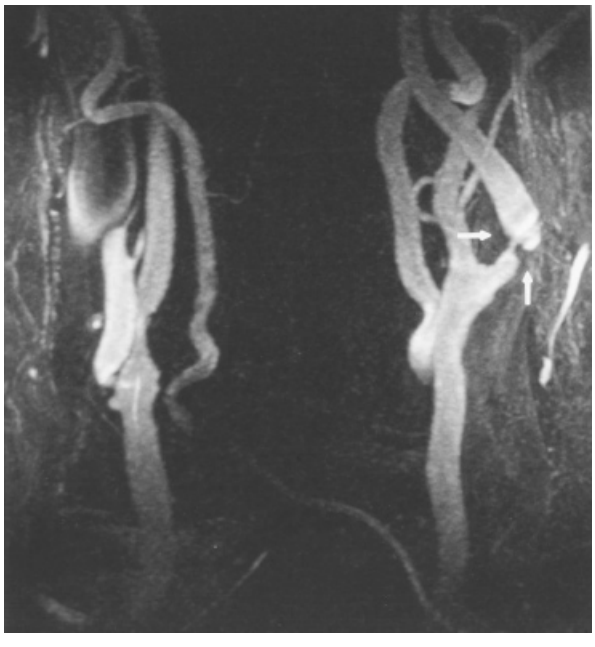

Figure 8 A front-to-back MIP image from a stack of coronal contrast enhanced 3D TOF MRA images of carotid arteries. Phased-array carotid coils were used to improve the signal-to-noise performance of the scanner at the carotid bifurcation region. The overall data acquisition time was 40 s. A tight stenosis is delineated on the left internal carotid artery (see arrow).

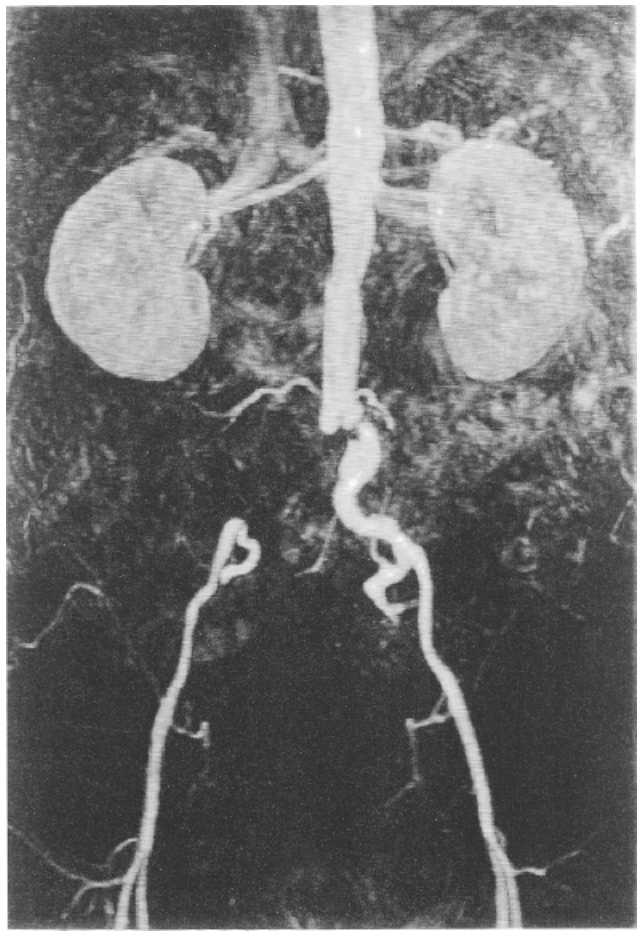

Figure 9 An abdominal TRICKS exam from a patient suffering from peripheral vascular disease. 3D TRICKS time frames were formed every 7 s after the injection of 15 mL of a gadolinium based contrast agent. Imaging parameters were: TR(ms)/TE(ms)/flip angle = 7.8/1.7/45, field of view = 48 cm × 36 cm. The peak arterial time frame demonstrates excellent depiction of the renal arteries as well as the infra-renal aorta and illiac arteries. 3D TRICKS is a time-resolved, contrast enhanced, 3D MR technique developed to image the passage of the contrast agent that: 1) increases the likelihood that an arterial-only 3D image set will be obtained. 2) Permits the passage of the contrast agent to be observed, and 3) allows temporal-processing techniques to be applied to yield additional information or improve image quality.[59–62] (Courtesy of Drs. Shannon Swan and Charles A. Mistretta of the University of Wisconsin, Madison, WI.)

5.2 Coronary Magnetic Resonance Angiography

Coronary artery disease (CAD) is the leading cause of death in many countries. X-ray contrast angiography is widely accepted as the definitive method of defining coronary anatomy. This procedure, however, is associated with significant radiation exposure, and thus carries risk for potentially serious complications. Additionally, the procedure is expensive. An alternative, noninvasive technique that could reliably provide both anatomic and functional information about the coronary circulation would be an important advance in imaging. MR coronary artery angiography was attempted almost at the initiation of MR imaging.[69] To date, however, even with significant improvement of scanner hardware, software, imaging technique and image quality, coronary MRA remains mainly a research tool used only at a limited number of institutions. The main reason is that the coronary tree is a small, convoluted, 3D structure that changes shape during heart contraction. During contraction, motion of the coronary tree involves both displacement and rotation. This motion is highly variable among different individuals depending on age, heart volume, breathing pattern, and global ventricular systolic function. In addition, there is complicated blood flow within the four chambers of the heart and in the coronary arteries. Thus, the technical challenges for imaging are to acquire high resolution images in a short time that is flow sensitive, yet motion insensitive.[7,70]

Advances in hardware (gradients, receiver coils) and software (data acquisition and image processing) now allow for imaging methods that can visualize the proximal portion of the major epicardial coronary arteries in the vast majority of individuals.[70,71] Although not yet standardized, current methods for coronary MRA combine fast imaging sequences with techniques for suppressing cardiac and respiratory motion. The basic technique is gradient echo, therefore the basic contrast mechanism is in-flow enhancement. Data acquisition can be divided into 2D and 3D techniques.[72–76] The advantages of each technique are summarized in previous sections. For removing the motion problem, almost all the techniques rely on cardiac gating. In addition, breath holding,[77–80] navigator pulses,[81–83] and complicated k-space trajectory design[84,85] all are used in one way or another to shorten imaging time or remove motion interference. Figure 10 is an example of using 2D breath-hold spiral technique to image the human coronary artery.

Figure 10 An image of the right main coronary artery (arrow) of a normal volunteer visualized using a 2D spiral technique. (Courtesy of Craig Meyer of Stanford University.) This image was acquired with breath-hold in 14 s using phased-array surface coils.

6 COMPARISON WITH OTHER ANGIOGRAPHY TECHNIQUES

MRA has become an extremely useful clinical tool. The inherent advantages of MRA include noninvasive, 3D, high resolution, flow sensitive, and no absolute need for contrast material. Before MR and ultrasonic imaging methods became available, invasive X-ray angiography was the method of choice for the visualization of all vascular anatomy. X-ray methods, however, have several undesirable properties including patient discomfort, morbidity, and cost. Ultrasonic methods are relatively safe and low cost, but are limited to a few clinical applications and are heavily dependent on the skill of the operator. A comparison of the effectiveness of techniques used for angiography is summarized in Table 3.[86]

Like any other radiological procedure, the proper role for a non-invasive MRA study is under considerable scrutiny in an increasingly cost-conscious health care system. Clinical experience to date has shown MRA to be effective in screening and follow-up of marginal pathological conditions, and in monitoring results of therapeutic intervention. MRA has a wide variety of options and techniques for the study of flowing blood: different methods for high spatial resolution, high temporal resolution, velocity sensitization, flow

Table 3 Effectiveness of techniques used for angiography

	MRA	Angiography	CT	Ultrasonagraphy
Aorta	++++	+++	+++	+++ (TEE)
Arch vessels	++++	+++	+++	+ (TEE)
Peripheral	++++	++++	+++	+++
Carotids	++++	+++	++	+++
Contrast	Not always needed for larger arteries[a]	Mandatory	Mandatory	Optional
Ionizing	No	Yes	Yes	No
Arterial	No	Yes	No	No (except for IVUS)

[a] Usually blood flow generates the needed contrast. If further contrast is desired, a gadolinium based agent is used intravenously. These agents contain no known allergens and have little if any side effects. CT, computed tomography; IVUS, intervascular ultrasonagraphy; TEE, transesophageal echocardiography. (Reproduced by permission of Biederman RWW, University of Alabama-Birmingham.)

quantification, visualizing pulsatility, and 3D registration with anatomic data acquired via alternate MR imaging methods are all used in MRA imaging.

ACKNOWLEDGMENTS

The author would like to acknowledge the following people for their contributions to the article: Zach Miller, K.R. Maravilla, M.D., Shannon Swan, M.D., Charles A. Mistretta, Ph.D., Craig Meyer, Ph.D., and General Electric Medical Systems.

ABBREVIATIONS AND ACRONYMS

CAD	Coronary Artery Disease
MIP	Maximum Intensity Projection
MR	Magnetic Resonance
MRA	Magnetic Resonance Angiography
PC	Phase-contrast
RF	Radiofrequency
TE	Echo Time
TOF	Time-of-flight
TR	Repetition Time
VENC	Velocity Encoding
2D	Two-dimensional
3D	Three-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Imaging, Functional • Magnetic Resonance in Medicine, High Resolution Ex Vivo • Magnetic Resonance, General Medical • Multinuclear Magnetic Resonance Spectroscopic Imaging

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General

REFERENCES

1. J.R. Singer, 'Blood Flow Rates by Nuclear Magnetic Resonance', *Science*, **130**, 1652 (1959).
2. E.L. Hahn, 'Detection of Seawater Motion by Nuclear Precession', *J. Geophys. Res.*, **65**, 776 (1960).
3. W.S. Hinshaw, P.A. Bottomley, G.N. Holland, 'Radiographic Thin-section Image of the Human Wrist by Nuclear Magnetic Resonance', *Nature*, **270**(22/29), 722–723 (1977).
4. L. Axel, 'Blood Flow Effects in Magnetic Resonance Imaging', *Amer. J. Radiol.*, **143**, 1157–1166 (1984).
5. C.L. Schultz, R.J. Alfidi, A.D. Nelson, S.Y. Kopiwoda, M.E. Clampitt, 'The Effect of Motion on Two-dimensional Fourier Transformation Magnetic Resonance Images', *Radiology*, **152**, 117–121 (1984).
6. V.J. Wedeen, R.A. Meuli, R.R. Edelman et al., 'Projective Imaging of Pulsatile Flow with Magnetic Resonance', *Science*, **230**, 946 (1985).
7. R.R. Edelman, 'MR Angiography: Present and Future', *Amer. J. Radiol.*, **161**, 1 (1993).
8. C.L. Dumoulin, 'Phase Contrast Angiography Techniques', *MRI Clinics of North America*, **3**(3), 399 (1995).
9. G.A. Laub, 'Time-of-flight Method of MR Angiography', *MRI Clinics of North America*, **3**(3), 391 (1995).

10. D. Atkinson, L. Teresi, 'Magnetic Resonance Angiography', *Mag. Reson. Quarterly*, **10**, 149 (1994).
11. W.G. Bradley, V. Waluch, 'Blood Flow: Magnetic Resonance Imaging', *Radiology*, **154**, 443–450 (1995).
12. R.R. Edelman, H.P. Mattle, B. Wallner et al., 'Extracranial Carotid Arteries: Evaluation with ''Black blood'' MR Angiography', *Radiology*, **177**, 45–50 (1990).
13. D. Atkinson, M. Brant-Zawadzki, G. Laub, 'Optimization Strategies Enhance Time-of-flight MRA', *Radiology*, **190**, 890–894 (1994).
14. F.W. Wehrli, A. Shimakawa, G.T. Gullberg et al., 'Time-of-flight MR Flow Imaging: Selective Saturation Recovery with Gradient Refocussing', *Radiology*, **160**, 781–785 (1986).
15. D.L. Parker, C. Yuan, D.D. Blatter, 'MR Angiography by Multiple Thin Slab 3D Acquisition', *Magn. Reson. Med.*, **17**, 434–451 (1991).
16. S.D. Wolff, J. Eng, R.S. Balaban, 'Magnetization Transfer Contrast: Method for Improving Contrast in Gradient Recalled Echo Images', *Radiology*, **179**, 133–137 (1991).
17. D.J. Atkinson, M. Brant-Zawadzki, G. Gillan, D. Purdy, G. Laub, 'Improved MR Angiography: Magnetization Transfer Suppression with Variable Flip Angle Excitation and Increase Resolution', *Radiology*, **190**, 890–894 (1994).
18. D.G. Nishimura, A. Macovski, J.L. Jackson, R.S. Hu, C.A. Stevick, L. Axel, 'Magnetic Resonance Angiography by Selective Inversion Recovery Using a Compact Gradient Echo Sequence', *Magn. Reson. Med.*, **8**, 96–103 (1988).
19. P. Schmalbrock, C. Yuan, D.W. Chakeres, J. Kohli, N.C. Pelc, 'Volume MR Angiography: Methods to Achieve Very Short Echo Times', *Radiology*, **175**, 861–865 (1990).
20. D. Chien, D. Atkinson, R. Edelman, 'Strategies to Improve Contrast in TurboFLASH Imaging: Reordered Phase Encoding and k-space Segmentation', *J. Magn. Reson. Imaging*, **1**, 63–70 (1991).
21. D.B. Richardson, A.E.H. Bampton, S.J. Riederer, J.R. MacFall, 'Magnetization-prepared MR Angiography with Fat Suppression and Venous Saturation', *J. Magn. Reson. Imaging*, **2**, 653–664 (1992).
22. D. Li, E.M. Haacke, J.P. Mugler, S. Berr, J.R. Brookeman, M.C. Hutton, 'Three-dimensional Time-of-flight MR Angiography Using Selective Inversion Recovery RAGE with Fat Saturation and ECG-triggering: Application to Renal Arteries', *Magn. Reson. Med.*, **31**, 414–422 (1994).
23. P.M. Pattany, L.C. Chiu, J.D. Lipcamon et al., 'Motion Artifact Suppression (MAST) for Magnetic Resonance Imaging', *J. Comput. Assist. Tomogr.*, **11**, 369–377 (1987).
24. Q.S. Xiang, O. Nalcioglu, 'A Formalism for Generating Multiparametric Encoding Gradients in NMR Tomography', *IEEE Trans. Med. Imag.*, **MI-6**, 14–20 (1987).
25. J.L. Durek, O.P. Simonetti, G.C. Hurst, 'Modified Gradients for Motion Suppression: Variable Echo Time and Variable Bandwidth', *Magn. Reson. Imaging*, **8**, 141–151 (1990).
26. P.R. Moran, 'A Flow Velocity Zeugmatographic Interlace for NMR Imaging in Humans', *Magn. Reson. Imaging*, **1**, 197 (1982).
27. M. O'Donnell, 'NMR Blood Flow Imaging Using Multiecho, Phase Contrast Sequences', *Med. Phys.*, **12**, 59–64 (1985).
28. C.L. Dumoulin, S.P. Souza, R.D. Darrow et al., 'Simultaneous Acquisition of Phase-contrast Angiograms and Stationary-tissue Images with Hadamard Encoding of Flow-induced Phase Shifts', *J. Magn. Reson. Imaging*, **1**, 399–404 (1991).
29. P. Turski, M. Bernstein, O. Boyko et al., 'Vascular Magnetic Resonance Imaging: Signa Applications Guide Volume III', *GE Medical Systems*, 1990.
30. R. Hausmann, J.S. Lewin, G. Laub, 'Phase-contrast MR Angiography with Reduced Acquisition Time: New Concept in Sequence Design', *J. Magn. Reson. Imaging*, **1**, 415–422 (1991).
31. N.J. Pelc, M.A. Bernstein, A. Shimakawa et al., 'Encoding Strategies for Three-direction Phase-contrast MR Imaging of Flow', *J. Magn. Reson. Imaging*, **1**, 405–413 (1991).
32. D.J. Bryant, J.A. Payne, D.N. Firmin et al., 'Measurement of Flow with NMR Imaging Using Gradient Pulse and Phase Difference Technique', *J. Comput. Assist. Tomogr.*, **8**, 588–593 (1984).
33. P. Van Dijk, 'Direct Cardiac NMR Imaging of Heart Wall and Blood Flow Velocity', *J. Comput. Assist. Tomogr.*, **8**, 429–436 (1984).
34. I.R. Young, G.M. Bydder, J.A. Payne, 'Flow Measurement by the Development of Phase Differences During Slice Formation in MR Imaging', *Magn. Reson. Med.*, **2**, 55 (1985).
35. J. Hennig, M. Muri, P. Brunner et al., 'Quantitative Flow Measurement with the Fast Fourier Flow Technique', *Radiology*, **166**, 237 (1988).
36. P.J. Keller, B.P. Drayer, E.K. Fram, K.D. Williams, C.L. Dumouline, S.P. Souza, 'MR Angiography with Two-dimensional Acquisition and Three-dimensional Display: Work-in-progress', *Radiology*, **173**, 527–532 (1989).
37. H.P. Mattle, K.C. Kent, R.R. Edelman, D.J. Atkinson, J.J. Skillman, 'Evaluation of the Extracranial Carotid Arteries: Correlation of Magnetic Resonance Angiography, Duplex Ultrasonography, and Conventional Angiography', *J. Vasc. Surg.*, **13**, 838–845 (1991).
38. A.W. Litt, E.M. Eidelman, R.S. Pinto et al., 'Diagnosis of Carotid Artery Stenosis: Comparison of 2DFT Time-of-flight MR Angiography with Contrast Angiography in 50 Patients', *Amer. J. Radiol.*, **156**, 611–616 (1991).
39. P.M. Ruggieri, T.J. Masaryk, J.S. Ross, M.T. Modic, 'Magnetic Resonance Angiography of the Intracranial

Vasculature', *Top. Magn. Reson. Imaging*, **3**, 23–33 (1991).
40. R.J. Sevick, J.S. Tsuruda, P. Schmalbrock, 'Three-dimensional Time-of-flight MR Angiography in the Evaluation of Cerebral Aneurysms', *J. Comput. Assist. Tomogr.*, **14**, 874–881 (1990).
41. J.P. Finn, A. Goldmann, R.R. Edelman, 'Magnetic Resonance Angiography in the Body', *Magn. Reson. Quarterly*, **8**, 1–22 (1992).
42. W. Erdman, H. Jayson, H. Redman, G. Miller, R. Parkey, R. Peshock, 'Deep Venous Thrombosis of Extremities: Role of MR Imaging in the Diagnosis', *Radiology*, **174**, 425–431 (1990).
43. R.S. Owen, J.P. Carpenter, R.A. Baum, L.J. Perloff, C. Cope, 'Magnetic Resonance Imaging of Angiographically Occult Runoff Vessels in Peripheral Arterial Occlusive Disease', *N. Engl. J. Med.*, **326**, 1577–1581 (1992).
44. C.L. Dumoulin, S.P. Souza, M.F. Walker et al., 'Three-dimensional Phase Contrast Angiography', *Magn. Reson. Med.*, **9**, 139–149 (1989).
45. S.P. Souza, C.L. Dumoulin, 'Dynamic Magnetic Resonance Angiography', *Dynamic Cardiovascular Imaging*, **1**, 126–143 (1987).
46. M.C. Henry-Feugeas, I. Idy-Peretti, B. Blanchet et al., 'Temporal and Spatial Assessment of Normal Cerebrospinal Fluid Dynamics with MR Imaging', *Magn. Reson. Imaging*, **11**, 1107 (1993).
47. G.L. Nayler, D.N. Firmin, D.B. Longmore, 'Blood Flow Imaging by Cine Magnetic Resonance', *J. Comput. Assist. Tomogr.*, **10**, 715 (1986).
48. J.F. Debatin, R.H. Ting, H. Wegmuller et al., 'Renal Artery Blood Flow: Quantitation with Phase-contrast MR Imaging with and Without Breath Holding', *Radiology*, **190**, 371 (1994).
49. K.C.P. Li, W.S. Whitney, C.H. McDonnell et al., 'Chronic Mesenteric Ischemia: Evaluation with Phase-contrast Cine MR Imaging', *Radiology*, **190**, 175 (1994).
50. M.B.M. Hofman, M. Kouwenhoven, M. Sprenger et al., 'Nontriggered Magnetic Resonance Velocity Measurement of the Time Average of Pulsatile Velocity', *Magn. Reson. Med.*, **29**, 648 (1993).
51. G.A. Laub, W.A. Kaiser, 'MR Angiography with Gradient Motion Refocusing', *J. Comput. Assist. Tomogr.*, **12**, 377–382 (1988).
52. C.M. Anderson, D. Saloner, J.S. Tsuruda, L.G. Shapeero, R.E. Lee, 'Artifacts in Maximum-intensity-projection Display of MR Angiographic Data', *Amer. J. Radiol.*, **154**, 623–629 (1990).
53. D. Saloner, W.A. Hanson, J.S. Tsuruda, R. vanTren, C.M. Anderson, R.E. Lee, 'Application of a Connected-voxel Algorithm to MR Angiographic Data', *J. Magn. Reson. Imaging*, **1**, 423–430 (1991).
54. H.E. Cline, C.L. Dumoulin, W.E. Lorensen et al., 'Volume Rendering and Connectivity Algorithms for MR Angiography', *Magn. Reson. Med.*, **18**, 384 (1991).
55. X. Hu, N. Alperin, D.N. Levin, K.K. Tan, M. Mengeot, 'Visualization of MR Angiographic Data with Segmentation and Volume-rendering Techniques', *J. Magn. Reson. Imaging*, **1**, 539–546 (1991).
56. M.R. Prince, 'Gadolinium-enhanced MR Aortography', *Radiology*, **191**, 155–164 (1994).
57. M.R. Prince, 'Body MR Angiography with Gadolinium Contrast Agents', *MRI Clin. North Am.*, **4**(1), 11–24 (1996).
58. H.J. Weinmann, R.C. Brasch, W.R. Press et al., 'Characteristics of Gadolinium-DTPA Complex: A Potential NMR Contrast Agent', *Amer. J. Radiol.*, **142**, 619–624 (1984).
59. F.R. Korosec, R. Frayne, T.M. Grist, C.A. Mistretta, 'Time-resolved Contrast-enhanced 3D MR Angiography', *Magn. Reson. Med.*, **36**, 345–351 (1996).
60. C.A. Mistretta, T.M. Grist, R. Frayne, F.R. Korosec, J.A. Polzin, 'Simulation of a Breath-hold Method for Time-resolved 3D Contrast Imaging', in Proc., ISMRM, 4th Annual Meeting, New York, 1996.
61. R. Frayne, T.M. Grist, F.R. Korosec, D.S. Willig, J.S. Swan, P.A. Turski, C.A. Mistretta, 'MR Angiography with Three-dimensional MR Digital Subtraction Angiography', *Top. Magn. Reson. Imaging*, **6**, 366–388 (1997).
62. C.A. Mistretta, T.M. Grist, F.R. Korosec, R. Frayne, D.C. Peters, Y. Mazaheri, T.J. Carrol, '3D Time-resolved Contrast Enhanced MR SDA: Advantages and Trade-offs', *Magn. Reson. Med.*, **40**, 571–581 (1998).
63. M.R. Prince, D.L. Narasimham, J.C. Stanley et al., 'Breath-hold Gadolinium-enhanced MR Angiography of the Abdominal Aorta and its Major Branches', *Radiology*, **197**, 785–792 (1995).
64. J.J. Snidow, M.S. Johnson, V.J. Harris et al., 'Three-dimensional Gadolinum-enhanced MR Angiography for Aortoiliac Inflow Assessment Plus Renal Artery Screening in a Single Breath Hold', *Radiology*, **198**, 725–732 (1996).
65. P. Steiner, J.F. Debatin, B. Romanowski et al., 'Optimization of Breath-hold 3D MR Pulmonary Angiography', *Radiology*, **201**(p), 201 (1996).
66. T.F. Hany, J.F. Bebatin, M. Schmidt et al., 'Three-dimensional MR Angiography of the Renal Arteries in a Single Breath-hold', *Radiology*, **201**(p), 218 (1996).
67. M.R. Prince, T.M. Grist, J.F. Debatin, *3D Contrast MR Angiography*, Springer-Verlag, Berlin, 16–17, 1997.
68. G. Krinsky, N. Rofsky, M. Flyer et al., 'Gadolinium-enhanced Three-dimensional MR Angiography of Acquired Arch Vessel Disease', *Amer. J. Radiol.*, **167**, 981–987 (1996).
69. L.M. Lieberman, R.E. Botti, A.D. Nelson, 'Magnetic Resonance of the Heart', *Radiol. Clin. North. Am.*, **22**, 847–858 (1984).

70. P.G. Danias, R.R. Edelman, W.J. Manning, 'Coronary MR Angiography', *Cardiology Clinics*, **16**(2), 207–225 (1998).
71. D.J. Pennell, J. Keegan, D.N. Firmin et al., 'Magnetic Resonance Imaging of Coronary Arteries: Technique and Preliminary Results', *Br. Heart J.*, **70**, 315–326 (1993).
72. A.J. Duerinckx, M.K. Urman, 'Two-dimensional Coronary MR Angiography: Analysis of Initial Clinical Results', *Radiology*, **193**, 731–738 (1994).
73. Y. Wang, T.M. Grist, F.R. Korosec et al., 'Respiratory Blur in 3D Coronary MR Imaging', *Magn. Reson. Med.*, **33**, 541–548 (1995).
74. J.C. Post, A.C. van Rossum, M.B.M. Hofman et al., 'Clinical Utility of Two-dimensional Magnetic Resonance Angiography in Detecting Coronary Artery Disease', *Eur. Heart. J.*, **18**, 426–433 (1997).
75. J.C. Post, A.C. van Rossum, M.B.M. Hofman et al., 'Three-dimensional Respiratory-gated MR Angiography of Coronary Arteries: Comparison with Conventional Angiography', *Amer. J. Radiol.*, **166**, 1399–1404 (1996).
76. C.H. Meyer, B.S. Hu, D.G. Nishimura et al., 'Fast Spiral Coronary Artery Imaging', *Magn. Reson. Med.*, **28**, 202–213 (1992).
77. W.J. Manning, W. Li, N.G. Boyle et al., 'Fast-suppressed Breath-hold Magnetic Resonance Coronary Angiography', *Circulation*, **87**, 94–104 (1993).
78. Y.L. Liu, S.J. Reiderer, P.J. Rossman et al., 'A Monitoring, Feedback, and Triggering System for Reproducible Breath-hold MR Imaging', *Magn. Reson. Med.*, **30**, 507–511 (1993).
79. M. Doyle, M.B. Scheidegger, R.G. de Graaf et al., 'Coronary Artery Imaging in Multiple 1-sec Breath-holds', *Magn. Reson. Imaging*, **11**, 3–6 (1993).
80. R.R. Edelman, W.J. Manning, D. Burstein et al., 'Coronary Arteries: Breath-hold MR Angiography', *Radiology*, **181**, 641–643 (1991).
81. D. Li, S. Kaushikka, E.M. Jaacke et al., 'Coronary Arteries: Three Dimensional MR Imaging with Retrospective Respiratory Gating', *Radiology*, **201**, 857–863 (1996).
82. M.F. Muller, M. Fleisch, R. Kroecker et al., 'Proximal Coronary Artery Stenosis: Three-dimensional MRI with Fat Saturation and Navigator Echo', *J. Magn. Reson. Imaging*, **7**, 644–651 (1996).
83. Y. Wang, P.J. Rossman, R.C. Grimm et al., 'Navigator-echo-based Real-time Respiratory Gating and Triggering for Reduction of Respiration Effects in Three-dimensional MR Angiography', *Radiology*, **198**, 55–60 (1996).
84. B.S. Hu, C.H. Meyer, A. Macovski et al., 'Multi-spiral Magnetic Resonance Coronary Angiography', *Proc. Pro. Int. Soc. Magn. Reson. Med.*, **1**, 176 (1996).
85. C.H. Meyer, B.S. Hu, D.G. Nishimura et al., 'Fast Spiral Coronary Artery Imaging', *Magn. Reson. Med.*, **28**, 202–213 (1992).
86. R.W.W. Biederman, A.R. Fuisz, G.M. Pohost, 'Magnetic Resonance Angiography', *Current Opinion in Cardiology*, **13**, 430–437 (1998).

Magnetic Resonance Imaging, Functional

Wolfgang Richter
National Research Council, Winnipeg, Canada

1 Introduction 1
2 History 2
3 Physiological Basis of Functional Magnetic Resonance Imaging 2
4 Hardware Aspects 3
4.1 Magnets 3
4.2 Magnetic Field Gradients and Shims 4
4.3 Radiofrequency 4
4.4 Stimulation Devices 5
5 Pulse Sequences 6
5.1 Pulse Sequences to Acquire One k-Space Line per Excitation 6
5.2 Pulse Sequences to Acquire Several k-Space Lines per Excitation 6
5.3 Pulse Sequences to Acquire the Complete k-Space in One Excitation 6
6 Safety Issues 7
6.1 Safety Issues Integral to the Examination 7
6.2 Safety Issues External to the Magnetic Resonance Imaging Examination 8
7 Noise and Artefacts 8
7.1 Environmental Noise 8
7.2 Instrumental Noise 8
7.3 Sample Noise 8
8 Experimental Design 10
8.1 Block Designs 10
8.2 Single Trial Designs 10
9 Data Analysis 11
9.1 Paradigmatic Methods 11
9.2 Non-parametric Methods 12
10 Examples of Applications 13
10.1 Sensory Tasks 13
10.2 Movement Tasks 15
10.3 Cognitive Tasks 16
10.4 Other Experiments 17
11 Comparison with Other Methods and Outlook 18
Abbreviations and Acronyms 18
Related Articles 18
References 19

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Functional magnetic resonance imaging (FMRI) is an analytical method for measuring brain activity while it occurs. FMRI was first demonstrated in 1992, but it has since become the most popular neuroimaging method. Its temporal resolution is of the order of seconds and hence superior to positron emission tomography (PET). Its spatial resolution is on the order of millimeters which makes it superior to both PET and electrophysiological methods such as electroencephalography (EEG). Furthermore, FMRI is noninvasive in the sense that no external contrast agent has to be used. FMRI contrast is based on the intrinisic blood oxygenation changes that occur at the site of brain activity in response to a specific task. The exact mechanism that links activity and signal change is currently not well understood and is an area of active research.

FMRI is subject to many experimental difficulties, however. A vexing problem is that of physiological (heartbeat and breathing) and gross motion. Gross motion is often coupled to the presentation of the stimulus and hence especially prone to producing artefactual activation. The analysis of the experimental data is not a standard procedure at present. While past research has generally used paradigmatic methods of analysis (hypothesis testing), nonparadigmatic (data driven) methods like fuzzy clustering analysis (FCA) or independent component analysis (ICA) have become important tools.

A more complete understanding of the physiological mechanisms leading to the activation signal, and a better grasp of the proper statistical treatment of the data, are likely to increase the power of FMRI even further.

1 INTRODUCTION

FMRI is arguably the most powerful neuroimaging technique available. Its comet-like rise in the field of neuroscience may be likened to the revolution that two-dimensional nuclear magnetic resonance (NMR) spectroscopy brought about for the determination of molecular structure in the early 1980s. Indeed, it has been observed time and time again that the potential applications of nuclear magnetism seem boundless, and the development of FMRI serves as yet another manifestation of this phenomenon. FMRI has filled many white spots on the map of the human brain already,

and it continues to do so by virtue of three outstanding advantages: spatial resolution, temporal resolution, and relative noninvasiveness.

To be sure, there are many other methods of measuring brain function, each of which may be superior to FMRI in any of these aspects. For example, single neuronal recording experiments deliver a spatial resolution which is orders of magnitude better than that of FMRI. Electroencephalographic (EEG) and magnetoencephalographic (MEG) measurements are capable of a temporal resolution that far exceeds that of FMRI. And finally, many standard psychological experiments are certainly even less invasive than FMRI, but they measure brain activity only in a very superficial manner. FMRI is capable of *simultaneous* temporal and spatial resolution on a scale appropriate for many questions of interest. Its temporal resolution is on the order of hundreds of milliseconds to a few seconds, which is a typical timescale for performing a cognitive task. Its spatial resolution is on the order of hundreds of micrometers to a few millimeters, which is the size of typical, anatomically defined brain areas. And finally, FMRI is invasive only insofar as it uses magnetic and electromagnetic fields, which penetrate the body but have no known adverse effects. This is why FMRI is the method of choice for innumerable experiments on human brain function.

FMRI experiments are subject to numerous technical difficulties. Therefore the method has mainly been used by groups with dedicated FMRI systems. Now most problems have been solved and FMRI experiments are being performed routinely on many clinical MRI (magnetic resonance imaging) systems. Given the abundance and continuing proliferation of MRI systems, both in North America and in other parts of the world, the importance of FMRI is likely further to increase for years to come.

2 HISTORY

FMRI rests on a foundation of over 50 years of NMR spectroscopy. Reports of the first successful NMR experiments were published in 1946,[1,2] and NMR subsequently became the premier method for the determination of molecular structure. NMR imaging was first demonstrated in 1973 in a phantom (sample).[3] In 1976, the first image of the human body (a finger) was produced;[4] followed in 1977 by an image of the thorax,[5] and, in 1978, of a human head.[6] The first clinical magnet was installed in 1981 and NMR imaging, rechristened MRI, has become a standard method for diagnostic imaging of all parts of the human body. The author estimates that there are between 10 000 and 20 000 clinical MRI machines in existence in the world.

Contrast in MRI is mostly based on the differential relaxation times of different types of tissue. FMRI is in principle not different from (clinical) MRI, since it also measures a relaxation time parameter (the apparent transverse relaxation time, T_2^*). FMRI, however, generates a *time course* by measuring the evolution of that quantity over the course of a behavioral experiment. It was not until the late 1980s that the potential connection between this parameter and brain activity was recognized. Furthermore, technical difficulties abound in the FMRI experiment. The expected signal from an activated brain region is only approximately 3% larger than the baseline signal; hence the demands on instrumental stability are not easily met. The first functional MRI experiments were based on previous work by Ogawa and Lee[7] and reported simultaneously by three groups in 1992;[8–10] all of these experiments involved sensory stimulation tasks. The scope of experiments has far exceeded those early ones; today, many different types of brain activation are being studied, most importantly, *cognitive tasks*. In a different application, FMRI has moved into the operating room. Avoiding particularly important cortical structures, like speech areas, is paramount in neurosurgery; because some of these areas are anatomically ill-defined, FMRI is often the method of choice for the localization of such areas. Furthermore, intraoperative FMRI gives the opportunity to follow anatomical changes in the brain during surgery.

3 PHYSIOLOGICAL BASIS OF FUNCTIONAL MAGNETIC RESONANCE IMAGING

As opposed to electrophysiological methods, like EEG, MEG, or single neuronal recording, FMRI is a rather *indirect* method of detecting neuronal activity in the brain. While the electrical activity of neurons is, in principle, expected to alter the magnetic properties of the surrounding tissue, this effect seems to be too small to be measured by FMRI. Instead, the FMRI signal is generated through a set of secondary effects of neuronal activity, comprised of changes in neuronal metabolism, blood flow, blood volume, and possibly other physiological parameters.

The physical quantity that is measured directly by FMRI, and hence generates the image contrast, is the apparent transverse relaxation time (T_2^*) of water protons; the mechanism linking brain activation and T_2^* is called the BOLD (blood oxygen level dependent) mechanism. FMRI can also be performed using other contrast mechanisms; however, among methods presently available, T_2^* gives the highest contrast. T_2^* is a time constant that describes the decay of transverse magnetization caused

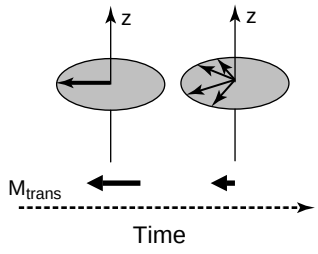

Figure 1 Mechanism of inhomogeneous broadening. Spins with different resonance frequencies acquire a phase difference; their vectorial sum decreases with time.

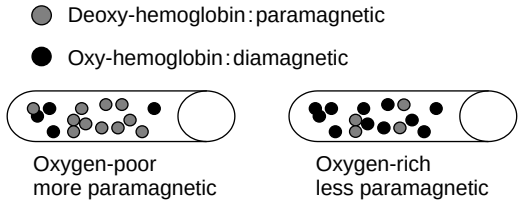

Figure 2 The origin of the BOLD effect. Increased blood oxygenation decreases the number of paramagnetic particles in the blood and hence increases homogeneity. Hence the T_2^* weighted signal increases.

by the dephasing of the individual spins (Figure 1). T_2^* contains contributions from the intrinsic transverse relaxation time and from susceptibility effects. The simplest model for the behavior of T_2^* upon brain activation takes into account only the magnetic properties of hemoglobin, which change depending on its oxygenation state. The hemoglobin molecule binds oxygen and transports it to the cells. In its oxygenated form, hemoglobin is diamagnetic, and in its deoxygenated form, it is paramagnetic (Figure 2). Hence deoxygenated hemoglobin forms a paramagnetic center, causing local susceptibility variations and a decrease in T_2^*. The magnitude of this effect is closely linked to the nature of the vasculature within the imaging voxel; it can be predicted that (1) the signal change from *tissue* is a quadratic function of magnetic field strength, (2) the signal change from *large vessels* is a linear function of the field strength, and (3) large vessels give rise to large signal changes, up to 20%.[11–13]

The complete physiological mechanism that links neuronal activity and T_2^* is complex and a matter of intense discussion. The following events are currently thought to take place:[14]

1. Immediately (tens to hundreds of milliseconds) after the onset of neuronal activation,
   (a) the oxygen metabolism of activated neurons increases,
   (b) vasodilation occurs,
   (c) local blood flow increases, and
   (d) a variety of other, nonvascular changes, such as changes in neural size, occur.

   Overall, these effects tend to counteract each other with respect to the BOLD signal, and what is commonly observed is either an initial small negative dip of the BOLD signal, or no measurable change.
2. Approximately 2 s after the onset of the activation, blood flow increases in excess of metabolic demand for oxygen, so that the deoxyhemoglobin concentration decreases, and the T_2^* weighted FMRI signal increases. This is the originally observed BOLD effect.
3. Finally, after neuronal activation ceases, blood flow decreases while oxygen metabolism is still elevated, and/or vasodilation persists after oxygen metabolism has reached the rest state. Both effects lead to the commonly observed negative BOLD undershoot before the signal returns to baseline permanently.

The most prominent predictions here are the initial "dip" of the BOLD signal and the final undershoot. The magnitude of those effects is expected to be small (on the order of 0.5% of the baseline signal) compared to the BOLD signal (approximately 3% of the baseline intensity). Both effects have indeed been observed, again in a visual stimulation paradigm.[15,16]

4 HARDWARE ASPECTS

4.1 Magnets

Since the days of the first NMR experiments, magnet technology has evolved tremendously. While the first commercial NMR magnets were room-temperature electromagnets, most magnets used today employ superconducting alloys, thereby eliminating problems of resistive heating. On the other hand, superconducting magnets have to use liquid helium as a coolant, which causes considerable operating expense. Furthermore, superconductivity breaks down at a magnetic flux density of more than 20 Tesla (20 T corresponds to a proton NMR frequency of 850 MHz). This is why the strongest magnet used for NMR (like the 33 T/32-mm bore magnet at the National High Magnetic Field Laboratory) is a room-temperature electromagnet; however, the power consumption of that magnet is approximately 19 MW (see National High Magnetic Field Laboratory and Florida State University, Tallahassee, 1999, www.nhmfl.gov).

The quest for higher magnetic fields in NMR spectroscopy is motivated by increased signal-to-noise-ratio (SNR) and spectral dispersion. SNR increases theoretically approximately with the magnetic field, $B_0^{5/2}$.

Linewidths and scalar couplings are largely independent of field strength (as they depend only on the relaxation times and coupling strength, respectively), but the chemical shift is a linear function of field strength. Hence the separation of lines increases at higher fields, which makes possible, for example, the study of macromolecules.

In FMRI, on the other hand, the advantages of higher field are controversial. Spectral dispersion is not an issue here because FMRI only looks at a single resonance line (that of the water protons). SNR increases, in principle, with magnetic field strength, because the physical mechanism of signal generation is the same as in NMR spectroscopy. However, technical issues prevent us from taking full advantage of the increased SNR. Most prominent here are the decreased relaxation times in tissue. At 1.5 T, T_2^* is on the order of 50 ms; at 4 T, it is approximately 30 ms. This means that the signal decays much faster, increasing demands on gradients and data acquisition hardware commensurately. However, the quantity of interest to functional MRI is the (BOLD) contrast-to-noise-ratio (CNR); it has been shown that CNR increases up to a field strength of at least 4 T.[17]

Most MRI systems in clinical use operate at a field strength of 1.5 T, although there are many 3-T imaging systems in research use (Figure 3). There are also a few 4-T systems in existence, and two ultra-high-field human systems: a 7-T system (at the University of Minnesota) (see Figure 4), and an 8-T system (at Ohio State University). The single most important feature that distinguishes magnets for human use from those for high-resolution NMR spectroscopy is physical size.

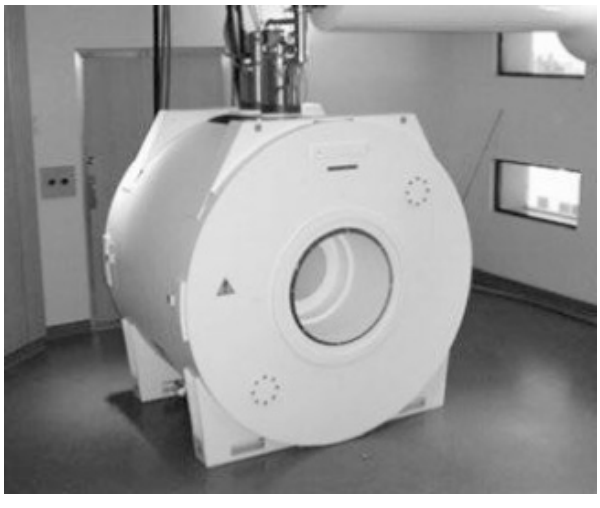

Figure 3 A 3-T head-only magnet in London, Canada. Note the small size of the magnet (length, 120 cm) which makes it relatively easy to site. The bore is wider at the end shown in order to accommodate the shoulders. The complete set-up also includes a patient table. (Reproduced by courtesy of John Saunders and Andrew Procca, Winnipeg, Canada.)

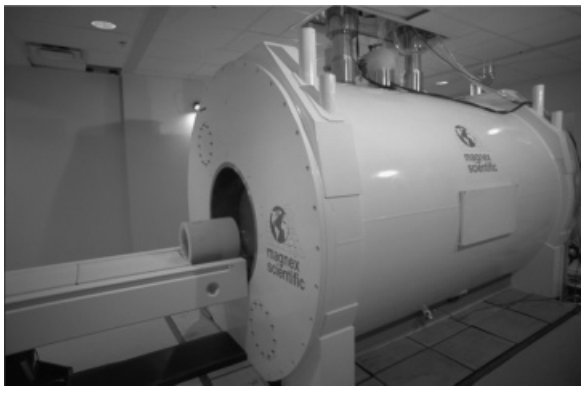

Figure 4 A 7-T human whole body MRI system in Minneapolis, Minnesota. (Reproduced by courtesy of Kamil Ugurbil and John Strupp, Minneapolis, Minnesota.)

The bore diameter of a high-resolution magnet is typically on the order of 5 cm, which accommodates a probe for a 5-mm NMR tube. A whole body human system typically requires a bore size of at least 60 cm in order to accommodate most people. It is being recognized, however, that functional MRI concentrates almost exclusively on the head (with a few notable exceptions); hence 3-T head-only systems are becoming quite popular now. Such systems require a bore diameter of less than 40 cm or so to accommodate the head and coil and are therefore easier to build and cheaper. These magnets may be very short and are therefore easy to site (see Figure 3).

4.2 Magnetic Field Gradients and Shims

In FMRI, spatial location is encoded by magnetic field strength and hence resonance frequency. Therefore, linear field gradients are applied along the three Cartesian axes. These field gradients are produced by specialized coils that are inserted into the magnet bore. A typical specification for gradients capable of echo planar imaging (EPI) is a rise time of 150 μs to a maximum amplitude of 30 mT m^{-1} (3 G cm^{-1}). The gradient rise time determines the imaging speed and the gradient strength determines the spatial resolution of the image. *Shims* are another set of specialized gradient coils. These are used to maximize the homogeneity of the magnetic field prior to any imaging experiment, with the sample or subject in the magnet.

4.3 Radiofrequency

The resonance frequency of protons at field strengths used for FMRI ranges from 64 MHz (at 1.5 T) to 168 MHz (at 4 T). The sample is excited at that frequency and the signal is received at that frequency as well. The

power used is on the order on 1 to several kilowatts. Radiofrequency (RF) is transmitted and received through RF coils specific to the body part to be imaged. In the case of FMRI, both surface coils and volume coils may be used. Surface coils exhibit higher sensitivity, but their active region may not penetrate the brain very deeply. If the brain region to be imaged is close to the surface of the head, for example the primary visual cortex, a surface coil may be the coil of choice. On the other hand, if the structure of interest is deep within the brain, for example the cingulate cortex, a volume coil will have to be used. Volume coils are all-purpose coils and may be the only coils available at many sites. Furthermore, with both improvements in imaging speed and the advancement of FMRI into the realm of psychology, experiments studying the whole brain simultaneously are becoming more common; in such experiments, volume coils (or arrays of local coils) are the only option. A surface and a volume coil are shown in Figures 5 and 6, respectively.

Figure 5 Dual-ring surface coil. This coil was constructed for a study of the thyroid gland. The coil is shown opened up for demonstration; when closed, it has the shape of a single cylinder. (Reproduced by courtesy of Scott King and Ian Smith, Winnipeg, Canada.)

4.4 Stimulation Devices

Most meaningful functional imaging experiments require some kind of interaction between the subject and the environment beyond the imaging system. Presenting an auditory or visual signal in an environment of high-magnetic field comprises its own set of technical challenges. To mention only the most obvious problem, many standard experiments in psychology are carried out with the aid of computers, but a computer can certainly not be brought into the magnet room. Two-way *auditory* communication between experimenter and subject is necessary for safety reasons and for information about the progress of the experiment. This is often achieved through a speaker/microphone set in or near the bore of the magnet. *Visual stimulation* is usually implemented by one of two methods. The simpler, but less versatile method is based on the optical projection of the image from a computer situated in the operator room onto a back-projection screen in the magnet room. That screen may be in front of the magnet or in the bore. It is viewed by the subject through a system of mirrors. In a more involved method, the computer-generated image is transformed into an optical image and transmitted through a set of fiber-optic cables into liquid crystal display (LCD) goggles worn by the subject. This system allows for stereo projection and a large visual field. *Auditory stimulation* is achieved by sending the audio signal into a pair of specialized headphones worn by the subject. Gradient noise (which may reach 110 or 120 dB) is an important issue here; often, a combination of headphones and intubated ear plugs is used. Both visual

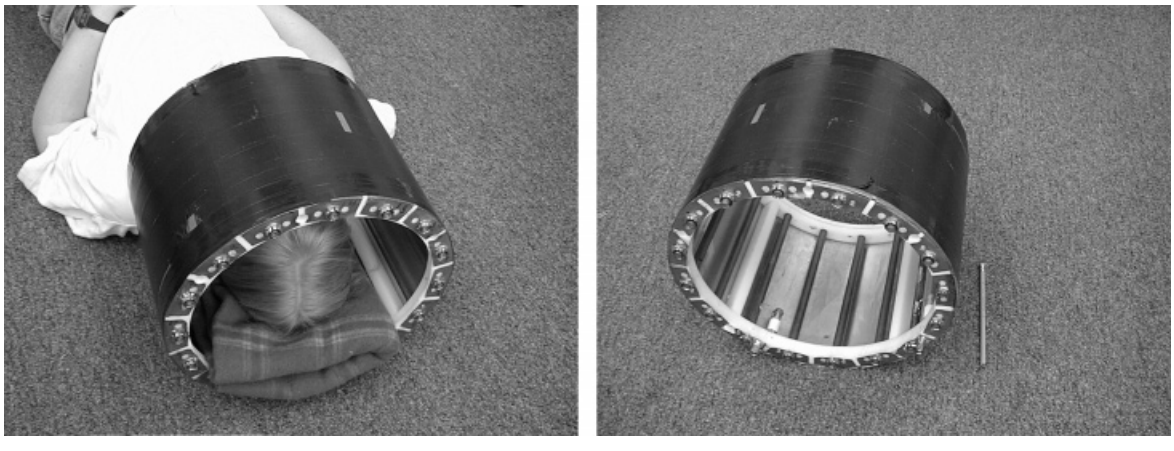

Figure 6 Volume head coil. The bird cage design is the most common design for such coils. (Reproduced by courtesy of Jarod Matwiy and Ian Smith, Winnipeg, Canada.)

and audio signal have to be routed through an RF filter panel into the shielded magnet room.

In addition to sending signals into the magnet room, many experiments require the recording of, or even interaction of stimulation software with, behavioral signals. For example, in one experiment a finger tapping task was investigated at high-temporal resolution.[18] In order to characterize the signal properly, the exact timing of the finger movement had to be measured. Therefore, subjects were given a button pad with one button for each finger. Every finger movement generated a digital signal which was sent outside the magnet room and recorded by a computer. Similarly, physiological data (breathing and heartbeat) are often measured, recorded, or used to trigger gated acquisitions.

5 PULSE SEQUENCES

MRI experiments can be conveniently described in an inverse space, commonly called ''k-space''. This space is analogous to the reciprocal space used in X-ray diffraction. Both in NMR spectroscopy and in MRI, the signal is acquired as a function of time and then decomposed into its constituent frequencies by Fourier transformation. In MRI, space is encoded by resonance frequency through the use of magnetic field gradients; hence Fourier transformation of the detected signal yields an image. k-Space is the space in which the signal is sampled; it has a Fourier transform relationship to physical space. The dimensionality and size of k-space is necessarily the same as that of real space; hence, in order to create a two-dimensional image with 64×64 pixels (for example, a slice through the head), we sample k-space in two dimensions with 64×64 time points.

Pulse sequences for MRI may be distinguished by the fraction of k-space that is sampled in a single excitation of the spin system. This, of course, largely determines the speed with which an image may be acquired. Not surprisingly, however, there is a tradeoff between speed (temporal resolution), spatial resolution, and SNR; hence every application may require a different pulse sequence.

5.1 Pulse Sequences to Acquire One k-Space Line per Excitation

A very common implementation of this concept is the FLASH (fast low-angle shot) sequence.[19] After a slice selective excitation pulse, a single phase-encode gradient is applied; this gradient determines the position of the acquired signal in the phase encode dimension of k-space. This gradient is followed by the readout gradient, which first dephases the signal, then rephases it, forming an echo. The echo then corresponds to a line in k-space. The pulse sequence is, therefore, repeated N times, where N is the desired number of pixels in the phase-encode dimension. Note that this pulse sequence will produce T_1 weighting if the magnetization is not fully relaxed before each excitation pulse; hence, for purse transverse weighting, a long repetition time has to be employed. Repetition times for FLASH sequences for *anatomic* imaging are typically on the order of a few milliseconds; for functional imaging, however, T_2^* weighting is required. The optimum contrast here is achieved when the echo time is on the order of T_2^* (tens of milliseconds), as shown below; the time needed to acquire 64 lines of k-space for a low-resolution image is on the order of several seconds! Hence FMRI using FLASH is usually not desirable, especially if we consider that there is almost always more than one slice of interest. However, FMRI can be and has been successfully performed by FLASH, as shown in section 10 below.

Similarly, spin-echo pulse sequences may be used for anatomic imaging. These pulse sequences exhibit T_2/T_1 weighting, as inhomogeneous broadening is refocused by the echo.

5.2 Pulse Sequences to Acquire Several k-Space Lines per Excitation

Both gradient echo and spin-echo pulse sequences may be modified by acquiring more than one echo. An example is the fast spin echo, or rapid acquisition with relaxation enhancement (RARE) method;[20] here the excitation is followed by a number of echoes, separated by an incremented gradient in the phase encode direction.

5.3 Pulse Sequences to Acquire the Complete k-Space in One Excitation

The fastest pulse sequences cover all of k-space in a single excitation. EPI sequences[21,22] traverse k-space in a rectangular pattern (see Figure 7). Spiral pulse sequences[23] traverse k-space in a spiral, simultaneously varying both transverse gradients. An example of a gradient-recalled echo (GRE) EPI pulse sequence is shown in Figure 8. The initial excitation pulse selects the slice of interest; the subsequent echo train then acquires the signal from that slice. The readout gradient oscillates continuously, forming an echo train, while the ''blipped'' phase encode gradient increments the phase between any two echoes.

Typical instrumental parameters for a high-performance system are as follows: maximum gradient strength is $30\,\mathrm{mT\,m^{-1}}$; rise time to maximum is $200\,\mu\mathrm{s}$; receiver bandwidth is 200 kHz. Note that, in EPI pulse sequences with linear sampling, the echo time is defined as the time from the first excitation pulse to the center of the echo train (32nd echo for 64×64 matrix size; 64th echo for 128×128 matrix size). As will be shown below, the optimal echo time for FMRI is on the order of T_2^*,

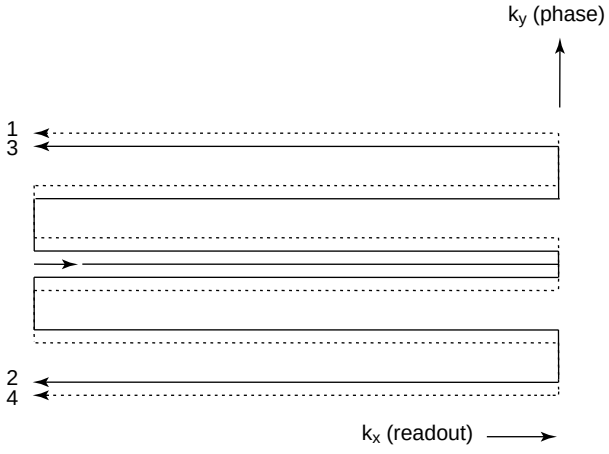

Figure 7 k-Space sampling trajectory for an EPI sequence. In this case, the sampling starts at the center and fills the phase-encode dimension in an alternating fashion (center-out sampling). Equally common is ''linear sampling'', where the phase encode dimension is filled linearly from bottom to top. Different sampling schemes are prone to various different artifacts. (Reproduced by courtesy of Seong-Gi Kim, Minneapolis, USA.)

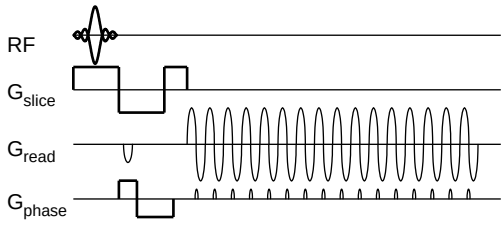

Figure 8 An EPI pulse sequence diagram. After slice selection, the read gradient oscillates to form the required number of echoes, while the (blipped) phase encode gradient is incremented between any two echoes. (Reproduced by courtesy of Seong-Gi Kim, Minneapolis, USA.)

which, in turn, ranges from about 25 ms (at 4 T) to about 50 ms (at 1.5 T). Hence the execution time of this pulse sequence (the time to acquire one slice) should be on the order of 100 ms at 1.5 T and 50 ms at 4 T. With the above parameters, this may be achieved at 4 T with a resolution of 64×64 pixels; if higher spatial resolution is required, the pulse sequence must be segmented (in segmented pulse sequences, k-space is divided into several segments, each of which is sampled in one acquisition). With a high-performance 1.5-T system, a resolution of 128×128 can usually be achieved in a single segment.

6 SAFETY ISSUES

Since the inception of MRI, the safety of the method has been intensely debated. Virtually all scientists agree that MRI is a very safe method if standard precautions are followed. Statistically, it is orders of magnitude more probable that a subject is hurt on the way to the examination, than during the MRI examination itself. It must be conceded, however, that nothing can be said about long-term side effects, because significant numbers of MRI examinations have been carried out only since about the mid-1980s.

There are several types of issue concerning safety. These may be categorized into safety issues that are integral to a standard examination and hence unavoidable and safety issues that are external to the examination and stem from improper procedures.

6.1 Safety Issues Integral to the Examination

There are three components to an MRI scan that may be the source of safety problems: the static magnetic field, electromagnetic radiation, and time-varying magnetic field gradients. Limiting values for these quantities are set by the various national regulatory bodies; in the USA, this is the Food and Drug Administration (FDA). FDA guidelines are presently (2000) as follows.[24]

6.1.1 Static Magnetic Field

For routine clinical use, the static magnetic field has long been limited to 2.0 T. However, a 3.0-T head-only scanner has been approved for this purpose. For research use, approval is given on a case-by-case basis; at present, the highest field strength in use for human experiments is 8.0 T. While no study has conclusively found adverse side effects of high-magnetic fields in humans, there has been anecdotal evidence of fatigue, dizziness, metallic taste and other nonspecific symptoms in the presence of very high-magnetic fields. Evidence to the contrary, however, was obtained when subjects were asked to fill out a standard exit questionnaire after they were subjects of a mock MRI study in a nonfunctional magnet.[25] The answers in the questionnaire were not significantly different from those obtained from actual experiments. However, as mentioned above, there is no well-controlled study that shows if the perceived symptoms are related to the magnetic field or not.

6.1.2 Electromagnetic Radiation

In order to excite the spin system, electromagnetic radiation is applied to the tissue in the active region of the coil. A large fraction of the energy associated with this radiation is absorbed by the tissue and may cause local heating. The permissible amount of energy is defined in terms of the specific absorption rate (SAR). There are presently two limits, either of which must be met:

1. The SAR must be less than $0.4\,\mathrm{W\,kg^{-1}}$ averaged over the body and less than $8.0\,\mathrm{W\,kg^{-1}}$ in any 1 g of tissue and less than $3.2\,\mathrm{W\,kg^{-1}}$ averaged over the head, or
2. The exposure to electromagnetic radiation must be such that it produces a core temperature increase of no more than 1 K and localized heating to no more than 38 °C in the head, 39 °C in the trunk, and 40 °C in the extremities. In some patients, however, these values may be exceeded.

The actual values are difficult to determine theoretically or experimentally. A common problem occurs with surface coils, where the spatial distribution of deposited power is not well-known. Consequentially, experimentalists often allow a considerable margin of error for these experiments.

6.1.3 Magnetic Field Gradients

These fast-switching gradients are probably the source of greatest concern. In fact, these gradients are largely responsible for the loud noise associated with MRI experiments. Possible biological effects include the illusion of flashing lights and stimulation of peripheral nerves. One of the following conditions must be met:

1. The system is not capable of producing gradient switching rates $\mathrm{d}B/\mathrm{d}t$ of more than $6\,\mathrm{T\,s^{-1}}$.
2. For longitudinal gradients, $\mathrm{d}B/\mathrm{d}t < 20\,\mathrm{T\,s^{-1}}$ if $t > 120\,\mu\mathrm{s}$, or

$$\mathrm{d}B/\mathrm{d}t < 2.4\,\mathrm{mT\,s^{-1}} \quad \text{if} \quad 12\,\mu\mathrm{s} < t < 120\,\mu\mathrm{s}, \quad \text{or}$$
$$\mathrm{d}B/\mathrm{d}t < 200\,\mathrm{T\,s^{-1}} \quad \text{for} \quad T < 12\,\mu\mathrm{s}$$

   For transverse gradient, limits are three times lower.
3. With an adequate margin of safety, the pulse sequence in question does not cause peripheral nerve stimulation.

6.2 Safety Issues External to the Magnetic Resonance Imaging Examination

In spite of the most careful precautions, ferromagnetic objects pose the greatest danger. Such objects can accidentally be brought near the magnet; particularly common are gas cylinders and metal buckets. Near the magnet, these objects can become projectiles and seriously hurt people and damage equipment.

Metal objects inside the subject's body can be dangerous as well. Dental fillings are usually not of concern here, especially when produced in North America or western Europe. However, other metal objects in the body (e.g. biomedical implants), or on the surface (e.g. earrings, or, in some cases, tatoos) may lead to inadmissibility. The danger represented by such objects is threefold. The objects may be dislodged if ferromagnetic and cause damage, the function of biomedical implants may be compromised and, for some geometries, currents may be induced in the object and cause heating. Hence subjects have to be screened carefully before each MRI examination, and a risk assessment has to be performed on a case-by-case basis. In the case of normal volunteers, the exclusion criteria will generally be much more stringent than in the case of a medically beneficial or necessary examination.

7 NOISE AND ARTEFACTS

As mentioned above, the BOLD effect is small (on the order of a few percent) and the signal is confounded by noise and other artefacts. (Note: in this context, an artefact is any unwanted signal.) We may conveniently distinguish between environmental noise, instrumental noise, and sample noise.

7.1 Environmental Noise

Environmental noise consists largely of electromagnetic radiation, either from ubiquitous sources, or from electronic devices in the vicinity of the MRI system; this is particularly common in a hospital environment. Environmental noise can be minimized through the use of a shielded room; signals intended to penetrate the RF shield, such as the MRI signal itself, are routed through filter panels.

7.2 Instrumental Noise

The bulk of the residual RF noise originates in the electronic components of the system itself. While this is in principle unavoidable, the design of ultra-low-noise electronic components, namely preamplifiers, is an area of active research.

7.3 Sample Noise

7.3.1 Physiological Motion

In the case of FMRI, the sample is a living system. Here the sample is arguably by far the most important source of artifacts. One extremely important source of sample noise is physiological in nature. Most prominent are the cardiac and respiratory cycles and swallowing.

Cardiac motion leads to a periodic change in blood volume, thereby affecting the BOLD signal. The frequency of this motion, however, is high (on the order of 1 Hz). In the majority of FMRI experiments, the acquisition frequency is less than 1 Hz; for example, the repetition time of a whole-brain experiment is typically on the order

of 3 s. Hence cardiac motion will usually be far undersampled and therefore not give rise to a systematic error. However, when the acquisition frequency is sufficiently high, cardiac motion may lead to an oscillation in the FMRI time courses. Even then, though, cardiac motion is not subject to voluntary control and hence is usually not coupled to the experimental paradigm. Therefore it will, again, not lead to a systematic error.

Respiratory motion affects the BOLD signal largely through the change in susceptibility in the brain which is caused by the changing geometry of the body. This motion occurs at a frequency on the order of 0.3 Hz, which is often nearly commensurate with the acquisition frequency. This may give rise to systematic artefacts. The issue of respiratory motion is further complicated by possible coupling between experimental paradigm and respiration, since respiration may indeed be voluntarily controlled. The same is true for *swallowing*, which, again, causes magnetic field variations in the brain. However, in the case of periodic physiological motion, retrospective signal correction may be performed if the contaminating signal is independently recorded during the experiment.[26]

7.3.2 Gross Motion

Possibly the most vexing source of sample noise is gross motion of the head. In areas of uniform intensity, motion will naturally have little effect on the signal; however, in the vicinity of large susceptibility gradients, like at the interface between sulci and gyri, near ventricles, and at the edge of the brain, gross motion will cause large signal fluctuations. Importantly, gross motion may be tightly coupled to the experimental paradigm, particularly in tasks involving overt movement of extremities or visual stimulation if the best viewing angle for the display is not in the most comfortable position for the subject. Hence gross motion may mask activation that is actually present or, which may be worse, mimic activation where there is in fact none. A single-slice or multislice experiment (as opposed to a true three-dimensional experiment) imposes a unique spatial direction on the sample; hence we have to distinguish between in-plane motion and out-of-plane motion, and also rotational and translational motion. In order to avoid erroneous measurements caused by gross motion, we have to consider four approaches, which may be used simultaneously. These are, in order of relative promise: (1) active motion correction during acquisition, (2) minimization of motion amplitude, (3) postacquisition motion correction, and (4) recognition of residual motion.

Active motion correction entails the measurement of motion and subsequent changing of gradient amplitudes in real time, such that the head remains immobile in the gradients' frame of reference. This method is still in its infancy, but commercial devices for this purpose are being developed.[27] The difficulty here is mainly the actual measurement of the motion with sufficient accuracy; this may be accomplished using laser beams.

The second approach largely entails immobilization of the subject. Devices that may be used for this purpose depend on the specific subject and on the nature of the experiment. The most effective immobilization devices are face masks and bite bars. Their use is not widespread, presumably because they may cause discomfort and anxiety in many subjects. Face masks are manufactured on-site from a thermoplastic material, which is heated and molded to the subject's face; holes are usually cut out for eyes, nose, and mouth. Both face masks and bite bars are rigidly attached to a head holder, which itself is attached to the patient table. Less effective for immobilization but ubiquitously used is a set of foam pads squeezed between the inside of the head holder and the subject's head. These can be individually adjusted for comfort; however, they restrict mainly translational motion along the magnet's x-axis and rotational motion about the z- and y-axes. The motion that is most likely to occur, however, is rotation about the x-axis (corresponding to a nodding of the head). For the two most common scanning directions (axial and coronal), this constitutes out-of-plane motion, which is difficult to correct for. Hence the immobilization devices used at present are unsatisfactory and will most certainly be developed further.

Postacquisition motion correction is now quite routinely performed; however, present algorithms leave much to be desired.[28] This correction actively takes into account the nonlinear effect of motion on the BOLD signal, or it may simply ''register'' the image with respect to a reference position. In general, out-of-plane motion is much more difficult to correct than in-plane motion.

It very important to recognize residual motion in a data set, if only to discard the experiment or limit the conclusions that can be reached. Artifacts caused by gross motion have some specific properties that may allow the experimenter to recognize them. If there were pure in-plane motion, we would expect the total signal intensity to be approximately conserved. For linear motion, we would further expect there to be a spatially symmetric pattern of positive and negative signal changes; for rotational motion, such a pattern might still be present over short distances. In the case of out-of plane (or ''through plane'') motion, a similar conservation rule does not apply; furthermore, a susceptibility gradient between two adjacent planes may often not be recognized as such, unless a three-dimensional reconstruction of the image is performed.

8 EXPERIMENTAL DESIGN

As with any experiment studying brain function, proper selection of control and activation conditions is crucial in FMRI. Furthermore, the specific question asked determines parameters like the pulse sequence used, the repetition time, and the spatial matrix size. As with any modality, spatial resolution, temporal resolution, and imaging volume compete with each other. Independent of that, we may distinguish two basic experimental designs: block designs and single trial designs.

8.1 Block Designs

The fundamental difference between block designs and single trial designs is that the former measure activation as an average over the execution of many tasks, while single trial experiments seek to devolve a complex activation pattern on the timescale of a single execution of a task. A basic block design consists of a series of alternating control and stimulation periods. The length of each block is large compared to the hemodynamic response time (which is on the order of a few seconds); hence this design only requires a simple binary reference function for data analysis (see below). Block-design experiments are appropriate when the timescale on which neuronal information is sought is long compared to the hemodynamic response time.

8.2 Single Trial Designs

Single trial experiments require a different design both for experiment and analysis. The duration of a single trial of a cognitive task may be a few seconds. If meaningful temporal characteristics of neuronal activity are to be determined on the timescale of a single trial, the finite hemodynamic response has to be taken into account. Several approaches have been used in the past: an ''averaged single trial'' method,[29] a mixed-trial technique,[30] and a true single-trial technique.[18,31–34]

In the averaged single trial method, FMRI signal acquisition is gated by the onset of the presentation of some task, independent of behavior. That approach is permissible in the limit where behavior and brain function do not vary significantly over repeated trials; this may limit the applicability to some experiments. In addition, insofar as intrinsic hemodynamic responses may be different in different regions, information about temporal differences in neural activity is necessarily limited. Thus comparing time courses across functionally specialized regions is sometimes problematic. Sequential neuronal activity cannot be easily determined simply from the difference in onset times in single averaged time courses. In spite of this difficulty, however, this approach is extremely valuable if information is sought on a temporal scale that is more coarse than the variations in hemodynamic response times.

The mixed-trial technique uses a sequence of random stimuli presented in rapid succession; the data are then analyzed by correlating the responses to the specific sequence pattern of each individual stimulus type. This method makes it possible to assign activation in different areas of the brain to different tasks or aspects of a task, but, again, the temporal resolution is limited to possible differences in hemodynamic response times between regions.

In the true single trial ''time resolved'' technique, hemodynamic response time differences between regions are canceled through the use of multiple individual trials, as shown in Figure 9. A task with a varying temporal parameter is carried out several times and the variation of the hemodynamic response is correlated with this variable parameter. In this manner, the unknown hemodynamic response time is canceled and activation may be assigned to an aspect of the task that does or does not vary with reaction time. This method was used for an investigation of Shepard and Metzler's mental rotation task; it was found that the superior parietal lobe and motor areas are involved in the very act of mental rotation.[33,34]

Because no averaging over several executions of a task is performed in time-resolved single trial FMRI, SNR can be poor. Thus, this method may often not be feasible, particularly at low-magnetic fields. This approach permits the determination of the specific function of given brain areas in the context of a single complex task, unlike methods that use indiscriminately averaged single time courses.

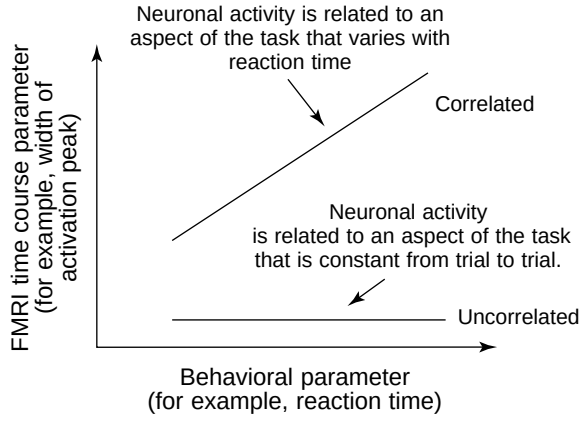

Figure 9 Schematic diagram of time-resolved (true single trial) FMRI. Instead of using a single measurement, data are recorded from a set of single trials with a variable temporal parameter. The relationship between the time course and this variable parameter can provide information about sequential neuronal processing *without* complete knowledge of the hemodynamic response function.

9 DATA ANALYSIS

The analysis of FMRI data is far from being a routine procedure. On the contrary, data analysis in itself is presently an area of intense research. The generic question that FMRI tries to answer may be, ''What is the activity in the brain that generates a certain behavior?'' There are two obstacles to answering this question. The first obstacle concerns the relationship between neuronal activity and the BOLD signal, which is not well understood at present. The second obstacle is the measured signal, which is different from the actual BOLD effect, as it is discrete in time and contaminated by noise.

The physiological origin of the FMRI is unclear, as discussed above, so that we cannot uniquely deduce brain activity from the time course of the BOLD signal in a given area. Conversely, we cannot accurately predict the time course from a postulated neuronal activity. In most FMRI experiments, this problem is simplified by the following implicit assumptions:

1. The activation state is a *binary* physical quantity; that is, a given area of the brain is either active or not active in relation to a task. A corollary of this statement is that, on the timescale of most experiments, activation is immediately switched on and off; it is further assumed that the BOLD signal shows the same behavior, though possibly after a (predictable) delay.
2. The location of the BOLD signal is indeed the location of activity, within the spatial resolution of the experiment.
3. The BOLD signal in distributed areas of the brain exhibits linear behavior; that is, the signal in one area is independent of that in the other areas.

The first assumption may be valid as long as the required temporal resolution is lower than the rise time of the hemodynamic response (a few seconds). For whole-brain studies, this is generally the case; however, if a specific area is to be investigated separately, we may want to scan as fast as 100 ms per imaging volume. In that case, the actual shape of the hemodynamic response may have to be taken into account. This is discussed further below.

The validity of the second assumption, again, depends on the required resolution. The BOLD signal is likely to contain both a metabolic and a flow component. The metabolic component is necessarily localized very close to the actual site of activation. This is why it is thought that the initial decrease in the BOLD signal, which is metabolic in nature, may give better spatial information than the conventional (positive) part of the BOLD signal.[15] The flow component of the signal, on the other hand, may originate several millimeters distant from the true locus of activation.

The third assumption, concerning spatial linearity, is possibly the least well explored. It is likely that this assumption is never strictly valid, as the brain is never ''quiet'', even during a so-called baseline condition. However, within the limits of our measurement and with clever task design, the assumption may often be valid for our purpose at present.

9.1 Paradigmatic Methods

A first-order approximation to the FMRI signal is that it is binary both in space and in time. This means that a given pixel is either active or not active at any given time. In *space*, the binary nature is dependent on voxel size; in a large voxel, we have to allow that an activated area only partially occupies a voxel (this is commonly called the ''partial volume effect'', which is a misnomer, because partial quantities are generally residual nonlinear effects). The signal intensity of that voxel will be the average of that of a quiet voxel and that of an activated voxel; therefore the relative signal change will be decreased and the voxel may or may not be found to be activated. This issue is essentially only important when voxels are large, which may be the case especially in the direction normal to the slice. In *time*, the binary nature of the signal is dependent on the acquisition frequency, as discussed above.

With these approximations, we are able to calculate easily a paradigmatic map voxel-by-voxel. The (usually implicit) null hypothesis for this calculation is that there is no temporal correlation between the time course of a given voxel and a behavioral vector. By inference, this means that that voxel is not significantly activated. The behavioral vector is usually derived from the presentation of the paradigm, or from the response of the subject. A common vector that is used consists of three baseline periods, separated by two activation periods. In a simple visual paradigm, for example, the visual stimulus would be turned on and off synchronously with this vector. Then, the correlation coefficient for each voxel time course with the behavioral vector is calculated. Commonly, Pearson's (linear) correlation coefficient is used, although robust (rank) correlation methods may sometimes be useful.

The proper threshold for such a correlation map is a matter of debate. Since this is a parametric method, we may expect both type I and type II errors. The type I error rate is given by the confidence interval that can be estimated from the correlation coefficient by, Equation (1)

$$p = 1 - \frac{2}{\sqrt{\pi}}\mathrm{erf}\left(cc\sqrt{\frac{N}{2}}\right. \qquad (1)$$

where p is the confidence level for the null hypothesis or the false positive rate, and cc is the correlation coefficient. This is strictly true for Gaussian noise and a good approximation for FMRI.[35]

The question remains what is the proper p-value for this experiment? In experimental science, a commonly used number is 0.05, allowing 5% false positive measurements. In the present case, however, we are performing a voxel-wise test and hence are making a *multiple comparison*. Indeed, each voxel represents a separate experiment in a statistical sense. Suppose we are performing an experiment on the whole brain, with $64 \times 64 \times 35$ matrix size (35 axial slices of thickness 5 mm would cover an average brain). There are on the order of 10^5 voxels in this volume. Hence a p-value of 0.05 would yield 5000 erroneously active voxels. Even if we consider that the brain, being an irregular body, typically occupies considerably less than half of the imaging volume, we are left with a number of type I errors that even exceeds the number of voxels that we expect to be truly activated.

An extreme solution to this problem is to apply the Bonferroni correction. In this correction, the p-value is reduced so that the probability of even a single type I error is small. In this case, we would divide the p-value by the number of voxels in the brain, which would yield a very small number. However, while this conservative procedure eliminates type I errors, it creates a large number of type II errors, and the experimenter may be left with no activated voxels altogether.

Better solutions to the significance problem take into account information beyond the time course of each voxel. This information is usually spatial, complementing the temporal information obtained from the time course. Spatial information is virtually always used implicitly, when the researcher looks at the map and decides which activated voxels ''make sense'' (coincide with a prior hypothesis about the location of activation). Naturally, this is hard to quantify and depends largely on intuition. A more objective approach is to use *spatial clustering*. The expectation here is that the volume of activation is significantly larger than the volume of a single voxel; hence true activation would produce at least two contiguous activated voxels. If p_1 is the probability of a type I error, then the probability of two type I errors in contiguous voxels is given by Equation (2)

$$p_2 = p_1[1 - (1 - p)^{26}] \tag{2}$$

since each voxel in the interior of the volume has 26 nearest neighbors. This, of course, does not make a distinction between neighbors that share a face, an edge or a corner; the contiguity of the latter two in this context is certainly a matter of debate.

From Equation (2), we find for $p_1 = 0.05, p_2 = 0.04$; for $p_1 = 0.01, p_2 = 0.002$; and for $p_1 = 0.001, p_2 = 3 \times 10^{-5}$. Clearly, with a p-value of 0.05, the requirement of spatial contiguity does not add significantly to our confidence. However, when lower p-values are used, spatial clustering improves the confidence level significantly; we can also require more than two contiguous active voxels to accept an activation in order to increase further our confidence interval. Hence, if we assume that the volume of true activation is always large compared to the voxel size, we increase our confidence interval without increasing the type I error frequency by spatial clustering. Hence the experimenter has to decide on the proper p-value and spatial clustering threshold, taking into account the actual voxel size, the anatomical nature of the area under investigation, and the total number of potentially activated voxels. There is no general rule which threshold should be applied.

An added difficulty is faced when the temporal resolution of the experiment is such that the rise and fall of the hemodynamic response has to be taken into account. The success of parametric methods now depends on the quality of the model function employed for the hemodynamic response. Various functional forms have been used for this purpose. One group[36] used a Gaussian function with two variable parameters (''lag'' and ''dispersion'') in order to extract temporal information from a given area. Another group[37] used a set of gamma functions to model the hemodynamic response. Yet another group[31,32] employed a set of trapezoidal reference functions with four variable parameters: onset time, rise time, plateau time, and fall time. It must be kept in mind, however, that it is not necessarily advantageous to describe the hemodynamic response function as accurately as possible, because this function may be variable across different areas of the brain. It is more appropriate to use a function which, with several variable parameters, does equal justice to all possible hemodynamic response functions.

9.2 Non-parametric Methods

Parametric methods are only appropriate if the temporal response to the stimulus is known. As technology improves, however, we would like to measure brain activity on a timescale shorter than the usual hemodynamic response time (that is, a few seconds). In fact, temporal information about neuronal processes in the subsecond range is just becoming available to FMRI. On this timescale, the hemodynamic response is not well characterized. Furthermore, we would often like to measure neuronal activity that is not obviously coupled to an external stimulus (''spontaneous'' activity); a notable example is brain activity during epileptic seizures. In such cases,

nonparametric methods are needed, which organize a set of time courses without regard to a paradigm.

A variety of nonparametric methods for FMRI data analysis have emerged. The three most prominent methods are principal component analysis (PCA), ICA, and FCA.

9.2.1 Principal Component Analysis

This method was introduced for neuroimaging data analysis in 1991.[38] In PCA, the data set is reorganized into eigenimages, where the first eigenimage contains the largest variance between pairs of voxels. For this method to be successful, the task-related signal has to be indeed an important source of variance in the data. This is not always the case, particularly in the presence of physiological noise or gross motion.

9.2.2 Independent Component Analysis

ICA has been applied to FMRI data analysis.[39] In ICA, the data set is organized into individual component processes; it is assumed that the component processes are spatially sparse (yielding focal activation maps) and spatially independent of one another. It has been shown[39] that this method is superior to PCA and correlation analysis for one on/off paradigm; however, the above-mentioned assumptions may not be true for all experiments.

9.2.3 Fuzzy Clustering Analysis

FCA[40,41] is an application of Bezdek's original algorithm.[42] Here a data set of N time points spans an N-dimensional space; each voxel time course is represented by a vector in that space. Voxel time courses are then grouped (clustered) by their proximity in that space; it is assumed that activated voxels behave sufficiently similarly to one another and sufficiently dissimilarly from other voxels that they will fall into small clusters. Examples of this method will be given below.

10 EXAMPLES OF APPLICATIONS

Since 1992 there has been an explosion in applications of FMRI; they are too numerous even to make an attempt to catalog completely. A literal definition of the scope of FMRI would encompass *all* physiological functions; in this sense, for example, a measurement of blood flow at rest would be categorized as FMRI. However, a more sensible operational definition is that FMRI measures *neuronal activity*, albeit through secondary effects like blood flow or the BOLD effect. This will be taken as the common property of all FMRI applications.

10.1 Sensory Tasks

The first FMRI experiments were carried out using visual stimulation.[8–10] There are several reasons for such a choice. Compared with other sensory modalities, the visual system occupies a large area of the human brain; hence we expect multifocal activation even with a simple visual task. Visual stimulation is relatively easy to achieve with commercial goggles or a back projection screen (see section 4), and is not prone to motion artifacts.

Figure 10 shows visual activation induced by a full-field black and white flashing checkerboard. The rest

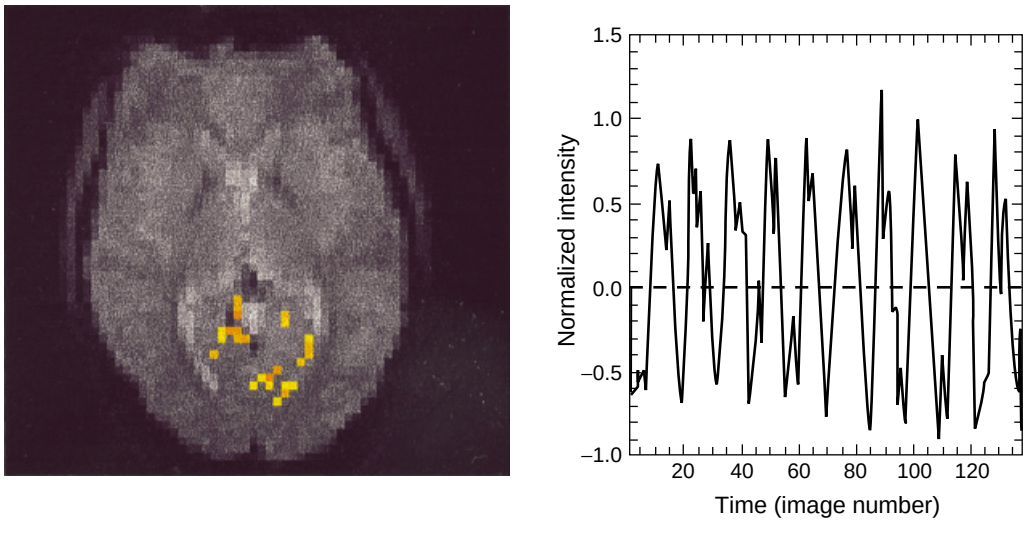

Figure 10 Activation in visual cortices from a flashing checkerboard paradigm. (Data are reproduced by courtesy of Keith Thulborn, Pittsburgh, USA; analysis is reproduced by courtesy of Ray Somorjai and Mark Jarmasz, Winnipeg, Canada.)

condition here consisted of a dark screen with a central fixation point.

Figure 11 shows a sagittal magnetic resonance image of the medial right hemisphere of the visual cortex, demonstrating ODCs.

In Figure 12, results from another visual stimulation experiment are shown. In addition to the BOLD signal, cerebral blood flow (CBF) and cerebral metabolic rate of oxygen ($CMRO_2$) are measured. Note that hot spots in the CBF and $CMRO_2$ maps correspond well with one another, while hot spots in the BOLD map are alienated from the former two.[43]

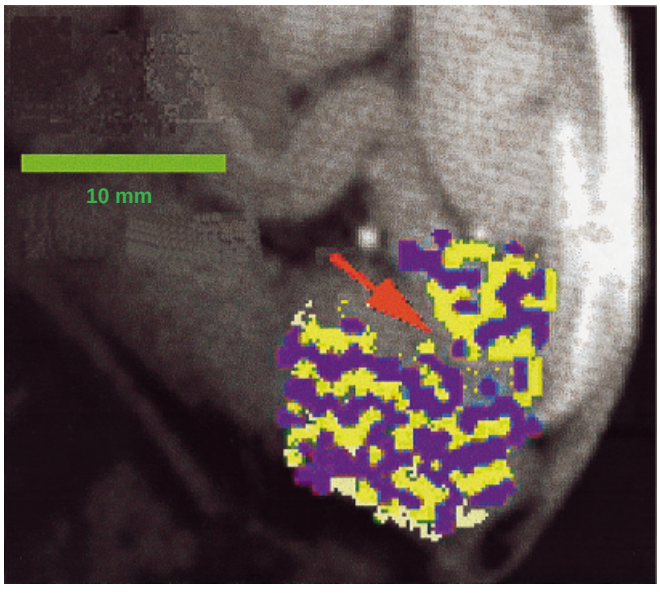

Figure 11 Sagittal magnetic resonance image of the medial right hemisphere of visual cortex, demonstrating ocular dominance columns (ODCs) localized using BOLD FMRI at 4 T. Blue (yellow) areas indicate ODCs corresponding to the left (right) eye. Note the exquisite (submillimeter) spatial resolution demonstrated in this experiment. The red arrow shows the calcarine fissure. (Reproduced by courtesy of Brad Goodyear and Ravi Menon, London, Canada.)

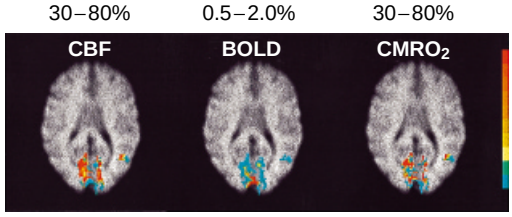

Figure 12 Activation in visual cortices from a visual paradigm. (Reproduced by courtesy of Seong-Gi Kim, Minneapolis, USA.)

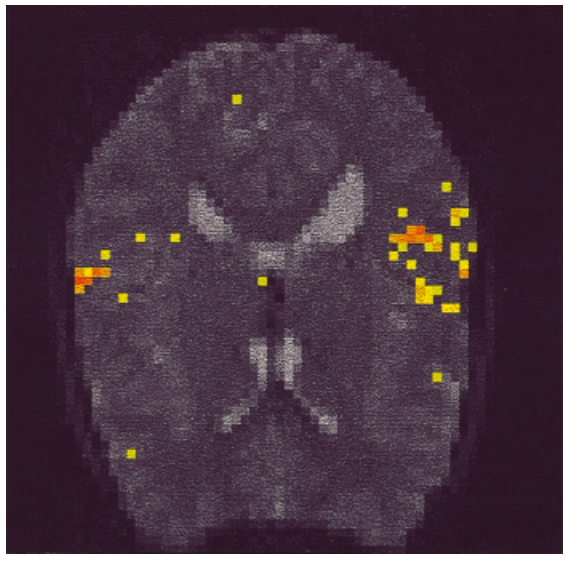

Figure 13 Activation during auditory stimulation. (Reproduced by courtesy of Bob Turner and Karl Friston, London, UK.)

Figure 13 shows activation during auditory stimulation.

Figure 14 shows results from a tactile stimulation of forehead and chin. Note how forehead stimulation produced more inferior activation along the postcentral gyrus relative to chin stimulation, consistent with

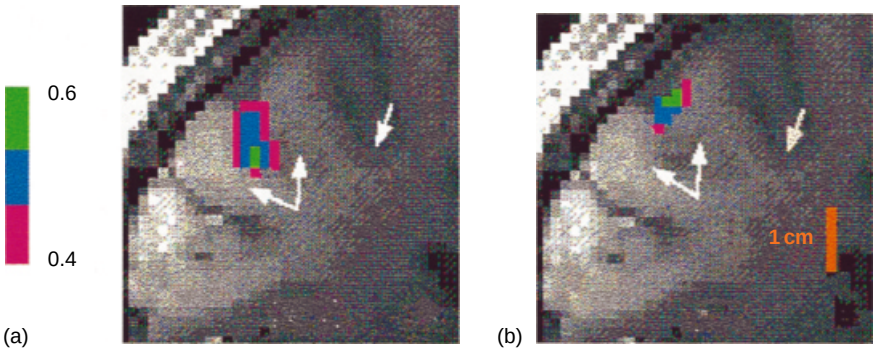

Figure 14 Activation from a tactile stimulation of (a) forehead and (b) chin. Double arrows indicate the postcentral gyrus; single arrows indicate the fundus of the central sulcus. (Reproduced by courtesy of Philip Servos, Waterloo, Canada.)

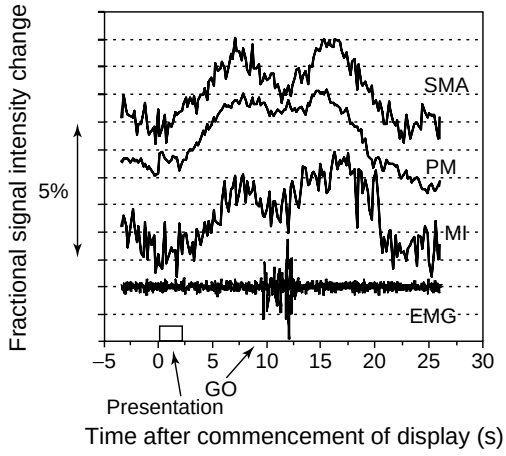

Figure 15 Time courses in motor areas from a single trial of a delayed finger movement task.

an inverted face representation of the somatosensory homunculus.[44]

10.2 Movement Tasks

A common problem for tasks involving overt movement is gross motion of the head (see section 7). Therefore, many movement tasks involve only finger movement, which does not generally lead to excessive head movement. An example is shown in Figure 15.[31] The task is a delayed finger movement task. The subject is shown four circles on the screen. During the instruction period, the four circles in the bottom row light up in a particular order (e.g. 1–2–4–3). The subject remembers this order during a variable delay period (between 0 and 7 s). Then, the upper circle lights up, serving as the "Go" signal, and the subject presses the four buttons on a button pad

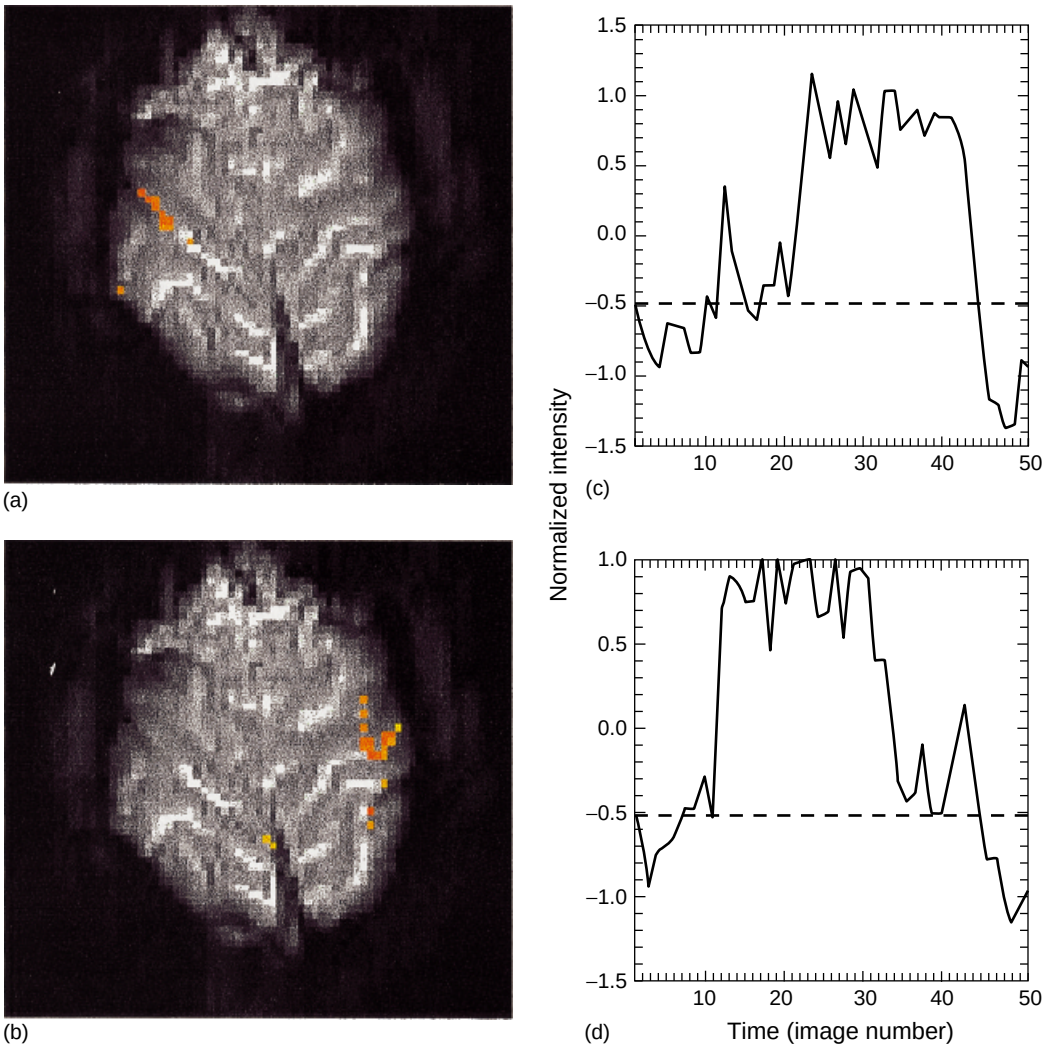

Figure 16 Activation from a finger tapping task. Note two activation clusters, one for the right hand (b), and one for the left hand (a). (Data are reproduced by courtesy of Ewald Moser, Vienna, Austria; analysis is reproduced by courtesy of Ray Somorjai and Mark Jarmasz, Winnipeg, Canada.)

in the prescribed order. Activation is seen in various motor areas (lateral premotor area (PM), supplementary motor area (SMA), and primary motor cortex (M1)). The time courses shown are from a single trial of one subject. Activation can be seen in all three areas during the delay period and during the execution period. The absence of movement during the delay period is monitored by EMG (electromyography) recording on the forearm. Note the exquisite temporal resolution of this experiment that permits the distinction between two separate components of a task (movement preparation and movement execution).

Figure 16 shows another finger movement task. After a rest period (10 volumes), the right hand starts moving. From time index 21 to 30, the movement is bimanual, and from 31 to 40, only the left hand moves. The data is analyzed by FCA (see section 9). Note how the right and left motor area are properly separated by this method. There is another cluster (not shown) that shows activity only during the bimanual movement period.

10.3 Cognitive Tasks

Figure 17 shows activation from a cognitively demanding finger movement task. In this case, the subject is asked to perform a series of sets of four finger movements without pattern repetition (i.e. 1234, 1423, 1243, etc.). In addition to motor areas, cingulate cortex and dorsolateral prefrontal cortex are active. Note that

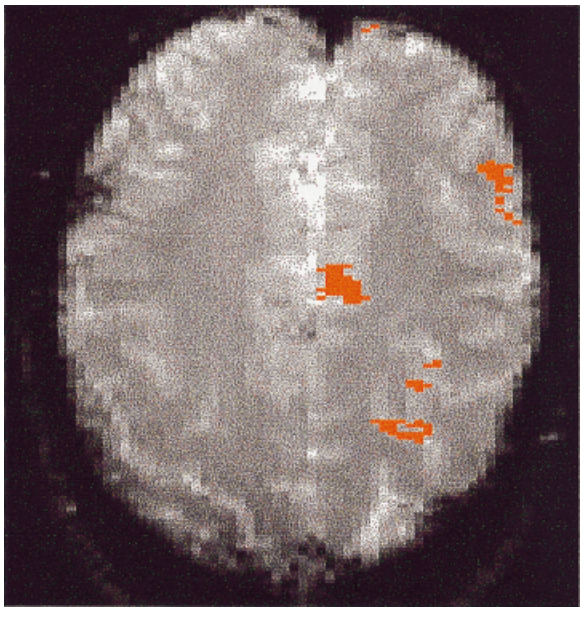

Figure 17 Activation pattern from a cognitively demanding finger movement task. (Reproduced by courtesy of Mike McIntyre, Winnipeg, Canada.)

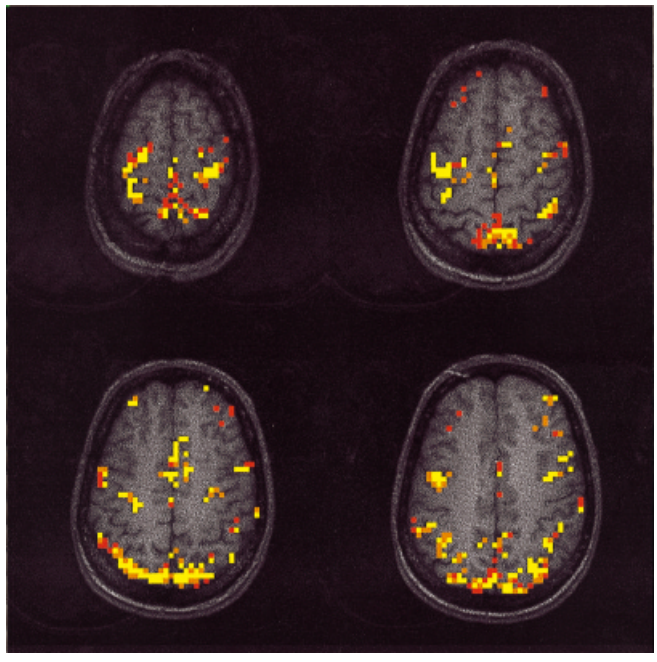

Figure 18 Activation map from a mental rotation task. Note the activity in frontal motor areas during a task that does not involve overt movement.

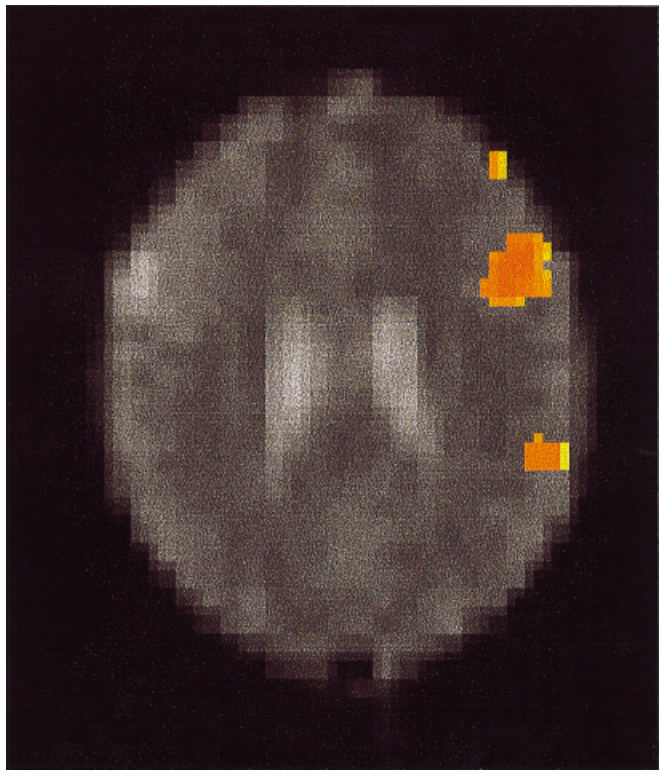

Figure 19 Activation during silent word generation. (Reproduced by courtesy of Stefan Posse, Jülich, Germany.)

this experiment is performed by FLASH imaging (see section 5).

Another cognitive task is Shepard and Metzler's mental rotation task.[32,34] Subjects are shown drawings of pairs of three-dimensional objects which are rotated with respect to one another. They then have to decide if the two objects are identical. It has been conjectured that the subject mentally rotates the two objects into congruence. Figure 18 shows the activation map from such a task; it can be seen that areas of the parietal lobe as well as frontal motor areas are active here.

In Figure 19, activation during a task involving silent word generation is shown.

10.4 Other Experiments

Figure 20 shows spontaneous activation in an epileptic patient during seizures. This is an excellent example of the use of paradigm-free methods (in this case, FCA) which are essential when there is no clear expectation of the time course of activation.

FMRI is not necessarily limited to the brain. Figure 21 shows time courses from activated pixels in the spinal chord during a hand movement paradigm. Spinal cord FMRI poses its own unique challenges, because the area to be imaged is very small and very prone to motion artifacts.

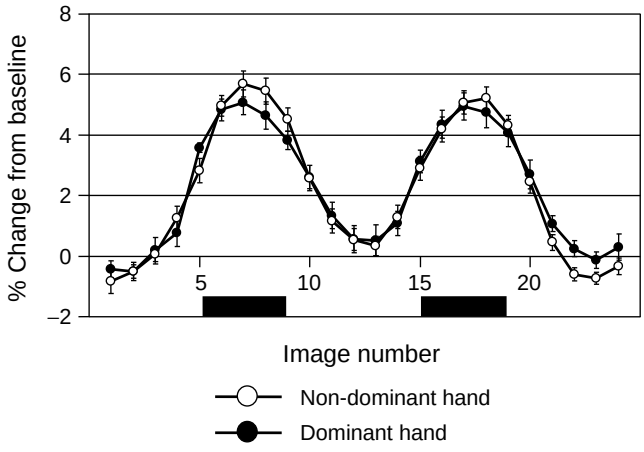

Figure 21 Average time course intensity data obtained from the cervical spinal cord in healthy volunteers performing a motor task with one hand. Equal length periods of rest and exercise (indicated by the black bars) were alternated. Magnetic resonance image data were acquired at 1.5 T with a GRE EPI technique. (Reproduced by courtesy of Patrick Stroman, Winnipeg, Canada.)

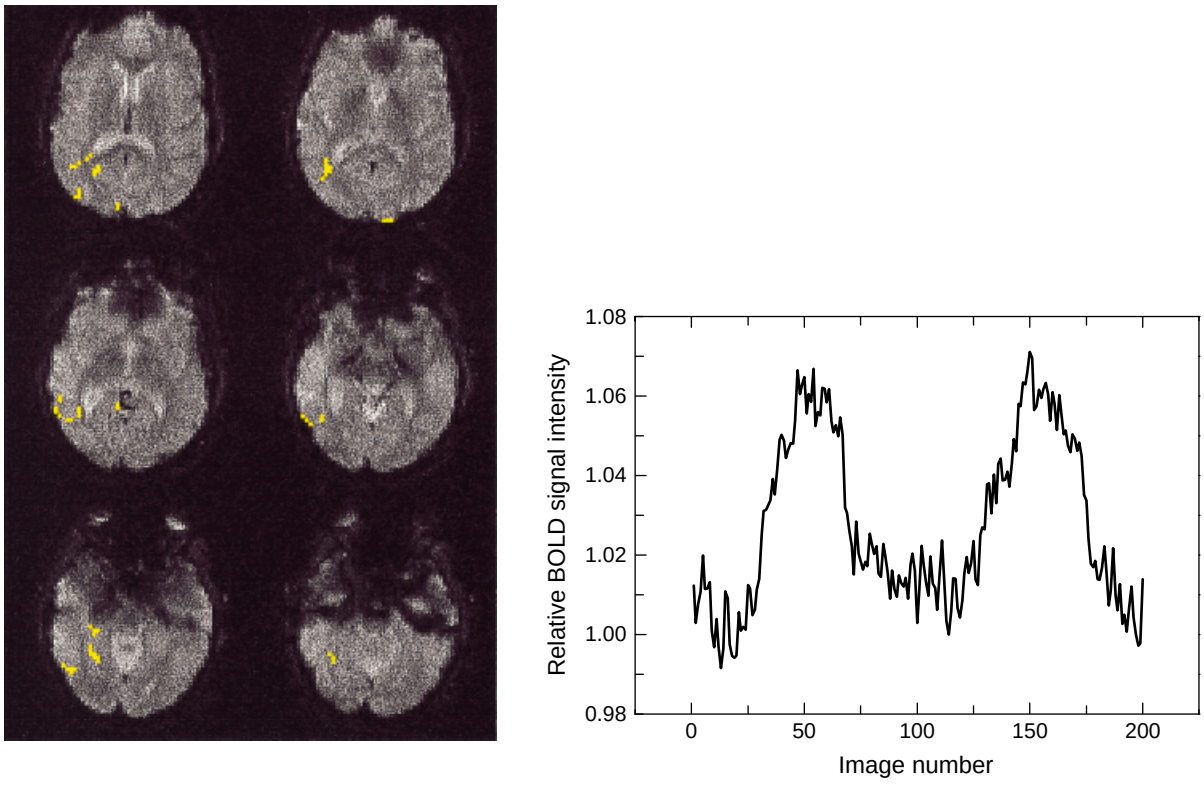

Figure 20 Spontaneous activation in an epileptic patient during seizures. (Data are reproduced by courtesy of Keith Thulborn, Pittsburgh, USA; analysis is reproduced by courtesy of Ray Somorjai and Mark Jarmasz, Winnipeg, Canada.)

11 COMPARISON WITH OTHER METHODS AND OUTLOOK

At present, FMRI occupies a prominent position among the various methods of studying brain function. Electrophysiological methods (for example, EEG or single neuronal recording) excel with a temporal resolution of milliseconds. In the case of EEG, however, the spatial resolution is centimeters at best and single neuronal recording studies, while extremely accurate in time and space, can only cover a very small number of neurons, and this only in animals or human patients. Older neuroimaging methods, like PET or single photon emission computed tomography (SPECT) generally fall behind FMRI both in temporal and spatial resolution.

Many issues in FMRI are unresolved at present. Most important here is that the physiological basis of the coupling between neuronal activity, oxygen consumption, blood flow, and the BOLD effect is not well understood. These questions are under intense scrutiny by some of the best research groups in the world and a better understanding of these effects will almost certainly lead to applications of FMRI with a spatial resolution on the order of hundreds of micrometers, and a temporal resolution on the order of hundreds of milliseconds. The increasing availability and quality of MRI machines primarily for diagnostic purposes is likely to help FMRI to remain the foremost neuroimaging method in the near future.

ABBREVIATIONS AND ACRONYMS

BOLD	Blood Oxygen Level Dependent
CBF	Cerebral Blood Flow
$CMRO_2$	Cerebral Metabolic Rate of Oxygen
CNR	Contrast-to-noise-ratio
EEG	Electroencephalography
EMG	Electromyography
EPI	Echo Planar Imaging
FCA	Fuzzy Clustering Analysis
FDA	Food and Drug Administration
FLASH	Fast Low-angle Shot
FMRI	Functional Magnetic Resonance Imaging
GRE	Gradient-recalled Echo
ICA	Independent Component Analysis
LCD	Liquid Crystal Display
M1	Primary Motor Cortex
MEG	Magnetoencephalographic
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
ODC	Ocular Dominance Column
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PM	Premotor Area
RARE	Rapid Acquisition with Relaxation Enhancement
RF	Radiofrequency
SAR	Specific Absorption Rate
SMA	Supplementary Motor Area
SNR	Signal-to-noise-ratio
SPECT	Single Photon Emission Computed Tomography

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Angiography • Magnetic Resonance in Medicine, High Resolution Ex Vivo • Multinuclear Magnetic Resonance Spectroscopic Imaging

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance • Electron Spin Resonance Spectroscopy • High-performance Liquid Chromatography Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Resonance in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
2. F. Bloch, W.W. Hansen, M.E. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
3. P.C. Lauterbur, 'Image Formation by Induced Local Interactions. Examples Employing Nuclear Magnetic Resonance', *Nature*, **242**, 190–191 (1973).
4. P. Mansfield, A.A. Maudsley, 'Medical Imaging by NMR', *Br. J. Radiol.*, **50**, 188–194 (1977).
5. R. Damadian, M. Goldsmith, L. Minkoff, 'FONAR Image of the Live Human Body', *Physiol. Chem. Phys.*, **9**, 97–100 (1977).
6. H. Clow, I.R. Young, *New Sci.*, **80**, 588 (1978).
7. S. Ogawa, T.M. Lee, 'Magnetic Resonance Imaging of Blood Vessels at High Fields: In Vivo and In Vitro Measurements and Image Simulation', *Magn. Reson. Med.*, **16**, 9–18 (1990).
8. P.A. Bandettini, E.C. Wong, R.S. Hinks, R.S. Tikofsky, J.S. Hyde, 'Time Course EPI of Human Brain Function During Task Activation', *Magn. Reson. Med.*, **35**, 107–113 (1992).
9. K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kemedy, B.E. Hoppel, M.S. Cohen, R. Turner, H.M. Cheng, T. Brady, B. Rosen, 'Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation', *Proc. Natl. Acad. Sci. USA*, **89**, 5675–5679 (1992).
10. S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.-G. Kim, H. Merkle, K. Ugurbil, 'Intrinsic Signal Changes Accompanying Sensory Stimulation: Functional Brain Mapping with Magnetic Resonance Imaging', *Proc. Natl. Acad. Sci. USA*, **89**, 5951–5955 (1992).
11. S. Ogawa, T.M. Lee, B. Barrere, 'The Sensitivity of Magnetic Resonance Image Signals of Rat Brain to Changes in the Cerebral Venous Blood Oxygenation', *Magn. Reson. Med.*, **29**, 205–210 (1993).
12. S. Ogawa, T.M. Lee, A.R. Kay, D.W. Tank, 'Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation', *Proc. Natl. Acad. Sci. USA*, **87**, 9868–9872 (1990).
13. S. Ogawa, R.S. Menon, D.W. Tank, S.G. Kim, H. Merkle, J.M. Ellermann, K. Ugurbil, 'Functional Brain Mapping by Blood Oxygenation Level Dependent Contrast Magnetic Resonance Imaging. A Comparison of Signal Characteristics with Biophysical Model', *Biophys. J.*, **64**, 803–812 (1993).
14. C. Janz, O. Speck, J. Hennig, 'Time Resolved Measurements of Brain Activation after a Short Visual Stimulus: New Results on the Physiological Mechanisms of the Cortical Response', *NMR Biomed.*, **10**, 222–220 (1997).
15. R.S. Menon, S. Ogawa, J.P. Strupp, P. Andersen, K. Ugurbil, 'BOLD-based Functional MRI at 4 Tesla Includes a Capillary Bed Contribution: Echo-planar Imaging Correlates with Previous Optical Imaging Using Intrinsic Signals', *Magn. Reson. Med.*, **33**, 453 (1995).
16. J. Hennig, C. Janz, O. Speck, T. Ernst, 'Functional Spectroscopy of Brain Activation Following a Single Light Pulse: Examination of Mechanism of the Fast Initial Response', *Int. J. Sys. Technol.*, **6**, 203 (1995).
17. J.S. Gati, R.S. Menon, K. Ugurbil, B.K. Rutt, 'Experimental Determination of the BOLD Field Strength Dependence in Vessels and Tissue', *Magn. Reson. Med.*, **38**, 296–302 (1997).
18. S.-G. Kim, W. Richter, K. Ugurbil, 'Limitations of Temporal Resolution in Functional MRI', *Magn. Reson. Med.*, **37**, 631–636 (1997).
19. A. Haase, 'Snapshot FLASH MRI: Application to T1-, T2-, and Chemical Shift Imaging', *Magn. Reson. Med.*, **13**, 77–89 (1990).
20. J. Hennig, A. Nauerth, H. Friedburg, 'RARE Imaging: A Fast Imaging Method for Clinical MR', *Magn. Reson. Med.*, **3**, 823–833 (1986).
21. P. Mansfield, 'Multi-planar Image Formation Using NMR Spin Echoes', *J. Phys. C: Solid State Phys.*, **10**, K55 (1977).
22. M.K. Stehling, R. Turner, P. Mansfield, 'Echo-planar Imaging: Magnetic Resonance Imaging in a Fraction of a Second', *Science*, **254**, 43 (1991).
23. C. Mayer, D. Nishimura, A. Macovski, 'Fast Spiral Coronary Artery Imaging', *Magn. Reson. Med.*, **28**, 202 (1992).
24. F.G. Shellock, E. Kanal, *Magnetic Resonance. Bioeffects, Safety, and Patient Management*, Lippincott-Raven, Philadelphia, 1996.
25. Kamil Ugurbil, Personal Communication.
26. X. Hu, T.H. Le, T. Parrish, P. Erhard, 'Retrospective Estimation and Correction of Physiological Fluctuation in Functional MRI', *Magn. Reson. Med.*, **34**, 201–212 (1995).
27. D.H. Wu, J.S. Lewin, J.L. Duerk, 'Inadequacy of Motion Correction Algorithms in Functional MRI: Role of Susceptibility-induced Artifacts', *J. Magn. Reson. Imag.*, **7**, 365–370 (1997).
28. H. Eviatar, B. Schattka, J.C. Sharp, J. Rendell, M.E. Alexander, 'Real Time Head Motion Correction for Functional MRI', *Proc. Int. Soc. Magn. Reson. Med.*, Philadelphia, 269 (1999).
29. R.L. Buckner, P.A. Bandettini, K.M. O'Craven, R.L. Savoy, S. Petersen, M.E. Raichle, B.R. Rosen, 'Detection of Cortical Activation During Averaged Single Trials of a Cognitive Task Using Functional Magnetic Resonance Imaging', *Proc. Natl. Acad. Sci.*, **93**, 14 878–14 883 (1996).
30. V.P. Clark, J.M. Maisog, J.V. Haxby, 'FMRI Study of Face Perception and Memory Using Random Stimulus Sequences', *J. Neurophysiol.*, **79**, 20 892–21 366 (1998).
31. W. Richter, P.M. Andersen, A.P. Georgopoulos, S.-G. Kim, 'Sequential Activity in Human Motor Areas During a Delayed Cued Finger Movement Task Studied by Time-resolved FMRI', *Neuroreport*, **8**, 1257–1261 (1997).

32. W. Richter, K. Ugurbil, A.P. Georgopoulos, S.-G. Kim, 'Time-resolved FMRI of Mental Rotation', *Neuroreport*, **8**, 3697–3702 (1997).
33. W. Richter, 'High Temporal Resolution FMRI at very High Field', *Top. Magn. Reson. Imag.*, **10**, 51–62 (1999).
34. W. Richter, R. Somorjai, R. Summers, M. Jarmasz, R.S. Menon, J.S. Gati, A.P. Georgopoulos, C. Tegeler, K. Ugurbil, S.-G. Kim, 'Motor Area Activity During Mental Rotation Studied by Time-resolved Single-trial FMRI', *J. Cognitive Neurosci.*, **12**, 310–320 (2000).
35. P.A. Bandettini, A. Jesmanowicz, E.C. Wong, J.S. Hyde, 'Processing Strategies for Time-course Data Sets in Functional MRI of the Human Brain', *Magn. Reson. Med.*, **30**, 161–173 (1993).
36. J.C. Rajapaske, F. Kruggel, J.M. Maisog, D.Y. von Cramon, 'Modeling Hemodynamic Response for Analysis of Functional MRI Time Series', *Hum. Brain Mapping*, **6**, 283–300 (1998).
37. K.J. Friston, P. Fletcher, O. Josephs, A. Holmes, M.D. Rugg, R. Turner, 'Event-related FMRI: Characterizing Differential Responses', *Neuroimage*, **7**, 30–40 (1998).
38. J.R. Moeller, S.C. Strother, 'A Regional Covariance Approach to the Analysis of Functional Patterns in Positron Emission Tomographic Data', *J. Cereb. Blood Flow Metab.*, **11**, A121–A135 (1991).
39. M.J. McKeown, S. Makeig, G.G. Brown, T.-P. Jung, S.S. Kindermann, A.J. Bell, T.J. Sejnowski, 'Analysis of FMRI Data by Blind Separation into Independent Spatial Components', *Hum. Brain Mapping*, **6**, 160–188 (1998).
40. R.L. Somorjai, M. Jarmasz, R. Baumgartner, W. Richter, 'Exploratory Analysis of fMR Images: Voxel Preselection via ''Self-similarity'' ', *Proc. Int. Soc. Magn. Reson. Med.*, Philadelphia, 1718 (1999).
41. R.L. Somorjai, M. Jarmasz, 'Exploratory Data Analysis of fMR Images: Philosophy, Strategies, Tools, Implementation', *Proc. Int. Soc. Magn. Reson. Med.*, Philadelphia, 1714 (1999).
42. J. Bezdek, 'Pattern Recognition with Fuzzy Objective Function Algorithms', Plenum, New York, 1981.
43. S.-G. Kim, E. Rostrup, H.B.W. Larsson, S. Ogawa, O.B. Paulson, 'Determination of Relative $CMRO_2$ from CBF and BOLD Changes: Significant Increase of Oxygen Consumption Rate During Visual Stimulation', *Magn. Res. Med.*, **41**, 1152–1161 (1999).
44. P. Servos, S.A. Engel, J. Gati, R. Menon, 'FMRI Evidence for an Inverted Face Representation in Human Somatosensory Cortex', NeuroReport, **10**, 1393–1395 (1999).

Magnetic Resonance in Medicine, High Resolution Ex Vivo

Ian C. Smith and Dorothea E. Blandford
Institute for Biodiagnostics, Winnipeg, Canada

1 Introduction 1
1.1 Biological Fluids 2
1.2 Biological Tissue 3
2 Diagnosis of Inborn Errors of Metabolism 4
3 Measurement of Renal Injury 5
4 Diagnosis of Neurological Disorders 7
5 Cancer Diagnosis 7
5.1 Colon Cancer 7
5.2 Thyroid Cancer 8
5.3 Brain Tumors 9
5.4 Ovarian Tumors 9
5.5 Prostate Cancer 9
5.6 Other Neoplastic Conditions 9
6 Conclusions 9
Abbreviations and Acronyms 9
Related Articles 10
References 10

Nuclear magnetic resonance (NMR) spectroscopy, or magnetic resonance spectroscopy (MRS), is a powerful physical technique to determine the quantity and structure of chemical compounds in a specimen. It involves the interaction of radiofrequency (RF) electromagnetic radiation with a collection of nuclei immersed in a strong magnetic field. It is unique in its ability to provide nondestructive chemical analyses in vivo and in vitro. Moreover, no or very little pretreatment of the specimen is required prior to making a measurement. Spectra take only a few minutes to acquire, and it is not necessary to preselect the metabolites of interest in order to detect and quantify them. Its limitations are a lack of availability of instruments and large databases of spectral changes correlated with pathological conditions, and a detection sensitivity of approximately 1 μ*M.*

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

NMR spectroscopy or MRS is a technique of vast potential. It is a powerful, nondestructive technique that can be used to determine the complete structural and conformational analysis of complex molecules, as well as the quantitative analysis of complex mixtures. NMR involves the absorption or emission of RF energy by nuclei in a magnetic field. The higher the magnetic field used, the greater is the detection sensitivity of the instrument.

High-resolution NMR has been widely used by scientists to characterize chemical compounds. The narrow resonances reveal fine structure due to spin–spin splitting, and a wide variety of chemically shifted resonances reveal the type and number of nuclear species. To illustrate. the sensitivity of NMR to chemical structure, the ^{1}H-NMR spectrum of ethyl benzene is shown in Figure 1. The spectrum contains individual groups of resonances due to the CH_3, CH_2, and aromatic protons. Integration of the areas under the multiplets yields the relative number of protons in each group, namely 3 : 2 : 5. The spin–spin splittings of the resonances indicate that the CH_3 and CH_2 groups are adjacent. The splittings of the aromatic protons are more complex, due to the similar magnitudes of the splittings and the differences in chemical shift of the types of protons – a so-called second-order spectrum. Multiple-dimensional NMR spectroscopy, in which one NMR parameter is portrayed along one axis, others along the other axes, have greatly aided the resolving power of NMR in complex situations. The most useful of these is correlated spectroscopy (COSY), where the chemical shifts of different resonances in a spectrum are correlated. More detailed discussions of the origin of NMR spectra are available.[1–3] In addition to ^{1}H, there is a variety of nuclei useful for biomedical applications. These are shown in Table 1 with their relevant detection properties. Their usefulness decreases in the order ^{1}H, ^{31}P, ^{13}C, ^{23}Na, ^{7}Li, ^{87}Rb, ^{19}F, ^{15}N.

Applications of high-resolution NMR to biomedical problems were hindered by the common view that the spectra would be difficult to obtain, the resonances would be broad, and overlaps of many broad resonances would make the spectra difficult to interpret. This has proved to be quite incorrect, and high-resolution NMR has been very successfully applied to biological fluids, tissue ex vivo, and living organisms.

One reason for this success was the rapid increase in availability of NMR instruments operating at high magnetic field. Detection sensitivity increases with the magnitude of the applied magnetic field, from linearly to the 1.75 power, depending on the nature of the specimen[4] (Table 2). Concomitantly, the spectral separation in hertz between chemically different species

Table 1 Properties of NMR active nuclei of use in biomedical research

Property	Nucleus							
	^{1}H	^{19}F	^{23}Na	^{31}P	^{87}Rb	^{7}Li	^{13}C	^{15}N
Frequency[a]	500	470	132	202	163	194	126	51
Detectability[b]	100	83	9	7	18	29	2	0.1
Isotopic abundance (%)	100	100	100	100	28	93	100	100

[a] Megahertz at 11.7 T.
[b] At 100% isotopic composition.

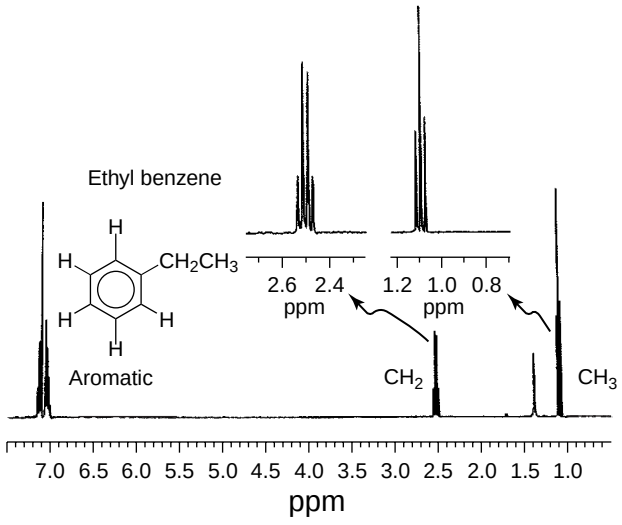

Figure 1 ^{1}H-NMR spectrum (360 MHz) of ethyl benzene (0.1%) in deuterochloroform.

Table 2 Signal-to-noise ratio (S/N) for various spectrometers using information supplied by manufacturers

Year	Spectroscopy frequency (MHz)	S/N[a]
1961	60	6
1965	100	30
1969	220	80
1978	200	300
1978	360	800
1985	500	3600
1989	600	6000
1994	750	12 000

[a] Strongest peak in the methylene proton signal of ethyl benzene, 1% by volume. Present day spectrometers use lower concentrations; the numbers have been scaled for comparison.

increases linearly with the field. A wide variety of multinuclear and multidimensional techniques have led to increased sensitivity and spectral simplification. Finally, the increasing megahertz available per unit cost has made the accessibility of high-field NMR much greater.

Some of the earliest applications to complex biomedical systems were made in the 1970s, principally with ^{31}P NMR.[5] A long period followed with emphasis placed on microorganisms and cell extracts. The observation of well-resolved ^{31}P spectra from humans in vivo[6] stimulated a new interest in NMR of tissue. NMR of biofluids evolved slowly but steadily during this time. Applications of high-resolution NMR in clinical and pharmaceutical chemistry,[7,8] as well as NMR of tissue,[9] have been reviewed.

1.1 Biological Fluids

A large number of biological fluids is accessible for MRS study ex vivo. Early applications of ^{1}H-MRS analysis demonstrating its utility involved analysis of urine and serum.[10–12] A large number of other biofluids, including cerebrospinal fluid (CSF),[13] amniotic fluid,[14,15] synovial fluid,[16,17] seminal plasma,[18,19] bile,[20,21] and tissue extracts[22,23] have since been examined.[8,24,25] ^{1}H-MRS analysis of biofluids offers several advantages: it requires little or no preparation of the sample; high-quality spectra, suitable for quantitative measurements, may be obtained in under 15 min; potentially all ^{1}H-containing metabolites (including acids, bases, and neutral compounds) may be measured by applying a suitable spectroscopic technique; and the possibility of studying simultaneously, in a single measurement, all of the more usually represented metabolites without any test preselection.

Biological fluids contain compounds of interest at relatively low concentrations in water. The ^{1}H resonance due to water is roughly 10^6 times more intense than the resonance of the compounds of interest. This creates a serious problem of dynamic range. Two approaches to resolve this difficulty have been applied. One is the increasing dynamic range of the analog-to-digital convertors (ADCs), reaching values as high as 32 bits ($2^{32} = 4.3 \times 10^9$, a significant range of discrimination). Furthermore, it has been found that saturation of the H_2O resonance by simple irradiation, or by a variety of pulse sequences, serves to suppress this resonance to tolerable amplitudes. One or both of these methods may be applied.

A second problem is the wide variety of compounds present in biological specimens. The high magnetic fields presently available, up to 19 T (800 MHz for ^{1}H), serve to diminish this problem to a tractable one.

A further problem is the simultaneous presence of species of high and low molecular weight. High-molecular-weight species rotate slowly, and thus have very broad resonances which may obscure those of the low-molecular-weight species. Fortunately, broad resonances are usually indicative of short transverse relaxation times (T_1), and these may be used in data manipulation to remove the confounding effect of the resonances due to high-molecular-weight species such as proteins, nucleic acids, and membrane lipids. This is demonstrated in Figure 2, where manipulation of the free induction decay leads to total suppression of the broad resonances due to proteins in plasma.

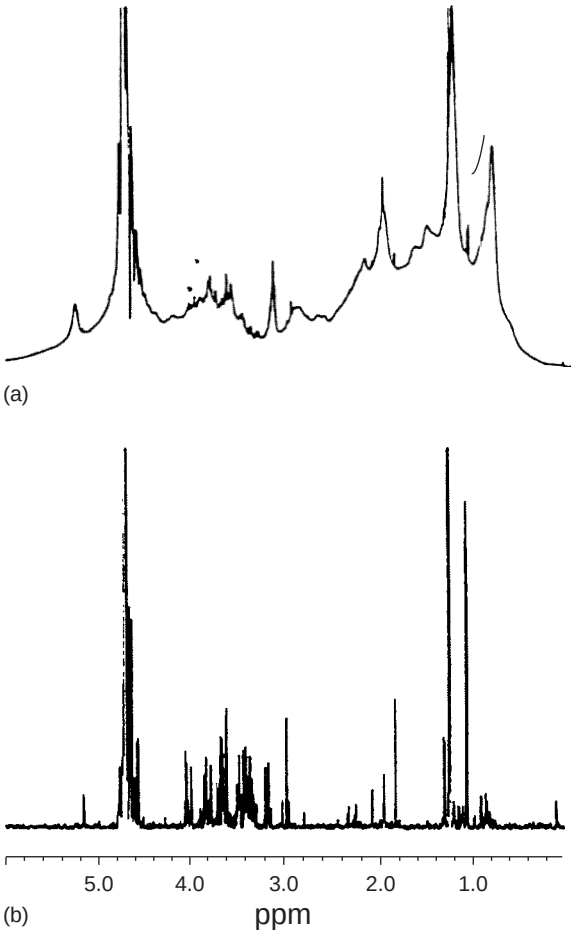

Figure 2 ^{1}H-NMR spectrum (360 MHz) of human blood plasma; the strong resonance due to H_2O has been reduced by irradiation at its frequency before commencing the spectral acquisition: (a) with no spectral manipulation; (b) after multiplication of the free induction decay by a sine-bell function to minimize broad resonances.

1.2 Biological Tissue

For many years it was believed that NMR of tissue was a hopeless pursuit. The seminal article by Mountford et al.,[26] demonstrating ^{1}H spectra of good resolution from the mammary tissue of rat, stimulated a resurgence of interest. High-resolution ^{1}H spectra (360 MHz) of human colon tissue were reported. The transverse relaxation times (T_2) of the lipid CH_2 resonance at 1.3 ppm appeared to be indicative of the metastatic potential of the hyperplastic tissue. The race was on! Since this initial observation, a wide variety of studies has been reported on human tissue ex vivo, suggesting that the high information content of the spectra should lead to effective clinical use of NMR in vivo.

Tissue differs drastically from biological fluids in that it is macroscopically solid. This leads to difficulty in optimizing the homogeneity of the applied magnetic field for optimal spectral resolution. A technique has been reported that minimizes this problem.[27] Figure 3 shows the positioning of a piece of human tissue within a capillary tube facilitating the optimization of field homogeneity and allowing an estimation of the volume of the specimen. This method has been used in most quantitative studies to date.

Ex vivo tissue is not dead tissue; it has, however, begun a progression toward death. In studies of human tissue ex vivo, great care must be taken to ensure that the spectra of the many specimens are taken at similar times after removal from the host. One way to ensure this is to freeze the samples at liquid N_2 temperature within minutes of excision. These frozen

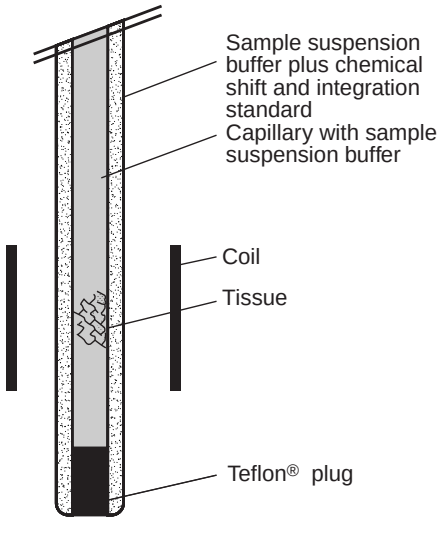

Figure 3 Diagram of the positioning of a biopsy specimen within a standard NMR tube of outside diameter 5 mm. (Reproduced by permission from Kuesel et al.[27])

specimens are stable for weeks. Likewise, a rigid protocol for warming the specimens and obtaining their spectra must be followed.

NMR spectra contain a wealth of information – some vital, some useful, some useless. It is often difficult to decide by visual inspection which regions of the spectra will be most useful in classifying the specimens. Ratios of peaks can be useful, but much time must be spent before the most useful ratios are found. In our experience computerized methods to select the most discriminatory region, and to perform the classification, are robust and accurate. The most exhaustive application of these methods is described in Somorjai et al.[28] Recently, a description of the regional selection process has appeared.[29]

For spectral classification, multivariate analysis has proved to be very effective. A variety of methods, such as linear discriminant analysis, quadratic discriminant analysis, neural networks, and genetic algorithms may be used. Very often the simplest method, linear discriminant analysis, is sufficient. In the most difficult cases, a number of methods may be used and finally entered into a meta classifier.[28]

It is crucial that the NMR spectra be adequately prepared for the classification process. They must be normalized for area, adjusted for chemical shift, and the most discriminatory regions determined. Data must be divided into a training set, in which the method is calibrated, and a test set, in which the accuracy of the method is tested. Regrettably, this has been done very rarely in the literature to date, yielding promising results but with a method that is not robust. To enhance robustness, the leave-one-out method should be implemented. The procedure is challenged by leaving out one data set, classifying it by exposure to the existing classifier, leaving another out and classifying it, until all data sets have been thus treated. Finally, a sufficient number of data sets must be used if the classifier is to be accurate, specific, and robust. The number required depends strongly on the types of spectra compared.

Using the procedures described above, experimental results from which no obvious discriminators can be seen, and in which there is a wide variation in spectra from a given class, can be accurately classified. The reader is referred to the articles in this *Encyclopedia* which deal with chemometrics encompassing some of the same techniques (**Chemometrics; Classical and Nonclassical Optimization Methods; Clustering and Classification of Analytical Data; Multivariate Calibration of Analytical Data; Second-order Calibration and Higher**). The remainder of this article presents an overview of the medical applications of high-resolution MRS ex vivo. Specifically, progress has recently been made towards the diagnosis of inborn errors of metabolism, the measurement of renal injury, diagnosis of neurological disorders, and cancer diagnosis.

2 DIAGNOSIS OF INBORN ERRORS OF METABOLISM

High-field ^{1}H spectroscopy of biofluids offers great potential for furthering the understanding of disease processes in humans. Historically, one of the most successful clinical applications of MRS has been the detection of a large number of metabolic disorders. In these disorders, the reduction or absence of activity of an enzyme or cofactor can have dramatic consequences for metabolism and its control. Many inherited metabolic disorders result in the accumulation of large amounts of organic intermediates produced proximal to the defective enzyme step, which eventually spill into the blood and urine. ^{1}H-MRS has been used to study the urinary excretion of such compounds. Initial studies demonstrated the ability to diagnose a number of metabolic disorders including histidinemia,[30] citrullinemia,[31] argininosuccinic acid lyase deficiency,[31,32] ornithine carbamoyl transferase deficiency,[31] cystinuria,[33] oxalic aciduria,[33] alkaptonuria,[30] multiple acyl CoA dehydrogenase deficiency (glutaric aciduria type II),[34] methylmalonic aciduria,[35] propionic aciduria,[35] porphyria,[33] 5-oxoprolinuria,[33,36,37] homocysteinemia,[38] trimethylaminuria,[39] and Fanconi syndrome.[33]

More recently, both ^{1}H- and ^{13}C-MRS have been used to identify and quantify the presence of elevated levels of galactonate in the urine of 27 patients with uridyltransferase-deficient galactosemia.[40] 750 MHz ^{1}H-NMR spectroscopy has also been used recently to examine the elevated levels of 2-hydroxyglutaric acid in the urine of a patient with the rare condition of 2-hydroxyglutaric aciduria.[41] Maschke et al.[42] reported the use of ^{1}H-MRS (300 MHz) for the analysis of trimethylamine (TMA) in the urine of a patient with trimethylaminuria and of other members of his family. In this study, 300 MHz ^{1}H-NMR spectroscopy was shown to have sufficient sensitivity, resolution, and linearity to allow the diagnosis of trimethylaminuria in a classical one-dimensional spectrum of the urine of the patient (Figure 4). Moreover, this technique is easier than the established classical biochemical methods that primarily use gas chromatography. The increased TMA signal, and the decreased trimethylamine-*N*-oxide (TMAO), as compared to a normal healthy subject, are clearly seen at 2.90 ppm and 3.27 ppm, respectively.

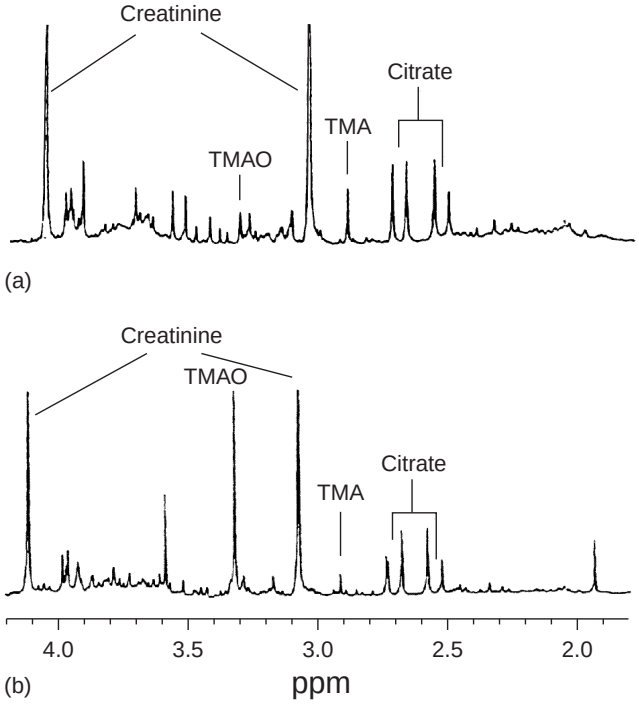

Figure 4 ^{1}H-NMR spectra (300 MHz) of urine samples: (a) a patient with trimethylaminuria; (b) a healthy subject. (Reprinted from Maschke, Wahl, Azaroual, Boulet, Crunelle, Jonbenotte, Foulard, Vermeersch and Lhermitte, '^{1}H-NMR Analysis of TMA in Urine for the Diagnosis of Fish-odour Syndrome', *Clin. Chim. Acta*, **263**, 139–146, Copyright (1997), with permission from Elsevier Science.)

Although urine appears to be the specimen of choice for the diagnosis of inborn errors of metabolism, the use of CSF has also been reported.[43,44] Derangements of organic acid concentrations in CSF may occur independent of the systemic metabolism, and thus, analysis of metabolites in CSF may be important in the diagnosis and follow-up of these patients. Figure 5 shows examples of the diagnostic power of ^{1}H-MRS, for relevant parts of the spectra from CSF samples of patients with isolated 3-methylcrotonyl-CoA-carboxylase deficiency, Canavan disease, and histidinemia.

A significant amount of data in the literature indicates that ^{1}H-MRS of urine, and possibly of CSF, is capable of diagnosing a vast number of inborn errors of metabolism. MRS not only provides information on endogenous biochemical processes, but is also capable of rapid quantitation of metabolites, using a small volume that requires little pretreatment. Furthermore, it is not necessary to preselect the metabolites of interest, as the technique allows simultaneous measurements of a range of components. There exists also, the possibility for the detection of some novel markers of these diseases, or insights into the underlying defects of these disorders.

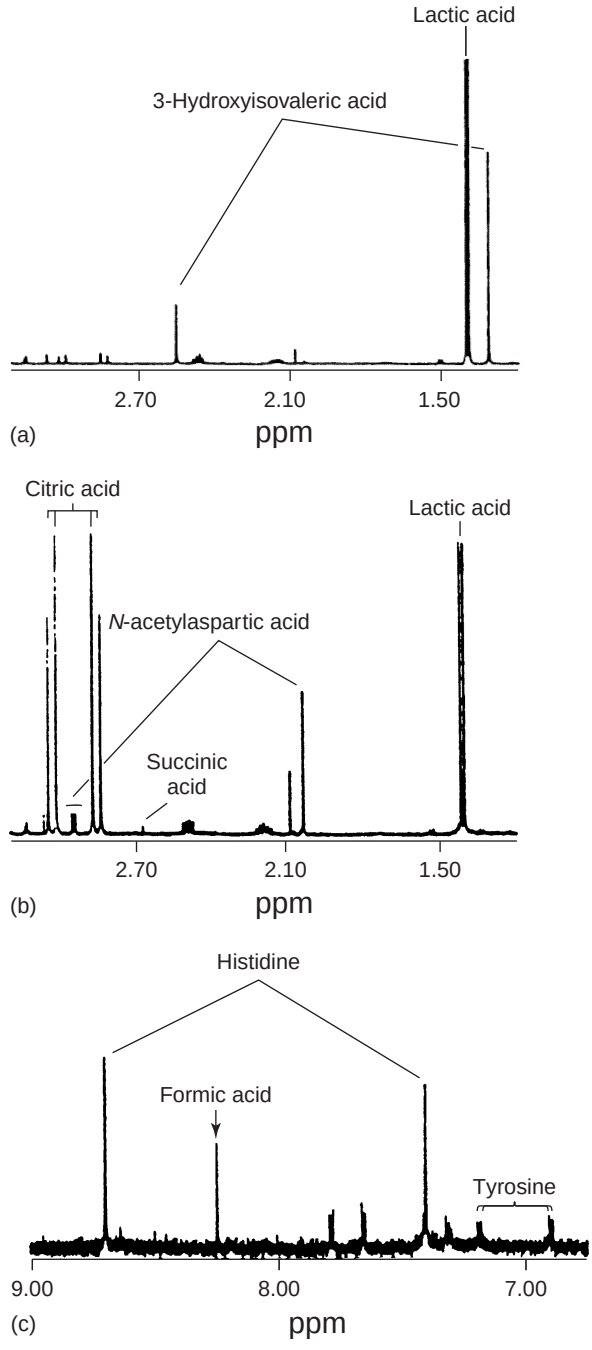

Figure 5 ^{1}H-NMR spectra (600 MHz) of CSF from patients with: (a) isolated 3-methylcrotonyl-CoA-carboxylase deficiency; (b) Canavan disease; and (c) histidinemia. (Reproduced by permission from Wevers et al.[44])

3 MEASUREMENT OF RENAL INJURY

Early studies of ^{1}H-MRS of urine demonstrated that in situations where renal damage is present, the urinary profile of the low-molecular-weight metabolites is altered.[45,46]

Furthermore, in studies of the effects of region-specific nephrotoxins, the application of ^{1}H-MRS urinalysis indicates that abnormal patterns of metabolites are associated with different sites of nephrotoxic actions. Renal proximal tubular toxins cause glycosuria, lactic aciduria, and amino aciduria; renal papillary toxins produce abnormal excretion patterns; and renal papillary necrosis produces an initial increase in TMAO and dimethylamine (DMA) excretion followed by a subsequent increase in *N*,*N*-dimethylglycine, succinate, and acetate, and a decrease in TMAO and 2-oxoglutarate.[47,48] This information is useful when assessing renal disease progression or nephrotoxicity associated with therapeutic agents. Subsequently, ^{1}H-MRS urinalysis has been used as a noninvasive diagnostic technique for tubular and papillary distortions in glomerulonephritis,[49,50] and for the assessment of the degree of chronic renal failure.[51]

The abnormal urinary excretion of specific enzymes and proteins as possible markers of renal allograft rejection has been studied extensively,[52,53] and found to be relatively nonspecific. Foxall et al.[54] used ^{1}H-MRS to investigate the pattern of metabolic changes associated with renal allograft dysfunction. As shown in Figure 6, the spectra of normal human urine showed signals for creatinine, glycine, citrate, alanine, lactate, and *N*-methylated metabolites in the chemical shift range of 3.1–3.3 ppm. The spectra of patients' urine collected following renal transplantation were considerably different. Compared to normal urine, the spectral pattern of urine from patients with an immediate functioning graft, urinary tract infection, renal tubular ischemia or with a nonfunctioning graft are widely different. Further studies associated an increased excretion of TMAO with biopsy confirmed acute graft rejection.[55] Thus, unlike the excretion of enzymes or proteins, ^{1}H-MRS provides both diagnostic and prognostic information.

As urine has been studied most extensively, a large number of the resonances in the spectrum of normal human urine have been assigned. The aim of a recent study was to standardize the analytical conditions, to quantify the major metabolites present in urine of normal subjects, and to evaluate changes due to physiological conditions such as feeding.[56] In this study, ^{1}H-MRS, operating at 300 MHz, studied serial urine samples from 50 normal subjects. In all specimens, creatinine, lactate, alanine, citrate, DMA, TMAO, glycine, and hippurate were found to be present. All metabolites were quantified on the basis of peak heights and were expressed as millimoles per mole of creatinine. The study of metabolic profiles in serial samples allowed evaluation of intra-individual variability and physiological changes due to feeding. The results suggest that every subject is characterized by a typical profile that does not change with time under physiological conditions and is independent of feeding. Thus, a ^{1}H-MRS urinary profile may be considered as an individual's metabolic fingerprint. Changes in the ^{1}H-MRS urinary profile of an individual may be indicative of metabolic changes or disturbances in renal function.

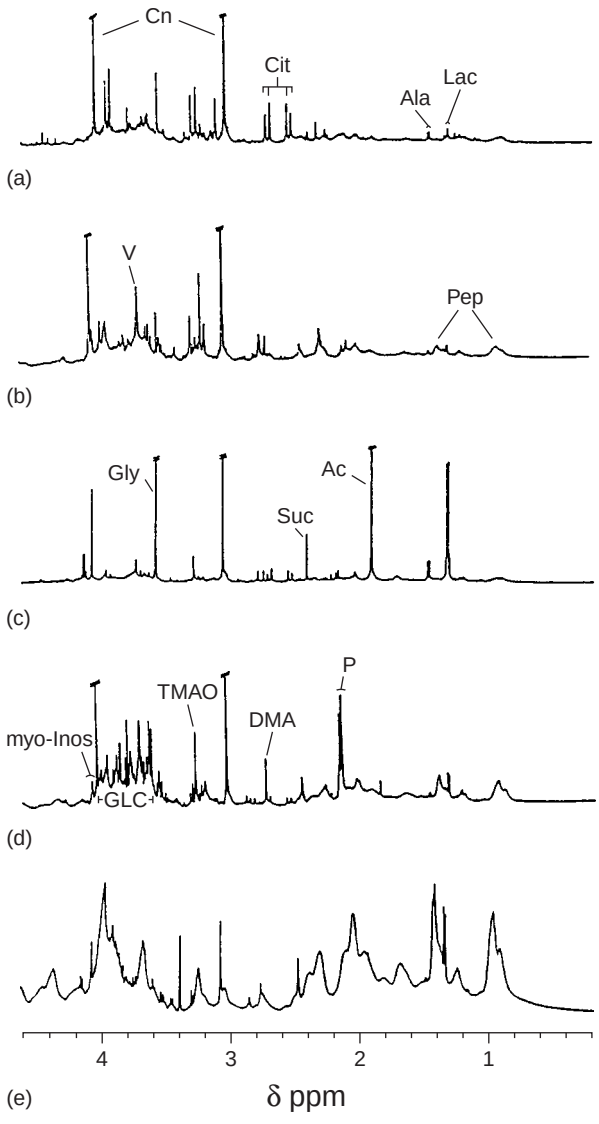

Figure 6 ^{1}H-NMR spectra (500 MHz) of (a) normal human urine, and urine collected from four patients 3 days following renal transplantation showing: (b) immediate functioning graft; (c) urinary tract infection; (d) renal tubular ischemia; and (e) non-functioning graft. Abbreviations are as follows: Ac = acetate, Ala = alanine, Ch = choline, Cit = citrate, DMA, GLC = glucose, Gly = glycine, Lac = lactate, myo-Inos = myo-inositol, Pep = peptides, P = acetaminophen (paracetamol) metabolites, Suc = succinate, TMAO, V = cyclosporin A drug metabolite. (Reprinted by permission of Blackwell Science, Inc., from Foxall et al.[54])

4 DIAGNOSIS OF NEUROLOGICAL DISORDERS

The use of high-resolution ^{1}H-MRS to evaluate brain metabolism through the metabolic profile of CSF constitutes a potentially powerful strategy to aid the differential diagnosis of neurological diseases. Although the biochemical composition of CSF has been well characterized by standard biochemical techniques, a number of studies have been done to evaluate the effectiveness of ^{1}H-NMR spectroscopy for examination of CSF as an aid to the biochemical diagnosis of central-nervous-system diseases. Early studies reported the assignments of a number of resonances.[13,43] Since then, a number of pattern-recognition approaches and discriminant analyses have been used to separate samples into different classes, and these used to differentiate between normal controls and subjects with various neurological disorders.[57–59]

Koschorek et al.[58] reported that the spectra of CSF of normal controls and subjects with tumors or multiple sclerosis (MS) can be perfectly separated, whereas those from subjects with disk herniations can be separated approximately 90% of the time using principal component analysis (PCA).

500 MHz ^{1}H-NMR spectra of postmortem CSF specimens from control subjects and patients with Alzheimer's disease were examined in a similar study.[60] PCA achieved partial separation of the groups; more formal statistical analysis suggested that citrate by itself was the best discriminator.

Spectra showing the metabolic brain profile induced by degenerative dementia are reported to be considerably different from those of control patients.[61] An increase in many metabolites has been observed, and for some (leucine–isoleucine) the concentration is increased threefold. Furthermore, some abnormal and unassigned resonances have been observed. These qualitative MRS abnormalities were not found to be correlated with any medical treatments.

In the CSF of patients with Huntington's chorea, 400 MHz ^{1}H-MRS analysis demonstrated a significant increase (approximately 60%) in the pyruvate concentration as compared to controls. However, in serum samples, no variation in this metabolite was detected.[61]

In studies of CSF from patients with MS, the findings among studies differ. Lynch et al.[62] reported that there were no significant differences between the levels of most metabolites, with the exception of acetate and formate which were increased and decreased respectively in patients with MS as compared to controls. Moreover, in 93% of patients with actively progressing MS, an unknown singlet peak at 2.82 ppm was found; it was presumed to be an *N*-methyl compound. Nicoli et al.[61] similarly reported that MS induced only slight modifications in the ^{1}H-MRS spectra, but these modifications differ from those reported earlier: an increase in lactate and fructose, and a decrease in creatinine and phenylalanine concentrations. Further, the unknown *N*-methyl compound described previously was not found in the CSF of any of the patients with MS. Interestingly, Aasly et al.[63] observed significantly lower levels of lactate and glutamine in MS patients as compared to controls. When the MS group was divided into two subgroups (chronic progressive and relapsing–remitting forms of MS), no significant differences were found in any of the parameters measured in the CSF, although there was a tendency toward lower levels of lactate and glutamine in the relapsing–remitting groups. What accounts for the different results in these studies is unclear; it may simply reflect the lack of a standardized method for specimen preparation and recording ^{1}H-MRS spectra from CSF. A standardized method for high-resolution ^{1}H-MRS of CSF needs to be adopted.[44]

5 CANCER DIAGNOSIS

Increased research in the field of cancer diagnosis and treatment over the past several years included a search for an inexpensive, accurate, and noninvasive screening test for early malignancy. It had been hoped that a simple screening test of plasma, using ^{1}H-MRS, might reveal a malignant condition. However, these investigations have been unsuccessful,[64,65] and attention has shifted toward the assessment of various tissue specimens, and extracts of tissue specimens, to determine whether ^{1}H-MRS can differentiate between malignant and nonmalignant diseases.

5.1 Colon Cancer

Colon cancer was the first disease to be approached in detail by ^{1}H-NMR.[66] A large number of spectra were obtained from colon tumors, and from tissue at the surgical margins of the removed colonic region to serve as normals. A very wide variety of spectra was obtained, and the longitudinal relaxation time of the lipid resonance at 1.3 ppm in the ^{1}H spectrum was used as an indicator of metastatic potential, whereas the intensities of resonances due to choline-containing compounds correlated with the degree of invasion of the cancer (Dukes A, B, C, or D). Subsequent studies on cultured human cell lines of different degrees of invasive capability showed that two-dimensional COSY ^{1}H spectra could be discriminatory.[67] Figure 7 shows the COSY spectra of these cell lines, with a cross-peak between 1.3 and 4.2 ppm whose intensity increases with the aggressivity of the cancer. It was postulated that this cross-peak is due to the sugar fucose, located in surface

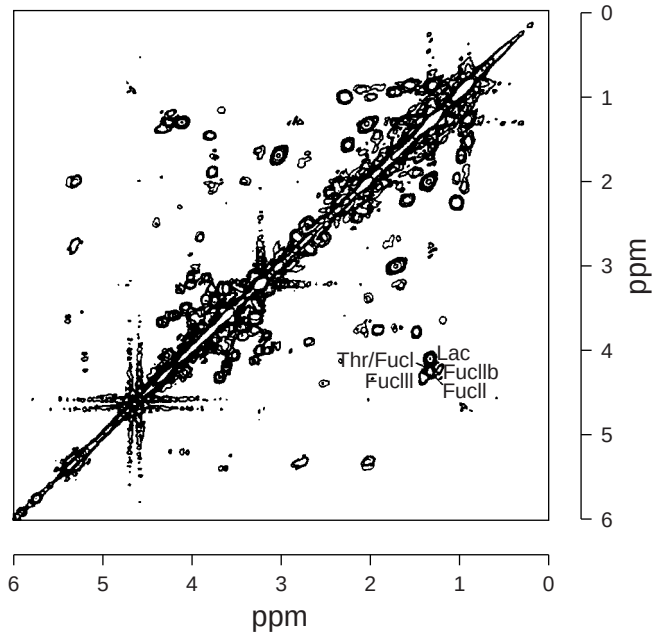

Figure 7 ^{1}H-NMR spectrum (360 MHz) COSY spectrum (magnitude mode, symmetrized) of excised tissue in 400 µL phosphate-buffered saline in D_2O, for a specimen of human ovarian serous carcinoma, poorly differentiated. Multiple cross-peaks are attributable to cell–surface fucosylation, namely Thr/Fuc I (1.33–4.27 ppm), which also contains a contribution from the amino acid threonine, Fuc II (1.25–4.28 ppm) Fuc IIb (1.19–4.20 ppm), and Fuc III (1.41–4.30 ppm), and correlated with tumor grade and loss of cellular differentiation. (Reproduced by permission from Blandford and Smith.[8])

antigens that interfere with the normal immunochemical recognition of cancer cells.

More recent studies of colon tissue revealed another critical element – the purity of the colon tissue. If care is not taken to exclude both muscle and subepithelial fat from the specimen, resonances from these underlying tissues may confound the analysis. A definitive study was performed using well-defined specimens, and rigorous analytical methods, showing that ^{1}H-NMR can yield very high levels of sensitivity and specificity for the detection of cancer.[22] More recently, this approach has been applied to the evaluation of other colonic disorders, such as Crohn's disease and inflammatory bowel disease.[68] The chemical discriminatory power of ^{1}H-NMR is expected to provide valuable insight into the progression of colon disease from dysplasia to malignancy, and to interpret the progression in terms of the oncogenic genes expressed during the progression.

5.2 Thyroid Cancer

A problem in thyroid cancer is the detection of follicular malignancy. Regular histopathological procedures cannot yield a conclusion. Only thyroidectomy, and analysis of the entire gland, can yield a definitive diagnosis of invasive disease. However, the patient has already lost the thyroid gland. ^{1}H-NMR studies on thyroid biopsies by both surgical and needle approaches demonstrated a high level of sensitivity to thyroid cancer, and provided a serious test of the multivariate methods for spectral

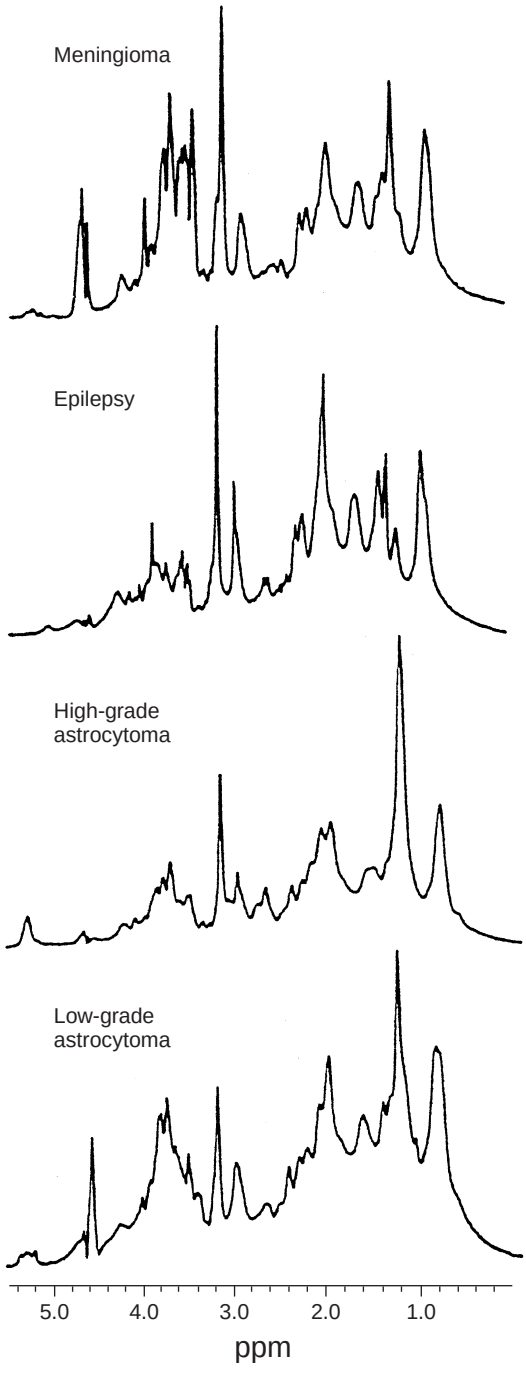

Figure 8 ^{1}H-NMR spectra (360 MHz) of biopsy specimens of human brain tumors, astrocytoma and the more aggressive glioblastoma. The large peak at 4.8 ppm is due to residual H_2O.

analysis.[28] This approach is now used in Australia to determine the therapy applied to neoplasms of the thyroid (C.E. Mountford, personal communication).

5.3 Brain Tumors

A variety of human brain neoplasms have been studied by this technique. Figure 8 shows the ^{1}H-NMR spectra of tissue from three different brain tumor types, and a patient with epilepsy. Relatively simple data analysis allowed accurate classification of these cancers, and gave cause for hope for a method to determine their degree of aggressivity.[29,69] The resonance at 1.3 ppm due to lipid was seen to be indicative of the degree of necrosis (dead tissue) in the tumor.[70] In vivo studies of brain tumors have now been reported,[71] and it is expected that ^{1}H-NMR will be a valuable technique for the planning of treatment for brain tumors.

5.4 Ovarian Tumors

^{1}H-NMR spectra of ovary biopsy specimens have shown good sensitivity to the presence of malignant disease by both resonance ratio[72] and multivariate methods.[73] In the latter study, sensitivity of the tumor to drug therapy was also detected by the method. It is hoped that in vivo use of this method will help detect ovarian cancer at an early stage, and define its potential resistance to therapy.

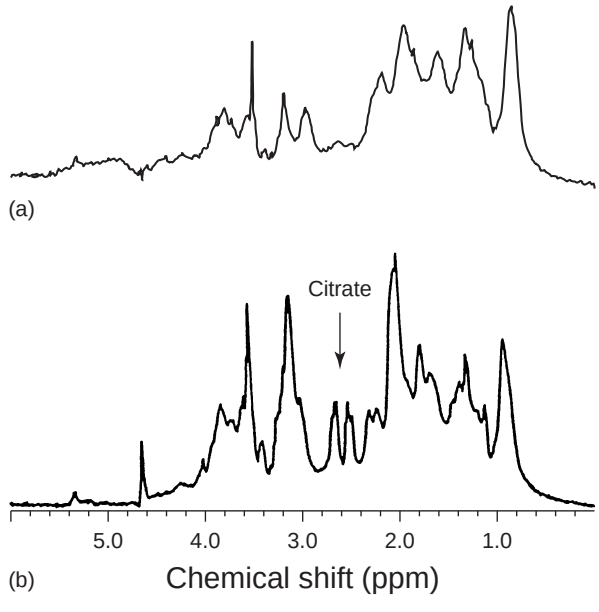

Figure 9 ^{1}H-NMR spectra of biopsy specimens from human prostate gland suffering from benign prostatic hyperplasia. The compositions of the tissue specimens were determined histologically to be (a) 95% stromal, 5% glandular, and (b) 90% glandular, 10% stromal. (Reproduced by permission from Hahn et al.[74])

5.5 Prostate Cancer

Screening of prostate tissue is currently done by transurethral or transcolonic biopsy. In both methods, the sampling is not representative of the entire gland. ^{1}H-NMR of prostate tissue has demonstrated remarkable discriminatory power.[74] Benign disease can be distinguished from cancer with sensitivity of 100%, and specificity of 95.5%. Furthermore, the method appears to be capable of distinguishing the stromal and glandular forms of benign prostatic hypertrophy (Figure 9). Recent transrectal ^{1}H-NMR spectra in vivo show spectra with similar detail, which promise to yield an accurate whole-gland diagnosis.[75]

5.6 Other Neoplastic Conditions

Studies are now under way on a variety of other tissues – cervix, esophagus, breast, kidney, and neck. Results are preliminary, but show a high degree of promise.

6 CONCLUSIONS

The range of applications of ^{1}H-NMR of biological fluids is wide, and is growing steadily. The study of the composition of physiological fluids and tissues, and the changes thereof in disease, are amenable to study by MRS. Although a greater availability of instruments, standardized methodology, a larger database of spectral changes correlated with pathological conditions, and more NMR-trained individuals in the clinical environment are necessary for routine use, the discriminating power of MRS can be exploited to attack difficult clinical problems. It can be used to determine the chemical components that characterize a particular disease.

The ex vivo ^{1}H-NMR of tissue technique has also enjoyed a remarkable degree of success in distinguishing cancer tissue from normal or otherwise abnormal tissue. The methods thus developed should lead to extension of the studies to human subjects in vivo with a high degree of accuracy. These future studies will be, of course, more time-consuming but it is expected that the savings in lives and treatment cost will readily justify the effort. Now is the time to proceed!

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Convertor
COSY	Correlated Spectroscopy
CSF	Cerebrospinal Fluid
DMA	Dimethylamine

MRS	Magnetic Resonance Spectroscopy
MS	Multiple Sclerosis
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
RF	Radiofrequency
S/N	Signal-to-noise
TMA	Trimethylamine
TMAO	Trimethylamine-*N*-oxide

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Multinuclear Magnetic Resonance Spectroscopic Imaging

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Food **(Volume 5)**
Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Imaging of Polymers • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Steel and Related Materials **(Volume 10)**
Nuclear Magnetic Resonance in Metals Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Nuclear Magnetic Resonance of Geological Materials and Glasses • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. S. Braun, H.O. Kalinowski, S. Berger, *100 and More Basic NMR Experiments*, VCH, Weinheim, 1996.
2. F.A. Bovey, *Nuclear Magnetic Resonance Spectroscopy*, Academic Press, New York, 1988.
3. R.S. Macomber, *NMR Spectroscopy: Essential Theory and Practice*, Harcourt Brace Jovanovich, Orlando, FL, 1988.
4. D.I. Hoult, R.E. Richards, in *Magnetic Resonance in Perspective*, ed. W.S. Brey, Academic Press, New York, 97–112, 1996.
5. I.K. O'Neill, C.P. Richards, 'Biological Phosphorus-31 NMR Spectroscopy', *Ann. Rep. NMR Spectrosc.*, **10A**, 133–236 (1980).
6. J.J.H. Ackerman, T.H. Grove, G.C. Wong, D.G. Gadian, G.K. Radda, 'Mapping of Metabolites in Whole Animals by ^{31}P-NMR Using Surface Coils', *Nature*, **283**, 167–170 (1980).
7. K. Brière, I.C.P. Smith, 'Magnetic Resonance Imaging and Spectroscopy: Chemistry In Vivo', in *Topics in Pharmaceutical Sciences: Proceedings of the 53rd International Congress of Pharmaceutical Sciences of F.I.P.*, September, 1993, eds. D.J.A. Crommelin, K.K. Midha, T. Nagai, Medpharm Scientific Publishers, Stuttgart, 263–277, 1994.
8. D. Blandford, I.C.P. Smith, 'Magnetic Resonance Spectroscopy in Clinical Chemistry: The Present and the Future', *J. Int. Fed. Clin. Chem.*, **10**, 6–13 (1998).
9. I.C.P. Smith, C.E. Mountford, 'Tissue NMR Ex Vivo', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 4776–4780, Vol. 8, 1995.
10. J.R. Bales, D.P. Higham, I. Howe, J.K. Nicholson, P.J. Sadler, 'Use of High-resolution Proton Nuclear Magnetic Resonance Spectroscopy for Rapid Multicomponent Analysis of Urine', *Clin. Chem.*, **30**, 426–432 (1984).
11. R.A. Iles, A.J. Hind, R.A. Chalmers, 'Use of Proton Nuclear Magnetic Resonance Spectroscopy in Detection and Study of Organic Acidurias', *Clin. Chem.*, **31**, 1795–1801 (1985).
12. M. Traube, J.L. Bock, J.L. Boyer, 'D-lactic Acidosis After Jejunoileal Bypass: Identification of Organic Anions By

Nuclear Magnetic Resonance Spectroscopy', *Ann. Intern. Med.*, **98**, 171–173 (1983).
13. B.C. Sweatman, R.D. Farrant, E. Holmes, F.Y. Ghauri, J.K. Nicholson, J.C. Lindon, '^{1}H-NMR Spectroscopy of Human Cerebrospinal Fluid: Effects of Sample Manipulation and Assignment of Resonances', *J. Pharm. Biomed. Anal.*, **11**, 651–664 (1993).
14. J.L. Bock, 'Metabolic Profiling of Amniotic Fluid by Proton Nuclear Magnetic Resonance Spectroscopy: Correlation with Fetal Maturation and Other Clinical Variables', *Clin. Chem.*, **40**, 56–61 (1994).
15. P.E. McGowan, J. Reglinski, R. Wilson, J.J. Walker, S. Wisdom, J.H. McKillop, 'Quantitative ^{1}H-NMR Analysis of Amniotic Fluid', *J. Pharm. Biomed. Anal.*, **11**, 629–632 (1993).
16. J.M. Duffy, J. Grimshaw, D.J.S. Guthrie, G.M. McNally, R.A.B. Mollan, P.L. Spedding, J. Trocha-Grimshaw, B. Walker, E. Walsh, '^{1}H-Nuclear Magnetic Resonance Studies of Human Synovial Fluid in Arthritic Disease States as an Aid to Confirming Metabolic Activity in the Synovial Cavity', *Clin. Sci.*, **85**, 343–351 (1993).
17. K. Albert, S. Michele, U. Günther, M. Fial, H. Gall, J. Saal, '^{13}C-NMR Investigation of Synovial Fluids', *Magn. Reson. Med.*, **30**, 236–240 (1993).
18. S. Hamamah, F. Seguin, C. Barthelemy, S. Akoka, A. Le-Pape, J. Lansac, D. Royere, '^{1}H-Nuclear Magnetic Resonance Studies of Seminal Plasma from Fertile and Infertile Men', *J. Reprod. Fertil.*, **97**, 51–55 (1993).
19. M.J. Lynch, J. Masters, J.P. Pryor, J.C. Lindon, M. Spraul, P.J.D. Foxall, J.K. Nicholson, 'Ultra High-field NMR Spectroscopic Studies on Human Seminal Fluid, Seminal Vesicle and Prostatic Secretions', *J. Pharm. Biomed. Anal.*, **12**, 5–19 (1994).
20. M. Spraul, M. Hoffman, J.C. Lindon, R.D. Farrant, M.J. Seddon, J.K. Nicholson, I.D. Wilson, 'Evaluation of Liquid Chromatography Coupled with High-field ^{1}H-NMR Spectroscopy for Drug Metabolite Detection and Characterization: The Identification of Paracetamol Metabolites in Urine and Bile', *NMR Biomed.*, **7**, 295–303 (1994).
21. T. Nishijima, M. Nishina, K. Fujiwara, 'Measurement of Lactate Levels in Serum and Bile Using Proton Nuclear Magnetic Resonance in Patients with Hepatobiliary Diseases: Its Utility in Detection of Malignancies', *Jpn. J. Clin. Oncol.*, **27**, 13–17 (1997).
22. T. Bezabeh, I.C.P. Smith, E. Krupnik, R.L. Somorjai, D.G. Kitchen, C.N. Bernstein, N.M. Pettigrew, R.P. Bird, K.J. Lewin, K.M. Brière, 'Diagnostic Potential for Cancer via Magnetic Resonance Spectroscopy of Colon Tissue', *Anticancer Res.*, **16**, 1553–1558 (1996).
23. S.E. Davies, J. Newcombe, S.R. Williams, W.I. McDonald, J.B. Clark, 'High-resolution Proton NMR Spectroscopy of Multiple Sclerosis Lesions', *J. Neurochem.*, **64**, 742–748 (1995).
24. I.C.P. Smith, D.E. Blandford, 'Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **67**, 509R–518R (1995).
25. O. Hendriksen, 'MR Spectroscopy in Clinical Research', *Acta Radiol.*, **35**, 96–116 (1994).
26. C.E. Mountford, G.L. May, P.G. Williams, M.H.N. Tattersall, H.N. Russell, J.K. Saunders, K.T. Holmes, R.M. Fox, J.R. Barr, I.C.P. Smith, 'Classification of Human Tumors by High-resolution Magnetic Resonance Spectroscopy', *Lancet*, March 22, 651–653 (1986).
27. A.C. Kuesel, T. Kroft, J.K. Saunders, M. Préfontaine, N. Mikhael, I.C.P. Smith, 'A Simple Procedure for Obtaining High-quality NMR Spectra of Semiquantitative Value from Small Tissue Specimens: Cervical Biopsies', *Magn. Reson. Med.*, **27**, 349–355 (1992).
28. R.L. Somorjai, A.E. Nikulin, N. Pizzi, D. Jackson, G. Scarth, B. Dolenko, H. Gordon, P. Russell, C.L. Lean, L. Delbridge, C.E. Mountford, I.C.P. Smith, 'Computerized Consensus Diagnosis: A Classification Strategy for the Robust Analysis of MR Spectra. 1. Application to ^{1}H Spectra of Thyroid Neoplasms', *Magn. Reson. Med.*, **33**, 257–263 (1995).
29. R.L. Somorjai, B. Dolenko, A.K. Nikulin, N. Pizzi, G. Scarth, P. Zhilkin, W. Halliday, D. Fewer, N. Hill, I. Ross, M. West, I.C.P. Smith, S.M. Donnelly, A.C. Kuesel, K.M. Brière, 'Classification of ^{1}H-MR Spectra of Human Brain Neoplasms: The Influence of Preprocessing and Computerized Consensus Diagnosis on Classification Accuracy', *J. Magn. Reson. Imaging*, **6**, 437–444 (1996).
30. S. Yamaguchi, 'The Future in Pediatric Clinical Chemistry: An Application of NMR Spectrometry to Analysis of Biological Fluids', *Prog. Clin. Biochem.*, **991**, 853–856 (1992).
31. S.P. Burns, D.A. Woolf, J.V. Leonard, R.A. Iles, 'Investigation of Urea Cycle Enzyme Disorders by ^{1}H-NMR Spectroscopy', *Clin. Chim. Acta*, **209**, 47–60 (1992).
32. S.P. Burns, R.A. Iles, 'An Investigation of Argininosuccinic Acid Anhydrides in Argininosuccinic Acid Lyase Deficiency by ^{1}H-NMR Spectroscopy', *Clin. Chim. Acta*, **221**, 1–13 (1993).
33. E. Holmes, P.J.D. Foxall, J.K. Nicholson, G.H. Neild, S.M. Brown, C.R. Beddell, B.C. Sweatman, E. Rahr, J.C. Lindon, M. Spraul, P. Neidig, 'Automatic Data Reduction and Pattern Recognition Methods for Analysis of ^{1}H Nuclear Magnetic Resonance Spectra of Human Urine from Normal and Pathological States', *Anal. Biochem.*, **220**, 284–296 (1994).
34. S. Burns, A. Johnson, R.A. Iles, 'Indirect Measurement of Betaine Homocysteine Methyltransferase Activity Using ^{1}H-NMR Spectroscopy', *Biochem. Soc. Trans.*, **19**, 402S (1991).
35. S. Kodama, M. Sugiura, H. Nakao, K. Kobayashi, M. Miyoshi, K. Yoshii, M. Komatsu, T. Sakurai, '^{1}H-NMR Studies of Urine in Propionic Acidemia and Methylmalonic Acidemia', *Acta Paediatr. Jpn.*, **33**, 139–145 (1991).

36. U. Engelke, A. Heerschap, R. Wevers, 'Screening for Inborn Errors of Metabolism Using ^{1}H-NMR Spectroscopy of Deproteinized Human Blood Plasma', *Proc. Soc. Magn. Reson. Med.*, **3**, 1163 (1993).
37. F.Y.K. Ghauri, H.G. Parkes, J.K. Nicholson, I.D. Wilson, D.P. Brenton, 'Asymptomatic 5-oxyprolinuria Detected by Proton Magnetic Resonance Spectroscopy', *Clin. Chem.*, **39**, 1341 (1993).
38. P. Lundberg, N.P.B. Dudman, P.W. Kuchel, D.E.L. Wilcken, '^{1}H-NMR Determination of Urinary Betaine in Patients with Premature Vascular Disease and Mild Homocysteinemia', *Clin. Chem.*, **4**, 275–283 (1995).
39. N.G.G.M. Abeling, A.H. VanGennip, H.D. Bakker, A. Heerschap, U. Engelke, R.A. Wevers, 'Diagnosis of a New Case of Trimethylaminuria Using Direct Proton NMR Spectroscopy of Urine', *J. Inherit. Metab. Dis.*, **18**, 182–184 (1995).
40. S.L. Wehrli, G.T. Berry, M. Palmieri, A. Mazur, L. Elsas, S. Segal, 'Urinary Galactonate in Patients with Galactosemia: Quantitation by Nuclear Magnetic Resonance Spectroscopy', *Pediatr. Res.*, **42**, 855–861 (1997).
41. E. Holmes, P.J.D. Foxall, M. Spraul, R.D. Farrant, J.K. Nicholson, J.C. Lindon, '750 MHz ^{1}H-NMR Spectroscopy Characterization of the Complex Metabolic Pattern of Urine From Patients With Inborn Errors of Metabolism: 2-Hydroxyglutaric Aciduria and Maple Syrup Urine Disease', *J. Pharm. Biomed. Anal.*, **15**, 1647–1659 (1997).
42. S. Maschke, A. Wahl, N. Azaroual, O. Boulet, V. Crunelle, M. Imbenotte, M. Foulard, G. Vermeersch, M. Lhermitte, '^{1}H-NMR Analysis of Trimethylamine in Urine for the Diagnosis of Fish-odour Syndrome', *Clin. Chim. Acta*, **263**, 139–146 (1997).
43. O.A.C. Petroff, R.K. Yu, T. Ogino, 'High-resolution Proton Magnetic Resonance Analysis of Human Cerebrospinal Fluid', *J. Neurochem.*, **47**, 1270–1276 (1986).
44. R.A. Wevers, U. Engelke, U. Wendel, J.G.N. de Jong, F.J.M. Gabreëls, A. Heerschap, 'Standardized Method for High-resolution ^{1}H-NMR of Cerebrospinal Fluid', *Clin. Chem.*, **41**, 744–751 (1995).
45. J.K. Nicholson, J.A. Timbrell, P.J. Sadler, 'Proton NMR Spectra as Indicators of Renal Damage', *Mol. Pharmacol.*, **27**, 644–651 (1984).
46. K.P.R. Gartland, J.C. Lindon, C.R. Beddell, J.K. Nicholson, 'Application of Pattern Recognition Methods to the Analysis of Toxicological Data Generated by NMR Spectroscopy', *Mol. Pharmacol.*, **39**, 629–642 (1991).
47. K.P.R. Gartland, R.S. Bonner, J.K. Nicholson, 'Investigations into the Biochemical Effects of Region-specific Nephrotoxins', *Mol. Pharmacol.*, **35**, 242–250 (1989).
48. E. Holmes, J.K. Nicholson, F.W. Bonner, B.C. Sweatman, C.R. Beddell, J.C. Lindon, E. Rahr, 'Mapping the Biochemical Trajectory of Nephrotoxicity by Pattern Recognition of NMR Urinalysis', *NMR Biomed.*, **5**, 368–372 (1992).
49. T.A. Lundina, T.L. Knubovets, K.R. Sedov, S.A. Markova, L.A. Sibeldina, 'Variability of Kidney Tubular Interstitial Distortions in Glomerulonephritis as Measured by ^{1}H-NMR Urinalysis', *Clin. Chim. Acta*, **214**, 165–173 (1993).
50. T.L. Knubovets, T.A. Lundina, L.A. Sibeldina, K.R. Sedov, '^{1}H NMR Analysis in Glomerulonephritis: A New Prognostic Criterion', *Magn. Reson. Imaging*, **10**, 127–134 (1992).
51. J.D. Bell, J.A. Lee, H.A. Lee, P.J. Sadler, D.R. Wilkie, R.H. Woodham, 'Nuclear Magnetic Resonance Studies of Blood Plasma and Urine from Subjects with Chronic Renal Failure: Identification of Trimethylamine-*N*-oxide', *Biochim. Biophys. Acta*, **1096**, 101–107 (1991).
52. C.T. Yuen, C.R.R. Corbett, P.R.N. Kind, 'Isoenzymes of Urinary *N*-acetyl-B-D-Glucosaminidase (NAG) in Patients with Renal Transplants', *Clin. Chim. Acta*, **164**, 339–350 (1987).
53. P.W. Mueller, V. Delaney, M.L. Macneil, S.P. Caudill, K.K. Steinberg, 'Indicators of Acute Renal Transplant Rejection in Patients Treated with Cyclosporin', *Clin. Chem.*, **36**, 759–764 (1990).
54. P.J.D. Foxall, G.J. Mellotte, M.R. Bending, J.C. Lindon, J.K. Nicholson, 'NMR Spectroscopy as a Novel Approach to the Monitoring of Renal Transplant Function', *Kidney Int.*, **43**, 234–245 (1993).
55. L. Le Moyec, A. Pruna, M. Eugène, J. Bedrossian, J.M. Idatte, J.F. Huneau, D. Tomé, 'Proton Nuclear Magnetic Resonance Spectroscopy of Urine and Plasma in Renal Transplantation Follow-up', *Nephron*, **65**, 433–439 (1993).
56. C. Zuppi, I. Messana, F. Forni, C. Rossi, L. Pennacchietti, F. Ferrari, B. Giardina, '^{1}H-NMR Spectra of Normal Urines: Reference Ranges of the Major Metabolites', *Clin. Chim. Acta*, **265**, 85–97 (1997).
57. A. Hiraoka, I. Miura, M. Hattori, I. Tominaga, K. Kushida, M. Maeda, 'Proton Magnetic Resonance Spectroscopy of Cerebrospinal Fluid as an Aid in Neuroradiolical Diagnosis', *Biol. Pharm. Bull.*, **17**, 1–4 (1994).
58. F. Koschorek, W. Offermann, J. Stelten, W.E. Braunsdorf, U. Steller, H. Gremmel, D. Leibfritz, 'High-resolution 1H-NMR Spectroscopy of Cerebrospinal Fluid in Spinal Diseases', *Neurosurg. Rev.*, **16**, 307–315 (1993).
59. R.D. Farrant, J.C. Lindon, E. Rahr, B.C. Sweatman, 'An Automatic Data Reduction and Transfer Method to Aid Pattern Recognition Analysis and Classification of NMR Spectra', *J. Pharm. Biomed. Anal.*, **10**, 141–144 (1992).
60. F.Y.K. Ghauri, J.K. Nicholson, B.C. Sweatman, J. Wood, C.R. Beddell, J.C. Lindon, N.J. Cairns, 'NMR Spectroscopy of Human Post Mortem Cerebrospinal Fluid: Distinction of Alzheimer's Disease from Control Using Pattern Recognition and Statistics', *NMR Biomed.*, **6**, 163–167 (1993).
61. F. Nicoli, J. Vion-Dury, S. Confort-Gouny, S. Maillet, J.-L. Gastaut, P.J. Cozzone, 'Cerebrospinal Fluid Metabolic Profiles in Multiple Sclerosis and Degenerative Dementias Obtained by High-resolution Proton Magnetic Resonance Spectroscopy', *C.R. Acad. Sci. 3*, **319**, 623–631 (1996).

62. J. Lynch, J. Peeling, A. Auty, G.R. Sutherland, 'Nuclear Magnetic Resonance Study of Cerebrospinal Fluid from Patients with Multiple Sclerosis', *Can. J. Neurol. Sci.*, **20**, 194–198 (1993).
63. J. Aasly, M. Gårseth, U. Sonnewald, J.-A. Zwart, L.R. White, G. Unsgard, 'Cerebrospinal Fluid Lactate and Glutamine are Reduced in Multiple Sclerosis', *Acta Neurol. Scand.*, **95**, 9–12 (1997).
64. M. Préfontaine, T. Kroft, M. Monck, J.K. Saunders, N.A. Mikhael, I.C.P. Smith, 'Evaluation of a Simple Line Width Test Involving Magnetic Resonance Spectroscopy of Plasma in Carcinoma of the Ovary', *Cancer*, **67**, 406–411 (1991).
65. A. Moreno, E. Escrich, M. Prats, N. Benito, J. Alonso, C. Arus, 'Study of the Ability of Proton Nuclear Magnetic Resonance Spectroscopy of Human Plasma to Differentiate Between Controls and Breast Cancer Patients', *Oncology*, **50**, 110–115 (1993).
66. I.C.P. Smith, E. Princz, J.K. Saunders, 'Magnetic Resonance Spectroscopy in Cancer Research', *J. Can. Assoc. Radiol.*, **41**, 32–38 (1990).
67. C.L. Lean, R.C. Newbold, D.A. Ende, E.K. Bokey, I.C.P. Smith, C.E. Mountford, 'Assessment of Human Colorectal Biopsies by ^{1}H-MRS: Correlation with Histopathology', *Magn. Reson. Med.*, **30**, 525–533 (1993).
68. T. Bezabeh, C.N. Bernstein, R.L. Somorjai, I.C.P. Smith, 'The Use of ^{1}H Magnetic Resonance Spectroscopy in the Study of Inflammatory Bowel Diseases: Distinguishing Ulcerative Colitis from Crohn's Disease', *Proc. Int. Soc. Magn. Reson. Med.*, **2**, 1279 (1997).
69. A. Rutter, H. Hugenholtz, J.K. Saunders, I.C.P. Smith, 'Classification of Brain Tumors by Ex Vivo ^{1}H-NMR Spectroscopy', *J. Neurochem.*, **64**, 1655–1661 (1995).
70. A.C. Kuesel, K.M. Brière, W.C. Halliday, G.R. Sutherland, S.M. Donnelly, I.C.P. Smith, 'Mobile Lipid Accumulation in Necrotic Tissue of High Grade Astrocytomas', *Anticancer Res.*, **16**, 1485–1490 (1996).
71. M.C. Preul, Z. Caramanos, D.L. Collins, J.-G. Villemure, R. LeBlanc, A. Olivier, R. Pokrupa, D.L. Arnold, 'Accurate Noninvasive Diagnosis of Human Brain Tumors by Using Proton Magnetic Resonance Spectroscopy', *Nat. Med.*, **2**, 323–325 (1990).
72. W.B. Mackinnon, P. Russell, G.L. May, C.E. Mountford, 'Characterization of Human Ovarian Epithelial Tumors (Ex Vivo) by Proton Magnetic Resonance Spectroscopy', *Int. J. Gynecol. Cancer*, **5**, 211–221 (1995).
73. J.C. Wallace, G.P. Raaphorst, R.L. Somorjai, C.E. Ng, M. Fung Kee Fung, M. Senterman, I.C.P. Smith, 'Classification of ^{1}H-MR Spectra of Biopsies from Untreated and Recurrent Ovarian Cancer Using Linear Discriminant Analysis', *Magn. Reson. Med.*, **38**, 569–576 (1997).
74. P. Hahn, I.C.P. Smith, L. Leboldus, C. Littman, R.L. Somorjai, T. Bezabeh, 'The Classification of Benign and Malignant Human Prostate Tissue by Multivariate Analysis of ^{1}H-MR Spectra', *Cancer Res.*, **57**, 3398–3401 (1997).
75. J.K. Kurhanowicz, D.B. Vigneron, S.J. Nelson, H. Hricak, J.M. Macdonald, P. Konety, P. Narayan, 'Citrate as an In Vivo Marker to Discriminate Prostate Cancer From Benign Prostatic Hyperplasia and Normal Prostate Peripheral Zone: Detection Via Localized Proton Spectroscopy', *Urology*, **45**, 459–466 (1995).

Magnetic Resonance, General Medical

Patrick Winter and Navin Bansal
University of Pennsylvania, Philadelphia, USA

1 Introduction **1**
1.1 History of Biomedical Nuclear Magnetic ResonanceNMR 1
1.2 Overview of Instrumentation 2
1.3 Advantages 6
1.4 Disadvantages 7
2 Proton Imaging **7**
2.1 Spatial Encoding and Imaging Pulse Sequences 7
2.2 Relaxation and Image Contrast 10
2.3 Image Artifacts 13
2.4 Advanced Imaging Techniques 14
2.5 Safety Issues 17
3 Spectroscopy **19**
3.1 Localization Techniques 19
3.2 ^{31}P Nuclear Magnetic Resonance SpectroscopyNMR 21
3.3 Water Suppression and ^{1}H Spectroscopy 22
3.4 Carbon-13 Spectroscopy and Decoupling 24
3.5 Sodium-23 Spectroscopy 25
3.6 Fluorine-19 Spectroscopy 26
4 Conclusions **27**
Acknowledgments **27**
List of Symbols **27**
Abbreviations and Acronyms **27**
Related Articles **28**
References **28**

Nuclear magnetic resonance (NMR) is the basic physical phenomenon in which certain nuclei absorb and emit radiofrequency (RF) radiation while in the presence of a magnetic field. The noninvasive nature of NMR techniques, including magnetic resonance imaging (MRI) and magnetic resonance spectroscopy (MRS), has allowed explosive developments in biomedical applications since the late 1970s. MRI of protons is now a widely accepted clinical imaging technique, whereas MRS of various nuclei, including ^{31}P, ^{1}H, ^{13}C, ^{23}Na, and ^{19}F, can provide a vast amount of metabolic and physiological information. MRI provides significant advantages over many other imaging techniques in terms of resolution, contrast, and its noninvasive nature. Although not as common in clinical medicine, MRS also provides valuable metabolic information that cannot be obtained by other techniques. This article reviews the fundamentals of spatial encoding, contrast, artifacts, and applications of MRI as well as some recent advances in the field. Some applications of biomedical MRS of various nuclei are also discussed.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © John Wiley & Sons Ltd

The sections "History of Biomedical NMR" and "Overview of Instrumentation" in this article are based in part on the online book, "*The Basics of MRI*", by Joseph P Hornak, http://www.cis.rit.edu/htbooks/mri/.

1 INTRODUCTION

NMR spectroscopy is widely used in chemistry to obtain detailed structural and motional information about complex molecules. Electromagnetic radiation in the frequency range of radio and television transmissions is used, often called RF radiation, which easily penetrates biological tissue without the adverse health effects of ionizing radiation, such as X-ray radiation. This unique feature has allowed explosive developments in noninvasive biomedical applications of NMR techniques since the late 1970s. MRI of protons (^{1}H) is now a widely accepted clinical imaging technique, primarily due to its ability to image soft tissue, such as the white matter and gray matter of the brain. NMR techniques can also provide a vast amount of metabolic and physiological information by observing ^{1}H and other nuclei in the tissue.

This article presents a concise introduction to the field of MRI including the history, instrumentation, advantages, and disadvantages of MRI technologies. The fundamentals of spatial encoding, contrast, artifacts and applications of MRI are reviewed, as well as some recent advances in the field. Also discussed is MRS of various nuclei, including ^{31}P, ^{1}H, ^{13}C, ^{23}Na, and ^{19}F. The intention is to cover the major topics in biomedical NMR and describe the general techniques and applications of MRI and MRS. The related articles listed at the end of this chapter provide more detailed descriptions of specific topics.

1.1 History of Biomedical Nuclear Magnetic ResonanceNMR

Bloch and Purcell, both of whom were awarded the Nobel Prize in 1952, discovered the magnetic resonance

phenomenon in 1946.[1,2] In the period between 1950 and 1970, NMR techniques were developed and used for chemical and physical molecular analysis. In 1973, Lauterbur proposed that NMR could be exploited to image the human body by the back-projection technique used in computed tomography (CT).[3] In 1975, Ernst proposed MRI using phase and frequency encoding, the current MRI technique.[4] Edelstein et al.. demonstrated the feasibility of imaging the body using this technique in 1980.[5] A single image could be acquired in approximately 5 min. During the same time period, the first localized NMR spectra from living animals were obtained.[6–9] By 1996, the image acquisition time had been reduced to a matter of seconds, without sacrificing too much image quality. In 1989, Dumoulin et al.. perfected magnetic resonance angiography, allowing flowing blood to be imaged without the use of contrast agents.[10] In 1989, echo-planar imaging (EPI) was introduced which permitted image acquisition at video rates (30 ms).[11] The EPI method has enabled the technique now referred to as functional brain imaging, which allows the regions of the brain responsible for thought and motor control to be mapped. In 1991, Richard Ernst was awarded the Nobel Prize in Chemistry for his achievements in pulsed NMR and MRI. Biomedical NMR is clearly a young, but growing, science.

1.2 Overview of Instrumentation

A magnetic resonance system used for clinical applications consists of several hardware components working together to form the desired images. The most noticeable component is the magnet (Figure 1), which houses the shim coils, gradient coils and RF coils. Additional hardware includes gradient amplifiers, RF transmitting and receiving electronics, and a host computer for integration and interfacing all the hardware and software components. A simple diagram of the major components

Figure 1 Photograph of a clinical 1.5 T superconducting magnet. (Reproduced by courtesy of Philips Medical Systems, Inc.).

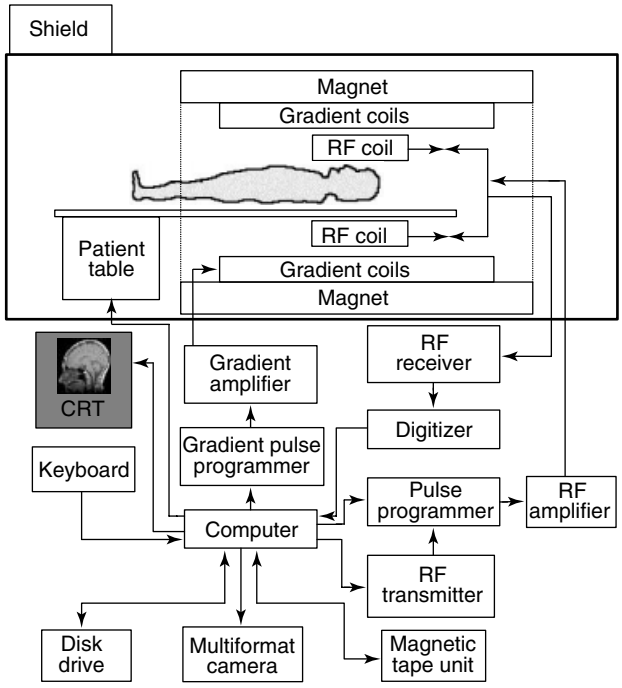

Figure 2 Simplified block diagram of an MRI scanner. The patient is positioned within the magnet, shims, gradients, and RF coil. A computer system controls the RF hardware as well as managing the image data..

of a magnetic resonance imager and a few of the major interconnections are shown in Figure 2. At the top are the components located in the scan room of a magnetic resonance imager. The magnet produces a large static magnetic field, denoted by B_0, for the imaging procedure. Within the bore of the magnet, the gradient coils produce spatial magnetic field gradients in B_0 along the x, y, and z directions. Within the gradient coils is the RF coil, which produces an oscillating magnetic field perpendicular to B_0, denoted B_1. The B_1 field is necessary to rotate the nuclear spins in the body, often by 90° or 180°. The RF coil is also used to detect the signal from the spins within the body. The patient is positioned within the magnet by a computer-controlled table. This table has a positioning accuracy of 1 mm. The scan room itself is RF shielded, usually by copper sheets within the walls to prevent RF emission as well as interference from outside RF signals. A magnetic shield made of thick steel plates also surrounds most scan rooms. This shield restricts the magnetic field from extending out of the scanner room. Self-shielded magnets are also now available, eliminating the need for large steel structures.

The computer on a magnetic resonance system controls all components of the imager. The RF components under computer control are the RF source and pulse programmer. The source produces a sine wave of the desired frequency. The pulse programmer shapes the

RF pulses into the desired form, usually an apodized SINC pulse (i.e. following the mathematical function $\mathrm{SINC}(t) = \sin(t)/t$ with t representing the variable time). The RF amplifier increases the pulse power from milliwatts to kilowatts. The computer also controls the gradient pulse programmer, which sets the shape and amplitude of each of the three gradient fields. The gradient amplifier increases the power of the gradient pulses to a level sufficient to drive the gradient coils. Some imagers also have an array processor, a device that is capable of performing a two-dimensional (2-D) Fourier transform in fractions of a second. The computer off-loads the Fourier transform to this faster device.

The operator of the imager gives input to the computer through a control console. An imaging sequence is selected and customized from the console. The operator can see the images on a video display located on the console or can make hard copies of the images on a film printer.

Further details about the magnet, shim coils, gradient coils, and RF coils on a magnetic resonance imager are discussed below.

1.2.1 Magnet

The magnets used for medical applications have an inside diameter of about 1 m and a horizontal orientation to facilitate moving the patient in and out of the magnet bore. Three types of magnets have been used in MRI, permanent, resistive and superconducting.

Permanent magnets are constructed of permanently magnetized ferromagnetic materials.[12] They are least expensive and have low maintenance costs because they do not require electrical power or expensive cryogens. Permanent magnets can be designed with three open sides, a configuration known as open MRI. This allows greater access and visibility for patients. Permanent magnets, however, are extremely heavy, weighing up to 100 tons and are limited to field strengths around 0.35 T or less. They tend to produce low-quality images and are usually not adequate for spectroscopic applications because of the relatively weak and inhomogeneous magnetic field. The low magnetic field strength makes a permanent magnet less hazardous to patients and staff but, if an accident occurs, a permanent magnet cannot be turned off. A further discussion of the safety issues associated with MRI equipment can be found in section 2.5.

Resistive electromagnets produce a magnetic field by running electrical current through a coil of wire. The magnetic field is oriented perpendicular to the plane of the coil. The power consumption required to achieve the necessary field strengths ($\sim$0.3 T) is similar to that of a large office building.[13] This produces a lot of heat, necessitating a cooling system to protect the patient and the magnet coils. A resistive electromagnet can be turned off when not in use or during an emergency. The high cost of the electricity and water cooling system required to operate a resistive electromagnet has prevented its widespread use in MRI.

Superconducting magnets work on the same principal as electromagnets. However, the coil windings are made of superconducting wire, often alloys of niobium and tin, that loses its electrical resistance as the temperature approaches absolute zero, $-273\,^{\circ}\mathrm{C}$.[12] With no resistance, no energy is lost as current runs through the wire. Therefore, no energy is required to keep the magnet running once the coil has been energized and superconducting magnets do not accrue the high-energy costs of resistive magnets. Although no heat is generated as current runs through the wire, a cryogen cooling system is required to keep the coil near absolute zero. This is achieved by immersing the coil in a bath of liquid helium. The helium chamber is surrounded by vacuum chambers and a liquid nitrogen chamber to reduce evaporation of the liquid helium.[14] In superconducting magnets, replenishing the evaporated cryogens represents the predominate operating cost. Most MRI scanners use a superconducting magnet because they can produce a very strong and uniform magnetic field over a large area. The maximum field strength currently used in clinical imaging is 1.5 T, whereas experimental scanners may exceed 4 T. A superconducting magnet is typically not shut down when not in use and will remain active even during power failures. However, they can quickly be turned off in case of an emergency. This requires specific safety precautions and specialized training of the MRI staff.

1.2.2 Shim Coils

The shim coils are used to optimize the magnetic field homogeneity over the imaging volume.[15] The field may change over time, when a patient enters the magnet or due to changes in position of metallic objects near the scanner room. A magnetic field inhomogeneity of just a few parts per million (ppm) leads to noticeable shading of magnetic resonance images, whereas greater inhomogeneities give rise to spatial distortions in the images. In general, NMR spectroscopy requires even better field homogeneity than does imaging. As many as 30 different shim coils may be used to adjust the field strength by carefully adjusting the amount of current in each coil until acceptable homogeneity is achieved.

1.2.3 Gradient Coils

Gradient coils are used to encode spatial information in the NMR signal during imaging and certain spectroscopy examinations. The coils are arranged along the three orthogonal axes so that the magnetic field can be manipulated in each direction. Gradients along other oblique axes can be implemented with combinations of the orthogonal gradients.[16]

Figure 3 shows the basic gradient coil design for obtaining the three orthogonal gradients. The gradient along the axis of the magnet (z direction) is achieved with a pair of circular coils, shown in Figure 3(b). The currents in the two coils flow in opposite directions creating a magnetic field gradient between them. The magnetic field of one coil adds to the main magnetic field while the field from the other coil subtracts from the main magnetic field. The x and y gradients are created by a pair of figure-of-eight coils, shown in Figure 3(c) and 3(d), respectively. Power is supplied to each of the gradient coils by independent computer-controlled gradient amplifiers. The gradient coils are positioned in the magnet such that the mid-point of each set is at the center of the magnet, denoted as the isocenter. Thus, the magnetic field is not changed at the isocenter when the gradients are switched on and off.

The four most important properties of a gradient system are: (a) the attainable gradient amplitude, (b) gradient linearity, (c) gradient rise time, and (d) eddy currents generated by the gradients. The maximum attainable gradient amplitude (the change in magnetic field per unit distance, often cited in units of gauss cm^{-1}) limits the minimum slice thickness and minimum field of view (FOV) that can be used. Gradient linearity refers to the uniformity of the slope along the gradient axis; nonlinearity yields image distortions. How fast a gradient can be powered from zero to full amplitude is referred to as the rise time, which should be as short as possible. The act of switching gradients on and off induces the formation of undesirable eddy currents in the metallic structures of the magnet. These eddy currents generate magnetic fields of their own, which dissipate at differing rates and can have deleterious effects on image quality. This problem can be overcome by driving the gradient coils with an empirically determined pulse shape, which cancels out eddy current contributions. A potentially more powerful approach is the use of self-shielded gradient coils. These coils are constructed such that the generated magnetic fields are confined to the interior of the coils, thus preventing the formation of eddy currents in the remainder of the magnet.

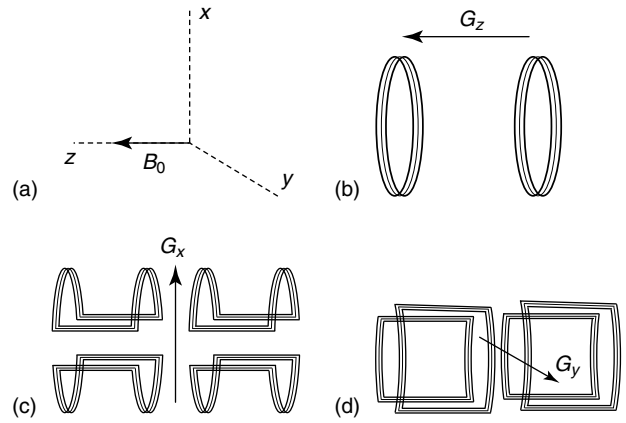

Figure 3 Diagram of three orthogonal gradient coils. (a) Coordinate system. (b) Two circular coils are used to apply a magnetic field gradient along the z-axis. (c) A pair of figure-of-eight coils creates a gradient along the x-axis. (d) A similar pair of figure-of-eight coils is oriented at 90° with respect to the x-axis to create the y-axis gradient..

1.2.4 Radiofrequency CoilsRF

A coil is also needed for transmission of RF radiation to excite the nuclei and to receive the emitted RF energy. As a basic analogy, the RF coil acts like an antenna. The design of coils can vary in size, shape, orientation and characteristics to suit their particular application. Two separate coils may be used for transmission and reception, or a single coil can serve both purposes. To illustrate the dual role of these coils, they are often called transceiver coils.

The RF coil has to be tuned and matched to maximize the efficiency. Like any antenna, the coil is tuned to maximize the sensitivity at the desired frequency and reduce the sensitivity to other frequencies. During excitation, this lowers the power requirements and minimizes patient exposure to unnecessary RF fields. This also maximizes the acquired signal while cutting down on the noise picked up by the coil. In addition, basic RF theory dictates that the maximum power is emitted from an antenna when the impedance of the antenna is matched to the rest of the electronics, typically 50 Ω. Tuning and matching the RF coil optimizes the performance of the MRI scanner producing the highest possible quality of images.

A circuit diagram of a simple coil is shown in Figure 4. The coil of wire is modeled as an inductor, labeled L in the circuit diagram. The capacitor C_t is used to tune the circuit, and the series capacitor C_m matches the circuit to 50 Ω.

The most basic coil, called a surface coil, is a simple loop of wire.[17] Figure 5 is a photograph of a variety of surface coils. They are very easy to build and can be placed directly over the sample for maximum sensitivity. However, the coil does not produce a uniform magnetic field. Only signals very close to the coil itself can be

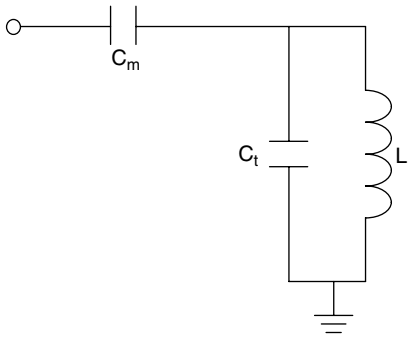

Figure 4 Circuit diagram of a simple RF coil. Capacitor C_t is used to tune the coil to the resonant frequency, and C_m matches the coil to the characteristic impedance of the RF electronics, typically 50 Ω..

Figure 5 Photograph of a variety of surface coils. Different sizes and shapes are used to image different parts of the body. The conductive loop is contained in a rigid plastic housing and the cables connect the coil to the RF electronics in the MRI scanner magnet. (Photo courtesy of Philips Medical Systems, Inc.).

excited and received. Therefore, surface coils are often used to receive signals from a localized area of the body. The coil must be positioned such that the magnetic field created by the coil is at right angles to the main external magnetic field. This makes the placement of a surface coil crucial to optimize the performance.

An advanced surface coil design is known as a phased-array coil. Numerous surface coils can be aligned to produce better field homogeneity along the plane of the structure. The coils can be controlled by switching circuitry[18] or by altering the phase characteristics of each element.[19] This reduces the noise contribution of each coil and significantly increases the available signal-to-noise ratio (SNR).

To image the body uniformly, many coil designs have been developed to produce a homogeneous magnetic field within a specified volume. An extension of the surface coil, seen in Figure 6(a), uses a series of loops to produce a

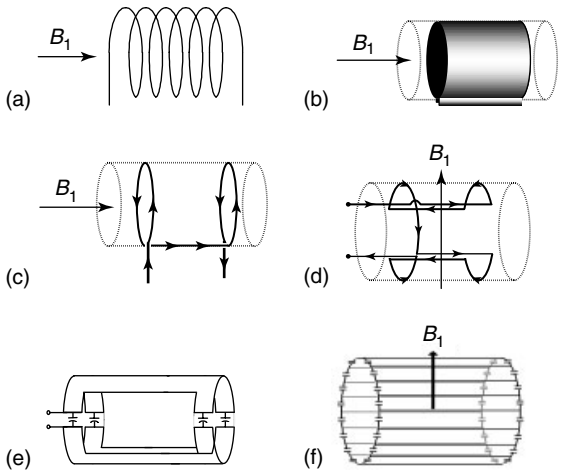

Figure 6 Common volume coils used in MRI. (a) The solenoid coil creates a uniform magnetic field along its axis. (b) The loop gap resonator uses a continuous conductive sheet to create a longitudinal magnetic field. This simple modification allows the loop gap resonator to be tuned to higher frequencies than a similar solenoid coil. (c) The Helmholtz coil is designed to allow easier access to the center of the coil while in the magnet, but still produces a magnetic field along the axis of the coil. (d) The saddle-shaped volume coil uses a simple design to produce a magnetic field perpendicular to the axis of the coil. (e) The slotted tube resonator also produces a perpendicular magnetic field with better homogeneity compared to a saddle shaped coil. (f) The birdcage coil offers superb field homogeneity and high sensitivity. As the birdcage produces a perpendicular magnetic field, this design is often employed for imaging the head and extremities..

uniform field oriented along its axis. This design, called a solenoid coil, maximizes both homogeneity and sensitivity and is relatively easy to construct. As the coil's magnetic field must be perpendicular to the main magnetic field, the axis of the solenoid coil must be perpendicular to the axis of the magnet. This orientation restricts access to the coil center, making it difficult to image large objects. In addition, the solenoid coil is not suited for high-frequency applications due to self-resonance problems.

The loop gap resonator[20] and the Helmholtz coil have been designed to overcome some of the limitations of the solenoid coil. The loop gap resonator, shown in Figure 6(b), consists of a conductive metal sheet wrapped in a cylindrical shape. The slight gap between the edges of the sheet forms an effective capacitance. This coil can be tuned to higher frequencies than a solenoid coil. However, the geometry of a loop gap resonator prohibits access to the interior of the coil once it is placed in the bore of the magnet. Figure 6(c) shows a two-turn solenoid design called a Helmholtz coil. This coil allows samples to extend beyond the coil and provides easier access to the coil center, but suffers from relatively low sensitivity and homogeneity. Both loop gap resonators and Helmholtz coils also produce a magnetic field along

their axes, requiring the coil to be oriented perpendicular to the axis of the magnet. This makes solenoid, loop gap, and Helmholtz coils inappropriate for most whole-body imaging applications on superconducting magnets.

Several coil designs produce magnetic fields perpendicular to their axes. These coils can be oriented along the magnet's axis and still produce a perpendicular magnetic field. The saddle-shaped coil shown in Figure 6(d) produces a perpendicular magnetic field, allowing easy access to both ends of the coil.[21] Another design is the slotted tube resonator shown in Figure 6(e).[22] This design produces a more homogeneous magnetic field than is possible with a saddle-shaped coil. The birdcage coil seen in Figure 6(f) represents a highly advanced design.[21] This coil offers the maximum homogeneity, sensitivity and accessibility to the coil center, allowing it to become one of the most popular designs in clinical MRI. However, a birdcage coil can be very difficult to design and construct.

An improvement in sensitivity by a factor of $\sqrt{2}$ can be accomplished by quadrature detection.[23] This requires that two RF coils in an orthogonal orientation be used for transmitting and receiving. If the coils are driven by RF pulses with a 90° phase shift, a rotating RF field will be produced rather than a field aligned along a single axis. The improved sensitivity can be very useful for large whole-body coils that are not close to the tissue under examination. The symmetry of the birdcage coil makes it perfectly suited for quadrature design.

There are often times when two nuclei need to be studied in biological examinations. The ability to tune a coil to more than one frequency allows examination of two nuclei in the exact same tissue under the exact same physiological states. This eliminates the need to move the patient or change the coil, reducing the time required for the examination. A simple way to construct a dual-tuned coil is shown in Figure 7.[24] This is equivalent to the single tuned circuit with the addition of a trap circuit in series with the tuning capacitor. At the lower frequency the trap circuit is inductive, whereas the trap is capacitive at the higher frequency. The series combination of the tuning capacitor and the inductive trap at the lower frequency tunes the circuit to the lower frequency. Likewise, the combination of the tuning and trap capacitors when driven by high frequency tunes the circuit to the higher frequency. Using similar design strategies, more trap circuits can be added to create triple, quadruple, or higher-tuned coils.

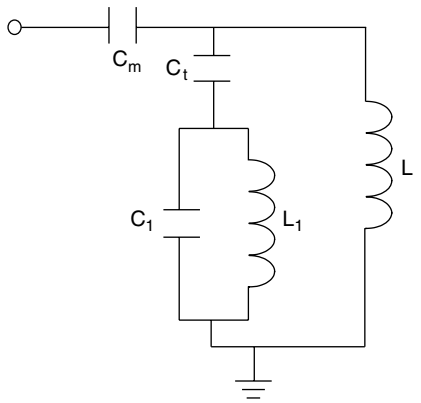

Figure 7 Circuit diagram for dual-tuned coil. The series combination of C_1 and L_1 forms a trap circuit, allowing the coil to resonate at two frequencies simultaneously..

1.3 Advantages

MRI has many advantages over other clinical imaging techniques and may be preferable to many chemical analysis techniques. Unlike many other imaging methods, MRI does not use ionizing radiation. X-ray and CT image the human body by recording the density of various structures.[25] The denser materials, such as bone, absorb more radiation than the soft tissues. Single photon emission computed tomography (SPECT) and positron emission tomography (PET) are two other imaging methods that employ ionizing radiation in a very different way. These techniques follow radioactively labeled metabolites through the body.[25] The resolution and SNR available with methods using ionizing radiation are often limited by the acceptable dose of radiation. In addition, cumulative exposure to radiation limits the number of times a patient can be imaged using these techniques. As MRI does not require the use of radioactive materials, the resolution and SNR do not suffer from the same restraints, and MRI can be used repeatedly without concern for cumulative radiation exposure.

Another strength of MRI is its high temporal and spatial resolution. Modern MRI techniques allow images to be collected within tens of milliseconds with resolutions of the order of millimeters. These feats cannot be achieved with classical radiographic methods. Other noninvasive techniques, such as ultrasound and optical imaging, also suffer from poor spatial resolution.

The contrast obtained from most imaging methods, including X-ray, CT, SPECT, PET, and ultrasound, depends upon the density of tissue or an exogenous contrast agent. However, MRI is much more versatile and can use the physical properties of tissue, such as density or relaxation, or specially designed contrast agents, such as relaxation agents. In addition, MRI can exploit numerous other forms of contrast, such as flowing blood, diffusive movement, or changes in blood oxygen saturation. This allows MRI to be employed in a wide variety of imaging applications.

The MRI technique has proved useful in the detection of many different types of brain diseases and disorders. It is often the preferred method clinically to diagnose and characterize brain tumors, infections, cerebrovascular disease, head trauma, intracranial hemorrhage, and white-matter disease. It can detect a wide variety of diseases including multiple sclerosis, herpes simplex virus type 1, meningitis, stroke and AIDS.

MRI has gained widespread acceptance as an invaluable diagnostic tool in many other parts of the body, including the cardiac, musculoskeletal, and abdominal systems. The soft-tissue contrast abilities, multiplanar imaging and large FOV, has made MRI a valuable tool for evaluating gynecological diseases. For example, MRI is 77–100% more sensitive in the detection of müllerian duct anomalies than classical ultrasound,[26–28] and has an accuracy greater than 90% for differentiating and staging various types of uterine tumors.[29–36] Clinical trials are currently evaluating three liver-specific MRI contrast agents: manganese *N*,*N*′-bis(pyridoxal-5-phosphate)ethylenediamine-*N*,*N*′-diacetic acid (Mn-DPDP), gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid (Gd-EOB-DTPA) and gadolinium benzyloxy propionictetraacetate dimeglumine (Gd-BOPTA-Dimeg).[37–42] These agents are preferentially taken up by hepatocytes within 30–120 min after intravenous injection and then cleared by the bile and renal systems. The rate of contrast uptake and the enhancement patterns can be used to characterize liver lesions.

NMR spectroscopy can be used to determine the chemical properties of tissues. It can detect which chemical species are present, how the body metabolizes various drugs, and how physiological perturbations alter tissue function. Unlike other analytical chemistry tools with biomedical applications, such as mass spectroscopy and atomic absorption, NMR does not require removal and destruction of the sample. Hence, NMR does not alter the tissue characteristics and allows repeated measurements from the same tissue. Similarly, some clinical methods, such as biopsy and histology, require invasive and potentially harmful techniques. Repeated measurements on the same sample cannot be performed with mass spectroscopy, atomic absorption, or histology. However, NMR can be performed noninvasively and serially on the same tissue.

1.4 Disadvantages

In the clinical setting MRI also has several disadvantages. The first is the complex and expensive instrumentation required to implement MRI. This equipment needs constant maintenance and care. The magnet, in particular, requires regular maintenance to replace the liquid helium and nitrogen that cools the superconducting coils. The equipment is so bulky that the patient must be transported to the imager. For some critical care patients this may be impossible. The stray magnetic fields and heavy magnets also require special housing considerations for a MRI scanner.

From a chemical analysis perspective, NMR has certain disadvantages compared to other methods. The sensitivity of NMR is much lower than found in mass spectroscopy and atomic absorption techniques. In addition, NMR may require complicated calibration routines if quantitative results are needed.

Although MRI uses only RF radiation, some people are not suited for imaging. People with pace makers, metallic implants or metallic prosthesises are not imaged because of the dangers related to strong magnetic fields. In addition, caution must be exercised in MRI examination of patients who are pregnant. As with all interventions in pregnancy, MRI should be used during the first trimester when it offers a definite advantage over other tests.

2 PROTON IMAGING

2.1 Spatial Encoding and Imaging Pulse Sequences

To produce an image of the human body, the magnetic resonance signal must be spatially encoded along three orthogonal axes. Magnetic field gradients are used to alter the frequency or phase of the magnetic resonance signal based upon its spatial origin. Thus, gradients allow the magnetic resonance data to be reconstructed into an image of the patient. While a magnetic field gradient is on, the resonant frequency of the nuclei will change according to position along the gradient axis. A stronger gradient will cause a larger change in the resonant frequencies. When the gradient is turned off, the spins return to the same frequency, but the phases are proportional to the position along the gradient axis. The strength and the duration of the magnetic field gradient determine the phase shift of the spins. In this way, spins ''remember'' that a gradient has been applied after it is turned off.

To demonstrate how the magnetic resonance signal is encoded in three dimensions, assume that the patient is lying on his back in the magnet, as shown in Figure 2. The coordinate system is defined such that the z axis is along the axis of the magnet, the x axis is up and down, and the y axis is left to right. It is assumed that images of axial slices through the patient are being collected. Therefore, slice selection will be along the z axis, frequency encoding will be along the x axis, and phase encoding will be along the y axis. These methods are easily modified to obtain sagittal (slice selection along the y axis, frequency

encoding along the z axis and phase encoding along the x axis) or coronal slices (slice selection along the x axis, frequency encoding along the z axis and phase encoding along the y axis). Figure 8 shows diagrams of the three most common medical imaging planes. In addition, these methods can be generalized to image any arbitrary plane, as described in the oblique imaging section.

The basic pulse sequence used in MRI is the spin–echo sequence shown in Figure 9.[4] A spin–echo sequence uses a 90° RF pulse to excite the nuclei and a 180° pulse to refocus inhomogeneities in the main magnetic field. In addition, the delay between the 90° and 180° pulses allows for spatial encoding of the magnetic resonance signal. The time between the 90° pulse and the formation of the echo is equal to the echo time (TE). The time between the 90° and 180° pulses is equal to TE/2. The time between the start of the last sequence and the start of the next sequence is the repetition time (TR). To encode the magnetic resonance signal in all three dimensions, three spatial encoding techniques are required: slice selection, frequency encoding, and phase encoding.

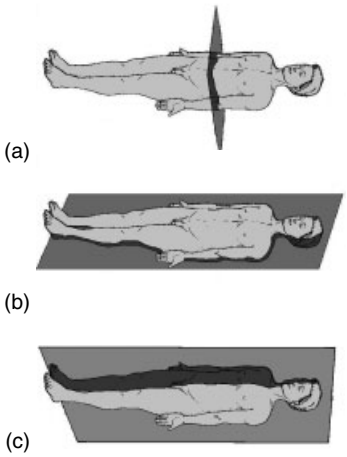

Figure 8 The three most common medical imaging planes shown with the patient in the supine position: (a) axial, (b) coronal, and (c) sagittal..

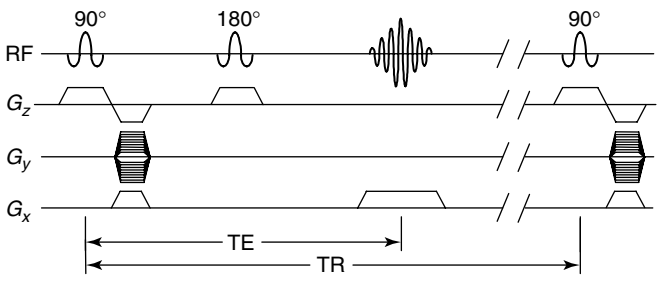

Figure 9 Spin–echo pulse sequence used in many MRI examinations..

2.1.1 *Slice Selection*

Slice selection is used to excite and receive signals from only a single thin plane along the z axis. This spatially encodes the signal in the first dimension. Slice selection requires the application of a magnetic field gradient along the z axis. This gradient alters the resonance frequency for nuclei at different positions along the z axis. An RF pulse can be applied that only excites a limited band of these resonance frequencies. Application of both a magnetic field gradient and a RF pulse with a narrow bandwidth will excite only those nuclei in the $x-y$ plane, matching the RF with the resonance frequency. Typically, the amplitude of the frequency-selective RF pulse has a SINC shape, causing the selected slice to have a rectangular profile. Other pulse shapes have also been used to obtain a uniform slice excitation. The position of the slice can be altered by changing the frequency of the band-limited RF pulse. The thickness of the slice can be changed by modifying the bandwidth of the RF pulse or by changing the magnitude of the gradient.

In Figure 9, both the 90° and 180° pulses are slice-selective pulses. In addition, the negative lobe on the z gradient is used to correct for the dephasing caused by the 90° slice-selective gradient. No rephasing is required after the 180° slice selective gradient because the 180° pulse causes the first half of the gradient to balance with the second half.

2.1.2 *Frequency Encoding*

The x dimension is encoded using a magnetic field gradient applied during data collection, called the readout gradient.[4] During data collection, all of the excited spins are emitting RF energy at a frequency proportional to the magnetic field. By applying a gradient along the x axis, the frequency of the magnetic resonance signal is proportional to the position along the x dimension. The gradient along the x axis in Figure 9 prior to application of the 180° pulse is to correct for dephasing caused by the readout gradient. These two gradients ensure that the net phase at the center of the readout gradient will be zero. Alternatively, the dephasing gradient may be placed after the 180° pulse, but the magnitude must be reversed to cancel the readout gradient.

2.1.3 *Phase Encoding*

Spatial encoding along the y axis requires phase encoding.[4] In Figure 9, the phase-encoding gradient is applied along the y axis between the 90° and 180° pulses. This gradient has no effect on the frequency of the acquired signal. Instead, the gradient imparts a phase that is proportional to the position along the y axis. In order

to generate a unique combination of phase and frequency for each pixel in the image plane, the imaging sequence is repeated several times with different phase-encoding gradient amplitudes.

2.1.4 *k-Space Representation of Magnetic Resonance Data*

The readout and phase-encoding magnetic resonance data is collected in a conceptual matrix known as k-space. The horizontal axis of k-space corresponds to the phase axis and the vertical axis corresponds to the frequency axis. The number of horizontal lines in k-space is equal to the number of phase-encoding steps and the number of points in each horizontal line is equal to the number of data points collected for each echo. In a spin–echo pulse sequence, one line of k-space is filled for each echo acquisition. The central horizontal line is collected when no phase-encoding gradient is applied. Positive phase-encoding gradients fill the upper half of k-space and negative phase-encoding gradients fill the lower half. As the phase-encoding gradients become stronger (in either the positive or negative direction) k-space lines further away from the center line are filled (top or bottom, respectively). Because a stronger phase-encoding gradient causes more dephasing, the top and bottom lines of k-space have less signal intensity than the middle lines. Similarly, the central portion of an echo has higher amplitude than the outer portion. Therefore, most of the magnetic resonance signal is concentrated around the center (both in the horizontal and vertical dimensions) of k-space.

It is important to remember that k-space does not correspond to the magnetic resonance image, that is, the top line of k-space is not the top line of the image. The central portion of k-space (with the highest signal) provides low-resolution information that corresponds to large areas of uniform signal intensity in the image. In comparison, the outer portions of k-space (with relatively little signal) provide high-resolution information that details fine aspects and borders of different objects in the image.

A spin–echo pulse sequence fills one line of k-space with every acquisition. Other pulse sequences may fill k-space in other ways. For instance, echo planar imaging pulse sequences fill all lines of k-space with a single acquisition. Typically, k-space is filled in a rectangular pattern, but some pulse sequences use spiral or radial patterns (which then must be interpolated onto the rectangular data points for discrete Fourier transformation). The data is stored in k-space and multidimensional Fourier transformations, as described in the next section, are performed to reconstruct the magnetic resonance images.

2.1.5 *Image Reconstruction*

The process of magnetic resonance image reconstruction is summarized in Figure 10. MRI produces a set of echo signals from each slice. Fourier transformation of each echo reconstructs the image information along the x axis. This process reveals the spatial information contained in the frequency content of the signal. To recover the

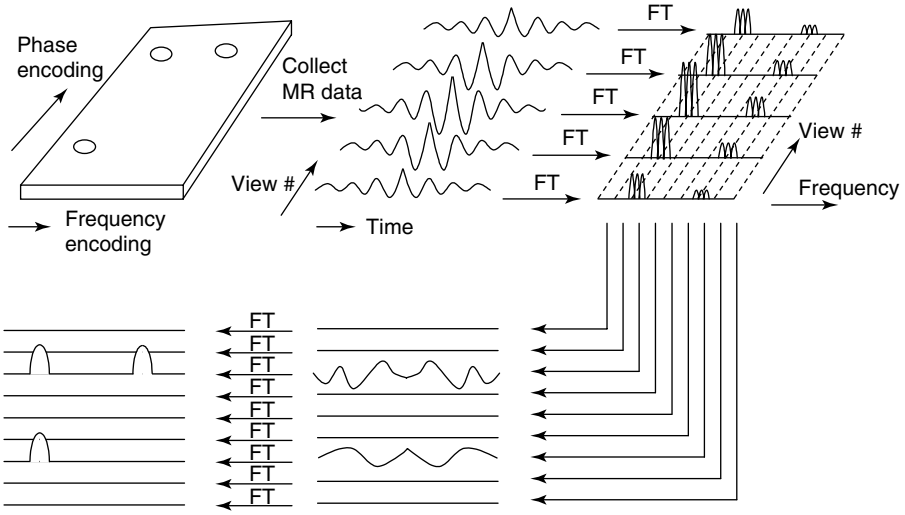

Figure 10 Simple diagram of magnetic resonance image reconstruction using 2-D Fourier transform. A slice containing three vials of water is imaged. A set of echoes is collected during magnetic resonance data acquisition. A one-dimensional Fourier transformation is applied to each echo, yielding a set of projections of the slice onto the frequency encoding axis. The rows and columns are interchanged to produce a new data set. A one-dimensional Fourier transform of this data set produces the image. Therefore, a 2-D Fourier transformation is used to reconstruct the image from the raw MRI data..

phase-encoded information, a second Fourier transform must be performed in the second dimension. This reveals the spatial information contained in the phase content of the signal. This 2-D Fourier transformation produces the image through a slice in the patient.

2.1.6 *Multislice Imaging*

If one needs to obtain images from multiple slices, one could repeat the above 2-D imaging experiment several times and obtain multiple images. However, this would increase the image data collection time substantially. Another approach is to excite and detect a different parallel slice during the TR time when one is waiting for T_1 processes to restore longitudinal magnetization in the first slice.[43] The only restriction is that the excitation used for one slice must not affect those from another slice. This can be accomplished by applying one magnitude slice selection gradient and changing the RF. In fact, excitation/detection may be carried out in many parallel slices during each TR interval and images from multiple slices can be collected without increasing the total data collection time.

2.1.7 *Oblique Imaging*

Oblique imaging is the production of images that lie between the conventional x, y, and z axes.[16] Oblique imaging is performed by applying linear combinations of the x, y, and z magnetic field gradients so as to produce a slice-selection gradient perpendicular to the imaged plane, a phase-encoding gradient along one edge of the imaged plane, and a frequency-encoding gradient along the remaining edge of the image. For example, Figure 11 shows an image slice lying along the x axis but passing between the z and y axes such that it makes a 30° angle with respect to the y axis and a 60° angle with the z axis. The following combination of gradients will be needed to image this slice: slice-selection gradient, $G_z = G_s \sin 60^\circ$ and $G_y = -G_s \cos 60^\circ$; phase-encoding gradient; $G_z = G_p \sin 30^\circ$ and $G_y = G_p \cos 30^\circ$; frequency-encoding gradient, $G_x = G_f$. In a similar fashion, double-oblique imaging is often employed in cardiac imaging to obtain long-axis or short-axis views of the heart.

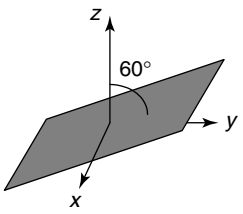

Figure 11 Example of an oblique imaging slice. This slice lies along the x axis, but forms a 60° angle with respect to the positive z axis and a 30° angle with respect to the positive y axis..

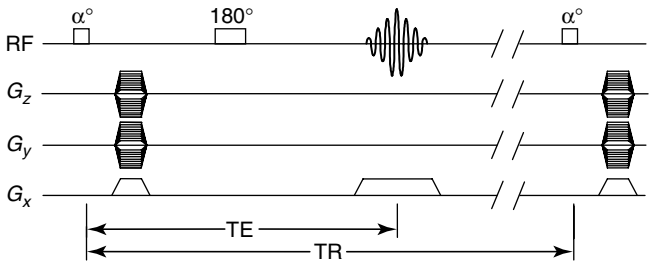

Figure 12 Pulse sequence used for collecting 3-D images. The slice-selective 90° and 180° pulses are replaced by hard pulses that excite the entire sample. The first RF pulse is labeled α because 3-D sequences often use excitation pulses with flip angles less than 90°. In addition, the slice-selective gradients are replaced by a second set of phase-encoding gradients..

2.1.8 *Volume Imaging*

The essential idea in volume imaging, or three-dimensional (3-D) imaging, is the replacement of slice-selective excitation with another phase-encoding process along the slice axis.[44] A 3-D imaging pulse sequence is shown in Figure 12. Each RF pulse excites the entire imaging volume instead of just one slice. This is followed by two phase-encoding gradients and one frequency-encoding gradient. A 3-D Fourier transformation of the data set yields a 3-D image, which may be arranged to yield sets of axial, sagittal, coronal, or oblique images.

In practice, it is often desirable to limit the 3-D acquisition to just a portion of the sensitive volume within the RF coil. This is accomplished by selecting a slab encompassing the slices to be imaged. The 3-D acquisition then results in subdividing the slab into thinner slices. The main drawback of 3-D imaging is long data collection time. However, in some instances when a short TR is used, 3-D imaging may produce better SNR compared to multislice 2-D imaging.

2.2 Relaxation and Image Contrast

As stated earlier, RF radiation at the resonance frequency excites the nuclei during a MRI examination. Relaxation of the nuclei is governed by two mechanisms, longitudinal and transverse relaxation. These relaxation effects are commonly used to generate contrast in magnetic resonance images.

2.2.1 Longitudinal Relaxation

The mechanism by which spins relax back to equilibrium is called longitudinal relaxation or spin–lattice relaxation. Longitudinal relaxation describes the process of nuclei returning to the lower energy state by losing energy to their surroundings. This process is also called spin–lattice relaxation to emphasize the exchange of energy between the excited spins and their surroundings. The return of spins to the lower energy state is an exponential process, given by Equation (1):

$$M_z = M_z^{\mathrm{o}}(1 - \mathrm{e}^{-t/T_1}) \tag{1}$$

where M_z^{o} is the net equilibrium magnetization, M_z is the longitudinal magnetization at time t and T_1 is the time constant for longitudinal relaxation. Spins are considered completely relaxed after 3–5 T_1 times.

2.2.2 Transverse Relaxation

Transverse relaxation, also called spin–spin relaxation, occurs when spins in the high and low energy state exchange energy but do not loose energy to the surrounding lattice. This results in loss of magnetization in the transverse plane. The rate of loss of transverse magnetization is also an exponential process given by Equation (2):

$$M_{xy} = M_{xy}^{\mathrm{o}}\, \mathrm{e}^{-t/T_2} \tag{2}$$

where M_{xy}^{o} is the initial transverse magnetization, M_{xy} is the transverse magnetization at time t, and T_2 is the time constant for transverse relaxation.

In addition to the molecular interactions, inhomogeneities in the main magnetic field also contribute to the decay of transverse magnetization. The time constant for the combination of these two factors is called T_2^* and is given by Equation (3):

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2^{\mathrm{inhomo}}} \tag{3}$$

2.2.3 Factors Influencing Relaxation

The T_1 relaxation is short when the motion of the nucleus (rotations and translations, or tumbling rate) matches that of the resonant frequency. As a result, T_1 relaxation is dependent on the main magnetic field strength, which specifies the resonant frequency. Higher magnetic fields are associated with longer T_1 times. T_2 is mostly field independent for the range of magnetic field strengths used in clinical magnetic resonance. When the motion of the nucleus is fast, the T_2 and T_1 times are approximately the same. The T_2 becomes considerably shorter than the T_1 as the motion of the nuclei become slow. The T_2^* time is always less than or equal to the T_2 time. The values of T_1 and T_2 at 1.5 T for some tissues in the human head are listed in Table 1.

2.2.4 Proton Density Weighted Images

Proton density images reflect the number of mobile hydrogen nuclei in the tissue. A spin–echo sequence with a long TR and a short TE will produce a proton density weighted image. The long TR allows all the spins to return to their relaxed state before the start of the next sequence. In addition, the short TE minimizes the effects of T_2 relaxation. Owing to the long TR required, the imaging time for proton density images can be very long. They also suffer from poor contrast because most tissues have essentially the same concentration of protons. A typical proton density image of the brain is shown in Figure 13. This image was collected with a TR of 6 s and a TE of 20 ms.

2.2.5 T_1-weighted Images

The contrast in T_1-weighted images is obtained through differences in T_1 relaxation times between various tissues. T_1-weighted images can be produced with a spin–echo

Table 1 Typical values of T_1 and T_2 relaxation in the human head measured at 1.5 T

Tissue	T_1 (ms)	T_2 (ms)
CSF	3000–6000	110–2000
White matter	600–800	61–100
Gray matter	1000–1300	61–109
Meninges	500–2200	50–165
Muscle	950–1820	20–67
Adipose	200–750	53–94

CSF, cerebrospinal fluid.

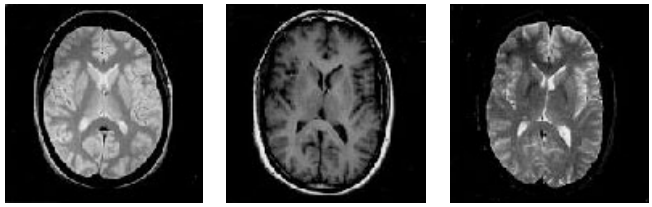

Figure 13 Examples of axial images through the brain, illustrating ventricles, gray matter and white matter. The proton density image was collected with TR/TE = 6000/20. For the T_1-weighted image TR/TE = 400/20; for the T_2-weighted image TR/TE = 4000/80. (Images courtesy of General Electric Medical Systems.).

sequence with a short TR and short TE. The short TE minimizes the effects of variations in T_2 values from different tissues. However, the short TR accentuates the effects of different T_1 values. Tissues with short T_1 times can fully relax back to equilibrium within the short TR time and appear bright on the final image. Tissues with long T_1 values cannot fully relax before the next acquisition and appear dark. A sample T_1-weighted image of the brain (with a TR of 400 ms and a TE of 20 ms) is shown in Figure 13. In such an image, fat is bright, CSF is dark and the gray matter of the brain is slightly darker than the white matter.

Other methods have been developed to produce T_1-weighted images under special circumstances. For instance, the inversion recovery sequence shown in Figure 14 enhances T_1 contrast and more clearly delineates gray and white matter. The inversion recovery sequence consists of a 180° pulse, which inverts the spins along the $-z$ direction, followed by an inversion recovery delay in front of the standard spin–echo pulse sequence. At a given inversion recovery delay value, the signal intensities of white and gray matter are more widely separated than in a T_1-weighted image obtained using a spin–echo pulse sequence. A sequence known as magnetization prepared rapid gradient echo imaging (MP/RAGE),[45] has been developed to obtain 3-D images with high T_1 contrast with a very short acquisition time.

2.2.6 T_2-weighted Images

Another way to increase MRI contrast is to produce a T_2-weighted image. T_2-weighted images are created when both the TR and TE times are long. As with proton density images, the long TR allows all the spins to relax back to equilibrium before the next imaging cycle, minimizing T_1 relaxation effects. However, the long TE time allows distinction to be made between tissues with different T_2 times. Tissues with a short T_2 will lose a great deal of signal during TE and will appear dark. However, tissues with a long T_2, will not relax as much during TE and will appear bright. A T_2-weighted image of the brain is shown in Figure 13. This image was collected with a TR of 4 s and a TE of 80 ms.

Like proton density images, the spin–echo T_2-weighted imaging sequence uses a long TR time. Thus, the acquisition time for T_2-weighted images can be very long. Sometimes the proton density sequence and the T_2-weighted sequence are combined to reduce the total examination time. In such a sequence, the TR time is kept long but two echoes are formed. The first echo is formed with a short TE and corresponds to the proton density image. The second echo is formed after a longer TE and corresponds to the T_2-weighted image.

As mentioned earlier, T_2^* is much shorter than T_2 and can obscure the action of spin–spin relaxation. The value of T_2^* is a function of local inhomogeneities in the main magnetic field. To obtain an image with T_2 weighting, the T_2^* effects must be removed from the image. The 180° pulse in a spin–echo sequence refocuses the phase shifts caused by inhomogeneities in the magnetic field, minimizing T_2^* weighting.

Under certain circumstances, a spin–echo pulse sequence is not suitable. For instance, a surface coil may not be able to produce a 180° pulse due to the B_1 field inhomogeneities. Gradient-echo pulse sequences,[46] such as the gradient-recalled acquisition at steady state (GRASS) pulse sequence shown in Figure 15, can create an echo signal without a RF refocusing pulse. The echo is formed by the joint action of the negative gradient and the read-out gradient along the frequency encoding axis. The negative gradient dephases the NMR signal, allowing the read-out gradient to rephase the signal and form an echo. If the negative gradient is equal to half the read-out gradient, the echo will be rephased at the center of the data collection window. The dephasing gradient can be either half the amplitude or half the duration of the read-out gradient to satisfy this condition.

Typically, a gradient echo can be formed with a shorter TE, 5–10 ms, than is possible with a spin–echo. Therefore, gradient echoes are often employed with fast imaging

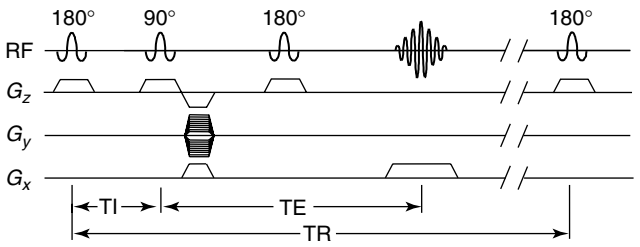

Figure 14 Pulse sequence for inversion recovery imaging. The first 180° pulse accentuates differences in T_1 relaxation. The inversion time (TI) controls the amount of T_1 weighting..

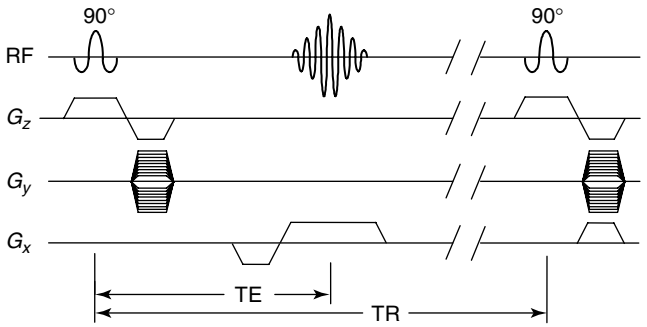

Figure 15 GRASS pulse sequence. This sequence uses a gradient echo instead the spin–echo of Figure 9. This sequence can produce TEs of <10 ms and produces T_2^*-weighted images..

pulse sequences. The B_0 field inhomogeneities are not refocused due to the lack of a 180° pulse, causing gradient echo images to be T_2^* weighted. Although this is typically an unwelcome consequence, certain imaging sequences, such as the blood oxygen level dependent (BOLD) effect, rely upon T_2^* weighting to provide the desired contrast. A further discussion of functional brain imaging using the BOLD effect is presented in section 2.4.2.

2.2.7 Contrast Agents

As with conventional CT imaging, intravenous contrast agents are often used in MRI. The most common contrast agent used in MRI is gadolinium diethylenetriamine pentaacetic acid (Gd-DTPA).[47–51] This agent decreases the relaxation times of nearby hydrogen nuclei. In proton density images, the contrast agent does not significantly alter the image. In T_1-weighted images, the T_1 is reduced and tissues exposed to the contrast agent appear brighter. In T_2-weighted images, the contrast agent decreases the T_2 of tissues containing contrast agent and so they appear darker. The additional contrast may be used to detect hemorrhage, disruption of the blood–brain barrier, or the increased vascularity associated with tumors. Because of the highly efficient relaxation mechanisms, the dose of contrast agent required for MRI is only about 1/20 that needed in CT imaging.[52–56]

2.2.8 Chemical Shift-selective Imaging

In some circumstances it may be necessary to image structures that are obscured by the bright signal of fat. This may be accomplished by exploiting the fact that fat and water protons are spectrally separated by about 3 ppm. It is therefore possible to use spectral saturation techniques such as used for water suppression (as described in section 3.3). As with water suppression, a number of methods have been developed utilizing RF pulses and magnetic gradients. In one case, the fat portion of the spectrum is saturated prior to the image sequence so that no fat signal can contribute to the image. Alternatively, an inversion recovery sequence can exploit the T_1 relaxation of fat so that the 90° excitation pulse is applied when the fat signal has no longitudinal component.

2.3 Image Artifacts

An image artifact is any false feature appearing on the magnetic resonance image caused by the imaging process. The source of the artifact can be from the magnetic resonance hardware or of a physiological origin.

The causes of hardware artifacts include B_0 inhomogeneity, gradient limitations, RF inhomogeneity, and RF quadrature effects. B_0 inhomogeneity of just a few ppm leads to noticeable shading of magnetic resonance images due to T_2^* effects, whereas greater inhomogeneities give rise to spatial distortions in the image.

Artifacts arising from problems with the gradient system are very similar to those from B_0 inhomogeneity. If an abnormal current passes through one of the gradient coils, the aspect ratio of the image will be incorrect and the shape of the object will be distorted. If a gradient coil is damaged, the direction of a gradient may not remain constant, which will make the object appear bent.

An RF inhomogeneity problem causes a variation in intensity across an image. Both nonuniform RF transmission and nonuniform reception can cause artifacts. Some RF coils have inherent variations in sensitivity and will always display artifacts. For instance, a surface coil is more sensitive to tissues close to the plane of the coil and the corresponding images will show higher signal in these tissues. The presence of a nonferromagnetic metal, such as dental work, may prevent the RF energy from passing into a tissue, leaving a signal void in the image. An example of a RF inhomogeneity artifact can be seen in Figure 16(a).

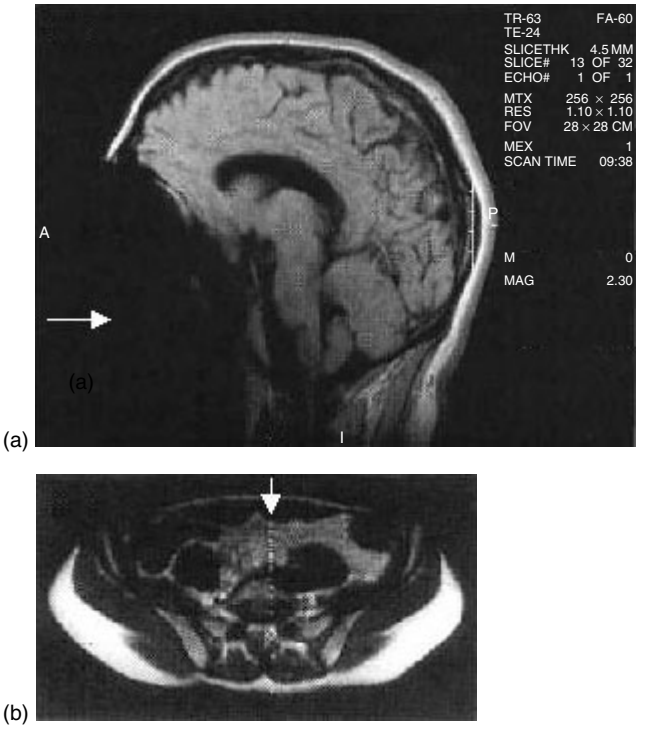

Figure 16 Examples of image artifacts arising from instrument limitations. (a) This sagittal head image displays a signal void (arrow) due to RF inhomogeneities caused by a large amount of dental work in the patient's mouth. (b) This axial image suffers from artifacts (arrow) caused by improper RF quadrature. The alternating black and white pattern of this artifact causes it to be known as a zipper artifact. (Images courtesy of Dr. William W. Orrison, Jr, University of Utah.).

RF quadrature artifacts arise from improper operation of the two channels of the detector. If one of the amplifiers has a DC offset in its output, the Fourier transformed data can display bright or dark spots in the center of the image, as seen in Figure 16(b). If one channel of the detector has a higher gain than the other, it will result in a ghosting of objects diagonally in the image.

Some natural processes that can cause image artifacts include motion, flow and chemical shift. Motion of the entire object during the imaging sequence generally results in a blurring of the entire image with ghost images in the phase encoding direction. Movement of a small portion of the imaged object results in blurring of that small portion of the object across the image. Immobilizing the patient or imaged object can minimize motion artifacts. Frequently, however, the patient's breathing or heartbeat causes the motion as in Figure 17(a). The solution in these cases is to gate the imaging sequence to the cardiac or respiratory cycle of the patient.

Blood flowing through a slice can experience a RF pulse and then flow out of the slice by the time the signal is recorded. This results in loss of signal from blood vessels as shown in Figure 17(b). In a multislice sequence, the slices could be positioned such that blood experiencing a 90° pulse in one slice can flow into another slice and experience a 180° rotation and into a third and contribute to the echo. In this case, the effect is usually that some slices have dark blood vessels and others have bright blood vessels.

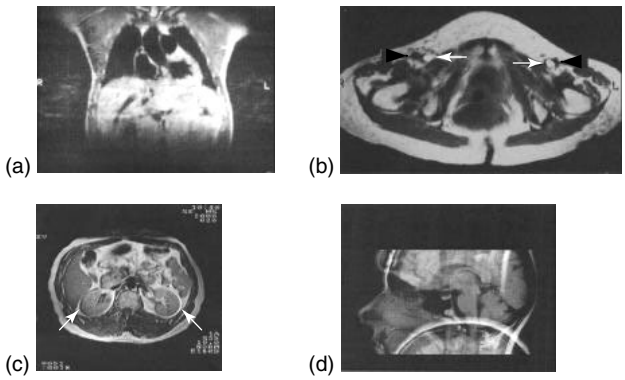

Figure 17 Examples of naturally occurring image artifacts. (a) This coronal image of the abdomen shows severe ghosting and blurring due to the patient's breathing during the acquisition. Such artifacts can be avoided by instructing the patient to hold their breath while the image is collected. (b) The femoral arteries in this pelvic image are black (black arrowheads) because of flow artifacts. The femoral veins (white arrows) appear bright because venous blood does not flow as quickly as arterial blood. (c) This axial image shows chemical shift artifacts (arrows), in which areas of the image representing fat are shifted relative to the areas that mainly contain water. (d) This image of the head shows a wrap-around artifact. The top portion of the head wraps around to the bottom of the image and the lower jaw and neck are wrapped-around to the top of the image. (Images (a), (c), and (d) courtesy of Dr. William W. Orrison, Jr, University of Utah. Image (b) courtesy of Dr. William G. Bradley, Jr, Long Beach Memorial Medical Center.).

A chemical shift artifact is caused by the difference in the resonance frequencies of fat and water. The artifact manifests itself as a misregistration between the fat and water pixels in an image. The magnitude of the effect increases with the B_0 field strength. At 0.5 T the effect is less than one pixel, whereas at 1.5 T the effect is approximately 2 pixels. Figure 17(c) shows an image with chemical shift artifacts. The portions of the image consisting mostly of fat are shifted slightly with respect to the portions that contain mostly water. This is most easily seen at the borders between fat and muscle.

In addition to the above causes, partial volume and wrap around artifacts may arise from improper instrument setting. A partial volume artifact is any artifact that is caused by the size of the image voxel. For example, a large voxel might contain a combination of fat and water. The signal intensity will be equal to the weighted average of the quantity of water and fat present in the voxel. Thus, resolution suffers because of multiple features present in a single-image voxel. Smaller voxels reduce partial volume artifacts by increasing the image resolution.

A wrap-around artifact is the occurrence of a part of the imaged anatomy, which is located outside of the FOV, inside of the FOV. A wrap-around artifact can be seen in Figure 17(d). This artifact is caused by the selected FOV being smaller than the size of the imaged object. More specifically, the digitization rate is less than the range of frequencies in the echo. Wrap-around artifacts are avoided by choosing a larger FOV, adjusting the position of the image center, or selecting an imaging coil that will not excite or detect spins from tissues outside of the desired FOV.

2.4 Advanced Imaging Techniques

As explained above, contrast can be enhanced by producing T_1-weighted or T_2-weighted images or by administering exogenous contrast agents. Other methods of contrast enhancement utilize flow, diffusion, or susceptibility effects. Flow and diffusion phenomena can be used to produce angiographic, perfusion-weighted, or diffusion-weighted images. The effects of susceptibility changes are often exploited in functional imaging. These phenomena can produce intrinsic contrast without the toxic contrast agents used in CT, such as iodine. Because of space limitation, none of the techniques will be covered in depth.

2.4.1 Flow-sensitive and Diffusion-sensitive Imaging

Flow- and diffusion-sensitive imaging can be used to produce angiographic images, perfusion-weighted images or diffusion-weighted images. Angiographic images display the signal from blood in the lumen of large and medium size vessels. Perfusion-weighted images are sensitive to the flow of blood through the microscopic capillary vessels. Diffusion-weighted images are sensitive to changes in the random thermal motion of water in the tissue. Techniques based on relaxation times or phase contrast are used for angiographic and perfusion-weighted imaging. Phase contrast techniques can also produce diffusion-weighted images.

Relaxation techniques use contrast agents or saturation methods to mark the incoming blood and trace its progress through the vessels. Exogenous contrast agents can be injected into the circulatory system to reduce the T_1 time of blood relative to the surrounding tissues.[57–63] When images are collected with a short TR value, the signal from the background tissues is suppressed due to its long T_1. Saturation techniques operate on a similar principle, but utilize endogenous spin labeling. After saturating the imaging slice, stationary tissues contribute very little signal to the image. However, the saturated spins in flowing blood are replaced by new fully relaxed spins. Therefore, an image produced shortly after saturation of the imaging slice will show only the unsaturated blood flowing into the slice.

Relaxation techniques can be used to monitor bulk flow through the macroscopic blood vessels or random flow through the capillary beds. For angiographic images the data is collected immediately following saturation or injection of contrast agent. Perfusion-weighted imaging requires a longer delay between saturation and image collection, to allow the unsaturated blood or contrast agent to enter the capillaries. These techniques cannot be used to produce diffusion-weighted images because of the incoherent nature of diffusion.

Phase-contrast methods can measure bulk flow, perfusion, or diffusion. To produce angiographic or perfusion-weighted images, phase contrast is achieved with a pulse sequence with a bipolar gradient pulse, as shown in Figure 18. A diffusion-weighted pulse sequence, as shown in Figure 19, also uses bipolar gradients. With a bipolar gradient, a positive gradient is turned on for a period of time then a negative gradient is applied for an equivalent amount of time. A bipolar gradient pulse has no net effect on stationary spins. The spins that move from one location to another will either gain or lose phase as a result of the gradients. Flow and diffusion are 3-D anisotropic processes. Therefore, phase contrast techniques must be applied in three directions. To accomplish this, images are collected with phase contrast gradients along each axis.

To produce a magnetic resonance angiographic phase-contrast image, two imaging sequences are performed in which the first has a positive bipolar gradient pulse and the second a negative bipolar gradient pulse. The raw data from the positive and negative bipolar gradient pulses are subtracted so that the signals from the stationary spins cancel, leaving only the moving spins.

Phase-contrast pulse sequences are also sensitive to perfusion. Because perfusion is random, the moving blood gains random phases. The average phase is zero, but the standard deviation of the phases is proportional to the perfusion. Plotting the log of signal intensity versus TE will yield a curve with two components. The slow component, associated with the diffusion of water in the interstitial and intracellular spaces, represents most (80–90%) of the total signal. The remaining fast

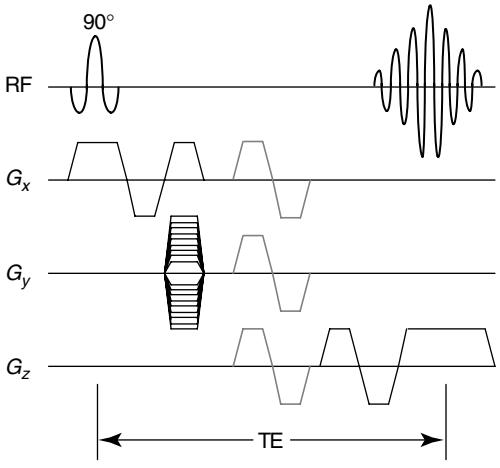

Figure 18 Phase contrast MRI angiography pulse sequence. A gradient echo is formed at time TE by the action of the slice-selective RF pulse and gradient. The flow encoding gradient is applied along the x, y, and z directions in different image acquisitions..

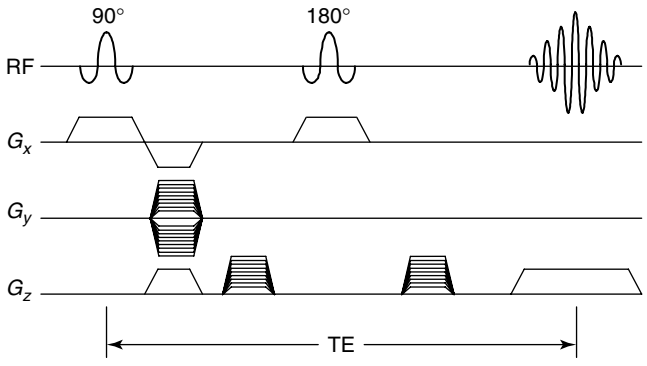

Figure 19 Pulse sequence sensitive to diffusion. Two matched but variable diffusion-weighting gradients are placed on either side of the slice-selective 180° pulse. In the absence of diffusion, these two gradients will cancel each other out and have no effect on the image. Diffusion causes dephasing of the spins which, in turn, decreases the signal amplitude..

component originates from blood flow through the capillary bed. However, exchange between the capillaries and the interstitial space can complicate the interpretation of such a plot.

Because most of the random motion in tissue is due to diffusion, only two phase-contrast images are needed with different gradient values to produce a diffusion weighted image. The ratio of the two images then yields a pixel-by-pixel map that is proportional to the diffusion coefficient. Although the gradient strength sets the sensitivity of a diffusion-weighted pulse sequence, the detectable diffusion coefficient is limited by the T_2 relaxation. An alternative method must be used to measure very long diffusion coefficients. Using stimulated echoes, it is possible to store the magnetization along the longitudinal axis. This allows longer diffusion times to be measured than is feasible with spin–echo pulse sequences.

Magnetic resonance angiography has been used extensively for the diagnosis of vascular disorders in the brain. A typical angiography image of the brain is seen in Figure 20. Arterial stenoses, tumor vascularity, aneurysms, and vascular malformations can be identified and characterized noninvasively by magnetic resonance angiography. The high quality of saturation techniques, also called time-of-flight methods,[64–69] has made MRI the method of choice for angiography. These methods have also been used in other parts of the body including the abdominal aorta, the vessels of the pelvis, and the gastrointestinal vessels. Any imaging sequence that is sensitive to blood flow is also sensitive to motion such as respiratory motion, patient movement or cardiac motion. For this reason, MRI has not been widely used to image the arterial tree of the heart. Radiographic imaging with a vascular contrast agent is still the most common method for mapping the coronary arteries.[70–72]

Although time-of-flight angiography has gained widespread acceptance, phase-contrast methods have a distinct advantage in detecting stenoses.[73] The narrowing of a vessel causes turbulence with no coherent flow. Time-of-flight methods show the narrowing of the vessel, but they do not detect turbulence. In phase-contrast images, however, areas of turbulence are dark because of the lack of bulk motion of the blood.

All the three techniques, angiography, perfusion-weighted imaging, and diffusion-weighted imaging, can be useful for diagnosis, characterization and treatment of stoke. During a stroke, blood flow in the brain drops due to blockage in the artery. Magnetic resonance angiography can be used to locate the blocked vessel, determine the extent of blockage, and predict which areas of the brain will be affected. Diffusion-weighted imaging can be used to assess the extent of ischemic damage within 30–60 min after a stroke.[74–78] A few minutes after a stroke, osmotic forces drive water into the cell causing intracellular edema, also known as cytotoxic edema. Because intracellular diffusion is slower than extracellular diffusion, cytotoxic edema decreases the average diffusion coefficient.[79–81] This leads to increased signal intensity[81–84] in the infarcted tissue on diffusion-weighted images. Figure 21 shows (a) proton density, (b) T_2-weighted, (c) T_1-weighted, and (d) diffusion-weighted images of the brain 12 h after a stroke. Whereas the other images show abnormal results, only the diffusion-weighted image clearly shows the extent of ischemic damage.

Following a stroke, the central area appears bright on diffusion-weighted images, but the surrounding regions

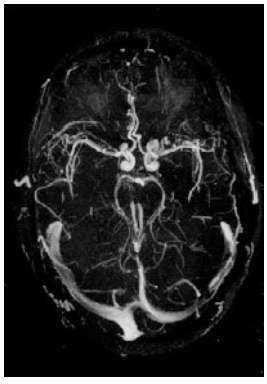

Figure 20 Angiographic magnetic resonance image of the brain. The background tissues are suppressed leaving only blood vessels in the image. (Image courtesy of Philips Medical Systems, Inc.).

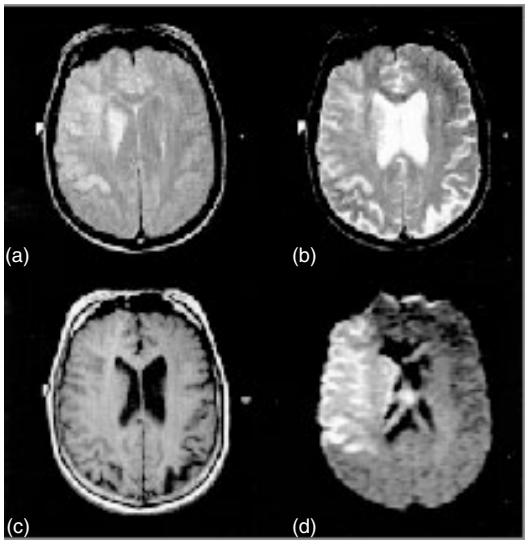

Figure 21 (a) Proton density, (b) T_2-weighted, (c) T_1-weighted, and (d) diffusion-weighted images of the brain 12 h after onset of stroke symptoms. Although all four images show abnormal results, only the diffusion-weighted image clearly shows the extent of ischemic tissues. (Image courtesy of General Electric Medical Systems.).

often appear normal. On perfusion-weighted images, no perfusion can be seen in the central area.[85–87] In addition, there is often decreased perfusion in the surrounding regions. This indicates a central infarcted area with decreased diffusion and no perfusion surrounded by a penumbra of normal diffusion but decreased perfusion. The penumbral regions often infarct later if not treated appropriately. Therefore, perfusion-weighted imaging can be used to prevent further ischemic damage to the brain.

In conclusion, magnetic resonance angiography can provide a detailed map of the blood vessels, allowing the blocked artery to be identified and treated. Diffusion-weighted images can be used to determine precisely the extent of ischemic tissues. Perfusion-weighted images can be used to predict areas at risk for future strokes, allowing suitable preventive treatments. Thus, all three methods can provide complementary information in the diagnosis and treatment of stroke.

2.4.2 *Functional Imaging*

Functional imaging can be used to study the activity of brain tissue. By monitoring which areas of the brain are stimulated under different tasks (such as visual stimulus, motor activity, auditory stimulus, mental tasks, and speech and language activities) the correlation between brain function and physical stimulation can be mapped. Most functional MRI in the brain is based upon the BOLD phenomenon.[88–91] Increases in neuronal activity are accompanied by increases in blood flow, volume, and oxygenation in the surrounding tissue. This increased oxygenation exceeds the increased metabolic demands of the brain tissue, causing the concentration of oxyhemoglobin in the blood to rise and the concentration of deoxyhemoglobin to decrease. Oxyhemoglobin is diamagnetic, whereas deoxyhemoglobin is paramagnetic. Therefore, deoxyhemoglobin acts like an intrinsic contrast agent. The increased oxygenation lowers the local concentration of deoxyhemoglobin, which in turn lowers the magnetic field inhomogeneities in the blood vessels. The signal arising from these areas of the brain decay at a slower rate and therefore appear brighter in images.

The ability to perform functional brain imaging came about through the development of EPI,[92] which enabled rapid imaging on the physiological timescale of brain activation. Figure 22 shows a typical EPI pulse sequence that can collect an entire image in only 0.1 s. In typical MRI, each phase-encoding step has to be individually executed to collect all the data needed for one image slice. For 256 phase-encoding steps, the pulse sequence has to run 256 times with a suitable delay between each step to avoid T_1 weighting. Instead of collecting the data

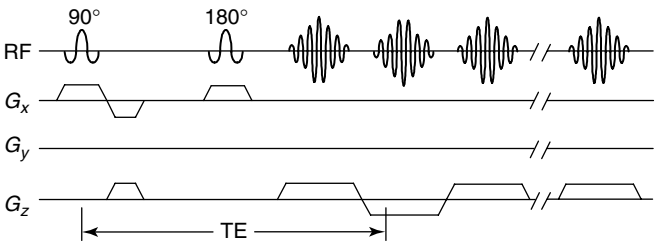

Figure 22 EPI pulse sequence, which allows the entire image to be collected with a single RF pulse. By flipping the z gradient between positive and negative values, a complete collection of echoes can be recorded..

in multiple phase-encoding steps, EPI collects all the data from a single excitation. To produce multiple echoes, the readout gradient is rapidly switched from positive to negative magnitudes. To achieve phase encoding, either the phase-encoding gradient is pulsed for each reversal of the readout gradient, or the phase-encoding gradient remains constant during data acquisition. To collect all of the data in one pulse run, the effective TEs tend to be very long. This causes the signal to be heavily T_2^* weighted, increasing the sensitivity of EPI to BOLD contrast.

Functional imaging has been used to characterize neurodegenerative diseases, such as Alzheimer's and Parkinson's, and cognitive disorders. Functional imaging has also been applied towards patients undergoing surgery to remove brain tumors. Brain tumors are unlike most other neoplastic growths in that the degree of malignancy is not the most important factor. Instead, the location of the tumor is more relevant to determining the treatment options and probable outcomes.[93] If located in a vital area of the brain, even the most benign tumors can be incapacitating or fatal. Surgery designed to remove the tumor must always weigh the benefits of removing as much of the tumor as possible against the risks of damaging essential brain tissue. Under pathological conditions, the location of brain functions may differ substantially from the normal neuroanatomic predictions. Without knowing where vital brain functions reside, the surgeon must be supremely careful in removing the tumor. However, functional MRI allows the doctor to map the exact locations of vital brain functions noninvasively before surgery. The surgeon can then make a better-informed decision, allowing the removal of the maximum amount of diseased tissue with the minimum risk of postoperative neurological deficits.

2.5 Safety Issues

Although MRI does not use ionizing radiation to produce images there are still some important safety considerations with which one should be familiar. These

relate to the use of strong magnetic fields, RF energy, and magnetic field gradients.[94]

2.5.1 Main Magnetic Field

The main magnetic field of a typical MRI scanner is about 30 000 times the strength of the earth's magnetic field. This strong magnetic field can literally pick up and pull ferromagnetic objects such as stretchers, oxygen bottles, fork-lift tines, heavy-duty floor buffers, mop buckets, and so on. The force exerted on a large metal object can injure or kill an individual. It can also seriously damage the magnet and imaging coils. A magnet can also pull smaller objects such as pagers, bobby pins and pens and injure a person or distort the magnetic field homogeneity.

Individuals with aneurysm clips, shrapnel or older ferromagnetic implants should not be imaged because magnetic forces can pull on these objects, cutting and compressing healthy tissue. The strong field also affects devices such as pacemakers. The magnetic reed switch in modern pacemakers is disturbed by strong magnetic fields resulting in possible deleterious effects to the patient. Some types of heart valves (e.g. Star–Edwards) are torqued in a magnetic field; however, this torque is less than the stresses that occur normally as a result of blood flow. Therefore, heart valves are not now considered an absolute contraindication for MRI.

Magnetic fields will also erase the magnetic strip of credit cards and magnetic storage media. Mechanical watches will "freeze up" in a strong field, sometimes permanently. Some metallic objects that are usually safe near a MRI machine are gold jewelry and eyeglass frames.

Although there is no evidence that magnetic and electric fields associated with MRI interfere with human development, caution must be exercised in MRI examination of patients who are pregnant. As with all interventions in pregnancy, MRI should be used during the first trimester when it offers a definite advantage over other tests.

The United States Food and Drug Administration (USFDA) safety guidelines state that field strengths not exceeding 2.0 T may be routinely used. Magnetic fields as large as 8 T, however, are currently employed for human research purposes. People with pacemakers must not be exposed to magnetic fields greater than 5 G. A 50 G magnetic field will erase magnetic storage media.

2.5.2 Radiofrequency RadiationRF

The RF power used in MRI is equivalent to the transmission of a small radio station (15–20 kW) and can cause heating of the tissue. The USFDA recommends that the exposure to RF energy be limited. The specific absorption rate (SAR) is the limiting measure:

$$\text{SAR} = \text{joules of RF}\,\text{s}^{-1}\,\text{kg}^{-1}\ \text{body weight} = \text{W}\,\text{kg}^{-1}$$

The SAR for the whole body must be less than $0.4\,\text{W}\,\text{kg}^{-1}$. It must be less than $3.2\,\text{W}\,\text{kg}^{-1}$ averaged over the head. Any pulse sequence must not raise the temperature by more than 1 °C and be no greater than 38 °C in the head, 39 °C in the trunk, and 40 °C in the extremities.

2.5.3 Varying Magnetic (Gradient) Fields

Time-varying magnetic fields induce electrical currents in conductors, as with an electrical generator. In patients with metallic implants, the potential exists for electrical currents to be induced in the metal, causing subsequent heating. Very rapidly changing magnetic fields, as may be achieved with EPI, have been shown to cause nerve stimulation.

The USFDA recommendations for the rate of change of magnetic fields state that the dB/dt for the system must be less than that required to produce peripheral nerve stimulation. Imaging gradients may also produce high acoustic noise due to mechanical motion of the gradients caused by switching high currents within the main magnetic field. The American Occupational Safety and Health Administration (OSHA) limits the peak acoustic noise to 200 Pa or 140 dB referenced to 20 mPa. Often, patients are instructed to wear protective earplugs to reduce the discomfort and possible hearing damage that can result from an MRI examination.

3 SPECTROSCOPY

MRS is an important aspect in the field of medical NMR, enabling much earlier identification of some diseases. With MRS, one is able to look inside the cell at its steady state metabolic processes, essentially performing a noninvasive chemical analysis. The information gathered can be interpreted spatially. As MRS is noninvasive, the same tissue can be examined again and again, thus making it an important tool in the diagnosis, prognosis and therapy monitoring of disease processes. Some of the more common nuclei that have been exploited by NMR to study biological tissue are listed in Table 2.

3.1 Localization Techniques

Clinical spectroscopy often requires the signal to be collected from only one particular area of tissue. Various methods have been devised to localize the acquired signal.

Table 2 Gyromagnetic ratio, spin quantum number, natural abundance, relative sensitivity, and absolute sensitivity of selected biologically important NMR nuclei

Nucleus	Gyromagnetic ratio, γ, (MHz T^{-1})	Spin quantum number, I	Natural abundance (%)	Relative sensitivity (%)	Absolute sensitivity (%)
^{1}H	42.57	1/2	99.98	100.0	100.0
^{19}F	40.05	1/2	100.00	83.0	83.0
^{23}Na	11.26	3/2	100.00	9.25	9.25
^{31}P	17.23	1/2	100.00	6.63	6.63
^{13}C	10.71	1/2	1.11	1.59	0.0176

The most common techniques include surface coil acquisition, depth-resolved surface coil spectroscopy (DRESS), image-selected in vivo spectroscopy (ISIS), stimulated echo acquisition mode spectroscopy (STEAM), point-resolved spectroscopy sequence (PRESS), and chemical shift imaging (CSI).

3.1.1 Surface Coil Acquisition

One of the easiest methods of localization involves the use of a surface coil, as introduced by Ackerman et al..[17] A surface coil only excites and receives signals from the immediate vicinity of the coil. By altering the size of the surface coil, the active volume can be changed. By moving the coil, a different area may be investigated. The main advantage of surface coil spectroscopy is the simplicity. No magnetic field gradients or specialized pulse sequences are required. In addition, a surface coil is one of the easiest RF coils to design and construct. A surface coil also offers superb SNR and filling factor. However, surface coils have several disadvantages. The surface coil must be in close physical contact with the tissue to be studied. This may even require the coil to be placed within the body (e.g. an endorectal coil used to study the prostate), greatly undermining the noninvasive nature of MRI. In addition, moving the coil to investigate other areas may be impractical. A surface coil does not produce a uniform RF field, being more sensitive to nearby tissues and less sensitive to deeper tissue. This may provide incomplete and unreliable localization.

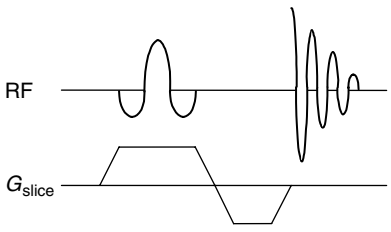

Figure 23 A pulse sequence used in combination with a surface coil for DRESS. The combination of a slice-selective pulse and a gradient perpendicular to the plane of the surface coil only excites nuclei in a disk-shaped volume parallel to the coil..

3.1.2 Depth-resolved Surface Coil SpectroscopyDRESS

A DRESS sequence, as shown in Figure 23,[95] has been developed to improve the localization characteristics of surface coils. This sequence uses a slice-selective pulse and a gradient perpendicular to the plane of the surface coil. Only nuclei in a disk-shaped volume parallel to the coil are excited and contribute signal. The thickness of the disk depends upon the bandwidth of the pulse and the magnitude of the gradient. The position is determined by the physical location within the magnet and the center frequency of the pulse. The DRESS sequence allows total suppression of signals arising from superficial tissues. This technique can be extended to acquire multislice data from the sensitive volume of the surface coil. This is called slice interleaved (SLIT) DRESS, also known by the acronym SLIT/DRESS.[96]

3.1.3 Image-selected In Vivo SpectroscopyISIS

The ISIS method was developed by Ordidge et al..[97] and is based on the selective inversion of magnetization from the volume of interest. In this method, data from a selected plane is obtained by acquiring two global spectra of the entire region excited by the RF coil. As shown in Figure 24, in the first experiment, the total signal is acquired from all nuclear spins following application of a single RF pulse. The second experiment inverts the spins in the selected slice with a slice-selective 180° pulse in combination with a magnetic field gradient. After a suitable inversion delay, this preparation pulse is followed by a signal-acquisition pulse as in the first experiment. Upon subtraction of the FID from the two experiments, the signals from the selected plane are added, whereas those from outside the plane cancel each other out. In a similar way, selection of a cube is achieved by incorporating three selective-inversion pulses, each in the presence of a magnetic field gradient along the x, y, and z axes. Eight experiments with different permutations of the three selective-inversion pulses are required for 3-D localization.

The main advantage of ISIS is that the spectra have very little T_2 weighting compared to other 3-D localization

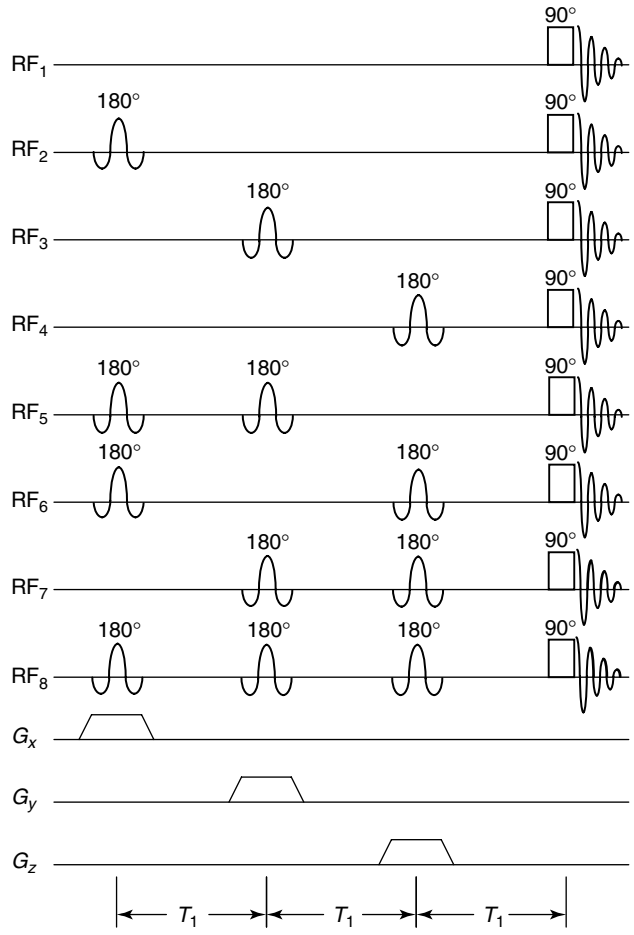

Figure 24 A set of eight pulse sequences used for ISIS. The signal is localized to a single cubic-shaped voxel by adding FIDs (free induction decays) from sequences 1, 5, 6 and 7 while subtracting FIDs from sequences 2, 3, 4 and 8. The resulting signal originates from only the intersection of the three orthogonal planes..

techniques. This allows localized spectra to be collected from metabolites with relatively short T_2 values, such as ^{31}P metabolites. Because signals are collected from each area and then added or subtracted from the total signal, ISIS is very sensitive to movement. If the patient moves between one acquisition and the next, the localization will be insufficient.

3.1.4 Point-resolved Spectroscopy

The PRESS pulse sequence seen in Figure 25 can achieve localization within a single acquisition.[98] The PRESS technique uses one slice-selective 90° pulse and two slice-selective 180° pulses to form a spin echo. The combination of the band-limited RF pulse and the magnetic field gradient only excites spins in a thin plane along the gradient axis. The three slice-selective pulses are aligned along the three cardinal directions, so that only spins at

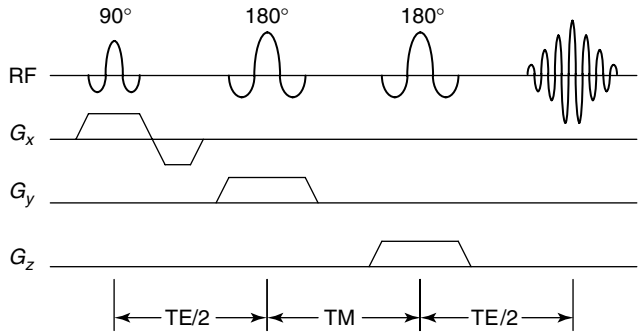

Figure 25 Pulse sequence for PRESS. The combination of the 90° and two 180° pulses forms a spin echo originating from the intersection of the three slice-selective gradients..

the intersection of these three planes contribute to the echo signal. This pulse sequence is suitable for nuclei that have long T_2 relaxation such as ^{1}H.

3.1.5 Stimulated Echo Acquisition Mode

Another single-shot localization pulse sequence is STEAM.[99] The STEAM pulse sequence shown in Figure 26 uses slice-selective pulses much like the PRESS sequence. However, STEAM uses three slice-selective 90° pulses to produce a stimulated echo from the spins at the intersection of the three planes. Because STEAM relies on the production of a stimulated echo, only half of the available magnetization can be acquired.[100] Thus, STEAM only produces half as much signal as a PRESS pulse sequence. However, STEAM allows a shorter TE than PRESS. Therefore, when studying compounds with short T_2 times, such as phosphorus metabolites, a STEAM pulse sequence may be preferable.

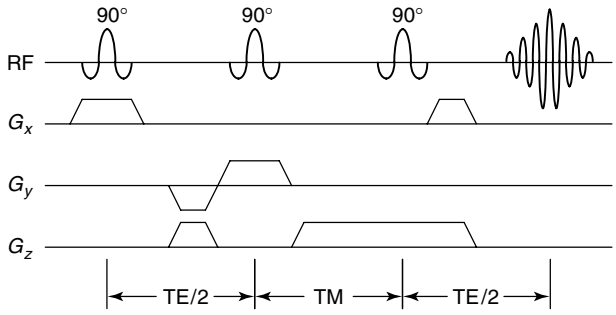

Figure 26 Pulse sequence for STEAM localization spectroscopy. The combination of the three 90° pulses forms a stimulated echo from a single voxel corresponding to the intersection of the three slice-selective gradients. As a stimulated echo is formed, only half of the available signal is obtained. A STEAM pulse sequence can achieve shorter TEs than a PRESS pulse sequence..

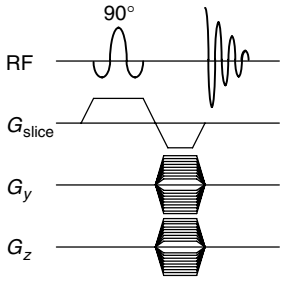

Figure 27 CSI pulse sequence. Spatial encoding in three dimensions is achieved with a slice-selective pulse and two phase-encoding gradients. Because there is no frequency-encoding gradient, spectral information is retained in the acquired signal..

3.1.6 Chemical Shift ImagingCSI

CSI, reported by Maudsley et al..[101] and Brown et al..,[102] is a technique for obtaining multiple localized spectra, simultaneously, from a set of voxels spanning the region of interest. As seen in Figure 27, it uses the same phase-encoding principles as 2-D and 3-D imaging techniques. Each spin is spatially encoded by phase-encoding gradients. There are no gradients during data collection, so frequency (spectral) information from each detectable compound is preserved. Although CSI can be used with one, two, or three dimensions of phase encoding, the simplest method is a 1-D CSI sequence with a surface coil to define the measured volume.

3.2 ^{31}P Nuclear Magnetic Resonance SpectroscopyNMR

Phosphorus NMR spectroscopy has been widely used to monitor noninvasively changes in high-energy metabolites. Several important metabolites can be observed in ^{31}P spectra including phosphocreatine (PCr), inorganic phosphate (P_i), adenosine triphosphate (ATP), phosphomonoesters (PME), and phosphodiesters (PDE). In vivo ^{31}P spectra from various vital organs are shown in Figure 28. The PCr resonance is normally set to 0.0 ppm. Thus, P_i is found at 3.7–5.2 ppm, γ-ATP around −2.7 ppm, α-ATP around −7.8 ppm, β-ATP around −16.3 ppm, and PDE 2–3 ppm from PCr.

Phosphorus has been widely used because of the importance of the observed signals. The major metabolite used for biological work including muscle contraction and ion transport across the cellular membrane is ATP. It is also used for the synthesis of many other required compounds, such as carbohydrates, proteins, lipids, and nucleic acids. However, PCr is often used in the body as a high-energy storage compound. PCr and ADP combine to form ATP and creatine (Cr). Thus, depleted stores of ATP can be restored by consuming PCr and creating Cr. The ratio of PCr to ATP has historically been used as an early indication of metabolic disturbance. A drop in oxidative phosphorylation will cause a decrease in the PCr/ATP ratio, indicating an impaired energy status. ^{31}P-NMR can also be used to quantitate tissue metabolites, calculate phosphorylation potential, measure free Mg^{2+} concentrations and measure chemical exchange.

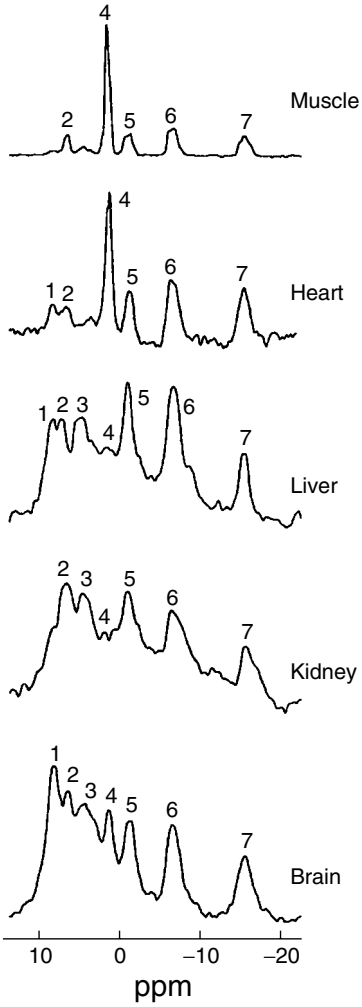

Figure 28 Phosphorus spectrum obtained from the skeletal muscle, heart, liver, kidney and brain. Peak assignments are as follows: 1, PME; 2, P_i; 3, PDE; 4, PCr; 5, γ-ATP; 6, α-ATP; 7, β-ATP.

Another strength of ^{31}P spectroscopy lies in the fact that the chemical shift of the P_i peak is sensitive to pH, has a pK_a value close to physiological pH, and is relatively insensitive to other ionic concentrations. In addition, P_i shows no spin–spin coupling and is found in adequate tissue concentrations.[103] P_i is found in the intracellular compartment, so pH measurements based on the chemical shift of P_i represent intracellular pH. The pH can be calculated from a measurement of the chemical shift of P_i relative to another peak. The reference peak is

generally chosen to be PCr, resulting in the Equation (4):

$$\mathrm{pH} = 6.75 + \log\left[\frac{\delta_{\mathrm{P_i-PCr}} - 3.28}{5.69 - \delta_{\mathrm{P_i-PCr}}}\right] \quad (4)$$

Alternatively, the α-ATP peak may be used as a chemical shift reference, leading to a slight modification (Equation 5):

$$\mathrm{pH} = 6.75 + \log\left[\frac{\delta_{\mathrm{P_i-\alpha\text{-}ATP}} - 10.85}{13.25 - \delta_{\mathrm{P_i-\alpha\text{-}ATP}}}\right] \quad (5)$$

Phosphorus NMR has been used to diagnose and characterize a number of different disorders throughout the body (including the heart, skeletal muscle, and brain). In addition, the ^{31}P signals have been used to monitor therapies designed to reverse the abnormal metabolic conditions.

A number of technical advancements, such as cardiac gating, phased array surface coils for maximum sensitivity and volume coils for uniform excitation, have allowed ^{31}P spectroscopy to be applied to heart disease. Significantly lower PCr/ATP ratios compared to healthy heart tissue can be seen in muscular dystrophy,[104] myocardial ischemia,[105,106] transplanted hearts,[107] and cardiac amyloidosis.[104] In 1992, Neubauer et al.. found a significant correlation between decreased PCr/ATP ratio and the New York Heart Association (NYHA) clinical classification of heart failure. Upon treatment, an improved NYHA classification was accompanied by an increase in PCr/ATP ratio.[108]

Phosphorus NMR spectroscopy can determine the metabolic state of skeletal muscle both at rest and during exercise. At rest, a low ratio of PCr to P_i may be indicative of mitochondrial myopathies, hypothyroidism, muscular dystrophy, myotonic dystrophy, denervating disorders or diabetes.[109–113] Several enzyme deficiencies may also be diagnosed by ^{31}P-NMR in conjunction with exercise. An unusually high rate of acidosis accompanied by a large drop in the PCr/P_i ratio indicates a disorder in an enzyme of the oxidative pathway. This disorder causes ATP to be produced by glycolysis, forming lactic acid and, hence, causing the high rate of acidosis. Conversely, a low rate of acidosis and an increase in the PME signal indicates disruption of one of the enzymes in the glycolytic pathway. As another example, spectra obtained from patients with McArdle's syndrome show rapid depletion of PCr, a low rate of acidosis, and no increase in the PME signal. These patients suffer from a disorder in which glycogen phosphorylase is missing, blocking glycogen metabolism, and causing rapid fatigue during exercise.[114] A phosphofructokinase deficiency can cause the PCr peak to decrease, the P_i peak to increase, and the PME peak to increase with no acidosis during exercise.[115] ^{31}P spectroscopy may be used as a screening tool to detect these disorders in neonates, allowing immediate administration of appropriate treatments.

Phosphorus spectroscopy has also been applied to a wide range of clinical applications in the brain. After birth asphyxia, a ^{31}P spectrum of the newborn brain shows reduced energy state and is predictive of outcome.[116–122] In adult stoke, ^{31}P spectra can be used to localize the extent of energy failure[123–125] and may indicate which areas are at risk of future strokes if not properly treated.[126]

Phosphorus NMR has also been used in the study of tumor diagnosis, progression and therapy. For many years tumors were assumed to have an acidic pH due to the conversion of carbohydrate into lactic acid. The advent of ^{31}P-NMR for noninvasive measurement of pH, however, has shown that tumors have an alkaline pH.[127] In addition, many human tumors show low PCr, high PME, and high PDE signals.[128] Changes in the ^{31}P spectra may be used to predict and monitor response to radiotherapy or chemotherapy. There is currently a 5-year eight-center study underway to determine the utility of ^{31}P during cancer treatment.

3.3 Water Suppression and ^{1}H Spectroscopy

Proton NMR spectroscopy has historically lagged behind ^{31}P spectroscopy despite the superior sensitivity and ease of application on clinical scanners. The fundamental difficulty in proton spectroscopy lies in the fact that the water ^{1}H signal, representing a concentration of 80 M in body tissues, greatly overshadows the signals arising from metabolites, often found in concentrations of about 1 mM. To achieve reasonable resolution of these small metabolite signals, the water signal must be suppressed from the spectrum. To exploit the high sensitivity of proton NMR spectroscopy clinically, many water-suppression techniques have been developed.

A common method of water suppression is selectively to saturate the water signal prior to obtaining the NMR spectrum.[129,130] This technique, called presaturation, requires a low-power RF pulse to be applied for a long duration followed by the normal spectroscopy sequence. The presaturation pulse must selectively excite only the water region of the spectrum. The pulse sequence then only excites the signals away from the water signal. Presaturation is very attractive for it's simplicity and ease of implementation. However, this technique also saturates signals close to the water peak. In addition, the exchangeable protons are also saturated and cannot be studied. Care must also be taken because presaturation pulses raise the SAR and may cause significant heating.

Another method of water suppression uses specialized RF pulses selectively to excite all of the signals except for the water. These pulses are sometimes called binomial pulses,[131,132] but many variations exist. The pulse is designed to sum to an effective 0° pulse for water, but a 90° pulse for all other signals. As the water is not saturated, exchangeable protons are visible by this method. Heating and RF safety concerns are not limiting factors as they are with presaturation methods.

A popular water suppression method that utilizes both selective pulses and magnetic field gradients is known as CHESS (chemical shift selective saturation).[133] A frequency-selective 90° pulse is applied to only the water signal followed by the application of a gradient that dephases the water signal. After repeated application of frequency-selective 90° pulses and gradients, very little water signal is detected. This method is easily implemented on most clinical scanners and reduces exposure to RF radiation.

The brain is ideal for ^{1}H-NMR studies because it is relatively stationary, uniform, spherical, and has low lipid concentrations within the tissues. Representative ^{1}H-NMR spectra with water suppression from normal and abnormal brain tissue are seen in Figure 29. Several biomedically important metabolites can be seen in these spectra.

The most intense peak in the ^{1}H spectrum from the normal brain is NAA at 2.02 ppm. As NAA is thought to exist only in neuronal cells,[134] it can be used to grade neuronal loss. Little or no NAA is found in brain tumors[135–138] or ischemic tissues following a stroke[139–145] as seen in Figure 29. In addition, levels of NAA decrease during progression of demyelinating diseases.[146] Levels of NAA can be significantly altered in diseases such as multiple sclerosis,[147–152] HIV,[153–155] and epilepsy.[156–163]

Analysis of NAA signal intensity can be complicated by several other resonances that overlap the NAA peak. In particular, the C(3) glutamate and glutamine peaks can add a significant amount of signal in spectra collected with short TE times. These contributions can be corrected by deconvoluting the NAA peak with nonoverlapping glutamate and glutamine peaks. The NAA peak also overlaps with *N*-acetyl glutamate in white matter.

Another large signal in the ^{1}H spectrum arises from Cho at 3.2 ppm. Cho is a component of the phospholipid metabolism and is found in cell membranes. The Cho peak may indicate membrane synthesis and cellular density. Increased Cho levels are found in tumors,[135–138] as seen in the spectra of Figure 29. Cho is also elevated in metabolic white matter diseases[146] and multiple sclerosis.[134,148,152] Demyelinating diseases are often associated with increased levels of Cho.[146,149] This may indicate that Cho can be released from NMR-invisible stores in myelin.

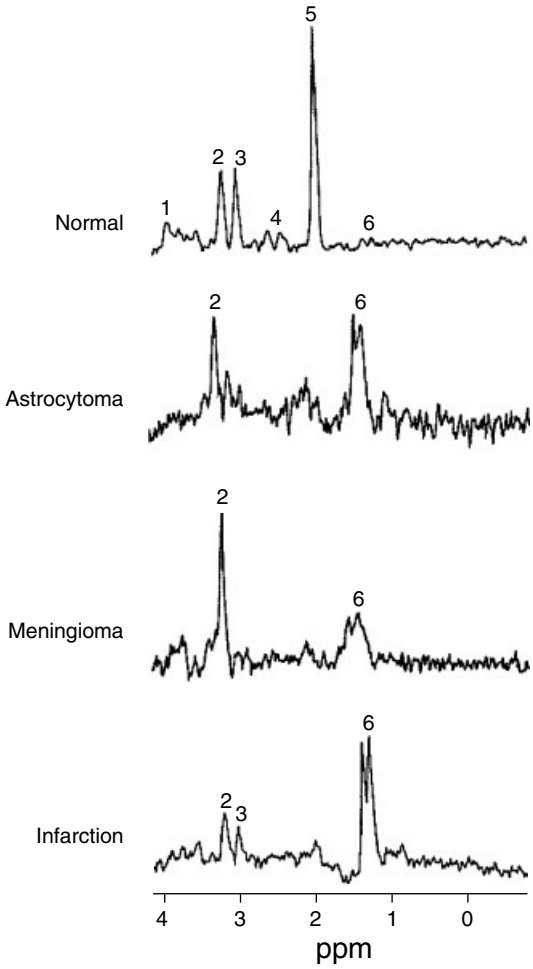

Figure 29 Proton spectra obtained with water suppression from the normal brain, an astrocytoma, a meningioma, and a cerebral infarction. Peak assignments are as follows: 1, PCr and Cr; 2, choline (Cho); 3, PCr and Cr; 4, NAA (*N*-acetylaspartate); 5, NAA; 6, lactate. (Spectra courtesy of Siemans Medical Systems, Inc.).

However, the Cho peak contains signals from free Cho, glycerol 3-phosphorylcholine, and phophorylcholine. Additional metabolites, such as phosphatidyl Cho, taurine, or betaine, may be found at 3.2 ppm under certain pathophysiological conditions. Analysis of ^{1}H and ^{31}P spectra may help separate some of the different contributions.

Lactate has been one of the most popular metabolites in ^{1}H spectroscopy. The lactate methyl doublet resonates at 1.33 ppm. Lactate, as a product of glycolysis, is used as an indicator of insufficient O_2 supply. As seen in Figure 29, very little is found in normal brain tissue, but lactate levels may increase 10–20 times in acute ischemia, such as immediately following a stroke, or as

a result of chronically reduced perfusion, such as found in tumors.[164–166] The regions of highest lactate content correspond with lowest NAA levels after stroke,[140,142–144] correlating greatest neuronal loss with the areas suffering from the lowest O_2 supply. Changes in lactate can also be used to monitor tumor growth or response to therapy. Macrophages are also believed to produce large amounts of lactate.[167–169] As macrophages quickly infiltrate the damaged tissue and remain for prolonged periods,[170–172] lactate levels may rise accordingly.

Cr, a metabolic marker, is seen at 3.03 ppm in ^{1}H spectra. Other resonances are also found at this chemical shift including PCr, γ-aminobutyric acid (GABA) and cytosolic macromolecules. Like NAA, the Cr resonance is decreased in the infarcted brain spectrum in Figure 29, indicating reduced metabolic activity. Cr levels can also be used to characterize tumor metabolism. Hypometabolic tumors show increased levels of Cr while hypermetabolic tumors show lower levels.[173]

In addition to analysis of elevated signals, the ratio of one signal to another can be used in the diagnosis. For example, spectra from AIDS patients show a characteristic reduction in the ratio of NAA to Cr and an increase in Cho to Cr.[174] These spectroscopic changes may be more sensitive to brain lesions than conventional MRI methods. Proton NMR has shown great promise towards distinguishing between low-grade and high-grade primary brain tumors. The degree of Cho elevation, lactate formation and reduction in Cr has been used to discriminate between various types of tumors. Preul et al.. correctly classified 104 of 105 proton spectra from normal volunteers and patients with the five most common adult supratentorial brain tumors using pattern recognition analysis.[175]

3.4 Carbon-13 Spectroscopy and Decoupling

Historically, complex metabolic pathways have been studied using radioactive carbon isotopes such as ^{11}C and ^{14}C. These techniques require several chemical or biochemical manipulations including quenching the cellular reactions, chromatographic separation, and carbon by carbon degradation of specific metabolites.[176] ^{13}C-NMR spectroscopy can yield similar information without the need for such complex maneuvers.

Carbon-13 is a stable carbon isotope with spin 1/2 and 1.1% natural abundance. The low occurrence of natural ^{13}C is both an advantage and a disadvantage. The low natural abundance yields low signals without enrichment, causing low SNRs and/or very long acquisition times. With the addition of enriched substrates, however, the lack of naturally occurring ^{13}C is a benefit. There is little interference from background signals and practically all of the signal observed can be assumed to originate from enriched substrates. Carbon-13 can be substituted into metabolic substrates without altering their chemical characteristics to follow metabolic pathways, measure the rates of specific pathways or to detect deviations from normal metabolism.[176] Despite the low sensitivity of ^{13}C-NMR, in vivo human experiments have been performed. To obtain a greater SNR, tissue extraction for high-resolution spectroscopy may be required.

NMR spectroscopy can reveal molecular structures through a phenomenon known as spin–spin coupling, also called J coupling or scalar coupling. Nuclei connected by chemical bonds exchange information about their magnetic states, causing a peak to split into a particular pattern of multiple smaller peaks. The separation between the peaks in these patterns is independent of the field strength. Therefore, spin–spin coupling is reported in terms of hertz. Homonuclear coupling occurs between two nuclei of the same species, that is, $-^{13}C^{13}C$. Heteronuclear coupling occurs between two nuclei of different species, that is, $-^{1}H^{13}C$. Chemically equivalent nuclei, however, do not produce spin–spin coupling. Heteronuclear coupling between protons and ^{13}C nuclei decreases the signal intensities of ^{13}C resonances. This reduces the available SNR and increases the complexity of ^{13}C spectra. Decoupling is achieved by supplying power at the ^{1}H frequency while collecting data at the ^{13}C frequency. This collapses the ^{13}C signals arising from single lines, increasing the SNR of the spectrum. A number of decoupling schemes, such as WALTZ (for wideband alternating phase low-power technique for zero residue splitting) and MLEV (named after the developer of the technique, Malcolm H. Levitt), are available to decouple a wide range of ^{1}H frequencies.[177] In addition, low power at the ^{1}H frequency can be applied in continuous wave mode at all other times to induce the nuclear Overhauser effect (NOE) causing a further increase in the SNR. To collect proton-decoupled ^{13}C spectra, either a dual-tuned $-^{1}H^{13}C$ coil or two separate singly tuned coils must be used.

Carbon-13 NMR has been used to detect natural abundance glycogen in the calf, liver, and brain.[178–183] As the principal means of carbohydrate storage in humans, glycogen provides a reservoir of glucose during strenuous muscular activity. The synthesis and degradation of glycogen plays a major role in the regulation of blood glucose levels during various metabolic conditions, such as rest, exercise, fed and fasted states.

Several inherited enzyme defects can lead to glycogen storage diseases. For instance, type IIIA glycogen storage disease causes increased glycogen concentrations in the liver and muscle and abnormal glycogen structures with short outer branches. In healthy normal subjects, the ^{13}C concentration of glycogen is of the order of 0.7–1.4 mM which borders the limit of NMR sensitivity.

Trained athletes or normal untrained subjects with special diets have increased glycogen concentrations detectable by NMR. In contrast, patients with type IIIA glycogen storage disease have approximately 2–3 times the amount of glycogen compared with trained athletes or subjects with special dietary preparation. In addition, the spectra obtained from patients with type IIIA glycogen storage disease show extra signals presumably due to the abnormal glycogen structure.[178] The conventional procedure for measuring glycogen levels is needle biopsy,[184,185] which is uncomfortable and may not reliably reflect changes during an extended sequence of events.

Carbon-13 NMR spectroscopy has also been applied to metabolism in the heart. The heart uses a mixture of fatty acids, lactate, and ketones to provide acetyl-CoA (coenzyme A) for entry into the citric acid cycle. The predominant substrate and the rates of metabolic flux are related to the health of the tissue.[186,187] Carbon-13 NMR allows distinction of normal from ischemic myocardium based upon the metabolic characteristics of the tissue. These methods may also be useful for prediction of a tissue's tolerance to ischemia. Suppression of fatty acid utilization has been shown to improve functional characteristics after reperfusion.[188,189]

A simple and straightforward analysis can be made from the homonuclear $-^{13}C^{13}C$ coupling patterns observed in ^{13}C spectra. As $-^{13}C^{13}C$ coupling patterns must be resolved, this method has only been performed with chemical extracts and isolated perfused organs to achieve high magnetic field homogeneity and SNR. This technique is known as isotopomer analysis.[190,191] For instance, with metabolism of [1,2-^{13}C]acetate and [3-^{13}C]lactate, the C(4) resonance of glutamate appears as nine lines that are due to $-^{13}C^{13}C$ coupling. Figure 30 shows a spectrum from an extract illustrating the resonance pattern including a singlet, C(4)S, two doublets with slightly different coupling constants, C(4)D34 and C(4)D45, and a quartet, C(4)Q. Two signals, C(4)S and C(4)D45, arise only from glutamate that is not enriched in the C(3) position. The other two signals, C(4)D34 and C(4)Q, can only arise when C(3) is enriched with a ^{13}C nucleus. The sum of the C(4)S and C(4)D34 resonance areas versus the sum of the C(4)D45 and C(4)Q resonance areas provides a direct measurement of the relative utilization of [3-^{13}C]lactate versus [1,2-^{13}C]acetate.[192]

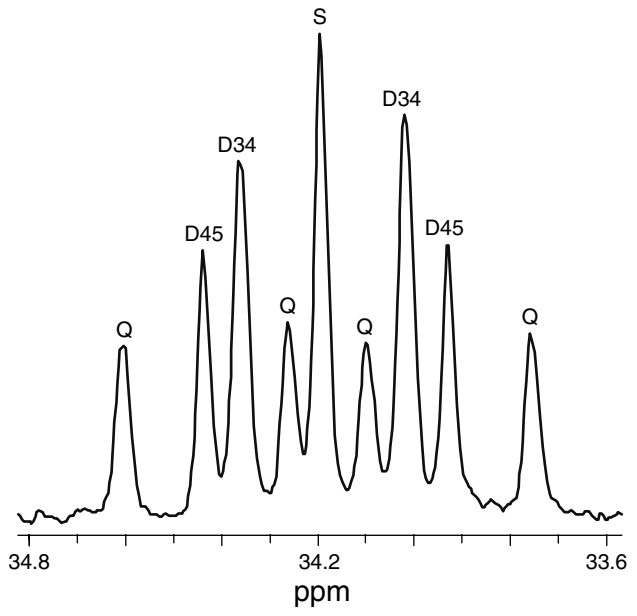

Figure 30 Typical ^{13}C spectrum collected from a tissue extract for isotopomer analysis. The resonance pattern shows a singlet, C(4)S, two doublets with slightly different coupling constants, C(4)D34 and C(4)D45, and a quartet, C(4)Q. (Spectrum courtesy of Dr. Craig R. Malloy, University of Texas, Southwestern Medical Center.).

3.5 Sodium-23 Spectroscopy

Healthy cells maintain an intracellular Na^+ concentration of 10–25 mM against an extracellular Na^+ concentration of 150 mM. The transmembrane Na^+ gradient is regulated to a large part by the action of the $-Na^+K^+$ ATPase, which pumps three Na^+ ions out and two K^+ ions in for each ATP molecule. The Na^+ gradient is required to drive many vital cellular processes, including maintenance of cell volume, facilitated transport of other ions and metabolites, contractile function, and transmission of nerve impulses. Many diseases, including hypertension,[193] manic depression,[194] cystic fibrosis,[195] cancer,[196] and sepsis[197] are associated with disruption of the transmembrane Na^+ gradient due to alterations in ion transport and exchange processes. Several methods have been developed to study Na^+ in biological samples, such as ion-selective microelectrodes, electromicroprobe X-ray analysis, fluorescent dyes, and whole tissue analysis.[196–198] However, only NMR has the potential for measuring Na^+ noninvasively.

Unlike the nuclei discussed so far, Na^+ has a spin quantum number of 3/2, resulting in biexponential T_2 relaxation in biological tissues. The fast relaxation component is of the order of 2 ms and represents up to 60% of the entire Na^+ signal. Civan and Shporer reported that 60% of the sodium in biological samples was invisible.[199] This may indicate that the percentage of sodium in biological samples with short T_2 relaxation cannot be detected with long spectrometer delay times or with long TEs. However, recent studies have shown identical results between NMR and conventional quantitative analysis, proving that Na^+ is 100% NMR visible.[200]

Because Na^+ exists only as an aqua cation in tissue, both intra- and extracellular ^{23}Na signals are coincident.

To study the transmembrane Na^+ gradient, the intra- and extracellular signals must be separated. Three common methods to differentiate between intra- and extracellular ^{23}Na utilize paramagnetic shift reagents, relaxation differences and multiple quantum (MQ) filters.

Paramagnetic shift reagents alter the chemical shift of extracellular ^{23}Na relative to intracellular ^{23}Na. The shift reagent thulium-1,4,7,10-tetraazacyclodecane-1,4,7,10-tetrakis(methylene phosphonate) ($TmDOTP^{5-}$) can be used to separate intra- and extracellular ^{23}Na signals in animal experiments,[200–204] but no shift reagent is currently available for clinical use.

Hilal et al.. used differences in relaxation times to distinguish between intra- and extracellular Na^+ in brain tumors.[205] Tumors are known to have increased concentrations of intracellular Na^+. The higher Na^+ concentration is needed to maintain proliferation and is related to the grade of malignancy.[206] Because of the high intracellular protein content, intracellular Na^+ has a faster T_2 decay than extracellular Na^+.[199,207] Hilal et al.. used images derived from FIDs and various TEs to map the fast decay fraction of tissues.[205] They produced putative intracellular Na^+ maps from the fraction of fast-decaying ^{23}Na and saw an excellent correlation between the putative intracellular Na^+ and tumor grade. They were able to differentiate between tumors, radiation necrosis, gadolinium-enhanced benign lesions, and gadolinium-enhanced malignant lesions. They reported that highly malignant gliomas (grade III) have intracellular Na^+ concentrations in the range 75–55 mM, whereas low-grade gliomas (grade I) have intracellular concentrations of 35–45 mM.

These sodium maps have shown several advantages compared to the conventional methods of proton MRI. The sodium techniques frequently reveal a much larger tumor than seen by proton MRI even with Gd-DTPA enhancement. The putative intracellular sodium images show high sodium concentrations only in neoplastic tissue. Edema, radiation necrosis and postoperative changes do not show up as bright areas. However, on T_2-weighted proton images, tumor, edema, and radiation necrosis are all hyperintense regions. Sodium MRI allows distinction between neoplastic and inflammatory tissues when proton MRI cannot resolve the two areas even with Gd-DTPA enhancement.[205]

The third method for discriminating between intra- and extracellular Na^+ uses MQ filters. As Na^+ has a spin quantum number of 3/2, there are four possible energy levels. This leads to three possible single quantum transitions, that is, $3/2 \leftrightarrow 1/2$, $-3/2 \leftrightarrow -1/2$ and $1/2 \leftrightarrow -1/2$. There are two possible double quantum transitions, that is, $3/2 \leftrightarrow -1/2$ and $1/2 \leftrightarrow -3/2$. There is also one triple quantum transition, that is, $3/2 \leftrightarrow -3/2$. The high concentration of macromolecules in the intracellular environment causes MQ transitions to originate mainly from Na^+ in the intracellular spaces. For this reason, MQ filters have been proposed to detect intracellular Na^+,[208] although some MQ-filtered signal does arise from the extracellular space.[209–211] In contrast to the single quantum Na^+ signal, however, the majority of the MQ signal comes from the intracellular space. In addition, the MQ-filtered signal is relatively insensitive to changes in extracellular Na^+ content.[212] These two facts lead to the MQ-filtered signal being useful for detecting and monitoring changes in intracellular Na^+. The low SNR available with MQ-filtered ^{23}Na-NMR has been the major obstacle in biological and clinical applications. Despite these difficulties, Dr. Boada at the University of Pittsburgh Medical Center has obtained triple quantum filtered ^{23}Na images of the human brain in about 18 min using a 3 T magnet. Representative single quantum and triple quantum filtered images can be seen in Figure 31.

3.6 Fluorine-19 Spectroscopy

Another nucleus that has been used is ^{19}F. Several advantages are associated with ^{19}F-NMR, such as 100% abundance, higher sensitivity than ^{31}P, wide spread of chemical shifts eliminating overlapping peaks, and no intrinsic background signals.[213–215] Fluorine-19 NMR has been used to monitor the metabolism of fluorinated medications. In particular, 5-fluorouracil (5-FU) can be detected as well as some of its metabolized products, including fluoro-β-alanine (FBAL) and some fluoronucleotides. Figure 32 shows ^{19}F spectra obtained from the liver of cancer patients during 5-FU treatment. The metabolism of 5-FU into FBAL is clearly seen. Thus, ^{19}F-NMR can be used to predict and monitor response to 5-FU therapies.[216–219]

Perfluorocarbons have historically been produced as potential blood substitutes. These compounds can be easily detected by ^{19}F-NMR. These compounds are

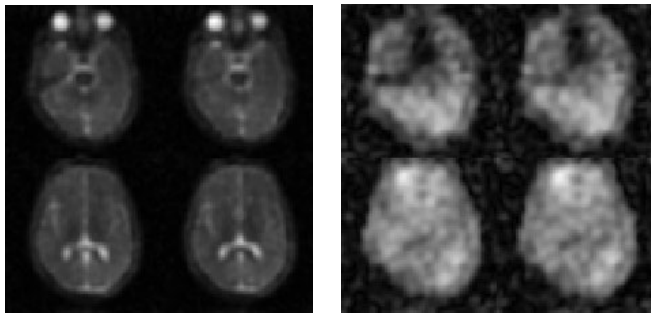

(a) (b)

Figure 31 (a) Single quantum and (b) triple quantum ^{23}Na images of the brain of a normal volunteer. (Images courtesy of Dr. Fernando Boada, University of Pittsburgh Medical Center.).

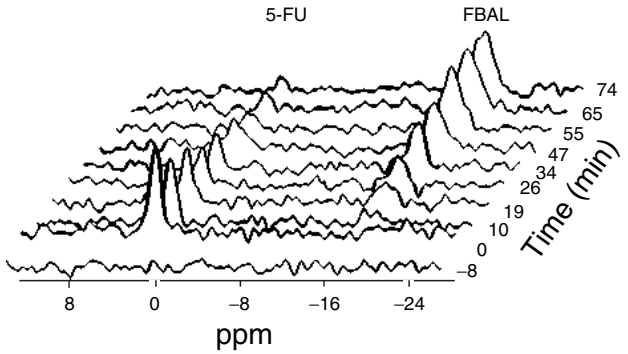

Figure 32 Fluorine-19 spectra obtained from the liver of a patient undergoing 5-FU treatment. The metabolism of 5-FU into FBAL is clearly seen. (Reproduced by courtesy of Siemans Medical Systems, Inc.).

also sensitive to many physiological parameters, such as oxygen tension, temperature, and pH.[213–224] The paramagnetism of oxygen alters the T_1 relaxation of the perfluorocarbon. The chemical shifts of various signals can also be dependent upon the temperature and/or the pH.

4 CONCLUSIONS

The application of NMR to the field of medicine has caused an explosion of new techniques and methods. The noninvasive nature of NMR has allowed stunning advancements in imaging the human body and investigating the biochemical properties of tissues. The distinct advantages of MRI over other imaging methods has caused it quickly to become the standard technique for imaging the brain, spinal cord, spine, joints, and other anatomical structures. The development of angiographic and functional imaging techniques proves that MRI is still developing and promises future advancements. Although not as routine as MRI, NMR spectroscopy has also proven useful in detecting and characterizing a wide variety of disease states. Although biomedical NMR is a young field it has gained unprecedented acceptance and is the method of choice for a wide variety of medical applications.

ACKNOWLEDGMENTS

The authors wish to thank the National Institutes of Health (HL 54574) and the Whitaker Foundation for providing financial support during the preparation of this chapter.

LIST OF SYMBOLS

B_0	the main magnetic field
B_1	the magnetic field produced by RF radiation transmitted by the RF coil
T_1	longitudinal relaxation time
T_2	transverse relaxation time
T_2^*	observed transverse relaxation time due to B_0 field inhomogeneities

ABBREVIATIONS AND ACRONYMS

ATP	Adenosine Triphosphate
BOLD	Blood Oxygen Level Dependent
CHESS	Chemical Shift Selective Saturation
Cho	Choline
CoA	Coenzyme A
Cr	Creatine
CSF	Cerebrospinal Fluid
CSI	Chemical Shift Imaging
CT	Computed Tomography
DRESS	Depth-resolved Surface Coil Spectroscopy
EPI	Echo-planar Imaging
FBAL	Fluoro-β-Alanine
FID	Free Induction Decay
FOV	Field of View
GABA	γ-Aminobutyric Acid
Gd-BOPTA-Dimeg	Gadolinium Benzyloxy Propionictetraacetate Dimeglumine
Gd-DTPA	Gadolinium Diethylenetriamine Pentaacetic Acid
Gd-EOB-DTPA	Gadolinium Ethoxybenzyl Diethylenetriamine Pentaacetic Acid
GRASS	Gradient-recalled Acquisition at Steady State
ISIS	Image-selected In Vivo Spectroscopy
Mn-DPDP	Manganese N,N'-bis(pyridoxal-5-phosphate)ethylenediamine-N,N'-diacetic Acid
MP/RAGE	Magnetization Prepared Rapid Gradient Echo Imaging
MQ	Multiple Quantum
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
NAA	N-Acetylaspartate

NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NYHA	New York Heart Association
OSHA	Occupational Safety and Health Administration
PCr	Phosphocreatine
PDE	Phosphodiesters
PET	Positron Emission Tomography
P_i	Inorganic Phosphate
PME	Phosphomonoester
ppm	parts per million
PRESS	Point-resolved Spectroscopy Sequence
RF	Radiofrequency
SAR	Specific Absorption Rate
SLIT	Slice Interleaved
SNR	Signal-to-noise Ratio
SPECT	Single Photon Emission Computed Tomography
STEAM	Stimulated Echo Acquisition Mode Spectroscopy
TE	Echo Time
$TmDOTP^{5-}$	Thulium-1,4,7,10-tetraaza-cyclodecane-1,4,7,10-tetrakis-(methylene phosphonate)
TR	Repetition Time
USFDA	United States Food and Drug Administration
2-D	Two-dimensional
3-D	Three-dimensional
5-FU	5-Fluorouracil

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Angiography • Magnetic Resonance Imaging, Functional • Magnetic Resonance in Medicine, High Resolution Ex Vivo • Multinuclear Magnetic Resonance Spectroscopic Imaging

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Multidimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. F. Bloch, W.W. Hansen, M.E. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
2. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
3. P.C. Lauterbur, 'Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance', *Nature (London)*, **242**, 190–191 (1973).
4. A. Kumar, D. Welti, R.R. Ernst, 'NMR Fourier Zeugmatography', *J. Magn. Reson.*, **18**, 69–83 (1975).
5. W.A. Edelstein, J.M.S. Hutchison, G. Johnson, T. Redpath, 'Spin Warp NMR Imaging and Applications to Human Whole-body Imaging', *Phys. Med. Biol.*, **25**, 751–756 (1980).
6. R.A. Challiss, M.J. Bleckledge, G.K. Radda, 'Spatial Heterogeneity of Metabolism in Skeletal Muscle In Vivo Studied by ^{31}P-NMR Spectroscopy', *Am. J. Physiol.*, **254**, C417–C422 (1988).
7. C.T. Burt, J. Koutcher, J.T. Roberts, R.E. London, B. Chance, 'Magnetic Resonance Spectroscopy of the Musculoskeletal System', *Radiol. Clin. N. Am.*, **24**, 321–331 (1986).
8. T. Jue, D.L. Rothman, J.A. Lohman, E.W. Hughes, C.C. Hanstock, R.G. Shulman, 'Surface Coil Localization of ^{31}P NMR Signals from Orthotopic Human Kidney and Liver', *Proc. Natl. Acad. Sci. USA*, **85**, 971–974 (1988).
9. C.C. Hanstock, D.L. Rothman, J.W. Prichard, T. Jue, R.G. Shulman, 'Spatially Localized ^{1}H NMR Spectra of Metabolites in the Human Brain', *Natl. Acad. Sci. USA*, **85**(6), 1821–1825 (1988).
10. C.L. Dumoulin, S.P. Souza, M.F. Walker, W. Wayle, 'Three-dimensional Phase Contrast Angiography', *Magn. Reson. Med.*, **9**, 139–149 (1989).
11. A.M. Howseman, M.K. Stehling, B. Chapman, R. Coxon, R. Turner, R.J. Ordidge, M.G. Cawley, P. Glover, P. Mansfield, R.E. Coupland, 'Improvements in Snapshot Nuclear Magnetic Resonance Imaging', *Br. J. Radiol.*, **61**, 822–828 (1988).
12. C. Kittel, *Introduction To Solid State Physics*, 5th edition, John Wiley & Sons, New York, 1976.

13. S.C. Bushong, *Radiologic Science for Technologists: Physics, Biology and Protection*, 5th edition, Mosby, St. Louis, 466, 1993.

14. P. Hanley, 'Superconducting and Resistive Magnets in NMR Scanning', in *NMR Imaging*, eds. R.L. Witcofski, N. Karstaedt, C.L. Partain, Bowman Grey School of Medicine, Winston-Salem, NC, 41–49, 1982.

15. W.A. Anderson, 'Electrical Current Shims for Correcting Magnetic Fields', *Rev. Sci. Instr.*, **32**, 241 (1961).

16. R.M. Slone, L.L. Buck, J.R. Fitzsimmons, 'Varying Gradient Angles and Offsets to Optimize Imaging Planes in MR', *Radiology*, **158**, 531 (1985).

17. J.J. Ackerman, T.H. Grove, G.G. Wong, D.G. Gadian, G.K. Radda, 'Mapping of Metabolites in Whole Animals by ^{31}P NMR Using Surface Coils', *Nature*, **283**, 167–170 (1980).

18. H. Requardt, J. Offermann, P. Erhard, 'Switched Array Coils', *Magn. Reson. Med.*, **13**, 385–397 (1990).

19. P.B. Roemer, W.A. Edelstein, C.E. Hayes, S.P. Souza, O.M. Mueller, 'The NMR Phased-array', *Magn. Reson. Med.*, **16**, 192–225 (1990).

20. W. Froncisz, J.S. Hyde, 'The Loop-gap Resonator: A New Microwave Lumped Circuit ESR Sample Structure', *J. Magn. Reson.*, **47**, 515–521 (1982).

21. C.E. Hayes, W.A. Edelstein, J.F. Schenck, O.M. Mueller, M. Eash, 'An Efficient, Highly Homogeneous Radiofrequency Coil for Whole-body NMR Imaging at 1.5 T', *J. Magn. Reson.*, **63**, 622–628 (1985).

22. D.W. Alderman, D.M. Grant, 'An Efficient Decoupler Coil which Reduces Heating in Conductive Samples in Superconducting Spectrometers', *J. Magn. Reson.*, **36**, 447 (1979).

23. G.M. Bydder, W.L. Curati, D.G. Gadian, 'Use of Closely Coupled Receiver Coils in Magnetic Resonance Imaging: Practical Aspects', *J. Comput. Assist. Tomogr.*, **9**, 987 (1985).

24. M.D. Schnall, V.H. Subramanian, J.S. Leigh, B. Chance, *J. Magn. Reson.*, **65**, 321 (1985).

25. M.P. Siedband, 'Medical Imaging Systems', in *Medical Instrumentation: Application and Design*, 2nd edition, ed. J.G. Webster, Houghton Mifflin Company, Boston, 630–693, Chapter 12, 1992.

26. J.S. Pellerito, S.M. Mccarthy, M.B. Doyle, M.G. Glickman, A.H. Decherney, 'Diagnosis of Uterine Anomalies: Relative Accuracy of Magnetic Resonance Imaging, Endovaginal Sonography and Hysterosalpingography', *Radiology*, **183**, 795–800 (1992).

27. B.M. Carrington, H. Hricak, R.N. Nuruddin, E. Secaf, R.K. Laros, E.C. Hill, 'Mullerian Duct Anomalies: Magnetic Resonance Imaging Evaluation', *Radiology*, **176**, 715–720 (1990).

28. G.S. Letterie, M. Haggerty, G. Lindee, 'A Comparison of Pelvic Ultrasound and Magnetic Resonance Imaging as Diagnostic Studies for Mullerian Tract Abnormalities', *Int. J. Fertil. Menopausal. Stud.*, **40**, 34–38 (1995).

29. A.S. Mark, H. Hricak, L.W. Heinrichs, M.R. Hendrickson, M.L. Winkler, J.A. Bachica, J.E. Stickler, 'Adenomyosis and Leiomyoma: Differential Diagnosis with Magnetic Resonance Imaging', *Radiology*, **163**, 527–529 (1987).

30. K. Togashi, H. Ozasa, I. Konishi, H. Itoh, K. Nishimura, I. Fujisawa, S. Noma, T. Sagoh, S. Minami, K. Yamashita, Y. Nakano, J. Konishi, T. Mori, 'Enlarged Uterus: Differentiation Between Adenomyosis and Leiomyoma with Magnetic Resonance Imaging', *Radiology*, **171**, 531–534 (1989).

31. C. Reinhold, S. Mccarthy, P.M. Bret, A. Mehio, M. Atri, R. Zakarian, Y. Glaude, Y. Liang, R.J. Seymour, 'Diffuse Adenomyosis: Comparison of Endovaginal US and Magnetic Resonance Imaging with Histopathologic Correlation', *Radiology*, **199**, 151–158 (1996).

32. Y. Yamashita, H. Mizutani, M. Torashima, M. Takahashi, K. Miyazaki, H. Okamura, H. Ushijima, H. Ohtake, T. Tokunaga, 'Assessment of Myometrial Invasion by Endometrial Carcinoma: Transvaginal Sonography vs Contrast-enhanced Magnetic Resonance Imaging', *Am. J. Roentgenol.*, **161**, 595–599 (1993).

33. H. Hricak, L.V. Rubinstein, G.M. Gherman, N. Karstaedt, 'Magnetic Resonance Imaging Evaluation of Endometrial Carcinoma: Results of an NCI Cooperative Study', *Radiology*, **179**, 829–832 (1991).

34. S.H. Kim, B.I. Choi, J.K. Han, H.D. Kim, H.P. Lee, S.B. Kang, J.Y. Lee, M.C. Han, 'Preoperative Staging of Uterine Cervical Carcinoma: Comparison of CT and MRI in 99 Patients', *J. Comput. Tomogr.*, **17**, 633–640 (1993).

35. L.L. Subak, H. Hricak, C.B. Powell, L. Azizi, J.L. Stern, 'Cervical Carcinoma: Computed Tomography and Magnetic Resonance Imaging for Preoperative Staging', *Obstet. Gynecol.*, **86**, 43–50 (1995).

36. H.H. Lien, V. Blomlie, K. Kjorstad, V. Abeler, O. Kaalhus, 'Clinical Stage I Carcinoma of the Cervix: Value of Magnetic Resonance Imaging in Determining Degree of Invasiveness', *Am. J. Roentgenol.*, **156**, 1191–1194 (1991).

37. S. Saini, R.R. Edelman, P. Sharma, W. Li, W. Mayo-Smith, G.J. Slater, P.J. Eisenberg, P.F. Hahn, 'Blood-pool Magnetic Resonance Contrast Material for Detection and Characterization of Focal Hepatic Lesions: Initial Clinical Experience with Ultrasmall Superparamagnetic Iron Oxide (AMI-227)', *Am. J. Roentgenol.*, **164**, 1147–1152 (1995).

38. P. Reimer, E.J. Rummeny, H.E. Daldrup, T. Balzer, B. Tombach, T. Berns, P.E. Peters, 'Clinical Results with Resovist: A Phase 2 Clinical Trial', *Radiology*, **195**, 489–496 (1995).

39. P. Reimer, E.J. Rummeny, K. Shasmi, T. Balzer, H.E. Daldrup, B. Tombach, T. Berns, P.E. Peters, 'Phase II Clinical Evaluation of Gd-EOB-DTPA: Dose, Safety Aspects and Pulse Sequence', *Radiology*, **199**, 177–183 (1996).

40. T. Murakami, R. Baron, M.S. Peterson, J.H. Oliver, P.L. Davis, S.R. Confer, M.P. Federle, 'Hepatocellular Carcinoma: Magnetic Resonance Imaging with Mangafodipir Trisodium (Mn-DPDP)', *Radiology*, **200**, 69–77 (1996).

41. N.M. Rofsky, J.P. Earls, 'Mangafodipir Trisodium Injection (Mn-DPDP). A Contrast Agent for Abdominal Magnetic Resonance Imaging', *Magn. Reson. Imag. Clin. N. Am.*, **4**, 73–85 (1996).

42. T.J. Vogl, S. Kummel, R. Hammerstingl, M. Schellenbeck, G. Schumacher, T. Balzer, W. Schwarz, P.K. Muller, W.O. Bechstein, M.G. Mack, O. Sollner, R. Felix, 'Liver Tumors: Comparison of Magnetic Resonance Imaging with Gd-EOB-DTPA and Gd-DTPA', *Radiology*, **200**, 59–67 (1996).

43. L.E. Crooks, D.A. Ortendahl, L. Kaufman, J. Hoenninger, M. Arakawa, J. Watts, C.R. Cannon, M. Brant-Zawadzki, P.L. Davis, A.R. Margulis, 'Clinical Efficiency of Nuclear Magnetic Resonance Imaging', *Radiology*, **146**, 123–128 (1983).

44. F.S. Buonanno, I.L. Pykett, T.J. Brady, P. Black, P.F. New, E.P. Richardson, W.S. Hinshaw, M. Goldman, G. Pohost, J.P. Kistler, 'Clinical Relevance of Two Different Nuclear Magnetic Resonance (NMR) Approaches to Imaging of a Low-grade Astrocytoma', *J. Comput. Assist. Tomogr.*, **6**, 529–535 (1982).

45. J.P. Mugler, J.R. Brookman, 'Three-dimensional Magnetization Prepared Rapid Gradient-echo Imaging (3D MP RAGE)', *Magn. Reson. Med.*, **15**, 152–157 (1990).

46. A. Haase, J. Frahm, D. Matthaei, W. Hanicke, K.D. Merboldt, 'FLASH Imaging: Rapid NMR Imaging Using Low Flip Angle Pulses', *J. Magn. Reson.*, **67**, 217–225 (1986).

47. S.H. Koenig, M. Spiller, R.D. Brown, G.L. Wolf, 'Relaxation of Water Protons by the Intra- and Extracellular Components of Blood Containing GdDTPA', *Magn. Reson. Med.*, **3**, 791–795 (1986).

48. R.C. Brasch, 'Contrast Enhanced Magnetic Resonance', *Magn. Reson. Med.*, **22**, 177–378 (1991).

49. S.M. Rocklage, D. Worah, S.-H. Kim, 'Metal Ion Release from Paramagnetic Chelates: What is Tolerable?', *Magn. Reson. Med.*, **22**, 216–221 (1991).

50. R.B. Lauffer, 'Paramagnetic Metal Complexes as Water Proton Relaxation Agents for NMR Imaging: Theory and Design', *Chem. Rev.*, **87**, 901–927 (1987).

51. P. Wedeking, S. Eaton, D.G. Covell, S. Nair, M.F. Tweedle, W.C. Eckelman, 'Pharmocokinetic Analysis of Blood Distribution of Intravenously Administered ^{153}Gd-labeled $Gd(DTPA)^{2-}$ and ^{99M}Tc(DTPA) in Rats', *Magn. Reson. Imaging*, **8**, 567–575 (1990).

52. W.W. Orrison, *Introduction to Neuroimaging*, Little, Brown, Boston, 65–234, 1989.

53. D. Norman, E.A. Stevens, S.D. Wing, V. Levin, T.H. Newton, 'Quantitative Aspects of Contrast Enhancement in Cranial Computed Tomography', *Radiology*, **129**, 683–688 (1978).

54. R.E. Latchaw, L.H.A. Gold, E.F. Tourje, 'A Protocol for the Use of Contrast Enhancement in Cranial Computed Tomography', *Radiology*, **126**, 681–687 (1978).

55. V.M. Runge, T.F. Schaible, H.A. Goldstein, M.L. Wood, V.M. Haughton, K.R. Maravilla, M.A. Weinstein, A.C. Price, S.M. Wolpert, D. Kaufman, 'Gd-DTPA: Clinical Efficacy', *RadioGraphics*, **8**(1), 147–158 (1988).

56. V.M. Runge, M.L. Wood, D. Kaufman, A.C. Price, 'Gd-DTPA: Future Applications with Advanced Imaging Techniques', *RadioGraphics*, **8**(1), 161–179 (1988b).

57. J.L. Creasy, R.R. Price, T. Presbrey, D. Goins, C.L. Partain, R.M. Kessler, 'Gadolinium-enhanced Magnetic Resonance Angiography', *Radiology*, **175**, 280–283 (1990).

58. D.W. Chakeres, P. Schmalbrock, M. Brogan, C. Yuan, L. Cohen, 'Normal Venous Anatomy of the Brain: Demonstration with Gadopentetate Dimeglumine in Enhanced 3-D Magnetic Resonance Angiography', *Am. J. Roengenol.*, **156**, 161–172 (1991).

59. G. Marchal, J. Michiels, H. Bosmans, P. Van Hecke, 'Contrast-enhanced MRA of the Brain', *J. Comput. Assist. Tomogr.*, **16**, 25–29 (1992).

60. V.M. Runge, J.E. Kirsh, C. Lee, 'Contrast-enhanced Magnetic Resonance Angiography', *J. Magn. Reson. Imag.*, **3**, 233–239 (1993).

61. W. Lin, E.M. Haacke, A.S. Smith, M.E. Clampitt, 'Gadolinium-enhanced High-resolution Magnetic Resonance Angiography with Adaptive Vessel Tracking: Preliminary Results in the Intracranial Circulation', *J. Magn. Reson. Imag.*, **2**, 277–284 (1992).

62. M.R. Prince, 'Body Magnetic Resonance Angiography with Gadolinium Contrast Agents', *Magn. Reson. Imag. Clin. N. Am.*, **4**, 11–24 (1996).

63. R. Levy, M.R. Prince, 'Arterial-phase Three-dimensional Contrast-enhanced Magnetic Resonance Angiography of the Carotid Arteries', *Am. J. Roentgenol.*, **167**, 211–215 (1996).

64. T.J. Masaryk, M.T. Modic, P.M. Ruggieri, J.S. Ross, G. Laub, G.W. Lenz, J.A. Tkach, E.M. Haacke, W.R. Selman, S.I. Harik, 'Three-dimensional (Volume) Gradient-echo Imaging of the Carotid Bifurcation: Preliminary Clinical Experience', *Radiology*, **171**, 801–806 (1989).

65. P.J. Keller, B.P. Drayer, E.K. Fram, K.D. Williams, C.L. Dumoulin, S.P. Souza, 'Magnetic Resonance Angiography with Two-dimensional Acquisition and Three-dimensional Display', *Radiology*, **173**, 527–532 (1989).

66. G. Marchal, H. Bosmans, L. Van Fraeyenhoven, G. Wilms, P. Van Hecke, A.L. Baert, 'Intracranial Vascular Lesions: Optimization and Clinical Evaluation of Three-dimensional Time-of-flight Magnetic Resonance Angiography', *Radiology*, **175**, 443–448 (1990).

67. P.M. Ruggieri, G.A. Laub, T.M. Masaryk, M.T. Modic, 'Intracranial Circulation: Pulse-sequence Considerations in Three-dimensional (Volume) Magnetic Resonance Angiography', *Radiology*, **171**, 785–791 (1989).

68. G.L. Parker, C. Yuan, D.D. Blatter, 'Magnetic Resonance Angiography by Multiple Thin-slab 3D Acquisition', *Magn. Reson. Med.*, **17**, 434–451 (1991).

69. A.W. Litt, E.M. Eidelman, R.S. Pinto, T.S. Riles, S.J. McLachlan, S. Schwartzenberg, J.C. Weinreb, I.I. Kricheff, 'Diagnosis of Carotid Artery Stenosis: Comparison of 2DFT Time-of-flight Magnetic Resonance Angiography with Contrast Angiography in 50 Patients', *Am. J. Neuroradiol.*, **12**, 149–154 (1991).

70. S.S. Yang, L.G. Bentivoglio, V. Maranhao, H. Goldberg, 'Blood Volume', in *From Cardiac Catheterization Data to Hemodynamic Parameters*, 3rd edition, F.A. Davis Co., Philadelphia, 81, 1988.

71. K.L. Gould, 'Detecting and Assessing Severity of Coronary Artery Disease in Humans', *Cardiovasc. Intervent. Radiol.*, **13**, 5–13 (1990).

72. Anonymous, 'Guidelines for Coronary Angiography: A Report of the American College of Cardiology/American Heart Association Task Force On Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Subcommittee on Coronary Angiography)', *Circulation*, **76**, 963A–977A (1987).

73. G. Karczmar, 'Imaging, Diffusion, Perfusion, and Flow', in *NMR in Physiology and Biomedicine*, ed. R.J. Gillies, Academic Press, New York, 25–42, 1994.

74. H.L. Lutsep, G.W. Albers, A. Decrespingny, G.N. Kamat, M.P. Marks, M.E. Moseley, 'Clinical Utility of Diffusion-weighted Magnetic Resonance Imaging in the Assessment of Ischemic Stroke', *Ann. Neurol.*, **41**, 574–580 (1997).

75. M.P. Marks, A. Decrespigny, D. Lentz, D.R. Enzmann, G.W. Albers, M.E. Moseley, 'Acute and Chronic Stroke: Navigated Spin–echo Diffusion-weighted Magnetic Resonance Imaging', *Radiology*, **199**, 403–408 (1996).

76. K. Minematsu, L. Li, C.H. Sotak, M.A. Davis, M.S. Fiandaca, 'Diffusion-weighted Magnetic Resonance Imaging: Rapid and Quantitative Detection of Focal Brain Ischemia', *Neurology*, **42**, 235–240 (1992).

77. S. Warach, D. Chien, D. Li, M. Ronthal, R.R. Edelman, 'Fast Magnetic Resonance Diffusion-weighted Imaging of Acute Human Stroke', *Neurology*, **42**, 1717–1723 (1992).

78. S. Warach, J. Sashe, R. Edelman, 'Clinical Outcome in Ischemic Stroke Predicted by Early Diffusion-weighted and Perfusion Magnetic Resonance Imaging: A Preliminary Analysis', *J. Cereb. Blood Flow Metab.*, **16**, 53–59 (1996).

79. M.E. Moseley, Y. Cohen, J. Mintorovitch, L. Chileuitt, H. Shimizu, J. Kucharczyk, M.F. Wendland, P.R. Weinstein, 'Early Detection of Regional Ischemia In Cats: Comparison of Diffusion- and T_2-weighted MRI and Spectroscopy', *Magn. Reson. Med.*, **14**, 330–346 (1990).

80. H. Benveniste, L.W. Hedlund, G.A. Johnson, 'Mechanism of Detection of Acute Cerebral Ischemia in Rats by Diffusion-weighted Magnetic Resonance Microscopy', *Stroke*, **23**, 746–754 (1992).

81. R.J. Sevick, F. Kanda, J. Minotorovitch, A.I. Arieff, T. Kucharczyk, J.S. Tsuruda, D. Norman, M.E. Moseley, 'Cytotoxic Brain Edema: Assessment with Diffusion-weighted Magnetic Resonance Imaging', *Radiology*, **185**, 687–690 (1992).

82. J. Mintorovich, G.Y. Yang, H. Shimizu, J. Kucharczyk, P.H. Chan, P.R. Wienstien, 'Diffusion-weighted Magnetic Resonance Imaging of Acute Focal Cerebral Ischemia: Comparison of Signal Intensity with Changes in Brain Water and Na^+, K^+ ATPase Activity', *J. Cereb. Blood Flow Metab.*, **14**, 332–336 (1994).

83. J. Kucharczyk, Z.S. Vexler, T.P. Roberts, H.S. Asari, J. Mintorovich, N. Derugin, A.D. Watson, M.E. Mosely, 'Echo-planar Perfusion-sensitive Magnetic Resonance Imaging of Acute Cerebral Ischemia', *Radiology*, **188**, 711–717 (1993).

84. K. Matsumoto, E.H. Lo, A.R. Pierce, L. Garrido, N.W. Kowell, 'Role of Vasogenic Edema and Tissue Cavitation in Ischemic Evolution on Diffusion-weighted Imaging: Comparison with Multiparameter Magnetic Resonance and Immunohistochemistry', *Am. J. Neuroradiol.*, **16**, 1107–1115 (1995).

85. A. Sorensen, F. Buonnano, R. Gonzalez, L. Schwamm, M. Lev, F. Huang-Hellinger, T. Reese, R. Weisskoff, T. Davis, N. Suwanwela, U. Can, J. Moreira, W. Copen, R. Look, S. Finkelstein, B. Rosen, W. Koroshetz, 'Hyperacute Stroke: Evaluation with Combined Multisection Diffusion-weighted and Hemodynamically Weighted Echo-planar Magnetic Resonance Imaging', *Radiology*, **199**, 391–401 (1996).

86. J. Kucharczyk, J. Mintorovich, H. Asgari, M. Moseley, 'Diffusion/Perfusion Magnetic Resonance Imaging of Acute Cerebral Ischemia', *Magn. Reson. Med.*, **19**, 311–315 (1991).

87. M. Moseley, J. Kucharczyk, J. Mintorovich, Y. Cohen, J. Kuranewicz, N. Derugin, H. Asgari, D. Norman, 'Diffusion-weighted Magnetic Resonance Imaging of Acute

Stroke: Correlation with T_2-weighted and Magnetic Susceptibility-enhanced Magnetic Resonance Imaging in Cats', *Am. J. Neuroradiol.*, **11**, 423–429 (1990).

88. K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, R. Turner, H.M. Cheng, T.J. Brady, B.R. Rosen, 'Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation', *Proc. Natl. Acad. Sci. USA*, **89**, 5675–5679 (1992).
89. R. Turner, P. Jezzard, H. Wen, K.K. Kwong, D. Le Bihan, T. Zeffiro, R.S. Balaban, 'Functional Mapping of the Human Visual Cortex at 4 and 1.5 Tesla Using Deoxygenation Contrast EPI', *Magn. Reson. Med.*, **29**, 277–279 (1993).
90. P.A. Bandettini, E.C. Wong, R.S. Hinks, R.S. Tikofsky, J.S. Hyde, 'Time Course EPI of Human Brain Function During Task Activation', *Magn. Reson. Med.*, **25**, 390–397 (1992).
91. J. Frahm, H. Bruhn, K. Merboldt, W. Hanicke, 'Dynamic Magnetic Resonance Imaging of Human Brain Oxygenation During Rest and Photic Stimulation', *J. Magn. Reson. Imag.*, **2**, 501–505 (1992).
92. M. Cohen, R. Weisskoff, 'Ultra Fast Imaging', *Magn. Reson. Med.*, **9**, 1–37 (1991).
93. W. Raushning, 'Brain Tumors and Tumorlike Masses: Classification and Differential Diagnosis', in *Diagnostic Neuroradiology*, ed. A. Osborn, Mosby, St Louis, 401–522, 1994.
94. J.S. Leigh, 'FDA Safety Guidelines for MRI Devices', *Resonance Newsletter SMRM*, **20**, 9 (1990).
95. P.A. Bottomly, T.H. Foster, R.D. Darrow, 'Depth-resolved Surface-coil Spectroscopy (DRESS) for *In Vivo* ^{1}H, ^{31}P and ^{13}C NMR', *J. Magn. Reson.*, **59**, 338–342 (1984).
96. P.A. Bottomly, L. Scott Smith, W.M. Leue, C. Charles, 'Slice-interleaved Depth-resolved Surface-coil Spectroscopy (SLIT/DRESS) for Rapid ^{31}P NMR *In Vivo*', *J. Magn. Reson.*, **64**, 347–351 (1985).
97. R.J. Ordidge, A. Connely, J.A.B. Lohman, 'Image-selected *In Vivo* Spectroscopy (ISIS). A New Technique for Spatially Selective NMR Spectroscopy', *J. Magn. Reson.*, **66**, 283–294 (1986).
98. P.A. Bottomly, 'Spatial Localization in NMR Spectroscopy *In Vivo*', *Ann. NY Acad. Sci.*, **508**, 333–348 (1987).
99. R. Kimmich, D. Hoepfel, 'Volume Selective Multipulse Spin Echo Spectroscopy', *J. Magn. Reson.*, **72**, 379–384 (1987).
100. E.L. Hahn, 'Spin Echoes', *Phys. Rev.*, **80**, 580–594 (1950).
101. A.A. Maudsley, S.K. Hilal, W.H. Perman, H.E. Simon, 'Spatially Resolved High-resolution Spectroscopy by "Four-dimensional" NMR', *J. Magn. Reson.*, **51**, 147–152 (1983).
102. T.R. Brown, B. Kincaid, K. Ugurbil, 'NMR Chemical Shift Imaging in Three Dimensions', *Proc. Natl. Acad. Sci. USA*, **79**, 3523–3526 (1982).
103. V. Seshan, N. Bansal, '*In Vivo* ^{31}P and ^{23}Na NMR Spectroscopy and Imaging', in *NMR Spectroscopy Techniques*, ed. M.D. Bruch, Marcel Dekker, New York, 557–607, 1996.
104. Y. Masuda, Y. Teteno, H. Ikehira, T. Hashimoto, F. Shishido, M. Sekiya, Y. Imazeki, H. Imai, S. Watanabe, Y. Ingaki, 'High-energy Phosphate Metabolism of the Myocardium in Normal Subjects and Patients with Various Cardiomyopathies – The Study Using ECG Gated Magnetic Resonance Spectroscopy with a Localization Technique', *Jpn. Circ. J.*, **56**, 620–626 (1992).
105. R.G. Weiss, P.A. Bottomley, C.J. Hardy, G. Gerstenblith, 'Regional Myocardial Metabolism of High-energy Phosphates During Isometric Exercise in Patients with Coronary Artery Disease', *N. Engl. J. Med.*, **323**, 1593–1600 (1990).
106. T. Yabe, K. Mitsunami, M. Okada, S. Morikawa, T. Inubushi, M. Kinoshita, 'Detection of Myocardial Ischemia by ^{31}P-Magnetic Resonance Spectroscopy During Hand-grip Exercise', *Circulation*, **89**, 1709–1716 (1994).
107. P.A. Bottomley, R.G. Weiss, C.J. Hardy, W.A. Baumgartner, 'Myocardial High-energy Phosphate Metabolism and Allograph Rejection in Patients with Heart Transplants', *Radiology*, **181**, 67–75 (1991).
108. S. Neubauer, T. Krahe, R. Schindler, M. Horn, H. Hillenbrand, C. Entzeroth, H. Mader, E.P. Kromer, G.A.J. Riegger, K. Lackner, G. Ertl, '^{31}P Magnetic Resonance Spectroscopy in Dilated Cardiomyopathy and Coronary Artery Disease', *Circulation*, **86**, 1810–1818 (1992).
109. Z. Argov, W.J. Bank, J. Maris, P. Peterson, B. Chance, 'Bioenergetic Heterogeneity of Human Mitochondrial Myopathies: Phosphorus Magnetic Resonance Spectroscopy Study', *Neurology*, **37**, 257–262 (1987).
110. Z. Argov, P.F. Renshaw, B. Boden, A. Winokur, J. Bank, 'Effects of Thyroid Hormones on Skeletal Muscle Bioenergetics: *In Vivo* Phosphorus-31 Magnetic Resonance Spectroscopy Study of Human and Rats', *J. Clin. Invest.*, **81**, 1695–1701 (1988).
111. W. Bank, Z. Argov, J.S. Leigh, B. Chance, 'The Value of ^{31}P NMR in the Diagnosis and Monitoring of Human Myopathies', *Ann. NY Acad. Sci.*, **508**, 448–450 (1987).
112. B. Chance, J.S. Leigh, D. Smith, S. Nioka, B.J. Clarck, 'Phosphorus Magnetic Resonance Spectroscopy of the Role of Mitochondria in the Disease Process', *Ann. NY Acad. Sci.*, **488**, 140–153 (1986).
113. D.W. Zochdone, R.T. Thompson, A.A. Drieger, M.J. Strong, D. Gravelle, C.F. Bolton, 'Metabolic Changes

in Human Muscle Denervation: Topical ^{31}P NMR Spectroscopy', *Mag. Reson. Med.*, **7**, 373–383 (1988).

114. B.D. Ross, G.K. Radda, D.G. Gadian, G. Rocker, M. Esini, J. Falconer-Smith, 'Examination of a Case of Suspected McArdles' Syndrome by ^{31}P Nuclear Magnetic Resonance', *New Engl. J. Med.*, **304**, 1338–1342 (1981).

115. B. Chance, S. Eleff, W. Bank, J.S. Leigh, R. Warnell, '^{31}P NMR Studies of Mitochondrial Function in Phosphofructokinase Deficient Human Skeletal Muscle', *Proc. Natl. Acad. Sci. USA*, **79**, 7714–7718 (1982).

116. D.P. Younkin, M. Delivoria-Papadopoulos, J.C. Leonard, V.H. Subramanian, S. Eleff, J.S. Leigh, B. Chance, 'Unique Aspects of Human Newborn Cerebral Metabolism Evaluated with Phosphorus Nuclear Magnetic Resonance Spectroscopy', *Ann. Neurol.*, **16**, 581–586 (1984).

117. P.L. Hope, E.O. Reynolds, 'Investigation of Cerebral Energy Metabolism in Newborn Infants by Phosphorus Nuclear Magnetic Resonance Spectroscopy', *Clin. Perinatol.*, **12**, 261–275 (1985).

118. P.A. Hamilton, P.L. Hope, E.B. Cady, D.T. Delpy, J.S. Wyatt, E.O. Reynolds, 'Impaired Energy Metabolism in Brains of Newborn Infants with Increased Cerebral Echodensities', *Lancet*, **1**, 1242–1246 (1986).

119. D.T. Delpy, M.C. Cope, E.B. Cady, J.S. Wyatt, P.A. Hamilton, P.L. Hope, S. Wray, E.O. Reynolds, 'Cerebral Monitoring in Newborn Infants by Magnetic Resonance and Near Infrared Spectroscopy', *Scand. J. Clin. Lab. Invest. Suppl.*, **188**, 9–17 (1987).

120. D. Azzopardi, J.S. Wyatt, E.B. Cady, D.T. Delpy, J. Baudin, A.L. Stewart, P.L. Hope, P.A. Himilton, E.O. Reynolds, 'Prognosis of Newborn Infants with Hypoxic-ischemic Brain Injury Assessed by Phosphorus Magnetic Resonance Spectroscopy', *Petiatr. Res.*, **25**, 445–451 (1989).

121. T. Kato, A. Tokumaru, T. O'uchi, I. Mikami, M. Umeda, K. Nose, T. Eguchi, M. Hasegawa, K. Okuyama, 'Assessment of Brain Death in Children by Means of P-31 Magnetic Resonance Spectroscopy: Preliminary Note. Work In Progress', *Radiology*, **179**, 95–99 (1991).

122. S.C. Roth, A.D. Edwards, E.B. Cady, D.T. Delpy, J.S. Wyatt, D. Azzopardi, J. Baudin, J. Townsend, A.L. Stewart, E.O. Reynolds, 'Relation Between Cerebral Oxidative Metabolism Following Birth Asphyxia and Neurodevelopmental Outcome and Brain Growth at One Year', *Dev. Med. Child. Neurol.*, **34**, 285–295 (1992).

123. K.M. Welch, S.R. Levine, J.A. Helpern, 'Pathophysiological Correlates of Cerebral Ischemia: The Significance of Cellular Acid Base Shifts', *Funct. Neurol.*, **5**, 21–31 (1990).

124. S.R. Levine, K.M. Welch, J.A. Helpern, M. Chop, R. Bruce, J. Selwa, M.B. Smith, 'Prolonged Deterioration of Ischemic Brain Energy Metabolism and Acidosis Associated with Hyperglycemia: Human Cerebral Infarction Studied by Serial ^{31}P NMR Spectroscopy', *Ann. Neurol.*, **23**, 416–418 (1988).

125. S.R. Levine, J.A. Helpern, K.M. Welch, A.M. Vande Linde, K.L. Sawaya, E.E. Brown, N.M. Ramadan, R.K. Deveshwar, R.J. Ordidge, 'Human Focal Cerebral Ischemia: Evaluation of Brain pH and Energy Metabolism with P-31 NMR Spectroscopy', *Radiology*, **185**, 537–544 (1992).

126. J.W. Prichard, 'The Ischemic Penumbra in Stroke: Prospects for Analysis by Nuclear Magnetic Resonance Spectroscopy. [Review]', *Research Publications–Association for Research in Nervous and Mental Disease*, **71**, 153–174 (1993).

127. J.R. Griffiths, 'Are Cancer Cells Acidic?', *Br. J. Cancer*, **64**, 425–427 (1991).

128. W. Negendank, 'Studies of Human Tumours by MRS: A Review', *NMR Biomed.*, **5**, 303–324 (1992).

129. J. Schaefer, 'Selective Saturation of Carbon-13 Lines in Carbon-13 Fourier Transform NMR Experiments', *J. Magn. Res.*, **6**, 670–671 (1972).

130. D.I. Hoult, 'Solvent Peak Saturation with Single Phase and Quadrature Fourier Transformation', *J. Magn. Res.*, **21**, 337–347 (1976).

131. V. Sklenar, Z. Starcuk, '1-2-1 Pulse Train: A New Effective Method of Selective Excitation for Proton NMR in Water', *J. Magn. Res.*, **50**, 495–501 (1982).

132. P.J. Hore, 'Solvent Suppression in Fourier Transform Nuclear Magnetic Resonance', *J. Magn. Reson.*, **54**, 539–542 (1983).

133. A. Haase, J. Frahm, W. Hanicke, D. Matthaei, '^{1}H NMR Chemical Shift Selective (CHESS) Imaging', *Phys. Med. Biol.*, **30**, 341–344 (1985).

134. D.H. Miller, S.J. Austin, A. Connelly, B.D. Youl, D.G. Gadian, W.I. McDonald, 'Proton Magnetic Resonance Spectroscopy of an Acute and Chronic Lesion in Multiple Sclerosis', *Lancet*, **1337**, 58–59 (1991).

135. J.R. Alger, J.A. Frank, A. Bizzi, M.J. Fulham, B.X. DeSouza, M.O. Duhaney, S.W. Inscoe, J.L. Black, P.C.M. van Zijl, C.T.W. Moonen, G. Di Chiro, 'Metabolism of Human Gliomas: Assessment with ^{1}H Magnetic Resonance Spectroscopy and ^{18}F-Deoxyglucose PET', *Radiology*, **177**, 633–641 (1990).

136. S.S. Gill, D.G.T. Thomas, N. Van Bruggen, D.G. Gadian, C.J. Peden, J.D. Bell, J. Cox, D.K. Menon, 'Proton Magnetic Resonance Spectroscopy of Intracranial Tumors: *In Vivo* and *In Vitro* Studies', *J. Comput. Assist. Tomogr.*, **14**, 497–504 (1990).

137. C.M. Segebarth, D.F. Balarieaux, P.R. Luyten, J.A. den Hollander, 'Detection of Metabolic Heterogeneity of Human Intracranial Tumors *In Vivo* by ^{1}H NMR Spectroscopic Imaging', *Magn. Reson. Med.*, **13**, 62–76 (1990).

138. K. Herholz, W. Heindel, P.R. Luyten, J.A. den Hollander, U. Pietrzyk, J. Vogas, Kugel, G. Friedmann, W.D. Heiss, '*In Vivo* Imaging of Glucose Consumption and Lactate Concentration in Human Gliomas', *Ann. Neurol.*, **31**, 319–327 (1992).

139. M.J. Fernstermacher, P.A. Narayana, 'Serial Proton Magnetic Resonance Spectroscopy of Ischemic Brain Injury in Humans', *Invest. Radiol.*, **25**, 1034–1039 (1990).

140. C.C. Ford, R.H. Griffey, N.A. Matwiyoff, G.A. Rosenberg, 'Multivoxel ^{1}H-MRS of Stroke', *Neurology*, **42**, 1408–1412 (1992).

141. S.R. Felber, F.T. Aichner, R. Sauter, F. Gerstenbraind, 'Combined Magnetic Resonance Imaging and Proton Magnetic Resonance Spectroscopy of Patients with Acute Stroke', *Stroke*, **23**, 1106–1110 (1992).

142. J.H. Duijin, G.B. Matson, A.A. Maudsely, J.W. Hugg, M.W. Weiner, 'Proton Magnetic Resonance Spectroscopic Imaging of Human Brain Infarction', *Radiology*, **183**, 711–718 (1992).

143. J.W. Hugg, J.H. Duijn, G.B. Matson, A.A. Maudsley, J.S. Tsuruda, D.F. Gelinas, M.W. Weiner, 'Elevated Lactate and Alkalosis in Chronic Human Brain Infarction Observed by ^{1}H and ^{31}P Magnetic Resonance Spectroscopic Imaging', *J. Cereb. Blood Flow Metab.*, **12**, 734–744 (1992).

144. G.D. Graham, A.M. Blamire, A.M. Howseman, D.L. Rothman, P.B. Fayad, L.M. Brass, O.A.C. Petroff, R.G. Shulman, J.W. Prichard, 'Proton Magnetic Resonance Spectroscopy of Cerebral Lactate and Other Metabolites in Stroke Patients', *Stroke*, **23**, 333–340 (1992).

145. P. Gideon, O. Henrikson, B. Sperlin, P. Christiansen, T.S. Olsen, H.S. Jorgensen, P. Arlien-Soborg, 'Early Time Course of *N*-Acetyl Aspartate, Creatine and Phosphocreatine and Compounds Containing Choline in the Brain After Acute Stroke: A Proton Magnetic Resonance Spectroscopy Study', *Stroke*, **23**, 1566–1572 (1992).

146. M.S. Van der Knapp, J. Van der Grand, P.R. Luyten, J.A. van Hollander, J.J.P. Nauton, J. Valk, '^{1}H and ^{31}P Magnetic Resonance Spectroscopy of the Brain in Degenerative Cerebral Disorders', *Ann. Neural.*, **31**, 202–211 (1992).

147. C.A. Davie, C.P. Hawkins, G.J. Barker, A. Brennan, P.S. Tofts, D.H. Miller, W.I. McDonald, 'Detection of Myelin Breakdown Products by Proton Magnetic Resonance Spectroscopy [Letter]', *Lancet*, **341**, 630–631 (1993).

148. D.L. Arnold, P.M. Matthews, G.S. Francis, J.R. Antel, 'Proton Magnetic Resonance Spectroscopy of Human Brain *In Vivo* in the Evaluation of Multiple Sclerosis: Assessment of Load of Disease', *Magn. Reson. Med.*, **14**, 154–159 (1990).

149. D.L. Arnold, P.M. Matthews, G.S. Francis, M.D. O'Connor, J.R. Antel, 'Proton Magnetic Resonance Spectroscopic Imaging for Metabolic Characterization of Demyelinating Plaques', *Ann. Neurol.*, **31**, 235–241 (1992).

150. H.B.W. Larsson, P. Christiansen, M. Jensen, J. Frederiksen, A. Heltberg, J. Olesen, O. Henriksen, 'Localized *In Vivo* Proton Spectroscopy in the Brain of Patients with Multiple Sclerosis', *Magn. Reson. Med.*, **22**, 23–31 (1991).

151. P.M. Matthews, F. Andermann, D.L. Arnold, 'A Proton Magnetic Resonance Spectroscopy Study of Focal Epilepsy in Humans', *Neurology*, **40**, 985–989 (1990).

152. J.S. Wolinsky, P.A. Narayana, M.J. Fenstermacher, 'Proton Magnetic Resonance Spectroscopy in Multiple Sclerosis', *Neurology*, **40**, 1764–1769 (1990).

153. L. Chang, B. Miller, D. McBride, M. Cornford, G. Oropilla, S. Buchthal, F. Chiang, H. Aronow, C.K. Beck, T. Ernst, 'Brain Lesions in Patients with AIDS: H-1 Magnetic Resonance Spectroscopy', *Radiology*, **197**, 525–531 (1995).

154. L. Chang, T. Ernst, 'Proton Magnetic Resonance Spectroscopy and Diffusion-weighted MRI in Focal AIDS Brain Lesions', in *Neuroimaging in AIDS II, a Special Issue of Neuroimaging Clinics of North America*, ed. J. Post, W.B. Saunders Company, 409–425, Vol. 7, 1997.

155. R.J.S. Chinn, I.D. Wilkinson, M.A. Hall-Craggs, M.N.J. Paley, R.F. Miller, B.E. Kendall, S.P. Newman, M.J.G. Harrison, 'Toxoplasmosis and Primary Central Nervous System Lymphoma in HIV Infection: Diagnosis with Magnetic Resonance Spectroscopy', *Radiology*, **197**, 649–654 (1995).

156. P.M. Matthews, G. Francis, J. Antel, D.L. Arnold, 'Proton Magnetic Resonance Spectroscopy for Metabolic Characterization of Plaques in Multiple Sclerosis', *Neurology*, **41**, 1251–1256 (1991).

157. J.W. Hugg, K.D. Laxer, G.B. Matson, A.A. Maudsley, M.W. Weiner, 'Neuron Loss Localizes Human Temporal Lobe Epilepsy by *In Vivo* Proton Magnetic Resonance Spectroscopic Imaging', *Ann. Neurol.*, **34**, 788–794 (1993).

158. S.N. Breiter, S. Arroyo, V.P. Matthews, R.P. Lesser, R.N. Bryan, P.B. Barker, 'Proton Magnetic Resonance Spectroscopy in Patients with Seizure Disorders', *Am. J. Neuroradiol.*, **15**, 373–384 (1994).

159. F. Cendes, F. Andermann, M.C. Pruel, D.L. Arnold, 'Lateralization of Temporal Lobe Epilepsy Based on Regional Metabolic Abnormalities in Proton Magnetic Resonance Spectroscopic Images', *Ann. Neurol.*, **35**, 211–216 (1994).

160. D.G. Gadian, A. Connelly, J.S. Duncan, J.H. Cross, F.J. Kirkham, C.L. Johnson, F. Vargha-Khadem, B.G. Nevile, G.D. Jackson, '^{1}H Magnetic Resonance

Spectroscopy in the Investigation of Intractable Epilepsy', *Acta Neurol. Scand. Suppl.*, **152**, 116–121 (1994).

161. H. Hetherington, R. Kuzniecky, J. Pan, G. Mason, R. Morawetz, C. Harris, E. Faught, T. Vaughan, G. Pohost, 'Proton Nuclear Magnetic Resonance Spectroscopic Imaging of Human Temporal Lobe Epilepsy at 4.1 T', *Ann. Neurol.*, **38**, 396–404 (1995).
162. T.C. Ng, Y.G. Comair, M. Xue, N. So, A. Majors, H. Kolem, H. Luders, M. Modic, 'Temporal Lobe Epilepsy: Presurgical Localization with Proton Chemical Shift Imaging', *Radiology*, **193**, 465–472 (1994).
163. P. Vainio, J.P. Usenius, M. Vapalahti, K. Partanen, R. Kalviainen, J. Rinne, R.A. Kauppinen, 'Reduced *N*-Acetylaspartate Concentration in Temporal Lobe Epilepsy by Quantitative ^{1}H MRS *In Vivo*', *Neuro Report*, **5**, 1733–1736 (1994).
164. H.A. Crockard, D.G. Gadian, R.S.J. Fackowiak, E. Proctor, K. Allen, S.R. Williams, R.W. Ross-Russel, 'Acute Cerebral Ischaemia: Concurrent Changes in Cerebral Blood Flow, Energy Metabolites, pH, and Lactate Measured with Hydrogen Clearance and ^{31}P and ^{1}H Nuclear Magnetic Resonance Spectroscopy II. Changes During Ischaemia', *J. Cereb. Blood Flow Metab.*, **7**, 394–402 (1987).
165. H. Bruhn, J. Frahm, M.L. Gyngell, K.D. Merboldt, W. Hanicke, R. Sauer, C. Hamburger, 'Non-invasive Differentiation of Tumors with use of Localized H-1 Magnetic Resonance Spectroscopy In Vivo: Initial Experience in Patients with Cerebral Tumors', *Radiology*, **172**, 541–548 (1989).
166. H. Bruhn, J. Frahm, M.L. Gyngell, K.D. Merboldt, W. Hanicke, R. Sauer, 'Cerebral Metabolism in Man After Acute Stroke: New Observations Using Localized Proton NMR Spectroscopy', *Magn. Reson. Med.*, **9**, 126–131 (1989).
167. J.F. Monroe, J.C. Shipp, 'Glucose Metabolism in Leukocytes from Patients with Diabetes Mellitus, with and without Hypercholesterolemia', *Diabetes*, **14**, 584–590 (1965).
168. P. Newsholme, S. Gordon, E.A. Newsholme, 'Rates of Utilization and Fates of Glucose, Glutamine, Pyruvate, Fatty Acids and Ketone Bodies by Mouse Macrophages', *Biochem. J.*, **24**, 631–636 (1987).
169. P. Newsholme, E.A. Newsholme, 'Rates of Utilization of Glucose Glutamine and Oleate and Formation of End Products by Mouse Peritoneal Macrophages in Culture', *Biochem. J.*, **26**, 211–218 (1989).
170. R.L. Swank, R.F. Hain, 'The Effect of Different Size Emboli on the Vascular System and Parenchyma of the Brain', *J. Neuropathol. Exp. Neurol.*, **11**, L289–L299 (1952).
171. J.H. Garcia, Y. Kamijyo, 'Cerebral Infarction: Evolution of Histopathological Changes After Occlusion of a Middle Cerebral Artery in Primates', *J. Neuropathol. Exp. Neurol.*, **33**, 408–421 (1974).
172. M. Nedergaard, 'Neuronal Injury I The Infarct Border: A Neuropathological Study in the Rat', *Acta Neuropathol. (Berlin)*, **73**, 267–274 (1987).
173. M.J. Fulham, A. Bizzi, M.J. Dietz, H.H. Shih, R. Raman, G.S. Sobering, J.A. Frank, A.J. Dwyer, J.R. Alger, G. DiChiro, 'Mapping of Brain Tumor Metabolites with Proton Magnetic Resonance Spectroscopic Imaging: Clinical Relevance', *Radiology*, **185**, 675–686 (1992).
174. L. Chang, '*In Vivo* Magnetic Resonance Spectroscopy in HIV and HIV-Related Brain Diseases', *Rev. Neurosci.*, **6**, 365–378 (1995).
175. M.C. Preul, Z. Caramanos, D.L. Collins, J.G. Villemure, R. Leblanc, A. Oliver, R. Pokrupa, D. Arnold, 'Accurate, Non-invasive Diagnosis of Human Brain Tumors by Using Proton Magnetic Resonance Spectroscopy', *Nature Med.*, **2**, 323–325 (1996).
176. J.L. Gallis, P. Canioni, 'MRS of the Liver', in *Magnetic Resonance Spectroscopy in Biology and Medicine: Functional and Pathological Tissue Characterization*, eds. J.D. DeCertaines, W.M.M.J. Bovee, F. Podo, Pergamon Press, New York, 345–368, 1992.
177. A.J. Shaka, J. Keeler, 'Broadband Spin Decoupling in Isotropic Liquids', in *Progress in NMR Spectroscopy*, eds. J.W. Emsley, J. Feeney, J. Sutcliffe, Pergamon Press, Oxford, 47, Vol. 19, 1987.
178. N. Beckmann, J. Seelig, H. Wick, 'Analysis of Glycogen Storage Disease by *In Vivo* ^{13}C NMR: Comparison of Normal Volunteers with a Patient', *Magn. Reson. Med.*, **16**, 150–160 (1990).
179. T. Jue, J.A. Lohman, R.J. Ordidge, R.G. Shulman, 'Natural Abundance ^{13}C NMR Spectrum of Glycogen in Humans', *Magn. Reson. Med.*, **3**, 337–379 (1987).
180. M.J. Avison, D.L. Rothman, E. Nadel, R.G. Shulman, 'Detection of Human-muscle Glycogen by Natural Abundance ^{13}C NMR', *Proc. Natl. Acad. Sci. USA*, **85**, 1634–1636 (1988).
181. A. Heerschap, P.R. Luyten, J.I. VanDerHeyden, L.J.M.P. Oosterwaal, J.A. DenHollander, 'Broadband Proton Decoupled Natural Abundance ^{13}C NMR Spectroscopy of Humans at 1.5 T', *NMR Biomed.*, **2**, 124–132 (1989).
182. M. Saner, G. McKinnon, P. Boesiger, 'Glycogen Detection by *In Vivo* ^{13}C NMR: A Comparison of Proton Decoupling and Polarization Transfer', *Magn. Reson. Med.*, **28**, 65–73 (1992).
183. T. Jue, D.L. Rothman, B.A. Tavitian, R.G. Shulman, 'Natural-abundance ^{13}C NMR Study of Glycogen Repletion in Human Liver and Muscle', *Proc. Natl. Acad. Sci. USA*, **86**, 1439 (1989).

184. E. Hultman, 'Muscle Glycogen in Man Determined in Needle Biopsy Specimens: Method and Normal Values', *Scand. J. Clin. Lab. Invest.*, **19**, 209–217 (1967).

185. L. Nilsson-Hison, E. Hultman, 'Liver and Muscle Glycogen in Man After Glucose and Fructose Infusion', *Scan. J. Clin. Lab. Invest.*, **33**, 5 (1974).

186. D.J. Herse, E.B. Chain, 'The Role of Glucose in the Survival and Recovery of the Anoxic Isolated Perfused Rat Heart', *Biochem. J.*, **128**, 1125–1133 (1972).

187. M.J. Rovetto, J.T. Whitmer, J.R. Neely, 'Comparison of the Effects of Anoxia and Whole Heart Ischemia on Carbohydrate Utilization in Isolated Working Rat Hearts', *Circ. Res.*, **32**, 699–711 (1973).

188. M.R. Laughlin, W.A. Petit, J.M. Dizon, R.G. Shulman, E.J. Barrett, '*In Vivo* NMR Measurement of Myocardial Glycogen Metabolism', *J. Biol. Chem.*, **263**, 2285–2291 (1988).

189. M.R. Laughlin, W.A. Petit, R.G. Shulman, E.J. Barrett, 'Measurement of Myocardial Glycogen Synthesis in Diabetic and Fasted Rats', *Am. J. Physiol.*, **258**, E184–E190 (1990).

190. T.E. Walker, C.H. Han, V.H. Kollman, R.E. London, N.A. Matwiyoff, '^{13}C Nuclear Magnetic Resonance Studies of the Biosynthesis by Microbacterium Ammoniaphilum Of L-Glutamate Selectively Enriched with Carbon-13', *J. Biol. Chem.*, **257**, 1189–1195 (1982).

191. S.M. Cohen, 'Simultaneous ^{13}C and ^{31}P NMR Studies of Perfused Rat Liver', *J. Biol. Chem.*, **258**, 14 294–14 308 (1983).

192. C.R. Malloy, A.D. Sherry, F.M.H. Jeffery, '^{13}C Nuclear Magnetic Resonance Methods for the Analysis of Citric Acid Cycle Metabolism in the Heart', in *Cardiovascular Applications of Magnetic Resonance*, ed. G.M. Pohost, Futura Publishing Co., Mount Kisco, NY, 261–270, 1993.

193. G. Clegg, D.B. Morgan, C. Davidson, 'The Heterogeneity of Essential Hypertension. Relation Between Lithium Efflux and Sodium Content of Erythrocytes and a Family History of Hypertension', *Lancet*, **2**, 891–894 (1982).

194. E. Dorus, G.N. Pandey, R. Shaughnessey, M. Galviria, E. Val, S. Ericksen, J.M. Davis, 'Lithium Transport Across Red Cell Membrane: A Cell Membrane Abnormality in Manic-depressive Illness', *Science*, **205**, 932–934 (1979).

195. E.J. Sorscher, J.L. Breslow, 'Cystic Fibrosis: A Disorder of Calcium-stimulated Secretion and Transepithelial Sodium Transport?', *Lancet*, **1**, 368–370 (1982).

196. L.M. Buja, H.K. Hagler, D. Parsons, K. Chien, R.L. Reynolds, J.T. Willerson, 'Alterations of Ultrastructure and Elemental Composition in Cultured Neonatal Rat Cardiac Myocytes after Metabolic Inhibition with Iodoacetic Acid', *J. Lab. Invest.*, **53**, 397–412 (1985).

197. N.M. Garcia, J.W. Horton, 'L-Arginine Improves Resting Cardiac Transmembrane Potential After Burn Injury', *Shock*, **1**, 354–358 (1994).

198. M.M. Sayeed, 'Alterations in Hepatic Na^+–K^+ Transport During Endotoxemia in Rats', *Am. J. Physiol.*, **247**, R465–R474 (1984).

199. M.M. Civan, M. Shporer, 'NMR of Sodium-23 and Potassium-39 in Biological Systems', in *Biological Magnetic Resonance*, eds. L.J. Berliner, J. Reuben, Plenum Press, New York, 1, Vol. 1, 1978.

200. P.M. Winter, V. Seshan, J.D. Makos, A.D. Sherry, C.R. Malloy, N. Bansal, 'Quantitation of Intracellular [Na^+] *In Vivo* by Using $TmDOTP^{5-}$ as an NMR Shift Reagent and Extracellular Marker', *J. Appl. Physiol.*, **85**(5), 1806–1812 (1998).

201. N. Bansal, M.J. Germann, I. Lazar, C.R. Malloy, A.D. Sherry, '*In Vivo* ^{23}Na Imaging and Spectroscopy of the Rat Brain During Infusion of the Shift Reagent $TmDOTP^{5-}$', *J. Magn. Reson. Imag.*, **2**, 385–391 (1992).

202. N. Bansal, M.J. Germann, V. Seshan, G.T. Shires, C.R. Malloy, A.D. Sherry, 'Thulium-1,4,7,10-Tetraazacyclodecane-1,4,7,10-Tetrakis(Methylene Phosphonate) as a ^{23}Na Shift Reagent for the *In Vivo* Rat Liver', *Biochemistry*, **32**, 5638–5643 (1993).

203. V. Seshan, M.J. Germann, P. Preisig, C.R. Malloy, A.D. Sherry, N. Bansal, '$TmDOTP^{5-}$ as a ^{23}Na Shift Reagent for the *In Vivo* Rat Kidney', *Magn. Reson. Med.*, **34**, 25–31 (1995).

204. Z. Xia, J.W. Horton, P.Y. Zhao, N. Bansal, E.E. Babcock, A.D. Sherry, C.R. Malloy, '*In Vivo* Studies of Cellular Energy State, pH and Sodium in Rat Liver after Thermal Injury', *J. Appl. Physiol.*, **76**, 1507–1511 (1994).

205. S.K. Hilal, C.H. Oh, I.K. Mun, A.J. Silver, 'Sodium Imaging', in *Magnetic Resonance Imaging*, eds. D.D. Stark, W.G. Bradley, Mosby, St. Louis, MO, 1091, 1991.

206. C.D. Cone, 'Unified Theory on the Basic Mechanism of Normal Mitotic Control and Oncogenesis', *J. Theor. Biol.*, **30**, 151–158 (1971).

207. H.J. Berendsen, H.T. Edzes, 'The Observation and General Interpretation of Sodium Magnetic Resonance in Biological Material', *Ann. NY Acad. Sci.*, **204**, 459–485 (1973).

208. J. Pekar, P.F. Renshaw, J.S. Leigh, 'Selective Detection of Intracellular Sodium by Coherence-transfer NMR', *J. Magn. Reson.*, **72**, 159–161 (1987).

209. L.A. Jelicks, R.K. Gupta, 'Double-quantum NMR of Sodium Ions in Cells and Tissues. Paramagnetic Quenching of Extracellular Coherence', *J. Magn. Reson.*, **81**, 586–592 (1989).

210. L.A. Jelicks, R.K. Gupta, 'Observation of Intracellular Sodium Ions by Double Quantum Filtered ^{23}Na NMR with Paramagnetic Quenching of Extracellular

Coherence by Gadolinium Tripolyphosphate', *J. Magn. Reson.*, **83**, 146–151 (1989).

211. L.A. Jelicks, R.K. Gupta, 'On the Extracellular Contribution to Multiple Quantum Filtered ^{23}Na NMR of Perfused Rat Heart', *Magn. Reson. Med.*, **29**, 130–133 (1993).

212. V. Seshan, A.D. Sherry, N. Bansal, 'Evaluation of Triple Quantum Filtered ^{23}Na NMR Spectroscopy in the *In Situ* Rat Liver', *Magn. Reson. Med.*, **38**, 821–827 (1997).

213. S.R. Thomas, 'The Biomedical Applications of Fluorine-19 NMR', in *Magnetic Resonance Imaging*, eds. C.L. Partain, R.R. Price, J.A. Patton, W.B. Saunders, New York, 1536–1552, Vol. 2, 1988.

214. M.J.W. Prior, R.J. Maxwell, J.R. Griffiths, 'Fluorine-19F NMR Spectroscopy and Imaging', in *In Vivo Magnetic Resonance Spectroscopy III*, eds. M. Rudin, J. Seelig, Springer-Verlag, New York, 102–130, 1992.

215. B.S. Selinsky, C.T. Burt, '*In Vivo* ^{19}F NMR', *Biol. Magn. Reson.*, **11**, 241–276 (1992).

216. A.N. Stevens, P.G. Morris, R.A. Iles, P.W. Sheldon, J.R. Griffiths, '5-Fluorouracil Metabolism Monitored *In Vivo* by ^{19}F NMR', *Br. J. Cancer*, **50**, 113–117 (1984).

217. W. Wolf, M.J. Albright, M.S. Silver, H. Weber, U. Reichardt, R. Sauer, 'Fluorine-19 NMR Spectroscopic Studies of the Metabolism of 5-Fluorouracil in the Liver of Patients Undergoing Chemotherapy', *Magn. Reson. Imaging*, **5**, 165–169 (1987).

218. M.J.W. Prior, R.J. Maxwell, J.R. Griffiths, '*In Vivo* ^{19}F NMR Spectroscopy of the Antimetabolite 5-Fluorouracil and its Analogues. An Assessment of Drug Metabolism', *Biochem. Pharmacol.*, **39**, 857–863 (1990).

219. W. Semmler, P. Bachert-Baumann, F. Guckel, F. Ermark, P. Schlag, W.J. Lorenz, G. van Kaick, 'Real-time Follow Up of 5-Fluorouracil Metabolism in the Liver of Tumor Patients by Means of F-19 Magnetic Resonance Spectroscopy', *Radiology*, **174**, 141–145 (1990).

220. D. Eidelberg, G. Johnson, D. Barnes, P.S. Tofts, D. Delpy, D. Plummer, W.I. McDonald, '^{19}F NMR Imaging of Blood Oxygenation in the Brain', *Magn. Reson. Med.*, **6**, 344 (1988).

221. J.E. Fishman, P.M. Joseph, T.F. Floyd, B. Mukherji, H.A. Sloviter, 'Oxygen Sensitive ^{19}F NMR Imaging of the Vascular System In Vivo', *Magn. Reson. Imaging*, **5**, 279–285 (1987).

222. R.P. Mason, R.L. Nunnaly, P.P. Antich, 'Tissue Oxygenation: A Novel Determination Using ^{19}F Surface Coil Spectroscopy of Sequestered Perfluorocarbon Emulsions', *Magn. Reson. Med.*, **18**, 71 (1991).

223. C.J. Deutsch, J.S. Taylor, 'Intracellular pH as Measured by ^{19}F NMR in Physiological NMR Spectroscopy: From Isoloated Cells to Man', *Ann. NY Acad. Sci.*, **508**, 33–47 (1987).

224. B.A. Berkowitz, J.T. Handa, C.A. Wilson, 'Perfluorocarbon Temperature Measurements Using ^{19}F NMR', *NMR Biomed.*, **5**, 65–68 (1992).

Multinuclear Magnetic Resonance Spectroscopic Imaging

Todd L. Richards
University of Washington, Seattle, USA

1 Introduction **1**
2 Techniques for Spectroscopic Imaging **2**
2.1 Hardware 2
2.2 Fourier Methods with Phase-encoding Gradients 2
2.3 Spectroscopic Imaging with Selective Irradiation 3
2.4 Spectroscopic Imaging of Nuclei with Extremely Short Spin–Spin Magnetic Resonance Relaxation Time 3
2.5 Echo-planar Spectroscopic Imaging 4
2.6 Data Analysis Algorithms for Spectral Quantification 4
3 Applications of Spectroscopic Imaging **5**
3.1 Phosphorus-31 5
3.2 Carbon-13 7
3.3 Fluorine-19 8
3.4 Multinuclear Single Resonance Imaging 10
4 Summary and Conclusions **10**
Abbreviations and Acronyms **11**
Related Articles **11**
References **11**

Magnetic resonance spectroscopic imaging (MRSI) is a technique for producing spatially resolved maps or images of different chemicals that are found in the body or tissue. These chemicals may be naturally inherent or may be introduced by injection, but they must be detectable by the magnetic resonance (MR) process (i.e. they must have nuclei that have a magnetic moment, and are mobile enough to give a free-induction decay (FID)). The MRSI technique is based on many of the same principles that are used for conventional magnetic resonance imaging (MRI) and also uses much of the same equipment as MRI. This chapter focuses on multinuclear applications of MRSI such as phosphorus-31, carbon-13, fluorine-19, sodium-23, xenon-129, helium-3, and boron-11. The MRSI method is an important technique for noninvasive measurement of biochemical and physiological changes in health and disease.

1 INTRODUCTION

An MR spectrum of nuclei such as ^{31}P or ^{13}C usually contains many peaks or resonances, and each resonance signal arises from the nuclei that are in slightly different local magnetic fields due to the molecular environment. This process is called chemical shift and is described in depth by Gadian.[1] In spectroscopic imaging, the resonances of the spectrum can be separated either by pulse sequences, designed to zoom in to a specific resonance of the spectrum, or by post-processing techniques. These different spectral separation procedures are discussed in section 2 of this article. A chemical of the body may have several resonances in the MR spectrum which can be used for imaging. For example, adenosine triphosphate (ATP) gives rise to three main resonances in the ^{31}P spectrum due to the different molecular environment of the three different phosphate groups in ATP. Each of these three resonances can be imaged, provided that there is sufficient spectral resolution for peak separation. In the case of ATP spectroscopic imaging, images can be created by line-fitting all of the resonances that belong to ATP and combining them into a single concentration value for each spatially resolved voxel. There are also chemicals and spectra that have only one resonance, such as boron-11.

Spectroscopic imaging is sometimes called chemical shift imaging (CSI) and was first developed by Brown et al. in 1982.[2] They described a pulse sequence that used phase-encoding gradients to spatially encode the MR signal, but acquired the rf (radiofrequency) signal without a readout gradient. Thus MRSI works by combining spectroscopic and imaging techniques into a single pulse sequence. Spectroscopic imaging also has greater signal-to-noise than single voxel techniques because the MR signal comes from the whole volume being imaged. Single-voxel spectroscopy (or point-by-point spectroscopy) suffers from the low sensitivity that occurs in all point-imaging techniques.[3,4] Lauterbur et al. in 1975 described a technique for using the increased sensitivity of whole-volume imaging combined with chemical shift spectroscopy.[5] Spectroscopic imaging is basically a technique for mapping the magnetic field distribution which is sensitive to changes in the magnetic field due to chemical shift. In 1979, Maudsley et al. developed a technique for mapping the magnetic field distribution but did not label it as spectroscopic imaging or CSI.[6] This technique is quite similar to the MR techniques used in CSI.[2,7–10] One main difference

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

between conventional MRI (used for clinical diagnosis) and MRSI is in the pulse sequence (a pulse sequence is software written to control the hardware of the MR process such as the rf transmission, rf signal reception, and x, y, and z magnetic field gradients). The pulse sequence used for conventional MRI is designed to image the single resonance of the proton nuclei in water and, as water is usually over a hundred times more concentrated than other hydrogen-containing compounds, no spectral-separation techniques are needed (except in the case of lipid suppression). Also for conventional MRI, a readout magnetic field gradient (a magnetic field gradient causes a linear change in the strength of the magnetic field in one spatial direction and is superimposed on the main magnetic field) is turned on during the acquisition of the rf signal. This readout gradient spreads out the frequencies of the water signal over several thousand cycles per second (Hz), so that the chemical shift information is almost completely obscured by the gradient. However, in spectroscopic imaging, chemical shift information can be preserved by acquiring the rf signal (FID) without a readout gradient. In order to obtain a high-quality MR spectrum, the main magnetic field needs to be very homogeneous and this allows spectral separation of the different resonances in the spectrum. If there is only one resonance in the spectrum then a readout gradient can be used in the pulse sequence to spatially map the chemical.

2 TECHNIQUES FOR SPECTROSCOPIC IMAGING

2.1 Hardware

The hardware for nonproton MRSI is much the same as used for conventional proton MRI. The hardware required here must include electronic components that are necessary for multinuclear MR spectroscopy and MRI. The common components of multinuclear MRSI and conventional MRI include the main magnet, the magnetic field gradient coils and amplifier, the computer, and digital-to-analog controller boards for each channel. However, the components that are different for multi-nuclear MRSI include the rf receiver and transmitter coils and the rf amplifier. For MRSI and single-voxel MR spectroscopy, rf coils are constructed (using resistors, capacitors, and inductors) to match the impedance and to tune the frequency to coincide with the resonant frequency for the specific nuclei of interest. For example, at 1.5 T, ^{31}P resonates at 25.8 MHz, ^{13}C at 16.05 MHz, ^{19}F at 60.08 MHz, and ^{11}B at 20.5 MHz. There are also commercial rf coils available that are double-tuned for both protons and the main nuclei of interest (hereafter referred to as the x nuclei, such as ^{31}P or ^{13}C). The rf amplifier needs to be broad-band enough in rf frequency to tune to each of the frequencies listed above. The ability to detect both protons and the x nuclei is necessary in MRSI because the stronger signal strength of the protons is used for real-time adjusting of the magnetic field homogeneity (hereafter referred to as shimming). Therefore, two rf channels are used so that one channel can be switched to protons during the shimming process and then switched to the x-nucleus channel for main signal acquisition. Another reason for having two rf channels is for proton-decoupling the x nuclei (proton-decoupling is commonly used for both ^{31}P and ^{13}C). The reason for decoupling the protons is to obtain spectra with higher signal-to-noise and also reduce the spectral splitting caused by j coupling (j coupling is caused by energy exchange between adjacent nuclei). For example, in molecules containing a $-CH_3$ group, the ^{13}C-NMR (nuclear magnetic resonance) spectrum will show this methyl carbon split into four peaks with the overall intensity also split among the four peaks at a ratio of 1 : 3 : 3 : 1. However, if this same molecule is measured with proton decoupling, there will be a single ^{13}C spectral peak with much higher intensity than the coupled spectrum. In proton-decoupling schemes, the proton channel is switched on and off even during pulse sequence. For rf signal detection, investigators have used surface coils,[11–13] bird-cage coils,[14–16] and saddle coils[17] for spectroscopic imaging.

2.2 Fourier Methods with Phase-encoding Gradients

CSI as proposed by Brown et al.[2] is performed using a pulse sequence with phase-encoding gradients. An example of a CSI pulse sequence is shown in Figure 1. In this pulse sequence, an excitation rf pulse places the net magnetic vector of the spins in the x–y plane, then the phase-encoding gradients are used to spatially encode the MR signal, and finally the MR signal is acquired without any gradients. This pulse sequence allows the chemical shift information of the MR signal to be preserved while the phase of the spins has been spatially encoded (the processing of this raw MR signal is discussed in section 2.3). This basic pulse sequence can be modified in the following several ways:

1. The excitation rf pulse can be either broad-band enough to excite the whole sample (not shown) or narrow band ($\sin x/x$ pulse) combined with a magnetic field gradient for slice selection (as shown in Figure 1). Using this same concept, CSI can be combined with pulse sequences such as the 90–180 spin-echo (shown in Figure 1), with point-resolved echo spectroscopy (PRESS; for cubic localized volume)[18] or with stimulated echo acquisition mode (STEAM; for cubic localized volume).[19]

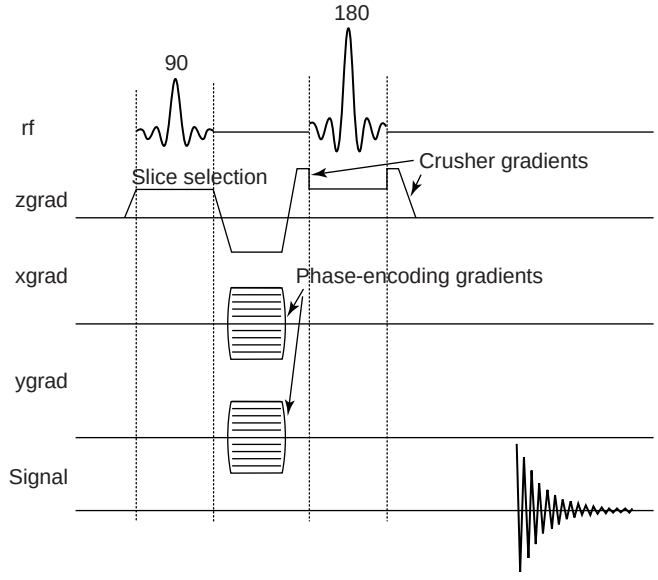

Figure 1 CSI pulse sequence using a spin-echo excitation, slice selection, and two dimensions of phase-encoded gradients (for encoding two dimensions of space).

2. The phase-encoding gradients can be used to encode spatial information in one, two (as shown in Figure 1) or three orthogonal directions, typically called the x, y, and z directions, where x is the subject's left to right direction, y is the subject's anterior to posterior direction (front to back), and z is the subject's superior to inferior (head to toe).
3. The spatial resolution of the spectroscopic imaging process can be controlled by the number of phase-encoded gradients for each spatial direction (typically, between 8 and 16 phase-encoded gradient steps are used).
4. Preparatory pulses can be added before the first rf excitation pulse for peak suppression, spatial suppression, or for spectral editing.

In a typical experiment, the pulse sequence is executed 16 times, while the x phase-encode gradient is incremented (the y phase-encode is held constant for the 16 x-value increments) and this whole set of steps is repeated 16 times for each y phase-encode step (i.e. a nested loop). Therefore, the total number of repetitions of the pulse sequence is 16×16, which is 256.

Each time the pulse sequence is executed for a unique set of x and y phase-encoded gradient values, the MR signal is digitized and stored on computer disk as a row of data. When the experiment is done there are 256 rows of raw time-domain data.

Computer processing of this data is used to reconstruct the spatial and chemical shift information from the time-domain MR signal. One possible algorithm has the following steps:

1. read and sort the raw data into a three-dimensional (3D) array of size (256,16,16);
2. filter with apodization and Fourier transform the data in the chemical shift dimension and put the result back into the 3D array;
3. filter with a Gaussian or sine-bell function and Fourier transform the data in the x direction;
4. filter with a Gaussian or sine-bell function and Fourier transform the data in the y direction;
5. form metabolite images or maps from each resonance of interest (this procedure is described in section 2.6).

2.3 Spectroscopic Imaging with Selective Irradiation

Bottomley et al.[20] developed a technique for imaging single resonances of the MR spectrum. This technique has the advantage of shorter acquisition times compared to conventional CSI because a read-out gradient can be on during the acquisition of the signal. However, this technique has the disadvantages of less signal-to-noise and also only one resonance can be scanned at a time compared to conventional CSI which is spatially localized with phase-encoded gradients (described in section 2.2). Here is how the pulse sequence works:

1. A chemically selective rf pulse is used to excite the resonance of interest and consists of a narrow-band rf pulse which is applied in the absence of imaging gradients.
2. Spatial mapping is achieved with imaging gradients for slice-selection, phase-encoding, and readout, as used in conventional MRI.

Other investigators have used chemical shift selective (CHESS) MRI to study protons[21–24] and also phosphocreatine (PCr).[25] Ernst et al. were able to perform direct imaging of PCr and inorganic phosphate (Pi) with or without selective excitation of PCr in the human brain and limb.[25]

2.4 Spectroscopic Imaging of Nuclei with Extremely Short Spin–Spin Magnetic Resonance Relaxation Time

For imaging nuclei such as boron-11 with T2 (spin–spin relaxation time) values less than 1 ms, spin echoes with phase encoding is not the best method because the MR signal will die out before the echo signal can be measured. Also, with only one peak visible in vivo, techniques for multiple-peak spectral acquisition are unnecessary for chemicals such as borocaptate sodium (BSH); therefore Glover et al.[26,27] developed a technique that uses

three-dimensional projection reconstruction (3DPR) to image boron in vivo. In this case, BSH was imaged because of its use in boron neutron capture therapy (BNCT) which is a treatment for brain cancers such as glioblastoma multiforme. The 3DPR technique has the advantage over CSI methods because acquisition of the MR signal can take place immediately after the rf excitation pulse. Lauterbur proposed the use of projection reconstruction for MRI in 1973.[28] Projection reconstruction is performed by measuring the MR signal in the presence of a readout gradient (which is Fourier transformed to give a spatial projection in one dimension), then the spatial direction of the readout gradient is varied in three dimensions to acquire several projections. These one-dimensional projections are then processed by a filtered back-projection, in order to spatially reconstruct the image.

2.5 Echo-planar Spectroscopic Imaging

Although fast spectroscopic imaging techniques such as echo-planar spectroscopic imaging (EPSI) have been mainly used for protons,[29–31] there are a few published uses of EPSI for nonproton use.[32,33] Spectroscopic imaging techniques that use phase-encoding gradients for all spatial localization suffer from long encoding times. Mansfield developed a rapid spectroscopic imaging technique called EPSI that used a much faster encoding scheme and employed spatial encoding interleaved into the spectral acquisition.[34] The EPSI technique uses echo-planar gradients during the acquisition of the FID MR signal and allows multiple points of *k*-space (a term used to describe the two- or three-dimensional space or matrix for the raw time domain MRI data before the Fourier transform; it includes the phase-encoded and readout-encoded dimensions) to be collected during one cycle of the pulse sequence (during one repetition time of the MR pulse sequence (TR) period). One example of this technique to achieve two spatial dimensions and one chemical shift dimension is described by Posse et al.[31] In this example, a slice-selective spin-echo signal is excited with a 90-tau-180 sequence (where 90° is the flip angle of the first transmission rf pulse; tau is the delay between the two pulses and 180° is the flip angle of the second transmission rf phase) and then a square-wave (echo-planar) gradient is turned on during the acquisition of the MR signal. This results in an MR signal with both spatial and chemical shift information (each individual echo contains the spatial information and the overall evolution of the signal contains the chemical-shift information). A phase-encoding gradient (in an orthogonal direction to the echo-planar gradient) is used to give another spatial dimension. After all the phase-encoded MR data is collected (during several TR periods), the MR data is sorted into a 3D array so that a three-dimensional Fourier transform (3DFT) can be performed. Other examples of EPSI have been described.[35–39]

2.6 Data Analysis Algorithms for Spectral Quantification

For spectroscopic imaging, there are usually too many spectra for manual quantification of peak areas; therefore, automated techniques have been developed for processing the raw data.[40,41] The following steps can be performed to generate a metabolite map, that is, an image of a specific chemical from the spectrum, from phase-encoded CSI:

1. Filter and Fourier transform the raw data to obtain spatially localized spectra (in the example above, there was a 3D matrix containing two dimensions of spatial information and one dimension of chemical shift information).
2. Phase each spectrum (i.e. make each resonance have a symmetric Lorentzian absorption-mode shape, as opposed to a dispersion-mode shape).
3. Reference the x axis so that each resonance can be identified based on the frequency scale.
4. Evaluate the area under the curve of each peak either by peak integration or by curve fitting methods.[40] Other spectral parameters that can measured and mapped include the frequency (relative to a specific peak), line-width (related to the T2), and phase of each peak.
5. Create a color or gray-level map from the peak area which was measured at each spatially resolved location (voxel).

Nelson and Brown developed an automated technique for peak detection and evaluation of peak areas called PIQABLE.[42] Barker et al. have reported the use of a nonlinear least-squares fitting routine (NLLS) for determining the amplitude, decay time, frequency, and phase of each resonance in the ^{31}P spectrum.[43] They were able to fit the following seven resonances: phosphomonoester (PME), Pi, phosphodiesters (PDEs), PCr, and the g, a, and b resonances of ATP.[43] Van den Boogaart et al. have reported on the use of two frequency domain methods (peak area integration and Lorentzian fitting (FITSPEC)) and two time domain methods (VARPOR and HLSVD) for in vivo ^{31}P-MRS (magnetic resonance spectroscopic) quantification.[44] Several other methods are available for spectral quantification.[12,41,45–50]

3 APPLICATIONS OF SPECTROSCOPIC IMAGING

3.1 Phosphorus-31

Phosphorus-31 MRSI can be used to study metabolites involved in energy metabolism such as ATP, PCr, Pi and sugar phosphates. In addition, ^{31}P can be used to study PMEs and PDEs involved in lipid metabolism. Brief summaries of ^{31}P applications to brain and muscle are discussed in this section.

3.1.1 Phosphorus-31 Spectroscopic Imaging of the Brain

An early description of ^{31}P spectroscopy imaging (SI) was provided by Maudsley and colleagues.[7,51] In their 1984 study, MRI techniques were applied to in vivo spectroscopic analysis of spatially resolved phosphorus spectra in the cat brain to determine whether changes associated with stroke could be detected. Two-dimensional images of separate phosphorus-containing compounds, as well as spectra arising from spatially localized points, demonstrated that metabolism within the tissue could be monitored in this manner. They also demonstrated preliminary results of phosphorus imaging of the human body in vivo.[51] Later, they were able to obtain ^{31}P spectroscopic images at 3.5 cm spatial resolution in the human brain with times of 37 min at 2.0 T.[52] Sasahira et al. reported ^{31}P-SI changes immediately after an ischemic insult, even though MRI and CT (computerized tomography) were negative.[53] Hubesch et al. reported ^{31}P metabolite concentrations were significantly reduced in tumors, infarcts, and deep white matter lesions.[54] Three reports have demonstrated the feasibility of interleaving ^{31}P and proton spectroscopic imaging.[14,55,56] This would be important because there is information that is mutually exclusive in both ^{31}P and proton spectroscopy. Gonen et al. developed a ^{31}P-SI technique that was a hybrid of two localized spectroscopy techniques, CSI and Hadamard spectroscopic imaging (HSI).[57] With this technique, they were able to shorten the acquisition time and also maintain intervoxel spectral isolation.[57] Gonen et al. also used a technique with polarization transfer from ^{1}H to *j*-coupled ^{31}P, called refocused insensitive nucleus enhancement by polarization transfer (RINEPT) and combined it with 3D CSI sequences on a clinical imager.[58] The technique was demonstrated on a phantom and in vivo human brain. Compared with direct ^{31}P Ernst-angle excitation, signal gains of up to $\times 1.8$ were obtained mainly as a result of spin–lattice relaxation time (T1) differences between ^{31}P and ^{1}H. Spectral interpretation was also simplified by editing out all non-proton-coupled ^{31}P signals.[58] Example ^{31}P-SI in the brain and the polarization transfer spectra are shown in Figure 2.

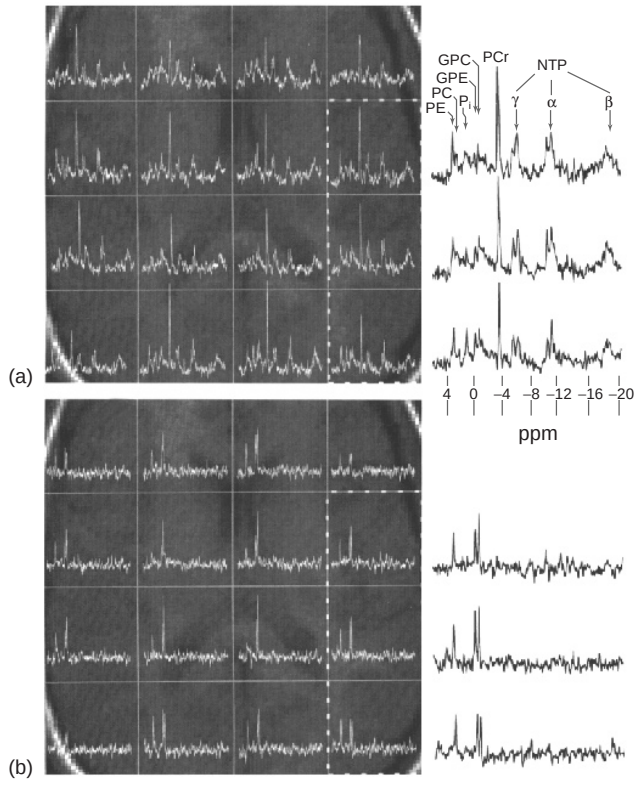

Figure 2 Two 4×4 arrays of ^{31}P spectra from an axial slice of back-to-back $8 \times 8 \times 8$ proton-decoupled 3D CSI on the brain of a female volunteer. The spectra, superimposed on an axial MRI for anatomical reference, are on a common vertical scale, span 750 Hz (−18 to 5 ppm), with a 3×3 cm grid to mark nominal voxel boundaries. (a) Direct excitation with an Ernst angle of 55°, at TR 1.5 s. (b) ^{31}P RINEPT. The spectra from the column of highlighted voxels in (a) and (b) were extracted, expanded, labeled, and displayed to the right of each image for greater detail. (Reprinted from O. Gonen, A. Mohebbi, R. Stoyanova, T.R. Brown, 'In Vivo Phosphorus Polarization Transfer and Decoupling from Protons in Three-dimensional Localized Nuclear Magnetic Resonance Spectroscopy of Human Brain', *Magn. Reson. Med.*, **37**(2), 301–306, Copyright (1997) by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

Jung et al. were able to measure the *j*-coupling constants of ATP in the brain using ^{31}P-SI.[59] Both the *j*-coupling constants and the chemical-shift difference between α-ATP and β-ATP were used to calculate the concentration of intracellular free magnesium in the brain.[59] A very recent report by Barker et al.[43] shows superb ^{31}P multislice SI in the human brain at 3 T (most clinical MRI scanners have magnets of only 1.5 T). Images of free magnesium concentrations and pH as well as PMEs, PDEs, Pi, PCr, and ATP were calculated from the SI data.[43] An example of ^{31}P metabolite images is shown in Figure 3.

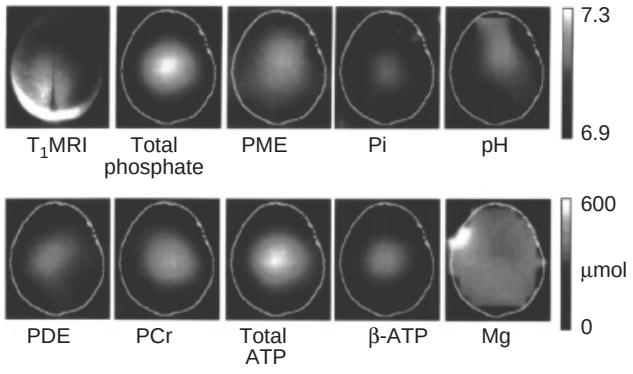

Figure 3 Phosphorus SI data of the human brain. Images of axial T1 localizer MRI, total phosphate, PME, Pi, pH, PDE, PCr, total ATP, β-ATP, and magnesium are presented. Metabolite images are overlaid with an outline of the skull from the T1 MRI. Phosphate images are windowed arbitrarily but on the same gray scale. Quantitative pH images (gray-level images with the image intensity displayed as a pH value) are shown on a scale ranging from pH 6.9 to 7.3. Magnesium images are on a scale ranging from 0 to 600 mM. Only the most superior slice of three anatomical slices is presented here from the original figure. (Reprinted from P.B. Barker, E.J. Butterworth, M.D. Boska, J. Nelson, K.M. Welch, 'Magnesium and pH Imaging of the Human Brain at 3.0 Tesla', *Magn. Reson. Med.*, **41**(2), 400–406, Copyright (1999) by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

In an effort to understand white matter MR signal hyperintensities in aging, Constans et al. used ^{31}P-SI and ^{1}H-SI.[60] They demonstrated that white matter lesions were associated with increased signal from choline-containing metabolites, no significant change of signal from *N*-acetylaspartate, and a trend to a decreased PME resonance.[60] Richardson et al. demonstrated an increased level of PME in the dyslexic brain compared to controls, using a four-dimensional SI technique.[61] Deicken et al. used ^{31}P-SI to demonstrate that schizophrenics had a significantly higher right relative to left temporal PCr/ATP ratio, PCr/Pi, and PCr, as well as significantly lower right relative to left temporal ATP.[62] Both the left temporal PCr and the degree of asymmetry of temporal lobe PCr were highly correlated with the thinking disturbance subscale of the Brief Psychiatric Rating Scale.[62] Husted et al. investigated myelin phospholipids in vivo in multiple sclerosis lesions and normal-appearing white matter by evaluating the spectral broad component from ^{31}P MRSI data.[63] They measured a reduced phospholipid broad component in both lesions and in normal-appearing white matter in multiple sclerosis subjects compared to control subjects, suggesting reduced myelin phospholipid concentration or altered relaxation times.[63]

Chu et al. used ^{31}P-SI to compare the phosphorous metabolite ratios in the mesial temporal lobe of healthy volunteers with the corresponding ratios in patients with temporal lobe epilepsy.[64] They were able to lateralize the seizure focus in temporal lobe epilepsy patients using various phosphorous metabolite ratios – PCr/Pi, PCr/γ-ATP, and γ-ATP/Pi – and to compare with clinical lateralization results. They found an average reduction of 15% in the PCr/Pi and γ-ATP/Pi ratios compared with the corresponding ratios in healthy volunteers in the entire mesial temporal lobe, and more than a 30% reduction in these two ratios in the anterior region of the epileptogenic mesial temporal lobe. The lateralization based on either the PCr/Pi or the γ-ATP/Pi ratio yielded a correspondence of 70–73% with the final clinical lateralization. Compared with the intracranial electroencephalogram (EEG), a 78% correspondence was found with the ^{31}P-NMR-based lateralization, whereas MRI provided a correspondence of only 33%, and scalp EEG provided a correspondence of only 56%. These results suggest the utility of adding the ^{31}P-NMR method to the group of noninvasive modalities used for presurgical decision-making in temporal lobe epilepsy patients.[64] Other investigators have encouraging results for the use of ^{31}P-SI in epilepsy.[54,65–70]

3.1.2 Phosphorus-31 Spectroscopic Imaging of the Heart and Skeletal Muscle

Bottomley et al. have used ^{31}P-MRSI in the heart to measure metabolite concentrations of PCr and ATP.[11] Schaefer et al. used ^{31}P-SI to measure abnormalities in patients with cardiomyopathy and left ventricular hypertropy.[71] Robitaille et al. were able to overcome surface coil rf field inhomogeneity artifacts using adiabatic pulses during ^{31}P-SI of the canine heart.[72] Auffermann et al. tested the concept that cine MRI and ^{31}P-MR spectroscopy might be used to provide a comprehensive evaluation of the functional and metabolic status of the myocardium in humans.[73] The result was that the ratios PDE/PCr and PDE/β-ATP were significantly higher in patients with dilated cardiomyopathy compared with healthy volunteers. They suggested that increased myocardial PDE might be a marker for abnormal myocardium.[73] Hetherington et al. used high-field (4 T) three-dimensional spectroscopic imaging (3DSI) to acquire PCr images of the human heart with 8 cc voxels.[74] The PCr images enabled observation of the septum, left ventricular free wall, apex, and skeletal muscle. Quantitative evaluation of the 50 myocardial voxels acquired from 10 studies of healthy adults revealed a PCr/ATP ratio of 1.80 ± 0.32 after correction for saturation effects.[74] An example of ^{31}P-SI data of the heart is shown in Figures 4 and 5.

Campbell et al. used ^{31}P-MRSI at 2 T to monitor high-energy phosphate metabolism over a 3-week period in a canine model of myocardial infarction and reperfusion.[75]

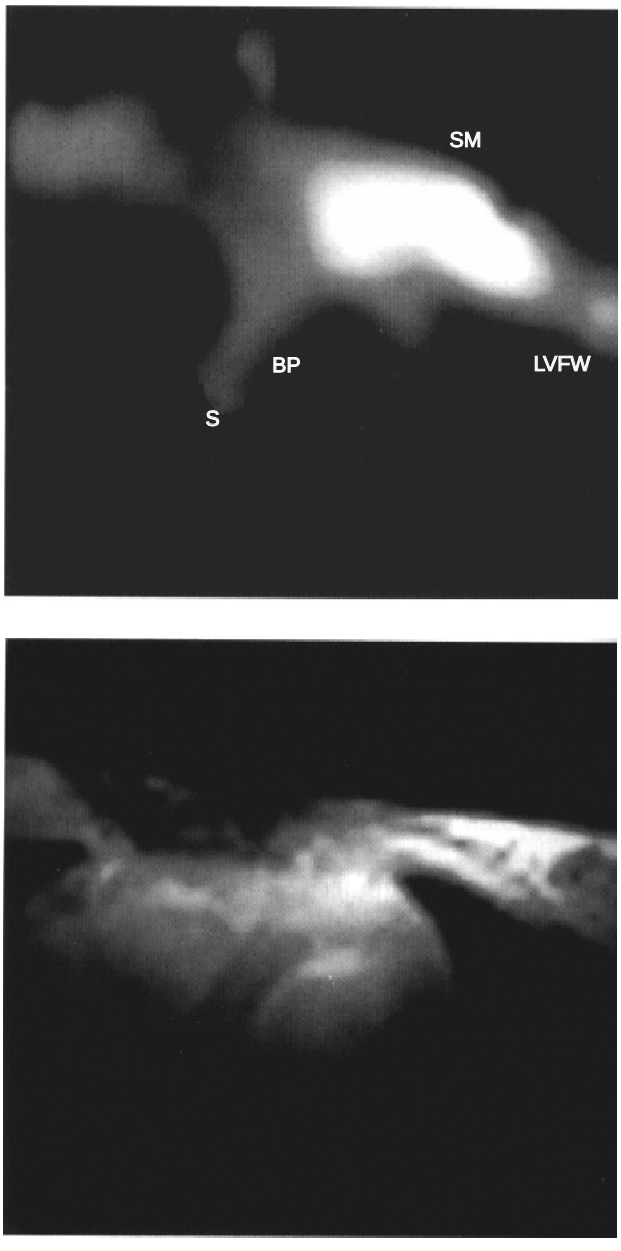

Figure 4 (a) A ^{31}P-SI of PCr in the human heart and surrounding tissue. The labeled areas are left ventricular free wall (LVFW), septum (S), skeletal muscle (SM), and blood pool (BP). (b) The corresponding MRI slice. (Reprinted from H.P. Hetherington, D.J. Luney, J.T. Vaughan, J.W. Pan, S.L. Ponder, O. Tschendel, D.B. Twieg, G.M. Pohost, '3D ^{31}P Spectroscopic Imaging of the Human Heart at 4.1 T', *Magn. Reson. Med.*, **33**(3), 427–431, Copyright (1995) by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

A metabolic protective effect was observed (versus controls) with both nitroglycerin and superoxide dismutase, manifested by a reduction in the degree of pH decline from baseline values and preservation of the ATP/total phosphate ratio during occlusion and reperfusion. This

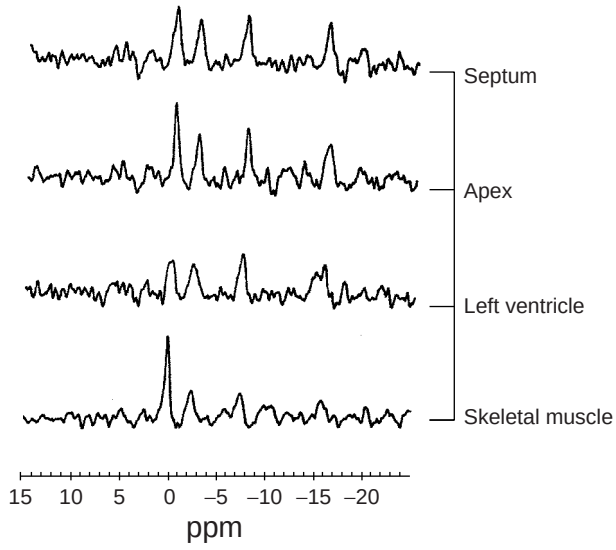

Figure 5 Single voxel (8 cc) spectra from the apex, septum, left ventricle, and skeletal muscle. (Reprinted from H.P. Hetherington, D.J. Luney, J.T. Vaughan, J.W. Pan, S.L. Ponder, O. Tschendel, D.B. Twieg, G.M. Pohost, '3D ^{31}P Spectroscopic Imaging of the Human Heart at 4.1 T', *Magn. Reson. Med.*, **33**(3), 427–431, Copyright (1995) by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

was an excellent demonstration that ^{31}P-SI can be used for the noninvasive assessment of myocardial metabolism in response to therapeutic intervention.[75] This same group also measured the ^{31}P metabolic effects during myocardial ischemia and sustained reperfusion of two other interventions (antianginal agents, diltiazem and propranolol).[76] Jeneson et al. used two-dimensional phosphorus chemical shift imaging (2D/^{31}P/CSI) to investigate macroscopic heterogeneity within the flexor digitorum profundus (FDP) muscle of the human forearm during exercise.[77] They measured differences in Pi and PCr within the FDP in regions of active and inactive fibers during flexion of individual fingers. These results of this study demonstrate the potential of 2D/^{31}P/CSI for in vivo investigation of intramuscular heterogeneity in human skeletal muscle.[77] A similar study was performed by Kushmerick et al., where the distribution of muscle fiber types were measured by ^{31}P-SI.[17] The result was that the PCr/ATP and PCr/Pi ratios were significantly higher in the medial gastrocnemias and tibialis anterior muscles than those for the lateral gastrocnemius and soleus.[17]

3.2 Carbon-13

The greatest advantage for carbon-13 spectroscopic imaging over protons and phosphorus is that a ^{13}C-labeled chemical can be introduced into the biological system and both the metabolic incorporation and breakdown can be measured noninvasively in biochemical pathways

such as in carbohydrate or lipid metabolism. This is possible because the natural occurrence of ^{13}C is only 1% of the carbon nuclei which are mostly ^{12}C (with no MR signal). It has the advantage over nuclear medicine techniques because not only can the presence of the label be spatially detected but also the chemical form of the label can be detected. Direct ^{13}C-SI has much less signal-to-noise ratio than both ^{1}H-SI and ^{31}P-SI. However, there are some techniques that allow the carbon nuclei to be indirectly observed through the adjacent protons. Van Zilj et al. measured the metabolism of [1-^{13}C]-glucose in the cat brain using gradient-enhanced proton-detected heteronuclear spectroscopy.[78] By observing the protons which were coupled to ^{13}C nuclei, they were able to detect [1-^{13}C]glucose, [3-^{13}C]lactate, [4-^{13}C]glutamine, [4-^{13}C]glutamate and the combined signals [2-^{13}C]glutamate/glutamine and [3-^{13}C]glutamate/glutamine, despite the fact that some of the associated proton resonances were close to the water signal.[78] Lee et al. proposed a scheme for acquiring ^{13}C spectroscopic images using inverse detection via polarization transfer through protons and they experimentally verified the technique by a phantom using a heteronuclear multiple-quantum coherence technique.[79] Using this scheme, they were able to gain signal and have an improvement of the time efficiency of data acquisition compared to the conventional 3DFT technique without sacrificing the most important attribute of ^{13}C, the enormity of the chemical shifts.[79]

Using direct detection of carbon-13, Hammer showed that proton-decoupled ^{13}C images had an increase in sensitivity over nondecoupled ^{13}C images because of the nuclear Overhauser effect (NOE) and elimination of multiple lines from scalar $^{13}C-^{1}H$ spin–spin couplings.[80] Morishita et al. used a CHESS pulse to image ^{13}C-labeled glucose in the rat head at 7 T and at a spatial matrix size of 64×64.[81] Beckmann et al. measured proton-decoupled ^{13}C-NMR spectra of the human head during hyperglycemic glucose clamping using intravenous infusions of [1-^{13}C]glucose in normal volunteers.[82] The spatial localization of the different regions of interest was confirmed by ^{13}C-NMR spectroscopic imaging with a time resolution of 9 min. They were able to measure the time course of metabolic breakdown products of [1-^{13}C]glucose and also ^{13}C-label incorporation into the C2, C3, and C4 resonances of glutamate/glutamine and into lactate in the human brain.[82] Carbon-13 SI in combination with ^{13}C labeled glucose makes it possible to obtain time-resolved, spatially selective, and chemically specific information from the human body.[82] An example of ^{13}C-spatially resolved spectra from the human brain is shown in Figure 6. All of these results suggest the possibility of measuring spatially resolved chemical information from the human body that may have a clinical application for detecting metabolic disorders.

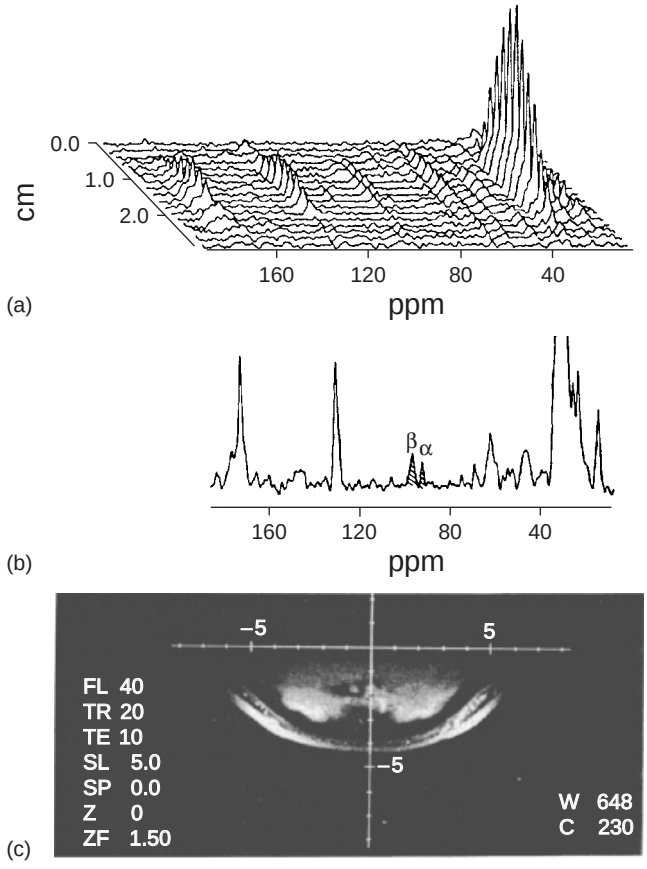

Figure 6 Surface coil ^{13}C-SI applied to the head of a normal volunteer after injection of glucose labeled at the C1 position. (a) Spatially resolved ^{13}C spectra – aquisition time = 9.0 minutes, TR = 1.05 s, eight averages, 64 phase-encoded steps (-5.0 to $5.0\,mT\,m^{-1}$), decoupling with Waltz 8. (b) Spectrum extracted from the ^{13}C map in (a), showing the good signal-to-noise ratio of the resonances corresponding to the enriched substance, despite the low measurement time. (c) MRI image (fast low-angle shot (used as an MR pulse sequence to acquire MR images less than about 5 s) FLASH) of the same region obtained with the surface coil. (Reprinted with permission from N. Beckmann, I. Turkalj, J. Seelig, U. Keller, '^{13}C-NMR for the Assessment of Human Brain Glucose Metabolism In Vivo', *Biochemistry*, **30**(26), 6362–6366, Copyright (1991) American Chemical Society.)

3.3 Fluorine-19

Fluorine MRI also has the advantage of no tissue background signal and the fluorine-containing compound has to be introduced from an external source. This section covers the applications of fluorine MRI to study anatomy, fluoro-deoxy-glucose metabolism, fluorine-containing anesthetics, tissue oxygen tension,

cerebral blood flow, and tumor metabolism of fluorine-containing anticancer drugs.

Schwarz et al. used ^{19}F-MRI in conjunction with perfluorononane to develop a new technique for gastrointestinal (GI) imaging.[83] Due to the lack of ^{19}F background signals, the contrast of the GI tract was only limited by the signal-to-noise ratio of the ^{19}F-MR images. Fluorine-19 RARE (rapid acquisition with refocused echoes) images of 1-mm slices with an in-plane resolution of $0.23 \times 0.23\,mm^2$ were obtained from the GI tract after oral perfluorononane administration. The passage of perfluorononane through the entire GI tract was monitored by repetitive MR measurements with a maximal time resolution of 38 s. The 3D surfaces of the GI tract were reconstructed and superimposed on corresponding ^{1}H-MR images, which provided complementary anatomical information.[83] Duong et al. used [^{19}F] 2-fluoro-2-deoxyglucose-6-phosphate (2FDG-6P) as a compartment-specific marker in normal and globally ischemic rat brain.[84] Assuming that 2FDG-6P reflects water motion, this study showed that water ADC (apparent diffusion coefficient) decreases in both spaces after ischemia, with the reduction of intracellular water motion being the primary source of diffusion-weighted contrast.[84] Venkatasubramanina et al. measured the distribution of two fluorinated anesthetics, halothane and isoflurane, in rabbit brain during brain uptake using ^{19}F-NMR CSI techniques.[85] Use of a short phase-encoding time and FID acquisition minimized T2-related signal losses in two-dimensional CSI. The spatial distribution of halothane and isoflurane in the brain was nonuniform and both anesthetics were detected predominantly in the cerebral cortex. Time-dependent variations in anesthetic concentration were observed in different brain regions. Xu et al. have also used ^{19}F-MRI to measure anesthetic distribution in the brain.[86] Pekar et al. used ^{19}F-MRI techniques to determine wash-in and wash-out curves of the inert, diffusable gas CHF_3 from 0.4-cc voxels in the cat brain.[87] Combined with expired air measurements of CHF_3, they were able to calculate cerebral blood flow values in the 0.4-cc voxels, and the blood flow images were registered with high-resolution ^{1}H MR images. Lee et al. proposed the use of ^{19}F imaging combined with the relaxation enhancement of the gadolinium contrast agent Gd-DTPA.[88]

There is a linear relationship between the T1 relaxation rate of fluorine-19 of a perfluorochemical (PFC) and the partial pressure of the oxygen (pO_2) dissolved in the PFC.[89] Tran et al. measured a linear relationship between the ^{19}F T1 relaxation rate (1/T1) of the 3.5% w/v emulsion and dissolved pO_2 and determined the slope of $0.0033\,s^{-1}\,mm\,Hg^{-1}$ and a correlation coefficient of 0.991.[89] In vivo measurements indicated a 20 mm Hg increase in intracellular pO_2 of liver phagocytes when the inspired gas was changed from 20% to 100% O_2.[89] Pratt et al. have also used a ^{19}F imaging technique (in this case a 3D spatial technique) for mapping oxygen tension.[90] Mason et al. developed a new technique for measuring tumor oxygen tension in vivo based on ^{19}F-NMR spin–lattice echo-planar imaging (EPI) relaxometry of hexafluorobenzene (HFB).[91] They were able to monitor dynamic changes in regional tumor oxygenation in response to respiratory challenge, and local pO_2 was determined using pulse burst saturation recovery ^{19}F-NMR/EPI on the basis of the spin lattice relaxation rate. Following irradiation (20 Gy), tumor oxygenation was significantly elevated and remained high for at least 10 h.[91] Brix et al. measured the extra- and intracellular uptake of the anticancer drug fluorouracil and its major catabolite. α-fluoro-β-alanine (FBAL) using gadolinium-enhanced ^{19}F-MRI.[92] The biodistributions of fluorouracil and FBAL were mapped by metabolic ^{19}F-MRI at about 6 min (early distribution phase) and 64 min (metabolic phase), respectively, after drug administration. They observed that the dominant fluorouracil signal intensity in the liver at 6 min arises from the intracellular compartment, which can be explained as retention of fluorouracil in hepatocytes.[92] Aboagye et al. have demonstrated the feasibility of noninvasive MRSI and SI (CSI) to detect activation of the prodrug 5-fluorocytosine (5-FCyt) to the cytotoxic species 5-fluorouracil (5-FU) by monoclonal antibody cytosine deaminase (CD) conjugates.[93] Other investigators have also used ^{19}F to measure oxygen tension.[36,94–96] After pretreatment of H2981 tumor-bearing mice with L6-CD, in vivo metabolism of 5-FCyt to 5-FU within the tumors was detected by ^{19}F-MRS; the chemical shift separation between 5-FCyt and 5-FU resonances was approximately 1.2 ppm. Whole body ^{19}F-CSI (6×6 mm in-plane resolution) of tumor-bearing mice demonstrated the highest signal intensity of 5-FU within the tumor region.[93] This study supports further development of noninvasive MR methods for preclinical and clinical monitoring of CD enzyme-prodrug therapies. Gonen et al. demonstrated simultaneous acquisition of ^{1}H-decoupled ^{31}P and ^{19}F-3D/CSI in the liver of a patient undergoing 5-FU chemotherapy. Both ^{31}P and ^{19}F shared the same voxel size (64 or 27 mL), bilevel ^{1}H-decoupling and 0.35 s TR. They concluded that simultaneous acquisition of ^{31}P and ^{19}F permits localized study of the influence of hepatic metabolism on the uptake and catabolism of fluoropyrimidine drugs.[97] A well-known radiopharmaceutical 2-deoxy-2-fluoro-D-glucose (FDG) widely used for positron emission tomography diagnosis in terms of glucose utilization, has been evaluated as an NMR pharmaceutical for cancer detection.[98] Kanazawa et al. studied the uptake and metabolism of FDG in the experimental tumor,

MH134, transplanted to the peritoneum of C_3H mice as an ascitic tumor and measured the signal by ^{19}F-NMR.[98] In these tissues, the 6-phosphate of the injected compound was converted reversibly to its epimer 2-deoxy-2-fluoro-D-mannose (FDM) and further to their nucleoside diphosphate (NDP) bound forms. Through the use of in vivo ^{19}F-NMR spectra and ^{19}F-CSI, the ^{19}F-NMR signal of NDP/FDM could be detected 1 day after the FDG injection and could be used as a target signal for tumor detection.[98]

3.4 Multinuclear Single Resonance Imaging

There are several MR-detectable nuclei that do not require SI techniques for spatial mapping because they produce only one resonance. This section covers MRI of sodium, xenon, helium, and boron.

3.4.1 Sodium Imaging

Sodium MRI has been investigated in medicine because of the importance of membrane integrity for maintaining different sodium concentrations in the intracellular and extracellular spaces in healthy tissue. The first in vivo NMR images of tissue sodium were shown by Hilal et al.[99] They were able to measure an increase in the regional sodium NMR signal after brain infarction in an animal stroke model.[99] Because sodium has a short T2 component, Ra et al. developed a technique for sodium imaging with a short echo time of 3.6 ms and this technique was used clinically for in vivo imaging of sodium-23 in the human head.[100] In living tissues, sodium exhibits at least two transverse relaxation constants, namely a short component T2s = 0.7–3.0 ms and a long component T2e = 16–30 ms. Their imaging method combined both projection reconstruction and Fourier encoding schemes and they claim that when using the shorter echo time, there is an increase in the MR signal and improved visibility of intraparenchymal sodium.[100] In another clinical study, Grodd and Klose noted that, in the brain, most of the sodium MR signal comes from the extracellular spaces such as the cerebral spinal fluid (CSF) spaces, whereas the normal parenchyma is not visible. Changes induced by encephalitis, ischemic infarction and tumors were detected by the increased sodium content.[101] Hashimoto et al. scanned patients with brain tumors using sodium MRI and concluded that sodium-23 MRI brought promising results in the evaluation of brain tumors and edema; however, discrimination and quantification of multiple T2 components and improvement of the spatial resolution were desperately needed.[102]

3.4.2 Noble Gases

The use of noble gases in MRI has been reviewed by Kauczor and colleagues.[103,104] In order to increase the MR signal strength of helium (^{3}He) and xenon (^{129}Xe) nuclei, the spin density is artificially enriched (hyperpolarization) by optical pumping and this results in a nonequilibrium polarization five orders of magnitude higher than the Boltzmann equilibrium. Hyperpolarized noble gases can then be applied as ''inhaled contrast media'' using a dedicated application device. A homogeneous distribution of signal intensity can be regarded as a normal finding in people without lung disease. Obstructive diseases and emphysematous changes lead to generalized or localized signal inhomogeneities. Studies using MRI of noble gases for early detection of obstructive lung diseases and disorders of distribution of ventilation in emphysema are promising.[104]

3.4.3 Boron Imaging

Noninvasive techniques for mapping the spatial distribution of boron-containing compounds have been of interest for investigators of BNCT.[26,27,105–111] Although the ^{10}B isotope is the form used with BNCT, ^{11}B-MR has higher signal-to-noise than ^{10}B-MR and can be used to map the spatial distribution. The compound $Na_2B_{12}H_{11}SH$ (BSH) has been suggested for BNCT of glioblastoma multiforme.[27] Bradshaw et al. used a 3DPR method to obtain in vivo ^{11}B images in a large canine brain tumor model and in a human infused with BSH. The 3DPR method demonstrates a signal-to-noise ratio that allows qualitative kinetic studies of the boron compound in normal and in tumor tissue of the head. They were able to obtain human boron images from a patient who underwent surgical resection and volumetric debulking of a large (7 cm) glioblastoma multiforme after BSH injection.[27] Bendel et al. have measured the longitudinal and transverse relaxation rates for the ^{11}B resonances in BSH.[105] These results were used as a basis for assessing the requirements and limitations of using ^{11}B-NMR to determine BSH concentrations in vivo. Bendel et al. also used surface coil ^{11}B-NMR spectroscopy of mice injected with BSH to demonstrate that the BSH MR signal was significantly higher in tumors, compared with muscle tissue, whereas the uptake and clearance kinetics were similar.[105] Richards et al. used a gradient echo pulse sequence (echo time of 0.8 ms) to image boron-11 in rabbits injected with BSH.[111] These studies show an important application of boron MRI that could be used in developing better pharmaceuticals for BNCT as well as for therapy planning during the clinical use of BNCT.

4 SUMMARY AND CONCLUSIONS

This article has discussed several MRSI techniques and applications for several different nuclei. Spectroscopic

imaging allows biomedical scientists and clinicians to map chemicals in the body noninvasively. There are important medical applications of MRSI that involve measuring the change in detectable biochemicals during a physiological perturbation or detecting a change in biochemicals, compared to normal controls, caused by diseases such as cancer or neurological disorders. This technique is suitable for detecting low-molecular-weight chemicals such as ATP or lactate, but is not suitable for directly detecting large-molecular-weight chemicals or chemicals in the solid state such as DNA, enzymes, or bone.

ABBREVIATIONS AND ACRONYMS

ADC	Apparent Diffusion Coefficient
ATP	Adenosine Triphosphate
BNCT	Boron Neutron Capture Therapy
BSH	Borocaptate Sodium
CD	Cytosine Deaminase
CHESS	Chemical Shift Selective
CSF	Cerebral Spinal Fluid
CSI	Chemical Shift Imaging
CT	Computerized Tomography
EEG	Electroencephalogram
EPI	Echo-planar Imaging
EPSI	Echo-planar Spectroscopic Imaging
FBAL	α-Fluoro-β-Alanine
FDG	2-Deoxy-2-Fluoro-D-Glucose
FDM	2-Deoxy-2-Fluoro-D-Mannose
FDP	Flexor Digitorum Profundus
FID	Free-induction Decay
FLASH	Fast Low-angle Shot
GI	Gastrointestinal
HFB	Hexafluorobenzene
HSI	Hadamard Spectroscopic Imaging
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopic
MRSI	Magnetic Resonance Spectroscopic Imaging
NDP	Nucleoside Diphosphate
NLLS	Nonlinear Least-squares Fitting Routine
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PCr	Phosphocreatine
PDE	Phosphodiester
PFC	Perfluorochemical
Pi	Inorganic Phosphate
PME	Phosphomonoester
PRESS	Point-resolved Echo Spectroscopy
RARE	Rapid Acquisition with Refocused Echoes
rf	Radiofrequency
RINEPT	Refocused Insensitive Nucleus Enhancement by Polarization Transfer
SI	Spectroscopy Imaging
STEAM	Stimulated Echo Acquisition Mode
TR	Repetition Time of the Magnetic Resonance Pulse Sequence
T1	Spin–Lattice Relaxation Time
T2	Spin–Spin Relaxation Time
2D/^{31}P/CSI	Two-dimensional Phosphorus Chemical Shift Imaging
2FDG-6P	2-Fluoro-2-Deoxyglucose-6-Phosphate
3D	Three-dimensional
3DFT	Three-dimensional Fourier Transform
3DPR	Three-dimensional Projection Reconstruction
3DSI	Three-dimensional Spectroscopic Imaging
5-FCyt	5-Fluorocytosine
5-FU	5-Fluorouracil

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Magnetic Resonance Angiography • Magnetic Resonance Imaging, Functional • Magnetic Resonance in Medicine, High Resolution Ex Vivo • Magnetic Resonance, General Medical

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction

REFERENCES

1. D.G. Gadian, *Nuclear Magnetic Resonance and its Application to Living Systems*, Oxford University Press, New York, 1982.
2. T.R. Brown, B.M. Kincaid, K. Ugurbil, 'NMR Chemical Shift Imaging in Three Dimensions', *Proc. Natl. Sci. USA*, **79**, 3523–3526 (1982).
3. P. Brunner, R.R. Ernst, 'Sensitivity and Performance Time in NMR Imaging', *J. Magn. Reson.*, **33**, 83–106 (1979).
4. I.L. Pykett, B.R. Rosen, 'Nuclear Magnetic Resonance: In Vivo Proton Chemical Shift Imaging. Work in Progress', *Radiology*, **149**(1), 197–201 (1983).
5. P.C. Lauterbur, D.K. Kramer, W.V. House, 'Zeugmatographic High Resolution Nuclear Magnetic Resonance Spectroscopy: Images of Chemical Inhomogeneity with Macroscopic Objects', *J. Am. Chem. Soc.*, **97**, 6866–6868 (1975).

6. A.A. Maudsley, A. Oppelt, A. Ganssen, 'Rapid Measurement of Magnetic Field Distributions Using Nuclear Magnetic Resonance', *Siemens Forsch. Entwickl.-Ber.*, **8**, 326–331 (1979).
7. A.A. Maudsley, S.K. Hilal, W.H. Perman, H.E. Simon, 'Spatially Resolved High Resolution Spectroscopy by Four Dimensional NMR', *J. Magn. Reson.*, **51**, 147–152 (1983).
8. S.J. Cox, P. Styles, 'Towards Biochemical Imaging', *J. Magn. Reson.*, **40**, 209–212 (1980).
9. L.D. Hall, V. Rajanayagam, S. Sukumar, 'Chemical-shift Resolved Tomography Using Four-Dimensional FT Imaging', *J. Magn. Reson.*, **61**, 188–191 (1985).
10. J.C. Haselgrove, V.H. Subramanian, J.S. Leigh, Jr, L. Gyulai, B. Chance, 'In Vivo One-dimensional Imaging of Phosphorus Metabolites by Phosphorus-31 Nuclear Magnetic Resonance', *Science*, **220**(4602), 1170–1173 (1983).
11. P.A. Bottomley, C.J. Hardy, P.B. Roemer, 'Phosphate Metabolite Imaging and Concentration Measurements in Human Heart by Nuclear Magnetic Resonance', *Magn. Reson. Med.*, **14**(3), 425–434 (1990).
12. B.J. Murphy, H. Jiang, R. Stoyanova, T.R. Brown, 'Quantification of Phosphorus Metabolites from Chemical Shift Imaging Spectra with Corrections for Point Spread Effects and B1 Inhomogeneity', *Magn. Reson. Med.*, **39**(3), 429–438 (1998).
13. M. Meininger, W. Landschutz, M. Beer, T. Seyfarth, M. Horn, T. Pabst, A. Haase, D. Hahn, S. Neubauer, M. von Kienlin, 'Concentrations of Human Cardiac Phosphorus Metabolites Determined by SLOOP P31 NMR Spectroscopy', *Magn. Res. Med.*, **41**(4), 657–663 (1999).
14. O. Gonen, J. Hu, B.J. Murphy, R. Stoyanova, T.R. Brown, 'Dual Interleaved ^{1}H and Proton-decoupled-^{31}P In Vivo Chemical Shift Imaging of Human Brain', *Magn. Reson. Med.*, **32**(1), 104–109 (1994).
15. O. Gonen, B.J. Murphy, R. Srinivasan, J. Hu, H. Jiang, R. Stoyanova, T.R. Brown, 'Simultaneous and Interleaved Multinuclear Chemical-shift Imaging, a Method for Concurrent, Localized Spectroscopy', *J. Magn. Reson. B*, **104**(1), 26–33 (1994).
16. D.B. Vigneron, S.J. Nelson, B.J. Murphy, D.A. Kelley, H.B. Kessler, T.R. Brown, J.S. Taylor, 'Chemical Shift Imaging of Human Brain: Axial, Sagittal, and Coronal P-31 Metabolite Images', *Radiology*, **177**(3), 643–649 (1990).
17. M.J. Kuschmerick, T. Richards, P. Proskauer, 'Distribution of Muscle Fiber in Human Lower Leg by P31 Spectroscopy', in *Proceedings from the Society of Magnetic Resonance in Medicine Annual Meeting, August 18–24, New York*, 890, 1990, Society of Magnetic Resonance in Medicine, Berkeley, CA.
18. P.A. Bottomley, 'Spatial Localization in NMR Spectroscopy In Vivo', *Ann. N.Y. Acad. Sci.*, **508**, 333–348 (1987).
19. J. Frahm, K.D. Merboldt, W. Hanicke, 'Localized Proton Spectroscopy Using Stimulated Echoes', *J. Magn. Res.*, **72**, 502–508 (1987).
20. P.A. Bottomley, T.H. Foster, W.M. Leue, 'In Vivo Nuclear Magnetic Resonance Chemical Shift Imaging by Selective Irradiation', *Proc. Natl. Acad. Sci. USA*, **81**(21), 6856–6860 (1984).
21. J. Frahm, A. Haase, W. Hanicke, D. Matthaei, H. Bomsdorf, T. Helzel, 'Chemical Shift Selective MR Imaging Using a Whole-body Magnet', *Radiology*, **156**(2), 441–444 (1985).
22. A. Haase, J. Frahm, W. Hanicke, D. Matthaei, '^{1}H NMR Chemical Shift Selective (CHESS) Imaging', *Phys. Med. Biol.*, **30**(4), 341–344 (1985).
23. D. Matthaei, J. Frahm, A. Haase, R. Schuster, H. Bomsdorf, 'Chemical-shift-selective Magnetic-resonance Imaging of Avascular Necrosis of the Femoral Head', *Lancet*, **1**(8425), 370–371 (1985).
24. D. Matthaei, A. Haase, J. Frahm, H. Bomsdorf, W. Vollmann, 'Multiple Chemical Shift Selective (CHESS) MR Imaging Using Stimulated Echoes', *Radiology*, **160**(3), 791–794 (1986).
25. T. Ernst, J.H. Lee, B.D. Ross, 'Direct ^{31}P Imaging in Human Limb and Brain', *J. Comput. Assist. Tomogr.*, **17**(5), 673–680 (1993).
26. G.H. Glover, J.M. Pauly, K.M. Bradshaw, 'Boron-11 Imaging with a Three-dimensional Reconstruction Method', *J. Magn. Res. Imag.*, **2**, 47–52 (1992).
27. K.M. Bradshaw, M.P. Schweizer, G.H. Glover, J.R. Hadley, R. Tippets, P.P. Tang, W.L. Davis, M.P. Heilbrun, S. Johnson, T. Ghanem, 'BSH Distributions in the Canine Head and a Human Patient Using ^{11}B MRI' (published erratum appears in *Magn. Reson. Med.*, **34**(4), 645 (1995)), *Magn. Reson. Med.*, **34**(1), 48–56 (1995).
28. P.C. Lauterbur, 'Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance', *Nature*, **242**, 190–191 (1973).
29. S. Posse, C. De Carli, B.D. Le, 'Three-dimensional Echo-planar MR Spectroscopic Imaging at Short Echo Times in the Human Brain', *Radiology*, **192**(3), 733–738 (1994).
30. S. Posse, G. Tedeschi, R. Risinger, R. Ogg, B.D. Le, 'High Speed ^{1}H Spectroscopic Imaging in Human Brain by Echo Planar Spatial–Spectral Encoding', *Magn. Reson. Med.*, **33**(1), 34–40 (1995).
31. S. Posse, S.R. Dager, T.L. Richards, C. Yuan, R. Ogg, A.A. Artru, G.H. Muller, C. Hayes, 'In Vivo Measurement of Regional Brain Metabolic Response to Hyperventilation Using Magnetic Resonance: Proton Echo Planar Spectroscopic Imaging (PEPSI)', *Magn. Reson. Med.*, **37**(6), 858–865 (1997).
32. B.R. Barker, R.P. Mason, N. Bansal, R.M. Peshock, 'Oxygen Tension Mapping with ^{19}F Echo-planar MR Imaging of Sequestered Perfluorocarbon', *J. Magn. Reson. Imaging*, **4**(4), 595–602 (1994).

33. S.J. Nelson, D.B. Vigneron, L.J. Star, J. Kurhanewicz, 'High Spatial Resolution and Speed in MRSI', *NMR Biomed.*, **10**(8), 411–422 (1997).
34. P. Mansfield, 'Spatial Mapping of Chemical Shift in NMR', *Magn. Reson. Med.*, **1**, 370–386 (1984).
35. Y. Bito, S. Hirata, T. Nabeshima, E. Yamamoto, 'Echo-planar Diffusion Spectroscopic Imaging', *Magn. Reson. Med.*, **33**(1), 69–73 (1995).
36. B.J. Dardzinski, C.H. Sotak, 'Rapid Tissue Oxygen Tension Mapping Using ^{19}F Inversion-recovery Echo-planar Imaging of Perfluoro-15-crown-5-ether', *Magn. Reson. Med.*, **32**(1), 88–97 (1994).
37. S. Hirata, Y. Bito, E. Yamamoto, 'Expansion of the Spectral Bandwidth by Spatial and Chemical Shift Selective Saturation in High-speed Magnetic Resonance Spectroscopic Imaging', *Magn. Reson. Med.*, **35**(4), 611–616 (1996).
38. G. Metzger, X. Hu, 'Application of Interlaced Fourier Transform to Echo-planar Spectroscopic Imaging', *J. Magn. Reson.*, **125**(1), 166–170 (1997).
39. R.V. Mulkern, P.S. Melki, H.S. Lilly, F.A. Hoffer, '1D Spectroscopic Imaging with rf Echo Planar (SIRFEN) Methods', *Magn. Reson. Imaging*, **9**(6), 909–916 (1991).
40. A.A. Maudsley, Z. Wu, D.J. Meyerhoff, M.W. Weiner, 'Automated Processing for Proton Spectroscopic Imaging Using Water Reference Deconvolution', *Magn. Reson. Med.*, **31**(6), 589–595 (1994).
41. I. Dologlou, H.S. Van, O.D. Van, 'Frequency-selective MRS Data Quantification with Frequency Prior Knowledge', *J. Magn. Reson.*, **130**(2), 238–243 (1998).
42. S.J. Nelson, T.R. Brown, 'The Accuracy of Quantification from 1D NMR Spectra Using the PIQABLE Algorithm', *J. Magn. Reson.*, **84**, 95–109 (1989).
43. P.B. Barker, E.J. Butterworth, M.D. Boska, J. Nelson, K.M. Welch, 'Magnesium and pH Imaging of the Human Brain at 3.0 Tesla', *Magn. Reson. Med.*, **41**(2), 400–406 (1999).
44. A. van den Boogaart, F.A. Howe, L.M. Rodrigues, M. Stubbs, J.R. Griffiths, 'In Vivo ^{31}P MRS: Absolute Concentrations, Signal-to-noise and Prior Knowledge', *NMR Biomed.*, **8**(2), 87–93 (1995).
45. T.R. Brown, R. Stoyanova, 'NMR Spectral Quantitation by Principal-component Analysis. II. Determination of Frequency and Phase Shifts', *J. Magn. Reson. B*, **112**(1), 32–43 (1996).
46. K. Derby, C. Hawryszko, J. Tropp, 'Baseline Deconvolution, Phase Correction, and Signal Quantification in Fourier Localized Spectroscopic Imaging', *Magn. Reson. Med.*, **12**(2), 235–240 (1989).
47. F. Podo, O. Henriksen, W.M. Bovee, M.O. Leach, D. Leibfritz, J.D. de Certaines, 'Absolute Metabolite Quantification by In Vivo NMR Spectroscopy: I. Introduction, Objectives and Activities of a Concerted Action in Biomedical Research', *Magn. Reson. Imaging*, **16**(9), 1085–1092 (1998).
48. H. Serrai, L. Senhadji, J.D. de Certaines, J.L. Coatrieux, 'Time-domain Quantification of Amplitude, Chemical shift, Apparent Relaxation Time T2, and Phase by Wavelet-transform Analysis. Application to Biomedical Magnetic Resonance Spectroscopy', *J. Magn. Reson.*, **124**(1), 20–34 (1997).
49. L. Van Hamme, A. van den Boogaart, S. Van Huffel, 'Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge', *J. Magn. Reson.*, **129**(1), 35–43 (1997).
50. T. Yabe, K. Mitsunami, T. Inubushi, M. Kinoshita, 'Quantitative Measurements of Cardiac Phosphorus Metabolites in Coronary Artery Disease by ^{31}P Magnetic Resonance Spectroscopy [see comments]', *Circulation*, **92**(1), 15–23 (1995).
51. A.A. Maudsley, S.K. Hilal, H.E. Simon, S. Wittekoek, 'In Vivo MR Spectroscopic Imaging with P-31. Work in Progress', *Radiology*, **153**(3), 745–750 (1984).
52. A.A. Maudsley, D.B. Twieg, M.D. Sappey, B. Hubesch, J.W. Hugg, G.B. Matson, M.W. Weiner, 'Spin Echo ^{31}P Spectroscopic Imaging in the Human Brain', *Magn. Reson. Med.*, **14**(2), 415–422 (1990).
53. M. Sasahira, K. Uchimura, K. Yatsushiro, T. Fujimoto, T. Asakura, 'Early Detection of Cerebral Infarction by ^{31}P Spectroscopic Imaging', *Neuroradiology*, **32**(1), 43–46 (1990).
54. B. Hubesch, D.S. Marinier, H.P. Hetherington, D.B. Twieg, M.W. Weiner, 'Clinical MRS Studies of the Brain', *Invest. Radiol.*, **24**(12), 1039–1042 (1989).
55. M.F. Arias, T. Javaid, R. Stoyanova, T.R. Brown, O. Gonen, 'Heteronuclear Multivoxel Spectroscopy of In Vivo Human Brain: Two-dimensional Proton Interleaved with Three-dimensional ^{1}H-decoupled Phosphorus Chemical Shift Imaging', *NMR Biomed.*, **9**(3), 105–113 (1996).
56. L.H. Chang, Y. Cohen, P.R. Weinstein, L. Chileuitt, T.L. James, 'Interleaved ^{1}H and ^{31}P Spectroscopic Imaging for Studying Regional Brain Injury', *Magn. Reson. Imaging*, **9**(2), 223–227 (1991).
57. O. Gonen, J. Hu, R. Stoyanova, J.S. Leigh, G. Goelman, T.R. Brown, 'Hybrid Three Dimensional (1D-Hadamard, 2D-chemical Shift Imaging) Phosphorus Localized Spectroscopy of Phantom and Human Brain', *Magn. Reson. Med.*, **33**(3), 300–308 (1995).
58. O. Gonen, A. Mohebbi, R. Stoyanova, T.R. Brown, 'In Vivo Phosphorus Polarization Transfer and Decoupling from Protons in Three-dimensional Localized Nuclear Magnetic Resonance Spectroscopy of Human Brain', *Magn. Reson. Med.*, **37**(2), 301–306 (1997).
59. W.I. Jung, A. Staubert, S. Widmaier, T. Hoess, M. Bunse, E.F. van, G. Dietze, O. Lutz, 'Phosphorus J-coupling Constants of ATP in Human Brain', *Magn. Reson. Med.*, **37**(5), 802–804 (1997).
60. J.M. Constans, D.J. Meyerhoff, D. Norman, G. Fein, M.W. Weiner, '^{1}H and ^{31}P Magnetic Resonance Spectroscopic Imaging of White Matter Signal Hyperintensity

areas in Elderly Subjects', *Neuroradiology*, **37**(8), 615–623 (1995).

61. A.J. Richardson, I.J. Cox, J. Sargentoni, B.K. Puri, 'Abnormal Cerebral Phospholipid Metabolism in Dyslexia Indicated by Phosphorus-31 Magnetic Resonance Spectroscopy', *NMR Biomed.*, **10**(7), 309–314 (1997).
62. R.F. Deicken, G. Calabrese, E.L. Merrin, S. Vinogradov, G. Fein, M.W. Weiner, 'Asymmetry of Temporal Lobe Phosphorous Metabolism in Schizophrenia: a 31 Phosphorous Magnetic Resonance Spectroscopic Imaging Study', *Biol. Psychiatry*, **38**(5), 279–286 (1995).
63. C.A. Husted, G.B. Matson, D.A. Adams, D.S. Goodin, M.W. Weiner, 'In Vivo Detection of Myelin Phospholipids in Multiple Sclerosis with Phosphorus Magnetic Resonance Spectroscopic Imaging', *Ann. Neurol.*, **36**(2), 239–241 (1994).
64. W.J. Chu, H.P. Hetherington, R.I. Kuzniecky, T. Simor, G.F. Mason, G.A. Elgavish, 'Lateralization of Human Temporal Lobe Epilepsy by ^{31}P NMR Spectroscopic Imaging at 4.1 T', *Neurology*, **51**(2), 472–479 (1998).
65. J. Van der Grond, J.R. Gerson, K.D. Laxer, J.W. Hugg, G.B. Matson, M.W. Weiner, 'Regional Distribution of Interictal ^{31}P Metabolic Changes in Patients with Temporal Lobe Epilepsy', *Epilepsia*, **39**(5), 527–536 (1998).
66. J.S. Duncan, 'Magnetic Resonance Spectroscopy', *Epilepsia*, **37**(7), 598–605 (1996).
67. P.A. Garcia, K.D. Laxer, J. Van der Grond, J.W. Hugg, G.B. Matson, M.W. Weiner, 'Phosphorus Magnetic Resonance Spectroscopic Imaging in Patients with Frontal Lobe Epilepsy', *Ann. Neurol.*, **35**(2), 217–221 (1994).
68. J.W. Hugg, K.D. Laxer, G.B. Matson, A.A. Maudsley, C.A. Husted, M.W. Weiner, 'Lateralization of Human Focal Epilepsy by ^{31}P Magnetic Resonance Spectroscopic Imaging', *Neurology*, **42**(10), 2011–2018 (1992).
69. J.W. Hugg, G.B. Matson, D.B. Twieg, A.A. Maudsley, M.D. Sappey, M.W. Weiner, 'Phosphorus-31 MR Spectroscopic Imaging (MRSI) of Normal and Pathological Human Brains', *Magn. Reson. Imaging*, **10**(2), 227–243 (1992).
70. D. Holtzman, R. Mulkern, R. Meyers, C. Cook, E. Allred, I. Khait, F. Jensen, M. Tsuji, P. Laussen, 'In Vivo Phosphocreatine and ATP in Piglet Cerebral Gray and White Matter During Seizures', *Brain Res.*, **783**(1), 19–27 (1998).
71. S. Schaefer, J.R. Gober, G.G. Schwartz, D.B. Twieg, M.W. Weiner, B. Massie, 'In Vivo Phosphorus-31 Spectroscopic Imaging in Patients with Global Myocardial Disease', *Am. J. Cardiol*, **65**(16), 1154–1161 (1990).
72. P.M. Robitaille, H. Merkle, E. Sublett, K. Hendrich, B. Lew, G. Path, A.H. From, R.J. Bache, M. Garwood, K. Ugurbil, 'Spectroscopic Imaging and Spatial Localization Using Adiabatic Pulses and Applications to Detect Transmural Metabolite Distribution in the Canine Heart', *Magn. Reson. Med.*, **10**(1), 14–37 (1989).
73. W. Auffermann, W.M. Chew, C.L. Wolfe, N.J. Tavares, W.W. Parmley, R.C. Semelka, T. Donnelly, K. Chatterjee, C.B. Higgins, 'Normal and Diffusely Abnormal Myocardium in Humans: Functional and Metabolic Characterization with P-31 MR Spectroscopy and Cine MR Imaging', *Radiology*, **179**(1), 253–259 (1991).
74. H.P. Hetherington, D.J. Luney, J.T. Vaughan, J.W. Pan, S.L. Ponder, O. Tschendel, D.B. Twieg, G.M. Pohost, '3D ^{31}P Spectroscopic Imaging of the Human Heart at 4.1 T', *Magn. Reson. Med.*, **33**(3), 427–431 (1995).
75. C.M. Campbell, G. Wisenberg, J. Sykes, R.T. Thompson, 'Controlled Reperfusion after Myocardial Ischemia in a Canine Model Monitored by Two-dimensional Phosphorus 31 Chemical Shift Spectroscopic Imaging', *Am. Heart. J.*, **133**(5), 508–516 (1997).
76. C.M. Campbell, G. Wisenberg, J. Sykes, R.T. Thompson, 'Noninvasive Assessment of Pharmaceutical Intervention During Myocardial Ischemia-reperfusion in a Canine Model Using Two-dimensional ^{31}P Chemical Shift Imaging', *Biochem. Cell Biol.*, **76**(2–3), 522–531 (1998).
77. J.A. Jeneson, S.J. Nelson, D.B. Vigneron, J.S. Taylor, B.J. Murphy, T.R. Brown, 'Two-dimensional ^{31}P-chemical Shift Imaging of Intramuscular Heterogeneity in Exercising Human Forearm Muscle', *Am. J. Physiol.*, C357–C364 (1992).
78. P.C. van Zijl, A.S. Chesnick, D. Des Pres, C.T. Moonen, J. Ruiz Cabello, P. van Gelderen, 'In Vivo Proton Spectroscopy and Spectroscopic Imaging of [1–^{13}C]-Glucose and its Metabolic Products', *Magn. Reson. Med.*, **30**(5), 544–551 (1993).
79. S.M. Lee, D.L. Tzou, G. Chiou, H.N. Yeung, 'An Alternative Scheme for ^{13}C Chemical-Shift Imaging via Inverse Detection of Protons through an MILS (or Inverse SLIM) Technique', *J. Magn. Reson.*, **130**(1), 102–110 (1998).
80. B.E. Hammer, 'Proton Decoupled ^{13}C NMR Imaging', *Magn. Reson. Imaging*, **7**(2), 235–240 (1989).
81. S. Morishita, M. Sumi, R. Nishimura, M. Takahashi, N. Iriguchi, 'Carbon-13 Chemical Shift Imaging of [1–^{13}C] Glucose Under Metabolism in the Rat Head In Vivo', *Radiat. Med.*, **10**(3), 94–100 (1992).
82. N. Beckmann, I. Turkalj, J. Seelig, U. Keller, '^{13}C-NMR for the Assessment of Human Brain Glucose Metabolism In Vivo', *Biochemistry*, **30**(26), 6362–6366 (1991).
83. R. Schwarz, M. Schuurmans, J. Seelig, B. Kunnecke, '^{19}F-MRI of Perfluorononane as a Novel Contrast Modality for Gastrointestinal Imaging', *Magn. Reson. Med.*, **41**(1), 80–86 (1999).
84. T.Q. Duong, J.J. Ackerman, H.S. Ying, J.J. Neil, 'Evaluation of Extra- and Intracellular Apparent Diffusion in

Normal and Globally Ischemic Rat Brain via ^{19}F-NMR', *Magn. Reson. Med.*, **40**(1), 1–13 (1998).
85. P.N. Venkatasubramanian, Y.J. Shen, A.M. Wyrwicz, 'Characterization of the Cerebral Distribution of General Anesthetics In Vivo by Two-dimensional ^{19}F Chemical Shift Imaging', *Magn. Reson. Med.*, **35**(4), 626–630 (1996).
86. Y. Xu, P. Tang, W. Zhang, L. Firestone, P.M. Winter, 'Fluorine-19 Nuclear Magnetic Resonance Imaging and Spectroscopy of Sevoflurane Uptake, Distribution, and Elimination in Rat Brain [see comments]', *Anesthesiology*, **83**(4), 766–774 (1995).
87. J. Pekar, L. Ligeti, T. Sinnwell, C.T. Moonen, J.A. Frank, A.C. McLaughlin, '^{19}F Magnetic Resonance Imaging of Cerebral Blood Flow with 0.4-cc Resolution', *J. Cereb. Blood Flow Metab.*, **14**(4), 656–663 (1994).
88. H. Lee, R.R. Price, G.E. Holburn, C.L. Partain, M.D. Adams, W.P. Cacheris, 'In Vivo Fluorine-19 MR Imaging: Relaxation Enhancement with Gd-DTPA', *J. Magn. Reson. Imaging*, **4**(4), 609–613 (1994).
89. H.T. Tran, Q. Guo, D.J. Schumacher, R.B. Buxton, R.F. Mattrey, '^{19}F Chemical Shift Imaging Technique to Measure Intracellular pO_2 In Vivo Using Perflubron', *Acad Radiol.*, **2**(9), 756–761 (1995).
90. R.G. Pratt, J. Zheng, B.K. Stewart, Y. Shiferaw, A.J. McGoron, R.C. Samaratunga, S.R. Thomas, 'Application of a 3D Volume ^{19}F-MR Imaging Protocol for Mapping Oxygen Tension (pO_2) in Perfluorocarbons at Low Field', *Magn. Reson. Med.*, **37**(2), 307–313 (1997).
91. R.P. Mason, S. Hunjan, D. Le, A. Constantinescu, B.R. Barker, P.S. Wong, P. Peschke, E.W. Hahn, P.P. Antich, 'Regional Tumor Oxygen Tension: Fluorine Echo Planar Imaging of Hexafluorobenzene Reveals Heterogeneity of Dynamics', *Int. J. Radiat. Oncol. Biol. Phys.*, **42**(4), 747–750 (1998).
92. G. Brix, M.E. Bellemann, L. Gerlach, U. Haberkorn, 'Intra- and Extracellular Fluorouracil Uptake: Assessment with Contrast-enhanced Metabolic F-19 MR Imaging', *Radiology*, **209**(1), 259–267 (1998).
93. E.O. Aboagye, D. Artemov, P.D. Senter, Z.M. Bhujwalla, 'Intratumoral Conversion of 5-Fluorocytosine to 5-Fluorouracil by Monoclonal Antibody–Cytosine Deaminase Conjugates: Noninvasive Detection of Prodrug Activation by Magnetic Resonance Spectroscopy and Spectroscopic Imaging', *Cancer Res.*, **58**(18), 4075–4078 (1998).
94. N.J. Baldwin, T.C. Ng, 'Oxygenation and Metabolic Status of KHT Tumors as Measured Simultaneously by ^{19}F Magnetic Resonance Imaging and ^{31}P Magnetic Resonance Spectroscopy', *Magn. Reson. Imaging*, **14**(5), 541–551 (1996).
95. J. Lutz, U. Noth, S.P. Morrissey, H. Adolf, R. Deichmann, A. Haase, 'In Vivo Measurement of Oxygen Pressure Using ^{19}F-NMR Imaging', *Adv. Exp. Med. Biol.*, **388**, 53–57 (1996).
96. U. Noth, S.P. Morrissey, R. Deichmann, H. Adolf, C. Schwarzbauer, J. Lutz, A. Haase, 'In Vivo Measurement of Partial Oxygen Pressure in Large Vessels and in the Reticuloendothelial System Using Fast ^{19}F-MRI', *Magn. Reson. Med.*, **34**(5), 738–745 (1995).
97. O. Gonen, B.J. Murphy, C.W. Li, S.K. Padavic, W.G. Negendank, T.R. Brown, 'Simultaneous 3D NMR Spectroscopy of Proton-decoupled Fluorine and Phosphorus in Human Liver during 5-Fluorouracil Chemotherapy', *Magn. Reson. Med.*, **37**(2), 164–169 (1997).
98. Y. Kanazawa, K. Umayahara, T. Shimmura, T. Yamashita, '^{19}F-NMR of 2-Deoxy-2-fluoro-D-glucose for Tumor Diagnosis in Mice. An NDP-bound Hexose Analog as a new NMR Target for Imaging', *NMR Biomed.*, **10**(1), 35–41 (1997).
99. S.K. Hilal, A.A. Maudsley, H.E. Simon, W.H. Perman, J. Bonn, M.E. Mawad, A.J. Silver, S.R. Ganti, P. Sane, I.C. Chien, 'In Vivo NMR Imaging of Tissue Sodium in the Intact Cat Before and After Acute Cerebral Stroke', *Ajnr. Am. J. Neuroradiol.*, **4**(3), 245–249 (1983).
100. J.B. Ra, S.K. Hilal, Z.H. Cho, 'A Method for In Vivo MR Imaging of the Short T2 Component of Sodium-23', *Magn. Reson. Med.*, **3**(2), 296–302 (1986).
101. W. Grodd, U. Klose, 'Sodium-MR-imaging of the Brain: Initial Clinical Results', *Neuroradiology*, **30**(5), 399–407 (1988).
102. T. Hashimoto, H. Ikehira, H. Fukuda, A. Yamaura, O. Watanabe, Y. Tateno, R. Tanaka, H.E. Simon, 'In Vivo Sodium-23 MRI in Brain Tumors: Evaluation of Preliminary Clinical Experience', *Am. J. Physiol. Imaging*, **6**(2), 74–80 (1991).
103. H. Kauczor, R. Surkau, T. Roberts, 'MRI Using Hyperpolarized Noble Gases', *Eur. Radiol.*, **8**(5), 820–827 (1998).
104. H.U. Kauczor, K. Markstaller, 'Magnetic Resonance Tomography with Inhalation of Polarized Noble Gases: new Perspectives in Functional Imaging Diagnosis of Emphysema', *Pneumologie*, **52**(10), 563–567 (1998).
105. P. Bendel, A. Frantz, J. Zilberstein, G.W. Kabalka, Y. Salomon, 'Boron-11 NMR of Borocaptate: Relaxation and In Vivo Detection in Melanoma-bearing Mice', *Magn. Reson. Med.*, **39**(3), 439–447 (1998).
106. P.R. Gavin, S.L. Kraft, R. Huiskamp, J.A. Coderre, 'A Review: CNS Effects and Normal Tissue Tolerance in Dogs', *J. Neurooncol.*, **33**(1–2), 71–80 (1997).
107. G.W. Kabalka, M. Davis, P. Bendel, 'Boron-11 MRI and MRS of Intact Animals Infused with a Boron Neutron Capture Agent', *Magn. Reson. Med.*, **8**(2), 231–237 (1988).
108. G.W. Kabalka, P. Bendel, M. Davis, D.N. Slatkin, P.L. Micca, 'Boron-11 Magnetic Resonance Imaging and Spectroscopy; Tools for Investigating Pharmacokinetics for Boron Neutron Capture Therapy', *Basic Life Sci.*, **50**, 243–249 (1989).
109. G.W. Kabalka, G.Q. Cheng, P. Bendel, P.L. Micca, D.N. Slatkin, 'In Vivo Boron-11 MRI and MRS Using

(B24H22S2)4- in the Rat', *Magn. Reson. Imaging*, **9**(6), 969–973 (1991).

110. G.W. Kabalka, C. Tang, P. Bendel, 'The Role of Boron MRI in Boron Neutron Capture Therapy', *J. Neuro-oncol*, **33**(1–2), 153–161 (1997).

111. T.L. Richards, K.M. Bradshaw, D.M. Freeman, C.H. Sotak, P.R. Gavin, 'Imaging with ^{11}B of Intact Tissues Using Magnetic Resonance Gradient Echoes', *Strahlenther Onkol.*, **165**(2–3), 179–181 (1989).

Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Michael G. Sowa, Lorenzo Leonardi, Anna Matas, Bernie J. Schattka, Mark D. Hewko, Jeri R. Payette, and Henry H. Mantsch
Institute for Biodiagnostics, Winnipeg, Canada

1 Introduction **1**
2 Optical Properties of Living Tissues **2**
2.1 Background 2
2.2 Tissue Absorption 3
2.3 Tissue Scattering 5
3 Instrumental Methods **7**
3.1 Broadband Spectroscopy 7
3.2 Discrete Wavelengths 9
3.3 Spectroscopic Imaging 10
4 Practical Considerations **11**
4.1 Pulsatile Modulation 11
4.2 Probe–Tissue Contact Pressure 13
4.3 Tissue Temperature 14
5 Selected Examples **16**
5.1 Forearm Ischemia–Reperfusion 16
5.2 Myocardial Ischemia–Reperfusion 17
5.3 Cerebral Oxygenation and Blood Flow 21
5.4 Near-infrared Assessment of Tissue Viability 22
6 Summary and Outlook **25**
Abbreviations and Acronyms **25**
Related Articles **26**
References **26**

In vivo near-infrared (NIR) spectroscopy has the potential of becoming an important tool in a number of areas in clinical medicine. Technological developments in photonics that have been spurred on by the communication revolution have set the stage for rapid advancement of optical and NIR spectroscopy based on noninvasive or minimally invasive medical diagnostic techniques. The goal of this article is to review the current capabilities and limitations of in vivo NIR spectroscopy and highlight the impact of these capabilities and limitations in selected areas where NIR spectroscopy is being used to address clinical problems. The optical properties of tissues are briefly reviewed, as are the instrumental methods available to the experimentalist. These properties and methods largely dictate the feasibility of an in vivo spectroscopic diagnostic approach and constrain the scope of problems that can be tackled using optical–NIR spectroscopy. Some of the more successful applications are described, including studies of tissue oxygenation, ischemia, and viability. A number of factors that can confound interpretation of in vivo NIR results are discussed. The number and magnitude of confounding influences that arise in in vivo spectroscopy can be daunting to the experimentalist and may represent the largest barrier in transforming in vivo spectroscopic measurements into clinically meaningful and reliable information. In vivo NIR spectroscopy abounds with opportunity and challenge.

1 INTRODUCTION

In vivo NIR spectroscopy involves analyzing the wavelength dependence of the attenuation of NIR light by living tissue, in situ. This is accomplished by measuring the light between the red end of the visible (600 nm) to the long wavelength NIR (2500 nm) of the electromagnetic spectrum that is diffusely reflected by the tissue and/or transmitted through the tissue. The attenuation of light by tissue is governed by both the light absorption properties and scattering characteristics of the tissue. Scattering is largely dictated by the structural properties of tissue such as boundaries and interfaces, cell count, particle size and shape. The light absorption characteristics of tissue are determined by the types and concentrations of constituent chromophore biomolecules. It is perhaps the complexity of the dynamic matrix of living tissue and the strong interplay between absorption and scatter-related contributions to the attenuation of NIR light by tissue that are the distinguishing elements of in vivo NIR spectroscopy. Despite the complexity, in vivo spectroscopy shares common goals with optical and infrared (IR) spectroscopic techniques applied to better defined and more ordered systems. The aims of most spectroscopic studies on tissues are to identify chromophores present in the sample, perform quantitative and/or semi-quantitative analysis on the chromophores present, and correlate changes in the bulk optical properties of the tissue to morphological or structural changes in the tissue. While this information is, in principal, latent in the NIR attenuation spectrum of tissue, the extraction of this information and its translation into medically relevant findings presents a major challenge in the analysis of in vivo NIR spectra.

The prospects for in vivo NIR spectroscopy are appealing and the stage is set for rapid advancement

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of noninvasive or minimally invasive optical and NIR spectroscopy based solutions to medical diagnostic problems. This prognosis is borne out by the explosion of recent developments in this area. This article is intended to highlight the current capabilities and limitations of in vivo NIR spectroscopy, provide a brief discussion of the more active application areas of the technology as well as a selected view of the recent literature. Finally, we will speculate on future developments and applications.

2 OPTICAL PROPERTIES OF LIVING TISSUES

2.1 Background

The continuous measurement of metabolic, physiological, or structural changes in tissue is of primary concern in many clinical and biomedical domains. Biochemical or physiological changes in the body are usually accompanied by medical problems. These changes lead to microscopic structural and chemical alterations that can influence morphology and function, creating symptoms that are clinically identifiable. The challenge for medical diagnostic techniques is to detect tissue changes at an earlier stage so that therapeutic interventions may have the greatest impact on the disease or problem.

Classical methods of in vivo assessment or diagnosis are either invasive, requiring a sample for histological examination[1] or the insertion of an electrochemical probe,[2] or they use ionizing radiation.[3] The increasing reliance on magnetic resonance imaging methods with the subsequent displacement of some traditional nuclear medicine techniques and X-ray based methods indicate the recent move toward methods that are less invasive and subject the patient to lower energy or lower radiation doses. An alternative approach to these techniques uses NIR light to probe tissue in a noninvasive manner. The NIR region of the electromagnetic spectrum is composed of wavelengths ranging from 700 to 2500 nm and has previously been used to measure hemoglobin (Hb),[4–6] myoglobin,[7,8] cytochrome aa_3,[9–11] and water[12–14] in tissues. The NIR spectra of some of these chromophores are depicted in Figure 1. The spectral region between 600 and 1300 nm has special significance in biomedical applications with respect to photodynamic therapy[15] and dosimetry.[16] The NIR region involves the study of high-energy vibrational transitions and low-energy electronic transitions. Many of the absorptions in this region result from overtones and combinations of fundamental mid-IR molecular stretching and bending vibrations.[17] The prominent overtone absorptions observed in the NIR

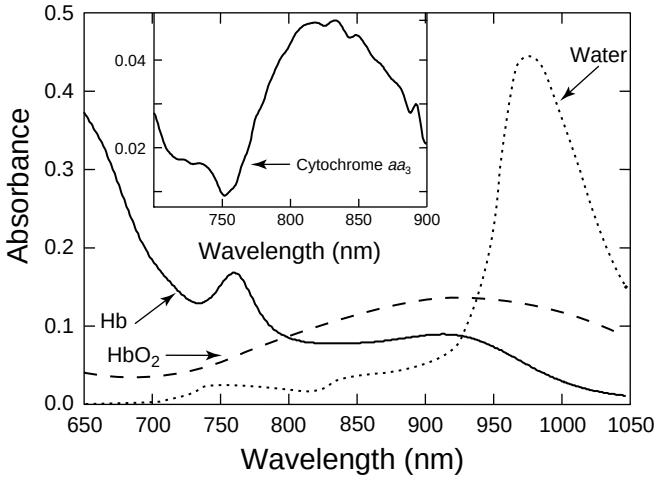

Figure 1 NIR absorbance spectra of Hb, oxyhemoglobin and water, with the NIR absorbance spectrum of cytochrome aa_3 inset.

are due to molecules that contain O–H, C–H, and N–H bonds.

NIR measurements of tissue chromophores can be made in one of two configurations: transmittance or reflectance. When NIR light of radiant power I_o strikes an absorbing compound, the emerging radiant power of the beam is decreased to I. The attenuation of light measured on the same axis as the incident light is known as transmittance. Samples following the Beer–Lambert relationship have an absorbance, A, as a function of the irradiated and transmitted light as shown in Equation (1):

$$A_\lambda = -\ln\left(\frac{I_\lambda}{I_{\lambda o}}\right) = \varepsilon_\lambda bc \qquad (1)$$

where ε_λ ($cm^{-1}\,M^{-1}$) is the wavelength-dependent extinction coefficient, b (cm) the sample thickness, and c (M) the concentration of chromophore in the medium. For a given chromophore, the extinction coefficients vary with wavelength, thus producing a wavelength-dependent spectrum. To simplify the relationship in tissue studies, the absorptivity and concentration terms are combined into a single parameter, the absorption coefficient defined as $\mu_a = \varepsilon c = A/b$ (cm^{-1}). The absorption coefficient describes the concentration of the chromophore in relation to the absorbance per unit length through the sample. Multiwavelength transmittance measurements using Equation (1) are applied to a variety of scatter-free solutions in routine clinical analysis. However, tissue chromophores are generally situated in a turbid or scattering media. An example of this is Hb in tissue.

Quantitative analyses in turbid media are difficult because of the spectral aberrations associated with multiple scattering events. When light strikes a sample containing scatterers, the light travels a distance greater

than the sample thickness as a result of the multiple scattering. Therefore, the Beer–Lambert relationship fails in a scattering medium since the actual pathlength is greater than the sample thickness. Scattering disperses the light in all directions, hence a small fraction of the scattered light that penetrates into the sample can be remitted back out to the surface. The remitted or diffusely reflected light collected by the detector has been attenuated as a result of the scattering as well as through absorption by the chromophores in the tissue. A measure of the reflected light provides spectral information on the scattering by the tissue and absorption by the tissue chromophores. Reflectance measurements rely on the sample containing scatterers to reflect the light back to the tissue surface. The type and direction of the scattering is strongly influenced by the size, shape and density of scattering particles present in the tissue. Absorption changes are easily measured provided the scattering properties of the tissue remain relatively constant throughout the observation. However, this condition is often violated when spectroscopic measurements are made on dynamic living tissue.

2.2 Tissue Absorption

Compared to the shorter ultraviolet (UV) and visible wavelength regions as well as the longer mid-IR wavelength range, the NIR wavelength range is ideally suited for in vivo tissue spectroscopy. This suitability stems from the absorption and scattering properties of tissue in the NIR. In the NIR range, the absorption by tissue chromophores is small in comparison to the mid-IR, visible and UV ranges. Although weaker tissue absorption in the NIR implies that less information is available spectroscopically, NIR light can traverse a greater distance in tissue before it is completely attenuated by chromophore absorption. This greater penetration depth of NIR light into tissue lies at the heart of its clinical utility. Whereas only the upper 10–20 μm of tissue are interrogated by mid-IR light before being completely absorbed, depths of several centimeters can be sampled in tissue using NIR light. This extended sampling depth increases the number of potential in vivo clinical applications of NIR spectroscopy relative to UV, visible and mid-IR spectroscopy. However, the weak absorption characteristics of tissue which enable deeper light penetration ultimately limit the sensitivity and type of information that can be retrieved using NIR spectroscopy.

The most clinically relevant tissue chromophores in the NIR are oxy- and deoxyhemoglobin, oxy- and deoxymyoglobin, oxidized and reduced cytochrome aa_3, and water. These chromophores provide vital tissue information related to oxygen delivery, storage, utilization, and tissue hydration.

Hb is often considered to be a respiratory pigment contained within red blood cells. One red blood cell contains approximately 265 million molecules of Hb.[18] Hb is an iron-containing protein composed of polypeptide chains. The oxygenated (HbO_2) and deoxygenated (Hb) forms have different extinction coefficients across the visible and short wavelength NIR regions, as seen in Figure 2. In the visible portion of the spectrum, HbO_2 has two peaks centered at 540 nm and 576 nm, whereas Hb has only one peak at 555 nm. Although the human body is robust in many respects, the physiological process of maintaining proper oxygen transport is a delicate and complex system. The relative oxygen saturation (O_2sat), a measure of the relative amount of oxygenated Hb to the total amount of Hb present (defined as $O_2\text{sat} = [HbO_2]/([HbO_2] + [Hb])$), provides a quantifiable measure of the oxygen transport in tissue. Using an unknown tissue spectrum and two or more of the extinction coefficients for oxy- and deoxyhemoglobin, the O_2sat for tissue can be determined.

Although the maximum extinction coefficients for HbO_2 and Hb lie in the visible portion of the spectrum, these values are not typically used for in vivo applications due to the presence of interfering chromophores in skin. For most in vivo NIR applications, spectra are obtained through the surface of the skin (transcutaneous). The major absorption in skin in the visible wavelength range results from melanin, a complex absorbing pigment found in the epidermal layer of the skin. Melanin is a highly effective absorber of light, especially in the UV region

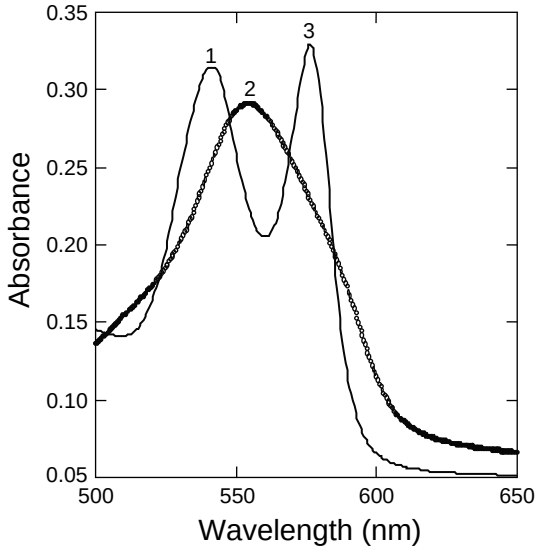

Figure 2 Visible absorbance spectra of Hb and oxyhemoglobin. The peaks labeled 1 and 3 correspond to Q_o and Q_v vibronic transitions of oxyhemoglobin while the broad peak labeled 2 is comprised of the corresponding vibronic transitions of deoxyhemoglobin.

of the spectrum. The absorption of light by melanin in the UV and visible regions directly affects the reflected and transmitted light through skin. However, in the NIR spectral region, the absorption from melanin is minimal while Hb and myoglobin are the dominant absorbers in the short wavelength NIR region (650–900 nm). Several methodologies have been used to determine the O_2sat in tissue employing the NIR wavelength region.[19]

The most common tissue chromophores found in muscle are Hb, myoglobin, cytochrome and water. Blood contains Hb, which carries oxygen to the tissue, where it is stored by myoglobin. Oxy- and deoxymyoglobin have distinct absorption spectra in the NIR.[20] To determine the oxygen saturation in muscle, i.e. the amount of oxygen present in muscle tissue in relation to the maximum storage capacity, the optical properties of myoglobin can be exploited in a similar manner to Hb by determining the myoglobin oxygen saturation. This parameter can be obtained by measuring the absorbed light at two or more different wavelengths, namely the oxy- and deoxy-absorption bands, and the oxygen saturation is calculated via a ratio of the oxy- absorption to the total oxy- and deoxy- absorptions.[20] Noninvasive quantification of intracellular oxygenation in muscle is useful in determining the physio-pathological state of patients in such applied areas as sports medicine or physiotherapy. In vivo NIR spectroscopy of myoglobin represents one step towards fully understanding the respiratory chain. Unfortunately, the visible–NIR absorption spectra of Hb and myoglobin are not easily distinguished. Therefore in blood perfused muscle, NIR oxygen saturation measurements represent the extent of both Hb and myoglobin oxygenation in the muscle.

Hb and myoglobin oxygenation measurements yield useful information about oxygen delivery and utilization. In some ways, the intracellular oxygen content can be inferred based on these measurements. A direct approach using cytochrome oxidase measurements is often preferred to determine intracellular oxygen content and utilization. Intracellular oxygen content, specifically oxygen consumption, is of primary importance in understanding physiology and pathophysiology when dealing with cellular neural damage,[21] cerebral activity,[22] and detection of hematomas[23] since proper brain function is highly dependent on oxygen content. Oxygen metabolism at the cellular level occurs at the mitochondria in the presence of cytochrome oxidase. Cytochrome oxidase is the terminal enzyme in the cellular respiratory chain located within the membrane of the mitochondria. The enzyme contains four redox active centers that change oxidation state during electron turnover of the enzyme. Oxygen consumption in tissue involves the transfer of electrons from the redox active centers to the oxygen. The transfer of the four electrons to oxygen and the subsequent reaction with cellular matter results in the production of four hydrogen ions, which leads to the formation of water. In the NIR region, the reduced and oxidized states of the Cu_A copper center of cytochrome oxidase have weak absorption bands.[9] In the oxidized state, cytochrome oxidase has a weak broad absorption band between 700 and 900 nm centered at 830 nm. This broad absorption band is de-emphasized in the reduced state of the enzyme. Since the total concentration of cytochrome oxidase does not change during oxidative phosphorylation, spectral measurements of the enzyme reflect the relative concentration of the oxidized and reduced species. As such, simply observing the spectral difference between the oxidized and reduced forms of the enzyme allows for concentration measurements of the intracellular oxygen consumption to be determined. The spectral difference in the oxidized and reduced forms of the enzyme is shown in Figure 1. Since the flux of reducing equivalents into the cytochrome electron transport chain is regulated by the rate of adenosine triphosphate hydrolysis or work, the application of NIR to detect the redox balance of cytochrome oxidase can provide a noninvasive method to monitor oxygen consumption and activity at a cellular level.

Although the extinction coefficients for cytochrome are similar in magnitude to Hb, as seen in Figure 1, cytochrome quantitation using the NIR is difficult. In general, the concentration of cytochrome oxidase is approximately one tenth the Hb concentration in most tissue.[24] Substantial work has been done to interpret NIR spectra and extract the cytochrome signal from the strong Hb absorption, with limited success.[19] Reliable qualitative measurement of the oxidation and reduction of the Cu_A copper metal center of cytochrome *c* oxidase using NIR spectroscopy remains an active and hotly contested area in in vivo spectroscopy.

The human body is composed of approximately two-thirds water. Although the majority of the water is contained within cells, approximately 8% is present in the bloodstream. In the NIR wavelength region of tissue spectra, the strongest absorption is due to the O–H water absorptions. The water spectrum in the NIR region, as shown in Figure 3, contains a strong combination band at 1940 nm, the first overtone of the O–H stretching vibration at 1440 nm, a weaker combination band at 1200 nm and second and third overtone vibrations at 980 and 740 nm.[25] Light absorption by tissue is mainly due to the water present in tissue, therefore making the NIR spectral region ideal for studying the properties of water in tissue. Water and its temperature dependent spectral variation has been a field of interest in biomedical tissue spectroscopy. The extinction coefficients of water are extremely temperature dependent, such that a

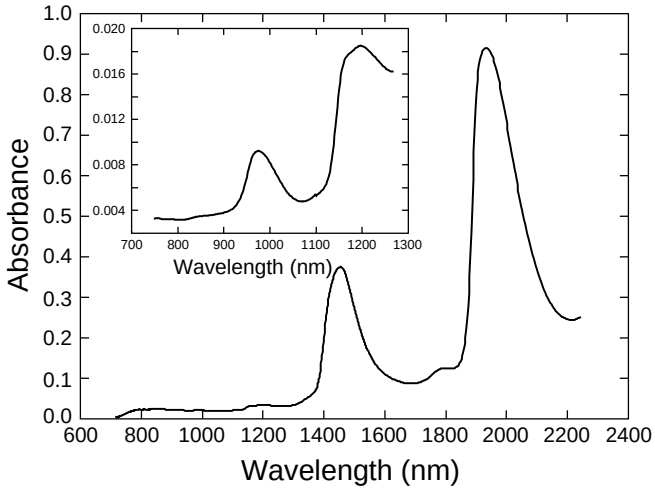

Figure 3 NIR absorbance spectrum of water.

temperature variation will cause a spectral shift in the NIR spectrum.[25] A temperature increase from 37 °C to 41 °C will cause the water spectrum to shift to shorter wavelengths (a hypsochromic shift). Therefore, one application of the water absorption band is as a noninvasive temperature monitoring system. Since the absorption of light is proportional to the chromophore content present in tissue, NIR can be used to determine quantitative and qualitative in vivo tissue hydration.[26] In some clinical measurements, water is assumed constant and can be used as a fixed constant absorber or internal standard to correct for pathlength variations arising from tissue scattering.[27,28] Assuming the water content in tissue is known, the absorption observed is in direct proportion to the optical pathlength traveled by the light in the tissue. Scattering and absorption at the water wavelength is assumed to be representative of those occurring at other wavelengths in the sample. However, this method is often difficult to implement since the water absorption may not reside in the same spectral region as the chromophore of interest and absorption from other chromophores in the tissue may also occur at the water wavelength. The limit in obtaining absolute quantitation in in vivo NIR spectroscopy originates from the multiple scattering of light within tissue resulting in an unknown pathlength and chromophore absorption.

2.3 Tissue Scattering

Absorption of light by a sample in a nonscattering medium provides a quantitative assay of the analyte present. However, the transport of light through complex samples such as tissues is governed by the light-scattering interactions in the medium. Scattering of light occurs as a direct result of the interaction of light with random variations in the refractive index or small particles in the medium, resulting in a dispersion of the light in all directions. In particular, scattering within tissue results from microscopic and macroscopic constituents such as cellular membranes, refractive index mismatching between intra- and extracellular fluids, mitochondria, ribosome and fat globules. Particles smaller than the wavelength of the incident light can scatter radiation elastically without a change in energy. Small particle scattering in which the scattering occurs in all directions is called Rayleigh scattering.[29] Rayleigh was the first to investigate from a theoretical standpoint the dependence of the scattered radiation intensity on the incident wavelength. In many quantitative analyses related to tissue, scattering has attributes which resemble scattering patterns described by Mie scattering.[30] Mie scattering is associated with scattering when the diameter of the particles is not much smaller than the incident wavelength. Derivable from Maxwell's equation, Mie scattering is a rigorous solution to the angular dependence of scattering for an arbitrary direction. However, both Rayleigh and Mie scattering provide only an approximate estimate of the wavelength-dependent scattering that occurs in tissue. The theories do not take into account multiple scattering or the presence of other nearby particles, an effect that becomes important for scattering in tissue.

When tissue is illuminated with NIR light, some of the light is absorbed by the tissue while a large portion of the light is diffusely scattered. The absorption coefficient, represented by μ_a, indicates the frequency of absorption events per unit length traveled. In tissue, absorption by chromophores is small, with typical values in the NIR ranging between 0.1 and 10 cm^{-1}.[30] The scattering coefficient, μ_s, represents the frequency of scattering events per unit length traveled and ranges from 100 to 1000 cm^{-1} in tissue, approximately two orders of magnitude greater than absorption.[30] The combined attributes of low absorption and high scattering in tissue allow for the light to penetrate deep within tissue, approximately 1–10 cm[31] enabling extraction of deep tissue information noninvasively. The high remittance of the scattered light makes the NIR ideal for studying biological and physiological characteristics. The penetration depth into tissue is dependent on the type of scattering, more specifically, the direction in which the scattered photons travel.

A single photon entering a tissue sample will experience many scattering events as it propagates through the medium. When light interacts with multiple scatterers, the light is altered from its original direction. These interactions of light with tissue are highly forward directed, i.e. light propagates in a forward direction. Scattered photons can be described by a scatter probability distribution function $P(\theta)$, where θ is the scattered

photon angle after the interaction with a single scattering event.[32] Phase distributions describe the scattering angle probability of the deflected light. The mean cosine angle of the phase distribution is used to describe the anisotropy of the scattered photon. This parameter, designated g, is called the anisotropy coefficient with values ranging from -1 (a totally back-scattered event) to 1 (a totally forward-scattering event). Isotropic scattering, $g = 0$, refers to a uniformly distributed scattering in all directions. Typical values of g for tissue in the NIR wavelength region range from 0.65 to 0.95.[30] The scattering anisotropy can be combined with the scattering coefficient to produce a single term, the reduced scattering coefficient, $\mu_s' = \mu_s(1 - g)$. Reduced scattering coefficients can be regarded as an effective isotropic scattering coefficient that represents the cumulative effect of several forward-scattering events. This reduced scattering coefficient parameter has special significance with respect to photon diffusion theory.[33,34]

A photon traveling through a tissue sample may experience varying degrees of scattering, resulting in long paths through the tissue sample. Therefore, one effect of scattering is to increase the pathlength traveled by the photon, thus increasing the absorption of the light by chromophores. Either photon time[35,36] or frequency[37,38] resolved techniques can obtain the distribution or mean paths a photon travels through a given scattering sample. A representative diagram of both techniques is shown in Figure 4. Photon time-of-flight methods, a time resolved technique, use short pulses of illumination (<50 ps) in which the light is diffusely scattered in the sample. Semiconductor, dye, or solid state lasers produce the ultrashort pulses at discrete wavelengths in time-of-flight instruments. The output pulses from the laser are split in two by a beamsplitter in which a fraction of the light enters into a reference detector and the remaining pulse is directed into the tissue sample and ultimately to a response detector. The reference and/or response detectors used with time correlated photon counting instruments are photomultiplier tubes (PMTs), microchannel plate-PMT, avalanche photodetector, and streak cameras. The time difference between the reference and response detectors is directly related to the time the photon spent in the tissue sample. Electronics are used to digitize and collect the output response for a single photon event or photon path through the tissue. Using a large number of photons, an intensity distribution is constructed of the sum of the number of photons with various times through the sample. A measurement of the photon time distribution effectively probes a series of pathlengths through the sample. In general, the light measured in a scattering sample has traveled considerably further than the direct distance between the input and output sources. The mean time, $\langle t \rangle$, corresponds to the center of mass of the photon time distributions as demonstrated in Figure 4(a). In essence, the detected photon distributions consist of the input response convolved with the absorption and scattering response from the tissue. Several methods exist to extract the chromophore concentration and scattering from the measured photon distribution.

Frequency domain or intensity modulated instruments also use a laser as a source except the intensity is modulated at radio frequencies with measurements of the intensity and phase made through the tissue sample. Small phase lag detection at radio frequencies are difficult, so practically all frequency resolved instruments employ a down-conversion scheme to bring the radio frequency signal down to intermediate frequencies where conventional phase detection methods can be applied. Knowledge of the phase shift and the modulation frequency can be used to determine the mean photon pathlength through the tissue sample. Applying this

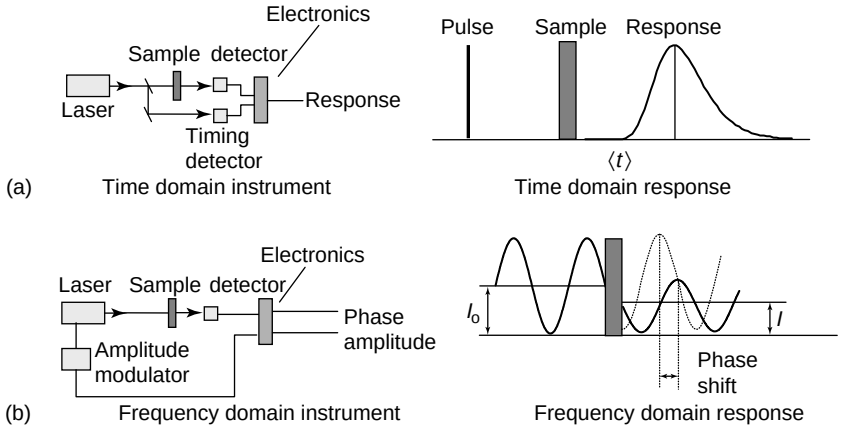

Figure 4 Time and frequency domain measurements. (a) A block diagram of a typical photon time-of-flight set-up with its corresponding measured time domain response. (b) A representative block diagram of a frequency modulated instrument and its measured amplitude and phase responses.

modified or corrected pathlength into the Beer–Lambert relationship permits the determination of chromophore concentrations in a scattering medium such as tissue.

3 INSTRUMENTAL METHODS

Discovery of IR radiation can be traced to an experiment undertaken by William Herschel in 1800. Herschel carried out a simple experiment to satisfy his curiosity as to which color within the visible spectrum was responsible for heat in sunlight. The experiment consisted of measuring the temperature from one color to the next using a thermometer and a glass prism to separate the colors of the sun's "white" light.[39] Nothing was observed until the thermometer was positioned below the red end of the spectrum. At that point a temperature change indicated energy was present beyond the red light. Because Herschel could not see that light, he named it IR, using the Latin prefix "infra" meaning below. William Herschel is credited for the discovery of the IR region and is often known as the "father" of NIR spectroscopy.[40] Contemporary NIR instruments have vastly improved from Herschels' initial experiment. NIR instruments can be used to measure tissue optical properties in vivo. These instruments can be classified as either broadband or discrete wavelength instruments, based on the wavelength selection method used.

3.1 Broadband Spectroscopy

Broadband spectroscopy refers to methods that observe the continuous spectral absorbance characteristics of a given sample within a given region of the electromagnetic spectrum (UV, visible, NIR, etc.). NIR broadband instruments for in vivo applications consist of three main components: a light source, a wavelength dispersing element, and a detection system. Light sources for broadband spectroscopy emit a wide spectral distribution of radiation. The most common light source for the NIR is the glass-encapsulated tungsten filament lamp. These sources provide radiation over the 400–2500 nm wavelength range being limited by the transmission characteristics of the glass. For samples requiring slightly higher light intensities than the tungsten lamp a quartz tungsten halogen lamp is available. Quartz tungsten halogen sources typically run at high temperatures thus producing high light throughputs. However, these sources tend to be less stable than normal tungsten lamps unless proper stabilizing electronics are incorporated into the lamp power supply.

Based on how wavelength selection is made, broadband NIR instruments can be classified as either dispersive or interferometric instruments. Dispersive wavelength selectors are the most commonly used, spatially spreading the light into its respective wavelengths using either prisms or diffraction gratings. Prisms separate the wavelengths spatially based on the wavelength dependence of the refractive index of the prism material. Light striking a prism is dispersed or separated in a nonlinear fashion such that shorter wavelengths (UV) are highly dispersed, while the longer wavelengths (NIR region) are less dispersed. Gratings, on the other hand, disperse the light into its wavelength components based on light diffraction and constructive wavelength interference rather than the variation in refractive index with wavelength. A diffraction grating is a plane or concave surface with ruled parallel, equally spaced grooves. Some of the basic designs of dispersive instruments are shown in Figure 5. Depicted in Figure 5(a) is the Bunsen design, used in early prism instruments. It uses two lenses and a prism for the wavelength selection. The separated wavelengths, illustrated as λ_1 and λ_2, are positioned on a single element detector by rotating the prism. Likewise, the Czerny–Turner design, depicted in Figure 5(b), requires the grating to rotate for wavelength

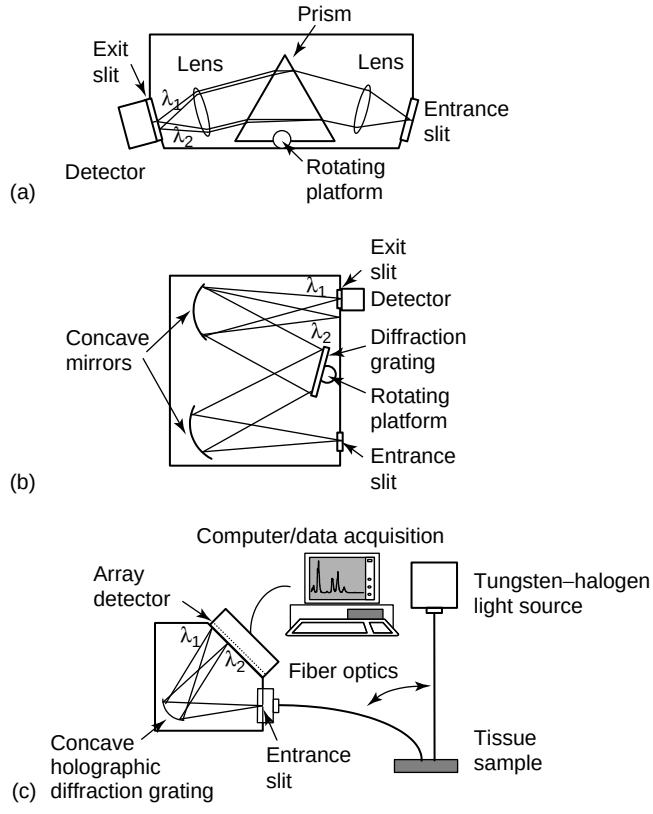

Figure 5 Representative NIR dispersive instruments incorporating (a) prism configuration (Bunsen design), (b) grating configuration (Czerny–Turner design) and (c) concave holographic diffraction grating design as the wavelength dispersive elements.

selection into the detector. Typical single element detectors used in the NIR region are PMTs and semiconductor based Si (360–1000 nm), PbS (900–2600 nm), InSb (1000–5000 nm), PtSi (1000–5000 nm) and InGaAs (1000–3000 nm). Prisms were commonly used in earlier NIR instruments for wavelength selection; however, recent instruments employ a grating design. Gratings offer a higher linear dispersion than prisms, typically tens of nanometers per millimeter. Holographic gratings have replaced mechanically grooved gratings. These have fewer defects, are easier to manufacture and have virtually perfect groove patterns. Additionally, modern instruments that incorporate gratings have a greater throughput across the entire NIR region.

Both the prism and grating designs of Figure 5(a) and (b) require lenses and/or mirrors to direct the light to the dispersive element. Wavelength selection is achieved by rotating the dispersion element. Optical components such as the mirror, lenses, and mechanical motion contribute noise and reduce instrument reproducibility. To circumvent these problems a ''nonmoving part'' dispersive instrument has been designed incorporating all the optical components into one system, as shown in Figure 5(c). The spectragraph of Figure 5(c) employs a concave holographic grating for both the dispersive and focusing elements, reducing the need for optical components, thus increasing the light throughput and reducing stray light within the instrument cavity. In place of the single element detector for single wavelength detection, modern dispersive instruments employ solid-state array detectors. These detectors consist of an array of a large number of closely packed miniature photoelectric semiconductor detectors. The detectors are either photodiode array or charge-coupled devices (CCDs). CCDs are solid-state sensors where the photon energy is converted and stored in a metal-oxide semiconductor. The greatest advantage of the array detectors is the immediate acquisition of the spectral response with readout times in the millisecond range. Since these detectors have multiple sensors, a rotating dispersive element is no longer necessary.

Interferometers are a class of instruments employing nondispersive devices in which the constructive and destructive interference of the light is exploited to obtain spectral information. Interferometers observe all optical frequencies from the source in a simultaneous and nondispersive manner. A classic example of an interferometer is the Michelson interferometer depicted in Figure 6(a). The instrument consists of several mirrors, a HeNe laser, photodiode detector, a beam splitter, and a NIR detector. Light from the source is split into two portions by the beam splitter. Half of the beam traverses the beam splitter and strikes a fixed mirror, while the remaining light is reflected to a moving mirror. When the mirrors are equidistant from the beam splitter i.e. no displacement, all wavelengths interfere constructively. Zero displacement is depicted as the large centered spike in Figure 6(b). As the mirror moves about this zero position, some of the wavelengths will constructively and destructively interfere. The degree of constructive/destructive interference is dependent on the mirror displacement and the wavelength of the incident radiation. The detected signal intensity versus the mirror displacement is known as an interferogram. In conjunction with the polychromatic

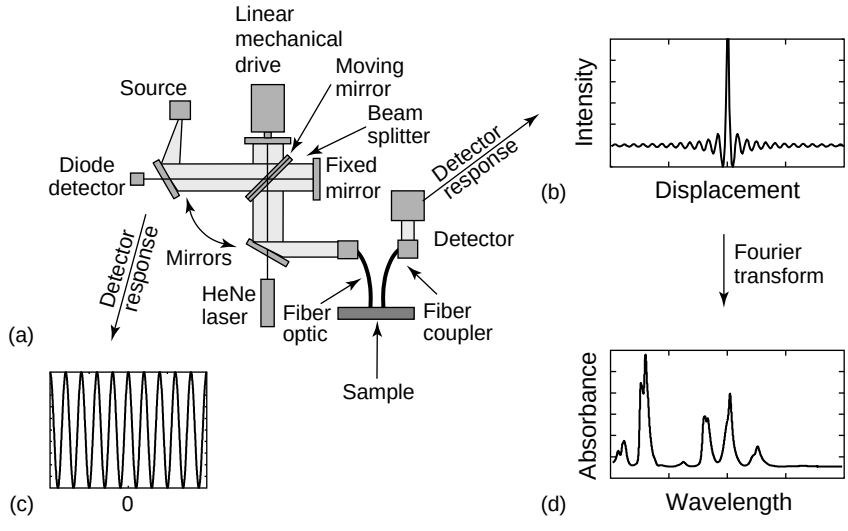

Figure 6 A typical Michelson interferometer. (a) General arrangement of the optical components of the Michelson interferometer. The interferometer produces an interferogram response illustrated in (b) for polychromatic light. Monochromatic light from a laser diode produces the response seen in (c). (d) Fourier transform of the interferogram to create the spectral response in the wavelength domain.

source, a monochromatic laser reference source follows a similar path through the interferometer mirrors. Since the laser is monochromatic, constructive/destructive interference from the laser produces bright and dark fields on the photodiode detector. The detector response for a monochromatic source is demonstrated in Figure 6(c). A dark field occurs when the moving mirror is placed $\lambda/4 \pm n\lambda/2$ ($n = 0, 1, 2, 3 \ldots$) away from the zeroth position. The detected intensity and mirror displacement for a monochromatic source is related to the wavelength by a cosine function, $\cos(4\pi x/\lambda)$. Polychromatic light consists of several wavelengths, each wavelength following a different cosine relation to produce constructive/destructive interference. The detected intensity for a polychromatic source is the sum of all the constructive/destructive interference from the various wavelengths emitted by the source. Shown in Figure 6(b) is a typical interferometric response for a polychromatic source. A Fourier integral can describe the intensity detected versus mirror displacement. Therefore, using the Fourier transform on the interferogram produces the spectral response in the wavelength domain as demonstrated in Figure 6(d).

The two greatest advantages of an interferometric instrument are the high spectral resolution that can be obtained at a fairly high signal-to-noise ratio (S/N) and the large light throughput. The improved S/N, in comparison to a dispersive single element detector, results from the multiplex advantage (Fellget's Advantage) of interferometric instruments. Interferometers permit many wavelengths to be observed simultaneously since the detector is exposed to all wavelengths during the scan process. To demonstrate this multiplexed advantage, a simple S/N comparison can be made with a dispersive single element detector scanning instrument. For the S/N comparison, both interferometer and scanning instruments are assumed to cover the same number of m wavelength points at comparable resolution and time. Since spectra are taken over a constant time interval, the number of photons observed with the interferometer instrument is m times more than the scanning system. Thus, the S/N improvement in the interferometer is a factor of $m^{1/2}$ better than the scanning instrument. Therefore, the multiplex advantage of interferometers occurs in situations where the detector noise is the limiting factor and signal averaging is necessary. Since interferometers have no entrance slits, the light throughput (Jacquinot's advantage) is significantly larger than dispersive instruments. The throughput can often be a factor of 10 to 100 times more than conventional dispersive instruments.

3.2 Discrete Wavelengths

Discrete wavelength spectroscopy is the simplest method to make in vivo tissue measurements. Many in vivo NIR commercial instruments incorporate filters or diode lasers to measure the status of living tissue. The most well known discrete wavelength instrument is the pulse oximeter, designed to measure the relative tissue oxygen content. Millikan was one of the first people to measure Hb/myoglobin oxygenation with the use of two filters in the green and yellow wavelength regions.[41] The filter wavelengths corresponded to the wavelength absorption of the oxygenated and deoxygenated forms of the chromophores. Spectroscopic instruments that employ filters for wavelength selection are typically referred to as spectrophotometers or photometers. Interference filters are the most commonly used method for discrete wavelength selection. Interference filters share some characteristics with infererometric methods employing destructive and constructive interference for wavelength selection. Interference filters are constructed in such a manner that when light strikes the filter, most of the light interferes destructively while a small band of wavelengths interfere constructively and pass through the filter. Depending on the filter design, the selected wavelength band can be wide covering a large wavelength range or extremely narrow (ca. 10 nm band).

The light sources of choice for discrete wavelengths when a narrow bandwidth is required are light emitting diodes (LEDs) and laser diodes. These sources are optoelectronic semiconductor devices that produce monochromatic light by electroluminescence. The material usually consists of GaAsP and GaP for wavelengths in the 380–800 nm range and GaAs for those beyond 800 nm. Oximetry is a prime example of the application of LED and laser diodes to determine blood oxygenation.[42] The application of discrete wavelengths to in vivo spectroscopy relies on prior knowledge of the spectral characteristics of the chromophore of interest. The choice of wavelengths to use for a particular application cannot be made arbitrarily and must be made based on the extinction coefficients of the chromophore. For example, in pulse oximetry the choice of wavelengths is based on the absorption coefficients of the oxygenated and deoxygenated species of Hb, as demonstrated in Figure 7. The absorbances measured at 660 nm and 940 nm are predominately due to the deoxygenated and oxygenated Hb, respectively. The appeal of semiconductor sources lies mainly in the narrow spectral bandwidth achievable (<7 nm), small package size (~5 mm × 5 mm), and large radiant output. However, the discrete wavelengths available are somewhat limited, which often leads to a less than optimal in vivo measurement.

Liquid crystal tunable filters (LCTFs) are an ideal discrete wavelength selection instrument with narrow bandwidths (5–50 nm bandwidths) tunable over a broad wavelength range (400–1100 nm). LCTFs are electronically controlled interference filters that exploit the

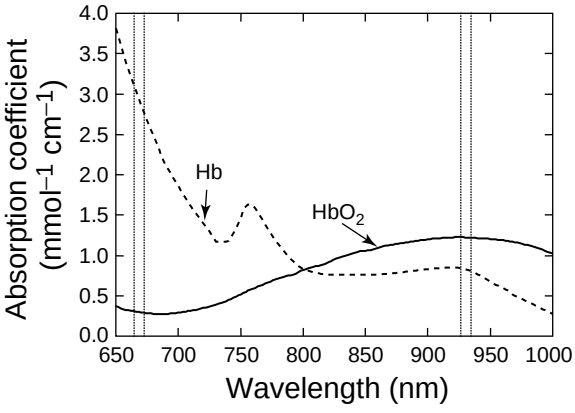

Figure 7 Single wavelength selections based on the absorption coefficients of oxy- and deoxyhemoglobin for pulse oximetry applications. The absorbances at 660 nm and 940 nm are predominantly due to the deoxygenated and oxygenated Hb, respectively.

polarization properties of birefringent materials to produce tunable discrete wavelength selection. As such, the LCTF has no moving mechanical parts and is capable of random access wavelength selection. This feature of the LCTF allows for discrete wavelength selection similar to interference filters with the tunability of a broadband spectroscopic instrument. LCTFs are typically used in applications demanding multiple discrete wavelength selection such as multi-probe fluorescence and tissue spectroscopy.

Related to LCTFs, acousto-optical tunable filters (AOTF) are another means of a nonmechanical method of discrete wavelength selection over a broad wavelength range in the NIR region. AOTFs consist of a carefully cut tellurium dioxide crystal in which acoustic waves at radio frequencies are used to isolate a single wavelength of light from a broadband source. The acoustic wave propagating within the crystal alters the refractive index of the crystal. The crystal behaves like an enhanced diffraction grating in which multiple diffraction occurs over the volume of the crystal to produce a single wavelength output. The wavelength output is dictated by the acousto-optical properties of the crystal and the input radio frequency. Discrete wavelength selection is achieved by varying the frequency of the radio wave that is coupled into the crystal. The unique feature of the AOTF is its ability to adjust its transmitted light throughput by varying the radio frequency power to the crystal.

3.3 Spectroscopic Imaging

In general, NIR spectroscopy using either broadband or discrete instruments is carried out at a single point on the in vivo tissue sample. The information acquired is localized to a distinct spot (1-cm diameter) on the tissue. Often this single point measurement is used to infer the regional health status of the tissue. However, there are clinical situations in which several such locations must be probed and assessed. Acquiring point spectroscopy at multiple locations introduces confounding variations and inaccuracies. Several methodologies have been developed to circumvent the limits of point spectroscopy, in particular, in vivo spectroscopic imaging. In vivo imaging requires a noninvasive method to collect pictures of the tissue at selected wavelengths. To add an imaging component to Fourier transform spectrometers, the interferometer has been modified to an imaging system incorporating a two-dimensional array at the image focal plane. The revised instrument permits the simultaneous collection of multiple spectra at multiple locations within the tissue sample. However, Fourier transform imaging spectrometers are generally limited to in vitro tissue samples.

CCDs, invented in 1970 by Boyle and Smith, are the most well known and used of the array detectors.[43] These detectors convert light or photons into an electrical charge that is collected and stored in a metal oxide semiconductor. The accumulated charge is a linear function of the incident and exposure time to the radiation. The CCD response represents the reflected spatial light intensity of a given object. In the scientific realm, CCDs are most often used for the detection of weak signals or sources. Unlike conventional CCD, scientific grade CCDs have a lower dark noise, higher responsivity, a larger dynamic range and higher resolution. The operating spectral range of a typical CCD detector is between 400 and 1100 nm. However, a CCD detector for in vivo spectroscopic applications is of little use without some means of wavelength selection. Filters, the simplest form of wavelength selection, are often used in conjunction with CCD detectors. The filters are usually mounted on the circumference of a rotating wheel, thus permitting the acquisition of several different wavelengths. Although filters are an excellent means of wavelength selection, the filter wheel involves mechanical motion, which is prone to displacement errors or low reproducibility. Theoretically, an infinite number of filters can be placed on the filter wheel; however, in practice only a limited number can be used, ordinarily six to eight filters. Recently, CCD detectors have been merged with LCTF technology to produce wavelength-dependent images as shown in Figure 8. Each pixel in the spectral image would have a corresponding NIR spectral response. Likewise, AOTFs can be used in combination with CCD technology to produce wavelength-dependent images. However, AOTF spectral images tend to suffer from lateral chromatic aberrations resulting from the input design of the crystal. Both LCTFs and AOTFs have the advantage of no mechanical motion, an unlimited

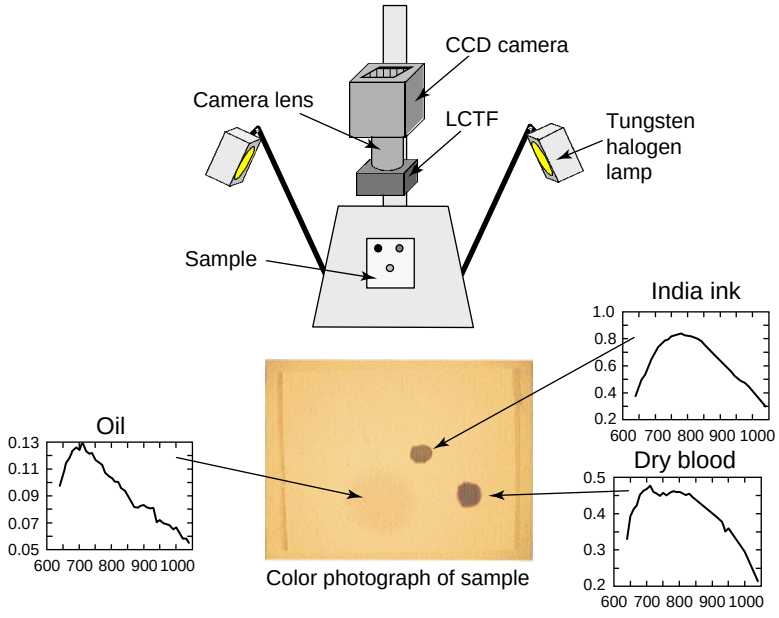

Figure 8 Schematic of a LCTF silicon CCD spectroscopic imaging system. The bright field image of a test card with a deposit of oil, blood and ink is presented along with the spectra associated with a pixel from each of the areas where a sample was deposited onto the card.

selection of wavelengths, and a fast response in the hundreds of milliseconds range.

4 PRACTICAL CONSIDERATIONS

NIR attenuation spectra of living tissue can potentially provide information related to the physiological processes being probed as well as the anatomical structure of the tissue being investigated. However, in vivo spectra are strongly influenced by physical factors related to the measurement configuration, as well as the instrument-to-tissue interface. Often these latter factors are confounding and of little interest to the experiment, yet these factors can be significant contributors to the variation in the spectrum or between spectra. Sources of confounding variation and their effects on in vivo spectra must clearly be understood and accounted for in the analysis of in vivo tissue spectra.

4.1 Pulsatile Modulation

One complicating factor is the modulation of the optical attenuation of tissue by pulsatile blood flow. While this well-known phenomenon is exploited in pulse oximetry[44–46] and photoplethysmography,[47] pulsatile modulation represents a source of ~1 Hz noise when only the underlying tissue absorptions are of interest. Pulsatile modulation results from the cardiac cycle of the circulatory system, which is responsible for the transport of oxygenated blood to tissue. During the systole phase of the cardiac cycle the left ventricle contracts and causes blood to be ejected from the heart. At diastole the left ventricle refills with blood. In the peripheral vasculature, the systole–diastole phases of the cardiac cycle result in a modulated dilation and contraction of arteries and arterioles. This small difference in blood volume between the systolic and diastolic phases of the cardiac cycle produces the observed modulation in the attenuation of light by the tissue. This pulse due to blood volume change is heavily dampened in the highly compliant venous system. The resulting pulsatile blood volume gives rise to a corresponding modulation in the optical attenuation which is largely localized in the arterial compartment of the circulatory system.

The pulsatile modulation depth of the optical attenuation of tissue depends on the vascular system being probed. In a highly vascularized tissue bed, the modulation depth can be as high as 15% of the signal. Figure 9 displays the mean ± SD of a time series of optical attenuation spectra collected over several hundred cardiac cycles at the index finger of the nonpreferred hand. Figure 10 displays the time course of the optical density of a finger tip at 800 nm over several cycles. The modulated optical signal has the same temporal contour as that measured by an arterial blood pressure catheter inserted percutaneously into a peripheral artery. The peak systolic and diastolic pressure points as well as the sharp incisura that coincides with the closure of the aortic valve and the cessation of ventricular ejection

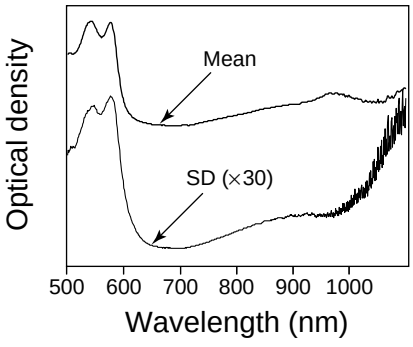

Figure 9 Mean and standard deviation of time series of 2048 attenuation spectra from the index finger of the nonpreferred hand. Instrument noise is the cause of increasing standard deviation beyond 1000 nm. Arterial modulation is the cause of increasing standard deviation of the oxyhemoglobin vibronic bands at 540 and 576 nm as well as in the 850 nm region due to the broad charge transfer band of oxyhemoglobin.

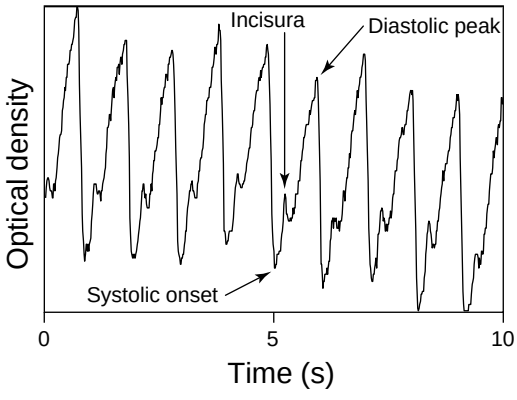

Figure 10 Time course of the optical density at 805 nm from the index finger of a subject. Both diastole and systole phase of each cardiac cycle as well as the sharp incisura are clearly visible.

are clearly visible in the noninvasive optical reflectance signal.

The variability of the heart rate leads to a nonstationary time series as a result of fluctuations in both amplitude and in the temporal domain of the arterial blood pressure waveform. The power spectral density calculated from the spectroscopic time series is shown in Figure 11. Power spectral analysis of the temporal optical waveform clearly displays the ~1 Hz pulsatile modulation component. Variability in the heart rate is responsible for the distribution of frequencies at approximately 1 Hz in the estimate of the power spectral density plot. Data analysis of arterial blood pressure waveforms based on autoregressive and cross-spectral analysis as well as bivariate spectral (phase and coherence spectra) analysis are useful as an aid in detection of relevant cardiovascular pathologies such as hypertension and diabetes.[48–50]

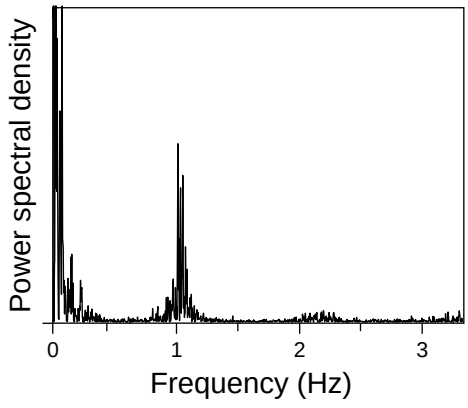

Figure 11 Power spectral density of 2048 attenuation spectra acquired at a sampling frequency of 6.6 Hz. Due to the nonsinusoidal near periodicity of the pulsatile modulation, overtones at approximately 2.2 and 3.3 Hz are clearly visible. The broadening of the bands at the fundamental and overtones is caused by heart rate variability.

The light scattering properties of tissue are slightly altered by pulsatile modulation of tissue blood volume. However, the primary effect is on the absorption characteristics of the medium and oxygenated Hb absorptions are preferentially modulated by the 1 Hz pulsatile component. The preferential pulsatile modulation of oxyhemoglobin absorption features is related to the much higher Hb oxygen saturation in the arterial compartment of the tissue compared to the venous system and capillary beds. Mixed venous blood Hb oxygenation ranges between 72 and 78% and roughly reflects a global tissue oxygen measure. Arterial Hb oxygen saturation typically ranges between 95 and 98%.[51] Pulse oximetry exploits this arterial compartment localization of the pulsatile modulation of the optical signal. Pulse oximetry monitors tissue reflectance or transmission at two wavelengths at which oxygenated and deoxygenated Hb have significantly different extinction coefficients. The fraction of oxygenated Hb is based on the magnitude of the differential absorption or reflectance at those two wavelengths. The static (DC) component of the signal is assumed to represent the fraction of oxygenated Hb within the volume of tissue which is being probed optically, while the time varying (AC) component of the reflectance response is largely attributed to the arterial compartment of that tissue volume. Arterial Hb oxygen saturation is empirically correlated with the AC component of the differential reflectance response. Thus, pulse oximetry provides a noninvasive measure of arterial Hb oxygen saturation.

Photoplethysmography determines peripheral arterial pressure waveforms by measuring the temporal shape of the optical reflectance over the cardiac cycle. Clinically, the magnitude of the pulsatile modulation of the optical signal correlates with the sufficiency of the arterial supply

of blood to the peripheral tissue. The entire arterial pressure waveform can be used to assess peripheral resistance, arterial compliance and wave propagation and reflection. Thus photoplethysmography provides a useful noninvasive clinical diagnostic tool in evaluating the blood supply to the extremities as well as accessing the severity of peripheral vascular disorders.[52–54]

Pulsatile modulation represents a source of 1 Hz variation in the in vivo NIR spectrum of human tissue. The spectral dependence of the pulsatile modulation of the optical attenuation of light between 500 and 1080 nm by tissue indicates that the modulated component of the signal provides information that is specific to the arterial compartment of the tissue that is being optically probed. While this has been exploited by pulse oximetry, spectroscopic analysis of the pulsatile modulated component of the tissue reflectance or transmission remains largely unexplored.[55] This latter aspect may provide a more reliable means of analyzing blood specific analytes noninvasively using NIR spectroscopy.

Power spectral analysis of the time series of optical attenuation spectra (Figure 11) reveals a number of low frequency noise contributions to the signal in addition to the ~1 Hz pulsatile component. One of the contributors to low frequency variation in the optical attenuation of tissue arises from respiration. Like the pulsatile component, modulation of the optical reflectance or transmission of tissue by respiration is site dependent. In addition, the experimental configuration can also have a profound effect on the magnitude of the respiration-induced variation of the optical signal. For instance, the contact pressure of the probe against the skin can be modulated by respiratory motion.[56,57]

4.2 Probe–Tissue Contact Pressure

A number of confounding variables arise in measurement configurations that require a probe to come in contact with tissue. Contact pressure between the optical probe and tissue represents a serious source of variation in the observed attenuation spectrum. The pressure exerted on the skin by the probe has a direct physical effect on the optical properties of the tissue. In some cases, it may also produce a blanching response in the underlying tissue.

The propagation of light through tissue is heavily influenced by scattering as a result of tissue inhomogeneity. Scattering is determined by the particle size, shape and distribution. As well as increasing the density of scattering centers in tissue, probe pressure against tissue squeezes out interstitial fluid and blood from the tissue. These factors modify the scattering properties of the underlying tissue and therefore influence the propagation of light in the tissue.

Capillary pressure in humans is in the range of 16 to 32 mmHg in the arteriolar limb and the forearm.[58] Physical pressures exerted on the capillaries that exceed capillary pressures impede the flow of blood through the capillaries. When this occurs, the pressure forces blood out of the underlying cutaneous vessels, thus changing the blood volume present in the underlying tissue. Visually, a reduced blood volume in the tissue causes the appearance of a white spot (or blanch) in lightly pigmented skin. In more heavily pigmented skin, the visible absorption of melanin prevents the identification of a blanched region. A change in the blood volume of tissue changes both the intensity of the Hb absorptions and alters the scattering properties of the underlying tissue. This results in a major perturbation of the NIR attenuation spectrum of the tissue. While blood flow to a region is impaired, cells are deprived of oxygen and nutrients and waste products accumulate. The condition of restricted blood flow is called ischemia. If prolonged, ischemia may result in necrosis of the affected tissue.

When the source of pressure is removed, blood flow resumes and circulation of blood is restored to the blanched area. There is an inverse relationship between the amount of pressure exerted on tissue and the length of time before this pressure damages the tissue.[59] If the time–pressure limits are exceeded, nonblanchable erythema may result. Nonblanchable erythema is a state in which pressure applied to skin does not cause a visible blanching response. This occurs as a result of tissue ischemia, caused by direct physical pressures exceeding capillary closure pressure for extended periods of time. Shear forces and friction also contribute to the development of nonblanchable erythema. Nonblanchable erythema is the first stage in the development of a pressure related wound or pressure sore.[60] As the blanching response cannot be observed in darkly pigmented skin, other less reliable indicators (such as edema or warmth) must be used to determine the presence of first stage pressure sores in individuals with dark skin.

In general, NIR spectroscopy does not involve probe-to-skin contact for lengths of time or at such pressures that nonblanchable erythema, or more importantly severe tissue damage would result. However, varying the contact pressure between the probe and skin will alter the degree of blanch response in tissue. It is important to note that both tissue scattering and tissue blood volume are altered in the region being probed spectroscopically as a result of the varying pressure applied to the surface of the tissue.

Analysis of the mean NIR reflectance response (1000–2400 nm) from skin and the probe-to-skin contact pressure indicates that probe-to-skin contact pressure is highly positively correlated (+0.85) with the mean reflectance response from skin. Fluctuations in the probe-to-skin contact pressure thus result in a varying NIR

reflectance and can represent a major source of variation in the reflectance response from tissue. The spectral dependence of the correlation between contact pressure and single-beam reflectance shows a distinct correlation pattern (Figure 12). If the spectral regions where water strongly absorbs are ignored, there is an overall negative slope in the positive correlation as a function of increasing wavelength. This behavior may be interpreted as the decreasing influence of tissue scatter on the overall light attenuation by tissue at longer wavelengths. Therefore, pressure induced changes in the tissue scatter function have a diminished effect on the attenuation of the longer wavelengths of NIR light by tissue.

The subtle slope change in the pressure–reflectance correlation spectrum is overshadowed by the large drop in reflectance–pressure correlation in the regions that are characterized by strong spectral absorptions. The most prominent feature is the large change in correlation over the OH stretch–bend combination region (1800–2000 nm). The OH stretch overtone/combination region (~1400 nm) displays a much smaller drop in pressure–reflectance correlation compared to the stretch–bend combination region. At wavelengths beyond 2200 nm the correlation drops dramatically as light attenuation by the tissue increases.

The anomalous reflectance–pressure correlations that occur in the region of the strong NIR absorptions of water arise from pressure induced blood volume changes. Fluid becomes displaced from the optical path when probe contact pressure is increased. Fluid displacement results in a decrease in the water absorption related attenuation of the NIR light. As the absorption coefficient of the water increases from the OH stretch overtone/combination region (~1400 nm) to the OH stretch–bend combination region (1800–2000 nm), the magnitude of this effect is expected to be greater at the longer wavelength water absorption band. In spectral regions where light attenuation is dominated by water absorption, pressure-induced fluid displacement and the resulting decrease in absorption offset the increase in NIR light attenuation due to higher probe pressure against the skin.

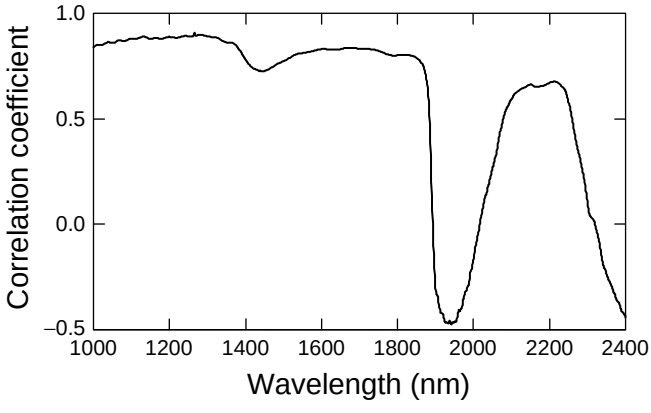

Figure 12 Pearson correlation coefficient between probe-to-skin contact pressure and skin reflectance measured between 1000 and 2400 nm on the abdomen of a group of six study volunteers. Each subject was measured for a 2 h period and the cumulative with-in and between subject pressure–reflectance correlation is reported.

Unlike the variation induced by pulsatile blood flow, the pressure exerted on the tissue by an optical probe is a variable that can be controlled by the experimenter. Another condition that the investigator must take note of is the effect of tissue temperature on spectroscopic measurements. This is another aspect where variability can be controlled.

4.3 Tissue Temperature

The optical properties of tissue are temperature dependent.[61] Qualitatively, the scattering of light by tissue can be thought of as arising from microscopic discontinuities in the refractive index within the tissue medium. For example, such discontinuities or boundaries exist at the interface between the cell membrane and the extracellular matrix. Since the refractive index of a medium is strongly temperature dependent, the scattering of light by tissue is expected to be sensitive to changes in tissue temperature. In addition to this direct influence on the optical scattering properties of tissue, temperature variations indirectly change the optical properties of tissues due to the potent vascular response to a local change in temperature. Since both the direct physical and physiological consequences of varying tissue temperature have significant effects on tissue optics, these effects must be taken into account in the analysis of in vivo NIR attenuation spectra of tissue.

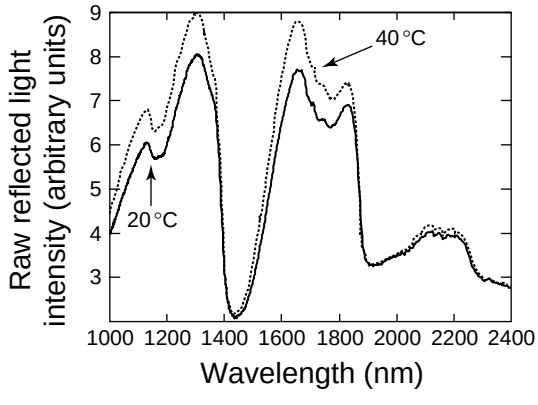

Figure 13 Representative traces of the reflectance response between 1000 and 2400 nm of the abdomen of a subject with a skin temperature of 20 and 40 °C.

As with the other confounding factors that have been discussed, the physiological response to a change in tissue temperature varies considerably depending on the type of tissue and the site.[61] For instance, thermoregulatory systems can significantly alter superficial skin blood flow and cutaneous blood volume when skin temperature is changed. In Figure 13 the reflectance response over the range 1000–2400 nm from the upper abdomen of a subject is compared at 20 °C and 40 °C. In Figure 14 the difference in the reflectance response of the upper abdominal site at 40 °C and 20 °C are compared over a study population consisting of 17 subjects. Figure 14 clearly indicates that both the magnitude and the wavelength dependence of the reflectance response to varying skin temperature are highly reproducible over the study population. The upper abdomen consistently had a higher reflectance at 40 °C. Figure 15 plots the correlation coefficient between skin temperature and reflectance responses from the abdomen over the course of a 2 h temperature ramping protocol carried out in a study group of 17 subjects. The reflectance response shows a significant high positive correlation (+0.77) with the temperature of the abdomen in the temperature ramping protocol. The high positive correlation between the reflectance response and skin temperature over most of the 1000–2300 nm wavelength region (Figure 15) is considered to be due to the temperature dependence of the scattering properties of the tissue. It is absent in the corresponding correlation spectra obtained from temperature ramp experiments using nonscattering aqueous media where the thermal volume displacement has been accounted for. Ignoring the spectral regions where water absorbs strongly, an overall negative slope in the positive correlation between reflectance and skin temperature is observed as a function of increasing wavelength. As noted in the previous section, the gradual drop in the correlation between skin temperature and reflectance at longer wavelengths arises from the decreased importance of scatter to the overall attenuation of light by the tissue. The decreased contribution from scattering at longer wavelengths results in a smaller general temperature effect on the overall attenuation of light by tissue in the spectral regions devoid of water absorptions.

Skin temperature–reflectance and pressure–reflectance correlation spectra differ in detail in the regions of the spectrum dominated by strong water absorptions. The observed decreased attenuation of light by water absorption as a function of probe contact pressure arises from the displacement of interstitial fluids and blood from the underlying tissues due to the applied pressure. The magnitude of this pressure effect was observed to be proportional to the extinction coefficient of water. However, the correlation between the reflectance response from skin and skin temperature over a water absorption band is not strictly a function of the extinction coefficient of water at each wavelength. Water absorption bands are known to be extremely temperature sensitive.[25,62,63] In particular there is a temperature dependent shift in the maxima of the water absorption bands. These band shifts give rise to the sharp derivative-type features and regions of negative correlation in the skin temperature–skin reflectance correlation spectra.

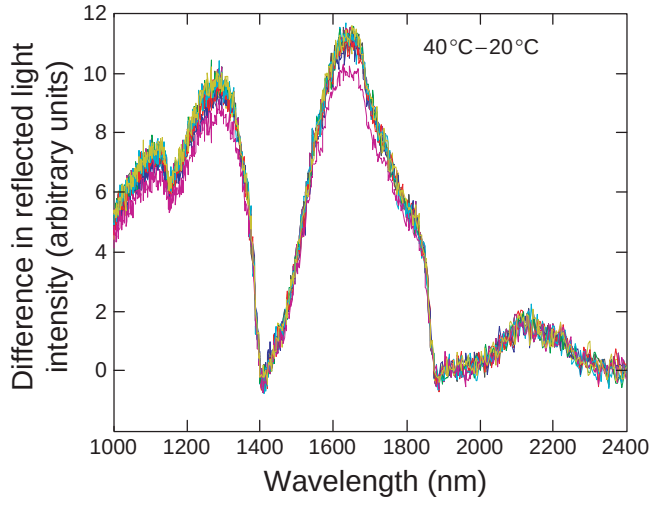

Figure 14 The difference in the reflected light intensity from a site on the abdomen when the skin is at 40 °C and 20 °C. The difference spectra from 17 subjects are overlaid to indicate the high inter-subject reproducibility of the temperature response of the skin reflectance.

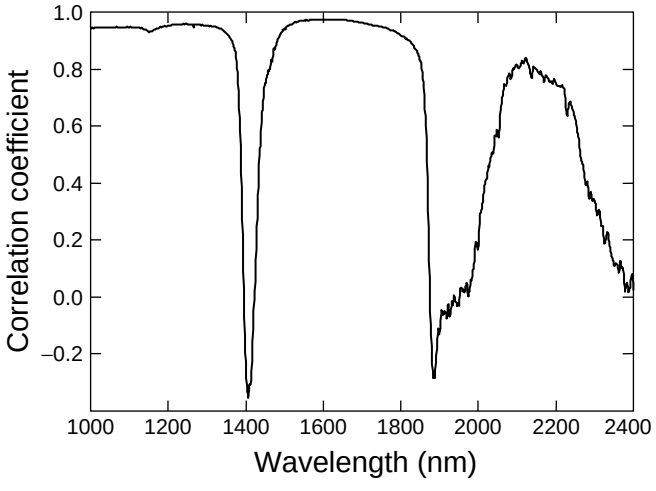

Figure 15 Pearson correlation coefficient between skin temperature and skin reflectance measured between 1000 and 2400 nm on the abdomen of a group of 17 study volunteers. Each subject was measured for a 2 h period over which the temperature was modulated between 20 and 40 °C. The cumulative with-in and between subject temperature–reflectance correlation is reported.

Variation of tissue temperature has a profound effect on the scattering properties of the medium. Superimposed on the effect of temperature on tissue scattering, temperature dependent spectral shifts in water absorption bands give rise to more subtle variations in tissue spectra. These combined effects are difficult to model and can represent a serious confounding factor in the realm of in vivo spectroscopy. Often they are simply ignored or overlooked. Of particular concern are situations where the variation in tissue temperature is correlated with the intervention or manipulation of interest that is being monitored spectroscopically.

5 SELECTED EXAMPLES

To date, in vivo NIR spectroscopy is primarily applied to the assessment of tissue perfusion and oxygenation of tissue. Tissue blood perfusion is a key element in the thermoregulation of tissue. Disruptions of blood flow, ischemia, or poor blood perfusion will often result in a drop in tissue temperature. Thus when NIR spectroscopy is used to determine the extent of ischemia or the degree of reperfusion, perturbations in the attenuation spectra related to tissue temperature must be considered. Preferably, they should be compensated for at the outset in the experimental design.

Detection and monitoring of poorly perfused (ischemic) and poorly oxygenated (hypoxic) tissue remains an important issue in many fields of medicine and biology. NIR spectroscopy offers a noninvasive means of assessing the balance between tissue oxygen delivery and oxygen utilization. Monitoring the relative concentrations of oxygenated and deoxygenated Hb provides a measure of th Hb oxygen saturation of tissues. NIR spectroscopy can determine information about the status of tissues at one point in time, or by monitoring a time series of spectra, it is possible to monitor changes as a function of time. Combining these capabilities and using an array detector, the spatial–temporal distribution of oxygenated and deoxygenated Hb in tissue can be imaged. Oxygen consumption through mitochondrial respiration can be inferred from changes in the visible–NIR spectra that arise due to a shift in the redox state of cytochrome oxidase during hypoxia. Reduced and oxidized cytochrome oxidase have distinct visible–NIR absorption spectra bands and thus, significant alterations in the ratio of reduced to oxidized enzyme in tissue can be determined spectroscopically. Time series of spectra are very useful for monitoring changes that occur during ischemia and reperfusion.

5.1 Forearm Ischemia–Reperfusion

The human forearm and leg have been convenient investigation sites for many of the noninvasive NIR studies that have examined the local response of tissue to manipulations of blood flow.[64–73] A disruption of blood flow (ischemia) to the arm can be produced by arresting circulation to the forearm with a pressure cuff, effectively impeding both arterial inflow and venous outflow. During ischemia, blood present in the forearm becomes increasingly deoxygenated. As deoxygenation progresses, the oxyhemoglobin doublet attenuates and is replaced by a single broad peak, centered on 564 nm (see Figure 2). This peak is characteristic of deoxyhemoglobin. NIR spectroscopy can detect similar oxygen related changes to the spectrum that occur during ischemia. The significant increase in optical density during ischemia in the region of the charge-transfer band of Hb at 760 nm is consistent with the increased fraction of Hb in ischemic tissue and is perhaps the clearest indicator in the NIR of the change in Hb oxygenation during ischemia. The relative concentrations of Hb and HbO_2 can be monitored by fitting the NIR attenuation spectra to the extinction coefficients of the chromophores.

Figures 16 and 17 compare the laser Doppler measure of blood flow throughout the ischemia–reperfusion protocol with the change in the calculated concentrations of [Hb] and [HbO_2] obtained from the ischemia–reperfusion time series of NIR spectra. The sum of [Hb] and [HbO_2] provides a measure of the total amount of Hb (tHb) present at any point in time. The steep drop in blood flow illustrated in Figure 16 corresponds to the inflation of the pressure cuff. As expected, [Hb] begins increasing as soon as the blood flow is cut off. The tissue consumes the limited amount of oxyhemoglobin, causing [HbO_2] to decrease during circulatory arrest to the forearm. It is apparent in

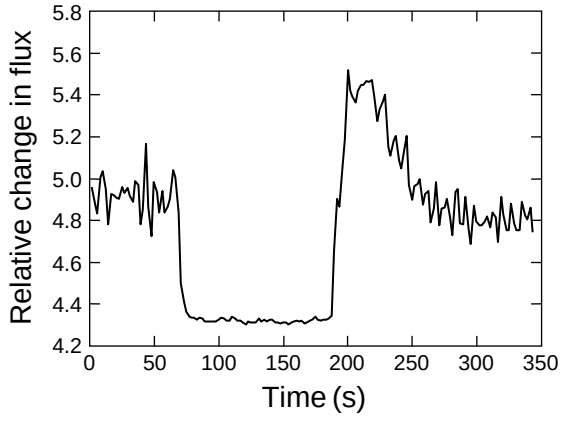

Figure 16 Laser Doppler flux measure of the change in forearm blood flow during a forearm ischemia–reperfusion protocol.

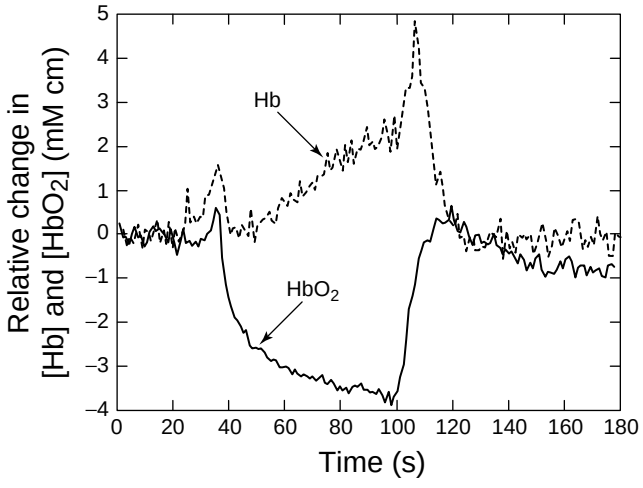

Figure 17 Change in the relative concentrations of oxy- and deoxyhemoglobin as measured by NIR spectroscopy. The spectral region between 740 and 840 nm was used to determine chromophore concentration changes. NIR spectra were measured at 2 s intervals over the same 350 s forearm ischemia–reperfusion protocol as presented in Figure 16.

Figures 16 and 17 that the changes in [Hb] and [HbO_2] are strongly correlated with the laser Doppler measure of blood flow to the forearm (−0.45 and +0.71, respectively).

An interesting result of both ischemia and hypoxia is hyperemia. In Figure 16 it can be seen that upon release of the pressure cuff blood flow increases dramatically, temporarily surpassing the preischemia level. This response is called reactive hyperemia; a similar reaction following hypoxia is termed hypoxemic hyperemia. As tissue oxygen concentration decreases during ischemia, the vascular system attempts to compensate by adjusting vascular resistance. Blood vessel walls contain a layer of smooth muscle that is very sensitive to the interstitial concentration of adenosine, a potent vasodilator. Under normal circumstances, the interstitial concentration of adenosine remains stable. However, during ischemia the normal washout of interstitial adenosine is prevented, causing it to accumulate. Additionally, the aerobic production of adenosine triphoshate (synthesized from adenosine) is arrested. This also contributes to the accumulation of interstitial adenosine. As a result of the accumulated adenosine, ischemia causes the blood vessels to become very dilated. When ischemia is relieved, the locally relaxed state of the vascular system permits an increased volume of blood to flow through that region, resulting in reactive hyperemia. As the excess interstitial adenosine is washed out, the vascular smooth muscle becomes less dilated and normal circulation resumes.[74]

During hyperemia, the increase in blood volume is naturally accompanied by an increase in tHb. As the influx of blood is arterial, it is highly oxygented. This causes an increase in [HbO_2]. Meanwhile, [Hb] decreases dramatically through venous washout. Note that in Figures 16 and 17 the flux, [Hb] and [HbO_2] return to their preischemic levels following hyperemia.

During ischemia the NIR tissue attenuation spectrum displays both an increased baseline offset and a quasi-linear wavelength dependent slope change. These features are characteristics of change in the scattering properties of the tissue. The increased blood volume associated with the hyperemic phase of the protocol has the opposite effect on the baseline offset and causes a reciprocal change in the slope of the spectrum. As the blood volume approaches the level that existed prior to ischemia, the spectrum offset and slope return to the level observed in the spectra acquired during the preischemia phase of the protocol.

NIR spectroscopic studies examining the response of the forearm to brief periods of ischemia followed by reperfusion clearly indicate that the measured changes in the attenuation spectrum of the tissue correlate with periods of changing forearm blood flow and blood volume, as well as oxygenation. The calculated changes in the concentrations of Hb, HbO_2 and therefore tHb and Hb oxygen saturation are weakly dependent on the range of wavelengths used to fit the spectra. The calculated concentration values are also dependent on the distance separating the light source and the light collector. Despite these shortcomings, crudely fitting the NIR attenuation spectra to the extinction coefficients of Hb and HbO_2 provides a trend monitor for the change in the relative concentrations of these species. Therefore this method provides a measure both of the oxygen delivery to the forearm and of the forearm blood volume. Early forearm studies demonstrated the potential of NIR spectroscopy to monitor the degree of tissue ischemia and the extent of reperfusion. Since then clinical interest has largely focused on developing NIR spectroscopy as a means of detecting cerebral ischemic events in high risk, low birth weight neonates as well as for monitoring skeletal muscle during exercise or the heart during surgery.

5.2 Myocardial Ischemia–Reperfusion

The ability of NIR spectroscopy to determine the oxygenation status of myoglobin makes it extremely useful for monitoring the consumption of oxygen by working muscle. Consequently, NIR spectroscopy is being increasingly applied in the area of sports medicine.[64,65,75–78] The heart muscle has also been studied by NIR spectroscopy.[79–82] Myocardial ischemia is extremely detrimental to cardiac function. Even a few minutes of ischemia may cause damage to the heart. Damage may contribute to exacerbation of infarction, impairment of contractile or conductive abilities or it may cause vascular

dysfunction.[83] Reperfusion of the heart may also cause damage, through mechanisms similar to those in muscle and brain tissue.[84] Myocardial ischemia occurs as a result of occlusion or a spasm of a coronary artery leading to a reduction or complete cessation of blood flow to the area perfused by the affected vessel. In addition, ischemia is encountered during surgical procedures requiring aortic cross clamping. NIR spectroscopy has been used to measure cardiac oxygenation[79–81] and tissue pH.[82]

Langendorff perfused rat hearts provide a useful model for examining myocardial ischemia–reperfusion. The absence of blood in this system permits the observation of myoglobin, without interference from Hb. When both Hb and myoglobin are present in a tissue sample, it is quite difficult to tell them apart, spectroscopically. In the absence of Hb, NIR spectroscopy can distinguish the fraction of oxygenated myoglobin based on the distinct spectral characteristics of oxy- and deoxy-myoglobin.

Figure 18 summarizes the results obtained from a series of spectroscopic and functional measurements made on eight Langendorff perfused rat hearts over the course of an ischemia–reperfusion protocol. Figure 18 compares the raw spectral means obtained from the full study group from the preischemia, ischemia and reperfusion time segments of the protocol while Figure 19 reports the corresponding mean difference spectra between the preischemia, ischemia and reperfusion segments of the protocol. Statistical analysis on the raw data indicates that the NIR attenuation ($\lambda > 650$ nm) changes significantly ($P < 0.05$) over the ischemia–reperfusion protocol. This effect arises largely from the change in the scattering properties of the tissue during ischemia and reperfusion relative to scattering properties of the myocardium prior to ischemia. In the short wavelength

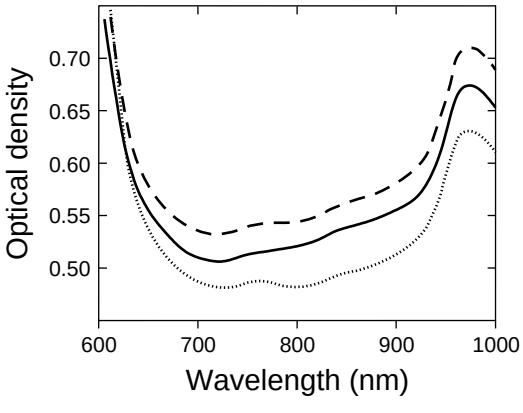

Figure 18 Mean NIR attenuation spectra from the left ventricle of a rat heart undergoing ischemia and reperfusion. Each trace represents the mean spectrum obtained prior to ischemia (——), during ischemia (· · · ·) and at reperfusion (- - - -). The mean spectra were calculated over a group of eight rats subjected to the same ischemia–reperfusion protocol.

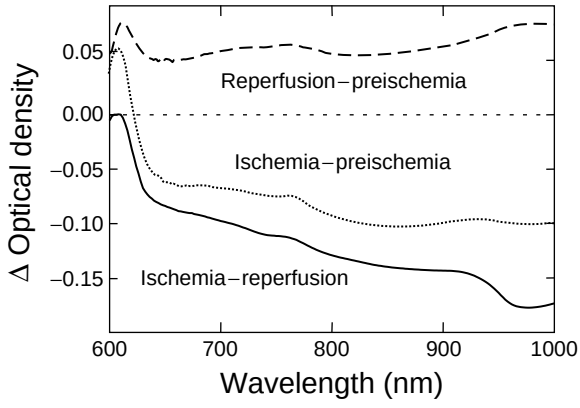

Figure 19 Mean difference spectra between the preischemia, ischemia and reperfusion segments of the myocardial ischemia–reperfusion protocol. Mean differences were calculated over a study group population of eight rats.

NIR ($\lambda = 650$–1000 nm), the attenuation of light by tissue is dominated by scattering. A small change in tissue scatter function can markedly affect the attenuation of NIR radiation by tissue. This is evident from the mean spectra and mean difference spectra presented in Figures 18 and 19 respectively. There is a significant offset and slope change in the NIR between preischemia, ischemia and reperfusion. Post-hoc means comparisons indicate that tissue scattering changes most dramatically at reperfusion resulting in decreased NIR reflectance from the heart.

The tissue scatter coefficient is highly temperature dependent. Strict temperature control throughout all segments of the experiment is crucial in order not to introduce temperature-related artifacts in the optical attenuation of the tissue. Statistical analysis indicates that the changes observed in the spectral time series are not significantly correlated with the measured temperature variations over the experiment and that heart temperature does not change significantly over the various segments of the ischemia–reperfusion protocol. The observed change in spectral offset and slope can be rationalized in terms of a varying mean photon pathlength over the ischemia–reperfusion protocol. These changes in the scattering properties of the tissue over the spectral time series would appear to arise as a direct result of ischemia–reperfusion on the heart as opposed to a secondary tissue temperature effect. Successfully accounting for the scatter-induced variation in the spectra would enable absolute myoglobin oxygenation changes to be determined by fitting the Mb and MbO_2 extinction coefficients to the scatter-corrected NIR spectra.

In vivo NIR spectra over the 700- to 900-nm region are usually fit to the extinction coefficients of three chromophores, [Hb, Mb], [HbO_2, MbO_2] and $[Cytaa_3]_{oxidized}$ minus $[Cytaa_3]_{reduced}$. The ischemia–reperfusion-related changes in the tissue-scattering function are often ignored,

or it is assumed that the scattering losses do not change over the protocol. The water contribution is also ignored despite the fact that water has two absorption bands in the spectral region used by most of the fitting routines. This omission is generally rationalized by the assumption that tissue water content is constant and does not change over the duration of the protocol. In Langendorff heart preparations, failing to account for the correlated changes in tissue scattering and water content confounds the calculated changes in the chromophore concentrations over the course of the ischemia–reperfusion protocol.

The correlated scatter-related changes in the attenuation spectrum present a particular problem when trying to determine changes in the redox state of Cu_A, the copper metal center of cytochrome *c* oxidase. The differential NIR absorption of the oxidized and reduced Cu_A center of the enzyme, [$Cytaa_3$], at around 830 nm are difficult to differentiate in the presence of tissue oxygenation changes and changes in the scattering characteristics of the tissue. Based on three chromophore fits without accounting for water and scatter changes in the spectra, it is by no means clear that the time series of redox state changes of [$Cytaa_3$] represents changes in the enzyme absorption characteristics or whether the changes are due to the scatter-related changes in the spectra or residuals from the fit to [Hb, Mb] and [HbO_2, MbO_2]. In general, the highly correlated response across the NIR effectively masks more subtle spectral features associated with enzyme Cu_A oxidation–reduction and makes unequivocal assignment of these redox-related changes in the Cu_A absorption feature of the enzyme difficult.

When superficial tissues are investigated, the visible heme $a + a_3$ transition can be used to monitor the metabolic activity of the mitochondria.[85,86] Upon reduction of the heme $a + a_3$ center of cytochrome *c* oxidase the heme α-band shifts from 590 to 605 nm and increases in intensity. The increase in intensity of the visible heme α-band upon reduction is the opposite behavior compared to NIR Cu_A absorption feature of the enzyme. Therefore the use of both the visible heme $a + a_3$ transition and the NIR Cu_A absorption feature would appear to be a prudent strategy to track mitochondrial oxygen utilization.

The heme α-band of cytochrome *c* oxidase lies on the shoulder of the intense visible Q_o Hb/myoglobin vibronic transition. The broad Q_o band of deoxyhemoglobin/myoglobin has a higher extinction coefficient in the long wavelength visible compared to the corresponding transition when the heme is oxygenated. Thus, simply monitoring the attenuation in the region between 590 and 620 nm will once again produce results that are heavily confounded by Hb/myoglobin oxygenation changes. Difference spectroscopy in which reference Q_o and Q_v Hb/myoglobin visible vibronic transitions are subtracted from the spectrum can be used to visualize changes in the heme α-band of cytochrome *c* oxidase. Line narrowing methods such as Fourier self deconvolution[87] and linear prediction extensions of the method[88] are often extremely useful in resolving overlapping spectral features and may offer a means of extracting the heme $a + a_3$ center oxidation changes more cleanly. Figure 20 compares visible attenuation spectra processed using the LOMEP[89] (linear shape optimized maximum entropy linear prediction) line narrowing procedure. The reduced heme α-band of cytochrome *c* oxidase becomes apparent in the processed spectra during myocardial ischemia. The integrated intensity of the isolated heme α-band is plotted over an ischemia–reperfusion protocol in Figure 21. In this Langendorff preparation reduction of the heme $a + a_3$ center which results in an increase in the integrated area of the isolated heme α-band closely tracks oxygen delivery to the heart. Higher attenuation at 610 nm upon ischemia relative to preischemia and reperfusion is consistent with the accumulation of reducing equivalents of the heme $a + a_3$ center in the absence of molecular oxygen. The positive optical density at 610 nm in the reperfusion minus preischemia difference spectra indicates that the redox balance of heme $a + a_3$ center of cytochrome *c* oxidase is not fully restored after reperfusion. A higher fraction of the enzyme is reduced relative to the preischemic state. The increased fraction of reduced enzyme following reperfusion may be a good indicator of the degree of functional damage to the heart caused by the ischemia–reperfusion sequence.

Chromophore concentration changes determined from the series of NIR spectra are consistent with the changes

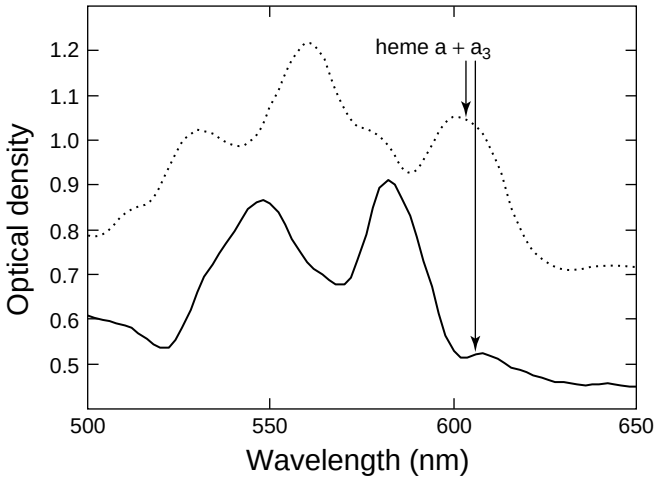

Figure 20 LOMEP resolution enhanced visible spectrum of the myocardium preischemia (——) and during ischemia (· · · ·). The heme α-band (heme $a + a_3$ transition) of cytochrome *c* oxidase increases in intensity upon reperfusion indicating an accumulation of reducing equivalents of the heme center during ischemia.

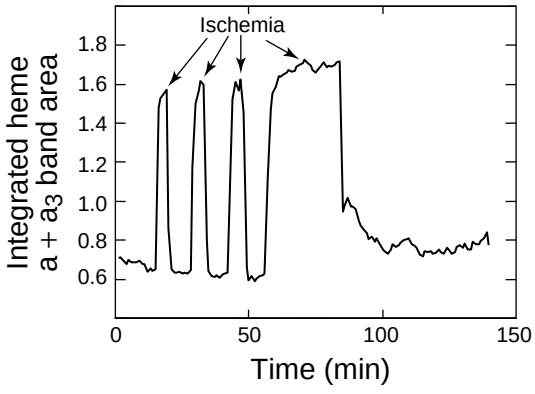

Figure 21 Time course of the integrated area of the heme α-band of cytochrome *c* oxidase. Reduction of the heme center of the enzyme results in an increase in heme α-band intensity. The three preconditioning ischemic events and the longer 20 min ischemic period are clearly present in the time trace. Following the 20 min ischemia, the redox state of the heme center of the enzyme does not fully return to its preischemia condition indicating cellular damage to the heart.

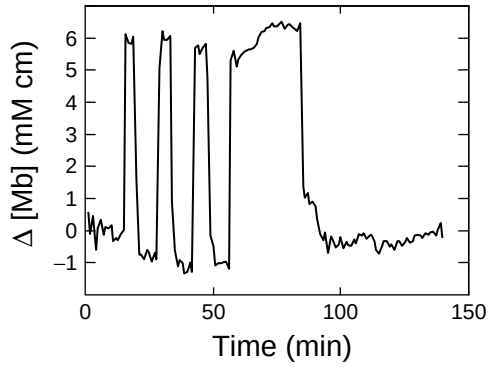

Figure 22 Changes in myoglobin concentration based on simultaneously fitting the extinction coefficents of $[Cytaa_3]_{oxidized} - [Cytaa_3]_{reduced}$, [Mb], $[MbO_2]$ and $[H_2O]$ as well as a scatter offset correction to the time series of NIR spectra covering the region between 740 and 880 nm.

observed for the visible heme α-band of reduced heme $a + a_3$ center of cytochrome *c* oxidase. As expected, Mb deoxygenation (Figure 22) closely tracks the oxygen delivery to the heart. The three 5 min preconditioning ischemic events and the 20 min ischemia are clearly evident in the [Mb] trace as well as the time trace for the heme α-band. The temporal fit to the NIR Cu_A absorption feature of cytochrome oxidase is strongly negatively correlated with the heme α-band of cytochrome *c* oxidase over the ischemia–reperfusion protocol. This is consistent with the known behavior of these transitions upon reduction of the metal centers of the enzyme. In addition to the heme α-band and the Cu_A absorption feature of cytochrome *c* oxidase being highly correlated, the transitions assigned to the enzyme are highly correlated with oxygen delivery to the heart. Thus in this preparation the heme $a + a_3$ and copper metal centers of cytochrome *c* oxidase mitochondria are reduced within a minute to changes in oxygen delivery to the heart. The last chromophore included in the fitting procedure was water. Generally the water content of tissue is considered to be static. However, based on the results of the chromophore fits to the attenuation spectra, the tissue water content displays a rapid and reversible change over the 5 min preconditioning ischemia segments but shows a marked drop over the 20 min ischemia (see Figure 23). Superimposed on this drop in water content during ischemia is a trend of increasing tissue water content over the course of the protocol. This is consistent with the increased production of heart edema (fluid accumulation) over time in the Langendorff preparation.

In addition to factors mentioned in the previous section, absolute measurements of oxy- and deoxymyoglobin concentrations are also confounded by the dependence of mean photon pathlength on the tissue scatter function. Changes in the tissue scattering properties alter the mean pathlength that photons travel through the tissue. In order to address this difficulty, several approaches have been suggested. The simplest approach eliminates the baseline offset between spectra and calculates the chromophore concentration changes from the time series of offset corrected spectra. The shortcoming of this approach is that it ignores the wavelength dependence of the tissue scatter function. The logical extension of this approach is to consider a linear as well as an intercept term to account for both the offset and slope changes between the spectra. This approach is often referred to as the multiplicative scatter correction. Obviously, it assumes a linear wavelength dependence for the tissue scatter

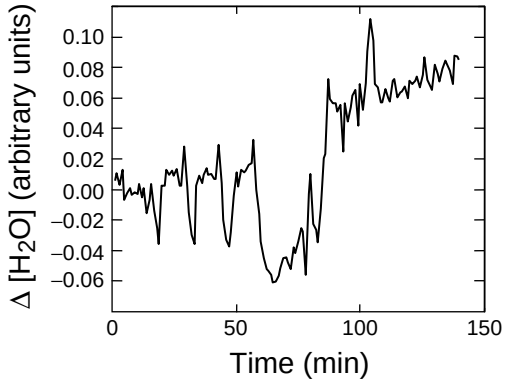

Figure 23 Change in the water content of the myocardium during the ischemia–reperfusion protocol. These results indicate an increased water content of the heart tissue over the time course of the protocol as well as a reversible loss of water during the ischemic phases of the protocol.

function. While this assumption does not hold over the full range of the NIR, the approximation seems reasonable when only a short wavelength segment of the NIR is used to fit the extinction coefficients of the major absorbing chromophores. A related scheme that has attracted much attention in the literature[27,28,90,91] uses the weak water absorption bands at 730 and 840 nm to perform a scatter correction. A second derivative of the spectrum is taken in order to eliminate the offset and the linear wavelength-dependent contribution of the scattering loss from the spectrum. The amplitude of the second derivative water absorption features are fit to a second derivative reference water spectrum to provide a scale factor that is then divided by the assumed water concentration to provide an effective photon pathlength (usually referred to as the differential pathlength factor). This method implies that tissue water concentration remains constant over the time series of spectral measurements and that changes in the second derivative amplitude of the 730 and 840 nm water features arise solely from the scatter dependence of the mean photon pathlength.

While the constant tissue water assumption may hold in a number of circumstances, as this example points out it is not always possible to make this assumption. In this example of myocardial ischemia–reperfusion, there is a significant increase in tissue edema over the course of the protocol. This leads to a severe breakdown of the constant tissue water assumption. At best in this example, the second derivative water correction provides a measure of the [Mb] to [H_2O] ratio in the tissue, allowing this trend to be monitored. For the reasons mentioned above, without an explicit determination of the optical scatter function of the sample tissue, the conversion of an attenuation spectrum into absolute chromophore concentrations is problematic. While baseline offset, multiplicative scatter correction and second derivative fit to water attempt to correct attenuation spectra for variations in the tissue scatter function each method has its limitations and shortcomings. Time and phase-resolved methods[92,93] offer a means to obtain scatter correction factors at discrete wavelengths that can then be used to correct conventional attenuation spectra. Increasingly these approaches, particularly phase-resolved measurements, are being relied upon to correct attenuation measurements for the scattering contribution to the measured loss. Another approach to determining changes in tissue scatter function is to simultaneously monitor several source–collector separations. This method effectively measures the attenuation losses at several different mean photon pathlengths. This approach, sometimes referred to as spatially resolved spectroscopy,[94] can be used to determine the scattering component of the attenuation loss, assuming that the different volumes of tissue that are sampled by each source–collector pair have the same effective scattering coefficient.

The conversion of NIR attenuation spectra of tissue into absolute measures of [Hb, Mb], [HbO_2, MbO_2] and $[Cytaa_3]_{oxidized}$ minus $[Cytaa_3]_{reduced}$ remains a key issue in the development of a reliable tool for the clinical assessment of oxygen delivery and cellular respiration. While the visible heme $a + a_3$ transition can be monitored with reasonable ease when the optical pathlength is relatively short, the strong attenuation of visible light by tissue limits the penetration depth to superficial tissues. The weak differential NIR absorption of the oxidized and reduced Cu_A center of cytochrome *c* oxidase is extremely difficult to extract in the presence of highly correlated spectral changes related to changes in Hb/myoglobin oxygenation and tissue scattering. This is particularly the case with deep tissue oxygenation applications of NIR spectroscopy such as cerebral oximetry for which few if any reference methods exist to aid in the calibration or refinement of algorithms aimed at providing absolute chromophore concentrations from NIR attenuation data.

5.3 Cerebral Oxygenation and Blood Flow

NIR spectroscopy has been extensively used to monitor the oxygenation of the brain.[95–98] The functional monitoring of cerebral activity following activation or stimulation using both NIR spectroscopy and magnetic resonance functional imaging techniques has recently received much attention.[99–102] Clinical monitoring of cerebral oxygenation in high-risk neonates[96,103–106] and during cardiopulmonary bypass[107,108] also remain important clinical applications of NIR spectroscopy.

Neonatal applications are mainly concerned with the incomplete development of autoregulatory control mechanisms governing cerebral blood flow. This commonly leads to incidents of cerebral ischemia and intraventricular hemorrhage in the low birth weight neonate population. These complications have devastating acute mortality and morbidity consequences on this population. Cerebral ischemia and intraventricular hemorrhage are also associated with a high risk of long-term neurological damage. There are a number of other conditions that place the oxygenation of the brain at risk in the neonatal-pediatric population, including birth asphyxia, traumatic head injury, near-drowning and cardiopulmonary bypass surgery. Currently there is no reliable or routine means of clinically monitoring cerebral oxygenation. NIR spectroscopy offers the potential of detecting cerebral circulation deficits by directly monitoring cerebral oxygenation, cerebral blood volume and blood flow. Considerable work has been carried out to develop a reliable clinical cerebral oximeter based on NIR spectroscopy.

NIR cerebral oximetry suffers from the confounding factors and complications that were described in the examples dealing with the investigation of superficial tissue oxygenation. However, deep tissue oximetry is complicated by additional factors. Oxygenation changes or changes in the optical properties of the surface and superficial tissues are generally not of interest in deep tissue oximetry applications. In cerebral oximetry applications, oxygenation changes of the scalp or changes in the optical properties of the skull must be discriminated against. The usual strategy that is employed is to ensure that the scalp and skull represent a minor component of the volume of tissue that is sampled by the collected photons. This can be achieved by an instrumental set-up in which the separation between the light source and the light detector/collector is adjusted to ensure that large tissue volumes are sampled. In human neonate cerebral oximetry, source–collector separations >1 cm are typical. Thus the tissue attenuation spectrum is heavily weighted to that of the cerebral tissue.

While it has been widely demonstrated that NIR spectroscopy can measure relative changes in cerebral oxygenation, absolute determination of [Hb] and [HbO_2] remain elusive. This causes absolute measurement of total Hb concentration or blood volume measurement to be problematic as well. In addition, cerebral Hb oxygenation is affected by neurological activity. In an ischemic region with minimal cerebral blood flow but abolished neuronal activity, Hb oxygenation may be normal despite the fact that the tissue is rapidly evolving to infarction. In these instances, a corresponding measure of cerebral blood flow, blood volume and mitochondrial oxygen utilization (redox status of cytochrome *c* oxidase) may be of more diagnostic value than a relative measure of cerebral oxygenation.

The strong absorption of visible light by tissue limits spectroscopic investigation of the heme α-band region of cytochrome *c* oxidase (590–620 nm) to superficial tissues. Instead the weak differential NIR absorption characteristics around 830 nm between the oxidized and reduced Cu_A center of the enzyme must be relied upon to infer mitochondrial oxygen use. Reliable in vivo measurement of mitochondrial oxygen use based on the NIR absorption of cytochrome *c* oxidase remains fraught with difficulties and has yet to be demonstrated under clinically realistic conditions.

The potential of NIR spectroscopy to measure cerebral blood flow has been demonstrated by exploiting indicator–dilution principles.[109–112] One approach introduces a bolus dose of a NIR fluorophore, indocyanine green (ICG), and follows the transit time of the fluorophore in the brain. ICG is rapidly eliminated thus generally only a single transit can be measured for each bolus injection of the fluorophore tracer. Usually multiple injections of ICG are administered and blood flow is calculated from a series of kinetic traces to try to provide a more reliable measure of cerebral blood flow. While ICG blood flow measurements have been used extensively for opthalmic applications, the procedure has not been widely adopted as a means of measuring cerebral blood flow. The invasiveness of the ICG administration makes the technique less suited to routine, bedside application.

A second, less invasive, strategy manipulates the arterial oxygenation by changing the fraction of inspired oxygen. The time for the measured cerebral oxygenation to respond to an abrupt change in arterial oxygenation has been used to provide a measure of cerebral blood flow. Both methods calculate cerebral flow values based on a modification of Fick's Principle in which the flow value is determined within the transit time of the tracer. When this assumption is fulfilled, venous outflow of the tracer can be ignored and the kinetic model is simplified. However the short tracer transit times in the brain, ~4 s, severely limits the quality of NIR measurement sequence that can be acquired. To date these measurements of cerebral blood flow have suffered from extremely high variability. This has led investigators to question the validity of cerebral blood flow values determined by this approach. More extensive kinetic modeling of the tracer impulse–response profiles may help provide greater consistency in the NIR-based measures of cerebral blood flow.

Considerable effort has gone into the development of NIR spectroscopic techniques to determine absolute cerebral oxygenation, blood volume, blood flow and the redox balance of the Cu_A center of cytochrome *c* oxidase. In spite of all this, enormous challenges remain. These NIR methods require further development and innovation to make them clinically practical and reliable and thereby fulfill the goal of having an inexpensive, portable bedside monitor able to detect cerebral circulatory deficits.

5.4 Near-infrared Assessment of Tissue Viability

There are clinical situations where the local oxygenation of tissue is of critical importance and it is in these situations that NIR spectroscopy may be the most useful. Determining the oxygenation of tissue that is near the surface of the skin is particularly accessible for NIR spectroscopic interrogation. Adequate oxygen delivery to wounds and reconstructed or transplanted tissue is requisite for proper healing. In the case of a burn injury, a key clinical question is whether the underlying tissue has the capacity to heal itself given proper wound care or whether the damage will extend over the full thickness of the skin. Full thickness injuries require tissue replacement usually in the form of skin grafting. Chronic poorly healing

wounds or dehiscent amputation sites generally suffer from inadequate blood perfusion (hypoperfusion) and poor oxygenation (hypoxia). Treatments are generally aimed at improving oxygen delivery to the wound site. Flaps and grafts form the basic procedures used in surgical reconstruction of skin and muscle. Following surgical elevation, the flap or graft experiences adverse circulatory changes that limits the supply of blood to the tissue and results in hypoxic sections of the graft or flap. Early and reliable assessment of the degree of compromised tissue perfusion has important clinical implications for the ultimate survival of tissue. The utility of NIR spectroscopy and imaging is being explored to determine peripheral vascular perfusion,[113–116] oxygenation of skin flaps and skin grafts[117–126] as well as its suitability as a general tool for the assessment of oxygenation and hemodynamics of superficial tissues during surgery.[127]

NIR spectroscopy is well suited to the assessment of oxygen delivery to compromised cutaneous tissue. The tissue under investigation in these applications is superficial, being at or near the surface of the skin. Small distances (<1 cm) between light source and collection optics (or detector) can be used to ensure that small tissue volumes are being predominantly sampled and thus near surface tissues are largely being probed in this configuration. This considerably simplifies the interpretation of the NIR spectral response compared to deep tissue investigations where the response from confounding superficial tissue is superimposed on the deep tissue spectral information. If only the upper surface of the tissue is being investigated, noncontact reflectance measurements can be made. This eliminates artifacts arising from the instrument–tissue interface. The shorter photon pathlengths that are typically sampled in the near surface applications enable a wider spectral range to be sampled. Visible (400–650 nm) and long wavelength NIR (1000–2500 nm) reflectance from tissues can be measured in the noncontact reflectance configuration or when source collector separations are on the order of 1 mm. Despite the fact that tissues are highly scattering in the NIR, image contrast is maintained for several scattering lengths. Thus NIR imaging can be used to resolve near surface features in tissues.

Clinicians empirically rely upon the visible appearance of tissue in the routine assessment of viability of tissue. Cyanosis, tissue swelling and desiccation provide important visual clues as to the status of skin flaps, chronic wounds or burn sites. Unfortunately, these differences only become clinically apparent several hours or days after the onset of the complication. The late manifestation of the clinical symptoms hampers early recognition of complications and often delays interventions aimed at alleviating the complication. Tissue swelling or the formation of edema affects the optical properties of the tissue in the NIR. Likewise tissue desiccation or dehydration along with cyanosis also profoundly affect the visible–NIR spectrum of tissue. Thus, well-perfused, healthy tissue and poorly perfused, compromised tissue will have different NIR spectral features.

The reversed McFarlane skin flap model provides an excellent model for demonstrating the capabilities of NIR spectroscopy and spectroscopic imaging of compromised cutaneous tissues. The model is illustrated in Figure 24. A pedicle skin flap is raised on the back of the rat. The base of the flap remains attached to an existing blood supply provided by the dorsal sacral vessels. The base of the flap is generally referred to as the vascular pedicle and provides the blood supply to the remainder of the skin flap. In this model, the flap is too long for the existing blood supply. The section of the flap proximal to the vascular pedicle remains well perfused and healthy while the section of the flap distal from the pedicle receives little if any blood supply. The central section of the flap is generally underperfused and ischemic. In the reverse McFarlane model, the distal section of the skin flap generally fails within 72 h of raising the flap while the proximal section remains healthy. The fate of the central section of the flap is less predictable and depends on the number and size of sacral vessels feeding the flap, the quality of the surgery and whether vascular complications arise as a result of the surgery or infection. NIR spectroscopy is best able to determine relative concentrations of oxy- and deoxyhemoglobin or changes in the concentrations of these species rather than providing an absolute concentration measurement. Similarly relative tissue water content or changes in tissue water content can be determined spectroscopically. Figure 25 compares NIR attenuation spectra of the central site of a rat's dorsum taken before and after elevation of the skin flap. Fitting a time series of attenuation spectra to the extinction coefficients of Hb, HbO_2, H_2O and $Cytaa_3$ provides a measure of the relative concentration of these species and the temporal change of

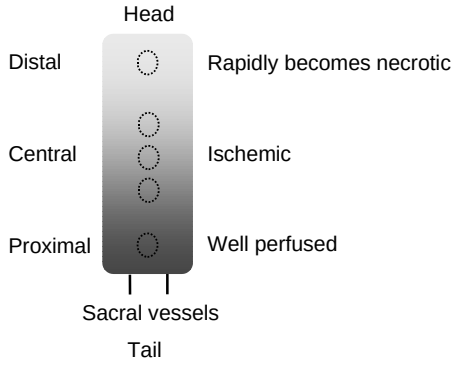

Figure 24 Schematic illustration of the dorsal reversed McFarlane skin flap model in the rat.

concentration of these chromophores in the tissue. Based on the relative changes in the Hb, HbO_2 concentrations, the degree of tissue oxygenation can be determined. Figure 26 presents the temporal change in the percent Hb oxygen saturation over a 120-h period measured at 6-h intervals at the proximal, central and distal sites on the dorsum of a rat. Hb oxygenation over the dorsum remains relatively constant prior to surgery. However, following surgery at the 60-h monitoring point the oxygenation of tissue at the central and distal sites drops dramatically. Following the initial drop in oxygenation after surgery the oxygenation of the central site remains relatively stable over the remaining 60 h of monitoring. Oxygenation at the distal site continues to drop after surgery. Tissue at and around this distal measurement site dies over the intervening 60 h of monitoring. Over the course of the 60-h postoperative monitoring period, oxygenation at the proximal site remains relatively unchanged from that measured at that site prior to surgery. While the ratio of concentrations of HbO_2 and Hb can provide a measure of tissue oxygenation, the sum of the concentrations provides a relative measure of tissue blood content (or blood volume). Figure 27 presents a time series of tHb images derived from NIR spectroscopic imaging sequences of the rat dorsum prior to surgical elevation of the flap and at 5, 60, 120 and 720 min following surgery. tHb images provide insight into the blood volume distribution in the skin flap. Prior to surgery the rat dorsum has a relatively uniform tHb distribution. Immediately following surgical elevation of the flap, a gradient in the tHb concentration exists along the length of the flap. tHb proximal to the vascular pedicle at the base of the flap is not significantly different from the tHb level of the surrounding uninvolved tissue of the rat dorsum. However, the tHb levels along the flap drop with increasing distance from the vascular pedicle. This blood volume (tHb) distribution persists in the flap for the duration of the experiment. These results are consistent with the tissue oxygenation findings. The NIR spectroscopic imaging and point spectroscopy clearly indicate that the central region of the flap remains ischemic and hypoxic following surgery but the tissue can generally cope with this level of underperfusion. The distal site shows a more dramatic drop in oxygenation and blood volume over the first 2 h following surgery and this site generally fails within the intervening 60 h. Figure 27 also shows the fractional change in tissue water

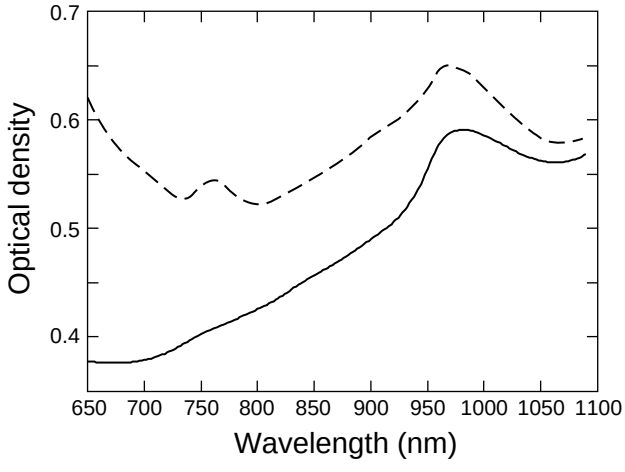

Figure 25 NIR attenuation spectra from a central site on the dorsum of a rat before (——) and after (- - - -) elevation of a skin flap.

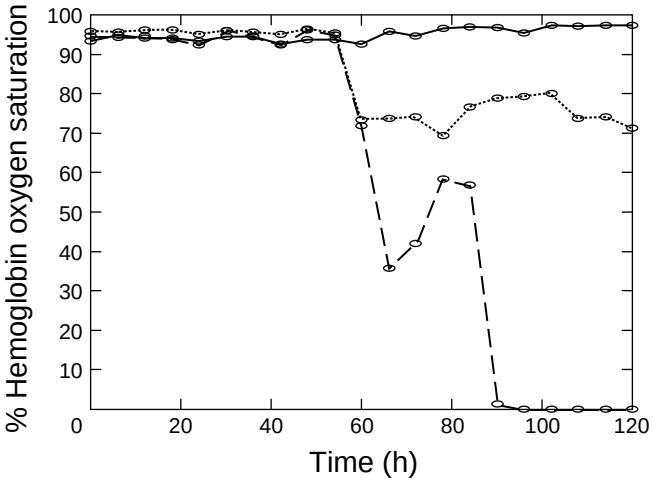

Figure 26 Change in percent Hb oxygenation at the proximal (——), central (· · · ·) and distal (- - - -) measurement sites on the rat dorsum measured over a 120 h monitoring period at 6 h intervals. Skin flap was elevated after 60 h of monitoring.

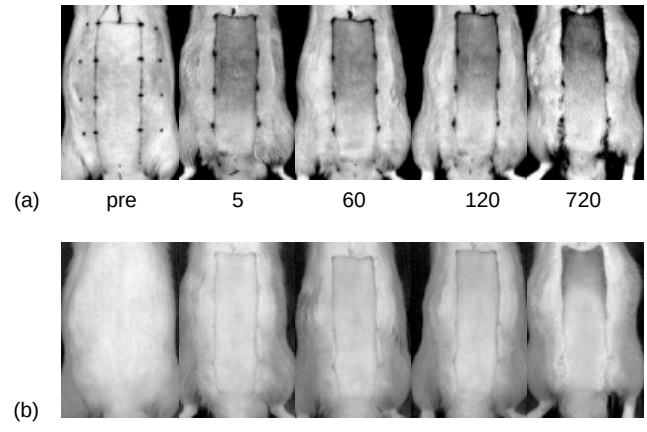

Figure 27 (a) Time series of total Hb images of a rat dorsum derived from spectroscopic imaging sequences taken immediately prior to elevation of the flap (0 min) and at 5, 60, 120 and 720 min after surgical elevation of the skin flap. Image intensity is proportional to the total Hb content in the tissue. (b) The corresponding time series of hydration images of the same flap. Image intensity is proportional to the water content of the tissue.

content over the same flap in which the tHb was reported. Following surgery the proximal site experiences a small increase in tissue water content. After 72 h, tissue water content at the proximal site returns to its pre-surgical level. The central site shows no significant change in hydration over the course of the experiment. Conversely, the water content at the distal site drops dramatically upon surgery and continues to drop over the first 12 h. The drop in blood volume or tHb, Hb oxygenation and tissue hydration are clear indicators of hypoperfusion of the distal sections of the flap. These indicators appear early in the post-operative period, well before there are overt clinical symptoms of tissue compromise. Figure 28 compares a NIR Hb oxygen image of a flap 1 h after surgery with a corresponding visual image and a visual image taken 72 h later. The regions of low oxygenation that are apparent in the 1-h oxygenation image correlate quite closely with the regions of the flap which go on to fail over the intervening 72 h. The 1-h visual image of the flap shows no sign of trouble.

NIR spectroscopy provides a nonsubjective means of classifying the changes in the optical properties of tissue which are characteristic of tissue compromise. Spectroscopic methods offer the promise of detecting and distinguishing these characteristic changes prior to the appearance of overt clinical signs of tissue health or failure. Thus, spectroscopic analysis potentially enables earlier detection of perfusion related complications thereby allowing prompt and effective intervention. NIR spectroscopic assessment of skin injury, cutaneous wounds and burns as well as surgically reconstructed areas of tissues should potentially be able to provide the clinician diagnostic information early in the clinical time course. Early and reliable diagnosis of perfusion and oxygenation related complications in tissue near the surface of the skin should improve clinical management of these tissues when compromised and thereby improve patient outcome.

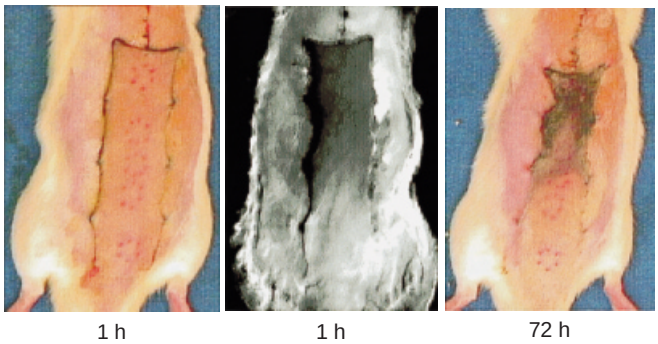

Figure 28 A comparison between the visible pictures of a reversed McFarlane flap taken at 1 and 72 h after surgery and a Hb oxygenation image of the same flap taken 1 h after surgery. The 1 h visible picture shows no sign of tissue compromise whereas the 1 h oxygenation image clearly shows that the distal end of the flap (upper half) is poorly oxygenated. The pattern in the 1 h oxygenation image resembles the outcome of the flap 72 h later.

6 SUMMARY AND OUTLOOK

NIR spectroscopy is being applied to diverse areas of physiology and medicine to study the dynamics of tissue and muscle oxygenation, mitochondrial oxygen utilization and edema or desiccation in living systems. Medical applications of the technique are developing rapidly and the clinical utility of the method will undoubtedly continue to grow in the future. However, as pointed out throughout this review, practitioners of in vivo NIR spectroscopy are faced with a number of tantalizing technical challenges and a myriad of confounding variables that can influence the measurements and complicate interpretation of the data. For instance, the goal of providing a quantitative and/or semiquantitative analysis on the chromophores present in the tissue, noninvasively using NIR spectroscopy remains illusive. In vivo measurement of mitochondrial oxygen utilization based on the differential NIR absorption of reduced and oxidized cytochrome *c* oxidase has also proven to be a difficult obstacle. Correlating changes in the bulk scattering properties of the tissue to morphological or structural changes in the tissue is hampered by the variety of confounding factors that influence light attenuation measurements made on living tissues. While there is little doubt that clinically useful information is, in principal, latent in the NIR attenuation spectrum of tissue, extracting this information reliably remains the major challenge in further expanding the clinical role of in vivo NIR spectroscopy.

ABBREVIATIONS AND ACRONYMS

AOTF	Acousto-optical Tunable Filter
CCD	Charge-coupled Device
Hb	Hemoglobin
ICG	Indocyanine Green
IR	Infrared
LCTF	Liquid Crystal Tunable Filter
LED	Light Emitting Diode
LOMEP	Linear Shape Optimized Maximum Entropy Linear Prediction
NIR	Near-infrared
PMT	Photomultiplier Tube
S/N	Signal-to-noise Ratio
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Fluorescence Spectroscopy In Vivo • Glucose, In Vivo Assay of • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Optical Coherence Tomography • Photodynamic Therapy

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry • Point-of-care Testing

Food **(Volume 5)**
Near-infrared Spectroscopy in Food Analysis

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Remote Sensing **(Volume 10)**
Hyperspectral Remote Sensing: Data Collection and Exploitation

Electronic Absorption and Luminescence **(Volume 12)**
Near-infrared Absorption/Luminescence Measurements

REFERENCES

1. V. Kumar, R.S. Cotran, S.L. Robbins, *Basic Pathology*, W. B. Saunders Co., Pennsylvania, 1992.
2. C.E.W. Hahn, 'Electrochemical Analysis of Clinical Blood-gases, Gases and Vapours', *Analyst*, **123**(6), 57R–86R (1998).
3. E.G. Chikirdin, 'Reviews: Developments in the Technology Basis of X-ray Diagnosis', *Biomed. Eng.*, **32**(1), 50–54 (1998).
4. W.G. Zijlstra, A. Buursma, W.P. Meeuwsen-vander Roest, 'Absorption Spectra of Human Fetal and Adult Oxyhemoglobin, De-oxyhemoglobin, Carboxyhemoglobin, and Methemoglobin', *Clin. Chem.*, **37**(9), 1633–1638 (1991).
5. R.N. Pittman, 'In Vivo Photometric Analysis of Hemoglobin', *Ann. Biomed. Eng.*, **14**, 119–137 (1986).
6. H. Liu, D.A. Boas, Y. Zhang, A.J. Yodh, B. Chance, 'Determination of Optical Properties and Blood Oxygenation in Tissue Using Continuous NIR Light', *Phys. Med. Biol.*, **40**, 1983–1993 (1995).
7. L.S.L. Arakaki, *An Optical Method for Myoglobin Oxygenation Measurements in the Blood-perfused Rat Hind Limb*, Doctoral Thesis, University of Washington, 1995.
8. O. Galkin, S. Buchter, A. Tabirian, A. Schulte, 'Pressure Effects on the Proximal Heme Pocket in Myoglobin Probed by Random and Near-infrared Absorption Spectroscopy', *Biophys. J.*, **73**(5), 2752–2764 (1997).
9. M. Cope, P. van der Zee, M. Essenpreis, S.R. Arridge, D.T. Delpy, 'Data Analysis for Near Infrared Spectroscopy of Tissue: Problems in Determining the Relative Cytochrome aa_3 Concentration', *Proc. SPIE*, **1431**, 251–262 (1991).
10. S. Wray, M. Cope, D.T. Delpy, J.S. Wyatt, E.O.R. Reynolds, 'Characterization of Near Infrared Absorption Spectra of Cytochrome aa_3 and Hemoglobin for the Noninvasive Monitoring of Cerebral Oxygenation', *Biochim. Biophys. Acta*, **933**, 184–192 (1988).
11. V. Quaresima, R. Springett, M. Cope, J.T. Wyatt, D.T. Delpy, M. Ferrari, C.E. Cooper, 'Oxidation and Reduction of Cytochrome Oxidase in the Neonatal Brain Observed by In Vivo Near-infrared Spectroscopy', *Biochim. Biophys. Acta*, **1366**, 291–300 (1998).
12. K.A. Martin, 'Direct Measurement of Moisture in Skin by NIR Spectroscopy', *J. Soc. Cosmet. Chem.*, **44**, 249–263 (1993).
13. K. Martin, 'In Vivo Measurements of Water in Skin by Near-infrared Reflectance', *Appl. Spectrosc.*, **52**, 1001 (1998).
14. P.L. Walling, J.M. Dabney, 'Moisture in Skin by Near-infrared Reflectance Spectroscopy', *J. Soc. Cosmet. Chem.*, **40**, 151–171 (1989).
15. W. Baumler, C. Abels, S. Karrer, T. Weiss, H. Messmann, M. Landthaler, R.M. Szeimies, 'Photo-oxidative Killing of Human Colonic Cancer Cells Using Indocyanine Green and Infrared Light', *Br. J. Cancer*, **80**(3–4), 360–363 (1999).
16. M.S. Patterson, E. Schwartz, B.C. Wislon, 'Quantitative Reflectance Spectrophotometry for the Noninvasive Measurement of Photosensitizer Concentration in Tissue During Photodynamic Therapy', *Proc. SPIE*, **1065**, 337–343 (1987).
17. J.D. Ingle, S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, New Jersey, 1988.
18. H. Curtis, N.S. Barnes, *Biology*, 5th edition, Worth Publishers, New York, 1989.
19. S.J. Matcher, C.E. Elwell, C.E. Cooper, M. Cope, D.T. Delpy, 'Performance Comparison of Several Published Tissue Near-infrared Spectroscopy Algorithms', *Anal. Biochem.*, **227**, 54–68 (1995).
20. L.S.L. Arakaki, M.J. Kushmerick, D.H. Burns, 'Myoglobin Oxygen Saturation Measured Independently of Hemoglobin in Scattering Media by Optical Reflectance Spectroscopy', *Appl. Spectrosc.*, **50**(6), 697–707 (1996).
21. F. Vernieri, N. Rosato, F. Pavri, F. Tibuzzi, F. Passavelli, P.M. Rossini, 'Near Infrared Spectroscopy and Transcranial Doppler in Monohemispheric Stroke', *Eur. Neurol.*, **41**(3), 159–162 (1999).
22. A. Villringer, J. Planck, J. Hock, L. Schleinkofer, U. Dirnagl, 'Near Infrared Spectroscopy (NIRS): A New Tool

to Study Hemodynamic Changes During Activation of Brain Function in Human Adults', *Neurosci. Lett.*, **154**, 101–104 (1993).
23. S.P. Gopinath, C.S. Robertson, R.G. Grossman, B. Chance, 'Near-infrared Spectroscopic Localization of Intracranial Hematomas', *J. Neurosurg.*, **79**(1), 43–47 (1993).
24. Y. Campos, M.A. Martin, T. Garcia-Silva, P. del Hoyo, J.C. Rubio, M. Castro-Gago, J. Garcia-Penas, J. Casas, A. Cabello, J.R. Ricoy, J. Arenas, 'Clinical Heterogeneity Associated with Mitochondrial DNA Depletion in Muscle', *Neuromuscul. Disord.*, **8**, 568–573 (1998).
25. F.O. Libnau, O.M. Kvalheim, A.A. Christy, J. Toft, 'Spectra of Water in the Near- and Mid-infrared Region', *Vib. Spectrosc.*, **7**, 243–254 (1994).
26. M.G. Sowa, J.R. Payette, H.H. Mantsch, 'Near Infrared Spectroscopic Assessment of Tissue Hydration Following Surgery', *J. Surg. Res.*, **86**, 62–69 (1999).
27. M. Cope, D.T. Delpy, S. Wray, J.S. Wyatt, E.O.R. Reynolds, 'A CCD Spectrophotometer to Quantitate the Concentration of Chromophores in Living Tissue Utilizing the Absorption Peak of Water at 975 nm', *Adv. Exp. Med. Biol.*, **248**, 33–40 (1989).
28. S.J. Matcher, M. Cope, D.T. Delpy, 'Use of Water Absorption Spectrum to Quantify Tissue Chromophores Concentration Changes in Near Infrared Spectroscopy', *Phys. Med. Biol.*, **39**(1), 177–196 (1994).
29. B.J. Berne, R. Percora, *Dynamic Light Scattering with Applications to Chemistry, Biology, and Physics*, Wiley-Interscience, New York, 1976.
30. W.F. Cheong, S.A. Prahl, A.J. Welch, 'A Review of the Optical Properties of Biological Tissues', *IEEE J. Quantum Electron.*, **26**(12), 2166–2185 (1990).
31. F. Faris, M. Thorniley, Y. Wickramasinghe, R. Houston, P. Rolfe, N. Livera, A. Spencer, 'Noninvasive In Vivo Near-infrared Optical Measurement of the Penetration Depth in the Neonatal Head', *Clin. Phys. Physiol. Meas.*, **12**(4), 353–358 (1991).
32. S.T. Flock, B.C. Wilson, M.S. Patterson, 'Total Attenuation Coefficients and Scattering Phase Functions of Tissues and Phantom Materials at 633 nm', *Med. Phys.*, **14**(5), 835–841 (1987).
33. S. Arridge, D.T. Delpy, 'The Theoretical Basis for the Determination of Optical Pathlengths in Tissue: Temporal and Frequency Analysis', *Phys. Med. Biol.*, **37**(7), 1531–1560 (1992).
34. S. Glasstone, M.C. Edlund, *The Elements of Nuclear Reactor Theory*, Van Nostrand, New York, 1952.
35. M.S. Patterson, B. Chance, B.C. Wilson, 'Time Resolved Reflectance and Transmittance for the Noninvasive Measurement of Tissue Optical Properties', *Appl. Opt.*, **28**(12), 2331–2336 (1989).
36. E.M. Sevick, B. Chance, L. Leigh, S. Nioka, M. Maris, 'Quantitation of Time and Frequency Resolved Optical Spectra for the Determination of Tissue Oxygenation', *Anal. Biochem.*, **195**(2), 330–351 (1991).
37. S. Fantini, M.A. Franceschini-Fantini, J.S. Maier, S.A. Walker, B. Barbieri, E. Gratton, 'Frequency-domain Multichannel Optical Detector for Noninvasive Spectroscopy and Oximeter', *Opt. Eng.*, **34**(1), 32–42 (1995).
38. M. Maris, A. Mayevsky, E. Sevick, B. Chance, 'Frequency Domain Measurements of Changes of Optical Pathlength During Spreading Depression in Rodent Brain Model', *Proc. SPIE*, **1431**, 136–148 (1991).
39. W. Herschel, 'Experiments of the Solar and Terrestrial Rays that Occasion Heat', *Philos. Trans. R. Soc. London*, **90**(XV), 293–326 (1800).
40. A.M.C. Davies, 'Subdivisions of the Infrared Region', *Appl. Spectrosc.*, **44**(4), 14A (1990).
41. G.A. Millikan, 'The Oximeter, an Instrument for Measuring Continuously the Oxygen Saturation of Arterial Blood in Man', *Rev. Sci. Instr.*, **13**, 434–444 (1942).
42. E.B. Merrick, T.J. Hayes, 'Continuous Noninvasive Measurement of Arterial Blood Oxygen Levels', *J. Hewlett Packard*, **28**, 2–9 (1976).
43. G.C. Holst, *CCD Arrays, Cameras, and Displays*, JCD Publishing, Winter Park, FL, 1998.
44. S. Kastle, F. Noller, S. Falk, A. Bukta, E. Mayer, D. Miller, 'A New Family of Sensors for Pulse Oximetry', *Hewlett-Packard Journal*, Article 7, 1–17 (1997).
45. J.A. Pologe, 'Pulse Oximetry: Technical Aspects of Machine Design', *Int. Anesthesiol. Clin.*, **25**, 137–175 (1987).
46. J.G. Webster (ed.), *Design of Pulse Oximeters*, Institute of Physics Publishing, Bristol, 1997.
47. L. Bernardi, S. Leuzzi, 'Laser Doppler Flowmetry and Photoplethysmography: Basic Principles and Hardware', in *Bioengineering of the Skin: Cutaneous Blood Flow and Erythema*, eds. E. Berardesca, P. Elsner, H.L. Maibach, CRC Press, Boca Raton, FL, Chapter 3, 1995.
48. J. Boddaert, D. Magula, J. Belmin, 'Diagnosis of Orthostatic Hypotension', *Lancet*, **352**(9141), 1705–1706 (1999).
49. L. Bernardi, L. Ricordi, P.L. Lazzari, P.L. Solda', A. Calciati, M.R. Ferrari, I. Vandea, G. Finardi, P. Fratino, 'Impaired Circadian Modulation of Sympathovagal Activity in Diabetes: A Possible Explaination for Altered Temporal Onset of Cardiovascular Disease', *Circulation*, **86**, 1443 (1992).
50. E. Heron, D. Chemla, J.L. Megnien, J.C. Pourny, J. Levenson, Y. Lecarpentier, A. Simon, 'Reactive Hyperemia Unmasks Reduced Compliance of Cutaneous Arteries in Essential Hypertension', *J. Appl. Physiol.*, **79**(2), 498–505 (1995).
51. T. Ahrens, K. Rutherford, *Essentials of Oxygenation*, Jones and Bartlett, Boston, MA, 1993.

52. A. Fronek, 'Photoplethysmography in the Diagnosis of Venous Disease', *Dermatol. Surg.*, **21**(1), 64–66 (1995).
53. C.E. Bird, M.H. Criqui, A. Fronek, J.O. Denenberg, M.R. Klauber, R.D. Langer, 'Quantitative and Qualitative Progression of Peripheral Arterial Disease by Noninvasive Testing', *Vasc. Med.*, **4**(1), 15–21 (1999).
54. S.A. Carter, R.B. Tate, 'Value of Toe Pulse Waves in Addition to Systolic Pressures in the Assessment of the Severity of Peripheral Arterial Disease and Critical Limb Ischemia', *J. Vasc. Surg.*, **24**(2), 258–265 (1996).
55. M. Kohl, C. Nolte, H.R. Heekeren, S. Horst, U. Scholz, H. Obrig, A. Villringer, 'Determination of the Wavelength Dependence of the Differential Pathlength Factor from Near-infrared Pulse Signals', *Phys. Med. Biol.*, **43**, 1771–1782 (1998).
56. C.E. Elwell, H. Owen-Reece, J.S. Wyatt, M. Cope, E.O.R. Reynolds, D.T. Delpy, 'Influence of Respiration and Changes in Expiratory Pressure on Cerebral Haemoglobin Concentration Measured by Near Infrared Spectroscopy', *J. Cereb. Blood Flow Metab.*, **16**, 353–357 (1996).
57. T. Shiga, E. Chihara, K. Tanabe, Y. Tanaka, K. Yamamoto, 'Real Time Monitoring of Pulsatile Change in Hemoglobin Concentration of Cerebral Tissue by a Portable Tissue Oximeter with a 10-Hz Sampling Rate', *SPIE*, **3194**, 435–442 (1997).
58. E.M. Landis, 'Micro-injection Studies of Capillary Blood Pressure in Human Skin', *Heart*, **15**, 15–29 (1930).
59. M. Kosiak, 'Etiology and Pathology of Ischemic Ulcers', *Arch. Med. Phys. Rehabil.*, **2**, 62–69 (1959).
60. J.R. Haalboom, J. den Boer, E. Buskens, 'Risk-assessment Tools in the Prevention of Pressure Ulcers', *Ostomy Wound Management*, **45**(2), 20–26, 28, 30–34 (1999).
61. S. Jaywant, B. Wilson, M. Patterson, L. Lilge, T. Flotte, J. Woolsey, C. McCulloch, 'Temperature Dependent Changes in the Optical Absorption and Scattering Spectra of Tissues: Correlation with Ultrastructure', *Proc. SPIE*, **1882**, 218–228 (1993).
62. H. Maeda, Y. Ozaki, M. Tanaka, N. Hayashi, T. Kojima, 'Near Infrared Spectroscopy and Chemometrics Studies of Temperature-dependent Spectral Variations of Water: Relationship between Spectral Changes and Hydrogen Bonds', *J. Near Infrared Spectrosc.*, **3**, 191 (1995).
63. J. Lin, C.W. Brown, 'Near-IR Fiber-optic Temperature Sensor', *Appl. Spectrosc.*, **47**, 62–68 (1993).
64. A. Hicks, S. McGill, R.L. Hughson, 'Tissue Oxygenation by Near-infrared Spectroscopy and Muscle Blood Flow During Isometric Contractions of the Forearm', *Can. J. Appl. Physiol.*, **24**, 216–230 (1999).
65. S. Homma, A. Kagaya, 'Detection of Oxygen Consumption in Different Forearm Muscles During Handgrip Exercise by Spatially Resolved NIR Spectroscopy', *Adv. Exp. Med. Biol.*, **428**, 327–332 (1997).
66. C.W. Yoxall, A.M. Weindling, 'Measurement of Venous Oxyhaemoglobin Saturation in the Adult Human Forearm by Near Infrared Spectroscopy with Venous Occlusion', *Med. Biol. Eng. Comput.*, **35**(4), 331–336 (1997).
67. R.A. De Blasi, M. Ferrari, A. Natali, G. Conti, A. Mega, A. Gasparetto, 'Noninvasive Measurement of Forearm Blood Flow and Oxygen Consumption by Near Infrared Spectroscopy', *J. Appl. Physiol.*, **76**, 1388–1393 (1994).
68. R.A. De Blasi, E. Quaglia, M. Ferrari, 'Skeletal Muscle Oxygenation Monitoring by Near Infrared Spectroscopy', *Biochem. Int.*, **25**, 241–248 (1991).
69. R.A. De Blasi, M. Cope, C. Elwell, F. Safoue, M. Ferrari, 'Noninvasive Measurement of Human Forearm Oxygen Consumption by Near Infrared Spectroscopy', *Eur. J. Appl. Physiol.*, **67**, 20–25 (1993).
70. R.A. De Blasi, M. Cope, M. Ferrari, 'Oxygen Consumption of Human Skeletal Muscle by Near Infrared Spectroscopy During Tourniquet-induced Ischemia in Maximal Voluntary Contraction', *Adv. Exp. Med. Biol.*, **317**, 771–777 (1992).
71. N.B. Hampson, C.A. Piantadosi, 'Near Infrared Monitoring of Human Skeletal Muscle Oxygenation During Forearm Ischemia', *J. Appl. Physiol.*, **64**, 2449–2457 (1988).
72. M.G. Sowa, J.R. Mansfield, G.B. Scarth, H.H. Mantsch, 'Noninvasive Assessment of Regional and Temporal Variations in Tissue Oxygenation by Near-infrared Spectroscopy and Imaging', *Appl. Spectrosc.*, **51**, 143–152 (1997).
73. J.R. Mansfield, M.G. Sowa, G.B. Scarth, R.L. Somorjai, H.H. Mantsch, 'Fuzzy C-means Clustering and Principal Component Analysis of Time Series from Near-infrared Imaging of Forearm Ischemia', *J. Comput. Med. Image. Graph.*, **21**, 299–308 (1997).
74. H.J. Granger, 'Autoregulation of Tissue Perfusion and Oxygenation', in *Tissue Perfusion and Organ Function: Ischemia/Reperfusion Injury*, eds. T. Kamada, T. Shiga, R.S. McCuskey, Elsevier, Amsterdam, 29–45, 1996.
75. B. Grassi, V. Quaresima, C. Marconi, M. Ferrari, P. Cerretelli, 'Blood Lactate Accumulation and Muscle Deoxygenation During Incremental Exercise', *J. Appl. Physiol.*, **87**, 348–355 (1999).
76. R. Boushel, F. Pott, P. Madsen, G. Radegran, M. Nowak, B. Quistorff, N. Secher, 'Muscle Metabolism from Near Infrared Spectroscopy During Rhythmic Handgrip in Humans', *Eur. J. Appl. Physiol.*, **79**, 41–48 (1998).
77. T.-K. Tran, N. Sailasuta, U. Kreutzer, R. Hurd, Y. Chung, P. Mole, S. Kuna, T. Jue, 'Comparative Analysis of NMR and NIRS Measurements of Intercellular PO_2 in Human Skeletal Muscle', *Am. J. Physiol.*, **276**, R1682–R1690 (1999).
78. Y. Bhambhani, S. Buckley, T. Susaki, 'Muscle Oxygenation Trends During Constant Work Rate Cycle Exercise in Men and Women', *Med. Sci. Sports Exerc.*, **31**, 90–98 (1999).

79. V.V. Kupriyanov, R.A. Shaw, B. Xiang, H. Mantsch, R. Deslauriers, 'Oxygen Regulation of Energy Metabolism in Isolated Pig Hearts: A Near-IR Spectroscopy Study', *J. Mol. Cell Cardiol.*, **29**, 2431–2439 (1997).
80. M.S. Thorniley, A. Lahiri, B. Glenville, C. Shurey, G. Baker, U. Ravel, J. Crawley, C.J. Green, 'Noninvasive Measurement of Cardiac Oxygenation and Haemodynamics During Transient Episodes of Coronary Artery Occlusion and Reperfusion in the Pig', *Clin. Sci.*, **91**, 51–58 (1996).
81. B.D. Groot, C.J. Zuurbier, J.H.G.M. van-Beek, 'Dynamics of Tissue Oxygenation in Isolated Rabbit Heart as Measured with Near-infrared Spectroscopy', *Am. J. Physiol.*, **45**, H1616–H1624 (1999).
82. B.R. Soller, R.H. Micheels, J. Coen, B. Parikh, L. Chu, C. Hsi, 'Feasibility of Noninvasive Measurement of Tissue pH Using Near-infrared Reflectance Spectroscopy', *J. Clin. Mon.*, **12**, 387–395 (1996).
83. B.F. Becker, C. Kupatt, P. Raschke, S. Zahler, 'Microcirculatory Disturbances due to Activated Leukocytes in Ischemic and Reperfusion Injury', in *Tissue Perfusion and Organ Function: Ischemia/Reperfusion Injury*, eds. T. Kamada, T. Shiga, R.S. McCuskey, Elsevier, Amsterdam, 97–107, 1996.
84. R. Salcido, J.C. Donofrio, S.B. Fisher, E.K. Legrand, J.M. Carney, R. Schosser, J. Rodgers, R. Liang, 'Evaluation of Ibuprofen for Pressure Ulcer Prevention: Application of a Rat Pressure Ulcer Model', *Adv. Wound Care*, **8**(4), 30–32, 34, 38–40 (1995).
85. Y. Nomura, A. Matsunaga, M. Tanura, 'Optical Characterization of Heme $a + a_3$ and Copper of Cytochrome Oxidase in Blood-free Perfused Rat Brain', *J. Neurosci. Meth.*, **82**, 135–144 (1998).
86. A. Matsunaga, Y. Nomura, S. Kuroda, M. Tanura, J. Nishihira, N. Yoshimura, 'Energy-dependent Redox State of Heme $a + a_3$ and Copper of Cytochrome Oxidase in Perfused Rat Brain in situ', *Am. J. Physiol.*, **275**, C1022–C1030 (1998).
87. J.K. Kauppinen, D.J. Moffatt, H.H. Mantsch, D.G. Cameron, 'Fourier Self Deconvolution: A Method for Resolving Intrinsically Overlapped Bands', *Appl. Spectrosc.*, **35**, 271–276 (1981).
88. O.-P. Sievanen, 'Estimating Optimum Filter Length in Linear Prediction', *Appl. Spectrosc.*, **51**(5), 718–720 (1997).
89. J.K. Kauppinen, D.J. Moffatt, H.H. Mantsch, 'Nonlinearity of the Maximum Entropy Method in Resolution Enhancement', *Can. J. Chem.*, **70**(12), 2887–2894 (1992).
90. S.J. Matcher, C.E. Cooper, 'Absolute Quantification of Deoxyhaemoglobin Concentration in Tissue Near Infrared Spectroscopy', *Phys. Med. Biol.*, **39**, 1295–1312 (1994).
91. C.E. Cooper, C.E. Elwell, J.H. Meek, S.J. Matcher, J.S. Wyatt, M. Cope, D.T. Delpy, 'The Noninvasive Measurement of Absolute Cerebral Deoxyhemoglobin Concentration and Mean Optical Pathlength in the Neonatal Brain by Second Derivative Near Infrared Spectroscopy', *Pediatr. Res.*, **39**(1), 32–38 (1996).
92. B. Chance, M. Cope, E. Gratton, N. Ramanujam, B. Tromberg, 'Review Article: Phase Measurement of Light Absorption and Scattering in Human Tissue', *Rev. Sci. Instrum.*, **69**(10), 3457–3481 (1998).
93. M. Oda, Y. Yamashita, T. Nakano, A. Suzuki, K. Shimizu, I. Hirano, F. Shimonura, E. Ohmae, T. Suzuki, Y. Tsuchiya, 'Near-infrared Time Resolved Spectroscopy System for Tissue Oxygenation Monitor', *Proc. SPIE*, **3597**, 611–617 (1999).
94. S. Suzuki, S. Takasaki, T. Ozaki, Y. Kobayashi, 'A Tissue Oxygenation Monitor Using NIR Spatially Resolved Spectroscopy', *Proc. SPIE*, **3597**, 582–592 (1999).
95. P.L. Madsen, N.H. Secher, 'Near-infrared Oximetry of the Brain', *Prog. Neurobiol.*, **58**(6), 541–560 (1999).
96. J.S. Soul, A.J. duPlessis, 'New Technologies in Pediatric Neurology. Near-infrared Spectroscopy', *Semin. Pediatr. Neurol.*, **6**(2), 101–110 (1999).
97. G. Nollert, T. Shin'oka, R.A. Jonas, 'Near-infrared Spectrophotometry of the Brain in Cardiovascular Surgery', *Thorac. Cardiovasc. Surg.*, **46**, 167–175 (1998).
98. H. Owen-Reece, M. Smith, C.E. Elwell, J.C. Goldstone, 'Near Infrared Spectroscopy', *Br. J. Anaesth.*, **82**, 418–426 (1999).
99. A. Villringer, B. Chance, 'Noninvasive Optical Spectroscopy and Imaging of Human Brain Function', *Trends Neurosci.*, **20**, 435–442 (1997).
100. T. Kato, A. Kamei, S. Takashima, T. Ozaki, 'Human Visual Cortical Function During Photic Stimulation Monitoring by Means of Near-infrared Spectroscopy', *J. Cereb. Blood Flow Metab.*, **13**, 516–520 (1993).
101. A. Kleinschmidt, H. Obrig, M. Requardt, K.-D. Merboldt, U. Dirnagl, A. Villringer, J. Frahm, 'Simultaneous Recording of Cerebral Blood Oxygenation Changes During Human Brain Activation by Magnetic Resonance Imaging and Near-infrared Spectroscopy', *J. Cereb. Blood Flow Metab.*, **16**, 817–826 (1996).
102. H.R. Heekeren, M. Kohl, H. Obrig, R. Wenzel, W. von Pannwitz, S.J. Matcher, U. Dirnagl, C.E. Cooper, A. Villringer, 'Noninvasive Assessment of Changes in Cytochrome *c* Oxidase Oxidation in Human Subjects During Visual Stimulation', *J. Cereb. Blood Flow Metab.*, **19**, 592–603 (1999).
103. B. Urlesberger, A. Kaspirek, G. Pichler, W. Muller, 'Apnoea of Prematurity and Changes in Cerebral Oxygenation and Cerebral Blood Volume', *Neuropediatrics*, **30**, 29–33 (1999).
104. F.F. Buchvald, K. Kesje, G. Greisen, 'Measurement of Cerebral Oxyhaemoglobin Saturation and Jugular Blood Flow in Term Healthy Newborn Infants by Near-infrared Spectroscopy and Jugular Venous Occlusion', *Biol. Neonate*, **75**, 97–103 (1999).
105. V. Dietz, M. Wolf, M. Keel, K.V. Siebeuthal, O. Baenziger, H. Bucher, 'CO_2 Reactivity of the Cerebral

Hemoglobin Concentration in Healthy Term Newborns Measured by Near Infrared Spectrophotometry', *Biol. Neonate*, **75**, 85–90 (1999).

106. P.E. Daubeney, D.C. Smith, S.N. Pilkington, R.K. Lamb, J.L. Monro, V.T. Tsang, S.A. Livesey, S.A. Webber, 'Cerebral Oxygenation During Paediatric Cardiac Surgery: Identification of Vulnerable Periods Using Near Infrared Spectroscopy', *Eur. J. Cardiothorac. Surg.*, **13**, 370–377 (1998).
107. A.J. McCleary, S. Gower, J.P. McGoldrick, J. Berridge, M.J. Gough, 'Does Hypothermia Prevent Cerebral Ischaemia During Cardiopulmonary Bypass?', *Cardiovasc. Surg.*, **7**, 425–431 (1999).
108. Y. Kadoi, F. Kawahara, S. Saito, T. Morita, F. Kunimoto, F. Goto, N. Fujita, 'Effects of Hypothermic and Normothermic Cardiopulmonary Bypass on Brain Oxygenation', *Ann. Thorac. Surg.*, **68**, 34–39 (1999).
109. I.G. Roberts, P. Fallon, F. Kirkham, P.M. Kirshbom, C.E. Cooper, M.J. Elliott, A.D. Edwards, 'Measurement of Cerebral Blood Flow During Cardiopulmonary Bypass with Near-infrared Spectroscopy', *J. Thorac. Cardiovasc. Surg.*, **115**, 94–102 (1998).
110. W.M. Kuebler, A. Sckell, O. Habler, M. Kleen, G.E.H. Kuhnle, M. Welte, K. Messmer, A.E. Goetz, 'Noninvasive Measurement of Regional Cerebral Blood Flow by Near-infrared Spectroscopy and Indocyanine Green', *J. Cereb. Blood Flow Metab.*, **18**, 445–456 (1998).
111. J. Patel, K. Marks, I. Roberts, D. Azzopardi, A.D. Edwards, 'Measurement of Cerebral Blood Flow in Newborn Infants Using Near Infrared Spectroscopy with Indocyanine Green', *Pediatr. Res.*, **43**, 34–39 (1998).
112. I.G. Roberts, P. Fallon, F.J. Kirkham, A. Lloyd-Thomas, C.E. Cooper, A.D. Edwards, M.J. Elliott, 'Estimation of Cerebral Blood Flow with Near Infrared Spectroscopy and Indocyanine Green', *Lancet*, **342**, 1425 (1993).
113. H.M. Kooijman, M.T.E. Hopman, W.N.J.M. Colier, J.A. van der Vliet, B. Oeseburg, 'Near Infrared Spectroscopy for Noninvasive Assessment of Claudication', *J. Surg. Res.*, **72**, 1–7 (1997).
114. T. Komiyama, H. Shigematsu, H. Yasuhara, T. Muto, 'An Objective Assessment of Intermittent Claudication by Near Infrared Spectroscopy', *Eur. J. Vasc. Surg.*, **8**, 294 (1994).
115. T.R. Cheatle, L.A. Potter, M. Cope, D.T. Delphy, P.D. Coleridge-Smith, J.H. Scurr, 'Near Infrared Spectroscopy in Peripheral Vascular Disease', *Br. J. Surg.*, **78**, 405–408 (1991).
116. P. Hickman, D.K. Harrison, S.D. Evans, J.J.F. Belch, P.T. McCollum, 'Use of Lightguide Reflectance Spectrophotometry in the Assessment of Peripheral Arterial Disease', *Adv. Exp. Med. Biol.*, **345**, 885 (1994).
117. M.S. Irwin, M.S. Thorniley, C.J. Dore, C.J. Green, 'Near Infrared Spectroscopy: A Noninvasive Monitor of Perfusion and Oxygenation within the Microcirculation of Limbs and Flaps', *Br. J. Plast. Surg.*, **48**, 14–22 (1995).
118. R.E. Hayden, M.A. Tavill, S. Nioka, T. Kitai, B. Chance, 'Oxygenation and Blood Volume Changes in Flaps According to Near-infrared Spectrophotometry', *Arch. Otolaryngol. Head Neck Surg.*, **122**, 1347–1351 (1996).
119. M.S. Thorniley, S.J. Sinclair, N.J. Barnett, C.B. Shurey, C.J. Green, 'The Use of Near Infrared Spectroscopy for Assessing Flap Viability During Reconstructive Surgery', *Br. J. Plast. Surg.*, **51**, 218–226 (1998).
120. M.F. Stranc, M.G. Sowa, B. Abdulrauf, H.H. Mantsch, 'Assessment of Tissue Viability Using Near Infrared Spectroscopy', *Br. J. Plast. Surg.*, **51**, 210–217 (1998).
121. Y. Tosa, W.P. Lee, N. Kollias, M.A. Randolph, J.W. May, Jr, 'Monoclonal Antibody to Intercellular Adhesion Molecule 1 Protects Skin Flaps Against Ischemia–Reperfusion Injury: An Experimental Study in Rats', *Plast. Reconstr. Surg.*, **101**, 1586–1594 (1998).
122. Y. Tosa, N. Kollias, W.P. Lee, J.W. May, Jr, 'Reduction of Ischemia–Reperfusion Injury by Monoclonal Antibody to Intercellular Adhesion Molecule-1', *Transplant. Proc.*, **28**(3), 1210–1211 (1996).
123. M.G. Sowa, J.R. Payette, M.F. Stranc, B. Abdulrauf, M.D. Hewko, J.R. Mansfield, H.H. Mantsch, 'Assessment of Tissue Viability by Near-IR Spectroscopy and Imaging', *SPIE*, **3257**, 199–207 (1998).
124. J.R. Payette, M.G. Sowa, S.L. Germscheid, M.F. Stranc, B. Abdulrauf, H.H. Mantsch, 'Noninvasive Diagnostics: Predicting Flap Viability with Near-IR Spectroscopy and Imaging', *Amer. Clin. Lab.*, **18**, 4–6 (1999).
125. J.R. Mansfield, M.G. Sowa, J.R. Payette, B. Abdulrauf, M.F. Stranc, H.H. Mantsch, 'Tissue Viability by Multispectral Near Infrared Imaging: A Fuzzy C-means Clustering Analysis', *IEEE Trans. Med. Imaging*, **17**, 1011–1016 (1998).
126. M.G. Sowa, J.R. Payette, M.D. Hewko, H.H. Mantsch, 'Visible–Near Infrared Multispectral Imaging of the Rat Dorsal Skin Flap', *J. Biomed. Opt.*, **4**, 474–481 (1999).
127. M.S. Thorniley, S. Simpkin, N.J. Barnett, P. Wall, K.S. Khaw, C.B. Shurey, S.J. Sinclair, C.J. Green, 'Applications of NIRS for Measurements of Tissue Oxygenation and Haemodynamics During Surgery', *Adv. Exp. Med. Biol.*, **411**, 481–493 (1997).

Optical Coherence Tomography

A.F. Fercher
University of Vienna, Austria

1 Introduction 1
2 Physical Basis 3
2.1 Backscattering and Diffraction Tomography 3
2.2 Time-domain and Fourier-domain Optical Coherence Tomography 5
2.3 Doppler Optical Coherence Tomography 8
2.4 Spectroscopic Optical Coherence Tomography 8
2.5 Resolution 8
3 Optical Coherence Tomography Schemes 9
3.1 Reflectometer Optical Coherence Tomography 9
3.2 Dual Beam Optical Coherence Tomography 9
3.3 Enface Optical Coherence Tomography 10
3.4 Dynamic Coherent Focus Optical Coherence Tomography 10
3.5 Doppler Optical Coherence Tomography 11
3.6 Endoscopic Optical Coherence Tomography 12
4 Implementation 13
4.1 Light Sources 13
4.2 Optical Delay Lines 13
4.3 Optical Detectors 14
5 Applications 14
5.1 Ophthalmology 14
5.2 Other Medical Fields and Optical Coherence Tomography Biopsy Imaging 16
5.3 Future Developments 17
Acknowledgments 17
Abbreviations and Acronyms 17
Related Articles 17
References 18

Optical coherence tomography (OCT) is a technique with which to peer inside a body noninvasively. Tissue structure defined by tissue absorption and scattering coefficients, and the speed of blood flow, are derived from the characteristics of light remitted by the body. Singly backscattered light detected by partial coherence interferometry (PCI) is used to synthesize the tomographic image coded in false colors. A prerequisite of this technique is a low time-coherent but high space-coherent light source, usually a superluminescent diode or a multimode semiconductor laser. This limits the wavelengths to the red and near-infrared (NIR) region from about 600 nm to 1300 nm, the so-called therapeutic window, where absorption ($\mu_a \approx 0.01\,mm^{-1}$) is small enough. Transverse resolution in OCT is diffraction limited, as in conventional imaging; depth resolution is limited by the coherence length of the light. Both figures are of the order of micrometers. Velocity resolution is of the order of $0.1\,mm\,s^{-1}$. A few instruments are commercially available. At present, OCT is mainly used in the medical field, in particular in ophthalmology. Owing to the high transmissivity of ocular media, the depth penetration is considerable. Corresponding applications in dermatology are somewhat hindered by the strong scattering of epidermic tissue ($\mu_s \approx 10^2\,mm^{-1}$). As OCT provides images with a resolution comparable to conventional histology, but in real time, it can be used as a biopsy technique in a wide range of biological systems to detect diseases. These include the tomographic imaging of the internal microstructure of in vitro atherosclerotic plaques, the tomographic real-time diagnostics for intraoperative monitoring, and in microsurgical intervention. Optical biopsy based on OCT also provides diagnostic information by differentiating the architectural morphology of urological tissue, gastrointestinal tissue, and respiratory tissue.

1 INTRODUCTION

Tomographic imaging techniques derive a slice image from the internal structure of a three-dimensional object. Tomographic imaging has a long association with X-ray diagnostic radiology. Early X-ray techniques used a moving X-ray source, together with a detector that moved so that only the selected plane of interest was imaged sharply. Out-of-focus planes were blurred. Modern tomographic techniques, such as computed tomography (CT), use mathematical techniques to generate the tomographic image from a series of discrete projections.

Tomographic imaging is performed with various types of radiation. The approach used to derive the slice image depends largely on the type of interaction of the radiation with the object. If X-rays or γ-rays are used as probes, straight-ray propagation can be assumed. Then the well-known projection technique (still commonly used in CT scanners but not so common in nuclear magnetic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

resonance (NMR) scanners) can be used to obtain the tomographic picture from straight-ray projections. Absorption and scattering lead to severe attenuation and diffusion of the radiation and the applicability of this concept is limited. Furthermore, if the structural elements to be resolved are of the order of the radiation wavelength, diffraction has to be taken into account, at least in principle.

Optical tomography techniques have been developed intensively since the early 1980s.[1] In optical tomography, tissue absorption and scattering coefficients are derived from the characteristics of light scattered and/or diffracted by the object. Three basic optical tomography approaches have been developed: (1) diffraction tomography,[2] reconstructing the object structure from the corresponding diffraction pattern; (2) diffuse optical tomography,[3] deriving spatial maps of absorption and scattering coefficients from the characteristics of multiply scattered light transmitted through the body; and (3) OCT,[4,5] where these data are derived from backscattered light detected by PCI. Instruments based on diffuse optical tomography are under development. Some instruments based on OCT are commercially available.

Optical tomography has, on the one hand, the important properties of being nonionizing and, at visible and NIR wavelengths, being strongly influenced by tissue chromophores and thus being related to the health status of the tissue. On the other hand, biological tissues are strongly absorbing and scattering media, and the penetration depth of optical radiation is small. Only at the spectral region from about 600 nm to 1300 nm are the absorption and scattering small enough to form the so-called therapeutic window[6] (Figure 1), which can also be used for OCT. Despite the rather small absorption coefficient of soft tissue ($\mu_a \approx 0.01\,\mathrm{mm}^{-1}$) scattering is high ($\mu_s \approx 10^2\,\mathrm{mm}^{-1}$) and therefore photons mainly propagate in a diffusive mode.

Light transmitted through tissue comprises ballistic, snake, and diffusive photons (Figure 2).[9] Ballistic photons are unscattered photons or photons strictly forwardscattered. Snake photons are scattered along a path close to a straight line. Ballistic and snake photons obey Beer's law. Diffusive photons propagate in a random mode and obey the radiative transport equation.

Ballistic and snake photons of a light-pulse-transmitting tissue arrive at the exit surface first. These forwardpropagating photons have a more or less deterministic path and, therefore, carry image information. Most optical forwardscattering tomography techniques under development rely on these image-bearing photons. Several gating techniques have been developed to isolate these photons from the usually more numerous diffuse photons. For example, a high-speed opto-electronic shutter

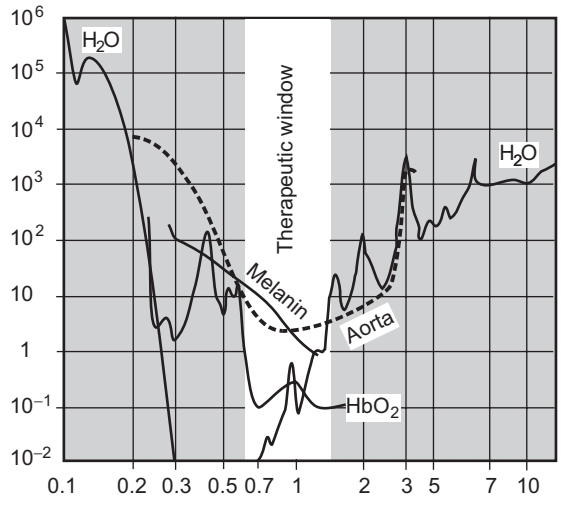

Figure 1 Spectral absorption of some tissue molecules: water (H_2O), oxyhemoglobin (HbO_2), melanin, and aortic tissue (dashed line; postmortem). The therapeutic window extends approximately from 600 to 1300 nm. H_2O and aortic tissue shown as the linear absorption coefficient ($1\,\mathrm{cm}^{-1}$); HbO_2 and melanin shown as the molar extinction coefficient ($10^3\,\mathrm{mol}^{-1}\,\mathrm{cm}^{-1}$). (Adapted from J.-L. Boulnois[7] and H.-P. Berlien et al.[8])

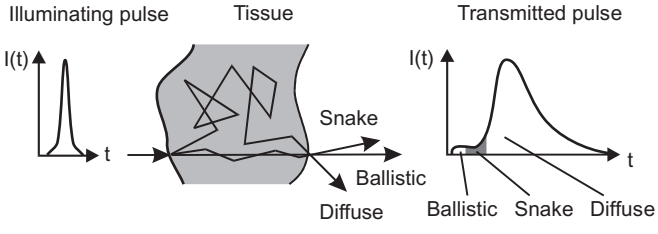

Figure 2 Light transmission in scattering tissue can be described in terms of ballistic, snake, and diffusive photon fractions, experiencing different modes of scattering and corresponding time delays of the transmitted pulse.

can be used to separate the early ballistic and snake photons and to block the diffusive photons. Other examples are Fourier gates, coherence gates, polarization gates, nonlinear optical effects, and holographic techniques.[10]

Unscattered photons can be used to obtain straight-ray projections. The mathematical problem of reconstructing a function from its straight-ray projections was first presented by J. Radon in 1917.[11] Its solution, the Fourier slice theorem, shows that some three-dimensional Fourier data from an object can be obtained from a two-dimensional Fourier transform of the projection. Because of its similarity to the Fourier diffraction theorem,[1] a closer look at this theorem is merited. From the Fourier transform (Equation 1) of the object function $F(x, y, z)$ (which, for example, in X-ray CT characterizes the two-dimensional distribution of the linear X-ray attenuation

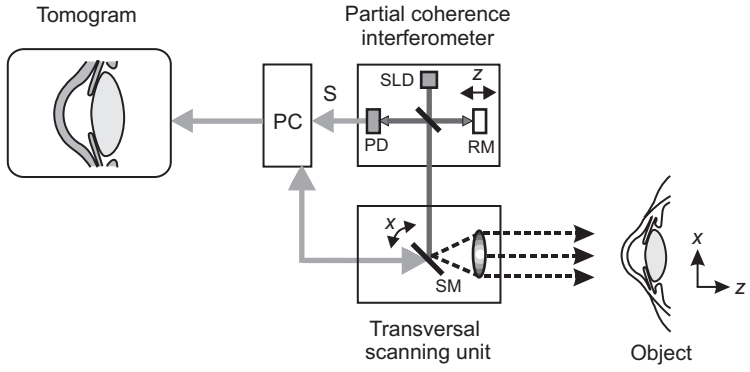

Figure 3 OCT arrangement. A partial coherence interferometer is used to perform optical A-scans along adjacent paths for a series of different x positions at the object. The tomographic image is synthesized from backscattered light data obtained with the interferometer.

coefficient)

$$\widehat{F}(u, v, w) = \mathrm{FT}\{F(x, y, z)\} = \int\int\int F(x, y, z) \times \exp[-2\pi i(ux + vy + wz)]\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \quad (1)$$

it follows readily that the projection $P(x, y) = \int F(x, y, z)\,\mathrm{d}z$ has the two-dimensional Fourier transform of Equation (2):

$$\mathrm{FT}_{x,y}\{P(x, y)\} = \int\int \left\{ \int F(x, y, z)\,\mathrm{d}z \right\} \exp[-2\pi i \times (ux + vy)]\,\mathrm{d}x\,\mathrm{d}y = \widehat{F}(u, v, 0) \quad (2)$$

Hence, some of the Fourier data of the object ($\widehat{F}(u, v, 0)$) can be obtained from a Fourier transform of projections. In the CT technique, a series of such projections at different directions is used to obtain depth resolution. A filtering step is applied to correct for the radial dependence of the Fourier data density introduced by the projection procedure.[1]

Diffuse photons do not propagate along straight rays. Several photon-transport models based on radiative transfer theory have been developed. A series of image-reconstruction algorithms has also been developed.[12] However, image-reconstruction techniques for optical imaging through diffuse media remain at a preliminary stage of development.[13]

OCT uses backscattered ballistic and snake photons. At present, its main application is in the field of medicine, where it yields high-quality images in a series of applications. However, the thickness of normal human organs does not permit the use of transmitted ballistic photons or transmitted snake photons. OCT is then limited to highly transparent objects, such as eye tissue and mucosal tissues, and uses light scattered backwards from object. OCT[14,15] uses PCI[16] to obtain so-called optical A-scans (a term borrowed from ultrasound technology). Optical A-scans obtained at a series of lateral locations are used to synthesize the tomographic image, as indicated in Figure 3.

2 PHYSICAL BASIS

2.1 Backscattering and Diffraction Tomography

In OCT the object is illuminated by a focused partially or low-coherent wave. In the context of OCT, the terms low coherence and partial coherence mean low time coherence but high space coherence. Low coherent light beams are usually obtained from multimode laser diodes and superluminescent diodes. These light sources emit Gaussian beams in the fundamental transverse mode. The minimal transverse radius of a focused Gaussian beam occurs at the beam waist and is given by Equation (3):

$$w_0 = \frac{\lambda}{\pi\vartheta} \quad (3)$$

At this radius the beam intensity falls to $1/\mathrm{e}^2$ of its central value.[17] The variable ϑ is the asymptotic angle of beam divergence (or aperture). A characteristic length in beam propagation direction is the confocal beam parameter or Rayleigh length z_0. At this distance the transverse beam radius increases to $\sqrt{2}w_0$ (Equation 4):

$$z_0 = \frac{\pi w_0^2}{\lambda} \quad (4)$$

Figure 4 depicts a Gaussian beam illuminating the anterior part of an eye positioned at the beam waist.

For simplification of the description the object depth is assumed of the order of the Rayleigh length of the illuminating beam. Then the illuminating light source wave has an approximately plane wavefront with wave

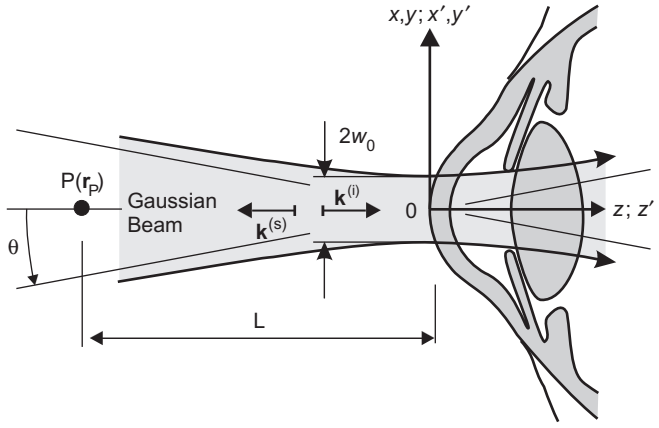

Figure 4 Backscattering. The object is illuminated by a focused Gaussian light beam with wave vector $\mathbf{k}^{(i)}$ and the beam waist at $z = 0$; the Gaussian beam is shown by its $1/e^2$ intensity contours. Backscattered light (wave vector $\mathbf{k}^{(s)}$) is detected at P.

vector $\mathbf{k}^{(i)}$. The electric field of the illuminating light source wave is represented as a complex scalar, i.e. polarization effects are ignored (Equation 5):

$$E_{LS}(\mathbf{r}, t) = A_{LS} \exp(i\mathbf{k}^{(i)}\mathbf{r} - iwt) \tag{5}$$

where A_{LS} is the light wave amplitude, $|\mathbf{k}^{(i)}| = k = 2\pi/\lambda = w/c$ is the wavenumber, λ is the wavelength, w is the frequency, c is the velocity of light, and $\mathbf{r}$ is the position vector. Let $E_0(\mathbf{r}, t)$ be the electric field wave scattered by the object. The sum of the two waves $E_{LS}(\mathbf{r}, t) + E_0(\mathbf{r}, t)$ satisfies the scalar Helmholtz equation. In the case of weakly scattering objects the scattered field can be obtained by the first Born approximation[18] as a volume integral extended over the illuminated object volume (Equation 6):

$$E_0(\mathbf{r}, t) = \frac{1}{4\pi} \int_{v(\mathbf{r}')} F_0(\mathbf{r}') E_{LS}(\mathbf{r}', t) G(|\mathbf{r} - \mathbf{r}'|)\, \mathrm{d}^3\mathbf{r}' \tag{6}$$

The Green's function of the Helmholtz equation is given by Equation (7):

$$G(|\mathbf{r} - \mathbf{r}'|) = \frac{\exp(ik^{(s)}|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} \tag{7}$$

Here $k^{(s)} = |\mathbf{k}^{(s)}| = k$, where $\mathbf{k}^{(s)}$ is the wave vector of the (coherently) scattered wave. Equation (6) can be considered as quantification of Huygens' principle: the Green's function represents the secondary wavelets which combine to form the scattered light. The term $F_0(\mathbf{r})$ is the scattering potential of the object, or the object structure. It determines the relative amplitudes and phases of the scattered wavelets (Equation 8):

$$F_0(\mathbf{r}) = k^2[m^2(\mathbf{r}) - 1] \tag{8}$$

where $m(\mathbf{r})$ is the complex refractive index of the object. In general, m will depend on the wavelength λ; here, m is assumed to be isotropic and independent of the wavelength. Furthermore, because in OCT the object is illuminated by a rather narrow light beam ($w_0 \ll L$ in Figure 4), far-field scattering is a reasonable approximation. In this case the exponent in G can be expressed as in Equation (9):[19]

$$ik^{(s)}|\mathbf{r} - \mathbf{r}'| = i\mathbf{k}^{(s)}(\mathbf{r} - \mathbf{r}') \tag{9}$$

Equation (6) then yields Equation (10),

$$E_0(\mathbf{r}, \mathbf{K}, t) = \frac{A_{LS}}{4\pi L} \exp(i\mathbf{k}^{(s)}\mathbf{r} - iwt) \times \int_{v(\mathbf{r}')} F_0(\mathbf{r}') \exp(-i\mathbf{K}\mathbf{r}')\, \mathrm{d}^3\mathbf{r}' \tag{10}$$

where $\mathbf{K} = \mathbf{k}^{(s)} - \mathbf{k}^{(i)}$ is the scattering vector. The wave scattered by the object is modified by the Fourier transform of the scattering potential of the object.[20] The Fourier variable in the frequency domain is $2/\lambda$. This is the far-field version of the Fourier diffraction theorem.[1] The variable L is the distance of the detection point P at position $\mathbf{r}_P$ from the origin near the object. Hence, the three-dimensional scattering potential $F_0(\mathbf{r})$ can be obtained from the scattered field $E_0(\mathbf{r}, \mathbf{K}, t)$ by an inverse Fourier transform. This is the physical basis of diffraction tomography.[2] Diffraction tomography is still in a rather immature state. A simplified one-dimensional version, however, is used in the optical A-scan of OCT.

There are some limitations in diffraction tomographic imaging, which also pertain to OCT (and other related imaging techniques). These can be seen from the data geometry in **K**-space, Figure 5. A general three-dimensional object will have Fourier data in a certain volume centered, for example, at the origin 0 of **K**-space. For any direction of the scattered light the scattering vector **K**, however, points onto the surface of the so-called Ewald sphere.[21,22] Only these Fourier data are available by scattering. Using a wavelength range from λ_1 to λ_2 gives access to Fourier data between the corresponding Ewald spheres.

The situation is even worse in the OCT A-scan. Here, backscattering along the illumination axis is used. If a wavelength range from λ_1 to λ_2 is used, Fourier components of the scattering potential in a high spatial frequency range (K_{1z}, K_{2z}), indicated by BS in Figure 5, can be accessed. Hence, only relatively abrupt changes in the scattering potential, occurring within a few wavelengths, can be seen by backscattering tomography. Therefore, OCT is a high-pass imaging technique.

For further simplification, we replace the integrations over x' and y' in Equation (10) by a constant factor W,

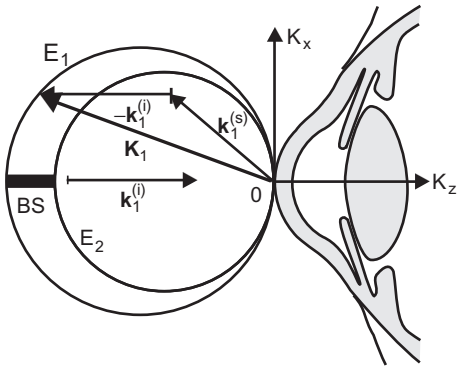

Figure 5 **K**-space geometry of backscattering; with 0 being the **K**-space origin: $\mathbf{k}_1^{(i)}$ is the wave vector of illuminating wave at wavelength λ_1; $\mathbf{k}_1^{(s)}$ is the wave vector of scattered wave at λ_1; $\mathbf{K}_1$ is the scattering vector corresponding to λ_1; E_1 and E_2 are the Ewald spheres corresponding to wavelengths λ_1 and λ_2; BS indicates the Fourier data accessible by backscattering along the illumination axis.

chosen proportional to the cross-section of the beam waist of the illuminating beam. Then the A-scan field backscattered by the object along the illumination axis at point P (namely $\mathbf{k}^{(s)} = -\mathbf{k}^{(i)}$; $|\mathbf{k}^{(s)}| = |\mathbf{k}^{(i)}| = k$) is

$$E_0(\mathrm{P}, k, t) = \frac{A_{\mathrm{LS}} W}{4\pi L} \exp(-ikz - iwt) \times \int_0^D F_0(z') \exp[iKz'] \, \mathrm{d}z' \qquad (11)$$

with $K = 2k = |\mathbf{K}|$. Equation (11) is the physical basis of Fourier domain OCT. The (complex) amplitude of the backscattered A-scan wave equals the (inverse) Fourier transform of the object structure.

Equation (11) can also be written in terms of a convolution (Equation 12):

$$\begin{aligned} E_0(z, k, t) &= \frac{A_{\mathrm{LS}} W}{4\pi L} \exp(-ikz - iwt) \times \int_0^D F_0(z') \exp[2ikz'] \, \mathrm{d}z' \\ &= \frac{A_{\mathrm{LS}} W}{8\pi L} \int_0^{2D} F_0\left(\frac{z'}{2}\right) \times \exp[-ik(z - z') - iwt] \, \mathrm{d}z' \end{aligned} \qquad (12)$$

This is a spatial convolution of the twofold dilated object structure with a backtraveling light wave (Equation 13):

$$E_0(z, k, t) \propto F_{0(1/2)}(z) \otimes E_{\mathrm{B}}(z, k, t) \qquad (13)$$

with (Equation 14)

$$E_{\mathrm{B}} \propto A_{\mathrm{LS}} \exp(-ikz - iwt) \qquad (14)$$

Here the convolution notation of Equation (15) has been used:

$$\begin{aligned} f_{(\alpha)}(t; p) \otimes g_{(\beta)}(t; p) &= \int_{-\infty}^{+\infty} f_{(1)}(\alpha x; p) g_{(1)}(t - \beta x; p) \, \mathrm{d}x \\ &= \int_{-\infty}^{+\infty} f(\alpha x; p) g(t - \beta x; p) \, \mathrm{d}x \end{aligned} \qquad (15)$$

Equation (13) is the basis of time-domain OCT. The backscattered A-scan wave equals the reflected light source wave convolved with the scattering potential of the object. Because of the twofold transit of the light, the z-extension of the scattering potential has to be enlarged by a factor of 2.

2.2 Time-domain and Fourier-domain Optical Coherence Tomography

2.2.1 Time Domain

In time-domain OCT the object structure encoded in the A-scan wave is deciphered with the help of a reference wave (Figure 6). The wave remitted by the object is brought to interference with a low-coherence reference wave traveling back from the reference mirror (Equation 16):

$$E_{\mathrm{R}}(z, k, t) \propto A_{\mathrm{LS}} \exp(-ikz - iwt) \qquad (16)$$

Thus, the resulting field strength is the sum of the object and reference waves (Equation 17):

$$E(z, k, t, \tau_{\mathrm{R}}) = E_0(z, k, t) + E_{\mathrm{R}}(z, k, t + \tau_{\mathrm{R}}) \qquad (17)$$

where a delay $\tau_{\mathrm{R}} = 2(d - s)/c$ between the two waves has been assumed.

The light intensity is defined as the cycle-averaged Poynting vector of the corresponding electromagnetic

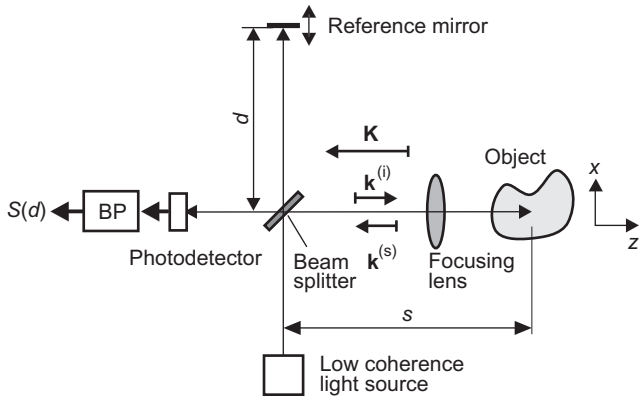

Figure 6 Michelson interferometer configuration for time-domain OCT. The A-scan is performed by scanning the time delay $\tau = (s - d)/c$ with the help of the reference mirror.

wave.[23] However, light waves used in OCT are not monochromatic but are quasimonochromatic waves with statistical properties and should be treated as stochastic processes.[24] This means that observables should be defined as ensemble averages. Usually, these ensemble averages can be replaced by time averages over many cycles of the corresponding observable; therefore, the light intensity is given by Equation (18)

$$I = \frac{1}{2}\varepsilon_0 cn\frac{1}{T}\int_{-T/2}^{T/2} E^*(t)E(t)\,\mathrm{d}t = \frac{1}{2}\varepsilon_0 cn\langle E^*(t)E(t)\rangle \tag{18}$$

where $T \gg 2\pi/w$, ε_0 is the electric permittivity, c is the velocity of light, and n is the phase refractive index.

Time-domain OCT is based on the transit time of wave groups. Wave groups show specific propagation properties. Their phase velocity equals the (phase) velocity of the corresponding carrier wave, whereas the amplitudes propagate with the group velocity.[25] The reference wave is a backtraveling wave directly derived from the light source wave by a beam splitter and reflection at the reference mirror (Equation 19):

$$E_\mathrm{R}(t) = \alpha E_\mathrm{B}(t) \tag{19}$$

where α is a constant depending on the properties of the beam splitter and the reference mirror. The object wave is also a backtraveling wave: E_B convolved with the scattering potential of the object (Equation 13). This convolution introduces a time delay τ_0 (Equation 20):

$$\tau_0 = \frac{\Delta z n_\mathrm{G}}{c} \tag{20}$$

where n_G is the group index and Δz is the additional pathlength introduced by the object.

Hence, the object wave (Equation 21) is

$$E_0(t) = \int_0^{2D} F_{0(1/2)}(t-\tau)E_\mathrm{B}(t)\,\mathrm{d}t = F_{0(1/2)}(\tau) \otimes E_\mathrm{B}(\tau) \tag{21}$$

The desired information is contained in the mutual coherence function of the re-emitted object wave and the reference wave in the interferogram term. Assume that the pathlengths of the reference beam and the object beam are matched. Then the intensity at the interferometer exit is (Equation 22)

$$I_\mathrm{E}(\tau_\mathrm{R}) = \tfrac{1}{2}\varepsilon_0 c\langle|E_\mathrm{R} + E_0|^2\rangle = I_\mathrm{R} + I_0 + \alpha\varepsilon_0 c\,\mathrm{Re}\{\langle E_\mathrm{B}^*(t)E_0(t+\tau)\rangle\} \tag{22}$$

Using the commutative law of convolution yields Equation (23):

$$I_\mathrm{E}(\tau) = \tfrac{1}{2}\varepsilon_0 c\langle|E_\mathrm{R} + E_0|^2\rangle = I_\mathrm{R} + I_0 + \alpha\varepsilon_0 c \times \mathrm{Re}\{\langle F_{0(1/2)}(\tau) \otimes E_\mathrm{B}^*(\tau)E_\mathrm{B}(\tau+\tau_\mathrm{R})\rangle\} \tag{23}$$

The time average $\langle\cdots\rangle$ depends on the time coherence function $\Gamma_\mathrm{LS}(\tau)$ or the complex degree $\gamma_\mathrm{LS}(\tau)$ of coherence of the source wave (Equation 24)[24]

$$\langle E_\mathrm{B}(t)E_\mathrm{B}^*(\tau+\tau_\mathrm{R})\rangle = I_\mathrm{B}\gamma_\mathrm{LS}(\tau_\mathrm{R}) \tag{24}$$

with (Equation 25)

$$\gamma_\mathrm{LS}(\tau) = \frac{\langle E_\mathrm{LS}^*(t)E_\mathrm{LS}(t+\tau)\rangle}{[\langle E_\mathrm{LS}^*(t)E_\mathrm{LS}(t)\rangle\langle E_\mathrm{LS}^*(t+\tau)E_\mathrm{LS}(t+\tau)\rangle]^{1/2}} = \frac{\frac{1}{2}\varepsilon_0 c\langle E_\mathrm{LS}^*(t)E_\mathrm{LS}(t+\tau)\rangle}{I_\mathrm{LS}} \tag{25}$$

Hence, besides being a constant factor, the interferogram term is the real part of the convolution of the object structure with the complex degree of the time coherence (Equation 26):

$$I_\mathrm{E}(\tau) = \tfrac{1}{2}\varepsilon_0 c\langle|E_\mathrm{R} + E_0|^2\rangle = I_\mathrm{R} + I_0 + 2I_\mathrm{B} \times \mathrm{Re}\{\gamma_\mathrm{LS}(\tau) \otimes F_{0(1/2)}(\tau)\} \tag{26}$$

Therefore, the object structure along the A-scan is obtained by scanning the time delay τ with the help of the reference mirror. The complex degree $\gamma_\mathrm{LS}(\tau)$ of the time coherence plays the role of a longitudinal or depth point spread function.

If the object is a single reflecting plane at z_0, this gives (Equation 27)

$$I(\tau) = I_0 + I_\mathrm{R} + 2\sqrt{I_0 I_\mathrm{R}}\,\mathrm{Re}\{\gamma_\mathrm{LS}(\tau)\} \tag{27}$$

Therefore, in PCI the complex degree $\gamma_\mathrm{LS}(\tau)$ of the time coherence plays the role of a longitudinal or depth point spread function.

2.2.1.1 Dispersion Most A-scans are performed through dispersive media, whereas the reference beam travels, for example, through air. Then the Fourier components of the A-scan wave and the reference wave experience different time delays and the two waves decorrelate. Hence, the complex degree $\gamma_\mathrm{LS}(\tau)$ of the time coherence broadens and the resolution decreases. If the length of a dispersive medium in one of the interferometer arms is L and the group dispersion of the medium is $\mathrm{d}n_\mathrm{G}/\mathrm{d}\lambda$, the width of the coherence envelope, after double passing through the medium, is given by Equation (28):

$$l_{\mathrm{C,m}} = \left(l_\mathrm{C}^2 + \left(\frac{\mathrm{d}n_\mathrm{G}}{\mathrm{d}\lambda}2L\Delta\lambda\right)^2\right)^{1/2} \tag{28}$$

Optimum resolution is achieved by placing a dispersion-compensating plane-parallel plate in the longer interferometer arm.

2.2.2 Fourier Domain: Complex Spectral Optical Coherence Tomography

In Fourier domain OCT the object structure can be obtained by several techniques. The most straightforward technique is to detect the (complex) K-spectrum of amplitude and phase data of the scattered field (Equation 11). The object structure is then obtained from an inverse Fourier transform (Equation 29):

$$F_0(z) \propto \mathrm{FT}^{-1}\{E_0(P, K)\} \tag{29}$$

The K-spectrum of the light backscattered in an A-scan can be obtained with the help of an interferometric spectrometer. This device is a spectrometer with phase-measuring capability (provided by the interferometer) as depicted in Figure 7.

In the interferometric spectrometer a detector array is used at the focal plane of the spectrometer optics. Known phase changes are induced between the object and reference beam by shifting the reference mirror by fixed amounts. From the spectral intensity data at the interferometer exit obtained at these reference mirror positions the spectral object wave phase and its intensity can directly be calculated.[28] Several algorithms are known that calculate the phase of a light wave using this technique; these algorithms also compensate phase-shift miscalibration and detector nonlinearity.[29]

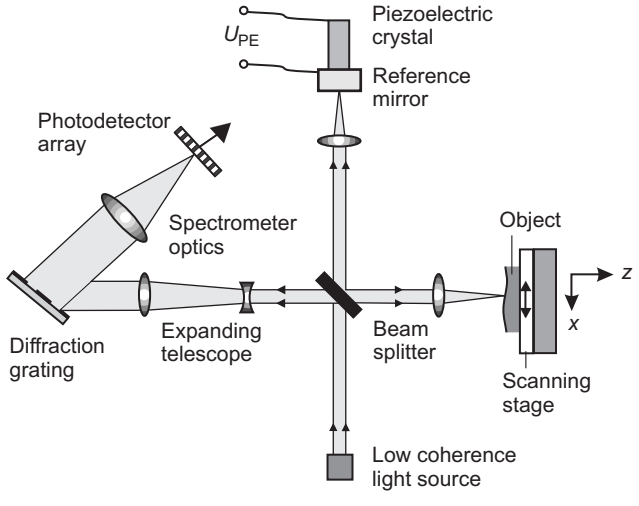

Figure 7 Fourier domain OCT with a Michelson interferometric spectrometer. U_{PE} is the piezo electric driving voltage used for phase measurement.

2.2.3 Fourier Domain: Spectral Interferometry Optical Coherence Tomography

In this technique the light exiting from the interferometer is detected by a photodetector array at the exit plane of a spectrometer in a similar way as in the complex interferometry technique. However, only the spectral intensity is recorded. No spectral phases are measured. Hence, a Fourier transform of the measured data does not yield the object structure but its autocorrelation. However, autocorrelation is not reversible. To obtain the object structure there are two possibilities. First, an additional singular light-remitting interface (reference mirror) is positioned at an optical path difference z_R from the object. Second, the object scattering potential can also be obtained if the object itself contains one interface with a relatively large reflectivity R acting as a reference mirror.[20] In both cases the scattering potential can be described as a sum of the object structure $F_0(z)$ plus a delta-like potential at z_R (with amplitude reflectivity R; Equation 30):

$$F(z) = F_0(z) + R\delta(z - z_R) \tag{30}$$

and the autocorrelation yields four terms (Equation 31):

$$\begin{aligned}\langle F^*(z)F(z+Z)\rangle &= \langle F_0^*(z)F_0(z+Z)\rangle \\ &\quad + \langle F_0^*(z)R\delta(z+Z-z_R)\rangle \\ &\quad + \langle R\delta^*(z-z_R)F_0(z+Z)\rangle \\ &\quad + \langle R^2\delta^*(z-z_R)\delta(z-z_R+Z)\rangle \\ &= \mathrm{ACF_F}(Z) + RF_0^*(z_R - Z) \\ &\quad + RF_0(z_R + Z) + R^2\delta(Z)\end{aligned} \tag{31}$$

An overlap between the four terms of the autocorrelation can be avoided by choosing the distance between the reference mirror and the object larger than the object depth $D : |z_R| > D$. Then the third term of Equation (31) yields a reconstruction of the object structure, centered at $Z = -z_R$.

2.2.4 Fourier Domain: Chirp Optical Coherence Tomography

Chirp OCT[30] (the name comes from the early days of the related electronic radar technique, when a sound produced by a pulse resembled the chirps of a bird in song), or wavelength tuning OCT[20] uses the tunable laser interferometry technique. Here the wavelength is tuned over a range of wavenumbers and the spectral light intensity remitted from the object is detected. Whereas in the spectral interferometry technique the intensity spectrum is simultaneously recorded by the detector array, here the same data are recorded sequentially by a single photodetector.

2.3 Doppler Optical Coherence Tomography

This technique combines OCT with Doppler velocimetry.[31] Doppler velocimetry detects the frequency shift w_D in the light scattered by moving particles (Equation 32):

$$w_D = \left(\mathbf{k}^{(s)} - \mathbf{k}^{(i)}\right)\mathbf{v} = \mathbf{K}\mathbf{v} \tag{32}$$

where $\mathbf{v}$ is the velocity of the moving particle. A frequency shift in the object beam leads to an additional phase term $f(t) = \exp(iw_D t)$ (Equation 33):

$$E_0(t) = f(t)E_B(t) \tag{33}$$

The intensity at the interferometer exit, with the reference field strength $E_B(t)$, is given by Equation (34):

$$\begin{aligned} I(\tau) &\propto \langle |E_B(t) + E_0(t+\tau)|^2 \rangle \\ &\propto I_R + I_0 + 2\mathrm{Re}\{\langle E_B^*(t)E_0(t+\tau)\rangle\} \end{aligned} \tag{34}$$

The frequency-dependent intensity or power spectrum is obtained from the Fourier transform of the correlation function of the electric fields (Equation 35):

$$\begin{aligned} I(w) &\propto \mathrm{FT}\{\langle E_B^*(t)E_0(t+\tau)\rangle\} \\ &= \mathrm{FT}\{f(t)\langle E_B^*(t)E_B(t+\tau)\rangle\} \\ &= \hat{f}(w) \otimes S(w) \end{aligned} \tag{35}$$

where $\hat{f}(w)$ is the spectrum of the light scattered by the moving particles and $S(w)$ is the power spectrum of the source light.

2.4 Spectroscopic Optical Coherence Tomography

One of the most powerful spectrometric techniques is Fourier spectrometry. This technique relies on the Fourier relationship between the frequency-dependent spectrum of a light beam and its first-order electric field correlation function (Equation 36):[23]

$$I_0(w) = \tfrac{1}{2}\varepsilon_0 cn\mathrm{FT}\{\langle E_0^*(t)E_0(t+\tau)\rangle\} \tag{36}$$

To use this technique in OCT the complete A-scan signal must be recorded. No rectification or low-pass filtering must be used. This signal has then to be split a into series of small time windows. Within each time window a Fourier transform yields the optical spectrum of the back-scattered light. The corresponding depth positions are obtained in the usual way.

2.5 Resolution

2.5.1 Optical Coherence Tomography Depth Resolution

In time-domain OCT the complex degree $\gamma_{LS}(\tau)$ of the time coherence plays the role of a longitudinal or depth point spread function (Equation 26). For example, most of the light sources used in OCT at present have an approximately Gaussian frequency spectrum. Therefore, they also have an approximately Gaussian degree of coherence (Equation 37),[23]

$$\gamma_{LS}(\tau) = \exp\left(-iw_0\tau - \tfrac{1}{2}\delta^2\tau^2\right) \tag{37}$$

with the full width at half-maximum (fwhm) value or coherence time $\tau_C = 1.177/\delta$, and the longitudinal coherence length $l_C = c\tau_C$ (Equation 38):

$$l_C = \frac{4\ln 2}{\pi}\frac{\lambda_0^2}{\Delta\lambda} \tag{38}$$

where λ_0 is the mean wavelength and $\Delta\lambda$ is the wavelength fwhm. In OCT, however, where the light travels along the same path twice (back and forth through the object) the depth resolution Δz is given by Equation (39):

$$\Delta z = \frac{l_C}{2} = \frac{2\ln 2}{\pi}\frac{\lambda_0^2}{\Delta\lambda} \tag{39}$$

For example, superluminescent diodes at a mean wavelength $\lambda_0 = 830\,\mathrm{nm}$, have spectral widths of about $\Delta\lambda = 25\,\mathrm{nm}$. Thus a corresponding coherence length of $l_C = 24\,\mu\mathrm{m}$ and a depth resolution of approximately $12\,\mu\mathrm{m}$ are obtained.

Depth resolution in Fourier domain OCT can be estimated with the help of the Fourier uncertainty relationship. For a Gaussian spectrum the fwhm values Δz and Δs of the Fourier variables are related by Equation (40):

$$\Delta z \Delta s = \frac{4\ln 2}{\pi} \tag{40}$$

With $s = 2/\lambda$ the same relationship as in time-domain OCT is obtained.

2.5.2 Optical Coherence Tomography Transverse Resolution

In imaging optics, transverse resolution is usually defined by the Rayleigh criterion. Two object points are resolved if the center of the transverse point spread function (Airy disk) of one object point falls on the first zero of the point spread function generated by the second.[32] Most beams used in OCT are approximately Gaussian beams in their fundamental transverse mode. In this

case, however, there is no zero intensity. If the transverse sampling rate is high enough, i.e. if it satisfies the sampling theorem, transverse resolution in OCT is determined by the transverse width of the beam performing the coherence scan. High transverse resolution demands that the beam is focused to a small beam waist radius w_0. This is the radius at which the beam intensity falls to $1/e^2$ of its central value.[17] The minimal transverse separation of two resolved image points is of the order of this value. Hence (see Equation 3), w_0 can be used to define the minimal transverse separation δ resolved in OCT (Equation 41):

$$\delta = \frac{\lambda}{\pi\vartheta} \tag{41}$$

A small beam waist radius demands a large asymptotic angle ϑ of beam divergence, but a high divergence results in a reduced confocal beam parameter z_0. Twice the confocal beam parameter can be used to define the depth of focus (DOF) (Equation 42):

$$\mathrm{DOF} = 2z_0 = 2\frac{\pi w_0^2}{\lambda} \tag{42}$$

Therefore, a compromise has to be found between the desired DOF and the desired transverse resolution δ. For example, a transverse resolution of $\delta = 20\,\mu\mathrm{m}$ at a mean wavelength of $\lambda_0 = 830\,\mathrm{nm}$ leads to a DOF of 3 mm, whereas a transverse resolution of $\delta = 2\,\mu\mathrm{m}$ reduces the DOF to only 30 µm.

The resolution arguments discussed above apply to all the OCT techniques described in this article.

3 OPTICAL COHERENCE TOMOGRAPHY SCHEMES

3.1 Reflectometer Optical Coherence Tomography

Many PCI techniques have been described. PCI has been used for many years to measure the thickness of thin films, as a position sensor, and to measure other features that can be converted to a displacement. At present all OCT techniques use PCI to perform the A-scan.

The standard PCI technique is depicted in Figure 8. The A-scan is performed by moving the reference mirror with a constant speed. Interference bursts occur at the interferometer exit, if any pathlength to a light-remitting site in the object arm coincides (within the coherence length) with the pathlength in the reference arm.

An alternating current (AC) signal U is detected at these positions by the photodetector, which has the Doppler frequency corresponding to the speed of the reference mirror. Therefore, noise-reducing electronic AC techniques can be used to obtain high signal quality.

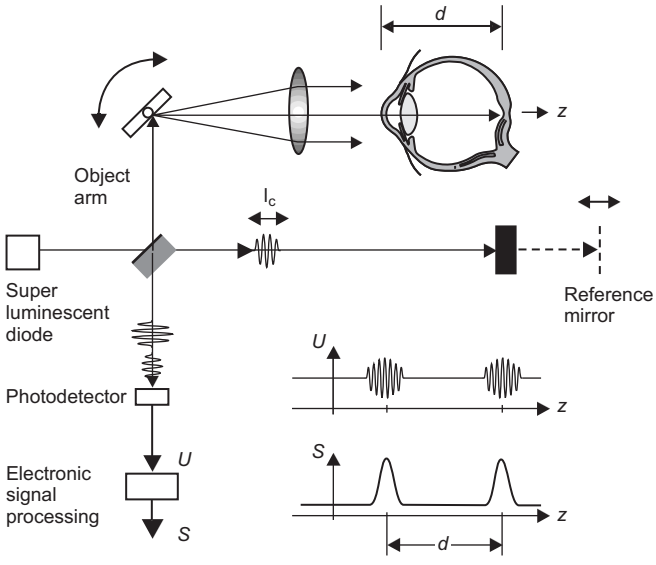

Figure 8 Reflectometer OCT: U is the electric photodetector signal; S is the interferometer A-scan signal. The A-scan is performed by moving the reference mirror at constant speed. For simplification only two interference bursts (a distance d apart) are indicated. The interferometer arm lengths are not to scale.

Amplification followed by band-pass frequency filtering, rectification, and low-pass frequency filtering is normally used to obtain the electronic interferometer signal S. This signal can either be used to determine distances between object interfaces or as the basic PCI depth scan signal for OCT.

At present, reflectometer OCT is by far the most frequently used OCT technique. One disadvantage of the reflectometer technique is that the reference mirror has to be shifted along the whole optical length of the object. However, as living samples move, the reference mirror is moved at high speed to avoid erroneous measurements.[33] A more basic solution to this problem is the dual-beam technique. In this technique a path difference generated in the interferometer is matched to the path difference in the object. Hence the measurement becomes more or less insensitive to object movements.

3.2 Dual Beam Optical Coherence Tomography

In this technique the object is illuminated by both beams exiting from a Michelson interferometer (or other two-beam interferometer). The interferometer splits the beam exiting from the light source into two components E and E', generates a path difference, and recombines the two components. One of the interferometer mirrors is moved by a stepper motor with a constant speed v, which causes a Doppler shift $f_D = 2v/\lambda$ of the light frequency of one beam component. The recombined beam leaves the

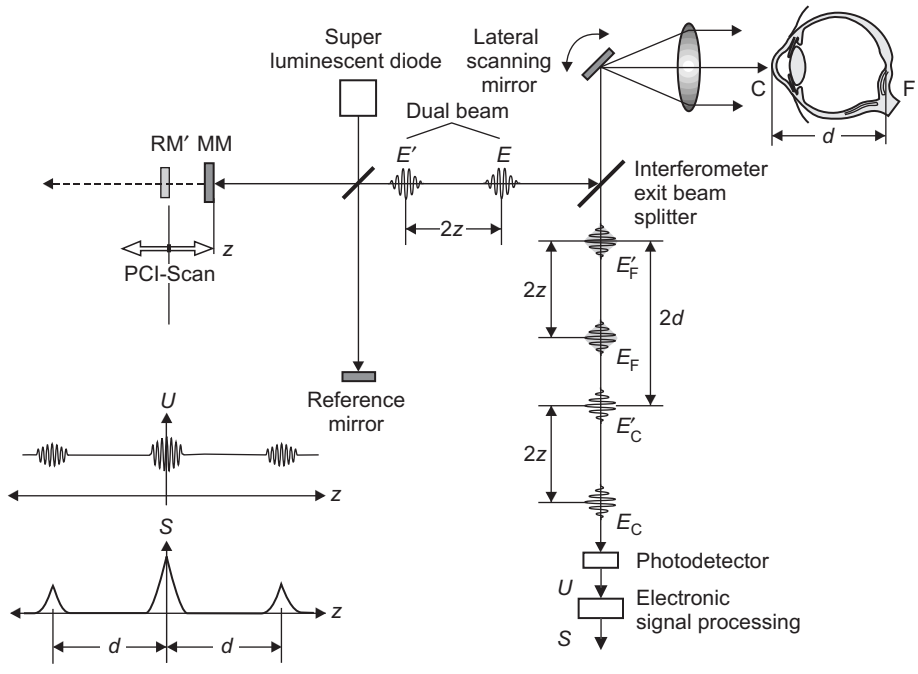

Figure 9 Dual-beam OCT: U is the electric photodetector signal, S is the interferometer A-scan signal. The A-scan is performed by moving the measurement mirror (MM) at constant speed, thus matching the path difference z of the Michelson interferometer to distances between light remitting sites of the object. E_F and E'_F are subcomponents remitted by the eye's fundus; E_C and E'_C are subcomponents remitted by the cornea.

interferometer as a coaxial dual beam. A lateral scanning mirror directs the dual beam to transversely adjacent positions at the object. Each of the two components E' and E of the dual beam that leave the interferometer towards the object generate subcomponents reflected at the object interfaces, separating regions of different refractive indices in the object. If the interferometer arm length difference coincides with an object interface distance, an interference burst occurs at the photodetector at a frequency f_D. The photodetector signal is amplified and filtered by a band-pass filter that transmits only signals with f_D. The envelope of the rectified signal (S in Figure 9) is the final A-scan signal and is recorded as a function of the interferometer's arm length difference z and azimuthal position by a personal computer.

The dual-beam scheme has the advantage of being relatively insensitive to object movements. It needs, however, a reference interface with high reflectivity.

3.3 Enface Optical Coherence Tomography

OCT generates depth images, i.e. images in planes normal to the object surface. Enface OCT generates transverse images. These images can be obtained by fixing the reference mirror and transversely scanning either the object or the sample arm of the interferometer.[34] A corresponding arrangement has also been described by Podoleanu et al.[35] Their experimental arrangement used a single-mode directional coupler, where the output fibers form a Michelson interferometer. The reference arm consists of a fixed microscope objective and a mirror mounted on a micrometer scanning stage. The sensing arm contains a microscope objective collimating the beam exiting from the fiber. A bidirectional galvanometer-driven scanner is used to transversely scan the target. Scanning the light beam across the (plane) target surface leads to a modulation of the corresponding pathlength and thus to a frequency modulation of the photoelectric signal, similar to that generated by the PCI depth scan.

A disadvantage of this technique is that, owing to the nonlinear pathlength dependence on the scanning angle, a frequency spread of the photoelectric signal occurs leading to a reduced signal-to-noise ratio. This disadvantage can be overcome by either frequency shifting the reference beam or laterally displacing the target beam from the axis of rotation of one of the galvanometer-driven mirrors, which leads to an additional modulation that is essentially linearly dependent on the deflection angle (Figure 10).

3.4 Dynamic Coherent Focus Optical Coherence Tomography

In contrast to conventional optical imaging, in OCT the depth or longitudinal resolution and transversal

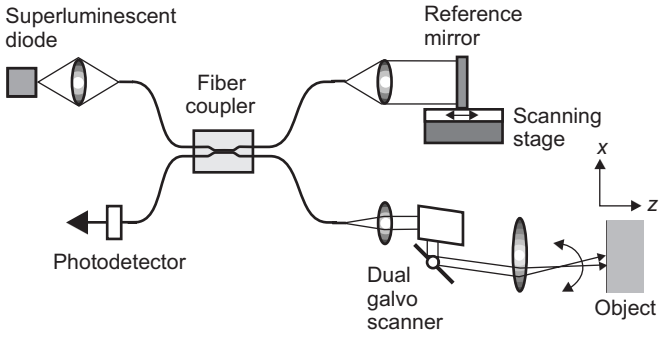

Figure 10 Optical scheme of enface OCT. The A-scan is performed by one of the galvo scanners in a direction (for example, the x direction) transverse to the illuminating beam. The scan depth is adjusted by the reference mirror.

resolution are uncoupled. OCT depth resolution is determined by half the coherence length l_C of the light used, whereas OCT transverse resolution, as for transverse resolution in lens optics, is limited by the aperture of the imaging light beam. Most OCT techniques generate depth tomograms (i.e. sections perpendicular to the object surface). Hence, it is not only the transverse resolution that counts but also the DOF (section 2.5). The DOF versus transverse resolution w_0 is nonlinear (Equation 42).

In order to maintain high transverse resolution throughout a reasonable object depth, the focusing lens can be shifted synchronously with the reference mirror during the depth scan.[36] However, this involves the (rapid) movement of at least two optical components. Alternatively, the dynamic coherent focus technique can be used, which only needs one moving mirror.[37] The optical principle is depicted in Figure 11.

The probing beam of a reflectometer or dual beam device is focused to a first focus by lens 1 at an oscillating mirror, positioned near the focal plane of another lens 2. The oscillating mirror performs the A-scan. The beam leaving lens 2 is turned round by a scanning unit (scanning mirror 2) and focused by lens 3 to the dynamic coherent focus within the object. By choosing the focal lengths of lenses 2 and 3, a preselected shift of the dynamic focus within the object can be realized.

Focusing the probing beam through a medium with a refractive index different from that of air causes an additional pathlength mismatch. First, the focus formed within the object is shifted by refraction at the interface between the air and the object (with phase refractive index n). Second, in OCT the optical pathlength in a medium equals the geometrical pathlength multiplied by the group refractive index n_G. Hence, the dynamic focus is at an optical depth given approximately by Equation (43):

$$z'' = nn_G z \tag{43}$$

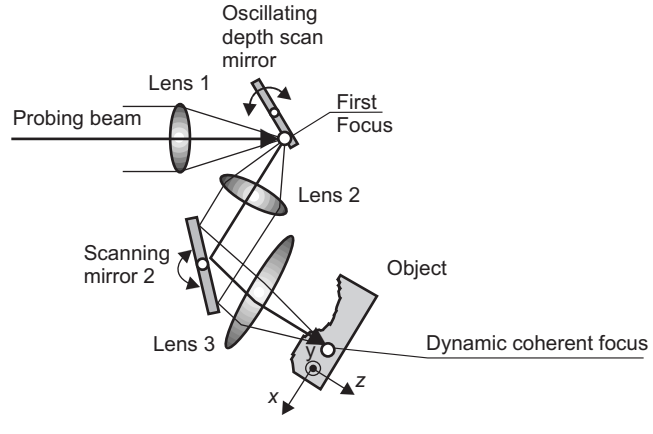

Figure 11 Dynamic coherent focus for enhanced depth of field OCT. The first focus is imaged by lenses 2 and 3 onto the object.

3.5 Doppler Optical Coherence Tomography

Two schemes have been used to obtain OCT maps of flow. In the so-called optical Doppler tomography (ODT) technique[31] lateral scans are performed with a constant velocity, followed by an incremental probe movement in the axial direction. In the so-called color Doppler optical coherence tomography (CDOCT) technique[38] the usual OCT scan mode is used, i.e. axial A-scans are performed at laterally adjacent locations.

3.5.1 Optical Doppler Tomography

In the ODT technique the light from the superluminescent diode and an aiming He–Ne laser beam are coupled into a fiber interferometer using a 2×1 coupler and then split equally into reference and target arms by a 2×2 fiber coupler (Figure 12). Piezoelectric transducer optical delay lines are used to modulate the optical

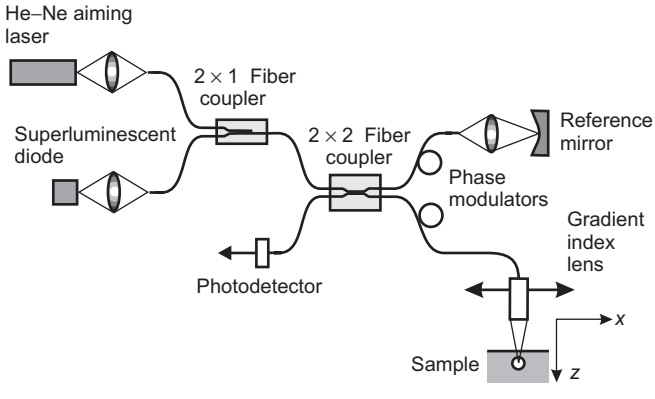

Figure 12 Diagram of ODT. Tomograms are formed from sequential lateral scans in the x direction followed by probe movement in the z direction.[31]

phase of the reference and probing beams. The probing beam is focused onto the sample by a gradient index lens with the optical axis oriented at 15–20° to the normal of the flow velocity (otherwise the Doppler shift vanishes, see Equation 32). Tomographic images are formed by laterally scanning the sample in a transversal direction followed by a sample movement in the z direction between the lateral scans. The interference fringe intensity at the interferometer exit is detected by a photodetector. After analog-to-digital conversion a short-time fast Fourier transform of the photodetector signal yields a spectrogram. From the signal at the phase modulation frequency a tomographic structural image is obtained, whereas from the Doppler frequency shift the flow velocity is determined.

3.5.2 Color Doppler Optical Coherence Tomography

In this technique, as in time-domain OCT, A-scans are performed at a series of laterally adjacent locations. Whereas in time-domain OCT the envelope of the interferometric signal is detected, CDOCT uses coherent detection of the interferometer signal. As in OCT the reference wave is Doppler frequency shifted by the constant motion of the reference mirror (Figure 13). The actual interferometric fringe frequency detected by the photodetector arises from the net sum of Doppler shifts generated by the moving reference mirror and moving scatterers in the sample. After band-pass filtering, coherent demodulation of the photodetector signal, and analog-to-digital conversion, short-time Fourier transform is applied for each depth scan. The resulting power spectra correspond to short-time sections of the A-scan. The local Doppler frequency is estimated from the centroid of each spectrum. The detected velocity is color coded to indicate magnitude and direction of flow.

Using a time-gating technique (periodic) dynamic changes in tissues have been recorded by CDOCT. For example, a series of frames of the beating heart of *Xenopus laevis* has been recorded. First, 1000 laterally oversampled A-scan OCT image with at least five A-scans per heartbeat was acquired while the sample was translated laterally. The frames were then composed from the 1000 A-scans obtained at different lateral positions but at the same segment of the cardiac cycle.[39]

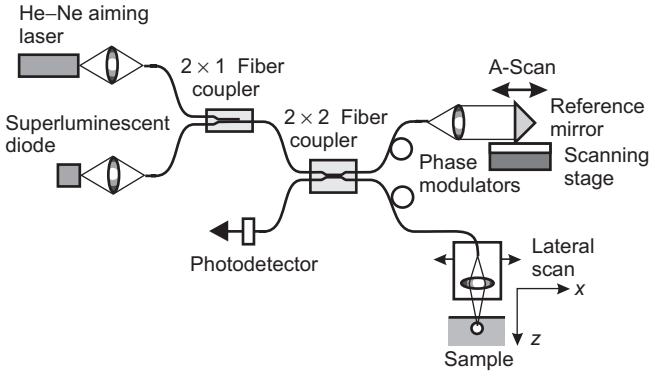

Figure 13 Color Doppler OCT imaging system.[38]

3.6 Endoscopic Optical Coherence Tomography

Essential to OCT imaging of internal organs is a catheter endoscope. In the OCT catheter endoscope depicted in Figure 14 light from the interferometer probing arm is coupled across a small air gap into the catheter optical fiber. The OCT catheter consists of a single-mode fiber attached to a focusing gradient index (grin) lens with a beam-directing microprism at the distal end, and an optical coupling element at its proximal end. Precise optical coupling between the two fibers is ensured by mounting the fibers in a precision mechanical ferrule assembly using two male fiber connectors and a male-to-male optical fiber coupler, typically used for rigidly connecting two fibers. The body of the catheter is composed of a flexible, rotating inner sleeve that fits loosely inside a stationary outer sheath. The optical fiber is fixed within the center of the inner sleeve, which is rotated as a unit with the optical connector.

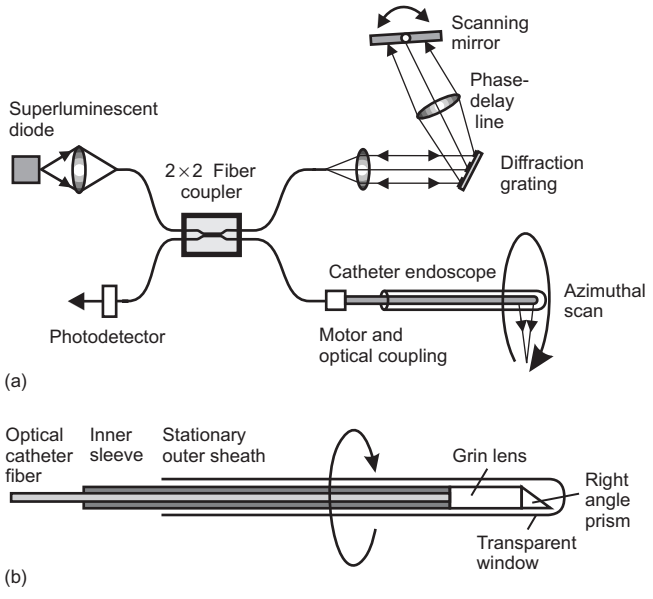

Figure 14 Diagram of OCT catheter endoscope with 1.1 mm diameter at the catheter window:[40] (a) OCT Michelson set-up; (b) distal end of endoscope catheter consisting of a single-mode fiber with a core diameter of 9 μm, a 0.2725-pitch gradient index lens, and a 0.5 mm right-angle microprism.

4 IMPLEMENTATION

Only a few components specific to OCT are mentioned here.

4.1 Light Sources

Depth resolution as well as spectrometric resolution in OCT are determined by the spectral width of the light source. Table 1 shows some examples of spectral widths and coherence lengths of different light sources; most have already been used in OCT.

The ideal OCT light source would yield radiation with perfect space coherence but low time coherence. In terms of laser optics this radiation must only contain longitudinal modes. This requirement is not generally met by low-coherence lasers. At present the most promising light sources in OCT are SLD and mode-locked solid-state lasers. Thermal light sources yield extremely low time-coherence radiation but suffers from a corresponding low space coherence. Filtering out spatially coherent waves reduces the available power to a few microwatts.

4.2 Optical Delay Lines

Time-domain OCT requires the scanning of a reference group delay to acquire the A-scan data. As many A-scans (500 or more) are acquired per tomographic image, rapid optical delay lines are needed. Many technical solutions have been proposed. As the total optical pathlength modulation needed is of the order of some millimeters, most delay lines use a mechanical solution (Table 2).

The first optical delay lines used a reference mirror mounted on a stepper-motor-driven linear translation stage. Scanning velocities of these systems are limited to some $10\,\mathrm{cm\,s^{-1}}$. For small scanning ranges piezoelectric actuators are suitable. However, their scanning range is limited to about 100 μm. Piezoelectric devices also suffer from drift and hysteresis.

Spinning glass cubes and spinning mirror pairs yield far higher scan velocities but suffer from a limited duty cycle. Spinning glass cubes introduce a delay-dependent dispersion. Spinning mirror pairs do not introduce dispersion. Moreover, a given dispersion in the collateral interferometer arm can be compensated by arranging a dispersion prism in front of the retroreflector, as indicated in Figure 15. High scan velocities and repetition rates can be achieved.

A standard optical delay line technique is a long fiber coiled around a cylindrical piezoelectric transducer used

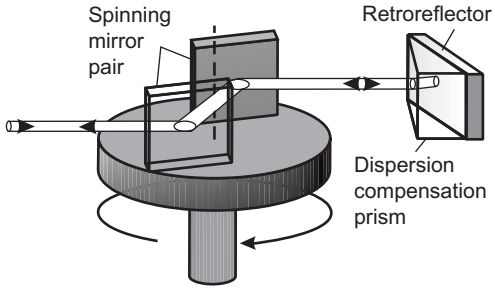

Figure 15 Spinning mirror pair used as optical delay line. Different dispersion prisms in front of the retroreflector can be used for matching a desired dispersion independently from the pathlength delay.

Table 1 Coherent power, mean wavelength λ_0, spectral width $\Delta\lambda$, and coherence length l_C of some light sources

Light source	Coherent power	λ_0 (μm)	$\Delta\lambda$	l_C	Ref.
Sunlight at sea level	few μW	0.650	0.5 μm	0.75 μm	37
Cd spectral lamp	few μW	0.644	0.0013 μm	28 cm	21
Ar laser pumped Ti sapphire fluorescence	2 μW	0.780	0.144 μm	3.7 μm	38
Mode-locked Ti/Al_2O_3 laser	400 mW	0.800	0.145 μm	3.8 μm	39
SLD	few mW	0.8	60 nm	9.4 μm	40
He–Ne laser	10 mW	0.633	10^{-3} nm	0.353 m	41

SLD, superluminescent semiconductor diodes.

Table 2 Parameters of optical delay lines

Delay line	Repetition rate	Duty cycle (%)	Scanning range (mm)	Independent phase and group delay	Ref.
Piezoelectric fiber stretcher	600 Hz	100	3	no	42
Rotating glass cube	80 Hz	50	3	no	43
Air-turbine-driven rotating glass cube	28.6 kHz	30	2	no	44
Grating-based tilting mirror	2 kHz	100	3	yes	45

to stretch it. High-speed data acquisition and path delays of a few millimeters are possible. The drawbacks of this technique are birefringence, polarization mismatch by the fiber, drifts and hysteresis caused by the piezoelectric transducer, and the high voltage needed to operate the device.

A high-speed delay line with independent phase and group delay has been decribed by Tearney et al.[49] From the Fourier shift theorem it follows that a linear phase ramp in the Fourier domain causes a delay in the time domain. Hence, a linear phase ramp is generated by placing a tilted mirror at the Fourier plane of a lens. Scanning the mirror angle produces a time-dependent optical group delay. Shifting the scanning mirror so that the center wavelength is offset from the axis of rotation generates an arbitrary heterodyne frequency (Figure 16).

High repetition rates can also be realized by oscillating devices such as galvanometer scanners. These devices can be operated with triangle waveforms at high frequencies (see Figure 11).[37]

4.3 Optical Detectors

OCT signals from biomedical objects are weak.[50] For low-light detection in the visible and NIR spectral range three basic detector technologies are available: the silicon (p-doped, intrinsic semiconductor, with n-doped layers) PIN photodiode, the silicon avalanche photodiode (APD), and the photomultiplier tube. Most OCT light sources operate in the red and NIR spectral region, where PIN diodes and APDs offer superior quantum efficiency, approximately two orders of magnitude larger than the photomultiplier tube. The internal gain of an APD can boost the signal above the noise of the preamplifier and/or the detector current. However, an improvement is realized only when preamplifier noise and dark-current noise are the limiting factor. If the signal is strong enough, the APD performs worse than the equivalent PIN diode.

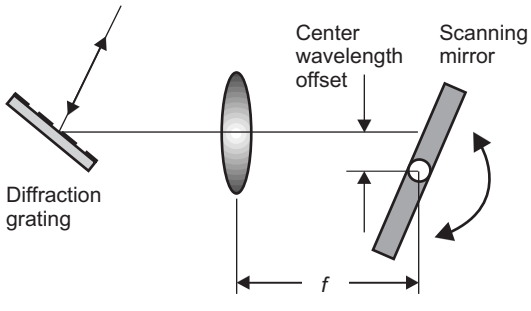

Figure 16 High-speed delay line with independent phase and group delays; f = focal length of the lens.

5 APPLICATIONS

OCT has found manifold applications in medicine but only a few technical applications, such as tomographic imaging of the fiber architecture of glass-reinforced polymer composites[51] and the detection of subsurface defects in ceramics.[52] Compared to related imaging techniques such as confocal imaging, OCT has the advantage in that its depth resolution is independent of the aperture of the scanning light beam. Therefore, OCT is useful in cases where high depth resolution is required but only a limited aperture is available. Because the OCT image is constituted from the tissue reflectivity of an infrared (IR) beam, precautions are required to interpret the cross-sectional images. As light is strongly scattered in most tissues, the imaging depth of OCT is limited to a few millimeters. OCT yields depth images and can be used to image NIR-transparent objects, such as tissues of the eye (Figure 17) and mucosal tissues.

5.1 Ophthalmology

Ophthalmology is the main field of OCT applications in medicine. OCT has been used to perform direct imaging of the ocular structure in the anterior and posterior segments of human eyes in vitro and in vivo.[53–55] OCT yields a cross-sectional image of the retina with a resolution comparable to a histological section in light microscopy.[56] OCT can discriminate the cross-sectional morphological features of the fovea and optic disk.[57] OCT was successful in staging macular holes and provided a quantitative measure of hole diameter and the amount of surrounding macular edema. OCT was also used to evaluate the vitreoretinal interface in a patient's fellow eyes and was able to detect small separations

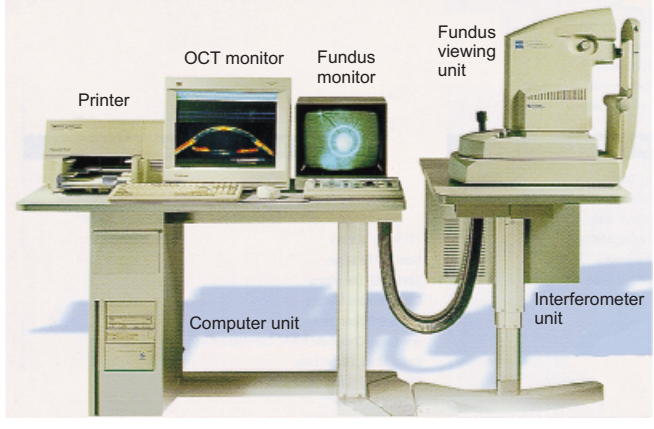

Figure 17 OCT scanner for two-dimensional cross-section imaging of the posterior segment of the eye. (Reproduced by courtesy of Carl Zeiss.)

of the posterior hyaloid from the retina.[58] Idiopathic juxtafoveal retinal telangiectasis[59] and degenerative retinoschisis from retinal detachment[60] have also been studied using OCT. The presence of a nonregular pattern of the retinal pigment epithelium was found by OCT in partial lipodystrophy.[61] The high depth resolution of OCT that is available even at the eye's fundus has been used to measure the thickness of fundus layers.[55,57,62,63]

Animal experiments have been performed to elucidate the relationship between OCT scans and retinal histology. In a direct comparison of OCT and histological images of the *Macaca mulatta* macula, the OCT images demonstrated reproducible patterns of retinal morphology that corresponded to the location of retinal layers seen on light microscopic overlays.[64] In normal and retinal degeneration chickens, quantitative OCT signals have also been shown to provide a predictable relationship between histology and pathology.[65]

Imaging of macular diseases with OCT and its correlation with fluorescein angiography and fundus examination has been studied in a series of pathologies, such as full- and partial-thickness macular hole, epiretinal membrane, macular edema, intraretinal exudate, idiopathic central serous chorioretinopathy, and detachments of the pigment epithelium and neurosensory retina.[66] Macular imaging with OCT has been shown to provide new information concerning posterior pole diseases and in thickness measurements, and is expected to be useful in studying internal layers of the retina.[67] Quantitative assessment of macular edema with OCT can be used for objectively monitoring retinal thickness with high resolution in patients with macular edema. OCT can be a sensitive diagnostic test for the early detection of macular thickening in patients with diabetic retinopathy.[68,69]

There is some doubt about the usefulness of OCT in patients with nonexsudative age-related macular disease (AMD). Although drusen, alteration of the retinal pigment epithelium, and secondary retinal changes could be identified, other structures such as basal laminar (linear) deposits could not be seen with this method.[70] In a study of age-related macular degeneration and choroidal neovascularization (CNV), OCT proved useful in quantitatively evaluating subretinal and intraretinal fluid, assessing possible subfoveal involvement of neovascularization, and in monitoring classic CNV before and after laser photocoagulation. OCT was, however, unable to detect CNV beneath serous pigment epithelial detachments.[71]

OCT has also been used to study optic disk pit maculopathy. OCT confirmed the two-layer structure of optic disk pit maculopathy and that the improvement in central vision after pneumatic displacement coincides with a reattachment of the outer limiting membrane in the macula. Furthermore, the hypothesis that the inner layer separation of the retina, which persists, provides a conduit for the continuous flow of fluid from the pit to the displaced retinal elevation has also been supported.[72,73]

OCT also provides information useful in surgery. For example, it has provided a structural assessment of the macula that was useful in the preoperative and postoperative evaluation of epiretinal membrane surgery.[74] OCT has also been used to document fundus changes after Ar laser photocoagulation treatment of retinal branch vein occlusion,[75] and to evaluate macular thickness changes after cataract surgery.[76]

As a result of its high depth resolution OCT is well suited to measure retinal layer thicknesses such as the nerve fiber layer (NFL). This layer is of relatively high reflectivity owing to its horizontally aligned components.[62] NFL thickness can be reproducibly measured using OCT[77] and has demonstrated a high degree of correlation with the functional status of the optic nerve, as measured by visual field examination.[78] In a study on focal defects of the NFL, OCT yielded a sensitivity of 65% and a specificity of 81%.[79] OCT has been used to detect NFL thinning in eyes with ONH drusen and glaucoma, and appears to be a sensitive and early indicator of NFL thinning.[80,81]

A study of the retinochoroidal structures seen in idiopathic polypoidal choroidal vasculopathy showed that cross-sectional OCT images may increase the understanding of the pathophysiology of this disease.[82] In vivo imaging of choroidal tumors by OCT has not yet been successfully demonstrated. Choroidal or choriocapillary reflectivity was nonspecifically lower than that of normal choroid but did not yield any additional information about tumor histology. OCT may only provide information about the retinal structure overlying prominent tumors and the extent of adjacent retinal detachment.[83] Retinal detachment has also been effectively quantified in central serous chorioretinopathy.[84]

In the anterior segment of the eye, corneal thickness and profile, anterior chamber depth and angle,

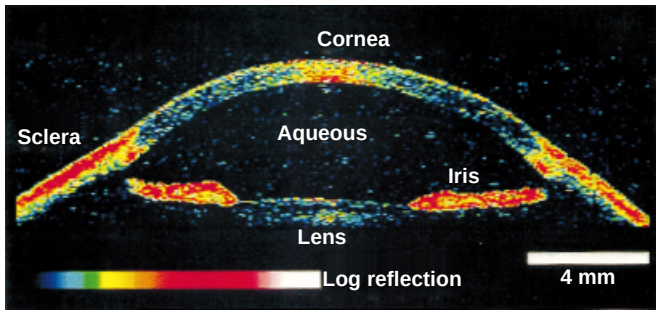

Figure 18 Anterior segment of a human eye. Time-domain OCT, $\lambda = 830\,\text{nm}$, $\Delta\lambda = 20\,\text{nm}$. (Reproduced by permission from Fujimoto et al.[55])

and iris thickness and profile have been measured (Figure 18). Furthermore, dense nuclear cataracts have been successfully imaged through their full thickness in a cold cataract model in calf eyes in vivo.[55,85] OCT has also been used to demonstrate and estimate tissue damages caused by thermokeratoplastical treatment of the cornea.[86,87]

5.2 Other Medical Fields and Optical Coherence Tomography Biopsy Imaging

Conventional biopsy and histopathology relies on microscopic inspection of excised tissue specimens. However, excision of a tissue specimen is often contraindicated or impossible. As OCT provides a resolution comparable to conventional histology, but in real time, it has a high potential for use in real-time optical biopsy.[88] OCT does not require excision of tissue with its associated complications, cost, and delay in obtaining a diagnosis. Moreover, large regions may be surveyed for pathology, and conventional biopsy can be reserved only for those regions in which abnormalities are detected by OCT imaging.

OCT biopsy imaging can be used in animal experiments to noninvasively image developing neural morphology and to assess the developing cardiovascular system.[89,90] Using a 1-mm diameter catheter endoscope, cross-sectional images of the rabbit gastrointestinal and respiratory tracts have been obtained at 10 μm resolution (Figure 19).[91] Images of human in vitro venous,[92] coronary artery,[93] and abdominal aorta[94] morphology have been obtained. The advantage of OCT as compared to intravascular ultrasound is a much better resolution with improved delineation between intimal layer and lipid-rich atherosclerotic plaque.[95]

Intraoperative OCT monitoring has been suggested by a study using a surgical microscope integrated with OCT to perform simultaneous imaging with enface visualization. Cross-sectional images were assembled to produce three-dimensional reconstructions of arterial anastomosis and helped identify sites of luminal obstruction.[96]

Other fields of medical OCT applications are:

- *Dermatology* OCT cross-sections showed different skin layers: stratum corneum, stratum germinativum, stratum papillare, epidermis, papillary ridges, papillary dermis, and sweat gland ducts (Figure 20).[97,98]
- *Gastroenterology* OCT images have demonstrated clear delineation of the mucosa and submucosa and muscularis in gastrointestinal tissues. Microscopic structures such as crypts, blood vessels, or esophagal glands in the submucosa and lymphatic nodules were observed.[99–102]
- *Dentistry* OCT has been shown to be able to differentiate between the various types of keratinized and nonkeratinized mucosa with high resolution. OCT can also provide detailed structural information on clinical abnormalities (caries and noncaries lesions) in teeth and provide guidance in dental restorative procedures.[103] A hand-held in vivo OCT device for use in the oral cavity has also been developed.[104] OCT images showed the gingival margin, periodontal sulcus, and the dento–enamel junction. The cemento–enamel junction was discernible in 64% of the images and the alveolar bone presumptively identified for 71% of the images.
- *Urology* In urological tissues, taken postmortem, microstructure was delineated, including the prostatic urethra, prostate, bladder, and ureter, with an axial resolution of 16 ± 1 μm, higher than any clinically available endoscopic intraluminal imaging

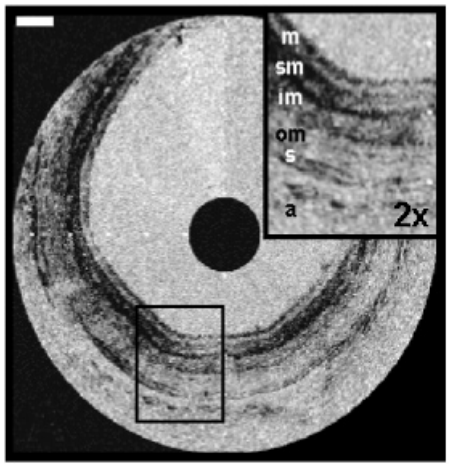

Figure 19 Endoscopic OCT image of the rabbit esophagus in vivo. Kerr-lens modelocked Cr^{4+}/forsterite laser: $\lambda = 1280$ nm; $\Delta\lambda = 75$ nm; bar = 500 μm; m = mucosa; sm = submucosa; im = inner muscular layer; om = outer muscular layer; s = serosa; a = adipose and vascular supportive tissue. (Reprinted with permission from G.J. Tearney, M.E. Brezinski, B.E. Bouma, S.A. Boppart, C. Pitris, F.J. Southern, J.G. Fujimoto, *Science*, **276**, 2037–2039 (1997). Copyright 1999 American Association for the Advancement of Science.)

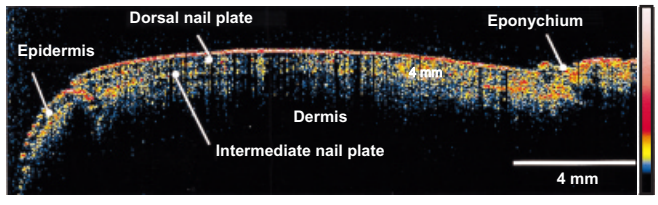

Figure 20 Optical tomogram of a human finger nail. Fourier domain OCT: $\lambda = 850$ nm; $\Delta\lambda = 18$ nm.

technology.[105] OCT provides noncontact high-resolution imaging of urological tissue architectural morphology, without the need for excisional biopsy.

- *Gynecology* OCT imaging of normal and neoplastic human cervical and uterine tissue showed that the structures of the normal ectocervix and endocervix, including epithelium, basal membrane, and glands, could be clearly identified. Because of its high resolution, OCT has the potential for powerful, minimally invasive assessment of the female reproductive tract.[106] Furthermore, micron-scale two- and three-dimensional images obtained of an in vitro oviduct obtained by OCT suggest that it may possible to identify and surgically treat tubal causes of infertility.[107]
- *Neurology* OCT offers increased resolution in the intraoperative identification of brain tumors and tumor margins. Changes in optical backscatter intensity could be used to identify regions of tumor and to locate tumor margins. A comparison of the optical coherence tomographic images with the histological slides showed that OCT can effectively differentiate normal human cortex from intracortical melanoma.[108]

5.3 Future Developments

Recent developments which might lead to important future applications are:

- *ODT* Many diseases are caused by ischemic conditions. Pharmacological intervention and therapy change blood flow. Hence, the quantitative assessment of tissue perfusion is of great importance. Nowadays, ultrasound Doppler techniques are well established in the clinic. Based on OCT principles, ODT offers high spatial resolution. So far, ODT imaging of blood flow dynamics and tissue structures in a vivo chick chorioallantoic membrane and rodent mesentery,[109] and throughout a beating *Xenopus laevis* heart and surrounding vasculature have been demonstrated.[110]
- *Spectroscopic OCT* Here the depth scan is performed stepwise. Each depth scan step signal is Fourier transformed. As in Fourier transform spectroscopy, the result yields the optical spectrum of the backscattered light, but in OCT the spectra obtained are depth localized.[111] A prerequisite for reasonable spectroscopic OCT, however, is a broadband light source delivering spatially coherent light. The related technique of absorption sensitive OCT imaging is based on a dual-wavelength OCT system; this technique has already been used for measuring the local concentration of water in an intralipid phantom.[112,113]
- *Birefringence-sensitive OCT* This technique can be used to characterize biological tissue and to map thermal tissue damage.[114,115] Polarization-sensitive OCT is a related technique.[116] It can be used to characterize dentin and enamel, to determine the birefringence and the optical axis in fibrous tissue, and may also find applications in material science for the investigation of polarization properties in opaque media such as ceramics and crystals.[117,118]

ACKNOWLEDGMENTS

Thanks are due to my colleagues at the Institute of Medical Physics. Our work is based on projects financed by the Austrian Science Foundation (FWF projects P7300-MED, P9781-MED, and P10316-MED).

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
AMD	Age-related Macular Disease
APD	Silicon Avalanche Photodiode
CDOCT	Color Doppler Optical Coherence Tomography
CNV	Choroidal Neovascularization
CT	Computed Tomography
DOF	Depth of Focus
fwhm	Full Width at Half-maximum
IR	Infrared
NFL	Nerve Fiber Layer
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
OCT	Optical Coherence Tomography
ODT	Optical Doppler Tomography
PCI	Partial Coherence Interferometry
PIN	p-Doped, Intrinsic Semiconductor, with n-Doped Layers
SLD	Superluminescent Semiconductor Diodes

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Particle Size Analysis **(Volume 6)**
Velocimetry in Particle Size Analysis

Surfaces **(Volume 10)**
Ellipsometry in Analysis of Surfaces and Thin Films

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Near-infrared Absorption/Luminescence Measurements

Infrared Spectroscopy **(Volume 12)**
Infrared Reflection–Absorption Spectroscopy • Microspectroscopy

REFERENCES

1. A.C. Kak, M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, 1988.
2. E. Wolf, 'Principles and Development of Diffraction Tomography', in *Trends in Optics*, ed. A. Consortini, Academic Press, San Diego & London, 83–110, 1996.
3. A. Yodh, B. Chance, 'Spectroscopy and Imaging with Diffusing Light', *Phys. Today*, **48**, 34–40 (1995).
4. D. Huang, E.A. Swanson, C.P. Lin, J.S. Schuman, W.G. Stinson, W. Chang, M.R. Hee, T. Flotte, K. Gregory, C.A. Puliafito, J.G. Fujimoto, 'Optical Coherence Tomography', *Science*, **254**, 1178–1181 (1991).
5. A.F. Fercher, 'Optical Coherence Tomography', *J. Biomed. Opt.*, **1**, 157–173 (1996).
6. J.A. Parrish, 'New Concepts in Therapeutic Photomedicine: Photochemistry, Optical Targeting and the Therapeutic Window', *J. Invest. Dermatol.*, **77**, 45–50 (1981).
7. J.-L. Boulnois, 'Photophysical Processes in Recent Medical Laser Developments: A Review', *Lasers Med. Sci.*, **1**, 47–66 (1986).
8. H.P. Berlien, G. Müller, 'Angewandte Lasermedizin', *Ecomed.*, (1989).
9. K.M. Yoo, R.R. Alfano, 'Time-resolved Coherent and Incoherent Components of Forward Light Scattering in Random Media', *Opt. Lett.*, **15**, 320–322 (1990).
10. R.R. Alfano (ed.), 'Advances in Optical Biopsy and Optical Mammography', *Ann. N. Y. Acad. Sci.*, **838**, 14–28 (1998).
11. C.M. Vest, 'Formation of Images from Projection: Radon and Abel Transforms', *J. Opt. Soc. Am.*, **64**, 1215–1218 (1974).
12. S.R. Arridge, P. van der Zee, M. Cope, D.T. Delpy, 'Reconstruction Methods for Infrared Absorption Imaging', *Proc. SPIE*, **1431**, 204–215 (1991).
13. S.R. Arridge, J.C. Hebden, 'Optical Imaging in Medicine: II. Modelling and Reconstruction', *Phys. Med. Biol.*, **42**, 841–853 (1997).
14. E.A. Swanson, J.A. Izatt, M.R. Hee, D. Huang, C.P. Lin, J.S. Schuman, C.A. Puliafito, J.G. Fujimoto, 'In vivo Retinal Imaging using Optical Coherence Tomography', *Opt. Lett.*, **18**, 1864–1866 (1993).
15. A.F. Fercher, C.K. Hitzenberger, W. Drexler, G. Kamp, H. Sattmann, 'In Vivo Optical Coherence Tomography', *Am. J. Ophthalmol.*, **116**, 113–114 (1993).
16. A.F. Fercher, E. Roth, 'Ophthalmic Laser Interferometry', *Proc. SPIE*, **658**, 48–51 (1986).
17. A. Gerrard, J.M. Burch, *Introduction to Matrix Optics Methods in Optics*, J. Wiley & Sons, London, 1975.
18. E. Wolf, 'Three-dimensional Structure Determination of Semi-transparent Objects from Holographic Data', *Opt. Commun.*, **1**, 153–156 (1969).
19. A.F. Fercher, H. Bartelt, H. Becker, E. Wiltschko, 'Image Formation by Inversion of Scattered Field Data: Experiments and Computational Simulation', *Appl. Opt.*, **18**, 2427–2439 (1979).
20. A.F. Fercher, C.K. Hitzenberger, G. Kamp, S.Y. El-Zaiat, 'Measurement of Intraocular Distances by Backscattering Spectral Interferometry', *Opt. Commun.*, **117**, 43–48 (1995).
21. R. Dändliker, K. Weiss, 'Reconstruction of the Three-dimensional Refractive Index from Scattered Waves', *Opt. Commun.*, **1**, 323–328 (1970).
22. J.M. Cowley, *Diffraction Physics*, North Holland, Amsterdam, 1975.
23. R. Loudon, *The Quantum Theory of Light*, Clarendon Press, Oxford, 1985.
24. M. Born, E. Wolf, *Principles of Optics*, Pergamon Press, New York, 1980.
25. P. Hariharan, *Optical Interferometry*, Australia Academic Press, Sydney, 1985.
26. C.K. Hitzenberger, W. Drexler, A. Baumgartner, A.F. Fercher, 'Dispersion Effects in Partial Coherence Interferometry', *Proc. SPIE*, **2981**, 29–36 (1997).
27. C.K. Hitzenberger, A. Baumgartner, W. Drexler, A.F. Fercher, 'Dispersion Effects in Partial Coherence Interferometry: Implications for Intraocular Ranging', *J. Biomed. Opt.*, **4**, 144–151 (1999).
28. A.F. Fercher, R. Leitgeb, C.K. Hitzenberger, H. Sattmann, M. Wojtkowski, 'Complex Spectral Interferometry OCT', *Proc. SPIE*, **3564**, 173–178 (1999).
29. J. Schmit, K. Creath, 'Extended Averaging Technique for Derivation of Error-compensating Algorithms in Phase-shifting Interferometry', *Appl. Opt.*, **34**, 3610–3619 (1995).
30. U.H.P. Haberland, V. Blazek, H.J. Schmitt, 'Chirp Optical Coherence Tomography of Layered Scattering Media', *J. Biomed. Opt.*, **3**, 259–266 (1998).
31. Z. Chen, T.E. Milner, X. Wang, S. Srinivas, J.S. Nelson, 'Optical Doppler Tomography: Imaging in vivo Blood Flow Dynamics Following Pharmacological Intervention and Photodynamic Therapy', *Photochem. Photobiol.*, **67**, 1–7 (1998).
32. J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1969.
33. E.A. Swanson, D. Huang, M.R. Hee, J.G. Fujimoto, C.P. Lin, C.A. Puliafito, 'High-speed Optical Coherence Domain Reflectometry', *Opt. Lett.*, **17**, 151–153 (1992).

34. J.A. Izatt, M.R. Hee, G.M. Owen, E.A. Swanson, J.G. Fujimoto, 'Optical Coherence Microscopy in Scattering Media', *Opt. Lett.*, **19**, 590–592 (1994).
35. A.Gh. Podoleanu, G.M. Dobre, D.J. Webb, D.A. Jackson, 'Coherence Imaging by use of a Newton Rings Sampling Function', *Opt. Lett.*, 1789–1791 (1996).
36. J.M. Schmitt, S.L. Lee, K.M. Yung, 'An Optical Coherence Microscope with Enhanced Resolving Power in Thick Tissue', *Opt. Commun.*, **142**, 203–207 (1997).
37. F. Lexer, A.F. Fercher, H. Sattmann, W. Drexler, S. Molebny, 'Dynamic Coherent Focus for Transversal Resolution Enhancement of OCT', *Proc. SPIE*, **3251**, 85–90 (1998).
38. J.A. Izatt, M.D. Kulkarni, S. Yazdanfar, J.K. Barton, A.J. Welch, 'In vivo Bidirectional Color Doppler Flow Imaging of Picoliter Blood Volumes using Optical Coherence Tomography', *Opt. Lett.*, **22**, 1439–1441 (1997).
39. S. Yazdanfar, M.D. Kulkarni, J.A. Izatt, 'High Resolution Imaging of in vivo Cardiac Dynamics using Color Doppler Optical Coherence Tomography', *Opt. Express*, **1**, 424–431 (1997).
40. G.J. Tearney, S.A. Boppart, B.E. Bouma, M.E. Brezinski, N.J. Weissman, J.F. Southern, J.G. Fujimoto, 'Scanning Single-mode Fiber Optic Catheter-endoscope for Optical Coherence Tomography', *Opt. Lett.*, **21**, 543–545 (1996).
41. H. Moseley, *Non-ionising Radiation*, Adam Hilger, Bristol and Philadelphia, 1988.
42. X. Clivaz, F. Marquis-Weible, R.-P. Salathe, '1.5 μm Resolution Optical Low Coherence Reflectometry in Biological Tissues', *Proc. SPIE*, **2083**, 338–346 (1993).
43. B.E. Bouma, G.J. Tearney, S.A. Boppart, M.R. Hee, M.E. Brezinski, J.G. Fujimoto, 'High Resolution Optical Coherence Tomographic Imaging using a Mode-locked $Ti:Al_2O_3$ Laser Source', *Opt. Lett.*, **20**, 1486–1488 (1995).
44. A.T. Semenov, V.K. Batovrin, I.A. Garmash, V.R. Shidlovsky, M.V. Shramenko, S.D. Yakubovich, '(GaAl)As SQW Superluminescent Diodes with Extremely Low Coherence Length', *Electron. Lett.*, **31**, 314–315 (1995).
45. D.R. Herriott, 'Some Applications of Lasers to Interferometry', in *Progress in Optics*, ed. E. Wolf, North Holland, 173–209, 1967.
46. G.J. Tearney, B.E. Bouma, S.A. Boppart, B.S.E.A. Golubovic, 'Rapid Acquisition of in vivo Biological Images by use of Optical Coherence Tomography', *Opt. Lett.*, **21**, 1408–1410 (1996).
47. C.B. Su, 'Achieving Variation of the Optical Pathlength by a Few Millimeters at Millisecond Rates for Imaging of Turbid Media and Optical Interferometry: A New Technique', *Opt. Lett.*, **22**, 665–667 (1997).
48. J. Szydlo, N. Delachenal, R. Gianotti, R. Wälti, H. Bleuler, R.P. Salathé, 'Air-turbine Driven Optical Low-coherence Reflectometry at 28.6-kHz Scan Repetition Rate', *Opt. Commun.*, **154**, 1–4 (1994).
49. G.J. Tearney, B.E. Bouma, J.G. Fujimoto, 'High-speed Phase- and Group-delay Scanning with a Grating-based Phase Control Delay Line', *Opt. Lett.*, **22**, 1811–1813 (1997).
50. A. McGregor, 'Silicone Avalanche Photodiodes for Low-light, High-speed Systems', in *The Photonics Directory*, eds. R.R. Afano et al., Laurin, Pittsfield, MA, H-110–H-112, 1992.
51. J.P. Dunkers, R.S. Parnas, C.G. Zimba, R.C. Peterson, K.M. Flynn, J.G. Fujimoto, B.E. Bouma, 'Optical Coherence Tomography of Glass Reinforced Polymer Composites', *Composites – Part A: Appl. Sci. Manufact.*, **30A**, 139–145 (1999).
52. M. Bashkansky, M.D. Duncan, M. Kahn, D. Lewis III, J. Reintjes, 'Subsurface Defect Detection in Ceramics by High-speed High-resolution Optical Coherence Tomography', *Opt. Lett.*, **22**, 61–63 (1997).
53. A.F. Fercher, C.K. Hitzenberger, W. Drexler, G. Kamp, H. Sattmann, 'In vivo Optical Coherence Tomography', *Am. J. Ophthalmol.*, **116**, 113–114 (1993).
54. E.A. Swanson, J.A. Izatt, M.R. Hee, D. Huang, C.P. Lin, J.S. Schuman, C.A. Puliafito, J.G. Fujimoto, 'In vivo Retinal Imaging using Optical Coherence Tomography', *Opt. Lett.*, **18**, 1864–1866 (1993).
55. J.A. Izatt, M.R. Hee, D. Huang, J.G. Fujimoto, E.A. Swanson, C.P. Lin, J.S. Schuman, C.A. Puliafito, 'Ophthalmic Diagnostics using Optical Coherence Tomography', *Proc. SPIE*, **1877**, 136–144 (1993).
56. N. Tanno, S. Kishi, 'Optical Coherence Tomographic Imaging and Clinical Diagnosis', *Medical Imaging Technol.*, **17**, 3–10 (1999).
57. M.R. Hee, J.A. Izatt, E.A. Swanson, D. Huang, J.S. Schuman, C.P. Lin, C.A. Puliafito, J.G. Fujimoto, 'Optical Coherence Tomography of the Human Retina', *Arch. Ophthalmol.*, **113**, 325–332 (1995).
58. M.R. Hee, C.A. Puliafito, C. Wong, E. Reichel, J.S. Duker, J.S. Schuman, E.A. Swanson, J.G. Fujimoto, 'Optical Coherence Tomography of Macular Holes', *Ophthalmology*, **102**, 748–756 (1995).
59. G. Trabucchi, R. Brancato, L. Pierro, U. Introini, C. Sannace, 'Idiopathic Juxtafoveolar Retinal Telangiectasis and Pigment Epithelial Hyperplasia: An Optical Coherence Tomographic Study', *Arch. Ophthalmol.*, **117**, 405–406 (1999).
60. M. Ip, C. Garza-Karren, J.S. Duker, E. Reichel, J.C. Swartz, A. Amirikia, C.A. Puliafito, 'Differentiation of Degenerative Retinoschisis from Retinal Detachment using Optical Coherence Tomography', *Ophthalmology*, **106**, 600–605 (1999).
61. G. Trabucchi, C. Sannace, U. Introini, R. Brancato, 'Partial Lipodystrophy with Associated Fundus Abnormalities: An Optical Coherence Tomography Study', *Br. J. Ophthalmol.*, **82**, 326 (1998).
62. W. Drexler, C.K. Hitzenberger, H. Sattmann, A.F. Fercher, 'Measurement of the Thickness of Fundus

Layers by Partial Coherence Tomography', *Opt. Eng.*, **34**, 701–710 (1995).
63. M. Baumann, R.C. Gentile, J.M. Liebmann, R. Ritch, 'Reproducibility of Retinal Thickness Measurements in Normal Eyes using Optical Coherence Tomography', *Ophthalmic Surgery and Lasers*, **29**, 280–285 (1998).
64. C. Toth, D.G. Narayan, S.A. Boppart, M.R. Hee, J.G. Fujimoto, R. Birngruber, C.P. Cain, C.D. DiCarlo, W.P. Roach, 'A Comparison of Retinal Morphology Viewed by Optical Coherence Tomography and by Light Microscopy', *Arch. Ophthalmol.*, **115**, 1425–1428 (1997).
65. Y. Huang, A.V. Cideciyan, G.I. Papastergiou, E. Banin, S.L. Semple-Rowland, A.H. Milam, S.G. Jacobson, 'Relation of Optical Coherence Tomography to Microanatomy in Normal and rd Chickens', *Invest. Ophthalmol. Vis. Sci.*, **39**, 2405–2416 (1998).
66. C.A. Puliafito, M.R. Hee, C.P. Lin, E. Reichel, J.S. Schuman, J.S. Duker, J.A. Izatt, E.A. Swanson, J.G. Fujimoto, 'Imaging of Macular Diseases with Optical Coherence Tomography', *Ophthalmology*, **102**, 217–229 (1995).
67. E. Pal, G. Givort, A. Laroche, P.O. Barale, S. Limon, M. Ullern, 'Macular Imaging with Optical Coherence Tomography', *J. Fr. Ophthalmol.*, **21**, 484–494 (1998).
68. M. Hee, C.A. Puliafito, J.S. Duker, E. Reichel, J.G. Coker, J.R. Wilkins, J.S. Schuman, E.A. Swanson, J.G. Fujimoto, 'Topography of Diabetic Macular Edema with Optical Coherence Tomography', *Ophthalmology*, **105**, 360–370 (1998).
69. M.R. Hee, C.A. Puliafito, C. Wong, J.S. Duker, E. Reichel, B. Rutledge, J.S. Schuman, E.A. Swanson, J.G. Fujimoto, 'Quantitative Assessment of Macular Edema with Optical Coherence Tomography', *Arch. Ophthalmol.*, **113**, 1019–1029 (1995).
70. C.W. Spraul, G.E. Lang, G.K. Lang, 'Die Bedeutung der optischen Kohärenz-Tomographie in der Diagnostik der altersbezogenen Makuladegeneration. Korrelation von fluoreszenzangiographischen mit OCT-Befunden', *Klein. Monatsbl. Augenheilkd.*, **212**, 141–148 (1998).
71. M.R. Hee, C.R. Baumal, C.A. Puliafito, J.S. Duker, E. Reichel, J.R. Wilkins, J.G. Coker, J.S. Schuman, E.A. Swanson, J.G. Fujimoto, 'Optical Coherence Tomography of Age-related Macular Degeneration and Choroidal Neovascularization', *Ophthalmol.*, **103**, 1260–1270 (1996).
72. H. Lincoff, I. Kreissig, 'Optical Coherence Tomography of Pneumatic Displacement of Optic Disc Pit Maculopathy', *Br. J. Ophthalmol.*, **82**, 367–372 (1998).
73. D. Krivoy, R. Gentile, J.M. Liebmann, Z. Stegman, J.B. Walsh, R. Ritch, 'Imaging Congenital Optic Disc Pits and Associated Maculopathy using Optical Coherence Tomography', *Arch. Ophthalmol.*, **114**, 165–170 (1996).
74. J.R. Wilkins, C.A. Puliafito, M.R. Hee, J.S. Duker, E. Reichel, J.G. Coker, J.S. Schuman, E.A. Swanson, J.G. Fujimoto, 'Characterization of Epiretinal Membranes using Optical Coherence Tomography', *Ophthalmol.*, **103**, 2142–2151 (1996).
75. N. Christoffersen, B. Sander, M. Larsen, 'Precipitation of Hard Exudate after Resorption of Intraretinal Edema after Treatment of Retinal Branch Vein Occlusion', *Am. J. Ophthalmol.*, **126**, 454–456 (1998).
76. P. Sourdille, P.Y. Santiago, 'Optical Coherence Tomography of Macular Thickness after Cataract Surgery', *J. Cataract. Refract. Surg.*, **25**, 256–261 (1999).
77. J.S. Schuman, T. Pedut-Kloizman, E. Hertzmark, M.R. Hee, J.R. Wilkins, J.G. Coker, C.A. Puliafito, E.A. Swanson, 'Reproducibility of Nerve Fiber Layer Thickness Measurements using Optical Coherence Tomography', *Ophthalmol.*, **103**, 1889–1898 (1996).
78. J.S. Schuman, M.R. Hee, C.A. Puliafito, C. Wong, T. Pedut-Kloizman, C.P. Lin, E. Hertzmark, J.A. Izatt, E.A. Swanson, J.G. Fujimoto, 'Quantification of Nerve Fiber Layer Thickness in Normal and Glaucomatous Eyes using Optical Coherence Tomography', *Arch. Ophthalmol.*, **113**, 586–596 (1995).
79. L. Pieroth, J.S. Schuman, E. Hertzmark, M.R. Hee, J.R. Wilkins, J. Coker, C. Mattox, R. Pedut-Kloizman, C.A. Puliafito, J.G. Fujimoto, E. Swanson, 'Evaluation of Focal Defects of the Nerve Fiber Layer using Optical Coherence Tomography', *Ophthalmology*, **106**, 570–579 (1999).
80. S. Roh, R.J. Noecker, J.S. Schuman, 'Evaluation of Coexisting Optic Nerve Head Drusen and Glaucoma with Optical Coherence Tomography', *Ophthalmol.*, **107**, 1138–1144 (1997).
81. S. Roh, R.J. Noecker, J.S. Schuman, I.T.R. Hedges, J.J. Weiter, C. Mattox, 'Effect of Optic Nerve Head Drusen on Nerve Fiber Layer Thickness', *Ophthalmol.*, **105**, 878–885 (1998).
82. H. Iijima, M. Imai, T. Gohdo, S. Tsukahara, 'Optical Coherence Tomography of Idiopathic Polypoidal Choroidal Vasculopathy', *Am. J. Ophthalmol.*, **127**, 301–305 (1999).
83. U. Schaudig, A. Hassenstein, A. Bernd, A. Walter, G. Richard, 'Limitations of Imaging Choroidal Tumors in vivo by Optical Coherence Tomography', *Graefes Arch. Clin. Exp. Ophthalmol.*, **236**, 588–592 (1998).
84. M.R. Hee, C.A. Puliafito, C. Wong, E. Reichel, J.S. Duker, J.S. Schuman, E.A. Swanson, J.G. Fujimoto, 'Optical Coherence Tomography of Central Serous Chorioretinopathy', *Am. J. Ophthalmol.*, **120**, 65–74 (1995).
85. J.A. Izatt, M.R. Hee, E.A. Swanson, C.P. Lin, D. Huang, J.S. Schuman, C.A. Puliafito, J.G. Fujimoto, 'Micrometer-scale Resolution Imaging of the Anterior Eye in vivo with Optical Coherence Tomography', *Arch. Ophthalmol.*, **112**, 1584–1589 (1994).

86. N. Koop, R. Brinkmann, E. Lankenau, S. Flache, R. Engelhardt, R. Birngruber, 'Optische Kohärenztomographie der Kornea und des vorderen Augenabschnitts', *Ophthalmology*, **94**, 481–486 (1997).
87. M.N. Asiyo-Vogel, N. Koop, R. Brinkmann, R. Engelhardt, R. Eggers, R. Birngruber, A. Vogel, 'Darstellung von LTK-Läsionen durch optische Kurzkohärenztomographie (OCT) und Polarisationsmikroskopie nach Sirius-Rot-Färbung', *Ophthalmology*, **94**, 487–491 (1997).
88. J.G. Fujimoto, M.E. Brezinski, G.J. Tearney, S.A. Boppart, B. Bouma, M.R. Hee, J.F. Southern, E.A. Swanson, 'Optical Biopsy and Imaging using Optical Coherence Tomography', *Nat. Med.*, **1**, 970–972 (1995).
89. S.A. Boppart, B.E. Bouma, M.E. Brezinski, G.J. Tearney, J.G. Fujimoto, 'Imaging Developing Neural Morphology using Optical Coherence Tomography', *J. Neurosci. Methods*, **70**, 65–72 (1996).
90. S.A. Boppart, G.J. Tearney, B.E. Bouma, J.F. Southern, M.E. Brezinski, J.G. Fujimoto, 'Noninvasive Assessment of the Developing Xenopus Cardiovascular System using Optical Coherence Tomography', *Proc. Natl. Acad. Sci. USA*, **94**, 4256–4261 (1997).
91. G.J. Tearney, M.E. Brezinski, B.E. Bouma, S.A. Boppart, C. Pitris, J.F. Southern, J.G. Fujimoto, 'In vivo Endoscopic Optical Biopsy with Optical Coherence Tomography', *Science*, **276**, 2037–2039 (1997).
92. G.J. Tearney, S.A. Boppart, B.E. Bouma, M.E. Brezinski, N.J. Weissman, F.J. Southern, J.G. Fujimoto, 'Scanning Single-mode Fiber Optic Catheter-endoscope for Optical Coherence Tomography', *Opt. Lett.*, **21**, 543–545 (1996).
93. M.E. Brezinski, G.J. Tearney, B.E. Bouma, J.A. Izatt, M.R. Hee, E.A. Swanson, J.F. Southern, J.G. Fujimoto, 'Optical Coherence Tomography for Optical Biopsy. Properties and Demonstration of Vascular Pathology', *Circulation*, **93**, 1206–1213 (1996).
94. M.E. Brezinski, G.J. Tearney, N.J. Weissman, S.A. Boppart, B.E. Bouma, M.R. Hee, A.E. Weyman, E.A. Swanson, J.F. Southern, J.G. Fujimoto, 'Assessing Atherosclerotic Plaque Morphology: Comparison of Optical Coherence Tomography and High Frequency Intravascular Ultrasound', *Heart*, **77**, 397–403 (1997).
95. F.X. Kleber, S. Döpfmer, T. Thieme, 'Invasive Strategies to Discriminate Stable and Unstable Coronary Plaques', *Eur. Heart J.*, **19**, C44–C49 (1998).
96. S.A. Boppart, B.E. Bouma, C. Pitris, G.J. Tearney, J.F. Southern, M.E. Brezinski, J.G. Fujimoto, 'Intraoperative Assessment of Microsurgery with Three-dimensional Optical Coherence Tomography', *Radiology*, **208**, 81–86 (1998).
97. J. Welzel, E. Lankenau, R. Birngruber, R. Engelhardt, 'Optical Coherence Tomography of the Human Skin', *J. Am. Acad. Dermatol.*, **37**, 958–963 (1997).
98. G. Häusler, M.W. Lindner, ' ''Coherence RADAR'' and ''spectral RADAR'' – New Tools for Dermatological Diagnosis', *J. Biomed. Opt.*, **3**, 21–31 (1998).
99. K. Kobayashi, J.A. Izatt, M.D. Kulkarni, J. Willis, M.V. Sivak, Jr, 'High-resolution Cross-sectional Imaging of the Gastrointestinal Tract using Optical Coherence Tomography: Preliminary Results', *Gastrointest. Endosc.*, **47**, 515–523 (1998).
100. G.J. Tearney, M.E. Brezinski, B.E. Bouma, S.A. Boppart, C. Pitris, J.F. Southern, J.G. Fujimoto, 'Optical Coherence Tomography: Lighting the Way Toward Optical Biopsy', *Gastroenterology*, **115**, 497–504 (1998).
101. G.J. Tearney, M.E. Brezinski, J.F. Southern, B.E. Bouma, S.A. Boppart, J.G. Fujimoto, 'Optical Biopsy in Human Gastrointestinal Tissue using Optical Coherence Tomography', *Am. J. Gastroent.*, **92**, 1800–1804 (1997).
102. M.B. Wallace, J. Van-Dam, 'Optical Coherence Tomography: Lighting the Way Around Optical Biopsy', *Gastroenterology*, **115**, 497–498 (1998).
103. F.I. Feldchtein, G.V. Gelikonov, V.M. Gelikonov, R.R. Iksanov, R.V. Kuranov, A.M. Sergeev, M.D. Gladkova, M.N. Ourutina, J.A. Warren, D.H. Reitze, 'In vivo OCT Imaging of Hard and Soft Tissue of the Oral Cavity', *Opt. Express*, **3**, 239–250 (1998).
104. B.W. Colston, M.J. Everett, L.B. DaSilva, 'OCT for Diagnosis of Periodontal Disease', *Proc. SPIE*, **3251**, 52–58 (1998).
105. G.J. Tearney, M.E. Brezinski, J.F. Southern, B.E. Bouma, S.A. Boppart, J.G. Fujimoto, 'Optical Biopsy in Human Urologic Tissue using Optical Coherence Tomography', *J. Urol.*, **157**, 1915–1919 (1997).
106. C. Pitris, A. Goodman, S.A. Boppart, J. Libus, J.G. Fujimoto, M.E. Brezinski, 'High-resolution Imaging of Gynecologic Neoplasms using Optical Coherence Tomography', *Obstet-Gynecol.*, **93**, 135–139 (1999).
107. J.M. Herrmann, M.E. Brezinski, B.E. Bouma, S.A. Boppart, C. Pitris, J.F. Southern, J.G. Fujimoto, 'Two- and Three-dimensional High-resolution Imaging of the Human Oviduct with Optical Coherence Tomography', *Fertility and Sterility*, **70**, 155–158 (1998).
108. S.A. Boppart, M.E. Brezinski, C. Pitris, J.G. Fujimoto, 'Optical Coherence Tomography for Neurosurgical Imaging of Human Intracortical Melanoma', *Neurosurgery.*, **43**, 834–841 (1998).
109. Z. Chen, T.E. Milner, X. Wang, S. Srinivas, J.S. Nelson, 'Optical Doppler Tomography: Imaging in vivo Blood Flow Dynamics following Pharmacological Intervention and Photodynamic Therapy', *Photochem. Photobiol.*, **67**, 1–7 (1998).
110. S. Yazdanfar, M.D. Kulkarni, J.A. Izatt, 'High Resolution Imaging of in vivo Cardiac Dynamics using Color Doppler Optical Coherence Tomography', *Opt. Express*, **1**, 424–431 (1997).
111. M. Kulkarni, J.A. Izatt, 'Spectroscopic Optical Coherence Tomography', *CLEO'96 Tech. Digest*, **9**, 59–60 (1996).

112. J.M. Schmitt, S.H. Xiang, K.M. Yung, 'Differential Absorption Imaging with Optical Coherence Tomography', *J. Opt. Soc. Am. A*, **15**, 2288–2296 (1998).
113. U.S. Sathyam, B.W. Colston, Jr, L.B. Da Silva, M.J. Everett, 'Evaluation of Optical Coherence Quantitation of Analytes in Turbid Media by use of Two Wavelengths', *Appl. Opt.*, **38**, 2097–2104 (1999).
114. M.J. Everett, K. Schoenenberger, B.W. Colston, L.B. Da Silva, 'Birefringence Characterization of Biological Tissue by use of Optical Coherence Tomography', *Opt. Lett.*, **23**, 228–230 (1998).
115. B.W. Colston, Jr, D.J. Maitland, L.B. Da Silva, M.J. Everett, K. Schoenenberger, 'Mapping of Birefringence and Thermal Damage in Tissue by use of Polarization-sensitive Optical Coherence Tomography', *Appl. Opt.*, **37**, 6026–6036 (1998).
116. J.M. Schmitt, S.H. Xiang, 'Cross-polarized Backscatter in Optical Coherence Tomography of Biological Tissue', *Opt. Lett.*, **23**, 1060–1062 (1998).
117. Xiao-Jun Wang, T.E. Milner, J.F. de Boer, Yi Zhang, D.H. Pashley, J.S. Nelson, 'Characterization of Dentin and Enamel by use of Optical Coherence Tomography', *Appl. Opt.*, **38**, 2092–2096 (1999).
118. J.F. De Boer, T.E. Milner, J.S. Nelson, 'Determination of the Depth-resolved Stokes Parameters of Light Backscattered from Turbid Media by use of Polarization-sensitive Optical Coherence Tomography', *Opt. Lett.*, **24**, 300–302 (1999).

Photodynamic Therapy

Beate Röder
Humboldt-University Berlin, Institute of Physics, Berlin, Germany

1 Introduction 1
2 History 2
3 Molecular Mechanism of Photodynamic Therapy 2
3.1 Types I, II, and III of Photosensitization 2
3.2 Photophysical Properties of Photosensitizers 3
3.3 Singlet Molecular Oxygen 3
4 Principle of Photodynamic Therapy 4
4.1 Photochemotherapy (Type I/III of Photosensitization) 5
4.2 Photodynamic Treatment (Type II Photosensitization) 5
4.3 Photoactivity from Higher Excited States of Photosensitizers 5
5 Photosensitizers 6
5.1 Exogenous: First, Second, and Third Generation of Photosensitizers 6
5.2 Endogenous: 5-Aminolevulinic Acid Therapy 10
6 Drug Delivery Systems 10
6.1 Carrier Molecules 10
6.2 Carrier Systems 11
6.3 Modular Carrier Systems 11
7 Common Light Sources 11
7.1 Lasers and Light-emitting Diodes 12
7.2 Lamps 12
7.3 Pulsed Lasers 12
8 Clinical Applications 12
8.1 Photodynamic Therapy of Skin Diseases and Cancer 12
8.2 Photodynamic Therapy of Other Diseases 13
9 Comparison and Combination with Other Methods 13
9.1 Chemotherapy 13
9.2 Radiotherapy 13
9.3 Hyperthermia 14
9.4 Cryotherapy 14
9.5 Immunotherapy 14
10 Outlook 14
Abbreviations and Acronyms 14
Related Articles 14
References 15

Photodynamic therapy (PDT) is an effective treatment of different types of cancer, virus infections, skin diseases and others using photosensitizers (PSs) which are nontoxic in the dark but become phototoxic following activation by low energy light. The PS is applied topically or systemically and should be selectively accumulated in the target tissue. The molecular mechanism of the therapy is the photodynamic effect: following light absorption energy is transferred from the PS to molecular oxygen generating singlet molecular oxygen ($^1O_2$) which causes direct or indirect photodamage of the target tissue.

In clinical therapy with first generation PSs like Photofrin® typically fluences of around 100 J cm^{-2} of red light are used. The light doses vary in dependence on the treated disease. Using PSs of second generation with very strong absorption in the red region like the very efficient m-tetrahydroxyphenylchlorin (m-THPC), smaller fluences, typically around 10 J cm^{-2}, are required. According to the different optical properties and the efficiency of selective accumulation of PS in the target tissue, the drug doses vary from 1–5 mg (first generation PS) to 0.1 mg (second generation PS) per kilogram body weight.

PDT is a very efficient method for the treatment of minor diseases, blood cancer, inoperable diseases like some types of bladder, brain or lung cancer, and breast carcinoma metastases. The side effects are less than in the case of chemo- or radiotherapy. The most common side effect is skin photosensitization. The strength of this effect depends on the PS used and also on the patients' skin type. Most of the new second generation PSs cause no or only weak skin phototoxicity. The therapeutic radius is limited by the penetration depth of the used light (in the order of some millimeters to centimeters). The method will not be very successful for the treatment of large solid tumors.

1 INTRODUCTION

Today it is generally accepted that PDT in the presence of a PS – in most cases these are tetrapyrroles – activated by red light (Figure 1) is an effective treatment for various types of cancer, skin diseases and virus infections. The most commonly used sensitizer in clinical therapy is a complex mixture of porphyrins, derived from hematoporphyrin, termed hematoporphyrin derivative (HPD). This PS is far from ideal, its tumor selectivity is very poor,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

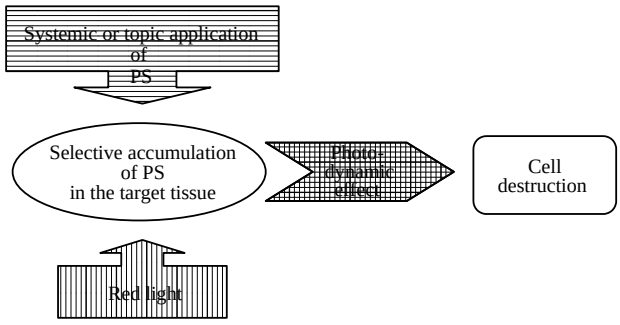

Figure 1 Principle of PDT.

and it has only a weak molar absorption coefficient in the red region of the spectrum where tissue penetration is high. In addition, the skin photosensitization caused by hematoporphyrin may continue up to 30 days as has been known since 1913.[1]

For these reasons in recent years beside the so-called first generation PS new compounds with strong photosensitizing properties, high absorption in the far red and less side effects, the so-called second generation PS were developed. Presently scientific interest is focused on the development of carrier systems that increase the selectivity of PS accumulation in the target tissue. For this photophysically optimized PS of the second generation are used to design together with units which direct them to biologically well defined places in the cells or on cell surfaces new drugs for PDT – the so-called third generation PS.

2 HISTORY

The photosensitizing activity has been known since the beginning of this century.[2] For the first time in 1903 the use of combined dye and illumination for the treatment of skin diseases was proposed.[3] These first reports were followed by broad research activity in the 1940s and 1950s with a lot of interesting results.[4,5] Nevertheless the use of the photodynamic effect was limited to skin diseases at this time.

In the mid-1970s this ''soft'' noninvasive therapy became of greater interest in medicine because of the very fast development of laser and fiber optics.[6,7] Using these new techniques it was possible now to treat all surfaces in the human body attainable with fiber optics of a diameter of some hundreds of micrometers.

3 MOLECULAR MECHANISM OF PHOTODYNAMIC THERAPY

The molecular mechanism of the photodynamic effect has been under investigation for a long time and a first scheme

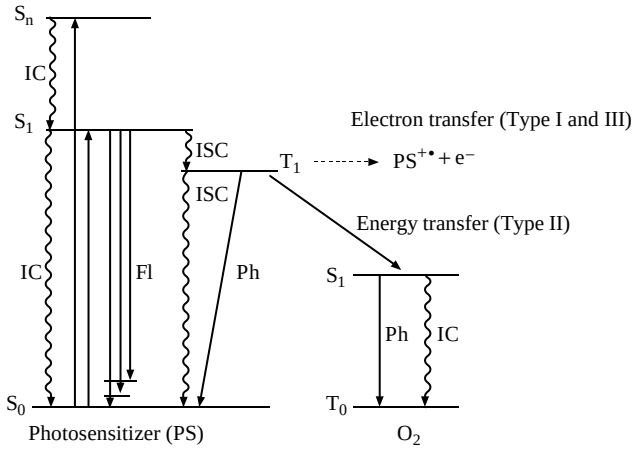

Figure 2 Primary molecular processes of photosensitization (Fl: fluorescence; Ph: phosphorescence; IC: internal conversion; ISC: intersystem crossing; S: singlet state; T: triplet state).

was given by Foote.[8] He postulated two types of photosensitization both occurring in the presence of molecular oxygen. The *Type I* process includes charge transfer and the *Type II* pathway energy transfer (EnT) processes. Later a third way of photosensitization the *Type III* which occurs in the absence of molecular oxygen and includes charge transfer was postulated by Laustriat.[9] Accepting this broad definition of photosensitization also photoinduced electron transfer processes (e.g. photosynthesis) can be treated as photosensitized reactions. The processes relevant for photosensitization occurring after light absorption by the PS molecule are summarized in Figure 2. After absorption of a light quantum the PS is excited directly to the first excited singlet state (S_1) or passes from a higher excited singlet state (S_n) to the S_1-state. In general energy and electron transfer processes can occur from this excited electronic state. Nevertheless the most important step in photosensitization is the process of intersystem crossing (ISC) converting the PS to the long-lived first excited triplet state (T_1). Because of the special properties of molecular oxygen (see section 3.3) a very efficient EnT from the PS in the T_1-state may occur resulting in the generation of singlet molecular oxygen ($^1O_2$). From the triplet state of the PS also charge transfer reactions are possible resulting in radical formation.

3.1 Types I, II, and III of Photosensitization

All three pathways of photosensitization start from the first excited triplet state of the PS. In Figure 3 these processes are shown schematically.

3.1.1 Type I

This mechanism occurs in the presence of molecular oxygen. Following electron transfer from the PS to

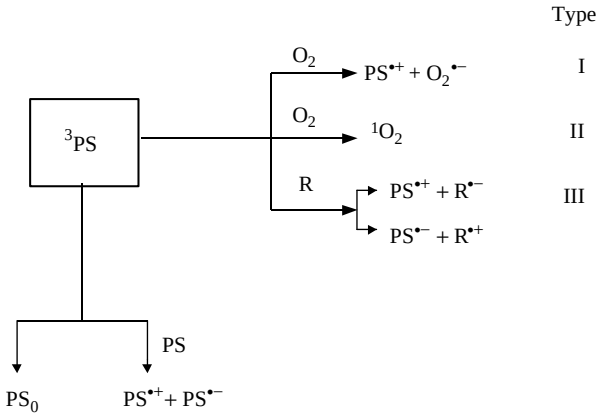

Figure 3 Types of photosensitization starting from the first excited triplet state of the PS (^{3}PS) with R – substrate molecules – and O_2 – molecular oxygen.

other molecules radicals are generated which interact in further dark reactions with biomolecules. One special case is the formation of the superoxide anion radical ($O_2^{-\bullet}$). This reaction may occur in polar solvents but it is rare since in most cases of PS–oxygen interaction singlet oxygen is generated. Since the first step is a PS – substrate interaction high substrate concentrations are required (otherwise PS–oxygen interactions (Type II) dominate). Under such conditions electron transfer from PS to substrate molecules (or inverse reaction, depending on the redox potential) followed by the reaction of the formed anion radical with molecular oxygen giving $O_2^{-\bullet}$ may occur. The initial interaction of PS and substrate molecules may also include hydrogen abstraction. The radical species formed can react with molecular oxygen giving peroxides.

3.1.2 Type II

The most important step is the EnT from the PS in its T_1-state to molecular oxygen. As a result singlet molecular oxygen ($^1O_2$) is produced. During this step the PS is transformed to the ground state (S_0) and can absorb another quantum of light. That means – the PS plays the role of a ''catalyst''. Using moderate light doses in PDT (which do not cause the photodestruction of the PS) every PS molecule can generate thousands of $^1O_2$ molecules. The photosensitized generated $^1O_2$ can interact directly with substrate molecules like proteins or lipids. Also the production of oxygen radicals in a high amount is possible[10,11] which permit additional phototoxic reactions. In other words: Type II mechanism is a very efficient catalytic chain reaction process. It needs only low amounts of PS to generate a strong phototoxic answer of the system.

The processes of Type I and II photosensitization take place in a system at the same time, which means that they are running in competition. The relative efficiency of both pathways is mainly defined by the physicochemical properties of the system. At normal oxygen concentration and neutral up to physiological pH-value in most cases the Type II mechanism will dominate. Using simple model systems a distinction between the two types of photosensitization can be made on the basis of different photoproducts.

3.1.3 Type III

This type is similar to the Type I in so far as radicals are also produced but in contrast to Type I photosensitization discussed above the reactions occur in the absence of molecular oxygen and/or at very high local concentrations of the reactants. The excited PS molecule directly interacts with the substrate molecule. Via electron transfer from the first excited singlet or triplet state of the PS to substrate molecules radicals are formed.

3.2 Photophysical Properties of Photosensitizers

From the scheme of the molecular processes of photosensitization discussed above (Figure 2) it can be deduced that potent PSs must possess a high triplet quantum yield and long triplet lifetime to obtain high singlet oxygen quantum yields. Such behavior is known from macrocyclic systems where the ISC is not longer forbidden because of the spin-orbit-coupling.[12] The activation energy of $^1O_2$ is 0.98 eV (22.4 kcal mol^{-1}), which means that potent PS should have triplet energies higher but near 0.9 eV. In other words compounds with high ISC quantum yield and strong absorption in the red or far red region (650–850 nm) could be effective PSs for both the Type I and Type II mechanisms. In general such compounds would be also effective PSs for Type III processes.

Additionally the above mentioned strong absorption in the red or far red region allows a high penetration depth of the incident light since biomolecules (excluding melanin) and water do not absorb radiation in this spectral range.

3.3 Singlet Molecular Oxygen

Singlet molecular oxygen plays a key role in PDT. It is the only activated oxygen species which can be generated by electronic excitation via EnT. The special properties and the important role of molecular oxygen results from its extraordinary electronic structure. This molecule has two unpaired electrons in the ground state. For this reason the ground state of molecular oxygen is a triplet state (T_0), in contrast to the most common molecules which have a ''closed shell'' conformation in the ground state (S_0). The first excited electronic state of molecular oxygen is

a singlet state ($S_1 : {}^1\Delta_g$ so-called: $^1O_2$). That means $^1O_2$ is generated as a result of a spin-forbidden ISC process. The direct excitation of this state with laser irradiation is possible but has very low quantum yields.[13]

For the generation of $^1O_2$ only a very small amount of energy (0.98 eV) is needed. The lifetime of $^1O_2$ strongly depends on the environment. In solution it can adopt values between 700 µs in CCl_4 and 3 µs in aqueous solution. An overview on singlet oxygen decay times in a very big number of solvents is given in Wilkinson et al.[14] To excite the second excited singlet state ($^1\Sigma_g^+$) which has a very short lifetime (about 10^{-10} s in solution), already 1.6 eV is needed. The first excited triplet state ($^3\Sigma_g^+$) is located in the ultraviolet (UV) region. Because of these energetic conditions only the first excited singlet state $^1\Delta_g$ of molecular oxygen ($^1O_2$) is of interest for biomedical applications.

For estimating the potential photosensitizing efficiency of a given compound the determination of its singlet oxygen quantum yield (Φ_Δ) is a useful tool. This value can be obtained from a kinetic analysis of the relevant processes shown in Figure 2. As a result we get Equation (1)

$$\Phi_\Delta = \{k_{ISC}(k_{rad} + k_{IC} + k_{ISC})^{-1}\} \times \{(k_{EnT}[O_2])k_{ISC} + k_{rad} + k_{EnT}[O_2])^{-1}\}S_\Delta \quad (1)$$

with the rate constants (k) of the ISC, internal conversion (IC), radiative deactivation (rad), and EnT. $[O_2]$ is the concentration of molecular oxygen. The factor S_Δ takes into account that not all PS in the triplet state are deactivated via EnT to molecular oxygen. For most molecules the value of S_Δ is much smaller than one but for the tetrapyrroles it approaches nearly one. Simplifying Equation (1) we get Equation (2):

$$\Phi_\Delta = \Phi_{ISC}\tau_T k_{EnT}[O_2]S_\Delta \quad (2)$$

From this equation it can be deduced that Φ_Δ depends not only on the triplet quantum yield (Φ_{ISC}) but also on the triplet lifetime (τ_T) of the PS under the given experimental conditions which are strongly dependent on the oxygen concentration and other environmental factors. Quantum yield and lifetime of the first excited triplet state of molecules are also influenced by the inner heavy atom effect. For this reason in general only a few atoms like zinc and magnesium should be used as central atoms in tetrapyrrolic PS.[15]

The transition of $^1O_2$ to the ground state is forbidden according to different selection rules. This results in a very low rate constant of the radiative deactivation and the observed lifetime mainly depends on the nonradiative deactivation processes. Depending on the solvent the quantum yields range between 10^{-5} and 10^{-6}.

Table 1 Radiative deactivation processes of singlet molecular oxygen ($^1\Delta_g$ and $^1\Sigma_g^+$)

Transition	Emission wavelength	Detection method
$^1\Delta_g \longrightarrow {}^3\Sigma_g^-$	1270 nm	Phosphorescence
$^1\Sigma_g^+ \longrightarrow {}^3\Sigma_g^-$	762 nm	Chemiluminescence
$2[^1\Delta_g] \longrightarrow {}^3\Sigma_g^- + {}^3\Sigma_g^-$	633.5 nm	Chemiluminescence
$[^1\Delta_g][^1\Sigma_g^+] \longrightarrow {}^3\Sigma_g^- + {}^3\Sigma_g^-$	478 nm	Chemiluminescence
$[^1\Sigma_g^+][^1\Sigma_g^+] \longrightarrow {}^3\Sigma_g^- + {}^3\Sigma_g$		

The production of $^1O_2$ can be measured using indirect (chemical) or direct (chemiluminescence, phosphorescence) methods. Chemical methods are based on the use of monitoring molecules like 1,3-diphenyl-isobenzofuran. From the obtained amount of photoproducts Φ_Δ can be calculated. These methods are useful only in homogeneous environments.[16] Mainly dimol reactions are used for the detection of $^1O_2$ via chemiluminescence[17] (Table 1). Nevertheless the small rate constant of the radiative deactivation is high enough for direct detection of the $^1O_2$-luminescence at 1270 nm. With the Ge-detectors available today steady-state as well as time-resolved detection in homogeneous and microheterogeneous environments is possible.[18,19] The direct detection of $^1O_2$-luminescence in vivo is impossible with the present experimental technique. The reason is the very short lifetime of $^1O_2$ in organisms caused by very efficient competitive quenching processes. Up to now only computations of the lifetime of $^1O_2$ in cells exist. Moan et al.[20] proposed a lifetime of 10–40 ns with a corresponding diffusion distance of 10–20 nm, and Meffert et al.[21] proposed diffusion distances of about 60 nm. For this reason indirect methods for the evidence of $^1O_2$ generation in cells like the measurement of consumed oxygen and the efficiency of phototoxic reactions[22] or the determination of the triplet lifetime of the PS[23] at different oxygen concentrations in the tissue are used.

4 PRINCIPLE OF PHOTODYNAMIC THERAPY

PDT is based on the use of the principle of photosensitization to get as much as possible phototoxicity. The advantages of this therapy are the use of low energy radiation which is not absorbed by biomolecules and water, and the catalytic action of the PS molecules. Strictly speaking PDT is based on the Type II pathway of photosensitization – the photodynamic effect. The principle of PDT is schematically shown in Figure 1. Nevertheless as pointed out before (see section 3) in general three mechanisms

of photosensitization can be distinguished. The Type I involves the generation of radicals in the presence of oxygen where the PS does not play a catalytic role. The Type III pathway describes the direct chemical reaction between photoexcited PS and the substrate molecules under deaerated conditions, the quantum yield is low and relatively high PS concentrations like in chemotherapy are needed. In other words no photodynamic effect occurs in these two cases of photosensitization. For this reason therapeutic treatments based on Type I and Type III mechanisms will be mentioned only briefly.

4.1 Photochemotherapy (Type I/III of Photosensitization)

In vivo it is not easy to discriminate between the Type I and Type II mechanism because they are competitive processes. In the ''classical'' photochemotherapy the photodynamic effect is not used and the phototoxic effects are therefore mainly due to Type I processes. Typical Type I agents are psoralens like the 8-methoxypsoralen used in the PUVA-therapy of psoriasis. Nevertheless they are also known to generate $^1O_2$.[24] Photosensitizing processes via Type III mechanism with remarkable quantum yields are known only from experiments carried out in solvents.[9] Classical photoinduced electron transfer processes are not regarded here.

4.2 Photodynamic Treatment (Type II Photosensitization)

As was mentioned before the limiting factors of the efficiency for this treatment are the photophysical properties of the PS, the oxygen concentration during the treatment and the localization of the PS molecules in the tissue. In this section we will focus on the mutual dependence of the first two factors.

PS used in PDT absorb light in the red or far red region. For this reason it is important to know the optical properties (absorption and scattering) of normal and tumor tissue in this region. Wilson et al.[25] gave an overview on direct and indirect methods for the measurement of optical properties of tissue. Profio et al.[26] investigated the energy fluence rates of light in dependence of absorption and anisotropic scattering for different geometries. Svaasand et al.[27] carried out detailed studies on the penetration depth of light with different wavelengths for normal tissue and human intracranial tumors. All these authors found penetration depths of some millimeters in the red region which was about 2–3 times larger than in the green/blue region. Using PSs of the second or third generation which absorb at wavelengths of about 680 nm to 780 nm the penetration depth may rise in dependence on the tissue up to centimeters.

In recent years the possibility of simultaneous two-photon activation of PS is under discussion. The advantage of this method is the use of near-infrared radiation from YAG-[28,29] or Ti : Sa-lasers[30] which once more allow a deeper penetration of the light. A review on the present knowledge and broad discussion of the advantages and problems of simultaneous two-photon stimulation is given in Fisher et al.[30]

Together with the photoactivation of the PS the question of the role of tumor oxygenation during photodynamic treatment becomes important. Foster et al.[31] investigated the oxygen consumption and diffusion effects in PDT and developed a mathematical model of the Type II mechanism. They found that during PDT carried out at incident light intensities of 50 mW cm^{-2} molecular oxygen can be consumed at rates of 6–9 μM s^{-1}. Thus, during photoirradiation the actual oxygen concentration may strongly decrease. It can become low enough to minimize or preclude $^1O_2$-mediated processes and as a result photodestruction of the PS may occur. Their analysis further suggests that the use of fractionated light (30 s exposure followed by 30 s dark periods) can overcome partially the oxygen depletion. In this regime a significantly longer delay in tumor growth was produced when compared to the regime of continuous illumination of the same total light dose.

The common photodynamic treatment is a continuous one-photon activation process. The PS doses differ depending on the absorption coefficient of the PS at the excitation wavelength. For PS of the first generation it is about 1–5 mg kg^{-1} body weight. The light doses used for these PSs at 610–620 nm are about 50–100 mW cm^{-2}.

Summarizing these facts we can conclude that the common treatment is not carried out under optimized conditions. Efficient photodynamic treatment should be done using fractionated simultaneous two-photon activation of PS.

4.3 Photoactivity from Higher Excited States of Photosensitizers

As was shown in Figure 2 the PS can be excited to higher electronic singlet states (S_n) using UV-light or sequential two-photon absorption (SPA). Proposing relatively long-lived S_n-states (some picoseconds) and low ISC-quantum yields it should be possible to generate radicals different from those generated via the Type I pathway.[32,33] First experiments with SPA on HPD[34] and protoporphyrin IX (PP)-dimethylester[35] showed that in this case not only Type I and II but also oxygen-independent reactions occur. Nevertheless until now SPA has not been used in clinical application.

5 PHOTOSENSITIZERS

Different classes of substances have been investigated for their photodynamic activity. It is well known that such substances like erythrosine or methylene blue which are used as dyes in cytology and dental medicine have high singlet oxygen quantum yields and may cause photodynamic effects.[36,37] Also some common toxins (adriamycin, daunomycin) used in the chemotherapy of cancer are able to generate $^1O_2$ under irradiation.[38]

A special problem is the therapy of brain tumors because of the blood–brain barrier. Beside tetrapyrroles also chalcogenapyrylium dyes are under investigation.[39,40] Also dyes with absorption in the green region of the spectrum like Merocyanine 540 are evaluated for their photosensitizing properties. Such dyes become important for the treatment of single cells, e.g. leukemic cells.[41]

Nevertheless the most important chemical class of potent PSs are tetrapyrroles. The large conjugated π-electron system guarantees high triplet quantum yields. Changes of the central atom, or of the substitution at different binding sites as well as the possibility to extend the macrocyclic system allow a wide range of modifications to fine-tune the electronic properties of these substances. Thus specific PS for the treatment of different diseases can be designed. An actual overview on the chemistry of potent porphyrin-based PS is given by Sternberg et al.[42]

For this reason the next section will focus on PS from the class of tetrapyrroles.

5.1 Exogenous: First, Second, and Third Generation of Photosensitizers

Because of the broad variety of tetrapyrroles a large number of possible PS were prepared during the last twenty years. An overview on the chemistry, and some aspects of photodynamic activity is given by Bonnett.[43] For this reason only selected important PSs will be discussed in this paper. The structural formula of some selected PSs belonging to the different groups of tetrapyrroles are shown in Figure 4. Some important photophysical parameters of some representative compounds are summarized in Table 2.

5.1.1 First Generation Photosensitizers

Porphyrins (Figure 4) were the initial PS in PDT. The history started with hematoporphyrin. Later HPD, a mixture of different porphyrins[57] was under detailed investigation. It emerged that the photoactive compounds are mainly dimers or higher aggregates. Such a mixture with a high content of the photoactive compounds is commercially available under the name Photofrin® I and II (see sections 1 and 8). Photofrin® was the first PS in clinical practice. Other porphyrins were also studied and it was found that coproporphyrin – Type III[58] – and protoporphyrin[59] have good photosensitizing properties. Berenbaum et al.[60] investigated the photodynamic activity of tetraphenylporphyrins (TPP) in vitro and Sacchini et al.[61] carried out a preclinical study using sulfonated TPP for the treatment of neoplastic tissue in topical application. The disadvantages of all these porphyrins are the weak absorption in the red region of the spectrum ($\varepsilon < 5000\,\mathrm{mol^{-1}\,cm^{-1}}$ at 630 nm) and serious side effects like skin photosensitization for 4 to 6 weeks.[62]

5.1.2 Second Generation Photosensitizers

It was the intention to overcome these disadvantages that led to the development of second generation PS. These PS were to combine strong absorption in the red or near infrared region together with high singlet oxygen quantum yields and weak side effects. The first group of tetrapyrroles investigated under these aspects were the benzoporphyrins. These compounds have a strong absorption around 692 nm[63] and their phototoxicity is 10–70 times higher[64] than that of HPD.

A very promising group of PS are systems that are based on reduced porphyrins, i.e. tetrapyrroles possessing a chlorin or bacteriochlorin skeleton. The lack of one or two double bonds in the tetrapyrrole macrocycle leads to a pronounced shift in the absorption spectra, generating systems with increased absorption in the red region, which is obviously very attractive for PDT.

Phorbides are a group of tetrapyrroles derived from the naturally occurring chlorophyll and have strong absorptions around 670 nm. A very efficient representative of this group is pheophorbide A. Its photodynamic activity was shown in vitro[65] and in vivo.[66,67] A broad variety of pheorbides and pyropheophorbides was studied for their photodynamic activity during the last few years. A review of these activities is given in Röder.[68]

Under similar conditions also the photodynamic activity of derivatives of bacteriochlorophyll and bacteriopheophorbide was investigated.[69,70] In comparison to chlorophyll derivatives these compounds have a strong red-shifted absorption at around 740 nm. This makes them very promising compounds for the treatment of melanoma tumors.[71,72]

Two other compounds derived from natural chlorin systems that interest has focused on are chlorin e_6 and mono-L-aspartylchlorin (Npe6). For both substances very good photoactivity in vitro and in vivo[73,74] was shown. Nowadays the in vivo activity of the artificial chlorin *m*-THPC is studied intensely.[75] First results of clinical trials

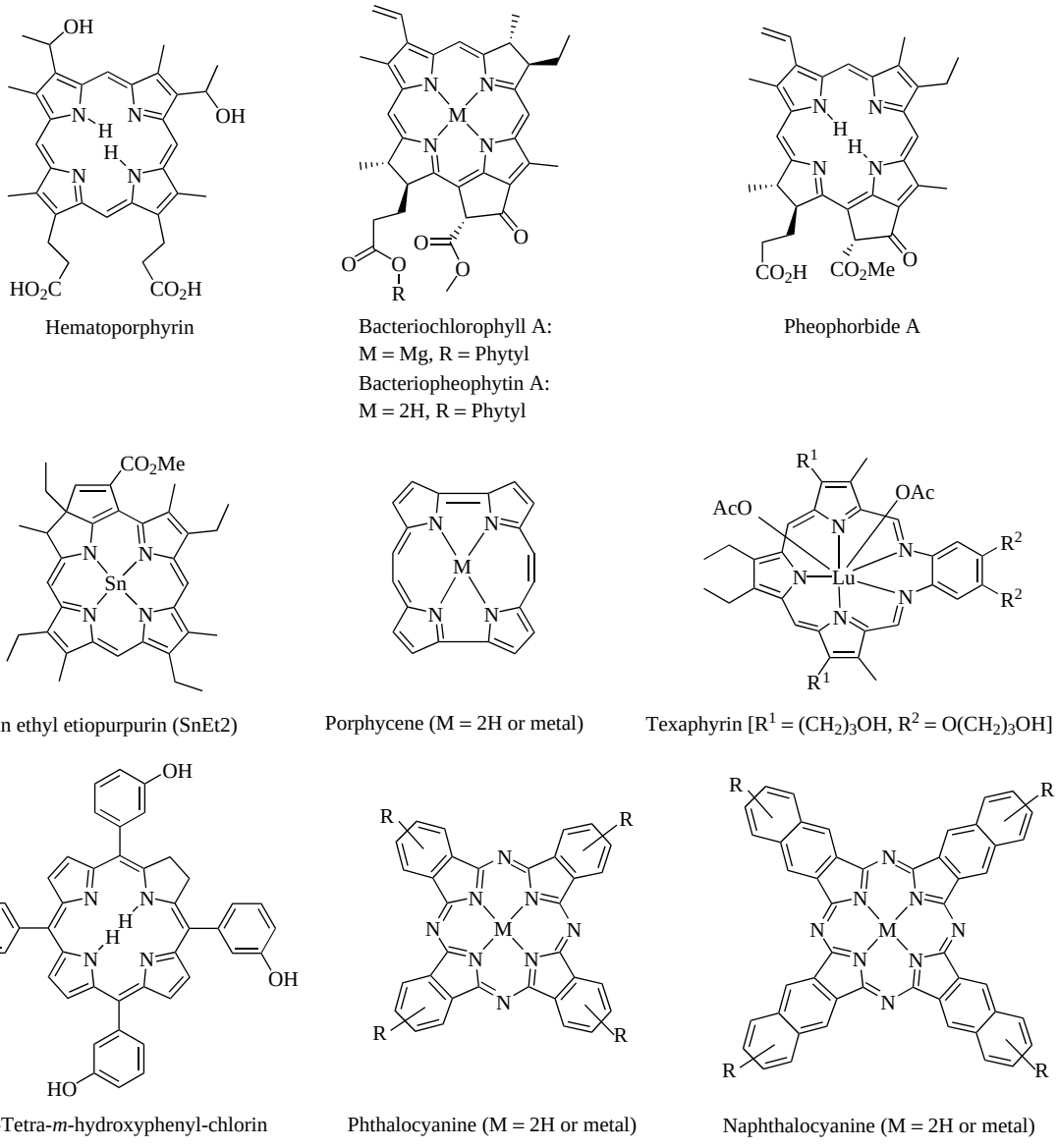

Figure 4 Structural formulae of some representative PSs from the different groups of tetrapyrroles.

showed that this compound is the most effective PS at this time (see section 8).

A lot of different compounds of the class of purpurins were synthesized and characterized. A few of them (e.g. SnNT2H2, SnET2, SnET2H2) show very good PDT response in vivo. The data of these compounds are summarized by Garbo.[76]

Another possibility to increase the strength of absorption is the extension of the macrocyclic ring system. As a result porphycenes,[77] texaphyrins[78] and sapphyrins[79] were synthesized and investigated. From this group the Lutetium texaphyrin, a water soluble PS, is the most promising compound. This molecule contains a 22π-electron macrocycle that causes a strong absorption at 740 nm. The complexation with a metal causes an increased ISC quantum yield together with very weak fluorescence quantum yields. First results of in vivo investigations are given by Kostenich et al.[80]

Because of the very strong absorption in the far red also the use of tetra-azaporphyrins – PCs and NPCs – was investigated. The photodynamic activity of different PCs and NPCs was demonstrated with a broad variety of compounds. Detailed studies were carried out using different Zn-PC,[81] or ClAl-PC.[82,83] Recently results are reported which suggest that using a silicon PC PS may avoid side effects like skin photosensitization.[84] The use of NPC is more problematic because of the absorption at wavelengths around 800 nm. This may cause a back

Table 2 Photophysical properties of some selected potent PSs (ε_{abs}: absorption coefficient at the absorption wavelength λ_{abs}; Φ_T: triplet quantum yield; Φ_Δ: singlet molecular quantum yield)

PS	ε_{abs} [$mol^{-1}\,cm^{-1}$]	λ_{abs} [nm]	Φ_T [%]	Φ_Δ [%]	Ref.
Porphyrins (P)					
Hemato-P-IX in methanol		610	0.90	0.60	44
in H_2O			0.56	<0.10	
Tetra-Phenyl-P		652	0.67	0.63	45
m-Tetra-hydroxy-phenyl-P		648	0.69	0.57	45
Phorbides					
Pheophorbide A	49.500	667	0.75	0.52	46
Bacteriochlorophylls					
Bacteriochlorophyll A	70.000	700	0.22	0.2–0.6	47
Bacteripheophytin A		600	0.64	0.47	47
Chlorins					
Chlorin e_6 (in benzene)		654		0.66	48
Mono-L-aspartyl chlorin e_6 (Npe6) in PBS		653		0.77	48
CHL I	42.300	660	0.57	0.58	49
CHL II	46.500	646	0.69	0.58	49
Purpurins					
Octaethylpurpurin (NT2)	41.400	700	0.81	0.65	50
Zn(II)etiopurpurin (ZnET2)	27.600	660	0.84	0.54	50
Porphycenes (PO)					
Porphycen	42.300	627	0.42	0.30	51
Tetra-*n*-propyl-PO	50.400	637		0.35	51
Texaphyrins (TP)					
Cd-TP (dicyano)	40.500	695	0.52	0.45	52
Cd-TP	45.900	759	0.88	0.78	52
PC					
t-Butyl-PC		702		0.16	53
t-Butyl-ClAl-PC		693		0.31	53
Tetra-sulfo-PC		660	0.22	0.52	54
Tetra-sulfo-ClAl-PC		680		0.20	54
Zn-PC	168.000	673	0.65	0.4–0.5	55
NPC					
Si-NPC		774	0.20	0.19	56
t-Butyl-NPC		784		0.17	53
t-Butyl-ClAl-NPC		785		0.41	53

PC, phthalocyanine; NPC, naphthalocyanine. (Partly reproduced with permission from B. Röder, 'Photosensibilisatoren für die Photodynamische Therapie', in *Angewandte Lasermedizin*, eds. H.-P. Berlien, H. Müller, Ecomed Verlag, München-Landsberg, 6. Erg. Lfg. 2/93, III-3.15, 1–12.)

transfer of energy from $^1O_2$ to the PS.[85] Nevertheless good photodynamic activity in vivo was shown for some NPC.[83,86] A review on NPC photosensitization is given by Kliesch et al.[87]

To date much fundamental research on the photosensitizing properties and photodynamic activity of tetrapyrroles has been carried out using a large variety of compounds, but only a few compounds have reached the stage of clinical trials.[74,88,89] These compounds together with their advantages and disadvantages, the selling company, and treated diseases are summarized in Table 3 (section 8).

5.1.3 Third Generation Photosensitizers

After twenty years of clinical PDT a variety of very efficient PS exist with excellent photophysical properties. The unsolved problem today is the ''right'' location of PS in the cell or at the cell surface. To solve this problem different strategies are employed. One possibility is the use of drugs specifically designed for

Table 3 PS of the first and second generation in clinical trials. (Data taken from several sources[127–136])

PS	Company	Advantages	Disadvantages	Appl.	Phase of clinical trial	Indications
Photofrin®	QLT, Vancouver, Canada	Easy to synthesize Easy to formulate Approved worldwide	Porphyrin mixture Weak absorption at 630 nm Skin phototoxicity (up to 4 weeks)	i.v.	Gouvern. approval Active Invest. Studies	Bladder Lung Esophagus Brain Head and Neck Lung Breast Bladder
ALA	DUSA, Pharmaceuticals, Mt. Kisco, NY, USA	Easy to formulate No skin phototoxicity	Pain associated during PDT Limited depth of tumor necrosis (about 1 mm)	Topical Oral	Phase 1	Psoriasis Actinic keratosis Basal cell carcinoma Bladder
	MEDAC, Hamburg, FRG			Topical	Phase 2	Basal cell carcinoma lung
BPD-MA	QLT, Vancouver, Canada	Absorption at long wavelengths (690 nm) No skin phototoxicity Clears rapidly (3 h incubation time)	Difficult to synthesize Clears rapidly (3 h incubation time)	i.v. Liposomes	Phase 1/2	Skin carcinoma Psoriasis
Npe6	Nippon Petrochemical Osaka, Japan	Easy to synthesize Easy to formulate No skin phototoxicity	No data	i.v. (Emulsion)	Phase 1	Skin tumors Lung
m-THPC	Scotia Pharmaceuticals, Guilford, UK	Very effective PS Easy to synthesize Absorption at long wavelengths (690 nm)	Skin phototoxicity 7–9 days incubation time before irradiation	i.v. Topical	Phase 1/2 Phase 1/2	ENT tumors Basal cell carcinoma
SnET2	PDT Inc. St Barbara, USA	Strong absorption in the red (660 nm)	No data	i.v. Liposomes	Phase 1/2	Metastatic breast carcinoma Basal cell carcinoma Kaposi's sarcoma
Zn-PC	QLT/Ciba Geigy Basel, Switzerland	Very strong absorption in the red (670 nm)	No data	i.v. Liposomes	Phase 1 Preclinical	Lung Esophagus tumors Psoriasis
Lu-Tex	PDT Inc. St Barbara, USA	No skin phototoxicity Strong absorption at 732 nm Effective PS	Pain associated in the skin during light treatment Multi-step synthesis	i.v.	Phase 1 Phase 2/3	Metastatic breast carcinoma Malignant Melanoma Kaposi's sarcoma Basal cell carcinoma Certain skin lesions
Sulfonated ClAl-PC	State Research Center for Lasertechn. Moscow, Russia	No data	No data	i.v. Topical	Phase 2	Breast Lung Skin
Pheo	Commercially not available (HU Berlin, FRG)	Easy to formulate Strong absorption at 677 nm Clears rapidly	Weak skin phototoxicity	Topical (Emulsion)	Preclinical	Psoriasis

(*continued overleaf*)

Table 3 (*continued*)

PS	Company	Advantages	Disadvantages	Appl.	Phase of clinical trial	Indications
Porphycene	Cytopharm-Glaxo NC, USA	No data	No data	Topical	Preclinical	Psoriasis
HPPH	Roswell Park Cancer Institute, USA	Easy to synthesize Easy to formulate Strong absorption at 660 nm No skin phototoxicity	No data	i.v.	Preclinical	Oesophageal cancer

ALA, 5-aminolevulinic acid; BPD-MA, benzoporphyrin derivative–monoacid ring A; Lu-Tex, lutetium–texaphyrin.

treatment of one disease. In this case the PS itself has a binding site with strong affinity to a special locus of the target cell type or is coupled with some molecule preferably accumulated by this cell type. The following two examples should explain this idea. To enhance the photodynamic activity of chlorin e_6 by its delivery to specific, sensitive intracellular compartments of target cells Akhlynina et al.[90] prepared a bovine serum albumin–insulin–chlorin e_6 complex that was bound specifically to the insulin receptors of a human hepatoma cell line. They found an enhancement of photodynamic activity of about two orders of magnitude. Water soluble serine conjugates of chlorophyll and bacteriochlorophyll were successfully used in the treatment of mice bearing melanoma tumors.[72]

Another possibility is the use of modular carrier systems (see section 6.3). Both strategies have the intention to get a selective accumulation and to increase at the same time the concentration of PS in the target tissue. Such PS having biological active sites for selective accumulation are developing and named third generation PS.

5.2 Endogenous: 5-Aminolevulinic Acid Therapy

The principle of ALA therapy differs from the other photodynamic treatments because a precursor of the active PS is applied. ALA is the natural precursor of endogenous porphyrins. The exogenously applicated ALA stimulates the intracellular production of PP and the endogenously produced PS is accumulated in malignant cells.[91,92] In 1990 the ALA therapy was proposed by Kennedy et al.[93] as a new phototherapeutic modality. It was found that ALA penetrates through human skin and topical application of ALA has been used for the treatment of different skin diseases in human: actinic keratoses, basal cell carcinoma,[94,95] and psoriasis[96] (see also Table 3, section 8). The disadvantages of this therapy are side effects like skin photosensitization and pain. Using animal models also the treatment of brain tumors,[97] and of viral infections[98] are reported. It has also been shown that intravenously injected ALA preferentially accumulates in melanoma.[99]

6 DRUG DELIVERY SYSTEMS

One of the crucial problems in any therapy is the selective action of the drugs used. All the discussed PSs of first and second generation do not exhibit active selective accumulation. Concentration gradients of PS accumulation between normal and tumor tissue are mainly caused by metabolic differences. As a result time-shifted accumulation peaks of PS in different tissues are found. These differences may become high (up to a relationship of 1 : 50), so that they can be used in photodynamic treatment. Anyhow such relations are not good enough for experienced clinical use. Moreover the nonspecific accumulation of PS is the reason for common PDT side effects. Another problem concerns the positioning of the PS molecules at the therapeutically important sites of the cells. To increase the selectivity and to avoid side effects different PS delivery systems were created based on different PS and tissue properties: immunotargeting, high molecular weight PS complexes, and lipophilic properties of PS. Nevertheless using the delivery systems described in the following sections (6.1 and 6.2) the problems could be solved only partly.

6.1 Carrier Molecules

The use of carrier molecules is directed on a selective and highly concentrated accumulation of PS molecules in the cells or at the cell surface. Different strategies have been developed which use the properties of malignant cells or of a special cell type.

Immunotargeting Covalent bonding of PS molecules to antibodies.[100] It has been demonstrated that the complexation of chlorin[100] with a monoclonal antibody does not influence the $^1O_2$ quantum yield. Later it was shown that PS–antibody complexes lose their immune activity if more than eight PS molecules are bound to one antibody.[101] Consequently only a defined quantity of PS given by the binding sites of antibodies on the cell surface can be delivered to the cells.

High Molecular Weight Photosensitizer Complexes
PSs with polyvinyl chains are used by different authors to deliver chlorins[101,102] or NPC.[87,103] Krinick et al.[102,104] proposed a delivery system containing water soluble polymer coupled via a degradable spacer to the PS (*meso*-chlorin e_6).

6.2 Carrier Systems

The common carrier systems – micelles and liposomes – use the lipophilic properties of PS. A broad number of different micelle forming detergents were investigated but mainly Tween 80 and Cremophor® EL were used for PS delivery.[105,106] In comparative studies of these carrier systems different pharmacokinetics for a NPC were found.[107] Woodburn et al.[108] reported an influence of the delivery systems on the physical properties of plasma lipoproteins, resulting in the formation of a lipoprotein degradation product.

It is known that serum lipoproteins play an important role in the in vivo transport of PSs. Mosley et al.[109] proposed the use of low-density lipoproteins (LDLs) as carrier systems for selective delivery of PS to tumor cells. Kessel[110] showed that LDL–PS complexes where the PS is associated with the lipid moiety of LDL are largely accumulated via a receptor-mediated endocytotic process Further investigations showed the existence of a correlation between the extent of sensitizer – LDL association and the efficiency of selective PS accumulation in tumor tissue. The extent of LDL association of PS strongly depends on the hydrophobicity of the PS. A detailed discussion on the role of LDL in the delivery of PS is given by Jori et al.[111] The delivery of Zn-PC to LDL using liposomes was investigated by Reddi et al.[112] They found that the extent of ZnPC accumulation depends on the content of cholesterol in the liposomes. The use of liposomes as PS delivery system is well accepted and used for sensitizers of the first and second generation: HPD,[113] BPD-MA,[114] pheophorbide,[115] PC,[112] NPC.[86]

All of these systems have the disadvantage that relatively large carrier systems have to be used to deliver one or only a small number of PS molecules to the target tissue. For this reason nowadays more efficient delivery systems which are able to carry a large number of PS molecules are under investigation. Moreover the selective accumulation of PS using the described delivery systems was not as high as expected.

6.3 Modular Carrier Systems

Nowadays the use of modular carrier systems and polyphasic tumor targeting (PTT) are proposed to get high concentrated and selective – biologically defined – accumulation of PS. The intention of this concept is the attachment of PS molecules to one addressee – and one multiplying unit – the so-called third generation PS.

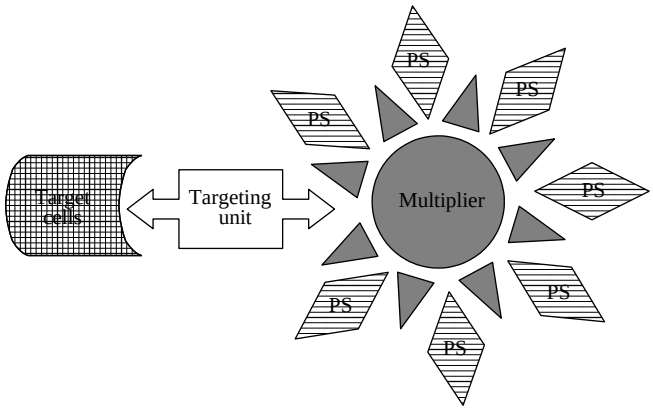

Figure 5 Schematic diagram of a modular carrier system loaded with PS molecules.

The sense of these systems is the spatial separation of PS from the localizing part of the carrier system and the attachment of a large number of PS molecules to one carrier unit (see Figure 5).

The idea to use modular carrier systems in the so-called PTT was suggested first by Moser.[116] He proposed a modular carrier system containing monoclonal antibodies as the addressee and an avidin–biotin complex (AB-C) as the multiplying unit. Every AB-C has four binding sites, three of them are saturated with PS molecules the fourth with the next AB-C. First in vitro experiments showed that the concept in general works very well.[89] But problems could arise using this modular carrier system containing several AB-C in vivo (since avidin is a large protein with a molecular weight of 70 000).

To avoid the immune problems arising from application of high molecular weight proteins the possibility of using other classes of molecules has been proved. In this context dendrimers were under investigation. Third generation starburst dendrimers were covalently linked on an average of twelve PS molecules[117] and their photophysical and photobiological properties were studied. First in vitro results suggest that dendrimers could become very efficient multiplying units. An additional advantage of this system consists in the light-induced abstraction of PS molecules from the dendrimer. That is, a large number of PS can be transferred directly to target tissue by one carrier unit which has no immune activity.

7 COMMON LIGHT SOURCES

The light doses used in clinical therapy depend on the photophysical properties of the PS and the tissue properties. Typically in the treatment with Photofrin® red

light with fluences between 50 and 100 J cm^{-2} are used. (At higher fluences hyperthermic effects may occur.) For the treatment of a tumor area of 10 cm^2 a power of approximately 0.3–0.6 W at the surface is needed.

According to the stronger absorption of second generation PS the fluences decrease to about 10 J cm^{-2} in the case of *m*-THPC. Also the wavelength of the used light is shifted to 670–740 nm, that means to lower energies.

In general dosimetry of light irradiated tissue is a rather complex problem. Various different applicators were developed to irradiate tissue surfaces of different geometry homogeneously (e.g. scattering spheres or fiber tips). However one of the crucial problems in PDT is the action radius of the photodynamic effect. This value strongly depends on the penetration depth of laser light, which may vary for different tissues. A practical approach to measure light fluence rates in vivo is given by Steiner et al.[118]

7.1 Lasers and Light-emitting Diodes

The common light sources in PDT are dye lasers pumped by conventional lasers, like Ar-, Cu-, or Nd : YAG (frequency doubled) lasers. Dye lasers typically deliver useful light in a spectral range of about 60 nm. This is a serious limiting factor in clinical practice where different PS for the treatment of different diseases should be used.

Nowadays the development of more cost-effective sources like light-emitting diodes (LED) could give the possibility to use optimum light sources for different treatments. The development of a novel LED type based on Al : Ga : arsenide allows the peak wavelength to be varied between 630 nm and 940 nm depending on the relative content of aluminum and gallium. Schmidt et al.[119] constructed a LED probe for in vivo studies containing 144 of such LED chips. The bandwidth of emitted light was 25 nm with the emission peak at 677 nm. The LED probe was placed in a conventional laser balloon adapter which was adjusted to deliver 363 J cm^{-2} for every treatment.

7.2 Lamps

Also the successful use of lamps, especially in the treatment of skin diseases is reported. Recent advantages in lamp miniaturization and of optical filter coatings led to the development of some special light sources for PDT. De Jode et al.[120] used a 5 W metal halide lamp delivering a 40 nm band in the red and 0.45 W at the end of a light guide. Whitehurst et al.[121] developed an equipment with a short arc lamp delivering a 30 nm band between 300 nm and 1100 nm with an output power of 1 W and 0.5 W via a light guide. First in vitro experiments suggest that the efficiency of this light source is similar to a cw Ar-pumped dye laser.

7.3 Pulsed Lasers

Beside cw-light sources an increasing interest in the use of short-pulsed lasers exists (see section 4.1). These lasers could be used in simultaneous two-step absorption of light. Because of the longer wavelengths of the used light deeper tissue penetration is suggested.[28–30,122] Solid-state lasers (Ti : Sa; Nd : YAG; Nd : YLF) with second-, third- or fourth-harmonic generating and optical parametric oscillators could become important light sources in PDT because of their variable emission between the deep UV and the near infrared region. These laser systems are still developing to higher power. Otherwise the development of strong absorbing PS results in smaller PDT light doses. For these reasons the importance of these laser systems in PDT will probably increase in the future.

8 CLINICAL APPLICATIONS

The modern history of clinical application of the photodynamic effect began in 1966 when Lipson et al.[123] for the first time treated one patient with HPD. The first systematic PDT study in human was reported by Dougherty et al.[124] in 1978. Today Photofrin® is approved worldwide and used in the treatment of different tumors and diseases (see Table 3). Overviews on clinical experiences with Photofrin® in recent years are given by Evensen,[125] Jocham[126] and Dougherty et al.[88] As was pointed out (see section 5) Photofrin® is far from an ideal PS. For this reason second generation PSs were developed and some of them are in clinical trials (Table 3).

8.1 Photodynamic Therapy of Skin Diseases and Cancer

PDT studies using new PS are at first mainly focused on the treatment of psoriasis and of cutaneous and subcutaneous tumors because of the possible topical administration of PS. Because of this, a great variety of results exists in this field but only some examples illustrating the typical behavior of these PS will be briefly discussed.

Collins et al.[137] reported on the treatment of 22 patients with plaque psoriasis using topical administered ALA and broadband visible radiation (10–40 mW cm^{-2}). They observed a clearance of psoriasis in some patients after a single treatment.

Szeimis et al.[94] treated actinic keratoses of ten patients with 36 lesions topically with ALA and red light (150 W cm^{-2}). They observed a complete remission in 71% of actinic keratoses located on the head.

Taber et al.[74] used the new PS Npe_6 for the treatment of patients with different superficial malignancies, including recurrent adenocarcinoma of the breast, basal cell carcinoma and squamous cell carcinoma (single i.v. injection of 0.5–3.5 mg kg^{-1} Npe_6; incubation time: 4 h; light: 25–100 J cm^{-2} at 664 nm). They observed an increased tumor response at higher PS doses (2.5 mg kg^{-1}) and no side effects, excluding a temporary skin photosensitivity.

Razum et al.[138] reported about a 95–100% response of basal cell carcinoma 12 weeks after the treatment with 1.2 mg kg^{-1} SnET2 (incubation time: 1 day; light: 200 J cm^{-2} at 660 or 664 nm).

BPD-MA[139] and Lu-Tex[140] were also successfully applied for the treatment of various types of skin diseases. Using Lu-Tex which is strongly absorbing light around 730 nm a complete response of subcutaneous melanoma tumors was found.

Clinical trials in different phases for the treatment of other tumors (see Table 3) using second generation PS are still running. Detailed and statistical reliable data are not available yet.

8.2 Photodynamic Therapy of Other Diseases

Beside the treatment of skin diseases and various types of cancer the photodynamic effect could be used also in the treatment of other diseases:

- viral infections (Herpes, Papilloma, HIV)
- Macula degeneration
- Bacterial infections in caries
- Arteriosclerosis
- Rheumatoid arthritis and other clinical indications.

Already preliminary studies with encouraging results have been done:

Antiviral activity: Photodynamic inactivation of viruses has been discussed as a useful tool for blood banking,[141] the successful treatment of Herpes simplex[97,142] as well as the antiviral efficiency of PSs[97,143] and immunoconjugates[144] against HIV have been shown.

Macula degeneration: Preliminary results obtained with BPD-MA[145] have been reported.

Antibacterial activity: Photosensitizing activity of PC on *Escherichia coli*,[146] of protoporphyrin on Candida cells[147] and of a deuteroporphyrin-hemin mixture on gram positive bacteria[148] are reported. Also photoactive antibacterial immunoconjugates containing Sn-chlorin e_6 were synthesized.[149] The photodynamic effect was also used to treat Streptococcus infections in caries.[150]

Arteriosclerosis: Preliminary investigations using animal models are reported.[151,152]

9 COMPARISON AND COMBINATION WITH OTHER METHODS

Compared with other tumor therapies PDT has the advantage of rather selective action. Serious side effects are not known. The observed long time skin phototoxicity of the first generation PS like Photofrin® could be overcome with the optimized second generation PS. The effective radius of PDT is determined by the penetration depth of the excitation light. That means PDT will be used mainly for the treatment of small or at the surface of some organ located diseases. This limiting value gives at the same time the advantages that even inoperable tumors can be treated successfully (e.g. bladder, lung). Compared with chemotherapy PDT is acting more selective at lower drug doses and without the well known complications. As it is shown in the following sections the therapeutic effect can be enhanced treating tumors with a combination of PDT with conventional therapies.

9.1 Chemotherapy

As already discussed molecular oxygen is needed to induce photodynamic effects. During PDT hypoxia in the treated tissue may arise due to the enhanced generation of $^1O_2$ under illumination. One possibility to avoid this situation is the fractionated illumination with light. Another possibility to achieve therapeutic effects under these conditions is the combination of PDT with chemotherapy using hypoxic toxins. Preliminary studies in tumor mice showed an enhanced therapeutic effect when both therapies were combined.[153] Baas et al.[154] reported on the combined treatment of four patients with mammary skin metastases. For this treatment the PS Photofrin® was combined with the hypoxic toxin mitomycin C. They found an enhanced effect even at reduced sensitizer and light doses.

9.2 Radiotherapy

It is well known that X-ray-irradiation may cause an increase of blood flow in some tumors for some days. Consequently reoxygenation of tumor cells occurs[155] and the cell sensitivity to further irradiation increases. The time needed for reoxygenation depends on the X-ray-dose and tumor type. Therefore different authors investigated the effect of combined PDT and radiotherapy. Graschew et al.[156] combined Photofrin® and local X-ray-irradiation for their studies on the mouse sarcoma tumor model. They found a significant enhancement of the therapeutic effect. The best result was obtained when the PDT irradiation followed 24 h after the radiotherapy. At this time tumor reoxygenation was maximized. Surprisingly they also observed some enhancement when PDT

preceded radiotherapy. Such effect was also described by Kostron et al.[157] They suggest a radiosensitizing effect of HPD.

9.3 Hyperthermia

It is also known that the efficiency of radiotherapy can be enhanced by hyperthermia. For this reason the effect of combined PDT and local hyperthermia in animal tumor models was studied by different authors. A comparative study on three different tumor models showed that the therapeutic effect was enhanced for simultaneous application of both therapies. For such treatment of the sarcoma 180 in mice a synergistic effect was observed.[153]

9.4 Cryotherapy

Morton et al.[158] carried out a study on patients with Bowen's disease to compare the efficacy of cryotherapy and PDT. ALA was used as a topically administered PS. They found a significantly higher probability for the clearance of lesions after one PDT treatment compared with cryotherapy. Following the treatment with cryotherapy side effects like ulceration, infection and also recurrent diseases were observed. Following PDT such complications did not occur.

9.5 Immunotherapy

The mechanism of tumor destruction caused by PDT treatment includes a variety of pathways leading to the death of tumor cells.[159] One of these events is the inflammation of the treated tissue associated with massive invasion of activated myeloid cells. Beside the tumor cell destruction an immune development process starts resulting in the generation of tumor-specific immune memory cells. In difference to the local character of PDT treatment this process can have a systemic character inducing an immune reaction of the organism. Thus, the specific cells generated by PDT can be used in further treatments to eradicate lesions (e.g. metastases) of the same cancer. Different regimes of immunotherapy were proofed in animal tumor models. It was found that PDT can be successfully combined with different immunotherapy regimes. A detailed discussion on the prospects of combined PDT-immunotherapy is given by Korbelik.[159]

10 OUTLOOK

The experienced use of PDT in the near future will depend mainly on the development of efficient selective PSs of the third generation which cause no side effects and on the development of low-cost light sources like LED. More and more new areas of PDT application become important. Beside the ''classical'' use in the treatment of tumors and various skin diseases the soft noninvasive PDT is also still under improvement for its use in the therapy of rheumatoid arthritis, arteriosclerosis, age-related macular degeneration processes, and of viral and bacterial infections. A better understanding of the processes leading to the destruction of tumor cells will result in optimized therapeutic regimes. This may include fractionated or SPA irradiation of the treated areas and also the combination with other established tumor therapies.

ABBREVIATIONS AND ACRONYMS

AB-C	Avidin–Biotin Complex
ALA	5-Aminolevulinic Acid
BPD-MA	Benzoporphyrin Derivative–Monoacid Ring A
EnT	Energy Transfer
HPD	Hematoporphyrin Derivative
IC	Internal Conversion
ISC	Intersystem Crossing
LDL	Low-density Lipoprotein
LED	Light-emitting Diodes
Lu-Tex	Lutetium–Texaphyrin
m-THPC	*m*-Tetrahydroxyphenylchlorin
NPC	Naphthalocyanine
PC	Phthalocyanine
PDT	Photodynamic Therapy
PP	Protoporphyrin IX
PS	Photosensitizer
PTT	Polyphasic Tumor Targeting
SPA	Sequential Two-photon Absorption
TPP	Tetraphenylporphyrins
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Single Biomolecule Detection and Characterization

Clinical Chemistry **(Volume 2)**
Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Phospholipids of Plasma Lipoproteins, Red Blood Cells and Atheroma, Analysis

of • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Fluorescence Imaging Microscopy • Fluorescence Lifetime Measurements, Applications of • Near-infrared Absorption/Luminescence Measurements • Phosphorescence Measurements, Applications of

General Articles **(Volume 15)**
Ultrafast Laser Technology and Spectroscopy

REFERENCES

1. F. Meyer-Betz, 'Untersuchungen über die biologische (photodynamische) Wirkung des Hämatoporphyrins und anderer Derivate des Blut- und Gallen-farbstoffes', *Dtsch. Arch. Klin. Med.*, **112**, 476–503 (1913).
2. H.v. Tappeiner, 'Über die Wirkung fluorescierender Stoffe auf Infusorien nach Versuchen von O. Raab', *Münch. Med. Wochenschr.*, **47**, 5–13 (1900).
3. H.v. Tappeiner, A. Jesionek, 'Therapeutische Versuche mit Fluorescierenden Stoffen', *Münch. Med. Wochenschr.*, **50**, 2042–2051 (1903).
4. H. Auler, G. Banzer, 'Untersuchungen über die Rolle der Poprhyrine bei geschwulstkranken Menschen und Tieren', *Z. Krebsforsch.*, **53**, 65–72 (1942).
5. J. Brugsch, *Porphyrine*, J.-A. Barth-Verlag, Leipzig, 1959.
6. D.R. Doiron, C.J. Gomer (eds.), *Porphyrin Localization and Treatment of Tumors*, Alan R. Liss, New York, 1984.
7. D. Kessel, T.J. Dougherty (eds.), *Porphyrin Photosensitization*, Plenum Press, New York, 1983.
8. C.S. Foote, 'Mechanisms of Photosensitized Oxidation', *Science*, **162**, 963–970 (1968).
9. G. Laustriat, 'Molecular Mechanisms of Photosensitization', *Biochim.*, **68**, 771–778 (1986).
10. R. Haseloff, B. Ebert, B. Röder, 'Generation of Free Radicals by Photoactivation of Pheophorbide A, Hematoporphyrin and Protoporphyrin', *J. Photochem. Photobiol. B: Biology*, **3**, 593–603 (1989).
11. B. Röder, 'Biologische Bedeutung Aktivierter Sauerstoffspezies', *Biol. Rundschau*, **25**, 273–284 (1987).
12. J.B. Birks, *Organic Molecular Photophysics*, Vol. 2, Wiley, London, New York, 1975.
13. S. Yoshida, H. Saito, T. Fujioka, 'New Singlet Oxygen Generator for Chemical Oxygen–Iodine Lasers', *Chem. Appl. Phys. Lett.*, **49**, 1143–1144 (1986).
14. F. Wilkinson, W.P. Helman, A.B. Ross, 'Rate Constants for the Decay and Reactions of the Lowest Electronically Excited Singlet State of Molecular Oxygen in Solution. An Expanded and Revised Compilation', *J. Phys. Chem. Ref. Data*, **24**, 663–911 (1995).
15. B. Röder, D. Näther, 'Photophysical and Biophysical Properties of Tetrapyrroles', *SPIE Proc. Ser. Vol.*, **1525**, 377–384 (1991).
16. W. Spiller, H. Kliesch, D. Wöhrle, St. Hackbarth, B. Röder, 'Singlet Oxygen Quantum Yields of Different Photosensitizers in Polar Solvents and Micellar Solutions', *JPP*, **2**, 145–158 (1998).
17. M.E. Murphy, H. Sies, 'Chemiluminescence: Approaches to Detecting and Determining Reactive Oxygen Species', *Anal. Proc.*, **27**, 217–219 (1990); A.A. Krasnovsky, Jr, Ch.S. Foote, 'Time-resolved Measurements of Singlet Oxygen Dimol-sensitized Luminescence', *J. Am. Chem. Soc.*, **115**, 6013–6016 (1993).
18. St. Oelckers, M. Sczepan, Th. Hanke, B. Röder, 'Time-resolved Detection of Singlet Oxygen Luminescence in Red Ghost Cell Suspensions', *J. Photochem. Photobiol. B: Biol.*, **39**, 219–223 (1997).
19. St. Oelckers, Th. Ziegler, I. Michler, B. Röder, 'Time-resolved Detection of Singlet Oxygen Luminescence in Red Ghost Cell Suspensions: Concerning a Signal Component that can be Attributed to $^1O_2$-luminescence from Inside of a Native Membrane', *J. Photochem. Photobiol. B: Biol.*, **53**, 121–127 (1999).
20. J. Moan, K. Berg, 'The Photodegradation of Porphyrins in Cells Can be Used to Estimate the Lifetime of Singlet Oxygen', *Photochem. Photobiol.*, **53**, 549–553 (1991).
21. H. Meffert, F. Böhm, B. Röder, N. Sönnichsen, '1st Angeregter Singulettsauerstoff am PUVA-Therapie-Effekt beteiligt?', *Dermatol. Monatsschr.*, **168**, 387–393 (1982).
22. J. Zilberstein, A. Bromberg, A. Frantz, V. Rosenbach-Belkin, A. Kritzman, R. Pfeferman, 'Light-dependent Oxygen Consumption in Bacterio-chlorophyll-serine-treated Melanoma Tumors: On-line Determination Using a Tissue-inserted Oxygen Microsensor', *Photochem. Photobiol.*, **65**, 1009–1012 (1997).
23. (a) A.P. Losev, V.N. Knukshto, I.N. Zhuravkin, 'Direct Phosphorescent Detection of Primary Event of Photodynamic Action', *SPIE Proc.*, **2133**, 254–258 (1994). (b) A.P. Losev, V.N. Knukshto, I.N. Zhuravkin, 'Phosphorescent Direct Detection of Hyperoxia and Hypoxia in Tumors During Photodynamic Action', *SPIE Proc.*, **2392**, 11–120 (1995).
24. A.A. Krasnovsky, Jr, V.L. Sukhorukov, S.Y. Egorov, A.Y. Potapenko, 'Generation and Quenching of Singlet Molecular Oxygen by Furocoumarins. Direct Measurements', *Stud. Biophys.*, **114**, 149–158 (1986).
25. B.C. Wilson, M.S. Patterson, S.T. Flock, 'Indirect Versus Direct Techniques for the Measurement of the Optical Properties of Tissues', *Photochem. Photobiol.*, **46**, 601–608 (1987).
26. A.E. Profio, D.R. Doiron, 'Transport of Light in Tissue in Photodynamic Therapy', *Photochem. Photobiol.*, **46**, 591–599 (1987).

27. L.O. Svaasand, R. Ellingsen, 'Optical Penetration in Human Intracranial Tumors', *Photochem. Photobiol.*, **41**, 73–76 (1985).
28. R.S. Bodaness, D.S. King, 'The Two-photon Induced Fluorescence of the Tumor Localizing Photosensitizer Hematoporphyrin Derivative via 1064 nm Photons from a 20 ns Q-switched Nd : YAG Laser', *Biochem. Biophys. Res. Commun.*, **126**, 346–351 (1985).
29. Y. Yamashita, F. Moriyasu, S. Ono, T. Kimura, K. Kajimura, H. Someda, N. Hamato, M. Nabeshima, M. Sakai, M. Okuma, 'Photodynamic Therapy Using Pheophorbide A and Q-switched Nd : YAG Laser on Implanted Human Hepatocellular Carcinoma', *Gast. Jap.*, **26**, 623–627 (1991).
30. W.G. Fisher, W.P. Partridge, Jr, C. Dees, E.A. Wachter, 'Simultaneous Two-photon Activation of Type-I Photodynamic Therapy Agents', *Photochem. Photobiol.*, **66**, 141–155 (1997).
31. Th.H. Foster, R.S. Murant, R.G. Bryant, R.S. Knox, S.L. Gibson, R. Hilf, 'Oxygen Consumption and Diffusion Effects in Photodynamic Therapy', *Rad. Res.*, **126**, 296–303 (1991).
32. K. Teuchner, A. Pfarrherr, H. Stiel, W. Freyer, D. Leupold, 'Spectroscopic Properties of Potential Sensitizers for New Photodynamic Therapy Start Mechanisms via Two-step Excited Electronic States', *Photochem. Photobiol.*, **57**, 465–471 (1993).
33. H. Stiel, K. Teuchner, A. Paul, W. Freyer, D. Leupold, 'Two-photon Excitation of Alkyl-substituted Magnesium Phthalocyanine: Radical Formation via Higher Excited States', *Photochem. Photobiol. A: Chem.*, **80**, 289–298 (1994).
34. A. Andreoni, 'Two-step Photoactivation of Hematoporphyrin by Excimer-pumped Dye-laser Pulses', *J. Photochem. Photobiol. B: Biol.*, **1**, 181–193 (1987).
35. T. Karu, V. Letokhov, 'Possible Benefits of Two-quantum Excitation in ALA-PDT', *J. Photochem. Photobiol. B: Biol.*, **23**, 261–262 (1994).
36. D.Y.C. Fung, R.D. Miller, 'Effect of Dyes on Bacterial Growth', *Appl. Microbiol.*, **5**, 793–799 (1973).
37. H. Okamoto, T. Iwase, T. Morioka, 'Dye-mediated Bacterial Effect of HeNe-laser Irradiation on Oral Microorganisms', *Lasers Surg. Med.*, **12**, 450–458 (1992).
38. A. Andreoni, E.J. Land, V. Malatesa, A.J. McLean, T.G. Truscott, 'Triplet State Characteristics and Singlet Oxygen Generation Properties of Anthracyclines', *Biochim. Biophys. Acta*, **990**, 190–197 (1989).
39. M.R. Detty, P.B. Merkel, 'Chalcogeniumpyrylium Dyes as Potential Photochemotherapeutic Agents. Solution Studies on Heavy Atom Effects on Triplet Yields, Quantum Efficiencies of Singlet Oxygen Generation, Rates of Reaction with Singlet Oxygen, and Emission Quantum Yields', *J. Am. Chem. Soc.*, **112**, 3845–3855 (1990).
40. D. Kessel, 'Intracellular Localization of a Chalcogeniumpyrylium Dye Probed by Spectroscopy and Sites of Photodamage', *Photochem. Photobiol.*, **53**, 73–76 (1991).
41. K.S. Gulliya, J.L. Matthews, 'Photosensitization of Leukemic Cells and Normal Bone Marrow Cells by 514 nm Laser Light and Effects of Laser Light on Migration Inhibition and Lymphokine Response', *Cell Biol. Int. Rep.*, **12**, 305–312 (1988).
42. E.D. Sternberg, D. Dolphin, Ch. Brückner, 'Porphyrin-based Photosensitizers for Use in Photodynamic Therapy', *Tetrahedron*, **45**, 4151–4202 (1998).
43. R. Bonnett, 'Photosensitizers of the Porphyrin and Phthalocyanine Series for Photodynamic Therapy', *Chem. Soc. Rev.*, **24**, 19–33 (1995).
44. G.J. Smith, 'The Effect of Aggregation on the Fluorescence and the Triplet State Yield of HP', *J. Photochem. Photobiol.*, **41**, 123–126 (1985).
45. R. Bonnett, D.J. McGarvey, A. Harriman, E.J. Land, T.G. Truscott, U.-J. Winfield, 'Photophysical Properties of *meso*-Tetraphenylporphyrin and some *meso*-Tetra(hydroxyphenyl)porphyrins', *Photochem. Photobiol.*, **48**, 271–276 (1988).
46. A.A. Krasnovsky, K.V. Neverov, S.Y. Egorov, B. Röder, T. Lewald, 'Photophysical Studies of Pheophorbide A and Pheophytin A – Phosphorescence and Photosensitized Singlet Oxygen Luminescence', *J. Photochem. Photobiol. B: Biol.*, **5**, 245–254 (1990).
47. C.F. Borland, D.J. McGarvey, T.G. Truscott, R.J. Codgell, E.J. Land, 'Photophysical Studies of Bacteriochlorophyll A and Bacteriopheophytin A, Two Derivatives of Bacteriochlorophyll A', *J. Photochem. Photobiol. B: Biol.*, **1**, 93–101 (1987).
48. J.D. Spikes, J.C. Bommer, 'Photosensitizing Properties of Mono-L-aspartyl chlorin e_6 (Npe6): A Candidate Sensitizer for the Photodynamic Therapy of Tumors', *J. Photochem. Photobiol. B: Biol.*, **17**, 135–143 (1993).
49. G. Schermann, A. Völcker, K. Seikel, R. Schmidt, H.-D. Brauer, F.-P. Montforts, 'Potential Photosensitizers for Photodynamic Tumor Therapy – I. Photophysical Properties of Two Chlorin Derivatives', *Photochem. Photobiol.*, **51**, 45–51 (1990).
50. A.R. Morgan, G.M. Garbo, T.G. Truscott, 'Photophysical and Photochemical Properties of Purpurins', *SPIE Proc.*, **997**, 139–144 (1988); C.F. Borland, D.J. McGarvey, A.R. Morgan, T.G. Truscott, 'Laser Flash Photolysis of Purpurins: Novel Potential Photosensitizers of Interest in Photodynamic Therapy', *J. Photochem. Photobiol., B: Biol.*, **2**, 427–434 (1988).
51. P.F. Aramendia, R.W. Redmond, S. Nonell, W. Schuster et al., 'The Photophysical Properties of Porphycenes: Potential PDT Agents', *Photochem. Photobiol.*, **44**, 555–559 (1986).
52. B.G. Maiya, A. Harriman, J.L. Sessler, G. Hemmi, T. Murai, T.E. Mallouk, 'Ground- and Excited-state Spectral and Redox Properties of Cadmium(II) Texaphyrin', *J. Phys. Chem.*, **93**, 8111–8115 (1989).
53. B. Röder, D. Näther, T. Lewald, M. Braune, C. Nowak, W. Freyer, 'Photophysical Properties and Photodynamic

Activity in vivo of Some Tetrapyrroles', *Biophys. Chem.*, **35**, 303–312 (1990).
54. W. Spiller, H. Kliesch, D. Wöhrle, St. Hackbarth, B. Röder, G. Schurpfeil, 'Singlet Oxygen Quantum Yields of Different Photosensitizers in Polar Solvents and Micellar Solutions', *JPP*, **2**, 145–158 (1998).
55. G. Jori, E. Reddi, R. Biolo, G. Valduga, 'Phthalocyanines as Phototherapeutic Agents for Tumors', *SPIE Proc.*, **1525**, 367–376 (1991).
56. P.A. Firey, M.A.J. Rodgers, 'Photoproperties of a Silicon Naphthalocyanine: A Potential Photosensitizer for PDT', *Photochem. Photobiol.*, **45**, 535–538 (1987).
57. T.J. Dougherty, 'Studies on the Structure of Porphyrins Contained in Photofrin® II', *Photochem. Photobiol.*, **46**, 569–573 (1987).
58. Y. Fujiwara, M. Otha, I. Kojima, 'Zusammensetzung zur Behandlung von Tumoren', Offenlegungsschrift DE 128 09 093, 1987.
59. N. Fujii, N. Iijima, 'Method of Treating Gastric Mammary, Lung and Uterus Tumors', US Patent 4.393.071, 1983.
60. M.C. Berenbaum, S.L. Akande, R. Bonnett, '*meso*-Tetra (Hydroxyphenyl) Porphyrins, a New Class of Potent Tumour Photosensitisers with Favourable Selectivity', *Br. J. Cancer*, **54**, 717–725 (1986).
61. V. Sacchini, E. Melloni, R. Marchesini, 'Preliminary Clinical Studies with PDT by Topical TPPS Administration in Neoplastic Skin Lesions', *Lasers. Surg. Med.*, **7**, 6–11 (1988).
62. O.J. Balchum, D.J. Doiron, 'Photoradiation Therapy of Endobronchial Lung Cancer', *Clin. Chest. Med.*, **6**, 255–275 (1985).
63. B. Ehrenberg, F.M. Johnson, 'Spectroscopic Studies of Tetrabenzoporphyrins: Mg TBP, Zn TBP and H TBP', *Spectrochim. Acta*, **46A**, 1521–1532 (1990).
64. A.M. Richter, B. Kelly, J. Chow, D.J. Liu, G.H.N. Towers, D. Dolphin, J.G. Levy, 'Preliminary Studies on a More Effective Phototoxic Agent than Hematoporphyrin', *J. Natl. Cancer Inst.*, **79**, 1327–1332 (1987).
65. A.B. Segelman, I.K. Hagen, S.A. Chernomorsky, 'Highly Purified Pheophorbide A as a Photosensitizer in Human Bladder Carcinoma in vitro', *SPIE Proc.*, **847**, 205–209 (1987).
66. B. Röder, G. Wischnewsky, S. Nicklisch, E. Slawaticki, H. Meffert, 'Mittel zur Behandlung von Hauterkrankungen und Tumoren', Patent WP 248 282, 8 October, 1984.
67. W. Allen, 'PDT-Laser Hyperthermia Combination to be Tested as Cancer Therapy. PDT Drug from Japan Could Replace HPD', *Laser Med. Surg. Adv.*, **4**, 17–19 (1988).
68. B. Röder, 'Pheophorbides', in *Photodynamic Tumor Therapy, Second and Third Generation Photosensitizers*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 35–42, 1998.
69. E.M. Beems, T.M.A.R. Dubbelman, J. Lugtneburg, J.A. van Best, M.F.M.A. Smeets, J.P.J. Boegheim, 'Photosensitizing Properties of Bacteriochlorophyllin A and Bacteriochlorin A, Two Derivatives of Bacteriochlorophyll A', *Photochem. Photobiol.*, **46**, 639–643 (1987).
70. R.K. Pandey, S. Constantine, T. Tsuchida, G. Zheng, C.J. Medforth, M. Aoudia, A.N. Kozyrev, M.A.J. Rodgers, H. Kato, K.M. Smith, T.J. Dougherty, 'Synthesis, Photophysical Properties, in vivo Photosensitizing Efficacy, and Human Serum Albumin Binding Properties of Some Novel Bacteriochlorins', *J. Med. Chem.*, **40**, 2770–2779 (1997).
71. J.J. Schuitmaker, G.F.J.M. Vrensen, J.L. van Delft, D. de Wolff-Rouendaal, T.M.A.R. Dubbelman, A. de Wolf, 'Morphologic Effects of Bacteriochlorin A and Light in vivo on Intraocular Melanoma', *Invest. Opthal. Visual Sci.*, **32**, 639–643 (1991).
72. V. Rosenbachbelkin, L. Chen, L. Fiedor, I. Tregug, F. Pavlotsky, V. Brumfeld, Y. Salomon, A. Scherz, 'Serine Conjugates of Chlorophyll and Bacteriochlorophyll – Phototoxicity in-vitro and Tissue Distribution in Mice Bearing Melanoma Tumors', *Photochem. Photobiol.*, **64**, 174–181 (1996).
73. G.A. Kostenich, I.N. Zhuravkin, A.V. Furmanchuk, E.A. Zhavrid, 'Photodynamic Therapy with Chlorin e_6. A Morphologic Study of Tumour Damage Efficiency in Experiment', *J. Photochem. Photobiol. B: Biol.*, **11**, 307–318 (1991).
74. S.W. Taber, V.H. Fingar, C.T. Coots, T.J. Wieman, 'PDT Using Mono-L-aspartyl Chlorin e6 (Npe6) for Treatment of Cutaneous Disease: A Phase I Clinical Study', *Clin. Cancer Res.*, **4**, 2741–2746 (1998).
75. S.A. Blant, A. Woodtli, G. Wagnieres, C. Fontolliet, H. Vandenbergh, P. Monnier, 'In-vivo Fluence Rate Effect in Photodynamic Therapy of Early Cancers with Tetra(m-Hydroxyphenyl)Chlorin', *Photochem. Photobiol.*, **64**, 963–968 (1996).
76. G.M. Garbo, 'Purpurins and Related Chlorins', in *Photodynamic Tumor Therapy, Second and Third Generation Photosensitizers*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 51–61, 1998.
77. M. Guardiano, R. Biolo, G. Jori, K. Schaffner, 'Tetra-*n*-propylporphycene as a Tumour Localizer: Pharmacokinetic and Phototherapeutic Studies on Mice', *Cancer Lett.*, **44**, 1–6 (1989).
78. A. Harriman, B.G. Maiya, T. Murai, G. Hemmi, J.L. Sessler, T.E. Mallouk, 'Metallotexaphyrins: A New Family of Photosensitizers for Efficient Generation of Singlet Oxygen', *J. Chem. Soc., Chem. Commun.*, 314–315 (1989).
79. B.G. Maiya, M. Cyr, A. Harriman, J.L. Sessler, 'In vitro Photodynamic Activity of Diprotonated Sapphyrin: A 22π-electron Pentapyrrolic Porphyrin-like Macrocycle', *J. Phys. Chem.*, **94**, 3597–3601 (1990).
80. G. Kostenich, A. Orenstein, Z. Malik, B. Ehrenberg, 'Preclinical Photodynamic Therapy Studies with

Endogenous and New Exogenous Photosensitizers', in *Photodynamic Tumor Therapy, Second and Third Generation Photosensitizers*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 101–114, 1998.
81. C. Milanesi, C. Zhou, R. Biolo, G. Jori, 'Zn(II) Phthalocyanine as a Photodynamic Agent for Tumours. II. Studies on the Mechanism of Photosensitised Tumour Necrosis', *Br. J. Cancer*, **61**, 846–850 (1990).
82. E.K. Matthews, Z.J. Cui, 'Photodynamic Action of Sulfonated Aluminum Phthalocyanine (SALPC) on ARA-2J Cells, a Carcinoma Cell Line of Rat Exorcine Pancreas', *Br. J. Cancer*, **61**, 695–701 (1990).
83. B. Röder, D. Näther, T. Lewald, M. Braune, C. Nowak, W. Freyer, 'Photophysical Properties and Photodynamic Activity in vivo of Some Tetrapyrroles', *Biophys. Chem.*, **35**, 303–312 (1990).
84. C.Y. Anderson, K. Freye, K.A. Tubesing, Y.S. Li, M.E. Kenney, H. Mukhtar, C.A. Elmets, 'A Comparative-analysis of Silicon Phthalocyanine Photosensitizer for in-vivo Photodynamic Therapy of RIF-1 Tumors in C3H Mice', *Photochem. Photobiol.*, **67**, 332–336 (1998).
85. F.A. Firey, W.E. Ford, J.R. Sounik, M.A.J. Rodgers, 'Silicon Naphthalocyanine Triplet State Oxygen: A Reversible Energy-transfer Reaction', *J. Am. Chem. Soc.*, **110**, 7626–7630 (1988).
86. V. Cuomo, G. Jori, B. Richter, M.E. Kenney, M.A.J. Rodgers, 'Liposome-delivered Si(IV)-naphthalocyanine as a Photodynamic Sensitiser for Experimental Tumours: Pharmacokinetic and Phototherapeutic Studies', *Br. J. Cancer*, **62**, 966–970 (1990).
87. H. Kliesch, A. Weitemeyer, U. Michelsen, M. Shopova, D. Wöhrle, 'Naphthalocyanines as Photosensitizers for PDT', in *Photodynamic Tumor Therapy, Second and Third Generation Photosensitizers*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 75–86, 1998.
88. T.J. Dougherty, C.J. Gomer, B.W. Henderson, G. Jori, D. Kessel, M. Korbelik, J. Moan, Q. Peng, 'Photodynamic Therapy', *J. Natl. Cancer Inst.*, **90**, 889–905 (1998).
89. A. Hirth, U. Michelsen, D. Wöhrle, 'Photodynamische Tumortherapie', *Med. Chem.*, **33**, 84–94 (1999).
90. T.V. Akhlynina, A.A. Rosenkranz, D.A. Jans, A.S. Sobolev, 'Insulin-mediated Intracellular Targeting Enhances the Photodynamic Activity of Chlorin e_6', *Cancer Res.*, **55**, 1014–1019 (1995).
91. Q. Peng, J. Moan, T. Warloe, M. Nesland, C. Rimington, 'Distribution and Photosensitizing Efficiency of Porphyrins Induced by Application of Exogenous 5-Aminolevulinic Acid in Mice Bearing Mammary Carcinoma', *Int J. Cancer*, **52**, 433–443 (1992).
92. Z. Malik, G. Kostenich, L. Roitman, B. Ehrenberg, A. Orenstein, 'Topical Application of 5-Aminolevulinic Acid, DMSO and EDTA: Protoporphyrin IX Accumulation in Skin and Tumours of Mice', *J. Photochem. Photobiol. B: Biol.*, **28**, 213–233 (1995).
93. J.C. Kennedy, R.H. Pottier, D.C. Pross, 'Photodynamic Therapy with Endogenous Protoporphyrin IX: Basic Principles and Present Clinical Experience', *J. Photochem. Photobiol. B: Biol.*, **6**, 143–148 (1990).
94. C. Fritsch, B. Verwohlt, K. Bolsen, T. Ruzicka, G. Goerz, 'Influence of Topical Photodynamic Therapy with 5-Aminolevulinic Acid on Porphyrin Metabolism', *Arch. Dermatol. Res.*, **288**, 517–521 (1996).
95. R.M. Szeimies, S. Karrer, A. Sauerwald, M. Landthaler, 'Photodynamic Therapy with Topical Application of 5-Aminolevulinic Acid in the Treatment of Actinic Keratoses: An Initial Clinical Study', *Dermatology*, **192**, 246–251 (1996).
96. M.R. Stringer, P. Collins, D.J. Robinson, G.I. Stables, R.A. Sheehan-Dare, 'The Accumulation of Protoporphyrin IX in Plaque Psoriasis after Topical Application of 5-Aminolevulinic Acid Indicates a Potential for Superficial Photodynamic Therapy', *J. Invest. Dermatol.*, **107**, 76–81 (1996).
97. K.M. Hebeda, A.E. Saarnack, M. Olivo, H.J.C.M. Sterenborg, J.G. Wolbers, '5-Aminolevulinic Acid Induced Endogenous Porphyrin Fluorescence in 9L and C6 Brain Tumours and in the Normal Rat Brain', *Acta Neurochir. (Wien)*, **140**, 503–513 (1998).
98. Z. Smetana, Z. Malik, A. Orenstein, E. Mendelson, E. Ben-Hur, 'Treatment of Viral Infections with 5-Aminolevulinic Acid and Light', *Las. Surg. Med.*, **21**, 351–358 (1997).
99. C. Fritsch, Ch. Abels, A.E. Goetz, W. Stahl, K. Bolsen, T. Ruzicka, G. Goerz, H. Sies, 'Porphyrins Preferentially Accumulate in a Melanoma Following Intravenous Injection of 5-Aminolevulinic Acid', *Biol. Chem.*, **378**, 51–57 (1997).
100. A.R. Oseroff, G. Ara, D. Ohuoha, J. Aprille, J.C. Bommer, M.L. Yarmush, J. Foley, L. Cincotta, 'Strategies for Selective Cancer Photochemotherapy: Antibody-targeted and Selective Carcinoma Cell Photolysis', *Photochem. Photobiol.*, **46**, 83–96 (1987).
101. C. Dressler, U. Möller, T. Lewald, H.-P. Berlien, B. Röder, H.-J. Risse, 'Introduction of a Simple Model for Testing Immunoconjugates with Photosensitizers', in *Photodynamic Therapy and Biomedical Lasers*, eds. P. Spinelli, M. Dal Fante, R. Marchesini, Elsevier Science, New York, 164–168, 1992.
102. N.L. Krinick, Y. Sun, D. Joyner, J.D. Spikes, R.C. Straight, J. Kopeček, 'A Polymeric Drug Delivery System for the Simultaneous Delivery of Drugs Activable by Enzymes and/or Light', *J. Biomater. Sci. Polym. Edn.*, **5**, 303–324 (1994).
103. C. Bellemo, G. Jori, B.D. Rihter, M.E. Kenney, M.A.J. Rodgers, 'Si(IV)-Naphthalocyanine: Modulation of its Pharmacokinetic Properties through the Use of Hydrophilic Axial Ligands', *Cancer Lett.*, **65**, 145–150 (1992).
104. J.D. Spikes, N.L. Krinick, J. Kopeček, 'Photoproperties of a Mesochlorin e_6-*N*-(2-Hydroxypropyl)methacrylamide Copolymer Conjugate', *J. Photochem. Photobiol. A: Chem.*, **70**, 163–170 (1993).

105. G.M. Garbo, 'The Use of Liposomes, Emulsions or Inclusion Complexes may Potentiate in vivo Effects of SnET2', *SPIE Proc.*, **1203**, 118–125 (1990).
106. D. Kessel, A. Morgan, G.M. Garbo, 'Sites and Efficacy of Photodamage by Tin Etiopurpurin in vitro Using Different Delivery Systems', *Photochem. Photobiol.*, **54**, 193–196 (1991).
107. M.M. Zuk, B.D. Rihter, M.E. Kenney, M.A.J. Rodgers, 'Effect of Delivery System on the Pharmacokinetics and Tissue Distribution of bis(Di-Isobutyl Octadecylsiloxy)Silicon 2,3-Naphthalocyanine', *Photochem. Photobiol.*, **63**, 132–140 (1996).
108. K. Woodburn, C.K. Chang, S. Lee, B. Henderson, D. Kessel, 'Biodistribution and PDT Efficacy of a Ketochlorin Photosensitizer as a Function of the Delivery Vehicle', *Photochem. Photobiol.*, **60**, 154–159 (1994).
109. S.T. Mosley, J.L. Goldstein, M.S. Brown, J.R. Falck, R.G.W. Anderson, 'Targeted Killing of Cultured Cells by Receptor-dependent Photosensitization', *Proc. Natl. Acad. Sci.*, **78**, 5717–5721 (1981).
110. D. Kessel, 'Porphyrin-lipoprotein Association as a Factor in Porphyrin Localization', *Cancer Lett.*, **33**, 183–188 (1986).
111. G. Jori, E. Reddi, 'The Role of Lipoproteins in the Delivery of Tumour-targeting Photosensitizers', *Int. J. Biochem.*, **25**, 1369–1375 (1993).
112. E. Reddi, S. Cernuschi, R. Biolo, G. Jori, 'Liposome- or LDL-administered Zn(II)-phthalocyanine as a Photodynamic Agent for Tumours III. Effect of Cholesterol on Pharmacokinetic and Phototherapeutic Properties', *Lasers Med. Sci.*, **5**, 339–343 (1990).
113. W. Li, K. Qian, W. Huang, X. Zhang, W. Chen, 'Water Soluble C60-Liposome and the Biological Effect of C60 to Human Cervix Cancer Cells', *Chin. Phys. Lett.*, **11**, 207–210 (1994).
114. C. Hadjur, G. Wagnieres, P. Monnier, H. Vandenbergh, 'EPR and Spectrophotometric Studies of Free-radicals (O-2-Center-Dot(-), Center-Dot-Oh, Bpd-Ma-Center-Dot(-)) and Singlet Oxygen', *Photochem. Photobiol.*, **65**, 416–421 (1997).
115. S. Kimura, 'The Mechanism of Photodynamic Action of Pheophorbide and its Application to Cancer-phototherapy and -diagnosis', *Photomed. Photobiol.*, **15**, 1–12 (1993).
116. J.G. Moser, A. Rübner-Heuermann, P. Oehr, H. Scheer, A. Verwoorts, S. Andrees, 'Carrier Systems in PDT: On the Way to Novel Anti-tumor Drugs', *SPIE Proc.*, **2325**, 92–99 (1994).
117. B. Röder, St. Hackbarth, G. Wöhlecke, 'Verfahren zur Nutzung von Dendrimeren als Multiplikatoren zur Verabreichung von Photosensibilisatoren', Patentantrag, Registration Nr. 199 36 997.6 (2 August, 1999), Germany.
118. R.W. Steiner, I.S. Melnik, M. Klotz, 'Irradiation Modalities and Irradiation Control', in *Photodynamic Tumor Therapy*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 137–150, 1998.
119. M.H. Schmidt, D.M. Bajic, K.W. Reichert II, T.S. Martin, G.A. Meyer, H.T. Whelan, 'Light-emitting Diodes as a Light Source for Intraoperative Photodynamic Therapy', *Neurosurg.*, **38**, 552–556 (1996).
120. M.L. DeJode, J.A. McGilligan, M.G. Dilkes, I. Cameron, P.B. Hart, M.F. Grahn, 'A Comparison of Novel Light Sources for Photodynamic Therapy', *Lasers Med. Sci.*, **12**, 260–268 (1997).
121. C. Whitehurst, K. Byrne, J.V. Moore, 'Development of an Alternative Light Source to Lasers for Photodynamic Therapy: 1. Comparative in Vitro Dose Response Characteristics', *Lasers Med. Sci.*, **8**, 259–267 (1993).
122. P. Lenz, 'In vivo Excitation of Photosensitizers by Infrared Light', *Photochem. Photobiol.*, **62**, 333–338 (1995).
123. R.L. Lipson, M.J. Gray, E.J. Baldes, 'Hematoporphyrin Derivative for Detection and Management of Cancer', *Proc. 9th Int. Cancer Congress*, Tokyo, Japan, 1966.
124. T.J. Dougherty, J.E. Kaufman, A. Goldfarb, K.R. Weishaupt, A. Mittleman, 'Photoradiation Therapy for the Treatment of Malignant Tumours', *Cancer Res.*, **38**, 2635–2638 (1978).
125. J.F. Evensen, 'The Use of Porphyrins and Non-ionizing Radiation for Treatment of Cancer', *Acta Oncol.*, **34**, 1103–1110 (1995).
126. D. Jocham, 'Clinical Experiences and Expectations', in *Photodynamic TumorTherapy*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 213–225 (1998).
127. D. Jocham, 'Clinical Experiences and Expectations', in *Photodynamic Tumor Therapy*, ed. J.G. Moser, Harwood Acad. Publ., 213–225, 1998.
128. A. Hirth, U. Michelsen, D. Wöhrle, 'Photodynamische Tumortherapie', *Med. Chem.*, **33**, 84–94 (1999).
129. T.J. Dougherty, C.J. Gomer, B.W. Henderson, G. Jori, D. Kessel, M. Korbelik, J. Moan, Q. Peng, 'Photodynamic Therapy', *J. Natl. Cancer Inst.*, **90**, 889–905 (1998).
130. R. Bonnett, 'Photosensitizers of the Porphyrin and Phthalocyanine Series for Photodynamic Therapy', *Chem. Soc. Rev.*, **24**, 19–33 (1995).
131. J.F. Evensen, 'The Use of Porphyrins and Non-ionizing Radiation for Treatment of Cancer', *Acta Oncol.*, **34**, 1103–1110 (1995).
132. C. Fritsch, B. Verwohlt, K. Bolsen, T. Ruzicka, G. Goerz, 'Influence of Topical Photodynamic Therapy with 5-Aminolevulinic Acid on Porphyrin Metabolism', *Acta Dermatol. Res.*, **288**, 517–521 (1996).
133. R.M. Szeimies, S. Karrer, A. Sauerwald, M. Landthaler, 'Photodynamic Therapy with Topical Application of 5-Aminolevulinic Acid in the Treatment of Actinic Keratoses: An Initial Clinical Study', *Dermatology*, **192**, 246–251 (1996).
134. S.W. Taber, V.H. Fingar, C.T. Coots, T.J. Wiemann, 'Photodynamic Therapy Using Mono-L-aspartyl Chlorin

e_6 (Npe6) for the Treatment of Cutaneous Diseases: A Phase I Clinical Study', *Clin. Cancer Res.*, **4**, 2741–2746 (1998).
135. B. Röder, 'Pheophorbides', in *Photodynamic Tumor Therapy, Second and Third Generation Photosensitizers*, ed. J.G. Moser, Harwood Acad. Publ., Amsterdam, 35–42, 1998.
136. R.K. Pandey, C.K. Herman, 'Shedding Some Light on Tumours', *Chem. Ind.*, 739–743 (1998).
137. P. Collins, D.J. Robinson, M.R. Stringer, G.I. Stables, R.A. Sheehan-Dare, 'The Variable Response of Plaque Psoriasis after a Single Treatment with Topical 5-Aminolaevulinic Acid Photodynamic Therapy', *J. Dermatol.*, **137**, 743–749 (1997).
138. N. Razum, A. Snyder, D. Doiron, 'SnET2: Clinical Update', *SPIE Proc.*, **2675**, 43–46 (1996).
139. J. Leung, 'Photosensitizers of Photodynamic Therapy', *Semin. Oncol.*, **21**, 4–10 (1994).
140. M. Renschler, A. Yuen, T. Panella, 'Photodynamic Therapy Trials with Lutetium Texaphyrin', abstract, *Photochem. Photobiol.*, **65**, 475 (1997).
141. M.M. Judy, 'Photodynamic Inactivation of Enveloped Viruses: Potential Application for Blood Banking', *J. Clin. Laser Med. Surg. Ber. GML*, 49–52 (1990).
142. M.M. Judy, J.L. Matthews, J.T. Newman, H.L. Skiles, R.L. Borlack, 'In vitro Photodynamic Inactivation of Herpes Simplex Virus with Sapphyrins: 22π-electron Porphyrin-like Macrocycle', *Photochem. Photobiol.*, **53**, 101–107 (1991).
143. D.W. Dixon, M.S. Kim, V. Kumar, G. Obara, L.G. Marzilli, R.F. Schinazi, 'Amino- and Hydroxytetraphenylporphyrins with Activity Against the Human Immunodeficiency Virus', *Antivir. Chem. Chemother.*, **3**, 279–282 (1992).
144. A.P. Savitsky, A.I. Turkin, E.W. Turkina, T.W. Tscherednikova, 'Podavlenije Virusnoi Infekzii s Pomoschtschju Fotoimmunotoksina', *Dokl. AN USSR*, **304**, 1258–1262 (1989).
145. E. Gragoudas, U. Schmidt-Erfurth, M. Sickenkey, 'Results and Preliminary Dosimetry of Photodynamic Therapy for Choroidal Neovascularization in Age-related Macular Degeneration in a Phase I/II Study', abstract, *Assoc. Res. Vision Ophthalm.*, **38**, 73 (1997).
146. G. Bertoloni, F. Rossi, G. Valduga, G. Jori, J. van Lier, 'Photosensitizing Activity of Water- and Lipid-soluble Phthalocyanines on *Escherichia coli*', *FEMS Microbiol. Lett.*, **71**, 149–156 (1990).
147. G. Bertoloni, B. Salvato, M. Dall Acqua, M. Vazzoler, G. Jori, 'Protoporphyrin Photosensitization of *Enterococcus hirae* and *Candida albicans* Cells', *Lasers Life Sci.*, **5**, 267–275 (1993).
148. Z. Malik, H. Ladan, Y. Nitzan, 'The Bactericidal Activity of a Deuteroporphyrin-hemin Mixture on Gram Positive Bacteria. A Microbiological and Spectroscopic Study', *J. Photochem. Photobiol. B: Biol.*, **6**, 419–430 (1990).
149. X.-M. Lu, A.J. Fischmann, E. Stevens, T.T. Lee, L. Strong, R.G. Tompkins, 'Sn-chlorin e_6 Antibacterial Immunoconjugates', *J. Immunol. Methods*, **156**, 85–99 (1992).
150. M. Wilson, J. Dobson, W. Harvey, 'Sensitization of *Streptococcus sanguis* by Light from a HeNe-laser', *Lasers Med. Sci.*, **8**, 69–73 (1993).
151. F. Litvack, W.S. Grundfest, J.S. Forrester, M.C. Fishbein, H.J.C. Swan, E. Gorday, D.M. Rider, I. Stuart, T.J. Pacala, J.B. Leudenslager, 'Effects of HPD and PDT on Atherosclerotic Rabbits', *Am. J. Cardiol.*, **56**, 667–671 (1985).
152. A. Visona, V. Cuomo, G. Zanetti, L. Lusiani, A. Pagnan, G. Jori, 'Accumulation of Zinc-Phthalocyanine in the Aorta of Atherosclerotic Rabbits', *Lasers Med. Sci.*, **4**, 167–170 (1989).
153. G. Graschew, M. Shopowa, 'Hypoxia, Misonidazole and Hyperthermia in Photodynamic Therapy of Tumors', *Lasers Med. Sci.*, **1**, 181–186 (1986).
154. P. Baas, I.P.J. van Geel, H. Oppelaar, M. Meyer, J.H. Beynen, N. van Zandwijk, F.A. Stewart, 'Enhancement of Photodynamic Therapy by Mitomycin C: A Preclinical and Clinical Study', *Br. J. Cancer*, **73**, 945–951 (1996).
155. R. Kallman, 'The Phenomena of Reoxygenation and its Implications for Fractionated Radiotherapy', *Radiology*, **105**, 135–140 (1972).
156. G. Graschew, M. Shopowa, 'Photodynamic Therapy and γ – Irradiation of Tumours: Effect of Tumour-cell Reoxygenation', *Lasers Med. Sci.*, **1**, 193–195 (1986).
157. H. Kostron, R.M. Swartz, D.C. Miller, R.L. Martuza, 'The Interaction of Hematoporphyrin Derivative, Light and Ionizing Radiation in a Rat Glioma Model', *Cancer*, **57**, 964–970 (1986).
158. C.A. Morton, C. Whitehurst, H. Moseley, J.H. McColl, J.V. Moore, R.M. Mackie, 'Comparison of Photodynamic Therapy with Cryotherapy in the Treatment of Bowen's Disease', *Br. J. Dermatol.*, **135**, 766–771 (1996).
159. M. Korbelik, 'Induction of Tumor Immunity by Photodynamic Therapy', *J. Clin. Laser Med. Surg.*, **14**, 329–334 (1996).

Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences

Yukihiro Ozaki
Kwansei-Gakuin University, Nishinomiya, Japan

Isao Noda
The Procter and Gamble Company, Cincinnati, USA

1 Introduction **1**
1.1 History of Two-dimensional Correlation Spectroscopy 1
1.2 Characteristics of Two-dimensional Correlation Spectroscopy 2
2 Theory **3**
2.1 Mathematical Treatment 3
2.2 Properties of Generalized Two-dimensional Correlation Spectra 3
3 Application to Biomedical Sciences **4**
3.1 Applications to Protein Research 5
3.2 Other Biomedical Applications 13
Abbreviations and Acronyms **17**
Related Articles **17**
References **17**

Two-dimensional (2-D) vibrational correlation spectroscopy offers a new opportunity for analyzing vibrational spectra. In the 2-D approach, spectral peaks are spread over the second dimension, thereby simplifying the visualization of complex spectra consisting of many overlapped bands, and enhancing spectral resolution. Correlation peaks appearing in the synchronous and asynchronous maps represent in-phase and out-of-phase variation tendencies of corresponding band intensities, respectively. 2-D correlation spectroscopy is very useful in unraveling infrared (IR), near-infrared (NIR), and Raman spectra of biological molecules and biomedical materials. For example, it has a powerful deconvolution ability for highly overlapped amide I, amide II, and amide III bands of proteins, enabling these bands to be assigned to various secondary structures. 2-D NIR correlation spectroscopy provides new insights into the hydration of proteins, and can also be used to investigate the order of perturbation-dependent structural changes.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

In 2-D correlation spectroscopy, a spectrum is obtained as a function of two independent spectral variables, such as wavenumbers.[1,2] Introduction of the second spectral dimension provides a new opportunity for analyzing spectra. 2-D correlation spectroscopy was initially evolved as an analytical data treatment method for sinusoidally varying dynamic IR signals detected in studies of molecular interactions.[1] In 1993 it was extended to generalized 2-D correlation spectroscopy, which uses a more generally applicable, yet reasonably simple, mathematical formalism.[2] The application of the newly proposed 2-D correlation spectroscopy to various other types of spectroscopy, such as Raman and NIR spectroscopy, is straightforward.[2]

1.1 History of Two-dimensional Correlation Spectroscopy

The basic working concept for 2-D vibrational correlation spectroscopy was developed much later than more well-established 2-D correlation techniques pioneered in the field of nuclear magnetic resonance (NMR) spectroscopy. Today it may seem surprising that this powerful yet simple idea, of obtaining a spectrum as a function of two independent spectral axes, was not applied to vibrational spectroscopy until a decade or so ago. In fact, several vibrational spectroscopists, inspired by the prolific expansion of successful 2-D NMR activities, had previously expressed some interest in such adaptations of multidimensional correlation schemes to their field. However, many believed that the fundamental differences in the timescales dealt with by different spectroscopic techniques were an insurmountable obstacle in applying the 2-D idea developed for NMR to common vibrational spectroscopy.

It is well known that the characteristic time span of molecular vibrational processes (picosecond range) is many orders of magnitude shorter than the typical spin relaxation times (microsecond range or even longer) encountered in NMR. In NMR, the double Fourier transformation of a set of time domain data collected under multiple-pulse excitations generates 2-D spectra. Direct adaptation of such a procedure, based on pulsed excitations, to conventional vibrational spectroscopy is rather difficult. Ordinary commercial IR spectrometers, for example, cannot adequately provide the rapid excitation and detection of vibrational relaxation responses needed to carry out such measurements. Thus, the specific experimental procedure developed for 2-D NMR had to be fundamentally modified before being applied to IR or Raman spectroscopy.

Early attempts at 2-D vibrational spectroscopy were still heavily influenced by the basic approach set by NMR, where only time-dependent changes in spectral signals are utilized for 2-D correlation. Thus, the first generation of 2-D vibrational correlation spectra were obtained from IR experiments based on the detection of various relaxation processes, which are much slower than vibrational relaxations but are closely associated with molecular-scale phenomena.[1] These slow relaxation processes can be studied with a conventional IR spectrometer using a simple time-resolved technique. Dynamic 2-D IR rheo-optics experiments, based on the IR dichroic measurement of orientational relaxation processes of molecular segments of various plastic films subjected to a small amplitude oscillatory deformation, gained considerable popularity, especially among polymer scientists. In addition to such mechanically stimulated experiments, similar 2-D IR works based on time-dependent IR signals induced by sinusoidally varying electrical or photoacoustic perturbations have also been tried.

In 1993 the concept of 2-D vibrational correlation spectroscopy was significantly expanded to include much more general forms of spectroscopic applications.[2] The mathematical procedure to yield 2-D correlation spectra from sinusoidally varying time-dependent spectral signals was modified to handle arbitrary time dependences which were much more complex than a simple sinusoidal wave. The type of spectral signals analyzed by the newly proposed 2-D correlation method became virtually limitless, covering IR, Raman, X-ray, ultraviolet (UV), fluorescence, and many more. Most importantly, it was finally recognized that this powerful 2-D correlation technique could be adopted for the analysis of spectral signals which change as functions not only of time but also of any other reasonable physical variables (such as temperature, concentration, pH, pressure, or any combination of these). Such a generalized correlation idea truly revolutionized the scope of potential applications for 2-D spectroscopy, especially in the field of IR and Raman spectroscopy.

1.2 Characteristics of Two-dimensional Correlation Spectroscopy

Figure 1 shows a basic scheme for constructing generalized 2-D correlation spectra from perturbation-induced dynamic fluctuations of spectroscopic signals.[2] In the generalized 2-D correlation method, there are many different types of external perturbations which could be used to stimulate a system of interest. When an external perturbation is applied to the system, various chemical constituents of the system are selectively excited. The excitation and subsequent relaxation processes towards the equilibrium are monitored with various kinds of electromagnetic probes. In order to generate generalized 2-D correlation spectra, a set of dynamic spectra must be calculated first (see section 2.1). Then, two kinds of correlation spectra, known as synchronous and asynchronous spectra, are constructed from the dynamic spectra.

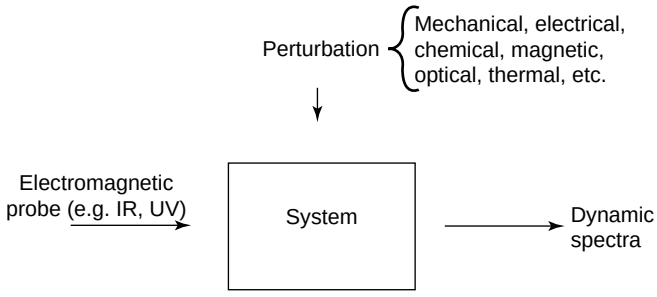

Figure 1 A general scheme for constructing generalized 2-D correlation spectra. (Reproduced from Noda[2] by permission of Society for Applied Spectroscopy. Copyright 1993.)

The new generalized 2-D correlation method can handle signals that fluctuate as an arbitrary function of time. Thus, a large number of time-dependent spectral variations associated with transient phenomena can be systematically analyzed. Furthermore, it has become apparent that the same 2-D correlation concept can be equally well applied to static systems of interest using any physical variables other than time, such as temperature, pressure, or even concentration.[3–21] This has opened up the possibility of introducing the versatility of 2-D correlation analysis to much wider ranges of applications, including electrochemistry, photochemistry, and reaction kinetics.

The major advantages of generalized 2-D correlation spectroscopy may be summarized as follows.

(1) Spectral resolution is enhanced by spreading spectral peaks over the second dimension. In other words, spectral information not readily accessible from one-dimensional (1-D) spectra is available.
(2) More detailed investigations of various inter- and intramolecular interactions are possible through selective correlation of peaks.
(3) The specific order of the spectral intensity changes can be probed.
(4) Band assignments can be made based upon correlations between various bands.
(5) Correlations between bands in two different spectroscopic techniques, such as Raman and IR spectroscopies, can be studied.

2 THEORY

2.1 Mathematical Treatment

The detailed mathematical treatment for generalized 2-D correlation spectroscopy has been described by Noda[2,21] elsewhere. Here, we describe the basic concept briefly. Let us consider a time-dependent fluctuation of spectral intensity $y(\nu, t)$ observed for a period of time between $-T/2$ and $T/2$. The variable ν can be any appropriate spectral variable, e.g. IR wavenumber, Raman shift, or even X-ray diffraction angle. In order to generate generalized 2-D correlation spectra from perturbation-induced dynamic fluctuations of spectroscopic signals, the dynamic spectrum must be calculated first. It is defined in Equation (1) as

$$\tilde{y}(\nu, t) = \begin{cases} y(\nu, t) - \bar{y}(\nu) & \text{for } -T/2 \le t \le T/2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $\bar{y}(\nu)$ is the reference spectrum. Although the selection of the reference spectrum is somewhat arbitrary, $\bar{y}(\nu)$ is usually set to be the static or time-averaged spectrum, defined as shown in Equation (2).

$$\bar{y}(\nu) = \frac{1}{T} \int_{-T/2}^{T/2} y(\nu, t)\, \mathrm{d}t \quad (2)$$

For many practical systems the reference spectrum may also be set to zero.

For the next step, it is necessary to Fourier transform the dynamic spectra measured in the time domain into the frequency domain.[2,21] The dynamic spectral intensity fluctuations $\tilde{y}(\nu_1, t)$ observed at some spectral variable ν_1 are Fourier transformed to $Y_1(w)$ as shown in Equation (3)

$$\begin{aligned} Y_1(w) &= \int_{-\infty}^{\infty} \tilde{y}(\nu_1, t)\, \mathrm{e}^{-iwt}\, \mathrm{d}t \\ &= Y_1^{\mathrm{Re}}(w) + iY_1^{\mathrm{Im}}(w) \end{aligned} \quad (3)$$

where $Y_1^{\mathrm{Re}}(w)$ and $Y_1^{\mathrm{Im}}(w)$ are the real and imaginary components of the complex Fourier transform of $\tilde{y}(\nu_1, t)$. The Fourier frequency w represents the individual frequency component of the time-dependent variation of $\tilde{y}(\nu_1, t)$. Similarly, $Y_2^*(w)$, the conjugate of the Fourier transform of dynamic spectral intensity $\tilde{y}(\nu_2, t)$ at spectral variable ν_2, is given by Equation (4).

$$\begin{aligned} Y_2^*(w) &= \int_{-\infty}^{\infty} \tilde{y}(\nu_2, t)\, \mathrm{e}^{+iwt}\, \mathrm{d}t \\ &= Y_2^{\mathrm{Re}}(w) - iY_2^{\mathrm{Im}}(w) \end{aligned} \quad (4)$$

Now, one can define the complex 2-D correlation intensity between $\tilde{y}(\nu_1, t)$ and $\tilde{y}(\nu_2, t)$ by Equation 5.

$$\Phi(\nu_1, \nu_2) + i\Psi(\nu_1, \nu_2) = \frac{1}{\pi T} \int_0^{\infty} Y_1(w) Y_2^*(w)\, \mathrm{d}w \quad (5)$$

The real and imaginary components of the complex 2-D correlation intensities, $\Phi(\nu_1, \nu_2)$ and $\Psi(\nu_1, \nu_2)$, are the generalized synchronous and asynchronous correlation spectra of the dynamic spectral intensity variations.[2,21]

The synchronous spectrum represents the overall similarity of the time-dependent behavior of spectral intensity variations measured at two separate wavenumbers, while the asynchronous spectrum represents the overall differences in dynamic spectra.

The synchronous and asynchronous spectra are explained in more detail later. It is important to note that the *time* used in the above 2-D correlation analysis can actually be regarded as a general variable which could be replaced by any other reasonable physical variable, such as temperature, pressure, or distance.[2]

This extension of the correlation concept, from a simple time-series analysis of transient spectral fluctuations to a more general scheme of spectral analysis, made 2-D correlation spectroscopy a much more versatile and universally applicable tool. While the formal definition of the 2-D correlation intensities is mathematically rigorous, it is rather cumbersome and time-consuming to calculate the actual correlation intensities based upon the above definition. Therefore, Noda[21] developed a new algorithm which allows 2-D correlation spectra to be calculated much more simply. The method does not rely on Fourier transformation but is based on the Hilbert transform, offering an easier and more efficient way for calculating the 2-D correlation. In this case the 2-D correlation spectra can be calculated according to Equations (6) and (7)

$$\Phi(\nu_1, \nu_2) = \frac{1}{T} \int_{-T/2}^{T/2} \tilde{y}(\nu_1, t)\tilde{y}(\nu_2, t)\, \mathrm{d}t \quad (6)$$

$$\Psi(\nu_1, \nu_2) = \frac{1}{T} \int_{-T/2}^{T/2} \tilde{y}(\nu_1, t)\tilde{z}(\nu_2, t)\, \mathrm{d}t \quad (7)$$

where the orthogonal spectrum $\tilde{z}(\nu_2, t)$ is a time-domain Hilbert transform of $\tilde{y}(\nu_2, t)$.

A more detailed mathematical treatment for calculating synchronous and asynchronous 2-D correlation spectra has been described.[21]

2.2 Properties of Generalized Two-dimensional Correlation Spectra

Figure 2(a) and (b) illustrates a schematic example of synchronous and asynchronous 2-D correlation spectra

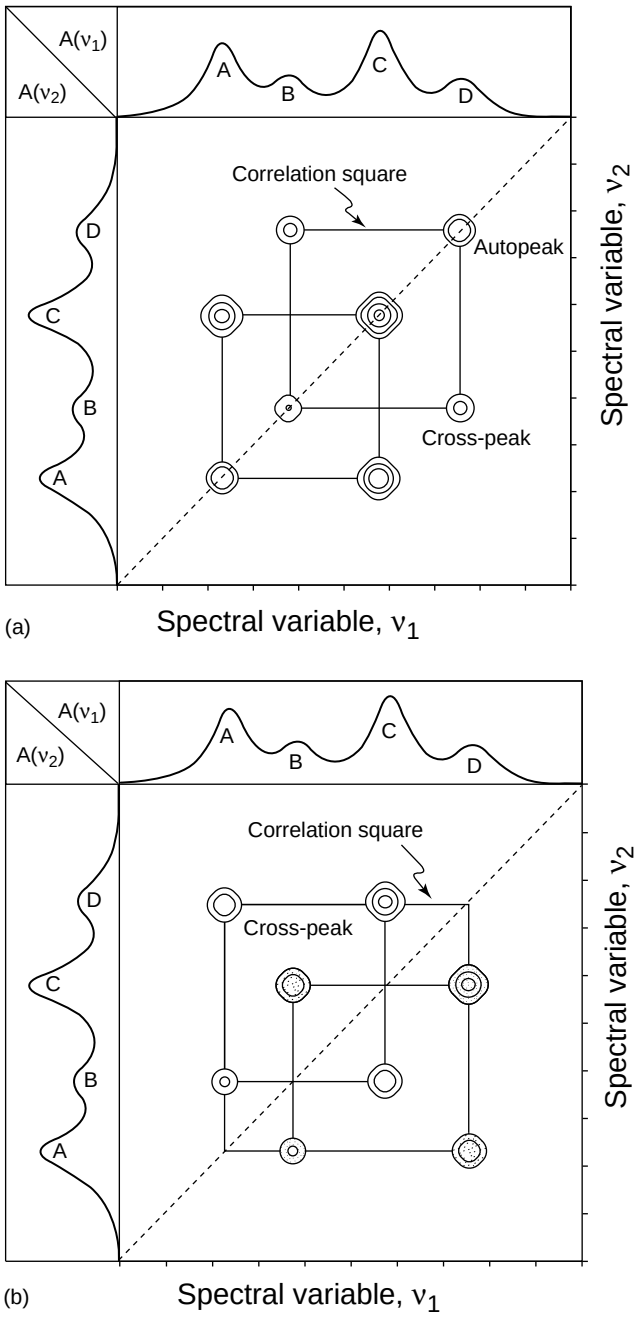

Figure 2 (a) Synchronous and (b) asynchronous 2-D correlation spectra. A 1-D reference spectrum is also provided at the top and side of the 2-D map. (Reproduced from Noda[2] by permission from Society for Applied Spectroscopy. Copyright 1993.)

plotted as a contour map.[2] The synchronous spectrum represents the simultaneous or coincidental changes of spectral intensities at ν_1 and ν_2, while the asynchronous spectrum represents sequential, or unsynchronized, variations. More specific properties of these correlation spectra are now discussed.

The synchronous spectrum is symmetric with respect to a diagonal line corresponding to spectral coordinates $\nu_1 = \nu_2$. The intensities of peaks located at diagonally opposing positions correspond to the autocorrelation functions of spectral intensity variations. These diagonal peaks are therefore referred to as autopeaks. In the spectrum shown in Figure 2(a), there are four autopeaks located at the spectral coordinates A, B, C, and D. The intensities of the autopeaks represent the overall extent of dynamic fluctuations of spectral signals. Cross-peaks located at the off-diagonal positions of a synchronous 2-D spectrum represent the simultaneous changes in signal intensities at two different wavenumbers. In the example spectrum, bands A and C are synchronously correlated, as are bands B and D. If the sign of the cross-peak is positive, the intensities at corresponding spectral variables increase (or decrease) together. On the other hand, the sign becomes negative if one of the spectral intensities is increasing while the other is decreasing.

The asynchronous 2-D correlation spectrum shown in Figure 2(b), which consists exclusively of off-diagonal cross-peaks, provides information complementary to the synchronous spectrum. It represents sequential changes in spectral intensities distinguished by different temperature (time, concentration, etc.) events measured at ν_1 and ν_2. Asynchronous cross-peaks develop only if the intensities of two dynamic spectral intensities vary out of phase with each other for some Fourier-frequency components of signal fluctuations. This feature is especially useful in differentiating between overlapped bands arising from different spectral origins or moieties. The sign of an asynchronous cross-peak can be either negative (as indicated by the shaded areas) or positive. It becomes positive if the intensity change at ν_1 occurs predominantly before ν_2 (i.e. in the earlier part of the sequentially arranged spectral data set). This rule, however, is reversed if $\Phi(\nu_1, \nu_2) < 0$. The example spectrum in Figure 2(b) indicates that the intensity changes at bands A and C occur before the changes at B and D (i.e. in the later part of the data set).

3 APPLICATION TO BIOMEDICAL SCIENCES

2-D correlation spectroscopy is a powerful tool in the analysis of complicated spectra of biological molecules and biomedical materials. For example, it enables the highly overlapped amide I, II, and III bands of proteins to be resolved into component bands ascribed to various secondary structures. It is also possible to investigate the correlations between amide I, II, and III

bands. Moreover, correlation between IR bands due to fundamental amide modes and NIR bands due to their combinations may be studied.

The first examples of the application of 2-D correlation spectroscopy to biomedical sciences were 2-D IR studies on the stratum corneum of human skin[1] and on keratin films from human hair.[20] Recently, generalized 2-D correlation IR spectroscopy has been extensively employed to investigate the secondary structure of proteins. Nabet and Pézolet[12] demonstrated the potential of 2-D IR correlation spectroscopy in exploring the conformation of proteins using hydrogen–deuterium (H–D) exchange. Sefara and Richardson[13] investigated the structure of β-lactoglobulin using both 2-D IR and 2-D NIR correlation spectroscopy. Wang et al.[14] reported 2-D NIR correlation analysis of concentration-dependent spectral changes in albumin solutions at various temperatures. New insight has been gained into the hydration and unfolding processes of secondary structures of ovalbumin by studying temperature-dependent correlation patterns in 2-D synchronous and asynchronous spectra. The 2-D correlation spectra for this study were constructed from concentration-perturbed NIR spectra at different temperatures. In addition to the standard 2-D IR and 2-D NIR analysis, 2-D infrared/near-infrared (IR/NIR) heterospectral correlation analyses of the thermal unfolding of ribonuclease A (RNase A) were carried out by Schultz et al.[15] The 2-D correlation analyses suggested that different structure elements in RNase A respond slightly differently to a temperature increase.

Gadaleta et al.[8] applied 2-D IR correlation spectroscopy to synthetic and biological apatites. They investigated the hydroxyapatite (HA) maturation, complex spectral contours undergoing changes as crystals ripen.

In section 3.1 the applications of 2-D correlation spectroscopy are described, by dividing them into two areas: applications to protein research, and other biomedical applications.

3.1 Applications to Protein Research

3.1.1 Two-dimensional Infrared Correlation Spectroscopy Study on the Secondary Structure of Proteins using Hydrogen–Deuterium Exchange

Fourier transform infrared (FTIR) spectroscopy has emerged as a valuable tool for qualitative and quantitative estimation of the secondary structure of proteins. Although many attempts have been made to assign the components of the amide I band to secondary structures such as α-helices and β-sheets, it is still a matter of controversy. 2-D IR correlation spectroscopy enhances the spectral resolution of the amide I and II regions and makes it possible to assign some of the amide I and II bands to given conformations.

Nabet and Pézolet[12] reported a 2-D FTIR correlation spectroscopy study of the secondary structure of myoglobin. They used H–D exchange of the amide protons as an external perturbation to generate the 2-D synchronous and asynchronous spectra. Because the amide protons associated with each conformation are not exchanged simultaneously, the different conformational contributions of the amide bands can be separated. Analysis of the synchronous and asynchronous maps of myoglobin shows that this method is very useful for unraveling the different components of the poorly resolved amide I, II, and II′ bands of proteins.

Microgram quantities of thin myoglobin films were deposited on attenuated total reflection (ATR) crystals, and H–D exchange was induced by hydrating the films with a flow of nitrogen containing D_2O vapor. In general, there are two kinds of amide groups involved in the deuteration kinetics: some amide groups that are readily accessible are exchanged rapidly at the beginning of the deuteration process, whereas those in structures that are less accessible to the solvent show slower exchange kinetics. Thus, to separate the fast kinetics from the slower ones more efficiently, different sampling times were used.

Figure 3(a) shows the synchronous 2-D IR correlation spectra of myoglobin calculated from the first 10 spectra recorded during the H–D exchange process, while Figure 3(b) shows the asynchronous spectra. Figure 4 depicts the corresponding spectra calculated from 10 spectra obtained approximately 1 h after the beginning of the H–D exchange process.[12] The amide I region of the synchronous correlation map for the rapidly exchanging protons (Figure 3a) shows three correlation peaks at 1675, 1640, and 1615 cm^{-1}. These amide I components are respectively assigned to the β-turns, random coil, and intermolecular β-sheets often found in aggregated proteins. Therefore, it seems that the amide groups associated with these three conformations are exchanged first during the deuteration process. The strongest peak in the synchronous map is observed in the amide II region at 1530 cm^{-1}, while in the amide II′ region two major peaks can be identified at 1440 and 1350 cm^{-1}.

The asynchronous map of myoglobin for the rapidly exchanging protons develops two cross-peaks in the amide I region, at 1675–1640 cm^{-1} and at 1640–1615 cm^{-1}, confirming that the three peaks at 1675, 1640, and 1615 cm^{-1} appearing in the synchronous map are due to three different conformations.

In the synchronous map for the slowly exchanging protons, one autopeak is observed at 1655 cm^{-1}, a frequency that is generally assigned to the amide I mode of the α-helix conformation. Thus, it is very likely that the amide protons of the α-helix conformation are exchanged more slowly than those associated with intermolecular β-sheets, random coil, and β-turns.[12] The other intense

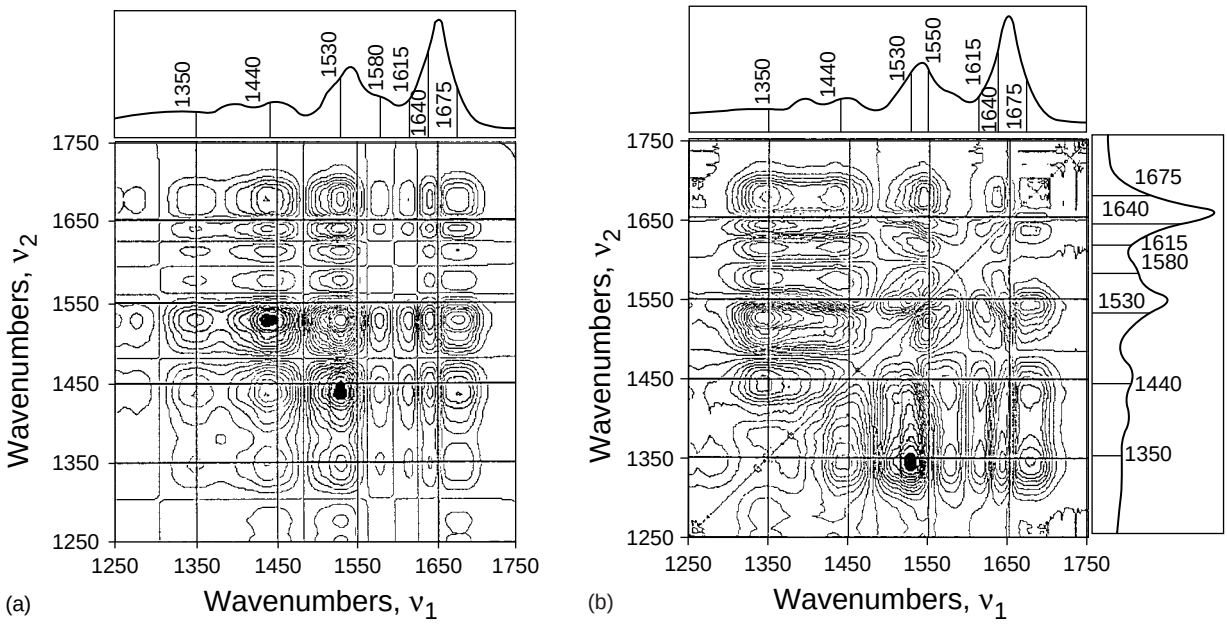

Figure 3 (a) Synchronous and (b) asynchronous 2-D IR correlation spectra of myoglobin generated from the first 10 spectra recorded during the H–D exchange process. Clear and dark peaks relative to the gray background are positive and negative, respectively. The 1-D spectra shown in two figures are the average of the 10 spectra. (Reproduced with permission from Nabet and Pézolet.[12])

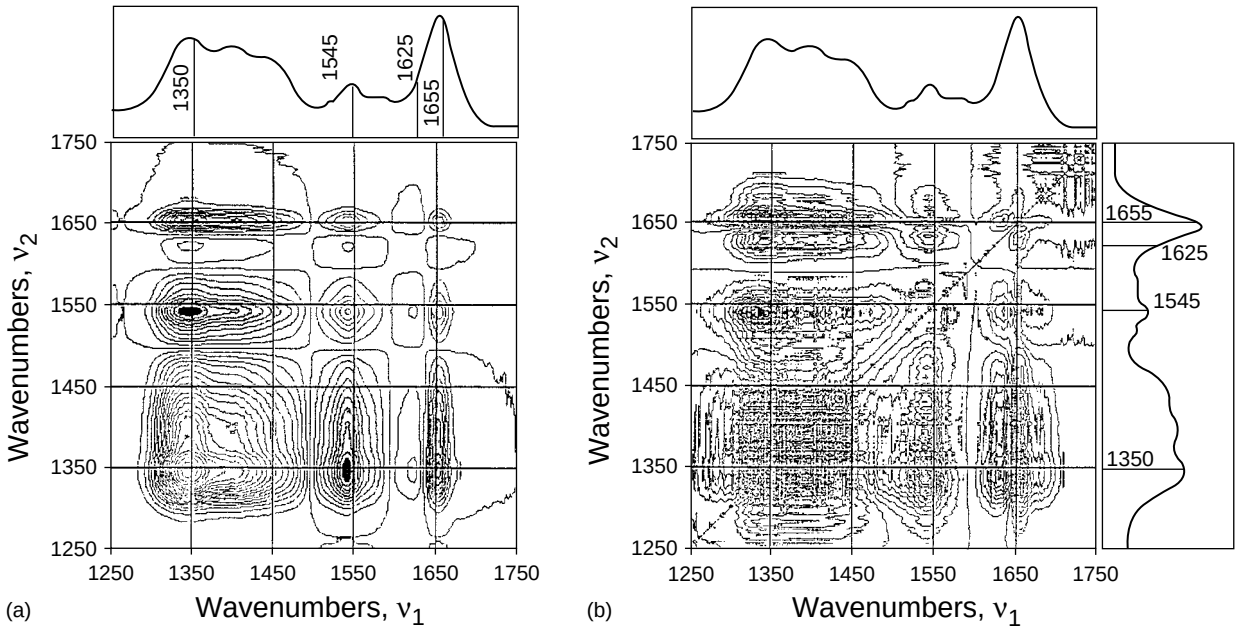

Figure 4 (a) Synchronous and (b) asynchronous 2-D IR correlation spectra of myoglobin constructed from 10 spectra obtained approximately 1 h after the beginning of the H–D exchange process. (Reproduced with permission from Nabet and Pézolet.[12])

peaks, observed at 1545 and 1345 cm^{-1}, may be assigned to the amide II and amide II′ modes of the α-helices of myoglobin, respectively.

The synchronous spectrum for the slowly exchanging system (Figure 4a) also has a weak component at 1625 cm^{-1}. This component could be due to the β-sheet conformation. Since the random coil and the β-turn structures do not develop the amide I components in the synchronous map calculated for the longer time domain, the H–D exchange rate for the β-sheet structure seems

to be slower than those for the random coil and β-turn structures.

3.1.2 *Two-dimensional Near-infrared Spectroscopy Study of Heat Denaturation of Ovalbumin in Aqueous Solutions*

Wang et al.[14] studied heat denaturation of ovalbumin in aqueous solutions using 2-D NIR correlation spectroscopy. NIR spectroscopy is useful in studying hydration and hydrogen bonds in proteins. In NIR spectra of aqueous protein solutions, a number of bands due to overtones and combination modes arising from water and amide groups of proteins can be observed. They are more sensitive to changes in the hydration and hydrogen bonds than IR bands ascribed to water and the amide groups. NIR spectroscopy has one serious problem, in that a number of bands overlap each other to make spectral analysis more difficult.

The purpose of the study by Wang et al.[14] was to investigate the potential of generalized 2-D NIR correlation spectroscopy in exploring subtle differences in NIR spectral responses concerned with structural changes in proteins. Temperature-dependent NIR spectra were measured for ovalbumin solutions over a temperature range from 45 to 80 °C, which covers the whole heat denaturation process. A series of dynamic NIR spectra modulated by the concentration at representative temperatures, such as 45, 76, and 80 °C, were used to construct 2-D synchronous and asynchronous correlation spectra.

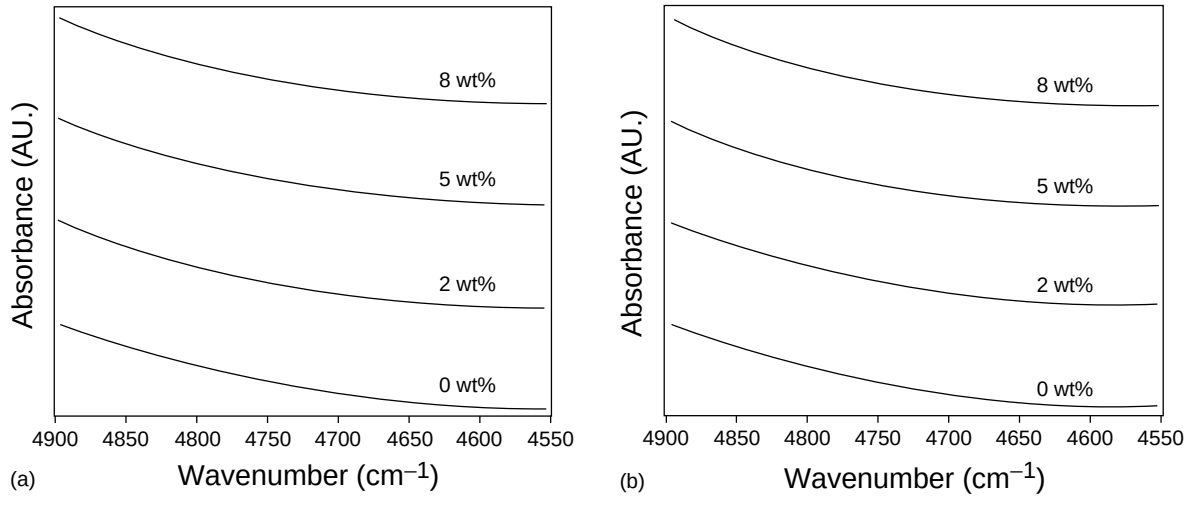

Figure 5 NIR spectra in the 4900–4550 cm^{-1} region for ovalbumin solutions with concentrations of 0, 2, 5, and 8 wt%: (a) at 45 °C; (b) at 80 °C. (Reprinted with permission from Wang et al.[14] Copyright 1998 American Chemical Society.)

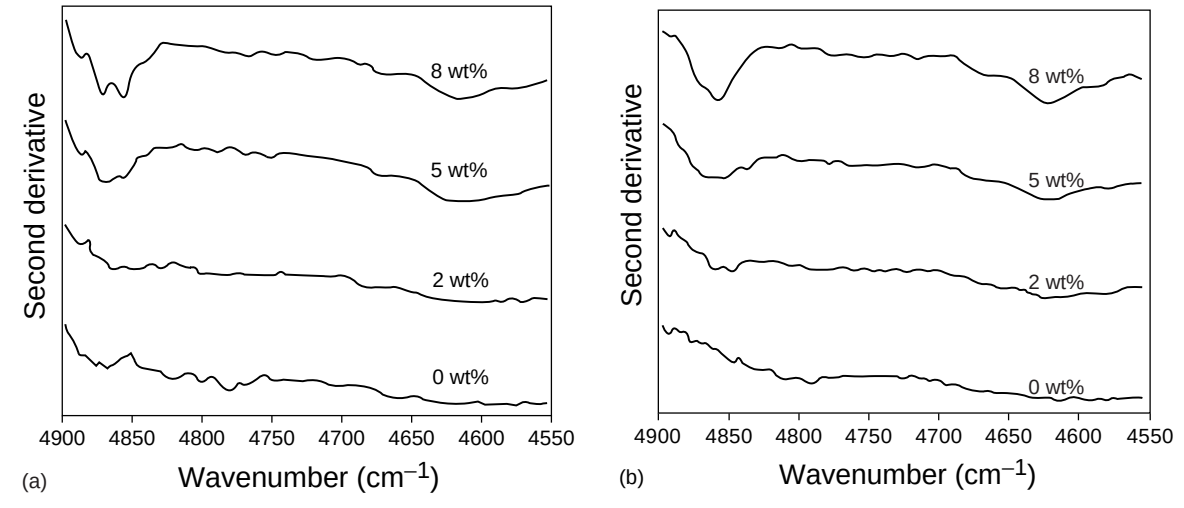

Figure 6 Second derivative NIR spectra in the 4900–4550 cm^{-1} region for ovalbumin solutions with concentrations of 0, 2, 5, and 8 wt%: (a) at 45 °C; (b) at 80 °C. (Reprinted with permission from Wang et al.[14] Copyright 1998 American Chemical Society.)

Figure 5(a) and (b) shows NIR spectra in the 4900–4550 cm^{-1} region for ovalbumin solutions with concentrations of 0, 2, 5, and 8 wt% measured at 45 and 80 °C, respectively, where the protein is in the natively folded and partially unfolded states.[14] It is not possible to identify any band in the spectra except for a tail of the strong absorption band of water near 5200 cm^{-1}. The corresponding second derivatives of the spectra in Figure 5(a) and (b) are presented in Figure 6(a) and (b), respectively.[14] Note that the calculation of the second derivative makes some bands visible. Two bands at 4850 and 4600 cm^{-1} in Figure 6(a) and (b) are assigned to combinations of free NH stretching of amide II (amide A/II) and intramolecular hydrogen-bonded NH stretching of amide II (amide B/II) of ovalbumin, respectively.

Figure 7 shows synchronous and asynchronous 2-D correlation maps constructed from the concentration-perturbed NIR spectra of ovalbumin solutions at 45 and 80 °C.[14] The synchronous spectra (Figure 7a and c) exhibit two strong autopeaks at 4850 and 4800 cm^{-1} at 45 °C. The two peaks probably correspond to the combination band of the amide A/II and to the second overtone of OH bending ($3\nu_2$) of water, respectively. The appearance of the autopeaks means that the intensities of these two bands vary most significantly as the concentration in the folded state increases. In addition, positive cross-peaks (4850 vs 4800 cm^{-1}) are seen in the synchronous spectra, indicating that their band intensities increase in phase (simultaneously).

Asymmetric cross-peaks at (4850, 4600) cm^{-1} and (4800, 4600) cm^{-1} dominate the asynchronous map in

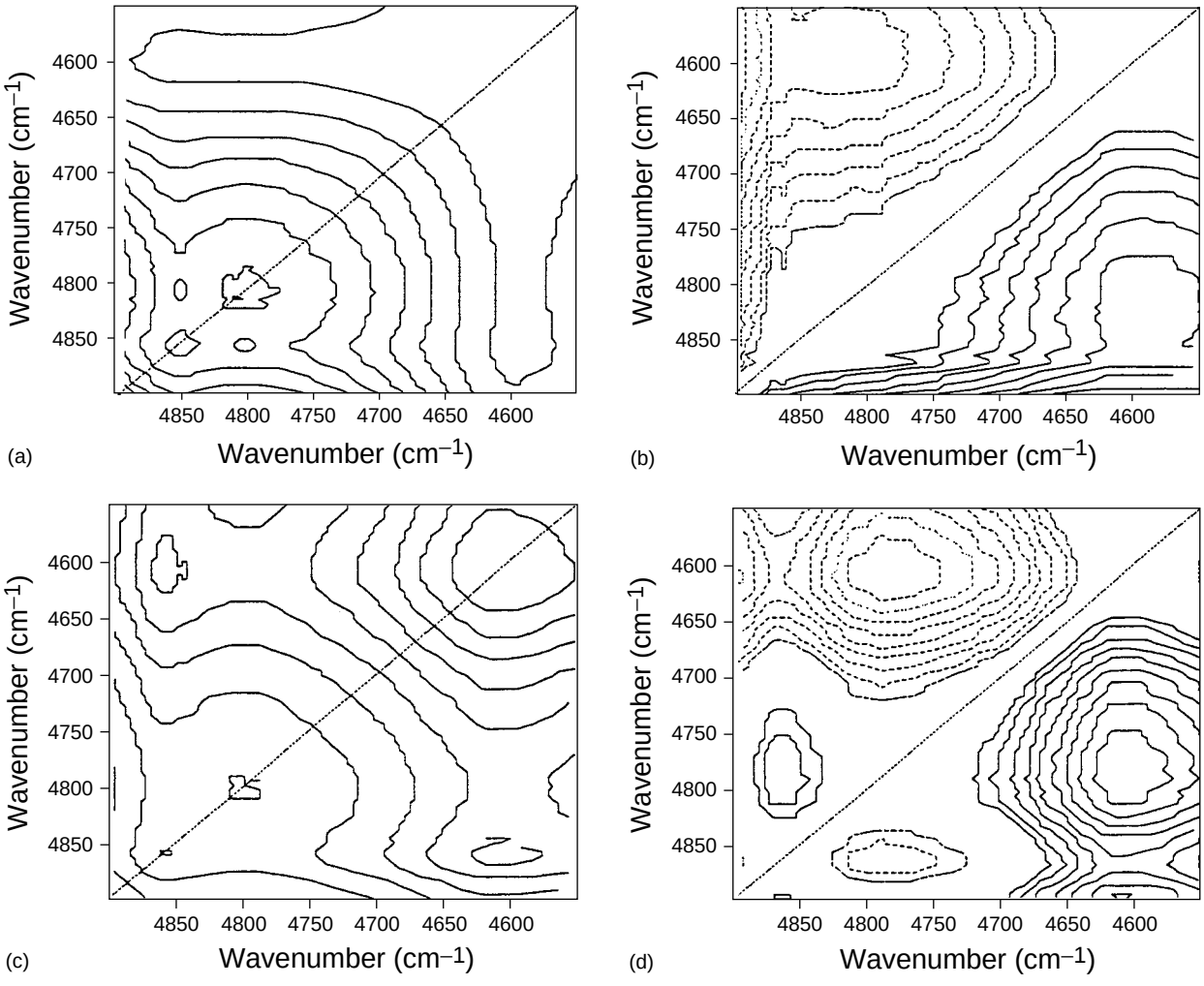

Figure 7 2-D NIR correlation spectra in the 4900–4550 cm^{-1} region, constructed from concentration-perturbed spectra of ovalbumin solutions at 45 °C: (a) synchronous; (b) asynchronous contour maps. 2-D NIR correlation spectra in the 4900–4550 cm^{-1} region, constructed from concentration-perturbed spectra of ovalbumin solutions at 80 °C: (c) synchronous, (d) asynchronous contour maps. Solid and dashed lines in the contour maps denote positive and negative correlation peaks, respectively. (Reprinted with permission from Wang et al.[14] Copyright 1998 American Chemical Society.)

Figure 7(b). The asynchronous spectrum reveals that the amide B/II at 4600 cm^{-1} shows an out-of-phase variation with the amide A/II and $3\nu_2$ bands. According to the rule proposed by Noda[2] for determining the sequential relationships between different bands, the negative peaks at (4850–4600) cm^{-1} and (4800, 4600) cm^{-1} in Figure 7(b) indicate that the intensity of the amide B/II band changes at a lower concentration than do those of the amide A/II and $3\nu_2$ bands.

A strong autopeak at 4610 cm^{-1} in the synchronous spectrum at 80 °C (Figure 7c) means that the intensity of the amide B/II band varies most significantly in the spectra of the unfolded state. The two autopeaks at 4850 and 4800 cm^{-1} in the synchronous spectrum at 45 °C (Figure 7a) become weak in the corresponding spectrum at 80 °C (Figure 7c). Instead, new positive cross-peaks appear at (4860, 4610) cm^{-1} in the synchronous map at 80 °C, indicating that the intensities of the amide A/II and the amide B/II bands vary in phase. The simultaneous intensity change between the amide A/II and the $3\nu_2$ in the folded state vanishes completely. Moreover, the $3\nu_2$ band changes out of phase with the amide A/II and amide B/II, as verified by the corresponding asynchronous spectrum (Figure 7d). The asynchronous spectrum of the denatured unfolded state at 80 °C (Figure 7d) develops two pairs of new cross-peaks at (4860, 4780) cm^{-1} and (4780, 4610) cm^{-1}. These peaks confirm the out-of-phase intensity variation of the $3\nu_2$ band with amide A/II and amide B/II.

Another important finding in the 2-D spectra is that the amide A/II and amide B/II bands shift upward by about 10 cm^{-1} while the $3\nu_2$ band shifts downward by some 20 cm^{-1} as the temperature increases from 45 to 80 °C. Wang et al.[14] ascribed these shifts to the thermally-induced unfolding of the secondary structure of ovalbumin.

A series of the 2-D correlation maps were calculated over a temperature range from 45 to 80 °C at intervals of 2 °C to investigate how the correlation plots vary with temperature during the heat denaturation process.[14] The 2-D correlation patterns below 65 °C and above 71 °C are very similar to those obtained at 45 and 80 °C, respectively. They do change markedly in a very narrow temperature range between 67 and 69 °C. It was concluded that during the hydration process ovalbumin is almost unchanged below 65 °C, then undergoes a sudden change from the native to the denatured state in the 67–69 °C region, and again changes little above 71 °C. The unfolding of secondary structures starts at about 69 °C and continues progressively until 80 °C.[14] Therefore, the change due to hydration takes place earlier than the unfolding process. It is likely that the hydration change in the critical temperature region between 67 and 69 °C brings about the unfolding process at 69–80 °C.

The 2-D NIR correlation spectroscopy study of ovalbumin also provided new insight into thermally-induced conformation changes in the secondary structure and changes in the hydrogen bonds of the amide groups.[14]

3.1.3 Conformational Changes of Ribonuclease A upon Thermal Unfolding Studied by Two-dimensional Infrared, Near-infrared, and Infrared/Near-infrared Correlation Spectroscopy

Schultz et al.[15] reported 2-D IR, 2-D NIR, and 2-D IR/NIR correlation spectroscopy studies on conformational changes of RNase A upon thermal unfolding. The IR and NIR spectra of RNase A were measured in a sodium cacodylate buffer made up in either H_2O or D_2O over a temperature range from 20 to 80 °C. Generalized 2-D correlation analysis was then applied to the spectra. One of the purposes of their study was to demonstrate the potential of NIR spectroscopy in monitoring structural changes of proteins in aqueous solutions. To establish correlations between NIR and IR bands, 2-D IR/NIR heterospectral correlation analysis was carried out. These 2-D correlation studies showed that the terminal unfolding of RNase A is not a completely cooperative process; rather it begins with some changes in the β-sheet structure, followed by the loss of α-helical structures, and then ends with the unfolding of the remaining β-sheets.[15]

Figure 8(a) and (b) shows original IR and NIR spectra of RNase A in H_2O buffer at 20 °C.[15] Spectra of the buffer solution (100 mmol sodium cacodylate, pH 2.1) in the corresponding regions are also shown. Figure 8(c) presents IR and NIR spectra of RNase A after subtraction of the buffer.[15] Subtraction of the buffer spectrum from that of the dissolved protein reveals a number of characteristic protein bands. The IR region is characterized by amide I (1643 cm^{-1}) and II (1546 cm^{-1}) bands while the NIR region is dominated by overtones and combination bands involving C–H, N–H, and O–H stretching vibrations. Schultz et al.[15] proposed the following band assignments based upon previous NIR studies of model compounds: 6616 and 6360 cm^{-1}, the first NH/OH stretching overtones; 5910 and 5767 cm^{-1}, the first overtones of CH_n stretching; 5060 cm^{-1}, a combination mode involving OH groups; 4866 and 4600 cm^{-1}, combination modes of NH groups; 4368 and 4050 cm^{-1}, combinations involving CH_n stretching modes.

Figure 9 illustrates IR spectra in the amide I and amide II regions and NIR spectra in the combination band region of RNase A in H_2O buffer measured at 20, 40, 50, and 80 °C.[15] The IR spectrum at 20 °C shows a strong amide I band at 1641 cm^{-1} and a shoulder at 1688 cm^{-1}. These bands indicate the presence of antiparallel β-strands in the native protein. There are major spectral changes between 35 and 50 °C, showing marked structural

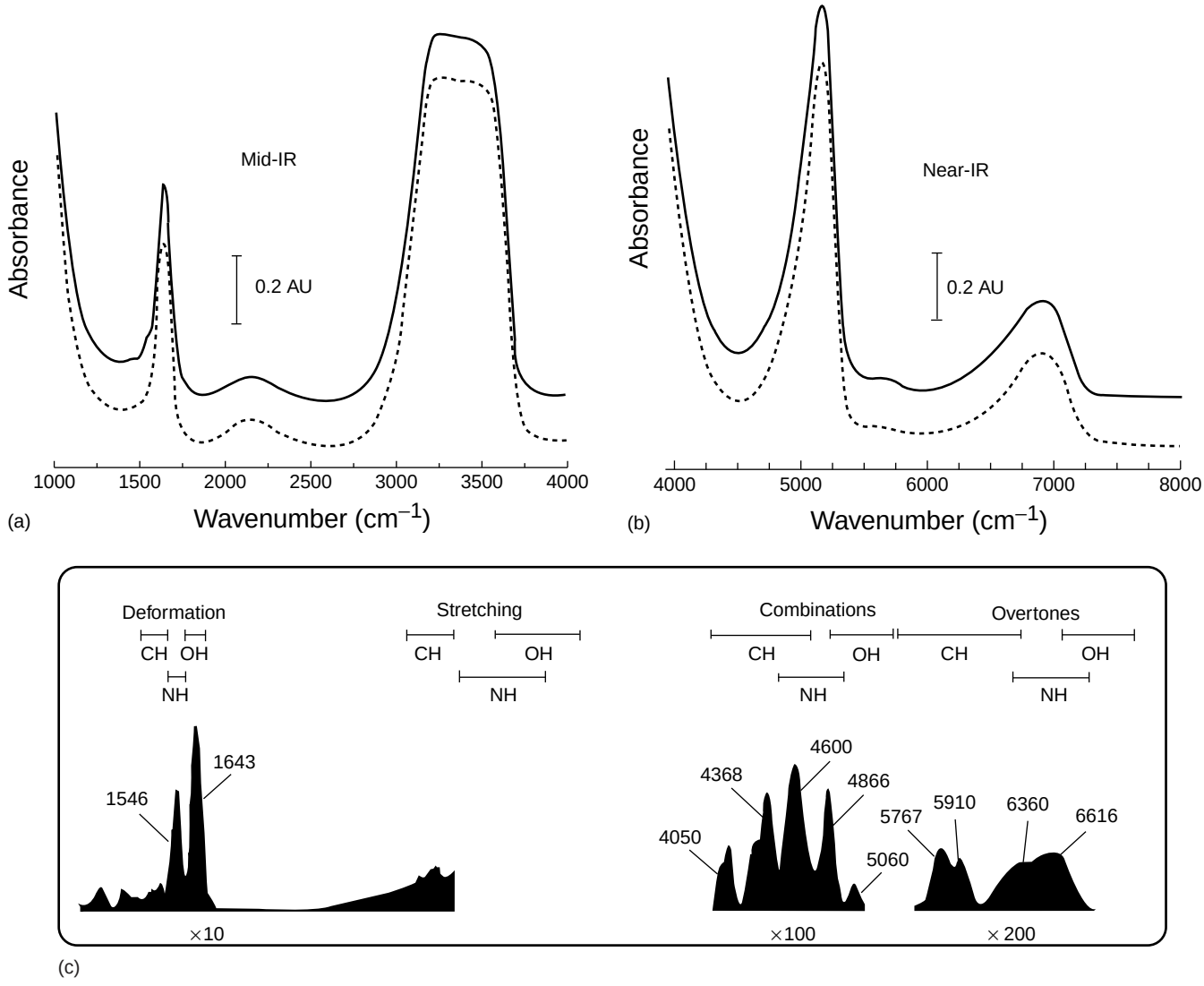

Figure 8 Original (a) IR and (b) NIR spectra of RNase A in H_2O buffer at 20 °C (solid lines). Spectra of buffer (100 mmol sodium cacodylate, pH 2.1) in the corresponding regions are shown by dotted lines. The IR spectra were measured in 6-μm cells, while 200-μm cells were used for measuring the NIR spectra. (c) IR and NIR spectra of RNase A after subtraction of the buffer. The absorbance scale for the difference spectra was expanded by a factor of 10, 100, or 200 compared with the scale of the original spectra. (C.P. Schultz, H. Fabian, H.H. Mantsch, *Biospectroscopy*, Copyright © 1998 Wiley-Liss, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

changes during unfolding of the protein. Figure 9(a) shows temperature-dependent intensity changes of the β-structure "marker" band at 1641 cm^{-1}. The intensity of the band changes mainly between 35 and 50 °C. The spectra measured above 70 °C show a broad, featureless amide I band contour. This contour is typical of a predominantly irregular protein structure.

The most pronounced spectral changes in the NIR spectra are observed in the 4900–4860 cm^{-1} region where a strong combination mode of amide A and amide II due to peptide groups is expected to appear.[15] This band is located at 4867 cm^{-1} at 20 °C. As the temperature is increased, this band loses intensity, becomes broader, and shifts to higher wavenumbers. The plot of its frequency versus temperature (Figure 9b) reveals that the frequency shift of the combination band has the same temperature dependence as the intensity change in the amide I band.

Figure 10 illustrates (a) synchronous and (b) asynchronous 2-D correlation spectra, generated from IR spectra of RNase A in H_2O buffer measured at various temperatures (20–70 °C).[15] For the 2-D correlation analyses, the IR and NIR spectra were transformed into Fourier self-deconvoluted spectra. The synchronous spectrum shows autopeaks in the amide I and II regions

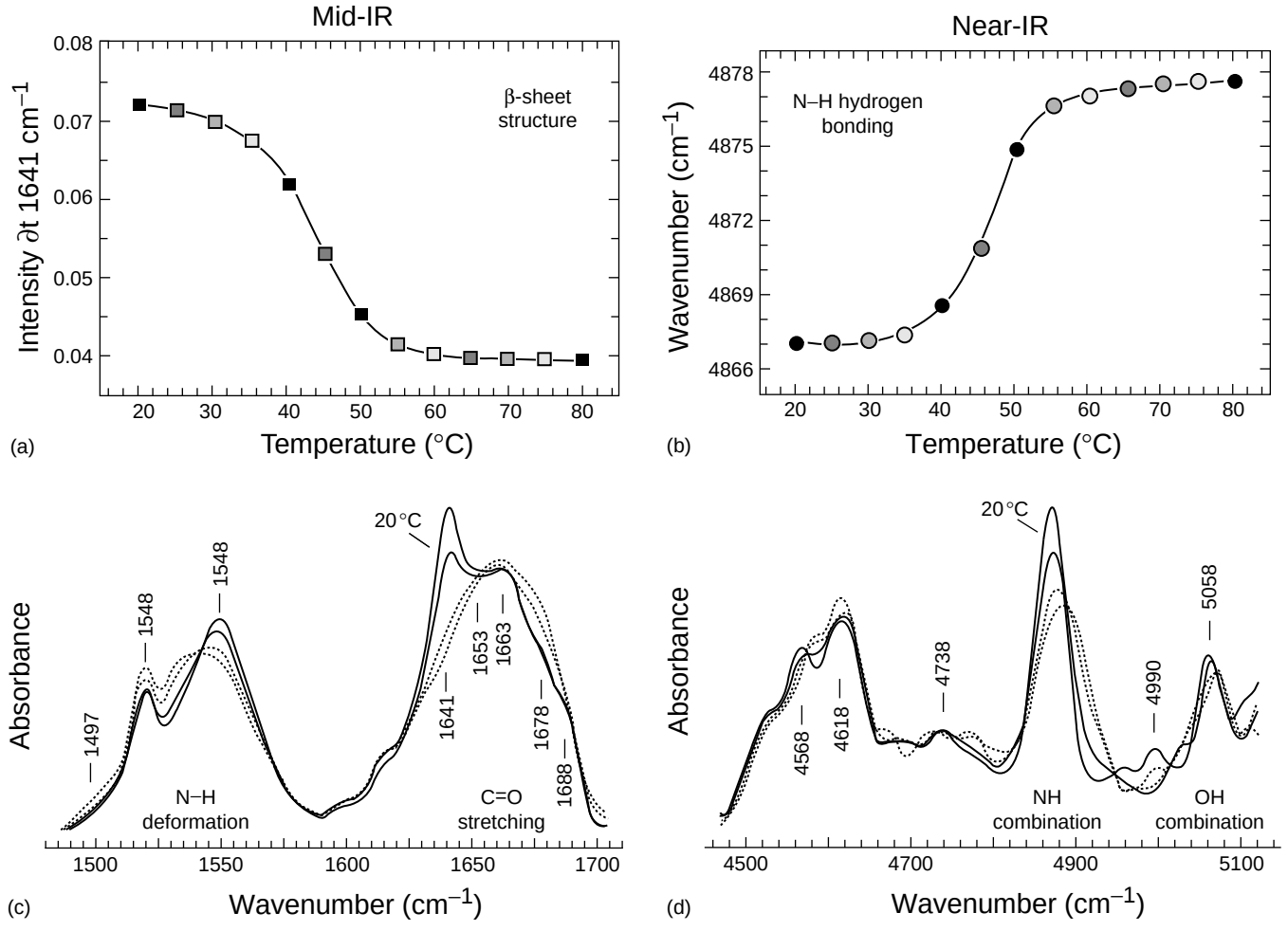

Figure 9 (a) Temperature dependence of the peak height intensity of the amide I band at 1641 cm^{-1} and (b) temperature dependence of the NH combination band frequency. (c) IR and (d) NIR spectra of RNase A in H_2O buffer measured at 20, 40, 50, and 80 °C. (C.P. Schultz, H. Fabian, H.H. Mantsch, *Biospectroscopy*, Copyright © 1998 Wiley-Liss, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

at 1688, 1661, 1652, 1641, 1547, and 1517 cm^{-1}. These autopeaks identify changes that occur in the secondary structure of the protein marked α for α-helix and β for β-sheet.

The asynchronous 2-D correlation spectrum enables differences in the stability of different secondary structure elements to be detected. It is not easy to detect such differences by conventional analysis of changes in the 1-D IR spectra of proteins. A close inspection of the asynchronous spectrum of RNase A in Figure 10(b) reveals that the unfolding of the protein does not occur in a single step.[15] Asynchronous cross-peaks at (1637, 1644) cm^{-1} suggest that there are two kinds of β-sheet structures in RNase A characterized by two low-frequency spectral features whose response to a change in temperature is slightly different. It was found that the component at 1637 cm^{-1} responds earlier during the protein unfolding process than the one at 1644 cm^{-1}. Probably, the slight difference in the frequency is ascribed to slightly different strengths of hydrogen bonding of the C=O groups.

The three α-helices of RNase A develop only very weak asynchronous cross-peaks (not visible in Figure 10b), indicating that the α-helices undergo an almost simultaneous unfolding process. One of the bands due to the α-helices (1656 cm^{-1}) shows a correlation cross-peak to the two bands of the β-sheet in the asynchronous spectrum, suggesting that the unfolding process of RNase A follows a sequence of three spectroscopic events involving the following bands: 1637/1692 cm^{-1} (step 1) $\rightarrow$ 1656 cm^{-1} (step 2) $\rightarrow$ 1644 cm^{-1} (step 3). This sequence means that the first step in the unfolding process is the partial unfolding of the stronger parts of the β-sheet, followed by the unfolding of the α-helices and then by the unfolding of the weaker parts of the β-sheet.

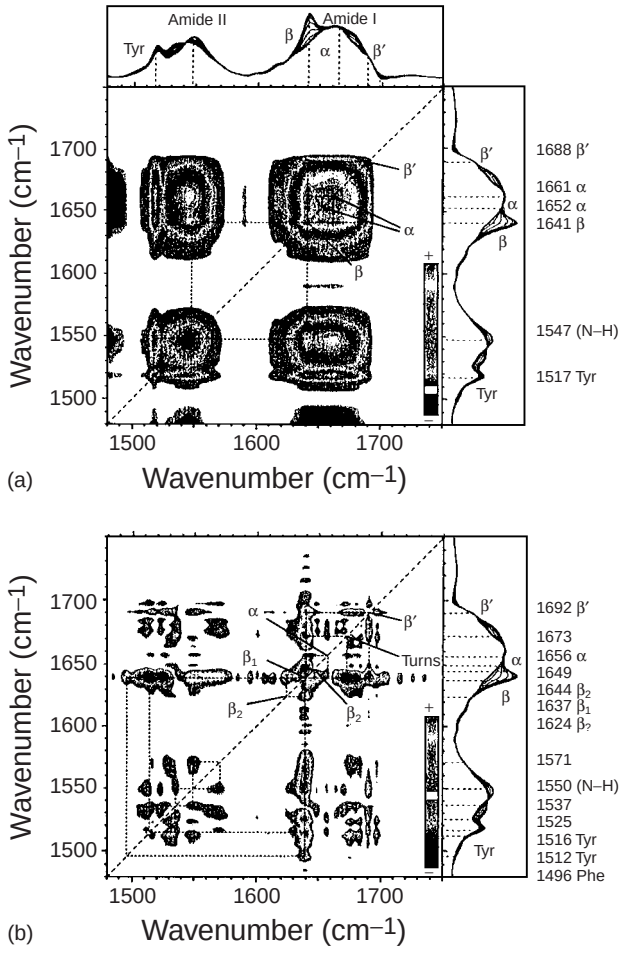

Figure 10 (a) Synchronous and (b) asynchronous 2-D IR correlation spectra of RNase A in H_2O buffer, pH 2.1, at 11 discrete temperatures between 20 and 70 °C. The correlation spectra are represented as intensity maps with a set of 1-D Fourier self-deconvoluted spectra shown outside the maps. (C.P. Schultz, H. Fabian, H.H. Mantsch, *Biospectroscopy*, Copyright © 1998 Wiley-Liss, Inc. Reprinted by permission of Wiley-VCH, Inc., a division of John Wiley & Sons, Inc.)

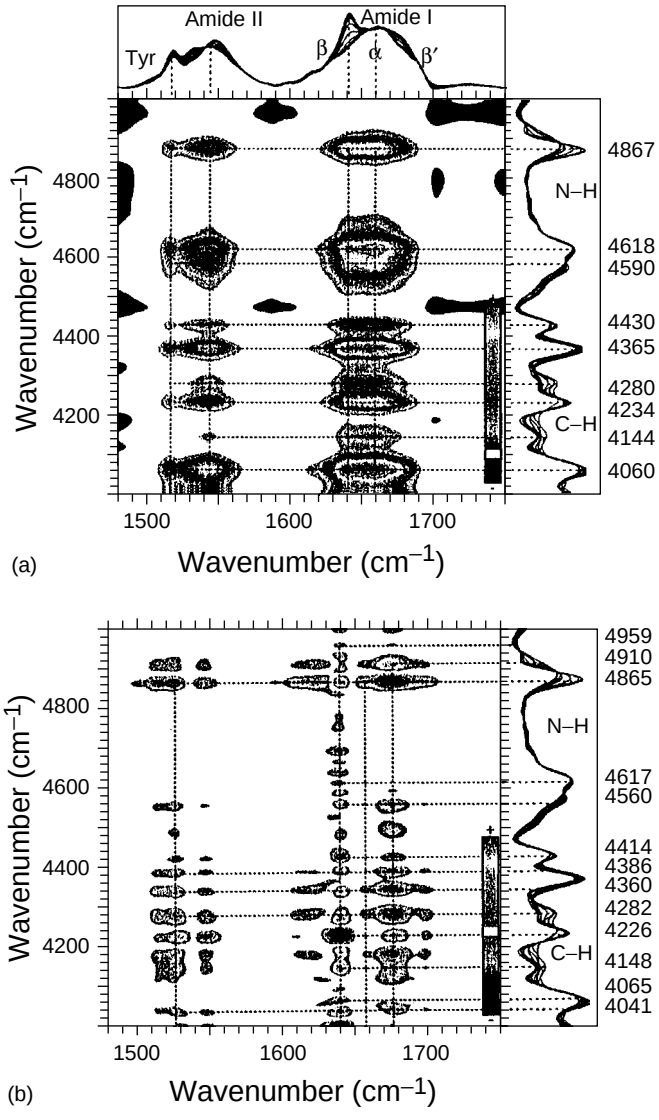

Figure 11 (a) Synchronous and (b) asynchronous 2-D IR/NIR correlation spectra of RNase A in H_2O buffer, pH 2.1, at 11 discrete temperatures between 20 and 70 °C. (C.P. Schultz, H. Fabian, H.H. Mantsch, *Biospectroscopy*, Copyright © 1998 Wiley-Liss, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

Figure 11 shows (a) synchronous and (b) asynchronous hetero 2-D IR/NIR spectra generated from the IR and NIR spectra of RNase A measured at 11 temperatures.[15] The 2-D correlation between the two spectral regions allows vibrations of a similar nature to be detected and band assignments in the NIR region to be made based upon established assignments in the IR region. In Figure 11 the positive cross-peaks between the bands in the amide I region and the N−H combination (amide A + amide II) band at 4867 cm^{-1} are notable. They indicate that the amide C=O groups (IR) are highly correlated with the amide N−H groups (NIR). An interesting point in the asynchronous spectrum is that five IR bands at 1675, 1641, 1548, 1525, and 1515 cm^{-1} have an out-of-phase correlation with most NIR bands. For example, the band at 1640 cm^{-1} due to the amide I mode of the β-sheet with slightly weaker hydrogen bonds and the 4865 or 4617 cm^{-1} bands share positive or negative asynchronous cross-peaks. These observations led Schultz et al.[15] to the conclusion that some structure elements in RNase A respond earlier and others respond later to an increase in temperature than the β-sheet structure. The overall conclusion reached by them was that the 2-D correlation analysis of IR and NIR spectra of proteins shows considerable promise of gaining more detailed structural insights into the secondary structure changes.

3.2 Other Biomedical Applications

3.2.1 Human Skin Stratum Corneum

The 2-D IR analysis of protein was first attempted for the dynamic rheo-optical study of proteinaceous components of a human skin stratum corneum specimen.[1] The small sample of human skin was mounted over an IR-transparent thin but robust fluorocarbon (Teflon™) substrate film. This composite sample was then acoustically stimulated with a small-amplitude strain at a frequency of 80 Hz to induce time-dependent reorientation of various submolecular constituents. The resulting dynamic IR dichroism spectra of the sample were then analyzed by the 2-D correlation method. Figure 12 show the amide I and II regions of the 2-D IR spectra for the proteinaceous component of stratum corneum.

The synchronous 2-D spectrum of human skin shown in Figure 12(a) clearly indicates the existence of temporal correlations between certain components of IR bands belonging to the amide I and II modes of vibrations (e.g. 1635 and 1530 cm^{-1} bands). The highly overlapped composite nature of the amide peaks is clearly demonstrated by the asynchronous spectrum shown in Figure 12(b), which shows the decoupling of the amide I peak into several independent bands. For example, the asynchronous cross-peaks at 1416 and 1635 cm^{-1} indicate the existence of separate component bands at these wavenumbers. It is interesting to note that while the dipole-transition moments for the different component bands of the amide I peak are reorienting independently of each other, some are coupled with the reorientation of certain amide II component bands.

The synchronously correlated IR band pairs in the human stratum corneum probably originate from common local conformations of polypeptide chains. Thus, it can be speculated that the 3_1 helix component of the amide I band, for example, will couple with the 3_1 helix component of the amide II band. Similar arguments may hold for other conformations. By systematically developing 2-D spectra of model polypeptides with known conformations, it may become possible to identify well resolved *2-D vibrational fingerprints* of protein conformations. 2-D IR spectroscopy alone obviously will not provide a complete answer to the fundamental questions of vibrational spectroscopy for a very complex biological sample. The technique, however, is a powerful analytical tool which should augment other spectroscopic studies to provide submolecular-level understanding of the system.

3.2.2 Human Hair Keratin

Rheo-optical dynamic 2-D IR spectroscopy is used to study the microstructure of human hair keratin films by carrying out 2-D correlation analysis to time-dependent spectral intensity fluctuations under a small-amplitude oscillatory strain.[20] The mechanical properties of human hair are very sensitive to changes in environmental conditions. For example, simple climatic changes in temperature and relative humidity (RH) can have a dramatic effect on hair set. Cosmetic treatments for hair such as shampooing, conditioning, and permanent waving are also known to affect the mechanical properties of hair fibers. However, surprisingly little is known about the fundamental mechanisms by which these physical changes occur.

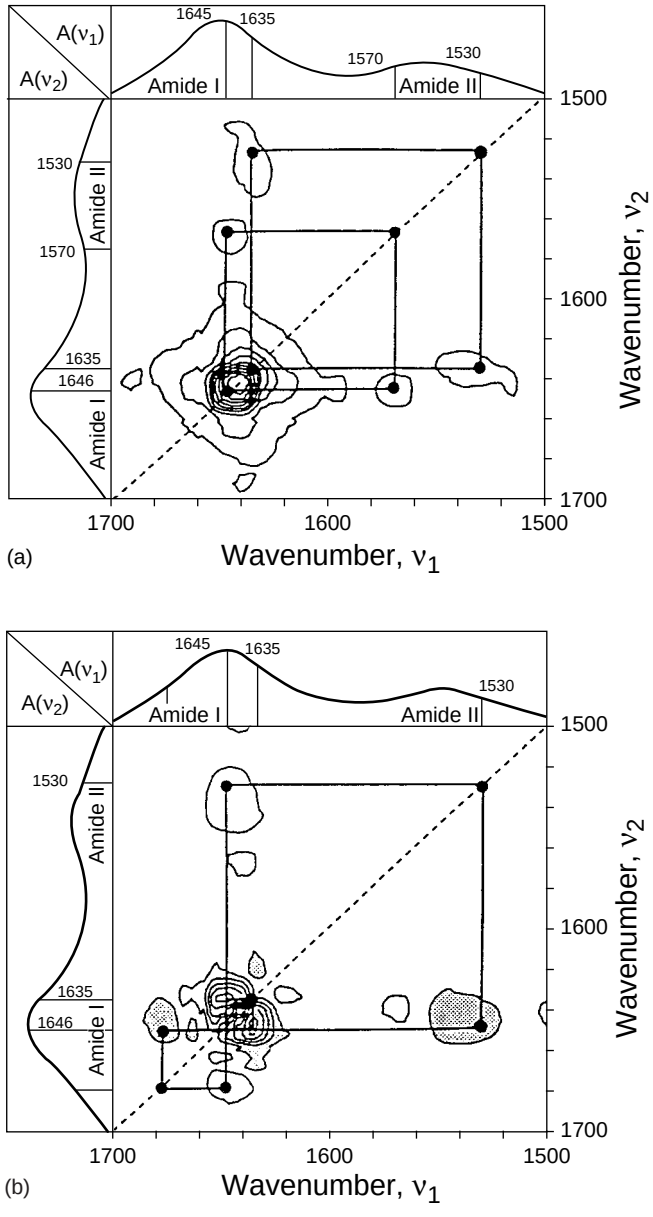

Figure 12 (a) Synchronous and (b) asynchronous 2-D IR spectra of human stratum corneum. (Reproduced by permission from Noda.[1])

2-D IR spectra of keratin films are sensitive to molecular-level changes induced by temperature, RH, and the addition active ingredients in hair treatments. 2-D IR spectra also provide an attractive spectral resolution enhancement of the naturally broad and highly overlapped amide I and II absorbance bands. The net result is an improved molecular-level understanding of the mechanical properties of human hair and identification of factors affecting the properties of these proteins. Thus, 2-D IR spectroscopy is selected to study the molecular-level origins of the properties of human hair under mechanical perturbations and systematic environmental changes.

A model hair specimen suitable for rheo-optical study was prepared as follows. Clean blended European blonde hair (National Hair, New York) was solubilized by first pulverizing it, and then breaking the disulfide bonds using 8 M urea and 50 mM dithiothreitol. After filtering, the keratin sulfhydryls were alkylated to form S-carboxymethyl keratin (SCMK). The SCMK was dialyzed and thin films were cast from a 10 mg mL^{-1} 50/50 water/isopropanol solution onto an IR-transparent polyhalocarbon substrate. Finished samples typically measured 4 × 2 cm with a thickness of 25–100 μm (not including the substrate). A 23-Hz dynamic strain with an amplitude of 50 μm was applied to the sample during the measurement. The jaws holding the sample were located in an environmental control chamber. This chamber allowed temperature control to within 0.5 °C and was modified to allow humidity control to within 2% RH. Keratin films were analyzed at temperatures from 30 to 60 °C and at various levels up to about 90% RH. All IR spectra were collected at a nominal spectral resolution of 8 cm^{-1}.

Figure 13(a) shows the asynchronous 2-D IR correlation spectrum of thin keratin (SCMK) films cast from solubilized hair in the amide I region at ambient temperature and humidity. Cross-peaks represent individual submolecular components reorienting at different rates in response to the externally applied perturbation. A clear resolution enhancement of this naturally broad IR absorbance band is achieved. The IR bands resolved by 2-D correlation analysis can be readily assigned to specific conformations, like α-helices (1661 and 1649 cm^{-1}), β-like extended chains and turns (1679, 1669, 1645, 1641, and 1620 cm^{-1}), as well as disordered structures (1656 cm^{-1}).[20] Obviously, such high-resolution mid-IR spectra of condensed-state specimens are not usually available using conventional (1-D) IR methods.

Figure 13(b) shows the corresponding synchronous 2-D IR spectrum of the same system. The synchronous spectrum in the amide I region is dominated by peak clusters of a distinct four leaf clover like pattern, consisting of two autopeaks with two adjacent negative cross-peaks, which is characteristic of α-helical responses.

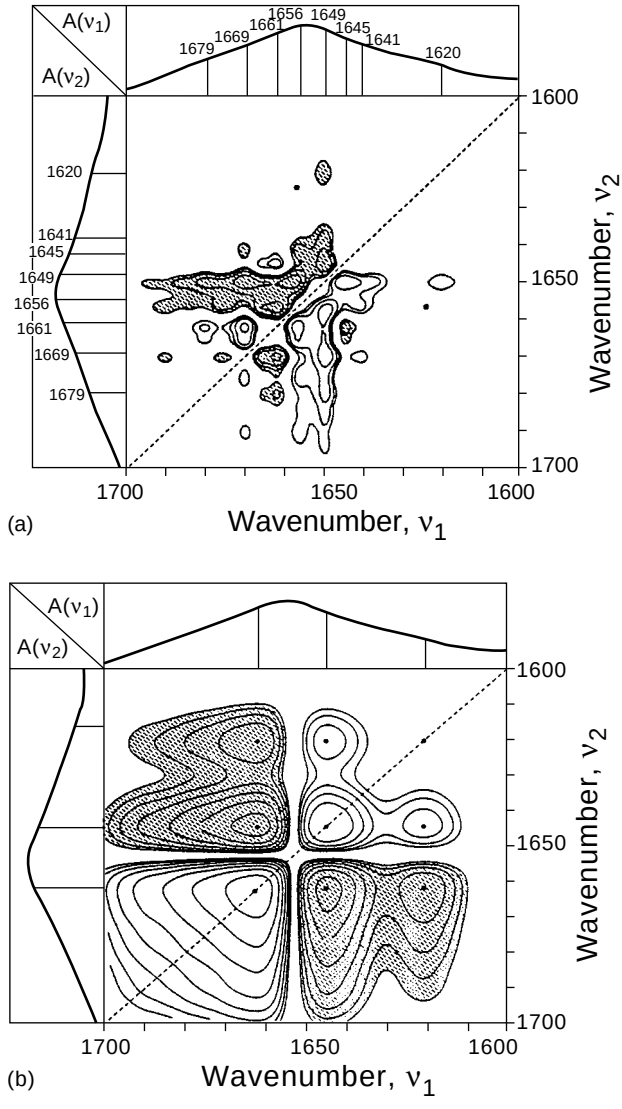

Figure 13 (a) Asynchronous and (b) synchronous 2-D IR spectra of human hair keratin in the amide I region.

This amide I region of the 2-D IR spectrum of the keratin changes markedly as a function of temperature. The changes observed correlate well with differential scanning calorimetry (DSC) data of actual hair, showing an irreversible thermal transition near 45 °C. The α-helical four leaf clover feature near 1650 cm^{-1} in the synchronous spectra remains relatively unchanged as a function of temperature, suggesting that other components such as random or β-like structures are responsible for the differences seen. The effects of high RH and treatment with naphthalene sulfonic acid (NSA) on hair keratin films have been examined by 2-D IR spectroscopy. We found a significant change in dynamic IR signal intensity for bands characteristic of α-helical responses at high RH and after NSA treatment.

3.2.3 Two-dimensional Infrared Correlation Spectroscopy of Synthetic and Biological Apatites

Studies on ultrastructural changes associated with the maturation of poorly crystalline HA in both synthetic and biological systems are very important for biomedical research. FTIR spectroscopy is one of the most powerful techniques for monitoring such changes. However, spectral analysis of the ν_1, ν_3 phosphate region ($1200-900\,cm^{-1}$) is not straightforward because the region is composed of several intrinsically overlapped bands. To overcome this problem, various techniques such as Fourier self-deconvolution, curve-fitting, and second derivative spectroscopy have been used. However, each of them has drawbacks.

Gadaleta et al.[8] demonstrated the utility of 2-D IR correlation spectroscopy for analysis and interpretation of the ν_1, ν_3 phosphate region in synthetic and biological HA. 2-D correlation analysis was used to investigate the maturation of poorly crystalline HA synthesized at both constant and variable pH. It was also used to explore spatial variations in the properties of the mineral found in the transitional region of normally calcified turkey leg tendons, as determined by IR microscopy.[8]

Figure 14 shows typical IR spectra of the ν_1, ν_3 phosphate region of HA synthesized at variable pH.[8] A number of bands overlap each other in this region, so that it is not always easy to identify the band positions. Figure 15(a) depicts the synchronous 2-D IR correlation map generated from a series of spectra monitoring the maturation of poorly crystalline HA synthesized at variable pH.[8] The synchronous map shows seven autopeaks at 1150, 1126, 1095, 1036, 1018, 999, and $961\,cm^{-1}$. The peaks at 1126, 1036, and $1018\,cm^{-1}$ correspond to vibrations of PO_4^{3-} in a nonstoichiometric/acid phosphate environment of poorly crystalline HA, while those at 1095, 999, and $961\,cm^{-1}$ correspond to vibrations of PO_4^{3-} in an apatitic/stoichiometric environment of poorly crystalline HA. Virtually all peaks due to stoichiometric HA are positively correlated with each other, as are peaks due to nonstoichiometric HA/acid phosphate. It should also be noted that the former peaks are all negatively correlated with each of the latter peaks. It was therefore concluded that the intensities of the peaks associated with stoichiometric HA simultaneously increase as peaks associated with nonstoichiometric/acid phosphate decrease.

Figure 15(b) illustrates the corresponding asynchronous plot.[8] In this study, interest in the asynchronous spectrum is due to the possibility of peak determination partly because the position of the bands can be used to characterize the environment of the phosphate ion and partly because the peak positions can be employed as input parameters for curve-fitting algorithms. In the asynchronous spectrum new bands are observed at 1075, 1045, and $1032\,cm^{-1}$.

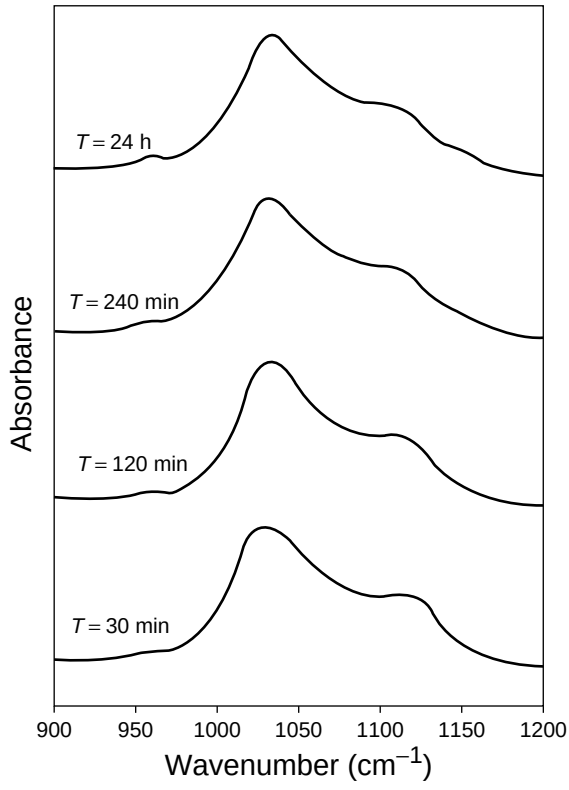

Figure 14 IR spectra of the ν_1, ν_3 phosphate region of HA synthesized at variable pH. Times indicated are the times when the aliquots were withdrawn. (S.J. Gadaleta, A. Gericke, A.L. Boskey, R. Mendelsohn, *Biospectroscopy*, **2**(3), 353–359, Copyright 1996 John Wiley & Sons, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

The 2-D correlation study suggested new assignments for a band at $1145\,cm^{-1}$.[8] This band, which had been previously attributed to HPO_4^{2-}-containing apatites, has now been assigned to PO_4^{3-} vibrations in an apatitic/stoichiometric environment of poorly crystalline HA.

Figure 16(a) and (b) presents the synchronous and asynchronous 2-D IR correlation spectra, respectively, for a series of IR microscopy spectra obtained from the transitional region of calcified turkey leg tendon.[8] Five autopeaks are observed at 1147, 1126, 1075, 1039, and $1019\,cm^{-1}$. The peaks at 1126, 1039, and $1019\,cm^{-1}$ correspond to vibrational modes of PO_4^{3-} in a nonstoichiometric/acid phosphate environment of poorly crystalline HA, whereas the peak at $1075\,cm^{-1}$ represents a mode due to PO_4^{3-} in an apatitic/stoichiometric environment of poorly crystalline HA. The mineral from the turkey tendon shows a 2-D pattern very similar to the synthetic HA prepared at variable pH. Gadaleta et al.[8] deduced from this comparison of the 2-D IR spectra of the synthetic and biological apatites that the

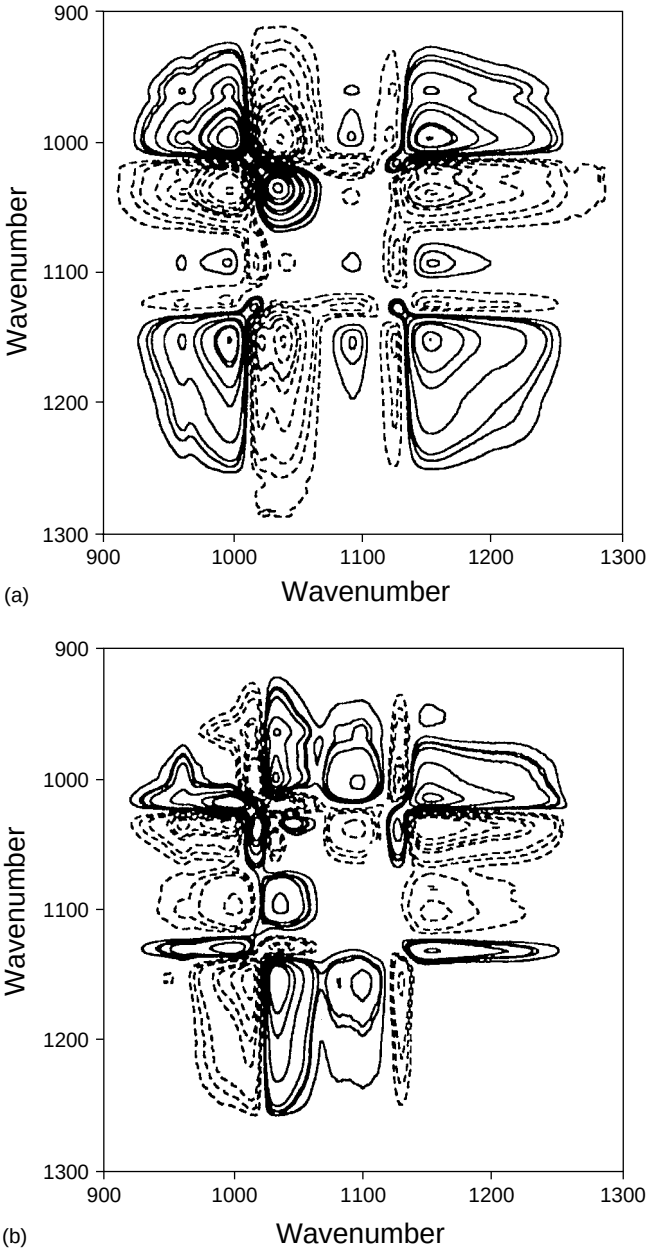

Figure 15 (a) Synchronous and (b) asynchronous 2-D IR correlation spectra of synthetic mineral prepared at variable pH. Solid lines represent positive peaks while dashed lines represent negative bands. (S.J. Gadaleta, A. Gericke, A.L. Boskey, R. Mendelsohn, *Biospectroscopy*, **2**(3), 353–359, Copyright 1996 John Wiley & Sons, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

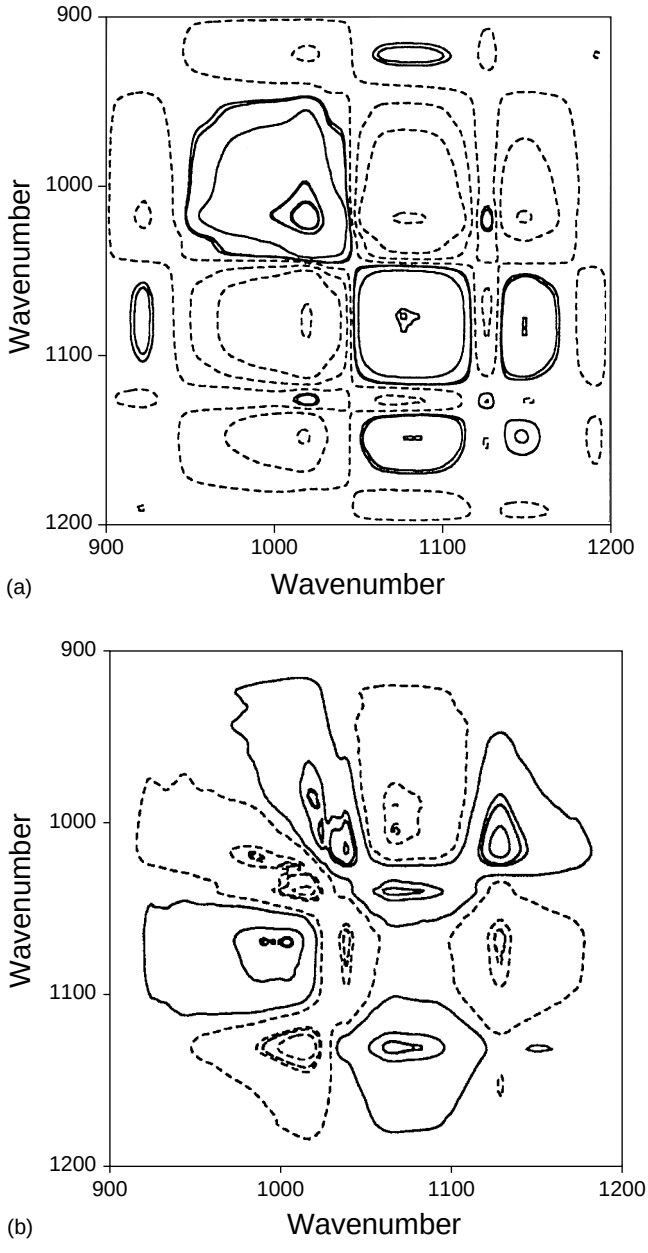

Figure 16 (a) Synchronous and (b) asynchronous 2-D IR correlation spectra of the mineral found in the transitional region of normally calcified turkey leg tendon. (S.J. Gadaleta, A. Gericke, A.L. Boskey, R. Mendelsohn, *Biospectroscopy*, **2**(3), 353–359, Copyright 1996 John Wiley & Sons, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

spatial variation of HA maturation of mineral along the tendon matches the temporal variation of HA formed at variable pH.

This study provided new insight into the subtle changes accompanying the maturation of synthetic and biological apatites, as investigated with IR spectroscopy.

As demonstrated in these examples, 2-D correlation vibrational spectroscopy has broad applicability in biomedical areas. There may be two directions for future applications. One direction is concerned with applications to basic science, for example 2-D correlation spectroscopy studies of biological macromolecules such as proteins, enzymes, and nucleic acids. Another direction aims at

more practical applications, including those to clinical medicine, food science, and pharmaceutical science. Even medical diagnosis by 2-D correlation spectroscopy might be possible. For example, it is possible to investigate the effects of drugs on cancer cells by drug concentration-dependent 2-D IR spectra of the cells.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflection
DSC	Differential Scanning Calorimetry
FTIR	Fourier Transform Infrared
HA	Hydroxyapatite
H–D	Hydrogen–Deuterium
IR	Infrared
IR/NIR	Infrared/Near-infrared
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
NSA	Naphthalene Sulfonic Acid
RH	Relative Humidity
RNase A	Ribonuclease A
SCMK	S-Carboxymethyl Keratin
UV	Ultraviolet
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Infrared Spectroscopy of Biological Applications

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Fourier Transform Raman Instrumentation

REFERENCES

1. I. Noda, 'Two-dimensional Infrared (2-D IR) Spectroscopy: Theory and Applications', *Appl. Spectrosc.*, **44**(4), 550–561 (1990).
2. I. Noda, 'Generalized Two-dimensional Correlation Method Applicable to Infrared, Raman, and Other Types of Spectroscopy', *Appl. Spectrosc.*, **47**(9), 1329–1336 (1993).
3. Y. Ozaki, I. Noda, 'Potential of Generalized Two-dimensional Correlation Spectroscopy in the Near Infrared Region', *J. NIR Spectrosc.*, **4**(1), 85–99 (1996).
4. Y. Ozaki, Y. Wang, 'Two-dimensional Near Infrared Correlation Spectroscopy: Principle and its Application', *J. NIR Spectrosc.*, **6**(1), 19–31 (1998).
5. I. Noda, Y. Liu, Y. Ozaki, M.A. Czarnecki, 'Two-dimensional Fourier-transform Near-infrared Correlation Spectroscopy Studies of Temperature-dependent Spectral Variations of Oleyl Alcohol', *J. Phys. Chem.*, **99**(10), 3068–3073 (1995).
6. Y. Liu, Y. Ozaki, I. Noda, 'Two-dimensional Fourier-transform Near-infrared Correlation Spectroscopy Study of Dissociation of Hydrogen-bonded N-Methylacetamide in the Pure Liquid State', *J. Phys. Chem.*, **100**(18), 7326–7332 (1996).
7. I. Noda, Y. Liu, Y. Ozaki, 'Two-dimensional Correlation Spectroscopy Study of Temperature-dependent Spectra Variations of *N*-Methylacetamide in the Pure Liquid State. 1. Two-dimensional Infrared Analysis', *J. Phys. Chem.*, **100**(21), 8665–8673 (1996).
8. S.J. Gadaleta, A. Gericke, A.L. Boskey, R. Mendelsohn, 'Two-dimensional Infrared Correlation Spectroscopy of Synthetic and Biologic Apatites', *Biospectrosc.*, **2**(3), 353–359 (1996).
9. M. Muller, R. Buchet, U.P. Fringeli, '2-D-FTIR ATR Spectroscopy of Thermo-induced Periodic Secondary Structural Changes of Poly-(L)-lysine: A Cross-correlation Analysis of Phase-resolved Temperature Modulation Spectra', *J. Phys. Chem.*, **100**(25), 10810–10825 (1996).
10. Y. Ozaki, Y. Liu, I. Noda, 'Two-dimensional Infrared and Near-infrared Correlation Spectroscopy: Applications to Studies of Temperature-dependent Spectral Variations of Self-associated Molecules', *Appl. Spectrosc.*, **51**(4), 526–535 (1997).
11. Y. Ozaki, Y. Liu, I. Noda, 'Two-dimensional Near-infrared Correlation Spectroscopy Study of Premelting Behavior of Nylon 12', *Macromolecules*, **30**(8), 2391–1399 (1997).
12. A. Nabet, M. Pézolet, 'Two-dimensional FTIR Spectroscopy: A Powerful Method to Study the Secondary Structure of Proteins Using H–D Exchange', *Appl. Spectrosc.*, **51**(4), 466–469 (1997).
13. N.L. Sefara, H.H. Richardson, 'Structural Characterization of β-Lactoglobulin in Solution Using Two-dimensional FT Mid-infrared and FT Near-infrared Correlation Spectroscopy', *Appl. Spectrosc.*, **51**(4), 536–540 (1997).

14. Y. Wang, K. Murayama, Y. Myojo, R. Tsenkova, N. Hayashi, Y. Ozaki, 'Two-dimensional Fourier Transform Near-infrared Spectroscopy Study of Heat Denaturation of Ovalbumin in Aqueous Solutions', *J. Phys. Chem. B*, **102**(34), 6655–6662 (1998).
15. C.P. Schultz, H. Fabian, H.H. Mantsch, 'Two-dimensional Mid-IR and Near-IR Correlation Spectra of Ribonuclease A: Using Overtones and Combination Modes to Monitor Changes in Secondary Structure', *Biospectrosc.*, **4**, 19–29 (1998).
16. M.A. Czarnecki, P. Wu, H.W. Siesler, '2-D FT-NIR and FT-IR Correlation Analysis of Temperature-induced Changes of Nylon 12', *Chem. Phys. Lett.*, **283**, 326–332 (1998).
17. M.A. Czarnecki, H. Maeda, Y. Ozaki, M. Suzuki, M. Iwahashi, 'Resolution Enhancement and Band Assignments for the First Overtone of OH Stretching Modes of Butanols by Two-dimensional Near-infrared Correlation Spectroscopy. 2. Thermal Dynamics of Hydrogen Bondings in *n*- and *tert*-Butyl Alcohol in the Pure Liquid State', *J. Phys. Chem. A*, **102**(46), 9117–9123 (1998).
18. I. Noda, G.M. Story, A.E. Dowrey, R.C. Reeder, C. Marcott, 'Applications of Two-dimensional Correlation Spectroscopy in Depth Profiling Photoacoustic Spectroscopy, Near-infrared Dynamic Rheooptics, and Spectroscopic Imaging Microscopy', *Macromol. Chem., Macromol. Sym.*, **119**, 1–13 (1997).
19. I. Noda, A.E. Dowrey, C. Marcott, 'Two-dimensional Infrared (2-D IR) Spectroscopy', in *Modern Polymer Spectroscopy*, ed. G. Zerbi, John Wiley & Sons, Weinheim, 1–32, 1998.
20. A.E. Dowrey, G.G. Hillebrand, I. Noda, C. Marcott, 'Dynamic Infrared Linear Dichroism (DIRLD) Spectroscopy of Human-hair Keratin', *SPIE*, **1145**, 156–157 (1989).
21. I. Noda, 'Dynamic Infrared Dichroism and Two-dimensional Correlation Spectroscopy', Doctoral Thesis, The University of Tokyo, 1997.

Biomolecules Analysis: Introduction

R.A. Meyers
RAMTECH Ltd, Tarzana, USA

Analysis of the various classes of biomolecules are covered in several sections of this Encyclopedia, including *Biomedical Spectroscopy*, *Carbohydrate Analysis*, *Clinical Chemistry*, *Nucleic Acids Structure and Mapping*, and *Peptides and Proteins*. Bioactive molecules analysis is covered in the section on *Pharmaceuticals and Drugs*, and some articles also directed to biomolecules can be found in the sections on *Forensic Science*, *Environment: Water and Waste*, and *Industrial Hygiene*.

The purpose of this section is to present bioanalytical methods which (a) are applicable to several biomolecule classes (e.g. proteins, lipids, and nucleic acids), (b) are applicable to biomembranes, cells and tissues, (c) cover chirality in biomolecules, or (d) provide analysis methods for small biomolecules (e.g. amino acids, neuromediators, and metabolites). The articles in this section contain information on sample preparation, matrix selection and detailed equipment descriptions for procedures, not covered elsewhere in the Encyclopedia.

The article on **Circular Dichroism in Analysis of Biomolecules** contains information on the available instrumentation for electronic and vibrational circular dichroism as well as experimental aspects for both qualitative and quantitative methods and a comparison of these methods with others for biomolecular analysis. The biomolecules treated include proteins, carbohydrates, porphyrin complexes and nucleic acids bases.

The **Fluorescence-based Biosensors** article includes instrumentation, interfacing biomolecules to devices, protein engineering for sensor design, applications to enzymes and enzyme-based sensors, fluorescent immunosensors, regulatory proteins, DNA-/RNA-based gene sensors, membrane-associated proteins and multi-analyte sensing.

High-performance Liquid Chromatography of Biological Macromolecules covers development modes, eluents, separation columns and chemistries, detection techniques, sample preparation and the validation of separation processes. Preferred techniques are presented for oligonucleotides, polynucleotides, peptides, proteins, and carbohydrates and their conjugates.

Infrared Spectroscopy of Biological Applications includes infrared spectroscopy of proteins, lipids, biomembranes, animal tissue, carbohydrates and plant material.

Mass Spectrometry in Structural Biology covers the use of matrix-assisted laser desorption/ionization time-of-flight in the analysis of lipids in membranes as well as proteins (collagen) and polysaccharides (e.g. glycosaminoglycans), from connective tissue and small molecules and common impurities in biological samples. Sample preparation is described and examples are given of the analysis of physiologically relevant molecules in comparison with other mass spectroscopy methods, such as electron ionization, chemical ionization, fast atom bombardment, and electrospray ionization.

The article on **Nuclear Magnetic Resonance of Biomolecules** contains information on instrumentation, spectral assignments and biomolecular structure determination including proteins, enzymes, nucleic acids and membranes. Information is also given on long-range distance interactions and angles.

Raman Spectroscopy in Analysis of Biomolecules presents details of sample handling and preparation, as well as the interpretation of Raman spectra with applications to amino acids and peptides, active centers in proteins, nucleic acids, lipids and biomembranes. Dynamics in biomolecules and microscopic analysis of biological systems are also covered.

The article on **Single Biomolecule Detection and Characterization** contains fundamental information on the interaction of light with single molecules and fluorescence correlation spectroscopy in solution, as well as the microscopy and detection of single molecules on surfaces using near-field microscopies and classical far-field microscopy. Preparation of biological macromolecules and laser micromanipulation are also covered, as is single molecule biophysics in biochemistry. The molecules analyzed include DNA in the molecular motors, kinesin, dynein, and actin.

Vibrational Optical Activity of Pharmaceuticals and Biomolecules covers the two principal instrumental forms for measuring the optical activity of biomolecules: vibration circular dichroism and Raman optical activity. Instrumentation is described from basic principles, and diagrams of typical equipment are presented. Examples include proteins, nucleic acids, and pharmaceuticals.

Voltammetry In Vivo for Chemical Analysis of the Living Brain describes electrode types and their preparation, as well as voltammetric techniques and surgical protocols. The substances targeted are neuromediators and metabolites: ascorbic acid, uric acid, dioxygen, nitric oxide, glucose, lactate, and glutamate.

Voltammetry In Vivo for Chemical Analysis of the Nervous System presents information on appropriate electrodes, recording methods, instrumentation and the preparation of cells, tissue slices, tissue pieces, and whole animals, as well as quantification and identification issues in comparison with other methods. Examples of specific bioactive compounds analyzed include dopamine, serotonin, norepinephrine and nitric oxides in cells, tissues, and whole animals.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Circular Dichroism in Analysis of Biomolecules

Petr Pančoška
University of Illinois at Chicago, Chicago, USA

1 Introduction **2**
1.1 Principles of Encoding/Decoding of the Biomolecular Features in Circular Dichroism Spectral Features 2
1.2 Advantages and Limitations 9
2 History and Available Instrumentation **10**
2.1 Single Beam Electronic Circular Dichroism Spectrometer 11
2.2 Dual Beam Electronic Circular Dichroism Spectrometer 11
2.3 Dispersive Vibrational Circular Dichroism Spectrometer 13
2.4 Fourier Transform Vibrational Circular Dichroism Spectrometer 14
2.5 Developmental Designs 15
3 General Experimental Aspects **16**
3.1 Instrument Calibration Procedures 16
3.2 Sampling Techniques 17
4 Structural Interpretation of Circular Dichroism Spectra of Biomolecules **18**
4.1 Quantitative Empirical Methods 18
4.2 Analysis of the Reference Set Properties Using Principal Components 19
4.3 Analysis of the Compatibility of the Analyzed Protein Spectrum and the Reference Set 23
4.4 Description of the Spectral Variance 25
4.5 Description of the Spectra–Structure Relation 26
5 Quantitative Relation of Circular Dichroism Spectral Information to Other Spectroscopic Techniques **33**
5.1 Two-dimensional Correlation Analysis of Series of Circular Dichroism Spectra 33
5.2 Neural Network Projections Between Circular Dichroism and Other Methods 33
6 Conclusions **34**
Acknowledgments **34**
Abbreviations and Acronyms **34**
Related Articles **34**
References **35**

Circular dichroism (CD) spectroscopy belongs to the family of chiroptical methods. These methods utilize the interaction of circularly polarized light (CPL) with chiral molecules and molecular systems to obtain more detailed information about their structure and electronic or vibrational states. CD spectrum $\Delta\varepsilon(\nu)$ is, in fact, the difference of two absorbance spectra: one measured with the left circularly polarized light (LCPL) and the other one recorded with the right circularly polarized light (RCPL) $\Delta\varepsilon(\nu) = \varepsilon_L(\nu) - \varepsilon_R(\nu)$. CD spectra can be measured as electronic circular dichroism (ECD) in spectral regions of electronic transitions (ultraviolet (UV)) through visible (VIS) and as vibrational circular dichroism (VCD) in the infrared (IR) spectral regions. An ECD spectrum can be also recorded as the difference excitation spectrum for fluorescence spectra excited by LCPL and RCPL, respectively (fluorescence-detected circular dichroism (FDCD)). The Raman optical activity (ROA) and circularly polarized luminescence spectroscopies complete the family of currently developed chiroptical methods. The differences $\Delta\varepsilon(\nu)$ measured as CD are typically $\sim 10^{-5}$ of sample absorbance. Special modifications of dispersive (for ECD and VCD) or Fourier transform infrared (FTIR) (for VCD) spectrometers are needed to measure CD with a reasonable signal-to-noise ratio (S/N). Polarization modulation of the incident light by a photoelastic modulator and synchronous electronic processing of the resulting photoelectric signal in the spectrometers are typically used for this purpose.

Molecular structure is encoded in the CD spectra because the chiral field of the CPL wave that induces a spectral transition in the chiral molecule can be observably altered both by the transition electron density redistribution (as in the conventional absorption spectroscopy) and by the transition magnetic field accompanying the molecular charge redistribution. The structure and conformation of the studied molecule define the relative orientation of the characteristic directions of these two effects. This relative orientation affects the probability of absorption of photons of RCPL and LCPL and, in turn, determines the CD sign, the primary information that is unique for CD spectroscopy. In the absolute value, CD intensity (''secondary'' unique information) is related to the angle of electric and magnetic transition moments and can, therefore, be converted into the molecular conformation once the suitable theoretical model is available. For complicated molecules (biomolecules, proteins, nucleic acids) where the corresponding theoretical calculations are too complex, the empirical interpretive methods based on reference sets of spectra measured for molecules

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

with known structures (from X-ray or nuclear magnetic resonance (NMR) experiments) are used with success. For example, using these empirical methods, the fractions of regular secondary structures (secondary structure fraction (SSF)) in globular proteins can be determined with a relative error of 3–7% and the number of secondary structure segments in proteins with a typical error of one to three segments per protein fold.

CD spectroscopy is highly sensitive to the polarization artifacts that can be related to the optical imperfections of the optical parts of the spectrometer. To achieve an acceptable S/N, the sample total absorbance A should be also controlled (typically $A < 0.8$). This, in effect, restricts the selection of usable solvents and largely determines the concentration ranges of studied samples.

1 INTRODUCTION

The study of differential response of molecules to circularly polarized radiation is a basis of the so-called chiroptical spectroscopic methods. The most frequently used chiroptical method is CD, which is known as ECD in UV/VIS and VCD in the IR frequency regions, respectively. The differential polarized absorption spectrum $\Delta\varepsilon(\nu)$ is what is measured experimentally as CD (Equation 1):

$$\Delta\varepsilon(\nu) = \varepsilon_{\mathrm{L}}(\nu) - \varepsilon_{\mathrm{R}}(\nu) \tag{1}$$

Here $\varepsilon_{\mathrm{L}}(\nu)$ and $\varepsilon_{\mathrm{R}}(\nu)$ are absorption spectra measured with LCPL and RCPL.[1–5] This phenomenon is the result of a difference in the interactions of three chiral fields: that of a molecule lacking symmetry (no rotation reflection axis (S_n where $n \geq 2$) can be present in its structure) and the electromagnetic fields of the two (chiral) states of the CPL (LCPL and RCPL). Similarly, the chiral molecular field of an electronically excited chiral molecule will generate unequal intensities of emitted LCPL and RCPL upon the radiative dissipation of the excitation energy.[6] There is also an observable differential Raman scattering of the photons of LCPL and RCPL by the polarizability of a chiral (i.e. optically active) molecule (ROA).[7–20] This differential phenomenon of parity-even observables gives rise to the diversity of chiroptical methods.[21] In this article, the discussion will be restricted to applications of ECD and VCD in biomolecular structural studies.

CD phenomena can be treated theoretically at different levels. Rotational strength R_{oj} is the quantity that explains the existence of CD for any optical transition from the ground (o) to the excited (j) state of a chiral molecule. This quantity can be expressed in terms of the area under the dichroic band by Equation (2):

$$R_{oj} = 7.663 \times 10^{-54} \quad \frac{\Delta\varepsilon(\lambda)}{\lambda}\, \mathrm{d}\lambda[CmJT^{-1}] \tag{2}$$

where the multiplicative factor results from the evaluation of constants in the theoretical formula for rotational strength expressed in SI units.

Derivation of the theoretical representation of rotational strength can be found in classical textbooks, as well as in recent articles.[21,22–27] These works have a common final result given by Equation (3).

$$\begin{aligned} R_{oj} &\sim \mathrm{Im}\langle\psi_o|\vec{\mu}|\psi_j\rangle\langle\psi_j|\vec{m}|\psi_o\rangle \sim \mathrm{Im}\,\vec{\mu}_{oj}\vec{m}_{oj} \\ &\sim \mathrm{Im}|\vec{\mu}_{oj}||\vec{m}_{oj}|\cos\alpha \end{aligned} \tag{3}$$

Here ψ_o and ψ_j are wave functions of the ground and the excited states of the studied molecule, respectively; $\vec{\mu}$ and $\vec{m}$ are operators of electric and magnetic dipole moments, and $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$ are the corresponding (transition) matrix elements. Because of the pure imaginary character of $\vec{m}_{oj}$, the observable CD is related to the imaginary part (Im) of the generally complex (dot) product of the electrical and magnetic transition dipoles $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$. The last form of Equation (3) for the rotational strength emphasizes the relationship between the sign of the experimentally observed CD and the relative orientation of the electric and magnetic transition dipole moments $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$. The CD sign is, therefore, the most informative feature of the effect.

Theoretical calculation of rotational strength from first principles is the purest way to achieve a rigorous interpretation of the experimental CD spectra. Unfortunately, the high quality of the wave function required for a reasonable quantitative comparison of calculation and experimental results restricts this possibility to relatively small molecules. However, development of quantum chemical methods is constantly moving the threshold of a ''small'' molecule toward larger and larger numbers of atoms.[30,31] Nevertheless, approximate and empirical or heuristic methods of CD spectra interpretation remain important for large biomolecular systems such as proteins or nucleic acids, where reasonable ab initio calculational results are beyond the current computational resources.

1.1 Principles of Encoding/Decoding of the Biomolecular Features in Circular Dichroism Spectral Features

Theoretical analysis of the rotational strength allows for the recognition of the so-called mechanisms of its origin.[24,25,32] These mechanisms are the dominant physical principles that convert selected structural features into the experimentally observable characteristics of CD spectra: sign, frequency at maximum, and intensity of the CD band (here ordered according to the significance of their informational content). Measurement and analysis of experimental data series on representative biomolecular sets are needed to identify the dominant

mechanism that best describes the chiroptical properties of a structurally related group of biomolecules.

This section summarizes those mechanisms that are important for biomolecular CD. Inherent chirality, the one-electron mechanism, and coupled oscillator models are based on classical analysis of rotational strength properties, which, in turn, uses the model of independent groups (chromophores). In biomolecules, these groups frequently correspond to monomers or the well-defined parts of monomers, such as amino acid residues, amide groups, side chains, nucleotides, nucleic acid bases, etc. This simplification facilitates the transfer of theoretical and experimental knowledge about the origin of rotational strength in these groups to polymers. The overall rotational strength of the whole system is then considered as a sum of the contributions from individual groups. This calculation either neglects (in the simplest approaches) or explicitly includes the effect of the intergroup interactions. Details of this approach can be found in numerous excellent papers.[33–36] Selected examples are given below to help in the qualitative understanding of the multiplicity of structure-encoding principles in CD spectra. Further factors that can affect the CD spectra, including the induced CD and polymer length, are also discussed. Theoretical aspects of interpreting the rotational strength for VCD are discussed briefly.

1.1.1 Inherent Chirality

If a molecule is inherently chiral, the rotational strength reflects the overall chirality of the individual, independent groups (chromophores) that are defined by the asymmetry of the whole molecule. A biologically interesting example of such a system can be the disulfide chromophore with the dihedral angle of the S–S bond deviating from 0° or 180°. The derivation of the relationship between that torsional angle and the CD $n \rightarrow \sigma^*$ band sign for an isolated disulfide bond transition is shown in Figure 1. Once this mechanism dominates, the experimental CD spectrum is monosignate with a maximum close to that of the corresponding absorption band. This CD spectrum might be relatively inert to changes of the substituents in the molecular structure as long as the chirality of skewed groups in the molecule is not inverted by such structural modifications.

1.1.2 One-electron Mechanism

In the one-electron mechanism approximation, the rotational strength can be considered to be the result of state mixing in individual achiral chromophores induced by a static perturbation of these states by the chiral molecular field. An example, explaining how the sign of the carbonyl chromophore $n \rightarrow \pi^*$ CD band is related to the position of the substituent that breaks its symmetry, is shown in Figure 2. If this mechanism dominates the rotational strength generation, then the sign of the dichroic band in the experimental CD spectrum can be used to determine the absolute stereochemistry of chiral centers in (small) biomolecules. These simple relationships between CD sign and molecular structure are known as sector rules. Sector rules associate the sign of a band in the CD spectrum with the location of the perturbing substituent group in a space sector (quadrant, octant) that surrounds the chromophore, as is shown in Figure 2. Historically, this was the first practical structural application of ECD spectroscopy.[39]

Algebraic theory of chirality functions reveals that these rules are only special cases of the general functional description of the perturbation required to induce chirality in the originally achiral molecule.[40,41] In Ruch's chirality function theory, the molecule is characterized by the point symmetry group of the achiral molecular skeleton, and by the number n of bonding sites on that skeleton that can accommodate ligands which will break the skeleton symmetry once in those positions. Algebraic analysis then determines the minimal number of different substituents needed for that to occur. This number can be found with the use of the so-called chirality function $\Phi(\overrightarrow{L})$ parameterized by the structural descriptors $\overrightarrow{L} \equiv \{l_1, l_2, \ldots, l_n\}$ of the perturbing groups. The theory proves that this function describes the sign of a perturbing group contribution to the CD spectra and can determine this sign as a function of the ligand position relative to the skeleton.

A chirality function is generated by projectors based on the skeleton point symmetry group and characters of the permutation group S_n, where the order n is equal to the number of bonding sites on the skeleton. The classical sector rule is meaningful only if the molecule can be characterized by the chirality function ${}_i\phi(l_i)$, which contains terms depending exclusively on the contributions from single perturbing groups in the individual bonding sites. In a general case, the chirality function contains terms $\phi(l_i, l_j, \ldots, l_k)$, which are dependent on the parameters of ligands from multiple bonding sites. With a molecule characterized by such a chirality function, a unique relationship of the CD band sign to a certain spatial position of a perturbing group in a space sector surrounding the chromophore cannot be simply found. An example of such a chirality function derived for an idealized skeleton of natural porphyrins is shown in Figure 3.[42,43]

If a rotational strength is generated predominantly by the one-electron mechanism, the observed CD will be a monosignate band in the frequency range of the corresponding absorption peak. The CD sign will be strongly dependent on the modifications of the substituent pattern in the skeleton.

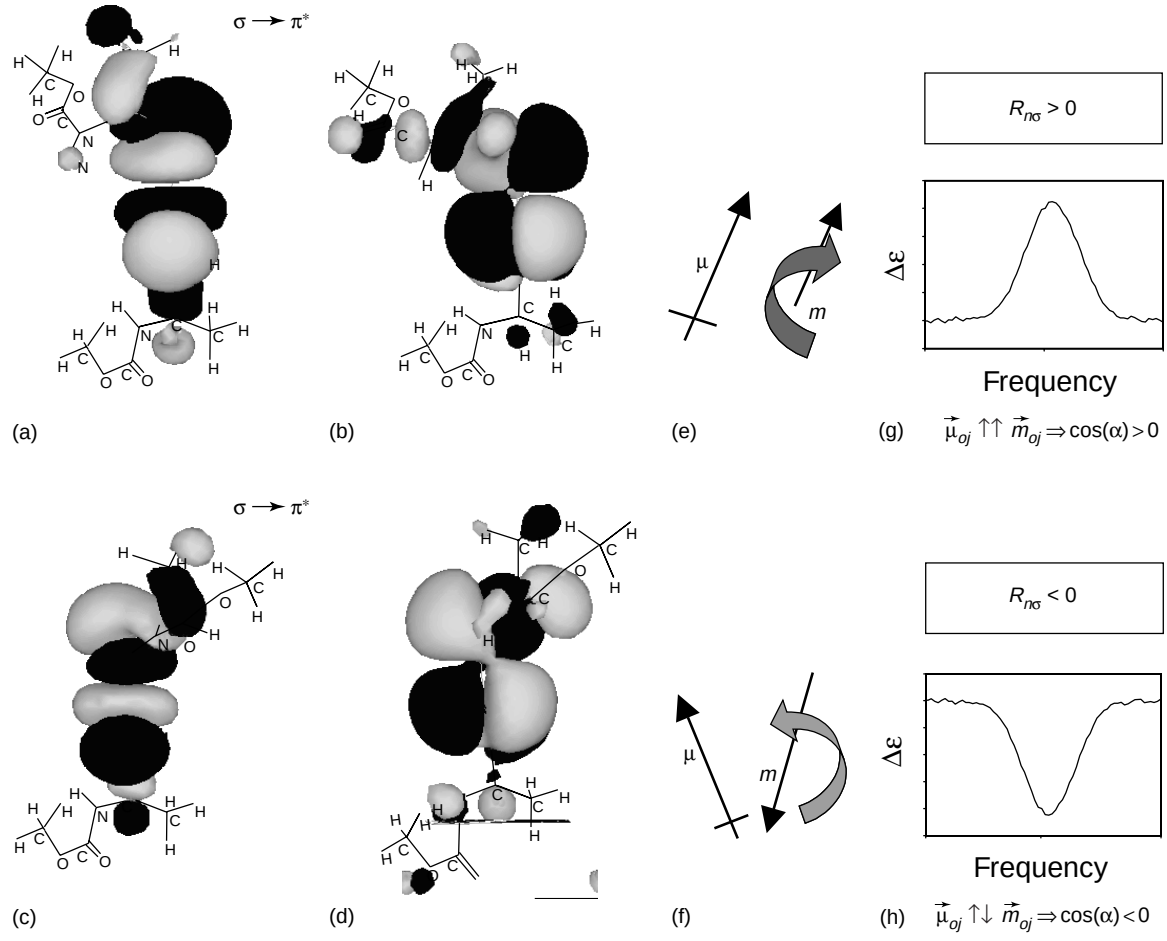

Figure 1 The mechanism by which the rotational strength is derived from the inherent chirality of the chromophore. An example of the relationship between the sign of the disulfide bond torsion angle and the sign of the dichroic band corresponding to the $\sigma \rightarrow \pi^*$ transition in this chromophore. The two σ orbitals (a, c) and π^* (b, d) of $\sigma \rightarrow \pi^*$ transition are shown for the two opposite values of S–S torsion angle (positive a, b; negative c, d) that are deviated from 0° and 180°. During the transition lifetime, and for both conformations, the electron density is redistributed from the bottom to the top of the chromophore as drawn. This change follows the ''rule'' of combination of involved orbital parts with identical and opposite phases as transition localization of partial excess and deficiency of the electron density, respectively.[34] The overall character of this redistribution can be described by the electric transition dipole moments $\vec{\mu}_{oj}$ shown as arrows (e, f). The twist of the S–S bond from planarity introduces the ''electron density'' rotation during this transition that generates the magnetic transition dipole needed for observable rotational strength. The direction of the $\vec{m}_{oj}$ can be determined by the ''right-hand rule'' assuming again that the rotation of the electron density follows the path of identical phases of the *o* and *j* orbitals and the geometry of their structurally based distortions. For a positive value of the S–S torsion angle, the vectors $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$ have parallel components (e) and, therefore, positive rotational strength is predicted (g). For a negative value of the S–S torsion angle, the vectors $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$ have mainly antiparallel components (f) and, therefore, negative rotational strength is predicted for the $\sigma \rightarrow \pi^*$ dichroic band (h). The orbitals were calculated using the ZINDO/S method (Hyperchem v. 5.3).[37,38]

1.1.3 Coupled Oscillator Models

All of the above mechanisms require involvement of the magnetic transition dipole $\vec{m}_{oj}$ to produce CD. Alternatively, the rotational strength can be explained without explicitly considering the magnetic transition dipole moment in structures with repeated monomeric units of oligo- and polymeric biomolecules. In these systems, the opposing CPL interacts differently with the chiral spatial distribution of purely electric transition dipoles of monomers. If the observed rotational strength is dominated by this ''coupled oscillator'' mechanism, the relation of the CD sign to the structure (conformation) is given by the scalar triple product of vectors $\vec{\mu}_i$, $\vec{\mu}_j$ and $\overrightarrow{T_{ji}} = \overrightarrow{T_j} - \overrightarrow{T_i}$. These vectors represent the electric transition dipoles in the monomers *i* and *j*, and the distance vector of the centers of mass of each $\vec{\mu}$, respectively. The rotational strength of the coupled oscillator system can be computed in the following way:

$$R_k = -\frac{\pi}{c} \sum_{i=1}^{N} \sum_{j=1}^{N} C_{ik} C_{jk} (\nu_j \overrightarrow{T_j} - \nu_i \overrightarrow{T_i})(\vec{\mu}_j \times \vec{\mu}_i) \qquad (4)$$

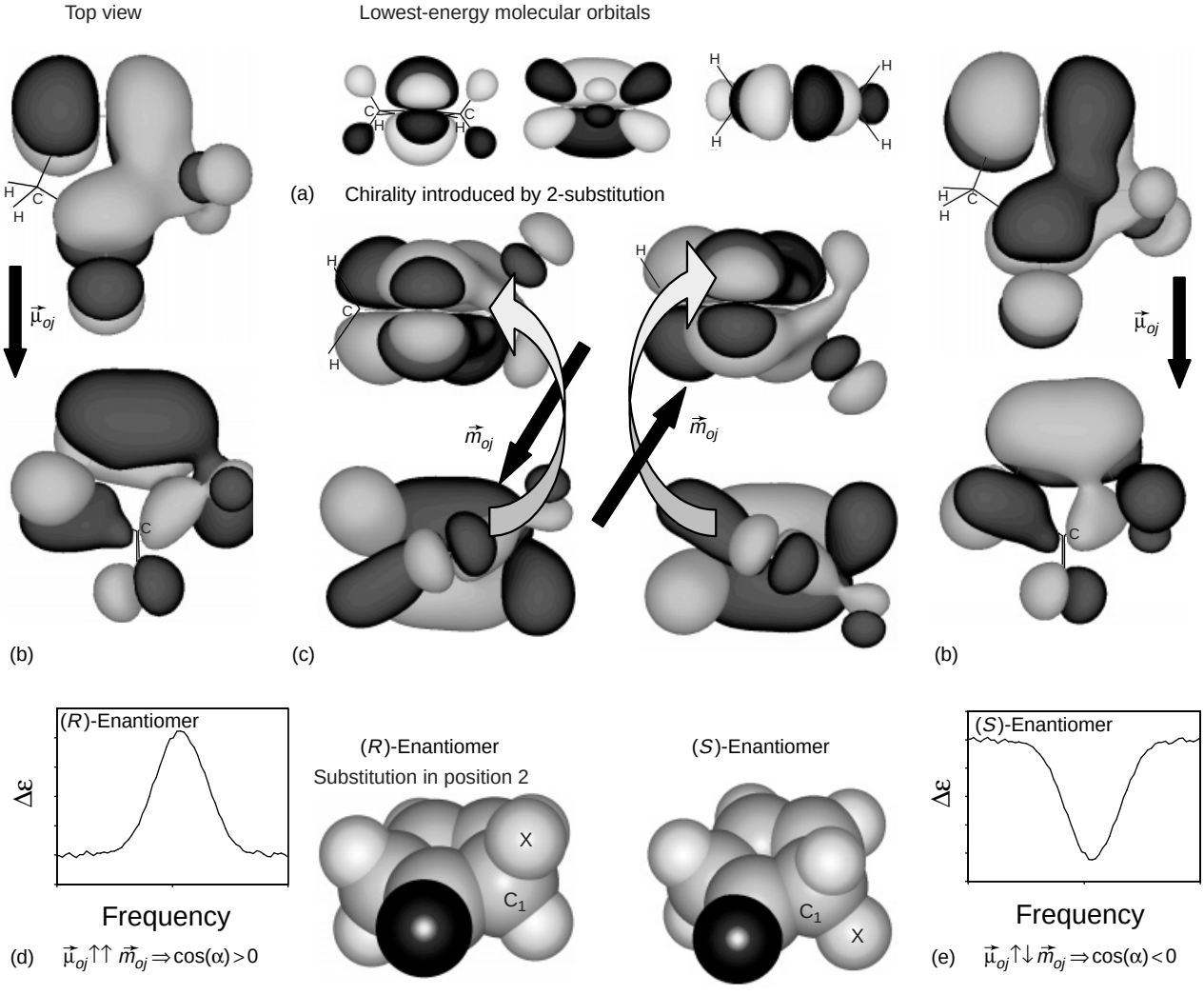

Figure 2 Explanation of the rotational strength using the one-electron mechanism. An example of the relationship of the absolute configuration of chiral center C_1 and the sign of the lowest-energy dichroic band of the ECD spectrum of a molecule containing the carbonyl chromophore. The lowest-energy molecular orbitals of achiral cyclopenten-1-one (n, π, and π^*) are shown looking in the C=O direction at the top of the figure. The substituent X in position 2 (see structures shown at the bottom of the figure) induces mixing of n and π orbitals, which provides the magnetic transition moment necessary for observing the nonzero rotational strength. The absolute configuration of the chiral center determines both the geometrical distortion and the phase pattern of the perturbed HOMO orbital that is induced by the mixing of the original orthogonal n and π orbitals. During the transition lifetime, the electron density is redistributed from the ring towards the oxygen atom for both enantiomers. This redistribution can be described by the electric transition dipole moment $\vec{\mu}_{oj}$ with the direction shown as the arrow labeled $\vec{\mu}_{oj}$ (see top view of the molecular orbitals). Distortion of the HOMO orbital by the 2-substituent perturbation introduces chirality into this transitional electron density redistribution that generates the transition magnetic dipole moment $\vec{m}_{oj}$. Its direction can be determined by the ''right-hand rule'' assuming that the electron density ''current'' follows the path of identical phases of the o and j orbitals. Consequently, the direction of $\vec{m}_{oj}$ is opposite for both enantiomers. For the (R)-isomer, the transition dipoles $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$ have parallel components and, therefore, a positive dichroic band is predicted. For the (S)-isomers, $\vec{\mu}_{oj}$ and $\vec{m}_{oj}$ have mainly antiparallel components and a negative CD band is predicted. The orbitals were calculated using the ZINDO/S method (Hyperchem v. 5.3).[37,38]

Here C_{ik} and C_{jk} are the eigenvector components of the dipole–dipole interaction matrix; N is the number of interacting transitions; and c is the speed of light. Electric transition dipoles are either determined from experimental data or calculated.[34,36,44]

For regular infinite secondary structures (such as α-helices), the selection rules based on the translational symmetry of the conformation are instrumental for the simplification of the double sum in Equation (4). For example, symmetry of an idealized infinite α-helical polypeptide eliminates all but three terms in Equation (4). These terms correspond to the three observed component bands of the CD of helical polypeptides and proteins, with polarization parallel and perpendicular to the helix main axis.[45,46]

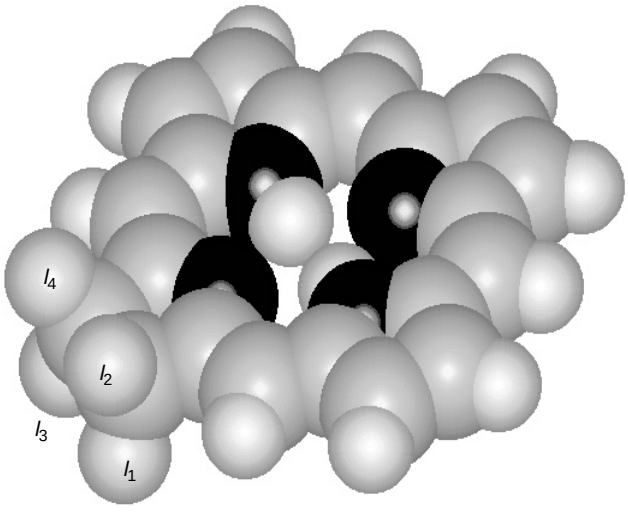

$$\Phi(\vec{L}) = \phi(l_1) - \phi(l_2) + \phi(l_3) - \phi(l_4) + \chi(l_1,l_2) + \chi(l_2,l_3) + \chi(l_3,l_4) + \chi(l_4,l_1)$$

Figure 3 Chirality function derived using Ruch's theory for porphine, the parent structure of naturally occurring porphyrins. Hydrogenation of the imidazole ring IV of porphine generates the ''molecular skeleton'' of C_{2v} symmetry with four ligand-binding sites l_1–l_4. Chirality function $\Phi(\vec{L})$, as shown in the figure, can be derived using the characters of the C_{2v} point group and the representation of permutation group S_4. Corresponding projectors define the polynomial structure of the overall chirality function in terms of the components $\phi(l_i)$ depending on the properties of single ligand and $\chi(l_i, \ldots, l_N)$, describing the contributions to the molecule chirality that arise from the interactions of ligand pairs. Descriptor l_i characterizes quantitatively the chemical composition of a ligand (in applications, parameter l_i can be the ligand diameter or some theoretically derived energy characteristic). The sector rule in the classical form can be observed for this molecular class only if all the functions $\chi(l_i, \ldots, l_N)$ vanish and only terms $\phi(l_i)$ remain in the chirality function.

A coupled oscillator mechanism can be successfully applied in semiquantitative biomolecular structural studies. Applications relying on this mechanism range from the determination of the absolute stereochemistry of aromatic compounds, through ECD study of absolute stereochemical assignments for polyols and sugars,[47] to the study of biopolymer conformation. The coupling of the monomer electric transition dipoles successfully explains the qualitative characteristics of the experimental CD spectra of proteins and nucleic acids in all regions.[48–56] The explanation of the opposite ECD signs for an idealized dinucleotide in the (*B*)- and (*Z*)-conformations is shown in Figure 4.[57] Dipole coupling described by this mechanism also explains the dichroic bands of aromatic side chains in the near UV region (360–260 nm) of protein ECD spectra. The extent of the dipole coupling is strongly dependent on the distance between the interacting aromatic rings, and this distance is determined by the protein tertiary structure. Therefore, the long-wavelength part of the protein ECD spectrum is primarily sensitive to this level of protein conformation.[58]

A unique indicator of the dominance of this mechanism in the experimental CD is the bisignate shape of the observed CD with the zero crossing near the maximum of the corresponding absorption band. In an ideal, energetically degenerate case, the intensities of both exciton components of the bisignate CD should be identical. If observed experimentally, such CD is called conservative. The nonconservative bisignate CD spectra with uneven intensity of the components are measured more frequently. Intensity redistribution can be explained by the multiple mixing of nondegenerate states that determines the coefficients C_{ij} in Equation (4).

1.1.4 Vibrational Circular Dichroism

The CD of vibrational transitions is weaker by one or two orders of magnitude compared with that of electronic transitions. This is a consequence of the inverse proportionality of $\Delta\varepsilon/\varepsilon$ to the wavelength.[21,23] The magnetic dipole transition moment that is needed for a theoretical description of the vibrational transition rotational strength can be obtained from the derivatives of the electric dipole moment surface. These derivatives should be calculated at the equilibrium nuclear geometry with respect to the displacement of nuclei from their equilibrium positions. The vibrational magnetic transition moment is related to the transition electron currents associated with the vibrational transition. The corresponding theoretical matrix elements are thus linearly related to the relative velocities of the vibrating nuclei. Although there have been attempts to produce pictorial descriptions of VCD mechanisms,[4] they were unsuccessful because of their limited validity.[30,31] A nondegenerate coupled oscillator mechanism can be used for semiquantitative structural analysis of the VCD of biopolymers such as nucleic acids and peptides.[49,52]

1.1.5 Induced Chirality

In addition to the induction of optical activity by the application of external electrical or magnetic fields (Kerr effect, magnetic circular dichroism (MCD)),[62,63] achiral molecules can become optically active through specific bonding to chiral counterparts (molecular matrices). An ordered and site-specific spatial relationship between the achiral molecule and its chiral molecular host is a necessary condition for observation of this so-called induced CD. The relative orientation of the chromophore and the host should be uniform in all realizations of such a specific supramolecular system in the solution, otherwise the variable signs of induced CD would lead to the

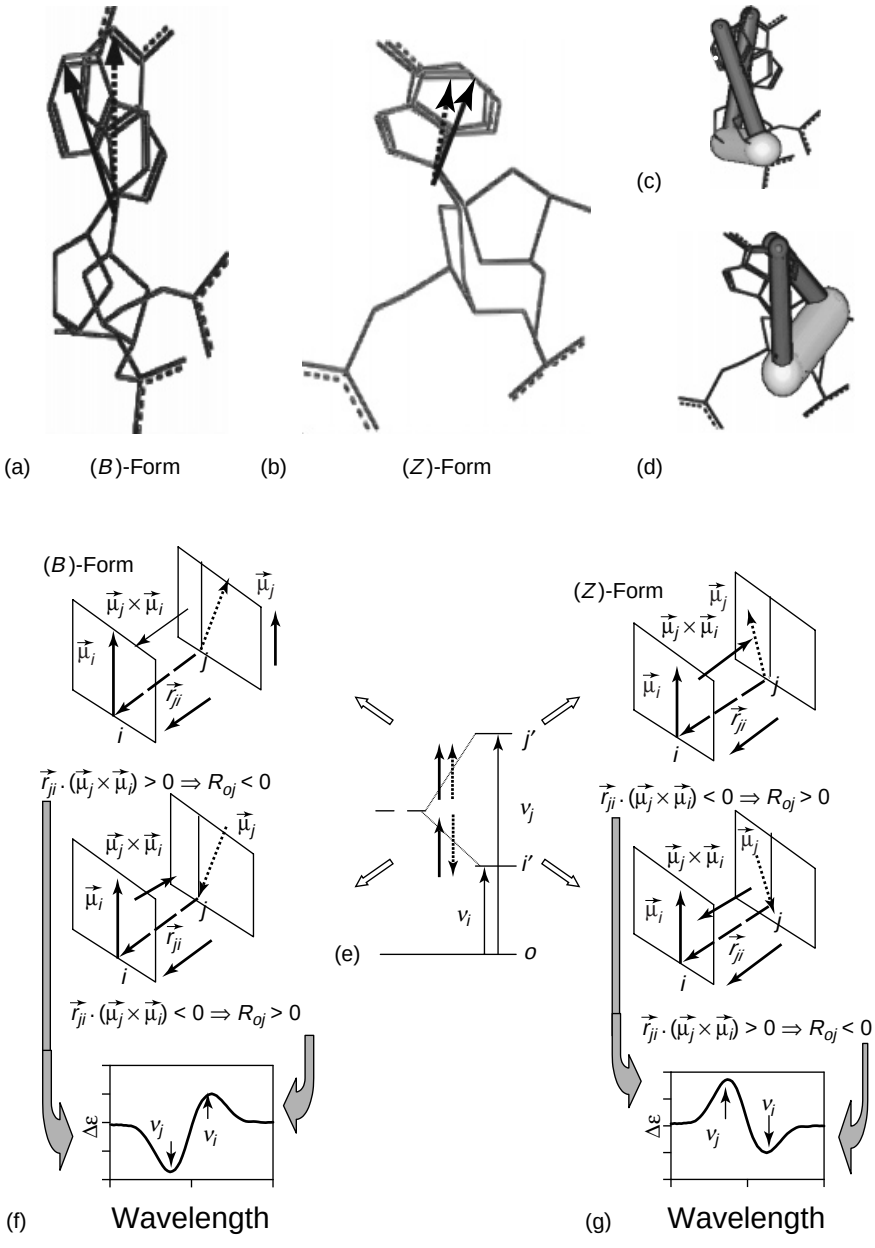

Figure 4 The use of a coupled oscillator mechanism to explain the origin of rotational strength. (a, b) The orientation of two neighboring bases in a polynucleotide in the (*B*)- and (*Z*)-forms. Arrows show the relative orientation of identical parts of the base structure. Full-line and dotted-line arrows are closer and more distant to the viewer, respectively. The opposite handedness of the two helical secondary structures is also shown using the ladder rendering (c, d). Perturbation of the degenerate electronic states of individual bases *i* and *j* generates two new states *i*′ and *j*′. One has higher and the other has lower energy than the degenerate original states. During the electronic transition, the higher-energy state corresponds to the parallel orientation of the electric transition dipole moments; the lower-energy state corresponds to their antiparallel orientation (e). Relative spatial orientation of the two transition electric dipoles $\vec{\mu}_j$ and $\vec{\mu}_i$ defines the directions of the distance vector $\vec{r}_{ji}$ and the vector product $\vec{\mu}_j \times \vec{\mu}_i$, respectively, that are needed for the determination of the sign of their scalar triple product, yielding the sign of the corresponding rotational strength. For the *B*-form (f) of the polynucleotide conformation, the exciton component of the bisignate dichroic band at shorter wavelength is negative. The triple product is in this case positive because the components of the $\vec{r}_{ji}$ and $\vec{\mu}_j \times \vec{\mu}_i$ vectors are parallel and rotational strength has the opposite sign to the triple vector product $\vec{r}_{ji} \cdot (\vec{\mu}_j \times \vec{\mu}_i)$ (see Equation 4). Using the same arguments, the exciton component at longer wavelength is predicted to be positive. The opposite signs of the corresponding components are found for the *Z*-form (g) as a result of the reversed chirality of the vectors in the $\vec{\mu}_j \times \vec{\mu}_i$ product. The predictions are in agreement with the experimental CD spectra of these DNA conformations.[57]

effective cancellation of the dichroic intensity. Examples of well-defined complexes that exhibit strong ECD upon formation are dye molecules that intercalate into a double helical nucleic acid host; carotenoid molecules fixed on the protein matrix in photosynthetic chlorophyll–protein complexes;[64,65] or complexes of cationic tetraphenyl porphyrins with sequential basic polymeric di-, tri-, and tetrapeptides.[42,66] Although theoretically equally possible, induced VCD has not yet been studied.

The dependence of induced CD on the molecular geometry and the chromophore interaction energy with the host matrix was theoretically analyzed using general symmetry and parity arguments.[67] The magnitude

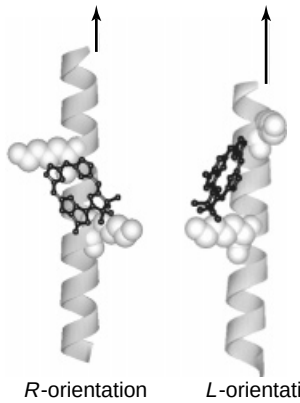

$$\Delta R = R^{R}_{ss'} - R^{L}_{ss'}$$

$$\Delta R = \frac{1}{2}\left(\sum_{ss'}\left(\rho^{\alpha R}_{ss'} + \rho^{\alpha L}_{ss'}\right)\left(\left\langle \phi^{+}_{s} \middle| \hat{R} \middle| \phi^{-}_{s'} \right\rangle + \left\langle \phi^{-}_{s} \middle| \hat{R} \middle| \phi^{+}_{s'} \right\rangle\right)\right)$$

Here $\phi^{\pm}_{s} = \frac{1}{\sqrt{2}}\left(\phi^{R}_{s} \pm \phi^{L}_{s}\right)$ and $\phi^{R,L}_{s}$ are wave functions characterizing the porphyrin in *R*- and *L*-orientation, respectively, and $\rho^{\alpha R}_{ss'}$ terms are density matrix elements $\rho^{\alpha R}_{ss'} = \left\langle \chi^{\alpha R}_{s} \middle| \chi^{\alpha R}_{s'} \right\rangle$ calculated using the wave functions $\chi^{\alpha R,\,L}_{s}$ characterizing the peptide in α-helical conformation with pigment bonded in *R*- or *L*-orientation.

(b)

Figure 5 Induced chirality of ordered molecular complexes of an achiral chromophore with a chiral polymeric matrix. Experimentally, induced CD in porphyrin transitions was observed for anionic and cationic tetraphenyl porphyrins complexed with polycationic sequential polypeptides and polynucleotides, respectively. (a) An idealized example is shown of the two complexes of an achiral porphine and α-helical polypeptides poly(L-Lys-L-Ala-L-Ala) and poly(L-Lys-L-Ala-L-Ala-L-Ala). The different helicity of the L-lysine side chain distributions on the backbone structure generates the enantiomorphic orientations *R* and *L* of the pigment with respect to the direction of the polypeptide helix (arrows in the two complexes). (b) A theoretical analysis of the difference in the induced rotational strength for the two conformations (ΔR) using the derived formula shows that this difference is dependent on the strength of the interaction between the chromophore and the peptide (density matrix elements $\rho^{\alpha R,L}_{ss'}$ are directly related to that interaction energy) that modulates the sum of the rotational strength operator matrix elements for the states s and s'.

difference was predicted for the induced CD of two mirror image geometries of the complex. This difference is modulated by the weight factor, depending on the interaction energy between the two parts of the chiral complex. A schematic explanation of this result is shown in Figure 5.

1.1.6 Length Dependence

There is yet another structural feature that can influence CD spectra of polymeric biomolecules. The CD intensity for these molecules depends nonlinearly on the biopolymer length, defined in terms of its monomeric unit number N. Taking this effect into account improved the quantitative predictions of SSF from ECD spectra.[5] The correction of the component spectra according to the length of the analyzed protein was implemented as an empirical function $(1 - C(\lambda)/N)$, which modifies the ECD band amplitude. Here, C is the wavelength-dependent parameter that is optimized experimentally. Theoretical analysis of this phenomenon showed that the above empirical function is only a special case of a more general

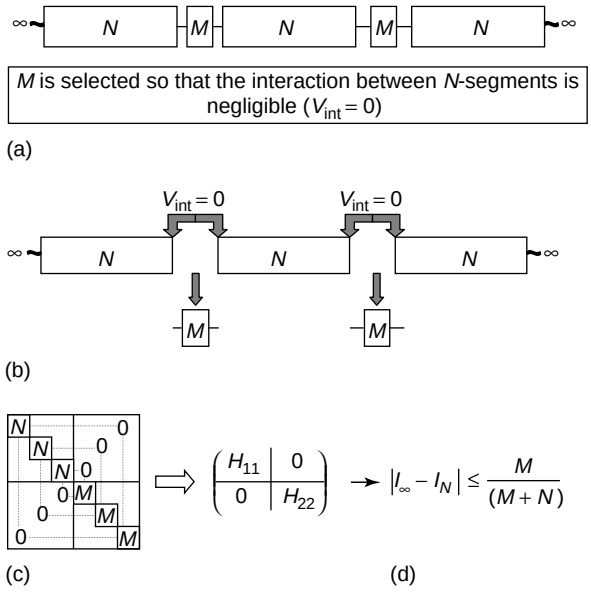

Figure 6 Steps in the derivation of length dependence of rotational strength (Equation 5). (a) The infinite length polypeptide is partitioned into two types of segments: those with the length N and those with length $M (N \gg M)$. (b) M is chosen in such a way that if the M-segment is removed from the chain, there is no interaction between the isolated N-segments. (c) This "structural" partitioning transforms the Hamiltonian of the system into the block-diagonal form where H_{11} and H_{22} describe N-segments and M-segments, respectively. In this form, the negative eigenvalue theorem[45] can be used to calculate the integrated density of states I_N corresponding to the Schrödinger equation with this Hamiltonian and given N. The limit of this solution for $N \to \infty$ regenerates the integrated density of states I_∞ and the basic inequality (d) is proven to hold. Equation (5) is the direct consequence of this relation.

functional form.[45] Figure 6 summarizes the background information for the physical basis of the derivation of the final formula. According to this formalism the length dependence of rotational strength $R_{oj}(\lambda, n)$ is

$$R_{oj}(\lambda, n) = R_{oj}(\lambda, \infty)\left(1 + \sum_{i=1}^{s}\left(\frac{C_i(\lambda)}{N^i}\right.\right. \qquad (5)$$

Here, $R_{oj}(\lambda, \infty)$ is the rotational strength of a given transition for an infinitely long polymer and $C_i(\lambda)$ are adjustable parameters, usually determined experimentally. Equation (5) extends the applicability of this intensity correction for chiroptical methods other than ECD. It was shown, in a comparative study of ECD and VCD of proline oligopeptides with systematically varied length, that VCD requires third-order terms in Equation (5) in order to achieve acceptable fit of observed intensity variation in amide I vibration. For the same peptides, the ECD intensity length dependence was satisfactorily described only by the first-order term.[68]

For proteins, it is more appropriate to interpret N in Equation (5) as the average length of the backbone segments folded into a given type of secondary structure. As a consequence of this effect, the CD of two proteins might be different even if they have identical secondary structure compositions distributed over a different number of segments (see Figure 7).[69,70]

1.2 Advantages and Limitations

If the structural problem in hand is qualitative, then the sign of the experimentally observed dichroic bands in any frequency region contains usually all the information that is needed. Examples of such applications are the differentiation between enantiomers, detection of opposite chirality in biopolymer regular structures (e.g. of (*B*)- and (*Z*)-conformations of nucleic acids), or any observation of induced CD.

The ability to observe CD is governed by unique selection rules that are different from those for absorption. Although this, in principle, leads to an observation of forbidden absorption transitions in CD, the dissymmetry of biomolecules reduces the importance of this CD advantage.

In spectral bands that are combinations of overlapping components, for example the IR regions of amide I or II where the protein amide group vibrations absorb, the differences in the CD signs can facilitate the identification of such components. Such a qualitative identification of

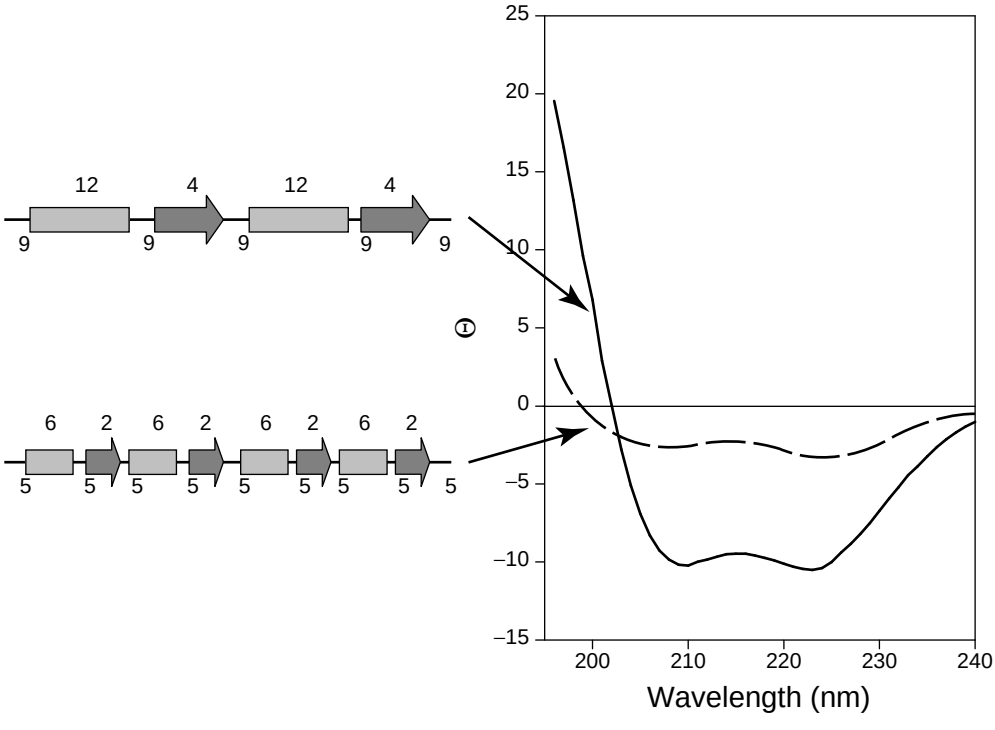

Figure 7 Demonstration of the effect of secondary structure segment length on the ECD spectra of proteins. Two proteins are shown with the same fractions of the secondary structure distributed differently over the secondary structure segments (rectangles, α-helices; arrows, β-strands; lines, loops; the numbers above segments indicate the number of amino acid residues in each segment). The predicted difference between the overall ECD spectra of the two proteins was generated using the characteristic helix, sheet, and ''other'' component spectra from the classical paper of Yang,[5] weighted in sum by the coefficients derived from Equation (5) with the experimentally derived parameters.[68]

a component that would be difficult to recognize in an ordinary absorption spectrum represents again the most effective application of CD spectroscopy. Attempts to utilize this qualitative advantage of CD quantitatively require deconvolving the CD spectral envelope into the signed component bands. This is a task that has perhaps even more serious algorithmic problems than the deconvolution of the corresponding absorbance spectra. The uncertainties in the determination of parameters of mutually canceling signed bands are serious and even the additional restrictions when CD and absorbance spectra are deconvolved simultaneously cannot significantly improve the situation.

Efficient methods of quantitative analysis of biomolecular CD spectra benefit primarily from the increased variability of the spectral bandshape given by the addition of sign to absorption bands. In this respect, the introduction of VCD was a major development because of the additional structural sensitivity of vibrational transitions. Besides their sign, the frequency at the maximum of dichroic bands in the IR region depends very specifically on the structure in general and on the conformation of biomolecules in particular. This facilitates, for example, better differentiation of helical and sheet spectral components of proteins by VCD than by ECD.[3,68,71–81]

Of all chiroptical methods, the ECD is the best established spectroscopic technique with commercially available instrumentation. In most biomolecular studies, ECD transitions are broad, heavily overlapped, and sensitive to relatively global structural features. The studied molecule undergoes a transition from a ground to an excited electronic state. ECD measurement typically requires about 2 mg or less of a biomolecular sample per experiment. The spectral region is accessible from 180 to 600 nm, which can be expanded to ~1000 nm with a photomultiplier change, which often requires a subsequent baseline realignment. The main experimental limitation for ECD measurement is the sample absorbance. It must be below 1.0 absorbance units for most CD spectrometers to ensure sufficient photon flux on the detector. Besides the oscillator strength of the studied biomolecular transitions, it is the choice of the solvent or buffer components that also interferes with the ECD detection below 220 nm. In a typical setting, the ECD measurement is relatively fast. It usually takes about 30 min to complete the experiment including the sample preparation and data pretreatment and storage. With a special attachment, ECD spectrometers can be used as selective detectors in chiral separations by high-performance liquid chromatography (HPLC).

VCD instrumentation is commercially available in the FTIR variant, and numerous dispersive instruments were built from commercial components in research laboratories.[3,4,61,71,77,82–84] In both variants, the details of the spectrometer optical design are very important. The differential absorption in the IR is 10–100 times smaller than the typical CD in the UV region, and the measured signal is, therefore, prone to a broad range of polarization artifacts. Transitions in the IR are generally well resolved and the studied molecules remain in the ground electronic state during the measurement. A VCD experiment typically requires concentrations of 50 mg in 2 mL while the practical sample volume needed for two or three measurements is about 100 μL. The concentrations of biomolecular samples measured in water solvents are even higher. Current VCD spectrometers can operate in the mid-IR spectral region (600–4000 cm^{-1}). The possibility of measurements in the other frequency regions requires a new design for the polarization optical elements and corresponding detector changes (see below). To obtain an acceptable S/N, the measurement is proportionally slower than in ECD, but this difference will become less important with further technical developments in VCD instrumentation. Absorption of the solvent, or buffer component, poses a more serious problem in VCD than in ECD. Absorption of water, which is the typical solvent in biomolecular studies, in the IR overlaps with the characteristic absorbance of proteins and nucleic acids. Deuteration of the solvent can open the needed spectral window but can also alter the vibration modes of the molecule involving exchangeable protons. This requires special attention, as incomplete deuteration in the solvent-inaccessible regions of biopolymer structures might complicate the VCD spectrum.[85] Under controlled experimental conditions, deuteration can be used as an additional factor in obtaining novel structural information.

2 HISTORY AND AVAILABLE INSTRUMENTATION

Differential absorption of RCPL and LCPL can be measured directly using the definition of CD in Equation (1) in an instrument that controls the state of the probing CPL. This principle is the method of choice in commercial CD spectrometers. Another physical observable that can be used to determine CD is the ellipticity of originally linearly polarized light after the interaction of its circularly polarized components with a circularly dichroic sample. The principal advantage of this approach compared with the previous method is its practically instantaneous response time, allowing measurements at the sub-microsecond level. This makes the method attractive for time-resolved CD spectral measurements even though it is technically more difficult.[4] Only the first technique will be discussed further.

It is customary to characterize the differential response of a chiral sample to LCPL and RCPL by the dissymmetry

factor $\Delta\varepsilon/\varepsilon = (\varepsilon_L - \varepsilon_R)/[(1/2)(\varepsilon_L + \varepsilon_R)]$. In a typical CD experiment, signals with dissymmetry factor of 10^{-4}–10^{-7} should be detected. To achieve an acceptable S/N, a proper averaging technique is needed. The signal to be averaged is generated by the high-frequency variation of the polarization state of the probing light from RCPL to LCPL. Various constructions of the polarization modulators were used for this purpose, for example Pockels cells, photoelastic modulators, and spinning quarter-wave plates.[4,86–89]

The photoelastic modulator occurs in most designs of CD spectrometers. It is a precisely dimensioned optical material, transparent for the studied light frequency range (typically quartz for ECD and CaF_2 or ZnSe for mid-IR regions, respectively), that is coupled to an alternating voltage-driven piezoelectric crystal. Periodic volume changes of the driver induce acoustically resonant changes of crystal optical properties. These in turn cause periodic variation in the retardation of the RCPL and LCPL components of the linearly polarized light passing through this system. For the modulator axis rotated in space by $\pi/4$ relative to the polarization plane of the incoming linearly polarized light, the output light polarization state oscillates between right and left circular polarization with the frequency of the driving voltage. A dichroic sample converts this polarization modulation into the detectable intensity modulation $I_{AC}(\nu)$. The light flux $I(\nu, t)$ after absorption by a sample is

$$I(\nu, t) = I_{DC}(\nu) + I_{AC}(\nu)\sin[\delta_{MAX}(\nu)\sin(w_M t)] \qquad (6)$$

The term $I_{DC}(\nu)$ is proportional to the absorbance of the sample, including the contribution from the instrument parts. The sine-of-sine function describes details of the time dependence of the state of the probing CPL ($\delta_{MAX}(\nu)$ is the maximum modulator retardation and w_M is the modulation frequency). Equation (6) can be rearranged into the following form that describes the change in CD with a photoelastic modulator by introducing Beer's law relating direct current (DC) and alternating current (AC) light flux signals to the input total light intensity (Equation 7):

$$\Delta\varepsilon(\nu) = \frac{g_{DC}}{g_{AC}}\frac{2}{2.303}2J_1[\delta_{MAX}(\nu)]\frac{I_{AC}(\nu)}{I_{DC}(\nu)} = K(\nu)\frac{I_{AC}(\nu)}{I_{DC}(\nu)} \qquad (7)$$

Here g_{AC} and g_{DC} are amplification gains for the two signals, respectively, and J_1 is the first-order Bessel function describing the fundamental frequency selection from the composite time dependence of the circular polarization (Equation 6) that is accomplished by the selective lock-in amplifier tuned to w_M. The optimization of the instrument constant $K(\nu)$ is performed at every frequency by the voltage U_λ derived from the monochromator wavelength that adjusts the retardation δ_{MAX}. The term $K(\nu)$ also includes the instrument function, which is normally determined by calibration.

2.1 Single Beam Electronic Circular Dichroism Spectrometer

Figure 8 shows the basic features of the ECD spectrometers available commercially. These use powerful light sources with intense radiation in the UV, which is the region of primary interest for biomolecular studies. The water-cooled high-pressure 250–450 W xenon discharge lamp is common. The light from the source is collected with a mirror and monochromatized. In the classical design, monochromators use a prism simultaneously as a dispersive and polarizing optical element. Only half of the total light intensity is used after the linear polarization of the probing light (the extraordinary ray is blocked). The linearly polarized and monochromatized light is converted into the alternatively left and right circularly polarized forms by a quartz photoelastic modulator before it enters the sample cell. The output light intensity is detected by the photomultiplier and the resulting signal is processed after preamplification by the detection electronics (lock-in) tuned to the photoelastic modulator carrier frequency. To facilitate processing, the I_{DC} part of the signal (Equation 6) is usually kept constant by real-time variation of the photomultiplier gain g_{DC}.

2.2 Dual Beam Electronic Circular Dichroism Spectrometer

An alternative design for a CD instrument, introduced by Olis (Bogart, GA), is shown in Figure 9. After the

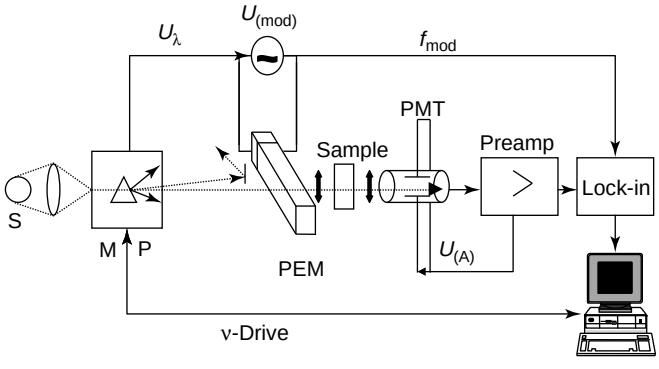

Figure 8 The generic spectrometer for ECD measurement. S, light source; MP, polarizing prism monochromator; PEM, photoelastic modulator; PMT, photomultiplier; U_λ, voltage ramp derived from the monochromator state and used to adjust the PEM modulation $U_{(mod)}$ to $\lambda/4$; f_{mod}, reference signal with the frequency of the PEM oscillation; $U_{(A)}$, voltage derived from the I_{DC} part of the photoelectric signal of PMT that is used to control the gain of PMT; ν-drive, signal for stepping motor of the monochromator.

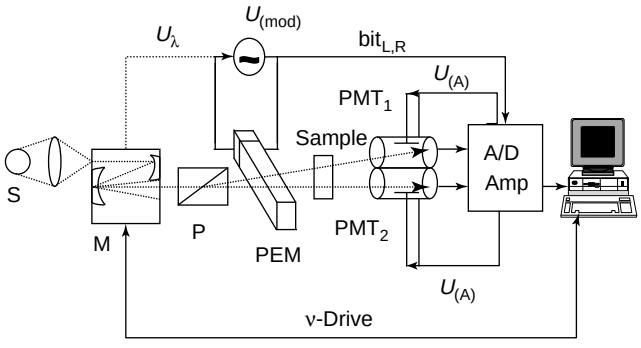

Figure 9 The dual beam spectrometer for ECD measurement. S, light source; M, grating monochromator; P, polarizer; PEM, photoelastic modulator; PMT_1 and PMT_2, photomultipliers; A/D amp, conversion of the analog signal into the digital form; bit_{LR}, information about the modulation state of the light used to tag the converted photocurrents; U_λ, optional voltage ramp derived from the monochromator state and used to adjust the PEM modulation $U_{(mod)}$ to λ/4; ν-drive, signal for stepping motor of the monochromator; $U_{(A)}$, voltage derived from the I_{DC} part of the photoelectric signal of PMT that is used to control the gain of PMT as needed for an effective analog-to-digital conversion.

wavelength selection by a grating monochromator, the photoelastic modulator modulates both ordinary and extraordinary linearly polarized light beams emerging from the MgF_2 Rochon polarizer. These two beams are at any given time in the oppositely circularly (or elliptically) polarized states through the perpendicular relative orientation of their linear polarization planes. In particular, when one beam is RCPL, the other one is LCPL and vice versa (see Figure 10a). The optical design ensures that both beams are passed through the sample cell, where the polarization modulation is converted into intensity modulation. The signals I_{D1} and I_{D2} that are independently detected by two photomultipliers can be described as

$$\frac{I_{D1}}{I_{01}} = \frac{1+Q(\nu)}{2} e^{\varepsilon(\nu)}(1+\Delta\varepsilon(\nu)\sin(\delta_0\sin(w_m t))) \quad (8a)$$

$$\frac{I_{D2}}{I_{02}} = \frac{1-Q(\nu)}{2} e^{\varepsilon(\nu)}(1-\Delta\varepsilon(\nu)\sin(\delta_0\sin(w_m t))) \quad (8b)$$

where I_{01} and I_{02} are input total intensities measured by detector 1 and 2, respectively; $Q(\nu)$ is Stokes parameter describing the effect of beam linear polarization, and $\varepsilon(\nu)$ and $\Delta\varepsilon(\nu)$ are overall absorbance and CD, respectively. Other symbols have the same meaning as in Equation (6).

To maximize the advantage of this design, which provides simultaneous information about the CD of the sample at any time point, the averaging strategy is modified.[90] The signals from both photodetectors are converted into digital form by two fast 14-bit analog-to-digital converter (ADC) circuits. Their 400 kHz conversion rate yields eight data points per cycle at 50 kHz modulator frequency. The most significant bit

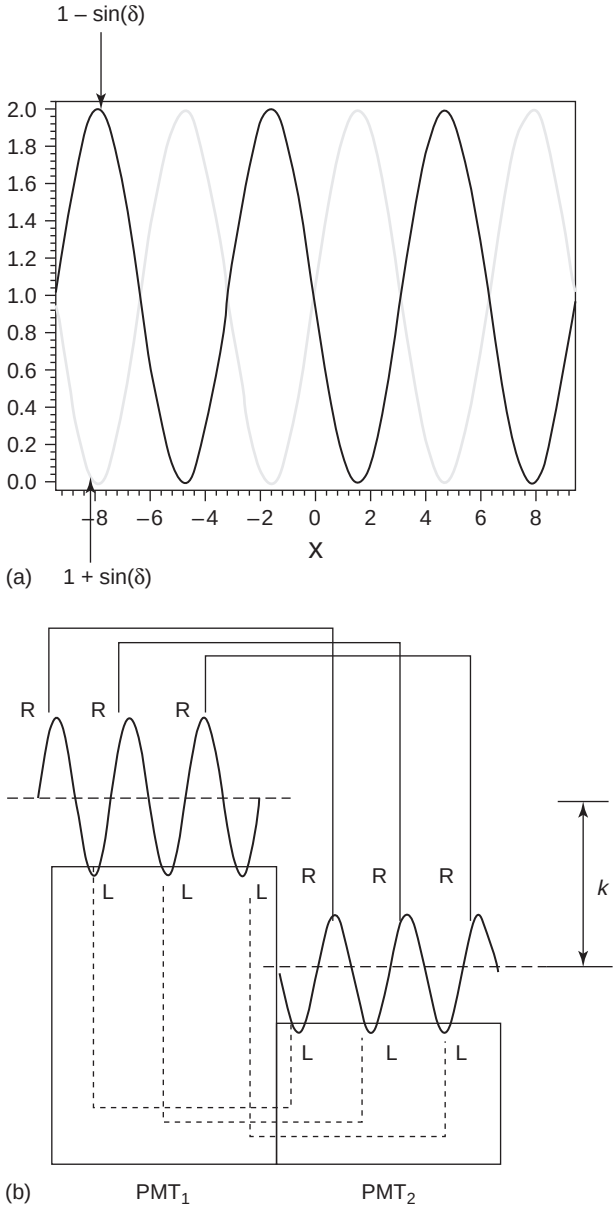

Figure 10 (a) Time dependence of the circular polarization modulation in the ordinary $(1+\sin(\delta))$ and extraordinary $(1-\sin(\delta))$ light beams of the dual beam ECD spectrometer. (b) The differences in the two optical channels of the dual beam ECD spectrometer (a simple multiplicative difference between the signals of photomultipliers 1 and 2) is compensated by the spectrometer-averaging algorithm (see text for details).

of each record is set to 1 or 0, indicating the state of the probing light (LCPL or RCPL) derived from the reference frequency of the photoelastic modulator ($bit_{L,R}$, Figure 9). The signals I_{01} and I_{02} are first used to normalize the intensities measured by individual detectors at each wavelength. The records corresponding to the RCPL and LCPL probing light beams are then averaged independently on the detector. The algorithm

uses only their digital tags for L and R recognition; consequently one-half of the data come from channel 1 and the other half from channel 2. The $Q(\nu)$ dependency in Equation (8) is effectively canceled by this averaging mode. CD is then calculated (Equation 9) using the ratio of determined RCPL and LCPL record averages

$$I_{R/L} = \frac{e^{-\varepsilon(\nu)}}{2}\left(1 \pm \Delta\varepsilon(\nu)\frac{2}{T}\int_0^{T/2}\sin(\delta_0 \sin(w_m t))\,dt\right)$$
$$= \frac{e^{-\varepsilon(\nu)}}{2}(1 \pm \Delta\varepsilon(\nu)\overline{\sin(\delta)}) \quad (9)$$

Here, $\overline{\sin(\delta)}$ is the frequency-dependent constant defined as the mean value of the functional description of modulation function. The final expression (Equation 10) describing this variant of CD assessment is derived from Equation (9) using the approximate formula for logarithms of small $\Delta\varepsilon(\nu)$ values:

$$\Delta\varepsilon(\nu) = \frac{1}{0.8686\overline{\sin(\delta)}}\log\frac{g_R}{g_L}\frac{I_{L0}}{I_{R0}} = K(\nu)\log\frac{I_L}{I_R} \quad (10)$$

There are several theoretical advantages of this CD measurement mode. First, the cancellation of the linear dichroism term $Q(\nu)$ justifies the use of the grating monochromator. The effect of the wavelength-dependent linear polarization of the probing light after the monochromatization by a grating is minimized. Second, the averaging mode ensures relative insensitivity of the CD to the fluctuations of the two photodetector signals. For example, if the signal level detected by one photomultiplier drops by a factor $k < 1.0$ (see Figure 10b), the averaging mode compensates for this difference in the two channels in the following way:

$$\Delta\varepsilon(\nu) = K(\nu)\log\frac{\sum_{i=1}^{N} I_{Li1} + k\sum_{i=1}^{N} I_{Li2}}{\sum_{i=1}^{N} I_{Ri1} + k\sum_{i=1}^{N} I_{Ri2}}$$
$$= K(\nu)\log\frac{(1+k)\sum_{i=1}^{N} I_{Li}}{(1+k)\sum_{i=1}^{N} I_{Ri}}$$
$$= K(\nu)\log\frac{\sum_{i=1}^{N} I_{Li}}{\sum_{i=1}^{N} I_{Ri}} = K(\nu)\log\frac{I_L}{I_R} \quad (11)$$

Similar argument holds even for more complex variations of signals in both channels. In these cases, the averaging mode effectively replaces the functional description of the fluctuation by the numerically determined mean values of that function. Constant values cancel out in the way described above.

Utilization of both ordinary and extraordinary light beams results in more effective use of the light source energy. There is, nevertheless, still restriction on the sample absorbance. The changes of the light intensity by the sample absorbance are compensated by the feedback variation of the high voltage on both photomultipliers. This is necessary for optimal use of the full digital range of the ADC. This instrument also offers selection between the classical monochromator and the rapid scanning, subtractive, double-grating monochromator with moving intermediate slit. In the classical variant, the maximal retardation of the photoelastic modulator is optimized for the output wavelength. In the rapid scanning mode, the wavelength change is too fast for the modulator to follow. The retardation is, therefore, set to be optimal for median wavelength of the scanned interval. The CD intensity is computationally corrected using a calibration curve as described below for VCD.

2.3 Dispersive Vibrational Circular Dichroism Spectrometer

There are two experimental set-ups for a measurement of VCD, the dispersive and Fourier transform (FT) instruments. In the former, the basic experimental scheme is the same in many aspects as that described for ECD spectrometers. A change of the frequency region dictates a replacement of the light source, the materials of the optical elements, and the detectors for operation in the IR spectral region. In the latter the interferometer of the FT instrument is used as the source of light. The interferometer introduces a modulation of the intensity of low frequencies that encodes the optical frequencies. The FT spectrometer is then extended by optical elements for high-frequency polarization modulation and by the additional electronics necessary for the detection of the weak-intensity modulation at the high frequency that is attributed to sample VCD.

The scheme of a generic dispersive VCD spectrometer is shown in Figure 11. The light sources that are used in practice include a high-temperature Nernst glower[64,85,87] or a carbon rod heated in an inert atmosphere to ~2400 K.[3,74,75] After the light has passed through the collection optics, the light intensity is modulated at low frequency (~100–200 Hz) by a mechanical chopper to improve the detection of signal I_{DC} corresponding to the sample absorbance. A grating monochromator, wire grid polarizer (typically on BaF_2 substrate), and photoelastic modulator with either CaF_2 or ZnSe crystals are used to prepare the probing light with high-frequency (~32 kHz)

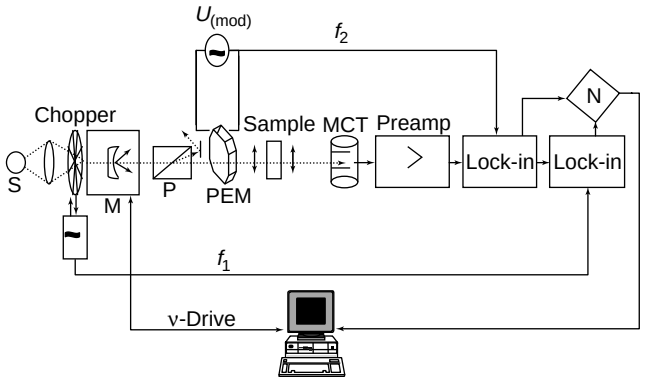

Figure 11 Spectrometer for VCD measurement. S, IR light source; M, grating monochromator; P, grid polarizer; PEM, photoelastic modulator; MCT, semiconductor photodetector; N, dynamic normalization circuit; f_1 and f_2, chopper and PEM modulation frequencies, used as a reference signal for the dual lock-in detection; $U_{(mod)}$, PEM modulation voltage; ν-drive, signal for the stepping motor of the monochromator.

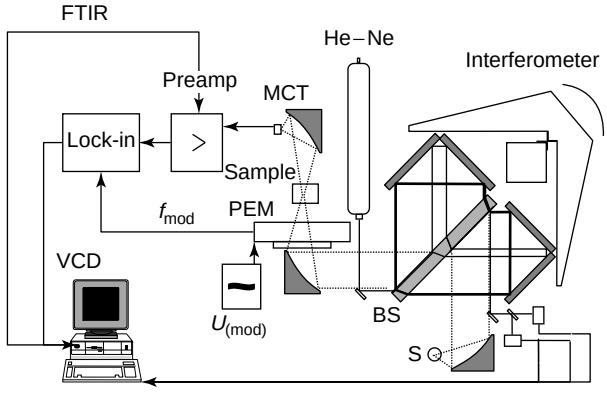

Figure 12 The FT spectrometer for VCD measurement. S, IR light source; BS, beam splitter; He–Ne, laser; PEM, polarizer/photoelastic modulator combination; MCT, semiconductor photodetector; $U_{(mod)}$, PEM modulation voltage; f_{mod}, PEM modulation frequency, used as reference signal for the lock-in detection; FTIR, low-frequency modulation of the photodetector signal by the interferometer.

modulation of its circularly polarized state. These materials have operation optima at $\sim$1200 cm^{-1} and $\sim$1000 cm^{-1} respectively and cutoff lower frequencies at $\sim$600 cm^{-1}. Barium fluoride is a sample cell material that is suitable for the water-based solutions used in biomolecular studies. The light is detected by the semiconductor detector, which is cooled by liquid nitrogen. Mercury cadmium telluride is suitable for mid-IR to $\sim$400 cm^{-1} whereas InSb is used for the near-IR. After amplification, the signals carried by low- and high-modulation frequencies are separated by two lock-ins. One amplifier operates at the reference frequency of the photoelastic modulator, the other at the chopper frequency. The two signals representing the transmission intensity and the polarization modulation intensity are processed in the dynamic normalization circuit. Alternatively, they can be separately put through an ADC and their ratio is then calculated numerically by the processing computer. These procedures replace the dynamic normalization by the photomultiplier gain in the ECD, which is impractical for semiconductor IR photodetectors. The ratio of AC and DC signals is scaled by the calibration factor $K(\nu)$ determined from the calibration procedure (see below), and the resulting VCD spectrum is stored.

2.4 Fourier Transform Vibrational Circular Dichroism Spectrometer

A generic FT spectrometer for VCD measurement is shown in Figure 12. A light from the IR source passes through the interferometer, which encodes the frequency information into the (low-frequency) intensity modulation. This light is then modulated at 30–50 kHz in a polarizer–photoelastic modulator optical system that is identical to that in the dispersive VCD spectrometer. After passing through the sample, the light is focused on the large-area semiconductor detector cooled by liquid nitrogen. Use of a lens for this purpose allows covering and integrating the signal uniformly over the photodetector active area. This reduces the sensitivity of the spectrometer to optical artifacts.[91] A signal from the photodetector is amplified and separated into the low- and high-frequency components. This step requires special care (use of filters and optimal gain selection) to avoid saturation of the spectrometer detection system and overloading of the processing electronics. A digitized low-frequency signal is processed by the standard FT routines to provide the I_{DC} part of Equation (7). The high-frequency signal is further selectively amplified by the lock-in that is tuned to the photoelastic modulator frequency. The resulting signal is fed into the special FT routines that determine the I_{AC} component of Equation (7). The interferogram defined the I_{AC} signal can be understood as the difference of the two interferograms corresponding to the absorbance spectra for oppositely CPL. Standard routines for the fast FT localize the zero path difference (ZPD) point of the interferogram using the unique intensity of the center burst. This causes problems for the difference interferograms representing VCD, which have the center burst area generally weak. Either an experimental or computational correction of the ambiguity of ZPD determination is needed.[92] In the FT calculation, the transferred phase algorithm is used. The required phase function can be determined experimentally using the stressed optical (ZnSe) plate instead of the sample and stored.[3] After the transformation of interferograms, the computer calculates the

VCD spectrum as the ratio of the I_{AC} and I_{DC} signals and scales it by the proper calibration factor (see below) corresponding to Equation (7). A variant of this ''fast-scan'' spectrometer was also developed for the step-scan FTIR spectrometers.[93]

2.5 Developmental Designs

2.5.1 Instruments and Methods

CD spectra can also be measured by monitoring the difference in the luminescence intensity that is emitted from a chiral sample after excitation by CPL. This technique is called FDCD; Figure 13 shows a typical design scheme.[25] In the FDCD apparatus, the excitation monochromator, polarizer, and photoelastic modulator are used to prepare light with alternating circularly polarized modulation. This light excites the fluorescence of the chiral molecule. The detection end of the optical system, where, optionally, the emission wavelength can be selected by the emission monochromator, collects emitted light. The fluorescence intensity is detected by the photomultiplier. After amplification of the photosignal by a lock-in that is in phase with the photoelastic modulator frequency, I_{AC} and I_{DC} components of the signal are ratioed and a CD spectrum is obtained. There is no signal outside the absorption band in all luminescence-based methods where, by comparison, absorbance techniques have high light intensity throughput. Modifications of the detection scheme are, therefore, needed to handle this trend of the I_{DC} signal.[6]

The potential advantages of FDCD are the possibility of measuring CD spectra for samples that are heavily light scattering and for selective measurements of CD spectra of components of a mixture, provided that there are sufficient spectral separation and a high quantum yield of their fluorescence. It is essential for the interpretation of FDCD spectra that they still represent the CD spectrum of the ground state of the studied molecule.

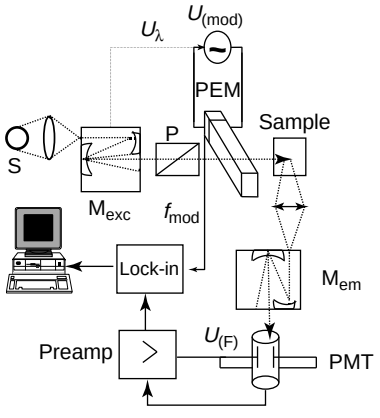

Figure 13 Spectrometer for FDCD spectroscopy. S, excitation light source; M_{exc}, excitation monochromator; P, polarizer; PEM, photoelastic modulator; M_{em}, emission monochromator; PMT, photomultiplier; U_λ, optional voltage ramp derived from the monochromator state and used to adjust the PEM modulation $U_{(mod)}$ to $\lambda/4$; f_{mod}, reference signal with the frequency of the PEM oscillation; $U_{(F)}$, signal used to control the gain of the photomultiplier.

The conformation of the excited state of a molecule can be examined using the spectral dependence of emitted CPL (Figure 14).[6] This instrument detects the difference in magnitudes of oppositely CPL components that are emitted by the chiral molecule. The photoelastic modulator and polarizer have an inverted role in this spectrometer compared with previous designs described. Here the photoelastic modulator converts circularly polarized components of the luminescence into modulated linear polarization of the emitted light. The polarizer in its proper orientation then converts this polarization into the intensity modulation that is processed by the lock-in amplification scheme in phase with the photoelastic modulator frequency. Alternatively, the dual modulation scheme with chopper, additional lock-in, and dynamic or computational normalization of absorbance and CD signals can be used.

CPL emission spectroscopy is the unique tool for excited-state structural and dynamic studies. Other promising areas for the applications of this technique are the enantioselective quenching of racemate luminescence by chiral quenchers and applications benefiting from the selectivity and unique physical mechanisms responsible for the luminescence phenomenon.

2.5.2 Spectrometer Components

The principal motivation for designing novel optical components for CD spectrometers is the attempt to broaden the spectral region in which CD spectra can be measured.

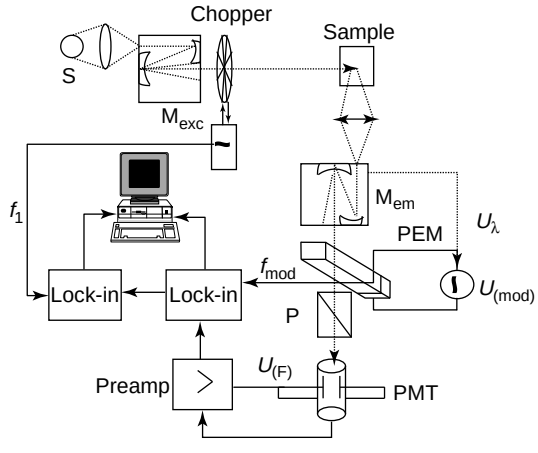

Figure 14 Spectrometer for CPL spectroscopy. Abbreviations as in Figure 13; f_1, reference signal with the chopper frequency.

The far-IR region (below 600 cm^{-1}) is one of the targets of these innovations. The 600 cm^{-1} low-frequency limit of current VCD spectrometers is principally a consequence of the combination of optical and mechanical properties of the IR-transparent materials that can be used as crystals in the photoelastic modulator. Two alternative techniques were tested as a basis for the replacement of the photoelastic modulator: the spinning quarter-wave plate polarization modulator and the polarizing interferometer. The principal scheme of the polarizing interferometer is shown in Figure 15.[98]

The key part of this design is the beam splitter, which is made of properly transparent optical material and covered by a polarizing grid. This addition makes the beam splitter transparent only for one light linear polarization state (S in Figure 15) while the other polarization, P, is reflected. Interferometer mirrors acts as the ''retro-mirrors'' that change the original polarization state of the light returned by them. Figure 15 explains the role of the polarizing beam splitter in bringing two perpendicularly linearly polarized beams to the coherent combination at the interferometer output. Periodic translation of the movable mirror introduces the phase retardation of these two polarized beams. Circular polarization of the monochromatic light is thus generated when the path difference equals $(4k+1)\lambda/2$ (LCPL) and $(4k+3)\lambda/2$ (RCPL), $k = 0, 1 \ldots$. The practical application requires a signal-processing method that properly selects interferogram components corresponding to these frequencies and an exceptionally stable spectrometer design.[61]

Another design of the polarization modulator brings an additional degree of flexibility beyond the frequency region extension. Its construction utilizes the simple fact that rotation of the quarter-wave plate perpendicularly to the path of the linearly polarized light produces variation in the circular polarization state twice within one full rotation cycle.[86,87] The mechanical design requires stable and virtually friction-free rotation of the plate with a frequency in the kilohertz range. This was achieved

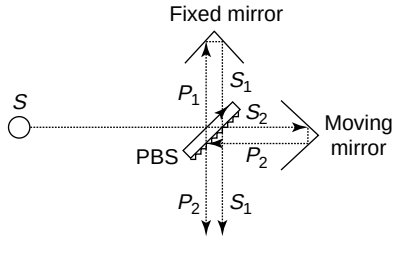

Figure 15 The polarizing interferometer. S, light source; PBS, beam splitter with polarizing grid that makes it transparent only for S-polarized state of the light; S_1, S_2, P_1, and P_2, polarization states of the light after interaction with the interferometer elements. Note the change of the polarization state at the mirrors.

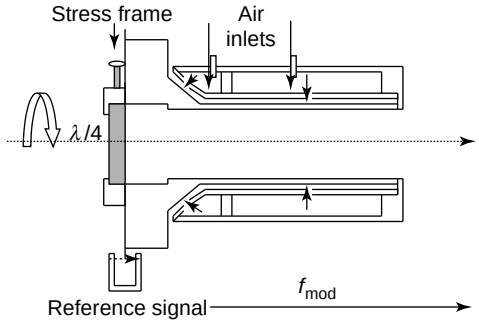

Figure 16 The spinning quarter-wave plate polarizer. Optical material of the plate is stressed in the stress frame, forming the quarter-wave retarder λ/4. Propeller gas is used to spin the plate holder. The reference signal f_{mod} is derived from the light-emitting diode – PIN photodiode pair and chopper-like blade on the spinning plate holder.

by modification of the design of the air-driven turbine that is used to spin the sample in NMR spectrometers (Figure 16). In addition to the possibility of varying the optical materials that can be used as the quarter-wave plate, this design offers the unique possibility of continuous variation of the modulation frequency simply by changing the pressure of the propeller gas.[86]

3 GENERAL EXPERIMENTAL ASPECTS

3.1 Instrument Calibration Procedures

The CD spectrometer needs calibration in all the above modifications. This is equivalent to the determination of the frequency-dependent constant $K(\nu)$ in Equations (7) and (10). Measurement of a standard sample with known dichroic properties is one alternative that is used for this purpose. A water solution of d-10-camphorsulfonic acid is now generally accepted as the calibration standard for classical ECD spectrometers. The protocol for reference sample purification and solution preparation has been summarized and critically reviewed by Yang et al.[5]

The optical properties of the multiple-wave birefringent plate (mechanically stressed ZnSe crystal) followed by the linear polarizer are used for the calibration of VCD spectrometers.[3,4,71,86] With this element replacing the sample, an intense CD signal is measured that can be described as

$$\frac{I_{AC}(\nu)}{I_{DC}(\nu)} = \frac{\pm 2J_1[\delta_{MAX}(\nu)]\sin[\delta_{BP}(\nu)]g_{AC}}{1 \pm J_0[\delta_{MAX}(\nu)]\cos[\delta_{BP}(\nu)]g_{DC}} \qquad (12)$$

with symbols having the same meaning as in Equation (7) and with different signs for different relative orientations of the birefringent plate and the polarizer. The points of equal intensity can be found for the pairs of formulae with proper choice of signs in Equation (12). As can be proved

by algebraic manipulation of the above equation together with Equation (7), the intensities of the calibrating device CD at these nonzero crossings are equal to the respective values of the $K(\nu)$. Practically, the intensities of the calibration signal at the midpoint between any two of its zero-crossings can be used as the $K(\nu)$. This follows from analysis of Equation (12) and eliminates the need for readjustment of the calibrating device. The same procedure is applicable for both dispersive and FTIR CD instruments. The advantage of this approach over the single- or even two-point calibration with a standard solution is that it can provide the calibration values as a function of wavelength. Calibration of a different kind is needed for the rapid-scanning dual beam ECD spectrometer where the maximum retardation of the photoelastic modulator cannot be wavelength-adjusted during the fast spectrum scan. The CD intensity has to be corrected in this case for the imperfections in the polarization state of the probing light over the measured frequency interval. Because of the described unique character of the data processing, no intensity calibration in the classical sense is needed for this spectrometer in either normal or rapid-scanning operational mode.

3.2 Sampling Techniques

There are numerous instrumental and practical aspects of ECD measurement that might affect the quality of the measured spectra. The lifetime of the light source is limited typically to about 1000–2500 h of operation. Degradation of the lamp output results in lower S/N and reduction in accessibility of the shortest wavelengths. It is recommended that the light throughput of every ECD spectrometer is continuously monitored in the well-defined state. Using the readings of the photomultiplier high voltage after the initialization period, which is proportional to the light throughput I_{DC} at defined wavelength, can conveniently do this. The mirror that collects the light from the lamp must be checked periodically and re-coated or replaced if necessary, as the harsh conditions in the lamp housing accelerate its degradation.

The optical system of the spectrometer should be flushed with dried nitrogen to facilitate measurements in UV and to prevent degradation of optical surfaces. UV coating of lenses and mirrors in the optical part of the instrument has become standard because it improves the instrument performance at short wavelengths.

Periodic confirmation of the accuracy of the wavelength reading with a reference sample is good practice. Sample cells, besides the normal protocol for measurement in UV region, should be purchased as strain-free to avoid polarization artifacts. The cleaning of the cell is also important for measurements at the highest instrument sensitivities ($\Delta A \sim 10^{-7}$). Highly volatile solvents (e.g. methanol) should not be used for cell washing as the heat of evaporation induces measurable strain in the quartz and the cell baseline is distorted. This alteration might persist for several hours or eventually become irreversible. The adsorption of solute on the cell surfaces is often significant for biopolymer samples and requires thorough cleaning of the cells in chromosulfuric acid or for largely cationic solutions in basic solutions. Checking the pH of the solution used to wash the cell after this treatment is a surprisingly sensitive indicator of the completeness of the cleaning process.

As the ECD spectrometer operates in the single beam mode, the position and orientation of the cell during the baseline and sample scan should be identical. Dependence of the ECD spectrum shape on the distance of the cell from the photomultiplier might indicate light scattering on the aggregates of the solute. A combination of the sample concentration and the cell pathlength is used to adjust the sample absorbance to 0.6–0.8. Artifacts related to the low light intensity and high gain on the photomultiplier are observed for higher absorbances.

As the instrumentation for VCD measurements is less standardized, the experimental protocols depend on the particular spectrometer. In the FTIR instruments, the saturation of the sensitive semiconductor detectors by the high light throughput might require use of optical filters, which reduces the multiplex advantage of this technique.

The successful measurement of VCD of biological samples requires careful choice of the solvent and cell window material and the cell construction. Windows of BaF_2 separated by Teflon or Mylar spacers and either demountable or refillable cells have been used for protein, peptide, and nucleic acid VCD studies.[94–97] The absorption of water or D_2O, which is a common solvent for these studies, dictates the small pathlengths (50 μm for D_2O and 6 μm for H_2O solutions). Demountable cells require typically 20–30 μL of 50 mg mL^{-1} concentrated protein solution in D_2O. Refillable cells are used for measurements with water solutions (more than 100 mg mL^{-1} protein is needed for separation of the solvent and protein signals) and may require proportionally larger sample volume. Demountable cells without spacers have been used.[101]

Baselines for VCD measurements are generally recorded under experimental conditions identical to those of the sample with a solvent-filled cell. Alternatively, a solution of poly(L,D)-lysine with the absorbance at 1650 cm^{-1} adjusted to match that of the measured sample can be used to minimize the impact of the absorption-related artifacts in protein VCD studies.[99,100] Simultaneous measurements of IR absorbance of the VCD sample allow normalization of the measured VCD

spectra. Absorbance of the sample at the maximum of the corresponding IR band or the area under the absorption spectrum is used to compensate for uncertain pathlength and concentration of the protein. This method assumes that the extinction coefficients for different molecules are identical and is, therefore, only an approximation in the absolute sense.

4 STRUCTURAL INTERPRETATION OF CIRCULAR DICHROISM SPECTRA OF BIOMOLECULES

4.1 Quantitative Empirical Methods

The earlier part of this chapter has discussed structural features of biomolecules and physical mechanisms that can generate observable changes in CD spectra in all frequency regions. The inverse problem, methods to derive the maximum quantitative structural information from CD spectra, will be critically assessed using as an example the empirical methods of CD spectral analysis of proteins. Many aspects of protein spectroscopic structural studies can be generalized and applied in the investigation of other biomolecules.

The key steps of the structure encoding and decoding in CD are shown in Figure 17. The molecular structure of any molecule is transformed into a CD spectrum through its interaction with the (polarized) light in a spectrometer. Details of the molecular structure are necessarily reduced in this step to the level given by the structural sensitivity of the chiroptical technique used. Without the quantitative theoretical model, necessary structure–spectra relationships cannot be determined by the analysis of a single spectrum. Empirical methods, therefore, use series of spectra measured for a set of reference molecules with known structures (e.g. determined by X-ray crystallography). Variation of the spectral bandshape that represents various aspects of molecular structure in the set of reference experimental spectra and the known structure of the reference molecules then constitute the input to the analysis. As a direct consequence of the limited structural sensitivity of the optical spectra, numerous structural features are necessarily convoluted and unresolved in the spectral representation. The goal of an empirical analysis method is, therefore, twofold. First, the structural features that induce observable and systematic changes of the spectral bandshape should be identified in the reference set of molecular structures. Second, the variability of the reference spectra should be parsed into components that correlate significantly with the trends of structural variation identified by the previous step. This correlation should be finally described in a mathematical form that is instrumental in predicting unknown molecular structures.

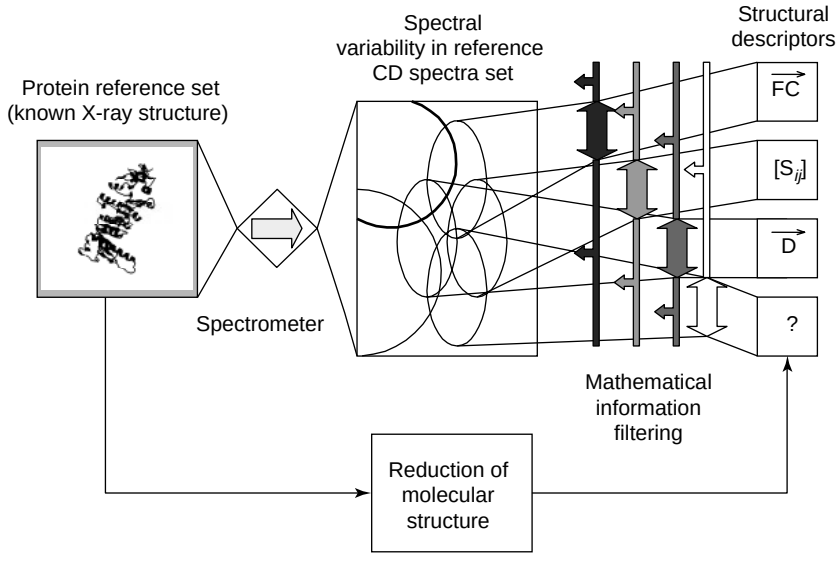

Figure 17 The steps in a selective empirical analysis of the informational content in biomolecular spectra. A representative set of reference molecules with known structure are projected through the selected spectroscopic method. The interdependencies in the reference spectra are analyzed and the bandshape variability in the set is defined by quantitative descriptors. An algorithm that reduces details of the reference molecular structures to a level compatible with the structural sensitivity of the spectroscopic method and quantifies them in terms of the structural descriptor components can be derived. An optimal and selective mathematical algorithm is finally developed to predict the various aspects of molecular structure from the spectral variability in the reference spectra and multiple structural descriptors which are derived from the known molecular structure of the reference molecules that can be related to different parts of the spectral variability in the reference set.

Spectra–structure correlation is developed by relating quantitative descriptors of spectral variability to quantitative descriptors of structural variability in the reference set. These two representations of the molecular structure should have a comparable level of resolution. Cartesian coordinates of reference molecule atoms need to be transformed into quantitative structural descriptors that represent the reduction of the atomic details to the spectroscopic resolution level. One example of such structural reduction is the determination of a protein secondary structure and subsequent calculation of its SSF.[101–106]

The SSF was historically considered as the only quantitative structural parameter that is extractable from ECD spectra of proteins.[107,108] Recent results showed that this assumption is too restrictive: there are structural features that go beyond the SSF and can be successfully predicted from CD spectra in various regions.[67,109,112] The following summary is formulated nontraditionally so that these new results can be incorporated into the general formulation of spectral analytical methods.

Any empirical interpretive method formulates a quantitative and selective (mathematical) transformation between the spectral variability and the variability of the chosen structural descriptor (Figure 17). The important first step in this formulation is to find the minimal number of independent components that are needed for a complete description of the spectral variability in the reference set. This information prevents overinterpretation of the experimental data and – more importantly – facilitates the decomposition of the experimental spectra into logical subsets, which are called spectral components, bands, or subspectra. Quantification of their contributions to the analyzed spectra is one of many ways for describing the spectral bandshapes numerically. These components can be also combined in various ways in the optimization of the correlation to structural descriptors.

Application of the same principle – determination of the number of independent components in the set of structural descriptors for the reference protein set – helps to reveal any interdependence between their values. The analysis can be simplified and additional criteria for the evaluation of the accuracy and precision of the predictions can be obtained, if any such relationship exists. The next section summarizes methods and examples suitable for such a development.

4.2 Analysis of the Reference Set Properties Using Principal Components

If we organize reference spectra column-wise into a matrix $[\Theta_i(\lambda)]$ with rows corresponding to frequencies or wavelengths (λ) and columns (i) attributed to the individual molecules (e.g. proteins) of the reference set, then the number of independent components in our spectral series is equal to the rank p of that matrix. Any mathematical method of rank determination can be used to find the p. Principal component analysis (PCA), one of the statistical methods of variance analysis, is the standard algorithm for this purpose.[110,111] This method was utilized not only for protein CD spectra analysis but also in numerous other spectroscopic applications.[112–115] PCA decomposes the reference spectra set into a linear combination of p orthogonal (linearly independent) basis functions (subspectra). In the mathematical formulation of PCA, the spectral intensities of the analyzed data matrix $[\Theta_i(\lambda)]$ are considered as variance from the common (zero) mean value. PCA basis functions are defined in this variance space by the requirement that the projections of the spectral intensities into them are maximal, that the basis functions are orthogonal, and that the correlations between the experimental spectra are preserved after the decomposition. This is the optimization problem with multiple boundary conditions that can be solved using the method of Lagrange multipliers. Its solution leads to the secular equation and determination of eigenvalues and eigenvectors of the correlation or covariance matrix of the experimental spectra.

With this understanding of the underlying statistical background, further development is straightforward matrix algebra. We have a set of N linear equations that define the decomposition of any CD spectrum $\Theta_i(\lambda)$ ($i = 1, \ldots, N$) from the reference set into the linear combination of p common independent subspectra $\phi_j(\lambda)$ ($j = 1, \ldots, p$):

$$\Theta_i(\lambda) = \sum_{j=1}^{p} C_{ij}\phi_j(\lambda) \tag{13}$$

To define the spectral correlation matrix that is needed for PCA algorithm, Equation (13) can be rewritten as

$$\Theta_i(\lambda) = A_i \sum_{j=1}^{p} c_{ij}\phi_j(\lambda) \tag{14}$$

where

$$A_i = \overline{\int_{\lambda_1}^{\lambda_2} \Theta_i^2(\lambda)\,\mathrm{d}\lambda} \tag{15}$$

is the normalization constant corresponding to the area of each decomposed spectrum.

The correlation matrix $[R]$ of the experimental spectra that is needed to solve the secular problem defined above can be calculated as

$$[R] = [w_i(\lambda)]^{\mathrm{T}}[w_i(\lambda)] \tag{16}$$

where

$$w_i(\lambda) = \frac{1}{A_i}\Theta_i(\lambda) = \sum_{j=1}^{p} c_{ij}\phi_j(\lambda) \tag{17}$$

and T is the transpose operator. The elements r_{ij} of the correlation matrix $[R]$ are, in fact, the numerically calculated values of the overlap integrals for all spectral pairs in the reference set:

$$r_{ij} = \int_{\lambda_1}^{\lambda_2} w_i(\lambda) w_j(\lambda)\, d\lambda \tag{18}$$

Any numerical algorithm can be substituted for the simplest method determining this integral using Equation (16). Principal components of $[\Theta_i(\lambda)]$ with the desired properties are found by diagonalization of matrix $[R]$ that yields eigenvector matrix $[q]$ and diagonal eigenvalue matrix $\Lambda_{ij}\delta_{ij}$

$$[q]^{-1}[R][q] = [\Lambda_{ij}\delta_{ij}] \tag{19}$$

where $[R]$ is the real, symmetric matrix and, therefore, $[q]^{-1} = [q]^{\mathrm{T}}$. Decomposition of Equation (17) can be found by multiplying the normalized experimental data matrix by unit matrix $[q][q]^{\mathrm{T}}$

$$[w_i(\lambda)] = [w_i(\lambda)][q][q]^{\mathrm{T}} \tag{20}$$

Comparison of Equations (17) and (20) defines the identities in Equations (21) and (22):

$$[\phi_j(\lambda)] = [w_i(\lambda)][q] \tag{21}$$

$$[c_{ij}] = [q]^{\mathrm{T}} \tag{22}$$

The series of Equations (13–19) mathematically transform $[\Theta_i(\lambda)]$ into the diagonal matrix of eigenvalues $[\Lambda_{ij}\delta_{ij}]$. It is known that these operations preserve the rank p of the analyzed set of CD spectra, which is thus defined as the number of nonzero eigenvalues in $[\Lambda_{ij}\delta_{ij}]$. Such a simple result can be obtained only for ideal data without noise. All eigenvalues are nonzero for real spectra with noise, and p is then defined as the minimal number of subspectra needed to reconstruct experimental spectra using Equation (14). Numerous criteria can be used for the determination of p. An excellent analysis of this problem studying the impact of experimental noise on the results of PCA is available.[110] For spectroscopic applications, it is practical to calculate the differences between the experimental and reconstructed spectra for a gradually increasing number of subspectra. Ordering of the subspectra according to their importance for the description of the reference set spectral features is needed for that purpose. Such an ordering is possible using the fact that according to the PCA statistical model the calculated eigenvalues are proportional to the fraction of the total variance accounted for by the associated basis function. Individual basis functions can be sorted according to their significance for the description of the spectral variability in the decomposed set using the eigenvalues associated with them. The minimal number of subspectra for which the differences between the experimental and the reference set spectra, reconstructed from the subspectra ordered according to their eigenvalues, do not exhibit significant spectral features then defines the value of p.

Once we know how to decompose the reference spectra set into the principal components, it is equally important to recognize any correlations or dependencies between the reference set structural descriptors or their components. This knowledge might contribute significantly to the method design and simplification. We can organize the structural descriptors of the reference molecules into matrix $[D_i(k)]$, which is constructed from k rows corresponding to individual descriptor elements (e.g. fractions of helix, sheet, etc.) in N columns representing the reference set molecules. This matrix can be subjected to the rank analysis as described above. The extent of the relation between the elements of three descriptors of protein structure is discussed below in the light of the rank analysis as an illustrative example.

We chose a representative set of protein structures from the protein data bank that spans all structural types. This set corresponded to 430 structures with good data quality and less than 25% sequence identity.[116] There are three types of structural descriptor that have been used until now for CD spectroscopic structural studies of proteins: SSF, fractions of distorted secondary structures,[109] and matrix descriptors for secondary structure segments determination.[69,70,117] Construction of these descriptors requires the secondary structure assignment of each amino acid residue that was found using the algorithm for the determination of the secondary structure of proteins.[101] Figure 18 summarizes the protocol for derivation of the last two descriptors from the secondary structure information. The elements of these descriptors were organized into matrices with 430 rows and corresponding numbers of columns for all proteins (five for SSF, including helix, sheet, turn, bend, and other; three for fractions of distorted secondary structures that included helix, sheet, and other; and nine for matrix descriptors of secondary structure segments, which were 3×3 matrices as shown in Figure 18). Table 1 summarizes the fractions of total variance in these structural descriptor matrices that is accounted for by individual principal components.

It is evident from the large fraction of variance described by the first principal component in these analyses that the helical, sheet, and other structural parts are not distributed completely randomly in the respective

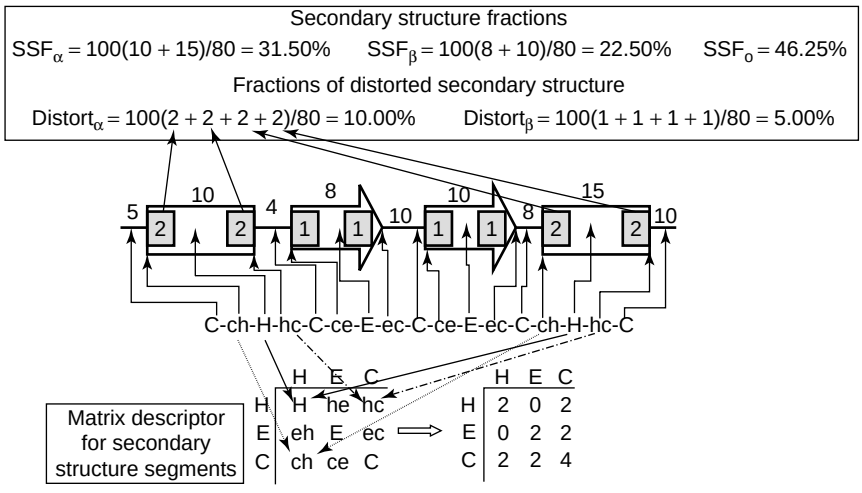

Figure 18 The derivation of descriptors related to the number of secondary structure segments in the protein fold. After the secondary structure assignment, the linear segment model is constructed (rectangles, α-helix; arrows, β-strands; lines, loops; numbers above the segments are the lengths of the segments in the amino acid residues; shaded rectangles and numbers inside the segments are the number of residues assumed to be distorted by the segment end-effects). SSFs and distorted SSFs are calculated as shown in the top part of the figure. The linear segment code is then generated (C, H, and E are ''central'' unperturbed parts of loop, helix, and strand segments, respectively; ch, hc, ce, ec, eh, and he are respective segment boundaries) and the numbers of all segment and boundary types are organized in the matrix scheme shown at the bottom.

Table 1 Fractions of total variance accounted for by most significant principal components in the analysis of structural descriptors for 430 representative proteins

Descriptor	Principal component (%)			
	1	2	3	4
SSF	79.9	10.5	5.5	2.2
Cumulative	79.9	90.4	95.9	98.1
Fractions of distorted secondary structure	88.7	11.1	0.2	
Cumulative	88.7	99.8	100	
Matrix (3 × 3) for segments of secondary structure	82.4	8.0	5.7	6.8
Cumulative	82.4	90.4	96.1	100
Matrix (5 × 5) for segments of secondary structure	74.2	6.1	3.9	2.2
Cumulative	74.2	80.4	84.2	86.4

protein folds. Figures 19–22 demonstrate graphically the extent of the descriptor element interdependencies that is the reason for the calculated result. It is clear from Figure 19(a) that for the SSF descriptor the dominating interdependence exists between helix and sheet fractions. This was also shown earlier by the neural network study.[118] This interdependency is generally overlooked, although it is not very surprising. It reflects the fact that most globular proteins are typically composed of about 70% ordered structure (helix and β-strands). Proteins made of turns and disordered structure are rare and, consequently, the fractions of bends and turns in the studied set do not exhibit such interdependence. This, in turn, results in projection of the helix–sheet correlation into their relations with ''other'' fractions (see Figure 19b). This is demonstrated for a typical content of ordered structures in a globular protein by the linearity of the helix–strand fraction dependence, with the average line cutting the fractional content α and fractional content β axes at 70.

Fractions of distorted secondary structures are significantly dependent on the total corresponding SSF values, as is seen in Figure 20(a) and (b). Correlation is stronger for strand than for helical segments. The helix–sheet interdependence discussed above is, therefore, projected also into significant correlation between fractions of distorted helices and strands (Figure 20c). Elements of matrix descriptor for determination of secondary structure segments are much less significantly related to the corresponding SSF values than is the previous descriptor (Figure 21). They also exhibit (with the exception of relations between numbers of strand and other segments) a lesser degree of interdependence, demonstrated in Figure 22.

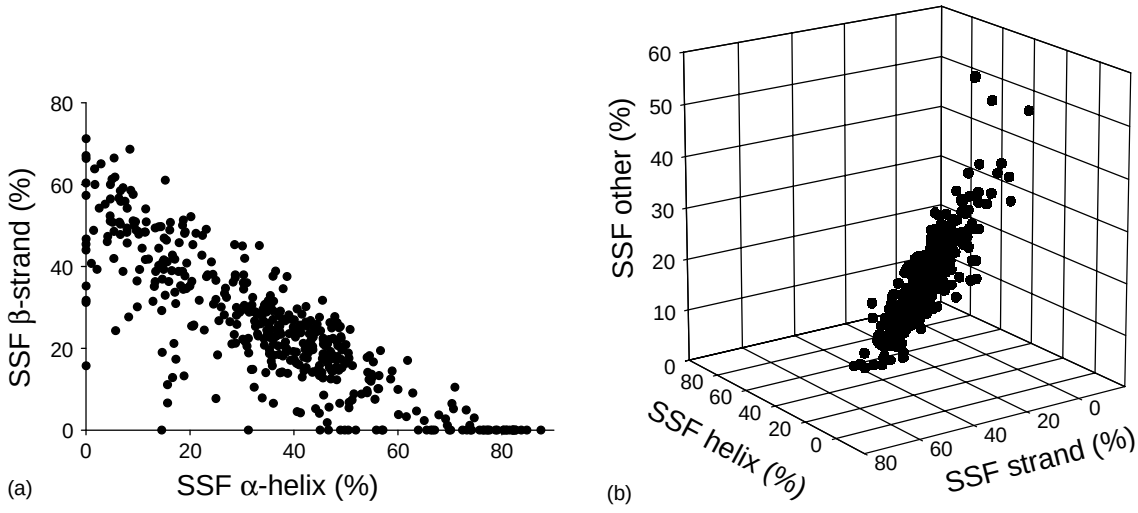

Figure 19 Interrelation of the SSF values for helix, sheet, and ''other'' in globular proteins. (a) Correlation of the sheet to helix SSFs in the set of 430 representative protein structures selected from a protein data bank and processed by the determination of the secondary structure of proteins algorithm[101] for secondary structure assignment of helices, sheets, bends, turns, and ''other''. (b) Correlation of fractions of helix, sheet, and other structures, demonstrating the minimal impact of the bend and turn fractions on the observed statistical dependency.

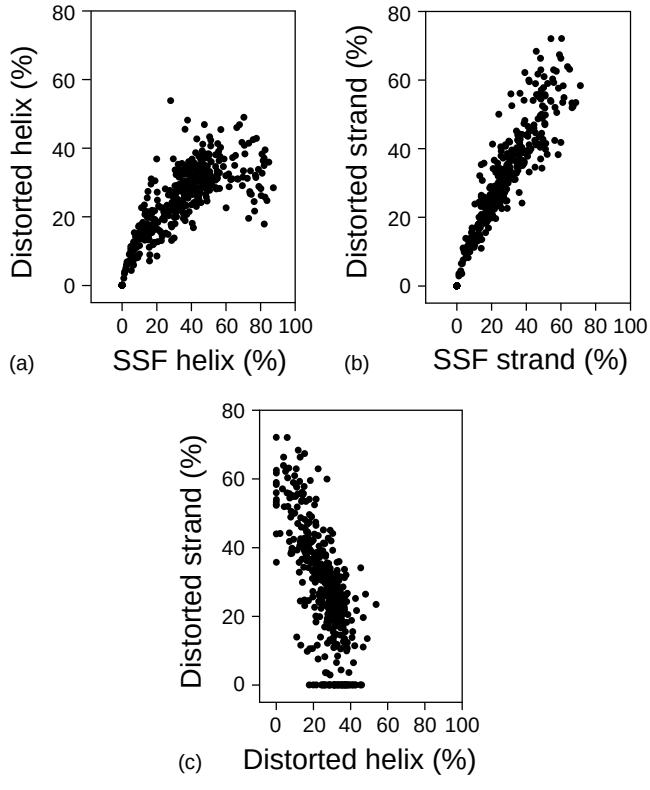

Figure 20 Interrelations of the distorted SSF values of helix and sheet segments in globular proteins. Fractions of secondary structures considered to be distorted can be correlated with the corresponding total SSF values found by the determination of the secondary structure of proteins algorithm: (a) helix–distorted helix; (b) sheet–distorted sheet. (c) Interrelation between the fractions of distorted sheet and distorted helix in the representative set of 430 protein data bank structures.

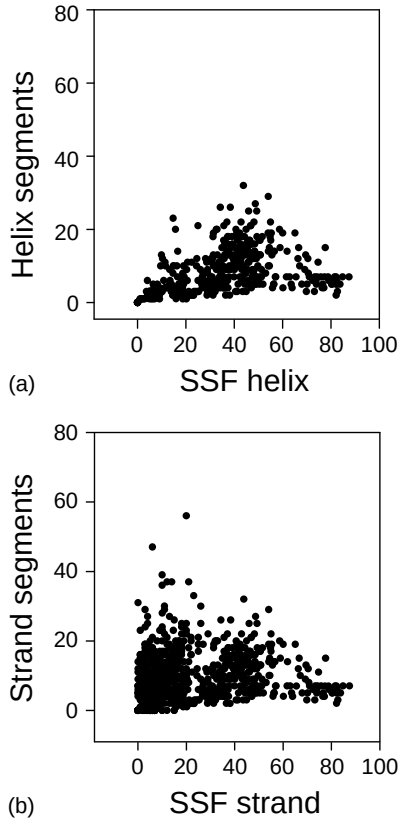

Figure 21 Interrelation of the number of helical and sheet segments to the fractions of helix and sheet in globular proteins. The number of (a) helical and (b) sheet segments (elements H and E of the matrix descriptor in Figure 18, respectively) are correlated with the helical and sheet SSFs for the representative set of 430 protein data bank structures.

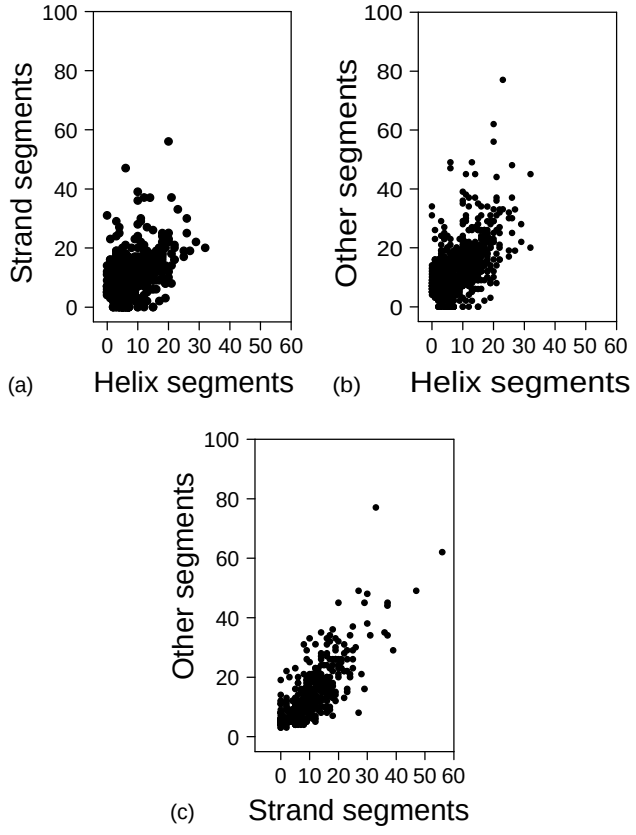

Figure 22 Interrelation of the number of helical, sheet, and "loop" segments in globular proteins. The number of (a) helical–sheet, (b) helical–loop, and (c) sheet–loop segments (elements H, E and C of the matrix descriptor in Figure 18, respectively) are correlated for the representative set of 430 protein data bank structures.

As a practical consequence of these approximate dependences, any empirical method of structural analysis of protein spectra should be able to predict multiple components of protein structural descriptors with errors smaller than the variance of the relations between these components. If this is not the case, it cannot be unambiguously determined if protein spectra really contain information about more than one component of that descriptor. For example, the absolute error in determination of the sheet fraction from CD spectra should be significantly smaller than 15%. Envelopes of the correlation plots (Figures 19–22) can be used to define similar practical error limits for other descriptor elements.

4.3 Analysis of the Compatibility of the Analyzed Protein Spectrum and the Reference Set

Before applying any empirical algorithm extracting structural information from the experimental CD spectrum of a sample with unknown structure, it is necessary to ensure that the analyzed spectrum is compatible with the reference spectra. Practically, such a testing can be based upon an algorithmized concept of similarity between the unknown and the reference spectra because the measured spectra are usually the only information available about the analyzed sample. Any quantitative pattern recognition or statistical method can be used for this purpose. The following examples are selected to demonstrate various strategies of compatibility testing.

One possibility is to calculate the correlation coefficients ("overlap integrals") $r_{\text{Unknown, ref}}$ of the studied CD spectrum $\Theta_{\text{Unknown}}(\lambda)$ with all other spectra $\Theta_{\text{ref}}(\lambda)$ in the reference set after normalization as (Equation 23):

$$r_{\text{Unknown, ref}} = \int_{\lambda_1}^{\lambda_2} w_{\text{Unknown}}(\lambda) w_{\text{ref}}(\lambda)\, d\lambda \qquad (23)$$

The unknown spectrum should exhibit statistically significant correlation to several or all reference CD spectra.

Because the problem has multiple dimensionality, cluster analysis is another algorithm that can be used for confirmation of compatibility of the analyzed and reference spectra. Clustering might be performed directly with the spectra (see Figure 23a) or using the principal component loadings C_{ij} or c_{ij} (see Equations 13 and 22). The last technique is more sensitive to unique features in the analyzed spectrum because principal component decomposition generates extra subspectra for features that are uncorrelated to spectral variance of other samples. High loadings of this extra subspectrum for the spectrum that contains this unique feature then discriminate clearly the outlier by its large generalized distance from the rest of the proteins in the reference set (Figure 23b).

Neural networks[119] can be also used to confirm the compatibility of analyzed and reference spectra. For this effect, the autoassociative network is trained to project spectra of the reference data set into themselves.[120] The general topology of such a projection neural network is shown in Figure 24(a). The network has n input and n output neurons, each corresponding to one wavelength of the spectra. During the training process, these neurons are filled with the respective intensities at the corresponding wavelengths. The network is trained (backpropagation is a suitable paradigm for this purpose) by optimization of synaptic weights matrices $\mathbf{w}_A$ and $\mathbf{w}_B$ that transmit the spectra through the potentials of a selected number of hidden-layer neurons. The number of neurons in the hidden layer can be optimized. This process should balance the number of degrees of freedom in the data used to train the network and its capability to generalize, i.e. to interpret new spectra. This is equivalent to finding a relation between the number of hidden neurons and the complexity of the analyzed data. The PCA of the weights exported from the pre-trained network with a

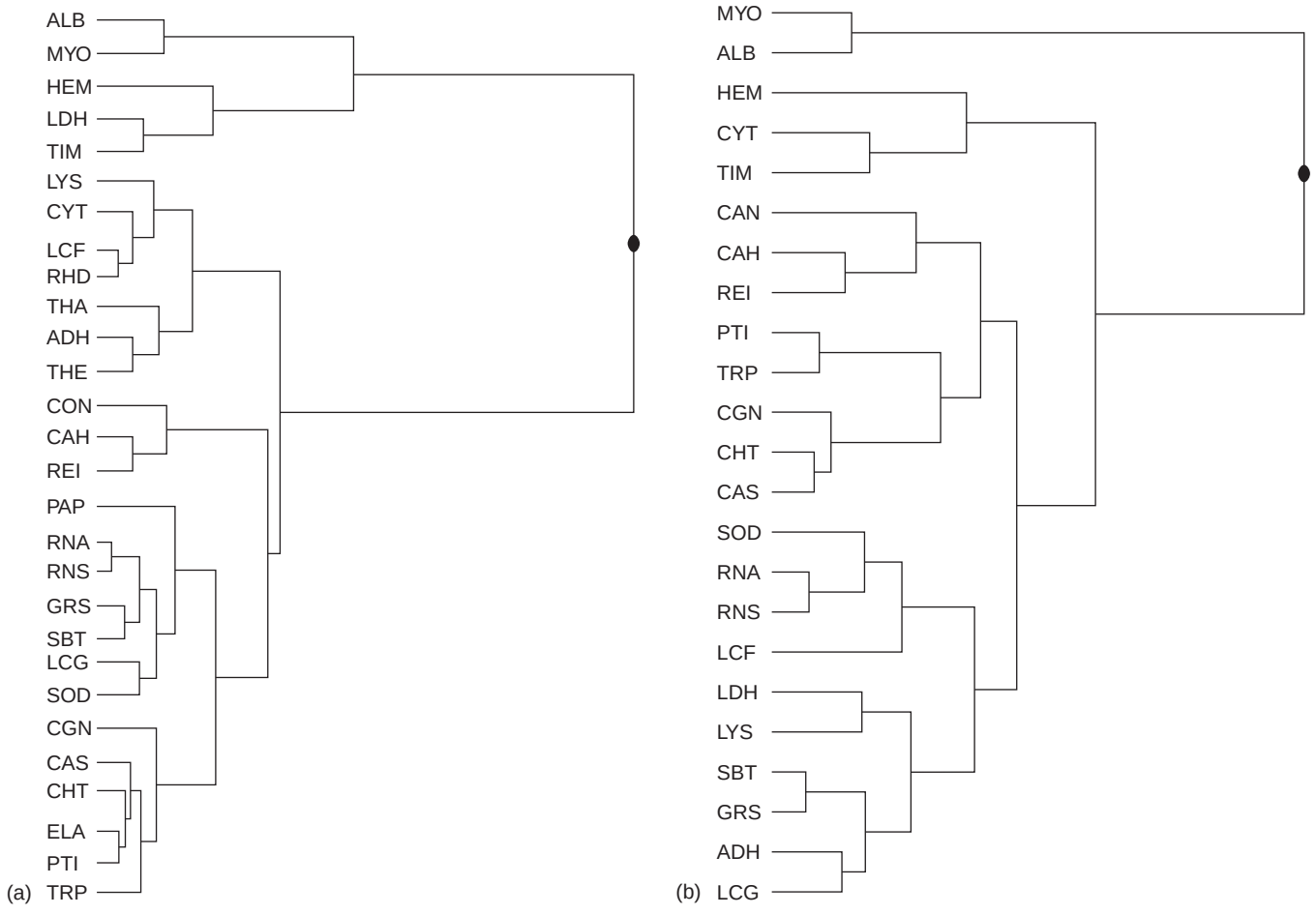

Figure 23 Comparison of the cluster analysis of 28 protein ECD spectra using direct and principal component representation of their bandshape. (a) Dendrogram characterizing the similarity in the reference set. Clustering was calculated with the 200 equidistantly digitized ECD intensities as input and with the Lance–Williams flexible clustering algorithm. (b) Lance–Williams flexible clustering of the reference protein ECD spectra using the loadings of the five principal components needed to decompose the spectra as in Equation (13). ALB, albumin; MYO, myoglobin; HEM, hemoglobin; LDH, lactate dehydrogenase; TIM, triose phosphate isomerase; LYS, lysozyme; CYT, cytochrome *c*; LCF, lactoferrin; RHD, rhodanese; THA, thaumatin; ADH, alcohol dehydrogenase; THE, thermolysin; CON, concanavalin A; CAH, carbonic anhydrase; REI, immunoglobulin; PAP, papain; RNA, ribonuclease A; RNS, ribonuclease S; GRS, glutathione reductase; SBT, subtilisin bpn; LCG, lactoglobulin; SOD, superoxide dismutase; CGN, chymotrypsinogen; CAS, α-casein; CHT, chymotrypsin; ELA, elastase; PTI, trypsin inhibitor; TRP, trypsin.

large number of hidden neurons (~60 for about 100 input spectral intensities) can be used for this purpose in the special case of spectra analysis. The method is able to find the rank p of the weight matrix, and the optimal network topology will, therefore, have p hidden-layer neurons.

The hidden layer actually decomposes the processed spectra into components that are in many aspects similar to subspectra discussed in the previous section. Equation (24) defines the actual calculation that the trained backpropagation neural network performs during the projection of the spectra $\Theta(\lambda)$. (The hyperbolic tangent is used as the transition function that determines the ''firing'' of each neuron and weight matrices $\mathbf{w}_A$ and $\mathbf{w}_B$ were obtained by training.)

$$\Theta(\lambda) = \tanh([\mathbf{w}_B]^T \tanh([\mathbf{w}_A]^T \Theta(\lambda))) \qquad (24)$$

If we substitute $\vec{\psi}$ for $\tanh([\mathbf{w}_A]^T\Theta(\lambda))$ and invert the tanh function, we can demonstrate the formal analogy of the network action with the PCA decomposition (Equation 14).

$$\operatorname{arctanh}(\Theta(\lambda)) = [\mathbf{w}_B]^T \vec{\psi} \approx \Theta(\lambda) + \frac{\Theta^3(\lambda)}{3} + \frac{\Theta^5(\lambda)}{5} + \cdots \qquad (25)$$

Equation (24) shows explicitly the nonlinear character of the network processing of the spectra. If only the first term of the expansion is preserved, then a matrix of weights $[\mathbf{w}_B]^T$ would play the role of the subspectra and the hidden neuron potentials $\vec{\psi}$

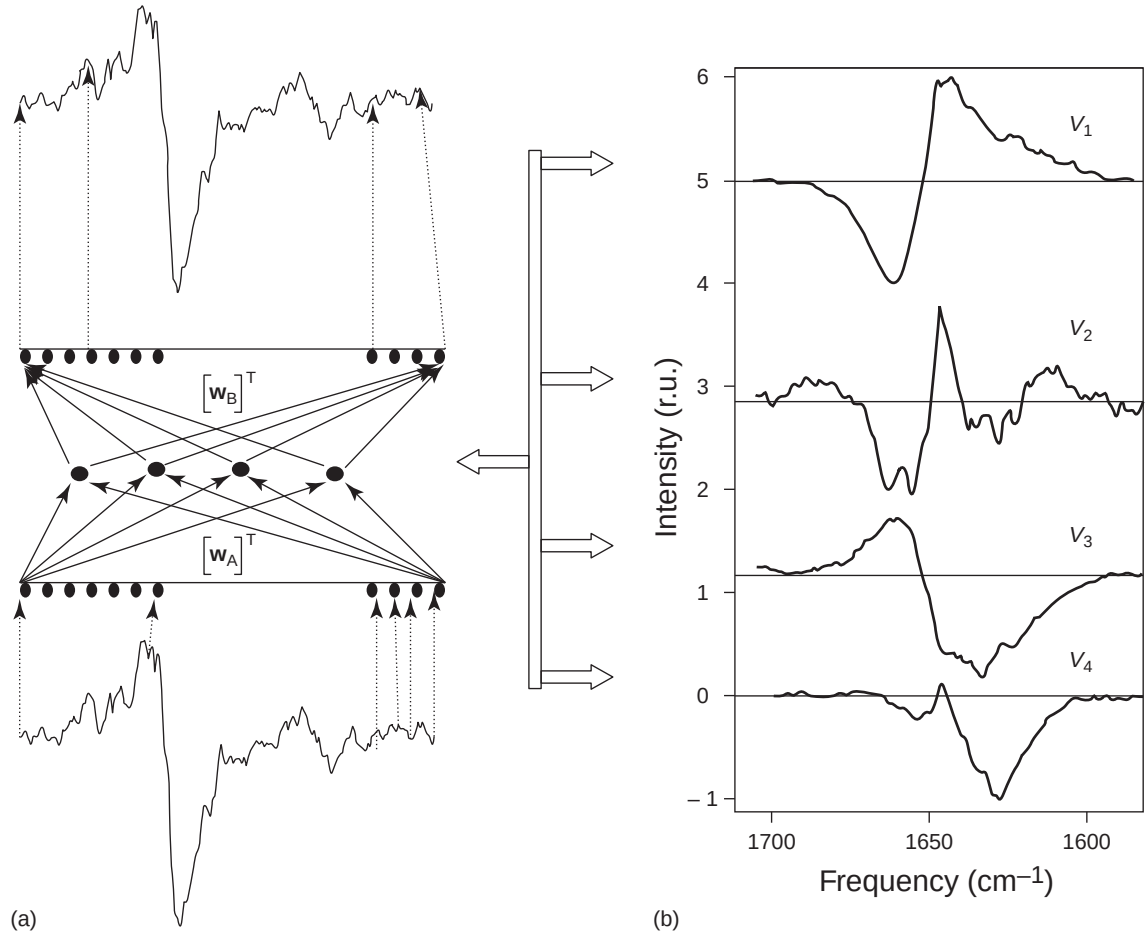

Figure 24 Reduction of protein VCD spectra into synaptic weight components of the autoassociative projection neural network. (a) The topology is shown of the autoassociative backpropagation neural network after optimization for projection of VCD spectra of 28 proteins through the single hidden layer using as input 100 equidistantly recorded VCD intensities. The desired output and the number of hidden layer neurons were optimized to four as described in the text. (b) A spectral representation of the network synaptic weights connecting the respective hidden layer neurons (V_1–V_4) to the output layer is given. As the output intensity is the sum of hidden layer potentials scaled by these weights, the picture demonstrates the decomposition of the experimental VCD spectra into network-generated component spectra (subspectra).

would correspond to a column of the coefficient matrix in Equation (22). Figure 24(b) demonstrates that this algebraic analysis corresponds to the actual network behavior. The connection weights between the hidden layer and output neurons in the autoassociative networks trained to project protein VCD spectra on themselves are plotted as a function of frequency. These plots of network weights can be definitely interpreted as the spectral components.

The spectrum of the unknown protein is submitted as input to the trained neural network with an optimized topology and its projected reconstruction is compared with the experimental prototype. If there are no significant differences, the spectrum of the analyzed protein can be considered to be compatible with the training set reference information.

4.4 Description of the Spectral Variance

The next step in the analysis is to describe quantitatively the variability of the bandshape of reference and analyzed spectra (Figure 17). There are numerous methods for encoding the shape of a spectrum into numbers that can be used as the input information for an algorithm correlating them to structural descriptors. These methods range from determination of the intensity at a selected wavelength to neural network algorithms.

1. The single wavelength intensity method. The molar ellipticity of the negative maximum at 222 or 208 nm is determined from protein ECD spectra.[107,108]
2. The similarity method. The correlation coefficient (overlap integral) of all spectra in the analyzed set with one reference spectrum is determined

using Equation (23).[124] In this method, the spectral bandshape is represented by a single number with magnitude between −1.0 and 1.0 that characterizes the similarity or dissimilarity of each spectrum to the reference one.

3. Cluster analysis method. Intensities at regularly spaced frequency intervals of the spectrum are determined;[100,122] the shape is represented ''as is'' in the form of a single vector of equidistantly spaced spectral intensities.
4. Band deconvolution method. The CD spectra are broken down into a series of p common component bands so the bandshape of any spectrum is encoded in a vector with $3p$ parameters defined by position, half-width, and amplitude of each component band.[123]
5. Methods using component spectra. These approaches deconvolve the spectra into a series of p component spectra. These spectral components might have different origins, which define the principal difference between various interpretive methods that utilize such spectral bandshape encoding. They might be derived from CD of model compounds (peptide ECD were used to decompose protein ECD); they might be represented by spectra of reference molecules with known structure,[124] or by the basis functions (subspectra) defined mathematically using PCA,[113,125,126] singular value decomposition,[103–105,127–132] simplex optimization,[133] or neural networks.[70,132,134] In all these methods, the bandshape of any spectrum is quantified by a set of p vectors with dimension identical to the number of points in the analyzed spectrum and one p-dimensional vector of the component spectra loadings (e.g. p subspectra $\phi_j(\lambda)$ with a row of coefficients C_{ij} in Equation (13) or matrix of weights and hidden-layer neuron potentials in analogy to Equation (25).
6. Neural network methods. These processing programs map spectra–structure relations using optimization of neuron potentials and synaptic weighting.

To further emphasize the bandshape variability, it might be advantageous to transform the reference and processed spectra into the difference variant by subtracting the set-average spectrum from them.[85,135] It should be noted that the quantitative bandshape variability decomposition with mathematically defined component spectra could be performed independently on the structural part of the problem. This is equivalent to analysis recognizing the internal correlations between the various spectral features, which can be then used with advantage in comparisons with the available structural information.

4.5 Description of the Spectra–Structure Relation

There is a validation protocol that is used almost universally to assess the degree of success during development of an empirical technique for biomolecular spectral analysis. It is called ''one-out testing'' and will be explained first. Its implementation requires that the spectral bandshape decomposition is already established together with the algorithm that projects the spectral components into the structural descriptor values. For ''one-out testing'', a new reference set of spectra and structural descriptors are created from the original ones by eliminating the spectrum and the structural descriptor of one reference molecule. This reduced set is used to optimize the projection function parameters in the spectra–structure relation. The left-out spectrum is then processed by the re-optimized relation as if it is an unknown sample. Structural descriptor predictions are compared with the known values for the ''one-out'' reference molecule and the prediction error is stored. This procedure is repeated by eliminating systematically all reference spectra and the corresponding descriptors one by one. A series of prediction errors is thus collected. The standard deviation σ_i is calculated from them as a quantitative characterization of the method prediction error. Alternatively, relative standard deviations, $\sigma_{i,\mathrm{rel}}$, can be used because their values are less dependent upon differences in the reference sets.[36,132,136,137] They are calculated from the standard deviation σ_i of ''one-out'' prediction of structural type i, and the dynamic range of the structural descriptor values in the reference set, $d_{i,\max}$ and $d_{i,\min}$ (Equation 26):

$$\sigma_{i,\mathrm{rel}} = \frac{\sigma_i}{d_{i,\max} - d_{i,\min}} 100\% \tag{26}$$

The next step in the development of the analysis method is to establish a relationship between spectral bandshape descriptors that were determined by the methods summarized above and the molecular structure descriptors of the reference molecules (Figure 17). Two complementary strategies can be applied. First, it can be assumed that all spectral components are related exclusively to the variability of the structural descriptor. A typical example of this approach is the assumption that the changes in ECD spectra of proteins have to be exclusively related to the variation in SSF. The second type of strategy postulates that only a subset of the spectral variability components can be related to the information encoded by the molecular structure reduction into the structural descriptor values (Figure 17).

The two strategies also employ different mathematical tools for establishing quantitative spectra–structure relations. In the first category, a dominant role is played by algebraic and optimization tools that are designed

to convert all spectral components and their numerical loadings into the descriptor values. They are based on the following set of assumptions.[137]

- Every studied CD spectrum is completely explained as a linear combination of component spectra. In the case of protein ECD, these component spectra characterize individual secondary structures. There is no spectral demonstration of the effect of the tertiary structure in this case.
- There are no differences between the averaged structures of the molecules studied in solution and in crystalline form (studied by spectroscopy and X-ray diffraction, respectively).
- Spectral contributions from chromophores or physical mechanisms that uniquely modify the spectral bandshape but are not universal for all processed spectra do not alter the analysis results. In the case of protein ECD spectra, this is equivalent to the assumption that contributions from nonpeptide chromophores (e.g. the aromatic side chains) are negligible.
- Structural types used to formulate the quantitative structural descriptor are characterized as rigid, ideal (secondary) structures. For proteins, this assumption is equivalent to neglecting the variability of ideal helical or sheet dihedral angles.

Mathematical algorithms for the second analysis strategy must contain some information filter that optimally selects the subset of all spectral components in the reference set and differentiates them from components that are unrelated to the question defined by the predicted structural descriptor. Examples of such algorithms are the variable selection method,[103,104,106,128–131] ridge regression algorithm,[124] complete multiple linear regression optimization,[66,101,109,112,117,125,136,138] and response analysis of the trained neural network.[69,70,132–135] Various mathematical methods have been used to relate the structural descriptors of proteins to the spectral bandshape characteristics (1–5 above); these are now described in more detail as a practical example of this stage in the methodology.

4.5.1 Single Wavelength Intensity Method

The simplest relations that are useful for rapid monitoring of protein conformational changes are the linear correlations between the ellipticity at 222 or 208 nm in a protein ECD spectrum and the fraction of α-helix.[110,111] The errors in the absolute predictions of helical fractions from these values are only moderately less accurate than those from more sophisticated methods. The reason for the success of this method is that the α-helical spectral bandshape dominates ECD once the protein contains 25% or more helical structure. Therefore, the intensity at a single wavelength is a good descriptor of the overall spectral bandshape. The disadvantage of this method is its limitation to prediction of only one secondary structure and its sensitivity to contributions from other chromophores if they contribute to the ECD intensity at the selected wavelength.

4.5.2 Similarity Method

The similarity method still relies on a single number that characterizes the spectral bandshape but utilizes all measured spectral points for that characterization. It was developed[121] to analyze the protein VCD in the amide III region (1100–1400 cm^{-1}). This approach is based on the evaluation of similarities of the spectral bandshapes between the reference and analyzed VCD spectra that are quantified by the value r_{ij} of overlap integrals between them (Equation 23). These r_{ij} values are calculated numerically by interpolating the product of the two spectra by natural cubic spline functions and integrating the result. Linear regression is then used to relate the r_{ij} values to reference protein SSFs. There is an optimization step involved in the determination of the final regression function $SSF_{\xi,i} = kr_{\mathrm{ref},i} + q$. There is no a priori indication as to which protein spectrum should be selected as the reference bandshape. Therefore, all possibilities should be tested systematically by selecting all protein spectra in the set as the reference bandshape. These tests should be performed independently for all secondary structure types and the linear equation with the lowest prediction error in ''one-out testing'' should be used for the processing of the VCD spectra of the unknown samples. This method is suitable for structural studies that use spectra with unavoidably large intensity errors. Overlap integrals r_{ij} are calculated from the spectra $w_i(\lambda) = \Theta_i(\lambda)/A_i$ normalized by the spectrum area A_i defined by Equation (15). Intensity variations are thus removed from r_{ij} and only the bandshape is involved in the similarity quantification. A disadvantage of the method is the lack of flexibility in the selection of the function representing the spectra–structure relation, which can have only one independent variable.

4.5.3 Cluster Analysis Methods

Spectral bandshape ''as is'' can be understood as a multicomponent vector characterizing the protein structure. In this model, every one of the n measured frequencies represents one dimension of a generalized n-dimensional space. The CD intensity at that frequency stands for a coordinate that defines the unique protein position in that space. Cluster analysis algorithms provide measures for distance determination in such multidimensional

spaces.[115,139,140] A generalized distance matrix can be determined in the CD spectra space by the application of these algorithms to the spectral matrix $[\Theta_i(\lambda)]$ (see section 4.3). Graphical representation of the generalized distance matrix, usually in the form of a dendrogram, is then used to visualize the results. The distances are ordered hierarchically and connected in a dendrogram scheme that uses the calculated relative similarity indices for all proteins (see Figure 23a).

Unsupervised clustering constitutes usually the first step in the pattern recognition or discrimination analysis. In an unsupervised clustering, the structural interpretation of the resulting dendrogram can be found by comparison with another clustering of the reference protein set that is based on the structural descriptors as generalized protein coordinates. Ideally, both spectrally based and structurally based clusters of the proteins in the reference set should be organized into similar hierarchical schemes. Selection of the various metrics and algorithms for cluster definition can be used to optimize this spectra–structure dendrogram similarity. This methodology was applied to protein ECD and VCD spectra characterization in terms of SSF values.[100,136] Characterization of protein tertiary structure folding class (all-α, α/β, $\alpha+\beta$, all-β, and denatured proteins) using clustering of their ECD spectra was implemented.[122]

In a supervised clustering, parameters of the generalized distance can be adjusted using the protein structural feature under investigation so that the resulting clusters are forced to reflect both the spectral and structural similarity. Both the applications mentioned above found that clustering similar to that based on the complete CD spectra can be obtained when only three to five principal component loadings C_{ij} of the experimental spectrum decomposition by PCA algorithm are used as the protein coordinates (see Equations 18 and 12). One study used this simplification of the protein spectral coordinates to define a decision function optimized to discriminate between the known tertiary structural classes of the reference protein set.[122] This function has the same number of parameters, W_i, $i = 1, \ldots, N$, as the number of data points in the processed spectrum. These parameters are found in the perceptron algorithm training that uses the spectra and the known fold types in the reference set. Assignment of the structural class for an unknown protein is based on the value of scalar product D between its spectrum $\Theta_i(\lambda)$ and parameter vector $\vec{W}$ (Equation 27):

$$D = \vec{W}\,\vec{\Theta}\,(\lambda) = \sum_{i=1}^{N} W_i \Theta(\lambda_i) + W_0 \qquad (27)$$

Clustering algorithms are suitable for studies of structural features that are difficult to encode quantitatively. Protein tertiary structure class characterization is an example of such a fuzzy structural concept. The user needs to handle the multiplicity of metrics for generalized distance calculations and multiplicity of clustering algorithms that can generate very different similarity schemes with the same inputs. Application of several algorithms and retaining only those clusters that are identified invariantly in all or most of the calculations might be tedious but it is a relatively robust strategy to avoid overinterpretation of a specific clustering.[100]

4.5.4 Band Deconvolution Methods

Deconvolution of the CD spectra into component bands became useful for protein structural studies only with the advent of VCD.[123] Frequency sensitivity of the IR bands is much greater than that of electronic transitions, especially in the amide I vibration region at 1550–1750 cm^{-1}, which is dominated by the carbonyl stretch mode. Unfortunately, the bisignate character of the CD spectrum increases deconvolution ambiguity, as is demonstrated in Figure 25. Two deconvolutions of the albumin VCD spectrum implements different numbers of peaks but are

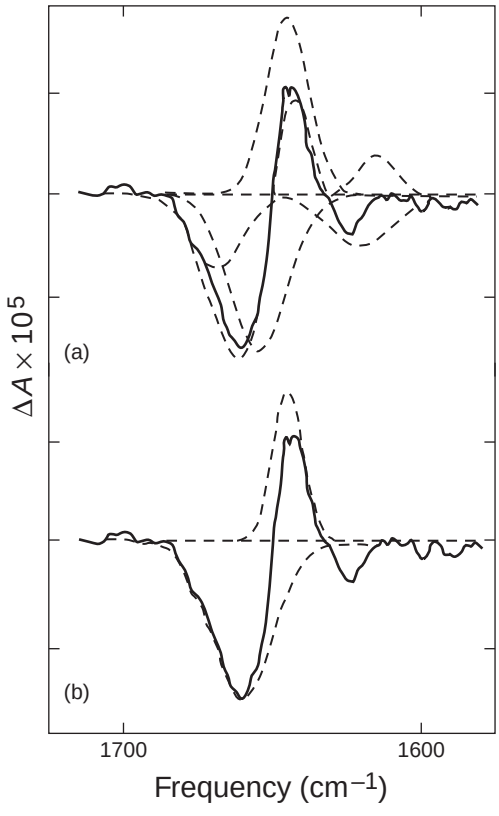

Figure 25 Problems in deconvolution of protein VCD spectra into the component bands. Deconvolution of VCD spectrum of hemoglobin into two sets of Gaussian bands give an equally significant description of the VCD bandshape whether (a) five band or (b) three-band deconvolution is used.

indistinguishable on the level of standard deconvolution error criteria. To retain acceptable and consistent solutions for all proteins in the set, a multistep procedure had to be applied in the deconvolution of the spectra of the 23-member protein VCD set that generated the example. Simultaneous processing of IR absorption spectra of the same proteins after their FT self-deconvolution and second-derivative calculation were used in the process. Multiple optimization cycles with selective band parameter restrictions were necessary. Despite this effort, component peak areas were uncorrelated to SSF values. The main contribution of VCD spectroscopy to the structural interpretation of protein vibrational spectra is the unambiguous proof that secondary structural assignments of various amide I component bands are, indeed, ambiguous. Amide I component bands with the same characteristic frequencies but opposite signs of CD components were clearly identified in the FTIR and VCD spectra of many proteins. This is a semiquantitative argument but its strength stems from the structural requirements that are necessary to change the VCD band sign. The bands with the opposite VCD signs cannot arise from the same type of structural elements even though the frequencies appear from FTIR measurement to be the same.

4.5.5 Methods with the Component Spectra

Methods that deconvolve reference set CD spectra into a linear combination of the component spectra (subspectra) use the most informative transformation of the experimental input. The independent or selective treatment of various spectral features becomes possible only at this level of spectral variability description. This transformation of the experimental spectra also facilitates testing of the multiple combinations of the component spectra against various protein structural features in the final relationship optimization. There are several conceptual differences characterizing algorithms that can be classified using the criteria listed in Table 2.

There are several ECD analysis methods that use subspectra of various origins with invariant shape for the unknown spectrum decomposition but no selection of spectral features or modification of the reference set is then performed.

4.5.5.1 Methods Using Experimental Circular Dichroism Spectra of Synthetic Polypeptides or Reference Proteins Multiple linear regression was used in classical methods to express the experimental CD by a linear combination of model peptide subspectra.[107,108] These algorithms exist in variants with and without consideration of nonnegativity and constraining the sum of predicted SSF to 100%. SSFs of α-helix, parallel and antiparallel β-sheet, turn, and "other" conformations can be determined with good relative prediction errors. In a further development, the subspectra derived from a reference protein ECD by solving the equation $[\Theta_i(\lambda)] = [\phi][\mathrm{SSF}_i]$ for $[\phi]$ with the X-ray values of SSF were used.[141] The subspectral intensity can be corrected for the protein length N by an empirical function of $1/N$ (see section 1.1.6). A variation of the above method that uses a ridge regression replaces the subspectra by the reference protein spectra.[124] The SSF prediction is then optimized by selective statistical weighting of contributions of reference ECD spectra to the spectra–structure relation.

Table 2 Summary of different approaches to basic steps of protein CD spectra analysis in terms of SSFs

Basic steps of the analysis	Method
Subspectra origin	Model peptides provide CD spectra Protein CD spectra Mathematical transformation of the experimental spectra
Deconvolution of an unknown spectrum	Independently on the reference set (the unknown spectrum has no impact on the subspectra) Together with the reference set spectra (unknown spectrum influences the subspectra bandshape)
Selection of spectral feature for structural predictions	No selection at all Variable selection of proteins in the reference set Variable selection of subspectra and their combinations Variable selection of the spectral frequency region that is used for processing Variable generation of multiple spectra–structure relationships optimized to predict structures of various reference proteins and use of similarity descriptors to find the relationship optimized for reference protein with the closest spectral bandshape to that of the unknown
Utilization of constraints on the structural descriptor values	Descriptor constraints are not a part of the algorithm; they are used only for postcalculational reliability evaluation Constraints are integral part of the algorithm
Utilization of subspectra $\phi_j(\lambda)$ or coefficients C_{ij} as the bandshape descriptors	Uses subspectra $\phi_j(\lambda)$ Uses coefficients C_{ij} Uses both subspectra $\phi_j(\lambda)$ and coefficients C_{ij} Uses coefficients C_{ij} decomposed into the intensity contribution (norm A_i, Equation 23), and bandshape contribution c_{ij}: $C_{ij} = A_i c_{ij}$

4.5.5.2 Methods Using Mathematically Derived Component Circular Dichroism Spectra The next group of methods uses subspectra derived by a purely mathematical transformation of reference CD spectra. The convex constraint method uses simplex optimization to find related intensity changes and groups them into subspectra that are then attributed to some secondary structural type according to the calculated bandshape.[133] It is allowed that one secondary structure can be characterized by more than one convex constraint subspectrum. Most other algorithms from this group are based on subspectra generated by projection of the reference spectra to an orthogonal basis using eigenvectors of correlation or covariance matrix (see section 4.3).[99,100,103–106,112,113,117,125,127–129,131,136,138,142] Singular value decomposition, variable selection method, self-consistent method, and restricted multiple linear regression implement various combinations of the features summarized in Table 2. All have been applied to protein ECD spectra to predict SSFs; only restricted multiple linear regression has been used for protein VCD spectra analysis. The self-consistent method is the only one from this group that was used for prediction of a structural descriptor other than SSF. All but restricted multiple linear regression use subspectra to find the structural descriptor. The major differences between these methods are the optimization principles on which they are based. All but restricted multiple linear regression solve a so-called generalized inverse equation that generates subspectra from matrices of structural descriptors and experimental CD spectra of the reference set. The unknown protein spectrum is then expressed as a linear combination of these subspectra with coefficients providing the structural fractions. In this sense, the methods follow the conventional model of decomposition of the experimental spectrum into secondary structure-related subspectra. These spectral components are the best least-squares approximation of the spectral response to fractions of various secondary structures in the reference set.

Differences of Individual Algorithms. In the singular value decomposition and variable selection method, the unknown spectrum is not processed with the reference set. To optimize the prediction performance, the program systematically selects various subsets of reference spectra and repeats the prediction cycle with monitoring of boundary conditions for the SSF. Acceptable solutions are averaged.

In the self-consistent method, the unknown spectrum is processed together with the reference spectra. This co-processing cannot be completed without the values of the unknown SSF. The method overcomes this difficulty by similarity-based initial guessing of the unknown SSF and iterative process. The subspectra determined by the modified variable selection method of singular value decomposition of the whole set are used to determine the first iteration of the unknown protein SSF. This result is substituted for the initial guess; a new set of subspectra is determined and the process is repeated until the SSF for the unknown protein does not change in consecutive cycles.

Using the Spectral Component Loadings Instead of the Subspectra. Restricted multiple linear regression uses a completely different approach. The first step is principal component decomposition (Equation 14) of the reference spectra together with one or more spectra of unknown proteins, which is mathematically exact and does not require any initial estimates or iterations. Calculated subspectra are used only as basis functions that provide coefficients C_{ij}, which characterize numerically the bandshape variability of both reference and unknown spectra for each analysis. This step is again exact, without any fitting or additional optimization. The number of these coefficients, which is initially equal to the number of processed spectra N, is reduced to a minimal number p necessary to reproduce the experimental bandshapes: typically from $N \sim 30$ to $p \sim 5–10$. To test the quality of the reconstruction, the difference between the experimental and reconstructed spectra is calculated and the differentials are monitored. The final set of $N \times p$ numbers (loadings in the factor analysis nomenclature) represents the pool of digitally encoded spectral components that are assumed to be dependent on various aspects of protein structure, not exclusively on the SSF.

The next step – restricted multiple linear regression – optimizes the combination of these components into a subset that best predicts the selected structural descriptor, SSF. The optimal restricted multiple linear regression relationship for each possible total number, n, of component loadings C_{ij}, $n = 1$ to p, is sought in the regression cycle in which all possible combinations of n-tuples generated from p loadings are used. Each combination of loadings in the given pool defines a different linear multivariate function, which is then tested for its accuracy as a descriptor of the correlation between C_{ij} and SSF values by the one-out testing. In this step, five to eight best models that yielded minimal prediction error and an optimal multivariate regression coefficient in fitting are found from the complete tested set and saved. These prediction errors for all stored models are then plotted against the number n of loadings C_{ij} or c_{ij} in the multiple linear regression models. The model with the best prediction standard deviation is used for the final processing of all proteins, including the unknown. If spectral representation is needed, selected optimal regression model parameters are used as coefficients in the linear combination of the subspectra selected in the optimization step to predict a

given secondary structural type. The advantage of this method is that there is no additional approximation in the decomposition of the unknown spectra. It is also possible to process more than one unknown spectrum simultaneously. The algorithm is fast and allows for optimizations including the complete search for combinations of as many as 16 subspectral coefficients in reasonable time. For this size of the pool, 65 535 regression calculations are solved. The method explicitly assumes that the subspectra needed for the experimental CD spectra reconstruction are not all related to SSFs.[69,70] Figure 26 graphically represents the results of component selection in the analysis of protein ECD and VCD spectra. Optimal prediction of all SSFs from ECD spectra requires a model with only one coefficient related to the most significant subspectrum. A model with four subspectral coefficients optimally predicts the protein helical and strand SSFs from VCD spectra. In both cases, the one-out method of testing proved that the incorporation of additional loadings into the models actually deteriorates the prediction error.

4.5.6 *Neural Network Methods*

Neural network algorithms are processing programs that create spectra–structure mapping by optimization of neuron potentials and weights of synapses between the

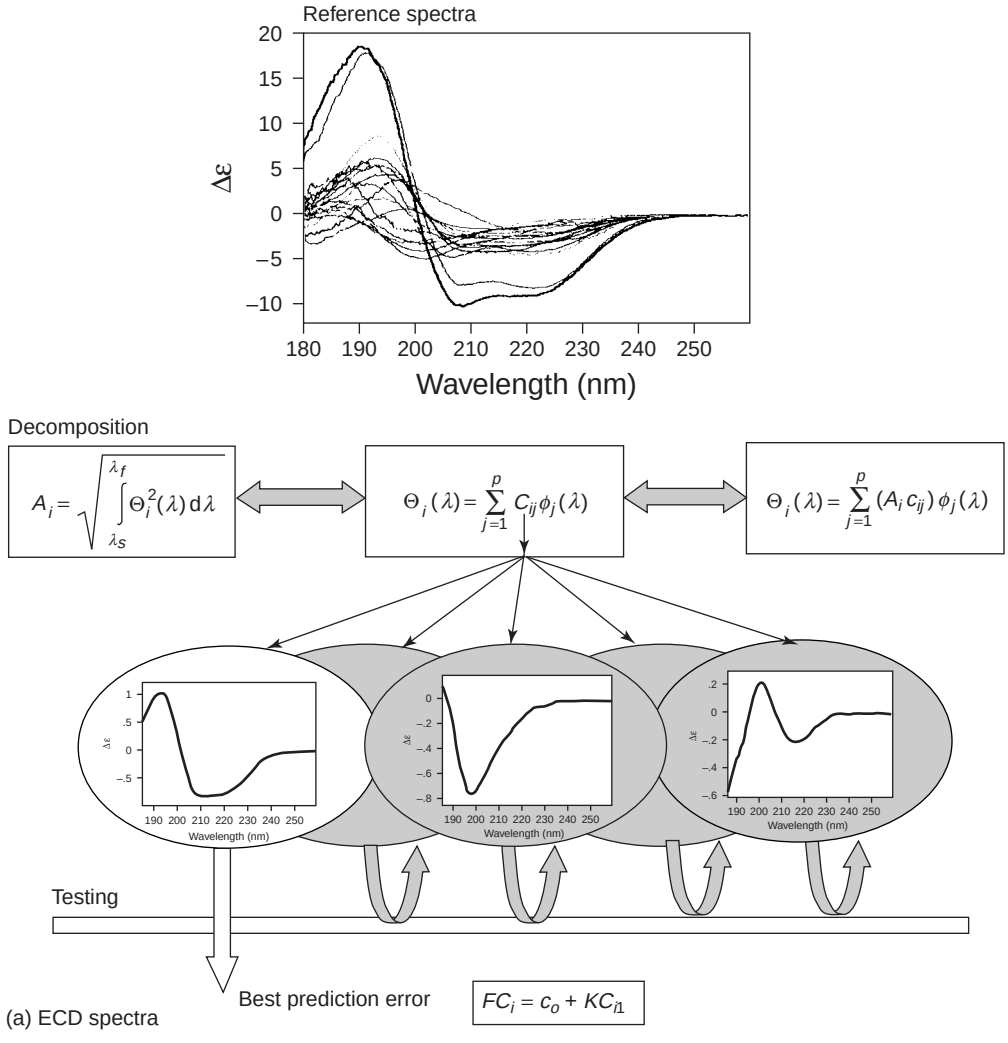

Figure 26 Results of the selective multiple regression between the principal component loadings of ECD and VCD spectra and sheet SSFs in proteins. Reference ECD (a) and VCD spectra (b) are decomposed into principal components that describe the contributions of orthogonal subspectra to the reconstruction of every experimental CD spectrum. Note that this decomposition facilitates the separation of intensity (norm A_i, Equation 15) and bandshape variation (loadings c_{ij} or C_{ij} in Equation 14). Loadings c_{ij} or C_{ij} are tested against their correlation to the sheet SSF in all possible combinations of linear multiple regression functions. Only the model that yields the best prediction error is retained. In the ECD analysis (a), loadings of only one subspectrum and inclusion of the norm A_i are needed for such an optimal relationship. For VCD (b), loadings of three subspectra are found to provide optimal predictions of the sheet fraction. Inclusion of additional components into the prediction equations substantially deteriorates the error.

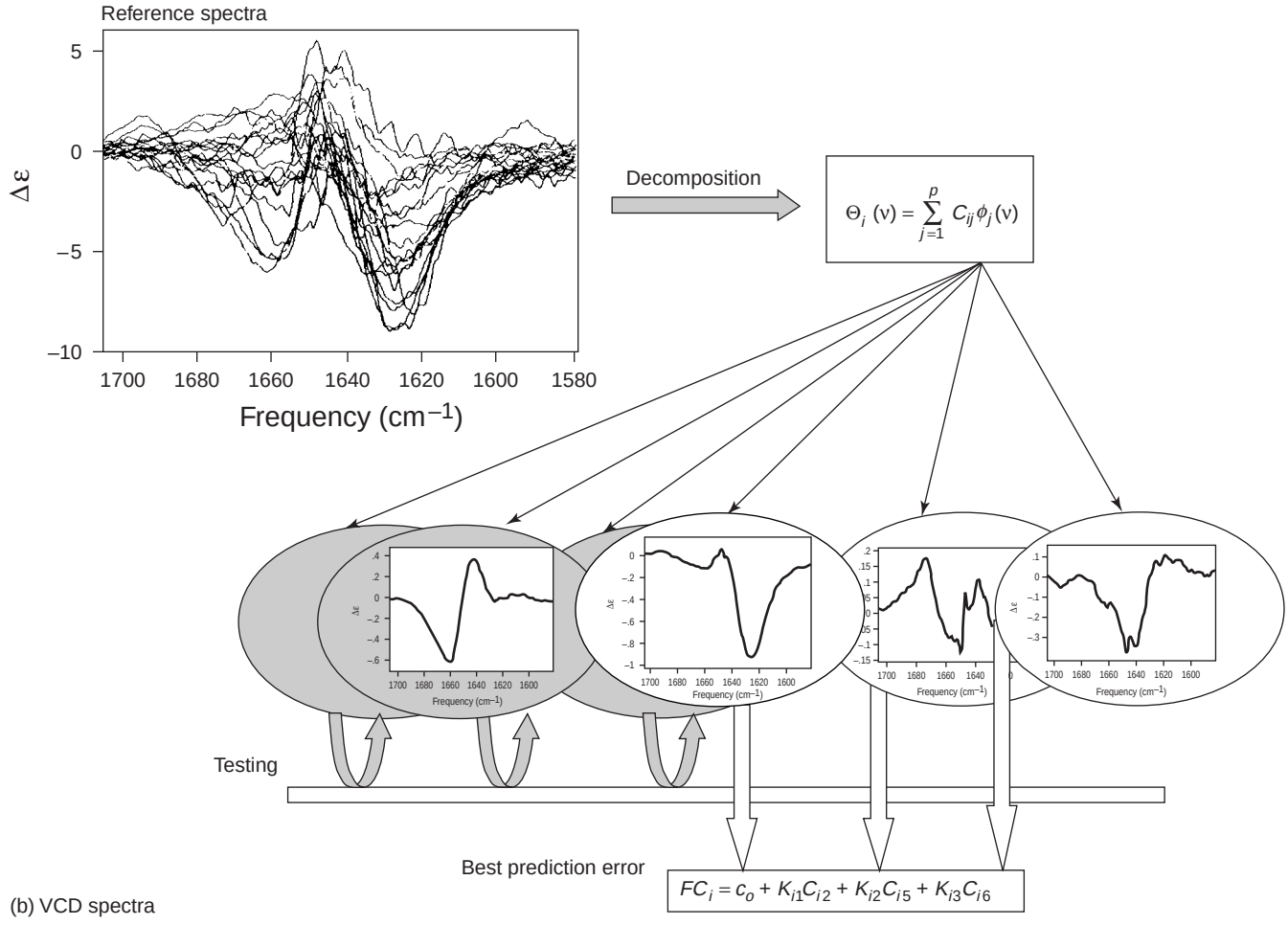

Figure 26 (*Continued*)

calculational nodes (neurons). The backpropagation and Kohonen self-organization map in conjunction with the prediction network are paradigms used to process the protein CD spectra.[132,134,143] Although the details of the decomposition of experimental spectra into subspectra are less transparent than in the above methods, subspectra can be represented either by the exported values of synaptic weight (as discussed in section 4.3) or by the interchange of input and output. The network can be trained to recognize spectra from the structural descriptor values. The subspectra related to particular structural types can be predicted by submitting a descriptor of pure conformation (e.g. 100% helical protein) to the trained network. The spectrum of the unknown protein cannot be processed together with the reference set. The method does not use any constraints on calculated descriptor values. Spectroscopic interpretation of the network mapping mathematics is similar to that described in section 4.3.

In all neural network methods for prediction of SSFs from the ECD spectra, the frequency interval used for analysis can be selected.[144] Another neural network backpropagation scheme has been used to predict elements of matrix descriptor for protein secondary structure segments (section 4.3).[69] This approach is an alternative to determining the number of secondary structure segments using the self-consistent method optimized to predict fractions of distorted secondary structure (Figure 18). Optimal performance is achieved when the unknown protein CD spectrum is first assigned to a subcategory of the reference set proteins. This assignment is based on clustering of the analyzed spectrum with those from the reference set proteins by using the SSF composition of the unknown protein that was previously determined from its spectrum. The improved prediction performance is achieved by using a re-trained network that minimizes the prediction error for protein with the CD spectrum most similar to the unknown one. The matrix descriptor provides in the diagonal elements equivalent structural information to that of the self-consistent method; the off-diagonal elements are information that is additional and complementary. The advantage of this descriptor is in its independence of the number of amino acid residues considered as distorted.

5 QUANTITATIVE RELATION OF CIRCULAR DICHROISM SPECTRAL INFORMATION TO OTHER SPECTROSCOPIC TECHNIQUES

5.1 Two-dimensional Correlation Analysis of Series of Circular Dichroism Spectra

The general features of the established two-dimensional correlation method were used for a comparative analysis of independent sets of protein optical spectra with respect to an ''internal'' structural perturbation.[145–147] The method used the fractional compositions of three major protein secondary structure types as the perturbation for obtaining heterospectral correlation maps for ECD, FTIR, and Raman spectra of the same set of proteins. The approach used a polynomial fit to the intensity variance at each frequency to generate the heterospectral correlation two-dimensional maps. This method generates an arbitrary number of points in the second dimension that represents the averaged response of the spectra to secondary structure variation in the set proteins. It emphasizes any correlations that arise while effectively eliminating or minimizing any asynchronous responses by appropriate rearrangement of the analyzed spectra. The algorithm improved the spectral resolution and generated new assignment of the spectral regions, which were mutually correlated through the protein structural feature used as a perturbation. The method is applicable for any spectroscopic study that results in statically measured reference sets of spectra that vary with respect to some other physical parameter, such as higher-level structure or even a biomedical diagnostic variable. An example of the heterospectral two-dimensional synchronous correlation map that uses variation of ECD spectral intensities and signs to identify bands in protein Raman spectra with different responses to the variation of helical and sheet SSF is shown in Figure 27.

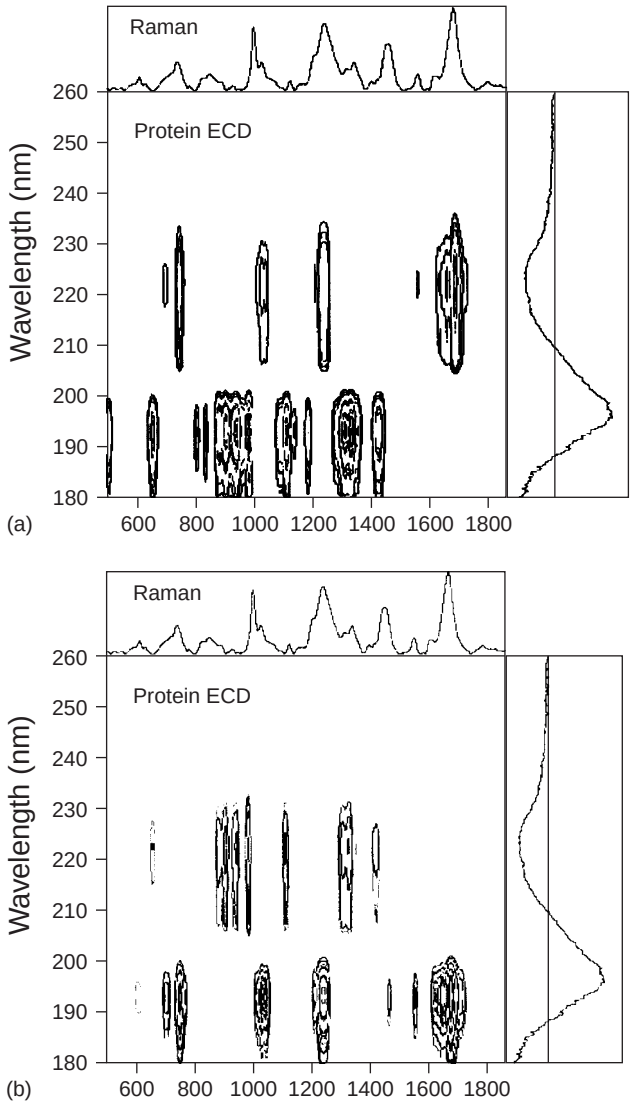

Figure 27 Two-dimensional heterocorrelation between the protein ECD and Raman spectral changes related to the variation of the helical SSF in the reference set. Positive (a) and negative (b) parts of the two-dimensional synchronous heterocorrelation map between protein ECD and Raman spectra calculated with respect to the change of helix fraction as the internal spectral perturbation. Extrema of these correlations indicate the Raman bands that are dependent on the helical content in the protein structure.

5.2 Neural Network Projections Between Circular Dichroism and Other Methods

The degree of interconvertibility of VCD spectra of proteins into ECD has been studied using heteroassociative rather than autoassociative neural networks.[123] In the calculation, the input layer neurons were loaded with either ECD or VCD normalized equidistant intensities. The network topology was then optimized in a two-step training procedure to predict the spectral bandshape of the protein spectrum measured by the complementary method. The optimal number of hidden layer neurons was determined by PCA of the exported weights, and by analysis of the network response to input variations. The results demonstrated that the two spectroscopic methods are not fully independent. Not all spectral features of VCD spectra were predictable from their ECD counterparts. The higher variability of the VCD bandshape, which stems from it being a vibrational mode with intrinsically higher-resolution components, leads to the higher sensitivity of VCD to structural variation. Similar studies of VCD spectra measured in D_2O and H_2O do not show the trends that occur in the ECD/VCD mapping and, therefore, confirm that measurement of vibrational spectra in either solution yielded equivalent structural sensitivities.

6 CONCLUSIONS

The presented discussion shows that CD spectroscopy is not only a technique for the analysis of secondary structure of proteins (as is the common perception of this method) but also a valuable spectroscopic tool for analytical applications. It encompasses multiple aspects of biomolecular properties, which can be studied through their interaction with CPL. With the new developments in instrumental, theoretical, and interpretive applications of the chiroptical methods, CD spectroscopy remains an invaluable tool in any bioanalytical study.

ACKNOWLEDGMENTS

The bulk of this work has been supported by a grant from the National Institutes of Health (GM30147), a Joint National Science Foundation grant between the Charles University Prague and the University of Illinois at Chicago (NSF INT 91-07588), a grant from the Grant Agency of Czech Republic (GACR 203-93-0714), and a Charles University Grant (GAUK 302). The author would like to thank Professor Timothy A. Keiderling and the Department of Chemistry at the University of Illinois at Chicago for continuing support. Cooperation with the Center for Discrete Mathematics and Computer Science (DIMATIA) at the Charles University Prague (Professor Jaroslav Nešetřil) was instrumental in developing many of the topics covered in this review. The assistance of staff at Olis Instruments Inc. (Atlanta, GA) in technical information exchange was also appreciated.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
ADC	Analog-to-digital Converter
CD	Circular Dichroism
CPL	Circularly Polarized Light
DC	Direct Current
ECD	Electronic Circular Dichroism
FDCD	Fluorescence-detected Circular Dichroism
FT	Fourier Transform
FTIR	Fourier Transform Infrared
HPLC	High-performance Liquid Chromatography
IR	Infrared
LCPL	Left Circularly Polarized Light
MCD	Magnetic Circular Dichroism
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
RCPL	Right Circularly Polarized Light
ROA	Raman Optical Activity
S/N	Signal-to-noise Ratio
SSF	Secondary Structure Fraction
UV	Ultraviolet
VCD	Vibrational Circular Dichroism
VIS	Visible
ZPD	Zero Path Difference

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy in Microbiology • Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Forensic Science **(Volume 5)**
Chiroptical Spectroscopy in Drug Analysis

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Fluorescence Spectroscopy in Peptide and Protein Analysis • Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis • Vibrational Spectroscopy in Drug Discovery, Development and Production

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data

Electronic Absorption and Luminescence **(Volume 12)**
Circular Dichroism and Linear Dichroism

Infrared Spectroscopy **(Volume 12)**
Interpretation of Infrared Spectra, A Practical Approach
• Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

REFERENCES

1. G.D. Fasman (ed.), *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York, 1996.
2. W.C. Johnson, Jr, 'Circular Dichroism Instrumentation', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D. Fasman, Plenum Press, New York, 635–652, 1996.
3. T.A. Keiderling, 'Vibrational Circular Dichroism. Comparison of Technique and Practical Considerations', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, eds. K. Krishnan, J.R. Ferraro, Academic Press, San Diego, CA, 203–284, 1990.
4. L.A. Nafie, T.A. Keiderling, P.J. Stephens, 'Vibrational Circular Dichroism', *J. Am. Chem. Soc.*, **98**, 2715–2723 (1976).
5. J.T. Yang, C.S.C. Wu, H.M. Martinez, 'Calculation of Protein Conformation from Circular Dichroism', *Meth. Enzymol.*, **130**, 208–269 (1986).
6. H.P.J.M. Dekkers, 'Circularly Polarized Luminescence: a Probe for Chirality in the Excited State', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D.Fasman, Plenum Press, New York, 121–152, 1996.
7. L.D. Barron, 'Raman Optical Activity: A Probe of Stereochemistry and Magnetic Structure', in *Vibrational Spectra and Structure*, ed. H.D. Bist, J.R. Durig, J.F. Sullivan, Elsevier, Amsterdam, 343–368, 1989.
8. L.D. Barron, L. Hecht, *Biomolecular Conformational Studies with Vibrational Raman Optical Activity*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, 235–266, 1993.
9. L.D. Barron, H.S.J. Ford, A.F. Bell, G. Wilson, L. Hecht, A. Cooper, 'Vibrational Raman Optical Activity of Biopolymers', *Faraday Discuss. Chem. Soc.*, **99**, 217–232 (1994).
10. L.D. Barron, L. Hecht, S.J. Ford, A.F. Bell, G. Wilson, 'Vibrational Raman Optical Activity of Biopolymers', *J. Mol. Struct.*, **349**, 397–400 (1995).
11. L.D. Barron, L. Hecht, A.D. Bell, 'Vibrational Raman Optical Activity of Biomolecules', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D. Fasman, Plenum Press, New York, 653–695, 1996.
12. L.D. Barron, L. Hecht, A.F. Bell, G. Wilson, 'Recent Developments in Raman Optical Activity of Biopolymers', *Appl. Spectrosc.*, **50**, 619–629 (1996).
13. S.J. Ford, Z.Q. Wen, L. Hecht, L.D. Barron, 'Vibrational Raman Optical of Alanyl Peptide Oligomers: A New Perspective on Solution Conformation', *Biopolymers*, **34**, 303–313 (1994).
14. S.J. Ford, A. Cooper, L. Hecht, G. Wilson, L.D. Barron, 'Vibrational Raman Optical Activity of Lysozyme: Hydrogen–Deuterium Exchange, Unfolding and Ligand Binding', *J. Chem. Soc., Faraday Trans.*, **91**, 2087–2093 (1995).
15. L. Hecht, L.D. Barron, 'Recent Developments in Raman Optical Activity Instrumentation', *Faraday Discuss. Chem. Soc.*, **99**, 35–47 (1994).
16. L.A. Nafie, C.G. Zimba, 'Raman Optical Activity and Related Techniques', in *Biological Applications of Raman Spectroscopy*, ed. T.G. Spiro, John Wiley & Sons, New York, 307–343, 1987.
17. L.A. Nafie, 'Theory and Measurement of Raman Optical Activity', in *Modern Nonlinear Optics*, eds. M. Evans, S. Kielich, John Wiley & Sons, New York, 105–149, Part 3, 1994.
18. G. Wilson, S.J. Ford, A. Cooper, L. Hecht, Z.Q. Wen, L.D. Barron, 'Vibrational Raman Optical Activity of alpha-Lactalbumin: Comparison with Lysozyme, and Evidence for Native Tertiary Folds in Molten Globule States', *J. Mol. Biol.*, **254**, 747–760 (1995).
19. Z.Q. Wen, L. Hecht, L.D. Barron, 'beta-Sheet and Associated Turn Signatures in Vibrational Raman Optical Activity Spectra of Proteins', *Prot. Sci.*, **3**, 435–439 (1994).
20. Z.Q. Wen, L. Hecht, L.D. Barron, 'alpha-Helix and Associated Loop Signatures in Vibrational Raman Optical Activity Spectra of Proteins', *J. Am. Chem. Soc.*, **116**, 443–445 (1994).
21. L.D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1989.
22. D.J. Caldwell, H. Eyring, *The Theory of Optical Activity*, John Wiley & Sons, New York, 1971.
23. A.D. Buckingham, 'The Theoretical Background to Vibrational Optical Activity', *Faraday Discuss. Chem. Soc.*, **99**, 1–12 (1994).
24. I. Tinoco, 'Theory of Optical Activity', *Adv. Chem. Phys.*, **4**, 113–185 (1962).
25. I. Tinoco, C. Bustamante, M.F. Maestre, 'The Optical Activity of Nucleic Acids and their Aggregates', *Annu. Rev. Biophys. Bioeng.*, **9**, 107–141 (1980).
26. E. Charney, *Theory of Chiroptical Methods*, Van Nostrand Reinhold, New York, 1980.

27. J. Zamastil, L. Skala, P. Pančoška, O. Bilek, 'New Approach to the Theory of Circular Dichroism', *Collect. Czech. Chem. Commun.*, **63**, 1187–1201 (1998).
28. R.D. Amos, N.C. Handy, K.J. Jalkanen, P.J. Stephens, 'Efficient Calculation of Vibrational Magnetic Dipole Transition Moments and Rotational Strengths', *Chem. Phys. Lett.*, **133**, 21 (1987).
29. R.D. Amos, K.J. Jalkanen, P.J. Stephens, 'Alternative Formalism for the Calculation of Atomic Polar Tensors and Atomic Axial Tensors', *J. Phys. Chem.*, **92**, 5571 (1988).
30. P.J. Stephens, F.J. Devlin, C.S. Ashvar, C.F. Chabalowski, M.J. Frisch, 'Theoretical Calculation of Vibrational Circular Dichroism Spectra', *Faraday Discuss. Chem. Soc.*, **99**, 103–119 (1994).
31. P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, 'Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra using Density Functional Force Fields', *J. Phys. Chem.*, **98**, 11623–11627 (1994).
32. I. Tinoco, *Radiat. Res.*, **20**, 133 (1963).
33. G. Snatzke, 'Circular Dichroism: An Introduction', in *Circular Dichroism*, eds. K. Nakanishi, N. Berova, R.W. Woody, VCH, New York, 1–38, 1994.
34. R.W. Woody, 'Optical Rotatory Properties of Biopolymers', *J. Polym. Sci., Macmol. Rev.*, **12**, 181–320 (1977).
35. R.W. Woody, 'Theory of Circular Dichroism of Proteins', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D. Fasman, Plenum Press, New York, 25–67, 1996.
36. R.W. Woody, 'Studies of Theoretical Circular Dichroism of Polypeptides: Contributions of Beta Turns', in *Peptides, Polypeptides and Proteins*, eds. M. Goodman, N. Lotan, John Wiley & Sons, New York, 338–350, 1974.
37. J.E. Ridley, M.C. Zerner, *Theor. Chim. Acta*, **42**, 223 (1976).
38. A.D. Bacon, M.C. Zerner, *Theor. Chim. Acta*, **53**, 21 (1979).
39. K. Nakanishi, N. Berova, R.W. Woody, *Circular Dichroism Principles and Applications*, VCH, New York, 1994.
40. E. Ruch, A. Schönhofer, 'Theory of Chirality Functions', *Theor. Chim. Acta*, **19**, 225–301 (1970).
41. E. Ruch, 'Theory of Chirality Functions for Molecules with Chiral Ligands', *Angew. Chem.*, **16**, 65–98 (1977).
42. P. Pančoška, K. Vacek, L. Souckova, 'Spectroscopy and Photoprocesses in Photosynthetic Model Systems', in *Proceedings of the 5th International Conference on Energy and Electron Transfer*, 259–269, 1986.
43. K. Vacek, P. Pančoška, M. Urbanova, L. Bednarova, 'New Trends in Spectroscopic Investigations of Photosynthetic Model Pigment–Protein Complexes', *Opt. Spectrosc. Chem. Biol.*, **22**, 253–257 (1990).
44. R.W. Woody, A.K. Dunker, 'Aromatic and Cystine Side-chain Circular Dichroism in Proteins', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D. Fasman, Plenum Press, New York, 109–157, 1996.
45. L. Skala, P. Pančoška, 'Interpolation Formula for Physical Properties of Polypeptides as a Function of the Number of Amino Acid Residues', *Chem. Phys.*, **125**, 21–30 (1988).
46. C.R. Cantor, P.R. Schimmell, *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function*, W.H. Freeman, San Francisco, CA, 1980.
47. P. Zhou, W.T. Wiesler, Y. Chen, N. Berova, K. Nakanishi, 'Natural Products and the Exciton Chirality Method', in *Proceedings of the 4th International Conference on Circular Dichroism*, eds. H. Klein, G. Snatzke, Ruhr-Universitat Bochum, Bochum, 1–17, 1991.
48. P.M. Bayley, *Prog. Biophys. Mol. Biol.*, **27**, 1–76 (1973).
49. M. Diem, M.R. Oboodi, C. Alva, 'Determination of Peptide Conformation via Vibrational Coupling: Application to Diastereoisomeric Alanyl Dipeptides', *Biopolymers*, **23**, 1917–1930 (1984).
50. M. Diem, 'Determining the Solution Conformation of Biological Moleculars by Infrared Circular Dichroism', *Spectroscopy*, **10**, 38–43 (1995).
51. W. Zhong, M. Gulotta, D.J. Goss, M. Diem, 'Infrared VCD in DNA Models. II. Experimental and Theoretical Results for B-Family Polymers', *Biochemistry*, **29**, 7485–7491 (1990).
52. P. Bour, T.A. Keiderling, 'Computational Evaluation of the Coupled Oscillator Model in the Vibrational Circular Dichroism of Selected Small Molecules', *J. Am. Chem. Soc.*, **114**, 9100–9105 (1992).
53. P. Bour, 'Computational Study of the Vibrational Optical Activity of Amides and Peptides', Ph.D. Thesis, Academy of Science, Prague, Czech Republic, 1993.
54. P. Bour, T.A. Keiderling, 'Ab Initio Simulation of the Vibrational Circular Dichroism of Coupled Peptides', *J. Am. Chem. Soc.*, **115**, 9602–9607 (1993).
55. P. Bour, T.A. Keiderling, P. Malon, 'Vibrational Circular Dichroism as a Tool for Determination of Peptide Secondary Structure', in *Peptides*, ed. H.L.L. Maia, Escom, Leiden, 517–518, 1995.
56. L.P. Wang, P. Pančoška, T.A. Keiderling, *Detection of Triple Helical Nucleic Acids with Vibrational Circular Dichroism*, Kluwer, Dordrecht, 1-81-2, 1993.
57. W.C. Johnson, Jr, 'CD of Nucleic Acids', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D. Fasman, Plenum Press, New York, 523–545, 1996.
58. J. Sopkova, V.M. Jacques, P. Pančoška, A. Lewit-Bentley, 'The Dynamic Behavior of Annexin V as a Function of Calcium Ion Binding: A Circular Dichroism, UV Absorption, and Steady-state and Time-resolved Fluorescence Study', *Biochemistry*, **33**, 4490–4499 (1994).
59. T.B. Freedman, L.A. Nafie, 'Theoretical Formalism and Models for Vibrational Circular Dichroism Intensity',

in *Modern Nonlinear Optics*, eds. M. Evans, S. Kielich, John Wiley & Sons, New York, 207–263, 1994.
60. T.B. Freedman, L.A. Nafie, 'Stereochemical Aspects of Vibrational Optical Activity', *Top. Stereochem.*, **17**, 113–206 (1987).
61. L.A. Nafie, 'Infrared and Raman Vibrational Optical Activity: Theoretical and Experimental Aspects', *Annu. Rev. Phys. Chem.*, **48**, 357–386 (1997).
62. A.M. Bracete, S. Kadkhodayan, M. Sono, A.M. Huff, C.F. Zhuang, D.K. Cooper, K.M. Smith, C.K. Chang, J.H. Dawson, 'Iron Chlorin-reconstituted Histidine-ligated Heme Proteins as Models for Naturally Occurring Iron Chlorin Proteins – Magnetic Circular Dichroism Spectroscopy as a Probe of Iron Chlorin Coordination Structure', *Inorg. Chem.*, **33**, 5042–5049 (1994).
63. M. Pawlikowski, T.A. Keiderling, 'Vibronic Coupling Effects in Magnetic Vibrational Circular Dichroism. A Model Formalism for Doubly Degenerate States', *J. Chem. Phys.*, **81**, 4765–4773 (1984).
64. J. Hladik, P. Pančoška, D. Sofrova, 'The Induced Chirality of Carotenoids and their Influence on the Conformation of Chlorophyll–Protein Complexes Isolated from the Cyanobacterium *Plectonema boryanum*. Absorption and Circular Dichroism Study', *Photosynthesis in Prokaryotes; Cell Differentiation and Function, Proc. Spec. FEBS Meeting*, ed. G.C.P. Papageorgiou, Lester, Elsevier, New York, 103–118, 1982.
65. J. Hladik, P. Pančoška, D. Sofrova, 'The Influence of Carotenoids on the Conformation of Chlorophyll–Protein Complexes Isolated from the Cyanobacterium *Plectonema boryanum*. Absorption and Circular Dichroism Study', *Biochim. Biophys. Acta*, **681**, 263–272 (1982).
66. P. Pančoška, M. Urbanova, L. Bednarova, K. Vacek, V.Z. Paschenko, S. Vasiliev, P. Malon, M. Kral, 'Models of Pigment–Protein Interactions in Photosynthetic Systems: Tetraphenylporphyrin Complexes with Polycationic Sequential Polypeptides. Absorption, Circular Dichroism and Fluorescence Properties', *Chem. Phys.*, **147**, 401–413 (1990).
67. P. Pančoška, L. Bednarova, V. Kapsa, 'Chirality of Ordered Molecular Complexes Consisting of Protein with Fixed Prosthetic Group', *Chem. Phys. Lett.*, **191**, 603–608 (1992).
68. T.A. Keiderling, P. Pančoška, 'Structural Studies of Biological Macromolecules Using Vibrational Circular Dichroism', in *Biomolecular Spectroscopy*, eds. R.E. Hester, R.J.H. Clarke, John Wiley & Sons, Chichester, Part B, 267–315, 1993.
69. P. Pančoška, V. Janota, T.A. Keiderling, 'Novel Matrix Descriptor for Secondary Structure Segments in Proteins: Demonstration of Predictability from Circular Dichroism Spectra', *Anal. Biochem.*, **267**, 72–83 (1999).
70. P. Pančoška, V. Janota, J. Nesetril, T.A. Keiderling, 'Novel Matrix Descriptor for Spectroscopic Determination of the Connectivity of Secondary Structure Segments in Proteins. Analysis of General Properties using Graph Theory', *KAM-DIMATIA Ser.*, **97–363**, 1–50 (1997).
71. T.A. Keiderling, 'Vibrational Circular Dichroism', *Appl. Spectrosc. Rev.*, **17**, 189–226 (1981).
72. T.A. Keiderling, S.C. Yasui, U. Narayanan, A. Annamalai, P. Malon, R. Kobrinskaya, L. Yang, 'Vibrational Circular Dichroism of Biopolymers', in *Spectroscopy of Biological Molecules: New Advances*, eds. E.D. Schmid, F.W. Schneider, F. Siebert, John Wiley & Sons, New York, 73–76, 1987.
73. T.A. Keiderling, S.C. Yasui, P. Malon, P. Lančoška, R.K. Dukor, P.V. Croatto, L. Yang, 'Techniques and Application of FTIR Vibrational Circular Dichroism', in *Proceedings of the 7th SPIE International Conference on Fourier Transform Spectroscopy*, ed. D.G. Cameron, 57–63, 1989.
74. T.A. Keiderling, P. Pančoška, 'Protein Conformational Studies with Vibrational Circular Dichroism', in *Spectroscopy of Biological Molecules*, eds. R.E. Hester, R.B. Girling, Royal Society of Chemistry, Cambridge, 121–124, 1991.
75. T.A. Keiderling, P. Pančoška, S.C. Yasui, M. Urbanova, R.K. Dukor, 'Conformational Studies of Proteins Using Vibrational Circular Dichroism', in *Proteins: Structure, Dynamics, Design*, eds. V. Renugopalakrishnan, P.R. Carey, I.C.P. Smith, S.G. Guang, A.C. Storer, Escom, Leiden, 165–170, 1991.
76. T.A. Keiderling, 'Vibrational Circular Dichroism of Peptides and Proteins', in *Infrared and Raman Spectroscopy of Biological Materials*, eds. B. Yan, H.U. Gremlich, M. Dekker, New York, 1999.
77. T.A. Keiderling, 'Vibrational Circular Dichroism Applications to Conformational Analysis of Biomolecules', in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G.D. Fasman, Plenum, New York, 555–598, 1996.
78. T.A. Keiderling, 'Vibrational Circular Dichroism Applications in Proteins and Peptides', in *NATO ASI in Biomolecular Structure and Dynamics*, ed. G. Vergoten, Loutraki, Greece, 1996.
79. T.A. Keiderling, 'Vibrational Circular Dichroism Spectroscopy of Peptides and Proteins', in *Circular Dichroism, Interpretation and Applications*, eds. K. Nakanishi, N. Berova, R. Woody, VCH, New York, 497–521, 1994.
80. T.A. Keiderling, B.L. Wang, M. Urbanova, P. Pančoška, R.K. Dukor, 'Empirical Studies of Protein Secondary Structure by Vibrational Circular Dichroism and Related Techniques – alpha-Lactalbumin and Lysozyme as Examples', *Faraday Discuss. Chem. Soc.*, 263–285 (1994).
81. T.A. Keiderling, 'Protein Structural Studies using Vibrational Circular Dichroism Spectroscopy', in *Determination of Protein Structure in Solution by Spectroscopic Methods*, ed. H. Havel, VCH, New York, 1995.

82. M. Diem, G.M. Roberts, O. Lee, O. Barlow, 'Design and Performance of an Optimized Dispersive Infrared Dichrograph', *Appl. Spectrosc.*, **42**, 20–27 (1988).
83. P. Xie, M. Diem, 'The Measurement of Dispersive Vibrational CD: Signal Optimization and Artifact Reduction', *Appl. Spectrosc.*, **50**, 675–680 (1996).
84. L.A. Nafie, 'Experimental and Theoretical Advances in Vibrational Optical Activity', *Adv. Infrared Raman Spectrosc.*, **11**, 49–93 (1984).
85. B. Baello, S. Stelea, P. Pančoška, T.A. Keiderling, 'Secondary Structure Fractions Using D–H exchange', unpublished results, 1999.
86. P. Malon, T.A. Keiderling, 'Spinning Quarter-wave Plate Polarization Modulator: Test of Feasibility for Vibrational Circular Dichroism Measurements', *Appl. Spectrosc.*, **50**, 669–674 (1996).
87. P. Malon, T.A. Keiderling, 'Theoretical Simulation of a Polarization Modulator based on Mechanically Rotating a Polarizing Element', *Appl. Optics*, **36**, 6141–6148 (1997).
88. C.C. Chen, P.L. Polavarapu, S. Weibel, 'New Design for Fourier Transform Infrared Vibrational Circular Dichroism Spectrometers', *Appl. Spectrosc.*, **48**, 1218–1223 (1994).
89. P.L. Polavarapu, *Vibrational Spectra: Principles and Applications with Emphasis on Optical Activity*, Elsevier, Amsterdam, 1998.
90. D. Cheng, J. Tallent, R.J. DeSa, 'High-speed Digital CD Spectropolarimeter', *Appl. Spectrosc.*, in press, 1999.
91. P. Malon, T.A. Keiderling, 'Fourier Transform Infrared Circular Dichroism and the Absorption Artifact Problem. Experimental Observations', *Appl. Spectrosc.*, **42**, 32–38 (1988).
92. R.K. Yoo, B. Wang, P.V. Croatto, T.A. Keiderling, 'Method for the Determination of the Zero Path Difference Position in a Fourier Transform Infrared Spectrometer. Application to Magnetic Vibrational Circular Dichroism', *Appl. Spectrosc.*, **45**, 231–236 (1991).
93. B. Wang, T.A. Keiderling, 'Observations on the Measurement of Vibrational Circular Dichroism with Rapid-scan and Step-scan FT–IR Techniques', *Appl. Spectrosc.*, **49**, 1347–1355 (1995).
94. V. Baumruk, T.A. Keiderling, 'Vibrational Circular Dichroism of Proteins in H_2O Solution', *J. Am. Chem. Soc.*, **115**, 6939–6942 (1993).
95. L.J. Wang, L.G. Yang, T.A. Keiderling, 'Vibrational Circular Dichroism of A-, B-, and Z-form Nucleic Acids in the PO(2)(–)Stretching Region', *Biophys. J.*, **67**, 2460–2467 (1994).
96. L. Wang, T.A. Keiderling, 'Vibrational Circular Dichroism Studies of the A-to-B Conformational Transition in DNA', *Biochemistry*, **31**, 10 265–10 271 (1992).
97. L. Wang, P. Pančoška, T.A. Keiderling, 'Detection and Characterization of Triple-helical Pyrimidine–Purine–Pyrimidine Nucleic Acids with Vibrational Circular Dichroism', *Biochemistry*, **33**, 8428–8435 (1994).
98. S.Y. Venyaminov, I.A. Baikalov, C.-S. Wu, J.T. Yang, 'Some Problems of CD Analysis of Protein Conformation', *Anal. Biochem.*, **198**, 250–255 (1991).
99. P. Pančoška, S.C. Yasui, T.A. Keiderling, 'Enhanced Sensitivity to Conformation in Various Proteins. Vibrational Circular Dichroism Results', *Biochemistry*, **28**, 5917–5923 (1989).
100. P. Pančoška, T.A. Keiderling, 'Systematic Comparison of Statistical Analyses of Electronic and Vibrational Circular Dichroism for Secondary Structure Prediction of Selected Proteins', *Biochemistry*, **30**, 6885–6895 (1991).
101. W. Kabsch, C. Sander, 'Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features', *Biopolymers*, **22**, 2577–2637 (1983).
102. M. Levitt, J. Greer, 'Automatic Identification of Secondary Structures in Globular Proteins', *J. Mol. Biol.*, **114**, 181–293 (1977).
103. W.C. Johnson, Jr, 'Secondary Structure of Proteins through Circular Dichroism Spectroscopy', *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 145–166 (1988).
104. W.C. Johnson, 'Analyzing the CD of Proteins for Secondary Structure', in *Molecular Chirality in Chemistry and Life Sciences*, CNR, Pisa, Italy, 15, 1997.
105. W.C. Johnson, 'Protein Secondary Structure and Circular Dichroism – A Practical Guide', *Proteins*, **7**, 205–214 (1990).
106. W.C. Johnson, 'Analysis of Circular-dichroism Spectra', *Meth. Enzymol.*, **210**, 426–447 (1992).
107. N. Greenfield, G.D. Fasman, 'Computed Circular Dichroism Spectra for the Evaluation of Protein Secondary Structure', *Biochemistry*, **8**, 4108–4116 (1969).
108. N.J. Greenfield, 'Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data', *Anal. Biochem.*, **235**, 1–10 (1996).
109. N. Sreerama, S.Y. Venyaminov, R.W. Woody, 'Estimation of the Number of α-Helical and β-Sheet Strand Segments in Proteins Using Circular Dichroism Spectroscopy', *Prot. Sci.*, **8**, 370–380 (1999).
110. E.R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons, New York, 1991.
111. R.J. Rummel, *Applied Factor Analysis*, Northwestern University Press, Evanston, IL, 1970.
112. P. Pančoška, E. Bitto, V. Janota, M. Urbanova, V.P. Gupta, T.A. Kiederling, 'Comparison of and Limits of Accuracy for Statistical Analyses of Vibrational and Electronic Circular Dichroism Spectra in Terms of Correlations to and Predictions of Protein Secondary Structure', *Protein Sci.*, **4**, 1384–1401 (1995).
113. R. Pribic, 'Principal Component Analysis of Fourier Transform Infrared and/or Circular Dichroism Spectra of Proteins Applied in a Calibration of Protein Secondary Structure', *Anal. Biochem.*, **223**, 26–34 (1994).
114. V. Baumruk, P. Pančoška, T.A. Keiderling, 'Predictions of Secondary Structure Using Statistical Analyses of

Electronic and Vibrational Circular Dichroism and Fourier Transform Infrared Spectra of Proteins in H_2O', *J. Mol. Biol.*, **259**, 774–791 (1996).
115. M.A. Sharaf, D.L. Illman, B.R. Kowalski, *Chemometrics*, John Wiley & Sons, New York, 1986.
116. U. Hobohm, M. Scharf, C. Sander, 'Representative Set of Protein Structures for Statistical Studies', *Prot. Sci.*, **1**, 409–417 (1992).
117. P. Pančoška, E. Bitto, V. Janota, T.A. Keiderling, 'Quantitative Analysis of Vibrational Circular Dichroism Spectra of Proteins. Problems and Perspectives', *Faraday Discuss. Chem. Soc.*, **99**, 287–310 (1995).
118. P. Pančoška, M. Blazek, T.A. Keiderling, 'Relationships Between Secondary Structure Fractions for Globular Proteins. Neural Network Analyses of Crystallographic Data Sets', *Biochemistry*, **31**, 10250–10257 (1992).
119. P.K. Simpson, *A Review of Artificial Neural Systems II: Paradigms, Applications and Implementations*, General Dynamics, San Diego, CA, 1988.
120. P. Pančoška, V. Janota, T.A. Keiderling, 'Interconvertibility of Electronic and Vibrational Circular Dichroism Spectra of Proteins: A Test of Principle Using Neural Network Mapping', *Appl. Spectrosc.*, **50**, 658–668 (1996).
121. B.I. Baello, P. Pančoška, T.A. Keiderling, 'Vibrational Circular Dichroism Spectra of Proteins in the Amide III Region. Measurement and Correlation of Bandshape to Secondary Structure', *Anal. Biochem.*, **250**, 212–221 (1997).
122. S.Y. Venyaminov, K.S. Vassilenko, 'Determination of Protein Tertiary Structure Class from Circular Dichroism Spectra', *Anal. Biochem.*, **222**, 176–184 (1994).
123. P. Pančoška, L. Wang, T.A. Keiderling, 'Frequency Analysis of Infrared Absorption and Vibrational Circular Dichroism of Proteins in D_2O Solution', *Prot. Sci.*, **2**, 411–419 (1993).
124. S.W. Provencher, J. Glockner, 'Estimation of Globular Protein Secondary Structure from Circular Dichroism', *Biochemistry*, **20**, 33–37 (1981).
125. P. Pančoška, I. Fric, K. Blaha, 'Modified Factor Analysis of the Circular Dichroism Spectra, Applied to a Series of Cyclodipeptides Containing L-Proline', *Collect. Czech. Chem. Commun.*, **44**, 1296–1311 (1979).
126. P. Pančoška, I. Fric, K. Blaha, 'Statistical Accuracy Test in Measurements of Circular Dichroism', *Collect. Czech. Chem. Commun.*, **44**, 1312–1317 (1979).
127. N. Sreerama, R.W. Woody, 'A Self-consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism', *Anal. Biochem.*, **209**, 32–44 (1993).
128. P. Manavalan, W.C. Johnson, Jr, 'Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular Dichroism Spectra', *Anal. Biochem.*, **167**, 76–85 (1987).
129. J.P. Hennessey, W.C. Johnson, Jr, 'Information Content in the Circular Dichroism of Proteins', *Biochemistry*, **20**, 1085–1094 (1981).
130. A. Toumadje, S.W. Alcorn, W.C. Johnson, Jr, 'Extending CD Spectra of Proteins to 168 nm Improves the Analysis for Secondary Structures', *Anal. Biochem.*, **200**, 321–331 (1992).
131. W.C. Johnson, Jr, 'Circular Dichroism and its Empirical Application to Biopolymers', *Meth. Biochem. Anal.*, **31**, 61–163 (1985).
132. N. Sreerama, R.W. Woody, 'Protein Secondary Structure from Circular Dichroism Spectroscopy – Combining Variable Selection Principle and Cluster Analyses with Neural Network, Ridge Regression and Self-consistent Methods', *J. Mol. Biol.*, **242**, 497–507 (1994).
133. A. Perczel, M. Hollosi, G. Tusnady, G.D. Fasman, 'Convex Constraint Analysis: A Natural Deconvolution of Circular Dichroism Curves of Proteins', *Prot. Eng.*, **4**, 669–679 (1991).
134. M.A. Andrade, P. Chacon, J.J. Merelo, F. Moran, 'Evaluation of Secondary Structure of Proteins from UV Circular Dichroism Spectra using an Unsupervised Learning Neural Network', *Prot. Eng.*, **6**, 383–390 (1993).
135. P. Pančoška, H. Fabian, G. Yoder, V. Baumruk, T.A. Keiderling, 'Protein Structural Segments and Their Interconnections Derived from Optical Spectra – Thermal Unfolding of Ribonuclease T-1 as an Example', *Biochemistry*, **35**, 13094–13106 (1996).
136. P. Pančoška, S.C. Yasui, T.A. Keiderling, 'Statistical Analysis of the Vibrational Circular Dichroism of Selected Proteins and Relationship to Secondary Structures', *Biochemistry*, **30**, 5089–5103 (1991).
137. M.C. Manning, R.W. Woody, 'Theoretical CD Studies of Polypeptide Helices: Examination of Important Electronic and Geometric Factors', *Biopolymers*, **31**, 569–586 (1991).
138. P. Pančoška, K. Blaha, 'Modified Facto: Analysis of the Circular Dichroism Spectra', in *2nd Czechoslovak–Polish Colloquium in Chemical Thermodynamics, Physical and Organic Chemistry*, ed. M. Siemion, Warsaw Technical University, Warsaw, Poland, 223–228, 1980.
139. D.L. Massart, L. Kaufman, 'The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis', in *Chemical Analysis*, John Wiley & Sons, New York, 1983.
140. L. Kaufman, D.L. Massart, 'Cluster Analysis', in *Chemometrics. Mathematics and Statistics in Chemistry*, ed. B.R. Kowalski, D. Reidel, New York, 393–401, 1984.
141. S. Brahms, J. Brahms, 'Determination of Protein Secondary Structure in Solution by Vacuum Ultraviolet Circular Dichroism', *J. Mol. Biol.*, **138**, 149–178 (1980).
142. R. Pribic, I.H.M van Stokkum, D. Chapman, P.T. Haris, M. Bloemendal, 'Protein Secondary Structure from Fourier Transform Infrared and/or Circular Dichroism Spectra', *Anal. Biochem.*, **214**, 366–378 (1993).
143. B. Dalmas, G.J. Hunter, W.H. Bannister, 'Prediction of Protein Secondary Structure from Circular Dichroism Spectra using Artificial Neural Network Techniques', *Biochem. Mol. Biol. Int.*, **34**, 17–26 (1994).

144. G. Bohm, R. Muhr, R. Jaenicke, 'Quantitative Analysis of Protein Far UV Circular Dichroism Spectra by Neural Networks', *Prot. Eng.*, **5**, 191–195 (1992).
145. J. Kubelka, P. Pančoška, T.A. Keiderling, 'Novel Use of a Static Modification of Two Dimensional Correlation Analysis II: Heterospectral Correlations of Protein Raman, FTIR and Circular Dichroism Spectra', *Appl. Spectrosc.*, **53**, 655–665 (1998).
146. J. Kubelka, P. Pančoška, T.A. Keiderling, 'Protein Spectra–Structure Relationship by Novel Modification of 2D Correlation Analysis. VCD Applications', in *Spectroscopy of Biological Molecules New Directions*, eds. J. Greve, G.J. Puppells, C. Otto, KAP, Dordrecht, 67–68, 1999.
147. P. Pančoška, J. Kubelka, T. Keiderling, 'Novel Use of a Static Modification of Two-dimensional Correlation Analysis. I. Comparison of the Secondary Structure Sensitivity of Raman, Electronic Circular Dichroism and FTIR Spectra of Proteins', *Appl. Spectrosc.*, **53**, 666–671 (1999).

Fluorescence-based Biosensors

John D. Brennan
McMaster University, Hamilton, Canada

1 Introduction 1
2 History 2
3 Instrumentation 2
4 Interfacing of Biomolecules to Devices 3
5 Applications of Fluorescence-based Biosensors 4
5.1 Applications of Protein Engineering to Sensor Design 4
5.2 Enzyme-based Sensors 4
5.3 Fluoroimmunosensors 6
5.4 Regulatory Proteins 9
5.5 Membrane-associated Proteins 12
5.6 DNA/RNA-based Genosensors 13
5.7 Multianalyte Sensing with Fluorescent Sensor Arrays 14
6 Comparison with other Methods 14
7 Summary and Future Prospects 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Biosensors, as defined by Biosensors and Bioelectronics*, are analytical devices ''incorporating a biological material, a biologically derived material or biomimetic intimately associated with or integrated within a physicochemical transducer or transducing microsystem''. Fluorescence-based biosensors are those devices that derive an analytical signal from the fluorescence emission process. Such biosensors may be used for a wide variety of tasks, including: detection of compounds of biomedical,[1] environmental[2] or defense interest,[3] on-line monitoring for process control,[4] monitoring of foodstuffs,[5] selective detection of compounds undergoing a chemical separation,[6] and screening of drug compounds.[7] Advantages of such devices include:[8] high selectivity; rapid response times; reusability; amenability to remote analysis; and immunity from electrical interferences. The selective nature of the complexation between the biomolecule and the analyte, combined with the small size of the device and the advantages of total internal reflection (TIR)-based spectroscopy,[9] also results in an ability to measure analytes in complex matrixes. Such samples may include highly scattering systems such as milk or whole blood,[10] or relatively inaccessible locations such as groundwater wells, or even intracellular environments.[11]*

The key limitation of such devices mainly centers around the poor stability of biological compounds, which can lead to a substantial drift in instrument response over time. There is also the potential for interferences related to biological autofluorescence, and the analyte-dependent sensitivity and limit of detection (LOD) for sensors, which rely on the nature of both the protein and the fluorescent probe utilized. Finally, in cases where immunological reagents are used, the devices can show a lack of reversibility, and may operate only as a ''one-shot'' screen, without the potential for continuous, quantitative analysis.

1 INTRODUCTION

In the past two decades, there has been explosive growth in the development and application of fluorescence-based biosensors. Such devices generally use a biological reagent, such as a protein or nucleic acid strand, to derive a selective interaction with an analyte of interest. To obtain a fluorescence signal, the analyte may be inherently fluorescent, may undergo a reaction, such as an enzyme-catalyzed hydrolysis, that leads to fluorescent products, or may compete with a fluorescently tagged analog for binding sites on a protein. Alternatively, the biological reagent may itself be fluorescently labeled. In this case, interactions between the analyte and biomolecule can produce fluorescence signals owing to alterations in the biomolecule structure or dynamics.

Fluorescence-based biosensors may utilize emission intensity,[8] emission wavelength,[12,13] fluorescence anisotropy[14] or fluorescence lifetime[15,16] measurements. Such signals may arise from selective perturbations that change the concentration of a fluorescent species, or those that alter polarity, pH, charge or viscosity, thereby changing the emission properties of the fluorophore. A detailed description of fluorescence principles and applications is given by Lakowicz,[17–19] while the theoretical basis of various fluorescence-based signals used for biosensors has been presented in detail by Lubbers.[20]

The remainder of this chapter begins with a brief outline of the history of optical sensors. This is followed by a description of some key aspects of sensor development, including instrument design, and a brief review of the methods used to interface biomolecules to optical devices. The fluorescence assay techniques associated with different types of biomolecules and the applications of various fluorescence-based biosensors are then described, with particular emphasis on areas where such sensors are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

uniquely suited. Finally, a comparison with other methods and future prospects for these devices are presented.

2 HISTORY

Fluorescence emission has been known since the turn of the century, and much of the theoretical basis of the process was described in the first half of this century. In the 1950s, commercial fluorescence instrumentation became available, and this led to numerous studies of biological systems and the development of several fluorescence-based biological assays.[21] In the early 1970s, Harrick reported the first use of total internal reflection fluorescence (TIRF) to measure the emission of dansyl-labeled bovine serum albumin (BSA) adsorbed onto a quartz waveguide.[22] Shortly thereafter, Kronic and Little[23] reported the first application of TIRF to immunosensing, wherein the interaction between a fluorescein-labeled antibody and a surface-immobilized hapten were quantitated. This work ushered in the modern era of fluorescence-based biosensing.

Over the years, numerous fluorescence-based biological assays have been adapted to a configuration suitable for sensing applications. Initially, biological reagents such as antibodies, enzymes, and membrane-bound receptors were used for sensor development,[24] since these species were available in high purity, or could be extracted from living tissue. More recently, species such as regulatory proteins,[25] RNA,[26] DNA[27] and nucleic acid aptamers[28] have been exploited for optical sensor development. In these cases, the development of new optical biosensors had to wait for advances in molecular biology and for the emergence of new synthetic tools, such as solid-phase synthesis of DNA, before the biomolecules could be obtained. At present, increased collaboration between sensors specialists (mainly chemists, physicists and engineers) and biomolecule specialists (biochemists and molecular biologists) is leading to a new era in sensor development, wherein biomolecules are being designed to have intrinsic signaling capabilities, engineered immobilization regions, or enhanced thermal stability. Aspects of these advances, and how they are extending the range of applications for such devices, are discussed below.

3 INSTRUMENTATION

Numerous instruments have been utilized for fluorescence-based biosensors, and it is not possible to describe all of the various configurations here. The majority of instrumentation used for fluorescence-based biosensors is based on optical fibers or planar waveguides. Figure 1

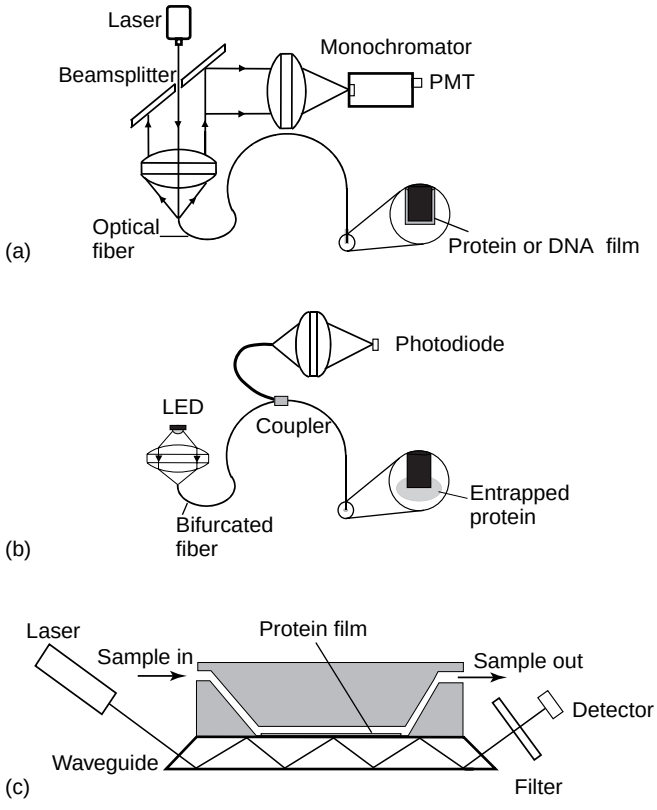

Figure 1 Examples of common instrumentation used for fluorescence-based biosensors. (a) Evanescent sensor using a perforated mirror beamsplitter. (b) Distal tip sensor using a bifurcated fiber beamsplitter. (c) Planar waveguide with built in flow channel. Note that the components shown in (a) and (b) are interchangeable. Each instrument could be used in distal tip or evanescent mode, with either beamsplitting method and either detector.

shows some of the typical instrumental configurations for fiber-optic sensors. Such devices consist of an excitation source, a beamsplitter to separate excitation and emission radiation, the fiber or waveguide with the immobilized biomolecule, and a detector and readout to process the signal. Typical sources include Xe lamps and a variety of lasers. More recently, light-emitting diodes (LEDs) have been used owing to their small size, low cost and useful spectroscopic characteristics, with some LEDs now emitting in the near UV region, where numerous probes absorb. Beamsplitters, as shown in Figure 1, generally consist of either a bifurcated optical fiber, a dichroic mirror, which separates light on the basis of wavelength, or a perforated mirror, which is generally used with collimated sources such as lasers. Detection of emitted radiation is usually done with a photomultiplier tube. However, recently developed sensors now utilize miniature detectors, including photodiodes, diode arrays, or charge-coupled device (CCD) cameras (for imaging of sensor arrays).

The optical fiber or waveguide is the heart of the fluorescence-based biosensor. Optical fibers provide a compact and rugged sensing device, and allow for remote spectroscopic measurements. A detailed description of the physical processes involved in transmission of light through an optical fiber is given elsewhere.[9] The properties of optical fibers which must be considered for sensor device development have been reviewed by Seitz,[29] while the principles of planar waveguides have been reviewed by Dessy.[30] The key aspect that makes these components of use for sensor development is their ability to propagate light via the process of TIR. TIR occurs when a ray of light that is propagating through a transparent medium of high refractive index encounters an interface with a medium of lower refractive index at an angle of incidence that is greater than a critical angle, which is given by the inverse sine of the ratio of external over internal refractive indices.

A consequence of the TIR process is that an evanescent field arises at the reflection points in the waveguide, and has the same wavelength characteristics as the waveguided radiation. The evanescent field intensity decays exponentially with distance into the less dense medium, with typical penetration depths being on the order of one-half of the wavelength of the guided radiation, or about 200–400 nm for visible radiation. In fabricating a fiber-optic chemical sensor the sample region is usually placed onto the surface of the waveguide in the form of a thin film, which interacts with the excitation radiation through the evanescent field. The short penetration depth of the exciting radiation results in preferential excitation of surface-bound species, and reduces interferences from bulk solution.

4 INTERFACING OF BIOMOLECULES TO DEVICES

To develop self-contained biosensors using biological components such as proteins or nucleic acids, the biorecognition element and, ideally, the fluorescent reporter group, must be put into intimate contact with a physicochemical transducer to allow excitation of fluorescence and detection of emission from the biorecognition element. For proteins, typical immobilization strategies have included physical adsorption of the protein onto the surface of a waveguide, various covalent immobilization strategies, and entrapment of the protein either within a semipermeable membrane or into an organic or inorganic polymer matrix.[31] For nucleic acids, single strands of DNA or RNA can either be grown onto surfaces one base at a time using solid-phase synthesis methods,[27] or intact strands can be immobilized either through interaction of biotinylated DNA with avidin-coated surfaces,[32] or by direct covalent attachment of the nucleic acid strands to a suitable surface.[33] The latter methods are preferred owing to difficulties associated with the solid-phase synthesis methods, including the likelihood of incomplete strand formation, and the problems with protecting group chemistry, which make it difficult to retain DNA on surfaces during deprotection of some bases.

Protein immobilization has been widely studied, and numerous reviews exist outlining the advantages and disadvantages of the various immobilization methods. Early examples of sensors utilizing TIRF were based on proteins that were adsorbed to surfaces.[14] However, while adsorption is a straightforward method for immobilization of biomolecules, many biomolecules adsorb in inactive orientations, or undergo irreversible denaturation upon adsorption to an inorganic surface.[34] Desorption of proteins subsequent to immobilization on a surface is also a problem. The constantly changing nature of the selective surface leads to an inability to maintain calibration of the sensing device.

A more common strategy for immobilization of a biomolecule is covalent linkage onto reactive sites on the surface of the transducer. The activating agents available for covalently linking a protein to a hydroxylated surface have been thoroughly reviewed by Weetall.[35] The functional groups on proteins that are suitable for binding to activated surfaces have been discussed by Srere and Uyeda.[36] Recent progress on protein immobilization has focused on methods for controlling the orientation of the immobilized biomolecule. Three approaches are commonly used for this purpose. The first is the use of a Protein A underlayer on a surface to aid in the immobilization of intact antibodies, which will bind through the complement region, resulting in the antigen binding regions facing toward the solution.[37] A second method is to place a streptavidin layer onto the surface, and to then bind selectively biotinylated proteins onto the surface to achieve orientational control. For example, Piervincenzi et al.[38] have demonstrated that single chains comprising the variable region of the antigen binding fragment (scFv) can be site-selectively modified with a peptide which a high affinity for streptavidin. This allowed immobilization of the scFv onto a streptavidin coated surface with a high degree of orientational control. A final method involves derivatizing the surface with a thiol-reactive species, and then reacting with a specific cysteine (Cys) within the protein. This method has been widely used for oriented immobilization of Fab fragments of antibodies.[39]

The other major method for interfacing biomolecules to optical surfaces is by entrapment in a polymeric matrix. Typical organic matrixes include nylon and

polyacrylamide.[40] In this case, the protein is not physically bonded to the surface, and orientation is not an issue. The porosity of the matrix is such that small molecules can enter, while the larger biomolecules remain entrapped. While organic materials have been relatively successful for entrapment of proteins, the reaction conditions employed are often harsh, and can result in substantial loss of biomolecule function.

An emerging method for protein immobilization is entrapment of biological components into inorganic silicate matrixes formed by a low-temperature sol–gel processing method.[41,42] The hydrated glass (sometimes referred to as a bioglass or biogel) contains a significant amount of entrapped water, and as a result, the biologically doped glasses maintain the activity of entrapped compounds, such as enzymes and antibodies, over periods of several months. A major advantage of these materials is that they are optically transparent, making them ideal for the development of chemical and biochemical sensors that rely on changes in a fluorescence signal.[43] Interested readers are referred to the reviews by Avnir et al.[41] and Dave et al.[42] for specific details pertaining to entrapment of biomolecules by this method.

5 APPLICATIONS OF FLUORESCENCE-BASED BIOSENSORS

The applications for which fluorescence-based biosensors may be used are numerous, and are limited only by the ability to develop a fluorescence-based assay that is sensitive to the presence of a desired analyte. Applications can be classified into those that are qualitative, simply involving the identification of a compound, and those that are quantitative, providing data about the concentration of a given analyte in a sample. Examples of qualitative applications include the use of biosensors as chromatographic detectors,[6] use of DNA arrays (gene chips) for the identification of genetic materials,[44] and the use of beads containing fluorescently tagged proteins for high-throughput drug screening.[7] While these are important applications for fluorescence-based biosensors, owing to space limitations this chapter will focus solely on those applications that are able to provide quantitative information about analyte concentrations.

5.1 Applications of Protein Engineering to Sensor Design

The ability to develop a biosensor for a specific analyte has generally been determined by the availability of purified biomolecules. At present, a wide variety of enzymes and antibodies are commercially available, and development of antibodies against specific analytes is fairly routine. However, even though such species are available, there are still a large number of analytes for which no biomolecules are available, and there is also a need to improve the currently available arsenal of selective proteins. The three main areas where protein engineering can aid in sensor development are:

1. selective placement of fluorescent labels into biomolecules to form reagentless sensors;
2. placement of reactive groups within the biomolecule to allow oriented attachment to surfaces;
3. engineering of thermostable proteins to help overcome difficulties associated with a loss of protein function over time.

In the first case, proteins can be prepared containing a single tryptophan (Trp) residue that is sensitive to analyte binding,[45] or a single reactive Cys residue which can be reacted with a thiol-selective fluorescent probe to provide an analyte-dependent reporter group.[46] Numerous examples of such proteins exist,[47] and reports are now appearing describing their use for sensor development.[48] The engineering of proteins containing reactive Cys residues is also useful for obtaining a selective site for attachment of proteins to derivatized surfaces.[49] It is also possible to selectively biotinylate a protein to control the orientation of binding to a streptavidin-coated surface, as described in section 4.[38] The final area of protein engineering, which is only now beginning to be explored, is the mutation of proteins to enhance their thermal stability. There are recent reports showing dramatic stabilization of protein structure and function as a result of a single residue mutation.[50] Such proteins are likely to be exploited for development of sensors with improved long-term stability.

5.2 Enzyme-based Sensors

Enzyme-based sensors have been developed using a variety of different signal transduction methods, including amperometric, potentiometric, absorbance and fluorescence approaches. Fluorescence-based enzyme sensors have been utilized for a variety of applications, as outlined in Table 1. These include biomedical monitoring (i.e. glucose and urea), monitoring of agents of environmental and defense interest (mainly enzyme inhibitors), and detection of heavy metals, which are often required cofactors for enzyme activity. The advantages of fluorescence-based detection strategies include freedom from electrical interferences, reduction of interferences related to matrix components with half-cell potentials similar to that of the reaction of interest

Table 1 Examples of enzyme-based biosensors

Analyte measured	Protein system used	Method of signal generation	Operating range	Refs.
Biomedical analytes				
Glucose	Glucose oxidase	Quenching by oxygen	0.7–10 mM	55
		FAD redox reaction	0.5–7 mM	63
Cholesterol	Cholesterol oxidase	Quenching by oxygen	0.15–3 mM	56
		FAD redox reaction	0.5–1.5 mM	63
Lactate	Lactate oxidase	Quenching by oxygen	1–60 mM	57
	Lactate dehydrogenase	NADH fluorescence	2–50 μM	59
Glutamate	Glutamate dehydrogenase	NADH fluorescence	0.2 μM[a]	58
Ethanol	Alcohol dehydrogenase	NADH fluorescence	0.9–9 mM	59
Pyruvate	Lactate dehydrogenase	NADH fluorescence	1–1000 μM	59
Urea	Urease	pH indicator dye	0.1–50 mM	53
ACh	FITC-acetylcholinesterase	pH indicator dye	0.05–3 mM	52
Penicillin	Penicillinase	pH indicator dye	0.1–10 mM	54
Environmental/defense				
Organophosphates	Organophosphate Hydrolase	FITC labeled enzyme	8–400 μM	60
	Acetylcholinesterase	Fluorogenic substrate	5–330 μg ml^{-1}	61
Zn, Cd, Co, Ni, Cu (all in +2 oxidation state)	Carbonic anhydrase	Binding of fluorescent inhibitor in presence of cofactor	10 pM to 100 nM, depends on ion	64–66

[a] Only lower LOD values were stated in these cases. FAD, flavin adenine dinucleotide; NADH, nicotinamide adenine dinucleotide, reduced form; ACh, acetylcholine; FITC, fluorescein isothiocyanate.

(a good example is ascorbate interferences in electrochemical glucose sensing), and the ability to monitor reactions involving species that are not electrochemically active. Disadvantages include background signals from inherently fluorescent biological components, and high background signals when sensing in opaque or highly scattering media.

Many enzyme-based fluorosensors have relied on the detection of a fluorescent signal change arising from the consumption of a fluorescent analyte, the production of a fluorescent product, or the change in signal from a secondary fluorescent reporter due to product formation.[51] In such cases, one can use the Michaelis–Menten equations to determine analyte concentration from the initial rate of reaction, provided that the Michaelis constant (K_m) for the enzyme is known under the conditions of use. Examples of this approach include a variety of sensors for ethanol, urea, ACh, glucose, lactate, glutamate, cholesterol, penicillin, and other species. In most cases, transduction occurs either by monitoring a change in pH using a pH-sensitive dye (ACh,[52] urea,[53] penicillin[54]), O_2 concentration via quenching of a stable fluorescent dye such as the tris(4,7-diphenyl-1,10-phenanthroline) ruthenium(II) complex (glucose,[55] cholesterol,[56] lactate[57] and others), or changes in NADH fluorescence (glutamate,[58] ethanol,[59] pyruvate[59]). More recently, enzyme-based fluorosensors have used for pesticide detection. For example, organophosphorus hydrolase (OPH) was used for reagentless fluorimetric determination of nine different organophosphates.[60] In this case, FITC was covalently bound to the enzyme to detect product dependent changes in pH in the local vicinity of the enzyme. In the case of paraoxon, the dynamic range extended from 25 to 400 μM with a detection limit of 8 μM.

An extension of the above approach is to monitor the concentration of enzyme inhibitors, which can be easily quantitated using the standard equations describing the effect of inhibitor concentration on reaction rates. Such sensors are particularly useful for detection of compounds of environmental or defense interest, since many such compounds act as enzyme inhibitors. A recent example of such a sensor is the use of sol–gel immobilized cholinesterase for detection of organophosphorous pesticides, which act as inhibitors of the enzyme.[61] Substances such as mercaban could be analyzed at levels ranging from 5 to 330 μg mL^{-1} in under 5 min. While such schemes are feasible, they rely on the presence of both the substrate (or a fluorogenic analog) and the analyte, and hence are not amenable to reagentless sensing.

Several reagentless enzyme-based sensors have also been developed utilizing the fluorescence of bound cofactors, such as FAD, for signal development.[62] A recent example is the use of changes in FAD fluorescence from sol–gel entrapped redox enzymes to detect glucose, cholesterol or L-amino acids.[63] The signal was based on the reduction of FAD to $FADH_2$, which produces a reversible change in the intensity of emission that is proportional to analyte concentration. Using this approach,

sensing of glucose or cholesterol could be done over a millimolar concentration range, which corresponds to the lower range of physiological interest for these species.

A promising alternative route to obtain a reagentless enzyme-based sensor is to use a fluorescent inhibitor that binds to the protein only in the presence of a specific cofactor. Thompson and Jones[64] initially demonstrated this approach for sensing Zn(II), based on the binding of a dansylamide inhibitor to the Zn(II) cofactor after the cofactor had bound to the active site of carbonic anhydrase. Initial studies used the inhibitor in solution and utilized fiber-optic detection. In this case, the concentration of metal ion over the range of 50 to 1000 nM was found to be proportional to the ratio of fluorescence intensities at two emission wavelengths, corresponding to emission from the bound and free inhibitor.

More recent work from Thompson et al.[65] has focused on the use of fluorescently labeled carbonic anhydrase for detection of picomolar to nanomolar levels of Zn(II), Co(II) and Cu(II) utilizing fluorescence anisotropy, and for monitoring of picomolar concentrations of Cu(II) and Zn(II), and nanomolar concentrations of Cd(II), Co(II), and Ni(II) using fluorescence lifetime[66] data as an analytical signal. In both cases, a fluorescent sulfonamide probe was bound to a single Cys residue that was placed near the enzyme active site via site-directed mutagenesis, making the sensor reagentless. This work represents one of the most sensitive methods for detection of heavy metals, and in some cases exceeds the detection limits of standard methods, such as atomic emission techniques using inductively coupled plasma sources.

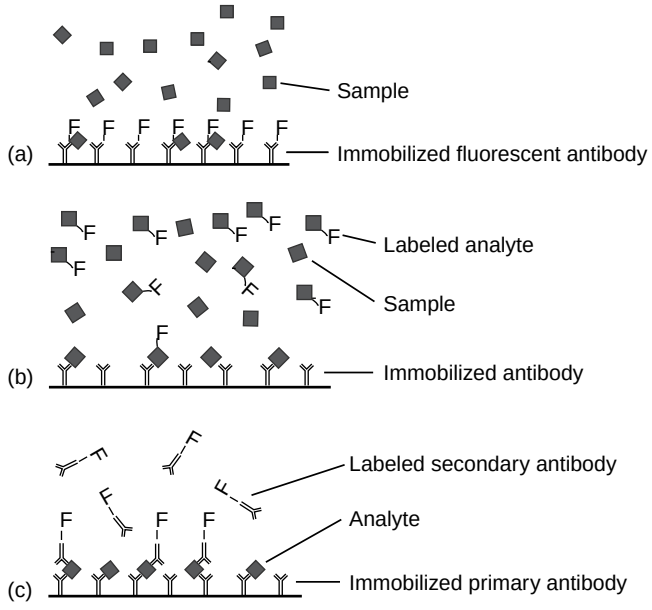

Figure 2 Immunoassay methods. (a) Direct assay using a fluorescently labeled primary antibody. Binding of the antigen changes the emission properties of the attached label, producing a signal. (b) Competitive assay using a fluorescently labeled analog of the analyte. A known amount of labeled analyte is mixed with the unlabeled sample. The fluorescence is quantitated after a washing step to remove labeled analyte that is free in solution. (c) Sandwich assay using a fluorescently labeled secondary antibody. The analyte is added first and binds to the primary antibody. An excess of labeled secondary antibody is added, and binds to the surface-bound antigen. Fluorescence is quantitated after a washing step to remove any unbound species.

5.3 Fluoroimmunosensors

Immunosensors comprise the largest single class of biosensors, and a wide variety of transduction methods have been reported for such sensors, including those based on surface plasmon resonance (SPR), electrochemical, absorbance and fluorescence strategies. Numerous examples of such sensors exist, and interested readers are referred to reviews on immunosensor properties and applications that have appeared in recent years.[67,68] Most work in the area of fluoroimmunosensors has centered on three common approaches, as outlined in Figure 2. The first is direct detection of fluorescent analytes. The second is the competitive assay format, wherein the analyte (which may be either an antibody or a hapten) competes with a known concentration of a fluorescently labeled analog for binding sites on a surface. The third common method is the sandwich assay format, wherein a large antigen with multiple antibody binding regions is first bound to a primary antibody present on a surface, and is then complexed with a second, fluorescently labeled antibody to derive a signal. As shown in Table 2, fluoroimmunosensors have been used for analytes of biomedical, defense and environmental interest. Most examples involve detection of proteins or small analytes. However, immunosensors have also be used to detect very large analytes, such as viruses and other pathogenic organisms.[69]

Examples of immunosensors which utilize the intrinsic fluorescence of the analyte are not widespread; however, they are used in the specific area of PAH detection. As an example, Alaire et al.[70] described a fiber-optic immunosensor for benzo[*a*]pyrene which relied on the direct detection of the bound analyte using evanescent excitation. In this case, a LOD of 20 nM and a dynamic range of spanning three orders of magnitude could be obtained with an equilibration time of 20 min, although a washing step was required to remove unbound analyte. In such cases, bound analyte can often be discriminated from free analyte using fluorescence polarization measurements, thereby removing the need for a washing step and improving response times.

Many more immunosensors utilize competitive immunosensing techniques, particularly for smaller analytes.

Table 2 Examples of fluoroimmunosensors

Analyte measured	Protein system used	Signal generation	Operating range	Refs.
Biomedical analytes				
Cocaine	Anti-cocaine IgG	Competitive assay	10–1000 ng mL^{-1}	72, 73
Human chorionic gonadotropin	Anti-human chorionic gonadotropin	Sandwich assay	3–500 ng mL^{-1}	86
Prostate-specific antigen	Anti-prostate-specific antigen	Sandwich assay	3.3 pM[a]	87
HSA	Dansylated anti-HSA	Direct assay	0.1–8 μg mL^{-1}	39
Myoglobin	Cascade-blue labeled anti-myoglobin	Energy transfer from heme to labeled antibody	80–3300 ng mL^{-1}	88
Environmental/defense				
Benzo[*a*]pyrene	Anti-benzo[*a*]pyrene IgG	PAH fluorescence	20–2000 nM	70
TNT	Cyt-5-TNT : anti-TNT	Competitive assay	10–1000 ng mL^{-1}	74
Phenytoin	Phycoerythrin–phenytoin and Texas red antibody	FRET-based competitive assay	1–250 μM	75
Pathogens/viruses				
Clostridium botulinum toxin A	TRITC-labeled antitoxin A	Sandwich assay	5–100 ng mL^{-1}	78
Yersinia pestis F1 antigen	Surface-expressed F1 antigen	Sandwich assay	5–100 ng mL^{-1}	79
Anthrax	Protective protein PA	Sandwich assay	5–100 ng mL^{-1}	80
Staphylococcus aureus	Protein A	Sandwich assay	1–100 ng mL^{-1}	81
Ricin	Anti-ricin antibody	Sandwich assay	0.1–250 ng mL^{-1}	82
NDV	FITC-labeled anti-NDV	Sandwich assay	5 ng ml^{-1} [a]	83
Rubella	Anti-rubella	Sandwich assay	30 IU mL^{-1} [a,b]	85
Staphylococcal enterotoxin B	Cyt-5 labeled anti-staphylococcal enterotoxin B	Sandwich assay	5–500 ng mL^{-1}	126

[a] Only lower LOD were stated in these cases.
[b] IU is infectious units.
IgG, immunoglobulin G; HSA, Human serum albumin; PAH, polyaromatic hydrocarbon; TNT, 2,4,6-trinitrotoluene; Cyt, cytochrome; FRET, fluorescence resonance energy transfer; TRITC, tetramethylrhodamine-5-isothiocyanate; NDV, Newcastle disease virus.

Examples of biomedical immunosensors utilizing the competitive assay format include one for the detection of human IgG using a planar waveguide,[71] and a series of sensors for the detection of cocaine,[72,73] all being based on displacement of a labeled analog from an immobilized antibody. In the latter case, cocaine could be detected directly in leaf extracts, with results that were in good agreement with those obtained using gas chromatography (GC). A drawback of this approach was that the sensor provided stable responses for only 14 days.

Competitive fluoroimmunosensors are also widely used for compounds of environmental and defense interest. For example, a competitive immunoassay for trinitrotoluene was reported, which had a LOD of 10 ng mL^{-1} using a 5 min equilibration time.[74] Astles and Miller[75] reported a phenytoin fiber-optic immunosensor utilizing an energy transfer-based signal. In this case, an antibody labeled with Texas red (acceptor) was mixed with B-phycoerythrin-labeled phenytoin (donor), and interaction of these species caused quenching of the phycoerythrin signal. Increased binding of free phenytoin resulted in a shift in antibody binding equilibria between labeled and unlabeled phenytoin, producing a proportional increase in fluorescence. Using this system, phenytoin could be detected over a dynamic range of 250 μM with a LOD of ~1 μM. An identical system was also described for detection of free theophylline, and gave similar analytical results.[76]

Competitive[77,78] assay methods have been used for the detection of pathogenic organisms of interest in both the food and defense industries. An example of a competitive assay involved the binding of botulinum toxin B to immobilized antibotulinum toxin B IgG. This reaction competed with the binding of biotinylated toxin. Higher levels of the analyte resulted in lower levels of bound the biotinylated analog, resulting in a lower fluorescence signal upon addition of FITC-labeled streptavidin.[77]

Examples of sandwich assays are also numerous. A key advantage of sandwich-based immunosensors is that they

are useful for larger species such as viruses and pathogenic organisms, which generally express antigenic proteins within their coat membranes. For example, a sandwich assay was used for the detection of *Clostridium botulinum* toxin A. Toxin A antibodies were covalently bound to a quartz optical fiber and incubated in solutions of toxin A and a TRITC labeled antitoxin A antibody. The fluorescence signal was directly proportional to the amount of toxin A present, and a detection limit of 5 ng mL^{-1} was achieved with a response time of 10 min.[78] Similar strategies have been utilized to detect the F1 surface antigen of *Yersinia pestis*,[79] the protective protein, PA, which is co-expressed with the anthrax toxins,[80] and protein A, which is a product secreted by *Staphylococcus aureus*.[81] In the first two cases, polyclonal or monoclonal antibodies were attached to the surface of a tapered optical fiber to capture the antigen, and TRITC-labeled monoclonal anti-F1 or anti-PA IgG was used for signal generation. In each case, an LOD of 5 ng mL^{-1} was reported using a 30 min incubation time, and the dynamic range extended to 100 ng mL^{-1} of antigen. In the latter case, antiprotein A was adsorbed to a plastic optical fiber, and was used to capture protein A. FITC-labeled antibody was then added to generate a signal. Using this approach, a LOD of 1 ng mL^{-1} was reported. A final example of the utility of the sandwich assay approach is the detection of the potently toxic protein ricin by a sandwich immunosensor.[82] An antiricin IgG was immobilized onto the surface of an optical fiber, and the binding of ricin was quantitated using a fluorescently tagged secondary antiricin antibody. The linear dynamic range of detection for ricin in buffer, using an avidin–biotin method for primary antibody immobilization, was 100 pg mL^{-1} to 250 ng mL^{-1}. The LOD for ricin in river water was slightly worse, at 1 ng mL^{-1}.

Detection of whole viruses has also been reported using fluoroimmunosensors. A fluorimetric sandwich assay for NDV was developed by Lee and Thompson[83] based on evanescent measurements using a fiber-optic probe within a flow cell. Anti-NDV antibody was immobilized onto the fiber, and was used to capture antigen. FITC-labeled anti-NDV antibody was then made to flow past the surface and bound to captured antigen. The LOD for this method was 5 ng mL^{-1}.

Immunosensors based on competitive or sandwich assay formats have enjoyed widespread use, primarily because they are amenable to a very large number of analytes, and are capable of detecting both small analytes (generally via competitive assay) and large biomolecules (using either competitive or sandwich techniques). However, as indicated above, such devices often provide slow response times, offer a limited dynamic range, show poor long-term stability, and can be tedious to use, owing to the need for washing steps. In some cases, TIRF has been used to discriminate between bound and unbound fluorescent species, and this has sometimes provided improved response times and a reduction in background interferences. Regardless of these successes, the need to add extra reagents, such as labeled analytes or antibodies, in order to quantitate analyte concentration is a serious problem for sensor development. In fact, a brief survey of the recent literature (1994 to 1999) has shown a steady decline in the number of reports describing fluoroimmunosensors, and a significant increase in immunosensors based on signals that do not require a secondary reagent for signal development. Such methods include SPR, piezoelectric and amperometric signal transduction methods.

A step toward a reagentless immunosensor has been the use of a fluorescence capillary fill device, shown in Figure 3, to introduce samples to a planar waveguide device. In this case, the antibody is bound to one face of a planar waveguide, and is used to capture analyte. All other necessary reagents, including fluorescently labeled antigens which compete with the analyte, are dosed within a thin polymer film that is on a parallel plate about 100 μm away from the immunosurface. A controlled volume of the sample is introduced by capillary action, and the presence of the buffer causes the doped film to release the reagents so that they are able to compete with the analyte for sites in the immunosurface.[84] Using such a device, species such as serum rubella antibody,[85] human chorionic gonadotropin[86] and prostate-specific antigen[87] have been detected, even in whole blood, with LODs ranging from 1.1×10^{-9} to 3.3×10^{-12} M.

Another advancement in fluoroimmunosensor design, also aimed at the development of reagentless sensor technology, has been the coupling of the analytical reporter group (i.e. fluorescent probe) directly to the immobilized antibody. In such a case, the selective binding of antigens or haptens to the sensor surface results in a direct change in the properties of the reporter group, without the need for the addition of any extra compounds. Such a design is reagentless, since the only species that needs to be present to produce a change in a signal is the analyte itself.

One of the first developments for reagentless immunosensor design was described by Bright et al. in 1990.[39] In this work, anti-HSA was incubated with HSA to block the active site of the antibody. The complex was then nonselectively reacted with dansylamine to place a solvent-sensitive fluorophore near the active site of the antibody. Removal of HSA was then carried out and the labeled antibody was immobilized near the distal tip of an optical fiber. The binding of HSA resulted in a large blue shift in the dansyl emission wavelength and an increase in emission intensity that was directly related to HSA concentration. The sensor could be regenerated

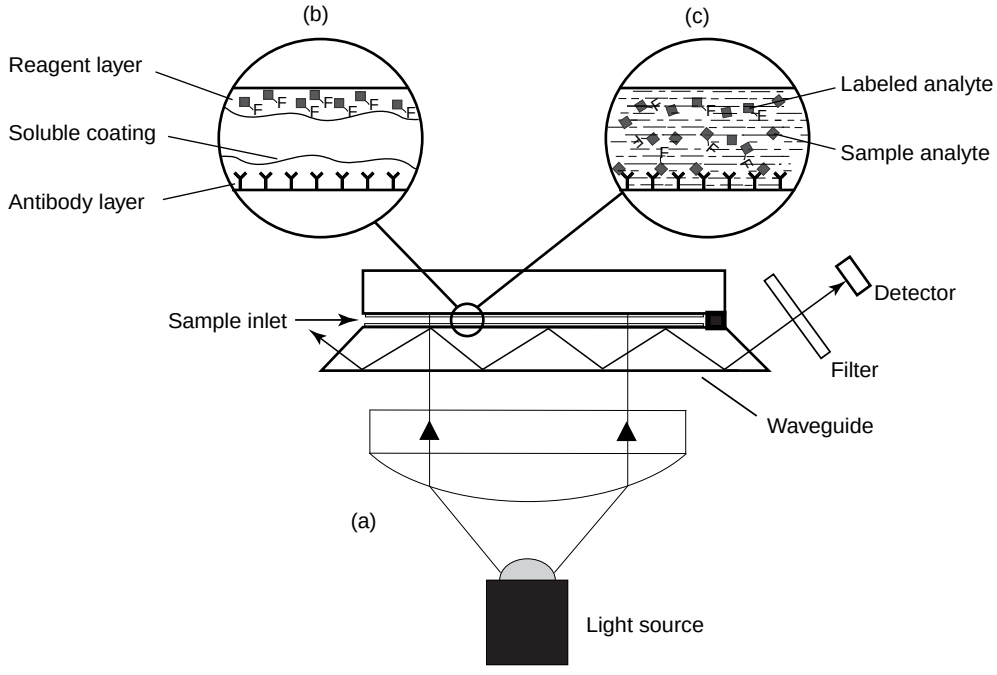

Figure 3 Schematic of the capillary fill device used for reagentless immunosensing. (a) Cutaway view of the device. Light directed into the waveguide causes direct excitation of the sample. Only light emitted by bound analyte undergoes TIR and is detected. (b) Enlarged view of the empty device showing the antibody and reagent layers. (c) Enlarged view of the filled device showing the dissolved reagent layer undergoing competitive binding with the analyte. Sandwich assays are also possible using this device. (Adapted from Badley et al.[84])

using a cheotropic solution and reused up to six times before significant changes in the fluorescence response were observed.

The use of labeled antibodies has continued to evolve in recent years. For example, myoglobin has been detected in a reagentless format using the anti-myoglobin antibody, which was labeled with the fluorescent dye cascade blue.[88] The labeled antibody was placed into a polyacrylamide gel layer at the distal tip of an optical fiber. Binding of myoglobin resulted in FRET from the cascade blue label to the heme group of the myoglobin. Using this approach, myoglobin could be measured in a reagentless fashion with a LOD of 83 ng mL^{-1}, and with a response time between 15 and 130 min.

Another exciting development which is likely to have an impact on the development of reagentless immunosensors is the use of protein engineering methods to prepare single-chain antibodies. For example, Stewart et al.[89] have developed an intrinsically fluorescent single chain antibody, which was used to generate a reagentless fluorescence response that was related to the binding of micromolar concentrations of Zn(II). The advantage of this scheme is that the fluorophore can be placed into a location that is highly sensitive to analyte binding, resulting in the ability to tailor the sensitivity of the device.

5.4 Regulatory Proteins

One of the recent advances in the field of biosensor design, which directly addresses the issue of reagentless fluorescence-based sensing, has been the use of engineered fluorescent peptides[90] and regulatory proteins[25] as selective recognition elements. In most cases, these proteins are site-selectively labeled with an environmentally sensitive fluorescent probe, or are linked to a green fluorescent protein (GFP) to form a chimera. Binding of the analyte results in a direct signal via specific interactions with the probe, or an indirect signal wherein a conformational change in the protein alters the environment of the probe or GFP, producing a measurable change in the fluorescence properties (intensity, FRET efficiency, emission wavelength or anisotropy). These biorecognition elements are attractive for sensor development since in many cases they bind analytes reversibly, and have binding constants which permit detection of analytes over a concentration range spanning picomolar up to micromolar or even millimolar levels.

While many of these biorecognition elements have not yet been placed onto transducers such as optical fibers and planar waveguides, they represent an important emerging class of biomolecules that are likely to revolutionize fluorescence-based biosensors. A key advantage of such

Table 3 Examples of regulatory protein systems and probes used for reagentless sensing

Analyte measured	Protein system used	Method of signal generation	Operating range	Ref.
Free fatty acids	Fatty acid-binding protein	Emission intensity ratio	1 nM to 20 μM	91
Nitric oxide	Cyt-*c*′	Quenching by nitric oxide	20 μM to 1 mM	93
Zn(II)	Zinc finger domains	FRET Emission ratio	0.1 to 1 mM	95
			7 pM to 65 nM	96
Ca(II)	Parvalbumin	Emission intensity	100 nM to 10 μM	97
Ca(II)	CaM	Emission intensity	100 nM to 10 μM	98
Ca(II)	Troponin C	Emission intensity	100 nM to 10 μM	99
Ca(II)	CaM-M13 chimera	FRET from BGFP-CaM to RGFP-MLCK	100 nM to 10 μM	102
Maltose	Maltose-binding protein	Emission intensity		
	S337C mutant		5 to 500 μM	46
	D95C mutant		0.05–10 μM	103
Glucose	Glucose-binding protein	Emission intensity	1 to 100 μM	107
P_i	Phosphate-binding protein	Emission intensity	0.2 to 30 μM	104
Drug compounds or ATP	Membrane-bound P-glycoprotein	Quenching of intensity	10 nM to 10 μM	111
Insulin	Membrane-bound IR	Emission intensity Rotational anisotropy	1–140 nM	112
Serotonin	$5HT_3$ serotonin receptor	Sandwich assay format	0.01–10 nM	113

CaM, calmodulin; BGFP, ''blue'' form of the GFP; RGFP, ''red'' form of GFP; MLCK, myosin light chain kinase; P_i, inorganic phosphate; ATP, adenosine triphosphate; IR, insulin receptor.

proteins is that they can be microinjected into cells to allow studies of intracellular events via changes in fluorescence responses.[25] Many of the applications of these biomolecules would not be possible using other techniques, and hence it is valuable to briefly review the utility of these systems here.

A number of fluorescence-based biosensors have been developed which rely on either a direct or indirect alteration in the environment surrounding the probe upon binding of analyte, as outlined in Table 3. Examples of sensors utilizing direct interactions between the analyte and protein include the use of a fluorescently labeled fatty-acid binding protein for detection of fatty acids,[91] the use of a labeled myosin light chain to examine myosin phosphorylation events intracellularly,[92] and the use of Cyt-*c*′ immobilized at the end of an optical fiber for fluorimetric detection of nitric oxide.[93] In the latter case, binding of NO to the pentacoordinate Fe site in Cyt-*c*′ resulted in a change in quenching of fluorescence from the porphyrin group. Alternatively, Cyt-*c*′ could be non-selectively labeled with a fluorescein derivative whose emission overlapped the absorbance band of the porphyrin so that the label acted as a donor while the heme group acted as an acceptor for FRET. The labeled protein showed a similar sensitivity for detection of NO as the unlabeled protein, but had a substantially improved LOD, which was determined to be 20 μM. An interesting aspect of this work was that the detection of NO was done with protein that was trapped in a polyacrylamide gel at the end of a nano-optrode, measuring only 200 nm in diameter at the tip, making it possible to do measurements in volumes as small as 10 pL.

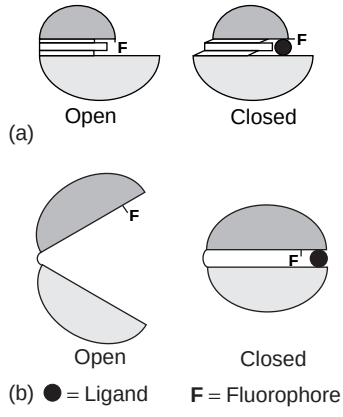

Figure 4 Examples of protein conformational motions involving (a) shear and (b) hinge movements. The binding of the ligand results in a change in the location and hence environment of the attached fluorescent probe, producing a signal. (Adapted from Brennan.[47])

Indirect sensors utilize fluorescently labeled proteins that undergo conformational changes on binding with analyte, resulting in a change in the environment of the attached probe, as shown in Figure 4. A wide variety of such proteins exist, and these display selectivity for a wide array of analytically interesting compounds. Examples of such proteins (with ligands in parentheses) include:[94] citrate synthase (oxaloacetate), hexokinase (glucose), alcohol dehydrogenase (ethanol), Trp repressor (Trp), the maltodextrin binding protein (maltose), the glucose/galactose binding protein (glucose), as well as proteins which bind ligands such as Cys, glutamine, P_i and vitamin B_{12}. Other examples include CaM (Ca^{2+}),

lactoferrin (Fe) and cAMP-dependent protein kinase. A number of these proteins, as well as a handful of peptides, have already been site-selectively labeled with fluorescent probes and used for both intracellular and in situ sensing applications, as described below.

Zn(II)-binding polypeptides (referred to as zinc finger domains) are known to undergo Zn(II)-dependent conformational changes. Walkup and Imperiali[90] synthesized a Zn(II)-binding polypeptide with a dansyl group placed within the cluster of residues forming a Zn(II)-induced hydrophobic pocket, and the resulting peptide showed a significant fluorimetric response to Zn(II)-induced structural changes. Ratiometric analysis, using emission wavelengths at 650 nm (Zn^{2+}-free) and 525 nm (Zn^{2+}-loaded) showed that the response was linear over the range from 0.1 to 1 mM Zn(II), and that the response was fully reversible. Godwin and Berg[95] reported a Zn(II)-sensing strategy based on a principle similar to that described above. In this case, the peptide was selectively labeled at one site with lissamine, and at another with fluorescein. Binding of Zn(II) resulted in a conformational change that brought the two probes into closer proximity, increasing the energy transfer efficiency from fluorescein to lissamine. The ratio of emission intensities at 521 nm and 596 nm was sensitive to Zn(II) binding in the concentration range from 0.1 to 1.0 mM.

More recent work by Walkup and Imperiali[96] has shown that the separation of the fluorescent reporter group from the Zn(II)-binding domain provides the advantage of being able to manipulate each part independent of the other.[96] In this work, it was demonstrated that the affinity of the metallopeptide for binding of Zn(II) could be manipulated by altering the Zn(II)-binding domain, without affecting the performance of the attached fluorophore. Using this approach, dissociation constants for binding of Zn(II) ranging from 7 pM to 65 nM were achieved.

Examples of sensors based on fluorescently labeled regulatory proteins are numerous, and have been recently reviewed.[47] A key application is the detection of Ca^{2+}, which has been done using a variety of labeled proteins, both in solution,[97,98] and in cells.[99] A few examples also exist describing the use of such proteins for the development of fiber-optic Ca^{2+} sensors.[48] In the majority of these cases, Ca^{2+} could be detected over the range spanning 100 nM up to 10 μM, with minimal interference from Mg^{2+} at levels up to 10 mM. Ions which had similar ionic radii to Ca^{2+}, particularly Cd^{2+} and Eu^{3+}, did result in substantial interferences when present at levels of 500 μM. The sensor response times were on the order of 15 min, which is longer than is desirable for a working sensor.

GFP-labeled Ca^{2+}-binding proteins have also been applied to sensing of intracellular Ca^{2+} levels. Examples of this strategy have recently been reviewed by Tsien and Miyawaki.[100] A few key examples will be presented here. One example of this is the use of a GFP-labeled CaM binding domain from MLCK to detect Ca^{2+}-loaded CaM in cells via an energy-transfer method.[101] The CaM-binding domain was labeled on one end with the BGFP, which acts as a donor, and at the other end with the RGFP, which acts as an acceptor. Binding of holo-CaM caused the GFP-labeled CaM-binding domain to adopt an extended conformation, spatially separating the BGFP and RGFP labels, and decreasing the energy transfer efficiency. Using this method, it was possible either to detect Ca^{2+} by having an excess of apo-CaM available in the solution containing the fluorescent indicator protein (denoted as FIP-CM_{SM}), or to detect holo-CaM directly. In another example, the C-terminus of RGFP-$MLCK_f$-BGFP (where $MLCK_f$ is the CaM-binding fragment of MLCK) was linked to the N-terminus of CaMCN (C and N terminal regions exchanged).[102] These chimeric proteins responded directly to changes in free intracellular Ca^{2+} concentration through a change in the energy transfer efficiency between the two forms of GFP.

Regulatory proteins have also been used to detect organic compounds. The first use of this approach was by Gilardi et al.,[46] who examined the effectiveness of a fluorescently labeled maltose binding protein (MBP) as a maltose sensor. A single Cys mutant was constructed by replacing a serine residue at position 337 (S337C), since this position was within the cleft where maltose binding occurred, and hence experienced a large environmental change on maltose binding. Attachment of thiol selective probes such as 4-[*N*-(iodoacetoxyethyl)-*N*-methylamino]-7-nitrobenz-2-oxa-1,3,-diazole (IANBD) or acrylodan resulted in a system that showed a substantial increase in label fluorescence upon binding of maltose (80% for Ac, 160% for IANBD). In optimal cases, maltose could be detected over a dynamic range of approximately 500 μM, with a LOD of approximately 5 μM. Response times were on the order of several hundred seconds which were independent of ligand concentration, indicating that the rate limiting step in generating a response was likely the conformational change itself. Other fluorescently labeled MBPs, such as a D95C mutant labeled with IANBD, have also been found to be sensitive to maltose binding. This particular mutant produced a 4.4-fold increase in fluorescence intensity when saturated with maltose, and could reversibly detect maltose over a range of c. 10 μM with a LOD of c. 50 nM.[103]

Brune et al.[104] used an engineered phosphate binding protein for the detection of P_i. In this case, a A197C mutant was prepared, and was labeled with the probe *N*-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide (MDCC). The protein undergoes a 5.2-fold increase in emission intensity in the presence of

saturating amounts of P_i. The new fluorescent protein biosensor was able to detect free P_i over a concentration range from 0 to 0.8 molar equivalents of protein with a dissociation constant, K_d, of 0.3 μM. The response was very fast (under 50 ms) and was remarkably selective. The only compound that proved to be a significant interferant was sodium arsenate, which bound with a K_d of 3 μM. The potential of labeled P_i-binding protein as an arsenate sensor remains to be explored. The rapid response of the new protein-based biosensor has permitted measurements including the release of P_i during ATP hydrolysis by actomyosin subfragment 1 ATPase[104] or myofibril ATPases,[105] and during the interaction of p21ras with the GTPase-activating proteins p120-GAP and neurofibromin.[106] In each case, stop-flow measurements were used to follow the reactions, and showed that the fluorescently labeled P_i-binding protein was able to follow extremely fast processes involving P_i release, which were occurring on the millisecond timescale.

The use of labeled regulatory proteins has also been applied to the development of a sensor for glucose.[107] In this case, single Cys mutants of the glucose/galactose binding protein were prepared and labeled with polarity sensitive probe. A L255C mutant labeled with acrylodan showed a 50% decrease in intensity and a slight red-shift in the emission maximum wavelength on binding of glucose. In this case, the sensor was responsive over a range of c. 2 μM and it had a LOD approaching the low nanomolar range. Use of a H152C labeled with IANBD produced a major change in signal (fourfold) on binding of glucose, and could measure glucose up to 100 μM, making it sensitive at the lower portion of the physiologically important range of glucose concentrations.

5.5 Membrane-associated Proteins

Another area of increasing importance is the use of membrane-associated proteins for the development of fluorescence-based biosensors. Recent examples include the use of the acetylcholine receptor (AChR) for detection of agonists and antagonists, the use of P glycoprotein (Pgp) to detect drug compounds, and the use of the IR to detect insulin. In the case of AChR, at least three approaches have been used. In the first approach, the protein was randomly labeled with nitrobenzoxadiazole and a fluorescence signal was generated as a result of an analyte-induced conformational change within the protein.[108] Using this approach, ACh could be detected at micromolar to millimolar levels. The second approach relied on perturbations of the surrounding membrane upon interaction of AChR with analytes. The membrane was doped with a small amount of a fluorescently labeled lipid, which responded to membrane perturbations resulting from the conformational change of the protein. This strategy was capable of detecting carbamylcholine over the range of 300 nM to 3 μM.[109] The final approach was to use a fluorescently labeled antagonist, FITC-α-bungarotoxin, as a labeled analog which competed with AChR-binding compounds for sites on a AChR-coated surface.[110] The sensor utilized AChR that was adsorbed onto the surface of an optical fiber, and the sample was excited via the evanescent wave. The sensor was able to detect unlabeled α-bungarotoxin over the concentration range spanning 1 to 1000 nM using an equilibration time of 5 min.

The detection of drug compounds was demonstrated using Pgp that was covalently labeled with 2-(4′-maleimidylanilino)naphthalene-6-sulfonic acid (MIANS) at both conserved Cys residues, one within each of the nucleotide binding domains.[111] ATP bound directly to the labeled nucleotide binding domains of Pgp, causing the fluorescence to be quenched by up to 15% owing to a direct interaction between ATP and the bound probe. Drug compounds, including vinblastine and colchicine, bound to a second site on the protein that was spatially removed from the nucleotide binding site. Drug binding induced a conformational change that resulted in a decrease of up to 60% in the fluorescence intensity. Using this approach, drug compounds could be detected at nanomolar concentrations over a range extending up to 10 μM.

Changes in the intrinsic Trp fluorescence of the IR have been used to detect insulin based on an analyte-induced conformational change in the protein.[112] Binding of insulin to IR caused a 35% increase in intrinsic Trp emission intensity, a slight blue-shift in the emission maximum, and an increase in the average rotational anisotropy of the protein. These spectroscopic changes were consistent with a change in the average Trp environment in which the probes became somewhat more buried into hydrophobic regions of the protein. The response was sensitive to insulin levels ranging up to 140 nM, and response times ranged from 1 to 20 min, becoming longer as the concentration of insulin decreased.

The majority of the examples cited above were not done in a sensor format, owing to the difficulty associated with immobilizing viable membrane-associated proteins on the surface of optical waveguides. However, two recent reports have appeared describing methods that may overcome this problem. In the first case, the 5-HT_3 serotonin receptor was expressed with a C-terminal histidine tag to allow the solubilized protein to be bound to the surface of a planar waveguide that was first derivatized with Ni(II)nitroloacetic acid.[113] The bound receptor was shown to retain its native, membrane-bound function, and the binding of a fluorescein-labeled serotonin analog could be detected over the concentration range of 10^{-11} to 10^{-8} M using evanescently excited fluorescence. Using a competitive assay format, the

binding of nonfluorescent species, such as serotonin, could be detected over the concentration range spanning 1 to 100 nM. The key limitation of the device was the extremely long equilibration times, which often reached several hours.

A second method for immobilization of membrane-bound proteins involved placing entire biomembrane fragments onto the surface of a planar waveguide that was part of a flow-cell.[114] The silica surface was first alkylated, and then a lipid layer was deposited by vesicle fusion. A Na,K-ATPase-rich membrane fragment, containing FITC-labeled protein, was then physisorbed onto the lipid surface. In the absence of Na^+, the sensor could detect K^+ over the concentration range of 10 μM to 10 mM, with response times of less than 1 min. Inclusion of Na^+ removed the K^+-dependent signal, showing the effect to be K^+ specific.

Another recent advancement that is likely to impact the development of sensors utilizing membrane-bound receptors in demonstration that whole organisms, such as plant[115] and animal[116] cells, can be successfully entrapped into sol–gel derived glasses. While no reports have yet emerged describing the use of such biomaterials for fluorescence biosensor development, the ability to cast films of these materials onto planar waveguides and optical fibers is likely to result in the development of fluorosensors using entrapped cells.

5.6 DNA/RNA-based Genosensors

While proteins are by far the most widely used biomolecules for sensor development, recent years have seen a dramatic increase in the number of sensors that utilize nucleic acids to derive selective interactions with analytes. DNA- and RNA-based devices, often termed genosensors, have been widely applied to the detection and even sequencing of DNA and RNA. In these cases, single-stranded DNA (or RNA) is placed on the surface of an optical fiber or waveguide and will bind to the complement strand in solution to form a double-stranded (ds) complex (dsDNA or dsRNA). Several approaches have been used for such sensors. One approach is to use a competitive sensing format, wherein a fluorescently tagged complement strand competes with unlabeled analyte for binding to immobilized DNA. This system could detect the labeled DNA with a LOD of 2×10^{-13} M, and the unlabeled analyte at 1.1×10^{-9} M.[117] An alternative approach is to do a sandwich assay. In this case, a nonfluorescent analyte first interacts with single-stranded DNA on a surface, forming the dsDNA. A second DNA strand, which is labeled with a fluorescent dye, is then added to the immobilized dsDNA and binds to a different complementary region of the analyte, forming a sandwich structure that can be quantitated on the basis of emission intensity. Using this approach, as little as 25 pM of *H. pylori* RNA could be detected in under 1 min.[118] A third method for deriving a fluorescence response, which does not rely on using labeled DNA, is the use of intercalating fluorescent dyes, such as ethidium bromide (EB). This probe will preferentially associate with the dsDNA, and will produce a fluorescent signal that is proportional to the amount of complementary DNA present in the sample. Piunno et al.[27,119] have used this approach to quantitate the binding of dA_{20} in solution to dT_{20} that was prepared on a fiber surface by solid-phase synthesis, and achieved detection limits of 86 ng mL^{-1}. This approach was also used to detect the formation of triple-helical T.AT on the surface of an optical fiber.[120]

An extension of this approach is to immobilize dsDNA and use it to detect the binding of carcinogenic species, such as decacycline and remazo brilliant blue, which intercalate into dsDNA.[121] In this approach, analytes displace EB which is intercalated into immobilized dsDNA, and hence produce a drop in fluorescence intensity. These species could be detected over one to two decades of concentration, with detection limits ranging from 50 ng mL^{-1} to 10 μg mL^{-1}.

Another application for nucleic acids is the selective binding of small ligands. In some cases DNA or RNA can take up specific three-dimensional structures that are capable of selectively catalyzing reactions (ribozymes, DNA catalysts), or that can reversibly bind analyte (aptamers). Thus far, no reports exist demonstrating the use of catalytic nucleic acids for sensor development. However, there are recent reports describing the use of nucleic acid aptamers for fluorescence-based sensing. In one case, high-affinity RNA was attached to an optical fiber via an avidin–biotin bridge. Binding of FITC-labeled L-adenosine could be detected using evanescent excitation.[26] A competitive assay was done to measure unlabeled L-adenosine, and was able to detect the analyte over a concentration range spanning three decades, with a detection limit of 100 nM.[17] Another recent example is the immobilization of fluorescently labeled DNA aptamers directly onto a surface to allow reagentless sensing of thrombin.[122] In this case, the binding of thrombin resulted in a significant increase in the anisotropy of the bound probe. The sensor could detect thrombin over three decades of concentration with a detection limit of 5 nM, and required only 10 μL of sample. Based on these values, absolute detection limits of 0.7 amol of thrombin were achieved.

The advantages of DNA based sensors, particular in comparison to other affinity-based systems such as antibodies, is that the binding is inherently reversible, and the biomolecule is far more stable than a protein. In fact, it has been reported that DNA sensors can withstand

sonication or autoclaving with no loss in sensitivity, and remain stable for periods of over a year.[119]

5.7 Multianalyte Sensing with Fluorescent Sensor Arrays

One of the most exciting advances in the area of fluorescence-based biosensors is the development of multianalyte sensors utilizing optical arrays. Such sensors combine two or more different selective chemical regions into a single device, and bring together the areas of device engineering and signal processing to create new devices. Early versions of multianalyte sensors were first described by Li and Walt.[123] This early work used two or three different selective reagents immobilized onto individual polymer pads placed at the tip of an optical fiber. Imaging with a CCD camera gave intensities from each spot, which could be analyzed to give the concentration of multiple analytes. Since this time, a number of multianalyte sensors, sometimes referred to as ''artificial noses'', have been reported, and the applications of such sensors have been reviewed.[124] Most systems have relied on spatial discrimination of individual elements within the array. For example, Narang et al.[125] have described fluoroimmunosensor arrays composed of different sensors for simultaneous detection of explosives such as hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) and TNT, and for simultaneous detection of clinically relevant species such as enterotoxin B (a common food poison), *Yersinia pestis* F1 antigen (an agent of plague), and D-dimer, which can cause pulmonary embolisms and myocardial infarction when present at high concentrations.[126] In the latter case, measurements could be done in whole blood, and the analytes could be quantitated over the physiologically relevant range covering 50 to 1000 ng mL^{-1} with detection times of 15 to 30 min.

Recent examples of arrays have used other methods of discriminating between different selective reactions within an array. In one case, randomly dispersed microbeads were prepared with different ratios of two fluorescent dyes.[127] Beads with ratio **A** were derivatized with chemistry **A**, those with ratio **B** had a different selective biomolecule **B** attached, and so on. Discrimination of different chemical reactions was done on the basis of different fluorescent signatures from the beads. Using such an approach, Walt recently demonstrated the ability to construct an array of beads within small cavities formed in individual fibers of an optical imaging fiber bundle, shown schematically in Figure 5. As many as 20 000 beads could be deposited and individually addressed. In this case, three different selective chemistries were used. However, based on the ability to distinguish different spectroscopic signals from differently modified beads, it should be possible to simultaneously detect up to 150 different analytes using this approach.

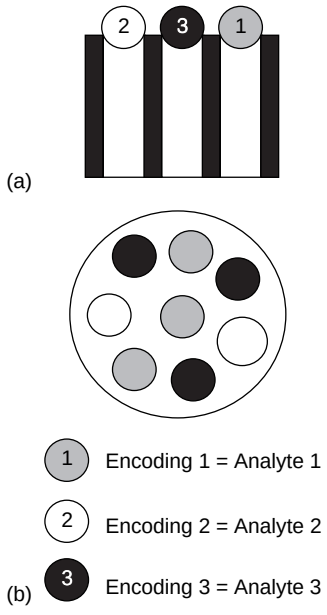

Figure 5 Schematic (a) side and (b) top view of a randomly addressable multianalyte sensor based on an optical imaging fiber and individual fluorescent beads. The pattern of fluorescent beads, as shown in (b), is imaged with a CCD camera. The encoding dye within the bead identifies the analyte associated with a given bead, while the intensity of the analyte signal from an individual bead can be used for analyte quantitation. (Adapted from Michael et al.[127])

6 COMPARISON WITH OTHER METHODS

The advantages and disadvantages of fluorescence-based biosensors can be determined either by comparison with sensors utilizing other transduction strategies, or by comparing such sensors to other analytical techniques. In the first case, it is prudent to compare the abilities of fluorescence-based sensors to those of sensors utilizing electrochemical, piezoelectric, or SPR techniques for signal transduction. In terms of sheer numbers of reports in the literature, fluorescence-based sensors are still a distant second compared with electrochemical sensors.[128]

Electrochemical sensors are still the only commercially available systems for home monitoring, and have the advantage of being able to detect species without the need for labeled analogs or analytes, or labeled proteins. Other advantages include immunity to background signals associated with stray light and biological autofluorescence. Electrochemical sensors often have similar analytical

characteristics (LOD, dynamic range) as fluorescence-based sensors, and suffer some of the same problems, particularly with regard to biomolecule stability. Other limitations inherent to electrochemical sensors are the need for a redox reaction for signal generation, and the inherent sensitivity to electrical and electrochemical interferences. Other areas where fluorescence signals offer unique advantages are in the development of intracellular sensing schemes, and the production of high-density arrays. Multianalyte sensors based on electrochemical signals have been described,[128] but are not as advanced as the arrays which utilize optical imaging fibers for discrimination of multiple analytes.

Comparison of fluorescence-based sensors to piezoelectric or SPR-based sensors shows some of the key advantages of the fluorescence transduction method. The major advantage of piezoelectric or SPR-based sensors is that they do not require labels. However, the lack of labels can also be seen as a drawback, since a limited amount of information is obtained from these techniques. In the case of SPR, one measures a change in ''optical mass'' (a combination of thickness and refractive index) upon analyte interaction with a surface. The signal itself is not information-rich, and it requires that the analyte have sufficient mass, generally over 10 000 MW, in order to produce a signal. Detection of small compounds such as drugs or metabolites is not possible with this method. Piezoelectric sensors rely on changes in mass at a surface, and are exquisitely sensitive when operated in the gas phase. However, most of the important applications for biosensors require operation in the liquid phase, where the operation of piezoelectric devices is not well understood.

A comparison with other analytical techniques can also be made. A key competitor is the enzyme-linked immunosorbent assay (ELISA), which is a standard sandwich assay where the secondary antibody is conjugated to an enzyme rather than a fluorescent probe. These systems can provide higher sensitivity than sandwich fluoroimmunoassays owing to the inherent amplification resulting from the enzyme catalyzed reaction. However, these systems are limited to analytes with molecular weights in excess of ~10 000 Da, since the analyte must have at least two sites for binding of antibodies. Fluoroimmunoassays, on the other hand, can be engineered to allow detection of virtually any substance. ELISA systems are also unsuitable for intracellular studies.

Other typical analytical methods for determination of organic compounds involve hyphenated techniques, such as liquid chromatography/mass spectrometry and GC/mass spectrometry. While these are extremely powerful techniques, and are often able to provide quantitative information for species present in relatively complex matrixes, there are still problems related to separation efficiency and matrix interferences (particularly if one wishes to analyze whole blood), and such methods are expensive and relatively time-consuming. Such methods are not well suited to continuous analysis, and the fact that they consume sample means that they are not reversible. Finally, such methods are not amenable analysis in remote locations, such as the inside of a cell.

7 SUMMARY AND FUTURE PROSPECTS

The development of fluorescence-based biosensors is a growing field that has not yet reached maturity. While most of the work in this area still occurs in the labs of academic institutions, a few commercial devices are now beginning to appear (for example the Ocean Optics oxygen and pH sensors), even though these are chemical sensors as opposed to true biosensors. The commonly employed methods of adapting standard enzyme and antibody-based assay technologies to a sensor format is slowly starting to wane, and a new generation of sensors utilizing proteins that are specifically engineered for sensing purposes is beginning to emerge. This new generation of fluorescence-based sensors is already starting to address some of the long-standing problems with sensor development, including the need for added reagents, and the lack of reversibility for some affinity-based sensors. Work is also under way in the development of improved proteins and immobilization technologies that may alleviate the problems associated with biomolecule stability. Key growth areas for this technology in the next decade will involve the development of quantitative gene chip technologies and improvements in multianalyte sensors, which may be used for quantitative identification of drug leads.

Recent advances in instrumentation are also likely to impact favorably on sensor development. These include the development of low cost fluorescence polarization-based instrumentation,[129] and the application of fluorescence lifetime measurements to fiber-optic sensing.[130] Such instruments reduce many of the problems associated with intensity-based measurements, such as artifacts arising from photobleaching and instrument drift. Other advances include interfacing of flow injection analysis systems to sensors to improve sample introduction,[131] combining sample preparation methods such as solid-phase micro extraction[132] or high-performance liquid chromatography[6] with sensors to improve selectivity, and the development of array-based instruments capable of multianalyte detection.[124] Each of these instrumentation advances makes new applications possible, and improves the reliability of sensors in established applications. Such developments are

sure to provide a bright future for fluorescence-based biosensors.

ABBREVIATIONS AND ACRONYMS

ACh	Acetylcholine
AChR	Acetylcholine Receptor
ATP	Adenosine Triphosphate
BGFP	''Blue'' form of the GFP
BSA	Bovine Serum Albumin
CaM	Calmodulin
CCD	Charge-coupled Device
Cys	Cysteine
Cyt	Cytochrome
ds	Double-stranded
EB	Ethidium Bromide
ELISA	Enzyme-linked Immunosorbent Assay
FAD	Flavin Adenine Dinucleotide
FITC	Fluorescein Isothiocyanate
FRET	Fluorescence Resonance Energy Transfer
GC	Gas Chromatography
GFP	Green Fluorescent Protein
HSA	Human Serum Albumin
IANBD	4-[*N*-(iodoacetoxyethyl)-*N*-methylamino]-7-nitrobenz-2-oxa-1,3,-diazole
IgG	Immunoglobulin G
IR	Insulin Receptor
LED	Light-emitting Diode
LOD	Limit of Detection
MBP	Maltose Binding Protein
MDCC	*N*-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide
MIANS	2-(4′-Maleimidylanilino)naphthalene-6-sulfonic acid
MLCK	Myosin Light Chain Kinase
NADH	Nicotinamide Adenine Dinucleotide, Reduced Form
NDV	Newcastle Disease Virus
OPH	Organophosphorus Hydrolase
P_i	Inorganic Phosphate
PAH	Polyaromatic Hydrocarbon
Pgp	P Glycoprotein
RDX	Hexahydro-1,3,5-trinitro-1,3,5-triazine
RGFP	''Red'' form of GFP
scFv	Single Chains Comprising the Variable Region of the Antigen Binding Fragment
SPR	Surface Plasmon Resonance
TIR	Total Internal Reflection
TIRF	Total Internal Reflection Fluorescence
TNT	2,4,6-Trinitrotoluene
TRITC	Tetramethylrhodamine-5-isothiocyanate
Trp	Tryptophan

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Immunoassay Techniques in Environmental Analyses

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

REFERENCES

1. A.G. Mignani, F. Baldini, 'Biomedical Sensors Using Optical Fibers', *Rep. Prog. Phys.*, **59**, 1–28 (1996).
2. R. Narayanaswamy, 'Chemical Transducers Based on Fiber Optics for Environmental Monitoring', *Sci. Total Environ.*, **135**, 103–113 (1995).
3. B.M. Paddle, 'Biosensors for Chemical and Biological Agents of Defense Interest', *Biosens. Bioelectron.*, **11**, 1079–1113 (1996).
4. S. Marose, C. Lindermann, R. Ulber, T. Scheper, 'Optical Sensor Systems for Bioprocess Monitoring', *Trends Biotechnol.*, **17**, 30–34 (1999).
5. J.H. Luong, P. Bouvrette, K.B. Male, 'Developments and Applications of Biosensors in Food Analysis', *Trends Biotechnol.*, **15**, 369–377 (1997).
6. H.A. Fishman, D.R. Greenwald, R.N. Zare, 'Biosensors in Chemical Separations', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 165–198 (1998).
7. G. Zlokarnik, 'Fluorescent Molecular Sensor for Drug Discovery', *Anal. Chem.*, **71**, 322A–328A (1999).
8. E.A.H. Hall, *Biosensors*, Prentice Hall, Englewood Cliffs, 1991.
9. W.F. Love, L.J. Button, R.E. Slovacek, 'Optical Characteristics of Fiberoptic Evanescent Wave Sensors: Theory and Experiment', in *Biosensors with Fiberoptics*, eds. B. Wise, L. Wingard, Humana Press, New York, 139–180, 1991.

10. A.E. Bruno, S. Barnard, M. Rouilly, A. Waldner, J. Berger, M. Ehrat, 'All-solid-state Miniaturized Fluorescence Sensor Array for the Determination of Critical Gases and Electrolytes in Blood', *Anal. Chem.*, **69**, 507–513 (1997).
11. S.L.R. Barker, Y. Zhao, M.A. Marletta, R. Kopelman, 'Cellular Applications of a Sensitive and Selective Fiber-optic Nitric Oxide Biosensor Based on a Dye-labeled Heme Domain of Soluble Guanylate Cyclase', *Anal. Chem.*, **71**, 2071–2075 (1999).
12. J.E. Whitaker, R.P. Haugland, F.G. Prendergast, 'Spectral and Photophysical Studies of Benzo[*c*]xanthene Dyes: Dual Emission pH Sensors', *Anal. Biochem.*, **194**, 330–344 (1991).
13. J.W. Parker, O. Laksin, C. Yu, M.-L. Lau, S. Klima, R. Fisher, I. Scott, B.W. Atwater, 'Fiber-optic Sensors for pH and Carbon Dioxide Using a Self-referencing Dye', *Anal. Chem.*, **65**, 2329–2334 (1993).
14. I. Gryczynski, Z. Gryczynski, J.R. Lakowicz, 'Polarization Sensing with Visual Detection', *Anal. Chem.*, **71**, 1241–1251 (1999).
15. H. Szmackinski, J.R. Lakowicz, 'Lifetime-based Sensing', in *Topics in Fluorescence Spectroscopy, Volume 4: Probe Design and Chemical Sensing*, ed. J.R. Lakowicz, Plenum Press, New York, 295–334, 1994.
16. Q. Chang, Z. Murtaza, J.R. Lakowicz, G. Roa, 'A Fluorescence Lifetime-based Solid-state Sensor for Water', *Anal. Chim. Acta*, **350**, 97–104 (1997).
17. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum Publishing, New York, 1983.
18. J.R. Lakowicz, *Topics in Fluorescence Spectroscopy, Volume 1: Techniques*, Plenum Publishing, New York, 1991.
19. J.R. Lakowicz, *Topics in Fluorescence Spectroscopy, Volume 2: Principles*, Plenum Publishing, New York, 1991.
20. D.W. Lubbers, 'Fluorescence-based Chemical Sensors', *Adv. Biosensors*, **2**, 215–260 (1992).
21. C. Reid, *Excited States in Chemistry and Biology*, Butterworth, London, 1957.
22. N.J. Harrick, G.I. Loeb, 'Multiple Internal Reflection Fluorescence Spectroscopy', *Anal. Chem.*, **45**, 687–691 (1973).
23. M.N. Kronic, W.A. Little, 'A New Immunoassay Based on Fluorescence Excitation by Internal Reflection Spectroscopy', *J. Immunol. Methods*, **8**, 235–240 (1975).
24. O.S. Wolfbeis, *Fiber Optic Chemical Sensors and Biosensors: Volume II*, CRC Press, Boca Raton, 1991.
25. K.A. Guiliano, P.L. Post, K.M. Hahn, D.L. Taylor, 'Fluorescent Protein Biosensors: Measurement of Molecular Dynamics in Living Cells', *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 405–434 (1995).
26. F. Kleinjung, S. Klussmann, V.A. Erdmann, F.W. Scheller, J.P. Furste, F.F. Bier, 'High-affinity RNA as a Recognition Element in a Biosensor', *Anal. Chem.*, **70**, 328–331 (1998).
27. P.A.E. Piunno, U.J. Krull, R.H.E. Hudson, M.J. Dahma, H. Cohen, 'Fiber Optic Biosensor for Fluorimetric Detection of DNA Hybridization', *Anal. Chim. Acta*, **288**, 205–214 (1994).
28. L.B. McGown, M.J. Joseph, J.B. Pitner, G.P. Vonk, C.P. Linn, 'The Nucleic Acid Ligand: A New Tool for Molecular Recognition', *Anal. Chem.*, **67**, 663A–668A (1995).
29. W.R. Seitz, 'Chemical Sensors Based on Immobilized Indicators and Fiber Optics', *CRC Crit. Rev. Anal. Chem.*, **19**, 135–173 (1988).
30. R. Dessy, 'Slab Waveguides in Chemistry', *Crit. Rev. Anal. Chem.*, **21**, 377–388 (1989).
31. W.H. Scouten, J.H.T. Luong, R.S. Brown, 'Enzyme or Protein Immobilization Techniques for Applications in Biosensor Design', *Trends Biotechnol.*, **13**, 178–185 (1995).
32. S.C. Huang, H. Swerdlow, K.D. Cladwell, 'Binding of Biotinylated DNA to Streptavidin-coated Polystyrene Latex', *Anal. Biochem.*, **222**, 441–449 (1994).
33. S.S. Ghosh, G.F. Musso, 'Covalent Attachment of Oligonucleotides to Solid Supports', *Nucleic Acids Res.*, **15**, 5353–5372 (1987).
34. E.T. Vandenberg, R.S. Brown, U.J. Krull, 'Immobilization of Proteins for Biosensor Development', in *Immobilized Biosystems: Theory and Practical Applications*, eds. I.A. Veliky, R.J.C. McLean, Blackie Academic & Professional, Glasgow, 129–231, 1994.
35. H.H. Weetall, 'Preparation of Immobilized Proteins Covalently Coupled Through Silane Coupling Agents to Inorganic Surfaces', *Appl. Biochem. Biotechnol.*, **41**, 157–188 (1993).
36. P.A. Srere, K. Uyeda, 'Functional Groups on Enzymes Suitable for Binding to Matrices', *Methods Enzymol.*, **44**, 11–19 (1976).
37. B. Lu, M.R. Smyth, R. O'Kennedy, 'Oriented Immobilization of Antibodies and its Applications in Immunoassays and Immunosensors', *Analyst*, **121**, 29R–32R (1996).
38. R.T. Piervincenzi, W.M. Reichert, H.W. Hellinga, 'Genetic Engineering of a Single-chain Antibody Fragment for Surface Immobilization in an Optical Biosensor', *Biosens. Bioelectron.*, **13**, 305–312 (1998).
39. F.V. Bright, T.A. Betts, K.S. Litwiler, 'Regenerable Fiber-optic Based Immunosensor', *Anal. Chem.*, **62**, 1065–1069 (1990).
40. K.F. O'Driscoll, 'Techniques of Enzyme Entrapment in Gels', *Methods Enzymol.*, **44**, 169–183 (1976).
41. D. Avnir, S. Braun, O. Lev, M. Ottolenghi, 'Enzymes and Other Proteins Entrapped in Sol–Gel Materials', *Chem. Mater.*, **6**, 1605–1614 (1994).
42. B.C. Dave, B. Dunn, J.S. Valentine, J.I. Zink, 'Sol–Gel Encapsulation Methods for Biosensors', *Anal. Chem.*, **66**, 1120A–1127A (1994).

43. J.D. Brennan, 'Using Intrinsic Fluorescence to Investigate Protein Entrapped in Sol–Gel Derived Materials', *Appl. Spectrosc.*, **53**, 106A–121A (1999).
44. J.G. Hacia, L.C. Brody, F.S. Collins, 'Applications of DNA Chips for Genomic Analysis', *Mol. Psychiatry*, **3**, 483–492 (1998).
45. K.K. Flora, J.D. Brennan, 'Fluorimetric Detection of Ca^{2+} Based on an Induced Change in the Conformation of Sol–Gel Entrapped Parvalbumin', *Anal. Chem.*, **70**, 4505–4513 (1998).
46. G. Gilardi, L.Q. Zhou, L. Hibbert, A.E.G. Cass, 'Engineering the Maltose Binding Protein for Reagentless Fluorescence Sensing', *Anal. Chem.*, **66**, 3840–3847 (1994).
47. J.D. Brennan, 'Preparation and Entrapment of Fluorescently-labeled Proteins for the Development of Reagentless Optical Biosensors', *J. Fluoresc.*, **9**, 295–312 (1999).
48. T.L. Blair, S.-T. Yang, T. Smith-Palmer, L.G. Bachas, 'Fiber-optic Sensor for Ca^{2+} Based on an Induced Change in the Conformation of the Protein Calmodulin', *Anal. Chem.*, **66**, 300–302 (1994).
49. W. Huang, J. Wang, D. Bhattacharyya, L.G. Bachas, 'Improving the Activity of Immobilized Subtilisin by Site-specific Attachment to Surfaces', *Anal. Chem.*, **69**, 4601–4607 (1997).
50. V. De Filippis, F. De Antoni, M. Frigo, P. Polverino de Laureto, A. Fontana, 'Enhanced Protein Thermostability by Ala → Aib Replacement', *Biochemistry*, **37**, 1686–1696 (1998).
51. W.R. Seitz, 'Transducer Mechanisms for Optical Biosensors. Part 1: The Chemistry of Transduction', *Comput. Methods Programs Biomed.*, **30**, 9–19 (1989).
52. K.R. Rogers, C.J. Cao, J.J. Valdes, A.T. Eldefrawi, M.E. Eldefrawi, 'Acetylcholinesterase Fiber-optic Biosensor for Detection of Anticholinesterases', *Fund. Appl. Toxicol.*, **16**, 810–820 (1991).
53. X. Xie, A.A. Suleiman, G.G. Guilbault, 'Determination of Urea in Serum by a Fiber-optic Fluorescence Biosensor', *Talanta*, **38**, 1197–1200 (1991).
54. M.S. Fuh, L.W. Burgess, G.D. Christian, 'Single Fiber-optic Fluorescence Enzyme-based Sensor', *Anal. Chem.*, **60**, 433–435 (1988).
55. Z. Rosenzweig, R. Kopelman, 'Analytical Properties and Sensor Size Effects of a Micrometer-sized Optical Fiber Glucose Biosensor', *Anal. Chem.*, **68**, 1408–1413 (1996).
56. M.D. Marazuela, B. Cuesta, M.C. Moreno-Bondi, A. Quejido, 'Free Cholesterol Fiber-optic Biosensor for Serum Samples with Simplex Optimization', *Biosens. Bioelectron.*, **12**, 233–240 (1997).
57. B.A.A. Dremel, W. Yang, R.D. Schmid, 'On-line Determination of Lactic Acid During Kefir Fermentation Based on a Fiber-optic Lactic Acid Biosensor and Flow Injection Analysis', *Anal. Chim. Acta*, **234**, 102–112 (1990).
58. J. Cordek, X. Wang, W. Tan, 'Direct Immobilization of Glutamate Dehydrogenase on Optical Fiber Probes for Ultrasensitive Glutamate Detection', *Anal. Chem.*, **71**, 1529–1533 (1999).
59. M.A. Arnold, J. Wangsa, 'Transducer-based and Intrinsic Biosensors', in *Fiber Optic Chemical Sensors and Biosensors*, ed. O.S. Wolfbeis, CRC Press, Boca Raton, 193–216, Vol. II, 1991.
60. K.R. Rogers, Y. Wang, A. Mulchandani, P. Mulchandani, W. Chen, 'Organophosphorus Hydrolase-based Assay for Organophosphate Pesticides', *Biotechnol. Prog.*, **15**, 517–521 (1999).
61. A. Navas Diaz, M.C. Ramos Peinado, 'Sol–Gel Cholinesterase Biosensor for Organophosphorous Pesticide Fluorimetric Analysis', *Sens. Actuators, B*, **38–39**, 426–431 (1997).
62. W. Trettnak, O.S. Wolfbeis, 'Fully Reversible Fiber-optic Glucose Biosensor Based on the Intrinsic Fluorescence of Glucose Oxidase', *Anal. Chim. Acta*, **221**, 195–203 (1989).
63. A.M. Hartnett, C.M. Ingersoll, G.A. Baker, F.V. Bright, 'Kinetics and Thermodynamics of Free Flavins and the Flavin-based Redox Active Site within Glucose Oxidase Dissolved in Solution or Sequestered within a Sol–Gel-derived Glass', *Anal. Chem.*, **71**, 1215–1224 (1999).
64. R.B. Thompson, E.R. Jones, 'Enzyme-based Fiber-optic Zinc Biosensor', *Anal. Chem.*, **65**, 730–734 (1993).
65. R.B. Thompson, B.P. Maliwal, V.L. Feliccia, C.A. Fierke, K. McCall, 'Determination of Picomolar Concentrations of Metal Ions Using Fluorescence Anisotropy: Biosensing with a 'Reagentless' Enzyme Transducer', *Anal. Chem.*, **70**, 4717–4723 (1998).
66. R.B. Thompson, B.P. Maliwal, C.A. Fierke, 'Selectivity and Sensitivity of Fluorescence-lifetime-based Metal Ion Biosensing Using a Carbonic Anhydrase Transducer', *Anal. Biochem.*, **267**, 185–195 (1999).
67. C.L. Morgan, D.J. Newman, C.P. Price, 'Immunosensors: Technology and Opportunities in Laboratory Medicine', *Clin. Chem.*, **42**, 193–209 (1996).
68. B. Hock, 'Antibodies for Immunosensors: A Review', *Anal. Chim. Acta*, **347**, 177–186 (1997).
69. B. Hock, 'Advances in Immunochemical Detection of Microorganisms', *Ann. Biol. Clin.*, **54**, 243–252 (1996).
70. J.P. Alaire, J.R. Bowyer, M.J. Sepaniak, A.H. Hoyt, T. Vo-Dihn, 'Fluorescence Monitoring of a Benzo[*a*]-pyrene Metabolite Using a Regenerable Immunochemical-based Fiber-optic Sensor', *Anal. Chim. Acta*, **236**, 237–244 (1990).
71. R.M. Sutherland, C. Dahne, J.F. Place, A. Ringrose, 'Optical Detection of Antibody–Antigen Reactions at a Glass–Liquid Interface', *Clin. Chem.*, **30**, 1533–1538 (1984).
72. R.A. Ogert, A.W. Kusterbeck, G.A. Wemhoff, R. Burke, F.S. Ligler, 'Detection of Cocaine Using the Flow Immunosensor', *Anal. Lett.*, **25**, 1999–2019 (1992).
73. A.R. Toppozada, J. Wright, A.T. Eldefrawi, M.E. Eldefrawi, E.L. Johnson, S.D. Emche, C.S. Helling, 'Evaluation of a Fiber Optic Immunosensor for Quantitating

Cocaine in Coca Leaf Extracts', *Biosens. Bioelectron.*, **12**, 113–124 (1997).
74. L.C. Shriver-Lake, K.A. Breslin, P.T. Charles, D.W. Conrad, J.P. Golden, F.S. Ligler, 'Detection of TNT in Water Using an Evanescent Wave Fiber-optic Biosensor', *Anal. Chem.*, **67**, 2431–2435 (1995).
75. J.R. Astles, W.G. Miller, 'Measurement of Free Phenytoin in Blood with a Self-contained Fiber-optic Immunosensor', *Anal. Chem.*, **66**, 1675–1682 (1994).
76. C.M. Hanbury, W.G. Miller, R.B. Harris, 'Antibody Characteristics for a Continuous Response Fiber Optic Immunosensor for Theophylline', *Biosens. Bioelectron.*, **11**, 1129–1138 (1996).
77. P. Kumar, J.T. Colston, E.D. Rael, J.J. Valdes, 'Detection of Botulinum Toxin Using a Evanescent Wave Immunosensor', *Biosens. Bioelectron.*, **9**, 57–63 (1994).
78. R.A. Ogert, E.J. Brown, B.R. Singh, L.C. Shriver-Lake, F.S. Ligler, 'Detection of *Clostridium Botulinum* Toxin A Using a Fiber Optic-based Biosensor', *Anal. Biochem.*, **205**, 306–312 (1992).
79. K.L. Coa, G.P. Anderson, F.S. Ligler, J. Ezzell, 'Detection of *Yersinia pestis* Fraction 1 Antigen with a Fiber Optic Biosensor', *J. Clin. Microbiol.*, **33**, 336–341 (1995).
80. D.C. Wijesuriya, G.P. Anderson, F.S. Ligler, 'A Rapid and Sensitive Immunoassay for Bacterial Cells', in *Proceedings of the 1993 ERDEC Scientific Conference on Chemical Defense Research*, eds. D.A. Berg, J.D. Williams Jr, P.J. Reeves, 16–19, 1994.
81. Y.H. Chang, T.C. Chang, E.F. Kao, C. Chou, 'Detection of Protein A Produced by Staphylococcus aureus with a Fiber Optic-based Biosensor', *Biosci. Biotechnol. Biochem.*, **60**, 1571–1574 (1996).
82. U. Narang, G.P. Anderson, F.S. Ligler, J. Burans, 'Fiber Optic-based Biosensor for Ricin', *Biosens. Bioelectron.*, **12**, 937–945 (1997).
83. W.E. Lee, H.G. Thompson, 'Fiber Optic Biosensor Assay of Newcastle Disease Virus', Defense Research Establishment Suffield, Canada, Suffield Report No. 580, 1–36.
84. R.A. Badley, R.A.L. Drake, I.A. Shanks, A.M. Smith, P.R. Stephenson, 'Optical Biosensors for Immunoassays: the Fluorescence Capillary Fill Device', *Philos. Trans R. Soc. London, Ser. B*, **316**, 143–160 (1987).
85. R.P. Parry, C.A. Love, G.A. Robinson, 'Detection of Rubella Antibody Using an Optics Immunosensor', *J. Virol. Methods*, **27**, 39–48 (1990).
86. J.K. Deacon, A.M. Thompson, A.L. Page, J.E. Stops, P.R. Roberts, S.W. Whiteley, J.W. Attridge, C.A. Love, G.A. Robinson, G.P. Davidson, 'An Assay for Human Chlorionic Gonadotropin Using the Fluorescence Capillary Fill Device', *Biosens. Bioelectron.*, **6**, 193–199 (1991).
87. P.B. Daniels, 'A Comparison of 3 Fluorophores for us in an Optical Biosensor for the Measurement of Prostate-specific Antigen in Whole-blood', *Sens. Actuators, B*, **27**, 447–451 (1995).
88. C.M. Hanbury, W.G. Miller, R.B. Harris, 'Fiber-optic Immunosensor for Measurement of Myoglobin', *Clin. Chem.*, **43**, 2128–2136 (1997).
89. J.D. Stewart, V.A. Roberts, M.W. Crowder, E.D. Getzoff, S.J. Benkovic, 'Creation of a Novel Biosensor for Zn(II)', *J. Am. Chem. Soc.*, **116**, 415–416 (1994).
90. G. Walkup, B. Imperiali, 'Design and Evaluation of a Peptidyl Fluorescent Chemosensor for Divalent Zinc', *J. Am. Chem. Soc.*, **118**, 3053–3054 (1996).
91. G.V. Richieri, R.T. Ogata, A.M. Kleinfeld, 'A Fluorescently Labeled Intestinal Fatty Acid Binding Protein: Interactions with Fatty Acids and its Use in Monitoring Free Fatty Acid', *J. Biol. Chem.*, **267**, 23 495–23 501 (1992).
92. P.L. Post, K.M. Trybus, D.L. Taylor, 'A Genetically Engineered, Protein-based Optical Biosensor of Myosin II Regulatory Light Chain Phosphorylation', *J. Biol. Chem.*, **269**, 12 880–12 887 (1994).
93. S.L. Barker, R. Kopelman, T.E. Meyer, M.A. Cusanovich, 'Fiber-optic Nitric Oxide-selective Biosensors and Nanosensors', *Anal. Chem.*, **70**, 971–976 (1998).
94. M. Gerstein, A.M. Lesk, C. Chothia, 'Structural Mechanisms for Domain Movements in Proteins', *Biochemistry*, **33**, 6739–6749 (1994).
95. H.A. Godwin, J.M. Berg, 'A Fluorescent Zinc Probe Based on Metal-induced Peptide Folding', *J. Am. Chem. Soc.*, **118**, 6514–6515 (1996).
96. G.K. Walkup, B. Imperiali, 'Fluorescent Chemosensors for Divalent Zinc Based on Zinc Finger Domains. Enhanced Oxidative Stability, Metal Binding Affinity, and Structural and Functional Characterization', *J. Am. Chem. Soc.*, **119**, 3443–3450 (1997).
97. F.G. Prendergast, M. Meyer, G.L. Carlson, S. Iida, J.D. Potter, 'Synthesis, Spectral Properties, and Use of 6-Acryloyl-2-dimethylaminonaphthalene (Acrylodan). A Thiol-selective, Polarity-sensitive Fluorescent Probe', *J. Biol. Chem.*, **258**, 7541–7543 (1983).
98. A.N. Watkins, F.V. Bright, 'Effects of Fluorescent Reporter Group Structure on the Dynamics Surrounding Cysteine-26 in Spinach Calmodulin: A Model Biorecognition Element', *Appl. Spectrosc.*, **52**, 1447–1456 (1998).
99. J.A. Putkey, W. Liu, X. Lin, S. Ahmed, M. Zhang, J.D. Potter, W.G.L. Kerrick, 'Fluorescent Probes Attached to Cys 25 or Cys 84 in Cardiac Troponin C are Differentially Sensitive to Ca^{2+}-Dependent Events in Vitro and in Situ', *Biochemistry*, **36**, 970–978 (1997).
100. R.Y. Tsien, A. Miyawaki, 'Seeing the Machinery of Live Cells', *Science*, **280**, 1954–1955 (1998).
101. V.A. Romoser, P.M. Hinkle, A. Persechini, 'Detection in Living Cells of Ca^{2+}-Dependent Changes in the Fluorescence of an Indicator Composed of Two Green Fluorescent Protein Variants Linked by a

Calmodulin – Binding Sequence. A New Class of Fluorescent Indicators', *J. Biol. Chem.*, **272**, 13 270–13 274 (1997).
102. A. Persechini, J.A. Lynch, V.A. Rosomer, 'Novel Fluorescent Indicator Proteins for Monitoring Free Intracellular Ca^{2+}', *Cell Calcium*, **22**, 209–216 (1997).
103. J.S. Marvin, E.E. Corcoran, N.A. Hattangadi, J.V. Zhang, S.A. Gere, H.W. Hellinga, 'The Rational Design of Allosteric Interactions in a Monomeric Protein and its Applications to the Construction of Biosensors', *Proc. Natl. Acad. Sci. USA*, **94**, 4366–4371 (1997).
104. M. Brune, J.L. Hunter, J.E.T. Corrie, M.R. Webb, 'Direct, Real-time Measurement of Rapid Inorganic Phosphate Release Using a Novel Fluorescent Probe and its Application to Actomyosin Subfragment 1 ATPase', *Biochemistry*, **33**, 8262–8371 (1994).
105. C. Lionne, M. Brune, M.R. Webb, F. Travers, T. Barman, 'Time Resolved Measurements Show that Phosphate Release is the Rate Limiting Step on Myofibrillar ATPases', *FEBS Lett.*, **364**, 59–62 (1995).
106. A.E. Nixon, M. Brune, P.N. Lowe, M.R. Webb, 'Kinetics of Inorganic Phosphate Release During the Interaction of p21ras with the GTPase-activating Proteins, p120-GAP and Neurofibromin', *Biochemistry*, **34**, 15 592–15 598 (1995).
107. J.S. Marvin, H.W. Hellinga, 'Engineering Biosensors by Introducing Fluorescent Allosteric Signal Transducers: Construction of a Novel Glucose Sensor', *J. Am. Chem. Soc.*, **120**, 7–11 (1998).
108. S.M. Dunn, M.A. Raftery, 'Multiple Binding Sites for Agonists on *Torpedo californica* Acetylcholine Receptor', *Biochemistry*, **21**, 6264–6272 (1982).
109. U.J. Krull, J.D. Brennan, R.S. Brown, S. Hosein, B.D. Hougham, E.T. Vandenberg, *Analyst*, **115**, 147–153 (1990).
110. K.R. Rogers, J.J. Valdes, M.E. Eldefrawi, 'Acetylcholine Receptor Fiber Optic Evanescent Fluorosensor', *Anal. Biochem.*, **182**, 353–359 (1989).
111. R. Lui, J. Sharom, 'Site-directed Fluorescence Labeling of P-Glycoprotein on Cysteine Residues in the Nucleotide Binding Domain', *Biochemistry*, **35**, 11 865–11 873 (1996).
112. J. Lee, P.F. Pilch, S.E. Shoelson, S.F. Scarlata, 'Conformational Changes of the Insulin Receptor Upon Insulin Binding and Activation as Monitored by Fluorescence Spectroscopy', *Biochemistry*, **36**, 2701–2708 (1997).
113. E.L. Schmid, A.-P. Tairi, R. Hovius, H. Vogel, 'Screening Ligands for Membrane Protein Receptors by Total Internal Reflection Fluorescence: The 5HT3 Serotonin Receptor', *Anal. Chem.*, **70**, 1331–1338 (1998).
114. M. Pawiak, E. Grell, E. Schick, D. Anselmetti, M. Ehrat, 'Functional Immobilization of Biomembrane Fragments on Planar Waveguides for the Investigation of Side-directed Ligand Binding by Surface-confined Fluorescence', *J. Chem. Soc., Faraday Trans.*, **111**, 273–288 (1998).
115. R. Campostrini, G. Carturan, R. Caniato, A. Piovan, R. Filippini, G. Innocenti, E.M. Cappelletti, 'Immobilization of Plant Cells in Hybrid Sol–Gel Materials', *J. Sol–Gel Sci. Technol.*, **7**, 87–97 (1996).
116. V.M. Sglavo, G. Carturan, R.D. Monte, M. Muraca, 'SiO_2 Entrapment of Animal Cells, Part I: Mechanical Features of Sol–Gel SiO_2 Coatings', *J. Mater. Sci.*, **34**, 1–4 (1999).
117. A.P. Abel, M.G. Weller, G.L. Duveneck, M. Ehrat, H.M. Widmer, 'Fiber-optic Evanescent Wave Biosensor for the Detection of Oligonucleotides', *Anal. Chem.*, **68**, 2905–2912 (1996).
118. S. Pilevar, C.C. Davis, F. Portugal, 'Tapered Optical Fiber Sensor Using Near-infrared Fluorophores to Assay Hybridization', *Anal. Chem.*, **70**, 2031–2037 (1998).
119. P.A. Piunno, U.J. Krull, R.H. Hudson, M.J. Damha, H. Cohen, 'Fiber-optic DNA Sensor for Fluorometric Nucleic Acid Determination', *Anal. Chem.*, **67**, 2635–2643 (1995).
120. A.H. Uddin, P.A. Piunno, R.H. Hudson, M.J. Damha, U.J. Krull, 'A Fiber Optic Biosensor for Fluorimetric Detection of Triple-helical DNA', *Nucleic Acids Res.*, **25**, 4139–4146 (1997).
121. P.C. Pandey, H.H. Weetall, 'Evanescent Fluorobiosensor for the Detection of Polyaromatic Hydrocarbon Based on DNA Intercalation', *Appl. Biochem. Biotechnol.*, **55**, 87–94 (1995).
122. R.A. Potyrailo, R.C. Conrad, A.D. Ellington, G.M. Hieftje, 'Adapting Selected Nucleic Acid Ligands (Aptamers) to Biosensors', *Anal. Chem.*, **70**, 3419–3425 (1998).
123. L. Li, D.R. Walt, 'Dual-analyte Fiber-optic Sensor for the Simultaneous and Continuous Measurement of Glucose and Oxygen', *Anal. Chem.*, **67**, 3746–3752 (1995).
124. T.A. Dickinson, J. White, J.S. Kauer, D.R. Walt, 'Current Trends in ''Artificial Nose'' Technology', *Trends Biotechnol.*, **16**, 250–258 (1998).
125. U. Narang, P.R. Gauger, A.W. Kusterbeck, F.S. Ligler, 'Multianalyte Detection Using a Capillary-based Flow Immunosensor', *Anal. Biochem.*, **255**, 13–19 (1998).
126. C.A. Rowe, S.B. Scruggs, M.J. Feldstein, J.P. Golden, F.S. Ligler, 'An Array Immunosensor for Simultaneous Detection of Clinical Analytes', *Anal. Chem.*, **71**, 433–439 (1999).
127. K.L. Michael, L.C. Taylor, S.L. Schultz, D.R. Walt, 'Randomly Ordered Addressable High-density Optical Sensor Arrays', *Anal. Chem.*, **70**, 1242–1248 (1998).
128. J. Janata, M. Jasowicz, P. Vanysek, D.M. DeVaney, 'Chemical Sensors', *Anal. Chem.*, **70**, 179R–208R (1998).
129. I. Gryczynski, Z. Gryczynski, J.R. Lakowicz, 'Polarization Sensing with Visual Detection', *Anal. Chem.*, **71**, 1241–1251 (1999).
130. R.S. Brown, J.D. Brennan, U.J. Krull, 'An Optical Fiber-based Spectrometer for Measurement of Fluorescence Lifetimes', *Microchem. J.*, **50**, 337–350 (1994).

131. E.H. Hansen, 'Principles and Applications of Flow Injection Analysis in Biosensors', *J. Mol. Recognit.*, **9**, 316–325 (1996).
132. R.S. Brown, I.S. Kozin, S.P. Tabash, 'Coupling of Solid Phase Micro-extraction Films to Optical Fibres for On-line Detection of Polycyclic Aromatic Hydrocarbons by Laser Induced Fluorescence', in *Proceedings of the Second Biennial International Conference on Chemical Measurements and Monitoring of the Environment*, eds. R. Clement, B. Burk, 441–446, 1998.

High-performance Liquid Chromatography of Biological Macromolecules

Knut Irgum
Department of Chemistry, Umeå University, Sweden

1 Introduction **1**

2 Development Modes in Column Chromatography **3**
2.1 Frontal Chromatography 3
2.2 Displacement Chromatography 3
2.3 Elution Chromatography 4
2.4 Viscous Heating and Column Thermostatting 5
2.5 Fundamental Chromatographic Thermodynamics 5
2.6 Macromolecules and Mass Transfer in Chromatography 7
2.7 Solubility Properties that Distinguish Macromolecules from Small Solutes 9
2.8 Stability of Biological Macromolecules 11

3 Mobile Phases – Eluent **12**

4 Separation Column **12**
4.1 Choice of Support Matrix Material 13
4.2 Polymer-coated Materials 18
4.3 The Biocompatibility Issue 19

5 Separation Chemistries Applicable to High-performance Liquid Chromatography of Biomolecules **20**
5.1 Size-exclusion Chromatography 20
5.2 Normal-phase Liquid Chromatography 24
5.3 Separation Modes Based on Hydrophobic Interactions 26
5.4 High-performance Liquid-affinity Chromatography 30
5.5 Ion-exchange Chromatography and Related Techniques 51
5.6 Chromatofocusing 57
5.7 Mixed-mode Interactions in the Chromatography of Biomolecules 59
5.8 Multidimensional Chromatography – Coupled Column Systems 61
5.9 Slalom Chromatography 62

6 Detection Techniques **62**
6.1 Labeling Reactions for Introduction of Detectable Properties 65

7 Sample Preparation **67**

8 Preferred Separation Techniques for Different Classes of Biomacromolecules **68**
8.1 Oligonucleotides and Polynucleotides 68
8.2 Peptides 69
8.3 Proteins 74
8.4 Carbohydrates and their Conjugates 79
8.5 Validation of Separation Processes 84

9 Trends in Separation Media for High-performance Liquid Chromatography of Biomacromolecules **84**
9.1 Nonporous and Micropellicular Separation Media 85
9.2 Perfusion-type Particles with Large Open or Gel-filled Pores 88
9.3 Media with Reduced Discontinuity – Membranes, Rolled Sheets, and Monoliths 89
9.4 Temperature-responsive Chromatography 91
9.5 Molecularly Imprinted Stationary Phases 92
9.6 Restricted Access Media 92

10 Conclusions **93**

Acknowledgments **93**

Abbreviations and Acronyms **93**

Related Articles **94**

References **94**

1 INTRODUCTION

Liquid chromatography (LC) with its modern counterpart, high-performance liquid chromatography (HPLC), has its roots in the biosciences. Although the discovery of LC is often attributed to the Russian botanist Mikhail Tswett shortly after the turn of the twentieth century,[1] liquid phase chromatographic separations can be traced back to Runge, Goppelsröder and Schönbein[2,3] in the 1860s. A recent survey[3] shows that numerous separations of biological macromolecules were done during the later part of the nineteenth century using ''capillary analysis'', a term used to describe a paper-chromatographic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

procedure remarkably similar to the technique for which Martin and Synge[2] were awarded the Nobel Prize in 1952. The term ''chromatography'' (from Greek *chroma* and *grafein*) was coined by Tswett because he used the technique for separating different-colored plant pigments. Although he may not have invented LC, Tswett should be given credit for having converted adsorption chromatography into a more practical procedure. This was done by packing insoluble inorganic particles in a column, to which mixed solutes were applied and separated by elution with a solvent, the characteristics of column LC.

Chromatography in the liquid phase was practiced with relatively large particles under gravity flow for many years, and the development into HPLC as we know it today took place during the late 1960s and the 1970s. Characteristic of HPLC is that columns are packed with small, evenly sized material, through which eluent is driven by hydraulic pressure, often as high as several tens of MPa. When HPLC was introduced in the 1970s, it was rapidly established as a refined way of performing liquid phase chromatography, and the rationale for coining the term ''high performance'' was obviously the vastly improved separation power that could be achieved using smaller and more narrowly size distributed separation materials. This provided a level of resolution unachievable with low-pressure or open-column systems commonly in use at that time. However, improved separation does come at a price and it was paid for by high pressure, which has sometimes been used as an expansion of the first two letters of HPLC.

In parallel with the developments in LC, a rapid evolution also took place in separation techniques based on partitioning from the gaseous phase. For small molecules that are volatile and stable enough to be separated based on differences in vapor pressure and polarity, gas chromatography (GC) is frequently the separation technique of choice. However, as volatility is connected to the molecular size and polar functional groups of the solute, GC cannot be used for separation of large and temperature-sensitive compounds. Biochemical and biomedical applications have therefore remained the primary application area for HPLC, and separation of biomacromolecules by HPLC is consequently a large and expanding field.

The term ''biological macromolecule'' is ambiguous. For instance, the *International Journal of Biological Macromolecules* defines its scope as ''proteins, carbohydrates, nucleic acids, viruses and membranes''. Of these structures, only the former three can be classified as solitary macromolecular entities. Viruses and membranes are multimolecular structures and will not be treated here, although many of the components that constitute such structures fit the scope of this article. Recent reviews of protein aggregate separation by chromatography following chemical cross-linking,[4] separation of bacteriophages and related particles,[5] and the use of cells and biomolecular assemblies as stationary phases[6] can serve as entry points to this literature.

It is also necessary to define a limit in the lower end of the molecular weight scale, a distinction that perhaps is even more ambiguous. One might argue, for example, that eicosanoids and triglycerides are macromolecules, as they are larger than many peptides, but they are considered as compound classes of their own and thus not included. Through development of advanced techniques such as solid phase synthesis and the polymerase chain reaction (PCR),[7] it has become possible to synthesize and multiply molecules entirely in vitro that are identical to biological macromolecules. Although these molecules can hardly fit the title of this treatise in a strict sense, they should still qualify as biological macromolecules. Consequently, the classes of compounds dealt with are natural and synthetic peptides and proteins, oligo- and polynucleotides, and oligomers and polymers of carbohydrates and their adducts with proteins. Worth noting at this stage is that many materials used for the separation of biomacromolecules are made from polymers or crystalline structures of biological origin, such as agarose,[8] dextran,[9] cellulose,[10] chitin,[11] chitosan,[12,13] diatomaceous earth,[14] and cattle bone.[15]

HPLC has also been launched commercially under different acronyms. An example is ''fast protein liquid chromatography'' (FPLC), which differs from regular HPLC mainly by being performed on instrumentation designed to suit the particular needs of biochemists. There are also techniques that are adjunct to HPLC, such as capillary electrochromatography (CEC),[16–19] micellar electrokinetic chromatography (MEKC),[20–23] countercurrent chromatography (CCC),[24–26] and some modes of field flow fractionation (FFF),[27,28] which could have been part of this overview. However, this treatise will have to limit itself to hydraulically driven techniques employing a liquid eluent, and a rigid solid or gel as a stationary phase.

Because of the importance of the area, HPLC of biological macromolecules has been the subject of a large number of books[29–41] and recent review articles of a general nature,[42–45] as well as works providing overviews of the different subtechniques of HPLC and separation of specific compound groups to a depth not possible in this article. References to works dealing with specific separation modes or analyte groups are given in the relevant sections. There is also a multitude of books covering HPLC theory in general,[46–49] as well as books describing HPLC with a distinctly practical approach.[50–58] These references should be useful both

to seasoned practitioners and newcomers to the field. Finally, the classic text by Snyder and Kirkland[59] is still well worth referring to. The number of references possible to provide in this format is limited. When choosing between ''original work'' and recent papers central to the technique, the latter have often been chosen, as this provides a better route into the literature.

2 DEVELOPMENT MODES IN COLUMN CHROMATOGRAPHY

There are several ways of applying samples to a chromatographic column and subsequently developing the chromatogram. The most common way of separating substances in HPLC is elution chromatography. However, other modes of development are also applicable in high-resolution bioseparations. The more important of these techniques are frontal chromatography and displacement chromatography. Both these techniques stem from the early days of chromatography but remain useful even today, in particular for preparative applications. In addition to the general elution principles outlined here, some separation chemistries may be amenable to more specific development schemes, such as chromatofocusing, which is a variation of gradient ion-exchange chromatography (IEC).

2.1 Frontal Chromatography

A frontal chromatographic experiment commences by equilibrating the separation material with a solution containing a replenisher that is weakly adsorbed onto the column. The column bed is then flushed continuously with a solution containing a mixture of the solutes to be separated. The separation is based on relative affinity to the separation material, and solutes that are interacting more strongly than the displacer with the stationary phase will replace the displacer compound and become attached to the column. When a larger volume of the solute mixture is applied to the column, solutes will act as displacers for each other. Consequently, the most strongly retained solute will occupy the sites at the top of the column, whereas more weakly interacting compounds are being forced further down to find their attachment sites. The process evolves as described in Figure 1.

The advantage of frontal chromatography is the high capacity, because a large part of the sorption capacity of the column is used simultaneously in the separation process. The main disadvantage is the inability to isolate compounds other than the most weakly retained solute, unless a separate eluent compound is used. In that case the technique changes to displacement chromatography.

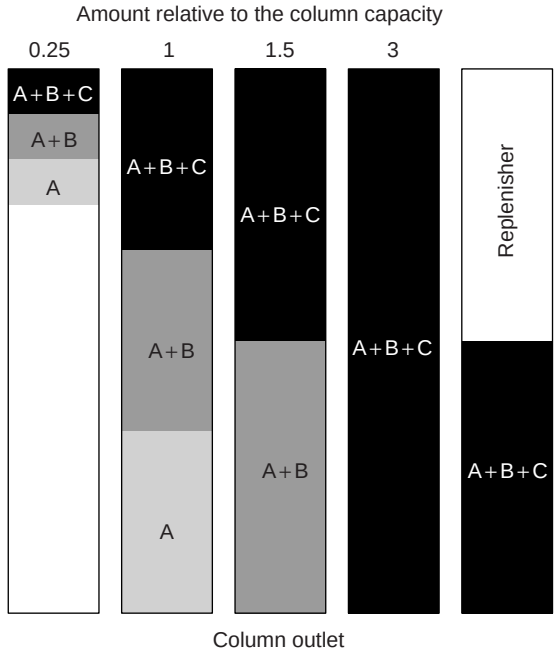

Figure 1 Diagram of the elution process in frontal chromatography. As the sample containing a mixture of solutes A, B, and C is pumped into the column, the solutes arrange themselves according to increasing affinity to the stationary phase, with the least strongly retained solute (A) eluting first, followed by more strongly retained compounds (B and C). The first solute (A) breaks through when the combined amount of solutes pumped equals the sorption capacity of the column. Only this solute emerges fully separated from the other. Subsequent bands contain the compound currently being eluted plus the previously eluted solutes at their concentrations in the sample. After a frontal separation, the column is conditioned for the next run by a replenisher (R) capable of removing the most strongly retained solute from the column.

2.2 Displacement Chromatography

As just mentioned, displacement chromatography is an extension of frontal chromatography and starts by loading the separation column with a sample slug that becomes adsorbed in a frontal chromatographic fashion. An eluent is thereafter applied, containing a displacer compound having a ligand affinity higher than that of any of the adsorbed solutes in the separation column. This sequentially elutes the solutes as adjacent rectangular bands in order of increasing affinity for the column ligands, entrained by the displacer. After the separation is completed, the separation column is prepared for the next cycle by flushing out the displacer with a regenerant solution and re-equilibration with the mobile phase. The procedure is shown in Figure 2. This can be a lengthy procedure unless a reaction can be used to alter the displacer affinity, as the regenerant should replace the displacer by a substantially more weakly attached compound. Although displacement

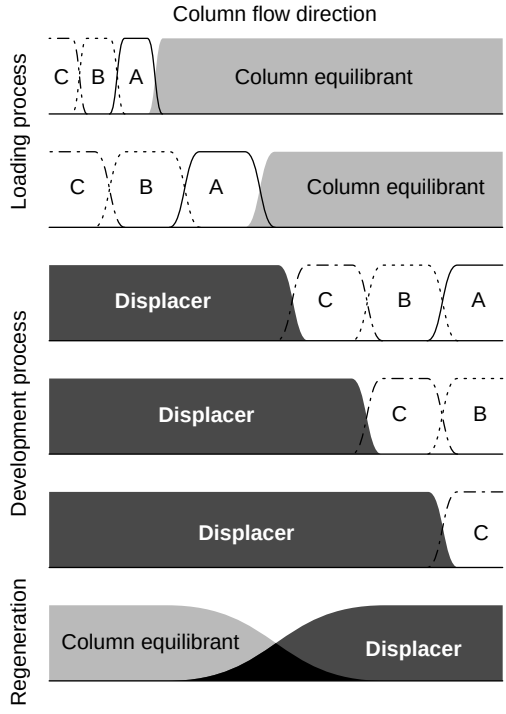

Figure 2 Diagram of the separation process in displacement chromatography. The column is loaded with a sample slug containing a mixture of solutes and is then flushed clean with a noneluting solution, followed by pumping of a displacer, which selectively elutes one or more of the absorbed solutes from the column. The solutes are displaced as bands, the widths of which are related to the solute concentration. After a separation run, the displacer must be removed from the column to enable sorption of solutes in the following run.

chromatography has a resemblance to gradient elution chromatography (see below), there are some significant differences that stem from the use of a displacer with substantially higher affinity than the solutes for the ligands. In gradient elution chromatography the solutes are (ideally) eluted as individual bands migrating at a velocity lower than the eluent and the eluted solutes are surrounded by the displacing compound. Displacement chromatography is therefore best described as gradient elution chromatography of severely overloaded samples using a strong eluent with two or more sequential components at limited concentration.

Displacement chromatography is an important development mechanism in affinity separation schemes, in downstream processing,[60] and in solid phase extraction (SPE) related to sample preparation where it is used to isolate solutes of interest and eliminate sample matrix components. An interesting recent development in displacement chromatography of proteins is the use of dendritic polymers[61] and block copolymers of methacrylic acid, methyl methacrylate, and 2-(dimethylamino)ethyl methacrylate[62] as eluting parties in hydrophobic interaction chromatography (HIC). Such designed displacing agents show a way toward biomimetic interactions with sorbents, providing for other selective elution techniques in the future.

2.3 Elution Chromatography

In elution chromatography, the sample is introduced in a finite volume plug at the head of the separation column, whereafter a suitable eluent is used to force the sample components through the column at solute-specific migration speeds, which are determined by the flow rate of the eluent and the distribution of solute between the mobile and stationary phases (Figure 3). The elution can be carried out isocratically (i.e. with eluents of constant composition) or by varying the concentration of one or more of the components in a stepwise or continuous gradient fashion, often involving several gradient segments. The similarity between displacement elution and gradient elution is obvious.

Isocratic elution is rarely used for elution of biomacromolecules in chromatographic modes involving sorption equilibria (partitioning), for reasons explained below. The separation efficiency would be too low due to mass transfer limitations caused by the low diffusion rates of

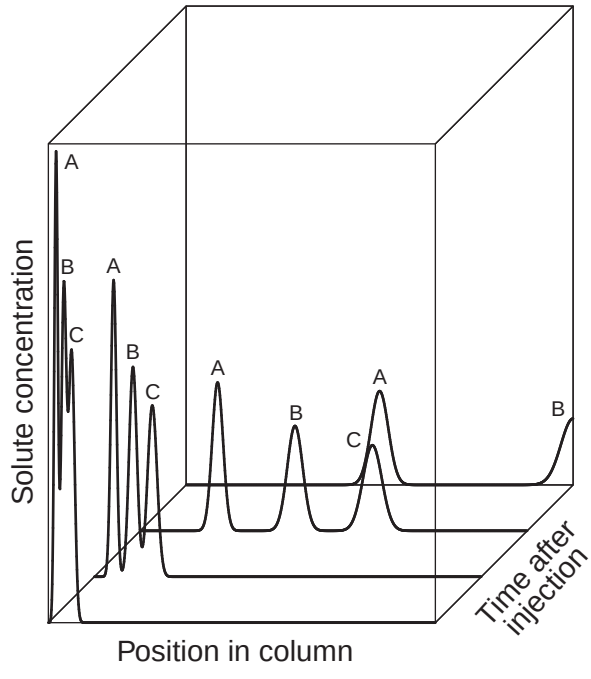

Figure 3 Diagram of the separation process in elution chromatography. A band of solutes is added to the top of the column and eluted with an eluent capable of developing different and suitably strong interactions between the solutes and the stationary phase. Solutes are eluted from the column as bands in order of increasing interaction strength with the column packing.

large molecules. Large differences in partition coefficients of macromolecules also make it difficult to separate even similar components using an isocratic eluent.

2.3.1 Gradient Elution

An eluent gradient[63] can be formed by varying the solvent composition in the eluent during a run, but more common in the elution of biomacromolecules is to vary the concentrations of salts, acids or bases, or other ancillary agent added to the solvent, which is often water. Gradient elution can also be accomplished by alterations in physical parameters such as temperature[64,65] and/or flow rate.[66] Linear gradients are often used, but modern instruments are capable of forming gradients of complex shapes with segments of different steepness, curvatures, step changes, and plateaus. Optimization of gradient elution is dependent on the chromatographic mode used, and a slow linear gradient is a sound first attempt for most biomacromolecule separations. If a fast gradient is used and solutes are adsorbed irreversibly, a judicious effort at improving the recovery would be to decrease the steepness of the gradient. The reason for this is best explained by studying the shape of the retention versus eluent concentration for curves B and D in Figure 46 below, along with the text adjoining that figure. Examples of gradients with many different shapes are found in the chromatograms presented later in this chapter.

Eluent demixing[67] is a phenomenon that takes place when the strong component of an eluent is retained on the separation material in a frontal chromatographic fashion, causing the eluent composition at the column outlet to deviate from the programmed shape. For instance, if a mobile phase with a linearly changing composition is pumped through the column, the shape of the concentration profile after passing the sorbent bed will often not reflect the composition at the column head, but instead emerge with a slower than intended increase during the initial part of the run followed by a steep step increase (a shock layer), after which the shape will follow the programmed shape with reasonable fidelity. The sudden concentration increase caused by demixing can lead to artifacts in the central part of a chromatogram, which can be seen as noticeably narrower peaks, shoulders, low recovery, or impaired resolution in this region.[68] The extent of demixing is dependent on depletion of the strong component from the stationary phase and will decrease if the weak eluent does not contain the strong eluent component, or if it is present at low concentration. IEC is particularly sensitive to demixing if eluent ions with different affinity for the fixed ionic groups are used in a gradient.[69] It is also important to thoroughly equilibrate the column with the weak eluent between each gradient run. If complete equilibration cannot be accomplished due to slow replacement of the stronger eluent component, the time between injections should at least be kept constant.

2.4 Viscous Heating and Column Thermostatting

Column temperature is also a source of concern. Properties of biological macromolecules are affected by temperature, and changes in protein conformation, melting of DNA, thermally controlled aggregation, etc. will influence the retention. Temperature is thus one of the optimization parameters (which is also evident from Equation 6). When the temperature is varied, attention should be paid to viscous heating, which occurs as the solvent forces its way through the packed bed[70] and can lead to a column temperature higher than expected. The power generated by the pump (i.e. the pressure drop times the flow rate) is converted into heat that is partly dissipated by heat exchange through the column walls and partly by the eluent itself. Due to this dual heat dissipation mechanism, both axial and radial temperature gradients can be encountered in columns. The magnitude of these thermal gradients will depend on back-pressure, flow rate, temperature, and thermal capacity of the eluent, the material, diameter and length of the column, and the preset temperature and efficiency of the column thermostatting device. Large-diameter columns with small particle-size sorbents are most affected by these effects.

2.5 Fundamental Chromatographic Thermodynamics

In order to rationalize many of the modern directions in chromatography of biological macromolecules, it is necessary to provide a brief background to the mechanisms that control retention and band broadening in elution HPLC. The data required to determine these thermodynamic and kinetic performance parameters of a column are acquired by running a set of test solutes under isocratic conditions. A chromatogram should thereby be obtained with peaks for the individual solutes, as shown in Figure 4.

By adding a suitable marker substance that passes the column without retention and is visible in the detector, we can determine the void time, t_0. The retention times (t_r) for the solutes must also be obtained, by measuring the time elapsed from the injection to the apex of the peak of the respective solute. (The retention time should actually be based on the center of gravity of the peak, but for Gaussian peaks like those shown in Figure 4 the peak apex and the center of gravity coincide. Determination of the dead volume is not uncomplicated either; general texts on HPLC discuss these problems.) For each peak, we then proceed by extracting the peak width, either at the base (t_w), or at half the peak height ($t_{w1/2}$). With these

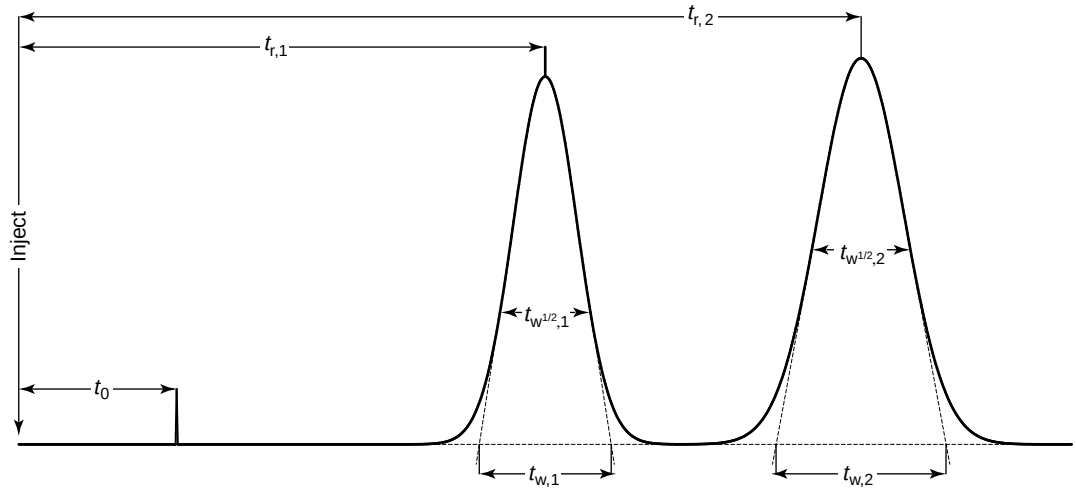

Figure 4 A simulated chromatogram showing the void time (t_0), retention times for solutes 1 ($t_{r,1}$) and 2 ($t_{r,2}$), and widths of these peaks at the base (t_w) and at half-height ($t_{w^{1/2}}$). These data form the basis for calculating the two most fundamental parameters in chromatography, i.e. the capacity factor (k') which is related to the column thermodynamics, and the number of theoretical plates (N) which describes the column efficiency and is mainly determined by the kinetics of the separation process.

data we can for each solute calculate its capacity factor (k'), which is a dimensionless number used to normalize the retention by making it independent of flow rate and column volume (Equation 1):

$$k' = \frac{t_r - t_0}{t_0} \tag{1}$$

What k' actually describes is the ratio of time spent by the solute in the stationary phase to the time spent in the mobile phase. For this to be accomplished, there has to be an identical ratio between the number of solute molecules in the stationary phase, n_s, and in the mobile phase, n_m, as the solute traverses the column (Equation 2):

$$k' = \frac{n_s}{n_m} \tag{2}$$

Considering the volumes of the mobile V_M and stationary V_S phases, and the phase ratio Φ which is the quotient of these two volumes, it is possible to express k' as Φ times the equilibrium constant K of the partitioning reaction, i.e. the ratio of the molar concentrations of solute X in the two phases of the column ($[X_S]$ and $[X_M]$, respectively), as in Equation (3):

$$k' = \frac{[X_S]V_S}{[X_M]V_M} = \Phi\frac{[X_S]}{[X_M]} = \Phi K \tag{3}$$

A prerequisite for obtaining separation between a pair of solutes is obviously that they have different k', and the degree of separation is expressed by the selectivity coefficient α as

$$\alpha = \frac{k'_2}{k'_1} \tag{4}$$

In Equation (4) the value of α is always ≥ 1, because k'_2 refers to the last eluted peak of the pair under consideration.

As K is related to overall energy change of the separation process through the Gibb's equation (Equation 5),

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -RT\ln K \tag{5}$$

we can relate k' directly to the fundamental thermodynamic parameters of the separation process through Equation (6):

$$\ln k' = \frac{\Delta H^\circ - T\Delta S^\circ}{-RT} + \ln \Phi \tag{6}$$

The selectivity according to Equation (4), i.e. how different solutes distribute between the mobile phase and the stationary phase, is thus related to the thermodynamics of the separation process. Equation (6) shows how the enthalpic and entropic contributions affect the separation, which can be mediated by sorption, repulsion, exclusion, or other processes in the column. Differences in Gibb's free energy of sorption are thus required for solutes to separate from each other according to Equation (7),

$$\log \alpha = \frac{|\Delta(\Delta G^\circ)|}{RT} = \frac{|\Delta(\Delta H^\circ) - T\Delta(\Delta S^\circ)|}{RT} \tag{7}$$

an energy difference that can obviously come from differences in heat of sorption (enthalpy, $\Delta(\Delta H^\circ)$) or from changes in order (entropy, $\Delta(\Delta S^\circ)$). Apart from their utility as separation tools, chromatographic techniques are consequently powerful means for obtaining fundamental thermodynamic parameters from sorption processes. For instance, van't Hoff plots ($\ln k'$ versus $1/T$)

are common when assessing the retention as a function of temperature.[71] An important point is that the parameters obtained in this way are valid only for the solutes tested, and under the specific conditions used during the test.

2.6 Macromolecules and Mass Transfer in Chromatography

Whereas the thermodynamic factors just discussed are responsible for establishing the selectivity of a separation process, its efficiency, i.e. how well an injected sample zone is kept from broadening as it traverses across the separation column, is controlled by kinetic and fluid dynamic factors. This is why the peak width data that were extracted in Figure 4 have not yet been used. In most separation modes the peak width is not related to differences in the column equilibria, but to limitations in mass transfer, sorption kinetics, and flow dynamics of the flow system downstream from the point of injection. Band broadening will occur whenever the zone initially injected is distributed across a larger part of the column than the originally injected zone. This dispersion can have its explanation in processes taking place in the column, but also in extracolumn effects originating from less than optimal system plumbing. Only the band broadening that occurs inside the column will be treated here. Guidelines to avoid extracolumn band broadening are found in general HPLC texts.

The common measure of separation efficiency of a separation column is the number of theoretical plates N, which can be derived from t_r and the time variance σ_t of an injected band or, alternatively, from the width of the peak at the base, t_w:

$$N = \left(\frac{t_r^2}{\sigma_t^2}\right) = 16\left(\frac{t_r}{t_w}\right)^2 = 16\left(\frac{t_0(1+k')}{t_w}\right)^2 \qquad (8)$$

The last equality in Equation (8) shows a way of obtaining N from k' and t_0, instead of t_r. An alternative to using t_w is also to use the width at half the height $t_{w1/2}$, in which case the constant 5.54 is used instead of 16.

The concept of theoretical plates is derived from distillation theory and visualizes the number of separation instances that a solute experiences on its way through the column. A common way to describe the quality of a column packing without having to know the length L of the column is to calculate the height equivalent to a theoretical plate, H, whose inverse relationship with N means it should be as low as possible to obtain good column efficiency (Equation 9):

$$H = \frac{L}{N} \qquad (9)$$

In order to compare column packings with different particle diameters, it is convenient to use the reduced plate height h, which is the height equivalent to a theoretical plate divided by the particle diameter (Equation 10),

$$h = \frac{H}{d_p} \qquad (10)$$

as well as the reduced velocity v, which is a product of the linear flow velocity u (equal to L divided by t_0) and the particle diameter, divided by the diffusion coefficient of the solute in the mobile phase D_m (Equation 11),

$$v = \frac{ud_p}{D_m} \qquad (11)$$

The factors contributing to the reduced plate height can then be modeled as functions of the reduced velocity by the van Deemter–Knox equation[72] (Equation 12):

$$h = Av^{1/3} + \frac{B}{v} + Cv \qquad (12)$$

The enemies of a good separation are all the factors that act towards an increase in h, and the important ones are shown schematically in Figure 5.

The factor A in Equation (12) relates to the band broadening resulting from eddy diffusion, whereby solutes are dispersed along the packing as the mobile phase percolates the stationary phase particles in a tortuous flow, where the different eluent flow streams experience pathways of varying flow resistance between the particles. The value of the constant A is roughly proportional to the particle diameter d_p (Equation 13):

$$A \propto d_p \qquad (13)$$

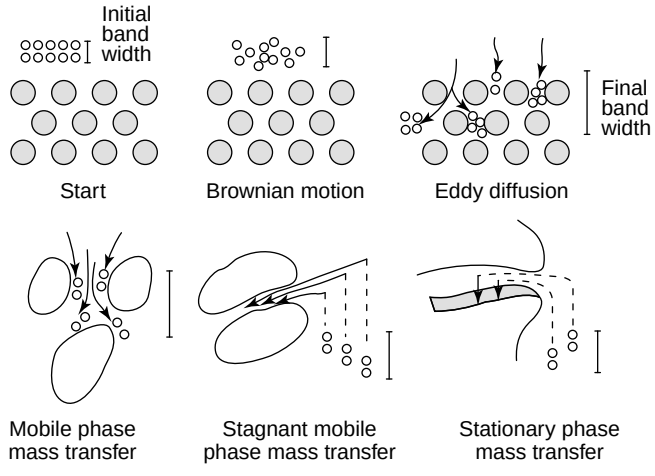

Figure 5 Diagrams of the factors that contribute to band broadening of a solute zone in a packed column. (Reproduced from C.T. Mant, R.S. Hodges, *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, 101, 1991, by permission of CRC Press.)

The eddy diffusion term is also affected by the shape of the particles and the quality of packing. A well packed column bed of small, uniformly shaped particles thus produces the lowest contribution to eddy diffusion.

The second term, B/v, relates to the axial diffusion in both phases. However, the most important contribution originates in the mobile phase, due to diffusion and Brownian motion. The inverse relationship with the reduced velocity reflects lower peak dispersion when the linear flow rate of the eluent is increased because the time available for diffusion decreases. The magnitude of B is also proportional to D_m (Equation 14):

$$B \propto D_m \tag{14}$$

Some representative values for diffusion coefficients of various biological compounds in water are listed in Table 1. It is evident that large molecules such as proteins may diffuse up to several orders of magnitude slower than small molecules.

At very low eluent flow rates, the factor B may become significant for small molecules. In separations of biological macromolecules, characterized by low diffusion coefficients, mass transfer resistance outweighs longitudinal diffusion for all practical separation situations. The separation efficiency for large molecules is thus largely controlled by the third term of the van Deemter–Knox equation, Cv, which relates band broadening to mass transfer resistance.

An important reason why peaks broaden can be found in local deviations from equilibrium across the solute band. If an ideal linear sorption isotherm is considered, which incidentally will hardly ever occur in the separation of biomacromolecules, the solute concentrations in the stationary phase should mirror concentrations in the mobile phase across the solute band as visualized in the left part of Figure 6, with a concentration ratio K according to Equation (3). This represents the ideal case. However, associated with the surface of a particle is a layer of nonmoving mobile phase, surrounded by a film of liquid flowing slower than the bulk flow. Whenever a solute diffuses from the bulk mobile phase into these regions of lower flow velocity, it will experience retardation relative to the bulk flow due to reasons unrelated to the partition process. Further, if the stationary phase is porous, there are regions in the intraparticle space where the mobile flow is essentially immobile (stagnant pools) through which a solute must diffuse in order to interact with the stationary phase. As the mobile phase flows continuously while these processes take place, the column equilibrium is perturbed and the concentration within the stationary phase will be different from that of the compound in the mobile phase. This eventually leads to dispersion of the injected solute along the column, as the system attempts to reach equilibrium. This process is illustrated in the right part of Figure 6. Deviation from local equilibrium depends on the overall kinetics of the partitioning reaction within the column and is particularly important for biomacromolecules, because of their low diffusion rates. Factors that make up for the factor C in Equation (12) involve contributions from different impeding processes (Equation 15):

$$C = \frac{k_1 d_p^2}{D_m} + \frac{k_2 d_s^2}{D_s} + k_3 \tag{15}$$

The first term is related to the mass transfer in the mobile phase, for which there is an inverse relationship with the solute diffusion rate in the mobile phase and a quadratic dependence on particle size. The second represents the band broadening due to mass transfer limitations in the stationary phase. Here again, an inverse relationship with

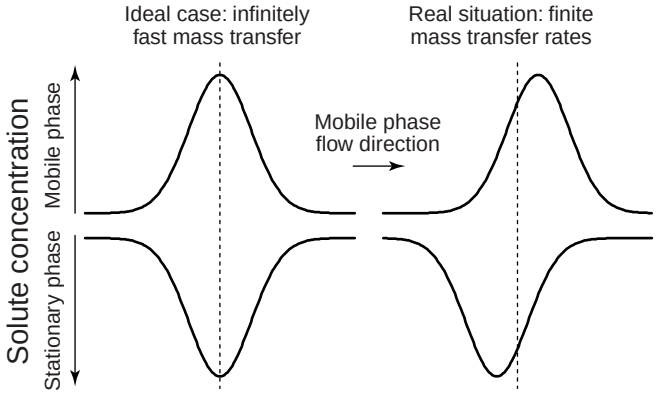

Figure 6 Illustration of the offset in concentration profiles across eluting peaks due to nonequilibrium conditions caused by the moving eluent in combination with mass transfer limitation. The left curve shows the ideal case, where the mobile and stationary phase concentrations are mirrored according to an ideal linear adsorption isotherm. In the right part of the figure, the mobile phase concentration profile runs ahead of the concentration profile in the stationary phase because of a finite mass transfer rate between the phases, leading to band broadening. (Reproduced by permission from Karger.[75])

Table 1 Diffusion coefficients D for some small molecules and biological macromolecules in water at 305 K[74]

Molecule	MW ($g\,mol^{-1}$)	$D(\times 10^{-5}\,cm^2\,s^{-1})$
Glycine	75	1.0
Sucrose	342	0.54
Milk lipase	6669	0.15
Insulin	24 430	0.077
γ-Globulin	153 100	0.042
α-Macroglobulin	820 000	0.025
Urinary mucoprotein	7 000 000	0.0034

the diffusion coefficient of the solute applies, this time in the stationary phase D_s. There is also a quadratic contribution from the stationary phase interaction layer thickness d_s. However, the interaction layer thickness can in general be neglected for biomacromolecules, because the interaction of these molecules with a rigid stationary phase typically takes place as a monolayer process at the surface of the pore system of the support. The third term is due to the kinetics of the actual interaction process, and can also be considered insignificant, unless the formation and breaking of covalent bonds or multiple strongly oriented interactions are involved in the sorption process, as in affinity chromatography.

In summary, the factors that lead to minimization of h, thus promoting a better separation efficiency, are:

- *Smaller particle size* – Minimizes the eddy diffusion and makes the mobile phase mass transfer more efficient. The penalty is a higher operating pressure and, ultimately, a risk of clogging the separation column.
- *Decreased mobile phase flow rates* – The influence of the mass transfer resistance term decreases as the flow rate is lowered. There is also an inverse relationship to flow rate in the factor B, but mass transfer resistance contributes far more to band broadening than longitudinal diffusion under normal HPLC flow rates – particularly for macromolecular solutes. The trade-off is slower separation and a higher risk of denaturation of proteins.
- *High solute diffusion rates in the mobile phase* – Solutes can partition more readily. At very low flow rates, increased D_m leads to more dispersion by longitudinal diffusion. Under flow rates typical of HPLC, this term is insignificant. Besides, D_m is a fundamental property of a solute, not an easily accessible experimental variable.
- *Low-viscosity mobile phases* – The D_m is a function of the mobile phase viscosity, which in turn depends on the temperature. Increased temperature usually leads to more efficient separations, but also increases the likelihood of conformational changes and denaturation of biopolymers. The same applies to the use of low-viscosity organic solvents.
- *Shallow and permeable stationary phase films* – Stationary phases with superficial (pellicular) interaction layers or particles with convective internal flow (see below) minimize the intraparticle diffusion pathways. Gel networks should be sufficiently permeable to avoid diffusion limitations.

More detailed treatments of the kinetics in LC of biopolymers can be found in appropriate texts.[76]

2.7 Solubility Properties that Distinguish Macromolecules from Small Solutes

In spite of the mass transfer limitations just mentioned, practitioners of HPLC familiar with the separation of small molecules may be surprised by the high separation efficiencies that can be achieved for biopolymers on columns of odd dimensions,[77,78] with particles of awkwardly large size, and under low back-pressure,[79] conditions quite different from those considered typical of HPLC of small molecules. The reason why such unconventional media formats can be used in biomacromolecule separations are to be looked for, at least partly, in the solvation mechanisms of macromolecules, and in utilization of multiple contact points for the interactions between the biomacromolecule and the stationary phase interaction layer.

The key to understanding how biomacromolecules behave in a chromatographic system is to recognize the mechanisms that keep these molecules in solution. Polypeptides and oligonucleotides are amphiphilic molecules having both hydrophilic and hydrophobic areas within the same molecule. This means that their surfaces exhibit areas with either polarity or heterogeneous charge distribution. Molecules in solution are solvated, i.e. they have to interact with the surrounding solvent molecules to stay in solution, and the process of transferring a molecule from the stationary phase into the mobile phase involves recreating a cavity for the molecule in the solvent. In polar solvents, where the intermolecular interactions are strong, breaking of bonds between solvent molecules can require substantial energy and the concomitant formation of bonds of similar strength to the solute. If the solute cannot provide for interactions acceptable by the solvent, a substantial amount of energy is required to form a cavity, and the transfer into the solvent is therefore unlikely. This, rather than the stationary phase hydrophobicity, explains why compounds without polar groups cannot dissolve in water, which is the highest-polarity solvent in common use in HPLC. Likewise, the process of dissolving a polar substance involves the breaking of polar bonds between solute molecules. This cannot be accomplished by a nonpolar solvent, which has no polar groups to offer as alternative bonding sites.

There is nothing unique to large molecules in relation to this general principle. The difference compared with small molecules lies in the concerted solvation action needed to stabilize a biomacromolecule with its many heterogeneous interaction sites. The stability of most biopolymers in solution is thus limited. Minor changes in the mobile phase are often sufficient to bring about an aggregation or a dissolution of a macromolecule, or to attach it to, or detach it from, a heterogeneous

surface. This affords excellent possibilities to achieve band separation, as determined by the selectivity factor α through changes in Gibb's free energy, according to Equation (7). The most important factors contributing to changes in Gibb's free energy are electrostatic and van der Waals type interactions, solvation effects, formation of the solvent cavity around the biopolymer, its self-association and heterogeneous associations in the absence of solvent, and deviations from ideality in the sorption–desorption process.[80]

The various energies involved in each interaction point between a macromolecular solute and the two competing phases are additive. The equilibrium process for a peptide or a protein P interacting with a stationary phase and its ligands L can thus be expressed by Equation (16),

$$\mathrm{P(Solv)}_a + n\mathrm{L(Solv)}_b \rightleftharpoons \mathrm{P(Solv)}_{(a-f)}n\mathrm{L} + (nb+f)\mathrm{Solv} \tag{16}$$

where $\mathrm{P(Solv)}_a$ represents a solute with a moles of bound solvent (typically water in biomacromolecule separations) or solvated ions, $n\mathrm{L(Solv)}_{(a-f)}$ represents the ligand attached to b moles of solvent or eluent ions, f is the number of solvent molecules or eluent ions released from the solute when binding to the ligands, and $\mathrm{P(Solv)}_{(a-f)}n\mathrm{L}$ represents the partially solvated solute attached to the immobilized ligands.

Interactions between the biomacromolecule and the mobile phase are equally important as its interactions with the ligands located on the sorbent surface. It has been found empirically that within the retention factor range $0 < k' < 20$, the relationship between k' and the displacing agent concentration c can be described by the following approximately linear relationship,

$$\log k' = A + B\left(\log \frac{1}{c}\right. \tag{17}$$

Equation (17) applies to widely different interaction modes,[81] such as ion exchange (see below for a more elaborate discussion of the Z value in this separation mode; Equation 18),

$$\log k' = \log k'_{0,\mathrm{IEX}} - Z\log[\mathrm{salt}] \tag{18}$$

reversed phase (RP) (Equation 19),

$$\log k' = \log k'_{0,\mathrm{RP}} - M\log[\mathrm{solvent}] \tag{19}$$

and separations based on hydrophobic interactions (Equation 20),

$$\log k' = \log k'_{0,\mathrm{HIC}} - H\log[\mathrm{salt}] \tag{20}$$

The terms $k'_{0,\mathrm{IEX}}$, $k'_{0,\mathrm{RP}}$, and $k'_{0,\mathrm{HIC}}$ represent the retention factors at unit concentrations of the displacing agents, where log(displacing agent) is zero, and the stoichiometric retention parameters Z, M, and H are the sign-inverted slopes of the plots of $\log k'$ versus the logarithms of the salt or solvent concentrations. These figures are related to the number of contact points involved in the interactions between the solute and the stationary phase. The values of Z, M, and H are typically in the range 2–10 for biopolymers, which means that their retention can be altered considerably by changes in the eluent strength, as illustrated by Figure 7.

As a consequence of this steepness of the k' versus eluent strength plots, biomacromolecular solutes often exhibit sharp transitions from soluble to associated or precipitated state according to the all-or-nothing principle coined by Tiselius,[82] producing what can be regarded as traffic light chromatography. High performance does therefore not necessarily require the use of small particles and high pressure in separation modes designed for biomolecules, where multiple strong interactions are exploited and controlled by gradient elution. Nor are long columns required or even advantageous in gradient elution,[83] as a single interaction instance is in principle sufficient for the capture and release required to accomplish separation. In such settings, high resolution can be achieved even using relatively large-size packing materials, shallow beds (sometimes even in the form of a membrane[84]), and even soft gels.

When a chromatographic method is being developed, the required separation can be accomplished through a combination of selectivity, retention, and efficiency

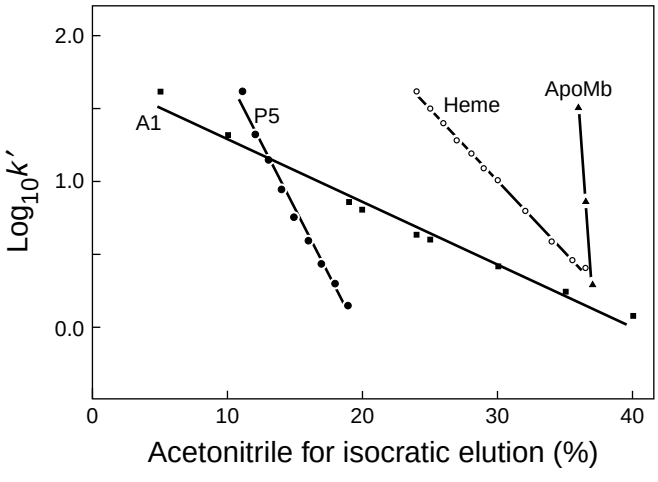

Figure 7 Effect of acetonitrile concentration on the $\log k'$ values of a small molecule represented by propiophenone (A1), a decapeptide (P5), the protein apomyoglobin (ApoMb), and the heme group of this protein (Heme) eluted isocratically on a Brownlee Labs Aquapore RP-300 C_8 (220×4.6 mm internal diameter (i.d.); 7 µm; 300 Å) column. (Reproduced from C.T. Mant, R.S. Hodges, *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, 7, 1991, by permission of CRC Press.)

suitable for that separation. For small molecules, the changes in solvation and sorption energy that occur when the solutes are transferred from solution to the stationary phase are often quite small and of similar magnitude for all compounds being separated. This means that the solutes will elute relatively close to each other in the chromatogram, enabling the use of isocratic elution. The principal impetus for such separations thus becomes efficiency, i.e. prevention of the narrowly spaced solute bands from dispersing into adjacent bands. In contrast, separations of biopolymers are most often controlled by selectivity. The separation chemistries are therefore more diverse and the process of finding suitable techniques and optimal conditions can consequently be arduous.

2.8 Stability of Biological Macromolecules

The stability of biopolymers, the need to preserve their native state during chromatographic separation, and ways of accomplishing this are important concerns for preparative and production-scale biomacromolecule separations. On the contrary, there is in analytical separations often no direct requirement to maintain the biomacromolecule in its native state, so long as the conformational changes are reproducible and manifest themselves in intermolecular interactions patterns that enable its separation from other solutes in the sample. Actually, a number of widely used analytical separation techniques such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE)[85] involve intentional denaturing of proteins in order to obtain a sample with more predictable properties.

The amphiphilic properties outlined above are responsible for the higher structural features of proteins. Whereas polar groups are exposed on the surface of a protein suspended in water to stabilize the molecule by solvation, the three-dimensional structure is maintained by intermolecular forces, partly from hydrogen bonding due to the peptide backbone, and partly from weak interaction between hydrophobic lateral chains. These interactions are usually stabilized by a relatively small number of salt bridges and covalent disulfide bridges. The structural stability originating from these concerted interactions can vary substantially – some proteins are so unstable that they can hardly be isolated from their natural environment without extensive denaturation,[86] whereas others are surprisingly robust and can resist quite extreme conditions. The most stable overall configuration is achieved at the point of balance where the internal bond strengths, molecular strain, solvation, and other intramolecular interactions yield the lowest Gibb's free energy kinetically accessible to the polypeptide chains. Anything that changes the relative strength of these forces acting upon the molecule challenges the native configuration. A system this complex is prone to respond to any action that causes an alteration in charge, solvation, or internal bonding pattern. This means that solvent activities, salt concentrations, and pH are factors that should be under close control in protein separations.

Interaction of a protein with a surface means that some of the external functional groups will be reoriented from solvent coordination onto interaction points at the support surface, which may result in a strain in the molecule. Following the initial contact, the protein will experience attractive and repulsive forces at the surface, which drive the molecule to flex and yield in order to minimize the overall energy in its new environment. During this process, new contact points may be found which alter and augment the strain. No denaturation occurs if the protein can withstand these ordeals without altering its internal bonding pattern. However, any interaction producing an appreciable retention is likely to cause changes in internal bonding patterns, in particular for very large molecules. These alterations can aggregate over time in a chain of small conformational changes that may eventually modify the molecule from its original state to such an extent that a spontaneous reorganization, rendering the macromolecule back to its original shape, is impossible upon release from the surface. If so, the protein has transformed into a denatured state, which usually has catastrophic consequences for its biological function. New conformations are also likely to display interaction patterns with the stationary phase that are different from those expressed by the natural molecule. This is exemplified in Figure 8, where four proteins are chromatographed on an interaction layer comprised of short alkyl chains.[87] In these chromatograms, the earlier-eluted peaks are the folded proteins and the later peaks are the unfolded (denatured) proteins. The peaks from the unfolded proteins increased at the expense of the folded proteins as the contact time between the proteins and the stationary phase was lengthened. Some stationary phases may exert such strong denaturing effects that proteins may fuse with their surface and cannot be eluted at all. If such denaturation is due to specific interactions with active residual sites on the interactive surface, sacrificing a readily available and inexpensive hydrophobic protein such as bovine serum albumin (BSA) can be a viable way to block the residual activity prior to using the column for separation of sensitive proteins.[88]

Much of the development in HPLC of biomacromolecules has consequently focused on finding soft interaction modes for proteins, in which the delicate interaction–deformation balance is optimal. Interactions that utilize fast and reversible binding to groups exposed at the surface of the protein are far less dangerous than interactions based on inversion of the protein structure, i.e. extensive exposure of hydrophobic groups. Likewise, the use of solvents or other mobile phase additives that

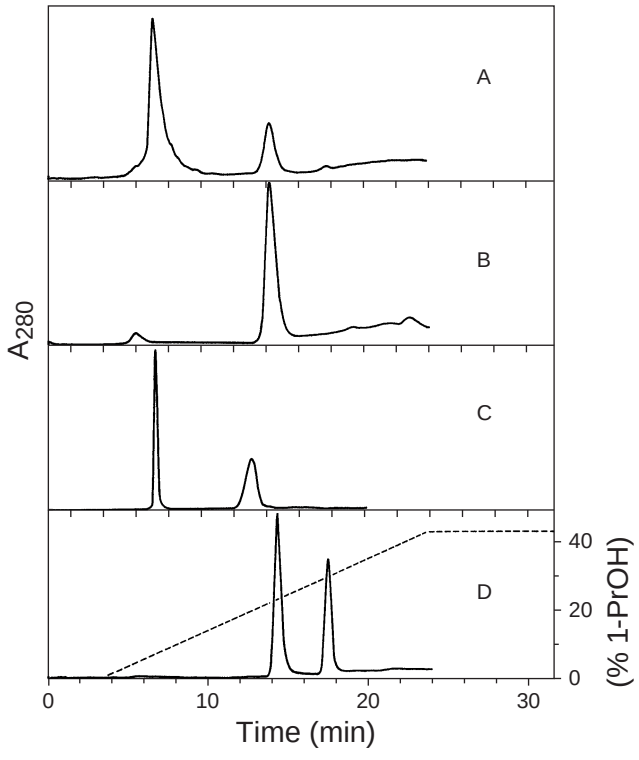

Figure 8 Reversed-phase chromatograms of (A) papain (PAPN), (B) α-chymotrypsinogen (CHTG), (C) lysozyme (LYSO), and (D) myoglobin (MYOG) on a 100 × 4.6 mm i.d. Vydac C_4 bonded phase silica (5.6 µm, 300 Å) column. Samples: 50 µg of each protein injected as 20-µL aliquots in water. Mobile phase A: 0.5% 1-propanol (1-PrOH) in 10 mM H_3PO_4, pH 2.1 (LYSO); 2% 1-propanol in 10 mM H_3PO_4, pH 2.1 (PAPN, MYOG); 2% 1-propanol in 1 mM HCl, pH 3.0 (CHTG). Mobile phase B: 45% 1-propanol in 10 mM H_3PO_4, pH 2.1 (LYSO, PAPN, MYOG); 45% 1-propanol in 1 mM HCl, pH 3.0 (CHTG). Gradient: 100% A to 100% B in 20 min; flow rate 1.0 mL min^{-1}. Column temperature 5 °C. (Reproduced from S. Lin, B.L. Karger, *J. Chromatogr.*, **499**, 94 (1990) by permission of Elsevier Science B.V.)

break the internal protein structure are also factors that will lead to excessive denaturation. Temperature and contact time with the surface are obviously important in these processes, as the majority of these changes are associated with activation energies sufficiently large to maintain the molecule in a native state in the absence of perturbations. Minimizing the separation time is therefore central for maintaining conformation, because most bioseparations are based on initial capture by highly retentive multipoint interaction processes followed by gradient elution.

3 MOBILE PHASES – ELUENT

An eluent for the separation of biomolecules should fulfill all general requirements for HPLC eluents:

- low viscosity to promote solute diffusion in the mobile phase and provide for a lower back-pressure;
- miscibility with other eluent components;
- an ability to wet or swell the stationary phase;
- compatibility with the detection mode;
- inertness relative to the solutes and the stationary phase;
- low volatility to prevent cavitation and outgassing in pumps and detectors;
- an ability to dissolve buffer salts;
- an ability to solvate the macromolecular solutes without causing undesirable denaturation or breakdown;
- low toxicity;
- a reasonable cost of both purchase and disposal.

Properties of a large number of eluent candidate solvents can be found in the literature.[55] Only a handful of organic solvents satisfy these stringent criteria. Water complies with most requirements and is also an ideal solvent from toxicological, environmental, and economic points of view. As most biomolecules discussed here exist in fully or partially aqueous phases, water is an obvious solvent choice, and in applications where proteins should be recovered in their native state, it is often the only solvent possible. Notable exceptions are membrane proteins, which need additional components for solubilization, and small peptides that require and can tolerate the use of polar organic solvents in reversed-phase high-performance liquid chromatography (RPHPLC). The most frequently used organic solvents in techniques requiring fully or partially nonaqueous eluents are acetonitrile and the lower alcohols (methanol, ethanol, and 2-propanol). In rare cases, alkanes and lower-chlorinated alkanes are also used.

In chromatography of small molecules, the polarity of an eluent and thereby its strength are usually altered by varying the mixture of organic solvents, often with water as one of the solvent components. In contrast, in the area of biomacromolecule separation, it is more common to alter the polarity of water without adding any less-polar solvents, just by using polar nonionic compounds such as urea or ethyleneglycol, or salts with cosmotropic or chaotropic ions to cause salting in and salting out of the solutes. These agents act mainly by altering the structure and activity of water, and affect the solvation and self-aggregation tendencies (collidal stability) of the biopolymers. They may also alter the retention by interacting with functional groups on the stationary phases.

4 SEPARATION COLUMN

The key to predictable high-performance separations in LC lies in a perfect control over the physicochemical

characteristics of the column and a sound insight into the processes that lead to sorption and desorption of biomacromolecules. Reproducibility in particle diameters, pore size distribution, and surface chemistry has improved tremendously over the years and particulate column packing materials are now almost exclusively made as spherical beads, ranging from micrometer-sized nonporous to gigaporous particles. Identification of porous properties as one of the key factors has led to materials with precisely controlled porosities, whereas the insight that chemical and morphological features of the interaction layers affect the reversibility in the biomacromolecule sorption has enabled new developments such as spatial location of orientated interaction layers with tailored interaction properties.

The most important decision that must be made during development of a separation method is the choice of separation mode, which directly implies the selection of a separation column. Choices have to be made with regard to support matrix, the functional groups and the morphology of the interaction layer (if any), and dimensions and material of the column. Factors that must be taken into consideration are the nature of the solutes and sample matrix, which determine the separation chemistry to be used, and the purpose of the separation, which determines the separation efficiency and sample capacity required. The sorption capacity of a separation material can be determined either as the static or as the dynamic capacity. The static capacity is measured in the batch mode, by measuring the adsorption of a protein probe such as BSA onto the material under conditions where sorption is strong. Dynamic capacity is the amount of material that the separation material will adsorb when pervaded by solution at a certain flow rate, i.e. under the conditions that the material will actually be used. There can be a substantial difference between the static and the dynamic capacities, reflecting the kinetics of the sorption process.

Other practical aspects influencing the choice of a column packing are compatibility with available detection modes and the required lifetime. When a method expected to have longevity is to be developed, it is furthermore essential that the same or at least similar column will be available over the lifetime of the method. This is particularly important in quality control, or in preparative separations such as downstream processing in biotechnology, where regulatory issues may restrict the degree of freedom in both the development and subsequent changes to the method. Industrial users are therefore adopting new techniques somewhat reluctantly, and proven and widespread principles are often preferred over new and uncommon techniques, although the latter may have better performance. More details are available in numerous monographs[89–92] and recent review articles.[93–98]

4.1 Choice of Support Matrix Material

Supports are best classified according to their morphology as soft, semirigid, or rigid gels. Soft gels (xerogels) are swollen, essentially homogeneous polymeric networks with solvent-imbibed interstices forming an isoporous system. Xerogels used in biomacromolecular separations with water as solvent are typically highly hydrophilic water-swollen networks that are termed hydrogels. In order to accommodate diffusion of biomacromolecular solutes, the cross-linking density of the material has to be low and, as a consequence, the rigidity of the network is insufficient to maintain its own structural integrity under stress. Xerogels therefore swell and shrink considerably with variations in solvent composition, salt strength, or other factors affecting the degree of internal solvation, and collapse when the solvent is removed. Collapsed xerogels are not always reswellable and permanent structural changes are often seen if the gels are allowed to dry. True xerogels are generally unable to withstand the differential pressures needed for HPLC separations.

In contrast, rigid gels (aerogels) are networks of stiff and dense material intersected by macropores into which solvents and solutes can penetrate. The rigid matrix volume of an aerogel does not change in solvents, and thus its dimensional and mechanical properties are retained in different eluents. The aerogel porosity also remains unchanged when the solvent is removed. Semirigid gels, such as partly swellable polymeric materials cross-linked through ordered crystalline segments, combine properties of both xerogels and aerogels. Although the amorphous regions swell in solvents and behave as a xerogel, the crystalline areas act as rigid aerogel-type reinforcing structures. Most synthetic polymers behave as semirigid aerogel–xerogel composites, as they swell in different solvents. The difference between xerogels and aerogels is shown schematically in Figure 9.

In interactive chromatography, the role of the support is usually to act as a carrier for the specific interaction layer, whereas in noninteractive (size exclusion) chromatography the support also constitutes the stationary phase. In either mode, only materials with well-controlled porous properties and surface chemistry can serve as high-efficiency HPLC sorbents. As outlined above, the prerequisites for efficient mass transfer are small-diameter particles, shallow interaction layers, and large enough pores for solutes to diffuse freely. In order to avoid pore crowding in interaction chromatography, a rule of thumb is that the average pore size should be 10 times the hydrodynamic diameter of the biomacromolecules under separation.[96,99]

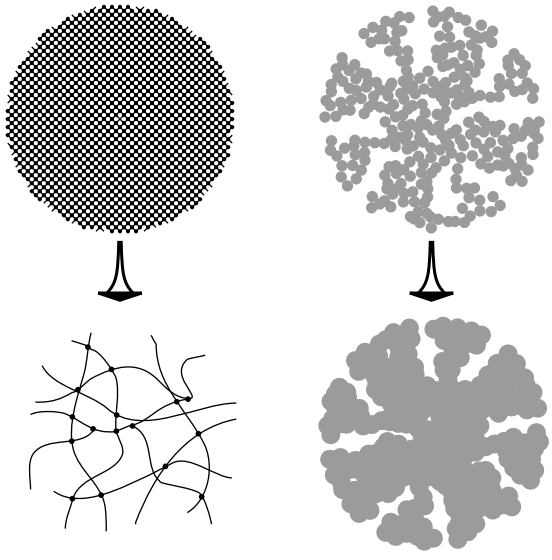

Figure 9 Diagrams of (a) soft gels (xerogels) and (b) rigid gels (aerogels). Xerogels are comprised of isoporous networks of randomly dispersed linear polymer chains connected by interchain cross-links at random intervals. The pores are sufficiently large to accommodate biopolymers. Aerogels are hard, impervious materials with high fractal dimension (i.e. resembling a cauliflower), where the particle is formed by an interconnected array of small microglobules. The voids between the microglobules are macropores that facilitate diffusion into the micropores.

The pore size also controls the capacity in interaction chromatography, as a result of the inverse relationship between the pore size and the interaction layer surface area, and because the interaction of biopolymers with sorbents is largely a monolayer process taking place on pore surfaces. This holds for relatively small biomacromolecules – when the size increases, pore diffusion becomes increasingly difficult and the interactions mainly take place on the perimeter of the particles, evident by a lowering of the overall loading capacity.[100] The porous structure of the support is also related to its mechanical stability, which is important because a column with efficient separation characteristics must also withstand a substantial differential pressure. As HPLC supports are packed in closed columns, the material should have a minimal tendency to compress and swell over time, and under changing buffer and solvent conditions.

The requirements outlined above are met by rigid materials (aerogels) with soft surfaces that do not cause denaturation and irreversible sorption of biomacromolecules. Unfortunately, the mechanical and chemical design parameters converge only rarely. The rational way to design a sorbent thus goes through preparation of a matrix with suitable physical characteristics, followed by outfitting its surface with an interactive layer whose characteristics are carefully designed for the targeted separation chemistry. These modifications frequently include steps leading to deactivation of the support, i.e. to elimination of groups that enable nonspecific interactions.

The principal support material in HPLC of small molecules is surface-modified porous silica gel. This material is also central as support material in HPLC of biomacromolecules, although organic polymers are rapidly gaining more ground as they facilitate the preparation of a wide variety of surface chemistries. Polymeric sorbents of sufficient mechanical strength can nowadays be prepared by the cross-linking of natural polymers, or by direct polymerization of functional and cross-linking monomers. Many cleverly designed composite materials have also been devised where the strongly sorptive silica surface is shielded by coating with organic polymers by shrink-wrapping or grafting procedures. The primary criterion for a successful surface interaction layer is that it should exhibit a reversible and predictable sorption, preferably obeying the Langmuir isotherm with the linear portion in the operational solute concentration range. It should also exhibit site homogeneity without residual strong interaction sites. Other important properties are chemical stability under the conditions encountered during their use, rapid column equilibration, and a long shelf-life.

4.1.1 Soft Gels Based on Biopolymers

Chromatographic bioseparation was for decades synonymous with soft gels operated at low linear flow velocities in cold rooms for hours or even days, producing moderate separations. The fundamental limitation of soft xerogels is their compressibility, which prevents the use of particles smaller than about 50 μm and limits the linear flow velocity to about 1 cm min^{-1}. The advent of HPLC with rigid particles capable of being pumped at flow velocities an order of magnitude faster than soft gels has revolutionized bioseparations. However, soft gels are still important in large-scale industrial biopolymer separations, and as interaction layers on rigid sorbents.

4.1.1.1 Cellulose Cellulose is an essentially linear polymer of 1,4-β-glucoside-linked D-glucose, whose insolubility in water and common solvents, in combination with low nonspecific interactions, makes it well suited as a stationary phase in LC, in particular in bioaffinity schemes. It is available in native microcrystalline form, or as macroporous particles produced by regeneration in organic solvents of cellulose that has been solubilized, e.g. as a xanthate. The glycoside bond is sensitive to hydrolysis but cross-linking with difunctional reagents

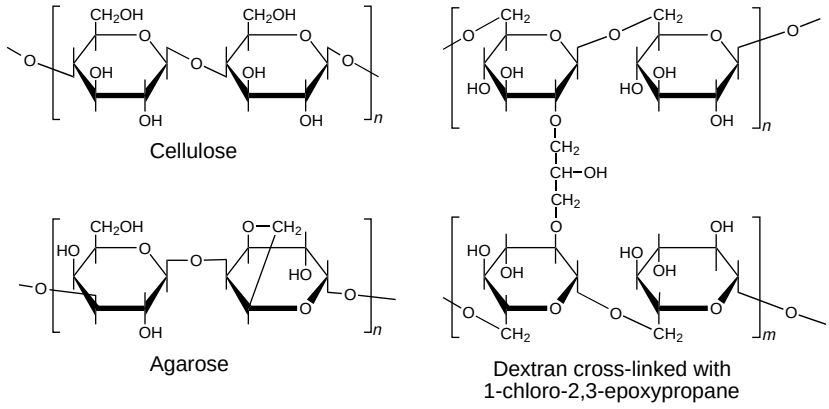

Figure 10 Structures of polysaccharides used as chromatographic supports. The dextran structure illustrates the cross-links introduced by a bifunctional reagent, in this case 1-chloro-2,3-epoxypropane.

such as 1-chloro-2,3-epoxypropane (epichlorohydrine) renders the cellulose more resistant to chemical attack and improves its mechanical properties. Underivatized and functionalized celluloses are available in the form of irregular particles, porous beads, and as membrane cartridges from a number of suppliers.

4.1.1.2 Dextran This material is a D-glucose polymer linked by 1,6-α-glucoside bonds, produced by fermentation of sucrose. The chains are branched by 1,2-, 1,3-, and 1,4-glucoside bonds, which effectively prevent crystallization. As opposed to cellulose, dextrans are soluble in water, and supports are made by cross-linking in suspension or by depositing layers of dextran cross-linked with difunctional reagents targeting the hydroxyl groups. Due to their hydrogel properties, cross-linked dextrans are soft and cannot be used under high flow rates. This prohibits their use in conventional HPLC columns. The principal use of dextran in biomacromolecule HPLC is therefore as surface layers on rigid gels.[101]

4.1.1.3 Agarose Agarose is a linear polysaccharide comprised of alternating 1,3-linked β-D-galactopyranose and 1,4-linked 3,6-anhydro-α-galactopyranose. Agarose dissolves readily in hot water, from which it solidifies on cooling below about 50 °C, a property that is used to prepare pellets or spherical agarose particles. These particles may subsequently be cross-linked with a difunctional reagent capable of forming hydrolytically stable covalent bridges between hydroxyl groups in different linear chains, thereby making the material insoluble. The cross-linking also improves the mechanical properties. Structures of agarose and a cross-linked dextran are shown in Figure 10.

On gelation, agarose forms an ordered three-dimensional structure with the polymer chains arranged to double helices that are ordered into fibrillar bundles by a combination of interchain hydrogen bonding and hydrophobic interactions. The result is a network of spatially segregated polymer aggregates, intersected by channels filled with water occupying as much as 100 times more volume than the polymer itself. These interaggregate pores provide a passage for diffusion of macromolecules and, as the pore size is inversely proportional to the agarose concentration,[102,103] exclusion limits can be conveniently varied over a wide range. The structure of agarose is visualized and compared to the structure of cross-linked dextran in Figure 11.[104]

The main drawback of agarose is its relatively low mechanical strength. Therefore agarose has traditionally been used in the format of large particles, unable to fulfill the criteria of a high-performance technique. In recent years, development of highly cross-linked agaroses has resulted in sorbents with substantially improved

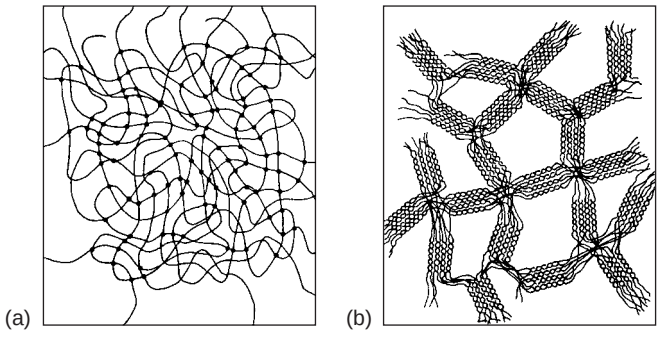

Figure 11 Representation of the cross-linked structures of (a) dextran and (b) agarose. The unordered dextran solution forms a random network with small pores and the polymer evenly distributed in the hydrogel, whereas the thermal gelation behavior of agarose leads to formation of an ordered network with the polymer arranged into fibrillar bundles and large through pores. Note that agarose gel aggregates may actually contain $10-10^4$ helices rather than the smaller numbers shown here. (Reproduced from S. Arnott et al., *J. Mol. Biol.*, **90**, 283 (1974) by permission of Academic Press.)

mechanical strength. Particles as small as 3 μm have been successfully produced and utilized for the separation of biopolymers.[105] Although these materials compress under flow, this deformation can be turned into an advantage, as controlled compression of a packed bed results in improved column efficiency due to reduced interparticle diffusion paths.[106]

An inherent and problematic property of biological macromolecules as chromatographic supports is their vulnerability to break-down by biological activity. Although some polysaccharides are subject to break-down by enzymatic action, it is particularly important to protect the supports from becoming consumed or contaminated by microbial activity. It is therefore necessary to add preservatives to eluents, and in particular to the solutions used for storing columns. A suitable preservative is 200 ppm sodium azide. This applies also to SEC (size-exclusion chromatography) stationary phases and other highly hydrophilic separation materials based on biomacromolecules.

4.1.2 Inorganic Oxides and Phosphates

The inorganic materials that have found use as HPLC stationary phases are crystalline or amorphous oxides or phosphates of metals or semimetals. Although crystalline materials possess a more predictable bulk and surface structure, they are often difficult to functionalize and their porous properties are hard to fine control. A critical parameter is also their total insolubility in the mobile phase, to prevent deterioration of separation properties and depletion of the packing material. The dissolution rate is a material property that depends on chemical factors, such as pH and presence of chelating agents and ions capable of dissolving the matrix or stripping the surface functionalities off the mobile phase. Temperature also has a substantial influence on the solubility (and thereby stability) of chromatographic matrices.

The surfaces of inorganic oxides typically used as supports carry ionic charges of varying density, which may exert Coulombic interactions with oppositely charged functional groups of solutes.[89] This material property is unpredictable and difficult to control, and is a function of the preparation procedure and the history of the sorbent, as well as mobile phase composition, in particular the pH. Shielding of the active groups located on the surface by functionalization or coating is therefore often necessary for obtaining reproducible results. The use of inorganic packing materials without surface functionalization is uncommon in bioseparation. An exception is hydroxyapatite and related materials.[107] The most common inorganic materials used as supports are listed in Table 2.

Although amorphous silicon dioxide (silica) occupies a unique position among the inorganic oxides, there are several other metal oxides and insoluble metal salts that have been used in HPLC of biomacromolecules. Alumina is a material with a long heritage in LC, whereas zirconia and titania have been introduced more recently. Common to these three oxides are their advantageous mechanical properties and their pH-stability superior to silica, which may be useful in affinity separations.[123] However, their high coordination numbers result in strong complexing abilities of their surfaces and tendencies to enter ligand exchange and ion-exchange reactions.[115,124] The Lewis acidity and basicity are generally also higher than those of silica, and titania has the highest Brönsted acidity of all oxides.[115] Lewis acidity leads to strong interactions with Lewis basic (electron donor) groups such as carboxylates and amines, unless the surface is deactivated. Among the nonsilica supports, zirconia seems to be the most promising, as it is stable to the entire pH range and amenable to facile surface functionalization.[119]

4.1.3 Porous Silica

Important factors contributing to the popularity of porous silica, in spite of the severe pH limitations, has been an excellent mechanical stability, good availability of spherical particles with well defined pore size and surface area, and the wealth of functionalization reactions that can be carried out on the surface silanols through formation of reasonably stable silyl ethers.

Silica is an amorphous hydrated silicon dioxide aerogel with the formula $SiO_2 \cdot xH_2O$. Preparation takes place through a sol-gel process where primary particles are coalesced into spherical clusters that are subsequently sintered and calcinated. This process renders a spherical micro- and mesoporous structure of high fractal dimension. After rehydration the surface contains hydroxyl groups of varying activity, in addition to siloxane bonds between adjacent silicon atoms. The surface of a hydrated silica has a weak Brönsted acidity due to dissociation of free silanol groups. In undissociated form, these silanols represent hydrogen donor groups, whereas closely spaced surface silanols associated by hydrogen bonding offer weak hydrogen-acceptor sites. These sites are depicted in Figure 12.

The relative abundance of these groups on the surface of silica is dependent on the calcination process and the degree of subsequent hydrolysis. The acidic silanols that are always present on hydrated silica surfaces dissociate to a net negative charge above the pH of the isoelectric point (p*I*), which is approximately at pH 3. The calcination process also brings out Lewis acidic and base groups onto the surface.[115] These active sites can cause nonspecific interactions with biopolymers and denaturation of proteins. The structural features of the silica supports are of critical importance in separation

Table 2 The most important insoluble inorganic materials used as HPLC supports[89]

Oxide	Application areas	Advantages	Drawbacks
Silicon dioxide (silica)[91]	Ubiquitous in HPLC as nonfunctionalized, surface functionalized, or polymer-coated spherical particles. Widely used in various biomacromolecule separation schemes.	Available in a wide variety of pore sizes, particle diameters, and continuous beds. Commercially available in a wide range of functionalizations from highly polar to highly hydrophobic.	Limited pH stability (2–8), although recently launched phases are more stable towards bases. Solubility of silica increases with temperature, limiting its use at high temperatures.
Aluminum oxide (alumina)	Useful for the separation of biomacromolecules only after a suitable coating.	Better pH stability (2–12) compared to silica.[108]	Strongly adsorptive Lewis acid sites must be coated for use with biomacromolecules.[109] Cannot be modified by silanization[110] and is therefore difficult to functionalize.[111]
Titanium dioxide (titania)[112,113,114]	Has a special affinity for organophosphates[114] and is therefore useful for the separation of nucleotides.[114]	Stable between pH 1 and 14[114,115] and amenable to functionalization by silylation. C_{18}-modified RP titania is more stable than silica.[116,117]	Strong Lewis and Brönsted acidity; strong Lewis basicity.[115] Unfunctionalized titania behaves as an amphoteric ion exchanger.[118] Limited availability of functional groups.
Zirconium dioxide (zirconia)[119]	Modified with ion-exchange groups for the separation of proteins, peptides and polynucleotides. Coated with nonpolar polymers for RPHPLC.	Stable between pH 1 and 14.[120]	Interacts with Lewis bases on solutes due to its Lewis acidity. Phosphates lower this tendency.[121] Limited availability compared to silica.
Hydroxyapatite[122]	Steric and amphiphilic ionic interactions.	Unique selectivity often capable of resolving subtly different proteins co-eluted in other modes. Nondenaturing elution conditions.	Relatively high solubility.

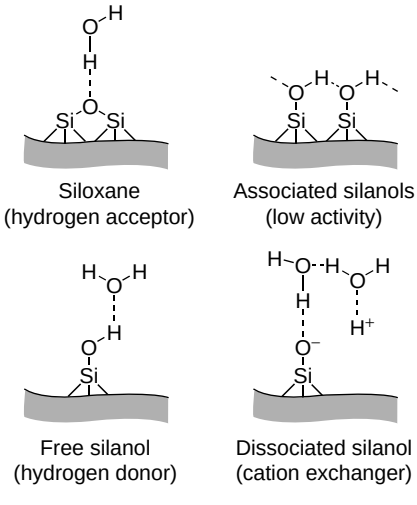

Figure 12 Representation of the sites with different activity existing on a calcinated and rehydrated silica surface.

of large molecules, and materials for separation of biomacromolecules are marketed in different nominal pore sizes up to 4000 Å.

4.1.4 Supports Based on Semirigid Synthetic Polymeric Materials

Separation media based on polymers are currently dominating size exclusion, ion-exchange, hydrophobic interaction, and affinity chromatography of biopolymers.[125] Numerous monomer systems are utilized, with the more important ones being styrene–divinylbenzene, and acrylic[126] or methacrylic acids,[127] esters or amides with hydrophilic,[94] functional,[128] or reactive[129] functionalities. Other examples of monomers used for the preparation of chromatographic packings are 4-vinylpyridine,[130] 1-vinylpyrrolidinone,[131] and vinyl acetate, the repeat units of which are partly hydrolyzed to poly(vinyl alcohol),[132] to mention a few. The main advantage of many polymer supports is their superior stability over the entire pH range.

Most synthetic carriers have a nonionic hydrocarbon backbone characterized by hydrophobic properties. However, hydrophilic lateral functionalities in some packings, such as amide in polyacrylamide or hydroxyl groups in poly(2-hydroxyethyl methacrylate), do not allow the biomacromolecules to reach the hydrophobic main chain.

In other polymers, the hydrophobicity can be mediated through surface functionalization or coating.

4.2 Polymer-coated Materials

The potential inherent in coating a rigid support with a thin polymeric layer of suitable interaction characteristics was mentioned above. Polybutadiene[133] (PBD) is often employed as a nonpolar polymeric coating on alumina,[134] zirconia,[119] and nonporous silica.[135] The cross-linked PBD chains lack labile groups and provide a hydrophobic layer which is stable over the entire aqueous pH range. It can therefore be used to take advantage of the wide operational pH range of alumina and, in particular, zirconia. For example, PBD remains unaffected even by 1 M methanolic NaOH at high temperatures,[136] a feature valuable in processes that require frequent sanitation. A problem with polymeric coatings is that the polymer layer affects the pore-size distribution of the support. For example, Reeder et al.[137] found that PBD coated on zirconia was preferentially located in the smaller pores, which also shows up in increased mass transfer resistance as the coating thickness is increased.[138] Several studies have furthermore shown that coated zirconia often exposes patches of the bare surface.[119,139] The combination of the high hydrophobicity of PBD and the strong Lewis acid properties of the uncoated surface patches yields strong interactions and irreversible adsorption of proteins.[109,134] However, these effects are mitigated if anions capable of forming strong complexes with zirconium ion, such as fluoride and phosphate, are added to the eluent. For example, phosphate-treated coated zirconia is a sorbent with both RP and cation-exchange characteristics, useful in mixed mode separation of proteins,[140,141] and zirconia treated with ethylenediamine-*N*,*N*′-tetramethylphosphonic acid acts as a cation-exchange support for the separation of proteins when used with eluents containing 2-(*N*-morpholino)ethanesulfonic acid and NaCl.[121]

Porous and nonporous silica have been noncovalently coated with homo- and copolymers of 2-hydroxyethyl methacrylate (HEMA), ethyl methacrylate, and octadecyl methacrylate, and been found useful for studies of protein unfolding in contact with sorbents in the RP mode.[142,143] Davankov et al.[144] developed a technique for coating porous silica with polystyrene for the preparation of efficient RP separation media for proteins and polypeptides. The polystyrene coatings were converted to strong cation and anion exchangers by treatment with chlorosulfonic acid and chloromethyl methyl ether followed by reaction with trimethylamine, respectively.[145] These sorbents separated a range of proteins and nucleotides by pH and salt concentration gradients. An interesting mixed-mode material was also prepared by consecutive chloromethylation, sulfonation, and amination, yielding a sorbent with both anion- and cation-exchange sites that could resolve peptides and proteins.

Coating of inorganic carriers is a rational way to establish the mechanical strength required to utilize the soft but highly biocompatible linear polysaccharides as HPLC supports. Simultaneous deactivation of the residual negative charges of silica and formation of a dextran layer in the pore system was achieved by Santarelli et al.[9] using dextran slightly modified with diethyl aminoethyl (DEAE) functionalities. Activation of these supports enabled immobilization of various ligands for HPLAC (high-performance liquid-affinity chromatography), such as protein A, concanavalin A and heparin. The use of these affinity materials for HPLAC separation by salt or pH gradient and competitive elution was shown.[146,147] Zirconia has also been coated with carboxymethylated dextran, followed by 1,4-butanediol diglycidyl ether cross-linking, as described by Dunlap and Carr.[148]

Poly(ethylene imine) (PEI) is another widely used coating useful for the preparation of packings for biopolymer separations. The polyimine is conveniently attached by electrostatic interaction with the acidic silanol groups of silica and cross-linked by reaction of the amino groups with bifunctional electrophilic reagents such as bisepoxides.[149] The PEI itself has weak anion exchange properties, but reactions targeting the amine groups can be used to afford other functionalities such as cation-exchange[150] and hydrophobic moieties.[151] This provides for an efficient way of making mixed-mode materials. Zirconia is also a suitable substrate for PEI coating.[152] Quaternized imino groups on PEI-coated zirconia has been used for the separation of oligonucleotides and oligodeoxynucleotides.[153] Separation of proteins has been demonstrated on a PEI-coated zirconia converted into a weak cation exchanger by cross-linking with poly(acrylic anhydride) or succinic anhydride similar to reactions on silica,[150] yielding carboxylic functionalities.[154]

A poly(vinylpyrrolidinone)-coated silica was prepared by Köhler[155] as a polar stationary phase for both normal and RPHPLC. Its hydrophilicity and advantageous mechanical properties are also suitable for the separation of biomolecules by size exclusion and hydrophobic interaction. Strong anion exchange coatings prepared by cross-linking and quaternization of poly(vinylimidazole) layers on underivatized or diol silica affords packings for protein separation.[156] Cation-exchange coatings based on poly(aspartic acid) have been around since the early 1980s.[157] The coating is accomplished by reacting aminopropyl silica with poly(succinimide) leading to ring opening, accompanied by a condensation to form an

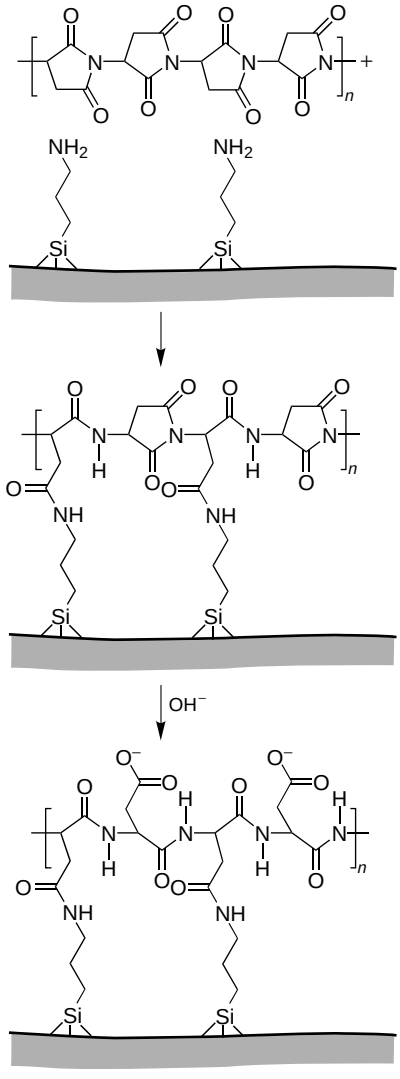

Figure 13 Reaction of poly(succinimide) with aminopropyl silica followed by hydrolysis gives a poly(aspartic acid) coating.

amide link to the silica. The remaining cyclic imides are thereafter hydrolyzed to give aspartic acid groups, according to Figure 13. The resulting coated silica is useful for weak cation-exchange separation of proteins.[158] If the initially formed layer of polysuccinimide is reacted with an amine instead of being hydrolyzed, this leads to the formation of relatively stable amides. Diethanolamine produces a hydrophilic poly(hydroxyethyl aspartamide) surface,[159] whereas reaction with alkylamines converts the polysuccinimide into weakly hydrophobic poly(alkyl aspartamides).[160] Strong cation exchangers based on poly(2-sulfoethyl aspartamide) have also been made in a similar manner.[161] Coating of porous silica with poly(methyl glutamate) was carried out by Bentrop et al.,[162] but total shielding of the silanol was not accomplished and protein separations were largely unsuccessful.

A grafting reaction has technically speaking taken place whenever a polymeric layer is covalently attached to a sorbent. A clear distinction between grafted layers and surface functionalized materials is therefore difficult to draw. To illustrate this point, the C_{18} functionality is, in a strict sense, as much grafted layer (an ethylene nonamer graft-to) as the polyethyleneimine and polysuccinimide layers mentioned above, and the Ce^{4+}-initiated tentacle chains[163] that have been brought to market in later years. It is quite clear that linear grafts can best produce the quite unique interaction layer properties characteristic of polymers, exemplified by thermally responsive separation materials (see below), but the absence of cross-linking is also accountable for the impractical chain collapse that occurs when linear alkyl phases are used in highly polar solvents.[164]

4.3 The Biocompatibility Issue

In view of the strongly sorptive properties of biopolymers, measures must be taken to avoid unintentional interaction of solutes with surfaces other than the stationary phase, in particular in trace analytical work. A consideration when a separation system is assembled is that many proteins interact with metal ions. For instance, stainless steel frits have been identified as a major source of iron contamination in HPLC eluents.[165] Oxidized metal surfaces (notably iron oxide) are usually Lewis acids, interacting strongly with electron-rich structures such as amines, which can have a negative influence on the chromatographic efficiency.[166] These interactions are significantly reduced when the column frits are replaced by stainless steel meshes, demonstrating that the porous surface of the frit cannot be ignored as a contributor to the overall retention characteristics of the column. Up to a third of the protein injected on an HPLC column in microgram quantities can be trapped on the frit[167] and the loss mechanism varies for different proteins.[168] Trumbore et al.[169] investigated the adsorption of proteins on stainless steel capillaries by scanning electron microscopy (SEM) and demonstrated that peak shapes were degraded, and the chromatographic behavior became erratic on repeated injections of biological samples. They concluded that proteins were adsorbed on the wall of the capillary and altered the surface flow dynamics. The anomalies were most pronounced at low ionic strengths and mobile phase flow rates, and removal of the deposited protein was found to be difficult. These finding are contradicted by successful use of conventional stainless steel HPLC equipment for purification of several enzymes.[170,171]

Uncertainty notwithstanding, most instrument manufacturers have moved away from stainless steel as a material for construction of bioseparation systems.

Titanium was used as an alternative for a time, but seems to be largely abandoned. A report stated that the metal was not of sufficient purity or inert enough to be used in biochromatography and recommended a completely metal-free system to be used instead.[172] If a metal system is used, then special concern should be given to groups with pronounced capabilities of binding to transition metal ions, such as phosphate. Provided mass spectrometry (MS) is not used in the detection step, HPLC methods for nucleotide analysis should therefore include phosphate as part of the buffer composition to saturate the binding sites on the metal surfaces.

Among the new materials which have been introduced as alternatives to stainless steel and a gained wide acceptance is poly(ether-ether-ketone) (PEEK), an aromatic polymer that owes its mechanical strength and inertness to its high crystallinity. Judging from its use as a material for capillary electrophoresis (CE) separation of peptides,[173] and its superiority over fused silica as transfer lines in MS,[174] PEEK features a low adsorption power for proteins. Extraction of PEEK parts may be necessary in order to remove metal impurities.[165] However, even PEEK may exhibit problems with adsorption, as observed for instance during loop injection of nonsteroid anti-inflammatory drugs.[175] In our experience, the material also suffers from fatigue after prolonged use with eluents containing high concentrations of acetonitrile, eventually leading to bursts. This may have prompted the recent introduction of gold-coated columns as a way of fighting protein adsorption.[176]

Whether biocompatible or not, most HPLC columns tend to adsorb small quantities of proteins in an irreversible and/or nonspecific way, and it is often recommended that an inexpensive protein is injected a few times until the integral of the detector signal remains constant before dilute or precious samples are run on a new column.

5 SEPARATION CHEMISTRIES APPLICABLE TO HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY OF BIOMOLECULES

The range of mechanisms applicable to the separation of biological macromolecules is vast and a wide range of separation chemistries has been developed exploiting fundamental properties of molecules and separation materials. Examples are hydrodynamic radius versus pore size, hydrophilic and hydrophobic properties, ionic attraction and repulsion, and specific interactions of solutes with immobilized functional entities of synthetic or biological origin. Chemistries are chosen in the context of the overall purpose of the separation work, and the isolation of biological macromolecules from crude biological feedstock does almost invariably require several cleaning steps for work-up and final polishing, regardless of whether the purpose of the overall separation scheme is analytical or preparative. As many biological macromolecules are sensitive to denaturation at phase boundaries and on contact with xenobiotic surfaces, the separation process itself may induce serious breakdown of fragile biomolecules. Loss of substance and/or biological activity can thus occur in each separation step, and also during handling of the solutions. Therefore, the overall separation process should be optimized to involve a minimum number of steps required to achieve the desired separation. Separation techniques are therefore preferably chosen so that fractions collected from early separations can be applied to the next separation column without excessive sample handling. A key to success is orthogonal selectivities, i.e. careful selection of a set of techniques where the mechanisms exploited in each step are complementary, so that the separation can be maximized in the minimum number of separations.

Obviously, high-performance techniques make this goal easier to achieve because the efficiency in each step is higher compared to open column separations. The high efficiency enables minimization of the cycle time in each separation and a considerable increase in the total throughput. Rapid separation has a positive effect on the yield and purity of the substances separated, as denaturation due to contact with sorbents and other surfaces in the separation equipment is time-dependent.

5.1 Size-exclusion Chromatography

SEC is the common name for the liquid chromatographic modes where molecules are separated on the basis of steric exclusion from the dimensional parameters of a well-defined gel or a pore network. The stationary phases are commonly referred to as gels, but bimodal porous materials are also often used in SEC. Separation is achieved as molecules of different sizes are either being allowed to penetrate or are excluded from the pores within the packing, depending on their hydrodynamic radius. Large molecules that cannot enter the pores are eluted in the interstitial void volume of the column V_0, whereas small molecules capable of freely diffusing into the solvent imbibed by the pore system are retarded and elute when the entire solvent volume of the column V_t has passed. The actual stationary phase is in this specific case the stagnant mobile phase.

Molecules of intermediate size are partly excluded and are fractionated based on their relative size (actually hydrodynamic radii) and eluted between V_0 and V_t in

order of decreasing molecular weight (Figure 14). SEC is the only separation mode in LC (except slalom chromatography, see below) where enthalpic contributions are intended to be absent from the retention mechanism. It is consequently a process driven by differences in entropy, and is characterized as a noninteractive chromatographic mode.

Under equilibrium conditions, the partition coefficient K_D, which is independent of column dimension, is defined for a molecule eluting at a specific volume V_e between V_0 and V_t by Equation (21):

$$K_D = \frac{V_e - V_0}{V_t - V_0} \tag{21}$$

This partition coefficient will always have a value between 0 and 1, and is dependent on the pore-size distribution of the stationary phase, and on the size and shape of the macromolecular solute.

As opposed to many other macromolecules, water-soluble proteins in their native state are folded into compact, globular entities, enabling the use of aqueous SEC for estimating their molecular size if SEC columns are calibrated by plotting log(molecular weight) of a set of protein standards against K_D.[177] What must be realized is that this procedure does not measure the molecular weight, but rather the hydrodynamic size of the proteins, as demonstrated by Figure 15. The graph shown is taken from a study by le Maire et al.,[178] where a number of water-soluble globular proteins and detergent-soluble membrane proteins of determined Stokes radii were used to establish SEC calibration curves. Several of the proteins in Figure 15 fall outside the calibration curve based on molecular weight, and the two groups of protein tested behaved differently and should therefore be calibrated with different sets of standards.[178]

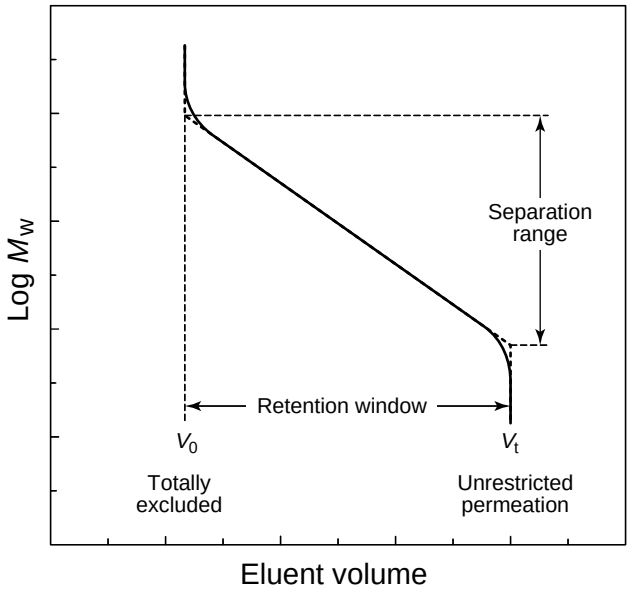

Figure 14 An idealized calibration curve in SEC. Molecules large enough to be completely excluded from the pore system are prohibited from entering the stationary phase, and are eluted in the interstitial void volume of the column V_0, whereas small molecules with access to the entire interior of the pores are eluted in the total column volume V_t. Molecules of intermediate hydrodynamic size are fractionated based on their relative diameter with larger molecules eluting ahead of smaller ones. In the absence of partitioning, all solutes must be eluted in this retention window.

The anomalies in Figure 15 are caused by deviation from spherical shape. The size of a protein can be determined by analytical ultracentrifugation[179] or by measurement of its diffusion coefficient D_p, and is expressed by the Stokes radius R_S, which is the size of spherical particle that would have the same frictional coefficient f_p as the protein:

$$f_p = \frac{RT}{ND_p} = 6\pi\eta_0 R_S \tag{22}$$

Using Equation (22), R_S can be calculated from f_p when the viscosity of the solvent η_0 is known.

However, because proteins deviate from sphericity and carry a certain amount of attached and encapsulated water, their actual frictional coefficients or Stokes radii are usually larger than would be the size of a perfect globule, R_{min}. Typical values of R_S/R_{min} for globular proteins are of the order of 1.2 but, due to asymmetrical shape, the actual size along the longest dimension can be substantially larger. Of the proteins deviating from the calibration curve in Figure 15(a), BSA has an R_S/R_{min} of 1.35, whereas the value for tyrosyl-tRNA synthase is 1.48. The anomalous retention behavior of fibrinogen is explained by its elongated, fibrillar shape, which prevents it from being characterized by a simple size parameter. An alternative for such molecules is a universal calibration based on intrinsic viscosimetric radii obtained from measurements of mass M and intrinsic viscosity η[180] (Equation 23):

$$R_\eta = \left(\frac{3[\eta]M}{10\pi N_a}\right)^{1/3} \tag{23}$$

where N_a is Avogadro's number. A detailed account of universal calibration for SEC[180–186] is beyond the scope of this article. The conclusion here is that aqueous SEC is capable of measuring the hydrodynamic size of biopolymers, but that this size cannot be directly translated into molecular weight.

To ascertain a predictable separation based on molecular size, i.e. to ensure that sieving is the sole mechanism governing the separation, intermolecular forces acting between the solute and the mobile phase have to be

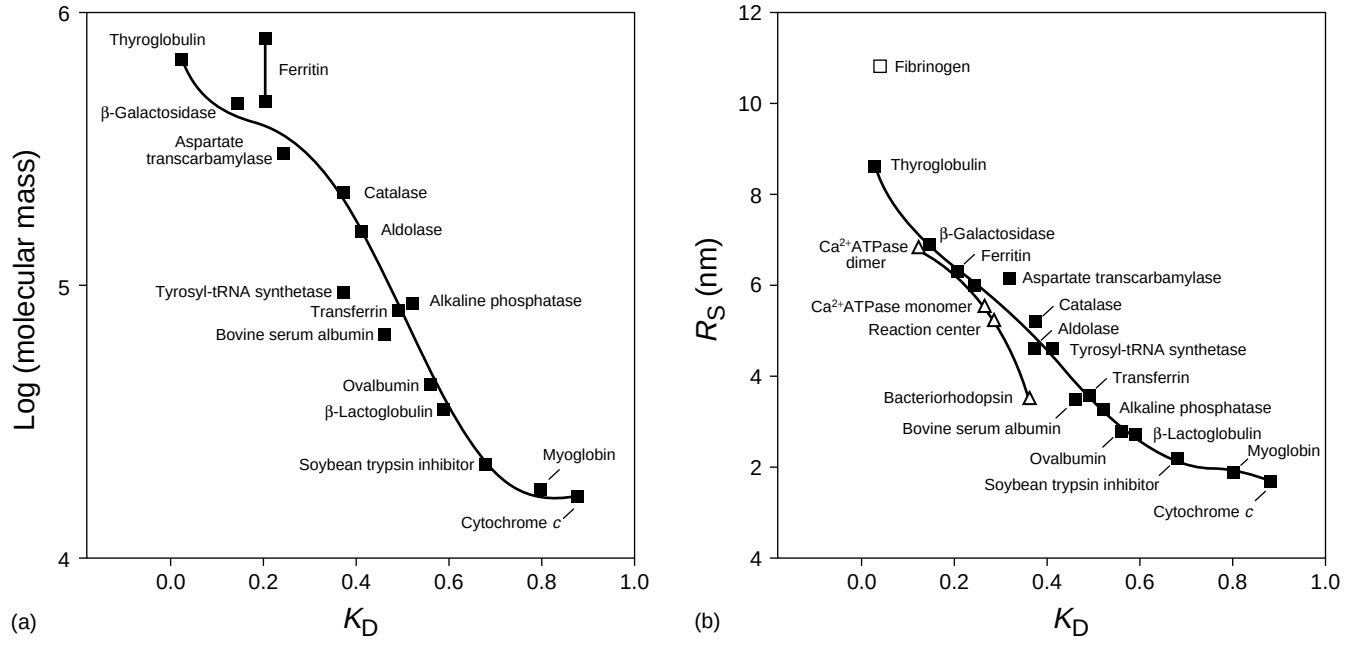

Figure 15 Calibration curves for some water-soluble proteins and solubilized membrane proteins on a TSK 3000SW SEC column based on (a) log(MW) and fricitional coefficient based on (b) Stokes radii as size parameters. The improved fit by calibrating against fricitional coefficient-based Stokes radius is due to deviation from sphericity. Fibrinogen has a very asymmetric shape and was excluded from the calibration. Bacteriorhodopsin, which also deviates from the calibration curve, binds a large amount of solubilizing detergent. (Reproduced from M. le Maire et al., *Anal. Biochem.*, **154**, 525–535 (1986) by permission of Academic Press.)

substantially stronger than the interactions between the solute and the stationary phase. This means that all partitioning reaction equilibrium constants should be as close to zero as possible. All solutes then elute in a defined retention range, the retention window. This window extends from the interstitial void volume of the packed bed (V_0; totally excluded large molecules) to the total solvent volume in the column (V_t; small solutes capable of penetrating all the pores; Figure 15). A peculiarity of SEC is consequently that all solutes have negative or zero capacity factors k' if the void volume is measured by conventional methods.

5.1.1 Stationary Phases for Size-exclusion Chromatography

When SEC is applied to the separation of biomacromolecules, aqueous buffers are utilized as mobile phases. Aqueous SEC is sometimes referred to as gel filtration chromatography, in particular by biochemists, but this term should be reserved for the low-pressure open-column variant. Columns for high-performance aqueous SEC, based on native or modified silica, hydrophilic synthetic polymers with well defined pore size, or cross-linked polysaccharides, are available from numerous manufacturers. A separation medium having pores of a specific size distribution is chosen to cover the molecular weight range required for the actual separation.

As SEC separations are based on size, it is an inherently generic concept for separation of macromolecular compounds. However, it suffers from rather low separation efficiency, resulting from the narrow separation window which dictates that all solutes should (ideally) elute within a very limited retention time-frame. In preparative SEC separations the limited resolution is dealt with by selecting a gel with suitable molecular size separation range, so that the solutes of interest fall within the range of maximum separation efficiency (typically around $K_D = 0.5$), whereas both larger and smaller molecules are eluted ahead of (close to V_0) or after the desired molecular size range (near V_t).

Although modern SEC stationary phases are designed to minimize partition interactions and serve only as a gateway allowing or denying access to the liquid phase trapped in its internal network,[187,188] most phases still have a low residual negative charge due to carboxylic groups or dissociated silica. These residual charges give rise to unintentional partition interactions,[189] and have in fact even enabled comparisons of ion-exchange interaction models to be carried out on SEC stationary phases.[189] Although ionic interactions are generally unwanted, they can be utilized for mixed-mode SEC separations involving an ion-exchange component.[190]

When silica is used as a separation material in SEC, the charges from dissociated silanols should be shielded

from interacting with the solutes. One way to compensate for the negative charge of silica is to use a shielding layer with a low positive charge.[9] Furthermore, although highly hydrophilic materials are typically used, the risk of weak hydrophobic interactions cannot be ruled out in fully aqueous media.

5.1.2 Applications and Practical Considerations

Gradient elution is not applicable to SEC, because the partition interaction variations that gradients are meant to develop would obviate the separation principle. As isocratic separations are dependent on the column efficiency, high-performance size-exclusion chromatography (HPSEC) columns are typically long and narrow and used at quite low linear flow rates, in contrast to columns used in interaction-based separation techniques. Selection of column and operating conditions is dependent on the application and the resolution required. The column packing should be selected so that all solutes of interest fall within the molecular weight fractionation range of the gel, as specified by the manufacturer. Separation of molecules with large differences in molecular size can thereby only be obtained if the separation material has a relatively wide span in pore sizes. However, this means that the separation power in the retention window becomes relatively poor and when the span in solute molecular weights is large, which often applies to carbohydrate polymers, a preferred alternative is to use several columns with different and more narrow molecular weight separation ranges in series.[191,192] For instance, four connected HPSEC columns have been used to characterize starch-based carbohydrates.[193] Also, three series-coupled columns were used in combination with differential refractive index (RI) and viscometry to separate and determine the shapes and aggregation states of complex plant cell-wall polysaccharides,[194] of amylopectin and amylose in microwave hydrolyzed starch,[195] and together with multiple angle laser light scattering (MALLS) detection for characterizing and obtaining information on the solution behavior of carbohydrate biological response modifiers[196] (Figure 16).

Coupled-column HPSEC is also used for studying protein aggregation, for instance, to determine toxic high-molecular-weight components present in therapeutic-grade glutaraldehyde-cross-linked polyhemoglobin,[197] and aggregates formed in heat-treated bovine β-lactoglobulin.[198] More commonplace separations are carried out on single columns, such as molecular weight determination of proteins, nucleic acid, plasmids, and less complex polysaccharides. Columns with a selectivity in the lower end of the molecular weight scale have also recently been marketed, providing good resolution of peptides. Among the more interesting of current developments are investigations of temperature-responsive

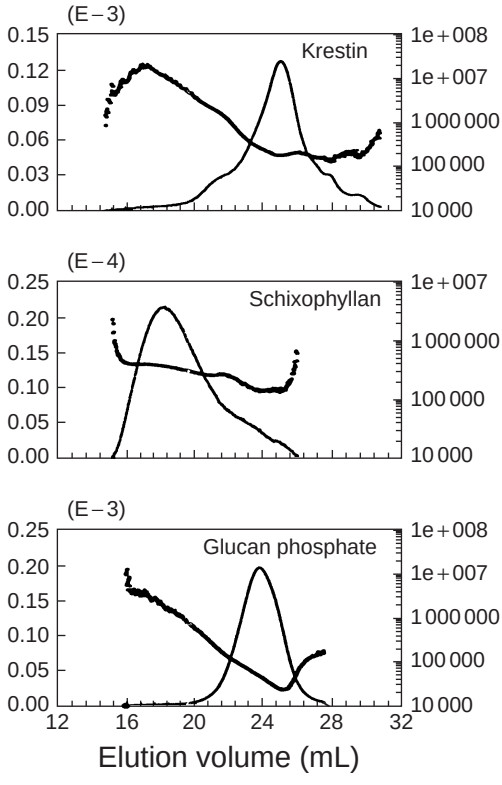

Figure 16 Molecular mass versus elution volume and concentration versus elution volume chromatograms for carbohydrate biological response modifiers, as established by SEC with MALLS and differential viscosity detection. Schizophyllan and glucan phosphate were analyzed by argon ion (488 nm) MALLS. Krestin was analyzed by helium–neon (633 nm) MALLS because of significant fluorescence at 488 nm. The Gaussian curves represent the concentration chromatograms, whereas the linear relationships represent the molecular mass versus elution volume chromatograms. (Reproduced from A. Müller et al., *J. Chromatogr. B*, **666**, 286 (1995) by permission of Elsevier Science B.V.)

materials as SEC stationary phases.[199] Apart from its use in preparative and analytical separations, aqueous SEC is also frequently used as a unit operation to accomplish desalting and buffer exchange in intermediate sample preparation steps.

The choice of eluent is relatively open, as long as sufficient salt is present to cancel ionic interactions, and the salt concentration does not exceed the level where proteins start to precipitate, or hydrophobic interactions (see below) come into play. Sodium chloride is the most commonly used salt at 0.1–0.3 M concentration, with pH being controlled by a suitable buffer with solution-phase interactions taken into account. The solute stability rather than the separation chemistry thus governs the mobile phase composition, which adds to the notion of SEC as being a very mild technique.

Sample loading on SEC columns requires some consideration. As in all elution HPLC techniques, the sample is applied in a finite volume on the column head. However, in contrast to most other biomacromolecule separation chemistries, which can tolerate relatively large sample volumes as long as they are loaded from solutions with weak elution power due to the stop-and-go control in gradient elution, there is generally no weak eluent at hand in SEC to accomplish a band sharpening at the column head. The sample loading capacity of SEC is inherently high because the total volume of the gel is used in the separation process but, in order to minimize band broadening, it is critical to keep the sample size low, or to use a sample preconcentration technique of some sort such as the on-line isotachophoretic technique described by Öfverstedt and Eriksson.[200,201] The sample should furthermore have a viscosity that is similar to that of the eluent.

Separation along basically the same selectivity dimension as SEC can be accomplished in SDS/PAGE with better resolution, wider molecular size window, and more degrees of freedom in the selection of operational parameters. The SDS/PAGE[85] technique is also used regularly for checking fractions eluted in SEC. The advantage of SEC over its electrophoretic counterpart lies in its soft and nondenaturing conditions, enabling facile collection of fractionated proteins in the native state for further separation or processing, as well as for fundamental studies of biopolymer behavior in solution.

Although nearly all separations of biological macromolecules in SEC take place with aqueous eluents, there is one exception worth mentioning. Linear polysaccharides, notably cellulose, can be difficult to dissolve. Few solvent systems thus exist which can be used for separation of these natural polymers, important in the pulp and paper industry. A solvent system that has proved capable of dissolving cellulosic polysaccharides and is possible to use as eluent in nonaqueous SEC is dimethylacetamide/LiCl.[202]

Summing up, the genuine value of SEC as an HPLC technique lies in its separation dimension, which is often complementary to other separation chemistries. In spite of its comparatively low resolution, SEC is therefore often employed in combination with other, more powerful HPLC techniques, where SEC usually is applied as the first of several chromatographic steps. In general, fractions from aqueous SEC can be applied directly to RP separations, or to IEC or HIC (see below) after dilution to lower the salt strength, or after addition of salt, respectively. Several monographs have been published over the last years dealing entirely or mainly with SEC.[203–205] The book edited by Wu[206] contains information on commercial separation materials, both from the manufacturer and user viewpoints.

5.2 Normal-phase Liquid Chromatography

LC on polar supports using less-polar solvents as eluents was the original mode used in ancient column chromatography by Tswett,[1] hence the term ''normal phase'' (NP), sometimes also ''straight phase''. Today, normal-phase high-performance liquid chromatography (NPHPLC) does not live up to its name and is among the less frequently used separation techniques in analytical chromatography. Because of the nonpolar solvents, and the high and often unpredictable retention of traditional inorganic oxide stationary phases towards polar functional groups, NP chromatography on conventional inorganic materials has been of limited use for biomacromolecules. The exception is peptides, for which several workers have demonstrated its utility.

Lerro et al.[207] used an aminopropylsilica column with an isopropylamine-modified chloroform–methanol gradient eluent to separate sticky amphiphilic synthetic peptides that were unresolvable by other methods. The peptides were separated on the basis of chain length, with recoveries better than 90%. Yoshida[208] recently demonstrated that columns with amide functionality (carbamoyl silica) are useful for peptide separation in NP mode with a gradient of aqueous acetonitrile containing 0.1% trifluoroacetic acid (TFA). These eluent conditions resemble those used for RPHPLC of peptides, but with a reversed gradient and a distinctly polar separation material. Hydrophilic peptides that were unretained in the RP mode could thus be retained and separated by NPHPLC, and the selectivity difference in the NP and RP modes made it possible to carry out two-dimensional separations of a protein digest (Figure 17). The retention coefficients established by linear multiple regression analysis of retention times of a large number of peptides can be used to predict retention times of peptides of known sequence.[209] A linear relationship has also been established between $\log k'$ and the logarithm of the volume fraction of mobile phase modifier, so that both isocratic and gradient retention times can be predicted after two initial gradient runs.[210]

Separation of very short peptides in NPHPLC has also been demonstrated recently by Armstrong et al. using silica gels modified with ristocetin A[211] or teicoplanin[212] as chiral stationary phases (CSPs).

5.2.1 Hydrophilic Interactions Chromatography

NP chromatography has been revived in hydrophilic interaction chromatography (HILIC) which shares much the same separation principle, i.e. a stationary phase that is more polar than the mobile phase. The acronym HILIC was coined by Alpert[159] and the main difference compared to NPHPLC is the use of more polar and thus stronger eluents, and of columns specifically developed

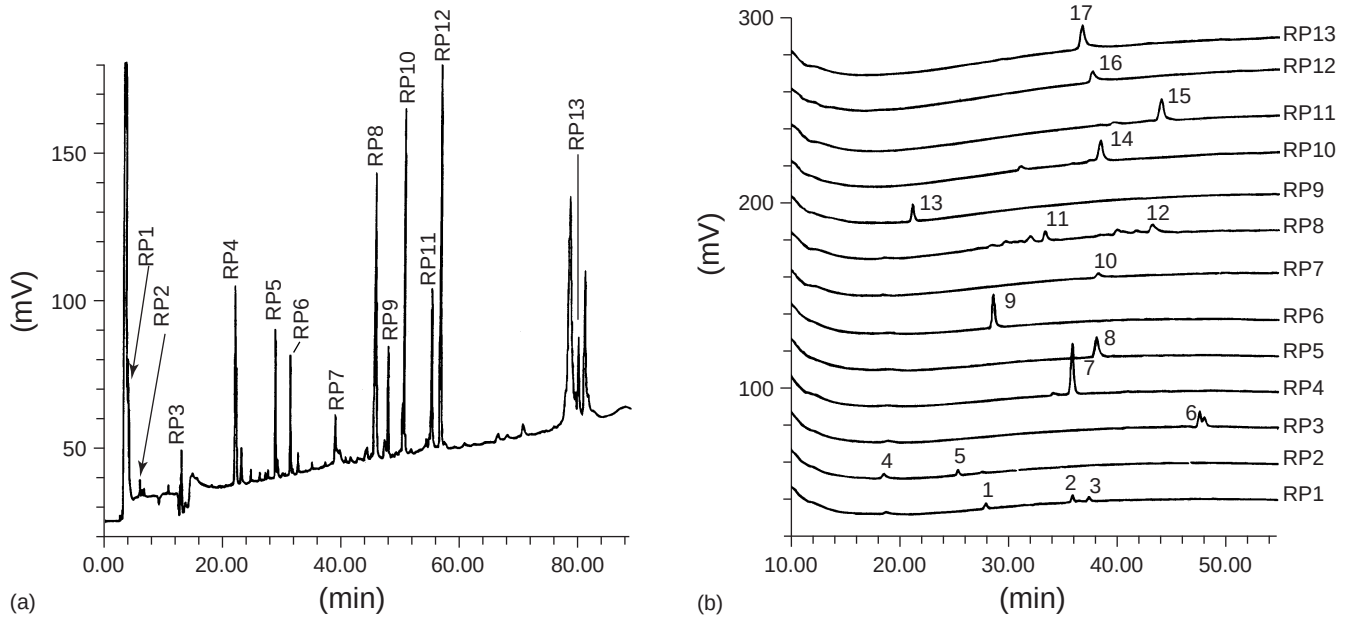

Figure 17 Complementary selectivities of RPHPLC and NPHPLC in peptide separation. (a) Chromatogram of a tryptic digest of concanavalin A on a TSK gel ODS-80Ts RP column. The digest was separated with a 90-min linear gradient of 5–55% acetonitrile in 0.1% TFA. Thirteen fractions were collected from this separation and subjected to NP separations producing the chromatograms shown in (b), eluted by 70-min linear reversed gradients of acetonitrile from 97% to 55% in 0.1% TFA. Both columns were 250 × 4.6 mm i.d. Peak identification: 1 = TAK, 2 = DQK, 3 = SK, 4 = SVR, 5 = VSSNGSPEGSSVGR, 6 = VSSNGSPEGSSVGR, 7 = SNSTHQTDALHPMFNQFSK, 8 = ETNTILSWSFTSK, 9 = VGTAHIIYNSVDK, 10 = VGLSASTGKYK, 11 = DILILQGDATTGTDGNLGLTR, 12 = LLGLFPDAN, 13 = ADTIVAVQLDTYDNTDIGDPSTPHIGIDIK, 14 = SPDSHPADGIAFFISNIDSSIPSGSTGR, 15 = LSAVVSYPNADATSVSYDVDLNDVLPEWVR, 16 = SAVVSYPNADATSVSYDVDLNDVLPEWVR, 17 = ALFYAPVHIWESSATVSAFEATFAFLIK. (Reproduced from T. Yoshida, *Anal. Chem.*, **69**, 3041 (1997) by permission of the American Chemical Society.)

for biomacromolecule separation. The packing originally devised by Alpert was a poly(succinimide)-coated silica modified by diethanolamine. More recently, materials with other functional groups have also been tested for HILIC, including a carbamoylated polymeric phase,[213] aminopropyl-derivatized silica,[214,215] and zirconia.[216]

The preferred organic solvent is acetonitrile, typically at 50% or higher admixture in water, but alcohols such as 2-propanol can also be used. However, as the alcohols are rather polar, the concentration required will be higher, leading to more viscous eluents. Addition of salts to the eluent is not required for nonelectrolyte solutes, but salt should be added to reduce residual electrostatic interactions with charged solutes. Gradient elution is achieved by decreasing the concentration of the organic solvent, thus increasing the concentration of the strong eluent component which is water. An alternative way of forming the gradient is to maintain the organic solvent admixture and increase the salt concentration. For highly hydrophilic solutes, a combination gradient with decreasing organic modifier content and increasing salt concentration can be advantageous. For peptides, the retention is a function of the hydrophilicity of the lateral groups, whereby basic moieties contribute most to the retention. The selectivity is thus opposite to that seen in RPHPLC, with Asn, Ser, and Gly being strong promoters of retention, whereas Phe and Leu provide little retention. The eluent pH can be used to control the selectivity, because at low pH only basic and phosphorylated groups will be charged, whereas both acidic and basic residues will be charged at a pH close to neutral. As basic groups are strong promoters of retention, a low pH eluent will elute peptides essentially according to an increasing number of basic residues. But as opposed to cation-exchange chromatography, a highly hydrophilic peptide can be more retained than a hydrophobic peptide with more basic residues. The selectivity of HILIC is therefore complementary to cation exchange. As the organic solvent is the weaker eluent component, samples should be prepared in a solution containing at least the same concentration of organic solvent as the eluent to obtain a band sharpening at the column inlet.[217] A separation using a polymeric HILIC column is shown in Figure 18.

The HILIC technique is particularly useful for the separations of biological macromolecules with pronounced capabilities of expressing hydrophilic interactions that are stable to the mostly organic eluents used. Compounds

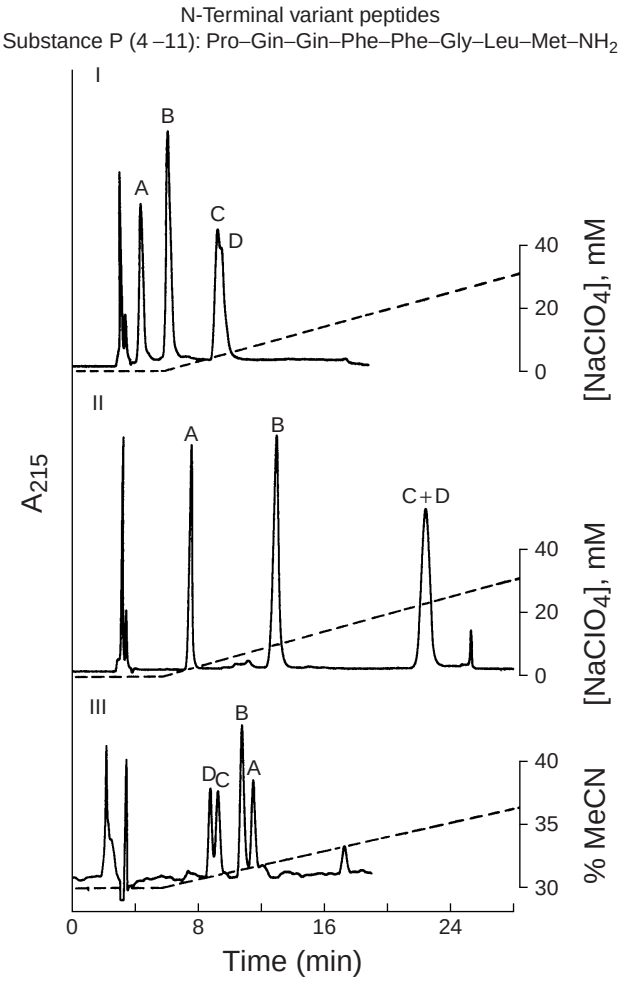

Figure 18 Gradient elution of substance P fragments from (I) a polyhydroxyethyl A column (HILIC mode); (II) a polysulfoethyl A column (HILIC mode); (III) a C_{18} column (RPC mode). Columns: 200 × 4.6 mm i.d.; flow rate, 0.8 mL min^{-1}; UV (ultraviolet) detection, 215 nm. Elution: (I) and (II) 90-min linear gradient, 0–100% eluent B; eluent A, 10 mM TEA-methylphosphonate (pH 3.0) in 80% MeCN; eluent B, 125 mM $NaClO_4$ in eluent A. (III) 90-min linear gradient, 0–100% eluent B; eluent A, 0.1% TFA in 30% MeCN; eluent B, 0.1% TFA in 55% MeCN. Peaks: Substance P fragments: A, (7–11); B, (6–11); C, (4–11); D, (5–11). MeCN = acetonitrile. (Reproduced from A.J. Alpert, *J. Chromatogr.*, **499**, 186 (1990) by permission of Elsevier Science B.V.)

separated according to the HILIC principle include nucleotides,[159,216] saponins,[214] complex carbohydrates and oligosaccharides,[159,216,218–221] peptides,[159,222–224] and basic proteins, notably histones.[159,213,225,226]

5.3 Separation Modes Based on Hydrophobic Interactions

RPHPLC is the most common separation technique for small organic compounds. The separation is characterized by the use of hydrophobic stationary phases in combination with relatively polar mobile phases, typically comprising mixtures of water-miscible organic solvents and water. A common misconception is that transfer of solutes between the mobile phase and the interaction layer is driven by interactive forces occurring mainly in the hydrophobic stationary phase. In reality, however, the retention process is largely driven by solvophobic forces, where the transfer of a solute to the stationary phase is effected by solvophobic expulsion from its cavity in the mobile phase, accompanied by displacement of eluent components from the stationary phase and release of solvent from the solute. The reverse process involves creation of a cavity in the solvent, solvation of the solute, and reoccupation of stationary phase surface by components of the eluent.[227] The selectivity is easily modulated by solute dissociation or ion-pair formation, widening the scope of this versatile stationary phase–solvent system. In addition to the eluent, altering and fine tuning of the selectivity can be achieved through the choice of a specific column from a wide variety of commercially available stationary phases with alkyl chains of different lengths, and by variation of the ligand type, density, and structure. The broad range of interactions makes RPHPLC suitable for the separation of drugs and other xenobiotics, as well as low-molecular-weight compounds of biological origin.

5.3.1 Reversed-phase High-performance Liquid Chromatography of Biological Macromolecules

The ubiquitous RPHPLC stationary phase is bonded-phase silica, due to the advantageous mechanical properties and solvent resistance offered by the silica matrix, and to the variety of well characterized chemistries available. The predominant functional groups are linear alkyl chains, in particular C_{18} (*n*-octadecyl), which provides a surface with relatively strong hydrophobicity and high capacity, suitable for interaction with small molecules. Less hydrophobic ligands such as octyl (C_8), butyl (C_4), phenyl, and cyanopropyl are also used, as are polymer-coated columns.[228] The nature of the functional groups plays an important role in the retention, and the factors that appear to affect the interactive process are hydrophobicity, flexibility, and the degree to which surface silanols are exposed in the contact surface.[229,230]

The versatility of RPHPLC is further derived from the large span in eluting powers offered by solvents such as alcohols, acetonitrile, and some ethers. These solvents can be used in combination with water to vary the retention over a very wide range, even with the same stationary phase, enabling excellent resolution of similar and structurally disparate substances. The preferred solvent is often acetonitrile because of the low viscosity and good UV transparency of its mixtures

with water. As the fraction of water in the eluents is relatively large, it is also possible to use conventional buffers for controlling retention. The retention can be controlled by the eluent pH to affect the degree of solute dissociation or protonation (ion suppression), or by additives such as short-chain ionic detergents, to promote the formation of hydrophobic ion pairs, which can be retained on the column.[231,232] These eluent manipulations enable selective retention adjustment for solutes with dissociation equilibria within and beyond the somewhat limited operational pH range of bonded-phase silica columns.

Although RPHPLC has been used for isolation and analysis of proteins,[99] it is rarely possible to recover the substances in their native or active state due to their interactions with the highly hydrophobic surfaces of conventional RPHPLC materials, and the denaturation power of the eluents. The retention process can also result in substantial loss of proteins that become tightly trapped on the separation material, apparently via hydrophobic domains existing in continuous regions on the surfaces of the protein.[233] The recovery of proteins depends largely on chain length of the hydrophobic ligand, with shorter ligands usually affording better yields and conformational integrity. Butyl and cyanopropyl silica are therefore preferred over longer alkyl chains and phenyl groups. The main application of RPHPLC in biopolymer separation is for natural, enzymatically derived, and synthetic peptides.[99,234–239] It has been found empirically that the retention factor of peptides eluted under linear sorption conditions can be expressed as a function of the mole fraction φ of the organic eluent component according to Equation (24),

$$\log k' = \log k'_{0,\mathrm{RP}} - S\varphi \qquad (24)$$

where S is the slope of the plot $\log k'$ versus φ, and $\log k'_0$ is the intercept, i.e. the extrapolated k' for hypothetical separation with no organic modifier added.[99] The dependency of k' on φ is affected by temperature and eluent pH and may have different shapes (see Figure 46). As a result of multiple contact point interactions, the values of S and $\log k'_{0,\mathrm{RP}}$ are usually larger for biomacromolecules than for small molecules, requiring the use of gradients to elute different biomacromolecules within a single run.

Separation in RPHPLC results from differences in the intrinsic hydrophobicity of the solutes, the eluotropic properties of the mobile phase, and the hydrophobic properties of the packing. The kinetics and strength of the interaction are determined by the number and type of exposed groups of the peptide and their spatial arrangement, as proteinaceous materials are known to interact with surfaces in an orientation-specific manner. Using cytochrome *c* and bovine growth hormone as model proteins, it was recently verified that the contact areas are localized on the protein surfaces and that the proteins retain their three-dimensional shape when attached to an RP sorbent.[233] For a sizable molecule such as a protein, the contact functionalities are thus usually those exposed at the protein surface, whereas most of the amino acid residues can take part in the retentive process that involves small peptides. Therefore retention of peptides that have not assumed a secondary structure can be predicted with reasonable accuracy.[240]

Acidic mobile phases are typically used for the separation of proteins and peptides in RPHPLC. This ensures that the surface silanols of the stationary phase are protonated, thus minimizing undesired ionic interactions with basic amino acid residues. The pH value is adjusted by adding about 0.1% of an acid such as TFA, heptafluorobutyric acid (HFBA), H_3PO_4, $HClO_4$, or HCl, with the choice of acid based on the detection mode and selectivity.[241] The volatile perfluorocarboxylic acids (TFA and HFBA) are advantageous for LC coupled to MS where the acid must not interfere with the ionization process. Hydrochloric acid has corrosive properties, which makes the use of inert systems mandatory. The acid anion also serves as an ion-pairing agent for the basic amino acid residues, implying that the separation selectivity can be altered by choosing acids with anions of suitable hydrophobicity. For example, HFBA increases the retention relative to TFA, whereas phosphoric acid renders the ion pairs more hydrophilic and lowers the retention. The perchlorate ion has chaotropic properties that increase the hydrophobicity of the peptides by forming ion pairs with basic residues[236] (Figure 19).

Many amphiphilic biopolymers, notably membrane proteins and peptides with long hydrophobic segments, require additional eluent components such as nonionic detergents to prevent the solutes from irreversible precipitation in the hydrophobic column.[99,242] However, if the solutes call for such measures in RPHPLC, it might be better to investigate other separation modes. Scaling up RPHPLC separations is relatively simple, and crude synthetic peptides are often separated on analytical columns to develop preparative HPLC procedures for product clean-up, followed by analysis of aliquots of purified material with a different eluent composition or another separation chemistry.[99]

5.3.2 Hydrophobic Interaction Chromatography

This is a variant of RP chromatography suitable for separation of proteins with maintained biological activity.[227,243–247] The concept was first described in the late 1940s as salting-out chromatography[248,249] and has

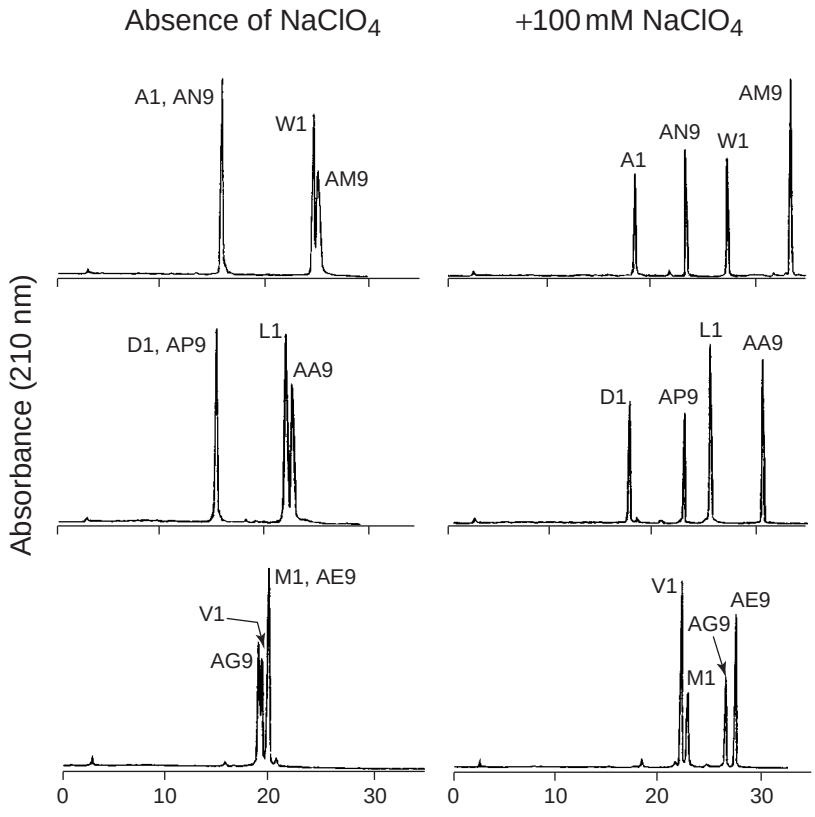

Figure 19 Separation of model synthetic peptides in the absence and presence of 100 mM $NaClO_4$. An aquapore RP-300 C_8 column (220 × 4.6 mm i.d.; 7 µm; 300 Å) was used with a linear (1% acetonitrile per minute) gradient from 20 mM aqueous H_3PO_4 (pH 2.0) to 20 mM H_3PO_4 in 50% aqueous acetonitrile. Sodium perchlorate addition took place in both eluents. (Reproduced from T.J. Sereda et al., *J. Chromatogr. A*, **776**, 163 (1997) by permission of Elsevier Science B.V.)

its roots in protein chemistry. Whereas organic solvents are needed to break the strong hydrophobic interactions in RPHPLC, HIC benefits from being carried out in completely aqueous eluents. Columns used in HIC are only moderately hydrophobic, and retention is based on the tendency of proteins to absorb on surfaces in the presence of cosmotropic salts that enable a salting out of proteins. These salts are positioned to the left in the Hofmeister series:[250,251]

$$SO_4^{2-} > Cl^- > Br^- > NO_3^- > ClO_4^- > I^- > SCN^-$$

The salt cation also affects the interaction strength, although to a lower degree, according to the following:

$$Mg^{2+} > Li^+ > Na^+ > K^+ > NH_4^+$$

Retention is a function of the extent of hydrophobic interaction, and varies with both the hydrophobicity of the stationary phase and the concentration of the cosmotropic salt. The commonly used retention-promoting salts are ammonium sulfate and sodium chloride. Samples are applied to the column in a neutral pH loading buffer containing the cosmotropic salt at relatively high concentration. Elution of adsorbed proteins is achieved by a gradient of decreasing concentration of the salt, with a buffer added at a relatively low concentration to maintain the eluent pH. The features of RPHPLC and HIC are compared in Table 3.

The traditional stationary phases for HIC are soft gels, modified to increase the hydrophobicity, and used in low pressure protein chromatography. However, in recent years a number of rigid phases have become available[252,253] based on silica,[160,252] cross-linked agarose,[106] and synthetic polymers,[254,255] with a variety of functional groups including short-chain alkyl chains, phenyl groups,[254] oligoethyleneglycol,[252] and intermediately polar groups such as poly(alkylaspartamides).[160] The polar–apolar characteristics of nine different RPHPLC and HIC columns were investigated by Rippel et al.[256] using homologous series as probes.

The default eluent anion is sulfate, whose solubility and salting-out effect is well known and suitable for most separations in HIC. However, selectivity is influenced by the salt type used in the inverse gradient, as well as by other additives affecting the conformation of the proteins under separation. The differences between sodium,

Table 3 Comparison of RPHPLC and HIC[95]

Parameter	RPHPLC	HIC
Stationary Phase		
Interaction mechanism	Dispersion	Dispersion
Ligand type	C_8-C_{18}	C_2-C_4, phenyl, ether
Ligand density	High	Low
Hydrophobicity	Strong	Moderate
Mobile Phase		
Type	Aqueous-organic (methanol, acetonitrile)	Aqueous solutions of salts
Operation	Gradient of increasing organic solvent contents	Gradient of decreasing salt concentration
Protein/peptide		
Structure on stationary phase	Unfolded	Native (folded)
Dominant feature	Overall hydrophobicity of the primary sequence	Surface hydrophobicity (tertiary structure preserved)
Loss of biological activity	Considerable	Small

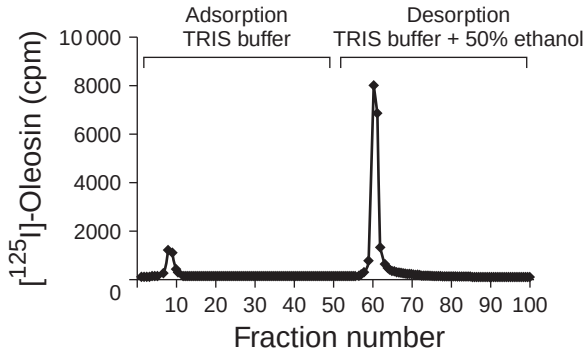

Figure 20 HIC separation with solvent elution. An octyl-Sepharose column (50 × 10 mm i.d.) was loaded with [^{125}I]-oleosin (0.022 mCi) at a flow rate of 0.5 mL min^{-1} in 20 mM tris(hydroxymethyl)aminomethane (TRIS)-HCl buffer (pH 7.4) with a step change to 20 mM TRIS-HCl buffer (pH 7.4) in 50% aqueous ethanol. The adsorbed [^{125}I]-oleosin was eluted with 50% ethanol, demonstrating its high hydrophobicity. (Reproduced from L. Pons et al., *J. Chromatogr. B*, **706**, 138 (1998) by permission of Elsevier Science B.V.)

potassium, or ammonium as counterions to sulfate was studied on a phenyl column by Kato et al.,[257,258] who found that sodium resulted in the most efficient protein retention, and that separations could be modeled to a certain extent by interchanging sodium and ammonium sulfate. Solvents and other chaotropic agents could also be used to speed up elution and to improve resolution while maintaining the proteins in their native state. Selection of optimal conditions for the separation of proteins by ammonium sulfate gradients can be done by employing isocratic elution with different eluents or using shallow gradients to determine the specific salt concentrations at which the specific protein elutes.[259] Polar organic compounds with an ability to alter the structure of water, such as ethylene glycol[260] and urea,[261] or solvents,[262,263] can also be added to the mobile phase to elute very hydrophobic peptides from the column (Figure 20). Alternatively, a surfactant can be added to the eluent[264] or used as a dynamic coating on the stationary phase.[265] The binding of detergents to HIC phases from low salt solutions also provides a way of removing detergents from protein solutions.

The major drawback of HIC is the relatively high concentration of salt present in the eluted proteins. Due to the separation mechanism, proteins elute close to their critical point[266] and are sensitive to adsorption on surfaces unless the concentration of cosmotropic salt is decreased. Desalting can be accomplished by SEC, by dialysis, or by ultrafiltration. However, none of these techniques are readily applicable to on-line detection, such as LC/MS. Buffers containing volatile salts such as ammonium acetate,[267] ammonium carbonate, and ammonium formate are possible, and best separations have been obtained with the acetate.[268] Addition of up to 6 M urea to an ammonium sulfate mobile phase for a silica column with ether groups has been shown to have a substantial effect on the retention of several proteins, attributed to conformational changes.[269]

Solvophobic theory has been extended to HIC. Melander et al.[270] found that a simple two-parameter model can be used to model the retention of proteins and peptides in linear gradient elution (decreasing salt or increasing solvent strength). Geng et al.[271] assumed a mechanism where water is the displacing substance. By varying the pH and the concentration of a series of eluent salts on ether type columns with alcohol or keto end groups, they were able to show that plots of $\log k'$ versus $\log[H_2O]$ for lysozyme and albumin were linear, with isotherms of a type that supported a stoichiometric displacement mechanism. Multivariate analysis was recently applied by Oscarsson and Karsnäs to study the effects of salt type and concentration on adsorption of serum protein to butyl, octyl, phenyl, and mercaptopyridine agaroses.[272] Different adsorption mechanisms were operating depending on adsorbent and buffer conditions. Other studies of HIC interactions have found that the retention of ionizable solutes is explainable by a modified double-layer adsorption model.[273,274] Apart from separation of proteins and peptides with minimal denaturation, HIC has also been used for the

separation of single stranded oligo-DNA and oligo-RNA nucleotide fragments.[275]

Although HIC is generally regarded as one of the softer separation techniques, capable of preserving the separated proteins in their native state,[95] problems with denaturation do occur. For example, De Frutos et al.[276] reported pH-dependent formation of multiple peaks in separations of β-lactoglobulins on a polyether silica, which was ascribed to denaturation, dissociation, or aggregation at different pH and temperature (Figure 21). Conformationally labile proteins elute as broadened peaks[277] and variations in temperature can be used to identify unfolding and assess its extent in HIC separations.[278]

Sample loading can also be problematic because the HIC mechanism requires the proteins to be close to their limit of solubility. Large sample volumes can be used, provided that the salt concentration in the sample is sufficiently high[279] but, for proteins of low solubility, multiple injections of sample with lower ionic strength is a better way of loading a larger sample onto the column.[252] Furthermore, precipitation of solutes on the column can occur if the protein concentration is high and the buffer pH is close to the p*I* of the proteins. In view of the retention-promoting effect of cosmotropic salts in HIC, it is obvious that these salts should be avoided in separation modes where nonspecific hydrophobic adsorption to stationary phases is detrimental, as in SEC and HPLAC. Conversely, an advantage of the high salt loading solutions is that eluates from ion-exchange separations can be applied to HIC columns without desalting.

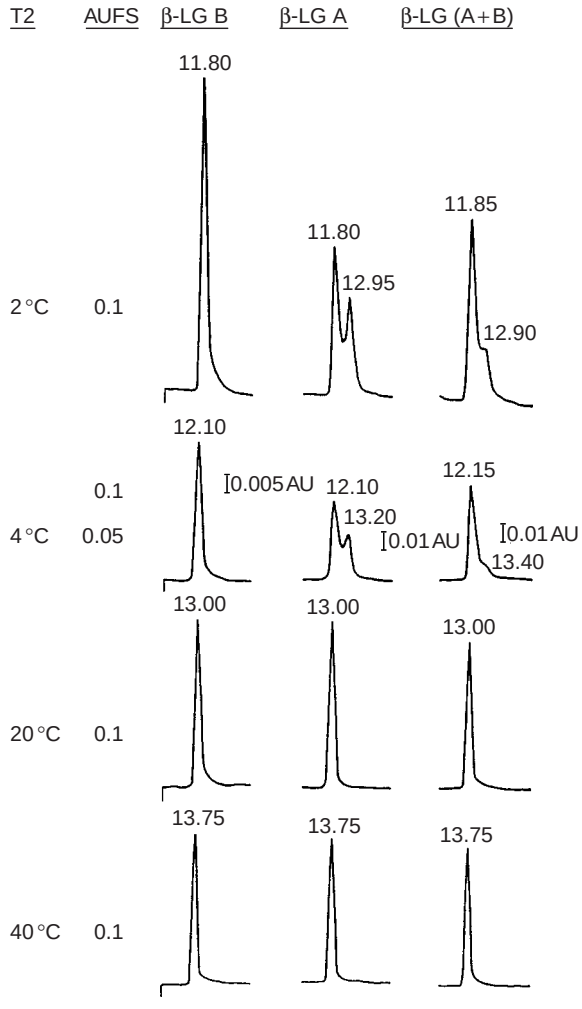

Figure 21 Hydrophobic interaction chromatograms of β-lactoglobulins at different column temperatures using a gradient 0–100% eluent B (0.5 M acetate buffer, pH 4.5) in 20 min at 1 mL min^{-1} flow rate. Eluent A: 3 M $(NH_4)_2SO_4$ dissolved in eluent B. Injections: 7.5-µL aqueous solution of β-lactoglobulins at 10 mg mL^{-1} concentration. UV detection at 280 nm. The numbers on top of the peaks correspond to their retention time in minutes. (Reproduced from M. De Frutos et al., *J. Chromatogr. A*, **778**, 46 (1997) by permission of Elsevier Science B.V.)

5.4 High-performance Liquid-affinity Chromatography

One of the characteristic features of biological macromolecules is their ability to enter highly selective and reversible interactions with other molecules. Examples of such interactions in vivo are the capture of antigens and haptens by antibodies, the attachment of substrates, inhibitors, cofactors and substrates to active sites of enzymes, binding of nucleic acids to complementary nucleotide bases, and the affinity of hormones and toxins for receptors. Such dissociable biospecific complexes form the basis for affinity chromatography and HPLAC.[36,103,280–288] Affinity chromatography is defined as a binary (on–off) separation technique based on specific interactions between solutes and ligands designed for a single compound or a narrow group of biological macromolecules.

The extraordinary specificity of bioaffinity interactions makes it possible to achieve the required separations even with low-efficiency separation materials such as soft gels, albeit at a penalty of long equilibration times and low flow rates. High binding constants also enable the use of sorbent layers with exceptional structure and extreme aspect ratios such as fibrous membranes[283,289] (Figure 22). Low-pressure techniques are therefore still extensively used in purification of biomolecules by affinity interactions. New high-efficiency separation materials have made HPLAC an interesting unit operation in various specific detection schemes, occupying a position between conventional separation science and dedicated immunochemical analysis methods.[290]

The operating principle of bioaffinity chromatography is shown in Figure 23, and encompasses a specific interaction of the molecule to be purified with an immobilized affinity ligand, leading to an attachment of

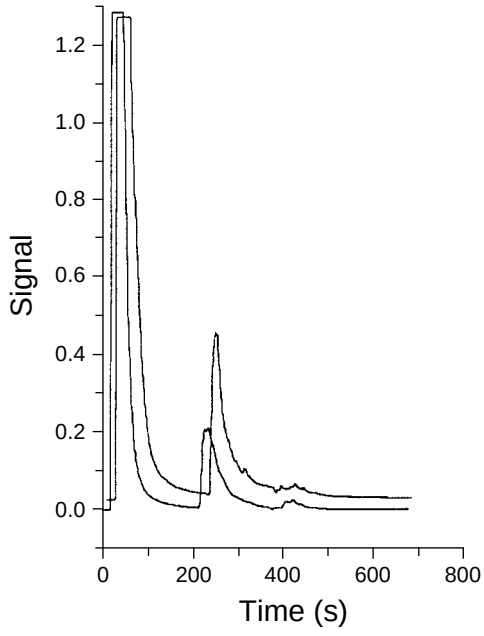

Figure 22 Fast isolation of protein G from 1:5 and 1:10 diluted cell lysate (solid and dotted elution curves, respectively) on a macroporous affinity disc. Adsorption buffer (A), PBS (pH 7.0); desorption buffer (B), 0.01 M HCl (pH 2.0); flow rate, 3 mL min^{-1}; probe volume, 5 mL. Step gradient: 0–2 min, 100% A; 2.0–2.1 min, 0–100% B; 2.1–6.0 min, 100% B; 6.0–6.1 min, 100–0% B; 6.1–10.0 min, 100% A. (Reproduced from C. Kasper et al., *J. Chromatogr. A*, **798**, 71 (1998) by permission of Elsevier Science B.V.)

the solute. The attachment step is followed by a washing step, whereby nonbonded molecules are rinsed from the column. The final step is the elution, where the selectively sequestered molecules are recovered by elution with competing substrate, or by applying conditions that lead to an altered conformation of the attached biological macromolecule that results in loss of specific interaction.

Affinity chromatography can thus be regarded more as a set of chemical operations aided by a heterogeneous surface than chromatography in a conventional sense. Therefore, optimal conditions must be sought for each step in this procedure. The ligand is chosen on the basis of its ability to form reversible complexes with the solutes to be separated, and should have a functional group through which it is attached to a support without seriously altering its biospecific binding properties. The binding constant should be sufficiently high, and the ligand should possess a specificity adequate for the desired separation. The association and dissociation rates of the binding reaction should be fast and the binding strength controllable over a wide range by simple and nondenaturing changes of conditions, so that solutes can be eluted in sharp bands without irreversible conformational changes being imparted on the solute. Ligands for HPLAC can either be biopolymers or low-molecular-weight compounds.

The association strength between the ligand and the solute is central. If the interaction is too weak, the solute will fail to become efficiently sequestered. In contrast, if the interaction strength is too high, it will be hard to find acceptable elution conditions. The specific multipoint interactions that have to be established and dissociated often have slower kinetics than more general interactions with less heterogeneous interaction layers.

5.4.1 Stationary Phases for High-performance Liquid-affinity Chromatography

The traditional bioaffinity supports have been soft hydrophilic gels, mainly cross-linked agarose, and the activation required to covalently immobilize the ligand has often been done by the user. In later years, a variety of pre-activated matrices have emerged, many of them based on alternative materials such as synthetic hydrophilic polymers and inorganic oxides. Rigid supports have enabled the evolution of bioaffinity chromatography into a much faster high performance technique.

The properties of the solid support matrix are very important in bioaffinity chromatography because they affect the stability of the immobilized ligand. Correct choice of matrix is thus essential for successful separation.

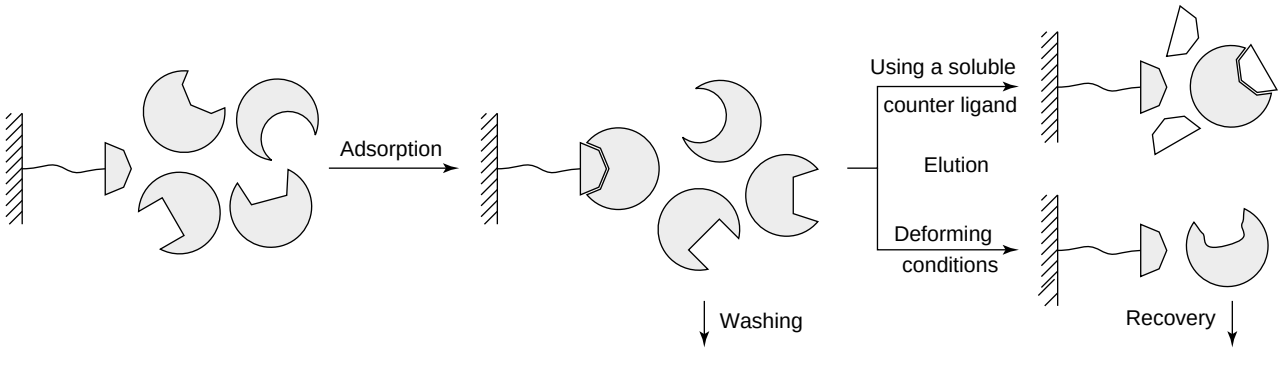

Figure 23 Diagram of the different steps involved in the use of bioaffinity chromatography.

Porath has pointed out the following properties for a suitable bioaffinity chromatography support:[291]

- insolubility;
- a large specific area with sufficient permeability in particles of suitable rigidity and suitable form;
- the absence of native sorption capacity, which translates into a hydrophilic character;
- a chemical reactivity sufficient to allow introduction of affinity ligands;
- a chemical stability under the conditions required for the attachment, adsorption, desorption, and regeneration, including resistance towards microbial and enzymatic attacks.

The size and accessibility of the pores are particularly important, as the porous system should be spacious enough to accommodate both the ligand and the target molecule, and to allow the biospecific interaction to occur undisturbed. Yet to maintain a reasonable capacity, the pore size should not be too large because of its inverse relationship with surface area. A network of evenly sized pores of the correct dimension is therefore required to obtain optimal kinetics and capacity. The surface properties of the pore system are also a key factor. During use, solutes should ideally form complexes only with the bonded affinant and exhibit minimal interaction with the support matrix. Because of the amphiphilic properties of biomacromolecules (which are advantageously used in HIC), a highly hydrophilic support is needed to diminish nonspecific interactions. For the same reason, ionic or ionogenic supports are seldom used. Because of this sensitivity to the nature of the support and the immobilization chemistry, several different matrices should be evaluated when HPLAC methods are developed.

Strong linkage of the affinant to the support is also very important. The support must therefore carry active groups, or groups capable of being activated to enable a stable and irreversible attachment. A large variety of immobilization chemistries have been developed, facilitating the attachment of various affinants. As the desired adsorption and desorption of solutes requires eluents with different pH and ionic strengths, both the sorbent and the ligand linkage must be stable under these conditions.

Highly hydrophilic polysaccharide supports such as agarose, dextran, cellulose, and less frequently starch are often used as supports. Materials based on mixtures of agarose and polyacrylamide, or cross-linked poly(tris(hydroxymethyl)acrylamide) are also commercially available. Recently developed bioaffinity supports include porous poly(styrene–divinylbenzene) beads coated with poly(vinyl alcohol) subsequently cross-linked and activated with epichlorohydrine, butanedioldiglycidyl ether,[292] or glutaraldehyde.[293] These materials are stable in strong bases and compare well with agarose-based materials.[294]

5.4.2 Activated Supports and Ready-to-use Media

Due to the almost limitless variety of affinity ligands, only a limited number of affinity columns with ready attached ligands are commercially available. These ligands are typically specific for families of affinants, such as carbohydrates, antibodies, etc. Media for more specific separations requiring less common ligands are still usually prepared by reacting the desired affinant with an activated support. This support is either purchased activated or given the required binding properties in the laboratory using a suitable reagent, typically chosen among the activation chemistries listed in Table 4. Activated materials are often delivered in bulk form, and can be characterized before use by titration to ascertain that the material has the specified binding capability. This can be done by attaching the ligand and monitoring the decrease in absorbance or another measurable parameter of the ligand in the attachment solution before and after reaction. The type of activation depends on the targeted functional groups on the ligand, its molecular size, the location of the affinity epitope on the ligand, etc. The choice of support matrix, spacer, activation chemistry, and immobilization conditions is a complex process, largely subject to trial and evaluation. A structured experimental approach is therefore advocated.

Immobilization chemistries on silica matrices are different from polysaccharide matrices in that the reactions used for alkyl hydroxyls in polysaccharide carriers cannot be applied directly to the silanol groups. The first activation step is therefore typically a liquid or gas phase[296] attachment of a reactive silane spacer, such as 3-glycidoxypropyltrimethoxysilane (GLYMO),[297] 3-aminopropyltrimethoxysilane,[298] (3-chloropropyl)trimethoxysilane,[299] or 3-isothiocyanato-propyltriethoxysilane.[298,300] The GLYMO activation is one of the most frequently used, because its reactive epoxy groups, introduced in a single-step reaction, can either be used for direct attachment,[301,302] or as a starting point for further reactions after an acid-catalyzed hydrolytic ring-opening into a diol.[303,304] Some other activation chemistries for silica are based on 1,1′-carbonyldiimidazole (CDI),[305,306] *N*-hydroxysuccinimide ester,[307] 4,6-diphenylthieno[3,4-*d*]-1,3-dioxol-2-one 5,5-dioxide,[308] oxalic or adipic dihydrazide,[309] covalent coupling of tresyl chloride-activated dextran to the sulfhydryl groups introduced on silica fibers by silylation with γ-mercaptopropyltrimethoxysilane,[310] and conversion into primary hydroxyl

Table 4 Some characteristics of the more frequently used synthetic activation chemistries used in liquid affinity chromatography[284,295]

Activation	Targeted functional group in ligand	Reaction conditions and properties
Cyanogen bromide (CNBr)	$-NH_2$ (proteins, peptides)	Prepared by reacting hydroxylic supports with CNBr under alkaline conditions, resulting in unstable cyanate ester and cyclic imidocarbonate. CNBr activation is particularly useful for agarose sorbents and activation is accompanied by slight cross-linking. Reaction with ligands takes place in aqueous solution at close to neutral pH and forms bonds mainly of the isourea type, which are positively charged ($pK \approx 9.5$). This imparts anion exchanger characteristics to the final affinity matrix. Stability of the isourea bond is limited and subject to aminolysis and to hydrolysis at alkaline pH, which makes CDI a better choice.
1-Ethyl-3-(3-dimethyl-aminopropyl) carbodiimide (EDC)	$-NH_2$ (amino acids, peptides)	Used for activating carboxyl groups to an *O*-acylisourea intermediate, facilitating formation of amide bonds with amines and thioesters with sulfhydryls. The intermediate is highly sensitive to hydrolysis and must be used immediately upon formation. The reaction has a wide scope in enabling activation of matrix carboxyls that form links to N and S nucleophiles on the ligands, or by activating the ligand and coupling to an amine support. EDC is also widely used for the cross-linking of biomolecules. No spacer is introduced by the EDC chemistry.
CDI	$-NH_2$ (amino acids, peptides)	Capable of activating carboxyls and hydroxyls on supports into reactive carbonylimidazoles and imidazole carbamates, both reacting with amines. Carbonylimidazole-activated supports yield amides with N-terminal and Lys ε-amines in proteins, whereas the imidazole carbamate reacts with nitrogen nucleophiles giving *N*-alkyl carbamate linkages. Neither link introduces charge to the matrix and they provide stable links. CDI provides more stable activated groups than DSC on hydroxyl supports. The activation reagent CDI is extremely sensitive to water.
N,*N*′-Disuccinimidyl carbonate (DSC)	$-NH_2$ (proteins, low M_W substances)	A hydroxyl-containing matrix reacted with DSC yields a hydroxy-succinimide carbonate, which reacts with amines under formation of carbamates. The ligand immobilization reaction can be carried out in neutral aqueous solution or organic solvents. The resulting link is considerably more stable than that formed by the CNBr activation and does introduce a charge to the matrix. Yields less cross-links than CDI.
Oxirane (epoxy) groups (difunctional epoxides or epichlorohydrine)	$-SH$ and $-NH_2$ (peptides, proteins) $-OH$ (carbohydrates)	Oxirane groups are introduced on hydroxylic supports by reaction with epichlorohydrin or diepoxides, which also causes cross-linking. Immobilization of carbohydrates requires alkaline conditions, which calls for a support capable of withstanding high pH. The use of a diepoxide results in concomitant introduction of a spacer arm. Coupling reactions require alkaline conditions (pH > 9 for amines and as much as 11–12 for hydroxyls). The reactivity against thiols is better and can be carried out at neutral pH. Elimination of residual epoxy functionality also requires relatively high pH. Mercaptoethanol can be used, but requires that the ligand lacks easily reduced R–S–S–R bonds.
Divinyl sulfone	$-SH$ and $-NH_2$ (peptides, proteins) $-OH$ (carbohydrates)	Application area similar to oxirane-activated materials, although the reactivity is higher and permits ligand coupling at pH 1–2 units lower. The final matrices are less stable to high pH conditions than with oxirane-activated materials.
Reactive sulfonic esters	$-NH_2$ and $-SH$ (proteins, low M_W substances)	Formed by reacting support hydroxyls with *p*-toluenesulfonyl (tosyl) chloride or 2,2,2-trifluoroethyl (tresyl) chloride, a reaction that is carried out in organic solvent, preferably acetone. This precludes gel matrices that do not swell in solvents. The ligand is attached directly to the carbon bearing the hydroxyl so the coupling chemistry provides no spacer. The stability of the activated support is limited, but storage in weak acid is possible for a limited time prior to ligand coupling. The coupling reaction can be accomplished in water or in organic solvents such as DMF. Tresyl-activated matrices are more reactive than tosyl, and ligand bonding can take place at refrigerator temperature in neutral pH.

(*continued overleaf*)

Table 4 (*continued*)

Activation	Targeted functional group in ligand	Reaction conditions and properties
Glutaraldehyde	$-NH_2$ (especially proteins)	Glutaraldehyde provides a spacer by itself. The activation chemistry is not well characterized and can lead to polymer formation on the support surface. The Schiff base formed is unstable but can be reduced to a stable imine by $NaBH_3CN$, subject to ligand compatibility.
Reactive disulfide groups	−SH (proteins, peptides, low M_W thiols)	An amine functional matrix is reacted with 2-iminothiolane, thereafter with 5,5′-dithiobis-(2-nitrobenzoic acid). The activated matrix is capable of reacting with sulfhydryl groups in proteins, peptides and small molecules. Links are subject to the lability of the −S−S− bond.
Adipic acid hydrazide	−CHO (low and high M_W aldehydes)	Adipic hydrazide is coupled via a hydrazide bond to an EDC-activated matrix. The resulting matrix is particularly useful for immobilizing glycoproteins, leaving critical active sites free.

Warning: Activation chemicals act by undergoing or promoting rapid reaction with functional groups present in biological material and can therefore be collectively regarded as a *highly toxic* category of substances. The table is only intended to provide an overview and cannot list all the risks involved in handling the chemicals involved. Consult material data safety sheets and other safety literature carefully before embarking upon any of the reactions outlined above.

groups[311,312] for coupling of proteins via hydroxyl-targeting activation reagents. Another viable route is the attachment of a polysaccharide layer[146] followed by one of the activation chemistries listed in Table 4.

The conditions during immobilization should be closely controlled and optimized for each immobilized ligand. Obviously, pH is among the more important variables, as the groups targeted on the biological macromolecules are polar groups that are conditionally nucleophilic, dependent on their charge state. For instance, thiolate ions and amines are far more susceptible to nucleophilic attack than thiols and ammonium ions. Type and concentration of buffer, other salts or nonionic agents added to the immobilization solution are also factors that should be controlled. High salt concentrations contribute to decreasing charge repulsion and can be useful in schemes aiming at oriented immobilization.[313] As bioaffinity heads towards designed interactions based on peptides and other bio-oligomers, there is a growing need to study the effect of the immobilization chemistry on the reaction micro-environment in the carrier–ligand–solute interaction region. Some of the important parameters are spacer length, ligand loading, mono- and multipoint ligand attachment, and targeted group in the biomimetic ligand.[314]

5.4.2.1 Spacer Arms Unexpectedly low binding strength and capacity can be due to steric hindrance if the ligands are immobilized directly onto the sorbent surface. These effects are more often seen for low-molecular-weight ligands and macromolecular affinants having the interaction sites located inside their coil, which is often the case with active sites of enzymes. The problem can be partly alleviated by the use of a spacer arm that extends the ligand from the carrier surface to improve its reach into the binding site of the target solute. The molecules used as spacers are bifunctional chains of varying length, with one functional group capable of attaching to the carrier and the other suitable for immobilization of the ligand. Some molecules that serve as spacers are shown in Table 5.

The use of an alkylene chain spacer introduces a hydrophobic moiety that may increase the undesired nonspecific interactions. The decision to use a spacer should therefore be made only if direct immobilization fails. In general, the shortest spacer that provides good results should be selected to avoid crowding the pore space. The spacer density is also important, as a very high coverage serves only to extend the level of the surface, without providing a real spacer effect. Extension of spacers is possible using reactions related to those used in solid-phase peptide synthesis. For example, supports activated with 6-aminocaproic acid can be extended by another 6-aminocaproic acid segment. Alternating use of a diaminoalkane and succinimide will produce a polyamide spacer that is less polar than the alkylene spacer of the same length. The polyamide with the shortest intra-amide distance is poly(glycine), but the use of a polypeptide as spacer can also lead to nonspecific interactions. It should be noted that many of the immobilization chemistries shown in Table 4, such as oxirane and glutaraldehyde, actually immobilize the ligand at a certain distance from the support surface without the use of a dedicated spacer molecule.

5.4.2.2 Capping of Residual Activity Residual activated groups remaining after the ligand has been immobilized can react with eluent and sample components and must therefore be quenched or capped. This is particularly important with large ligands imposing steric hindrance in the immobilization process, and for activation chemistries

Table 5 Examples of spacer arms used to extend the distance from the support surface to the immobilized ligand[284]

Ligand	Spacer structure	Comments
3,3′-Iminobis(propylamine)	$NH_2-CH_2-CH_2-CH_2-NH_2-CH_2-CH_2-CH_2-NH_2$	Soluble in water and most organic solvents. Reacts with matrices capable of binding amines. Should be used in substantial excess in relation to the reactive group density of the support to avoid cross-linking. Yields amine functionalities on a nine-atom spacer.
1,6-Diaminohexane	$NH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-NH_2$	Similar to 3,3′-iminobis(propylamine), except that the linear six-carbon chain provides a more hydrophobic support. Small ligands are most affected by increased hydrophobicity.
6-Aminocaproic acid	$NH_2-CH_2-CH_2-CH_2-CH_2-CH_2-COOH$	Provides a terminal carboxylic group on supports activated to react with amines. The carboxylic group is activated for further reactions with amines with, e.g., 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). Some carboxylic groups remain after the reaction, increasing the hydrophilicity and rendering the ligand-bound sorbent with a weak cation-exchange character.
1,3-Diamino-2-propanol	$NH_2-CH_2-CH(OH)-CH_2-NH_2$	Reactions and use similar to 3,3′-iminobis-(propylamine) and 1,6-diaminohexane, except that the spacer has very low hydrophobicity.
Succinic anhydride	$HOOC-CH_2-CH_2-COOH$	Reacts with amine groups, and can as such be used alternately with diamine reagents for extending the chain length of spacers.
1,2-Diaminoethane	$NH_2-CH_2-CH_2-NH_2$	Used for convenient conversion of carboxylic groups into amines. Steric problems are small and the short alkylene chain provides little hydrophobicity. Can be used to transamidate polyacrylamide, forming primary amine functionalities.

Warning: Spacer molecules are difunctional compounds designed to react with moieties present in biological molecules and should as such be regarded as toxic by default. Laboratory protocols for attachment of these and other ligands to various activated matrices can be found in the literature.[284]

that are stable in aqueous media. Fortunately, most activation chemistries are sensitive to hydrolysis and the simplest way of deactivating these supports can be to leave them suspended in the coupling buffer for a few days after the ligand has been immobilized. This approach does not work for supports activated with amino and carboxylic groups, which introduce an ion exchanger character to the matrices. It should be kept in mind that capping agents targeting particular functional groups may also attack similar groups on the immobilized ligand, with a risk of altering its properties.

The capping agent is usually a small hydrophilic molecule containing a polar functional group. Examples of these compounds are glycine, TRIS, or ethanolamine, which react through their amine groups. The latter is also useful for quenching of carboxylate functionalities. Amine functionalities can be capped with acetic anhydride, whereas mercaptoethylamine, mercaptoethanol, or cysteine react readily with sulfhydryl groups. In general, quenching agents with terminal hydroxyls provide the most efficient blocking, and capping agents with dual functionalities such as glycine, cysteine, or mercaptoethylamine should only be used when the new functionality introduced is desired.[284]

5.4.2.3 Oriented Immobilization The immobilization of high-molecular-weight ligands such as protein, which possess more than one attachment site, can occur in several different ways and through a variety of functionalities distributed along the chain. Immobilization through a

functional group of, or in close vicinity to, the active site may block the access of compounds and decrease the possibilities of interacting with solutes. These problems were first identified in the immobilization of antibodies, and as a result several techniques have been developed for oriented immobilization. Among them are attachment of antibodies via immobilized Protein A and Protein G, proteins on immobilized antibodies, glycoproteins using their carbohydrate moieties, the use of site-directed mutagenesis, and attachment by means of boronate affinity, metal complexes, and biotin–avidin or streptavidin conjugates.[315]

5.4.3 Specific Ligands and Practical Considerations

The boronate group of 3-phenylboronic and other immobilized boronic acids[316,317] is a group-selective small ligand of wide scope through its reversible formation of cyclic boronate esters with coplanar adjacent *cis* hydroxyls (*cis*-diols). This forms the basis for efficient separation of low-molecular-weight compounds such as nucleotides, nucleosides, catecholamines, and carbohydrates,[318] as well as biomacromolecules. Competitive elution is conveniently carried out by adding a low-molecular-weight carbohydrate to the elution buffer. Due to its relatively broad range of target molecules, boronate affinity is often used as a sample preparation and enrichment step for other separative modes. Series coupling with RPHPLC is facile, because the elution conditions in boronate affinity chromatography are rarely effective for eluting the purified substances on an RP column. Bands are thereby sharpened even though relatively large eluate injections are made. The boronate ligand can also serve as an immobilizing group for diol ligands when preparing other affinity chromatography packings. Enzyme substrates can for instance be attached successfully to matrices through this chemistry.

Examples of the use of boronate chromatography are the separation of glycoproteins[319] (e.g. amylases,[320] hemoglobins,[321] and glycosylated albumin[322]) by elution with sorbitol[321,322] or step gradients of increasing acidity,[323] and RNA.[324] The specificity for diols can also be used for selective enrichment of smaller biomolecules and xenobiotics, such as diadenosine polyphosphates in human platelets,[325] ribonucleosides in urine and serum,[326] norepinephrine, epinephrine and dopamine in urine,[327] and the antiviral drug 1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide (Ribavirin) in serum, plasma and cerebrospinal fluid.[328] There are numerous other examples in the literature of preconcentration of nucleosides and related compounds on boronate phases. Relatively uncomplicated preparation procedures are also published, based on reaction of 3-aminoboronic acid with epoxidated[329] or 3-chloropropylated[330] silica.

5.4.3.1 Biomimetic Interactions The interaction of the glycopeptide antibiotic teicoplanin with a D-Ala-D-Ala-6-aminohexanoyl ligand on agarose[331] is a representative of a biomimetic interaction with a small immobilized molecule, imitating a bacterial cell wall component.[332] An interesting observation made by Berthod et al. is that, when used as a chiral separation ligand for amino acids and dipeptides, teicoplanin has a higher preference for L-Ala-D-Ala than for D-Ala-D-Ala.[333] Along the same lines, Pingali et al.[334] attached peptide ligands by direct synthesis onto an amine-functionalized perfusion-type material. The peptide ligands synthesized were GPRP, an effective small synthetic affinity ligand for fibrinogen,[335] and CAQCHTVEK, a cytochrome *c* peptide segment to which a heme group was covalently attached by linkage to the cysteine thiols in order to achieve biomimetic binding to albumins.[336] Both resins worked as expected and could be eluted by lowering the pH or by increasing the salt concentration. A representative of steroid ligands is a silica column with attached dexamethasone, used by Formstecher et al.[337] for the purification of specific antidexamethasone immunoglobulins (Igs) from rabbit antisera. An opposite small molecule interaction worth mentioning is the group-specific affinity of immobilized tomatine, a tomato glycoalkaloid comprising a steroidal part with a four-sugar carbohydrate side-chain, with steroids.[338]

5.4.3.2 Biotin–Avidin A strong small molecule/biomacromolecule interaction is the biotin–avidin binding[339] which is often used in immunoassay protocols. The binding is too intimate to be of direct practical use in affinity chromatography, and the biotin–avidin scheme is therefore probably best utilized as a means for oriented immobilization.[315] The strong binding has also prompted the development of reversible[340] and photocleavable ligands,[341] as well as soluble heterobifunctional biotin affinity–metal chelating polymers capable of expressing biotin–avidin binding and reversible attachment to an immobilized metal ion affinity chromatography (IMAC) sorbent[342] (see below). The strep-tag technique is a way of using this interaction, where the expensive avidin or streptavidin is replaced by a recombinant core streptavidin comprising a nine amino-acid peptide sequence with streptavidin-binding activity.[343] Fusion proteins are eluted by addition of biotin to the buffer. Evolving developments along the same lines are fusion of peptide sequences specifying biotinylation[344] and a truncated rat neurotensin receptor.[345] The avidin–biotin interaction has also been used for immobilizing liposomes to stationary phases (Figure 24) for studies of drug–membrane partitioning.[346]

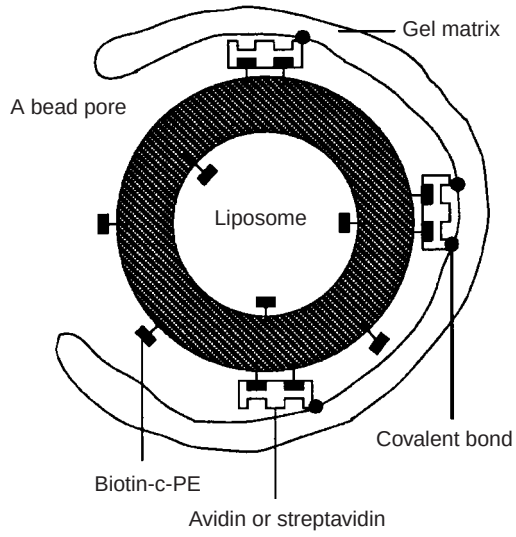

Figure 24 Diagram of a biotinylated liposome immobilized by multiple-site (strept)avidin–biotin binding in a gel matrix. (Reproduced from Q. Yang et al., *J. Chromatogr. B*, **707**, 140 by permission of Elsevier Science B.V.)

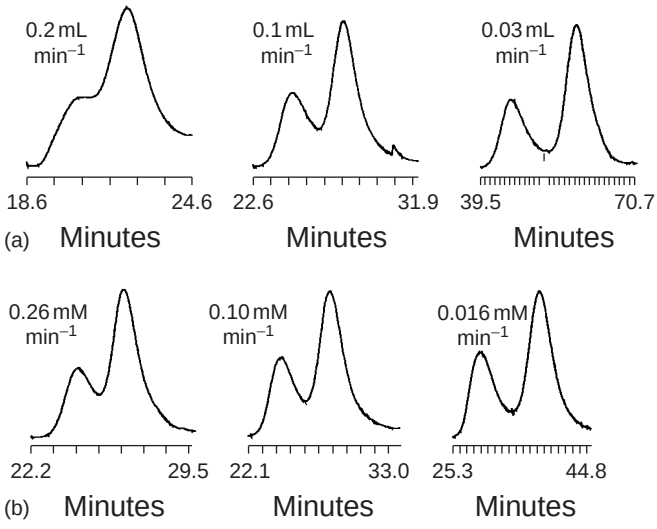

Figure 25 Effect of (a) eluent flow rate and (b) gradient steepness on the separation of recombinant glutathione *S*-transferase enzymes rP1-1 and rA1-1 on a HEMA-based *S*-octyl glutathione affinity column (50 × 3 mm i.d.). Mobile phases: A, 10 mM sodium phosphate (pH 6); B, 200 mM NaCl in A; C, 5 mM *S*-hexylglutathione in B. Samples were loaded in A, 0–9 min, 0.1 mL min^{-1}; washed in B, 10–14 min, 0.8 mL min^{-1}; and the gradient of 0–100% C commenced at 15 min. Detection: 280 nm. (Reproduced from J.B. Wheatley et al., *J. Chromatogr. A*, **676**, 88 (1994) by permission of Elsevier Science B.V.)

5.4.3.3 Enzyme–Substrate/Product/Inhibitor Interactions Enzyme–substrate/product/inhibitor interactions provide a means for the separation of enzymes according to the affinity principle. It is difficult to draw clear distinctions between these types of mechanisms, as most enzymes are inhibited not only by their products but also by immobilized substrates, cofactors, and by compounds of high structural similarity to any of these ligands. An example of a substrate interaction is the use of glutathione immobilized via an *S*-octyl spacer on a cross-linked 2-hydroxyethylmethacrylate carrier for affinity separation of glutathione *S*-transferase isoenzymes.[347] Elution is effected by adding glutathione modified with *S*-alkyl groups of different length to the eluent and the eluate from the affinity separation step can be coupled directly via an on-line coupling to an RPHPLC column.[348] The best results are obtained with very low flow rates and shallow gradients in the elution step from the affinity column, as seen in Figure 25. This is a good illustration of the poor kinetics in affinity chromatography compared to many other modes of chromatography, due to the typical slow dissociation rates of protein–ligand complexes.[349]

The interaction of anhydrochymotrypsin with tryptophan immobilized via a six-carbon spacer, based on aromatic amino acids being the cleavage points for trypsin enzymes, has been used for separating anhydrochymotrypsin from related active enzymes.[350] Cholesterol has analogously been used as a ligand for purification of cholesterol-esterifying enzymes.[351] The utilization of enzyme–substrate interactions in chromatography and other branches of analytical chemistry was recently reviewed by Freitag.[352]

An example of a cofactor interaction is the nucleotide analog ligand 8-((6-aminohexyl)amino)-2′-phosphoadenosine-5′-diphosphoribose, which produces a high and specific retention for nicotinamide adenine dinucleotide phosphate $(NADP)^+$-dependent enzymes when attached to epoxy activated silica.[353] Detachment of the retained enzymes can take place either by a salt gradient, or by competitive elution with $NADH^+$. The reverse principle is also possible, namely immobilization of an enzyme for retaining cofactors and xenobiotic mimetics, as exemplified by the immobilization of alcohol dehydrogenase for separation of adenine nucleosides, adenine nucleotides, and triazine dyes.[304] Considering the nature of these compounds and the separation efficiency obtained, this separation could probably be better solved with other HPLC techniques.

5.4.3.4 Enzyme–Inhibitor Complexes Examples of enzyme–inhibitor complexes being used in affinity schemes are replete in the literature. These reactions are normally highly specific and the sorbents are therefore dedicated for the target solute and synthesized by the operator on the basis of unfunctionalized or preactivated sorbents. Interactions could be divided into

three categories: protein–protein interactions, and interactions between small molecules (natural or synthetic) where the small molecule is either the ligand, or the separated solute. One of the most widely used protein–protein interactions is the trypsin–trypsin inhibitor complex, and conditions have been published for immobilizing bovine pancreatic trypsin inhibitor to silica[312] and HEMA copolymer particles[354] for affinity chromatographic purification of trypsin and related serine proteases. This type of interaction can be used with either the enzyme or its inhibitor as ligand, providing a means for purifying and separating both the serine proteases and their inhibitors. Similar schemes are available for many other enzymes. Immobilized anhydrotrypsin serves as an example both of a modified enzyme as an affinity ligand, and an enzyme–inhibitor interaction. Anhydrotrypsin is prepared from bovine trypsin by transforming a serine residue at the catalytic site into a dehydroalanine residue. This bereaves the transformed enzyme of its catalytic activity, but a strong affinity against trypsin digestion products is maintained, so that peptides with Arg, Lys, or *S*-aminoethyl cysteine residues at the C-terminus are adsorbed under weakly acidic conditions with high selectivity.[355] A related scheme based on immobilized anhydroelastase, purified on the product analog ligand $(Arg)_3$ on a hexyl spacer, has also been described as a tool for purification and search for other naturally occurring protease inhibitors.[356] An advanced, ternary indirect immunoaffinity enzyme–inhibitor scheme was recently devised for chromatographic studies of the role of proteases from the same leguminous seeds that are the sources of their inhibitors. Salt elution was used to dissociate the enzymes from a multicatalytic protease complex purified from winged bean seeds, using an anti-(winged bean) chymotrypsin inhibitor immunoglobulin G (IgG) as ligand.[357]

4-Aminobenzamidine and other benzamidines are examples of synthetic enzyme inhibitors that inhibit the activity of the serine proteinases and provide a strong retention for a number of these enzymes when immobilized onto silica due to interactions with their active sites.[358,359] Another example of a group of synthetic inhibitory ligands are peptide boronic acids, which target chymotrypsin-like proteinases and elastolytic enzymes. Columns with Ala-Ala-boroPhe and Ala-Ala-boroVal attached through their N-terminals have a preferential retention for bovine α-chymotrypsin and porcine pancreatic elastase, respectively.[360] A series of thymidine derivatives showing competitive inhibition on thymidine kinase have been attached to agarose carriers and evaluated as affinity media in various orientations with a retained ability to bind thymidine kinase.[361] Chaiken discussed the use of antisense peptides in affinity technology[362] and gave an overview of trends towards molecular recognition and tagging techniques in bioseparations.[363] An example of a natural peptide–protein interaction is the Ca^{2+}-dependent affinity of calmodulin for melittin, a 26-residue α-helical peptide present in bee venom that was coupled to GLYMO-activated silica by Foster and Jarrett[364] and used for affinity chromatography of calmodulin.

Synthetic anthraquinone dyes can be considered as a group of ligands with a serendipitous biomimicry, based on the resemblance of the core structure of these dyes, borrowed from the textile dye industry, with adenosine nucleotides and adenine enzyme cofactors. A large number of dyes have been evaluated as affinity ligands,[365] but only a limited number, in particular Cibacron Blue F3G-A[366,367] and Procion Red HE-3B,[366] have become widely used for the purification and analytical separation of nucleotide binding of enzymes such as kinases, dehydrogenases, and ATPases, but also nonenzyme plasma proteins,[295,367–369] in which case the retention should be considered as emanating from less specific mixed-mode interactions.[370] Some examples of synthetic affinity ligands are shown in Figure 26.

Proteins are bound to the dye ligands by specific interactions with the protein active site or through various nonspecific interactions.[372] In general, protein elution from dye ligands can take place by salt gradients, sometimes with altered pH, but enzymes are often eluted more efficiently and at higher purity if a biospecific eluent component such as a substrate[373] or substrate analog,[374] an inhibitor,[371,375] or a product or product analog[376] is added to the eluent either continuously at a low concentration or in pulses. Cofactors such as nucleotides are the most common biospecific displacers,

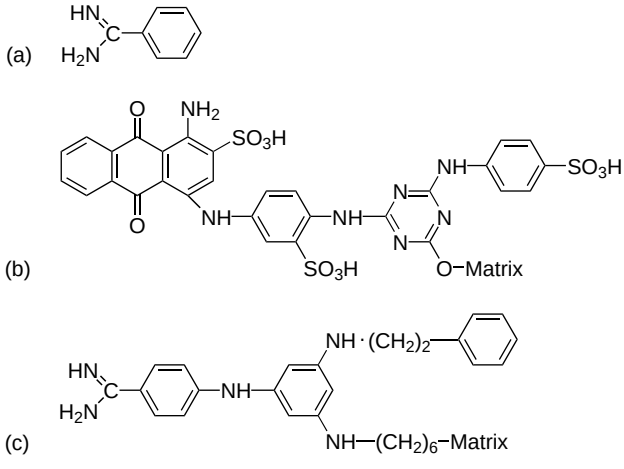

Figure 26 Structures of (a) the synthetic inhibitor benzamidine, (b) the synthetic dye Cibacron Blue, and (c) a recently designed affinity ligand for trypsin-like serine proteases. (Reproduced from C. Koch et al., *J. Chromatogr. B*, **718**, 46 (1998) by permission of Elsevier Science B.V.)

and sometimes an oligonucleotide may also be needed to develop a biospecific solution phase interaction.[377] Selectivity to different cofactors has been used as a tool for separating related enzymes and for differentiating between active and inactive forms of enzymes.[378] Some metal ions (Zn^{2+}, Co^{2+}, Mn^{2+}, Ni^{2+}, Cu^{2+}, Ca^{2+}, Mg^{2+}, and Al^{3+}) are capable of promoting the binding of certain proteins to dye ligands, whereby chelating agents are used for elution.[379] Ethylene glycol[380] and chaotropic ions[381] may also be employed in eluent schemes to effect elution or for removal of strongly bound components from the affinity matrix. Another synthetic dye ligand is the *N*-methyl acridinium group,[382] which in different incarnations has been used for purification of acetylcholinesterase.[383] Carroll et al.[384] recently presented a convenient way of preparing an acridinium affinity sorbent by attaching 9-amino-1,2,3,4-tetrahydroacridine (tacrine) to an epoxy carrier in a one-step reaction. Dye affinity chromatography is a mature and highly valuable technique in view of the high selectivities that can be obtained with this set of relatively inexpensive ligands. Stationary phases with various standard dyes are available from a number of suppliers, in the form of conventional soft bioseparation gels and as HPLC columns.

Recent developments in dye affinity chromatography have focused on refining the interaction modes and understanding the interaction processes. Chemistries for immobilization of dye ligands via epoxide-activated spacers instead of the conventional coupling to surface hydroxyls were recently investigated by Scoble and Scopes,[385] who showed that the epoxide immobilization route gave materials with considerably higher ligand density. Mattiasson et al.[386] addressed the nonspecific interactions of dye affinity sorbents by shielding the ligands with layers of nonionic, water-soluble polymers such as poly(vinylpyrrolidone) or poly(vinyl alcohol). The same authors also reviewed new developments in dye affinity for bioprocessing, dealing specifically with new dyes and conditions for the protein–ligand attachment and detachment steps.[387] An interesting extension of this work is their use of water-soluble polymers for protein displacement in dye–ligand chromatography.[388] It was demonstrated that porcine muscle lactate dehydrogenase (LDH) could be efficiently displaced from a blue agarose gel by poly(*N*-vinyl pyrrolidinone) and poly(*N*-vinyl caprolactam).

As the latter polymer is thermoresponsive, it could be precipitated from the eluate and recovered by centrifugation at 45 °C. PEI was found to be an even more efficient displacer, but requires more elaborate regeneration of the dye–ligand column. Bohácová et al.[389] investigated the solution-phase interactions between seven anthraquinone dyes with and without the triazine moiety on the activity of LDH at different pH. The inhibitory action was found to be dependent on configuration similarity and spatial arrangement of acidic and basic groups in NADH and the tested dyes, which in turn was dependent on the dye acid–base properties (Figure 27).

5.4.3.5 Bioaffinity Chromatography Antibodies or Igs are multimeric polypeptides expressed by cells as recognitive elements for mediating an immunological response to foreign material. The specific affinity of antibodies is usually very high against biomacromolecules of nonhomogeneous composition and antibodies are therefore capable of highly specific sequestering of proteins, polysaccharides, and other structurally heterogeneous macromolecules from biological fluids. This forms the basis for bioaffinity chromatography. An antibody will be produced against a compound only if the four immunogenicity criteria are met, i.e. foreignness, molecular size, chemical composition, and complexity.[390] However, it is possible to induce the formation of antibodies against haptens, small molecules that are antigenic but nonimmunogenic because of their insufficient molecular weight, by conjugating hapten molecules to a large protein carrier and using the conjugate to induce the immunogenic reaction. The specificity of antibodies against small molecules is usually lower, and a considerable cross-reactivity is quite common. This lack of specificity is problematic in nonseparative immunoanalysis schemes, but can be advantageous for obtaining separation of closely related substances in immunoaffinity chromatography.

Antibodies are either polyclonal or monoclonal. Polyclonal antibodies are produced by immunizing a living organism, monitoring its immunoresponse, and harvesting the produced antibodies from serum. Trapping and purification[391] are often done by the aid of mixed-bed ion-exchange, hydroxyapatite, or HIC, or by some of the IgG-specific affinity schemes described later in this section.[392] Characteristic of polyclonal antibodies is a large heterogeneity in binding strength to the antigen, often to different regions (epitopes) of the antigen. This property can make polyclonal antibodies valuable in enrichment schemes, as exemplified by the on-line extraction of the benzodiazepine flunitrazepam and its main metabolites,[393] and isolation of aflatoxins,[394] but they are less useful as mediators of analytical separation due to the substantial variation in binding strengths towards single antigens or haptens. Monoclonal antibodies are produced in immortal cell lines (hybridomas) that are subjected to the antigen or a hapten-carrier conjugate and grown. Individual cells are thereafter selected and used to establish a cloned source of identical antibodies that are produced as long as the cell line is kept alive. It should be realized that each hybridoma cell produces different antibodies following the initial immunization, and that the

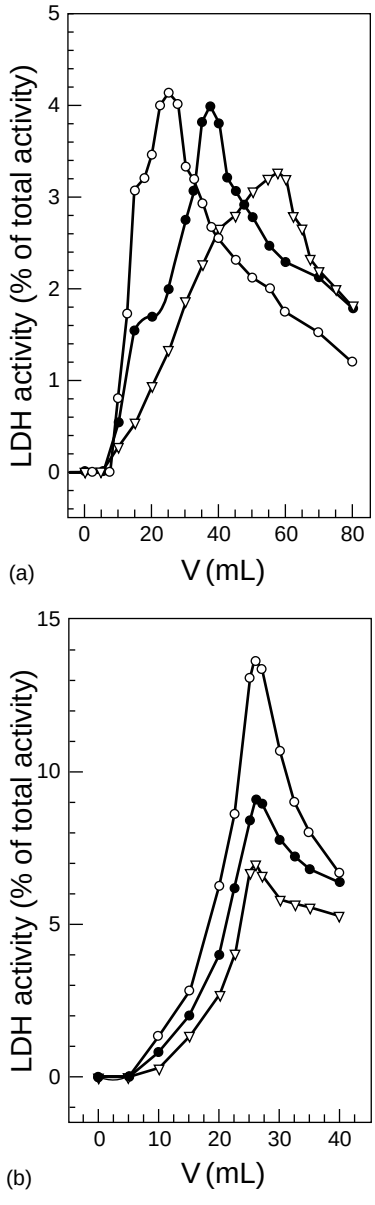

Figure 27 Interaction of LDH with porous cellulose beads functionalized with the nontriazine dye 1-amino-4-[4-(1-sulfonyl-ethyl-2-sulfoxy)]-2-(9,10-anthraquinone)-sulfonic acid (C.I.19) and the triazine-containing dye 1-amino-4-{4-[3-chloro-2,4,6-triazinyl-amino-5-(3-sulfanilino)-3-sulfanilino]}-2-(9,10-anthraquinone)-sulfonic acid (C.I.2) using zonal affinity chromatography. (a) Zonal chromatography of LDH on (C.I.2)–cellulose beads (138 mM per liter of gel) using elution with a mobile phase containing different concentrations of NADH ($\nabla = 100$ mM, $\bullet = 300$ mM, $\circ = 500$ mM). (b) Zonal chromatography of LDH on (C.I.2)–cellulose beads (389 mM per liter of gel) using elution with different concentrations of NADH ($\nabla = 20$ mM, $\bullet = 50$ mM, $\circ = 100$ mM). Eluent: 20 mM phosphate buffer, pH 7.0. (Reproduced from V. Bohácová et al., *J. Chromatogr. B*, **715**, 278 (1998) by permission of Elsevier Science B.V.)

process of selecting a particular cell for cloning is important. It is thus possible to select an optimal monoclonal antibody based on the desired characteristic. Off-the-shelf monoclonal antibodies are commercially available against many analytically important antigens that must be determined at very low concentrations. There are also commercial enterprises that will undertake development and production of custom monoclonal antibodies, but the cost of such reagents can be prohibitively high.

The antibody used for analytical recognition is usually IgG, which is a class of tetrameric polypeptides with M_W around 150 kDa. Pepsin digestion of intact IgG produces a fragment $F(ab)_2$ of approximately 100 kDa containing the antigen recognition elements of the antibody. The remainder is the constant (Fc) region which does not carry specific recognition elements. Both intact antibodies and $F(ab)_2$ fragments can be used as selective elements on immunosorbents. However, because only the Fc fragment is glycosylated, the possibility of oriented attachment through carbohydrate-directed attachment chemistries (Figure 28) is lost if the Fc fragment of the IgG is removed. The most sensitive part of the procedure is the immobilization of the antibody to the carrier, which could be either a preactivated or an in-situ-activated matrix. The carrier surface properties, the immobilization chemistry, and the conditions used during the immobilization will determine the biological activity of the interaction layer, i.e. to what extent the antibodies are capable of interacting with the antigen. Techniques for oriented immobilization[315] should be considered if immunosorbents with acceptable binding capacity and site homogeneity cannot be produced with conventional chemical immobilization. The kinetics involved in immobilization of antibodies to HPLC supports were recently investigated by Oates et al.,[395] for the coupling of periodate-treated IgG to dihydrazide-activated silica. Important factors are the coupling site densities on the antibodies and the support, the amount of antibody in relation to the support, and the reaction slurry density, i.e. the degree of dilution by the coupling buffer. It was found that the density of the activated site on the support has a low influence on the binding rate, which is instead determined by the availability of coupling groups on the antibodies. Enhanced binding is often seen for proteins on epoxy carriers in the presence of cosmotropic salts, explained by an association of the protein to the carrier surface according to a HIC mechanism. The high concentration achieved on the surface by salting-out the protein is supposed to drive the immobilization reaction to a faster completion.[301,302] Although amine-based buffers are generally to be avoided when immobilizing proteins onto supports activated with electrophilic groups, immobilization can still apparently be enhanced by more than molar concentrations of ammonium sulfate[302] (Figure 29).

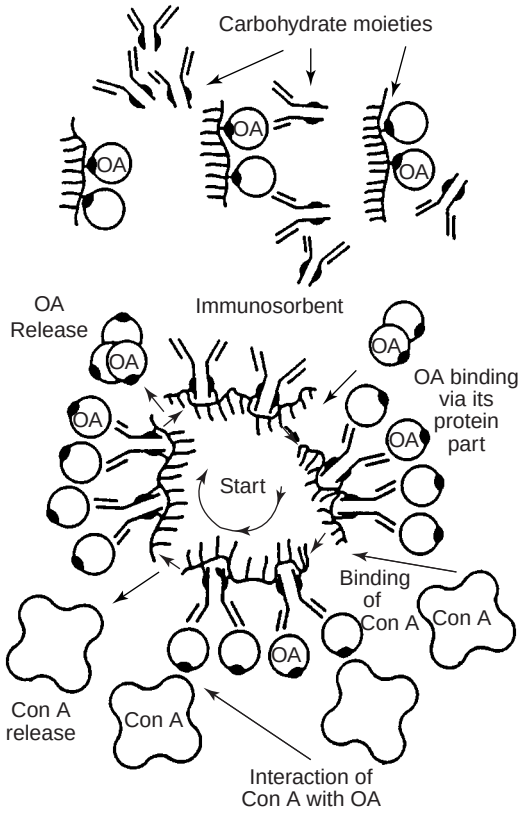

Figure 28 An example of oriented immobilization of IgG, mediated by a hydrazide attachment chemistry targeting the carbohydrate moiety situated in the nonspecific Fc part of the antibody. The IgGs used in this illustration are raised against ovalbumin (OA). After attachment to the IgG, OA is used for isolation of concanavalin A (Con A). By attaching a protein to a monoclonal antibody in this manner, the interaction surface of the IgG-bound protein will be on the surface opposite the antibody epitope, and hence an oriented immobilization is achieved. (Reproduced from J. Turková, *J. Chromatogr. B*, **722**, 21 (1999) by permission of Elsevier Science B.V.)

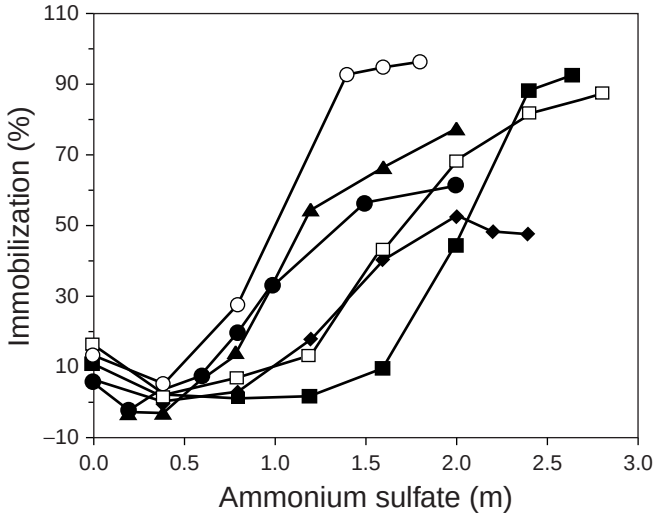

Figure 29 Percentage protein (0.33 to 1 mg mL^{-1} in 10 mM phosphate buffer, pH 7) immobilized on Hydropore EP 12 epoxy affinity supports versus the concentration of ammonium sulfate added to the protein solution. See the original reference for experimental details. ○ = IgG_1k, = transferrin, = α-lactalbumin, ■ = myoglobin, ● = BSA, △ = carbonic anhydrase, = β-lactoglobulin B. (Reproduced from J.B. Wheatley, D.E. Schmidt, Jr, *J. Chromatogr.*, **644**, 13 (1993) by permission of Elsevier Science B.V.)

Protein A and Protein G are bacterial surface receptors from *Staphylococcus aureus* and *Streptococcus* sp., respectively, which are capable of binding to the Fc fragment of Igs from many species. The specificity of native Fc-binding proteins differs, with an overlap among Ig classes, subclasses and species where the Igs are raised. By the use of recombinant techniques, it has become possible to produce A/G fusion proteins which combine the binding capabilities of Protein A and Protein G. All three proteins are commercially available. The important feature is the capability of binding to the Fc part of an IgG, which means that these bacterial receptors can be used for oriented immobilization of intact IgGs and ensuring that the recognitive element is protruding from the surface of the immunosorbent. It is also useful in cases where the IgG has to be replaced frequently on the immunosorbent.[396] Another technique that can accomplish oriented immobilization of IgG is anti-IgG antibodies. The drawback of both these techniques is the relatively high cost of the attachment chemistries.

Examples of affinity separations based on antibodies are numerous in the literature and often tangential to immunochemical analysis techniques, where the ultimate purpose of using the biospecific interaction is not to produce a chromatogram, but to isolate and facilitate the detection of a biomacromolecule or a hapten at very low concentrations. A few illustrations of chromatographic uses of immobilized antibodies are in the enrichment of mycotoxins (aflatoxins,[397,398] ochratoxin,[398] fumonisins,[399] and lysergic acid diethylamide (LSD)[400]). A different kind of specificity was utilized by Kasper et al.,[289] who immobilized human IgG to a macroporous glycidyl methacrylate-*co*-ethylene dimethacrylate polymer disk and used the binding ability of IgG to isolate *E. coli* recombinant Protein G. Histidine was introduced some years ago as a pseudoaffinity ligand. The initial target was IgG,[401] which can be retained and eluted under mild conditions from particulate sorbents or poly(ethylene vinyl alcohol) hollow fibres. Histidine ligands have subsequently proved to be useful for enrichment and purification of many other proteins.[402] Subtractive affinity chromatography on immobilized antihuman albumin and antihuman transferrin IgGs was described by Wheatley[403] as an alternative way of purifying IgG from other plasma proteins. Some HPLC

techniques have also been used to monitor the production of antibodies in hybridoma cultures, by using nonporous silica columns with immobilized Fc receptors.[404]

5.4.3.6 Chromatographic Studies of Protein–Drug Interactions Studies of the binding of small drugs to immobilized proteins is also a field of biomacromolecule separation that falls within the scope of this article. In reality, this is an inverse affinity chromatography scheme, where the small ligand molecule is presented in the liquid phase to a sequestered biomacromolecule acting as the stationary phase. Knowledge that protein components of blood serum bind drugs differently[405] necessitates studies of drug–protein binding[406] in pharmacology. Affinity chromatography on immobilized proteins has been established as a fast and reliable way of conducting these studies, focused on a narrow range of proteins known as carriers of small molecules in blood. A number of recent papers can serve as entry points into this literature. Kaliszan[286] reviewed how chemometrics in combination with affinity chromatographic retention studies of drug analytes on biomacromolecular ligands can be used to establish quantitative structure–retention relationships relevant to molecular pharmacology and rational drug design. Layers of keratin[407] and collagen[408] covalently bound to silica serve as models for percutaneous penetration. Silica-immobilized melanin is used to elucidate the pharmacologically and toxicologically important binding of many drugs and toxic substances to this localized molecule.[409–411] Hage et al. have published a number of studies on interactions between native and modified human serum albumin (HSA), and in a recent work they characterize the specificity and cross-reactivity of a set of drugs that have evolved as indicators for minor binding sites in HSA.[412] Another major transporter in serum is α_1-acid glycoprotein (orosomucoid), which binds basic substances preferentially and has been immobilized on agarose and used for selective but nonspecific clean-up of β-antagonists from urine.[413] These techniques should be regarded as a set of tools that can be used in combination with immobilized biomimetic components such as liposomes, proteoliposomes and biomembrane vesicles,[414] and immobilized artificial membrane chromatography.[415,416] The latter technique is based on partitioning to a phospholipid bilayer similar to a biological membrane (Figure 30), deposited on silica, and has become a valuable tool for screening drug–membrane interactions,[417] and for determining hydrophobic peptide–membrane interactions.[418]

5.4.3.7 Designed Synthetic Biomimetic Ligands A trend in specific bioaffinity chromatography is towards designed synthetic biomimetic ligands, which is in line with concurrent drug–protein interaction research. An important additional driving force for this development is the need for highly selective, sturdy, and inexpensive separation techniques for down-stream processing in the biotechnologies.[419] The anthraquinone dyes described above are intriguing candidate substrates for ligand modification, in order to establish new interactions that are tailored to particular solutes. Consequently, a substantial amount of work has been devoted to the systematic modification of anthraquinone dyes, aiming at developing ligands with more well-defined biomimetic properties.[420–422] A more ab initio approach was taken in a recently published work by Li et al.,[423] where a series of low-molecular-weight ligands mimicking the binding ability of Protein A to IgG were synthesized, based on

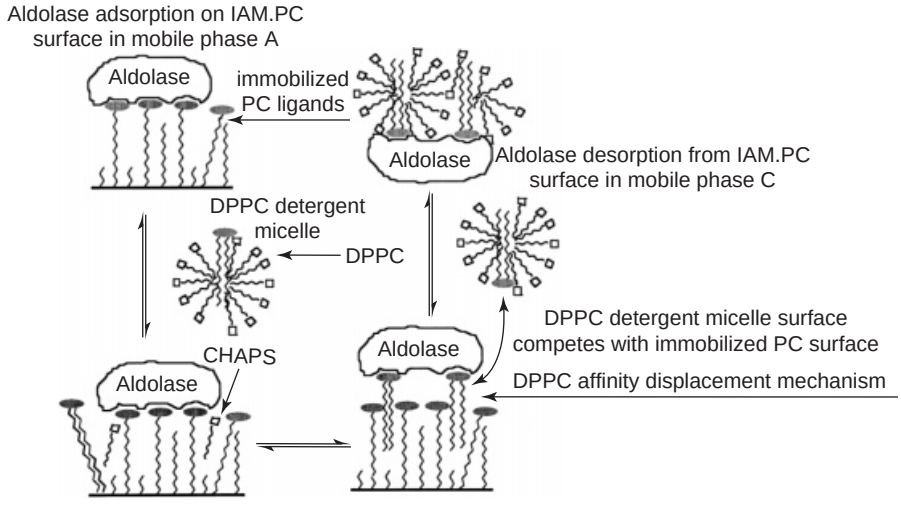

Figure 30 Rat liver aldolase purification by immobilized artificial membrane HPLC using 3-[(cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) and dipalmitoylphosphatidyl choline (DPPC) detergent mixtures. (Reproduced from H. Liu et al., *J. Chromatogr. B*, **703**, 61 (1997) by permission of Elsevier Science B.V.)

X-ray crystallographic studies of the specific recognition and complexation region between Protein A and the Fc fragment of IgG. Immobilized to an agarose carrier, one of these synthetic ligands, ''artificial Protein A'', has the ability to bind several IgGs, and can be used to purify IgG from human plasma.

A related approach is to determine amino acid sequences of antibodies, to synthesize these sequences, and to immobilize these synthetic antibodies as biospecific recognition elements. An example of such work is by Lasonder et al.,[424] who also used a surface plasmon resonance biosensor to study the interactions. In the area of metal interaction affinity chromatography, Hutchens and Yip[425] used repeating units of the (GHHPH) amino acid sequence to produce a sorbent with similar binding capacity for Cu^{2+} and Zn^{2+} that was used for selective adsorption and isolation of proteins from human plasma. Based on these results, they suggested that designed metal chelating ligands could be used to elucidate biologically relevant metal-ion-dependent interactions and transfer events in vitro.

5.4.3.8 Lectins A ligand group complementing boronate media for affinity separation of oligosaccharides are lectins, a family of carbohydrate-binding proteins extracted from plants, seeds, and fungi. Lectins derived from different plants target different sugar sequences and result in different selectivities for carbohydrates. For instance, concanavalin A (Con A) binds selectively to internal and nonreducing α-mannosyl residues, whereas wheat germ agglutinin binds to terminal *N*-glucosamine groups.[426] Jacalin, extracted from seeds of the jack fruit (*Artocarpus integrifolia*), has a galactose-specific interaction,[427] providing an affinity for proteins with O-linked oligosaccharides.[428] A useful immunochemical feature is its ability to isolate IgA,[429] which has been used to deduce the structure of the IgA core oligosaccharide by enzymatic scission and MS.[430] A micropreparative isolation technique for jacalin based on weak anion exchange chromatography has been published.[431] A wide spectrum of additional lectins have been purified from various sources, showing a substantial differences in their oligosaccharide binding patterns. Release of oligosaccharides from a lectin column is usually accomplished by displacement with asialic or sialic mono- or oligosaccharides or their derivatives, singly or sequential, depending on the interaction pattern of the lectin ligand.

Lectins have been used in HPLAC since the technique was in its infancy,[432] and many applications deal with the enrichment and separation of small carbohydrate compounds such as nucleotides[433] and substituted monosaccharides.[434] However, the real potential of lectins is a result of the discovery and characterization of lectins with different oligosaccharide interaction patterns from many different biological sources. Lectin interactions have thereby developed into powerful and indispensable tools in modern glycobiology, a branch of science devoted to elucidating the role of carbohydrates in cellular function and recognition. Immobilized lectins are thus used for separation of glycolipids, glycoproteins, and recombinant products based on their glycosylation patterns, and for fractionation of glycopeptides and oligosaccharides derived from glycoproteins. The combination of several lectins in multidimensional separation systems has lately become customary for elucidating complex carbohydrate structures bound to proteins,[435,436] for instance when used in a multidimensional LC/MS separation system to reduce the complexity of an endoproteinase LysC digest of a recombinant glycoprotein[436] (Figure 31). The preparation of lectin columns and the selectivity of different immobilized lectins in affinity chromatography was recently reviewed by Endo.[437] An extensive listing of lectins and a review of their use in affinity chromatography has also been published in an on-line journal.[438]

5.4.3.9 Sugars and Sugar Derivatives A good illustration that affinity chemistries can be inverted is that lectins are purified on affinity columns with mono- or oligosaccharide ligands resembling their target. For instance, immobilized glucosamine has been employed as a ligand for the micropreparative HPLAC purification of Con A,[439] and a glycoprotein can serve the role as the oligosaccharide ligand for separating lectins.[440] Honda et al.[441] attached glucamine, maltamine, and lactamine to epoxy-activated methacrylate resins and evaluated these as affinity supports for lectins. Con A, lentil lectins, and pea lectins were sequestered by the maltamine resin, whereas ricinus agglutinin, peanut agglutinin, and soy bean agglutinin had a strong specific affinity for the lactamine resin. Using aqueous acetonitrile eluents, they could also use immobilized sugars to group separate pentoses and 6-deoxyhexoses from hexoses. In later works, they used reductively aminated oligosaccharides prepared by hydrazinolysis of ovalbumin[442] and homologous oligosaccharide series[443] as ligands for similar separation purposes.

Lakhiari et al.[444–446] have presented a series of papers where *N*-acetylneuramic acid (NANA; sialic acid) is immobilized onto porous silica that has been deactivated by coating with a DEAE-modified dextran layer and activated with several different activation chemistries capable of immobilizing carbohydrates. The rationale for synthesizing these sorbents was the presence of sialic acids in cellular insulin receptor structure, and the materials did indeed express a quasi biospecific interaction with insulin which made it possible to adsorb and separate insulins specifically. The interaction pattern of these materials

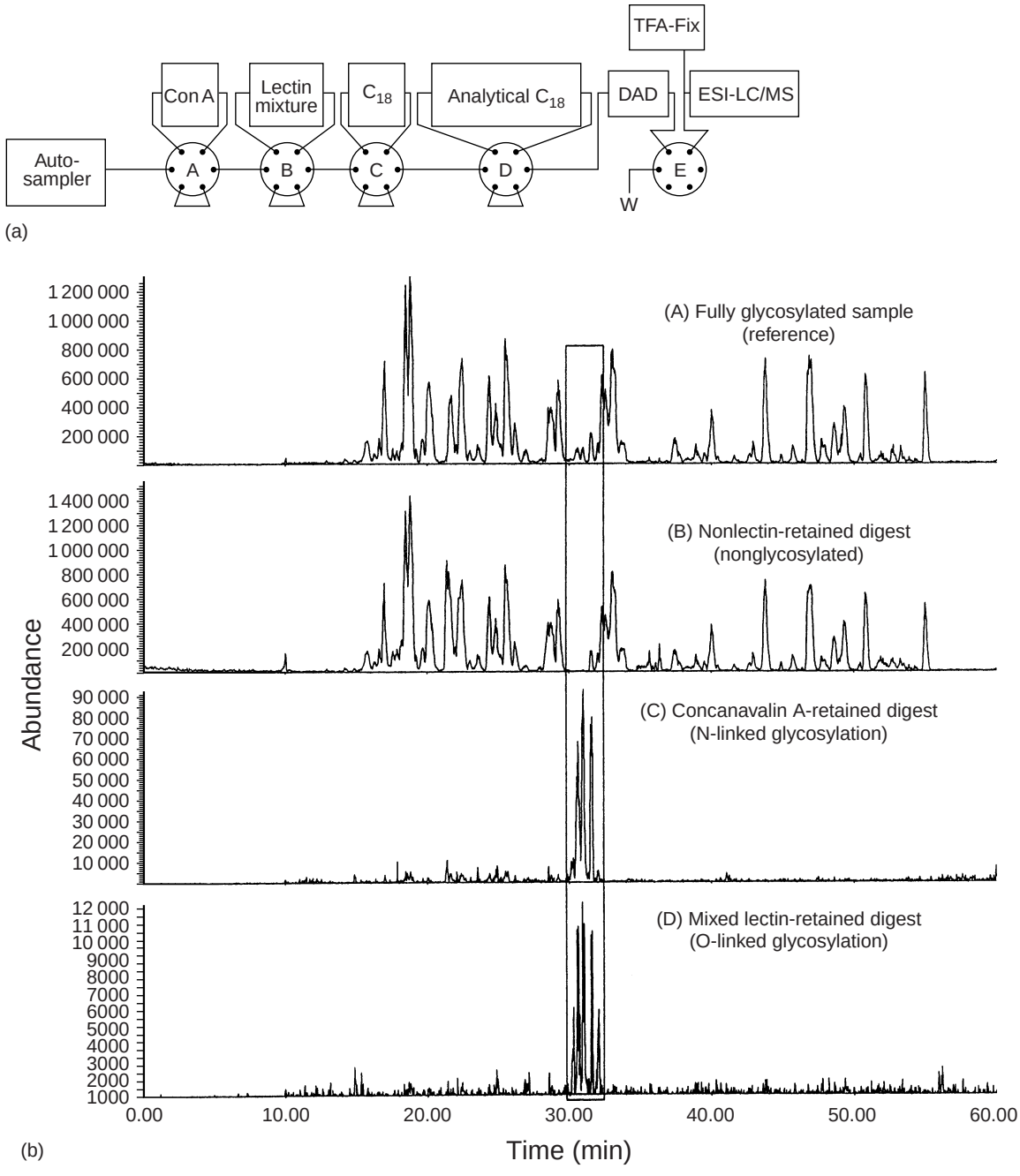

Figure 31 (a) Diagram of the multidimensional LC/MS system used for obtaining the glycopeptide maps shown in Figure 31(b). In the loop positions of three automatically acutated six-port injection valves (A–C) are situated precolumns (10 × 2.1 mm i.d.) with: A, immobilized concanavalin A (Con A); B, an immobilized mixture of four lectins (*Datura stramonium* agglutin, jacalin, peanut lectin, and amarathin); and C, a 5-μm silica-based C_{18} stationary phase. The precolumns are connected in series while loaded, then rinsed and desorbed individually onto the analytical column (250 × 2 mm i.d. C_{18}; 3 μm; 120 Å). A diode-array detector (DAD) was used to monitor the separation and an electrospray ionization (ESI) LC/MS instrument was used for final detection after propionic acid/isopropanol (TFA-fix) had been added post-column. (b) Electrospray LC/MS maps of glycopeptides from endoproteinase LysC digest of DSPAα1, a recombinant DNA-derived glycoprotein. The peptide maps are the following: (A) entire digest without affinity prefractionation; (B) nonglycosylated peptides retained on the C_{18} precolumn; (C) N-linked glycopeptides retained on the Con A precolumn; (D) O-linked glycopeptides retained on a mixed lectin precolumn. The changes inside the boxed region show different patterns with each precolumn. (Reproduced from A. Apffel et al., *J. Chromatogr. A*, **750**, 40 (1996) by permission of Elsevier Science B.V.)

is dependent on the immobilization chemistry, a feature predicted to be useful for studying interactions between coated silica grafted with NANA and insulin.[446] Caron et al.[447] studied the binding of biotinylated human brain lectin A to a ganglioside immobilized on a silica coated with DEAE-dextran with the aid of a streptavidin horseradish peroxidase conjugate detection scheme.

A charged carbohydrate ligand with a broad scope is heparin, a 5–30 kDa mucopolysaccharide composed of repeating α-L-dipyranuronic acid 2-sulfate and 2-deoxy-2-sulfamino-α-D-glucopyranose-6-sulfate dimers. It has been used for the separation of lipases, phospholipases and kinases,[448] growth factors,[449–451] steroid receptors,[452] DNA-binding proteins,[453] restriction endonucleases and RNA polymerases,[454] serum coagulation proteins,[455,456] lipoproteins (LPs),[457] heparin-binding proteins of smooth muscle cell membranes,[458] and selenium-containing proteins.[459] Immobilized heparin has two principal interaction modes – either as a biospecific affinity ligand, as in interaction with coagulation factors, or as a high-capacity cation exchanger, due to its sulfate groups. Elution with salt gradients is prevailing for both interaction modes. Björklund and Hearn prepared high-performance heparin affinity columns by reductive amination coupling of heparin to amino-functionalized silica[460] and discussed the pore filling problem associated with a head-on attachment of these relatively long polyelectrolytes. Ishimura et al.[461] have shown that the heparin-type specificity is not unique to heparin; a similar interaction pattern with heparin-binding proteins could be obtained on polymeric epoxy precolumns onto which they had attached a β-cyclodextrin sulfate, a known synthetic heparin substitute[462] (Figure 32). They also found that elution was easier from this material, and that the hydrolytic stability was better than for the heparin ligand. Heparin-like selectivity for human basic fibroblast growth factor was also seen by Anspach et al.[463] on a sulfonic acid tentacle-type cation exchanger, which had the advantage of being stable under cleaning-in-place conditions.

Nonsulfated cyclodextrins have also been used as affinity ligands, for the purification of glycogen-bound protein phosphatase G,[464] fungal[465] and recombinantly[466] produced α-amylases, β-amylase,[467] and cyclomaltodextrin glucanotransferase fusion proteins.[468] Elution is usually accomplished with buffers containing β-cyclodextrin.

5.4.3.10 Oligonucleotides and Related Compounds

Oligonucleotides and related compounds have been used for decades as affinity ligands for the isolation of DNA-binding proteins involved in transcription, replication, and recombination. The DNA initially used as ligands was nonspecific and derived from calf thymus.

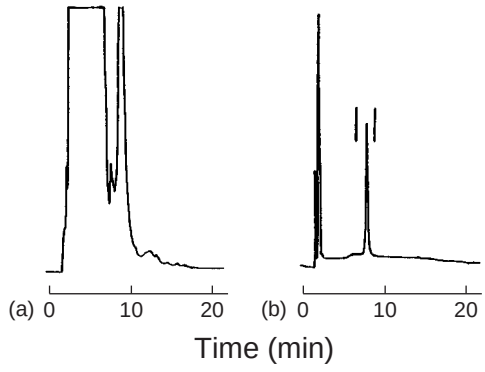

Figure 32 Reversed-phase HPLC separation of lysozyme after selective enrichment from chicken egg white diluted 1 : 9 in water (50 μL injected): (a) without precolumn; (b) with a β-cyclodextrin sulfate precolumn. Nonbinding proteins were eluted from the precolumn by washing with water for 5 min followed by two consecutive injections of 500 μL 20 mM TRIS-HCl buffer, pH 7.5 with 180 mM NaCl with a 2-min interval between injections. The precolumn was then backflushed into the RP system using an isocratic eluent of 35% (v/v) acetonitrile–aqueous 10 mM H_3PO_4. (Reproduced from K. Ishimura et al., *J. Chromatogr. A*, **769**, 213 (1997) by permission of Elsevier Science B.V.)

Modern DNA affinity chromatography uses specific deoxyoligonucleotide sequences that are synthesized and multimerized before attachment to agarose gels[469] or core-shell latex[470] supports. Quantitative zonal DNA affinity chromatography has been used to determine the equilibrium constants of proteins binding nonspecifically to DNA.[471] Jarrett et al. have developed high-performance DNA affinity supports based on silica and reviewed their use,[472] whereas Wheatley et al.[313] have used epoxide-activated HEMA resins and pointed out the advantage of using a high concentration of phosphate buffer when immobilizing double stranded DNA via a 5′-mercaptohexyl spacer. A dihydrazide activation chemistry that could be used to immobilize RNA was developed by Ruhn et al.[309] Yashima et al.[473] prepared a series of nucleic acid mimetic stationary phases by graft polymerizing the vinylic monomers poly(9-vinyladenine), poly(9-adenylethyl methacrylate), and poly(thymylethyl methacrylate) onto porous silica that had been activated with 3-trimethoxysilylpropyl methacrylate. The resulting columns were used for separating nucleosides and nucleotide dimers and oligomers according to an HIC-like mechanism (Figure 33).

Another nonantibody biomacromolecular interaction implemented on solid carriers[474] is the separation of mixtures of RNA and DNA on immobilized histones, basic proteins that bind strongly to DNA. By using a rather concentrated salt gradient, oligonucleotides can be fractionated into partly degraded and intact RNA emerging as two separate peaks at different eluent salt

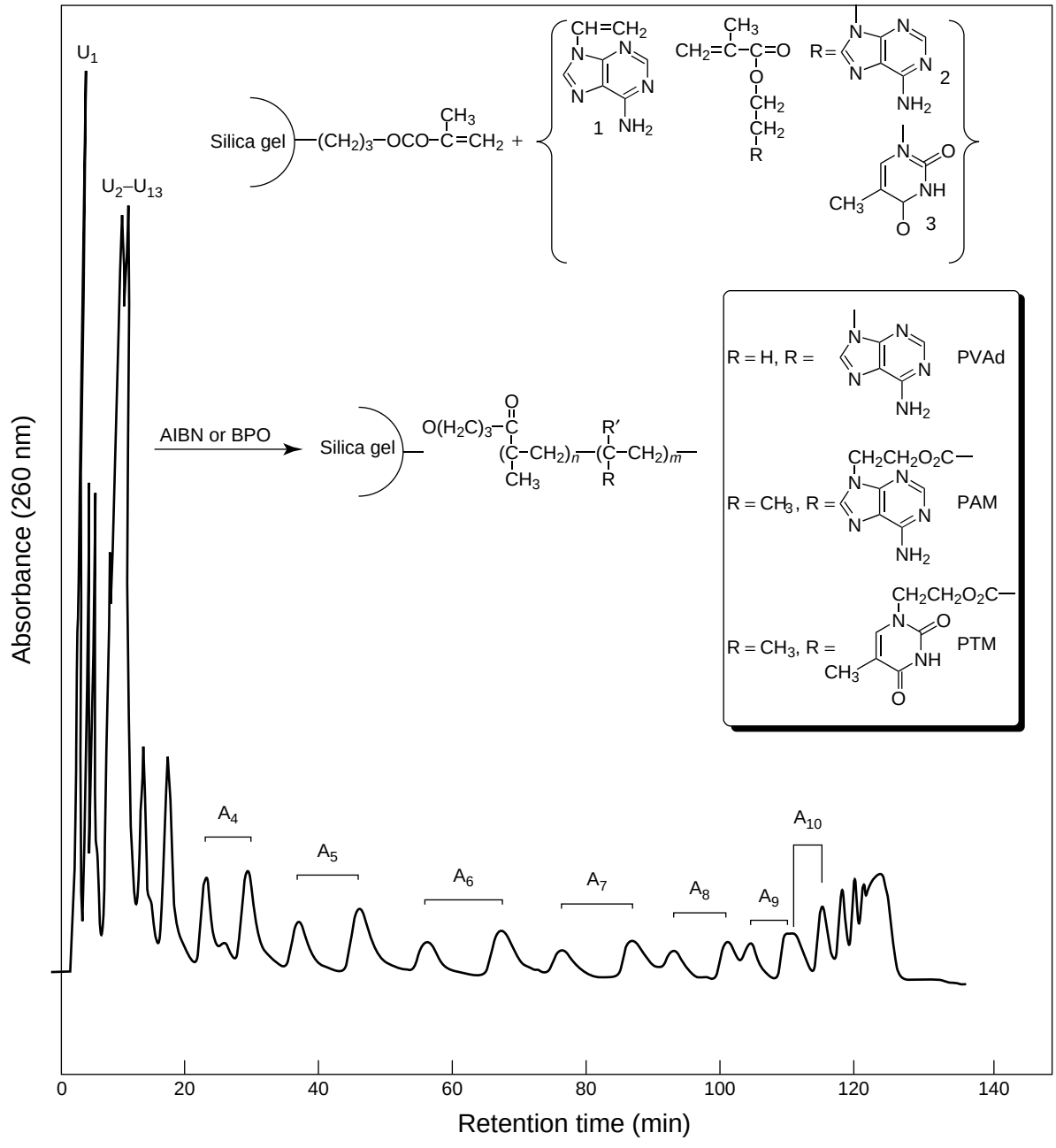

Figure 33 Separation of oligo(A) (DP = 1–12) and oligo(U) (DP = 1–13) mixtures (terminal 3′- and a mixture of 2′- and 3′-phosphates) on an Si-PTM (poly(thymylethyl methacrylate)) column (125 × 4.6 mm i.d.). Eluents: linear gradient 0–120 min from 1.5 M ammonium acetate (pH 7.0) (100% at 0 min) to 100% water at a flow rate of 0.5 mL min^{-1}; temperature 30 °C. The inset shows the synthesis routes to the grafted nucleic acid mimetic stationary phases produced in this work. (Reproduced from E. Yashima et al., *J. Chromatogr.*, **603**, 113, 117 (1992) by permission of Elsevier Science B.V.)

concentrations, followed by a single DNA peak at even higher eluent strength.

5.4.3.11 Screening of Combinatorial Libraries Although a large number of natural ligands have already been developed and evaluated, the quest for new ligands continues as the number of new targets grows rapidly. Libraries of both small molecules and mid-sized peptides have been prepared and screened for activity towards different biomolecules. This approach was primarily used in order to discover new drugs. However, these findings have also afforded new affinants for the preparation of powerful separation media for affinity chromatography of biopolymers.[475,476] Immunochromatographic screening

is based on immobilization of a receptor molecule that is then challenged with a combinatorial library of candidate compounds. The most retained compound is selected for further evaluation. This comprises their preparation in larger scale and immobilization on a support to afford a separation medium for affinity chromatography of the original receptor. Techniques such as introduction of receptors into in vitro biomembrane mimetic structures like liposomes have also started to emerge.[477]

5.4.3.12 Aptamers Combinatorial technologies developed during the 1990s have made it possible to isolate oligonucleotide sequences known as ''aptamers'',[478] which can enter highly specific affinity interactions with a host of target molecules in an antibody-like fashion.[479] These oligonucleotides are produced from random libraries by systematic evolution of ligands by exponential enrichment (SELEX),[480] based on repeated use of affinity chromatography with the immobilized target as ligand followed by enzymatic amplification. Because the PCR technique is used for multiplication of the evolved aptamers, this technique is a highly promising prospect for preparing synthetic biospecific affinity ligands at will. The first application of an aptamer as a ligand in affinity chromatography was published by Romig et al.[481] Using a DNA aptamer selected for its affinity to human L-selectin, they achieved good purification and recovery of a recombinant human L-selectin–Ig fusion protein from Chinese hamster ovary cell-conditioned medium. Aptamers have recently also proven their utility in biosensors,[482] and in the future we should expect to see more applications of the aptamer technology in affinity chromatography.[483,484]

5.4.3.13 Immobilized Metal Ion Affinity Chromatography IMAC[485] is a versatile separation technique for proteins and peptides based on their interaction with certain metal ions sequestered on chelating agents attached to sorbents, and is known under several other prefixes such as metal chelate, metal-ion interaction, and ligand-exchange chromatography. It occupies a position between biospecific affinity and more general techniques of low specificity such as ion exchange. The interaction principle of IMAC is somewhat similar to hydroxyapatite chromatography both in scope and applications. Both techniques are based on metal-ion affinity and are useful for separating and purifying structurally similar proteins.[486–490] Whereas hydroxyapatite is limited to a defined crystalline medium, with few possibilities of altering the interaction pattern, IMAC offers far more degrees of freedom by the choice of support type, chelating group, metal ligand, loading buffers, and eluent conditions. For detailed mechanistic aspects of protein separation by IMAC, the reader is referred to the literature,[425,491–497] and to the references dealing with new ligands and eluents below. The applications are numerous and disparate, and IMAC has been reviewed repeatedly.[498–505] Practical guides are also available in the literature.[505,506] This article will therefore concentrate on recent developments in separation media and studies aimed to gain a better understanding of the separation mechanisms.

The chelating groups present on the IMAC stationary phase are designed to bind metal ions known to interact with amino acid side chain functional groups that are exposed on the surface of proteins. Soft-to-borderline metal ions such as Cu^{2+}, Ni^{2+}, and Zn^{2+} (listed in their general order of decreasing interaction strength with proteins)[507] mainly target histidine, cysteine, and tryptophan, but to some extent also surface-accessible aspartic and glutamic acid residues.[508] The α-amino group of a peptide can also contribute significantly to the retention under certain conditions.[509–512] The binding of a protein is not related to a particular affinity for metal ions, but is dependent on the number and spatial arrangement of these amino acid residues on the protein surface. The use of spacer arms is therefore not as critical in IMAC as in modes where the ligand should reach into the reaction centrum of a protein. Numerous metal-ion chelating groups have been attached to sorbents for use in IMAC, and a critical factor is that the metal ion should be presented in such a way that it is accessible to the protein solutes. The classical chelating group that was introduced by Porath[485] is iminodiacetate (IDA). A chelating group with different properties, also introduced by Porath's group, is tris(carboxymethyl)ethylenediamine (TED).[513] Both moieties are illustrated in Figure 34.

A difference between these two groups is the number of coordination sites occupied by the chelating agent. Whereas IDA is a tridentate chelator, TED is pentacoordinate. This means that fewer coordination sites are available for interaction with the protein when TED is used as chelating group. A consequence is that metal

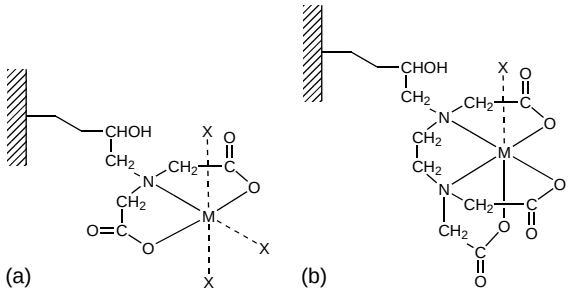

Figure 34 Two frequently used chelating groups for IMAC sorbents with metal ligands attached: (a) IDA and (b) TED. Both groups have been immobilized through bonding to a support activated by an epoxy group. Note the difference in available coordination sites on the metal ion.

ions sequestered with TED normally have a lower retention. This difference also reflects the kind of metal ions attached to the different chelating groups. Whereas IDA is typically used with divalent transition metal ions of lower coordination number, TED was originally developed for use with the group III ions Al^{3+}, Ga^{3+}, In^{3+}, and Tl^{3+}.[513] Ramadan and Porath have developed a sorbent based on glycinehydroxamate for use with Fe^{3+} as metal ligand,[514] although IDA has remained a popular chelating group for this metal ion. Hochuli et al.[515] introduced nitrilotriacetate (NTA) as a quadridentate alternative to IDA, especially suitable for hexacoordinate metal ions such as Cu^{2+} and Ni^{2+}, leaving two valencies vacant for biopolymer binding. The transition metal ion-mediated retention seen with dye affinity sorbents by Hughes et al.[379] should strictly also be classified as an IMAC-type interaction.

An IDA-functionalized silica for high-performance IMAC was developed and evaluated by El Rassi and Horváth.[516] Wirth et al.[297] developed a new silylation reagent carrying a precursor of the IDA group, which was used to prepare porous and nonporous silica with IDA chelating groups. By using another silylating agent in competition with the IDA silane for available silanols, the density of IDA groups on the surface could be varied, and the IMAC capacity of the porous sorbent leveled off when the coverage with IDA ligand was about 50% of the maximum obtainable. The importance of ligand density control was also demonstrated by Todd et al., in studies of the retention mechanisms of cytochrome *c* histidine variants constructed by site-directed mutagenesis on Cu^{2+}-loaded IDA ligands.[517] Changes in the number of accessible histidines on the protein surface influences the interaction strength, and when multiple interaction points are involved, the binding constant becomes strongly dependent on the availability of metal sites for interaction. The metal ion–ligand stoichiometry is also important; at sufficient chelating ligand density and with less than half the equivalent of metal ions, these become doubly chelated and inaccessible for interaction with the proteins.[517] Further studies on engineered proteins with imidazole added to the eluent have revealed that even a protein with only one surface histidine had an interaction pattern indicating coordination to multiple Cu^{2+} sites on the IMAC support at neutral pH.[518]

Several new chelating moieties have been developed in later years. One of these is a poly(vinylimidazole) with a complexing ability for Cu^{2+}.[519] Metalloprotoporphyrins with different metal centers on silica were found to have strong axial metal–nitrogen interactions with L-histidine, where the metal ion dependence follows the order $Fe^{3+} > Ni^{2+} > Cu^{2+} > Zn^{2+} \approx Cd^{2+}$.[520] These metalloprotoporphyrins phases are also involved in mixed-mode interaction with contributions from $\pi-\pi$ bonding, evident from retention for phenylalanine residues that are normally indifferent in IMAC. An advantage with the protoporphyrin-based chelating groups is the permanent bonding of the metal ion. A silica furnished with β-diketoamine groups by reacting 3-aminopropylated silica with 3-bromopentanedione has been loaded with Cu^{2+} and used for separating α-lactalbumin from milk whey.[521]

Hearn et al. have developed and evaluated a variety of chelating groups. The first in this series was an *O*-phosphoserine (OPS)-modified agarose that showed a stronger chelating capability compared to IDA for Fe^{3+} and Al^{3+} and a selectivity different from the conventional chelating agents.[522] Chelating groups based on 1,4,7-triazacyclononane (tacn) to which the borderline metal ions Cu^{2+}, Ni^{2+}, Zn^{2+}, Co^{2+}, Mn^{2+}, and Cr^{3+} were attached have been found useful for chromatography of several standard protein probes, as well as proteins in partially fractionated human plasma preparations (Figure 35). In connection with these borderline metal ions, the tacn-based sorbents showed a new selectivity toward proteins, compared to previous chelating groups based on IDA, NTA, or OPS.[523] In two recent studies, they make use of 2,6-diaminomethylpyridine[524] and *N*-(2-pyridylmethyl)aminoacetate[525] as chelating moieties in connection with Cu^{2+}. The different selectivities offered by all these new ligands refine IMAC as a separation principle, but in spite of their demonstrated

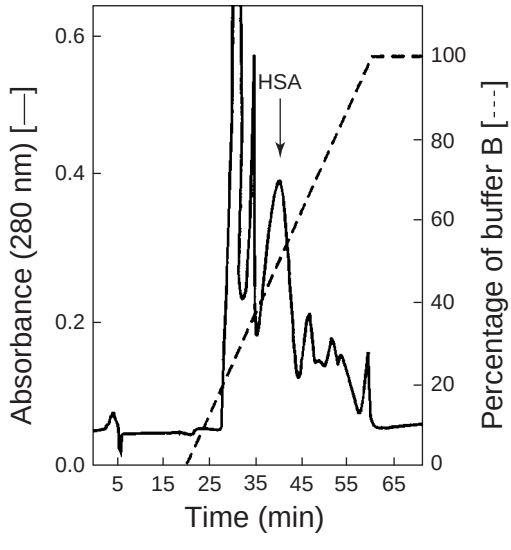

Figure 35 Elution profile for defibrinylated human serum protein preparation separated on an *im*-Cu^{2+}-tacn-Sepharose CL-6B sorbent using a salt gradient from 20 mM sodium acetate–100 mM NaCl, pH 4.0, to 20 mM sodium acetate–1 M NaCl, pH 4.0, at a flow rate of 0.5 mL min^{-1}. (Reproduced from W. Jiang et al., *Anal. Biochem.*, **255**, 53 (1998) by permission of Academic Press.)

utility, commercial availability is a problem with all the less-common chelating sorbents. Therefore, IDA remains the most widely used sorbent functionality and is available from several sources. Development of an IMAC separation method can be simplified by choosing as the initial system a commercially available chelating support with IDA groups and a strongly interacting metal ion such as Cu^{2+}. If elution becomes difficult, the less retentive metal ions Ni^{2+} and Zn^{2+} should be tested. If these systems do not work, consider alternative chelating groups and metal ions.

As in other interaction chromatographic modes based on polar functional groups, eluent molecules (solvent or buffer substances) will occupy the metal coordination sites in the absence of proteins interacting with these sites. Protein sorption is therefore connected with a metal ion–ligand exchange. It has to be realized that IMAC sorbents are inherently ionic and thus exhibit a conventional ion-exchange capacity,[495] in particular when the immobilized metal ion is a hard Lewis acid.[497] In order to harvest the specific interaction patterns offered by the lone pair Lewis base–Lewis acid ligand exchange mechanism, eluents should therefore always contain salt, and the salt concentration is not uncritical.[526] An implication is that IMAC can be used to separate proteins eluted from ion-exchange columns without intervening desalting, provided that proper adjustments of pH and salt concentration are made before the sample is loaded onto the IMAC column.

There are several ways to control the retention in IMAC and the easiest and most controllable way of eluting the adsorbed proteins from the column is through gradients of decreasing pH,[486,527] or increasing concentration of imidazol.[507] Samples are normally loaded from neutral or weakly alkaline buffers. The pH gradients operate by changing the degree of ionization of histidine and other interacting groups on the protein surface, whereas the imidazole gradients act by competing for the metal chelating binding sites. Gradient pH elution works with most retained proteins, unless the binding is too strong, in which case a less-retentive chelating group could be considered. Phosphorylated amino acids, phosphopeptides and phosphoproteins interacting with Fe^{3+}–IDA are exceptions,[528] where elution is accomplished by increasing the buffer pH or the phosphate concentration,[529] or by using phosphoserine or a displacing metal ion such as Mg^{2+}.[530] This specific type of interaction is particularly valuable for trapping low concentrations of phosphoproteins for MS analysis.[531] Displacement of proteins from Cu^{2+} complexes can take place by substituting ammonium for the loading buffer salt cation[486] (which should obviously not be ammonium ion or an amine such as TRIS). Imidazole is also a widely used eluent with Ni^{2+} and Zn^{2+} IDA chelates, where its advantage over pH gradient elution is that the metal ions remain on the sorbent during elution.[507] Elution by decreasing pH and increasing imidazole gradients produce similar elution pattern, as shown by Yip et al.,[532] who investigated the interactions of a large number of synthetic peptides (up to 42 amino acids long) on a polymeric high-performance IMAC sorbent with IDA groups and Cu^{2+}, Ni^{2+}, and Zn^{2+} as metal ions. A third way of eluting is to add a chelating agent such as ethylenediamine tetraacetic acid (EDTA) to the eluent. This procedure will also release the metal ion from the immobilized chelate, in which case the column must be reloaded before next use. However, leakage of metal ions does take place with most eluents, in particular if compounds having a strong interaction with the metal ligands such as glycine or primary amines are present in the eluent,[507] so a prudent practice is to always reload between runs, unless a very stable chelate is used in the column. A reloading cycle typically comprises stripping off remaining metal ions with a complexing agent such as EDTA at neutral pH, followed by reloading of metal ions onto the chelating groups on the sorbent.

The IMAC technique has become part of a unique biotechnological purification scheme through a technology known as ''His-tagging''.[533,534] This tagging takes place by extending the nucleotide sequence coding for a protein to be produced by recombinant technology by the codons for a chelating peptide, generating the gene for a chimeric metal-chelating peptide–protein that can be cloned and expressed. The metal-chelating peptide fusion handle is thereafter used for IMAC purification on an Ni^{2+}–NTA resin, whereafter it is removed by chemical or enzymatic cleavage so that the protein regains its natural amino acid sequence. Inclusion of a terminal sequence consisting of six histidines is most common,[534] although fewer histidines in controlled sequences provide for weaker interactions and possibilities of using softer elution conditions.[535–537] A technique related to His-tagging has recently been used to bestow IMAC retention to DNA produced in a PCR scheme.[538] The His-tag IMAC interaction has also been used for oriented immobilization of antibodies for immunoaffinity chromatography.[539] A recent paper furthermore describes the use of a His-tag in combination with anti-His-tag antibodies for a tandem affinity (IMAC and immunoaffinity) separation of His-tagged proteins.[540] An example of an IMAC separation based on his-tagging is shown in Figure 36.

Metal ion affinity can be used for reversible immobilization of native proteins to affinity-sorbent surfaces, as shown by the preparation of an affinity phase with variable and controlled amounts of Con A using Cu^{II} chelated to IDA porous silica.[541] Although a wide range of eluent conditions could be used without loss of the

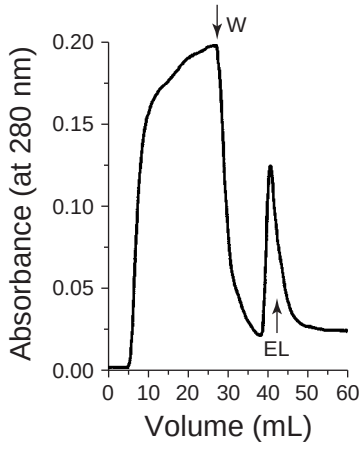

Figure 36 Anti-His-tag immunoaffinity chromatography of a His-tagged mini-antibody (a homodimerized single-chain Fv fragment with a His_6-tag fused to the C-terminus via a linker; sequence scFv-GGSGGAPGHHHHHH) after preliminary purification by IMAC. The affinity stationary phase contains an anti-His-tag antibody (3D5) immobilized onto an Affi-Hz hydrazide support. W denotes the start of washing buffer and EL the start of elution with 1 M imidazole, monitored at 280 nm. (Reproduced from K.M. Müller et al., *Anal. Biochem.*, **259**, 60 (1998) by permission of Academic Press.)

lectin, its detachment could still rapidly be accomplished by eluting with low pH or EDTA. The existence of dual functionalities produced a mixed retention mechanism, characteristic of both IMAC and lectin affinity chromatography.

A different kind of metal interaction chromatography takes place on organomercurial resins, whereupon nucleosomes containing transcriptionally active chromatin unfold and uncover the sulfhydryl groups of a histone, enabling its interaction with reagents having affinity for sulfhydryl groups.[542]

5.4.4 Salt-promoted/Thiophilic Adsorption Chromatography

Porath et al.[543] discovered in 1985 that thioether–sulfone gels, prepared by activation of agarose by divinylsulfone and coupling of 2-mercaptoethanol (Figure 37), have an enhanced affinity towards certain human serum proteins, in particular Igs. The retention appeared to be unique and quite different from hydrophobic interaction, and it works for all three major classes of Igs and their subclasses. This particularly mild technique, which was termed

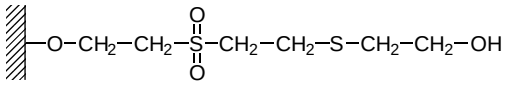

Figure 37 Thioether–sulfone functional group of Porath's original thiophilic gel.[543]

thiophilic adsorption chromatography (TAC), provides for efficient recovery of pure Igs at neutral pH from complex biological matrices.[544] Synthesis of surface-functionalized rigid material for high-performance TAC has also been described;[147,545,546] TAC has also been carried out in the membrane format.[547]

The TAC method is a typical on–off affinity technique. The sample is loaded onto the column in a 10–100 mM buffer at neutral pH, to which a 0.5–1 M solution of a cosmotropic salt, typically sodium, potassium, or ammonium sulfate, has been added. The buffer type does not appear to have any significant influence on the result. Examples of buffers are phosphate, borate, and zwitterionic aminosulfonic acids. Although the target is retained, the nonadsorbed molecules are washed away by the running salt-containing buffer. Elution occurs once the mobile phase is changed in a single step to the buffer solution only, or by a gradient of decreasing concentration of the adsorption-promoting salt.[548] Figure 38 shows a separation of mouse Ig subclasses.[147] This facile attachment and detachment of Igs furthermore allows TAC media to be used as immobilization sorbents in flow immunoassays.[549–552]

Apart from Ig purification, thiophilic adsorption sorbents have also found use as stationary phases for the separation of amylases[553,554] and for purification of allergens from horse sweat.[555] Experiments by Noel et al.[556] have shown that the presence of a thioether moiety is not essential to promote adsorption of Igs by thiophilic stationary phases. Instead, their ability to retain Igs appears to be connected

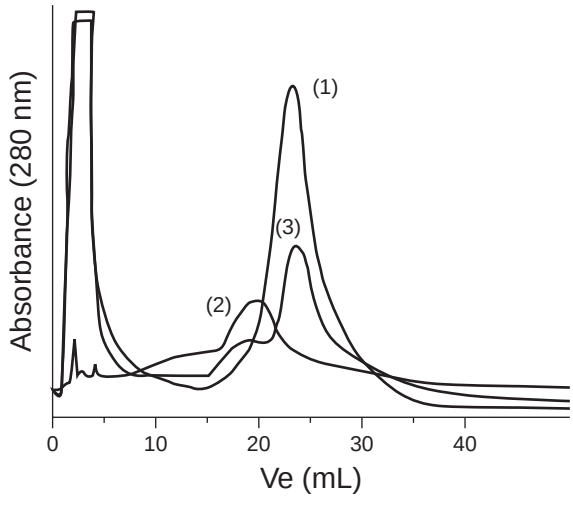

Figure 38 Elution profile of mouse ascitic fluid containing IgG3 (1), pure IgG1 (2), and their mixture (3) on a thiophilic column (125 × 4 mm i.d.) based on a DEAE–dextran-coated silica. (Reproduced from A. Serres et al., *J. Chromatogr. A*, **711**, 156 (1995) by permission of Elsevier Science B.V.)

with the divinylsulfone activation/cross-linking, making "thiophilic" a somewhat inappropriate name for what is in reality a salt-promoted retention effected by varying concentrations of cosmotropic anions (compare HIC). This observation was subsequently verified by others.[272,557,558]

Several new separation media have recently been synthesized, partly as a result of these findings, by attaching propylamine, allylamine, aminoacetonitrile, 2-aminocrotonitrile, 1,2-diamino-maleonitrile and tricyanoaminopropene to 1,4-butanediol diglycidyl ether and divinyl sulfone activated agarose.[559] In this study it was found that the nature of the spacer arm and cyano-substitution in the ligand influenced the affinity for proteins, and that the agarose activated by divinyl sulfone and reacted with tricyanoaminopropene had the highest adsorption for proteins. This ligand was consequently studied further and was found to have unique adsorption properties for proteins with binding being both dependent and independent of promotion by salt, which led to a proposed electron donor–acceptor adsorption mechanism for the observed interactions.[560] Results by Schwarz et al.[546] and Scholz et al.[557,561] show that heterocyclic ligands may be a viable route to selective sorption of Igs from solutions of low salt concentration. It thus appears that the salt-promoted "thiophilic" adsorption principle is not limited to antibodies, but is of a more general nature.

5.4.5 Multidimensional Affinity Chromatography

The high specificity of affinity chromatography, the relative insensitivity to variations in buffer conditions during the binding step, and the highly specific elution schemes make it possible to combine several affinity chemistries in a single separation. For instance, immobilized anti-HSA and protein A were bound to different silica columns coupled in tandem for capturing albumin and IgG from human serum.[562] Wheat germ agglutinin was used to separate sialoglycoproteins and the individual fractions were subjected to IMAC on a Cu^{2+}-loaded IDA silica.[563] A technique termed cascade-mode multiaffinity chromatography has been developed on traditional soft media to exploit the group-resolving power of affinity chromatography.[564] A column train consisting of (a) a Zn^{2+}–TDA IMAC gel, (b) a thiophilic gel, (c) Zn^{2+} chelated to dipicolylamine (DPA) agarose, (d) a hexyl-thioether agarose, and (e) Ni^{2+}–DPA agarose was capable of trapping >99% of the protein in human serum from the same high salt-containing buffer. Following adsorption, each column was connected to other column types for further separation in a minimized number of operations, so that essentially complete fractionation of serum protein was accomplished with high resolution, recovery, and capacity.

5.5 Ion-exchange Chromatography and Related Techniques

Ionic separation materials have been used in biopolymer separations for more than half a century and IEC (or electrostatic interaction chromatography as it should probably be called) has become one of the most widely used techniques for analytical and preparative separation of proteins.[565] It is also applicable to other charged biomacromolecules. The popularity comes from an excellent separation power, good possibilities of tuning the separations by selecting columns containing a variety of commercially available stationary phases with many different ionic functional groups, a soft interaction mode due to the buffered fully aqueous salt gradients typically used for elution, and a high sample capacity. The IEC technique is therefore often used in the early preparation stages where overloading is more likely to occur. Ion exchange is also often part of mixed-mode techniques (see below), where an additional separation dimension is allowed to contribute to the overall separation. Books dedicated to IEC are not abundant in the recent literature,[566] but chapters on IEC are found in most books on HPLC of biological macromolecules.[565,567–570]

The separation principle is based on electrostatic interactions between fixed ionic charges on the stationary phase and oppositely charged functional groups on the surface of the solute molecules. The retention is controlled by the charges of the stationary phase and the solute, and the ionic strength of the medium. Ions that are electrostatically attached to the ion-exchange groups are referred to as counterions, whereas ions of the same charge sign as the ion exchanger are termed co-ions. The fixed charges of the ion-exchange material can be either anionic, in which case the material is a cation exchanger, or cationic, giving the properties of an anion exchanger. Ion-exchange materials are also classified as weak or strong, depending on whether the ionic functional group has a dissociation or protonation equilibrium in the operational pH range, or maintains its charge when pH is changed. A compilation of the most common ion-exchange sorbent functionalities used for biomacromolecule separations is presented in Table 6.

Due to adjacent group effects and changes in solvation and cross-linking, there is a wide span in apparent pK values of individual functional groups in weak ion exchangers, manifested in shallow titration curves, often with steps. The pK values in Table 6 are therefore only approximate. This functional group heterogeneity affects weak anion-exchange materials in particular, where for instance materials based on cross-linked PEI[150] contain a mixture of primary, secondary, and tertiary amine groups, whereas DEAE materials contain tandem functionalities where one DEAE group is reacted on top of another

Table 6 Typical functionalities and key properties of the most common ion-exchange sorbent types for biopolymer chromatography

Ion exchanger type	Functional group	Representative functional group[a]	Designation[b]	p*K*[c]	pH range[d]
Weak cation exchanger (WCX)	Carboxyl	Support−COO^-	CM	≈5	6–13
Strong cation exchanger (SCX)	Sulfoalkyl	Support−$CONH-C(CH3)_2CH_2SO_3^-$	S	<1	2–13
Weak anion exchanger (WAX)	Diethylaminoethyl	Support−$CONH-(CH_2)_2N(C_2H_5)_2H^+ X^-$	DEAE	≈11	2–10
Strong anion exchanger (SAX)	Quaternary ammonium	Support−$CONH-(CH_2)_2N(CH_3)_3^+$	Q	>13	2–12

[a] The linkages are those often used on polysaccharide materials. Functional groups attached to synthetic polymers may be slightly different.
[b] Common abbreviation used by column manufacturers to identify the functional group.
[c] p*K* values of polymer-bound ionic groups are approximate and dependent on the functional group structure, the polymer matrix, steric factors, ionic strength, etc. Titration curves are typically shallow, which means that the net surface charge cannot be calculated directly from the p*K*.
[d] Operational pH range determined by the ionization equilibrium of the ionic/ionogenic group. Other factors may limit this range.

in a quaternizing reaction, leading to a relatively wide distribution in p*K*.[565] The titration curve of a weak anion exchange resin can thus take the shape shown in Figure 39.

From these curves, it is also apparent that the charge density of a weak anion exchanger is substantially influenced by the ionic strength of the medium, which typically changes during gradient elution. Optimization of separations on weak ion exchangers can therefore be complicated and strong ion exchangers (the S or Q types) are often preferred because they provide a more predictable retention through their lack of disturbing dissociation and protonation equilibria.[572] The retention at varying pH is thereby governed only by changes in charge of the solute. Another parameter affecting the retention is the charge density of the ion-exchange material, which is one of the factors controlling the selectivity.[149,573]

The versatility stems from the fact that practically all biomacromolecules, with asialic polysaccharides being the notable exception, have functional groups that are

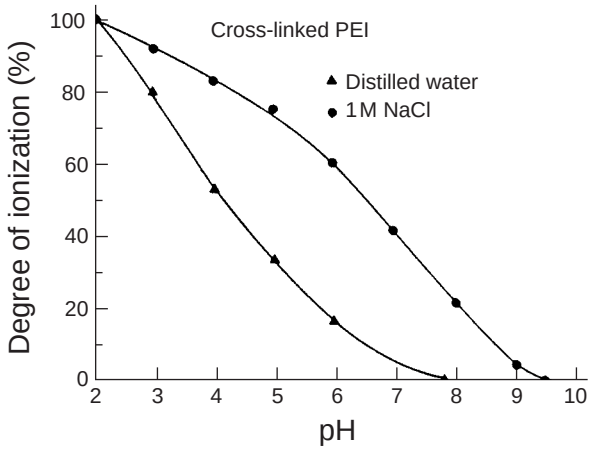

Figure 39 Ionization of a pellicular PEI weak anion exchanger as a function of pH and ionic strength. The curves are obtained by titrating with 0.1 M HCl in both distilled water and 1.0 M NaCl. (Reproduced from W. Kopaciewicz et al., *J. Chromatogr.*, **266**, 14 (1983) by permission of Elsevier Science B.V.)

charged, or can become charged through dissociation or protonation equilibria within the normal operating pH range of biopolymer separations. This sensitivity to changes in pH provides a key to tuning the selectivity in IEC and also provides for different IEC separation dimensions for the same biomacromolecule. A protein is characterized by its p*I*, where its net charge is equal to zero. The dissociation of acidic groups and protonation of basic groups on the protein surface produces a negative charge when the pH is increased above the p*I* value, and conversely results in a positive charge when the pH is lower than p*I*. A protein can therefore be separated on both anion- and cation-exchange columns, provided its p*I* is not extreme and the pH stability is sufficient. In principle, a protein should not have retention, neither on an anion exchanger nor on a cation exchanger, at its exact p*I*,[574] as illustrated in Figure 40.

The steepness of the titration curve is quite different among proteins, as is the p*I*, and even subtle differences, such as the substitution or deamidation of a single-surface amino acid, can cause a sufficient difference in charge properties to allow separation by IEC.[575] The charge properties are mediated by acidic and basic lateral groups of the protein, as listed in Table 7.

Basic proteins thus have a high p*I* and an excess of basic groups. Histones and chicken egg white lysozyme (p*I* 11.3–11.6) are among the most strongly basic proteins. Acidic proteins have an excess of acidic groups and thus low p*I*s. For instance, pepsin has a p*I* of close to 1. Most proteins have p*I*s lower than 7 and are therefore negatively charged at neutral pH. Carboxylate groups from sialic acid in sialoglycoproteins, γ-carboxyglutamate in blood coagulation factors, and phosphoserine in phosphoproteins also contribute to the charge of such proteins.[565] The charge properties of a protein are also dependent on the conformation. The addition of structure-altering agents such as solvents, detergents, water-soluble polymers, and chaotropic and cosmotropic ions can therefore selectively alter the retention of a protein in IEC.[565] Very rough estimates of the p*I* of a

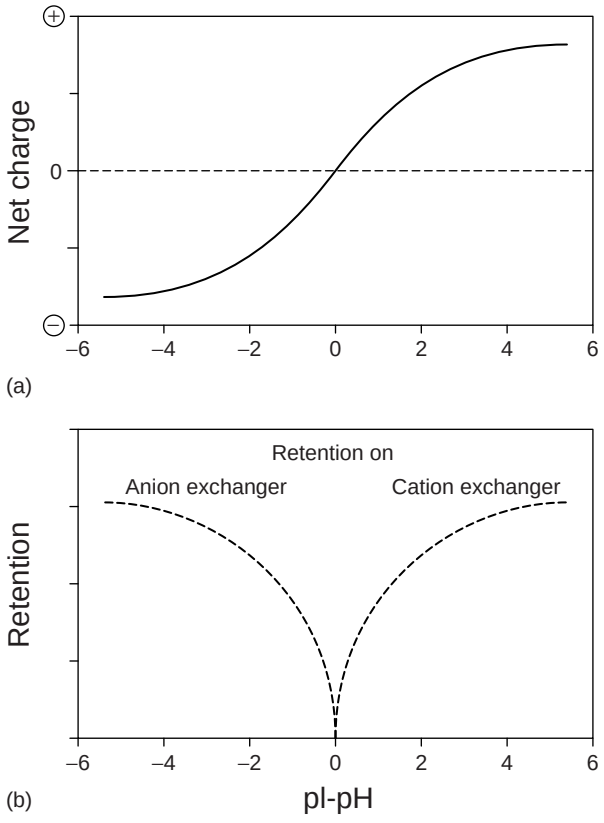

Figure 40 Diagram of (a) the net charge and (b) retention factor dependency on the mobile phase pH in relation to the p*I* of a protein, assuming that proteins are retained based on their net charge only. (Adapted from G. Choudhary, C. Horváth, *Methods Enzymol.*, **270**, 55 (1996) by permission of Academic Press.)

protein can be calculated[576] from amino acid side-chain dissociation and protonation constants if the primary structure is known. Without knowledge of the p*I*, a sound approach to IEC separation of an unknown protein would be to load the protein from a slightly basic solution onto an anion exchanger and use a salt gradient at this pH.

As just mentioned, a noncharged molecule should not interact with an ion exchanger through electrostatic mechanisms. However in practice many proteins do interact significantly with ion-exchange sorbents at their p*I*, causing a retention that is commonly attributed to an inhomogeneous charge distribution on the molecule surface.[571] A biomacromolecular solute with zero net charge can thus have charge patches of either sign, providing ''contact surfaces'' with the ion exchange material. For this reason it is often postulated in the literature that proteins have preferred orientations in their interactions with charged surfaces. Because ionic interactions act over relatively long ranges and through gradients that extend from the surface, this heterogeneous charge distribution can enable the molecule to get attached to both anion- and cation-exchange resins at its p*I*, using different orientations. The steepness of the titration curve appears to be involved in determining the extent of electrostatic interaction at the solute p*I*.[577]

A central parameter in the discussion of retention in IEC is the Z number,[571] which is used to visualize the number of charged sites involved in the interaction between the solute and the ionic stationary phase. The Z number is determined as the steepness of the relationship between $\log k'$ and the logarithm of the molality of the eluent obtained from a series of isocratic elution experiments. Because Z numbers usually assume rather high values, high resolutions can be obtained in IEC as a result of the reasons discussed above for hydrophobic separation techniques.

In its most simplistic form, the force between the charges on the sorbent involved in the interaction Z_s, and the corresponding charges on a biopolymer solute Z_p, separated from each other by the distance r, follows Coulomb's law (Equation 25),

$$F = \frac{Z_s Z_p}{Dr^2} \tag{25}$$

where D is the dielectric constant of the solvent. Ionic interaction forces are therefore weaker in water with

Table 7 Charged amino acid lateral groups and some other charge-promoting groups in proteins[565]

Group	Structure	pK_a		Occurrence in proteins (%)
		Proteins	Free amino acids	
Arginine	Guanidino	12.0	12.5	5.7
Aspartic acid	Carboxylate	3.9–4.0	3.7	5.3
Cysteine	Thiol	9.0–9.5	10.6	1.73
Glutamic acid	Carboxylate	4.3–4.5	4.3	6.2
Histidine	Imidazole	6.0–7.0	6.1	2.2
Lysine	ε-Amino	10.4–11.1	10.7	5.7
Tyrosine	Phenol	10.0–10.3	10.4	3.2
α-Amino	Amino	6.8–8.0	–	–
α-Carboxyl	Carboxylate	3.5–4.3	–	–

a dielectric constant close to 80, compared to typical water-miscible organic solvents used as eluents, whose dielectric constants are in the range 20–30. Concerning the retention mechanism for IEC when operated in fully aqueous solution, there are two conflicting views. The traditional retention model, which has its roots in the 1955 Boardman and Partridge's paper,[574] is based on the conception that a stoichiometric exchange of counterions takes place on the sorbent when the protein docks at the surface. This model, referred to as the KRFR model, was formalized and verified on experimental data over a limited concentration range by Kopaciewicz et al.[571,578] In a large number of subsequent works by many authors, this empirical model has served as a tool for modeling the retention within the eluent concentration and sorbent charge density ranges normally used in IEC, and has provided data that are meaningful in the context of pI and charge densities on protein surfaces. The KRFR model is based on the linear relationship that appears to exist between $\log k'$ and $-\log I$, where each protein has a characteristic slope, Z, interpreted as the number of ionic groups on the protein that are involved in the interaction process. The Z numbers are therefore dependent on the charge status of the protein and hence on the pH. The KRFR model has later been expanded to include the shorter range van der Waals interactions,[579] in order to account for retention observed on ion-exchange materials at high salt concentrations due to the hydrophobic effect.

A rather different view on the retention process was presented by Ståhlberg et al.,[573] who applied ideas from particle interaction in colloidal science to explain the retention in terms of long-range interactions based on averaging the charge of the protein and treating the sorbent surface as a charged flat slab (the SJH model). In terms of predicting retention, the main difference is that the SJH model forecasts $\log k'$ to be a linear function of the logarithm of the inverse square root of the ionic strength I, instead of the logarithm of the inverse first power of I, as predicted by the KRFR model. The first version of the SJH model could account only for the long-range electrostatic interactions. Subsequent refinements have also taken van der Waals interactions[580] and charge regulation theory[577] into account. The current model can couple the slope of the titration curves of proteins at the pI to retention at zero net charge, which can take place even on sorbents with the same net charge as the protein[572,581] (Figure 41). A mechanistic treatment of the KRFR model (the RUL model)[582] has been introduced, taking into account protein size and net charge and adopting a simplified form of the sphere–plate concept, on which the SJH model is based. Similar to the KRFR model, the RUL model explains $\log k'$ as being a linear function of $\log I$. However, according to Ståhlberg's recent review of retention models for IEC,[583] the RUL

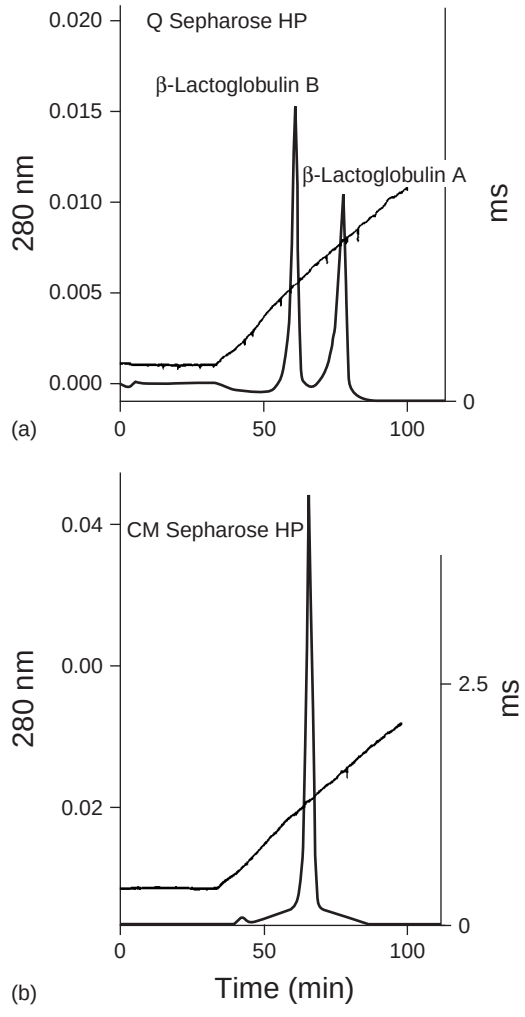

Figure 41 Gradient elution of β-lactoglobulin (Lg) A and B (p$I \approx 5.1$ and 5.2, respectively) near their pIs with (a) anion-exchange chromatography and (b) cation-exchange chromatography. The gels (see legend in Figure) were packed in glass columns (150 × 8 mm i.d.) and eluted with a gradient from 30 to 500 mM NaCl in 10 mM acetate buffer, pH 5.2, at a temperature of 25 °C and a flow rate of 1 mL min^{-1}. The gradient rate was chosen to accomplish baseline separation of LgA and LgB. (Reproduced from S. Yamamoto, T. Ishihara, *J. Chromatogr. A*, **852**, 33 (1999) by permission of Elsevier Science B.V.)

model is applicable only when the distances between the protein and the ionic surface are long. The SJH model is supported by studies of ionic retention on SEC materials with very low charge density,[190] a situation that cannot be treated by the KRFR model and its extensions. Yet at present the tradition model continues to prevail,[584,585] and as pointed out by Roth et al.,[582] the experimental data that were correlated on a log versus log square root plot to verify the SJH model can also be fitted to the KRFR model with a log–log plot.

Among the more interesting developments in IEC during the 1990s are tentacle gels and perfusion media. The

latter behave similar to other IEC resins, except that the mass transfer is improved, and are treated in a separation section of this text. Tentacle ion-exchange materials are produced by surface-initiated graft polymerization of linear oligomers combs, comprised of monomers with lateral ionic groups comprising 5–50 charged moieties per ligand unit, amounting to a length of up to ≈10 nm in the fully extended state.[163] These tentacle gels were found to absorb proteins to a degree which by far exceeds a monolayer coverage, and exhibit up to four times higher protein sorption capacity compared to corresponding surface functionalized materials.[88] A seemingly rational explanation for this high capacity is that the protein must be adsorbed in multilayers or dissolved within the polymer layer.[88] However, this cannot be contemplated in view of more recent theories of protein interaction with ion-exchange sorbents founded in colloid science, as discussed above. Recent experiments by Ratnayake and Regnier,[586] based on a series of tentacle-grafted cation-exchange sorbents synthesized with varying ligand density and tentacle length, have shown that the retention mechanism of tentacle ion exchangers is similar to surface functionalized materials. It was also concluded that some cross-linking exists in the tentacle layers prepared by free radical grafting, so that not all the polyacrylate chains are in fact tentacles. Another investigation has linked the higher capacity of tentacle gels to the increased flexibility of the ionic functional groups.[587]

Hearn's group has also conducted several studies on tentacle ion exchangers. When comparing tentacle type and conventional polymeric surface functionalized anion-exchange material,[588] they found that the retention followed a $\log k'$ versus $\log I$ relationship, and that some proteins were not, or only weakly, retained in spite of an eluent pH that was several units above the protein p*I*s. There were also significant differences between a silica-based and a polymer-based tentacle material used in these tests. In a study of salt and pH effects on band broadening,[589] the two types of tentacle materials tested behaved quite differently. These differences were attributed to a dynamic interplay between the tentacle ligands and the protein solutes. On studying the adsorption of different proteins in batch experiments with a silica-based tentacle cation exchanger,[590] it was found that most tested proteins were absorbed according to an ion-exchange mechanism. Lysozyme and insulin were the only exceptions, for which the sorption capacity increased with the concentration of salt. This odd retention behavior was interpreted as a hydrophobic effect. In a comparative study of two grafted strong cation-exchange materials[584] – a tentacle-type material and a polysulfoethyl aspartamide-coated silica – the Z numbers for the tested proteins were found to be higher for the tentacle gel, indicating that the flexible interaction layers interact with proteins through larger contact areas. The hydrophobic contribution to the retention also appeared to be higher on the polysulfoethyl aspartamide material. In a study of the bandwidth characteristics of the same sorbents at different temperatures,[591] band broadening induced by unfolding generally occurred at a higher temperature for the tentacle material. At lower temperatures, the tentacle material appeared to undergo morphological changes, which influenced the bandwidth of eluted proteins.

An advantage of tentacle gels that has been demonstrated in several difficult protein separations is the low hydrophobic interaction, which results from shielding of the underlying support by an extended layer of ionic polymer combs.[163] For instance, Rögner[592] showed that an early tentacle-type material provided the best separation power for membrane-protein complexes isolated from cyanobacterial thylakoid membranes with detergents above the critical micelle concentration (CMC). Fricke et al.[593] used the low hydrophobicity to purify casein-cleaving membrane proteinase on a strong anion-exchange tentacle material. Purification of these solubilized membrane proteinases could not be done with conventional techniques because the proteins bound irreversibly to all the other stationary phases tested. A high retention ascribed to self-aggregation and strong hydrophobic interactions was also seen with the tentacle materials, but this was solved by an unusual elution scheme with 2-propanol and a decreasing salt gradient. Tentacle materials have superior capabilities of resolving proteins with subtle differences. For example, a cation-exchange tentacle material baseline separated bovine, horse, and rabbit cytochrome *c* variants, native ribonuclease A and its deamidated Asp67 and isoAsp67 forms, and monoclonal antibody variants differing by a single lysine at the heavy chain C terminus, and hemoglobin variants[575] (Figure 42).

Furthermore, flexible ionic tentacles have a unique biomimetic resolving capability for proteins known to interact in vivo with biopolymers having regularly spaced charges. The heparin-mimicking capability of sulfonated tentacle materials has been demonstrated by Cacia et al.,[594] who separated recombinant human deoxyribonuclease I (rhDNase) variants differing in deamidation at a single residue, and by Anspach et al.,[463] who used the heparin-like structure for purification of recombinant human basic fibroblast growth factor.

Separation of proteins and peptides is the main application of IEC, but other charged biopolymers such as polynucleotides,[595] glycoprotein oligosaccharides,[596] and neutral carbohydrates can also be resolved by ion exchange, the latter by high pH anion chromatography.[567] Separations of proteins have been made using anion and cation columns in series and with a singular

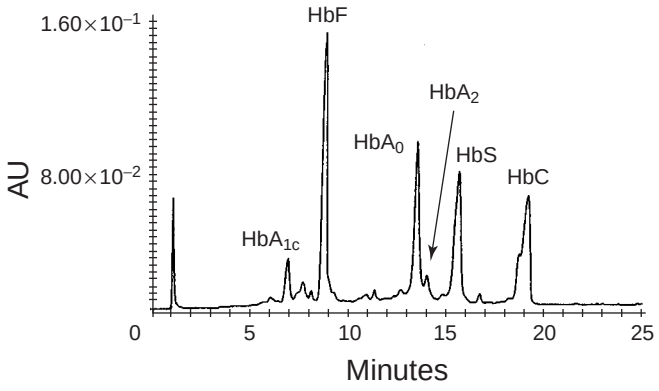

Figure 42 CEX chromatographic separation of four hemoglobin (Hb) variants (fetal, sickle cell, normal, and C hemoglobin) on a 250 × 4 mm ProPac SC-10 cation exchanger. Eluent A = 50 mM sodium phosphate and 2 mM KCN to pH 6.0; eluent B = 0.5 M NaCl, 50 mM sodium phosphate and 2 mM KCN to pH 6.0. Gradient = 0–50% B in 30 min at 1 mL min^{-1}. (Reproduced from M. Weitzhandler et al., *J. Chromatogr. A*, **828**, 370 (1998) by permission of Elsevier Science B.V.)

column packed with a blend of anion and cation exchangers.[597,598] However, these dual mode operations do not appear to be widely used.

Binding of biopolymers to the stationary phase takes place from a loading buffer that has its pH adjusted so that the charge of the proteins is opposite to that of the immobilized charged groups of the sorbent (below the p*I* for cation exchangers, and above the p*I* for anion exchangers). The salt concentration should also be sufficiently low to avoid elution of the solutes during loading. Sharpening of the injected band takes place at the top of the column and elution does not occur until the mobile phase composition meets the requirement for solute detachment. Conditions used to promote elution are either an increase in salt concentration or, less commonly, a change in pH towards the p*I* of the adsorbed solutes. A combination of both is also possible. Addition of organic solvents to the eluent can also be used to alter selectivity.[599]

The selectivity can to some extent also be controlled by the type of displacer ions. Kopacievicz and Regnier[572] investigated a number of different salts and classified the displacing capacity of eluent ions from weak to strong. They concluded that the best salts were those where the ions had intermediate eluting capacity, as they were able both to spread out the chromatogram and provide good protein recovery. The same group has also investigated the combined use of salts and solvents to affect retention in HIC and IEC.[260] Hearn et al. studied the effect of different displacing salts on retention and band broadening of proteins in isocratic anion chromatography.[600] The retention order was found to correlate well with decreasing hydrated ionic radii of the eluent anions and the retention was also affected by the type of cation, an effect attributed to preferential interactions with the protein solute. Band broadening was related to Hofmeister effects on the protein and on the double layer interface between the stationary and mobile phases. Malmquist and Lundell employed principal component analysis to find patterns in protein retention data using a large number of salts with cation and anion exchangers at different pH.[601] They concluded that there is a limited possibility of using salts to establish selectivity in IEC, but the most significant effects seen in their experiments were due to changes in the eluent strength of the starting buffer and the alteration of the apparent gradient slopes due to the use of buffer ions of different valence. They also point out that salt effects are more pronounced for late-emerging peaks, which are eluted in a higher salt concentration, and that chaotropic ions are the best candidates for establishing a unique salt-mediated selectivity. Due to the pronounced influence of pH on the charge state of biomacromolecular solutes, proper buffering is obviously very important in IEC. Apart from intended interaction between the ion-exchange surface and the biomacromolecular solutes under study, interactions between solutes and absorbed proteins already present on the surface, for instance due to fouling of the column, should be taken into consideration.[585]

5.5.1 Hydroxyapatite Chromatography

Hydroxyapatite, $Ca_5(PO_4)_3OH$, is a crystalline solid formed by thermal treatment of calcium phosphate[602] in sodium hydroxide solution.[603] The surface affords amphoteric interactions, mediated by the calcium ions and the oxygens attached to the phosphate atoms in the crystal, and the crystal surface offers interaction sites with different properties.[604] Carboxylic groups, basic moieties, and phosphate groups of various biopolymers interact with hydroxyapatite crystals, forming the basis for soft high-capacity multimodal separations.[605] The technique is best characterized as a steric IEC, made possible by the regularly spaced ionic binding sites at the crystalline surface of the sorbent.[606] The separation properties of hydroxyapatite are unique and offer an unparalleled selectivity that can be used to separate proteins that co-elute in other chromatographic modes. As the individual interactions are relatively weak, this method is not well suited for the separation of low-molecular-weight compounds that are only weakly retained. In contrast, the best results were observed for intermediate to large proteins and nucleic acids. The separation mechanism of hydroxyapatite columns, which seems to rely on the formation of complexes between proteins and ions in the mobile phase, was found to control the adsorption of proteins.[607] These processes

can also alter the ionic composition of the hydroxyapatite surface.[122]

The most important applications are separation of subtly different proteins such as antibodies and their fragments,[608] isoenzymes,[609] protein toxins,[608] histone complexes,[610] phosphoproteins,[565] LPs,[611] and solubilized membrane proteins.[612] Hydroxyapatite chromatography can also be used for the separation of polynucleotides such as t-RNA[613] and high-molecular-weight RNA,[614] and glycosaminoglycans.[615] Protein samples are loaded onto the column at a neutral pH from a low concentration phosphate buffer, and the elution is accomplished by a phosphate buffer with increasing concentration at neutral or slightly alkaline pH. Depending on the kind of groups taking part in the interaction and the number of such groups suitably exposed on the solute, the final buffer strength needed to elute the solutes typically varies from 50 to 500 mM. Low column temperatures should be avoided when the more concentrated phosphate buffers are used to prevent the salt from crystallization. The eluent acts by competing with the proteins for the calcium binding sites. Elution can also be effected by other calcium-binding anions such as citrates[565] and volatile ammonium hydrogen carbonate.[616] The common buffer salts such as NaCl or $(NH_4)_2SO_4$ have only a small effect on the retention. Therefore, protein mixtures isolated using other ion-exchange techniques in a NaCl gradient can be loaded directly onto hydroxyapatite columns without previous desalting.[565] A protein separation on a spherical hydroxyapatite column is shown in Figure 43.

Conventional hydroxyapatite suffers from limited mechanical and chemical stability and these properties do not improve even after sintering.[617] The search for mechanically stable materials with hydroxyapatite features had resulted in porous silica functionalized with diol groups and covered with a thin hydroxyapatite layer.[618] Another alternative to hydroxyapatite is ceramic fluoroapatite,[619] which is more stable than hydroxyapatite and has similar chromatographic properties. However, it does not appear to be widely used. The advantageous selectivity of hydroxyapatite has also prompted the investigation of less closely related materials such as lead phosphate hydroxyapatite,[620,621] various metal metaphosphates,[622] and deformed nonporous agarose coated with ferric oxyhydroxide,[623] or aluminum and zirconium (hydr)oxide,[624] onto which phosphate was stoichiometrically bound. All these materials show some features similar to hydroxyapatite, but also large differences. Calcium hydroxyapatite from bones was also evaluated.[15] A different use of the protein-retentive power of hydroxyapatite is in deproteination steps, as part of the sample preparation procedures in the analysis of small molecules.[625]

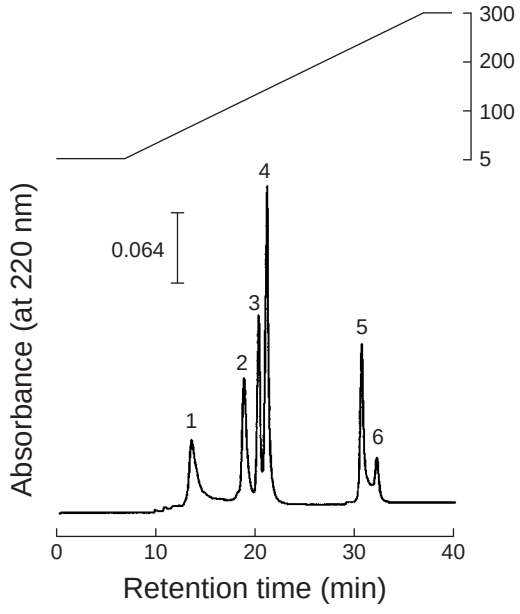

Figure 43 Separation of standard proteins on a column (135 × 7.8 mm i.d.) packed with spherical hydroxyapatite using a 30-min linear gradient from 5 to 300 mM potassium phosphate buffer (pH 6.8) at 1 mL min^{-1}, and UV detection at 220 nm. Peaks: 1 = BSA, 2 = myoglobin, 3 = ribonuclease A, 4 = lysozyme, 5 = reduced cytochrome *c*, 6 = oxidized cytochrome *c*. (Reproduced from F. Honda et al., *J. Chromatogr. A*, **813**, 27 (1998) by permission of Elsevier Science B.V.)

5.6 Chromatofocusing

Sluyterman et al.[626,627] introduced in 1978 a protein separation technique that can be viewed as a hybrid of IEC and isoelectric focusing. Chromatofocusing, as the technique is called, combines high capacity and high resolution, and has an added benefit of providing an estimate of the p*I* of the eluted proteins,[628,629] which elute at a pH that is determined primarily by their p*I* and secondarily by charge microinhomogeneity on the protein surface.[630] Chromatofocusing is based on the loading of proteins onto a weak anion-exchange column of appropriate buffering capacity that has been equilibrated with a buffer, defining the upper end of the pH gradient to be explored. The proteins are thereafter eluted isocratically with a solution of an oligomeric polyampholyte buffer (a so-called polybuffer), which carries a set of buffering moieties designed to give an even buffering capacity over the separation pH range. The pH of the polyampholyte elution buffer is lower than that of the equilibration buffer and defines the lower end of the separation pH interval. During elution, the polybuffer is titrated by the weak ion-exchange groups on the stationary phase, and pH gradients are thereby formed along the column, both in the mobile phase and on the separation phase. The pH of the eluent decreases in a nearly linear fashion as the

elution proceeds. In general, the starting pH should be above the p*I* of the proteins to be separated (usually 0.5–1 pH unit higher), although this may not always be necessary. Best separations are achieved with shallow pH gradients, which are accomplished by a small pH difference between the equilibrium buffer and the elution buffer, low elution buffer concentration relative to the capacity of the ion exchanger, or a comparatively long separation column. It is essential that the salt strength of the eluting buffer is kept low, so that conventional ion-exchange elution by salt displacement is kept at a minimum. Resolution down to 0.01 p*I* unit differences can be achieved with commercial high-performance chromatofocusing (HPCF) media.[630] A diagram of the gradients developed along an HPCF column during elution with a polybuffer is shown in Figure 44.

Chromatofocusing is one of the mildest of the separation techniques based on pH gradients, due to the fact that proteins are kept in a buffering environment and are not exposed to extreme pHs during the run. In this lies also a risk of aggregation or precipitation of proteins on the column, because the protein surface charge at p*I* is very low, and the colloidal stability of the biomacromolecular solute is hence at its lowest. However, compared to isoelectric focusing, chromatofocusing provides good possibilities of using additives to prevent protein aggregation. Urea, glycerol, ethylene glycol, zwitterionic components, and nonionic detergents can be added at high to moderate concentrations, whereas ionic additives have to be limited to a total concentration of 20–30 mM to avoid salt-induced elution.[630] Another advantageous property of chromatofocusing is high capacity, since a large fraction of the column is used concurrently in the separation process. The separation efficiency is comparable to IEC, but HPCF can be done in shorter time on comparably sized materials.[631]

The columns used for separation are either purposely synthesized to contain a mixture of tertiary amine and quaternary ammonium groups, or anion-exchange columns based on polyethyleneimine. Establishment of chromatofocusing as a high-performance separation technique was done by Hutchens et al.[632] who used PEI-functionalized silica anion exchangers for analytical and preparative separation of estrogen receptor isoforms at good resolution and recovery. Coincidentally, Fägerstam et al.[633] were working with polymer-based mixed strong/weak anion exchangers for fast high-resolution chromatofocusing of human serum proteins, and for studying protein microheterogeneity. Chromatofocusing does not necessarily have to be carried out on anion exchangers. Monkarsh et al.[634] used, for example, a preparative strong cation-exchange sulfopropyl silica column in a quasichromatofocusing set-up with a gradient of acetate/phosphate buffers at increasing pH to resolve two 154-amino-acid fragments of recombinant human interleukin-1α differing by deamidation of one Asn (p*I* 5.1 and 5.3, respectively). In this 75 mL min^{-1} eluent flow rate preparative separation, the use of inexpensive buffer components represents a significant saving

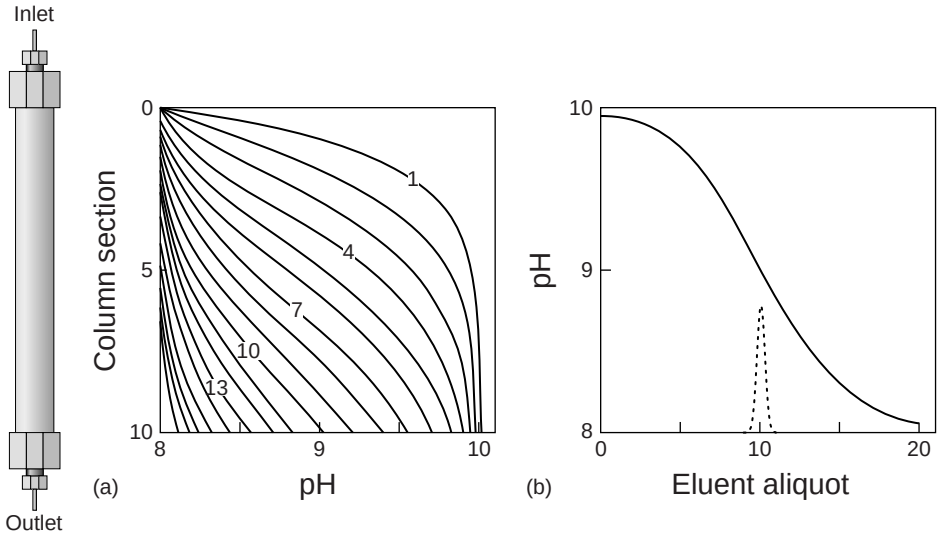

Figure 44 Diagram of the evolution of the mobile and stationary phase pH gradients in chromatofocusing at a starting pH of 10 and a focusing buffer of pH 8. Each curve in (a) represent a gradient inside the column after the stated numbers of eluent aliquots have been pumped. The intersection of these curves with column section 10 (the bottom of the diagram, representing the column outlet) is used to construct the pH profile at the column outlet, as shown in (b). The peak shown as a hatched curve on (b) depicts a protein with p*I* (i.e. elution pH) 9. (Reproduced from L.A. Sluyterman, J. Wijdeness, *J. Chromatogr.*, **150**, 19 (1978) by permission of Elsevier Science B.V.)

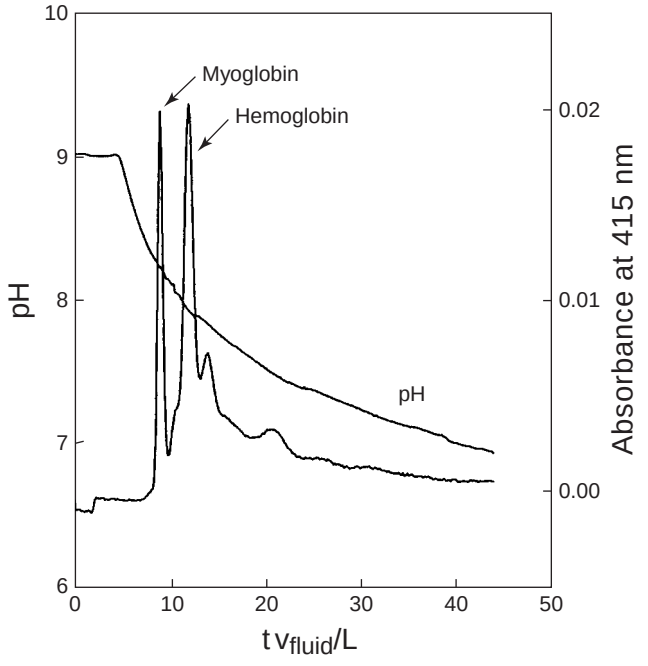

Figure 45 Separation of horse myoglobin and bovine hemoglobin on a PBE 94 column presaturated with 0.02 M DEA at pH 9. The eluent was a mixture of 20 mM DEA, 2 mM TRIS and 4 mM imidazole, pH 7, at 0.75 mL min^{-1} flow rate. (Reproduced from R.C. Bates, D.D. Frey, *J. Chromatogr. A*, **814**, 53 (1998) by permission of Elsevier Science B.V.)

over polybuffers. Carboxylic acid weak cation exchangers are also common.

As just hinted, the main drawback of chromatofocusing has been the expensive polyampholyte eluents and special columns that are marketed as a package solution. Furthermore, oligomeric ampholytes are not unproblematic, and unexpected and unwanted interactions have been observed with some protein separations (for a review see Hutchens[630]). Efforts have therefore been made ever since the early days of chromatofocusing to substitute the oligobuffers with simpler buffering components and to use conventional ion-exchange columns for chromatofocusing. Hearn and Lyttle[635] were among the first to evaluate common amphoteric and nonamphoteric buffers in mixtures as development solutions for chromatofocusing on weak anion and cation-exchange columns. Gradients were indeed formed, but their shapes were strongly influenced by the pK_a values of the individual buffer components and did not describe the desired smooth and linear pH changes as a function of elution volume. This work was followed up by Hutchens et al., who developed[636] and applied[637] several multibuffer systems containing up to 34 different buffering substances, and were able to achieve practically linear gradients through the nearly continuous distribution in buffer pK_a, from pH 3.5 to 8.1.

All these studies have been based on internal pH gradient formation, i.e. the isocratic elution buffer and the weak ion-exchange resin are titrated by each other to produce pH gradients both in the mobile phase and on the resin. Results presented by Liu and Anderson[638,639] show that the pH gradient elution technique (termed external pH gradient formation by Sluyterman and Elgersma[626]) can produce linear gradients at the column exit with a weak anion-exchange column using only a few buffer components. Frey et al.[640] developed optimized buffer systems which did not require a buffering weak anion-exchange resin to form a retained pH gradient, and also showed that they could be used for generating quasilinear internal gradients with weak anion-exchange materials[641] (Figure 45).

Linear gradients are necessary when the sample contains many components that are to be separated but, in preparative chromatofocusing, the desired separation can often be achieved with a buffer system showing a few distinct transitions in the right pH intervals. This can also alleviate problems with tailing peaks. Such step gradients can be produced by an internal pH gradient formation with few retained buffer components on a strong ion exchanger.[642,643]

A practical guide to chromatofocusing has been written by Li and Hutchens.[644] Hutchens has also discussed the theory and applications of chromatofocusing, and listed commercially available stationary phases.[630]

5.7 Mixed-mode Interactions in the Chromatography of Biomolecules

Due to their heterogeneous nature, biological macromolecules rarely behave in a chromatographic system as if their retention was controlled by a single separation mode, for which the partitioning between the mobile and stationary phase is expected to change monotonically as a function of changes in the concentration of displacing agent. In reality, the retention is governed to a large extent by factors related to mobile phase interactions, such as conformation changes, precipitation, and ion pair formation. Examples of unintentional multimodality and ways to circumvent this have been discussed in the section on SEC, where enthalpic interaction should be completely avoided, and in HPLAC, where the biospecific interactions should ideally be the only retentive mechanism in operation. The extent of multimodal interaction depends both on the separation mode and on the nature of the solute, and will often result in concave dependencies of retention on the displacer concentration, as illustrated in Figure 46.[81]

These curves illustrate that minimum retention is only obtained in a certain range of displacer concentration. At concentrations below this point, retention is determined

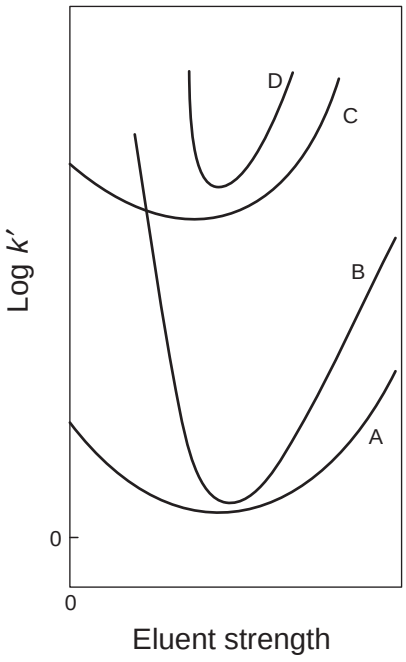

Figure 46 Diagram of typical relationships between retention factor and displacer concentration in HPLC involving multimode interactions. The concave shapes of the elution curves are due to the interplay between the different factors. Fastest elution is achieved in a retention window where the combination of intermolecular interactions between the solute and the mobile and stationary phases favors the solvated state, rather than adsorption to the stationary phase. Curves A to D are explained in the text. (Reproduced from M.T.W. Hearn, C.T. Mant, R.S. Hodges (eds.), *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, 105–122, by permission of CRC Press.)

by the eluent strength in a conventional sense, as set forth in Equations (18–20). The increase in retention beyond the minimum is due mainly to interactions taking place in the mobile phase, resulting in aggregation of the solutes and their precipitation on the stationary phase. In reality, most biomolecule separations are therefore achieved by means of mixed partitioning and precipitation chromatography.[645] The shapes of the four curves in Figure 46 are explained as follows.[81] Shape A is typical of small polar peptides separated under RP or ion-exchange conditions, and is characterized by a shallow shape according to Equation (18) or (19). The plot has a relatively small k' value as the displacer concentration approaches zero. Retention of these compounds resembles that of small molecules, which can usually be separated isocratically over a wide range of eluent concentrations. Shape B is typical for many polypeptides and fibrous and globular proteins in RPHPLC and in HIC, as well as for most separations in IEC. The most salient feature of shape B is the narrow elution range, which efficiently precludes isocratic elution and makes the separation sensitive to the sample volume and matrix. The retention plot of curve C also exhibits a shallow dependency on concentration, similar to curve A but with a substantially higher value of $\log k'$ at the minimum. The shallow flanks are indicative of a low number of contact points, each having a strong interaction. Curves of type C are typical of hydrophobic peptides under RP conditions, certain proteins in IEC, and some HPLAC and HIC schemes. The most problematic case in represented by shape D, wherein the minimum is sharp and located at high k' value. Therefore, the elution is difficult and low recoveries can be expected. This behavior is characteristic of many membrane proteins in RPHPLC and IEC, and their separation in RPHPLC may require addition of zwitterion pairing agents or detergents in IEC.

If gradient elution is used under sorption conditions described by curves B or D, good recoveries call for the use of relatively shallow gradients, as the solute is capable of traversing the column only within the gradient segment where its retention is at its lowest. If the eluent composition is changing too rapidly, the solute will simply be trapped by the flanks on either side of the eluent concentration of minimum retention, as it is overtaken by the rapidly changing gradient. This will effectively reattach the protein to the stationary phase before it has a chance of being eluted. For that reason, the separation chemistry should be chosen so that k' at minimum retention is unity or less. Once the optimal concentration has been found, recovery can be further improved by applying a more shallow gradient or an isocratic plateau around the critical peak. With Figure 46 fresh in mind, it should be noted that separations showing minima in retention versus eluent strength provide two possibilities of gradient elution: in the conventional fashion, i.e. by increasing the eluent strength (which means increasing eluent concentration except in HIC); and by a retrogradient starting from the opposite eluent composition.[81]

Mixed-mode interactions have been intentionally exploited in order to establish unique selectivities.[646] For instance, Mant et al.[223,224,647–649] have used cation exchangers with hydrophilic supports, operated with organic solvents and salt as eluents, to radically alter resolution between peptides on a single stationary phase (Figure 47). Hirata et al.[650] obtained unique selectivities for peptides on a poly(vinyl alcohol) column with residual carboxylic groups using aqueous acetonitrile as eluent. Floyd and Hartwick used a hydrophobic ion exchanger to separate oligonucleotides in mixed mode,[651] and also presented a review of mixed-mode interactions for separation of biomolecules.[652]

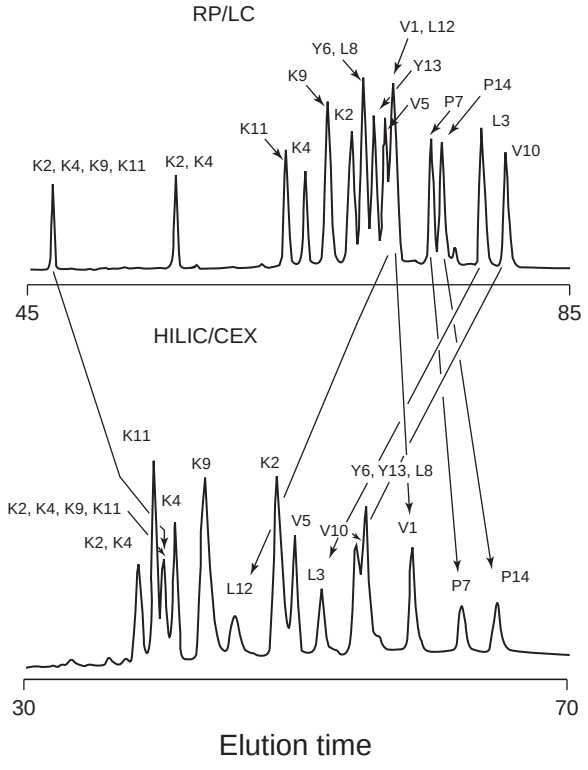

Figure 47 Comparison of RP and HILIC/cation-exchange (CEX) elution profiles of cyclic peptides. RP conditions: linear A–B gradient (0.5% acetonitrile min^{-1}) at a flow rate of 1 mL min^{-1}, where eluent A is 0.05% aqueous TFA, pH 3, and eluent B is 0.05% TFA in acetonitrile. Column temperature 70 °C. HILIC/CEX conditions: linear A–B gradient (2.5 mM $NaClO_4$ per minute, following 5 min of isocratic elution with 100% buffer A) at a flow rate of 1 mL min^{-1}, where buffer A is 20 mM aqueous triethylammonium phosphate (TEAP), pH 3; buffer B is buffer A containing 400 mM $NaClO_4$; acetonitrile is present in both buffers at concentrations of 90% (v/v) in buffer A and 80% (v/v) in buffer B. Column temperature 30 °C. Peptides were detected by their absorbance at 210 nm in both modes. (Reproduced from C.T. Mant et al., *J. Chromatogr. A*, **816**, 88 (1998) by permission of Elsevier Science B.V.)

5.8 Multidimensional Chromatography – Coupled Column Systems

As mentioned above, an orthogonality in separation dimensions is usually necessary to achieve good separations of complex mixtures. Furthermore, as proteins may be thermodynamically unstable but kinetically stable to the chromatographic conditions, separation time and manual handling of solutes are factors working against good overall recoveries. Coupling several separation steps together in a multidimensional system is therefore rational in the separation of biomacromolecules. The distinction between multidimensional and multimode chromatography is sometimes confusing. Multimode chromatography is based on the use of a single separation material with more than one separation principle in operation, a mixture of sorbents in the same column, or columns with different separation dimensions plumbed tandem and flushed by the same eluent.[516] Multidimensional chromatography involves the sequential use of columns with different selectivities in separate runs, operated with a heart-cut or a flush technique to transfer a select portion of the compounds eluting from the first column into the second separation step, a concept that is described by Regnier and Huang.[240]

As opposed to small molecules, where quite incompatible solvents and separation materials have to be used in the different separation steps required to achieve orthogonal selectivities, it should be more facile to couple separation techniques for biomacromolecules because most separations are carried out in fully aqueous eluents. However, whereas coupled-column techniques are frequently used in, for instance, environmental HPLC analyses,[653] surprisingly few automated multidimensional HPLC systems have been described for separation of biomacromolecules. A couple of recent examples are coupling of size-exclusion and anion-exchange chromatography in the analysis of poly- and oligosaccharides,[654] and a process application utilizing an immunosorbent with immobilized protein A connected in series with a quaternary ammonium anion exchanger.[655] An automated system has also been demonstrated for protein primary structure determination, comprising immunoaffinity chromatography, automatic buffer exchange, proteolysis with immobilized trypsin, on-line concentration/desalting, and microbore LC/MS.[656] A few more examples are given in the peptides section.

Timing of eluent transfer to the second column is critical in multidimensional chromatography. It requires experimentally determined retention times and a dedicated instrumentation with exact timing and low dead volume values, and quite complex systems have been designed for use in automated purification, desalting, hydrolysis, and RPHPLC.[657] A persisting problem is drift in retention times of the primary column due to ageing and fouling. But by exploiting the steep elution curves typical of most biological macromolecules, it is often possible to find a binary gradient composition on the first column, so that the retentions of the solutes to be transferred are altered from near infinite to practically nil. Such transitions are usually more dependent on mobile phase conditions than on the state of the separation column. This enables reproducible, efficient and nearly quantitative transfer of the solutes onto the head of the second column where the separation material and the conditions should be the cunning trap that is required to re-establish a sharp injection band.[240] This entrapment is also favored by the U-shaped elution curves, where the minima seldom coincide for different separation chemistries. A particularly promising

field for multidimensional techniques is in combination with MS analysis. Commercial systems with such capabilities have started to emerge.

5.9 Slalom Chromatography

A rather odd separation principle, slalom chromatography,[658] was discovered some years ago, but does not seem to have gained widespread use. Although performed on HPLC equipment, it can in a strict sense not be categorized as a chromatographic technique because the separation takes place entirely in the mobile phase with the packing acting only as a flow-perturbing element. Solutes applicable are large double-stranded DNA molecules (up to 50 kbp), which are separated by means of columns packed with spherical particles similar to those used for high-performance aqueous SEC. The obtained separation is based on molecular size, but smaller molecules are eluted ahead of larger molecules (Figure 48), which is contrary to the selectivity obtained with SEC materials. This inverted retention is due to a hydrodynamic phenomenon occurring predominantly in the mobile phase, whereby larger molecules are delayed because they find it less facile to negotiate optimal flow passages between the particles of the column packing.[659] The technique has a slight similarity to hydrodynamic chromatography,[660] a technique that exploits the nonhomogeneous hydrodynamic fields that exist near the packing surfaces in hydraulically driven flow systems due to the viscous forces. Selective enrichment of larger macromolecules occurs close to the stationary phase due to directed migration transverse to the flow direction, and the technique is used for the separation of high-molecular-weight synthetic polymers and other colloidal particles.

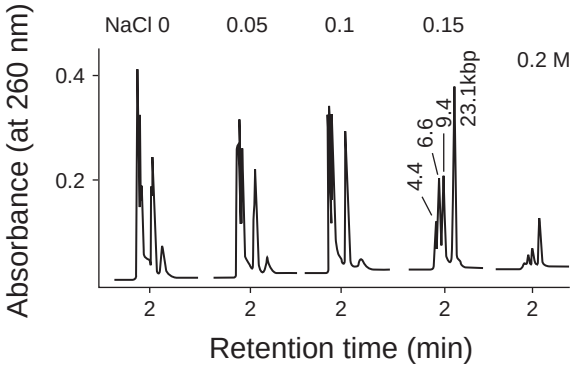

Figure 48 Slalom chromatographic separations of λ/*Hind*III fragments (0, 0.1, 0.05, 0.15, 0.2, 4.4, 6.6, 9.4 and 23.1 kbp) on a Capcell-Pak C1 column. Eluents: 10 mM sodium phosphate buffer, pH 6.8, 1 mM EDTA with NaCl added to 0, 0.05, 0.1, 0.15, or 0.2 M concentration (left to right); flow rate 1.0 mL min^{-1}. (Reproduced from J. Hirabayashi, K. Kasai, *J. Chromatogr. A*, **722**, 137 (1996) by permission of Elsevier Science B.V.)

The dependence on a shear force makes the separation in slalom chromatography reliant on the flow rate of the solution acting as the hydrodynamic selector, i.e. no separation is achieved at low flow rates, with all solutes eluted in the column void volume, and the separation improves steadily as the flow rate in increased. There is also a temperature dependence opposite to normal, with better separations as the temperature is lowered.[659] This technique thus deviates from conventional principles in chromatography in virtually all respects.

The choice of stationary phase appears to be of limited importance,[659,661] but the separation is dependent on minimal partition interaction with the solutes. The particle size determines the molecular size window of the separation[659] – smaller size range DNA molecules are better separated on small diameter packings, whereas larger particles provide improved resolution of large DNA molecules. The solutions used to bring about the hydrodynamic separation are comparatively low ionic strength salt solutions – either purely aqueous, or with a low admixture of acetonitrile.[661] A chapter on slalom chromatography is included in a recently published monograph.[41]

The principle involves accumulation of macromolecular solutes on the surface of the material, and the risk of precipitation should therefore not be neglected, although for molecules with pronounced polyelectrolyte behavior such as double-stranded DNA the risk should be small. It should also be kept in mind that at high flow, the shear forces may reach levels where very large solutes can start to fragment.[662]

6 DETECTION TECHNIQUES

Analytical chromatographic separations would be futile unless a detector device that can provide quantitative (and preferably also qualitative) information is available for monitoring substances eluting from the column. The choice of detector is more important than it might appear at first. It should not only be sensitive enough to detect the eluted species in the required concentration interval, but also comprise an integral part of the overall selectivity of the method. Accordingly, detectors based on a variety of principles have been developed for use in HPLC, and their principles and characteristics have been covered in several monographs.[663–666] A variety of often conflicting criteria determine the applicability of a detector. The sensitivity should be sufficient, with a linear response spanning the entire analyte concentration range. It should give stable signals and show a good reproducibility at a low noise level. The response should be fast and the internal volume minimal with a flow geometry optimized

to minimize band broadening. Low sensitivity to changes in eluent temperature and RI are important, particularly in gradient elution. When hyphenated systems are used, or when the separated compounds are collected for further analysis, is should be verified that the detector does not alter the solutes on passage through the detector. The reliability and user-friendliness, i.e. the level of expertise required to operate the detector, are also very important in real laboratory settings, so that the operator can easily identify a malfunction and understand how to correct it.

The detector should either give identical response to all solutes, or a highly selective response towards one or more classes of solutes. As mentioned, the detector is an important part of the overall selectivity optimization of an analytical separation. If the detector is sufficiently selective yet has a general response, the chromatographic separation process could in principle be disposed of, because the detector would be capable of both resolving and detecting the analytes in a non-chromatographic process. Examples of such detectors are mass spectrometers, notably matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOFMS), described elsewhere in this encyclopedia.

Detectors are usually classified according to their response pattern as general or selective. A general response has a wide scope, as a single detector can be used to detect many different compounds, but also requires that the compounds presented to the detector are completely resolved. A distinction is often made between bulk property and solute property detectors, where a bulk property detector measures the changes in a bulk property of the eluent, such as RI, as a solute elutes. Characteristic of bulk property detectors is their low selectivity and sensitivity, but this is not always a drawback. Worse is their sensitivity to variations in eluent composition, which effectively precludes their use in gradient elution. Solute property detectors are designed to measure changes in a property expressed by the eluted solutes, but not by the eluent. By tuning the detector to a common property of the solutes, it is then possible to determine these in the presence of an eluent of changing concentration. A selective detector is usually tunable, whereby a discrimination of co-eluting compounds may be possible.

Most detectors are spectroscopic, i.e. they measure the interaction of solutes with electromagnetic radiation. Different wavelength regions provide access to different kinds of information and offer unique possibilities to tune the selectivity. Quantitative information is denser in some spectral regions, whereas other regions offer higher absorptivity and thereby better sensitivity, and freedom in the choice of solvents. These factors explain the persistent use of UV spectrophotometry, although the information density provided by electron excitation in solution is rather meagre. A universal feature of proteinaceous matter is the peptide bond, which absorbs strongly in the far UV, below 220 nm. A narrow window for general detection thus exists in the wavelength range 210–220 nm, below which noise and solvent absorption tends to become troublesome. Disulfide bonds absorb in the 250 nm region, and the aromatic side-chains of phenylalanine, tyrosine, and tryptophan absorb in the wavelength range 250–300 nm. The absorption band of nucleotide bases lies in the vicinity of 260 nm. Carbohydrates lack chromophores in the accessible UV region. When several absorption bands are available, the longer wavelength absorptivity is often weaker. However, the actual signal-to-noise ratio, as determined by the presence and variability of spurious peaks in the vicinity of the solute (chemical background noise), is usually better because nonsolutes rarely absorb UV light at long wavelengths. Furthermore, intensive low-wavelength UV radiation can cause bond dissociation and trigger other photochemical processes that alter the solutes on passage through the detector.

DADs are absorbance spectrophotometers capable of measuring the entire UV wavelength range simultaneously in real time using integrated photodiodes as sensing elements.[665,667] Their price is low when considering the flexibility and the amount of information provided, and modern DADs are good alternatives to conventional single-wavelength UV detectors. There is a principal and important difference between multiwavelength DAD and conventional single-wavelength UV spectroscopic detectors, evident in Figure 49. In a single-wavelength detector, the eluate is illuminated by a single wavelength and the transmitted light is measured by a single photodetector. In order for multiwavelength detection to be feasible in a DAD, the sample has to be illuminated with light of the entire spectral range being measured (usually 190–380 nm), which means that photochemically sensitive substances are prone to undergo photochemical reactions[668] on their passage through the detector. In a purely analytical setting this may not be troublesome, unless artifact signals are produced from irreproducible bleaching or formation of new compounds in the detector cuvette. However, photochemical processes must be considered if the separated compounds are collected for further investigations.

Fluorescence detection is based on emission of light following photoexcitation and is a native property of some biological molecules, notably polynucleotides and aromatic peptide lateral groups. Fluorescent properties can also be introduced in molecules by derivatization (see below). As opposed to spectrophotometry, where the measurement is based on log transformation of subtractively measured absorbed light, fluorescence is at low solute concentration a linear function of the solute concentration and the intensity of the incident beam. The

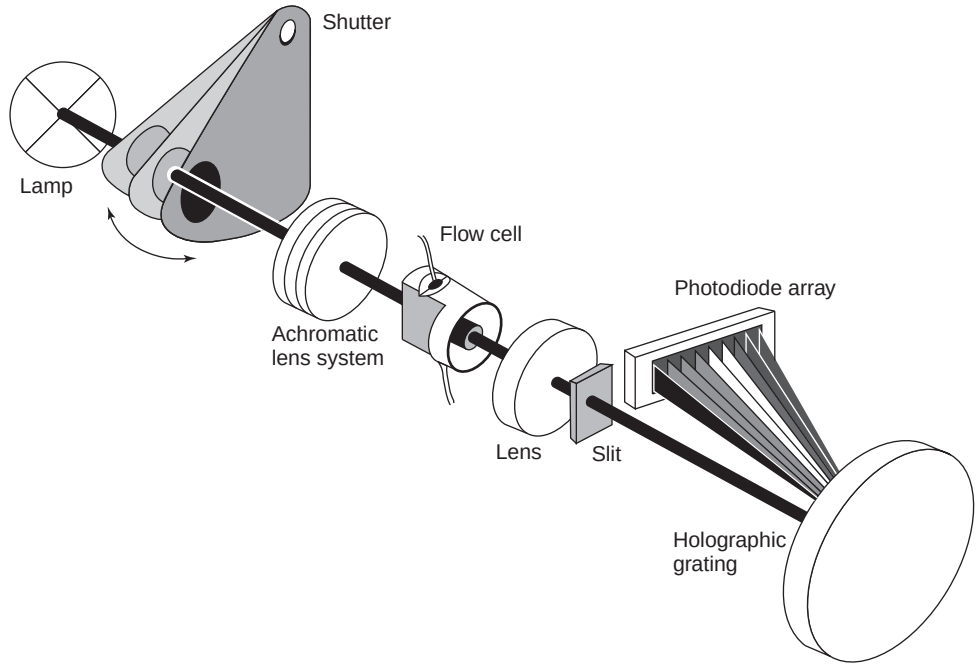

Figure 49 Optical and flow cell components of a single-beam diode array detector using reversed optics, meaning that all the light from the lamp passes the flow cell before reaching the wavelength-dispersive element, where it is separated into its spectral components and projected onto the photodiode array. (Reproduced by permission of Hewlett-Packard Company.)

emitted light is collected perpendicular to the excitation beam and the need to extract the signal by subtraction is thereby eliminated, enabling sensitive and selective detection. Selectivity is further enhanced as few molecules absorbing light are capable of fluorescing. A suitable entry point to this literature can be the recent review of Tao and Kennedy on laser-induced fluorescence detection in microcolumn separations.[669]

Chiroptic detection based on measurement of circular dichroism (CD), i.e. the difference in absorption of the clockwise and counter-clockwise rotated vector components of a polarized light beam passing through a medium containing a chiral component, has been used for a long time in HPLC.[670] The CD principle is also widely used for determining the secondary structure of proteins, because of the chirality of amino acids and the spontaneous organization of polypeptide chains in either left- and right-twisted α-helices and β-pleated sheets, or as a random coil, producing a characteristic optical rotation and CD spectra. Thus, CD detectors can provide information on the conformation of eluted proteins.[671] The main drawback of CD detectors is their limited sensitivity, which restricts their use to separations carried out on high-capacity materials, where the concentration of protein is relatively high.[672]

Vibrational spectra carry more information than the electronic excitation spectra obtained in the ultraviolet/visible (UV/VIS) region. However, infrared spectroscopy detection is not a viable alternative, because the solvents used for the separation of biomacromolecules are not transparent in the infrared region. Raman spectroscopy has been used sporadically as a detection principle in HPLC. For instance, purine bases have been detected by on-line surface-enhanced Raman spectroscopy after mixing with a silver sol.[673]

Modern high-field nuclear magnetic resonance (NMR) spectroscopy provides an extremely powerful tool set for characterization of biological macromolecules and their interactions. However, although direct coupling of NMR to HPLC has been successfully utilized for structural elucidation of small molecules,[674] the long time required for obtaining spectra for large molecules has prevented the use of LC/NMR for on-line characterization of eluted biomacromolecules. When NMR spectroscopy is used in combination with HPLC of biomacromolecules, eluted peaks should therefore instead be collected and subjected to off-line spectroscopy.

RI detectors measure changes in the index of refraction of the eluent induced by the solute. In principle, all solutes affect the eluent RI, but the sensitivity varies, and compounds lacking absorption or other properties enabling selective detection can often be detected by RI detection. The drawback of RI is that it is a bulk property of the eluent, which makes the detector signal very sensitive to eluent composition and effectively precludes gradient elution. As the majority of all separation schemes

for biomacromolecules require changing eluent composition, RI detection is usually impossible. An exception is high-molecular-weight carbohydrates, for which separation often takes place in SEC and where the lack of chromophores has been a persisting problem. Differential RI detection is thus still used, often hyphenated with other measurement principles. Recent developments have alleviated the temperature sensitivity and improved the detection limit substantially.[675]

Related to RI detection, although with a radically different sensitivity and degree of sophistication, is surface plasmon resonance, a principle used in the leading biosensor system. The small biosensor detection volume makes it suitable for use as an HPLC detector,[676] in particular in affinity chromatography.[677,678] This and other biospecific detection principles have been reviewed.[679]

Due to their size, biomacromolecules are good scatterers of light in solution, which forms the basis for the low-angle laser light scattering (LALLS) and MALLS principles that have been introduced to LC detectors in recent years. Both of these measurement principles can provide information on the molecular weight of macromolecules, thus supplementing external calibration in SEC. The additional information provided by multiple measurement angles in MALLS also provides information on the radius of gyration, so that characterization of both the molecular mass and the shape of biomacromolecules can be carried out on line.[680] An example of a SEC separation with MALLS and differential viscosity detection is shown above (Figure 16). A commercial detector is also available where the scattered light is measured at right angles (RALLS). This detector is combined with an RI detector and a viscosity detector based on measuring the differential pressure across a detector cell containing multiple capillaries for determining association and rheological phenomena of high polymers.

A different scattering principle is used in evaporative light scattering detection (ELSD), where the signal is based on light scattered from a dried aerosol produced by spraying the eluent into a desolvating chamber and evaporating the solvent and other eluent components. Nonvolatile solutes form microparticles that are detected in the gas phase as they pass the detection cell. The technique has become particularly popular in the separation of lipids and related substances, and for carbohydrates. An entry point to the literature can be a recent review of its use in combinatorial chemistry.[681]

Many electrochemical detectors exist, but the mode particularly interesting for biological macromolecules is pulsed amperometric detection (PAD),[682] because it is capable of detecting polar aliphatic compounds containing alcohol, amine, and thiol groups that are not easily detected by photometric or fluorimetric techniques. The PAD technique uses a pulse sequence to adsorb compounds on a platinum or gold electrode surface and then selectively oxidize the deposited compounds, followed by a cleaning sequence where adsorbed material is oxidatively removed from the surface prior to the next pulse sequence. The strongly adsorptive properties of noble metal electrodes, which are problematic under nonpulsed conditions, are thereby mastered and turned into an advantage and the PAD technique has become the preferred non-MS detection method for detecting carbohydrates.

Almost general yet highly selective, MS is close to fulfilling all requirements of an ideal detector. Its extremely high sensitivity coupled with a multitude of ionization and fragmentation techniques makes it capable of providing sufficient information to resolve the structures of complicated biological macromolecules. The ease of using a mass spectrometric detector has improved vastly, and it is now an accessible technique even outside the most sophisticated laboratories, which makes it a serious competitor to the techniques listed above. However, the cost of buying and operating an MS detector is still high compared to the alternatives. Traditional detectors will therefore still exist alongside MS for a long time where their performance is adequate. The LC/MS combination forms an extremely powerful tool for structural elucidation of biomacromolecules, a topic dealt with in other parts of this encyclopedia.

6.1 Labeling Reactions for Introduction of Detectable Properties

Intrinsic properties such as UV absorption, native fluorescence, or electrochemical activity form the basis of most detection schemes in HPLC and sufficient detection limits can usually be obtained with substances in their native form. However, in trace work, the sensitivity achievable by the use of intrinsic properties can be insufficient, and it may therefore be necessary to resort to derivatization, i.e. enhancing the detectability of solutes by introduction of a detectable property through chemical reactions. These reactions can be carried out either before the sample is injected (precolumn), or after the sample has been separated (postcolumn). Derivatization techniques targeting functional groups present in biomacromolecules have been described in several books.[668,683–686] Two recent reviews by Krull et al.[687,688] focus on derivatization and detection of proteins, Koller and Eckert have reviewed derivatization of peptides,[689] and Boppana and Miller-Stein gave an outline of pre- and postcolumn derivatization techniques for sensitive detection of peptide drugs.[690] A sensitive separation of peptides with ninhydrin postcolumn derivatization is shown in Figure 50.

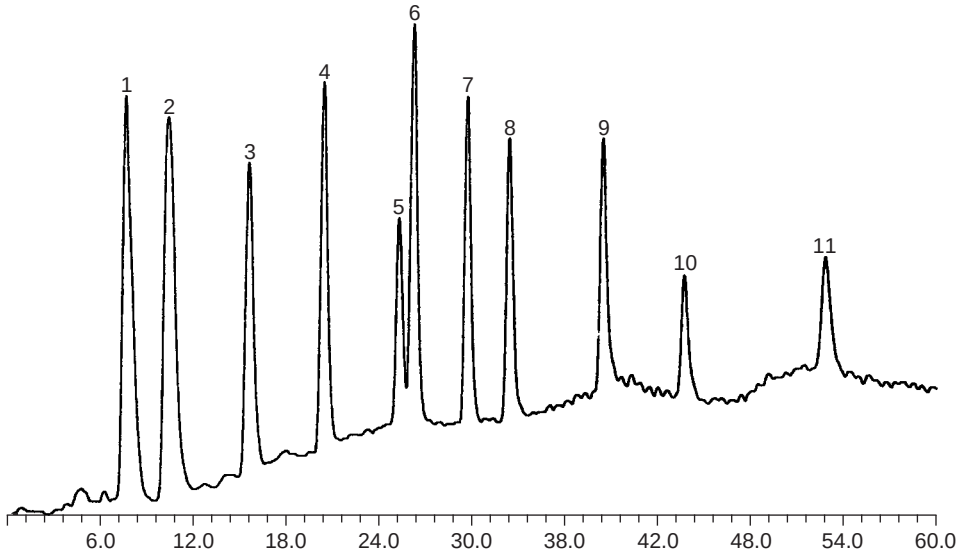

Figure 50 Endogenous peptides separated on a C_8 silica column (250 × 2.0 mm i.d.; 5 μm) and detected by postcolumn reaction with ninhydrin. Mobile phase: monochloroacetate buffer (65 mM, pH 3.2) with 10 mM sodium octanesulfonate and methanol (55 : 45 v/v initial; following injection, the methanol was increased to 70% in 30 min and held for 30 min). Temperature 60 °C. Peak identities and amounts injected: 1 = desglycylamide vasopressin (39 pmol), 2 = Arg-vasopressin (41 pmol), 3 = luteinizing-hormone releasing hormone (18 pmol), 4 = angiotensin II (20 pmol), 5 = bombesin (13 pmol), 6 = bradykinin (8 pmol), 7 = angiotensin III (22 pmol), 8 = angiotensin I (16 pmol), 9 = atrial natriuretic peptide (7 pmol), 10 = ACTH (6 pmol), 11 = growth-hormone releasing factor (4 pmol). (Reproduced from V.K. Boppana, C. Miller-Stein, *Anal. Chim. Acta*, **352**, 67 (1997) by permission of Elsevier Science B.V.)

Precolumn derivatization uses polar functional groups in the solute to attach a tag, i.e. a chromophore, a fluorophore, an electrophore, or a moiety within a characteristic fragmentation pattern in MS. The derivatized analyte is then separated and detected through the properties of the attached tag. A wide variety of labeling reagents are available, targeting amino groups, hydroxyl groups of different reactivity, carboxylic acids, keto and aldehyde groups, etc. The tags originally introduced were promoting UV absorbance, but over the years derivatization for fluorescence detection has become more common. The detection limits obtainable by fluorescence derivatization rival those obtainable in MS, so fluorescence detection may be an alternative worth considering if the added identification and structural elucidation possibilities of MS are not needed. Precolumn derivatization reagents and their break-down products usually display the detectable property, and excess reagent will therefore be visible in the detector, although eluted at a different retention time. The reagent is therefore usually added only in moderate excess relative to the target functional groups, to avoid large spurious peaks in the chromatograms. At low analyte concentrations, reaction speed will be radically decreased because both the reagent and solute are reactants in a bimolecular reaction. Ways to circumvent this are to add more reagent and remove the excess reagent by extraction, or to allow it to react with a solid carrier containing functional groups reacting with the reagent. When considering whether or not to apply a derivatization scheme, it must be taken into account that precolumn derivatization alters the retention properties of the analytes, so that existing separation schemes have to be modified. An altered retention pattern can also be the reason for carrying out a derivatization, e.g. by reversible attachment of chiral fluorescent groups[691] or fluorescent groups capable of expressing affinity interactions.[692] Derivatization of biopolymers is complicated by the possibility of multiple derivatization on several reactive groups, leading to multiple products from a single solute molecule. Although this carries a potential for more sensitive detection, it is likely to lead to splitting of solute peaks in the separation step. A monumental handbook of derivatization techniques for HPLC was published recently.[693]

Postcolumn derivatization is based on separating the analytes in their native form, then subjecting them to a derivatization reaction after the sample has passed the column. This means that existing separation schemes can be used, provided the eluent is compatible with the postcolumn derivatization reaction. The reaction is carried out on line by adding one or more reagents to the eluate and allowing sufficient time for the reaction to take place before the sample is passed into the detector. Successful use of postcolumn derivatization calls for reagents acting as a precursor of the detectable property, i.e. the reagent itself should not be visible

to the detector, but be converted into a detectable moiety on reaction with specific functional groups in the solutes. As in precolumn derivatization, reaction kinetics is central and reaction time must be gained by delaying the eluent–reagent mixture on its way to the detector. Various means of accomplishing delays without excessive band broadening have been described in the literature but some deterioration of the separation efficiency is inevitable. The books referenced above provide more information on applicable techniques.

7 SAMPLE PREPARATION

General guidelines for preparing samples are difficult to give, because the diversity of biological macromolecules and the matrices in which they exist is infinite. A few key concepts are nevertheless important to convey. The most prominent feature of material derived from living organisms is biological activity, and cellular processes are characterized by a continuous reconstruction through simultaneous enzyme-catalyzed metabolic and catabolic processes. Metabolic break-down of macromolecular components by controlled enzymatic processes is a central part of a system where building units are extensively recycled and, in a living cellular system, sensitive molecules are separated from metabolic processes by membranes and other barriers. A controlled exchange of material takes place with the surrounding tissue or medium through these barriers in processes driven by energy. When the system is disturbed, it will attempt to accommodate to the changes by altering the cellular processes, and if the perturbations are sufficiently large, the system collapses and the cells will die. However, this does not mean that all processes taking place in the cell will cease, only that the pathways and the mechanisms will be different, and in the absence of added energy the metabolic processes will prevail. As the energy required to maintain order ceases to flow, components normally kept separate can come in contact.

What this dramatic account of impending chaos and disorder intends to tell is the importance of knowing the history of a sample, and to control processes that lead to alterations of samples. In general, a biological sample should be subjected to analysis as soon as possible after isolation. If this is not possible, appropriate measures should be taken to ensure that the components of interest are not subject to changes by biological or chemical processes. There are numerous ways to store a sample, and the more common are freezing and lyophilization. Both these processes will lead to disruption of membranes in tissue samples, so that most components will be in contact with each other when the sample is thawed or rehydrated. Such samples are therefore particularly prone to alterations. Of particular concern are proteolytic enzymes, which exist in all biological systems. If left to operate on a cell suspension, they will rapidly lead to significant alterations. Other enzymatic processes can also cause changes to the samples. Obviously, this breakdown or alteration will largely cease if the analytes are separated from the enzymes responsible for catalyzing the changes, or if these enzymes are inhibited. Metal-complexing agents such as EDTA and EGTA (ethyleneglycol bis(aminoethyl)tetraacetic acid) are widely used and protease inhibitors are available singly or in mixtures for adding to samples in order to prevent break-down. Hassel et al. recently published a study of ways to eliminate unwanted proteolysis during extraction of proteins from protease-rich tissue,[694] and came up with a rather complicated mixture of inhibitors and metal-complexing agents. However, this should only be regarded as a first aid, as addition of reagents complicates the matrix and addition of foreign biological material compromises the integrity of the sample. Also, protection by recognized synthetic or biological protease inhibitors cannot be entirely relied upon, as is evident from accounts in the literature.[695,696] The key rule is to keep the samples on ice, and to initiate the preliminary sample clean-up steps as soon as possible, in order to separate deleterious components from the analytes. Keeping samples cold also helps maintain the biological activity of sensitive proteins that may be rapidly denatured when the cellular environment is disrupted. Detergents may be needed to get and keep solutes in solution, such as use of the mild nondenaturing detergent Triton X-114 as a phase-partitioning agent in the purification of plant proteins.[697] Micelles may also be valuable, as in the extraction of hydrophobic peptides.[698] Partitioning in aqueous two-phase systems is a batch-operated soft-separation process for cells, organelles, and membranes,[699,700] that can be used as a sample preparation step for high-performance separation.[414,701]

Reviews have been published on sample preparation and column regeneration in protein separations,[702] sample preparation and separation of food-derived peptides,[703] the use of solid-phase extraction for preparation of catecholamines and biologically active peptides,[704] on hydrolysis and amino-acid composition analysis of proteins,[705] and on enzymatic digestion of DNA for determination of nucleotide composition.[706] Solid-phase extraction on a weak cation exchanger has for example been used to quantitatively extract and concentrate peptides from plasma samples with simultaneous removal of endogenous interferents.[707]

Other processes leading to alterations of a sample are changes in the redox state and the influence of light. Although the latter factor is simple to control, it is more difficult to control the redox potential in a sample,

which can have significant impact on thiols, disulfide bridges, conjugated phenols, and other easily oxidized or reduced moieties in biological macromolecules. Addition of 2-mercaptoethanol or dithiothreitol is common in order to prevent oxidation of methionine[708] or to break disulfide bonds in difficult-to-extract proteins, such as hair.[709] The final culprit is surface adsorption of trace amounts of biopolymers, which should be controlled by the same intuition that guides the choice of stationary phases and eluents in LC.

8 PREFERRED SEPARATION TECHNIQUES FOR DIFFERENT CLASSES OF BIOMACROMOLECULES

Descriptions of the various techniques presented above have listed numerous examples of separations for each separation mode. The other dimension into this vast field is application-oriented. The exceptional variation in biological macromolecules and the size of the field make it impossible to cover applications except in summary, and the best way is to guide the reader to recent relevant reviews.

8.1 Oligonucleotides and Polynucleotides

Separation of polynucleotides is nowadays a stronghold of electrophoretic separation techniques, the most important being Southern and Northern analysis,[710] and capillary gel or zone electrophoresis.[711,712] However, liquid chromatographic techniques have been important in nucleic acid research, often in combination with electrophoretic techniques,[713] and are still widely used for clean-up, sample preparation, and for analytical separations.[714] Polymeric nucleic acids are linear polyanions, where the chains are held together by phosphodiester links between the 3′ and 5′ hydroxy groups in the (deoxy)riboses of adjacent nucleotides. The negative charge comes from the phosphodiester links, but there are also positive charges from some of the pyrimidine or purine bases of the nucleotides. In a homopolymer, the charge will increase linearly with length but, as natural nucleic acids are heteropolymers, charge cannot be directly correlated to the length of the polynucleic acid. Separations are also complicated by the formation of complexes between nucleic acids. These interactions can take the form of hair-pin turns within the same molecule, bonding between adjacent chains, or association to hydrophobic sample components.

Both hydrogen bonding and hydrophobic forces are important in the intermolecular interactions of nucleic acids. Presence of water diminishes the importance of hydrogen bonds, but in oligonucleotides the bases are stacked because of hydrophobic π-electron interactions between the aromatic bases. This stacking prevents water from accessing the hydrogen bonding sites of the bases, and strengthens the interchain binding. The hydrophobic interactions can be broken by solvents, so that water is provided a better access. Water then competes with the bases for the available hydrogen bonding sites, which weakens the bonding between base pairs. Many RPHPLC solvents are capable of mediating this breaking of hydrogen bonds. Because of this dual mechanism, these solvents can be used not only to control the retention in RPHPLC, but also to increase the retention in anion-exchange chromatography of oligonucleotides. Interchain hydrogen bonding is furthermore affected by the structure of water, whereby chaotropic salts tend to break up hydrogen bonding, whereas cosmotropic salts increase the structure and should be avoided in the separation of ''sticky'' oligonucleotides. Increased temperature also promotes dissociation of bonds between oligonucleotide chains. Applicable HPLC separation schemes are consequently based on hydrophobic interactions (RP), ionic interactions (anion exchange), and hydrophilic interaction. Metal systems should be avoided because phosphate groups bind strongly to metals. Addition of phosphate to the buffer will decrease interactions with metals. A separation of a DNA digest by anion exchange is shown in Figure 51.

Retention in RPHPLC is due to the aromatic character of the bases, and can be altered by the addition of an alkylamine ion pair reagent to the eluent, effectively transforming the separation into a mixed-mode reversed-phase ion-exchange chromatography (RPIEC). Separation materials with shorter alkyl chains are often preferred over octadecyl silica, as the retention with the latter phases tends to be high. Except for very short oligonucleotides, elution must be done with gradients. In anion-exchange chromatography, retention is controlled by salt concentration, pH, and admixture of solvent, which is recommended to break hydrophobic interactions. Because the charge is increasing rapidly with increasing size, elution has also here to be accomplished by gradients. Contrary to RP, retention in anion chromatography increases with solvent concentration.[716]

A comprehensive coverage of nucleic acid separations was the subject of a special issue of *Journal of Chromatography* (Vol. 806:1). It is significant that a majority of the papers dealt with electrophoretic or other nonchromatographic techniques, but HPLC is nevertheless being used. Two of the chromatography papers were reviews and treated the use of micropellicular anion-exchange and ion-pair RP stationary phases and optimization of gradient elution parameters in the separation of double stranded DNA[715] (includes a table comparing anion-exchange and ion-pair RP DNA separations), and displacement chromatography in the purification of

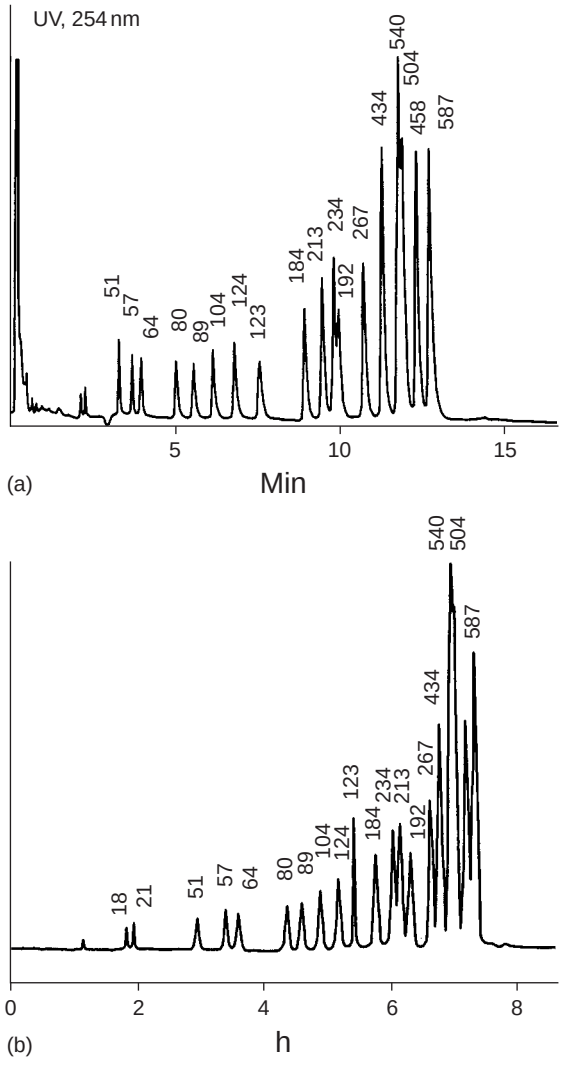

Figure 51 Separation of a pBR322 DNA-*Hae*III digest on (a) a micropellicular and (b) a macroporous anion-exchange stationary phase. Conditions in (a): column, TSKgel DEAE-NPR (2.5 µm, 35 × 4.6 mm i.d.); mobile phase, (A) 20 mM TRIS-HCl, pH 9.0, (B) 1 M NaCl in 20 mM TRIS-HCl, pH 9.0; linear gradient, 25–45% B in 0.1 min, 45–50% B in 2.9 min, 50–100% B in 57 min; flow-rate, 1.5 mL min^{-1}; temperature, 25 °C; sample, 4.8 mg of pBR322 DNA-*Hae*III digest. Conditions in (b): column, Mono Q (10 µm, 700 Å pore size); mobile phase, (A) 20 mM TRIS-HCl, pH 8.3, (B) 1 M sodium chloride in 20 mM TRIS-HCl, pH 8.3; linear gradient, 40–65% B in 2.2 h, 65–75% B in 5.8 h; flow rate, 0.15 mL min^{-1}; ambient temperature; sample, 20 mg of pBR322 DNA-*Hae*III digest. (Reproduced from C.G. Huber, *J. Chromatogr. A.*, **806**, 9 (1998) by permission of Elsevier Science B.V.)

synthetic oligonucleotides and nucleic acids.[717] The analysis of DNA adducts by capillary HPLC and CE coupled to MS was reviewed by Apruzzese and Vouros,[718] who discussed the entire analysis chain with extraction, hydrolysis, enzymatic digestion, and solid-phase extraction or HPLC as sample preparation steps. The use of LC/MS for the characterization of nucleosides, nucleotides, modified nucleotides, oligonucleotides and DNA adducts was reviewed by Esmans et al.,[719] collating the procedures for the analysis of oligonucleotides in a diagram. The preferred technique is HPLC with electrospray MS. Techniques for the detection of damage to DNA were reviewed by Cadet and Weinfeld[720] and by Frenkel and by Klein,[721] and the use of HPLC for recognizing covalent modifications to nucleic acid by anticancer drugs was reviewed by Cummings et al.[722] Size-dependent chromatographic separation of nucleic acids, in particular slalom chromatography, was reviewed by Kasai,[658] and the use of nucleic acid polymers immobilized on solid supports for affinity chromatographic isolation of polynucleotides and polynucleotide binding proteins was the subject of a review by Jarrett.[472]

The use of three acrylate-based SEC columns with different molecular weight cut-offs connected in series was described by Franek and Koutnik to separate oligonucleosomal DNA fragments from apoptotic animal cells with molar ammonium acetate as eluent.[723] Itoh et al. used a fluoroalkyl silica that was converted into a mixed-mode sorbent by treatment with trioctylmethylammonium chloride for separation of various nucleic acids by gradient elution with either an acetonitrile solvent gradient or a perchlorate salt gradient.[724] Good separations could be obtained by both elution mechanisms. The mixed mode could be used for separations of up to 40-mer oligoadenylates, and for oligocytidylates, oligodeoxynucleotides, and t-RNA. Chromatography on strong anion-exchange columns has been used for the preparative purification of supercoiled plasmid DNA.[595] Ion-pair HPLC could resolve three components in a reverse transcriptase chain reaction product, whereas only two components were detected by agarose gel electrophoresis.[725] An example of RP separations of DNA digests is shown in Figure 52.

Transfer-RNA has been separated by gradient elution on hydroxyapatite, and it was found that increased pH caused a stronger bonding of t-RNA to the stationary phase, a retention dependence opposite to proteins.[613] A mixed-mode separation on aminopropyl silica modified with alkyl groups, mainly according to the HIC mode, has also been used for t-RNA.[726,727] High-molecular-weight RNA has also been separated on hydroxyapatite following dissolution in guanidinium isothiocyanate.[614] A way to accomplish 5′-aminoalkyl labeling of DNA oligomers through *N*-hydroxysuccinimide ester activation and separation of the resulting derivatives was described by Wang et al.[728] Mixed-mode separation of nucleic acids has also been reviewed.[729]

8.2 Peptides

Peptides are small proteinaceous compounds, ubiquitous in living creatures and in the natural environment.

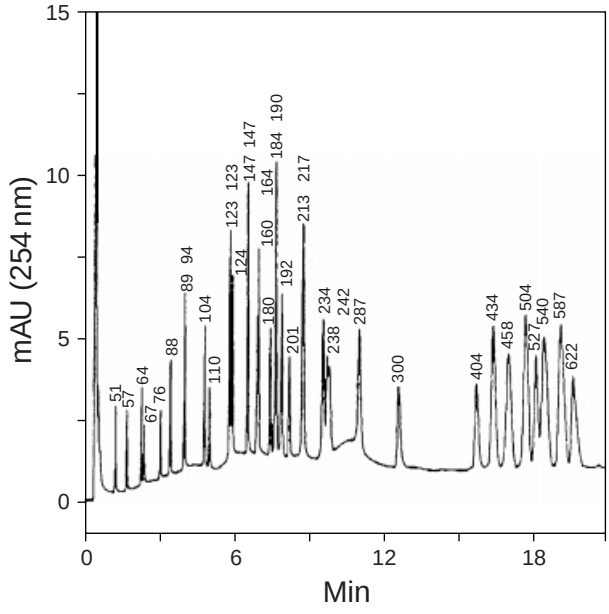

Figure 52 Chromatogram of pBR322-*Hae*III digest (0.75 µg) and pBR322-*Msp*I digest (0.4 µg) on a PS-DVB-C_{18} RPHPLC column (50 × 4.6 mm i.d.) with elution optimized for separation of the large number of fragments (size indicated by labels). Mobile phase composition: A = 0.1 M triethylammonium acetate (TEAA), pH 7.0; B = 0.1 M TEAA pH 7.0, 25% acetonitrile. Gradient: linear 37–55% B in 6 min; 55–65% B in 14 min; flow rate, 1 mL min^{-1}; temperature, 55 °C; detection, UV as 254 nm. (Reproduced from P.A. Doris et al., *J. Chromatogr. A*, **806**, 49 (1998) by permission of Elsevier Science B.V.)

When actively produced by organisms, peptides tend to have highly specific tasks and their biological activity is usually expressed at very low concentrations. Isolation of peptides is therefore important in biosciences, both for analytical purposes and as preparative steps in biological and biomedical research. Because of their potential activity, it is also essential that peptides synthesized for research or therapeutic use are efficiently purified before being used in biological systems,[730] a task often accomplished by preparative HPLC. Peptide separation is furthermore a central part of peptide mapping schemes (Figure 53) for elucidating protein structure.[731,732]

The distinction between a peptide and a protein is somewhat arbitrary, but it is common to refer to amino acid sequences with 50 residues or less as peptides. The properties differ widely in this range – dipeptides are small molecules and behave chromatographically as such, with shallow retention curves and retention more or less directly related to the properties of the constituent amino acids. Large peptides with M_W of several thousand daltons differ little from proteins in their chromatographic behavior and follow the multiple interaction mechanisms outlined in the introduction. Under favorable conditions, peptides lack secondary structure such as α-helices up to a size of about 15 amino acids. When peptides exceed this length, the likelihood of secondary and tertiary structure increases[733] and, towards the upper size limit, peptide chains are prone to fold in order to internalize side-chains of low polarity.[734] This has a pronounced effect

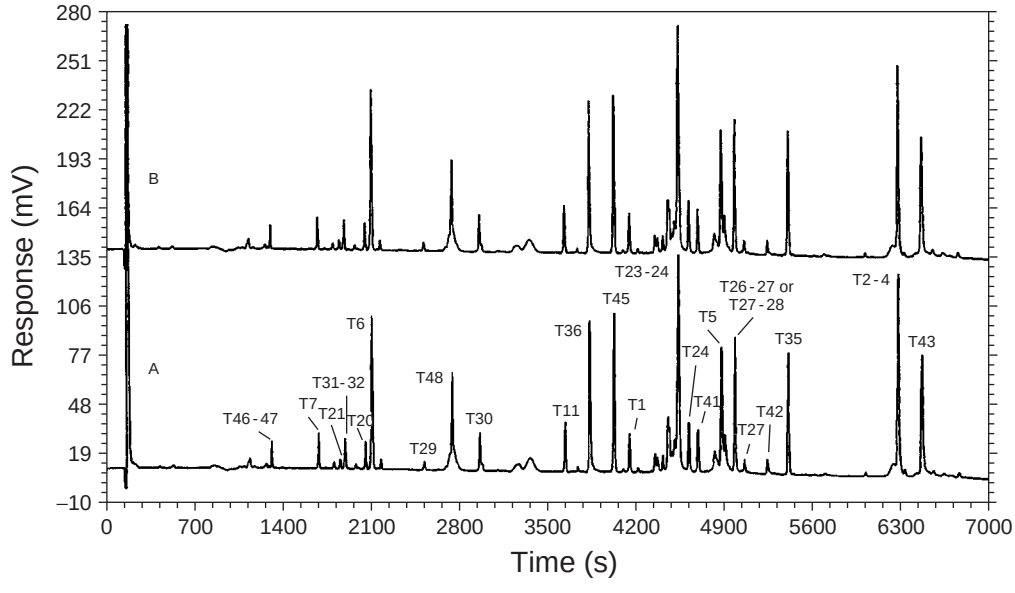

Figure 53 Optimized and (A) aseptically and (B) nonaseptically processed (2-h trypsin digest at ambient temperature) protein C peptide map on a Zorbax 300SB-C_{18} (4.6 × 250 mm) column at 35 °C. Mobile phase A, 0.1% TFA in H_2O; mobile phase B, 0.08% TFA in 80% acetonitrile and 20% H_2O, at a flow rate of 1 mL min^{-1}. Samples injected: 100 µL; 0.25 mg mL^{-1}. Detection: UV at 214 nm (1 AU/V). (Labels refer to tryptic fragments used in a validation procedure.) (Reproduced from A.D. Carlson et al., *Anal. Chim. Acta*, **352**, 227 (1997) by permission of Elsevier Science B.V.)

Table 8 Summary of HPLC separation techniques commonly used to separate peptides

Separation mode	Use	Typical eluents
SEC	First separation step, isolation of peptides from proteins and other large molecules. Columns with pores separating from 100 to 7000 Da are available for analytical separations.	Water with the addition of buffer and salt at a concentration sufficiently high to suppress adsorption tendencies due to residual ionic charge on the stationary phase.
IEC	Gaining in popularity because small changes in pH can give large changes in selectivity. Ion exchange materials with high separation efficiency have also become available.	Cation-exchange mode is favored, at least with silica-based materials. Buffers are typically weakly acidic. Elution can be carried out by salt or pH gradients.
RPHPLC	The most popular technique because of the high separation efficiency, the wide scope, and the relative stability of peptides, compared with proteins.	Gradients of water-miscible organic solvents (acetonitrile is often preferred) in acidic buffers. The buffer anion can have an effect on selectivity.

on the apparent hydrophobicity and the dissociation constants of ionizable side-chains in the native state. The presence of disulfide bridges also contributes to conformational changes, and depends on the oxidation state of the peptide. All factors leading to alterations in conformation must be considered and brought under control when a method is developed, because they may have a large impact on the chromatographic behavior of a peptide.[125]

Separation schemes for peptides are often complex, especially when the analyte exists in a biological matrix. Combinations of chromatographic techniques with orthogonal separation dimensions are therefore regularly employed. Because of the relatively low molecular weight and reasonable stability of peptides, a variety of HPLC subtechniques which would lead to irreversible sorption or denaturation of larger biomacromolecules may be applicable. The most popular separation modes are SEC, IEC, and RPHPLC. The initial separation steps usually aim at isolating the molecular weight fraction of interest from proteins and low M_W matrix components. Final separations of the peptide fraction are then carried out with techniques that capitalize on differences in hydrophobicity or net ionic charge. Gel electrophoresis and capillary zone electrophoresis (CZE) are useful for verifying the purity of separated peptides, as their separation principle based on differential migration in electric fields is markedly different from chromatographic partitioning. Table 8 summarizes the most important HPLC techniques used in the separation of peptides. As can be inferred from their biological significance, separation of peptides has been the subject of numerous monographs[73,735–738] and a large number of review articles, of which a select number are referred to below. Recent practical reviews have been written by Mant and Hodges,[125] and by Lundell.[739]

SEC is often used as the first step in a multistep peptide separation protocol or for separation of peptides from proteins. The most frequently used columns are silica

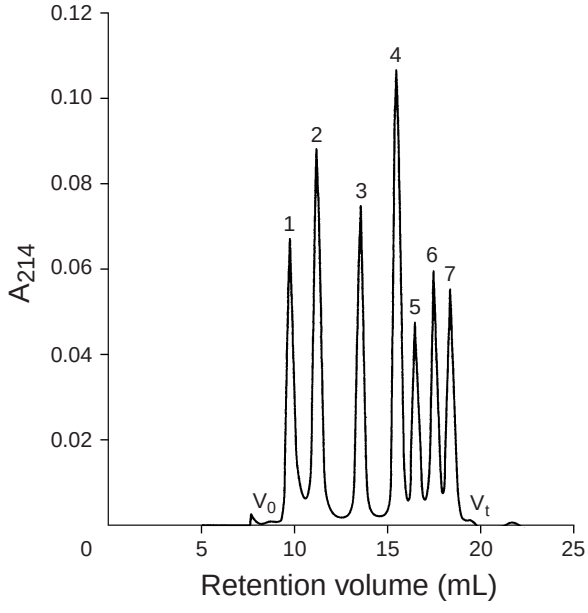

Figure 54 Separation of standards on a Superdex Peptide HR 10/30 column (10 × 306 mm). Sample: 25 µL, 0.2 mg mL^{-1} of each polypeptide (1–6) and 7.8 mg mL^{-1} of glycine (7). Mobile phase: 20 mM phosphate buffer, pH 7.2, containing 0.25 M NaCl; flow rate 0.5 mL min^{-1}. Standards, with molecular weights (Da) in parentheses: 1, Cytochrome *c* (12 500); 2, aprotinin (6500); 3, gastrin I (2126); 4, substance P (1348); 5, Gly_6 (360); 6, Gly_3 (189); Gly (75). (Reproduced from G.B. Irvine, *Anal. Chim. Acta*, **352**, 392 (1997) by permission of Elsevier Science B.V.)

with pore sizes of 60–150 Å. These columns are designed for SEC of proteins and the molecular weight range is thus inoptimal for separation of peptides. Polymer-based columns designed for optimal separation of the molecular weight range 100–7000 Da are commercially available and have been successfully used in the SEC separation of peptides[740] (Figure 54). The general precautions outlined above apply to peptides. Self-association and interaction with other biomacromolecular solutes can also affect their separation. The use of denaturing conditions,

such as by the addition of urea, may therefore be necessary to ascertain that the peptides are monomeric. The theoretical and practical aspects of high-performance SEC of peptides was recently reviewed by Irvine.[741]

RPHPLC of peptides is typically run on silica columns with linear alkyl functional groups (C_4 to C_{18}) of 100–300 Å pore size, with gradients of 2-propanol, acetonitrile or methanol. Some TFA or another acid (see above) is added to the eluent in order to suppress the dissociation of surface silanols, and to accomplish an ion pairing with basic amino acid side-chains. Due to the reasons outlined under mixed-mode interactions above, gradient parameters may be critical to success. A gradient should always start from a low concentration of organic modifier, not pure water, because the layer of alkyl chains on RP stationary phases can collapse in entirely aqueous eluents. The best separations are usually obtained with relatively shallow, linear gradients. The final concentration of organic modifier is typically between 50% and 100%, which means that the solubility of ionic mobile phase additives should be checked at the highest modifier concentration. Extremely hydrophobic peptides may require special eluents and solvents.[742] The column temperature can be used to fine tune the retention through temperature-induced conformation changes of peptides in contact with the stationary phase. A slightly increased temperature can also improve the separation efficiency because the eluent viscosity decreases, affecting the diffusion coefficients positively. A comparison of RPHPLC and transient isotachophoretic CZE for the determination of peptides in plasma has been published.[707] The related but softer separation principle of HIC is useful for separation of larger peptides, whose secondary structure is creating problems in RPHPLC, or if denaturation is an issue.[99]

Synthetic peptides are regularly purified by RPHPLC (Figure 55), although other techniques are also used, such as cation exchange,[161,743,744] mixed-mode HIC/cation exchange,[224,647–649] ion pair chromatography,[745] and IMAC,[495] specifically for isolating peptides with metal-binding N-terminal amino acids (His, Trp, or Cys).[746] The conventional eluent for RPHPLC of peptides is an aqueous acetonitrile gradient acidified with TFA. Addition of perchlorate ions to the mobile phase was recently demonstrated by Sereda et al.[236] to have a pronounced effect on the retention of synthetic amphiphilic α-helical and nonhelical peptides in C_8 RPHPLC, where phosphoric acid was used to acidify an acetonitrile gradient. The general trend seen on adding perchlorate was an increased retention, accompanied by changes in selectivity. These effects were ascribed to ion-pair formation between amino acids with a positively charged side-chain and the perchlorate ions, which was also confirmed by calculations. The selectivity of the eluent containing

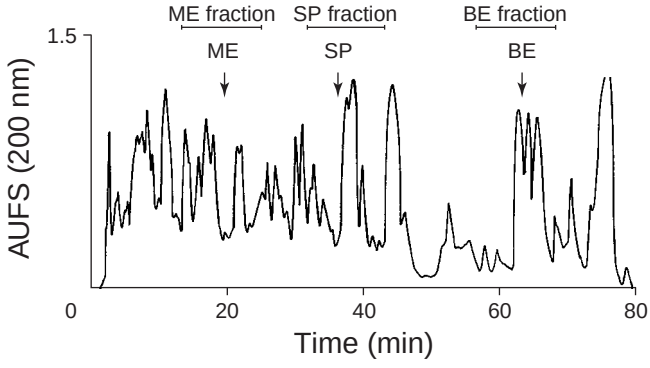

Figure 55 Gradient RPHPLC chromatogram of the peptide-enriched fraction obtained from one human pituitary. The arrows indicate the retention time of synthetic peptides, and the brackets indicate collected peptide fractions. (Reproduced from D.M. Desiderio, *Anal. Chim. Acta*, **352**, 89 (1997) by permission of Elsevier Science B.V.)

aqueous phosphoric acid–perchlorate–acetonitrile was also markedly different from aqueous acetonitrile–TFA. Addition of chaotropic anions to the eluent thus provides a tool for altering the retention compared to that obtained with acetonitrile–TFA gradients. Chiral separations[211,691] may be required for determination of the enantiomeric purity of peptides in the pharmaceutical, alimentary, and agricultural industry.

An entire volume of *Analytica Chimica Acta* was recently devoted to the chromatography of peptides. Among the reviews presented was a survey by Štulik et al. of RP, HIC, SEC, IEC, affinity, and CSP for peptide analysis by HPLC.[95] The paper contains an outline of how the chromatographic behavior of a peptide is influenced by its physicochemical properties and lists a number of manufacturers providing columns with different chemistries for separation of peptides. Characterization and purification of peptides by high-performance strong cation-exchange chromatography was reviewed by Crimmins.[569] A commonly used stationary phase is sulfoethyl aspartamide. Separations are carried out at an acidic pH to promote the cationic character of basic groups. Among the synthetic analytes used as examples are N-terminally blocked and disulfide linked peptides. Separation of model peptides subjected to limited proteolysis was reviewed by Rholam and Cohen,[747] with particular emphasis on cleaving site structure in order to establish models for peptide hormone catabolism. Monitoring proline-specific endo- and exo-peptidase activity by HPLC assays was reviewed by Harada.[234] Characterization of human hemoglobins by the differences in their polypeptide chains was reviewed by Huisman,[748] discussing separations by cation exchange and RPHPLC on short alkyl chain columns. Meriluoto[235] summarized the separation of microcystins and nodularins, classes of cyclic

hepatotoxic peptides with tumor-promoting properties produced by water-blooming cyanobacteria. Emphasis is placed on the conditions for extraction from different types of sample matrices, separation by anion exchange, and by RPHPLC with conventional and internal surface materials. A comparison is also made with other methods for determination of these peptides. A strategy for screening of combinatorial peptide libraries was reviewed by Hammond,[238] focusing on peptide sequences immobilized on a carrier and used as affinity ligands in highly specific purification schemes for proteins. The interaction of the target amino acid sequence FLLVPL with fibrinogen (blood clotting Factor I) was described and its use for purification of the protein was discussed.

Neuropeptide separation has been treated in several recent reviews. Rissler[749] addressed sample preparation, separation, and determination of low levels of peptides related to substance P, an undecapeptide involved in the transmission of pain. Hurst and Zagon[750] reviewed the sample preparation, separation and detection of enkephalins. Desiderio presented a review of separation and detection of endogenous neuropeptides[237] in various biological tissues and fluids, and Desiderio and Zhu[239] reviewed the sample collection, SPE clean-up and HPLC separation in the determination of Met-enkephalin and β-endorphin from pituitary gland tissue (Figure 55). Both papers discussed radioimmunoassay (RIA) and tandem MS, among other detection techniques, and concluded that tandem mass spectrometry (MS/MS) was superior for detecting these peptides.

Multidimensional separations are required to separate highly complex peptide mixtures. An LC/LC/MS system for mapping fragments from proteolytic digests was thus constructed by Jorgenson et al.,[192] based on conventional diameter analytical HPLC columns. The first separation dimension was by SEC, using six 30-cm long columns in series with 0.1% aqueous TFA as the mobile phase. Fractions from this column train were transferred to an RPHPLC separation step using two columns packed with nonporous C_{18} modified silica with an aqueous acetonitrile gradient containing 0.05% TFA (Figure 56). Peptide fragments from ovalbumin and BSA were separated with this system and identified by electrospray MS. Another LC/LC/MS technique (cation exchange followed by RP) was also found to be promising for the separation and identification of undigested proteins.[751]

Coupling HPLC with electrophoretic separation techniques adds yet another separation dimension because of the radically different separation principle of CE compared to interaction chromatography. Moore and Jorgenson separated fluorescein isothiocyanate (FITC)-tagged peptide fragments from trypsin digests of hen

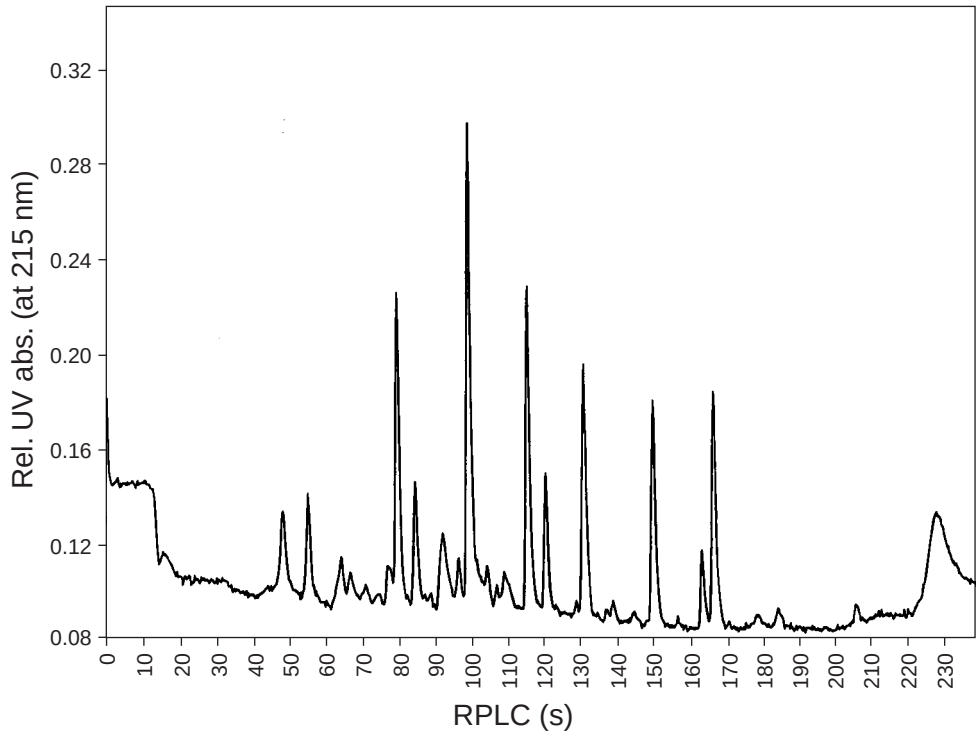

Figure 56 Single coupled-column RPHPLC chromatogram extracted from a SEC separation (retention time interval 90–94 min), showing separation of tryptic peptides as detected at 215 nm. (Reproduced from G.J. Opticek et al., *Anal. Chem.*, **69**, 2287 (1997) by permission of American Chemical Society.)

ovalbumin by size exclusion on a hydrophilic macroporous 2-hydroxyethyl methacrylate/ethylene dimethacrylate (HEMA/EDMA) copolymer column, using 85% methanol–100 mM ammonium acetate as mobile phase at low flow rates. The fragments eluted from the SEC column were sampled repetitively on-line and subjected to separation on a narrow bore 5-μm C_8 column with a fast aqueous acetonitrile gradient. The exit of this column was in turn linked to a vertical fused-silica CZE capillary, onto which sampling was triggered by a high speed optical gating injection system to obtain a comprehensive three-dimensional peptide separation.[752,753] Another automated application of HPLC is in the chromatography-based assay of enzyme activities, reviewed by Lambeth and Muhonen.[754] The potential of multiple separation techniques in combination with selective and sensitive detection is large, but the design and operation of multi-dimensional systems requires considerable expertise and it is not yet a feasible technique for most analytical laboratories. It should also be mentioned that matrix-assisted laser desorption/ionization (MALDI) and direct injection electrospray MS are nonchromatographic separation techniques that can be favorably used for peptide samples in relatively sparse matrices.[755]

Detection-oriented techniques for identification and determination of peptides containing arginine, cysteine, proline, tryptophan, or tyrosine residues were reviewed by Cui et al.[756] A number of selective derivatization reagents were discussed, targeting each residue, and electrochemical (tyrosine-containing peptides), UV, and fluorimetric detection is used to identify the residues based on the properties introduced by the reagents.

8.3 Proteins

Chromatographic separation of proteins can take place either by interaction or by size exclusion. Interaction chromatography is based on contact between the amino acid residues at the outer surface of the protein and the stationary phase, resulting in interactions whose strength is adjusted by varying the mobile phase conditions. Proteins differ in stability and their conformation can be significantly altered during separation, with changes in their surface properties as result. Which residues will be exposed is of course determined by the basic structural features of the protein, but also to a large extent by the chromatographic conditions. The factors important for controlling the protein structure and hence the exposed surfaces during separation are the stationary phase (the nature of the ligands and their density, the chemical nature and porous properties of the matrix), the mobile phase composition (pH, solvents, concentration and type of added salts or nonionic agents affecting the water structure, presence of detergents or other additives), and also physical factors (temperature and flow rate). A general strategy for the separation of proteins is therefore difficult to formulate.[99]

The best guidance is obtained by gathering as much information as possible on the proteins to be separated. Amino acid sequence, chemical nature, size, helicity and three-dimensional structure, and knowledge of the biological function are factors that can be useful in selecting an optimal separation scheme. If the proteins have been characterized, data are likely to be present in databases that are freely available on the Internet. One such database, which is referenced here in spite of the volatile nature of the Internet, is the ExPaSy Molecular Biology Server.[576] The data that can be collected or computed on this server can be very useful for designing protein separations. Among the modules available is a huge database of protein sequences, families and domains, two-dimensional polyacrylamide gel electrophoresis data, three-dimensional images of proteins and other biological macromolecules, and enzyme nomenclature. Tools for protein identification and characterization, facilities for similarity, pattern and profile searches, primary structure analysis and secondary structure prediction, and tertiary structure and transmembrane region detection also exist on this server. Calculations can be done from the database entries, or from user-entered primary structures. ExPaSy also provides links to a large number of other servers with additional databases and computational facilities.

Advanced protein separation techniques was the topic of a recent review issue of *Journal of Chromatography*, including overviews of new packing materials,[96] nonporous sorbents,[97] permeable packings,[98] and protein interactions at sorbent interfaces.[757] Strategies were reviewed for separation of recombinant proteins,[758] sensitive enzymes,[86] modified hemoglobins,[759] lipoproteins,[611] glycoproteins,[760] glycated and carbonylated proteins,[319] and the isolation of allergens.[761] Special techniques such as labeling reactions[688] and the studies of drug–protein interactions[762] were also part of this issue. A number of books are also devoted to chromatography of proteins.[51,73,737,738,763–766] A recent monograph by Sadana[767] deals with the mechanisms for different bioseparation processes and how they affect protein stability. The molecular basis of surface activity and protein adsorption is discussed and strategies for avoiding inactivation are outlined. Oscarsson[757] has reviewed the factors leading to partitioning of biomacromolecules to sorbents, with particular attention to protein–surface interactions, and describes a scanning probe microscopy technique for studying such interactions at the nanometer level. Beeckmans[768] has reviewed the use of chromatographic techniques for studying protein–protein interactions, and provides an extensive practical section on support activation and immobilization.

8.3.1 Membrane Proteins

The group of proteins associated with cellular membranes differ in their way of attaching to the lipid bilayer. Integral membrane proteins are attached either by covalently bound lipid chains, or by having a consecutive sequence of amino acids with low-polarity side-chains sufficiently long to extend through the lipid membrane, whereas the extrinsic membrane proteins are associated with these proteins, or the head groups of the bilayer lipids. The salient amphiphaticity makes it difficult to detach integral proteins from the lipid membrane, and the use of detergents is necessary in order to solubilize the membrane and get the proteins into solution. The difficulty of finding appropriate conditions increases with the number of components in the system, and a series of different steps are often needed in order to prepare injectable samples of the different proteinaceous components of membranes. Strategies for finding these conditions, as well as choosing separation technique and detergents, are outlined by Josic et al.[769,770] Tools for predicting transmembrane regions in a protein are available through the ExPaSy server[576] and might be useful in method development.

Once the proteins have been solubilized, they are amenable to separation by practically the entire array of separation chemistries applicable to proteins. There are some particular problems that should be kept in mind, such as microheterogeneity due to differences in glycosylation pattern, irreversible adsorption to the column due to nonspecific interaction, the tendencies to precipitate or aggregate, and the strain put on the system because of the high concentrations of salt, detergents, or other denaturing reagents that may have to be added to the eluents to overcome these tendencies. Subject to the nature of the protein, mild detergents can be sufficient to suppress aggregation and precipitation, in which case separation can be made under nondenaturing conditions with IEC, hydroxyapatite, HIC, and various affinity chemistries. Otherwise, or if the selected separation chemistries are SEC and RPHPLC, the addition of strong, denaturing detergents, urea, or other denaturing agents is normally necessary. Solubilizing detergents also tend to dynamically modify the stationary phase by adsorption, leading to substantial changes in interaction patterns.[769] Lundahl et al.[771] evaluated hydroxyapatite columns of different brands for separation of integral membrane proteins solubilized in SDS (sodium dodecyl sulfate) and found considerable differences in elution pattern (Figure 57).

von Jagow and Schgger have presented a monograph devoted entirely to practical aspects of membrane protein purification.[764] An interesting ancillary development that should be mentioned in connection with membrane proteins is the use of immobilized liposomes, proteoliposomes, or biomembranes for studies of interactions of proteins with lipid surfaces.[6,414]

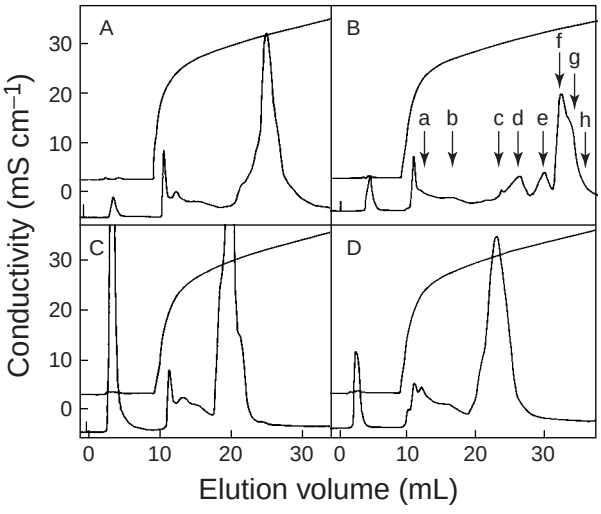

Figure 57 Chromatograms of solubilized human erythrocyte integral (intrinsic) membrane proteins in complex with SDS using hydroxyapatite HPLC columns of four different manufacturers: (A) Tonen Taps-020810, (B) PENTAX SH-0710F, (C) A-7610 and (D) TSKgel HA-1000. The SDS–polypeptide complexes (0.5 mg) were eluted with a phosphate buffer concentration gradient at a flow rate of 0.8 mL min^{-1}. The height of each panel corresponds to 0.12 absorbance units at 280 nm. Arrows indicate fractions that were analyzed by electrophoresis. Conductivities of 10, 20 and 30 mS cm^{-1} at 25 °C correspond to phosphate concentrations in the eluent of 85, 205, and 370 mM, respectively. (Reproduced from P. Lundahl et al., *J. Chromatogr.*, **604**, 98 (1992) by permission of Elsevier Science B.V.)

8.3.2 Lipoproteins

Serum LPs are complexes between proteins and lipids found in blood serum. The core consists of hydrophobic components such as triglycerides or cholesterol esters, shielded by an amphiphilic layer that is comprised mainly of proteins (Figure 58).

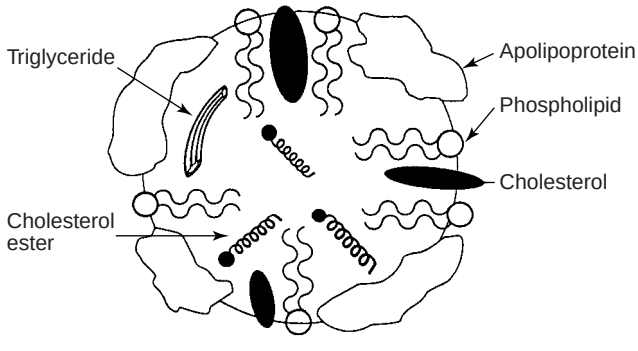

Figure 58 Diagram of an LP particle. (Reproduced from T. Takey, W.C. Purdy, *J. Chromatogr. B*, **671**, 238 (1995) by permission of Elsevier Science B.V.)

LPs are classified as high-density lipoproteins (HDLs), low-density lipoproteins (LDLs), and very-low-density lipoproteins (VLDLs), determined by the relative buoyancy in a tedious, complicated, and time-consuming gradient ultracentrifugation process. Takey and Purdy[772] reviewed preparative and analytical separations of LPs, using SEC, affinity chromatography, and HPLC techniques for resolution of structurally different LPs. A review of separation strategies for LPs was also recently presented by Shibusawa,[611] who concluded that the preferred separation technique for HPLC of LPs is hydroxyapatite chromatography with step changes in phosphate concentration at pH 7.4 for elution (Figure 59). This technique can separate all three classes of LPs, but is unable to resolve HDLs from serum albumin and globulin. A complementary technique is CCC, which offers the very mild separation conditions required to preserve the structures of the aggregates during separation. Combined use of these two techniques with CCC as a preliminary fractionation step for final separation by hydroxyapatite chromatography is capable of resolving all three classes of LP.[611] Note that CCC is not an HPLC technique, and it requires highly specialized and quite complicated equipment.

8.3.3 Recombinant Proteins

Proteins expressed in prokaryotic organisms impose particular challenges on chromatographic separations for purification and analytical purposes. Glycosidation patterns of recombinant proteins are often different from proteins expressed in eukaryotic systems, protein folding is an issue, and the extraction of proteins that are excreted by eukaryotic cells but remain inside prokaryote organisms may call for rigorous conditions

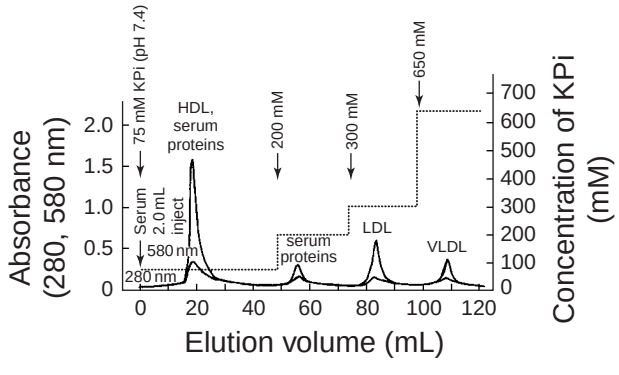

Figure 59 Stepwise elution profile of human serum LPs on a Bio-Gel HTP DNA grade hydroxyapatite column (250×10 mm i.d.) using 25, 200, 300, and 650 mM potassium phosphate buffer at pH 7.4 as eluents in a step gradient fashion, at a flow rate of 0.2 mL min^{-1}. Sample: 2.0 mL human serum. (Reproduced from Y. Shibusawa, *J. Chromatogr. B*, **699**, 427 (1997) by permission of Elsevier Science B.V.)

for solubilization or separation[773,774] of inclusion bodies. The use of column chromatography for the purification of recombinant proteins was reviewed by Jungbauer and Boschetti[775] with special emphasis on validation and on long-term stability in the chromatographic separation processes. Evangelista Dyr and Suttnar[758] reviewed methodologies applicable to the purification of recombinant proteins, including sample preparation and assays methods, and compared the techniques with those used in the purification of conventional proteins.

A correct disulfide bonding pattern is central for folding of recombinant proteins and verification of correct disulfide linkages should therefore be part of the quality control scheme for recombinant proteins. Lei et al.[776] described the use of tryptic mapping of recombinant human growth hormone on an immobilized enzyme cartridge, followed by on-line separation of fragments by TFA/acetonitrile RPHPLC (Figure 60). The MALDI/TOFMS technique was used to verify the absence of peptides with mismatched disulfide bridges. Ikegaya et al.[777] used the Edman degradation and pepsin mapping followed by HPLC on C_{18} columns to elucidate the exact positions of the disulfide bridges in a purified recombinant HSA. General characterization of recombinant proteins is often performed with CE instead of by HPLC.[778,779]

8.3.4 Enzymes

In analytical separations of biomolecules, it is often sufficient to establish stable and reproducible conditions so that the molecule under study elutes as a single well-shaped peak at a predictable retention time, separated from other biomacromolecular components of the sample. Denaturation and loss of biological activity that accompanies conformation changes is therefore often an intentional part of the sample preparation. It may also be induced by the choice of mobile phase components or stationary phases that lead to controlled denaturation. Enzymes are different, because the purpose of a separation is typically to recover the enzyme in its native state with maintained biological activity, in particular in preparative high-performance separations. The in vitro stability of enzymes varies, with some being quite unstable. Special techniques are therefore often required when the goal of the separation is to preserve biological activity, a topic addressed in a review by Kaufmann.[86] Among the techniques advocated are perfusion chromatography and the use of tentacle gels,[163] to accomplish fast chromatography on soft interaction media.

Toribio et al.[780] have reviewed recently developed protocols for the purification of pyridine nucleotide-dependent enzymes. All of the referred protocols made use of chromatographic steps, with affinity

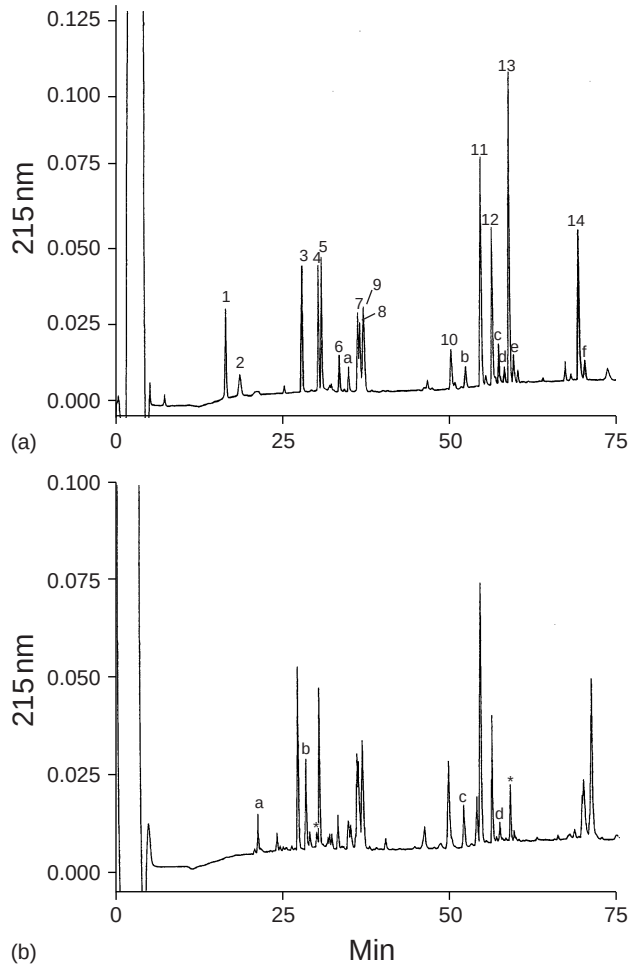

Figure 60 Tryptic map of (a) native and (b) reduced recombinant human growth hormone. The protocol and conditions for reduction, digestion, and RP chromatography are found in the original paper. Labels refer to tryptic fragments; lower case letters correspond to incomplete tryptic digests and peaks marked with an asterisk correspond to unreduced disulfide bridged fragments. (Reproduced from J. Lei et al., *J. Chromatogr. A*, **808**, 125, 127 (1998) by permission of Elsevier Science B.V.)

chromatography being employed in more than nine out of ten occasions. Among the ligands used were Cibacron Blue F3GA, Procion Red HE3b, Orange A, and 2′,5′-ADP. Frequently encountered were also IEC, SEC and HIC, whereas other chromatographic modes were used more sporadically. The sequence in which the chromatographic steps are performed is important to make best use of the separation dimension and sample capacity offered by each technique, leading to rather complicated schemes like the one outlined in Figure 61. The authors recommended the use of perfusion sorbents and also considered the development of tailor-made molecularly imprinted stationary phases specific for individual enzymes as a goal for the future. Toribio et al. have also reviewed purification techniques for glutathione-related enzymes (reductase, peroxidase and transferase).[171] The hegemony of chromatography for preparative purposes was demonstrated by its utility in all of the reviewed papers, whereas only 3% of the procedures used preparative techniques based on electrophoresis or isoelectric focusing. The need to utilize more than one separation dimension in an overall purification scheme was obvious – two or more chromatographic techniques with different selectivities were used in more than 90% of the protocols, and in more than half of the protocols three or more chromatographic dimensions were used.

The purification of carbonic anhydrase isozymes from different sources was reviewed by Bergenhem,[781] with emphasis on application of their sulfonamide complexes in single-step affinity schemes. Kopperschläger and Kirchberger[782] reviewed methods for the separation of LDH and its isoforms, concentrating on affinity separation techniques using natural and pseudo-biospecific ligands. Methods for the separation of superoxide dismutase isoenzymes were reviewed by Michalski[783] along with techniques for monitoring their purification and measuring their activity. Lin et al.[784] provided an overview of methods for the purification of hydroxysteroid dehydrogenases, mainly by HIC and dye-affinity chromatography. Sample preparation, prefractionation, and separation techniques for mammalian cytochrome enzymes based on IEC, hydroxyapatite, HIC, and immobilized metal affinity systems were reviewed by Roos.[785] Multistep separations using SEC, IEC, hydroxyapatite chromatography, and RPHPLC for the isolation and characterization of mammalian catechol-*O*-methyltransferase were reviewed by Tilgmann and Ulmanen.[786] A wide array of chromatographic techniques (RPHPLC, IEC, HIC, TAC, and affinity schemes based on sorbents with phenylboronate, immobilized metal ions, and cyclodextrin and starch as biospecific ligands) applicable to the separation of α- and β-amylases and their isoenzymes from complex extracts was reviewed by Brena et al.,[320] Moriyama and Ikeda[787] reviewed the assay of clinically relevant amylase enzymes, acting on glycosidic bonds.

The purification of proteases using HIC and affinity schemes with cyclic peptide antibiotics and immobilized dyestuffs as protease-specific ligands was reviewed by Ibrahim-Granet and Bertrand.[788] Hamada[789] reviewed the use of preparative HPLC for the large-scale production of carbohydrate hydrolases, proteases, and transglutaminases from food, microbial and plant sources for industrial applications. Parmentier et al.[790] reviewed the purification of glutathione synthetase, glutathione reductase, γ-glutamylcysteine synthetase, and γ-glutamyltranspeptidase by the use of multiple LC separations. Isolation of transglutaminase isoforms from tissue and body fluids along with procedures for purification of γ-glutamyltransferases were reviewed by Wilhelm

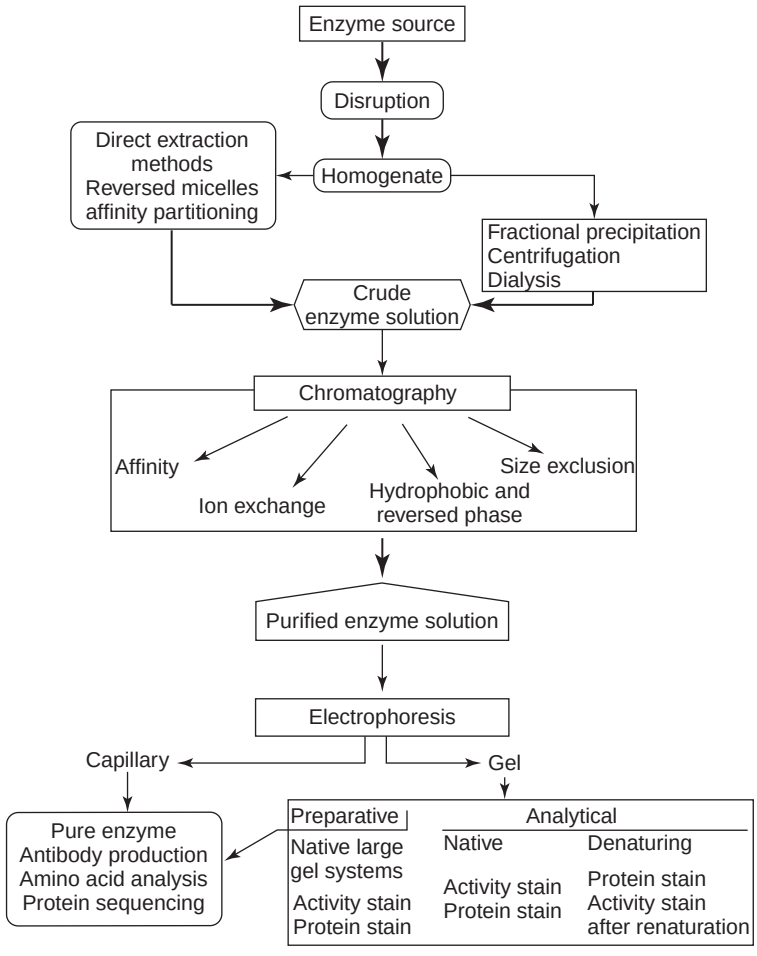

Figure 61 General purification scheme suitable for $NADP^+$-dependent enzymes according to Toribio et al.[780] (Reproduced from *J. Chromatogr. B*, **684**, 2 (1996) by permission of Elsevier Science B.V.)

et al.,[791] particularly addressing the problems caused by formation of irreversible aggregates. The most widely used separation techniques were SEC and IEC. Affinity schemes based on heparin for the purification of lipases, phospholipases, and kinases were reviewed by Farooqui et al.[448] Focus was also placed on techniques for attaching heparin to supports and detecting its presence after immobilization. Raspi[792] presented a review of sample preparation, purification, and assays for isolating and determining kallikrein proteinases from tissues and body fluids. Knudsen et al.[793] reviewed chromatographic and electrophoretic separation strategies for eukaryotic DNA topoisomerases isoforms, and the techniques used for determining their catalytic activity. The added benefit of high-sensitivity metal analysis by coupling a separation of metal-containing enzymes to inductively coupled plasma mass spectrometry (ICPMS)[794] is illustrated in Figure 62.

The availability of assay techniques is central for determining the biological activity of enzymes during and after isolation. Many such assays are based on HPLC, and a few overviews covering different HPLC assays have appeared in the recent literature.[754,795–797]

8.3.5 Protein Folding and Unfolding Studies using High-performance Liquid Chromatography

By virtue of the large variations in surface properties and the possibility of controlling the interaction time and strength by manipulating the eluent parameters, liquid chromatographic techniques lend themselves to studies of protein folding and unfolding, the processes responsible for denaturation and irreversible adsorption of proteins on sorbent interfaces. Modulation of the hydrophobicity of polymethacrylate coatings on porous and nonporous silica enabled Hanson et al.[142] to mediate selective unfolding of proteins on the sorbent surface, conformational changes that made it possible to alter the separation pattern of the protein probes lysozyme, cytochrome *c*, and myoglobin. Similar effects were seen for synthetic peptides with defined secondary, tertiary, and quaternary structures.[143] Benedek[798] studied the unfolding

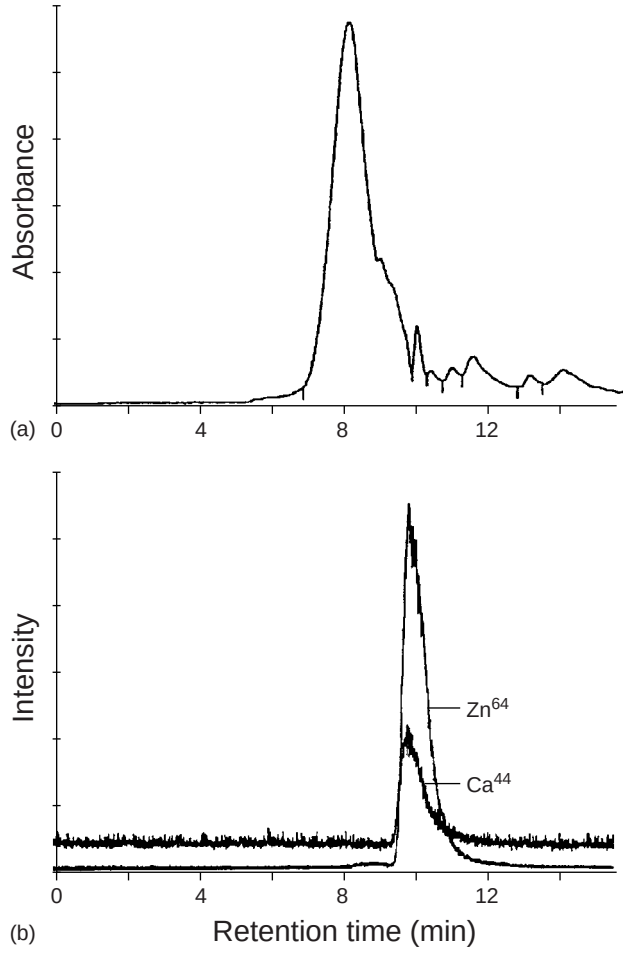

Figure 62 Size-exclusion HPLC and metal ion determination of an alanyl-aminopeptidase from *Pseudomonas aeruginosa*: (a) UV detection at 254 nm; (b) corresponding chromatogram using ICPMS detection showing distribution of Zn and Ca. (Reproduced from I. Leopold, B. Fricke, *Anal. Biochem.*, **252**, 284 (1997) by permission of Academic Press.)

of recombinant human brain-derived neurotrophic factor on butyl columns by varying the temperature and incubation time in acidic RP eluents, and found that the data supported a process wherein the dimeric protein is first attached to the surface, then dissociated into its monomeric units, which are subsequently unfolded. The unfolding mechanisms were different in acetonitrile and 1-propanol eluents. Purcell et al.[799] used peptides as probes with butyl and octadecyl RP stationary phases over a wide range of temperatures, and found that the bandwidth could be correlated with secondary structure and dynamics of conformational changes. Richards et al.[277] studied the retention behavior and conformational states of cytochrome *c* of different origin in RPHPLC on C_4 and C_{18} columns by using aqueous TFA/acetonitrile eluents at different gradient times and elution temperatures. In a simultaneously published work, the same group[800] made use of variations in the 4σ bandwidths of the eluted peaks to study how the secondary and tertiary structures contributed to the interaction of these proteins with the stationary phases.

Hodges et al.[801] correlated the retention differences of two synthetic amphiphatic α-helical peptides differing in three residues (valine and isovaline, respectively) and on a C_8 column with CD measurements of their helical stability. A study on how the recoveries of albumin, fibrinogen, and IgG varied with the time the proteins were allowed to be in contact with nonporous strong anion and cation-exchange materials was done by Goheen and Hilsenbeck,[570] by adjusting the salt gradient steepness. They found that recoveries were correlated to the residence times and ascribed their results to an initial weak binding followed by a stronger attachment, whose time dependence is due to the time required by the protein to unfold on the charged surface (Figure 63). Related to the area of protein folding is also the studies of chaperonins, proteins that assist in or accelerate protein folding and prevent aggregation, the purification of which was recently reviewed by Quaite-Randall and Joachimiak.[802]

8.4 Carbohydrates and their Conjugates

Polysaccharides play very important and highly diverse roles in biological systems, ranging from basic energy

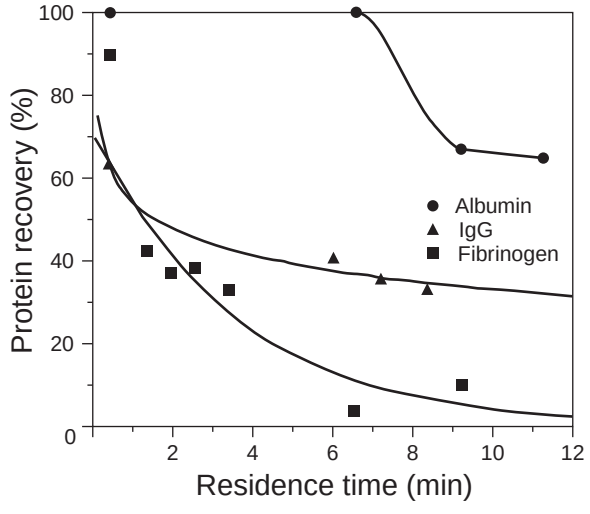

Figure 63 Protein recovery with time on a quaternary amine support. Proteins were introduced individually to the column under the conditions of pH 7.4, 37 °C, and a flow rate of 1.0 mL min^{-1}. Recoveries were determined from peak areas by their absorbance measured at 280 nm. Residence times were varied by either prolonging the gradient time (from 0 to 1 M NaCl in 5 mM TRIS) or maintaining the [NaCl] = 0 condition for various periods of time. Both methods produced similar results. (Reproduced from S.C. Goheen, J.L. Hilsenbeck, *J. Chromatogr. A*, **816**, 92 (1998) by permission of Elsevier Science B.V.)

storage, over structurally reinforcing and impact-damping elements, to complex signaling and recognition systems. Furthermore, biologically derived carbohydrate macromolecules have a large economic importance in industry, as rheological modifiers in the food and cosmetic sectors, and as a basis for the paper pulping and the vegetal fiber textile industries. Process and quality control in these industries calls for ways of isolating and characterizing polysaccharides, which are often of very high molecular weight, with respect to differences in composition, branching, and charge. Polynucleotides are, although treated separately here, in principle also carbohydrate polymers, although their pendant nucleotide bases provide other possibilities for separation. In other areas of the bioscience, substantial interest is presently devoted to glycoconjugates of polypeptides and lipids. Their characterization requires techniques for separating relatively small polysaccharides cleaved from the conjugates, and also of smaller oligosaccharide fragments and single monosaccharides after enzymatic cleavage. The differences sought in these compounds are often subtle and highly demanding on the separation chemistry. Techniques used for studies of carbohydrates are consequently highly diverse.

The basic property of polymeric carbohydrates is their monomeric structure, i.e. the kind and sequence of sugars incorporated. Branching is also central, as is charge derived from the presence of carboxylic groups, sulfate moieties, or amino groups on the sugars. When the polysaccharide reaches a certain length, these properties will have a large impact on its behavior and also its HPLC separability. Very large polysaccharides are almost exclusively separated by HPSEC, and the first problem facing the analyst is to ascertain that the sample is properly dissolved, and that the association between chains is minimal in the separation solvent. Most large polysaccharides will dissolve in water and can be separated in aqueous eluents, although both the dissolution and separation will often require heating to overcome associations that manifest in gelation behavior. Although silica can be used for the separation of many carbohydrates, it will not withstand elevated temperatures over a long time. Polymeric stationary phases are therefore preferred. Solute dissolution usually takes place quite readily when the polysaccharides are charged or highly branched, whereas uncharged polysaccharides with little or no branching form crystalline structures that require more vigorous conditions for dissolving and keeping the polymers from aggregating in solution. For example, cellulose and chitin take several days to dissolve by swelling in *N*,*N*-dimethylacetamide/LiCl. This is the only solvent system known to dissolve underivatized cellulose, and is also used in the HPSEC separation[803] (Figure 64).

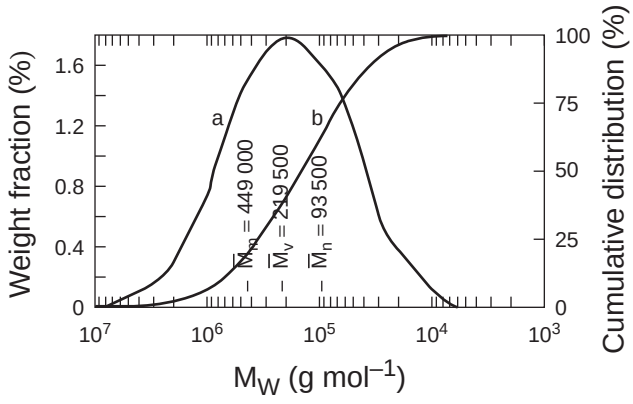

Figure 64 Size-exclusion chromatographic (a) mass fraction and (b) cumulative M_r distribution plots for cellulose linters powder dissolved in the LiCl/DMAC solvent system. (Reproduced from M. Strlic et al., *J. Chromatogr. A*, **805**, 96 (1998) by permission of Elsevier Science B.V.)

Another problem in SEC of high-M_W charged polysaccharides is their high intrinsic viscosity, caused by their expanded configuration in solution. Macromolecular crowding will occur if the injected concentrations are too high in a SEC separation, due to entanglement of the polymers chains approaching the pore system. This solute-induced restricted access to the pore space produces peak distortions, which effectively maximizes the sample concentration. For very high molecular weights the maximum acceptable concentration can be as low as 0.02%. With injected concentrations this low detection becomes difficult, augmented by the absence of natural chromophores in polysaccharides. Introduction of a detectable group, such as a fluorescent group, onto high-molecular-weight polysaccharides by derivatization has yielded only limited success, because the labeling becomes less efficient as the molecular size increases.[805] Molecular weight calibration of polysaccharides in SEC is furthermore a source of concern. The shape factors of polysaccharides vary quite strongly, so different polysaccharide molecular weight standards are not directly interchangeable. Pullulans and dextrans can be acceptable as calibration standards for compounds with similar solution behavior, but a better alternative to external calibration is to use a MALLS detector (see above), or a combination of some other scattering and viscosity measurements, to determine the molecular weight on-line. Carbohydrate separation by SEC, including considerations for selection of columns, has been reviewed by Churms.[806]

More separation options are available for oligosaccharides and smaller polysaccharides. Although sugars are very polar, RPHPLC of uncharged oligosaccharides is possible on regular alkylsilane columns. However, the retention for underivatized oligosaccharides is so low that almost totally aqueous eluents have to be used,

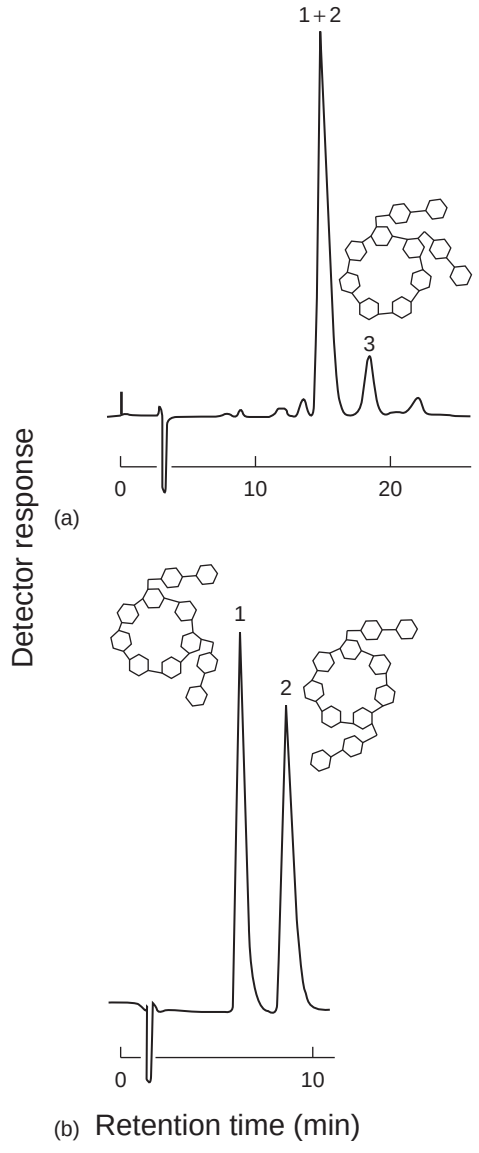

Figure 65 Reversed-phase separation of positional isomers of dimaltosyl-β-cyclodextrin on (a) a YMC-Pack A-312 ODS column with methanol–water (6 : 94) at 35 °C and (b) a Hypercarb with acetonitrile–water (17 : 83) at 50 °C. Peaks: $1 = 6^1$, 6^3-, $2 = 6^1$, 6^4-, and $3 = 6^1$, 6^2-di-*O*-(α-maltosyl)-β-cyclocextrin. Flow rate 1 mL min^{-1}. RI detection. (Reproduced from K. Koizumi, *J. Chromatogr. A*, **720**, 124 (1996) by permission of Elsevier Science B.V.)

conditions that are known to result in phase collapse and eventual expulsion of eluent from the pore system. The bonded phase density also has a profound effect on retention, as does the presence of uncapped silanol groups.[227] Dissociation of carboxylic carbohydrates such as pectin and its fragments can be suppressed by the use of acidic eluents.[807] The retention for acidic carbohydrates can be increased by formation of ion pairs[232] with cationic tetraalkylammonium ions, in a process that can be regarded either as the result of partitioning of hydrophobic ion pairs formed in the liquid phase, or a dynamic coating of hydrophobic cations onto the RP sorbent, effectively transferring it into a cation-exchange sorbent. Porous graphitized carbon (PGC) offers separation according to the RP principle. The retention is higher compared to alkylsilane columns and the PGC columns have a more pronounced selectivity, which appears to be substantially influenced by the solute shape. The resolution of anomeric forms and closely related oligosaccharides is therefore possible,[808] as illustrated in Figure 65.

Of opposed selectivity to RPHPLC is HILIC, practiced on polar stationary phases with aqueous eluents containing relatively high concentrations organic solvent, typically acetonitrile. A wide variety of stationary phases have been used for HILIC separation of diverse carbohydrates, ranging from monosaccharides to relatively large sialic and asialic oligosaccharides. The retention shows a marked influence on the nature and position of sugar units, and HILIC is therefore valuable in separations of hetero-oligosaccharides, exemplified by an ability to separate recombinant human interferon-γ based on site-specific glycosylation microheterogeneity (Figure 66). The use of HILIC for separation of carbohydrates has been reviewed.[220]

Classical anion-exchange chromatography has been used for a long time in the separation of carbohydrates.

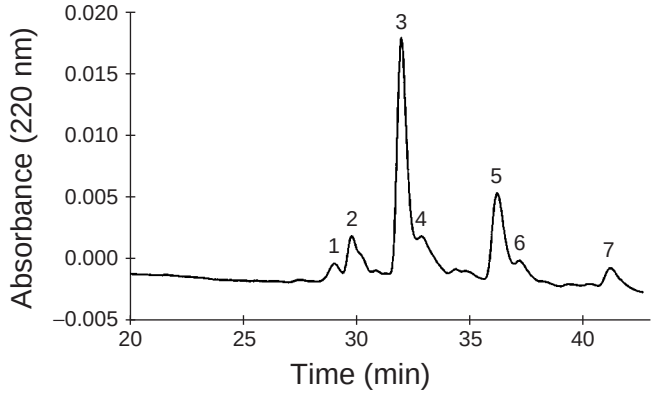

Figure 66 A HILIC chromatogram of tryptic glycopeptides of Asn^{25} from recombinate human interferon-γ on a polyhydroxyethyl aspartamide column (150 × 1 mm i.d.). The glycopeptide fractions from a preceding RP separation step were mixed with mobile phase A (85% acetonitrile, 15% water, 10 mM triethylamine; pH 6.0) and then loaded onto the HILIC column at 0.1 mL min^{-1}. Gradient elution: 50 μL min^{-1} with a 60-min gradient from 100% mobile phase A to 100% mobile phase B (20% acetonitrile, 80% water, 10 mM triethylamine, 25 mM $NaClO_4$; pH 6.0). Detection: UV, 220 nm. Manually collected fractions were identified by MALDI/TOFMS. (Reproduced from J. Zhang, D.I.C. Wang, *J. Chromatogr. B*, **712**, 77 (1998) by permission of Elsevier Science B.V.)

The anionic properties of sialic carbohydrates provide for direct separation, but neutral carbohydrates need to be complexed by borate to form anionic species that are retained on anion-exchange resins. This technique is now largely obsoleted by high-pH anion-exchange chromatography (HPAEC) using strongly alkaline eluents to promote a dissociation of the carbohydrate hydroxyls, so that neutral sugars are converted into anionic species.[596] Thus, HPAEC is applicable to both sialic and asialic oligosaccharides without the need to add a separate complexing agent. Hydroxide ion is a weak eluent, so solutes are eluted from the column by an increasing concentration of a pushing ion, typically acetate or nitrate (Figure 67). The retention is largely a function of the number of hydroxyl groups, corresponding to the degree of polymerization. Obviously, high-pH eluents necessitate the use of hydrolysis-stable polymeric columns. An advantage of the alkaline eluent is that the separated sample can be detected by a PAD detector (see above) without postcolumn addition of base. The disadvantages of high-pH eluents are epimerization and degradation of the samples, to which there is no good solution. Still, HPAEC has secured a central position among modern HPLC techniques for oligosaccharide separation. Sialylated oligosaccharides can be separated from asialics by elution at a lower pH, although base must then be added to enable detection by PAD. The HPAEC technique has been reviewed by Lee.[567] Hydroxyapatite is also an ionic separation mode applicable to the separation of glycosaminoglycans.[615]

Affinity chromatography of carbohydrates is possible either through the formation of cyclic boronate esters with aminophenylboronic affinity phases, which is a technique particularly valuable for sample work-up and enrichment, or by the use of immobilized lectins[437] (see affinity section). Glycoaffinity chromatography and its applications in studies of biological recognition has been reviewed by Caron et al.[811]

Absence of detectable groups is a concern not only for high-molecular-weight carbohydrates, as outlined above. Saturated carbohydrates absorb light only at very low UV wavelengths, where the chemical noise from other sample compounds and solvent contamination becomes problematic. Monounsaturated sugars produced by enzymatic cleavage have their absorption wavelength somewhat shifted, but still in the low UV. Derivatization reactions aimed an enhancing the detectability of oligo- and polysaccharides have therefore been employed for many years. Derivatization could aim at any of the polar groups in polysaccharides, but this is not practical for several reasons, except in GC and SEC where the purpose is to reduce the polarity of sugars and oligosaccharides.

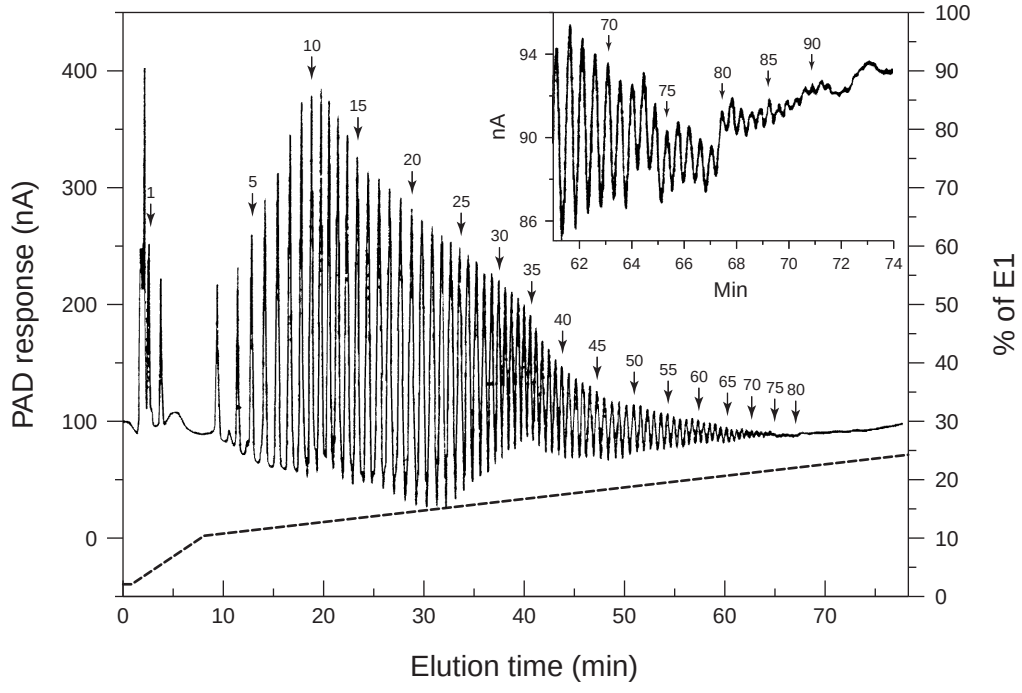

Figure 67 Separation of Neu5Ac oligo/polymers derived from colominic acid (equivalent to 100 μg) on a CarboPac PA-100 column using an eluent containing 0.1 M NaOH with a superimposed 1 M $NaNO_3$ (E1) gradient, as indicated in the figure. Detection was by pulsed amperometry, using a gold working electrode. The inset shows an enlarged view of the retention time interval between 61 and 74 min. Peaks are labeled with degree of polymerization. (Reproduced from Y. Zhang et al., *Anal. Biochem.*, **250**, 247 (1997) by permission of Academic Press.)

Instead the reducing end of the polysaccharide is targeted,[812] which is also used to deal with a separation problem in interactive chromatography of small sugars, namely the splitting of α- and β-anomers and their on-column isomerization leading to badly shaped double peaks.[813] This can be eliminated by derivatization schemes using reductive amination (see Figure 68). Introduction of a nonpolar group in the terminal end of the oligosaccharide also has a positive effect by increasing the interaction with RP sorbents. Postcolumn derivatization as a means of modifying carbohydrates for spectrophotometric, fluorimetric, or electrochemical detection in HPLC has been reviewed by Honda.[814]

In glycosylated proteins, the borderline between a protein and a polysaccharide is diffuse, indicated by their

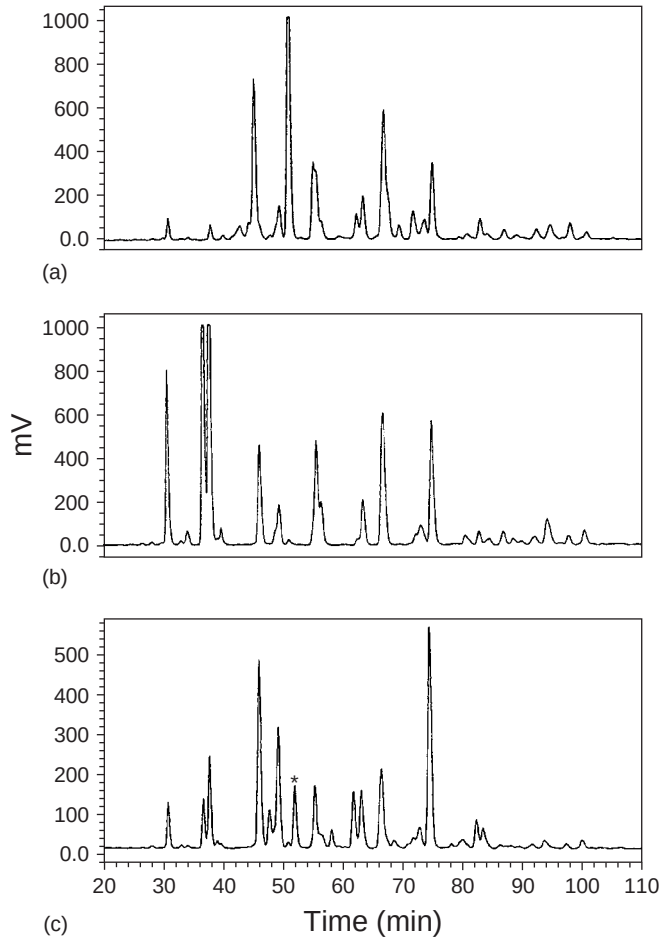

Figure 68 Normal-phase HILIC chromatography of oligosaccharides in human milk derivatized by reductive amination with 2-aminoacridone on a GlycoSep N column (250 × 3.9 mm i.d.): (a) blood group O, (b) blood group A, and (c) blood group O. Mobile phases: A, acetonitrile; B, 250 mM ammonium formate, pH 4.4. Gradient: 20–58% B over 152 min. Fluorescence detection at 428/525 nm. (Reproduced from J. Charlwood et al., *Anal. Biochem.*, **273**, 275 (1999) by permission of Academic Press.)

nomenclature. Glycoproteins comprise anywhere from 1% to 80% of their weight as oligo- and polysaccharides attached to their surfaces, which should intuitively lead to a substantial change in their interaction patterns. The influence of this carbohydrate modification is surprisingly low, however, and glycoproteins are therefore isolated on the basis of their protein properties (charge and hydrophobicity), rather than on the basis of their carbohydrate properties.[227] Affinity chromatographic schemes,[760] hydroxyapatite chromatography[816] (Figure 69), or CE[817] are useful when glycosylation pattern heterogeneities are sought. The use of HPAEC for separation of glycopeptides obtained in tryptic maps was recently review by Townsend et al.[818] Proteoglycans, which have a protein core – usually a glycoprotein – may contain up to 95% oligo- and polysaccharides[819] of 200 or more disaccharide units, and reach molecular weights in excess of 1 MDa.[820] Their extraction from tissue requires special consideration and separation is mostly carried out on SEC or nonHPLC techniques.[820] Glycosaminoglucans (GAGs) – the oligosaccharide chains of proteoglycans – are composed of alternating uronic acid and *N*-acetylhexosamine residues, which give the molecules a net negative charge, augmented by sulfate groups attached to hydroxyls. According to Alberts et al.,[819] these oligosaccharides have an almost limitless heterogeneity potential, which obviously makes chromatographic separations and other means of characterization difficult. The complicated biochemical pathways involved in the biosynthesis of oligo- and polysaccharides attached to lipids and proteins are indicators of their important function, but most of these biological functions are still not known.[821] Glycobiology is consequently an important and growing area in molecular biology, dealing with elucidation of the cellular functions of glycoconjugates. The first steps in this work is to characterize the carbohydrate moieties. Although modern MS techniques have evolved into extremely powerful tools for glycobiology,[822] the complexity of carbohydrates and their conjugates still calls for a battery of separation and detection steps coupled with chemical and biological modifications to accomplish the most complicated characterizations. The use of CZE[823,824] and supercritical fluid chromatography[825] has been reviewed and is not covered here.

Characterization of oligosaccharides conjugated to other biomolecules requires that they are detached from their protein, peptide, or lipid components, a feat nowadays accomplished by enzymes. Prior to cleaving the carbohydrate portion from glycoproteins, these are usually broken down into peptides by proteases, so that steric hindrance in the enzymatic detachment is minimized. Detachment of N-linked oligosaccharides is feasible by several enzymes, either at the

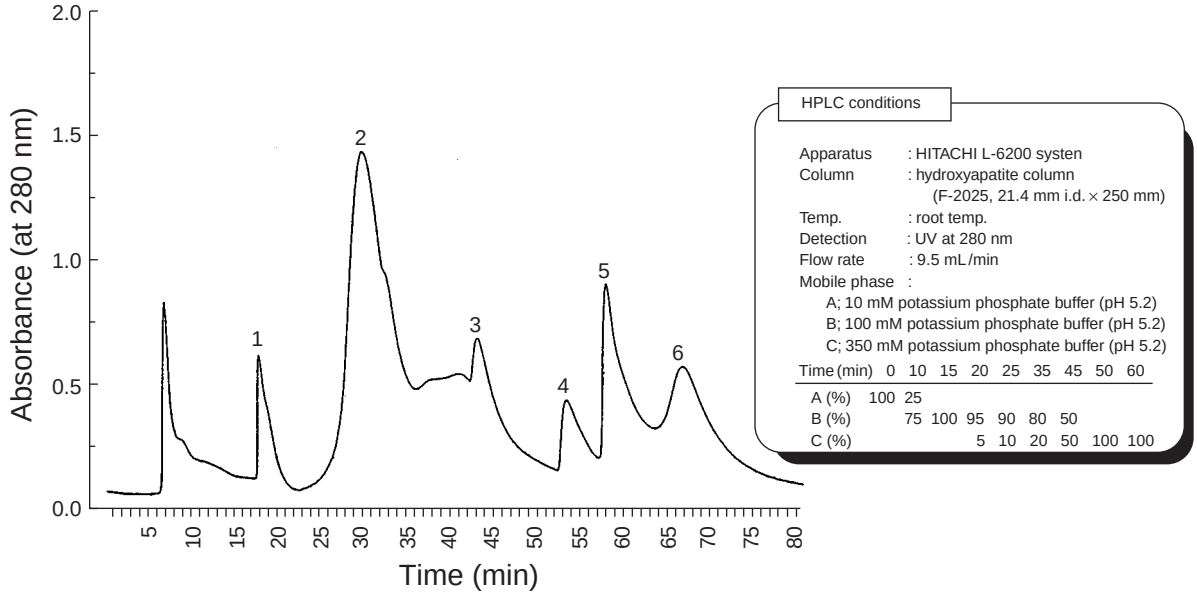

Figure 69 Chromatogram of the glycoforms of α_1-acid glycoprotein from the serum of healthy subjects separated by high-performance hydroxyapatite chromatography. For conditions, see inset. (Reproduced from S. Kishino et al., *J. Chromatogr. B*, **703**, 3 (1997) by permission of Elsevier Science B.V.)

Asn linkage by glycosamidase resulting in an intact oligosaccharide, or by an endoglycosidase leading to an oligosaccharide deprived of the *N*-acetylglycosamine linkage. No well-characterized enzyme systems are available today that are capable of general release of O-linked oligosaccharides.[826] Released fragments are usually subjected to further enzymatic cleavage by exo- or endoglycosidases of high specificity, targeting particular glycosidic linkage patterns in the oligosaccharide, whereafter the fragments are separated and identified.[827] Large amounts of structural information can be produced in a short time if arrays of enzymes are used. The most recent developments in this area dispose of the chromatographic separation step entirely and carry out the enzymatic reactions directly on a MALDI target.[828] Yet modern biomacromolecule chromatography is indispensable in glycobiology, because the purity of the enzymes used is obviously of paramount importance to the reliability of the data produced.

A combination of separation techniques into multidimensional separation schemes is required to elucidate the complicated structures of oligosaccharides in biological systems, and even three-dimensional separation maps with orthogonal selectivities have been described.[221] Several books on carbohydrate analysis have also been published during the 1990s.[829–831]

8.5 Validation of Separation Processes

From the preceding text, it is obvious that HPLC is becoming a viable alternative to many conventional techniques for preparative purification of proteins and other biotechnologically derived products, due to the high speed and separation potential. Whenever a chromatographic technique is adopted as part of a production scheme for a therapeutic formulation, stringent rules apply to the validation of the separation method. Aspects such as assessing the long-term consistency of retention, the possibility of cleaning and sterilizing separation columns in-place and the possible deterioration of phases due to such actions leading to release of leachable compounds are treated in a recent review by Jungbauer and Broschetti.[775]

9 TRENDS IN SEPARATION MEDIA FOR HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY OF BIOMACROMOLECULES

During the 1990s, there has been a tremendous development in new separation media for biological macromolecules, in particular proteins.[96,832] Low diffusion coefficients are the Achilles' heel of large-molecule chromatography, leading to a steep decrease in efficiency resulting from the contribution to factor *C* (Equation 15) of the van Deemter–Knox equation (Equation 12). The activities aimed at improving the separation efficiency have therefore focused on the physical configurations of sorbents to enhance the mass transfer. Much less has been done in the area of chemistries of the separation media.

Efforts to improve mass transfer can be divided into three major areas:

- Decreasing the particle size to diminish the intraparticle diffusion length.
- Avoiding the intraparticle diffusion using active sites on nonporous supports (micropellicular materials).
- Facilitation of the intraparticle mass transfer by convection (Figure 70).[98]

9.1 Nonporous and Micropellicular Separation Media

Supports may be classified according to the spatial distribution of pores as nonporous, superficially porous, and totally porous.[96] Micrometer-sized nonporous silica particles were introduced in the 1980s as a means of alleviating the problem of mass transfer limitations that results from the slow intraparticle diffusion. Although the advantage of elimination of intraparticle diffusion could be demonstrated,[833,834] these nonporous particles have an extremely low surface area compared to porous media. As the separation proceeds mainly at the surface, both very low sample capacity and retention are typical for these packings. Based on a solid SiO_2 density of 2.2 g cm^{-3}, the surface area of a packed bed of 5 µm spherical particles is as low as 0.6 m^2 mL^{-1}, and even if the particle size is decreased to 1 µm, the surface area is increased to only 3 m^2 mL^{-1}. This is about two orders of magnitude lower surface area than the typical porous silicas useful for the separation of biological macromolecules. A comparison of porous and nonporous microparticulate RP silicas used in the separation of two closely related peptides also shows that the pellicular material is more sensitive to changes in the mobile phase due to the overall low coverage with interaction groups.[835] In a study of the effect of packing particle size on column efficiency in C_{18} RP separation of proteins on nonporous and porous silica, and on nonporous silica of different particle size,[836] all columns appeared to have essentially the same efficiencies, due to the binary sorption conditions during the steep gradients (Figures 71 and 72).

There are still substantial advantages inherent in the use of nonporous media. Solid particles are extremely hard and can tolerate very high differential pressure across the packed beds. The nonporous surface with a well-defined topography enables high accessibility of the interaction layer and the absence of pores ensures that separations are not affected by exclusion processes. In specific applications, the low retention can be used to an advantage, as it makes the use of mild eluents mandatory. In combination with a low column void volume, and the inherent high efficiency, columns packed with nonporous particles provide for very fast separations. This appears to be the principal advantage of nonporous microparticles.[832] Fast separation implies that denaturation can be kept at a minimum. Protein separations can be carried out at elevated temperatures

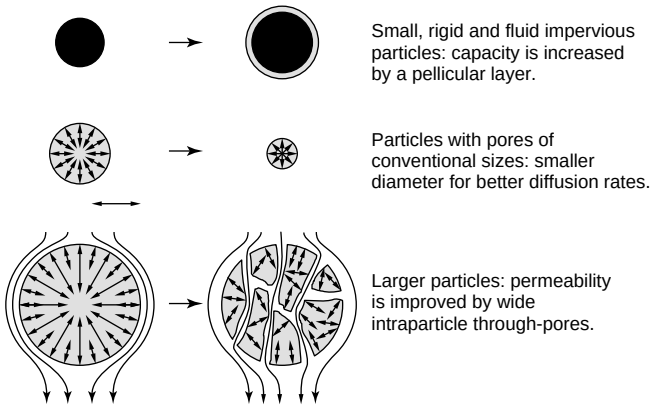

Figure 70 Overview of the routes taken to improve the separation efficiency for biological macromolecules by reducing the mass transfer resistance in the partitioning process. (Reproduced from A.E. Rodrigues, *J. Chromatogr. B*, **699**, 48 (1997) by permission of Elsevier Science B.V.)

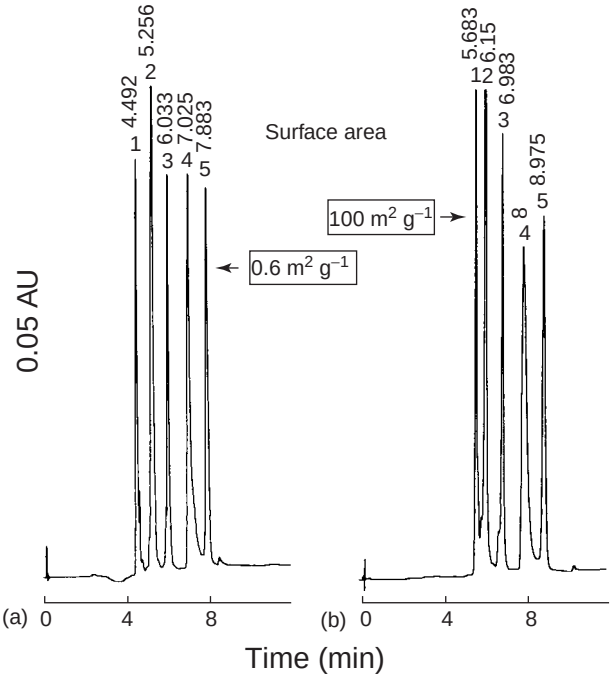

Figure 71 Proteins separated on (a) 5-µm nonporous bonded C_{18}, and (b) 7-µm macroporous bonded C_{18} silica columns by conventional gradient elution at 2 mL min^{-1} and at ambient temperature. Eluents: (A) 0.1% TFA in water and (B) 0.1% TFA in 90% aqueous acetonitrile. Linear gradient from 10% to 100% B in 20 min. Peak identities: 1, insulin; 2, cytochrome *c*; 3, lysozyme; 4, BSA; 5, carbonic anhydrase; 6, insulin B chain; 7, α-lactalbumin. (Reproduced from N. Nimura et al., *J. Chromatogr.*, **585**, 209 (1991) by permission of Elsevier Science B.V.)

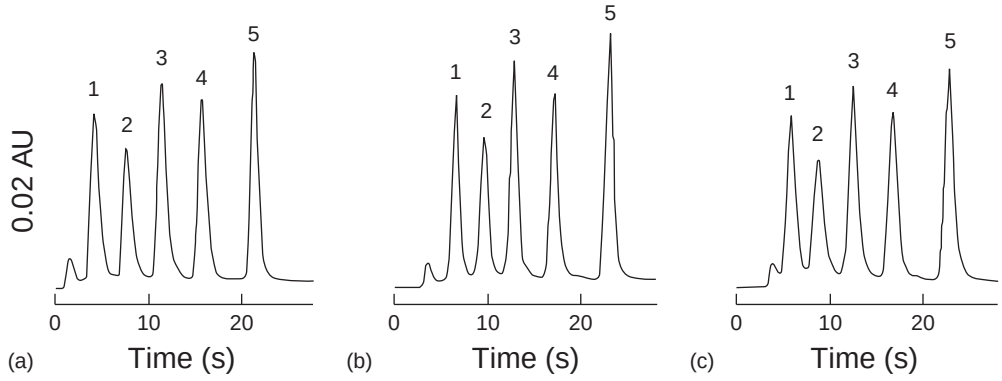

Figure 72 Fast protein separations on bonded C_{18} nonporous silica with particle diameter (a) 2 µm (b) 5 µm, and (c) 20 µm at 4 mL min^{-1} flow rate and 75 °C. Linear gradient from 22% to 100% eluent B (Figure 71) in 48 s. Peak identities: 1, insulin B chain; 2, insulin; 3, lysozyme; 4, α-lactalbumin; 5, carbonic anhydrase. Figures 71 and 72. (Reproduced from N. Nimura et al., *J. Chromatogr.*, **585**, 209 (1991) by permission of Elsevier Science B.V.)

to further enhance the separation. The decreased eluent viscosity promotes both diffusion and interaction kinetics and lowers the back-pressure. An example of a rapid separation under these conditions is shown in Figure 73. It should be noted that this separation is carried out under conditions that are denaturing the analytes even at normal temperature. Thus, the use of high temperature does not alter the situation. However, denaturing does not appear to be a major issue in analytical separation unless the purpose of the separation is elucidation of tertiary or higher protein structure, and can therefore be acceptable as long as the peaks are eluted singly in narrow bands at predictable retention times, and provided that the stationary phases are not irreversibly damaged by precipitated protein.

Obviously, separations on nonporous packings are limited to analytical and micropreparative scale because of the low loading capacity. Another consequence of the low capacity is the need for sensitive detection. Although this is partly compensated for by the high concentration of the analyte in narrow peaks, separations in columns packed with nonporous materials require more sensitive detection techniques.

The development of current micropellicular separation media represents a revival of technology of the early days of HPLC.[832] Many pellicular sorbents of that time were provided with thicker interaction layers compared to the current monolayer functionalized silicas. Today, the term micropellicular separation media is used to describe all smooth, nonporous particles that are monolayer functionalized,[838] or covered with polymer shrink-wraps,[135,142,143] superficially porous layers,[839] polymer tentacles,[163] and electrostatically attached layers of polymeric latex particles.[840] Pellicular layers can also have a polymeric core and an inorganic interaction layer.[15] Monolayers have little effect on increase of

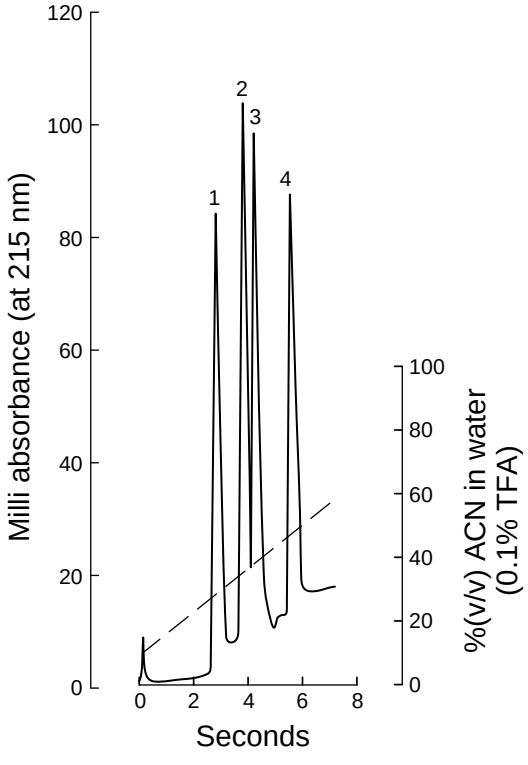

Figure 73 Separation of (1) ribonuclease, (2) cytochrome *c*, (3) lysozyme and (4) β-lactoglobulin on a 30 × 4.6 mm i.d. column packed with 2-µm nonporous octadecyl silica using a 12-s gradient from 10% to 90% (v/v) acetonitrile (ACN) in water containing 0.1% TFA at a temperature of 120 °C and a flow rate of 5 mL min^{-1}. Column inlet pressure 240 bar. (Reproduced from H. Chen, Cs. Horváth, *Anal. Methods Instrum.*, **1**, 213–222 (1993) by permission of John Wiley & Sons.)

the loading capacity and merely serve the purpose of outfitting the surface area with a suitable interaction layer. This is similar to the processes used for functionalization of internal surfaces of porous beads. The same applies

to smooth and dense polymer shrink-wraps regardless of their thickness, as biomolecules typically interact only with the surface. However, the difference in loading capacity compared to a porous sorbent is lower than may be expected, because the interior surface of a porous material would be largely inaccessible to a slowly diffusing protein at the separation speeds used with nonporous microparticles,[100] evident also in Figure 71. A majority of the polymer coating techniques described earlier in the text have been used for coating smooth microparticles.

Most separation chemistries have been implemented on micropellicular media except, for obvious reasons, SEC. Unger showed that RP separation of proteins and peptides can be done with high efficiency on C_{18}, C_8, C_4, C_2, and phenyl-modified monodisperse 1.5-μm nonporous silica,[838] and that hydrophobic interaction chromatographic interaction layers can be established on these supports activated by diethylene glycol allyl methyl ether and acetamidopropyltriethoxysilane.[841] The recovery of proteins (pepsin, catalase, and horseradish peroxidase) in their native state on the RP materials showed the expected dependence on ligand chain length, solvent admixture, and contact time.[842] Coating with PBD was later shown to be a useful alternative for providing a hydrophobic coating on nonporous silica.[135]

Examples of affinity chromatographic separations are glycoproteins and Igs on columns with immobilized protein A, concanavalin A, or wheat-germ agglutinin,[100] and the separation of cytochrome *c*, lysozyme, and chymotrypsinogen A by rapid metal-interaction chromatography on IDA ligands charged with Co^{2+}, Fe^{3+} or Ni^{2+} ions.[843] Preparation of micropellicular silica phases with RP, anion exchange, and hydrophobic interaction surface layers and their use for rapid separation of proteins was demonstrated by Kalghatgi.[844] A micropellicular sorbent for HILIC based on aminofunctionalized nonporous silica was recently used by Lin and Lee for separation of a fructo–oligosaccharide mixture.[215]

Micropellicular sorbents prepared from 3-μm cross-linked polystyrene beads were used for separation of proteins in RP mode at different pH and temperature by Maa and Horváth.[845] Although some swelling was observed in organic solvents, typical RP eluents did not have any negative effect on the efficiency, which was found to be best at the extreme ends of the tested pH scale. Polystyrene beads were also functionalized with DEAE groups and used in anion-exchange mode for fast separation of DNA fragments with gradients of aqueous buffers as eluents.[846] Letourneur et al.[847] converted nonporous polystyrene into phosphate-type ion exchangers. The pellicular separation media most widely used are the ion exchange resins with agglomerated latex coating first described by Small et al.[840] for the separation of small ions by IEC. Modern tentacle-type sorbents with application areas directed towards separations of biomacromolecules[575] originated from this work.

A different type of nonporous medium, deswollen and cross-linked agarose beads, was described by Hjertén and Liao.[848,849] These particles are prepared by suspending macroporous agarose beads in a solvent mixture that contains dioxane and chloroform, followed by cross-linking of the shrunken beads with 1,4-butanediol diglycidyl ether. This treatment leads to particles with diameters as small as 3 μm, with sufficient mechanical stability to be packed and used for separation. However, the cross-linking renders the agarose more hydrophobic and makes it suitable for HIC.[848] Treatment with glycidol enhances the hydrophilicity of the surface. Functionalization to an anion exchanger and its use in IEC of a number of proteins was also demonstrated[849] (Figure 74). Yet another variety of nonporous support are 0.5-μm quartz fibers, which were silylated with mercaptopropyltrimethoxysilane and coupled with tresyl chloride-activated dextran by Wikström and Larsson to afford affinity supports. After derivatization with a NAD (nicotinamide adenine dinucleotide) ligand, this medium was used for rapid processing of enzymes from dilute solutions.[310]

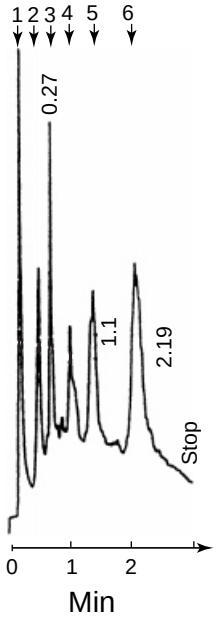

Figure 74 Separation of proteins on a compressed, nonporous anion exchange column (62 × 6 mm i.d.; 15 μm) prepared by reacting with dimethylamine a macroporous agarose that had been rendered nonporous, cross-linked, and activated by 1,4-butanediol diglycidylether in solvent. Linear gradient over 2.5 min from (A) 10 mM TRIS-HCl, pH 8.5, to (B) 180 mM sodium acetate in (A), at a flow rate of 4 mL min^{-1}. Proteins: 1, myoglobin; 2, haemoglobin; 3, transferrin; 4, ovalbumin; 5, serum albumin; 6, phycoerythrin. (Reproduced from J.L. Liao, S. Hjertén, *J. Chromatogr.*, **457**, 180 (1988) by permission of Elsevier Science B.V.)

Two reviews have been published on the use of nonporous stationary phases in biomacromolecule separations. Lee[97] focuses on the preparation and functionalization of nonporous sorbents and the dynamics of protein separation and offers a comprehensive listing of commercially available nonporous sorbents; the topic of Huber[715] is separation of double-stranded DNA on micropellicular stationary phases by anion-exchange or ion pair chromatography. Chen and Horváth placed micropellicular sorbents in a wider perspective of HPLC development in their review of high-speed HPLC of peptides and proteins.[832]

9.2 Perfusion-type Particles with Large Open or Gel-filled Pores

As outlined in the introductory part of this article, the main obstacle against high efficiency separations, with a sample capacity sufficient for preparative applications on conventional porous sorbents, is the intraparticle mass transfer resistance, i.e. the pools of stagnant mobile phase that the solutes have to move across in order to interact with the internal pore surfaces. One way of enhancing the mass transport in porous media, through augmented diffusivity by convection, was described in the chemical engineering literature in the early 1980s.[850] After the concept was applied to protein separation media in the late 1980s,[851] the technique became known as perfusion chromatography. The underlying principle is quite straightforward. By ascertaining that the sorbent has a sufficiently large pore size, the differential pressure across the particle forces part of the eluent through the sorbent particle, in addition to the flow around it through the interstitial voids. This convective transparticle flow enhances the mass transfer through the macroporous space within the beads, so that the solutes are brought more quickly into contact with the interacting sites. A transparticle flow of only a few percent of the total flow is sufficient to increase the efficiency substantially. Characteristic of convective transport is that the factor C in the van Deemter–Knox equation levels off at a critical flow rate, and that the efficiency beyond that point becomes nearly independent of the eluent flow velocity. In other words, the column throughput per unit time increases almost linearly with eluent flow rate, without a band-broadening penalty. Another beneficial effect of the more efficient mass transport in the convective regime is that, unlike conventional porous sorbents, the useful dynamic capacity of the sorbent is preserved at high flow rates. The theory of convection-enhanced internal mass transport in particles has been worked out by Rodrigues et al.,[98,850,852–859] Liapis et al.,[860–871] and by Carta et al.[855,872–876] For example, the extended version of the van Deemter–Knox equation (the Rodrigues equation) adequately describes the mass transfer enhancement for biomacromolecules in perfusive sorbents (Figure 75).

Perfusive sorbents are typically used as relatively large particles (10–40 μm) in small to medium scale preparative separations. They can therefore be operated in the augmented transport regime at high linear flow rates without excessive back-pressure. The most valuable feature of perfusion chromatography is the speed of separation, which allows high throughput processing using relatively small columns operated in automated systems.

There are currently two different types of augmented diffusivity materials on the market, contrasting with respect to the medium occupying the pore system. The material first introduced[851] is manufactured from a megaporous styrene–divinylbenzene sorbent with a bimodal pore size distribution. These sorbents contain very wide pores with a diameter of 600–800 nm transecting clusters of spheroid microglobules containing 50–150 nm small diffusion pores that also account for the majority of the interaction surface area. Another material, introduced a few years later, is a soft-gel-in-a-rigid-shell aerogel–xerogel composite, comprising a rigid hyperporous polystyrene–silica composite that has the pore space filled with a lightly cross-linked soft hydrogel, which is also the carrier of the interactive groups.[877] The presence of a hydrogel inside the pore system increases the capacity[878] and lowers the eluent flow rate where the transition from the diffusive to the convective operation takes place, so that the efficiency is independent of flow rate in practically the entire operational range.[856]

Superporous agarose, prepared by Gustavsson and Larsson[879] using the double emulsifier technique, has also the attributes of a perfusion sorbent, although the size of both particles (300–500 μm) and the through-pores

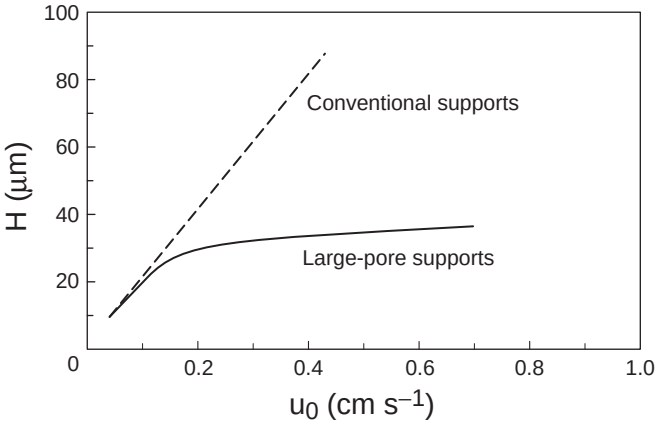

Figure 75 Typical flow rate dependence of the height of a theoretical plate (H) for conventional porous and perfusion supports. (Reproduced from A.E. Rodrigues, *J. Chromatogr. B*, **699**, 50 (1997) by permission of Elsevier Science B.V.)

(25–75 μm) are considerably larger than those of the rigid synthetic porous polymers discussed above. For this reason it hardly qualifies as an HPLC sorbent. In fact, the intraparticle superpores are so large that the convective flow can be monitored by videomicroscopy of dyed microparticles moving through the pores.[880] The advantages of intraparticle flow within these beads have been demonstrated in ion-exchange,[879] hydrophobic interaction,[881] and affinity chromatography.[882] A similar material was also used for the preparation of continuous monolithic beds[883] (see below). The applications of superporous agarose are in large-scale purifications.

9.3 Media with Reduced Discontinuity – Membranes, Rolled Sheets, and Monoliths

Although the lowest theoretical interparticular volume of perfectly packed uniformly sized spherical beads is calculated to be about 27% of the total column volume, the void volumes of even the best packed columns are in practice substantially larger. Separation in a medium which is essentially a single piece and does not contain interparticular voids normally contributing to peak broadening has been treated theoretically[884] but experimental work remained scarce as a result of the lack of better alternatives to beads.

The problem of interparticular volume does not exist in systems in which a membrane is used as the separation medium. Both theoretical calculations and experimental results[283] clearly demonstrate that membrane systems can be operated in a dead-end, or filtration mode, at much higher flow rates than packed beds, because all the substrate solution flows through the support and the mass transfer is much faster as a result of this convective flow. Obviously, a reasonable compromise between the membranes with their fast mass transfer and beads with high capacity may result in separation media with enhanced properties.

The simplest available porous membrane is filter paper made of pure cellulose fibers. Numerous modification methods have been described for this material and applications of microcrystalline cellulose powders in chromatography of biopolymers are well known. These may be the most important reasons for the use of modified cellulose sheets (membranes) in various separation devices. As the binding capacity of a single sheet would be too small, commercial membrane cartridges contain several membranes stacked and placed in specifically designed cartridges. The pore size of these membranes is an important variable that effects the separation of proteins.[885] As a result of their large width and low height, membrane cartridges allow for high volumetric throughputs at low back-pressure. The residence time within the cartridge is generally short and thus the danger of denaturation of biopolymers during the separation process is largely reduced. Numerous analytical and preparative separations of proteins in various chromatographic modes have been performed in membrane cartridges with very good results.[78,886,887]

Theory predicts a dramatic increase in efficiency for adsorptive membrane devices compared to that of beds packed with particles, perhaps by a factor of 10.[887] However, the nonuniform flow and dead volumes typical of many current cartridges[888] do not fully utilize the intrinsic separation ability of the membranes, and therefore their design must be improved to compete successfully with packed columns. The problem of distorted flow and voids between the membranes in the stack does not exist in cartridges that accommodate only a single thicker disk of porous material. Additionally, Belenkii's theory of critical distance in adsorption chromatography leads to a conclusion that even a very short layer of a separation medium such as the disk can be sufficient for an efficient separation of large molecules.[889] These provisions led to the development of rigid disks for rapid chromatography of biopolymers in the late 1980s.[77]

Flat slabs and rods of porous material are easily prepared in a single step by free-radical polymerization within a closed mold. Disks are cut with a blade from the sheets or machined from the rod and placed into special cartridges. If needed, the porous material can be provided with desired functionalities by chemical modification. Typically, the molded material is wetted with the solution of a reagent and the modification is performed in a flask.[890] Josic et al. studied numerous applications of rigid macroporous disks[891,892] and found that a well-designed cartridge in combination with optimized disks dramatically improved the performance and allowed for a rapid chromatography of proteins. The separations of protein mixtures in different chromatographic modes, such as ion-exchange and hydrophobic interaction, were easily achieved in a few seconds[892] (Figure 76). Macroporous disks using this technology have been commercialized.

Disks of regenerated cellulose provided with ion-exchange functionalities and placed in disposable cartridges are also commercially available. These devices are characterized by very large through-pores that allow direct purification of dirty feedstocks such as homogenized yeast cells at high flow rates.[893]

Ladisch fabricated stationary phases prepared from rolled flat soft materials–woven textile fabric.[894] The cylindrical roll is inserted into a typical HPLC column blank. The number of chemistries is limited by the variety of fibers. Aramid columns behave as a hydrophobic interaction phase, whereas cotton fabrics transformed to an ion-exchange derivative can be used in IEC.[895] Rolled columns are very stable and have much better flow

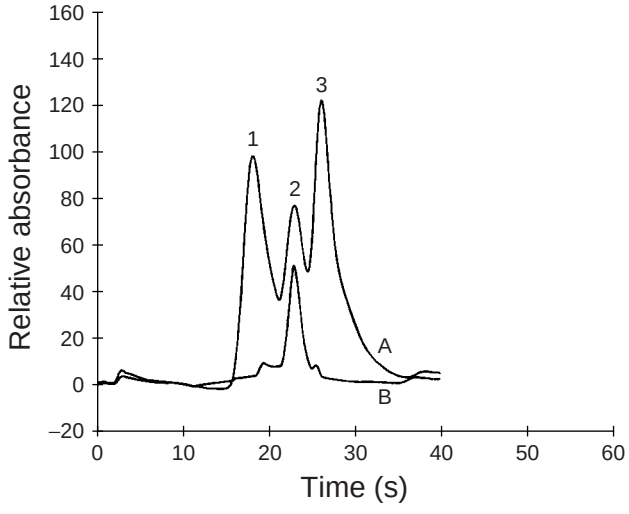

Figure 76 Separation on a rigid macroporous disk applied to in-process analysis during α_1-antitrypsin (AAT) production. (A) A sample containing 5 mg mL^{-1} of AAT, with 5 mg mL^{-1} each of HSA and transferrin (Tf) as impurities. (B) Purified AAT (5 mg mL^{-1}). The separation element was a 25-mm diameter, 3-mm thick GMA-EDMA disk with DEAE functionality. Buffer A, 10 mM TRIS-HCl, pH 7.4; Buffer B, 1 M NaCl in buffer A. Gradient: 0–100% Buffer B in 30 s at 5 mL min^{-1}; back-pressure, 800 kPa; room temperature. Detection: UV, 280 nm. Peak identities: 1, Tf; 2, AAT; 3, HSA. (Reproduced from A. Strancar et al., *Anal. Chem.*, **68**, 3486 (1996) by permission of American Chemical Society.)

properties than columns packed with less than 1-mm short fibers.[896] As expected, the efficiency (plate height) does not change with flow velocity because the mass transfer is substantially enhanced by convection of the mobile phase through the stationary phase. Kennedy and Paterson used a similar approach and rolled a rod from a paper-like cellulose sheet.[897] After cyanogen bromide or CDI activation, Protein A and Protein G were immobilized onto the cellulose and the column was successfully used for the affinity separation of IgG.

Slightly cross-linked polyacrylamide beads have been known for a few decades as excellent packing for the low-pressure LC of biomacromolecules. The general perception is that their swelling and softness prevent them from use in tall chromatographic columns. However, a study of the resolving power of another soft gel material – agarose beads – revealed that compressed beds of these particles had better efficiency than the original columns.[105] Hjertén et al. transformed this experience into the preparation of highly cross-linked (30–60%) methylenebisacrylamide copolymers with functional monomers that were packed into columns and compressed to a fraction of their original volume before use in chromatographic separations.[127] Although plugs have a porosity of about 66%, this porosity mainly represents large pores, whereas the highly cross-linked and compressed nodules are nonporous for proteins.[898] Therefore, separations of large molecules are rather fast. This technology was later extended to the preparation of microcolumns. Materials for chromatographic separations in different modes are obtained either directly by copolymerization that involve different charged or hydrophobic monomers or using in situ modifications.[899] These compressed beds are now commercialized.[900,901] Both quaternary ammonium and sulfonic acid columns for IEC are claimed to have good stability and a high dynamic binding capacity that does not change with an increase in flow velocity. For example, the strong cation-exchange type binds 42 mg of bovine IgG per milliliter of column volume.

Another procedure in which rigid macroporous materials of any desired shape are prepared in their final size is in situ polymerization, developed in the early 1990s.[902] As can be implied, these separation media are not packed but molded within the confines of the tube of a chromatographic column. In this procedure, the column tube (sealed at one end) is filled with the polymerization mixture; then it is sealed at the other end and heated in a water bath. Once the polymerization is complete, the seals are removed and the column is provided with fittings, attached to the HPLC system, and washed. Obviously, an ideal monolith should contain both large pores to achieve a mobile phase flow through the column at a reasonable back-pressure (convection), and a connected network of shorter and smaller pores for diffusion. Indeed, an extensive study of the effects of individual variables of bulk polymerization on the porous properties of products, with particular attention paid to temperature (polymerization kinetics) and porogen composition (phase separation process), led to highly porous monolithic materials with a bimodal pore size distribution that is completely different from that found for macroporous beads prepared from identical polymerization mixtures. Typically, these monoliths contain not only small diffusive pores, but also a large number of flow-through pores with diameters in the range of 700 to 2000 nm.[903,904] This structure accelerates mass transfer of the sample within the porous material by convection, resulting in faster chromatographic runs without sacrificing the separation power of the monolithic column.

Poly(styrene–divinylbenzene) monoliths are well suited for the rapid separation of proteins in the RP mode with an efficiency on par with micropellicular materials, although at a substantially lower back-pressure[905] (Figure 77). An application showing the potential of this concept is the preparation of monolithic styrene–divinylbenzene in pulled electrospray needles, enabling separation and direct on-line determination of peptides in an MS system with low background noise.[906] Rigid reactive poly(glycidyl methacrylate–ethylene dimethacrylate) monoliths can be easily

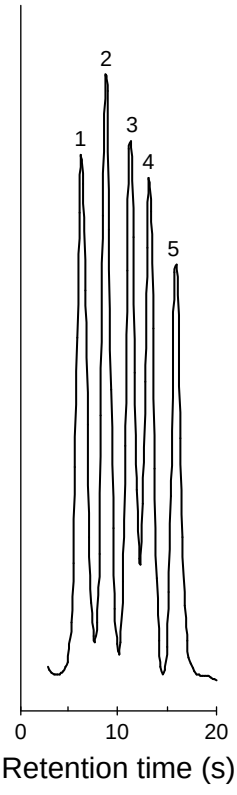

Figure 77 Rapid RP separation of (1) ribonuclease, (2) cytochrome *c*, (3) BSA, (4) carbonic anhydrase, and (5) ovalbumin on a 50 × 4.6 mm i.d. monolithic poly(styrene–divinylbenzene) column at a flow rate of 10 mL min^{-1}. (Reproduced from S. Xie et al., *J. Chromatogr.*, **865**, 174 (1999) by permission of Elsevier Science B.V.)

provided with special functionalities such as ion-exchange or hydrophobic groups by chemical modification. The experimental work done so far confirms the potential of these molded continuous supports. In addition to RP, HIC, IEC, and precipitation chromatography, these materials have also been used for CEC.[907]

Tanaka et al. reported the preparation of C_{18}-modified silica-based monolithic columns in the mid-1990s.[909] A single piece of porous material was obtained by hydrolytic polymerization of tetramethoxysilane in the presence of poly(ethylene oxide) and catalytic amounts of acetic acid in a plastic mold. The material was removed from the mold, treated subsequently with aqueous ammonia and nitric acid, washed with methanol, and dried. The final structural features of the porous silica were obtained after 2 h of heating to 600 °C. The material was finally machined to the column shape and placed in a cartridge. Although this process is more complex than the direct in situ polymerization used for the preparation of monolithic columns from porous organic polymers, the columns afforded efficiencies of up to 96 000 plates per meter for alkylbenzenes[909,910] and a remarkable 20 000 plates per meter for the separation of insulin at a high flow velocity of 5 mm s^{-1}.[911] This excellent performance appears to arise primarily from the specific morphology of these materials, which have an unusually high level of pore connectivity and possess a larger volume of mesopores (Figure 78).

Among the numerous advantages of monolithic columns are their ruggedness, the versatility of their chemistries, and their excellent efficiency in the separation of biological and synthetic molecules.

9.4 Temperature-responsive Chromatography

A significant recent trend in biological macromolecule separations is the use of interaction surfaces capable of undergoing rapidly and controlled morphological changes in response to changes in their environment, which in turn alters the interaction characteristics. The best studied of these phenomena are the intra- and intermolecular interactions of side-chains of linear synthetic polymers, which in some polymers carrying side-chains of suitable polarity and shape causes an abrupt temperature-dependent morphological transition known as a critical solution temperature transition. In response to increased temperature, these transitions can be both expansion–dissolution and contraction–precipitation. The distinct temperature-dependent phase phenomena observed for biopolymers, such as thermoreversibility of agarose and gelatin gels, and the melting of double-stranded DNA, are examples of critical solution temperature behavior. These kinds of reversible changes can be utilized in temperature-responsive chromatography.

The best known example of a temperature-responsive synthetic polymer is poly(*N*-isopropylacrylamide) (poly(NIPAAm)). Its thermal responsiveness has

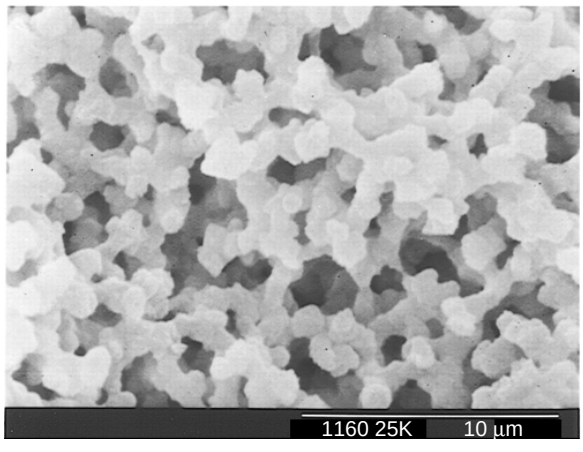

Figure 78 Scanning electron micrograph of a porous monolithic silica rod. (Reproduced from H. Minakuchi et al., *Anal. Chem.*, **68**, 3499 (1996) by permission of the American Chemical Society.)

found use in immunoassay,[912] for modulating[913] and mimicking[914] enzyme activity, in electrochemical biosensor technology,[915] and in LC.[916–918] The characteristic that makes poly(NIPAAm) particularly interesting is the convenient location of the lower critical solution temperature at 32 °C, where drastic changes of the retention properties are observed.[128] The isopropyl group of NIPAAm has a hydrophobic character well suited for the HIC separation of proteins. This application was verified on a continuous-bed sorbent[919] and grafted layers on controlled pore glass[920] (Figure 79). The temperature-controlled retention allows the HIC separations to proceed with isocratic eluents, thus excluding the tedious equilibration with the starting mobile phase typical of gradient elutions.

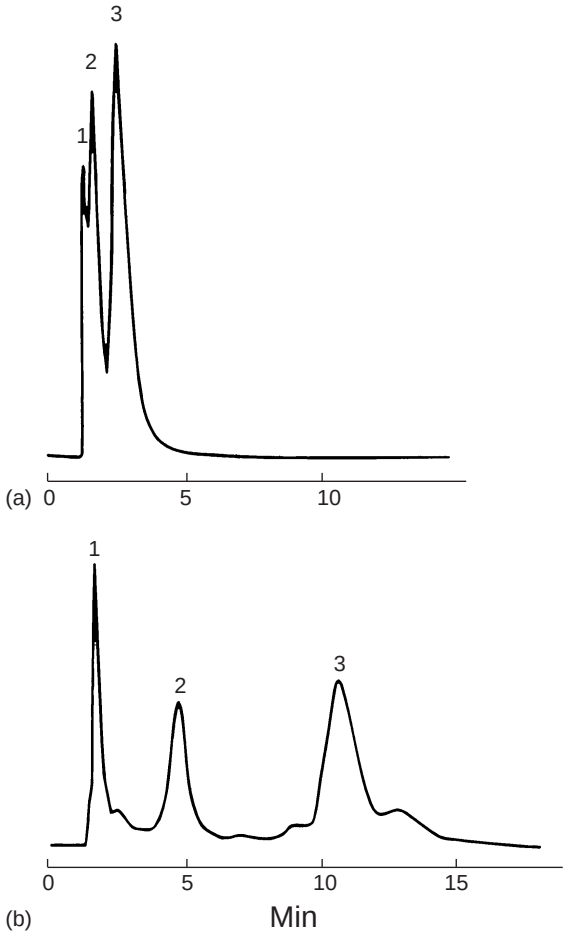

Figure 79 Chromatograms of a polypeptide mixture on aminopropyl silica modified with a semitelechelic copolymer of *N*-isopropylacrylamide (96.8 mol %) and butyl methacrylate (3.2 mol %). Mobile phase: 0.5 M aqueous NaCl (pH 2.1), at (a) 5 °C and (b) 30 °C. Peak identities: (1) insulin chain A, (2) β-endorphin fragment 1–27, (3) insulin chain B. (Reproduced from H. Kanazawa et al., *Anal. Chem.*, **69**, 829 (1997) by permission of the American Chemical Society.)

9.5 Molecularly Imprinted Stationary Phases

Molecularly imprinted polymers (MIPs), conceptualized by Pauling as early as 1940,[921] have been the subject of systematic studies since the 1970s.[922] The driving force for pursuing MIPs as HPLC matrices is to prepare stationary phases with the ultimate selectivity – recognition of a single molecular species. This is accomplished by polymerization of a mixture of monomers around a template molecule, so that the three-dimensional shape of the template is imprinted in the polymer structure. The technique relies upon polar groups in the template being able to organize complementary functional groups on the monomers prior and during solidification of the system. This spatial organization is subsequently frozen by a dense cross-linking. The polymer block is thereafter ground and the template molecules extracted from the pores. The empty cavities have an ability to selectively sequester the template molecules and the spatially organized imprint can recognize chirality. The utility of MIP sorbents has mainly been demonstrated in the separation of racemic mixtures of small chiral compounds.[923] Although Pauling foresaw the preparation of ''synthetic antibodies'', it has not yet been possible to prepare MIPs with recognition capabilities for biomacromolecules approaching those found in biological systems. As recognition depends on sequential formation and breaking of several interaction points located in exactly shaped cavities, sorption and desorption reactions in sterically restricted locations become rate-limiting steps and the kinetics of current MIPs is therefore slow. Site inhomogeneity producing severe tailing is also often seen in demonstrated separations. Judging from the large numbers of reviews that have been published recently,[923–929] MIPs have not lost their appeal. However, the persistent challenges in their development are to obtain recognition for large molecules, and to accelerate the interaction kinetics. It appears that radically different designs of MIPs are needed to obtain recognition of biomacromolecules.[930]

9.6 Restricted Access Media

Restricted access media (RAM), also known as internal-surface reversed-phase (ISRP) sorbents, are porous beads where the functionalization of the perimeter is different from the internal pore system. The outward-facing surface of the particle is typically hydrophilic, whereas the inner surface is hydrophobic or carrying groups for specific interactions with solutes capable of entering the porous system.[931] They can be considered as mixed-mode sorbents with spatially segregated hydrophobicity controlled by a size-exclusion mechanism. Such materials are obviously not very useful for separation of biological macromolecules but their utility in separation

schemes for large molecules lies in the possibility of selectively interacting with small molecules that should be removed from macromolecules.[932] Internal-surface materials have also proven useful in the separation of cyclic peptides.[235]

10 CONCLUSIONS

It is virtually impossible to provide a complete and nuanced overview of an area as vast as the one covered by this article. Yet, it has hopefully been possible to communicate not only the basic problems associated with separation of biological macromolecules in LC, but also the possibilities created by recent developments in the area. An important distinction must be made between analytical separations and micro- or macropreparative separations by HPLC, a gap that has been narrowed through modern materials that merge soft separation techniques previously reserved for slow preparative separations into the HPLC format. The major tradeoffs are, as always, between separation efficiency, separation time, and capacity. A recurring issue has been ways of preventing denaturing, i.e. to preserve biological activity of the separated substances. Again, this issue is usually more important to biochemists than to analytical chemists, as long as the separation techniques are capable of resolving the solutes and presenting them to the detector in a form which allows identification and quantification. Although separation chemistries have been treated as if they were entirely separate from each other, biomacromolecules are themselves governing the interaction modes to a large extent. Their chromatographic behaviors are therefore usually a mix between different separation modes, where the combined properties of the solute, the support, the interaction groups, and the eluent are interacting in providing a retention according to a multitude of interaction mechanisms. By having provided abbreviated descriptions of each of these modes, the reader should be better prepared to make all the choices required in developing a method. However, in the end all biological macromolecules are unique, and any generalization is prone to fail when faced with the immense variability of biological systems.

ACKNOWLEDGMENTS

The author is greatly indebted to Frantisek Švec for carefully reading and commenting on most parts of the manuscript, as well as for providing a textual input on continuous separation media.

ABBREVIATIONS AND ACRONYMS

BSA	Bovine Serum Albumin
CCC	Countercurrent Chromatography
CD	Circular Dichroism
CDI	1,1′-Carbonyldiimidazole
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CMC	Critical Micelle Concentration
CSP	Chiral Stationary Phase
CZE	Capillary Zone Electrophoresis
DAD	Diode-array Detector
DEAE	Diethyl Aminoethyl
DPA	Dipicolylamine
EDTA	Ethylenediamine Tetraacetic Acid
EGTA	Ethyleneglycol Bis(aminoethyl)-tetraacetic Acid
ELSD	Evaporative Light Scattering Detection
FFF	Field Flow Fractionation
FITC	Fluorescein Isothiocyanate
FPLC	Fast Protein Liquid Chromatography
GAG	Glycosaminoglucans
GC	Gas Chromatography
GLYMO	3-Glycidoxypropyltrimethoxysilane
HDL	High-density Lipoprotein
HEMA	2-Hydroxyethyl Methacrylate
HEMA/EDMA	2-Hydroxyethyl Methacrylate/ Ethylene Dimethacrylate
HFBA	Heptafluorobutyric Acid
HIC	Hydrophobic Interaction Chromatography
HILIC	Hydrophilic Interaction Chromatography
HPAEC	High-pH Anion-exchange Chromatography
HPCF	High-performance Chromato-focusing
HPLAC	High-performance Liquid-affinity Chromatography
HPLC	High-performance Liquid Chromatography
HPSEC	High-performance Size-exclusion Chromatography
HSA	Human Serum Albumin
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IDA	Iminodiacetate
IEC	Ion-exchange Chromatography
Ig	Immunoglobulin
IgG	Immunoglobulin G
IMAC	Immobilized Metal Ion Affinity Chromatography

ISRP	Internal-surface Reversed-phase
LALLS	Low-angle Laser Light Scattering
LC	Liquid Chromatography
LDH	Lactate Dehydrogenase
LDL	Low-density Lipoprotein
LP	Lipoprotein
LSD	Lysergic Acid Diethylamide
MALDI	Matrix-assisted Laser Desorption/ Ionization
MALDI/TOFMS	Matrix-assisted Laser Desorption/ Ionization Time-of-flight Mass Spectrometry
MALLS	Multiple Angle Laser Light Scattering
MEKC	Micellar Electrokinetic Chromatography
MIP	Molecularly Imprinted Polymer
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NAD	Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NANA	*N*-Acetylneuramic Acid
NMR	Nuclear Magnetic Resonance
NP	Normal Phase
NPHPLC	Normal-phase High-performance Liquid Chromatography
NTA	Nitrilotriacetate
OPS	*O*-Phosphoserine
PAD	Pulsed Amperometric Detection
PBD	Polybutadiene
PCR	Polymerase Chain Reaction
PEEK	Poly(ether-ether-ketone)
PEI	Poly(ethylene imine)
PGC	Porous Graphitized Carbon
p*I*	Isoelectric Point
poly(NIPAAm)	Poly(*N*-isopropylacrylamide)
RAM	Restricted Access Media
RI	Refractive Index
RIA	Radioimmunoassay
RP	Reversed Phase
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RPIEC	Reversed-phase Ion-exchange Chromatography
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size-exclusion Chromatography
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SEM	Scanning Electron Microscopy
SPE	Solid Phase Extraction
TAC	Thiophilic Adsorption Chromatography
tacn	1,4,7-Triazacyclononane
TED	Tris(carboxymethyl)ethylenediamine
TFA	Trifluoroacetic Acid
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VLDL	Very-low-density Lipoprotein

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures

Liquid Chromatography **(Volume 13)**
Biopolymer Chromatography

REFERENCES

1. M.S. Tswett, *Chromatographic Adsorption Analysis. Selected Works*, Horwood, Hemel Hempstead, 1990.
2. The Nobel Foundation, *Nobel Lectures, including Presentation Speeches and Laureates' Biographies. Chemistry 1942–1962*, Elsevier, Amsterdam, 1964.
3. P. Jössang, 'The Inventor of Chromatography (1861): F. Goppelsröder, Professor at the University of Basel', *Sci. Tribune*, October (1996).
4. K. Loester, D. Josic, 'Analysis of Protein Aggregates by Combination of Crosslinking Reactions and Chromatographic Separations', *J. Chromatogr. B, Biomed. Sci. Appl.*, **699**, 439–461 (1997).
5. P. Serwer, G.A. Griess, 'Advances in the Separation of Bacteriophages and Related Particles', *J. Chromatogr. B*, **722**, 179–190 (1999).
6. A. Lundqvist, P. Lundahl, 'Chromatography on Cells and Biomolecular Assemblies', *J. Chromatogr. B*, **699**, 209–220 (1997).
7. R. Rapley, B.D.M. Theophilus, I.S. Bevan, M.R. Walker, 'Fundamentals of the Polymerase Chain Reaction: A Future in Clinical Diagnostics?', *Med. Lab. Sci.*, **49**, 119–128 (1992).

8. S. Hjerten, 'Agarose Gels in h.p.l.c. Separation of Biopolymers', *Trends Anal. Chem.*, **3**, 87–90 (1984).
9. X. Santarelli, D. Müller, J. Jozefonvicz, 'Dextran-coated Silica Packings for High-performance Size-exclusion Chromatography of Proteins', *J. Chromatogr.*, **443**, 55–62 (1988).
10. H.Y. Aboul-Enein, 'Recent Applications of Polysaccharide-type Chiral Stationary Phases for the Enantioselective Analysis of Chiral Drugs with One and Two Stereogenic Centres', *Biomed. Chromatogr.*, **12**, 116–119 (1998).
11. J. Nahlik, I. Derdowska, W. Neugebauer, G. Kupryszewski, 'Application of Chitin and Chitosan from Antarctic Krill (*Euphausia superba*) as a Support in Thin-layer Chromatography', *Chem. Anal. (Warsaw)*, **30**, 39–47 (1985).
12. V. Carunchio, A.M. Girelli, A. Messina, M. Sinibaldi, 'Chitosan-coated Silica Gel as a New Support in High-performance Liquid Chromatography', *Chromatographia*, **23**, 731–735 (1987).
13. Y.C. Shi, Y.M. Jiang, D.X. Sui, Y.L. Li, T. Chen, L. Ma, Z.T. Ding, 'Affinity Chromatography of Trypsin Using Chitosan as Ligand Support', *J. Chromatogr. A*, **742**, 107–112 (1996).
14. G.A. Maresh, B.S. Dunbar, 'Simple, Inexpensive Method for the Salt-free Concentration of Proteins for Analysis by Two-dimensional Gel Electrophoresis', *Electrophoresis (Weinheim)*, **9**, 54–57 (1988).
15. F. Honda, H. Honda, M. Koishi, T. Matsuno, 'Properties of Cattle bone Powder-coated Composite Particles as High-performance and Open Column Liquid-chromatographic Column Packings', *J. Chromatogr. A*, **813**, 21–33 (1998).
16. A.L. Crego, A. Gonzalez, M.L. Marina, 'Electrochromatography', *Crit. Rev. Anal. Chem.*, **26**, 261–304 (1996).
17. L.A. Colon, Y. Guo, A. Fermier, 'Capillary Electrochromatography', *Anal. Chem.*, **69**, 461A–467A (1997).
18. L.A. Colon, K.J. Reynolds, R. Alicea-Maldonado, A.M. Fermier, 'Advances in Capillary Electrochromatography', *Electrophoresis*, **18**, 2162–2174 (1997).
19. M.G. Cikalo, K.D. Bartle, M.M. Robson, P. Myers, M.R. Euerby, 'Capillary Electrochromatography', *Analyst*, **123**, 87R–102R (1998).
20. J. Vindevogel, P. Sandra, *Introduction to Micellar Electrokinetic Chromatography*, Huethig, Heidelberg, 1992.
21. G.M. Janini, H.J. Issaq, G.M. Muschik, 'Electrokinetic Chromatography without Electro-osmotic Flow', *J. Chromatogr. A*, **792**, 125–141 (1997).
22. M.A. Strege, A.L. Lagu, 'Micellar Electrokinetic Chromatography of Proteins', *J. Chromatogr. A*, **780**, 285–296 (1997).
23. K. Otsuka, S. Terabe, 'Micellar Electrokinetic Chromatography', *Bull. Chem. Soc. Jpn.*, **71**, 2465–2481 (1998).
24. Y. Ito, W.D. Conway, *High-speed Countercurrent Chromatography*, John Wiley & Sons, Chichester, 1995.
25. W.D. Conway, R.J. Petroski, *Modern Countercurrent Chromatography*, (ACS Symp. Ser., Vol. 593), American Chemical Society, Washington, DC, 1995.
26. Y. Ma, Y. Ito, 'Recent Advances in Peptide Separation by Countercurrent Chromatography', *Anal. Chim. Acta*, **352**, 411–427 (1997).
27. S. Levin, 'Field-flow Fractionation in Biomedical Analysis', *Biomed. Chromatogr.*, **5**, 133–137 (1991).
28. J. Janca, 'Field-flow Fractionation: Analytical and Micro-preparative Methodology', *Mikrochim. Acta*, **111**, 135–162 (1993).
29. J.C. Giddings, *Biotechnology Applications and Methods*, (Adv. Chromatogr. Ser., Vol. 29), Marcel Dekker, New York, 1989.
30. K.M. Gooding, F.E. Regnier (eds), *HPLC of Biological Macromolecules. Methods and Applications*, Dekker, New York, 1990.
31. W.S. Hancock (ed.), *High Performance Liquid Chromatography in Biotechnology*, John Wiley & Sons, Chichester, 1990.
32. Cs. Horváth, J.G. Nikelly (eds), *Analytical Biotechnology: Capillary Electrophoresis and Chromatography*, (ACS Symp. Ser., Vol. 434), American Chemical Society, Washington, DC, 1990.
33. NATO Advanced Study Institute on Chromatography and Membrane Processes, *Chromatographic and Membrance Processes in Biotechnology*, (NATO ASI Ser., Ser. E, Appl. Sci., Vol. 204), Kluwer, Dordrecht, 1991.
34. M.T.W. Hearn (ed.), *HPLC of Proteins, Peptides, and Polynucleotides: Contemporary Topics and Applications*, VCH, New York/Weinheim, 1991.
35. Cs. Horváth, L.S. Ettre (eds), *Chromatography in Biotechnology*, (ACS Symp. Ser., Vol. 529), American Chemical Society, Washington, DC, 1993.
36. T.T. Ngo (ed.), *Molecular Interactions in Bioseparations*, Plenum, New York, 1993.
37. G.T. Tsao (ed.), *Advances in Biochemical Engineering/Biotechnology. Vol. 49. Chromatography*, Springer Verlag, New York, 1993.
38. E.D. Katz, *High-performance Liquid Chromatography. Principles and Methods in Biotechnology*, John Wiley and Sons, Chichester, 1995.
39. B.L. Karger, W.S. Hancock (eds), *Methods of High-resolution Separation and Analysis of Biological Macromolecules: Fundamentals*, (Methods Enzymol., 270), Academic Press, San Diego, 1996.
40. B.L. Karger, W.S. Hancock (eds), *Methods of High-resolution Separation and Analysis of Biological Macromolecules: Applications*, (Methods Enzymol., 271), Academic Press, San Diego, 1996.
41. H.Y. Aboul-Enein (ed.), *Analytical and Preparative Separation Methods of Biomacromolecules*, Marcel Dekker, New York, 1999.
42. C. Schöneich, S.K. Kwok, G.S. Wilson, S.R. Rabel, J.F. Stobaugh, T.D. Williams, D.G. Vander Velde,

'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **65**, R67–R84 (1993).
43. C. Schöneich, A.F.R. Hühmer, S.R. Rabel, J.F. Stobaugh, S.D.S. Jois, C.K. Larive, T.J. Siahaan, T.C. Squier, D.J. Bigelow, T.D. Williams, 'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **67**, R155–R181 (1995).
44. A.F.R. Hühmer, G.I. Aced, M.D. Perkins, R.N. Gürsoy, D.S.S. Jois, C. Larive, T.J. Siahaan, C. Schöneich, 'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **69**, R29–R57 (1997).
45. C.K. Larive, S.M. Lunte, M. Zhong, M.D. Perkins, G.S. Wilson, G. Gokulrangan, T. Williams, F. Afroz, C. Schöneich, T.S. Derrick, C.R. Middaugh, S. Bogdanowich-Knipp, 'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **71**, R389–R423 (1999).
46. S. Lindsay, *High-performance Liquid Chromatography*, John Wiley and Sons, Chichester, 1992.
47. W.J. Lough, I.W. Wainer, *High-performance Liquid Chromatography. Fundamental Principles and Practice*, Chapman and Hall, London, 1995.
48. R.M. Smith, *Retention and Selectivity in Liquid Chromatography*, (J. Chromatogr. Lib., Vol. 57), Elsevier, Amsterdam, 1995.
49. E. Katz, R. Eksteen, P. Schoenmakers, N. Miller, *Handbook of HPLC*, Marcel Dekker, New York, 1998.
50. R.W.A. Oliver, *HPLC of Macromolecules: a Practical Approach*, IRL Press, Oxford, 1989.
51. A. Kenney, S. Fowell (eds), *Practical Protein Chromatography*, (Meth. Mol. Biol, Vol. 11), Humana Press, Totowa, 1992.
52. B.A. Bidlingmeyer, *Practical HPLC Methodology and Applications*, John Wiley and Sons, Chichester, 1993.
53. V.R. Meyer, *Practical High-performance Liquid Chromatography*, John Wiley and Sons, Chichester, 1993.
54. M.C. McMaster, *HPLC: a Practical User's Guide*, VCH, New York, 1994.
55. P. Sadek, *The HPLC Solvent Guide*, John Wiley and Sons, New York, 1996.
56. J. Swadesh, *HPLC: Practical and Industrial Applications*, CRC Press, Boca Raton, 1996.
57. L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, John Wiley and Sons, Chichester, 1997.
58. T. Hanai, *HPLC: a Practical Guide*, Royal Society of Chemistry, Cambridge, UK, 1999.
59. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, John Wiley and Sons, Chichester, 1980.
60. R. Freitag, J. Breier, 'Displacement Chromatography in Biotechnological Downstream Processing', *J. Chromatogr. A*, **691**, 101–112 (1995).
61. G. Jayaraman, Y.F. Li, J.A. Moore, S.M. Cramer, 'Ion-exchange Displacement Chromatography of Proteins – Dendritic Polymers as Novel Displacers', *J. Chromatogr. A*, **702**, 143–155 (1995).
62. R.C. Ruaan, D. Hsu, W.Y. Chen, H. Chen, M.S. Lin, 'Protein Separation by Hydrophobic-interaction Chromatography Using Methacrylic Block Copolymers as Displacers', *J. Chromatogr. A*, **824**, 35–43 (1998).
63. P. Jandera, J. Churacek, *Gradient Elution in Column Liquid Chromatography: Theory and Practice*, Elsevier, Amsterdam, 1985.
64. I.Y. Galaev, C. Warrol, B. Mattiasson, 'Temperature-induced Displacement of Proteins from Dye-affinity Columns Using an Immobilized Polymeric Displacer', *J. Chromatogr. A*, **684**, 37–43 (1994).
65. F. Houdiere, P.W.J. Fowler, N.M. Djordjevic, 'Combination of Column Temperature Gradient and Mobile Phase Flow Gradient in Microcolumn and Capillary Column High-performance Liquid Chromatography', *Anal. Chem.*, **69**, 2589–2593 (1997).
66. V. Lesins, E. Ruckenstein, 'Gradient Flow Programming: a Coupling of Gradient Elution and Flow Programming', *J. Chromatogr.*, **467**, 1–14 (1989).
67. A. Velayudhan, M.R. Ladisch, 'Role of the Modulator in Gradient Elution Chromatography', *Anal. Chem.*, **63**, 2028–2032 (1991).
68. A. Velayudhan, R.L. Hendrickson, M.R. Ladisch, 'Simultaneous Concentration and Purification Through Gradient Deformation Chromatography', *Am. Ind. Chem. Eng. J.*, **41**, 1184–1193 (1995).
69. A. Velayudhan, M.R. Ladisch, 'Effect of Modulator Sorption on Gradient Shape in Ion-exchange Chromatography', *Ind. Eng. Chem. Res.*, **34**, 2805–2810 (1995).
70. A. Brandt, G. Mann, W. Arlt, 'Temperature Gradients in Preparative High-performance Liquid Chromatography Columns', *J. Chromatogr. A*, **769**, 109–117 (1997).
71. M.T.W. Hearn, G.L. Zhao, 'Investigations into the Thermodynamics of Polypeptide Interaction with Nonpolar Ligands', *Anal. Chem.*, **71**, 4874–4885 (1999).
72. J.H. Knox, 'Practical Aspects of LC Theory', *J. Chromatogr. Sci.*, **15**, 352–364 (1977).
73. C.T. Mant, R.S. Hodges (eds), *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, 1991.
74. R.A. Freitas, Jr, *Nanomedicine: Basic Capabilities*, Landes Bioscience, Austin, Table 3.3, Volume 1, 1999.
75. B.L. Karger, 'The Relationship of Theory and Practice in High-speed Liquid Chromatography', in *Modern Practice of Liquid Chromatography*, ed. J.J. Kirkland, John Wiley & Sons: New York, 25, 1971.
76. O. Mikeš, *High-performance Liquid Chromatography of Biopolymers and Biooligomers. Part A: Principles, Materials and Techniques*, (J. Chromatogr. Lib., Vol. 41A), Elsevier, Amsterdam, 1988.
77. T.B. Tennikova, B.G. Belenkii, F. Svec, 'High-performance Membrane Chromatography. A Novel Method of Protein Separation', *J. Liq. Chromatogr.*, **13**, 63–70 (1990).
78. J.L. Coffman, D.K. Roper, E.N. Lightfoot, 'High-resolution Chromatography of Proteins in Short Columns

and Adsorptive Membranes', *Bioseparation*, **4**, 183–200 (1994).
79. X. Santarelli, F. Domergue, G. Clofent-Sanchez, M. Dabadie, R. Grissely, C. Cassagne, 'Characterization and Application of New Macroporous Membrane Ion Exchangers', *J. Chromatogr. B*, **706**, 13–22 (1998).
80. M.T.W. Hearn, 'Current Status and Future Challenges of High-performance Liquid Chromatographic Techniques for Biopolymer Analysis and Purification', in *HPLC of Proteins, Peptides, and Polynucleotides: Contemporary Topics and Applications*, ed. M.T.W. Hearn, VCH, New York/Weinheim, 1–35, 1991.
81. M.T.W. Hearn, 'High-performance Liquid Chromatography of Peptides and Proteins: Quantitative Relationships for Isocratic and Gradient Elution of Peptides and Proteins', in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, eds C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 105–122, 1991.
82. A. Tiselius, 'Electrophoresis and Chromatography of Proteins and Polypeptides', *Angew. Chem.*, **67**, 245–251 (1956).
83. N.I. Dubinina, O.I. Kurenbin, T.B. Tennikova, 'Peculiarities of Gradient Ion-exchange High-performance Liquid Chromatography of Proteins', *J. Chromatogr. A*, **753**, 217–225 (1996).
84. M.B. Tennikov, N.V. Gazdina, T.B. Tennikova, F. Svec, 'Effect of Porous Structure of Macroporous Polymer Supports on Resolution in High-performance Membrane Chromatography of Proteins', *J. Chromatogr. A*, **798**, 55–64 (1998).
85. U.K. Laemmeli, 'Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4', *Nature*, **227**, 680–685 (1970).
86. M. Kaufmann, 'Unstable Proteins: How to subject them to Chromatographic Separations for Purification Procedures', *J. Chromatogr. B*, **699**, 347–369 (1997).
87. S.W. Lin, B.L. Karger, 'Reversed-phase Chromatographic Behavior of Proteins in Different Unfolded States', *J. Chromatogr.*, **499**, 89–102 (1990).
88. R. Janzen, K.K. Unger, W. Müller, M.T.W. Hearn, 'Adsorption of Proteins on Porous and Nonporous Poly(ethyleneimine) and Tentacle-type Anion-exchangers', *J. Chromatogr.*, **522**, 77–93 (1990).
89. K.K. Unger (ed.), *Packings and Stationary Phases in Chromatographic Techniques*, Dekker, New York, 1990.
90. R.P.W. Scott, *Liquid Chromatography Column Theory*, John Wiley & Sons, Chichester, 1992.
91. R.P.W. Scott, *Silica Gel and Bonded Phases: Their Production, Properties and Use in LC*, John Wiley and Sons, Chichester, 1993.
92. U.D. Neue, *HPLC Columns. Theory, Technology, and Practice*, Wiley-VCH, Weinheim, 1997.
93. K.K. Unger, R. Janzen, G. Jilge, 'Packings and Stationary Phases for Biopolymer Separations by HPLC', *Chromatographia*, **24**, 144–154 (1987).
94. J. Coupek, I. Vins, 'Hydroxyethyl Methacrylate-based Sorbents for High-performance Liquid Chromatography of Proteins', *J. Chromatogr. A*, **658**, 391–398 (1994).
95. K. Štulìk, V. Pacáková, J. Suchánková, H.A. Claessens, 'Stationary Phases for Peptide Analysis by High-performance Liquid Chromatography: a review', *Anal. Chim. Acta*, **352**, 1–19 (1997).
96. M. Leonard, 'New Packing Materials for Protein Chromatography', *J. Chromatogr. B*, **699**, 3–27 (1997).
97. W.C. Lee, 'Protein Separation Using Non-porous Sorbents', *J. Chromatogr. B*, **699**, 29–45 (1997).
98. A.E. Rodrigues, 'Permeable Packings and Perfusion Chromatography in Protein Separation', *J. Chromatogr. B*, **699**, 47–61 (1997).
99. M.I. Aguilar, M.T.W. Hearn, 'High-resolution Reversed-phase High-performance Liquid Chromatography of Peptides and Proteins', *Methods Enzymol.*, **270**, 3–27 (1996).
100. L. Várady, K. Kalghatgi, Cs. Horváth, 'Rapid High-performance Affinity Chromatography on Micropellicular Sorbents', *J. Chromatogr.*, **458**, 207–215 (1988).
101. G. Matthijs, E. Schacht, 'Dextran-coated Silica and its Behaviour in High-performance Size-exclusion Chromatography', *J. Chromatogr. A*, **755**, 1–10 (1996).
102. S. Hjerten, K.O. Eriksson, 'High-performance Molecular Sieve Chromatography of Proteins on Agarose Columns: the Relation Between Concentration and Porosity of the Gel', *Anal. Biochem.*, **137**, 313–317 (1984).
103. J. Turkova, *Bioaffinity Chromatography*, 2nd edition (J. Chromatogr. Library, Vol. 55), Elsevier, Amsterdam, 1993.
104. S. Arnott, A. Fulmer, W.E. Scott, I.C.M. Dea, R. Moorhouse, D.A. Rees, 'The Agarose Double Helix and its Function in Agarose Gel Structure', *J. Mol. Biol.*, **90**, 269–284 (1974).
105. S. Hjertén, Z.Q. Liu, D. Yang, 'Some Studies on the Resolving Power of Agarose-based High-performance Liquid-chromatographic Media for the Separation of Macromolecules', *J. Chromatogr.*, **296**, 115–120 (1984).
106. S. Hjertén, J. Mohammad, K.O. Eriksson, J.L. Liao, 'General Methods to render Macroporous Stationary Phases Nonporous and Deformable, exemplified with Agarose and Silica Beads and their use in High-performance Ion-exchange and Hydrophobic-interaction Chromatography of Proteins', *Chromatographia*, **31**, 85–94 (1991).
107. K.M. Gooding, H. Hagestam Freiser, 'High-performance Size-exclusion Chromatography of Proteins', in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, eds C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 135–144, 1991.
108. C. Laurent, H.A.H. Billiet, L. De Galan, 'On the Use of Alumina in High-performance Liquid Chromatography with Aqueous Mobile Phases at Extreme pH', *Chromatographia*, **17**, 253–258 (1983).

109. L. Sun, A.V. McCormick, P.W. Carr, 'Study of the Irreversible Adsorption of Proteins on Polybutadiene-coated Zirconia', *J. Chromatogr. A*, **658**, 465–473 (1994).
110. G. Schomburg, 'Stationary Phases in High-performance Liquid Chromatography. Chemical Modification by Polymer Coating', *LC-GC*, **6**, 36–50 (1988).
111. J.J. Pesek, H.D. Lin, 'Evaluation of Synthetic Procedures for the Chemical Modification of Alumina for HPLC', *Chromatographia*, **28**, 565–568 (1989).
112. M. Kawahara, H. Nakamura, T. Nakajima, 'Titania and Zirconia: Possible new Ceramic Microparticulates for High-performance Liquid Chromatography', *J. Chromatogr.*, **515**, 149–158 (1990).
113. U. Truedinger, G. Mueller, K.K. Unger, 'Porous Zirconia and Titania as Packing Materials for High-performance Liquid Chromatography', *J. Chromatogr.*, **535**, 111–125 (1990).
114. H. Matsuda, H. Nakamura, T. Nakajima, 'New Ceramic Titania: Selective Adsorbent for Organic Phosphates', *Anal. Sci.*, **6**, 911–912 (1990).
115. M. Gruen, A.A. Kurganov, S. Schacht, F. Schueth, K.K. Unger, 'Comparison of an Ordered Mesoporous Aluminosilicate, Silica, Alumina, Titania and Zirconia in Normal-phase High-performance Liquid Chromatography', *J. Chromatogr. A*, **740**, 1–9 (1996).
116. K. Murayama, H. Nakamura, T. Nakajima, K. Takahashi, A. Yoshida, 'Preparation and Evaluation of Octadecyl Titania as Column-packing Material for High-performance Liquid Chromatography', *Microchem. J.*, **49**, 362–367 (1994).
117. K. Murayama, H. Nakamura, T. Nakajima, K. Takahashi, A. Yoshida, 'Alkali-resistance of Octadecyl Titania, a new Packing Material for Reversed-phase High-performance Liquid Chromatography', *Anal. Sci.*, **10**, 497–499 (1994).
118. K. Tani, Y. Suzuki, 'Effect of Octadecyl-modification of Retention when Using Titania as a Support', *J. Liq. Chromatogr. Relat. Technol.*, **19**, 3037–3048 (1996).
119. J. Nawrocki, M.P. Rigney, A. McCormick, P.W. Carr, 'Chemistry of Zirconia and its use in Chromatography', *J. Chromatogr. A*, **657**, 229–282 (1993).
120. M.P. Rigney, E.F. Funkenbusch, P.W. Carr, 'Physical and Chemical Characterization of Microporous Zirconia', *J. Chromatogr.*, **499**, 291–304 (1990).
121. A.M. Clausen, P.W. Carr, 'Chromatographic Characterization of Phosphonate Analogue EDPTA-modified Zirconia Support for Biochromatographic Applications', *Anal. Chem.*, **70**, 378–385 (1998).
122. T. Kawasaki, 'Hydroxyapatite as a Liquid-chromatographic Packing', *J. Chromatogr.*, **544**, 147–184 (1991).
123. H.J. Wirth, K.O. Eriksson, P. Holt, M. Aguilar, M.T.W. Hearn, 'Ceramic-based Particles as Chemically Stable Chromatographic Supports', *J. Chromatogr.*, **646**, 129–141 (1993).
124. K. Tani, Y. Suzuki, 'Investigation of the Ion-exchange Behaviour of Titania: Application as a Packing Material for Ion Chromatography', *Chromatographia*, **46**, 623–627 (1997).
125. C.T. Mant, R.S. Hodges, 'Analysis of Peptides by High-performance Liquid Chromatography', *Methods Enzymol.*, **271**, 3–51 (1996).
126. J.L. Liao, C.M. Zeng, A. Palm, S. Hjerten, 'Continuous Beds for Microchromatography: Detection of Proteins by a Blotting Membrane Technique', *Anal. Biochem.*, **241**, 195–198 (1996).
127. S. Hjertén, J.L. Liao, R. Zhang, 'High-performance Liquid Chromatography on Continuous Polymer Beds', *J. Chromatogr.*, **473**, 273–275 (1989).
128. K. Hosoya, K. Kimata, T. Araki, N. Tanaka, J.M.J. Fréchet, 'Temperature-controlled High-performance Liquid Chromatography Using a Uniformly Sized Temperature-responsive Polymer-based Packing Material', *Anal. Chem.*, **67**, 1907–1911 (1995).
129. V. Smigol, F. Svec, J.M.J. Fréchet, 'Novel Uniformly Sized Polymeric Stationary Phase with Hydrophilized Large Pores for Direct Injection HPLC determination of Drugs in Biological Fluids', *J. Liq. Chromatogr.*, **17**, 891–911 (1994).
130. A. Sugii, K. Harada, 'Evaluation of Vinylpyridine and Vinylpyridinium Polymers as Column Packings for High-performance Liquid Chromatography', *J. Chromatogr.*, **544**, 219–232 (1991).
131. B.H. Yang, L.M. Fan, Y. Wei, R.N. Shen, S.X. Jiang, L.R. Chen, 'Divinylbenzene–Vinylpyrrolidone Copolymer–Silica Packing for Chromatographic Separation of Proteins', *Fenxi Huaxue*, **26**, 1267–1270 (1998).
132. L. Battinelli, E. Cernia, M. Delbo, G. Ortaggi, A. Pala, S. Soro, 'New class of Poly(vinyl alcohol) Polymers as Column-chromatography Stationary Phases for *Candida rugosa* lipase Isoforms Separation', *J. Chromatogr. A*, **753**, 47–55 (1996).
133. G. Schomburg, J. Koehler, H. Figge, A. Deege, U. Bien-Vogelsang, 'Immobilization of Stationary Liquids on Silica Particles by Gamma-radiation', *Chromatographia*, **18**, 265–274 (1984).
134. J.E. Haky, A. Raghani, B.M. Dunn, 'Comparison of Polybutadiene-coated Alumina and Octadecyl-bonded Silica for Separations of Proteins and Peptides by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **541**, 303–315 (1991).
135. M. Hanson, K.K. Unger, G. Schomburg, 'Non-porous Polybutadiene-coated Silicas as Stationary Phases in Reversed-phase Chromatography', *J. Chromatogr.*, **517**, 269–284 (1990).
136. J. Li, Y. Hu, P.W. Carr, 'Fast Separations at Elevated Temperatures on Polybutadiene-coated Zirconia Reversed-phase Material', *Anal. Chem.*, **69**, 3884–3888 (1997).
137. D.H. Reeder, J. Li, P.W. Carr, M.C. Flickinger, A.V. McCormick, 'Models for Polybutadiene Pore Wall Coatings in Porous Zirconia', *J. Chromatogr. A*, **760**, 71–79 (1997).

138. J. Li, P.W. Carr, 'A study of the Efficiency of Polybutadiene-coated Zirconia as a Reversed-phase Chromatographic Support', *Anal. Chem.*, **69**, 2193–2201 (1997).
139. J.W. Li, D.H. Reeder, A.V. McCormick, P.W. Carr, 'Factors Influencing Polybutadiene Deposition Within Porous Chromatographic Zirconia', *J. Chromatogr. A*, **791**, 45–52 (1997).
140. L.F. Sun, P.W. Carr, 'Mixed-mode Retention of Peptides on Phosphate-modified Polybutadiene-coated Zirconia', *Anal. Chem.*, **67**, 2517–2523 (1995).
141. L. Sun, P.W. Carr, 'Chromatography of Proteins Using Polybutadiene-coated Zirconia', *Anal. Chem.*, **67**, 3717–3721 (1995).
142. M. Hanson, K.K. Unger, C.T. Mant, R.S. Hodges, 'Polymer-coated Reversed-phase Packings with Controlled Hydrophobic Properties. I. Effect on the Selectivity of Protein Separations', *J. Chromatogr.*, **599**, 65–75 (1992).
143. M. Hanson, K.K. Unger, C.T. Mant, R.S. Hodges, 'Polymer-coated Reversed-phase Packings with Controlled Hydrophobic Properties. II. Effect on the Selectivity of Peptide Separations', *J. Chromatogr.*, **599**, 77–85 (1992).
144. V.A. Davankov, A.A. Kurganov, K.K. Unger, 'Reversed-phase High-performance Liquid Chromatography of Proteins and Polypeptides on Polystyrene-coated Silica Supports', *J. Chromatogr.*, **500**, 519–530 (1990).
145. A.A. Kurganov, V.A. Davankov, K.K. Unger, 'Ion-exchange High-performance Liquid Chromatography of Nucleotides and Polypeptides on new types of Ion-exchange Sorbents, based on Polystyrene-coated Silicas', *J. Chromatogr.*, **548**, 207–214 (1991).
146. F.L. Zhou, D. Müller, X. Santarelli, J. Jozefonvicz, 'Coated Silica Supports for High-performance Affinity Chromatography of Proteins', *J. Chromatogr.*, **476**, 195–203 (1989).
147. A. Serres, D. Müller, J. Jozefonvicz, 'Purification of Monoclonal Antibodies on Dextran-coated Silica Support Grafted by Thiophilic Ligand', *J. Chromatogr. A*, **711**, 151–157 (1995).
148. C.J. Dunlap, P.W. Carr, 'Synthesis and Chromatographic Characterization of Dextran-coated Zirconia High-performance Liquid-chromatographic Stationary Phases', *J. Chromatogr. A*, **746**, 199–210 (1996).
149. W. Kopaciewicz, M.A. Rounds, F.E. Regnier, 'Stationary Phase Contributions to Retention in High-performance Anion-exchange Protein Chromatography: Ligand Density and Mixed-mode Effects', *J. Chromatogr.*, **318**, 157–172 (1985).
150. W. Kopaciewicz, F.E. Regnier, 'Synthesis of Cation-exchange Stationary Phases Using an Adsorbed Polymeric Coating', *J. Chromatogr.*, **358**, 107–117 (1986).
151. W. Kopaciewicz, F.E. Regnier, 'Synthesis of an Adsorbed Reversed-phase Packing Material for the Separation of Proteins and Peptides', *J. Chromatogr.*, **358**, 119–128 (1986).
152. C. McNeff, Q.H. Zhao, P.W. Carr, 'High-performance Anion Exchange of Small Anions with Polyethyleneimine-coated Porous Zirconia', *J. Chromatogr. A*, **684**, 201–211 (1994).
153. C. McNeff, P.W. Carr, 'High-performance Anion-exchange Chromatography of Oligonucleotides and Oligodeoxynucleotides on Quaternized Polyethyleniminecoated Zirconia', *Anal. Chem.*, **67**, 2350–2353 (1995).
154. Y. Hu, P.W. Carr, 'Synthesis and Characterization of new Zirconia-based Polymeric Cation-exchange Stationary Phases for High-performance Liquid Chromatography of Proteins', *Anal. Chem.*, **70**, 1934–1942 (1998).
155. J. Köhler, 'Poly(vinylpyrrolidone)-coated Silica: a Versatile, Polar Stationary Phase for HPLC', *Chromatographia*, **21**, 573–582 (1986).
156. R. Lemque, C. Vidal Madjar, M. Racine, J. Piquion, B. Sebille, 'Anion-exchange Chromatographic Properties of α-Lactalbumin eluted from Quaternized Polyvinylimidazole. Study of the Role of the Polymer Coating', *J. Chromatogr.*, **553**, 165–177 (1991).
157. A.J. Alpert, 'Cation-exchange High-performance Liquid Chromatography of proteins on Poly(aspartic acid)–Silica', *J. Chromatogr.*, **266**, 23–37 (1983).
158. C.N. Ou, G.J. Buffone, G.L. Reimer, A.J. Alpert, 'High-performance Liquid Chromatography of Human Haemoglobins on a new Cation Exchanger', *J. Chromatogr.*, **266**, 197–205 (1983).
159. A.J. Alpert, 'Hydrophilic-interaction Chromatography for the Separation of Peptides, Nucleic Acid and other Polar Compounds', *J. Chromatogr.*, **499**, 177–196 (1990).
160. A.J. Alpert, 'High-performance Hydrophobic-interaction Chromatography of Proteins on a series of Poly(alkyl aspartamide)–Silicas', *J. Chromatogr.*, **359**, 85–97 (1986).
161. A.J. Alpert, P.C. Andrews, 'Cation-exchange Chromatography of Peptides on Poly-(2-sulphoethyl aspartamide)–Silica', *J. Chromatogr.*, **443**, 85–96 (1988).
162. D. Bentrop, J. Kohr, H. Engelhardt, 'Poly(methyl glutamate)-coated Surfaces in HPLC and CE', *Chromatographia*, **32**, 171–178 (1991).
163. W. Müller, 'New Ion-exchangers for the Chromatography of Biopolymers', *J. Chromatogr.*, **510**, 133–140 (1990).
164. J.W. Burns, S.E. Bialkowski, D.B. Marshall, 'Molecular Interactions at Octadecylated Chromatographic Surfaces', *Anal. Chem.*, **69**, 3861–3870 (1997).
165. G. Weber, 'Measurement of the Speciation of Iron in the Nanogram range: Investigation of Chromatographic Peaks induced by Iron Blanks', *Anal. Chim. Acta*, **283**, 354–360 (1993).
166. P.C. Sadek, P.W. Carr, L.W. Bowers, 'Significance of Metallophilic and Silanophilic Interactions in Reversed-phase HPLC', *J. Liq. Chromatogr.*, **8**, 2369–2386 (1985).
167. P.C. Sadek, P.W. Carr, L.D. Bowers, L.C. Haddad, 'Radiochemical Study of Irreversible Protein Loss on High-performance Liquid Chromatography Column Frits', *Anal. Biochem.*, **144**, 128–131 (1985).

168. P.C. Sadek, P.W. Carr, L.D. Bowers, L.C. Haddad, 'A Radiochemical Study of Irreversible Adsorption of Proteins on Reversed-phase Chromatographic Packing Materials', *Anal. Biochem.*, **153**, 359–371 (1986).
169. C.N. Trumbore, R.D. Tremblay, J.T. Penrose, M. Mercer, F.M. Kelleher, 'Unusual Flow Behaviour and Losses of Proteins in High-performance Liquid Chromatography Capillary Tubing', *J. Chromatogr.*, **280**, 43–57 (1983).
170. M. Herold, G.P. Rozing, J.L. Curtis, 'Recovery of Biologically Active Enzymes after HPLC Separation', *BioTechniques*, **10**, 656–662 (1991).
171. F. Toribio, E. Martinez-Lara, P. Pascual, J. Lopez-Barea, 'Methods for Purification of Glutathione Peroxidase and Related Enzymes', *J. Chromatogr. B*, **684**, 77–97 (1996).
172. W.D. Beinert, W. Mueller, R. Spatz, K. Schneider, 'Advantages of Metal-free HPLC Systems in Biochromatography', *GIT Fachz. Lab.*, **34**, 580–583 (1990).
173. P.J. Schneider, H. Engelhardt, 'Novel Applications for Organic Polymer Capillaries', *J. Chromatogr. A*, **802**, 17–22 (1998).
174. B. Law, D. Temesi, J. Kenworthy, K. Pitts, 'Unexpected Interactions with Silica Capillary Transfer Lines in HPLC-MS', *LC-GC Int.*, **11**, 254 (1998).
175. P. Hambleton, W.J. Lough, J. Maltas, M.J. Mills, 'Unusual Analyte Adsorption Effects on Inert LC Components', *J. Liq. Chromatogr.*, **18**, 3205–3217 (1995).
176. A. Maisch, 'Efficient High-speed HPLC with Gold Columns', *LaborPraxis*, **21**(3), 56–58 (1997).
177. M. le Maire, A. Ghazi, J.V. Møller, 'Gel Chromatography as an Analytical Tool for Characterization of Size and Molecular Mass of Proteins', *ACS Symp. Ser.*, **635**, 36–51 (1996).
178. M. le Maire, L.P. Aggerbeck, L.P. Monteilhet, J.P. Andersen, J.V. Møller, 'The use of High-performance Liquid Chromatography for the Determination of Size and Molecular Weight of Proteins: a Caution and a List of Membrane Proteins Suitable as Standards', *Anal. Biochem.*, **154**, 525–535 (1986).
179. J. Mlynar, S. Sarkanen, 'Renaissance in Ultracentrifugal Sedimentation Equilibrium Calibrations of Size exclusion Chromatographic elution Profiles', *ACS Symp. Ser.*, **635**, 379–400 (1996).
180. M. Potschka, 'A Soft-body Theory of Size Exclusion Chromatography', *ACS Symp. Ser.*, **635**, 67–87 (1996).
181. H. Benoit, Z. Grubizic, P. Rempp, D. Dekker, J.G. Zillox, 'Liquid-phase Chromatographic Study of Branched and Linear Polystyrenes of Known Structure', *J. Chim. Phys.*, **63**, 1507–1514 (1966).
182. J.C. Giddings, E. Kucera, C.P. Russell, M.N. Myers, 'Statistical Theory for the Equilibrium Distribution of Rigid Molecules in Inert Porous Networks. Exclusion Chromatography', *J. Phys. Chem.*, **72**, 4397–4408 (1968).
183. E.F. Casassa, 'Comments on Exclusion of Polymer Chains from Small Pores and its Relation to Gel Permeation Chromatography', *Macromolecules*, **9**, 182–185 (1976).
184. M. Potschka, 'Mechanism of Size-exclusion Chromatography. I. Role of Convection and Obstructed Diffusion in Size-exclusion Chromatography', *J. Chromatogr.*, **648**, 41–69 (1993).
185. R. Nave, K. Weber, M. Potschka, 'Universal Calibration of Size-exclusion Chromatography for Proteins in Guanidinium Hydrochloride including the High-molecular-mass Proteins Titin and Nebulin', *J. Chromatogr. A*, **654**, 229–246 (1993).
186. J.P. Busnel, C. Degoulet, J.F. Tassin, 'Modeling of Size Exclusion Chromatography by Monte Carlo Simulation', *ACS Symp. Ser.*, **635**, 52–66 (1996).
187. S. Hjertén, B. Wu, J. Liao, 'High-performance Liquid-chromatographic Matrix Based on Agarose Cross-linked with Divinyl Sulphone', *J. Chromatogr.*, **396**, 101–113 (1987).
188. D. Horik, J. Hradil, M.J. Benes, 'Packings for Size Exclusion Chromatography: Preparation and Some Properties', *ACS Symp. Ser.*, **635**, 190–210 (1996).
189. V.A. Romano, T. Ebeyer, P.L. Dubin, 'Influence of Net Protein Charge and Stationary Phase Charge on Protein Retention in Size Exclusion Chromatography', *ACS Symp. Ser.*, **635**, 88–102 (1996).
190. C.H. Cai, V.A. Romano, P.L. Dubin, 'Ionic Strength Dependence of Protein Retention on Superose 12 in SEC-IEC Mixed Mode Chromatography', *J. Chromatogr. A*, **693**, 251–261 (1995).
191. H. Engelhardt, 'Size-exclusion Chromatography of Proteins', in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, eds C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 145–153, 1991.
192. G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, 'Two-dimensional SEC/RPLC Coupled to Mass Spectrometry for the Analysis of Peptides', *Anal. Chem.*, **69**, 2283–2291 (1997).
193. D.S. Jackson, R.D. Waniska, L.W. Rooney, 'Molecular-weight Patterns of Nageli Amylodextrins', *Starch-Starke*, **44**, 59–61 (1992).
194. M.L. Fishman, D.T. Gillespie, B. Levaj, 'Structural-analysis of Aggregated Polysaccharides by High-performance Size-exclusion Chromatography Viscometry', *ACS Symp. Ser.*, **521**, 314–325 (1993).
195. M.L. Fishman, P.D. Hoagland, 'Characterization of Starches Dissolved in Water by Microwave-heating in a High-pressure Vessel', *Carbohyd. Polym.*, **23**, 175–183 (1994).
196. A. Müller, H.A. Pretus, R.B. McNamee, E.L. Jones, I.W. Browder, D.L. Williams, 'Comparison of the Carbohydrate Biological Response Modifiers Krestin, Schizophyllan and Glucan Phosphate by Aqueous Size Exclusion Chromatography with Inline Argon-ion Multi-angle Laser Light Scattering Photometry and Differential Viscometry Detectors', *J. Chromatogr. B*, **666**, 283–290 (1995).
197. J. Simoni, G. Simoni, M. Feola, 'Chromatographic Analysis of Biopolymers Distribution in ''Poly-haemoglobin'',

an Intermolecularly Crosslinked Haemoglobin Solution', *Anal. Chim. Acta*, **279**, 73–88 (1993).
198. M.A.M. Hoffmann, G. Sala, C. Olieman, K.G. deKruif, 'Molecular Mass Distributions of Heat-induced β-Lactoglobulin Aggregates', *J. Agric. Food Chem.*, **45**, 2949–2957 (1997).
199. H. Lakhiari, T. Okano, N. Nurdin, C. Luthi, P. Descouts, D. Müller, J. Jozefonvicz, 'Temperature-responsive Size-exclusion Chromatography Using Poly(*N*-isopropylacrylamide) Grafted Silica', *Biochim. Biophys. Acta*, **1379**, 303–313 (1998).
200. L.G. Öfverstedt, K.O. Eriksson, 'Pre-column Concentration of Protein Samples by Displacement Electrophoresis (Isotachophoresis) followed by High-performance Molecular Sieve Chromatography on a Packed Agarose Column', *Anal. Biochem.*, **137**, 318–323 (1984).
201. H.P.J. Bennett, in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, eds R.S. Hodges, C.T. Mant, CRC Press, Boca Raton, 163–167, 1991.
202. A.M. Strigel, J.D. Timpa, 'Size Exclusion Chromatography of Polysaccharides in Dimethylacetamide–Lithium Chloride', *ACS Symp. Ser.*, **635**, 366–378 (1996).
203. B.J. Hunt, S.R. Holding (eds), *Size Exclusion Chromatography*, Blackie, Glasgow, 1989.
204. M. Potschka, P.L. Dubin (eds), *Strategies in Size Exclusion Chromatography*, (ACS Symp. Ser., Vol. 635), American Chemical Society, Washington, DC, 1996.
205. C.S. Wu, *Column Handbook for Size Exclusion Chromatography*, Academic Press, San Diego, 1998.
206. C.S. Wu, *Handbook of Size-exclusion Chromatography*, (Chromatogr. Sci., Vol. 69), Marcel Dekker, New York, 1995.
207. K.A. Lerro, R. Orlando, H.Z. Zhang, P.N.R. Usherwood, K. Nakanishi, 'Separation of the Sticky Peptides from Membrane Proteins by High-performance Liquid Chromatography in a Normal-phase System', *Anal. Biochem.*, **215**, 38–44 (1993).
208. T. Yoshida, 'Peptide Separation in Normal Phase Liquid Chromatography', *Anal. Chem.*, **69**, 3038–3043 (1997).
209. T. Yoshida, 'Calculation of Peptide Retention Coefficients in Normal-phase Liquid Chromatography', *J. Chromatogr. A*, **808**, 105–112 (1998).
210. T. Yoshida, 'Prediction of Peptide Retention Time in Normal-phase Liquid Chromatography', *J. Chromatogr. A*, **811**, 61–67 (1998).
211. K.H. Ekborg Ott, Y. Liu, D.W. Armstrong, 'Highly Enantioselective HPLC Separations Using the Covalently Bonded Macrocyclic Antibiotic, Ristocetin A, Chiral Stationary Phase', *Chirality*, **10**, 434–484 (1998).
212. D.W. Armstrong, Y.B. Liu, K.H. Ekborg-Ott, 'A Covalently Bonded Teicoplanin Chiral Stationary Phase for HPLC Enantioseparations', *Chirality*, **7**, 474–497 (1995).
213. M.A. Strege, 'Hydrophilic-interaction Chromatography – Electrospray Mass Spectrometry Analysis of Polar Compounds for Natural Product Drug Discovery', *Anal. Chem.*, **70**, 2439–2445 (1998).
214. J.A. Beutler, 'Separation of High Molecular Weight Saponins of *Archidendron ellipticum* by Hydrophilic Interaction Chromatography', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2415–2426 (1997).
215. S.C. Lin, W.C. Lee, 'Separation of Fructo-oligosaccharide Mixture by Hydrophilic Interaction Chromatography Using Silica-based Micropellicular Sorbents', *J. Chromatogr. A*, **803**, 302–306 (1998).
216. J. Yu, Z. El Rassi, 'Preparation of Amino-zirconia Bonded Phases and their Evaluation in Hydrophilic Interaction Chromatography of Carbohydrates with Pulsed Amperometric Detection', *J. High Resolut. Chromatogr.*, **17**, 773–778 (1994).
217. Partly adapted from *Care and Feeding for your HILIC Column*, Product documentation, PolyLC, Columbia, MD.
218. A.J. Alpert, M. Shukla, A.K. Shukla, L.R. Zieske, S.W. Yuen, M.A.J. Ferguson, A. Mehlert, M. Pauly, R. Orlando, 'Hydrophilic-interaction Chromatography of Complex Carbohydrates', *J. Chromatogr. A*, **676**, 191–202 (1994).
219. R.R. Townsend, P.H. Lipniunas, C. Bigge, A. Ventom, R. Parekh, 'Multimode High-performance Liquid Chromatography of Fluorescently Labelled Oligosaccharides from Glycoproteins', *Anal. Biochem.*, **239**, 200–207 (1996).
220. S.C. Churms, 'Recent Progress in Carbohydrate Separation by High-performance Liquid Chromatography based on Hydrophilic Interaction', *J. Chromatogr. A*, **720**, 75–91 (1996).
221. N. Takahashi, 'Three-dimensional Mapping of N-linked Oligosaccharides Using Anion-exchange, Hydrophobic and Hydrophilic-interaction modes of High-performance Liquid Chromatography', *J. Chromatogr. A*, **720**, 217–225 (1996).
222. A.R. Oyler, B.L. Armstrong, J.Y. Cha, M.X. Zhou, Q. Yang, R.I. Robinson, R. Dunphy, D.J. Burinsky, 'Hydrophilic Interaction Chromatography on Aminosilica Phases complements Reversed-phase High-performance Liquid Chromatography and Capillary Electrophoresis for Peptide Analysis', *J. Chromatogr. A*, **724**, 378–383 (1996).
223. C.T. Mant, J.R. Litowski, R.S. Hodges, 'Hydrophilic-interaction/Cation-exchange Chromatography for Separation of Amphipathic α-Helical Peptides', *J. Chromatogr. A*, **816**, 65–78 (1998).
224. C.T. Mant, L.H. Kondejewski, R.S. Hodges, 'Hydrophilic-interaction/Cation-exchange Chromatography for Separation of Cyclic Peptides', *J. Chromatogr. A*, **816**, 79–88 (1998).
225. H. Lindner, B. Sarg, W. Helliger, 'Application of Hydrophilic-interaction Liquid Chromatography to the Separation of Phosphorylated H1 Histones', *J. Chromatogr. A*, **782**, 55–62 (1997).
226. H. Lindner, B. Sarg, C. Meraner, W. Helliger, 'Separation of Acetylated Core Histones by Hydrophilic-interaction

Liquid Chromatography', *J. Chromatogr. A*, **743**, 137–144 (1996).
227. Z. El Rassi, 'Recent Progress in Reversed-phase and Hydrophobic-interaction Chromatography of Carbohydrate Species', *J. Chromatogr. A*, **720**, 93–118 (1996).
228. M. Hanson, A. Kurganov, K.K. Unger, V.A. Davankov, 'Polymer-coated Reversed-phase Packings in High-performance Liquid Chromatography', *J. Chromatogr. A*, **656**, 369–380 (1993).
229. K. Albert, E. Bayer, 'Characterization of Bonded Phases by Solid-state NMR Spectroscopy', *J. Chromatogr.*, **544**, 345–370 (1991).
230. I. Yarovsky, M.I. Aguilar, M.T.W. Hearn, 'Influence of the Chain Length and Surface Density on the Conformation and Mobility of *n*-Alkyl Ligands Chemically Immobilized onto a Silica Surface', *Anal. Chem.*, **67**, 2145–2153 (1995).
231. M.T.W. Hearn (ed.), *Ion-pair Chromatography: theory and Biological and Pharmaceutical Applications*, Dekker, New York, 1985.
232. G. Schill, 'High-performance Ion-pair Chromatography', *J. Biochem. Biophys. Meth.*, **18**, 249–270 (1989).
233. M.I. Aguilar, D.J. Clayton, P. Holt, V. Kronina, R.I. Boysen, A.W. Purcell, M.T.W. Hearn, 'RP HPLC Binding Domains of Proteins', *Anal. Chem.*, **70**, 5010–5018 (1998).
234. M. Harada, 'High-performance Liquid-chromatographic Determination of Peptidase Activity toward Proline-containing Peptides', *Anal. Chim. Acta*, **352**, 179–185 (1997).
235. J. Meriluoto, 'Chromatography of Microcystins', *Anal. Chim. Acta*, **352**, 277–298 (1997).
236. T.J. Sereda, C.T. Mant, R.S. Hodges, 'Use of Sodium Perchlorate at low pH for Peptide Separations by Reversed-phase Liquid Chromatography. Influence of Perchlorate Ion on apparent Hydrophilicity of Positively Charged Amino-acid Side-chains', *J. Chromatogr. A*, **776**, 153–165 (1997).
237. D.M. Desiderio, 'Analysis of Endogenous Neuropeptides by Reversed-phase High-performance Chromatography and Mass Spectrometry', *Anal. Chim. Acta*, **352**, 85–102 (1997).
238. D.J. Hammond, 'Identification of Affinity Ligands from Peptide Libraries and their Applications', *Chromatographia*, **47**, 475–476 (1998).
239. D.M. Desiderio, X. Zhu, 'Quantitative Analysis of Methionine Enkephalin and β-Endorphin in the Pituitary by Liquid Secondary-ion Mass Spectrometry and Tandem Mass Spectrometry', *J. Chromatogr. A*, **794**, 85–96 (1998).
240. F. Regnier, G. Huang, 'Future Potential of Targeted Component Analysis by Multidimensional Liquid Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **750**, 3–10 (1996).
241. C.T. Mant, R.S. Hodges, 'Optimization and Prediction of Peptide Retention Behavior in Reversed-phase Chromatography', in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, eds. C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 277–305, 1991.
242. S. Welling-Wester, M. Feijlbrief, D.G.A.M. Koedijk, M.A. Braaksma, B.R.K. Douma, G.W. Welling, 'Effect of Different Amounts of the Non-ionic Detergents C10E5 and C12E5 Present in Eluents for Ion-exchange High-performance Liquid Chromatography of Integral Membrane Proteins of Sendai virus', *J. Chromatogr.*, **646**, 37–44 (1993).
243. Y. Kato, 'High-performance Hydrophobic Interaction Chromatography of Proteins', *Adv. Chromatogr.*, **26**, 97–115 (1987).
244. N. Cooke, P. Shieh, N. Miller, 'High-performance Hydrophobic-interaction Chromatography of Proteins', *LC-GC*, **7**, 954, 956–958 (1989).
245. L. Szepesy, G. Rippel, 'High-performance Hydrophobic Interaction Chromatography of Proteins', *LC GC Int.*, **5**, 24–29 (1992).
246. K. Leopold, 'Hydrophobic Interaction Chromatography', *GIT Fachz. Lab.*, **38**, 455–456, 459–460 (1994).
247. S.L. Wu, B.L. Karger, 'Hydrophobic Interaction Chromatography of Proteins', *Methods Enzymol.*, **270**, 27–47 (1996).
248. A. Tiselius, 'Adsorption Separation by Salting Out', *Arkiv Kemi, Mineral. Geol.*, **26B**, 1–5 (1948).
249. J. Porath, 'Gel Filtration of Proteins, Peptides, and Amino Acids', *Biochim. Biophys. Acta*, **39**, 193–207 (1960).
250. K.D. Collins, M.W. Washabaugh, 'The Hofmeister Effect and the Behavior of Water at Interfaces', *Quart. Rev. Biophys.*, **18**, 323–422 (1985).
251. M.G. Cacace, E.M. Landau, J.J. Ramsden, 'The Hofmeister Series: Salt and Solvent effects on Interfacial Phenomena', *Quart. Rev. Biophys.*, **30**, 241–277 (1997).
252. R.G. Hatch, 'Chromatography of Proteins on a Silica-based Support with Polyoxyethylene Glycol Ligands', *J. Chromatogr. Sci.*, **28**, 210–214 (1990).
253. P. Hubert, R. Mathis, E. Dellacherie, 'Polymer Ligands for Mild Hydrophobic Interaction Chromatography – Principles, Achievements and Future Trends', *J. Chromatogr.*, **539**, 297–306 (1991).
254. P. Smidl, I. Kleinmann, J. Plicka, V. Svoboda, 'Hydrophobic-interaction Chromatography of Proteins on Separon HEMA [Hydroxyethylmethacrylate]. II. Influence of Sorbent Modification on Efficiency of Separation', *J. Chromatogr.*, **523**, 131–138 (1990).
255. M. Bryjak, A. Trochimczuk, J. Bryjak, 'Evaluation of Terpolymers of 2-Hydroxyethyl Methacrylate, Dodecyl Methacrylate and Ethylene Glycol Dimethacrylate as Matrices for Hydrophobic Interaction Chromatography', *Angew. Makromol. Chem.*, **242**, 147–157 (1996).
256. G. Rippel, E. Alattyani, L. Szepesy, 'Characterization of Stationary Phases used in Reversed-phase and Hydrophobic Interaction Chromatography', *J. Chromatogr. A*, **668**, 301–312 (1994).

257. Y. Kato, T. Kitamura, T. Hashimoto, 'Operational Variables in High-performance Hydrophobic Interaction Chromatography of Proteins on TSKgel Phenyl-5PW', *J. Chromatogr.*, **298**, 407–418 (1984).
258. T. Kitamura, Y. Kato, 'Effect of the Type of Salt on Separation of Proteins in High-performance Hydrophobic Interaction Chromatography on TSKgel Phenyl-5PW', *J. High Resolut. Chromatogr.*, **8**, 306–307 (1985).
259. J. Plicka, P. Smidl, I. Kleinmann, V. Svoboda, 'Hydrophobic-interaction Chromatography of Proteins on Separon HEMA [Hydroxyethylmethacrylate]. III. Selection of suitable Gradient Conditions for the Separation of Proteins by Hydrophobic-interaction Chromatography', *J. Chromatogr.*, **523**, 139–150 (1990).
260. M.L. Heinitz, L. Kennedy, W. Kopaciewicz, F.E. Regnier, 'Chromatography of Proteins on Hydrophobic Interaction and Ion-exchange Chromatographic Matrices: Mobile Phase Contributions to Selectivity', *J. Chromatogr.*, **443**, 173–182 (1988).
261. A. Onishi, M.O. Proudlove, 'Isolation of Beer Foam Polypeptides by Hydrophobic Interaction Chromatography and their Partial Characterization', *J. Sci. Food Agric.*, **65**, 233–240 (1994).
262. T. Arakawa, L.O. Narhi, 'Solvent Modulation in Hydrophobic Interaction Chromatography', *Biotechnol. Appl. Biochem.*, **13**, 151–172 (1991).
263. L. Pons, A. Olszewski, J.L. Gueant, 'Characterization of the Oligomeric Behavior of a 16.5 kDa Peanut Oleosin by Chromatography and Electrophoresis of the Iodinated Form', *J. Chromatogr. B, Biomed. Sci. Appl.*, **706**, 131–140 (1998).
264. J.J. Buckley, D.B. Wetlaufer, 'Surfactant-mediated Hydrophobic Interaction Chromatography of Proteins: Gradient Elution', *J. Chromatogr.*, **518**, 99–110 (1990).
265. Y.L. Sing, Y. Kroviarski, S. Cochet, D. Dhermy, O. Bertrand, 'High-performance Hydrophobic Interaction Chromatography of Proteins on Reversed-phase Supports coated with Non-ionic Surfactants of Polyoxyethylene type. Purification of a Fungal Aspartic Proteinase', *J. Chromatogr.*, **598**, 181–187 (1992).
266. J. Huddleston, J.C. Abelaira, R.D. Wang, A. Lyddiatt, 'Protein Partition Between the Different Phases comprising Poly(ethylene glycol)–Salt Aqueous Two-phase Systems, Hydrophobic Interaction Chromatography and Precipitation: a Generic Description in terms of Salting-out Effects', *J. Chromatogr. B, Biomed. Sci. Appl.*, **680**, 31–41 (1996).
267. T. Konishi, M. Kamada, H. Nakamura, 'Evaluation of Ammonium Acetate as a Volatile Buffer for High-performance Hydrophobic-interaction Chromatography', *J. Chromatogr.*, **515**, 279–283 (1990).
268. H. Nakamura, T. Konishi, M. Kamada, 'Use of Volatile Buffers in High-performance Hydrophobic-interaction Chromatography of Proteins', *Anal. Sci.*, **6**, 137–138 (1990).
269. W.K. Feng, X.Q. Jiang, X.D. Geng, 'The Influence of Urea on the Retention of Protein in Hydrophobic Interaction Chromatography', *J. Liq. Chromatogr.*, **18**, 217–226 (1995).
270. W.R. Melander, D. Corradini, Cs. Horváth, 'Salt-mediated Retention of Proteins in Hydrophobic-interaction Chromatography. Application of Solvophobic Theory', *J. Chromatogr.*, **317**, 67–85 (1984).
271. X. Geng, L. Guo, J. Chang, 'Study of the Retention Mechanism of Proteins in Hydrophobic Interaction Chromatography', *J. Chromatogr.*, **507**, 1–23 (1990).
272. S. Oscarsson, P. Karsnäs, 'Salt-promoted Adsorption of Proteins onto Amphiphilic Agarose-based Adsorbents. II. Effects of Salt and Salt Concentration', *J. Chromatogr. A*, **803**, 83–93 (1998).
273. P.M. Brandts, W.J. Gelsema, C.L. De Ligny, 'Influence of Additives to the Eluent on the Hydrophobic Interaction Chromatography of Simple Compounds. IV. Influence of Electrolytes on the Retention of Normal Alkanols at 25 Degree', *J. Chromatogr.*, **437**, 337–350 (1988).
274. P.M. Brandts, W.J. Gelsema, C.L. De Ligny, 'Influence of Additives to the Eluent on Hydrophobic Interaction Chromatography of Simple Compounds. V. Influence of pH and Electrolytes on the Retention of *n*-Alkylammonium Ions and *n*-Alkyl Carboxylates', *J. Chromatogr.*, **438**, 181–195 (1988).
275. K. Makino, H. Wada, H. Ozaki, T. Takeuchi, H. Hatano, T. Fukui, K. Noguchi, Y. Yanagihara, 'Separation of Single-stranded Oligonucleotides on Poly(vinyl alcohol) Gel Column', *Chromatographia*, **20**, 713–716 (1985).
276. M. De Frutos, A. Cifuentes, J.C. Dìez-Masa, 'Multiple Peaks in High-performance Liquid Chromatography of Proteins. β-Lactoglobulins eluted in a Hydrophobic Interaction Chromatography System', *J. Chromatogr. A*, **778**, 43–52 (1997).
277. K.L. Richards, M.I. Aguilar, M.T.W. Hearn, 'Comparative Study of the Retention Behaviour and Stability of Cytochrome *c* in Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. A*, **676**, 17–31 (1994).
278. A.W. Purcell, M.I. Aguilar, M.T. Hearn, 'Probing the Binding Behavior and Conformational States of Globular Proteins in Reversed-phase High-performance Liquid Chromatography', *Anal. Chem.*, **71**, 2440–2451 (1999).
279. W. Feng, X. Zhu, L. Zhang, X. Geng, 'Retention Behaviour of Proteins under Conditions of Column Overload in Hydrophobic Interaction Chromatography', *J. Chromatogr. A*, **729**, 43–47 (1996).
280. P.D.G. Dean, W.S. Johnson, F.A. Middle (eds), *Affinity Chromatography: a Practical Approach*, IRL Press: Oxford, UK, 1985.
281. P.O. Larsson, 'High-performance Liquid Affinity Chromatography', *Methods Enzymol.*, **104**, 212–223 (1987).

282. M.A. Vijayalakshmi, O. Bertrand (eds), *Protein–Dye Interactions: Development and Applications*, Elsevier Applied Science, London, 1989.
283. E. Klein, *Affinity Membranes. Their Chemistry and Performance in Adsorptive Separation Processes*, John Wiley & Sons, Chichester, 1991.
284. G.T. Hermanson, A.K. Mallia, P.K. Smith, *Immobilized Affinity Ligand Techniques*, Academic Press, San Diego, 1992.
285. T. Kline (ed.), *Handbook of Affinity Chromatography*, (Chromatogr. Sci. Ser., Vol. 63), Marcel Dekker, New York, 1993.
286. R. Kaliszan, 'Retention Data from Affinity High-performance Liquid Chromatography in view of Chemometrics', *J. Chromatogr. B*, **715**, 229–244 (1998).
287. D.R. Hall, D.J. Winzor, 'Potential of Biosensor Technology for the Characterization of Interactions by Quantitative Affinity Chromatography', *J. Chromatogr. B*, **715**, 163–181 (1998).
288. J. Turková, 'Bioaffinity Chromatography' in *Analytical and Preparative Separation Methods of Biomacromolecules*, ed. H.Y. Aboul-Enein, Marcel Dekker, New York, 1999.
289. C. Kasper, L. Meringova, R. Freitag, T.B. Tennikova, 'Fast Isolation of Protein Receptors from Streptococci G by means of Macroporous Affinity Discs', *J. Chromatogr. A*, **798**, 65–72 (1998).
290. M. de Frutos, S.K. Paliwal, F. Regnier, 'Analytical Immunology', *Methods Enzymol.*, **270**, 82–101 (1996).
291. J. Porath, 'General Methods and Coupling Procedures', *Methods Enzymol.*, **34**, 13–30 (1974).
292. M. Leonard, C. Fournier, E. Dellacherie, 'Poly(vinyl alcohol)-coated Macroporous Polystyrene Particles as Stationary Phases for the Chromatography of Proteins', *J. Chromatogr. B*, **664**, 39–46 (1995).
293. D.C. Nash, G.E. McCreath, H.A. Chase, 'Modification of Polystyrenic Matrices for the Purification of Proteins. Effect of the Adsorption of Poly(vinyl alcohol) on the Characteristics of Poly(styrene divinylbenzene) Beads for use in Affinity Chromatography', *J. Chromatogr. A*, **758**, 53–64 (1997).
294. D.C. Nash, H.A. Chase, 'Modification of Polystyrenic Matrices for the Purification of Proteins. II. Effect of the Degree of Glutaraldehyde–Poly(vinyl alcohol) crosslinking on various Dye Ligand Chromatography Systems', *J. Chromatogr. A*, **776**, 55–63 (1997).
295. J. Carlsson, J.C. Jansson, M. Sparrmann, 'Affinity Chromatography', in *Protein Purification. Principles, High Resolution Methods, and Applications*, 2nd edition, eds J.C. Janson, L. Rydén, Wiley-Liss, New York, 375–442, 1998.
296. P. Wikström, C.F. Mandenius, P.O. Larsson, 'Gas Phase Silylation, a Rapid Method for Preparation of High-performance Liquid Chromatography Supports', *J. Chromatogr.*, **455**, 105–117 (1988).
297. H.J. Wirth, K.K. Unger, M.T. Hearn, 'Influence of Ligand Density on the Properties of Metal–Chelate Affinity Supports', *Anal. Biochem.*, **208**, 16–25 (1993).
298. J.N. Kinkel, B. Anspach, K.K. Unger, R. Wieser, G. Brunner, 'Separation of Plasma Membrane Proteins of Cultured Human Fibroblasts by Affinity Chromatography on Bonded Microparticulate Silicas', *J. Chromatogr.*, **297**, 167–177 (1984).
299. E. Hagemeier, K.S. Boos, E. Schlimme, K. Lechtenboerger, A. Kettrup, 'Synthesis and Application of a Boronic Acid-substituted Silica for High-performance Liquid Affinity Chromatography', *J. Chromatogr.*, **268**, 291–295 (1983).
300. F.B. Anspach, H.J. Wirth, K.K. Unger, P. Stanton, J.R. Davies, M.T. Hearn, 'High-performance Liquid Affinity Chromatography with Phenylboronic Acid, Benzamidine, Tri-L-alanine, and Concanavalin A Immobilized on 3-Isothiocyanatopropyltriethoxysilane-activated Non-porous Monodisperse Silicas', *Anal. Biochem.*, **179**, 171–181 (1989).
301. J.B. Wheatley, D.E. Schmidt, Jr, 'Salt-induced Immobilization of Proteins on a High-performance Liquid-chromatographic Epoxide Affinity Support', *J. Chromatogr.*, **644**, 11–16 (1993).
302. J.B. Wheatley, D.E. Schmidt, Jr, 'Salt-induced Immobilization of Affinity Ligands onto Epoxide-activated Supports', *J. Chromatogr. A*, **849**, 1–12 (1999).
303. F.E. Regnier, R. Noel, 'Glycerolpropylsilane Bonded Phases in the Steric Exclusion Chromatography of Biological Macromolecules', *J. Chromatogr. Sci.*, **14**, 316–320 (1976).
304. K. Nilsson, P.O. Larsson, 'High-performance Liquid Affinity Chromatography on Silica-bound Alcohol Dehydrogenase', *Anal. Biochem.*, **134**, 60–72 (1983).
305. M.T. Hearn, 'Dynamic Multizoning Effects in Biospecific Affinity Chromatography on Porous Supports: Evaluation of Activation and Ligand Coupling Effects with Different Support Materials', *J. Chromatogr.*, **376**, 245–257 (1986).
306. M.T. Hearn, '1,1′-Carbonyldiimidazole-mediated Immobilization of Enzymes and Affinity Ligands', *Methods Enzymol.*, **135**, 102–117 (1987).
307. H.W. Jarrett, 'Development of *N*-Hydroxysuccinimide Ester Silica, a Novel Support for High-performance Affinity Chromatography', *J. Chromatogr.*, **405**, 179–189 (1987).
308. M. Hill, B. Arrio, 'Activation of Matrices by 4,6-Diphenylthieno[3,4-*d*]-1,3-dioxol-2-one 5,5-dioxide. High-performance Liquid Affinity Chromatographic Separations', *J. Chromatogr.*, **589**, 101–108 (1992).
309. P.F. Ruhn, S. Garver, D.S. Hage, 'Development of Dihydrazide-activated Silica Supports for High-performance Affinity Chromatography', *J. Chromatogr. A*, **669**, 9–19 (1994).
310. P. Wikström, P.O. Larsson, 'Affinity Fiber – a New Support for Rapid Enzyme Purification by High-performance

Liquid Affinity Chromatography', *J. Chromatogr.*, **388**, 123–134 (1987).
311. K. Ernst-Cabrera, M. Wilchek, 'Silica Containing Primary Hydroxyl Groups for High-performance Affinity Chromatography', *Anal. Biochem.*, **159**, 267–272 (1986).
312. K. Ernst-Cabrera, M. Wilchek, 'Coupling of Ligands to Primary Hydroxyl-containing Silica for High-performance Affinity Chromatography. Optimization of Conditions', *J. Chromatogr.*, **397**, 187–196 (1987).
313. J.B. Wheatley, M.H. Lyttle, M.D. Hocker, D.E. Schmidt, Jr, 'Salt-induced Immobilization of DNA Oligonucleotides on an Epoxide-activated High-performance Liquid-chromatographic Affinity Support', *J. Chromatogr. A*, **726**, 77–90 (1996).
314. P.G. Mingalyov, N.V. Orishchenko, A.Yu. Fadeev, 'Study of Correlations Between Fine Structure of Bonded Layer and Affinity Properties of Silicas with Attached Cyclic Oligopeptides', *J. Chromatogr. A*, **777**, 249–259 (1997).
315. J. Turková, 'Oriented Immobilization of Biologically Active Proteins as a Tool for Revealing Protein Interactions and Functions', *J. Chromatogr. B*, **722**, 11–31 (1999).
316. V. Adamek, X.C. Liu, Y.A. Zhang, K. Adamkova, W.H. Scouten, 'New Aliphatic Boronate Ligands for Affinity Chromatography', *J. Chromatogr.*, **625**, 91–99 (1992).
317. X.C. Liu, W.H. Scouten, 'New Ligands for Boronate Affinity Chromatography', *J. Chromatogr. A*, **687**, 61–69 (1994).
318. R.P. Singhai, S. De Silva, M. Shyamali, 'Boronate Affinity Chromatography', *Adv. Chromatogr.*, **31**, 293–335 (1992).
319. I. Miksik, Z. Deyl, 'Post-translational Non-enzymic Modification of Proteins II. Separation of Selected Protein Species after Glycation and other Carbonyl-mediated Modifications', *J. Chromatogr. B*, **699**, 311–345 (1997).
320. B.M. Brena, C. Pazos, L. Franco-Fraguas, F. Batista-Viera, 'Chromatographic Methods for Amylases', *J. Chromatogr. B*, **684**, 217–237 (1996).
321. S. Hjertén, J.P. Li, 'High-performance Liquid Chromatography of Proteins on Deformed Non-porous Agarose Beads. Fast Boronate Affinity Chromatography of Haemoglobin at Neutral pH', *J. Chromatogr.*, **500**, 543–553 (1990).
322. K. Yasukawa, F. Abe, N. Shida, Y. Koizumi, T. Uchida, K. Noguchi, K. Shima, 'High-performance Affinity Chromatography System for the Rapid, Efficient Assay of Glycosylated Albumin', *J. Chromatogr.*, **597**, 271–275 (1992).
323. T. Koyama, K. Terauchi, 'Synthesis and Application of Boronic Acid-immobilized Porous Polymer Particles: a Novel Packing for High-performance Liquid Affinity Chromatography', *J. Chromatogr. B*, **679**, 31–40 (1996).
324. N. Singh, R.C. Willson, 'Boronate Affinity Adsorption of RNA: Possible Role of Conformational Changes', *J. Chromatogr. A*, **840**, 205–213 (1999).
325. J. Jankowski, W. Potthoff, M. van der Giet, M. Tepel, W. Zidek, H. Schlueter, 'High-performance Liquid-chromatographic Assay of the Diadenosine Polyphosphates in Human Platelets', *Anal. Biochem.*, **269**, 72–78 (1999).
326. E. Schlimme, K.S. Boos, E. Hagemeier, K. Kemper, U. Meyer, H. Hobler, T. Schnelle, M. Weise, 'Direct Clean-up and Analysis of Ribonucleosides in Physiological Fluids', *J. Chromatogr.*, **378**, 349–360 (1986).
327. A. Rudolphi, K.S. Boos, D. Seidel, 'Coupled-column HPLC Analysis of Free Urinary Catecholamines Using Restricted Access Affinity Precolumn and Microparticulate Nonporous Silica Analytical Column', *Chromatographia*, **41**, 645–650 (1995).
328. G.G. Granich, D.J. Krogstad, J.D. Connor, K.L. Desrochers, C. Sherwood, 'High-performance Liquid Chromatography (HPLC) Assay for Ribavirin and Comparison of the HPLC Assay with Radio-immunoassay', *Antimicrob. Agents. Chemother.*, **33**, 311–315 (1989).
329. L. Hansson, M. Glad, C. Hansson, 'Boronic Acid–Silica: a New Tool for Purification of Catecholic Compounds On-line with Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **265**, 37–44 (1983).
330. E. Hagemeier, K.S. Boos, E. Schlimme, K. Lechtenboerger, A. Kettrup, 'Synthesis and Application of a Boronic Acid-substituted Silica for High-performance Liquid Affinity Chromatography', *J. Chromatogr.*, **268**, 291–295 (1983).
331. E. Riva, N. Ferry, A. Cometti, G. Cuisinaud, G.C. Gallo, J. Sassard, 'Determination of Teicoplanin in Human Plasma and Urine by Affinity and Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **421**, 99–110 (1987).
332. M.S. Westwell, D.H. Williams, 'The Fight Against Antibiotic Resistant Bacteria', *Chemtech*, **26**, 210–215 (1996).
333. A. Berthod, Y. Liu, C. Bagwill, D.W. Armstrong, 'Facile Liquid Chromatographic Enantioresolution of Native Amino Acids and Peptides Using a Teicoplanin Chiral Stationary Phase', *J. Chromatogr. A*, **731**, 123–137 (1996).
334. A. Pingali, B. McGuinness, H. Keshishian, F.W. Jing, L. Varady, F. Regnier, 'Peptides as Affinity Surfaces for Protein Purification', *J. Mol. Recogn.*, **9**, 426–432 (1996).
335. C. Kuyas, A. Haeberli, P. Walder, P.W. Straub, 'Isolation of Human Fibrinogen and its Derivatives by Affinity Chromatography on Gly-Pro-Arg-Pro-Lys-Fractogel', *Thromb. Haemost.*, **63**, 439–444 (1990).
336. P.A. Adams, R.D. Goold, A.A. Thumser, 'Heme-Peptide/Protein Interactions: the Binding of Heme Octa and Undecapeptides, and Microperoxidase-8 and -11, to Human Serum Albumin', *J. Inorg. Biochem.*, **37**, 91–103 (1989).

337. P. Formstecher, H. Hammadi, N. Bouzerna, M. Dautrevaux, 'Rapid Purification of Antisteroid Antibodies by High-performance Liquid Affinity Chromatography', *J. Chromatogr.*, **369**, 379–390 (1986).
338. I. Csiky, L. Hansson, 'High-performance Liquid Affinity Chromatography (HPLAC) of Sterols with Tomatine Chemically Bonded to Micro-particulate Silica', *J. Liq. Chromatogr.*, **9**, 875–886 (1986).
339. E.A. Bayer, M. Willchek, 'Avidin–Biotin Technology', *Methods Enzymol.*, **184** (1990).
340. H.L. Ball, G. Bertolini, S. Levi, P. Mascagni, 'Purification of Synthetic Peptides with the Aid of Reversible Chromatographic Probes', *J. Chromatogr. A*, **686**, 73–83 (1994).
341. C. Thiele, F. Fahrenholz, 'Photo-cleavable Biotinylated Ligands for Affinity Chromatography', *Anal. Biochem.*, **218**, 330–337 (1994).
342. G. Ehteshami, J. Porath, R. Guzman, 'Interactions and Applications of Soluble Heterobifunctional Affinity Chelating Polymers in Immobilized Metal Affinity Chromatography', *J. Mol. Recogn.*, **9**, 733–737 (1996).
343. T.G.M. Schmidt, A. Skerra, 'One-step Affinity Purification of Bacterially Produced Proteins by Means of the ''Strep Tag'' and Immobilized Recombinant Core Streptavidin', *J. Chromatogr. A*, **676**, 337–345 (1994).
344. P.J. Schatz, 'Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-modifying Enzyme: a 13 Residue Consensus Peptide Specifies Biotinylation in *Escherichia coli*', *Biotechnology*, **11**, 1138–1143 (1993).
345. J. Tucker, R. Grisshammer, 'Purification of a Rat Neurotensin Receptor Expressed in *Escherichia coli*', *Biochem. J.*, **317**, 891–899 (1996).
346. Q. Yang, X.Y. Liu, S.I. Ajiki, M. Hara, P. Lundahl, J. Miyake, 'Avidin–Biotin Immobilization of Unilamellar Liposomes in Gel Beads for Chromatographic Analysis of Drug–Membrane Partitioning', *J. Chromatogr. B*, **707**, 131–141 (1998).
347. J.B. Wheatley, M.K. Kelley, J.A. Montali, C.O.A. Berry, D.E. Schmidt, Jr, 'Examination of Glutathione *S*-Transferase Isoenzyme Profiles in Human Liver Using High-performance Affinity Chromatography', *J. Chromatogr. A*, **663**, 53–63 (1994).
348. J.B. Wheatley, J.A. Montali, D.E. Schmidt, Jr, 'Coupled Affinity–Reversed-phase High-performance Liquid Chromatography Systems for the Measurement of Glutathione *S*-Transferases in Human Tissues', *J. Chromatogr. A*, **676**, 65–79 (1994).
349. J.B. Wheatley, B. Hughes, K. Bauer, D.E. Schmidt, Jr, 'Study of Chromatographic Parameters for Glutathione *S*-Transferases on a High-performance Liquid Chromatography Affinity Stationary Phase', *J. Chromatogr. A*, **676**, 81–90 (1994).
350. K. Shimura, K. Kasai, S. Ishii, 'High-performance Affinity Chromatography of Anhydrochymotrypsin on a Hydrophilic Vinyl-polymer Gel Coupled with Tryptophan', *J. Chromatogr.*, **350**, 265–272 (1985).
351. J. Hradec, F. Franek, P. Dufek, 'Purification of Cholesterol-esterifying Enzymes from Rat Liver Cytosol by High-performance Liquid Chromatography', *J. Chromatogr. B*, **681**, 55–62 (1996).
352. R. Freitag, 'Utilization of Enzyme–Substrate Interactions in Analytical Chemistry', *J. Chromatogr. B*, **722**, 279–301 (1999).
353. J. Alhama, J. Lopez-Barea, F. Toribio, 'High-performance Affinity Chromatography of $NADP^+$ Dehydrogenases from Cell-free Extracts Using a Nucleotide Analogue as General Ligand', *J. Chromatogr.*, **586**, 51–59 (1991).
354. J. Stovickova, V. Podrazky, K. Trnavsky, 'Affinity Chromatography of Serine Proteinases from the Bovine Intervertebral Disc on BPTI-Separon HEMA 1000', *J. Chromatogr.*, **575**, 229–235 (1992).
355. T. Kumazaki, K. Terasawa, S. Ishii, 'Affinity Chromatography on Immobilized Anhydrotrypsin: General Utility for Selective Isolation of C-Terminal Peptides from Protease Digests of Proteins', *J. Biochem. Tokyo*, **102**, 1539–1546 (1987).
356. T. Kumazaki, M. Kobayashi, S. Ishii, 'Anhydroelastase: Enhanced Affinity Toward Product-type Ligands Revealed by Affinity Chromatography', *J. Mol. Recognit.*, **1**, 93–98 (1988).
357. R. Usha, M. Singh, 'Purification of a Multicatalytic Protease Complex from Developing Winged Bean Seeds by Indirect Immunoaffinity Chromatography', *Protein Expr. Purif.*, **15**, 48–56 (1999).
358. M.P. Vivarat-Perrin, B. Sebille, C. Vidal-Madjar, 'Study of Affinity Supports Based on Reactive Polymers Immobilized on Silica: Affinity Constant Determination from Isocratic Zonal Elution', *J. Chromatogr.*, **584**, 3–9 (1992).
359. S. Khamlichi, D. Müller, R. Fuks, J. Jozefonvicz, 'Specific Adsorption of Serine Proteases on Coated Silica Beads Substituted with Amidine Derivatives', *J. Chromatogr.*, **510**, 123–132 (1990).
360. D.E. Zembower, C.L. Neudauer, M.J. Wick, M.M. Ames, 'Peptide Boronic Acids. Versatile Synthetic Ligands for Affinity Chromatography of Serine Proteinases', *Int. J. Pept. Protein. Res.*, **47**, 405–13 (1996).
361. R.A. Beck, B. Munch-Petersen, M. Dolker, L. Cloos, G. Tyrsted, K. Eger, 'Ligands for the Affinity Chromatography of Mammalian Thymidine Kinase. 1: Strategy, Synthesis and Evaluation', *Pharm. Acta Helv.*, **71**, 279–291 (1996).
362. I. Chaiken, 'Interactions and Uses of Antisense Peptides in Affinity Technology', *J. Chromatogr.*, **597**, 29–36 (1992).
363. C. Jones, A. Patel, S. Griffin, J. Martin, P. Young, K. O'Donnell, C. Silverman, T. Porter, I. Chaiken, 'Current Trends in Molecular Recognition and Bioseparation', *J. Chromatogr. A*, **707**, 3–22 (1995).
364. W.S. Foster, H.W. Jarrett, 'Melittin–Silica, a High-pressure Affinity Chromatography Resin for Calmodulin', *J. Chromatogr.*, **403**, 99–107 (1987).

365. H. Mottl, W. Keck, 'Rapid Screening of a Large Number of Immobilized Textile Dyes for the Purification of Proteins: Use of Penicillin-binding Protein 4 of *Escherichia coli* as a Model Enzyme', *Protein Expr. Purif.*, **3**, 403–9 (1992).
366. P.D. Dean, D.H. Watson, 'Protein Purification Using Immobilised Triazine Dyes', *J. Chromatogr.*, **165**, 301–19 (1979).
367. S. Subramanian, 'Dye–Ligand Affinity Chromatography: the Interaction of Cibacron Blue F3GA with Proteins and Enzymes', *CRC. Crit. Rev. Biochem.*, **16**, 169–205 (1984).
368. L. Miribel, E. Gianazza, P. Arnaud, 'The Use of Dye–Ligand Affinity Chromatography for the Purification of Non-enzymatic Human Plasma Proteins', *J. Biochem. Biophys. Methods*, **16**, 1–15 (1988).
369. D.M. Worrall, 'Dye–Ligand Affinity Chromatography', *Methods. Mol. Biol.*, **59**, 169–76 (1996).
370. G.M. Finette, Q.M. Mao, M.T. Hearn, 'Comparative Studies on the Isothermal Characteristics of Proteins Adsorbed under Batch Equilibrium Conditions to Ion-exchange, Immobilised Metal Ion Affinity and Dye Affinity Matrices with Different Ionic Strength and Temperature Conditions', *J. Chromatogr. A*, **763**, 71–90 (1997).
371. C. Koch, L. Borg, K. Skjödt, G. Houen, 'Affinity Chromatography of Serine Proteases on the Triazine Dye Ligand Cibacron Blue F3G-A', *J. Chromatogr. B*, **718**, 41–46 (1998).
372. P.M. Boyer, J.T. Hsu, 'Protein Purification by Dye–Ligand Chromatography', *Adv. Biochem. Eng. Biotechnol.*, **49**, 1–44 (1993).
373. V. Dommes, W. Luster, M. Cvetanovic, W.H. Kunau, 'Purification by Affinity Chromatography of 2,4-Dienoyl-CoA Reductases from Bovine Liver and *Escherichia coli*', *Eur. J. Biochem.*, **125**, 335–341 (1982).
374. R.K. Scopes, K. Griffiths-Smith, 'Use of Differential Dye-ligand Chromatography with Affinity Elution for Enzyme Purification: 6-Phosphogluconate Dehydratase from *Zymomonas mobilis*', *Anal. Biochem.*, **136**, 530–534 (1984).
375. R.M. Pinto, M.J. Costas, A. Fernandez, J. Canales, J.A. Garcia-Agundez, J.C. Cameselle, 'Dinucleoside Tetraphosphatase from Human Blood Cells. Purification and Characterization as a High Specific Activity Enzyme Recognized by an Anti-rat Tetraphosphatase Antibody', *FEBS Lett.*, **287**, 85–88 (1991).
376. R.K. Scopes, 'Use of Differential Dye–Ligand Chromatography with Affinity Elution for Enzyme Purification: 2-Keto-3-deoxy-6-phosphogluconate Aldolase from *Zymomonas mobilis*', *Anal. Biochem.*, **136**, 525–529 (1984).
377. J.L. Drocourt, D.C. Thang, M.N. Thang, 'Blue-dextran – Sepharose Affinity Chromatography: Recognition of a Polynucleotide Binding Site of a Protein', *Eur. J. Biochem.*, **82**, 355–362 (1978).
378. C.E. Grimshaw, 'Chromatographic Separation of Activated and Unactivated forms of Aldose Reductase', *Arch. Biochem. Biophys.*, **278**, 273–276 (1990).
379. P. Hughes, C.R. Lowe, R.F. Sherwood, 'Metal Ion-promoted Binding of Proteins to Immobilized Triazine Dye Affinity Adsorbents', *Biochim. Biophys. Acta*, **700**, 90–100 (1982).
380. S. Subramanian, 'Separation of Chymosin and Pepsin in Calf Rennet by Dye–Ligand Affinity Chromatography', *Prep. Biochem.*, **17**, 297–312 (1987).
381. L. Miribel, P. Arnaud, 'Fractionation of Plasma Proteins Using Dye–Ligand Chromatography: Elution Profiles on Immobilized Green TSK-AF', *J. Biochem. Biophys. Methods*, **14**, 291–302 (1987).
382. F.M. Vallette, D.J. Marsh, F. Muller, J. Massoulie, B. Marcot, C. Viel, 'Comparative Affinity Chromatography of Acetylcholinesterases from Five Vertebrate Species', *J. Chromatogr.*, **257**, 285–296 (1983).
383. J.L. Taylor, W.K. Allmond, C.M. Himel, 'Affinity Chromatography of Acetylcholinesterase from Electrophorus Electricus Electroplax. Investigations on 9-Aminoalkylacridine Affinity Ligands', *J. Chromatogr.*, **257**, 275–284 (1983).
384. R.T. Carroll, J.L. Grimm, T.W. Hepburn, M.R. Emmerling, 'Purification of Acetylcholinesterase by Tacrine Affinity Chromatography', *Protein Expr. Purif.*, **6**, 389–393 (1995).
385. J. Scoble, R. Scopes, 'Well Defined Dye Adsorbents for Protein Purification', *J. Mol. Recogn.*, **9**, 728–732 (1996).
386. B. Mattiasson, I.Yu. Galaev, N. Garg, 'Polymer-shielded Dye-affinity Chromatography', *J. Mol. Recognit.*, **9**, 509–514 (1996).
387. N. Garg, I.Yu. Galaev, B. Mattiasson, 'Dye-affinity Techniques for Bioprocessing: Recent Developments', *J. Mol. Recognit.*, **9**, 259–274 (1996).
388. I.Yu. Galaev, P. Arvidsson, B. Mattiasson, 'Protein Displacement in Dye-ligand Chromatography Using Neutral and Charged Polymers', *J. Mol. Recognit.*, **11**, 255–260 (1998).
389. V. Bohácová, P. Docolomanský, A. Breier, P. Gemeiner, A. Ziegelhöffer, 'Interaction of Lactate Dehydrogenase with Anthraquinone Dyes: Characterization of Ligands for Dye-ligand Chromatography', *J. Chromatogr. B*, **715**, 273–281 (1998).
390. J. Kuby, *Immunology*, 3rd edition, Freeman, New York, 1997.
391. H. Leibl, R. Tomasits, H.M. Wolf, M.M. Eibl, J.W. Mannhalter, 'Method for the Isolation of Biologically Active Monomeric Immunoglobulin A from a Plasma Fraction', *J. Chromatogr. B*, **678**, 173–180 (1996).
392. D. Josic, W. Hofmann, R. Habermann, J.D. Schulzke, W. Reutter, 'Isolation of Immunoglobulins and Their use in Immunoaffinity HPLC', *J. Clin. Chem. Clin. Biochem.*, **26**, 559–568 (1988).
393. I. Deinl, L. Angermaier, C. Franzelius, G. Machbert, 'Simple High-performance Liquid-chromatographic

Column-switching Technique for the Online Immunoaffinity Extraction and Analysis of Flunitrazepam and Its Main Metabolites in Urine', *J. Chromatogr. B*, **704**, 251–258 (1997).
394. T. Urano, M.W. Trucksess, S.W. Page, 'Automated Affinity Liquid Chromatography System for Online Isolation, Separation and Quantitation of Aflatoxins in Methanol/Water Extracts of Corn or Peanuts', *J. Agric. Food Chem.*, **41**, 1982–1985 (1993).
395. M.R. Oates, W. Clarke, E.M. Marsh, D.S. Hage, 'Kinetic Studies on the Immobilization of Antibodies to High-performance Liquid Chromatographic Supports', *Bioconjug. Chem.*, **9**, 459–465 (1998).
396. D.S. Hage, 'Affinity Chromatography: a Review of Clinical Applications', *Clin. Chem.*, **45**, 593–615 (1999).
397. E. Ioannou-Kakouri, M. Christodoulidou, E. Christou, E. Constandinidou, 'Immunoaffinity Column-HPLC. Determination of Aflatoxin M1 in Milk', *Food. Agric. Immun.*, **7**, 131–137 (1995).
398. M. Nakajima, H. Tsubouchi, M. Miyabe, Y. Ueno, 'Survey of Aflatoxin B1 and Ochratoxin A in Commercial Green Coffee Beans by High-performance Liquid Chromatography Linked with Immunoaffinity Chromatography', *Food Agric. Immunol.*, **9**, 77–83 (1997).
399. C.M. Maragos, G.A. Bennett, J.L. Richard, 'Affinity-column Clean-up for the Analysis of Fumonisins and Their Hydrolysis Products in Corn', *Food Agric. Immunol.*, **9**, 3–12 (1997).
400. J.M. Francis, D.H. Craston, 'Development of a Stand-alone Affinity Clean-up for Lysergic Acid Diethylamide in Urine', *Analyst*, **121**, 177–182 (1996).
401. X. Wu, K. Haupt, M.A. Vijayalakshmi, 'Separation of Immunoglobulin G by High-performance Pseudo-bioaffinity Chromatography with Immobilized Histidine. I. Preliminary Report on the Influence of the Silica Support and the Coupling Mode', *J. Chromatogr.*, **584**, 35–41 (1992).
402. M.A. Vijayalakshmi, 'Histidine Ligand Affinity Chromatography', *Mol. Biotechnol.*, **6**, 347–357 (1996).
403. J.B. Wheatley, 'Multiple Ligand Applications in High-performance Immunoaffinity Chromatography', *J. Chromatogr.*, **603**, 273–278 (1992).
404. I. Gaillard, A. Martial, H. Pinton, A. Marc, J.M. Engasser, M. Fick, 'Monitoring of Proteins in Hybridoma Cultures by HPLC', *Biotechnol. Tech.*, **8**, 215–220 (1994).
405. E. Pike, P. Kierulf, B. Skuterud, J.E. Bredesen, P.K. Lunde, 'Drug Binding in Sera Deficient in Lipoproteins, Albumin or Orosomucoid', *Br. J. Clin. Pharmacol.*, **16**, 233–239 (1983).
406. B. Sebille, R. Zini, C.V. Madjar, N. Thuaud, J.P. Tillement, 'Separation Procedures Used to Reveal and Follow Drug–Protein Binding', *J. Chromatogr.*, **531**, 51–77 (1990).
407. M. Turowski, R. Kaliszan, 'Keratin Immobilized on Silica as a New Stationary Phase for Chromatographic Modelling of Skin Permeation', *J. Pharm. Biomed. Anal.*, **15**, 1325–1333 (1997).
408. M. Turowski, R. Kaliszan, 'Collagen Immobilised on Silica Derivatives as a new Stationary Phase for HPLC', *Biomed. Chromatogr.*, **12**, 187–192 (1998).
409. B.S. Larsson, 'Interaction Between Chemicals and Melanin', *Pigment Cell Res.*, **6**, 127–33 (1993).
410. A. Radwanska, T. Frackowiak, H. Ibrahim, A.F. Aubrey, R. Kaliszan, 'Chromatographic Modelling of Interactions between Melanin and Phenothiazine and Dibenzazepine Drugs', *Biomed. Chromatogr.*, **9**, 233–237 (1995).
411. R. Knörle, E. Schinz, T.J. Feuerstein, 'Drug Accumulation in Melanin: an Affinity Chromatography Study', *J. Chromatogr. B*, **714**, 171–179 (1998).
412. A. Sengupta, D.S. Hage, 'Characterization of Minor Site Probes for Human Serum Albumin by High-performance Affinity Chromatography', *Anal. Chem.*, **71**, 3821–3827 (1999).
413. G. Brambilla, M. Fiori, I. Curiel, L. Serpe, P. Gallo, 'α1-Acid Glycoprotein Affinity Columns for the Clean-up of Adrenergic Drugs', *Analyst*, **123**, 2693–2696 (1998).
414. P. Lundahl, C.M. Zeng, C. Lagerquist Hägglund, I. Gottschalk, E. Greijer, 'Chromatographic Approaches to Liposomes, Proteoliposomes and Biomembrane Vesicles', *J. Chromatogr. B*, **722**, 103–120 (1999).
415. H. Liu, D.E. Cohen, C. Pidgeon, 'Single Step Purification of Rat Liver Aldolase Using Immobilized Artificial Membrane Chromatography', *J. Chromatogr. B*, **703**, 53–62 (1997).
416. C. Pidgeon, U.V. Venkataram, 'Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids', *Anal. Biochem.*, **176**, 36–47 (1989).
417. G.W. Caldwell, J.A. Masucci, M. Evangelisto, R. White, 'Evaluation of the Immobilized Artificial Membrane Phosphatidylcholine. Drug Discovery Column for High-performance Liquid Chromatographic Screening of Drug–Membrane Interactions', *J. Chromatogr. A*, **800**, 161–169 (1998).
418. E. Krause, M. Dathe, T. Wieprecht, M. Bienert, 'Noncovalent Immobilized Artificial Membrane Chromatography, an Improved Method for Describing Peptide–Lipid Bilayer Interactions', *J. Chromatogr. A*, **849**, 125–133 (1999).
419. N. Labrou, Y.D. Clonis, 'The Affinity Technology in Downstream Processing', *J. Biotechnol.*, **36**, 95–119 (1994).
420. C.R. Lowe, S.J. Burton, N.P. Burton, W.K. Alderton, J.M. Pitts, J.A. Thomas, 'Designer Dyes: 'Biomimetic' Ligands for the Purification of Pharmaceutical Proteins by Affinity Chromatography', *Trends Biotechnol.*, **10**, 442–448 (1992).
421. N.E. Labrou, Y.D. Clonis, (a) 'Biomimetic Dye Affinity Chromatography for the Purification of Bovine Heart Lactate Dehydrogenase', *J. Chromatogr. A*, **718**, 35–44 (1995); (b) 'Biomimetic-dye Affinity Chromatography

for the Purification of Mitochondrial L-Malate Dehydrogenase from Bovine Heart', *J. Biotechnol.*, **45**, 185–194 (1996).
422. N.E. Labrou, E. Eliopoulos, Y.D. Clonis, 'Molecular Modeling for the Design of a Biomimetic Chimeric Ligand. Application to the Purification of Bovine Heart L-Lactate Dehydrogenase', *Biotechnol. Bioeng.*, **63**, 322–332 (1999).
423. R. Li, V. Dowd, D.J. Stewart, S.J. Burton, C.R. Lowe, 'Design, Synthesis, and Application of a Protein A Mimetic', *Nat. Biotechnol.*, **16**, 190–195 (1998).
424. E. Lasonder, W. Bloemhoff, G.W. Welling, 'Interaction of Lysozyme with Synthetic Anti-lysozyme D1.3 Antibody Fragments Studied by Affinity Chromatography and Surface Plasmon Resonance', *J. Chromatogr. A*, **676**, 91–98 (1994).
425. T.W. Hutchens, T.T. Yip, 'Synthetic Metal-binding Protein Surface Domains for Metal Ion-dependent Interaction Chromatography. II. Immobilization of Synthetic Metal-binding Peptides from Metal Ion Transport Proteins as Model Bioactive Protein Surface Domains', *J. Chromatogr.*, **604**, 133–141 (1992).
426. L. Stryer, *Biochemistry*, 3rd edition, Freeman, New York, 298, 345, 1988.
427. Y. Pilatte, G. Rostoker, K. Violleau, M. Basta, C.H. Hammer, 'Further Identification of Human Plasma Glycoproteins Interacting with the Galactose-specific Lectin Jacalin', *J. Chromatogr. B*, **668**, 1–11 (1995).
428. G.L. Hortin, B.L. Trimpe, 'Lectin Affinity Chromatography of Proteins Bearing O-linked Oligosaccharides: Application of Jacalin-agarose', *Anal. Biochem.*, **188**, 271–277 (1990).
429. M.J. Cullina, J.F. Greally, 'Novel Lectin-based Enzyme-linked Immunosorbent Assay for the Measurement of IgA1 in Serum and Secretory IgA1 in Secretions', *Clin. Chim. Acta*, **216**, 23–38 (1993).
430. H. Iwase, A. Tanaka, Y. Hiki, T. Kokubo, I. Ishii-Karakasa, J. Nishikido, Y. Kobayashi, K. Hotta, 'Application of Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry to the Analysis of Glycopeptide-containing Multiple O-Linked Oligosaccharides', *J. Chromatogr. B*, **709**, 145–149 (1998).
431. S.G. De Simone, R. Santos, M.F. Araujo, R.T. Pinho, 'Preparative Isolation of the Lectin Jacalin by Anion-exchange High-performance Liquid Chromatography', *J. Chromatogr. A*, **688**, 357–362 (1994).
432. A. Borchert, P.O. Larsson, K. Mosbach, 'High-performance Liquid Affinity Chromatography on Silica-bound Concanavalin A', *J. Chromatogr.*, **244**, 49–56 (1982).
433. M. Tokuda, M. Kamei, S. Yui, F. Koyama, 'Rapid Resolution of Nucleotide Sugars by Lectin Affinity High-performance Liquid Chromatography', *J. Chromatogr.*, **323**, 434–438 (1985).
434. D.J. Anderson, R.R. Walters, 'Affinity-chromatographic Examination of a Retention Model for Macromolecules', *J. Chromatogr.*, **331**, 1–10 (1985).
435. C. Schaumann, F. Oesch, K.K. Unger, R.J. Wieser, 'Analytical Technique for Studying the Structure of Glycoprotein *N*-Glycans', *J. Chromatogr.*, **646**, 227–234 (1993).
436. A. Apffel, J.A. Chakel, W.S. Hancock, C. Souders, T. M'Timkulu, E. Pungor, Jr, 'Application of Multidimensional Affinity High-performance Liquid Chromatography and Electrospray Ionization Liquid Chromatography/Mass Spectrometry to the Characterization of Glycosylation in Single-chain Plasminogen Activator. Initial Results', *J. Chromatogr. A*, **750**, 35–42 (1996).
437. T. Endo, 'Fractionation of Glycoprotein-derived Oligosaccharides by Affinity Chromatography Using Immobilized Lectin Columns', *J. Chromatogr. A*, **720**, 251–261 (1996).
438. S.N.A. Di Virgilio, 'High Performance Lectin Affinity Chromatography for Fractionation and Sequence Determination of Oligosaccharides', in *Lectins, Biology, Biochemistry, Clinical Biochemistry*, eds. E. van Driessche, S. Beeckmans, T.C. Bøg-Hansen, Textop, Hellerup, Denmark, Vol. 12, 1998. http://plab.ku.dk/tcbh/Lectins12/DiVirgilio/paper.htm
439. R.R. Walters, 'Mini-columns for Affinity Chromatography', *Anal. Chem.*, **55**, 1395–1399 (1983).
440. T. Manabe, N. Higuchi, T. Okuyama, Y. Mukaiyama, 'High-performance Affinity Chromatography of Human Serum Concanavalin A-binding Proteins', *J. Chromatogr.*, **431**, 45–54 (1988).
441. S. Honda, S. Suzuki, K. Kakehi, 'Separation of Carbohydrates and Lectins on Carbohydrate-immobilized Resins', *J. Chromatogr.*, **396**, 93–100 (1987).
442. S. Honda, K. Suzuki, S. Suzuki, K. Kakehi, 'High-performance Affinity Chromatography of Carbohydrate-binding Proteins by Two-stage Separation on a Resin Carrying a number of Oligosaccharides', *Anal. Biochem.*, **169**, 239–245 (1988).
443. K. Kakehi, Y. Kojima, S. Suzuki, S. Honda, 'Effect of Chain Length of Oligosaccharide in High-performance Affinity Chromatography of a Lectin on Oligosaccharide-immobilized Columns', *J. Chromatogr.*, **502**, 297–304 (1990).
444. H. Lakhiari, D. Müller, J. Jozefonvicz, 'Interaction Mechanisms Between Insulin and *N*-Acetylneuraminic Acid in Affinity Chromatography', *J. Chromatogr. A*, **711**, 93–103 (1995).
445. H. Lakhiari, E. Legendre, D. Müller, J. Jozefonvicz, 'High-performance Affinity Chromatography of Insulin on Coated Silica Grafted with Sialic Acid', *J. Chromatogr. B*, **664**, 163–173 (1995).
446. H. Lakhiari, J. Jozefonvicz, D. Müller, 'Influence of the Nature of Coupling Agents on Insulin Adsorption on Supports Grafted with Sialic Acid for High-performance Affinity Chromatography', *J. Chromatogr. B*, **706**, 33–41 (1998).

447. M. Caron, R. Joubert-Caron, J.R. Cartier, A. Chadli, D. Bladier, 'Study of Lectin–Ganglioside Interactions by High-performance Liquid Affinity Chromatography', *J. Chromatogr.*, **646**, 327–333 (1993).
448. A.A. Farooqui, H.C. Yang, L.A. Horrocks, 'Purification of Lipases, Phospholipases and Kinases by Heparin–Sepharose Chromatography', *J. Chromatogr. A*, **673**, 149–158 (1994).
449. M.T.W. Hearn, M. Guthridge, J. Bertolini, 'Separation of Heparin-binding Growth Factors by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **397**, 371–378 (1987).
450. F. De Ceuninck, J. Willeput, M. Corvol, 'Purification and Characterization of Insulin-like Growth Factor II (IGF II) and an IGF II Variant from Human Placenta', *J. Chromatogr. B*, **666**, 203–214 (1995).
451. A. Seeger, U. Rinas, 'Two-step Chromatographic Procedure for Purification of Basic Fibroblast Growth Factor from Recombinant *Escherichia coli* and Characterization of the Equilibrium Parameters of Adsorption', *J. Chromatogr. A*, **746**, 17–24 (1996).
452. P. Blanchardie, P. Lustenberger, J.L. Orsonneau, S. Bernard, 'Resolution of the Molecular Forms of Rat Liver Glucocorticoid Receptor by Affinity Chromatography on Immobilized Heparin', *Biochimie*, **66**, 505–511 (1984).
453. C. Hizel, J.C. Maurizis, P. Rio, Y. Communal, J. Chassagne, D. Favy, Y.J. Bignon, D.J. Bernard-Gallon, 'Isolation, Purification and Quantification of BRCA1 Protein from Tumour Cells by Affinity Perfusion Chromatography', *J. Chromatogr. B*, **721**, 163–170 (1999).
454. C.I. Hammond, M.J. Holland, 'Purification of Yeast RNA Polymerases Using Heparin Agarose Affinity Chromatography. Transcriptional Properties of the Purified Enzymes on Defined Templates', *J. Biol. Chem.*, **258**, 3230–3241(1983).
455. D. Josic, F. Bal, H. Schwinn, 'Isolation of Plasma Proteins from the Clotting Cascade by Heparin Affinity Chromatography', *J. Chromatogr.*, **632**, 1–10 (1993).
456. M. Björklund, M.T.W. Hearn, 'Characterization of Silica-based Heparin Affinity Sorbents from Equilibrium-binding Studies on Plasma Fractions Containing Thrombin', *J. Chromatogr. A*, **762**, 113–133 (1997).
457. L.K. Hennessy, S.T. Kunitake, M. Jarvis, R.L. Hamilton, G. Endeman, A. Protter, J.P. Kane, 'Isolation of Subpopulations of High-density Lipoproteins: Three Particle Species Containing apoE and Two Species Devoid of apoE that have Affinity for Heparin', *J. Lipid Res.*, **38**, 1859–1868 (1997).
458. A.S. Clairbois, D. Letourneur, D. Müller, J. Jozefonvicz, 'High-performance Affinity Chromatography for the Purification of Heparin-binding Proteins from Detergent-solubilized Smooth Muscle Cell Membranes', *J. Chromatogr. B*, **706**, 55–62 (1998).
459. H. Koyama, K. Omura, A. Ejima, Y. Kasanuma, C. Watanabe, H. Satoh, 'Separation of Selenium-containing Proteins in Human and Mouse Plasma Using Tandem High-performance Liquid Chromatography Columns Coupled with Inductively Coupled Plasma Mass Spectrometry', *Anal. Biochem.*, **267**, 84–91 (1999).
460. M. Björklund, M.T.W. Hearn, 'Synthesis of Silica-based Heparin-affinity Adsorbents', *J. Chromatogr. A*, **728**, 149–169 (1996).
461. K. Ishimura, K. Fukunaga, T. Irie, K. Uekama, T. Ohta, H. Nakamura, 'Application of a β-Cyclodextrin Sulfate-immobilized Precolumn to Selective On-line Enrichment and Separation of Heparin-binding Proteins by Column-switching High-performance Liquid Chromatography', *J. Chromatogr. A*, **769**, 209–214 (1997).
462. J. Folkman, P.B. Weisz, M.M. Joullie, W.W. Li, W.R. Ewing, 'Control of Angiogenesis with Synthetic Heparin Substitutes', *Science*, **243**, 1490–1493 (1989).
463. F.B. Anspach, H. Spille, U. Rinas, 'Purification of Recombinant Human Basic Fibroblast Growth-factor – Stability of Selective Sorbents under Cleaning In-place Conditions', *J. Chromatogr. A*, **711**, 129–139 (1995).
464. S. Wera, M. Bollen, W. Stalmans, 'Purification and Characterization of the Glycogen-bound Protein Phosphatase from Rat Liver', *J. Biol. Chem.*, **266**, 339–345 (1991).
465. S. Yu, T.M. Christensen, K.M. Kragh, K. Bojsen, J. Marcussen, 'Efficient Purification, Characterization and Partial Amino Acid Sequencing of two α-1,4-Glucan Lyases from Fungi', *Biochim. Biophys. Acta*, **1339**, 311–320 (1997).
466. N. Juge, J.S. Andersen, D. Tull, P. Roepstorff, B. Svensson, 'Overexpression, Purification, and Characterization of Recombinant Barley α-Amylases 1 and 2 Secreted by the Methylotrophic Yeast *Pichia pastoris*', *Protein Expr. Purif.*, **8**, 204–214 (1996).
467. A. Totsuka, C. Fukazawa, 'Affinity Purification of β-Amylases Originating from Plant Using Cyclomaltohexaose-immobilized Sepharose 6B in the Presence of Ammonium Sulfate', *Protein Expr. Purif.*, **4**, 333–336 (1993).
468. J. Hellman, P. Mantsala, 'Construction of an *Escherichia coli* Export-affinity Vector for Expression and Purification of Foreign Proteins by Fusion to Cyclomaltodextrin Glucanotransferase', *J. Biotechnol.*, **23**, 19–34 (1992).
469. J.T. Kadonaga, 'Purification of Sequence-specific Binding Proteins by DNA Affinity Chromatography', *Methods Enzymol.*, **208**, 10–23 (1991).
470. T. Wada, H. Watanabe, H. Kawaguchi, H. Handa, 'DNA Affinity Chromatography', *Methods Enzymol.*, **254**, 595–604 (1995).
471. E.S. Jenuwine, S.L. Shaner, 'Equilibrium Constants for Nonspecific Binding of Proteins to DNA may be Obtained by Quantitative Zonal DNA Affinity Chromatography', *Anal. Biochem.*, **242**, 228–233 (1996).
472. H.W. Jarrett, 'Affinity Chromatography with Nucleic Acid Polymers', *J. Chromatogr.*, **618**, 315–339 (1993).
473. E. Yashima, T. Shiiba, T. Sawa, N. Miyauchi, M. Akashi, 'High-performance Affinity Chromatography of Oligonucleotides on Nucleic Acid Analogue Immobilized

Silica Gel Columns', *J. Chromatogr.*, **603**, 111–119 (1992).
474. C.P. Holstege, M.L. Pickaart, L.L. Louters, 'Separation of RNA and DNA by Histone Affinity High-performance Liquid Chromatography', *J. Chromatogr.*, **455**, 401–405 (1988).
475. A.W. Czarnik, 'Combinatorial Chemistry. What's in it for Analytical Chemists?', *Anal. Chem.*, **70**, 378A–386A (1998).
476. L. Van Hijfte, G. Marciniak, N. Froloff, 'Combinatorial Chemistry, Automation and Molecular Diversity: New Trends in the Pharmaceutical Industry', *J. Chromatogr. B*, **725**, 3–15 (1999).
477. C.P. Woodbury, Jr, D.L. Venton, 'Methods of Screening Combinatorial Libraries Using Immobilized or Restrained Receptors', *J. Chromatogr. B*, **725**, 113–137 (1999).
478. A.D. Ellington, J. Szostak, 'In Vitro Selection of RNA Molecules that Bind Specific Ligands', *Nature*, **346**, 818–822 (1990).
479. S.D. Jayasena, 'Aptamers: an Emerging Class of Molecules that Rival Antibodies in Diagnostics', *Clin. Chem.*, **45**, 1628–1650 (1999).
480. C. Tuerk, L. Gold, 'Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase', *Science*, **249**, 505–150 (1990).
481. T.S. Romig, C. Bell, D.W. Drolet, 'Aptamer Affinity Chromatography: Combinatorial Chemistry Applied to Protein Purification', *J. Chromatogr. B*, **731**, 275–284 (1999).
482. R.A. Potyrailo, R.C. Conrad, A.D. Ellington, G.M. Hieftje, 'Adapting Selected Nucleic Acid Ligands (Aptamers) to Biosensors', *Anal. Chem.*, **70**, 3419–3425 (1998).
483. L.C. Bock, L.C. Griffin, J.A. Latham, E.H. Vermaas, J.J. Toole, 'Selection of Single-stranded DNA Molecules that Bind and Inhibit Human Thrombin', *Nature*, **355**, 564–566 (1992).
484. P. Bridonneau, Y.F. Chang, A.V. Buvoli, D. O'Connell, D. Parma, 'Site-directed Selection of Oligonucleotide Antagonists by Competitive Elution', *Antisense Nucleic Acid Drug Dev.*, **9**, 1–11 (1999).
485. J. Porath, J. Carlsson, I. Olsson, G. Belfrage, 'Metal Chelate Affinity Chromatography, a New Approach to Protein Fractionating', *Nature*, **258**, 598–99 (1975).
486. A. Figueroa, C. Corradini, B. Feibush, B.L. Karger, 'High-performance Immobilized-metal Affinity Chromatography of Proteins on Iminodiacetic Acid Silica-based Bonded Phases', *J. Chromatogr.*, **371**, 335–352 (1986).
487. E.S. Hemdan, Y.J. Zhao, E. Sulkowski, J. Porath, 'Surface Topography of Histidine Residues: a Facile Probe by Immobilized Metal Ion Affinity Chromatography', *Proc. Natl. Acad. Sci. USA*, **86**, 1811–1815 (1989).
488. F. Herve, F. Fouache, C. Marche, J.P. Tillement, 'Abnormal Microheterogeneity Detected in one Commercial α1-Acid Glycoprotein Preparation Using Chromatography on Immobilized Metal Affinity Adsorbent and on Hydroxyapatite', *J. Chromatogr. B*, **688**, 35–46 (1997).
489. K.Y. Jiang, O. Pitiot, M. Anissimova, H. Adenier, M.A. Vijayalakshmi, 'Structure–Function Relationship in Glycosylated α-Chymotrypsin as Probed by IMAC and IMACE', *Biochim. Biophys. Acta*, **1433**, 198–209 (1999).
490. H. Wu, D.F. Bruley, 'Homologous Human Blood Protein Separation Using Immobilized Metal Affinity Chromatography: Protein C Separation from Prothrombin with Application to the Separation of Factor IX and Prothrombin', *Biotechnol. Prog.*, **15**, 928–931 (1999).
491. M. Belew, T.T. Yip, L. Andersson, J. Porath, 'Interaction of Proteins with Immobilized Cu^{2+}. Quantitation of Adsorption Capacity, Adsorption Isotherms and Equilibrium Constants by Frontal Analysis', *J. Chromatogr.*, **403**, 197–206 (1987).
492. T.W. Hutchens, T.T. Yip, J. Porath, 'Protein Interaction with Immobilized Ligands: Quantitative Analyses of Equilibrium Partition Data and Comparison with Analytical Chromatographic Approaches Using Immobilized Metal Affinity Adsorbents', *Anal. Biochem.*, **170**, 168–182 (1988).
493. T.W. Hutchens, T.T. Yip, 'Protein Interactions with Immobilized Transition Metal Ions: Quantitative Evaluations of Variations in Affinity and Binding Capacity', *Anal. Biochem.*, **191**, 160–168 (1990).
494. T.W. Hutchens, T.T. Yip, 'Differential Interaction of Peptides and Protein Surface Structures with Free Metal Ions and Surface-immobilized Metal Ions', *J. Chromatogr.*, **500**, 531–542 (1990).
495. M. Belew, J. Porath, 'Immobilized Metal Ion Affinity Chromatography. Effect of Solute Structure, Ligand Density and Salt Concentration on the Retention of Peptides', *J. Chromatogr.*, **516**, 333–354 (1990).
496. T.W. Hutchens, R.W. Nelson, C.M. Li, T.T. Yip, 'Synthetic Metal-binding Protein Surface Domains for Metal ion-dependent Interaction Chromatography. I. Analysis of Bound Metal Ions by Matrix-assisted UV Laser Desorption Time-of-flight Mass Spectrometry', *J. Chromatogr.*, **604**, 125–132 (1992).
497. M. Zachariou, M.T. Hearn, 'Application of Immobilized Metal Ion Chelate Complexes as Pseudocation Exchange Adsorbents for Protein Separation', *Biochemistry*, **35**, 202–211 (1996).
498. A.J. Fatiadi, 'Affinity Chromatography and Metal Chelate Affinity Chromatography', *CRC Crit. Rev. Anal. Chem.*, **18**, 1–44 (1987).
499. J. Porath, 'High-performance Immobilized-metal-ion Affinity Chromatography of Peptides and Proteins', *J. Chromatogr.*, **443**, 3–11 (1988).
500. J. Porath, 'IMAC: Immobilized-metal-ion Affinity-based Chromatography', *Trends Anal. Chem.*, **7**, 254–259 (1988).

501. J.W. Wong, R.L. Albright, N.H.L. Wang, 'Immobilized-metal-ion Affinity Chromatography (IMAC)–Chemistry and Bioseparation Applications', *Sep. Purif. Methods*, **20**, 49–106 (1991).
502. J. Porath, 'Immobilized Metal Ion Affinity Chromatography', *Protein Expr. Purif.*, **3**, 263–281 (1992).
503. T.T. Yip, T.W. Hutchens, 'Immobilized Metal Ion Affinity Chromatography', *Mol. Biotechnol.*, **1**, 151–164 (1994).
504. T.T. Yip, T.W. Hutchens, 'Immobilized Metal Ion Affinity Chromatography', *Methods Mol. Biol.*, **59**, 197–210 (1996).
505. L. Kågedal, 'Immobilized Metal Ion Affinity Chromatography', in *Protein Purification. Principles, High Resolution Methods, and Applications*, 2nd edition, eds. J.C. Janson, L. Rydén, Wiley-Liss, New York, 311–342, 1998.
506. T.T. Yip, T.W. Hutchens, 'Immobilized Metal Affinity Chromatography', in *Practical Protein Chromatography*, eds. A. Kenney, S. Fowell, (Meth. Mol. Biol, Vol. 11), Humana Press, Totowa, 17–31, 1992.
507. M. Belew, T.T. Yip, L. Andersson, R. Ehrnström, 'High-performance Analytical Applications of Immobilized Metal Ion Affinity Chromatography', *Anal. Biochem.*, **164**, 457–465 (1987).
508. M. Zachariou, M.T. Hearn, 'Protein Selectivity in Immobilized Metal Affinity Chromatography Based on the Surface Accessibility of Aspartic and Glutamic Acid Residues', *J. Protein Chem.*, **14**, 419–430 (1995).
509. L. Andersson, E. Sulkowski, 'Evaluation of the Interaction of Protein α-Amino Groups with M(II) by Immobilized Metal Ion Affinity Chromatography', *J. Chromatogr.*, **604**, 13–17 (1992).
510. P. Hansen, G. Lindeberg, L. Andersson, 'Immobilized Metal Ion Affinity Chromatography of Synthetic Peptides. Binding via the α-Amino Group', *J. Chromatogr.*, **627**, 125–135 (1992).
511. P. Hansen, G. Lindeberg, 'Importance of the α-Amino Group in the Selective Purification of Synthetic Histidine Peptides by Immobilised Metal Ion Affinity Chromatography', *J. Chromatogr. A*, **690**, 155–159 (1995).
512. P. Hansen, L. Andersson, G. Lindeberg, 'Purification of Cysteine-containing Synthetic Peptides via Selective Binding of the α-Amino Group to Immobilized Copper(II) and Nickel(II) Ions', *J. Chromatogr. A*, **723**, 51–59 (1996).
513. J. Porath, B. Olin, B. Granstrand, 'Immobilized-metal Affinity Chromatography of Serum Proteins on Gel-immobilized Group III A Metal Ions', *Arch. Biochem. Biophys.*, **225**, 543–547 (1983).
514. N. Ramadan, J. Porath, 'Fe^{3+}-Hydroxamate as Immobilized Metal Affinity-adsorbent for Protein Chromatography', *J. Chromatogr.*, **321**, 93–104 (1985).
515. E. Hochuli, H. Döbeli, A. Schacher, 'New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighbouring Histidine Residues', *J. Chromatogr.*, **411**, 177–184 (1987).
516. Z. El Rassi, Cs. Horváth, 'Metal Chelate-interaction Chromatography of Proteins with Iminodiacetic Acid-bonded Stationary Phases on Silica Support', *J. Chromatogr.*, **359**, 241–253 (1986).
517. R.J. Todd, R.D. Johnson, F.H. Arnold, 'Multiple-site Binding Interactions in Metal-affinity Chromatography. I. Equilibrium Binding of Engineered Histidine-containing Cytochrome *c*', *J. Chromatogr. A*, **662**, 13–26 (1994).
518. R.D. Johnson, R.J. Todd, F.H. Arnold, 'Multipoint Binding in Metal-affinity Chromatography II. Effect of pH and Imidazole on Chromatographic Retention of Engineered Histidine-containing Cytochrome *c*', *J. Chromatogr. A*, **725**, 225–235 (1996).
519. M.C. Millot, F. Herve, B. Sebille, 'Retention Behaviour of Proteins on Poly(vinylimidazole)-Copper(II) Complexes Supported on Silica: Application to the Fractionation of Desialylated Human α1-Acid Glycoprotein Variants', *J. Chromatogr. B*, **664**, 55–67 (1995).
520. J. Xiao, M.E. Meyerhoff, 'Retention Behavior of Amino Acids and Peptides on Protoporphyrin-silica Stationary Phases with Varying Metal Ion Centers', *Anal. Chem.*, **68**, 2818–2825 (1996).
521. A. Gambero, L.T. Kubota, Y. Gushikem, C. Airoldi, J.M. Granjeiro, E.M. Taga, E.F.C. Alcántara, 'Use of Chemically Modified Silica with β-Diketoamine Groups for Separation of α-Lactoalbumin from Bovine Milk Whey by Affinity Chromatography', *J. Colloid Interface Sci.*, **185**, 313–316 (1997).
522. M. Zachariou, I. Traverso, M.T. Hearn, '*O*-Phosphoserine as a new Chelating Ligand for Use with Hard Lewis Metal Ions in the Immobilized-metal Affinity Chromatography of Proteins', *J. Chromatogr.*, **646**, 107–20 (1993).
523. W. Jiang, B. Graham, L. Spiccia, M.T.W. Hearn, 'Protein Selectivity with Immobilized Metal Ion-tacn Sorbents: Chromatographic Studies with Human Serum Proteins and Several Other Globular Proteins', *Anal. Biochem.*, **255**, 47–58 (1998).
524. H. Chaouk, M.T. Hearn, 'Examination of the Protein Binding Behaviour of Immobilised Copper(II)-2,6-diaminomethylpyridine and Its Application in the Immobilised Metal Ion Affinity Chromatographic Separation of Several Human Serum Proteins', *J. Biochem. Biophys. Methods*, **39**, 161–177 (1999).
525. H. Chaouk, M.T. Hearn, 'New Ligand, *N*-(2-Pyridylmethyl)aminoacetate, for Use in the Immobilised Metal Ion Affinity Chromatographic Separation of Proteins', *J. Chromatogr. A*, **852**, 105–115 (1999).
526. W. Jiang, M.T. Hearn, 'Protein Interaction with Immobilized Metal Ion Affinity Ligands Under High Ionic Strength Conditions', *Anal. Biochem.*, **242**, 45–54 (1996).
527. Y. Nakagawa, T.T. Yip, M. Belew, J. Porath, 'High-performance Immobilized-metal-ion Affinity Chromatography of Peptides: Analytical Separation of Biologically

Active Synthetic Peptides', *Anal. Biochem.*, **168**, 75–81 (1988).
528. G. Muszynska, G. Dobrowolska, A. Medin, P. Ekman, J.O. Porath, 'Model Studies on Iron(III) Ion Affinity Chromatography. II. Interaction of Immobilized Iron(III) Ions with Phosphorylated Amino Acids, Peptides and Proteins', *J. Chromatogr.*, **604**, 19–28 (1992).
529. L. Andersson, J. Porath, 'Isolation of Phosphoproteins by Immobilized Metal (Fe^{3+}) Affinity Chromatography', *Anal. Biochem.*, **154**, 250–254 (1986).
530. G. Muszynska, L. Andersson, J. Porath, 'Selective Adsorption of Phosphoproteins on Gel-immobilized Ferric Chelate', *Biochemistry*, **25**, 6850–6853 (1986).
531. P. Cao, J.T. Stults, 'Phosphopeptide Analysis by Online Immobilized Metal-ion Affinity Chromatography–Capillary Electrophoresis–Electrospray Ionization Mass Spectrometry', *J. Chromatogr. A*, **853**, 225–235 (1999).
532. T.T. Yip, Y. Nakagawa, J. Porath, 'Evaluation of the Interaction of Peptides with Cu(II), Ni(II), and Zn(II) by High-performance Immobilized Metal Ion Affinity Chromatography', *Anal. Biochem.*, **183**, 159–171 (1989).
533. M.C. Smith, T.C. Furman, T.D. Ingolia, C. Pidgeon, 'Chelating Peptide-immobilized Metal Ion Affinity Chromatography. A New Concept in Affinity Chromatography for Recombinant Proteins', *J. Biol. Chem.*, **263**, 7211–7215 (1988).
534. J. Crowe, H. Döbeli, R. Gentz, E. Hochuli, D. Stuber, K. Henco, '6xHis-Ni-NTA Chromatography as a Superior Technique in Recombinant Protein Expression/Purification', *Methods Mol. Biol.*, **31**, 371–387 (1994).
535. A.V. Patwardhan, G.N. Goud, R.R. Koepsel, M.M. Ataai, 'Selection of Optimum Affinity Tags from a Phage-displayed Peptide Library. Application to Immobilized Copper(II) Affinity Chromatography', *J. Chromatogr. A*, **787**, 91–100 (1997).
536. V. Gaberc-Porekar, V. Menart, S. Jevsevar, A. Vidensek, A. Stalc, 'Histidines in Affinity Tags and Surface Clusters for Immobilized Metal-ion Affinity Chromatography of Trimeric Tumor Necrosis Factor α', *J. Chromatogr. A*, **852**, 117–128 (1999).
537. V.V. Kronina, H.J. Wirth, M.T. Hearn, 'Characterisation by Immobilised Metal Ion Affinity Chromatographic Procedures of the Binding Behaviour of Several Synthetic Peptides Designed to have High Affinity for Cu(II) Ions', *J. Chromatogr. A*, **852**, 261–272 (1999).
538. C. Min, G.L. Verdine, 'Immobilized Metal Affinity Chromatography of DNA', *Nucleic Acids Res.*, **24**, 3806–3810 (1996).
539. P. Lötscher, L. Mottlau, E. Hochuli, 'Immobilization of Monoclonal Antibodies for Affinity Chromatography Using a Chelating Peptide', *J. Chromatogr.*, **595**, 113–119 (1992).
540. K.M. Müller, K.M. Arndt, K. Bauer, A. Pluckthun, 'Tandem Immobilized Metal-ion Affinity Chromatography/Immunoaffinity Purification of His-tagged Proteins – Evaluation of two Anti-His-tag Monoclonal Antibodies', *Anal. Biochem.*, **259**, 54–61 (1998).
541. Z. El Rassi, Y. Truei, Y.F. Maa, Cs. Horváth, 'High-performance Liquid Chromatography with Concanavalin A Immobilized by Metal Interactions on the Stationary Phase', *Anal. Biochem.*, **169**, 172–180 (1988).
542. P. Allegra, R. Sterner, D.F. Clayton, V.G. Allfrey, 'Affinity Chromatographic Purification of Nucleosomes Containing Transcriptionally Active DNA Sequences', *J. Mol. Biol.*, **196**, 379–388 (1987).
543. J. Porath, F. Maisano, M. Belew, 'Thiophilic Adsorption–a New Method for Protein Fractionation', *FEBS Lett.*, **185**, 306–310 (1985).
544. T.W. Hutchens, J. Porath, 'Thiophilic Adsorption of Immunoglobulins – Analysis of Conditions Optimal for Selective Immobilization and Purification', *Anal. Biochem.*, **159**, 217–26 (1986).
545. T.W. Hutchens, 'Thiophilic Adsorption Chromatography', in *Practical Protein Chromatography*, eds. A. Kenney, S. Fowell, (Meth. Mol. Biol, Vol. 11), Humana Press, Totowa, 1–15, 1992.
546. A. Schwarz, F. Kohen, M. Wilchek, 'Novel Heterocyclic Ligands for the Thiophilic Purification of Antibodies', *J. Chromatogr. B*, **664**, 83–88 (1995).
547. U.B. Finger, J. Thommes, D. Kinzelt, M.R. Kula, 'Application of Thiophilic Membranes for the Purification of Monoclonal Antibodies from Cell Culture Media', *J. Chromatogr. B*, **664**, 69–78 (1995).
548. F.T. Kreutz, D.S. Wishart, M.R. Suresh, 'Efficient Bispecific Monoclonal Antibody Purification Using Gradient Thiophilic Affinity Chromatography', *J. Chromatogr. B*, **714**, 161–170 (1998).
549. S. Oscarsson, G. Chaga, J. Porath, 'Thiophilic Adsorbents for RIA and ELISA Procedures', *J. Immunol. Meth.*, **143**, 143–149 (1991).
550. D.A. Palmer, M. Evans, J.N. Miller, 'Thiophilic Gels: an Alternative to Protein A and G for Use in Flow-injection Immunoassays', *Anal. Proc.*, **31**, 123–125 (1994).
551. D.A. Palmer, J.N. Miller, 'Thiophilic Gels: Applications in Flow-injection Immunoassays for Macromolecules and Haptens', *Anal. Chim. Acta*, **303**, 223–230 (1995).
552. D.A. Palmer, M. Evans, J.N. Miller, 'Flow-injection Immunoassay for Albumin Using Thiophilic Gels', *Anal. Biochem.*, **222**, 281–283 (1994).
553. C. Pazos, L. Franco-Fraguas, F. Batista-Viera, 'Salt Promoted Adsorption Chromatography of Malted Barley Amylases', *Chromatographia*, **38**, 232–234 (1994).
554. C. Pazos, L. Franco-Fraguas, F. Batista-Viera, 'Wheat β-Amylase Behaviour Regarding Salt-promoted Adsorption Processes', *Chromatographia*, **48**, 209–214 (1998).
555. H. Goubran-Botros, J. Rabillon, C. Gregoire, B. David, J.P. Dandeu, 'Thiophilic Adsorption Chromatography: Purification of Equ c2 and Equ c3, two Horse Allergens from Horse Sweat', *J. Chromatogr. B*, **710**, 57–65 (1998).

556. R.J. Noel, W.T. O'Hare, G. Street, 'Thiophilic Nature of Divinylsulfone Crosslinked Agarose', *J. Chromatogr. A*, **734**, 241–246 (1996).
557. G.H. Scholz, P. Wippich, S. Leistner, K. Huse, 'Salt-independent Binding of Antibodies from Human Serum to Thiophilic Heterocyclic Ligands', *J. Chromatogr. B*, **709**, 189–196 (1998).
558. P.P. Berna, N. Berna, J. Porath, S. Oscarsson, 'Comparison of the Protein Adsorption Selectivity of Salt-promoted Agarose-based Adsorbents. Hydrophobic, Thiophilic and Electron Donor–Acceptor Adsorbents', *J. Chromatogr. A*, **800**, 151–159 (1998).
559. P.P. Berna, J. Porath, 'Cyanocarbons as Ligands for Electron Donor Acceptor Chromatography of Human Serum Proteins', *J. Chromatogr. A*, **753**, 57–62 (1996).
560. P.P. Berna, J. Porath, 'Salt-independent Adsorption of Human Serum Proteins on Cyanocarbon Gels', *J. Chromatogr. B*, **693**, 277–285 (1997).
561. G.H. Scholz, S. Vieweg, S. Leistner, J. Seissler, W.A. Scherbaum, K. Huse, 'A Simplified Procedure for the Isolation of Immunoglobulins from Human Serum Using a Novel Type of Thiophilic Gel at Low Salt Concentration', *J. Immunol. Meth.*, **219**, 109–118 (1998).
562. D.S. Hage, R.R. Walters, 'Dual-column Determination of Albumin and Immunoglobulin G in Serum by High-performance Affinity Chromatography', *J. Chromatogr.*, **386**, 37–49 (1987).
563. D. Corradini, Z. El Rassi, Cs. Hórvath, G. Guerra, W. Horne, 'Combined Lectin-affinity and Metal-interaction Chromatography for the Separation of Glycophorins by High-performance Liquid Chromatography', *J. Chromatogr.*, **458**, 1–11 (1988).
564. J. Porath, P. Hansen, 'Cascade-mode Multiaffinity Chromatography. Fractionation of Human Serum Proteins', *J. Chromatogr.*, **550**, 751–764 (1991).
565. E. Karlsson, L. Rydén, J. Brewer, 'Ion-exchange Chromatography', in *Protein Purification. Principles, High Resolution Methods, and Applications*, 2nd edition, eds. J.C. Janson, L. Rydén, Wiley-Liss, New York, 145–205, 1998.
566. S. Yamamoto, K. Nakanishi, R. Matsuno, *Ion Exchange Chromatography of Proteins*, (Chromatogr. Sci., Vol. 43), Marcel Dekker, New York, 1988.
567. Y.C. Lee, 'Carbohydrate Analyses with High-performance Anion-exchange Chromatography', *J. Chromatogr. A*, **720**, 137–149 (1996).
568. G. Choudhary, Cs. Horváth, 'Ion-exchange Chromatography', *Methods Enzymol.*, **270**, 47–82 (1996).
569. D.L. Crimmins, 'Strong Cation-exchange High-performance Liquid Chromatography as a Versatile Tool for the Characterization and Purification of Peptides', *Anal. Chim. Acta*, **352**, 21–30 (1997).
570. S.C. Goheen, J.L. Hilsenbeck, 'High-performance Ion-exchange Chromatography and Adsorption of Plasma Proteins', *J. Chromatogr. A*, **816**, 89–96 (1998).
571. W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, 'Retention Model for High-performance Ion-exchange Chromatography', *J. Chromatogr.*, **266**, 3–21 (1983).
572. W. Kopaciewicz, F.E. Regnier, 'Mobile Phase Selection for the High-performance Ion-exchange Chromatography of Proteins', *Anal. Biochem.*, **133**, 251–259 (1983).
573. J. Ståhlberg, B. Jönsson, Cs. Horváth, 'Theory for Electrostatic Interaction Chromatography of Proteins', *Anal. Chem.*, **63**, 1867–1874 (1991).
574. N.K. Boardman, S.M. Partridge, 'Separation of Neutral Proteins on Ion-exchange Resins', *Biochem. J.*, **59**, 543–552 (1955).
575. M. Weitzhandler, D. Farnan, J. Horvath, J.S. Rohrer, R.W. Slingsby, N. Avdalovic, C. Pohl, 'Protein Variant Separations by Cation-exchange Chromatography on Tentacle-type Polymeric Stationary Phases', *J. Chromatogr. A*, **828**, 365–372 (1998).
576. Swiss Institute of Bioinformatics, Genève, Switzerland. *ExPASy Molecular Biology Server* (http://www.expasy.ch/).
577. J. Ståhlberg, B. Jönsson, 'Influence of Charge Regulation in Electrostatic Interaction Chromatography of Proteins', *Anal. Chem.*, **68**, 1536–1544 (1996).
578. W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, 'Evaluation of a Retention Model for High-performance Ion-exchange Chromatography Using two Different Displacing Salts', *J. Chromatogr.*, **283**, 37–45 (1984).
579. W.R. Melander, Z. El Rassi, Cs. Horváth, 'Interplay of Hydrophobic and Electrostatic Interactions in Biopolymer Chromatography. Effect of Salts on the Retention of Proteins', *J. Chromatogr. A*, **469**, 3–27 (1989).
580. J. Ståhlberg, B. Jönsson, Cs. Horváth, 'Combined Effect of Coulombic and van der Waals Interactions in the Chromatography of Proteins', *Anal. Chem.*, **64**, 3118–3124 (1992).
581. S. Yamamoto, T. Ishihara, 'Ion-exchange Chromatography of Proteins near the Isoelectric Points', *J. Chromatogr. A*, **852**, 31–36 (1999).
582. C.M. Roth, K.K. Unger, A.M. Lenhoff, 'Mechanistic Model of Retention in Protein Ion-exchange Chromatography', *J. Chromatogr. A*, **726**, 45–56 (1996).
583. J. Ståhlberg, 'Retention Models for Ions in Chromatography', *J. Chromatogr. A*, **855**, 3–55 (1999).
584. F.W. Fang, M.I. Aguilar, M.T.W. Hearn, 'Influence of Temperature on the Retention Behaviour of Proteins in Cation-exchange Chromatography', *J. Chromatogr. A*, **729**, 49–66 (1996).
585. W.S. Xu, F.E. Regnier, 'Protein–Protein Interactions on Weak-cation-exchange Sorbent Surfaces During Chromatographic Separations', *J. Chromatogr. A*, **828**, 357–364 (1998).
586. C.K. Ratnayake, F.E. Regnier, 'Study of Protein Binding to a Silica Support with a Polymeric Cation-exchange Coating', *J. Chromatogr. A*, **743**, 15–23 (1996).

587. M.A. Hashim, K.H. Chu, P.S. Tsan, 'Determination of Lumped Rate Coefficients of Proteins: Effect of Adsorption', *Chem. Eng. Technol.*, **19**, 137–142 (1996).
588. M.T.W. Hearn, A.N. Hodder, F.W. Fang, M.I. Aguilar, 'Retention Behavior of Proteins with Macroporous Tentacle-type Anion-exchangers', *J. Chromatogr.*, **548**, 117–126 (1991).
589. F.W. Fang, M.I. Aguilar, M.T.W. Hearn, 'Evaluation of Bandwidth Behavior of Proteins Chromatographed on Tentacle-type Anion-exchangers', *J. Chromatogr.*, **599**, 163–170 (1992).
590. J.R. Xie, M.I. Aguilar, M.T.W. Hearn, 'Studies on the Adsorption Capacities of Proteins with a Tentacle-type Ion-exchanger and their Relationship to the Stoichiometric Retention Parameter Z', *J. Chromatogr. A*, **711**, 43–52 (1995).
591. F.W. Fang, M.I. Aguilar, M.T.W. Hearn, 'Temperature-induced Changes in the Bandwidth Behaviour of Proteins Separated with Cation-exchange Adsorbents', *J. Chromatogr. A*, **729**, 67–79 (1996).
592. M. Rögner, 'Purification of Membrane–Protein Complexes Isolated from a Cyanobacterial Thylakoid Membrane by High-performance Liquid-chromatography', *J. Chromatogr.*, **512**, 219–226 (1990).
593. B. Fricke, T. Buchmann, S. Friebe, 'Unusual Chromatographic Behavior and One-step Purification of a Novel Membrane Proteinase from *Bacillus cereus*', *J. Chromatogr. A*, **715**, 247–258 (1995).
594. J. Cacia, C.P. Quan, M. Vasser, M.B. Sliwkowski, J. Frenz, 'Protein Sorting by High-performance Liquid Chromatography. I. Biomimetic Interaction Chromatography of Recombinant Human Deoxyribonuclease I on Polyionic Stationary Phases', *J. Chromatogr.*, **634**, 229–239 (1993).
595. D.M.F. Prazeres, T. Schluep, C. Cooney, 'Preparative Purification of Supercoiled Plasmid DNA Using Anion-exchange Chromatography', *J. Chromatogr. A*, **806**, 31–45 (1998).
596. R.R. Townsend, M.R. Hardy, 'Analysis of Glycoprotein Oligosaccharides Using High-pH Anion-exchange Chromatography', *Glycobiology*, **1**, 139–147 (1991).
597. Z. El Rassi, C. Horváth, 'Tandem Columns and Mixed-bed Columns in High-performance Liquid Chromatography of Proteins', *J. Chromatogr.*, **359**, 255–264 (1986).
598. Y.F. Maa, F.D. Antia, Z. El Rassi, Cs. Horváth, 'Mixed-bed Ion-exchange Columns for Protein High-performance Liquid Chromatography', *J. Chromatogr.*, **452**, 331–345 (1988).
599. P. Tauc, S. Cochet, E. Algiman, I. Callebaut, J.P. Cartron, J.C. Brochon, O. Bertrand, 'Ion-exchange Chromatography of Proteins: Modulation of Selectivity by Addition of Organic Solvents to Mobile Phase. Application to Single-step Purification of a Proteinase Inhibitor from Corn and Study of the Mechanism of Selectivity Modulation', *J. Chromatogr. A*, **825**, 17–27 (1998).
600. M.T.W. Hearn, A.N. Hodder, M.I. Aguilar, 'Influence of Different Displacer Salts on the Retention and Bandwidth Properties of Proteins Separated by Isocratic Anion-exchange Chromatography', *J. Chromatogr.*, **443**, 97–118 (1988).
601. G. Malmquist, N. Lundell, 'Characterization of the Influence of Displacing Salts on Retention in Gradient Elution Ion-exchange Chromatography of Proteins and Peptides', *J. Chromatogr.*, **627**, 107–124 (1992).
602. A. Tiselius, 'Chromatography of Proteins on Calcium Phosphate Columns', *Arkiv Kemi*, **7**, 443–449 (1954).
603. A. Tiselius, S. Hjertén, Ö. Levin, 'Protein Chromatography on Calcium Phosphate Columns', *Arch. Biochem. Biophys.*, **65**, 132–155 (1956).
604. T. Kawasaki, M. Niikura, Y. Kobayashi, 'Fundamental Study of Hydroxyapatite High-performance Liquid Chromatography. III. Direct Experimental Confirmation of the Existence of two Types of Adsorbing Surface on the Hydroxyapatite Crystal', *J. Chromatogr.*, **515**, 125–148 (1990).
605. B. Pavlu, 'HPLC Purification of Monoclonal Antibodies', in *HPLC of Proteins, Peptides, and Polynucleotides: Contemporary Topics and Applications*, ed. M.T.W. Hearn, VCH, New York/Weinheim, 599–623, 1991.
606. T. Kawasaki, M. Niikura, Y. Kobayashi, 'Fundamental Study of Hydroxyapatite High-performance Liquid Chromatography. II. Experimental Analysis on the Basis of the General Theory of Gradient Chromatography', *J. Chromatogr.*, **515**, 91–123 (1990).
607. K. Makino, B.M. Kim, 'Separation of Proteins on Hydroxyapatite Columns', *J. Chromatogr. Sci.*, **27**, 659–664 (1989).
608. S. Hjertén, J. Lindeberg, B. Shopova, 'High-performance Adsorption Chromatography of Proteins on 2-μm Spherical Beads of Hydroxyapatite', *J. Chromatogr.*, **440**, 305–313 (1988).
609. M. Kastner, D. Neubert, 'Isolation of Cytochrome P-450 Components from Marmoset Liver Microsomes by High-performance Liquid Chromatography', *J. Chromatogr.*, **587**, 117–126 (1991).
610. A. Zagariya, S. Khrapunov, W. Zacharias, 'Rapid Method for the Fractionation of Nuclear Proteins and their Complexes by Batch Elution from Hydroxyapatite', *J. Chromatogr.*, **648**, 275–278 (1993).
611. Y. Shibusawa, 'Lipoproteins: Comparison of Different Separation Strategies', *J. Chromatogr. B*, **699**, 419–437 (1997).
612. T. Ichimura, N. Ikuta, Y. Uda, T. Horigome, S. Omata, 'Separation of Membrane Proteins Solubilized with a Nondenaturing Detergent and a High Salt Concentration by Hydroxyapatite High-performance Liquid Chromatography', *Anal. Biochem.*, **224**, 250–255 (1995).
613. Y. Yamakawa, K. Miyasaka, T. Ishikawa, Y. Yamada, T. Okuyama, 'High-performance Liquid Chromatography of Transfer Ribonucleic Acids on Spherical

Hydroxyapatite Beads. II. Effects of pH and Sodium Chloride on Chromatography', *J. Chromatogr.*, **515**, 563–572 (1990).
614. S. Hori, S. Ohtani, K. Miyasaka, T. Ishikawa, H. Tanabe, 'Separation of High-molecular Mass RNA by High-performance Liquid Chromatography on Hydroxyapatite', *J. Chromatogr.*, **515**, 611–619 (1990).
615. H. Narita, Y. Takeda, K. Takagaki, T. Nakamura, S. Harata, M. Endo, 'Identification of Glycosaminoglycans Using High-performance Liquid Chromatography on a Hydroxyapatite Column', *Anal. Biochem.*, **232**, 133–136 (1995).
616. T. Kadoya, 'High-performance Liquid Chromatography of Proteins on a Ceramic Hydroxyapatite with Volatile Buffers', *J. Chromatogr.*, **515**, 521–525 (1990).
617. G.W. Boers, J.J. Kettenes Van den Bosch, A. Bult, 'Applications of HY-Apatite in Liquid Chromatography', *Anal. Chim. Acta*, **279**, 89–94 (1993).
618. G. Bruno, F. Gasparrini, D. Misiti, E. Arrigoni-Martelli, M. Bronzetti, 'High-performance Liquid-chromatographic Separation of Biomolecules Using Calcium Phosphate Supported on Macroporous Silica Microparticles', *J. Chromatogr.*, **504**, 319–333 (1990).
619. T. Sato, T.S. Su, T. Ogawa, T. Okayama, 'Ceramic Fluoroapatite Beads for Chromatography of Proteins and Amino-acids', *Chromatography*, **19**, 191–199 (1998).
620. A. Benmoussa, M. Mikou, J.L. Lacout, A.M. Siouffi, 'Lead Phosphate Hydroxyapatite High-performance Liquid Chromatography', *J. Chromatogr. A*, **694**, 486–491 (1995).
621. A. Benmoussa, J.L. Lacout, P.R. Loiseau, M. Mikou, 'Calcium Phosphate Hydroxyapatite and Lead Phosphate Hydroxyapatite for Separation by Liquid Chromatography', *Chromatographia*, **42**, 177–180 (1996).
622. S. Inoue, N. Ohtaki, 'Metaphosphates as Packing Materials for Biochromatography', *J. Chromatogr. A*, **679**, 247–253 (1994).
623. S. Hjertén, I. Zelikman, J. Lindeberg, J.L. Liao, K.O. Eriksson, J. Mohammad, 'High-performance Adsorption Chromatography of Proteins on Deformed Non-porous Agarose Beads Coated with Insoluble Metal Compounds. I. Coatings: Ferric Oxyhydroxide with Stoichiometrically Bound Phosphate', *J. Chromatogr.*, **481**, 175–186 (1989).
624. S. Hjertén, I. Zelikman, J. Lindeberg, M. Lederer, 'High-performance Adsorption Chromatography of Proteins on Deformed Non-porous Agarose Beads Coated with Insoluble Metal Compounds. II. Coating: Aluminium and Zirconium (Hydr)oxide with Stoichiometrically Bound Phosphate', *J. Chromatogr.*, **481**, 187–199 (1989).
625. H. Iwase, K. Gondo, T. Koike, I. Ono, 'Novel Precolumn Deproteinization Method Using a Hydroxyapatite Cartridge for the Determination of Theophylline and Diazepam in Human Plasma by High-performance Liquid Chromatography with Ultra-violet Detection', *J. Chromatogr. B*, **655**, 73–81 (1994).
626. L.A. Sluyterman, O. Elgersma, 'Chromatofocusing: Isoelectric Focusing in Ion-exchange Columns. I. General Principles', *J. Chromatogr.*, **150**, 17–30 (1978).
627. L.A. Sluyterman, J. Wijdeness, 'Chromatofocusing: Isoelectric Focusing in Ion-exchange Columns. II. Experimental Verification', *J. Chromatogr.*, **150**, 31–44 (1978).
628. M.O. Funk, M.A. Whitney, E.C. Hausknecht, E.M. O'Brien, 'Resolution of the Isoenzymes of Soybean Lipoxygenase Using Isoelectric Focusing and Chromatofocusing', *Anal. Biochem.*, **146**, 246–251 (1985).
629. B.M. Moore, W.H. Flurkey, 'Mini-chromatofocusing of Plant and Fungal Polyphenoloxidases', *Anal. Biochem.*, **172**, 504–508 (1988).
630. T.W. Hutchens, 'Chromatofocusing', in *Protein Purification. Principles, High Resolution Methods, and Applications*, 2nd edition, eds J.C. Janson, L. Rydén, Wiley-Liss, New York, 207–238, 1998.
631. F.J. Burgos, R. Pascual, J. Vendrell, C.M. Cuchillo, F.X. Aviles, 'The Separation of Pancreatic Procarboxypeptidases by High-performance Liquid Chromatography and Chromatofocusing', *J. Chromatogr.*, **481**, 233–243 (1989).
632. T.W. Hutchens, R.D. Wiehle, N.A. Shahabi, J.L. Wittliff, 'Rapid Analysis of Estrogen Receptor Heterogeneity by Chromatofocusing with High-performance Liquid Chromatography', *J. Chromatogr.*, **266**, 115–128 (1983).
633. L.G. Fägerstam, J. Lizana, U.B. Axio-Fredriksson, L. Wahlström, 'Fast Chromatofocusing of Human Serum Proteins with special Reference to α1-Antitrypsin and Gc-Globulin', *J. Chromatogr.*, **266**, 523–532 (1983).
634. S.P. Monkarsh, E.A. Russoman, S.K. Roy, 'Separation of Interleukins by a Preparative Chromatofocusing-like Method', *J. Chromatogr.*, **631**, 277–280 (1993).
635. M.T.W. Hearn, D. Lyttle, 'Buffer-focusing Chromatography [of proteins] Using Multi-component-electrolyte Elution Systems', *J. Chromatogr.*, **218**, 483–495 (1981).
636. T.W. Hutchens, C.M. Li, P. Besch, 'Development of Focusing Buffer Systems for Generation of Wide-range pH Gradients during High-performance Chromatofocusing', *J. Chromatogr.*, **359**, 157–168 (1986).
637. T.W. Hutchens, C.M. Li, P. Besch, 'Performance Evaluation of a Focusing Buffer Developed for Chromatofocusing on High-performance Anion-exchange Columns', *J. Chromatogr.*, **359**, 169–179 (1986).
638. Y. Liu, D.J. Anderson, 'Gradient Chromatofocusing High-performance Liquid Chromatography. II. Theoretical Aspects', *J. Chromatogr. A*, **762**, 47–54 (1997).
639. Y. Liu, D.J. Anderson, 'Gradient Chromatofocusing High-performance Liquid Chromatography. I. Practical Aspects', *J. Chromatogr. A*, **762**, 207–217 (1997).
640. D.D. Frey, A. Barnes, J. Strong, 'Numerical Studies of Multicomponent Chromatography Using pH Gradients', *AIChE J.*, **41**, 1171–1183 (1995).
641. R.C. Bates, D.D. Frey, 'Quasi-linear pH Gradients for Chromatofocusing Using Simple Buffer Mixtures: Local

Equilibrium Theory and Experimental Verification', *J. Chromatogr. A*, **814**, 43–54 (1998).
642. D.D. Frey, 'Local Equilibrium Behavior of Retained pH and Ionic Strength Gradients in Preparative Chromatography', *Biotechnol. Progr.*, **12**, 65–72 (1996).
643. J.C. Strong, D.D. Frey, 'Experimental and Numerical Studies of the Chromatofocusing of Dilute Proteins Using Retained pH Gradients formed on a Strong Base Anion-exchange Column', *J. Chromatogr. A*, **769**, 129–143 (1997).
644. C.M. Li, T.W. Hutchens, 'Chromatofocusing', in *Practical Protein Chromatography*, eds A. Kenney, S. Fowell, (Meth. Mol. Biol, Vol. 11), Humana Press, Totowa, 1992, 237–248, 1992.
645. M. Petro, F. Švec, I. Gitsov, J.M.J. Fréchet, 'Molded Monolithic Rod of Macroporous Poly(styrene-*co*-divinylbenzene) as a Separation Medium for High-performance Liquid Chromatography of Synthetic Polymers: ''On-column'' Precipitation–Re-dissolution Chromatography as an Alternative to Size-exclusion Chromatography of Styrene Oligomers and Polymers', *Anal. Chem.*, **68**, 315–321 (1996).
646. E. Schweinheim, H. Engelhardt, 'Investigations on ''Mixed-mode'' Retention Behaviour of Stationary Phases with Mixed Ligands for the Separation of Proteins', *Fresenius' Z. Anal. Chem.*, **330**, 397–398 (1988).
647. B.Y. Zhu, C.T. Mant, R.S. Hodges, 'Hydrophilic-interaction Chromatography of Peptides on Hydrophilic and Strong Cation-exchange Columns', *J. Chromatogr.*, **548**, 13–24 (1991).
648. B.Y. Zhu, C.T. Mant, R.S. Hodges, 'Mixed-mode Hydrophilic and Ionic Interaction Chromatography Rivals Reversed-phase Liquid Chromatography for the Separation of Peptides', *J. Chromatogr.*, **594**, 75–86 (1992).
649. J.R. Litowski, P.D. Semchuk, C.T. Mant, R.S. Hodges, 'Hydrophilic Interaction/Cation-exchange Chromatography for the Purification of Synthetic Peptides from Closely Related Impurities: Serine Side-chain Acetylated Peptides', *J. Pept. Res.*, **54**, 1–11 (1999).
650. N. Hirata, M. Kasai, Y. Yanagihara, K. Noguchi, 'Separation of Peptides by Multimode Chromatography on a Column Packed with Vinyl Alcohol Copolymer Gel', *J. Chromatogr.*, **434**, 71–82 (1988).
651. T.R. Floyd, S.E. Cicero, S.D. Fazio, T.V. Raglione, S.H. Hsu, S.A. Winkle, R.A. Hartwick, 'Mixed-mode Hydrophobic Ion Exchange for the Separation of Oligonucleotides and DNA Fragments Using HPLC', *Anal. Biochem.*, **154**, 570–577 (1986).
652. T.R. Floyd, R.A. Hartwick, 'Mixed-interaction Stationary Phases in HPLC', *High Perform. Liq. Chromatogr.*, 445–490 (1986).
653. E.A. Hogendoorn, P. van Zoonen, 'Coupled-column Reversed-phase Liquid Chromatography in Environmental Analysis', *J. Chromatogr. A*, **703**, 149–166 (1995).
654. T. Suortti, 'Coupled Size-exclusion Chromatography–Anion-exchange Chromatography in the Analysis of Poly- and Oligosaccharides', *J. Chromatogr. A*, **763**, 331–335 (1997).
655. Y. Fang, L. Dumont, B. Larsen, 'Real-time Isoform Analysis by Two-dimensional Chromatography of a Monoclonal Antibody during Bioreactor Fermentations', *J. Chromatogr. A*, **816**, 39–47 (1998).
656. Y.L.F. Hsieh, H.Q. Wang, C. Elicone, J. Mark, S.A. Martin, F. Regnier, 'Automated Analytical System for the Examination of Protein Primary Structure', *Anal. Chem.*, **68**, 455–462 (1996).
657. M.M. Bushey, J.W. Jorgenson, 'Automated Instrumentation for Comprehensive Two-dimensional High-performance Liquid Chromatography of Proteins', *Anal. Chem.*, **62**, 161–167 (1990).
658. K. Kasai, 'Size-dependent Chromatographic Separation of Nucleic Acids', *J. Chromatogr.*, **618**, 203–221 (1993).
659. J. Hirabayashi, N. Ito, K. Noguchi, K. Kasai, 'Slalom Chromatography: Size-dependent Separation of DNA Molecules by a Hydrodynamic Phenomenon', *Biochemistry*, **29**, 9515–9521 (1990).
660. H. Small, 'Discoveries Concerning the Transport of Colloids and New Forms of Chromatography', *Acc. Chem. Res.*, **25**, 241–246 (1992).
661. J. Hirabayashi, K. Kasai, 'Applied Slalom Chromatography. Improved DNA Separation by the Use of Columns Developed for Reversed-phase Chromatography', *J. Chromatogr. A*, **722**, 135–142 (1996).
662. E.V. Chubarova, V.V. Nesterov, 'Behavior of Macromolecules in Nonhomogeneous Hydrodynamic Fields: Degradation Mechanism of Macromolecules', *ACS Symp. Ser.*, **635**, 127–156 (1996).
663. D. Parriott (ed.), *A Practical Guide to HPLC Detection*, Academic Press, London, 1992.
664. G. Patonay (ed.), *HPLC Detection. Newer Methods*, VCH, Weinheim, 1993.
665. L. Huber, S.A. George (eds), *Diode Array Detection in HPLC*, Marcel Dekker, New York, 1993.
666. R.P.W. Scott, *Chromatographic Detectors. Design, Function, and Operation*, Marcel Dekker, New York, 1996.
667. G.P. Rozing, 'Diode Array Detection', *Methods Enzymol.*, **270**, 201–232 (1996).
668. I.S. Krull (ed.), *Reaction Detection in Liquid Chromatography*, Dekker: New York, 1986.
669. L. Tao, R.T. Kennedy, 'Laser-induced Fluorescence Detection in Microcolumn Separations', *Trends Anal. Chem.*, **17**, 484–491 (1998).
670. S.A. Westwood, D.E. Games, L. Sheen, 'Use of Circular Dichroism as a High-performance Liquid-chromatograph Detector', *J. Chromatogr.*, **204**, 103–107 (1981).
671. N.J. Greenfield, 'Applications of Circular Dichroism in Protein and Peptide Analysis', *Trends Anal. Chem.*, **18**, 236–244 (1999).
672. A. Mannschreck, 'Online Measurement of Circular Dichroism Spectra during Enantioselective Liquid Chromatography', *Trends Anal. Chem.*, **12**, 220–225 (1993).

673. R. Sheng, F. Ni, T.M. Cotton, 'Determination of Purine Bases by Reversed-phase High-performance Liquid Chromatography Using Real-time Surface-enhanced Raman Spectroscopy', *Anal. Chem.*, **63**, 437–442 (1991).
674. D.L. Olson, M.E. Lacey, J.V. Sweedler, 'NMR Detection for Trace Analysis and Capillary Separations', *Anal. Chem.*, **70**, 257A–264A (1998).
675. A. Hanning, J. Roeraade, 'A Spectroscopic Refractometer for Temperature-independent Refractive Index Detection', *Anal. Chem.*, **69**, 1496–1503 (1997).
676. G. Cepria, J.R. Castillo, 'Surface Plasmon Resonance-based Detection–an Alternative to Refractive Index Detection in High-performance Liquid Chromatography', *J. Chromatogr. A*, **759**, 27–35 (1997).
677. E. Nice, M. Lackmann, F. Smyth, L. Fabri, A.W. Burgess, 'Synergies Between Micropreparative High-performance Liquid Chromatography and an Instrumental Optical Biosensor', *J. Chromatogr. A*, **660**, 169–185 (1994).
678. I. Blikstad, L.G. Fagerstam, R. Bhikhabhai, H. Lindblom, 'Detection and Characterization of Oligosaccharides in Column Effluents Using Surface Plasmon Resonance', *Anal. Biochem.*, **233**, 42–49 (1996).
679. J. Emneus, G. Marko Varga, 'Biospecific Detection in Liquid-chromatography', *J. Chromatogr. A*, **703**, 191–243 (1995).
680. I.V. Astafieva, G.A. Eberlein, Y.J. Wang, 'Absolute Online Molecular Mass Analysis of Basic Fibroblast Growth Factor and its Multimers by Reversed-phase Liquid Chromatography with Multi-angle Laser Light Scattering Detection', *J. Chromatogr. A*, **740**, 215–229 (1996).
681. B.H. Hsu, E. Orton, S.Y. Tang, R.A. Carlton, 'Application of Evaporative Light Scattering Detection to the Characterization of Combinatorial and Parallel Synthesis Libraries for Pharmaceutical Drug Discovery', *J. Chromatogr. B*, **725**, 103–112 (1999).
682. D.C. Johnson, D. Dobberpuhl, R. Roberts, P. Vandeberg, 'Pulsed Amperometric Detection of Carbohydrates, Amines and Sulfur Species in Ion Chromatography–the Current State of Research', *J. Chromatogr.*, **640**, 79–96 (1993).
683. R.W. Frei, K. Zech (eds), *Selective Sampling Handling and Detection in High-performance Liquid Chromatography. Part A*, Elsevier, Amsterdam, 1987.
684. K. Zech, R.W. Frei, (eds), *Selective Sample Handling and Detection in High-performance Liquid Chromatography. Part B*, Elsevier, Amsterdam, 1989.
685. H. Lingeman, W.J.M. Underberg (eds), *Detection-oriented Derivatization Techniques in Liquid Chromatography*, Dekker, New York, 1990.
686. K. Blau, J. Halket, *Handbook of Derivatives for Chromatography*, John Wiley and Sons, Chichester, 1993.
687. M. Szulc, R. Mhatre, J. Mazzeo, I.S. Krull, 'Detection in Liquid Chromatography', *Methods Enzymol.*, **270**, 175–201 (1996).
688. I.S. Krull, R. Strong, Z. Sosic, B.Y. Cho, S.C. Beale, C.C. Wang, S. Cohen, 'Labelling Reactions Applicable to Chromatography and Electrophoresis of Minute Amounts of Proteins', *J. Chromatogr. B*, **699**, 173–208 (1997).
689. M. Koller, H. Eckert, 'Derivatization of Peptides for their Determination by Chromatographic Methods', *Anal. Chim. Acta*, **352**, 31–59 (1997).
690. V.K. Boppana, C. Miller-Stein, 'High-performance Liquid-chromatographic Determination of Peptide Drugs in Biological Fluids by means of Pre- and Post-column Fluorescence Derivatization Techniques', *Anal. Chim. Acta*, **352**, 61–69 (1997).
691. T. Suzuki, T. Watanabe, T. Toyo'oka, 'Discrimination of D/L-Amino-acid in Peptide Sequence based on Fluorescent Chiral Derivatization by Reversed-phase Liquid Chromatography', *Anal. Chim. Acta*, **352**, 357–363 (1997).
692. P. Mascagni, H.L. Ball, G. Bertolini, 'Selective Purification of Synthetic Proteins by the use of FMOC- and Biotin-based Reversible Chromatographic Probes', *Anal. Chim. Acta*, **352**, 375–385 (1997).
693. G. Lunn, L. Hellwig, A. Cecchini, *Handbook of Derivatization Reactions of HPLC*, Wiley, New York, 1998.
694. M. Hassel, G. Klenk, M. Frohme, 'Prevention of Unwanted Proteolysis During Extraction of Proteins from Protease-rich Tissue', *Anal. Biochem.*, **242**, 274–275 (1996).
695. K.J. Sweadner, 'Trypsin Inhibitor Paradoxically Stabilizes Trypsin Activity in Sodium Dodecyl Sulphate, facilitating Proteolytic Fingerprinting', *Anal. Biochem.*, **194**, 130–135 (1991).
696. F.S. Wu, M.Y. Wang, 'Extraction of Proteins for Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis from Protease-rich Plant Tissues', *Anal. Biochem.*, **139**, 100–103 (1984).
697. A. Sanchez Ferrer, M. Perez Gilabert, E. Nunez, R. Bru, F. Garcia Carmona, 'Triton X-114 Phase Partitioning in Plant Protein Purification', *J. Chromatogr. A*, **668**, 75–83 (1994).
698. H. Tani, T. Kamidate, H. Watanabe, 'Micelle-mediated Extraction', *J. Chromatogr. A*, **780**, 229–241 (1997).
699. H. Walter, G. Johansson, 'Partitioning in Aqueous Two-phase Systems: an Overview', *Anal. Biochem.*, **155**, 215–242 (1986).
700. H. Walter, G. Johansson, D.E. Brooks, 'Partitioning in Aqueous Two-phase Systems: Recent Results', *Anal. Biochem.*, **197**, 1–18 (1991).
701. T. Agasøster, 'Aqueous Two-phase Partitioning Sample Preparation Prior to Liquid Chromatography of Hydrophilic Drugs in Blood', *J. Chromatogr. B*, **716**, 293–298 (1998).
702. C.T. Wehr, 'Sample Preparation and Column Regeneration in Biopolymer Separations', *J. Chromatogr.*, **418**, 27–50 (1987).

703. T. Herraiz, 'Sample Preparation and Reversed Phase High-performance Liquid Chromatography Analysis of Food-derived Peptides', *Anal. Chim. Acta*, **352**, 119–139 (1997).
704. F.J. Al-Shammary, 'Sample Preparation of Catecholamines and Biologically Active Peptides by Solid-phase Extraction', *J. High. Resolut. Chromatogr.*, **13**, 309–311 (1990).
705. M. Fountoulakis, H.W. Lahm, 'Hydrolysis and Amino-acid Composition Analysis of Proteins', *J. Chromatogr. A*, **826**, 109–134 (1998).
706. T. Kaneko, K. Katoh, M. Fujimoto, M. Kumagai, J. Tamaoka, Y. Katayama-Fujimura, 'Determination of the Nucleotide Composition of a Deoxyribonucleic Acid by High-performance Liquid Chromatography of its Enzymic Hydrolysate: a Review', *J. Microbiol. Methods*, **4**, 229–240 (1986).
707. J.C. Waterval, H. Krabbe, J. Teeuwsen, A. Bult, H. Lingeman, W.J. Underberg, 'Comparison Between Transient Isotachophoretic Capillary Zone Electrophoresis and Reversed-phase Liquid Chromatography for the Determination of Peptides in Plasma', *Electrophoresis*, **20**, 2909–2916 (1999).
708. S. Linde, J.H. Nielsen, B. Hansen, B.S. Welinder, 'High-performance Liquid Chromatography of Rat and Mouse Islet Polypeptides: Potential Risk of Oxidation of Methionine Residues during Sample Preparation', *J. Chromatogr.*, **530**, 29–37 (1990).
709. C. Offidani, S. Strano Rossi, M. Chiarotti, 'Improved Enzymic Hydrolysis of Hair', *Forensic. Sci. Int.*, **63**, 171–174 (1993).
710. R.A. Kroczek, 'Southern and Northern Analysis', *J. Chromatogr.*, **618**, 133–145 (1993).
711. Y. Baba, T. Matsuura, K. Wakamoto, M. Tsuhako, 'Comparison of High-performance Liquid Chromatography with Capillary Gel Electrophoresis in Single-base Resolution of Polynucleotides', *J. Chromatogr.*, **558**, 273–284 (1991).
712. M.G. Khaledi, *High-performance Capillary Electrophoresis: Theory, Techniques and Applications*, Wiley-Interscience, New York, 1998.
713. C. Sumita, Y. Baba, K. Hide, N. Ishimaru, K. Samata, A. Tanaka, M. Tsuhako, 'Comparative Study of Non-porous Anion-exchange Chromatography, Capillary Gel Electrophoresis and Capillary Electrophoresis in Polymer Solutions in the Separation of DNA Restriction Fragments', *J. Chromatogr. A*, **661**, 297–303 (1994).
714. J.R. Thayer, R.M. McCormick, N. Avdalovic, 'High Resolution Nucleic Acid Separations by High-performance Liquid Chromatography', *Methods Enzymol.*, **271**, 147–174 (1996).
715. C.G. Huber, 'Micropellicular Stationary Phases for High-performance Liquid Chromatography of Double-stranded DNA', *J. Chromatogr. A*, **806**, 3–30 (1998).
716. O. Mikeš, *High-performance Liquid Chromatography of Biopolymers and Biooligomers. Part B: Separation of Individual Compound Classes*, (J. Chromatogr. Lib., Vol. 41B), Elsevier, Amsterdam, 1988.
717. R.R. Deshmukh, W.E. Leitch, II, D.L. Cole, 'Application of Sample Displacement Techniques to the Purification of Synthetic Oligonucleotides and Nucleic Acids: a Mini-review with Experimental Results', *J. Chromatogr. A*, **806**, 77–92 (1998).
718. W.A. Apruzzese, P. Vouros, 'Analysis of DNA Adducts by Capillary Methods Coupled to Mass Spectrometry: a Perspective', *J. Chromatogr. A*, **794**, 97–108 (1998).
719. E.L. Esmans, D. Broes, I. Hoes, F. Lemiere, K. Vanhoutte, 'Liquid Chromatography/Mass Spectrometry in Nucleoside, Nucleotide and Modified Nucleotide Characterization', *J. Chromatogr. A*, **794**, 109–127 (1998).
720. J. Cadet, M. Weinfeld, 'Detecting DNA Damage', *Anal. Chem.*, **65**, 675A–682A (1993).
721. K. Frenkel, C.B. Klein, 'Methods used for Analyses of ''Environmentally'' Damaged Nucleic Acids', *J. Chromatogr.*, **618**, 289–314 (1993).
722. J. Cummings, R.C. French, J.F. Smyth, 'Application of High-performance Liquid Chromatography for Recognition of Covalent Nucleic Acid Modification with Anticancer Drugs', *J. Chromatogr.*, **618**, 251–276 (1993).
723. F. Franek, P. Koutnik, 'Separation of Oligonucleosomal DNA Fragments from Apoptotic Animal Cells Using a Triad of Sephacryl Columns', *J. Chromatogr. A*, **806**, 61–65 (1998).
724. H. Itoh, T. Kinoshita, N. Nimura, 'High-performance Liquid-chromatographic Separation of Nucleic Acids on a Fluorocarbon-bonded Silica Gel Column', *J. Chromatogr. A*, **662**, 95–99 (1994).
725. P.A. Doris, P.J. Oefner, B.S. Chilton, A. Hayward-Lester, 'Quantitative Analysis of Gene Expression by Ion-pair High-performance Liquid Chromatography', *J. Chromatogr. A*, **806**, 47–60 (1998).
726. R. Bischoff, L.W. McLaughlin, 'Mixed-mode Chromatographic Matrices for the Resolution of Transfer Ribonucleic Acids', *J. Chromatogr.*, **317**, 251–261 (1984).
727. R. Bischoff, L.W. McLaughlin, 'Nucleic Acid Resolution by Mixed-mode Chromatography', *J. Chromatogr.*, **296**, 329–337 (1984).
728. Z. Wang, N. Bian, X. Qian, R.W. Giese, '*N*-Hydroxysuccinimide Ester Labeling 5'-Aminoalkyl DNA Oligomers: Reaction Conditions and Purification', *J. Chromatogr. A*, **806**, 93–95 (1998).
729. L.W. McLaughlin, 'Mixed-mode Chromatography of Nucleic Acids', *Chem. Rev.*, **89**, 309–319 (1989).
730. D.D. Smith, A.M. Hanly, 'Purification of Synthetic Peptides by High Performance Liquid Chromatography', *Methods Mol. Biol.*, **73**, 75–87 (1997).
731. A.D. Carlson, J.D. Hofer, R.M. Riggin, 'Development of an Optimized Peptide Map for Recombinant Activated Human Protein C by Means of an Experimental Design Strategy', *Anal. Chim. Acta*, **352**, 221–230 (1997).

732. G.J. Opiteck, J.W. Jorgenson, M.A. Moseley, III, R.J. Anderegg, 'Two-dimensional Microcolumn HPLC Coupled to a Single-quadrupole Mass Spectrometer for the Elucidation of Sequence Tags and Peptide Mapping', *J. Microcol. Sep.*, **10**, 365–375 (1998).
733. M.I. Aguilar, S. Mougos, J. Boublik, J. Rivier, M.T.W. Hearn, 'Effect of D-Amino-acid Substitutions on the Reversed-phase High-performance Liquid Chromatography Retention Behaviour of Neuropeptide Y[18–36] Analogues', *J. Chromatogr.*, **646**, 53–65 (1993).
734. E. Krause, S. Rothemund, M. Beyermann, M. Bienert, 'Monitoring of α-Helical Secondary Structures in Peptides by Reversed-phase HPLC of Replacement Sets', *Anal. Chim. Acta*, **352**, 365–374 (1997).
735. S. Blackburn, *CRC Handbook of Chromatography. Peptides*, CRC Press, Boca Raton, 1986.
736. R.B. Holman, A.J. Cross, M.H. Joseph, *High-performance Liquid Chromatography in Neuroscience Research*, John Wiley and Sons, Chichester, 1993.
737. P. Matsudaira (ed.), *A Practical Guide to Protein and Peptide Purification for Microsequencing*, 2nd edition, Academic Press, San Diego, 1993.
738. J.R. Chapman, *Protein and Peptide Analysis by Mass Spectrometry*, Humana Press, Totowa, USA, 1996.
739. N. Lundell, 'A Peptide in a Haystack – Strategies for a Biochromatographer', *LC–GC Int.*, **8**, 636, 638–640, 642–644, 646–647 (1995).
740. L. Hagel, 'Characteristics of Modern Media for Aqueous Size Exclusion Chromatography', *ACS Symp. Ser.*, **635**, 274–327 (1996).
741. G.B. Irvine, 'Size-exclusion High-performance Liquid Chromatography of Peptides: a Review', *Anal. Chim. Acta*, **352**, 387–397 (1997).
742. R. Bollhagen, M. Schmiedberger, E. Grell, 'High-performance Liquid-chromatographic Purification of Extremely Hydrophobic Peptides: Transmembrane Segments', *J. Chromatogr. A*, **711**, 181–186 (1995).
743. C.T. Mant, R.S. Hodges, 'Separation of Peptides by Strong Cation-exchange High-performance Liquid Chromatography', *J. Chromatogr.*, **327**, 147–155 (1985).
744. T.W. Burke, C.T. Mant, J.A. Black, R.S. Hodges, 'Strong Cation-exchange High-performance Liquid Chromatography of Peptides. Effect of Non-specific Hydrophobic Interactions and Linearization of Peptide Retention Behaviour', *J. Chromatogr.*, **476**, 377–389 (1989).
745. M. Gazdag, K. Mihályfi, S. Görög, 'Ion-pair RP-HPLC Separation of Thymopoietin Fragments and Related Peptides', *Anal. Chim. Acta*, **352**, 231–237 (1997).
746. G. Lindeberg, H. Bennich, A. Engström, 'Purification of Synthetic Peptides. Immobilized Metal Ion Affinity Chromatography (IMAC)', *Int. J. Pept. Protein Res.*, **38**, 253–259 (1991).
747. M. Rholam, P. Cohen, 'Peptides as Tools for Studying Propeptide and Proprotein Processing', *Anal. Chim. Acta*, **352**, 155–178 (1997).
748. T.H.J. Huisman, 'High-performance Liquid-chromatographic Analysis of Human Haemoglobins and Their Polypeptide Chains: Its Use in the Identification of Variants', *Anal. Chim. Acta*, **352**, 187–200 (1997).
749. K. Rissler, 'Sample Preparation, High-performance Liquid Chromatographic Separation and Determination of Substance P-related Peptides', *J. Chromatogr. B*, **665**, 233–270 (1995).
750. W.J. Hurst, I.S. Zagon, 'Isolation, Separation and Detection of Enkephalins: a Review of Methods for High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Liq. Chromatogr.*, **18**, 2943–2967 (1995).
751. G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, R.J. Anderegg, 'Comprehensive Online LC-LC-MS of Proteins', *Anal. Chem.*, **69**, 1518–1524 (1997).
752. A.W. Moore, Jr, J.W. Jorgenson, 'Comprehensive Three-dimensional Separation of Peptides Using Size Exclusion Chromatography–Reversed-phase Liquid Chromatography–Optically Gated Capillary Zone Electrophoresis', *Anal. Chem.*, **67**, 3456–3463 (1995).
753. A.W. Moore, Jr, J.W. Jorgenson, 'Rapid Comprehensive Two-dimensional Separations of Peptides via Reversed-phase LC–Optically Gated Capillary Zone Electrophoresis', *Anal. Chem.*, **67**, 3448–3455 (1995).
754. D.O. Lambeth, W.W. Muhonen, 'High-performance Liquid Chromatography-based Assays of Enzyme Activities', *J. Chromatogr. B*, **656**, 143–157 (1994).
755. D.N. Nguyen, G.W. Becker, R.M. Riggin, 'Protein Mass Spectrometry: Applications to Analytical Biotechnology', *J. Chromatogr. A*, **705**, 21–45 (1995).
756. H. Cui, J. Leon, E. Reusaet, A. Bult, 'Selective Determination of Peptides Containing Specific Amino-acid Residues by High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr. A*, **704**, 27–36 (1995).
757. S. Oscarsson, 'Factors Affecting Protein Interaction at Sorbent Interfaces', *J. Chromatogr. B*, **699**, 117–131 (1997).
758. J. Evangelista Dyr, J. Suttnar, 'Separation Used for Purification of Recombinant Proteins', *J. Chromatogr. B*, **699**, 383–401 (1997).
759. F. Frantzen, 'Chromatographic and Electrophoretic Methods for Modified Haemoglobins', *J. Chromatogr. B*, **699**, 269–286 (1997).
760. S. Kishino, K. Miyazaki, 'Separation Methods for Glycoprotein Analysis and Preparation', *J. Chromatogr. B*, **699**, 371–381 (1997).
761. V. Pacakova, K. Stulik, M. Ticha, 'High-performance Separations in Isolation and Characterization of Allergens', *J. Chromatogr. B*, **699**, 403–418 (1997).
762. D.S. Hage, S.A. Tweed, 'Recent Advances in Chromatographic and Electrophoretic Methods for the study of Drug–Protein Interactions', *J. Chromatogr. B*, **699**, 499–525 (1997).

763. R.A. Copeland, *Methods for Protein Analysis*, Chapman and Hall, London, 1994.
764. G. von Jagow, H. Schgger, *A Practical Guide to Membrane Protein Purification*, Academic Press, San Diego, 1994.
765. J. Walker, *The Protein Protocols Handbook*, Humana Press, Totowa, 1996.
766. J.C. Janson, L. Rydén (eds), *Protein Purification. Principles, High Resolution Methods, and Applications*, 2nd edition, Wiley-Liss, New York, 1998.
767. A. Sadana, *Bioseparation of Proteins*, Academic Press, San Diego, 1997.
768. S. Beeckmans, 'Chromatographic Methods to Study Protein–Protein Interactions', *Methods*, **19**, 278–305 (1999).
769. D. Josic, W. Reutter, 'Sample Preparation and Application of Different HPLC Methods for the Separation of Hydrophobic Membarne Proteins', in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, eds C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 231–238, 1991.
770. D. Josic, K. Zeilinger, 'Membrane Proteins', *Methods Enzymol.*, **271**, 113–135 (1996).
771. P. Lundahl, Y. Watanabe, T. Takagi, 'High-performance Hydroxyapatite Chromatography of Integral Membrane-proteins and Water-soluble Proteins in Complex with Sodium Dodecyl-sulfate', *J. Chromatogr.*, **604**, 95–102 (1992).
772. T. Takey, W.C. Purdy, 'Chromatographic Techniques for the Isolation and Purification of Lipoproteins', *J. Chromatogr. B*, **671**, 237–253 (1995).
773. S.K.R. Williams, G.M. Raner, W.R. Ellis, Jr, J.C. Giddings, 'Separation of Protein Inclusion Bodies from *Escherichia coli* Lysates Using Sedimentation Field-flow Fractionation', *J. Microcolumn Sep.*, **9**, 233–239 (1997).
774. S.G. Walker, A. Lyddiatt, 'Aqueous Two-phase Systems as an Alternative Process Route for the Fractionation of Small Inclusion Bodies', *J. Chromatogr. B*, **711**, 185–194 (1998).
775. A. Jungbauer, E. Boschetti, 'Manufacture of Recombinant Proteins with Safe and Validated Chromatographic Sorbents', *J. Chromatogr. B*, **662**, 143–179 (1994).
776. J. Lei, D.A. Chen, F.E. Regnier, 'Rapid Verification of Disulfide Linkages in Recombinant Human Growth Hormone by Tandem Column Tryptic Mapping', *J. Chromatogr. A*, **808**, 121–131 (1998).
777. K. Ikegaya, M. Hirose, T. Ohmura, K. Nokihara, 'Complete Determination of Disulfide Forms of Purified Recombinant Human Serum Albumin, Secreted by the Yeast *Pichiapastoris*', *Anal. Chem.*, **69**, 1986–1991 (1997).
778. K.A. Denton, S.A. Tate, 'Capillary Electrophoresis of Recombinant Proteins', *J. Chromatogr. B*, **697**, 111–121 (1997).
779. A. Pantazaki, M. Taverna, C. Vidal-Madjar, 'Recent Advances in the Capillary Electrophoresis of Recombinant Glycoproteins', *Anal. Chim. Acta*, **383**, 137–156 (1999).
780. F. Toribio, J. Alhama, J. Lopez-Barea, 'Methods for Chromatographic and Electrophoretic Separation and Assay of $NADP^+$ Oxidoreductases', *J. Chromatogr. B*, **684**, 1–23 (1996).
781. N. Bergenhem, 'Chromatographic and Electrophoretic Methods Related to the Carbonic Anhydrase Isozymes', *J. Chromatogr. B*, **684**, 289–305 (1996).
782. G. Kopperschläger, J. Kirchberger, 'Methods for the Separation of Lactate Dehydrogenases and Clinical Significance of the Enzyme', *J. Chromatogr. B*, **684**, 25–49 (1996).
783. W.P. Michalski, 'Chromatographic and Electrophoretic Methods for Analysis of Superoxide Dismutases', *J. Chromatogr. B*, **684**, 59–75 (1996).
784. S.X. Lin, B. Xie, B. Li, M. Zhou, 'Chromatography of Hydroxysteroid Dehydrogenases', *J. Chromatogr. B, Biomed. Sci. Appl.*, **684**, 99–105 (1996).
785. P.H. Roos, 'Chromatographic Separation and Behaviour of Microsomal Cytochrome P-450 and Cytochrome *b*5', *J. Chromatogr. B*, **684**, 107–131 (1996).
786. C. Tilgmann, I. Ulmanen, 'Purification Methods of Mammalian Catechol-*O*-methyltransferases', *J. Chromatogr. B*, **684**, 147–161 (1996).
787. T. Moriyama, H. Ikeda, 'Hydrolases Acting on Glycosidic Bonds: Chromatographic and Electrophoretic Separations', *J. Chromatogr. B*, **684**, 201–216 (1996).
788. O. Ibrahim-Granet, O. Bertrand, 'Separation of Proteases: Old and New Approaches', *J. Chromatogr. B*, **684**, 239–263 (1996).
789. J.S. Hamada, 'Large-scale High-performance Liquid Chromatography of Enzymes for Food Applications', *J. Chromatogr. A*, **760**, 81–87 (1997).
790. C. Parmentier, P. Leroy, A. Visvikis, A. Nicolas, 'Liquid-chromatographic Methods for the Purification and Activity Measurement of Glutathione-related Enzymes', *LC–GC Int.*, **9**, 356, 360, 362, 366, 368, 370, 372 (1996).
791. B. Wilhelm, A. Meinhardt, J. Seitz, 'Transglutaminases: Purification and Activity Assays', *J. Chromatogr. B*, **684**, 163–177 (1996).
792. G. Raspi, 'Kallikrein and Kallikrein-like Proteinases: Purification and Determination by Chromatographic and Electrophoretic Methods', *J. Chromatogr. B*, **684**, 265–287 (1996).
793. B.R. Knudsen, T. Straub, F. Boege, 'Separation and Functional Analysis of Eukaryotic DNA Topoisomerases by Chromatography and Electrophoresis', *J. Chromatogr. B*, **684**, 307–321 (1996).
794. I. Leopold, B. Fricke, 'Inhibition, Reactivation, and Determination of Metal Ions in Membrane Metalloproteases of Bacterial Origin Using High-performance Liquid Chromatography Coupled On-line with Inductively Coupled Plasma Mass Spectrometry', *Anal. Biochem.*, **252**, 277–285 (1997).

795. E.H.J.M. Jansen, P. De Fluiter, 'Detection of the Enzymic Activity of Cytochrome P-450 Enzymes by High-performance Liquid Chromatography', *J. Chromatogr.*, **580**, 325–346 (1992).
796. G.W. Welling, A.J. Scheffer, S. Welling-Wester, 'Determination of Enzyme Activity by High-performance Liquid Chromatography', *J. Chromatogr. B*, **659**, 209–225 (1994).
797. J.A. Boutin, 'Tyrosine Protein Kinase Assays', *J. Chromatogr. B*, **684**, 179–199 (1996).
798. K. Benedek, 'Kinetics of Recombinant Human Brain-derived Neurotropic Factor Unfolding under Reversed-phase Liquid Chromatography Conditions', *J. Chromatogr.*, **646**, 91–98 (1993).
799. A.W. Purcell, M.I. Aguilar, M.T.W. Hearn, 'Dynamics of Peptides in Reversed-phase High-performance Liquid Chromatography', *Anal. Chem.*, **65**, 3038–3047 (1993).
800. K.L. Richards, M.I. Aguilar, M.T.W. Hearn, 'Effect of Protein Conformation on Experimental Bandwidths in Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. A*, **676**, 33–41 (1994).
801. R.S. Hodges, B.Y. Zhu, N.E. Zhou, C.T. Mant, 'Reversed-phase Liquid Chromatography as a Useful Probe of Hydrophobic Interactions Involved in Protein Folding and Protein Stability', *J. Chromatogr. A*, **676**, 3–15 (1994).
802. E. Quaite-Randall, A. Joachimiak, 'Purification of Chaperonins', *J. Chromatogr. B*, **722**, 153–177 (1999).
803. M. Hasegawa, A. Isogai, F. Onabe, 'Size-exclusion Chromatography of Cellulose and Chitin Using Lithium Chloride-*N*,*N*-dimethylacetamine as a Mobile Phase', *J. Chromatogr.*, **635**, 334–337 (1993).
804. M. Strlic, J. Kolar, M. Zigon, B. Pihlar, 'Evaluation of Size-exclusion Chromatography and Viscometry for the Determination of Molecular Masses of Oxidised Cellulose', *J. Chromatogr. A*, **805**, 93–99 (1998).
805. J. Chmelik, J. Chmelikova, M.V. Novotny, 'Characterization of Dextrans by Size-exclusion Chromatography on Unmodified Silica Gel Columns, with Light-scattering Detection, and Capillary Electrophoresis with Laser-induced Fluorescence Detection', *J. Chromatogr. A*, **790**, 93–100 (1997).
806. S.C. Churms, 'Recent Progress in Carbohydrate Separation by High-performance Liquid Chromatography based on Size Exclusion', *J. Chromatogr. A*, **720**, 151–166 (1996).
807. M.E. Vazquez-Blanco, M.L. Vazquez-Oderiz, J. Lopez-Hernandez, J. Simal-Lozano, M.A. Romero-Rodriguez, 'HPLC Determination of Pectins in Raspberries as Galacturonic Acid and Optimization Using Forward Optical Scanning', *J. Chromatogr. Sci.*, **31**, 477–479 (1993).
808. K. Koizumi, 'High-performance Liquid-chromatographic Separation of Carbohydrates on Graphitized Carbon Columns', *J. Chromatogr. A*, **720**, 119–126 (1996).
809. J. Zhang, D.I.C. Wang, 'Quantitative Analysis and Process Monitoring of Site-specific Glycosylation Microheterogeneity in Recombinant Human Interferon-γ from Chinese Hamster Ovary Cell Culture by Hydrophilic Interaction Chromatography', *J. Chromatogr. B*, **712**, 73–82 (1998).
810. Y. Zhang, Y. Inoue, S. Inoue, Y.C. Lee, 'Separation of Oligo/polymers of 5-*N*-Acetylneuraminic Acid, 5-*N*-Glycolylneuraminic Acid, and 2-Keto-3-deoxy-D-glycero-D-galacto-nonionic acid by High-performance Anion-exchange Chromatography with Pulsed Amperometric Detector', *Anal. Biochem.*, **250**, 245–251 (1997).
811. M. Caron, A.P. Seve, D. Bladier R. Joubert-Caron, 'Glycoaffinity Chromatography and Biological Recognition', *J. Chromatogr. B*, **715**, 153–161 (1998).
812. S. Hase, 'Precolumn Derivatization for Chromatographic and Electrophoretic Analysis of Carbohydrates', *J. Chromatogr. A*, **720**, 173–182 (1996).
813. T. Nishikawa, S. Suzuki, H. Kubo, H. Ohtani, 'On-column Isomerization of Sugars During High-performance Liquid Chromatography: Analysis of the Elution Profile', *J. Chromatogr. A*, **720**, 167–172 (1996).
814. S. Honda, 'Postcolumn Derivatization for Chromatographic Analysis of Carbohydrates', *J. Chromatogr. A*, **720**, 183–199 (1996).
815. J. Charlwood, D. Tolson, M. Dwek, P. Camilleri, 'A Detailed Analysis of Neutral and Acidic Carbohydrates in Human Milk', *Anal. Biochem.*, **273**, 261–277 (1999).
816. S. Kishino, A. Nomura, M. Saitoh, M. Sugawara, K. Iseki, A. Kitabatake, K. Miyazaki, 'Single Step Isolation Method for Six Glycoforms of Human α1-Acid Glycoprotein by Hydroxylapatite Chromatography and Study of their Binding Capacities for Disopyramide', *J. Chromatogr. B*, **703**, 1–6 (1997).
817. K. Kakehi, S. Honda, 'Analysis of Glycoproteins, Glycopeptides and Glycoprotein-derived Oligosaccharides by High-performance Capillary Electrophoresis', *J. Chromatogr. A*, **720**, 377–393 (1996).
818. R.R. Townsend, L.J. Basa, M.W. Spellman, 'Identification and Characterization of Glycopeptides in Tryptic Maps by High-pH Anion-exchange Chromatography', *Methods Enzymol.*, **271**, 135–147 (1996).
819. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, *Molecular Biology of the Cell*, 3rd edition, Garland Publishing, New York, 608, 1994.
820. H. Savolainen, 'Isolation and Separation of Proteoglycans', *J. Chromatogr. B*, **722**, 255–262 (1999).
821. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, *Molecular Biology of the Cell*, 3rd edition, Garland Publishing, New York, 971, 1994.
822. D.J. Harvey, 'Matrix-assisted Laser Desorption-ionization Mass Spectrometry of Oligosaccharides and Glycoconjugates', *J. Chromatogr. A*, **720**, 429–446 (1996).
823. S. Suzuki, S. Honda, 'A Tabulated Review of Capillary Electrophoresis of Carbohydrates', *Electrophoresis*, **19**, 2539–2560 (1998).

824. N.K. Karamanos, A. Hjerpe, 'A Survey of Methodological Challenges for Glycosaminoglycan/Proteoglycan Analysis and Structural Characterization by Capillary Electrophoresis', *Electrophoresis*, **19**, 2561–2571 (1998).
825. M. Lafosse, B. Herbreteau, L. Morin-Allory, 'Supercritical-fluid Chromatography of Carbohydrates', *J. Chromatogr. A*, **720**, 61–73 (1996).
826. R.A. O'Neill, 'Enzymic Release of Oligosaccharides from Glycoproteins for Chromatographic and Electrophoretic Analysis', *J. Chromatogr. A*, **720**, 201–215 (1996).
827. S. Prime, J. Dearnley, A.M. Ventom, R.B. Parekh, C.J. Egde, 'Oligosaccharide Sequencing Based on Exo- and Endoglycosidase Digestion and Liquid-chromatographic Analysis of the Products', *J. Chromatogr. A*, **720**, 263–274 (1996).
828. J. Colangelo, R. Orlando, 'On-target Exoglycosidase Digestions-MALDI MS for Determining the Primary Structures of Carbohydrate Chains', *Anal. Chem.*, **71**, 1479–1482 (1999).
829. P.E. Shaw, *CRC Handbook of Sugar Separations in Foods by HPLC*, CRC Press, Boca Raton, 1988.
830. S.C. Churms, *CRC Handbook of Chromatography: Carbohydrates*, CRC Press, Boca Raton, FL, Vol. I, 1982 and Vol. II, 1990.
831. Z. El Rassi, *Carbohydrate Analysis*, (J. Chromatogr. Lib., Vol. 58), Elsevier, Amsterdam, 1995.
832. H. Chen, Cs. Horváth, 'High-speed High-performance Liquid Chromatography of Peptides and Proteins', *J. Chromatogr. A*, **705**, 3–20 (1995).
833. K.K. Unger, G. Jilge, J.N. Kinkel, M.T.W. Hearn, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by High-performance Liquid Chromatography. II. Performance of Non-porous Monodisperse 1.5 μm Silica Beads in the Separation of Proteins by Reversed-phase Gradient-elution High-performance Liquid Chromatography', *J. Chromatogr.*, **359**, 61–72 (1986).
834. K.K. Unger, G. Jilge, R. Janzen, H. Giesche, J.N. Kinkel, 'Non-porous Micro-particulate Supports in High-performance Liquid Chromatography (HPLC) of Biopolymers: Concepts, Realization and Prospects', *Chromatographia*, **22**, 379–380 (1986).
835. D. Corradini, K. Kalghatgi, Cs. Horváth, 'Effect of Mobile Phase Additives on Peptide Retention in Reversed-phase Chromatography with Pellicular and Totally Porous Sorbents', *J. Chromatogr. A*, **728**, 225–233 (1996).
836. N. Nimura, H. Itoh, T. Kinoshita, N. Nagae, M. Nomura, 'Fast Protein Separation by Reversed-phase High-performance Liquid-chromatography on Octadecylsilyl-bonded Nonporous Silica-gel–Effect of Particle-size of Column Packing on Column Efficiency', *J. Chromatogr.*, **585**, 207–211 (1991).
837. H. Chen, Cs. Horváth, 'Rapid Separation of Proteins by Reversed-phase HPLC at Elevated Temperatures', *Anal. Methods Instrum.*, **1**, 213–222 (1993).
838. G. Jilge, R. Janzen, H. Giesche, K.K. Unger, J.N. Kinkel, M.T.W. Hearn, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by High-performance Liquid Chromatography. III. Retention and Selectivity of Proteins and Peptides in Gradient Elution on Non-porous Monodisperse 1.5 μm Reversed-phase Silicas', *J. Chromatogr.*, **397**, 71–80 (1987).
839. J.J. Kirkland, 'Superficially Porous Silica Micro-spheres for the Fast High-performance Liquid Chromatography of Macromolecules', *Anal. Chem.*, **64**, 1239–1245 (1992).
840. H. Small, T.S. Stevens, W.C. Bauman, 'Novel Ion Exchange Chromatographic Method Using Conductimetric Detection', *Anal. Chem.*, **47**, 1801–1809 (1975).
841. R. Janzen, K.K. Unger, H. Giesche, J.N. Kinkel, M.T.W. Hearn, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by High-performance Liquid Chromatography. V. Performance of Non-porous Monodisperse 1.5 μm Bonded Silicas in the Separation of Proteins by Hydrophobic-interaction Chromatography', *J. Chromatogr.*, **397**, 91–97 (1987).
842. R. Janzen, K.K. Unger, H. Giesche, J.N. Kinkel, M.T.W. Hearn, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by High-performance Liquid Chromatography. IV. Mobile Phase and Surface-mediated Effects on Recovery of Native Proteins in Gradient Elution on Non-porous, Monodisperse 1.5 μm Reversed-phase Silicas', *J. Chromatogr.*, **397**, 81–89 (1987).
843. G.K. Bonn, K. Kalghatgi, W.C. Horne, Cs. Horváth, 'Rapid Metal-interaction Chromatography of Proteins and Peptides on Micropellicular Sorbents', *Chromatographia*, **30**, 484–488 (1990).
844. K. Kalghatgi, 'Micropellicular Stationary Phases for Rapid Protein Analysis by High-performance Liquid Chromatography', *J. Chromatogr.*, **499**, 267–278 (1990).
845. Y.F. Maa, Cs. Horváth, 'Rapid Analysis of Proteins and Peptides by Reversed-phase Chromatography with Polymeric Micro-pellicular Sorbents', *J. Chromatogr.*, **445**, 71–86 (1988).
846. Y.F. Maa, S.C. Lin, Cs. Horváth, U.C. Yang, D.M. Crothers, 'Rapid High-performance Liquid Chromatography of Nucleic Acids with Polystyrene-based Micropellicular Anion Exchangers', *J. Chromatogr.*, **508**, 61–73 (1990).
847. D. Letourneur, V. Migonney, D. Müller, M. Jozefowicz, 'Phosphorylated Polystyrene Resins in High-performance Ion-exchange Chromatography', *J. Chromatogr.*, **589**, 87–91 (1992).
848. S. Hjertén, J.L. Liao, 'High-performance Liquid Chromatography of Proteins on Compressed, Non-porous Agarose Beads. I. Hydrophobic-interaction Chromatography', *J. Chromatogr.*, **457**, 165–174 (1988).
849. J.L. Liao, S. Hjertén, 'High-performance Liquid Chromatography of Proteins on Compressed, Non-porous Agarose Beads. II. Anion-exchange Chromatography', *J. Chromatogr.*, **457**, 175–182 (1988).

850. A.E. Rodrigues, 'Intraparticle Convection in Chromatographic Permeable Packings', *ACS Symp. Ser.*, **635**, 157–172 (1996).
851. N.B. Afeyan, N.F. Gordon, I. Mazsaroff, L. Varady, S.P. Fulton, Y.B. Yang, F.E. Regnier, 'Flow-through Particles for the High-performance Liquid-chromatographic Separation of Biomolecules: Perfusion Chromatography', *J. Chromatogr.*, **519**, 1–29 (1990).
852. A.E. Rodrigues, J.C. Lopes, Z.P. Lu, J.M. Loureiro, M.M. Dias, 'Importance of Intraparticle Convection in the Performance of Chromatographic Processes', *J. Chromatogr.*, **590**, 93–100 (1992).
853. A.E. Rodrigues, A.M.D. Ramos, J.M. Loureiro, M. Diaz, Z.P. Lu, 'Influence of Adsorption Desorption-kinetics on the Performance of Chromatographic Processes Using Large-pore Supports', *Chem. Eng. Sci.*, **47**, 4405–4413 (1992).
854. G. Carta, A.E. Rodrigues, 'Diffusion and Convection in Chromatographic Processes Using Permeable Supports with a Bidisperse Pore Structure', *Chem. Eng. Sci.*, **48**, 3927–3935 (1993).
855. A.E. Rodrigues, Z.P. Lu, J.M. Loureiro, G. Carta, 'Peak Resolution in Linear Chromatography. Effects of Intraparticle Convection', *J. Chromatogr. A*, **653**, 189–198 (1993).
856. A.E. Rodrigues, J.M. Loureiro, C. Chenou, M. Rendueles de la Vega, 'Bioseparations with Permeable Particles', *J. Chromatogr. B*, **664**, 233–240 (1995).
857. A.E. Rodrigues, C. Chenou, M. Rendueles de la Vega, 'Protein Separation by Liquid Chromatography Using Permeable POROS Q/M Particles', *Chem. Eng. J. Biochem. Eng. J.*, **61**, 191–201 (1996).
858. M. Rendueles de la Vega, C. Chenou, J.M. Loureiro, A.E. Rodrigues, 'Mass Transfer Mechanisms in Hyper D Media for Chromatographic Protein Separation', *Biochem. Eng. J.*, **1**, 11–23 (1998).
859. V.M.T.M. Silva, A.E. Rodrigues, 'Adsorption and Diffusion in Bidisperse Pore Structures', *Ind. Eng. Chem. Res.*, **38**, 4023–4031 (1999).
860. A.I. Liapis, M.A. McCoy, 'Theory of Perfusion Chromatography', *J. Chromatogr.*, **599**, 87–104 (1992).
861. M.A. McCoy, A.I. Liapis, K.K. Unger, 'Applications of Mathematical-modeling to the Simulation of Binary Perfusion Chromatography', *J. Chromatogr.*, **644**, 1–9 (1993).
862. A.I. Liapis, M.A. McCoy, 'Perfusion Chromatography–Effect of Micropore Diffusion on Column Performance in Systems Utilizing Perfusive Adsorbent Particles with a Bidisperse Porous Structure', *J. Chromatogr. A*, **660**, 85–96 (1994).
863. A.I. Liapis, Y. Xu, O.K. Crosser, A. Tongta, 'Perfusion Chromatography – the Effects of Intraparticle Convective Velocity and Microsphere Size on Column Performance', *J. Chromatogr. A*, **702**, 45–57 (1995).
864. G.A. Heeter, A.I. Liapis, 'Perfusion Chromatography–Performance of Periodic Countercurrent Column Operation and its Comparison with Fixed-bed Operation', *J. Chromatogr. A*, **711**, 3–21 (1995).
865. Y. Xu, A.I. Liapis, 'Modelling and Analysis of the Elution Stage of ''Perfusion Chromatography''–Effects of Intraparticle Convective Velocity and Microsphere Size on System Performance', *J. Chromatogr. A*, **724**, 13–25 (1996).
866. G.A. Heeter, A.I. Liapis, 'Multi-component Perfusion Chromatography in Fixed Bed and Periodic Counter Current Column Operation', *J. Chromatogr. A*, **734**, 105–123 (1996).
867. G.A. Heeter, A.I. Liapis, 'Affinity Adsorption of Adsorbates into Spherical Monodisperse and Bidisperse Porous Perfusive and Purely Diffusive Adsorbent Particles Packed in a Column–Parameter Estimation in the Laplace Transform Domain', *J. Chromatogr. A*, **760**, 55–69 (1997).
868. G.A. Heeter, A.I. Liapis, 'Estimation of Pore Diameter for Intraparticle Fluid Flow in Bidisperse Porous Chromatographic Particles', *J. Chromatogr. A*, **761**, 35–40 (1997).
869. G.A. Heeter, A.I. Liapis, 'Model Discrimination and Estimation of the Intraparticle Mass Transfer Parameters for the Adsorption of Bovine Serum Albumin onto Porous Adsorbent Particles by the Use of Experimental Frontal Analysis Data', *J. Chromatogr. A*, **776**, 3–13 (1997).
870. J.J. Meyers, A.I. Liapis, 'Network Modeling of the Intraparticle Convection and Diffusion of Molecules in Porous Particles Packed in a Chromatographic Column', *J. Chromatogr. A*, **827**, 197–213 (1998).
871. A.I. Liapis, H. Sadikoglu, O.K. Crosser, 'Frontal Chromatography of Proteins: the Effect on Column Performance of the Restricted Diffusion of Molecules in Porous Chromatographic Adsorbents', *J. Chromatogr. A*, **828**, 345–356 (1998).
872. G. Carta, 'Linear Driving-force Approximation for Intraparticle Diffusion and Convection in Permeable Supports', *Chem. Eng. Sci.*, **50**, 887–889 (1995).
873. M.A. Fernandez, W.S. Laughinghouse, G. Carta, 'Characterization of Protein Adsorption by Composite Silica–Polyacrylamide Gel Anion Exchangers. 2. Mass Transfer in Packed Columns and Predictability of Breakthrough Behavior', *J. Chromatogr. A*, **746**, 185–198 (1996).
874. M.A. Fernandez, G. Carta, 'Characterization of Protein Adsorption by Composite Silica–Polyacrylamide Gel Anion Exchangers. 1. Equilibrium and Mass Transfer in Agitated Contactors', *J. Chromatogr. A*, **746**, 169–183 (1996).
875. R.K. Lewus, F.H. Altan, G. Carta, 'Protein Adsorption and Desorption on Gel-filled Rigid Particles for Ion Exchange', *Ind. Eng. Chem. Res.*, **37**, 1079–1087 (1998).
876. R.K. Lewus, G. Carta, 'Binary Protein Adsorption on Gel-composite Ion-exchange Media', *Am. Ind. Chem. Eng. J.*, **45**, 512–522 (1999).

877. E. Boschetti, 'Advanced Sorbents for Preparative Protein Separation Purposes', *J. Chromatogr. A*, **658**, 207–236 (1994).
878. J. Horvath, E. Boschetti, L. Guerrier, N. Cooke, 'High-performance Protein Separations with Novel Strong Ion Exchangers', *J. Chromatogr. A*, **679**, 11–22 (1994).
879. P.E. Gustavsson, P.O. Larsson, 'Superporous Agarose, a new Material for Chromatography', *J. Chromatogr. A*, **734**, 231–240 (1996).
880. P.E. Gustavsson, A. Axelsson, P.O. Larsson, 'Direct Measurement of Convective Fluid Velocities in Superporous Agarose Beads', *J. Chromatogr. A*, **795**, 199–210 (1998).
881. P.E. Gustavsson, A. Axelsson, P.O. Larsson, 'Superporous Agarose Beads as a Hydrophobic Interaction Chromatography Support', *J. Chromatogr. A*, **830**, 275–284 (1999).
882. E. Pålsson, A.L. Smeds, A. Petersson, P.O. Larsson, 'Faster Isolation of Recombinant Factor VIII SQ, with a Superporous Agarose Matrix', *J. Chromatogr. A*, **840**, 39–50 (1999).
883. P.E. Gustavsson, P.O. Larsson, 'Continuous Superporous Agarose Beds for Chromatography and Electrophoresis', *J. Chromatogr. A*, **832**, 29–39 (1999).
884. A.I. Liapis, 'Affinity Adsorption Separations in High Performance Liquid Chromatography and Perfusion Chromatography Systems: the Effects of Pore-size Distribution and Fractal Pores on Column Performance', *Math. Model. Sci. Comput.*, **1**, 397–414 (1993).
885. F.T. Sarfert, M.R. Etzel, 'Mass Transfer Limitations in Protein Separations Using Ion-exchange Membranes', *J. Chromatogr. A*, **764**, 3–20 (1997).
886. H. Splitt, I. Mackenstedt, R. Freitag, 'Preparative Membrane Adsorber Chromatography for the Isolation of Cow Milk Components', *J. Chromatogr. A*, **729**, 87–97 (1996).
887. D.K. Roper, E.N. Lightfoot, 'Separation of Biomolecules Using Adsorptive Membranes', *J. Chromatogr. A*, **702**, 3–26 (1995).
888. E.N. Lightfoot, A.M. Athalye, J.L. Coffman, D.K. Roper, T.W. Root, 'Nuclear Magnetic Resonance and the Design of Chromatographic Separations', *J. Chromatogr. A*, **707**, 45–55 (1995).
889. B.G. Belenkii, A.M. Podkladenko, O.I. Kurenbin, V.G. Maltsev, D.G. Nasledov, S.A. Trushin, 'Peculiarities of Zone Migration and Band Broadening in Gradient Reversed-phase High-performance Chromatography of Proteins with Respect to Membrane Chromatography', *J. Chromatogr. A*, **645**, 1–15 (1993).
890. T.B. Tennikova, F. Svec, 'High-performance Membrane Chromatography: Highly Efficient Separation Method for Proteins in Ion-exchange, Hydrophobic Interaction and Reversed-phase Modes', *J. Chromatogr.*, **646**, 279–288 (1993).
891. Dj. Josic, Y.P. Lim, A. Strancar, W. Reutter, 'Application of High-performance Membrane Chromatography for Separation of Annexins from the Plasma Membranes of Liver and Isolation of Monospecific Polyclonal Antibodies', *J. Chromatogr. B*, **662**, 217–226 (1994).
892. A. Strancar, P. Koselj, H. Schwinn, Dj. Josic, 'Application of Compact Porous Disks for Fast Separations of Biopolymers and In-process Control in Biotechnology', *Anal. Chem.*, **68**, 3483–3488 (1996).
893. Dj. Josic, A. Strancar, 'Application of Membranes and Compact, Porous Units for the Separation of Biopolymers', *Ind. Eng. Chem. Res.*, **38**, 333–342 (1999).
894. Y. Yang, A. Velayudhan, C.M. Ladisch, M.R. Ladisch, 'Protein Chromatography Using a Continuous Stationary Phase', *J. Chromatogr.*, **598**, 169–180 (1992).
895. K.H. Hamaker, S.L. Rau, R. Hendrickson, J. Liu, C.M. Ladisch, M.R. Ladisch, 'Rolled Stationary Phases: Dimensionally Structured Textile Adsorbents for Rapid Liquid Chromatography of Proteins', *Ind. Eng. Chem. Res.*, **38**, 865–872 (1999).
896. A. Singh, N.G. Pinto, 'Polymeric Short-fiber Chromatographic Supports for Downstream Processing of Biomolecules', *React. Polym.*, **24**, 229–242 (1995).
897. J.F. Kennedy, M. Paterson, 'Application of Cellulosic Fast-flow Column Filters to Protein Immobilization and Recovery', *Polym. Int.*, **32**, 71–81 (1993).
898. J.L. Liao, R. Zhang, S. Hjertén, 'Continuous Beds for Standard and Micro High-performance Liquid Chromatography', *J. Chromatogr.*, **586**, 21–26 (1991).
899. S. Hjertén, 'Standard and Capillary Chromatography, Including Electrochromatography, on Continuous Polymer Beds (monoliths), Based on Water-soluble Monomers', *Ind. Eng. Chem. Res.*, **38**, 1205–1214 (1999).
900. T.L. Tisch, R. Frost, J.L. Liao, W.K. Lam, A. Remy, E. Scheinpflug, C. Siebert, H. Song, A. Stapleton, 'Biochemical Separations by Continuous-bed Chromatography', *J. Chromatogr. A*, **816**, 3–9 (1998).
901. S. Vogt, R. Freitag, 'Displacement Chromatography Using the UNO Continuous Bed Column as a Stationary Phase', *Biotechnol. Progr.*, **14**, 742–748 (1998).
902. F. Svec, J.M.J. Fréchet, 'Continuous Rods of Macroporous Polymer as High-performance Liquid Chromatography Separation Media', *Anal. Chem.*, **64**, 820–822 (1992).
903. F. Svec, J.M.J. Fréchet, 'Molded Rigid Monolithic Porous Polymers: an Inexpensive, Efficient, and Versatile Alternative to Beads for the Design of Materials for Numerous Applications', *Ind. Eng. Chem. Res.*, **38**, 34–48 (1999).
904. C. Viklund, F. Svec, J.M.J. Fréchet, K. Irgum, ' ''Molded'' Porous Materials with High Flow Characteristics for Separation or Catalysis: Control of Porous Properties During Polymerization in Bulk Solution', *Chem. Mater.*, **8**, 744–750 (1996).
905. F. Svec, J.M.J. Fréchet, 'New designs of Macroporous Polymers and Supports: from Separation to Biocatalysis', *Science*, **273**, 205–211 (1996).

906. R.E. Moore, L. Licklider, D. Schumann, T.D. Lee, 'A Microscale Electrospray Interface Incorporating a Monolithic, Poly(styrene–Divinylbenzene) Support for On-line Liquid Chromatography/Tandem Mass Spectrometry Analysis of Peptides and Proteins', *Anal. Chem.*, **70**, 4879–4884 (1998).
907. E.C. Peters, K. Lewandowski, M. Petro, F. Svec, J.M.J. Fréchet, 'Chiral Electrochromatography with a ''Moulded'' Rigid Monolithic Capillary Column', *Anal. Commun.*, **35**, 83–86 (1998).
908. S. Xie, R.W. Allington, F. Svec, J.M.J. Fréchet, 'Rapid Reversed-phase Separation of Proteins and Peptides Using Optimized ''Molded'' Monolithic Polystyrene–Divinylbenzene Columns', *J. Chromatogr. A*, **865**, 169–174 (1999).
909. H. Minakuchi, K. Nakanishi, N. Goga, N. Ishizuka, N. Tanaka, 'Octadecylsilylated Porous Silica Rods as Separation Media for Reversed-phase Liquid Chromatography', *Anal. Chem.*, **68**, 3498–3501 (1996).
910. K. Cabrera, G. Wieland, D. Lubda, K. Nakanishi, N. Soga, H. Minakuchi, K.K. Unger, 'SilicaROD – a New Challenge in Fast High-performance Liquid Chromatography Separations', *Trends Anal. Chem.*, **17**, 50–53 (1998).
911. H. Minakuchi, K. Nakanishi, N. Soga, N. Ishizuka, N. Tanaka, 'Effect of Skeleton Size on the Performance of Octadecylsilylated Continuous Porous Silica Columns in Reversed-phase Liquid Chromatography', *J. Chromatogr. A*, **762**, 135–146 (1997).
912. N. Monji, A.S. Hoffman, 'Novel Immunoassay System and Bio-separation Process Based on Thermal Phase Separating Polymers', *Appl. Biochem. Biotechnol.*, **14**, 107–120 (1987).
913. N.L. Eremeev, A.V. Kukhtin, 'Stimuli-sensitive Hydrogel Material for Biosensor-chemical Trigger', *Anal. Chim. Acta*, **347**, 27–34 (1997).
914. Y.Z. Li, N. He, X. Wang, W.B. Chang, Y.X. Ci, 'Mimicry of Peroxidase by Immobilization of Haemin on *N*-Isopropylacrylamide-based Hydrogel', *Analyst*, **123**, 359–364 (1998).
915. T. Tatsuma, K. Saito, N. Oyama, 'Enzyme Electrodes Mediated by a Thermoshrinking Redox Polymer', *Anal. Chem.*, **66**, 1002–1006 (1994).
916. H. Kanazawa, Y. Kashiwase, K. Yamamoto, Y. Matsushima, A. Kikuchi, Y. Sakurai, T. Okano, 'Temperature-responsive Liquid Chromatography. 2. Effects of Hydrophobic Groups in *N*-Isopropylacrylamide Copolymer-modified Silica', *Anal. Chem.*, **69**, 823–830 (1997).
917. H. Kanazawa, K. Yamamoto, Y. Kashiwase, Y. Matsushima, N. Takai, A. Kikuchi, Y. Sakurai, T. Okano, 'Analysis of Peptides and Proteins by Temperature-responsive Chromatographic System Using *N*-Isopropylacrylamide Polymer-modified Columns', *J. Pharm. Biomed. Anal.*, **15**, 1545–1550 (1997).
918. H. Kanazawa, Y. Matsushima, T. Okano, 'Temperature-responsive Chromatography', *Trends Anal. Chem.*, **17**, 435–440 (1998).
919. C.M. Zeng, J.L. Liao, K.I. Nakazato, S. Hjertén, 'Hydrophobic-interaction Chromatography of Proteins on Continuous Beds Derivatized with Isopropyl Groups', *J. Chromatogr. A*, **753**, 227–234 (1996).
920. A.E. Ivanov, L.S. Zhigis, E.V. Kurganova, V.P. Zubov, 'Effect of Temperature upon the Chromatography of Proteins on Porous Glass, Chemically Coated with *N*-Isopropylacrylamide Copolymer', *J. Chromatogr. A*, **776**, 75–80 (1997).
921. L. Pauling, 'A Theory of the Structure and Process of Formation of Antibodies', *J. Am. Chem. Soc.*, **62**, 2643–2657 (1940).
922. G. Wulff, 'Molecular Recognition in Polymers Prepared by Imprinting with Templates', *ACS Symp. Ser.*, **308**, 186–230 (1986).
923. M. Kempe, K. Mosbach, 'Molecular Imprinting Used for Chiral Separations', *J. Chromatogr. A*, **694**, 3–13 (1995).
924. S. Rimmer, 'Synthesis of Molecular Imprinted Polymer Networks', *Chromatographia*, **47**, 470–474 (1998).
925. O. Ramström, L. Ye, M. Krook, K. Mosbach, 'Applications of Molecularly Imprinted Materials as Selective Adsorbents: Emphasis on Enzymatic Equilibrium Shifting and Library Screening', *Chromatographia*, **47**, 465–469 (1998).
926. L. Schweitz, L.I. Andersson, S. Nilsson, 'Molecular Imprint-based Stationary Phases for Capillary Electrochromatography', *J. Chromatogr. A*, **817**, 5–13 (1998).
927. P.K. Owens, L. Karlsson, E.S.M. Lutz, L.I. Anderson, 'Molecular Imprinting for Bio- and Pharmaceutical Analysis', *Trends Anal. Chem.*, **18**, 146–154 (1999).
928. K. Ensing, T. de Boer, 'Tailor-made Materials for Tailor-made Applications: Application of Molecular Imprints in Chemical Analysis', *Trends Anal. Chem.*, **18**, 138–145 (1999).
929. B. Sellergren, 'Polymer- and Template-related Factors Influencing the Efficiency in Molecularly Imprinted Solid-phase Extractions', *Trends Anal. Chem.*, **18**, 164–174 (1999).
930. T. Takeuchi, D. Fukuma, J. Matsui, 'Combinatorial Molecular Imprinting: an Approach to Synthetic Polymer Receptors', *Anal. Chem.*, **71**, 285–290 (1999).
931. J. Haginaka, 'Drug Determination in Serum by Liquid Chromatography with Restricted-access Stationary Phases', *Trends Anal. Chem.*, **10**, 17–22 (1991).
932. K.S. Boos, A. Rudolphi, S. Vielhauer, A. Walfort, D. Lubda, F. Eisenbeiss, 'Alkyl-diol Silica (ADS) – Restricted Access Precolumn Packings for Direct-injection and Coupled-column Chromatography of Biofluids', *Fresenius J. Anal. Chem.*, **352**, 684–690 (1995).

Infrared Spectroscopy of Biological Applications

B.H. Stuart
University of Technology, Sydney, Australia

1 Introduction **1**
2 Sampling Methods **2**
2.1 Transmission Methods 2
2.2 Reflectance Methods 5
2.3 Microsampling Methods 7
2.4 Gas Chromatography/Infrared 8
2.5 Temperature Cells 8
3 Spectrum Interpretation **8**
3.1 Group Frequencies 8
3.2 Complicating Factors 9
3.3 Identification of Unknown Materials 12
4 Quantitative Analysis **12**
4.1 Spectrum Manipulation 12
4.2 Determination of Concentration 15
4.3 Simple Analysis 16
4.4 Multicomponent Analysis 16
5 Applications **18**
5.1 Proteins and Peptides 18
5.2 Lipids 23
5.3 Biomembranes 25
5.4 Animal Tissue 26
5.5 Carbohydrates 27
5.6 Pharmaceuticals 28
5.7 Foods 29
5.8 Plant Material 29
Abbreviations and Acronyms **30**
Related Articles **31**
References **31**

Infrared (IR) spectroscopy has proved to be a powerful tool for studying biological molecules and the application of this technique to biological problems is continually expanding, particularly with the advent of Fourier transform infrared (FTIR) spectroscopy in recent decades. One of the great advantages of IR spectroscopy is that virtually any sample in virtually any state can be studied. Liquids, solutions, pastes, powders, films, fibers, gases and surfaces can all be examined with a judicious choice of sampling technique. Biological systems such as proteins, peptides, lipids, biomembranes, carbohydrates, pharmaceuticals, foods and both plant and animal tissues have all been successfully characterized using IR spectroscopy.[1–4]

IR spectrometers have been commercially available since the 1940s. At that time the instruments relied on prisms to act as dispersive elements, but by the mid 1950s, diffraction gratings had been introduced into dispersive machines. The most significant advances in IR spectroscopy, however, have come about with the introduction of Fourier transform spectrometers. This type of instrument employs an interferometer and exploits the well established mathematical process of Fourier transformation. FTIR spectroscopy has dramatically improved the quality of IR spectra and minimized the time required to obtain data. Also, with improvements to computers in recent years, IR spectroscopy has made great strides.

1 INTRODUCTION

In recent years FTIR spectroscopy has found increasing favor in laboratories.[2] This more recent method is based on the old idea of the interference of radiation between two beams to yield an interferogram. An interferogram is a signal produced as a function of the change of pathlength between the two beams. The two domains of distance and frequency are interconvertible by the mathematical method of Fourier transformation. Although the basic optical component of FTIR instruments, the Michelson interferometer, has been known for almost a century, it was not until advances in computing that the technique could be successfully applied. The basic components of a FTIR spectrometer are shown schematically in Figure 1. The radiation emerging from the source is passed through an interferometer to the sample before reaching a detector. Upon amplification of the signal, in which high frequency contributions have been eliminated by a filter, the data are converted to a digital form by an analog-to-digital converter and transferred to the computer for Fourier transformation.

The most common interferometer used is a Michelson interferometer, and consists of two perpendicularly plane mirrors, one of which can travel in a direction perpendicular to the plane (Figure 2). A semi-reflecting film, the beam splitter, bisects the planes of these two mirrors. If a collimated beam of monochromatic radiation of wavelength λ cm is passed into an ideal beam splitter, 50% of the incident radiation will be reflected to one of the mirrors and 50% will be transmitted to the other mirror. The two beams are reflected from these mirrors, returning to the beam splitter where they recombine and interfere. Fifty percent of the beam reflected from the fixed mirror is transmitted through the beam splitter and 50% is reflected back in the direction of the source. The beam which emerges from the interferometer at 90° to the input beam is called the transmitted beam and this is the beam detected in FTIR spectrometry.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

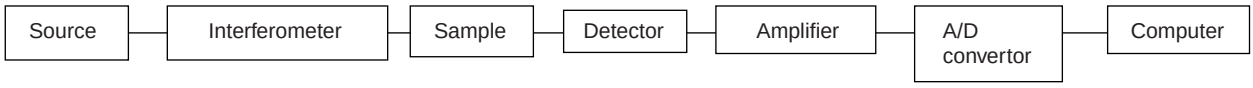

Figure 1 The components of a FTIR spectrometer.

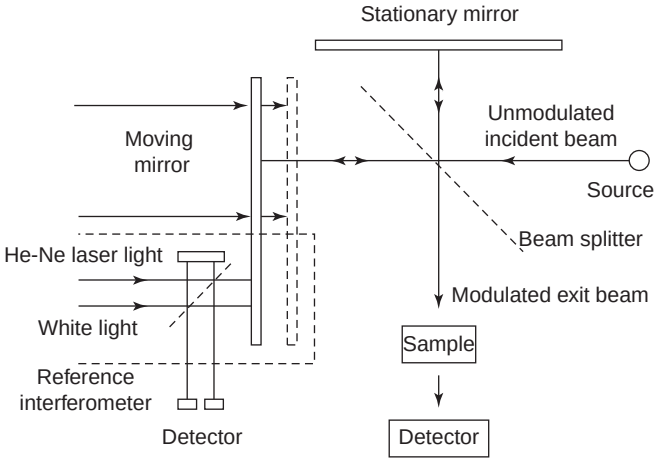

Figure 2 A Michelson interferometer.

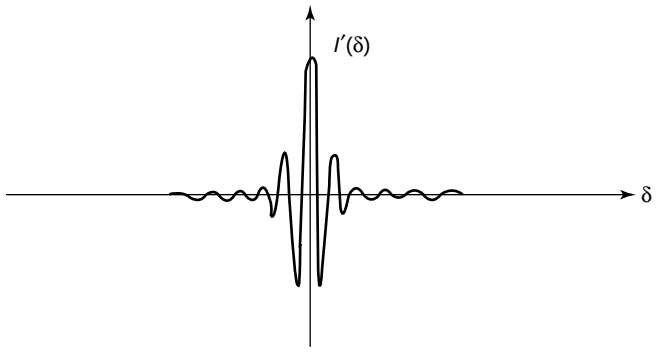

Figure 3 Interferograms for polychromatic radiation.

The moving mirror produces an optical path difference between the two arms of the interferometer. For path differences of $(n + 1/2)\lambda$, the two beams interfere destructively in the case of the transmitted beam and constructively in the case of the reflected beam. The resultant interference pattern is shown in Figure 3 for a source of polychromatic radiation. The former is a simple cosine function but the latter is of a more complicated form because it contains all the spectral information of the radiation falling on the detector. The essential experiment to obtain a FTIR spectrum is to produce an interferogram with and without a sample in the beam and transforming the interferograms into spectra of (a) the source with sample absorptions and (b) the source without sample absorptions. The ratio of the former and the latter corresponds to a double beam dispersive spectrum.

2 SAMPLING METHODS

IR spectroscopy is a versatile experimental technique and it is relatively easy to obtain spectra from biological samples in solution or the solid state.[1] In this section we shall study how samples can be introduced into the instrument, the equipment required to obtain spectra and the pretreatment of samples will be discussed.

2.1 Transmission Methods

Transmission spectroscopy is the oldest and most basic IR method. The method is based upon the absorption of IR radiation at specific wavelengths as it passes through a sample. Liquid or solid forms can be analyzed using this technique.

There are several different types of transmission solution cells available. Fixed pathlength sealed cells are useful for volatile liquids, but cannot be taken apart for cleaning. Semi-permanent cells are demountable so that the windows can be cleaned. A semi-permanent cell is illustrated in Figure 4. The spacer is usually made of polytetrafluoroethylene (PTFE, Teflon®) and is available in a variety of thicknesses, allowing one cell to be used for various pathlengths. Variable pathlength cells incorporate a mechanism for continuously adjusting the pathlength and a vernier scale allows accurate adjustment. All these cell types are filled using a syringe and the syringe ports are sealed with PTFE plugs before sampling. An important consideration in the choice of IR cells is the type of window material. The material

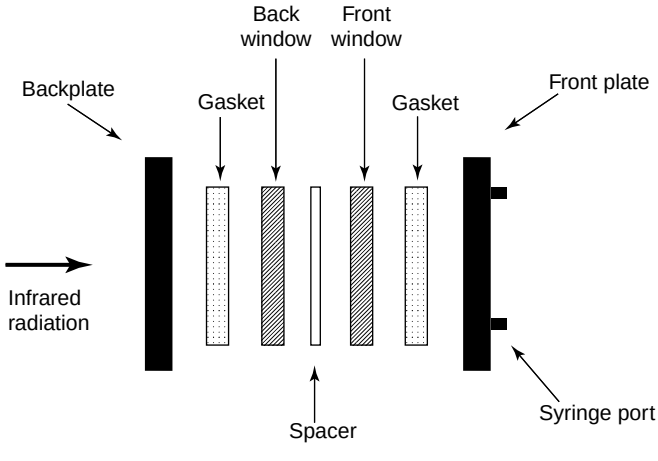

Figure 4 A semi-permanent cell.

Table 1 Summary of some optical materials used in transmission IR spectroscopy

Window material	Useful range (cm^{-1})	Refractive index	Properties
NaCl	40 000–600	1.5	Soluble in water; slightly soluble in alcohol; low cost; fair resistance to mechanical and thermal shock; easily polished
KBr	43 500–400	1.5	Soluble in water and alcohol; slightly soluble in ether; hygroscopic; good resistance to mechanical and thermal shock
CaF_2	77 000–900	1.4	Insoluble in water; resists most acids and bases; does not fog; useful for high pressure work
BaF_2	66 666–800	1.5	Insoluble in water; soluble in acids and NH_4Cl; does not fog; sensitive to thermal and mechanical shock
KCl	33 000–400	1.5	Similar properties to NaCl but less soluble; hygroscopic
CsBr	42 000–250	1.7	Soluble in water and acids; hygroscopic
CsI	42 000–200	1.7	Soluble in water and alcohol; hygroscopic

must be transparent to the incident IR radiation and normally alkali halides are used in transmission methods. The cheapest material is sodium chloride (NaCl), but other commonly used materials are listed in Table 1. Liquid films provide a quick method of examining liquid samples. A drop of liquid may be sandwiched between two IR plates which are then mounted in a cell-holder. The type of spectrum expected from a liquid film is illustrated in Figure 5, which shows the spectrum of chloroform, a solvent commonly used for biological samples.

Before producing an IR sample in solution, a suitable solvent must be chosen. In choosing a solvent there is a number of factors to be considered. A solvent must obviously be able to dissolve the compound. It should also be as nonpolar as possible to minimize solute–solvent interactions. In addition, the solvent should not strongly absorb IR radiation. If quantitative analysis is required it is necessary to use a cell of known pathlength. A guide to pathlength selection for different solution concentrations is shown in Table 2.

Water is a commonly used solvent for biological samples and NaCl cannot be used as a IR window material as it is very soluble in water. Small pathlengths (~0.010 mm) are available and help reduce the intensity of the very strong IR modes produced in the water spectrum. The small pathlength also produces a small sample cavity, allowing samples in milligram quantities to be examined. Certain difficulties arise when using water as a solvent in IR spectroscopy. The IR modes of water are very intense and may overlap with the sample modes of interest. This problem may be overcome by substituting water with deuterium oxide (D_2O). The IR modes of D_2O occur at different frequencies to those observed for water because of the mass dependence of the vibrational frequency. Table 3 lists the characteristic bands observed for both H_2O and D_2O.

There are three general methods for examining solid samples in transmission IR spectroscopy: alkali halide discs, mulls and films. A choice of method depends very much on the nature of the sample to be examined. The use of alkali halide discs involves mixing a solid sample (a few mg) with a dry alkali halide powder (100–200 mg). The mixture is usually ground with an agate mortar and pestle and subjected to a pressure of about 10 ton in^{-2} (1.575×10^5 kg m^{-2}) in an evacuated die. This sinters the mixture and produces a clear transparent disc. The most commonly used alkali halide is potassium bromide (KBr), which is completely transparent in the mid-IR region.

Certain factors need to be considered when preparing alkali halide discs. The ratio of the sample to alkali halide is important. Surprisingly little sample is needed and approximately 2 to 3 mg of sample with about 200 mg of halide should be used. The disc must not be too thick or too thin. Thin discs are fragile and difficult to handle, while thick discs transmit too little radiation. A disc of about 1 cm diameter made from about 200 mg of material usually results in a good thickness of about 1 mm. If the crystal size of the sample is too large excessive scattering of radiation results, particularly so at high wavenumbers. The crystal size must be reduced, normally by grinding the solid using a mortar and pestle. If the alkali halide is not perfectly dry this results in the appearance of bands due to water. Contributions due to water are difficult to avoid, and the alkali halide should be kept desiccated and warm prior to use.

The preparation of mulls for solid samples involves grinding the sample then suspending it (about 50 mg) in 1–2 drops of a mulling agent. This is followed by further grinding until a smooth paste is obtained. Ideally a mulling agent should be IR transparent between 4000 and 600 cm^{-1}, but no such agent exists. The most commonly used mulling agent is liquid paraffin (Nujol®), and its

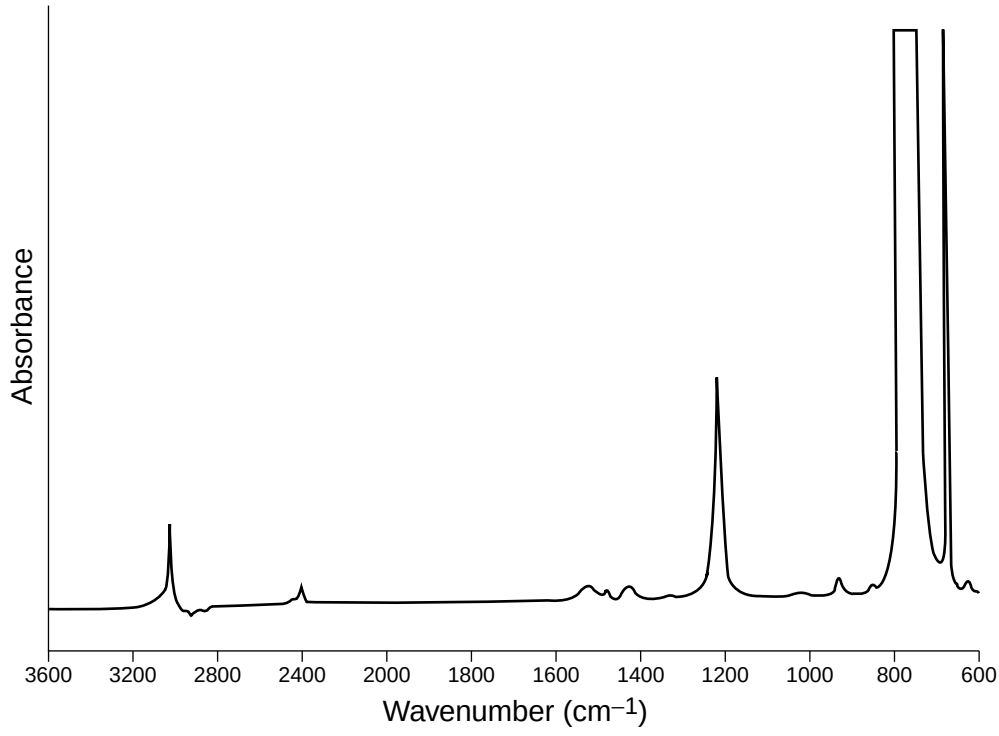

Figure 5 The IR spectrum of chloroform.

Table 2 Pathlength selection for solution cells

Concentration (%)	Pathlength (mm)
>10	0.05
1–10	0.1
0.1–1	0.2
<0.1	>0.5

Table 3 The major IR bands of water and deuterium oxide

Frequency (cm^{-1})	Assignment
3920	O–H stretching
3490	O–H stretching
3280	O–H stretching
1645	H–O–H bending
2900	O–D stretching
2540	O–D stretching
2450	O–D stretching
1215	D–O–D bending

spectrum is shown in Figure 6. Although the mull method is quick and easy, there are some points to consider. The ratio of the sample to mulling agent must be appropriate as if there is too little sample and there will be no sign of the sample in the spectrum. Too much sample and a thick paste will be produced and no radiation will be transmitted. A rough guide to mull preparation is to use a micro spatula tip of sample to 2–3 drops of mulling agent. If the crystal size of the sample is too large this leads to scattering of radiation, which gets worse at the high frequency end of the spectrum. If the mull is not spread over the whole plate area, the beam of radiation passes part through the mull and part through only the plates, producing a distorted spectrum. Care should also be taken not to use too much or too little mull between the IR plates. Too little leads to a very weak spectrum showing only the strongest absorption bands. Too much mull leads to poor transmission of radiation so that the baseline may be at 50% or less. It is sometimes possible to reduce the energy of a reference beam to a similar extent by use of an attenuator. Beam attenuators work somewhat like a venetian blind, illustrated by Figure 7.

Films can be produced by either solvent casting or by melt casting. In solvent casting the sample is dissolved in an appropriate solvent (the concentration depends on the required film thickness). A solvent must be chosen which not only dissolves the sample, but will also produce a uniform film. The solution is poured onto a leveled glass plate (such as a microscope slide) or a metal plate and spread to uniform thickness. The solvent may then be evaporated in an oven and, once dry, the film can be stripped from the plate. However, care must be taken

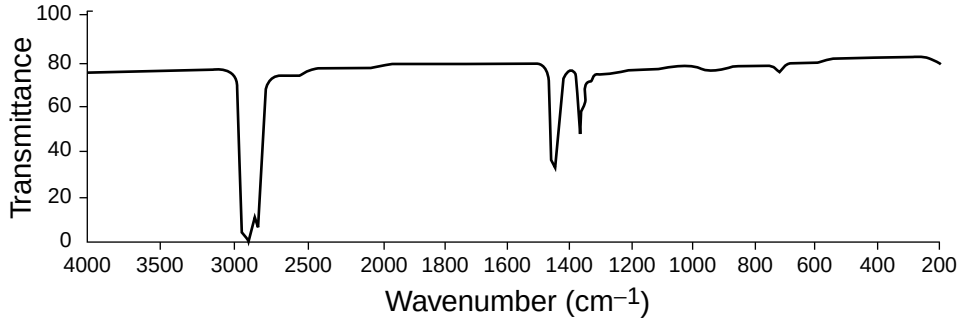

Figure 6 The IR spectrum of liquid paraffin (Nujol®).

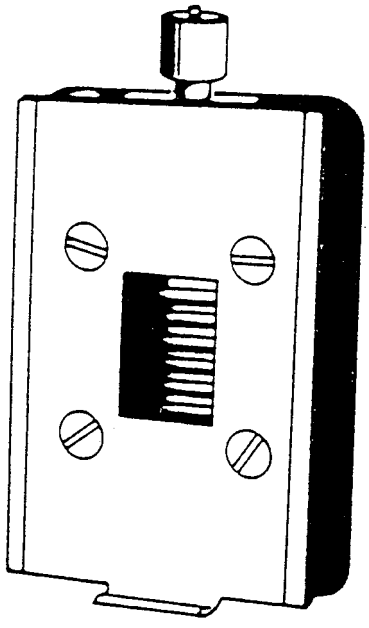

Figure 7 A beam attenuator.

as heating biological samples may cause degradation. Alternatively, it is possible to cast a film straight onto the IR window to be used. Solid samples which melt at relatively low temperatures without decomposition can be prepared by melt casting. A film is prepared by hot pressing the sample in a hydraulic press between heated metal plates.

2.2 Reflectance Methods

Reflectance techniques can be used for samples which are difficult to analyze by the normal transmittance methods. Reflectance methods can be divided into two categories. Internal reflectance measurements can be made using an attenuated total reflectance (ATR) cell in contact with the sample. There are also external reflectance measurements which involve an IR beam reflected directly from the sample surface.

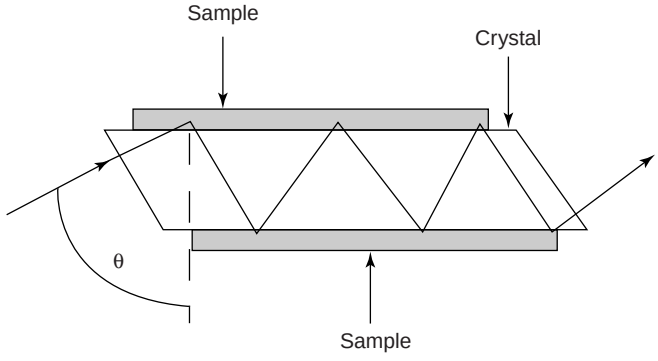

Figure 8 An ATR cell.

ATR spectroscopy utilizes the phenomenon of total internal reflection (Figure 8).[5] A beam of radiation entering a crystal will undergo total internal reflection when the angle of incidence at the interface between the sample and crystal is greater than the critical angle. The critical angle is a function of the refractive indices of the two surfaces. The beam penetrates a fraction of a wavelength beyond the reflecting surface and when a material which selectively absorbs radiation is in close contact with the reflecting surface, the beam loses energy at the wavelength where the material absorbs. The resultant attenuated radiation is measured and plotted as a function of wavelength by the spectrometer and gives rise to the absorption spectral characteristics of the sample. The depth of penetration in ATR is a function of wavelength (λ), the refractive index of the crystal and the angle of incident radiation (θ). The depth of penetration, d_p, for a nonabsorbing medium is given by Equation (1):

$$d_p = \frac{\lambda/n_1}{\{2\pi[\sin\theta - (n_1/n_2)^2]^{1/2}\}} \quad (1)$$

where n_1 is the refractive index of the sample and n_2 is the refractive index of the ATR crystal. The crystals used in ATR cells are made from materials which have low solubility in water and are of very high refractive index. Such materials include zinc selenide (ZnSe), germanium

Table 4 Materials used as ATR crystals and their properties

Window material	Useful range (cm^{-1})	Refractive index	Properties
KRS-5 (thallium/ iodide)	17 000–250	2.4	Soluble in bases; slightly soluble in water; insoluble in acids; soft; highly toxic (handle with gloves)
ZnSe	20 000–500	2.4	Insoluble in water, organic solvents, dilute acids and bases
Ge	5000–550	4.0	Insoluble in water; very brittle

(Ge) and thallium/iodide (KRS-5). The properties of these commonly used materials for ATR crystals are summarized in Table 4. Different designs of ATR cells allow both liquid and solid samples to be examined. It is also possible to set-up a flow-through ATR cell by including an inlet and outlet into the apparatus. This allows for continuous flow of solutions through the cell and permits spectral changes with time to be monitored. Multiple internal reflectance (MIR) and ATR are similar techniques, but MIR produces more intense spectra from multiple reflections. While a prism is usually used in ATR work, MIR uses specially shaped crystals which cause many internal reflections, typically 25 or more (Figure 9).

In external reflectance incident radiation is focused onto the sample and two forms of reflectance can occur: specular and diffuse. External reflectance measures the radiation reflected from a surface. The material

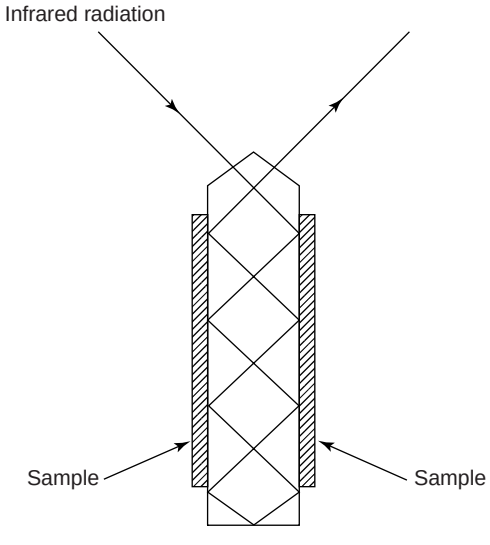

Figure 9 A MIR cell.

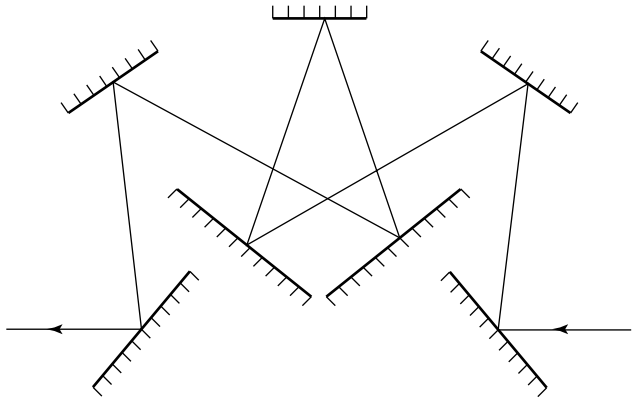

Figure 10 Specular reflectance.

must, therefore, be reflective or be attached to a reflective backing. A particularly useful application for this technique is the study of surfaces.

Specular reflectance occurs when the reflected angle of incident radiation equals the angle of incidence (Figure 10). The amount of light reflected depends on the angle of incidence, the refractive index, surface roughness and absorption properties of the sample. For most materials the reflected energy is only 5–10%, but in regions of strong absorptions the reflected intensity is greater. The resultant data appears different from normal transmission spectra, as derivative-like bands result from the superposition of the normal extinction coefficient spectrum with the refractive index dispersion (based upon Fresnel's relations from physics). However, the reflectance spectrum can be corrected using a Kramers–Kronig transformation (or K–K transformation). The corrected spectrum appears like the familiar transmission spectrum.

Increased pathlengths through thin coatings can be achieved using grazing angles of incidence (up to 85°). Grazing angle sampling accessories allow measurements to be made on samples over a wide range of angles of incidence. Solid samples, particularly coatings on reflective surfaces, are simply placed on a flat surface. The technique is also commonly used for liquid samples which can be poured into a PTFE trough. Oriented films on the liquid surface can be investigated using this material. The grazing angle approach can be used to study surfactants, proteins, steroids and phospholipids in biological membranes.

In external reflectance, the energy which penetrates one or more particles is reflected in all directions. This component is called diffuse reflectance. In the diffuse reflectance infrared technique, commonly called (DRIFT), a powdered sample is mixed with KBr powder. The DRIFT cell reflects radiation to the powder and collects the energy reflected back over a large angle.

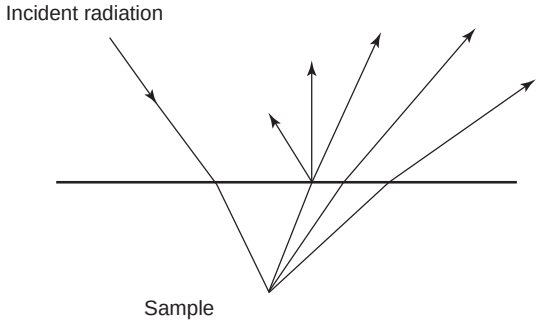

Figure 11 Diffuse reflectance.

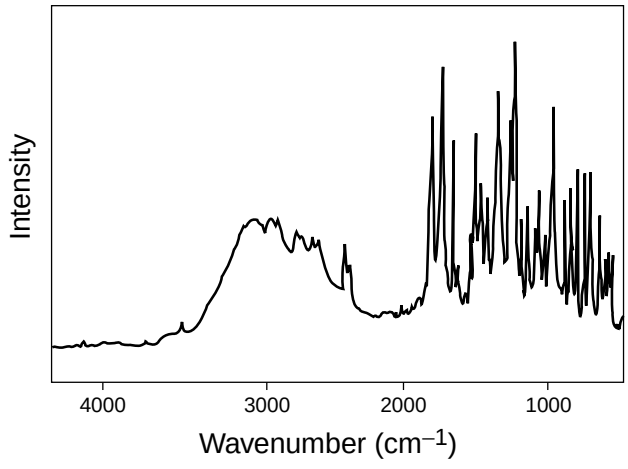

Figure 12 The diffuse reflectance spectrum of aspirin.

Diffusely scattered light can be collected directly from a sample or, alternatively, using an abrasive sampling pad. DRIFT is particularly useful for sampling powders or fibers. Figure 11 shows diffuse reflectance from the surface of a sample. Kubelka and Munk[6] developed a theory describing the diffuse reflectance process for powdered samples which relates the sample concentration to the scattered radiation intensity. The Kubelka–Munk equation is Equation (2):

$$\frac{(1-R_\infty)^2}{2R_\infty} = \frac{c}{k} \tag{2}$$

where R_∞ is the absolute reflectance of the layer, c is the concentration and k is the molar absorption coefficient. Figure 12 shows the diffuse reflectance spectrum of aspirin. An alternative relationship between the concentration and the reflected intensity is now widely used in near-IR diffuse reflectance spectroscopy. The relationship is analogous to the Beer–Lambert law (Equation 3):

$$\log\left(\frac{1}{R}\right. = k'c \tag{3}$$

where k' is a constant.

2.3 Microsampling Methods

It is possible to combine an IR spectrometer with a microscope facility in order to study very small quantities.[7] In recent years there have been considerable advances in FTIR microscopy with samples as small as 25 μm being characterized. In FTIR microscopy, the microscope sits above the FTIR sampling compartment. IR radiation from the spectrometer is focused onto a sample placed on a standard microscope $x-y$ stage. After passing through the sample, the IR beam is collected by a Cassegrain objective which produces an image of the sample within the barrel of the microscope. A variable aperture is placed in this image plane. The radiation is then focused on a small area mercury cadmium telluride detector by another Cassegrain condenser. The microscope also contains glass objectives to allow visual inspection of the sample. In addition, by switching mirrors in the optical train, the microscope can be converted from transmission mode to reflectance mode.

If a microscope facility is not available, there are other special sampling accessories available to allow examination of microgram or microlitre amounts. This is accomplished using a beam condenser so that as much as possible of the beam passes through the sample. Microcells are available with volumes of around 4 μL and pathlengths up to 1 mm.

There are a number of microsampling cells available. A diamond anvil cell (DAC) uses two diamonds to compress a sample to a thickness suitable for measurement and increase the surface area. This technique can be used at normal atmospheric pressures, but it may also be applied to study samples under high pressures and improve the quality of the spectrum of trace samples.

A MIR cell can also be used as this technique can produce strong spectra. The ability to obtain spectra from trace amounts of materials is illustrated in Figure 13,

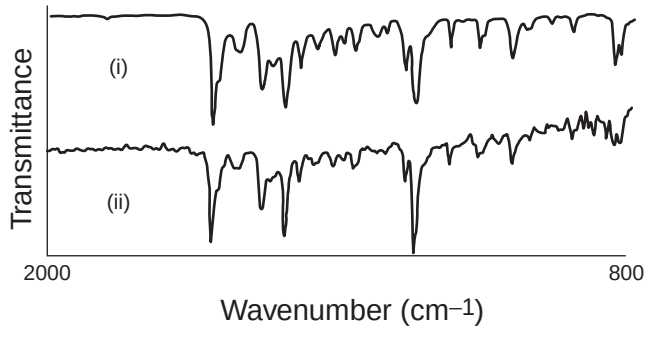

Figure 13 The IR spectra of phenacetin deposited on a micro MIR crystal: (i) 1 μg sample, spectrum recorded in 30 s; (ii) 50 ng sample, spectrum recorded in 2 min (resolution of 4 cm^{-1} in both cases). (Reproduced with permission from a Perkin-Elmer sales leaflet.)

which shows the IR spectra of the pain relieving drug phenacetin deposited onto a micro MIR crystal.

2.4 Gas Chromatography/Infrared

IR spectroscopy has been combined with various other analytical techniques. Gas chromatography/infrared spectroscopy (GC/IR) allows the identification of the components eluting from gas chromatography (GC).[8] GC/IR has certain advantages over, say, gas chromatography/mass spectrometry (GC/MS). While GC/MS is able to distinguish easily between compounds of different mass, it is unable to differentiate structural isomers of the same molecular mass. By comparison, GC/IR can easily distinguish isomers.

Figure 14 shows a schematic diagram of a typical for combining system GC with FTIR. The apparatus consists of a FTIR spectrometer, a gas chromatograph, light pipe, detector and a computer for analysis. The light pipe interfaces the GC with the spectrometer. The pipe is usually made of quartz and should be approximately one-half the width of the GC peaks for the best compromise between chromatographic resolution and IR sensitivity. For capillary column GC, the light pipe usually has a 1 mm inner diameter and for packed columns a light pipe of inner diameter 2–3 mm is used.

The nature of this technique requires that interferograms are collected in short time intervals. Gram–Schmidt chromatograms are created from the interferogram intensities and reflect the IR absorption. The Gram–Schmidt traces are a plot of IR absorbance versus time. The absorbance spectra of the eluted species are calculated by performing a Fourier transformation on the interferogram.

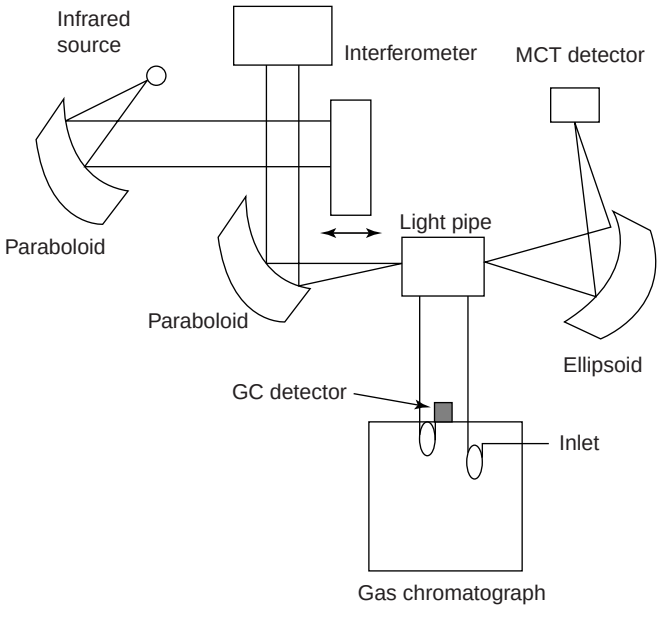

Figure 14 A combined GC and FTIR system.

IR spectroscopy has been combined with other established analytical techniques. For instance, with thermogravimetric analysis (TGA) which is a technique which involves measuring the change of mass of a sample when it is heated. While TGA can provide quantitative information about a decomposition process, it is unable to identify the decomposition products. However, TGA and IR have been combined to provide a complete qualitative and quantitative characterization of thermal decomposition processes.

2.5 Temperature Cells

Variable temperature cells can be obtained which are controlled to 0.1 °C in the range −180 to 250 °C. An electrical heating system is used for temperatures above ambient, and liquid nitrogen with a heater for low temperatures. These cells can be used to study phase transitions and the kinetics of reactions. As well as transmission temperature cells, variable temperature ATR cells and cells for microsampling are available.

3 SPECTRUM INTERPRETATION

Spectrum interpretation is simplified by the fact that the bands which appear can be assigned to particular parts of the molecule, producing what are known as group frequencies. The mid-IR region is normally treated as four regions, depending upon the type of group frequency. However, of certain factors which complicate IR spectra. Phenomena such as overtone and combination bands, Fermi resonance and hydrogen bonding can introduce additional, and sometimes misleading, information into the spectrum. It is important to be aware of these factors before tackling the interpretation of a given spectrum.

3.1 Group Frequencies

The mid-IR spectrum can be divided into four regions and the nature of a group frequency may generally be determined by the region in which it is located. These regions are defined as follows:

1. X−H stretching region (4000–2500 cm^{-1})
2. triple bond region (2500–2000 cm^{-1})
3. double bond region (2000–1500 cm^{-1})
4. fingerprint region (1500–600 cm^{-1}).

All the fundamental vibrations in the region 4000–2500 cm^{-1} may be attributed to X−H stretching. O−H stretching produces a broad band which occurs in the

range 3700–3600 cm^{-1}. By comparison, N–H stretching is usually observed between 3400 and 3300 cm^{-1}. This absorption is generally much sharper than O–H stretching and may therefore be differentiated. Compounds containing the NH_2 group usually show a doublet structure, while secondary amines show one sharp band. C–H stretching bands from aliphatic compounds occur in the range 3000–2850 cm^{-1}. They are moderately broad and show medium intensity. It is also possible to resolve asymmetrical and symmetrical C–H stretching absorptions of a CH_3 group which usually occur at about 2965 and 2880 cm^{-1}, respectively, while the corresponding absorptions for a CH_2 group occur at 2930 and 2860 cm^{-1}. If the C–H bond is adjacent to a double bond or aromatic ring, the C–H stretching frequency increases and absorbs between 3100 and 3000 cm^{-1}. This is useful in distinguishing purely aliphatic compounds, but must be used with care as for many compounds containing a small number of aromatic hydrogens and many aliphatic C–H bonds the peaks above 3000 cm^{-1} may only appear as a shoulder on the stronger aliphatic absorption and may be obscured. However, evidence for the presence of an aromatic ring may be obtained by examination of other regions of the spectrum. Deuterated compounds would be expected to show C–D stretching at a factor of 0.73 less due to the change in atomic mass. That is, at about 2130 cm^{-1}.

Triple bond stretching absorptions fall in the 2500–2000 cm^{-1} region because of the high force constants of the bonds. C≡C bonds absorb between 2300 and 2050 cm^{-1}, while the nitrile group (C≡N) occurs between 2300 and 2200 cm^{-1}. These groups may be distinguished since C≡C stretching is normally very weak, while C≡N stretching is of medium intensity. These are the most common absorptions in this region, but you may come across some X–H stretching absorptions, where X is a more massive atom like phosphorus or silicon. These absorptions usually occur near 2400 and 2200 cm^{-1}, respectively.

The principal bands in the 2000–1500 cm^{-1} region are due to C=C and C=O stretching. Carbonyl stretching is one of the easiest absorptions to recognize in an IR spectrum. It is usually the most intense band in the spectrum, and depending on the type of C=O bond, occurs in the region 1830–1650 cm^{-1}. Note also that metal carbonyls may absorb above 2000 cm^{-1}. C=C stretching is much weaker and occurs at around 1650 cm^{-1}. This band is often absent for symmetry or dipole moment reasons. C=N stretching also occurs in this region and is usually stronger. The N–H bending vibration in amines occurs between 1630 and 1500 cm^{-1} and is usually strong. Before assigning a band always check the N–H stretching region above 3000 cm^{-1} to avoid possible confusion.

At frequencies of values greater than 1500 cm^{-1} it is generally possible to assign each absorption band in an IR spectrum. This is not so for most absorptions observed below 1500 cm^{-1}. This region is referred to as the fingerprint region, since quite similar molecules give different absorption patterns at these frequencies.

It has been assumed so far that each band in an IR spectrum can be assigned to a particular deformation of the molecule, the movement of a group of atoms or the bending or stretching of a particular bond. This is possible for many bands, particularly stretching vibrations of multiple bonds which are well behaved. However, many vibrations are not so well behaved and may vary by hundreds of wavenumbers even for similar molecules. This applies to most bending and skeletal vibrations, which absorb in the 1500–650 cm^{-1} region, for which small steric or electronic effects in the molecule lead to large shifts. A spectrum of a molecule may have a hundred or more absorption bands present, but there is no need to assign the vast majority. The spectrum can be regarded as a 'fingerprint' of the molecule. Most single bonds absorb at similar frequencies and hence the vibrations couple. The observed pattern will depend on the carbon skeleton, and the resulting bands will originate from the oscillation of large parts of the skeleton, or the skeleton and the attached functional groups. C–C stretching frequencies may also couple with C–H bending vibrations. The C–O stretching frequency is one of the bands that can be useful for identification purposes. If no intense band appears in the fingerprint region, you can usually be sure that no C–O bonds are present. The frequency is rather variable, occurring anywhere between 1400 and 1000 cm^{-1}. Aromatic rings give rise to two bands at 1600 and 1500 cm^{-1}. They are usually sharp, but are of variable intensity and occasionally the band at 1600 cm^{-1} splits into a doublet. Aromatic rings and alkenes give rise to other bands which are perhaps the most useful in this region. They are out-of-plane C–H bending vibrations which occur between 1000 and 700 cm^{-1}.

The information provided by group frequencies can be summarized in what are known as a correlation tables. Figure 15 shows a correlation table for organic molecules. The lines in the correlation table represent the regions of the spectrum where bands appear for the relevent class of molecule.

3.2 Complicating Factors

Although it is a useful assumption to be make, it is not true that all bands in a spectrum can simply be associated with a particular bond or part of a molecule. Certain complicating factors need to be taken into account. One such factor is the presence of overtone and combination bands. Overtone bands in an IR spectrum are multiples of the fundamental absorption frequency. The energy levels for overtones of IR modes are illustrated in Figure 16. The

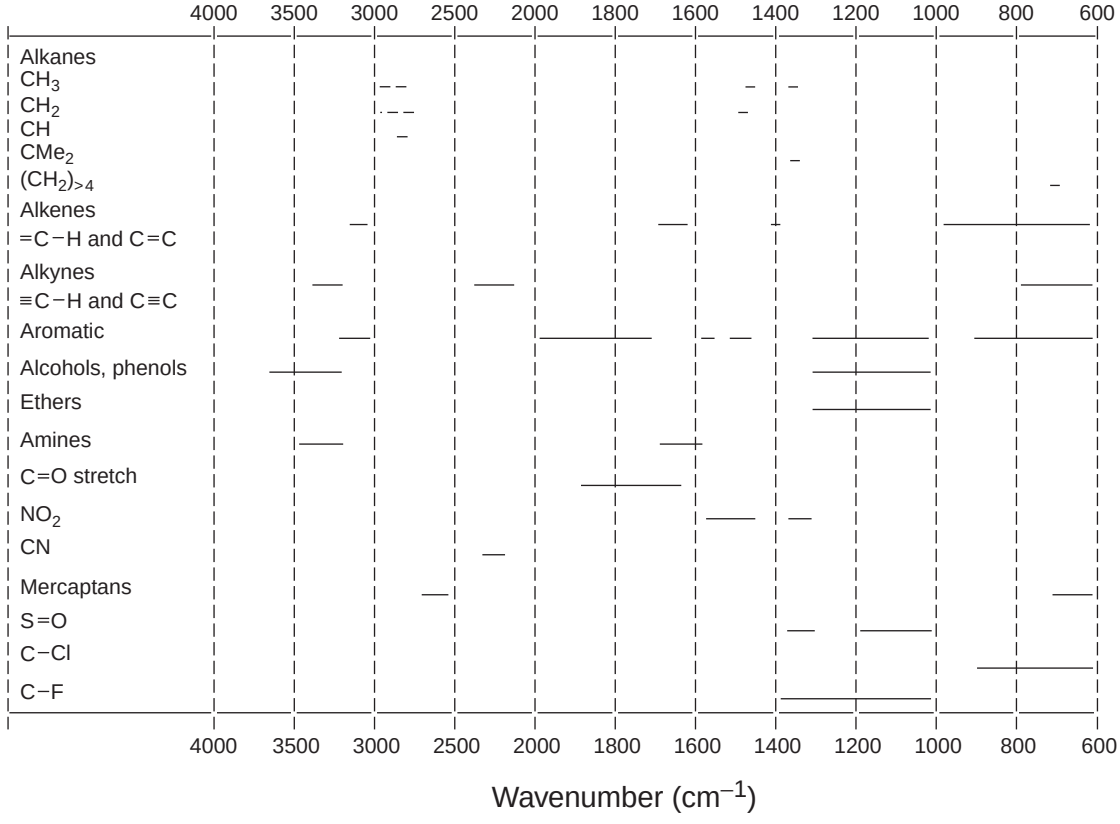

Figure 15 A correlation table for organic molecules.

energy required for the first overtone is twice the fundamental, assuming evenly spaced energy levels. Since the energy is proportional to the frequency absorbed and this is proportional to the wavenumber, the first overtone will appear in the spectrum at twice the wavenumber of the fundamental. Combination bands arise when two fundamental bands absorbing at $\bar{\nu}_1$ and $\bar{\nu}_2$ absorb energy simultaneously. The resulting band will appear at $(\bar{\nu}_1 + \bar{\nu}_2)$ wavenumbers. The absorptions observed in the near-IR region (13 000–4000 cm^{-1}) are overtones or combinations of the fundamental stretching bands that occur in the region 3000–1700 cm^{-1}. The bands involved are usually due to C–H, N–H, O–H, P–H or S–H stretching. The resulting bands in the near-IR are usually weak in intensity and the intensity generally decreases by a factor of 10 from one overtone to the next. The bands in the near-IR are often overlapped, making them less useful than the mid-IR region for qualitative analysis. However, there are important differences between the near-IR positions between functional groups and these differences can often be exploited for quantitative analysis. There are some classes of biological materials which can be successfully analyzed using near-IR spectroscopy, such as foods.

The Fermi resonance effect usually leads to two bands appearing close together when only one is expected.

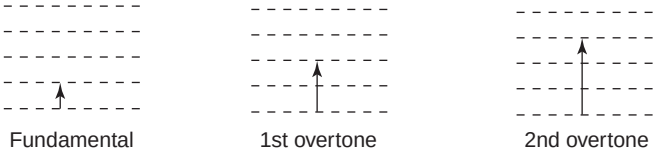

Figure 16 Energy levels for fundamental and overtone IR bands.

When an overtone or a combination band has the same frequency as or a similar frequency to a fundamental, two bands appear split either side of the expected value and are of about equal intensity. The effect is greatest when the frequencies match, but it is also present when there is a mismatch of a few tens of wavenumbers. The two bands are referred to as a Fermi doublet.

The presence of hydrogen bonding is of great importance in biological molecules. For instance, the biological activity of DNA relies on this type of bonding. Hydrogen bonding can be defined as the attraction which occurs between a highly electronegative atom carrying a nonbonded electron pair (such as fluorine, oxygen or nitrogen) and a hydrogen atom, itself bonded to a small highly electronegative atom. An example of this type of bonding is illustrated by the interactions between

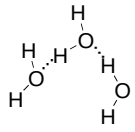

Figure 17 Hydrogen bonding of water molecules.

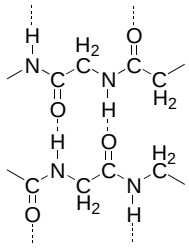

Figure 18 Intramolecular hydrogen bonding in a protein.

water molecules in Figure 17. This is an example of intermolecular hydrogen bonding. It is also possible for a hydrogen bond to form between appropriate groups within the one molecule. This is known as intramolecular hydrogen bonding and is illustrated by the protein structure shown in Figure 18. The segments shown belong to the same protein molecule.

Many solvents are capable of forming hydrogen bonds to solutes. To form a hydrogen bond a proton donor group and an electron donor group are needed. Three classes of solvent exist which may lead to trouble when used for hydrogen bonding studies is solution: compounds which contain hydrogen donor groups (e.g. halogenated compounds which contain a sufficient number of halogens to activate the hydrogens present, such as chloroform); compounds which contain nonbonded electron pairs (e.g. ethers, aldehydes, tertiary amines); compounds which contain both types of group (e.g. water and alcohols). There are only a few solvents which do not have the above characteristics, such as carbon tetrachloride (CCl_4) and carbon disulfide (CS_2). These still contain lone electron pairs but being on S and Cl are less available, and any interaction will therefore be extremely weak.

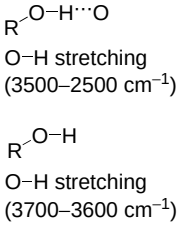

Figure 19 The effect of hydrogen bonding on a O–H stretching vibration.

Hydrogen bonding is a very important effect in IR spectroscopy. In IR spectroscopy the frequencies of vibration of bonds depend on the masses of the atoms in the bond and the bond stiffness. Hydrogen bonding influences the bond stiffness and so alters the frequency of vibration. For example, for a hydrogen bond in an alcohol, the O–H stretching vibration in a hydrogen bonded dimer is observed in the range 3500–2500 cm^{-1}, rather than the usual range 3700–3600 cm^{-1} (Figure 19). Apart from solvent effects, concentration and temperature also affect the degree of hydrogen bonding in a compound. The lower the concentration, the less chance there is of two molecules colliding. It follows that the degree of hydrogen bonding decreases with decreasing concentration. Increasing temperature means that each molecule will have more energy on average and hence weak associative forces, such as hydrogen bonds, are likely to be broken. This should lead to a lesser degree of hydrogen bonding, and thus changes in frequency to greater values would be observed for groups forming the hydrogen bond.

Another issue to be considered is coupling. Vibrations in the skeleton of molecules become coupled. In other words, they are not restricted to one or two bonds, but may involve a large part of the carbon backbone and oxygen or nitrogen atoms if present. The energy levels mix, resulting in the same number of vibrational modes, but at different frequencies, and bands can no longer be assigned to one bond. This is very common and occurs when adjacent bonds have closely similar frequencies. Coupling commonly occurs between C–C stretching, C–O stretching, C–N stretching, C–H rocking and C–H wagging motions. A further requirement is that to be strongly coupled the motions must be in the same part of the molecule. There are three factors which determine whether coupling will take place: the groups must be adjacent; the frequencies must be similar; and the vibrations must be in the same part of the molecule. It turns out that all vibrations which absorb in the region 1500–650 cm^{-1} will couple to an extent that critically depends on the shape of the molecule. Large groups, the presence of double and triple bonds and electronic effects will influence coupling. Hence it is difficult to predict the extent of this coupling and this leads to great difficulties in the assignment of bands in this region of the spectrum. This phenomenon accounts for the wide range of frequencies over which C–C, C–O and C–N stretching vibrations absorb. The most important consequence of coupling is that small changes in structure leads to major changes in the IR spectrum in the 1500–650 cm^{-1} region. It is very unlikely, therefore, that two different molecules will have exactly the same pattern of bands in this region.

3.3 Identification of Unknown Materials

There are a few general rules that can be stated to help in the use of an IR spectrum for the determination of a structure. However, the most effective way to learn is through practice. The following is a suggested strategy for spectrum interpretation:

1. Look first at the high-frequency end of the spectrum ($>1500\,cm^{-1}$) and concentrate initially on the major bands.
2. For each band, short list the possibilities using a correlation chart.
3. Use the lower frequency end of the spectrum for the confirmation or elaboration of possible structural elements.
4. Do not expect to be able to assign every band in the spectrum.
5. Cross-check wherever possible. For example, and aldehyde should absorb near $1730\,cm^{-1}$ and in the region $2900-2700\,cm^{-1}$.
6. Exploit negative evidence as well as positive evidence. For example, if there is no band in the $1850-1600\,cm^{-1}$ region, it is most unlikely that a carbonyl group is present.
7. Band intensities should be treated with some caution. Under some circumstances they may vary considerably for the same functional group.
8. Take care when using small frequency changes. These can be influenced by whether the spectrum was run as a solid or liquid or in solution. Some bands are very solvent sensitive in solutions.
9. Subtract the solvent bands before interpretation – these could be confused with bands from the sample.

Advances in computer retrieval techniques have extended the range of information available from an IR instrument by allowing comparison of an unknown spectrum with a bank of known compounds. This was always possible in the past by manually searching libraries of spectra.[9] Once the compound was classified the atlas of spectra could be searched for an identical or very similar spectrum. The advent of more efficient computers has speeded this search process. Programs widely available work by hunting through stored data to match intensities and frequencies of absorption bands. The computer will then output the best fits and the names of the compounds found. There is a wide range of spectral libraries of particular interest to life scientists. For instance, libraries for the following types of samples are currently available: proteins, peptides, enzymes, sugars, carbohydrates, fatty acids, glycerides, steroids, food additives, drugs and pharmaceuticals.

It may not always be possible by examination of the IR spectrum of a compound alone to identify it unequivocally. It is normal to use IR in conjunction with other techniques, such as chromatographic methods, mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy and the variety of other spectroscopic techniques.

4 QUANTITATIVE ANALYSIS

Quantitative IR spectroscopy suffers certain disadvantages when compared with other analytical techniques and it tends to be confined to specialist applications. However, there are certain applications where it is used because it is cheaper or faster. The technique is often used for the analysis of one component of a mixture, especially when the compounds in the mixture are alike chemically or have very similar physical properties (for example, structural isomers). In these instances analysis using ultraviolet/visible (UV/VIS) spectroscopy is difficult because the spectra of the components will be nearly identical. Chromatographic analysis may be of limited use because separation, of say isomers, is difficult to achieve. The IR spectra of isomers are usually quite different in the fingerprint region. Another advantage of the IR technique is that it is nondestructive and requires only a relatively small amount of sample.

4.1 Spectrum Manipulation

There are a number of techniques available to users of modern IR spectrometers which help with both qualitative and quantitative interpretation of spectra. It is usual in quantitative IR spectroscopy to use a baseline joining the points of lowest absorbance on the peak, preferably in reproducibly flat parts of the absorption line. The absorbance difference between the baseline and the top of the band is then used. A baseline construction is shown in Figure 20.

Noise in a spectrum can be diminished by smoothing. After a spectrum is smoothed it becomes similar to the result of an experiment at a lower resolution. The features are blended into each other and the noise level decreases. A smoothing function is basically a convolution between the spectrum and a vector whose points are determined by the degree of smoothing you choose to apply. Generally, you will be asked to enter a degradation factor, which will be some positive integer. A low value, say one, will produce only subtle changes, while a high value has a more pronounced effect on your spectrum.

The most straightforward method of analysis for complex spectra is difference spectroscopy. Difference spectroscopy can be carried out simply by subtracting

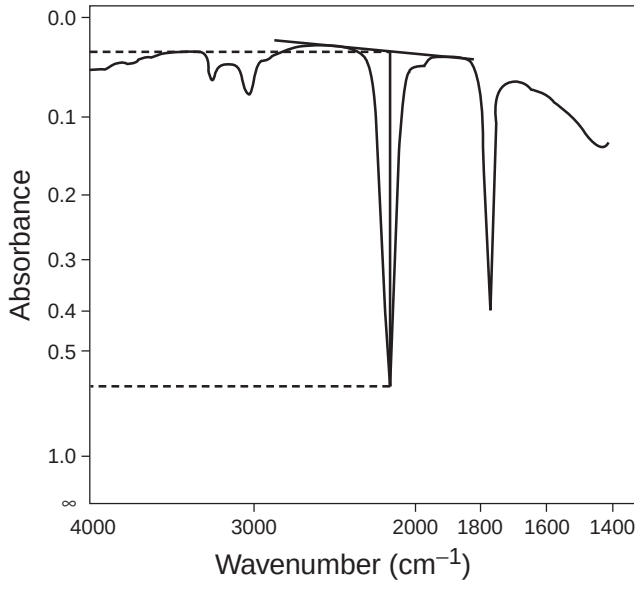

Figure 20 A baseline correction.

the IR spectrum of one component of the system from the combined spectrum to leave the spectrum of the other component. If the interaction between components results in a change in the structural properties of either one or both of the components, the changes will be observed in the difference spectra. The changes may manifest themselves via the appearance of positive or negative peaks in the spectrum.

Spectral subtraction can be applied to numerous applications and can be used for the data collected for solutions. In order to obtain the spectrum of a solution it is necessary to record spectra of both the solution and the solvent alone. The solvent spectrum may then be subtracted from the solution spectrum. There are two approaches – the solvent alone in the cell may be used as a reference in a double beam dispersive experiment, or the solvent spectrum may be recorded separately and then subtracted if using a single beam instrument. Figure 21, where the strong IR spectrum of water has been removed from a relatively weak spectrum of a 1% w/v solution of aspirin, illustrates how subtraction can be very useful in FTIR spectroscopy. Care must be taken when subtracting spectra. The concentration of the solvent alone is greater in than the solvent in the solution and negative peaks may appear in the regions of solvent absorption. In certain circumstances, the spectrum due to the solvent may be very intense, making simple subtraction impossible. This situation can make it difficult to investigate the sample spectrum as solvent bands may overlap with the region under investigation. In such experiments, ATR provides a suitable substitute. The nature of the ATR technique produces a less intense solvent contribution to the overall IR spectrum and so solvent spectra can be more readily subtracted from the sample spectrum of interest. Subtraction can also be applied to solid samples and is especially useful for mulls. The spectrum of the mulling agent can be subtracted, giving the spectrum of the solid only.

Spectra can also be differentiated. Figure 22 shows a single absorption peak and its first and second

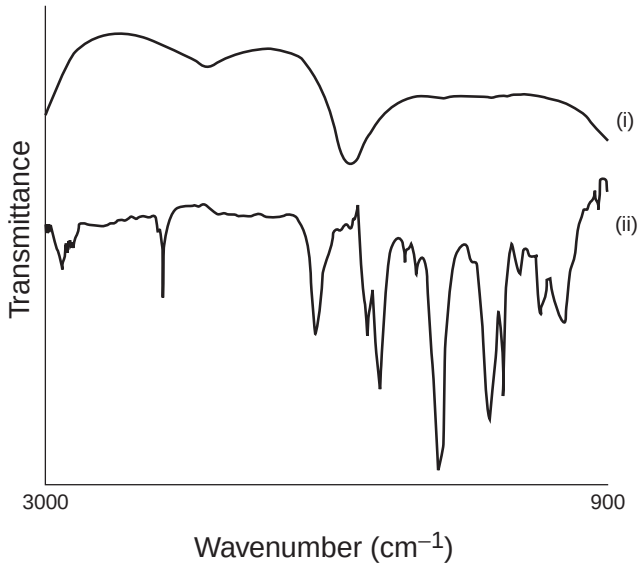

Figure 21 The IR spectra of 1% w/v aspirin in water (i) and after subtraction of the water spectrum (ii). (Reproduced with permission from a Perkin-Elmer sales leaflet.)

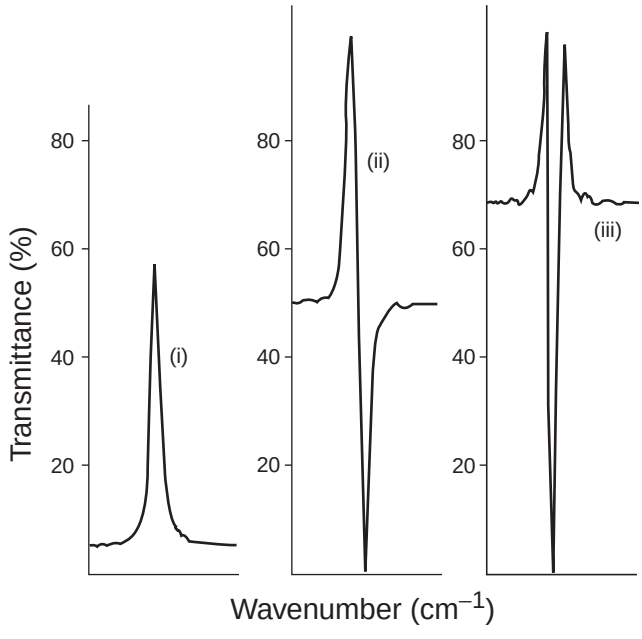

Figure 22 A single absorption peak (i), the first derivative (ii) and the second derivative (iii).

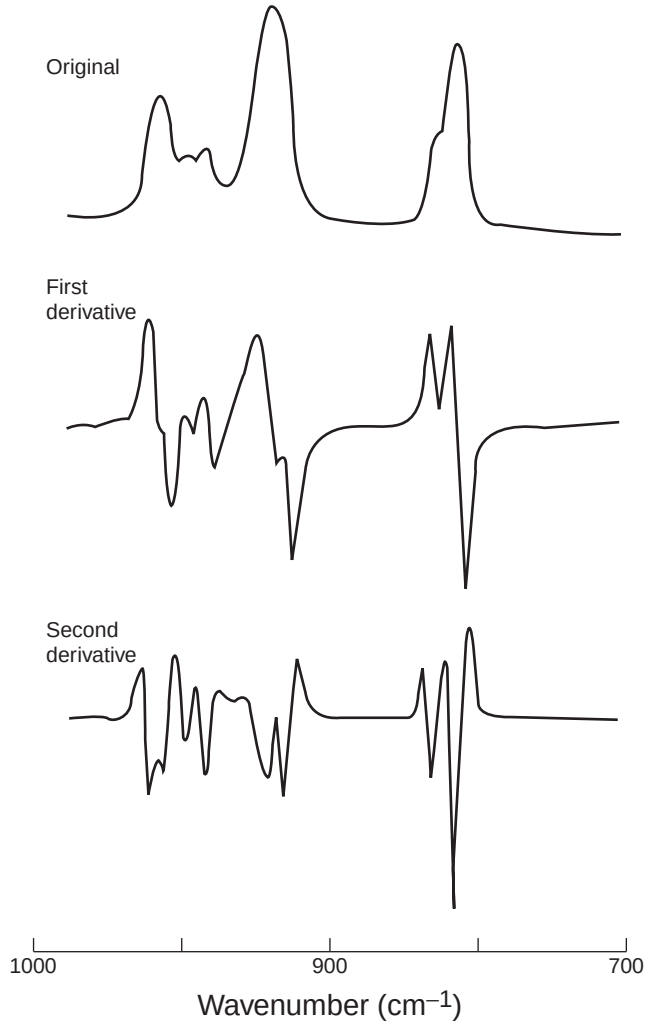

Figure 23 A complex absorption band with first and second derivatives. (Reproduced with permission from the Perkin-Elmer application note on derivative spectroscopy.)

derivative. Derivative techniques have long been used in quantitative UV/VIS spectroscopy, the benefits of the technique being two-fold. Resolution is enhanced in the first derivative since we are looking at changes in gradient. The second derivative gives a negative peak for each band and shoulder in the absorption spectrum. The advantage of derivatization is more readily appreciated for more complex spectra and Figure 23 shows how differentiation can be used to resolve and locate peaks in an envelope. Note that sharp bands are enhanced at the expense of broad ones and this may allow selection of a peak, even when there is a broad band beneath. With modern FTIR spectrometers it is possible to apply what is known as Fourier derivation.[10] During this process the spectrum is transformed to an interferogram. It is then multiplied by an appropriate weighting function and finally it is retransformed to give the derivative. This technique provides more sensitivity.

Deconvolution is the process of compensating for the intrinsic linewidths of modes in order to resolve overlapping bands. The technique yields spectra that have much narrower bands and is able to distinguish closely-spaced features. The instrumental resolution is not increased, but the ability to differentiate spectral features can be significantly improved. This is illustrated by Figure 24, which shows a broad band before and after deconvolution has been applied. The peaks at quite close frequencies are now easily distinguished. The deconvolution technique generally involves several steps: computation of an interferogram of the sample by computing the inverse Fourier transform of the spectrum; multiplication of the interferogram by a smoothing function and by a function consisting of a Gaussian–Lorentzian band shape; Fourier transformation of the modified interferogram. The deconvolution procedure is typically repeated iteratively for best results. At iteration, the lineshape is adjusted in an attempt to provide narrower bands without excessive distortion. There are three parameters that can be adjusted to tune the lineshape. The proportion of Gaussian and Lorentzian lineshape may be altered. The scale of these components is adjusted depending on the predicted origins of the band shape. For instance, it will vary depending on whether you

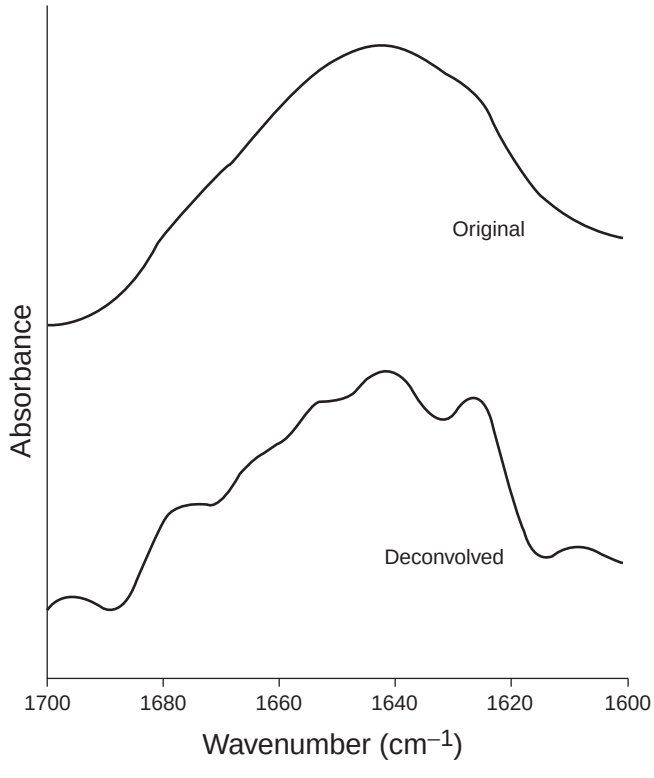

Figure 24 A broad IR band before and after deconvolution.

are examining a solid or a liquid or a gas. The half-width is the width of the lineshape and is normally the same as or larger than the intrinsic linewidth (full width at half height) of the band. If the value specified is too small, the spectrum tends to show only small variations in intensity. If too large, distinctive negative side lobes are produced. The narrowing function is the degree of narrowing attempted on a scale 0–1. If specified too small, the resulting spectrum will not be significantly different from the original. If too large, the spectrum may produce false peaks as noise starts to be deconvolved.

Quantitative values for band areas of heavily overlapped bands can be achieved by using curve-fitting procedures. Many curve-fitting procedures are based on a least-squares minimization procedure. Least squares curve-fitting covers the general class of techniques whereby one attempts to minimize the sum of the squares of the difference between an experimental spectrum and a computed spectrum generated by summing the component curves. Generally, the procedure involves entering the values of the frequencies of the component bands (determined using derivatives and/or deconvolution) and then the program determines the best estimate of the parameters of the component curves. Apart from the obvious variables of peak height and width, the type of band shape needs to be considered. The class of band shape of an IR spectrum depends on the type of sample. A choice of Gaussian, Lorentzian or a combination of these band shapes is usually applied. Figure 25 illustrates the curve-fitting process.

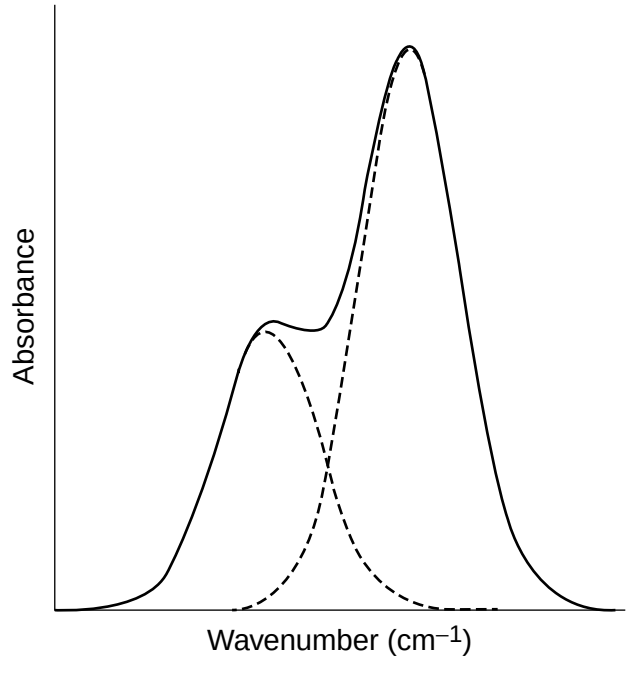

Figure 25 Curve-fitting of overlapping IR bands.

4.2 Determination of Concentration

According to the Beer–Lambert law the absorbance of a solution is directly proportional to the thickness and the concentration of the sample (Equation 4):

$$A = \varepsilon cl \tag{4}$$

where A is the absorbance of the solution, c is the concentration and l the path length of the sample. The constant of proportionality is usually given the symbol epsilon, ε, and referred to as the molar absorptivity. The absorbance is equal to the difference between the logarithms of the intensity of the light entering the sample (I_0) and the intensity of the light transmitted (I) by the sample (Equation 5):

$$A = \log_{10}\left(\frac{I_0}{I}\right. \tag{5}$$

Absorbance is therefore dimensionless. Transmittance is defined by Equation (6) as:

$$T = \frac{I}{I_0} \tag{6}$$

The Beer–Lambert law tells us that a plot of absorbance against concentration should be linear with a gradient of εl and pass through the origin. In theory, to analyze a solution of unknown concentration, solutions of known concentration need to be prepared, a suitable peak chosen, the absorbance at this frequency measured, and a graph plotted (a calibration graph). The concentration of the compound in solution can be read, given that its absorbance is known.

However, there are a few factors which need to be considered first when taking this approach. When preparing the solutions of known concentration, the concentrations have to give sensible absorbance values – not too weak and not too intense. Choosing a suitable absorption peak is crucial. The technique should be as sensitive as possible so an intense peak should be chosen. However, often IR spectra have many, sometimes overlapping peaks. A peak isolated from others, with a high molar absorptivity is required. A further problem that sometimes arises, especially in the spectra of solid samples, is the presence of asymmetric bands. In these cases, peak height cannot be used because the baseline will vary from sample to sample. Peak area measurements should be used instead. Many FTIR spectrometers have accompanying software which can carry out these calculations. Quantitative measurements need to be carried out on absorbance spectra rather than on transmittance spectra. Thus, transmittance spectra need to be converted to absorbance spectra. Most spectrometers have this simple process incorporated into their software.

4.3 Simple Analysis

The quantitative analysis of a component in solution can be successfully carried out given that there is a suitable band in the spectrum of the component of interest. As discussed previously, the band chosen for analysis should: have a high molar absorptivity; not overlap with other peaks from other components in the mixture or solvent; be symmetrical; and give a linear calibration plot of absorbance versus concentration.

Most quantitative IR methods of analysis use the intensities of the C=O, N−H or O−H groups. The C=O stretching band is the most commonly used because it is a strong band in a spectral region relatively free of absorption by other functional groups. In addition, the carbonyl band is not as susceptible as the O−H and N−H bands to chemical change or hydrogen bonding.

Simple quantitative analysis may be illustrated using the example of aspirin dissolved in chloroform. The best peak to choose in this example is the C=O stretching band of aspirin observed at 1764 cm^{-1} because it is an intense peak and lies in a region where there is no interference from the chloroform spectrum. The next step is to draw a calibration plot of absorbance versus aspirin concentration. A series of chloroform solutions of known aspirin concentration were prepared and the IR spectrum of each solution was recorded using a 0.1 mm NaCl transmission cell. The intensity of the 1764 cm^{-1} band was measured for each solution and the results are listed in Table 5. This data can then be graphed to produce a linear plot, as shown in Figure 26. The next stage in the analysis is to determine the amount of aspirin present in a chloroform solution of unknown concentration. The IR spectrum of the unknown sample was recorded. The intensity of the 1764 cm^{-1} carbonyl band was measured and an absorbance of 0.351 was recorded. Examination of the calibration graph shows that this corresponds to a concentration of 67 mg mL^{-1}.

IR spectroscopy can be used to measure the number of functional groups in a molecule, say the number of −OH or $-NH_2$ groups. It has been found that the molar absorptivity of the bands corresponding to the group is proportional to the number of groups. That is, each group has its own intensity which does not vary drastically from molecule to molecule.

Table 5 Calibration data for aspirin solutions

Concentration (mg mL^{-1})	Absorbance at 1764 cm^{-1}
0	0.000
25	0.158
50	0.285
75	0.398
100	0.501

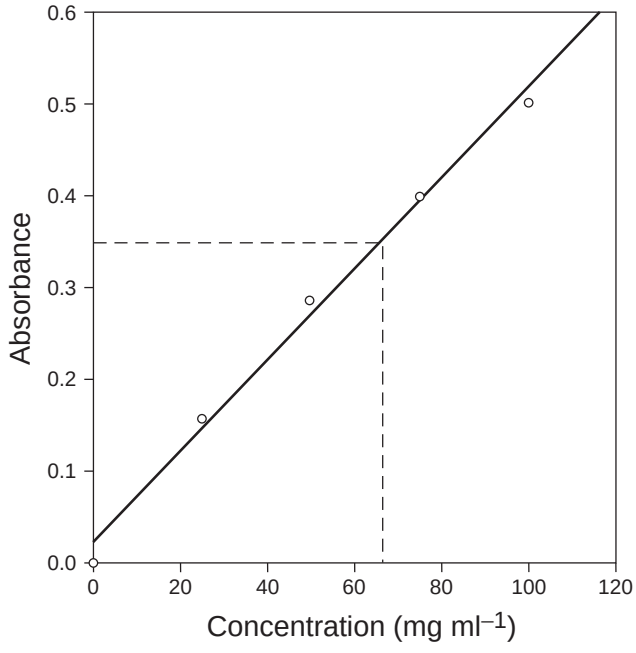

Figure 26 Calibration graph for aspirin solutions.

Solid mixtures can also be analyzed using IR spectroscopy. Solids are more susceptible to errors because of scattering. These analyses are usually carried out with KBr discs or in mulls. The problem here is the difficulty in measuring the path length. However, this measurement becomes unnecessary when an internal standard is used. Addition of a constant known amount of an internal standard is made to all samples and calibration standards. The calibration curve is obtained by plotting the ratio of the absorbance of the analyte to that of the internal standard, against the concentration of the analyte. The absorbance of the internal standard varies linearly with the sample thickness. The discs or mulls must be made under exactly the same conditions to avoid intensity changes or shifts in band positions. The standard must be carefully chosen and it should ideally have a simple spectrum with very few bands; be stable to heat and not absorb moisture; be easily reduced to a particle size less than the incident radiation without lattice deformation; be nontoxic, giving clear discs in a short time; and be readily available in the pure state. Some common standards used include calcium carbonate, sodium azide, naphthalene and lead thiocyanate.

4.4 Multicomponent Analysis

The analysis of a component in a complex mixture presents special problems. Quantitative analysis of a multicomponent system is illustrated by its application to

Table 6 Calibration data for drug mixture

Component	Concentration ($mg\,mL^{-1}$)	C=O stretching frequency (cm^{-1})	Absorbance
Aspirin	90	1764	0.217
Phenacetin	65	1511	0.185
Caffeine	15	1656	0.123

the simultaneous determination of aspirin, phenacetin and caffeine in tablets. These drug components can be determined by dissolving the sample in chloroform. Each of these components show distinct carbonyl bands. First the absorbances of the carbonyl bands of each of the drug components need to be measured using standard solutions. Typical results are listed in Table 6. Now the Beer–Lambert law (Equation 4) can be applied to determine the molar absorptivity for each component. Each of the standards was studied using a 0.1 mm path length NaCl transmission cell (Equations 7–9). For aspirin:

$$\varepsilon = \frac{A}{cl} = \frac{0.217}{[90\,mg\,mL^{-1} \times 0.1\,mm]} = 2.41 \times 10^{-2}\,mL\,mg^{-1}\,mm^{-1} \quad (7)$$

For phenacetin:

$$\varepsilon = \frac{0.185}{[65\,mg\,mL^{-1} \times 0.1\,mm]} = 2.85 \times 10^{-2}\,mL\,mg^{-1}\,mm^{-1} \quad (8)$$

For caffeine:

$$\varepsilon = \frac{0.123}{[15\,mg\,mL^{-1} \times 0.1\,mm]} = 8.20 \times 10^{-2}\,mL\,mg^{-1}\,mm^{-1} \quad (9)$$

Examination of the IR spectrum of the unknown sample produced under the standard sample conditions resulted in the data listed in Table 7. The concentrations of each component may be determined by once again applying the Beer–Lambert Law (Equations 10–12).

For aspirin:

$$c = \frac{A}{\varepsilon l} = \frac{0.207}{[2.41 \times 10^{-2}\,mL\,mg^{-1}\,mm^{-1} \times 0.1\,mm]} = 85\,mg\,mL^{-1} \quad (10)$$

For phenacetin:

$$c = \frac{0.202}{[2.85 \times 10^{-2}\,mL\,mg^{-1}\,mm^{-1} \times 0.1\,mm]} = 71\,mg\,mL^{-1} \quad (11)$$

Table 7 Absorbance values for unknown drug mixture

Frequency (cm^{-1})	Absorbance
1764	0.207
1511	0.202
1656	0.180

For caffeine:

$$c = \frac{0.180}{[8.20 \times 10^{-2}\,mL\,mg^{-1}\,mm^{-1} \times 0.1\,mm]} = 22\,mg\,mL^{-1} \quad (12)$$

It should be pointed out that there could be a certain degree of error in these values. There can be error due to the fact that it is difficult to select a baseline for the analysis because of overlapping peaks. Also, any nonlinearity of the Beer–Lambert law plots has been ignored, having used one value to determine the constant of proportionality above for each solution.

Alternatively, this same problem can be determined by spectral subtraction. If a sample of the additive is available, mixtures of the starting material and product can be separated by subtraction of the spectra of the mixture and additives. This is achieved by subtracting the fraction of aspirin, then phenacetin and finally caffeine from the unknown sample spectrum to null the corresponding absorptions. The subtraction factors can be read from the spectrometer computer. This second method is, of course, much faster.

Studies of the trans unsaturation in fats also provide a good illustration of multicomponent analysis. Naturally occurring vegetable fats and oils are a mixture of triglycerides of the general structure shown in Figure 27. This is a tri-ester of glycerol. The acids are long chain and can be saturated or unsaturated. The unsaturation is always present as the cis isomer. These cis bonds can be isomerized to the trans configuration during extraction or subsequent processing. For example, oxidation or partial hydrogenation can lead to isomerization. It is commercially important for labeling to determine this trans content. It is difficult to separate cis from trans isomers using other techniques such as GC and an IR method is commonly used. The IR method has the convenience of no weighing being required, nor are any

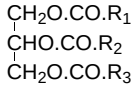

Figure 27 The molecular structure of triglycerides.

accurate volumes needed. It can therefore be used for very small samples.

The configuration of alkenes can be determined from the frequency of their C–H bending vibrations. Cis isomers absorb between 840 and 700 cm^{-1} while trans isomers absorb between 1000 and 930 cm^{-1}. This band can be used as the basis for an analytical method. It is normal to hydrolyze the triglyceride mixture to glycerol and a mixture of fatty acids, and the latter then converted to their methyl esters. It is this methyl ester mixture which is analyzed. The formulae of methyl oleate, the cis isomer and methyl elaidate, the trans isomer, are given in Figure 28. Both geometrical isomers absorb strongly at 1163 cm^{-1}, the C–O stretching frequency from the ester group. Cis isomers absorb weakly at 965 cm^{-1} while trans isomers absorb strongly at this frequency. It can be shown (Equation 13) that:

$$\%\text{trans} = K\left(\frac{A_{965}}{A_{1163}} - f\right) \qquad (13)$$

where K and f are constants, and A_{965} and A_{1163} are the absorbances of the solution at these frequencies, respectively. Note that no concentration or path length terms appear in this equation. This is the equation of a straight line of gradient K and of y-intercept, f, in a plot of % trans versus the ratio of absorbances at the two frequencies, 965 and 1163 cm^{-1}. A series of solutions of known trans composition are required, the ratio of absorbances then measured and the % trans content is then plotted against these values.

The improvement in computer technology associated with spectroscopy has also led to the expansion of quantitative IR spectroscopy. The application of statistical methods to the analysis of experimental data is known as chemometrics. There is a large amount of information now available regarding chemometrics and a detailed discussion is beyond the scope of this article. However, there are several multivariate data analytical methods

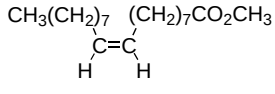

Methyl oleate
(*cis*)

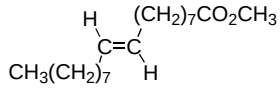

Methyl elaidate
(*trans*)

Figure 28 The molecular structures of methyl oleate and methyl elaidate.

that are used for the analysis of FTIR data.[11] The most commonly used methods in IR spectroscopy are: classical least squares (CLS); inverse least squares (ILS); partial least squares (PLS); and principal component regression (PCR). CLS methods (also referred to as K-matrix methods by IR spectroscopists) use a calibration model representing the physical law that describes the variation in the spectra with composition. For example, the Beer–Lambert Law where absorbance is represented as a linear function of the component concentrations. ILS methods (also known as P-matrix methods) differ in that the data are empirically modeled without using a direct physical model. These calibration methods are not restricted to the use of the same number of components as the individual components present in the spectral region being analyzed. The newer PCR and PLS methods are two factor analysis based. Factor analysis methods are used to factor the spectral data matrix into the product of two small matrices in order to simplify the data representation. This allows a more stable solution to be achieved, compared to the use of the ILS approach.

5 APPLICATIONS

IR spectroscopy has been applied to quite a wide range of biological systems and a number of these are examined here. These include proteins, peptides, lipids, biomembranes, animal tissue, carbohydrates, pharmaceuticals, foods and plant material.

5.1 Proteins and Peptides

The IR spectra of proteins exhibit absorption bands associated with their characteristic amide group, the structural unit common to all molecules of this type (shown in Figure 29).[12] An isolated planar amide group gives rise to five in-plane modes and one out-of-plane normal mode. The in-plane modes are due to C=O stretching, C–N stretching, N–H stretching, N–H bending, O–C–N bending, while the out-of-plane mode is due C–N torsion. Characteristic bands of the amide groups of protein chains are similar to absorption bands exhibited by secondary amides in general, and are labeled amide bands. There are nine such bands, called amide A, amide B and amide I–VII, in order of decreasing frequency and these bands are summarized in Table 8. Some of these bands are more useful for conformation

Figure 29 The amide group of proteins.

Table 8 Characteristic IR amide bands of proteins

Designation	Approximate frequency (cm^{-1})	Nature of vibration
A	3300	N–H stretching in resonance with
B	3110	with overtone (2 × amide II)
I	1653	80% C=O stretching; 10% C–N stretching; 10% N–H bending
II	1567	60% N–H bending; 40% C–N stretching
III	1299	30% C–N stretching; 30% N–H bending; 10% C=O stretching; 10% O=C–N bending; 20% methyl-CH_3 stretching
IV	627	40% O=C–N bending; 30% methyl-CH_3 stretching and 30% other
V	725	N–H bending
VI	600	C=O bending
VII	200	C–N torsion

studies than others and the amide A, amide I and amide II bands have been the most frequently used. Knowledge of peptide bond vibrations is based mainly on normal coordinate calculations using *N*-methylacetamide as a model compound.

The most useful IR band for the analysis of the secondary structure of proteins in aqueous media is the amide I band, occurring between approximately 1700 and 1600 cm^{-1}.[13] The amide I band represents (80%) C=O stretching vibration of the amide group coupled to the in-plane N–H bending and C–N stretching modes. The exact frequency of this vibration depends on the nature of hydrogen bonding involving the C=O and N–H groups and this is determined by the particular secondary structure adopted by the protein. Proteins generally contain a variety of domains containing polypeptide fragments in different conformations. As a consequence, the observed amide I band is usually a complex composite, consisting of a number of overlapping component bands representing helices, β-structure, turns and random structures.

The amide II band represents mainly (60%) N–H bending, with some C–N stretching (40%). Like the amide I band, it is possible to split the amide II band into components depending on the secondary structure of the protein. The position of the amide II band is sensitive to deuteration, shifting from around 1550 cm^{-1} to a frequency of 1450 cm^{-1}. The amide II band of the deuterated protein overlaps with the H–O–D bending vibration, making it difficult to obtain information about the conformation of this band. However, the remainder of the amide II band at 1550 cm^{-1} may provide information about the accessibility of solvent to the peptide backbone. Hydrophobic environments or tightly ordered structures, such as α-helix or β-sheet, reduce the chance of exchange of the amide N–H proton.

Secondary amides in the trans configuration exhibit two characteristic bands at about 3300 cm^{-1} and 3100 cm^{-1}, while those in this configuration show two bands at about 3200 cm^{-1} and 3100 cm^{-1}. The band at 3300–3200 cm^{-1} is the amide A band and is due to the stretching mode of the N–H bond which is hydrogen bonded. The amide B is at 3100 cm^{-1} and is due to the Fermi resonance of the N–H stretching vibration with the overtone of the amide II vibration for the trans configuration and with the combination of the C–O stretching and N–H in-plane bending vibrations for the cis configuration. The amide A and B bands are very sensitive to hydrogen bonding, but because of the mixed nature of these modes correlations between frequency shifts and hydrogen bonding are more complex than would be the case for an unperturbed N–H fundamental.

The amide III and the amide VII bands have not been studied as extensively as the amide A, B, I and II modes. The amide III and amide IV vibrations are complex in-plane modes with no more than about 40% of the potential energy associated with any single internal displacement coordinate. The amide IV and VII vibrations cannot even be said to be localized in the amide group and normal coordinate analysis of system containing several repeat units are necessary to interpret their relationship to structure.

Of the low frequency modes, the amide V vibration has been the most useful in structural investigations since it involves N–H out-of-plane bending and depends considerably on the backbone conformation. The amide V mode represents N–H out-of-plane bending and depends considerably on the backbone conformation. The band appears at around 600 cm^{-1} for an α-helical conformation, at around 650 cm^{-1} for unordered structures and at around 700 cm^{-1} for β-structure.

The amide IV, VI and VII modes have not yet been studied in detail and little information is available regarding their dependence on conformation. The amide IV band represents 40% O=C–H bending and 30% methyl–C stretching, the amide VI represents C=O out-of-plane bending. The amide VII band represents C–N torsion. Both of these modes have low intensities, making information difficult to obtain. Also, information is difficult to extract about the amide VII band because of its low frequency.

The amide III band is heavily mixed, representing 30% C–N stretching, 30% N–H in-plane bending and 20% methyl–C stretching. The amide III mode is also sensitive to secondary structure, but also gives rise to a very weak IR band at about 1300 cm^{-1}. Due to different selection rules, the amide VIII band is much stronger in the corresponding Raman technique, making Raman

Table 9 The characteristic frequencies of amino acid side chains

Amino acid	Frequency (cm^{-1})	Assignment
Alanine	1465	CH_2 bending
Valine	1450	CH_3 asymmetric bending
Leucine	1375	CH_3 symmetric bending
Serine	1350–1250	O–H deformation
Aspartic acid	1720	C=O stretching
Glutamic acid	1560	CO_2^- asymmetric stretching
	1415	CO_2^- symmetric stretching
Asparagine	1650	C=O stretching
Glutamine	1615	NH_2 bending
Lysine	1640–1610, 1550–1485	NH_3^+ deformation
	1160, 1100	NH_3^+ rocking
Phenylalanine	1602, 1450, 760, 700	Ring vibrations
Tyrosine	1600, 1450	Ring vibrations

spectroscopy a more suitable technique for studying this vibrational mode.

In general, a spectrum can only be properly analyzed if all the absorption bands are considered because there are no isolated vibrations involving only particular groups of atoms. However, for complex molecules it is possible to draw conclusions from the analyses of selected bands if the corresponding vibrations can be regarded as localized and this is true for many of the characteristic amide bands. It should be pointed out that the characteristic amide bands are not the only bands to appear in the spectra of proteins or simple model compounds. There are bands associated with C–H, C–N and C–C linkages in the spectra of model compounds.

In addition, the IR contributions of the side chains of the amino acids which constitute the protein must also be considered. Amino acid side chains exhibit IR modes that are often useful for investigating the local group in a protein. Fortunately these contributions have been found to be small in D_2O compared to the contributions made by the amide I band. It is also important to be aware of the location of such modes as they may be confused with amide vibrations. But the arginyl residue is the only residue which makes a significant contribution in the 1700–1600 cm^{-1} region. However, even the bands at 1586 and 1608 cm^{-1} due to the arginyl residue do not contribute greatly compared to the amide I contributions. The characteristic side chain IR frequencies of amino acids are summarized in Table 9.

Quantitative analysis of the secondary structures of polypeptides may be carried out. The amide I band of proteins and peptides consists of a series of overlapped component bands due to the secondary structures present in such molecules. Resolution enhancement of the amide I band allows the identification of various structures present in a protein or peptide. Derivatives and deconvolution can be used to obtain such information. Figure 30 illustrates the use of these methods to analyze the amide I mode of a peptide. However, simply measuring the peak heights of the resulting spectra does not provide an accurate measure of the proportion of secondary structure. In Fourier deconvolved spectra the band shapes are distorted and the extent of this distortion depends on the widths of the component bands and the parameters use to compute the deconvolved spectra, so the resulting peak heights are complex functions of a number of parameters. Similarly, the peak heights in the derivative spectra

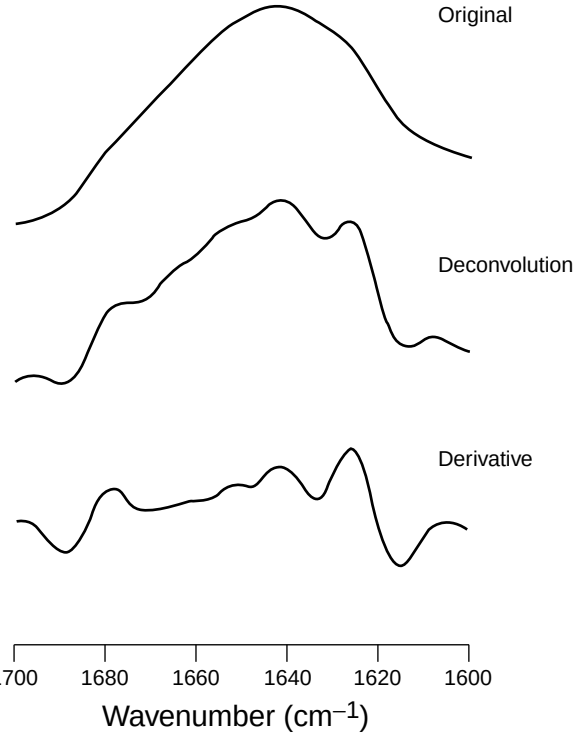

Figure 30 The original, deconvolved and derivative spectra of a peptide amide I band.

Table 10 The amide I band frequencies and secondary structure assignments for standard proteins

Protein	β-Structure				Helix	Random		Turns and bends			
Carbonic anhydrase	1636		1625	1679	1653	1645	1660	1668		1690	
Carboxypeptidase	1635	1628	1623	1679	1652	1642	1665	1671		1690	
Casein						1644		1671			
α-Chymotrypsin	1637		1627	1674	1654	1647	1665		1681	1688	
Chymotrypsinogen	1637		1627	1674	1654	1646	1664		1682	1688	1696
Concanavalin A	1639	1634	1623	1671	1653	1646	1659		1684		
Elastase	1639	1633	1620	1673	1657	1645	1665		1683		
Immunoglobulin G	1637		1624	1672	1651		1659			1688	
α-Lactalbumin	1637		1627	1676	1652	1644	1665		1684		
β-Lactalglobulin		1634	1623	1679	1654	1646	1685		1685		1692
Lysozyme		1630		1673	1654	1641	1666		1682		
Papain	1640	1632	1621	1679	1654	1646	1662	1670		1687	
Ribonuclease A	1637	1628		1676	1655	1646	1665		1684	1688	
Trypsin	1638		1627	1673	1654	1646	1664		1681		
Trypsinogen	1636		1624	1675	1654	1646	1663		1684		1693
Trypsin inhibitor	1636			1675	1656	1646	1664		1685		
Mean	1637	1631	1624	1675	1654	1645	1663	1670	1683	1688	1694
Rms deviation	1.4	2.5	2.4	2.5	1.5	1.6	2.2	1.4	1.5	1.1	1.7
Maximum deviation	3	3	4	4	3	4	4	2	2	2	2

cannot be used as a quantitative measure because they are functions of both intensities and the widths of the original bands. The best method used for the estimation of protein secondary structure involves band-fitting the amide I band. The parameters required, the number of component bands and their positions, are obtained from the resolution enhanced spectra. The fractional areas of the fitted component bands are directly proportional to the relative proportions of structure that they represent. The percentages of helix, β-structure and turns are estimated by addition of the areas of all the component bands assigned to each of these structures and expressing the sum as a fraction of the total amide I area. The assumption is made that the intrinsic absorptivities of the amide I bands corresponding to different structures are identical.

IR spectroscopy has been used in a large number of studies of proteins in a range of environments. Some examples of the proteins analyzed are listed in Table 10, which details the deconvolved amide I frequencies and secondary structure assignments made for a series of proteins in D_2O.[14] These assignments are based on the fact that the secondary structures of these globular proteins have been very well characterized by X-ray crystallography. The characteristic frequencies of the secondary structures of proteins have also been estimated from normal coordinate calculations on model peptides and proteins of known structure.[13] The frequencies are summarized in Table 11. These assignments are by no means exact and some bands may appear outside these ranges for some proteins, particularly when solvent interactions are considered.

An example of the quantitative approach to protein analysis is illustrated by Figure 31 which shows the amide I band of the enzyme lysozyme in D_2O. The amide I band of this protein shows nine component bands and the relative

Table 11 The characteristic amide I band frequencies of protein secondary structure

Frequency (cm^{-1})	Assignment
1695–1670	Intermolecular β-structure
1690–1680	Intramolecular β-structure
1666–1659	Three-turn helix
1657–1648	α-Helix
1645–1640	Random coil
1640–1630	Intramolecular β-structure
1625–1610	Intermolecular β-structure

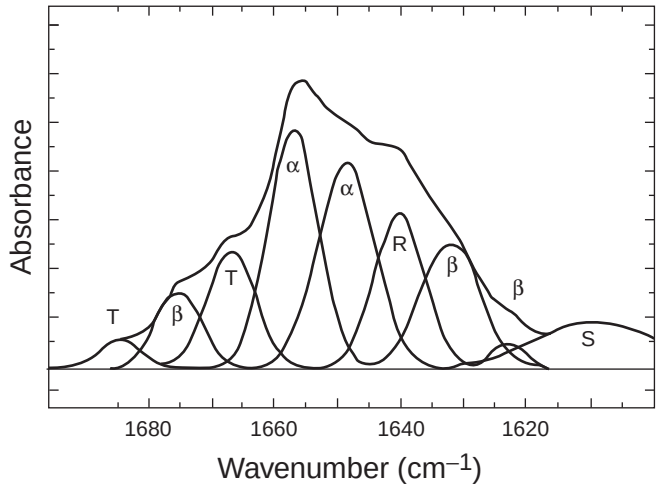

Figure 31 The curve-fit amide I band of lysozyme in D_2O.

Table 12 Analysis of the amide I band of lysozyme in D_2O

Frequency (cm^{-1})	Relative area (%)
1623	1
1632	15
1640	16
1648	24
1657	24
1667	11
1675	7
1684	2
1693	<1

areas of these components are listed in Table 12. The components may be assigned to the various types of secondary structures. The bands at 1623 and 1632 cm^{-1} are characteristic of β-structure, as is the band at 1675 cm^{-1}. The bands at 1667, 1684 and 1693 cm^{-1} occur at frequencies characteristic of turns and bends. The band at 1610 cm^{-1} is due to arginyl side chain vibrations and the 1640 cm^{-1} band may be assigned to random coil. The two remaining bands at 1648 and 1657 cm^{-1} are due to the presence of α-helix. Usually in protein IR spectra only one component due to α-helix is observed. However, X-ray data indicates the presence of two types of helix in lysozyme: α-helix and 3-turn helix. These different helices vibrate at different frequencies. The 1657 cm^{-1} is assigned to α-helix, while the 1648 cm^{-1} band is due to 3-turn helix. These assignments give a quantitative estimate of 48% helix, 23% β-structure, 13% turns and 16% random coil in lysozyme.

Peptides are often derived from the larger proteins, and a similar quantitative approach can be applied to these molecules. However, their small size must be taken into account when applying such methods. Small peptides are not necessarily capable of producing the same secondary structures observed in proteins. The CiT4 peptide is derived from a nervous system protein and consists of a sequence of 26 amino acids. The amide I band of CiT4 peptide in D_2O has been deconvolved and curve-fit and the results are illustrated in Figure 32. Table 13 lists the wavenumber and the fractional area of each component of the amide I band of CiT4 in D_2O. The major component in Figure 32 is observed at 1644 cm^{-1}, contributing to 73% of the total amide I band area. This frequency is characteristic of random coil in proteins and peptides. Two smaller component bands appear at 1664 and 1674 cm^{-1}, respectively, and both can be attributed to turns and bends in the peptide structure. The weak component at 1617 cm^{-1} is at a frequency associated with arginine side chain contributions.

Table 13 Analysis of the amide I band of the CiT4 peptide in D_2O

Frequency (cm^{-1})	Relative area (%)
1644	73
1664	16
1674	11

The solvent environment in which proteins and peptides are recorded affects the secondary structures observed for these molecules. For instance, the solvent trifluoroethanol (TFE) is more polar than water. The FTIR spectrum of the CiT4 peptide in TFE/D_2O is shown in Figure 33. The frequency and the fractional area of

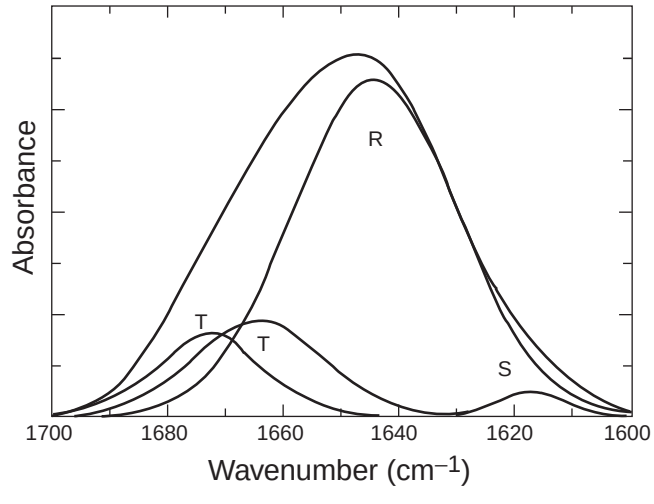

Figure 32 The curve-fit amide I band of the CiT4 peptide in D_2O.

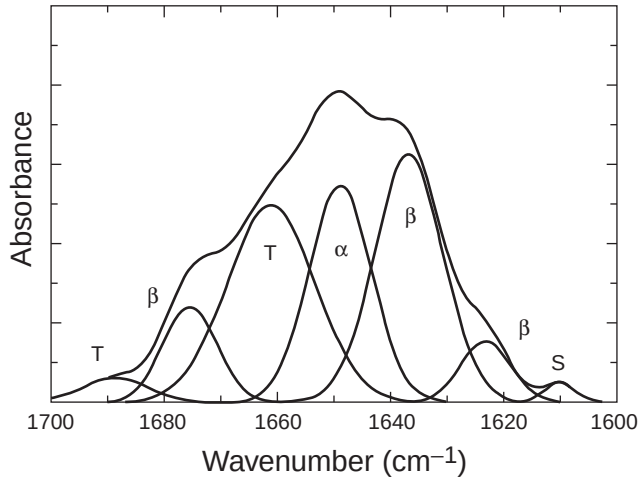

Figure 33 The curve-fit amide I band of the CiT4 peptide in TFE/D_2O.

Table 14 Analysis of the amide I band of the CiT4 peptide in TFE/D_2O

Frequency (cm^{-1})	Relative area (%)
1623	7
1637	26
1649	21
1661	30
1675	10
1689	6

each component of the amide I band of CiT4 in TFE solution are listed in Table 14. There are significant changes observed to the amide I band of CiT4 in the presence of TFE. There is a notable change in the shape of the amide I band and six component bands are observed. The small component at 1610 cm^{-1} occurs in the frequency range normally attributed to side chain contributions. The components at 1623 and 1637 cm^{-1} are assigned to low-frequency β-structure, with a high-frequency component of β-structure also observed at 1675 cm^{-1}. These bands contribute to 43% of the total band area. The component in Figure 33 observed at 1655 cm^{-1}, contributing to 21% to the total amide I band area, is assigned to α-helix. The remainder of the components in Figure 33 at 1661 and 1689 cm^{-1} are assigned to turns and bends in the peptide secondary structure. Analysis of the spectrum of CiT4 in TFE/D_2O solution indicates differences in the secondary structure of the peptide when compared to that observed for the peptide in the pure aqueous solution. On addition of TFE, a more structured conformation of CiT4 is indicated. Notably α-helical structure and β-structure appear to be formed in the less polar environment provided by TFE. Thus, a change in environment, in this case solvent, can cause notable changes to a peptide conformation observable using IR spectroscopy.

IR techniques can also be used to study protein adsorption onto surfaces. For example, a flow-through ATR cell can be used ex vivo with flowing blood. Blood–surface interactions are of great importance when medical polymers, such as those used in heart valves and artificial organs, are implanted into the body. When polymers come into contact with blood, complex reactions take place and can result in the formation of a blood clot. IR analysis has shown in ex vivo studies of dogs that during the early stages the proteins albumin and glycoprotein are present. Immediately after this observation increased amount of fibrinogen appear. As the adsorption process continues, albumin is replaced by other proteins until a blood clot is formed.

5.2 Lipids

Many of the lipids found in nature contain phosphorus and are classed as phospholipids. One class of phospholipid are the phosphoglycerides and the general structure of this type of lipid is illustrated in Figure 34. Phospholipids are organized in lipid bilayers of about 40–80 Å in thickness, where the polar head group points towards the aqueous phase and hydrophobic tails point towards the tails of a second layer. The chains can be in an all trans conformation which is referred to as the gel phase. A liquid crystalline phase is obtained when the chain also contains gauche C–C groups.

The IR spectra of phospholipids can be divided into spectral regions which originate from molecular vibrations of the hydrocarbon tail, the interface region and the head-group.[15] The major IR modes due to phospholipids are summarized in Table 15. The hydrocarbon tail gives rises to acyl chain modes. The most intense vibrations in the IR spectra of lipid systems are the CH_2 stretching vibrations. These give rise to bands in the region 3100 to 2800 cm^{-1}. The CH_2 asymmetric and symmetric stretching modes at 2920 and 2851 cm^{-1}, respectively, are generally the strongest bands in the lipid spectra. The frequencies of these bands are conformation-sensitive and respond to changes of the trans/gauche ratio in the acyl chains. This is also the case for the vibrational modes due to the terminal CH_3 groups at 2956 cm^{-1} (asymmetric stretching) and 2873 cm^{-1} (symmetric stretching). The C=C–H stretching bands due to unsaturated acyl chains are found at 3012 cm^{-1}. The bands due to methylene and methyl groups occur in the 1500–1350 cm^{-1} region. At around 1470 cm^{-1} there are bands due to CH_2 bending and the number and frequency of these bands are dependent on acyl chain packing and conformation. While the asymmetric

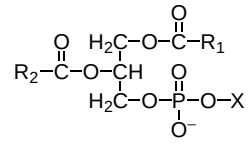

phosphatidylserine

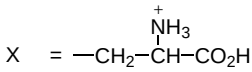

phosphatidycholine

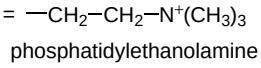

phosphatidylethanolamine

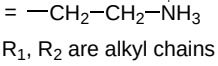

R_1, R_2 are alkyl chains

Figure 34 Phospholipid structure.

Table 15 Characteristic IR bands of lipids

Approximate frequency (cm^{-1})	Nature of vibration
3010	=C−H stretching
2956	CH_3 asymmetric stretching
2920	CH_2 asymmetric stretching
2870	CH_3 symmetric stretching
2850	CH_2 symmetric stretching
1730	C=O stretching
1485	$(CH_3)_3N^+$ asymmetric bending
1473, 1472, 1468, 1463	CH_2 scissoring
1460	CH_3 asymmetric bending
1405	$(CH_3)_3N^+$ symmetric bending
1378	CH_3 symmetric bending
1400–1200	CH_2 wagging band progression
1228	PO_2^- asymmetric stretching
1170	CO−O−C asymmetric stretching
1085	PO_2^- symmetric stretching
1070	CO−O−C symmetric stretching
1047	C−O−P stretching
972	$(CH_3)_3N^+$ asymmetric stretching
820	P−O asymmetric stretching
730, 720, 718	CH_2 rocking

deformation modes of the CH_3 group are obscured by the scissoring bands, the symmetric deformation mode appears at 1378 cm^{-1}. The polymethylene chains show a series of bands in the region 1380–1180 cm^{-1} but these are obscured in the spectra of phospholipids because they overlap with the strong PO_2^- asymmetric stretching band around 1228 cm^{-1}. There is also a series of CH_2 rocking bands in the 1150–700 cm^{-1} region. The most intense band of this group is at 720 cm^{-1} and overlaps with a strong H_2O band, so it is necessary to observe it in D_2O. Deuteration of lipids result in C−D stretching frequencies in the 2000–2300 cm^{-1} region, a region free from interfering absorption bands. As for the C−H stretching bands, the C−D stretching bands are sensitive to the conformation and the trans/gauche ratio of the deuterated acyl chain segments. Bands around 2195 and 2090 cm^{-1} are due to the CD_2 asymmetric and symmetric stretching vibrations, respectively. Bands at 2212, 2169 and 2070 cm^{-1} arise from the CD_3 group. Deuteration also simplifies the fingerprint region of the IR spectra as the methylene wagging and rocking-twisting band progressions disappear from their usual position in the 1400–700 cm^{-1}.

Spectral modes arising from the head-group and interfacial region also provide useful information.[16] The most useful IR bands for studying the interfacial region of lipid assemblies are the ester group vibrations, particularly the C=O stretching bands in the 1750–1700 cm^{-1} region. In diacyl lipids this region consists of at least two bands originating from the two ester carbonyl groups. A band at 1742 cm^{-1} is assigned to the C=O mode of the *sn*-1 chain with a trans conformation in the carbon–carbon bond adjacent to the ester grouping, while the 1728 cm^{-1} C=O frequency of the *sn*-2 chain suggests the presence of a gauche band in that position. The frequency difference observed reflects the structural inequivalence of the chains, with the *sn*-1 chain initially extending in a direction perpendicular to the *sn*-2 chain and then developing a gauche bend in order to render the two chains parallel. For example, the C=O stretching region of the spectrum of the lipid phosphatidylserine (PS) is shown in Figure 35. The band at 1624 cm^{-1} is due to the asymmetrical CO_2^- stretching vibration of PS. The band at 1737 cm^{-1} is broad, consisting of two ester C=O stretching bands at 1742 and 1728 cm^{-1} and a small band at 1701 cm^{-1} which is assigned to a hydrogen-bonded carbonyl group. The higher frequency band at 1742 cm^{-1}, associated with the *sn*-1 chain of PS is the more intense band, indicating that a trans conformation of the chain about the C−C bond is favored.

The choline, ethanolamine and serine groups have characteristic modes. The CH_3 asymmetric stretching modes of the $(CH_3)_3N^+$ group are around 3040 cm^{-1} and are out of the range of the acyl chain CH_3 bands. The corresponding methyl bending vibrations of the choline group occur at higher frequencies, compared to those of the acyl chain methyl groups. The C−N stretching bands are found between 1040 and 800 cm^{-1}. There are also several characteristic phosphate group vibrations. There are strong bands due to the P=O stretching vibration: an asymmetric PO_2^- stretch around 1228 cm^{-1} and a symmetric PO_2^- stretch around 1085 cm^{-1}. There are P−O stretching modes in the region 900–800 cm^{-1}, but these are obscured by strong water absorption in this region.

In certain phospholipid membranes which contain unsaturated acyl chains, the typical lamellar liquid crystalline phase converts to a micellar nonlamellar phase

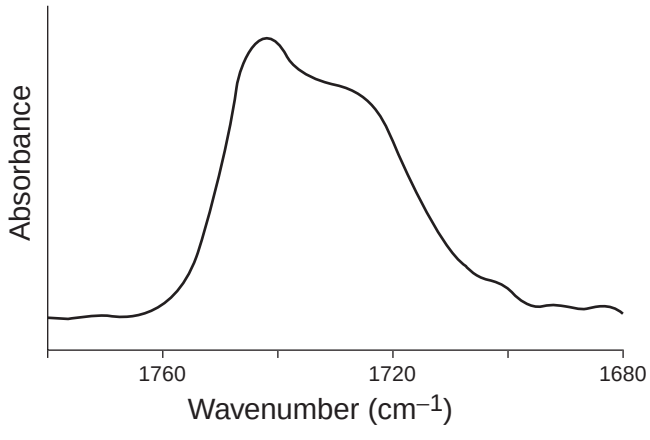

Figure 35 The C=O stretching region of PS in D_2O.

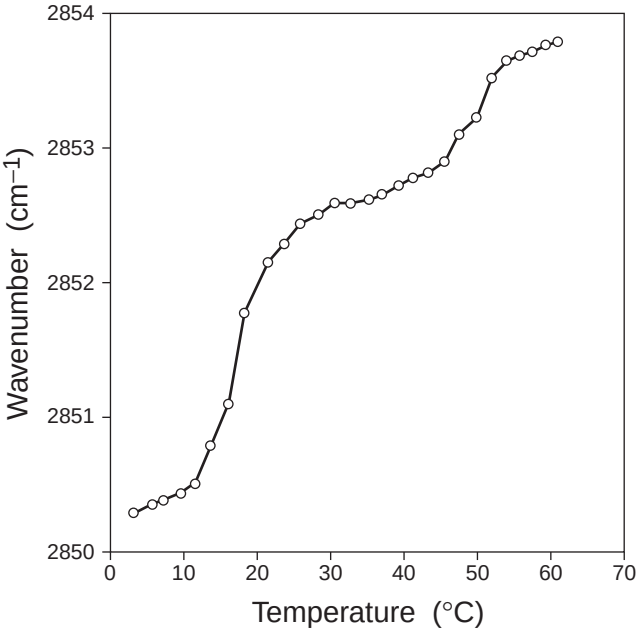

Figure 36 The temperature dependence of the symmetric CH_2 stretching band of PE.

upon heating.[17] Such a thermally-induced transition involves a major structural rearrangement. Temperature studies of the IR spectra of phospholipids provide a sensitive means of studying such transitions in lipids. Figure 36 shows the temperature dependence of the frequency of the symmetric CH_2 stretching band in the spectra of lipid membranes obtained from phosphatidylethanolamine (PE). The increasing frequency with temperature indicates an increasing concentration of gauche bands in the acyl chains and this leads to the formation of the nonbilayer phase at higher temperatures. Figure 36 shows a frequency shift of about $2\,cm^{-1}$ at 18 °C and this is associated with the gel to liquid crystal phase transition. An additional frequency shift of approximately $1\,cm^{-1}$ at 50 °C is associated with a transition to the micellar phase. Both these transitions have been observed to be reversible.

Quantitative IR analysis can be carried out on blood serum to determine the relative amounts of lipid present.[18] Triglycerides, phospholipids and cholesteryl esters are the classes of lipid which occur in blood serum. These compounds occur naturally in concentrations which make IR analysis attractive, and the necessary preliminary separation is simple. These classes of compounds can be characterized using IR spectroscopy by their carbonyl bands. The peak maxima are:

triglyceride $1742\,cm^{-1}$
phopholipid $1737\,cm^{-1}$
cholesteryl ester $1723\,cm^{-1}$.

However, the carbonyl peaks are heavily overlapped and these are too close to allow analysis by three simultaneous equations, but a least-squares method can be used to separate the components. The concentrations of these lipid components are usually in the range 0.03–0.3% in human blood, and standard solutions can be prepared in chloroform.

5.3 Biomembranes

FTIR spectroscopy is particularly useful for probing the structure of membrane proteins.[17] Until recently a lack of adequate experimental techniques has been the reason for the poor understanding of the secondary structure of most membrane proteins. X-ray diffraction requires high quality crystals and these are not available for many membrane proteins. Circular dichroism (CD) has been widely used for studying the conformation of water-soluble proteins, but problems arise for its use for membrane proteins. The light scattering effect may distort CD spectra and lead to substantial errors in their interpretation. Also, the reference spectra used for the analysis of CD spectra are based on globular proteins in aqueous solution and may not be applicable to membrane proteins in the hydrophobic environment of lipid bilayers. IR spectroscopy is largely unaffected by the aforementioned problems. The technique can be used to study the secondary structure of proteins both in their native environment, as well as after reconstitution into model membranes. IR spectroscopy offers several advantages for studies of protein–lipid interaction. Information about lipid conformation and protein secondary structure from the same sample can be obtained in a single experiment.

Myelin basic protein (MBP) is a major protein of the nervous system and has been studied using FTIR spectroscopy in both aqueous solution and after reconstitution in myelin lipids.[19] The amide I band of MBP in D_2O solution (deconvolved and curve-fit) illustrated in Figure 37. The amide band shows a small band at $1616\,cm^{-1}$ which occurs at a frequency normally attributed to amino acid side chain contributions. The largest component band observed at $1647\,cm^{-1}$ is assigned to random coil and accounts for 86% of the total band area. The remaining component band at $1671\,cm^{-1}$ is assigned to turns and bends in the protein. Table 16 lists the frequency and the fractional area of each component of the amide I band of MBP in D_2O. The FTIR spectrum of MBP protein after reconstitution in myelin with all proteins removed was also recorded. The results of deconvolution and band-fitting the amide I band of MBP in myelin are shown in Figure 38. The frequency and the fractional area of each component of the amide I band of MBP in myelin are listed in Table 17. There are dramatic changes

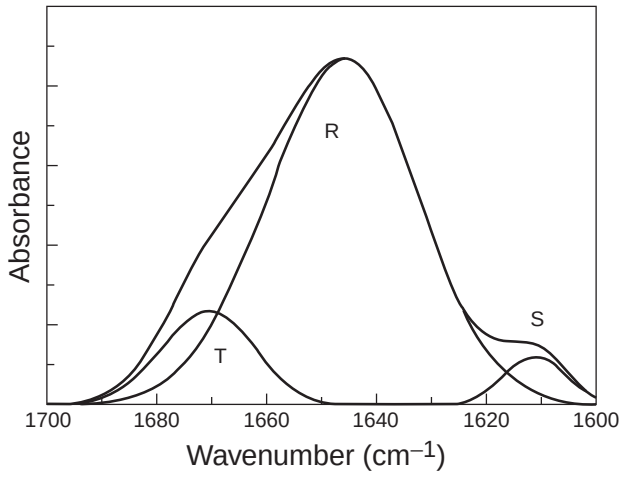

Figure 37 The curve-fit amide I band of MBP in D_2O.

Table 16 Analysis of the amide I band of MBP in D_2O

Frequency (cm^{-1})	Relative area (%)
1647	86
1671	14

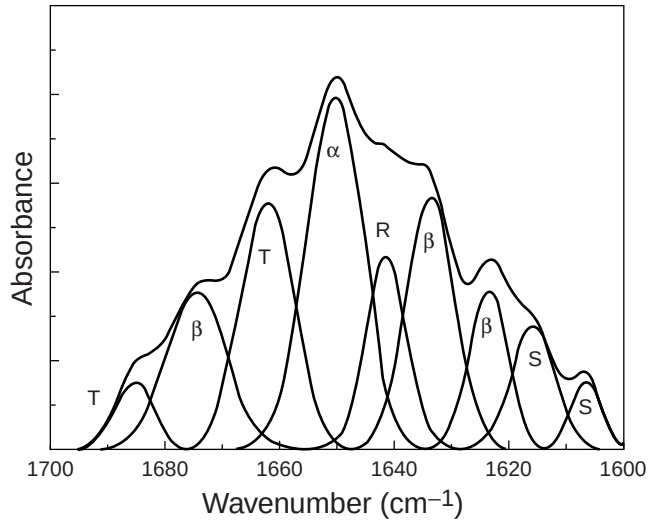

Figure 38 The curve-fit amide I band of MBP in myelin.

observed to the amide I band of MBP when the protein is complexed with myelin. In Figure 38 the two small components at 1606 and 1616 cm^{-1} occur in the frequency range attributed to side chain contributions. The appearance of the band at 1624 cm^{-1} indicates the presence of water-bonded β-structure. Also, the component at 1634 cm^{-1} is assigned to β-structure, with a high-frequency component due to β-structure observed at 1675 cm^{-1}. These bands

Table 17 Analysis of the amide I band of MBP in myelin

Frequency (cm^{-1})	Relative area (%)
1624	8
1634	14
1642	9
1651	22
1663	14
1675	11
1686	3

contribute to 40% of the total band area. This estimation indicates that a notable amount of β-structure is formed when the results are compared to those obtained for MBP in D_2O where no β-structure was observed. The component at 1651 cm^{-1}, contributing to 28% of the total amide I band area, is assigned to α-helix in the protein. Thus, reconstitution of MBP in myelin produces a significant amount of α-helix in the protein in the predominantly lipid environment. Again, none of this type of structure was observed for the protein in an aqueous environment. The small component at 1642 cm^{-1} appears at a frequency normally attributed to random coil. The remainder of the components in Figure 38 at 1663 and 1686 cm^{-1} are assigned to turns and bends in the protein. Thus, the importance of the conditions in which biological molecules are examined in the IR needs to be emphasized. It is observed here with the example of MBP that the protein that this protein is very sensitive to the surroundings. The latter lipid environment more closely mimics the native environment, and thus presumably gives a far better indication of the native structure.

5.4 Animal Tissue

In recent years IR spectroscopy has been used in clinical analysis.[20,21] In particular, high-pressure (pressure-tuning) IR spectroscopy has been developed as a successful technique for the evaluation of malignancy in human tissues. The pressure dependence on parameters such as frequency, intensity and band shape can provide further information about the structural changes associated with malignancy. For example, it is possible to detect cervical cancer arising from a premalignant state termed dysplasia by using IR band differences among normal, dysplastic and malignant cervical cells. A study of the pressure dependence of the frequencies of bands indicates that there are extensive changes in the degree of hydrogen bonding of phosphodiester groups of nucleic acids and the C–OH groups of proteins. Pressure studies also indicate changes in the degree of disorder of methylene chains of lipids in the malignant tissue. The IR spectra of tissue

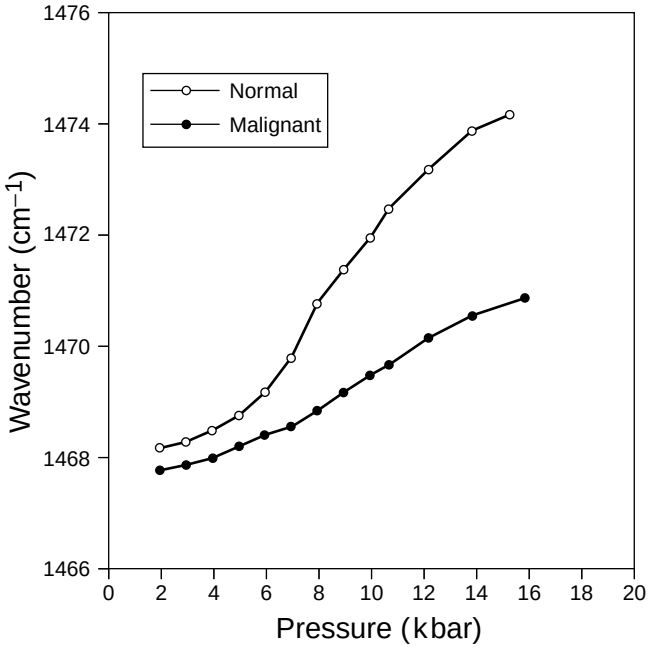

Figure 39 The pressure dependence of the CH_2 bending mode frequency of the methylene chain of lipids for normal and malignant tissue. (Reproduced by permission of P.T.T. Wong from P.T.T. Wong, R.K. Wong, T.A. Caputo, T.A. Godwin, B. Rigas, *Proc. Natl. Acad. Sci. USA*, **88**, 10 988–10 992 (1991).)

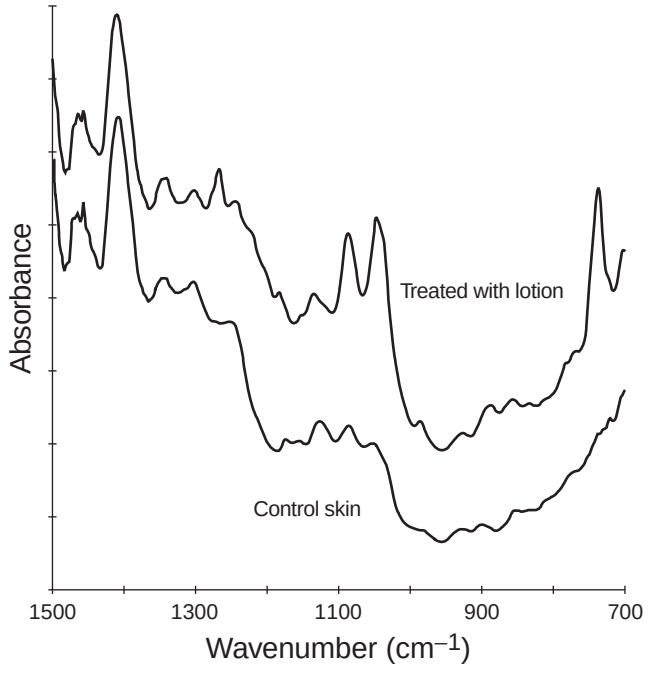

Figure 40 The FTIR microscopic spectrum of skin before and after treatment with emollient followed by washing. (Reprinted from R.G. Messerschmidt, M.A. Harthcock (eds.), *Infrared Microspectroscopy: Theory and Applications*, 1998 by courtesy of Marcel Dekker, New York.)

samples with dysplasia demonstrate the same changes to a lesser degree than those observed for the cancer samples. Figure 39 shows the pressure dependence of the CH_2 bending mode frequency of the methylene chain of lipids for both normal and malignant tissue. Pressure increases this frequency because it induces ordering of the methylene chains in the lipid bilayers, thus increasing interchain interactions. In malignant cervical tissue, pressure induces a smaller shift in the frequency of this mode compared with normal tissue. The difference indicates that in cervical cancer the methylene chains of lipids are more disordered than in normal cervical tissue.

Skin samples are often used to study the interactions of various chemical mixtures, such as lotions, with living tissue.[7] FTIR microscopy offers a valuable tool for these types of analyses as living skin is relatively expensive and difficult to obtain. This technique has been used to study a piece of living skin treated with a conditioning emollient. The sample was then subjected to various washes. Figure 40 shows the FTIR microscopic spectrum of the skin ($200 \times 200\,\mu m$) after treatment with the emollient followed by washing. Figure 40 also shows the spectrum of a piece of control skin. Comparison of the spectra indicates the appearance of a band near $730\,cm^{-1}$ due to the long chain aliphatic hydrocarbon component of one of the lotions constituents. Also, there is the notable appearance of a band at $1040\,cm^{-1}$ due to hydroxy C–O stretching related to one of the active ingredients in the emollient. This example demonstrates how an IR microscope can be utilized as a probe for measuring the presence of chemicals on skin.

5.5 Carbohydrates

The structure of monsaccharides has been successfully studied using IR spectroscopy. For example, IR spectroscopy can be used to clarify the structure of the simple sugar glucose. Although glucose exists almost entirely in the cyclic form, in solution it appears to be in equilibrium with a minute amount of the noncyclic form. Figure 41 shows these glucose structures. IR spectroscopy has been used to show that the acyclic form of this sugar is not present to any appreciable extent because it equilibrates to give different cyclic forms. The acyclic form of the glucose has an aldehyde group present and should

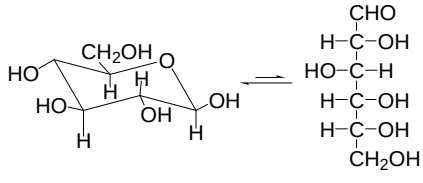

Figure 41 Glucose structures.

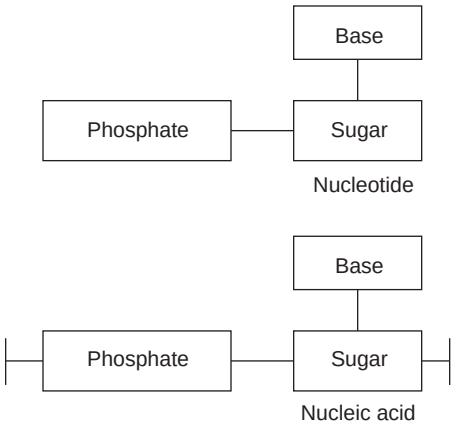

Figure 42 Nucleotides and nucleic acids.

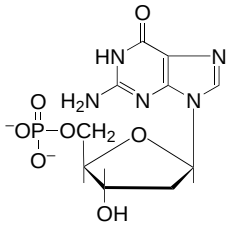

Figure 43 The structure of GMP.

obviously produce a significant carbonyl absorption. As the spectrum does not show a carbonyl absorption it is clear that the acyclic form cannot be present.

Nucleic acids are vital molecules which carry the genetic code and are responsible for its expression by protein synthesis. The two types of nucleic acids are deoxyribonucleic acids (DNA) and ribonucleic acids (RNA). The components of nucleic acids can be obtained by hydrolysis. Partial hydrolysis of a nucleic acid produces nucleotides, which consist of a base, a sugar and a phosphate group. Nucleotides are monomers for nucleic acid polymers, illustrated in Figure 42. Raman spectroscopy has been used extensively for the investigation of nucleic acids, but it is also possible to study these molecules using IR spectroscopy. The IR spectra of nucleic acids can be divided into the modes due to the constituent base, sugar and phosphate groups. One of the principal deoxyribonucleotides is deoxyguanosine monophosphate (GMP), the structural formula of which is shown in Figure 43. The major IR modes of GMP are listed in Table 18.[22]

5.6 Pharmaceuticals

IR spectroscopy has been extensively used in both qualitative and quantitative pharmaceutical analysis.[23] There is widespread use of FTIR for the identification

Table 18 The major IR bands of GMP

Frequency (cm^{-1})	Assignment
3540, 3512	Asymmetric NH_2 stretching
3456, 3436	Symmetric NH_2 stretching
3331	O−H stretching
3211	NH_2 stretching
3145	C−H stretching
3088, 3028	N−H stretching
2987	CH_2 stretching
1692	C=O stretching
1597	C=C stretching, C−N stretching
1535	C−N stretching, C=O stretching
1479	C−H bending, C−N stretching
1358	Pyrimidine ring vibrations
1234, 1204	C−N stretching, C−H bending
1070, 972	PO_3^{2-} stretching
805	P−O stretching
780	Ring stretching
620	N−H out-of-plane deformation
525	Pyrimidine ring deformation

of drugs, excipients and the raw materials used in the manufacture of pharmaceuticals. FTIR spectroscopy has been used to differentiate the structures of penicillins. The general structure of penicillins is shown in Figure 44, where the R group depends on the nature of the particular penicillin. Table 19 lists the major characteristic IR bands of penicillins recorded in D_2O. Quantitative analysis can be readily performed using the intensity of the β-lactam carbonyl stretching band.

GC/IR is an appropriate technique for drug analysis as it can be used for isomer separation or contaminant detection.[24] The technique is emerging as a complementary method to GC/MS, which is generally accepted as the most definitive method for identification of drugs of abuse in urine, for instance. Amphetamines are one class of drug

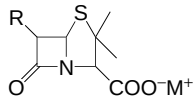

Figure 44 Penicillin structure.

Table 19 The characteristic IR bands of penicillins in D_2O solution

Frequency (cm^{-1})	Assignment
1762–1757	β-lactam C=O stretching
1656–1625	Amide I
1445–1435	Amide II
1601–1595	Asymmetric COO^- stretching
1403–1397	Symmetric COO^- stretching

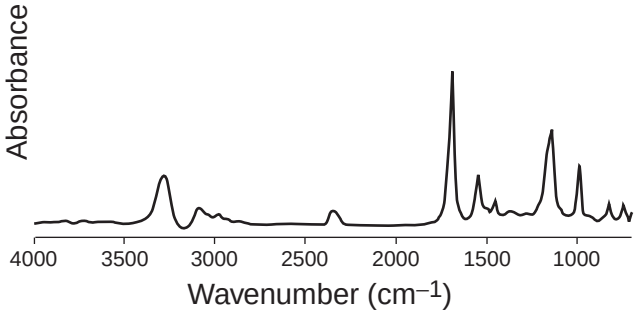

Figure 45 The GC/IR spectrum of an amphetamine derivative. (Reproduced by permission of Preston Publications from K.S. Kalasinsky, B. Levine, M.L. Smith, J. Magluilo, T. Schaeffer, *J. Anal. Toxicology*, **17**, 359–363 (1993).)

which have been successfully differentiated using GC/IR. Amphetamines are structurally similar molecules which can be easily misidentified. In particular, there are some common 'designer drug' derivatives of amphetamine. However, although such similar compounds cannot be differentiated by their mass spectra, there are prominent differences in their IR spectra. Figure 45 illustrates the FTIR spectrum of an amphetamine derivative detected using GC/IR.

5.7 Foods

Near-IR spectroscopy has proved to be a very useful technique for the investigation of foods.[25] The major IR modes in this region in the spectra of foods are due to O–H, N–H and C–H groups. These groups appear due to the presence of proteins, carbohydrates and fats in food. Near-IR has been used in a wide range of studies in food analysis, including the protein content of wheat and meat, analysis of fiber, and oil in mayonnaise. Figure 46 shows the near-IR spectrum of dehydrated tomato soup and this illustrates the modes of common food components.

Wine analysis is also feasible with near-IR spectroscopy. Near-IR provides a relatively straightforward method of analyzing ethanol in wine. The O–H bond in ethanol produces a near-IR mode that is easily distinguished from the O–H mode of water. In the 1980s the wine industry was troubled by several cases of adulteration using diethylene glycol and methanol. Methanol is easily distinguished from ethanol in the near-IR. Figure 47 illustrates that a 1% addition of methanol to a wine sample is easily distinguished from a 1% addition of ethanol. Thus, near-IR spectroscopy has potential as a method for the routine screening of wine.

Near-IR is also an effective approach to the analysis of artificial sweeteners. The near-IR spectrum of saccharin is shown in Figure 48. As aromatic compounds are not very common in food products so modes due to those

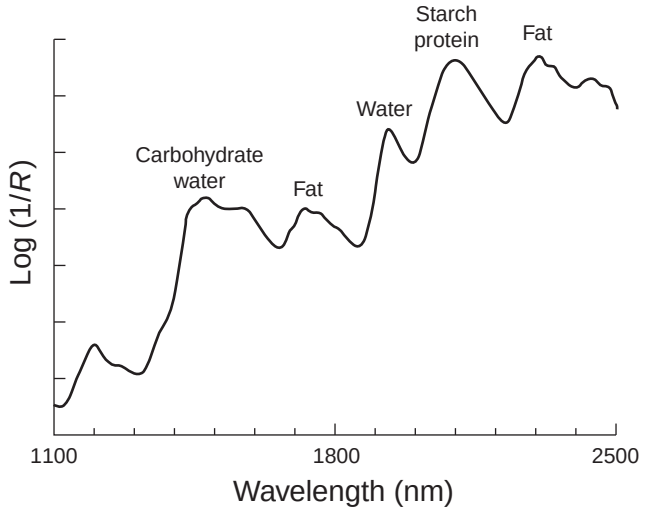

Figure 46 The near-IR spectrum of dehydrated tomato soup. (Reproduced by permission of the Royal Society of Chemistry from C.S. Creaser, A.M.C. Davies (eds.), *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, Cambridge, 1988.)

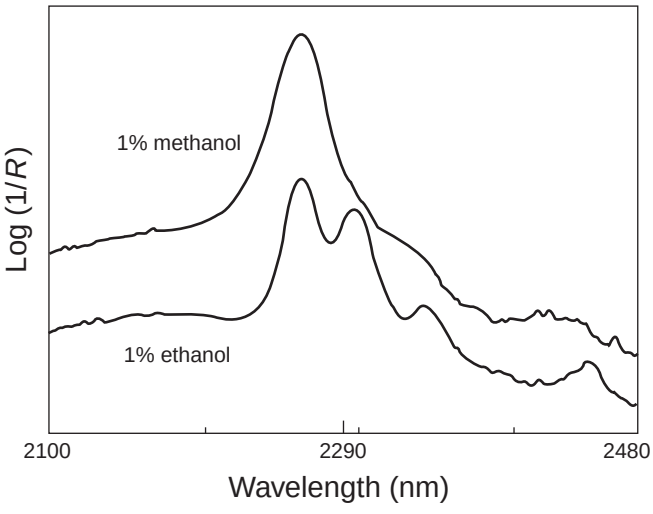

Figure 47 The near-IR spectrum of wine samples containing 1% methanol and 1% ethanol. (Reproduced by permission of the Royal Society of Chemistry from C.S. Creaser, A.M.C. Davies (eds.), *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, Cambridge, 1988.)

compounds are useful. Although saccharin is found at low concentrations in foods, it is normally added as a concentrated solution and this can be monitored by near-IR measurements in the presence of other ingredients.

5.8 Plant Material

FTIR spectroscopy has emerged as a suitable tool for the investigation of plant material.[7] For instance,

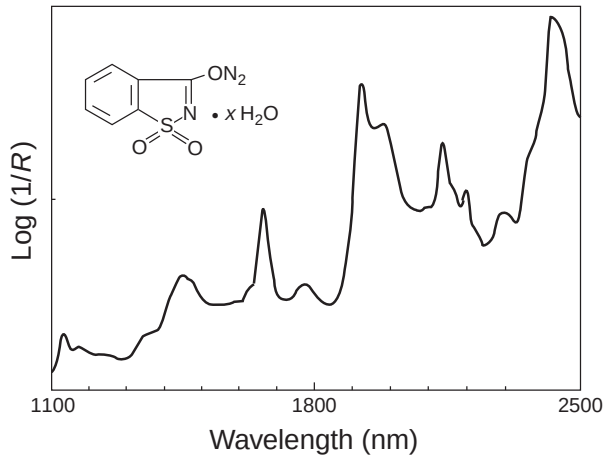

Figure 48 The near-IR spectrum of saccharin. (Reproduced by permission of the Royal Society of Chemistry from C.S. Creaser, A.M.C. Davies (eds.), *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, Cambridge, 1988.)

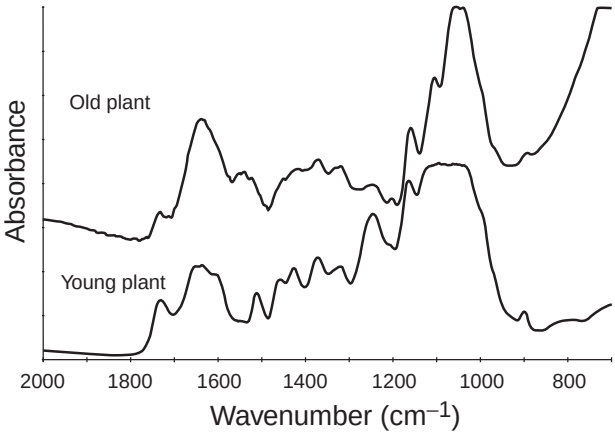

Figure 49 The FTIR microscopic spectra of old and young leaves of alfalfa. (Reprinted from R.G. Messerschmidt, M.A. Harthcock (eds.), *Infrared Microspectroscopy: Theory and Applications*, 1998, by courtesy of Marcel Dekker, New York.)

it can be used to examine the quality of animal food, such as alfalfa. There is a relationship between the quality of alfalfa as a food supplement and the plant age. IR microscopy can be used to compare an old plant with a young plant. Figure 49 shows the spectra of an old and young leaf (both $100 \times 100\,\mu m$). At first sight it is difficult to discern any major differences between these spectra. However, spectral subtraction illustrates some important differences. Figure 50 shows the difference spectrum obtained by subtracting the spectrum of the young plant from that of the older plant. Positive bands indicate a higher relative concentration in the older leaf and negative bands indicate a high concentration in the younger plant. A carbonyl band near $1700\,cm^{-1}$ indicates a high

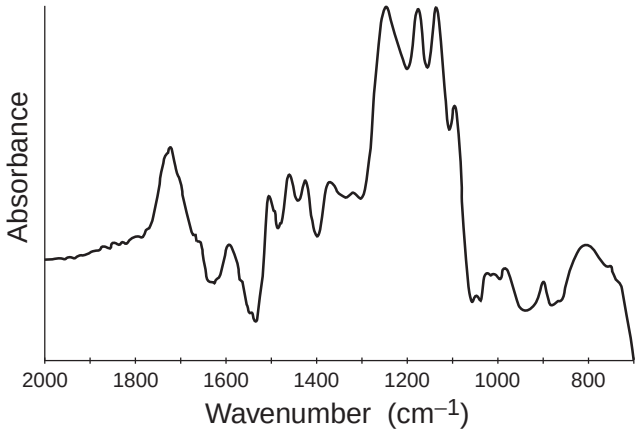

Figure 50 The difference spectrum of young alfalfa from old alfalfa. (Reprinted from R.G. Messerschmidt, M.A. Harthcock (eds.), *Infrared Microspectroscopy: Theory and Applications*, 1998, by courtesy of Marcel Dekker, New York.)

concentration in the older plant. A higher protein content in the young plant is suggested by the negative amide I and amide II bands at 1650 and $1545\,cm^{-1}$, respectively. Thus, Figure 50 tells us that there is a significant difference in the cellulosic nature of the two samples, with the older plant having a higher relative cellulose content.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflectance
CD	Circular Dichroism
CLS	Classical Least Squares
DAC	Diamond Anvil Cell
DRIFT	Diffuse Reflectance Infrared Technique
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/IR	Gas Chromatography/Infrared
GC/MS	Gas Chromatography/Mass Spectrometry
GMP	Deoxyguanosine Monophosphate
ILS	Inverse Least Squares
IR	Infrared
MBP	Myelin Basic Protein
MIR	Multiple Internal Reflectance
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PCR	Principal Component Regression
PE	Phosphatidylethanolamine
PLS	Partial Least Squares
PS	Phosphatidylserine
PTFE	Polytetrafluoroethylene
TFE	Trifluoroethanol
TGA	Thermogravimetric Analysis
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Gas Chromatography/Infrared Spectroscopy • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. B.H. Stuart, *Biological Applications of Infrared Spectroscopy*, John Wiley & Sons, Chichester, 1997.
2. R.E. Hester, R.B. Girling (eds.), *Spectroscopy of Biological Molecules*, Royal Society of Chemistry, Cambridge, 1991.
3. J.R. Durig (ed.), *Chemical, Biological and Industrial Applications of Infrared Spectroscopy*, John Wiley & Sons, Chichester, 1985.
4. T. Theophanides (ed.), *Fourier Transform Infrared Spectroscopy: Industrial, Chemical and Biochemical Applications*, D. Riedel Publishing, Dordrecht, 1984.
5. N.J. Harrick, 'Surface Chemistry from Spectral Analysis of Totally Internally Reflected Radiation', *J. Chem. Phys.*, **64**, 1110–1114 (1960).
6. M.P. Fuller, P.R. Griffiths, 'Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry', *Anal. Chem.*, **50**, 1906–1909 (1978).
7. R.G. Messerschmidt, M.A. Harthcock (eds.), *Infrared Microspectroscopy: Theory and Applications*, Marcel Dekker, New York, 1988.
8. P.R. Griffiths, J.A. de Haseth, L.V. Azarraga, 'Capillary GC/FTIR', *Anal. Chem.*, **55**, 1361A–1387A (1983).
9. C.J. Pouchert, *The Aldrich Infrared Library of Infrared Spectra*, Aldrich Chemical Company, Milwaukee, 1981.
10. J.K. Kauppinen, D.J. Moffatt, H.H. Mantsch, D.G. Cameron, 'Fourier Self-Deconvolution: A Method for Resolving Intrinsically Overlapped Bands', *Appl. Spectrosc.*, **35**, 271–276 (1981).
11. J.K. Ferraro, K. Krishnan (eds.), *Practical Fourier Transform Infrared Spectroscopy*, Academic Press, San Diego, 1990.
12. J. Bandekar, 'Amide Modes and Protein Conformation', *Biochim. Biophys. Acta*, **1120**, 123–143 (1992).
13. J.L.R. Arrondo, A. Muga, J. Castresana, F.M. Goni, 'Quantitative Studies of the Structure of Proteins in Solution by Fourier Transform Infrared Spectroscopy', *Prog. Biophys. Mol. Biol.*, **59**, 23–56 (1993).
14. D.M. Byler, H. Susi, 'Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra', *Biopolymers*, **25**, 469–487 (1986).
15. A. Watts, J.J.H.H.M. De Pont (eds.), *Progress in Protein–Lipid Interactions 2*, Elsevier, Amsterdam, 1986.
16. E. Mushayakarara, I.W. Levin, 'Determination of Acyl Chain Conformation at the Lipid Interface Region', *J. Phys. Chem.*, **86**, 2324–2327 (1982).
17. M. Jackson, H.H. Mantsch, 'Biomembrane Structure from FTIR Spectroscopy', *Spectrochim. Acta Rev.*, **15**, 53–69 (1993).
18. P. Bhandara, Y. Mendelson, E. Stohr, R.A. Peura, 'Glucose Determination in Simulated Blood Serum Solutions by Fourier Transform Infrared Spectroscopy – Investigation of Spectral', *Appl. Spectrosc.*, **48**, 271–273 (1994).
19. B.H. Stuart, 'A Fourier Transform Infrared Spectroscopic Study of the Secondary Structure of Myelin Basic Protein in Reconstituted Myelin', *Biochem. Mol. Biol. Int.*, **38**, 839–845 (1996).
20. P.T.T. Wong, R.K. Wong, T.A. Caputo, T.A. Godwin, B. Rigas, 'Infrared Spectroscopy of Exfoliated Human Cervical Cells: Evidence of Extensive Structural Changes During Carcinogenesis', *Proc. Natl Acad. Sci. USA*, **88**, 10 988–10 992 (1991).
21. B. Rigas, S. Morgello, I.S. Goldman, P.T.T. Wong, 'Human Colorectal Cancers Display Abnormal Fourier Transform Infrared Spectra', *Proc. Natl Acad. Sci. USA*, **87**, 8140–8144 (1990).
22. J.R. Durig (ed.), *Analytical Applications of FTIR to Molecular and Biological Systems*, D. Reidel Publishing, Dordrecht, 1980.
23. R.E. Schirmer, *Modern Methods of Pharmaceutical Analysis*, CRC Press, Boca Raton, 1991.
24. K.S. Kalasinsky, B. Levine, M.L. Smith, J. Magluilo, T. Schaefer, 'Detection of Amphetamine and Methamphetamine in Urine by Gas Chromatography/Fourier Transform Infrared (GC/FTIR) Spectroscopy' *J. Anal. Toxicol.*, **17**, 359–364 (1993).
25. C.S. Creaser, A.M.C. Davies (eds.), *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, Cambridge, 1988.

Mass Spectrometry of Biological Molecules

Beate Fuchs, Klaus Arnold and Jürgen Schiller
University of Leipzig, Leipzig, Germany

1 Introduction **1**
1.1 What is Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry 2
1.2 Coupling Matrix-assisted Laser Desorption/Ionization with Other Mass Analyzers 6
1.3 Biological Molecules and Related Chemistry 7
2 Advantages of the Analysis of Biomolecules **9**
3 Interpretation of a Given Mass Spectrum **10**
4 Preparing Sample–Matrix Mixtures **11**
4.1 How to Choose an Appropriate Matrix 11
4.2 Finding the Appropriate Solvent–Analyte Mixture 11
4.3 Mixing the Matrix with the Analyte 13
5 Recording Mass Spectra **13**
5.1 Important Instrumental Parameters 14
5.2 Laser Power 14
5.3 Standards of Mass Calibration 14
6 Application of Mass Spectrometry to the Analysis of Biomolecules **15**
6.1 Small Molecules and Common Impurities of Biological Samples 15
6.2 Proteins 16
6.3 Carbohydrates/Polysaccharides 18
6.4 Lipids 22
6.5 Other Compounds 27
6.6 Matrix-assisted Laser Desorption/Ionization Imaging 28
7 Quality of Mass Spectra **28**
7.1 Mass Resolution and Mass Accuracy 28
7.2 Signal-to-noise Ratio 29
7.3 Observed Fragmentations 29
7.4 Troubleshooting 29
8 Comparison with other Mass Spectrometric Methods **29**
9 Summary and Method Development **30**
Acknowledgments **30**
Abbreviations and Acronyms **31**
Chemical Abstract Service **31**
Related Articles **31**
References **32**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

Mass spectrometric methods may serve as powerful analytical methods in life sciences and medicine. Mass spectrometry (MS) is comparable in sensitivity to techniques based on chromatography or electrophoresis but has superior resolution. Although there are a number of suitable ionization methods, allowing the analysis of biological molecules that are of high molecular weight and low volatility, there are mainly two ''soft-ionization'' techniques that confer only a minimum of degradation of the analyte: MALDI (matrix-assisted laser desorption/ionization) and ESI (electrospray ionization).

The following article provides a short overview on the advantages and drawbacks of MALDI analysis of biological samples. Whereas most applications of MALDI are in the field of protein analysis, the analyses of carbohydrates and lipids are also comprehensively discussed. It is shown that MALDI is a useful tool especially in the field of lipid analysis, whereas with carbohydrate analysis, still a number of problems have to be solved. MALDI is a rapid and very sensitive (from the low picomolar to the attomolar range) tool that tolerates high amounts of impurities (in contrast to ESI), which are typically present in biological samples. A further advantage is that mainly singly charged ions are formed; this enables MALDI to be applied even to the analysis of crude mixtures although ion yields may differ significantly in dependence on the molecules of interest. In comparison with ESI, MALDI detects compounds of higher molecular weights more sensitively, whereas ESI is more suitable for labile compounds. Therefore, the combination of both methods is, in many cases, the ideal solution for a number of analytical problems.

1 INTRODUCTION

Chemists and biochemists were searching for a long time for a simple method to accurately and sensitively determine the molecular masses of naturally occurring biopolymers such as proteins or polysaccharides. Established methods in this field today are chromatographic techniques (gel filtration) or gel electrophoresis (sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE)). However, both methods are relatively time-consuming and do not provide high mass accuracy.[1]

MS is superior to all other techniques for mass resolution. Highlights in the development of MS, showing its unusual high mass resolution, are the exact determination of atomic weights and the discovery of new isotopes of a given element. The main problem in MS, however, is that the sample has to be evaporated to form intact, isolated, ionized molecules in the gas phase.

Unfortunately, because of their high polarity and often extremely high molecular weight (and the resulting low volatility), most biopolymers cannot be ionized by the ''classical'' electron impact (EI) mass spectroscopy method,[2] which is widely used for the analysis of small molecules with molecular weights up to about 1000 Da.[3] In this approach, most information is obtained from typical fragmentation patterns and the intensity of the molecular ion is often very weak or is even completely absent for biopolymers.

Since the mid-1970s, new volatilization and ionization techniques that also allow the analysis of high-molecular-weight polar biopolymers have been developed. Although mass spectrometers were already being manufactured during or just after World War II by several companies, some time passed until the EI technique was extended with the invention of the field ionization (FI) and field desorption (FD) techniques in which the sample was ionized by a strong electric field gradient, resulting in less fragmentation.[4] Although the molecular masses of polymers that could be analyzed by this technique were rather low, this was the initial step toward making MS also suitable for biological samples. At much the same time, the potential of the technique of chemical ionization (CI) (discovered by Thompson in 1913) was recognized. In this approach, the analyte is ionized through its reaction with gas molecules (e.g. ammonia).[5] This technique is much more gentle than EI and generates less fragment ions, allowing the determination of the molecular weight by the detection of the molecular ion. Secondary ion mass spectrometry (SIMS),[6] plasma desorption (PD) mass spectrometry,[7] and laser desorption (LD) mass spectrometry[8] soon followed. Fast atom bombardment (FAB) is very successful for the analysis of highly polar, low-volatility compounds but has the disadvantage of producing a large amount of fragmentation.[2] An overview of the most important ionization techniques, their benefits, and drawbacks is given in Table 1.

Two rather new ''soft-ionization'' techniques[9] (i.e. with only a very small amount of fragmentation) are ESI[2,10] and MALDI,[11] which now are successful MS methods since both are suitable for the analysis of large biopolymers. A further advantage is the rather simple and less-expensive experimental design allowing simple time-of-flight (TOF) instruments to be used.[3] Therefore, compared to classical instruments, such modern mass spectrometers are rather small.

In 2002, the inventors of these techniques, Tanaka et al.[12] and Fenn et al.[13] were awarded the Nobel Prize for the development of MALDI and ESI, respectively.[14] It should be noted that in the very first ''MALDI'' study, ultrafine cobalt powder was used as the ''matrix''.[12] Nowadays, however, organic matrix compounds are nearly exclusively used for many different reasons. The latter approach was developed by Franz Hillenkamp and Michael Karas in Germany and was published already in 1988.[15]

Although ESI and MALDI yield comparable (often complementary) information, this article mainly deals with the application of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOF MS) for the analysis of biologically relevant samples. MALDI/TOF is primarily established in the field of protein[16–18] and peptide analysis.[19] Terms such as ''proteomics''[20] or even ''peptidomics''[21] are nowadays widely established, although their exact definition still remains to be specified. Nevertheless, its applicability for carbohydrate[22,23] and lipid analysis[24,25] is also described and compared with other MS methods.

1.1 What is Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry

A comprehensive description and the theoretical basics of classical MS and the corresponding experimental equipment are discussed in detail in other articles in this Encyclopedia. Here, only a few comments on the special features of MALDI/TOF MS are given. This is one of the fastest and most accurate methods to determine a number of polymer (mainly biopolymers) characteristics, especially their molecular weight and the corresponding weight distribution.[16]

MALDI is based (in most cases) on the utilization of a matrix that absorbs ultraviolet (UV) light. Although lasers with other emission characteristics (e.g. in the infrared (IR) range) are also available today, the focus is still on UV lasers because they are nearly exclusively available on commercial mass spectrometers. The use of an appropriate matrix was independently introduced in 1988 by Hillenkamp and Karas[15] and Tanaka et al.[12] The technique has become an established tool in biochemical analysis, and the number of publications on the principles and applications of MALDI for biological problems is constantly growing (Figure 1).

There has been a tremendous increase in the number of publications dealing with this topic and it is nearly impossible to open a recent issue of a journal such as *Analytical Chemistry* or *Analytical Biochemistry* without finding at least one paper on MALDI/TOF MS. It is also evident from Figure 1 that there are less papers dealing

Table 1 Overview of the most important ionization techniques used in modern mass spectrometry

Method	General	Advantages	Drawbacks
Gas-phase ionization			
EI	Sample is ionized by EI. Most information is derived from characteristic fragmentation patterns. Mass limit is about 1000 Da	Applicable to nearly all volatile compounds. Can be easily coupled to separation techniques (e.g. GC/MS). High reproducibility	Compound must be thermally volatile and stable. Molecular ions are very often not detectable or extremely weak
CI	A reagent gas (e.g. ammonia) is ionized by an electron beam. The gas ionizes subsequently the sample. Mass limit is ca. 1000 Da	Molecular ion often detectable. Fragmentation strongly reduced in comparison to EI	Mass spectra highly dependent on gas type, gas pressure, and sample. Sample must be thermally stable
Field desorption and ionization			
FD	Sample is evaporized by a strong electric field. Typical mass range about 3–5 kDa	Little fragmentation. Simple mass spectra	Sensitive to contamination. Relatively slow measurements.
FI	Sample is evaporated by electron tunneling. Mass limit ca. 1000 Da	Little fragmentation. Simple mass spectra	Sample must be thermally volatile
Particle bombardment			
FAB	Sample is dissolved in a liquid matrix and ionized by bombardment with fast atoms (often argon atoms). For molecules up to 6 kDa	Rapid and simple. Composition of sample is less important	Analyte must be soluble in the matrix. High fragmentation and high background. Difficult to apply for mixtures
SIMS	Same principle as for FAB, but ions are used instead of fast atoms	Same as for FAB, but with higher sensitivity for higher masses (up to 13 kDa)	Fragmentation still more expressed than with FAB
Atmospheric pressure ionization (spray methods)			
ESI	Sample is sprayed across a high potential difference. Heat and gas flow are used to desolvate the ions. Maximum mass in the MDa range	Very easy to combine with e.g. chromatography. Very low background. Works best for high polar compounds	Very sensitive to contaminants. Multiple-charged species are formed. Apolar compounds difficult to analyze
Laser desorption			
MALDI	Ionization by laser irradiation of an additional matrix compound. Maximum mass ca. 1000 kDa	Rapid and accurate mass determination. Tolerates moderate quantities of impurities	Requires a mass analyzer, which is compatible with pulsed ionization techniques. Low reproducibility

GC/MS, gas chromatography/mass spectrometry.

with MALDI (white bars) than with ESI (black bars). The reader should note that this is not due to the lower importance of MALDI but due to the fact that ESI is also applicable to very small molecules as no matrix is used. Such molecules are difficult to analyze by MALDI only because of the significant contribution of matrix peaks in this mass region. This topic has been recently reviewed, for instance, in the excellent book by Hillenkamp and Peter-Katalinić.[26]

The most important observation for the development of MALDI was the minimal degree of fragmentation (in comparison with other MS techniques like LD) of the analyte achieved by using a suitable matrix compound. The matrix has two important tasks: to absorb the laser energy and to separate the analyte molecules from each other to avoid cluster formation. Therefore, the matrix and the sample are mixed in an appropriate solvent with an approximately 1000–10 000-fold molar excess of the matrix. The main requirements of a ''good'' matrix are its absorption properties at the laser wavelength used, the low extent of fragment ion formation, and good mixing properties with the analyte to give homogeneous crystallization.[27] Finally, the yield of matrix peaks should be very moderate, particularly if the analysis of small molecules is planned.[26] The sample–matrix mixture is subsequently placed onto a special sample probe tip (because it is irradiated by the laser, it is often called the *target*), which, for most instruments, allows the investigation of a number of samples. The required amount of sample is normally in the

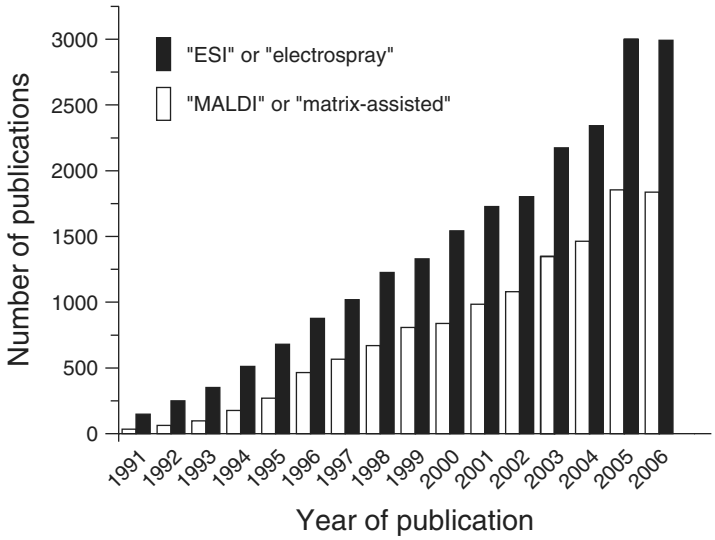

Figure 1 Number of scientific papers containing in their abstract the terms ''MALDI'' or ''matrix-assisted'' (white bars) or ''ESI'' or ''electrospray'' (black bars) and the respective year of their publication. Data were taken from the ''Web of Science'' database.

picomol range. In Figure 2, a typical MALDI sample plate (PerSeptive, Framingham, MA; Voyager Technology) is shown alongside a standard 20-μL Eppendorf pipette tip for comparison.

As MALDI is regarded as a ''high-throughput'' method, larger ''targets'' resembling 384-well microtiter plates are nowadays available from many companies. Already a 0.5–2 μL of analyte–matrix mixture is sufficient to fill a well of the sample plate. The solvent is subsequently removed by air drying (or more rapidly by heating or drying under a slight vacuum) and cocrystallized analyte molecules dispersed within matrix molecules are obtained. Although this is still the most common sample preparation method, it is evident that cocrystals between matrix and analyte are never completely homogeneous. Therefore, the achievable signal intensity depends on the position where the sample is hit by the laser beam (''sweet spot phenomenon''). As this problem does not apply for liquids, the utilization of liquid matrices has been proposed.[28] However, the use of liquid matrices is still associated with rather low mass resolution, high background signals, potential instrument contamination, and poor ionization efficiency and has, therefore, not yet replaced the established matrix compounds.[26]

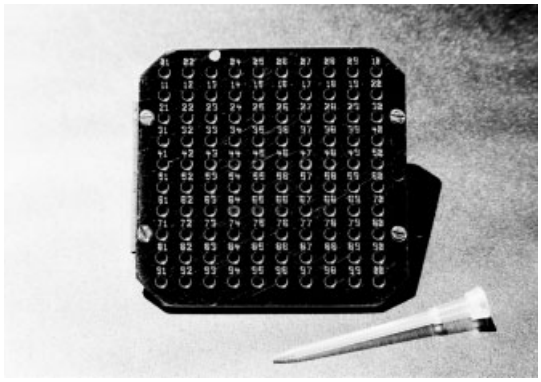

Figure 2 Typical sample plate (Voyager) for MALDI/TOF analysis comprising a 10 × 10 matrix of 1-μL wells etched into the stainless steel with gold cover. The dimensions of the plate are 57 × 57 mm and the individual wells are 2.54 mm in diameter.

A method called *surface-assisted laser desorption/ionization* (*SALDI*) that applies two-phase matrices was also recently introduced.[29] In addition to glycerol, a few liquid matrices, for instance, 3-nitrobenzyl alcohol, nitrophenyl octyl ether, and thioglycerol were already successfully used and detection sensitivities comparable to that by conventional crystalline matrices could be obtained. Finally, there are also protocols that use ''matrix-free'' approaches. This does not absolutely mean that there is no matrix at all, but there is no common organic, UV-absorbing molecule. The introduction of matrix-free desorption/ionization on electrochemically etched silicon surfaces (''DIOS'') has allowed the analysis of analytes with masses as low as 100 Da[30]; it is also known that silylation of the used porous silicon results in much-improved desorption/ionization on electrochemically etched silicon surfaces mass spectrometry (DIOS-MS) performance.[31] It is expected that further improvements of the understanding of the MALDI-ionization process (that is not yet clear at all) will also significantly increase the knowledge on more useful attempts of sample preparation.[32]

A description of the processes occurring at the sample in the mass spectrometer is given in Figure 3.

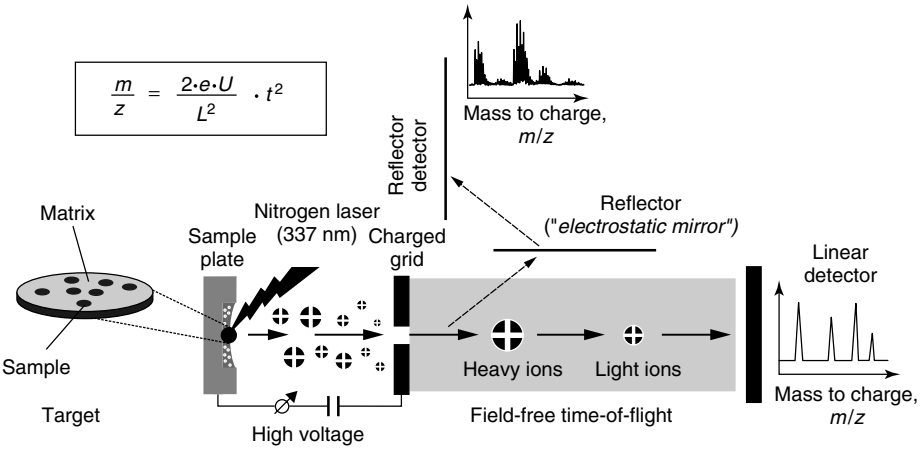

Figure 3 Schematic representation of events occurring in a conventional MALDI/TOF mass spectrometer. The analyte–matrix mixture is evaporated and ionized by laser irradiation. These ions are accelerated in an electric field and then allowed to drift in a field-free pathway. Over this distance, a separation between low-mass and high-mass ions occurs. The ''time-of-flight'' (t) depends on the flight path (L), the mass of the ion (m), its energy ($e\,U$), and the number of charges (z). (Figure was published in ref. 25, © Elsevier 2004.)

When the pulsed laser beam (we focus exclusively on N_2 lasers emitting at 337 nm) hits the probe, the laser energy is transferred to the UV-absorbing matrix, which is partially vaporized and carries intact analyte molecules into the vapor phase. In this expanding gas cloud, all ions ablated from the sample plate have nearly the same velocity of about 500 m s^{-1} and a half-width distribution of approximately 100 m s^{-1} (strongly dependent on the used matrix).[16] During the expansion of this gas cloud, protons and ions like Na^+ are exchanged between the analyte (often a protein, but lipids, many carbohydrates, deoxy ribonucleic acid (DNA), and artificial polymers can be analyzed in the same way) and the matrix molecules, leading to the formation of positively and negatively charged analyte molecules. Unfortunately, despite its importance for MALDI, the process of ion formation is only poorly understood so far, and further physical and chemical investigations are absolutely required.[11,33] However, an important difference between MALDI and the classic ionization techniques is that MALDI provides pulsed ion generation and not continuous ion formation. Requirements for the properties of the laser are stringent: to avoid thermal decomposition of the thermally labile molecules, the energy must be transferred within a very short time. Typical laser pulse widths are in the range 1–100 ns.[16]

Once formed, the ions are accelerated in a strong electric field, with typical accelerating voltages of approximately 20 kV. After passing a charged grid, the molecules drift freely over a field-free space, which is, for most MALDI/TOF spectrometers, in the range of 0.50–3 m. The length of this field-free TOF pathway influences the achievable mass resolution: ions of low mass arrive at the detector faster than do those of higher mass; consequently, ions of different masses can be separated. An appropriate recorder at the end of the flight tube produces a signal upon the impact of each ion group.[16] The TOF spectrum is a simple recording of the detector signal as a function of time. The relationship of time (t) with the square root of the mass-to-charge ratio $(m/z)^{1/2}$ is used to calculate the ion mass from the measured TOF, from which a conventional mass spectrum (intensity over the m/z ratio) is obtained. It must be emphasized that the ionization process does not only form positively charged ions.[11] Some molecules, especially acidic molecules carrying negatively charged groups, are more easily detected as negative ions than as positive ions. Switching from the detection of positive ions to the observation of negative ions is simply performed by inverting the polarity of the accelerating field.

Although a comprehensive theoretical treatment is outside the scope of this review, a few comments on the parameters influencing the mass resolution of MALDI/TOF mass spectrometers are necessary. Mass resolution ($m/\Delta m$) is a measure of a spectrometer's capability to produce separate signals from ions of similar mass. From the equation given in Figure 3, it is evident that the length of the TOF path (L) influences the mass resolution since the separation of the individual ions occurs over this distance. Therefore, resolution can be enhanced using longer flight tubes. However, to construct the small spectrometers with relatively short flight tubes that are preferred, a certain trick is used: ions are reflected at the end of the flight path by a so-called electrostatic mirror and reach the ''reflector detector'' at the opposite end.[16] A schematic comparison between the linear mode and the reflector mode is given in Figure 4.

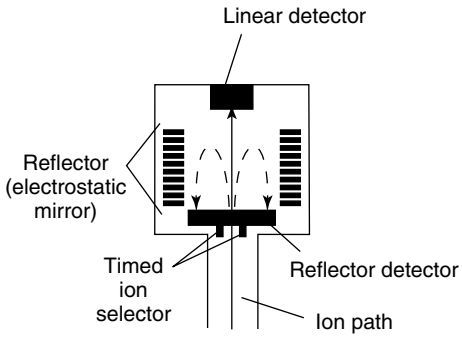

Figure 4 Comparison between different configurations of MALDI/TOF mass spectrometers. The classical ''linear'' mode and the ''reflector mode'' of detection are shown.

Most modern mass spectrometers can be operated in both ways. Unfortunately, the gain of mass resolution in the reflector mode is accompanied by a loss of sensitivity since the total number of ions is spread over a wider range. Mass resolution obtained with common MALDI/TOF spectrometers is relatively poor in comparison with other MS methods, since ions exhibit a broad kinetic energy distribution that is largely a result of the initial velocity imparted to the ions during the desorption/ionization process. This leads to considerably broadened peaks in the mass spectrum. To minimize the influence of the initial velocity of the ions, modern MALDI/TOF spectrometers can be run in a special ''delayed extraction'' (DE) mode, where the velocity distribution of formed ions is minimized and, therefore, the mass resolution is enhanced. In short, a gradient field, not a constant electric field, is used for the acceleration of the ions. This field is high close to the ion source and decreases with the distance. Accordingly, the faster ions experience a weaker field, whereas the slower ions experience a stronger field. If the parameters are carefully adjusted, all ions with the same m/z value arrive at the detector at the same time and peak broadening effects can be avoided. This technique is also termed *time-lag energy focusing*.[34] Further information on this technique, which is now available on many mass spectrometers, can be found in some further reports.[16,34] The spectra of compounds with a relatively low molecular weight (up to 4 kDa) that are shown below were recorded with DE.

Of course, the determination of the molecular weight alone does not provide structural information and tandem mass spectrometry (MS/MS) is often required. Tandem mass spectrometry is the process of selecting an ion, causing it to fragment and obtaining a mass spectrum of the resulting fragment ions. Although MALDI devices with real MS/MS are nowadays commercially available, they are seldom used, as they are expensive. However, if MALDI-TOF devices with a reflectron are available, it is possible to record ''post source decay'' (PSD) spectra that may also be helpful to obtain some structural information on the compound of interest: A given parent ion penetrates to the back of the reflectron and may be easily focused onto the detector. Fragment ions of the parent ion do not penetrate as deeply and consequently, are not as well focused. This problem can be overcome by acquiring several spectra (known as *segments*) at reduced reflectron voltages. The different mass regions are then stitched together to form a single ''PSD'' spectrum representing principally an MS/MS spectrum. This ''PSD'' must be clearly differentiated from ''in source decay'' that occurs directly in the ion source and primarily contributes massively to the ''chemical spectral noise''.

TOF/TOF instruments may be additionally combined with ''collision-induced dissociation'' (CID): an inert gas (Ar or He at a certain pressure is commonly used) is introduced into a special collision cell between both TOF analyzers. The parent ion of interest is selected and transmitted into the collision cell, where collisions between the gas molecules and the parent ions will occur. By this energy transfer, fragmentations of the parent ions occur, leading to the generation of charged and neutral fragments that may help structural elucidation.

Further details on fragmentation processes are available in the excellent book by Hillenkamp and Peter-Katalinić.[35]

1.2 Coupling Matrix-assisted Laser Desorption/Ionization with Other Mass Analyzers

As already mentioned above, MALDI is characterized by a pulsed, but not a continuous generation of ions. Therefore, historically MALDI has always been coupled to TOF analyzers, since these analyzers are most suitable for monitoring the pulsed events taking place in the ion source. However, at least for more sophisticated analytical problems, there has been a strong need to combine the unique properties of the MALDI-ionization technique to other mass analyzers. Some of these reasons are as follows:

- to obtain more accurate mass determinations than achievable by simple TOF analysis;
- to improve the isolation selectivity of parent ions in MS/MS;
- to have access to multistage fragmentation (MS^n).

There were two important problems that had to be solved. The first problem was the pulsed ion generation of the MALDI source, but this problem could be overcome with the invention of more rapid pulsing lasers. Lasers pulsing in the range of kilohertz (thousand times per second) provide nearly a continuous ion stream. The second problem was to avoid the decay of the generated ions prior to their analysis. This problem was solved

by ''cooling'' the ions of interest by collisional damping before their extraction by increasing the gas pressure in the ion source. These attempts are known as *intermediate pressure* or *atmospheric pressure* MALDI.[36]

The comparison between vacuum and (in particular) atmospheric pressure MALDI extends further the knowledge of fragment ion generation and may help identify the structure of unknown compounds. However, peak intensities are generally higher in the vacuum MALDI spectra. A comprehensive comparison of vacuum MALDI and ''atmospheric pressure'' MALDI regarding ''proteomics'' was recently provided.[36]

Although a comprehensive discussion of the different mass analyzers used in modern MS is clearly outside the scope of this article, a short summary is provided in Table 2. A more comprehensive treatment of this important topic is available in the recent book by Hillenkamp and Peter-Katalinić.[35] Different ionization methods as well as vacuum and atmospheric pressure detection were recently compared.[37]

1.3 Biological Molecules and Related Chemistry

In this article, the more general term ''structural biology'' is used to describe the molecules responsible for the highly ordered structures of living beings, e.g. the membranes of cells and connective tissues.[38] Interestingly, a relatively small number of similar, ubiquitous molecules achieve the structural integrity seen in living organisms. This article gives a short overview of such basic molecules, including lipids (particularly phospholipids (PLs)), proteins, and carbohydrates (especially polysaccharides).

1.3.1 Lipids

All higher organisms comprise a large number of cells that are highly differentiated into specific cell

Table 2 Overview of the most important mass analyzers used in mass spectrometry

Instrument	m/z Range	Mass resolution	Mass accuracy	Sensitivity	Remarks
Quadrupole (Q)	2–4000	1×10^3	0.1 Da[a]	0.5–1 pmol	Often combined with IT. One disadvantage is that only one ion is measured at a time. Thus, much available information is discarded. This can be overcome by the so-called o-tof, where the fragment ions are accelerated orthogonally and all of them are detected by the TOF analyzer
MALDI/TOF (Reflectron)	Unlimited	2×10^4	20 ppm	1–10 fmol (peptide) 1–5 pmol protein	The TOF analyzer is still the most useful analyzer for MALDI because it is most suitable for pulsed ion generation. The introduction of very rapid pulsing lasers (1000 Hz), however, made coupling with other detectors possible. Nowadays, TOF/TOF devices are also available (large devices and very expensive!)
IT	20–6000	1×10^3	0.1 Da	10–20 fmol	Often combined with Quadrupole Ion Trap (QIT) and often used as inexpensive mass analyzers. Ions can be fragmented by collision with e.g. He gas and the daughter ions analyzed within the trap. Selected daughter ions can undergo further fragmentation, thus allowing MS^n. Linear trap and orbitrap were developed to overcome problems with mass accuracy of standard IT devices
FTICR	>15 000	1×10^6	~0.1 ppm	20 attomol	May be regarded as combination of IT and a superconducting magnet. The strength of the magnetic field is very important and influences MS quality. Therefore a tendency to higher field is evident. Highest recorded mass resolution of all mass spectrometers. However, devices are large and very expensive. Not a high-throuhput method

IT, ion trap; FTICR, Fourier transform ion cyclotron resonance.
[a] Depends on the mass window being used.

types to fulfill individual tasks. However, all cells have certain common structures: cells are separated from the extracellular space by a membrane consisting of a highly ordered lipid bilayer. This cell membrane is essential for the transport of molecules in and out of the cell. The membrane is composed of a small number of different lipid classes **(1–3)**, mainly phosphatidylcholine (**(2)**; PC), phosphatidylserine (**(1)**; PS), and phosphatidylethanolamine (**(3)**; PE). All these lipids possess, besides their relatively hydrophilic polar headgroup, two lipophilic fatty acyl residues. It is very common that saturated and unsaturated acyl residues are located in *sn*-1 and in *sn*-2 positions, respectively (stereospecific numbering).[25]

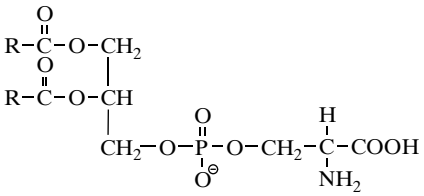

(1)

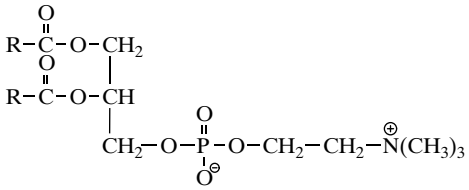

(2)

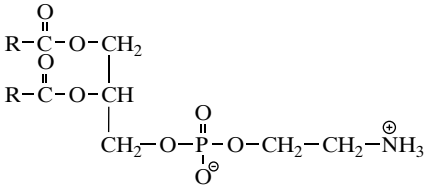

(3)

This marked amphiphilicity enables such lipids to form the characteristic bilayer in an aqueous environment. The formation of this bilayer is a unique feature of PLs. One main problem in the analysis of lipids of natural origin is that they consist of a variety of different fatty acyl residues, derived from the saturated palmitic acid (16:0) up to the highly unsaturated docosahexaenoic acid (22:6).[38] The acronym (x:y) is the usual nomenclature for fatty acids and lipids. Here, "x" represents the number of carbon atoms and "y" the number of double bonds in the fatty acyl residues of the lipid. Sometimes, if no further distinction is required, both residues of a given lipid are combined in this nomenclature. For instance, "36:2" may indicate the presence of two oleoyl (18:1) residues or one stearoyl (18:0) and one linoleoyl (18:2) acyl residue.[39]

Although investigations of lipids have been neglected over a long period of time, there is growing evidence that lipids like phosphatidylinositol phosphates or phosphatidic acid (PA) and diacylglycerols (DAGs) (**(4–6)**, respectively) play an important role in signal transduction processes, where they serve as messenger molecules.[40] The increasing interest in lipids and their physiological relevance is also reflected by the introduction of the term "lipidomics".[41,42]

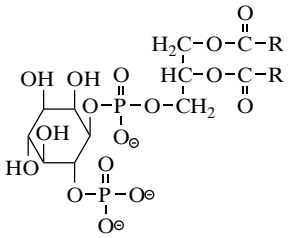

(4)

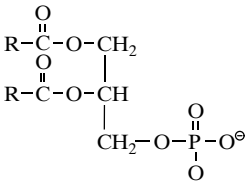

(5)

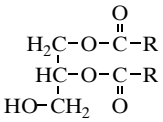

(6)

1.3.2 Proteins

Proteins play an important functional role in all living organisms, especially as biocatalysts and receptors. Additionally, proteins are the macromolecules that allowed the evolution of higher animals, since without the introduction of an active transport mechanism for oxygen (hemoglobin and myoglobin), the size of an organism would be strictly limited by diffusion processes.[38] Finally, proteins form the largest part of the dry weight of the human body, since proteins are also important components of muscles and of connective tissues. In tissues such as skin, cartilage, bones, and tendons, collagen comprises a larger moiety, providing the mechanical stability of such tissues.[43,44] The analysis of proteins and peptides is one of the main applications of MALDI/TOF MS and there are some excellent reviews on this topic.[16,17,45,46] Consequently, protein analysis will be treated relatively lightly here and only some important examples will be given. The interested reader is also referred to some recent reviews dealing with mass spectrometric investigations of the structural biology of proteins.[35,47]

1.3.3 Carbohydrates

Simple carbohydrates like glucose, fructose, or saccharose are important ''fuels'' for living organisms.[43] The polysaccharides mainly fill important structural roles. For example, they constitute the largest part of the exoskeleton of insects and marine organisms in the form of chitin and its deacetylated form, chitosan **(7)**, a positively charged polysaccharide. It is thought that chitosan is the most abundant polysaccharide on earth.[43]

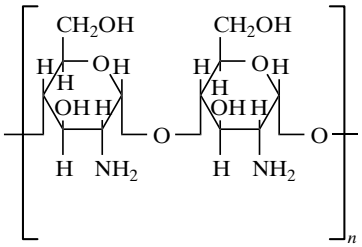

(7)

A similar, but negatively charged, polysaccharide in the flora is alginate **(8)**, which is a widespread component of plants, e.g. brown algae.[43] Polysaccharides like dextran **(9)** or starch are the storage form of carbohydrates in many living organisms; in addition, naturally occurring carbohydrates can be used therapeutically, for example, heparin **(10)** as an anticoagulant. Although the contribution of polysaccharides such as chondroitin sulfate (CS) and hyaluronic acid (HA) ((**11**) and (**12**), respectively) in the connective tissues is relatively low, many diseases are thought to be linked with damage to carbohydrates in the connective tissue, e.g. rheumatic diseases with cartilage degradation.[48] Since the number of patients suffering from rheumatic diseases is constantly growing, such processes are of great socioeconomic significance, and the analysis of relevant polysaccharides by MALDI/TOF will also be discussed.[49]

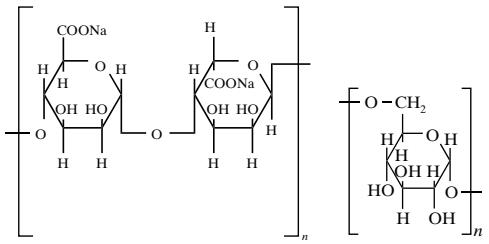

(8) **(9)**

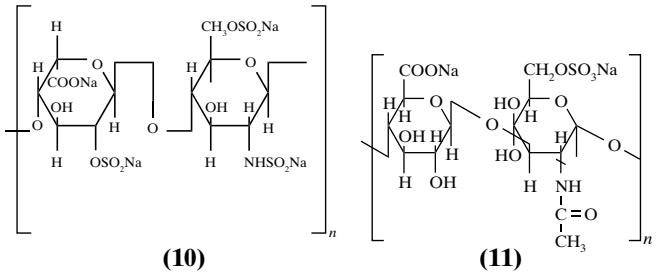

(10) **(11)**

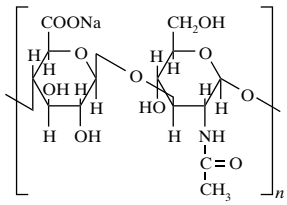

(12)

2 ADVANTAGES OF THE ANALYSIS OF BIOMOLECULES

MALDI/TOF MS offers a nearly unique possibility to analyze very small quantities of biological (detection limits down to the zeptomole (10^{-21} mole) range have already been reported),[50] highly polar, and less-volatile macromolecules, such as proteins and carbohydrates. In comparison with other MS techniques, the extent of fragmentation reactions is extremely low and the molecular ions (i.e. ''quasimolecular'' ions) are detectable in most cases. This is an advantage in the analysis of crude mixtures such as extracts of biological tissues or body fluids. However, one should always be aware that individual molecules in mixtures may be detected with different sensitivities.[35] A further advantage of MALDI/TOF MS is that mainly singly charged molecular ions are formed.[16] This leads to mass spectra that are easy to interpret, since each individual molecule results in a single peak. It makes MALDI/TOF a suitable tool for the analysis of mixtures without previous separation of the individual components, although the suppression of less sensitively detectable compounds by more sensitively detectable compounds should be always considered. This will be outlined below in more detail.

Finally, MALDI/TOF mass spectra can be easily and quickly recorded and no extensive sample clean-up is needed.[51] The commonly used buffers for sample preparation do not markedly influence the quality of the mass spectra. In this sense, MALDI is obviously superior to similar MS methods such as ESI, which is strongly affected by such impurities.[10] Nevertheless, one should always bear in mind that all salts (in particular, polyvalent ions as SO_4^{2-}, Ca^{2+}, PO_4^{3-}, etc.), buffers, and detergents deteriorate the achievable signal intensities and line widths.[52] The solvent of choice is, therefore, distilled water.

3 INTERPRETATION OF A GIVEN MASS SPECTRUM

Although MALDI/TOF mass spectra are generally easy to interpret, a few remarks on the characteristic features of MALDI/TOF spectra are useful. All modern MALDI/TOF mass spectrometers directly calculate the mass-to-charge ratio of the unknown compound from the TOF, which is the actually measured quantity. The number of counts (i.e. the number of ions reaching the detector) is normally plotted against the mass-to-charge ratio. In most standard measurements, the number of counts must not be overestimated since this value is influenced by many parameters, in particular, the applied laser strength.[53] The number of counts is also difficult to reproduce because of the inhomogeneity of the analyte–matrix mixture[54] and, therefore, the inhomogeneity of crystal formation.

The accessible mass resolution of MALDI/TOF spectra depends strongly on the individual compounds: high-molecular-weight compounds do not yield isotopically resolved peaks, but these can be obtained if substances with a lower molecular weight (up to about 4 kDa) are investigated. As a typical example of a compound yielding isotopically resolved peaks, the expanded region of a spectrum of an aqueous solution of adenosine triphosphate (ATP) recorded with 2,5-dihydroxybenzoic acid (DHB) as matrix is shown (Figure 5).

Since ATP possesses negatively charged phosphate groups, positive (Figure 5a) as well as negative ions (Figure 5b) are produced upon laser irradiation. MALDI/TOF spectra often represent the content of ions within the sample, leading to characteristic adduct ion signals. As one can clearly see from the mass difference of 22 Da, both spectra in Figure 5 were recorded in the presence of large quantities of Na^+, which compete with protons for the binding site at the phosphate group ($M - H^+ + Na^+$). The fact that ATP gives more than a single (isotopically resolved) peak is very typical for a biological environment. According to its monoisotopic mass of 551 Da for the disodium salt, the corresponding protonated ion ($M + H^+$; $m/z = 552$) is easily found in the positive ion spectrum. However, this peak is accompanied by four additional peaks arising from an exchange of the protons with Na^+ ($m/z = 574$ and 596) or the two Na^+ with protons ($m/z = 508$ and 530). The presence of five peaks is, therefore, typical for compounds with four negative charges. The negative ion mode provides a simpler spectrum, yielding only three intense peaks that are shifted compared with that seen in the positive ion mode to lower masses (less than 2 Da) as a consequence of the loss of two protons (e.g. $m/z = 528$ instead of 530).

Of course, the sample may be purified prior to MALDI/TOF mass spectrometric characterization and the Na^+ concentration may be decreased. However, according to our experience, it is nearly impossible to remove Na^+ ions completely from biological samples. In cases of mixtures, where, for instance, H^+ and Na^+ adducts overlap and obscure an important mass region of the spectrum, the addition of an excess of CsCl is a convenient method of shifting masses of interest.[55]

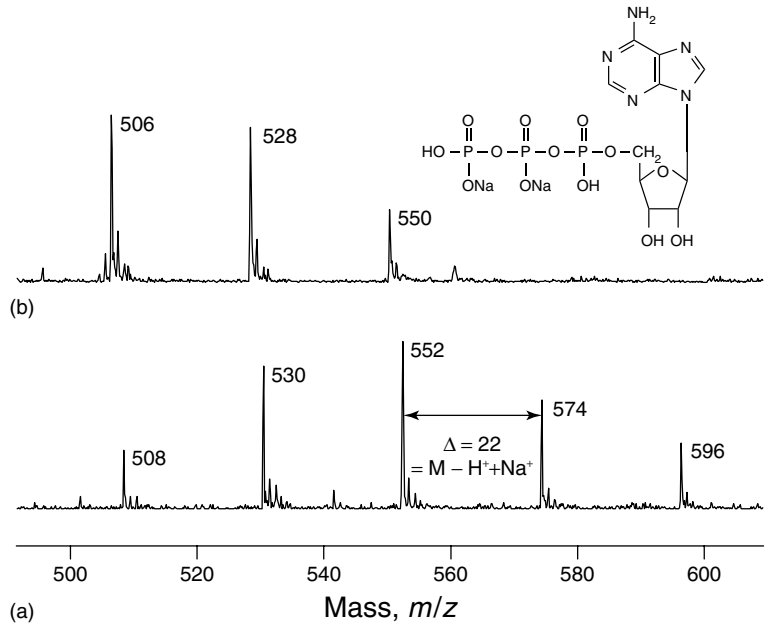

Figure 5 Positive (a) and negative (b) ion spectra of a 0.1 mg mL^{-1} sample of ATP in pure water. Spectra were recorded with a 10 mg mL^{-1} aqueous DHB matrix in the reflector mode of the mass spectrometer.

4 PREPARING SAMPLE–MATRIX MIXTURES

In most instrumental analytical procedures, the quality of the resulting spectra is influenced by a variety of different parameters, which have to be set appropriately; however, in MALDI/TOF methodology there are only two parameters that have a major impact on the quality of the mass spectra. These are the homogeneity of the sample–matrix cocrystals and the applied laser intensity; these determine the extent of fragmentation and, therefore, the quality of the spectra.[56]

Consequently, it is wise not to underestimate the preparation of the analyte–matrix layer because this is the most important step in MALDI/TOF analysis, although this procedure sounds simple.

4.1 How to Choose an Appropriate Matrix

The first MALDI/TOF mass spectra of proteins were successfully recorded by Karas and Hillenkamp[15] using nicotinic acid as a matrix. However, this substance is no longer considered to possess ideal properties for MALDI/TOF analysis since it exhibits a strong tendency to fragment. A large number of useful matrices are now available for different applications[11,27,57] and the MALDI user must carefully choose the matrix that is appropriately adapted for his requirements. Lists of suitable matrix compounds are, for instance, available in the MALDI book by Hillenkamp and Peter-Katalinić.[35]

Since all the compounds suitable as matrices have both advantages and disadvantages, in this article only a small number of common matrix compounds that are currently used in our laboratory and their most important analytical applications are discussed (Table 3).

Since we use a pulsed nitrogen laser emitting at 337 nm, all matrix compounds must show strong absorptions at this wavelength; however, there are also a number of matrix compounds that are suitable for IR lasers (e.g. glycerol or the widely used tris(hydroxymethyl)aminomethane (TRIS) buffer).[27]

Finding the best matrix is a rather empiric process and only a few guidelines can be given. No single matrix compound is suitable for all analytes.[27] Generally, DHB (**(13)**, gentisic acid) is useful for small, particularly highly polar, molecules (e.g. for carbohydrates), since it has only a low tendency to fragmentation and represents a very ''fast'' matrix, i.e. the arising ions possess high initial velocity.[11] DHB dissolved in a solvent such as methanol is also suitable for the quantitative analysis of lipids.[56] Although different isomers of dihydroxybenzoic acid are known, only the 2,5-isomer is suitable as MALDI matrix.[58]

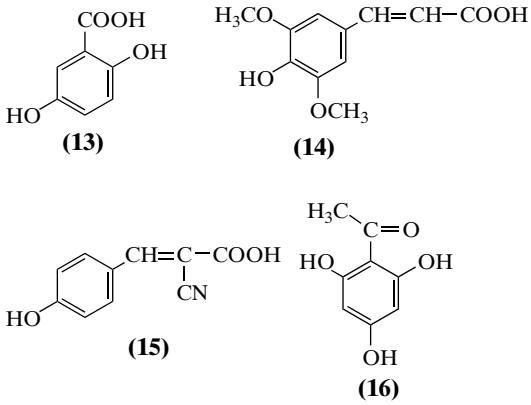

An additional advantage of DHB is that it can be used in the positive as well as the negative ion mode. In the field of protein and peptide research, however, DHB is used less often. In these cases, sinapinic acid (**(14)**; 3,5-dimethoxy-4-hydroxycinnamic acid)[59] or α-cyanohydroxycinnamic acid **(15)** for smaller proteins[60] should be used in the positive ion mode and 2,4,6-trihydroxyacetophenone **(16)** when the negative ion mode is of special interest.[61] Mixtures of different matrices are also sometimes recommended.[62]

One should note that nowadays there are also ''matrix-free'' approaches that enable the evaluation of even small masses without interfering matrix signals. This topic has been very recently reviewed.[63] A quite similar approach is the use of graphite or even pencil lead as ''matrix''.[64]

4.2 Finding the Appropriate Solvent–Analyte Mixture

The importance of the homogeneity of the analyte–matrix crystals in determining the quality of the MALDI spectra is a fact that is often neglected. Since the analyte–matrix cocrystallization from a solution containing both compounds depends on the solvent used, the choice of the solvent is most important. A ''good'' solvent in the context of MALDI/TOF MS is a solvent in which both analyte and matrix are readily soluble. It is also advisable to use highly volatile solvents, whereas the application of solvents with a high boiling point (e.g. dimethyl sulfoxide, DMSO) is not recommended. Since the matrix must isolate the analyte molecules from each other (to prevent aggregation) in addition to allowing ion formation, the sample–matrix mixture should homogeneously crystallize when the solvent evaporates.[16] In a poor solvent, one component (the matrix or the analyte) would crystallize prior to the other one, resulting in separation of analyte-rich and matrix-rich compartments. In many cases, mixtures of different solvents (e.g. trifluoroacetic acid (TFA), acetonitrile (ACN), and water)[53] have to be used, since the matrix is more readily soluble in less-polar

Table 3 Guidelines for the preparation of suitable analyte-sample mixtures for the most important classes of biological molecules

Compound	Matrix/solvent	Conditions	Remarks
Peptides + proteins	Sinapinic acid (10 mg mL^{-1}) in a mixture of TFA, water, and ACN (1:4:5, v:v:v, or similar volume ratios). Sample preferably in pure water. α-Cyano-hydroxy-cinnamic acid can also be used	Cover 1-μ L sample solution with about 1.5-μ l matrix solution. No phosphate should be used for the sample. Preferably use buffers like HEPES or TMEDA	Allow the sample–analyte mixture to crystallize slowly at ambient temperature (the application of a slight vacuum may help improve the homogeneity of crystallization)
Carbohydrates	Use a saturated DHB solution in water (about 10 mg mL^{-1}) for high-molecular-weight carbohydrates or water–methanol (ethanol) mixtures for low-molecular-weight compounds. The use of arabinose osazone is also convenient in a number of cases	Highly dependent upon the chemical structure of the carbohydrate under investigation (the analysis of neutral carbohydrates is much more easier than that of strongly acidic ones):	Since carbohydrates and matrix differ considerably in their solubilities, the mixture tends to separate. Crystallization is normally very poor. The recrystallization of the mixture from a small quantity of ethanol may help improve the homogeneity of crystallization (difficult to perform!)
	2,4,6-Trihydroxyacetophe-none is a useful matrix for the negative ion mode	1. mix sample with matrix directly on the sample plate 2. premix matrix and sample in a small Eppendorf tube The sample/analyte mixture should be dried under a warm stream of air (hair-dryer)	
Lipids	DHB (0.5 M; 77.1 mg mL^{-1}) in methanol (ethyl acetate may also be used, if less-polar lipids have to be analyzed)	Place ca. 1 μ L of lipid sample on the sample plate, followed by an addition of 1 μ L of matrix solution (caution: organic solvents will readily "spride" over the metal surface). Under an air stream, very homogeneous analyte–matrix crystals are formed	Lipid and matrix compounds are readily soluble in the same (organic) solvent. This enhances the homogeneity of crystallization considerably. This makes lipids one of the most easy to analyze substance class
	Sample should be dissolved in chloroform or in $CHCl_3$–CH_3OH mixtures		

ACN, acetonitrile; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-ethanesulfonic acid; TFA, trifluoroacetic acid; TMEDA, *N*,*N*,*N*′,*N*′-tetramethylethylenediamine; DHB, 2,5-dihydroxybenzoic acid.

solvents such as ethanol or ACN, whereas the analyte often requires highly polar solvents such as water (often also in the presence of a certain amount of acid). The higher the difference between the polarity of the analyte and that of the matrix, the more difficult it is to obtain a homogeneous sample–matrix mixture. Table 3 lists some appropriate solvents or solvent mixtures for different substance classes.

Since spectra are mainly recorded in the positive ion mode, a cationizing species is often added along with the matrix to increase the concentration of ionized species.[16] Although biological samples nearly always contain cations such as alkali metals (Na^+ and K^+), the addition (of low amounts) of TFA has been found to be useful for biological samples. Finally, mass spectra of biological samples always reflect the natural ratio between Na^+ and K^+ (e.g. in the used buffer). That means that H^+, Na^+, and K^+ adducts are often simultaneously detectable in biological samples. However, an equilibrium exists between these different forms and that can be strongly

influenced by the addition of ions that occur naturally only at negligible levels (e.g. Cs^+).[55]

4.3 Mixing the Matrix with the Analyte

There are two different methods for preparing a sample spot for MALDI. The first method is to mix sample with matrix and to transfer this mixture subsequently to the sample plate. Although frequently recommended, this method should only be used if pure solvents, not solvent mixtures, are used. With mixtures of solvents, one less-soluble component (usually the analyte) precipitates and is not easily detected. Consequently, direct mixing of analyte and matrix on the sample plate is often advisable.[65] About 1 μL of analyte is transferred to the MALDI target, followed by 1 μL of matrix solution, which should be prepared shortly before use. Matrix solutions should never be stored over several hours since they change with time.

It is strongly recommended to add analyte before matrix; if the matrix is added before the analyte it may start drying before it mixes with the analyte, and this may result in uneven sample–matrix crystals. Touching the sample plate with the fingers or with the pipette tip must also be avoided since this may cause uneven crystallization. Finally, the surface of the sample plate is rather sensitive to mechanical or chemical treatments. During cleaning of the sample plate, actions such as brushing that could damage the surface of the plate should be avoided, as damage can also lead to uneven crystallization. The sample plate should not be put in solutions containing high concentrations of detergents or strongly alkaline or acidic conditions for more than a few minutes.

Recently, the so-called add-on technique has been introduced, which confers enhanced sensitivity.[65,66] Here, only the matrix is primarily transferred to the sample plate. After crystallization of the matrix, the sample is added on top of the matrix crystal and gives a large signal when such a ''hot spot'' is directly hit by the laser and no signal when the laser does not. This technique has been successfully used in the field of lipid analysis.[66] Unfortunately, the fluctuations from shot to shot are high.[56]

As mentioned above, MALDI tolerates considerable amounts of buffer compounds, detergents, and salts. However, some common anions like sulfate or phosphate as well as high amounts of Na^+ strongly decrease molecular ion formation, whereas the influence of organic, nonionic buffers such as *N*-(2-hydroxyethyl)piperazine-*N*′-ethane sulfonic acid (HEPES) is only weak. The influence of different commonly used buffers and salts has recently been thoroughly investigated.[67]

If the sample cannot be desalted prior to the measurement (e.g. by ion-exchange chromatography), the analyte–matrix mixture can be directly ''washed'' on the sample plate. This requires that the impurities are more readily soluble than the analyte itself. In this case, a small amount of water is added to the dried analyte–matrix mixture and is removed after about 10 s by sucking it off with a pipette. This process can be repeated several times. Another important method for the purification of contaminated samples prior to MALDI analysis is drop dialysis.[68] Here, a few microliters of a protein sample are deposited on a membrane floating on top of pure water. This drop is retrieved after about 30 min with a micropipette. However, desalting the sample prior to its application to the sample plate is preferred. Very recently, specially designed pipette tips containing a layer of an ion-exchange resin, which allows the direct desalting of the sample, became commercially available (ZipTip™, Millipore). The protein solution of interest is prepared in diluted TFA. Under these conditions, proteins are positively charged and bind to the negatively charged resin, whereas all further neutral molecules and strong electrolytes as salts are not bound. After one or two washing steps of the resin, an alkaline solution is used for re-elution of the protein from the resin. Using this method, proteins can be easily purified and concentrated.

Highly diluted samples cannot be concentrated easily by evaporating the solvent, as potential impurities would be concentrated too. For that purpose, special ''anchor'' sample plates were developed and are commercially available now. These special targets have small hydrophilic islands, typically of 100–500-μm diameter, placed on a hydrophobic surface.[69] The purpose of the hydrophobic surface is the prevention of the spreading of the sample solution over a large area: during crystallization, the applied solution contracts onto the hydrophilic islands, thereby concentrating the matrix and the analyte into a defined volume. If impurities are significantly less polar than the analyte of interest, a separation between both may be also achieved by this simple procedure.

5 RECORDING MASS SPECTRA

On a modern MALDI/TOF mass spectrometer, the acquisition of mass spectra is very simple because only a few parameters have a major impact on the quality of the spectra. As a guideline for recording the very first MALDI/TOF spectra, it is strongly recommended to use the acquisition files provided by the manufacturer and a sample of known composition. The first aim should be to obtain a spectrum with a sufficient signal-to-noise ratio (S/N) and an adequate mass resolution as a starting point for further optimization of the parameter set.

5.1 Important Instrumental Parameters

All adjustable parameters are highly dependent upon the type of spectrometer used; therefore only some general guidelines can be given. Since modern mass spectrometers often provide the possibility of working in a number of different acquisition modes (positive/negative ion detection mode, continuous or DE and, finally, reflector or linear mode), the chemical composition and the molecular weight of the analyte are of special interest.

For molecules exceeding a molecular weight of about 4000 Da, the easiest way to implement the linear mode with continuous extraction should be used. This method provides the highest sensitivity and only a low amount of fragmentation[16] because the path of flight is short. However, the peaks are broader than in the DE mode because of the significant velocity distribution of ions that is least partially compensated using the DE mode. Since high-mass molecules yield broad peaks through strong adduct formation, mass resolution is not so important. For such analytes, the accelerating voltage should also be set to its maximum value (often 25 kV), providing the highest sensitivity. The so-called ''low mass gate'' is used to avoid the saturation of the detector with low-molecular-weight ions derived from matrix degradation. This gate prevents the activation of the detector until ions greater in mass than the selected value pass the detector. Setting the low mass gate too close to the peaks of interest should be avoided because under these circumstances peak shapes as well as the achievable mass accuracy may be negatively affected. If an expected peak at lower mass is missing or is poorly resolved, the same spectrum should be recorded without the low mass gate and the individual spectra compared. All other parameters that have to be set for the use of DE and the reflectron (both recommended for smaller molecules) are given in the manual of the spectrometer. It is outside the scope of this review to give a comprehensive overview on this topic.

5.2 Laser Power

Besides homogeneous analyte–matrix crystallization, the applied laser power is the second parameter that has a major impact on the quality of the mass spectra. It is strongly recommended to set this parameter as low as possible within the power levels required. Whereas higher laser power often confers an enhanced signal-to-noise (S/N) ratio, it also causes more fragmentation and a decreased mass resolution; using a lower laser power (only a small amount over the threshold level) improves the mass resolution, especially the quality of the baseline. Furthermore, increased laser power leads to enhanced intensities of metal and matrix adducts, confers enhanced formation of doubly and triply charged ions (at one half and one-third of the expected molecular mass) and, finally, leads to the formation of dimers (at the double molecular mass). The required laser power strongly depends on the substance class that is under investigation. For example, it has been shown that the minimum, or threshold, laser irradiance necessary to produce protein ions from a sample is about 10 mJ cm^{-2} for a 10-ns laser pulse width.[16] Near the threshold, ion production falls off with the fifth power of laser irradiance.[16] Generally, a laser intensity of about 10% over threshold is a suitable setting for most measurements.[70]

Whereas oligopeptides give a strong signal even at relatively low laser intensity, acidic carbohydrates require maximum laser intensity. Increased laser strength is also required when a large amount of ionic impurities are present. The applied laser intensity is the main reason why spectra obtained at different MALDI spectrometers are difficult to compare: most suppliers provide only a relative scale of the used laser intensity and, therefore, these settings are difficult to compare with those on another mass spectrometer.[58]

5.3 Standards of Mass Calibration

In MALDI/TOF MS, calibration adjusts the mass assignment for the acquired data. The software of most mass spectrometers includes a default calibration routine (only the matrix has to be specified) that provides an adequate mass accuracy (deviation approximately 0.1%) for many applications. However, if optimum mass accuracy is required, a calibration based on the observed TOF of a known mass of a special calibration standard should be performed. This standard must have a similar mass and should be of similar structure to the analyte.[71] For example, a lipid should serve as calibration standard when lipids are investigated.[56] The calibration standard may be measured individually (external calibration) or be included in the analyte (internal calibration). The basic equations used in the calibration process were described in more detail by Juhasz et al.[72] These authors determined the initial velocity of ions in specific matrices and introduced corresponding correction terms. This correction is particularly necessary if different delay times are used in the DE mode.[72]

For the first experiments to calibrate the mass spectrometer, the use of a known low-molecular-weight substance (e.g. an oligopeptide that gives highly resolved peaks) is strongly recommended. Using this substance, the effects of different settings on the quality of the calibration file should be tested. The use of the linear mode and the reflectron mode requires different calibration files and sometimes (particularly if larger targets are used) differences are also found depending on the sample position on the sample plate.

Mass accuracy is typically given in parts per million (ppm). For example, 50 ppm means that the real mass of an ion measured to 1000 Da has a predicted error margin of ±50 mDa. One should note that the mass accuracy is also determined by the number of data points that are sampled for each mass spectrum as the number of points determines the center of a peak as accurately as it can be determined.[73]

6 APPLICATION OF MASS SPECTROMETRY TO THE ANALYSIS OF BIOMOLECULES

The following short overview of the most important, biologically relevant substance classes describes experiments with isolated components, where possible spectra of biological samples, such as body fluids or cell suspensions, are described as examples of the applications as well as the limitations of MALDI/TOF analysis.

6.1 Small Molecules and Common Impurities of Biological Samples

Although not originally developed for such purposes, MALDI/TOF MS is also suitable for the detection of low-molecular-weight compounds with masses of a few kDa occurring typically in biological samples (e.g. carbohydrates). Such low-molecular-weight compounds (e.g. oligopeptides) were the first compounds where the possibilities of quantifying MALDI/TOF mass spectra were demonstrated.[74] Although low-molecular-weight compounds often interfere with the matrix, such compounds are usually easily detected because their molecular weight is known. Typical examples of such molecules are reduced nicotinamide adenine dinucleotide (NADH), ATP (Figure 5), different kinds of lipids and PLs, and carbohydrates like saccharose or glucose, all of which can be found in most cell and tissue extracts.

Because of its high sensitivity, MALDI also detects a number of impurities in biological samples.[75] A very common impurity, especially in commercially available biochemicals, is poly(ethylene glycol) (PEG). This polymer plays a major role as an additive in a number of biological processes,[76] such as the precipitation of proteins from solutions. Fortunately, PEG can be easily distinguished as it has a typical peak pattern, reflecting the molecular weight of the repeating unit and the distribution of the molecular weight.[34] PEG is mainly synthesized from ethylene glycol by polycondensation or from ethylene oxide by anionic ''living'' polymerization. The molecular weight of its repeating unit $-CH_2-CH_2-O-$ is 44 Da and this can be identified among the peaks of the spectrum.

In Figure 6, the influence of the acquisition parameters and the molecular mass of PEG is shown. The spectra in (a) and (b) were obtained with a PEG 1500 sample, whereas (c) used a PEG 6000 sample.

Spectra (b) and (c) were recorded in the linear mode, and (a) was recorded using a reflectron. It is evident that (a) exhibits improved mass resolution since the reflectron enhances the flight path and compensates for the differences in flight time of ions with equal mass-to-charge ratios but different initial velocity.

The linear mode spectrum is characterized by decreased mass resolution but higher sensitivity. There may be slight differences between the determined mass-to-charge ratios in the linear and the reflectron mode (cf. Figure 6a and b); consequently, different calibration files have to be used for the two modes.[16] Although the use of the reflector mode would be valuable for the analysis of synthetic polymers and, therefore, for industrial research, polymers with a higher degree of polymerization cannot be measured in this way. As shown in Figure 6 for spectra obtained in the same manner but with differing degrees of polymerization of PEG (b and c), the larger polymer (average 6 kDa) has a broad mass distribution that prevents identification of other individual molecular species. However, it has been shown that the application of more sophisticated techniques (such as DE conditions) still allows the acquisition of highly resolved mass spectra of polymers (e.g. PEG 20 000).[34] Again, the matrix used plays the most important role and great care is needed to select a suitable compound (DHB is good for many synthetic polymers).[75] Detergents based on ethylene glycol such as triton, Brij, Myrj, Tween, and Tagat are often used to solubilize membrane proteins. Their mass spectrometric characteristics were investigated very recently, and it could be clearly shown that all these detergents give very typical mass spectra, allowing the differentiation of the individual compounds.[77]

In addition to such high-molecular-weight impurities, most biological samples contain a large number of compounds (buffer salts, detergents, small carbohydrates like glucose, amino acids, etc.) that possess a molecular weight similar to the matrix. Although these compounds can be eliminated to some extent by the low mass gate, a large amount of such molecules may affect the quality of mass spectra by adduct formation and may considerably reduce ion formation.[66]

Finally, it must be kept in mind that most matrix compounds undergo a slow polymerization process, even in the solid state and especially in aqueous solution. For example, cinnamic acid derivatives yield upon aging (owing to the presence of the alkenic residue) intense peaks of polymeric products (between about 11 and 18 kDa). Therefore, it is advisable that all matrix solutions are always freshly prepared. Additionally, the starting

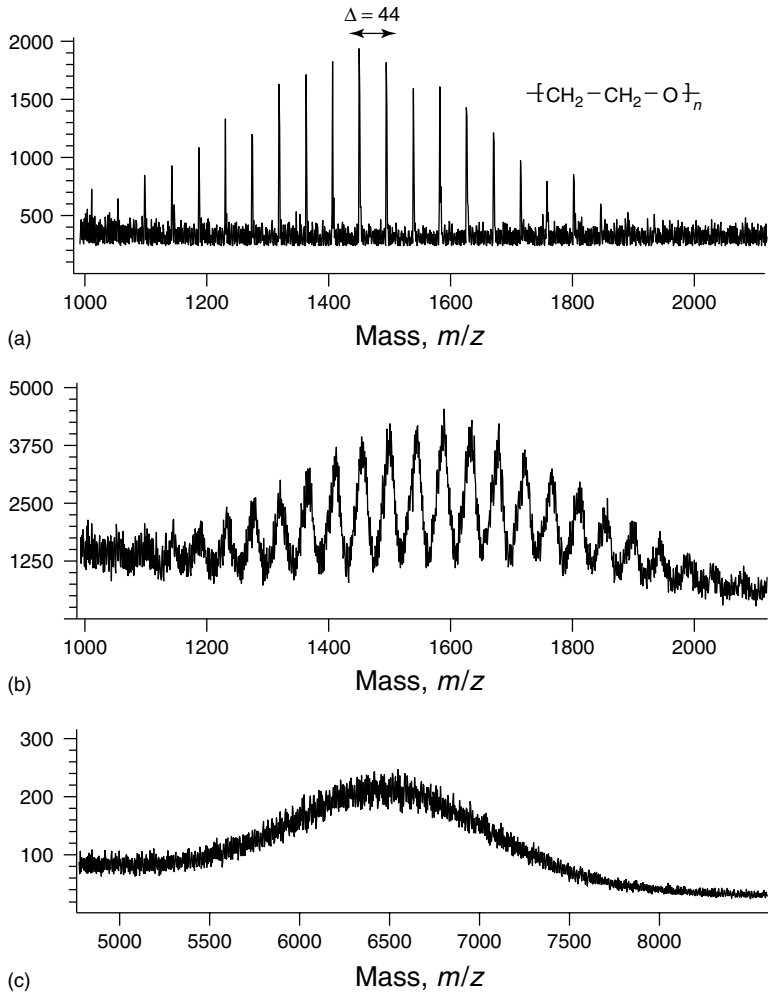

Figure 6 Comparison of the MALDI/TOF mass spectra of commercially available PEG 1500 (a,b) and PEG 6000 (c). Spectrum (a) was recorded with a reflectron and in the DE mode, whereas (b) and (c) were recorded in the linear mode. A 0.1 mg mL^{-1} PEG sample and a 10 mg mL^{-1} DHB solution in H_2O were used. Note the higher intensities but significantly enhanced line widths if the spectra are recorded under ''linear'' conditions.

material should be replaced after a fixed time by a new charge.

It should also be realized that UV-absorbing compounds can be characterized by MALDI/TOF even in the absence of an additional matrix.[78] Under these conditions, the compound under investigation serves simultaneously as matrix and as analyte. Here, the term *LD* spectrum instead of *MALDI* should be used. Unfortunately, this is only possible for compounds of molecular weights up to about 3 kDa. Beyond this size, the energy needed for resonant absorption is greater than the energy required for dissociation of the molecule.[16]

6.2 Proteins

One of the main applications of MALDI/TOF MS is the investigation of proteins, especially oligopeptides.[11,16–18] The main advantages of MALDI/TOF in this context are the simple experimental setup, a rapid procedure, and, finally, the high mass resolution (e.g. in comparison with electrophoresis). These advantages of MALDI/TOF have already led to the introduction of highly automated mass spectrometers for the analysis of unknown proteins.[59] Here, the protein of interest is subjected to different enzymatic digestions, and the digests (usually obtained by treatment with trypsin) are subsequently analyzed by MALDI/TOF (''peptide mapping''). The concept of peptide mass mapping for protein identification by combination of MS and protein sequence database searching was already proposed in 1993 by different research groups.[79] This peptide mass fingerprinting (PMF) approach of protein analysis is also referred to as the ''bottom-up'' approach. Since there are many papers[80] available on this subject, this is not outlined here in more detail. Detailed descriptions of

methodological approaches[81] as well as excellent reviews are also available.[82] Suppliers of mass spectrometers have also significantly contributed to the progress in this field, and nowadays, highly automated devices equipped with nano-LC systems coupled to Robot MALDI-spotters are commercially available.

The reader should note that the reliable analysis of peptide mixtures obtained by enzymatic digestion of a certain protein requires very high mass accuracy as inaccuracy increases the number of hits upon database searching significantly. Although very large databases are nowadays available, not all obtained peptides will be actually available. Therefore, ''de novo sequencing'' of such an unknown peptide by using the interpretation of peptide fragment patterns is often necessary. Finally, isobaric amino acids such as leucine or isoleucine are still difficult to differentiate by MS methods.

As one representative example of a MALDI/TOF mass spectrum of an intact protein, Figure 7 shows the positive ion mass spectrum of a commercially available serum albumin (molecular weight approximately 67 kDa).

This spectrum is relatively simple to interpret and consists only of three different peaks. Whereas the most intense peak at the highest mass-to-charge ratio represents the singly charged molecular ion, the remaining two peaks at about 33 and 22 kDa correspond to the doubly and triply charged molecular ions, which appear with decreasing intensity compared with the singly charged molecular ion. Such multiple-charged ions are found in most cases when proteins are investigated and are affected by the applied laser intensity only to a small extent. By comparison, the formation of protein dimers, which is clearly reflected by the formation of peaks at the double mass-to-charge ratio (for albumin at approximately 134 kDa, data not shown), is highly influenced by the applied laser intensity and can be minimized by using a lower intensity.[17] A high extent of dimer formation even at moderate laser intensity may also indicate that the protein concentration within the sample is too high and that the sample should be further diluted with the solvent or the matrix.

One further recent application of MALDI/TOF MS is the detection of covalently modified amino acids in larger proteins (''posttranslational modification''). More than 200 different posttranslational modifications are currently known, and they give rise to a characteristic mass increase of certain amino acid residues, and thereby, the peptides obtained after an enzymatic digestion of the protein.[83] For instance, the detection of lysine methylation,[84] tyrosine sulfation,[85] and tyrosine phosphorylation[86,87] has been reported. Since phosphorylation events in a biological environment correlate with protein kinase activity, this also allows enzyme activities to be determined by MS.[88]

Unfortunately, proteins containing large proportions of carbohydrates (e.g. glycoproteins) are still rather refractory to MS analysis. The problem here is that the signals of glycosylated proteins are extremely broad and reflect the large amount of heterogeneity in the carbohydrate moiety.[16] MALDI/TOF mass spectrometers are, however, generally able to resolve individual glycoforms only for small proteins containing a limited number of glycans, preferably attached only to one site. One example is ribonuclease B (15 kDa) with five mannose glycans at a single glycosylation site.

Treatment of such glycosylated proteins with specific glycosidases and the subsequent remeasurement of the mass spectra often provides useful information on the number and the chemical structure of the carbohydrates that are removed by the enzyme.[89]

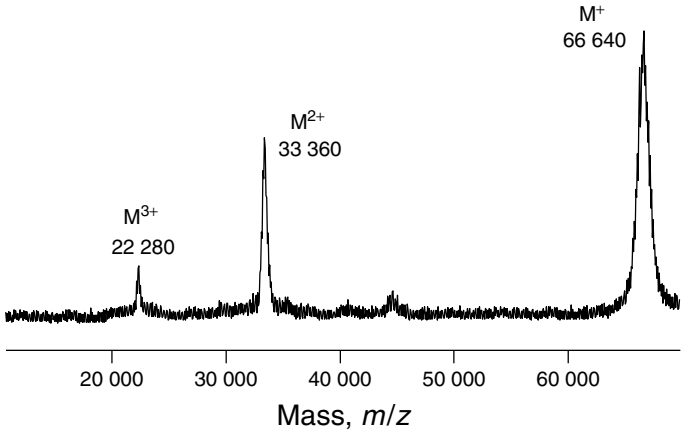

Figure 7 MALDI/TOF mass spectrum of a 1 mg mL^{-1} solution of human serum albumin. For this spectrum 1 μL of protein solution (1 mg mL^{-1}) was transferred to the sample plate followed by 1 μL of matrix solution (10 mg mL^{-1} sinapinic acid in a water–ACN–TFA mixture).

A further important advantage of MALDI in the field of protein analysis is that the quaternary structure of a protein can be elucidated.[90,91] For example, the tetrameric, noncovalently bound subunits of proteins like glucose isomerase and streptavidin could be clearly detected by MALDI/TOF.[16] Dissociation and association of such complexes can easily be followed when different solvents are used. Pure water favors the association of the complexes, whereas the addition of organic (less polar) solvents such as ethanol leads to dissociation of the protein complexes.[16] This exciting application of MS is still in its infancy, but currently, the number of papers dealing with the detection of noncovalent complexes is increasing strongly.[92] Very recently, it was shown that noncovalent protein–protein complexes can be easily analyzed if neutral matrix solutions (instead of commonly acidified solutions) are used.[92] For a comprehensive review of this topic, see the excellent paper by Bolbach.[93]

Traditionally, hydrogen exchange has been used in conjunction with nuclear magnetic resonance (NMR). However, it is now also possible to use MS to measure hydrogen exchange. The method is based on the fact that each protein contains acidic hydrogen atoms capable of exchanging with the solvent (−OH, −NH, −SH). If D_2O is used as the solvent, the heavier mass of the deuterium is incorporated into the protein. This increase in mass can be easily monitored by MS.[94] By measuring the hydrogen exchange rates, conclusions can be drawn about the dynamics of the protein as, primarily, changes of the protein backbone are monitored by this technique. Therefore, this method is very useful for characterizing protein folding or unfolding as well as potential interactions between the protein and the DNA, membranes or surfaces. It is to be noted that the largest complexes studied so far by this technique were viral capsids with molecular masses of the order of 20 MDa![95]

The most important advantages of MS analysis over SDS/PAGE are the simple experimental setup (no gel has to be prepared) and the high mass resolution with enhanced mass accuracy. However, electrophoresis is, in most cases, more sensitive than MALDI/TOF, especially when the very sensitive silver-staining procedure is used.[96] The detection limit of MALDI can be illustrated using an example. A $10\,\mu g\,mL^{-1}$ protein solution can be easily and routinely analyzed and provides a sufficient S/N. Because only $1\,\mu L$ of this sample is required (cf. Table 3), about 10 ng of protein on the sample plate is sufficient for analysis. Assuming a molecular weight of 67 kDa, one can calculate that about 0.15 pmol of protein is enough for MALDI analysis. Some recent reports have shown that femtomole[97] and even attomole[98] quantities of proteins give reasonable MALDI spectra, at least under idealized conditions. However, unfortunately, the calculation above indicates that although very low amounts of analyte on the sample plate are sufficient, relatively concentrated protein solutions have to be used.

As already mentioned, MALDI/TOF is only affected to a low extent by impurities such as detergents (sodium dodecyl sulfate (SDS)) or relatively high salt concentrations and can, therefore, be used for the analysis of crude protein mixtures[99] derived, for example, by enzymatic digestion of biological tissues. As a result, perhaps the most important application of MALDI/TOF is the analysis of protein mixtures. The low amount of fragmentation products and the presence of mainly singly charged molecular ions make the interpretation of MALDI mass spectra very simple. This makes MALDI/TOF a competitor for well-established techniques of protein chemistry such as electrophoresis. For instance, it has been shown that the individual collagen types can be easily differentiated by means of MS subsequent to thermal denaturation and trypsin digestion.[100]

Unfortunately, it must be noted that most biological samples are not sufficiently concentrated for a direct MALDI/TOF mass spectrometric analysis (e.g. the enzymes released by different cell types upon stimulation).[101] Under these conditions, only highly abundant proteins like human serum albumin can be detected; the signals arising from proteins occurring in lower concentrations are suppressed by the highly abundant proteins. One of the major advantages of MALDI/TOF MS is, however, that proteins separated by classical electrophoresis can be subsequently used for MALDI/TOF analysis[102]; in this way, the advantages of protein separation and MS methods can be combined. In recent years, special interest has been paid to such applications and a number of publications on this topic are now available. Nevertheless, the main problem in the field of protein analysis is still the rather low reproducibility of MALDI/TOF mass spectra. Different approaches have been tried to improve the homogeneity of the analyte–matrix crystals, for example by the application of nitrocellulose membranes[103] and electrospray sample preparation.[104]

6.3 Carbohydrates/Polysaccharides

Whereas MALDI/TOF MS has become an established tool in protein chemistry, there are not many reports on the feasibility of MALDI/TOF for the analysis of carbohydrates.[23] This lack of experimental data is primarily because of two reasons. First, many proteins have a rather fixed molecular weight (if posttranslational modifications are neglected), whereas naturally occurring polysaccharides possess a considerable distribution of their molecular weights. This results in a decreased concentration of the individual species and, consequently,

sensitivity for carbohydrate detection is considerably diminished.[105,106] Second, many physiologically relevant polysaccharides do not occur as such, but are mostly esterified with phosphorus or sulfuric acid.[107] The resulting phosphates and sulfates possess a higher polarity than the uncharged polysaccharides, and desorption as well as ion formation is strongly decreased. Because this derivatization leads to serious problems, at least under standard experimental conditions, up to now, mainly low-molecular-weight oligosaccharides like dextran that carry no additional charges have been characterized.[106] A typical spectrum of an aqueous solution of dextran (which is a homopolymer composed of glucose residues that are mainly $1 \rightarrow 6$ glycosidically linked)[43] recorded in an aqueous DHB matrix is shown in Figure 8.

Dextran shows a mass spectrum that resembles the PEG spectrum, with the mass difference of the two individual peaks at 162 Da.[106] This value corresponds to the molecular weight of the glucose subunit (180 Da) minus one water molecule (18 Da), which is eliminated by the formation of the glycosidic linkage. Therefore, the mass spectrum can be directly used to estimate the number of repeating units of the polysaccharide although smaller molecules are more sensitively detected than larger ones. In Figure 8(b), the matrix as well as the dextran solution has been prepared in deuterated water (D_2O) to show the effects of isotope exchange. As clearly seen, the use of the deuterated solvent shifts all peaks to higher masses and multiplies the number of peaks (mass difference is 1 Da) through the competition of hydrogen and deuterium for the number of exchangeable protons within the dextran molecule. This behavior enables the operator to estimate the number of acidic residues within the polysaccharide. The theoretical basis and some applications of such exchange processes were recently discussed.[108] At the same time, the deuterium/hydrogen exchange results in a poorer signal because of increased noise owing to the lower numbers of the individual species.

Although commonly used matrices like DHB can be used for the analysis of small carbohydrates, methods for the improvement of sensitivity and resolution of saccharides are still needed. However, very recently, it has been shown that high-molecular-weight carbohydrates (~70 kDa) can also be analyzed by MALDI/TOF if an extremely high matrix/analyte ratio is used (DHB in about 10^5-fold molar excess over the analyte).[109] Unfortunately, the resulting mass resolution is extremely poor and the only information obtainable is an approximate estimation of the molecular weight. Consequently, for carbohydrate analysis, new matrix compounds yielding higher mass resolution and an enhanced S/N are essential for future applications. One recent proposal for this problem has been the use of arabinose osazone as matrix.[110,111] The idea behind this is that osazones (the reaction products between reducing carbohydrates and phenylhydrazine) are classical well-crystallizing compounds that are often used for the derivatization of carbohydrates. It is, therefore, assumed that osazones possess ideal crystallization properties in a mixture with different carbohydrates. The authors demonstrated that the use of arabinose osazone gave spectra with an improved S/N compared with spectra obtained with DHB.[110] An additional advantage of this application is the very high tolerance of the osazone matrix toward salts and other contaminants.

A very intriguing example for the MALDI/TOF analysis of complex carbohydrates was an investigation of

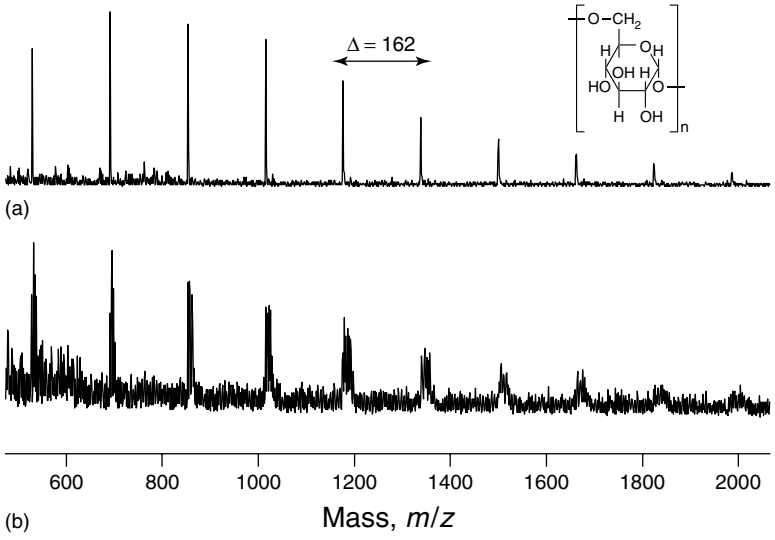

Figure 8 Comparison of the MALDI/TOF mass spectra of commercially available dextran (MW ≈ 1500). Both spectra were recorded with a reflectron and in the DE mode. (a) Matrix and analyte dissolved in H_2O. (b) Matrix and analyte dissolved in D_2O.

milk.[112] Although initial separation of the different carbohydrate classes was required, the authors succeeded in the identification of a large number of different, so far unknown, carbohydrates. It is especially interesting that, even under these nonoptimized conditions, polysaccharides up to a molecular weight of about 8 kDa could be detected and their structure elucidated.

It is even harder to analyze acidic polysaccharides. Although some attempts have been made to characterize strongly acidic carbohydrates such as heparin **(10)**[113,114] or CS **(11)**[115] by MS, all these studies used a previous derivatization of the carbohydrate sample to enhance desorption and volatility. For MALDI/TOF analysis of acidic carbohydrates, it is strongly recommended to add an artificial, positively charged (caused by e.g. a high arginine content) oligopeptide to the carbohydrate solution.[116] This peptide binds noncovalently to the negatively charged groups of the polysaccharide, reducing the negative charge of the analyte. The resulting ionic complexes easily desorb and can be conventionally analyzed by MALDI/TOF. However, all molecular weights have to be corrected for the (known) mass of the peptide. Comparable spectra may also be obtained by the use of positively charged spermine.[117]

Very recently, it was found that strongly acidic oligosaccharides (up to a molecular weight of about 2 kDa) can be directly characterized by MALDI/TOF. Such oligosaccharides are usually produced by the enzymatic treatment of the corresponding polysaccharide with specific or less-specific saccharidases, for example, hyaluronidase[118,119] or chondroitinase.[119,120] If this digestion is carefully performed, the reproducibility of mass spectra is high; it was shown, in the case of selected hyaluronan oligosaccharides, that quantitative evaluation of spectra is even possible.[121] In Figure 9, some selected spectra of HA-6 recorded at different concentrations are shown.

The spectra (A) correspond to the positive ion spectra, whereas spectra (B) represent the negative ion spectra. It is obvious that the negative ion detection mode is more sensitive: considering the lowest concentration of HA-6 (traces, Figure 9c and f), the positive ion (Figure 9c) spectrum is characterized by a poor S/N ratio that is close to the detection limit. It is also to be noted that the peaks detectable in Figure 9(a) (marked with asterisks) at $m/z = 1233$ and 1245 are not caused by HA, but are stemming from the applied DHB matrix ($7\,\mathrm{DHB} - 6\,\mathrm{H}^+ + 7\,\mathrm{Na}^+$). It is well known that DHB, under conditions of laser irradiation, gives some charged oligomers.[122] By defining any arbitrary S/N ratio as threshold, detection limits can be easily obtained. Selected data are shown in Figure 10:

It is evident that the detectability decreases when the mass of the HA of interest increases, indicating that larger amounts are required when HA oligosaccharides with higher masses are investigated. It is also clear that this effect is more pronounced if the positive ion spectra are analyzed: since each polymer repeat unit (composed of one glucuronic acid and one *N*-acetylglucosamine unit) adds one negative charge, the sensitivity difference between the positive and negative detection mode increases if the mass of the investigated HA increases.[121]

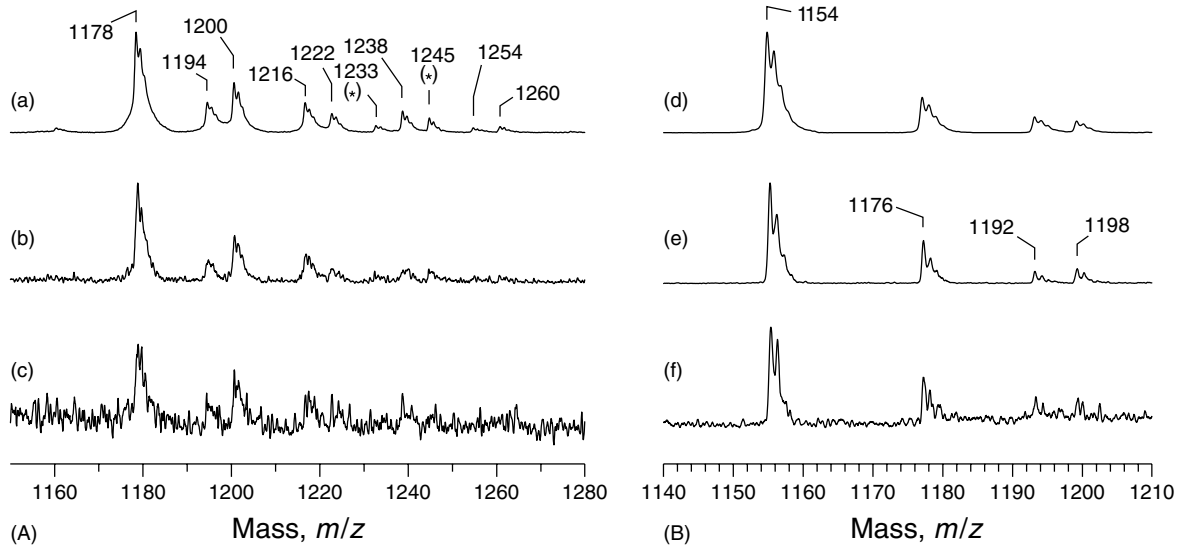

Figure 9 Positive (A) and negative ion (B) MALDI/TOF mass spectra of aqueous solutions of HA-6 relative to the amount of HA on the MALDI target. Solutions of 3.13 ng (a,d), 0.41 ng (b,e), and 0.05 ng (c,f) were deposited onto the sample plate. All spectra were recorded in the linear detection mode for optimum sensitivity. All peaks are marked according to their individual m/z ratios. Peaks arising from the DHB matrix are marked with asterisks. (Figure was published in ref. 121, © Elsevier 2006.)

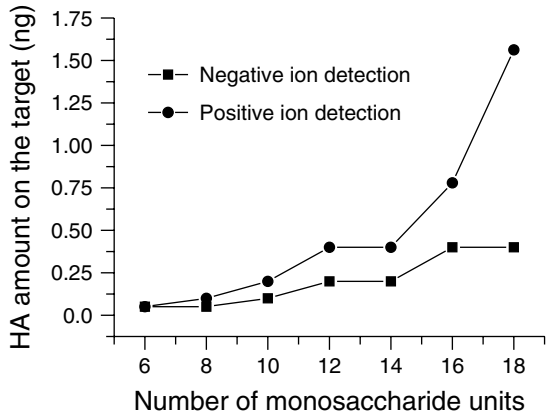

Figure 10 Detection thresholds (amount of HA on the sample plate) relative to the size (number of monosaccharide units) of the individual investigated hyaluronan fragments. The deviations were estimated to be of the order of ±5%. An S/N ratio of 3:1 was regarded as the detection limit. It is to be noted that the considerable deviations between positive and negative detection mode when HA fractions with a higher molecular weight were investigated. (Figure was published in ref. 121, © Elsevier 2006.)

Since our group is mainly interested in the investigation of degradation processes of the extracellular matrix of cartilage during rheumatic diseases, the analysis of negatively charged oligosaccharides would be of great significance for progress in this field. It is known that the strongly acidic polysaccharides of cartilage (HA, CS, and keratan sulfate) are depolymerized during inflammation, under the influence of reactive oxygen species (ROS)[49] and different enzymes.[123] Owing to the large variety of ROS and enzymes that both occur under in vivo conditions, however, the investigation of such processes is very difficult. Nevertheless, MS has proven usefulness for the investigation of degradation processes of polysaccharides of cartilage by ROS on a model level.[124] However, one should always keep in mind that a direct analysis of acidic carbohydrates with molecular weights higher than a few kilodaltons is impossible, and details of previous enzymatic degradation are necessary to obtain defined oligosaccharides. A similar approach based an ESI MS and selective enzymatic digestion was also used to analyze the hyaluronan content in CS preparations.[125]

One promising approach for the cure of degenerative joint diseases is the transplantation of stem cells or chondrocytes (the cartilage cells that synthesize the extracellular matrix) into the joint of the affected patient to replace the damaged cartilage. An alternative is the transplantation of bioengineered cartilage and this interesting topic has been recently reviewed.[126] Unfortunately, it is very difficult to assess under which conditions chondrocytes generate a maximum of extracellular matrix, as the standard histological approaches provide only a semiquantitative information. It was shown recently that the de novo synthesis of CS may be easily determined by MALDI/TOF MS subsequent to cartilage degradation by chondroitinase ABC (an enzyme that cleaves CS A, B, and C) that produces a single unsaturated disaccharide from the CS of cartilage that can be very easily monitored by means of a positive and – more sensitive – negative ion MS.[127] In these studies, it could be shown that not only the amount of CS can be easily determined but information on the ratio between the 4- and the 6-sulfate isomer (that occur both under in vivo conditions) can also be obtained, as the fragment ion spectra from the 4- and the 6-sulfate isomer differ significantly.[128]

This is schematically shown in Figure 11.

Chondroitin sulfate was completely digested with chondroitinase ABC and the negative ion spectra directly recorded from the digest. The protonated and the sodiated forms of the resulting unsaturated disaccharide are shown in Figure 11. Both ions were subjected to further PSD analysis: it is obvious that both spectra differ considerably, although the same molecule, but only different adducts are analyzed. In particular, the peaks at $m/z = 282$ and 300 that appear exclusively in the upper spectrum are of interest because they are derived exclusively from the 6-sulfate and the 4-sulfate, respectively. The reference data were obtained by using electrospray ionization/collision-induced dissociation (ESI/CID).[128] However, such a sophisticated equipment is obviously not absolutely required, as the same information can be also obtained by using the simple PSD technique that is routinely available on all MALDI devices equipped with a reflectron.[129]

However, this method has some disadvantages too: for instance, PSD suffers from relatively low sensitivity, and only moderate resolution can be obtained. Additionally, the need to stitch together spectra derived from several reflector voltages makes calibration rather difficult. It is also important to note that the extent of fragmentation depends significantly on the applied matrix. ''Hot'' matrices such as α-cyano-4-hydroxy cinnamic acid (4-HCCA) usually catalyze considerable fragmentation, whereas ''cooler'' matrices as DHB or Trihydroxyacetophenone (THAP) do not. A more comprehensive discussion of this aspect was provided by Pfenninger et al.[130]

In summary, so far the main application of MALDI/TOF in the field of carbohydrate research is for the analysis of well-defined carbohydrates occurring in complex molecules such as glycolipids or glycoproteins. There are only few reports on the analysis of naturally occurring, acidic oligo or polysaccharides. The main problem, besides the distribution of molecular weights,

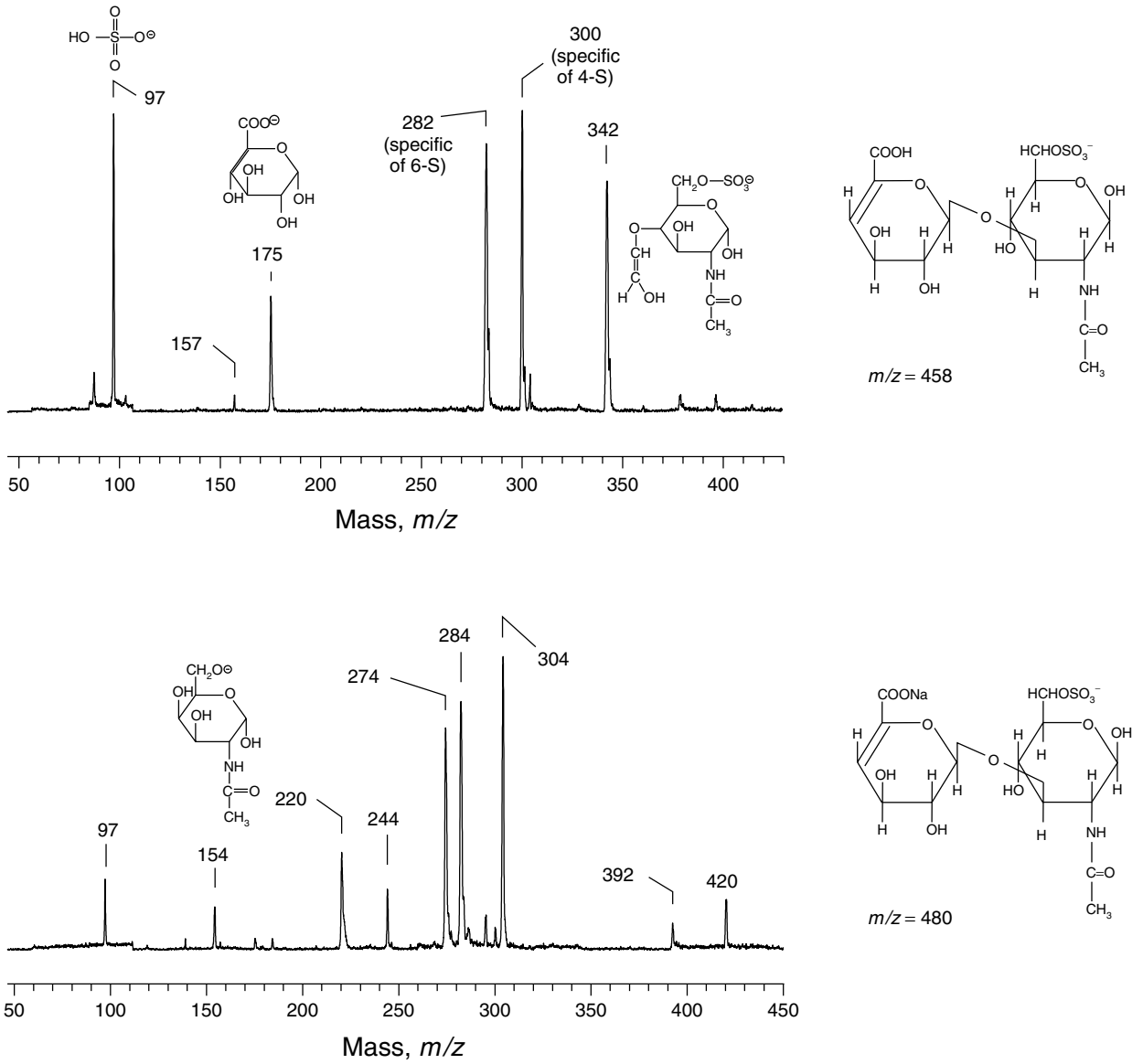

Figure 11 Negative ion post source decay (PSD) MALDI/TOF mass spectra of chondroitin sulfate digested with the enzyme chondroitinase ABC. The fragmentation of the protonated and the sodiated parent ions was investigated. It is to be noted that both spectra are clearly different and provide some peaks allowing the assignment of the 4- and the 6-sulfate. For further details see the articles by Schulz et al.[127] and Zaia et al.[128]

is the low extent of desorption of carbohydrates, leading to low sensitivity. We hope that these difficulties can be minimized by the application of new matrix compounds and further developments in MS devices.[131,132]

6.4 Lipids

Although the importance of lipids has tended to be overlooked, interest has recently increased with the realization that rarer lipids have important roles in biological modulation.[133] While common PLs like PC **(2)**, PS **(1)**, and PE **(3)** are essential constituents of the cell membranes,[38] rare lipids like phosphatidylinositol phosphates **(4)**, DAGs **(6)**, and PA **(5)** are assumed to play a very important role in signal transduction processes.[40] Finally, highly unsaturated fatty acids, e.g. arachidonic acid, are important for the synthesis of essential molecules like prostaglandins or leukotrienes.[24]

Whereas nearly all available ionization methods of MS have been already used in MS lipid analysis, nowadays, only two of them play a major role, namely, ESI[134] and APCI[135] (atmospheric pressure chemical ionization). Reviews on the advantages and drawbacks of the individual methods are also available.[1] MALDI/TOF

MS is also increasingly used in the field of lipid analysis, although it took some time to explore the advantages and capabilities of this method after some pioneering studies in the middle of the 1990s.[136,137] This is somewhat surprising, as the most important requirement of MALDI/TOF MS, an even crystallization of analyte and matrix, is fulfilled by lipids: these apolar molecules are insoluble in water and, therefore, all operations may be performed in organic solvents only, where a majority of common UV MALDI matrices are also easily soluble. This provides very homogeneous analyte–matrix crystals. In Figure 12, two different wells on a MALDI sample plate containing crystals of an aqueous analyte–matrix mixture and an organic analyte–matrix mixture are shown.

Figure 12(a) shows a typical example of a carbohydrate–matrix (aqueous DHB) mixture with uneven crystallization; Figure 12(b) shows a lipid–matrix mixture for which a 1-μL sample of lipid in chloroform was transferred to the sample plate, and a 1-μL droplet of the matrix solution (DHB in methanol) was subsequently added. It is clear that the latter mixture provides by far more homogeneous crystallization accompanied by good shot-to-shot reproducibility.

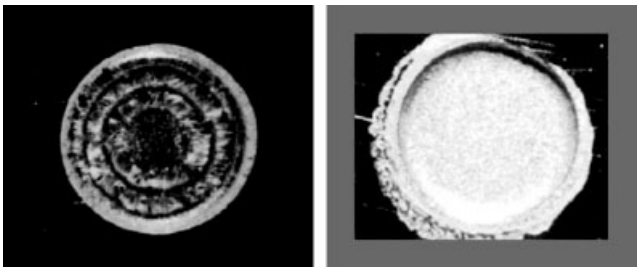

Figure 12 Comparison of two different wells of the MALDI sample plate. (a) The crystallization pattern of a typical aqueous carbohydrate sample with DHB as matrix. (b) A typical dried mixture of DHB (0.5 M in methanol) and a lipid, crystallized from chloroform.

To give the reader an idea about typical spectral patterns of lipids, some selected positive and negative ion (indicated by "+" and "−") MALDI-TOF mass spectra of three physiologically relevant PLs (PC 16:0/18:1, PE 16:0/18:1, and PS 16:0/18:1) as well as triolein are shown in Figure 13.

Considering first the positive ion spectra, significant differences are obvious: PC 16:0/18:1 provides two peaks

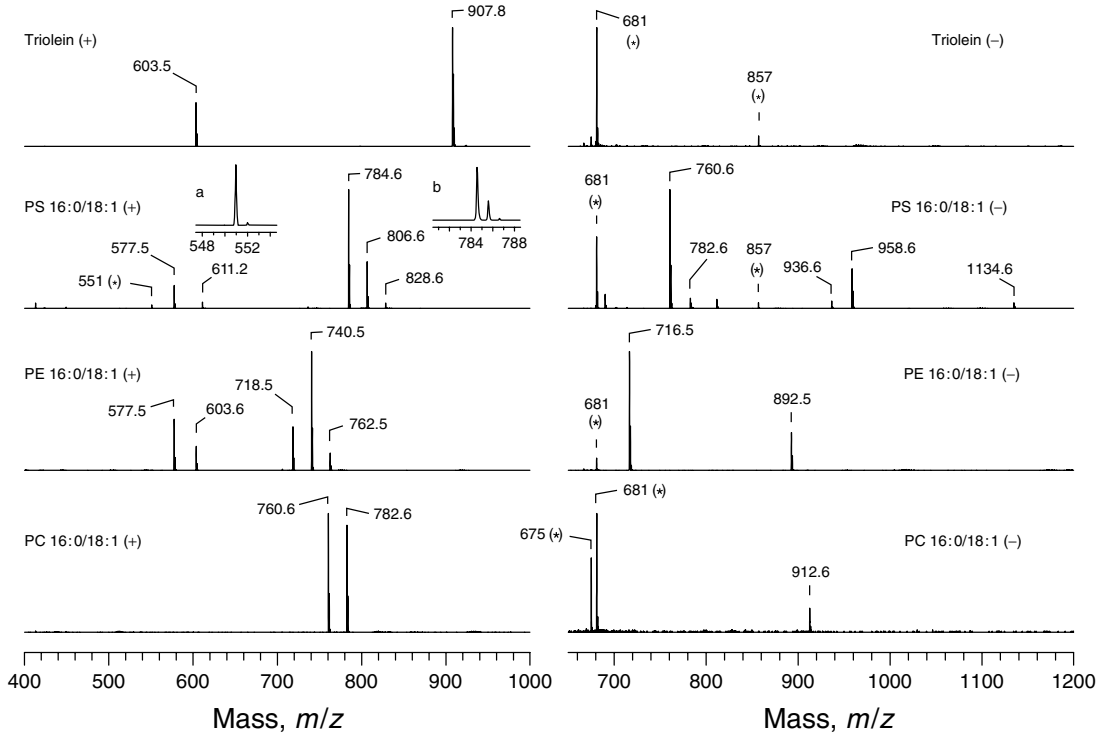

Figure 13 Selected positive and negative ion (indicated by "+" and "−") MALDI-TOF mass spectra of three physiologically relevant phospholipids (PC 16:0/18:1, PE 16:0/18:1, and PS 16:0/18:1) as well as triolein. It is to be noted that all lipids are detectable as positive ions, whereas only acidic lipids provide significant yields of negative ions. The more marked contribution of matrix signals (marked with asterisks) to the negative ion spectra is obvious. In the inserts labeled "a" and "b", the expanded regions of the peak at $m/z = 551$ and 784.6 are shown. It is obvious that the different chemical formula of the DHB matrix and the lipid are clearly reflected by the distribution of the isotopes.

at $m/z = 760.6$ and 782.6 according to the H^+ and the Na^+ adduct of the neutral PC molecule. There are no further signals that would point out fragmentation. This is in agreement with our previous work[25] and the work from other groups.[138] There are not even traces of lysophosphatidylcholine (LPC) that is generated from PC by the loss of an acyl residue. This makes MALDI a useful tool to determine LPC – an important lipid second messenger – in biological samples.[139]

In contrast, the spectrum of PE 16:0/18:1 is more difficult to interpret. This is amazing, since PC as well as PE are both zwitterionic lipids with comparable properties. However, one must consider that PE possesses a $-NH_3$ group with exchangeable protons[25]: the peaks at $m/z = 718.5$ and 740.5 correspond to the H^+ and the Na^+ adduct of the PE as in the case of PC. In contrast, the peak at $m/z = 762.5$ corresponds to an exchange of one H^+ by Na^+, i.e. this peak can be assigned to $PE - H^+ + 2Na^+$. Additionally, in contrast with PC, the PE has a higher tendency to lose its polar head group, and this results in the peak at $m/z = 577.5$. The same fragment (but with lower intensity) is also detected in the case of PS 16:0/18:1 that possesses the same fatty acyl composition as the PE. In the case of the PS, different adducts ($m/z = 784.6$, 806.6, and 828.6) are also detected. This is not surprising, as the PS is an acidic PL and there are different possibilities to compensate its negative charge to obtain a singly charged positive ion. A survey of the mass spectrometric properties of different lipid classes was recently published[25] and their tendency to give fragment ions was discussed.[140] It is notable that the different adducts of a certain lipid have different tendencies to give fragment ions.[42] For instance, it was shown in the example of PC that the Na^+ adducts give much more fragment ions and are, therefore, much more useful for structural investigations than the corresponding H^+ adducts that are dominated by the loss of the polar headgroup.[138] A similar investigation is also available for the fragmentation behavior of PE.[141]

Surprisingly, the triolein does not give a mixture of H^+ and Na^+ adducts, but the Na^+ adduct is exclusively detected at $m/z = 907.8$. Although this behavior of triacylglycerols (as well as DAGs)[56] is well known,[142] only very recently, a potential explanation has been provided.[143] It has been suggested that the H^+ adducts are generated in similar amounts as the Na^+ adducts. However, their stability is much lower in comparison to the Na^+ adducts. Accordingly, the H^+ adducts do not "survive" the TOF distance and yield the fragment ion clearly detectable at $m/z = 603.5$ corresponding to the loss of one acyl residue. Consequently, to decrease the yield of fragment ions, it has also been suggested to perform the sample preparation under alkaline conditions, where the yield of H^+ adducts is much lower.[143]

Generally, the yield of alkali adducts depends strongly on the alkali ion content of the sample. This means, for example, that Na^+ adducts may be replaced by K^+ adducts when an excess of K^+ is added to the lipid solution.[24] This may also be useful when problems with overlapping peaks (H^+, Na^+, and differences in acyl compositions) have to be solved, although, in that case, the use of Cs^+ seems still more advisable.[144]

Generally, the negative ion mode is less frequently used for the characterization of lipids. There are two reasons: (i) not all lipids are detectable as negative ions and (ii) the matrix (DHB is most often used for lipids)[25] provides more intense signals in comparison with the positive ion spectra.[58] Accordingly, because the triolein does not possess acidic groups, it is not detectable as negative ion. The detected signals at $m/z = 681$ and 857 are caused by oligomerization of the applied DHB matrix.[122] Of course, PS ($m/z = 760.6$) and PE ($m/z = 716.5$) can also be detected as negative ions because of their exchangeable protons, respectively. Surprisingly, there are additional peaks at higher masses, for instance, at $m/z = 892.5$ in the case of the PE.[141] This signal is caused by a cluster ion between the lipid (detected as negative ion at $m/z = 716.5$) and the applied DHB matrix (atomic mass of the corresponding sodium salt: 176). Although the formation of matrix ion clusters was not yet comprehensively studied, it is of interest to note that such clusters with PC enable the detection of PC as negative ion at $m/z = 912.6$.[145] This cluster ion formation seems a unique feature of DHB, because by using other matrix compounds as 2,4,6-trihydroxyacetophenone[146] or *para*-nitroaniline (PNA),[147] matrix lipid clusters could not be observed.

Generally, the presence of PC in lipid mixtures in similar amounts as other lipids causes problems, as PC is very sensitively detectable as positive ion because of the permanent positive charge of its choline headgroup. Therefore, its presence may lead to the suppression of the signals of further lipid classes.[122] Unfortunately, this does not only hold for the positive ion detection mode but the detection of negatively charged lipids as negative ions is also influenced by the presence of PC.[148] A detailed investigation of the detection limits of different important PLs was reported by Gellermann et al.[149]

Although the preferred detection of lipid classes with quaternary ammonia groups (PC, sphingomyelin, and the corresponding lysolipids) may also be regarded as an advantage because the spectra are significantly simplified if only selected lipid classes are detected, this is simultaneously a disadvantage: previous separation into the individual lipid classes is required if a more detailed lipid analysis is to be performed. As thin-layer

chromatography (TLC)[150] is an established analytical tool in lipid research, there is an increasing interest in the direct coupling between MALDI and TLC, enabling the direct detection of PLs on the TLC plate subsequent to separation. Such approaches were already established for glycolipids,[151] but only minor attempts were made to develop a suitable method of PL analysis by direct MALDI/TLC coupling.

Two different approaches are so far available: the first one used a home-built orthogonal MALDI/TOF device with an IR laser[152] and glycerol as matrix, whereas the second approach applied a commercially available device equipped with a UV laser.[153] Both reports indicated comparable sensitivities and resolutions. The use of the dye primuline is recommended for staining the individual lipids on the TLC plates because this dye does not bind covalently to lipids and, therefore, does not alter their molecular weights.[154]

Quantitative analysis of MALDI/TOF mass spectra of lipids can be performed by three different methods[25]: (i) by the use of a known internal standard, (ii) by the comparison of the peak intensity of the analyte with a known matrix peak, or (iii) by using the S/N ratio.[155] Method (i) seems most suitable for DAG[56] and PC analysis, whereas (ii) has so far been only used for apolar lipids, especially triacylglycerols.[142] Finally, method (iii) is suitable for the analysis of acidic PLs, e.g. phosphoinositides,[148] and lysophospholipids.[155] Nevertheless, the determination of absolute amounts of different PL classes in mixtures by MALDI/TOF MS is still difficult and spectroscopic techniques as ^{31}P NMR seem superior for that application.[156]

There are obviously several important applications of lipid analysis by MALDI/TOF MS. First, the rough lipid composition of an unknown sample can be easily and quickly estimated. To achieve this, the sample is extracted by known procedures with mixtures of methanol and chloroform, plus hydrochloric acid, if more polar lipids have to be investigated.[157] After separation of the organic and the aqueous layer, the chloroform layer can be immediately applied to the MALDI target without any further purification, since common impurities like salts and buffer components remain in the aqueous layer. However, the organic layer should be washed several times with distilled water.[66]

In Figure 14, this technique was applied to obtain PL profiles of different cell lines (murine stem cells, astroglia, and rat hepatoctyes) and one selected tissue (human liver). Both, positive (A) and negative (B) ion spectra are shown and it is obvious that a clear differentiation of all samples can be easily made at the first glance.

As already indicated above, the detection of PE in mixtures is a major problem as this PL is only scarcely detectable as negative ion,[141] but is suppressed in the positive ion mode by lipids with quaternary ammonia groups that are more sensitively detectable.[122] To be

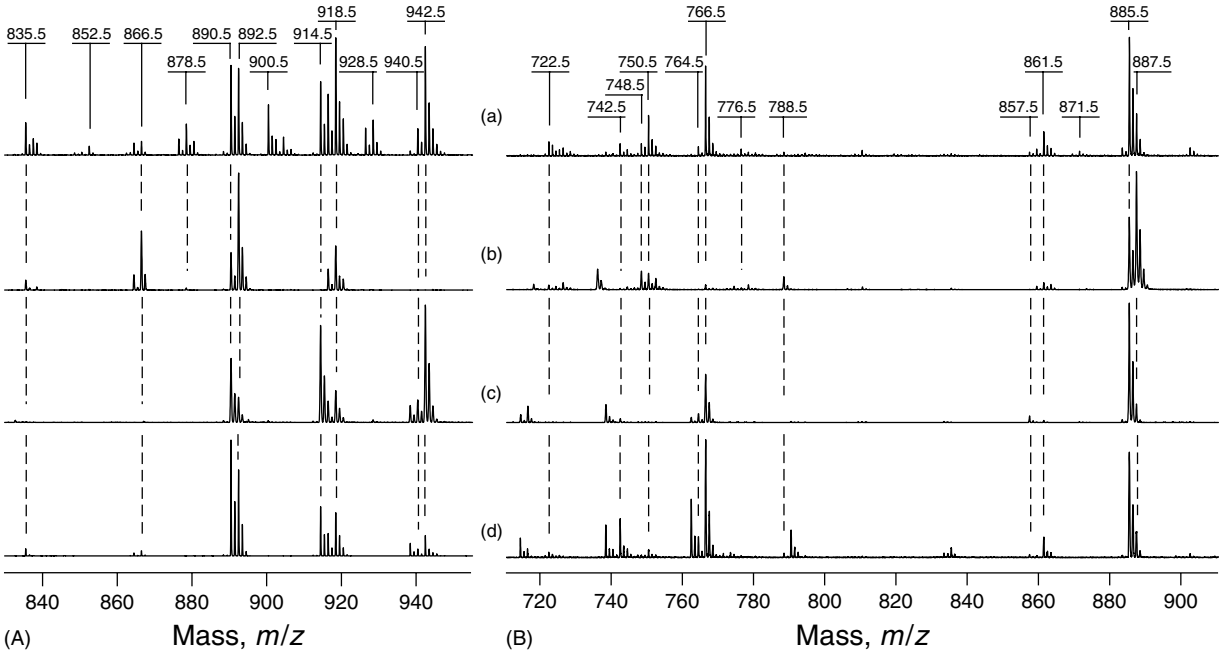

Figure 14 Positive (A) and negative (B) ion MALDI-TOF mass spectra of organic extracts of murine stem cells (a), astroglia (b), rat hepatoctyes (c), and human liver (d). All spectra were recorded with *para*-nitroaniline (PNA) as matrix. To minimize peak overlap between H^+ and Na^+ adducts on the one hand, and differences in fatty acyl compositions on the other, all solutions were saturated with CsCl.[144]

able to detect PE in the negative ion mode and to avoid suppression effects of further negatively charged PL (e.g. phosphatidylinositole (PI) at $m/z = 885.6$)[148] or PL with quaternary ammonia groups, PNA was used as matrix.[147] Furthermore, for the detection of positive ions, an excess of CsCl was added to the matrix solution to assign peaks unequivocally and to avoid overlap between H^+, Na^+, and K^+ adducts of lipids with different acyl compositions.[144]

The mass spectra of the murine stem cells indicate the most pronounced diversity of PLs, particularly in the positive ion mode. Furthermore, stem cells contain high amounts of ether-linked PL (cf. the peaks at $m/z = 878.5$, 900.5, and 928.5) and highly unsaturated acyl chain residues (the peak at $m/z = 942.5$, for instance, represents PC 18:0/20:4). It is obvious that such compounds are nearly completely absent in the spectrum of the astroglia (Figure 14b). In contrast, the mass spectrum of the astroglia extract contains a significant concentration of PC 16:0/16:0 ($m/z = 866.5$) that is rather typical of this cell line and represents a completely saturated lipid species. In the same way, the positive ion spectra of hepatocytes (Figure 14c) and the human liver extract (Figure 14d) can be easily differentiated. It is also obvious that the negative ion spectra allow a very fast estimation of the PE content: for instance, astroglia (Figure 14c) have only a very small PE content, whereas the liver (Figure 14d) and the stem cells (Figure 14a) possess a significant portion of PE, particularly PE 18:0/20:4 ($m/z = 766.5$).

Although from this simple approach only a limited extent of information can be obtained, this type of analysis can be performed within a few minutes and, therefore, information on crude sample composition can be obtained in a very fast and convenient way. Many further cell and tissue extracts were already investigated in this way, including granulocytes,[24] human spermatozoa,[158] animal spermatozoa (bull),[159] and synovial fluid as well as serum[160] from patients suffering from rheumatoid arthritis. A small textbook with the molecular weights of many relevant lipid classes and selected mass spectrometric properties of lipids is also available.[161]

Finally, it should also be noted that MALDI/TOF MS analysis is not limited to organic extracts, but tissue slices may be also directly analyzed. For instance, Jones et al. have shown by analyzing different tissue sections of mouse brain, heart, and liver by MALDI Fourier-transform mass spectrometry (FTMS) that the quantitative data on lipid composition are superior in comparison to chromatographic reference data and provide higher accuracy.[162] Although reports on ''whole cell'' lipid analysis are also available,[163] there is so far only a single paper dealing with ''whole animal'' analysis. *Daphnia Galeata* with a diameter of a few millimeters were analyzed by MALDI/TOF MS by placing the animals directly onto the sample plate.[164] Although reasonable spectra could be recorded allowing the assignment of many lipids, a rather poor S/N ratio was obtained because of poor sample purity.

Another important application of lipid analysis by MALDI/TOF MS is the examination of reactions accompanied by lipid peroxidation, which was primarily performed by ESI in the past.[165,166] It is well known that lipids containing double bonds react quickly with ''ROS'' like hypochlorite[167] or hydroxyl radicals ($HO^{\bullet}$) under the formation of (to a greater or lesser extent) characteristic products.[168]

To investigate such reactions, MALDI/TOF analysis is very useful, because it can provide fast information on molecular weights and quantitative information on the extent of lipid modification.[169] As a typical example, the conversion of PC 18:0/18:1 (its monoisotopic mass is 785.6) to the corresponding chlorohydrine under the influence of the HOCl-generating system myeloperoxidase–H_2O_2–Cl^- is shown (Figure 15). Myeloperoxidase (MPO) is the enzyme that catalyzes the generation of HOCl from H_2O_2 and Cl^- under inflammatory in vivo conditions.

It is evident that chlorohydrine formation[169] can be easily followed by the increase of the peak at $m/z = 840.6$, which corresponds exactly to the molecular mass of the product expected upon the addition of hypochlorous acid to PC 18:0/18:1 (788.6 + 52 (HOCl)), if only the main abundant isotope of chlorine (^{35}Cl) is used for the calculation.[168] As expected, the corresponding Na^+ adduct is also detectable at $m/z = 862.5$. Further studies have also provided evidence that besides the chlorohydrines, LPC (lacking the unsaturated fatty acyl residue) is also generated by the reaction of PC with HOCl.[170] However, the content of double bonds in the unsaturated fatty acyl residue seems to determine the yield of LPC: docosahexaenoyl residues (22:6) provide the most pronounced LPC generation, whereas oleoyl (18:1) residues provide only small amounts of LPC.[170] Similar studies were also performed with plasmalogens that can – despite their rather labile alkenyl ether group – also easily be characterized by MALDI-TOF MS.[171] It must be stressed, however, that the addition of acid (to enhance the yield of H^+ adducts) to the matrix must be strictly avoided, as plasmalogens are extremely sensitive to even traces of acids and decompose under the generation of LPC.[159]

The very last comment on the analysis of lipids seems trivial, but neglecting this fact may give rise to a number of serious problems. This problem concerns details of the handling of the lipid samples. Lipid analysis usually requires the use of chloroform as solvent. Compared with most other widely used solvents, $CHCl_3$ is a rather aggressive chemical that reacts with many

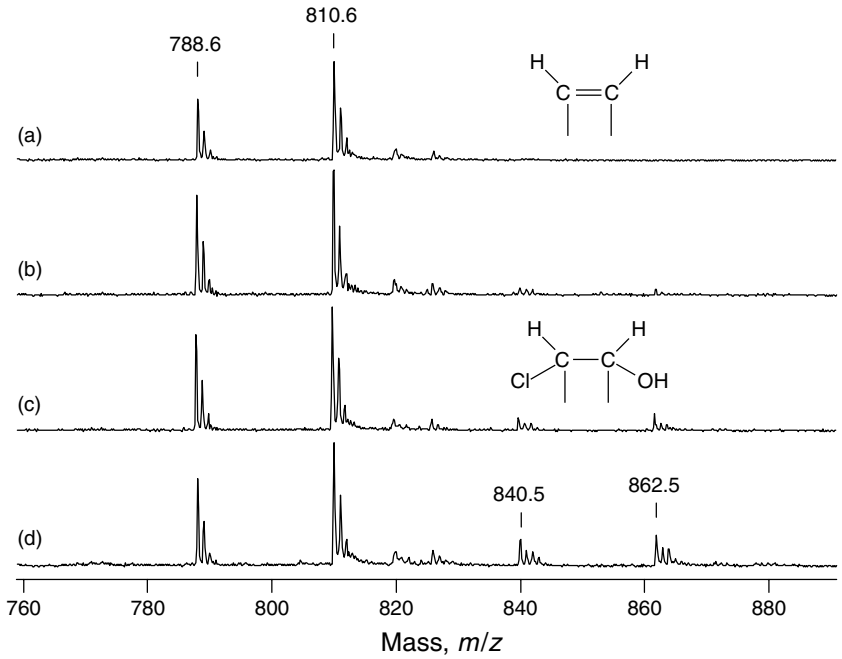

Figure 15 Changes in the mass spectrum of 1-stearoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (trace (a)) after incubation with the myeloperoxidase–hydrogen peroxide–chloride system at pH 6.0. Phospholipid liposomes (0.03 mg mL^{-1}) prepared in 0.14 mol L^{-1} NaCl, 50 mmol L^{-1} phosphate were incubated with hydrogen peroxide (three additions of 0.1 mmol L^{-1} in intervals of 10 min) and MPO (35 nmol L^{-1}) for 60 min (b) or 120 min (c). In trace (d), all additions of MPO and hydrogen peroxide have been repeated after 1 h. The total incubation time in sample (d) was 2 h. All concentrations are final ones. (Figure was published in ref. 169, © Elsevier 2001.)

substances, including polymers. Therefore, it is strongly recommended to avoid any plastic material and to use exclusively glassware as reaction vessels. Additionally, glass syringes instead of plastic pipette tips, should be used. Figure 16 shows the MALDI/TOF (positive ion) mass spectra of $CHCl_3$ in which colored pipette tips had been stored for a few minutes.

The yellow pipette tip particularly (trace in Figure 16a) results in a number of impurities; these most likely correspond to the plasticizers in the material of the pipette tip. This fact should be carefully taken into account even if the molecular weights of the impurities do not fall into the range of the substances of interest. Since such plasticizers give a strong signal, particularly for the positive ion spectra, the intensity of the peaks of interest is considerably diminished.[66]

6.5 Other Compounds

Some studies have examined the feasibility of MS for the investigation of nucleic acids,[172] especially of smaller oligomeric units of ribonucleic acid (RNA).[173] These

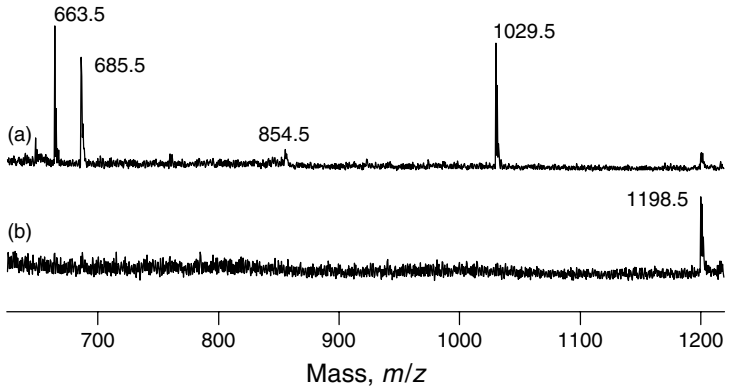

Figure 16 Positive ion MALDI/TOF mass spectra of standard pipette tips, which were incubated for a few minutes with chloroform: (a) extract of yellow tips; (b) extract of blue tips.

compounds are rather difficult to analyze because they are characterized by a significant negative charge density caused by their phosphate residues. Therefore, low ion yields, fragmentation, and signal heterogeneity due to multiple cationization of the phosphate backbone are the main problems that need to be solved for successful performance. Although there are an increasing number of reports on this topic,[174] its comprehensive treatment is outside the scope of this article.

Many studies available so far have used IR lasers as the ionization process by IR MALDI, which is known to be softer in comparison with UV MALDI.[35] Using IR MALDI, DNA samples up to about 20 kDa have already been successfully analyzed.[175] If UV MALDI is to be used, 3-HPA (3-hydroxy-picolinic acid) is the matrix of choice for DNA analysis. Surprisingly, the yield of positive and negative ions is comparable despite the considerable negative charge of DNA and RNA if 3-HPA is used.[176] Nevertheless, sample preparation is still more important than in the case of other substances.[177]

The MALDI/TOF analysis of other substance classes, for instance, carotenoids and analogous compounds, is described elsewhere.[178]

6.6 Matrix-assisted Laser Desorption/Ionization Imaging

Over the past decade, a new analytical method ''MALDI imaging'' for examining biological samples and for visualizing analytical information has evolved.[179] Clearly, this topic is out of the scope of the present article, but some comments seem necessary. For a more detailed review see the article by McDonnell and Heeren.[179]

This new technique allows one to visualize the two-dimensional concentration profiles of defined biomolecules in or on the surface of biological samples, and it is expected to have a great impact on, for example, proteomics and metabolomics research in the future because of the available spatial resolution. One should, however, also note that such attempts have been performed since many years and are, therefore, not completely new. Even about 30 years ago, soon after powerful UV lasers first became commercially available, a new method was developed that permitted the mass spectrometric analysis of biological material, by using a highly focused laser beam.[180] This ''laser microprobe mass analysis'' (''LAMMA'') technique already provided a spatial resolution of about 1 μm and not much progress was made as modern ''MALDI imaging'' devices (using nitrogen lasers with $\lambda = 337$ nm) could provide a resolution of only about 0.6 μm.

Although resolution seems quite limited in comparison with modern microscopic techniques, there are some recent encouraging applications. For instance, by imaging studies of human glioblastoma slices, it is possible to differentiate the tumor region from the ischemic and necrotic areas of the lesion on the basis of localization of different proteins.[181] Very recently, ''MALDI imaging'' has also been successfully used to differentiate normal from the cancerous regions of man's prostate by evaluating specific protein expression patterns.[182] From these examples, it seems most likely that MALDI imaging will have a great future development.[179]

7 QUALITY OF MASS SPECTRA

Usually, a ''good'' mass spectrum should exhibit an excellent mass resolution and a reasonable S/N.[183] However, every user must decide what information is most important, since the resolution can be – at least to a certain extent – compromised for increased sensitivity and vice versa.

7.1 Mass Resolution and Mass Accuracy

Reasonable mass resolution is often of particular importance, for example, to derive the elemental composition of an unknown compound. This is of particular relevance if the masses of peptides obtained by enzymatic hydrolysis of an unknown protein are assigned by database searching.[35] If mass accuracy is relatively low, many potential peptide structures are obtained. The higher the mass accuracy, the lower is the number of ''hits'' in the database! Unfortunately, highly resolved mass spectra cannot be routinely acquired and require the use of sophisticated – and expensive – mass spectrometers, for instance, equipped with FTICR (Fourier transform ion cyclotron resonance). A more detailed description of this technique, which was recently reviewed, is outside the scope of this article.[184]

Independent of the mass analyzer, each user of a mass spectrometer should be aware that the number of sampled data points is essential for high resolution. For example, if one wants to determine the mass of an unknown substance with an accuracy up to the third decimal point, at least 1000 data points/Da are required to achieve this. This means that for highly resolved mass spectra, the mass range must be minimized to obtain a sufficient number of data points. Commercially available MALDI/TOF devices provide mass resolutions of a few thousand (linear mode) up to about 20 000 in the reflectron mode under DE conditions.[35] This indicates that mixtures of molecules with monoisotopic masses of 1000 and 1000.05 can still be resolved. The mass accuracy of many commercial devices is of the order of 20 ppm, i.e. if a compound with a mass of 1000 Da is investigated, the mass can be provided with 1000 ± 0.02. One should also note that changes of the

instrument settings to increase sensitivity (e.g. increasing the reflectron voltage) require new calibration of the instrument.[35]

High mass resolution is not only important for the assignment of unknown peptides obtained by digestion of proteins, but is also typically required to determine whether a peak in a lipid spectrum, for example, corresponds to a K^+ adduct (M + 39) or the Na^+ adduct of the corresponding epoxide (M + 23 + 16). Since each possibility leads to slightly different masses (M + 38.9637 or M + 38.9847), exact mass determinations would allow the clear discrimination of both compounds.

However, the same information can also be obtained by a simple exchange experiment, using an excess of K^+ or Cs^+ [144] instead of Na^+.

7.2 Signal-to-noise Ratio

A reasonable S/N is highly desirable, for example, to identify low amounts of a certain compound of a mixture. Improving the S/N is a rather empirical process since it depends on a number of parameters that have to be carefully set. Such parameters include the analyte concentration, the matrix concentration, the analyte/matrix ratio, the applied laser intensity, the parameter setup, and so on. Here, for each substance and each sample, one has to find individually the best conditions. However, MALDI/TOF differs in a certain respect from most other analytical methods: an enhancement of the sample concentration does not necessarily increase the S/N. If the matrix/analyte ratio reaches a suboptimal level, the signal of the analyte is strongly suppressed![25]

7.3 Observed Fragmentations

In contrast with classical (EI) MS, where the inspection of fragmentation products mainly allows conclusions on chemical structure, in MALDI/TOF methodology the extent of fragmentation can be kept to a minimum and, usually, the (quasi) molecular ions provide the most important information. Therefore, some caution is needed not to obscure the molecular ion. In particular, the matrix should always be present in a large excess over the analyte to prevent a ''direct hit'' of the laser on the analyte, which would cause degradation of the sample.[24] It is also a widespread error that increased laser intensity leads to an improved S/N. In most cases, enhanced laser intensity is accompanied by a strong distortion of the baseline,[56] which may prevent the observation of small quantities of analyte. Basically, using a different matrix and changing the pH of the applied solutions slightly may be sufficient to decrease the extent of fragment ion generation.

7.4 Troubleshooting

If the acquired mass spectra are not in accordance with the expected sample composition, or, even worse, no signal at all can be detected, there are a few general changes that can be tried.

For most mass spectra, the average of a fixed number of different laser shots is calculated. It is possible to observe the individual spectra and try to lead the laser beam to a ''hot spot''. However, it may happen that all analyte on the sample plate has been spent (in some cases, the ablation from the sample plate can be seen). If this occurs (usually after about 30 ''shots'' on the same spot), the laser beam should be moved to a new point on the sample plate or more analyte should be applied to the sample plate.

A serious problem (especially in aqueous samples) is the disturbing effect of impurities. If the degree of contamination with salts, buffers, or detergents is not known, the following procedure is recommended. Prepare a comparably concentrated reference sample (e.g. $1\,mg\,mL^{-1}$ serum albumin in pure water). This sample should give an intense, reproducible signal. Dilute the albumin sample with distilled water and with the sample of interest, and record the mass spectra. If the spectra differ markedly in signal intensity and mass resolution, the sample is contaminated and should be purified before further analysis. Dialysis, microdialysis, or applying ZipTips™ are often already sufficient to overcome this problem. Apolar impurities like detergents can be easily removed by binding to commercially available apolar beads. It is also important to note that the ''limit of detection'' (LOD) is mainly limited by the level of chemical background noise and hence, cannot be improved significantly by averaging over a larger number of laser exposures.[152]

8 COMPARISON WITH OTHER MASS SPECTROMETRIC METHODS

Although, in the opinion of these authors, the MALDI/TOF technique is superior to other MS techniques in many respects, it must also be compared with other MS methods for advantages, drawbacks, and applicability (Table 1).

Ionization of the sample by EI[2] provides well-reproducible and simple-to-quantify mass spectra, but also a high extent of fragmentation. This is the reason why the molecular ion is only rarely detected (or is very weak).[3] Most ionization techniques also exhibit this problem and are, therefore, not suitable for the analysis of high-molecular-weight compounds exceeding masses of a few kilodaltons. However, EI is still the

ionization of choice for quantitative investigations, and gas chromatography/mass spectrometry (GC/MS) is still a very important method. The reason is that EI generates radical cations. The yield of these radical cations depends primarily on the available functional groups within the analyte. In contrast, using ionization methods that give ''quasimolecular'' ions under the addition of cations to the analyte, the ion yield is much more dependent on the acidity or basicity of the analyte. This is the reason why internal standards should always be used for quantitative investigations by MALDI or ESI MS.

Both MALDI and ESI represent ''soft-ionization'' techniques giving only very minor degradation of the sample,[16] and both these techniques are further compared. ESI generally provides more accurate mass information on low-mass compounds than MALDI and can be very easily combined with chromatographic techniques. By comparison, MALDI is more sensitive in many cases, can also be applied to very large biomolecules (larger than 100 kDa as TOF detectors basically enable an unlimited mass range), and is relatively insensitive toward salt concentrations in the millimolar range. Additionally, MALDI yields simpler spectra, since ESI produces a number (20 or more) of molecular ions, from which the mass of the analyte can be calculated. The ESI process seems softer than MALDI and ESI has also the considerable advantage that the analyte is investigated in solution, whereas for MALDI cocrystallization with the matrix is required. Labile molecules may not survive this procedure and this is why the analysis of noncovalent complexes is still a domain of ESI MS.[185]

Both methods, MALDI and ESI, often give complementary information. Therefore, it is more beneficial to ask ''Can I use both methods to obtain an optimum of information?'' rather than to ask ''Which method should I use?''.

9 SUMMARY AND METHOD DEVELOPMENT

The most important fundamentals of MALDI/TOF MS have been described, and its applicability to the investigation of different selected substance classes has been discussed. The comparison between MALDI and ESI has been emphasized because these relatively new methods are both ''soft-ionization'' techniques that yield information mainly on the intact molecular ion. This makes both methods very suitable for the analysis of mixtures. However, it must be considered that ion yields may differ from compound to compound, and less sensitively detectable compounds may be suppressed by more sensitively detectable ones.

One of the most important applications of MALDI/TOF is protein analysis, very often in combination with other biochemical methods prior to MALDI analysis. Proteins with a molecular weight close to 1000 kDa have already been successfully characterized.[101,186] Subsequent to an enzymatic digestion of the protein, MALDI is especially well suited for so-called peptide mapping. The high sensitivity and the adequate mass resolution of MALDI/TOF make it a strong competitor for classical electrophoresis. The most current trend in this field is automatization. Mass spectrometers will soon achieve complete, nearly fully automated, sequence analysis of proteins. However, a major drawback is the still relatively low reproducibility of MALDI/TOF mass spectra owing to the limited homogeneity of crystallization between the matrix and the analyte. Here, the search for further useful matrices is essential.

In comparison with protein analysis, there are by far fewer applications of MALDI/TOF for the analysis of carbohydrates[23] and lipids.[25] Whereas lipids can be very easily and accurately characterized, carbohydrates are still a major problem. Acidic carbohydrates, particularly, show a very weak tendency to desorb and are, therefore, very difficult to analyze. To solve these problems, some efforts are necessary, and the search for suitable matrices has to be continued. Further developments in MALDI/TOF MS will follow increasing knowledge of the matrix and its interaction with the analyte.[2,11]

At present, a major drawback in the field of MALDI MS is the limited, quantitative analysis of MALDI/TOF mass spectra. Although there were some recent reports on this,[56,74] one major problem is still the limited homogeneity of the cocrystals between matrix and analyte. Further experimental progress is expected in this area as this would be also important to establish MS imaging as a reliable diagnostic tool.[179]

ACKNOWLEDGMENTS

The authors wish to thank all colleagues and friends who helped them in writing this article. We especially wish to thank Dr Holger Spalteholz, Dr Jacqueline Leßig, Dr Matthias Müller, and Ms Rosmarie Süß. The kind and helpful advice of Dr Suckau and Dr Schürenberg (Bruker Daltonics, Bremen) is also gratefully acknowledged.

This work was supported by the German Research Council (DFG Schi 476/5-1, Schi 476/7-1 as well as the former HBFG program enabling the purchase of a Bruker ''Autoflex'' device) and the Federal Ministry of Education and Research (Grant BMBF 0313836).

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
APCI	Atmospheric Pressure Chemical Ionization
ATP	Adenosine Triphosphate
CI	Chemical Ionization
CID	Collision-induced Dissociation
CS	Chondroitin Sulfate
DAG	Diacylglycerol
DE	Delayed Extraction
DHB	2,5-Dihydroxybenzoic Acid
DIOS	Desorption/Ionization on Electrochemically Etched Silicon Surfaces
DIOS-MS	Desorption/Ionization on Electrochemically Etched Silicon Surfaces Mass Spectrometry
DMSO	Dimethyl Sulfoxide
DNA	Deoxy Ribonucleic Acid
EI	Electron Impact
ESI	Electrospray Ionization
ESI-CID	Electrospray Ionization/Collision-induced Dissociation
FAB	Fast Atom Bombardment
FD	Field Desorption
FI	Field Ionization
FTICR	Fourier Transform Ion Cyclotron Resonance
GC/MS	Gas Chromatography/Mass Spectrometry
3-HPA	3-Hydroxy-picolinic Acid
HA	Hyaluronic Acid (Hyaluronan and Hyaluronate as synonyms)
HEPES	*N*-(2-Hydroxyethyl)piperazine-*N*′-ethane sulfonic Acid
IR	Infrared
IT	Ion Trap
LAMMA	Laser Microprobe Mass Analysis
LD	Laser Desorption
LOD	Limit of Detection
LPC	Lysophosphatidylcholine
MALDI	Matrix-assisted Laser Desorption/Ionization
MPO	Myeloperoxidase
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass-to-charge Ratio
NADH	Nicotinamide Adenine Dinucleotide
PA	Phosphatidic Acid
PC	Phosphatidylcholine
PD	Plasma Desorption
PE	Phosphatidylethanolamine
PEG	Poly(ethylene glycol)
PI	Phosphatidylinositole
PL	Phospholipid
PMF	Peptide Mass Fingerprinting
PNA	*para*-Nitroaniline
ppm	parts per million
PS	Phosphatidylserine
PSD	Post Source Decay
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
SALDI	Surface-assisted Laser Desorption/Ionization
SDS	Sodium dodecyl sulfate
SDS/PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIMS	Secondary Ion Mass Spectrometry
S/N	Signal-to-noise
TFA	Trifluoroacetic Acid
TLC	Thin-layer Chromatography
TMEDA	*N*,*N*,*N*′,*N*′-tetramethylethylene-diamine
TOF	Time-of-flight
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

CHEMICAL ABSTRACT SERVICE

Adenosine triphosphate disodium salt	987-65-5
Albumin	9048-46-8
Chondroitin 4-sulfate	9082-079
Chondroitinase ABC	9024-13-9
α-Cyanohydroxycinnamic acid	28166-41-8
Dextran	9004-54-0
2,5-Dihydroxybenzoic acid	490-79-9
Dipalmitoylglycerol	30334-71-5
Dipalmitoylphosphatidic acid	7091-44-3
Palmitoyl-oleoylphosphatidylcholine	59491-62-2
Palmitoyl-oleoyphosphatidylethanolamine	26662-94-2
Palmitoyl-oleoyphosphatidylserine	321863-21-2
Heparin	9041-08-1
Hyaluronic acid	9067-32-7
Hyaluronidase	37326-33-3
Hydrogen Peroxide	7722-84-1
HEPES	7365-45-9
Hypochlorous Acid	7681-52-9
Myeloperoxidase	9003-99-0
p-Nitroaniline	100-01-6
Poly(ethylene glycol)	25322-68-3
Sinapinic acid	530-59-6
2,4,6-Trihydroxyacetophenone	480-66-0
Triolein	122-32-7

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Proteoglycan and Acidic Polysaccharide Analysis

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Literature of Mass Spectrometry • Time-of-flight Mass Spectrometry

REFERENCES

1. B.L. Peterson, B.S. Cummings, 'A Review of Chromatographic Methods for the Assessment of Phospholipids in Biological Samples', *Biomed. Chromatogr.*, **20**, 227–243 (2006).
2. J.R. Chapman, 'Ionization Methods', in *Encyclopedia of Analytical Science*, ed. A. Townsend, Academic Press, London, Vol. 5, 1995.
3. J.R. Chapman, *Practical Organic Mass Spectrometry – A Guide for Chemical and Biochemical Analysis*, John Wiley & Sons, Chichester, 1993.
4. H.D. Beckey, *Principles of Field Desorption Mass Spectrometry*, Pergamon Press, Oxford, 1977.
5. A.G. Harrison, *Chemical Ionization Mass Spectrometry*, CRC Press, Boca Raton, 1983.
6. A. Benninghoven, W.K. Sichtermann, 'Detection, Identification and Structural Investigation of Biologically Important Compounds by Secondary Ion Mass Spectrometry', *Anal. Chem.*, **50**, 1180–1804 (1978).
7. R.D. Macfarlane, D.F. Torgerson, 'Californium-252 Plasma Desorption Mass Spectroscopy', *Science*, **191**, 920–925 (1976).
8. R.J. Cotter, 'Laser Mass Spectrometry: An Overview of Techniques, Instruments and Applications', *Anal. Chim. Acta*, **195**, 45–59 (1987).
9. A.F. Casy, 'Mass Spectrometry as an Aid to the Identification of Ergots and Dihydroergots: Comparison of Hard and Soft Ionization Techniques', *J. Pharm. Biomed. Anal.*, **12**, 41–46 (1994).
10. S.A. Lorenz, E.P. Maziarz, T.D. Wood, 'Electrospray Ionization Fourier Transform Mass Spectrometry of Macromolecules: The First Decade', *Appl. Spectrosc.*, **53**, 18A–36A (1999).
11. R. Zenobi, R. Knochenmuss, 'Ion Formation in MALDI Mass Spectrometry', *Mass Spectrom. Rev.*, **17**, 337–366 (1999).
12. K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, 'Protein and Polymer Analysis up to m/z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **2**, 151–153 (1988).
13. C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, 'Electrospray Interface for Liquid Chromatographs and Mass Spectrometers', *Anal. Chem.*, **57**, 675–679 (1985).
14. A. Cho, D. Normile, 'Nobel Prize in Chemistry. Mastering Macromolecules', *Science*, **298**, 527–528 (2002).
15. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**, 2299–2301 (1988).
16. F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**, 1193A–1203A (1991).
17. W.T. Moore, 'Laser Desorption Mass Spectrometry', *Methods Enzymol.*, **289**, 520–542 (1997).
18. M. Mann, G. Talbo, 'Developments in Matrix-Assisted Laser Desorption/Ionization Peptide Mass Spectrometry', *Curr. Opin. Biotechnol.*, **7**, 11–19 (1996).
19. D.H. Patterson, G.E. Tarr, F.E. Regnier, S.A. Martin, 'C-terminal Ladder Sequencing Via Matrix-Assisted Laser Desorption Mass Spectrometry Coupled with Carboxypeptidase Y Time-Dependent and Concentration-Dependent Digestion', *Anal. Chem.*, **67**, 3971–3978 (1995).
20. S. Hu, J.A. Loo, D.T. Wong, 'Human Body Fluid Proteome Analysis', *Proteomics*, **6**, 6326–6353 (2006).
21. M. Soloviev, P. Finch, 'Peptidomics, Current Status', *J. Chromatogr. B*, **815**, 11–24 (2005).
22. D.J. Harvey, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Oligosaccharides and Glycoconjugates', *J. Chromatogr. A*, **720**, 429–446 (1996).
23. D.J. Harvey, 'Analysis of Carbohydrates and Glycoconjugates by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: An Update Covering the Period 1999–2000', *Mass Spectrom. Rev.*, **25**, 595–662 (2006).
24. J. Schiller, J. Arnhold, S. Benard, M. Müller, S. Reichl, K. Arnold, 'Lipid Analysis by Matrix-Assisted Laser Desorption and Ionization Mass Spectrometry: A Methodological Approach', *Anal. Biochem.*, **267**, 46–56 (1999).
25. J. Schiller, R. Süß, J. Arnhold, B. Fuchs, J. Leßig, M. Müller, M. Petković, H. Spalteholz, O. Zschörnig, K. Arnold, 'Matrix-Assisted Laser Desorption and Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry in Lipid and Phospholipid Research', *Prog. Lipid Res.*, **43**, 449–488 (2004).

26. L. Cohen, E.P. Go, G. Siuzdak, 'Small Molecule Desorption/Ionization Mass Analysis', in *MALDI MS – A Practical Guide to Instrumentation, Methods and Application*, eds. F. Hillenkamp, J. Peter-Katalinić, Wiley-VCH, Weinheim, 299–337, 2007.
27. M.C. Fitzgerald, G.R. Parr, L.M. Smith, 'Basic Matrices for the Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Proteins and Oligonucleotides', *Anal. Chem.*, **65**, 3204–3211 (1993).
28. D.S. Cornett, M.A. Duncan, I.J. Amster, 'Liquid-Mixtures for Matrix-Assisted Laser-Desorption', *Anal. Chem.*, **65**, 2608–2613 (1993).
29. M.J. Dale, R. Knochenmuss, R. Zenobi, 'Graphite/Liquid Mixed Matrices for Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **68**, 3321–3329 (1996).
30. J. Wei, J.M. Buriak, G. Siuzdak, 'Desorption-Ionization Mass Spectrometry on Porous Silicon', *Nature*, **399**, 243–246 (1999).
31. S.A. Trauger, E.P. Go, Z. Shen, J.V. Apon, B.J. Compton, E.S.P. Bouvier, M.G. Finn, G. Siuzdak, 'High Sensitivity and Analyte Capture with Desorption/Ionization Mass Spectrometry on Silylated Porous Silicon' *Anal. Chem.*, **76**, 4484–4489 (2004).
32. R. Knochenmuss, 'Ion Formation Mechanisms in UV-MALDI', *Analyst*, **131**, 966–986 (2006).
33. A.J. Hoteling, W.F. Nichols, D.J. Giesen, J.R. Lenhard, R. Knochenmuss, 'Electron Transfer Reactions in Laser Desorption/Ionization and Matrix-Assisted Laser Desorption/Ionization: Factors Influencing Matrix and Analyte Ion Intensities', *Eur. J. Mass Spectrom.*, **12**, 345–359 (2006).
34. R.M. Whittal, D.C. Schriemer, L. Li, 'Time-Lag Focusing MALDI Time-of-Flight Spectrometry for Polymer Characterization: Oligomer Resolution, Mass Accuracy, and Average Weight Information', *Anal. Chem.*, **69**, 2734–2741 (1997).
35. F. Hillenkamp, J. Peter-Katalinić, *MALDI MS – A Practical Guide to Instrumentation, Methods and Application*, Wiley-VCH, Weinheim, 2007.
36. C. Mayrhofer, S. Krieger, E. Raptakis, G. Allmaier, 'Comparison of Vacuum Matrix-Assisted Laser Desorption/Ionization (MALDI) and Atmospheric Pressure MALDI (AP-MALDI) Tandem Mass Spectrometry of 2-Dimensional Separated and Trypsin-Digested Glomerular Proteins for Database Search Derived Identification', *J. Proteome Res.*, **5**, 1967–1978 (2006).
37. E. Pittenauer, M. Zehl, O. Belgacem, E. Raptakis, R. Mistrik, G. Allmaier, 'Comparison of CID Spectra of Singly Charged Polypeptide Antibiotic Precursor Ions Obtained by Positive-Ion Vacuum MALDI IT/RTOF and TOF/RTOF, AP-MALDI-IT and ESI-IT Mass Spectrometry', *J. Mass Spectrom.*, **41**, 421–447 (2006).
38. L. Stryer, *Biochemie*, Vieweg, Braunschweig, 1988.
39. E. Fahy, S. Subramaniam, H.A. Brown, C.K. Glass, A.H. Merrill Jr, R.C. Murphy, C.R. Raetz, D.W. Russell, Y. Seyama, W. Shaw, T. Shimizu, F. Spener, G. van Meer, M.S. Vannieuwenhze, S.H. White, J.L. Witztum, E.A. Dennis, 'A Comprehensive Classification System for Lipids', *J. Lipid Res.*, **46**, 839–861 (2005).
40. M.N. Hodgkin, T.R. Pettitt, A. Martin, R.H. Mitchell, A.J. Pemberton, M.J. Wakelam, 'Diacylglycerols and Phosphatidates: Which Molecular Species are Intracellular Messengers?', *Trends Biochem. Sci.*, **23**, 200–204 (1998).
41. T. Kolter, 'Lipids and Lipidomics', *Angew. Chem. Int. Ed. Engl.*, **45**, 5910–5911 (2006).
42. J. Schiller, R. Süß, B. Fuchs, M. Müller, O. Zschörnig, K. Arnold, 'MALDI-TOF MS in Lipidomics', *Front. Biosci.*, **12**, 2568–2579 (2007).
43. G. Ebert, *Biopolymere*, Teubner, Stuttgart, 1993.
44. J. Schiller, B. Fuchs, K. Arnold, 'The Molecular Organization of Polymers of Cartilage in Health and Disease', *Curr. Org. Chem.*, **10**, 1771–1789 (2006).
45. P. Chaurand, F. Luetzenkirchen, B. Spengler, 'Peptide and Protein Identification by Matrix-Assisted Laser Desorption Ionization (MALDI) and MALDI-Post-Source Decay Time-of-Flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **10**, 91–103 (1999).
46. W. Pusch, M. Kostrzewa, 'Application of MALDI-TOF Mass Spectrometry in Screening and Diagnostic Research', *Curr. Pharm. Des.*, **11**, 2577–2591 (2005).
47. M. Sharon, C.V. Robinson, 'The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes', *Annu. Rev. Biochem.*, **76**, 167–193 (2007).
48. L.S. Lohmander, M.W. Lark, L. Dahlberg, L.A. Walakovits, H. Roos, 'Cartilage Matrix Metabolism in Osteoarthritis: Markers in Synovial Fluid, Serum and Urine', *Clin. Biochem.*, **25**, 167–174 (1992).
49. J. Schiller, B. Fuchs, J. Arnhold, K. Arnold, 'Contribution of Reactive Oxygen Species to Cartilage Degradation in Rheumatic Diseases: Molecular Pathways, Diagnosis and Potential Therapeutic Strategies', *Curr. Med. Chem.*, **10**, 2123–2145 (2003).
50. J.S. Page, S.S. Rubakhin, J.V. Sweedler, 'Single-Neuron Analysis Using CE Combined with MALDI MS and Radionuclide Detection', *Anal. Chem.*, **74**, 497–503 (2002).
51. F. Hucho, P. Franke, 'ESI- und MALDI-Massenspektrometrie-Routinemethoden in der Proteinbiochemie', *Biospektrum*, **6**, 43–45 (1995).
52. K.O. Bornsen, 'Influence of Salts, Buffers, Detergents, Solvents, and Matrices on MALDI-MS Protein Analysis

in Complex Mixtures', *Methods Mol. Biol.*, **146**, 387–404 (2000).

53. S.L. Cohen, B.T. Chait, 'Influence of Matrix Solution Conditions on the MALDI/MS Analysis of Peptides and Proteins', *Anal. Chem.*, **68**, 31–37 (1996).
54. A.I. Gusev, W.R. Wilkinson, A. Proctor, D.M. Hercules, 'Improvement of Signal Reproducibility and Matrix/Comatrix Effects in MALDI Analysis', *Anal. Chem.*, **67**, 1034–1041 (1995).
55. J.-F. Gal, P.-C. Maria, L. Massi, C. Mayeux, P. Burk, J. Tammiku-Taul, 'Cesium Cation Affinities and Basicities', *Int. J. Mass Spectrom.*, **267**, 7–23 (2007).
56. S. Benard, J. Arnhold, M. Lehnert, J. Schiller, K. Arnold, 'Experiments Towards Quantification of Saturated and Polyunsaturated Diacylglycerols by Matrix-Assisted Laser Desorption and Ionization Time-of-Flight Mass Spectrometry', *Chem. Phys. Lipids*, **100**, 115–125 (1999).
57. C. Pan, S. Xu, H. Zhou, Y. Fu, M. Ye, H. Zou, 'Recent Developments in Methods and Technology for Analysis of Biological Samples by MALDI-TOF-MS', *Anal. Bioanal. Chem.*, **387**, 193–204 (2007).
58. J. Schiller, R. Süß, B. Fuchs, M. Müller, M. Petković, O. Zschörnig, H. Waschipky, 'The Suitability of Different DHB Isomers as Matrices for the MALDI-TOF MS Analysis of Phospholipids: Which Isomer for What Purpose?', *Eur. Biophys. J.*, **36**, 517–527 (2007).
59. P. Onnerfjord, S. Ekstrom, J. Bergquist, J. Nilsson, T. Laurell, G. Marko-Varga, 'Homogeneous Sample Preparation for Automated High Throughput Analysis with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **13**, 315–322 (1999).
60. R.C. Beavis, B.T. Chait, 'Cinnamic Acid Derivatives as Matrices for Ultraviolet Laser Desorption Mass Spectrometry of Proteins', *Rapid Commun. Mass Spectrom.*, **3**, 432–435 (1989).
61. U. Pieles, W. Zurcher, M. Schar, H.E. Moser, 'Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry: A Powerful Tool for the Mass and Sequence Analysis of Natural and Modified Oligonucleotides', *Nucleic Acids Res.*, **21**, 3191–3196 (1993).
62. M.D. Mohr, K.O. Bornsen, H.M. Widmer, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: Improved Matrix for Oligosaccharides', *Rapid Commun. Mass Spectrom.*, **9**, 809–814 (1995).
63. D.S. Peterson, 'Matrix-Free Methods for Laser Desorption/Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **26**, 19–34 (2007).
64. C. Black, C. Poile, J. Langley, J. Herniman, 'The Use of Pencil Lead as a Matrix and Calibrant for Matrix-Assisted Laser Desorption/Ionisation', *Rapid Commun. Mass Spectrom.*, **20**, 1053–1060 (2006).
65. O. Vorm, P. Roepstorff, M. Mann, 'Improved Resolution and Very High Sensitivity in MALDI-TOF of Matrix Surfaces Made by Fast Evaporation', *Anal. Chem.*, **66**, 3281–3287 (1994).
66. T.R. Klein, D. Kirsch, R. Kaufmann, D. Riesner, 'Prion Rods Contain Small Amounts of Two Host Sphingolipids as Revealed by Thin-Layer Chromatography and Mass Spectrometry', *Biol. Chem.*, **379**, 655–666 (1998).
67. J. Yao, J.R. Scott, M.K. Young, C.L. Wilkins, 'Importance of Matrix: Analyte Ratio for Buffer Tolerance Using 2,5-Dihydroxybenzoic Acid as a Matrix in Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Mass Spectrometry and Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight', *J. Am. Soc. Mass Spectrom.*, **9**, 805–813 (1998).
68. H. Gorisch, 'Drop Dialysis: Time Course of Salt and Protein Exchange', *Anal. Biochem.*, **173**, 393–398 (1988).
69. M. Schuerenberg, C. Luebbert, H. Eickhoff, M. Kalkum, H. Lehrach, E. Nordhoff, 'Prestructured MALDI-MS Sample Supports', *Anal. Chem.*, **72**, 3436–3442 (2000).
70. R.C. Beavis, B.T. Chait, 'Factors Affecting the Ultraviolet Laser Desorption of Proteins', *Rapid Commun. Mass Spectrom.*, **3**, 233–237 (1989).
71. W.E. Ens, Y. Mao, R. Meyer, 'Properties of Matrix-Assisted Laser Desorption. Measurements with a Time-to-Digital Converter', *Rapid Commun. Mass Spectrom.*, **5**, 117–123 (1991).
72. P. Juhasz, M.L. Vestal, S.A. Martin, 'On the Initial Velocity of Ions Generated by Matrix-Assisted Laser Desorption Ionization and its Effect on the Calibration of Delayed Extraction Time-of-Flight Mass Spectra', *J. Am. Soc. Mass Spectrom.*, **8**, 209–217 (1997).
73. L. Sleno, D.A. Volmer, A.G. Marshall, 'Assigning Product Ions from Complex MS/MS Spectra: The Importance of Mass Uncertainty and Resolving Power', *J. Am. Soc. Mass Spectrom.*, **16**, 183–198 (2005).
74. M.W. Duncan, G. Matanovic, A. Cerpa-Poljak, 'Quantitative Analysis of Low Molecular Weight Compounds of Biological Interest by Matrix-Assisted Laser Desorption Ionization', *Rapid Commun. Mass Spectrom.*, **7**, 1090–1094 (1993).
75. K.J. Wu, R.W. Odom, 'Characterizing Synthetic Polymers by MALDI MS', *Anal. Chem.*, **70**, 456A–461A (1998).
76. J.M. Harris, 'Introduction to Biotechnical and Biomedical Applications of Poly(Ethylene Glycol)', in *Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications*, ed. J.M. Harris, Plenum Press, New York, 1992.
77. K. Raith, C.E. Schmelzer, R.H. Neubert, 'Towards a Molecular Characterization of Pharmaceutical Excipients: Mass Spectrometric Studies of Ethoxylated Surfactants', *Int. J. Pharm.*, **319**, 1–12 (2006).

78. J. Schiller, J. Arnhold, J. Schwinn, H. Sprinz, O. Brede, K. Arnold, 'Differences in the Reactivity of Phthalic Hydrazide and Luminol with Hydroxyl Radicals', *Free Radical Res.*, **30**, 45–57 (1999).

79. W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, 'Identifying Proteins from Two-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases', *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5011–5015 (1993).

80. R. Aebersold, M. Mann, 'Mass Spectrometry-Based Proteomics', *Nature*, **422**, 198–207 (2003).

81. M. Kussmann, P. Roepstorff, 'Sample Preparation Techniques for Peptides and Proteins Analyzed by MALDI-MS', *Methods Mol. Biol.*, **146**, 405–424 (2000).

82. J. Reinders, U. Lewandrowski, J. Moebius, Y. Wagner, A. Sickmann, 'Challenges in Mass Spectrometry-Based Proteomics', *Proteomics*, **4**, 3686–3703 (2004).

83. O.N. Jensen, 'Interpreting the Protein Language Using Proteomics', *Nat. Rev. Mol. Cell Biol.*, **7**, 391–403 (2006).

84. J.X. Yan, J.C. Sanchez, P.A. Binz, K.L. Williams, D.F. Hochstrasser, 'Method for Identification and Quantitative Analysis of Protein Lysine Methylation Using Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry and Amino Acid Analysis', *Electrophoresis*, **20**, 749–754 (1999).

85. J.L. Wolfender, F. Chu, H. Ball, F. Wolfender, M. Fainzilber, M.A. Baldwin, A.L. Burlingame, 'Identification of Tyrosine Sulfation in Connus Pennaceus Conotoxins a-PnIA and a-PnIB: Further Investigation of Labile Sulfo- and Phosphopeptides by Electrospray, Matrix-Assisted Laser Desorption/Ionization (MALDI) and Atmospheric Pressure MALDI Mass Spectrometry', *J. Mass Spectrom.*, **34**, 447–454 (1999).

86. L.F. Bonewald, L. Bibbs, S.A. Kates, A. Khatri, J.S. McMurray, K.F. Medzihradszky, S.T. Weintraub, 'Study on the Synthesis and Characterization of Peptides Containing Phosphorylated Tyrosine', *J. Pept. Res.*, **53**, 161–169 (1999).

87. J.M. Asara, J. Allison, 'Enhanced Detection of Phosphopeptides in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using Ammonium Salts', *J. Am. Soc. Mass Spectrom.*, **10**, 35–44 (1999).

88. S.A. Gerber, C.R. Scott, F. Turecek, M.H. Gelb, 'Analysis of Rates of Multiple Enyzmes in Cell Lysates by Electrospray Ionization Mass Spectrometry', *J. Am. Chem. Soc.*, **121**, 1102–1103 (1999).

89. Y. Yang, R. Orlando, 'Simplifying the Exoglycosidase Digestion/MALDI/MS Procedures for Sequencing N-Linked Carbohydrate Side Chains', *Anal. Chem.*, **68**, 570–572 (1996).

90. A.S. Woods, J.C. Buchsbaum, T.A. Worrall, J.M. Berg, R.J. Cotter, 'Matrix-Assisted Laser Desorption/Ionization of Noncovalently Bound Compounds', *Anal. Chem.*, **67**, 4462–4465 (1995).

91. S. Jespersen, W.M. Niessen, U.R. Tjaden, J. van der Greef, 'Basic Matrices in the Analysis of Non-Covalent Complexes by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry', *J. Mass Spectrom.*, **33**, 1088–1093 (1998).

92. F. Song, 'A Study of Noncovalent Protein Complexes by Matrix-Assisted Laser Desorption/Ionization', *J. Am. Soc. Mass Spectrom.*, **18**, 1286–1290 (2007).

93. G. Bolbach, 'Matrix-Assisted Laser Desorption/Ionization Analysis of Non-Covalent Complexes: Fundamentals and Applications', *Curr. Pharm. Des.*, **11**, 2535–2557 (2005).

94. J.R. Engen, D.L. Smith, 'Investigating Protein Structure and Dynamics by Hydrogen Exchange MS', *Anal. Chem.*, **73**, 256A–265A (2001).

95. L. Wang, L.C. Lane, D.L. Smith, 'Detecting Structural Changes in Viral Capsids by Hydrogen Exchange and Mass Spectrometry', *Protein Sci.*, **10**, 1234–1243 (2001).

96. F. Gharahdaghi, C.R. Weinberg, D.A. Meagher, B.S. Imai, S.M. Mische, 'Mass Spectrometric Identification of Proteins from Silver-Stained Polyacrylamide Gel: A Method for the Removal of Silver Ions to Enhance Sensitivity', *Electrophoresis*, **20**, 601–605 (1999).

97. K. Gevaert, H. de Mol, J.L. Verschelde, J. van Damme, S. de Boeck, J. Vandekerckhove, 'Novel Techniques for Identification and Characterization of Proteins Loaded on Gels in Femtomole Amounts', *J. Protein Chem.*, **16**, 335–342 (1997).

98. T. Solouki, J.A. Marto, F.M. White, S. Guan, A.G. Marshall, 'Attomole Biomolecule Mass Analysis by Matrix-Assisted Laser Desorption/Ionization Fourier Transform Ion Cyclotron Resonance', *Anal. Chem.*, **67**, 4139–4144 (1997).

99. J. Schiller, J. Arnhold, K. Arnold, 'Nuclear Magnetic Resonance and Mass Spectrometric Studies on the Action of Proteases on Pig Articular Cartilage', *Z. Naturforsch.*, **53C**, 1072–1080 (1998).

100. G. Zhang, A. Sun, W. Li, T. Liu, Z. Su, 'Mass Spectrometric Analysis of Enzymatic Digestion of Denatured Collagen for Identification of Collagen Type', *J. Chromatogr. A*, **1114**, 274–277 (2006).

101. B.E. Chong, D.B. Wall, D.M. Lubmann, S.J. Flynn, 'Rapid Profiling of E. Coli Proteins up to 500 kDa from Whole Cell Lysates Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 1900–1908 (1997).

102. X. Liang, J. Bai, Y.-H. Liu, D.M. Lubman, 'Characterization of SDS/PAGE-Separated Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **68**, 1012–1018 (1996).

103. L.M. Preston, K.K. Murray, D.H. Russell, 'Reproducibility and Quantitation of Matrix-Assisted Laser Desorption Ionization Mass Spectrometry: Effects of Nitrocellulose on Peptide Ion Yields', *Biol. Mass Spectrom.*, **22**, 544–550 (1993).

104. R.R. Hensel, R.C. King, K.G. Owens, 'Electrospray Sample Preparation for Improved Quantitation in Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 1785–1793 (1997).

105. D.J. Harvey, P.M. Rudd, R.H. Bateman, R.S. Bordoli, K. Howes, J.B. Hoyes, R.G. Vickers, 'Examination of Complex Oligosaccharides by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry on Time-of-Flight and Magnetic Sector Instruments', *Org. Mass Spectrom.*, **29**, 753–765 (1994).

106. B. Stahl, M. Steup, M. Karas, F. Hillenkamp, 'Analysis of Neutral Oligosaccharides by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **63**, 1463–1466 (1991).

107. D.I. Papac, A. Wong, A.J.S. Jones, 'Analysis of Acidic Oligosaccharides and Glycopeptides by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry', *Anal. Chem.*, **68**, 3215–3223 (1996).

108. T. Wyttenbach, M.T. Bowers, 'Gas Phase Conformations of Biological Molecules: The Hydrogen/Deuterium Exchange Mechanism', *J. Am. Soc. Mass Spectrom.*, **10**, 9–14 (1999).

109. C. Hao, X. Ma, S. Fang, Z. Liu, S. Liu, F. Song, J. Liu, 'Positive- and Negative-Ion Matrix-Assisted Laser Desorption Mass Spectrometry of Saccharides', *Rapid Commun. Mass Spectrom.*, **12**, 345–348 (1998).

110. P. Chen, A.G. Baker, M.V. Novotny, 'The Use of Osazones as Matrices for the Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Carbohydrates', *Anal. Biochem.*, **244**, 144–151 (1997).

111. Y. Mechref, A.G. Baker, M.V. Novotny, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Neutral And Acidic Oligosaccharides with Collision-Induced Dissociation', *Carbohydr. Res.*, **313**, 145–155 (1998).

112. B. Stahl, S. Thurl, J. Zeng, M. Karas, F. Hillenkamp, M. Steup, G. Sawatzki, 'Oligosaccharides from Human Milk as Revealed by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Biochem.*, **223**, 218–226 (1994).

113. A.J. Rhomberg, S. Ernst, R. Sasisekharan, K. Biemann, 'Mass Spectrometric and Capillary Electrophoretic Degradation of Heparin-Like Glycosaminoglycans', *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 4176–4181 (1998).

114. T. Ii, M. Kubota, S. Okuda, T. Hirano, M. Orashi, 'Negative-Ion Fast Atom Bombardment Tandem Mass Spectrometry for Characterization of Sulfated Unsaturated Disaccharides from Heparin and Heparan Sulfate', *Glycoconjugate J.*, **12**, 162–172 (1995).

115. K. Takagaki, K. Kojima, M. Majima, T. Nakamura, I. Kato, M. Endo, 'Ion Spray Mass Spectrometric Analysis of Glycosaminoglycan Oligosaccharides', *Glycoconjugate J.*, **9**, 174–179 (1992).

116. P. Juhasz, K. Biemann, 'Utility of Non-Covalent Complexes in the Matrix-Assisted Laser Desorption Ionization Mass Spectrometry of Heparin-Derived Oligosaccharides', *Carbohydr. Res.*, **270**, 131–147 (1995).

117. Y. Mechref, M.V. Novotny, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Acidic Glycoconjugates Facilitated by the Use of Spermine as a Co-Matrix', *J. Am. Soc. Mass Spectrom.*, **9**, 1293–1302 (1998).

118. K. Takagaki, T. Nakamura, J. Izumi,H. Saitoh, M. Endo, K. Kojima, I. Kato, M. Majima, 'Characterization of Hydrolysis and Transglycosylation by Testicular Hyaluronidase Using Ion-Spray Mass Spectrometry', *Biochemistry*, **33**, 6503–6507 (1994).

119. J. Schiller, J. Arnhold, S. Benard, S. Reichl, K. Arnold, 'Cartilage Degradation by Hyaluronate Lyase and Chondroitin ABC Lyase: A MALDI/TOF Mass Spectrometric Study', *Carbohydr. Res.*, **318**, 116–122 (1999).

120. H. Toyoda, K. Motoki, M. Tanikawa, K. Shinomiya, H. Akiyama, T. Imanari, 'Determination of Human Urinary Hyaluronic Acid, Chondroitin Sulphate and Dermatan Sulphate as their Unsaturated Disaccharides by High-Performance Liquid Chromatography', *J. Chromatogr.*, **565**, 141–148 (1991).

121. K. Busse, M. Averbeck, U. Anderegg, K. Arnold, J.C. Simon, J. Schiller, 'The Signal-to-Noise Ratio as a Measure of HA Oligomer Concentration: A MALDI-TOF MS Study', *Carbohydr. Res.*, **341**, 1065–1070 (2006).

122. M. Petković, J. Schiller, M. Müller, S. Benard, S. Reichl, K. Arnold, J. Arnhold, 'Detection of Individual Phospholipids in Lipid Mixtures by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry: Phosphatidylcholine Prevents the Detection of Further Species', *Anal. Biochem.*, **289**, 202–216 (2001).

123. N. Hilbert, J. Schiller, J. Arnhold, K. Arnold, 'Cartilage Degradation by Stimulated Human Neutrophils: Elastase is Mainly Responsible for Cartilage Damage', *Bioorg. Chem.*, **30**, 119–132 (2002).

124. L. Soltes, R. Mendichi, G. Kogan, J. Schiller, M. Stankovska, J. Arnhold, 'Degradative Action

of Reactive Oxygen Species on Hyaluronan', *Biomacromolecules*, **7**, 659–668 (2006).

125. O.M. Saad, R.A. Myers, D.L. Castleton, J.A. Leary, 'Analysis of Hyaluronan Content in Chondroitin Sulfate Preparations by Using Selective Enzymatic Digestion and Electrospray Ionization Mass Spectrometry', *Anal. Biochem.*, **344**, 232–239 (2005).

126. R.M. Schulz, A. Bader, 'Cartilage Tissue Engineering and Bioreactor Systems for the Cultivation and Stimulation of Chondrocytes', *Eur. Biophys. J.*, **36**, 539–568 (2007).

127. R. Schulz, S. Höhle, G. Zernia, M. Zscharnack, J. Schiller, A. Bader, K. Arnold, D. Huster, 'Analysis of Extracellular Matrix Production in Artificial Cartilage Constructs by Histology, Immunocytochemistry, Mass Spectrometry, and NMR Spectroscopy', *J. Nanosci. Nanotechnol.*, **6**, 2368–2381 (2006).

128. J. Zaia, X.Q. Li, S.Y. Chan, C.E. Costello, 'Tandem Mass Spectrometric Strategies for Determination of Sulfation Positions and Uronic Acid Epimerization in Chondroitin Sulfate Oligosaccharides', *J. Am. Soc. Mass Spectrom.*, **14**, 1270–1281 (2003).

129. T. Yamagaki, Y. Ishizuka, S. Kawabata, H. Nakanishi, 'Analysis of Glycosidic Linkages in Saccharide Compounds by Post-Source Decay Fragment Methods in Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectroscopy', *Rapid Commun. Mass Spectrom.*, **11**, 527–531 (1997).

130. A. Pfenninger, M. Karas, B. Finke, B. Stahl, G. Sawatzki, 'Matrix Optimization for Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Oligosaccharides from Human Milk', *J. Mass Spectrom.*, **34**, 98–104 (1999).

131. M.T. Cancilla, S.G. Penn, C.B. Lebrilla, 'Alkaline Degradation of Oligosaccharides Coupled with Matrix-Assisted Laser Desorption/Ionization Fourier Transform Mass Spectrometry: A Method for Sequencing Oligosaccharides', *Anal. Chem.*, **70**, 663–672 (1998).

132. Z. Lam, M.B. Comisarow, G.G.S. Dutton, 'Structural and Reaction Assignments for Some Common Polysaccharides Using Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Anal. Chem.*, **60**, 2304–2306 (1988).

133. R.W. Gross, X. Han, 'Unlocking the Complexity of Lipids: Using Lipidomics to Identify Disease Mechanisms, Biomarkers and Treatment Efficacy', *Future Lipidol.*, **1**, 539–547 (2006).

134. M. Pulfer, R.C. Murphy, 'Electrospray Mass Spectrometry of Phospholipids', *Mass Spectrom. Rev.*, **22**, 332–364 (2003).

135. W.C. Byrdwell, 'Atmospheric Pressure Chemical Ionization Mass Spectrometry for Analysis of Lipids', *Lipids*, **36**, 327–346 (2001).

136. D.J. Harvey, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Phospholipids', *J. Mass Spectrom.*, **30**, 1333–1346 (1995).

137. J.A. Marto, F.M. White, S. Seldomridge, A.G. Marshall, 'Structural Characterization of Phospholipids by Matrix-Assisted Laser Desorption/Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **67**, 3939–3984 (1995).

138. K.A. Al-Saad, W.F. Siems, H.H. Hill, V. Zabrouskov, N.R. Knowles, 'Structural Analysis of Phosphatidylcholines by Post-Source Decay Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **14**, 373–382 (2003).

139. M. Petković, J. Müller, M. Müller, J. Schiller, K. Arnold, J. Arnhold, 'Application of Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry for Monitoring the Digestion of Phosphatidylcholine by Pancreatic Phospholipase A_2', *Anal. Biochem.*, **308**, 61–70 (2002).

140. K.A. Al-Saad, V. Zabrouskov, W.F. Siems, N.R. Knowles, R.M. Hannan, H.H. Hill, 'Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry of Lipids: Ionization and Prompt Fragmentation Patterns', *Rapid Commun. Mass Spectrom.*, **17**, 87–96 (2003).

141. B. Fuchs, C. Schober, G. Richter, R. Süß, J. Schiller, 'MALDI-TOF MS of Phosphatidylethanolamines: Different Adducts Cause Different Post Source Decay (PSD) Fragment Ion Spectra', *J. Biochem. Biophys. Methods*, **70**, 689–692 (2007).

142. G.R. Asbury, K. Al-Saad, W.F. Siems, R.M. Hannan, H.H. Hill, 'Analysis of Triacylglycerols and Whole Oils by Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **10**, 983–991 (1999).

143. J. Gidden, R. Liyanage, B. Durham, J.O. Lay, 'Reducing Fragmentation Observed in the Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometric Analysis of Triacylglycerols in Vegetable Oils', *Rapid Commun. Mass Spectrom.*, **21**, 1951–1957 (2007).

144. J. Schiller, R. Süß, M. Petković, N. Hilbert, M. Müller, O. Zschörnig, J. Arnhold, K. Arnold, 'CsCl as an Auxiliary Reagent for the Analysis of Phosphatidylcholine Mixtures by Matrix-Assisted Laser Desorption and Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS)', *Chem. Phys. Lipids*, **113**, 123–131 (2001).

145. J. Schiller, R. Süß, M. Petković, O. Zschörnig, K. Arnold, 'Negative-Ion Matrix-Assisted Laser Desorption and Ionization Time-of-Flight Mass Spectra of Complex Phospholipid Mixtures in the Presence of Phosphatidylcholine: A Cautionary Note on Peak Assignment', *Anal. Biochem.*, **309**, 311–314 (2002).

146. G. Stubiger, O. Belgacem, 'Analysis of Lipids Using 2,4,6-Trihydroxyacetophenone as a Matrix for MALDI Mass Spectrometry', *Anal. Chem.*, **79**, 3206–3213 (2007).

147. R. Estrada, Y.C. Yappert, 'Alternative Approaches for the Detection of Various Phospholipid Classes by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry', *J. Mass Spectrom.*, **39**, 412–422 (2004).

148. M. Müller, J. Schiller, M. Petković, W. Oehrl, R. Heinze, R. Wetzker, K. Arnold, J. Arnhold, 'Limits for the Detection of (Poly-)Phosphoinositides by Matrix-Assisted Laser Desorption and Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS)', *Chem. Phys. Lipids*, **110**, 151–164 (2001).

149. G.P. Gellermann, T.R. Appel, P. Davies, S. Diekmann, 'Paired Helical Filaments Contain Small Amounts of Cholesterol, Phosphatidylcholine and Sphingolipids', *Biol. Chem.*, **387**, 1267–1274 (2006).

150. J.C. Touchstone, 'Thin-Layer Chromatographic Procedures for Lipid Separation', *J. Chromatogr. B*, **671**, 169–195 (1995).

151. J. Guittard, X.L. Hronowski, C.E. Costello, 'Direct Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Glycosphingolipids on Thin Layer Chromatographic Plates and Transfer Membranes', *Rapid Commun. Mass Spectrom.*, **13**, 1838–1849 (1999).

152. A. Rohlfing, J. Muthing, G. Pohlentz, U. Distler, J. Peter-Katalinić, S. Berkenkamp, K. Dreisewerd, 'IR-MALDI-MS Analysis of HPTLC-Separated Phospholipid Mixtures Directly from the TLC Plate', *Anal. Chem.*, **79**, 5793–5808 (2007).

153. B. Fuchs, J. Schiller, R. Süß, M. Schürenberg, D. Suckau, 'A Direct and Simple Method of Coupling Matrix-Assisted Laser Desorption and Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) to Thin-Layer Chromatography (TLC) for the Analysis of Phospholipids from Egg Yolk', *Anal. Bioanal. Chem.*, **389**, 827–834 (2007).

154. J. Schiller, R. Süß, B. Fuchs, M. Müller, O. Zschörnig, K. Arnold, 'Combined Application of TLC and Matrix-Assisted Laser Desorption and Ionisation Time-of-Flight Mass Spectrometry (MALDI-TOF MS) to Phospholipid Analysis of Brain', *Chromatographia*, **57**(Suppl. 1), 297–302 (2003).

155. M. Petković, J. Schiller, J. Müller, M. Müller, K. Arnold, J. Arnhold, 'The Signal-to-Noise Ratio as the Measure for the Quantification of Lysophospholipids by Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry', *Analyst*, **126**, 1042–1050 (2001).

156. J. Schiller, M. Müller, B. Fuchs, K. Arnold, D. Huster, 'P-31 NMR Spectroscopy of Phospholipids: From Micelles to Membranes', *Curr. Anal. Chem.*, **3**, 283–301 (2007).

157. J. Folch, M. Lees, G.H.S. Stanley, 'A Simple Method for the Isolation and Purification of Total Lipids from Animal Tissues', *J. Biol. Chem.*, **226**, 497–509 (1957).

158. J. Schiller, J. Arnhold, H.-J. Glander, K. Arnold, 'Lipid Analysis of Human Spermatozoa and Seminal Plasma by MALDI-TOF Mass Spectrometry and NMR Spectroscopy – Effects of Freezing and Thawing', *Chem. Phys. Lipids*, **106**, 145–156 (2000).

159. J. Schiller, K. Müller, R. Süß, J. Arnhold, C. Gey, A. Herrmann, J. Leßig, K. Arnold, P. Müller, 'Analysis of the Lipid Composition of Bull Spermatozoa by MALDI-TOF Mass Spectrometry – A Cautionary Note', *Chem. Phys. Lipids*, **126**, 85–94 (2003).

160. B. Fuchs, J. Schiller, U. Wagner, H. Häntzschel, K. Arnold, 'The Phosphatidylcholine/Lysophosphatidylcholine Ratio in Human Plasma is an Indicator of the Severity of Rheumatoid Arthritis: Investigations by ^{31}P NMR and MALDI-TOF MS', *Clin. Biochem.*, **38**, 925–933 (2005).

161. R.C. Murphy, *Mass Spectrometry of Phospholipids: Tables of Molecular and Product Ions*, Illuminati Press, Denver, 2002.

162. J.J. Jones, S. Borgmann, C.L. Wilkins, R.M. O'Brien, 'Characterizing the Phospholipid Profiles in Mammalian Tissues by MALDI FTMS', *Anal. Chem.*, **78**, 3062–3071 (2006).

163. J.J. Jones, M.J. Stump, R.C. Fleming, J.O. Lay, C.L. Wilkins, 'Investigation of MALDI-TOF and FT-MS Techniques for Analysis of Escherichia Coli Whole Cells', *Anal. Chem.*, **75**, 1340–1347 (2003).

164. Y. Ishida, O. Nakanishi, S. Hirao, S. Tsuge, J. Urabe, T. Sekino, M. Nakanishi, T. Kimoto, H. Ohtani, 'Direct Analysis of Lipids in Single Zooplankter Individuals by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **75**, 4514–4518 (2003).

165. C.M. Spickett, A.R. Pitt, A.J. Brown, 'Direct Observation of Lipid Hydroperoxides in Phospholipid Vesicles by Electrospray Mass Spectrometry', *Free Radical Biol. Med.*, **25**, 613–620 (1998).

166. C.M. Spickett, Chlorinated Lipids and Fatty Acids: An Emerging Role in Pathology', *Pharmacol. Ther.*, **115**, 400–409 (2007).

167. J. Arnhold, O.M. Panasenko, J. Schiller, Y.U.A. Vladimirov, K. Arnold, 'The Action of Hypochlorous Acid on Phosphatidylcholine Liposomes in Dependence on the Content of Double Bonds. Stoichiometry and NMR Analysis', *Chem. Phys. Lipids*, **78**, 55–64 (1995).

168. O. Zschörnig, C. Bergmeier, R. Süß, K. Arnold, J. Schiller, 'Human Low-Density Lipoprotein Oxidation: Hypochlorous Acid Leads to the Generation of Lysophosphatidylcholines under Acidic Conditions', *Lett. Org. Chem.*, **1**, 381–390 (2004).

169. J. Arnhold, A.N. Osipov, H. Spalteholz, O.M. Panasenko, J. Schiller, 'Effects of Hypochlorous Acid on Unsaturated Phosphatidylcholines', *Free Radical Biol. Med.*, **31**, 1111–1119 (2001).

170. J. Arnhold, A.N. Osipov, H. Spalteholz, O.M. Panasenko, J. Schiller, 'Formation of Lysophospholipids from Unsaturated Phosphatidylcholines under the Influence of Hypochlorous Acid', *Biochim. Biophys. Acta*, **1572**, 91–100 (2002).

171. J. Leßig, J. Schiller, J. Arnhold, B. Fuchs, 'Hypochlorous Acid-Mediated Generation of Glycerophosphocholine from Unsaturated Plasmalogen Glycerophosphocholine Lipids', *J. Lipid Res.*, **48**, 1316–1324 (2007).

172. F. Kirpekar, E. Nordhoff, K. Kristiansen, P. Roepstorff, S. Hahner, M. Karas, F. Hillenkamp, 'Matrix Assisted Laser Desorption/Ionization Mass Spectrometry of Enzymatically Synthesized RNA up to 150 kDa', *Nucleic Acids Res.*, **22**, 3866–3870 (1994).

173. P.A. Limbach, P.F. Crain, J.A. Closkey, 'Characterization of Oligonucleotides and Nucleic Acids by Mass Spectrometry', *Curr. Opin. Biotechnol.*, **6**, 96–102 (1995).

174. J. Ragoussis, G.P. Elvidge, K. Kaur, S. Colella, 'Matrix-Assisted Laser Desorption/Ionisation, Time-of-Flight Mass Spectrometry in Genomics Research', *PLoS Genet.*, **2**, e100 (2006).

175. S. Berkenkamp, F. Kirpekar, F. Hillenkamp, 'Infrared MALDI Mass Spectrometry of Large Nucleic Acids', *Science*, **281**, 260–262 (1998).

176. K.J. Wu, A. Steding, C.H. Becker, 'Matrix-Assisted Laser Desorption Time-of-Flight Mass Spectrometry of Oligonucleotides Using 3-Hydroxypicolinic Acid as an Ultraviolet-Sensitive Matrix', *Rapid Commun. Mass Spectrom.*, **7**, 142–146 (1993).

177. S. Sauer, 'The Essence of DNA Sample Preparation for MALDI Mass Spectrometry', *J. Biochem. Biophys. Methods*, **70**, 311–318 (2007).

178. R. Kaufmann, T. Wingerath, D. Kirsch, W. Stahl, H. Sies, 'Analysis of Carotenoids and Carotenol Fatty Acid Esters by Matrix-Assisted Laser Desorption Ionization (MALDI) and MALDI-Post-Scource-Decay Mass Spectrometry', *Anal. Biochem.*, **238**, 117–128 (1996).

179. L.A. McDonnell, R.M. Heeren, 'Imaging Mass Spectrometry', *Mass Spectrom. Rev.*, **26**, 606–643 (2007).

180. F. Hillenkamp, E. Unsöld, R. Kaufmann, R. Nitsche, 'Laser Microprobe Mass Analysis of Organic Materials', *Nature*, **256**, 119–120 (1975).

181. M. Stoeckli, P. Chaurand, D.E. Hallahan, R.M. Caprioli, 'Imaging Mass Spectrometry: A New Technology for the Analysis of Protein Expression in Mammalian Tissues', *Nat. Med.*, **7**, 493–496 (2001).

182. K. Schwamborn, R.C. Krieg, M. Reska, G. Jakse, R. Knuechel, A. Wellmann, 'Identifying Prostate Carcinoma by MALDI-Imaging', *Int. J. Mol. Med.*, **20**, 155–159 (2007).

183. M. Vetsal, P. Juhasz, 'Resolution and Mass Accuracy in Matrix-Assisted Laser Desorption Ionization-Time-of-Flight', *J. Am. Soc. Mass Spectrom.*, **9**, 892–911 (1998).

184. B. Bogdanov, R.D. Smith, 'Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up', *Mass Spectrom. Rev.*, **24**, 168–200 (2005).

185. T.D. Veenstra, 'Electrospray Ionization Mass Spectrometry in the Study of Biomolecular Non-Covalent Interactions', *Biophys. Chem.*, **79**, 63–79 (1999).

186. M. Karas, U. Bahr, A. Ingendoh, F. Hillenkamp, 'Laserdesorptions-Massenspektrometrie von Proteinen mit Molmassen zwischen 100 000 und 250 000 Dalton', *Angew. Chem.*, **101**, 805–806 (1989).

Nuclear Magnetic Resonance of Biomolecules

Jeremy N.S. Evans
Washington State University, Pullman, USA

1 Introduction **1**
1.1 How Nuclear Magnetic Resonance is Used to Study Biomolecules 1
1.2 Practical Considerations 1
1.3 Instrumentation 2
2 Biomolecular Spectral Assignments **2**
2.1 Two-dimensional Nuclear Magnetic Resonance Experiments 2
2.2 Three-dimensional and Four-dimensional Experiments 5
2.3 Sequence-specific Assignments 7
3 Biomolecular Structure Determination **10**
3.1 Assigning Nuclear Overhauser Effect Intensities 10
3.2 Use of *J* Couplings to Obtain Torsion Angles 10
3.3 Distance Geometry 11
3.4 Molecular Dynamics and Simulated Annealing 12
3.5 Energy Minimization 14
3.6 Procedure for Structure Determination 15
3.7 Assessing the Accuracy and Precision of the Final Structure 15
4 Obtaining Long-range Distance and Angle Information **16**
4.1 Use of Shift Tensors to Obtain Torsion Angles 16
4.2 Use of Residual Dipolar Couplings to Obtain Distance and Angle Restraints 17
4.3 Solid-state Nuclear Magnetic Resonance in Biomolecular Structure Determination 17
5 Examples of Biomolecular Nuclear Magnetic Resonance **21**
5.1 Proteins and Nucleic Acids 21
5.2 Enzymes 23
5.3 Membranes 26
Acknowledgments **29**
Abbreviations and Acronyms **29**
Related Articles **30**
References **30**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Nuclear magnetic resonance (NMR) spectroscopy exploits a property of the nucleus known as nuclear spin, which exhibits angular momentum and as such generates a local magnetic field that can be influenced by a larger static external field. In the presence of a large external magnet, the nuclear spins align with the field, with a slight excess opposed to the field. That small excess can be flipped to alignment with the field in the presence of a radiofrequency (RF) pulse of appropriate frequency. Relaxation of those spins back to the ground state results in emission of RF radiation whose frequency is an indication of the local electron density of the molecule in which the spin resides. Such information can ultimately be directly related to molecular structure. NMR spectroscopy is unique in that it can be used to image the whole human body at one extreme, and determine the three-dimensional (3D) structure of a biomolecule within the body at the other extreme. The principal advantages of NMR are that: (i) the complete molecular structure can be determined; (ii) the technique spans all principal states of matter (solid, liquid, gas); (iii) the method is nondestructive and can be used noninvasively, as in a clinical setting. The principal disadvantages of NMR are that: (i) the technique is very insensitive (typically samples below 50 μM are difficult to study, although recent advances in nanoprobe and cryprobe technology are rapidly revising these numbers); (ii) in liquids there are molecular weight limits of around 50 kDa – although there are new methods being developed that may have a significant impact on this; (iii) in solids only limited structural information can be obtained to date.

1 INTRODUCTION

1.1 How Nuclear Magnetic Resonance is Used to Study Biomolecules

The methods of NMR spectroscopy have proved to be an essential analytical method for the determination of the structures of biomolecules, and in the study of biomolecular function. This article provides an overview of the principal NMR methods that can be employed in structure–function studies of biomolecules, with the main focus on structure determination. Many texts discuss these topics in greater detail than can be considered here.[1–10]

1.2 Practical Considerations

The success or failure of any biomolecular NMR project is completely dependent upon the quality of spectra generated by the particular samples under study. For

example, although it is generally true that proteins with $M_r > 50$ kDa are largely intractable projects by liquid-state NMR, there are examples of proteins with $M_r < 10$ kDa that give impossibly broad spectra, and proteins with $M_r > 60$ kDa that, with appropriate measures, give reasonably well-resolved spectra. Thus no NMR project should be ruled in or out without extensive preliminary tests of conditions to investigate whether the project is viable. In this sense, biomolecular NMR spectroscopy is similar to X-ray crystallography, where obtaining crystals is equivalent to getting a spectrum, and getting diffraction-quality crystals is equivalent to getting well-resolved spectra. Just as the X-ray crystallographer searches a wide variety of crystallization conditions, including testing temperature, pH, ionic strength, solvent and so on, so the NMR spectroscopist must carry out similar kinds of studies.

1.3 Instrumentation

For most biomolecular NMR studies, a high-field NMR spectrometer is necessary, at least 500 MHz (for ^{1}H, 11.7 T) for liquids and 400 MHz (9.4 T) wide-bore (89 mm) for solids – bigger is generally better (although there are reasons for doing studies at particular field strengths, as discussed later). For a biomolecular liquid-state NMR structure determination, generally a 600 MHz (14 T) instrument is preferable, whereas for biomolecular solid-state NMR structure determination (although there are advantages in going to higher fields) there are also technical issues that presently limit the kinds of experiments possible. The highest commercially available field strength (in early 1999) is 800 MHz (18.8 T), although a 900 MHz (21.2 T) instrument is scheduled for completion by summer 2000. To date, 800 MHz has become the field strength of choice in leading protein NMR laboratories doing liquid-state NMR, and a number of solid-state NMR laboratories are also exploring this field strength in their work.

2 BIOMOLECULAR SPECTRAL ASSIGNMENTS

The first step in biomolecular NMR spectroscopy, in particular by ^{1}H liquid-state NMR, is the assignment of the NMR spectrum. For this, a general strategy for identification and assignment of resonances is usually adopted, which employs a number of two-dimensional (2D) or higher (3D, four-dimensional (4D)) NMR experiments. Some of the basic experiments are discussed here, followed by an account of how they are employed in carrying out sequence-specific assignments. For large biomolecules these approaches fail and alternative methods are considered in section 2.3.3 for such situations. Once the assignments have been obtained, then this information can be used to solve the biomolecular structure (section 3) or used to interpret biochemical events (section 5.2).

2.1 Two-dimensional Nuclear Magnetic Resonance Experiments

2.1.1 Correlation Spectroscopy

The most important 2D NMR experiment is the correlation spectroscopy or COSY experiment, introduced by Jeener in 1971. The sequence consists of $90^\circ - t_1 - 90^\circ$ and, in the presence of homonuclear scalar J coupling, it gives information about coupling networks. In a coupled system, the second pulse of the experiment causes magnetization that arose from one transition during t_1 to be redistributed amongst all the others with which it is associated. With this sequence of pulses, in the presence of coupling, magnetization evolves with different frequencies during the two time intervals because coherence is transferred between multiplets during the second pulse.

Another important variant on the basic COSY sequence is double-quantum-filtered correlation spectroscopy (i.e. DQFCOSY), in which an extra $\pi/2$ pulse is added immediately after the end of the COSY sequence (Equation 1):

$$\frac{\pi}{2}_{\phi} - t_1 - \frac{\pi}{2}_{\phi} - \Delta - \frac{\pi}{2}_{x} - t_2(\text{acquire}) \quad (1)$$

where Δ is a short phase switching delay, and multiple quantum coherences that existed before the third pulse are converted back to observable magnetization. With a suitable choice of phase cycling, signals that have arisen from different orders of multiple quantum coherence can be separated out. For instance, double quantum coherence is twice as sensitive to phase changes in its excitation sequence as is single quantum coherence. Therefore by shifting the phase ϕ of the first two pulses (in general, *all* pulses before the coherence is created) by 90°, the phase of the detected signal which came via double quantum coherence is inverted. Inverting the receiver phase (i.e. subtracting the 90° from the 0° experiment) then selects the component which passed through double quantum coherence. To select p-quantum coherence, the rule is to step the phases of the excitation pulses through the sequence $0, 180/p, \ldots (2p-1) \times 180/p$, alternating the receiver phase for each scan. The advantages are considerable improvement in spectral simplification. Singlets are largely eliminated in DQFCOSY. Furthermore the diagonal *and* the cross-peaks are in pure absorption

phase, improving the resolution. The disadvantage is that DQFCOSY is less sensitive by a factor of 2 than conventional COSY.

2.1.2 Total Correlation Spectroscopy

Total correlation spectroscopy (TOCSY) or its variant homonuclear Hartmann–Hahn (HOHAHA) spectroscopy is an experiment that relies on cross-polarization (CP) rather than coherence transfer, and whose pulse sequence is shown in Figure 1.

The spin-locking field causes cross-polarization, or an oscillatory exchange of spin-locked magnetization between two spins, provided that the effective local RF fields experienced by the two spins are identical. This causes the spins to become temporarily equivalent. The spin-lock sequences are termed "isotropic mixing" (with no effective field) or "Hartmann–Hahn mixing" (with an effective field, i.e. this is the HOHAHA experiment). Isotropic mixing leads to spectra with mixed phases (phase modulated), whereas in Hartmann–Hahn mixing, pure phases are obtained. For a simple two-spin system, complete exchange of magnetization occurs for contact times (or spin-lock mixing times) equal to $1/(2J)$, where J is the two-spin scalar coupling constant. For biological macromolecules, which involve large spin systems, there is no simple analytical solution, and in practice spectra are usually obtained with a variety of spin-lock mixing times.

2.1.3 Heteronuclear Single Quantum Coherence and Heteronuclear Multiple Quantum Coherence

One of the most important experiments in biomolecular NMR is the heteronuclear single quantum coherence (HSQC) correlation experiment, and its more general form, the heteronuclear multiple quantum coherence (HMQC)[11] experiment (sometimes referred to as the "forbidden echo" experiment).[12–14] These experiments exploit the intrinsically greater sensitivity of protons to indirectly detect a heteronucleus such as ^{13}C and ^{15}N. The basic HMQC pulse sequence is shown in Figure 2.

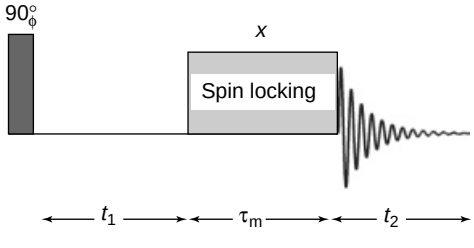

Figure 1 The TOCSY pulse sequence. (Adapted from ref. 8 with permission.)

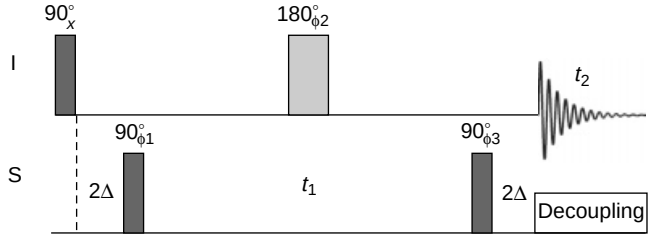

Figure 2 The basic HMQC pulse sequence. The delay Δ is usually set to ~20% less than $1/4(J_{IS})$. (Adapted from ref. 8 with permission.)

In terms of product operators, a heteronuclear spin system IS is represented by Equation (2),

$$I_z \xrightarrow{\text{A}} -2I_xS_y \xrightarrow{t_1} -2I_xS_y \cos w_S t_1 \xrightarrow{\text{B}} -I_y \cos w_S t_1 \tag{2}$$

where A and B are operators representing the effect of the pulse sequence before and after the t_1 period, respectively, and the pulse phases are assumed to be $+x$. Equation (2) indicates that the proton signal (I_y) which is detected will be in-phase and amplitude modulated by w_S, giving rise to a single absorption line in the 2D spectrum. Considering a second coupled proton M, then the effect of the small homonuclear coupling J_{IM} during the short delays Δ can be neglected. During t_1, dephasing caused by J_{SM} coupling is refocused by the 180° pulse. However, the effect of any J_{IM} coupling is not refocused, because both coupled nuclei are influenced by the 180° pulse. Therefore the effect of the pulse sequence is given by Equation (3):

$$\begin{aligned} I_z \xrightarrow{\text{A}t_1\text{B}} & -I_y \cos \pi J_{IM} t_1 \cos w_S t_1 \\ & + 2I_xM_z \sin \pi J_{IM} t_1 \cos w_S t_1 \end{aligned} \tag{3}$$

The detected signal will consist of four multiplet components at $w_S \pm \pi J_{IM}$ and $w_I \pm \pi J_{IM}$, and will contain both in-phase absorptive (I_y) and antiphase dispersive ($2I_xM_z$) components.

The pulse sequence of the HSQC experiment is shown in Figure 3.

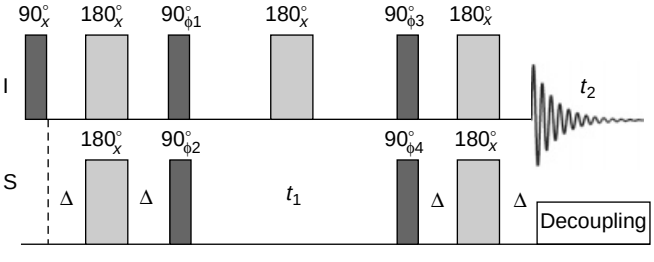

Figure 3 The HSQC pulse sequence. (Adapted from ref. 8 with permission.)

This is a variation on an experiment introduced by Bodenhausen and Ruben[15] and employs two insensitive nuclear enhancement by polarization transfer (INEPT)-type[16,17] transfers to transfer magnetization from the protons to the low-γ nucleus and back to the protons. The INEPT sequence transfers I_{spin} (e.g. ^{1}H) I_z magnetization into antiphase S-spin magnetization. Decoupling the I-spin during t_1 is accomplished by application of a 180° pulse at the midpoint of t_1. A subsequent INEPT transfer reconverts the transverse S magnetization into observable I magnetization. Equation (4) expresses this in terms of product operators:

$$I_z \xrightarrow{\text{INEPT}} -2I_zS_y \xrightarrow{t_1} -2I_zS_y \cos w_S t_1 \xrightarrow{\text{reverse INEPT}} -I_x \cos w_S t_1 \tag{4}$$

where for clarity only those terms that contribute to the final spectrum are retained, and the phase of the pulses are given that correspond to the first step in the phase cycle. Pure absorptive lineshapes are obtained. Furthermore the linewidth is narrower than with the HMQC experiment, because it is not affected by homonuclear or heteronuclear J couplings and determined by the average relaxation rates of S_x and I_zS_y. There are a number of advantages that HSQC has over HMQC, principally in terms of sensitivity.[18–20] It is for this reason that HSQC has emerged as the gold standard for determining the viability of a biomolecular NMR project. A ^{1}H–^{15}N HSQC is routinely obtained on a protein to establish whether a particular sample gives well-resolved and well-dispersed spectra.

2.1.4 Nuclear Overhauser Effect Spectroscopy

The 2D experiment that permits the detection of the nuclear Overhauser effect (NOE) is the NOESY (nuclear Overhauser effect spectroscopy) pulse sequence given in Equation (5):

$$\frac{\pi}{2}_{\phi} - t_1 - \frac{\pi}{2}_{\phi} - \tau_m - \frac{\pi}{2}_{\phi} - \text{Acquire}(t_2) \tag{5}$$

The first 90°_x pulse creates transverse xy magnetization, and the spins now precess during t_1 as in the COSY sequence. The second 90°_x pulse rotates components of the magnetization along the $-z$ axis. During the subsequent mixing period τ_m, z-magnetization components exchange under the influence of cross-relaxation. The transverse components are not required and are removed by phase cycling. The magnetization vectors therefore have a component in the z direction only after these procedures. Finally, a third 90°_x pulse regenerates observable magnetization. If this sequence is repeated for a larger value of t_1, the magnetization vectors would dephase further, and a smaller $-z$ component is thus created, which would pass through zero and become positive for increasing values of t_1. The changes are "read" by the final 90° pulse. The cross-peaks are generated from the magnetization transfer (cross-relaxation) between two spins.

Note that in any practical application, a variety of cross-peaks can be generated by mechanisms other than cross-relaxation, and these cause confusion in interpreting the data. In addition, chemical exchange cross-peaks occur (which can be valuable if chemical exchange is known to be taking place), but can be a source of undesirable confusion. Furthermore, in a weakly coupled system the cross-relaxation pathways arising from coherence transfer between J-coupled spins has to be considered. By analogy with the COSY pulse sequence, the system remains the same until after the second 90° pulse, when the density operator takes the form of Equation (6):

$$\begin{aligned}\sigma_3 = [&-I_z \cos w_I t_1 + I_x \sin w_I t_1 - S_z \cos w_S t_1 \\ &+ S_x \sin w_S t_1] \cos \pi J_{IS} t_1 - [-2I_xS_y \cos w_I t_1 + 2I_yS_x \\ &\times \cos w_S t_1 - 2I_zS_y \sin w_I t_1 + 2I_yS_z \sin w_S t_1] \sin \pi J_{IS} t_1\end{aligned} \tag{6}$$

The terms $2I_xS_y$ and $2I_yS_x$ consist of a superposition of double- and zero-quantum coherences, which can be shown by expanding σ_3 (Equation 7):

$$\begin{aligned}\sigma_3 = [&-I_z \cos w_I t_1 + I_x \sin w_I t_1 - S_z \cos w_S t_1 \\ &+ S_x \sin w_S t_1] \cos \pi J_{IS} t_1 + \tfrac{1}{2}[(2I_xS_y + 2I_yS_x)^{[1]} \\ &- (2I_yS_x + 2I_xS_y)^{[2]}] \cos w_I t_1 + \tfrac{1}{2}[(2I_xS_y + 2I_yS_x)^{[3]} \\ &+ (2I_yS_x + 2I_xS_y)^{[4]}] \cos w_S t_1 + 2I_zS_y \sin w_I t_1 \\ &+ 2I_yS_z \sin w_S t_1 \quad \sin \pi J_{IS} t_1\end{aligned} \tag{7}$$

where the terms labeled [1] and [3] represent double quantum coherence, and those labeled [2] and [4] represent zero quantum coherence. As we are only interested in longitudinal magnetization components during the mixing period τ_m, it is necessary to remove all the components except $-I_z \cos w_I t_1$ and $-S_z \cos w_S t_1$. In order to do this, we can phase cycle the first two pulses by 90° each: xx, yy, $-x$, $-x$, $-y$, and $-y$, and add each separate response, then this will cancel coherences of order 1, 2, 3, 5, ... and leave the zero quantum coherences. Thus at the beginning of the mixing period Equation (8) holds:

$$\begin{aligned}\sigma_3 = &(-I_z \cos w_I t_1 - S_z \cos w_S t_1) \cos \pi J_{IS} t_1 \\ &+ (I_yS_x - I_xS_y)[\cos w_S t_1 - \cos w_I t_1] \sin \pi J_{IS} t_1\end{aligned} \tag{8}$$

The zero quantum term $(I_yS_x - I_xS_y)$ evolves in a way that depends on the difference in the chemical shifts, and the longitudinal $(-I_z)$ terms mix under the effects of cross-relaxation. After the final 90°_x pulse, the observable magnetization is given by Equation (9):

$$\begin{aligned}\sigma_4^{\text{obs}} = [&I_y a_{II}\cos w_I t_1 + S_y a_{SS}\cos w_S t_1 + I_y a_{SI}\cos w_S t_1 \\ &+ S_y a_{IS}\cos w_I t_1]\cos \pi J_{IS} t_1 + (I_z S_x - I_x S_z) \\ &\times \cos(w_I - w_S)\tau_{\text{m}}(\cos w_S t_1 - \cos w_I t_1)\sin \pi J_{IS} t_1\end{aligned} \tag{9}$$

The terms in σ_4 represent diagonal peaks proportional to the mixing coefficients a_{II} and a_{SS}, and NOE cross-peaks proportional to a_{SI} and a_{IS}. These are all in phase with respect to J_{IS}. Also, there are diagonal and cross-peaks that are antiphase with respect to J_{IS}. These are the J cross-peaks derived from zero quantum coherence that cannot be removed by phase cycling, and have to be eliminated by other methods (e.g. random incremental variation of τ_{m}).

For a three-spin system, I, S, and M, the I-spin diagonal peak intensity governed by a_{II} decays with increasing τ_{m}, whereas the S- and M-spin cross-peak intensities, generated by a_{IS} and a_{IM}, increase before decaying with increasing τ_{m}. The importance of the time development of cross-peak intensities is that the initial build-up rates of the S- and M-spin cross-peaks (i.e. at $\tau_{\text{m}} = 0$) are proportional to σ_{IS} and σ_{IM}. If, therefore, a_{IM} corresponds to a known fixed internuclear distance r_{IM}, r_{IS} can be calculated from the r^{-6} dependence of σ (Equation 10):

$$\frac{r_{IM}}{r_{IS}} = \left(\frac{\sigma_{IS}}{\sigma_{IM}}\right)^{\frac{1}{6}} \tag{10}$$

This is the principal method by which quantitative distance information is obtained in biomolecular NMR. A series of NOESY spectra is recorded as a function of mixing time in order to generate experimental plots similar to those obtained experimentally, as shown above. Alternatively, if one can be certain that the measured cross-peak intensities in a single NOESY experiment for a given mixing time were developed within the initial rate approximation, then quantitative distance information can be obtained with good accuracy by measurement of relative cross-peak intensities rather than initial rates (Equation 11):

$$\frac{r_{IM}}{r_{IS}} = \left(\frac{a_{IS}}{a_{IM}}\right)^{\frac{1}{6}} \tag{11}$$

Note that this is true for the initial rate approximation only.

2.2 Three-dimensional and Four-dimensional Experiments

One of the problems with ^{1}H NOESY NMR of proteins is that all the NOEs must be resolved. This becomes increasingly difficult with higher-molecular-weight proteins. To solve this problem, a whole family of pulse sequences have been introduced that involve three time variables rather than two, so that 3D Fourier transformation leads to a cube rather than a contour map. They involve using an isotopic label for indirect detection via HMQC/HSQC methods in combination with a type of COSY or NOESY sequence. Almost all of these 3D (and 4D) pulse sequences rely on heteronuclear and homonuclear couplings for correlating resonances from

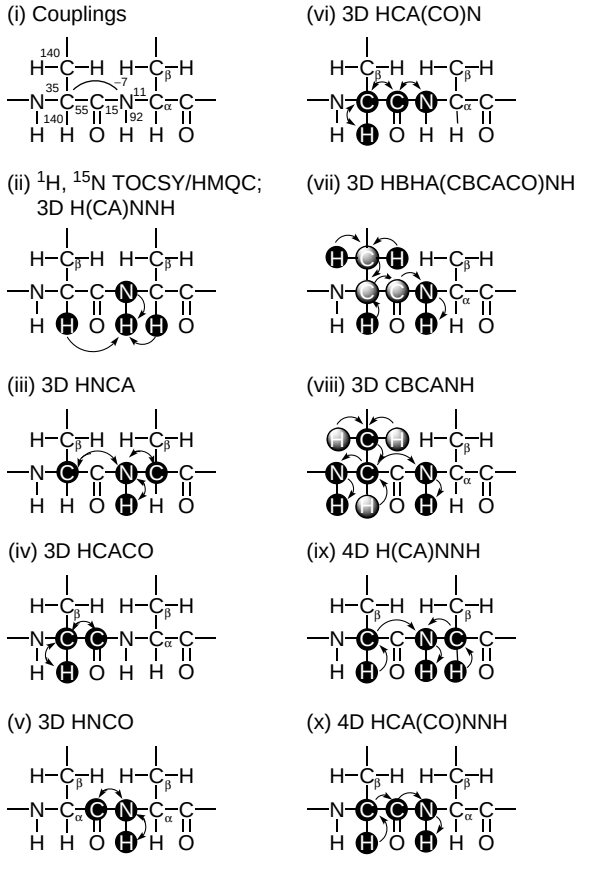

Figure 4 A dipeptide segment of a protein backbone with typical values of homonuclear and heteronuclear J couplings, and showing the correlations obtained using the selection of experiments indicated. Nuclei for which the chemical shift is measured in the 3D (or 4D) experiment are marked by dark shaded circles. Nuclei involved in the magnetization transfer pathway, but not observed, are marked by light shaded circles. Magnetization transfer is marked by curved solid lines, and the direction of the transfer indicated with arrows. Note that sequences involving small J couplings are generally not suitable for large proteins and are usually less sensitive. (Adapted from ref. 8 with permission.)

one residue to the next, in for example an oligopeptide, oligonucleotide or oligosaccharide. Rather than provide a detailed description of all the pulse sequences used in modern biomolecular NMR, they are summarized in Figure 4. The *J* couplings involved, and the correlations which each sequence can obtain, are illustrated in Figure 5 for a dipeptide fragment.

The majority of experiments involve a heteronucleus, and coherence is transferred from one nucleus to the next, usually ending up on ^{1}H, which is then detected. Such coherence transfers are achieved using INEPT and related approaches. As is clear from Figure 5, the amide protons are correlated with the α carbons using the HNCA[21–25] and HN(CO)CA[25,26] experiments. The HNCA provides intraresidue correlations (H_i^N N_i^H $C_i^\alpha\{C_{i-1}^\alpha\}$), whereas the HN(CO)CA provides exclusively interresidue correlations (H_i^N N_i^H C_{i-1}^α). The latter experiment is more sensitive and is complementary to the HNCA experiment. The amide protons are correlated with the α protons using the H(CA)NNH, HN(CA)H, HN(CA)HA, and HN(COCA)HA experiments. The H(CA)NNH,[27,28] HN(CA)H,[29] and HN(CA)HA[30] experiments all provide the same correlations (H_i^N N_i^H $H_i^\alpha\{H_{i-1}^\alpha\}$). The HN(CA)NNH situation is not very sensitive for $M_r >$ 20 kDa, and HN(CA)HA is significantly more sensitive and has been used to assign a 27-kDa protein.[31] The HN(CA)H has even more improved sensitivity, particularly for larger proteins (>30 kDa). The HN(COCA)HA[32] correlates H_i^N N_i^H H_{i-1}^α, and is analogous to the HN(CO)CA, providing exclusively interresidue correlations, and is therefore complementary

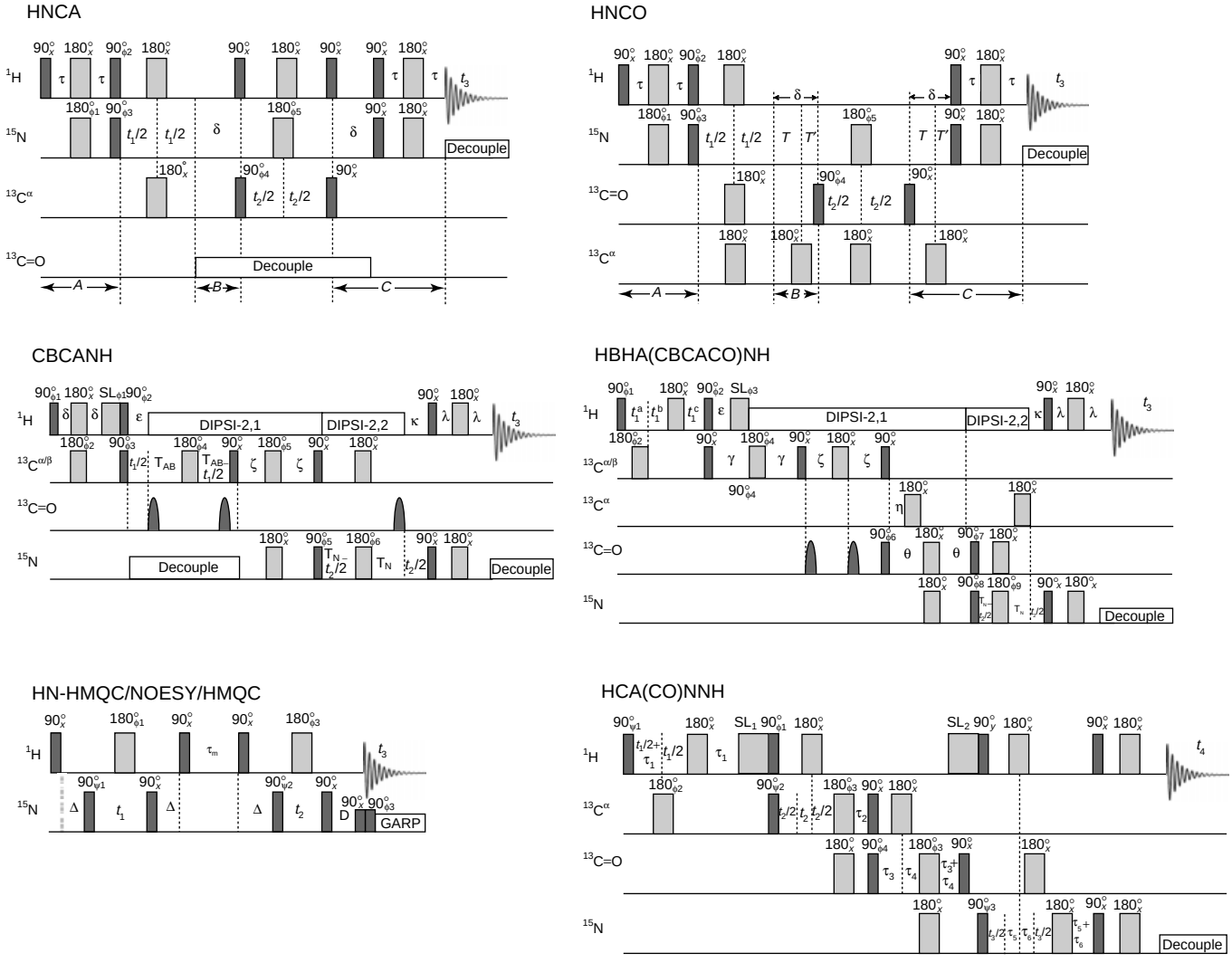

Figure 5 Some examples of 3D and 4D pulse sequences. These are presented in their simplest form, without pulse field gradients or constant-time implementations. (Adapted from ref. 8 with permission.)

to the H(CA)NNH, HN(CA)H, and HN(CA)HA experiments. The amide protons are correlated with the carbonyl carbons using the HNCO and HN(CA)CO experiments. The HNCO[21,22,33,34] experiment correlates H_i^N N_i^H C'_{i-1}, and its high sensitivity makes it attractive in assignment strategies. The HN(CA)CO[25,35] provides intraresidue and some weaker interresidue correlations, but is much less sensitive, and therefore complementary to the HNCO experiment. The amide protons are correlated with the α and β protons using the CBCA(CO)NH, CBCANH, HNCACB, HBHA(CBCACO)NH, and HBHA(CBCA) NH experiments. The CBCA(CO)NH[33,36] experiment correlates H_i^N N_i^H $C_{i-1}^{\alpha,\beta}$, and is an extension of the HN(CO)CA experiment, as it correlates both the C^β and C^α chemical shifts with the amide from the previous residue. The CBCANH[37] experiment correlates H_i^N N_i^H $C_i^{\alpha,\beta}\{C_{i-1}^{\alpha,\beta}\}$, and is an extension of the HNCA experiment, providing mainly intraresidue correlations, but with some interresidue correlations. Although in principle this one experiment can provide complete assignments for C^β, C^α, ^{15}N, and H^N resonances, like the HNCA experiment, the sensitivity is limited, and it is mainly useful for smaller proteins ($M_r < 20$ kDa). The HNCACB[33,38] experiment provides the same information as the CBCANH, but is more sensitive. By using these experiments in conjunction with the HBHA(CBCACO)NH[39] (H_i^N N_i^H $H_{i-1}^{\alpha,\beta}$) and the HBHA(CBCA)NH[40] (H_i^N N_i^H $H_i^{\alpha,\beta}\{H_{i-1}^{\alpha,\beta}\}$) experiments, the assignment can be extended to the carbon-attached protons as well. In all these experiments, both the α and β correlations appear in the same dimension.

The HCACO[21,22,41–43] experiment (H_i^α C_i^α C'_i) and the HCA(CO)N[21,22,41] experiment (H_{i-1}^α C_{i-1}^α N_i^H) are both usually acquired with a sample in D_2O, in order to allow observation of those α protons otherwise obscured by the water resonance. This has the disadvantage of requiring two different samples in which small differences in preparation conditions can lead to differences in chemical shifts, and the presence of the deuterium leads to isotope shifts. However, the HCACO is more sensitive than other experiments such as the HN(CA)HA. Also, the HCA(CO)N is useful for identifying residues preceding proline residues. There are also several other 3D sequences that have been proposed, such as the HN(CA)NNH[44] (H_i^N N_i^H N_{i+1}^H), H(NCA)NNH[44] (H_i^N N_i^H N_{i+1}^H), H(N)CACO[45] (H_i^N C_i^α C'_i), and HA[CAN]HN[46] ($H_i^N[N_i^H + C_i^\alpha]$ H_i^α). These have not been as widely used in protein assignment strategies. Finally there is also the HCACOCANH (H_i^N N_i^H C'_i) experiment,[47] which appears to be gaining in popularity.

The most common experiments for side-chain assignment are the HCCH COSY[48] and the HCCH TOCSY.[49–51] These provide essentially the same information (H_i^x C_i^x H_i^y with C^x and C^y being one bond apart) as the 2D COSY/TOCSY experiments, except that the COSY/TOCSY transfer is between carbon rather than proton nuclei. The third dimension correlates the shifts of the ^{13}C nuclei attached to the protons on the diagonal which reduces the overlap. However, one disadvantage is that they are generally acquired with the sample dissolved in D_2O. One way round this is the HNCCH TOCSY[52] experiment, which correlates the amide nitrogens, the α carbons, and the side-chain protons on a sample in H_2O. Alternatively, there are the C(CO)NH (H_i^N N_i^H $C_{i-1}^{x,y,z,\cdots}$) and H(CCO)NH (H_i^N N_i^H $H_{i-1}^{x,y,z,\cdots}$) experiments,[53] which are both more sensitive.

There are a number of 4D experiments, although they all suffer from poor sensitivity. It is a necessity that 4D experiments use pulsed field gradients for coherence selection [54–56] with a minimum of phase cycling. They have been used as the only source of backbone assignments.[57] The HCANNH (H_i^N N_i^H $H_i^\alpha\{C_{i-1}^\alpha$ $H_{i-1}^\alpha\}$) and HCA(CO)NNH (H_i^N N_i^H H_i^α C_{i-1}^α H_{i-1}^α) experiments[58] are complementary, and can provide complete sequential assignments for all the backbone atoms apart from the carbonyl carbons. The HCANNH experiment gives intraresidue correlations and a few interresidue correlations; the HCA(CO)NNH gives exclusively interresidue correlations. Improved constant-time versions have also been introduced.[59] The HNCAHA and HN(CO)CAHA experiments[60,61] together provide the correlations, but use different pulse sequences. In general the HCANNH is preferable for small proteins, whereas the HNCAHA is more sensitive for proteins with $M_r > 20$ kDa. The HCC(CO)NNH[62] experiment (H_i^N N_i^H $C_{i-1}^{x,y,z,\cdots}$ $H_{i-1}^{x,y,z,\cdots}$) is based on the HCA(CO)NNH and allows the assignment of the side-chain as well as the α nuclei.

The use of these experiments in carrying out sequential assignments is outlined in the next section.

2.3 Sequence-specific Assignments

2.3.1 Sequential Assignments for Proteins with Relative Molecular Mass below 15 kDa

The standard notation used for distances in proteins is that the distance between the hydrogen atoms A and B located in the amino acid residues in the sequence positions i and j, respectively, is denoted by $d_{AB}(i, j)$. This notation can be applied to all hydrogen atoms, but in practice the distances between different backbone hydrogen atoms and between backbone hydrogens and βCH_n groups are of particular interest. The standard approach for the sequential assignments of a small ($M_r < 15$ kDa) protein is to carry out the following four steps:

1. The spin systems of nonlabile protons in individual amino acids are identified for the native protein in D_2O by DQFCOSY and TOCSY experiments.
2. From studies in H_2O, the identity of amino acid spin systems of the individual residues are completed through identification of the J connectivities with the labile protons.
3. Sequentially neighboring amino acid ^1H spin systems are identified from observation of the sequential NOE connectivities $d_{\alpha N}$ or d_{NN}, or possibly $d_{\beta N}$.
4. The aim of steps 1–3 is to identify groups of ^1H-NMR lines corresponding to peptide segments that are sufficiently long to be unique in the primary sequence, or a single residue which is unique. Sequential assignments are then obtained by matching the peptide segments thus identified by NMR with the corresponding segments in the independently determined (not by NMR) amino acid sequence.

For unique residues, the sequential assignment follows directly from the identification of the spin system. For the resonances of unique dipeptide segments, sequential assignments result if a sequential connectivity between the ^1H spin systems of the two residues can be established. The unique residues and dipeptide segments can then be used as reference locations in the primary structure, so that further sequential identification steps directly provide the sequential assignments for the adjacent residues. The characteristic chemical shifts for random extended chain structures (sometimes inappropriately called random coil) are available.[8] Also the amino acid residues display characteristic cross-peak fine structures in their DQFCOSY spectra. Using this type of information, generally all but the side chains of Lys and Arg beyond the βCH_2 can be readily identified for proteins with $M_r \leq 15\,000$. Other techniques (e.g. TOCSY) are needed for the complete assignment of all the resonances for Lys and Arg.

The NOEs characteristic of short distances can be identified by comparing NOESY and COSY spectra. Common NOEs correspond to the sequential distances for the i to the $i+1$ residues. The importance of observing the NH signals in H_2O is clear. Finally, comparison of the α helix and β sheet distances shows that a regular secondary structure can be spotted and categorized by its NOE pattern. α-Helices have particularly short sequential d_{NN} values, and β sheets have short $d_{\alpha N}$ values. Even more characteristic are the short $d_{\alpha N}(i, i+3)$ and $d_{\beta N}(i, i+3)$ distances found in helices, and the short interstrand $d_{\alpha N}(i, j)$, $d_{NN}(i, j)$ and $d_{\alpha\alpha}(i, j)$ distances found in β sheets. Indeed, even without any sequential assignments, the presence of $d_{\alpha\alpha}$ NOEs can be used as a crude diagnosis for antiparallel β sheets.

2.3.2 *Sequential Assignments for Proteins with Relative Molecular Mass in the Range 15–50 kDa*

For larger proteins the problem of overlapping resonances and increased linewidths can become severe, and the solution is to turn to isotopic labeling and 3D heteronuclear NMR. This is because the heteronuclear couplings, $^1J_{CH}$ (125–160 Hz) and $^1J_{NH}$ (≈92 Hz), are much larger than $^3J_{HH}$, and frequently as much as 50–90% of the magnetization can be transferred from protons to their directly coupled heteronuclei.[63] Consequently, such heteronuclear shift correlation experiments are highly sensitive. The strategy for making 3D heteronuclear NMR assignments relies exclusively on homonuclear and heteronuclear J coupling, which avoids the use of NOE connectivities, and makes use of several different experiments illustrated in Figure 5. The strategy was introduced originally by Markley et al.[64,65] The advantage of avoiding NOE-based strategies for the assignment of larger protein spectra is that it is largely independent of protein conformation, and relies only on one-bond coupling. However, each sequential connectivity depends upon a single correlation. Degeneracy of carbon and nitrogen resonances can be more hampering than degeneracy of proton resonances in the case of the NOE-based assignments, where observation of more than one of the $d_{\alpha N}$, d_{NN} and $d_{\beta N}$ connectivities can resolve ambiguities. Therefore, these heteronuclear strategies can also be combined with homonuclear NOESY experiments, although the presence of isotopic labeling, for example with ^{13}C, can exacerbate the linewidth problem for large proteins because the $^1H-^{13}C$ dipolar interaction causes additional proton line broadening.

Typical values for the relevant coupling constants are indicated in Figure 5. Equally important are the magnitudes of the t_2 relaxation times, which determine the linewidth and depend approximately linearly on the rotational correlation time, and on the degree of internal mobility and local conformation. Generally, for residues that do not have a high degree of internal mobility in a globular protein of $M_r \approx 20$ kDa at 35 °C, the linewidths are about 12 Hz for the NH proton, for ^{15}NH nitrogen ≈7 Hz (^1H coupled) and ≈4 Hz (^1H decoupled), ≈15 Hz for $^{13}C^{\alpha}$, and ≈25 Hz for a ^{13}C-attached $^1H^{\alpha}$. The linewidth of the carbonyl is dominated by chemical shift anisotropy and therefore is proportional to the square of $\mathbf{B}_0$, and ≈6 Hz at 11.7 T. Comparison of these values with those shown in Figure 5 indicate that for a 20 kDa protein most one-bond J couplings are significantly larger than the linewidths, which means that magnetization can be transferred efficiently from one spin to its directly attached coupling partner. Use of a combination of these experiments permits the complete sequential assignments for proteins up to a maximum

$M_r \approx 50$ kDa. For proteins with favorable resonance dispersion and sensitivity, the HNCA, HCACO, and HN(CA)CO experiments may not be necessary to complete the assignments.[66]

Common assignment strategies for heteronuclear triple resonance experiments often start with correlating $^1H/^{15}N$ shift pairs for backbone amides (from HSQC or, if too overlapped, HNCO experiments) with other nuclei via the HNCA experiment. Although other correlations can in principle be achieved using a HN(CO)CA experiment, in practice a combination of other experiments are usually carried out. For example, the HNCA, TOCSY/HMQC, NOESY/HMQC, HBHA(CO)NH, CBCA(CO)NH, and CBCANH together give correlations between the $^1H/^{15}N$ amide resonances and the intraresidue/sequential C^α, C^β, H^α, and H^β shifts. However, using the C^α and C^β shifts and the results from a C(CO)NH experiment, the amino acid type can be determined for each $^1H/^{15}N$ pair. This can be confirmed using the 4D HN(COCA)NH experiment.[67] For larger proteins, a typical strategy would be to use HNCA, HN(CO)CA, and HNCO experiments to correlate the NH_i pairs to the C_i^α, $\{C_{i-1}^\alpha\}$, and C'_{i-1} shifts with high sensitivity. The HN(CA)CO experiment can provide intraresidue C'_i correlations, and TOCSY/HMQC can provide H_i^α correlations, although with lower sensitivity. The HCACO can be used to correlate the H_i^α, C_i^α, and C'_i shifts for all residues other than glycine. The assignments obtained by this means can then be checked against the HN(CA)HA, HA(CACO)NH, and the side-chain resonances of the TOCSY/HMQC experiment. The glycine resonances can be checked against an appropriately optimized (for CH and CH_2 groups) HCACO experiment. The α-carbon shifts of glycine can generally be found from the HCCH COSY and correlated with NOESY/TOCSY/HMQC. These strategies can always break down at any point, and alternative experiments can be employed to fill in the gaps.

The key to carrying out multidimensional NMR studies on proteins with $M_r < 50$ kDa, and structural investigations on selected regions of proteins with $M_r > 50$ kDa, involves the use of isotopic labeling. This can be either universal labeling or selective labeling. The isotopes commonly employed under these circumstances are ^2H, ^{13}C, and ^{15}N. For isotopic labeling to be practical, it is essential that the protein of interest is overexpressed at a sufficiently high level in an organism that is easy to handle. Generally this means that the gene encoding a particular protein should be inserted into a plasmid with a good promoter, and most conveniently in a suitable strain of *Escherichia coli*. A reasonable target for levels of overexpression for ^{13}C,^{15}N double labeling should be in the region of 10–50 mg of purified protein from 1 L of cells. Lower levels of overexpression can be tolerated for universal labeling with ^{15}N, as $^{15}NH_4Cl$ is relatively cheap.

Deuterium labeling can be used to simplify ^1H-NMR spectra of proteins.[68,69] The modern approach has focused on reducing the $^1H-^1H$ dipolar interaction through fractional deuterium labeling.[70–72] There are a number of advantages to the use of deuteration for NMR studies of larger proteins. Random deuteration helps to reduce spin diffusion, and incorporation of stereospecifically deuterated amino acids helps to resolve degenerate proton resonances and enable quantitative NOE analysis. Chiral deuteration can aid in stereospecific assignments, but is rarely employed because it is experimentally difficult.

2.3.3 Methods for Extending the Molecular Weight Limit

There are a number of approaches for extending the molecular weight limit that NMR imposes on the study of biomolecules. At the root of the problem is the fact that larger molecular weight particles tumble more slowly, generally leading to longer correlation times (τ_c), which in turn leads to broader lines. Therefore approaches that narrow these broad lines are what is required. One rather obvious approach is to stop the molecular tumbling by going to the solid state, and then to employ the methods of solid-state NMR such as cross-polarization/magic-angle spinning (CP/MAS) to narrow the lines. These techniques throw away important information, such as the dipolar coupling, so techniques to reintroduce these couplings are necessary (section 4) for solid-state NMR to be useful in structural biology. In solution, a number of approaches have been introduced which can be used separately or in concert. Use of deuterium labeling, as discussed in the previous section, helps to reduce linewidths substantially. Also Hahn spin–echo sequences can be used to edit with respect to broad lines in the presence of narrow lines. Recently a new approach called transverse relaxation optimized spectroscopy (TROSY) has been introduced[73] which relies on the cancellation of two competing effects at high magnetic field strength. These effects are the dipolar relaxation and chemical shift anisotropic relaxation. Some TROSY variants on the complete set of standard 3D protein NMR pulse sequences have been introduced.[74,75] It has been suggested[73] that these approaches will be applicable to proteins up to 500 kDa. In practice the benefits depend on the system under study, and in some situations TROSY versions of 3D experiments are preferable for large proteins even at low field strengths (500–600 MHz ^1H), and in some cases the significant loss of sensitivity with TROSY is too high a price to pay.

3 BIOMOLECULAR STRUCTURE DETERMINATION

The application of NMR spectroscopy both in solution and in the solid state can generate a variety of distance and angle restraints that can be used in computing the 3D structure of a biomolecule. These restraints can only be interpreted once the spectral assignments have been made, and may be obtained in a number of ways. The accuracy of distance and angle measurements depends largely on the technique, with those methods that yield the lowest number of restraints tending to give more accurate data, although in the case of solid-state NMR data this usually defines the local rather than the global structure. However, accuracy is not always required in order to determine a structure with high precision. Protein structures can be obtained with high precision with relatively poorly defined distance restraints, such as those obtained from NOESY experiments. However, the same does not hold for DNA, for which accurate distances are required. Furthermore, the use of NOEs is generally confined to those molecules whose molecular weight is above 50 000 (or whose overall rotational correlation time τ_c is long). For larger molecules alternative strategies have to be employed, which include solid-state NMR techniques, residual dipolar coupling measurements, and TROSY approaches (sections 2.3.3 and 4.2).

3.1 Assigning Nuclear Overhauser Effect Intensities

Nuclear Overhauser enhancements can be used to obtain distances between protons up to about 5 Å apart. In general, NOESY spectra are obtained at a variety of mixing times, τ_m. This enables the region where the NOE build-up rates are linear to be determined. Note that at long mixing times or for large biomolecules ($M_r > 50\,000$), spin diffusion takes place, in which cross-relaxation is propagated across the molecule from one nucleus to another. This can be avoided by ensuring that the initial rate condition is fulfilled. The cross-peak intensities in NOESY spectra can be measured to an accuracy of about 10%. As the NOE build-up rate depends on r^{-6}, where r is the distance between the two protons, it might be hoped that distances can be obtained very accurately. However, local internal motion, together with other contributory forms of relaxation or exchange, are the reasons why high accuracy is not obtainable in practice.

Thus, in general NOESY cross-peak intensities are classified into four different categories, with distances of 1.8–2.7 Å, 1.8–3.3 Å, 1.8–5.0 Å, and 1.8–6.0 Å for strong, medium, weak, and very weak NOEs. Although this classification is somewhat arbitrary, it works well in practice. Usually the NOE cross-peak intensities are calibrated against some internal standard fixed length, and the choice of such a standard is important because local motion can lead to erroneous results. For example, calibration using the δ and ε protons of an aromatic ring is quite common, and yet it is not uncommon for aromatic side chains (at least in proteins) to give NOEs which vary too much to be reliable. Care has to be taken to ensure that there is relative molecular motion in the functional group being used for distance calibration. Strong NOEs usually appear as intense cross-peaks in NOESY spectra even at short mixing times (e.g. $\tau_m = 50$ ms for a protein whose M_r is below 20 000). In the same spectra, medium NOEs appear as weak cross-peaks. The weak NOEs appear only at longer mixing times (e.g. $\tau_m = 200$ ms for a protein whose M_r is below 20 000) and may be due to indirect effects. The lower bound is kept the same for each category because local motion could severely attenuate the NOE, even for short distances. These distance bounds or restraints are used as input into computer algorithms for the calculation of the 3D structure (sections 3.3 and 3.4). Even though only ranges of distances for a given pair of atoms can be found, these restrict drastically the number of possible structures, and refinement protocols (section 3.6) ensure that a low-energy structure can be determined.

3.2 Use of *J* Couplings to Obtain Torsion Angles

Judicious use of the Karplus relation ($J = A + B\cos\theta + C\cos^2\theta$) yields information on the dihedral angles.[76,77] The constants in the Karplus relationship have been determined empirically. In proteins, the couplings of interest are the three-bond couplings; in particular $^3J_{H\alpha,NH}$ is related to the angle ϕ, and $^3J_{H\alpha,H\beta}$ and $^3J_{N,H\beta}$ are related to χ_1. These relationships have been calibrated on known structures[78,79] and are given in Equation (12) (in Hertz):

$$\begin{aligned}
^3J_{H\alpha,HN}(\phi) &= 6.4\cos^2(\phi - 60^\circ) \\
&\quad - 1.4\cos(\phi - 60^\circ) + 1.9 \\
^3J_{H\alpha,H\beta2}(\chi_1) &= 9.5\cos^2(\chi_1 - 120^\circ) \\
&\quad - 1.6\cos(\chi_1 - 120^\circ) + 1.8 \\
^3J_{H\alpha,H\beta3}(\chi_1) &= 9.5\cos^2\chi_1 - 1.6\cos\chi_1 + 1.8 \\
^3J_{N,H\beta2}(\chi_1) &= -4.5\cos^2(\chi_1 + 120^\circ) \\
&\quad + 1.2\cos(\chi_1 + 120^\circ) + 0.1 \\
^3J_{N,H\beta3}(\chi_1) &= -4.5\cos^2(\chi_1 - 120^\circ) \\
&\quad + 1.2\cos(\chi_1 - 120^\circ) + 0.1
\end{aligned} \tag{12}$$

Using these equations, the dihedral angles ϕ and χ_1 may be calculated from both homonuclear and heteronuclear couplings.[80] There are corresponding important dihedral

angles in other types of biomolecules, such as in DNA and carbohydrates.

Accurate measurement of J couplings in an overlapping one-dimensional spectrum of a biomolecule can be very difficult. Two-bond and three-bond coupling constants may be extracted with care from the antiphase multiplet structure of the cross-peaks in phase-sensitive COSY or DQFCOSY. A better approach is to use one of a number of modified COSY experiments which simplify the cross-peak multiplet structure, such as ECOSY[81] (exclusive correlation spectroscopy), or PECOSY[82] (primitive exclusive correlation spectroscopy). The disadvantage of the ECOSY experiment is that it has very low sensitivity. The PECOSY experiment is at least four times more sensitive than the ECOSY experiment.

For larger biomolecules, measurement of J couplings from the PECOSY spectrum is difficult because of extensive resonance overlap. In these cases, an estimate of the values for J couplings may be obtained with ^{15}N-labeled material, from the cross-peaks in a 3D ^{1}H–^{15}N TOCSY/HMQC spectrum. The extent of magnetization transfer during the Hartmann–Hahn spin-lock depends on the magnitude of the scalar J coupling.

The measured coupling constant represents the time average (on the 100-ms timescale) of the actual J coupling. Thus the interpretation of a measured J coupling can be hampered by uncertainty as to whether it represents a J value that corresponds to a single dihedral angle, or whether it represents a rapid averaging of J values from quite different dihedral angles. When a biological molecule does not exist as a single conformation, but as an equilibrium of rapidly interconverting conformers, the NOE and J contain time-averaged distance and angle information. Both averaging processes are strongly nonlinear in their dependence on geometry, and the presence of multiple conformations can lead to NOEs and J couplings that cannot simultaneously be consistent with a single conformation.

The stereospecific assignment of resonances in a biomolecule, for example the β-methylene protons or methyl resonances of leucine or valine, can be extremely important in defining local conformations. Early NMR structures were obtained by treating NOEs involving chiral protons with indistinguishable chemical shifts ($<$0.01 ppm) as a single distance to the geometric center between the two methylene protons, with a 1 Å correction added to the upper bound of the measured distance to correct for the difference in position of the so-called pseudo-atom relative to the true proton. Similarly the distance to the geometric center of a methyl group was used with a corresponding increase of 2.4 Å to the upper bound for the measured distance. Instead of using pseudo-atoms, a more rigorous approach is to match the experimental NOEs and J coupling constants with those calculated for conformations present in a database. Both possible stereospecific assignments are considered, and if the database contains only conformations consistent with one of the two possibilities, then the correct assignment, together with the allowed ranges for ϕ, ψ, and χ_1, can be determined. In practice only about 50% of the possible stereospecific assignments can be made in an unambiguous manner.

The $^{3}J_{\mathrm{HN}\alpha}$ can be used in crude calculations to provide ϕ backbone torsion-angle restraints, which for proteins can be classified into two ranges: $^{3}J_{\mathrm{HN}\alpha} < 6$ Hz corresponding to $-35^\circ < \phi < -85^\circ$; and $^{3}J_{\mathrm{HN}\alpha} > 8$ Hz corresponding to $-80^\circ < \phi < -175^\circ$. Refinement of these numbers using the methods already outlined, including more accurate ϕ, ψ, and χ_1 torsion angles and stereospecific assignments, can almost double the number of experimental restraints that are used as input to distance geometry and related algorithms. As a general rule of thumb, helices exhibit $^{3}J_{\mathrm{NH}\alpha} < 6$ Hz, β-sheets exhibit $^{3}J_{\mathrm{NH}\alpha} > 8$ Hz, and random extended chains exhibits $^{3}J_{\mathrm{NH}\alpha} = 6$–8 Hz. Note, however, that deviations from this rule occur when proline is present.

3.3 Distance Geometry

Distance geometry[83–85] involves taking the restraints together with the covalent structure, and creating a matrix of interatomic distances between all atoms so as to be consistent with the input. This approach is called the metric matrix distance geometry and the set of distances from n-dimensional space are projected into 3D Cartesian coordinate space. This method operates in distance space, whereas there are also real space methods which operate in either Cartesian coordinate space directly, or in torsion-angle space.[86] The algorithm involves no initial priming with any kind of initial approximate structure. Indeed most versions start from a set of randomly generated structures with the restraints embedded. It is therefore a very unbiased way of determining structure. Mathematically, the metric matrix distance geometry approach calculates a selection of trial distances to replace the upper and lower bounds in the distance matrix **D**. The trial distances d_{ij} are usually selected randomly although they must still satisfy the bounds (Equation 13):

$$l_{ij} \leq d_{ij} \leq u_{ij} \tag{13}$$

Next the distance is calculated from each point i to the center of mass, denoted as point 0, using Lagrange's theorem (Equation 14):

$$d_{i0}^2 = \frac{1}{N}\sum_{j=1}^{N} d_{ij}^2 - \frac{1}{N^2}\sum_{j=2}^{N}\sum_{k=1}^{j-1} d_{jk}^2 \tag{14}$$

The metric matrix g_{ij} is defined as the scalar product of the two vectors, $\vec{u}_i$ and $\vec{u}_j$, and can be easily calculated from the distances using the law of cosines (Equation 15):

$$g_{ij} \equiv \vec{u}_i \cdot \vec{u}_j = \tfrac{1}{2}(d_{i0}^2 + d_{j0}^2 - d_{ij}^2) \tag{15}$$

Calculating the metric with respect to the center of mass means that no coordinates are emphasized, and all the coordinates are comparable in magnitude. Repeating for all atom pairs yields the metric matrix **G**. The metric matrix can also be thought of as a matrix containing the projections of all atom pairs. The eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots$ are determined and ranked according to magnitude. The three largest eigenvalues are selected and are by nature mutually orthogonal. The corresponding eigenvectors of the real symmetric metric matrix **G** compose an N-column matrix **W**, where N is called the embedding dimensionality and is usually three. The largest eigenvalues are chosen as the principal axes so as to minimize scatter into higher dimensions. Thus the embedded coordinates, v_{ij}, for point i in N dimensions, where $j = 1, 2, \ldots, N$ are given by Equation (16):

$$v_{ij} = \lambda_N^{\frac{1}{2}} w_{ij} \tag{16}$$

One difficulty with metric matrix distance geometry is that distances alone cannot define the chirality of the structure, so that mirror images (local or global) of the correct structure can occur. In general, these can easily be rejected at the substructure stage as the chirality of single amino acids (L) and helices (right-handed) is known. A further problem associated with the method is that it does not sample efficiently the conformational space consistent with the stereochemical and experimental restraints. As a result, the atomic root mean square (RMS) distribution of a series of calculated structures tends to be underestimated, particularly in regions that are poorly defined by the experimental data. Subjecting these structures to restrained molecular dynamics (RMD), for example, results in an increase in the atomic RMS distribution, while improving the agreement with the experimental data and reducing the van der Waals energy.

In methods involving minimization in torsion angle space, bond lengths and angles are kept fixed during the minimization and only the torsion angles are varied. To ensure that correct folding occurs, the restraints have to be introduced gradually during the calculation. This is achieved in a variety of ways, so that the distances between residues further and further apart in the sequence are incorporated in successive cycles of minimization. In general, the structures found by these methods tend to have inordinately high energies and have to be subjected to extensive energy minimization with only minimal changes in atomic RMS deviations.

3.4 Molecular Dynamics and Simulated Annealing

When restrained energy minimization methods are used, inevitable local energy minima are encountered which can lead to inaccurate structures. To circumvent this, RMD are usually employed. This involves including NMR restraints in one of the many molecular dynamics simulation programs. Molecular dynamics[87,88] solves Newton's equations of motion (Equation 17),

$$F_i = m_i a_i \tag{17}$$

where F_i is the force, m_i is the mass, and a_i is the acceleration of atom i. The force on atom i can be computed directly from the derivative of the potential energy V with respect to the coordinates r_i. The energy can be expressed in the explicitly differentiable form of Equation (18):

$$-\frac{\mathrm{d}V}{\mathrm{d}r_i} = m_i \frac{\mathrm{d}^2 r_i}{\mathrm{d}t_i^2} \tag{18}$$

Therefore, with an adequate expression for the potential energy and known masses, this differential equation can be solved for future positions in time, t_i. In general, this can be solved only approximately, as V is usually a complex function of the coordinates of all (or many) of the atoms (i.e. $V = V(r_1, r_2, r_3 \ldots r_N)$). The temperature can be calculated from the atomic velocities (Equation 19),

$$\frac{3N}{2} k_{\mathrm{B}} T = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \tag{19}$$

where k_{B} is Boltzmann's constant, m_i and v_i are the mass and velocity of atom i, and N is the number of atoms (and $3N$ is the number of degrees of freedom). For a simulation at constant energy, the temperature fluctuates due to the interconversion of kinetic and potential energy. If the temperature is held constant, then the atomic velocities can be adjusted accordingly. If the pressure is held constant, the volume must be allowed to fluctuate by rescaling the interatomic distances.The total potential energy, V_{total}, is usually defined (Equation 20) as the sum of a number of terms:

$$V_{\mathrm{total}} = V_{\mathrm{bond}} + V_{\mathrm{angle}} + V_{\mathrm{dihedr}} + V_{vdW} + V_{\mathrm{coulomb}} + V_{\mathrm{NMR}} \tag{20}$$

where V_{bond}, V_{angle}, and V_{dihedr} keep bond lengths, angles and dihedral angles at their equilibrium values. (The first five terms are empirical energy terms describing the physical interactions between the atoms, whereas the last term is a means of including the NMR information, but does not correspond to any real physical force.) They are

summarized in Equations (21–23):

$$V_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2} K_b (b - b_0)^2 \tag{21}$$

$$V_{\text{angle}} = \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \tag{22}$$

$$V_{\text{dihedr}} = \sum_{\text{dihedr}} K_\phi [1 + \cos(n\phi - \delta)] \tag{23}$$

These are pseudoharmonic potentials that constrain bond lengths (b), bond angles (θ), and the rotamer angles (ϕ, δ) for staggered and eclipsed conformations, and K is a constant. The van der Waals and electrostatic interactions are described by V_{vdW} and V_{coulomb},

$$V_{vdW} = \sum_{\text{pairs } (ij)} \frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^6} \tag{24}$$

$$V_{\text{coulomb}} = \sum_{\text{pairs } (ij)} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \tag{25}$$

where Equation (24) is the Lennard–Jones potential, containing repulsive and attractive terms (C is a constant), and Equation (25) describes the coulombic interactions between two charged particles (i, j) with partial charges q that are a distance r_{ij} apart in a dielectric medium described by $\varepsilon_0 \varepsilon_r$. The term V_{NMR} contains the NMR restraints, and has the effect of pulling the protons that show an NOE interaction closer to the measured distance r_{ij}. Similarly, V_{NMR} may also contain J-coupling information by including a torsion term. These potentials are also pseudo harmonic functions of similar form to Equations (21–23). Distance constraints which can be reasonably accurately determined may therefore be defined as Equation (26):

$$V_{\text{NOE}} = \begin{array}{ll} K_1 (r_{ij} - r_{ij}^0)^2 & \text{if} \quad r_{ij} > r_{ij}^0 \\ K_2 (r_{ij} - r_{ij}^0)^2 & \text{if} \quad r_{ij} < r_{ij}^0 \end{array} \tag{26}$$

where r_{ij} and r_{ij}^0 are the calculated and experimental interproton distances respectively, and K_1 and K_2 are force constants given by Equation (27):

$$K_1 = \frac{k_{\text{B}} TS}{[2(\Delta_{ij}^+)^2]} \quad \text{and} \quad K_2 = \frac{k_{\text{B}} TS}{[2(\Delta_{ij}^-)^2]} \tag{27}$$

where k_{B} is Boltzmann's constant, T is the absolute temperature of the simulation (*not* the experiment), S is a scale factor, and Δ_{ij}^+ and Δ_{ij}^- are the positive and negative error estimates, respectively, of r_{ij}. If, however, only ranges of distances can be specified, then the distance restraints are incorporated into a pseudo square-well potential (Equation 28):

$$V_{\text{NOE}} = \begin{array}{ll} K_{\text{NOE}} (r_{ij} - r_{ij}^u)^2 & \text{if} \quad r_{ij} > r_{ij}^u \\ 0 & \text{if} \quad r_{ij}^l \le r_{ij} \le r_{ij}^u \\ K_{\text{NOE}} (r_{ij} - r_{ij}^l)^2 & \text{if} \quad r_{ij} < r_{ij}^l \end{array} \tag{28}$$

where r_{ij}^u and r_{ij}^l are the upper and lower limits, respectively, of the target distances obtained from experiment, and K_{NOE} is the NOE force constant, which is typically chosen to be of the order of $1000\,\text{kJ mol}^{-1}\text{nm}^{-2}$. Similarly, torsion angle restraints can be incorporated into pseudo square-well potentials (Equation 29):

$$V_{\text{tor}} = \begin{array}{ll} K_{\text{tor}} (\phi_i - \phi_i^u)^2 & \text{if} \quad \phi_i > \phi_i^u \\ 0 & \text{if} \quad \phi_i^l \le \phi_i \le \phi_i^u \\ K_{\text{tor}} (\phi_i - \phi_i^l)^2 & \text{if} \quad \phi_i < \phi_i^l \end{array} \tag{29}$$

where ϕ_i^u and ϕ_i^l are the upper and lower limits of the target range of a particular torsion angle obtained from experiment, ϕ_i is its calculated value, and K_{tor} is the torsion force constant, which is typically chosen to be of the order of $4000\,\text{kJ mol}^{-1}\text{rad}^{-2}$. To ensure that the experimental restraints are the dominating factor in determining the conformation of the molecule, it is very important that the force constants for the restraints are set sufficiently high that the experimental data are satisfied within the precision of the measurements. At the same time, the contribution from the empirical energy function should be such that for any individual RMD structure the deviations from ideal geometry are small, and the nonbonded interactions are good (i.e. the Lennard–Jones potential is negative). This can be determined from a few trial calculations. Thus, convergence on the structure is guided by the requirement to minimize NOE or other restraint violations. The number of distance restraint violations N_{viol} is counted when, for example, $r_{ij} \ge r_{ij}^0 + 1$, which would allow for 1 Å fluctuations. Another parameter that can be minimized in addition to N_{viol} is the sum of the distances in excess of the constraints Δr_{viol}, as given by Equation (30):

$$\sum \Delta r_{\text{viol}} = \sum_{k=1}^{N_{\text{viol}}} (r_{ij})_k - [(r_{ij}^0)_k + 1] \tag{30}$$

where the sum runs over all those interproton (or pseudo-atom) distances for which N_{viol} is defined. Although an arbitrary structure may be used for RMD calculation, in practice a starting structure obtained from distance geometry and energy minimization is often used. Because of the kinetic energy present in the protein during the dynamics simulation, the problem of local minima can be overcome relatively easily.

The RMD approach requires a relatively large amount of computation time compared to distance geometry

methods. This problem can be overcome by using a simplified potential energy function, where all nonbonded contact interactions are described by a single van der Waals repulsion term. Also, by using a cut-off distance, in which nonbonded interactions for pairs of atoms that are separated by a distance greater than some reasonable value (e.g. 5–10 Å), the number of nonbonded interactions is decreased significantly.

Simulated annealing[89] involves raising the temperature of the system followed by slow cooling in order to overcome local minima and locate the global minimum region of the target function. It is computationally more efficient than RMD and yields structures of similar quality. The potentials are very similar to RMD and again Newton's laws of motion are solved as a function of time. However, in implementations found in commercial programs, the nonbonded interaction potential is modified so that there is a simple van der Waals repulsion term with a variable force constant K_{rep} (Equation 31):

$$V_{repel} = \begin{cases} 0 & \text{if} \quad r \geq sr_{min} \\ K_{rep}(s^2 r_{min}^2 - r^2)^2 & \text{if} \quad r < sr_{min} \end{cases} \qquad (31)$$

The values of r_{min} are given by the sum of the standard values of the van der Waals radii between two atoms as represented by the Lennard–Jones potential used in, for example, the empirical energy function CHARMM (Chemistry at Harvard Molecular Mechanics). A van de Waals scale factor, s, typically should be set to 0.8 to account for the fact that interatomic separations slightly smaller than the sum of the hard-sphere van der Waals radii can easily occur due to the attractive component of the van der Waals interaction. Further potentials and force constants can be introduced for refinement against chemical shift and torsion angle databases, and for inclusion of residual dipolar couplings.[90]

3.5 Energy Minimization

The next step in biomolecular structure determination is the refinement of the distance geometry-derived Cartesian coordinates. Energy minimization is generally used to find local minima in the potential energy for the structure. Commonly employed energy minimization algorithms[91,92] offer choices of one of the following methods: steepest descent, conjugate gradient (also known as Powell[93]), and Newton–Raphson. All these assume that the energy surface is approximately harmonic, and physically adjust the atomic positions in Cartesian space over the energy surface, deeper into the local potential well. Although the rate of descent for the steepest descent method is initially fast, it is nonconvergent and often results in large oscillations about the minimum. This may lead to structures that have climbed out of the local well to a higher energy point. The conjugate gradient method approaches the minimum more slowly, and results in a converged minimum point energy. However, for very poor starting structures, it is more likely to converge.

In general, for biomolecules, 100 or so steps of steepest descent minimization are followed by 500–1000 steps of either conjugate gradient or modified Newton–Raphson methods. These energy minimization steps can be carried out at a variety of the steps involved in generating NMR structures of biomolecules. Restraints can be included in some of the implementations of these algorithms. In fact, too many minimization steps without restraints (and without including solvent molecules) may lead to structures that are far from either X-ray crystal or NMR structures. Therefore, although these methods are used to generate theoretical model structures, without restraints, their indiscriminate application to biological problems is likely to lead to meaningless structures that are not at all helpful in understanding their function.

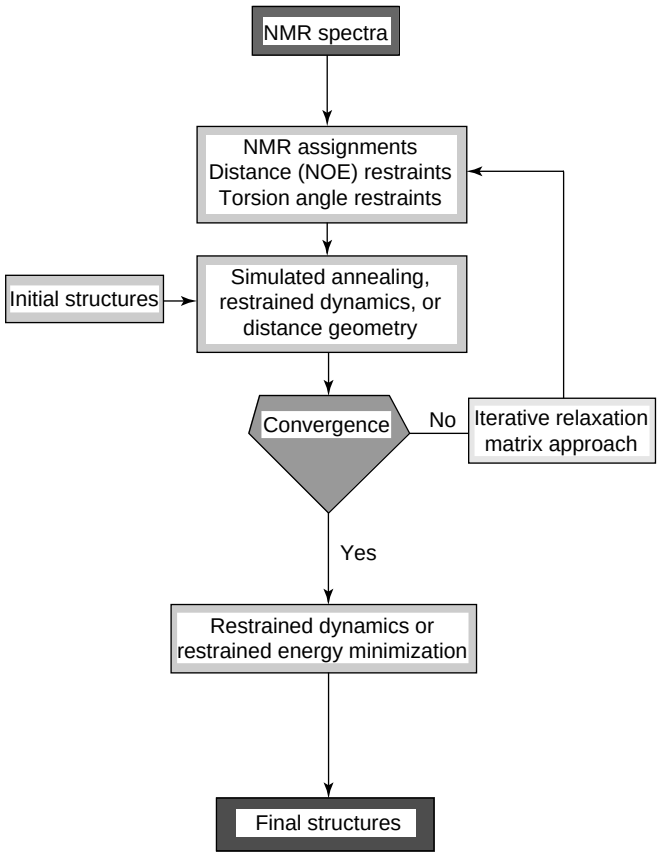

Figure 6 The strategy for solving the 3D structure of biomolecules on the basis of NMR data. (Adapted from ref. 8 with permission.)

3.6 Procedure for Structure Determination

The typical procedure in obtaining an NMR structure is as follows. The NMR spectra are assigned using 2D, 3D, and 4D NMR experiments, and from these a list of distance restraints derived from NOEs and/or dipolar couplings are generated, together with torsion angle restraints derived from J couplings and/or dipolar couplings. This list is used to determine a set of NMR force potentials that, together with a seed of randomly generated initial extended chain structures, form the input files for distance geometry and/or RMD/simulated annealing calculations (Figure 6).

When the structures converge on a well-defined structure (the next section provides a definition of ''well-defined''), then either the whole family of structures or a calculated average structure can be subjected to further restrained dynamics or energy minimization, to generate the final structures. Generally it is preferable to keep the whole family of structures, and submit them to the Protein Data Bank as the final structure. Such families inform users of the structural data as to the areas of the structure that are not well-defined. However, if the structures do not converge, then a sometimes lengthy iteration of going back to the original data to check assignments (both in terms of chemical shifts and NOE intensities) and recalculation of structures ensues. This can be supplemented with use of iterative relaxation matrix approaches to back-calculate the NOESY spectra, and comparison with the experimental data can lead to a semi-automated way of recalculating the structures. This method seems to work for relatively small molecules, but generally fails in biomolecules whose M_r exceeds 15 kDa. Refinement of the structure continues until some criteria of accuracy are met.

3.7 Assessing the Accuracy and Precision of the Final Structure

As already discussed, the typical procedure for obtaining NMR structures involves using a large number of different starting structures for the distance geometry and/or RMD/simulated annealing. If a significant fraction of the calculated structures converge, satisfying all the NMR restraints, and show a root-mean-square deviation (RMSD) from each other of less than 2 Å, then this indicates that the calculated structures must be close to the actual structure. The RMSD is generated by superimposing the centroids of the structures, and calculating a rotation that minimizes the overall value (Equation 32):

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_i')^2} \qquad (32)$$

where N is the number of atoms being compared, and r_i and r_i' are the atomic coordinates for the standard structure and the rotated structure, respectively. For comparisons of families of NMR structures, each of the structures is rotated so as to provide the best fit with a predetermined member of the family. The RMSD values are reported for all pairwise structure comparisons. If serious violations of the NMR restraints remain, or if the RMSD between the various structures is too large, then reanalysis of the NMR spectra is necessary. A further check can be made using the iterative relaxation matrix approach and back-calculating the NOESY spectrum from the structure. The precision of the measurements, for example how narrowly the bounds for NOE-derived distances are defined, will improve the RMSD between the set of calculated structures and therefore the accuracy of the structure. Those NMR protein structures that show a relatively low level of detail typically have RMSDs of at least 1.5–2 Å. Higher-resolution structures show RMSDs of the order of 0.8 Å for the backbone atoms, 1 Å for backbone atoms plus some interior side-chain atoms, and around 1.5 Å for all the atoms in the molecule. Such improvements in resolution can be attributed in part to the availability of stereospecific assignments, which have the effect of narrowing the distance bounds. Relatively few restraints are sufficient to define the global fold in proteins, whereas the same is not true for nucleic acids, where there are no possible long-range restraints. Here iterative relaxation matrix approaches to refinement of the precision of the distance measurements from NOESY cross-peak volumes is very important.

Another method that is beginning to be used in assessing the accuracy of a structure is the R factor, or structure factor, which, by analogy with the term in X-ray crystallography, is the normalized mean deviation between the structure factors derived from the model and the experimental data. For NMR data, this usually involves[94] a direct measure of the fit between the experimental and theoretical NOE intensities when the final model is subjected to back-calculation of the NOESY spectrum. In its simplest, most general form the NMR R factor is defined as in Equation (33):

$$R = \frac{\sum_{i,j} W_{ij}(\tau_m)\ A_{ij}^{\text{calc}}(\tau_m) - A_{ij}^{\text{exp}}(\tau_m)}{\sum_{i,j} W_{ij}(\tau_m) A_{ij}^{\text{exp}}(\tau_m)} \qquad (33)$$

where $A_{ij}^{\text{calc}}(\tau_m)$ and $A_{ij}^{\text{exp}}(\tau_m)$ are the elements connecting protons i and j of the theoretical and experimental NOE intensity matrices for a given mixing time τ_m. Weight factors $W_{ij}(\tau_m)$ are included to account for measurement errors. A number of other definitions of the R factor have appeared, and some of the available algorithms calculate

several different types of R factor simultaneously. The convergence limit on as low an R factor as achievable in an iterative relaxation matrix structure calculation is usually used as the stopping point for the refinement. However, as for X-ray crystallography, the R factor only measures the agreement of the calculated structure with the experimental data, so that a low value does not necessarily mean an accurate structure if the precision of the measurements is poor. One type of R factor recently introduced[95] that attempts to circumvent this problem relies on cross-validation, which is a statistical method that estimates the quality of the fit to the observed data without making any assumptions about the distribution of errors in the NMR data.

Despite best efforts at obtaining precise distance measurements for use in calculating accurate structures, there are often still regions of a molecular structure for which NMR restraints are absent. These regions are therefore disordered during all the refinement procedures, and it is tempting to suggest that this region has considerable conformational flexibility. In other words, a common argument is that absence of NMR restraints means an absence of structure. But the negative observation of a less accurately determined structure in a certain region (resulting from the lack of NOEs for example) is not sufficient to establish that a structure is disordered and conformationally labile. This can be addressed by measuring backbone dynamics using ^{15}N-labeled protein and measuring heteronuclear ^{15}N T_1, $T_{1\rho}$, and T_2 relaxation rates and ^{1}H–^{15}N NOEs, so that the order parameter (S^2) can be calculated.[96–99] This is a more rigorous way of determining whether molecular motion is the real reason for the absence of NOEs. Similarly, side-chain dynamics can be probed using measurements of ^{13}C T_1, $T_{1\rho}$, and T_2 relaxation rates and ^{1}H–^{13}C NOEs and ^{2}H T_1 and $T_{1\rho}$ relaxation rates.[100,101]

4 OBTAINING LONG-RANGE DISTANCE AND ANGLE INFORMATION

The need for a long-range distance and angle information is clear when liquid-state NMR is only capable of furnishing distances of the order of 5 Å, and angle information for spins involved in direct scalar J-coupling interactions. Here a number of approaches are emerging that have the potential to address this significant shortcoming of NMR as a biomolecular structural technique.

4.1 Use of Shift Tensors to Obtain Torsion Angles

The origin of the chemical shift is that the moving electric charges of the electron cloud around a nucleus induces a local magnetic field which opposes the applied field. Thus the effective field at the nucleus is (Equation 34):

$$\mathbf{B}_{\text{eff}} = \mathbf{B}_0(1 - \sigma_{\text{iso}}) \tag{34}$$

The nucleus is said to be *shielded*, and the extent of the shielding is given by the shielding constant, also called the isotropic chemical shift tensor, σ_{iso}. This is directly related to the electron density ρ at a distance r from the nucleus by Lamb's equation (Equation 35):

$$\sigma_{\text{iso}} = \frac{4\pi e^2}{3mc^2} \int_{-\infty}^{\infty} r\rho(r)\,\mathrm{d}r \tag{35}$$

There are three principal components of the shift tensor, σ_{11}, σ_{22} and σ_{33}, and the isotropic shift tensor σ_{iso} is given by Equation (36):

$$\sigma_{\text{iso}} = \tfrac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \tag{36}$$

The shift tensor σ_{iso} is related to the Larmor frequency w_0 by Equation (37):

$$w_0 = \frac{\gamma}{2\pi}\mathbf{B}_0(1 - \sigma_{\text{iso}}) \tag{37}$$

and to the chemical shift by Equation (38):

$$\delta = 10^6(\sigma_{\text{ref}} - \sigma_{\text{sample}}) \tag{38}$$

By appropriate choice of a coordinate system, the isotropic chemical shift σ_{iso}, which is a 3×3 matrix, or second rank tensor, may be converted to three principal elements, σ_{11}, σ_{22} and σ_{33}, which serve to characterize the 3D nature of the shielding. From these elements it is possible to define the shielding anisotropy $\Delta\sigma$ and the shielding asymmetry parameter η (Equations 39 and 40):

$$\Delta\sigma = \sigma_{33} - \tfrac{1}{2}(\sigma_{11} + \sigma_{22}) \tag{39}$$

$$\eta = \frac{\sigma_{22} - \sigma_{11}}{\sigma_{33} - \sigma_{\text{iso}}} \tag{40}$$

The chemical shielding tensor ellipsoid is illustrated in Figure 7, together with the relationship between the principal axis system and the laboratory frame (the frame in which $\mathbf{B}_0$ resides). The principal elements of the shift tensor can be obtained by rotational side-band intensity analysis.[102] The parameter $\Delta\sigma$ provides an estimate of the departure from spherical symmetry of the electrons around the nucleus.

In solution, the isotropic molecular motion averages the shift anisotropy (see Equation 36), except in cases where there is partial ordering of the sample. In the solid state, the chemical shift anisotropy will be manifested according to the state of the sample:

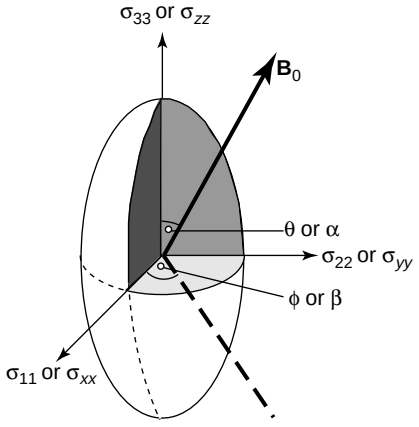

Figure 7 The chemical shielding ellipsoid, which is used to indicate that different orientations of the magnetic field relative to the molecular framework result in different resonance positions for the same chemical species. (Adapted from ref. 8 with permission.)

1. In a single crystal in a fixed orientation to the magnetic field, a single sharp line will be observed for each magnetically unique orientation of a particular nucleus with respect to the field direction, and the positions of these lines will change as the orientation of the crystal is changed.
2. For a powdered sample, each random crystallite orientation will give rise to a line, and the sum of these generate a broad line, the shape of which will depend on the principal elements of the shielding tensor (Figure 7).[103]

When two of the elements are identical, the shielding pattern is said to be axially symmetric. Such a situation can arise either by chance, or from the presence of molecular motion which partially averages the shift anisotropy in one direction.

When the principal elements of the shift tensor have been determined for a given pair of nuclei, their mutual orientation can be determined relative to the laboratory frame and to the molecular frame, thereby enabling torsion angles to be determined. This can be achieved by relating the tensor orientations of the two nuclei back to the molecular axis, or by fixing one tensor relative to the molecular axis and determining other unknown tensors by measuring some angle-dependent property, such as the dipolar coupling (sections 4.2, 4.3.2, and 4.3.4).

4.2 Use of Residual Dipolar Couplings to Obtain Distance and Angle Restraints

Distance and angle information can be obtained from NMR through the determination of dipolar couplings. The dipole–dipole interaction is averaged to zero in an isotropically tumbling liquid. However, for a biological macromolecule, tumbling can be anisotropic under conditions of high magnetic fields or in the presence of a liquid crystalline phase.[104–108] These methods partially orient the liquid-state sample through the use of lipid bicelles,[109] filamentous phage,[110] viruses,[111] or purple membranes.[112] Under these circumstances, residual dipolar couplings can be measured, which can give restraints relative to a molecular frame, enabling domains to be oriented relative to one another, and leading to more accurate distance measurements. They can be measured using pulse sequences which measure the contribution to J couplings that are due to residual dipolar couplings, such as the in-phase/antiphase (IPAP) HSQC and HNCO pulse sequences.[105,107] Such residual dipolar couplings can then be incorporated into simulated annealing protocols for structure refinement,[108,113] using Equation (41) for dipolar couplings between two nuclei, P and Q, relative to the molecular alignment tensor:

$$D_{\mathrm{PQ}}(\theta,\phi) = \mathbf{S}\frac{\mu_0}{4\pi}\frac{\gamma_{\mathrm{P}}\gamma_{\mathrm{Q}}h}{4\pi^2 r_{\mathrm{PQ}}^3} \times \left\{A_{\mathrm{a}}(3\cos^2\theta - 1) + \frac{3}{2}A_{\mathrm{r}}\sin^2\theta\cos 2\phi\right\} \quad (41)$$

where **S** is the generalized order parameter for internal motion of the PQ vector, γ_{P} and γ_{Q} are the gyromagnetic ratios of P and Q, h is Planck's constant, r_{PQ} is the distance between P and Q, and θ and ϕ are the cylindrical coordinates describing the orientation of the PQ vector in the principal axis system of **A**. Values for the axial (A_{a}) and rhombic (A_{r}) components of **A** depend on the shape of the protein and vary with the bicelle concentration. These approaches are already having a significant impact on NMR structure determination,[90,114–117] and are likely to revolutionize liquid-state biomolecular structure determination in general. However, there still remains the need for accurate distance measurements in larger macromolecular assemblies, or in unoriented soluble proteins whose molecular weight exceeds the current liquid-state NMR limits, and here solid-state NMR has a lot to offer.

4.3 Solid-state Nuclear Magnetic Resonance in Biomolecular Structure Determination

As the dipolar coupling is not averaged to zero in the solid state, it can be a valuable source of distance information from NMR in addition to NOEs. Torsion angles can also be determined very accurately from the angular dependence of the dipolar coupling, and can be used to determine the structure of a small biomolecule. Distances and torsion angles calculated from the measured dipolar coupling constant in the solid

state can be significantly more accurate than distances or angles determined from NOEs and J couplings in solution. For example, distances can be measured with accuracies of the order of ± 0.1–0.5 Å, and angles of the order of $\pm 3°$. The main disadvantage is that currently only pairwise distances or angles between isolated spins can be measured, so that there is a serious paucity of constraints compared with solution-state methods. Therefore, these approaches involve generating a very limited set of accurate torsion angle or distance measurements which can be used to define parts of a molecular structure. Furthermore, in the case of the distance measurement methods, far longer distances (up to 18 Å) can be measured than for NOEs where the limit is about 6 Å. For instance, functionally crucial distances may be measured, such as between a substrate and an enzyme, or a protein and DNA, or an antibody and its antigen. These techniques have also been reviewed elsewhere at greater length.[8,118]

4.3.1 *Homonuclear Dipolar Recoupling to Obtain Distances*

Most modern solid-state NMR approaches reduce linewidths which are broadened by large proton–proton dipolar couplings by use of magic-angle spinning (MAS). This technique causes all the dipolar couplings (both homonuclear and heteronuclear) in the dipolar Hamiltonian to fluctuate sinusoidally during the rotor cycle, so that integrating the Hamiltonian over one rotor cycle gives zero. A number of methods have been introduced recently which selectively reintroduce the dipolar coupling (hence ''recoupling'') for strongly coupled homonuclear spins and use this to measure selected distances in biomolecules, particularly those with $M_r > 50\,000$. Almost all the techniques work by manipulating the dipolar Hamiltonian in such a way that it integrates to a nonzero value over the rotor cycle, scaled by a factor determined by the particular experiment. The first technique, called rotational resonance (or R^2), introduced by Griffin et al.[119,120] relies on the fact that magnetization transfer between two homonuclear spins is most efficient when the chemical shift difference equals an integral of the MAS frequency, which can place the spinning side-band of one resonance overlapping with the isotropic resonance of another.[121] This phenomenon is called the rotational resonance condition, and is given by Equation (42),

$$\Delta v = n w_r \tag{42}$$

where Δv is the chemical shift difference of the cross-relaxing resonances, n is an integer, and w_r is the spinning speed. Clearly the requirement for isotropic resonances to be well separated places some restrictions on the applicability of this method. Griffin et al. made use of this phenomenon using isotopic labels, although others have used it at natural abundance.[122,123] The pulse sequence (Figure 8) involves an exchange time for magnetization transfer to occur. If the two rotationally coupled spins are also close in space, and therefore dipolar recoupled, their resonance intensities will vary as a function of the mixing time due to a transfer of magnetization. When spinning at the $n = 1$ rotational resonance condition, the initial rate of magnetization transfer is dominated by the strength of the dipolar coupling. As the value of n increases, the rate of transfer decreases and contributions from the orientations of the chemical-shift tensors become more significant, making interpretation of the transfer rate more difficult. Therefore the $n = 1$ experiments are the most useful, although it is still necessary to know the isotropic shifts of the two resonances, the zero quantum transverse decay T_2^{ZQ}, the principal values of the shift tensors (which are known for many functional groups[124]) and, less important, the mutual orientation of the two tensors. The zero-quantum T_2^{ZQ} decay influences magnetization transfer, because the difference polarization is continuously exchanged with zero quantum coherence under the influence of dipolar coupling, and may be measured from Hahn spin–echo experiment. Theoretical simulations of strongly coupled sites also predict T_2^{ZQ} and oscillations in the transfer curves that result from the relative orientations of the exchanging sites.[125]

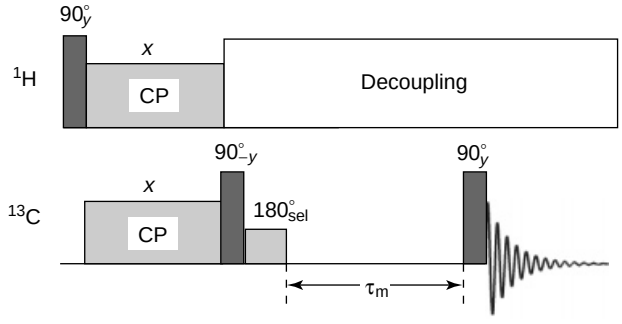

Figure 8 Pulse sequence for rotational resonance experiment. After cross-polarization (CP), the $90°_{-y}$ pulse returns the ^{13}C magnetization to the z axis (a flipback pulse). The weak $180°_{sel}$ inverts selectively one of the two resonances separated by Δv, and after a variable mixing times, τ_m, the spectrum is recorded. (Adapted from ref. 8 with permission.)

As the magnetization transfer is proportional to the dipolar coupling (D), measurement of D can be used to calculate internuclear distance r according to Equation (43):

$$D = \frac{\mu_0}{4\pi} \frac{\gamma_I \gamma_S \hbar}{r^3} \tag{43}$$

where γ_I and γ_S are the gyromagnetic ratios for the two spins involved (in the case of the R^2 experiment, this means two homonuclear spins, so $\gamma_I = \gamma_S$, although obviously this can apply to two heteronuclear spins), $\hbar$ is Planck's constant divided by 2π, and μ_0 is the permittivity of a vacuum. The experimental data fit with simulated curves to an accuracy of ± 0.4 Å for an isolated spin pair (e.g. $^{13}C-^{13}C$ or $^{31}P-^{31}P$), and the limits of distances which may be detected can be as high as 18 Å for $^{19}F-^{19}F$ pairs. These distances can be used as restraints in distance geometry and/or simulated annealing calculations.

A number of other homonuclear recoupling methods have been introduced, including RFDR (rotor-driven dipolar recoupling),[126] SEDRA (simple excitation for the dephasing of rotational amplitudes),[127] DRAMA (dipolar recovery at the magic angle)[128] and its variants,[129,130] MELODRAMA (melding of spin-locking and DRAMA),[131] DRAWS (dipolar recoupling with a windowless sequence),[132] CROWN (combined rotation with nutation),[133] HORROR (homonuclear rotatory resonance),[134] C7 (7-fold symmetric homonuclear recoupling),[135] and POST-C7 (compensated 7-fold symmetric homonuclear recoupling).[136]

4.3.2 Homonuclear Dipolar Recoupling to Obtain Angles

The rotational resonance experiment can also be used to measure torsion angles.[137] At higher orders of the rotational resonance condition, $n \geq 3$, the magnetization exchange rate is not only dependent on the internuclear distances but also very strongly dependent upon the relative orientation of the two chemical shift tensors (provided they are both large). McDermott et al.[138] have used this approach to measure dihedral angles between two adjacent nonprotonated carbons in phosphoglycolate bound to triosephosphate isomerase, to an accuracy of $\pm 10^\circ$. They made use of the fact that the intensity of the zero quantum side-bands depends systematically on the dihedral angle, which can be measured by determining the magnetization exchange rates at $n = 4$.

There are alternative approaches to R^2 for mutually orienting two homonuclear shift tensors, and thereby obtaining torsion angles. There are also new techniques emerging for orienting the chemical shift tensors of two dipolar coupled spins with respect to one another, using magnetization transfer,[139] including rotational resonance,[137,138] or multiple quantum coherence,[140,141] including C7.[142–144]

4.3.3 Heteronuclear Dipolar Recoupling to Obtain Distances

The principal heteronuclear technique introduced by Schaefer et al.[145–147] is called rotational echo double resonance (REDOR), and relies on the dephasing of magnetization of one spin in the presence and absence of dipolar coupling to another heteronucleus, and subsequent refocusing as a function of the MAS frequency. The variation of intensity for the one nucleus in the presence and absence of dephasing of the other nucleus is directly proportional to the dipolar coupling constant. The pulse sequence is outlined in Figure 9.

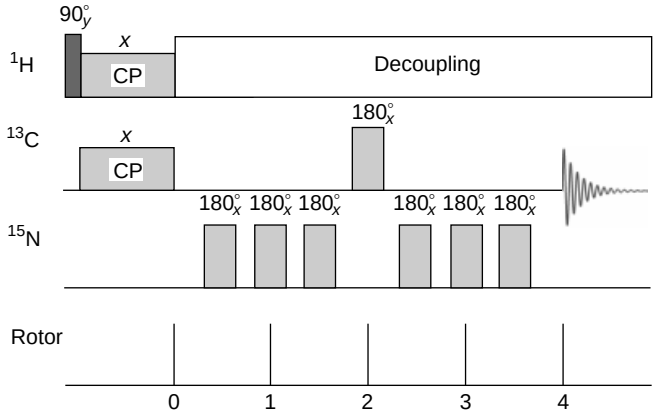

Figure 9 Pulse sequence for a version of REDOR ^{13}C-NMR. Two equally spaced 180° ^{15}N pulses per rotor period result in dephasing of transverse carbon magnetization produced by cross-polarization (CP) transfer from dipolar-coupled protons. The 180° ^{13}C pulse replaces the ^{15}N pulse in the middle of the dephasing period and refocuses isotropic ^{13}C chemical shift differences at the beginning of data acquisition. High power ^1H-decoupling is maintained throughout the ^{15}N dephasing sequence and the acquisition. (Adapted from ref. 8 with permission.)

The first spectrum is obtained using a standard cross-polarization pulse sequence with a π pulse on the observed nucleus (e.g. ^{13}C) in the middle of the evolution period. During this period, the observable magnetization evolves under the influence of the chemical shifts and the heteronuclear dipolar interaction. The 180° pulse refocuses both interactions, leading to a signal S during the acquisition period. The second spectrum is obtained with an additional train of π pulses on the dipolar-coupled spin (e.g. ^{15}N). These pulses affect the observed signal by preventing rotational refocusing of the dipolar interaction. The magnetization is therefore not completely refocused, and the signal intensity drops by an amount ΔS. For weak dipolar coupling, the change in signal intensity is related to the distance between the coupled spins by Equation (44)

$$\frac{\Delta S}{S} = KD^2 N_c^2 \nu_r^{-2} \tag{44}$$

where N_c is the number of rotor cycles during the evolution period, ν_r is the spinning speed, D is the

dipolar coupling, and K is a dimensionless constant. If performed at relatively slow spinning speeds and over several rotor cycles, the size of the difference signal can be increased. This experiment therefore yields the internuclear distance, which for an isolated spin pair (e.g. $^{13}C-^{15}N$ or $^{31}P-^{13}C$) can be determined to an accuracy of around ± 0.1 Å.

One advantage this experiment has over the current homonuclear methods is that a universal curve can be used for the calculation of distances. There is no dependence on the values or orientations of the chemical shift tensors, or on the zero-quantum T_2^{ZQ}. Therefore once a particular instrument has been calibrated with a known sample, such a curve can be used to read off dipolar couplings directly. Furthermore, the distances that can be measured can be quite considerable, for example <5 Å for $^{13}C-^{15}N$, <8 Å for $^{13}C-^{31}P$, and <12 Å for $^{13}C-^{19}F$. Also REDOR can be applied[148–153] to quadrupolar nuclei, for example ^{2}H and ^{14}N.

Other variants of REDOR include experiments involving three or more S spins, such as the transferred-echo double resonance (TEDOR) experiment,[154] which can be used to edit out the natural abundance background signals by selecting unique spin triplets. A third rare spin, such as ^{2}H, ^{19}F, or ^{31}P, is used to distinguish the signal arising from a specific $^{13}C-^{15}N$ pair from that of the natural abundance background, although they require specialized quadruple resonance instrumental capabilities. Another interesting development has been the application by Mueller et al.[155,156] of a Bessel function transform to deconvolute multiple time-dependent dephasing signals, and thereby measure multiple dipolar couplings simultaneously.[157]

The REDOR technique has been further developed, including the REDOR/TOSS (total side-band suppression) pulse sequence,[158,159] alternative phase-cycling schemes,[160–164] and more recently the REDOR/TPPM (two-pulse phase modulation) pulse sequence.[165] Finally, the approach called REDOR-editing[166] can be used to edit a spectrum with a large natural abundance background with respect to a dipolar coupled spin pair that are unique to a biological molecule.

4.3.4 Heteronuclear Dipolar Recoupling to Obtain Angles

This method was developed by Opella, Cross et al., and determines the torsion angles, for example in the polypeptide backbone of a protein, as a means to determine structure based solely on orientational restraints.[167–169] We will consider the theory of this technique as applied to proteins. Although it could be applied to other classes of biomolecules such as DNA, to date its use has been limited primarily to membrane proteins.

The α carbons of adjacent amino acids in a polypeptide chain are joined by the amide C–N bonds. The six atoms that form this peptide linkage lie roughly in a plane, and the relative orientation of adjacent planes is defined by the ϕ and ψ torsion angles (see Figure 10). The secondary structure of the peptide backbone can be determined by establishing the sequential orientation of each peptide plane relative to a common axis.

Measurements of both the dipolar and chemical shift interactions are necessary to limit the number of possible orientations and define the peptide structure.[167,169–171] The orientation dependence of the dipolar interaction is given by Equation (45):

$$\Delta v = D(3\cos^2\theta - 1) \tag{45}$$

where Δv is the observed dipolar splitting, D is the dipolar coupling constant, and θ is the angle between the internuclear vector connecting the spins and the external magnetic field. The dipolar interactions point along the bond axes and are axially symmetric. The observed dipolar splitting ranges from zero for bonds aligned at the magic angle, to $2D$ for bonds parallel to the z axis of the magnetic field. One of the problems in determining peptide plane orientations is that the experimental Δv specifies two or four different values for θ depending on the size of D relative to Δv. Consequently, several bond orientations need to be measured to arrive at a unique solution for the peptide-plane orientation. The $^{15}N-^{1}H$ and $^{15}N-^{13}C$ bond orientations are the most readily measured dipolar interactions because they involve spin-$\frac{1}{2}$ nuclei, although ^{14}N quadrupole and dipole splittings can also be measured and used as orientation restraints.[172,173] Additional restraints for peptide-plane orientations can be derived from the amide ^{15}N and carbonyl ^{13}C chemical-shift tensors. The orientation dependence of the chemical

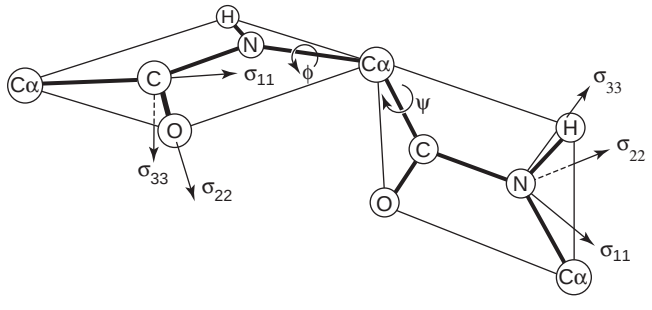

Figure 10 Illustration of the peptide planes of a dipeptide and the appropriate orientations of the amide ^{15}N and carbonyl ^{13}C chemical shift tensors with respect to the molecular frame. The ϕ and ψ angles define the orientation of the planar peptide linkages and are determined by measuring the N–C and N–H bond orientations, as well as the orientation of the ^{15}N and ^{13}C shift tensors. (Reprinted from ref. 103 with permission.)

shift interaction is given by Equation (46):

$$\sigma = \sigma_{11}\cos^2\alpha\sin^2\beta + \sigma_{22}\sin^2\alpha\cos^2\beta + \sigma_{33}\cos^2\beta \quad (46)$$

where σ is the observed chemical shift, σ_{11}, σ_{22}, and σ_{33} are the principal components of the chemical shift tensor, and α and β are the Euler angles relating the principal axis system of the chemical-shift tensor to the laboratory frame (Figure 7). Although the dipolar interaction points along the internuclear axis and is related to the magnetic field axis by a single rotation about an angle θ, the orientation of the chemical-shift tensor must be determined relative to a molecular axis and relative to the magnetic field axis. The standard approach for establishing the orientation of a chemical-shift tensor relative to a molecular axis system has been through NMR studies of single crystals in which the molecular axis orientation can be determined independently using X-ray methods.[174–176] More recent methods have been developed that make use of the axial orientation of dipolar interactions as a convenient frame of reference.[177–180]

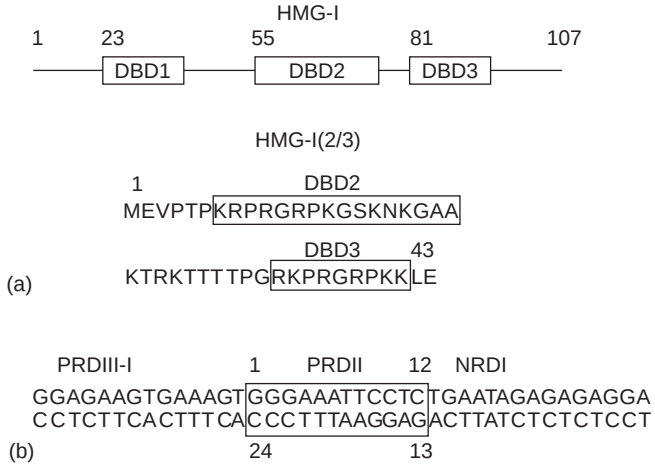

Figure 11 (a) Organization of human HMG-I(Y) showing the positions of the three DBDs and the sequence of HMG-I(Y)(2/3) (which consists of residues 50–91). (b) Partial sequence of the interferon-β promoter showing the positions of the PRDII, PRDIII-I and NRDI elements. The dodecamer used to make the complex with HMG-I(Y)(2/3) is boxed. (Adapted from ref. 181 with permission.)

5 EXAMPLES OF BIOMOLECULAR NUCLEAR MAGNETIC RESONANCE

There are essentially limitless ways that NMR spectroscopy can be and has been applied to the study of biomolecules. Rather than attempting to provide an exhaustive treatment here (which has been covered more fully elsewhere[8]), this section highlights a few examples to illustrate the use of NMR to study proteins, nucleic acids and enzymes, and one elegant example from the literature of a membrane protein structure determination.

5.1 Proteins and Nucleic Acids

A brief example of the ways in which NMR can be applied to the structure determination of a protein–DNA complex is recent work on the high-mobility group (HMG) I/Y protein. Using high-resolution multinuclear multidimensional solution-state NMR spectroscopy, the structure of the protein–DNA complex of the HMG I/Y protein complexed to a 12-mer oligonucleotide was determined.[181] This protein is relatively small (M_r = 11 kDa, 107 residues), so at first sight would be amenable to standard ^{1}H-NMR COSY and NOESY approaches outlined in section 2.3.1. However this protein turns out to be unusual in that, in the absence of DNA, the protein is largely unstructured, and contains a large number of lysine, arginine, and proline residues, making assignments difficult. In the presence of DNA, the protein becomes partially structured, and binds largely to the minor groove of DNA, although it still remains a difficult assignment problem, particularly when complexed to 12-mer DNA. Therefore it was necessary to generate ^{13}C/^{15}N-labeled protein truncated to contain two of the three DNA-binding domains (DBDs) (Figure 11). A large number of standard 3D NMR experiments were carried out on the protein complexed with 12-mer DNA (Table 1).

Representative strip plots from 3D CBCANH and CBCA(CO)NH spectra are shown in Figure 12. Use

Table 1 NMR of protein–DNA complexes

Protein assignments
2D ^{15}N HSQC, 3D ^{13}C-edited ^{15}N HSQC
3D HCCH-COSY, HCCH-TOCSY
3D ^{15}N-edited HOHAHA
3D HNHA, H(CA)NH, HNCA
3D HN(CO)CA, HNCO, HCACO
3D HCA(CO)N, CBCA(CO)NH, HN(CO)HB
3D CBCANH, HBHA(CO)NH, HACAHB
3D HBHANH, C(CO)NH, H(CCO)NH
3D ^{15}N/^{15}N-edited NOESY
4D ^{15}N/^{13}C-edited NOESY
4D ^{13}C/^{13}C-edited NOESY/ROESY
DNA assignments
2D ^{12}C, ^{14}N-filtered COSY
2D ^{12}C, ^{14}N-filtered HOHAHA
2D ^{12}C, ^{14}N-filtered NOESY/ROESY
Intermolecular complex assignments
3D ^{15}N-edited, ^{14}N, ^{12}C-filtered NOESY
3D ^{13}C-edited, ^{12}C-filtered NOESY

ROESY, rotating frame Overhauser effect spectroscopy.

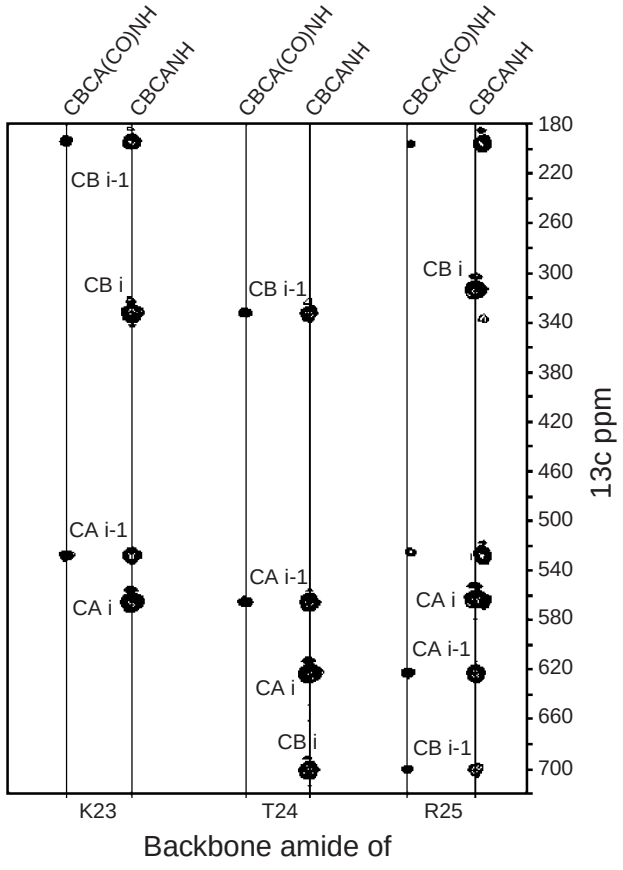

Figure 12 Representative CBCANH and CBCA(CO)NH strips from HMG-I(Y)(2/3)–PRDII DNA complex, illustrating the assignment procedure.

of ^{13}C-separated, ^{12}C-filtered, and ^{15}N-separated, ^{12}C-filtered NOESY experiments[182–191] has enabled identification of over 100 intermolecular NOE contacts between the protein and the DNA (Figure 13). These experiments rely on the ability to select either the species that is isotopically labeled, or the species that is not, as illustrated in Figure 14. Using ^{12}C/^{12}C-filtered NOESY experiments the DNA spectrum was assigned. A representative region is shown in Figure 15. One discovery with this sample was that only one HMG-I(Y) DBD was bound to the 12-mer DNA operator at any one time, with a difference between the two bonding affinities for the two DBDs in the truncated protein used for the structure determination. The structures of each DBD bound to DNA have been determined, as illustrated in Figure 16. A summary of the structural statistics is given in Table 2. Note that the region between the two DBDs is involved in protein–protein interactions, but because the DNA operator used in this structure determination was too short, no structural information could be obtained for this region.[181]

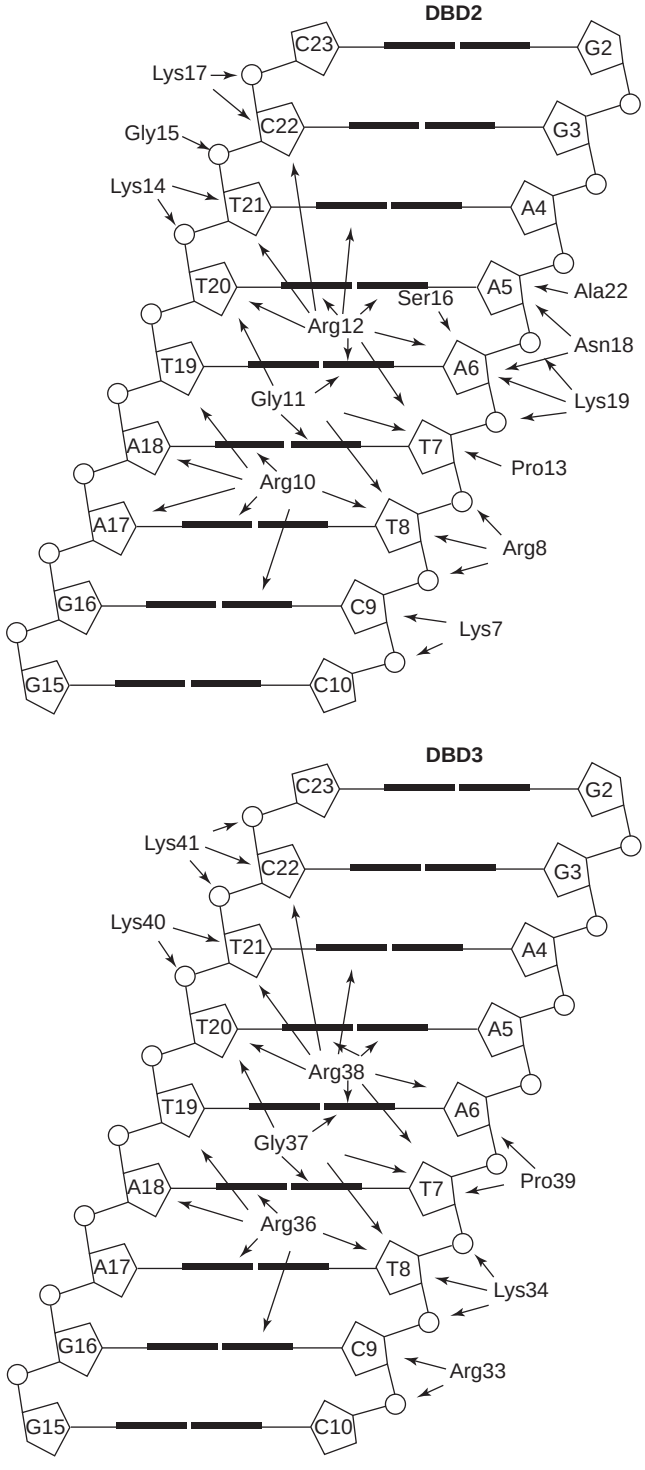

Figure 13 Summary of the DNA contacts involving DBD2 and DBD3. The DNA is represented as a cylindrical projection viewed from the minor groove side. Bases are indicated as thick lines, the deoxyribose sugar rings as pentagons, and the phosphates as circles. Contacts involving amino acid side-chains and backbone amides are indicated by solid and dashed arrows, respectively. (Adapted from ref. 181 with permission.)

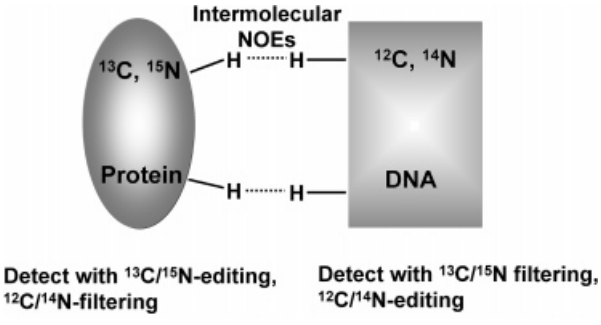

Figure 14 The principles of 3D NMR filtering experiments.

5.2 Enzymes

In addition to NMR being useful in determining protein structure, it can be used to probe functionally critical regions of proteins, such as encountered in the active site of an enzyme. An example of this is work on the enzyme 5-enolpyruvyl-shikimate-3-phosphate (EPSP) synthase, which catalyzes the reaction outlined in Scheme 1; EPSP synthase is of interest as the primary target for the broad-spectrum herbicide glyphosate (*N*-(phosphonomethyl)glycine (GLP)) which is the principal ingredient of Roundup®.

This enzyme has been extensively studied by kinetic and biophysical methods since the late 1980s. Direct observation of the enzyme-bound intermediate complex by solution-state NMR was first reported in 1989,[192,193] and later confirmed by another laboratory.[194] This is illustrated in Figure 17. Some site-directed mutants of the enzyme by steady-state kinetics, kinetic isotope effect analysis, fluorescence spectroscopy, and solution-state ^{31}P-, ^{13}C-, and ^{15}N-NMR spectroscopy have since been characterized.[195–198] The enzyme has also been characterized qualitatively by solid-state REDOR NMR spectroscopy,[199] and a quantitative REDOR study has been completed.[200]

In time-resolved solid-state NMR, the enzyme and substrate are mixed together very rapidly and fired into a secondary cryogen such as liquid propane cooled to about 85 K.[201–206] This rapid freeze–quench generates a snow that can be packed in a solid-state NMR rotor for subsequent NMR analysis. In this manner enzymatic

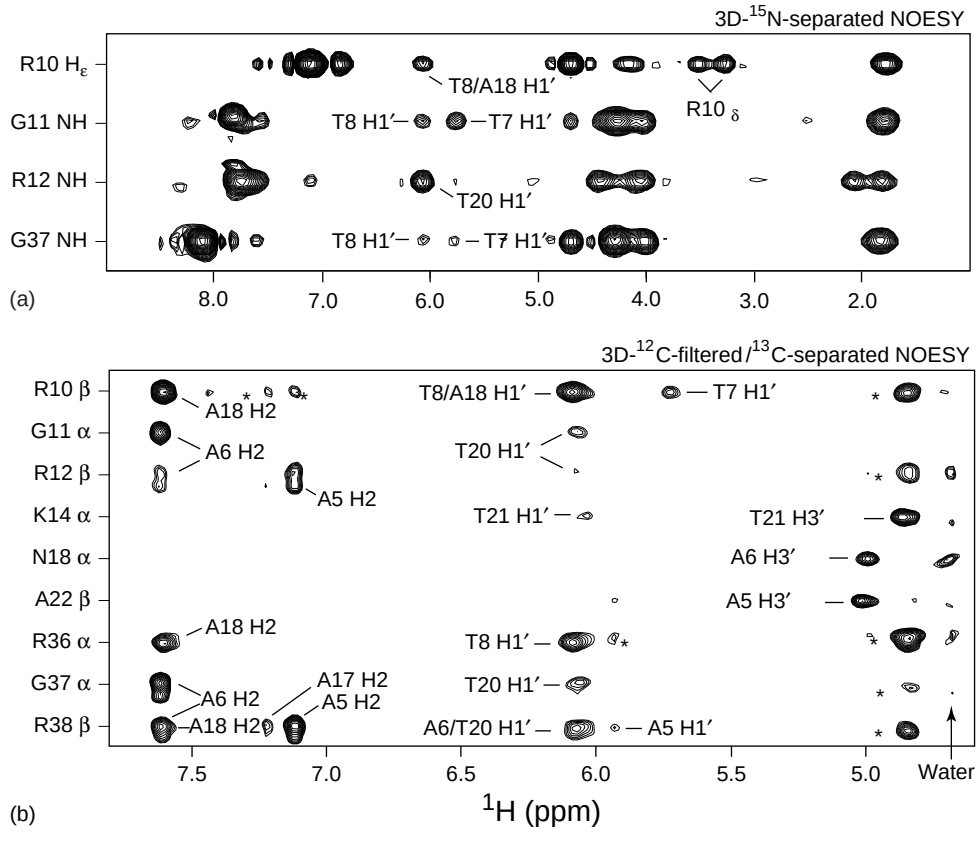

Figure 15 Composite of (a) ^{15}N–H and ^{13}C–H strips taken from a 3D ^{15}N-separated NOE spectrum and (b) a 3D ^{13}C-separated/^{12}C-filtered NOE spectrum. Cross-peaks marked with an asterisk have their maxima on another plane. The cross-peaks between Arg 38(CβH) and the H2 protons of A17 and A18, designated in italics, arise from a minor (<10%) population in which DBD3 is bound in the opposite orientation. (Adapted from ref. 181 with permission.)

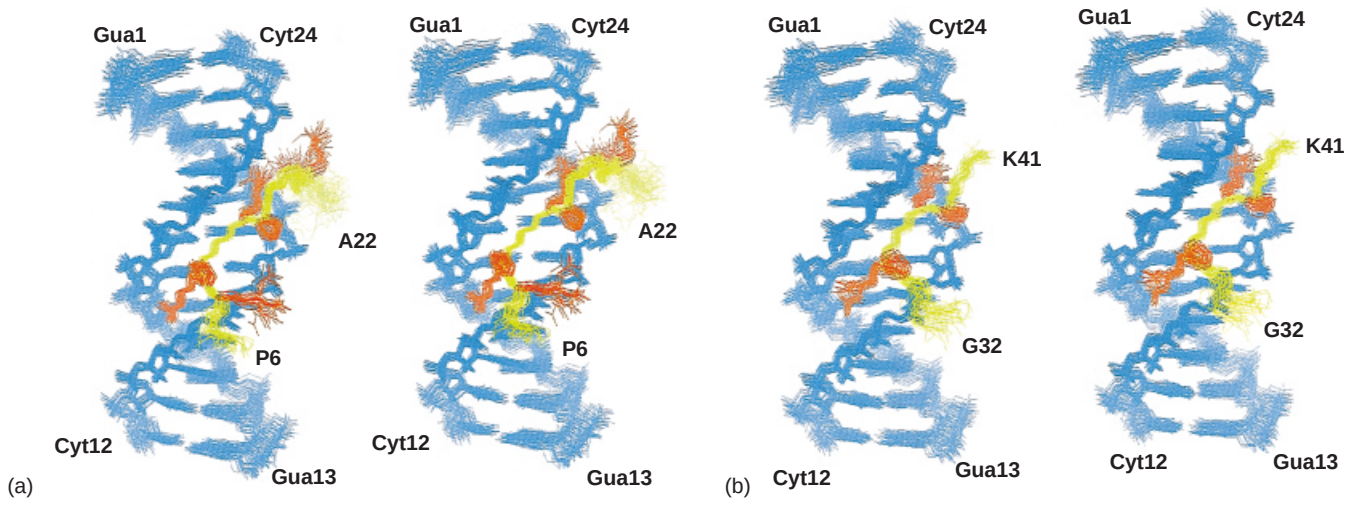

Figure 16 Stereo view of superposition of HMG-I(Y)–DNA complex showing 35 NMR structures of (a) DBD2 and (b) DBD3. The backbone (N, Cα, C) atoms of the DBDs are shown in yellow; side-chains of Arg8, Pro9, Arg10, Arg12, Pro13, Lys17 and Asn18 for DBD2, and Pro35, Arg36, Arg38 and Pro39 of DBD3 are shown in red; and the DNA is shown in blue. (Adapted from ref. 181 with permission.)

reactions can be stopped with a time resolution of about 6 ms along the reaction coordinate, and the relatively stable and therefore populated ground-state species examined as a function of time. Concerns about the fate of the protein under these conditions of freezing have been addressed,[207,208] and at the protein concentrations and freezing rates (10^5 K s^{-1}, or millisecond time regime) that were employed, there is significant dispersal of the solute in the frozen water.[204] Furthermore the frozen water is probably largely amorphous,[204] with the protein itself acting in a similar manner to a cryoprotectant,[209] thereby reducing the formation of hexagonal ice that is detrimental to the protein. The specific activity of EPSP synthase employed in these freeze–quench experiments was assayed before and after rapid freezing and found to be unchanged at the high protein concentrations used. The time-resolved solid-state NMR technique has been successfully applied to EPSP synthase[201–203,205] and uridine diphosphate *N*-acetyl glucosamine enolpyruvyl transferase (UDP-NAG EPT).[206] The work with EPSP synthase (Figure 18) showed that it was possible to use solid-state NMR to follow the build-up of the intermediate (with concomitant loss of substrate and build-up of product) under pre-steady-state kinetic conditions.[205] These studies confirmed the intermediates reported by Anderson and Johnson,[210] determined by chemical quench methods.

Time-resolved solid-state NMR has been coupled with REDOR for qualitative structural studies of complexes of EPSP synthase[199] and UDP-NAG EPT,[211] and REDOR used as an editing technique in combination with time-resolved solid-state NMR to detect the transient enzyme–intermediate of EPSP synthase.[166] REDOR dephasing on a nucleus unique to the substrate or intermediate can be used to illuminate a sphere of a radius defined by the limit of the technique, and to light up selected residues within the vicinity if they are appropriately isotopically labeled (Figure 19). Following a report for a new mechanism for EPSP synthase,[212] quantitative REDOR measurements on several species formed during the enzymatic reaction proved the proposed new mechanism to be incorrect.[200] The X-ray structure of the unliganded form of EPSP synthase has been reported,[213] but to date there has been no published report of the structure in the presence of substrates. It is known that substrates induce a conformational change, and it is thought that the protein shifts to a closed form. Therefore, several attempts have been made to characterize the structure of the closed conformation by NMR.[214–218] In order to carry out intramolecular and intermolecular distance measurements by dipolar recoupling techniques, it is necessary to have a fixed site on the protein whose location has been established unambiguously, and make measurements either within the protein or to substrates whose location is also known. The weakness of some work in this field is ambiguity as to the location of one or other partner in the dipolar-coupled spin pair. This is a nontrivial problem, because techniques for site-specific isotopic labeling of proteins are generally insufficiently developed to be able to generate NMR quantities of sample. Uniform isotopic labeling is problematic because making sequence-specific assignments by solid-state NMR is not yet possible for proteins of 46 kDa. This laboratory is in the process of introducing a number of novel approaches to site-specific labeling and carrying out dipolar recoupling measurements to generate a limited set of

Table 2 Structural statistics[a]

Parameter	DBD2 ⟨SA⟩	DBD3 ⟨SA⟩
RMSDs from NOE interproton distance restraints (Å)[b]		
All (624) (501)	0.017 ± 0.003	0.014 ± 0.001
Protein		
interresidue sequential ($\|i-j\|=1$) (71/31)	0.024 ± 0.014	0.00005 ± 0.0003
interresidue medium range ($1 < \|i-j\| \leq 5$) (4/2)	0.02 ± 0.027	0.016 ± 0.025
intraresidue (64/2)	0.0006 ± 0.007	0.007 ± 0.012
DNA		
intraresidue (249/249)	0.0007 ± 0.002	0.0004 ± 0.001
sequential intrastrand (119/1 19)	0.012 ± 0.004	0.010 ± 0.004
interstrand (44/44)	0.011 ± 0.007	0.011 ± 0.006
Protein-DNA (73/34)[c]	0.031 ± 0.007	0.042 ± 0.008
RMSDs from DNA H-bond restraints (Å) (42/42)[d]	0.0205 ± 0.008	0.022 ± 0.006
RMSDs from distance restraints to phosphates (5/4)[e]	0.0008 ± 0.004	0.003 ± 0.010
RMSDs from "repulsive" restraints (Å) (20/28)[f]	0.025 ± 0.011	0.010 ± 0.008
RMSDs from experimental dihedral restraints (deg) (172/153)[b]	0.093 ± 0.057	0.190 ± 0.096
RMSDs from experimental coupling constants		
$3J_{\mathrm{HN}\alpha}$ (Hz) (13/6)[b]	0.58 ± 0.08	0.827 ± 0.046
$3J_{\mathrm{coco}}$ (Hz) (8/3)[b]	0.38 ± 0.11	0.308 ± 0.007
RMSDs from experimental ^{13}C shifts		
^{13}C α (ppm) (21/9)	0.873 ± 0.082	0.926 ± 0.096
^{13}C β (ppm) (18/8)	0.534 ± 0.072	0.511 ± 0.109
Deviations from idealized covalent geometry		
bonds(Å) (1229/999)	0.006 ± 0.0007	0.006 ± 0.0005
angles (°) (2238/1814)	0.863 ± 0.027	0.923 ± 0.017
impropers (°) (603/496)[g]	0.497 ± 0.057	0.511 ± 0.062
E_{L-J} (kcal mol^{-1})[h]	-402 ± 7	-36 ± 7
Coordinate precision (Å)[i]		
protein backbone plus DNA	0.62 ± 0.12	0.63 ± 0.15
all protein atoms plus DNA	0.72 ± 0.11	0.77 ± 0.14
protein backbone	0.57 ± 0.17	0.62 ± 0.20
all protein atoms	0.98 ± 0.19	1.25 ± 0.25
DNA	0.58 ± 0.14	0.58 ± 0.16

[a] ⟨SA⟩ is the final set of 35 simulated annealing structures for each complex. $\overline{SA}$ are the mean structures for the two complexes obtained by averaging the coordinates of the individual SA structures best fitted to each other (with respect to residues 6–19 of the protein DBD2 and residues 32–41 of the protein DBD3, and base pairs 1–12 of the DNA). $(\overline{SA})_{\mathrm{r}}$ are the restrained regularized mean coordinates for the two complexes obtained by restrained regularization of the mean SA coordinates. The number of terms for the various restraints is given in parentheses with the first number referring to the DBD2–DNA complex and the second to the DBD3–DNA complex.

[b] None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, $^3J_{\mathrm{HN}_\alpha}$ coupling constant violations greater than 2 Hz, or $^3J_{\mathrm{coco}}$ coupling constant violations greater than 0.5 Hz.

[c] The number of NOEs involving $(\quad r^{-6})^{-1/6}$ sum restraints to multiple DNA protons is 27 for the DBD2–DNA complex and nine for the DBD3–DNA complex. The effect of such restraints is to permit a close interproton distance contact to be obtained from a proton of the protein to whichever of the protons on the DNA is specified in the restraint, the $(\quad r^{-6})^{-1/6}$ sum restraint being satisfied providing only that at least one of these DNA protons is close to the specified proton of the protein.

[d] The hydrogen bond restraints within the DNA were used to maintain Watson–Crick base pairing.

[e] The Nζ atom of a lysine or guanidino nitrogen atoms of arginines were restrained within 7 Å of a DNA phosphate atom when NOEs from a residue (Lys7, Arg8, Lys14, Lys17, Lys19, Arg33 and Lys41) or when structure calculations and ^{15}N relaxation measurements (Lys34 and Lys40) indicated that residue interacts with a DNA phosphate. $(\quad r^{-6})^{-1/6}$ sum restraints were used and in each case a choice of two adjacent phosphate atoms was given.

[f] In the final stages of the structure calculations two types of repulsive distance restraints, with a lower bound of 4 Å (and an unrestrained upper bound), were introduced to facilitate convergence. First, repulsive distance restraints were used to prevent energetically unfavorable proximity of hydrogen bond donors to other donors, and hydrogen bond acceptor groups to other acceptors: there were 11 such restraints for the DBD2–DNA complex and 10 for the DBD3–DNA complex. Second, a small number of repulsive restraints were also used to prevent structures being generated which predicted short interproton distances (<3 Å) for which no corresponding NOEs could be observed: there were nine such restraints for the DBD2–DNA complex and 18 for the DBD3–DNA complex.

[g] The improper torsion restraints serve to maintain planarity and chirality.

[h] E_{L-J} is the calculated Lennard–Jones van der Waals energy and is not included in the target function for simulated annealing or restrained minimization.

[i] The precision of the coordinates is defined as the average atomic RMS difference between the 35 individual simulated annealing structures of each complex and the mean coordinates $\overline{SA}$. The values refer to residues 6–19 and 32–41 of DBD2 and DBD3, respectively, and to base pairs 1–12 of the DNA.

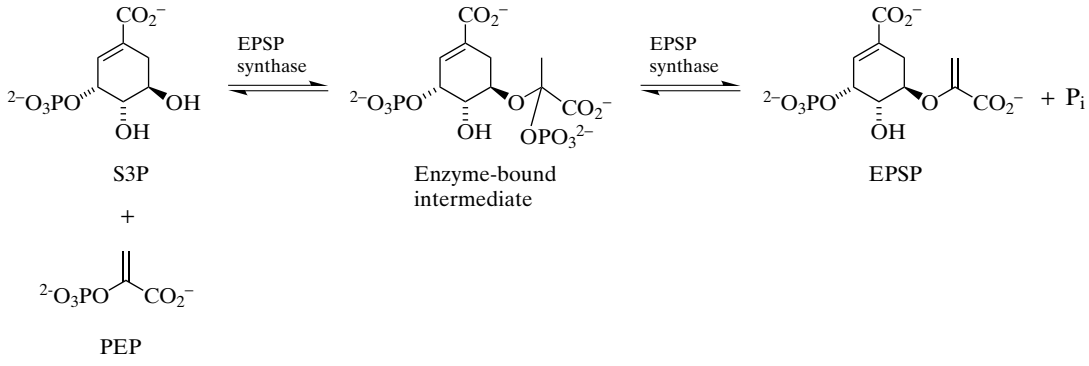

Scheme 1

accurate long-range distance restraints. These are being used in conjunction with short-range NOE-derived distance restraints from solution-state NMR data to solve the structure of the closed conformation of EPSP synthase. Figure 20 shows the HSQC spectrum of an N-terminal fragment of EPSP synthase, and Figure 21 shows the divide-and-conquer strategy taken to solve such a large protein structure.

5.3 Membranes

As membranes and membrane proteins cannot easily be crystallized, NMR spectroscopy is a particularly powerful method for probing their structure and function.[219–221] This section considers the membrane protein gramicidin A, whose structure has been determined in the presence of lipid bilayers.

Gramicidin A is a small hydrophobic peptide composed of 15 alternating L- and D-amino acids ($M_r \approx$ 1600).[222] The peptide dimerizes and forms channels that conduct monovalent cations. There are substantial differences between two X-ray crystal structures of the peptide,[223,224] and the bilayer structure that has been proposed on the basis of circular dichroism

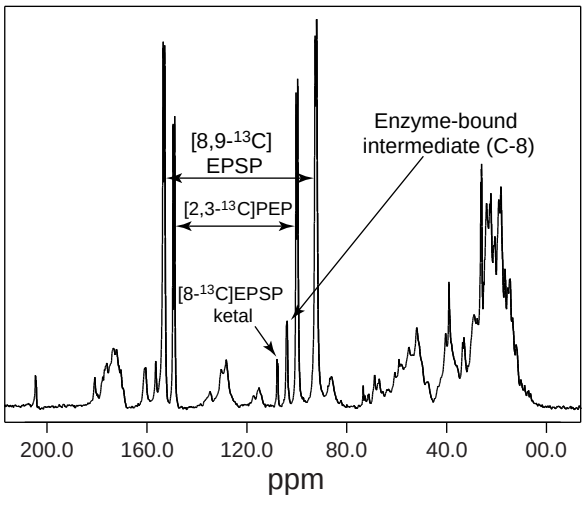

Figure 17 ^{1}H-decoupled ^{13}C solution-state NMR spectrum of EPSP synthase plus substrates (S3P, shikimate-3-phosphate, and [2,3-$^{13}C_2$]PEP, phosphoenolpyruvate) under equilibrium conditions, showing the resonances due to the transient enzyme–intermediate.

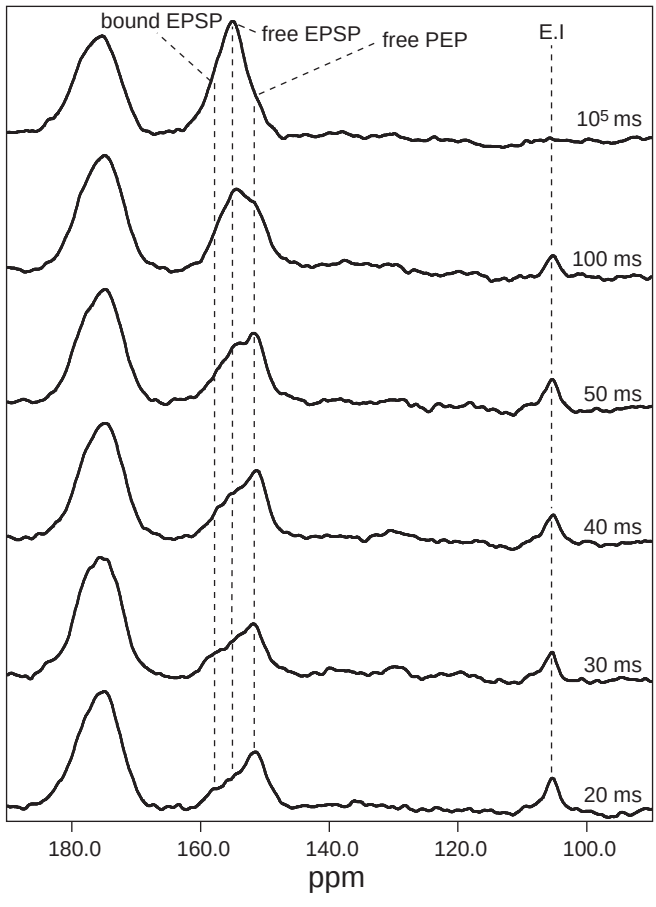

Figure 18 Time-resolved solid-state NMR spectra of the enzyme EPSP synthase plus [^{13}C]-labeled substrate PEP at 233 K, showing the transient formation of the enzyme–intermediate and the formation of product (EPSP) as a function of reaction time. (Adapted from ref. 205 with permission.)

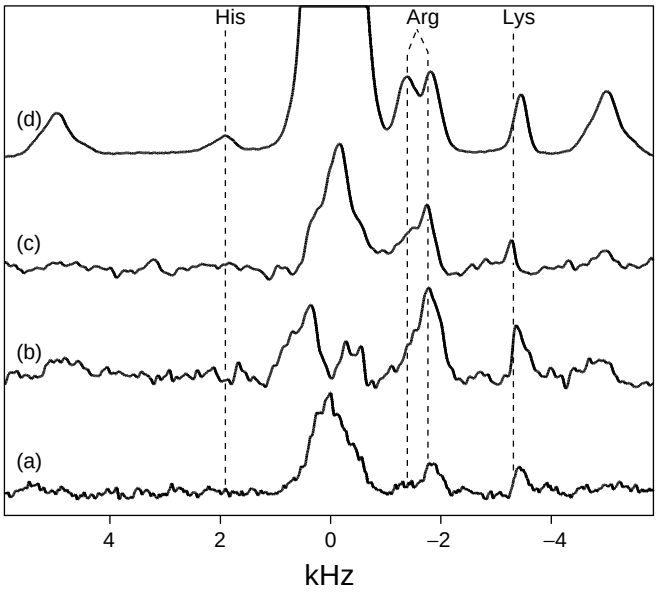

Figure 19 Nitrogen-15-detected, ^{31}P-dephased CP/MAS solid-state REDOR difference (ΔS) NMR spectra of [U–^{15}N]EPSP synthase • S3P plus a variety of inhibitors/substrates: (curve a) GLP at 295K; (b) 3-(*Z*)-fluoro-PEP at 273K; (c) [2-^{13}C]PEP under time-resolved conditions at 223K; (d) the full echo spectrum (S_0) of the sample in (a). Resonances due to a subset of lysine and arginine residues within a 6-Å radius of the substrates suggest that these residues are located in the enzyme-active site. A histidine residue putatively located in the enzyme-active site is clearly not detectable. Unfortunately, although the numbers of proximal lysine and arginine residues can be estimated, no site-specific assignments can be made from these data. (Adapted from ref. 192 with permission.)

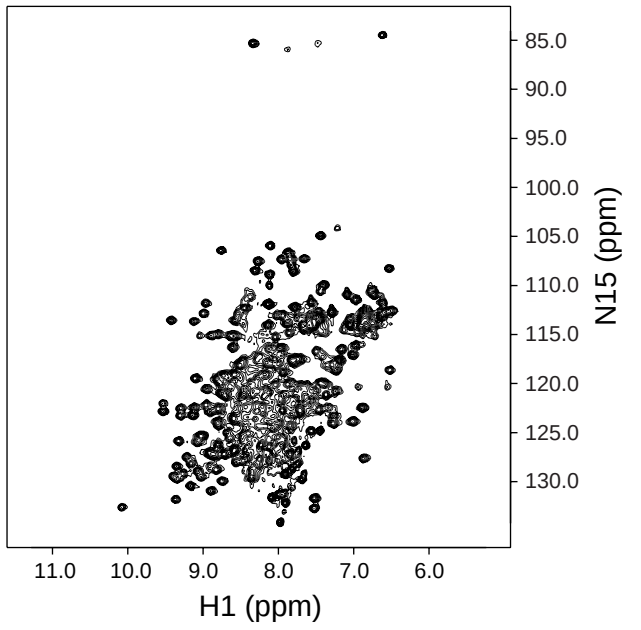

Figure 20 The ^{15}N-HSQC of the N-terminal domain of EPSP synthase. The spectrum displays reasonably good dispersion and resolution for a 23-kDa protein (the size of the N-terminal fragment). However, even when spread in three dimensions, as with an HNCO experiment, it is not sufficiently well-resolved for a complete assignment. Deuterium labeling is necessary to achieve adequate resolution.

measurements. Such a controversy is amenable to solution by NMR methods.

Cross et al. have conducted a set of experiments using orientation-dependent solid-state NMR studies to address the structure of gramicidin in membranes.[170,225–227] They examined by ^{15}N- and ^{13}C-NMR samples of gramicidin, specifically [^{15}N]- and [^{13}C]-labeled at particular residues, and oriented between glass plates in lipid bilayers. They succeeded in determining the structure by measurement of dipolar splittings and calculating ϕ and ψ torsion angles. Figure 22 shows the spectra from gramicidin, ^{15}N-labeled at Ala3 and Leu4. The measurements are of the Gly2–Ala3 and Ala3–Leu4 peptide bonds and the dipolar splittings ($\Delta\nu$) of the amide ^{15}N resonances resulting from the directly bonded carbonyl carbons. The top spectra show the spectra of oriented gramicidin singly ^{15}N-labeled at the amide nitrogens of Ala3 (left) and Leu4 (right), and the lower spectra show the dipolar splitting of the ^{15}N by ^{13}C when the peptide bonds are doubly labeled. The observed dipolar splittings are 670 Hz for the ^{13}C-Gly2–^{15}N-Ala3 bond and 820 Hz for the ^{13}C-Ala3–^{15}N-Leu4 bond. An accurate

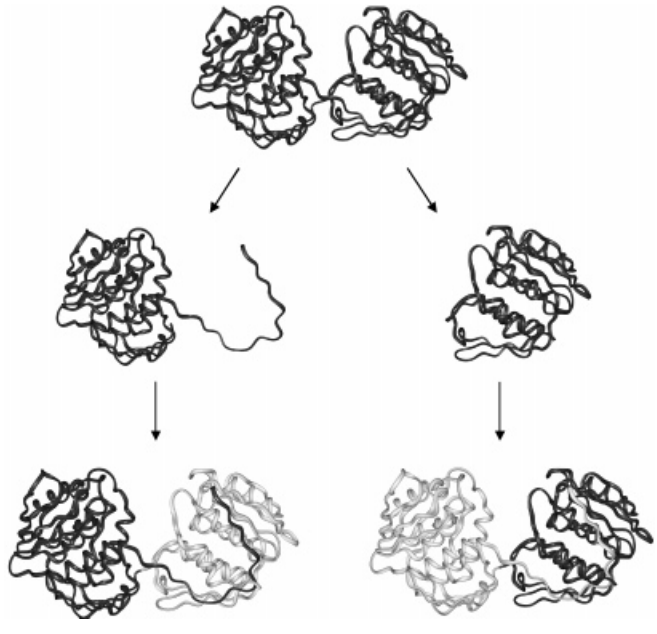

Figure 21 The divide-and-conquer strategy for attempting to solve the solution-state NMR structure of EPSP synthase. The N- and C-terminal domains are expressed separately, and after reconstitution can be isotopically labeled separately, thereby halving the resonance overlap problem for this 46-kDa protein.

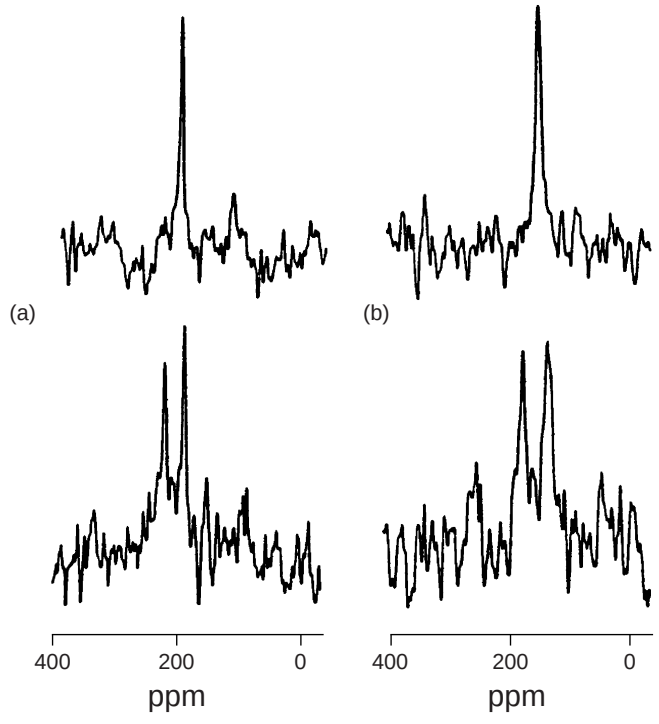

Figure 22 ^{15}N spectra of gramicidin A in DMPC (dimyristoylphosphatidylcholine) bilayers. The top spectra are of gramicidin singly labeled with ^{15}N at the amides of (a) Ala3 and (b) Leu4, and the bottom spectra illustrate the dipolar splitting observed when the peptide bonds are doubly labeled with ^{13}C. (Reprinted from ref. 170 with permission.)

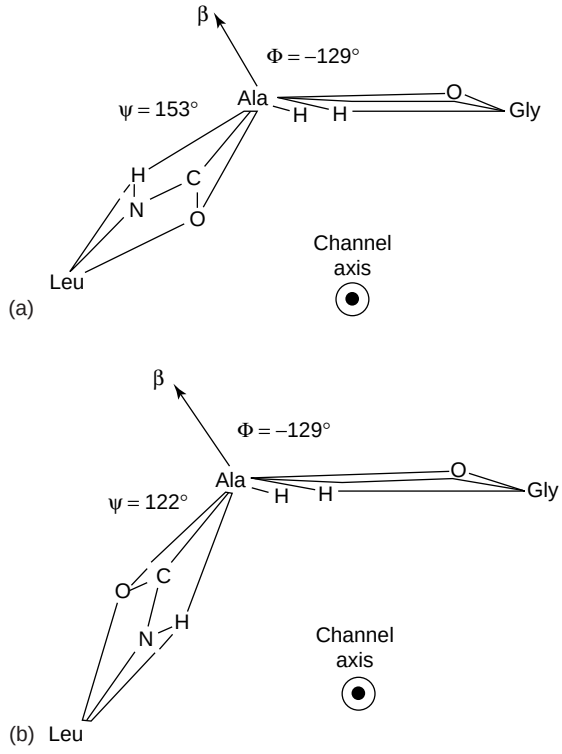

Figure 23 Two possible structures for the peptide planes at Ala3 in gramicidin A. Both correspond to right-handed β-type helices but differ in the orientation of the Ala3 carbonyl group, which points toward the gramicidin channel axis (*left*) or away from it (*right*). (Reprinted from ref. 170 with permission.)

bond length is the only additional information needed to calculate the angle θ. A 1.34-Å C–N bond length taken from the crystal structure of alanylalanine translates into a dipolar coupling constant D of 1.26 kHz, so from Equation (45), these data define four possible orientations for each C–N bond because D is larger than the observed splittings.

To limit these possibilities, independent restraints on the peptide plane orientations are obtained from the ^{15}N–^{1}H dipolar couplings and the ^{15}N chemical shift tensor. The ^{15}N–^{1}H dipolar couplings are determined using separated local field experiments.[228–232] These experiments are designed to separate the dipolar couplings from the chemical shifts. They can also be used to relate the chemical shift and dipolar coupling tensors and, as the latter is readily related to the molecular coordinate system, the orientation of the principal components of the shift anisotropy can be determined directly. In gramicidin, the separated local field experiments yield only two possible orientations for each N–H bond, because the ^{15}N–^{1}H dipolar coupling constant is smaller than the observed splittings. As the orientation of the ^{15}N chemical shift tensor relative to the molecular frame has been established independently by reference to model compounds,[177–180,229] the observed ^{15}N chemical shifts provide a way to discriminate between several of the orientations of the N–H and N–C bonds implied by the dipolar couplings. These data, together with additional geometric restraints, define two possible sets of torsion angles for the Ala3 position with errors in the ϕ and ψ angles of $\pm 6^\circ$ and $\pm 5^\circ$ respectively (Figure 23).

The two structures defined by the Ala3 torsion angles are right-handed β-type helices. The difference between the two structures is the orientation of the carbonyl group with respect to the channel axis, which in the upper structure (a) points towards the channel ($\psi = 153^\circ$) and in the lower structure (b) points away from it ($\psi = 122^\circ$). Structure (a) is the favored model because it is biologically sensible, because the selectivity of cations over anions by gramicidin might arise because of the partial negative charges of the backbone carbonyls, which could solvate the cations after they have been stripped of water as they enter the channel.

An alternative approach has been adopted by Schaefer et al.,[233] who used 2D REDOR on Val1-[1-^{13}C]Gly2-[^{15}N]Ala3-gramicidin A in multilamellar dispersions

of DMPC to measure dipolar couplings and infer from these four possible angles between the Gly2-Ala3 $^{13}C-^{15}N$ peptide bond. This information supported a right-handed single-stranded helical dimer as a structural model for gramicidin in multilamellar dispersions.

We have given a detailed description of how torsion angle restraints were derived for one set of dipolar splittings. Cross et al.[234–237] have carried out a systematic determination of the structure of the backbone and side-chains for gramicidin in a lipid environment using 141 orientational restraints derived from measurements of ^{15}N chemical shift anisotropy, $^{15}N-^{1}H$ and $^{15}N-^{13}C$ dipolar splitting, and ^{2}H quadrupole splitting. The final structure, shown in Figure 24, was calculated using an algorithm similar to simulated annealing on a set of 40 random structures, and had RMSDs of $\pm 5^{\circ}$, which reflects the very high precision of the solid-state NMR measurements.

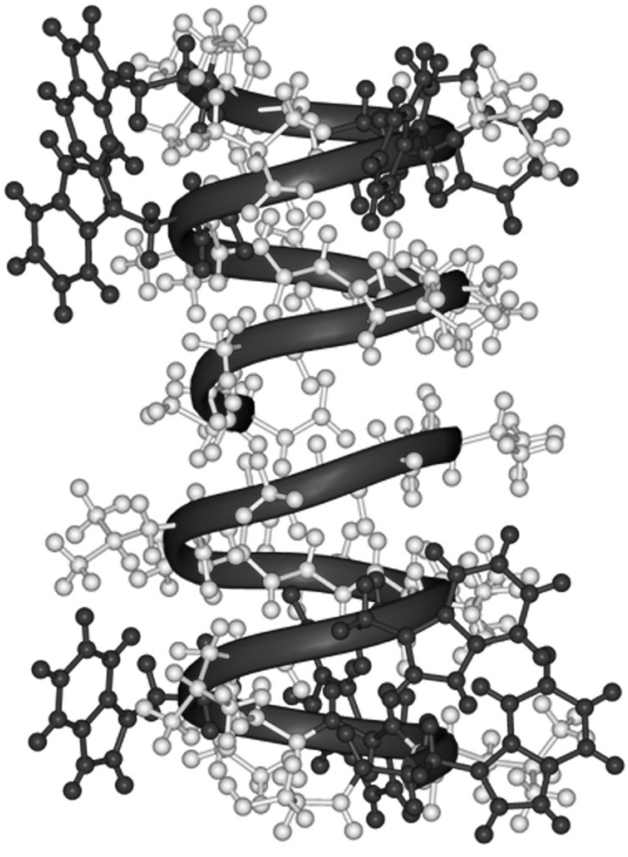

Figure 24 Structure of gramicidin A in a membrane bilayer as determined by solid-state NMR.

ACKNOWLEDGMENTS

I should like to thank members of my research group past and present (Paul Barlow, Richard Appleyard, Wendy Shuttleworth, Cecilia Ramilo, Yan Li, Jaroslav Zajicek, Tatyana Igumenova, Dan Mitchell, David Jakeman, Greg Helms, and Melissa Pohl) for their help with much of the work described here. I should also like to acknowledge Angela Gronenborn, Marius Clore, Jeff Huth, Carole Bewley, Dan Garrett, Frank Delaglio (The National Institutes of Health) and Mark Nissen and Ray Reeves (WSU) for their kind help. This work was supported by NIH grant GM43215. The WSU NMR Center equipment was supported by NIH grants RR0631401 and RR12948, and NSF grants CHE-9115282 and DBI-9604689.

ABBREVIATIONS AND ACRONYMS

C7	7-Fold Symmetric Homonuclear Recoupling
CHARMM	Chemistry at Harvard Molecular Mechanics
COSY	Correlation Spectroscopy
CP/MAS	Cross-polarization/Magic-angle Spinning
CROWN	Combined Rotation with Nutation
DBD	DNA-binding Domain
DMPC	Dimyristoylphosphatidylcholine
DQFCOSY	Double-quantum-filtered Correlation Spectroscopy
DRAMA	Dipolar Recovery at the Magic Angle
DRAWS	Dipolar Recoupling with a Windowless Sequence
ECOSY	Exclusive Correlation Spectroscopy
EPSP	5-Enolpyruvyl-shikimate-3-phosphate
GLP	*N*-(Phosphonomethyl)glycine
HMG	High-mobility Group
HMQC	Heteronuclear Multiple Quantum Coherence
HOHAHA	Homonuclear Hartmann–Hahn
HORROR	Homonuclear Rotatory Resonance
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclear Enhancement by Polarization Transfer
IPAP	In-phase/Antiphase
MAS	Magic-angle Spinning
MELODRAMA	Melding of Spin-locking and DRAMA

NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PECOSY	Primitive Exclusive Correlation Spectroscopy
PEP	Phosphoenolpyruvate
POST-C7	Compensated 7-fold Symmetric Homonuclear Recoupling
REDOR	Rotational Echo Double Resonance
RF	Radiofrequency
RFDR	Rotor-driven Dipolar Recoupling
RMD	Restrained Molecular Dynamics
RMS	Root Mean Square
RMSD	Root-mean-square Deviation
ROESY	Rotating Frame Overhauser Effect Spectroscopy
SEDRA	Simple Excitation for the Dephasing of Rotational Amplitudes
S3P	Shikimate-3-phosphate
TEDOR	Transferred-echo Double Resonance
TOCSY	Total Correlation Spectroscopy
TOSS	Total Side-band Suppression
TPPM	Two-pulse Phase Modulation
TROSY	Transverse Relaxation Optimized Spectroscopy
UDP-NAG EPT	Uridine Diphosphate *N*-Acetyl Glucosamine Enolpyruvyl Transferase
2D	Two-dimensional
3D	Three-dimensional
4D	Four-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance, General Medical

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. O. Jardetsky, G.C.K. Roberts, *NMR in Molecular Biology*, Academic Press, New York, 1981.
2. K. Wüthrich, *NMR in Biological Research: Peptides and Proteins*, Elsevier, New York, 1976.
3. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, 1986.
4. G.C.K. Roberts, *NMR of Macromolecules*, ed. G.C.K. Roberts, IRL Press, Oxford, 1993.
5. D.G. Reid, *Protein NMR Techniques*, ed. D.G. Reid, Humana Press, Totawa, New Jersey, 1997.
6. J.L. Markley, S.J. Opella, *Biological NMR Spectroscopy*, eds. J.L. Markley, S.J. Opella, Oxford University Press, New York, 1997.
7. T.L. James, *Nuclear Magnetic Resonance in Biochemistry*, Academic Press, New York, 1975.
8. J.N.S. Evans, *Biomolecular NMR Spectroscopy*, Oxford University Press, Oxford, 1995.
9. R.A. Dwek, *NMR in Biochemistry*, Oxford University Press, Oxford, 1973.
10. J. Cavanagh, W.J. Fairbrother, A.G. Palmer, III, N.J. Skelton, *Protein NMR Spectroscopy*, Academic Press, San Diego, CA, 1996.
11. A. Bax, M.F. Summers, '^{1}H and ^{13}C Assignments from Sensitivity Enhanced Detection of Heteronuclear Multiple-bond Connectivity by 2D Multiple Quantum NMR', *J. Am. Chem. Soc.*, **108**, 2093–2094 (1986).
12. A.G. Redfield, 'Stimulated Echo NMR Spectra and their use for Heteronuclear Two-dimensional Shift Correlation', *Chem. Phys. Lett.*, **96**, 537 (1983).
13. A. Bax, R.H. Griffey, B.L. Hawkins, 'Correlation of Proton and Nitrogen-15 Chemical Shifts by Multiple Quantum NMR', *J. Magn. Reson.*, **55**, 301 (1983).
14. L. Müller, 'Sensitivity Enhanced Detection of Weak Nuclei using Heteronuclear Multiple Quantum Coherence', *J. Am. Chem. Soc.*, **101**, 4481–4484 (1979).

15. G. Bodenhausen, D.J. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185 (1980).
16. D.P. Burum, R.R. Ernst, 'Net Polarization Transfer via a J-Ordered State for Signal Enhancement of Low-sensitivity Nuclei', *J. Magn. Reson.*, **39**, 163–168 (1980).
17. G.A. Morris, R. Freeman, 'Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer', *J. Am. Chem. Soc.*, **101**, 760–762 (1979).
18. T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, I.D. Campbell, 'Comparison of Techniques for ^{1}H-Detected Heteronuclear ^{1}H-^{15}N Spectroscopy', *J. Magn. Reson.*, **87**, 488–501 (1990).
19. A. Bax, M. Ikura, L.E. Kay, D.A. Torchia, R. Tschudin, 'Comparison of Different Modes of Two-dimensional Reverse Correlation NMR for the Study of Proteins' *J. Magn. Reson.*, **86**, 304–318 (1990).
20. L.E. Kay, P. Keifer, T. Saarinen, 'Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity', *J. Am. Chem. Soc.*, **114**, 10663–10665 (1992).
21. M. Ikura, L.E. Kay, A. Bax, 'A Novel Approach for Sequential Assignment of Larger Proteins: Heteronuclear Triple-resonance Three-dimensional NMR Spectroscopy. Application to Calmodulin', *Biochemistry*, **29**, 4659–4667 (1990).
22. L.E. Kay, M. Ikura, R. Tschudin, A. Bax, 'Three-dimensional Triple-resonance NMR Spectroscopy of Isotopically Enriched Proteins', *J. Magn. Reson.*, **89**, 496–514 (1990).
23. S. Grzesiek, A. Bax, 'Improved 3D Triple-resonance NMR Techniques', *J. Magn. Reson.*, **96**, 432–440 (1992).
24. B.T. Farmer, R.A. Venters, L.D. Spicer, M.G. Wittekind, L. Müller, 'A Refocused and Optimized HNCA: Increased Sensitivity and Resolution in Large Macromolecules', *J. Biomol. NMR*, **2**, 195–202 (1992).
25. T. Yamazaki, W. Lee, C.H. Arrowsmith, D.R. Muhandiram, L.E. Kay, 'A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ^{15}N, ^{13}C, ^{2}H Labeled Proteins with High Sensitivity', *J. Am. Chem. Soc.*, **116**, 11655–11666 (1994).
26. A. Bax, M. Ikura, 'An Efficient 3D NMR Technique for Correlating the Proton and ^{15}N Backbone Amide Resonances with the α-Carbon of the Preceding Residue in Uniformly ^{15}N/^{13}C Enriched Proteins', *J. Biomol. NMR*, **1**, 99–104 (1991).
27. G.T. Montelione, G. Wagner, 'Conformation-independent Sequential NMR Connections in Isotope-enriched Polypeptides by ^{1}H-^{13}C-^{15}N Triple-resonance Experiments', *J. Magn. Reson.*, **87**, 183–188 (1990).
28. L.E. Kay, M. Ikura, A. Bax, 'The Design and Optimization of Complex NMR Experiments. Application to a Triple-resonance Pulse Scheme Correlating Hα, NH and ^{15}N Chemical Shifts in ^{15}N-^{13}C-Labeled Proteins', *J. Magn. Reson.*, **91**, 84–92 (1991).
29. S. Seip, J. Balbach, H. Kessler, 'An Improved Technique for Correlating Backbone and Amide Protons with ^{15}H and Hα Protons (HN(CA)H) in Isotopically Enriched Proteins', *J. Magn. Reson.*, **100**, 406–410 (1992).
30. R.T. Clubb, V. Thanabal, G. Wagner, 'A New 3D HN(CA)HA Experiment for Obtaining Fingerprint H^{N}-H^{α} Cross-peaks in ^{15}N- and ^{13}C-Labeled Proteins', *J. Biomol. NMR*, **2**, 203–210 (1992).
31. R.H. Fogh, D. Schipper, R. Boelens, R. Kaptein, '^{1}H, ^{13}C and ^{15}N NMR Backbone Assignments of the 269-Residue Serine Protease PB92 from *Bacillus alcalophilus*', *J. Biomol. NMR*, **4**, 123–128 (1994).
32. R.T. Clubb, G. Wagner, 'A Triple-resonance Pulse Scheme for Selectively Correlating Amide $^{1}H^{N}$ and ^{15}N Nuclei with the $^{1}H^{\alpha}$ Proton of the Preceding Residue', *J. Biomol. NMR*, **2**, 389–394 (1992).
33. D.R. Muhandiram, L.E. Kay, 'Gradient-enhanced Triple-resonance Three-dimensional NMR Experiments with Improved Sensitivity', *J. Magn. Reson. B*, **103**, 203–216 (1994).
34. L.E. Kay, G.Y. Xu, T. Yamakazi, 'Enhanced-sensitivity Triple-resonance Spectroscopy with Minimal H_2O Saturation', *J. Magn. Reson. A*, **109**, 129–133 (1994).
35. R.T. Clubb, V. Thanabal, G. Wagner, 'A Constant-time Three-dimensional Pulse Scheme to Correlate Intraresidue $^{1}H^{N}$, ^{15}N and ^{13}C′ Chemical Shifts in ^{15}N-^{13}C-labeled Proteins', *J. Magn. Reson.*, **97**, 213–217 (1992).
36. S. Grzesiek, A. Bax, 'Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Experiments', *J. Am. Chem. Soc.*, **114**, 6291–6293 (1992).
37. S. Grzesiek, A. Bax, 'An Efficient Experiment for Sequential Backbone Assignment of Medium-sized Isotopically Enriched Proteins', *J. Magn. Reson.*, **99**, 201–207 (1992).
38. M. Wittekind, L. Müller, 'HNCACB, a High Sensitivity 3D NMR Experiment to Correlate Amide-proton and Nitrogen Resonances with the Alpha- and Beta-carbon Resonances in Proteins', *J. Magn. Reson. B*, **101**, 201–205 (1993).
39. S. Grzesiek, A. Bax, 'Amino Acid Type Determination in the Sequential Assignment Procedure of Uniformly ^{13}C/^{15}N-Enriched Proteins', *J. Biomol. NMR*, **3**, 185–204 (1993).
40. A.C. Wang, P.J. Lodi, J. Qin, G.W. Vuister, A.M. Gronenborn, G.M. Clore, 'An Efficient Triple-resonance Experiment for Proton-directed Sequential Backbone Assignment of Medium-sized Proteins', *J. Magn. Reson. B*, **105**, 196–198 (1994).
41. R. Powers, A.M. Gronenborn, G.M. Clore, A. Bax, 'Three-dimensional Triple-resonance NMR of ^{13}C/^{15}N-Enriched Proteins Using Constant-time Evolution', *J. Magn. Reson.*, **94**, 209–213 (1991).

42. S. Grzesiek, A. Bax, 'The Origin and Removal of Artifacts in 3D HCACO Spectra of Proteins Uniformly Enriched with ^{13}C', *J. Magn. Reson. B*, **102**, 103–106 (1993).
43. F. Löhr, H. Rüterjans, 'Improved NMR Experiments for the Correlation of Backbone-carbon with Aliphatic-proton Resonances in Proteins', *J. Magn. Reson. B*, **109**, 80–87 (1995).
44. R. Weisemann, H. Ruterjans, W. Bermel, '3D Triple-resonance NMR Techniques for the Sequential Assignment of NH and ^{15}N Resonances in ^{15}N- and ^{13}C-Labeled Proteins', *J. Biomol. NMR*, **3**, 113–120 (1993).
45. S. Seip, J. Balbach, H. Kessler, 'A Simple Way for Sequential Assignment in Isotopically Enriched Proteins Using a H(N)CACO Correlation', *J. Biomol. NMR*, **3**, 233–237 (1993).
46. T. Szyperski, G. Wider, J.H. Bushweller, K. Wüthrich, '3D ^{13}C-^{15}N-Heteronuclear Two-spin Coherence Spectroscopy for Polypeptide Backbone Assignments in ^{13}C-^{15}N-Double-labeled Proteins', *J. Biomol. NMR*, **3**, 127–132 (1993).
47. F. Löhr, H. Rüterjans, 'A New Triple-resonance Experiment for the Sequential Assignment of Backbone Resonances in Proteins', *J. Biomol. NMR*, **6**, 189–197 (1995).
48. A. Bax, G.M. Clore, P.C. Driscoll, A.M. Gronenborn, M. Ikura, L.E. Kay, 'Practical Aspects of Proton–Carbon–Carbon–Proton Three-dimensional Correlation Spectroscopy of ^{13}C-Labeled Proteins', *J. Magn. Reson.*, **87**, 620–627 (1990).
49. A. Bax, G.M. Clore, A.M. Gronenborn, '^{1}H-^{1}H Correlation via Isotropic Mixing of ^{13}C Magnetization, a New Three-dimensional Approach for Assigning ^{1}H and ^{13}C Spectra of ^{13}C-Enriched Proteins', *J. Magn. Reson.*, **88**, 425–431 (1990).
50. L.E. Kay, G.-Y. Xu, A.U. Singer, D.R. Muhandiram, J.D. Forman-Kay, 'A Gradient-enhanced HCCH-TOCSY Experiment for Recording Side-chain ^{1}H and ^{13}C Correlations in H_2O Samples of Proteins', *J. Magn. Reson. B*, **101**, 333–337 (1993).
51. M. Sattler, M.G. Schwendinger, J. Schleucher, C. Griesinger, 'Novel Strategies for Sensitivity Enhancement in Heteronculear Multi-dimensional NMR Experiments Employing Pulsed Field Gradients', *J. Biomol. NMR*, **6**, 11–22 (1995).
52. R. Weisemann, F. Lohr, R. Ruterjans, 'HNCCH-TOCSY, a Triple Resonance Experiment for the Correlation of Backbone $^{13}C^{\alpha}$ and ^{15}N Resonances with Aliphatic Side-chain Proton Resonances and for Measuring Vicinal ^{13}CO, $^{1}H^{\beta}$ Coupling Constants in Proteins', *J. Biomol. NMR*, **4**, 587–593 (1994).
53. S. Grzesiek, J. Anglister, A. Bax, 'Correlation of Backbone Amide and Aliphatic Side-chain Resonances in ^{13}C/^{15}N-Enriched Proteins by Isotropic Mixing of ^{13}C Magnetization', *J. Magn. Reson. B.*, **101**, 114–119 (1993).
54. J. Keeler, R.T. Clowes, A.L. Davis, E.D. Laue, 'Pulsed-field Gradients: Theory and Practice', *Methods Enzymol.*, **239**, 145–207 (1994).
55. L.E. Kay, 'Pulsed Field Gradient Multi-dimensional NMR Methods for the Study of Protein Structure and Dynamics in Solution', *Prog. Biophys. Mol. Biol.*, **63**, 277–299 (1995).
56. L.E. Kay, 'Field Gradient Techniques in NMR Spectroscopy', *Curr. Opin. Struct. Biol.*, **5**, 674–681 (1995).
57. S.L. Campbell-Burk, P.J. Domaille, M.A. Starovasnik, W. Boucher, E.D. Laue, 'Sequential Assignment of the Backbone Nuclei, (^{1}H, ^{15}N, and ^{13}C) of c-H-ras p21 (1-166)-GDP using a Novel 4D NMR Strategy', *J. Biomol. NMR*, **2**, 639–646 (1992).
58. W. Boucher, E.D. Laue, S.L. Campbell-Burk, P.J. Domaille, 'Four-dimensional Heteronuclear Triple Resonance NMR Methods for the Assignment of Backbone Nuclei in Proteins', *J. Am. Chem. Soc.*, **114**, 2262–2264 (1992).
59. W. Boucher, E.D. Laue, S.L. Campbell-Burk, P.J. Domaille, 'Improved 4D NMR Experiments for the Assignment of Backbone Nuclei in ^{13}C/^{15}N Labeled Proteins', *J. Biomol. NMR*, **2**, 631–637 (1992).
60. L.E. Kay, M. Wittekind, M.A. McCoy, M.S. Friedrichs, L. Müller, '4D NMR Triple-resonance Experiments for Assignment of Protein Backbone Nuclei using Shared Constant-time Evolution Periods', *J. Magn. Reson.*, **98**, 443–450 (1992).
61. E.T. Olejniczak, R.X. Xu, A.M. Petros, S.W. Fesik, 'Optimized Constant-time 4D HNCAHA and HN(CO)-CAHA Experiments. Applications to the Backbone Assignments of the FKBP/Ascomycin Complex', *J. Magn. Reson.*, **100**, 444–450 (1992).
62. R.T. Clowes, W. Boucher, C.H. Hardman, P.J. Domaille, E.D. Laue, 'A 4D HCC(CO)NNH Experiment for the Correlation of Aliphatic Side-chain and Backbone Resonances in ^{13}C/^{15}N-Labeled Proteins', *J. Biomol. NMR*, **3**, 349–354 (1993).
63. S.W. Fesik, E.R.P. Zuiderweg, 'Heteronuclear Three-dimensional NMR Spectroscopy. A Strategy for the Simplification of Homonuclear Two-dimensional NMR Spectra', *J. Magn. Reson.*, **78**, 588–593 (1988).
64. B.H. Oh, W.M. Westler, P. Darba, J.L. Markley, 'Protein Carbon-13 Spin Systems by a Single Two-dimensional Nuclear Magnetic Resonance Experiment, *Science*, **240**, 908–911 (1988).
65. W.M. Westler, B.J.Y.H. Stockman, M. Kainosho, J.L. Markley, 'Correlation of Carbon-13 and Nitrogen-15 Chemical Shifts in Selectively and Uniformly Labeled Proteins by Heteronuclear Two-dimensional NMR Spectroscopy', *J. Am. Chem. Soc.*, **110**, 6256 (1988).
66. A. Bax, S. Grzesiek, 'Methodological Advances in Protein NMR', *Acc. Chem. Res.*, **26**, 131–138 (1993).

67. S. Grzesiek, J. Anglister, H.A.B. Ren, '^{13}C Line Narowing by ^{2}H Decoupling in ^{2}H/^{13}C/^{15}N-Enriched Proteins. Application to Triple Resonance 4D J Connectivity of Sequential Amides', *J. Am. Chem. Soc.*, **115**, 4369–4370 (1993).
68. J.L. Markley, I. Putter, O. Jardetzky, 'High-resolution Nuclear Magnetic Resonance Spectra of Selectively Deuterated Staphylococcal Nuclease', *Science*, **161**, 1249–1251 (1968).
69. H.L. Crespi, R.M. Rosenberg, J.J. Katz, 'Proton Magnetic Resonance of Proteins Fully Deuterated Except for ^{1}H-Leucine Side Chains', *Science*, **161**, 795–796 (1968).
70. D.M. LeMaster, 'Deuterium Labeling in NMR Structural Analysis of Larger Proteins', *Q. Rev. Biophys.*, **23**, 133–174 (1990).
71. D.M. LeMaster, F.M. Richards, 'NMR Sequential Assignment of *Escherichia coli* Thioredoxin Utilizing Random Fractional Deuteration', *Biochemistry*, **27**, 142–150 (1988).
72. C.H. Arrowsmith, R. Pachter, R.B. Altman, S.B. Iyer, O. Jardetzky, 'Sequence-specific ^{1}H NMR Assignments and Secondary Structure in Solution of *Escherichia coli* trp Repressor', *Biochemistry*, **29**, 6332–6341 (1990).
73. K. Pervushin, R. Riek, G. Wider, K. Wuthrich, 'Attenuated T_2 Relaxation by Mutual Cancellation of Dipole–Dipole Coupling and Chemical Shift Anisotropy Indicates an Avenue to NMR Structures of Very Large Biological Macromolecules in Solution', *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371 (1997).
74. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wuthrich, 'TROSY in Triple-resonance Experiments: New Perspectives for Sequential NMR Assignment of Large Proteins', *Proc. Natl. Acad. Sci. USA*, **95**, 13585–13590 (1998).
75. M. Salzmann, G. Wider, K. Pervushin, H. Senn, K. Wüthrich, 'TROSY-type Triple-resonance Experiments for Sequential NMR Assignments of Large Proteins', *J. Am. Chem. Soc.*, **121**, 844–848 (1999).
76. M. Karplus, 'Vicinal Proton Coupling in Nuclear Magnetic Resonance', *J. Am. Chem. Soc.*, **85**, 2870 (1963).
77. M. Karplus, 'Contact Electron-spin Coupling of Nuclear Magnetic Moments', *J. Chem. Phys.*, **30**, 11 (1959).
78. A. Pardi, M. Billeter, K. Wuthrich, 'Calibration of the Angular Dependence of the Amide Proton-C Alpha Proton Coupling Constants, 3JHN Alpha, in a Globular Protein. Use of 3JHN Alpha for Identification of Helical Secondary Structure, *J. Mol. Biol.*, **180**, 741–751 (1984).
79. A. DeMarco, M. Llinas, K. Wüthrich, '^{1}H-^{15}N Spin–Spin Coupling in Alumichrome', *Biopolymers*, **17**, 2727–2742 (1978).
80. G.T. Montelione, M.E. Winkler, P. Rauenbühler, G. Wagner, 'Accurate Measurements of Long-range Heteronuclear Coupling Constants from Heteronuclear 2D NMR Spectra of Isotope-enriched Proteins', *J. Magn. Reson.*, **82**, 198 (1989).
81. C. Griesinger, O.W. Sørensen, R.R. Ernst, 'Two-dimensional Correlation of Connected NMR Transitions', *J. Am. Chem. Soc.*, **107**, 6394–6396 (1985).
82. L. Müller, 'PECOSY, a Simple Alternative to ECOSY', *J. Magn. Reson.*, **72**, 191–196 (1987).
83. I.D. Kuntz, J.F. Thomason, C.M. Oshiro, 'Distance Geometry', *Methods Enzymol.*, **177**, 159–204 (1989).
84. T.F. Havel, 'An Evaluation of Computational Strategies for use in the Determination of Protein Structure from Distance Constraints obtained by Nuclear Magnetic Resonance', *Prog. Biophys. Mol. Biol.*, **56**, 43–78 (1991).
85. G.M. Crippen, 'A Novel Approach to Calculation of Conformation: Distance Geometry', *J. Comput. Phys.*, **24**, 96 (1977).
86. W. Braun, 'Distance Geometry and Related Methods for Protein Structure Determination from NMR Data', *Q. Rev. Biophys.*, **19**, 115–157 (1987).
87. J.A. McCammon, S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, 1987.
88. R.M. Scheek, W.F. van Gunsteren, R. Kaptein, 'Molecular Dynamics Simulation Techniques for Determination of Molecular Structures from Nuclear Magnetic Resonance Data', *Methods Enzymol.*, **177**, 204–218 (1989).
89. M. Nilges, G.M. Clore, A.M. Gronenborn, 'Determination of Three-dimensional Structures of Proteins from Interproton Distance Data by Dynamical Simulated Annealing from a Random Array of Atoms. Circumventing Problems Associated with Folding', *FEBS Lett.*, **239**, 129–136 (1988).
90. G.M. Clore, A.M. Gronenborn, 'New Methods of Structure Refinement for Macromolecular Structure Determination by NMR', *Proc. Natl. Acad. Sci. USA*, **95**, 5891–5898 (1998).
91. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.
92. R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, Vol. 1, 1980.
93. M.J.D. Powell, 'Restart Procedures for the Conjugate Gradient Method', *Math. Program.*, **12**, 241–254 (1977).
94. C. Gonzalez, J.A.C. Rullman, A.M.J.J. Bonvin, R. Boelens, R. Kaptein, 'Toward an NMR *R* Factor', *J. Magn. Reson.*, **91**, 659–664 (1991).
95. A.T. Brunger, G.M. Clore, A.M. Gronenborn, R. Saffrich, M. Nilges, 'Assessing the Quality of Solution Nuclear Magnetic Resonance Structures by Complete Cross-validation', *Science*, **261**, 328–331 (1993).
96. B.L. Grasberger, A.M. Gronenborn, G.M. Clore, 'Analysis of the Backbone Dynamics of Interleukin-8 by ^{15}N Relaxation Measurements', *J. Mol. Biol.*, **230**, 364–372 (1993).
97. G. Barbato, M. Ikura, L.E. Kay, R.W. Pastor, A. Bax, 'Backbone Dynamics of Calmodulin Studied by ^{15}N Relaxation Using Inverse Detected Two-dimensional

NMR Spectroscopy: The Central Helix is Flexible', *Biochemistry*, **31**, 5269–5278 (1992).
98. L.E. Kay, D.A. Torchia, A. Bax, 'Backbone Dynamics of Proteins as Studied by ^{15}N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease', *Biochemistry*, **28**, 8972–8979 (1989).
99. N.A. Farrow, R. Muhandiram, A.U. Singer, S.M. Pascal, C.M. Kay, G. Gish, S.E. Shoelson, T. Pawson, J.D. Forman-Kay, L.E. Kay, 'Backbone Dynamics of a Free and Phosphopeptide-complexed Src Homology 2 Domain Studied by ^{15}N NMR Relaxation', *Biochemistry*, **33**, 5984–6003 (1994).
100. D. Yang, A. Mittermaier, Y.K. Mok, L.E. Kay, 'A Study of Protein Side-chain Dynamics from New ^{2}H Autocorrelation and ^{13}C Cross-correlation NMR Experiments: Application to the N-Terminal SH3 Domain from drk', *J. Mol. Biol.*, **276**, 939–954 (1998).
101. V.A. Daragan, K.H. Mayo, 'A Simple Approach to Analyzing Protein Side-chain Dynamics from ^{13}C NMR Relaxation Data', *J. Magn. Reson.*, **130**, 329–334 (1998).
102. J. Herzfeld, A.E. Berger, 'Sideband Intensities in NMR Spectra of Samples Spinning at the Magic Angle', *J. Chem. Phys.*, **73**, 6021–6030 (1980).
103. S.O. Smith, O.B. Peersen, 'Solid-state NMR Approaches for Studying Membrane Protein Structure', *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 25–47 (1992).
104. N. Tjandra, A. Bax, 'Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium, *Science*, **278**, 1111–1114 (1997).
105. N. Tjandra, A. Bax, 'Measurement of Dipolar Contributions to 1JCH Splittings from Magnetic-field Dependence of J Modulation in Two-dimensional NMR Spectra', *J. Magn. Reson.*, **124**, 512–515 (1997).
106. N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, 'Use of Dipolar ^{1}H-^{15}N and ^{1}H-^{13}C Couplings in the Structure Determination of Magnetically Oriented Macromolecules in Solution', *Nat. Struct. Biol.*, **4**, 732–738 (1997).
107. M. Ottiger, F. Delaglio, A. Bax, 'Measurement of J and Dipolar Couplings from Simplified Two-dimensional NMR Spectra', *J. Magn. Reson.*, **131**, 373–378 (1998).
108. G.M. Clore, A.M. Gronenborn, N. Tjandra, 'Direct Structure Refinement Against Residual Dipolar Couplings in the Presence of Rhombicity of Unknown Magnitude', *J. Magn. Reson.*, **131**, 159–162 (1998).
109. C.R. Sanders, J.P. Schwonek, 'Characterization of Magnetically Orientable Bilayers in Mixtures of Dihexanoylphosphatidylcholine and Dimyristoylphosphatidylcholine by Solid-state NMR', *Biochemistry*, **31**, 8898–8905 (1992).
110. M.R. Hansen, L. Mueller, A. Pardi, 'Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions', *Nat. Struct. Biol.*, **5**, 1065–1074 (1998).
111. G.M. Clore, M.R. Starich, A.M. Gronenborn, 'Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-shaped Viruses', *J. Am. Chem. Soc.*, **120**, 10571–10572 (1998).
112. J. Sass, F. Cordier, A. Hoffmann, M. Rogowski, A. Cousin, J.G. Omichinski, H. Löwen, S. Grzesiek, 'Purple Membrane Induced Alignment of Biological Macromolecules in the Magnetic Field', *J. Am. Chem. Soc.*, **121**, 2047–2055 (1999).
113. G.M. Clore, A.M. Gronenborn, A. Bax, 'A Robust Method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information', *J. Magn. Reson.*, **133**, 216–221 (1998).
114. M. Cai, Y. Huang, R. Zheng, S.Q. Wei, R. Ghirlando, M.S. Lee, R. Craigie, A.M. Gronenborn, G.M. Clore, 'Solution Structure of the Cellular Factor BAF Responsible for Protecting Retroviral DNA from Autointegration', *Nat. Struct. Biol.*, **5**, 903–909 (1998).
115. C.A. Bewley, K.R. Gustafson, M.R. Boyd, D.G. Covell, A. Bax, G.M. Clore, A.M. Gronenborn, 'Solution Structure of Cyanovirin-N, a Potent HIV-inactivating Protein', *Nat. Struct. Biol.*, **5**, 571–578 (1998).
116. D.S. Garrett, Y.J. Seok, A. Peterkofsky, A.M. Gronenborn, G.M. Clore, 'Solution Structure of the 40 000 M_r Phosphoryl Transfer Complex Between the N-Terminal Domain of Enzyme I and HPr', *Nat. Struct. Biol.*, **6**, 166–173 (1999).
117. A.C. Drohat, N. Tjandra, D.M. Baldisseri, D.J. Weber, 'The Use of Dipolar Couplings for Determining the Solution Structure of Rat apo-S100B($\beta\beta$)', *Prot. Sci.*, **8**, 800–809 (1999).
118. J.M. Griffiths, R.G. Griffin, 'Nuclear Magnetic Resonance Methods for Measuring Dipolar Couplings in Rotating Solids', *Anal. Chim. Acta*, **283**, 1081–1101 (1993).
119. D.P. Raleigh, M.H. Levitt, R.G. Griffin, 'Rotational Resonance in Solid-state NMR', *Chem. Phys. Lett.*, **146**, 71–76 (1988).
120. D.P. Raleigh, F. Creuzet, S.K.D. Gupta, M.H. Levitt, R.G. Griffin, 'Measurement of Internuclear Distances in Polycrystalline Solids–Rotationally Enhanced Transfer of Nuclear-spin Magnetization', *J. Am. Chem. Soc.*, **111**, 4502–4503 (1989).
121. E.R. Andrew, A. Bradbury, R.G. Eades, V.T. Wynn, 'Nuclear Cross-relaxation Induced by Specimen Rotation', *Phys. Lett.*, **4**, 99 (1963).
122. M.G. Columbo, B.H. Meier, R.R. Ernst, 'Rotor-driven Spin Diffusion in Natural-abundance ^{13}C Spin Systems', *Chem. Phys. Lett.*, **146**, 189 (1988).
123. W.E.J.R. Maas, W.S. Veeman, 'Natural Abundance ^{13}C Spin Diffusion Enhanced by Magic-angle Spinning', *Chem. Phys. Lett.*, **149**, 170 (1988).
124. W.S. Veeman, 'Carbon-13 Chemical Shift Anisotropy', *Prog. NMR Spectrosc.*, **16**, 193 (1984).

125. M.H. Levitt, D.P. Raleigh, F. Creuzet, R.G. Griffin, 'Theory and Simulations of Homonuclear Spin Pair Systems in Rotating Solids', *J. Chem. Phys.*, **92**, 6347–6364 (1990).
126. A.E. Bennett, J.H. Ok, R.G. Griffin, S. Vega, 'Chemical-shift Correlation Spectroscopy in Rotating Solids–Radio Frequency-driven Dipolar Recoupling and Longitudinal Exchange', *J. Chem. Phys.*, **96**, 8624–8627 (1992).
127. T. Gullion, S. Vega, 'A Simple Magic Angle Spinning NMR Experiment for the Dephasing of Rotational Echoes of Dipolar Coupled Homonuclear Spin Pairs', *Chem. Phys. Lett.*, **194**, 423–428 (1992).
128. R. Tycko, G. Dabbagh, 'Measurement of Nuclear Magnetic Dipole-dipole Couplings in Magic Angle Spinning NMR', *Chem. Phys. Lett.*, **173**, 461–465 (1990).
129. R. Tycko, G. Dabbagh, 'Double-quantum Filtering in Magic-angle Spinning NMR Spectroscopy: An Approach to Spectral Simplification and Molecular Structure Determination', *J. Am. Chem. Soc.*, **113**, 9444–9448 (1991).
130. R. Tycko, S.O. Smith, 'Symmetry Principles in the Design of Pulse Sequences for Structural Measurements in Magic-angle Spinning Nuclear Magnetic Resonance', *J. Chem. Phys.*, **98**, 932–943 (1993).
131. B.-Q. Sun, P.R. Costa, D. Kocisko, P.T. Lansbury, Jr, R.G. Griffin, 'Internuclear Distance Measurements in Solid State Nuclear Magnetic Resonance: Dipolar Recoupling via Rotor Synchronized Spin-locking', *J. Chem. Phys.*, **102**, 702–707 (1995).
132. D.M. Gregory, D.J. Mitchell, J.A. Stringer, S. Kiihne, J.C. Shiels, J. Callahan, M.A. Mehta, G.P. Drobny, 'Windowless Dipolar Recoupling: The Detection of Weak Dipolar Couplings Between Spin 1/2 Nuclei with Large Chemical Shift Anisotropies', *Chem. Phys. Lett.*, **246**, 654–663 (1995).
133. J.M. Joers, R. Rosanke, T. Gullion, J.R. Garbow, 'Detection of Dipolar Interactions by CROWN NMR', *J. Magn. Reson. A*, **106**, 123–126 (1994).
134. N.C. Nielsen, H. Bildsøe, H.J. Jakobsen, M.H. Levitt, 'Double-quantum Homonuclear Rotary Resonance: Efficient Dipolar Recovery in Magic-angle Spinning Nuclear Magnetic Resonance', *J. Chem. Phys.*, **101**, 1805 (1994).
135. Y.K. Lee, N.D. Kurar, M. Helmle, O.G. Johannessen, N.C. Nielsen, M.H. Levitt, 'Efficient Dipolar Recoupling in the NMR of Rotating Solids. A Sevenfold Symmetric Radiofrequency Pulse Sequence', *Chem. Phys. Lett.*, **242**, 304–309 (1995).
136. M. Hohwy, H.J. Jakobsen, M. Edén, M.H. Levitt, N.C. Nielsen, 'Broadband Dipolar Recoupling in the Nuclear Magnetic Resonance of Rotating Solids: A Compensated C7 Pulse Sequence', *J. Chem. Phys.*, **108**, 2686–2694 (1998).
137. A.E. McDermott, F. Creuzet, R. Gebhard, K. Vanderhoef, M.H. Levitt, J. Herzfeld, J. Lugtenburg, R.G. Griffin, 'Determination of Internuclear Distances and the Orientation of Functional-groups by Solid-state NMR–Rotational Resonance Study of the Conformation of Retinal in Bacteriorhodopsin', *Biochemistry*, **33**, 6129–6136 (1994).
138. Y. Tomita, E.J. Oconnor, A. McDermott, 'A Method for Dihedral Angle Measurement in Solids – Rotational Resonance NMR of a Transition-state Inhibitor of Triose Phosphate Isomerase', *J. Am. Chem. Soc.*, **116**, 8766–8771 (1994).
139. D.P. Weliky, R. Tycko, 'Determination of Peptide Conformations by Two-dimensional Magic-angle Spinning NMR Exchange Spectroscopy', *J. Am. Chem. Soc.*, **118**, 8487–8488 (1996).
140. G.P. Drobny, *Experimental NMR Conference*, Asilomar, CA, March 17–22, 1996.
141. K. Schmidt-Rohr, 'Torsion Angle Determination in Solid ^{13}C-Labeled Amino Acids and Peptides by Separated-local-field Double-quantum NMR', *J. Am. Chem. Soc.*, **118**, 7601–7603 (1996).
142. X. Feng, Y.K. Lee, D. Sandström, M. Edén, H. Maisel, A. Sebald, M.H. Levitt, 'Direct Determination of a Molecular Torsional Angle by Solid-state NMR', *Chem. Phys. Lett.*, **257**, 314–320 (1996).
143. X. Feng, M. Edén, A. Brinkmann, H. Luthman, L. Eriksson, A. Gräslund, O.N. Antzutkin, M.H. Levitt, 'Direct Determination of a Peptide Torsion Angle ψ by Double-quantum Solid-state NMR', *J. Am. Chem. Soc.*, **119**, 12006–12007 (1997).
144. X. Feng, P.J.E. Verdegem, Y.K. Lee, D. Sandström, M. Edén, P. Bovee-Geurts, W.J. de Grip, J. Lugtenburg, H.J.M. de Groot, M.H. Levitt, 'Direct Determination of a Molecular Torsional Angle in the Membrane Protein Rhodopsin by Solid-state NMR', *J. Am. Chem. Soc.*, **119**, 6853–6857 (1997).
145. G.R. Marshall, D.D. Beusen, K. Kociolek, A.S. Redlinski, M.T. Leplawy, Y. Pan, J. Schaefer, 'Determination of a Precise Interatomic Distance in a Helical Peptide by REDOR NMR', *J. Am. Chem. Soc.*, **112**, 963–966 (1990).
146. T. Gullion, J. Schaefer, 'Rotational-echo Double-resonance NMR', *J. Magn. Reson.*, **81**, 196–200 (1989).
147. T. Gullion, J. Schaefer, 'Detection of Weak Heteronuclear Dipolar Coupling by Rotational-echo Double-resonance Nuclear Magnetic Resonance', *Adv. Magn. Reson.*, **13**, 57–83 (1989).
148. A. Schmidt, R.A. McKay, J. Schaefer, 'Internuclear Distance Measurement Between Deuterium (I = 1) and a Spin-1/2 Nucleus in Rotating Solids', *J. Magn. Reson.*, **96**, 644–650 (1992).
149. C.P. Grey, W.S. Veeman, A.J. Vega, 'Rotational Echo N-14/C-13/H-1 Triple-resonance Solid-state Nuclear Magnetic Resonance – a Probe of C-13–N-14 Internuclear Distances', *J. Chem. Phys.*, **98**, 7711–7724 (1993).
150. C.P. Grey, A. Eijkelenboom, W.S. Veeman, 'N-14 Population Transfers in 2-Dimensional C-13-N-14-H-1 Triple-resonance Magic-angle Spinning Nuclear

Magnetic Resonance Spectroscopy', *Solid State Nucl. Magn. Reson.*, **4**, 113–120 (1995).
151. C.P. Grey, W.S. Veeman, 'The Detection of Weak Heteronuclear Coupling between Spin-1 and Spin-1/2 Nuclei in Mas NMR – N-14/C-13/H-1 Triple Resonance Experiments', *Chem. Phys. Lett.*, **192**, 379–385 (1992).
152. T. Gullion, 'Measurement of Dipolar Interactions Between Spin-1/2 and Quadrupolar Nuclei by Rotational-echo, Adiabatic-passage, Double-resonance NMR', *Chem. Phys. Lett.*, **246**, 325–330 (1995).
153. T. Gullion, 'Detecting C-13–O-17 Dipolar Interactions by Rotational-echo, Adiabatic-passage, Double-resonance NMR', *J. Magn. Reson. A*, **117**, 326–329 (1995).
154. S.M. Holl, G.R. Marshall, D.D. Beusen, K. Kociolek, A.S. Redlinski, M.T. Leplawy, R.A. McKay, S. Vega, J. Schaefer, 'Determination of an 8-Angstrom Interatomic Distance in a Helical Peptide by Solid-state NMR-spectroscopy', *J. Am. Chem. Soc.*, **114**, 4830–4833 (1992).
155. K.T. Mueller, T.P. Jarvie, D.J. Aurentz, B.W. Roberts, 'The REDOR Transform – Direct Calculation of Internuclear Couplings from Dipolar-dephasing NMR Data', *Chem. Phys. Lett.*, **242**, 535–542 (1995).
156. K.T. Mueller, T.P. Jarvie, D.J. Aurentz, B.W. Roberts, 'The REDOR Transform – Direct Calculation of Internuclear Couplings from Dipolar-dephasing NMR Data (Vol. 242, Pg. 535, 1996)', *Chem. Phys. Lett.*, **254**, 281–282 (1996).
157. T.P. Jarvie, G.T. Went, K.T. Mueller, 'Simultaneous Multiple Distance Measurements in Peptides via Solid-state NMR', *J. Am. Chem. Soc.*, **118**, 5330–5331 (1996).
158. Y. Li, J.N.S. Evans, 'Determination of a Distance between 2 Heteroatoms by the Rotational-echo Double-resonance Experiment with Total Suppression of Side-band', *J. Magn. Reson. A*, **109**, 256–259 (1994).
159. Y. Li, J.N.S. Evans, 'Theoretical-study of Rotational-echo Double-resonance and Related Experiments', *J. Chem. Phys.*, **101**, 10211–10216 (1994).
160. T.I. Igumenova, D.J. Mitchell, J.N.S. Evans, 'Comparative Studies of Phase-cycling Schemes for Multiple Pi-pulse Sequences', *J. Magn. Reson.*, **127**, 144–146 (1997).
161. Y. Li, J.N.S. Evans, 'A New Multiple π-Pulse Sequence in Solid-state NMR', (correction to *Chem. Phys. Lett.*, **241**, 79–83 (1995)), *Chem. Phys. Lett.*, **246**, 527–528 (1995).
162. Y. Li, J.N.S. Evans, 'A New Multiple π-Pulse Sequence in Solid-state NMR', *Chem. Phys. Lett.*, **241**, 79–83 (1995).
163. Y. Li, J.N.S. Evans, 'The Importance of Xy-8 Phase Cycling in the Rotational-echo Double-resonance Experiment with Total Side-band Suppression', *J. Magn. Reson. A*, **116**, 150–155 (1995).
164. Y. Li, J.N.S. Evans, 'Multiple-π-pulse Sequences in Biological Solid-state NMR – Enhancement of Sensitivity by Permutation of the Phases', *J. Magn. Reson. B*, **111**, 296–299 (1996).
165. D.J. Mitchell, J.N.S. Evans, 'Improved Heteronuclear Decoupling in REDOR with the use of TPPM', *Chem. Phys. Lett.*, **292**, 656–659 (1998).
166. D.J. Mitchell, D.L. Jakeman, T.I. Igumenova, W.A. Shuttleworth, K.D. Miller, J.N.S. Evans, 'Time-resolved Solid-state REDOR-edited NMR Detection of a Transient Enzyme-intermediate', *Chem. Commun.*, 1019–1020 (1997).
167. S.J. Opella, P.L. Stewart, K.G. Valentine, 'Protein Structure by Solid-state NMR Spectroscopy', *Q. Rev. Biophys.*, **19**, 7–49 (1987).
168. T.A. Cross, S.J. Opella, 'Protein Structure by Solid-state NMR', *J. Am. Chem. Soc.*, **105**, 306–308 (1983).
169. T.A. Cross, 'Solid-state Nuclear Magnetic Resonance Approach for Determining the Structure of Gramicidin A without Model Fitting', *Biophys. J.*, **49**, 124–126 (1986).
170. Q. Teng, L.K. Nicholson, T.A. Cross, 'Experimental Determination of Torsion Angles in the Polypeptide Backbone of the Gramicidin A Channel by Solid State Nuclear Magnetic Resonance', *J. Mol. Biol.*, **218**, 607–619 (1991).
171. L.K. Nicholson, Q. Teng, T.A. Cross, 'Solid-state Nuclear Magnetic Resonance Derived Model for Dynamics in the Polypeptide Backbone of the Gramicidin A Channel', *J. Mol. Biol.*, **218**, 621–637 (1991).
172. S.J. Opella, P.L. Stewart, 'Solid-state Nuclear Magnetic Resonance Structural Studies of Proteins', *Methods Enzymol.*, **176**, 242–275 (1989).
173. S.J. Opella, 'Solid-state NMR of Biological Systems', *Annu. Rev. Phys. Chem.*, **33**, 533–562 (1982).
174. S. Pausak, A. Pines, J.S. Waugh, 'Carbon-13 Chemical Shielding Tensors in Single-crystal Durene', *J. Chem. Phys.*, **59**, 591–595 (1973).
175. P. Lauterbur, 'Anisotropy of the ^{13}C Chemical Shift in Calcite', *Phys. Rev. Lett.*, **1**, 343–344 (1958).
176. R.G. Griffin, J.D. Ellet, Jr, M. Mehring, J.G. Bullitt, J.S. Waugh, 'Single Crystal Study of the ^{19}F Shielding Tensors of a Trifluoromethyl Group', *J. Chem. Phys.*, **57**, 2147–2155 (1972).
177. C.J. Hartzell, T. Pratum, G.P. Drobny, 'Mutual Orientation of Three Magnetic Tensors in a Polycrystalline Dipeptide by Dipole-modulated ^{15}N Chemical Shift Spectroscopy', *J. Chem. Phys.*, **87**, 4324–4331 (1987).
178. C.J. Hartzell, M. Whitfield, T.G. Oas, G.P. Drobny, 'Determination of the ^{15}N and ^{13}C Chemical Shift Tensors of L-[^{13}C]Alanyl-L-[^{15}N]-alanine from the Dipole-coupled Powder Patterns', *J. Am. Chem. Soc.*, **109**, 5966–5969 (1987).
179. T.G. Oas, C.J. Hartzell, F.W. Dahlquist, G.P. Drobny, 'The Amide ^{15}N Chemical Shift Tensors of Four Peptides Determined from ^{13}C Dipole-coupled Chemical Shift Powder Patterns', *J. Am. Chem. Soc.*, **109**, 5962–5966 (1987).

180. T.G. Oas, C.J. Hartzell, T.J. McMahon, G.P. Drobny, F.W. Dahlquist, 'The Carbonyl ^{13}C Chemical Shift Tensors of Five Peptides Determined from ^{15}N Dipole-coupled Chemical Shift Powder Patterns', *J. Am. Chem. Soc.*, **109**, 5962–5966 (1987).
181. J.R. Huth, C.A. Bewley, M.S. Nissen, J.N. Evans, R. Reeves, A.M. Gronenborn, G.M. Clore, 'The Solution Structure of an HMG-I(Y)–DNA Complex Defines a new Architectural Minor Groove Binding Motif', *Nat. Struct. Biol.*, **4**, 657–665 (1997).
182. A.M. Gronenborn, G.M. Clore, 'Structures of Protein Complexes by Multidimensional Heteronuclear Magnetic Resonance Spectroscopy', *Crit. Rev. Biochem. Mol. Biol.*, **30**, 351–385 (1995).
183. G.W. Vuister, S.-J. Kim, C. Wu, A. Bax, '2D and 3D NMR Study of Phenylalanine Residues in Proteins by Reverse Isotopic Labeling', *J. Am. Chem. Soc.*, **116**, 9206–9210 (1994).
184. M.H. Werner, M.E. Bianchi, A. Gronenborn, G.M. Clore, 'NMR Spectroscopic Analysis of the DNA Conformation Induced by the Human Testis Determining Factor SRY', *Biochemistry*, **34**, 11998–12004 (1995).
185. G. Otting, K. Wüthrich, 'Heteronuclear Filters in Two-dimensional [^1H,^1H]-NMR Spectroscopy: Combined use with Isotope Labeling for Studies of Macromolecular Conformation and Intermolecular Interactions', *Rev. Biophys.*, **23**, 39–96 (1990).
186. S.W. Fesik, R.T. Gampe, Jr, H.L. Eaton, G. Gemmecker, E.T. Olejniczak, P. Neri, T.F. Holzman, D.A. Egan, R. Edalji, R. Simmer et al., 'NMR Studies of [U-^{13}C]-cyclosporin A Bound to Cyclophilin: Bound Conformation and Portions of Cyclosporin Involved in Binding', *Biochemistry*, **30**, 6574–6583 (1991).
187. M. Ikura, G.M. Clore, A.M. Gronenborn, G. Zhu, C.B. Klee, A. Bax, 'Solution Structure of the Calmodulin-target Peptide Complex by Multidimensional NMR', *Science*, **256**, 632–638 (1992).
188. M. Ikura, A. Bax, 'Isotope Filtered 2D NMR of a Protein–Peptide Complex. Study of a Skeletal Muscle Myosin Light Chain Kinase Fragment Bound to Calmodulin', *J. Am. Chem. Soc.*, **114**, 2344–2440 (1992).
189. L.E. Kay, G.M. Clore, A. Bax, A.M. Gronenborn, 'Four-dimensional Heteronuclear Triple-resonance NMR Spectroscopy of Interleukin-1 beta in Solution', *Science*, **249**, 411–414 (1990).
190. E.R.P. Zuiderweg, A.M. Petros, S.W. Fesik, E.T. Olejniczak, 'Four-dimensional [^{13}C, ^1H, ^{13}C, ^1H]HMQC/NOE/HMQC NMR Spectroscopy: Resolving Tertiary NOE Distance Restraints in Spectra of Larger Proteins', *J. Am. Chem. Soc.*, **113**, 370–372 (1991).
191. G.W. Vuister, G.M. Clore, A.M. Gronenborn, R. Powers, D.S. Garrett, R. Tschudin, A. Bax, 'Increased Resolution and Improved Quality in Four-dimensional $^{13}C/^{13}C$ Separated HMQC/NOE/HMQC Spectra Using Pulse Field Gradients', *J. Magn. Reson. B*, **101**, 210–213 (1993).
192. J.N.S. Evans, *NMR and Enzymes, in Pulsed Magnetic Resonance: NMR, ESR and Optics (A Recognition of E.L. Hahn)*, Oxford University Press, Oxford, 1992.
193. P.N. Barlow, R.J. Appleyard, B.J.O. Wilson, J.N.S. Evans, 'Direct Observation of the Enzyme Intermediate Complex of 5-Enolpyruvylshikimate-3-phosphate Synthase by C-13 NMR-spectroscopy', *Biochemistry*, **28**, 7985–7991 and correction p. 10093 (1989).
194. K.S. Anderson, R.D. Sammons, G.C. Leo, J.A. Sikorski, A.J. Benesi, K.A. Johnson, 'Observation by C-13 NMR of the EPSP Synthase Tetrahedral Intermediate Bound to the Enzyme Active-site', *Biochemistry*, **29**, 1460–1465 (1990).
195. W.A. Shuttleworth, M.E. Pohl, G.L. Helms, D.L. Jakeman, J.N.S. Evans, 'Site-directed Mutagenesis of the Putative Active Site Residues of EPSP Synthase', *Biochemistry*, **38**, 296–302 (1999).
196. W.A. Shuttleworth, C.D. Hough, K.P. Bertrand, J.N.S. Evans, 'Over-production of 5-Enolpyruvylshikimate-3-phosphate Synthase in *Escherichia coli*–Use of the T7-Promoter', *Prot. Eng.*, **5**, 461–466 (1992).
197. W.A. Shuttleworth, J.N.S. Evans, 'Site-directed Mutagenesis and NMR-studies of Histidine-385 Mutants of 5-Enolpyruvylshikimate-3-phosphate Synthase', *Biochemistry*, **33**, 7062–7068 (1994).
198. W.A. Shuttleworth, J.N.S. Evans, 'The H385N Mutant of 5-Enolpyruvylshikimate-3-phosphate Synthase–Kinetics, Fluorescence, and Nuclear-magnetic-resonance Studies', *Arch. Biochem. Biophys.*, **334**, 37–42 (1996).
199. Y. Li, R.J. Appleyard, W.A. Shuttleworth, J.N.S. Evans, 'Time-resolved Solid-state REDOR NMR measurements on 5-Enolpyruvylshikimate-3-phosphate Synthase', *J. Am. Chem. Soc.*, **116**, 10799–10800 (1994).
200. D.L. Jakeman, D.J. Mitchell, W.A. Shuttleworth, J.N.S. Evans, 'On the Mechanism of EPSP Synthase', *Biochemistry*, **37**, 12012–12019 (1998).
201. J.N.S. Evans, R.J. Appleyard, W.A. Shuttleworth, 'Time-resolved Solid-state NMR: Small Molecules and Enzymes in Rapidly Frozen Solution', *Bull. Magn. Reson.*, **14**, 81–85 (1992).
202. J.N.S. Evans, R.J. Appleyard, W.A. Shuttleworth, 'Detection of an Enzyme Intermediate Complex by Time-resolved Solid-state NMR Spectroscopy', *J. Am. Chem. Soc.*, **115**, 1588–1590 (1993).
203. J.N.S. Evans, 'Time-resolved Solid-state NMR of Enzyme–Substrate Interactions', in *Encyclopedia of NMR Volume 7*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 4757–4763, 1995.
204. R.J. Appleyard, J.N.S. Evans, 'Solid-state NMR of Glycine in Frozen Solution', *J. Magn. Reson. B*, **102**, 245–252 (1993).
205. R.J. Appleyard, W.A. Shuttleworth, J.N.S. Evans, 'Time-resolved Solid-state NMR spectroscopy of 5-Enolpyruvylshikimate-3-phosphate Synthase', *Biochemistry*, **33**, 6812–6821 (1994).

206. C. Ramilo, R.J. Appleyard, C. Wanke, F. Krekel, N. Amrhein, J.N.S. Evans, 'Detection of the Covalent Intermediate of UDP-N-Acetylglucosamine Enolpyruvyl Transferase by Solution-state and Time-resolved Solid-state NMR Spectroscopy', *Biochemistry*, **33**, 15071–15079 (1994).
207. O. Fennema, 'Proteins at Low Temperatures', *Adv. Chem. Ser.*, **180**, (1979).
208. F. Franks, *Biophysics and Biochemistry at Low Temperatures*, Cambridge University Press, Cambridge, 1985.
209. L. Bachmann, E. Mayer, in *Cryotechniques in Biological Electron Microscopy*, eds. R.A. Stenbrecht, K. Zierold, Springer-Verlag, Berlin, 1987.
210. K.S. Anderson, J.A. Sikorski, K.A. Johnson, 'A Tetrahedral Intermediate in the EPSP Synthase Reaction Observed by Rapid Quench Kinetics', *Biochemistry*, **27**, 7395–7406 (1988).
211. Y. Li, F. Krekel, C.A. Ramilo, N. Amrhein, J.N.S. Evans, 'Time-resolved Solid-state REDOR NMR Studies of UDP N-Acetylglucosamine Enolpyruvyl Transferase', *FEBS Lett.*, **377**, 208–212 (1995).
212. D.R. Studelska, L.M. McDowell, M.P. Espe, C.A. Klug, J. Schaefer, 'Slowed Enzymatic Turnover Allows Characterization of Intermediates by Solid-state NMR', *Biochemistry*, **36**, 15555–15560 (1997).
213. W.C. Stallings, S.S. Abdelmeguid, L.W. Lim, H.S. Shieh, H.E. Dayringer, N.K. Leimgruber, R.A. Stegeman, K.S. Anderson, J.A. Sikorski, S.R. Padgette, G.M. Kishore, 'Structure and Topological Symmetry of the Glyphosate Target 5-Enol-Pyruvylshikimate-3-phosphate Synthase–a Distinctive Protein Fold', *Proc. Natl. Acad. Sci. USA*, **88**, 5046–5050 (1991).
214. D.R. Studelska, C.A. Klug, D.D. Beusen, L.M. McDowell, J. Schaefer, 'Long-range Distance Measurements of Protein Binding Sites by Rotational-echo Double-resonance NMR', *J. Am. Chem. Soc.*, **118**, 5476–5477 (1996).
215. L.M. McDowell, C.A. Klug, D.D. Beusen, J. Schaefer, 'Ligand Geometry of the Ternary Complex of 5-Enolpyruvylshikimate-3-phosphate Synthase From Rotational-echo Double-resonance NMR', *Biochemistry*, **35**, 5395–5403 (1996).
216. L.M. McDowell, A. Schmidt, E.R. Cohen, D.R. Studelska, J. Schaefer, 'Structural Constraints on the Ternary Complex of 5-Enolpyruvylshikimate-3-phosphate Synthase from Rotational-echo Double-resonance NMR', *J. Mol. Biol.*, **256**, 160–171 (1996).
217. L.M. McDowell, J. Schaefer, 'High-resolution NMR of Biological Solids', *Curr. Opin. Struct. Biol.*, **6**, 624–629 (1996).
218. A.M. Christensen, J. Schaefer, 'Solid-state NMR Determination of Intramolecular and Intermolecular P-31–C-13 Distances for Shikimate 3-Phosphate and [1-C-13]Glyphosate Bound to Enolpyruvylshikimate-3-phosphate Synthase', *Biochemistry*, **32**, 2868–2873 (1993).
219. J.H. Davis, 'The Description of Membrane Lipid Conformation, Order and Dynamics by 2H-NMR', *Biochim. Biophys. Acta*, **737**, 117–171 (1983).
220. I.C.P. Smith, (ed.), *Structure and Properties of Cell Membranes*, CRC Press, Boca Raton, Florida, Vol. III, 1985.
221. S.J. Opella, 'NMR and Membrane Proteins', *Nat. Struct. Biol.*, **4** Suppl, 845–848 (1997).
222. B.A. Wallace, 'Gramicidin Channels and Pores', *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 127–157 (1990).
223. D.A. Langs, 'Three-dimensional Structure at 0.86 Å of the Uncomplexed Form of the Transmembrane Ion Channel Peptide Gramicidin A', *Science*, **241**, 188–191 (1988).
224. B.A. Wallace, K. Ravikumar, 'The Gramicidin Pore: Crystal Structure of a Cesium Complex', *Science*, **241**, 182–187 (1988).
225. F.D. Moll, T.A. Cross, 'Optimizing and Characterizing Alignment of Oriented Lipid Bilayers Containing Gramicidin D', *Biophys. J.*, **57**, 351–362 (1990).
226. L.K. Nicholson, T.A. Cross, 'Gramicidin Cation Channel: An Experimental Determination of the Right-handed Helix Sense and Verification of Beta-type Hydrogen Bonding', *Biochemistry*, **28**, 9379–9385 (1989).
227. L.K. Nicholson, Q. Teng, T.A. Cross, 'Solid-state Nuclear Magnetic Resonance Derived Model for Dynamics in the Polypeptide Backbone of the Gramicidin A Channel', *J. Mol. Biol.*, **218**, 621–637 (1991).
228. T.A. Cross, S.J. Opella, 'Protein Structure by Solid State Nuclear Magnetic Resonance. Residues 40 to 45 of Bacteriophage fd Coat Protein', *J. Mol. Biol.*, **182**, 367–381 (1985).
229. M. Munowitz, R.G. Griffin, G. Bodenhausen, T.H. Huang, 'Two-dimensional Rotational Spin-echo Nuclear Magnetic Resonance in Solids: Correlation of Chemical Shift and Dipolar Interactions', *J. Am. Chem. Soc.*, **103**, 2529–2533 (1981).
230. M. Munowitz, W.P. Aue, R.G. Griffin, 'Two-dimensional Separation of Dipolar and Scaled Isotropic Chemical Shift Interactions in Magic Angle NMR Spectra', *J. Chem. Phys.*, **77**, 1686–1689 (1982).
231. A.C. Kolbert, M.H. Levitt, R.G. Griffin, 'An Improved Approach to Two-dimensional Dipolar-chemical-shift Spectroscopy in Rotating Solids', *J. Magn. Reson.*, **85**, 42–49 (1989).
232. Q. Teng, T.A. Cross, 'The In Situ Determination of the ^{15}N Chemical-shift Tensor Orientation in a Polypeptide', *J. Magn. Reson.*, **85**, 439–447 (1989).
233. A.W. Hing, J. Schaefer, 'Two-dimensional Rotational-echo Double Resonance of Val1-[1-^{13}C]Gly2-[^{15}N]Ala3-gramicidin A in Multilamellar Dimyristoylphosphatidylcholine Dispersions', *Biochemistry*, **32**, 7593–7604 (1993).
234. R.R. Ketchem, W. Hu, T.A. Cross, 'High-resolution Conformation of Gramicidin A in a Lipid Bilayer by Solid-state NMR', *Science*, **261**, 1457–1460 (1993).

235. W. Hu, K.C. Lee, T.A. Cross, 'Tryptophans in Membrane Proteins: Indole Ring Orientations and Functional Implications in the Gramicidin Channel', *Biochemistry*, **32**, 7035–7047 (1993).
236. R.R. Ketchem, K.C. Lee, S. Huo, T.A. Cross, 'Macromolecular Structural Elucidation with Solid-state NMR-derived Orientational Constraints', *J. Biomol. NMR*, **8**, 1–14 (1996).
237. R.R. Ketchem, B. Roux, T.A. Cross, 'High-resolution Polypeptide Structure in a Lamellar Phase Lipid Environment from Solid State NMR Derived Orientational Constraints', *Structure*, **5**, 1655–1669 (1997).

Raman Spectroscopy in Analysis of Biomolecules

Gediminas Niaura
Institute of Chemistry, Vilnius, Lithuania

1 Introduction **2**
2 History **3**
3 Comparison of Raman Methods **4**
4 Interpretation of Raman Spectra **4**
4.1 Combination of Several Raman Techniques 4
4.2 Isotopic Substitution 7
4.3 Mutation 8
4.4 Depolarization Ratio Measurements 8
5 Experimental **9**
5.1 Sample Handling 9
5.2 Frequency and Intensity Calibration 9
6 Applications to Biomolecules **11**
6.1 Amino Acids and Peptides 11
6.2 Active Centers in Proteins 15
6.3 Nucleic Acids 19
6.4 Lipids and Biomembranes 23
6.5 Dynamics in Biomolecules 25
6.6 Raman Microscopic Analysis of Biological Systems 26
7 Future Developments of Raman Spectroscopy **26**
8 Comparison with Other Spectroscopic Methods **27**
8.1 Structural Information 27
8.2 Sensitivity 27
Acknowledgments **27**
Abbreviations and Acronyms **27**
Related Articles **28**
References **29**

Raman spectroscopy (RS) is a vibrational spectroscopy technique, based on the phenomenon of inelastic light scattering from matter (the Raman effect). It is an irreplaceable tool for the study of biological events at the molecular level and for identification of biomolecules. There are several reasons for this. First of all, RS measures the vibrations of atoms. This implies that the positions of the bands, widths and intensities are sensitive to the molecular structure, and minor perturbations induced by intramolecular or intermolecular interactions can easily be followed. Vibrations of some molecular groups are very characteristic (characteristic vibrations) and therefore can be used for the identification of certain groups or even whole molecules. In addition, work using isotope labeling, model compounds, site-directed mutagenesis and normal coordinate analysis has led to the identification of so-called marker bands, which serve as a bridge in establishing spectrum–structure correlations. Second, water causes weak Raman scattering, and consequently biomolecules can be studied in their natural environment without solvent interference. The third reason for the growing interest in RS is related to the considerable intensity enhancement (up to 10^6) observable in resonance Raman spectroscopy (RRS) and surface-enhanced Raman spectroscopy (SERS). Important structural units, such as chromophoric active centers, aromatic amino acid residues, bases of nucleic acids, and peptide bonds, can be studied selectively by exploring the resonance Raman (RR) phenomenon and tuning the excitation wavelength (λ_{ex}) into the electronic absorption band of certain structural elements. Such an approach allows the major problem in vibrational spectroscopy of biomolecules to be overcome – namely interpretation difficulties due to overlapping of numerous bands. Because the Raman effect is inherently weak, the future for analytical applications is closely connected with both RRS and SERS. The combination of these two methods might be especially productive. Fourthly, because the Raman effect is instantaneous (timescale in the order of 10^{-14} s), and due to the recent advances in laser technology, the time-resolved action of biomolecules down to picosecond resolution can be followed, retaining the value of structural information. The ability to focus a laser beam down to the dimensions of λ_{ex} is the fifth advantage of RS. Combined with RRS or SERS, this approach allows the composition of a single living cell to be analyzed.

To explore fully the advantages of RS, the limitations of the technique must also be considered. First of all, the probability of Raman scattering is very low, so the effect is weak and high concentrations of biomolecules are required. Secondly, in RR experiments excitation within the electronic absorption band often causes photodegradation of the active center in the biomolecule. To avoid such problems, low laser powers, moving samples and independent inspection of the sample integrity are often employed. The same problems are also present in the SERS field. Overheating by the laser beam could be an obstacle in Fourier transform Raman (FTR) studies. Finally, many active centers in biomolecules, or impurities in the sample, exhibit intense fluorescence (Fl), obscuring the Raman spectra.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The Raman effect arises from the interaction of incident photons with electrons of the matter under investigation. During this interaction the photons can lose (Stokes' process) or gain (anti-Stokes' process) energy equal to the vibrational energy of the atoms. Consequently the vibrational energy of the atoms increases or decreases. Such communication is possible for the motions of atoms, which modulate the polarizability of the molecule. Intense Raman bands will be observed from nonpolar groups, particularly aromatic rings, the vibrations of which produce considerable modulation of polarizability. The resulting Raman spectrum, presented in wavenumbers (cm^{-1}) as the difference between the excited and emitted photon energy, is the vibrational spectrum of the molecule. The effect is very weak, because the energy exchange probability is low. Only one in $\sim 10^8$ of the incident laser photons is converted to a Raman photon.[1] The timescale of the scattering process is in the order of 10^{-14} s, so the effect is extremely fast.[2] In the Raman process it is important that the vibrations of the atoms are not probed directly by the incident photons, but through the interaction with the electron cloud. Thus the state of the electrons is also significant and when the energy of incident photons approaches the electronic transition the intensity in Raman spectra is enhanced up to 10^6 times (RR effect).[3] As we will see this unique possibility determines the face of modern biomolecular RS.[1–9]

RS has the following main advantages which make it attractive for biologists:[1–8]

(1) structural sensitivity
(2) intrinsic weak scattering from water
(3) RR and surface-enhanced Raman (SER) phenomenon
(4) possibility to design various kinds of time-resolved experiments with picosecond resolution
(5) possibility to focus laser beam down to the dimensions of λ_{ex}.

Vibrations of atoms depend on the mass and force field. Even minor perturbations in the force field induced by intermolecular or intramolecular interactions are reflected in the spectra, permitting the analysis of structural conformations, bonding strength, and phase transitions. In addition, isotopic labeling perturbs the mass of the system and the location of particular vibrational modes in the spectra can be identified. The motions of atoms in some groups are typical of that particular group, and are hardly affected by the rest of molecule or the environment (characteristic vibrations).[1] Such bands are very useful for instantaneous identification of certain atomic groups in a biomolecule from its spectrum. Tremendous efforts in establishing correlations between spectra and structure, using isotope labeling, model compounds, site-directed mutagenesis and normal coordinate analysis, have led to the identification of certain marker bands, whose position or intensity is directly dependent on the structure and state of the particular group. For example, marker bands provide information on the redox and spin state of the heme group in heme proteins,[7] the secondary structure of polypeptides,[6,7] the state and bonding of amino acid residues,[6] the structure and stacking of nucleic acid bases,[6] and melting phenomena in phospholipid bilayers.[6,7] Structural transformations of biomolecules can be followed by RS in water, i.e. in their natural environment, without serious interference from the solvent. Water causes weak Raman scattering, since vibrations of the highly polar O–H bonds produce only minor polarizability changes. Although Raman spectra mainly contain structural information, the application of conventional RS to real systems is hindered by the complexity of biomolecules. Indeed, proteins routinely contain thousands of atoms. Because the number of normal vibrations for an N-atom molecule is $3N - 6$ (or $3N - 5$ for linear molecules), it is clear that in general spectra contain too much information and consist of many overlapping bands, which cause serious interpretation problems. Fortunately, resonance enhancement (when λ_{ex} lies within the electronic absorption band) and surface enhancement (at roughened Ag, Au and Cu surfaces) effects have been discovered.[9] In RR studies only the vibrations of chromophoric groups are enhanced in intensity (up to 10^6 times), considerably simplifying the spectra. Because different structural units exhibit absorption at different wavelengths, tuning the λ_{ex} makes it possible to selectively probe each unit. Combination of surface and resonance enhancements in surface-enhanced resonance Raman spectroscopy (SERRS) appears extremely promising for analytical purposes, as in some cases concentrations of analytes can be reduced down to $\sim 10^{-10}$ M, and also because of considerable Fl damping. Reports on the detection of single dye molecules have appeared.[10,11] Unlike Fl studies, it is important that the vibrational spectrum from a single molecule contains structural information and characteristic vibrations. The development of reliable and more reproducible sample preparation protocols will expand the analytical applications. The function of biomolecules cannot be understood without knowledge of the evolution of structural transformations over time, as well as the formation of various transient states and species. Time-resolved RS[9] is particularly attractive for fundamental biophysical and biochemical studies because of the combination of molecular level information intrinsic to vibrational spectroscopy with time resolution from milliseconds to picoseconds. Events faster then 1 ps can

also be followed by time-resolved RS, although due to the uncertainty principle the spectral resolution will be reduced. Achievements in this field of RS are directly connected to recent developments in pulsed laser technology. Another field of biospectroscopical research has emerged from the combination of vibrational information with spatial resolution, in the form of micro-Raman spectroscopy (MRS).[9] The laser beam can be focused to the dimensions of light wavelength. Consequently, spatial resolution depends on the wavelength of λ_{ex}. For biologists this opens the possibility of examining compositions of the components of living cells. The combination of RR and/or SER effects with MRS appears to be extremely promising for clinical research.[12]

The advantages of RS for biomolecular studies cannot be fully explored without considering the following main limitations:[1,2,9]

(1) Raman scattering is a weak phenomenon;
(2) the sample may undergo photodegradation by the laser beam;
(3) the sample may be heated by the laser beam;
(4) Fl.

Because Raman scattering is weak, high concentrations of the biomolecules are required. Typical concentrations of nucleotide or peptide groups in nonresonance RS should be in the order of 0.03–0.1 M.[13] In RRS and SERS these concentrations can typically be reduced down to 10^{-8}–10^{-5} M. However, the influence of the laser beam on the sample must be very seriously considered in RRS and SERS experiments. Absorption of photons by the molecular group under investigation (in RRS) or by the surface (in SERS) often leads to photodegradation or denaturation of the biomolecule due to the increased temperature. Overheating problems also should be considered in FTR experiments as water absorbs laser beams at 1064 nm. Fl is intrinsic for many chromophoric active centers in biomolecules, and can obscure the Raman spectrum completely, due to its higher efficiency (up to 10^6 times). Additional source of Fl are traces of impurities in the sample. This problem has seriously restricted routine applications of RS to biomolecules. Tuning λ_{ex} in the near-infrared (NIR) region as well as application of FTR techniques[1,9,14] may overcome this limitation.

In summary, RS in its various forms provides valuable information on the structure of the molecule and its interactions. In particular, certain structural units and chromophoric centers can be probed selectively by exploring the RR approach. In addition, because of the fast response time rapid structural transformations can be followed. Such information is the basis for a long-standing goal of biochemists and biophysicists – to understand the mechanisms of biological phenomena at a molecular level. The structural sensitivity is also extremely attractive for analytical purposes, as it can be used to identify certain biologically active functional groups or biomolecules in different environments and mixtures. A major breakthrough in analytical applications is expected from the combination of RRS and SERS. The possibility of detecting biomolecules at trace level is becoming a reality in universities and research institutes. The development of more reproducible and reliable surface preparation protocols will shift this field to more applied laboratories and clinical units.

The goal of this article to provide orientation on the advantages and limitations of RS, and thus to stimulate the application of this technique to the analysis of biomolecules. We hope to provide an idea of the structural information extractable by different kinds of RS.

Characteristic vibrations and marker bands are the basis of many RS works, and are provided in many sections. It is believed that those bands serve first of all as illustrations in spectra–structure correlations, but can also be used independently in some studies using RS. Selected examples will demonstrate what kind of problems in the analysis of biomolecules can be addressed by RS.

2 HISTORY

The Raman effect was discovered experimentally in 1928 by Raman and Krishnan,[15] and independently by Landsberg and Mandelstam.[16] The importance of the discovery was immediately appreciated by the scientific community, and in 1930 Raman received the Nobel Prize in physics. In 1952, the resonance enhancement effect was observed experimentally by Shorygin.[17] These results were presented in 1953, at a meeting in Gmunden, Austria.[4] In the early, mercury lamp history of RS the biological applications were secondary. Not until 1958 did Garfinkel and Edsall[18] record the first Raman spectrum of a protein, lysozyme. They used a high-pressure mercury lamp for the scattering excitation and photographic plates for the detection. The important step towards a deeper understanding of the structure of proteins came with the determination of the normal vibrational modes of the peptide group in *N*-methylacetamide by Miyazawa et al. in 1958.[19] Further attempts in this direction led to the elucidation of some useful correlations between the amide vibrational spectra and the secondary structure. The discovery of lasers stimulated tremendous growth of Raman applications and in 1968 a report on the first laser-excited Raman spectra of proteins appeared.[20] Two years later Long and Loehr[21] introduced the resonance enhancement effect for the analysis of the

active site of the iron–sulfur protein, rubredoxin. Further application of RRS to biomolecules (heme proteins) led to the discovery of the inverse-scattering polarization phenomenon by Spiro and Strekas[22] in 1972. This discovery stimulated the development of RR theory and had an impact on general RS. The possible use of RS in model membrane studies was first demonstrated in 1971 by Lipperd and Peticolas.[23] At this time several new and powerful Raman techniques emerged. Barron and Buckingham[24] predicted Raman optical activity (ROA) in 1971. Two years later the effect was observed experimentally by Barron et al.[25] An interesting and unpredicted discovery, the SER effect, was observed[26] in an electrochemical cell in 1974. In 1980, after understanding the main features of the effect, first surface-enhanced resonance Raman (SERR) spectra of the adsorbed proteins cytochrome *c* (cyt *c*) and myoglobin (Mb) were reported,[27] at a silver electrode, from solutions containing 10^{-6} M concentrations of proteins. Further development of this technique led to single-molecule detection sensitivity.[10,11]

It is clear from this brief historical overview that biomolecular RS is open for fascinating new discoveries and powerful applications. For example, in 1998 the first ROA spectra from DNA and RNA were reported by Bell et al.[28] opening the new field of direct study of chiral elements in biomolecules.

3 COMPARISON OF RAMAN METHODS

The characteristic mark of modern RS is the variety of techniques used to explore the Raman effect.[9] Table 1 summarizes the main advantages and limitations of methods useful for biomolecular applications.[29–36] To overcome the selectivity and sensitivity limitations characteristic of classical nonresonant RS, the RR and ultraviolet resonance Raman (UV/RR) approaches are extremely powerful. Most importantly, by tuning the λ_{ex} different parts of biomolecules can be selectively probed. This approach will be demonstrated in section 4. SERRS allows adsorbed species to be detected at concentrations in solution of the order of 10^{-10} M. Today this is the most sensitive vibrational technique. However, the price for such sensitivity is the adsorption requirement, which frequently causes denaturation or structural transformations.[33] To preserve the function and structure of biomolecules at surfaces lowering of the temperature or modification of the electrode surface by preadsorbed organic molecules or ions has been suggested.[32,33] In addition, the possibility of photodegradation should be considered seriously in SERRS, SERS, and RR experiments. Various ways of moving or rotating the sample can effectively decrease the destructive influence of the laser beam,[2] and will be considered in more detail. Finally, FTR instruments are very attractive for routine analysis of rather concentrated samples, due to the absence of the Fl problem and their simplicity to use.

4 INTERPRETATION OF RAMAN SPECTRA

Assignment of the bands is the bridge between recorded spectra and normal vibrations of molecules. It is the first step necessary for the application of RS in analysis of biomolecules. The major problem is the complexity of the spectra. In this section we will demonstrate how the combination of several Raman techniques, isotopic substitution, site-directed mutagenesis and depolarization ratio (ρ) measurements is able to considerably simplify the task. Normal-coordinate analysis,[1,2] although a very powerful tool for the assignment of bands, is out of scope of this article and will be not discussed here.

4.1 Combination of Several Raman Techniques

One of the most serious limitations preventing the application of classical vibrational spectroscopy to biomolecules is that the spectra are too complex and overlapped, due to the large number of atoms in the molecule. An elegant solution to this limitation is the application of several Raman techniques to the same sample in combination, providing the possibility of seeing the molecule at different angles. The unique power of combined visible resonance Raman spectroscopy (VIS/RRS) and ultraviolet resonance Raman spectroscopy (UV/RRS) is demonstrated in Figure 1.[31] Electronic absorption transitions from different protein structural units can be observed in the absorption spectrum:

(1) heme charge transfer (CT) band at 600 nm that contains transitions involving the metal ion and the axial ligands;

(2) strong $\pi \rightarrow \pi^*$ Soret transition in heme group at 400 nm;

(3) absorption of aromatic amino acid residues (Trp and Tyr) in the 260–220 nm range;

(4) peptide backbone amide $\pi \rightarrow \pi^*$ transition below 220 nm.

As can be seen from Figure 1, by tuning λ_{ex} from 600 nm to 210 nm it is possible to selectively probe different structural units in the protein:

(1) iron–fluoride (axial ligand) stretches and heme in-plane ring vibrations;

Table 1 Comparison of different Raman methods

Method	Sensitivity and requirements for sample	Advantages	Limitations	Refs.
Nonresonant RS	0.1–0.03 M of nucleotide or peptide group; solid, liquid samples; biomolecules in solutions	Most biological systems can be studied; Fl can be avoided by using red λ_{ex}	Low sensitivity; low selectivity; interpretation difficulties due to complexity of the spectrum from biomolecules	1, 2, 13
Nonresonant RDS	10^{-3} M in protein complexes	Most biological systems can be studied; Fl can be avoided by using red λ_{ex}; probe of ligand–protein, substrate–enzyme, and H-bonding interactions; frequency shift accuracy $\sim$0.1 cm^{-1}; estimation of isotopic frequency shifts; photolabile ligands can be studied	Low sensitivity	29, 30
VIS/RRS	10^{-4}–10^{-7} M of biomolecules	Selective probe of chromophoric groups; high sensitivity	Photodegradation; Fl; chromophore is required; expensive lasers	1, 2
UV/RRS	10^{-3}–10^{-4} M of nucleotide or peptide group; molecules in solution	Selective probe of groups absorbing in 190–300 nm region (peptide group, aromatic amino acid residue, bases of DNA and RNA); high sensitivity; no interference from Fl for $\lambda_{ex} < 260$ nm	Photodegradation; expensive lasers	31
SERS	10^{-4}–10^{-8} M of biomolecules; roughened Ag, Au, Cu surfaces; adsorbed biomolecules	High sensitivity; groups interacting with surface can be probed selectively; adsorption geometry; dependence of the intensity from molecular group on the distance from the electrode; reduction of Fl; excellent ability to study redox behavior of biomolecules in electrochemical cell	Adsorption-induced denaturation; photodegradation; thermodegradation; only adsorbed molecules can be probed	2, 9, 32
SERRS	10^{-6}–10^{-10} M of biomolecules; roughened Ag, Au, Cu surfaces; adsorbed biomolecules	Extremely high sensitivity; selective probe of chromophoric active centers; orientation of chromophores with respect to the electrode surface; intensity dependent on the distance from electrode; reduction of Fl; excellent ability to study redox behavior of biomolecules in electrochemical cell	Adsorption-induced denaturation; photodegradation; thermodegradation; only adsorbed molecules can be probed	9, 32, 33
FTRS	Excitation in NIR region; interferometer	Elimination of Fl; high resolution; excellent spectral subtraction possibilities (can see 1% components); Stokes and anti-Stokes regions can be collected simultaneously (useful for obtaining spectroscopic temperature of the sample); both IR and Raman capabilities in one instrument; ease of use for nonspecialist, safe method, does not required special skills	Low sensitivity; absorption in the NIR spectral region of the analyte; water absorption; overheating problems	2, 9, 14
TR[3]	10^{-3}–10^{-5} M of biomolecules in solution	Picosecond timescale resolution; Raman spectra from excited state molecules; selective probe of chromophoric groups	Photodegradation; expensive lasers	2, 9, 34

(*continued overleaf*)

Table 1 (*continued*)

Method	Sensitivity and requirements for sample	Advantages	Limitations	Refs.
MRS	Biomolecules in films; biomolecules adsorbed at surfaces of (combined with SERS); sensitivity depends on combination with RR or SERS/SERRS	Spatial resolution 0.1–2 μm; low sample mass (picogram range)	Photodegradation	2, 9, 12, 35
ROA	0.1–0.01 M of nucleotide or peptide group; biomolecules in solution	Sensitive to 3-D structure (or tertiary fold in the case of proteins) of chiral molecules and molecular groups	Low sensitivity; high acquisition times	9, 36

RDS, Raman difference spectroscopy; VIS/RRS, visible resonance Raman spectroscopy; UV/RRS, ultraviolet resonance Raman spectroscopy; FTRS, Fourier transform Raman spectroscopy; TR^3, time-resolved resonance Raman; IR, infrared; 3-D, three-dimensional.

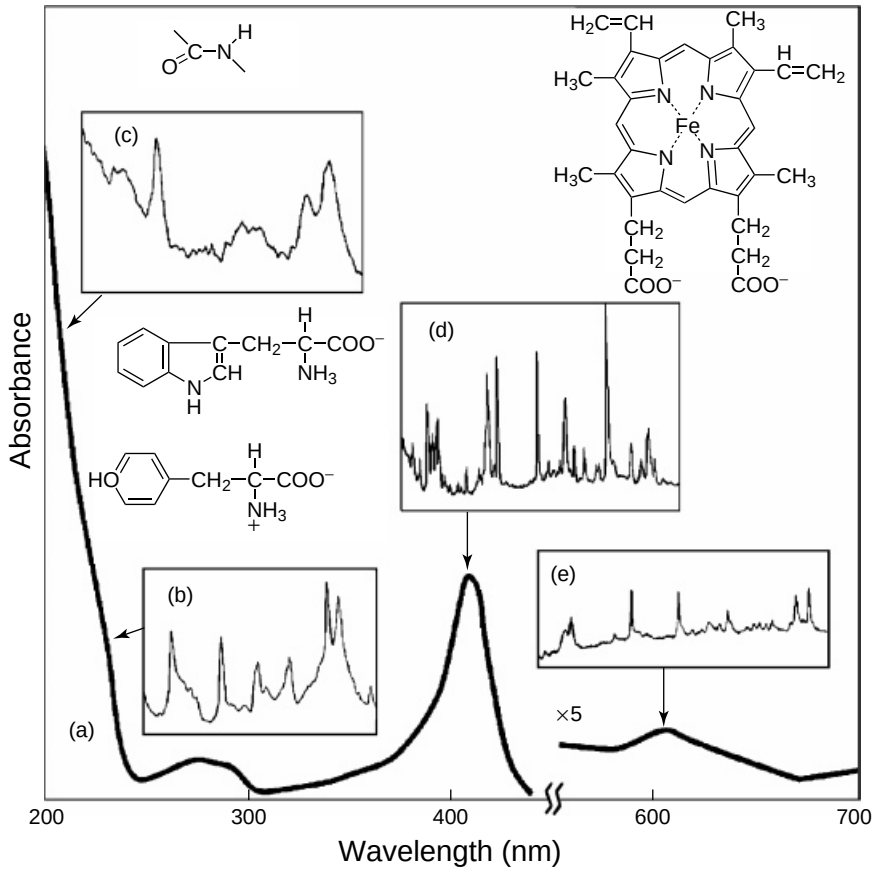

Figure 1 Selective probe of different structural units of sperm whale Mb by RRS: (a) absorption spectrum; (b) ultraviolet (UV) excitation at ~225 nm enhances tyrosine (Tyr) and tryptophan (Trp) bands (structures shown); (c) excitation further in the UV region enhances amide vibrations whose intensities and frequencies depend on the protein secondary structure (structure of amide group shown). Excitation within the heme absorption bands in the visible spectral region (400–600 nm) enhances vibrations only of (d) the heme ring (structure shown) and (e) vibrations between the heme iron and its axial ligands. No protein vibrational modes are observed. [Reproduced by permission of American Chemical Society from S.A. Asher, *Anal. Chem.*, **65**(4), 201A–210A (1993).]

(2) heme in-plane ring vibrations;

(3) vibrations of aromatic rings in Trp and Tyr;

(4) amide backbone vibrations.

Detailed information on the binding of the ligand to the metal ion, the spin and redox states of the heme group, the state of Tyr and Trp residues, and the secondary structure of the protein can be collected from such investigations.

The second promising combination is used for studies of adsorbed biomolecules. It is straightforward to apply both SERRS and SERS techniques to molecules containing chromophoric groups, by changing λ_{ex}. As has been demonstrated[33] details on the state of the heme group (chromophore) can be derived from SERRS studies of the adsorbed protein cyt *c*, while information on coadsorbed anions, the state of amino acid residues close to the surface, or groups possibly involved in binding to the surface, can be extracted from SERS.[37] It is important to note that the state of the heme, which is probed by SERRS, in many cases is determined by the interactions of the protein with the surface or with adsorbed ions, which are probed by SERS. Thus complete and unique information on the metal/protein interface can be collected.

4.2 Isotopic Substitution

Vibrational frequencies are determined by the force field and masses of the atoms. Isotopic substitution selectively affects the mass of the vibrating atoms at a particular place in the molecule, but in general does not influence the bonding properties between atoms. Consequently, isotopic exchange of atoms is reflected in the frequencies as isotopic shift (Δ). Isotopic substitution plays a key role in solving the following problems:

(1) assignment of the vibration of a particular molecular group in the spectrum;

(2) coupling between vibrations;

(3) identification of the ligand and its binding mode (for example, it is possible to determine the number of identically bonded atoms using isotopically labeled ligand atoms).

As interpretation of the vibrational spectrum is one of the major problems in biospectroscopy, direct assignment of the vibrations of particular molecular groups in the spectrum through isotopic substitution is an indispensable tool in RS. The largest Δ is observed when H is replaced by D. The simplest way to produce such an exchange in biomolecules is by using D_2O as a solvent. The hydrogen atoms attached to N, O, and S atoms are mobile and will spontaneously exchange with deuterium provided these groups have sufficient accessibility to the solvent. The frequency ratio can be evaluated by using Equation (1) for two atom molecules:

$$\nu(m_1m_2) = \frac{1}{2\pi}\sqrt{k\left(\frac{m_1+m_2}{m_1m_2}\right)} = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad (1)$$

where m_1 and m_2 are the masses of the atoms, k is the force constant, and μ is the reduced mass. The calculation for the particular ν(NH)/ν(ND) case is shown in Equation (2).

$$\frac{\nu(\mathrm{NH})}{\nu(\mathrm{ND})} = \frac{\sqrt{(14+1)/(14\times 1)}}{\sqrt{(14+2)/(14\times 2)}} \approx 1.37 \qquad (2)$$

This means that the stretching vibration of the N–H bond, for example in the amine group of the lysine amino acid residue, will shift from $\sim$3300 cm^{-1} to $\sim$2360 cm^{-1}. From Table 3 in section 6.1, we can see that such a shift is indeed observed and helps to assign vibration of this group in the spectra. Coupling between vibrations arises when atoms of similar masses are connected in the molecular group by similar strength bonds.[2] For example, separate stretching vibrations of N–H, C–C, or C–O bonds cannot be found in the $-NH-CH_2-CH_2-O-$ group. Isotopic exchange of $^{14}N/^{15}N$ in this molecular group would lead to shifts in positions of several bands, thus indicating those vibrational modes and how much they are coupled with the motion of the N atom. The extent of the coupling can be recognized roughly from the Δ value (the stronger the coupling, the larger the Δ value). An interesting example is coupling between ring vibrational modes with N–H deformation in the imidazole ring of histidine (His). As shown in section 6.1, perturbations in coupling between vibrations induced by the replacement of the N–H group by the N–D group in acidic D_2O solutions result in completely different Raman spectra, featuring a very strong peak at 1408 cm^{-1}. This band from the protonated imidazole ring in D_2O solutions serves as the best His residue marker band for analytical purposes. Blue copper proteins are the other important examples of vibrational coupling. As shown in section 6.2, the isotopic $^{32}S/^{34}S$ and $^{14}N/^{15}N$ substitutions in the molecular group $Cu-S-C_\beta-C_\alpha-N$ induce clear shifts in all the observed low-frequency RR modes, showing that all the bands are generated by the ν(Cu–S) stretching vibration coupled with the deformations of the $Cu-S-C_\beta-C_\alpha-N$ moiety. In studying coupling phenomena, isotopic substitution experiments combined with normal coordinate analysis[1,2] are the most powerful tools.

Figure 2 demonstrates how isotopic substitution helps to understand protein–ligand bonding in cytochrome *bd*

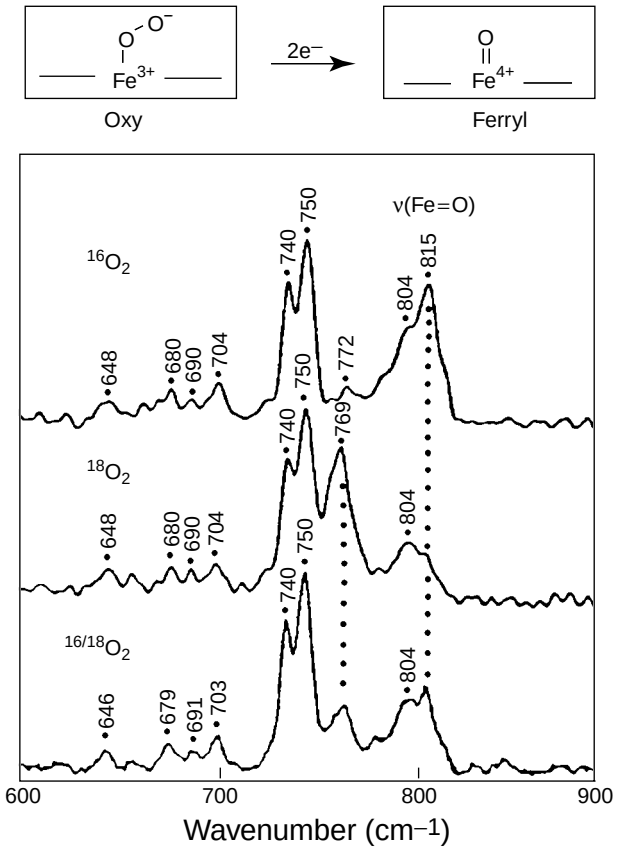

Figure 2 Identification of a ferryl intermediate in cytochrome *bd* oxidase. [Reproduced by permission of ACS from T.M. Loehr, ACS Symposium Series 692, eds. E.I. Solomon, K.O. Hodgson, ACS, Washington, 136–153, 1998.]

oxidase.[38] The reduced enzyme interacts with O_2 and the nature of the formed species was identified by RRS. In Figure 2 the ferryl intermediate Fe(IV)=O was identified using $^{18}O_2$, $^{16}O_2$ and mixed-isotope $^{18}O^{16}O$ dioxygen species. As can be seen, the position of one band clearly depends on the O_2 isotope used, and shifts from 815 cm^{-1} to 769 cm^{-1} upon $^{16}O_2 \rightarrow {}^{18}O_2$ substitution. The intensity of the peaks remains unchanged. In this frequency range two vibrational modes are expected: ν(O–O) from Fe–O–O^- species or ν(Fe=O) from Fe(IV)=O species. Unambiguous identification of Fe(IV)=O comes from the third experiment, using mixed-isotope $^{18}O^{16}O$ species. The vibrational modes $\nu(^{18}O-{}^{18}O)$, $\nu(^{16}O-{}^{16}O)$, and $\nu(^{16}O-{}^{18}O)$, should have different frequencies; thus if Fe–O–O^- species are present a new band should appear in the $^{16}O^{18}O$ experiment midway ($\sim$792 cm^{-1}) between the bands observed in the $^{16}O_2$ and $^{18}O_2$ experiments. Because no peak was observed in the vicinity of 790 cm^{-1}, the assignment of the observed bands to the ν(Fe=O) vibrations is definitive.

4.3 Mutation

Replacement of the entire amino acid residue in a protein through mutation combined with RS opens a way to study

(1) properties of individual amino acid residues[29]
(2) effects of mutation on protein structure[29]
(3) architecture and spectroscopy of the active centers in proteins.[39]

The ionization state of carboxyl groups has been monitored by comparing nonresonant Raman difference spectra of the wild-type protein and a mutant in which the carboxyl group was replaced by serine.[29] New insights into the spectroscopy of the active site of the blue copper protein, azurin (Az) have been revealed by the mutation of the His ligand to glycine (Gly).[39] In the active site of Az, the Cu ion is strongly coordinated by three ligands: one S atom from the cysteine (Cys) residue, and two N atoms from imidazole rings of His residues. The problem lies in deciding which of the bands in multiple RR spectra correspond to the Cu–N vibration. It was shown that mutant Az is able to bind exogenous ligands.[39] This approach opens the way to constructing an active site in which the Cu ion coordinates ^{15}N-labeled imidazole, simply by introducing isotopically labeled imidazole in the solution containing mutated protein. No isotope dependence was observed in the RR spectra of Az mutant reconstituted with ^{15}N-imidazole, indicating that His ligands do not contribute significantly to the spectrum of Az.

4.4 Depolarization Ratio Measurements

The depolarization ratio (ρ) is defined as shown in Equation (3)

$$\rho = \frac{I_\perp}{I_\parallel} \tag{3}$$

where $I_\perp$ and $I_\parallel$ refer to the intensities of scattered radiation polarized perpendicular and parallel to the incident light, respectively. Such measurements can be performed very easily for biomolecules in solution by placing the analyzer in the path of scattered Raman light between the sample and slit. Two spectra must be collected at perpendicular and parallel rotations of the analyzer. Thus ρ can be estimated for each band in the spectra. Important information about the symmetry of vibrational modes, which is critical in performing assignments of the bands, can be extracted from the ρ value

- $0 < \rho < 3/4$ for totally symmetric (polarized) vibrations
- $\rho = 3/4$ for nontotally symmetric (depolarized) vibrations

- $3/4 < \rho < \infty$ for anomalous (inverse) polarized vibrations (can only be observed in RR spectra).

The phenomenon of anomalous polarization was discovered in studies of heme proteins, hemoglobin and cyt *c*.[22] Polarization measurements were critical in making assignments in the RR spectra of complex heme proteins. Totally symmetric (A_{1g}), nontotally symmetric (B_{1g} and B_{2g}) and anomalously polarized (A_{2g}) bands were distinguished.

5 EXPERIMENTAL

The classical Raman spectrometer consists of the following main components:

(1) excitation source (laser);
(2) sample illumination;
(3) sample holder;
(4) scattered radiation collection system;
(5) monochromator or spectrograph;
(6) detection system [photomultiplier or charge-coupled device (CCD) camera].

The focus in this section will be on sampling schemes, avoiding laser effects on the integrity of the biomolecules, and calibration of the spectrometer.

5.1 Sample Handling

Laser-induced photodegradation of the sample is a major concern in RRS, UV/RRS, SERS, and SERRS experiments (Table 1). The principal solution used to avoid this effect is based on moving the laser spot relative to the sample. Such a procedure decreases the illumination time per sample volume or area. Rotating liquid sample cells,[40] liquid samples flowing through capillaries, rotating electrodes,[41] moving electrochemical cells,[42] and scanning beam techniques[43] have been developed. Some of them are shown in Figure 3.

5.2 Frequency and Intensity Calibration

Various methods[44–46] used for frequency calibration are compared in Table 2. The best accuracy can be achieved using neon emission lines. This method works very well for both one-channel detection systems (using photomultipliers) and multichannel detectors.[44] The small red commercial pilot lamp is easily available. It is convenient to record Ne lines and Raman spectra of the sample simultaneously. After the calibration Ne lines can be subtracted using the reference spectrum recorded with a blocked laser beam. An extensive list of Ne emission lines is available[44] in the wavelength range from 313 nm to 890 nm. As the Ne line method is an absolute wavenumber method, the exact λ_{ex} is required

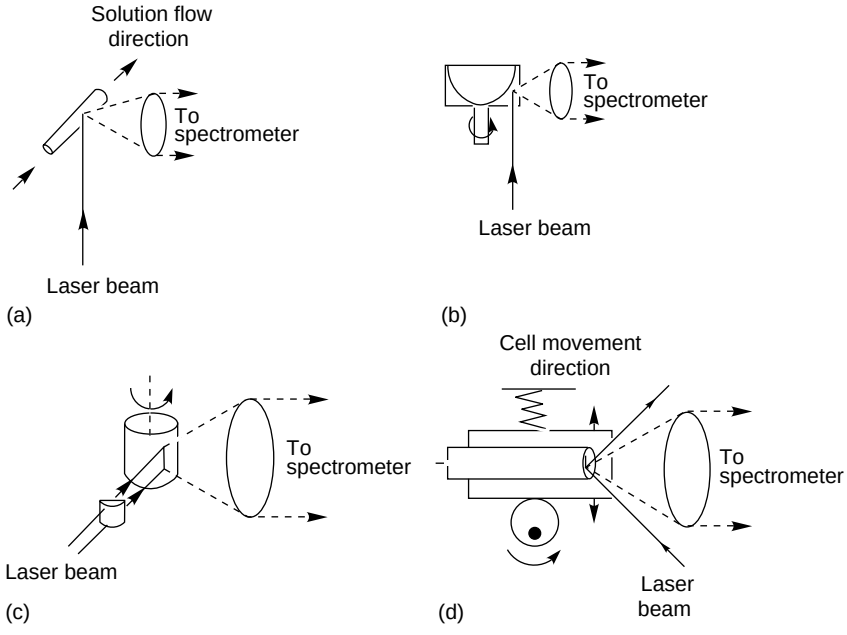

Figure 3 Sample schemes employed to avoid laser-induced photodegradation of the sample: (a) flowing capillary cell for liquids; (b) rotating cylindrical electrode for liquids; (c) rotating cylindrical electrode with use of cylindrical lens for surface Raman studies; (d) moving electrochemical cell for surface Raman studies. [(c) Reproduced by permission of John Wiley & Sons from P. Hildebrandt, K.A. Macor, R.S. Czernuszewicz, *J. Raman Spectrosc.*, **19**, 65–69 (1988); (d) reproduced by permission of John Wiley & Sons from G. Niaura, A.K. Gaigalas, V.L. Vilker, *J. Raman Spectrosc.*, **28**, 1009–1011 (1997).]

Table 2 Comparison of methods for frequency calibration[a]

Method	Accuracy (cm^{-1})	Advantages	Limitations
Internal standard (solvent bands or specially added compounds)	1	Simplicity; determined frequencies are temperature-independent; intense bands of tetrahedral anions may be used in aqueous solutions: SO_4^{2-} $\nu_s = 981\,cm^{-1}$; ClO_4^- $\nu_s = 935\,cm^{-1}$	Low accuracy; possible interaction of biomolecules with standard; narrow frequency range
Indene (Raman shift standard)	0.5	High accuracy; independent of λ_{ex} (useful for dye lasers)	Cannot be measured simultaneously with the sample; not useful for the high-frequency Raman calibrations ($1800-2700\,cm^{-1}$)
Laser plasma lines	<1	High accuracy; high density of lines; wide frequency range	Limited by the laser used
Neon emission lines (absolute wavenumber standard)	0.2	Very high accuracy; wide frequency range; best for multichannel detectors; possible simultaneous recording with the sample spectra	Accurate λ_{ex} value required (for dye lasers λ_{ex} must be determined)

[a] Based on Ferraro and Nakamoto,[2] Laserna,[9] Hamaguchi,[44] Strommen and Nakamoto,[45] and Fryling et al.[46]

for recalculating the positions of the lines to the Raman shifts. If the precise λ_{ex} is not known (for instance when working with dye lasers) it should be also calibrated against the Ne lines. The exact λ_{ex} value is not required for Raman shift methods, because of the well-known peak positions of standard samples. A spectrum from indene, which is recommended by the International Union of Pure and Applied Chemistry (IUPAC) for frequency calibration, is shown in Figure 4. Indene should be purified before use by vacuum distillation.[2]

Intensity calibration is a complicated issue in RS. Two aspects of the problem are considered in this section: estimation of the analyte concentration based on the intensity of the particular Raman band, and calibration of spectrometer sensitivity with respect to the wavelength (correction of relative intensities in the spectra). The intensity of the Raman scattering of a particular vibrational mode from the analyte molecule (I_a) depends on many factors,[2,45] as shown in Equation (4):

$$I_a = I_0 \nu^4 K(\nu) A(\nu) \beta_a C_a \tag{4}$$

where I_0 is the intensity of the incident laser radiation, ν is the Raman frequency, $K(\nu)$ is the overall spectrometer response, $A(\nu)$ is the self-absorption of the medium (important for RR studies), β_a is the molar Raman scattering parameter of the analyte molecules, and C_a is the concentration of the analyte molecules. Because the parameters $K(\nu)$, $A(\nu)$, and β_a are difficult to evaluate, relative rather than absolute intensities are employed in RS for C_a determination. For this purpose many internal intensity standards may be used, such as the tetrahedral anions ClO_4^- ($935\,cm^{-1}$ band) and SO_4^{2-} ($981\,cm^{-1}$ band), or water bands ($\delta(HOH) = 1640\,cm^{-1}$, $\nu(OH) = 3400\,cm^{-1}$, and $\delta(DOD) = 1204\,cm^{-1}$, and $\nu(OD) = 2500\,cm^{-1}$). If parameters referring to internal standard molecules are denoted by $K(\nu')$, $A(\nu')$, β_s, and C_s, the relative intensity $I_a(rel) = I_a/I_s$ can be expressed as shown in Equation (5):

$$I_a(rel) = \frac{\nu^4 K(\nu) A(\nu) \beta_a C_a}{(\nu')^4 K(\nu') A(\nu') \beta_s C_s} = \text{constant} \times C_a \tag{5}$$

The value of the constant may be determined easily by introducing a known amount of the analyte C_a(known). The calibration curves of Raman intensity versus concentration are linear.[1]

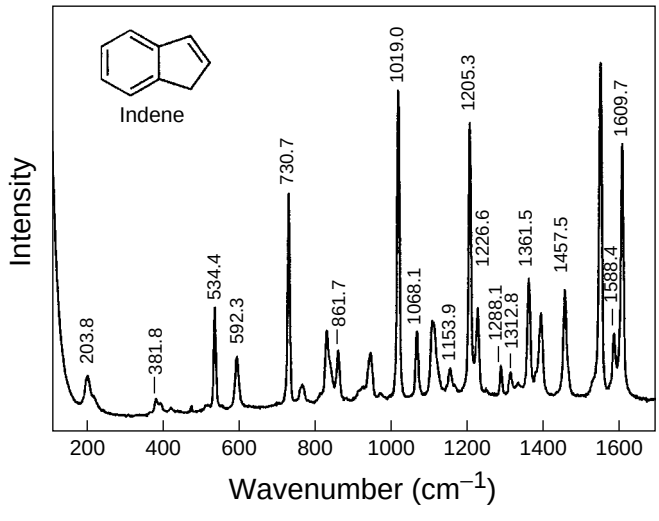

Figure 4 The Raman spectrum from indene used as wavenumber calibration standard. Frequencies of well-defined bands are given. [Reproduced by permission of Marcel Dekker from H. Hamaguchi, *Appl. Spectrosc. Rev.*, **24**(1&2), 137–174 (1988).]

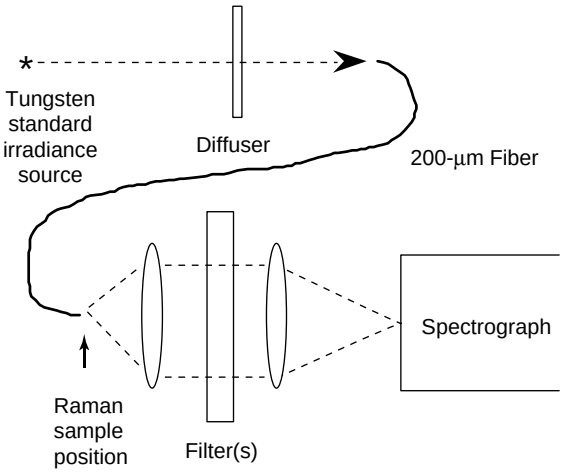

Figure 5 Apparatus for generating white light emission for the sensitivity calibration of Raman spectrometers. [Reproduced by permission of Society for Applied Spectroscopy from M. Fryling, C.J. Frank, R.L. McCreery, *Appl. Spectrosc.*, **47**(12), 1965–1974 (1993).]

Several methods have been proposed for the correction of spectrometer response with wavelength:

(1) tungsten lamps;[45,46]
(2) standard lamps from the National Institute of Standards and Technology (NIST);
(3) pure rotational spectrum from D_2;[44]
(4) Fl spectrum from quinine.[44]

The discreteness of the lines and narrow frequency range limit the application of the rotational Raman spectrum from D_2. The Fl method suffers from a rather narrow spectral range (400–600 nm) and the requirement to use a UV laser for the excitation of the radiation. The combination of a tungsten lamp and fiber optics was demonstrated[46] to be a promising choice as the white light source for intensity calibrations (Figure 5). For the Raman intensity correction to the instrument response, the measured Raman spectrum should be divided by the white spectrum recorded from the tungsten lamp at the same experimental conditions.[46]

6 APPLICATIONS TO BIOMOLECULES

This section will describe the applications of various kinds of RS to biological systems. Starting from the spectroscopic characterization of simple structural elements using marker bands, we will show how this knowledge can be applied to the molecular-level characterization of more diverse biological systems. The emphasis will be on the spectrum–structure correlations provided in the numerous tables.

Applications of RS to biomolecular studies are rapidly increasing. A number of excellent reviews have been published concerning studies of peptides,[47–51] active centers of proteins,[52–59] nucleic acids,[60–62] membranes,[63–66] dynamics of biomolecules,[67–69] and microscopic applications.[70]

6.1 Amino Acids and Peptides

Amino acids are the fundamental building blocks of proteins and therefore attract much attention from vibrational spectroscopists. The Raman technique was first applied to amino acids by Garfinkel and Edsall in 1958.[71] From then on these systems have been extensively studied by nonresonance Raman,[47,72–78] UV/RR,[48,49,79] ROA,[36] and SERS.[32] The information obtained in such studies can be divided into three categories: identification, structural and conformational analysis, and interactions.

Precise identification of amino acid residues in proteins is a difficult task even for such a structure-sensitive technique as RS. This is because many functional groups present in proteins contribute to the spectra, and as a result the observed bands are broad and overlap. To overcome these difficulties a combination of several approaches is very useful. We will concentrate on three of them:

(1) isotopic solution H_2O/D_2O exchange
(2) variation of pH
(3) application of the UV/RR technique.

The usefulness of such an approach can be seen from Table 3, where characteristic vibrations of the main functional groups of amino acids in different environments are summarized. The frequencies of the vibrations of N–H, O–H or S–H bonds containing exchangeable hydrogen atoms are lowered by a factor of approximately $\sqrt{2}$ going from H_2O to D_2O solutions. Such a big shift in principle allows identification of the amine ($-NH_2$), ammonium ($-NH_3^+$) and sulfhydryl ($-SH$) groups (Table 3). For example, the sulfhydryl group can be immediately identified by the observation of the strong band in the 2560–2580 cm^{-1} region in H_2O solutions, which shifts to 1850–1880 cm^{-1} in D_2O. However, $\nu_s(NH_2)$ falls in the region of strong water scattering, and subtraction of a pure water spectrum is recommended. It should be noted also that full width at half-maximum (fwhm) for the N–H vibration in an $-NH_2$ group is considerably lower than vibrations of water. More serious problems are related to identification of the ammonium group. A high rate of exchange of protons and extended H-bonding interactions result in very high fwhm for these bands in addition to their

Table 3 Characteristic vibrations of the main functional groups in amino acids[a]

Group	Frequency (cm^{-1})		Approximate description	Structural information
	1650–1750 (w)		ν(C=O)	Ionization state; H-bond strength
	1250–1300 (w)		ν(C−O)	
	620–640 (m)		δ(CCO)	
	1395–1420 (s)		ν_s(COO)	Ionization state; coordination type with metal cations
	1610–1560 (w)		ν_{as}(COO)	
	H_2O	D_2O		
$-NH_2$	3290–3344 (vs)	2432–2480	$\nu_s(NH_2)/\nu_s(ND_2)$	Ionization state; H-bond strength
	1590–1627 (m)	1187–1214	$\delta(NH_2)/\delta(ND_2)$	
$-NH_3^+$	3000–3100 (w,br)	2130–2220	$\nu_s(NH_3^+)/\nu_s(ND_3^+)$	Ionization state
	1590–1660 (mw)	1140–1230	$\delta_{as}(NH_3^+)/\delta_{as}(ND_3^+)$	
	1490–1530 (m)	1064–1180	$\delta_s(NH_3^+)/\delta_s(ND_3^+)$	
		λ_{ex} (nm)		
		488 200		
Phe residue	624	(w) –	δ(ring)	
	1004	(vs) (s)	ν(ring breathing)	Intensity standard; marker band
	1032	(m) (m)	δ(CH) in plane	Marker band
	1207	(m) (m)	$\nu(C_6H_5-C)$	Marker band
	1586	(w) (vs)	ν(ring)	Marker band in UV/RR
	1606	(m) (vs)	ν(ring)	Marker band in UV/RR
	Tyr	Tyrosinate anion		
	pH 0.6	pH 12.06		
Tyr residue	645	644	δ(ring)	
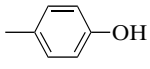	827 (w)	832 (s)	FR doublet: ν(ring) + 2 × δ(ring) out-of-plane	*I*(850)/*I*(830) sensitive to H-bonding, ionization state of OH group and hydrophobic interactions; ratio decreases with increase in OH group proton donor strength
	847 (s)	854 (s)	FR doublet: ν(ring) + 2 × δ(ring) out-of-plane	
	1182 (vw)	1175 (w)	δ(CH) in-plane	
	1210 (m)	1209 (s)	ν(ring-C)	
	1267 (s)	1270 (s)	ν(ring-O) + δ_s(ring)	
	1600 (w, m[b])	1555 (w, m[b])	ν_{as}(ring) + δ(OH)	Frequency is sensitive to the strength of H-bond interactions (UV/RR with λ_{ex} = 239 nm)
	1619 (w, s[b])	1602 (m)	ν_s(ring) + δ(OH)	

(*continued overleaf*)

Table 3 (*continued*)

Group	Frequency (cm^{-1})		Approximate description	Structural information
	H_2O	D_2O		
Trp residue	759 (s)	754	ν(indole ring breathing)	Marker band
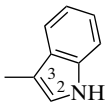	879 (w)	859	δ(NH) + δ(ring)	H-bonding of NH
	1012 (s)	1012	ν(benzene ring breathing)	Marker band
	1342 (m)	–	FR doublet	1342/1360 cm^{-1} doublet is sensitive to the hydrophobicity of local environment
	1360 (m)	1352	FR doublet	
	1435 (m)	1383	δ(NH) + ν(CC)	H/D sensitive
	1552 (s)	1550	$\nu(C_2{-}C_3)$ pyrrole ring	Marker band
	1579 (w)	1574	ν(benzene ring)	
	1622 (w, nr)	1618	ν(benzene ring)	
	$HisD_2^+$	HisD		
His residue	1602 (m)	1565 (m)	$\nu(C_4{=}C_5)$	Conformation of ring; ionization state; metal coordination site marker
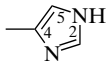	1408 (vs)		$\nu(N{-}C_2{-}N) + \delta_{ip}(ND)$	Ionization state and H-bonding; marker band
Tautomer I	1264 (s)	1274 (s) (I) 1258 (sh) (II)	ν(ring breathing)	Tautomer sensitive
	1110 (m)	1098 (m)	δ_{ip}(CH)	
Tautomer II	$HisH_2^+$	HisH		
	1627 (m)	1585 (w) (II) 1568 (m) (I)	$\nu(C_4{=}C_5)$	Tautomer sensitive; ionization state
	1484 (s)	1440 (s)	δ(ring) + $\delta(C_2H) + \delta_{ip}$(NH)	Ionization state
	1264 (s)	1282 (m) (I) 1260 (m) (II)	ν(ring breathing)	Tautomer sensitive
	1193 (s)	1160 (m)	ν(ring) + δ_{ip}(NH)	Ionization state
	994 (m)	1004 (m) (II) 983 (m) (I)	δ_{ip}(CH) in-plane	Tautomer sensitive
	H_2O	D_2O		
Cys residue	2560–2580 (vs)	1850–1880	ν(S–H)/ν(S–D)	Very characteristic marker band; H-bonding of SH
$-H_2C-SH$	590–735 (s)		ν(C–S)	**T/G** conformation; **T** mode has higher frequency
$-S-S-$	490–560 (s)		ν(S–S)	**T/G** conformation
Met residue	655, 724		ν(C–S) **T**	**T/G** conformation of $S-CH_3$ group
$-H_2C-S-CH_3$	700		ν(C–S) **G**	

[a] Based on Tu,[47] Harada and Takeuchi,[48] Thamann,[49] Krimm and Bandekar,[50] Tensmeyer and Kauffman,[51] Siamwiza et al.,[72] Ashikawa and Itoh,[73] Harada et al.,[74] Takeuchi et al.,[75] Fischer and Eysel,[76] Miura et al.,[77] Yue et al.,[78] and Rava and Spiro.[79]
[b] Intensity in UV/RR spectra.
FR, fermi resonance; ν_s, symmetric stretching; ν_{as}, asymmetric stretching; δ_s, symmetric deformation; δ_{as}, asymmetric deformation; δ_{ip}, in-plane deformation; **T**, trans; **G**, gauche; s, strong; w, weak; m, medium; sh, shoulder.

low intensity. In this situation the combined analysis of $\delta_s(NH_3^+)/\delta_s(ND_3^+)$ vibrations (Table 3) might be helpful. It is not only direct vibrations of the bond involving hydrogen atoms that can be probed by the H_2O/D_2O exchange approach. For example, some bands from the indole ring of Trp residue and the imidazole ring from His residue clearly shift on deuteration (Table 3).[73–75] The effect is associated with the coupling of certain ring vibrations with deformational N–H motion. In the case of the Trp residue two strong features at 759 and 1012 cm^{-1} can serve as the marker bands, and a slight shift of the first band to 754 cm^{-1} in D_2O solutions can serve as a reliable confirmation of the assignment, but the analysis of His residues strongly requires isotopic addition experiments. Three main phenomena control the shape and structure of the vibrational spectrum of the imidazole ring in His, depending on environmental conditions:

(1) extensive coupling of ring vibrational modes with N–H deformation;
(2) increases in local symmetry from C_s to C_{2v} upon protonation of both N atoms in the imidazole ring (as a result the spectrum is considerably simplified);[75]
(3) tautomerism of the neutral imidazole ring[73] (Table 3).

The protonated, positively charged imidazolium ring can be undoubtedly identified by the strong and narrow band at 1408 cm^{-1} which appears in D_2O solutions (Table 3). This band clearly dominates in the spectrum and serves as the best spectral signature for His residues. The symmetric stretching vibration $\nu_s(COO)$ from carboxylate groups and the deformational motion $\delta(CH_2)$ from methylene groups are possible interferences, although those bands do not change position in H_2O and D_2O solutions. Monitoring of the imidazole ring pK_a is the second valuable application of the 1408 cm^{-1} feature. On the other hand, tautomerism should be investigated in H_2O solutions by using the three tautomer-sensitive bands indicated in Table 3. There are also three bands sensitive to the ionization state of the imidazole group in H_2O solutions, located at 1627, 1484, 1193 cm^{-1} and 1568, 1440, 1160 cm^{-1} in protonated and deprotonated forms of the imidazole ring, respectively. Successful use of these bands for pK_a evaluation of His residues in proteins by employing RDS has been demonstrated by Yue et al.[78] Next, the advantages of UV/RRS for analysis of amino acids will be considered. Four aromatic amino acids, Phe (phenylalanine), Tyr, Trp and His, containing electronic transitions in the 190–300 nm wavelength region (Table 4) can be studied by the RR technique, which explores 10–1000-fold enhancement in intensity of the Raman bands due to the resonance effect.[31,79,80] By choosing an appropriate λ_{ex} it is possible to selectively probe particular amino acid residues in the protein, since to a very rough approximation the resonance enhancement is proportional to ε^2.[38] For example, Trp and Tyr can be examined selectively using λ_{ex} at 218 and 200 nm respectively (Table 4).

The second most important structural element in proteins is the peptide bond, which connects together amino acids. Thus in proteins not only are vibrations characteristic of individual amino acid residues observed, but also additional bands associated with particular peptide bonds (Figure 6). Normal vibrational modes of peptide groups were first determined by Miyazawa et al.[19] in 1958 using the model compound *N*-methylacetamide. Further experimental and theoretical efforts in this field led to the elucidation of important correlations between the amide vibrational spectra and the secondary structure of proteins.[47,50,51] Recent advances in this field are associated with the application of resonance enhancement of the amide vibrations using λ_{ex} in the UV region.[48,49]

Three in-plane vibrations of the peptide group are important in Raman studies: amide I, amide II and amide III (Figure 6). In D_2O solutions, exchange of hydrogen with deuterium takes place and the corresponding amide vibrations are labeled as amide I′, amide II′, and amide III′. As can be seen from Figure 6, the amide I vibration is concentrated on the motion of the C=O

Table 4 Molar enhancement factors ($\times 10^{-2}$) measured relative to the 3400 cm^{-1} H_2O band for selected Raman bands of the aromatic amino acids[a]

Amino acid	Electronic absorption λ_{max} (nm) and $\varepsilon \times 10^{-3}$ (M^{-1} cm^{-1})			Raman band (cm^{-1})	λ_{ex} (nm)			
					200	218	240	266
Phe	185 (36)	207 (7)	255 (0.3)	1000	36	96	8	–
Tyr (pH 7)	193 (36)	222 (7)	273 (1.3)	1617	202	126	30	–
Tyr (pH 11)	–	240 (11)	293 (2.4)	1602	437	152	53	–
Trp	218 (34)	273 (5.5)	287 (4.5)	1016	29	506	53	205
His (pH 8)	207 (5)	–	–	1575	8	32	4	–

[a] Data from Rava and Spiro,[79] and Caswell and Spiro.[80]
ε, Extinction coefficient.

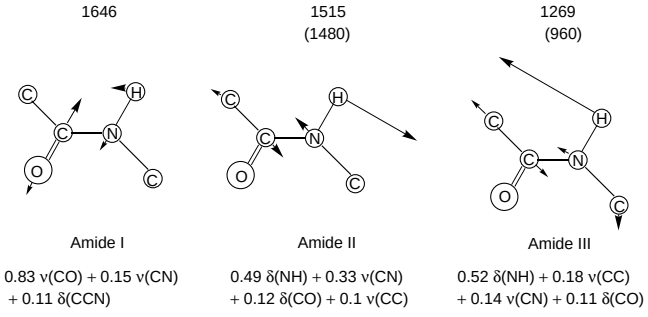

Figure 6 Calculated vibrational amide modes of *N*-methylacetamide. Vibrational frequencies in H_2O (D_2O) and potential energy distributions (all contributions ≥ 0.1 are included) are also shown. ν, Stretching; δ, deformation. [Reproduced by permission of Academic Press from S. Krimm, J. Bandekar, *Adv. Protein. Chem.*, **38**, 181–364 (1986).]

group, while the other two modes have higher contributions from the in-plane bending of N–H, and therefore can be easily detected by their considerable frequency reduction in D_2O solutions. The amide II mode, usually very weak in nonresonant RS, appears as an intense band in UV/RRS, especially in D_2O solutions. The high contribution from the stretching C–N vibration results in the considerable resonance enhancement of this special vibration. Particular enhancement of the amide II′ mode comes from the transformation of this vibration to the ν(CN) mode upon H→D exchange. The H-bonding between the hydrogen of the NH group and the oxygen of the CO group (Figure 6) is an essential component in determining the secondary structure of proteins. The H-bonding strength and rotation reflect on the amide frequencies. Thus different protein structures can be distinguished based on these frequencies, as shown in Figure 7. The frequency ranges presented demonstrate the basic trends in spectrum–structure correlations, although in each particular case such correlations should be considered with some caution.[50] A selective probe of amide vibrations is possible through the UV/RR approach with $\lambda_{ex} < 220$ nm. The electronic absorption spectrum of the peptide group is characterized by three transitions:[49] weak $n \rightarrow \pi^*$ at 220 nm; strong $\pi \rightarrow \pi^*$ at 190 nm; and $n \rightarrow \sigma^*$ at 165 nm. Excitation within the $\pi \rightarrow \pi^*$ transition enhances vibrational modes significantly coupled to ν(C–N), i.e. amide II and III vibrations (Figure 6). Excitation within the $n \rightarrow \sigma^*$ absorption selectively enhances the amide I mode.[49] A selective probe of amide bands in Mb with $\lambda_{ex} < 220$ nm is shown clearly in Figure 1.

6.2 Active Centers in Proteins

Active centers in proteins are mainly responsible for the particular biological function of the biomolecule.

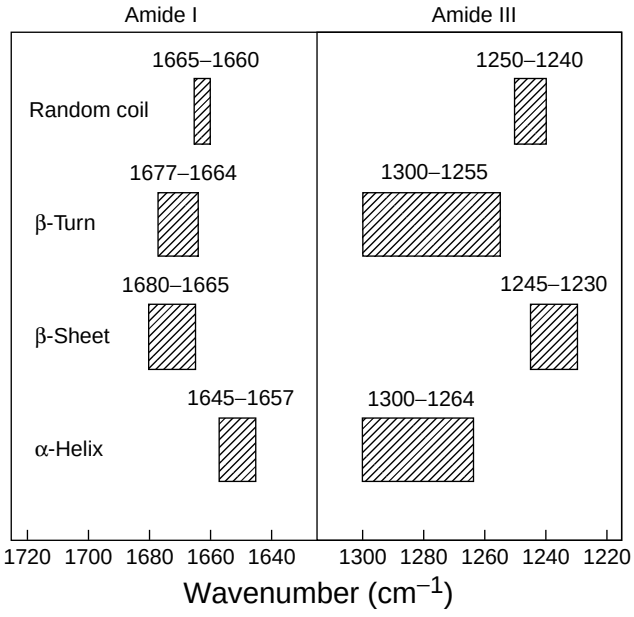

Figure 7 Correlation between amide I and III wavenumbers and protein conformations. There are always exceptions in band wavenumbers, but the ranges shown in the figure usually cover the majority of bands for a given conformation. [Reproduced by permission of John Wiley & Sons from A.T. Tu, in *Spectroscopy of Biological Molecules*, eds. R.J. Clark, R.E. Hester, John Wiley & Sons, Chichester, 47–112, 1986.]

In metalloproteins, metal ions coordinated by protein amino acid residues or certain molecular groups serve as a catalytic center of enzyme action or are involved in the electron transfer process. In many cases such active centers are chromophores and thus can be studied selectively by the RR technique. Unique information on the structure and dynamics of active centers can be extracted from the vibrational spectra. In this section, selected examples are used to illustrate applications of RRS to metalloproteins.

Heme proteins are the classical examples used in RR studies, because of their important biological functions and the possibility of observing high-quality spectra even in dilute solutions. Indeed, the first biomolecular application of RRS was performed on heme proteins by Spiro and Strekas in 1972.[22] The heme group, which consists of the porphyrin ring system and a coordinately bound iron ion (see structure in Figure 1), has intense electronic π–π^* absorption bands, because of the extended π system of the porphyrin ring. The main features of the heme electronic spectrum can be summarized as follows:[54]

- an intense band near 400 nm (Soret or B band, $\varepsilon \approx 10^5\ M^{-1}cm^{-1}$);

- a pair of peaks (about ten times weaker) at ~520 and ~550 nm (β and α or Q_v and Q_0 bands respectively);
- weak CT transitions.

Because coordination of nitrogen atoms from the porphyrin ring to the central metal ion influences the structure of the macrocycle, the observed resonance-enhanced vibrational spectrum contains unique information on the state of the central ion. Vibrations of axial ligands such as CO, CN^-, O_2, F^-, OH^- and NO can also be studied due to coupling with electronic $\pi-\pi^*$ transitions.[52,53] Excitation within CT transition also results in enhancement of the axial ligand modes.[52] The heme group can be considered in the first approximation as a planar structure with D_{4h} symmetry. Further, if the substitutents are treated as point masses, the in-plane and out-of-plane vibrational modes can be represented according to the symmetry shown in Equations (6) and (7), respectively.[54]

$$\Gamma_{\text{in-plane}} = 9A_{1g} + 8A_{2g} + 9B_{1g} + 9B_{2g} + 18E_u \quad (6)$$

$$\Gamma_{\text{out-of-plane}} = 3A_{1u} + 6A_{2u} + 5B_{1u} + 4B_{2u} + 8E_g \quad (7)$$

Because of the presence of the center of symmetry in an idealized heme group, the mutual exclusion principle operates for the Raman and IR selection rules. Only the ''gerade'' (g) modes are Raman-active, while ''ungerade'' (u) modes are IR-active. The allowed Raman modes can be recognized in the spectra by different polarization properties: A_{1g} modes are polarized, B_{1g} and B_{2g} modes are depolarized, and A_{2g} modes are anomalously polarized. It should be noted that the phenomenon of ''anomalous polarization'' was first observed by Spiro and Strekas for the heme proteins hemoglobin and cyt *c*.[22] The observed spectral pattern depends on λ_{ex}. There is a relationship between the symmetry of resonantly enhanced vibrational modes and the nature of the electronic transition. For a single molecule the Raman intensity I_R is given by Equation (8)

$$I_R \sim I_0 \nu_R^4 \sum_{p,\sigma} (\alpha_{p\sigma})_k^{\ 2} \quad (8)$$

where I_0 is the intensity of incident light, ν_R is the Raman wavenumber, and $(\alpha_{p\sigma})_k$ is the polarizability tensor component for the k transition. The summation is over molecular coordinates p and σ. For RR scattering, in the first approximation the polarizability tensor can be described by two terms as shown in Equation (9).

$$(\alpha_{p\sigma})_k \approx A + B \quad (9)$$

Depending on the nature of the electronic transition, term *A* or *B* dominates. The main features of terms *A* and *B* can be summarized as follows.

- Term *A* is associated with vibrational interactions with the excited electronic state through Franck–Condon overlapping and usually is a leading contributor to RR intensity. Fundamentals, overtones and combination bands of totally symmetric modes dominate in the spectra. For the heme group this resembles excitation within the B band and resonance enhancement of A_{1g} modes and overtones.
- Term *B* is associated with the vibronic coupling of one resonantly excited state to another excited state. As a result, fundamental transitions of nontotally symmetric vibrations are resonantly enhanced. This is often the case when excitation is within a weakly allowed electronic transition which is vibronically coupled with a strong electronic transition. For the heme group this resembles excitation within Q bands and resonant enhancement of B_{1g}, B_{2g} and A_{2g} modes.

Thus by choosing the appropriate λ_{ex}, the A_{1g} or B_{1g}, B_{2g} and A_{2g} modes can be investigated separately, considerably simplifying the interpretation. In real biological systems the symmetry of the heme group is distorted from the idealized D_{4h} as a result of asymmetric substitutions and deformation of the planar heme structure due to interactions with the protein matrix. Indeed, many in-plane IR-active E_u and out-of-plane deformation modes have been observed in cyt *c* RR spectra due to the protein-induced distortion of the heme group.[81] When assignment of the RR spectrum is performed, unique information on the state of the heme metal ion can

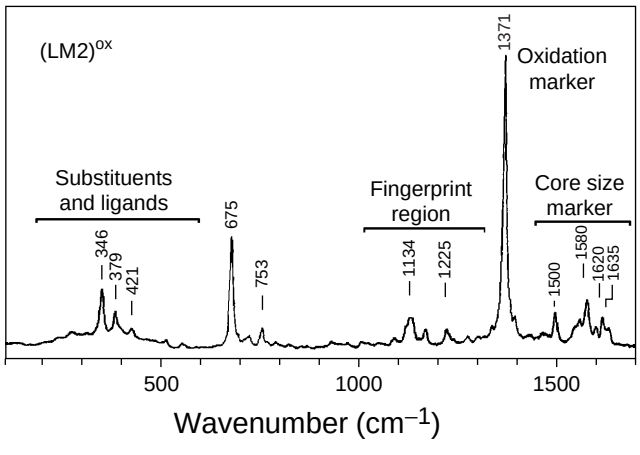

Figure 8 Overview of RR spectrum of the substrate-free ferric cytochrome P-450 LM2 [$(LM2)^{ox}$] excited at 407 nm. [Reproduced by permission of Akademie Verlag from P. Hildebrandt, in *Relationships Between Structure and Function of Cytochrome P-450 – Experiments, Calculations, Models*, eds. K. Ruckpaul, H. Rein, Akademie Verlag, Berlin, Chapter 6, 166–215, 1992.]

Table 5 Oxidation state and spin state marker bands of heme group[a]

Mode	Approximate interpretation	Fe^{3+}			Fe^{2+}			Comments
		LS 6C	HS 5C	HS 6C	LS 6C	HS 5C	HS 6C	
ν_4 (A_{1g})	ν_s (C–N) Pyr half-ring breathing	1373	1373	1370	1359	1357	1355 1373 (CO) Mb	Porphyrin π^* electron density marker: (1) oxidation state marker band (ν_4 decreases upon reduction); (2) axial ligation of π acid ligands CO, NO, O_2 marker band (ν_4 increases upon ligation)
ν_3 (A_{1g})	ν_s (CC)	1502	1491	1480	1493	1471	1463	Porphyrin core size marker bands: (1) spin state marker band (transition from LS to HS increases frequency); (2) CN marker band (transition from 6C to 5C increases frequency)
ν_{19} (A_{2g})	ν_a (CC)	1586	1571	1560	1583	1550		
ν_{10} (B_{1g})	ν_a (CC)	1640	1626	1610	1617	1604		

[a] Based on Spiro and Lee,[54] Hildebrandt and Stockburger,[55] and Spiro and Czernuszewicz.[56]
CN, coordination number; LS, low-spin; HS, high-spin; 6C, six-coordinated; 5C, five-coordinated; Pyr, pyrrole.

be derived by analyzing certain marker bands. A typical RR spectrum from a heme group excited within the B electronic band is presented in Figure 8. Totally symmetric A_{1g} vibrations dominate in this spectrum. Representative marker bands at different states of the heme iron ion are summarized in Table 5. Three main reasons are responsible for the sensitivity of the high-frequency in-plane stretching modes to the state of the Fe ion.[56]

- The frequency of the C–N and C–C stretching vibrations of the heme group depends on the electron density on the antibonding π^* orbitals of the porphyrin ring. Higher electron density weakens the order of C–C and C–N bonds, and as a result the frequency decreases. Transition from Fe^{3+} to Fe^{2+} increases electron occupation on π^* orbitals and can be recognized by the lowering of frequency (Table 5). On the other hand, all chemical effects causing withdrawal of the electrons from π^* orbitals cause increases in frequency. Such effects refer to the axial ligation of the π acid ligands as CO, NO and O_2.
- Stretching frequencies of the porphyrin ring depend on the core size of the porphyrin cavity. The core size depends on two factors: the spin state of the Fe ion, and the coordination number (CN). In the high-spin (HS) state two antibonding d orbitals of the Fe ion contain an electron and the porphyrin ring cavity is expanded compared with the low-spin (LS) state, where the antibonding d orbitals are empty. Expanding the core size decreases stretching frequencies (Table 5). The CN of the central Fe ion influences the core size as follows: in the 6C state the Fe ion is located in the heme plane, while in the 5C state the central ion is dislocated from the heme plane towards the fifth ligand. Such a dislocation decreases the core size and as a result, stretching frequencies increase (Table 5).
- Twisting and tilting of Pyr rings in the heme group form the idealized plane structure induced by the protein matrix.[56,81]

The axial ligation of the heme group can be studied not only through the high-frequency stretching vibrations of the porphyrin ring, but also directly by analyzing the stretching $\nu(M-AB)$, bending $\delta(M-A-B)$ and internal $\nu(A-B)$ modes (Figure 9) of two atomic AB ligands bonded with the metal ion M. Such modes were observed in RR spectra excited within the B electronic transition.[52,53] Important structural information about the bonded AB ligand, including the angle $M-A-B$ and the strength of the $M-A$ bond, can be derived from such studies.[52,53]

The RR spectra of heme proteins can be observed because of $\pi-\pi^*$ electronic transitions in the porphyrin

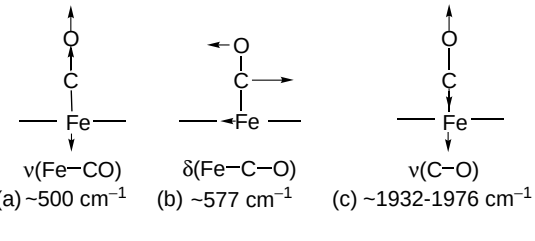

Figure 9 Schematic representation of metal–ligand (a) stretching, (b) bending and (c) internal ligand vibrations in the complex formed between CO and the iron ion of the heme group. Data for (c) from Yu and Kerr.[53] [(a) and (b) reproduced by permission of Academic Press from N.-T. Yu, *Methods Enzym.*, **130**, 350–409 (1986).]

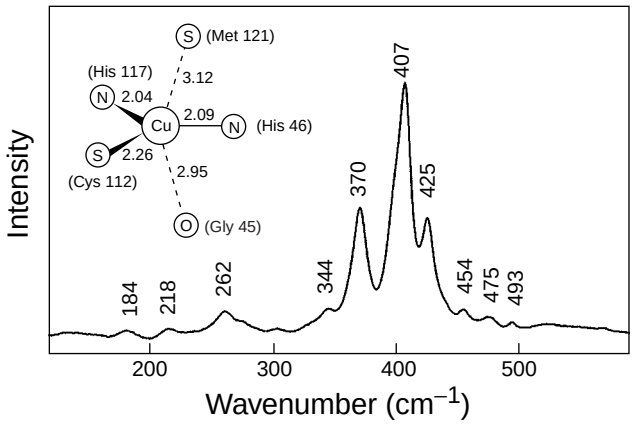

Figure 10 Overview of low-frequency RR spectra of blue copper protein Az (1 mM) from *Pseudomonas aeruginosa* obtained with 647 nm excitation. Integration time, 200 s. Insert shows the active site structure; bond lengths indicated in Å. (Data for insert from Nar et al.[82])

system. A different resonance enhancement mechanism in the Raman spectra operates for other metalloproteins, where the metal ion is directly ligated to the protein amino acid residues. Interesting examples for vibrational spectroscopy are the ''blue copper'' proteins (plastocyanin, Az, stellacyanin and others). The biological function of these proteins, electron transport, is closely related to the geometry of the active site. The structure of the Cu active center in Az can be described as trigonal-bipyramidal (Figure 10). Three strongly bonded ligands, namely the S atom of Cys, and two N imidazole ring atoms from His groups, are arranged in planar trigonal configuration. Two weakly bonded ligands, the S atom from methionine and the O atom from the peptide carbonyl of Gly, lie in the axial position with respect to the trigonal plane.[82] The axial ligands turn the geometry from trigonal to trigonal-bipyramidal. Such a structure is a compromise between the tetragonal geometry usually observed for Cu(II) coordination compounds, and the tetrahedral arrangement characteristic of the Cu(I) compounds, and is believed to be responsible for the fast electron-transfer rates typical for blue copper proteins. In the Cu(II) oxidation state these biomolecules exhibit intense electronic absorption bands around 600 nm assigned to the electronic $S(Cys)(\pi) \rightarrow Cu(II)(d_{x^2-y^2})$ CT transition. The RR spectrum produced by the excitation within this transition consists of several intense bands in the low frequency region (200–500 cm^{-1}), where metal–ligand vibrational modes are expected (Figure 10). Because the electronic transition involves CT from the S(Cys) to Cu(II) ion, the Cu–S bond properties are considerably affected by the absorption, and resonant enhancement in the Raman spectrum is concentrated on vibrations of the Cu–S bond. However, not one but several clear bands are observed experimentally. The origin of the bands has been discussed extensively.[56,58,83,84] In the earlier studies these bands were attributed to vibrations of Cu–N and Cu–S bonds.[83] Later, strong coupling between stretching ν(Cu–S) and deformation δ(SCC) vibrations from the Cys moiety was suggested to explain the observed lower isotopic $^{63}Cu \rightarrow {}^{65}Cu$ frequency shift in stellacyanin compared with the value predicted by normal coordinate analysis.[84] Recent extensive isotopic substitution experiments[39,56,58] (Table 6) provide evidence that all the bands observed in the 200–500 cm^{-1} frequency range are generated by stretching vibrations of the Cu–S(Cys) bond dramatically coupled with deformations of the $Cu-S-C_\beta-C_\alpha-N$ moiety. Multiple shifts induced by all $^{14}N \rightarrow {}^{15}N$ substitutions demonstrate the participation of the remote N atom of the Cys linkage

Table 6 Isotopic shifts of some Az Raman frequencies[a]

Modes	$^{32}S \rightarrow {}^{34}S$	$^{63}Cu \rightarrow {}^{65}Cu$	$^{14}N \rightarrow {}^{15}N$ (all)	$H_2O \rightarrow D_2O$	Structure and atoms (in bold) responsible for isotopic shifts
348			–2.5		
372	–0.4	–0.6	–0.9	–1	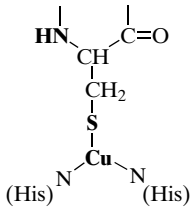
400		–0.6	–2.7	–2	
408	–3.8	–0.6	–1.4	–1	
428	–1.4	–0.2	–3.8	–1.5	
456			–2.6	–0.7	
476			–2.8	–1.5	
494			–0.5		

[a] Data from Andrew et al.[39]

in the coupling phenomenon (Table 6), because no isotopic shifts were observed with labeled N atoms of the imidazole ring in mutant Az.[39] Experiments with engineered active sites in mutant Az proteins[85] revealed the sensitivity of the ν(Cu–S) frequency to the length of the Cu–S(Cys) bond and thus to the geometry of the active center. The highest frequency, in the vicinity of 430–405 cm^{-1}, corresponds to the trigonal coordination (Figure 11). Weak axial interactions by the L' ligand cause a slight reduction in frequency. As bonding between the Cu ion and the L' ligand becomes stronger, and the geometry of the active sites changes through tetrahedral to tetragonal, the ν(Cu–S) frequency undergoes a further reduction to 320–300 cm^{-1} in the pure tetragonal arrangement. As happens frequently with Raman studies of biological molecules, the extended coupling observed between vibrational modes and ν(Cu–S) frequency–structure correlations goes far beyond the problems of blue copper proteins and has an impact on the vibrational spectroscopy of coordinated compounds in general.

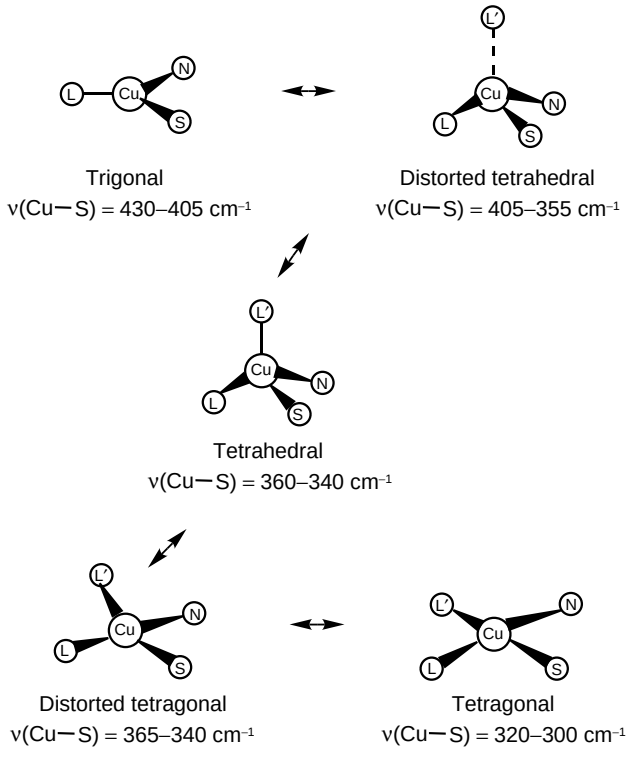

Figure 11 Correlation between coordination geometries in mononuclear copper cysteinate proteins and frequencies of Cu–S(Cys) bond stretching vibrations. Ligand movements required for interconversion between different geometries are shown by arrows. [Reproduced by permission of American Chemical Society from C.R. Andrew, J. Sanders-Loehr, *Acc. Chem. Res.*, **29**(8), 365–372 (1996).]

6.3 Nucleic Acids

Applications of RS for the analysis of nucleic acids have been increasing rapidly since the pioneering work of Lord and Thomas in 1967.[86] Excellent reviews of this extremely complex field are available.[60–62] Several different and powerful Raman techniques, such as nonresonant RS,[60–62] UV/RR,[49,87] RDS,[88] SERS,[32] and ROA,[28] have been applied in this field, providing different structural flavors for these biomolecules. Excitation in the UV region holds several advantages.[87]

(1) The overall spectrum is simplified, because only the vibrational modes from heterocyclic components of nucleic acids (purine and pyrimidine bases) are preferentially enhanced.

(2) The relative intensity of bases depends on λ_{ex}, and thus monitoring of particular bases becomes possible. For example, the guanine (G) and adenine (A) residues can be followed selectively using excitation at 250 and 240 nm.[87]

(3) Some modes (for example the $\delta(NH_2)$ mode which is important in H-bonding studies) which are heavily overlapped with conventional excitation can be observed using the UV/RR approach.[87]

Recently, a powerful new Raman technique called vibrational ROA was introduced in the field of nucleic acid studies by Bell et al.[28] The ability of the technique to provide new structural information on conformations of the sugar ring and phosphate backbone has been demonstrated.

The large number of Raman studies on native nucleic acids as well as on model compounds has led to the identification of certain marker bands whose frequency or intensity are sensitive to the structure. In this section we will show what kind of information about nucleic acids can be extracted from the Raman approach. We will try to briefly describe the physical mechanisms leading to the sensitivity of certain Raman modes to the structure and interactions of nucleic acids. Starting from the description of known marker bands we will demonstrate how this method works to solve more practical problems related to interactions of nucleic acids with drugs and proteins.

Four important kinds of information can be extracted from RS studies:

(1) H-bonding between bases;
(2) noncovalent interactions between bases (base stacking);
(3) structural conformations;
(4) intermolecular interactions of proteins, drugs, and metal ions with nucleic acids.

In order to understand the complex vibrational spectra of nucleic acids, the lower molecular weight building blocks must be considered first. Nucleic acids are composed of nucleotides. The main components of nucleotides are a heterocyclic base, a sugar, and a phosphate group. Strong features from heterocycles and bands from symmetric vibrations of the phosphate group dominate in the Raman spectra of nucleic acids. Consequently, only those bands will be considered in this section. Characteristic vibrations of the main groups of nucleic acids as well as important marker bands are summarized in Table 7. Interpretation of the spectra is based on isotopic H/D, $^{14}N/^{15}N$, and $^{16}O/^{18}O$ substitution, comparison with model compounds, normal coordinate analysis, and application of the UV/RR approach.[60–62,86,87] H-bonding interactions can be followed by RS only if certain vibrational modes are concentrated on the motions of atomic groups involved in the bonding. The possible sites for H-bonding interactions are the NH and NH_2 donor groups and C=O and N acceptor groups in base residues as well as oxygen atoms of the phosphate group. Because protons of donor groups are easily exchangeable by deuterons, vibrations of these groups can be unambiguously distinguished by frequency shifts upon H_2O/D_2O solution exchange (Table 7). The strength of the UV/RR approach should be mentioned in this context. While bending vibrations of NH_2 groups (δNH_2) were not resolved using visible excitation,[86] these modes become prominent for A, cytosine (C), and G bases with UV-excited spectra[87] (Table 7). Vibrations of carbonyl groups ($\nu C{=}O$) occur in a rather narrow frequency range (1650–1700 cm^{-1}) and differentiation between $C_2{=}O$ and $C_4{=}O$ modes in thymine (T) and uracil (U) bases was performed by ^{18}O labeling and exploring the high sensitivity of the intensity of bands in the UV/RR excitation range (200–266 nm) on λ_{ex} (Table 7). Noncovalent interactions between bases (stacking) can be followed effectively by UV/RRS. In this case changes in intensity of the bands associated with the vibrations of purine and pyrimidine rings are monitored. Briefly, the mechanism can be described as follows: stacking of heterocyclic rings suppresses ε of the electronic absorption transition in the UV region. Because the RR enhancement is proportional to ε^2, the corresponding intensity of UV/RR bands markedly decreases. For example, the intensity of the breathing motion of the imidazole ring at 730 cm^{-1} in an A base residue decreases upon stacking (Table 7). The third important issue for molecular biologists, where the Raman technique can be valuable, is the structure of nucleic acids. The double helix structure is flexible and several conformational forms labeled as *A*, *B*, *C*, and *Z* were distinguished. Coupled X-ray diffraction (XRD) and Raman studies led to the elucidation of structural conformation marker bands[60,61,88] (Table 7). The conformational sensitivity of the imidazole ring breathing modes at 730 and 679 cm^{-1} observed for A and G base residues (Table 7) arises from the orientation effects around the N–R bond. From two observed symmetric phosphate stretching bands, the vibration of bonded oxygen atoms ($\nu_s PO_2'$) located at 740–850 cm^{-1} is the most sensitive marker band of structural conformations, due to the influence of rotation around the $R{-}O'$ bond on the frequency. The characteristic vibrations of different structural conformers useful for Raman examinations are summarized in Table 8.

Next, application of RS for the analysis of intermolecular interactions between nucleic acids and drugs, proteins or metal ions will be considered. The H-bonding marker bands and vibration of exposed oxygen atoms from the phosphate group ($\nu_s PO_2$) are usually affected by ligand binding. The power of vibrational spectroscopy (Fourier transform infrared (FTIR) and RDS) was recently demonstrated by analysis of drug (aspirin) interactions with DNA.[89] Upshifts in Raman frequencies indicated that at low drug concentrations, aspirin binds through the three main sites in DNA:

(1) the phosphate backbone PO_2 group ($\nu_s PO_2$ shifted from 1093 cm^{-1} in the free state to 1095 cm^{-1} upon binding);

(2) the A N_7 atom (bands at 1342 and 1304 cm^{-1} shifted to 1344 and 1306 cm^{-1});

(3) the T O_2 atom (band at 1661 cm^{-1} shifted to 1670 cm^{-1}).

The Raman investigations of binding between the model dinucleotide compound and protein led to the identification of specific binding sites.[90] In this work UV-excited (240–250 nm) RR spectra have shown the interaction of protein ribonuclease with dinucleotide through H-bonding involving the particular N_3 atom in the C base residue.[90] The marker band at 1526 cm^{-1} was found to be specifically upshifted in bonded dinucleotides (Table 7). Formation of complexes between nucleic acids and proteins was extensively investigated by Benevides et al.[88] Specific marker bands for bonded DNA were recognized. Interactions of drugs (at low concentrations) with DNA can be effectively examined through the SERS approach.[32]

The potential of the SERS technique in analysis of components of nucleic acids at the trace level was also demonstrated.[91,92] RNA[91] and DNA[92] bases were successfully detected in real time in a combined SERS–flow injection analysis (FIA) system. Figure 12 demonstrates the sensitivity of the SERS technique.[92] Characteristic bands from the adenosine base can be easily recognized after a 50-μL injection of solution containing as little as 800 pmol of biomolecules. In

Table 7 Characteristic Raman frequencies of main groups in nucleic acids[a]

Group	Frequency		Approximate description	Structural information
	H_2O	D_2O		
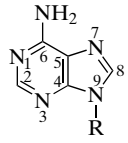	1604 w[b]	1190 m[b]	δNH_2 / δND_2	H-bonding of NH_2 group
	1581 s	1578 s	νC_5C_4, νC_4N_3	
	1509 m	1518 w		
A base	1482 m	1482 m	Vibration of imidazole ring (δC_2H, νN_9C_8, δC_8H)	
	1423 w	1425 w	νC_4N_9, δC_8H	
	1376 m	1376 m		
	1339 vs	1337 vs	Vibration of imidazole ring (νN_7C_5, νC_8N_7)	Marker band of nucleoside conformation
	1309 s	1309 s	νN_9C_8, νN_3C_2, δC_8H	
	1254 w		νN_1C_6, $\nu C_6N_6'$	
	730 m	720	Breathing of imidazole ring ($\delta N_7C_8N_9$, νN_9R)	Marker band for structural conformations; intensity decreases upon stacking; intensity increases with λ_{ex} = 218 nm
	H_2O	D_2O		
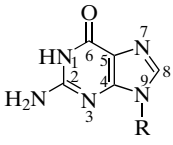	1679 m	1671	νC_6O, νC_5C_6	H-bonding of O_6; shifts to lower frequency when O_6 accepts H-bond
	1603 m[c]	1207 m[c]	$\delta NH_2/\delta ND_2$	H-bonding of NH_2
	1579 s	1579 vs	νC_4N_3, νC_5C_4	
G base	1539 w	1539 m	νC_4N_9, νN_7C_5	
	1489 vs	1480 m	Vibration of imidazole ring (δC_8H, νN_9C_8, νC_8N_7)	H-bonding of N_7; shifts to lower frequency when N_7 accepts a hydrogen bond
	1419 w	1411		
	1364 m	1353 m	νC_8N_7, νN_1C_6	
	1326 m	1329	Vibration of imidazole ring (δC_8H, νC_8N_7)	Marker band of structural conformations
	1179 w	1174 vw		
	854 w	834 vw	νN_9R, νN_3C_2	
	679 w		Breathing of imidazole ring ($\delta N_7C_8N_9$, $\delta C_5N_7C_8$)	Marker band of structural conformations
	H_2O	D_2O		
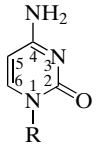	1652 w	1649 w	νC_2O, νC_2N_3	H-bonding
	1605 w[b]	1176 w[b]	δNH_2, $\nu C_4N_4'/\delta ND_2$	H-bonding of NH_2
	1528 s	1526 m	νN_3C_4, νN_1C_2	N_3 atom H-bonding marker band; blue shift reflects bonding
C base	1500 w	1507 s[b]	νN_1C_2, νN_1C_6, νC_2N_3	
	1294 vs	1300	νN_1C_6, νC_5C_6	Marker band of structural conformations
	1250 vs	1250 m	δC_6H, $\nu C_4N_4'$	Marker band of structural conformations
	1210 m			
	988 w	1037 m	ρNH_2, δC_6H	H-bonding of NH_2
	782 vs	771 s	Ring breathing (νN_1R, νC_4N, $\nu C_4C_5C_6$)	Marker band of structural conformation

(*continued overleaf*)

Table 7 (*continued*)

Group	Frequency		Approximate description	Structural information
	H_2O[b]	D_2O[b]		
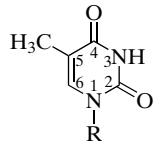 T base	1700 w[d]	1700 w[d]	νC_2O	H-bonding of C_2O
	1655 vs	1657 vs	νC_2O, νC_4O, νC_5C_6	H-bonding of C_2O and C_4O
	1484 w	1482 w		
	1379 vs	1379 vs	νC_6C_5, νC_5C_4, νC_4O	
		1309 m		
		1267 m		
	1244 s	1244 m		Marker band of structural conformations
	1185 m	1159 m		H-bonding
	1023 w	1023 w		
	965 w			
	785 m	785 m	Ring breathing	Marker band of structural conformations
	745 w	745 w		
	H_2O	D_2O		
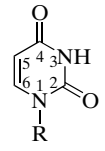 U base	3180 vw[e]	2230 m[e]	$\nu NH/\nu ND$	H-bonding of N_3H
	1686 w[c]	1700 s	νC_2O, νC_2N_3	H-bonding
	1674 m	1654 vs	νC_5C_4, νC_4O	H-bonding
	1628 w	1617 w		
	1476 w[f]	1471 w[f]	νN_1C_2	
	1394 m	1397 m	δN_3H, νC_4O	H-bonding and coordination of N_3H and C_4O groups
	1230 vs		δC_6H, νC_2N_3	H-bonding and coordintion of N_3H group
		1247 vs	νC_4O, δN_3D	D-bonding
		1200 w		
		1138 m	νN_1R	
	1000 w		δC_6H	
		992 m		
	783 m	780 m	Ring breathing	
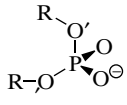 Phosphate group	740–850		$\nu_s PO_2'$	Marker band of structural conformations
	1090–1100		$\nu_s PO_2$	Intensity standard; complexation with metal ions and drugs

[a] Based on Carey,[3] Nishimura and Tsuboi,[60] Thomas,[61] Peticolas,[62] Lord and Thomas,[88] Fodor et al.,[87] Benevides et al.,[86] Neault et al.,[89] and Molina et al.[90]
[b] Observed with $\lambda_{ex} = 266$ nm.
[c] Observed with $\lambda_{ex} = 218$ nm.
[d] Observed with $\lambda_{ex} = 200$ nm.
[e] Observed in crystalline phase.
[f] Very strong with $\lambda_{ex} = 218$ nm.
ν, stretching; ν_s, symmetric stretching; δ, deformation; ρ, rocking.
vw, very weak; w, weak; m, medium; s, strong; vs, very strong.

Table 8 Frequencies (cm^{-1}) of base residues, phosphate, and methylene groups characteristic of nucleic acid structural conformations[a]

Group	Structural conformations				
	A (DNA, RNA)	*B* (DNA)	*C* (DNA)	*Z* (DNA)	*Z* (RNA)
T	642 777 1239	665 748 1208	–	–	–
G	666 1318	682 1333	670	625 1316	640
A	727	727	–	624 729	–
C	780	782 1250	–	–	–
Phosphate					
(ν_s PO_2')	806–813	825–842	875	742–748	–
(ν_s PO_2)	1099 ± 1	1092 ± 1	–	1095 ± 1	–
Methylene					
(δ CH_2)	1418 ± 2	1422 ± 2	–	1425 ± 2	–

[a] Based on data from Twardowski and Anzenbacher,[8] and Benevides et al.[88]

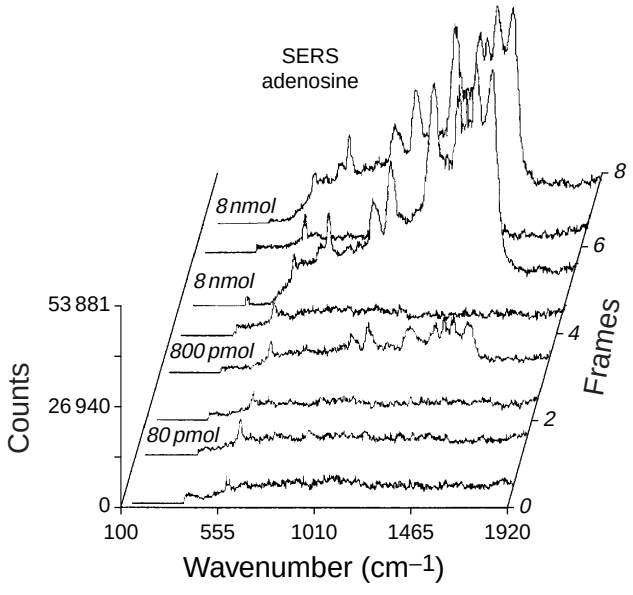

Figure 12 Limits of detection data for the SERS technique under flowing conditions: 50-μL injections of adenosine just prior to frames 2, 4, 6, and 8 representing 80 pmol, 800 pmol, 8 nmol, and 8 nmol, respectively. Frame 2 is below the detection limits for 10 mW of optical power and 10-s integration times. Frames 6 and 8 show the reproducibility for repeat injections. [Reproduced by permission of Society for Applied Spectroscopy from N.J. Pothier, R.K. Force, *Appl. Spectrosc.*, **48**(4), 421–425 (1994).]

addition to the sensitivity, the reproducibility is also impressive and encouraging.

6.4 Lipids and Biomembranes

Lipid bilayers are the general components of biological membranes and serve as a matrix for proteins. RS has proved to be valuable in providing information on phase transitions and order in lipid bilayers as well as on interactions of lipids with membrane perturbants such as proteins or drugs.[63–66] In this section two main problems will be considered:

(1) the relationship between lipid bilayer order parameters and Raman marker bands, and
(2) perturbations induced in the lipid structure by drugs and proteins.

Table 9 summarizes the phospholipid characteristic vibrations and marker bands useful in the structural analysis of lipid bilayers. The structure of a typical phospholipid, dipalmitoylphosphatidylcholine (DPPC), is also shown. Three order parameters, I_{2935}/I_{2880}, I_{2850}/I_{2880}, and I_{1088}/I_{1130} have been shown to be useful for monitoring the self-organization of lipids, the main elements of membrane structure. This is because the C–H stretching (2800–3000 cm^{-1}) and C–C skeletal (1000–1200 cm^{-1}) modes are sensitive to the intermolecular (between chains) and intramolecular (within chain) interactions to a different degree. While C–C skeletal stretches reflect only the changes in **T/G** conformation (intramolecular order), the C–H modes are sensitive to both **T/G** (intramolecular) and lateral chain–chain (intermolecular) interactions. Phase transitions between the bilayer gel and the liquid crystalline phase can be sensitively monitored using the dependence of the appropriate order parameter (Table 9) on the temperature.[63–65]

Interactions of polypeptides, proteins, nucleic acids or drugs with lipid bilayers induce changes in the bilayer structure affecting marker bands and providing the possibility of studies by RS.[65] Depending on the interaction type, ordering or disordering of lipid bilayers was observed. For example it was established that the charged polypeptide poly-L-lysine raises the transition temperature of zwitterionic phosphatidylcholines by 7 °C, suggesting that the interaction is stronger in the gel phase than in the liquid crystalline phase.[65] By contrast, an intrinsic membrane protein, the coat protein of filamentous phage viruses, lowers the phase transition of lecithin bilayers. In general, Raman data indicate that intrinsic membrane proteins, interacting over a large surface area with acyl chains, induce a higher number of **G** conformers within chains and lower the mobility of the chains. On the other hand, the extrinsic proteins interact mainly through the charged groups and in general increase the transition temperature.[65]

Table 9 Characteristic vibrations and marker bands of lipids[a]

Frequency (cm^{-1})	Assignment	Structural information	Order parameters
Structure of typical phospholipid molecule (DPPC)			
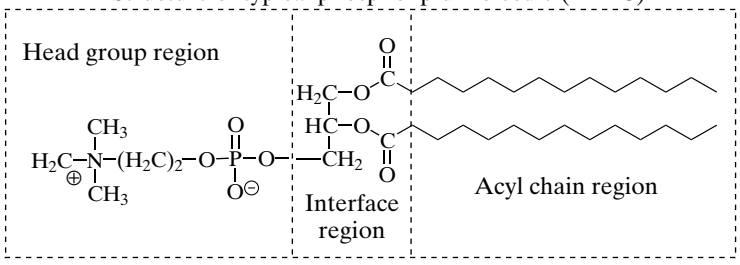			
		Acyl chain region	
2956	$\nu_s(CH_3)$		(1) Parameter I_{2935}/I_{2880} monitors acyl chain disorder/order arising from latera chain–chain interactions and **G/T** isomerization (inter- and intramolecular interactions)
~2935 br	$\nu_s(CH_3)$ FR	Both intra- and intermolecular interactions	
~2847 br	$\nu_s(CH_2)$ FR with 2 $\delta(CH_2)$		
2880 vs	$\nu_{as}(CH_2)$	Both intra- and intermolecular interactions	(2) Parameter I_{2850}/I_{2880} directly monitors acyl chain disorder/order arising from lateral chain–chain interactions (intermolecular interactions)
2847 s	$\nu_s(CH_2)$	Both intra- and intermolecular interactions	
1457 vs	$\delta(CH_2)$	Intermolecular order; acyl chain packing; lattice subcell configurations	(3) Parameter I_{1088}/I_{1130} monitors extent of **G/T** isomerization (intramolecular interactions)
1441 vs	$\delta(CH_2)$		
1296 vs	$\iota(CH_2)$ twist		
1062 s	ν(CC) **T** (out-of-phase)	Intramolecular order (**T/G** changes)	
1100	ν(CC) **T** (in-phase)	Intramolecular order (**T/G** changes)	
1130 s	ν(CC) **T** (in-phase)	Intramolecular order (**T/G** changes)	
~1088 br	ν(CC) **G**	Intramolecular order (**T/G** changes)	
0–400 w	Longitudinal acoustic modes		
		Interface region	
1720 w	ν(CO)		
860–890 w	ν(CC)	Sensitive to packing restrictions at the interface region	
		Head group region	
3042 w	$\nu_{as}(CH_3)$		
~1080 m–w	$\nu_s(PO_2)$ PO_2^- group	Interactions with phosphate group	
1253 vw	$\nu_{as}(PO_2)$ PO_2^- group		
700–800 w	$\nu_s(CN)$	**T/G** conformations of choline group	

[a] Based on Levin,[63] Carey and Salares,[64] Yager and Gaber,[65] and Harrad.[66]
ν_s, symmetric stretching; ν_{as}, asymmetric stretching; δ, deformation; ι, twist.
br, broad; vs, very strong; s, strong; w, weak; m–w, medium–weak; vw, very weak.

6.5 Dynamics in Biomolecules

Understanding the specific functions of biomolecules requires that the dynamics of conformational transitions and bonding with transient species are described in molecular detail. That is, techniques sensitive not only to the structure of biomolecules but also having nanosecond or even picosecond time resolution must be applied. Recent progress in laser technology and optical detection has enabled time-resolved Raman (TR^2) spectroscopy to become one of the most powerful methods in the study of ultrafast biomolecular events.[38,67–69] Due to the intrinsic weakness of the Raman process, the resonance-enhancement phenomenon is usually explored in such studies, and the technique is labeled as TR^3 spectroscopy. There are two essential steps in TR^3 experiments: initiation of the process, and monitoring of the process. Based on the process initiation protocols, the following methods can be distinguished:[67]

(1) rapid-mixing methods;

(2) light-initiated methods using continuous wave (CW) lasers:

- rapid-flow method
- spinning-cell method;

(3) light-initiated methods using pulsed lasers:

- single-pulse method
- dual-pulse method (''pump'' laser beam for initiation of the process and a different wavelength ''probe'' beam for process monitoring).

This TR^3 strategy was successfully applied to numerous biological systems, including heme enzymes, photosynthetic reaction centers, and visual pigments.[67] In the case of heme enzymes, the fast light-driven de-ligation of the CO adduct of the Fe(II) form serves as the initiation step. The following structural transformations of the protein matrix can be examined using either a probe beam at Soret absorption wavelengths to monitor the ν(Fe–N(His)) mode evolution, or a UV-probe beam to monitor changes in aromatic amino acid residues by exploring the UV/RR phenomena. The relationship between the structure of the protein matrix and the ligation state of the heme group in hemoglobin was demonstrated by Rodgers et al.[93] In this work the photolysis of hemoglobin-bound CO was initiated by the 419-nm laser pulse and the delayed 230-nm pulse served for the selective excitation of the transient UV/RR spectra from the Trp and Tyr residues in protein. Similar transient UV/RR spectra of the aromatic amino acid residue Tyr excited using a 240-nm probe beam were successfully applied for the monitoring of millisecond or even nanosecond structural changes in Bacteriordopsin initiated by the 515-nm pump beam.[94] As another interesting example to illustrate how the dynamics of biomolecules can be seen through RRS, the folding and unfolding reaction of cyt *c* will be considered briefly. Recently, a rapid solution-mixer with a 100-μs time resolution was coupled with RRS, providing the ability to monitor the folding/unfolding dynamics of cyt *c* with sub-millisecond resolution.[95] The initiation of folding was achieved by rapid dilution of denaturant-unfolded protein with an excess of buffer solution. The low-frequency mode (at 394 cm^{-1}) of the ferrous heme form associated with the vibration of the covalent thioether linkage between the heme and the polypeptide [$\delta(C_\beta–C_\alpha–S)$][81] served as the folding/unfolding marker band. Figure 13 demonstrates the development of the low-frequency spectrum during the folding reaction.[95] In the native folded state this band is intense, because of the twisting and tension of the heme group induced by the surrounding polypeptide. This band virtually disappears in the unfolded state, due to the more planar structure of the heme group. It is clear

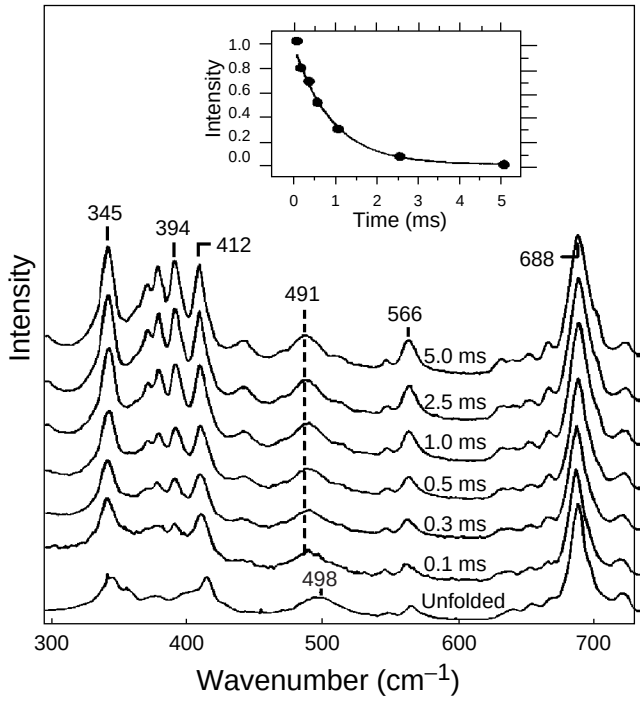

Figure 13 Time evolution of the low-frequency RR spectra of cyt *c* following the initiation of folding. The ferrous CO-bound form of the protein was unfolded in 7.0 M guanidine hydrochloride (Gdn-HCl) and refolded in 0.7 M Gdn-HCl at neutral pH. The insert is the decay curve of the unfolded form obtained by monitoring the line at 394 cm^{-1}. The line at 491–498 cm^{-1} is assigned as the Fe–CO stretching mode. [Reproduced by permission of American Chemical Society from S.-R. Yeh, S. Han, D.L. Rousseau, *Acc. Chem. Res.*, **31**(11), 727–736 (1998).]

from Figure 13 that microscopic details of the folding reaction can be monitored in real time by RRS.

6.6 Raman Microscopic Analysis of Biological Systems

The structural sensitivity inherent to vibrational spectroscopy is combined with spatial resolution of 0.1–2 μm in MRS. Such a combination makes the MRS technique extremely attractive for applications in molecular biology such as analysis of cell components,[35,70] drug distribution,[96] or identification of the constituents of the cell membrane.[70] Recent introduction of UV/RRS in this field has considerably increased the selectivity and sensitivity of the MRS technique for molecules absorbing in the 180–300-nm region.[35] There is a growing interest in the application of the MRS approach to in vivo analysis of biomaterials.

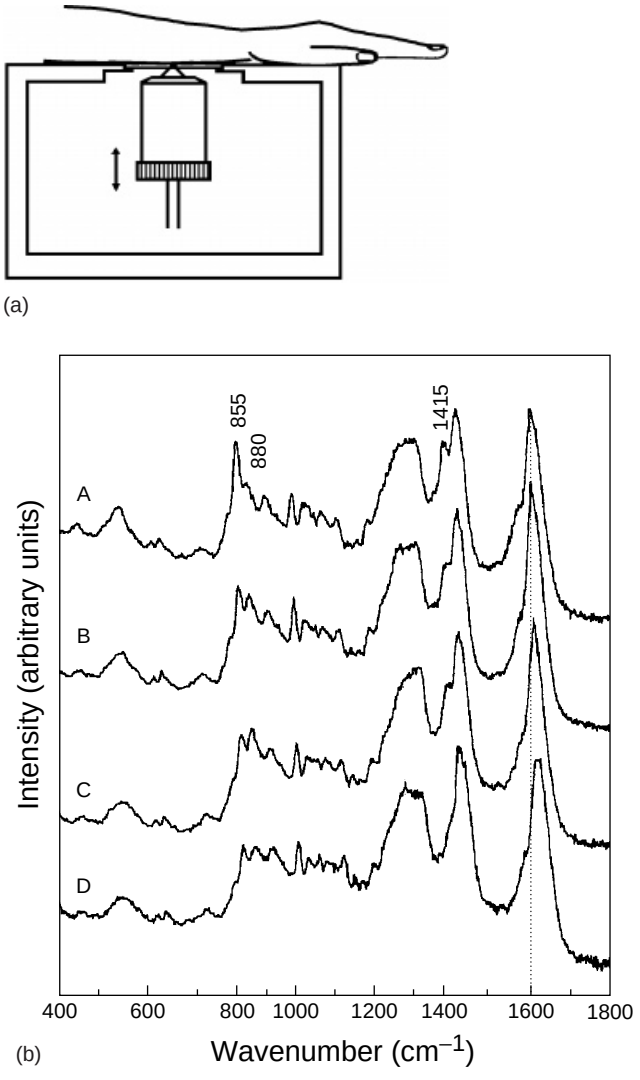

Figure 14 In vivo RS measurements of human skin: (a) inverted microscope used for measurements; (b) in vivo Raman spectra of the stratum corneum at different anatomical regions measured at the skin surface. The spectra are from a single person: line A, fingerprint; line B, thenar; line C, dorsal surface of the middle finger; line D, volar aspects of the forearm. Spectrum A was measured in 180 s, all other spectra in 120 s. Objective 63 × 0.7. Excitation wavelength 850 nm, power 100 mW. [Reproduced by permission of John Wiley & Sons from P.J. Caspers, G.W. Lucassen, R. Wolthuis, H.A. Bruining, G.J. Puppels, *Biospectroscopy*, **4**, 31–39 (1998).]

Caspers et al.[97] have studied human skin in vitro and in vivo using a confocal Raman microspectrometer. The inverted microscope principle shown in Figure 14(a) was used for the in vivo Raman measurements. A CaF_2 window serves as support for the sample of interest. The required depth with the 9-μm resolution was achieved by translating the objective vertically. It was possible to record Raman spectra with 0-μm depth directly from the skin. Differences in several spectral regions (1640–1680 cm^{-1} amide I region; 1240–1300 cm^{-1} amide III region; bands at 1415 cm^{-1}, 855 and 880 cm^{-1}) can be easily identified from Figure 14(b) by comparing Raman spectra from different places on the skin. Variations in human moisturizing factors and lipid contents were suggested as possible reasons for some observed spectral variations. These results demonstrate the possibility of obtaining noninvasive vibrational spectra from living skin that contain information about the composition and structure of different places and layers of the skin.

7 FUTURE DEVELOPMENTS OF RAMAN SPECTROSCOPY

Advances in laser technology will make UV lasers cheaper. Because electronic systems of important biological structures such as bases of nucleic acids, aromatic amino acid residues, and peptide groups have electronic transitions in the 180–300 nm range, the applications of UV/RRS will rapidly expand.

There is some hope that progress in the preparation of more reproducible and stable surfaces for SERS will expand the application of this technique to monitoring traces of biomolecules in more applied investigations.

With a further reduction in the price of combined FTR/FTIR instruments, routine characterization of biosamples using vibrational spectra, especially at higher concentrations, will become common.

The progress in quantitative analysis by RS is connected to the development of simple sensitivity-calibration procedures and the evolution of detection systems.

The combination of highly sensitive Raman techniques (RR, UV/RR and SERS/SERRS) with the submicrometer spatial resolution ability of MRS is encouraging,

and holds promise for the wide expansion of vibrational spectroscopy to the analysis of single living cell components, leading to direct biomedical applications. However, problems with the laser-induced photodegradation of the sample will be apparent in this field.

8 COMPARISON WITH OTHER SPECTROSCOPIC METHODS

In this section the structural information and sensitivity of RS will be compared with IR, ultraviolet/visible (UV/VIS) absorption, and Fl techniques. An important feature of modern RS is the variety of techniques for exploring the Raman phenomenon (Table 1). Thus in making any comparison the technique will be specified. We will focus our attention on RS and IR techniques.

8.1 Structural Information

The main disadvantages of UV/VIS and Fl techniques are their lack of structural sensitivity. In contrast, two vibrational spectroscopic techniques, RS and IR, are able to provide rich structural information on the molecular systems. The resolution of vibrational spectroscopies is several orders of magnitude higher than for UV/VIS or Fl. For this reason we will concentrate on a comparison of the IR and RS methods. The first advantage of RS is the weak Raman scattering from water, in contrast to IR absorption. Therefore biomolecules can be studied in their natural environment. Secondly, resonance and surface-enhancement effects are possible for RS, providing the possibility of selectively monitoring the structure of certain molecular groups or adsorbed species. This is extremely important for large molecular systems such as biomolecules, due to interpretation problems. Third, ρ measurements provide an elegant way for assigning the vibrations according to their symmetry. Fourth, Raman spectra can be easily recorded over a wide frequency range ($100-4000\,cm^{-1}$), in contrast to IR spectra. Thus structurally important, low-frequency, metal–ligand vibrations can be studied using RS.

8.2 Sensitivity

The main disadvantage of RS compared with IR for analytical applications is the difficulty in establishing the analyte concentration from the Raman intensity, because the value of the Raman cross-section of an analyte molecule is usually unknown or is difficult to estimate. In addition, Raman intensities depend on many factors.[1] For this reason internal standard methods are usually employed in RS.[1] The sensitivity of IR spectroscopy, for example in studies of protein–ligand binding, is of the order of $\sim 10^{-3}$ M[52] (by monitoring the ligand internal mode); this is higher than nonresonance RS. In the case of RRS, however, the same protein–ligand systems can be monitored by low-frequency vibrations at concentrations in the order of $\sim 10^{-5}$ M.[52] The most sensitive technique, used for the determination of trace levels of fluorophores or even single molecules, is Fl spectroscopy. However, single molecules have also been detected recently using SERS and SERRS.[10,11] High sensitivity and also structural specificity were evident from these experiments.

Recently the sensitivity limits for the analysis of proteins with nonresonant RS and RDS were considered using a CCD detection system and an axial transmissive Raman spectrometer.[98] Good spectra were collected from 100-μM concentrations of protein. Finally, as can be seen from Figure 12, it is possible to detect adenosine at the 800-pmol level in a 50-μL sample using the SERS technique.

ACKNOWLEDGMENTS

The author would like to thank Dr Adolfas Gaigalas, Dr Vincent Vilker, and Dr Vytas Reipa from the Bioprocess and Engineering group at NIST (Gaithersburg, USA) for stimulating discussions and support during his stay at NIST and later. He also thanks Dr Rimas Jakubėnas for useful discussions and assistance in preparing figures.

ABBREVIATIONS AND ACRONYMS

A	Adenine
Az	Azurin
C	Cytosine
CCD	Charge-coupled Device
CN	Coordination Number
CT	Charge Transfer
CW	Continuous Wave
Cys	Cysteine
cyt *c*	Cytochrome *c*
DPPC	Dipalmitoylphosphatidylcholine
FIA	Flow Injection Analysis
FR	Fermi Resonance
Fl	Fluorescence
FTIR	Fourier Transform Infrared
FTR	Fourier Transform Raman
FTRS	Fourier Transform Raman Spectroscopy
fwhm	Full Width at Half-maximum
G	Guanine
G	Gauche
Gdn-HCl	Guanidine Hydrochloride

Gly	Glycine
His	Histidine
HS	High-spin
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
LS	Low-spin
Mb	Myoglobin
MRS	Micro-Raman Spectroscopy
NIR	Near-infrared
NIST	National Institute of Standards and Technology
Phe	Phenylalanine
Pyr	Pyrrole
RDS	Raman Difference Spectroscopy
ROA	Raman Optical Activity
RR	Resonance Raman
RRS	Resonance Raman Spectroscopy
RS	Raman Spectroscopy
SER	Surface-enhanced Raman
SERR	Surface-enhanced Resonance Raman
SERRS	Surface-enhanced Resonance Raman Spectroscopy
SERS	Surface-enhanced Raman Spectroscopy
T	Thymine
T	Trans
TR^2	Time-resolved Raman
TR^3	Time-resolved Resonance Raman
Trp	Tryptophan
Tyr	Tyrosine
U	Uracil
UV	Ultraviolet
UV/RR	Ultraviolet Resonance Raman
UV/RRS	Ultraviolet Resonance Raman Spectroscopy
UV/VIS	Ultraviolet/Visible
VIS/RRS	Visible Resonance Raman Spectroscopy
XRD	X-ray Diffraction
3-D	Three-dimensional
5C	Five-coordinated
6C	Six-coordinated

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy in Microbiology • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by • Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Infrared Spectroscopy of Biological Applications • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Drugs of Abuse, Analysis of • Infrared Spectroscopy in Clinical Chemistry • Laboratory Instruments in Clinical Chemistry, Principles of • Lipid Analysis for Important Clinical Conditions • Nucleic Acid Analysis in Clinical Chemistry

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Enzyme Analysis and Bioassays in Food Analysis • Lipid Analyses in Food • Proteins, Peptides, and Amino Acids Analysis in Food

Industrial Hygiene **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Molecules, Properties and Detection of Single • DNA Structures of Biological Relevance, Studies of Unusual Sequences • RNA Tertiary Structure

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis • Protein–Drug Interactions • Protein–Oligonucleotide Interactions

Pharmaceuticals and Drugs **(Volume 8)**
Vibrational Spectroscopy in Drug Discovery, Development and Production

Process Instrumental Methods **(Volume 9)**
Raman Spectroscopy in Process Analysis

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Kinetic Determinations **(Volume 12)**
Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation • Raman Microscopy and Imaging • Raman Scattering, Fundamentals

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quantitative Spectroscopic Calibration • Traceability in Analytical Chemistry

REFERENCES

1. *Infrared and Raman Spectroscopy: Methods and Applications*, eds. B. Schrader, D. Bougeard, John Wiley & Sons, New York, 1995.
2. J.R. Ferraro, K. Nakamoto, *Introductory Raman Spectroscopy*, Academic Press, Boston, 1994.
3. P.R. Carey, *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York, 1982.
4. A. Tu, *Raman Spectroscopy in Biology. Principles and Applications*, John Wiley & Sons, New York, 1982.
5. F.S. Parker, *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*, Plenum Press, New York, 1983.
6. *Spectroscopy of Biological Systems*, eds. R.J. Clark, R.E. Hester, Advances in Spectroscopy, John Wiley & Sons, Chichester, Vol. 13, 1986.
7. *Biological Applications of Raman Spectroscopy*, ed. T.G. Spiro, John Wiley & Sons, New York, Vol. 1–3, 1987 (Vol. 1 and 2) and 1988 (Vol. 3).
8. J. Twardowski, P. Anzenbacher, *Raman and IR Spectroscopy in Biology and Biochemistry*, Ellis Horwood Ltd, Chichester, 1995.
9. *Modern Techniques in Raman Spectroscopy*, ed. J.J. Laserna, John Wiley & Sons, Chichester, 1996.
10. S. Nie, S.R. Emory, 'Probing Single Molecules and Single Nanoparticles by Surface-enhanced Raman Scattering', *Science*, **275**, 1102–1106 (1997).
11. K. Kneipp, Y. Wang, H. Kneipp, L.T. Perelman, I. Itzkan, R.R. Dasari, M.S. Feld, 'Single Molecule Detection Using Surface-enhanced Raman Scattering (SERS)', *Phys. Rev. Lett.*, **78**(9), 1667–1670 (1997).
12. V. Deckert, D. Zeisel, R. Zenobi, T. Vo-Dinh, 'Near-field Surface-enhanced Raman Imaging of Dye-labeled DNA with 100-nm Resolution', *Anal. Chem.*, **70**(13), 2646–2650 (1998).
13. W.L. Peticolas, 'Raman Spectroscopy of DNA and Proteins', *Methods Enzymol.*, **246**, 389–416 (1995).
14. S. Keller, B. Schrader, A. Hoffmann, W. Schrader, K. Metz, A. Rehlaender, J. Pahnke, M. Ruwe, W. Budach, 'Application of Near-infrared/Fourier Transform Raman Spectroscopy in Medical Research', *J. Raman Spectrosc.*, **25**, 663–671 (1994).
15. C.V. Raman, K.S. Krishnan, 'A New Type of Secondary Radiation', *Nature*, **121**, 501–502 (1928).
16. G. Landsberg, L. Mandelstam, 'Eine neue Erscheinung bei der Lichtzerstreuung in Kristallen', *Naturwissenschaften*, **28**, 557–558 (1928).
17. P.P. Shorygin, 'The Resonance Combinatorial Light Scattering', *Proc. Acad. Sci. USSR.*, **87**(2), 201–204 (1952) (in Russian).
18. D. Garfinkel, J.T. Edsall, 'Raman Spectra of Amino Acids and Related Compounds. X. The Raman Spectra of Certain Peptides and of Lysozyme', *J. Am. Chem. Soc.*, **80**, 3818–3823 (1958).
19. T. Miyazawa, T. Shimanouchi, S. Mizushima, 'Normal Vibrations of N-methylacetamide', *J. Chem. Phys.*, **29**(3), 611–616 (1958).
20. M.C. Tobin, 'Raman Spectra of Crystalline Lysozyme, Pepsin and Alpha Chymotrypsin', *Science*, **161**(3836), 68–69 (1968).
21. T.V. Long, T.M. Loehr, 'The Possible Determination of Iron Coordination in Nonheme Iron Proteins Using Laser Raman Spectroscopy. Rubredoxin', *J. Am. Chem. Soc.*, **92**(21), 6384–6386 (1970).
22. T.S. Spiro, T.C. Strekas, 'Resonance Raman Spectra of Hemoglobin and Cytochrome c: Inverse Polarization and Vibronic Scattering', *Proc. Natl. Acad. Sci. USA.*, **69**(9), 2622–2626 (1972).
23. J.L. Lipperd, W.L. Peticolas, 'Laser Raman Investigation of the Effect of Cholesterol on Conformational Changes in Dipalmitoyl Lecithin Multilayers', *Proc. Natl. Acad. Sci. USA.*, **68**(7), 1572–1576 (1971).
24. L.D. Barron, A.D. Buckingham, 'Rayleigh and Raman Scattering from Optically Active Molecules', *Mol. Phys.*, **20**, 1111–1119 (1971).
25. L.D. Barron, M.P. Bogaard, A.D. Buckingham, 'Raman Scattering by Circularly Polarized Light by Optically Active Molecules', *J. Am. Chem. Soc.*, **95**(2), 603–605 (1973).
26. M. Fleischmann, P.J. Hendra, A.J. McQuillan, 'Raman Spectra of Pyridine Adsorbed at a Silver Electrode', *Chem. Phys. Lett.*, **26**, 163–166 (1974).
27. T.M. Cotton, S.G. Schultz, R.P. Van Duyne, 'Surface-enhanced Resonance Raman Scattering from Cytochrome *c* and Myoglobin Adsorbed on a Silver Electrode', *J. Am. Chem. Soc.*, **102**(27), 7960–7962 (1980).
28. A.F. Bell, L. Hecht, L.D. Barron, 'Vibrational Raman Optical Activity of DNA and RNA', *J. Am. Chem. Soc.*, **120**(23), 5820–5821 (1998).
29. R. Callender, H. Deng, 'Nonresonance Raman Difference Spectroscopy: A General Probe of Protein Structure, Ligand Binding, Enzymatic Catalysis, and the Structures of Other Biomacromolecules', *Ann. Rev. Biophys. Biomol. Struct.*, **23**, 215–245 (1994).
30. P.R. Carey, 'Raman Spectroscopy in Enzymology: The First 25 Years', *J. Raman Spectrosc.*, **29**, 7–14 (1998).
31. S.A. Asher, 'UV Resonance Raman Spectroscopy for Analytical, Physical and Biophysical Chemistry', *Anal. Chem.*, **65**(4), 201A–210A (1993).
32. T.M. Cotton, J.-H. Kim, G.D. Chumanov, 'Application of Surface-enhanced Raman Spectroscopy to Biological Systems', *J. Raman Spectrosc.*, **22**, 729–742 (1991).

33. T.M. Cotton, J.-H. Kim, R.E. Holt, 'Surface-enhanced Resonance Raman Scattering (SERRS) Spectroscopy: A Probe of Biomolecular Structure and Bonding at Surfaces', *Adv. Biophys. Chem.*, **2**, 115–147 (1992).
34. J.R. Kincaid, 'Structure and Dynamics of Transient Species Using Time-resolved Resonance Raman Spectroscopy', *Methods Enzymol.*, **246**, 460–501 (1995).
35. V. Pajcini, C.H. Munro, R.W. Bormett, R.E. Witkowski, S.A. Asher, 'UV Raman Microspectroscopy: Spectral and Spatial Selectivity with Sensitivity and Simplicity', *Appl. Spectrosc.*, **51**(1), 81–86 (1997).
36. L.D. Barron, L. Hecht, A.F. Bell, 'Raman Optical Activity: An Incisive New Probe of the Structure and Dynamics of Biomolecules', *Sci. Progr.*, **81**(1), 17–34 (1998).
37. G. Niaura, A.K. Gaigalas, V.L. Vilker, 'Non-resonant SERS Study of the Adsorption of Cytochrome *c* on a Silver Electrode', *J. Electroanal. Chem.*, **416**, 167–178 (1996).
38. T.M. Loehr, 'Recent Advances in Resonant Raman Spectroscopy', in *Spectroscopic Methods in Bioinorganic Chemistry*, eds. E.I. Solomon, K.O. Hodgson, American Chemical Society, Washington, ACS Symposium Series 692, 136–153, 1998.
39. C.R. Andrew, J. Han, T. den Blaauwen, G. van Pouderoyen, E. Vijgenboom, G.W. Canters, T.M. Loehr, J. Sanders-Loehr, 'Cysteine Ligand Vibrations are Responsible for the Complex Resonance Raman Spectrum of Azurin', *J. Biol. Inorg. Chem.*, **2**, 98–107 (1997).
40. W. Kiefer, H.J. Bernstein, 'A Cell for Resonance Raman Excitation with Lasers in Liquids', *Appl. Spectrosc.*, **25**(4), 500–501 (1971).
41. P. Hildebrandt, K.A. Macor, R.S. Czernuszewicz, 'Novel Cylindrical Rotating Electrode for Anaerobic Surface-enhanced Raman Spectroscopy', *J. Raman Spectrosc.*, **19**, 65–69 (1988).
42. G. Niaura, A.K. Gaigalas, V.L. Vilker, 'Moving Spectroelectrochemical Cell for Surface Raman Spectroscopy', *J. Raman Spectrosc.*, **28**, 1009–1011 (1997).
43. S. Mamedov, A. Kisliuk, S. Loheider, D. Quitmann, 'New Scanning Beam Technique for Raman Spectroscopy', *Appl. Spectrosc.*, **49**(8), 1199–1200 (1995).
44. H.-O. Hamaguchi, 'Calibration Multichannel Raman Spectrometers', *Appl. Spectrosc. Rev.*, **24**(1 & 2), 137–174 (1988).
45. D.P. Strommen, K. Nakamoto, *Laboratory Raman Spectroscopy*, John Wiley & Sons, New York, 1984.
46. M. Fryling, C.J. Frank, R.L. McCreery, 'Intensity Calibration and Sensitivity Comparisons for CCD/Raman Spectrometers', *Appl. Spectrosc.*, **47**(12), 1965–1974 (1993).
47. A.T. Tu, 'Peptide Backbone Conformation and Microenvironment of Protein Side Chains', in *Spectroscopy of Biological Systems*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, 47–112, 1986.
48. I. Harada, H. Takeuchi, 'Raman and Ultraviolet Resonance Raman Spectra of Proteins and Related Compounds', in *Spectroscopy of Biological Systems*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, 113–175, 1986.
49. T.J. Thamann, 'Probing Local Protein Structure with Ultraviolet Resonance Raman Spectroscopy', in *Spectroscopic Methods for Determination of Protein Structure in Solution*, ed. H.A. Havel, VCH, New York, 96–134, 1996.
50. S. Krimm, J. Bandekar, 'Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins', *Adv. Protein Chem.*, **38**, 181–364 (1986).
51. L.G. Tensmeyer, E.W. Kauffman, II, 'Protein Structure as Revealed by Nonresonance Raman Spectroscopy', in *Spectroscopic Methods for Determination of Protein Structure in Solution*, ed. H.A. Havel, VCH, New York, 69–95, 1996.
52. N.-T. Yu, 'Resonance Raman Studies of Ligand Binding', *Methods Enzymol.*, **130**, 350–409 (1986).
53. N.-T. Yu, E.A. Kerr, 'Vibrational Modes of Coordinated CO, CN^-, O_2 and NO', in *Biological Applications of Raman Spectroscopy*, ed. T.G. Spiro, John Wiley & Sons, New York, 39–95, Vol. 3, 1988.
54. T.G. Spiro, X.-Y. Li, 'Resonance Raman Spectroscopy of Metalloporphyrins', in *Biological Applications of Raman Spectroscopy*, ed. T.G. Spiro, John Wiley & Sons, New York, 1–37, Vol. 3, 1988.
55. P. Hildebrandt, M. Stockburger, 'Conformational Studies of Biological Molecules by Surface Enhanced Raman Spectroscopy', in *Raman Spectroscopy Sixty Years On, Vibrational Spectra and Structure*, Elsevier, Amsterdam, eds. H.D. Bist, J.R. During, J.F. Sullivan, 443–466, Vol. 17A, 1989.
56. T.G. Spiro, R.S. Czernuszewicz, 'Resonance Raman Spectroscopy of Metalloproteins', *Methods Enzymol.*, **246**, 416–460 (1995).
57. P. Hildebrandt, in *Relationships Between Structure and Function of Cytochrome P-450–Experiments, Calculations, Models*, eds. K. Ruckpaul, H. Rein, Akademie Verlag, Berlin, 166–215, 1992.
58. C.R. Andrew, J. Sanders-Loehr, 'Copper–Sulfur Proteins: Using Raman Spectroscopy to Predict Coordination Geometry', *Acc. Chem. Res.*, **29**, 365–372 (1996).
59. P. Carey, 'Raman Spectroscopy in Enzymology: The First 25 Years', *J. Raman Spectrosc.*, **29**, 7–14 (1998).
60. Y. Nishimura, M. Tsuboi, 'Local Conformations and Polymorphism of DNA Duplexes as Revealed by Their Raman Spectra', in *Spectroscopy of Biological Systems*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, Advances in Spectroscopy, 177–232, Vol. 13, 1986.
61. G.J. Thomas, Jr, 'Applications of Raman Spectroscopy in Structural Studies of Viruses, Nucleoproteins and Their Constituents', in *Spectroscopy of Biological Systems*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, Advances in Spectroscopy, Chapter 5, 233–309, Vol. 13, 1986.

62. W.L. Peticolas, 'Raman Spectroscopy of DNA and Proteins', *Methods Enzymol.*, **246**, 389–416 (1995).
63. I.W. Levin, *Vibrational Spectroscopy of Membrane Assemblies*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Heyden, Advances in Infrared and Raman Spectroscopy, 1–48, Vol. 11, 1984.
64. P.R. Carey, V.R. Salares, *Raman and Resonance Raman Studies of Biological Systems*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Heyden, Advances in Infrared and Raman Spectroscopy, 1–58, Vol. 7, 1984.
65. P. Yager, B.P. Gaber, 'Membranes', in *Biological Applications of Raman Spectroscopy, Raman Spectra and Conformations of Biological Macromolecules*, ed. T.G. Spiro, John Wiley & Sons, New York, 203–261, Vol 1, 1987.
66. M. Harrad, 'Study of the Configuration of Phospholipids (As Models for Biological Membranes) by Raman Spectroscopy: Intensity and Polarization', in *Spectroscopy of Biological Molecules*, eds. A.J.P. Alix, L. Bernard, M. Manfait, John Wiley & Sons, Chichester, 265–270, 1985.
67. J.R. Kincaid, 'Structure and Dynamics of Transient Species Using Time-resolved Resonance Raman Spectroscopy', *Methods Enzymol.*, **246**, 460–501 (1995).
68. M.R. Ondrias, M.C. Simpson, R.W. Larsen, 'Time-resolved Resonance Raman Spectroscopy', in *Modern Techniques in Raman Spectroscopy*, ed. J.J. Laserna, John Wiley & Sons, Chichester, 343–386, 1996.
69. S.E.J. Bell, 'Time-resolved Resonance Raman Spectroscopy', *Analyst*, **121**, 107R–120R (1996).
70. G. Turrell, P. Dhamelincourt, 'Micro-Raman Spectroscopy', in *Modern Techniques in Raman Spectroscopy*, ed. J.J. Laserna, John Wiley & Sons, Chichester, 109–142, 1996.
71. D. Garfinkel, J.T. Edsall, 'Raman Spectra of Amino Acids and Related Compounds. VIII. Raman and Infrared Spectra of Imidazole, 4-Methylimidazole and Histidine', *J. Am. Chem. Soc.*, **80**(15), 3807–3812 (1958).
72. M.N. Siamwiza, R.C. Lord, M.C. Chen, 'Interpretation of the Doublet at 850 and 830 cm^{-1} in the Raman Spectra of Tyrosil Residues in Proteins and Certain Model Compounds', *Biochemistry*, **22**(14), 4870–4876 (1975).
73. I. Ashikawa, K. Itoh, 'Raman Spectra of Polypeptides Containing L-Histidine Residues and Tautomerism of Imidazole Side Chain', *Biopolymers*, **18**, 1859–1876 (1979).
74. I. Harada, T. Miura, H. Takeuchi, 'Origin of the Doublet at 1360 and 1340 cm^{-1} in the Raman Spectra of Tryptophan and Related Compounds', *Spectrochim. Acta*, **42A**(2/3), 307–312 (1986).
75. H. Takeuchi, Y. Kimura, I. Koitabashi, I. Harada, 'Raman Bands of *N*-Deuterated Histidinium as Markers of Conformation and Hydrogen Bonding', *J. Raman Spectrosc.*, **22**, 233–236 (1991).
76. W.B. Fischer, H.E. Eysel, 'Polarized Raman Spectra and Intensities of Aromatic Amino Acids Phenylalanine, Tyrosine and Tryptophan', *Spectrochim. Acta*, **48A**(5), 725–732 (1992).
77. T. Miura, T. Satoh, A. Hori-i, H. Takeuchi, 'Raman Marker Bands of Metal Coordination Sites of Histidine Side Chains in Peptides and Proteins', *J. Raman Spectrosc.*, **29**, 41–47 (1998).
78. K.T. Yue, M. Lee, J. Zheng, R. Callender, 'The Determination of the pK_a of Histidine Residues in Proteins by Raman Difference Spectroscopy', *Biochim. Biophys. Acta*, **1078**, 296–302 (1991).
79. R.P. Rava, T.G. Spiro, 'Resonance Enhancement in the Ultraviolet Raman Spectra of Aromatic Amino Acids', *J. Phys. Chem.*, **89**(10), 1856–1861 (1985).
80. D.S. Caswell, T.G. Spiro, 'Ultraviolet Resonance Raman Spectroscopy of Imidazole, Histidine, and Cu(imidazole)$_4^{2+}$: Implications of Protein Studies', *J. Am. Chem. Soc.*, **108**(21), 6470–6477 (1986).
81. S. Hu, I.K. Morris, K.M. Smith, T.G. Spiro, 'Complete Assignment of Cytochrome *c* Resonance Raman Spectra via Enzymatic Reconstruction with Isotopically Labeled Hemes', *J. Am. Chem. Soc.*, **115**(26), 12 446–12 458 (1993).
82. H. Nar, A. Messerschmidt, R. Huber, M. van de Kamp, G.W. Canters, 'Crystal Structure of Oxidized *Pseudomonas aeruginosa* Azurin at pH 5.5 and pH 9.0. A pH-induced Conformational Transition Involves a Peptide Bond Flip', *J. Mol. Biol.*, **221**, 765–772 (1991).
83. O. Siiman, N.M. Young, P.R. Carey, 'Resonance Raman Spectra of Blue Copper Proteins and the Nature of Their Copper Sites', *J. Am. Chem. Soc.*, **98**(3), 744–748 (1976).
84. L. Nestor, J.A. Larrabee, G. Woolery, B. Reinhammar, T.G. Spiro, 'Resonance Raman Spectra of Blue Copper Proteins: Assignments from Normal Mode Calculations and Copper-63/Copper-65 and H_2O/D_2O Shifts for Stellacyanin and Laccase', *Biochemistry*, **23**(6), 1084–1093 (1984).
85. C.R. Andrew, H. Yeom, J.S. Valentine, B.G. Karlsson, N. Bonander, G. van Pouderoyen, G.W. Canters, T.M. Loehr, J. Sanders-Loehr, 'Raman Spectroscopy as an Indicator of Cu–S Bond Length in Type 1 and Type 2 Copper Cysteinate Proteins', *J. Am. Chem. Soc.*, **116**(25), 11 489–11 498 (1994).
86. R.C. Lord, G.J. Thomas, Jr, 'Raman Spectral Studies of Nucleic Acids and Related Molecules – I. Ribonucleic Acid Derivatives', *Spectrochim. Acta*, **23A**, 2551–2591 (1967).
87. S.P.A. Fodor, R.P. Rava, T.R. Hays, T.G. Spiro, 'Ultraviolet Resonance Raman Spectroscopy of the Nucleotides with 266-, 240-, 218-, 200-nm Pulsed Laser Excitation', *J. Am. Chem. Soc.*, **107**(6), 1520–1529 (1985).
88. J.M. Benevides, M.A. Weiss, G.J. Thomas, Jr, 'An Altered Specificity Mutation in the λ Repressor Induces Global Reorganization of the Protein–DNA Interface', *J. Biol. Chem.*, **629**, 10 869–10 878 (1994).
89. J.F. Neault, M. Naoui, M. Manfait, H.A. Tajjmir-Riahi, 'Aspirin–DNA Interaction Studied by FTIR and Laser Raman Difference Spectroscopy', *FEBS Lett.*, **382**, 26–30 (1996).

90. M. Molina, Y. Wang, R. Purrello, T.G. Spiro, 'Dinucleotide Binding to RNase Probed with UV Resonance Raman Spectroscopy', *J. Raman Spectrosc.*, **22**, 205–209 (1991).
91. F. Ni, R. Sheng, T.M. Cotton, 'Flow Injection Analysis and Real-time Detection of RNA Bases by Surface-enhanced Raman Spectroscopy', *Anal. Chem.*, **62**(18), 1958–1963 (1990).
92. N.J. Pothier, R.K. Force, 'Detection of Biologically Important Compounds in Flowing Aqueous Streams by Surface-enhanced Raman Spectroscopy at a Silver Electrode', *Appl. Spectrosc.*, **48**(4), 421–425 (1994).
93. K.R. Rodgers, C. Su, S. Subramaniam, T.G. Spiro, 'Hemoglobin R → T Structural Dynamics from Simultaneous Monitoring of Tyrosine and Tryptophan Time-resolved UV Resonance Raman Signals', *J. Am. Chem. Soc.*, **114**(10), 3697–3709 (1992).
94. J.B. Ames, M. Ros, J. Lugtenburg, R.A. Mathies, 'Time-resolved Ultraviolet Resonance Raman Studies of Protein Structure–Application to Bacteriorhodopsin', *Biochemistry*, **31**(23), 5328–5334 (1992).
95. S.-R. Yen, S. Han, D.L. Rosseau, 'Cytochrome *c* Folding and Unfolding: A Biphasic Mechanism', *Acc. Chem. Res.*, **31**(11), 727–736 (1998).
96. N.M. Sijtsema, S.D. Wouters, C.J. de Grauw, C. Otto, J. Greve, 'Confocal Direct Imaging Raman Microscope: Design and Applications in Biology', *Appl. Spectrosc.*, **52**(3), 348–355 (1998).
97. P.J. Caspers, G.W. Lucassen, R. Wolthuis, H.A. Bruining, G.J. Puppels, 'In Vitro and In Vivo Raman Spectroscopy of Human Skin', *Biospectroscopy*, **4**, 31–39 (1998).
98. J. Dong, D. Dinakarpandian, P.R. Carey, 'Extending the Raman Analysis of Biological Samples to the 100 Micromolar Concentration Range', *Appl. Spectrosc.*, **52**(8), 1117–1122 (1998).

Single Biomolecule Detection and Characterization

Karl Otto Greulich
Institute of Molecular Biotechnology, Jena, Germany

1 Introduction 1
2 History 1
3 Preparation and Handling of Single Molecules 2
3.1 Microfluidics for Single Molecule Handling 2
3.2 Preparation of Single Biological Macromolecules 2
3.3 Laser Micromanipulation for Handling of Individual Macromolecules 2
4 Interaction of Light with Single Molecules 3
5 Fluorescence Correlation Spectroscopy as a Quasi Single Molecule Technique 3
6 Microscopy of Single Molecules 3
6.1 Near-field Microscopies 3
6.2 Very Highly Sensitive Fluorescence: Dyes and Detectors 4
6.3 Classical Far-field Microscopy for High Temporal Resolution 4
7 Single (Nonbiological) Molecule Spectroscopy 4
7.1 Reaching the Limits of Spectroscopy 4
7.2 A Direct View on Quantum Mechanical Properties 5
7.3 Magnetic Resonance of Single Molecules 5
8 Single Molecule Biophysics and Biochemistry 5
8.1 Direct Observation of Single Molecule Reactions 5
8.2 Nanometers and Piconewtons: Mechanical Properties of Single Macromolecules 5
8.3 The Tiniest Motors: Kinesin, Dynein and Myosin 5
8.4 Receptor–Ligand Binding on a Single Molecule Basis 6
8.5 DNA as a Molecular Individuum 6
Abbreviations and Acronyms 6
Related Articles 6
References 7

In order to detect and analyze single biomolecules, near-field microscopes and conventional far-field microscopes are used. Spectroscopic experiments are now possible to test predictions of quantum mechanics in a very direct way. With biological macromolecules, chemical reactions can be seen directly in the microscope. Mechanical properties of individual macromolecules can be measured with very high precision. This allows a very basic process of life, the generation of motion, to be understood in its finest details. Finally, single molecule experiments allow the DNA molecule to be studied in its full, practically unlimited individuality.

1 INTRODUCTION

The ultimate goal of sensitivity in analytical chemistry is single molecule detection, characterization and particularly the study of reactions of individual molecules with each other. The era of this grade of perfection of analytical chemistry has dawned with the end of the passing millenium and certainly the new millenium will become the time of single molecule physics, chemistry and biology. By the end of 1999, in addition to a number of reviews, special issues of regular journals[1–4] on single molecules and a book[5] are available in the literature.

2 HISTORY

One of the first single molecule experiments (1961) in analytical biochemistry is the study of the reaction of a single β-galactosidase molecule held in a microdroplet.[6] Widespread applications became possible in the 1980s, with the advent of near-field microscopies which allowed imaging, for example, of the orbitals of benzene molecules.[7] It came as a surprise that these micrographs (probably nanograph would be the better expression) revealed a threefold symmetry rather than a sixfold symmetry as one might have expected from the Kekulé formula. Later, individual atoms and molecules could be manipulated with the tips of such scanning microscopes (for these microscopy techniques see below). These pure detection techniques did not yet allow spectroscopy, since they were monochromic. Color and thus spectroscopy in the sense that it is typically used by the analytical chemist were not yet widely used. This problem was alleviated by the advent of the near-field scanning optical microscope. Then it became possible to use the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

wavelength of an individual molecule which had already been exploited in bulk analytical or physical chemistry to obtain spectral information. The biggest surprise was the finding that normal – though high quality – fluorescence microscopes could be used for single molecule analysis. This was thought to contradict the Abbe criterion which states that the resolution of a classical far-field microscope is not better than 50% of the wavelength of the light used for illumination (0.5 μm for green light). The mistake of the sceptics was that they confused resolution with reproducibility of localization. When the molecules were prepared with sufficiently small concentration onto a surface, they had an average distance of several micrometers from each other: they no longer had to be spatially resolved, they just had to be visualized. Visualization of individual molecules is possible down to sizes of a few nanometers (see for example page 274 of Greulich[8]). The price one has to pay is that the molecules appear too large, but for many applications this drawback is acceptable. The classical fluorescence microscope has two significant advantages compared to the above-mentioned near-field microscopies. It is fast (standard temporal resolution is video frequency, 40 ms, and even this can be speeded up), i.e. the dynamics of molecules and reactions becomes experimentally accessible and preparation techniques can be easily adapted from bulk analytical methods. Thus, with the advent of classical fluorescence microscopy the path was paved to studying real single molecule reactions. Some studies dealt with small reactions such as the catalytic oxidation of CO on platinum 111.[9] Here it turned out that the individual molecule reaction just reflected what was known from bulk. A different situation was expected for biomolecules: since theoretical knowledge of protein structures had predicted that all individual enzymes of a species would differ slightly but distinctly from each other, it was no surprise that they revealed different reaction rates.[10] A particularly attractive target for single molecule studies is the DNA molecule. DNA is so variable that it is not possible to synthesize one copy of each possible DNA molecule only 120 basepairs in length (a typical gene would be much longer). The mass of the universe would not suffice (see for example page 161 of Greulich[8]). Due to this virtually unlimited variability, analysis of DNA single molecules, unlike conventional genome-megaprojects, may finally provide full information on DNA as a molecular individuum.

3 PREPARATION AND HANDLING OF SINGLE MOLECULES

Two types of single molecule solutions are interesting for single molecule research: highly diluted solutions of small molecules (single chromophores) and, on the other hand, biological macromolecules stained with a large number of chromophores.

3.1 Microfluidics for Single Molecule Handling

Typical bulk dye solutions are millimolar to micromolar. A solution with one molecule in the typical volume of 1 fL which is used in single molecule research corresponds to 10^{15} mol L^{-1} or approximately a nanomolar solution. These small concentrations can be obtained by conventional serial dilution. In order to prevent loss of single molecules to vessel walls, suitable materials are helpful, for example hydrophobic ones for hydrophilic molecules and vice versa. A problem is to set the time point for a reaction start. One approach is to premix the reaction partners, dilute them subsequently, transfer them to a surface and then observe the reaction. With such a strategy, the first minutes of the reaction are lost and therefore only slow reactions can be observed. An alternative approach is to bring the reaction partners onto a surface in separate nanodroplets and to start the reaction by fusing the latter at a preselected time. Here, the reaction can be observed a few seconds after mixing.

3.2 Preparation of Single Biological Macromolecules

A completely different approach is possible when biological macromolecules such as DNA, microtubules or actin filaments are involved. Such molecules can be coupled to micrometer sized polystyrene or silica spheres which act as a sort of handle. They can be handled micromechanically, or, more gently and precisely, by optical tweezers.[8] One successful strategy to start a reaction of such a macromolecule with a reaction partner is to use caged compounds: for example, many enzymes acting on DNA require magnesium. When the latter is present as a caged compound (and thus is inaccessible for the enzyme) the reaction cannot proceed. When the caged compound is disrupted by irradiating it with ultraviolet light, the magnesium is released and the reaction starts immediately.

3.3 Laser Micromanipulation for Handling of Individual Macromolecules

An interesting tool for handling of individual macromolecules are optical tweezers: a continuous infrared laser of moderate power, which is focused in a microscope to the theoretical (diffraction) limit. Light pressure and gradient forces push microscopic objects in the focus of the laser. By moving the focus, these objects can be moved along complex trajectories. Since only forces exerted by light are acting, no sticking of individual molecules to microtools or disruption by the latter occurs.

4 INTERACTION OF LIGHT WITH SINGLE MOLECULES

A single, light absorbing molecule is a unique quantum system which interacts with the light field. Such an interaction not only allows us to obtain unique information on the molecule, but also on the intrinsic properties of light. For example, a number of important experiments on which quantum mechanics are based only exist so far as ''Gedankenexperiment'' or have, at most, been experimentally verified only in a disputable way. Surprisingly for many, the famous quantum mechanical version of the double slit experiment is among these unsolved problems. This crucial experiment, where it is assumed that individual photons hit the double slit and interfere with each other or with themselves, has never been performed since it is difficult to produce single photons. In most cases wavetrains are produced by attenuating a laser, and this is not what the Gedankenexperiment needs. With single molecules one has for the first time the perspective to produce and use individual photons safely.

5 FLUORESCENCE CORRELATION SPECTROSCOPY AS A QUASI SINGLE MOLECULE TECHNIQUE

There are basically two strategies for the observation of single molecules: one may embed them in some sort of matrix and observe them at the surface of a substrate material, or one may study them directly in solution. The latter technique requires extreme dilution of the molecules and simultaneously extreme miniaturization of the observation volume. This has been achieved by investigating solutions of fluorescing molecules with a diffraction limited fluorescence microscope, i.e. in an observation volume of less than 1 fL. Basically, in such experiments the velocity of a molecule migrating through the tiny observation volume is measured. One detects a single molecule in a given volume, and after a preset time of waiting one looks to see if it is still there. Slow particles will be found several times, fast particles will be found only twice or three times. Mathematically, this is described by the correlation function (Equation 1)

$$G(t) = \sum I(t)^* \, I(t + T) \qquad (1)$$

where T is the waiting time.

$G(t)$ is related to the speed of the molecule which can be used to calculate the diffusion constant of the molecule. The latter can be used to estimate the molecular weight. Such experiments are often used to measure the binding of two individual molecules. This is possible when the two binding partners differ by at least a factor of 8–10 in size, for example the binding of an antigen to an antibody or the hybridization of a small DNA or RNA oligonucleotide to a long DNA or RNA molecule.[11] In such experiments, the smaller binding partner is made fluorescent. One will see the small partner as a fast diffusing object and the complex of small with large partner as a slow object. The correlation analysis not only allows the absolute migrating velocities to be measured, but also the ratio of free to bound molecules, i.e. the binding constants according to the mass action law. Since one needs to measure a number of molecules and molecular complexes to obtain such binding constants, fluorescence correlation spectroscopy is not a true single molecule technique, but it is a technique for very small concentrations. True single molecule techniques can be based on microscopies as described in the following.

6 MICROSCOPY OF SINGLE MOLECULES

6.1 Near-field Microscopies

The near-field microscopes all work according to a common basic principle: a sort of ultrafine stylus scans an object matrixwise and subsequently assembles an image from the data it has obtained. This data assembly takes some time and therefore the temporal resolution of these microscopes is limited to seconds or somewhat below seconds. In turn, the spatial resolution is better than with any other tool available to the analytical chemist or biochemist. The reason for this is that the stylus comes very close to the object under observation. It is the same principle as the stethoscope of a medical doctor: there, acoustic waves of a beating heart with wavelengths of meters can be exploited to localize the heart with an accuracy of centimeters. Therefore, the optical version of the near-field microscopes, the SNOM (scanning near-field optical microscope), has also been called the optical stethoscope. It has a typical spatial resolution of 60 to 80 nm and can be used as a highly localized spectroscope. The second near-field microscope used in the analysis of single biomolecules is the atomic force microscope (AFM). Here, a mixture of van der Waals, electrostatic and other forces is used to detect the molecule. Since these are all short range forces, one has to get closer to the molecule under observation and the spatial resolution is better than with the SNOM for ideal (hard) objects in the subnanometer range. The near-field microscope with the best resolution is the scanning tunneling microscope which, under favorable conditions, can detect submolecular structures as mentioned already in the historical overview. It needs electrically conducting objects.

6.2 Very Highly Sensitive Fluorescence: Dyes and Detectors

The near-field techniques for the detection of single molecules often require specific preparation protocols. Also, the tip of the detecting stylus is very sensitive and can easily break. The typical practical lifetime of such tips are hours or days at most. Therefore, near-field microscopies may easily become time-consuming and expensive. This may be one reason why classical far-field microscopy has got its chance in single molecule analysis. Here, light is the carrier of information and comparably gentle interactions can be used to study individual molecules. A disadvantage is that only molecules with a significant fluorescence yield or nonfluorescing molecules after labeling with a fluorescence dye can be investigated. Among biomolecules, NADH, flavines and other natural compounds occurring in biological cells are used. For biological macromolecules, a large number of fluorescing dyes is available which are known not, or only moderately, to modify the macromolecule's function. Many of these dyes have been developed to match with their absorption maximum the working wavelength of standard lasers.

Detecting single molecules by a fluorescence microscope is not so much a problem of the number of emitted photons. Since a molecule can, in principle, emit a light quantum every 10 ns, up to 100 million quanta might be expected from a single molecule per second. Two problems hamper this process: first, a dye molecule will decay after a few hundred up to a million cycles of excitation and emission into nonfluorescing derivative. Second, in order to excite a dye molecule at typical concentrations (see above), some 10^8 photons will pass the detection volume. In other words, one has a large number of unwanted photons per useful photon for the detection of the molecule. Optical filtering techniques are not completely sufficient to identify the useful photon. This is the major challenge for classical light microscopy in single molecule analytics.

Part of these problems may be overcome by replacing classical fluorescence dyes by nanocrystals or quantum dots. These are nanometer sized clusters of, for example, zinc sulfide capped cadmium selenide.[12] Their absorption and emission behavior is solely determined by their size: small clusters are blue, larger ones are red. An advantage of these quantum dots is their very good absorption and emission behavior; a disadvantage is their bulkiness.

6.3 Classical Far-field Microscopy for High Temporal Resolution

In spite of these problems, classical far-field microscopy is attractive since it provides high temporal resolution. When a fluorescence microscope is equipped with a standard TV CCD camera, one can obtain 25 images per second. This corresponds to a temporal resolution of 40 ms. Special cameras are much faster and millisecond resolution is easily available when one is willing to go a step beyond standard CCD recording. For low level light, as in single molecule research, cooled CCD cameras integrating over several image frames can be used, but then the temporal resolution is decreased correspondingly. When extreme sensitivities at high temporal resolution are required, single photon counting cameras are available. These are in principle arrays of microphotomultipliers, often with 250×250 pixel elements. In almost all cases the reduced (as compared to standard CCD cameras with 700×500 pixels) image size is completely sufficient, since single molecule observation requires only parts of a full visual field in a light microscope. For an overview on such cameras see Pilarczyk.[13]

Color information can be added by using a true color camera but they do not usually have the sensitivity described above. More efficient is the use of optical filters. Fast switching filter wheels and computer programs for exactly overlaying the different colors are available on the market.

7 SINGLE (NONBIOLOGICAL) MOLECULE SPECTROSCOPY

Since the energy states in molecules are not indefinitely sharp, absorption and emission lines have a nonzero width. From quantum mechanics, one can relate these widths with lifetimes of a specific energy state and with other properties of a molecule. The shape of such a line can be described by a Lorentzian curve. Line widths in bulk experiments detect an average over many slightly differing molecules – they are ''inhomogeneously broadened'' – i.e. they are much wider than the informative Lorentzians. Therefore, the advent of single molecule techniques represents a milestone for molecular spectroscopy. Even time resolved fluorescence studies on single molecules[14–16] and three-dimensional imaging are now possible,[17] not only in solid state, but also in solution.[18]

7.1 Reaching the Limits of Spectroscopy

One of the first steps towards single molecule spectroscopy was achieved with pentacene or terrylene molecules embedded into a *p*-terphenyl single crystal matrix (see for example chapter 1 in Basche et al.[5]). Many of these experiments were performed in the cold. Such experiments showed that very narrow line widths (10 MHz at 1 K) broadened by a factor of 100 when the temperature was increased by only 10° (chapter 2 in Basche et al.[5]).

7.2 A Direct View on Quantum Mechanical Properties

One of the probably most convincing single chromophore experiments deals with a central phenomenon of spectroscopy and quantum mechanics. A fluorescing dye molecule absorbs light and, after typically a few nanoseconds, re-emits it at a longer wavelength. Occasionally, by intersystem crossing, it goes into a triplet state. Quantum mechanics predicts that this state lives theoretically forever, in practice microseconds to hours (the latter is known from the phosphorescence of the face of a watch). In bulk experiments this prediction could be verified, but never directly seen. Several groups have now shown this process on the single molecule level: terrylene molecules were embedded in their *p*-terphenyl matrix and observed over a long period. During most of the time, they absorbed, re-emitted, re-absorbed and so on. Occasionally, they were ''off'' for a longer period. In some cases, this molecular blinking was attributed to transient oxidation. In other experiments oxidation could be excluded and the blinking was attributed to the quantum mechanic transition into the long lived exited triplet state. As long as the molecule is in this state, no further absorption and re-emission can occur: the molecule remains dark. This is the first direct visualization of a single-triplet transition: for the first time the observer has a direct view on a true quantum system (see chapter 1, page 52 in Basche et al.[5]).

7.3 Magnetic Resonance of Single Molecules

While near-field microscopy and optical far-field techniques represent the major tools for single molecule research, a few additional methods exist. One of the most important of these is the magnetic resonance technique. Usually, nuclear magnetic resonance (NMR) works with radiowaves and electron spin resonance with microwaves. In both cases, the energy per photon is small; one needs many of them to detect a spin and therefore the methods are not sensitive enough for single molecule detection. However, one can also detect the spin resonances optically (optically detected magnetic resonance, ODMR), and then single spin sensitivity can be achieved. This and other improvements of magnetic resonance techniques now allow single spins to be detected and analyzed.[19]

8 SINGLE MOLECULE BIOPHYSICS AND BIOCHEMISTRY

Single molecule techniques are even more important in studies of biological macromolecules. For some biological macromolecules the individual differences are not only interesting for reasons of basic research. Such differences may be important also for health and disease. For example, the malfunction of an enzyme may cause irregularities. Such a malfunction may develop by a sort of evolution and may only gradually turn into a true disease form. Early detection of such changes on the single molecule basis will probably allow early diagnosis and thus may help to prevent the outbreak of a disease.

8.1 Direct Observation of Single Molecule Reactions

The conformational dynamics and cleaving mechanisms of enzymes can be directly observed at surfaces.[20] The action of individual enzyme molecules in solution becomes accessible, for example, when a nonfluorescent substrate is converted into a fluorescent product or vice versa, or when the reaction with nonfluorescent partners can be made visible by a subsequent fluorescence reaction as is general practice in medical diagnosis or in nutrition chemistry. For example, the conversion of nonfluorescing NAD^+ into fluorescing NADH by individual molecules of lactate dehydrogenase has been observed by this approach. In the two droplet approach as described above, the reaction of some ten molecules has been observed. By the premix technique, quantized reaction rates have been found. The rates were multiple integers of a basic reaction rate. This was interpreted as the reaction catalyzed by one, two ... up to six enzyme molecules.[21] For this type of study the term ''single molecule enzymology'' has been coined.[22]

8.2 Nanometers and Piconewtons: Mechanical Properties of Single Macromolecules

The mechanical properties of filamentous molecules can be measured directly with optical tweezers. For this purpose again microbeads are biochemically coupled to the ends of the molecules and then pulled apart either by the stylus of an AFM or by optical tweezers. The force required to elongate such a molecule is measured.

8.3 The Tiniest Motors: Kinesin, Dynein and Myosin

Self driven motion and reproduction are two characteristic features of life. Both depend on molecular transport processes. In many of these processes, either dynein or kinesin molecules travel along rail-like filamentous structures called microtubules. Correspondingly, myosin molecules move along actin filaments. The action of a muscle, for example, results from the concerted operation of thousands of such tiny motors, which in most cases convert the chemical energy of an adenosine triphosphate molecule into motion. Typical single molecule experiments bind the rails (microtubules or actin) to a glass surface of a microscope slide and observe the single, fluorescently labeled motor protein gliding along the rails.

Alternatively, the motors are bound and the filamentous rails are held with double optical tweezers in suspension. When the molecular motor transports the rail, force is exerted on the optical tweezers, which can be measured with piconewton accuracy. Such experiments have shown that the force exerted by single motor proteins is of the order of a few piconewtons and the step width with which it moves along a rail is 8 nm for kinesin and probably also for dynein, while it may range from a few nanometers up to more than 10 nm for native and genetically modified myosin molecules with a different neck region. For kinesin, single molecule experiments under permanent load showed that models derived from bulk data, where permanent load of in vivo processes could not be simulated, cannot be correct. In these experiments, Michaelis–Menten kinetics, a very basic mechanism of biochemistry, could be measured on the single molecule level.[23] For myosin, per nanometer of a structural part called the neck region, the step width is 0.5 nm.[24] Similar experiments have been performed with RNA polymerase from *E. coli*. A surprise was that this enzyme exerts a force more than double that of the motor proteins, in spite of the fact that its task is not to generate motion but to read genetic information from DNA into RNA. A possible explanation for this finding is that this enzyme does not read from naked DNA but from DNA complexed with proteins (histones, nucleosomes) and that the force is required to read its way through these obstacles. The RNA polymerase experiments, similar to the kinesin experiments described above, also showed the importance of true single molecule experiments for the elucidation of the basic mechanisms of generation of motion.[25] Perhaps it is only an experimental curiosity: knots have even been tied into actin molecules.[26]

8.4 Receptor–Ligand Binding on a Single Molecule Basis

The imaging of single molecule diffusion[27] and the counting of single protein molecules at surfaces represent important steps for limiting the sensitivity analysis of disease processes. This may be binding to hormone receptors or antigen–antibody binding. A technique to count such binding processes has been developed by Löscher.[28] Even more detailed are studies on the forces between individual antibodies and their antigens.[29] The perspective of such studies is not only to learn details on the underlying binding process, but also to develop a force sensing microscope.

8.5 DNA as a Molecular Individuum

The unlimited variability of DNA molecules and the resulting need of single molecule studies has already been addressed above. Mechanical properties of individual DNA molecules can be studied for example by stretching them first in a shear flow and then observing by a fast video microscope camera how these filamentous structures collapse into random coils.[30] When DNA is overstretched, mechanical properties become sequence dependent.[31] A special double stranded DNA construct, which is overwound and thus supercoiled, can be stretched and thereby denatures locally, i.e. the double strand becomes opened into two single stranded stretches of DNA.[32] A similar experiment separated two complementary strands of DNA mechanically, whereby obviously the double helix was unwound without prior overwinding. The required separation force is different for GC- and AT-rich sequence elements and thus information on the base composition of a single DNA molecule can be obtained with an accuracy of 100–500 bases.[33] Since DNA exists in different structural forms, for example B-DNA and Z-DNA with different lengths per basepair, structural transitions can be exploited to construct a nanomechanical device.[34] Studies on the behavior of single DNA molecules in gels have a quite different motivation.[35] Here, the mechanism of gel electrophoresis of large DNA fragments is studied.

Of particular interest is the direct observation by light microscopy of DNA cutting, since this is a process occurring in many processes in life sciences. The observation of such reactions in electrophoresis gels has been termed optical mapping.[36] Phosphodiester bond breaks have been directly observed in a dual beam optical microscope[37] and the binding of RecA molecules to a single DNA molecule held by double beam optical tweezers was visualized directly in a light microscope.[38] The enzymatic cleavage of an individual Lambda phage DNA molecule by the restriction endonuclease Apa1 has been observed.[39,40] One of the most beautiful single molecule experiments is perhaps the visualization of an RNA polymerase molecule running along a DNA molecule. The processivity of this enzyme can thus be directly seen, without any model assumptions.[41]

ABBREVIATIONS AND ACRONYMS

AFM	Atomic Force Microscope
NMR	Nuclear Magnetic Resonance
ODMR	Optically Detected Magnetic Resonance
SNOM	Scanning Near-field Optical Microscope

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Nucleic Acid Analysis in Clinical Chemistry

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science • Microspectrophotometry in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Molecules, Properties and Detection of Single • Optical Mapping in Genomic Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Lifetime Measurements, Applications of

REFERENCES

1. K.O. Greulich, E. Klose, H. Neumann, 'Single Molecule Techniques: a New Era in Life Sciences', *Exp. Tech. Phys.*, **41**(2), 157–311 (1995).
2. R. Erdmann, C. Seidel, 'Single Molecule Detection; Basics and Applications in the Life Sciences', *Bioimaging*, **5**(1), 1–2 (1997).
3. R. Erdmann, J. Enderlein, C. Seidel, 'Single Molecule Detection; Basics and Applications in the Life Sciences', *Bioimaging*, **6**(3), 1–2 (1998).
4. R. Erdmann, J. Enderlein, C. Seidel, 'Single Molecule Detection and Ultrasensitive Analysis in the Life Sciences', *Cytometry*, **36**(3), 162–264 (1999).
5. T. Basche, W.E. Moerner, M. Orrit, U.P. Wild, *Single Molecule Optical Detection, Imaging and Spectroscopy*, VCH, Weinheim, 1997.
6. B. Rothmann, 'On the Reaction of Single Beta Galactosidase Molecules in Microdroplets', *Proc. Natl. Acad. Sci. USA*, **47**(1), 1–6 (1961).
7. Editorial on H. Ohtani, C. Chiang, R.J. Wilson, 'Image Reveals an Atom's Secret', *Photon. Spectra*, **88**(9), 48 (1988).
8. K.O. Greulich, *Micromanipulation by Light in Biology and Medicine: the Laser Microbeam and Optical Tweezers*, Birkhäuser, Basel, 1999.
9. J. Wintterlin, S. Völkering, T.W. Janssens, T. Zambelli, G. Ertl, 'Atomic and Macroscopic Reaction Rates of a Surface Catalyzed Reaction', *Science*, **278**, 1931–1934 (1997).
10. Q. Xue, E. Yeung, 'Difference in the Chemical Reactivity of Individual Molecules of an Enzyme Reaction', *Nature*, **373**, 681–683 (1995).
11. M. Kinjo, R. Rigler, 'Ultrasensitive Hybridization Analysis Using Fluorescence Correlation Spectroscopy', *Nucl. Acids Res.*, **23**, 1795–1799 (1995).
12. W.C.W. Chan, S. Nie, 'Quantum Dot Bioconjugates for Ultrasensitive Nonisotopic Detection', *Science*, **281**, 2016–2018 (1996).
13. G. Pilarczyk, S. Monajembashi, C. Hoyer, V. Uhl, K.O. Greulich, 'Fluorescence Microscopy and the Reactions of Single Molecules', in: *Applied Fluorescence in Chemistry, Biology and Medicine*, eds. W. Rettig, B. Stehmel, S. Schrader, H. Seifert, Springer, Berlin, 417–438, 1999.
14. W.P. Ambrose, P.M. Goodwin, J.C. Martin, R.A. Keller, 'Alterations of Single Molecule Fluorescence Lifetimes in Near Field Optical Microscopy', *Science*, **265**, 364–367 (1994).
15. J.J. Macklin, J.K. Trautmann, T.D. Harris, L.E. Brus, 'Imaging and Time Resolved Spectroscopy of Single Molecules at an Interface', *Science*, **272**, 255–258 (1996).
16. L. Brand, C. Eggeling, C. Zander, K.H. Drexhage, C.A.M. Seidel, 'Single Molecule Identification of Coumarin 120 by Time Resolved Fluorescence Detection: Comparison of One and Two Photon Excitation in Solution', *J. Phys. Chem. A*, **101**, 4313–4321 (1997).
17. R.M. Dickson, D.J. Norris, Y.L. Tzeng, W.E. Moerner, 'Three Dimensional Imaging of Single Molecules Solvated in Pores of Polyacrylamide Gels', *Science*, **274**, 966–969 (1996).
18. D.Y. Chen, N.J. Dovichi, 'Single Molecule Detection in Capillary Electrophoresis: Molecular Shot Noise as a Fundamental Limit to Chemical Analysis', *Anal. Chem.*, **68**, 690–696 (1996).
19. J. Wrachtrup, C. von Borczyskowski, J. Bernard, R. Brown, M. Orrit, 'Hahn Echo Experiments on a Single Triplet Electron Spin', *Chem. Phys. Lett.*, **254**, 262–267 (1995).
20. T. Ha, A.Y. Ting, J. Liang, W.B. Caldwell, A.A. Deniz, D.S. Chemla, P.G. Schultz, S. Weiss, 'Single Molecule Fluorescence Spectroscopy of Enzyme Conformational Dynamics and Cleavage Mechanism', *Proc. Natl. Acad. Sci. USA*, **96**, 893–898 (1999).
21. V. Uhl, G. Pilarczyk, K.O. Greulich, 'Fluorescence Microscopic Observation of Catalysis by Single or Few LDH1 Enzyme Molecules', *Biol. Chem.*, **379**, 1175–1180 (1998).
22. X.S. Xie, H.P. Lu, 'Single Molecule Enzymology', *J. Biol. Chem.*, **274**, 15 967–15 970 (1999).
23. K. Visscher, M.J. Schnitzer, S.M. Block, 'Single Kinesin Molecules Studied with a Molecular Force Clamp', *Nature*, **400**, 184–189 (1999).
24. C. Ruff, M. Furch, B. Brenner, D.J. Manstein, E. Meyhöfer, 'The Step Size of Myosin is Directly Determined by the Length of The Neck Domain', *Proc. Natl. Acad. Sci. USA*, (2000) submitted.
25. M.D. Wang, M.J. Schnitzer, H. Yin, R. Landick, J. Gelles, S.M. Block, 'Force and Velocity Measured for Single Molecules of RNA Polymerase', *Science*, **282**, 902–907 (1997).
26. Y. Arai, R. Yasuda, K.I. Akashi, Y. Harada, H. Miyata, K. Kinosita, H. Ito, 'Tying a Molecular Knot with Optical Tweezers', *Nature*, **399**, 446–448 (1999).
27. T. Schmidt, G.J. Schütz, W. Baumgartner, H.J. Gruber, H. Schindler, 'Imaging of Single Molecule Diffusion', *Proc. Natl. Acad. Sci. USA*, **93**, 2926–2929 (1996).

28. F. Löscher, S. Böhme, J. Martin, S. Seeger, 'Counting of Single Protein Molecules at Interfaces and Applications of this Technique in Early Stage Diagnosis', *Anal. Chem.*, **70**, 3202–3205 (1998).
29. R. Ros, F. Schwenger, D. Anselmetto, M. Kubon, R. Schäfer, A. Plückthun, L. Tiefenauer, 'Antigen Binding Forces of Individually Addressed Single Chain Fv Antibody Molecules', *Proc. Natl. Acad. Sci. USA*, **95**, 7402–7405 (1998).
30. D.E. Smith, H.P. Babcock, S. Chu, 'Single Polymer Dynamics in Steady Shear Flow', *Science*, **283**, 1724–1727 (1999).
31. M. Rief, H. Clausen-Schaumann, H.E. Gaub, 'Sequence-dependent Mechanics of Single DNA Molecules', *Nature Struct. Biol.*, **6**, 346–349 (1999).
32. T.R. Strick, V. Croquette, D. Bensimon, 'Homologous Pairing in Stretched Supercoiled DNA', *Proc. Natl. Acad. Sci. USA*, **95**, 10 579–10 583 (1998).
33. B. Essevaz-Roulet, U. Bockelmann, F. Heslot, 'Mechanical Separation of the Complementary Strands of DNA', *Proc. Natl. Acad. Sci. USA*, **94**, 11 935–11 940 (1997).
34. C. Mao, W. Sun, Z. Shen, N.C. Seeman, 'A Nanomechanical Device Based on the B-Z Transition of DNA', *Nature*, **397**, 144–146 (1999).
35. S. Guerrieri, S.B. Smith, C. Bustamante, 'Trapping of Megabase-sized DNA Molecules During Agarose Gel Electrophoresis', *Proc. Natl. Acad. Sci. USA*, **96**, 453–458 (1999).
36. Y.K. Wang, E. Huff, D.C. Schwartz, 'Optical Mapping of Site Directed Cleavages on Single DNA Molecules by the RecA Assisted Restriction Endonuclease Technique', *Proc. Natl. Acad. Sci. USA*, **92**, 165–169 (1995).
37. W.A. Lyon, M.M. Fang, W.E. Haskins, S. Nie, 'A Dual Beam Optical Microscope for Observation and Cleavage of Single DNA Molecules', *Anal. Chem.*, **70**, 1743–1748 (1998).
38. M.L. Bennink, O.D. Schärer, R. Kanaar, K. Sakata Sogawa, J.M. Schins, J.S. Kanger, B.G. de Grooth, Jan Greve, 'Single Molecule Manipulation of Double Stranded DNA Using Optical Tweezers: Interaction Studies of DNA with Rec A and YOYO-1', *Cytometry*, **36**, 200–208 (1999).
39. C. Hoyer, S. Monajembashi, K.O. Greulich, 'Laser Manipulation and UV Induced Single Molecule Reactions of Individual DNA Molecules', *J. Biotech.*, **52**, 65–73 (1996).
40. B. Schäfer, B. Nasanshargal, S. Monajembashi, H. Gemeinhardt, K.O. Greulich, V. Uhl, 'Study of Single Molecule Reactions with Classic Light Microscopy', *Cytometry*, **36**, 209–216 (1999).
41. Y. Harada, T. Funatsu, K. Murakami, Y. Nonoyama, A. Ishihama, T. Yanagida, 'Single Molecule Imaging of RNA Polymerase–DNA Interactions in Real Time', *Biophys. J.*, **76**, 709–715 (1999).

Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Rina K. Dukor
Vysis Inc., Downers Grove, USA, and BioTools, Elmhurst, USA

Laurence A. Nafie
Syracuse University, USA, and BioTools, Elmhurst, USA

1 Introduction 1
1.1 Definitions of Vibrational Circular Dichroism and Raman Optical Activity 2
1.2 Comparison of Vibrational Circular Dichroism and Raman Optical Activity 2
2 History and Instrumentation 3
2.1 Vibrational Circular Dichroism Instrumentation 3
2.2 Raman Optical Activity Instrumentation 3
3 Instrument Operation and Performance 3
3.1 Fourier Transform Vibrational Circular Dichroism Instrumentation and Measurement 3
3.2 Examples of Vibrational Circular Dichroism Spectra 5
3.3 Baseline Flatness and Stability 5
3.4 Intensity Calibration and Reproducibility 7
3.5 Signal-to-noise Ratio 8
4 Biological Examples 8
4.1 Proteins 8
4.2 Nucleic Acids 10
4.3 Sugars 10
5 Pharmaceutical Examples 10
6 Areas of Application 11
6.1 Enantiomeric Excess 11
6.2 Absolute Configuration 12
7 Conclusions 13
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 14
References 14

This article describes recent progress in the field of vibrational optical activity (VOA) as a probe of the structural properties of pharmaceutical and biological molecules. A strong emphasis is placed on vibrational circular dichroism (VCD) owing to its more advanced state of development. Raman optical activity (ROA) is included in the first two sections for completeness; other reviews give additional information about ROA.[1–3,8] The article focuses on the practical aspects of VCD spectral measurement and interpretation. This is supplemented by examples that serve to illustrate the principal areas of VOA application. VCD is defined as the difference in the absorbance of the left circularly polarized (LCP) versus right circularly polarized (RCP) infrared (IR) radiation for a chiral molecule undergoing a vibrational transition. A pair of enantiomers will produce VCD spectra that are equal and opposite in sign and a racemic mixture will have a null VCD signal. VCD can be measured for all kinds of chiral molecules, irrespective of their size. In practice, measurements are often carried out in solution but, with the new advances in instrumentation, it is now possible to measure spectra of solids and mulls. Compared to other optical techniques such as electronic circular dichroism (CD) and IR absorption, VCD is unique because it combines the optical activity property of CD with the rich structural fingerprint region of IR. The discovery and first measurements of VCD occurred in the early 1970s. Although over a dozen practitioners have published close to a 1000 papers since then, it is only recently that VCD instrumentation has become available commercially for nonspecialists. Applications span a variety of fields from chemical and pharmaceutical to biological. The VCD technique can be used for the determination of absolute configuration of small chiral molecules or larger natural products. It can also be used to follow a chiral synthesis both for stereochemistry and for optical purity, and to study the secondary structure of large proteins and small peptides, and the conformation of nucleic acids and sugars. Some recent reports have also shown the unique sensitivity of ROA in the study of viruses and in protein-folding experiments. VOA is now fulfilling its promise of becoming a technique that is broadly used for stereochemical and conformational studies of all varieties of chiral molecules, both natural and synthetic.

1 INTRODUCTION

VOA is defined as the differential response of a molecule to LCP versus RCP radiation during a vibrational transition.[1,2,3] There are two principal forms of VOA, an IR form referred to as VCD, and a Raman form known as ROA. VCD is the extension of CD measured for electronic transitions from the visible and ultraviolet regions into the IR region where vibrational transitions

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

occur. For Raman scattering, there is no classical analog, thus ROA is measured as circularly polarized (CP) intensity differences in the incident laser radiation, the scattered radiation, or both. As with CD, VOA is zero unless the sample possesses molecular chirality, either through its constituent chiral molecules or through a chiral spatial arrangement of nonchiral molecules.

1.1 Definitions of Vibrational Circular Dichroism and Raman Optical Activity

There is only one form of VCD as it involves the differential response of a molecule undergoing a one-photon transition. The spectra are typically measured as the difference in the absorbances of a molecule for LCP minus RCP light (Equation 1):[4]

$$\Delta A(\bar{\nu}) = A_{\mathrm{L}}(\bar{\nu}) - A_{\mathrm{R}}(\bar{\nu}) \tag{1}$$

where $\bar{\nu}$ is the wavenumber frequency in units of cm^{-1}. Alternatively, the effects of path length l and concentration C may be eliminated by using the absorption coefficient (Equation 2):

$$\Delta\varepsilon(\bar{\nu}) = \frac{\Delta A(\bar{\nu})}{lC} = \varepsilon_{\mathrm{L}}(\bar{\nu}) - \varepsilon_{\mathrm{R}}(\bar{\nu}) \tag{2}$$

The ROA method is more complex than VCD both experimentally and theoretically. There are four distinct forms of ROA involving CP radiation and an additional four involving linearly polarized light, although the latter are more specialized and will not be discussed further. If we designate the polarization state of the incident laser beam with a superscript and that of the scattered radiation with a subscript, the four forms of CP/ROA can be written as in Equations (3–6):[5]

$$\Delta I_{\alpha}(\bar{\nu}) = I_{\alpha}^{\mathrm{R}}(\bar{\nu}) - I_{\alpha}^{\mathrm{L}}(\bar{\nu}) \tag{3}$$

$$\Delta I^{\alpha}(\bar{\nu}) = I_{\mathrm{R}}^{\alpha}(\bar{\nu}) - I_{\mathrm{L}}^{\alpha}(\bar{\nu}) \tag{4}$$

$$\Delta I_{I}(\bar{\nu}) = I_{\mathrm{R}}^{\mathrm{R}}(\bar{\nu}) - I_{\mathrm{L}}^{\mathrm{L}}(\bar{\nu}) \tag{5}$$

$$\Delta I_{II}(\bar{\nu}) = I_{\mathrm{L}}^{\mathrm{R}}(\bar{\nu}) - I_{\mathrm{R}}^{\mathrm{L}}(\bar{\nu}) \tag{6}$$

Here, the intensities are a function of the wavenumber shift of the scattered radiation. The form of ROA given in Equation (3) is the one first discovered and measured since the 1970s. It is called incident circularly polarized (ICP) ROA because only the polarization state of the incident beam is switched between RCP and LCP states. The polarization state of the scattered light α is a fixed linearly polarized or unpolarized state. Note that the definition of ICP/ROA is opposite that of VCD with respect to RCP and LCP radiation. This is the result of historical convention, and not the result of any deep underlying principle. In fact, recent work on the theory of resonance Raman optical activity (RROA) has revealed a direct connection between the electronic CD of the resonant state and the observed RROA in which the magnitudes are the same and the sign correlation is opposite owing to the opposite definitions of CD and ROA.[6] The other three forms of ROA involve changing the polarization state of the scattered radiation between RCP and LCP states. Equation (4) describes scattered circularly polarized (SCP) ROA in direct analogy to ICP/ROA. The other two forms of the ROA in Equations (5) and (6) describe the in-phase and out-of-phase forms of dual circularly polarized (DCP) ROA, designated with subscripts I and II, respectively (i.e. DCP_I and DCP_{II}).

Another variable of ROA is the scattering angle. All early measurements of ROA were carried out for right-angle scattering. In the 1980s, the advantages of backscattering ROA were realized[7] and, currently, nearly all measurements of ROA are carried out with backscattering geometry, either as unpolarized ICP/ROA, DCP_I/ROA, or unpolarized SCP/ROA.[8]

1.2 Comparison of Vibrational Circular Dichroism and Raman Optical Activity

Both VCD and ROA were discovered in the early 1970s.[1] Each has evolved both experimentally and theoretically to a sophisticated state. Each is very sensitive to the stereostructural features of chiral molecules in solution, and they have similar differences and advantages to their parent spectroscopies IR absorption and Raman scattering. In fact, neither VCD nor ROA can be successfully measured unless an excellent parent IR or Raman spectrum can first be measured. Owing to the more complex nature of carrying out measurements and theoretical analyses, ROA has lagged behind VCD in applications to structural analysis. Further, both the instrumentation for the routine measurement of VCD and software for the ab initio calculation of VCD are now commercially available. Thus, VCD is much more accessible for use by nonspecialists for practical applications, in both academic and industrial settings.

Despite these differences in instrumental and theoretical commercial availability, VCD and ROA are unique complementary forms of VOA, each having its own advantages or disadvantages in particular kinds of applications. The situation is analogous to the complementary nature of IR and Raman spectroscopy in general where differences in sensitivity, solvent interference, and sampling have a role in the selection of the preferred approach. It has been shown that VCD and ROA spectra are strongly noncorrelated in the signs and relative intensities of the bands. The only common element is the structure and vibrational motion of the molecules under study. Thus, adding an ROA spectrum to a study where the VCD spectrum has already been obtained yields new

useful information. Both ROA and VCD have been used for conformational elucidation, optical purity measurements, and the determination of absolute configuration. In both VCD and ROA the abundance of vibrational transitions, each with its own probe of the stereochemistry of the chiral molecule, provides a wealth of structural information.

2 HISTORY AND INSTRUMENTATION

2.1 Vibrational Circular Dichroism Instrumentation

The first measurements of VCD were achieved in the early 1970s with instruments that were relatively crude by today's standards, but they demonstrated the existence of VCD as a natural phenomenon that could be used to learn about the detailed structure and dynamics of chiral molecules.[9,10,11] Later improvements in VCD spectrometers extended the wavelength coverage from the region of hydrogen stretching modes into the mid-IR region where a greater variety of vibrational transitions could be observed.[12] A particularly important advance was the advent of Fourier transform infrared (FTIR) methods for VCD measurement.[13] Virtually all modern VCD spectrometers, including all commercially available models, are now FTIR/VCD spectrometers.

During the 1990s, accessory modules for the measurement of FT/VCD (Fourier transform/vibrational circular dichroism) have been offered by several manufacturers of FTIR spectrometers. In the case of Bomem, Inc., a completely dedicated stand-alone, factory-aligned FT/VCD spectrometer has become commercially available, the Chiralir™, which has opened the way for widespread applications of VCD spectroscopy not only in academia but also in industry.[14]

The principal applications of VCD spectroscopy include measurements of the conformation, absolute configuration, and enantiomeric excess of chiral molecules.[4] Most of the molecules of interest for study with VCD are biological in origin. Many are molecules of pharmaceutical interest. The most unique application of VCD is its ability to determine absolute configuration in conjunction with ab initio calculations. Remarkably close matches have been achieved between experimental VCD spectra and the corresponding spectra calculated from first principles using quantum mechanical calculations.

2.2 Raman Optical Activity Instrumentation

Early ROA instrumentation was based on scanning dispersive spectrometers.[15,16] Light levels were low as nonholographic grating technology was used and two or three grating stages were necessary for the reduction of stray Rayleigh radiation. As a consequence, spectra spanning $1000\,cm^{-1}$ or longer required a day or more to measure. The experimental set-up used in these early measurements was always right-angle depolarized ICP measurements.

Several advances over the years changed this situation dramatically. These included the advent of backscattering geometry, back-thinned charge-coupled detectors, and holographic transmission gratings. Currently, ROA spectra spanning $2000\,cm^{-1}$ for the most favorable samples can be discerned in less than 10 s.[17,18] High-quality spectra for these samples can be obtained in approximately 1 min. This advance in speed and sensitivity has opened the way for more sophisticated applications of ROA, such as studying the folding properties of proteins as a function of temperature and other perturbations.

Most recently, attempts are underway to reduce ROA instrumentation to routine operation. These efforts include the introduction of fiber optic coupling to eliminate the effects of beam wander and loss of the precise alignment needed for artifact-free measurements.[8] When this is achieved, ROA is likely to join VCD as a technique available to chemists and biologists wishing to study the conformation of molecules of chemical, biological, and medical interest.

3 INSTRUMENT OPERATION AND PERFORMANCE

3.1 Fourier Transform Vibrational Circular Dichroism Instrumentation and Measurement

3.1.1 Basic Principles

The measurement of a VCD spectrum begins with the measurement of the IR spectrum of a sample. If the IR spectrum cannot be measured with good spectral quality, there is little chance of measuring the corresponding VCD spectrum. The general recommendation is to obtain the IR spectrum with a decadic absorbance in the range 0.1 to 1.0. The expression for the absorbance, $A(\bar{\nu})$, is given by Equation (7):

$$A(\bar{\nu}) = -\log_{10} \frac{I(\bar{\nu})}{I_0(\bar{\nu})} \tag{7}$$

An absorbance of 1.0 corresponds to a percentage transmission value (100 times the ratio in square parentheses) of 10%, whereas an absorbance of 0.1 corresponds to a percentage transmission of about 80%. If the absorbance is greater than 1.0, too little light reaches the detector, whereas if the absorbance is less than 0.1 the sample absorbs too little of the incident IR beam. The optimum value of the absorbance for VCD measurements is approximately 0.4.

In order to measure a good VCD spectrum, it is essential to consider additional issues beyond the quality of the parent IR spectrum. These include the quality of the VCD baseline, the selection of the spectral region, and the setting of the polarization modulator. In addition, for FT/VCD measurements, the appropriate phase correction must be used to obtain the desired spectrum.

3.1.2 Instrument Block Diagram

A simple block diagram for the measurement of FT/VCD in the mid-IR region is given in Figure 1. The design corresponds to the so-called double modulation approach to FT/VCD measurement. Here two interferograms are present simultaneously at the detector, one for the ordinary IR spectrum and the other for the VCD spectrum. The method involves measuring each of the interferograms and ratioing the raw VCD spectrum to the raw transmission spectrum of the sample. IR radiation from an interferometer first passes through an optical filter to isolate the spectral region of interest. The beam continues through a linear polarizer that defines a single state of polarization and then on to a ZnSe photoelastic modulator (PEM) that modulates the polarization between LCP and RCP states at a frequency of 37 kHz. The chiral sample is placed directly after the PEM and the beam is focused with a lens to a HgCdTe detector that is cooled with liquid nitrogen. The detector signal is preamplified and divided into two paths. One is the ordinary IR path that represents the FTIR interferogram of the sample. The expression for the IR transmission interferogram is given by Equation (8),

$$I_{DC}(\delta) = \int_0^\infty I_{DC}(\bar{\nu}) \cos[2\pi\bar{\nu}\delta - \theta_{DC}(\bar{\nu})]\, d\bar{\nu} \qquad (8)$$

where the subscript DC refers to the constant nature of this interferogram relative to the high-frequency polarization modulation. Here δ is the path difference in the interferometer and $\theta_{DC}(\bar{\nu})$ is the Fourier phase function. After digitization, phase correction, and Fourier transformation, one obtains the IR (DC) transmission

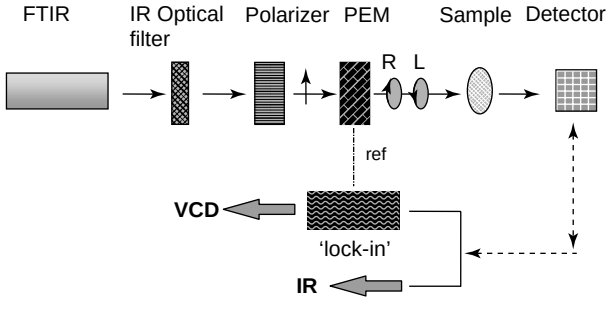

Figure 1 Optical–electronic diagram for a typical FT/VCD spectrometer.

spectrum of the sample given in terms of the average transmissions for LCP and RCP radiation as (Equation 9)

$$I_{DC}(\bar{\nu}) = \tfrac{1}{2}[I_R(\bar{\nu}) + I_L(\bar{\nu})] \qquad (9)$$

The other path is to the lock-in amplifier, as shown in Figure 1, referenced to the PEM modulation frequency. The output of the lock-in is the interferogram associated with the VCD which is given by Equation (10)[19]

$$I_{AC}(\delta) = \int_0^\infty I_{AC}(\bar{\nu}) \exp(-2V\bar{\nu}\tau) \cos[2\pi\bar{\nu}\delta - \theta_{AC}(\bar{\nu})]\, d\bar{\nu} \qquad (10)$$

where the exponential factor describes the effect of the lock-in time constant τ of the signal, and V is the mirror velocity of the interferometer. After this interferogram is digitized, phase corrected, and Fourier transformed, one obtains the instrumental VCD (AC) transmission spectrum (Equation 11),

$$I_{AC}(\bar{\nu}) = \tfrac{1}{2}[I_R(\bar{\nu}) - I_L(\bar{\nu})] \exp(-2V\bar{\nu}\tau) \sin\alpha_M(\bar{\nu}) \qquad (11)$$

where the retardation angle of the PEM varies sinusoidally at the PEM frequency and is given by Equation (12):

$$\alpha_M(\bar{\nu}) = \alpha_M^0(\bar{\nu}) \sin 2\pi\nu_M t \qquad (12)$$

When the DC and AC transmission spectra in Equations (9) and (11) are ratioed, the result is proportional to the VCD spectrum of the sample, according to Equation (13):

$$\frac{I_{AC}(\bar{\nu})}{I_{DC}(\bar{\nu})} = 2J_1[\alpha_M^0(\bar{\nu})] \exp(-2V\bar{\nu}\tau) 1.15\Delta A(\bar{\nu}) \qquad (13)$$

Here the efficiency of the PEM at a particular wavenumber frequency is given by the first-order Bessel function, $J_1[\alpha_M^0(\bar{\nu})]$. The final VCD spectrum can be obtained after a VCD calibration spectrum is collected using a multiple waveplate with axes parallel or perpendicular to the stress axis of the PEM followed by a polarizer oriented parallel or perpendicular to the polarizer in Figure 1. There are four possible settings for the waveplate and polarizer, and the VCD intersection points of the curves they produce define a calibration plot given by Equation (14):

$$\left.\frac{I_{AC}(\bar{\nu})}{I_{DC}(\bar{\nu})}\right|_{CAL} = \pm 2J_1[\alpha_M^0(\bar{\nu})] \exp(-2V\bar{\nu}\tau) \qquad (14)$$

Division of this curve into the ratio in Equation (13) isolates the properly calibrated VCD spectrum $\Delta A(\bar{\nu})$. The phase correction used for the VCD interferogram to eliminate $\theta_{AC}(\bar{\nu})$ before Fourier transformation is obtained by a special procedure that requires the transfer of the phase correction from a monosignate VCD spectrum from a stressed optical plate. This is necessary

because the self-phase correction of the VCD spectrum using conventional FTIR software leads to sign errors when the raw VCD spectrum changes sign.

3.2 Examples of Vibrational Circular Dichroism Spectra

The VCD and IR spectra of (*s*)-(−)-α-pinene between 1350 and 800 cm^{-1} are presented in Figure 2. The spectrum was measured as a neat liquid with a 73 μm BaF_2 cell and the Chiralir™ VCD spectrometer. The VCD is half the difference between 20 minute collections of the (−) and (+) enantiomers as discussed further below. Above the VCD spectrum on the same intensity scale is the noise spectrum associated with the VCD spectrum. Note that the VCD spectrum is approximately four orders of magnitude smaller than the corresponding parent absorbance spectrum. Positive VCD bands, according to the definition of ΔA given in Equation (1), absorb LCP radiation more strongly than RCP radiation, and vice versa for negative VCD bands. The VCD spectrum of α-pinene is routinely used as an intensity standard and a basis for comparing the performance of different VCD instruments. This molecule can be sampled in high concentration as a neat liquid and has a rigid stereoconformation due to its fused ring structure. The VCD spectrum of α-pinene is relatively large and not difficult to measure with high signal quality.

A second example of a VCD spectrum is given in Figure 3 where the IR, VCD, and VCD noise spectra of L-valine in H_2O are presented. The spectra were measured on the Chiralir™ at a concentration of 250 mg mL^{-1} in aqueous HCl solution using a 12 μm path length cell. This is a more difficult case for two reasons. First, the

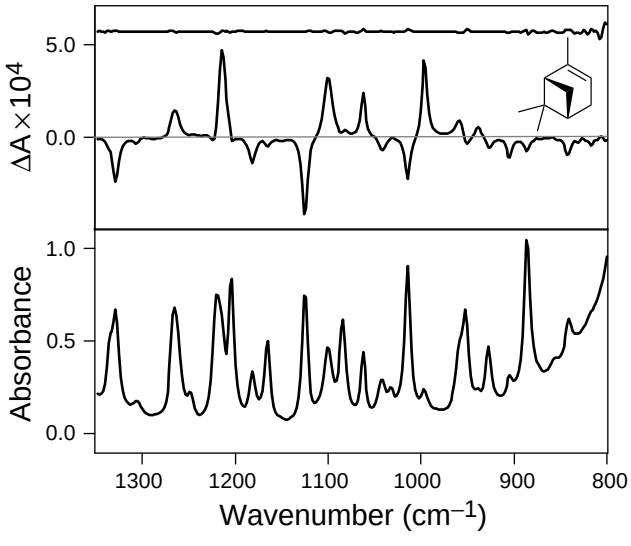

Figure 2 Absorbance (bottom), VCD (middle) and VCD noise (top) spectra of (*S*)-(−)-α-pinene as a neat liquid in a 73-μm BaF_2 cell. Collection time for each enantiomer is 20 min.

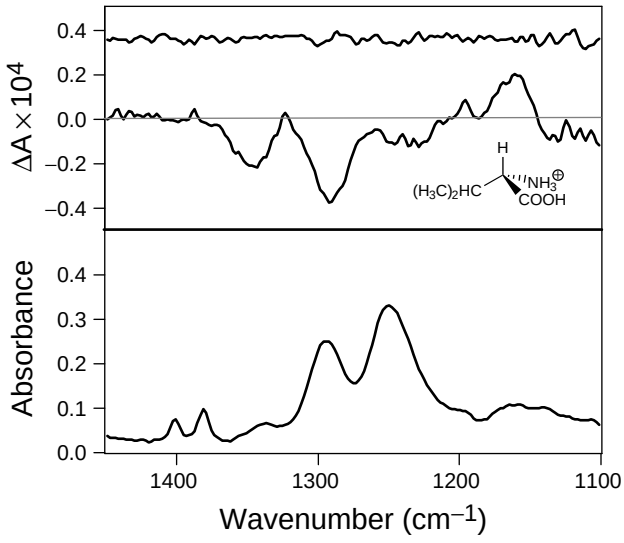

Figure 3 Absorbance (bottom), VCD (middle) and VCD noise (top) spectra of L-valine measured in H_2O/HCl solution at a concentration of 250 mg mL^{-1} and a path length of 12 μm. Collection time for each enantiomer is 1 h.

largest VCD features are 10 times smaller than the largest features for α-pinene and, second, the solvent absorbs some of the IR beam, thus reducing the amount of available IR radiation for measuring the VCD spectrum. Therefore, even though the VCD spectrum for L-valine was collected for 1 h, the signal-to-noise ratio (S/N) is lower than that observed for α-pinene in 20 min of collection. In the following sections we examine in further detail some of the more practical aspects of acquiring final calibrated VCD spectra.

3.3 Baseline Flatness and Stability

Because VCD signals are so small, the instrument baseline at the scale of VCD sensitivity may not correspond to true zero. In Figure 4, the VCD spectra of both enantiomers of α-pinene, and a third spectrum from a solution that is nearly the racemic mixture but still has a few percent of the (−) enantiomer present, are shown. This near racemic spectrum is very nearly the baseline of the empty instrument, and except for a few deviations where the spectrum has large VCD spectral features is identical, to within the noise of the true racemic VCD spectrum. Notice that all three curves pass through the common points where the VCD has a zero value. The baseline shown is relatively flat and close to the true zero of the instrument. Specifying the instrument baseline in the absence of a sample is a criterion for the quality of the VCD instrument. If little care is taken in the choice of optical elements or the alignment of these elements, poor VCD baselines with large deviations from zero on the VCD intensity scale may be encountered. The

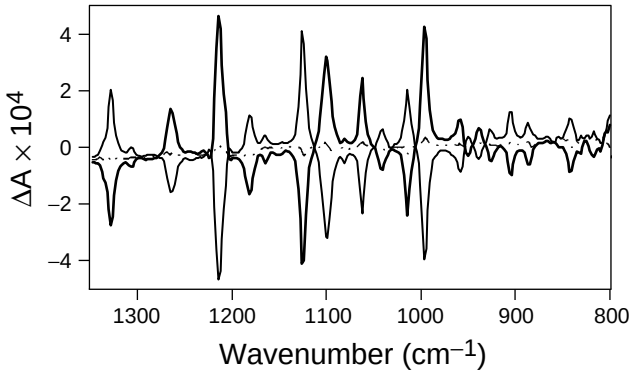

Figure 4 The VCD spectra of (−)-, (+)-, and (±)-α-pinene, neat liquid, at 73-μm path length, showing opposite VCD for the (+) and (−) enantiomers, and a near racemic mixture exhibiting a slight percentage enantiomeric excess bias toward the (−)-enantiomer.

baseline in Figure 4 corresponds to a very good VCD baseline.

There are two principal sources of baseline deviations in the measurement of a VCD spectrum. One is the empty beam VCD baseline. This is the starting point of all VCD measurements. A second type of baseline deviation, or artifact, varies (in a complex way) with the absorbance of the sample. This second kind of artifact can be isolated, if it is present, by comparing the empty beam VCD spectrum, or the solvent VCD spectrum, with the VCD of the racemic mixture of the chiral sample under investigation. We make this comparison in Figure 5 for valine in H_2O, where in this case the empty beam includes the sample cell and H_2O solvent. The VCD spectrum of racemic valine should be zero since the sample has no net chirality. Inspection of Figure 5 reveals that the baseline VCD spectrum of the solvent and racemic valine are indeed very close and substantially within the noise level of the VCD measurement. Hence, there is only a baseline deviation of the first kind here, and this is easily removed by subtraction of the solvent VCD. If both kinds of baseline deviation were present, then the correct VCD spectrum could only be obtained by subtraction of the VCD spectrum of the racemic sample or, better still if available, the opposite enantiomer.

Although, it wasn't stated explicitly earlier, opposite enantiomers have opposite VCD spectra with respect to the baseline for the sample. This is illustrated in Figure 6 for L-valine and D-valine. The final spectrum in Figure 3 was obtained by subtracting the VCD spectrum of D-valine from that of L-valine and dividing by two. This subtraction removes all sources of artifacts. In many cases, however, both enantiomers are not available. For such cases, it is important that the VCD instrument is not sensitive to absorption artifacts, and ideally that the empty instrument or solvent artifacts are as small as possible.

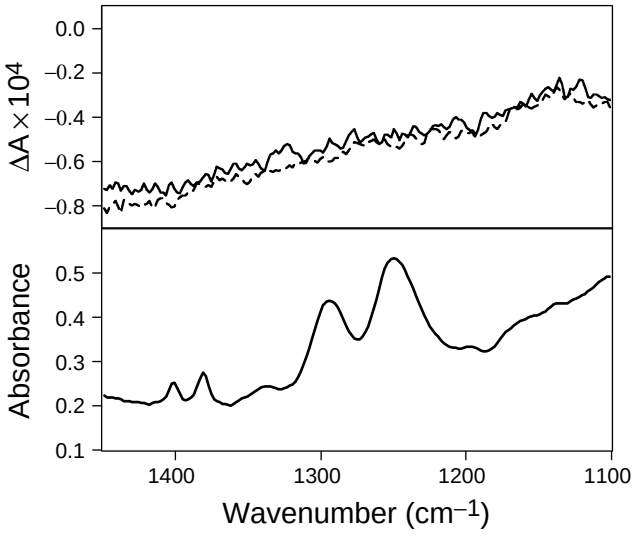

Figure 5 Comparison of the baselines for the solvent H_2O (dashed) and racemic valine (solid) in H_2O/HCl showing that both are the same to within the noise level. Spectra were obtained under the conditions of Figure 3. Absorbance spectrum shown is for the raw DL-valine prior to water subtraction.

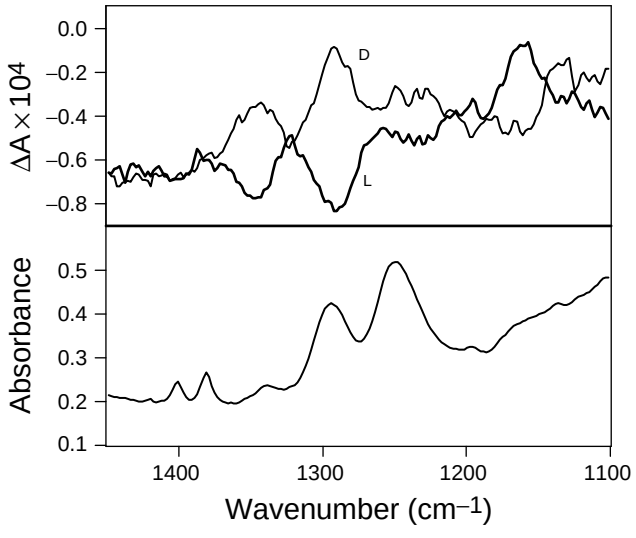

Figure 6 Comparison of the VCD spectra of L-valine and D-valine in H_2O/HCl prior to baseline correction. Spectra were obtained under the conditions of Figure 3. Absorbance spectrum shown is for raw L-valine prior to water subtraction.

It has been observed that absorption artifacts increase in importance (and annoyance) as the empty instrument baseline deviates more significantly from true zero.

The reliable subtraction of VCD spectra obtained from successive measurements, requires good instrument stability over the time of both measurements. Stability can be checked by running a series of VCD spectra to look for variations or baseline wander outside the expected

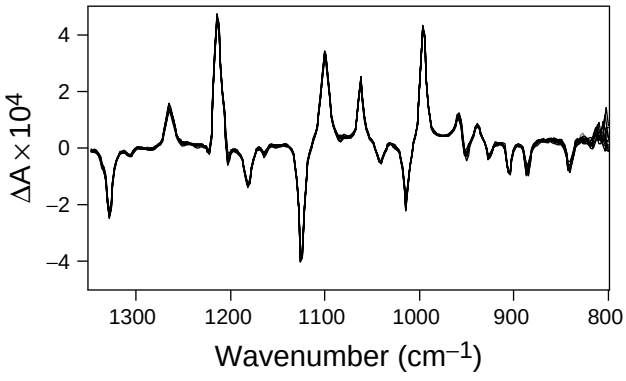

Figure 7 Superposition of 12 VCD spectra of (−)-α-pinene, neat liquid, for 20-min collection time per measurement.

noise level. This can be checked in several ways. One is to determine a noise curve for each VCD measurement that is the result of dividing the VCD collection into two successive blocks. Adding these blocks yields the final VCD spectrum and subtracting removes all the VCD features but leaves only the effects of noise and baseline instability. Examples of such noise curves are presented in Figures 2 and 3. From these the noise level across the spectrum is more apparent than it is in the spectrum itself, and second, the flatness and position of the noise curve at the zero level attests to the stability of the two measurements.

Another approach to testing VCD instrument stability is presented in Figure 7 where the results of 12 successive 20-min measurements of (−)-α-pinene are presented. The spectra were measured over a 4-h period without any interruptions or sample changes. A high level of reproducibility is observed with no systematic drift outside the noise level of the measurements. Below approximately 850 cm^{-1} the noise level increases owing to absorption in the beam from BaF_2 windows in the sample cell.

3.4 Intensity Calibration and Reproducibility

As mentioned above, Equation (14) describes a calibration curve obtained from the nonzero intersection points of a pair of VCD calibration curves. These intersection points and the resulting calibration represent unit VCD intensity at each wavenumber frequency. In Figure 8 the two curves are shown with their connected intersection points. The resulting calibration curve is shown for a setting of the PEM with maximum VCD efficiency near 1100 cm^{-1}. For the region near this frequency the calibration curve is nearly the same value. The decrease in VCD efficiency toward higher frequencies is the result of lower values of the Bessel function and the exponential lock-in decay factor which disadvantages higher frequencies.

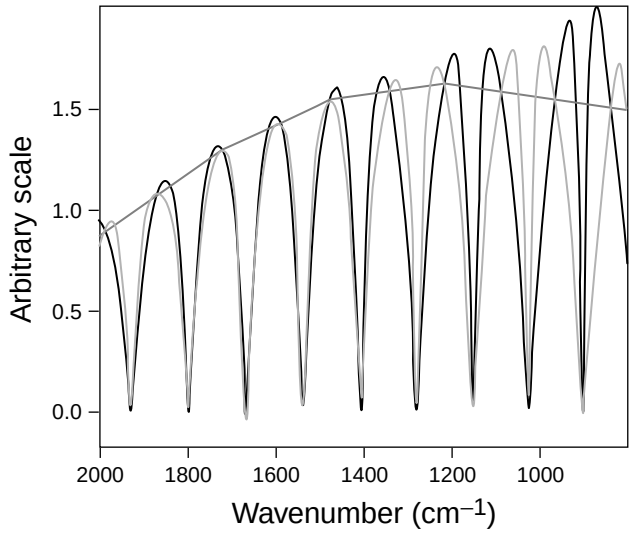

Figure 8 The VCD waveplate curves and the VCD calibration curve derived from connecting intersection points on the waveplate curves.

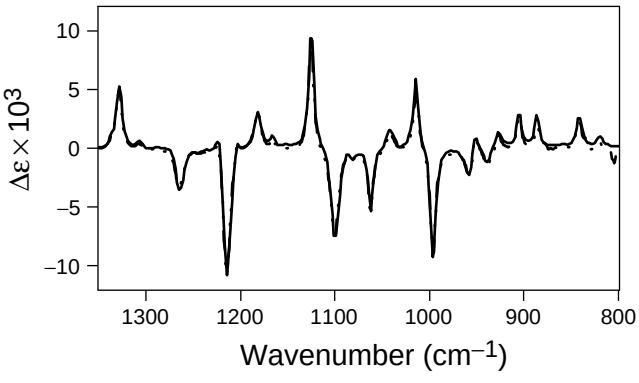

Figure 9 Comparison of the absolute intensities of the VCD spectrum of (+)-α-pinene measured on the Chiralir™ at the University of Southern California (solid line) and a home-built instrument at Syracuse University (dashed line).

Presently, work is in progress to define an absolute intensity standard for VCD based on the spectrum of (−)-α-pinene. Spectra for this sample and one other have been obtained from a number of laboratories where VCD instruments are located. To illustrate the level of agreement obtained for two such instruments, Figure 9 shows VCD spectra of (−)-α-pinene from two laboratories, the laboratory of Philip Stephens at the University of Southern California using the Chiralir™ (solid line) and the laboratory of one of the co-authors at Syracuse University using a homebuilt VCD instrument based on Nicolet 850 spectrometer (dashed line). Both VCD spectra are presented in terms of the absorption coefficient, $\Delta\varepsilon$, to eliminate the effects of differing path length, and hence different Δ absorbance values, for the

two measurements. These two VCD measurements from instruments of different design are in relatively close agreement. Differences that do exist arise from variations in the way the calibration measurement was executed and possible errors in the path lengths of the cells used. It is expected that with sufficient care, agreement for the absolute intensity values of the (−)-α-pinene VCD spectrum in the mid-IR region can be determined to within a few percent. Once this has been accomplished, a new way to calibrate VCD, namely relative to the (−)-α-pinene spectrum, can be established.

3.5 Signal-to-noise Ratio

An important variable in the measurement of VCD spectra is the S/N. All VCD spectra, no matter how large, take longer to acquire than the corresponding FTIR absorbance spectra. The disadvantage of four orders of magnitude is overcome in part owing to the sophisticated electronic pathway that permits isolation and amplification of the VCD signal independently of processing the ordinary IR spectrum. The VCD spectra are generally too small to obtain by simply obtaining the absorbance spectra of LCP and RCP separately and subtracting.

To obtain a VCD spectrum at the desired S/N, one needs to decide how long to collect the spectrum. Figure 10 shows the VCD and noise spectra for a single enantiomer, with no baseline correction, of (−)-α-pinene after 1 min of collection, 5 min of collection, and 20 min of collection. Dramatic increases in the S/N can be seen. The general rule of signal averaging for FTIR and FT/VCD is that the S/N improves as the square root of the increase in time of collection. Thus the 5 min scan has lower noise than the 1 min scan by a factor of $1/\sqrt{5}$, and the improvement gained in collecting for 20 min instead of 5 is a factor of 2. For long scans it advisable to collect the spectra in blocks, say 1 h each, so that if the liquid nitrogen runs out, the sample degrades, or some undesired change in the measurement occurs, the prior blocks of data will be available and safe.

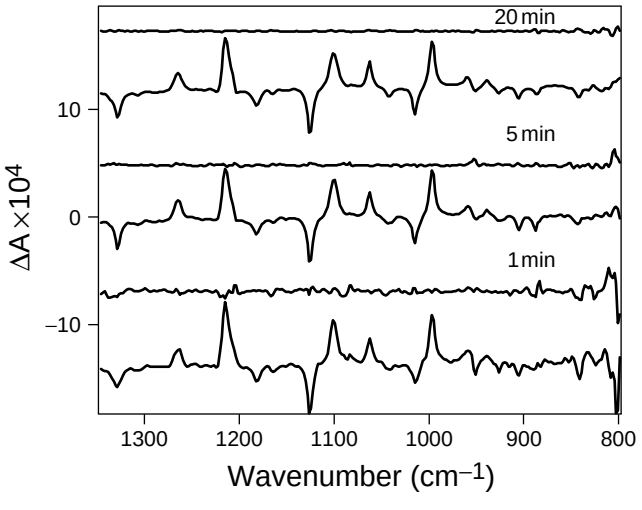

Figure 10 The VCD spectra and noise curves (above) for (−)-α-pinene under the sampling conditions of Figure 2 for collection times listed showing improvement in spectral quality with increased collection time.

4 BIOLOGICAL EXAMPLES

Naturally occurring biological molecules of all varieties can be measured by VCD. These include amino acids, as illustrated above in the case of L-valine, peptides, proteins, nucleic acids, and sugars. This section provides further examples of biological molecules of several classes. An important issue to resolve for the sampling of any biological sample is the solvent. Water (H_2O or D_2O) is an important solvent for biological samples. All but a few solvents have regions of intense absorption in particular regions of the mid-IR. A useful preliminary exercise when considering a solvent is to measure the transmission spectrum of the pure solvent near the path length of interest for the sample measurements. In Figure 11 the transmission spectra for H_2O and D_2O are presented where, for the sake of direct comparison, a path length of 12 μm is used for both solvents. In practice, shorter path lengths are generally used for H_2O and longer ones for D_2O.

4.1 Proteins

Proteins exhibit strong VCD in the amide I and II regions of the IR spectrum, and the chirality of the constituent amino acids drives the overall sense of the helical structure.[20,21] In Figure 12 four examples of protein VCD are shown. In Figure 12(a) and 12(b) the VCD

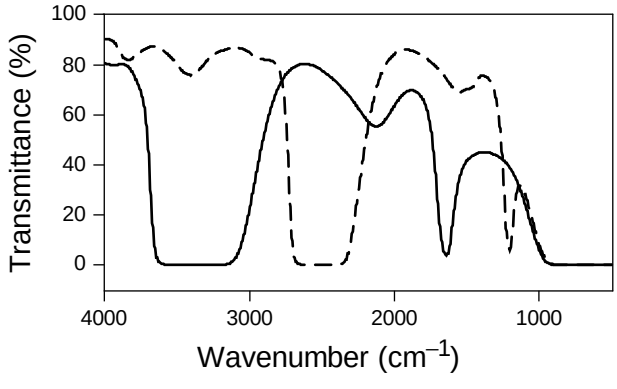

Figure 11 Percentage transmission spectra for H_2O (solid) and D_2O (dashed) at a path length of 12 μm.

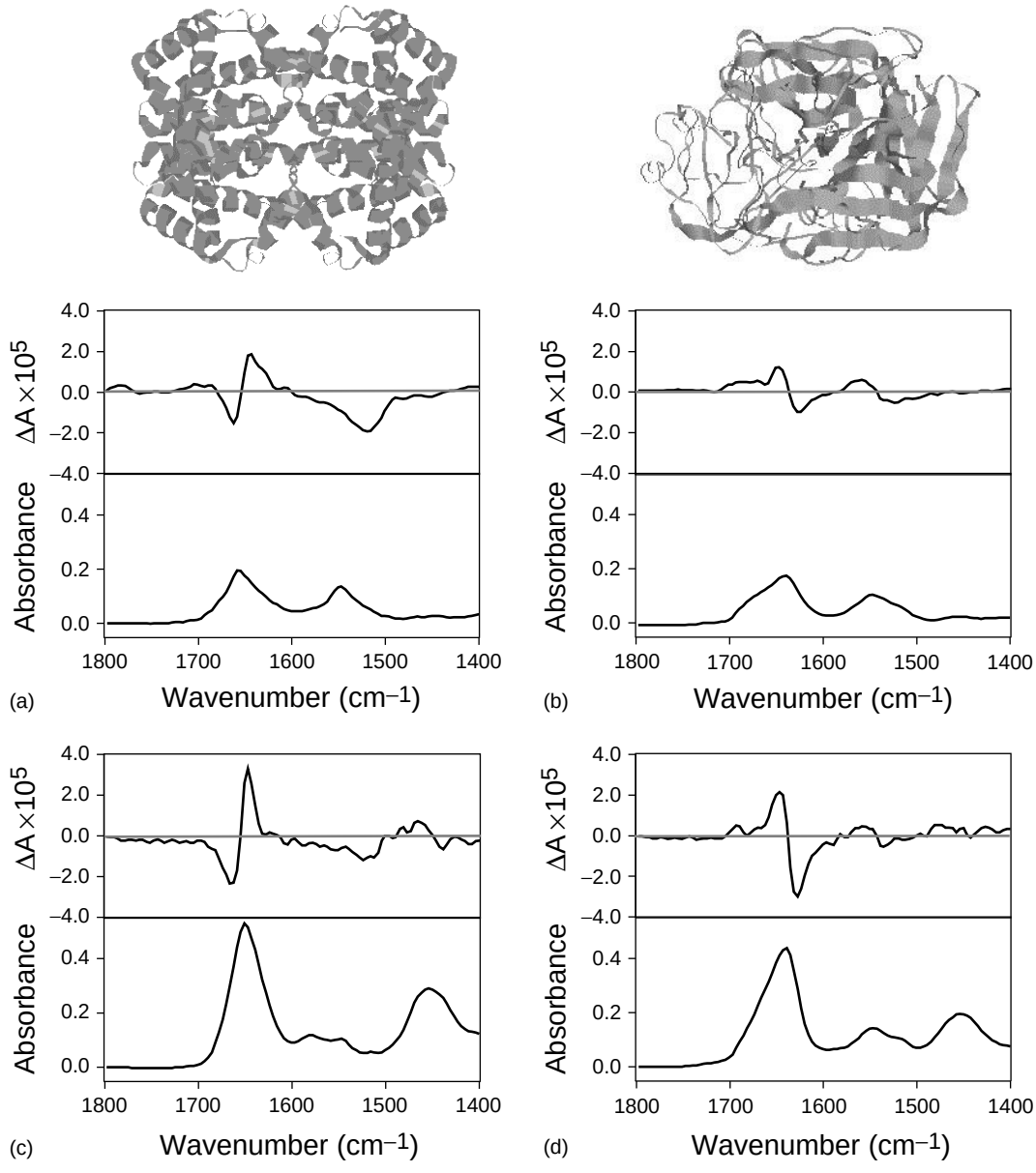

Figure 12 Absorbance and VCD spectra in the amide I and II regions for two proteins: (a) hemoglobin in H_2O, (b) chymotrypsin in H_2O, (c) hemoglobin in D_2O, and (d) chymotrypsin in D_2O. Spectra in H_2O were measured at a concentration of 150–200 mg mL^{-1} and path lengths of 4–6 μm and spectra in D_2O were measured at a concentration of 20 mg mL^{-1} and 80-μm path lengths.

and IR spectra for hemoglobin and α-chymotrypsin in H_2O, respectively, are presented, and the corresponding spectra in D_2O are given in Figure 12(c) and 12(d). The measurements were made with the Chiralir™ at a spectral collection time of 1 h for hemoglobin and 3 h for chymotrypsin at a resolution of 8 cm^{-1}. For the H_2O measurements the protein concentrations were quite high at 150–200 mg mL^{-1} and correspondingly small path lengths of approximately 4–6 μm. For the D_2O measurements the concentration was lowered by a factor of 10, down to the biologically more relevant value of 20 mg mL^{-1} and the path length was increased to 80 μm. The absorbance spectra in all cases are corrected for solvent absorption, and the two prominent bands, near 1650 cm^{-1} and 1530 cm^{-1} are the amide I and II bands, respectively. In D_2O, small changes occur in the amide I region and the amide II band shifts to lower frequency, near 1450 cm^{-1}, although some unexchanged amide groups still remain. The secondary structure of hemoglobin under these conditions is predominantly α-helix, whereas the structure of α-chymotrypsin is primarily β-sheet. Both the VCD and IR spectra provided here

are typical for proteins with this secondary structure, and the distinct differences in these two cases reflect the sensitivity of VCD to the conformation of protein structure.

4.2 Nucleic Acids

Nucleic acids are chiral by virtue of the sugar–phosphate backbone which in turn drives the helical structure of the entire molecule. The VCD technique is sensitive to all regions of nucleic acids,[22] and in particular to the base stacking regions between 1750 and 1550 cm^{-1}. In Figure 13, the VCD and IR spectra of salmon sperm DNA dissolved in D_2O is presented. Under these conditions, the DNA helix is in the B-form conformation. The measurements were made with the Chiralir™ at a spectral collection time of 2 h at a resolution of 8 cm^{-1}. The concentration of DNA was 15 mg mL^{-1} and the path length was 75 μm. The VCD and IR spectra were corrected by subtraction of the corresponding background D_2O spectrum.

4.3 Sugars

Sugars have multiple centers of chirality and exhibit strong VCD spectra. Their conformation is easily studied in a variety of solvents. In Figure 14, the VCD and IR spectra of α-D-glucose in dimethylsulfoxide-d_6 (DMSO-d_6) is presented.[23] The solvent VCD and IR background spectrum were subtracted to obtain the result presented. The lower spectral limit is 1100 cm^{-1} because the solvent absorbs strongly below that point. The concentration used was 0.5 M or 90 mg mL^{-1} and the path length was 70 μm.

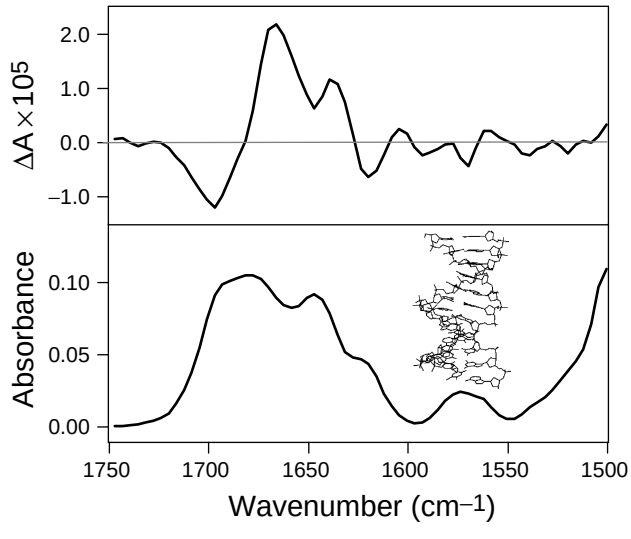

Figure 13 Absorbance and VCD spectra of salmon sperm B-DNA in D_2O solution at a concentration of 15 mg mL^{-1}, 75 μm path length, 2-h collection time.

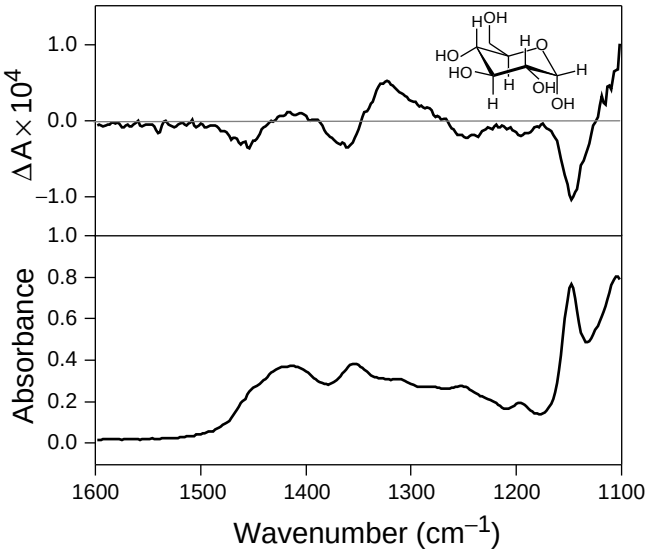

Figure 14 Absorbance and VCD spectra of α-D-glucose in DMSO-d_6 solution at a concentration of 90 mg mL^{-1} and a path length of 70 μm.

The spectral collection time was 3 h using the Chiralir™ VCD spectrometer at Vanderbilt University. The VCD from sugars have strong contributions in the mid-IR from modes associated with C–O and C–C stretching and various bending modes.

5 PHARMACEUTICAL EXAMPLES

Increasingly the development of chirally pure pharmaceutical molecules is gaining importance. These molecules offer a rich variety of structural motifs for intended purposes in many different areas of application. In this section, two examples of chiral pharmaceutical molecules are provided as examples of the structure complexity in this class of molecules. The VCD and IR spectra are equally varied and offer the prospect of studying and specifying the stereochemistry of these interesting molecules.[24]

Ibuprofen is a common over-the-counter analgesic molecule. The VCD and IR spectra of (*R*)-(−)-ibuprofen are presented in Figure 15 in the region between 1600 and 800 cm^{-1}. The spectrum was obtained using a 75-μm cell at 4 cm^{-1} resolution on the Chiralir™. The collection time was 1 h per enantiomer where the final spectrum is half the subtraction of the two enantiomers. Ibuprofen was sampled as a solution in $CDCl_3$ at a concentration of 250 mg mL^{-1}. The absorbance spectrum was corrected for solvent absorption.

As a second example, the VCD and IR spectra of (*R*)-warfarin in $CDCl_3$ are presented in Figure 16. The spectra

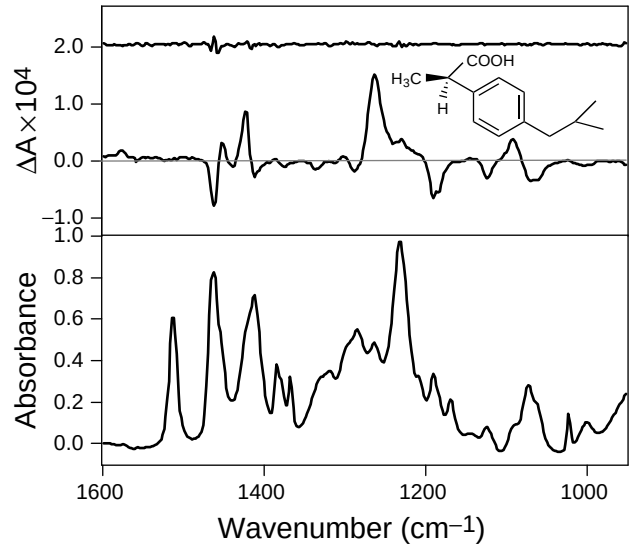

Figure 15 Absorbance (bottom), VCD (middle), and VCD noise (top) spectra of (*R*)-(−)-ibuprofen in $CDCl_3$ at a concentration of 250 mg mL^{-1}, path length of 75 μm and a collection time of 1 h for each enantiomer.

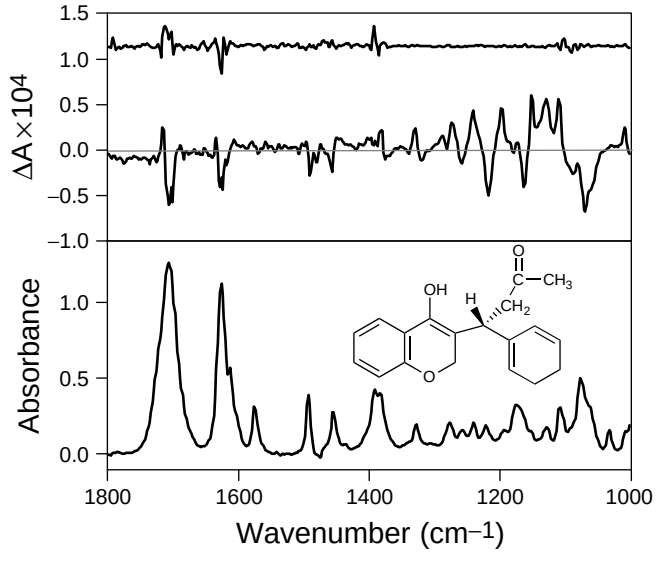

Figure 16 Absorbance (bottom), VCD (middle), and VCD noise (top) spectra of (*R*)-warfarin in $CDCl_3$ at a concentration of 8.6 mg mL^{-1}, path length of 500 μm and a collection time of 3 h for a single enantiomer.

were measured on the Chiralir™ at 4 cm^{-1} resolution as a single enantiomer in a 3 h collection period. The concentration of the sample was 8.6 mg mL^{-1} at a path length of 500 μm. The VCD intensity in this spectrum is approximately 10 times smaller than that of α-pinene as can be discerned from the VCD intensity scale for similar IR absorbance intensities. This demonstrates that molecules with relatively small VCD can still be measured with good signal quality. Note the richness of the vibrational band structure for this molecule across a wide range of frequencies.

6 AREAS OF APPLICATION

VCD, and ROA, can be used for a wide variety of stereostructural analyses. The applications of VCD fall into three areas. The first is conformational analysis, and we have seen in the previous figures examples of VCD spectra where the principal objective in obtaining the spectra has been to study the solution conformations of the selected biological molecules. This is the oldest and most basic use of VCD spectra. Since the mid 1970s, VCD has been used as a research tool in academia to study the conformational properties of molecules where the absolute configurations were well known and the optical purity was assumed to be as high as possible, typically near 100%.

The other two areas of application of VCD are more practical and have come to the fore with the increasing interest by the chemical and pharmaceutical industries in the potential of VCD for determining the optical purity of manufactured samples and the absolute configuration of new chiral molecules intended for medicinal application. This section provides examples of the use of VCD in these two areas.

6.1 Enantiomeric Excess

Prior to VCD there were primarily three techniques used for the determination of enantiomeric excess. The oldest method is the measurement of optical rotation. If the rotational angle for the pure compound is known, measurement of the optical rotation provides a measure of optical purity. The principal shortcomings of this methods are that, first, only one spectral datum is associated with the measurement and, second, the value is very sensitive to the density of the sample and hence its temperature. Optical rotation has been replaced over the years by two new techniques. One is NMR shift reagent measurements and the other is chiral chromatography. In the NMR method, a chiral NMR shift reagent is added to the sample as a perturbant. Opposite enantiomers in the sample will experience shifts in some of their NMR resonances in opposite directions, and the areas of these shifted bands can be compared to determine the relative percentages of the two enantiomers. The drawbacks here involve finding suitable shift reagents, and difficulties arising when multiple chiral centers are present. In addition, the sample is altered in the process of measurement. In chiral chromatography a physical separation of the two enantiomers is effected by a chiral substrate in a column. If

the relative areas in the separated bands from the column are measured then the optical purity can be determined. The accuracy of this method for favorable cases is in the range of 0.1% enantiomeric excess. The percentage enantiomeric excess is defined as the difference in the measured quantities for two enantiomers divided by their sum and multiplied by 100. The principal difficulties in chiral chromatography are the time and expense of finding an efficient chiral substrate and maintaining the column.

VCD offers an alternative to these methods that has certain advantages, with some disadvantages. As shown above, VCD spectra have multiple bands that permit multiple points of chiral interrogation of the sample. Unlike optical rotation, VCD measurements are not particularly sensitive to temperature, and any changes in density that do occur are not multiplied by long path lengths and can be cancelled with the same changes that occur in the parent IR spectrum. In contrast to NMR shift reagents no foreign additive to the sample is needed to make the measurement and, further, no physical separation of enantiomers is required as is the case for chiral chromatography. One merely needs a good IR spectrum and a VCD instrument. The sample may be recovered entirely from the measurement. With appropriate sampling techniques, it should be possible to use VCD with multiple sampling or even flow sampling techniques. The principal disadvantage of VCD is the practical limit of approximately 1–0.5% ee obtained to date; however, improved techniques using chemometric methods are being developed.

An example of the use of VCD for a relatively difficult sample is given in Figure 17. Here are shown a set of VCD spectra for L-valine in HCl/H_2O in the region 1400–1200 cm^{-1} measured with the Chiralir™. Two bands are used for analysis, but many data points across each spectral curve are used in the determinations. The solutions were prepared by volumetric quantitative methods to be 100, 90, 80, 60, 40 and 20% ee. The VCD

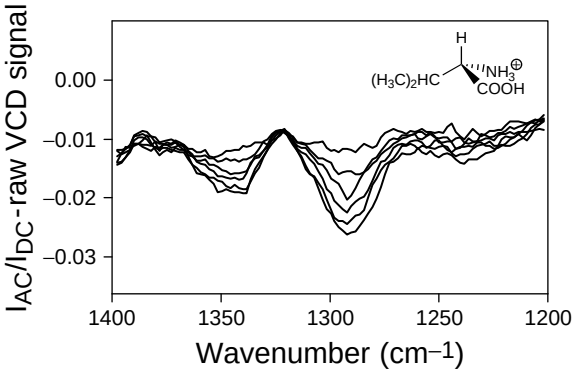

Figure 17 The VCD spectra of L-valine in H_2O/HCl under conditions of Figure 3 for single enantiomer at percentage ee values of 100, 90, 80, 60, 40, and 20.

data were fitted to these known values using partial least squares (PLS) analysis and the optimized values obtained for these samples, as determined by VCD, were 100.5, 88.5, 78.5, 59.6, 44.3, 18.5% ee. The error range for these six determinations is 0.4–4.3% and the standard error (the square root of a result of 5 divided into the sum of the squares of the errors) is 2.26%. Unknown errors may also be present in preparation of samples, although they are presumed to be smaller than VCD spectral determination of percentage ee. Samples with higher intrinsic noise levels have produced errors as low as 0.4% ee, but the case here may be a more typical result. Thus VCD is available as a straightforward method with which to reliably determine percentage ee under a variety of nonperturbing sampling conditions. If only a qualitative monitor of the percentage ee, say to within 2% ee, and the sample integrity were desired, results could be available off-line on a 10–20 min cycle time or less depending on the S/N of the VCD measurement.

6.2 Absolute Configuration

Presently, there is only one conventional spectroscopic method for the determination of the absolute configuration of an unknown chiral molecule without reference to a prior structure or correlation. That is single-crystal X-ray diffraction. For this method the molecule of interest must form a pure single crystal of sufficient size for the X-ray analysis. In most cases use of a heavy atom crystal dopant is necessary to establish the absolute phase, and hence the absolute stereochemistry.

VCD now offers a viable alternative to X-ray analysis. As VCD spectra are usually measured in the solution phase, it is not necessary to grow a crystal or dope the crystal for analysis. The absolute stereochemistry is established by comparing the solution-phase VCD spectrum to the results of an ab initio quantum chemistry calculation. The absolute configuration used for the calculation is known. If the measured VCD agrees with the theoretical spectrum, the absolute configuration of the sample molecule is the same as that calculated, and if the experimental signs are opposite to the corresponding theoretical ones, so then is the absolute configuration of the sample molecule. In addition to this definitive method of determination of absolute configuration without reference to a standard compound, VCD can be used empirically to show the stereochemical relationship between two closely related molecules.

In principle, electronic CD could also be used in the same way for absolute configuration determination; however, there are several difficulties. First, there are usually only a few bands rather that a dozen or so for points of comparison. Second, calculating the electronic CD spectrum requires good quantum mechanical descriptions of both the ground and the excited electronic states,

whereas only the former is needed for VCD calculations; VCD arises from only the vibrational modulation of the ground electronic state and is not a two-electronic-state phenomenon. It is well known that excited electronic states are much more difficult to calculate than the corresponding ground electronic state. Finally, not all molecules have electronic states in the accessible region of the electronic spectrum, as is the case below.

In Figure 18, the calculated and observed VCD and IR spectra for (*S*)-(−)-*trans*-pinane is shown. The absolute stereochemistry of this molecule is also displayed in the figure. Note that the molecule is a pure hydrocarbon and has no functional groups. Hence its electronic CD spectrum would be difficult, if not impossible, to measure with a conventional electronic CD instrument. Inspection of the figure reveals a high level of agreement, both in signs, relative magnitudes, and absolute magnitudes of all the bands. No adjustable parameters were invoked in the calculation, only a choice of bandwidth for the theoretical intensities. The experimental spectrum was obtained with the Chiralir™ from a neat liquid using a path length of 55 μm and 4 cm^{-1} resolution. The theoretical spectrum was obtained using the commercial software program Gaussian 98 using density functional theory, a 6-31G(d) basis set, B3LYP hybrid functionals, 54 h of cpu time for the frequency and VCD calculation on a Silicon Graphics SGI-02 workstation. Because this molecule is rigid, the determination of the equilibrium conformation is simplified to one conformer, but absolute configuration determination by VCD has been successfully demonstrated for cases where more than one conformer is important or potentially important.

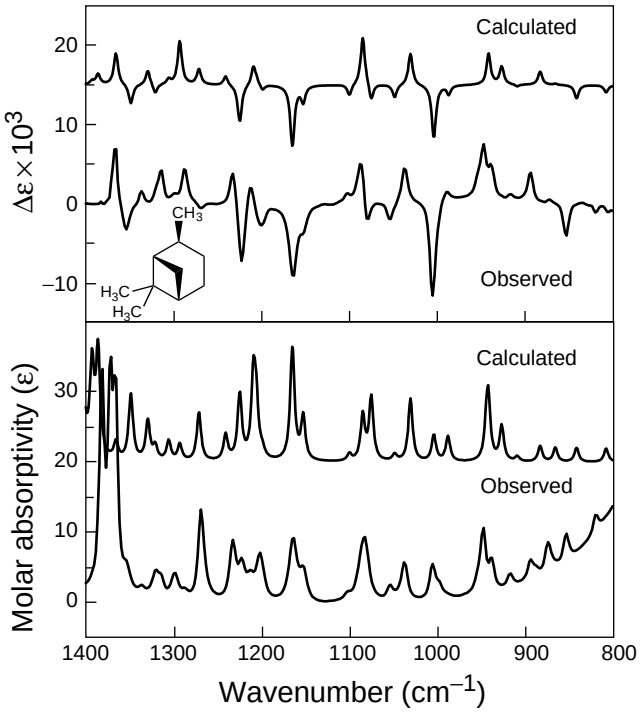

Figure 18 Comparison of the absorbance and VCD spectra of (*S*)-(−)-*trans*-pinane for experimentally measured (observed) and theoretically predicted (calculated) spectra.

7 CONCLUSIONS

VOA, consisting of IR/VCD and vibrational ROA, has evolved to a point of routine stereochemical application. This article has introduced the basic principles of these two closely related techniques and has illustrated the principal areas of application of VCD, namely conformational analysis, optical purity measurement, and absolute configuration determination. The availability of commercial instrumentation for VCD such as the Chiralir™ from Bomem/BioTools and ab inito molecular orbital software for VCD intensity calculation in Gaussian 98 from Gaussian Inc. have hastened the advance of VCD for widespread application beyond the laboratories of the VCD specialists. Similar advances will eventually be made for ROA, and then both approaches to VOA will be available for broad use in the study of chiral molecules.

ACKNOWLEDGMENTS

We would like to thank Dr Jean-Rene Roy of Bomem Inc., Professor Teresa B. Freedman of Syracuse University, Professor Prasad L. Polavarapu of Vanderbilt University, Dr Frank Devlin of University of Southern California, Dr Holger Hartwig of DSM, and Dr Anders Holmen of Astra/Zeneca for assistance in measuring or the calculation of some of the spectra presented in this paper.

ABBREVIATIONS AND ACRONYMS

CD	Circular Dichroism
CP	Circularly Polarized
DCP	Dual Circularly Polarized
FTIR	Fourier Transform Infrared
FT/VCD	Fourier Transform/Vibrational Circular Dichroism
ICP	Incident Circularly Polarized
IR	Infrared
LCP	Left Circularly Polarized
PEM	Photoelastic Modulator
PLS	Partial Least Squares
RCP	Right Circularly Polarized

ROA	Raman Optical Activity
RROA	Resonance Raman Optical Activity
SCP	Scattered Circularly Polarized
S/N	Signal-to-noise Ratio
VCD	Vibrational Circular Dichroism
VOA	Vibrational Optical Activity

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Circular Dichroism in Analysis of Biomolecules • Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules

Forensic Science **(Volume 5)**
Chiroptical Spectroscopy in Drug Analysis

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis • Peptide Diastereomers, Separation of

Pesticides **(Volume 7)**
Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis • Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis • Vibrational Spectroscopy in Drug Discovery, Development and Production

Chemometrics **(Volume 11)**
Chemometrics

Electronic Absorption and Luminescence **(Volume 12)**
Circular Dichroism and Linear Dichroism

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs

REFERENCES

1. L.A. Nafie, 'Infrared and Raman Vibrational Optical Activity: Theoretical and Experimental Aspects', *Ann. Rev. Phys. Chem.*, **48**, 357–386 (1997).
2. L.A. Nafie, 'Vibrational Optical Activity', *Appl. Spectrosc.*, **50**(5), 14A–26A (1996).
3. L.A. Nafie, G.-S. Yu, X. Qu, T.B. Freedman, 'Comparison of IR and Raman Forms of Vibrational Optical Activity', *Faraday Discuss.*, **99**, 13–34 (1994).
4. L.A. Nafie, T.B. Freedman, 'Vibrational Circular Dichroism: An Incisive Tool for Stereochemical Applications', *Enantiomer*, **3**, 283–297 (1998).
5. L.A. Nafie, D. Che, 'Theory and Measurement of Raman Optical Activity', in *Modern Nonlinear Optics, Part 3*, eds. M. Evans, S. Kielich, John Wiley & Sons, New York, 105–149, Vol. 85, 1994.
6. L.A. Nafie, 'Theory of Resonance Raman Optical Activity: The Single-electronic State Limit', *Chem. Phys.*, **205**, 309–322 (1996).
7. W. Hug, 'Instrumental and Theoretical Advances in Raman Optical Activity', in *Raman Spectroscopy*, eds. J. Lascombe, P.V. Huong, John Wiley & Sons, New York, 3–12, 1982.
8. W. Hug, 'Raman Optical Activity, Spectrometers', in *Encyclopedia of Spectroscopy and Spectrometry*, eds. J. Lindon, G. Tranter, J. Holmes, Academic Press, London, 1966–1976, 1999.
9. G. Holzwarth, E.C. Hsu, H.S. Mosher, T.R. Faulkner, A. Moscowitz, 'Infrared Circular Dichroism of Carbon–Hydrogen and Carbon–Deuterium Stretching Modes. Observations', *J. Am. Chem. Soc.*, **96**, 251–252 (1974).
10. L.A. Nafie, J.C. Cheng, P.J. Stephens, 'Vibrational Circular Dichroism of 2,2,2-Trifluoro-1-phenylethanol', *J. Am. Chem. Soc.*, **97**, 3842 (1975).
11. L.A. Nafie, T.A. Keiderling, P.J. Stephens, 'Vibrational Circular Dichroism', *J. Am. Chem. Soc.*, **98**, 2715–2723 (1976).
12. C.N. Su, V.J. Heintz, T.A. Keiderling, 'Vibrational Circular Dichroism in the Mid-infrared', *Chem. Phys. Lett.*, **73**, 157–159 (1980).
13. E.D. Lipp, C.G. Zimba, L.A. Nafie, 'Vibrational Circular Dichroism in the Mid-infrared Using Fourier Transform Spectroscopy', *Chem. Phys. Lett.*, **90**, 1–5 (1982).
14. L.A. Nafie, F. Long, T.B. Freedman, H. Buijs, A. Rilling, J.-R. Roy, R.K. Dukor, 'The Detemination of Enantiomeric Purity and Absolute Configuration by Vibrational Circular Dichroism Spectroscopy', in *Fourier Transform Spectroscopy: 11th International Conference*, ed. J.A. deHaseth, American Institute of Physics, Woodbury, NY, 432–434 Vol. 430, 1997.

15. L.D. Barron, M.P. Bogaard, A.D. Buckingham, 'Raman Scattering of Circularly Polarized Light by Optically Active Molecules', *J. Am. Chem. Soc.*, **95**, 603–605 (1973).
16. W. Hug, S. Kint, G.F. Bailey, J.R. Scherer, 'Raman Circular Intensity Differential Spectroscopy. The Spectra of (−)-α-Pinene and (+)-α-Phenylethylamine', *J. Am. Chem. Soc.*, **97**, 5589–5590 (1975).
17. M. Vargek, T.B. Freedman, L.A. Nafie, 'Improved Backscattering Dual Circular Polarization Raman Optical Activity Spectrometer with Enhanced Performance for Biological Applications', *J. Raman Spectrosc.*, **28**, 627–633 (1997).
18. L. Hecht, L.D. Barron, 'Instrumentation for Raman Optical-activity Measurements', *J. Mol. Struct.*, **347**, 449–458 (1995).
19. L.A. Nafie, 'Polarization Modulation FTIR Spectroscopy', in *Advances in Applied FTIR Spectroscopy*, ed. M.W. Mackenzie, John Wiley & Sons, New York, 67–104, 1988.
20. V. Baumruk, T.A. Keiderling, 'Vibrational Circular Dichroism of Proteins in H_2O Solution', *J. Am. Chem. Soc.*, **115**, 6939 (1993).
21. T.A. Keiderling, 'Vibration Circular Dichroism Spectroscopy of Peptides and Proteins', in *Circular Dichroism, Principles*, eds. K. Nakanishi, N.D. Berova, R.W. Woody, VCH, New York, 497–521, 1994.
22. L. Wang, T.A. Keiderling, 'Vibrational Circular Dichroism Studies of the A-to-B Conformational Transition in DNA', *Biochemistry*, **31**(42), 10 265–10 271 (1992).
23. P.K. Bose, P.L. Polavarapu, 'Vibrational Circular Dichroism of Monosaccharides', *Carbohydr. Res.*, **319**, 172–183 (1999).
24. T.B. Freedman, F. Long, M. Citra, L.A. Nafie, 'Hydrogen-stretching Vibrational Circular Dichroism Spectroscopy: Absolute Configuration and Solution Conformation of Selected Pharmaceutical Molecules', *Enantiomer*, **4**, 103–119 (1999).

Voltammetry In Vivo for Chemical Analysis of the Living Brain

Robert D. O'Neill
University College Dublin, Dublin, Ireland

John P. Lowry
National University of Ireland, Maynooth, Ireland

1 Introduction **2**
1.1 Intact Brain Preparation 2
1.2 Other Brain Preparations 3
1.3 Legal and Ethical Considerations 4
2 History **4**
2.1 Original Interference Problem 5
2.2 Analytical Problems of Voltammetry In Vivo 5
3 Substances Targeted with Voltammetry In Vivo **6**
4 Electrode Types and Preparation **6**
4.1 Auxiliary and Reference Electrodes 6
4.2 Platinum 7
4.3 Carbon Epoxy 7
4.4 Carbon Fiber 7
4.5 Carbon Paste 8
4.6 Biosensors 9
5 Voltammetric Techniques Used In Vivo **10**
5.1 Chronoamperometry 11
5.2 Linear Ramps 11
5.3 Pulsed Techniques 12
5.4 Constant Potential Amperometry 12
6 Surgical Protocol for Voltammetry In Vivo Experiments **12**
6.1 General Considerations 12
6.2 Anesthesia 12
6.3 Surgery and Recovery 13
6.4 Assessment of Animal Health 14
6.5 Experimental Conditions 14
6.6 Hazard Assessment 14
7 Detection of Substances in Brain Extracellular Fluid **14**
7.1 Electroactive Neuromediators 14
7.2 Electroactive Neuromediator Metabolites 15
7.3 Ascorbic Acid 16
7.4 Uric Acid 17
7.5 Dioxygen 17
7.6 Nitric Oxide 17
7.7 Glucose and Lactate 18
7.8 Glutamate 19
8 Comparison with Brain Microdialysis **20**
Acknowledgments **21**
Abbreviations and Acronyms **21**
Related Articles **21**
References **22**

The mammalian brain is the most complex structure known to science. Our behavior, feelings, thoughts, and maybe even consciousness itself, may be a reflection of the interplay of electrical and chemical signaling that create these states, and how the physical brain gives rise to the properties of mind remains a major unanswered question. It is clear, however, that many of the drugs used empirically in the treatment of neurological disorders, such as Parkinson's disease, schizophrenia and depression, as well as mind-altering substances of abuse, have specific chemical actions on nerve cells in the brain, highlighting the importance of chemical signaling in the functioning of neural networks.

A growing number of methodologies are being developed, including sampling, spectroscopic and electrochemical, to study neurochemical phenomena in the living brain. One such set of techniques focuses on the detection of substances released from nerve cells, using amperometric electrodes and voltammetry in vivo (VIV) techniques. By implanting a microvoltametric electrode in a specific brain region, applying a suitable potential profile and recording the resulting faradaic current, changes in the concentration of a variety of substances in the extracellular fluid (ECF) can be monitored with sub-second time resolution over extended periods. This allows investigations of the functions and roles of specific neurochemicals in neuronal signaling, drug actions, and well-defined behaviors. The main limitations associated with VIV is, on the one hand, the limited number of ECF species that are electroactive, and on the other, the fact that electroactive substances present tend to oxidize at similar potentials, interfering with each other's detection. These restrictions are being overcome by ongoing developments, including the design of biosensors to broaden the range of target analytes and permselective membranes to block interference.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Intact Brain Preparation

Mammalian behavior is a reflection of the activity of the central nervous system (CNS), and in particular the brain. The brain is adept in a wide range of functions, including: sensory input and processing; behavioral output or response; subtle colorings of the animal with feelings such as depression, anxiety and reward; memory; and in higher animals it is involved in phenomena such as language and imaginative thought.[1]

The substrates of these diverse functions are complex networks of complex units – nerve cells or neurons. Each neuron consists essentially of a cell body from which extend a large number of cytoplasmic projections. One of these projections, the axon, is usually much longer than the others, and may bifurcate several times; the remaining, smaller, extensions are the dendrites. The ends of the axon (nerve terminals) are normally swollen and contain specialized organelles for storing chemical messengers. Most neurons are electrically isolated from each other, and form networks by interfacing through anatomically well-defined gaps (synapses), usually between the terminal of one neuron and the cell body or dendrites of another (Figure 1).

Although the processing of electrical signals may be considered the primary function of nerve cells in the brain,[2] intercellular chemical signaling plays a major role in determining the properties of neural networks. The individual nerve cells normally communicate their state of excitation or inhibition to other neurons by releasing chemical messengers across their synapses, and the CNS uses a wide variety of intercellular chemical messengers (neuromediators) in synaptic signaling. (The term neuromediator is used here to include classical neurotransmitters, neuromodulators and neurohormones.) In addition, each released neuromediator substance can bind to a range of different receptor proteins embedded in the membrane of the target cell. Some receptors (ionotropic), when activated, open ion channels that lead to millisecond changes in the electrical potential across the membrane. Others (metabotropic) are linked to enzymes located on the intracellular side of the membrane; when activated these receptors can either accelerate or inhibit the production of intracellular 'second messenger' systems that can change the condition of the target neuron over extended periods of time. This combination of diverse neuromediators, receptors and mechanisms of interaction are some of the factors that enable the CNS to function over a vast spectrum of timescales, producing responses ranging from reflex actions to memories lasting many decades.

Figure 1 shows a schematic representation of intercellular chemical signaling at a synapse between a nerve terminal and target neuron. In the resting state, the nerve cell membrane (lipid bilayer containing a variety of attached and transmembrane proteins) is polarized at about −70 mV (negative on the inside) due to a transmembrane concentration gradient of K^+ (high inside) and the presence of K^+-selective

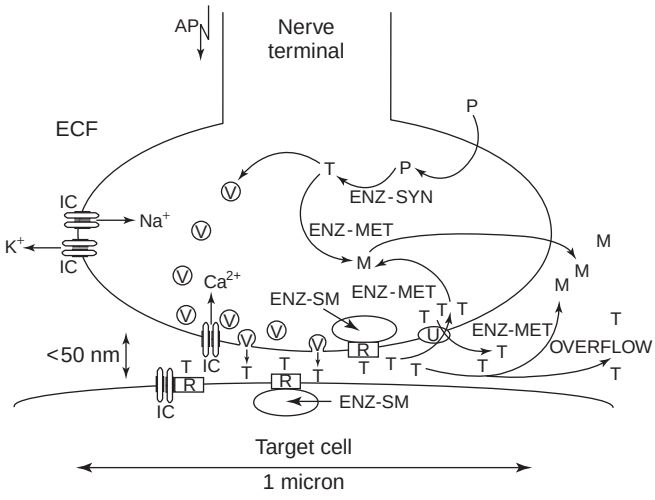

Figure 1 Schematic representation of intercellular chemical signaling at a synapse between a nerve terminal and target neuron. Neuromediator substance (T); precursor of T (P); enzymes involved in the synthesis of T (ENZ-SYN); vesicles for storing T (V); metabolites of T (M); enzymes involved in metabolism of T (ENZ-MET); action potential (AP); ion channels (IC); receptor complexes (R); second messenger enzyme (ENZ-SM); ion-coupled carrier uptake sites (U). For dopamine (DA)-releasing terminals in the striatum, nucleus accumbens and frontal cortex that have been the subject of many VIV studies, the following neurochemistry is involved. The precursor amino acid tyrosine is converted to 3,4-dihydroxyphenylalanine (DOPA) in the rate-determining synthetic step by tyrosine hydroxylase and then to DA by DOPA decarboxylase. The presynaptic metabolic enzyme is monoamine oxidase (MAO) (type A, MAO-A) that acts on DA to produce 3,4-dihydroxyphenylacetic acid (DOPAC). Both DA and DOPAC can be metabolized outside the terminal (on the postsynaptic membrane and in glial cells) by catechol-*O*-methyl transferase (COMT) to produce 3-methoxytyramine (3MT) and homovanillic acid (HVA) respectively. Because 3MT has a higher affinity for the DA uptake carrier than DA itself,[29] it is not surprising that the bulk of 3MT is further degraded by MAO to yield HVA, so that DOPAC and HVA are the main DA metabolites found in the ECF. The main presynaptic DA receptor is the D-2 receptor, linked to adenylate cyclase. Activation by DA leads to a decrease in the level of cyclic adenosine monophosphate (cAMP) in the terminal, an inhibition of Ca^{2+} release from intracellular stores, and the opening of K^+ channels, leading to hyperpolarization; two principal effects of these changes are a decrease in the activity of tyrosine hydroxylase lowering the synthesis rate of DA, and a decrease in DA release (negative feedback). The postsynaptic D-1 receptor is also linked to adenylate cyclase, but here activation by DA leads to stimulation of cAMP production. (Adapted from O'Neill[13] with permission from Royal Society of Chemistry.)

ion channels that are permanently open (leakage channels).

A precursor, such as an amino acid, is taken from the ECF into the terminal where one or a number of enzymes transform it into the neuromediator substance characteristic of the particular neuronal pathway. Although most of the newly synthesized neuromediator is taken up into storage vesicles, some will be inactivated before reaching the vesicles by metabolic enzymes present in the terminal. When an action potential (rapid depolarization, followed by repolarization of the membrane potential; total duration about 5 ms) propagates down an axon to the nerve terminal, voltage-sensitive Na^+ ion channels open, causing depolarization of the terminal membrane and the opening of voltage-sensitive Ca^{2+} channels, the influx of Ca^{2+}, and ultimately to the fusion of vesicles that release their neuromediator content into the synapse; a few milliseconds later, the opening of voltage-sensitive K^+ channels results in repolarization. This Ca^{2+}-dependent exocitotic release of neuromediator is surprisingly fast, occurring within microseconds of the start of the action potential in the terminal.[3]

Once in the synapse, the neuromediator molecules may bind to a range of different receptor proteins on either the presynaptic or postsynaptic membrane. On the presynaptic side, receptors may be linked to a 'second messenger' enzyme involved in regulating the activity of the rate-determining enzyme used in the synthesis of neuromediator (negative feedback control). Postsynaptically, receptors may also be linked to a second messenger system or to ion channels for Na^+, K^+, Ca^{2+} or Cl^- that control the membrane potential of the target neuron. Inactivation of the chemical signal is achieved either by re-uptake of neuromediator into the terminal (and other sites, such as glial cells) through specific ion-coupled carrier uptake sites (also capable of releasing cytoplasmic neuromediator by a Ca^{2+}-independent mechanism) and subsequently metabolized, or may be metabolized extracellularly. The metabolites can diffuse into the ECF that also contains small concentrations of neuromediator that overflow from the synapse. The concentrations of metabolites and neuromediator in the ECF can also be regulated by the rate of their excretion into the cerebrospinal fluid or across the blood–brain barrier (BBB).

One of the key elements in understanding the role of each neuromediator substance in neural networks is a knowledge of release rate, and Adams et al.[4] suggested that voltammetric electrodes implanted directly in the intact brain might be used to monitor electroactive neuromediators in the CNS (see section 2). However, electrodes generally employed for neurochemical analysis using VIV are ~5–300 μm in diameter and monitor an average concentration of analyte in the medium surrounding cells, the ECF, and not directly in synapses that are orders of magnitude smaller (typically 20–30 nm from terminal to target, see Figure 1). It is not clear at present what the relationship is between release of different neuromediators into the synapse and resulting ECF levels, although balance between release and re-uptake (Figure 1) is involved. It is also debatable whether neuromediators released into a synapse have additional extrasynaptic functions, so that the synapse might be designed to 'leak'. For example, extrasynaptic receptors have been reported for the main excitatory CNS neuromediator, glutamate,[5] as well as evidence for the diffusion of neuromediators, such as glutamate, outside the synaptic cleft.[6] An additional factor is that some neuromediators, such as norepinephrine (NE), are not released into anatomically well-defined synapses, but from simple swellings along the axon, so that their sphere of influence may extend far (up to 100 μm)[7] into the ECF.

Neuromediators and their metabolites in the ECF have been studied using VIV and a variety of perturbations that fall into two main categories (see Figure 1): those that cause depolarization of the terminal, including electrical stimulation, application of extracellular K^+, administration of agents that bind to receptors linked to ligand-gated Na^+ channels (e.g. glutamate), and behavioral activation (e.g. tail pinch); and those that cause changes in ECF levels by mechanisms not involving depolarization directly, such as inhibition of synthesis (e.g. α-methyl-*p*-tyrosine), displacing neuromediators from the terminal (e.g. amphetamine), inhibition of metabolic enzymes (e.g. pargyline), or blocking re-uptake (e.g. cocaine). A number of reviews on the principles and applications of voltammetric analysis of the ECF have been published, and a representative selection is cited here.[8–22] Original forms of some of the concepts presented in this article can be found in previous reviews.[13,19,21]

1.2 Other Brain Preparations

As a complementary approach to voltammetric recording in the living brain, the brain slice preparation has proved useful for studying the properties of individual neural pathways using voltammetric techniques.[15,20,22–25] A slice of brain containing the projection of interest (often the dopaminergic nigrostriatal system) is maintained in a physiological medium in vitro. This allows direct manipulation of the pathway with, for example, electrical or chemical stimulation, and the application of drugs affecting release, such as neuromediator synthesis inhibitors and receptor ligands (Figure 1), with reproducible results reported for up to 9 h.[26] Details of this preparation are given in article: **Neurotransmitters, Electrochemical Detection of**.

Pioneered by Schenk et al.,[27] another preparation of brain tissue, that has been used in voltammetric studies in vitro, is obtained by dissecting a particular brain region (usually the striatum) followed by its disruption, using a rotating disk electrode (RDE), to produce a suspension of whole tissue. With the RDE held at a constant potential to detect neuromediators, analogs or other agents added to either this suspension or the more classical synaptosomal preparation of striatal tissue, the mechanism of DA uptake,[28] ligand structure/activity relationships[29] and interactions with drugs of abuse[30] have been studied.

Understanding the functioning of brain neurons can also be advanced by studying the properties of other cell types[31–33] and peripheral neurons[34–38] isolated in vitro. Investigations of the chemistry within the cytoplasm of individual cells in vitro are now possible using submicron carbon ring, carbon fiber or Pt electrodes,[39–41] and intracellular Ni^{2+},[42] oxygen,[43] DA[44,45] and serotonin[46] have been monitored electrochemically. The majority of voltammetric reports on the behavior of isolated cells, however, have concentrated on the mechanism of release of catecholamines from adrenal medullary chromaffin cells[47–49] and serotonin from mast cells,[50,51] and signals associated with the release of neuromediator from single vesicles have been described.[32,47,52,53] Recent work on the release of peptide hormones from single cells has also been reported.[54,55]

1.3 Legal and Ethical Considerations

Legislation controlling the use of animals in scientific experiments varies from country to country, with some countries having strict legal requirements and others having poor, or in some cases no, legislation. For example, in the United Kingdom, legislation is enforced by the Home Office and is covered by the Animals (Scientific Procedures) Act 1986. The latter operates on a dual licence system: a personal licence for individual workers, specifying procedures and types of animals that may be used according to competence, training, qualifications and experience; and a project licence covering each specific programme of work. In the United States, the system is generally based on institutional licences being required for all projects and procedures. Large funding agencies for biomedical research, such as The Wellcome Trust (UK) and the National Institutes of Health (USA), also have very specific regulations concerning the welfare of animals and proper codes of practice. In addition to national legislation and institutional requirements, there are also international laws, such as European Directive 86/609/EEC, relating to the protection of animals used for experimental and other scientific purposes. Generally, all such policies are designed to control procedures that may cause pain, suffering, distress or lasting harm to animals used in scientific procedures. Provisions generally cover several areas, including a specification of the programme of work, the purpose and scientific justification for the work, a full description of the procedures involved (experimental protocol), the types of animals to be used, and an assessment of the potential severity of each procedure and the scientific programme as a whole.

The ethics of using animals in scientific experiments has become a very important issue. All use of animals in research for human benefit creates a dilemma; the justification for using the animal depends on it being different from the human, while the validity of the results obtained depends on the similarity of the animals, and their responses, to those of the human. Important considerations are:

- Can one justify what one is doing and does the potential benefit derived from the results outweigh the value of the animal? Potential benefit must be weighed against any pain, suffering (e.g. isolation or confinement) or lasting harm that may be caused to the animals used.
- Is it possible to achieve the same benefit by either carrying out an in vitro experiment, using a mathematical model, or using a different animal species lower in the phylogenetic scale?
- Can one use fewer animals by better experimental design, or by collaborating with colleagues so that as many tissues as possible from each animal are utilized?
- Can one alter the experimental method in some way to decrease the animal's potential suffering, for example, by an alteration in surgical technique, improvements in housing or bedding or by altering the drug dosage regime for a reduced volume or reduced frequency?

The last three considerations are often referred to as 'the three Rs' (replacement, reduction, and refinement) and form the backbone of any ethical review of animal research. Since this section is intended only to provide general guidelines relating to legislation and the ethics of using animals, the reader is urged to consult and examine texts for a more detailed consideration of the law and the ethics of animal use before embarking on such research.[56–58]

2 HISTORY

Although reports of voltammetry in the living brain go back at least as far as 1958 when Leland C. Clark demonstrated the feasibility of voltammetric recording in brain tissue in vivo for oxygen and ascorbic acid (AA)

detection,[59,60] the concept of applying voltammetric techniques to monitor changes in the concentration of electroactive neuromediators in brain ECF is generally attributed to Ralph N. Adams et al.[4] The paper describes attempts to detect neuromediator DA in anesthetized rat brain, using a carbon paste electrode (CPE) and cyclic voltammetry (CV). Even at this early stage, the authors successfully identified many of the problems and factors affecting the shape of voltammograms recorded in brain tissue. Thus, reference is made to depletion of the ECF around the electrode by the scan leading to thin-layer behavior, tortuous diffusion in the tissue, heterogeneous distribution of target analytes in various compartments in the brain, overlap of different signals, and perturbation of the tissue by the implanted electrodes. Quite inspirationally, the authors also suggested a number of strategies to address some of these problems, such as the use of permselective membranes and immobilized enzymes, tactics that proved successful many years later.

2.1 Original Interference Problem

Despite the aim to detect neuromediator DA in this first study using VIV, Adams et al.[4] concluded correctly that the voltammograms recorded were due to a number of different compounds present in the ECF, oxidizing at similar potentials at the CPE, and that AA, not DA, was the principal contributor. The initial success, however, was soon undermined by a widespread assumption that the concentration of AA in brain ECF was essentially constant. This led to conclusions that changes observed in the voltammetric signal were due to other species such as DA and serotonin. It was not until the early 1980s that this assumption had to be abandoned when two independent studies[61,62] showed that systemic administration of a commonly used DA-releasing drug, amphetamine, also caused changes in the concentration of AA in brain ECF.

Having established that ECF brain AA concentrations respond to a wide variety of stimuli, including pharmacological, electrical and behavioral (see section 7.3 and review by O'Neill[63] for details), efforts began in earnest to resolve the different contributions present in brain voltammograms. Two basic approaches have been used: eliminating undesired signals (e.g. those of AA and neuromediator metabolites), principally by means of chemical modification of the electrode surface; and separating the voltammogram components into identifiable signals using a combination of surface modification, applied potential waveforms, and mathematical analysis of the current; examples of validation of specific sensors are described in section 4. These developments in VIV over the past two decades have led to its establishment as a valuable tool in many aspects of neuroscience research, including anatomical, pharmacological and behavioral studies.[8–22]

2.2 Analytical Problems of Voltammetry In Vivo

It should be stressed that the application of VIV involves far more than a straightforward transfer of expertise gained from the practice of voltammetry in vitro. In addition to being anatomically complicated and containing a wide range of electroactive substances, brain tissue presents a complex chemical environment that includes surfactants (lipids), electrodes poisons (proteins), electrocatalysts such as glutathione and AA, and a tissue matrix that both restricts mass transport to the electrode surface and reacts physiologically to the presence of the probe.[64] Thus, the behavior of a particular electrode and voltammetric technique in vitro might well be different to its behavior in vivo, because both modification of the electrode surface by components of the tissue,[65] and the differences in mass transport,[66–69] could lead to different responses (sensitivity, selectivity, stability, etc.) in the two environments.[70] Therefore, detailed characterization in vitro of parameters such as detection limits is of little use, and may be misleading in an application like VIV where the major issue is that of selectivity rather than sensitivity.

Indeed, the complexity of the system is such that some forms of VIV may have been characterized in error in the literature, and subtle differences in methodologies might lead to an inability to reproduce reported characteristics. It is advisable, therefore, that each laboratory

Table 1 Electrochemical and concentration data for the most common species detected voltametrically or amperometrically in brain ECF

Compound	Charge[a]	$E_{1/2}$ versus SCE[b]	ECF concentration[c]
AA	−	−100 to 400	100–500 μM
DA	+	100 to 250	1–50 nM
NA	+	100 to 250	1–50 nM
DOPAC	−	100 to 250	1–20 μM
5HT	+	300 to 400	1–10 nM
5HIAA	−	300 to 400	1–10 μM
UA	−	300 to 400	1–50 μM
3MT	+	500 to 600	<1 μM
HVA	−	500 to 600	1–10 μM
NO	0	−900 and +1000	10–100 nM
O_2	0	−400 to −600	40–80 μM
Glucose	0	enzyme catalyzed	0.5–2 mM
Lactate	−	enzyme catalyzed	0.5–2 mM
Glutamate	−	enzyme catalyzed	1–10 μM

[a] At physiological pH (7.4).
[b] Very rough guide; depends on electrode material and its state of activation.
[c] Very rough guide of nonstimulated mean values; depends on animal species, brain region and level of anesthesia; values most appropriate for rat striatum.
5HT, 5-hydroxytryptamine; 5HIAA, 5-hydroxyindoleacetic acid; NO, nitric oxide; SCE, saturated calomel electrode; UA, uric acid.

considering adopting VIV, having decided on the form that promises to be suitable for the neurochemical study in question, should either characterize the technique in detail in vivo or, at least, test it with a series of well-documented pharmacological challenges. For example, a sensor designed to detect striatal DA must be relatively unresponsive to other ECF species, especially AA and catechol metabolites (see Table 1). The first selectivity test should be against AA, because of its several thousand-fold excess over DA in brain ECF. This selectivity can be ascertained by, for example, systemic administration of a DA-receptor agonist, such as apomorphine, that causes decreases in striatal ECF DA levels while increasing those of striatal AA.[71] If the sensor passes this test, then systemic administration of a MAO inhibitor (see Figure 1 legend), such as pargyline or amphetamine, can be used to test selectivity for DA versus catechol metabolites by increasing the concentration of DA in the ECF and decreasing catechol metabolite levels.[61,72] Note, however, that this second test can only be interpreted correctly if the sensor has passed the first selectivity test, because drugs like pargyline and amphetamine also cause increases in striatal AA concentration in the ECF.[61,62,72] The route of drug administration in these tests is also important. Particularly, systemic injection of drugs (intraperitoneally, subcutaneously, etc.) can have very different effects compared with their local perfusion directly into the brain.[62,73] An additional precaution is that of timescale. A technique demonstrated to detect DA selectively on a fast timescale (seconds) cannot be assumed to do so on slower timescales (minutes) because changes in neuromediator concentrations tend to be much faster than those of their metabolites.[74] These points will be developed below.

3 SUBSTANCES TARGETED WITH VOLTAMMETRY IN VIVO

One of the main problems associated with producing identifiable voltammetric signals directly in vivo is that the ECF contains a wide variety of electroactive species that oxidize at similar potentials on many electrode materials. Moreover, baseline ECF concentrations of the main target molecules – neuromediators – are orders of magnitude lower (often nanomolar) than those of other endogenous electroactive species (see Table 1). Factors affecting determination of ECF concentrations are discussed elsewhere.[21]

Electroactive compounds in the ECF include: small, highly permeable, species such as dioxygen and NO; AA, that has a range of general antioxidant and specific neurochemical functions; neuromediator catecholamines such as DA and NE, as well as their metabolites DOPAC, 3MT and HVA; the 5-hydroxyindole neuromediator, 5HT, (serotonin) and its metabolite 5HIAA; amino acids and the purine metabolite UA. Electroinactive or badly behaved electroactive species, such as glutamate, glucose and lactate, can now be targeted in vivo using biosensors, especially enzyme-modified electrodes.[16,21]

The range of oxidation potentials for ECF analytes (against SCE) and estimated baseline extracellular concentrations are given in Table 1. In general, the higher the potential needed to detect the analyte of interest, the greater the risk of interference, so that at the potential needed to detect NO, say, interference by amino acids such as tyrosine and tryptophan, and molecules containing these moieties, needs to be considered. One feature that has helped separate voltammetric signals for AA and neuromediator metabolites from those of the neuromediators themselves is their ionic form at physiological pH. Because the former are anionic and the latter cationic at pH 7.4, modification of the electrode surface by anion repelling agents can increase the selectivity of the sensors for the parent neuromediator. The problems of VIV signal resolution have also been tackled with some success by choice of working electrode material, fine-tuning the shape of the applied potential, modification of the surface properties of the working electrode, and mathematical treatment of the recorded current.[21]

4 ELECTRODE TYPES AND PREPARATION

4.1 Auxiliary and Reference Electrodes

A wide variety of auxiliary and reference electrodes have been used for VIV. The main criterion is that the electrodes are small to minimize tissue damage upon implantation into the tissue, the most common placement site. Minimal disruption can also be achieved by locating these electrodes between the brain and skull (see section 6), although care is needed to ensure that these stay connected to a conducting medium. Using metal items already in place for other functions, such as a stainless steel guide cannula[4] or a supporting screw[62] as the auxiliary electrode, will also reduce damage. Recordings made in a metal frame or cage do not normally require explicit earthing of the animal, although this may be needed for recordings where plastic containers are used.

The reference electrode needs to be chosen with greater care because of the need for chemical reversibility at its surface. A common form is AgCl-coated Ag wire housed in a pulled micropipette containing a gelatinous solution of NaCl. This has the advantage that very small tip sizes are achievable, but disadvantages include the formation of bubbles in the tip and gradual leaching of the NaCl over extended periods in the tissue. In part to overcome these

two problems, the implantation of AgCl-coated Ag wire directly into the tissue is quite common. The consequent problem here appears to be the dissolution of the AgCl in the tissue, although a Nafion®-coating has been reported to extend the life of this type of implanted reference electrode.[75] Rather surprisingly, the most stable form of reference electrode for VIV is a bare Ag wire that can provide stable signals over weeks of recording;[76] the exact nature of the chemical reversibility is not clear, but may involve either Ag/Ag^+ or Ag/AgCl couples. Sample VIV data shown in sections 4.5 and 7.2 were recorded with this type of reference electrode.

The following description gives details of one approach to preparing auxiliary and reference electrodes for VIV, using Teflon®-coated silver wire (200-μm bare diameter). The Teflon® coating is removed from both ends to expose a 2–3 mm length of bare wire. An earth wire is also prepared using the same procedure. A gold connector is soldered to one end and the other end is gently manipulated, using a small tweezers, to produce a characteristic hook shape in the case of the auxiliary and earth. This enables them to be placed between the skull and dura (see section 6). The end of the reference is manipulated to produce a perpendicular kink about 1–2 mm from the end of the wire. This prevents the wire from penetrating beyond the cortical tissue. Once prepared, all electrodes (including the working electrodes) are inserted into a plastic socket via the gold pins (see section 6) and secured with dental acrylate. The design of various types of working electrodes for VIV is described below.

4.2 Platinum

Coated Pt wire normally forms the basis of Pt working electrodes used for VIV.[21] Either disk (100–300 μm diameter) or cylinder (~1 mm length) electrodes can be generated simply by removal of selected segments of the insulation, usually Teflon®. Diameters as low as 15 to 25 μm have been reported,[77,78] but in these cases additional mechanical support, usually a glass capillary, is required. To protect from fouling, Pt surfaces can be coated with a variety of membranes such as precast or cast cellulose acetate,[79] recast or electrodeposited perfluorosulfonated ionomers,[80] or electropolymerized films that can be as thin as 10 nm.[81,82]

4.3 Carbon Epoxy

A mixture of carbon paste (see section 4.5) and epoxy resin, force-filled into a glass capillary (~50–200 μm in diameter) and heat cured, provides a robust carbon epoxy electrode (CEE) for implantation in brain tissue. In the unmodified form, CEEs show little discrimination between AA, metabolites and neuromediators and early reports using CEEs with chronoamperometry for voltammetric detection of DA and 5HT in vivo may have been affected by changes in AA concentration.[83,84] Electrochemically pretreated, however, by the application of a potential high enough to electrolyze water, AA and catechol signals have been well-separated with pulse voltammetries, apparently by the cleaning,[85] roughening and oxidation of the electrode surface.[86] Anion-repelling polymer membranes (especially Nafion®) on the electrode surface that inhibit AA and metabolite access to the electrode[87,88] have also been used to improve the selectivity of CEEs for DA.[23] Although not widely used now for VIV, CEEs continue to be developed as forms of rigid carbon polymeric biocomposites.[89]

4.4 Carbon Fiber

Prepared by sealing single[90,91] or multiple[92,93] strands of carbon fiber into glass,[94] silica[95] or Teflon®[96] capillaries, carbon fiber electrodes (CFEs) are the most commonly used electrodes for VIV, principally because of their small diameter (5–50 μm) and easily varied length (0 to 500 μm). Sealing the fiber into the support is the most critical step in the preparation of CFEs and this has been achieved by a variety of methods, including the use of epoxy resin and, more recently, by the electrophoresis of paint.[55]

Armstrong-James and Millar[90] and Ponchon et al.[97] introduced CFEs in 1979, and their usefulness increased dramatically when it was demonstrated that resolution of AA (as peak 1) and catechol (as peak 2) could be achieved with pulse[61,98,99] and staircase[100] voltammetries following modification of the surface by application of either a triangular potential[91] or constant current[70,101] before use. Electrochemical and spectroscopic studies indicate that these treatments are likely to improve the performance of CFEs by increasing electron transfer rates[102–104] and adsorption of analyte[105,106] on the surface. Treatments have been explained in terms of a combination of electrode cleaning, increased area of active graphite edges from cracking, and oxidation of the surface.[104,107,108] It appears that the important parameter of the treatment is not the applied potential, but the current density passed, and the most reproducible results have been obtained under galvanostatic rather than potentiostatic conditions.[101,109] This factor may explain the difficulty some laboratories have experienced in producing good electrical treatment of CFEs, especially if the potentiostat used was not designed for the relatively high currents associated with this process.[110] Although electrochemical activation still remains the most common form of CFE treatment, laser,[102] heat[104,111] and chemical[112,113] treatments have also been successfully used. Other important factors shaping the effectiveness

of these modifications include the pH[114,115] and the buffer ions[102,116] of the treatment solution, phosphate buffer pH 7.4 being the most common medium.

Although electrical and other treatments of CFEs enable separation of AA and catechol signals, the problem of separating the two catechols (small DA and large DOPAC currents in DA-rich, NE-poor brain regions, such as dorsal striatum) is more difficult, but has been achieved using a number of different strategies. Nafion® coating of either treated or untreated fiber disks or cylinders repels AA and DOPAC, and allows both K^+-evoked[117,118] and electrically stimulated[119,120] DA release to be observed, although AA and DOPAC may interfere with the detection of slower drug-induced changes in DA.[74]

In addition to AA, DA and DOPAC detection, CFEs have also been used to monitor the 5-hydroxyindoles, 5HT and 5HIAA. By adjusting the electrical treatment parameters, the relative sensitivities and resolution of CFEs can be varied and a voltammetric peak (peak 3), containing contributions from 5HT, 5HIAA and UA, can be recorded in the ECF.[121–125] In acute experiments shortly after implantation of the CFE, the main contributors to peak 3 are 5HIAA (about 70%) and UA (about 30%) in both brain striatum[122] and spinal cord.[126] Peak 3 recorded 7 days following implantation, however, appears to be due entirely to 5HIAA.[127] See section 7.4 for further discussion of this phenomenon.

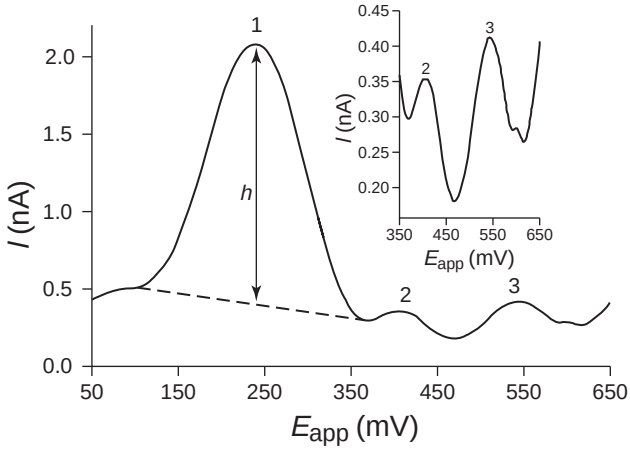

Figure 2 Staircase voltammogram (background subtracted) recorded at $5\,mV\,s^{-1}$ with a 160-μm diameter CPE in the striatum of a freely-moving rat several weeks after implantation. Peak 1 is essentially due to AA, peak 2 to 5HIAA, and peak 3 to the DA metabolite HVA. Increases in the extracellular concentration of catechols, DA and DOPAC, increase the current between peaks 1 and 2, considerably changing the potential of these peaks for large increases. Changes in the height, *h*, of peak 1, when there is no significant change in its peak potential, reflect variations in the concentration of AA in the ECF. When 320-μm diameter CPEs are used under the same conditions peak 2 is due to about 90% UA and 10% 5HIAA. Inset: Close-up of peaks 2 and 3 recorded with a 160-μm diameter CPE showing that, although these signals are of the order of 10^{-10} amps, they are measurable. Changes in the height of peak 3 can be used to monitor HVA in the ECF; a time course at 12-min intervals for this signal over a 3-day period is shown in Figure 4.

4.5 Carbon Paste

Carbon paste, prepared by thoroughly mixing carbon powder with either Nujol[4] or silicone oil[62] can be easily packed into cavities made by pulling the insulation over fine metal wires, producing disk electrodes of diameters as small as 160 μm, although diameters of about 300 μm have been generally used for VIV. A carbon/Nujol paste electrode with CV was used in one of the first reports of voltammetry in brain tissue with the aim of studying DA,[4] and a single peak was observed in rat striatum, attributed mainly to the oxidation of AA. Decreasing the scan rate to $10\,mV\,s^{-1}$ and semidifferentiation of the current later revealed three rather broad waves on the striatal voltammogram.[128] Further developments of the technique by the use of computer-controlled equipment,[129] silicone oil paste,[62] an additional reduction of the sweep rate to $5\,mV\,s^{-1}$, abandoning semidifferentiation, and subtraction of the background current recorded in situ, yielded CPE voltammograms consisting of three well-separated peaks in striatal tissue[130] (see Figure 2).

Peak-shaped signals are obtained by this non-differential method due to both thin-layer behavior in the compartment around the implanted electrode[66,131,132] and adsorption of the analytes.[65,133] Because of the lack of clear resolution between AA and catechols by CPEs, a wide variety of techniques have been used to characterize the CPE signals recorded in rat striatum in vivo: theoretical analysis;[132] electrochemical studies in vitro;[62,129] microinfusion of candidate substances[129,134] and metabolizing enzymes[133,135–137] beside electrodes implanted in brain tissue; lesions to striatal afferent pathways;[71,135,138] pharmacological perturbations using drugs with well-characterized actions;[71,134,136,139] and comparison with other in vivo techniques, such as microdialysis.[133,140]

These experiments, reviewed recently,[64] indicate that peak 1 (see Figure 2) is due mainly to AA; moreover changes in the height of peak 1 result from variations in the extracellular concentration of AA, provided there is no significant shift in the relative potentials of peaks 1 and 2. When such potential shifts occur they are not due to electrocatalysis but to the algebraic combination of the AA peak (at the normal potential of peak 1) and the

catechol peak some 60 mV higher, and difference voltammograms can be used to analyze the changes in terms of fluctuations in the catechol concentration.[129] Under most circumstances (some exceptions being stroke[141] and death[142,143]) these latter changes are mainly due to DOPAC, because the level of DA in striatal ECF has been estimated by a range of techniques to be less than 50 nM.[72,144,145] The main component of peak 2 is UA with a minor contribution from 5HIAA when ~300 μm CPEs are used,[133] whereas peak 2 recorded with 160 μm CPEs is due to 5HIAA almost exclusively.[139] Changes in the height of peak 3 are directly proportional to changes in the extracellular concentration of methylated metabolites of DA. Normally this is HVA,[71] but when MAO is inhibited pharmacologically, 3MT may contribute[140] (see Figure 1).

The main advantage of CPEs is their stability over several months of recording in vivo.[8] It appears that lipids present in the tissue remove pasting oil from the surface[19] and reverse poisoning caused by proteins.[70] This factor may also undermine the use of CPEs that have been chemically modified with lipophiles, such as stearic acid, for the detection of DA[146,147] because such species can be removed from the electrode surface following contact with brain tissue,[148] although this point is still a subject for debate.[17] The principal disadvantage of CPEs is their comparative large size, and neurochemical studies have been limited to relatively large brain regions such as striatum,[149–151] nucleus accumbens,[151–153] cortex,[154,155] pallidum[156] and hippocampus.[151,155,157] None of these brain areas can be considered an homogeneous array of cells – even the striatum consists of an array of substructures[158] – and so the development of an electrode with the dimensions and sensitivity of CFEs and the stability of CPEs is an on-going objective for VIV.

4.6 Biosensors

A biosensor can be defined as a device that involves the immobilization of a sensitive and selective biological element on, or within close proximity of, an analytical detector. The range of biological components that have been used in the development of biosensors is extensive; typical examples include enzymes, plant and animal tissues, microbes and antibodies. The type of analytical detector employed is also wide ranging (e.g. electrodes, optical fibers and crystals) and is generally used to classify the biosensor. The fundamental principles, designs and wide ranging applications of biosensors have been described in detail in comprehensive reviews in the literature.[16,159,160]

The primary reason for developing biosensors for applications in vivo is the ability to measure biologically important compounds that are either electroinactive or only poorly electroactive at analytically useful potentials. Such compounds include glucose, lactate and glutamate (see Table 1). Enzyme-modified electrodes, particularly amperometric devices, are the most thoroughly investigated sensors in the biosensor field. Their origin dates back to the first reported biosensor (for glucose) by Clark and Lyons in 1962.[161] Prolific research on the development and applications of enzyme-modified electrodes over the intervening three decades has yielded sensors for a large variety of analytes, including glucose, ethanol, cholesterol, bile acids, choline, purines and amino acids.[159,160] However, despite the significant advantage of continuous real-time monitoring afforded by these electrochemical devices, only a small percentage have actually been used in vivo for chemical analysis in the living brain. This is primarily due to the many difficulties associated with performing direct measurements in such complex environment as brain ECF (see section 2.2). Important considerations include the immunological response to the implanted sensor, the presence of electrode poisons and endogenous interferences, restriction of mass-transport by brain tissue, effects of depletion, and the availability of suitable enzymes for the specific analytes of interest. The sensors that have been successfully used for VIV have all been amperometric flavoprotein oxidase-modified electrodes for either glucose, lactate, glutamate or choline.[21]

The immobilization of the enzyme component onto the electrode surface has been achieved using various strategies. The most common methods include: direct physical adsorption; covalent bonding (e.g. via carbodiimide) and/or cross-linking (e.g. via glutaraldehyde); entrapment behind pre-cast (e.g. cellulose acetate), cast (e.g. polyurethane), and electrodeposited (e.g. Nafion®) membranes; and more recently, entrapment within electrosynthesized polymeric matrices (e.g. polypyrrole and polyphenylenediamine).[16,159,160]

The classic and most common enzyme-modified electrode is the glucose biosensor incorporating the oxidoreductase enzyme, glucose oxidase (GOx, isolated from *Aspergillus niger*, E.C. 1.1.3.4). This type of sensor has become the model system primarily due to the importance of glucose monitoring in the disease diabetes mellitus, and the fact that glucose determination in various body fluids, such as blood, plasma and urine, remains one of the most common analyses performed in clinical laboratories. Additionally, GOx has good substrate specificity, a high turnover rate, and excellent stability that makes it ideally suited for use in a biosensor.[162] GOx is a flavin-containing enzyme that catalyses the oxidation of D-glucose in the presence of oxygen, producing hydrogen peroxide:

$$\beta\text{-D-glucose} + \text{GOx/FAD} \longrightarrow \text{D-glucono-}\delta\text{-lactone} + \text{GOx/FADH}_2$$

$$\text{GOx/FADH}_2 + \text{O}_2 \longrightarrow \text{GOx/FAD} + \text{H}_2\text{O}_2$$

where FAD is the oxidized form of the prosthetic group, flavin adenine dinucleotide. The development of amperometric enzyme electrodes has focused on the measurement of the glucose signal generated by the reaction of the reduced enzyme ($GOx/FADH_2$), and can be categorized broadly in terms of first, second and third generation devices.

Classic or 'first generation' enzyme-modified electrodes involved monitoring either the consumption of oxygen,[163] or the formation of hydrogen peroxide.[164] These devices suffered from a number of drawbacks including fluctuations in response due to low and/or variable oxygen tension, and interference from endogenous electroactive species due to the large applied overpotential (typically +700 mV versus SCE) required for hydrogen peroxide oxidation. Despite this, hydrogen peroxide detecting sensors still remain the most common type of enzyme-modified electrode, and over the years a number of attempts have been made to overcome these problems. In particular, nonphysiological electron transfer mediators, immobilized on the electrode surface or within the enzyme layer, have been used to replace oxygen, the natural or physiological enzyme mediator. Such artificial mediators must exhibit several important characteristics: reversible electron transfer kinetics at the electrode surface, and rapid homogeneous kinetics for the enzymatic reaction in solution; possess a low overpotential for regeneration of the oxidized form of the mediator; and be stable in both oxidized and reduced forms. Many such redox mediators have been investigated and found to meet these criteria. Typical examples include ferrocinium ions, quinones, ferrocyanide, ruthenium compounds and tetrathiafulvalene.[21] However, although these 'second generation' devices have the advantage of a low operating potential, and elimination of oxygen from the reaction scheme, they do suffer from a number of problems. These include leaching of mediator from the electrode surface or enzyme layer,[165] toxicity in biological tissues,[166] and electrochemical interference, e.g. oxidized ferrocenes can be reduced by AA.[162] Additionally, the complete insensitivity to oxygen tension sometimes claimed for this type of sensor has been questioned for certain mediators.[167] More recently, another form of mediated system that has found widespread usage is that of entrapping the enzyme in a redox polymer (e.g. poly(ethylene glycol) diglycidyl ether) that can serve as both the mediator and enzyme immobilization matrix.[168,169]

Eliminating the mediator from the reaction scheme and achieving the simplest approach of direct electron transfer from the reduced enzyme to an electrode has proved extremely difficult. This is generally accepted to be because the distance between the FAD active site and the electrode is such that direct electron transfer to the electrode takes place slowly or not at all. Consequently, very few mediatorless or 'third generation' enzyme electrodes have as yet been devised. The first report of such a device was by Albery et al.[170] using one-dimensional organic conducting salt electrodes, e.g. tetrathiafulvalenium tetracyanoquinodimethanide (TTF^+TCNQ^-), to eliminate the mediator from the reaction scheme. However, the mechanism of electron transfer at these organic salt electrodes is now generally accepted as being a mediated process involving species such as $TCNQ^0$.[171] Since then, successful direct electrical communication between GOx and an electrode surface has been claimed by several groups using a variety of techniques. These include direct electron transfer between covalently bound GOx and a cyanuric chloride modified graphite electrode,[172] between GOx immobilized on glassy carbon modified by aminophenylboronic acid,[173] and between GOx immobilized via adsorption and a variety of cross-linking reagents onto platinized carbon paper electrodes.[174] Urea has been used to modify the enzyme structure by opening the active site, thus making it more accessible to a ferrocyl acetamide-modified electrode, composed of arrays of electron-transfer relays.[175] This type of chemical modification of the enzyme has led to the development of 'electrically wired' enzymes.[176] A hydrophobic interaction between GOx and TTF^0 molecules has also been reported with the resultant opening of a direct path of electrical communication between a Pt electrode and the flavin redox center.[177]

Finally, before a biosensor can be used reliably in vivo it is necessary to perform a detailed characterization in vitro of the device. Such a characterization typically includes studies of sensitivity, linear calibration range, response time, interference properties, and stability.[178,179] However, it is also very important that a detailed in-vivo characterization be performed in the target environment before the sensor is used in direct physiological investigations.[180] This is to ensure that the properties obtained in the artificial in-vitro environment are maintained in the more complex living biological milieu.

5 VOLTAMMETRIC TECHNIQUES USED IN VIVO

In VIV one is usually interested in how the concentration of a particular neurochemical is changing with time

on a scale from milliseconds to days. The choice of technique will therefore be restricted by the time resolution needed, in addition to other factors such as how well the working electrode can discriminate between the different electroactive species present in the ECF. Since the vast majority of substances of interest in VIV studies undergo oxidation, rather than reduction reactions, the descriptions that follow will be for oxidative voltammetry. The forms of voltammetry most often used for neurochemical analysis in vivo have been reviewed.[8–22]

Different forms of voltammetry can be categorized on the basis of the time resolution achieved for detecting changes in the concentration of species in the ECF. Fast techniques with resolutions of the order of seconds or less are chronoamperometry, fast cyclic voltammetry (FCV), differential pulse amperometry (DPA) and constant potential amperometry (CPA), and these methods are generally used to detect stimulated changes in neuromediator overflow. Slow techniques with time resolutions of the order of several minutes are linear sweep voltammetry (LSV), staircase voltammetry (SCV), differential pulse voltammetry (DPV) and differential normal pulse voltammetry (DNPV) at low scan rates.[13,21] The latter are usually used in studies of neuromediator metabolites as well as AA and UA. The minimum interval between scans is limited not only by the sweep time but also by the time taken for the concentration of the electrolyzed species in the compartment around the implanted electrode tip to return to its unperturbed value by means of tortuous diffusion through the tissue matrix (see Figure 2).[69]

5.1 Chronoamperometry

This is one of the simplest forms used for VIV and was one of the first.[83] Initially, the potential is held at a sufficiently low level where no oxidation occurs. A square pulse is then applied for between typically 100 ms and 1 s to a value corresponding to a high overpotential for the particular analyte of interest, and the current sampled close to the end of the pulse (usually integrated over a short time interval) where the ratio of faradaic to charging current is large. Because the initial currents associated with chronoamperometry are among the largest produced by VIV, it is interesting to note that simultaneous recording of nerve cell firing during this,[181] and other relatively high current forms of VIV,[182,183] indicate that voltammetry in nervous tissue does not appreciably interfere with the normal electrical properties of neurons.

Although chronoamperometry has an advantage for VIV of good temporal resolution (up to 200 ms[184]), a major limitation in its simplest form is the inability to differentiate between compounds with similar oxidation potentials. Thus, because AA tends to oxidize at least to some extent at lower potentials than monoamine neuromediators at unmodified electrodes, early in-vivo chronoamperometric data may have been contaminated with contributions from AA. Two later innovations have revived interest in chronoamperometry: the use of anion-repelling polymer membranes (especially Nafion®) on the electrode surface that inhibit AA and metabolite access to the electrode; and the development of reverse-pulse chronoamperometry that allows discrimination between different species oxidized based on the ratio of their oxidation/reduction response.[21,24,185–187]

5.2 Linear Ramps

The general feature of this group of voltammetric techniques is the application of an essentially linearly increasing potential from an initial value where no oxidation occurs to a final value at a fixed scan rate (5 to 100 $mV\,s^{-1}$ for VIV), allowing the sequential oxidation of a range of analytes. The simplest example is LSV where a truly linear (from an analog device) potential is used. Because many VIV studies now use computer-controlled digital equipment,[72,100,129] the potential is incremented in discrete steps so that SCV is obtained. In studies with CFEs, where relatively large steps are involved (10 mV), a differential form of analysis is used where the current measured on a given step is subtracted from the next value to give a peak-shaped voltammogram with cylindrical microelectrodes.[100,188]

If the potential is scanned back to the initial value after reaching the maximum point, the technique is CV. This approach is used very rarely at slow scan rates for VIV because most of the oxidation reactions are chemically irreversible in vivo.[4] However, FCV at sweeps of about 300 $V\,s^{-1}$, where the oxidized form is reduced before the following chemical reaction can occur appreciably, is now widely used for VIV because the different components in the ECF can be distinguished on the basis of the shape of their oxidation–reduction currents.[18,22,120,189–195]

The main advantage of the slow ramp techniques is that a number of separate signals associated with different neurochemicals of interest can be obtained in the same voltammogram using either CPEs (see Figure 2) or CFEs.[21] A consequent disadvantage is the time taken to sweep the potential (up to 2 min at 5 $mV\,s^{-1}$) as this restricts the time resolution of the concentration determination to a minimum of about 5 min for 5 $mV\,s^{-1}$[157] and 2 min for 100 $mV\,s^{-1}$[100] to allow for the slow recovery of ECF levels of analytes electrolysed from the compartment around the electrode tip. For FCV the situation is reversed because a good time resolution of 100 ms has been achieved,[196] but only

a single substance can be detected under the conditions of any given experiment.

5.3 Pulsed Techniques

There is a wide variety of pulsed voltammetric techniques available:[13,21] DPV, where a pulse of fixed amplitude is superimposed on a linearly increasing ramp, and the current, sampled just before the pulse, is subtracted from that measured close to the end of the pulse; normal pulse voltammetry (NPV) in which a pulse of increasing magnitude is applied relative to a constant resting baseline potential, and the current sampled near the end of the pulse; DNPV, where a double pulse of fixed potential difference is applied with increasing amplitude from a constant baseline; and DPA, where a double pulse of fixed potential difference is applied at fixed amplitude and the difference in current at the two pulses plotted as a function of time.

The first pulsed voltammetric form for VIV with carbon electrodes was NPV.[97] DPV, however, has been widely used, although DNPV is also common and DPA has been used more frequently in recent reports.[197–199] The main advantage of the pulse approach is that the different parameters of the potential form – pulse amplitude, duration and separation – can be adjusted to provide maximum resolution of the target analyte based on differences in its kinetic and adsorption properties on the electrode.[99,200] In addition, capacitance currents can be minimized if the subtraction is carried out before the amplification stage. Square wave voltammetry that involves sequential measurement of oxidation and reduction currents[201] has not been used widely in VIV, but has an advantage that good peak resolution can be achieved at relatively fast scan rates of about $125\,mV\,s^{-1}$[202] compared with about $10\,mV\,s^{-1}$ for DPV and DNPV.

More recently, differential multipulse amperometry (DMPA) has been developed and applied to neurochemical studies.[21,203] Here potentials determined using DNPV are used to construct a voltage profile consisting of four short pulses (two pairs) repeated indefinitely. The pulse potentials were chosen so that the difference in amperometric response between the first two pulses was due to DA, whereas 5HT was detectable using the second pair. Sub-second detection of these two neuromediators appears possible using this approach, although the data available to date is limited.

5.4 Constant Potential Amperometry

In CPA, the current is monitored essentially continuously at a fixed potential where the redox reaction of the analyte of interest occurs, usually under diffusion-limiting conditions. This form has recently been used for VIV because of its high time resolution, but applications are limited to situations where the oxidation of only one analyte can be assumed with some certainty as in the detection of AA at low potentials,[137,151,204] and catecholamines following stimulation of specific pathways in vivo[22,205–207] and from isolated cells in vitro.[22,47,50,55,94] The specificity of biosensors makes them ideal devices for use with CPA, and this will be discussed separately below.

6 SURGICAL PROTOCOL FOR VOLTAMMETRY IN VIVO EXPERIMENTS

6.1 General Considerations

Before carrying out any surgical procedure it is important to think carefully about the facilities that are available and whether they are adequate for what is to be achieved. Implantation of microelectrodes for VIV requires good aseptic conditions with proper sterilization of equipment and instruments. Surgical procedures should be as atraumatic as possible. It is important to learn how to handle surgical instruments correctly and to avoid damaging tissues by grasping them with excessive force or with inappropriate instruments. Surgery on small animals such as rodents should be performed in a designated surgery room. A recovery area should be provided where animals can be taken after surgery for postoperative care and observation while recovering from the procedures and anesthetic.

Male Sprague–Dawley rats weighing between 200 and 300 g (mature young adults) are generally used in VIV experiments. The animals are usually housed in a temperature-, humidity- and light-controlled environment, with free access to food and water. Sample protocols for VIV experiments are described in the following sections.

6.2 Anesthesia

Successful anesthesia does not depend simply on the types of drugs, doses, and routes of administration used. A good standard of animal care must also be maintained, both pre- and postoperatively, including reduction in stress and provision of pain control. All these factors must be taken into account when designing an anesthetic regime. A typical regime used to induce anesthesia for implanting microelectrodes is as follows.[208] A combination of the opiate analgesic fentanyl ($0.32\,mg\,mL^{-1}$)/fluanisone (Hypnorm, $0.1\,mg\,mL^{-1}$) and the benzodiazepine midazolam (Hypnovel, $5\,mg\,mL^{-1}$), mixed 1 : 1 : 2 with sterile water, is injected intraperitoneally (i.p.) at a volume of

3.3 mL kg^{-1}. This results in surgical anesthesia within 4–5 min, that is maintained for the 45 min of the surgical procedure. Both corneal and hind limb withdrawal reflexes are tested to confirm the level of surgical anesthesia. In order to complete the cementing process at the end of surgery (see section 6.3) 1–3 mg kg^{-1} of midazolam is administered intramuscularly. On completion of the surgery and cementing, an injection of naloxone (0.1 mg kg^{-1}, i.p.) is given to reverse the opiate effect and prevent respiratory depression. Post-operative analgesia is provided in the form of a single injection (0.1 mg kg^{-1}, subcutaneously) of Vetergesic (buprenorphine) given immediately following the surgery.

6.3 Surgery and Recovery

Once surgical anesthesia is established, the animal is placed in a stereotaxic frame. Body temperature is maintained at 37 °C throughout surgery with an isothermal heating pad. A midline incision 1–2 cm in length is made to reveal characteristic skull markings, bregma and lambda, which are used as coordinate reference points. The underlying muscle and fascia are scraped away and the skull levelled between bregma and lambda. A schematic of a typical surgical and implanting process is shown in Figure 3(b). Artery clamps are used to hold the incision open. With the aid of a dissecting microscope, small burr holes are made in the skull to enable implantation of small support screws and the recording electrodes. The latter are positioned in the target brain area according to coordinates obtained from a stereotaxic atlas.[209] For example, for rat striatum (left and right) that is the most widely studied region for VIV, the coordinates are: anterior/posterior, +1.0 mm from bregma; mediolateral, ±2.5 mm from bregma; and dorsoventral, −5.0 mm from the dura. The reference electrode (8T Ag wire, 200-μm bare diameter) is placed in the cortex, the auxiliary electrode (8T Ag wire) between the skull and dura, and an earth wire (8T Ag wire, applicable with plastic animal bowls) attached to one

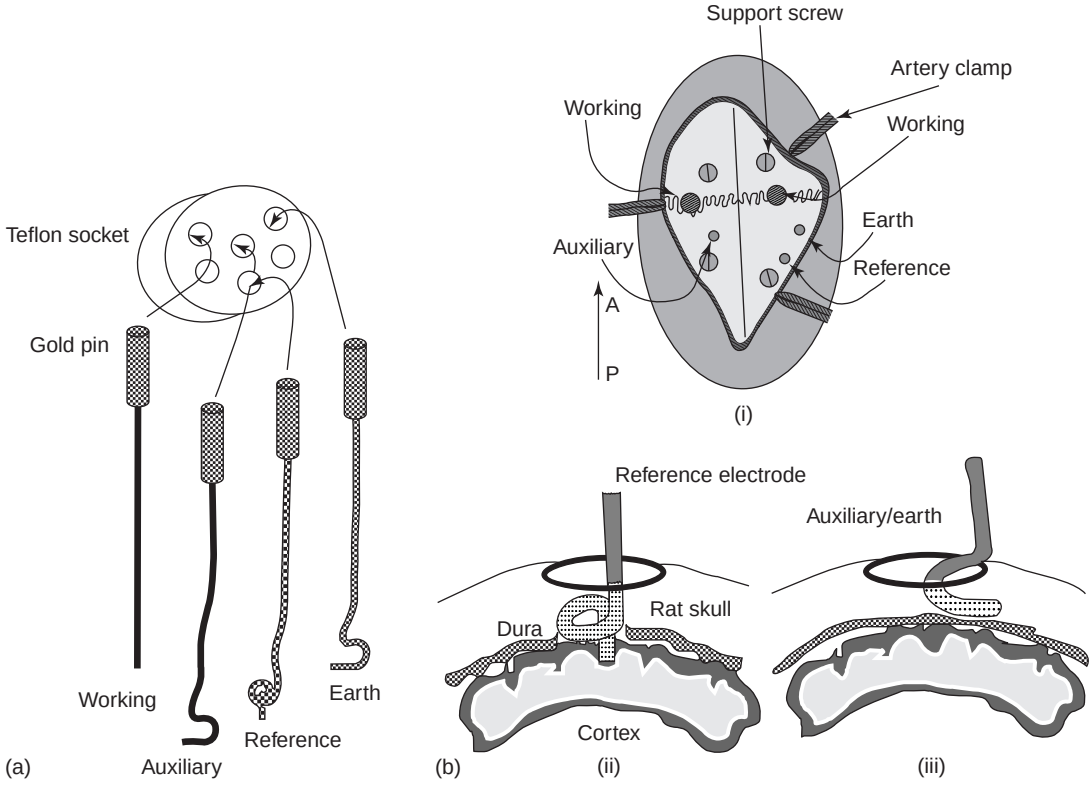

Figure 3 (a) Schematic representation of the electrode assembly suitable for VIV. A gold connector is soldered to one end of each electrode. Each electrode is inserted into a six-pin Teflon® socket via the gold pins. This socket, which is secured to the skull with dental acrylate after implantation, can then be connected through a flexible screened six-core cable, mounted through a swivel (allowing free movement) above the animal's head, to the electrochemical instrumentation. The design of the various types of electrodes is described in section 4. (b) Schematic representation of a surgical and implanting process: (i) A midline incision 1–2 cm in length is made over bregma (cross) and the underlying muscle and fascia scraped away. Artery clamps are used to hold the incision open. Small burr holes are made in the skull to enable implantation of support screws and the recording electrodes. (ii) Positioning of the reference electrode (200-μm bare Ag wire) in the cortex. (iii) Positioning of the auxiliary electrode or earth wire (200-μm bare Ag wire) between the skull and dura. The auxiliary may also be simply attached to one of the support screws.

of the support screws. All electrodes are secured with dental acrylate. The clamps are removed and a second application of acrylate serves to close the incision and to further stabilize the electrodes.

Once the dental acrylate has sufficiently hardened the animal is removed from the stereotaxic frame and returned to its home cage to recover, with its body temperature controlled while the analgesic effect of the opiate persists. At the end of this period, if signs of pain are observed, an additional injection of buprenorphine is given subcutaneously. The animal room is normally maintained on a 12/12 hour light/dark cycle.

6.4 Assessment of Animal Health

All animals should be monitored routinely following surgery for any signs of pain or distress. During the recovery period food and water should be freely available. A post-operative record should be kept that includes information on urine, faeces, food and water intake, drugs administered, and stress scoring. This practice ensures that only healthy, pain free animals become experimental subjects. Generally, a post-surgery period of 24 h is necessary and sufficient for recovery.

6.5 Experimental Conditions

There are two main categories of experimental conditions for VIV. The first, in terms of technical simplicity, is when recording takes place while the animal is still in the stereotaxic frame under the influence of anesthetic; this is termed an 'acute' preparation. The following factors must be taken into account when analysing data recorded under these conditions: concentrations of analytes may be effected by acute tissue damage; the BBB may not be intact; and the effects of the specific anesthetic agents on the concentration of target analytes.[21]

The second, and now more common, 'chronic' preparation involves allowing the animal to recover from anesthesia (see section 6.3). Implanted electrodes can be connected to the recording equipment through, for example, a six-pin Teflon® socket (see Figure 3a), and a flexible screened six core cable, that is mounted through a swivel above the rat's head. This arrangement allows relatively free movement of the animal. Possible complications with this preparation include gliotic reaction by the tissue and perturbations of the chemical environment around the implanted electrodes (see section 7.4).

6.6 Hazard Assessment

As with all procedures in a laboratory, it is now considered wise to undertake and document a safety hazard assessment of the materials and actions involved before carrying out the procedure, in which the hazards are identified, the risks assessed and control measures put in place to minimise the risks. Particular hazards for VIV experiments include: being bitten and contracting disease; developing allergic reactions in the respiratory system and skin; and being injured by scalpels and needles. Control measures associated with these hazards should include the careful handling of animals, appropriate inoculation of personnel, the wearing of face masks and gloves, and the careful use and disposal of sharps.

7 DETECTION OF SUBSTANCES IN BRAIN EXTRACELLULAR FLUID

The brain operates at a range of levels – molecular, cellular, regional and as a whole – and the choice of voltammetric technique depends to a large extent on what one wants to study. Microdialysis has also been developed to sample the ECF for neuroactive species, usually followed by separation with high-performance liquid chromatography (HPLC), and either electrochemical or spectroscopic detection (see section 8). An advantage of microdialysis is the wide range of compounds: neuromediators, metabolites, amino acids, drugs, etc. that can be collected in each sample and separated reliably using recent advances in HPLC technology.[210] The main disadvantages of dialysis are the relatively large probe size available (minimum $\sim 200\,\mu m$) and the poor time resolution, ranging from 2–10 min, although about 1 min has been achieved recently with capillary electrophoretic separation.[211] Although voltammetry has been limited until recently to electroactive analytes (a limitation that is being overcome with the development of enzyme-modified electrodes; see section 4.6), its advantages of small probe size and high temporal resolution, which have been achieved in many studies, allow investigations not possible with microdialysis.

The vast majority of neurochemical studies employing voltammetry have been carried out using rats, although there are examples of other species being investigated, such as snails,[212] turtles,[213] mice,[214] hamsters,[92] guinea-pigs,[215] cats,[207] sheep[216] and non-human primates.[217]

7.1 Electroactive Neuromediators

Although brain slices offer a convenient and effective way of manipulating individual neuronal pathways to study their pharmacology and physiology, the function and response of a given projection in vivo depends to a large extent on the myriad of connections to and from other pathways, many of which are severed in the slice.

Voltammetry has therefore been used in vivo to investigate neuromediator function in the context of the living brain, and in many cases its relationship to behavior.[21] However, because the baseline extracellular concentrations of electroactive monoamine neuromediators are of the order of nanomolar[72,144,145,218,219] (Table 1) due to efficient reuptake mechanisms[218] (Figure 1), many of the techniques available are not sensitive enough to detect these levels. A number of approaches have been used, therefore, to monitor artificially high levels of neuromediator overflow in the ECF, mainly in acute, anesthetized preparations.

High doses of the MAO inhibitor, pargyline, greatly reduce DOPAC levels in the ECF (see Figure 1) so that peak 2 recorded with CFEs and DPV is abolished.[61] The introduction of DNPV,[200] which shows a higher sensitivity for DA, revealed a small DA signal close to the potential of peak 2 in pargyline-treated rat striatum, and later the development of DPA[99] allowed the detection of DA with high time resolution in a number of brain regions under similar conditions.[214,220,221] In other approaches, chronoamperometry, FCV or CPA with CFEs has been used to monitor stimulated DA[22,195,222,223] and NE[7,205,224] release in a number of brain areas.[21] A successful model based on release and uptake parameters[225] has been used to understand the relationship between release and stimulation frequency for the dopaminergic nigrostriatal pathway, and how the effectiveness of drugs used clinically, including neuroleptics, can depend on firing rate.[144,226] In contrast, baseline (non-stimulated) ECF levels have been measured for DA as low as 2 nM in striatum,[219] and for 5HT[227] in frontal cortex (5 nM) and raphe nuclei (10 nM) using Nafion®-coated electrochemically-treated CFEs with DPV, and an important study with these electrodes has shown that extracellular concentrations of 5HT and 5HIAA are not necessarily correlated.[228] Although promising, these electrodes have not found widespread use for VIV, presumably due to reproducibility and stability problems.

Until recently, these approaches had a limitation that anesthesia was involved during the VIV recordings. A range of general anesthetics are known to affect the firing rate and responsiveness of neurons and the release, uptake and metabolism of a variety of neuromediator substances. This is not surprising because the widely distributed γ-aminobutyric acid (GABA)-receptor/chloride-channel complex may be a common mediator in some aspects of anesthetic action.[229] Indeed, specific examples of different (sometimes opposite) VIV results in the presence and absence of anesthetics have been reported.[149,157,230] The finding, therefore, that basal concentrations of DA and 5HT can be detected with mathematical deconvolution[231] of the DNPV signals recorded with electrochemically treated CFEs in unanesthetized, freely moving rats is an important advance in VIV.[21] More recently, replaceable electrode systems and other strategies have been applied to monitor behaviorally-related changes in the concentration of monoamine neuromediators with CFEs, including sexual behavior,[232] free-choice novelty,[233] and electrical self-stimulation of brain reward centers.[195]

7.2 Electroactive Neuromediator Metabolites

A principal aim of VIV is the generation of a signal as closely related as possible to neuromediator release in the synapse. An important question that has yet to be resolved is which species in the ECF is the closest reflection of synaptic release. The main candidate is the neuromediator itself overflowing into the ECF (see Figure 1). However, there are populations of neuromediator receptors outside the synapse and transmitter in the ECF may have a different function to that in the synapse. Moreover, there appear to be cases where environmental stimulation (e.g. presentation of food after a fast) causes activation of nigro-striatal DA cells without associated increases in striatal ECF DA levels.[234] The link between extracellular metabolite levels and release rate is also problematic. In the case of the monoaminergic systems, DOPAC and 5HIAA are more likely to reflect synthesis rate than release (see Figure 1). However, COMT (one of the enzymes required for HVA production) is located outside the nerve terminal[235] (see Figure 1). It has been suggested, therefore, that HVA may be a better index of DA release than DOPAC, and that HVA is a fair index in the absence of drugs, at least on a timescale of minutes.[13]

A principal advantage of some metabolite signals is that they are stable over several days of recording with the same electrode, and therefore allow investigation of the relationships between neurochemical events and observed behavior over extended periods.[76] A disadvantage is that changes in metabolite concentrations are much more sluggish than that of the corresponding neuromediator with lags of up to 30 min in the peak responses.[236] The main VIV techniques used to monitor neuromediator metabolite levels in the ECF are CFEs with either SCV, DPV or DNPV (DOPAC as peak 2 and 5HIAA as peak 3) and CPEs with SCV (HVA as peak 3; see Figure 2). The CFE 5HIAA signal has been used to detect 5HT metabolism in the suprachiasmatic nuclei, raphe nuclei and hypothalamus over the sleep/wake cycle, and as a result of stress.[21]

Diurnal variations in DA metabolism using the CPE HVA signal have been recorded in frontal cortex, nucleus accumbens and striatum and correlated with simultaneously monitored motor activity (Figure 4). Rats are nocturnal and show a high level of locomotion during the dark period. Electrochemical detection of HVA

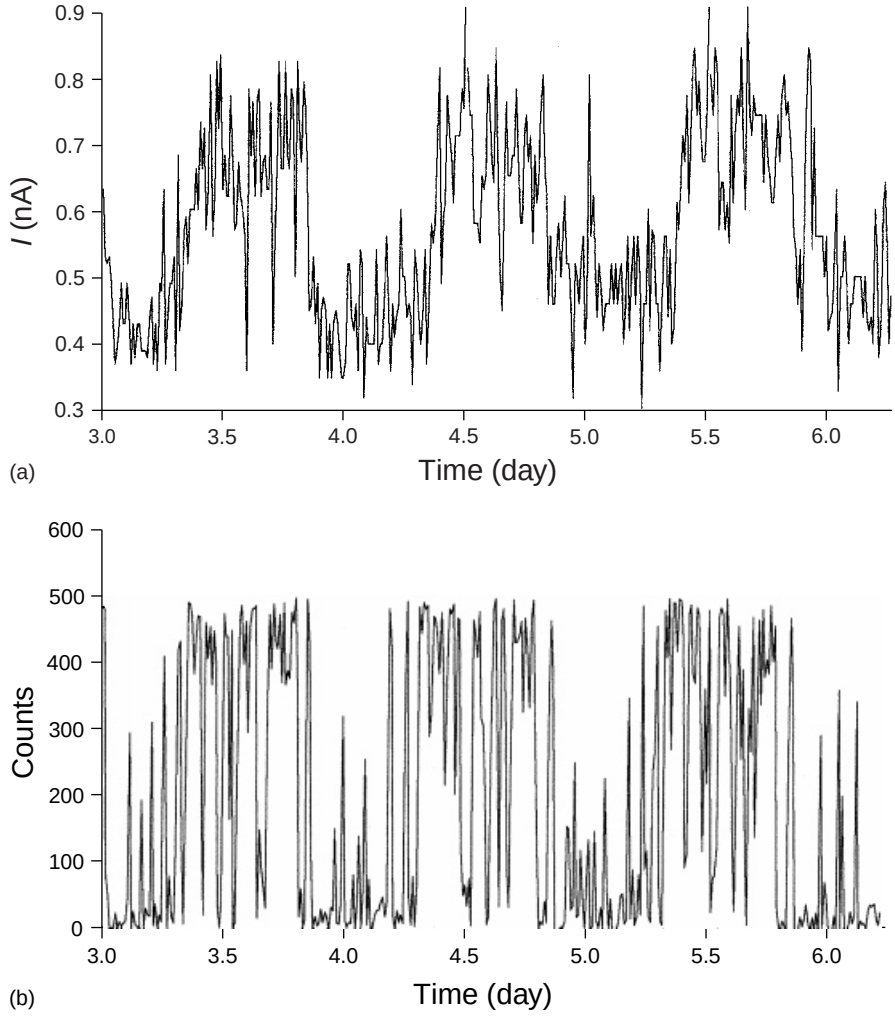

Figure 4 Time course over 3 days of changes in the methylated DA metabolite signal (peak 3, see Figure 2) recorded with a CPE in the striatum of a freely-moving rat at 12-min intervals (a) and simultaneously monitored motor activity (b). Each day's recording started at 12.00 h, with lights off from 20.00 h to 08.00 h. The period shown corresponds to days 3 to 6 of a longer recording period, and demonstrates the stability of CPEs in brain tissue. Similar data have been used to compare the activity of DA terminals in different brain regions during motor behavior, and to test observed correlations between these two parameters with physiological and pharmacological perturbations.

shows that, associated with this high activity phase, DA transmission in certain forebrain areas is also elevated. Correlation analysis indicates that DA transmission in the latter two brain structures, but not in frontal cortex, is closely involved in motor behavior.[154] This correlation can be disrupted for striatal, but not accumbens, HVA by valium-type anxiolytic drugs such as the benzodiazepine, flurazepam.[237] Changes in HVA associated with trained lever-pressing behavior have also been reported for striatum and nucleus accumbens.[238]

7.3 Ascorbic Acid

With its high concentration and ease of oxidation, AA is probably the simplest molecule to detect and monitor in brain ECF using VIV techniques. Microdialysis methodologies, however, are not as fortunate in that they appear to have difficulty recovering and detecting changes in brain AA. This may be due to the presence in brain tissue of non-dialysable dehydroascorbate reductase that normally protects AA from oxidation in the ECF,[239] and to the oxidation of AA in the buffer, catalysed by trace amounts of heavy metal ion impurities.[240] These factors, together with the continuous depletion of ECF AA caused by perfusion, and the difficulty in separating AA successfully from the solvent front at the HPLC stage, may explain such analytical difficulties.[63]

Voltammetric AA signals have revealed that the extracellular concentration of AA in discrete brain areas of awake animals changes with the animal's behavior

and responds to the administration of drugs. On the basis of experiments in vitro with synaptosomes, a model relating AA in the ECF with the uptake of excitatory amino acids through a heteroexchange mechanism has been proposed to explain these variations.[241] The in vivo applicability of the model has been tested by a number of independent laboratories using a variety of lesion, stimulation, microinfusion and other pharmacological experiments, and so far has not been invalidated.[63,157,242–245] It remains to be seen, however, whether this idea will survive more detailed testing in vivo, and whether AA will prove a useful tool for investigating in vivo the functions of glutamate, a ubiquitous neurotransmitter in the CNS.

7.4 Uric Acid

The origin and significance of changes in brain extracellular UA levels are not clearly understood, not least because the concentration of UA in the ECF depends to a great extent on the experimental conditions.[246] Until recently it was believed that xanthine oxidase, the enzyme involved in the metabolism of hypoxanthine to UA, was not present in brain tissue and thus there have been few reports on the production and significance of this purine metabolite in the CNS. There is now, however, direct evidence that UA is produced in the mammalian brain. Furthermore, the use of in vivo monitoring techniques has shown that, in addition to spontaneous diurnal variations, a number of different stimuli give rise to changes in the extracellular concentration of brain UA, including: electroconvulsive shock; ischemia; systemic and intracerebral administration of drugs; feeding and drinking behavior; and motivated lever pressing.[246] It appears, however, that the concentration of UA in brain ECF depends markedly on the size of the probe used and on the length of time the probe is in contact with the tissue.[139]

Thus, a possible explanation has emerged for the apparently paradoxical results for UA and 5HIAA obtained with CPEs and CFEs. For several hours after an electrode is implanted the concentration of 5HIAA and urate are similar in the low micromolar range. However, the vast majority of CPE data have been recorded in awake freely-moving animals several days following implantation. With large CPEs (320-μm diameter), the concentration of UA increases to about 50 μM by 24 h after surgery and remains high for months; peak 2 is much larger compared with the acute condition and reflects mostly (about 90%) UA.[133] When chronically implanted small electrodes are used such as CFEs or 160-μm diameter CPEs, the low level of UA observed under acute conditions actually decreases to below the limit of detection (about 1 μM[127,139]). This phenomenon appears to be linked to the lesser extent of glial reaction to the smaller electrodes, and suggests that the low micromolar level of UA detected in the ECF under acute conditions may arise from leakage across the damaged BBB.[246]

It is unclear whether extracellular UA reflects glucose utilization or neuromediator adenosine metabolism, although the weight of evidence to date favors the latter.[246] Changes in the concentration of UA detected in the ECF under acute conditions (including studies using removable electrodes and dialysis cannulae in awake animals) may reflect variations in blood flow, because the BBB is damaged temporarily and blood contains relatively high UA levels. For chronically implanted large probes (>160-μm diameter), tissue reaction results in the extracellular concentration of UA being significantly higher than normal. One import of these data is that tissue perturbation caused by the implant is highlighted, as is the need to continue to develop small stable in vivo probes.

7.5 Dioxygen

Molecular oxygen was one of the first substances detected voltammetrically in vivo, both in brain[59,60] and peripheral tissues,[247] using noble metal and carbon electrodes that continue to be the electrode materials of choice for this analyte.[180,248–250] The spread of mean concentrations reported for oxygen in cerebral tissue is quite tight, ranging from 40 μM,[251] through 50 μM[252] and 60 μM,[253] to 80 μM.[254,255] The anatomical distribution of concentrations reported within the brain is wide, however, depending on the depth of penetration of the sensor into the tissue[256] and the heterogeneity of the tissue.[251,257] Brain tissue oxygen responds to a range of perturbations including electrical stimulation[254] and neuromediator release,[258] with more rapid turnover in cerebral cortex compared with other brain areas.[252] There is good evidence that oxygen levels in brain ECF are positively and strongly correlated with cerebral blood flow,[198,259,260] and it is now possible to monitor continuous brain tissue oxygen, carbon dioxide, pH, and temperature with simultaneous recording in the blood, using multisensor probes.[261]

7.6 Nitric Oxide

NO is a relatively recently recognized neuromediator responsible for a variety of physiological responses,[262–264] and initial attempts at its electrochemical detection in biological tissues have been reviewed.[262,265–267] One of difficulties associated with detecting this free radical is its short half-life (<1 min) in tissues containing oxygen, and informal feedback from users of NO sensors (including commercial ones) suggests dissatisfaction with their

performance. Although NO can be reduced at about −900 mV versus SCE with a gas membrane-covered Pt electrode,[265] interference by molecular oxygen at this potential means that reduction mode is not suitable for NO detection in vivo. Monitoring is therefore normally carried out at high anodic potentials. A range of electrode materials have been employed at between 0.6 and 1 V, including Pt,[268] glassy carbon[269] and CFEs,[270–272] with metalloporphyrins being used in many cases to electrocatalyze the oxidation (~300 mV lower than at bare electrodes), and polymers such as Nafion® to block interference by anionic nitrogen oxides and other anions. The concept of tissue concentration of such a reactive species is questionable, but estimates lie in the range of 10–100 nM[268,273] for brain ECF, emphasizing the need for high sensitivity and selectivity for NO sensors used for VIV. The need for high potentials to detect NO at the low levels present in tissue, even when electrocatalysis is exploited, makes it difficult to avoid interference from a range of species, including AA, UA, and electroactive amino acids such as tyrosine and tryptophan, so that caution is needed in interpreting VIV data for NO. Despite these limitations, brain NO detection in vivo continues to be seen as an important objective, and reports of monitoring NO across sleep/wake cycles,[274] in response to heatstroke,[275] and following NO administration[276] are available.

7.7 Glucose and Lactate

As outlined in section 4.6, the first reported biosensor was a glucose sensor developed by Clark and Lyons in 1962.[161] Despite the prolific research on such devices since this initial work, it was not until the mid-1970s that biosensors for direct applications in the living brain were reported. A glucose biosensor developed by Silver and co-workers was one of the first enzyme-modified electrodes demonstrated to be suitable for insertion into brain tissue.[277] The electrode was prepared by simply dipping a 0.1-μm diameter Pt-coated glass micropipette electrode into an aqueous solution of GOx. The electrode, operated at +600 mV (versus Ag/AgCl), had a response time of 2–3 s that later increased to >30 s when a protective nitrocellulose membrane was added. Because these were early first generation hydrogen peroxide detecting devices, there were difficulties in the quantification of the glucose concentration arising from fluctuating oxygen and/or AA concentrations.

Later, Boutelle et al. constructed a very stable GOx-modified electrode, based on the $TTF^{+}TCNQ^{-}$ electrodes of Albery et al.,[170] for applications in brain tissue.[278] GOx was immobilized by adsorption from solution. The electrode (250 μm) was held at 0 V (versus Ag/AgCl) and the signal from AA was found to decrease following both enzyme adsorption and implantation into brain tissue. Additionally, a differential method (involving the subtraction of the current recorded at an adjacent denatured enzyme electrode) was employed to minimize any possible interferences from fluctuating levels of AA. The response time of the sensor was of the order of 20 s, and no dramatic degradation of response was observed following 6 h periods of use in brain tissue. These biosensors were used to monitor glucose in the brains of freely moving rats and were shown to be capable of monitoring changes in glucose following a local application of insulin.

More recently, Silver and Erecinska have continuously monitored extracellular glucose in the brains of anesthetized rats in acute experiments with two types of substrate-specific microelectrodes (tips < 1 μm) in a number of different physiological and pathological conditions.[279] Electrodes used in normoxic brain were needle electrodes polarized at +600 mV (versus Ag/AgCl) and prepared by electrodeposition of Pt-black on gold or Pt surfaces. The enzyme was immobilized by dipping in a solution of GOx, air-dried and cross-linked on the metal with glutaraldehyde. Finally, the tips were coated with four thin layers of cellulose diacetate, dried and dipped in a solution of ascorbate oxidase (to reduce AA interference). In hypoxic studies, mediator glucose electrodes were used with ferrocene ($E^{0'}$, +165 mV versus SCE) as the electron acceptor. Extracellular glucose levels were shown to increase in hyperglycemia and decrease in hypoglycemia, paralleling changes in blood sugar. Increased neuronal activity, and in particular spreading depression, evoked triphasic alterations in the extracellular glucose concentration. Limitation in oxygen supply (hypoxia) led to a decline in the extracellular concentration of glucose with decreases to undetectable levels seen in ischemia despite the use of mediator electrodes. The extracellular concentration of glucose was estimated at 2.4 mM, in contrast to a value of 0.5 mM obtained using quantitative microdialysis.[280]

The electrochemical polymerization of a selective poly(*o*-phenylenediamine) (PPD) layer onto Pt electrodes has also recently been reported to produce interference-free sensors having a high enzyme activity and a fast response time (about 1 s), making them ideally suited to detecting analytes directly in brain ECF.[179,180] These sensors are generally prepared quite simply by potentiostatic polymerization from an enzyme/monomer mixture and operated at +700 mV (versus SCE) to detect the enzymatically produced hydrogen peroxide. Lowry and co-workers have reported a detailed characterization of Pt/PPD/GOx biosensors (125 μm) both in vitro[179] and in vivo in the striatum of freely moving rats.[180] Confirmation that these biosensors respond to changing glucose levels in vivo was obtained by i.p. injection of

insulin and local administration of glucose (via an adjacent microdialysis probe). Interference studies involved administering AA and subanesthetic doses of ketamine i.p. Both resulted in increased extracellular AA levels with ketamine also increasing oxygen. No significant change in the Pt/PPD/GOx signal was observed in either case indicating that changes in AA and oxygen, the principal endogenous interferences, have minimal effects on the response of these first generation biosensors. They also exhibited excellent stability in vivo and have been used continuously under chronic conditions (14 days). These results were attributed to a pre-implantation conditioning procedure[180] that helped overcome the general problem of loss of sensitivity (usually between 20–50%) following implantation. Furthermore, it was confirmed that depletion of analyte by the sensor was not a problem by examining the effects of perfusion, via an adjacent microdialysis probe, of drugs such as veratridine, that increases energy metabolism by stimulating Na^+/K^+-ATPase.[281] The same group also implanted combined biosensors and microdialysis probes into the striatum of freely-moving rats to determine the ECF concentration of glucose using a modification of the Lönnroth no-net-flux quantitative microdialysis technique.[281] A value of about 0.35 mM was obtained that is in good agreement with the value of 0.47 mM determined using microdialysis.[280] Recently Pt/PPD/GOx sensors have also been used to study the effects of mild hypoxia and hyperoxia,[282] neuronal activation (tail pinch),[208] and anesthesia on brain ECF glucose.[283]

Other recent work on ECF glucose monitoring using biosensors includes that of Hu and Wilson[284] who used a needle-type biosensor (~100 μm) to monitor changes in brain glucose in anesthetized rats. The sensor has a rigorous construction protocol involving Nafion®, cellulose acetate, bovine serum albumin, and polyurethane.[285] It had a response time of about 5 s and its usefulness was demonstrated by examining the time-dependent interstitial glucose concentration in the rat hippocampus in response to KCl depolarization and by stimulation of glutamate neurons through the perforant pathway. Shram et al.[286] have also reported the use of a glucose biosensor in the living brain. The sensor was constructed by immobilizing GOx onto the surface of a carbon fiber either by cross-linking with gluteraldehyde (the latter, in given conditions, was also covered with an additional membrane of Nafion® or cellulose acetate to eliminate proteins) or by enzyme entrapment in electropolymerized films of *m*-phenylenediamine. These glucose sensors were operated using DNPV and were reported to be insensitive to interfering substances such as AA. With properly chosen pulse parameters, the response to glucose resulted in a peak at a potential of about 1 V (versus Ag/AgCl) due to the oxidation of the enzymatically produced hydrogen peroxide. Extracellular brain concentrations were monitored in anesthetized rats during administration of insulin and glucagon. These sensors were later used in a detailed investigation of brain glucose in anesthetized rats under normal and hyperglycemic conditions.[287]

One of the principal motivations of the above work to develop and utilize a reliable, selective and rapid response glucose biosensor in the living brain is the fact that glucose is the major source of energy for the mammalian brain under normal physiological conditions. However, over the past few years there has been increasing evidence that lactate, synthesized and released by astrocytes, may also serve as an important energy substrate for neurons in the adult brain.[288] This prompted Hu and Wilson[289] to construct a lactate biosensor by substituting lactate oxidase for GOx in their original biosensor design.[284] The resultant sensor was used to monitor brain extracellular lactate in anesthetized animals and changes due to neuronal activity were examined in the dentate gyrus of the hippocampus after electrical stimulation of the perforant pathway. Shram and co-workers have also reported a lactate-specific biosensor based on a CFE coated with lactate oxidase.[290] Operated with their previously reported DNPV method for glucose[287] they performed preliminary in vivo experiments in the cortex of the anesthetized rat.

An interesting new approach to in vivo neurochemical monitoring that has also been applied to monitoring ECF lactate is the dialysis electrode reported by Albery and co-workers.[291] This system combines the advantages of in vivo monitoring using voltammetry with those of microdialysis: working, reference and auxiliary electrodes are inserted into the tip of a microdialysis probe and the chosen enzyme introduced by slow perfusion using a microdialysis pump (as in a conventional microdialysis experiment). This arrangement achieves the good sensitivity, time resolution and continuous output of an implanted electrode, with the added advantage that fresh enzyme or mixtures of enzymes can be added. The dialysis electrode is now available commercially (Sycopel International Ltd., UK) and has been used by Ikegami et al.[292] to monitor the concentration of extracellular lactate in the cerebral cortex of urethane-anesthetized rats. In addition to the lactate specific lactate oxidase, ascorbate oxidase was used to help eliminate AA interference.

7.8 Glutamate

L-Glutamic acid is a key excitatory amino acid neurotransmitter in the mammalian brain, and the measurement of ECF glutamate is a key step in elucidating its role in CNS physiological and pathological states. Although microdialysis collection with ex situ detection of glutamate has been widely used,[293,294] the ability to

monitor continuous real-time glutamate dynamics has only recently been realized due to the availability of the newly isolated enzyme, glutamate oxidase (GLUOx), and the development of GLUOx-modified electrodes. Because the concentration of glutamate in brain ECF (about 3 μM[295]) is much less than that for glucose, the problem of interference by reducing agents, especially AA, is more severe. Consequently, in comparison to ECF glucose, reports of brain glutamate monitoring with implanted biosensors have been more limited.

Wilson and co-workers have used glutamate sensitive microelectrodes to monitor changes in ECF glutamate in the brains of anesthetized rats.[296] Constructed in a similar manner to their glucose biosensor,[284,289] these devices also have underlying layers of both Nafion® and cellulose acetate to help eliminate interference problems. GLUOx and ascorbate oxidase were co-immobilized by droplet evaporation from a bovine serum albumin/glutaraldehyde solution. The sensor was operated at +600 mV (versus Ag/AgCl), had a rapid response time (about 1 s), and a high degree of sensitivity (<2 μM) and selectivity. Placement of the electrode in the dentate gyrus of the hippocampus led to the detection of both KCl-induced release of glutamate and release induced by electrical stimulation of the perforant pathway, despite a 50% decrease in sensitivity observed immediately following implantation in tissue.

A glutamate sensitive dialysis electrode (see section 7.7) has also been used for the continuous measurement of glutamate efflux in the brain of a freely moving rat following a 5-min tail pinch.[291] A Nafion®-coated Pt working electrode was held at a potential of +650 mV (versus Ag/AgCl) and the enzymatic production of hydrogen peroxide by GLUOx entrapped within the dialysis membrane monitored. Electrochemical interference from AA was eliminated through the use of ascorbate oxidase. The response time in vivo of this sensor was of the order of 1 min. Several other groups have also used this type of biosensor to monitor brain glutamate. Walker and co-workers monitored both glutamate and AA in rat hippocampus after perforant path stimulation in a study of epileptic discharge after seizure.[245] Asai et al. used a dialysis electrode to monitor extracellular glutamate in rat striatum and demonstrated two distinct phases of glutamate release during early severe brain ischemia.[297] The problem of oxygen interference under such conditions was assumed to be overcome by perfusion of the enzyme solution resulting in a stable oxygen supply at the reaction site on the electrode surface. Both latter studies were performed in acute preparations with response times of between 10–15 s. The Nafion® coating was replaced with a PPD layer to help further reduce interference.

More recently, a new glutamate biosensor has been reported that involves the simple entrapment of GLUOx on a platinum electrode surface (120 μm) with a PPD film, in a similar manner to that reported by Lowry et al.[179,180] for glucose. A detailed characterization in vitro of this sensor has been performed indicating similar characteristics to the Pt/PPD/GOx sensor, and suggesting ideal suitability for monitoring in vivo.[298] Preliminary experiments in vivo to detect extracellular glutamate in the awake, freely moving rat at 1 s intervals using a mild stressor (10-s tail pinch) have been reported for this device.[299]

Another analyte on which some preliminary monitoring in the brain has been performed using enzyme-modified electrodes is choline. Michael and co-workers developed amperometric biosensors for peroxide, choline and acetylcholine by immobilizing horseradish peroxidase (HRP), choline oxidase (CHOx), and acetylcholinesterase in a cross-linked redox polymer deposited on glassy carbon electrodes.[300] Co-immobilization of HRP and the model GOx was used to establish the feasibility of highly efficient bienzyme sensors at low substrate levels. Replacing GOx with CHOx produced sensors with submicromolar detection limits and a linear response up to 0.8 mM choline. Addition of acetylcholinesterase to the sensors generated a relatively small response to acetylcholine that demonstrated the feasibility of trienzyme sensors. Similarly prepared CHOx-electrodes based on 7 and 10-μm diameter CFEs have been used by the same group to monitor choline transients in vivo in the rat brain following local injection of choline near the electrode.[301,302] The electrodes were operated at −0.1 V versus SCE and exhibited a decrease in sensitivity of less than 25% following 5 h of exposure to brain tissue. The response time of all these sensors was less than 30 s, with 2 s response times achieved in some cases.

8 COMPARISON WITH BRAIN MICRODIALYSIS

As an alternative to VIV for neurochemical analysis in intact tissue, in vivo microdialysis involves the perfusion of the tissue (using a permselective membrane), collection of the dialysate, and subsequent analysis following separation (usually by HPLC or capillary electrophoresis). Electrochemical detection is often used to quantify the concentration of species present in dialysates, although other forms of detection are available, such as spectroscopic and refractive index measurements. Recent reviews of microdialysis methodologies cover issues such as comparison with other sampling techniques,[303] technological advances,[304–306] critical analyses of the physiological significance of species in the dialysate,[307–311] studies of peripheral physiology,[312] pharmacokinetics,[313–315]

calibration issues,[316] biotechnology applications,[317,318] behavioral studies,[319,320] and applications in the human brain.[321]

The microdialysis technique has been applied to humans in clinical studies, including head injury,[322,323] ischemia,[324,325] pharmacokinetics,[326,327] glucose metabolism,[328,329] neuromediator function,[330] and for monitoring during brain surgery.[331] The majority of these have been investigations of peripheral physiology, such as metabolite kinetics in skeletal muscle,[332] histamine release in skin,[333] and the effect of drugs[334] and physical exercise[335] on fat metabolism. The greater analytical certainty associated with cerebral microdialysis has also lead to clinical applications in the brain.[321,336] Of particular interest has been the role of excitatory amino acid neurotransmitters[337] and dioxygen[338] in the development of brain tissue damage following head injury. Neurochemical correlates of epileptic seizures have also been studied using the dialysis technique to quantify the involvement of glucose metabolism[339] and neuromediators[340,341] in this phenomenon.

An advantage of microdialysis is that a wide range of compounds – neuromediators, metabolites, amino acids, drugs, etc. – can be collected in each sample and separated reliably using recent advances in HPLC and electrophoretic technologies. The main disadvantages of dialysis at present are the relatively large probe size available (minimum about 200 μm) and the poor time resolution, ranging from 1–10 min. Voltammetry, on the other hand, has a more limited range of analytes, and often VIV studies are restricted to monitoring one analyte at a time. Its advantage of small probe size and high temporal resolution, however, allows investigations not possible with microdialysis.

ACKNOWLEDGMENTS

Financial assistance from Enterprise Ireland (Grant Nos. SC/98/450 and SC/99/149), University College Dublin and the National University of Ireland, Maynooth, is gratefully acknowledged. We are grateful to Dr Maddalena Miele for artwork (Figure 3).

ABBREVIATIONS AND ACRONYMS

AA	Ascorbic Acid
BBB	Blood–Brain Barrier
cAMP	Cyclic Adenosine Monophosphate
CEE	Carbon Epoxy Electrode
CFE	Carbon Fiber Electrode
CHOx	Choline Oxidase
CNS	Central Nervous System
COMT	Catechol-*O*-Methyl Transferase
CPA	Constant Potential Amperometry
CPE	Carbon Paste Electrode
CV	Cyclic Voltammetry
DA	Dopamine
DMPA	Differential Multipulse Amperometry
DNPV	Differential Normal Pulse Voltammetry
DOPA	3,4-Dihydroxyphenylalanine
DOPAC	3,4-Dihydroxyphenylacetic Acid
DPA	Differential Pulse Amperometry
DPV	Differential Pulse Voltammetry
ECF	Extracellular Fluid
FAD	Flavin Adenine Dinucleotide
FCV	Fast Cyclic Voltammetry
GABA	γ-Aminobutyric Acid
GLUOx	Glutamate Oxidase
GOx	Glucose Oxidase
HPLC	High-performance Liquid Chromatography
HRP	Horseradish Peroxidase
HVA	Homovanillic Acid
LSV	Linear Sweep Voltammetry
MAO	Monoamine Oxidase
NE	Norepinephrine
NO	Nitric Oxide
NPV	Normal Pulse Voltammetry
PPD	Poly(*o*-phenylenediamine)
RDE	Rotating Disk Electrode
SCE	Saturated Calomel Electrode
SCV	Staircase Voltammetry
TCNQ	Tetracyanoquinodimethanide
TTF	Tetrathiafulvalenium
UA	Uric Acid
VIV	Voltammetry In Vivo
3MT	3-Methoxytyramine
5HIAA	5-Hydroxyindoleacetic Acid
5HT	5-Hydroxytryptamine

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Glucose Measurement

Electroanalytical Methods **(Volume 11)**
Neurotransmitters, Electrochemical Detection of • Pulse Voltammetry • Ultrafast Electrochemical Techniques

REFERENCES

1. S. Reinis, J.M. Goldman, *The Chemistry of Behavior*, Plenum Press, New York, 1982.
2. D. Zecevic, 'Multiple Spike-initiation Zones in Single Neurons Revealed by Voltage-sensitive Dyes', *Nature*, **381**, 322–325 (1996).
3. B.L. Sabatini, W.G. Regehr, 'Timing of Neurotransmission at Fast Synapses in the Mammalian Brain', *Nature*, **384**, 170–172 (1996).
4. P.T. Kissinger, J.B. Hart, R.N. Adams, 'Voltammetry in Brain Tissue – A New Neurophysiological Measurement', *Brain Res.*, **55**, 209–213 (1973).
5. O.P. Ottersen, A.S. Landsend, 'Organization of Glutamate Receptors at the Synapse', *Eur. J. Neurosci.*, **9**, 2219–2224 (1997).
6. D.M. Kullmann, F. Asztely, 'Extrasynaptic Glutamate Spillover in the Hippocampus: Evidence and Implications', *Trends Neurosci.*, **21**, 8–14 (1998).
7. K. Mitchell, A.F. Oke, R.N. Adams, 'In Vivo Dynamics of Norepinephrine Release–Reuptake in Multiple Terminal Field Regions of Rat Brain', *J. Neurochem.*, **63**, 917–926 (1994).
8. C.A. Marsden, M.H. Joseph, Z.L. Kruk, N.T. Maidment, R.D. O'Neill, J.O. Schenk, J.A. Stamford, 'In Vivo Voltammetry – Present Electrodes and Methods', *Neuroscience*, **25**, 389–400 (1988).
9. J.A. Stamford, 'In Vivo Voltammetry – Prospects for the Next Decade', *Trends Neurosci.*, **12**, 407–412 (1989).
10. R.N. Adams, 'In Vivo Electrochemical Measurements in the CNS', *Prog. Neurobiol.*, **35**, 297–311 (1990).
11. A.G. Ewing, T.G. Strein, L. Yi, 'Analytical Chemistry in Microenvironments: Single Nerve Cells', *Accounts Chem. Res.*, **25**, 440–447 (1992).
12. K.T. Kawagoe, J.B. Zimmerman, R.M. Wightman, 'Principles of Voltammetry and Microelectrode Surface States', *J. Neurosci. Methods*, **48**, 225–240 (1993).
13. R.D. O'Neill, 'Microvoltammetric Techniques and Sensors for Monitoring Neurochemical Dynamics In Vivo – a Review', *Analyst*, **119**, 767–779 (1994).
14. A.A. Boulton, G.B. Baker, R.N. Adams (eds.), 'Voltammetric Methods in Brain Systems', in *Neuromethods*, Humana Press, New Jersey, 1995.
15. Z.L. Kruk, J.J. O'Connor, 'Fast Electrochemical Studies in Isolated Tissues', *Trends Pharmacol. Sci.*, **16**, 145–149 (1995).
16. P. Pantano, W.G. Kuhr, 'Enzyme-modified Microelectrodes for In Vivo Neurochemical Measurements', *Electroanalysis*, **7**, 405–416 (1995).
17. C.D. Blaha, A.G. Phillips, 'A Critical Assessment of Electrochemical Procedures Applied to the Measurement of Dopamine and its Metabolites During Drug-induced and Species-typical Behaviors', *Behav. Pharmacol.*, **7**, 675–708 (1996).
18. J.A. Stamford, J.B. Justice, 'Probing Brain Chemistry', *Anal. Chem.*, **68**, A359–A363 (1996).
19. R.D. O'Neill, 'Neurochemical Analysis Using Sensors', in *Biosensors in the Body: Continuous In Vivo Monitoring*, ed. D.M. Fraser, John Wiley & Sons, Chichester, 171–196, 1997.
20. R.A. Clark, S.E. Zerby, A.G. Ewing, 'Electrochemistry in Neuronal Microenvironments', in *Electroanalytical Chemistry*, Marcel Dekker, New York, 227–294, 1998.
21. R.D. O'Neill, J.P. Lowry, M. Mas, 'Monitoring Brain Chemistry In Vivo: Voltammetric Techniques, Sensors and Behavioral Applications', *Crit. Rev. Neurobiol.*, **12**, 69–127 (1998).
22. D.J. Michael, R.M. Wightman, 'Electrochemical Monitoring of Biogenic Amine Neurotransmission in Real Time', *J. Pharmaceut. Biomed. Anal.*, **19**, 33–46 (1999).
23. J.O. Schenk, E. Miller, M.E. Rice, R.N. Adams, 'Chronoamperometry in Brain Slices: Quantitative Evaluations of In Vivo Electrochemistry', *Brain Res.*, **277**, 1–8 (1983).
24. Y. Wang, C.Y. Cheng, J.Y. Wang, J.C. Lin, 'Enhanced Antinociception of Clonidine in Spontaneously Hypertensive Rats Involves a Presynaptic Noradrenergic Mechanism', *Pharmacol. Biochem. Behav.*, **59**, 109–114 (1998).
25. C.C. Toner, J.A. Stamford, 'Characteristics of the NMDA Receptor Modulating Hypoxia/Hypoglycaemia-induced Rat Striatal Dopamine Release In Vitro', *Eur. J. Pharmacol.*, **340**, 133–143 (1998).
26. P. Palij, D.R. Bull, M.J. Sheehan, P.P.A. Humphrey, J. Millar, J.A. Stamford, Z.L. Kruk, 'A New Technique for Quantitative Measurement of Stimulated Endogenous Dopamine Release in Brain Slices Using Voltammetry', *Br. J. Pharmacol.*, **94**, 347P (1988).
27. J.O. Schenk, T.A. Patterson, J.S. McElvain, 'Rotating Disc Voltammetric Measurements in Neurobiology and Neuropharmacology', *Trends Anal. Chem.*, **9**, 325–330 (1990).
28. J.S. McElvain, J.O. Schenk, 'A Multisubstrate Mechanism of Striatal Dopamine Uptake and its Inhibition by Cocaine', *Biochem. Pharmacol.*, **43**, 2189–2199 (1992).
29. S.M. Meiergerd, J.O. Schenk, 'Striatal Transporter for Dopamine: Catechol Structure–Activity Studies and Susceptibility to Chemical Modification', *J. Neurochem.*, **62**, 998–1008 (1994).
30. S.M. Meiergerd, J.S. McElvain, J.O. Schenk, 'Effects of Cocaine and Repeated Cocaine Followed by Withdrawal', *Biochem. Pharmacol.*, **47**, 1627–1634 (1994).

31. C.J. Hatton, C. Peers, 'Electrochemical Detection of K^+-evoked Quantal Secretory Events from Isolated Rat Type I Carotid Body Cells', *Exp. Physiol.*, **82**, 415–418 (1997).
32. E.R. Travis, R.M. Wightman, 'Spatio-temporal Resolution of Exocytosis from Individual Cells', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 77–103 (1998).
33. C.D.T. Bratten, P.H. Cobbold, J.M. Cooper, 'Single Cell Measurements of Purine Release Using a Micromachined Electroanalytical Sensor', *Anal. Chem.*, **70**, 1164–1170 (1998).
34. C. Mermet, F.G. Gonon, L. Stjaerne, 'On-line Electrochemical Monitoring of the Local Noradrenaline Release Evoked by Electrical Stimulation of the Sympathetic Nerves in Isolated Rat Tail Artery', *Acta Physiol. Scand.*, **140**, 323–329 (1990).
35. F. Gonon, J.X. Bao, M. Msghina, M.F. Suaud-Chagny, L. Stjaerne, 'Fast and Local Electrochemical Monitoring of Noradrenaline Release from Sympathetic Terminals in Isolated Rat Tail Artery', *J. Neurochem.*, **60**, 1251–1257 (1993).
36. J.X. Bao, 'Sympathetic Neuromuscular Transmission in Rat Tail Artery: A Study Based on Electrochemical, Electrophysiological and Mechanical Recording', *Acta Physiol. Scand.*, **148**, 1–58 (1993).
37. T.P. Doyle, D.F. Donnelly, 'Effect of Na^+ and K^+ Channel Blockade on Baseline and Anoxia-induced Catecholamine Release from Rat Carotid Body', *J. Appl. Physiol.*, **77**, 2606–2611 (1994).
38. J.A. Brock, M. Bridgewater, T.C. Cunnane, 'Beta-adrenoceptor Mediated Facilitation of Noradrenaline and Adenosine 5′-Triphosphate Release from Sympathetic Nerves Supplying the Rat Tail Artery', *Br. J. Pharmacol.*, **120**, 769–776 (1997).
39. Y.T. Kim, D.M. Scarnulis, A.G. Ewing, 'Carbon-ring Electrodes with 1-μm Tip Diameter', *Anal. Chem.*, **58**, 1782–1786 (1986).
40. A.G. Ewing, T.G. Strein, Y.Y. Lau, 'Analytical Chemistry in Microenvironments – Single Nerve Cells', *Acc. Chem. Res.*, **25**, 440–447 (1992).
41. T.K. Chen, Y.Y. Lau, D.K.Y. Wong, A.G. Ewing, 'Pulse Voltammetry in Single Cells Using Platinum Microelectrodes', *Anal. Chem.*, **64**, 1264–1268 (1992).
42. T. Malinski, F. Bailey, J.R. Fish, F. Kiechle, 'Determination of Nickel Accumulation in Single Biological Cells Using Porphyrinic Microsensors', *Anal. Chim. Acta*, **249**, 35–41 (1991).
43. Y.Y. Lau, T. Abe, A.G. Ewing, 'Voltammetric Measurement of Oxygen in Single Neurons Using Platinized Carbon Ring Electrodes', *Anal. Chem.*, **64**, 1702–1705 (1992).
44. Y.Y. Lau, J.B. Chien, D.K.Y. Wong, A.G. Ewing, 'Characterization of the Voltammetric Response at Intracellular Carbon Ring Electrodes', *Electroanalysis*, **3**, 87–95 (1991).
45. S. Kitayama, K. Morita, T. Dohi, 'Uptake and Release of Dopamine Through the Rat Dopamine Transporter Expressed in Xenopus laevis Oocyte: Evaluation by Voltammetric Measurement of Intracellular Dopamine Concentration', *Neurosci. Lett.*, **211**, 132–134 (1996).
46. A. Meulemans, B. Poulain, G. Baux, L. Tauc, 'Changes in Serotonin Concentration in a Living Neurone: a Study by On-line Intracellular Voltammetry', *Brain Res.*, **414**, 158–162 (1987).
47. A. Albillos, G. Dernick, H. Horstmann, W. Almers, G.A. Detoledo, M. Lindau, 'The Exocytotic Event in Chromaffin Cells Revealed by Patch Amperometry', *Nature*, **389**, 509–512 (1997).
48. R. Borges, E.R. Travis, S.E. Hochstetler, R.M. Wightman, 'Effects of External Osmotic Pressure on Vesicular Secretion from Bovine Adrenal Medullary Cells', *J. Biol. Chem.*, **272**, 8325–8331 (1997).
49. Q. Xin, R.M. Wightman, 'Simultaneous Detection of Catecholamine Exocytosis and Ca^{2+} Release from Single Bovine Chromaffin Cells Using a Dual Microsensor', *Anal. Chem.*, **70**, 1677–1681 (1998).
50. J.M. Finnegan, K. Pihel, P.S. Cahill, L. Huang, S.E. Zerby, A.G. Ewing, R.T. Kennedy, R.M. Wightman, 'Vesicular Quantal Size Measured by Amperometry at Chromaffin, Mast, Pheochromocytoma, and Pancreatic β-Cells', *J. Neurochem.*, **66**, 1914–1923 (1996).
51. K. Pihel, S.C. Hsieh, J.W. Jorgenson, R.M. Wightman, 'Quantal Corelease of Histamine and 5-Hydroxytryptamine from Mast Cells and the Effects of Prior Incubation', *Biochemistry*, **37**, 1046–1052 (1998).
52. R.A. Clark, A.G. Ewing, 'Quantitative Measurements of Released Amines from Individual Exocytosis Events', *Mol. Neurobiol.*, **15**, 1–16 (1997).
53. K.D. Kozminski, D.A. Gutman, V. Davila, D. Sulzer, A.G. Ewing, 'Voltammetric and Pharmacological Characterization of Dopamine Release from Single Exocytotic Events at Rat Pheochromocytoma (PC12) Cells', *Anal. Chem.*, **70**, 3123–3130 (1998).
54. C.D. Paras, R.T. Kennedy, 'Amperometry and Cyclic Voltammetry of Tyrosine and Tryptophan-containing Oligopeptides at Carbon Fiber Microelectrodes Applied to Single Cell Analysis', *Electroanalysis*, **9**, 203–208 (1997).
55. X.J. Zhang, B. Ogorevc, M. Rupnik, M. Kreft, R. Zorec, 'Cathophoresis Paint Insulated Carbon Fiber Ultramicro Disc Electrode and its Application to In Vivo Amperometric Monitoring of Quantal Secretion from Single Rat Melanotrophs', *Anal. Chim. Acta*, **378**, 135–143 (1999).
56. M.E. Cooper, *Introduction to Animal Law*, Academic Press, London, 1987.
57. J. Hampson, *Laboratory Animal Protection Laws in Europe and North America*, Royal Society Prevention Cruelty Animals, Horsham, UK, 1985.
58. J.A. Smith, K.M. Boyd, *Lives in the Balance: The Ethics of Using Animals in Biological Research*, Oxford University Press, Oxford, 1991.

59. L.C. Clark, Jr, G. Misrahy, R.P. Fox, 'Chronically Implanted Polarographic Electrodes', *J. Appl. Physiol.*, **13**, 85–91 (1958).
60. L.C. Clark, Jr, C. Lyons, 'Studies of a Glassy Carbon Electrode for Brain Polarography with Observations on the Effect of Carbonic Anhydrase Inhibition', *Ala. J. Med. Sci.*, **2**, 353–359 (1965).
61. F. Gonon, M. Buda, R. Cespuglio, M. Jouvet, J.F. Pujol, 'Voltammetry in the Striatum of Chronic Freely Moving Rats: Detection of Catechols and Ascorbic Acid', *Brain Res.*, **223**, 69–80 (1981).
62. R.D. O'Neill, R.A. Grunewald, M. Fillenz, W.J. Albery, 'Linear Sweep Voltammetry with Carbon Paste Electrodes in the Rat Striatum', *Neuroscience*, **7**, 1945–1954 (1982).
63. R.D. O'Neill, 'The Measurement of Brain Ascorbate In Vivo and its Link with Excitatory Amino Acid Neurotransmission', in *Voltammetric Methods in Brain Systems*, eds. A.A. Boulton, G.B. Baker, R.N. Adams, Humana Press, New Jersey, 221–268, 1995.
64. R.D. O'Neill, 'Sensor – Tissue Interactions in Neurochemical Analysis with Carbon Paste Electrodes In Vivo', *Analyst*, **118**, 433–438 (1993).
65. D.E. Ormonde, R.D. O'Neill, 'The Oxidation of Ascorbic Acid at Carbon Paste Electrodes. Modified Response Following Contact with Surfactant, Lipid and Brain Tissue', *J. Electroanal. Chem.*, **279**, 109–121 (1990).
66. H.Y. Cheng, J. Schenk, R. Huff, R.N. Adams, 'In Vivo Electrochemistry: Behaviour of Micro Electrodes in Brain Tissue', *J. Electroanal. Chem.*, **100**, 23–31 (1979).
67. M.A. Dayton, A.G. Ewing, R.M. Wightman, 'Diffusion Processes Measured at Microvoltammetric Electrodes in Brain Tissue', *J. Electroanal. Chem.*, **146**, 189–200 (1983).
68. C. Amatore, R.S. Kelly, E.W. Kristensen, W.G. Kuhr, R.M. Wightman, 'Effects of Restricted Diffusion at Ultramicroelectrodes in Brain Tissue. The Pool Model: Theory and Experiment for Chronoamperometry', *J. Electroanal. Chem.*, **213**, 31–42 (1986).
69. C. Nicholson, E. Sykova, 'Extracellular Space Structure Revealed by Diffusion Analysis', *Trends Neurosci.*, **21**, 207–215 (1998).
70. D.A. Kane, R.D. O'Neill, 'Major Differences in the Behaviour of Carbon Paste and Carbon Fiber Electrodes in a Protein – Lipid Matrix: Implications for Voltammetry In Vivo', *Analyst*, **123**, 2899–2903 (1998).
71. R.D. O'Neill, M. Fillenz, 'Detection of Homovanillic Acid In Vivo Using Microcomputer-controlled Voltammetry: Simultaneous Monitoring of Rat Motor Activity and Striatal Dopamine Release', *Neuroscience*, **14**, 753–763 (1985).
72. J.L. Gonzalez-Mora, J.A. Sanchez-Bruno, M. Mas, 'Concurrent On-line Analysis of Striatal Ascorbate, Dopamine and Dihydroxyphenylacetic Acid Concentrations by In Vivo Voltammetry', *Neurosci. Lett.*, **86**, 61–66 (1988).
73. R.L. Wilson, R.M. Wightman, 'Systemic and Nigral Application of Amphetamine Both Cause an Increase in Extracellular Concentration of Ascorbate in the Caudate Nucleus of the Rat', *Brain Res.*, **339**, 219–226 (1985).
74. D.J. Wiedemann, A. Basse-Tomusk, R.L. Wilson, G.V. Rebec, R.M. Wightman, 'Interference by DOPAC and Ascorbate During Attempts to Measure Drug-induced Changes in Neostriatal Dopamine with Nafion®-coated, Carbon-fiber Electrodes', *J. Neurosci. Methods*, **35**, 9–18 (1990).
75. F. Moussy, D.J. Harrison, 'Prevention of the Rapid Degradation of Subcutaneously Implanted Ag/AgCl Reference Electrodes Using Polymer Coatings', *Anal. Chem.*, **66**, 674–679 (1994).
76. M. Fillenz, R.D. O'Neill, 'Effects of Light Reversal on the Circadian Pattern of Motor Activity and Voltammetric Signals Recorded in Rat Forebrain', *J. Physiol. (London)*, **374**, 91–101 (1986).
77. U. Gustafsson, F. Sjoeberg, 'Serotonin – One Possible Link Between Oxygen Metabolism and the Regulation of Blood Flow in the Brain?', *Int. J. Microcirc. Clin. Exp.*, **16**, 143–146 (1996).
78. M.W. Uhl, P.M. Kochanek, J.K. Schiding, J.A. Melick, E.M. Nemoto, 'The Regional Cerebral Blood Flow Response to Cortical Microelectrode Insertion is Neutrophil Dependent', *Adv. Exp. Med. Biol.*, **317**, 701–705 (1992).
79. R. Sternberg, D.S. Bindra, G.S. Wilson, D.R. Thevenot, 'Covalent Enzyme Coupling on Cellulose Acetate Membranes for Glucose Sensor Development', *Anal. Chem.*, **60**, 2781–2786 (1988).
80. Z.H. Fan, D.J. Harrison, 'Permeability of Glucose and Other Neutral Species Through Recast Perfluorosulfonated Ionomer Films', *Anal. Chem.*, **64**, 1304–1311 (1992).
81. C. Malitesta, F. Palmisano, L. Torsi, P.G. Zambonin, 'Glucose Fast-response Amperometric Sensor Based on Glucose Oxidase Immobilized in an Electropolymerized Poly(*o*-phenylenediamine) Film', *Anal. Chem.*, **62**, 2735–2740 (1990).
82. Y. Ohnuki, H. Matsuda, T. Ohsaka, N. Oyama, 'Permselectivity of Films Prepared by Electrochemical Oxidation of Phenol and Amino-aromatic Compounds', *J. Electroanal. Chem.*, **158**, 55–67 (1983).
83. J.C. Conti, E. Strope, R.N. Adams, C.A. Marsden, 'Voltammetry in Brain Tissue: Chronic Recording of Stimulated Dopamine and 5HT Release', *Life Sci.*, **23**, 2705–2716 (1978).
84. R. Huff, R.N. Adams, 'Dopamine Release in Nucleus Accumbens and Striatum by Clozapine: Simultaneous Monitoring by In Vivo Electrochemistry', *Neuropharmacology*, **19**, 587–590 (1980).
85. H.H. Zhang, L.A. Coury, 'Effects of High-intensity Ultrasound on Glassy Carbon Electrodes', *Anal. Chem.*, **65**, 1552–1558 (1993).

86. L. Falat, H.Y. Cheng, 'Voltammetric Differentiation of Ascorbic Acid and Dopamine at an Electrochemically Treated Graphite-epoxy Electrode', *Anal. Chem.*, **54**, 2108–2111 (1982).
87. G.A. Gerhardt, A.F. Oke, G. Nagy, B. Moghaddam, R.N. Adams, 'Nafion®-coated Electrodes with High Selectivity for CNS Electrochemistry', *Brain Res.*, **290**, 390–395 (1984).
88. M.P. Brazell, R.J. Kasser, K.J. Renner, J. Feng, B. Moghaddam, R.N. Adams, 'Electrocoating Carbon Fiber Microelectrodes with Nafion® Improves Selectivity for Electroactive Neurotransmitters', *J. Neurosci. Methods*, **22**, 167–172 (1987).
89. S. Alegret, 'Rigid Carbon–Polymer Biocomposites for Electrochemical Sensing – A Review', *Analyst*, **121**, 1751–1758 (1996).
90. M. Armstrong-James, J. Millar, 'Carbon Fiber Microelectrodes', *J. Neurosci. Methods*, **1**, 279–287 (1979).
91. F. Gonon, C.M. Fombarlet, M. Buda, J.F. Pujol, 'Electrochemical Treatments of Pyrolytic Carbon Fiber Electrodes', *Anal. Chem.*, **53**, 1386–1389 (1981).
92. C. Forni, A. Nieoullon, 'Electrochemical Detection of Dopamine Release in the Striatum of Freely Moving Hamsters', *Brain Res.*, **297**, 11–20 (1984).
93. J.P. Rivot, D. Pointis, J.M. Besson, 'In Vivo Electrochemical Detection of 5-Hydroxyindole within the Trigeminal Nucleus Caudatus of Freely Moving Rats: the Effect of Morphine', *Brain Res.*, **463**, 275–283 (1988).
94. A. Schulte, R.H. Chow, 'Cylindrically Etched Carbon-fiber Microelectrodes for Low-noise Amperometric Recording of Cellular Secretion', *Anal. Chem.*, **70**, 985–990 (1998).
95. A.H. Swiergiel, V.S. Palamarchouk, A.J. Dunn, 'A New Design of Carbon Fiber Microelectrode for In Vivo Voltammetry Using Fused Silica', *J. Neurosci. Methods*, **73**, 29–33 (1997).
96. X.J. Zhang, B. Ogorevc, 'Poly(tetrafluoroethylene) Film Housing of Carbon Fibers Using Capillary-pull Technology for One-stage Fabrication of Carbon Disk Ultramicroelectrodes and their Characterization', *Anal. Chem.*, **70**, 1646–1651 (1998).
97. J.L. Ponchon, R. Cespuglio, F. Gonon, M. Jouvet, J.F. Pujol, 'Normal Pulse Polarography with Carbon Fiber Electrodes for In Vitro and In Vivo Determination of Catecholamines', *Anal. Chem.*, **51**, 1483–1486 (1979).
98. F. Gonon, M. Buda, R. Cespuglio, M. Jouvet, J.F. Pujol, 'In Vivo Electrical Detection of Catechols in the Rat Neostriatum: Dopamine or DOPAC?', *Nature*, **286**, 902–904 (1980).
99. F. Marcenac, F. Gonon, 'Fast In Vivo Monitoring of Dopamine Release in the Rat Brain with Differential Pulse Amperometry', *Anal. Chem.*, **57**, 1778–1779 (1985).
100. C. Oh, T.W. Gardiner, G.V. Rebec, 'Blockade of Both D1- and D2-dopamine Receptors Inhibits Amphetamine-induced Ascorbate Release in the Neostriatum', *Brain Res.*, **480**, 184–189 (1989).
101. T.Z. Peng, H.H. Lu, G.Q. Liu, Y.P. Cao, 'An Improved Electrochemical Procedure for Treating Carbon Fiber Electrodes for In Vivo Determination of some Neurotransmitters', *Anal. Lett.*, **25**, 795–805 (1992).
102. T.G. Strein, A.G. Ewing, 'In Situ Laser Activation of Carbon Fiber Microdisk Electrodes', *Anal. Chem.*, **63**, 194–198 (1991).
103. C.D. Allred, R.L. McCreery, 'Adsorption of Catechols on Fractured Glassy Carbon Electrode Surfaces', *Anal. Chem.*, **64**, 444–448 (1992).
104. G.M. Swain, T. Kuwana, 'Anodic Fracturing and Vacuum Heat Treated Annealing of Pitch-based Carbon Fibers', *Anal. Chem.*, **64**, 565–568 (1992).
105. S. Sujaritvanichpong, K. Aoki, K. Tokuda, H. Matsuda, 'Electrochemical Behavior of Dopamine at Carbon Fiber Electrodes', *J. Electroanal. Chem.*, **198**, 195–203 (1986).
106. P.M. Kovach, M.R. Deakin, R.M. Wightman, 'Electrochemistry at Partially Blocked Carbon-fiber Microcylinder Electrodes', *J. Phys. Chem.*, **90**, 4612–4617 (1986).
107. B.W. Chun, C.R. Davis, Q.A. He, R.R. Gustafson, 'Development of Surface Acidity During Electrochemical Treatment of Pan-carbon Fibers', *Carbon*, **30**, 177–187 (1992).
108. E. Desimoni, G.I. Casella, T.R.I. Cataldi, A.M. Salvi, T. Rotunno, E. Dicroce, 'Remarks on the Surface Characterization of Carbon Fibers', *Surf. Interface Anal.*, **18**, 623–630 (1992).
109. Y.M. Xie, P.M.A. Sherwood, 'X-ray Photoelectron Spectroscopic Studies of Carbon Fibers. 15. Electrochemical Treatment on Pitch-based Fibers by Potentiostatic and Galvanostatic Methods', *Appl. Spectrosc.*, **45**, 1158–1165 (1991).
110. J.X. Feng, M. Brazell, K. Renner, R. Kasser, R.N. Adams, 'Electrochemical Pretreatment of Carbon Fibers for In Vivo Electrochemistry: Effects on Sensitivity and Response Time', *Anal. Chem.*, **59**, 1863–1867 (1987).
111. R.M. Wightman, M.R. Deakin, P.M. Kovach, W.G. Kuhr, K.J. Stutts, 'Methods to Improve Electrochemical Reversibility at Carbon Electrodes', *J. Electrochem. Soc.*, **131**, 1578–1583 (1984).
112. F.E. Woodard, D.E. McMackins, R.E.W. Jansson, 'Electrosorption of Organics on Three Dimensional Carbon Fiber Electrodes', *J. Electroanal. Chem.*, **214**, 303–330 (1986).
113. E. Theodoridou, J.O. Besenhard, H.P. Fritz, 'Chemically Modified Carbon Fiber Electrodes. Part 1. Bulk-functionalized Carbon Fibers', *J. Electroanal. Chem.*, **122**, 67–71 (1981).
114. Z. Hahn, R. Cespuglio, H. Faradji, M. Jouvet, 'Factors Influencing the Properties of Voltammetric Carbon Fiber Electrodes: the Importance of the pH of the Medium Used for the Electrical Treatment and of the Resin Coating of the Fibers', *J. Biochem. Biophys. Methods*, **11**, 265–275 (1985).
115. P. Hulthe, K. Johannessen, L. Svensson, J. Engel, 'A New Pretreatment Method for Carbon Fiber Microelectrodes

Enhances the Selectivity for Dopamine', *J. Neurosci. Methods*, **29**, 297 (1989).
116. M. Nakahara, K. Shimizu, 'Effects of Electrolyte on the Structure of Pyrolytic Graphite Surfaces in Anodic Oxidation', *J. Mater. Sci.*, **27**, 1207–1211 (1992).
117. R.J. Kasser, K.J. Renner, J.X. Feng, M.P. Brazell, R.N. Adams, 'Spreading Depression Induced by 100 mM KCl in Caudate is Blocked by Local Anaesthesia of the Substantia Nigra', *Brain Res.*, **475**, 333–344 (1988).
118. R.J.W. Pentney, A. Gratton, 'Effects of Local Delta and Mu Opioid Receptor Activation on Basal and Stimulated Dopamine Release in Striatum and Nucleus Accumbens of Rat: An In Vivo Electrochemical Study', *Neuroscience*, **45**, 95–102 (1991).
119. S.R. Jones, S.J. O'Dell, J.F. Marshall, R.M. Wightman, 'Functional and Anatomical Evidence for Different Dopamine Dynamics in the Core and Shell of the Nucleus Accumbens in Slices of Rat Brain', *Synapse*, **23**, 224–231 (1996).
120. P.S. Cahill, Q.D. Walker, J.M. Finnegan, G.E. Mickelson, E.R. Travis, R.M. Wightman, 'Microelectrodes for the Measurement of Catecholamines in Biological Systems', *Anal. Chem.*, **68**, 3180–3186 (1996).
121. R. Cespuglio, H. Faradji, F. Riou, M. Buda, F. Gonon, J.F. Pujol, M. Jouvet, 'Differential Pulse Voltammetry in Brain Tissue. II. Detection of 5-Hydroxyindoleacetic Acid in the Rat Striatum', *Brain Res.*, **223**, 299–311 (1981).
122. F. Crespi, T. Sharp, N. Maidment, C.A. Marsden, 'Differential Pulse Voltammetry In Vivo – Evidence that Uric Acid Contributes to the Indole Oxidation Peak', *Neurosci. Lett.*, **43**, 203–207 (1983).
123. K. Taguchi, Y. Suzuki, 'The Response of the 5-Hydroxyindole Oxidation Current to Noxious Stimuli in the Spinal Cord of Anesthetized Rats: Modification by Morphine', *Brain Res.*, **583**, 150–154 (1992).
124. H.W. Clement, D. Gemsa, W. Wesemann, 'The Effect of Adrenergic Drugs on Serotonin Metabolism in the Nucleus Raphe Dorsalis of the Rat, Studied by In Vivo Voltammetry', *Eur. J. Pharmacol.*, **217**, 43–48 (1992).
125. L. Imeri, M.G. De Simoni, R. Giglio, A. Clavenna, M. Mancia, 'Changes in the Serotonergic System During the Sleep–Wake Cycle: Simultaneous Polygraphic and Voltammetric Recordings in Hypothalamus Using a Telemetry System', *Neuroscience*, **58**, 353–358 (1994).
126. J.P. Rivot, E. Noret, L. Ory-Lavollee, J.M. Besson, 'In Vivo Electrochemical Detection of 5-Hydroxyindoles in the Dorsal Horn of the Spinal Cord: the Contribution of Uric Acid to the Voltammograms', *Brain Res.*, **419**, 201–207 (1987).
127. R. Cespuglio, N. Sarda, A. Gharib, H. Faradji, N. Chastrette, 'Differential Pulse Voltammetry In Vivo with Working Carbon Fiber Electrodes: 5-Hydroxyindole Compounds or Uric Acid Detection', *Exp. Brain Res.*, **64**, 589–595 (1986).
128. R.F. Lane, A.T. Hubbard, C.D. Blaha, 'Application of Semi-derivative Electroanalysis to Studies of Neurotransmitters in the Central Nervous System', *J. Electroanal. Chem.*, **95**, 117–122 (1979).
129. R.D. O'Neill, M. Fillenz, W.J. Albery, N.J. Goddard, 'The Monitoring of Ascorbate and Monoamine Transmitter Metabolites in the Striatum of Unanaesthetised Rats Using Microprocessor-based Voltammetry', *Neuroscience*, **9**, 87–93 (1983).
130. R.D. O'Neill, M. Fillenz, W.J. Albery, 'The Development of Linear Sweep Voltammetry with Carbon Paste Electrodes In Vivo', *J. Neurosci. Methods*, **8**, 263–273 (1983).
131. H.Y. Cheng, 'Compartment Model for Chronoamperometric Measurement In Vivo', *J. Electroanal. Chem.*, **135**, 145–151 (1982).
132. W.J. Albery, M. Fillenz, R.D. O'Neill, 'The Compartment Model for Chronically Implanted Voltammetric Electrodes in the Rat Brain', *Neurosci. Lett.*, **38**, 175–180 (1983).
133. R.D. O'Neill, J.L. Gonzalez-Mora, M.G. Boutelle, D.E. Ormonde, J.P. Lowry, A. Duff, B. Fumero, M. Fillenz, M. Mas, 'Anomalously High Concentrations of Brain Extracellular Uric Acid Detected with Chronically Implanted Probes: Implications for In Vivo Sampling Techniques', *J. Neurochem.*, **57**, 22–29 (1991).
134. M.H. Joseph, H. Hodges, J.A. Gray, 'Lever Pressing for Food Reward and In Vivo Voltammetry: Evidence for Increases in Extracellular Homovanillic Acid, the Dopamine Metabolite, and Uric Acid in the Rat Caudate Nucleus', *Neuroscience*, **32**, 195–201 (1989).
135. R.D. O'Neill, M. Fillenz, R.A. Grunewald, M.R. Bloomfield, W.J. Albery, C.M. Jamieson, J.H. Williams, J.A. Gray, 'Voltammetric Carbon Paste Electrodes Monitor Uric Acid and not 5HIAA at the 5-Hydroxyindole Potential in the Rat Brain', *Neurosci. Lett.*, **45**, 39–46 (1984).
136. K. Mueller, R. Palmour, C.D. Andrews, P.J. Knott, 'In Vivo Voltammetric Evidence of Production of Uric Acid by Rat Caudate', *Brain Res.*, **335**, 231–235 (1985).
137. M.G. Boutelle, L. Svensson, M. Fillenz, 'Rapid Changes in Striatal Ascorbate in Response to Tail-pinch Monitored by Constant Potential Voltammetry', *Neuroscience*, **30**, 11–17 (1989).
138. R.D. O'Neill, R.A. Grunewald, M. Fillenz, W.J. Albery, 'The Effect of Unilateral Cortical Lesions on the Circadian Changes in Rat Striatal Ascorbate and Homovanillic Acid Levels Measured In Vivo Using Voltammetry', *Neurosci. Lett.*, **42**, 105–110 (1983).
139. A. Duff, R.D. O'Neill, 'Effect of Probe Size on the Concentration of Brain Extracellular Uric Acid Monitored with Carbon Paste Electrodes', *J. Neurochem.*, **62**, 1496–1502 (1994).
140. M.H. Joseph, A.M.J. Young, 'Pharmacological Evidence, Using In Vivo Dialysis, that Substances Additional to Ascorbic Acid, Uric Acid and Homovanillic

Acid Contribute to the Voltammetric Signals Obtained in Unrestrained Rats from Chronically Implanted Carbon Paste Electrodes', *J. Neurosci. Methods*, **36**, 209–218 (1991).

141. A. Slivka, T.S. Brannan, J. Weinberger, P.J. Knott, G. Cohen, 'Increase in Extracellular Dopamine in the Striatum During Cerebral Ischemia: a Study Utilizing Cerebral Microdialysis', *J. Neurochem.*, **50**, 1714–1718 (1988).
142. A.G. Vulto, T. Sharp, U. Ungerstedt, D.H.G. Versteeg, 'Rapid Postmortem Increase in Extracellular Dopamine in the Rat Brain as Assessed by Brain Microdialysis', *J. Neurochem.*, **51**, 746–749 (1988).
143. J.L. Gonzalez-Mora, N.T. Maidment, T. Guadalupe, M. Mas, 'Post-mortem Dopamine Dynamics Assessed by Voltammetry and Microdialysis', *Brain Res. Bull.*, **23**, 323–327 (1989).
144. K.T. Kawagoe, P.A. Garris, D.J. Wiedemann, R.M. Wightman, 'Regulation of Transient Dopamine Concentration Gradients in the Microenvironment Surrounding Nerve Terminals in the Rat Striatum', *Neuroscience*, **51**, 55–64 (1992).
145. W.H. Church, J.B. Justice, Jr, 'Rapid Sampling and Determination of Extracellular Dopamine In Vivo', *Anal. Chem.*, **59**, 712–716 (1987).
146. C.D. Blaha, R.F. Lane, 'Chemically Modified Electrode for In Vivo Monitoring of Brain Catecholamines', *Brain Res. Bull.*, **10**, 861–864 (1983).
147. P.A. Broderick, 'State-of-the-art Microelectrodes for In Vivo Voltammetry', *Electroanalysis*, **2**, 241–251 (1990).
148. P.D. Lyne, R.D. O'Neill, 'Stearate-modified Carbon Paste Electrodes for Detecting Dopamine In Vivo: Decrease in Selectivity Caused by Lipids and Other Surface-active Agents', *Anal. Chem.*, **62**, 2347–2351 (1990).
149. J.A. Clemens, L.A. Phebus, 'Changes in Brain Chemistry Produced by Dopaminergic Agents: In Vivo Electrochemical Monitoring Reveals Opposite Changes in Anaesthetized vs Unanaesthetized Rats', *Brain Res.*, **267**, 183–186 (1983).
150. K. Mueller, C. Haskett, 'Effects of Haloperidol on Amphetamine-induced Increases in Ascorbic Acid and Uric Acid as Determined by Voltammetry In Vivo', *Pharmacol. Biochem. Behav.*, **27**, 231–234 (1987).
151. M.G. Boutelle, L. Svensson, M. Fillenz, 'Effect of Diazepam on Behavior and Associated Changes in Ascorbate Concentration in Rat Brain Areas: Striatum, n. Accumbens and Hippocampus', *Psychopharmacology*, **100**, 230–236 (1990).
152. K. Mueller, 'Repeated Administration of High Doses of Amphetamine Increases Release of Ascorbic Acid in Caudate but not Nucleus Accumbens', *Brain Res.*, **494**, 30–35 (1989).
153. M.P. Brazell, S.N. Mitchell, M.H. Joseph, J.A. Gray, 'Acute Administration of Nicotine Increases the In Vivo Extracellular Levels of Dopamine, 3,4-Dihydroxyphenylacetic Acid and Ascorbic Acid Preferentially in the Nucleus Accumbens of the Rat: Comparison with Caudate-putamen', *Neuropharmacology*, **29**, 1177–1185 (1990).
154. R.D. O'Neill, M. Fillenz, 'Simultaneous Monitoring of Dopamine Release in Rat Frontal Cortex, Nucleus Accumbens and Striatum: Effect of Drugs, Circadian Changes and Correlations with Motor Activity', *Neuroscience*, **16**, 49–55 (1985).
155. K. Mueller, P.M. Kunko, 'The Effects of Amphetamine and Pilocarpine on the Release of Ascorbic and Uric Acid in Several Rat Brain Areas', *Pharmacol. Biochem. Behav.*, **35**, 871–876 (1990).
156. C. Haskett, K. Mueller, 'The Effects of Serotonin Depletion on the Voltammetric Response to Amphetamine', *Pharmacol. Biochem. Behav.*, **28**, 381–384 (1987).
157. R.D. O'Neill, M. Fillenz, L. Sundstrom, J.N.P. Rawlins, 'Voltammetrically Monitored Brain Ascorbate as an Index of Excitatory Amino Acid Release in the Unrestrained Rat', *Neurosci. Lett.*, **52**, 227–233 (1984).
158. L.J. May, R.M. Wightman, 'Heterogeneity of Stimulated Dopamine Overflow Within Rat Striatum as Observed with In Vivo Voltammetry', *Brain Res.*, **487**, 311–320 (1989).
159. A.P.F. Turner, I. Karube, G.S. Wilson (eds.), *Biosensors: Fundamentals and Applications*, Oxford University Press, New York, 1987.
160. A.E.G. Cass, *Biosensors: A Practical Approach*, Oxford University Press, New York, 1990.
161. L.C. Clark, Jr, C. Lyons, 'Electrode Systems for Continuous Monitoring in Cardiovascular Surgery', *Ann. NY Acad. Sci.*, **102**, 29–45 (1962).
162. R. Wilson, A.P.F. Turner, 'Glucose Oxidase: an Ideal Enzyme', *Biosensors Bioelectron.*, **7**, 165–185 (1992).
163. S.J. Updike, G.P. Hicks, 'The Enzyme Electrode', *Nature (London)*, **214**, 986–988 (1967).
164. G.G. Guilbault, *Analytical Uses of Immobilised Enzymes*, Marcel Dekker, New York, 1984.
165. B. Grundig, C. Krabisch, 'Electron Mediator-modified Electrode for the Determination of Glucose in Fermentation Media', *Anal. Chim. Acta*, **222**, 75–81 (1989).
166. S.K. Beh, G.J. Moody, J.D.R. Thomas, 'Studies on Enzyme Electrodes with Ferrocene and Carbon Paste Bound with Cellulose Triacetate', *Analyst*, **116**, 459–462 (1991).
167. N. Martens, E.A.H. Hall, 'Model for an Immobilized Oxidase Enzyme Electrode in the Presence of Two Oxidants', *Anal. Chem.*, **66**, 2763–2770 (1994).
168. B.A. Gregg, A. Heller, 'Redox Polymer Films Containing Enzymes. 1. A Redox-conducting Epoxy Cement: Synthesis, Characterization, and Electrocatalytic Oxidation of Hydroquinone', *J. Phys. Chem.*, **95**, 5970–5975 (1991).

169. M.V. Pishko, A.C. Michael, A. Heller, 'Amperometric Glucose Microelectrodes Prepared Through Immobilization of Glucose Oxidase in Redox Hydrogels', *Anal. Chem.*, **63**, 2268–2272 (1991).
170. W.J. Albery, P.N. Bartlett, D.H. Craston, 'Amperometric Enzyme Electrodes. Part II. Conducting Salts as Electrode Materials for the Oxidation of Glucose Oxidase', *J. Electroanal. Chem.*, **194**, 223–235 (1985).
171. P.N. Bartlett, P.R. Birkin, F. Palmisano, G. Debenedetto, 'A Study on the Direct Electrochemical Communication Between Horseradish Peroxidase and a Poly(aniline) Modified Electrode', *J. Chem. Soc., Faraday Trans.*, **92**, 3123–3130 (1996).
172. R.M. Ianniello, T.J. Lindsay, A.M. Yacynych, 'Differential Pulse Voltammetric Study of Direct Electron Transfer in Glucose Oxidase Chemically Modified Graphite Electrodes', *Anal. Chem.*, **54**, 1098–1101 (1982).
173. K. Narasimhan, L.B.J. Wingard, 'Enhanced Direct Electron Transfer with Glucose Oxidase Immobilized on (Aminophenyl)boronic Acid Modified Glassy Carbon Electrode', *Anal. Chem.*, **58**, 2984–2987 (1986).
174. H.P. Bennetto, D.R. DeKeyzer, G.M. Delaney, A. Koshy, J.R. Mason, G. Mourla, L.A. Razack, J.L. Stirling, C.F. Thurston, D.J. Anderton, W.H. Mullen, 'A Glucose Oxidase Electrode for Amperometric Determination of Glucose', *Int. Ind. Biotech.*, **8**, 5–10 (1988).
175. Y. Degani, A. Heller, 'Direct Electrical Communication Between Chemically Modified Enzymes and Metal Electrodes. 2. Methods for Bonding Electron-transfer Relays to Glucose Oxidase and D-Amino-acid Oxidase', *J. Am. Chem. Soc.*, **110**, 2615–2620 (1988).
176. A. Heller, 'Electrical Wiring of Redox Enzymes', *Acc. Chem. Res.*, **23**, 128–134 (1990).
177. P.N. Bartlett, 'Some Studies of Electrodes made from Single Crystals of TTF.TCNQ', *J. Electroanal. Chem.*, **300**, 175–189 (1991).
178. J.P. Lowry, R.D. O'Neill, 'Partial Characterization in Vitro of Glucose Oxidase-modified Poly(phenylenediamine)-coated Electrodes for Neurochemical Analysis In Vivo', *Electroanalysis*, **6**, 369–379 (1994).
179. J.P. Lowry, K. McAteer, S.S. El Atrash, A. Duff, R.D. O'Neill, 'Characterization of Glucose Oxidase-modified Poly(phenylenediamine)-coated Electrodes In Vitro and In Vivo: Homogeneous Interference by Ascorbic Acid in Hydrogen Peroxide Detection', *Anal. Chem.*, **66**, 1754–1761 (1994).
180. J.P. Lowry, M. Miele, R.D. O'Neill, M.G. Boutelle, M. Fillenz, 'An Amperometric Glucose-oxidase/poly(*o*-phenylenediamine) Biosensor for Monitoring Brain Extracellular Glucose: In Vivo Characterisation in the Striatum of Freely-moving Rats', *J. Neurosci. Methods*, **79**, 65–74 (1998).
181. F. Hefti, D. Felix, 'Chronoamperometry In Vivo: Does it Interfere with Spontaneous Neuronal Activity in the Brain', *J. Neurosci. Methods*, **7**, 151–156 (1983).
182. T. Nakazato, S. Hosoda, A. Akiyama, 'A Triangular Conditioning Voltage Wave does not Influence Spontaneous Neuronal Activity in the Rat Striatum', *J. Neurosci. Methods*, **46**, 69–72 (1993).
183. F. Crespi, T. England, E. Ratti, D.G. Trist, 'Carbon Fiber Micro-electrodes for Concomitant In Vivo Electrophysiological and Voltammetric Measurements: No Reciprocal Influences', *Neurosci. Lett.*, **188**, 33–36 (1995).
184. E.A. Kiyatkin, 'Cocaine Enhances the Changes in Extracellular Dopamine in Nucleus Accumbens Associated with Reinforcing Stimuli: A High-speed Chronoamperometric Study in Freely Moving Rats', *Eur. J. Neurosci.*, **5**, 284–291 (1993).
185. R.M. Sullivan, A. Gratton, 'Relationships Between Stress-induced Increases in Medial Prefrontal Cortical Dopamine and Plasma Corticosterone Levels in Rats: Role of Cerebral Laterality', *Neuroscience*, **83**, 81–91 (1998).
186. T. Shinba, Y. Andow, T. Shinozaki, N. Ozawa, K. Yamamoto, 'Phasic Increase of Monoamine-related Electrochemical Signal in the Rat Caudate Nucleus Following Conditioned Auditory Stimulation During the Reaction-time Task', *Brain Res.*, **781**, 284–290 (1998).
187. G.A. Gerhardt, C. Ksir, C. Pivik, S.D. Dickinson, J. Sabeti, N.R. Zahniser, 'Methodology for Coupling Local Application of Dopamine and Other Chemicals with Rapid In Vivo Electrochemical Recordings in Freely-moving Rats', *J. Neurosci. Methods*, **87**, 67–76 (1999).
188. R.C. Pierce, A.J. Clemens, C.P. Grabner, G.V. Rebec, 'Amphetamine Promotes Neostriatal Ascorbate Release via a Nigro-thalamo-cortico-neostriatal Loop', *J. Neurochem.*, **63**, 1499–1507 (1994).
189. X.J. Zhang, C.F. Wang, X.Y. Zhou, 'Voltammetry of Dihydroxyphenylalanine (L-DOPA) Using a Nafion®-coated Carbon Fiber Ultramicroelectrode Array', *Anal. Chim. Acta*, **265**, 27–34 (1992).
190. S.D. Young, A.C. Michael, 'Voltammetry of Extracellular Dopamine in Rat Striatum During ICSS-like Electrical Stimulation of the Medial Forebrain Bundle', *Brain Res.*, **600**, 305–307 (1993).
191. K. Pihel, T.J. Schroeder, R.M. Wightman, 'Rapid and Selective Cyclic Voltammetric Measurements of Epinephrine and Norepinephrine as a Method to Measure Secretion from Single Bovine Adrenal Medullary Cells', *Anal. Chem.*, **66**, 4532–4537 (1994).
192. B.P. Jackson, S.M. Dietz, R.M. Wightman, 'Fast-scan Cyclic Voltammetry of 5-Hydroxytryptamine', *Anal. Chem.*, **67**, 1115–1120 (1995).
193. M.A. Bunin, C. Prioleau, R.B. Mailman, R.M. Wightman, 'Release and Uptake Rates of 5-Hydroxytryptamine in the Dorsal Raphe and Substantia Nigra Reticulata of the Rat Brain', *J. Neurochem.*, **70**, 1077–1087 (1998).

194. G.V. Rebec, 'Real-time Assessments of Dopamine Function During Behavior: Single-unit Recording, Iontophoresis, and Fast-scan Cyclic Voltammetry in Awake, Unrestrained Rats', *Alcohol Clin. Exp. Res.*, **22**, 32–40 (1998).
195. Z.L. Kruk, S. Cheeta, J. Milla, R. Muscat, J.E.G. Williams, P. Willner, 'Real Time Measurement of Stimulated Dopamine Release in the Conscious Rat Using Fast Cyclic Voltammetry: Dopamine Release is not Observed During Intracranial Self Stimulation', *J. Neurosci. Methods*, **79**, 9–19 (1998).
196. R.M. Wightman, C. Amatore, R.C. Engstrom, P.D. Hale, E.W. Kristensen, W.G. Kuhr, L.J. May, 'Real-time Characterization of Dopamine Overflow and Uptake in the Rat Striatum', *Neuroscience*, **25**, 513–523 (1988).
197. H.P.M. Horikawa, T. Nakazato, O. Hikosaka, 'Duration of Catalepsy Correlates with Increased Intrastriatal Sulpiride', *Eur. J. Pharmacol.*, **326**, 15–21 (1997).
198. J.P. Lowry, M.G. Boutelle, M. Fillenz, 'Measurement of Brain Tissue Oxygen at a Carbon Paste Electrode can Serve as an Index of Increases in Regional Cerebral Blood Flow', *J. Neurosci. Methods*, **71**, 177–182 (1997).
199. T. Nakazato, A. Akiyama, 'In Vivo Electrochemical Measurement of the Long-lasting Release of Dopamine and Serotonin Induced by Intrastriatal Kainic Acid', *J. Neurochem.*, **69**, 2039–2047 (1997).
200. F.G. Gonon, F. Navarre, M.J. Buda, 'In Vivo Monitoring of Dopamine Release in the Rat Brain with Differential Pulse Voltammetry', *Anal. Chem.*, **56**, 573–575 (1984).
201. M.M. Murphy, J.J. O'Dea, J. Osteryoung, 'Pulse Voltammetry at Microcylinder Electrodes', *Anal. Chem.*, **63**, 2743–2750 (1991).
202. L. Svensson, J. Zhang, K. Johannessen, J.A. Engel, 'Effect of Local Infusion of Glutamate Analogs into the Nucleus Accumbens of Rats: An Electrochemical and Behavioral Study', *Brain Res.*, **643**, 155–161 (1994).
203. M. Mas, J.L. Gonzalez-Mora, 'Monitoring Brain Neurotransmitter Release During Sociosexual Interactions', in *Monitoring Molecules in Neuroscience*, eds. J.L. Gonzalez-Mora, R. Borges, M. Mas, University of La Laguna, Tenerife, 308–309, 1996.
204. M. Miele, M. Fillenz, 'Changes in Extracellular Brain Ascorbate in Rat Striatum in Response to Administration of Non-volatile Anaesthetic Agents', *Br. J. Anaesth.*, **78**, 588–590 (1997).
205. K. Mitchell, R.N. Adams, 'Comparison of the Effects of Voltage-sensitive Calcium Channel Antagonism on the Electrically Stimulated Release of Dopamine and Norepinephrine In Vivo', *Brain Res.*, **604**, 349–353 (1993).
206. C. Dugast, M.F. Suaud-Chagny, F. Gonon, 'Continuous In Vivo Monitoring of Evoked Dopamine Release in the Rat Nucleus Accumbens by Amperometry', *Neuroscience*, **62**, 647–654 (1994).
207. D.G. Buerk, S. Lahiri, D. Chugh, A. Mokashi, 'Electrochemical Detection of Rapid DA Release Kinetics During Hypoxia in Perfused-superfused Cat CB', *J. Appl. Physiol.*, **78**, 830–837 (1995).
208. J.P. Lowry, M. Fillenz, 'Evidence for Uncoupling of Oxygen and Glucose Utilization During Neuronal Activation in Rat Striatum', *J. Physiol. (London)*, **498**, 497–501 (1997).
209. G. Paxinos, C. Watson, *The Rat Brain in Stereotaxic Coordinates*, Academic Press, Sidney, 1986.
210. J.M. Kauffmann, J.C. Vire, 'Pharmaceutical and Biomedical Applications of Electroanalysis – A Critical Review', *Anal. Chim. Acta*, **273**, 329–334 (1993).
211. M.W. Lada, R.T. Kennedy, 'Quantitative In Vivo Monitoring of Primary Amines in Rat Caudate Nucleus Using Microdialysis Coupled by a Flow-gated Interface to Capillary Electrophoresis with Laser-induced Fluorescence Detection', *Anal. Chem.*, **68**, 2790–2797 (1996).
212. A. Meulemans, J.P. Mothet, A. Schirar, P. Fossier, L. Tauc, G. Baux, 'A Nitric Oxide Synthase Activity is Involved in the Modulation of Acetylcholine Release in Aplysia Ganglion Neurons: A Histological, Voltammetric and Electrophysiological Study', *Neuroscience*, **69**, 985–995 (1995).
213. M.E. Rice, J. Cammack, 'Anoxia-resistant Turtle Brain Maintains Ascorbic Acid Content In Vitro', *Neurosci. Lett.*, **132**, 141–145 (1991).
214. M.K. Sanghera, F. Crespi, K.F. Martin, D.J. Heal, W.R. Buckett, C.A. Marsden, 'Biochemical and In Vivo Voltammetric Evidence for Differences in Striatal Dopamine Levels in Inbred Strains of Mice', *Neuroscience*, **39**, 649–656 (1990).
215. M.E. Rice, C.D. Richards, S. Nedergaard, J. Hounsgaard, C. Nicholson, S.A. Greenfield, 'Direct Monitoring of Dopamine and 5-HT Release in Substantia Nigra and Ventral Tegmental Area In Vitro', *Exp. Brain Res.*, **100**, 395–406 (1994).
216. B.A. Baldwin, M.W. Hill, J. Osborne, D.F. Sharman, 'Voltammetric Recording from the Brain of a Conscious Sheep', *J. Physiol. (London)*, **343**, 5P–5P (1983).
217. G.A. Gerhardt, W.A. Cass, J. Hudson, M. Henson, Z.M. Zhang, A. Ovadia, B.J. Hoffer, D.M. Gash, 'In Vivo Electrochemical Studies of Dopamine Overflow and Clearance in the Striatum of Normal and MPTP-treated Rhesus Monkeys', *J. Neurochem.*, **66**, 579–588 (1996).
218. R.S. Kelly, R.M. Wightman, 'Detection of Dopamine Overflow and Diffusion with Voltammetry in Slices of Rat Brain', *Brain Res.*, **423**, 79–87 (1987).
219. F. Crespi, C. Moebius, 'In Vivo Selective Monitoring of Basal Levels of Cerebral Dopamine Using Voltammetry with Nafion® Modified (NA-CRO) Carbon Fiber Microelectrodes', *J. Neurosci. Methods*, **42**, 149–161 (1992).
220. M.F. Suaud-Chagny, M. Buda, F.G. Gonon, 'Pharmacology of Electrically Evoked Dopamine Release Studied in the Rat Olfactory Tubercle by In Vivo Electrochemistry', *Eur. J. Pharmacol.*, **164**, 273–283 (1989).
221. M.F. Suaud-Chagny, K. Chergui, G. Chouvet, F. Gonon, 'Relationship Between Dopamine Release in the Rat

Nucleus Accumbens and the Discharge Activity of Dopaminergic Neurons During Local In Vivo Application of Amino Acids in the Ventral Tegmental Area', *Neuroscience*, **49**, 63–72 (1992).
222. P.A. Garris, J.R.C. Christensen, G.V. Rebec, R.M. Wightman, 'Real-time Measurement of Electrically Evoked Extracellular Dopamine in the Striatum of Freely Moving Rats', *J. Neurochem.*, **68**, 152–161 (1997).
223. Y. Lu, J.L. Peters, A.C. Michael, 'Direct Comparison of the Response of Voltammetry and Microdialysis to Electrically Evoked Release of Striatal Dopamine', *J. Neurochem.*, **70**, 584–593 (1998).
224. P. Bickford-Wimer, K. Pang, G.M. Rose, G.A. Gerhardt, 'Electrically-evoked Release of Norepinephrine in the Rat Cerebellum: An In Vivo Electrochemical and Electrophysiological Study', *Brain Res.*, **558**, 305–311 (1991).
225. L.J. May, W.G. Kuhr, R.M. Wightman, 'Differentiation of Dopamine Overflow and Uptake Processes in the Extracellular Fluid of the Rat Caudate Nucleus with Fast-scan In Vivo Voltammetry', *J. Neurochem.*, **51**, 1060–1069 (1988).
226. R.M. Wightman, J.B. Zimmerman, 'Control of Dopamine Extracellular Concentration in Rat Striatum by Impulse Flow and Uptake', *Brain Res. Rev.*, **15**, 135–144 (1990).
227. F. Crespi, K.F. Martin, C.A. Marsden, 'Measurement of Extracellular Basal Levels of Serotonin In Vivo Using Nafion®-coated Carbon Fiber Electrodes Combined with Differential Pulse Voltammetry', *Neuroscience*, **27**, 885–896 (1988).
228. F. Crespi, J.C. Garratt, A.J. Sleight, C.A. Marsden, 'In Vivo Evidence that 5-Hydroxytryptamine (5-HT) Neuronal Firing and Release are not Necessarily Correlated with 5-HT Metabolism', *Neuroscience*, **34**, 139–144 (1990).
229. D.L. Tanelian, P. Kosek, I. Mody, M.B. MacIver, 'The Role of the $GABA_A$ Receptor/Chloride Channel Complex in Anesthesia', *Anesthesiology*, **78**, 757–776 (1993).
230. F. Crespi, P.E. Keane, M. Morre, 'Anaesthesia Abolishes the Effect of Valproate on Extracellular 5HIAA, DOPAC and Ascorbate as Measured in Rat Striatum by Differential Pulse Voltammetry', *Br. J. Pharmacol.*, **87**, 279–283 (1986).
231. J.L. Gonzalez-Mora, T. Guadalupe, B. Fumero, M. Mas, 'Mathematical Resolution of Mixed In Vivo Voltammetry Signals. Models, Equipment, Assessment by Simultaneous Microdialysis Sampling', *J. Neurosci. Methods*, **39**, 231–244 (1991).
232. M. Mas, B. Fumero, R.D. O'Neill, J.L. Gonzalez-Mora, 'Fixed Versus Removable Microdialysis Probes for In Vivo Neurochemical Analysis: Implications for Behavioral Studies. Reply', *J. Neurochem.*, **64**, 1900–1901 (1995).
233. G.V. Rebec, J.R.C. Christensen, C. Guerra, M.T. Bardo, 'Regional and Temporal Differences in Real-time Dopamine Efflux in the Nucleus Accumbens During Free-choice Novelty', *Brain Res.*, **776**, 61–67 (1997).
234. K. Inoue, N. Kiriike, M. Okuno, H. Ito, Y. Fujisaki, T. Matsui, Y. Kawakita, 'Scheduled Feeding Caused Activation of Dopamine Metabolism in the Striatum of Rats', *Physiol. Behav.*, **53**, 177–181 (1993).
235. L. Naudon, N. Dourmap, I. Leroux-Nicollet, J. Costentin, 'Kainic Acid Lesion of the Striatum Increases Dopamine Release but Reduces 3-Methoxytyramine Level', *Brain Res.*, **572**, 247–249 (1992).
236. K.J. Renner, L. Pazos, R.N. Adams, 'In Vivo Voltammetric Evidence for the Detection of Norepinephrine Release in the Thalamus of Freely Moving Rats', *Brain Res.*, **577**, 49–56 (1992).
237. N. Brose, R.D. O'Neill, M.G. Boutelle, M. Fillenz, 'Dopamine in the Basal Ganglia and Benzodiazepine-induced Sedation', *Neuropharmacology*, **27**, 589–595 (1988).
238. M.H. Joseph, H. Hodges, 'Lever Pressing for Food Reward and Changes in Dopamine Turnover and Uric Acid in Rat Caudate and Nucleus Accumbens Studied Chronically by In Vivo Voltammetry', *J. Neurosci. Methods*, **34**, 143–149 (1990).
239. R.C. Rose, 'Cerebral Metabolism of Oxidized Ascorbate', *Brain Res.*, **628**, 49–55 (1993).
240. J.C. Deutsch, J.F. Kolhouse, 'Ascorbate and Dehydroascorbate Measurements in Aqueous Solutions and Plasma Determined by Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **65**, 321–326 (1993).
241. M. Fillenz, R.D. O'Neill, R.A. Grunewald, 'Changes in Extracellular Brain Ascorbate Concentration as an Index of Excitatory Amino Acid Release'. in *Monitoring Neurotransmitter Release During Behavior*, eds. M.H. Joseph, M. Fillenz, I.A. Macdonald, C.A. Marsden, Ellis Horwood, Chichester, 144–163, 1986.
242. J. Cammack, B. Ghasemzadeh, R.N. Adams, 'The Pharmacological Profile of Glutamate-evoked Ascorbic Acid Efflux Measured by In Vivo Electrochemistry', *Brain Res.*, **565**, 17–22 (1991).
243. G.V. Rebec, R.C. Pierce, 'A Vitamin as Neuromodulator: Ascorbate Release into the Extracellular Fluid of the Brain Regulates Dopaminergic and Glutamatergic Transmission', *Prog. Neurobiol.*, **43**, 537–565 (1994).
244. M. Miele, M.G. Boutelle, M. Fillenz, 'The Physiologically Induced Release of Ascorbate in Rat Brain is Dependent on Impulse Traffic, Calcium Influx and Glutamate Uptake', *Neuroscience*, **62**, 87–91 (1994).
245. M.C. Walker, P.T. Galley, M.L. Errington, S.D. Shorvon, J.G.R. Jefferys, 'Ascorbate and Glutamate Release in the Rat Hippocampus after Perforant Path Stimulation: A ''dialysis electrode'' Study', *J. Neurochem.*, **65**, 725–731 (1995).
246. R.D. O'Neill, J.P. Lowry, 'On the Significance of Brain Extracellular Uric Acid Detected with In-vivo Monitoring Techniques: A Review', *Behav. Brain Res.*, **71**, 33–49 (1995).

247. L.C. Clark, Jr, E.W. Clark, 'Epicardial Oxygen Measured with a Pyrolytic Graphite Electrode', *Ala. J. Med. Sci.*, **1**, 142–148 (1964).
248. D.W. Luebbers, 'Oxygen Electrodes and Optodes and their Application In Vivo', *Adv. Exp. Med. Biol.*, **388**, 13–34 (1996).
249. A. Dittmar, S. Mangin, C. Ruban, W.H. Newman, H.F. Bowman, V. Dupuis, G. Delhomme, N.F. Shram, R. Cespuglio, N. Jaffrezic-Renault, P. Roussel, D. Barbier, C. Martelet, 'In Vivo and In Vitro Evaluation of Specially Designed Old and Carbon Fiber Oxygen Microelectrodes for Living Tissues', *Sens. Actuators B - Chem.*, **44**, 316–320 (1997).
250. B. Jedlinska, A. Mellstroem, M. Hartmann, K. Joensson, 'Comparison of Tissue Oxygen-tension Measurements by Different Devices. An Experimental Study in Pigs', *Scand. J. Clin. Lab. Invest.*, **58**, 63–71 (1998).
251. R. Murr, S. Berger, L. Schuerer, K. Peter, A. Baethmann, 'A Novel, Remote-controlled Suspension Device for Brain Tissue PO_2 Measurements with Multiwire Surface Electrodes', *Pflugers Arch.*, **426**, 348–350 (1994).
252. P.K. Nair, D.G. Buerk, J.H. Halsey, Jr, 'Comparison of Oxygen Metabolism and Tissue pO_2 in Cortex and Hippocampus', *Stroke*, **18**, 616–622 (1987).
253. A. Zauner, R. Bullock, X. Di, H.F. Young, 'Brain Oxygen, CO_2, pH, and Temperature Monitoring: Evaluation in the Feline Brain', *Neurosurgery*, **37**, 1168–1176 (1995).
254. D.B. McCreery, W.F. Agnew, L.A. Bullara, T.G. Yuen, 'Partial Pressure of Oxygen in Brain and Peripheral Nerve During Damaging Electrical Stimulation', *J. Biomed. Eng.*, **12**, 309–315 (1990).
255. T. Kayama, T. Yoshimoto, S. Fujimoto, Y. Sakurai, 'Intratumoral Oxygen Pressure in Malignant Brain Tumor', *J. Neurosurg.*, **74**, 66–69 (1991).
256. H. Baumgartl, U. Heinrich, D.W. Lubbers, 'Oxygen Supply of the Blood-free Perfused Guinea-pig Brain in Normo- and Hypothermia Measured by the Local Distribution of Oxygen Pressure', *Pflugers Arch.*, **414**, 228–234 (1989).
257. D.W. Luebbers, H. Baumgaertl, 'Heterogeneities and Profiles of Oxygen Pressure in Brain and Kidney as Examples of the pO_2 Distribution in the Living Tissue', *Kidney Int.*, **51**, 372–380 (1997).
258. R.T. Kennedy, S.R. Jones, R.M. Wightman, 'Simultaneous Measurement of Oxygen and Dopamine: Coupling of Oxygen Consumption and Neurotransmission', *Neuroscience*, **47**, 603–612 (1992).
259. M.R. Gaab, B. Poch, V. Heller, 'Oxygen Tension, Oxygen Metabolism, and Microcirculation in Vasogenic Brain Edema', *Adv. Neurol.*, **52**, 247–256 (1990).
260. J.B. Zimmerman, R.T. Kennedy, R.M. Wightman, 'Evoked Neuronal Activity Accompanied by Transmitter Release Increases Oxygen Concentration in Rat Striatum In Vivo but not In Vitro', *J. Cereb. Blood Flow Metab.*, **12**, 629–637 (1992).
261. M. Menzel, A. Rieger, S. Roth, J. Soukup, I. Furka, I. Miko, P. Molnar, C. Peuse, C. Hennig, J. Radke, 'Comparison Between Continuous Brain Tissue pO_2, pCO_2, pH, and Temperature and Simultaneous Cerebrovenous Measurement Using a Multisensor Probe in a Porcine Intracranial Pressure Model', *J. Neurotrauma*, **15**, 265–276 (1998).
262. F.L. Kiechle, T. Malinski, 'Nitric Oxide: Biochemistry, Pathophysiology, and Detection', *Am. J. Clin. Pathol.*, **100**, 567–575 (1993).
263. E.M. Schuman, D.V. Madison, 'Nitric Oxide and Synaptic Function', *Annu. Rev. Neurosci.*, **17**, 153–183 (1994).
264. J. Zhang, S.H. Snyder, 'Nitric Oxide in the Nervous System', *Annu. Rev. Pharmacol. Toxicol.*, **35**, 213–233 (1995).
265. D. Christodoulou, S. Kudo, J.A. Cook, M.C. Krishna, A. Miles, M.B. Grisham, R. Murugesan, P.C. Ford, D.A. Wink, 'Electrochemical Methods for Detection of Nitric Oxide', *Methods Enzymol.*, **268**, 69–83 (1996).
266. F. Bedioui, S. Trevin, J. Devynck, 'Chemically Modified Microelectrodes Designed for the Electrochemical Determination of Nitric Oxide in Biological Systems', *Electroanalysis*, **8**, 1085–1091 (1996).
267. A.J. Gow, S.R. Thom, C. Brass, H. Ischiropoulos, 'Electrochemical Detection of Nitric Oxide in Biological Systems', *Microchem. J.*, **56**, 146–154 (1997).
268. K. Shibuki, 'An Electrochemical Microprobe for Detecting Nitric Oxide Release in Brain Tissue', *Neurosci. Res.*, **9**, 69–76 (1990).
269. M. Pallini, A. Curulli, A. Amine, G. Palleschi, 'Amperometric Nitric Oxide Sensors: a Comparative Study', *Electroanalysis*, **10**, 1010–1016 (1998).
270. M.N. Friedemann, S.W. Robinson, G.A. Gerhardt, '*o*-Phenylenediamine-modified Carbon Fiber Electrodes for the Detection of Nitric Oxide', *Anal. Chem.*, **68**, 2621–2628 (1996).
271. B. Fabre, S. Burlet, R. Cespuglio, G. Bidan, 'Voltammetric Detection of NO in the Rat Brain with an Electronic Conducting Polymer and Nafion® Bilayer-coated Carbon Fiber Electrode', *J. Electroanal. Chem.*, **426**, 75–83 (1997).
272. K.M. Mitchell, E.K. Michaelis, 'Multimembrane Carbon Fiber Electrodes for Physiological Measurements of Nitric Oxide', *Electroanalysis*, **10**, 81–88 (1998).
273. L.Q. Mao, G.Y. Shi, Y. Tian, H.Y. Liu, L.T. Jin, K. Yamamoto, S.G. Tao, J.Y. Jin, 'A Novel Thin-layer Amperometric Detector Based on Chemically Modified Ring-disc Electrode and its Application for Simultaneous Measurements of Nitric Oxide and Nitrite in Rat Brain Combined with In Vivo Microdialysis', *Talanta*, **46**, 1547–1556 (1998).
274. S. Burlet, R. Cespuglio, 'Voltammetric Detection of Nitric Oxide (NO) in the Rat Brain: Its Variations Throughout the Sleep–Wake Cycle', *Neurosci. Lett.*, **226**, 131–135 (1997).

275. F. Canini, L. Bourdon, R. Cespuglio, A. Buguet, 'Voltametric Assessment of Brain Nitric Oxide During Heatstroke in Rats', *Neurosci. Lett.*, **231**, 67–70 (1997).
276. M. Tristani-Firouzi, E.G. DeMaster, B.J. Quast, D.P. Nelson, S.L. Archer, 'Utility of a Nitric Oxide Electrode for Monitoring the Administration of Nitric Oxide in Biologic Systems', *J. Lab. Clin. Med.*, **131**, 281–285 (1998).
277. I.A. Silver, 'A Microglucose Electrode', in *Ion and Enzyme Electrodes*, ed. M. Kessler, Urban and Schwarzenberg, Munich, 189–192, 1976.
278. M.G. Boutelle, C. Stanford, M. Fillenz, W.J. Albery, P.N. Bartlett, 'An Amperometric Enzyme Electrode for Monitoring Brain Glucose in the Freely Moving Rat', *Neurosci. Lett.*, **72**, 283–288 (1986).
279. I.A. Silver, M. Erecinska, 'Extracellular Glucose Concentration in Mammalian Brain: Continuous Monitoring of Changes During Increased Neuronal Activity and upon Limitation in Oxygen Supply in Normo-, Hypo-, and Hyperglycemic Animals', *J. Neurosci.*, **14**, 5068–5076 (1994).
280. L.K. Fellows, M.G. Boutelle, M. Fillenz, 'Extracellular Brain Glucose Levels Reflect Local Neuronal Activity: A Microdialysis Study in Awake, Freely Moving Rats', *J. Neurochem.*, **59**, 2141–2147 (1992).
281. J.P. Lowry, R.D. O'Neill, M.G. Boutelle, M. Fillenz, 'Continuous Monitoring of Extracellular Glucose Concentrations in the Striatum of Freely Moving Rats with an Implanted Glucose Biosensor', *J. Neurochem.*, **70**, 391–396 (1998).
282. J.P. Lowry, M. Demestre, M. Fillenz, 'Relation Between Cerebral Blood Flow and Extracellular Glucose in Rat Striatum During Mild Hypoxia and Hyperoxia', *Dev. Neurosci.*, **20**, 52–58 (1998).
283. M. Fillenz, J.P. Lowry, 'The Relation Between Local Cerebral Blood Flow and Extracellular Glucose Concentration in Rat Striatum', *Exp. Physiol.*, **83**, 233–238 (1998).
284. Y.B. Hu, G.S. Wilson, 'Rapid Changes in Local Extracellular Rat Brain Glucose Observed with an In Vivo Glucose Sensor', *J. Neurochem.*, **68**, 1745–1752 (1997).
285. Y.N. Zhang, G.S. Wilson, 'In Vitro and In Vivo Evaluation of Oxygen Effects on a Glucose Oxidase Based Implantable Glucose Sensor', *Anal. Chim. Acta*, **281**, 513–520 (1993).
286. L.I. Netchiporouk, N.F. Shram, N. Jaffrezic-Renault, C. Martelet, R. Cespuglio, 'In Vivo Brain Glucose Measurements: Differential Normal Pulse Voltammetry with Enzyme-modified Carbon Fiber Microelectrodes', *Anal. Chem.*, **68**, 4358–4364 (1996).
287. N.F. Shram, L.I. Netchiporouk, C. Martelet, N. Jaffrezic-Renault, R. Cespuglio, 'Brain Glucose: Voltammetric Determination in Normal and Hyperglycaemic Rats Using a Glucose Microsensor', *NeuroReport*, **8**, 1109–1112 (1997).
288. N. Westergaard, U. Sonnewald, A. Schousboe, 'Metabolic Trafficking Between Neurons and Astrocytes: The Glutamate Glutamine Cycle Revisited', *Dev. Neurosci.*, **17**, 203–211 (1995).
289. Y.B. Hu, G.S. Wilson, 'A Temporary Local Energy Pool Coupled to Neuronal Activity: Fluctuations of Extracellular Lactate Levels in Rat Brain Monitored with Rapid-response Enzyme-based Sensor', *J. Neurochem.*, **69**, 1484–1490 (1997).
290. N.F. Shram, L.I. Netchiporouk, C. Martelet, N. Jaffrezic-Renault, C. Bonnet, R. Cespuglio, 'In Vivo Voltammetric Detection of Rat Brain Lactate with Carbon Fiber Microelectrodes Coated with Lactate Oxidase', *Anal. Chem.*, **70**, 2618–2622 (1998).
291. W.J. Albery, M.G. Boutelle, P.T. Galley, 'The Dialysis Electrode – a New Method for In Vivo Monitoring', *J. Chem. Soc., Chem. Commun.*, 900–901 (1992).
292. Y. Ikegami, M. Maeda, A. Yokota, Y. Hayashida, 'Cerebral Extracellular Lactate Concentration and Blood Flow During Chemical Stimulation of the Nucleus Tractus Solitarii in Anesthetized Rats', *Brain Res.*, **758**, 33–38 (1997).
293. M.H. Millan, T.P. Obrenovitch, G.S. Sarna, S.Y. Lok, L. Symon, B.S. Meldrum, 'Changes in Rat Brain Extracellular Glutamate Concentration During Seizures Induced by Systemic Picrotoxin or Focal Bicuculline Injection: an In Vivo Dialysis Study with On-line Enzymatic Detection', *Epilepsy Res.*, **9**, 86–91 (1991).
294. H.L. Rowley, K.F. Martin, C.A. Marsden, 'Determination of In Vivo Amino Acid Neurotransmitters by High-performance Liquid Chromatography with *o*-phthalaldehyde-sulphite Derivatisation', *J. Neurosci. Methods*, **57**, 93–99 (1995).
295. M. Miele, M. Berners, M.G. Boutelle, H. Kusakabe, M. Fillenz, 'The Determination of the Extracellular Concentration of Brain Glutamate Using Quantitative Microdialysis', *Brain Res.*, **707**, 131–133 (1996).
296. Y. Hu, K.M. Mitchell, F.N. Albahadily, E.K. Michaelis, G.S. Wilson, 'Direct Measurement of Glutamate Release in the Brain Using a Dual Enzyme-based Electrochemical Sensor', *Brain Res.*, **659**, 117–125 (1994).
297. S. Asai, Y. Iribe, T. Kohno, K. Ishikawa, 'Real Time Monitoring of Biphasic Glutamate Release Using Dialysis Electrode in Rat Acute Brain Ischemia', *NeuroReport*, **7**, 1092–1096 (1996).
298. M.R. Ryan, J.P. Lowry, R.D. O'Neill, 'Biosensor for Neurotransmitter L-Glutamic Acid Designed for Efficient Use of L-Glutamate Oxidase and Effective Rejection of Interference', *Analyst*, **122**, 1419–1424 (1997).
299. J.P. Lowry, M.R. Ryan, R.D. O'Neill, 'Behaviorally Induced Changes in Extracellular Levels of Brain Glutamate Monitored at 1 s Resolution with an Implanted Biosensor', *Anal. Commun.*, **35**, 87–89 (1998).
300. M.G. Garguilo, N. Huynh, A. Proctor, A.C. Michael, 'Amperometric Sensors for Peroxide, Choline, and Acetylcholine Based on Electron Transfer Between

Horseradish Peroxidase and a Redox Polymer', *Anal. Chem.*, **65**, 523–528 (1993).
301. M.G. Garguilo, A.C. Michael, 'An Enzyme-modified Microelectrode that Detects Choline Injected Locally into Brain Tissue', *J. Am. Chem. Soc.*, **115**, 12218–12219 (1993).
302. M.G. Garguilo, A.C. Michael, 'Quantitation of Choline in the Extracellular Fluid of Brain Tissue with Amperometric Microsensors', *Anal. Chem.*, **66**, 2621–2629 (1994).
303. R.D. Myers, A. Adell, M.F. Lankford, 'Simultaneous Comparison of Cerebral Dialysis and Push–Pull Perfusion in the Brain of Rats: A Critical Review', *Neurosci. Biobehav. Rev.*, **22**, 371–387 (1998).
304. L.A. Dawson, 'Capillary Electrophoresis and Microdialysis: Current Technology and Applications', *J. Chromatogr. B Biomed. Sci. Appl.*, **697**, 89–99 (1997).
305. C.E. MacTaylor, A.G. Ewing, 'Critical Review of Recent Developments in Fluorescence Detection for Capillary Electrophoresis', *Electrophoresis*, **18**, 2279–2290 (1997).
306. M.I. Davies, C.E. Lunte, 'Microdialysis Sampling Coupled On-line to Microseparation Techniques', *Chem. Soc. Rev.*, **26**, 215–222 (1997).
307. C. Humpel, T. Ebendal, L. Olson, 'Microdialysis: A Way to Study In Vivo Release of Neurotrophic Bioactivity: A Critical Summary', *J. Mol. Med.*, **74**, 523–526 (1996).
308. G. Di Chiara, G. Tanda, E. Carboni, 'Estimation of In-vivo Neurotransmitter Release by Brain Microdialysis: The Issue of Validity', *Behav. Pharmacol.*, **7**, 640–657 (1996).
309. P. Osborne, 'Interpretation of Microdialysis Data', *J. Neurochem.*, **68**, 1773–1774 (1997).
310. W. Timmerman, B.H.C. Westerink, 'Brain Microdialysis of GABA and Glutamate: What Does it Signify?', *Synapse*, **27**, 242–261 (1997).
311. B.H.C. Westerink, W. Timmerman, 'Do Neurotransmitters Sampled by Brain Microdialysis Reflect Functional Release?', *Anal. Chim. Acta*, **379**, 263–274 (1999).
312. L. Groth, 'Cutaneous Microdialysis – Methodology and Validation', *Acta Derm. Venereol. (Stockh)*, **76**(Suppl 197), 1–61 (1996).
313. W.F. Elmquist, R.J. Sawchuk, 'Application of Microdialysis in Pharmacokinetic Studies', *Pharm. Res.*, **14**, 267–288 (1997).
314. E.C.M. De Lange, M. Danhof, A.G. De Boer, D.D. Breimer, 'Methodological Considerations of Intracerebral Microdialysis in Pharmacokinetic Studies on Drug Transport Across the Blood–Brain Barrier', *Brain Res. Rev.*, **25**, 27–49 (1997).
315. M.I. Davies, 'A Review of Microdialysis Sampling for Pharmacokinetic Applications', *Anal. Chim. Acta*, **379**, 227–249 (1999).
316. J.A. Stenken, 'Methods and Issues in Microdialysis Calibration', *Anal. Chim. Acta*, **379**, 337–358 (1999).
317. N. Torto, T. Laurell, L. Gorton, G. Markovarga, 'Recent Trends in the Application of Microdialysis in Bioprocesses', *Anal. Chim. Acta*, **374**, 111–135 (1998).
318. R. John, M.J. John, 'Evaluation of the Microdialysis Electrode Concept as a Generic Biosensing Technology', *Electroanalysis*, **10**, 1125–1129 (1998).
319. M. Mas, 'Neurobiological Correlates of Masculine Sexual Behavior', *Neurosci. Biobehav. Rev.*, **19**, 261–277 (1995).
320. B.H.C. Westerink, 'Brain Microdialysis and its Application for the Study of Animal Behavior', *Behav. Brain Res.*, **70**, 103–124 (1995).
321. C. Hamani, M.S. Luer, M. Dujovny, 'Microdialysis in the Human Brain: Review of its Applications', *Neurol. Res.*, **19**, 281–288 (1997).
322. L. Persson, J. Valtysson, P. Enblad, P.E. Waerme, K. Cesarini, A. Lewen, L. Hillered, 'Neurochemical Monitoring Using Intracerebral Microdialysis in Patients with Subarachnoid Hemorrhage', *J. Neurosurg.*, **84**, 606–616 (1996).
323. J.C. Goodman, A.B. Valadka, S.P. Gopinath, M. Cormio, C.S. Robertson, 'Lactate and Excitatory Amino Acids Measured by Microdialysis are Decreased by Pentobarbital Coma in Head-injured Patients', *J. Neurotrauma*, **13**, 549–556 (1996).
324. R. Kanthan, A. Shuaib, R. Griebel, H. Miyashita, J. Kalra, 'Glucose-induced Decrease in Glutamate Levels in Ischemic Human Brain by In-vivo Microdialysis', *Neurosci. Lett.*, **209**, 207–209 (1996).
325. P. Enblad, J. Valtysson, J. Andersson, A. Lilja, S. Valind, G. Antoni, B. Langstroem, L. Hillered, L. Persson, 'Simultaneous Intracerebral Microdialysis and Positron Emission Tomography in the Detection of Ischemia in Patients with Subarachnoid Hemorrhage', *J. Cereb. Blood Flow Metab.*, **16**, 637–644 (1996).
326. M. Mueller, O. Haag, T. Burgdorff, A. Georgopoulos, W. Weninger, B. Jansen, G. Stanek, H. Pehamberger, E. Agneter, H.G. Eichler, 'Characterization of Peripheral-compartment Kinetics of Antibiotics by In Vivo Microdialysis in Humans', *Antimicrob. Agents Chemother.*, **40**, 2703–2709 (1996).
327. M.T. O'Connell, F. Tison, N.P. Quinn, P.N. Patsalos, 'Clinical Drug Monitoring by Microdialysis: Application to Levodopa Therapy in Parkinson's Disease', *Br. J. Clin. Pharmacol.*, **42**, 765–769 (1996).
328. E. Hagstroem-Toft, S. Enoksson, E. Moberg, J. Bolinder, P. Arner, 'Absolute Concentrations of Glycerol and Lactate in Human Skeletal Muscle, Adipose Tissue, and Blood', *Am. J. Physiol. Endocrinol. Metab.*, **273**, E584–E592 (1997).

329. T. Vering, S. Adam, H. Drewer, C. Dumschat, R. Steinkuhl, A. Schulze, E.G. Siegel, M. Knoll, 'Wearable Microdialysis System for Continuous In Vivo Monitoring of Glucose', *Analyst*, **123**, 1605–1609 (1998).
330. R. Kanthan, A. Shuaib, R. Griebel, H. El-Alazounni, H. Miyashita, J. Kalra, 'Evaluation of Monoaminergic Neurotransmitters in the Acute Focal Ischemic Human Brain Model by Intracerebral In Vivo Microdialysis', *Neurochem. Res.*, **21**, 563–566 (1996).
331. H. Baechli, H. Langemann, A. Mendelowitsch, B. Alessandri, H. Landolt, O. Gratzl, 'Microdialytic Monitoring During Cerebrovascular Surgery', *Neurol. Res.*, **18**, 370–376 (1996).
332. R.C. Hickner, J.S. Fisher, A.A. Ehsani, W.M. Kohrt, 'Role of Nitric Oxide in Skeletal Muscle Blood Flow at Rest and During Dynamic Exercise in Humans', *Am. J. Physiol. Heart Circ. Physiol.*, **273**, H405–H410 (1997).
333. L.J. Petersen, K. Winge, E. Brodin, P.S. Skov, 'No Release of Histamine and Substance P in Capsaicin-induced Neurogenic Inflammation in Intact Human Skin In Vivo: A Microdialysis Study', *Clin. Exp. Allergy*, **27**, 957–965 (1997).
334. J. Galitzky, M. Lafontan, J. Nordenstroem, P. Arner, 'Role of Vascular Alpha-2 Adrenoceptors in Regulating Lipid Mobilization from Human Adipose Tissue', *J. Clin. Invest.*, **91**, 1997–2003 (1993).
335. B. Stallknecht, L. Simonsen, J. Buelow, J. Vinten, H. Galbo, 'Effect of Training on Epinephrine-stimulated Lipolysis Determined by Microdialysis in Human Adipose Tissue', *Am. J. Physiol. Endocrinol. Metab.*, **269**, E1059–E1066 (1995).
336. R. Kanthan, A. Shuaib, G. Goplen, H. Miyashita, 'A New Method of In-vivo Microdialysis of the Human Brain', *J. Neurosci. Methods*, **60**, 151–155 (1995).
337. R. Kanthan, A. Shuaib, 'Clinical Evaluation of Extracellular Amino Acids in Severe Head Trauma by Intracerebral In Vivo Microdialysis', *J. Neurol. Neurosurg. Psychiatry*, **59**, 326–327 (1995).
338. L. Hillered, L. Persson, 'Parabanic Acid for Monitoring of Oxygen Radical Activity in the Injured Human Brain', *NeuroReport*, **6**, 1816–1820 (1995).
339. M.J. During, I. Fried, P. Leone, A. Katz, D.D. Spencer, 'Direct Measurement of Extracellular Lactate in the Human Hippocampus During Spontaneous Seizures', *J. Neurochem.*, **62**, 2356–2361 (1994).
340. A. Hamberger, B. Nystroem, S. Larsson, H. Silfvenius, C. Nordborg, 'Amino Acids in the Neuronal Microenvironment of Focal Human Epileptic Lesions', *Epilepsy Res.*, **9**, 32–43 (1991).
341. H. Carlson, E. Ronne-Engstroem, U. Ungerstedt, L. Hillered, 'Seizure Related Elevations of Extracellular Amino Acids in Human Focal Epilepsy', *Neurosci. Lett.*, **140**, 30–32 (1992).

Voltammetry In Vivo for Chemical Analysis of the Nervous System

Greg A. Gerhardt
Chandler Medical Center, University of Kentucky, Lexington, USA

Jason J. Burmeister
1238 Fillmore Street, Denver, USA

1 Introduction **2**
2 Electrodes **3**
2.1 Working Electrodes 4
2.2 Reference and Auxiliary Electrodes 6
3 Recording Methods **6**
3.1 Amperometry 7
3.2 Chronoamperometry 7
3.3 Cyclic Voltammetry 8
3.4 Linear Sweep Voltammetry 10
3.5 Rotating Disk Recordings 10
3.6 Differential Pulse Voltammetry 10
3.7 Differential Normal Pulse Voltammetry 11
4 Instrumentation **11**
4.1 Commercial 11
4.2 Laboratory Designed 11
5 Preparations **12**
5.1 Cells in Culture 12
5.2 Brain Slices/Tissue Pieces 12
5.3 Whole Animals 12
6 Applications **12**
6.1 Release of Neurotransmitters and Neurochemicals 13
6.2 Uptake/Reuptake and Diffusion 13
7 Quantification and Identification Issues **13**
7.1 Recording Methods 14
7.2 Electrode Coatings 14
8 Comparison with Other Methods **16**
8.1 Microdialysis 16
8.2 Fiber Optic Sensors 16
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **17**
References **17**

Since their introduction in the early 1970s, in vivo voltammetric methods have provided researchers with a means of studying the dynamics of neurotransmitter release mechanisms in whole animals and brain slices, and more recently in single cells in culture. Using microelectrodes (5–30 μm OD, outside diameter), generally made of carbon, investigators have studied the signaling dynamics of neurotransmitters and neuromodulators such as dopamine (DA), norepinephrine (NE) and serotonin (5-HT), and more recently nitric oxide (NO) and glutamate. Unlike the complimentary technique known as microdialysis, in vivo voltammetric methods allow for the very rapid (1–200 Hz) and spatially resolved (5–30 by 30–150 μm as compared to 200–300 μm by 1–4 mm) measurement of the dynamic properties of neurochemicals. However, the routine detection limits of such methods are in the 25–50-nM range for analytes such as DA, 5-HT and NO, and these do not rival the picomolar detection limits of methods used to analyze microdialysis samples.

These seemingly straightforward methods have been plagued by methodological problems, primarily dealing with the selective and sensitive measurements of neurotransmitters. Microelectrodes and associated recording techniques must be able to measure and identify selectively the analytes of interest without the use of powerful separation and quantitation methods such as high-performance liquid chromatography coupled to electrochemical detection (HPLC/EC). Improvements in microelectrode technologies and recording methods have greatly enhanced the utility of such methods for a variety of studies.

The purpose of this chapter is to give readers a general introduction to in vivo voltammetric methods that can be applied to studies of neurotransmitter and neuromodulator signaling dynamics in brain tissues and in cells in culture. We discuss the types of microelectrodes that are now used to record neurochemicals as well as surface modifications that are carried out on these microelectrodes to make them more sensitive and selective for different neurochemicals. The different recording methods and some of their inherent strengths and weaknesses are reviewed as well as the instrumentation currently used to record from microelectrodes for such measures. Finally, we discuss some of the applications of these methods and how they compare with other in vivo recording methods such as microdialysis and fiber optic sensors. This review does not focus on the biological systems that have been studied with these methods and the results of such studies. The major purpose is to give the reader an overview of in vivo voltammetry methods and how they can be used to study neurotransmitter and neuromodulator signaling dynamics in biological systems. Future developments in this field hold great promise for developing a large number of new microelectrodes for detecting a variety of neurochemicals in order to gain an understanding of chemical signaling,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

which is an integral part of central nervous system (CNS) biology and all biological systems.

1 INTRODUCTION

The concept of using in vivo voltammetric recordings for the analysis of neurotransmitters and neurochemicals, such as DA, in the CNS was pioneered by Ralph Adams in the early 1970s.[1] The basic concept of the methodology is seemingly straightforward and is illustrated schematically in Figure 1. A voltammetric microelectrode is implanted in the brain region that contains the nerve endings, varicosities or dendrites of interest. More recently, this approach has been used in slices of brain tissue and in cell culture. Voltammetric recordings are carried out using a recording method with a sensitive potentiostat connected to the microelectrodes to measure the neurotransmitter of interest. Thus, the release of a neurotransmitter can be measured and related to the function of the CNS. Unfortunately, there are a number of methodological considerations that have needed refinement in order to carry out these seemingly simple measures reliably. In particular, the type of recording electrode and the choice of voltammetric recording method are critical for the rapid, sensitive, selective and reliable detection of certain neurotransmitters and neurochemicals.

During the 1990s, the techniques of FSCV, FCV and HSC were refined for rapid in vivo electrochemical measurement of neurotransmitters in brain slices and in the brains of mice, rats and monkeys.[2–12] In contrast to microdialysis techniques, which permit measurement of basal extracellular levels and slow (ca. 1–20 min) changes in extracellular monoamines and metabolites, electrochemical recordings using 5–30-μm diameter microelectrodes allow for rapid (1–200 Hz) measurement of neurotransmitters and neuromodulators in discrete brain nuclei.[2–14] In addition, such methods can dynamically monitor neurotransmitter release and clearance kinetics.[3–7] Using rapid voltammetric recording methods our laboratory and others have measured DA, NE and 5-HT release and uptake.[2–8,12,15–24] Improvements in recording methods have also allowed for the further identification of the detected neurotransmitter species, which has greatly increased the reliability of such measures.[2,3,7–9,12–25] Moreover, numerous laboratories have been using amperometric recordings to measure release of neurotransmitters from individual cells in culture.[26–32] Amperometric methods with modified microelectrodes have also been used to measure NO.[33,34] Finally, future developments in microelectrode

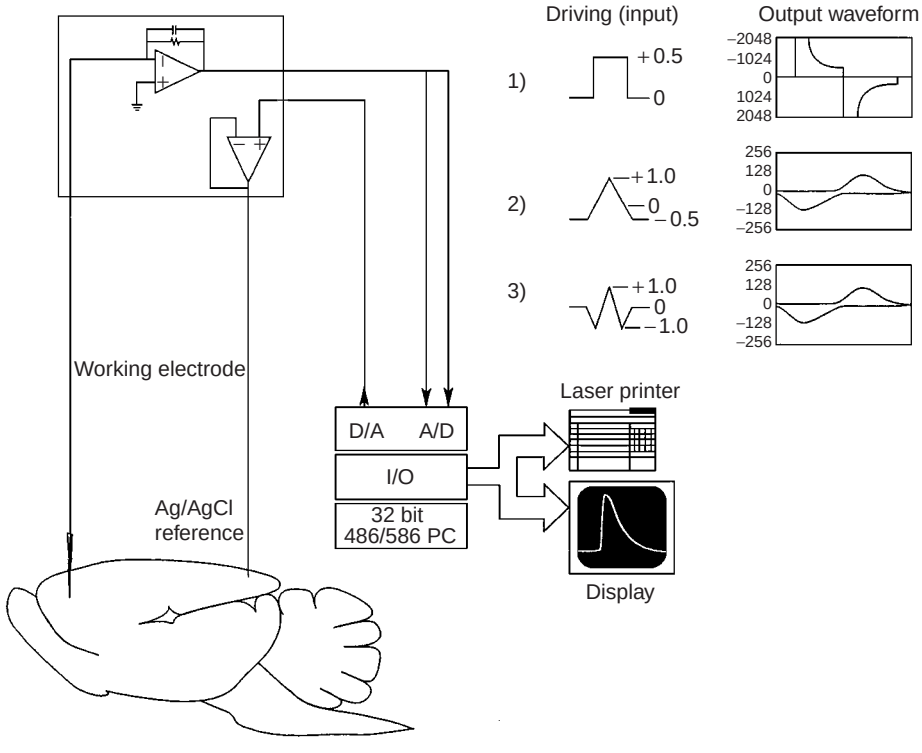

Figure 1 Schematic diagram illustrating the basic instrumentation needed to perform in vivo voltammetric recordings. Waveforms 1, 2, and 3 show the input potentials and output waveforms for high-speed chronoamperometry (HSC), fast-scan cyclic voltammetry (FSCV), and fast cyclic voltammetry (FCV), respectively.

technologies hold promise for routinely measuring other neurotransmitters and neurochemicals such as glutamate, lactate, glucose and γ-aminobutyric acid (GABA).[35–41]

This article is intended for those who want to gain a general knowledge of the principles, materials and techniques that underlie the current methodologies that are used to study CNS tissues and cells in culture using voltammetric methods coupled with microelectrodes. The current technologies that are used to measure neurotransmitters and neuromodulators, such as DA, NE, 5-HT, NO, glutamate and other neurochemicals, will be reviewed. We have endeavored to include mostly current data and references, but we have not tried to be comprehensive in our citation of all literature pertaining to this topic. There are additional reviews and review chapters that are recommended for those interested in the use of these methods.[2,12,25,42] We have undoubtedly omitted some important information, but we hope that this article will be of value to numerous scientists who want to use these methods to study the dynamics of neurotransmitters and neurochemicals in biological systems.

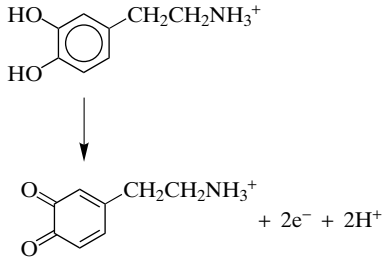

Scheme 1

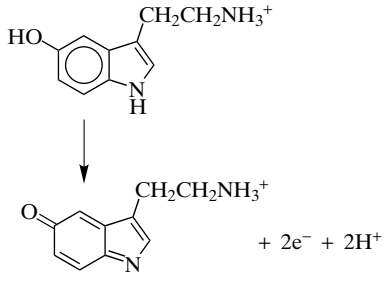

Scheme 2

2 ELECTRODES

As seen in Figure 1, the major tools that are used to carry out voltammetric recordings are microelectrodes that are implanted into CNS tissue or, in the case of single cell recordings, placed very close to the cells. The recording or working electrode is the electrode where the detection of the neurotransmitters or neurochemicals occurs. This is perhaps the single most important component of the recording system. A variety of different types of microelectrode, ranging in size from 5 to 200 μm in diameter, are used for such studies (see below). Reference and auxiliary electrodes are needed to carry out the measures without a drift in the potential that is applied to the working electrode.[43] A basic understanding of electrochemistry at solid electrodes can be obtained from Bard and Faulkner[43] or Adams.[44] Basically, as a function of the voltammetric method and the magnitude of the potentials applied to the electrode surface (see section 3), electrochemical reactions occur at the surface of the microelectrode. The basic oxidation reaction for DA is depicted in Scheme 1.

DA can be oxidized to form an *o*-quinone that under certain circumstances can be reduced back to DA. The reaction is a two-electron process. The oxidation reaction for 5-HT is rather different and again involves a two-electron reaction as seen in Scheme 2.

These reactions can occur at the surface of the working electrodes provided the voltammetric method (section 3) produces potentials at the microelectrode surface that are sufficiently large to cause the oxidation of these compounds and if the analyte is adjacent to the surface of the microelectrode. The current flow caused by these chemical reactions or faradaic current is proportional to the concentration of the analyte. While one equation cannot be used to describe the current response of all voltammetric recordings as they are related to the recording method (see section 3 and Bard and Faulkner[43]), the basic response of a large electrode at potentials > 100 mV beyond the oxidation potential for a given analyte can be described in Equation (1)

$$i = nFAC_bM \tag{1}$$

where i is the current measured, n is the number of electrons for a given chemical reaction, F is the Faraday constant, A is the total active recording surface area of the electrode, C_b is the bulk solution concentration of the analyte and M is a general purpose term encompassing mass transport of analyte to the electrode surface. In contrast to large electrodes, microelectrodes have both linear and spherical diffusion of analyte to the microelectrode tip and thus, if the response of the electrode is described accurately, an additional set of terms must be applied to model the electrode response completely.[45] However, to understand the basic concepts of the measurements, the current recorded is directly proportional to the concentration of the analyte or neurotransmitter (C_b), the active recording area of the probe (A), the number of electrons per molecule (n) and the mass transport of the analyte to the surface of the microelectrode. In an unstirred solution such as the extracellular space of the brain or in a quiet

solution next to a cell, a substitution can be made to Equation (1) to give Equation (2)

$$i = XC_b \quad (2)$$

where X, the calibration factor for a given microelectrode, can be determined pre- and/or postexperimentation by varying C_b and measuring i under solution conditions that mimic the extracellular space of the brain (artificial cerebrospinal fluid) or the media in which cells are studied. Thus, a hallmark of many voltammetric recording methods is that the electrodes respond linearly to changes in concentration of the analyte. This forms the analytical foundation for the use of these methods in biological systems.

Using the aforementioned terms, the i recorded from oxidation or reduction of an analyte, such as DA, is more accurately depicted for chronoamperometry (CA) using a modified form of the Cottrell equation seen in Equation (3).[45]

$$i = \frac{nFAC_bD^{1/2}}{\pi^{1/2}t^{1/2}} + 4nFr\pi C_bD \quad (3)$$

where t is the time of the measurement of i in seconds, D is the diffusion coefficient of the analyte, and r is the radius of the microelectrode.

The selectivity of recordings for the measurement of neurotransmitters such as DA in biological systems is a major issue. Modifications to microelectrode surfaces (see section 7) and the use of different recording methods (see section 3) have enhanced the selectivity of the voltammetric methods. In addition, selection of applied potentials can minimize interferents. However, voltammetric methods are inherently selective, as very few molecules in the extracellular space of the brain can be caused to oxidize or reduce. This is both an inherent strength and weakness of these methods. Although not comprehensive, Table 1 summarizes some endogenous and exogenous compounds that can be measured using voltammetric recording techniques using low or high oxidation potentials.[46,47] We urge caution because the actual oxidation potentials of these compounds are dependent on the recording electrode used, the recording method and any modifications of the microelectrode surface.

There are a variety of working, reference and auxiliary electrodes that are used to record neurotransmitters and neurochemicals. In the following sections we have tried to describe the basic compositions of the different electrodes and the advantages of each type.

2.1 Working Electrodes

The working or recording electrode used for in vivo voltammetric measurements in CNS tissues is generally fabricated to achieve a tip diameter of 5–200 μm, depending on the application. There are a number of types, and the choice of the size and type is often based on the molecule that will be measured and the recording preparation.

Table 1 Endogenous electroactive compounds and exogenous drug compounds that are electroactive, which are encountered in neurochemical studies[a,46,47]

Easy to oxidize	Harder to oxidize
DA	Melatonin
NE	Metanephrine
5-HT	Normetanephrine
Epinephrine	MHPG
DOPAC	5-HTP
L-DOPA	Tyramine
AA	Octopamine
5-HIAA	Tryptamine
Uric acid	Tyrosine
DHPG	Tryptophan
	Hydrogen peroxide
	NO
	3-MT
	HVA
	Neuroleptics: Clozapine, chlorpromazine
	Neuropeptides: Neurotensin, enkephalin
	Antidepressants: Imipramine, desipramine, mianserin
Catecholamine-related drugs: α-Methyldopa, α-Methyl DA, α-Methyl NE, Carbidopa	MAO inhibitors: Tranylcypromine, pargyline, α-methyl tryptamine
Other Drugs: Acetaminophen	β-Blockers: Sotalol, timolol, atenolol

[a] Compounds in bold are of most interest.
DOPAC, 3,4-dihydroxyphenylacetic acid; DOPA, 3-(3,4-dihydroxyphenyl)alanine; AA, ascorbic acid; 5-HIAA, 5-hydroxyindoleacetic acid; DHPG, 3,4-dihydroxyphenylglycol; MHPG, 3-methoxy-4-hydroxyphenylglycol; 5-HTP, 5-hydroxytryptophan; 3-MT, 3-methoxytyramine; HVA, homovanillic acid; MAO, monoamine oxidase.

2.1.1 Carbon Paste

Carbon paste electrodes are one of the oldest forms of electrodes used for in vivo electrochemical recordings in CNS tissues. They are formed by packing carbon paste into the end of a Teflon®-coated wire[48] or by using a mixture of carbon epoxy that is packed into a glass capillary.[45] Carbon paste is prepared by mixing powdered graphite with a hydrophobic material such as liquid paraffin, silicone oil, or wax. Stearic acid is sometimes added to the paste to improve the electrode's recording properties.[48] Graphite pastes can be mixed

with epoxy to form a graphite–epoxy mixture[45,46] or used as they are to make the Teflon® wire-type electrodes. A commercially available graphite–epoxy mixture is used for some types of graphite–epoxy electrodes (PX-grade Graphpoxy, Dylon, Industries, Cleveland, OH). The resulting electrodes have tip diameters of 100–250 μm, depending on the resulting diameter of the graphite–epoxy filled glass capillary or the diameter of the Teflon®-coated wire. While useful for certain applications, the size of these electrodes precludes their use for many applications.

2.1.2 Carbon Fiber Electrodes

Working electrodes made from carbon fiber or carbon monofilament are the most common and widely used forms of electrodes for CNS recordings. Basically, all of these electrodes are generally made by sealing one or more carbon fibers or carbon monofilaments in glass capillaries.[8,10,12,49,50] For certain single cell recordings, the fiber is sealed in plastic pipette tips.[51,52] Appropriate electrical connections to the fibers are made using silver pastes,[53] graphite pastes[8,10,54] or salt solutions.[55] There are numerous manufacturers of carbon fiber or carbon monofilament materials that range in size from 5 to 30 μm in diameter. Typical manufacturers and part numbers for popular fibers and one monofilament are as follows: 10-μm diameter (P-55, Amoco, Greenville, SC),[55] 7–8-μm diameter (type HM or XA-S, unsized, Courtaulds),[6,19] 8–12-μm diameter (SOFICAR, Paris France),[17] 30-μm carbon monofilament (#80036-001, Textron Systems Division, Wilmington, MA).[8,10] In addition, a number of companies sell prefabricated carbon fiber-based electrodes. Some of these are Quanteon, LLC (Lexington, KY), World Precision Instruments (Sarasota, FL), Dagan Corp. (Minneapolis, MN), ALA Scientific (New York, NY), MPB Electrodes (London, UK) and Axon Instruments (Foster City, CA).

Although often similar in size, the recording properties of different fibers can be quite different. This is attributed to the type of fiber and the methods and/or cleaning steps used to prepare the fibers. The degree of oxygen on the surface of the carbon will affect the recording properties of the fibers.[56] It is a good idea to use carbon fibers and carbon monofilaments that have been previously characterized. As stated, the fabrication steps are quite similar for different carbon fiber electrodes (CFEs); however, the choices of glass, connecting wire, cements and other materials can be quite diverse. Typical fabrication methods for a variety of fiber electrodes are given by various authors.[12,42,49,50]

One type of carbon monofilament-based electrode used for studies in our laboratory is constructed by cementing single ''as pyrolyzed'' 30-μm diameter single carbon monofilaments (#800036-001; Textron Systems Division, Wilmington, MA) inside a pulled glass capillary. A single glass capillary with a tip diameter of approximately 5 μm is pulled from a 2-cm length of (4.0-mm OD, 2.6-mm ID (inside diameter)) borosilicate glass tubing (Glass Warehouse, #26005) using a Narishige PE-2 puller. The tip of the pulled glass capillary is bumped to an inner diameter of 40–45 μm. A single carbon fiber (5 cm in length) is then passed completely through so that the fiber extends well beyond the tubing at both ends. A small drop of Epoxylite™ (The Epoxylite™ Corp., Westerville, OH) is placed on the carbon fiber at the exposed end of the glass capillary, and the fiber is then pulled into the tubing so that it is sealed well in the tubing with no excess Epoxylite™ on the exposed carbon fiber. This carbon fiber/glass capillary assembly is cured overnight in an oven set at 100 °C to dry the Epoxylite™. The open end of the larger tubing is then packed with a graphite–epoxy mixture (Graphpoxy™, Dylon Industries, Berea, OH) by smearing a layer of Graphpoxy™ on a microscope slide and tapping the back end of the electrode repeatedly through the Graphpoxy™ to the glass slide. It is generally not possible to fill the entire void with Graphpoxy™, but it is possible to see at least 3–4 mm of packed material through the side of the larger tubing. The tapping procedure results in breaking off the back end of the graphite fiber flush with the end of the larger tubing. The excess glass is cut off, exposing a glass capillary filled with Graphpoxy™ with a maximum OD of 500–750 μm. A 10-cm length of 36-gauge lacquer-coated copper wire is prepared by scraping 5 mm of coating from one end of the wire. The wire is then inserted into the back end of the larger tubing, through the Graphpoxy™. This results in a good electrical connection between the carbon fiber and the copper wire. Finally, a small drop of Epoxylite™ is placed on the copper wire at the point of attachment to the larger piece of glass tubing and the microelectrode is again baked overnight at 100 °C to cure the Graphpoxy™.

Once sealed in glass, the carbon monofilament or CFEs can be configured for a variety of applications. They are either cut and sometimes polished to form an oval disk electrode[57] or trimmed to form a short cylinder (30–150 μm in length), which is perhaps the most popular form of CFE. A scanning electron micrograph of the tip of a 30-μm diameter cylindrical tip carbon monofilament electrode is shown in Figure 2. In addition to the extended cylinder, the tips of the electrodes can be etched or the electrodes can be electrically pretreated to enhance their recording properties.[17]

2.1.3 Carbon Arrays

A novel type of array electrode that holds promise for future applications is manufactured using semiconductor

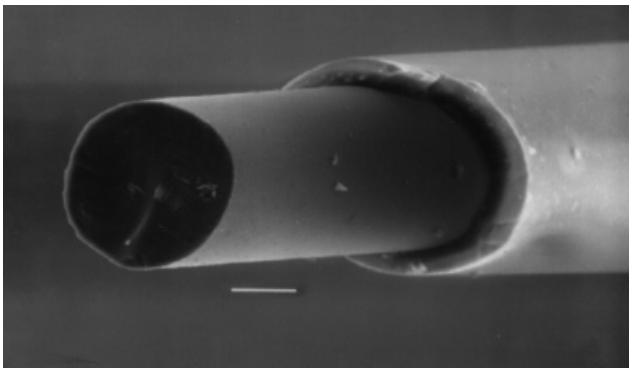

Figure 2 Scanning electron micrograph of the tip (30-μm OD, cylindrical tip) of a single carbon monofilament microelectrode. The scale bar is 10 μm.

fabrication techniques on substrates of silicon. The potential advantages of array electrodes include: (1) reproducible recording sites, (2) the ability to record from multiple brain sites from a single probe, (3) spatially defined recording sites for recording from layered brain structures, (4) the ability to stimulate and record single unit or chemical activity electrically from the same recording sites, and (5) reduced cost compared to handmade electrodes. These electrodes have been characterized[58] and used in vivo.[59] Briefly, these electrodes are fabricated using photolithography to pattern recording sites, connecting lines and bonding pads on to an insulated silicon surface. The metalized connecting lines are insulated and then the recording sites are sputter-coated with carbon. At present, the recording characteristics of the carbon-sputtered multisite probes are worse than single carbon fiber microelectrodes.[58,59]

2.1.4 Noble Metals

Platinum (Pt) or platinum–iridium (Pt–Ir) wires, often sealed in glass, form the foundation of a number of electrodes that are being developed to measure other neurotransmitters and neurochemicals by using oxidative enzymes attached to electrode surfaces.[35–41,60,61] These microelectrodes can be made ranging from 5 to 250 μm in diameter. In general, Pt and Pt–Ir are not ideal surfaces for measuring molecules like DA, NE and 5-HT. However, they are excellent for measuring peroxide formed from oxidative enzymes or enzyme cascades. A particularly unique type of recording electrode has been developed that is actually a combination of a microdialysis probe and a platinum electrode and has a 230-μm OD. These electrodes are called dialysis electrodes and they are sold by Sycopel International Ltd (Tyne and Wear, UK).[62] They are very versatile, as they can be filled with a variety of oxidative enzymes to make them sensitive to a number of analytes, such as glutamate and lactate.

2.2 Reference and Auxiliary Electrodes

A stable reference is an important element of the electrochemical measurement because the potential of the working or recording electrode is changed relative to a stable reference electrode. By contrast, many laboratories do not use an auxiliary electrode, because of the small currents and low impedance of the recording solutions. However, some instruments still work in a three-electrode (auxiliary, reference, working) versus two-electrode (reference, working) mode. Generally miniature forms of reference electrodes and auxiliary electrodes are used so that they can be directly implanted into CNS tissue. However, for brain slice and single cell recordings, standard size reference electrodes are often employed.

2.2.1 Ag/AgCl Reference Electrodes

The miniature Ag/AgCl electrode is the most common reference electrode used for voltammetric recordings in CNS tissues.[4,6,10,15,20] These can be made by using a short length (4–6 cm) of Teflon®-coated (250-μm OD) silver wire. The Teflon® is removed from 3–5 mm on one end. A voltage of +0.5–+1 V is applied to the Ag wire that is submerged in 1-M HCl for about 5–10 min using a platinum wire as the cathode. The Ag/AgCl reference should be a silver gray color. A standard size Ag/AgCl reference that is used for slice and single cell recordings is the RE-5 (BAS, Bioanalytical Systems, Lafayette, IN).

2.2.2 $Hg/HgCl_2$ Sodium Saturated Calomel Electrodes

The sodium saturated calomel electrode (SSCE) is a less popular reference, but it is still used.[12] Mercury is in contact with a solution that is saturated with mercury(I) chloride and NaCl. These can be made in-house or purchased from private manufacturers.

2.2.3 Auxiliary Electrodes

As previously stated, auxiliary electrodes are not always used for voltammetric recordings with microelectrodes. A standard auxiliary electrode is a Teflon®-coated silver wire (250 μm diameter) that has 3–5 mm of the Teflon® removed from one end. This is placed in contact with brain tissue or cell supernatant, or for whole animal recordings the Ag wire is connected to the skull via a screw that is in contact with the dura. Alternatively, a stainless steel wire in contact with CNS tissue or the perfusion medium is used.[24]

3 RECORDING METHODS

The choice of recording method is perhaps the single most confusing aspect of voltammetric recordings in CNS

tissues and cells. In this section we have tried to explain the different recording methods and how they have been used to study neurotransmitters and neurochemicals in CNS tissues and single cells. The methods are all based on the way the potential is varied at the working electrode to record the oxidation or reduction of analytes, such as DA. Basically, as depicted in the inset to Figure 1, recording methods usually involve applying voltages to the working electrodes using square waveforms, triangle waveforms or a combination of triangles and squares. In addition, the voltage can be constantly applied for certain applications. The choice of the recording method for a given study will depend on the temporal resolution, potential confounding substances, the size of the electrode needed and the sensitivity required for the measurements. In addition, the availability of instrumentation to perform the different methods and personal preferences must also be considered.

3.1 Amperometry

Amperometry involves the measurement of current at a constant fixed potential and is the simplest of the electrochemical techniques. The current can be monitored continuously therefore this technique has the ability to record events that occur very fast (200–1000 Hz). In addition, since the voltage is applied to the working electrode constantly, the nonfaradaic background current recorded from the electrode is very low, allowing for very sensitive measurements. Amperometry is used with HPLC/EC systems, which accounts, in part, for the high sensitivity of these methods. Amperometric recordings, because of the high sampling resolution, are used to study catecholamines released from single cells.[27–32,63] In addition, amperometry is useful for measuring compounds such as hydrogen peroxide, NO, and ascorbate. Hydrogen peroxide is of interest because it is the product of oxidase enzymes such as glutamate oxidase.[35–40] Enzyme modified electrodes will be discussed later.

The danger of fixing the potential is that analytes and analyte oxidation products readily adhere to the electrode. This poisoning of the electrode changes its recording properties. Also, when using a fixed potential the source of the analytical signal cannot be determined because any species that is electroactive at the potential applied to the working electrode will contribute to the current (see Table 1). Cahill and Wightman overcame this problem by the use of two electrodes to differentiate between ascorbate and catecholamines.[64] The first electrochemically modified carbon fiber potential was set at +0.05 V versus SSCE. This electrode was selective for ascorbate. A second unmodified CFE was set at +0.65 V and detected ascorbate and catecholamines. Currents obtained from both electrodes were used to determine the species being released from bovine adrenal medullary cells. Thus, a combination of electrodes can be used to overcome some of the limitations of amperometry. However, in general, amperometric recordings are used to study very rapid (<1 s) release of catecholamines from single cells and to carry out studies of NO and measures involving oxidase enzyme-based electrodes in CNS tissues.

3.2 Chronoamperometry

CA is one of the oldest methods used for in vivo electrochemical measurements in CNS tissue and is based on the work of Cottrell, as discussed by Bard and Faulkner.[43] An applied potential is instantaneously stepped from a resting value to a higher potential that is usually 50–100 mV in excess of the oxidation potential of an analyte. The potential is held at this higher value for ca. 1 s then it is stepped back to the resting potential. Current is measured after the potential has been stepped, during the time that the nonfaradaic background current of the electrode has declined (ca. 100 ms). Current increases dramatically as the potential is stepped from its resting value. This is caused by the charging current of the electrode and the electroactive species that is being oxidized. The charging current decays with time as the potential is held at its higher value (see Waveform 1 in Figure 1). An advantage of this method is that the current recorded from the analyte is directly proportional to the concentration of the analyte and therefore the current can be summed or averaged over most of the measurement, once the charging current has declined for ca. 5–20 ms. Another advantage of this method is that the background charging current of the electrode is a minor component of the signal, which enhances the signal-to-noise of the recording method. These square wave potential steps have been repeated at 60-s intervals to monitor events that occur as a function of time using stearate-modified carbon paste electrodes. Typical resting and applied potentials were −0.15 V and +0.25 V versus Ag/AgCl reference, respectively when using stearate-modified carbon paste electrodes to measure extracellular DA.[65–67]

3.2.1 High-speed Chronoamperometry

High-speed chronoamperometric recordings are a rapid version of the aforementioned methods originally described by Cottrell.[43] They have been used to measure DA, NE, 5-HT and NO in a variety of preparations. These methods are advantageous because they are rapid, sensitive and record data that can be used further to identify the detected compound(s).

A carbon fiber microelectrode is held at a resting potential of −0.20–0.0 V versus Ag/AgCl reference; electrochemical recordings are made in a suitable buffer

solution to simulate CNS tissue. The voltage is then instantaneously stepped to an oxidation voltage of +0.45–+0.55 V. This voltage step is chosen to exceed the oxidation potential of the potential analyte, in this case the biogenic amine neurotransmitters such as DA and 5-HT, by at least +0.10 V. This produces a rapid change in the current recorded by the microelectrode that rapidly decays as a function of time. This charging current is allowed to decay for 4–20 ms before the background current of the microelectrode is measured. The square-wave voltage step is held at the oxidation voltage for a fixed time period of 20, 50 or 100 ms, during which the oxidation current is sampled for 15, 40 or 80 ms. The square-wave voltage step is then instantaneously returned to the resting potential. Similar to the case of the first oxidation step, the return of the potential to the resting voltage produces a reduction reaction at the surface of the electrode. A rapid decrease in the reduction current represents the charging current of the recording and this is allowed to decay for 4–20 ms. The resting voltage is maintained for a fixed period that is identical to the prior square-wave step of 20, 50 or 100 ms. Like the oxidation current step, the current is sampled for a fixed time period of 15, 40 or 80 ms. In our studies, a 50-kHz sampling rate is employed to measure the oxidation and reduction current portions of the signals. As the current is integrated, the correct term for this type of recording is high-speed chronocoulometry.[8] The symmetric voltage steps are continued at a set time period to produce signals with a sampling resolution of 5 Hz (100-ms oxidation, 100-ms reduction; 5 per second), 10 Hz (50-ms oxidation, 50-ms reduction; 10 per second) or 25 Hz (20-ms oxidation, 20-ms reduction; 25 per second). In theory, the signal-to-noise ratio of such recordings is greatly enhanced because small changes in current due to the detection of an analyte are not masked by large background current. In addition, potential changes in the background current of an electrode that do not relate to increases or decreases in an analyte are potentially less prone to produce recording artifacts, owing to the lower background current levels. Typical detection limits for DA and 5-HT are 25 nmol using a 30-μm diameter carbon cylinder electrode.[68–70]

In a similar way to FSCV methods (see below), HSC recordings yield information regarding the major contributor to the electrochemical signal.[8,68–70] The detection of DA yields reduction to oxidation (red/ox) current ratios with Nafion™-coated electrodes between 0.5 and 0.9 depending on the thickness of the Nafion™ film and the drying temperature of the Nafion™-coating procedure.[8,10,68–70] In contrast, AA produces red/ox ratios of 0.0 and 5-HT produces ratios of 0.0 to 0.2.[8,15,18,68–70] Thus, somewhat analogous to the waveforms seen with FSCV (see below),[7,12] DA, AA and 5-HT produce distinctive red/ox ratios which can be used further to identify the detection of these compounds in brain tissue.

Many recordings in CNS tissues and other applications do not require temporal resolutions of 5–25 Hz. Therefore, for more standard recordings, the signals are averaged over a 1- or 2-s time period.[8,68–70] This further enhances the signal-to-noise ratio of the recording methods and decreases the number of data points that must be stored by a microcomputer recording system.

3.2.2 Double Potential Step

Double potential step CA or differential pulse amperometry (DPA) involves stepping the potential from a resting potential to a higher potential that oxidizes easily oxidizable, interfering species such as ascorbate. The current measured at this time is a result of the background current and any easily oxidized interfering species. The potential is again stepped to the final potential. The current measured at this point is caused by the background, interfering species and analyte. The difference in current between the prepulse and final potentials is recorded. In this way, the contribution due to the background and interfering species can be minimized. The main drawback of DPA is the lower temporal resolution compared to FCV or HSC (scan rate = 1 Hz). Using DPA, DA can be measured with untreated CFEs using a resting potential of −240 mV and a final potential of +200 mV. The measuring pulse duration was 400 ms with a measuring pulse amplitude and prepulse duration of 50 mV and 100 ms, respectively.[17,70] The electrodes were 50 times more sensitive to DA than AA and the detection limit was 0.5 μM. Electrochemical pretreatment of the electrodes was seen to lower the detection limit to 5 nM at the expense of a great decrease in response time (10–30 s).

3.3 Cyclic Voltammetry

In cyclic voltammetry (CV) the current is measured as the potential is steadily ramped up from a starting potential to the switching potential where it is ramped down to the starting potential (see the triangle wave 2 in Figure 1). The scan rate is typically between 10 and 300 mV s^{-1}. CV is classically used as a qualitative technique for determining red/ox potentials and diffusion coefficients of molecules. Typically the starting potential is lower than the oxidation potential of the analyte of interest (ca. −0.5–0.0 V). In an unstirred solution, as the potential becomes more positive the recorded current becomes more negative as the electroactive species begins to be oxidized. If the electrode is large enough a peak will be seen at the oxidation potential due to depletion of the electroactive species near the electrode surface. If the electrode is small, a plateau will be seen because spherical

diffusion will be sufficient to bring fresh molecules to the electrode surface to be oxidized. After the potential ramp is reversed, the current begins to rise as the oxidized species that has not diffused away from the electrode begins to be reduced. If the electrode is sufficiently small, no reduction peak will be seen because all of the reduced species will have diffused away from the electrode surface. While used as a tool to characterize working electrodes, normal CV is rarely used for in vivo electrochemical recordings due to its slow recording rates. For a complete discussion of CV refer to Bard and Faulkner.[43]

3.3.1 Fast-scan Cyclic Voltammetry (Triangular)

FSCV uses the same triangular-shaped ramps in potential as normal CV but at very high scan rates. This is one of the most popular in vivo voltammetric methods due to its excellent sensitivity, time resolution and compound identification capabilities.[3,7,12,14,16] Typical resting and switching potentials are −0.4 and +1.0 V, respectively, for DA. With FSCV, much higher scan rates are used compared with CV (300–900 V s^{-1}). The high scan rates are coupled to 10-Hz repetition rates to track rapid biological events. However, high scan rates cause large background currents, which can increase the limit of detection (LOD). To obtain an analytically useful signal, the voltammograms are background-subtracted to make the analyte peaks (faradaic current) apparent. A background voltammogram or series of voltammograms is selected and subtracted from subsequent scans during a recording session.

Repetitive CVs, often 10 Hz, are sampled to obtain time-dependent concentration profiles. The current is averaged over a desired potential range corresponding to oxidation (or reduction) of the analyte. The interval between successive scans is at least 5–10 times the duration of each individual scan to allow the diffusion layer to return to its prescan state. Typically analytes are sampled up to approximately 8 μm from the electrode surface.[12]

The voltage ranges of the oxidation and reduction peaks on the voltammograms or the shapes of the voltammograms can be used to identify compounds that are responsible for a given signal. As with HSC, the ratio of the oxidation and reduction peaks can be used to help identify the analyte. In addition the number of peaks can be used to identify the source of the sensor response. Although their oxidation peaks were overlapping (+0.53 V), DA was distinguished from 5-HT because DA had a single reduction peak at −0.18 V but 5-HT had two reduction peaks at −0.05 and −0.56 V. A dual wave triangular potential was used with resting and switching potentials of −0.7 and +1.3 V, respectively. The scan rate was 813 V s^{-1} and the triangular waveform was repeated at 4 Hz.[71,72] In another recent report, three-dimensional plots of voltage versus time with current plotted in false color were used to visualize the source of the sensor response. This was useful for distinguishing between DA, 5-HT, and pH changes during FSCV recordings.[73]

Measurements of DA using FSCV have been shown to be dependent upon divalent cation concentrations such as Ca^{2+} and Mg^{2+} and changes in pH.[12,74] Also, the oxidation of DA with double potential step CA and CV yields higher currents than would be expected from a species that was free to diffuse away from the electrode following oxidation.[75] These studies support the hypothesis that oxidation of DA at CFEs involves significant adsorption of the analyte. Unlike FSCV, HSC does not seem to be affected by changes in divalent cations and no effects from pH changes have been reported.[8,68–70]

DA has been most extensively measured using FSCV. In addition NE and 5-HT have been quantified.[3,12,14,16,76] Recordings for NE are inherently less sensitive than those for DA.[12] Also NE gives a similar waveform signature to DA. 5-HT is readily discerned from NE, DA and other neurochemicals. In contrast to DA where the scan rate was 300 V s^{-1} and the resting and switching potentials were −0.4 and 1.0 V, respectively, the potential was scanned from 0.2 to 1.0 to −0.1 to 0.2 V with a scan rate of 1000 mV s^{-1}.[3,76] By using this alternate waveform and scan rate the follow-up reaction where surface adsorption occurs was minimized. Nafion™-coated electrodes were 20 times more sensitive to 5-HT than DA. Typical detection limits for DA and 5-HT are 25–50 nM with these methods. Finally, histamine and oxygen can be quantified using FSCV.[12,77]

3.3.2 Fast Cyclic Voltammetry ($1\frac{1}{2}$ Triangle)

Another approach involving CV recordings uses a triphasic ramp voltage as the waveform (see Waveform 3 in Figure 1).[20,22–24,78] This technique has been called $1\frac{1}{2}$ triangle or FCV. With this "W"-shaped waveform the potential decreases from the resting potential to a lower potential. The potential then increases to the switching potential, which is higher than the resting potential. The potential is finally ramped down to the lower potential then back to the resting potential. Each scan typically lasts 15–20 ms and the scan rate is 300 V s^{-1} or higher. The frequency of scans is usually 1 Hz. The resting, lower, and upper potentials are 0, −1, and +1 V, respectively. Using this approach, DA has an oxidation peak at +0.6 V and a reduction peak at −0.2 V at certain carbon fiber microelectrodes.[79] With this waveform the detection limit of such electrodes has been reported to be about 200 nM. Increased sensitivity for DA has been shown when the upper potential was increased to +1.4 V, which was attributed to increased analyte adsorption and activation of the carbon surface.[80]

3.3.3 Sine-wave Cyclic Voltammetry

Another CV method involves the use of sine wave driving voltages. Sine wave-based CV techniques are those that involve using a nonlinear scanning input voltage instead of the linear ramps and pulses used in other techniques. The slopes of the voltage ramps are constantly changing and the potential is essentially held at the switching potentials instead of being rapidly switched. The first example is continuous scan cyclic voltammetry (CSCV). CSCV was developed to increase the temporal resolution of voltammetry. The input voltage was a 100-Hz sine wave with a peak potential of +1.0 V and a minimum potential of −0.5 V. Background subtraction was again used and the current was sampled every 10 ms. An advantage of CSCV over FCV is that fast scan rates can be used without losing sensitivity. Sample and hold circuits were used and set at +0.59 and −0.27 V for measuring DA. Some microelectrodes were able to detect DA with a detection limit of 10 nM.[81] Interestingly, investigators have not taken advantage of the AC signal properties of these recordings to increase sensitivity or to further identify analytes.

The second example involves a triphasic sine wave like that used with FCV. The triphasic sine wave (2 Hz) was used to overcome the artifact seen when the polarity of a linear potential ramp is switched. The resting, upper and lower potentials were 0, −1, and +1.3 V, respectively. The potential followed a sinusoidal wave down to the lower potential where it climbed to the upper potential. The sine wave went to the lower potential then back to resting potential. DA yielded single oxidation (+0.6 V) and reduction peaks while NO showed a single oxidation (+1.2 V) and two reduction peaks. DA and NO were quantified simultaneously by following the current at their respective oxidation potentials with little cross-talk. The reduction peaks were used to identify the analytes.[82] The sine-wave techniques have the advantage of reducing instrumental artifacts caused by the rapid switching of potentials when using high scan rates with the triangle waveforms.

3.4 Linear Sweep Voltammetry

In linear sweep voltammetry (LSV) current is measured while the voltage is steadily ramped, usually at slow rates of 10–20 mV s^{-1}. Using some electrodes, a peak is observed for an electroactive species in unstirred solutions caused by depletion of the species near the electrode. The electroactive species is only partially replenished by diffusion. The linear sweep is applied at regular intervals (ca. 5–10 min) and is not a fast recording technique. A semiderivative circuit is used to improve the resolution between peaks for LSV. Although LSV and semiderivative LSV[48,83] have been used to track in vivo changes in neurotransmitters, the speed of the technique limits its utility for many applications.

3.5 Rotating Disk Recordings

Rotating disk electrode voltammetry (RDEV) has been applied to measuring uptake and release of neurotransmitters, such as DA and NE, in brain tissue minces and cells in culture. RDEV is unique because larger glassy carbon electrodes (3 mm) are used and the electrode actually stirs the sample solution. A constant potential is applied to the electrode which is usually >100 mV past the peak oxidation potential of the analyte of interest. The analyte is brought to the working electrode surface by a constant stirring motion produced by spinning the electrode at a constant rate. The Levich equation can be used to describe the response of the electrode.[43] These techniques have been applied to studying the kinetics of neurotransmitter uptake in tissue pieces and cell suspensions.[84,85]

Using these methods, DA was measured as the limiting current recorded at a glassy carbon RDE (rotating disk electrode) at an applied potential of +0.45 V versus Ag/AgCl reference. The rotation was 2000 rpm.[84] Both NE and DA have been measured in cell suspensions using RDEV.[85] In another study, a two-electrode system was employed to measure two analytes. DA and *m*-tyramine transport in cell suspensions was investigated using two RDEs. The potential of the first electrode was held at +0.4 V versus Ag/AgCl reference. Current measurements at this electrode corresponded to the oxidization of DA alone. The second electrode's potential was set at +0.65 V to measure both DA and *m*-tyramine, with both electrodes rotating at 4000 rpm.[86]

3.6 Differential Pulse Voltammetry

In differential pulse voltammetry (DPV), there is a linear ramp in potential with superimposed steps in potential that happen at regular intervals.[17,43] The current is measured for short time periods before and at the end of each pulse. The difference between these is plotted versus potential to yield voltammograms with pronounced peak shapes. The power of this technique is in resolving between compounds with similar oxidation potentials. The drawback of DPV is the temporal resolution: about one scan per minute.

DPV has been used to measure DOPAC in both the nucleus accumbens and striatum in the rat CNS with two electrically pretreated pyrolitic 12-μm CFEs.[87] One scan was completed every 5 min, as the potential range was −0.2–+0.35 V with a scan rate of 5 mV s^{-1}. The pulse amplitude was 50 mV with a frequency of 2 Hz. Individual oxidation peaks were observed for AA, catecholamines, indolamines, and uric acid. The separation of different

compounds is a clear advantage of this methodology. However, its slow sampling rate precludes the use of DPV for many applications.

3.7 Differential Normal Pulse Voltammetry

Differential normal pulse voltammetry (DNPV) involves stepping the potential from a resting potential to a higher potential then stepping the potential to a slightly higher potential. The potential is returned to the resting potential and the previous potential steps are repeated while sequentially increasing the step size. The difference between the prepulse and final potential values is recorded. The advantage of this technique over DPV is that the potential returns to the resting potential between measurements.[88]

DNPV has been used to measure DOPAC.[89] A scan was completed every 2 min from −0.26 to +0.16 V versus Ag/AgCl reference. The scan rate was 3–4 mV/0.4 ms with a prepulse duration of 70–80 ms. The pulse duration was 40 ms and pulse amplitude was 30 mV. The detection limit for DOPAC was 1 μM at a potential of +0.055 V. Dihydroxyphenylethylene glycol (DOPEG) was detected at +0.08 V (LOD = 0.3 μM).

4 INSTRUMENTATION

The measurement of neurotransmitters and neurochemicals with microelectrodes requires specialized instruments capable of making often fast, repetitive (1–1000 Hz) and low-noise current measurements. Typical currents are in the low picoampere to mid-nanoampere range. There are no commercial instruments available that are capable of performing all of the aforementioned methods. We have tried to identify instrumentation that has the capability to carry out many of the different electrochemical recording methods.

4.1 Commercial

Several companies sell very versatile potentiostats that are often controlled by Windows-based software. Some of these are: Bioanalytical Systems – BAS 100B/W and CV-50W (West Lafayette, IN), Radiometer Analytical S.A. – Volta Lab 40 (Lyon, France), EG&G Princeton Applied Research – Model 394 (Wellesley, MA), and Cypress Systems, Inc. – CS-1200 (Lawrence, KS). These are all computer-controlled electrochemical analyzers capable of performing amperometry, CV, LSV, CA and a variety of other electrochemical methods. However, these instruments are more suited to microelectrode characterization studies, as they cannot perform many of the fast cyclic waveforms and currently they are not capable of analyzing and storing repetitive measures at 1–100 Hz recording rates.

Axon Instruments, Inc. (Foster City, CA) produces a potentiostat called the Gene Clamp 500 that is ideal for fast amperometric recordings. This system can be linked to commercially designed software for data analysis and storage. In addition, the Dagan Corporation (Minneapolis, MN) makes the Chem-Clamp that is suitable for amperometric recordings. There are additional commercial patch-clamp amplifiers that can be used for amperometric recordings. These amplifier systems are generally used for single cell recordings involving very rapid (<50 ms) current measurements.[26–32,90]

Several companies market systems designed for high-speed in vivo electrochemical measurements. All of these instruments are optimized for rapid recordings using microelectrodes, but they cannot perform all of the aforementioned methods. SOLEA Tacussel (Villeurbanne, France) makes the Biopulse, which is an analog instrument for carrying out DPV, DNPV and DPA measurements. Cypress Systems sells the EI-400, which is a low-noise potentiostat that is generally used for FSCV measurements. It now has an internal waveform generator and custom, Windows 98-based software is available. P.D. Systems (West Molesey, Surrey, UK) sells the Millar Voltammeter, which is an analog system designed primarily for FSV recordings. GMA Technologies, Inc. (Vancouver, Canada) markets a microcomputer-controlled potentiostat called the E-Chempro electrochemical instrument. This instrument can carry out LSV and CA recordings. Harvard Apparatus (Holliston, MA) now markets the IVEC-10, which is a microcomputer-controlled system for performing HSC, FSCV, FCV and amperometry. This system is a ''turn-key'' microcomputer-controlled potentiostat that has menu-driven software that is very easy to use. The current availability and capabilities of all of the aforementioned instruments should be gained from the manufacturers and laboratories that use these instruments.

4.2 Laboratory Designed

Many laboratories build or modify their in vivo electrochemical recording systems. The typical laboratory-designed system consists of a two-stage amplification system, a general-purpose potentiostat, an external waveform generator (if required), and an output device such as a strip chart recorder, sample and hold amplifier, and/or a microcomputer. The preamplifier is located as close as possible to the microelectrodes to minimize noise. The second amplifier is often located within the potentiostat. Software can be written or commercially available programs can be employed to store the data. External waveform generators can cause excessive noise in these

systems, therefore proper grounding and shielding must be used.[12,49] Low-pass filtering can be used to eliminate high-frequency noise. This can be done using analog filters or can be carried out digitally using microcomputer software.

5 PREPARATIONS

By far the largest number of voltammetric recordings in CNS tissues have been carried out with respect to the nigrostriatal and mesolimbic DA pathways in the CNS of rats, mice, monkeys and guinea pigs.[1,2,4–7,9–14,16–20,22] Voltammetric measures of DA have involved studies of electrical stimulation of the medial forebrain bundle and the effects of drugs that alter DA neurons in anesthetized rats and more recently in free-moving rats.[7,12,14,16,20,22,42,57] The low extracellular levels of DA, NE and 5-HT have limited voltammetric studies of resting levels or basal levels of the neurotransmitters, which can be carried out with microdialysis methods. This has resulted in the extensive use of electrical and chemical stimuli to cause the release of DA so that it can be reliably measured. In contrast to the extensive number of studies of DA, there is a growing body of data regarding NE and 5-HT release from brain areas of the rat such as the hippocampus, substantia nigra, raphe nucleus, cerebellum, locus coeruleus and the dorsal raphe.[3,6,8,17,19,21,23]

5.1 Cells in Culture

Wightman et al.[91] introduced the concept of using fast amperometric recordings to study release of neurotransmitters, such as NE, from single cells. This is a rapidly growing area of study that uses $\leq$5-μm diameter disk recording electrodes coupled with amperometric or FSCV recordings to study properties of quantal release from individual cells in culture. A number of studies have been carried out on PC-12 cells, adrenal chromaffin cells, mast cells and engineered cells.[26–32,63,91] These studies demonstrate the very rapid recording capabilities of the amperometric recording methods. Although it can be argued that many of these cells are not necessarily good models of CNS neurons, a variety of mechanistic studies are being carried out to understand further the fundamental properties of exocytotic release.

5.2 Brain Slices/Tissue Pieces

Perhaps one of the most rapidly growing areas using voltammetric recording techniques involves the use of brain slices from rats and mice. This methodology has a number of advantages over studying the CNS using anesthetized or freely moving animals. First, the slice preparation allows for visual control of all electrode placements. This is particularly advantageous when one is studying layered structures such as the hippocampus and hard-to-hit nuclei such as the substantia nigra and the dorsal raphe.[3,6,8,19,70] Secondly, the slice preparations can be superfused with known concentrations of drugs for pharmacological studies. Third, the slice preparation makes it easier to study the brains of smaller animals such as mice. Finally, the slice preparation is unanesthetized tissue. Disadvantages involve the fact that outputs and inputs to brain structures are severed and that the viability of the slices can be a problem.

5.3 Whole Animals

Anesthetized rats have been used extensively to study the dynamics of DA neuronal systems using voltammetric techniques.[4,5,13–18,25,42] Potential disadvantages of this preparation are the effects of anesthesia on the measurements and the inability to know the exact concentrations of drugs that affect CNS function. However, the anesthetized preparation is still powerful for studies of the more intact CNS.

The other major preparation for voltammetric studies involves the use of freely moving animals. In these studies the recording electrodes are often implanted days prior to measurement or a microdrive cannulae system is attached to the skull of the rat.[7,10,11,22,48,64–66,92] These measurements often involve the use of special miniature potentiostats that are placed close to the animal's head.[7,22,93] Studies have been carried out to measure electrically-evoked release of DA and potassium-induced release of DA. In addition, a number of studies in freely moving rats have been carried out regarding the effects of drugs of abuse, such as cocaine and amphetamine. These studies are very time-consuming compared with brain slice or cell culture studies, but hold the promise of understanding more of the underlying properties of neurotransmitter signaling in freely moving and behaving rats.[93,94]

6 APPLICATIONS

The majority of in vivo voltammetric studies involves the insertion of the microelectrodes into the brain regions of interest in the intact brain or brain slices (see Figure 1). In the case of cells in culture, the microelectrodes are positioned as close as possible to individual cells. Voltammetric measurements are then started to achieve a stable background. Once stable measurements are recorded, the effects of an event such as electrical stimulation or chemical stimulation of the neurons or

cells is triggered and studied. Additional events are triggered to determine the effects of drugs used to alter the neuronal systems. The numerous applications and results obtained with these methods are beyond the scope of this article. We suggest that the interested investigator should read additional review articles and book chapters that concentrate on applications of this technology.[2,12,42]

6.1 Release of Neurotransmitters and Neurochemicals

The major reason for the development of in vivo electrochemical recording methods was to measure the release of neurotransmitters such as DA, NE and 5-HT in the intact CNS. Because of the very low basal release levels of these neurotransmitters, the majority of in vivo voltammetric studies to date have relied on ways of activating the release of these compounds from nerve endings, varicosities and dendrites by using electrical stimulation.[3,6,7,14,16,19,20,22] In addition, chemical stimuli such as *d*-amphetamine, excess potassium, veratridine, and tyramine have been extensively used to cause the release of neurotransmitters.[8,9,59,68–70] Moreover, in freely moving animals, systemically administered drugs and even self-administered drugs have been used to activate release processes.[10,11,49,64–66] An example of the dramatic differences seen between potassium-evoked release of DA and *d*-amphetamine-induced overflow of DA which were recorded from a rat striatal brain slice is seen in Figure 3. These data also demonstrate the power of the temporal resolution afforded by these recording techniques.

6.2 Uptake/Reuptake and Diffusion

The major mechanism for inactivating released neurotransmitters is through membrane-bound proteins known as transporters. These efficient proteins shuttle the released neurotransmitter back into the neuron and inactivate the signal produced by the neurotransmitter.[16,94] Voltammetric methods are ideally suited to studying the function of the DA, NE and 5-HT neuronal transporters.[3–5,7–9,13–16,18,20,55,57] This is carried out using FSCV and HSC in whole animals and brain slices, and RDEV in brain minces or cells. In addition, the diffusional properties of DA, NE and 5-HT in different brain regions can be studied by these methods. This is an emerging area of study, which requires the speed and spatial resolution of the voltammetric methods. An example of the measurement of DA uptake in brain slices of the rat striatum and the effects from cocaine is seen in Figure 4. Interestingly, Wightman and co-workers have developed a model based on electrical stimulation of the slice or intact animal that allows for measurement of both release and uptake parameters in the same preparation.[3,7,14,16,57]

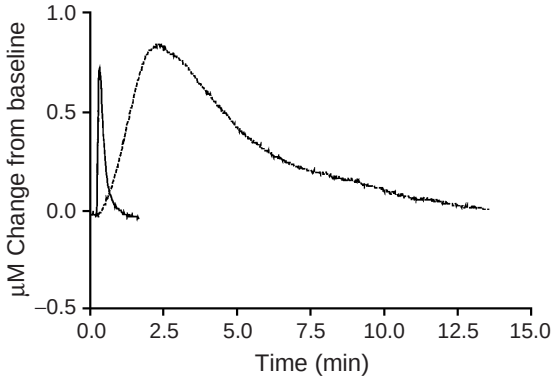

Figure 3 HSC recordings of the effects of potassium-evoked overflow (KCl, ——) and *d*-amphetamine-induced (- - - -) efflux of DA in the rat striatal slice. Potassium and *d*-amphetamine were added to the superfusion bath at time 0 for 2 min. Note the long time course of the *d*-amphetamine-induced release of DA.

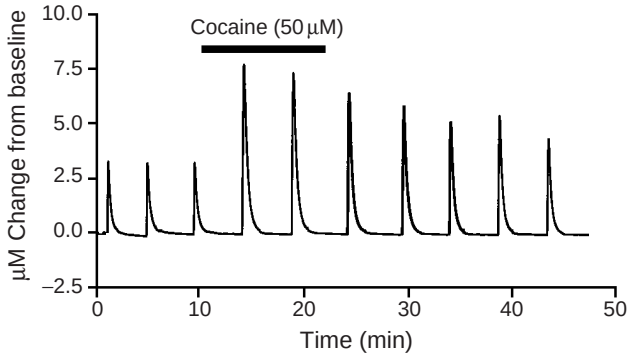

Figure 4 HSC recordings of the effects of cocaine (50 µM) on DA uptake in a brain slice from the rat striatum. DA was ejected from a glass micropipette at 5-min intervals. Cocaine was applied during the time period denoted by the black bar above the recording.

7 QUANTIFICATION AND IDENTIFICATION ISSUES

Perhaps the greatest controversy that has surrounded voltammetric measurements in CNS tissues and cells in culture concerns the identity of the detected molecules. Although voltammetric methods are inherently selective compared to many methods because of the relatively few electroactive compounds present in brain tissues and cells coupled with the use of low-oxidation potentials to limit the number of analytes that can be measured (see Table 1), there has always been concern that ''you don't know what you are measuring with the in vivo voltammetric recordings.'' Unlike microdialysis methods that use powerful separation techniques such as HPLC to separate molecules prior to detection, the microelectrodes

can see a number of molecules at the same time. This can confound the interpretation of in vivo voltammetric data. This has prompted the extensive development of electrode coatings (see below) to enhance further the selectivity of the microelectrodes for neurotransmitters or neurochemicals of interest. In fact, the real challenge for future development of microelectrodes is to develop microelectrodes and methods that are specific to a given molecule. In reality, any coating produces a diffusion layer that slows the response time of the microelectrode. This has limited the use of surface-modified electrodes for certain studies, such as rapid single cell recordings. However, electrode surface coatings that can solve some of the specificity problems are still under development.

7.1 Recording Methods

The best tools to use to identify further the detected molecule during fast recordings ($\geq$1 s) have involved the use of FSCV, FCV and the HSC recording methods.[6–12,14–16,19–20,22–24,68–70] As pointed out in section 3, the reduction/oxidation current relationship gives signatures for the detection of DA or 5-HT as compared to interferents such as AA and monoamine metabolites such as 5-HIAA and DOPAC. For example, DA yields distinctive cyclic voltammograms with the rapid voltammetry methods as well as distinctive reduction/oxidation current ratios using HSC recordings.[6–12,19,20,22–25,68–70] This is true for the measurement of 5-HT as well. Confirming the detection of NE versus DA is more difficult.[8,12] These chemical ''signatures'' are reliable measurements because they are fundamental properties of the detected compound.

7.2 Electrode Coatings

Electrode coatings to enhance the selectivity of microelectrodes for certain analytes and to make them sensitive to compounds that are not readily oxidized or reduced is a large area of scientific growth in the area of microelectrode development. In addition, such coatings can protect the surfaces of microelectrodes from fouling. Unfortunately, such layers can decrease the response times of the recordings, due to the diffusion layers created by these surface coatings.

7.2.1 Nafion™

Nafion™ is a Teflon® derivative that has sulfonic acid groups substituted into the polymer matrix. The sulfonate groups are negatively charged and this property has been used extensively to repel anions from electrode surfaces and prevent microelectrodes from fouling. Nafion™ films are one of the most widely used methods for improving the selectivity of voltammetric recordings in CNS tissues.[8,12,14,55,57,59,68–70] As well as repelling anions, Nafion™ is thought to concentrate cations like DA at the electrode surface. Nafion™ is commercially available as suspensions in alcohol (5%, Aldrich Chemical Company). Electrodes are simply dipped into this mixture or Nafion™ is electroplated on to the electrode surface.[55] Many methods of drying have been investigated including air drying,[4,5,18] use of heat guns,[12,14] and oven baking.[68–70] The effect upon electrode sensitivity of the various coating and drying methods can vary between laboratories and experimenters. However, Nafion™ has been consistently used to enhance the selectivity of CFEs or carbon monofilament microelectrodes by $\geq$100 : 1 for DA, NE and 5-HT compared with AA, uric acid, DOPAC and 5-HIAA. The tradeoff with Nafion™ is the thicker the coating, the greater the selectivity for cations versus anions but the slower the response time.

7.2.2 Electropolymerized Coatings

(Poly)phenylenediamine (PPD) has been used as a coating to achieve selectivity and prevent fouling on electrodes. Selectivity is attained due to size exclusion. The electropolymerized coating has holes that allow small molecules like NO and H_2O_2 to diffuse to the electrode surface to be oxidized while blocking larger organic molecules such as DA from the electrode. The *ortho*, *para*, and *meta*-isomers of phenylenediamine have been investigated as well as combinations with catechol and resorcinol.[95] We have developed a microelectrode based on a Nafion™-coated, *o*-phenylenediamine-modified CFE to measure NO selectively. Good selectivity over ascorbate, DA and nitrite has been observed. HSC (resting potential (E_o) = 0, applied potential (E_{max}) = +0.9 V) and amperometry (E = +0.9 V) were used to measure NO selectively.[34,96]

Overoxidized films have been proposed as a substitute for Nafion™ films. When oxidized, polypyrrole is a cationic conducting polymer. When overoxidized, the cationic groups are replaced by carbonyl and carboxylic acid groups. These oxygen-containing groups repel anions such as ascorbate while attracting cations like DA. These electrodes were shown to be more sensitive to DA than Nafion™-coated electrodes with FSCV.[97]

7.2.2 Porphryin

A prophyrinic-based microsensor for measuring NO was introduced by Malinski and Taha. Polymeric porphyrin was electrochemically deposited using CV on to carbon fibers from monomeric nickel tetrakis(3-methoxy-4-hydroxyphenyl)porphyrin (TMHPP). NO showed a peak at +0.63 V while NO_2^- had a peak at +0.8 V. A layer of Nafion™ was used to overcome the nitrate (NO_2^-) interference. DPV and amperometry were used to measure

NO with detection limits of 20 and 10 nM (signal-to-noise ratio = 3), respectively. The tip of certain types of microelectrode can be made small enough to penetrate individual cells.[33] This sensor has been applied to measurements of NO in rat brain[98] and in arteries.[99]

7.2.3 Enzymes

Enzymes can be coated or immobilized on to electrodes to measure analytes that are not electroactive. This is another area of major growth in this field. The analyte is enzymatically converted to an electroactive product, which can be monitored at the electrode surface. Enzymes are often very selective for a particular substrate, which contributes to the selectivity of such enzyme-based sensors.

Oxidase and dehydrogenase enzymes are commonly attached to the surfaces of the microelectrodes. Oxidase enzymes produce hydrogen peroxide as one of their products and the oxidation of hydrogen peroxide is monitored at the electrode. Glucose, lactate, choline, and glutamate oxidase enzymes have been made commercially available. Microelectrodes for all of these molecules have been reported. Because oxygen is consumed in the reaction, microelectrodes built around this scheme sometimes have noticeable oxygen dependence. The second class of enzyme is dehydrogenase enzymes. They oxidize the substrate while reducing NAD^+ (nicotinamide adenine dinucleotide (oxidized form)) to NADH (nicotinamide adenine dinucleotide (reduced form)). Schemes for detecting NADH have been developed. An excess of enzyme is usually immobilized on to the electrode to increase sensor lifetime.

7.2.3.1 L-Glucose Much work has been done to develop sensors to measure L-glucose.[100–106] This research has been motivated by the need to monitor continuously blood glucose levels of people suffering from diabetes. Typically, amperometry with an applied potential between +0.6 and +0.7 V is used to measure peroxide at a platinized carbon,[107] platinum[60,61] or platinum iridium electrode.[35] Using this microelectrode design, glucose has been measured in the rat brain.[40,41,108]

7.2.3.2 Lactate Similarly, lactate oxidase can be used to convert lactate into electroactive peroxide. Enzyme-modified CFEs have been used to measure lactate in the rat brain.[38,39]

7.2.3.3 Choline Peroxide can be measured at a carbon electrode by using horseradish peroxidase and an osmonium poly(vinylpyridine) redox polymer to shift the recording potential to −0.1 V versus SSCE. In this scheme, leaching of enzymes and the redox mediator is minimized. Another advantage is that interferents are minimized due to the low recording potential.[109] A miniaturized version of this sensor was applied to the measurement of local injections of choline in rat brain. Choline oxidase was included in the polymer to sense choline.[110] The detection limit was 1 μm for choline and response times were from 2 to 30 s.

7.2.3.4 L-Glutamate By using dehydrogenase enzymes, detection of NADH or NAD(P)H (nicotinamide adenine dinucleotide phosphate (reduced form)) can be used to track changes in glutamate. However, high overpotentials are required to oxidize NAD(P)H (+1.1 V at carbon[111] and +1.3 V at platinum).[112] High overpotentials are undesirable because of surface fouling, oxidation of interferents present in the sample, and high background currents. The major challenge is to oxidize NADH or NAD(P)H at lower potentials. Electron mediators, which lower the required potential to that of the mediator, have been employed for this purpose. Alvarez-Crespo et al. have described a glutamate sensor based upon a modified carbon paste electrode that was coated with PPD.[112] The nonconductive PPD film not only limits interferents from reaching the electrode and inhibits leaching of soluble species from the electrode surface, but in this case it allowed a lower potential (0 V versus Ag/AgCl reference) to be used for detection of the NAD(P)H produced from the glutamate dehydrogenase enzymatic reaction. An LOD of 3.8 μM was observed. Curulli et al. coated electrodes with 1,2-, 1,3-, 1,4-diaminobenzene (DAB) and 4-aminobiphenyl in the presence of pyrroloquinolinequinone (PQQ) using CV to obtain a layer through which only small molecules can pass to the electrode surface.[113] PQQ acts as an electron mediator for oxidation of NADH. Assembling and evaluation of new dehydrogenase enzyme electrode probes obtained by electropolymerization of aminobenzene isomers and PQQ on gold, platinum and carbon electrodes were examined. The use of the electropolymerized layer increased selectivity. PQQ made the electrodes capable of catalyzing the oxidation of the reduced form of NADH at a potential of +0.2 V. Glutamate dehydrogenase was entrapped in the polymer, which made the electrode sensitive to glutamate with a detection limit of 5 μM.

Enzymes can also be immobilized on to the electrode surface using glutaraldehyde as a cross-linker. Pan and Arnold developed a glutamate sensor consisting of a Pt electrode (diameter = 1.6 mm) coated with Nafion™ followed by glutaraldehyde cross-linked glutamate oxidase.[60] The anionic polymer Nafion™ was used to repel anions such as ascorbate while allowing hydrogen peroxide to be detected at the electrode surface. They reported an LOD of 0.3 μM with linearity up to 800 μM.

Entrapment of the enzyme within a polymer matrix has also been employed.[61] O'Neill et al.[61] have reported a glutamate sensor based on a 60-μm diameter platinum wire electrode that was dip-coated in a glutamate oxidase solution (200 units mL^{-1}, phosphate buffer, pH 7.4). *o*-Phenylenediamine was electropolymerized to entrap the enzyme. This method was an efficient use of enzyme compared to copolymerization.[61]

Enzymes can also be covalently attached to the electrode surface. Avidin was used to attach biotin-modified enzymes to a biotin-modified CFE.[114] In addition, membranes with immobilized enzymes have been attached to larger electrodes.

Elimination of interferents is generally done using coatings of Nafion™,[60] cellulose acetate,[35] and PPD.[61] Enzymes themselves can be used to eliminate interferents. Hu and Wilson et al. developed a glutamate electrode based on a two-enzyme design.[35] Selectivity was obtained by coating the Pt–Ir wire (diameter = 0.17 mm) with the anionic polymers Nafion™ and cellulose acetate and incorporating ascorbate oxidase into the enzyme layer. Ascorbate oxidase removed the interferent AA. Excellent sensitivity and selectivity were reported; however, the two-enzyme design gave the sensor a larger oxygen dependence.

8 COMPARISON WITH OTHER METHODS

8.1 Microdialysis

Microdialysis involves implanting a probe (200–300 μm in diameter by 0.5–4.0 mm in length) which has a dialysis membrane that will allow transfer of compounds of a given molecular weight range. Fluid is pumped through the microdialysis probe at flow rates of 0.3–2.0 μL min^{-1}. The fluid, usually artificial cerebrospinal fluid, is collected at regular intervals (ca. 5–20 min). The compounds of interest that diffuse into the dialyzing fluid are analyzed off-line by HPLC/EC, liquid chromatography with fluorescence detection or gas chromatography combined with mass spectrometry detection. Some people have considered microdialysis methods to be comparable to in vivo voltammetric methods. In reality, these methods are complimentary. In vivo voltammetric methods sample much smaller areas (5–30 μm) at sampling rates of often 1–200 Hz. In addition, the response of a microelectrode is not impeded by the slow exchange of materials that occurs at the microdialysis membrane. Moreover, the microdialysis methods can modify the extracellular space composition. Studies suggest that the microdialysis probe can greatly change the properties of neurons adjacent to the probe.[115] However, typical detection limits for voltammetric recordings of DA and 5-HT with microelectrodes are in the 25–50-nM range. This is poor compared to the routine picomole LODs achieved by analytical methods, such as HPLC/EC, that are used to measure neurotransmitters in microdialysis fractions. Thus, in vivo voltammetric methods have greater spatial and temporal resolution compared to microdialysis, but lack the sensitivity and ability to separate and quantitate numerous compounds that exist simultaneously in the extracellular space of the brain.

8.2 Fiber Optic Sensors

Fiber optic sensors have been used to measure neurotransmitters. Light of a desired wavelength is directed through a single optical fiber or a bundle. The light interacts with the chemical sensing end of the probe that is in contact with the sample. Light that has been modified by the compound of interest is collected with the same or different fibers then measured at a detector. Optical detection has the potential advantage of reduced electrical noise compared to electrochemical detection. In addition, multiple wavelengths can be used to identify compounds selectively. The disadvantage is that they are in general less sensitive than electrochemical detection and the instrumentation involved often precludes their use in freely moving animals. In addition, the sizes of fiber optic probes are too large for recordings of single cells. Fiber optic sensors have been reported for many compounds including NO[116,117] and glutamate.[118–120] To date, voltammetric-based recording methods have been more widely utilized for neurochemical detection than fiber optic sensors.

ACKNOWLEDGMENTS

This work was supported by United States Public Health Service grants NS09199, AG06434 and a Level II Research Scientist Development Award (MH01245) to Greg Gerhardt from the National Institutes of Mental Health. In addition, we thank the National Science Foundation for generous support through grant #DBI-9730899.

ABBREVIATIONS AND ACRONYMS

AA	Ascorbic Acid
CA	Chronoamperometry
CFE	Carbon Fiber Electrode
CNS	Central Nervous System
CSCV	Continuous Scan Cyclic Voltammetry

CV	Cyclic Voltammetry
DA	Dopamine
DAB	Diaminobenzene
DHPG	3,4-Dihydroxyphenylglycol
DNPV	Differential Normal Pulse Voltammetry
DOPA	3-(3,4-Dihydroxyphenyl)alanine
DOPAC	3,4-Dihydroxyphenylacetic Acid
DOPEG	Dihydroxyphenylethylene Glycol
DPA	Differential Pulse Amperometry
DPV	Differential Pulse Voltammetry
FCV	Fast Cyclic Voltammetry
FSCV	Fast-scan Cyclic Voltammetry
GABA	γ-Aminobutyric Acid
HPLC/EC	High-performance Liquid Chromatography Coupled to Electrochemical Detection
HSC	High-speed Chronoamperometry
HVA	Homovanillic Acid
ID	Inside Diameter
LOD	Limit of Detection
LSV	Linear Sweep Voltammetry
MAO	Monoamine Oxidase
MHPG	3-Methoxy-4-hydroxyphenylglycol
NAD^+	Nicotinamide Adenine Dinucleotide (Oxidized Form)
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
NAD(P)H	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
NE	Norepinephrine
NO	Nitric Oxide
OD	Outside Diameter
PPD	(Poly)phenylenediamine
PQQ	Pyrroloquinolinequinone
RDE	Rotating Disk Electrode
RDEV	Rotating Disk Electrode Voltammetry
SSCE	Sodium Saturated Calomel Electrode
TMHPP	Tetrakis(3-methoxy-4-hydroxyphenyl)porphyrin
3-MT	3-Methoxytyramine
5-HIAA	5-Hydroxyindoleacetic Acid
5-HT	Serotonin
5-HTP	5-Hydroxytryptophan

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Voltammetry In Vivo for Chemical Analysis of the Living Brain

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Neurotransmitters, Electrochemical Detection of • Pulse Voltammetry • Selective Electrode Coatings for Electroanalysis • Ultrafast Electrochemical Techniques

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

REFERENCES

1. P.T. Kissinger, J.B. Hart, R.N. Adams, 'Voltammetry in Brain Tissue: A New Neurophysiological Measurement', *Brain Res.*, **55**(1), 209–213 (1973).
2. R.N. Adams, 'In Vivo Electrochemical Measurements in the CNS', *Prog. Neurobiol.*, **35**(4), 297–311 (1990).
3. M.A. Bunin, R.M. Wightman, 'Quantitative Evaluation of 5-Hydroxytryptamine (Serotonin) Neuronal Release and Uptake: An Investigation of Extrasynaptic Transmission', *J. Neurosci.*, **18**(13), 4854–4860 (1998).
4. W.A. Cass, G.A. Gerhardt, R.D. Mayfield, P. Curella, N.R. Zahniser, 'Differences in Dopamine Clearance and Diffusion in Rat Striatum and Nucleus Accumbens Following Systemic Cocaine Administration', *J. Neurochem.*, **59**(1), 259–266 (1992).
5. W.A. Cass, G.A. Gerhardt, K. Gillespie, P. Curella, R.D. Mayfield, N.R. Zahniser, 'Reduced Clearance of Exogenous Dopamine in Rat Nucleus Accumbens, but not in Dorsal Striatum, Following Cocaine Challenge in Rats Withdrawn from Repeated Cocaine Administration', *J. Neurochem.*, **61**(1), 273–283 (1993).
6. S.J. Cragg, C.R. Hawkey, S.A. Greenfield, 'Comparison of Serotonin and Dopamine Release in Substantia Nigra and Ventral Tegmental Area: Region and Species Differences', *J. Neurochem.*, **69**(6), 2378–2386 (1997).
7. P.A. Garris, J.R. Christensen, G.V. Rebec, R.M. Wightman, 'Real-time Measurement of Electrically Evoked Extracellular Dopamine in the Striatum of Freely Moving Rats', *J. Neurochem.*, **68**(1), 152–161 (1997).
8. G.A. Gerhardt, 'Rapid Chronocoulometric Measurements of Norepinephrine Overflow and Clearance in CNS Tissues', in *Neuromethods–Voltammetric Methods in Brain Systems*, Humana, New Jersey, 117–151, Vol. 27, 1995.
9. G.A. Gerhardt, W.A. Cass, J. Hudson, M. Henson, Z. Zhang, A. Ovadia, B.J. Hoffer, D.M. Gash, 'In Vivo Electrochemical Studies of Dopamine Overflow and Clearance in the Striatum of Normal and MPTP-treated Rhesus Monkeys', *J. Neurochem.*, **66**(2), 579–588 (1996).

10. A. Gratton, R.A. Wise, 'Drug- and Behavior-associated Changes in Dopamine-related Electrochemical Signals During Intravenous Cocaine Self-administration in Rats', *J. Neurosci.*, **14**(7), 4130–4146 (1994).
11. A. Gratton, 'In Vivo Analysis of the Role of Dopamine in Stimulant and Opiate Self-administration', *J. Psychiatry Neurosci.*, **21**(4), 264–279 (1996).
12. K.T. Kawagoe, J.B. Zimmerman, R.M. Wightman, 'Principles of Voltammetry and Microelectrode Surface States', *J. Neurosci. Meth.*, **48**, 225–240 (1993).
13. E.J. Cline, C.E. Adams, G.A. Larson, G.A. Gerhardt, N.R. Zahniser, 'Medial Dorsal Striatum is More Sensitive than Lateral Dorsal Striatum to Cocaine Inhibition of Exogenous Dopamine Clearance: Relation to [3*H*]Mazindol Binding, but not Striosome/Matrix', *Exp. Neurol.*, **134**(1), 135–149 (1995).
14. L.J. May, R.M. Wightman, 'Heterogeneity of Stimulated Dopamine Overflow within Rat Striatum as Observed with In Vivo Voltammetry', *Brain Res.*, **487**(2), 311–320 (1989).
15. L.C. Daws, G.M. Toney, D.J. Davis, G.A. Gerhardt, A. Frazer, 'In Vivo Chronoamperometric Measurements of the Clearance of Exogenously Applied Serotonin in the Rat Dentate Gyrus', *J. Neurosci. Meth.*, **78**(1–2), 139–150 (1997).
16. B. Giros, M. Jaber, S.R. Jones, R.M. Wightman, M.G. Caron, 'Hyperlocomotion and Indifference to Cocaine and Amphetamine in Mice Lacking the Dopamine Transporter', *Nature*, **379**(6566), 606–612 (1996).
17. F. Gonon, 'Monitoring Dopamine and Noradrenaline (Norepinephrine) Release in Central and Peripheral Nervous Systems with Treated and Untreated Carbon-fiber Electrodes', in *Neuromethods–Voltammetric Methods in Brain Systems*, Humana, New Jersey, 117–151, Vol. 27, 1995.
18. J. Luthman, M.N. Friedemann, B.J. Hoffer, G.A. Gerhardt, 'In Vivo Electrochemical Measurements of Serotonin Clearance in Rat Striatum: Effects of Neonatal 6-Hydroxydopamine-induced Serotonin Hyperinnervation and Serotonin Uptake Inhibitors', *J. Neural Transmitters*, **104**(4–5), 379–397 (1997).
19. M.E. Rice, C.D. Richards, S. Nedergaard, J. Hounsgaard, C. Nicholson, S.A. Greenfield, 'Direct Monitoring of Dopamine and 5-HT Release in Substantia Nigra and Ventral Tegmental Area In Vitro', *Exp. Brain Res.*, **100**(3), 395–406 (1994).
20. J.A. Stamford, Z.L. Kruk, P. Palij, J. Millar, 'Diffusion and Uptake of Dopamine in Rat Caudate and Nucleus Accumbens Compared Using Fast Cyclic Voltammetry', *Brain Res.*, **448**(2), 381–385 (1988).
21. P. Brun, M.F. Suaud-Chagny, F. Gonon, M. Buda, 'Differential Effects of Desipramine on Direct- and Sensory-evoked Noradrenaline Release in Thalamic Locus Coeruleus Terminals', *Eur. J. Pharmacol.*, **235**(2–3), 205–210 (1993).
22. Z.L. Kruk, S. Cheeta, J. Milla, R. Muscat, J.E. Williams, P. Willner, 'Real Time Measurement of Stimulated Dopamine Release in the Conscious Rat Using Fast Cyclic Voltammetry: Dopamine Release is not Observed During Intracranial Self Stimulation', *J. Neurosci. Meth.*, **79**(1), 9–19 (1998).
23. L.F. Callado, J.A. Stamford, 'Alpha 2A- But Not Alpha 2B/C-Adrenoceptors Modulate Noradrenaline Release in Rat Locus Coeruleus: Voltammetric Data', *Eur. J. Pharmacol.*, **366**(1), 35–39 (1999).
24. M.M. Iravani, J. Millar, Z.L. Kruk, 'Differential Release of Dopamine by Nitric Oxide in Subregions of Rat Caudate Putamen Slices', *J. Neurochem.*, **71**(5), 1969–1977 (1998).
25. J.A. Stamford, J.B. Justice, Jr, 'Probing Brain Chemistry', *Anal. Chem.*, **68**(11), 359A–363A (1996).
26. K.T. Kawagoe, J.A. Jankowski, R.M. Wightman, 'Etched Carbon-fiber Electrodes as Amperometric Detectors of Catecholamine Secretion from Isolated Biological Cells', *Anal. Chem.*, **63**(15), 1589–1594 (1991).
27. R.A. Clark, A.G. Ewing, 'Quantitative Measurements of Released Amines from Individual Exocytosis Events', *Mol. Neurobiol.*, **15**(1), 1–16 (1997).
28. A. Elhamdani, Z. Zhou, C.R. Artalejo, 'Timing of Dense-core Vesicle Exocytosis Depends on the Facilitation L-Type Ca Channel in Adrenal Chromaffin Cells', *J. Neurosci.*, **18**(16), 6230–6240 (1998).
29. E.R. Travis, R.M. Wightman, 'Spatio-temporal Resolution of Exocytosis from Individual Cells', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 77–103 (1998).
30. R.J. Fisher, R.D. Burgoyne, 'The Effect of Transfection with Botulinum Neurotoxin C1 Light Chain on Exocytosis Measured in Cell Populations and by Single-cell Amperometry in PC12 Cells', *Pflugers Arch.*, **437**(5), 754–762 (1999).
31. M. Anderova, A.D. Duchene, J.G. Barbara, K. Takeda, 'Vasoactive Intestinal Peptide Potentiates and Directly Stimulates Catecholamine Secretion from Rat Adrenal Chromaffin Cells', *Brain Res.*, **809**(1), 97–106 (1998).
32. L.M. Gutierrez, A. Gil, S. Viniegra, 'Preferential Localization of Exocytotic Active Zones in the Terminals of Neurite-emitting Chromaffin Cells', *Eur. J. Cell Biol.*, **76**(4), 274–278 (1998).
33. T. Malinski, Z. Taha, 'Nitric Oxide Release from a Single Cell Measured In Situ by a Porphyrinic-based Microsensor', *Nature*, **358**, 676–678 (1992).
34. M.N. Friedemann, S.W. Robinson, G.A. Gerhardt, '*O*-Phenylenediamine-modified Carbon Fiber Electrodes for the Detection of Nitric Oxide', *Anal. Chem.*, **68**(15), 2621–2628 (1996).
35. Y. Hu, K.M. Mitchell, F.N. Albahadily, E.K. Michaelis, G.S. Wilson, 'Direct Measurement of Glutamate Release in the Brain Using a Dual Enzyme-based Electrochemical Sensor', *Brain Res.*, **659**, 117–125 (1994).
36. W.J. Albery, 'Molecular Recognition and Molecular Sensors', *Ciba. Found. Symp.*, **158**, 55–67 (1991).

37. O. Niwa, R. Kurita, T. Horiuchi, K. Torimitsu, 'Small-volume On-line Sensor for Continuous Measurement of γ-Aminobutyric Acid', *Anal. Chem.*, **70**(1), 89–93 (1998).
38. Y. Hu, G.S. Wilson, 'A Temporary Local Energy Pool Coupled to Neuronal Activity: Fluctuations of Extracellular Lactate Levels in Rat Brain Monitored with Rapid-response Enzyme-based Sensor', *J. Neurochem.*, **69**(4), 1484–1490 (1997).
39. N.F. Shram, L.I. Netchiporouk, C. Martelet, N. Jaffrezic-Renault, C. Bonnet, R. Cespuglio, 'In Vivo Voltammetric Detection of Rat Brain Lactate with Carbon Fiber Microelectrodes Coated with Lactate Oxidase', *Anal. Chem.*, **70**(13), 2618–2633 (1998).
40. L.I. Netchiporouk, N.F. Shram, N. Jaffrezic-Renault, C. Martlet, R. Cespuglio, 'In Vivo Brain Glucose Measurements: Differential Normal Pulse Voltammetry with Enzyme-modified Carbon Fiber Microelectrodes', *Anal. Chem.*, **68**(24), 4358–4364 (1996).
41. Y. Hu, G.S. Wilson, 'Rapid Changes in Local Extracellular Rat Brain Glucose Observed with an In Vivo Glucose Sensor', *J. Neurochem.*, **68**(4), 1745–1752 (1997).
42. A.A. Boulton, G.B. Baker, R.N. Adams, 'Voltammetric Methods in Brain Systems', in *Neuromethods*, Humana Press, New Jersey, Vol. 27, 1995.
43. A.J. Bard, L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, John Wiley & Sons, New York, 1980.
44. R.N. Adams, *Electrochemistry at Solid Electrodes*, Marcel Dekker, New York, 1969.
45. J.O. Schenk, R.N. Adams, 'Chronoamperometric Measurements in the Central Nervous System', *IBRO Handbooks Series: Methods in the Neurosciences*, John Wiley & Sons, New York, 193–208, Vol. 6, 1984.
46. C.A. Marsden, M.P. Brazell, N.T. Maidment, *IBRO Handbooks Series: Methods in the Neurosciences*, John Wiley & Sons, New York, 127–152, Vol. 6, 1984.
47. P. Huettl, G.A. Gerhardt, 'Multicomponent Quantitations of Monoamines and Related Compounds in Biological Fluids and CNS Tissues by HPLC Coupled with Multi-electrode Electrochemical Detection', *Prog. HPLC-HPCE*, **6**, 213–225 (1997).
48. P.A. Broderick, M.F. Piercey, 'Neurochemical and Behavioral Evidence Supporting (+)-AJ 76 as a Potential Pharmacotherapy for Cocaine Abuse', *J. Neural Transmitters*, **105**(10–12), 1307–1324 (1998).
49. J. Millar, 'Simultaneous In-Vivo Voltammetric and Electrophysiological Recording with Carbon Fiber Microelectrodes', in *Methods in Neurosciences*, ed. M. Conn, Academic Press, New York, Vol. 4, 1991.
50. F. Gonon, M. Buda, J.F. Pujol, 'Treated Carbon Fiber Electrodes for Measuring Catechols and Ascorbic Acid', in *Measurement of Neurotransmitter Release In Vivo*, ed. C.A. Morsden, John Wiley & Sons, Chichester, 153–171, 1984.
51. Z. Zhou, S. Misler, R.H. Chow, 'Rapid Fluctuations in Transmitter Release from Single Vesicles in Bovine Adrenal Chromaffin Cells', *Biophys. J.*, **73**(3), 1543–1552 (1996).
52. Z. Zhou, S. Misler, 'Amperometric Detection of Quantal Secretion from Patch-clamped Rat Pancreatic β-Cells', *J. Biol. Chem.*, **271**(1), 270–277 (1996).
53. J.A. Stamford, P. Palij, C. Davidson, C.M. Jorm, J. Millar, 'Simultaneous 'Real-time' Electrochemical and Electrophysiological Recording in Brain Slices with a Single Carbon-fiber Microelectrode', *J. Neurosci. Meth.*, **50**, 279–290 (1993).
54. P.S. Cahill, Q.D. Walker, J.M. Finnegan, G.E. Mickelson, E.R. Travis, R.M. Wightman, 'Microelectrodes for the Measurement of Catecholamines in Biological Systems', *Anal. Chem.*, **68**, 3180–3186 (1996).
55. M.E. Rice, C. Nicholson, 'Measurement of Nanomolar Dopamine Diffusion Using Low-noise Perfluorinated Ionomer Coated Carbon Fiber Microelectrodes and High-speed Cyclic Voltammetry', *Anal. Chem.*, **61**, 1805–1810 (1989).
56. R.L. McCreery, 'Carbon Electrode Surface Chemistry', in *Voltammetric Methods in Brain Systems*, Humana Press, New Jersey, 1–26, 1995.
57. P.A. Garris, R.M. Wightman, 'Diffusional Differences in Dopamine Release, Uptake, and Diffusion Measured by Fast-scan Cyclic Voltammetry', in *Voltammetric Methods in Brain Systems*, Humana Press, New Jersey, 179–220, 1995.
58. G. Sreenivas, S.S. Ang, I. Fritsch, W.D. Brown, G.A. Gerhardt, D.J. Woodward, 'Fabrication and Characterization of Sputtered-carbon Microelectrode Arrays', *Anal. Chem.*, **68**, 1858–1864 (1996).
59. C. van Horne, B. Bement, B.J. Hoffer, G.A. Gerhardt, 'Multichannel Semiconductor-based Electrodes for In Vivo Electrochemical and Electrophysiological Studies in Rat CNS', *Neurosci. Lett.*, **120**, 249–252 (1990).
60. S. Pan, M.A. Arnold, 'Selectivity Enhancement for Glutamate with a Nafion™/Glutamate Oxidase Biosensor', *Talanta*, **43**, 1157–1162 (1996).
61. M.R. Ryan, J.P. Lowry, R.O. O'Neill, 'Biosensor for Neurotransmitter L-Glutamic Acid Designed for Efficient Use of L-Glutamate Oxidase and Effective Rejection of Interference', *Analyst*, **122**, 1419–1424 (1997).
62. W.J. Albery, M.G. Boutelle, P.T. Galley, 'The Dialysis Electrode: A New Method for In Vivo Monitoring', *J. Chem. Soc., Chem. Commun.*, **12**, 900–901 (1992).
63. X. Quan, R.M. Wightman, 'Simultaneous Detection of Catecholamine Exocytosis and Ca^{2+} Release from Single Bovine Chromaffin Cells Using a Dual Microsensor', *Anal. Chem.*, **70**, 1677–1681 (1998).
64. P.S. Cahill, R.M. Wightman, 'Simultaneous Amperometric Measurement of Ascorbate and Catecholamine Secretion from Individual Bovine Adrenal Medullary Cells', *Anal. Chem.*, **67**(15), 2599–2605 (1995).

65. P. Di Ciano, C.D. Blaha, A.G. Phillips, 'Conditioned Changes in Dopamine Oxidation Currents in the Nucleus Accumbens of Rats by Stimuli Paired with Self-administration or Yoked-administration of D-Amphetamine', *Eur. J. Neurosci.*, **10**(3), 1121–1127 (1998).
66. P. Di Ciano, C.D. Blaha, A.G. Phillips, 'The Relation Between Dopamine Oxidation Currents in the Nucleus Accumbens and Conditioned Increases in Motor Activity in Rats Following Repeated Administration of D-Amphetamine or Cocaine', *Eur. J. Neurosci.*, **10**(3), 1113–1120 (1998).
67. S.B. Floresco, C.R. Yang, A.G. Phillips, C.D. Blaha, 'Basolateral Amygdala Stimulation Evokes Glutamate Receptor-dependent Dopamine Efflux in the Nucleus Accumbens of the Anaesthetized Rat', *Eur. J. Neurosci.*, **10**(4), 1241–1251 (1998).
68. A.F. Hoffman, G.A. Gerhardt, 'Differences in Pharmacological Properties of Dopamine Release Between the Substantia Nigra and Striatum: An In Vivo Electrochemical Study', *J. Pharmacol. Exp. Ther.*, **289**, 455–463 (1999).
69. M.A. Hebert, C.G. van Horne, B.J. Hoffer, G.A. Gerhardt, 'Functional Effects of GDNF in Normal Rat Striatum: Presynaptic Studies Using in vivo Electrochemistry and Microdialysis', *J. Pharmacol. Exp. Ther.*, **279**, 1181–1190 (1996).
70. M.F. Suaud-Chagny, P. Brun, M. Buda, F. Gonon, 'Fast In Vivo Monitoring of Electrically Evoked Dopamine Release by Differential Pulse Amperometry with Untreated Carbon Fiber Electrodes', *J. Neurosci. Meth.*, **45**, 183–190 (1992).
71. M.E. Rice, S.J. Cragg, S.A. Greenfield, 'Characteristics of Electrically Evoked Somatodendritic Dopamine Release in Substantia Nigra and Ventral Tegmental Area In Vitro', *J. Neurophysiol.*, **77**(2), 853–862 (1997).
72. S.J. Cragg, M.E. Rice, S.A. Greenfield, 'Heterogeneity of Electrically Evoked Dopamine Release and Reuptake in Substantia Nigra, Ventral Tegmental Area, and Striatum', *J. Neurophysiol.*, **77**(2), 863–873 (1997).
73. D. Michael, E.R. Travis, R.M. Wightman, 'Color Images for Fast-scan CV Measurements in Biological Systems', *Anal. Chem.*, **70**(17), 586A–592A (1998).
74. J. Kume-Kick, M.E. Rice, 'Dependence of Dopamine Calibration Factors on Media Ca^{2+} and Mg^{2+} at Carbon-fiber Microelectrodes Used with Fast-scan Cyclic Voltammetry', *J. Neurosci. Meth.*, **84**, 55–62 (1998).
75. A.C. Michael, J.B. Justice, Jr, 'Oxidation of Dopamine and 4-Methylcatechol at Carbon Fiber Disk Electrodes', *Anal. Chem.*, **59**, 405–410 (1987).
76. B.P. Jackson, S.M. Dietz, R.M. Wightman, 'Fast-scan Cyclic Voltammetry of 5-Hydroxytryptamine', *Anal. Chem.*, **67**, 1115–1120 (1995).
77. K. Pihel, S. Hsieh, J.W. Jorgenson, R.M. Wightman, 'Electrochemical Detection of Histamine and 5-Hydroxytryptamine at Isolated Mast Cells', *Anal. Chem.*, **67**, 4514–4521 (1995).
78. J. Millar, T.G. Barnett, 'Basic Instrumentation for Fast Cyclic Voltammetry', *J. Neurosci. Meth.*, **25**, 91–95 (1988).
79. Z.L. Kruk, S. Cheeta, J. Milla, R. Muscat, J.E.G. Williams, P. Willner, 'Real Time Measurement of Stimulated Dopamine Release in the Conscious Rat Using Fast Cyclic Voltammetry: Dopamine Release is not Observed During Intracranial Self Stimulation', *J. Neurosci. Meth.*, **79**, 9–19 (1998).
80. S. Hafizi, Z.L. Kruk, J.A. Stamford, 'Fast Cyclic Voltammetry: Improved Sensitivity to Dopamine with Extended Oxidation Scan Limits', *J. Neurosci. Meth.*, **33**, 41–49 (1990).
81. J. Millar, J.J. O'Connor, S.J. Trout, Z.L. Kruk, 'Continuous Scan Cyclic Voltammetry (CSCV): A New High-speed Electrochemical Method for Monitoring Neuronal Dopamine Release', *J. Neurosci. Meth.*, **43**, 109–118 (1992).
82. M.M. Iravani, J. Millar, Z.L. Kruk, 'Differential Release of Dopamine by Nitric Oxide in Subregions of Rat Caudate Putamen Slices', *J. Neurochem.*, **71**, 1969–1977 (1998).
83. P.A. Broderick, 'In Vivo Electrochemical Studies of Gradient Effects of (SC) Cocaine on Dopamine and Serotonin Release in Dorsal Striatum of Conscious Rats', *Pharmacol. Biochem. Behav.*, **46**, 973–984 (1993).
84. S.M. Meiergerd, J.O. Schenk, 'Kinetic Evaluation of the Commonality Between the Site(s) of Action of Cocaine and Some Other Structurally Similar and Dissimilar Inhibitors of the Striatal Transporter of Dopamine', *J. Neurochem.*, **63**, 1683–2692 (1994).
85. W.B. Burnette, M.D. Bailey, S. Kukoyi, R.D. Blakely, C.G. Trowbridge, J.B. Justice, Jr, 'Human Norepinephrine Transporter Kinetics Using Rotating Disk Electrode Voltammetry', *Anal. Chem*, **68**, 2932–2938 (1996).
86. N. Chen, J.B. Justice, Jr, 'Cocaine Acts as an Apparent Competitive Inhibitor at the Outward-facing Conformation of the Human Norepinephrine Transporter: Kinetic Analysis of Inward and Outward Transport', *J. Neurosci.*, **18**(24), 10257–10268 (1998).
87. R. Rivest, C.A. Marsden, 'Muscarinic Antagonists Attenuate Neurotensin-stimulated Accumbens and Striatal Dopamine Metabolism', *Neuroscience*, **47**(2), 341–349 (1992).
88. M.F. Suaud-Chagny, R. Cespuglio, J.P. Rivot, M. Buda, F. Gonon, 'High Sensitivity Measurement of Brain Catechols and Indoles In Vivo Using Electrochemically Treated Carbon-fiber Electrodes', *J. Neurosci. Meth.*, **48**, 241–250 (1993).
89. J.-Y. Gillion, F. Richard, L. Quintin, J.-F. Pujol, B. Renaud, 'Pharmacological and Functional Evidence for Extracellular 3,4-Dihydroxyphenylacetic Acid as an Index of Metabolic Activity of the Adrenergic Neurons: An In Vivo Voltammetry Study in the Rat Rostral Ventrolateral Medulla', *Neuroscience*, **37**(2), 421–430 (1990).

90. E.N. Pothos, V. Davila, D. Sulzer, 'Presynaptic Recording of Quanta from Midbrain Dopamine Deurons and Modulation of the Quantal Size', *J. Neurosci.*, **18**(11), 4106–4118 (1998).
91. D.J. Leszczyszyn, J.A. Jankowski, O.H. Viveros, E.J. Diliberto, Jr, J.A. Near, R.M. Wightman, 'Nicotinic Receptor-mediated Catecholamine Secretion from Individual Chromaffin Cells. Chemical Evidence for Exocytosis', *J. Biol. Chem.*, **265**(25), 14 736–14 737 (1990).
92. G.A. Gerhardt, C. Ksir, C. Pivik, S.D. Dickinson, J. Sabeti, N.R. Zahniser, 'Methodology for Coupling Local Application of Dopamine and Other Chemicals with Rapid In Vivo Electrochemical Recordings in Freely-moving Rats', *J. Neurosci. Meth.*, **87**, 67–76 (1999).
93. P.A. Garris, M. Kilpatrick, M.A. Bunin, D. Michael, Q.D. Walker, R.M. Wightman, 'Dissociation of Dopamine Release in the Nucleus Accumbens from Intracranial Self-stimulation', *Nature*, **398**(6722), 67–69 (1999).
94. G. Rudnick, J. Clark, 'From Synapse to Vesicle: The Reuptake and Storage of Biogenic Amine Neurotransmitters', *Chim. Biophys. Acta*, **1144**, 249–263 (1993).
95. R.J. Geise, J.M. Adams, N.J. Barone, A.M. Yacynych, 'Electropolymerized Films Prevent Interferences and Electrode Fouling in Biosensors', *Biosens. Bioelec.*, **6**, 151–160 (1991).
96. D.A. Smith, A.F. Hoffman, D.J. David, C.E. Adams, G.A. Gerhardt, 'Nicotine-evoked Nitric Oxide Release in the Rat Hippocampal Slice', *Neurosci. Lett.*, **255**, 127–130 (1998).
97. K. Pihel, Q.D. Walker, R.M. Wightman, 'Overoxidized Polypyrole-coated Carbon Fiber Microelectrodes for Dopamine Measurements with Fast-scan Cyclic Voltammetry', *Anal. Chem.*, **68**, 2084–2089 (1996).
98. Z.G. Zhang, M. Chopp, F. Bailey, T. Malinski, 'Nitric Oxide Changes in the Rat Brain after Transient Middle Cerebral Artery Occlusion', *J. Neurol. Sci.*, **128**, 22–27 (1995).
99. M. Gerova, S. Mesaros, F. Kristek, M. Kittova, T. Malinski, 'NO Concentration in the Periendothelial Area of the Femoral Artery of Dog Measured In Vivo', *Physiol. Res.*, **47**(3), 169–175 (1998).
100. C.A. Quinn, R.E. Connor, A. Heller, 'Biocompatible, Glucose-permeable Hydrogel for In Situ Coating of Implantable Biosensors', *Biomaterials*, **18**(24), 1665–1670 (1997).
101. D.W. Schmidtke, A. Heller, 'Accuracy of the One-point In Vivo Calibration of ''Wired'' Glucose Oxidase Electrodes Implanted in Jugular Veins of Rats in Periods of Rapid Rise and Decline of the Glucose Concentration', *Anal. Chem.*, **70**(10), 2149–2155 (1998).
102. J.G. Wagner, D.W. Schmidtke, C.P. Quinn, T.F. Fleming, B. Bernacky, A. Heller, 'Continuous Amperometric Monitoring of Glucose in a Brittle Diabetic Chimpanzee with a Miniature Subcutaneous Electrode', *Proc. Natl. Acad. Sci. USA.*, **95**(11), 6379–6382 (1998).
103. G. Kenausis, Q. Chen, A. Heller, 'Electrochemical Glucose and Lactate Sensors Based on ''Wired'' Thermostable Soybean Peroxidase Operating Continuously and Stably at 37 °C', *Anal. Chem.*, **69**(6), 1054–1060 (1997).
104. D.S. Bindra, Y. Zhang, R. Sternberg, D.R. Thevenot, D. Moatti, G. Reach, G.S. Wilson, 'Design and In Vitro Studies of a Needle-type Glucose Sensor for Subcutaneous Monitoring', *Anal. Chem.*, **63**, 1692–1696 (1991).
105. V. Thome-Duret, G. Reach, M.N. Gangnerau, F. Lemonnier, J.C. Klein, Y. Zhang, Y. Hu, G.S. Wilson, 'Use of a Subcutaneous Glucose Sensor to Detect Decreases in Glucose Concentration Prior to Observation in Blood', *Anal. Chem.*, **68**(21), 3822–3826 (1996).
106. V. Thome-Duret, B. Aussedat, G. Reach, M.N. Gangnerau, F. Lemonnier, J.C. Klein, Y. Zhang, Y. Hu, G.S. Wilson, 'Continuous Glucose Monitoring in the Free-moving Rat', *Metabolism*, **47**(7), 799–803 (1998).
107. E. Tamiya, Y. Sugiura, Y. Amou, I. Karube, A. Ajima, R.T. Kado, M. Ito, 'Direct Monitoring of Glutamate Released from Cerebellar Neurons Using Platinized Carbon Disk Microelectrode Modified with Glutamate Oxidase-bound Membrane', *Sens. Mater.*, **7**, 249–259 (1995).
108. N.F. Shram, L.I. Netchiporouk, C. Martelet, N. Jaffrezic-Renault, R. Cespuglio, 'Brain Glucose: Voltammetric Determination in Normal and Hyperglycaemic Rats Using A Glucose Microsensor', *Neuroreport*, **8**(5), 1109–1112 (1997).
109. M.G. Garguilo, N. Huynh, A. Proctor, A.C. Michael, 'Amperometric, Sensors for Peroxide, Choline, and Acetylcholine on Electron Transfer between Horseradish Peroxidase and a Redox Polymer', *Anal. Chem.*, **65**, 523–528 (1993).
110. M.G. Garguilo, A.C. Michael, 'Quantification of Choline in the Extracellular Fluid of Brain Tissue with Amperometric Microsensors', *Anal. Chem.*, **66**, 2621–2629 (1994).
111. J. Moiroux, P.P. Elving, 'Effects of Adsorption, Electrode Material and Operational Variables on the Oxidation of Dihydronicotinamide Adenine Dinucleotide at Carbon Electrodes', *Anal. Chem.*, **50**, 1056–1062 (1978).
112. S.L. Alvarez-Crespo, M.J. Lobo-Castanon, A.J. Miranda-Ordieres, P. Tunon-Blanco, 'Amperometric Glutamate Biosensor Based on Poly(*o*-phenylenediamine) Film Electrogenerated on to Modified Carbon Paste Electrodes', *Biosens. Bioelectron.*, **12**(8), 739–747 (1997).
113. A. Curulli, I. Carelli, O. Trischitta, G. Palleschi, 'Assembling and Evaluation of a New Dehydrogenase Enzyme Electrode Probes Obtained by Electropolymerization of Aminobenzene Isomers and PQQ on Gold, Platinum, and Carbon Electrodes', *Biosens. Bioelectron.*, **12**(9–10), 1043–1055 (1997).
114. P. Pantano, W.G. Kuhr, 'Dehydrogenase-modified Carbon-fiber Microelectrodes for the Measurement

of Neurotransmitter Dynamics. 2. Covalent Modification Utilizing Avidin–Biotin Technology', *Anal. Chem.*, **65**(5), 623–630 (1993).

115. H. Yang, J.L. Peters, A.C. Michael, 'Coupled Effects of Mass Transfer and Uptake Kinetics on In Vivo Microdialysis of Dopamine', *J. Neurochem.*, **71**, 684–692 (1998).
116. S.L. Barker, R. Kopelman, 'Development and Cellular Applications of Fiber Optic Nitric Oxide Sensors Based on a Gold-adsorbed Fluorophore', *Anal. Chem.*, **70**(23), 4902–4906 (1998).
117. S.L. Barker, R. Kopelman, T.E. Meyer, M.A. Cusanovich, 'Fiber-optic Nitric Oxide-selective Biosensors and Nanosensors', *Anal. Chem.*, **70**(5), 971–976 (1998).
118. J. Cordek, X. Wang, W. Tan, 'Direct Immobilization of Glutamate Dehydrogenase of Optical Fiber Probes for Ultrasensitive Glutamate Detection', *Anal. Chem.*, **71**, 1529–1533 (1999).
119. S. Kar, M.A. Arnold, 'Fiber-optic Ammonia Sensor for Measuring Synaptic Glutamate and Extracellular Ammonia', *Anal. Chem.*, **64**, 2438–2443 (1992).
120. A. Wang, M.A. Arnold, 'Dual-enzyme Fiber-optic Biosensor for Glutamate Based on Reduced Nicotinamide Adenine Dinucleotide Luminescence', *Anal. Chem.*, **64**, 1051–1055 (1992).

Carbohydrate Analysis: Introduction

Martin F. Chaplin
South Bank University, London, UK

1 Historical Review 3
2 Analysis 6
Abbreviations and Acronyms 6
References 6

The term ''carbohydrate analysis'' tends to have a different meaning for the different types of carbohydrates. For low-molecular-weight molecules (monosaccharides and simple oligosaccharides) it is generally taken to mean their separation and quantification, whereas for larger molecules and glycans derived from glycoproteins and glycolipids it may involve these factors but it is primarily the determination of the structure that is usually important. There are a number of electronic resources that are available to help clarify the rather specialized language used by the carbohydrate analyst and to provide databases of carbohydrate structures and nuclear magnetic resonance (NMR) spectra (Table 1).

Carbohydrates form a very diverse group of molecules usually made up from basic structures that commonly have a backbone of five or six carbon atoms and possess several alcohol groups and a hemiacetal or, more rarely, a hemiketal group. Less commonly found are carbohydrate units containing a nine-carbon backbone, whereas those with seven and eight carbons are sometimes found in bacteria. There are a number of isomers possible for these structures. As an example, a carbohydrate with six carbon units $(CH_2O)_6$ may be an aldehyde or ketone. The aldehyde form has four chiral centers allowing 16 isomers (2^4), each of which may cyclize to form either five- (furanose) or six (pyranose)-membered rings, which additionally introduces a further chiral center so totaling 64 possible structural entities. The ketone form with one less chiral position may consist of 32 structurally distinct entities. These units (*monosaccharides*) can join together by means of acetal or ketal (*glycosidic*) linkages to form oligomers (*oligosaccharides*) and polymers (*polysaccharides*), both of which may be termed glycans. These linkages can form in a number of ways between the monosaccharide units utilizing any of the hydroxyl groups and forming unbranched and multiply branched structures, such branching being either essentially ordered or disordered. Different monosaccharides may be joined together in an ordered or apparently random manner to form a mainly homogeneous or heterogeneous family of structural entities. These structures may also be linked to proteins and/or lipids as the glycans in glycoproteins, glycosaminoglycans and glycolipids. As these carbohydrates are not directly genetically coded but depend on the sequential action of enzymes, which are only capable of recognizing the local structural arrangement of up to about 12 residues or so, they are commonly found to be heterogeneous. This may be a microheterogeneity, as evident in glycoprotein glycans where most of the structure remains fixed but a variety of structures exist due to missing residues at the molecule's periphery. Alternatively, there may be sequence and linkage heterogeneity as found in heteropolysaccharides or the molecular weight heterogeneity of homopolysaccharides. Polysaccharides are inherently highly polydisperse.

Each natural glycan appears to have one of three main purposes: as a food store; playing a structural role; or as a bearer of information in molecular or cellular recognition. Their structure determines, and will be determined by, their use. Although it is recognized that some simple low-molecular-weight glycans play a structural role as part of glycoproteins where they may alter the secondary, tertiary, and quaternary structure, many structural glycans are relatively stable, of high molecular weight, anisotropic and have a single backbone with little or no complex branching. Food glycans are easily hydrolyzed and have a high proportion of nonreducing termini. Information-bearing glycans often have very complex but specific and relatively low-molecular-weight structures. Carbohydrates play a major part in all living organisms. Understanding the structure of carbohydrates helps our understanding of these biological roles.

The analytical problems surrounding the carbohydrates are clearly orders of magnitude more difficult than that of protein analysis. Indeed, it is apparent that for many polysaccharides no two molecules are likely to be identical and the concept of ''structure'' must be interpreted differently from that for proteins. There is a further level of complexity if the conformation of the carbohydrates in solution is sought because, in contrast to protein structures, these are generally not fixed but fluctuate widely and may change irreversibly during isolation or with changes in the temperature, solvent, pH, ionic strength, concentration or co-solutes.

Fortunately, although the analytical task remains formidable, the full range of monosaccharide, ring size and linkage diversity is not found, with most glycans being constructed out of a very limited subset of possible molecular building blocks (Figure 1). For example, not counting equilibrium structures, there are 30 different D-glucose dimers and 1120 different D-glucose trimers possible

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Electronic resources for carbohydrate analysis

Basic stereochemistry terminology	http://www.chem.qmw.ac.uk/iupac/stereo/
CARBBANK (structural database)	http://128.192.9.29/carbbank/CarbBank.htm
Carbohydrate nomenclature	http://www.chem.qmw.ac.uk/iupac/2carb/
Glycolipid nomenclature	http://www.chem.qmw.ac.uk/iupac/misc/glylp.html
Glycoprotein nomenclature	http://www.chem.qmw.ac.uk/iupac/misc/glycp.html
Glycosciences hyperglossary	http://www.vei.co.uk/tgn/glossary/
Glycosyl hydrolase nomenclature	http://www.expasy.ch/enzyme/
Glycosyl hydrolase families	http://www.expasy.ch/cgi-bin/lists?glycosid.txt
SUGABASE (NMR database)	http://www.boc.chem.ruu.nl/sugabase/sugabase.html

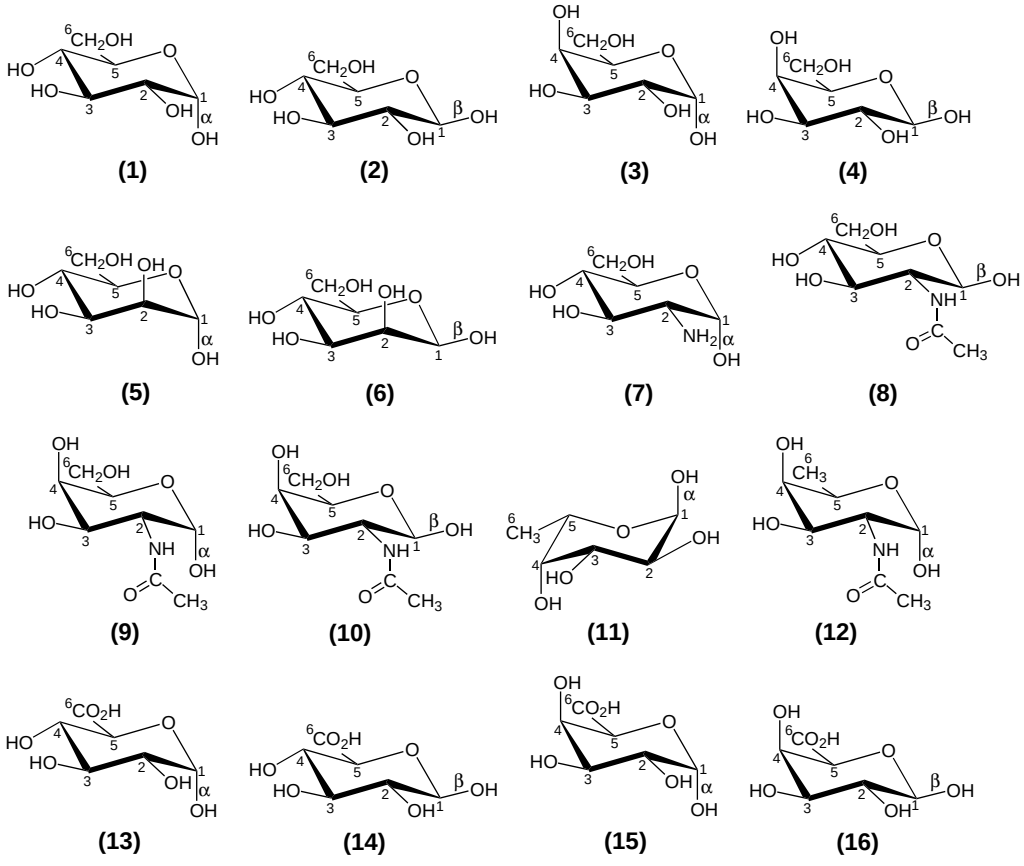

Figure 1 The commonest monosaccharides present in glycans, with the base form of their common abbreviations in parentheses: **(1)** α-D-glucose (Glc); **(2)** β-D-glucose; **(3)** α-D-galactose (Gal); **(4)** β-D-galactose; **(5)** α-D-mannose (Man); **(6)** β-D-mannose; **(7)** α-D-glucosamine (GlcN); **(8)** *N*-acetyl-β-D-glucosamine (GlcNAc); **(9)** *N*-acetyl-α-D-galactosamine (GalNAc); **(10)** *N*-acetyl-β-D-galactosamine; **(11)** α-L-fucose (Fuc); **(12)** *N*-acetyl-α-D-fucosamine (FucNAc); **(13)** α-D-glucuronic acid (GlcA); **(14)** β-D-glucuronic acid; **(15)** α-D-galacturonic acid (GalA); **(16)** β-D-galacturonic acid; **(17)** α-L-arabinose (Ara); **(18)** β-L-arabinose; **(19)** α-D-xylose (Xyl); **(20)** β-D-xylose; **(21)** α-L-iduronic acid (IdoA); **(22)** α-L-rhamnose (Rha); **(23)** 3,6-anhydro-α-D-galactose; **(24)** 3,6-anhydro-α-L-galactose; **(25)** β-D-fructose (Fru); **(26)** *N*-acetyl-β-L-talosaminuronic acid; **(27)** α-L-galactose; **(28)** β-D-galacto-furanose; **(29)** *N*-acetyl-α-neuraminic acid (NeuAc); **(30)** β-L-arabinose; **(31)** 3-deoxy-α-D-*manno*-oct-2-ulopyrasonic acid (Kdo); **(32)** β-D-mannuronic acid (ManA); **(33)** α-L-guluronic acid (GulA). Structures **(17)**, **(18)**, **(25)** and **(28)** are furanoses, the rest being pyranoses. Structures **(1–6)**, **(25)**, **(27)** and **(28)** are hexoses, **(7)** is an aminohexose, **(8–10)** are *N*-acetamidohexoses, **(11)**, **(12)** and **(22)** are deoxyhexoses, **(13–16)**, **(21)**, **(32)** and **(33)** are hexuronic acids and **(17–20)** and **(30)** are pentoses. Structures **(2)**, **(4–6)**, **(8)**, **(10)**, **(11)**, **(14)**, **(15)**, **(17)**, **(20)**, **(22)** and **(29)** are relatively common, whereas **(7)**, **(12)**, **(16)**, **(18)**, **(23)**, **(24)**, **(26–28)** and **(30–33)** are found only rarely, the remainder occurring occasionally. Other monosaccharide residues (not shown) are found very rarely, with microorganisms in particular producing a wide variety of rare monosaccharide units. Compounds are also found as derivatives with some or all of sulfate, phosphate or acetyl groups or as their methyl ethers.

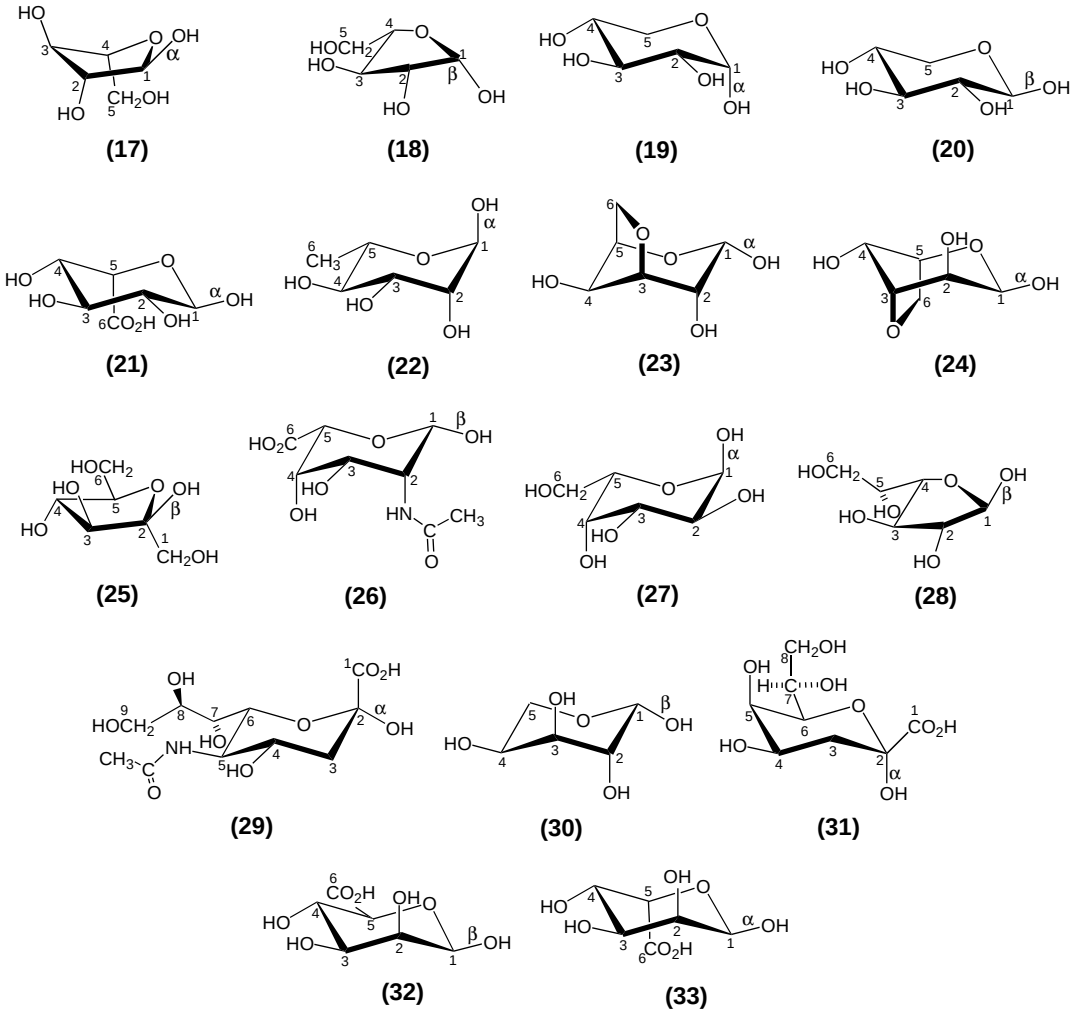

Figure 1 (*Continued*)

but only six dimers and three trimers are commonly encountered (maltose, cellobiose, isomaltose, gentiobiose, α,α-trehalose, sophorose, cellotriose, maltotriose and panose). However, the analysis of polysaccharides and glycans remains challenging and must be tackled in a systematic manner.

1 HISTORICAL REVIEW

It is surprising that carbohydrates, given the present difficulties in their structural analysis, were amongst the first natural products to be isolated, quantified and to have their structures determined. The straightforward, if flawed, analysis of mixtures of glucose, fructose and sucrose, using the rotation of polarized light, facilitated the early studies on enzyme kinetics. This determination was used by Berthelot[1] in 1860 in the first proof of the catalysis of biological reactions by nitrogenous material (enzymes), by Sørenson[2] to show the effect of pH on enzyme (invertase) activity in 1909 and by Michaelis and Menten[3] in 1913 to underpin the early development of enzyme kinetics. The determination of specific rotation is still used as a confirmatory technique and in those rare instances where mixtures of both anomeric forms of component sugars are present in complex glycans and their ratio is required.

Before powerful chromatographic methods evolved, a number of colorimetric assays were developed involving the reducing power of the low-molecular-weight sugars or by coupling to suitable molecules such as phenol after acidic dehydration.[4] Some of these assays were developed further to give different responses with different monosaccharides. They remain of considerable use to this day, particularly where just the concentration of a known carbohydrate is required. Where selectivity is required, these assays are replaced by colorimetric or fluorimetric assays involving enzymatic conversions.

Early analytical work in the 1940s and 1950s centered on the use of paper chromatography to separate the component sugars from acid-hydrolyzed polysaccharides, with much effort being applied to the selection of the developing solvents and the detection methodology. Although paper chromatography is easy and inexpensive, can separate a wide variety of structural types, including oligosaccharides and some polysaccharides, without their prior derivatization and generally ensures that no components go undetected, it is only rarely used nowadays. In the 1960s and 1970s, paper chromatography was developed into thin-layer chromatography with its greater resolving power, and in the 1980s to high-performance thin-layer chromatography. These improved methodologies possess very much greater resolving power and improved quantification but are little used in modern carbohydrate analysis, apart from their important utility for separating glycolipids, having been generally outclassed by high-performance liquid chromatography (HPLC) techniques.

Another development from paper chromatography was the use of column chromatography. Initially these consisted of charcoal or cellulose, and then the ion exchange and ion exclusion matrices that have evolved over the years into the high-performance columns used today. Although these separation techniques generally work well, a major problem has been in the detection of carbohydrates. Unlike proteins, carbohydrates have no natural ultraviolet absorption, color or fluorescence. Free aldehyde or ketone groups do absorb ultraviolet light but only below about 200 nm, and then rather inefficiently. This does allow their detection but only when noninterfering eluents, such as pure water or acetonitrile, are used. Refractive index detectors have been commonly used but these are rather insensitive and easily confused by changes in eluent composition such as solute gradients. Many pre- and post-column derivatization processes have been devised in order to facilitate sensitive spectrophotometric determination. Relatively recently, the introduction of pulsed amperometric detection has driven forward the use of HPLC for carbohydrate analysis. This methodology in which about 1% of carbohydrate alcohol groups are oxidized at a gold electrode when in alkaline solution, has proved to be both sensitive and reliable. In this system the gold electrode is utilized and cleaned about once every 1 s by a series of oxidative and reductive electrical pulses. This methodology has been coupled to the development of alkali-stable anion exchange resins for use in high-performance anion exchange chromatography (HPAEC).

Improved separation techniques are constantly being developed with very high resolutions and capable of distinguishing even quite closely related glycans. When coupled to detection by mass spectrometry (MS), they can form a powerful part of the carbohydrate analytical toolbox.

Gas chromatographic methods were commonly used for the analysis of monosaccharides in the 1960s, when their high sensitivity and resolution placed them well ahead of the, then current, liquid chromatographic methods. Their development depended on methodology, based on trimethylsilylation or the acetylation of their alditols, to convert the inherently involatile carbohydrates into volatile derivatives. Although currently often replaced by HPAEC, gas chromatography methods are still used because the technique is well understood. Gas chromatography is particularly useful for the analysis of partially methylated alditol acetates produced by methylation analysis and where coupled MS is necessary.

The absolute configuration (D or L) of monosaccharides was an important part of early analyses, involving a considerable degree of effort and ingenuity. This was achieved via the relatively large-scale purification of hydrolysis-derived monosaccharides, followed by their chemical characterization. Where possible, the configuration was decided or confirmed by X-ray crystallography. Once sufficient absolute structures had been established, materials were determined by comparison of their specific optical rotation. Much simpler chiral chromatographic methods now exist, but absolute configurations are only rarely specifically established, because most work is on materials where this is already known or may be assumed from, for example, the use of enzymes of known specificity.

Linkage analysis was first determined, in the 1940s, by means of the analysis of fragments produced by periodate. This method, which depends on the ability of periodate to cleave between neighboring unlinked alcohol groups, is still used occasionally to help solve otherwise intransigent problems. A useful development of the technique is the Smith degradation,[5] where (nowadays) borohydride reduction of the resultant aldehydes and chromatography of the resultant alcohols follow oxidation. The introduction of a powerful and efficient methylation procedure by Hakomori[6] in 1964 allowed linkage positions to be determined more explicitly, and developments of this methodology, coupled with the use of gas chromatography/mass spectrometry (GC/MS), are commonly used today. Linkage analysis normally must be combined with other techniques such as partial hydrolysis and NMR if more than localized structural arrangements are to be determined.

Determination of the anomeric configuration has always been achieved by the use of specific enzymes. The use of enzymes also confirms the absolute configurations. Sequence analysis has always been problematic. Glycoprotein glycans, where the monosaccharide units are linked together in precise and generally well-understood sequences, can be investigated successfully by the controlled and sequential addition of specific

exoglycosidases. Advances over the years have concerned the range of available enzymes and the more detailed knowledge concerning their specificity. Although suitable enzymes are often not easily available, this is an important area of research and there is likely to be considerable progress in the future. Recently, NMR has been used to distinguish the anomeric configurations and sequence of oligosaccharides in many cases.

The determination of the molecular weight of carbohydrate-containing polymers has been decidedly difficult until recently. This is because they may be very large, they hold on to large amounts of water, they may possess very anisotropic structures and there are few ''standard'' molecules with known structures that are suitable for comparison with the unknown substances. Analytical ultracentrifugation was commonly used in the 1940s and 1950s but is rarely used nowadays. Size exclusion chromatography has proved to be very useful in determining the size of proteins but is of limited use for large carbohydrates, except as a means of separation, owing to the very different ways in which carbohydrate-containing molecules behave in solution and the typical inapplicability of the available calibration molecules. Although light scattering was one of the earliest methods available, in the 1940s, for determining molecular weights in solution, it is only the recent introduction of multi-angle laser light scattering (MALLS) that has made this technique widely accepted. MALLS allows the relatively straightforward and accurate determination of the absolute weight-averaged molecular weight (M_w) of large carbohydrate-containing molecules without the need for standards. In its most powerful mode it can be used as part of HPLC detection to monitor continuously the molecular weights and shapes of the materials eluted. Number-averaged molecular weights (M_n) and polydispersity (M_w/M_n: a measure of the breadth of the molecular weight distribution) can also be estimated by this method.

Although NMR has been used for many years, the one-dimensional spectrum of most carbohydrates is generally too complex to be assigned fully. There remain, however, sufficient ''reporter'' groups for much useful information to be obtained. In particular, the monosaccharides present, their anomeric conformations and ring size (i.e. pyranose or furanose) and some indication of the linkages present are usually determinable. With the development of multidimensional techniques and powerful (600-MHz) instrumentation, NMR is proving to be a dominant technique for determining complex protein structures but has not yet had a similar level of success with large carbohydrates. The spectra for medium or large molecules are extremely complex and difficult to interpret but NMR has proved valuable for determining or confirming the structure of oligosaccharide fragments, and a database of structural assignments is being built up. NMR is capable of determining the sequence, linkage position and anomeric configuration. Although it is, potentially, a most powerful method for determining the conformation of carbohydrates, this use of NMR is still in its infancy with respect to glycans. A major drawback for the technique is its lack of sensitivity, particularly as many polysaccharides have limited solubility, form gels or are relatively inflexible in solution. Often such molecules give rise to poor spectra owing to their unfavorable relaxation properties.

The specific binding of antibodies or lectins to carbohydrates has been used to show the existence and availability of particular substructures within complex carbohydrates and may be used in carbohydrate separation and purification. As the family of known lectins has grown over the years and their characterization has improved, they have become useful tools for distinguishing between closely related glycans without the need for full analyses.

MS can be used to determine the molecular weight of low- and medium-molecular-weight carbohydrates of up to about 10 kDa. For some molecules, tandem mass spectrometry (MS/MS) can also give compositional, sequence and linkage information. Little success has been reported for higher molecular weight polysaccharides, perhaps due to the inherent heterogeneity of carbohydrate samples. The main restriction with MS is that isomeric monosaccharides, such as glucose and galactose, cannot be distinguished. The technique can be very sensitive, with only subpicomole amounts being required. Electrospray ionization MS may be used as a detection system for HPLC, capillary electrophoresis and capillary isoelectric focusing. Such systems are very powerful in that they can achieve two-dimensional separations based on MS and chromatography.

The area of conformational analysis remains one of the most intransigent problems in carbohydrate analysis. This is partly because of the vast number of possible local conformational potential energy minima possible: for just four 1→4-linked hexopyranoses joined together, there are about 3^{25} (approximately 1 trillion) possible local potential energy minima. For medium and large glycans the number of local minima is so vast that it is possible for only a fraction of the conformations to be sampled during their existence. In addition, a significant number of these conformations are in deep potential energy wells. Most sizeable molecules will therefore consist of mixtures, with little chance that any two molecules are conformationally identical unless they possess long-range order such as the presence of helices. NMR is capable of determining the conformation of these carbohydrates but will deliver time-averaged conformations that may or may not be true reflections of the actual conformations. In such circumstances, although there may be little meaning to any one ''averaged'' conformation, knowledge concerning the typical conformational motifs can be a very useful

aid in discussing the glycan's molecular interactions and functional properties. Additionally, where multiple conformational states exist and are in rapid equilibrium (i.e. usually), the lack of unique torsional angles and time averaging of the NMR data combine to blur the conclusions that can be drawn. Molecular modeling is helping for making inroads into solving this conformation problem. Its present restricted usefulness, due to the vast memory and computational requirements, is being rapidly overcome.

2 ANALYSIS

No single method suffices for the analysis of glycans, in contrast to protein analysis where either X-ray crystallography or NMR may be sufficient. In addition, crystalline protein structures appear to be very similar to their structure in solution, whereas carbohydrates that take up very different conformations with no clearly defined structure in solution may form a regular structure in the solid state. Isolation and purification are also more likely to affect the structure and conformation of high-molecular-weight carbohydrates than proteins. The structural determination of most complex carbohydrates utilizes a wide variety of methodologies, often including the use of enzymes, methylation analysis, MS and NMR; these methods complement and reinforce each other. A complete structural determination of an unknown glycan, subsequent to its isolation and purification, involves the following stages, which are often undertaken approximately in the order given:

- determination of the constituent monosaccharides, their absolute configuration (D or L), their ring size and any derivatization they have undergone, such as sulfation, methylation or acetylation;
- determination of the molecular size or molecular size distribution;
- determination of the anomeric configuration of the linkages (α or β);
- determination of the linkage positions between the monosaccharide units;
- determination of their sequence, including any repetitive pattern in the structure and the degree of any heterogeneity;
- determination of any preferred conformation of parts or all of the molecule(s) present.

The more important analytical methods appropriate to different carbohydrates are given in Table 2. A modern carbohydrate analytical facility clearly involves the availability of a range of expensive instrumentation and staff skilled in a variety of chemical and biochemical techniques.

Table 2 Major analytical methods for carbohydrates

Monosaccharides	Colorimetric assays, GC/MS, HPAEC
Oligosaccharides	HPAEC, high-performance thin-layer chromatography, methylation analysis,[a] MS, NMR
Polysaccharides	Enzymatic cleavage,[a] methylation analysis, size exclusion chromatography plus MALLS
Glycolipids	HPLC, high-performance thin-layer chromatography, methylation analysis, MS, NMR
Glycoprotein glycans	Capillary electrophoresis, enzymatic cleavage, HPLC, methylation analysis, MS, NMR, use of lectins
Glycosaminoglycans	Capillary electrophoresis, enzymatic cleavage, extractive methods

[a] Methods involving cleavage produce monosaccharides and/or oligosaccharides.

This section of the Encyclopedia (i.e. carbohydrate analysis) has articles on all the main classes of carbohydrates selected on the basis of class rather than methodology. Each class (**Monosaccharides and Sugar Alcohol Analysis; Disaccharide, Oligosaccharide and Polysaccharide Analysis; Proteoglycan and Acidic Polysaccharide Analysis; Glycoprotein Analysis: Using Nuclear Magnetic Resonance; Glycoprotein Analysis: General Methods** and **Glycolipid Analysis**) has its own particular analytical direction, although all involve analysis of their component monosaccharides. They may use related methodology for part of their analytical schemes but they also use their own characteristic methods based on their distinct structures and properties.

ABBREVIATIONS AND ACRONYMS

GC/MS	Gas Chromatography/Mass Spectrometry
HPAEC	High-performance Anion Exchange Chromatography
HPLC	High-performance Liquid Chromatography
MALLS	Multi-angle Laser Light Scattering
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance

REFERENCES

1. P.E.M. Berthelot, 'On the Glucose Fermentation of Cane Sugar', *C. R. Acad. Sci.*, **50**, 980–984 (1860).
2. S.P.L. Sørenson, 'Enzymes Studies II. Measurement and Significance of Hydrogen Ion Concentration in Enzyme Processes', *Medd. Carlsberg Lab.*, **8**, 1–153 (1909).

3. L. Michaelis, M.L. Menten, 'The Kinetics of Invertin Action', *Biochem. Z.*, **49**, 333–369 (1913).
4. M. Dubois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, 'Colorimetric Method for Determination of Sugars and Related Substances', *Anal. Chem.*, **28**, 350–356 (1956).
5. F. Smith, J.W. Van Cleve, 'Reduction of the Products of Periodate Oxidation of Carbohydrates. I. Hydrogenation with Raney Nickel of the Dialdehydes from the Methyl Glycopyranosides', *J. Am. Chem. Soc.*, **77**, 3091–3096 (1955).
6. S. Hakomori, 'A Rapid Permethylation of Glycolipids, and Polysaccharide Catalyzed by Methylsulfinyl Carbanion in Dimethyl Sulfoxide', *J. Biochem.*, **55**(2), 205–208 (1964).

Disaccharide, Oligosaccharide and Polysaccharide Analysis

Anne D. Blackwood and Martin F. Chaplin
South Bank University, London, UK

1 Introduction **1**
2 Extraction **4**
3 Component Analysis **4**
4 Enzyme and Colorimetric Assays **5**
5 Biochemical Methods of Analysis **5**
5.1 Separation Methods 5
5.2 Methylation Analysis 10
5.3 Periodate Oxidation 10
5.4 Hydrolysis with Specific Glycosidases 11
6 Molecular Weight Analysis **12**
6.1 Size Exclusion Chromatography 13
6.2 Light Scattering 13
7 Physical Methods of Analysis **15**
7.1 Mass Spectrometry 15
7.2 Nuclear Magnetic Resonance Spectroscopy 17
7.3 Infrared Spectroscopy 21
List of Symbols **21**
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

Disaccharides, oligosaccharides and polysaccharides form a very diverse and complex family of biologically important macromolecules. In order to discover their structure it is necessary to determine not only monosaccharides present and their linkage positions and sequence, but also the anomeric configuration of linkages, the ring size (furanose or pyranose), the absolute configuration (D or L) and identify any other substituents present. There is no one method that can be used to determine their fine structure; instead we must use a combination of analytical methods to gain as much information as we can. For analysis of disaccharide and oligosaccharide mixtures, separation and quantification techniques such as colorimetric and enzymatic assays, thin-layer chromatography (TLC) or high-performance anion-exchange chromatography may be applied. For polysaccharide analysis separation and extraction techniques, component analysis, methylation analysis, glycosidic hydrolysis, mass spectrometry (MS) methods, and nuclear magnetic resonance (NMR) spectroscopy are required to determine the fine structure and size exclusion chromatography (SEC) coupled with light scattering (LS) detectors to determine the molecular size distribution.

1 INTRODUCTION

Disaccharides are compounds in which two monosaccharide units are joined by glycosidic linkages. Oligosaccharides are broadly defined as carbohydrate polymers containing up to twenty residues. According to the number of units, they are called trisaccharides, tetrasaccharides, pentasaccharides etc. where the number of units is the degree of polymerization (DP, e.g. tetrasaccharides have a DP of 4). The borderline between oligosaccharides and polysaccharides is not precise, however, and oligosaccharides are generally molecules with precisely defined structures whereas polysaccharides have inexactly defined length or structure.

Polysaccharides form a diverse and complex family of biological macromolecules. The source and structure of some commonly occurring polysaccharides are given in Table 1. They are essentially condensation polymers of monosaccharide units joined by glycosidic linkages. Polysaccharides may be composed of a single sugar unit (termed *homopolymers*) or may contain two or more constituent monosaccharides (*heteropolymers*) joined in a variety of ways. The polymers may be linear, branched or even cyclic in nature and this has important consequences for their physicochemical nature. Polysaccharides are abundant in nature and found in nearly every living organism. They can function as storage polysaccharides, such as starch in plants and glycogen in animals, structural elements in the cell walls of bacteria and plants and skeletal elements in arthropods. They can also be found linked to proteins and lipids, and have been implicated in cell–cell recognition and molecular targeting. Glycoproteins and glycolipids will not be reviewed here as they are discussed in separate articles.

Polysaccharides occurring in nature often have high molecular weights, frequently well in excess of 1000 kDa. They are usually polydisperse in nature and may contain a wide range of closely related structures. The generic term for all polysaccharides is *glycans*. They are often individually classified by adding '*an*' to the end of the constituent prefix for the parent sugar, for example *glucans* for polymers of glucose, *xylans* for polymers of xylose and so on. However many common unsystematic names, such as cellulose, pectin, amylose and inulin, are still in use alongside the systematic ones. Heteropolymers are commonly classified

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Source and structure of some important polysaccharides

Main sources	Polysaccharide	Polymer structure
Land plants	Arabinoxylan	β-(1 → 4)-linked D-xylose with α-L-(1 → 2), (1 → 3) or (1 → 2)(1 → 3)-linked L-arabinose
	Cellulose	β-(1 → 4)-linked D-glucan
	Galactomannan	α or β linked D-mannan substituted with α or β linked D-galactose
	Glucomannan	β-(1 → 4)-linked D-mannose with β-(1 → 4)-linked D-glucose
	Guar gum	α-(1 → 4)-linked D-galactomannan
	Gum arabic	L-Arabino-, α-(1 → 3)- and α-(1 → 6)-D-galactan with glucuronic acid side chains
	Inulin	β-(2 → 1)-linked D-fructans
	Pectin	Linear and branched α-(1 → 4)-linked galacturonan
	Starch	α-(1 → 4)-linked D-glucose either linear (amylose) or branched (amylopectin) with α-(1 → 6)-linked D-glucose
	Xylan	β-(1 → 4)-linked D-xylose often with substituent groups attached
Marine plants	Agar	Sulfated DL-galactans, contains 3,6-anhydrogalactose units
	Alginates	Linear β-(1 → 4)-D-mannuronan and α-L-guluronan
	Carrageenan	Sulfated β-(1 → 3), α-(1 → 4)-D-galactans
	Laminaran	β-(1 → 3)-linked D-glucan
Micro-organisms	Curdlan	β-(1 → 3)-linked D-glucan
	Dextran	α-(1 → 6), (1 → 3)-linked D-glucan, and sometimes α-(1 → 2) and α-(1 → 4) links
	Levan	β-(2 → 6), (2 → 1)-linked D-fructan
	Mannan	α-(1 → 2), α-(1 → 6)-linked D-mannans, sometimes α-(1 → 3) links
	Nigeran	α-(1 → 3), (1 → 4)-linked D-glucan
	Pullulan	α-(1 → 4), (1 → 6)-linked D-glucan
	Xanthan gum	β-(1 → 4)-linked D-glucan with mannose and glucuronic acid side chains
Animals	Chitin	β-(1 → 4)-linked *N*-D-acetylglucosamine
	Glycogen	α-(1 → 4)-linked D-glucose with some α-(1 → 6)-linked branch points

by placing the branched or lowest unit first and the parent backbone sugar last; for example, rhamnogalactans are polymers with a galactose backbone and rhamnose side-chains.

Polysaccharides can be classified in terms of their monosaccharide composition, type of glycosidic linkages present, sequence of units along the backbone, the relative degree and spacing of the substituted groups along the chain and the DP. Many of the structure–function relationships of carbohydrates have been poorly elucidated due to the difficulty in extraction, purification and fine structure characterization of these complex polymers. However the availability of highly specific glycosidases to break down the polysaccharides backbone and recent advances in techniques such as high-performance liquid chromatography (HPLC), MS and NMR spectroscopy, have enabled a number of fine structures to be elucidated. It is hoped that detailed characterization of these molecules will lead to a greater understanding of their biological role.

Disaccharides and oligosaccharides form a varied class of hydrophilic low-molecular-weight molecules (Table 2). A number of these are nonreducing sugars used as mobile stores to provide for the synthetic needs of their source organism without risk of nonspecific reactions otherwise associated with free reducing termini. Others are mainly found as the hydrolytic products of polysaccharides and other glycans.

In order to discover the structure of glycans, it is necessary to determine not only the monosaccharides present, their linkage positions and sequence, but also the anomeric configuration of linkages, the ring size (furanose or pyranose), the absolute configuration (D or L) and to identify any other substituents present. There is no one method that can be used to determine the fine structure of polysaccharides; instead we must use a combination of analytical methods to gain as much information as we can. For analysis of disaccharide and oligosaccharide mixtures, separation and quantification techniques such as colorimetric and enzymatic assays (section 4), TLC, section 5.1.1) or HPLC, (section 5.1.2) may be applied. For polysaccharide analysis separation and extraction techniques (section 2), component analysis (section 3), methylation analysis (section 5.2), glycosidic hydrolysis (section 5.3), MS methods (section 7.1) and ^{1}H and ^{13}C NMR spectroscopy (section 7.2) are required to determine the fine structure and SEC (section 6.1) coupled with LS (section 6.2) to determine the molecular size distribution.

Table 2 The more commonly encountered disaccharides, trisaccharides and oligosaccharides

Common name	CAS number	Structure
Ajugose		α-D-Gal*p*-(1 → 6)-[α-D-Gal*p*-(1 → 6)]$_3$-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*
Cellobiose	528-50-7	β-D-Glc*p*-(1 → 4)-D-Glc
Cellotriose	6817-81-8	β-D-Glc*p*-(1 → 4)-β-D-Glc*p*-(1 → 4)-D-Glc
Chacotriose	550-72-1	α-L-Rha*p*-(1 → 2)-[α-L-Rha*p*-(1 → 4)]-D-Glc
Chitobiose	577-76-4	β-D-Glc*pN*Ac-(1 → 4)-D-Glc*N*Ac
α-Cyclodextrin[a]	10016-20-3	→ 4)-[α-D-Glc*p*]$_6$-(1–
β-Cyclodextrin[b]	7585-39-9	→ 4)-[α-D-Glc*p*]$_7$-(1–
γ-Cyclodextrin[c]	17465-86-0	→ 4)-[α-D-Glc*p*]$_8$-(1–
Galactosucrose		β-D-Fru*f*-(2 ↔ 1)-α-D-Gal*p*
Gentianose		β-D-Glc*p*-(1 → 6)-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*
Gentiobiose	554-91-6	β-D-Glc*p*-(1 → 6)-D-Glc
Isokestose[d]	470-69-9	β-D-Fru*f*-(2 → 1)-β-D-Fru*f*-(2 ↔ 1)-α-D-Glc*p*
Isomaltose	499-40-1	α-D-Glc*p*-(1 → 6)-D-Glc
Isomaltotriose	3371-50-4	α-D-Glc*p*-(1 → 6)-α-D-Glc*p*-(1 → 6)-D-Glc
Isopanose		α-D-Glc*p*-(1 → 4)-[α-D-Glc*p*-(1 → 6)]-D-Glc
Isotrehalose[e]		β-D-Glc*p*-(1 ↔ 1)-β-D-Glc*p*
Kestose[f]	562-68-5	β-D-Fru*f*-(2 → 6)-β-D-Fru*f*-(2 ↔ 1)-α-D-Glc*p*
Kojibiose		α-D-Glc*p*-(1 → 2)-D-Glc
Lactose	63-42-3	β-D-Gal*p*-(1 → 4)-D-Glc
Lactulose	4618-18-2	β-D-Gal*p*-(1 → 4)-D-Fru
Laminaribiose	34980-39-7	β-D-Glc*p*-(1 → 3)-D-Glc
Laminaritriose		β-D-Glc*p*-(1 → 3)-β-D-Glc*p*-(1 → 3)-D-Glc
Maltose	69-79-4	α-D-Glc*p*-(1 → 4)-D-Glc
Maltotriose	1109-28-0	α-D-Glc*p*-(1 → 4)-α-D-Glc*p*-(1 → 4)-D-Glc
Maltulose		α-D-Glc*p*-(1 → 4)-D-Fru
Manninotriose	13382-86-0	α-D-Gal*p*-(1 → 6)-α-D-Gal*p*-(1 → 6)-D-Glc
Melezitose	597-12-6	α-D-Glc*p*-(1 → 3)-β-D-Fru*f*-(2 ↔ 1)-α-D-Glc*p*
Melibiose	585-99-9	α-D-Gal*p*-(1 → 6)-D-Glc
N-Acetyllactosamine	32181-59-2	β-D-Gal*p*-(1 → 4)-D-Glc*N*Ac
Neokestose	3688-75-3	β-D-Fru*f*-(2 → 6)-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*
Nigerose	34980-39-7	α-D-Glc*p*-(1 → 3)-D-Glc
Nystose	13133-07-8	β-D-Fru*f*-(2 → 1)-β-D-Fru*f*-(2 → 1)-β-D-Fru*f*-(2 ↔ 1)-α-D-Glc*p*
Palatinose		α-D-Glc*p*-(1 → 6)-D-Fru
Panose	33401-87-5	α-D-Glc*p*-(1 → 6)-α-D-Glc*p*-(1 → 4)-D-Glc
Planteobiose		α-D-Gal*p*-(1 → 6)-D-Fru
Planteose		α-D-Gal*p*-(1 → 6)-β-D-Fru*f*-(2 ↔ 1)-α-D-Glc*p*
Primaverose		β-D-Xyl*p*-(1 → 6)-D-Glc
Raffinose	512-69-6	α-D-Gal*p*-(1 → 6)-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*
Rutinose	90-74-4	α-L-Rha*p*-(1 → 6)-D-Glc
Solatriose	528-40-5	α-L-Rha*p*-(1 → 2)-[β-D-Glc*p*-(1 → 3)]-D-Gal
Sophorose	534-46-3	β-D-Glc*p*-(1 → 2)-D-Glc
Stachyose	470-55-3	α-D-Gal*p*-(1 → 6)-α-D-Gal*p*-(1 → 6)-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*
Sucrose	57-50-1	β-D-Fru*f*-(2 ↔ 1)-α-D-Glc*p*
α,α-Trehalose	99-20-7	α-D-Glc*p*-(1 ↔ 1)-α-D-Glc*p*
Turanose	547-25-1	α-D-Glc*p*-(1 → 3)-D-Fru
Umbelliferose		α-D-Gal*p*-(1 → 2)-α-D-Gal*p*-(1 ↔ 2)-β-D-Fru*f*
Verbascose	546-62-3	α-D-Gal*p*-(1 → 6)-[α-D-Gal*p*-(1 → 6)]$_2$-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*

[a] α-cyclomaltohexaose.
[b] β-cyclomaltoheptaose.
[c] γ-cyclomaltooctaose.
[d] 1-kestose.
[e] β,β-trehalose.
[f] 6-kestose.

2 EXTRACTION

The first step in structural polysaccharide analysis is the extraction and separation of a single polysaccharide species from extraneous components. Polysaccharide material must be separated from low-molecular-weight material such as inorganic salts and fat and from other high-molecular-weight material commonly found associated with polysaccharides such as protein. Unless the level of these extraneous components is already known in the sample material, then these must be determined prior to extraction. The key stage is the selective extraction of the appropriate polysaccharide in as high a yield as possible, with minimum effect on the polysaccharide fine structure. The amount of carbohydrate present at each stage of the purification procedure can be monitored using a non-specific carbohydrate assay such as the phenol–sulfuric assay.[1]

Water-extractable polysaccharides can be solubilized in aqueous solutions, and low-molecular-weight material removed by dialysis, ultrafiltration or treatment with ion exchange resins. Other carbohydrate components may be selectively degraded at this stage using the appropriate glycosidase e.g. starch can be removed by enzymatic degradation with α-amylases. The enzymes used must not cause degradation of the polysaccharide of interest. Polysaccharide material can then be isolated from aqueous solutions by graduated fractionation with, for example, ammonium sulfate or ethanol or by lyophilization.

Water-unextractable polysaccharide material can often be solubilized in alkali. If required, extraneous components can be removed prior to alkaline extraction e.g. fat can be solubilized in an appropriate solvent such as chloroform–methanol (95 : 5 v/v) using Soxhlet apparatus and protein can be denatured by suspension in aqueous sodium dodecylsulfate. The extraction step must be carefully designed to minimize base-catalyzed hydrolysis, β-elimination and Lobry de Bruyn–Eckenstein transformations (degrading the reducing end of the molecule), which may otherwise occur under alkaline conditions. It is useful to use an extractant, if possible, which will preferentially select for the carbohydrate of interest. This will simplify the purification procedure. An example is the use of saturated barium hydroxide solution, containing a critical concentration of sodium borohydride, to selectively extract insoluble arabinoxylans from wheat cell wall material.[2] If a sufficiently selective extractant cannot be found, then it may be necessary to introduce further purification steps, such as separation on ion-exchange resins e.g. diethylaminoethyl (DEAE)-cellulose, in order to separate polysaccharide species or provide a sample of sufficient homogeneity for analysis.

3 COMPONENT ANALYSIS

There are many different stereoisomers of hexoses or pentoses that are possible but fortunately for the analyst only a few are commonly found in polysaccharides. These are listed in Table 3. Analysis of the monosaccharide composition of carbohydrates requires complete hydrolysis of all the glycosidic linkages present in the sample. Polysaccharides containing furanoside units such as fructans and arabinans can be completely and rapidly hydrolyzed by treatment with mild acid. Polysaccharides containing pentose sugars as pyranoside units and even more so those containing hexopyranoside units are rather more resistant to hydrolysis. They are still relatively easily hydrolyzed using e.g. 2 M HCl for a period of 2–5 h at 100 °C. If the polysaccharide material is not water-soluble, it may be pre-treated with 72% sulfuric acid for a short period at low temperature to facilitate solubilization followed by dilution and hydrolysis as above. Some residues involved in very acid-resistant glycosidic linkages may not be released by mild acid hydrolysis and may require more stringent techniques. For example uronic-acid-containing polysaccharides, particularly with the aldobiuronic acid type of linkage, require combined treatment of methanolysis (2 M HCl in methanol) followed by hydrolysis with trifluoroacetic acid to release all the components present.[3] The glycosidic linkages of aminosugar residues are also very resistant to acid hydrolysis. In general α-glycosidic bonds are more acid-labile than β-glycosidic bonds.

Analysis of the released monosaccharide components can be identified, after any appropriate derivatization, by a number of chromatographic techniques including TLC, gas chromatography (GC) or HPLC. Driven by improvements, particularly in the type and sensitivity of detectors available, the method of choice for analyzing component sugar composition is probably high-performance anion-exchange chromatography coupled with pulsed amperometric detection (HPAEC/PAD).[4] Column chromatography of the monosaccharides and their derivatives are discussed in

Table 3 Some common sugars occurring in polysaccharides

Type	Monosaccharide
Pentoses	L-Arabinose, D-xylose
Hexoses	D-Galactose, D-glucose, D-mannose, D-fructose, L-galactose
6-Deoxyhexoses	L-Rhamnose
Hexuronic acids	D-Glucuronic acid, D-galacturonic acid, 4-*O*-methyl-D-glucuronic acid
Hexosamines	*N*-Acetyl-D-glucosamine, *N*-acetyl-D-galactosamine

full in the article **Monosaccharides and Sugar Alcohol Analysis**.

4 ENZYME AND COLORIMETRIC ASSAYS

There are a number of rapid and convenient colorimetric assays available for determination of carbohydrate groups in intact molecules. These tend to be based either on determination of the number of reducing end groups, or on the creation of a chromophore by complexation with the sugar under acidic dehydrating conditions. General carbohydrate assays such as the phenol–sulfuric acid assay are often used to monitor the carbohydrate content of column effluents to give a quick and quantitative measure of, for example, free and combined glucose.[1] Calibration with appropriate standards is necessary, as different sugars will give varying color responses with this assay. More specific assays can be used to detect the presence of particular kinds of sugar such as hexoses, pentoses, uronic acids or aminosugars. For example, pentose can be measured with a modified version of the L-cysteine–sulfuric acid assay,[5] uronic acids using a modification of the carbazole assay[6] and hexoses by any of the general carbohydrate assays. For reducing sugars either the Nelson–Somogyi method[7] or the more convenient dinitrosalicylic acid assay can be used.[8] As with many of the less specific colorimetric assays the reducing sugar assays do not differentiate between sugars and can give varying color responses. Accurate quantitative determination of mixtures of monosaccharides or hydrolyzed polymer structures may be better determined using the more powerful column chromatography techniques such as HPLC.

There are a number of specific colorimetric enzymatic assays available for the determination of individual carbohydrate groups. These are ideal for a quantitative analysis of known carbohydrates in, for example, complex mixtures of food samples. There are a large number of enzymes available for oligosaccharide and polysaccharide analysis, a full summary of which is given in section 5.4. For example determination of starch or sucrose may be achieved by enzymatic hydrolysis to monosaccharides with glucoamylase or invertase respectively, followed by enzymatic determination of glucose using the hexokinase/dehydrogenase assay.[9] There are commercially available kits for the enzymatic detection of sucrose, maltose, lactose, and starch. There are also an increasing number of commercial biosensors available not just for glucose analysis[10] but also for other carbohydrates such as maltose, sucrose and starch for which suitable enzyme systems exist.[11] These are particularly useful for monitoring the effluent in on-line automated HPLC processes.

5 BIOCHEMICAL METHODS OF ANALYSIS

5.1 Separation Methods

5.1.1 Thin-layer Chromatography

TLC can be used as a convenient and semi-quantitative method of analyzing the oligosaccharide hydrolysates of polysaccharides. It can provide a rapid preliminary screening of mixtures of monosaccharides and oligosaccharides giving an indication of the component sugars present and of their relative amounts. This gives valuable information about the composition of the native polysaccharide. Due to the sensitivity of the procedure, even minor components can be detected. Carbohydrates have been separated using a number of different solid supports and an array of solvent systems,[12,13] but no single method is universal and trial-and-error must sometimes be adopted to find appropriate separation conditions. Thin-layer plates can be prepared in the laboratory but, for reproducibility, are best purchased pre-coated with microcrystalline cellulose or silica gel as the solid support. Sugars are separated essentially by liquid–liquid partition, with some element of adsorption in the case of silica gel. Impregnation of silica gel with inorganic salts has been shown to improve resolution with sugar separations. For more complex mixtures of sugars where the R_f values are small, resolution can be improved by performing multiple ascents where the chromatogram is eluted, dried and re-eluted using the same eluent a number of times.

The relative movement of the carbohydrate (R_f) and the efficiency of separation may be affected by laboratory conditions such as humidity, temperature and chromatographic tank size. Variations in the coating batch and coating method may also adversely affect R_f values. Hence, the R_f values for standards are always determined alongside the unknowns and values quoted in the literature used for guidance only. In general, the lower the molecular weight or the more hydrophobic the carbohydrate the higher the R_f value.

Detection of carbohydrates for TLC is commonly based on dehydration and complexation after the chromatogram has been sprayed with an appropriate reagent and heated to 100–150 °C for 5–10 min. Suitable reagents include those based on diphenylamine/aniline/phosphoric acid or naphthoresorcinol/ethanol/sulfuric acid. A wide variety of colored responses is given, depending on the nature of the sugars present. TLC can be used semi-quantitatively using scanning densitometric analysis. The interpretation of such results may be misleading as color development will vary with spray technique, temperature and duration of heating and traces of salts, proteins and other impurities present in the sample.

TLC has been used to separate complex mixtures of cyclodextrins and their derivatives on silica gel.[14] Cyclodextrins, their glucosyl derivatives, and linear oligosaccharides were separated by eluting with mobile phases containing aqueous ammonia and either acetonitrile (ACN) or 1-propanol (Table 4). Compounds with different numbers of substituents on the cyclodextrins could be clearly separated, though the system did not allow separation of individual isomers.

Oligosaccharides may be purified on the milligram scale by high-performance thin-layer chromatography (HPTLC). HPTLC plates are available pre-coated from commercial suppliers, and are characterized by a much narrower particle size and particle size range resulting in improved resolution. Analysis by densitometry is also more satisfactory with HPTLC plates than conventional TLC plates.

5.1.2 High-performance Liquid Chromatography

For the qualitative and quantitative separation and analysis of mixtures of disaccharides and oligosaccharides the two most widely used techniques available to the researcher are HPLC and GC. The choice between the two methods is not often clear cut, but in general HPLC has been preferred for the separation and analysis of oligosaccharide mixtures, while GC has been used for more complex mixtures of monosaccharides and disaccharides. Recent advances in HPLC techniques, particularly in the type and sensitivity of detectors available, have seen it become the method of choice for many different types of analyses. In practice, though, GC is still in common use and there are many good examples of disaccharide and oligosaccharide separations in the literature.[15,16] SEC, used for the separation of polysaccharides and preliminary screening of mixtures of oligosaccharides and unknown samples, is discussed in section 6.1.

Table 4 Resolutions of cyclodextrins, their glucosyl derivatives, and linear oligosaccharides by TLC on Kieselgel 60

Compound	R_f (ACN system)[a]	R_f (propanol system)[a]
α-Cyclodextrin	0.18	0.21
β-Cyclodextrin	0.14	0.18
γ-Cyclodextrin	0.11	0.14
Glucosyl-β-cyclodextrin[b]	0.08	0.11
Maltosyl-β-cyclodextrin[b]	0.06	0.09
Dimaltosyl-β-cyclodextrin[b]	0.03	0.05
Glucose	0.23	0.31
Maltose	0.19	0.26
Maltotriose	0.15	0.24
Maltotetraose	0.12	0.19
Maltopentaose	0.09	0.15
Maltohexaose	0.07	0.12
Maltoheptaose	0.06	0.08

[a] TLC systems: ACN–water–concentrated aqueous ammonia (6:3:1 v/v), 1-propanol–ethyl acetate–water–concentrated aqueous ammonia (6:1:3:1 v/v).

[b] Product of Ensuiko Sugar Refining Co., Yokohama. There are hundreds of potential isomers for these groups of cyclodextrin derivatives; the authors have separated them on the basis of the number of substituents rather than within these sub-groups. (Reproduced from *Carbohydr. Res.*, **275**, 1–7, © 1995 with permission from Elsevier Science.)

The term 'HPLC' is conventionally applied to columns having plate numbers above 2000, the typical range being 2000–20 000. For analytical applications, spherical packing materials generally offer better reproducibility and column stability over irregular-shaped particles. Particle sizes ranging from 3 to 20 μm are available. The most popular particle size for analytical columns is 5 μm, but the use of 3-μm particles in short columns is increasing. The retentivity and capacity of a packing material is related to its pore size. For most general analytical methods (analyte molecular weights less than 3 kDa) pore sizes of between 6 and 13 nm are typical.

The type and commercial availability of HPLC columns and detectors has expanded tremendously in recent years. For carbohydrate analysis the two most common types of packing materials are amino-bonded silica-based materials and metal-loaded polymeric substrates such as cross-linked styrene–divinylbenzene. The silica-based packing materials have a high rigidity and do not swell in any solvent. Their weakness is limited stability in aqueous mobile phases at high pH. The polymeric-based packings are more compressible, but they are compatible with all mobile phases including the entire aqueous pH range. Both the inorganic and the polymeric based materials offer a wide range of separation modes including reversed phase, ion exchange, hydrophilic interaction, hydrophobic interaction and SEC.

Some of the commercially available columns and typical analyses associated with them are given in Table 5. It should be borne in mind that there are many commercially available columns that will perform similar functions, and the list in Table 5 is by no means exhaustive. Choice of appropriate column for given separation is often due as much to personal experience, cost and availability of materials as choice of separation chemistry.

In reversed phase chromatography, where separation is by means of hydrophobic interactions, the stationary phase is nonpolar and the mobile phase is polar.[17] Typical stationary phases are silica-based bonded phases with aliphatic hydrocarbon ligands, for example octadecylsilane (ODS or C_{18}). Other packings for reverse phase chromatography include graphitized carbon or styrene–divinylbenzene. The mobile phase for carbohydrates is commonly a mixture of ACN and water. Graphitized carbon columns show exceptional physical and chemical stability and can be used throughout the

Table 5 Some commercially available HPLC columns and their modes of separation

Matrix[a]	Separation mode	Typical analysis	Mobile phase	Commercially available columns
Polymeric resins				
S-PS (Na^+ form)	Size exclusion and ligand exchange	Di-, oligo- and polysaccharides	Water	Shodex sugar KS (Showa Denko)
S-DVB (Ag^+ form)	Ion-moderated partition	Oligosaccharides	Water	Aminex HPX-42A (Bio-Rad)
	Ion exchange	Oligosaccharides	Water	Rezex RSO (Phenomenex)
S-DVB (Ca^{2+} form)	Ion-moderated partition	High-fructose corn syrup	Water	Aminex HPX-87C (Bio-Rad) Sugar-Pak 1 (Waters)
VA (NH_2 bonded)	Normal phase	Oligosaccharides	ACN–water	Asahipak NH2P-50 (Showa Denko)
Pellicular resin (Quaternary ammonium)	Ion exchange	Disaccharides Oligosaccharides	NaOH	CarboPak PA 10 (Dionex) CarboPak PA 100 (Dionex)
Methacrylate	Size exclusion	Polysaccharides	Water	TSK G4000 PW (TOSOH) Ultrahydrogel (Waters)
Silica				
(Amino-propylsilane bonded)	Reverse phase	Di-, and oligosaccharides	ACN–water	Carbohydrate (Waters) Partisil PAC (Whatman) LiChrosorb NH_2 (Merck) Capcell Pak NH_2 (Shiseido)
(Glycerylpropyl) (Organosilane)	Size exclusion	Oligo- and polysaccharides	Water	LiChrospher Diol (Merck) Zorbax SE-250 (Hewlett-Packard)
(β-Cyclodextrin)	Ligand exchange	Di- and oligosaccharides	ACN–water	Cyclobond 1 (Rainin)
Graphitized carbon	Hydrophobic interaction	Di- and oligosaccharides	ACN–water	Hypercarb (Shandon)

[a] Type of bonded phase or functional group is shown in brackets. VA, vinyl alcohol; S-PS, sulfonated polystyrene; S-DVB, sulfonated styrene divinyl benzene.

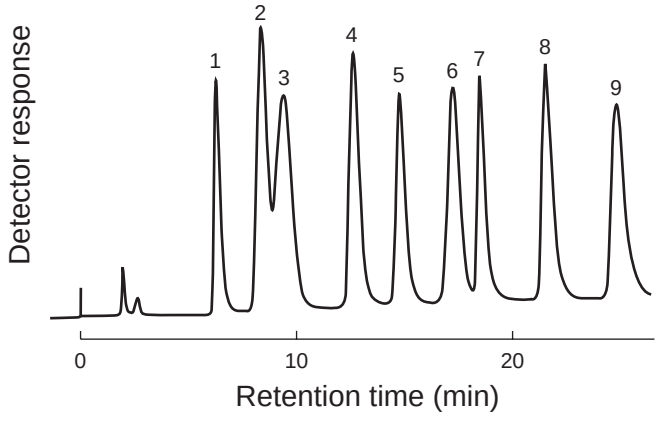

Figure 1 Separation of gluco-disaccharides on a Hypercarb column by linear gradient elution (0–15 min, 1 ml min^{-1}; pulsed amperometric detection (PAD), ambient temperature) with mM sodium hydroxide solution containing 1.5–5% ACN. 1, trehalose; 2, isomaltose; 3, kojibiose; 4, maltose; 5, nigerose; 6, gentiobiose; 7, sophorose; 8, cellobiose; and 9, laminaribiose. (Reprinted from *Carbohydr. Res.*, **215**, 67–80, © 1991 with permission from Elsevier Science.)

whole pH range. Such a column has been shown capable of separating nine isomeric disaccharides using gradient elution with dilute alkali containing small amounts of ACN (Figure 1).[18]

Separation based on hydrophilic interaction chromatography (HILIC) is another popular method of carbohydrate analysis.[19] HILIC uses polar stationary phases, such as microparticulate silica, with mobile phases ranging in polarity from water through ACN to hexane. Weak anion-exchangers based on bonded phase silicas containing primary, secondary and tertiary amines are the most common type of packing used for carbohydrate analysis. Retention increases with the hydrophilicity of the stationary phase and the hydrophobicity of the mobile phase. An alternative to using amino-bonded stationary phases for the separation of neutral oligosaccharides has been recently demonstrated using a β-cyclodextrins bonded phase column.[20] Separation of neutral oligosaccharides derived from starch, cellulose, pullulan, xylan, inulin, and mannan can be achieved using ACN–water mobile phases. Retention times are dependent on both the monosaccharide composition and glycosidic linkage type. The fast separation of inulin-derived oligosaccharides, up to DP 13, in ACN–water phases is demonstrated in Figure 2(a), with the effect of a higher ratio of ACN in the mobile phase shown in Figure 2(b). The β-cyclodextrin bonded phase column appears to have similar selectivity but greater durability compared to the amino-bonded phase silica gels. Ion-exchange chromatography,

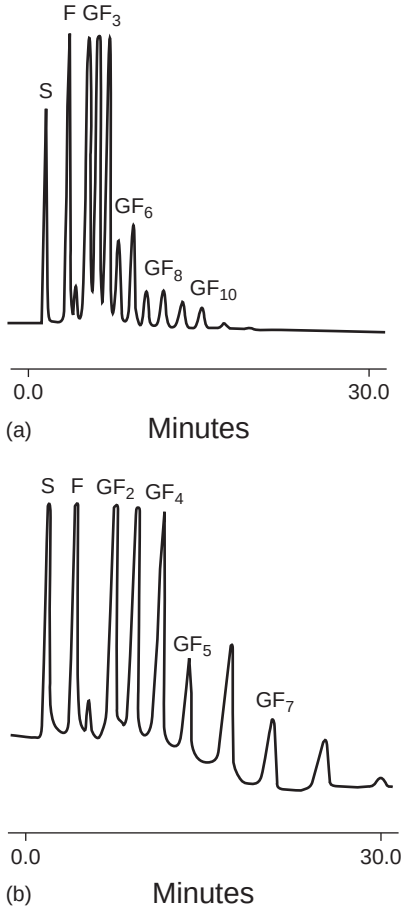

Figure 2 Separation of inulin-derived oligosaccharide standards using (a) ACN–water (70 : 30, v/v) and (b) ACN–water (75 : 25, v/v). F = fructose, GF_2 = 1-kestose (DP 3), GF_3 = DP 4, etc. using a β-cyclodextrin bonded phase column (Cyclobond I, 250 mm × 4.6 mm i.d.); operated at room temperature and 1 ml min^{-1} flow rate with refractive index (RI) detection. S represents the solvent peak. (Reprinted from *J. Chromatogr., A*, **648**, 131–137, © 1993 with permission from Elsevier Science.)

based on strong anion-exchange resins (high-performance anion-exchange chromatography methods, HPAEC), is a powerful technique, which is discussed separately in section 5.1.3.

Monosaccharides, produced by enzymatic or chemical hydrolysis of polysaccharides, are often derivatized prior to HPLC analysis by the introduction of stable chemical groups such as 1-phenyl-3-methyl-5-pyrazolone (PMP). This can improve both resolution and sensitivity. A list of some common derivatives used is available in the article on **Monosaccharides and Sugar Alcohol Analysis**.

Ultraviolet (UV) detectors are not commonly used for carbohydrate analysis, though it is possible to monitor sugars directly at wavelengths below 200 nm. However at such short wavelengths, low sensitivity and background interference levels make the use of UV detectors often impractical. RI detectors work on the principle of differential refractometry and have, until recently, been the mainstay of HPLC carbohydrate detection. However, RI detectors are temperature dependent and cannot be used with a gradient, as they are extremely sensitive to changes in mobile phase concentration. RI can be useful for routine analysis of, for example, food samples where sensitivity in excess of 0.05% w/v is not generally required. If greater sensitivity is necessary then alternative detectors must be found. Pulsed amperometric detectors now offer much more versatility and sensitivity than RI detectors. These detectors are commonly used in tandem with anion-exchange chromatography, and are discussed in more detail in the next section.

5.1.3 High-performance Anion-exchange Chromatography Methods

In recent years HPAEC has become the method of choice for the separation of disaccharide and oligosaccharide mixtures. Coupled with PAD, HPAEC/PAD offers high resolution at picomole sensitivity. Specifically developed for the analysis of carbohydrates, HPAEC/PAD takes advantage of the weakly acidic nature of carbohydrates under alkaline conditions for highly selective separations on strong anion exchange resins. Anion-exchange chromatography of the partially ionized carbohydrates is carried out on polymeric, nonporous pellicular resins that are ideally suited to conditions of high pH. Separation is due to the degree of ionization of individual saccharides, but can also be based on anomers, positional isomers and molecular size. Retention times can be manipulated by altering the strength of the sodium hydroxide concentration or by addition of the more strongly eluting acetate anion. Typical analysis conditions will use mobile phase concentrations of 1–150 mM NaOH (pH 11–13) with gradient elution in sodium acetate for more strongly bound analytes.

Carbohydrates can be electrochemically oxidized at the surface of a gold or platinum electrode by the application of a positive potential (E_1). The initial current produced is proportional to the analyte concentration. As the oxidative carbohydrate products will lead to poisoning of the electrode surface over time, it is necessary to clean the electrode surface between measurements using a pulsed sequence of potentials. The electrode potential is raised to a sufficient degree to oxidize the surface and remove carbohydrate oxidation products (E_2). To reduce the surface back to gold, a third potential is then applied (E_3). A typical cyclic sequence for carbohydrate analysis by pulsed amperometry is shown in Figure 3.[21] Although Dionex were the first company to commercially manufacture pulsed amperometric detectors, a number of other companies now produce their own versions.

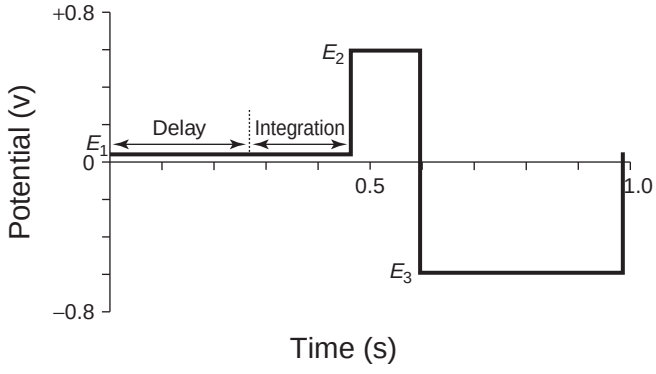

Figure 3 Typical cyclic sequence for pulsed amperometry. The potentials (E_1, E_2, E_3) and duration of the phases are shown.

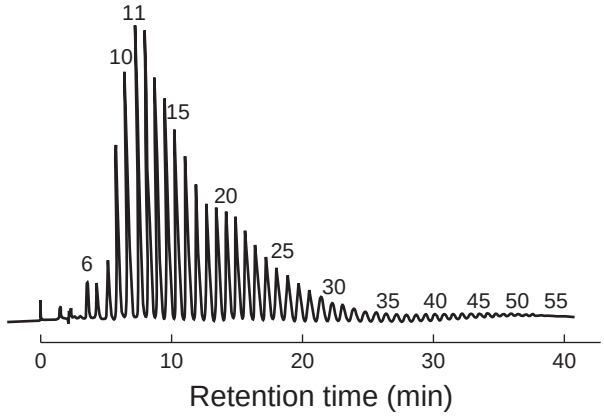

Figure 4 HPAEC elution profile of the components of debranched wheat amylopectin on a HPIC-AS6 (250 × 4 mm i.d.) column using a gradient of 200 mM sodium acetate at the start, 250 mM at 2 min, 300 mM at 10 min and 400 mM at 40 min in 150 mM sodium hydroxide solution (1 ml min^{-1}; ambient. temperature) and PAD. The number on each peak indicates its DP. (Reprinted from *J. Chromatogr.*, **585**, 233–238, © 1991 with permission from Elsevier Science.)

The advantage of detection with PAD for carbohydrates is that no pre- or post-column derivatization is necessary. Sample preparation is also simplified because only oxidizable analytes will be detected by PAD, and the sensitivity for carbohydrates is orders of magnitude greater than for possible contaminant species.[21] Although HPAEC is often carried out at high pH, some separations are better carried out at neutral or acidic pH. While PAD sensitivity will decrease below pH 11, this problem can be circumvented by the post-column addition of base, usually 300–500 mM NaOH, to maintain pH through the detector. Unlike RI detectors, PAD is not sensitive to changes in salt concentration. It does show slight sensitivity to variable low NaOH concentrations such that post-column addition of base may also be necessary when running low concentration NaOH gradients.

There are a number of good reviews on HPAEC/PAD techniques in the literature.[22,23] The technique has been used to estimate the chain length of different amylopectins after enzymatic digestion by isoamylase (Figure 4).[24] Although the relative detector response changes with the DP, this can be overcome by using saccharides with appropriate chain lengths as standards. Recently a method coupling SEC with HPAEC has been developed to distinguish between highly branched and linear oligoglucans, as well as linear isomeric gluco-oligosaccharides.[25] This method may be very useful in overcoming problems separating smaller oligosaccharides from the larger polymeric mass during the early stages of enzymatic digestion of polysaccharides.

5.1.4 Capillary Electrophoresis

Capillary electrophoresis (CE) is a relatively new technique that provides rapid, highly efficient separations for carbohydrates.[26,27] Electrophoresis is the movement of charged substances in a conductive medium, usually aqueous, under the influence of an applied electric field. Charged species are separated on the basis of their electrophoretic mobility (charge : mass ratio) and the electroosmotic flow (migration of positive buffer ions to the cathode under the influence of the applied field). Separation can be manipulated by changing either buffer composition or pH. The technique is also applicable to neutral polysaccharides, which can be charged either through complexation with for example borate ions, or by raising the pH of their microenvironment. CE is complimentary to separation of oligosaccharides by GC and HPLC techniques. Like HPLC, it is suited to the analysis of both polar and nonpolar carbohydrates and it may replace many of the HPLC techniques in the near future. CE offers fast and efficient separation, relatively affordable and durable capillary columns, requires small sample volumes and has low reagent consumption. However, it is not generally useful for preparative scale applications.

CE is available in a number of modes including capillary zone electrophoresis (CZE), isoelectric CE, and micellar electrokinetic CE. Typical CE equipment consists of high-voltage power supply, buffer reservoirs, a capillary column, and a detector. CE is usually performed using fused silica capillaries that are 50 or 70 μm i.d., 375 μm o.d. and between 30 and 100 cm long. Fused silica capillaries are commercially available from a number of suppliers and much longer ones have been used in GC for a number of years. The major limitation in the use of CE for carbohydrate analysis, as with other techniques, has been detector sensitivity. Advances in separation and detection techniques in the last five years including PAD and fluorescence detection has seen yoctomole (10^{-24} mole) sensitivity become a realistic target.[28]

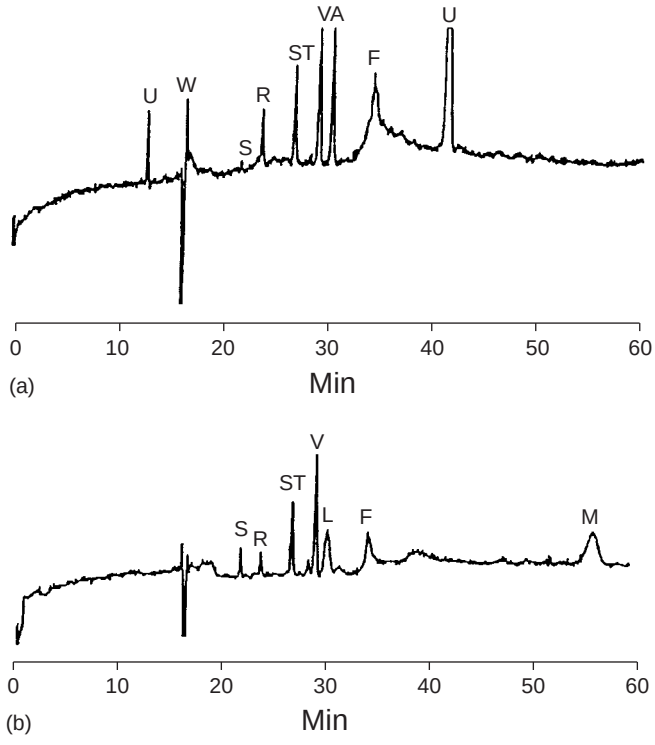

Figure 5 CZE of a mixture of oligosaccharides (a) and partially purified pea extract (b), using the following separation conditions: 100 mM $Na_2B_4O_7$, pH 9.9, 50 °C, 10 kV. S, sucrose; R, raffinose; ST, stachyose; V, verbascose; A, ajugose; L, lactose; M, melibiose; W, water front; U, unknown; F, 'false peak'. (Reprinted from *J. Chromatogr., A*, **652**, 517–524, © 1993 with permission from Elsevier Science.)

Separation of the raffinose family of oligosaccharides (Figure 5) has been demonstrated using the capillary zone method of electrophoresis, where the carbohydrates are charged using their cis-diol borate complexes.[29] Partial purification using ion exchangers before CZE appeared essential to achieve good separation (Figure 5b). Separation by high-performance CE of the borate-complexed oligosaccharides offered several advantages over HPLC separation using RI methods, including on-column UV detection, ease of operation, low cost per analysis and the use of nontoxic chemicals.

5.2 Methylation Analysis

Methylation analysis is one of the most commonly used techniques in structural carbohydrate chemistry.[30] Methylation analysis will give information on the positions at which sugar residues are substituted and their relative abundance in the polysaccharide, though it gives no information on the sequence that these residues appear in, nor on their anomeric conformation. Methylation analysis involves methylation of all the free hydroxyl groups in the polysaccharide, hydrolysis of the methylated polysaccharide to its component parts, reduction of these methylated monosaccharides to alditols, acetylation of the alditols, and subsequent qualitative and quantitative analysis of the partially methylated alditol acetate derivatives by GC or GC/MS. The positions of the free hydroxyl groups in the partially methylated sugars indicate where the residue was substituted.

The main problems associated with the technique are incomplete methylation of the polysaccharide, oxidative degradation of some hydroxyl groups and contamination from 'impure' reagents and solvents. Methylation analysis is commonly based on a modification of the Hakomori procedure, where alkoxide groups are readily generated by the presence of a strong nucleophile ensuring complete methylation of all free hydroxyl groups.[31] The polysaccharide is dissolved in dry dimethyl sulfoxide (DMSO) in the presence of a strong base such as sodium hydride, 1-butyl lithium, potassium *tert*-butoxide or powdered sodium hydroxide generating production of the dimsyl anion. Methylation of the alkoxide groups is usually carried out with methyl iodide (Scheme 1). Latterly methods based on the use of potassium dimsyl have become more widespread, and it is reported to give at least equal or better results than the sodium salt.[32] At the end of the methylation reaction, water is added and the polysaccharide can be recovered from the aqueous phase after extraction with chloroform. The methylated polysaccharide is then reduced to its component sugars by hydrolysis (trifluoroacetic, hydrochloric or sulfuric acid may be used). At this stage the components have free hydroxyl groups in the positions where they were linked to each other and methyl ether groups where the original polysaccharide had free hydroxyl groups. Reduction of the methylated monosaccharides in sodium borodeuteride, to label the reducing ends, is followed by peracetylation with acetic anhydride in pyridine (1 : 1 v/v). Finally the volatile, partially methylated alditol acetates are analyzed by GC coupled to a mass spectrometer (GC/MS, see Scheme 1).[33]

For oligosaccharide substrates it is often advantageous to convert the reducing termini of the oligosaccharides to their alditols before methylation, in order to reduce side reactions.

5.3 Periodate Oxidation

Periodate oxidation is another important tool in structural polysaccharide chemistry and can be used to determine glycosidic linkage. In the first step periodate oxidizes adjacent hydroxyl groups in carbohydrate residues. Formic acid is produced when the polysaccharide unit contains three hydroxyl groups on adjacent carbon atoms. Oxidation is followed by reduction to polyalcohols with sodium borohydride followed by mild acid hydrolysis.

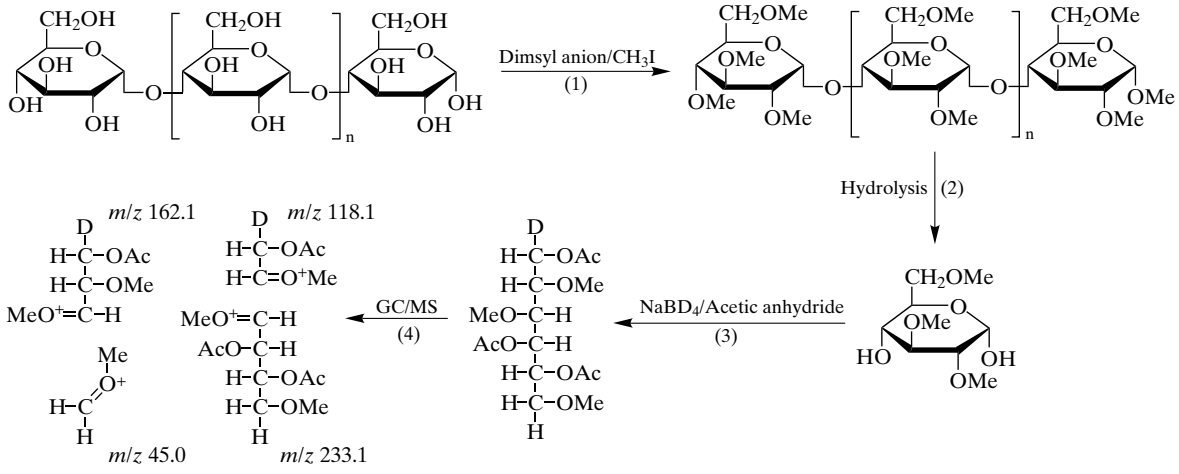

Scheme 1 Methylation analysis of glycan: methylation of all free hydroxyl groups using dimsyl anion and methyl iodide (1) is followed by hydrolysis of the methylated polysaccharide to its component sugars (2), reduction with sodium borodeuteride and acetylation with acetic anhydride (3). The major cleavage fragments of the partially methylated alditol acetate as analyzed by GC/MS are shown (4). Other fragments arise from these by elimination of neutral molecules such as acetic acid (−60.0 Da) and methanol (−32.0 Da).

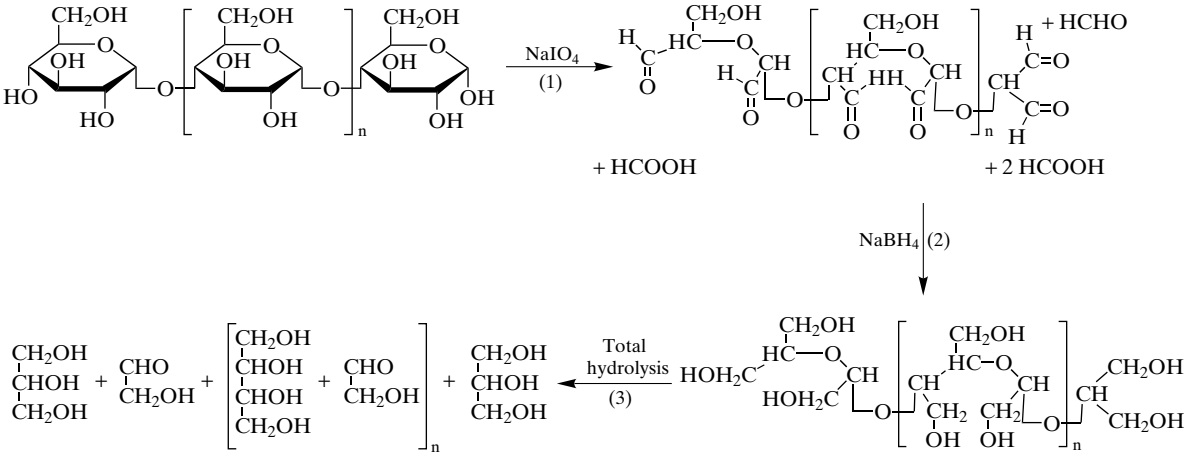

Scheme 2 Periodate oxidation of 1,4 α-linked glucan: oxidation of the glucan with sodium periodate (1) is followed by reduction with sodium borohydride (2). After acid hydrolysis (3) the products can be derivatized and analyzed by GC.

Degradation of periodate-oxidized amylose, after reduction and hydrolysis, is demonstrated in Scheme 2. Since periodate oxidation requires adjacent free hydroxyls, if the glucan in Scheme 2 was (1 → 3)-linked then only the terminal residues at both the reducing and nonreducing ends would be oxidized. Periodate oxidation is normally carried out in the pH range 3–5 to avoid acid hydrolysis and to minimize side-reactions involving nonselective oxidations. Periodate can cause over-oxidation of certain products causing misinterpretation of results. Where the main structural features of a polysaccharide are already known, periodate oxidation can be a useful method.

An important modification of this procedure, the Smith degradation, selectively degrades polysaccharides to oligosaccharides or smaller polysaccharides. Controlled hydrolysis with dilute acid is carried out at ambient temperatures. The acetals of oxidized sugar rings are the most labile during the hydrolysis step. The products may range from the low-molecular-weight products shown in Scheme 2 to monosaccharide, oligosaccharide and polysaccharide glycosides that can be usefully put through the procedure again.

5.4 Hydrolysis with Specific Glycosidases

Glycosidases are an important tool in structural determination because of their ability to specifically cleave carbohydrate linkages in disaccharides, oligosaccharides and polysaccharides. It is their ability to cleave specific

Table 6 Specificity of some important glycosidases used in polysaccharide analysis

Common name	Enzyme commission (EC) number	Specificity
α-Amylase	3.2.1.1	Internal hydrolysis of α-(1 → 4)-glucans
β-Amylase	3.2.1.2	Terminal hydrolysis of maltose from α-(1 → 4)-glucans
Glucoamylase	3.2.1.3	Terminal hydrolysis of α-linked glucose units
Cellulase	3.2.1.4	Internal hydrolysis of β-(1 → 4)-glucans
Inulinase	3.2.1.7	Internal hydrolysis of β-(2 → 1)-fructans
Endo–β-xylanase	3.2.1.8	Internal hydrolysis of β-(1 → 4)-xylans
Dextranase	3.2.1.11	Internal hydrolysis of α-(1 → 6)-glucans
Chitinase	3.2.1.14	Internal hydrolysis of *N*-acetyl-β-(1 → 4)-glucosamine links in chitin
Polygalacturonase	3.2.1.15	Internal hydrolysis of galacturonic acid in pectins
α-Glucosidase	3.2.1.20	Terminal hydrolysis of α-linked glucose
β-Glucosidase	3.2.1.21	Terminal hydrolysis of β-linked glucose
α-Galactosidase	3.2.1.22	Terminal hydrolysis of α-linked galactose
β-Galactosidase	3.2.1.23	Terminal hydrolysis of β-linked galactose
α-Mannosidase	3.2.1.24	Terminal hydrolysis of α-linked mannose
β-Mannosidase	3.2.1.25	Terminal hydrolysis of β-linked mannose
Invertase	3.2.1.26	Terminal hydrolysis of β-linked fructofuranose
N-Acetyl-β-glucosaminidase	3.2.1.30	Terminal hydrolysis of β-linked Glc*N*Ac in glycans
β-Glucuronidase	3.2.1.31	Terminal hydrolysis of β-linked glucuronic acid in polysaccharides
β-Xylosidase	3.2.1.37	Terminal hydrolysis of β-linked xylose
Pullulanase	3.2.1.41	Hydrolysis of reducing terminal α-linked L-arabinose
α-L-Arabinofuranosidase	3.2.1.55	Hydrolysis of α-(1 → 6)-linked glucose in predominantly α-(1-4)-glucans
Levanase	3.2.1.65	Internal hydrolysis of β-(2 → 6)-fructans
Isoamylase	3.2.1.68	Internal hydrolysis of α-(1 → 6)-linked glucose in predominantly α-(1-4)-glucans
β-Mannanase	3.2.1.78	Internal hydrolysis of (1 → 4)-linked mannose in mannans, galactomannans and glucomannans

residues without side-reactions, yielding unmodified oligosaccharides that is important for the determination of native polysaccharide structure. Mixtures of di- and oligosaccharides can then be separated and identified by column chromatography and NMR. Knowledge of the mode of action of the enzymes used allows a prediction for native polysaccharide fine structure. This technique is hindered only by the availability of suitable, single-activity enzymes. Complete hydrolysis of a complex polysaccharide will involve the use of more than one type of enzyme. A list of some of the glycosidases commonly used in polysaccharide analysis is given in Table 6.

6 MOLECULAR WEIGHT ANALYSIS

Oligosaccharides and polysaccharides usually possess one, and only one, reducing end group per molecule. Where these can be quantified, the number of molecules present can be determined. If the total carbohydrate present is also determined, the average chain length and the molecular weight may be calculated. There are very sensitive colorimetric assays available for reducing end groups including the use of oligosaccharide dehydrogenase, an enzyme specific for reducing end groups.[34] Alternatively, quantitative ^{1}H NMR using the terminal anomeric reporter groups may determine them. These methods allow molecular weights up to about 20 kDa (NMR) to over 100 kDa (colorimetric) to be determined. However, the presence of even tiny amounts of monosaccharide impurities may render the result invalid due to the number averaging of the resultant molecular weights.

The molecular weight (M) of pure macromolecules may be determined by ultracentrifugation, so long as both the sedimentation coefficient (s) and the diffusion coefficient (D) are determined under the same conditions, using the relationship shown in Equation (1).

$$\frac{s}{D} = \frac{M(1 - \bar{v}\rho)}{RT} \tag{1}$$

where T is the temperature (K), R is the gas constant, $\bar{v}$ is the partial specific volume (the effective volume per unit mass) of the solute and ρ is the solution density. Although this method is still sometimes used, the method remains arduous compared with SEC and laser LS methodology.

6.1 Size Exclusion Chromatography

SEC has proved to be a useful technique for separating carbohydrates on the basis of their molecular weight.[35] Molecules are eluted in order of decreasing molecular size from a stationary phase containing a wide range of different size pores, allowing greater access and hence slower elution for smaller molecules. SEC is particularly successful in separating members of oligosaccharide families where it often allows the exact number of component sugars to be determined (Figure 6). Where the oligosaccharides come from a mixture of oligosaccharide families, the assignment of molecular size is less reliable in the absence of known standards. SEC is less successful at determining the molecular size of polysaccharides where it depends on the use of similar carbohydrates, of known molecular weight, for calibration. Unless these molecules are monodisperse and have almost identical structures to those of the unknowns, the resultant calibration may be grossly in error. This has been shown to be the case where the molecular weights can be independently determined by, for example, LS (see section 6.2). Use of secondary calibration in the SEC, where polysaccharides of known molecular weight but dissimilar structure are used, has been shown to overestimate the molar mass of chitosans[36] and capsular polysaccharides by two- and threefold (Figure 7).[37] Improvement can be made if 'universal calibration' is used involving the use of the product of the intrinsic viscosity and the molecular weight ($\eta\overline{M}_w$, a measure of the dynamic volume of the polysaccharide) in place of the molecular weight alone.[38] However, this involves the additional use of a viscometric detector and gives incorrect results if the calibrating and unknown polysaccharides are of different conformational types (e.g. random coil versus stiff rod). Although high-performance size exclusion chromatography (HPSEC) uses columns having theoretical plate numbers of about 10 000, even under optimal conditions they are only capable of resolving polysaccharide peaks where there is at least a twofold difference in molecular weights or sizes present. There are two reasons for this: (1) the resolving power of SEC columns is relatively poor for high-molecular-weight polysaccharides, and (2) such polysaccharides usually consist of a range of molecular species with a wide molecular weight distribution. In spite of these limitations, SEC is proving to be a powerful technique for the analysis of polysaccharide polydispersity when used in combination with on-line LS molecular weight determination.

6.2 Light Scattering

Light is scattered from molecules in solution in a way that depends on the concentration and molecular size of the molecules present. This has been known for over 50 years

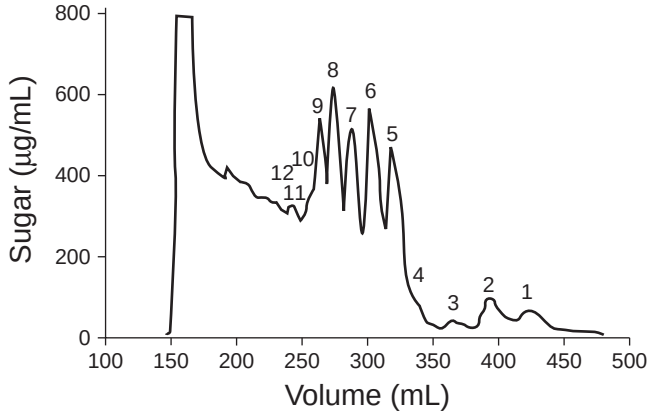

Figure 6 Separation of oligomeric arabinoxylans by SEC on a Bio-Gel P-2 at 60 °C. The number of the fraction corresponds to its DP. Fractions 1–3 are xylose, xylobiose and xylotriose respectively, Fractions 4–10 are mainly Xyl_3Ara, Xyl_4Ara, Xyl_5Ara/Xyl_4Ara_2, Xyl_5Ara_2, Xyl_6Ara_3/Xyl_7Ara_2, Xyl_6Ara_4/Xyl_7Ara_3, respectively. (Reprinted from *Carbohydr. Res.*, **249**, 369–382, © 1993 with permission from Elsevier Science.)

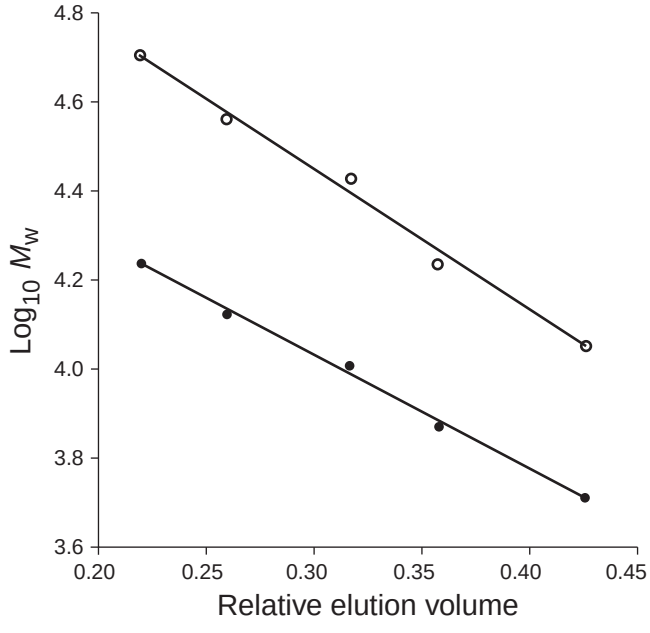

Figure 7 Comparison of molar-mass determination of capsular polysaccharide fractions by SEC using dextrans for calibration and by multiangle laser light scattering (MALLS). (Reproduced from D'Ambra et al.[37] by permission of Academic Press, Inc.)

but it is only recently that methods have become widely used for polysaccharides. In particular the use of MALLS has enabled the accurate nondestructive determination of the molecular weight, molecular weight distribution and radius of gyration of many polysaccharides without the need for known calibrating standards. When combined with the separating power of HPSEC, it forms an

important tool for the polysaccharide analyst. A typical SEC/MALLS system includes an RI detector, placed after the MALLS detector, for determining the concentration of the eluting polysaccharides. Generally there is no need for the simultaneous UV detection, which is so useful in the SEC/MALLS of proteins, as polysaccharides rarely possess absorbing groups.

LS from molecules in solution is described by Equation (2).[39]

$$\frac{Kc}{R_\theta} = \frac{1}{M_w}\left(1 + \frac{16\pi^2}{3\lambda^2}\langle r_g^2\rangle \sin^2\left(\frac{\theta}{2}\right. \right. + f_4 \sin^4\left(\frac{\theta}{2}\right. \quad + \cdots \quad + 2A_2c \tag{2}$$

where:

K is an optical parameter given in Equation (3).

$$K = \frac{4\pi^2 n^2 (\mathrm{d}n/\mathrm{d}c)^2}{\lambda^2 N_A} \tag{3}$$

c is the sample concentration, calculated from the differential RI response, R_θ is the excess intensity of scattered light at angle θ, and M_w is the weight average molar mass (molecular weight, Equation 4).

$$M_w = \frac{\sum_i n_i M_i^2}{\sum_i n_i M_i} \tag{4}$$

where M_i is the mass of the ith polysaccharide chain and n_i is the number of chains with that mass.

λ is the wavelength of the scattered light in vacuum. A_2 is the second virial coefficient, a measure of the solute solvent interaction. The $2A_2c$ term is usually negligible in SEC separations due to typically low concentrations. A_2 is characteristically about $1–2 \times 10^{-5}$ mol ml g^{-2}. At polysaccharide (M_w 10^6) concentrations of about 1 mg ml^{-1}, ignoring this term gives about a 2–4% error. It may, however, be significant for charged polysaccharides at low ionic strength.

N_A is Avogadro's number, n is the RI. ($\mathrm{d}n/\mathrm{d}c$) is the RI increment of the solute, which is usually determined prior to HPSEC using a range of polysaccharide concentrations. This value varies between polysaccharides and is somewhat less (~0.13–0.15 ml g^{-1}) than that for proteins (~0.18–0.19 ml g^{-1}).

$\langle r_g^2\rangle$ is the z-average mean square of the radius of gyration (R_G, Equation 5)

$$R_G^2 = \langle r_g^2\rangle = \frac{\sum_i n_i M_i^2 \langle r_g^2\rangle_i}{\sum_i n_i M_i^2} \tag{5}$$

R_G reflects the average displacement of each residue from the molecular centre. Typically for a linear polysaccharide (M_w 10^6) this would be approximately 6 nm if spherically packed (R_G = radius $\times \sqrt{3/5}$), 940 nm if as an extended stiff rod (R_G = length/$\sqrt{12}$), and 17 nm (R_G = unit length $\times \sqrt{\text{units}/6}$) if as a random coil. The $\sin^4(\theta/2)$ and higher terms (not shown in Equation 2) are not explicitly used as the determination of the molar mass and the radius of gyration involve extrapolation to zero angle. For smaller molecules with R_G up to 10 nm, the $\sin^2(\theta/2)$ term is ignored as well as this simplification only introduces errors of less than about 1%. The lack of sensitivity to scattering angle at low R_G means that it cannot be accurately determined when smaller than about a sixtieth of the wavelength of the light used (i.e. <10 nm). Ignoring this term in $\sin^2(\theta)$ allows single-angle LS (Equation 6) data (often at 90°) to be used.

$$\mathrm{LS} = K_{LS} c M_w \left(\frac{\mathrm{d}n}{\mathrm{d}c}\right.^2 \tag{6}$$

where K_{LS} is a detector calibration constant. As the RI signal (Equation 7) is given by

$$\mathrm{RI} = K_{RI} c \left(\frac{\mathrm{d}n}{\mathrm{d}c}\right. \tag{7}$$

where K_{RI} is the instrumental calibration constant, the molar mass (M_w) can be determined from these two readings once the polysaccharide's RI increment has been determined (Equation 8).

$$\frac{1}{M_w} = K' \frac{\mathrm{d}n}{\mathrm{d}c}\frac{\mathrm{RI}}{\mathrm{LS}} \tag{8}$$

K' is a calibration constant determined by using previously characterized similar polysaccharides.

At low angles, the angular dependence of LS depends only on the radius of gyration and is independent of molecular conformation or branching. This is of particular benefit in the study of polysaccharides, which may take up a wide variety of conformations in solution. Often the molecular weight and radius of gyration are determined by plotting (Kc/R_θ) against $\sin^2(\theta/2)$ (Zimm plot) to yield an often almost linear curve whose intercept gives M_w and whose slope at low angles gives $\langle r_g^2\rangle$. For branched polysaccharides the radius of gyration can provide a measure of the degree of branching if similar but unbranched polysaccharides are available. When repeated at various polysaccharide concentrations, or across a SEC run, other parameters such as the number average molecular weight ($\overline{M}_n$, Equation 9), polydispersity (Equation 10), molecular weight distribution and weight-averaged radius of gyration ($\overline{R}_G$, Equation 11)

may also be determined.

$$\overline{M}_n = \frac{\sum_i n_i M_i}{\sum_i n_i} = \frac{\sum_i c_i}{\sum_i \frac{c_i}{M_i}} \quad (9)$$

$$\text{polydispersity} = \frac{\overline{M}_w}{\overline{M}_n} \quad (10)$$

$$\overline{R}_G^2 = \overline{\langle r_g^2 \rangle} = \frac{\sum_i n_i M_i \langle r_g^2 \rangle_i}{\sum_i n_i M_i} = \frac{\sum_i c_i \langle r_g^2 \rangle_i}{\sum_i c_i} \quad (11)$$

M_w is always greater or equal to M_n so the polydispersity, which measures the breadth of the molecular weight distribution present, is always greater or equal to unity.

MALLS detectors simultaneously measure the LS at a number of angles and extrapolate via software to zero angle. There are also low-angle laser light scattering (LALLS) devices which only measure at very low angles (~3–7°). The choice between these systems depends on a number of factors including sensitivity to extraneous noise. It has been found that, for most polysaccharides, the molecular weights vary across HPSEC peaks (Figure 8). Molecular weights of polysaccharides that are insoluble in water can be determined using MALLS with other solvents (e.g. DMSO) without necessarily using HPSEC.

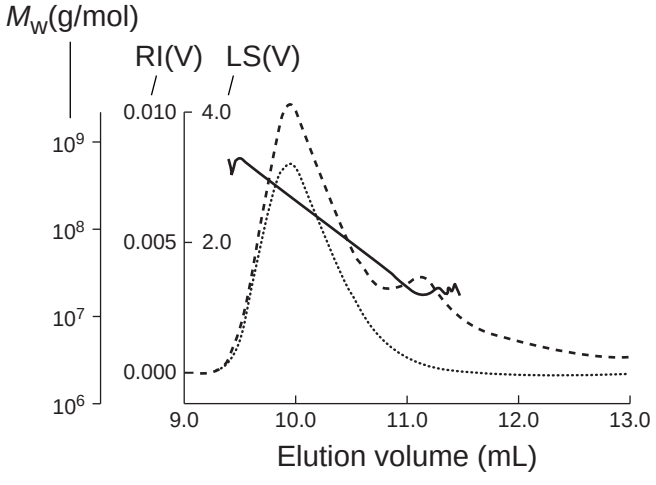

Figure 8 HPSEC/MALLS of maize starch. The HPSEC was performed using two columns (TSK gel G4000 SWXL 7.8 mm i.d. × 30 cm and TSK gel G3000 SWXL 7.8 mm i.d. × 30 cm, TosoHaas) in series at 30 °C. (· · · · ·) LS, volts; (- - - - -) RI, volts; (————) molecular weight (M_w) from MALLS. Note that the second peak is almost lost using LS as this depends not only on concentration but also on molecular size. (Reproduced from Bello-Pérez et al.[55] with permission of Academic Press, Inc.)

Even highly insoluble polysaccharides such as cellulose may be analyzed (e.g. in dimethylacetamide containing LiCl), if they can be dissolved in suitable solvents, although such solvents may not be compatible with the HPSEC matrices. MALLS detection may be used in conjunction with other HPLC separation techniques (e.g. reverse phase chromatography) but as it determines the weight-averaged molecular weight it will give less meaningful results if a wide range of molecular masses, or few species but with very disparate molecular masses, are present. MALLS can also be used for determining the molar mass and molecular weight distribution of oligosaccharides.[40]

7 PHYSICAL METHODS OF ANALYSIS

7.1 Mass Spectrometry

MS was one of the first methods used to investigate carbohydrate structure. Its principle of operation is to create gas-phase ions, separate them in space or time based on their mass-to-charge (m/z) ratio and to quantify them. Originally it assisted in linkage analysis by identifying the methylated alditol acetates produced in the methylation analysis of glycans and separated by GC. The electron impact/mass spectrometry ionization method (EI/MS), often utilized in GC, is highly energetic causing extensive degradation and, usually, the absence of molecular ions. This lack of molecular ions is of little importance as it is the fragmentation pattern that is so important to methylation analysis, clearly demonstrating which hydroxyl groups were free and have been methylated, and which were involved in linkages and have been acetylated. As EI/MS requires volatile samples, oligosaccharides are too polar unless derivatized. Development of the gentler fast atom bombardment ionization source allows oligosaccharides, both derivatized and underivatized, to be analyzed in a similar manner but where molecular ions and large fragments, produced by cleavage between the glycosidic units, are produced. The reducing end of oligosaccharides can be labeled with ^{18}O for MS by catalyzed exchange with $H_2{}^{18}O$ (e.g. by using 2-aminopyridine or other epimerization catalyst), enabling fragments involving this end to be distinguished. Larger oligosaccharides and polysaccharides may make use of the even milder ionization techniques involving matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). As these produce molecular ions without interference from breakdown products, the molecular weight distribution can be clearly demonstrated. MS must be combined with other methods of analysis, as it cannot distinguish between isomeric sugar residues and cannot easily provide the linkage and stereochemical information of NMR.

7.1.1 Fast Atom Bombardment/Mass Spectrometry

FAB/MS (fast atom bombardment/mass spectrometry) can be used optimally on nanomolar amounts of oligosaccharides and small polysaccharides up to about 15 kDa[41] although usually requiring microgram amounts of oligosaccharides of less than about 5 kDa. Xe or Cs ion or Xe atom guns are all suitable for carbohydrates. Carbohydrates are generally, but not necessarily, derivatized by permethylation or peracetylation as then they are easier to purify, the sensitivity is improved by as much as two orders of magnitude and the fragmentation pattern is more predictable. However, the use of underivatized carbohydrates removes the risk of any loss of labile substituents. The spectra usually contain the molecular ions, which may be protonated or cationized with ammonium, sodium or potassium. This allows cyclic glycans, such as the cyclodextrins, to be identified, as their molecular mass is 18 (H_2O) less than the corresponding linear glycan. The major inter-glycosidic fragmentation processes may result in hydrolytic-like fragments from either side of the glycosidic link or elimination-like oxonium ion fragments from the anomeric side only (Scheme 3). Two fragmentation pathways dominate the spectrum. Permethylated and peracetylated glycans cleave on nonreducing side of glycosidic bonds to give oxonium ions (Scheme 3a), a cleavage that is particularly favored by *N*-acetyl-hexosamines. Native carbohydrates commonly give β-cleavage where a hydrogen atom is taken from the neighbor glycoside and retained on either the reducing (Scheme 3b) or nonreducing end (Scheme 3c). The residue masses are given in Table 7. Linkage analysis can be achieved from FAB/MS by the analysis of rarer cross-ring fragmentation products or by utilizing partial

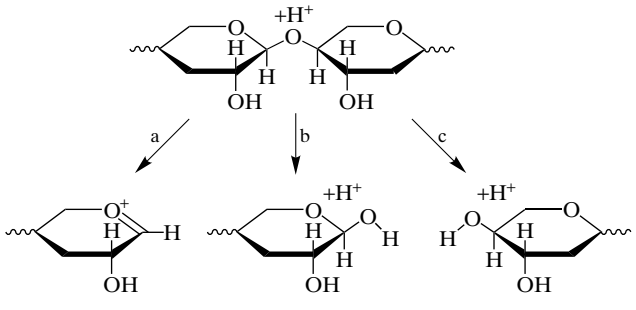

Scheme 3 The major inter-glycosidic fragmentation processes. (a) 'A type' cleavage; (b) β-cleavage with charge retention on the reducing end; (c) β-cleavage with charge retention on the nonreducing end. In β-cleavage the hydrogen atom has been abstracted from the position β to the glycosidic oxygen on the uncharged fragment and producing an enol/keto group (e.g. the 3-position in (b) and the 2-position in (c)) in unmethylated oligosaccharides.

Table 7 Exact mass data for carbohydrate residues in oligosaccharides and polysaccharides utilizing the most abundant isotopes; also given is the maximum number of free hydroxyl groups if the residues are not terminal (there will be one more if terminal) and the extra mass involved for each additional methylated and acetylated group

Residue	Exact mass	Hydroxyls available
Pentose	132.0432	2
Hexose	162.0528	3
Deoxyhexose	146.0579	2
Hexuronic acid	176.0321	3
N-Acetyl hexosamine	203.0794	2
N-Acetyl neuraminic acid	291.0954	4
O-Methyl replacing hydroxyl group	14.0157	
O-Acetyl group replacing hydroxyl group	42.0106	
Nonreducing end	1.0078	
Reducing end	17.0027	
^2H replacing ^1H	1.0063	
^{18}O replacing ^{16}O	2.0042	

acid or enzyme hydrolysates. Anomeric configurations can be determined by FAB/MS due to the different rates of their chromium trioxide oxidation in peracetylated oligosaccharides, β-linkages are equatorial and are more rapidly oxidized than α-linkages so forming a keto-ester, breaking the ring, and resulting in a mass change of +14 (+O −2H).

7.1.2 Matrix-assisted Laser Desorption Ionization Mass Spectrometry

MALDI can vaporize and ionize large molecules and is capable of determining molecular weights, in exceptional cases, of up to *m*/*z* 300 kDa using time-of-flight (TOF) detection. It requires only femtomoles to picomoles of material, needs no sample derivation and tolerates more buffer and detergent than other MS methods. Although able to handle proteins, it has not had similar success, so far, with natural polysaccharides in the upper part of this range (>25 kDa) and seems to work best with oligosaccharides below 5 kDa (Figure 9). This may be partly due to the polydispersity of such samples and the inherently poor resolution of MALDI/TOF/MS.

MALDI ionizes by using a UV laser pulse to ablate a UV-absorbing matrix which carries some of the large molecules into the gas phase in an ionized form. The matrix commonly used for carbohydrates is 2,5-dihydroxybenzoic acid (DHB), although several other matrices (e.g. 3-aminoquinoline, α-cyano-4-hydroxycinnamic acid, 2-hydroxy-5-methoxybenzoic acid, 2,6-dihydroxyacetophenone) are often tried in order

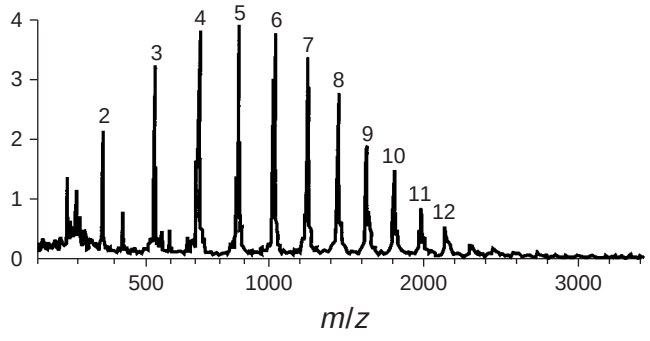

Figure 9 MALDI/MS of mixed isomeric onion (*Allium cepa* L.) fructan oligosaccharides showing the DP of the predominant monosodium adduct ions $[M + Na]^+$. The difference between sequential peaks is 162.1 Da. (Reproduced from Stahl et al.[56] with permission of Academic Press, Inc.)

to optimize the spectra. Anions, such as sulfate, are added by some workers in order to obtain negatively charged ions, which are analyzed by MS using the negative mode. The matrix is usually present in about a hundred-fold molar excess over the carbohydrate. Intact oligosaccharides can be recovered and sequenced directly, by MALDI/TOF, from sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE) separated glycoproteins.[42] Gel pieces containing glycoprotein are denatured, reduced *S*-alkylated and endoglycosidase digested, then extracted with water, cleaned up using micro-columns, dried and subjected to MALDI/TOF. Neutral oligosaccharides give $[M + Na]^+$ and smaller $[M + K]^+$ peaks. Acidic residues, such as sialic and uronic acids, form both positive and negative ions and readily form salts with alkali metal ions resulting in multiple peaks. These adverse effects may be overcome by prior esterification.

The mass accuracy of MALDI (0.1–0.01%) is generally lower than other MS methods, but the recent linking of the method to collision-induced dissociation/Fourier transform mass spectrometry (CID/FTMS) extends this to 0.001% for oligosaccharides.[43]

7.1.3 Electrospray Ionization/Mass Spectrometry

ESI allows production of molecular ions directly from samples in solution and is thus suitable for direct interfacing to HPLC and CE. It is far less able to handle buffers and detergents than MALDI. Multiply charged ions are usually produced which is an advantage on naturally multiply charged proteins and glycoproteins (up to about 150 kDa) but a disadvantage on polydisperse neutral glycans. Although ESI has good sensitivity (picomole) and accuracy (0.005%), the *m*/*z* range of their normal quadrupole detectors is only up to 3 kDa allowing its use only on glycopeptides and smaller oligosaccharides and lipopolysaccharides.

7.1.4 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) utilizes two or more mass selection devices in series so that the known ion from one analyzer may be selected and induced to give daughter fragments which are determined using a second analyzer. It can provide complete sequence, branching and linkage information.[44] If higher energy collision-induced dissociation (CID) is achieved, the choice between pyranose and furanose ring systems may be distinguished due to carbon–carbon bond cleavage of the side chain leaving ring oxonium ions.

MS/MS can be used with FAB (fast atom bombardment), ESI[45] and MALDI. The use of post-source decay (PSD) in MALDI/TOF is proving particularly useful for the linkage analysis of complex oligosaccharides where a database of spectral information can be built up, such as with glycoprotein *N*-glycans.[46] Similar data can be obtained by ESI CID. This is because the extent and distribution pattern of fragments is highly reproducible. Multiple fragmentation occurs inter-glycosidically with retention of Na^+ cationized charge on either the reducing or nonreducing fragment in a similar manner to that in FAB/MS (see section 7.1.1).

7.2 Nuclear Magnetic Resonance Spectroscopy

NMR is proving to be the most powerful technique for the structural analysis of oligosaccharides and polysaccharides. It is particularly useful in finding unusual and unusually substituted sugars. Each proton (^1H nucleus) attached directly to carbon in a monosaccharide has its own characteristic environment. However as the number of carbohydrate units increases in the molecule so does the number of magnetically distinct nuclei. Pentoses and hexoses in oligosaccharides and polysaccharides have eight or ten distinct protons per residue, respectively. Two or three, respectively, of these are associated with alcohol hydroxyl groups. This means that they have approximately the same number of protons as proteins of similar size; significantly more if the hydroxyl protons are counted in. They also have approximately the same number of carbon atoms. Nitrogen atoms are absent in most oligosaccharides and polysaccharides but, even where present in large amounts (e.g. in chitosans), there are significantly fewer than in similarly sized proteins. ^1H and ^{13}C NMR spectra of carbohydrates may, therefore, be extremely complex. Additionally, many of these environments are often similar, giving a poor spectral dispersion with overlapping resonances. As in protein NMR, the low natural abundance of ^{13}C (1.11%) and ^{15}N (0.37%) and

these isotopes' lower natural sensitivities compared to ^{1}H (^{13}C, 1.6%; ^{15}N, 0.1%), means that NMR utilizing these atoms is much improved if the carbohydrate has been isotopically enriched.

The key features of carbohydrate NMR are due to the downfield shifts (to higher ppm) due to the spatial proximity of oxygen atoms. As the anomeric carbon is bonded to two oxygen atoms, the anomeric (C-1) carbon (^{13}C) atoms and protons are easily distinguishable and provide important inroads into the structural determination. Other groups also provide well-resolved resonances. The methyl protons and carbon (^{13}C) atoms in deoxysugars are upfield (i.e. at lower ppm) and the carboxylate carbon (^{13}C) atoms in uronic acids are downfield. Free amino groups give pH dependent shifts, upfield of hydroxyls, whilst *N*-acetyl and *O*-acetyl groups give upfield methyl groups.

In NMR the quality of the spectrum depends on two relaxation times. The longitudinal relaxation time T_1 (seconds) describes the rate at which the magnetization returns to equilibrium along the magnetic field (z direction) and correlates with the overall rotational tumbling of the carbohydrate and its intramolecular flexibility. The transverse relaxation time T_2 (seconds) describes the rate of decay of the effective magnetization observed in the receiver (xy directions) and correlates with the dynamic properties of the molecule (e.g. it reduces with increasing molecular size). Line width is inversely proportional to T_2. The large size of polysaccharides leads to slow tumbling, hence broad lines due to fast T_2 spin–spin relaxation. This reduces the sensitivity, broadens the spectral lines and generally results in poor performance of 2-D techniques that incorporate delays to allow coherence (same phase) transfer. Although spectral improvements can be achieved by heating (~70–85 °C), albeit with the introduction of extra 'noise', this may have conformational consequences. For example, multihelical structures give very poor spectra, which can be improved by heating and/or alkali but the treatment results in them having a different conformation. Cations (e.g. Ca^{2+}) may also cause gelling with carboxylic-acid-containing polysaccharides, so reducing performance. A high viscosity is also associated with poor NMR spectra. In general, the larger the polysaccharide, the less resolved its NMR spectrum. Resolution may be improved by partial hydrolysis (e.g. 10 mM HCl, 100 °C, 10 min) but this may give fragments that are not representative of the original structure. For example, partial acid hydrolysis of arabinoxylans strips off the labile arabinofuranose groups leaving bare β-xylan chains. If available, partial hydrolysis using well-characterized endo-glycosidases is preferred.

Polysaccharides are usually run in D_2O ($^{2}H_2O$). Working in D_2O eliminates the less well-defined exchangeable hydroxyl protons and the vast excess (110 M) of H_2O protons, although significant HOD (4.4 mM) resonance remains even in 99.996% D_2O solution. Intramolecular hydrogen bonding may be investigated by dissolving the carbohydrate in per-deuterated DMSO when the hydroxyl (and amine) protons do not exchange with solvent and give useful information. It must be noted, however, that sugars may behave differently in this solvent than in aqueous solution. Inter-residue hydrogen bonding is important in determining the overall conformations of carbohydrates but is difficult to determine by NMR. In some cases hydrogen bonding may be determined from proton exchange in mixed $^{1}H_2O$ and $^{2}H_2O$ solvents where solvent exposure may be quantified from the differential shifts in proton and carbon spectra.

The anomeric protons can provide information concerning the types of residues present, their anomeric (α or β) configuration and their ring size (pyranose or furanose). Details of the NMR spectra of many carbohydrates are available from a database.[47] Although 1-D proton NMR spectra can be used as fingerprints to enable structural determination where databased structures exist, they are generally too complex to fully assign even for relatively simple oligosaccharides (Figure 10a). ^{13}C Spectra are simpler but are also usually too complex to use without the availability of fully assigned NMR spectra from suitable known structures for guidance. The structures of unknown carbohydrates generally demand a variety of through-bond correlated spectroscopy (COSY) and through-space nuclear Overhauser effect spectroscopy (NOESY) 2-D NMR techniques using high field spectrometers (600 MHz) for their elucidation.

Although NMR is an essentially nondestructive method relatively large amounts of material are required and may be lost or degraded if required for long periods. NMR is between three and six orders of magnitude less sensitive than MS. 1-D NMR requires between about 10 nmol (600 MHz) to 80 nmol (300 Mz) and 2-D NMR requires somewhat more. Oligosaccharides have small proton resonance linewidths when compared to proteins which allow 2-D NMR spectra to be obtained at low (~100 μM) concentrations.

^{13}C NMR offers simplified spectra with greater spectral dispersion, which are often best for preliminary information. Complete assignment of ^{1}H NMR and ^{13}C NMR resonances usually starts with the well-resolved anomeric H-1 or C-1 peak using 1-D NMR. The proton resonances can then be assigned sequentially using through-bond techniques;[48] the H-2 by using 2-D time proportional phase incrementation/double quantum filtered/correlation spectroscopy (TPPI/DQF/COSY), then the H-3 using 2-D one-step relayed coherence transfer spectroscopy (RECSY) after eliminating the

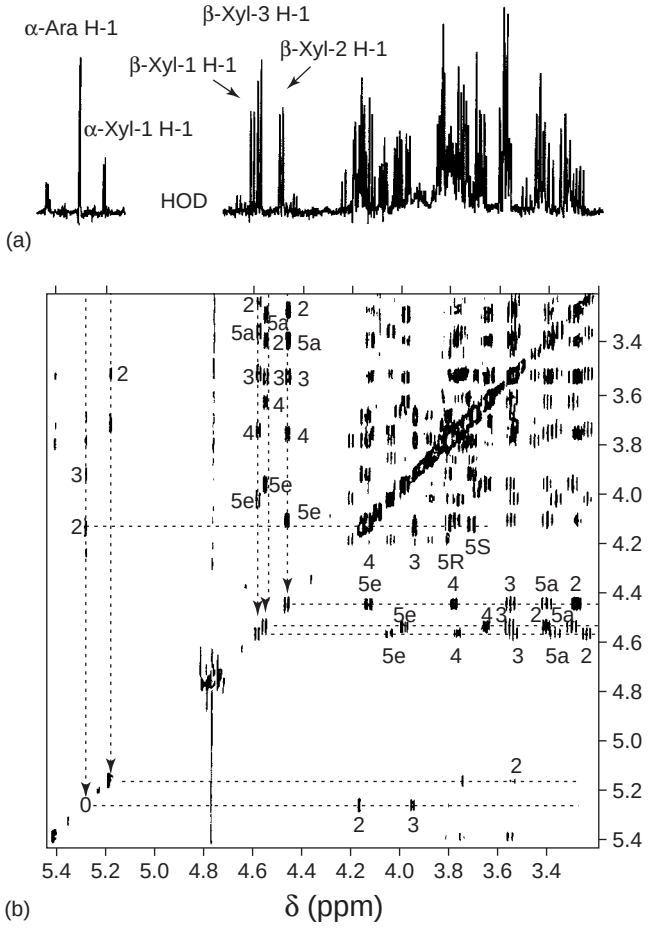

Figure 10 ^{1}H NMR spectrum (600 MHz) of the arabinoxylan fragment α-L-Ara*f*-(1 → 2)-β-D-Xyl*p*-(1 → 4)-β-D-Xyl*p*-(1 → 4)-D-Xyl*p*. (a) 1-D NMR showing the distinctive anomeric peaks, mainly as doublets. (b) 2-D HOHAHA NMR showing how off-diagonal assignments of each ring system can be made once the anomeric proton is assigned. Peaks on the diagonal are equivalent to the 1-D NMR spectrum. (Reproduced from *Carbohydr. Res.*, **254**, 245–255, © 1994 by permission from Elsevier Science.)

H-2 signal, followed finally by the H-4 using 2-D time proportional phase incrementation/total correlation spectroscopy (TPPI/TOCSY) after eliminating both the H-2 and H-3 signals. All the protons associated with each carbohydrate residue can be identified using 2-D homonuclear Hartmann–Hahn spectroscopy (HOHAHA), or its close relation, total correlation spectroscopy (TOCSY), in D_2O. The cross-peaks for all ring protons H-2–H-6 (for hexoses) associated with the anomeric proton signal can be distinguished (Figure 10b), as the method utilizes a number of repetitive pulses that permit magnetization to flow freely from one proton to another at a rate determined by their ^3J coupling. TQF/COSY is particularly used to find groups of protons that couple, such as H-4/H-5e/H-5a in pentopyranoses and H-5/H-6/H-6′ in hexopyranoses, but are the most difficult to assign by other correlation techniques. Linkage positions can be found by comparison between related oligosaccharides and noting the field shifts.[49]

The anomeric conformation can often be determined directly from the chemical shift by comparison with known structures. It may be confirmed from the size of the anomeric $^1J_{C-H}$ coupling constant, which is typically 170 Hz for axial hydrogen and 160 Hz if equatorial. The $^3J_{H-H}$ coupling constants obey a Karplus relationship (Equation 12).

$$^3J = A\cos^2(\theta) - B\cos(\theta) + C \qquad (12)$$

A, *B* and *C* are fitted parameters derived from a number of known trainer molecules. The cosine nature of Karplus curves means that it gives one to four possible angles every 360° for a given ^3J (Figure 11). Simple molecular modeling can usually make choice between these as often only one conformation is structurally reasonable. Karplus curves are not exact as other factors, such as bond lengths, bond angles and the nature and orientation of groups attached to the atoms, all have an effect. They can, however, generally give the torsion angles to within about 30° and allow the ring conformation (e.g. pyranose chair conformation $^4C_1$ or $^1C_4$) to be easily verified. Several Karplus parameters are available from different sources, which have been optimized on their

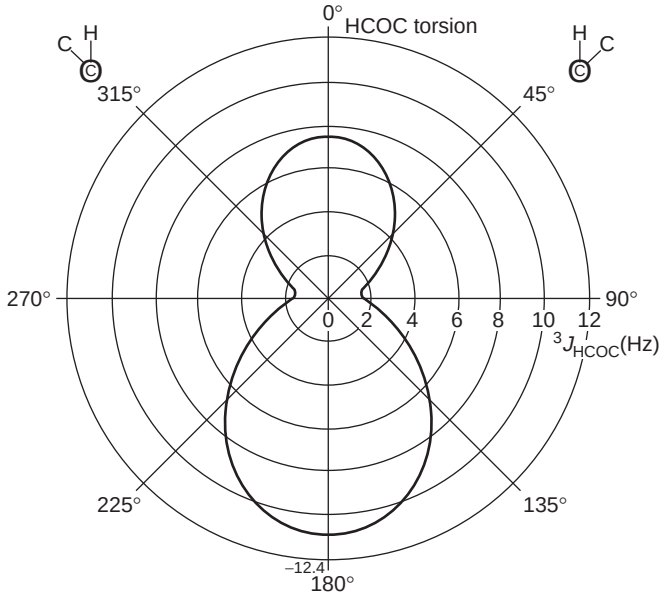

Figure 11 The Karplus relationship $^3J = A\cos^2(\theta) - B\cos(\theta) + C$. For the coupling constants involving the glycosidic linking torsions ($^3J_{COCH}$), *A*, *B* and *C* are 7.6, 1.7 and 1.6 Hz respectively.[58] It can be noted that for any experimental value of ^3J there may be up to four possible torsional angles (e.g. a $^3J_{COCH}$ of 6 Hz may be derived from a torsion of 28°, 131°, 229° or 332°).

own subset of structures or with the aid of quantum mechanical methods. As many compounds often have similar torsional angles for similar structures and a complete range of torsions (0–360°) are rarely available in the model compounds, these parameters may give very different results when away from their training angles. However, even if the parameters are very different, the equation generally gives good correspondence for the commonly found angles. The $^3J_{H-H}$ coupling constants are often able to confirm the molecular species and anomeric configurations present. For example, the anomeric configuration of glucopyranose and galactopyranose are easily observed, from the $^3J_{H1-H2}$ coupling constant, as the H-1 and H-2 protons of the beta anomers are trans-diaxial (~180°) giving large couplings whereas those in alpha anomers are gauche (~60°) with small couplings. Mannopyranose anomers are undecided by this method as both are gauche. D-xylose and D-glucose are characterized by large coupling constants for H-2, H-3, H-4 and H-5 (5a in xylose) arising from their mutually trans-diaxial orientations. Karplus curves giving the relationship of ^3J coupling constants to the H–C–C–H torsion are given in Haasnoot et al.[50]

The power of NMR is most useful when it is used to determine the sequence, linkage positions and conformation of carbohydrates. This can only be attempted when there has been a full assignment of the ^{1}H and ^{13}C spectra and it is preferred that the carbohydrate has been enriched with the ^{13}C isotope. Linkages can be determined from the $^3J_{CH}$ scalar coupling through and across glycosidic bonds using heteronuclear multiple bond correlation (HMBC) and from interglycosidic ^{1}H NOE (nuclear Overhauser effect) effects using NOESY or ROESY (rotating frame nuclear Overhauser effect correlated spectroscopy). This also allows the sequence to be determined from these through-space connectivities. However, with increasing size above about 20 glycosyl residues, it becomes increasingly difficult to delineate branch locations unambiguously. ^{1}H NOE effects give those protons that are in close proximity (less than about 0.5 nm, the signal being inversely proportional to the sixth power of the inter-proton distance). They can be ambiguous particularly at branch points in certain conformations where protons, neighbouring the linkage site, may lie closer than the linkage protons although further apart in the primary structure. NOE effects are also difficult to interpret in the presence of significant internal molecular motion, due to positional averaging. Although NMR is difficulty to apply to many polysaccharides, the presence of repeating units can be identified from the NMR of oligomers and quantified in polymers.

Linkage positions may also be discovered by determination of the position of the free hydroxyl groups, giving information similar to that obtained from methylation analysis (section 5.2), but nondestructively. ^{1}H NMR spectra in H_2O (~10% D_2O) and D_2O are compared to identify the position(s) of the hydroxyl groups.[51] Hydroxyl proton signals are observed, downfield of anomeric protons, in H_2O at low temperature (e.g. −16 °C in H_2O–perdeuterated acetone 4 : 1 v/v) when the pH is adjusted at 6.4 ± 0.2 to reduce the hydroxyl proton exchange with H_2O that is catalyzed above pH 7. Where the hydroxyl protons and their associated C–H aliphatic protons are coupled there is line broadening. If the linewidth comparison fails to give the hydroxyl groups due to overlapping peaks then the COSY CH–OH cross-peaks may be used (in H_2O).

The biological roles of oligosaccharides and polysaccharides are related to their conformations. This mainly depends on the trans-glycosidic torsional angles connecting the carbohydrate units (Scheme 4). Conformational analysis requires independent data for each independent torsion angle (ϕ, ψ and sometimes w) and, where a mixture of conformations exists, each relative population occupancy. These may be available from inter-residue proton NOE (e.g. between the anomeric proton and the proton on the linkage carbon, H–C_1–O–C_n–H) and the $^3J_{COCH}$, $^3J_{COCC}$, $^3J_{CCCH}$ and $^2J_{COC}$ coupling constants (Figure 11) but the cosine nature of the Karplus equation for ^3J coupling means modeling may be required. Whilst the conformation of rigid oligosaccharides can be directly derived from NOESY data, the upper and lower bounds of the possible distances from NOESY often enclose a large number of ϕ, ψ combinations. It is usual to combine NMR, molecular dynamics and random walk molecular mechanics (RAMM) in order to have confidence in the resulting conformation. The additional information available from molecular modeling is particularly useful for polysaccharides where these coupling constants may not be measurable. Modeling is also necessary in oligosaccharides where there are mixed conformational populations present. It should be noted that multiple conformations are often in fast equilibrium on the NMR timescale and NMR data are effectively time-averaged, which may give rise to unrealistic putative torsional angles.

A continuous probability distribution (CUPID) approach can analyze the rotamer populations from NMR spin–spin couplings and NOE enhancements.[52] This gives results reported as typically ±30–40° for the dihedrals and ±0.2 for the probabilities. This method requires an amount of experimental data that is normally only available for rotations around carbon–carbon bonds such as the C5–C6 bond in hexopyranoses.

Solid state ^{13}C NMR has been used on simple highly ordered polysaccharides, such as cellulose, but few detailed conclusions can be drawn.

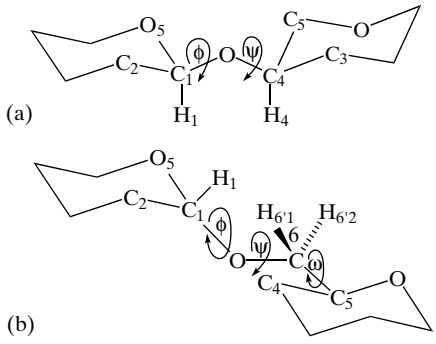

Scheme 4 The torsions and possible coupling constants involved in the anomeric linkages in oligosaccharides and polysaccharides. (a) Torsions ϕ and ψ are defined as the clockwise torsions $O_rC_aO_aC_n$, and $C_aO_aC_nC_{n-1}$ respectively, where r, a and n represent ring, anomeric and structural number respectively. Alternative torsions are sometimes used as they are easier to determine experimentally; ϕ_H is $H_aC_aO_aC_n$ (about $\phi + 120°$; relative angles are not exact as the bond angles are not exactly tetrahedral) and ψ_H is $C_aO_aC_nH_n$ (about $\psi - 120°$), The relevant coupling constants are $^3J_{H_1C_1OC_4}$ giving ϕ_H, $^3J_{C_1OC_4H_4}$ giving ψ_H, $^3J_{C_1OC_4C_3}$ giving ψ and $^3J_{C_2C_1OC_4}$. (b) An exocyclic link; the additional linking torsion w is defined as the clockwise torsion $O_aC_nC_{n-1}C_{n-2}$. The relevant coupling constants are $^3J_{H_1C_1OC_6}$ giving ϕ_H, $^3J_{C_1OC_6C_5}$ giving ψ, $^3J_{C_1OC_6H_{6'1}}$, $^3J_{C_1OC_6H_{6'2}}$, $^3J_{C_2C_1OC_6}$, $^3J_{H_{6'1}C_6C_5C_4}$ and $^3J_{H_{6'2}C_6C_5C_4}$.

7.3 Infrared Spectroscopy

Fourier transform infrared (FTIR) is a rapid non-destructive and quantitative method for determining oligosaccharides and polysaccharides. Pure samples of the glycans and any interfering substances must be available for comparison. FTIR determines the whole spectrum at once, easily allowing a number of scans to be averaged to achieve high sensitivity and resolution. The complexity of the spectra means that it is not suitable for totally unknown mixtures where there may be interference from unknown materials.[53] Analysis is achieved from the absorbance at a number of differing wavelengths ($\sim$900–1200 cm^{-1}, e.g. the C–O stretching vibration) after comparison with the absorbance matrix obtained using the calibration materials.

LIST OF SYMBOLS

A_2	Second Virial Coefficient
c	Sample Concentration
D	Diffusion Coefficient
J	Coupling Constant
K	Optical Parameter
M	Molecular Weight
M_n	Number-averaged Molecular Weight
M_w	Weight-averaged Molecular Weight
$\overline{M}_w$	Mean Weight-averaged Molecular Weight
m/z	Mass to Charge Ratio
n	Refractive Index
N_A	Avogadro's Number
R	Gas Constant
R_θ	Excess Intensity of Scattered Light at Angle θ
R_f	Movement of a Solute Relative to the Solvent Front
R_G	Radius of Gyration
s	Sedimentation coefficient
T	Temperature

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
CE	Capillary Electrophoresis
CID	Collision-induced Dissociation
CID/FTMS	Collision-induced Dissociation/ Fourier Transform Mass Spectrometry
COSY	Correlated Spectroscopy
CUPID	Continuous Probability Distribution
CZE	Capillary Zone Electrophoresis
DEAE	Diethylaminoethyl
DHB	2,5-Dihydroxybenzoic Acid
DMSO	Dimethyl Sulfoxide
DP	Degree of Polymerization
EI/MS	Electron Impact/Mass Spectrometry
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FAB/MS	Fast Atom Bombardment/Mass Spectrometry
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HILIC	Hydrophilic Interaction Chromatography
HMBC	Heteronuclear Multiple Bond Correlation
HOHAHA	Homonuclear Hartmann–Hahn Spectroscopy
HPAEC	High-performance Anion-exchange Chromatography Methods
HPAEC/PAD	High-performance Anion-exchange Chromatography coupled with Pulsed Amperometric Detection
HPLC	High-performance Liquid Chromatography
HPSEC	High-performance Size Exclusion Chromatography

HPTLC	High-performance Thin-layer Chromatography
LALLS	Low-angle Laser Light Scattering
LS	Light Scattering
MALDI	Matrix Assisted Laser Desorption Ionization
MALLS	Multiangle Laser Light Scattering
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
ODS	Octadecylsilane
PAD	Pulsed Amperometric Detection
PMP	1-Phenyl-3-Methyl-5-Pyrazolone
PSD	Post-source Decay
RAMM	Random Walk Molecular Mechanics
RECSY	Relayed Coherence Transfer Spectroscopy
RI	Refractive Index
ROESY	Rotating Frame Nuclear Overhauser Effect Correlated Spectroscopy
SDS/PAGE	Sodium Dodecyl Sulfate/ Polyacrylamide Gel Electrophoresis
S-DVB	Sulfonated Styrene Divinyl Benzene
SEC	Size Exclusion Chromatography
S-PS	Sulfonated Polystyrene
TLC	Thin-layer Chromatography
TOCSY	Total Correlation Spectroscopy
TOF	Time-of-flight
TPPI/DQF/COSY	Time Proportional Phase Incrementation/Double Quantum Filtered/Correlation Spectroscopy
TPPI/TOCSY	Time Proportional Phase Incrementation/Total Correlation Spectroscopy
UV	Ultraviolet
VA	Vinyl Alcohol

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • High-performance Liquid Chromatography of Biological Macromolecules • Infrared Spectroscopy of Biological Applications • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Glycoprotein Analysis: General Methods • Glycoprotein Analysis: Using Nuclear Magnetic Resonance • Monosaccharides and Sugar Alcohol Analysis • Proteoglycan and Acidic Polysaccharide Analysis

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Dietary Fiber Analysis as Non-starch Polysaccharides • Enzyme Analysis and Bioassays in Food Analysis • Liquid Chromatography in Food Analysis • Near-infrared Spectroscopy in Food Analysis • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials • Starch Analysis in Food

Particle Size Analysis **(Volume 6)**
Light Scattering, Classical: Size and Size Distribution Characterization

Polymers and Rubbers **(Volume 9)**
Size-exclusion Chromatography of Polymers

Pulp and Paper **(Volume 9)**
Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis • Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Liquid Chromatography **(Volume 13)**
Biopolymer Chromatography • Thin-layer Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. M. Dubois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, 'Colorimetric Method for Determination of Sugars and Related Substances', *Anal. Chem.*, **28**, 350–356 (1956).
2. H. Gruppen, R.J. Hamer, A.G.J. Voragen, 'Barium Hydroxide as a Tool to Extract Pure Arabinoxylans from Water-insoluble Cell Wall Material of Wheat Flour', *J. Cereal Sci.*, **13**(3), 275–290 (1991).
3. G.A. De Ruiter, H.A. Schols, A.G.J. Voragen, F.M. Rombouts, 'Carbohydrate Analysis of Water-soluble

Uronic Acid-containing Polysaccharides with High-performance Anion-exchange Chromatography using Methanolysis Combined with TFA Hydrolysis is Superior to four other Methods', *Anal. Biochem.*, **207**(1), 176–185 (1992).
4. Y.C. Lee, 'High-performance Anion-exchange Chromatography for Carbohydrate Analysis', *Anal. Biochem.*, **189**(2), 151–162 (1990).
5. Z. Dische, 'Spectrophotometric Method for the Determination of Free Pentose and Pentose in Nucleotides', *J. Biol. Chem.*, **181**, 379–392 (1949).
6. K.A. Taylor, J.G. Buchanan-Smith, 'A Colorimetric Method for the Quantitation of Uronic Acid and a Specific Assay for Galacturonic Acid', *Anal. Biochem.*, **201**(1), 190–196 (1992).
7. M. Somogyi, 'Determination of Blood Sugar', *J. Biochem.*, **195**, 19–23 (1952).
8. P. Bernfield, 'Amylases, α and β', *Methods Enzymol.*, **1**, 149–158 (1955).
9. Anon, in *Methods of Enzymatic Food Analysis*, Boehringer Mannheim, GmbH, Germany, 1989.
10. C.E. Uhegbu, K.B. Lim, H.L. Pardue, 'Initial Studies of a New Approach to the Design and Use of Enzyme-based Reactor/Sensor Systems: Amperometric System for Glucose', *Anal. Chem.*, **65**(18), 2443–2451 (1993).
11. J. Karkalas, R.F. Tester, 'Continuous Enzymic Determination of α-Glucan in Eluates from Gel-Chromatography Columns', *J. Cereal. Sci.*, **15**(2), 175–180 (1992).
12. M. Ghebregzabher, S. Rufini, B. Monaldi, M. Lato, 'Thin-layer Chromatography of Carbohydrates', *J. Chromatogr.*, **127**, 133–162 (1976).
13. S.C. Churms, in *CRC Handbook of Chromatography*, eds. G. Zweig, J. Sherma, CRC Press, Boca Raton, Florida, 1–272, 1982.
14. J. Jindrich, J. Pitha, B. Lindberg, 'Separation of Cyclodextrins and their Derivatives by Thin-layer and Preparative Column Chromatography', *Carbohydr. Chem.*, **275**(1), 1–7 (1995).
15. D.J. Folkes, A. Brookes, in *Glucose Syrups: Science and Technology*, eds. S.Z. Dziedzic, M.W. Kearsey, Elsevier Applied Science, London, 1984.
16. H. Traitler, S. del Vedovo, T.F. Schewizer, 'Gas-Chromatographic Separation of Sugars by On-column Injection on Glass-Capillary Columns', *J. High Resol. Chromatogr. Chromatogr. Comm.*, **7**(10), 558–562 (1984).
17. E. Rajakyla, 'Use of Reversed-phase Chromatography in Carbohydrate Analysis', *J. Chromatogr.*, **353**, 1–12 (1986).
18. K. Koizumi, Y. Okada, M. Fukuda, 'High-performance Liquid Chromatography of Mono- and Oligosaccharides on a Graphitized Carbon Column', *Carbohydr. Res.*, **215**(1), 67–80 (1991).
19. S.C. Churms, 'Recent Progress in Carbohydrate Separation by High-performance Liquid Chromatography Based on Hydrophilic Interaction', *J. Chromatogr. A*, **720**(1–2), 75–91 (1996).
20. P.J. Simms, R.M. Haines, K.B. Hicks, 'High-performance Liquid Chromatography of Neutral Oligosaccharides on a β-Cyclodextrin Bonded Phase Column', *J. Chromatogr.*, **648**(1), 131–137 (1993).
21. Dionex, Technical Note 21: 'Optimal Settings for Pulsed Amperometric Detection of Carbohydrates Using Dionex Pulsed Electrochemical and Amperometric Detectors', Dionex Corporation, Salt Lake City, UT, USA, 1998.
22. C. Corradini, 'Recent Advances in Carbohydrate Analysis by High-performance Anion-exchange Chromatography Coupled with Pulsed Amperometric Detection', *Ann. Chim.*, **84**(9–10), 385–396 (1994).
23. Y.C. Lee, 'Carbohydrate Analyses with High-performance Anion-exchange Chromatography', *J. Chromatogr.*, **720**(1–2), 137–149 (1996).
24. K. Koizumi, M. Fukuda, S. Hizukuri, 'Estimation of the Distributions of Chain Length of Amylopectins by High-performance Liquid Chromatography with Pulsed Amperometric Detection', *J. Chromatogr.*, **585**(2), 233–238 (1991).
25. T. Suortti, 'Coupled Size-exclusion Chromatography-anion-exchange Chromatography in the Analysis of Poly- and Oligosaccharides', *J. Chromatogr. A.*, **763**(1–2), 331–335 (1997).
26. Z. El Rassi, Y. Mechref, 'Recent Advances in Capillary Electrophoresis of Carbohydrates', *Electrophoresis*, **17**, 275–301 (1996).
27. S. Suzuki, S. Honda, 'A Tabulated Review of Capillary Electrophoresis of Carbohydrates', *Electrophoresis*, **19**(15), 2539–2560 (1998).
28. M.J. Davies, E.F. Hounsell, 'Carbohydrate Chromatography: Towards Yoctomole Sensitivity', *Biomed. Chromatogr.*, **10**(6), 285–289 (1996).
29. A.M. Arentoft, S. Michaelson, H. Sorenson, 'Determination of Oligosaccharides by Capillary Zone Electrophoresis', *J. Chromatogr. A.*, **652**, 517–524 (1993).
30. A. Jay, 'The Methylation Reaction in Carbohydrate Analysis', *J. Carbohydr. Chem.*, **15**(8), 897–923 (1996).
31. S.I. Hakomori, 'A Rapid Permethylation of Glycolipid, and Polysaccharide Catalyzed by Methylsulfinyl Carbanion in Dimethyl Sulfoxide', *J. Biochem.*, **55**(2), 205–208 (1964).
32. L.R. Philips, B.A. Fraser, 'Methylation of Carbohydrates with Dimsyl Potassium in Dimethyl Sulfoxide', *Carbohydr. Res.*, **90**(1), 149–152 (1981).
33. P.J. Harris, R.J. Henry, A.B. Blakeney, B.A. Stone, 'An Improved Procedure for the Methylation Analysis of Oligosaccharides and Polysaccharides', *Carbohydr. Res.*, **127**(1), 59–73 (1984).
34. G.S. Nilsson, M. Anderson, T. Ruzgas, L. Gorton, 'Oligosaccharide Dehydrogenase-catalyzed Assay for the Determination of Polysaccharides', *Anal. Biochem.*, **265**, 151–156 (1998).
35. S.C. Churms, 'Recent Progress in Carbohydrate Separation by High-performance Liquid Chromatography

based on Size Exclusion', *J. Chromatogr. A*, **720**, 151–166 (1996).
36. R.G. Beri, J. Walker, E.T. Reese, J.E. Rollings, 'Characterization of Chitosans via Coupled Size-exclusion Chromatography and Multiple-angle Laser Light-scattering Technique', *Carbohydr. Res.*, **238**, 11–26 (1993).
37. A.J. D'Ambra, J.E. Baugher, P.E. Concannon, R.A. Pon, F. Michon, 'Direct and Indirect Methods for Molar-Mass Analysis of Fragments of the Capsular Polysaccharide of *Haemophilus influenzae* Type b', *Anal. Biochem.*, **250**, 228–236 (1997).
38. A.M. Striegel, R.D. Plattner, J.L. Willett, 'Dilute Solution Behavior of Dendrimers and Polysaccharides: SEC, ESI/MS and Computer Modelling', *Anal. Chem.*, **71**, 978–986 (1999).
39. P.J. Wyatt, 'Light Scattering and the Absolute Characterization of Macromolecules', *Anal. Chim. Acta*, **272**, 1–40 (1993).
40. J.E. Knobloch, P.N. Shaklee, 'Absolute Molecular Weight Distribution of Low-molecular Weight Heparins by Size-exclusion Chromatography with Multiangle Laser Light Scattering Detection', *Anal. Biochem.*, **245**, 231–241 (1997).
41. A. Dell, A.J. Reason, K-H. Khoo, M. Panico, R.A. McDowell, H.R. Morris, 'Mass Spectrometry of Carbohydrate-containing Biopolymers', *Methods Enzymol.*, **230**, 108–132 (1994).
42. B. Küster, S.F. Wheeler, A.P. Hunter, R.A. Dwek, D.J. Harvey, 'Sequencing of *N*-Linked Oligosaccharides directly from Protein Gels: In-Gel Deglycosylation followed by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry and Normal-phase High-performance Liquid Chromatography', *Anal. Biochem.*, **250**, 82–101 (1997).
43. K. Tseng, L.L. Lindsay, S. Penn, J.L. Hedrick, C.B. Lebrilla, 'Characterization of Neutral Oligosaccharide-Alditols from *Xenopus laevis* Egg Jelly Coats by Matrix-assisted Laser Desorption Fourier Transform Mass-spectrometry', *Anal. Biochem.*, **250**, 18–28 (1997).
44. N. Viseux, E. de Hoffmann, B. Domon, 'Structural Assignment of Permethylated Oligosaccharide Subunits Using Sequential Tandem Mass Spectrometry', *Anal. Chem.*, **70**, 4951–4959 (1998).
45. A.S. Weiskopf, P. Vouros, D.J. Harvey, 'Electrospray Ionization-Ion Trap Mass Spectrometry for Structural Analysis of Complex *N*-Linked Glycoprotein Oligosaccharides', *Anal. Chem.*, **70**, 4441–4447 (1998).
46. J.C. Rouse, A-M. Strang, W. Yu, J.E. Vath, 'Isomeric Differentiation of Asparagine-linked Oligosaccharides by Matrix-assisted Laser Desorption-Ionization Post-source Decay Time-of-Flight Mass Spectrometry', *Anal. Biochem.*, **256**, 22–46 (1998).
47. CARBANK, *http://128.192.9.29/carbbank/CarbBank.htm*
48. S.B. Levery, M.S. Toledo, E. Suzuki, M.E.K. Salyan, S. Hakomori, A.H. Straus, H.K. Takahashi, 'Structural Characterization of a new Galactofuranose-containing Glycolipid Antigen of *Paracoccidioides brasiliensis*', *Biochem. Biophys. Res. Comm.*, **222**, 639–645 (1996).
49. N. Shibata, M. Onozawa, N. Tadano, Y. Hinosawa, A. Suzuki, K. Ikuta, H. Kobayashi, S. Suzuki, Y. Okawa, 'Structure and Antigenicity of the Mannans of *Candida famata* and *Candida saitoana*: Comparative Study with the Mannan of *Candida guilliermondii*', *Arch. Biochem. Biophys.*, **336**(1), 49–58 (1996).
50. A.G. Haasnoot, F.A.A.M. DeLeeuw, C. Altona, 'The Relationship between Proton–Proton Coupling Constants and Substituent Electronegativities-I', *Tetrahedron*, **36**(19), 2783–2792 (1981).
51. S. Sheng, R. Cherniak, H. van Halbeek, 'A 1H NMR Spectroscopic Approach to the Unambiguous Determination of Glycosyl Linkage Positions in Oligosaccharides', *Anal. Biochem.*, **256**, 63–66 (1998).
52. Z. Dzakula, W.M. Westler, J.L. Markley, 'Continuous Probability Distribution (CUPID) Analysis of Potentials for Internal Rotations', *J. Magn. Reson. B*, **111**, 109–126 (1996).
53. E.M. Kemsley, L. Zhuo, M.K. Hammouri, R.H. Wilson, 'Quantitative Analysis of Sugar Solutions using Infrared Spectroscopy', *Food Chem.*, **44**, 299–304 (1992).
54. F.J.M. Kormelink, R.A. Hoffmann, H. Gruppen, A.G.J. Voragen, J.P. Kamerling, J.F.G. Vliegenhart, 'Characterisation by ^{1}H NMR Spectroscopy of Oligosaccharides Derived from Alkali-extractable Wheat-flour Arabinoxylan by Digestion with Endo-(1 → 4)-β-D-Xylanase III from *Aspergillus awamori*', *Carbohyd. Res.*, **249**, 369–382 (1993).
55. L.A. Bello-Pérez, P. Roger, B. Baud, P. Colonna, 'Macromolecular Features of Starches Determined by Aqueous High-performance Size Exclusion Chromatography', *J. Cereal Sci.*, **27**, 267–278 (1998).
56. B. Stahl, A. Linos, M. Karas, F. Hillenkamp, M. Steup, 'Analysis of Fructans from Higher Plants by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Biochem.*, **246**, 195–204 (1997).
57. R.J. Viëtor, R.A. Hoffmann, S.A.G.F. Angelino, A.G.J. Voragen, J.P. Kamerling, J.F.G. Vliegenhart, 'Structures of Small Oligomers Liberated from Barley Arabinoxylans by Endoxylanase from *Aspergillus awamori*', *Carbohyd. Res.*, **254**, 245–255 (1994).
58. J.E. Anderson, A.I. Ijeh, 'Eclipsed Ground-state Conformations for Methoxycyclohexanes with Adjacent Methyl-Group Substitution – An NMR Criterion and Molecular Mechanics Calculations', *J. Chem. Soc.*, Perkin, **2**(9), 1965–1967 (1994).

Glycolipid Analysis

Ian M. Morrison
The Scottish Crop Research Institute, Invergowrie, UK

1 Introduction **1**
2 Extraction and Purification **2**
2.1 Extraction 3
2.2 General Purification 3
2.3 Separation into Neutral Glycolipid and Acidic Glycolipid (Ganglioside) Fractions 4
2.4 Separation of Glycolipid Mixtures 4
3 Chemical Methods of Analysis **6**
3.1 Composition of Glycolipids 6
3.2 Linkage Analysis of Carbohydrate Chains 8
3.3 Anomeric Configuration of Carbohydrate Residues 9
3.4 Linkage Analysis of Fatty Acids 9
4 Biochemical Methods of Analysis **9**
4.1 Linkage Analysis and Anomeric Configuration via Specific Glycosidases 9
4.2 Linkage Analysis via Immunological Methods 10
5 Physical Methods of Analysis **10**
5.1 Mass Spectroscopy 10
5.2 Nuclear Magnetic Resonance Spectroscopy 14
Acknowledgments **15**
Abbreviations and Acronyms **15**
Related Articles **16**
References **16**

Glycolipids are conjugated macromolecules that have a carbohydrate (oligosaccharide) component covalently bound to a lipid (diacylglycerol, ceramide, sterol or polyprenol) component. They are present both in procaryotes and eucaryotes but the carbohydrate components, as well as the lipid components, vary considerably between different animal, plant and microbial cells. The biological specificity of a glycolipid is due mainly to the carbohydrate structure. The amphipathic nature of glycolipids means that a single extraction scheme is impractical for all the different molecular species while special techniques are required to differentiate between the neutral glycolipids and acidic glycolipids (gangliosides). Thin-layer chromatography (TLC) and high-performance liquid chromatography (HPLC) are the major purification methods. Chemical methods are used to determine the carbohydrate and lipid profiles as well as the linkage analysis of both types of components and are used in conjunction with gas chromatography (GC), HPLC and TLC, sometimes coupled with mass spectroscopy (MS). Enzymatic methods can be used to determine the linkage sequence and anomeric configurations of the oligosaccharide structures and the same parameters are being determined by immunological methods with increasing frequency. Enzymatic methods are also used to establish the linkage analysis of the fatty acid in the diacylglycerol species. However, various MS and both ^{1}H- and ^{13}C-NMR (nuclear magnetic resonance) methods, but mostly the former, are now invariably used to confirm the full structural features of an unknown glycolipid. The type of MS method used, matrix-assisted laser desorption/ionization (MALDI), fast atom bombardment (FAB) or tandem mass spectrometry (MS/MS) will depend on the type of glycolipid.

1 INTRODUCTION

Glycoconjugate is the general term applied to macromolecules, usually biological, that have a carbohydrate component covalently bound to a noncarbohydrate component. The most widespread examples are glycoproteins and glycolipids. The same complex carbohydrate component can be associated with both examples. Glycolipids are found both in procaryotes and eucaryotes but the complex carbohydrate components, as well as the compounds that are considered to be the lipid components, vary considerably between the different animal, plant and microbial cells. Some simple glycolipids have been synthesized on an industrial scale and have commercial applications. The glyco components are complex oligosaccharide chains that may involve frequent branching. The carbohydrate components can also carry noncarbohydrate substituents such as acetate or sulfate groups linked by covalent chemical bonds. The lipid component is far more variable since many types of compound are considered as lipids. Diacylglycerol, ceramides (amides derived from a fatty acid and a sphingosine base), sterols and polyprenols are the most abundant. The biological specificity of a glycolipid is due more to the overall carbohydrate structure and influenced less by the lipid component.

In glycoglycerolipids, glycerol is always in the *sn* configuration and fatty acids are attached at C-1 (*sn*-1) and C-2 (*sn*-2) by ester bonds: it is unusual for the same

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

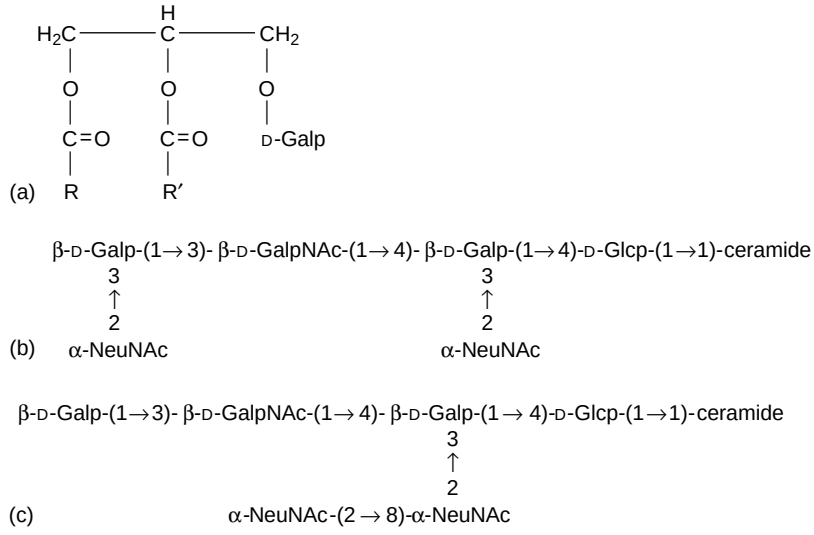

Figure 1 General structure of (a) a D-galactosyldiacylglyceride and (b) and (c) the two related disialogangliosides, G_{D1a} and G_{D1b}, respectively.

fatty acid to occupy both positions. The carbohydrate components, mostly frequently D-galactose or oligomers of this sugar residue, are always joined at C-3 (*sn*-3) by a glycosidic bond. A D-galactosyldiacylglyceride is shown in Figure 1(a). For the cerebrosides, the sphingosine base is normally unsaturated and its primary hydroxy group is involved in a glycosidic bond with the reducing end of the carbohydrate chain. The disialogangliosides G_{D1a} and G_{D1b} are shown in Figures 1(b) and 1(c). The cerebrosides can be subdivided into neutral cerebrosides and acidic cerebrosides (gangliosides). The neutral cerebrosides contain mainly neutral hexose and hexosamine residues while the gangliosides are characterized by the presence of the nine-carbon sugar, neuraminic acid, as well as the same sugars as are found in the neutral cerebrosides. Sulfatides are a special form of cerebrosides containing sulfate ester groups. As more sensitive detection methods are used, the range and complexity of carbohydrates present in glycolipids is increasing. D-Galactose has been found in the more unusual furanose form while residues of the acidic sugar, D-glucuronic acid, have also been detected. *N*-Acetylneuraminic acid is not the only sialic acid, *N*-glycolyl and N-O-disubstituted derivatives now being frequently detected, while other sugars, such as 3-deoxy-D-*glycero*-D-*galacto*-2-nonulosonic acid (KDN), are no longer considered to be rare. Molecular species can be found from the same source which have the same oligosaccharide chain along with the same sphingenine base but have different fatty acids present. The linkage between sterols and the carbohydrate chains are also through glycosidic bonds. The polyprenol-types are unique amongst the glycolipids as the link is via a phosphate bridge. Sterol- and polyprenol-types are not present in tissues and cells at high concentrations but the latter are widespread, being involved as transporters for the carbohydrate components of biologically active molecules.

2 EXTRACTION AND PURIFICATION

The very nature and definition of the name 'glycolipid' confirms that care must be taken during isolation and purification. Although carbohydrates are frequently, although not exclusively, water-soluble, the high concentration of hydroxy groups in their structures makes them hydrophilic. In contrast, lipids, by definition, are water-insoluble organic substances that can be extracted with nonpolar solvents. All glycolipids are, therefore, amphipathic. The combination of hydrophilic and hydrophobic regions in the same molecule accounts for many of their biological functions, especially in membranes, but does make their isolation and characterization more difficult.

The amphipathic nature means that the solvent system will generally require a polar and a nonpolar component. However, as the percentage of polar component to nonpolar component in a specific glycolipid (or lipoglycan) will depend on its composition and could approach either 0% or 100%, the ideal solvent system for any particular glycolipid needs to have similar polarity. No single solvent or mixture of solvents will, therefore, be capable of extracting all the glycolipids from a single biological sample.

Another paramount consideration when isolating glycolipids is the lability of the glycolipid in the extracting or purification solvent. The different covalent bonds in glycolipids have different labilities under acidic and alkaline

conditions. Glycosidic bonds between the sugar residues in the oligosaccharide component as well as to the lipid component are relatively stable under alkaline conditions but are cleaved in acidic conditions, especially at elevated temperatures. All glycosidic bonds are not equally labile in acid and this factor can be used in structural analyses. Other components, particularly esters substituents, are less stable in alkaline conditions than acidic conditions. Unless some structural information has been previously obtained, both acidic and alkaline conditions should be avoided during extraction. Enzymic hydrolysis may also occur if precautions are not taken. Before biological samples are extracted, enzymes, if present, need to be inactivated. This is particularly relevant for the plant glycolipids since phospholipases and galactolipases may be present. Under these circumstances, a prior treatment with boiling propan-1-ol is recommended.[1] Furthermore, care needs to be taken to prevent oxidation of unsaturated fatty acids as well as to minimize the co-extraction of plant pigments that can interfere with subsequent fractionation.

2.1 Extraction

Even though animal glycolipids are mainly glycosphingolipids and plant glycolipids are mainly monoglycosyldiacylglycerols, most of the extraction procedures are based on the related methods of Folch et al.[2] and Bligh and Dyer[3] that were published in the late 1950s. Both procedures recommend that the fresh tissue is extracted with a chloroform–methanol mixture for up to 1 h at temperatures from 4° to 25 °C (sometimes up to 100 °C is reported), with 4 °C being preferred. Temperatures above ca. 25 °C ought to be avoided unless labile constituents are known to be absent. Extraction periods as low as 3 min have been used. Periods of over 1 h are only recommended when fresh material is not available and lyophilized samples have to be used.

The recommended chloroform–methanol ratio depends on the proportions of lipid to carbohydrate in the glycolipid. Ratios from 1:2 through 1:1 to 2:1 have all been published and many authors cite sequential extractions with the different ratios. When a suspension of fresh tissue is used at a ratio of 3.75 volumes solvent to 1 volume tissue, the chloroform–methanol (1:2 v/v) mixture gives a one-phase extract containing chloroform–methanol–water (1:2:0.8 v/v/v). On adding more water and chloroform, two phases are produced with the glycolipids being found in the lower chloroform-rich phase and the nonlipid components being partitioned into the upper water-rich phase. The chloroform-rich phase is dried under vacuum and should be immediately fractionated or, if this is not practical, stored in a water-free form in chloroform–methanol (2:1) at −20 °C. When an unknown source has being extracted, the water-rich phase should be examined to see if any highly polar glycolipids are present in that fraction. Clearly, even using chloroform : methanol mixtures, there is no single procedure that can be recommended for the extraction of all glycolipids from every biological source and, ideally, the method that is used requires optimization for the source being investigated.

Alternative solvent systems have been proposed. When chloroform–methanol (1:2) is used, 0.4 M sodium acetate can been used instead of water and is reported to give improved results when gangliosides are present.[4] Another method, also reported to be particularly applicable to gangliosides, uses 0.01 M phosphate buffer in tetrahydrofuran.[5] For the extraction of higher plant glycolipids, fresh tissue is fragmented in dry ice and homogenized in boiling propan-2-ol for 1–2 min. This procedure is recommended when degradative enzymes may be present. It is also reported to minimize peroxidation.

2.2 General Purification

In parallel with the extraction of glycolipids, no single method can be used to separate all the glycolipids in an extract from co-extracted contaminants. Each procedure must be tested for the particular fraction required. When the chloroform–methanol extracts obtained from either the Folch et al.[2] or Bligh and Dyer[3] procedures are separated into a two-phase system, the lower phase contains the neutral glycolipids, particularly those with short carbohydrate chains, as well as neutral lipids and phospholipids, while the upper phase contains not only the gangliosides and the neutral glycolipids with longer carbohydrate chains but also contaminating glycoproteins and glycopeptides. When gangliosides are present, the water-rich upper layer can be dialyzed against distilled water at 4 °C for 2–3 days and the ganglioside fraction recovered by lyophilization of the retentate.

Many procedures have been published for general purification ranging from precipitation with propan-2-one to countercurrent distribution but the vast majority of applications use column chromatography. While diethylaminoethyl (DEAE) matrices have been used in the chromatographic methods, most procedures use silica gel when significant quantities of material are available. After applying the crude extract, such as the lower phase from a chloroform–methanol extract, to the column, neutral lipids and pigments are eluted by washing with chloroform and the glycolipids can be eluted by washing with chloroform–propan-2-one (1:1 v/v) or propan-2-one alone. Gradient elution can be used to provide an initial fractionation. Phospholipids, if present, can be recovered by subsequent elution with methanol–chloroform.

An adaptation of this procedure, initially described for gangliosides, was proposed by Williams and McCluer,[6] who used solid phase extraction (SPE) columns packed with a matrix containing bonded octadecylsilyl (ODS) groups. These cartridge columns (Sep-Pak™) are loaded with the ganglioside mixture present in the water-rich upper layer from a chloroform–methanol extract. Washing with water, under slight either positive or negative pressure, removes the ions, and the gangliosides are removed by washing with methanol or chloroform–methanol.

As well as ODS-bonded matrices, cartridges are now available with a range of matrices such as silica gel as well as both cation- and anion-exchange materials. Being produced under standard commercial conditions, the reproducibility between columns can almost be guaranteed. With care and suitable washing, the cartridges can be used a number of times. The use of SPE in all aspects of lipid analysis has been reviewed.[7] Gel permeation matrices have also been used for general purification procedures. The water-compatible Sephadex G-25[8] and the matrix designed for nonaqueous solvents, Sephadex LH-20,[9] have both been used.

2.3 Separation into Neutral Glycolipid and Acidic Glycolipid (Ganglioside) Fractions

The ability of the acidic sugar residue(s) in the ganglioside to bind to DEAE-groups attached to various supports is the main method and a procedure[10] using DEAE–Sephadex is typical. The authors recommend that the acetate form of the gel and the column be prewashed with chloroform–methanol–water (15 : 30 : 4 v/v/v) before loading the sample in the same solvent mixture. The neutral glycolipids are washed off using the same solvent and the gangliosides are then eluted with 0.8 M sodium acetate in methanol.

Since individual gangliosides may contain more than one acidic (sialic acid) residue, DEAE-matrices can also be used to fractionate mono-, di- and trisialoganglioside mixtures.[11] This procedure recommends that the crude ganglioside mixture is dispersed in chloroform–methanol (1 : 1) and mixed with the acetate form of DEAE–Sephadex A-25 with shaking. The suspension is diluted with methanol–water (15 : 4 v/v) before being packed into the column. The neutral lipids are eluted with chloroform–methanol–water (15 : 30 : 4 v/v/v) and neutral glycolipids with methanol. The gangliosides can then be fractionated using a concave gradient of 0–0.45 M ammonium acetate in methanol. Ganglioside species containing a single sialic acid residue are eluted first and those with 3–4 residues require the higher molarity of ammonium acetate.

2.4 Separation of Glycolipid Mixtures

2.4.1 Thin-layer Chromatographic Methods

TLC procedures for lipid separations have been reviewed[12] and their applications in glycolipid analyses are numerous. Using the term high-performance thin-layer chromatography (HPTLC) as a special form of this technique highlights the availability and reproducible quality of very fine grade silica gels. It has many advantages over HPLC. The fineness results in superior resolution. These grades of silica gel cannot be used in HPLC due to the pressure requirements, while very sensitive but destructive or corrosive reagents can be used for detection purposes. The plates are readily impregnated with additives such as Ag^+ or borate ions that enhance the separation of specific components such as unsaturated or vicinal dihydroxy-containing compounds.

Neutral solvent systems, such as chloroform–methanol–water (65 : 25 : 4 v/v/v), are recommended for neutral glycolipids. Derivatization of the glycolipid by, for example, acetylation is not reported to give any improvement to the separation. Any acceptable reagent that can detect nonreducing sugars, such as orcinol/H_2SO_4, can be used to visualize the components. When carrying out preparative TLC, care should be taken to prevent the plates drying out: quantitative recovery is not achieved even when ultrasonic treatment is used.

The best methods for the separation of gangliosides are also based on silica gel and detection is achieved with the resorcinol/HCl spray reagent. However, a basic solvent system (e.g. chloroform–methanol–water–ammonia (60 : 35 : 7 : 1 v/v/v/v)), or one containing Ca^{2+} ions (e.g. chloroform–methanol–0.2% aqueous $CaCl_2$ (60 : 40 : 9 v/v/v)) gives optimum separations in a one-dimensional system.[13]

Two-dimensional (2-D)/TLC is frequently used for analyzing glycolipid mixtures, particularly when only small quantities of material are available (see, for example, Sonnino et al.).[14] For neutral glycolipids, chloroform–methanol–0.2% $CaCl_2$ (60 : 35 : 8 v/v/v) has been recommended for both dimensions while for ganglioside molecular species, chloroform–methanol–0.2% $CaCl_2$ (55 : 45 : 10 v/v/v) has been recommended for the first dimension and propan-1-ol–water–25% NH_4OH (75 : 25 : 5 v/v/v) for the second.

A simple and quantitative purification procedure, based on TLC, is termed TLC blotting.[15] After 2-D HPTLC, the separated components are treated with primuline and visualized under ultraviolet (UV) light. The bands are marked with a colored drawing pencil then transferred, by TLC blotting, onto a polyvinylidene difluoride (PVDF) membrane. The major advantage is the yield and purity of the products. Direct extraction of an HPTLC plate rarely recovers more than 50% of the

lipid component while yields of 80–90% are regularly achieved using the PVDF membrane. Another benefit of PVDF membrane technology is the development of more sensitive detection methods. When the glycolipids on the PVDF membrane are mildly oxidized, they can be conjugated with streptavidin–horseradish peroxidase and, using enhanced chemiluminescence, are clearly visible at a concentration factor of 10 lower than detected by the orcinol reagent.[16] The method can be generally applied to the use of a specific antibody for detection and, since the staining procedure is nondestructive, the glycolipids can still be used for further molecular analyses.

A further adaptation of 2-D TLC, especially applicable to gangliosides, is termed TLC mapping.[17] The glycolipid mixture is first chromatographed in one dimension on a narrow NH_2-silica gel HPTLC strip using chloroform–methanol–1% diethylamine (50 : 47 : 15 v/v/v). This strip is then clamped to a 2-D TLC plate and developed with chloroform–methanol–25% NH_4OH (4 : 10 : 6 v/v/v). The partially separated glycolipids are transferred from the strip to the plate and chromatographed in the first dimension in a simultaneous process. The plate can then be run in the second dimension with chloroform–methanol–0.2% $CaCl_2$ as above. The method allows the determination of membrane constituents, particularly those that are present at low concentrations. A typical TLC map of the gangliosides from rat brain is shown in Figure 2.

2.4.2 High-Performance Liquid Chromatographic Methods

Because of the expansion in the use of HPTLC (see section 2.4.1), the applications using HPLC have not expanded as greatly as anticipated a decade ago. Columns that can compete with HPTLC by incorporating Ag^+ and borate ions have been developed and gradient elution at different temperatures is easy to perform. However, the pressure limits on the fineness of the silica gel that can be used as well as the detection limits are major obstacles. The IR detector is too insensitive while UV detection at 190–210 nm is nonspecific.

For native glycolipids, HPLC on silica is used for the separation of lipid classes while reversed-phase or bonded columns are used for the separation of molecular species.[18] Native neutral glycolipids can be separated on porous silica using a gradient involving propan-2-ol–hexane–water. For native gangliosides, silica gel and DEAE-linked glass beads have been used but the preferred method uses Lichrosorb-NH_2 or μ-Bondapak-NH_2 columns. Molecular species within mono-, di- and trisialoganglioside fractions can be separated on a single run.[19]

The introduction through chemical modification of the more UV-sensitive benzoyl or *p*-nitrobenzoyl groups into the glycolipids has greatly reduced the sensitivity problem. These groups also change the polarity of the glycolipid and allow a widely different set of solvent systems to be used. While the *p*-nitrobenzoyl derivatives are more UV-sensitive, the benzoyl derivatives usually give a better separation. The method used to prepare the derivatives is also important and depends on whether the native glycolipid needs to be recovered intact. Benzoyl chloride in pyridine gives a fully substituted product but treatment with benzoic anhydride and 4-dimethylaminopyridine as catalyst results only in O-benzoylation. The fully benzoylated derivative gives better separations than the O-benzoylated product. However, attempts to recover the

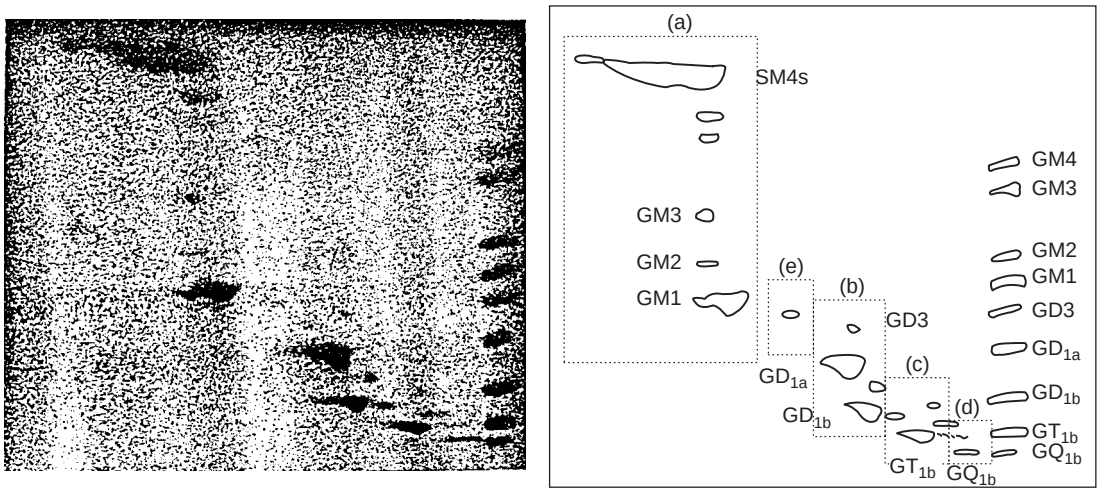

Figure 2 2-D thin layer chromatogram (TLC map) of the ganglioside mixture isolated from rat brain:[17] (a) to (d) refer to the mono- to tetrasialoganglioside fractions, respectively, and (e) refers to an, as yet, unidentified ganglioside fraction. (Reprinted with permission of Academic Press, New York from *Anal. Biochem.*, **227**, 195–200 (1995).)

native glycolipid with mild alkali only remove *O*-benzoyl groups: the *N*-benzoyl groups on either the sphingosine base or any amino sugars are stable.[20]

More recent developments with moving wire flame ionization detectors offer increased sensitivity for native glycolipids and the evaporative light-scattering detector is also a sensitive detection method when separating lipid classes, including glycolipids. This latter technique has been used to separate the nine phospho- and glycolipid classes present in plant membranes.[21]

2.4.3 Supercritical Fluid Chromatographic Methods

SFC is a relatively new technique that has been adapted by Reinhold et al.[22] for the separation of glycolipids. In general, thermally labile and nonvolatile compounds can be analyzed at lower temperatures than are possible with GC while the use of a flame ionization detector gives the method increased sensitivity over HPLC. A permethylated ganglioside mixture containing the mono-, di- and trisialogangliosides, G_{M1}, $G_{D1a,b}$ and G_{T1b}, has been separated on an SB-phenyl-5™ column at 120 °C using CO_2 as the supercritical fluid. The separation is shown in Figure 3. Possibly the main strength of the method is the ability to couple directly to a mass spectrometer and use the CO_2 along with NH_3 as chemical ionization reagent (see section 5.1).

2.4.4 Capillary and Thin-layer Electrophoretic Methods

These techniques have mainly been applied to the carbohydrate components released from glycoconjugates in general. The capillary and thin layer electrophoresis of both underivatized and derivatized carbohydrates have been reviewed by Oefner and Scherz,[23] while Jackson[24] has reviewed the high-resolution polyacrylamide gel electrophoresis of fluorophore-labeled carbohydrates. Gangliosides have been separated by capillary zone electrophoresis using cyclodextrin.[25]

3 CHEMICAL METHODS OF ANALYSIS

3.1 Composition of Glycolipids

3.1.1 Carbohydrate Profiles

The glycosidic bonds present in any polymeric carbohydrate, including glycolipids, are hydrolyzed under acidic conditions. Hydrolysis is most efficient when carried out under homogeneous conditions but the amphipathic nature of all glycolipids means that a single process cannot be recommended for all glycolipids. Glyolipids with a high proportion of carbohydrate residues are more likely to be soluble in an aqueous acidic solution while those with a small proportion of carbohydrate residues require

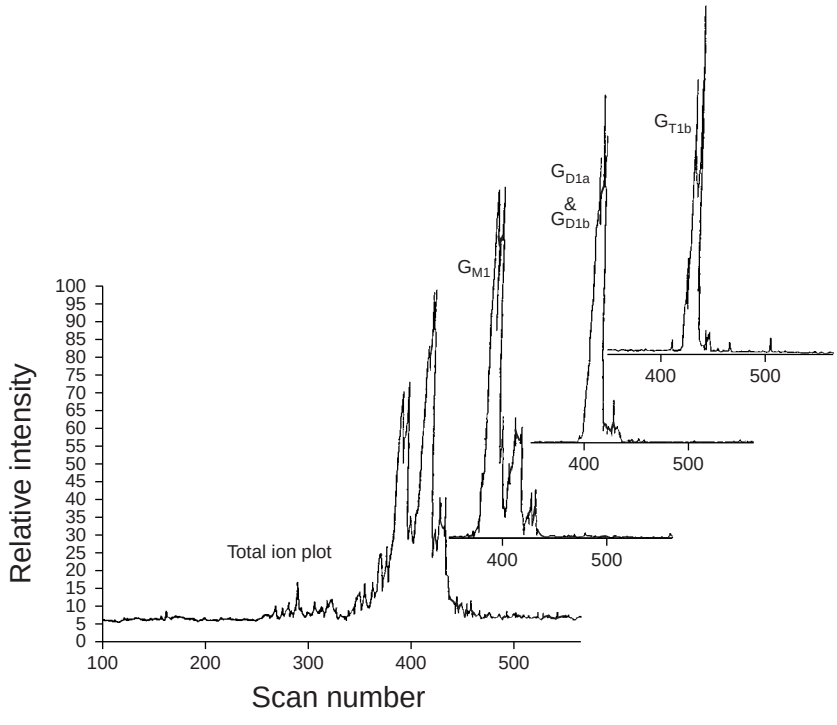

Figure 3 Supercritical fluid chromatography (SFC) mass spectrometry of the permethylated gangliosides G_{M1}, $G_{D1a,b}$ and G_{T1b} from bovine brain using NH_3/CO_2 as CI reagent gas.[22] (Reprinted with permission of Academic Press, New York from *Anal. Biochem.*, **193**, 24–34 (1991).)

a less polar solvent such as an acidic water–methanol mixture.

Furthermore, the glycosidic bonds of each type of carbohydrate residue found in glycolipids are not hydrolyzed at the same rate. In general, those involving amino sugars (e.g. galactosamine and its naturally occurring *N*-acetyl derivative) are relatively stable and require high concentrations of H^+ ions (ca. 0.3 M) for longer periods of time and at elevated temperatures. Glycosidic bonds involving the pyranose forms of most hexoses (e.g. gluco- and galactopyranose) are less stable. They still require high concentrations of H^+ ions (ca. 0.3 M) at elevated temperatures but not for such long periods. The precise conditions will depend on the sugar residue, its ring size (pyranose or furanose), its anomeric configuration (α or β) and the position of linkage to the aglycone. The aglycone can be another sugar residue, the sphingosine base, glycerol or whatever the lipid is. The most labile sugar residues found in glycolipids are the sialic acids or any furanosyl forms of sugars. In particular, galactofuranosyl groups have been identified in glycolipids from many microbiological sources.

In practice, it is not possible simply to use the conditions required to hydrolyze the most stable glycosidic bond. The reducing sugars released by hydrolysis are themselves unstable when exposed for prolonged periods to acidic conditions at elevated temperatures. The sialic acids present in gangliosides are particularly unstable. It must also be remembered that noncarbohydrate substituents attached to sugar residues, such as the *N*-acetyl groups on amino sugars and acetyl or sulfate groups present on neutral sugar residues, will also be hydrolyzed under acidic conditions. Preliminary analyses may be necessary to select the optimum conditions for a particular glycolipid.

Hydrolysis procedures involving H_2SO_4, HCl and trifluoracetic acid have been recommended but one using glacial acetic acid–0.15 M H_2SO_4 (9 : 1) at 80 °C for 6 h is frequently used. After neutralization of the H^+ or removal via an ion-exchange resin, the dried hydrolysate is reduced with borohydride to the corresponding alditol. Borodeuteride is used if subsequent analysis will include mass spectrometry. This procedure simplifies the chromatography by converting the α/β anomers of the pyranose/furanose forms into a single product. The alditol is then acetylated with acetic anhydride and products analyzed by GC or GC/MS. Many variations have been published but the general procedure of Watanabe et al.[26] is recommended. Improved gas chromatographic procedures have been published (see, for example, Blakeney et al.)[27] and milder acetylation conditions involving the use of 1-methylimidazole as catalyst[28] have been recommended. These hydrolysis conditions are necessary to release amino sugars but sialic acid residues are destroyed.

Methanolysis is recommended for gangliosides but can be used for other glycolipids. The glycolipid is treated with 1.0 M methanolic HCl at 85 °C for 18 h in a sealed tube. This procedure hydrolyzes the glycosidic bonds but simultaneously protects the vulnerable hemiacetal group by creating the methyl glycosides. After neutralization with Ag_2CO_3 or ion-exchange resin, and re-N-acetylation if amino sugars are present, the trimethylsilyl (TMS) derivatives of the methyl glycosides are prepared by conventional methods[29] and analyzed by GC.[30] The methanolysis procedure is acceptable for terminally-linked sialic acids but the conditions need to be made less severe when internal sialic acid residues are present. Since sialic (neuraminic) acid residues can be present carrying either *N*-acetyl or *N*-glycolyl substituents, as well as some residues being disubstituted, specific procedures involving GC, HPLC or colorimetric analyses have been reviewed.[31]

An interesting scheme has been proposed by Wiesner and Sweeley[32] where the glycolipid is methanolyzed with 0.75 M HCl in methanol. Using a glycolipid that contained a sialic acid residue and an amino sugar as well as neutral sugars, methanolysis was complete in 2 h at 110 °C. Instead of the TMS derivative, peracetylation introduced the more stable acetyl group. In contrast to other procedures, the individual sugar residues, the fatty acids and the sphingosine moieties could all be determined by a single GC separation in 45 min. This procedure is likely to be universally adopted.

The carbohydrate profiles from the plant monoglycosyldiacylglycerols are now known to be more complex than originally reported. In addition to the three carbohydrate moieties, β-D-galactopyranosyl, 6′-*O*-α-D-galactopyranosyl-β-D-galactopyranosyl and sulfoquinovosyl frequently cited, higher homologues of the second have been found as well as all β and mixed α/β linked galactosyl oligosaccharides.

3.1.2 Fatty Acid Profiles

The release of the individual fatty acids present in glycolipids is subject to the same general problems that were discussed for the carbohydrate profiles (see section 3.1.1). A purely aqueous hydrolysis is not practical and the most generally recommended method involves heating a methanol–H_2O–conc HCl (29 : 4 : 3) mixture at 80 °C for 18 h in a sealed tube under N_2. After extraction of the fatty acids (and some methyl esters), all the fatty acids are esterified with 1.5 M methanolic HCl and determined by capillary GC on a BP1 or BPX70 column.[33]

Methyl esters have been prepared using different reagents such as BF_3/methanol, $AlCl_3$/methanol and diazomethane. Pentafluorobenzyl esters have significant advantages if the more-sensitive electron capture detector is used for the GC whereas propan-2-yl esters give better

resolution for the different *trans* isomers. If the analysis involves GC/MS, the introduction of an N-containing ester group, such as a picolinyl or pyrrolidide group, is preferred for the MS.[34] When hydroxy fatty acids are present, the peak width of these acids is increased and tailing also occurs. The quality of the chromatograms and, therefore, the accuracy of quantification are improved by trimethylsilylating the hydroxy groups of the fatty acid methyl esters. The retention times are also reduced.

Reversed-phase HPLC offers an alternative strategy for the analysis of individual fatty acids and the topic has been reviewed.[35] The methyl esters are not sufficiently sensitive for most HPLC detection systems but the introduction of a phenacyl group allows more sensitive detection in the UV while introduction of the 4-bromomethyl-7-methoxycoumarin group allows detection by fluorescence. Ag^+ ion chromatography has special benefits when double bonds are present and the presence of borate ions assists the separation of any dihydroxy acids.

3.1.3 Other Components

The sphingosine bases are released from glycosphingolipids by treatment with methanol–H_2O–conc HCl (29 : 4 : 3 v/v/v) as described above for the release of fatty acids (see section 3.1.2). They remain in the acidic aqueous phase when the fatty acids and esters are extracted with petroleum ether but can then be extracted with diethyl ether when the pH is raised to >8. They are then determined by GC of their TMS derivatives[36] but the scheme proposed by Wiesner and Sweeley[32] for carbohydrate profiles (see section 3.1.1) can be used to determine the sphingosine bases simultaneously.

Sterols are released and isolated by a similar method to the sphingosine bases. They can be analyzed directly by GC but the acetate derivatives give improved resolution and analysis times. Glycerol is quantitatively released from plant glycolipids by the action of $LiAlH_4$ in diethyl ether. After purification, the glycerol content can be determined, without derivatization, by GC. It can also be determined spectrophotometrically[37] using the enzymes pyruvate kinase, lactate dehydrogenase and glycerokinase with appropriate cofactors.

3.2 Linkage Analysis of Carbohydrate Chains

Linkage analysis of the carbohydrate chain in a glycolipid will not give an unambiguous structure of the carbohydrate component of a glycolipid. It will give an indication of the relative position of each sugar residue present and, by analogy with known structures, the unknown structure may be derived. All nonreducing terminal residues are identified as well as main chain residues that have a single substitution point and branch points that are doubly substituted. The pattern is mainly determined after permethylation of the glycolipid (or the oligosaccharide derived from it). The product is then hydrolyzed to the partially methylated sugars, reduced, acetylated and the products determined by GC or GC/MS. The essential requirement is that every hydroxy group must be substituted. This may not always occur for native glycolipids; under these circumstances, it is preferable to release the oligosaccharide chain prior to methylation.

3.2.1 Isolation of Oligosaccharide Chains

A number of procedures have been used to isolate intact oligosaccharide chains from native glycolipids. Most rely on the presence of a double bond in most naturally occurring sphingosine bases. In the procedure using osmium tetroxide/periodic acid,[38] the glycolipid is first acetylated before the oxidizing agents break the double bond in the sphingosine base. The rump of the base is finally removed from the oligosaccharide chain with sodium methoxide, which also removes the acetyl-protecting groups to leave the intact oligosaccharide. Recoveries are not complete: the recovery from a ceramide with three sugar residues was 70% but <50% from a ceramide with two sugar residues. Ozone can also be used to split the unsaturated bond in the sphingosine base without the need to acetylate and the aldehydic rump is removed with aqueous base.

Trifluoroacetolysis results in different products depending on the ratio of trifluoroacetic acid to trifluoroacetic anhydride.[39] O-trifluoroacetylation is complete and amino functions are all N-trifluoroacetylated. *N*-Acetyl groups undergo transamidation to *N*-trifluoroacetyl groups but sialic acid residues are hydrolyzed from the gangliosides. If an unsaturated sphingosine base is present, the oligosaccharide is released, although yields are not quantitative. A 1 : 1 ratio is recommended for monoglycosylceramides, giving a better yield of glucose or galactose, while a 1 : 50 ratio gave a better yield with a triglycosylceramide.

3.2.2 Methylation Procedures

The ability to methylate oligosaccharide chains fully in intact glycolipids or those released from glycolipids (see section 3.2.1) with essentially no degradation is based on the work of Hakomori.[40] The method uses the methylsulfinyl anion, generated from NaH and dimethylsulfoxide (DMSO), to ionize all the hydroxy groups on a carbohydrate and these are converted to methyl ethers by reacting with iodomethane. Over the years, many modifications have been suggested, including a cleaner preparation of the methylsulfinyl anion from KH or potassium *t*-butoxide. The hazardous nature of the preparation of these carbanions, particularly that

prepared from KH, needs to be stressed. Although relatively stable as dispersions in mineral oil, pure hydrides are extremely reactive and must only be used in small quantities and with great care.

As for the carbohydrate profiles (see section 3.1.1), the composition of the carbohydrate component of the glycolipid, and therefore the stability of the individual glycosidic linkages, will affect the hydrolysis conditions employed to depolymerize the permethylated glycolipid. If amino sugars are present, stronger hydrolysis conditions are needed such as glacial acetic acid–5 M H_2SO_4 (19 : 1 v/v) at 80 °C for 18 h. When only neutral sugars (or sialic acids) are present, permethylated glycolipids can be depolymerized by heating with 95% formic acid at 100 °C for 2 h. Conventional methods are used to clean up the products. 2 M Trifluoroacetic acid at 121 °C for 1 h is a convenient alternative when only neutral sugars are present: the acid can be conveniently removed in a stream of N_2.

Most applications use GC or GC/MS to identify and quantify the products and the alditol acetate method is that of choice (see section 3.1.1). The procedure has been reviewed.[41] A more recent advance has been selectively to hydrolyze the permethylated glycolipid at a labile bond. In this procedure, the oligosaccharides released are reduced, the generated hydroxy groups are converted to ethyl ethers (as for methylation) and products separated by HPLC. GC/MS or MS alone allows structural features of the oligosaccharides to be determined.

3.3 Anomeric Configuration of Carbohydrate Residues

There are no chemical methods currently available that will definitively identify the anomeric configurations of the carbohydrate residues in a glycolipid, and biochemical and/or physical methods (see sections 4 and 5) are far more reliable.

3.4 Linkage Analysis of Fatty Acids

In plant glycoglycerolipids, as well as variations in the structures of the carbohydrate components (see section 3.1.1), the fatty acids esterified to the glycerol backbone can vary. The structure, isolation and analysis of plant glycolipids have been reviewed.[42] Irrespective of the nature of the carbohydrate, it is always attached to C-3 of glycerol in the *sn* configuration. The fatty acids, therefore, are attached to C-1 and C-2 and the overall pattern as well as the distribution of individual fatty acids in each molecular species is biologically and industrially important.

Lipases have been isolated from a number of sources that are regioselective for the *sn*-1 ester group and so are capable of releasing the intact acyl group from C-1 (*sn*-1) of all glycoglycerolipids except those containing sulfoquinovosyl groups. This fatty acid can be directly determined (see section 3.1.2) and the lysoglycolipid, which still contains the acyl group attached to C-2 (*sn*-2) can then be hydrolyzed and this fatty acid also determined. Only one extracellular lipase, isolated from *Rhizopus arrhizus*, is capable of carrying out the above reaction with the sulfoquinovosyl-containing glycoglycerolipids.

4 BIOCHEMICAL METHODS OF ANALYSIS

4.1 Linkage Analysis and Anomeric Configuration via Specific Glycosidases

Exo-glycosidases act by cleaving the glycosidic bond of that specific glycoside when present in a nonreducing terminal position. The procedure is best explained by referring to an actual example. A pentaglycosylceramide of the globo series has the structure shown in Figure 4.[43] The glycosidic bond involving the nonreducing terminal α-*N*-acetyl-D-galactosaminosyl residue is specifically removed by the action of the purified α-*N*-acetyl-D-galactosaminidase from hog liver. In common with most glycolipid methods, a general protocol, applicable to all possible situations, cannot be given but some recommendations can still be made. First, the purity of the enzyme is paramount. Even traces of other glycosidase activities will yield erroneous results. The amphipathic nature of

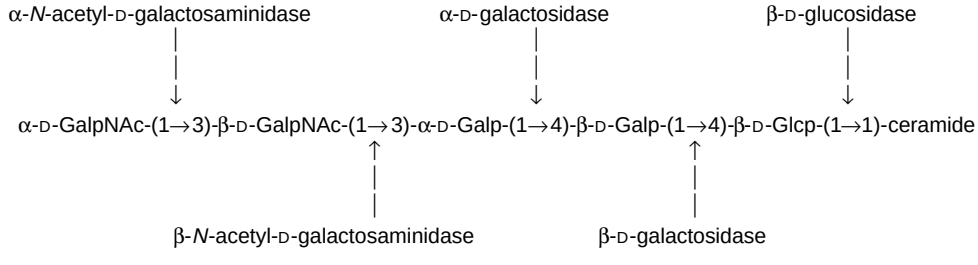

Figure 4 Structure of the pentaglycosylceramide of the globo series showing how the sequential action of an α-*N*-acetyl-D-galactosaminidase, a β-*N*-acetyl-D-galactosaminidase, an α-D-galactosidase, a β-D-galactosidase and a β-D-glucosidase are required to determine the oligosaccharide sequence and anomeric configurations.

glycolipids means that a detergent such as sodium taurocholate is usually incorporated in the reaction buffer. The method is not just applicable to any nonreducing terminal residues. Addition of chloroform–methanol to the aqueous solution of reaction products gives a two-phase system with the released sugar residue in the upper aqueous phase and the remainder of the glycolipid, now containing one fewer sugar residue, in the lower phase. The released sugar residue is recovered and determined (see section 3.1.1) and the new glycolipid can be recovered and treated with another *exo*-glycosidase.

The method does have to be treated with some caution. If the internal α-D-galactosyl residue had been another α-*N*-acetyl-D-galactosaminosyl residue, its glycosidic bond would not have been susceptible to the *exo*-glycosidase. However, if the internal β-*N*-acetyl-D-galactosaminosyl residue had been another α-*N*-acetyl-D-galactosaminosyl residue, both α-*N*-acetyl-D-galactosaminosyl residues would have been hydrolyzed at the same time. The method is straightforward but, for an unknown glycolipid, can be time-consuming. The method, therefore, allows the carbohydrate sequence to be determined as well as the anomeric configuration of the different sugar residues. It cannot give any reliable information on the position of linkage to the next sugar residue in the sequence. Highly purified *exo*-glycosidases are now commercially available for each of the anomeric configurations of most of the sugar residues found in neutral glycolipids.

The gangliosides pose a greater problem. Sialidases have been isolated and purified but their activity is reported to be dependent on the environment of each sialic acid as well as the reaction medium. The presence of *N*-glycolyl and disubstituted neuraminic acids, as well as the more common *N*-acetyl residues, affects the specificity.

4.2 Linkage Analysis via Immunological Methods

The use of immunological methods has greatly expanded and the technique owes much to the work of Hakomori and Young.[44] One of the earliest reports[45] used monoclonal antibodies that had been raised against the terminal oligosaccharide sequences β-D-Galp-(1→4)-β-D-GlcNAc-(1→R) and α-L-Fucp (1→2)-β-D-Galp-(1→4)-β-D-GlcNAc-(1→R). Glycolipids carrying these sequences could be differentiated from each other. Because of their specificity, monoclonal antibodies are able to detect the individual sugar residue and its anomeric configuration as well as the linkage arrangement within a relatively short segment of the nonreducing terminus of the oligosaccharide component. As more specific monoclonal antibodies are produced, the range of sequences that can be characterized will increase. Furthermore, the combination of detection through the use of antibodies coupled with TLC and TLC blotting onto PVDF membranes leads to a level of detection that would not have been considered feasible even in the late 1980s.

As the general methods proposed are very similar to the original ones, two more recent examples are cited to show what is possible. Using monoclonal antibodies to type1 precursor, H-type 1, Le-a and Le-b epitopes, small quantities of Le-b blood group glycolipids were detected in a Lewis negative individual.[46] Previous studies by Marcus's group had produced seven murine monoclonal antibodies that were active against Le-x. Using a phage display library, two novel antibodies have been produced whose affinities are 10^2 higher than the originals.[47] As well as binding to Le-x, one of the new antibodies will bind to sialyl Le-x while the other will also bind to neolactosylceramide.

5 PHYSICAL METHODS OF ANALYSIS

5.1 Mass Spectroscopy

Of all the analytical methods and techniques that have been applied to the analysis of glycolipids, those involving MS have made the greatest impact since the late 1980s. Indeed, it is now very rare to see a research paper on glycolipids without some MS-related analysis or characterization. The theory of the different MS techniques, electrospray ionization (ESI), MALDI, FAB and so on, will not be addressed: there are many reviews currently available. This section will try to summarize the relevant techniques that can be applied to the analysis of glycolipids, giving as many examples from as wide a range of glycolipid types as possible.

One of the major advantages of MS is the very small sample size that can be analyzed. Another major asset is the ability to couple MS to other analytical instruments. GC/MS has been available for many years. The interface between MS and SFC is straightforward and LC/MS (liquid chromatography MS) is now in routine use. The earlier unreliability of the coupling interface between the two systems has largely been solved. Even the analysis of spots or bands on HPTLC is now compatible with MS largely through the use of blotting onto PVDF membranes (see section 2.4.1). Special mention will be made of combined methods such as collision induced dissociation FAB/(CID)/MS/MS (see later).

5.1.1 Matrix-assisted Laser Desorption/Ionization Methods

The technique of MALDI, as applied to oligosaccharides and glycoconjugates, has been reviewed[48] while specific results for glycosphingolipids[49] have been published.

Glycolipids, as well as the oligosaccharides derived from them, give mainly $[M + Na]^+$ ions which is in contrast to glycoproteins where $[M + H]^+$ ions are more abundant. Resolution is generally superior with ion cyclotron resonance and magnetic sector instruments over time-of-flight (TOF) instruments. Neutral glycolipids are able to produce strong spectra from a number of different matrices but the gangliosides do undergo extensive fragmentation due to the destruction of the sialic acid residues.

The best matrices for neutral glycosphingolipids are 2,5-dihydroxybenzoic acid (2,5-DHB), α-cyano-4-hydroxycinnamic acid and esculetin, particularly in the positive-ion mode. When *N*-acetylhexosaminyl residues were not present in the glycolipid and 2,5-DHB was used as matrix, fragmentation occurred between the carbohydrate and lipid revealing the structure of these components. With other matrices, loss of the acylamide from the ceramide was dominant and gave the distribution of C atoms between the acyl and sphingosyl groups. In contrast, when *N*-acetylhexosaminyl residues were present, fragmentation was not as specific but did allow much structural information to be gathered. For the gangliosides, loss of CO_2 and sialic acids both occurred while other fragment ions, derived from glycosidic cleavage, identified the carbohydrate sequence. The low-resolution MALDI spectra of the ganglioside G_{M2}, obtained with the matrices 2,5-DHB and esculetin are shown in Figure 5.

Other reports from Harvey's group have compared the structural information that can be obtained about the oligosaccharides using different methods of fragmentation, ways of stabilizing the sialic acid residues in gangliosides when using positive-ion MALDI methods and how the structure of the oligosaccharide will affect the signal strength in both TOF and magnetic sector instruments. A new technique, called delayed extraction MALDI, has been applied to the analysis of cerebrosides and sulfatides.[50] Both types of glycolipid were deacylated by microwave-mediated saponification and the long chain base components as well as the lysoglycosphingolipids could be characterized. Unlike classical methods, which only deacylate cerebrosides, the sulfatides are also deacylated and fewer by-products occur than are detected after acidic methanolysis.

A novel application for the sequencing of neutral oligosaccharides has been the combination of MALDI/MS with *exo*-glycosidase digestion.[51] The molecular mass profile is obtained by MALDI/TOF/MS using 2,5-DHB and 1-hydroxyquinoline as the matrix. The sample is then recovered from the target and the matrix removed by droplet dialysis. *Exo*-glycosidase digestion follows with the buffer and enzyme being removed by droplet dialysis and absorption onto a Nafion membrane, respectively, before a further MALDI/TOF/MS analysis. The cycle can be repeated as necessary. The method is designed to give more reliability to the sequence analysis as well as being much faster.

5.1.2 Fast Atom Bombardment Methods

The analysis of glycolipids has been simplified by the use of soft ionization techniques such as FAB. The structural characterization of the sulfoquinovosyl diacylglycerols from a cyanobacterium, dissolved in chloroform–methanol, in both positive ion (containing *m*-nitrobenzyl alcohol) and negative ion (containing triethanolamine) modes is shown in Figure 6.[52] The former gives $[M - H + 2Na]^+$ ions and the two main peaks at m/z 865 and 839 correspond to a total fatty acid composition of 34 : 1 and 32 : 0, respectively. The fatty acid profile is confirmed by the negative ions, $[M - H]^-$, at 819 and 793.

Another specific technique is the direct negative ion FAB/MS of glycosphingolipids separated on TLC plates.[53] The plates are moved within the ion source by a motor-driven probe. The technique has been optimized on neutral glycolipids containing five- and seven-sugar residues but has also been able to give a sequence determination for a highly branched eight-sugar glycolipid. As well as information on the oligosaccharide, molecular species with different ceramide compositions could also be detected.

5.1.3 Electrospray Ionization Mass Spectrometry

Electrospray ionization mass spectrometry (ESIMS) is another type of soft ionization that has been shown to give molecularly related ions with high sensitivity and accuracy. It is reported to be very accommodating for polysialogangliosides and the spectrum of a tetrasialoganglioside has been published.[54] There are references to the technique alone but it is likely to be more useful in conjunction with other secondary techniques such as MS/MS/CID/MS/MS.

5.1.4 Liquid Secondary Ion Mass Spectrometry

This technique has been used for sulfated gangliosides.[55] From the normal negative liquid secondary ion mass spectrometry (LSIMS) spectra, monosulfated species gave $[M - H]^-$ ions while bis-sulfated species gave $[M + Na - H]^-$ ions. In the high-energy CID spectra, only ions containing sulfate esters were detected, with major differences depending whether the sulfate group was attached to a nonreducing terminal residue or was present on an inner hexose residue. The low-energy CID spectra were very simple with only one or two product ions, specific for each sulfated glycolipid, being detected. Combining the data, oligosaccharide structures differing in location and number of sulfate esters could be distinguished from each other.

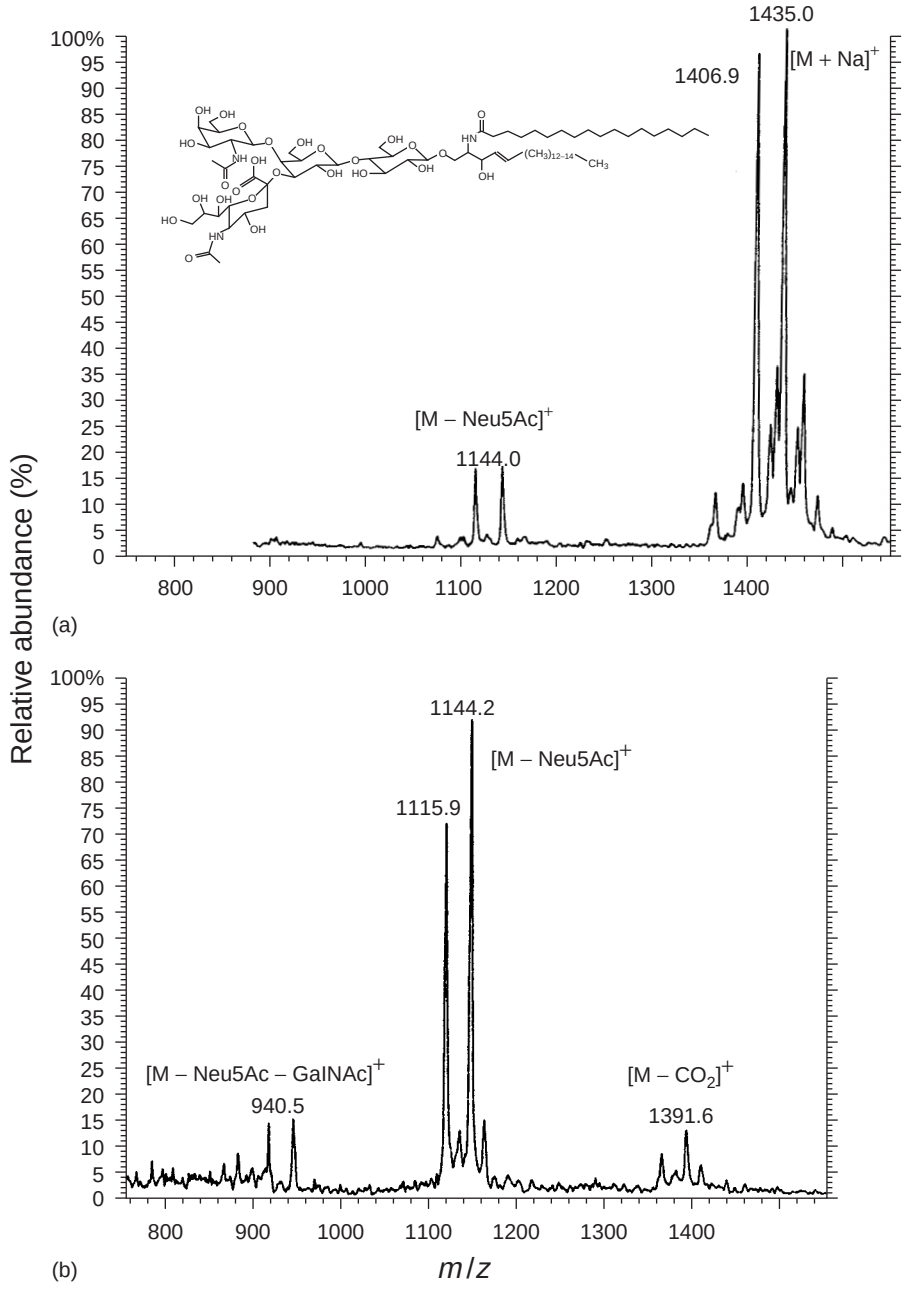

Figure 5 MALDI mass spectra of the ganglioside G_{M2} with 2,5-DHB (a) and esculetin (b) as the matrices.[49] (Reprinted with permission of John Wiley & Sons, Chichester from *J. Mass Spectrom.*, **30**, 1311–1324 (1995).)

5.1.5 Tandem Mass Spectrometry

MS/MS, as its name implies, employs two stages of mass analysis. The general aim is to produce particular, often molecular, ions in one system and then to fragment these ions in another system. The fragment ions produced in the second system will be characteristic of the compound. In theory, two isomeric compounds could produce molecular ions of similar mass but should be differentiated by the fragmentation pattern. The technique, as applied to glycolipids, has been reviewed[56] and some examples are given below.

The sulfoquinovosyldiacylglycerols, discussed under FAB (see section 5.1.2),[52] had a major molecular ion $[M - H + 2Na]^+$ at m/z 865 in the positive-ion FAB spectrum. The high mass region, obtained from the CID breakdown of this ion (see Figure 7: an MS/MS system), had ions at m/z 739, 725, 711, 697, 683, 669 and 655 ($^3I_1$) which are characteristic of an 18:1 unsaturated fatty acid group, namely oleic acid, at C-1. The ions at

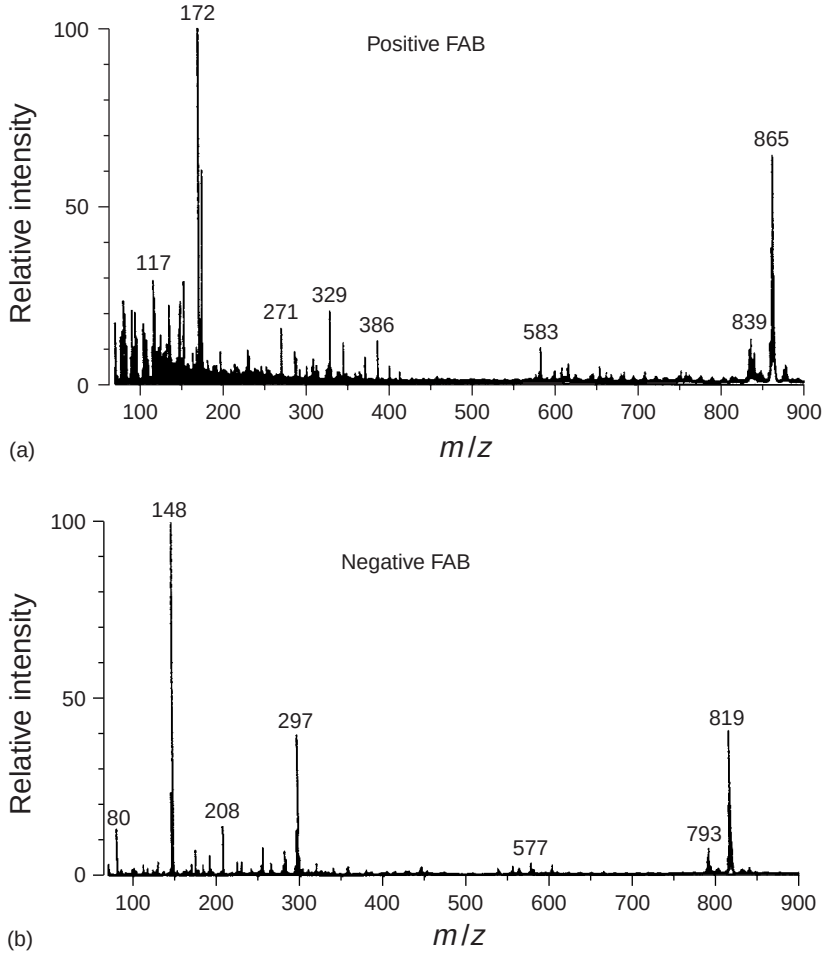

Figure 6 FAB mass spectra of the sulfoquinovosyl diacylglycerides from the wild-type bacterium *Synechocystis* sp. PCC 6803 in positive ion (a) and negative ion (b) modes.[52] (Reprinted with permission of John Wiley & Sons, Chichester from *J. Mass Spectrom.*, **32**, 968–977 (1997).)

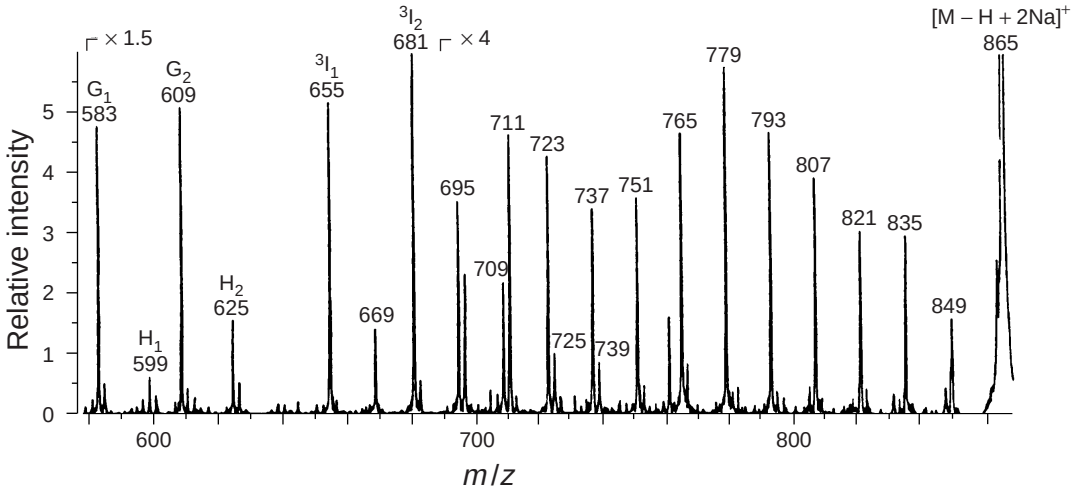

Figure 7 High mass region of the FAB/CID/MS/MS spectrum of the positive ion (*m/z* 865) from Figure 6(a) that confirms the total structure of the major sulfoquinovosyldiacylglyceride.[52] (Reprinted with permission of John Wiley & Sons, Chichester from *J. Mass Spectrom.*, **32**, 968–977 (1997).)

Table 1 ^{1}H Chemical shifts (δ) and $J_{1,2}$ coupling constants for the pentaglycosylceramide from Figure 4 and its lower homologues[59]

Glycosphingolipid	δ	$J_{1,2}$	δ	$J_{1,2}$	δ	$J_{1,2}$	δ	$J_{1,2}$	δ	$J_{1,2}$
1	–	–	–	–	–	–	–	–	4.10	7.7
2	–	–	–	–	–	–	4.22	7.3	4.17	7.7
3	–	–	–	–	4.81	4.0	4.27	7.7	4.16	8.1
4	–	–	4.52	8.1	4.81	3.6	4.26	7.7	4.16	7.7
5	4.74	3.6	4.56	8.5	4.82	3.6	4.27	7.7	4.17	7.7

1. β-D-Glcp-(1→1)-ceramide
2. β-D-Galp-(1→4)-β-D-Glcp-(1→1)-ceramide
3. α-D-Galp-(1→4)-β-D-Galp-(1→4)-β-D-Glcp-(1→1)-ceramide
4. β-D-GalpNAc-(1→3)-α-D-Galp-(1→4)-β-D-Galp-(1→4)-β-D-Glcp-(1→1)-ceramide
5. α-D-GalpNAc-(1→3)-β-D-GalpNAc-(1→3)-α-D-Galp-(1→4)-β-D-Galp-(1→4)-β-D-Glcp-(1→1)-ceramide

m/z 737, 723, 709, 695 and 681 ($^3I_2$) are due to a 16:0 saturated fatty acid, palmitic acid, at C-2. In contrast, the predominant molecular ion $[M + Na]^+$ at *m/z* 801 from monogalactosyldiacylglycerol contained linoleic (18:2) acid residues at both C-1 and C-2.

The gangliosides discussed in the ESIMS section (see section 5.1.3)[54] were further characterized by CID MS/MS. Using both negative and positive ion modes, the isomeric disialogangliosides, G_{D1a} and G_{D1b}, were readily distinguishable showing that the method should be applicable to any positional isomers of glycolipids. A final example is the characterization of glycosphingolipids from bovine erythrocytes.[57] By comparison with authentic samples, detection levels in the ranges 10^{-15}–10^{-9} mole are quoted for mono- and difucosylated glycolipids of the neolactose series. These glycolipids are very minor components found in the central nervous system but are reported to be of major biological importance.

5.2 Nuclear Magnetic Resonance Spectroscopy

5.2.1 ^{1}H Nuclear Magnetic Resonance

NMR spectroscopy has been applied to the analysis of intact glycolipids, their peracetyl and permethyl derivatives as well as oligosaccharides released from them (see section 3.2.1) but, compared with mass spectrometry, it is less usable owing to the relatively large amount of material needed for analysis. The choice of procedure is dependent on the information required. However, when any possible interrelationship between the different components of the glycolipid are required, the intact and underivatized glycolipid must be used. The ^{1}H signals from the ceramide component can all be unambiguously assigned and do not overlap with any of the carbohydrate signals. For the oligosaccharide component, the ^{1}H signals from the glycosidic protons are well separated from all the other ring protons.

Dabrowski, Hanfland and Egge[58] reviewed the earlier applications. Although the data is quite old, the chemical shifts and coupling constants given in Table 1 are reliable enough to confirm the structure of the carbohydrate component of a pentaglycosylceramide of the globo-series (see Figure 4) and its lower homologues.[59] Even with a low field instrument, the size of the coupling constant of each glycosidic proton gives its anomeric configuration. In the example cited in Table 1, the coupling constants for the α- and β-*N*-acetyl-D-galactosaminosyl and α- and β-D-galactosyl residues are distinct. Apart from sugars with the manno-configuration, α-glycosidic protons have $J_{1,2}$ coupling constants of 3–4 Hz while β-anomers show values of 7–9 Hz. As with other procedures, the amphipathic nature of glycolipids means that special solvent systems are necessary for dissolution. Many reports use 2% D_2O in deuterated DMSO while others use pyridine-d_5.

The introduction of instruments with field strengths up to 800 MHz can give unambiguous results with as little as 100 μg of glycolipid. Future generations of instruments are unlikely to improve significantly on this lower limit. For a straightforward 1-D ^{1}H spectrum, 400 MHz instruments allow the coupling constants to be obtained not only of the anomeric protons but all of the protons attached to the ring carbon atoms to be assigned and determined. Thus, an unambiguous assignment of each sugar residue and its configuration is obtained.

The theory and practice of NMR spectroscopy has advanced greatly since the mid 1970s with the introduction of 2-D spectroscopy and related methods and there are many general reviews (see, for example, Goldman).[60] Few are specifically applied to glycolipids although animal glycerosphingolipids have been reviewed[61] and plant glycodiacylglycerols[42] are included in a general review. The different types of 2-D NMR spectra can be classed under shift correlation spectroscopy (COSY) methods or spin-lattice relaxation methods. The latter are due to the time-dependent interaction between two spins

producing a cross-reaction, or nuclear Overhauser effect spectroscopy (NOESY). There are further refinements of each technique and for more information relevant papers should be consulted.

As one example, the conformational analysis of the ganglioside G_{D1a} has been determined by a NOESY method (see Figure 8 for 2-D spectra).[62] The inter-residual hydrogen distances calculated from the NMR data were in good agreement with the lowest energy model previously suggested. The internal and branched sialic acid residue is stacked underneath the *N*-acetyl-D-galactosamine residue with a hydrogen bond between the carboxy group of the sialic acid and the amino group of the *N*-acetyl-D-galactosamine residue. Both sialic acid residues are orientated almost perpendicular to the plane of the neutral tetrasaccharide core. By contrast, the structural analyses of the oligosaccharide chains of glycolipids were determined by magnetic transfer methods.[63] COSY and homonuclear Hartmann Hann (HOHAHA) methods were able to extract subspectra of constituent monosaccharides and identify their sugar type.

Figure 8 Slices from the 2-D NOESY spectrum of the ganglioside G_{D1a} at 200, 400 and 800 ms corresponding to the chemical shift of H-3^{f}_{ax} (a–c) and H-3^{e}_{ax} (d–f), respectively.[62] (Reprinted with permission of the American Chemical Society from *J. Am. Chem. Soc.*, **113**, 3236–3246 (1991).)

NMR has been used to understand the arrangement of glycolipids in bilayer membranes.[64] The sphingolipid metabolites behaved as single populations of lipid amphiphiles that were uniformly dispersed in the membrane and displayed rapid symmetric motion about their long molecular axis. The whole area of the application of NMR to lipid phase behavior and lipid diffusion has been reviewed.[65]

5.2.2 ^{13}C Nuclear Magnetic Resonance

Far more structural information can be obtained from a ^{13}C-NMR spectrum than from ^{1}H-NMR. Owing to the low natural abundance of the ^{13}C isotope, the technique does suffer from the need for relatively large quantities of the native glycolipid. A specific example of the technique was the elucidation of the structure of the digalactosyldiacylglycerol present in oat grain.[66] By using ^{13}C-NMR, as well as ^{1}H-NMR, the identity of the glycosyl moieties in the polar head group, the linkage position between the groups and the anomeric configuration of the moieties were all determined.

^{13}C-NMR has also been used to determine the positional distribution of fatty acids in glycoglycerolipids from different plant families.[67] The acyl carbonyl shifts of fatty acids attached at C-1 are downfield from those of C-2. In addition, the shifts also depend on the distance to the nearest double bond. Hence, plant families with high proportions of all-*cis*-Δ-7,10,13-hexadecatrienoic and all-*cis*-Δ-6,9,12,15-octadecatetraenoic acids could be determined and the results were in agreement with the distribution as determined using a lipase.

ACKNOWLEDGMENTS

The author wishes to thank the following for permission to publish data and figures from published material: Academic Press, New York, for permission to publish data for Figures 2 and 3, John Wiley & Sons, Chichester, for permission to publish data for Figures 5, 6 and 7 and American Chemical Society for permission to publish data for Figure 8.

ABBREVIATIONS AND ACRONYMS

CID	Collision Induced Dissociation
COSY	Correlation Spectroscopy
DEAE	Diethylaminoethyl
DMSO	Dimethylsulfoxide
ESI	Electrospray Ionization

ESIMS	Electrospray Ionization Mass Spectrometry
FAB	Fast Atom Bombardment
GC	Gas Chromatography
HOHAHA	Homonuclear Hartmann Hann
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
KDN	3-Deoxy-D-*glycero*-D-*galacto*-2-nonulosonic Acid
LSIMS	Liquid Secondary Ion Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/ Ionization
MS	Mass Spectroscopy
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
ODS	Octadecylsilyl
PVDF	Polyvinylidene Difluoride
SFC	Supercritical Fluid Chromatography
SPE	Solid Phase Extraction
TLC	Thin-layer Chromatography
TMS	Trimethylsilyl
TOF	Time-of-flight
UV	Ultraviolet
2-D	Two-dimensional
2,5-DHB	2,5-Dihydroxybenzoic Acid

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis • Monosaccharides and Sugar Alcohol Analysis

REFERENCES

1. M. Kates, *Techniques of Lipidology*, 2nd edition, Elsevier, Amsterdam, 1986.
2. J. Folch, M. Lees, G.H. Sloane-Stanley, 'A Simple Method for the Isolation and Purification of Total Lipids from Animal Tissues', *J. Biol. Chem.*, **226**, 497–509 (1957).
3. E.G. Bligh, W.J. Dyer, 'A Rapid Method of Total Lipid Extraction and Purification', *Can. J. Biochem. Physiol.*, **37**, 911–917 (1959).
4. B.L. Slomiany, A. Slomiany, 'Branched Blood Group A-active Fucolipids of Hog Gastric Mucosa', *Biochim. Biophys. Acta*, **486**, 531–540 (1977).
5. G. Tettamanti, F. Bonali, S. Marchesini, V. Zambotti, 'New procedure for the extraction, purification, and fractionation of brain gangliosides', *Biochim. Biophys. Acta*, **296**, 160–170 (1973).
6. M.A. Williams, R.H. McCluer, 'The Use of Sep-Pak™ C_{18} Cartridges During the Isolation of Gangliosides', *J. Neurochem.*, **35**, 266–269 (1980).
7. W.W. Christie, 'Solid-phase Extraction Columns in the Analysis of Lipids', in *Advances in Lipid Methodology – One*, ed. W.W. Christie, Oily Press, Ayr, 1–17, 1992.
8. R.E. Wuthier, 'Purification of Lipids from Nonlipid Contaminants on Sephadex Bead Columns', *J. Lipid Res.*, **7**, 558–561 (1966).
9. H. Baumann, E. Nudelman, K. Watanabe, S. Hakomori, 'Neutral Fucolipids and Fucogangliosides of Rat Hepatoma HTC and H35 Cells, Rat Liver, and Hepatocytes', *Cancer Res.*, **39**, 2637–2643 (1979).
10. R.K. Yu, R.W. Ledeen, 'Gangliosides of Human, Bovine, and Rabbit Plasma', *J. Lipid Res.*, **13**, 680–686 (1972).
11. T. Momoi, S. Ando, Y. Magai, 'High Resolution Preparative Column Chromatographic System for Gangliosides using DEAE-Sephadex and a New Porous Silica, Iatrobeads', *Biochim. Biophys. Acta*, **441**, 488–497 (1976).
12. J.C. Touchstone, 'Thin-layer Chromatographic Procedures for Lipid Separation', *J. Chromatogr., B Biomed. Applic.*, **671**, 169–195 (1995).
13. S. Hakomori, 'Chemistry of Glycosphingolipids', in *Handbook of Lipid Research 3: Sphingolipid Biochemistry*, eds. J.N. Kanfer, S. Hakomori, Plenum Press, New York and London, 1–65, 1983.
14. S. Sonnino, R. Ghidoni, V. Chigorno, M. Masserini, G. Tettamanti, 'Recognition by Two-dimensional Thin-layer Chromatography and Densitometric Quantification of Alkali-labile Gangliosides from the Brain of Different Animals', *Anal. Biochem.*, **128**, 104–114 (1983).
15. T. Taki, T. Kasama, S. Handa, D. Ishikawa, 'A Simple and Quantitative Purification of Glycosphingolipids and Phospholipids by Thin-layer Chromatography Blotting', *Anal. Biochem.*, **223**, 232–238 (1994).
16. L.E. Hernández, N.J. Brewin, B.K. Drøbak, 'Detection of Glycolipids as Biotinylated Derivatives using Enhanced Chemiluminescence', *Anal. Biochem.*, **241**, 59–66 (1996).
17. K. Watanabe, M. Nishiyama, 'A New Type of Two-dimensional Thin-layer Chromatography (TLC Mapping) for Analysis of Acidic Glycosphingolipid Molecular Species', *Anal. Biochem.*, **227**, 195–200 (1995).
18. F.B. Jungalwala, M.D. Ullman, R.H. McCluer, 'High-performance Liquid Chromatography of Glycosphingolipids in Brain Disease', in *Chromatography of Lipids*

in Biomedical Research and Clinical Diagnosis, ed. A. Kuksis, Journal of Chromatography Library, Elsevier, Amsterdam, 348–377, Vol. 37, 1987.

19. M.M. Whalen, G.C. Wild, W.D. Spall, R.J. Sebring, 'Separation of Underivatized Gangliosides by Ion-exchange High-performance Liquid Chromatography', *Lipids*, **21**, 267–270 (1986).
20. R.H. McCluer, J.E. Evans, 'Quantitative Analysis of Brain Galactosylceramides by High-performance Liquid Chromatography of their Perbenzoyl Derivatives', *J. Lipid Res.*, **17**, 412–418 (1976).
21. G.A. Picchioni, A.E. Watada, B.D. Whitaker, 'Quantitative High-performance Liquid Chromatography Analysis of Plant Phospholipids and Glycolipids Using Light-scattering Detection', *Lipids*, **31**, 217–221 (1996).
22. M.V. Merritt, D.M. Sheeley, V.N. Reinhold, 'Characterization of Glycosphingolipids by Supercritical Fluid Chromatography–Mass Spectrometry', *Anal. Biochem.*, **193**, 24–34 (1991).
23. P. Oefner, H. Scherz, 'Capillary Electrophoresis and Thin-layer Electrophoresis of Carbohydrates', in *Advances in Electrophoresis*, eds. A. Chrambach, M.J. Dunn, B.J. Radola, VCH, Weinheim, 155–224, Vol. 7, 1994.
24. P. Jackson, 'The Analysis of Fluorophore Labeled Saccharides by High-resolution Polyacrylamide Gel Electrophoresis', in *Advances in Electrophoresis*, eds. A. Chrambach, M.J. Dunn, B.J. Radola, VCH, Weinheim, 225–262, Vol. 7, 1994.
25. Y.S. Yoo, Y.S. Kim, G.-J. Jhon, J. Park, 'Separation of Gangliosides Using Cyclodextrin in Capillary Zone Electrophoresis', *J. Chromatogr.*, **652**, 431–439 (1993).
26. K. Watanabe, S. Hakomori, R.A. Childs, T. Feizi, 'Characterization of a Blood Group I-active Ganglioside. Structural Requirements for I and i Specificities', *J. Biol. Chem.*, **254**, 3221–3228 (1979).
27. A.B. Blakeney, P.J. Harris, R.J. Henry, B.A. Stone, 'A Simple and Rapid Preparation of Alditol Acetates for Monosaccharide Analysis', *Carbohydr. Res.*, **113**, 291–299 (1983).
28. S. Iwata, T. Narui, K. Takahashi, S. Shibata, 'Separation of Carbohydrates by IC on an Aqueous Silica-gel Column', *Carbohydr. Res.*, **133**, 157–162 (1984).
29. R.E. Chambers, J.R. Clamp, 'Assessment of Methanolysis and Other Factors Used in the Analysis of Carbohydrate-containing Materials', *Biochem. J.*, **125**, 1009–1018 (1971).
30. M.F. Chaplin, 'A Rapid and Sensitive Method for the Analysis of Carbohydrate Components in Glycoproteins Using Gas-liquid Chromatography', *Anal. Biochem.*, **123**, 336–341 (1982).
31. M.F. Chaplin, 'Monosaccharides', in *Carbohydrate Analysis: A Practical Approach*, 2nd edition, eds. M.F. Chaplin, J.F. Kennedy, IRL Press, Oxford, 1–41, 1994.
32. D.A. Wiesner, C.C. Sweeley, 'Microscale Analysis of Glycosphingolipids by Methanolysis, Peracetylation, and Gas Chromatography', *Anal. Biochem.*, **217**, 316–322 (1994).
33. W.W. Christie, *Gas Chromatography and Lipids*, Oily Press, Ayr, 1989.
34. D.J. Harvey, 'Mass Spectrometry of Picolinyl and Other Nitrogen-containing Derivatives of Fatty Acids', in *Advances in Lipid Methodology – One*, ed. W.W. Christie, Oily Press, Ayr, 19–80, 1992.
35. B. Nikolova-Damyanova, 'Reversed-phase High-performance Liquid Chromatography: General Principles and Application to the Analysis of Fatty Acids and Triacylglycerols', in *Advances in Lipid Methodology – Four*, ed. W.W. Christie, Oily Press, Dundee, 193–251, 1997.
36. Y. Fujino, M. Ohnishi, 'Constituents of Ceramide and Ceramide Monohexoside in Rice Bran', *Chem. Phys. Lipids*, **17**, 275–289 (1976).
37. M. Eggstein, E. Kuhlmann, 'Triglycerides and Glycerol (Alkaline Saponification)', in *Methods of Enzymatic Analysis*, ed. H.U. Bergmeyer, Academic Press, New York, 1871–1878, Vol. 4, 1974.
38. D.L. MacDonald, L.M. Patt, S. Hakomori, 'Notes on Improved Procedures for the Chemical Modification and Degradation of Glycosphingolipids', *J. Lipid Res.*, **21**, 642–645 (1980).
39. A. Lundblad, S. Svensson, B. Löw, L. Messeter, B. Cedergren, 'Release of Oligosaccharides from Human Erythrocyte Membranes of Different Blood Group P Phenotypes by Trifluoroacetolysis', *Eur. J. Biochem.*, **104**, 323–330 (1980).
40. S. Hakomori, 'Rapid Permethylation of Glycolipids and Polysaccharides, Catalysed by Methylsulfinyl Carbanion in Dimethyl Sulfoxide', *J. Biochem. (Tokyo)*, **55**, 205–208 (1964).
41. S.B. Levery, S. Hakomori, 'Microscale Methylation Analysis of Glycolipids Using Gas Chromatography Chemical Ionisation Mass Fragmentography Ion Monitoring', in *Methods in Enzymology*, ed. V. Ginsburg, Academic Press, New York, 13–25, Vol. 138, 1987.
42. E. Heinz, 'Plant Glycolipids: Structure, Isolation and Analysis', in *Advances in Lipid Methodology – Three*, ed. W.W. Christie, Oily Press, Dundee, 211–332, 1996.
43. K. Stellner, H. Saito, S. Hakomori, 'Determination of Aminosugar Linkages in Glycolipids by Methylation. Aminosugar Linkages of Ceramide Pentasaccharides of Rabbit Erythrocytes and Forssman Antigen', *Arch. Biochem. Biophys.*, **155**, 464–472 (1973).
44. S. Hakomori, W.W. Young, 'Glycolipid Antigens and Genetic Markers', in *Handbook of Lipid Research. 3. Sphingolipid Biochemistry*, eds. J.N. Kanfer, S. Hakomori, Plenum Press, New York and London, 381–436, 1983.
45. W.W. Young, J. Portoukalian, S. Hakomori, '2 Monoclonal Anti-carbohydrate Antibodies Directed to Glycosphingolipids with a Lacto-N-glycosyl Type-II Chain', *J. Biol. Chem.*, **21**, 967–972 (1981).

46. S.M. Henry, P-A. Jovall, S. Ghardashkhani, M.L. Gustavsson, B.E. Samuelsson, 'Structural and Immunochemical Identification of Le-b Glycolipids in the Plasma of a Group O Le(a-b-) Secretor', *Glycoconjugate J.*, **12**, 309–317 (1995).
47. Q. Dinh, N-P. Weng, M. Kiso, H. Ishida, A. Hasegawa, D.M. Marcus, 'High Affinity Antibodies Against Le-x and Sialyl Le-x from a Phage Display Library', *J. Immunol.*, **157**, 732–738 (1996).
48. D.J. Harvey, 'Matrix-assisted Laser Desorption-ionisation Mass Spectrometry of Oligosaccharides and Glycoconjugates', *J. Chromatogr., A*, **720**, 429–446 (1996).
49. D.J. Harvey, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Sphingo- and Glycosphingolipids', *J. Mass Spectrom.*, **30**, 1311–1324 (1995).
50. T. Taketomi, A. Hara, K.-I. Uemura, E. Sugiyama, 'Rapid Method of Preparation of Lysoglycosphingolipids and their Confirmation by Delayed Extraction Matrix-assisted Laser Desorption/Ionisation Time-of-flight Mass Spectrometry', *J. Biochem. (Tokyo)*, **120**, 573–579 (1996).
51. B. Küster, T.J.P. Navan, D.J. Harvey, 'Rapid Approach for Sequencing Neutral Oligosaccharides by Exoglycosidase Digestion and Matrix-assisted Laser Desorption/Ionisation Time-of-flight Mass Spectrometry', *J. Mass Spectrom.*, **31**, 1131–1140 (1966).
52. Y.H. Kim, J.S. Yoo, M.S. Kim, 'Structural Characterization of Sulfoquinovosyl, Monogalactosyl and Digalactosyl Diacylglycerols by FAB/CID/MS/MS', *J. Mass Spectrom.*, **32**, 968–977 (1997).
53. B. Lanne, B.M. Olsson, T. Larsson, K.A. Karlsson, 'Fast Atom Bombardment Mass Spectrometric Analysis of Glycosphingolipids by Direct Desorption from Thin Layer Chromatography Plates: Application of the Method to a Receptor-active 8-sugar Glycolipid', *Eur. Mass Spectrom.*, **2**, 361–368 (1996).
54. T. Ii, Y. Ohashi, Y. Nagai, 'Structural Elucidation of Underivatised Gangliosides by Electrospray-ionization Tandem Mass Spectrometry', *Carbohydr. Res.*, **273**, 27–40 (1995).
55. K. Tadano-Aritomi, H. Kubo, P. Ireland, M. Okuda, T. Kasami, S. Handa, I. Ishizuka, 'Structural Analysis of Mono- and Bis-sulfated Glycosphingolipids by Negative Liquid Secondary Ion Mass Spectrometry with High- and Low-energy Collision-induced Dissociation', *Carbohydr. Res.*, **273**, 41–52 (1995).
56. C.E. Costello, J.E. Vath, 'Tandem Mass Spectrometry of Glycolipids', in *Methods in Enzymology*, ed. J.A. McCloskey, Academic Press, New York, 738–768, Vol. 163, 1990.
57. H. Perreault, X.L. Hronowski, O. Koul, J. Street, R.H. McCluer, C.E. Costello, 'High Sensitivity Mass Spectral Characterization of Glycosphingolipids from Bovine Erythrocytes, Mouse Kidney and Fetal Calf Brain', *Int. J. Mass Spectrom. Ion Processes*, **169**, 351–370 (1997).
58. J. Dabrowski, P. Hanfland, H. Egge, 'Analysis of Glycosphingolipids by High-resolution Proton Nuclear Magnetic-resonance Spectroscopy', in *Methods in Enzymology*, ed. V. Ginsburg, Academic Press, New York and London, 69–86, Vol. 83, 1980.
59. J. Dabrowski, P. Hanfland, H. Egge, 'Structural Analysis of Glycosphingolipids by High-resolution Proton Nuclear Magnetic Resonance Spectroscopy', *Biochemistry*, **19**, 5652–5658 (1980).
60. M. Goldman, *Quantum Description of High-resolution NMR in Liquids*, Oxford University Press, Oxford, 1992.
61. T.A.W. Koerner, J.H. Prestegard, R.K. Yu, 'Oligosaccharide Structure by Two-dimensional Proton Nuclear Magnetic-resonance Spectroscopy', in *Methods in Enzymology*, ed. V. Ginsburg, Academic Press, New York, 38–59, Vol. 138, 1987.
62. S. Sabesan, J.O. Duus, T. Fukunaga, K. Bock, S. Ludvigsen, 'NMR and Conformational Analysis of Ganglioside G_{D1a}', *J. Am. Chem. Soc.*, **113**, 3236–3246 (1991).
63. F. Inagaki, 'Structural Analysis of Oligosaccharide Chains of Glycolipids by Magnetization Transfer Methods', *J. Synthetic Org. Chem. (Japan)*, **51**, 502–515 (1993).
64. A.C. Rigby, K.R. Barber, C.W.M. Grant, 'H-2-NMR Study of 2 Probe-labeled Glycosphingolipid-derived Signaling Modulators in Bilayer Membranes', *Biochim. Biophys. Acta*, **1240**, 75–82 (1995).
65. G. Lindblom, 'Nuclear Magnetic Resonance Spectroscopy and Lipid Phase Behavior and Lipid Diffusion', in *Advances in Lipid Methodology – Three*, ed. W.W. Christie, Oily Press, Ayr, 133–209, 1996.
66. J.B. Hauksson, M.H.J. Bergquist, L. Rilfors, 'Structure of Digalactosyldiacylglycerol from Oats', *Chem. Phys. Lipids*, **78**, 97–102 (1995).
67. B.W.K. Diehl, H. Herling, I. Riedl, E. Heinz, '^{13}C-NMR Analysis of the Positional Distribution of Fatty Acids in Plant Glycolipids', *Chem. Phys. Lipids*, **77**, 147–153 (1995).

Glycoprotein Analysis: General Methods

Henri Debray, Jean-Claude Michalski, and Geneviève Spik
Université des Sciences et Technologies de Lille, France

1 Introduction **2**
2 Glycan Structures: An Overview **2**
2.1 Carbohydrate–Protein Linkage 2
2.2 Carbohydrate Composition 3
2.3 N-linked Glycoproteins 3
2.4 O-linked Glycoproteins 6
2.5 Structure and Biosynthesis of Glycosylphosphatidylinositol Anchors 8
3 Glycoprotein Isolation and Characterization **8**
3.1 Mucin Isolation 8
3.2 Immunoaffinity Chromatography 9
3.3 Isolation of Glycoforms by Anion-exchange Chromatography 9
3.4 Isolation of Glycoforms by Lectin Affinity Chromatography 9
3.5 Characterization of Glycoprotein Glycosylation 15
4 Isolation and Characterization of Glycopeptides and Glycans **16**
4.1 Proteolytic Digestion of Glycoprotein 16
4.2 Methods for Glycan Release 17
4.3 Fractionation of Glycopeptides and Glycans 18
4.4 Structural Characterization of Glycans 24
5 Conclusions **27**
Acknowledgments **27**
Abbreviations and Acronyms **27**
Related Articles **28**
References **28**

Glycosylation is a common co- and post-translational modification of a protein which may have profound effects on protein structure and function. Many biologically interesting proteins are glycosylated at their asparagine, serine, and threonine residues. Glycosylation has now been recognized as being more ubiquitous and structurally varied than all other types of post-translational modifications combined. The glycosylation pattern of a glycoprotein is not random but is dependent on cell type, physiological conditions and, in some cases, disease states. Carbohydrate chains or glycans attached to the peptide backbone are classified according to the nature of the linkage to the peptide and their structure; both of these parameters allow the definition of common cores. In addition to the extreme structural diversity and complexity of glycan structures, one feature of protein glycosylation is the presence at a single site of any one of a number of glycan structures. This property gives rise to an extremely heterogeneous population of glycoproteins, termed glycoforms. The complexity of the glycan structures, the multiple substitutions (microheterogeneity) at glycosylation sites, and the structural diversity associated with the protein backbone itself represent an enormous task for analytical structural studies.

Indirect methods such as immunological approaches using antibodies or lectins provide some structural information concerning the nature of glycans. Prior to their structural study, glycans have to be released from the protein backbone by enzymatic or chemical methods. Different endoenzymes such as N-glycanase® allow a complete deglycosylation of glycoproteins, providing a pool of oligosaccharides which may be further fractionated. Establishment of the site heterogeneity and assessment of the different glycoforms require isolation and analysis of the individual glycopeptides generated after hydrolysis with specific proteases. Fractionation of glycans or glycopeptides can be achieved by different high-performance liquid chromatography (HPLC) or electrophoretic procedures. Lectins, due to their extreme specificities with regard to monosaccharide or oligosaccharide motifs, represent powerful tools for sequential glycan fractionation. Traditionally, a complete structural analysis of the glycan structures, including determination of the carbohydrate sequence and the sugar linkage, has been a tedious, multidisciplinary task. Characterization and determination of the relative amount of individual constituent monosaccharides is generally obtained by gas–liquid chromatography (GLC) after hydrolysis of the oligosaccharide chain. Information concerning the nature of linkages and branching points is furnished by the methylation procedure. Nevertheless, these chemical sequencing methods often necessitate milligrams of material. To progress from this situation, sensitive physicochemical or enzymatic methods have been developed. The nuclear magnetic resonance (NMR) study of oligosaccharides, even of the less sensitive of the physicochemical techniques, is extremely powerful because it allows determination in a mixture of glycans, the complete structure of individual glycans including nature of monosaccharides, sequence, type of linkages, anomericity and branching points. The recent advances in mass

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

spectrometry (MS) techniques, mainly matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI), have provided very sensitive means for the analysis of oligosaccharides and glycopeptides. The gain in sensitivity associated with a minimum number of sample manipulation steps allows work at the submicrogram level. The use of enzyme reagents is rapidly becoming popular in modern carbohydrate analysis because of their inherent selectivities for a sugar substrate and its linkage type. Exoglycosidases can be used in conjunction with other methods such as MALDI/MS or fluorophore-assisted carbohydrate electrophoresis (FACE) separation. Complete structural information is now feasible for low-microgram and submicrogram quantities of a glycoprotein; this quality of performance places the field of carbohydrate analysis closer to the sequencing methods available for other biomolecules, proteins or nucleic acids.

1 INTRODUCTION

In 1907, The American Society of Biological Chemists and the American Physiology Society defined the glycoproteins as ''compounds of the protein molecule with a substance or substances containing carbohydrate other than nucleic acid''. In 1986, the International Union of Pure and Applied Chemistry–International Union of Biochemistry (IUPAC–IUB) Joint Commission on Biochemical Nomenclature[1] described glycoconjugates as various types of compound consisting of carbohydrates covalently linked with other types of chemical constituent. This general definition of glycoconjugates includes glycoproteins, glycopeptides, glycolipids, peptidoglycans, and lipopolysaccharides.

This article refers only to glycoproteins and glycopeptides, which are compounds containing one or more oligosaccharides or glycans covalently linked to the polypeptide chain by a glycosyl linkage.

According to the type of glycosylation, the family of glycoproteins may be subdivided into glycoproteins, mucins, and proteoglycans. The glycans linked to the polypeptide chain of the proteoglycans are called glycosaminoglycans. The glycosaminoglycans are linear polymers of alternating residues of hexosamine and uronic acid or galactose. They are often sulfated at various positions.[2,3] A survey of methodological challenges for glycosamino glycan/proteoglycan analysis has been published recently.[4] The glycoproteins are ubiquitous components. They result from an enzymatic addition of glycans to the proteins biosynthesized by viruses, many archaebacteria, some eubacteria, plants, and animals. The glycosylation step of a protein represents one of the most important post-translational modifications of this protein.

Glycoproteins are widely distributed in biological fluids mainly in serum, intracellularly, and at cell surfaces. The mucins are highly glycosylated proteins found at the surface of all mucosa. Mucin-like components have been identified at cell surfaces, mainly in cancer cells. Significant advances in the understanding of the structure and the biosynthesis of glycoproteins have revealed that different glycans may be linked to an individual glycosylation site, leading to a mixture of glycovariants or glycoforms within a single polypeptide chain.[5] In parallel with our increasing knowledge of the glycoprotein structures, important functional roles of glycans have been recognized, not only in the modulation of the conformation and the physicochemical properties of the protein, but also in the biological interactions of the glycoforms with free or membrane target molecules. Most of the recent advances in structure and function of the glycoproteins have been extensively reviewed.[6–11]

Recombinant antibodies, hormones, and enzymes are now produced for clinical applications. A connection between the biological functions of these recombinant glycoproteins and the structure of their glycans has been established.[12] Arising from the importance of the development of genetic engineering techniques for the production of recombinant glycoproteins, techniques are becoming necessary for rapid and reliable determination of the glycan structure as well as for the glycan distribution among the different glycosylation sites.

This article describes the general features of the glycan structures found in glycoproteins and mucins, and a combination of chromatographic techniques is proposed for their isolation and structural characterization.

2 GLYCAN STRUCTURES: AN OVERVIEW

Recent advances in the physical methods used to determine the primary structure of glycoprotein glycans now enable the analysis of very small amounts of material. This has rapidly increased the number of glycan structures, more than 1500 of which are now gathered in the CarbBank. The diversity of these glycans is mainly explained by the nature of the covalent carbohydrate–protein linkage, the composition in monosaccharides, the branching of the structures, and the origin of the glycoproteins.

2.1 Carbohydrate–Protein Linkage

For a long time the only two types of covalent linkages known were *N*-glycosylated (alkali stable) and *O*-glycosylated (alkali labile). In the N-linked glycans, the nitrogen of the side chain of an asparagine is linked to the β-anomeric carbon of an *N*-acetylglucosamine

Table 1 Linkage between glycan and polypeptide chain

Type	Mono-saccharide	Amino acid	Occurrence
N-Glycoside	β-GlcNAc	Asn	Common
	β-GalNAc	Asn	Archaebacteria
	β-Glc	Asn	Archaebacteria
	α/β-Glc	Asn	Animals
	L-Rha	Asn	Eubacteria
O-Glycoside	α-GalNAc	Ser/Thr	Animals (mucin type)
	β-GalNAc	Ser/Thr	Animals
	β-GlcNAc	Ser/Thr	Animals (intracellular type)
	α-Xyl	Ser	Animals (collagen type)
	α-Fuc	Ser/Thr	Animals
	β-Gal	Hyp	Animals
	α-Glc	Tyr	Animals (glycogenin type)
	β-Glc	Ser/Thr	Animals, eubacteria
	α-Man	Ser/Thr	Animals, yeasts
	α-Gal	Hyp/Ser	Plants, eubacteria
	β-L-f-Ara	Hyp	Plants
	β-Gal	Tyr	Eubacteria
C-Glycoside	α-Man	Trp	Human RNase

GlcNAc, *N*-acetyl-D-glucosamine; GalNAc, *N*-acetyl-D-galactosamine; Glc, D-glucose; Xyl, D-xylose; Fuc, L-fucose; Gal, D-galactose; Man, D-mannose; f-Ara, L-furanoarabinose; Hyp, hydroxyproline; Tyr, tyrosine; Trp, tryptophan.

and constitutes the 2-acetamido-*N*-(L-aspart-4-yl)-2-deoxy-β-D-glucopyranosylamine or N^4-(*N*-acetyl-β-D-glucosaminyl)asparagine linkage. The asparagine residue belongs to the peculiar sequon AsnX (Ser or Thr, and rarely Cys) of a polypeptide chain (Asn = L-asparagine; Ser = serine; Thr = threonine; Cys = cysteine). Here X is any amino acid except proline (which completely blocks the glycosylation) and some amino acids such as tryptophan, glutamic acid, aspartic acid, or leucine (which create conformational constraints and partially inhibit the glycosylation). All Asn residues included in the triplet are not glycosylated, because the glycosylation depends on the accessibility of the sequon to the oligosaccharyltransferase, which transfers the complete immature oligosaccharide to the asparagine. In bacterial glycoproteins, the *N*-acetylglucosamine linked to Asn is replaced by an *N*-acetylgalactosamine, a glucose, or a rhamnose residue (Table 1).[13] In *O*-glycosidically linked glycans, the anomeric carbon of a monosaccharide is attached through an α- or β-*O*-glycosidic linkage to the hydroxyl oxygen of serine, threonine, hydroxylysine, hydroxyproline or tyrosine (Table 1).[13] The *N*-acetyl-α-D-galactosaminyl-Ser or -Thr is the most common carbohydrate–protein linkage found in the mucins. Proposed peptide motifs containing Ser and Thr specifying *O*-glycosylation sites have been described[14] and a database containing the sequences around *O*-glycosylation sites is on the Internet. Recently, a fundamentally different kind of protein glycosylation has been found in which the C2 atom of tryptophan is attached by a C–C linkage to the anomeric carbon of an α-mannose residue.[15] This *C*-mannosylation of tryptophan has been shown by tandem electrospray MS analysis in human ribonuclease 2. In addition to *N*-, *O*-, and *C*-glycosyl linkages, in membrane glycoproteins, the anchor of the glycolipid glycosylphosphatidylinositol (GPI) is conjugated to the protein *C*-terminus through an amide bond.[16]

2.2 Carbohydrate Composition

A landmark for the carbohydrate composition of a mature *N*-glycosylated protein is the presence of the monosaccharides D-mannose (Man), *N*-acetyl-D-glucosamine (GlcNAc), D-galactose (Gal), L-fucose (Fuc), *N*-acetyl-D-galactosamine (GalNAc), D-xylose (Xyl), L-furanoarabinose (f-Ara), and sialic acids which are various derivatives of neuraminic acid,[17] mainly *N*-acetyl-neuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc). During biosynthesis,[18] immature glycans contain one to three residues of Glc involved in the trisaccharide structure Glc(α1–2)Glc(α1–3)Glc(α1–3) linked to the oligomannosidic type glycan $Man_9GlcNAc_2$. The presence of glucuronic acid (GlcA) is not common in N-linked glycans. In O-linked glycans, the main monosaccharides are GalNAc, Gal, GlcNAc, Fuc, and sialic acids. Even if mannose has been found in some O-linked glycans, it is considered to be unusual here. Other peculiar monosaccharides,[13] such as the deaminated neuraminic acid 2-keto-3-deoxynononic acid (Kdn) and various *O*-sulfated, *O*-phosphorylated, *O*-methylated and methyl *O*-phosphorylated monosaccharides, have been identified either in O- or N-linked glycans. The monosaccharides found in glycoproteins have the pyranose configuration and sometimes the unusual furanose configuration.

2.3 N-linked Glycoproteins

2.3.1 General Structure of N-*Glycans*

The structure of N-linked glycans is very heterogeneous; however, all glycans possess a pentasaccharide core constituted of two *N*-acetylglucosamine and three mannose residues (Figure 1). This backbone may be substituted either by one *N*-acetylglucosamine (known as the intersecting GlcNAc residue), one xylose, or one to two fucose residues. The complexity of the glycan structures results from the addition to Man(α1–3) and Man(α1–6) of one to five *N*-acetylglucosamine residues. In fact, these

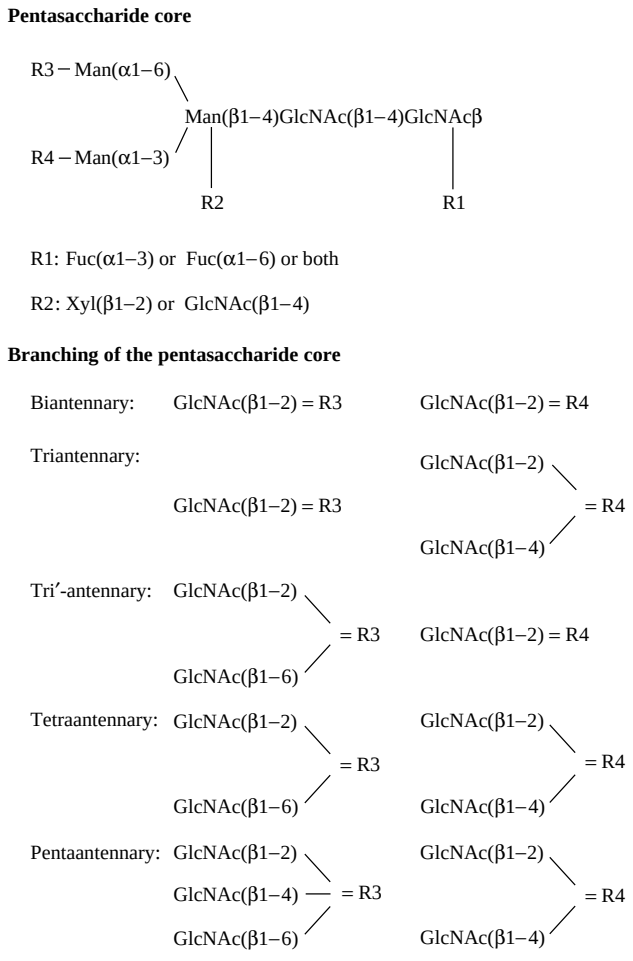

Figure 1 Modifications of the pentasaccharide core of *N*-glycosylated proteins.

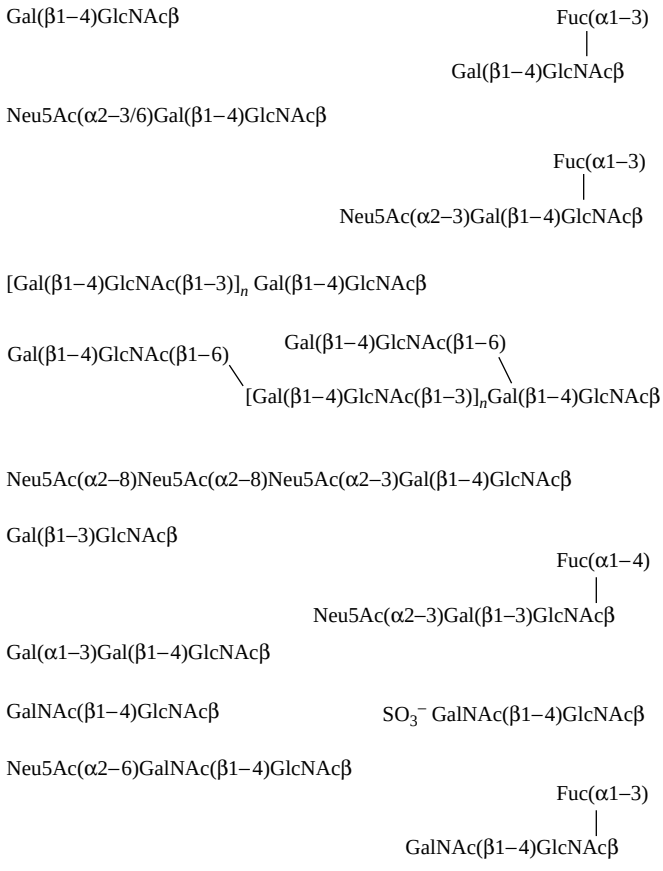

Figure 2 Side chains in *N*-acetyllactosaminic or complex-type glycans.

GlcNAc initiate a branching and generate di-, tri-, tri′-, tetra-, and pentaantennary glycans[19] (Figure 1). In mature glycans, extension of the branching is realized by addition of various peripheral structures including Gal, GlcNAc, GalNAc, Fuc, and sialic acids in order to complement the antennae (Figure 2). The glycan structures obtained are called *N*-acetyllactosamine or complex type, and when the extension consists of several *N*-acetyllactosamine units (Gal(β1−4)GlcNAc) associated by a (β1−3) linkage, the structures are called polylactosamine type.

Analysis of the three-dimensional structure of glycans has revealed that the pentasaccharide core is rigid whereas the antennae are mobile. The glycans have interactions with the polypeptide chain and occupy an important volume at the surface of the glycoprotein.[20,21]

2.3.2 Biosynthesis of N-*Glycans*

Biosynthesis of the N-linked glycans and the control of glycosylation induced by the conformation of the polypeptide chain are now well established. This biosynthesis involves, in the first step, the transfer of the oligosaccharide ($Glc_3Man_9GlcNAc_2$) from a dolichyloligosaccharide precursor to the nascent polypeptide chain. Mature glycans are obtained following the activity of trimming glycosidases (α-glucosidases and α-mannosidases) and specific glycosyltransferases found in the endoplasmic reticulum and the Golgi apparatus.[22,23] During the biosynthesis, it may be noticed that the glycan structure is affected by the compartments traversed within a cell and also by the speed with which the glycoprotein passes through these different compartments. Examples of oligomannose or high mannose, hybrid and biantennary *N*-acetyllactosamine glycan structures obtained during the biosynthesis pathway and found in the structure of native glycoproteins are given in Figure 3.

2.3.3 Variability of Glycosylation

A systematic and comparative study on glycan primary structures of transferrin from biological fluids of different species has revealed that the glycosylation pattern of the individual transferrin molecule depends

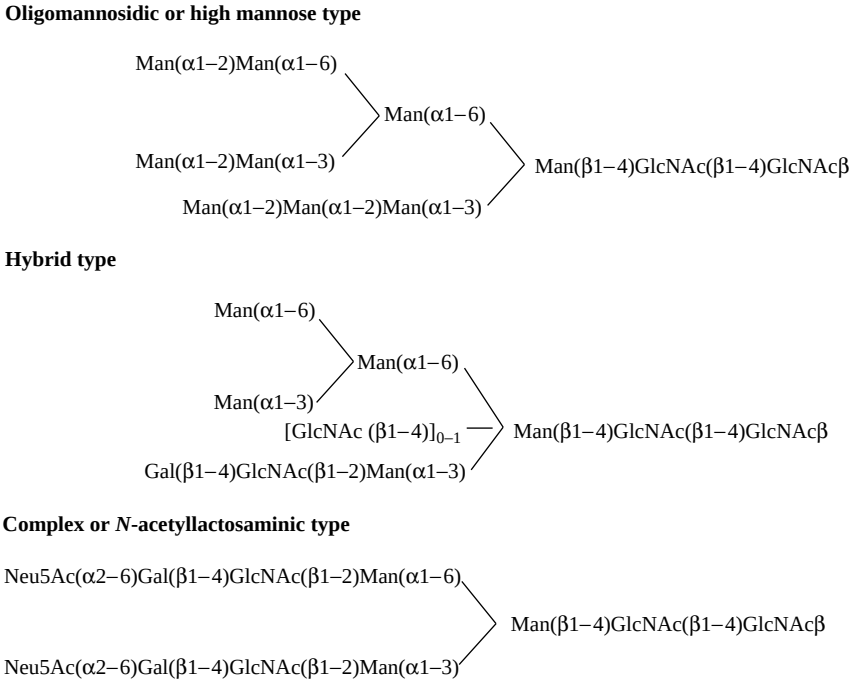

Figure 3 Representative structures of the three types of *N*-glycosidically linked glycans.

on the glycosyltransferase activities of the cell and of the species.[24] The cell specificity, tissue specificity and species specificity of glycosylation are now well established and have important implications.[5,21] For instance, the glycosylation pattern of recombinant glycoproteins is highly dependent on the glycosyltransferase activities of the cell. As only few bacteria possess glycosylation capacity, yeast, insect cells, plants, and animal cell lines have often been used as systems for expressing recombinant glycoproteins.[25] Examples of typical glycan structures biosynthesized by various cells are given in Figure 4.[26–30] The high variability of the glycan structures found in a given recombinant glycoprotein expressed in various cells may provide an explanation for some divergences in the physicochemical, biochemical, immunological and functional properties of the resultant recombinant glycoforms.

A specific set of glycans is attached to a unique glycosylation site. For instance, in human serum transferrin, the two glycosylation sites are occupied by either bi-, tri-, tri′, or tetraantennary glycans leading to different glycoforms.[24] These glycoforms have been separated using lectin-affinity chromatography, anion-exchange chromatography, isoelectric focusing or ESI/MS. In normal physiological conditions, the percentages of these glycoforms are reproducible. However, high variations in these percentages appear during development, differentiation, and under modifications of the normal physiological state. The nature of the different factors involved in the alterations of glycosylation is not yet known; however, the modulatory effects of hormones or cytokines during pregnancy[31] or inflammation[32] have been described.

2.3.4 Biological Roles of N-linked Glycans

The *N*-glycans are recognized as components having many functional roles[33–36] and the new science of glycobiology has been developed. It is difficult to predict the function of a given glycan, because the same glycan may be associated with various functions in different glycoproteins. Of the numerous functions attributed to *N*-glycans, some are important to the glycoprotein itself, mainly for its folding, solubility, protease sensitivity, antigenicity, intracellular location, and biological half-life. In general, the peripheral sequence of a glycan represents the most important functional part of this glycan. This peripheral part may be involved in specific carbohydrate–carbohydrate interactions or carbohydrate–protein interactions which generate most of the biological processes in which the glycoproteins are involved. For instance, the mechanism of recognition is responsible for the following: binding of asialoglycoproteins to the hepatic cell receptor which specifically recognizes terminal Gal and GalNAc residues and which is involved in the clearance of the asialoglycoproteins;[37] the specific binding of pathogens to the cells of the immune system;[38] or the specific targeting of enzymes possessing a glycan with Man-6P to the lysosomes.[39] An aberrant glycosylation of glycoproteins may alter

N-linked glycan of the S-layer glycoprotein of *Halobacterium halobium*[26]

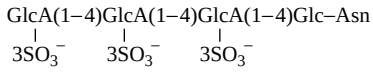

N-linked glycan found in honeybee venom phospholipase A_2[27]

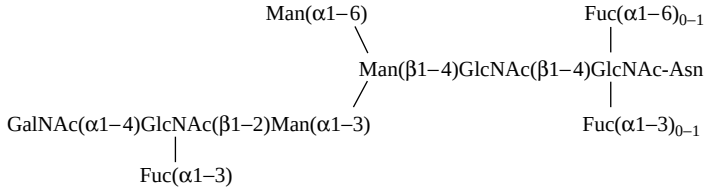

N-linked glycan from plant glycoprotein[28,29]

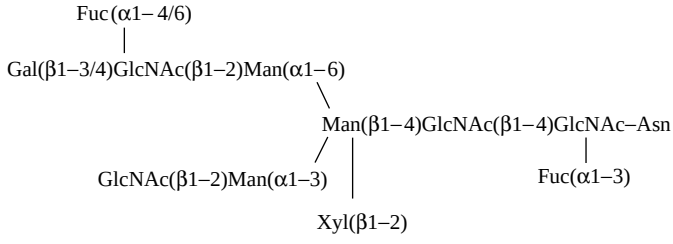

N-linked glycan from Yeast[30]

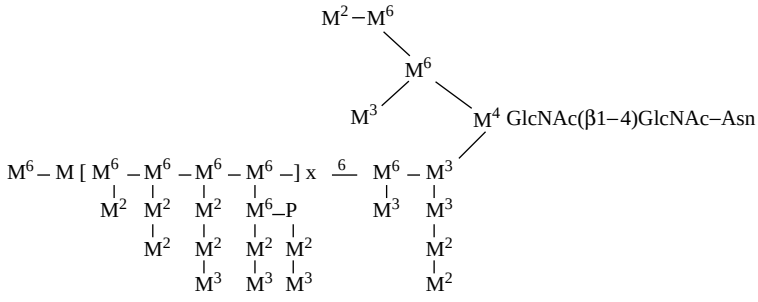

Figure 4 Examples of *N*-glycosidically linked glycans.

the intra- and intermolecular functions of the glycans. The science of glycopathology is now developing and some correlations between the modifications of glycosylation patterns and pathology have been established.[40] A better understanding of the fine variations in glycan structures and in the relative proportions of the glycoforms, at different stages of a disease, appear relevant to diagnosis.

Progress in genetic engineering has allowed the biosynthesis of a multitude of recombinant glycoproteins used for pharmaceutical applications. The most important problem encountered concerns the biosynthesis of a panel of glycoforms, which should be similar to the panel biosynthesized in vivo. Variations in glycosylation of recombinant glycoproteins may induce alteration in intercellular targeting and clearance[41] and may therefore inhibit their therapeutical activity.

2.4 O-linked Glycoproteins

2.4.1 General Structure of O-linked Glycans

In glycoproteins such as yeast glycoproteins,[30] secretory immunoglobulin A,[42] and glycophorin,[43] a limited number of O-linked glycans is found associated with N-linked glycans, whereas in mucins, several hundred to several thousand of these glycans are clustered in highly glycosylated domains.[44] The structure of the *O*-glycans is also very heterogeneous. The diversity depends on the nature of the core and the chain length, which may comprise 1–20. All these glycans are based on the same structural model, made of three distinct regions: the core, the backbone, and the periphery. To date, eight cores including the GalNAc attached to the hydroxy amino acid and Gal, GlcNAc and GalNAc residues have been characterized (Figure 5).

Cores

1 Gal(β1–3)GalNAcα

2 GlcNAc(β1–6) / GalNAcα / Gal(β1–3)

3 GlcNAc(β1–3)GalNAcα

4 GlcNAc(β1–6) / GalNAcα / GlcNAc(β1–3)

5 GalNAc(α1-3)GalNAcα

6 GlcNAc(β1–6)GalNAcα

7 Gal(β1–6) / GalNAcα / Gal(β1–3)

8 GalNAc(α1–6)GalNAcα

Backbones

Gal(β1–3)GlcNAcβ (Type I)

Gal(β1–4)GlcNAcβ (Type II)

Gal(β1–3)GalNAcα (Type III)

Gal(β1–3)GalNAcβ (Type IV)

Gal(β1–3)Galβ (Type V)

Gal(β1–4)Glcβ (Type VI)

Figure 5 Cores and backbones found in *O*-glycosidically linked glycans.

Structure of *O*-mannose-linked glycans of α-Dystroglycan[45]

Neu5Ac(α2–3)Gal(β1–4)GlcNAc(β1–2)Man/Ser or Thr

Structure of an *O*-glycan present in the jelly coat of the amphibian *Xenopus laevis*[46]

Fuc(α1–3)Fuc(α1–4)Kdn(α2–6) / GalNAcol

GalNAc(β1–3)Gal(β1–4) GlcNAc(β1–3)

Fuc(α1–2)

Structure of *O*-glycans found in (a) resting and (b) activated lymphocytes[47]

Neu5Ac(α2–6)

(a) Neu5Ac(α2–3)Gal(β1–3)GalNAcol

Neu5Ac(α2–3)Gal(β1–3)GlcNAc(β1–6)

(b) Neu5Ac(α2–3)Gal(β1–3)GalNAcol

Structure of an *O*-glycan found in respiratory mucin from a patient with cystic fibrosis[48]

HO_3S

Neu5Ac(α2–3)Gal(β1–4)GlcNAc(β1–6)

Fuc(α1–3) / GalNAcol

Neu5Ac(α2–3)Gal(β1–3)

Figure 6 Examples of *O*-glycosidically linked glycans.

Linked to the cores are six types of backbone, the most common being the *iso*-*N*-acetyllactosamine unit (Gal(β1–3)GlcNAc; type I) and the *N*-acetyllactosamine unit (Gal(β1–4)GlcNAc; type II). The periphery is made of different monosaccharides (Gal, GalNAc, Fuc, NeuAc) which may be assembled in repeated units. Examples of typical and peculiar O-linked glycans found in various glycoproteins are given in Figure 6.[45–48]

2.4.2 Biosynthesis of O-linked Glycans

In contrast to the biosynthesis of N-linked glycans, the biosynthesis of *O*-glycans is not performed by a transfer from a dolichyl–oligosaccharide precursor to the protein, but by a sequential addition of a monosaccharide from a nucleotide monosaccharide. In mucins, the UDP-GalNAc polypeptide α-*N*-acetylgalactosaminyltransferase is the enzyme which conjugates the first GalNAc residue to the completely folded apoprotein. Several types of this enzyme have been characterized. The eight cores result from the addition of Gal, GlcNAc, and GalNAc residues by α- or β-glycosyltransferases, whereas the backbones are synthesized by the action of two types of β-galactosyltransferase and two types of β-*N*-acetyl-glucosaminyltransferase. Addition of the monosaccharides found in the peripheral part is performed essentially by α-glycosyltransferases. The numerous glycosyltransferases involved in the biosynthesis of *O*-glycans are specific and almost localized in the Golgi apparatus.[49,50]

2.4.3 Biological Roles of O-linked Glycans

Like the N-linked glycans, the O-linked glycans are involved in intra- and intermolecular interactions and these interactions are increased by the fact that the *O*-glycans are often found in glycoproteins and mucins as clusters.[44,51] The presence of a cluster of sialylated and sometimes sulfated *O*-glycans confers to the polypeptide chain an extended conformation, a protection against the protease activity and heat denaturation. *O*-Glycans present in membrane glycoproteins have been recognized as receptors for viruses, bacteria, and pathogens. Antigenic determinants responsible for A, B, H, Le^x, Le^y, and

Type 1	
Fuc(α1–2)Gal(β1–3)GlcNAc	Blood group H (O)
GalNAc(α1–3)[Fuc(α1– 2)]Gal(β1–3)GlcNAc	Blood group A
Gal(α1–3)[Fuc(α1–2)]Gal(β1–3)GlcNAc	Blood group B
Gal(β1–3)[Fuc(α1–4)]GlcNAc	Blood group Lewis^a
Neu5Ac(α2–3)Gal(β1–3)[Fuc(α1–4)]GlcNAc	Sialyl Lewis^a
Fuc(α1–2)Gal(β1–3)[Fuc(α1–4)]GlcNAc	Blood group Lewis^b
Type 2	
Fuc(α1–2)Gal(β1–4)GlcNAc	Blood group H (O)
GalNAc(α1–3)[Fuc(α1–2)]Gal(β1–4)GlcNAc	Blood group A
Gal(α1–3)[Fuc(α1–2)]Gal(β1–4)GlcNAc	Blood group B
Gal(β1–4)[Fuc(α1–3)]GlcNAc	Lewis^x
Neu5Ac(α2–3)Gal(β1–4)[Fuc(α1–3)]GlcNAc	Sialyl Lewis^x
Fuc(α1–2)Gal(β1–4)[Fuc(α1–3)]GlcNAc(β1–3)Gal	Lewis^y
Poly-*N*-acetyllactosaminoglycan	
$[Gal(\beta1–4)GlcNAc(\beta1–3)]_n$	Blood group i
Gal(β1–4)GlcNAc(β1–6) Gal(β1–4)GlcNAc Gal(β1–4)GlcNAc(β1–3)	Blood group I

Figure 7 Structure of blood group determinants.

sialyl Le^x blood group activities are largely distributed among the glycoproteins containing O-linked glycans[52] and are located at the periphery of the structures (Figure 7). *O*-Glycosylation plays an important role in the mechanism of cell–cell recognition. For instance, in the mechanism of leukocyte adhesion to activated endothelial cells, specific structures of *O*-glycans are recognized by animal lectins called selectins.[53] During fertilization, the spermatozoon binds to *O*-glycans located in the mucins of the zona pellucida of the ovocyte. The specificity of these *O*-glycan structures should offer an explanation for species specificity in fertilization.[54] Alterations of *O*-glycosylation result in various pathologies, in cancer development, and in metastasis,[55] and the *O*-glycan structures are generally decreased.

2.5 Structure and Biosynthesis of Glycosylphosphatidylinositol Anchors

These anchors are complex structures sharing a common structure in which the inositol from a membrane glycerophosphatidylinositol is linked to a glucosamine residue from the linear tetrasaccharide Man(α1–2) Man(α1–6)Man(α1–4)GlcNH_2(α1–6). The diversity of GPI anchor structure depends on substitution of the core. Although little is known about the enzymes involved in the biosynthesis of a GPI anchor, the different steps in the biosynthesis of the trypanosomal GPI anchors have been well established.[56] The methods used to isolate and identify the complete structure of a GPI anchor are not described here.

3 GLYCOPROTEIN ISOLATION AND CHARACTERIZATION

Many methods are available for glycoprotein isolation. The conventional chromatographic methods of glycoprotein fractionation are based on molecular mass (gel permeation chromatography), surface charge interactions (ion-exchange chromatography) and surface hydrophobic interactions (hydrophobic chromatography). More selective glycoprotein isolation methods are based on interactions of the protein moiety with specific monoclonal or polyclonal antibodies (immunoaffinity chromatography) and the interactions of glycan moiety with lectins (lectin-affinity chromatography).

The use of these different column chromatography techniques (alone or in combination, performed at low or high pressure) can isolate pure glycoproteins (in a single step or in several steps). The purified glycoproteins are often composed of a mixture of glycoforms which differ by their degree of sialylation and therefore by their isoelectric point. To isolate the glycoforms, additional methods are often necessary based on electrophoresis, such as capillary electrophoresis, sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE), and isoelectric focusing.

This review concentrates first on immunoaffinity chromatography, which can be used (if specific antibodies are available) before other methods to extract a pure glycoprotein from a sample. Secondly, anion-exchange chromatography and lectin-affinity chromatographic methods, which are able to separate the glycoforms, are considered.

3.1 Mucin Isolation

3.1.1 Isolation of Mucins Requires Specific Methods

Due to their high molecular weight and their extreme diversity both in peptidic backbone and carbohydrate content, the isolation of mucins requires specific isolation methods.[57–59] Mucin can be collected from mucosal cells by scraping of the tissue surface using guanidine hydrochloride/phosphate-buffered saline

(PBS) at pH 7.0. Generally mucins are contaminated with proteoglycans, nucleic acids, and other glycoproteins. The first steps of purification consist of molecular-weight-based fractionation using repeated ultracentrifugation on a cesium chloride gradient in the presence of concentrated guanidine solutions. The samples are centrifuged at 100–150 000 *g* for 24–48 h in order to generate a density gradient from 1.3 to 1.6 g mL^{-1}. Under these conditions, proteoglycans, nucleic acids, and mucins are well separated.

Further purifications steps of individual mucins can be achieved according to molecular size using gel filtration on Sepharose™ CL2B or Sephacryl S1000. As a result of their charge diversity in relation to their sialic acid and/or sulfate content, individual mucin glycoforms can be separated by anion exchange chromatography on an anion exchanger such as diethylaminoethyl (DEAE). All of these chromatographic procedures should be carried out in the presence of strong denaturating agents such as guanidine hydrochloride buffer, or by introduction of reducing agents such as dithiothreitol (1% w/v).

If subunits are required, reduction and alkylation of mucin samples are necessary. Finally, strategies based on immunoaffinity procedures, using specific antibodies directed against the tandem repeat of the protein backbone such as the Pro-Asp-Thr-Arg-Ala-Pro-Gly-Ser (PDTRAPGS) region, have been recently developed and may be used in the isolation of abnormal mucins expressed in pathological states such as cancer.

3.2 Immunoaffinity Chromatography

A one-step purification of glycoprotein may be performed using specific protein monoclonal or polyclonal antibodies immobilized on different supports. Numerous methods have been used for preparing immobilized antibodies. Protein immobilization to CNBr-activated gel (agarose, Sepharose™, Ultrogel®) is the most popular choice.[60,61] A major drawback found in protein affinity gel preparation by activation of Sepharose™ with CNBr concerns the introduction of charged groups into the matrix. This is avoided when the proteins are immobilized on polyacrylic-hydrazido-Sepharose™ support.[62] Immobilized proteins can also be prepared with activated agarose-containing *N*-succinimide esters that react with amino groups of the proteins.[63]

In addition to these general techniques used for all kinds of proteins and many lectins, oriented immobilization methods have been described for antibodies. The different techniques of oriented immobilization of antibodies have been reviewed, such as the use of immobilized protein A or G, the use of immobilized antibodies against the Fc constant region of the antibodies to be immobilized, and the use of immobilized antibodies through the carbohydrate moiety after acid periodic oxidation.[64]

The selection methods for the most appropriate monoclonal antibody to be immobilized and precise methods for binding, washing and the elution steps have been reviewed.[65] In general, once the glycoprotein has been captured by the specific immobilized antibodies, elution of the glycoprotein is performed in acidic conditions and, to preserve the biological activity of the purified antibodies, eluted antibodies must be immediately neutralized.

Immunoaffinity chromatography is also used for detection of analytes as they elute from other types of chromatographic columns. Advances in analytical applications of immunoaffinity chromatography have been recently reviewed.[66]

3.3 Isolation of Glycoforms by Anion-exchange Chromatography

The glycoforms often differ by their charge and can be separated preparatively by using anion-exchange chromatography after careful optimization of the analytical conditions. For instance, to separate the 14 glycoforms of recombinant human erythropoietin, different methods of fractionation were used with the best results being obtained by DEAE-Sephacel chromatography.[67] Anion-exchange HPLC has been used to fractionate mucins and to identify glycoforms.[68] A very rapid separation of monosialylated and disialylated serum transferrin glycoforms was obtained by ion-exchange chromatography on a Mono Q HR 10/10 column using fast protein liquid chromatography (FPLC).[69]

3.4 Isolation of Glycoforms by Lectin Affinity Chromatography

3.4.1 Problems of Lectin Specificity

Lectins are sugar-binding proteins or glycoproteins of nonimmune origin that can agglutinate cells and/or precipitate glycoconjugates.[70]

They possess at least two sugar-binding sites, the presence of which explains their ability to precipitate polysaccharides, glycoproteins, or glycolipids, and why they agglutinate cells.

Lectins have been now isolated from viruses, bacteria, plants, and from a variety of animal tissues and species. However, only plant lectins are now commonly used in research and in clinical laboratories for the detection and fractionation of soluble and membrane glycoproteins, as well as glycopeptides or glycans derived from these glycoproteins.[71–73]

Lectins are often still classified according to the monosaccharide that inhibits the interaction between a lectin and a cell, or which allows the specific elution of a

Table 2 Lectins commonly used for glycoprotein study, classified according to their monosaccharide specificity

Lectin	Abbreviation
α-Man, α-Glc	
Canavalia ensiformis	Con A
Lens culinaris	LCA
β-Gal, β-GalNAc	
Ricinus communis	RCA I
	RCA II
Glycine max (soybean)	SBA
Arachis hypogaea (peanut)	PNA
α-Gal, α-GalNAc	
Griffonia simplicifolia I	GSA I
Dolichus biflorus	DBA
β-GlcNAc	
Triticum vulgare (wheat germ)	WGA
α-Fuc	
Aleuria aurantia	AAA
Lotus tetragonolobus	LTA
Ulex europeus I	UEA I
α-Neu5Ac	
Limulus polyphemus (limulin)	

Con A, concanavalin A; DBA, *Dolichus biflorus* agglutinin; GSA I, *Griffonia simplicifolia* agglutinin I; LCA, *Lens culinaris* agglutinin; LTA, *Lotus tetragonolobus* agglutinin; PNA, *Arachis hypogaea* agglutinin; RCA I, *Ricinus communis* agglutinin I; RCA II, *Ricinus communis* agglutinin II; SBA, soybean agglutinin; UEA I, *Ulex europeus* agglutinin I; WGA, wheat germ agglutinin.

bound glycoconjugate from an immobilized lectin column (Table 2).

Analysis of sequence data of plant lectins by protein sequencing and cDNA cloning, together with X-ray crystallography data and molecular modeling, result in classification by their three-dimensional structures, not taking into account their monosaccharide-binding specificities.[74–76] In this way seven groups of plant lectins can be distinguished:

- The legume lectins with a 12-stranded β-sandwich structure.
- The monocot-binding lectins with a 12-stranded β-barrel structure.
- The Jacalin-like lectins with a β-prism fold.
- The Amaranth-like lectins.
- The chitin-binding lectins made of hevein-like domains.
- The type II ribosome-inactivating proteins (RIPs) which contain an enzymatically active A chain and a B chain with lectin activity.
- The cucurbitaceae phloem lectins, which are a small family of chitin-binding lectins, found in the phloem exudate of cucurbitaceae species.

The structural basis of selective monosaccharide recognition by plant lectins has been investigated by X-ray crystallography and is now well documented.[77,78] It explains in terms of hydrogen bonds the low affinity (dissociation constant, K_d, in the 0.1–10 mM range) found in the interaction between a monosaccharide and a given lectin. However, it is well known that, in most cases, complex oligosaccharides are several-thousand-fold more potent inhibitors of lectins than monosaccharides.[79,80]

These high binding selectivities towards complex oligosaccharides, first determined by biochemists, are now also explained by X-ray crystallography studies in terms of extended binding sites on a lectin. In addition to some amino acids forming the monosaccharide-binding site, other amino acid residues are involved in additional direct and water-mediated contacts between oligosaccharides and the lectin surface.[78,81,82]

It is very important to know the complex specificity before using a lectin – as a probe in the exploration of cell surface glycoconjugates or as a tool to fractionate various glycoconjugates by affinity chromatography.

3.4.2 Determination of the Complex Specificity of Lectins

The specificity of lectins towards monosaccharides or oligosaccharides of glycoproteins or glycolipids can be determined by inhibition experiments in which monosaccharides or oligosaccharides of known structures are tested for their ability to inhibit either hemagglutination[80] or precipitation of polysaccharides.[83] However, a major drawback of these methods is that they require large quantities of oligosaccharides that are often available only in limited amounts.

A second approach to determine the complex specificity of a lectin is to study the affinity of the immobilized lectin towards complex oligosaccharides.[84,85] In this approach, lectins are first immobilized on insoluble matrices. Small columns containing 2–10 mL of the immobilized lectins are then prepared and equilibrated in an appropriate buffer; the behavior of radiolabeled oligosaccharide or glycopeptides of known structure applied to these columns is then examined.

This second approach is of particular interest because it calls for only small amounts of oligosaccharides and, once the precise specificity of an immobilized lectin is established, it becomes possible to predict the primary structure of glycopeptides or oligosaccharides displaying similar elution profiles.

Specificities of lectins can be also studied using more sophisticated methods including isothermal titration microcalorimetry (ITC)[86] or real-time interaction by surface plasmon resonance with biosensors.[87]

3.4.3 *Specificity of Lectins Towards Complex Oligosaccharides*

From all these studies, important conclusions can be drawn about the carbohydrate-binding specificity of plant lectins. Lectins can interact with internal oligosaccharidic sequences of a glycan. Very few lectins recognize terminal nonreducing monosaccharides on a glycan and most of them recognize internal sequences of oligosaccharides. Lectins identical in terms of monosaccharide specificity can display different complex specificity. Both concanavalin A (Con A) and *Lens culinaris* agglutinin (LCA) display the same monosaccharide specificity for α-Man or α-Glc. However, as shown in Table 3, they display a higher affinity for different *N*-glycosylpeptides and related oligosaccharides. In particular, the presence of a fucose residue at the C-6 position of the GlcNAc residue involved in the *N*-glycosylamine linkage in a biantennary *N*-acetyllactosamine-type glycopeptide enhances the affinity of LCA for this glycan and is an important determinant for the glycan-binding specificity of the lectin.[80,88]

Moreover, a given lectin can recognize very different oligosaccharide sequences. For example, Con A has a strong affinity for oligomannoside type as well as for some hybrid-type glycans and a weaker one for biantennary *N*-acetyllactosamine-type glycans (Table 3).

These complex glycan specificities of plant lectins result from the existence of extended carbohydrate-binding sites on the lectins.

It is also worth noting that different lectins can recognize different saccharide sequences on the same glycan. As these sequences are common to numerous glycoproteins, results obtained with immobilized lectins during a fractionation of glycoproteins or by Western blot analysis of glycoproteins with lectins must be interpreted with caution.

The spatial conformation of glycans may affect their interaction with many lectins. Some lectins, such as LCA or wheat germ agglutinin (WGA) show a stronger affinity for *N*-glycosylpeptides containing the recognized oligosaccharide sequences than for the glycans released from these glycopeptides either by chemical or enzymatical cleavages and still containing the oligosaccharide determinants. This can be explained by the fact that the attachment of a glycan to asparagine imposes on the trisaccharidic core sequence Man(β1–4)GlcNAc(β1–4)GlcNAc(β1-N)Asn a rigid structure that favors the specific recognition of the oligosaccharidic sequences on the glycan by a given lectin.[89] Some lectins such as *Ricinus communis* agglutinin I (RCA I) present a higher affinity for oligosaccharide sequences substituted by sialic acid residues bound in the α-2,6 position of galactose rather than in the α-2,3 position. This can be related to the higher mobility of sialic acid residues around the α-2,6 linkage.[80,90]

However, for *Phaseolus vulgaris* phytohemagglutinins (PHAs) with erythro-agglutinating or leuco-agglutinating activities (E4-PHA and L4-PHA, respectively), this higher mobility of the sialic acid residues around the α-2,6 linkage can mask the internal oligosaccharidic determinants. This can explain why sialylated glycans in the α-2,6 position of galactose are not recognized by these lectins, whereas sialylated glycans in the α-2,3 position of galactose interact with them.[91]

The concept of a larger space mobility of oligosaccharide sequences around an α-1,6-glycosidic linkage compared to the higher rigidity imposed by the α-1,3 linkage can explain the higher affinity of some lectins such as LCA or L4-PHA for tri′-antennary *N*-acetyllactosamine-type glycans (Table 3) in which one of the α-Man residues is substituted by the third antennae at the C-2 and C-6 positions, rather than for triantennary glycans containing another α-Man substituted at C-2 and C-4 positions.[88,92]

Due to their extended carbohydrate-binding sites, most lectins will display very precise specificity towards well-defined oligosaccharide sequences belonging to either *N*- or *O*-glycosylproteins. Therefore, classifications of lectins according to monosaccharide specificity must be superseded by classifications taking into account the concept of dominant oligosaccharides as recognized by lectins at either internal or peripheral positions on a glycan. However, a few lectins will present a specificity towards terminal nonreducing dominant monosaccharides.[93–105]

Table 3 lists the specificity of some commonly used lectins towards oligosaccharide sequences belonging to *N*-glycosylproteins.

3.4.4 *Use of Immobilized Lectins for Fractionation of Glycoproteins*

3.4.4.1 Immobilization of Lectins Numerous methods have been proposed for preparing immobilized lectins and many of them are now available from different suppliers (Pharmacia Biotech, E. Y. Laboratories, Sigma Chemical Co, Vector Laboratories). Lectins immobilized to CNBr-activated agarose are the most popular and can be easily prepared at a density of 2–10 mg of lectin per milliliter of settled gel (agarose, Sepharose™ or Ultrogel®) activated with CNBr according to the procedures of March et al.[60] or Kohn and Wilchek.[61]

Immobilization of lectins on polyacrylic-hydrazido-Sepharose™ support according to Wilchek and Miron[62] does not introduce charged groups into the matrix, a major drawback found in lectin affinity gels prepared by activation of Sepharose™ with CNBr.

Table 3 Specificity of commonly used lectins towards oligosaccharide sequences belonging to *N*-glycosylproteins

Con A

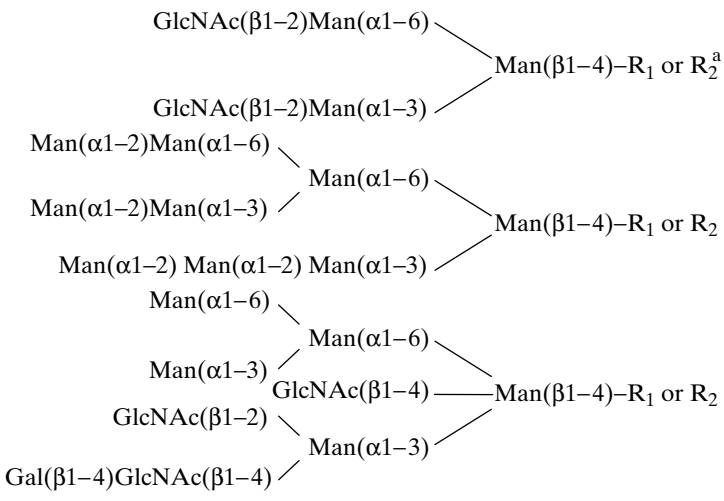

Lens culinaris

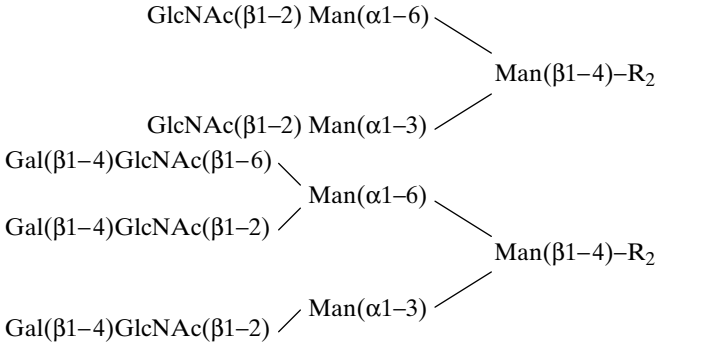

RCA I[b]

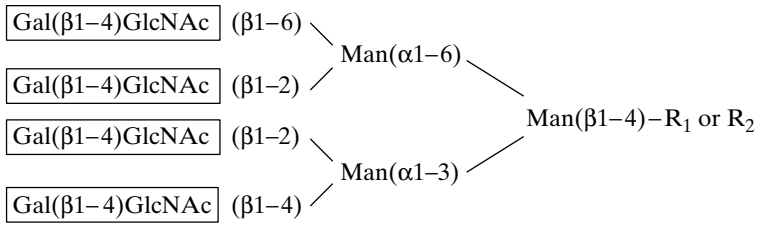

Leukoagglutinating lectin from *Phaseolus vulgaris* (L4–PHA)

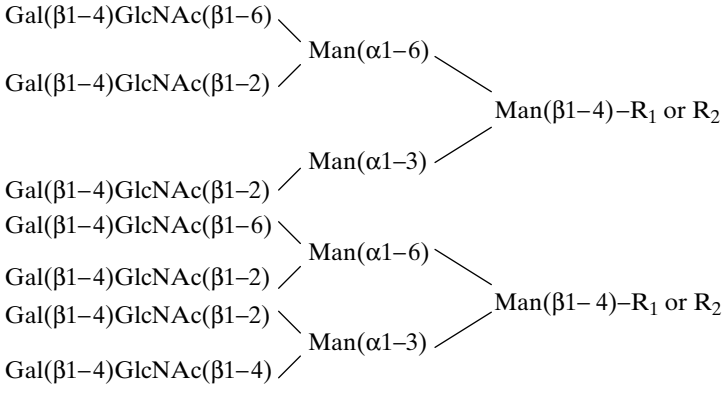

Table 3 (*continued*)

WGA[b]

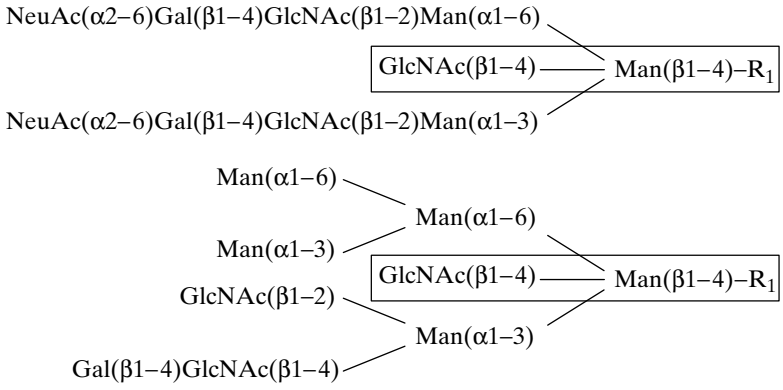

Erythroagglutinating lectin from *Phaseolus vulgaris* (E4 – PHA)

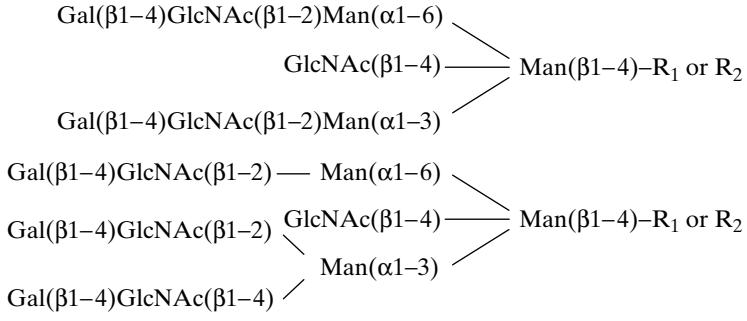

DSA[c]

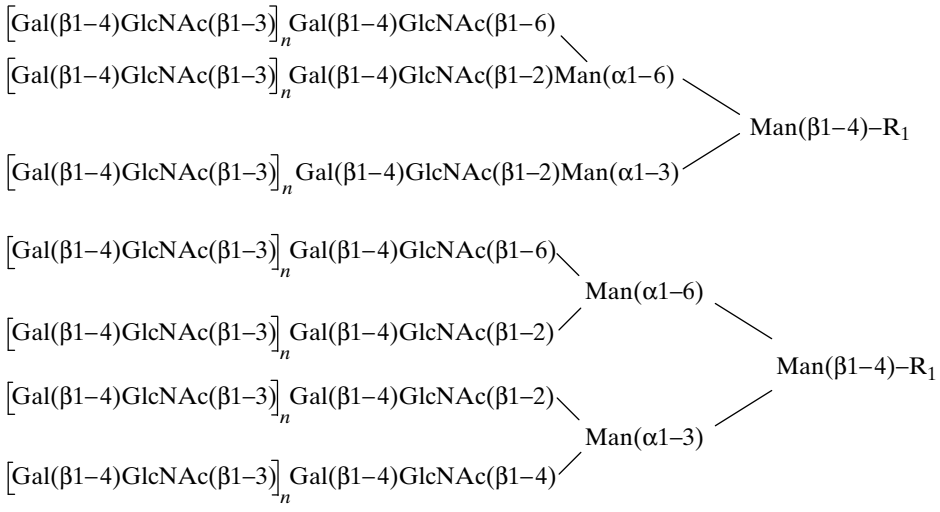

[a] R_1 = GlcNAc(β1–4)GlcNAc(β1–N)Asn; R_2 = GlcNAc(β1–4)[Fuc(α1–6)]GlcNAc(β1–N)Asn.
[b] Sequences in boxes are the minimal oligosaccharide structure necessary for lectin recognition.
[c] DSA, *Datura stramonium* agglutinin.

Immobilized lectins can be prepared with activated agarose containing *N*-hydroxysuccinimide esters which react with amino groups of lectins.[63] These activated gels are available from BioRad as Affi-Gel® 10 or Affi-Gel® 15.

All these gels can be used in low-pressure chromatography with flow rates of 5–20 mL h^{-1}. However, lectins can be also immobilized on activated supports compatible with HPLC (flow rate of 1 mL min^{-1}). This technique combines the high speed and resolution

characteristics of HPLC with the selectivity of biospecific interactions.[106,107]

3.4.4.2 Fractionation of Glycoproteins on Immobilized Lectins Affinity chromatography on immobilized lectins is now widely used to fractionate either soluble or membrane glycoproteins, and many examples have been reviewed elsewhere.[97,108–113] Highlighted here are some particular points which can be useful during such fractionations.

Variations in the amount, as well as in the quality of the lectin immobilized on a gel, represent important factors influencing the binding of glycoproteins. For example, affinity differences are often observed between different commercially available immobilized Con A. This requires the calibration of the lectin column with well-known glycoproteins before using a new batch of immobilized lectin.

Three fractions are generally obtained, reflecting relative affinities of the glycoproteins for the immobilized lectins.

The nonreactive glycoproteins are eluted at the void volume of the column in the equilibration buffer. This fraction, when the exact capacity of the immobilized lectin is not known, must be submitted to a new cycle of absorption and elution on the same column, to be certain that the immobilized lectin was not saturated during the first run. In order to limit nonspecific interactions between glycoproteins and immobilized lectins, equilibration buffers must possess a moderate ionic strength (0.1–1.0 M in NaCl, for example).

The weakly reactive glycoproteins give fraction(s) that are recovered by elution with the equilibration buffer as retarded fraction(s). In this case, the separation of weakly interacting glycoproteins can be obtained by using a long and thin lectin column; this is more efficient than a wider and shorter column containing the same amount of immobilized lectin. The spatial conformations of the native glycoproteins or nonspecific hydrophobic interactions may modulate the accessibility of some glycans to the lectin and can explain the presence of some artefactual weakly reactive fractions. These fractions are absent when the glycoproteins are reduced and alkylated before fractionation on the immobilized lectin.[114]

Strongly reactive glycoproteins are specifically eluted after extensive washing with the equilibration buffer containing the competitive monosaccharide (0.1–0.5 M). After extensive washing with the equilibration buffer, the immobilized lectin can be used again with the same efficiency. Stronger lectin–glycoprotein interactions may result either in very high affinity of the lectin for some oligosaccharidic determinants of a glycoprotein or in the multivalent interactions between the immobilized lectin and a very high density of a saccharidic determinant of the glycoprotein recognized by the lectin (avidity effect). In the particular case of immobilized Con A, the recovery of high-affinity glycoproteins can be improved by raising the temperature of the 0.5 M methyl-α-D-glucoside or α-D-mannoside solution to 37 °C or even 60 °C. However, very often these strong interactions could be reversed only with nonspecific desorption processes which may cause irreversible denaturation of the immobilized lectin. Such nonspecific elutions can be performed by pH changes with either 20 mM acetic acid or diaminopropane solutions or with 0.02–0.1 M borate buffers (pH 8.0), without denaturation of the lectin.

High-yield recovery can also be obtained by heating the immobilized lectin–glycoprotein complex for 3 min at 100 °C in buffer containing 5% (w/v) sodium dodecyl sulfate (SDS) and 8 M urea, but with irreversible inactivation of the lectin.

The sequential use of immobilized lectins, with different and well-defined specificities towards saccharidic determinants, can be applied to fractionate complex mixtures of glycoproteins into classes depending on their affinities for the different lectins. Immobilized Con A, LCA, and WGA are the most commonly used.

When lectin affinity chromatography is used to fractionate membrane glycoproteins, the most important problem is the stability of the immobilized lectin in the detergent solution utilized for membrane protein extraction. This detergent must be present during all further purification steps of the solubilized membrane glycoproteins. However, this detergent may dissociate the lectin into subunits or change their active conformation. As shown by Lotan et al.,[115] nonionic detergents such as Triton® X-100 at 0.1% concentration in the running buffer remain the most compatible with lectin affinity chromatography of membrane glycoproteins.

3.4.4.3 Limitations of Lectin Affinity Chromatography in Purification of Glycoproteins Affinity chromatography on immobilized lectins rarely permits the purification of a given soluble or membrane glycoprotein from complex mixtures in a single step. This method usually fractionates glycoproteins into different classes containing similar glycans. As different lectins will be able to recognize different saccharidic sequences, but belonging to the same glycan, sequential affinity chromatography on different immobilized lectins will not improve the fractionation significantly.

However, a few membrane or soluble glycoproteins have been isolated in one step from complex mixtures by lectin affinity chromatography. For example, immobilized WGA allows the purification of either glycophorin, the major human erythrocyte membrane glycoprotein,[116] or the insulin receptor[117] in a single step. Laminin, a major

component of basement membrane, can be purified in one step on an immobilized *Griffonia simplicifolia* agglutinin I (GSA I-B4) lectin column.[118]

Human IgA1 and IgD can be isolated in one step from human sera on an immobilized *Artocarpus integrifolia* lectin (Jacalin) column which strongly interacts with the *O*-glycosidically linked Gal(β1–3)GalNAc determinants present in the hinge region of these glycoproteins.[119–122]

α-2 Macroglobulin from human sera is the only glycoprotein with oligomannosidic glycans recognized and eluted with a 0.1 M methyl-α-D-mannoside solution from an immobilized *Galanthus nivalis* mannose-specific lectin, isolated from the snowdrop bulb.[123] However, lectin affinity chromatography usually fractionates glycoproteins into different classes containing similar carbohydrate determinants. Nevertheless, when a given glycoprotein can be purified in a first step from other contaminating glycoproteins, either by conventional biochemical procedures or by immunoaffinity chromatography, lectin affinity chromatography represents a very powerful tool with which to unravel the glycan microheterogeneity inside this glycoprotein.[124]

3.5 Characterization of Glycoprotein Glycosylation

3.5.1 Glycan Detection Kits

Protein glycosylation can be easily and sensitively detected using different detection systems based on immunological labeling of glycoproteins on blots.[125–128] The procedures are based on oxidation of the adjacent hydroxyl groups of the monosaccharide constituents to aldehyde by mild periodate treatment and subsequent derivatization of the resultant aldehyde groups with either biotin or digoxigenin (DIG) via a hydrazide group. Biotin or DIG can be subsequently detected using either streptavidin or antidigoxigenin – alkaline phosphatase conjugate. An adaptation of the method consists in using biotin or DIG-labeled lectins, specific to the different constituent monosaccharides.

These methods are commonly used as a first step to assess both the occurrence and nature of glycans.

3.5.2 Detection of Glycoproteins with Lectins after Separation by Polyacrylamide Gel Electrophoresis

The high resolving power of polyacrylamide slab gel electrophoresis either under native or denaturing conditions[129] or two-dimensional gel electrophoresis combining isoelectric focusing and SDS–polyacrylamide gel electrophoresis[130] allows excellent separations of protein and glycoprotein mixtures. These glycoproteins can be detected with lectins of well-defined specificity either by direct incubation of the gels with lectins or by incubation of the electrophoretically separated glycoproteins transferred onto various immobilizing matrices with lectins.

3.5.2.1 Direct Application of Lectins to Polyacrylamide Gels

Use of Radioiodinated Lectins.[131] After electrophoresis, the slab gel slices are fixed in solution with or without 0.1% Coomassie brilliant blue and washed in a destaining solution containing 7.5% methanol and 7.5% acetic acid in water. After equilibration in Tris-buffered saline (TBS) (0.05 M Tris · HCl, pH 7.5; 0.15 M NaCl; 0.1% NaN_3), gel slices are put in a flat, humidified plastic box, overlaid with a radioiodinated lectin diluted in the equilibration buffer to $0.1–2.0 \times 10^7$ cpm mL^{-1} and incubated for 1–20 h.

After incubation, unbound lectin is washed out in several changes of equilibration buffer. Gel slices are then dried out and the lectin-reactive glycoproteins visualized by autoradiography.

In the overlay method, there is only binding of the radioiodinated lectin to glycoproteins on the overlaid face of the gel. Total immersion of the gel slices in the lectin solution will label both gel faces, but more lectin solution will be necessary.

For a given lectin, parallel gel slices must be incubated and washed in the presence or absence of 0.1 M competitive monosaccharide to detect any nonspecific binding.

On a gel slice, different glycoproteins can be revealed by a lectin because they possess the same oligosaccharide determinant recognized by the lectin. A single glycoprotein on a gel can be revealed by several lectins because of the possible glycan microheterogeneity of this glycoprotein, but also because different saccharide sequences belonging to a single glycan can be recognized by different lectins.

Use of Peroxidase-labeled Lectins. After electrophoresis, separated glycoproteins can be also reacted with peroxidase-coupled lectins and detected by color development of the enzyme reaction.[132] Lectins can be easily coupled to horseradish peroxidase and many peroxidase-coupled lectins are commercially available. However, some lectins, such as Con A, can be used without any coupling procedure with peroxidase because of their multivalency and of their great affinity for the N-linked glycans of horseradish peroxidase. In the first step, the lectin will bind to the separated glycoproteins on the gel and, in the second step, the glycoprotein-bound lectin will bind to horseradish peroxidase, which will be visualized by the chromogenic reaction with 3,3′-diaminobenzidine as the substrate.

3.5.2.2 Detection of Glycoproteins after Transfer Onto Immobilizing Matrices The transfer of electrophoretically separated proteins and glycoproteins onto immobilizing matrices was first described in 1979 by Renart et al.[133] and Towbin et al.[134] Since then, protein blotting has become the method of choice for detecting and characterizing proteins in complex mixtures.[135,136]

After electrophoresis, the separated proteins and glycoproteins are transferred onto different immobilizing matrices under conditions that maintain the electrophoretic pattern. Electrotransfer is generally preferred to capillary transfer because of the increased speed and the improved resolution of the bands. Many immobilizing matrices have been described but the most popular are the cellulose nitrate (CN) and the microporous polyvinylidene difluoride (PVDF) membranes on which transferred proteins are strongly bound by interfacial (hydrophobic) interactions.

Transfer conditions must be defined with care. The composition of the transfer buffers, with or without methanol which promotes hydrophobic binding by stripping SDS from proteins and with or without SDS which increases the elution of high-molecular-weight proteins from the gel as well as the running conditions, depends on the proteins to be transferred.

After the transfer of proteins to immobilizing matrices, all the unoccupied potential binding sites of the matrices must be blocked to avoid unspecific binding of the detecting molecules. Electrotransferred glycoproteins can be visualized by incubation of the matrices with radioiodinated lectins,[137] with biotin-conjugated lectins revealed with avidin-peroxidase, with peroxidase-coupled lectins or with DIG-labeled lectins[138] revealed with anti-DIG antibodies conjugated either to alkaline phosphatase, to horseradish peroxidase or labeled with fluorescence markers.

In particular, the detection of glycoproteins on blots with DIG-labeled lectins represents a very sensitive method based on the use of the plant steroid DIG as a hapten which will be detected by antibodies (anti-DIG) with high affinity towards this hapten. DIG-labeled lectins as well as DIG antibody (Fab fragments) conjugated with alkaline phosphatase are available from Boehringer Mannheim (Mannheim, Germany).

Together with blots with electrotransferred glycoproteins, dot blots, prepared by spotting on nitrocellulose membrane glycopeptide-containing fractions obtained by different chromatographic methods, can be revealed by the DIG–lectin/anti-DIG–alkaline phosphatase system.

It is also noteworthy that, as electrotransferred glycoproteins are strongly bound on nitrocellulose or PVDF membranes, treatment of the immobilized glycoproteins with various endo- or exoglycosidases of known specificities, in combination with DIG–lectins of well-defined specificities, can provide structural information on the glycan moieties using minute amount of glycoproteins. However, reference glycoproteins with well-characterized glycans must then be included in the analyzed membrane in order to verify the efficiency of either the enzymatic treatment(s) and/or the lectin specificity.

3.5.3 Crossed Affinoimmunoelectrophoresis of Glycoproteins

Crossed affinoimmunoelectrophoresis (CAIE) combines the interaction between a lectin and glycoproteins in the first-dimension electrophoretic step with crossed electrophoresis into an antibody-containing gel in the second dimension.[139] This sensitive technique can give important information about the interaction between a given lectin and the glycoprotein to be purified because a good correlation is generally observed between results obtained by CAIE and by affinity chromatography on the immobilized lectin.

Above all, the method rapidly reveals the glycan microheterogeneity of the glycoprotein to be purified. This method can also be used for the characterization of membrane glycoproteins solubilized in nonionic detergents because both lectin–membrane glycoprotein interactions and antigenicity of membrane glycoproteins are not affected by nonionic detergents such as Triton® X-100.[140,141]

Finally, determination of glycan microheterogeneity of serum glycoproteins by CAIE can be used for a rapid and discriminating diagnosis of various pathological conditions.[142]

4 ISOLATION AND CHARACTERIZATION OF GLYCOPEPTIDES AND GLYCANS

4.1 Proteolytic Digestion of Glycoprotein

In an individual glycoprotein each attachment site frequently accommodates different glycans generating a set of glycoforms. Examination of the glycosylation site heterogeneity requires the proteolysis of the glycoprotein to obtain glycopeptides of different size, which may be HPLC separated or purified and subsequently analyzed by MS. Each individual glycopeptide will contain a unique peptide sequence and a single oligosaccharide chain.

The choice of the protease will determine the number of peptides (or glycopeptides) produced and the length of the peptidic chain (directly related to the hydrophobicity of the glycopeptides) and for this reason will considerably influence the choice of the HPLC elution conditions.

4.1.1 Pronase Digestion

Pronase, which is a mixture of endopeptidases with broad specificity, can be used to obtain glycoasparagines or glycopeptides with a very short peptide chain. After purification, the structure of the glycan moieties can be established either directly or after their release by chemical or enzymatic methods.

4.1.2 Tryptic and Chymotryptic Digestion

In order to minimize the peptide heterogeneity of the glycopeptide, the enzymes of choice will be the highly specific proteases such as trypsin or chymotrypsin. Proteolytic hydrolysis may be performed either on native glycoproteins or after reduction and alkylation which generally produces a better substrate for protease digestion.

Trypsin, chymotrypsin or endoproteinase Glu C (*Staphylococcus aureus* V8 protease) are used when a longer polypeptide chain is required or when it appears necessary to access the different glycan structure to each of the glycosylation sites present on the polypeptide chain.

4.2 Methods for Glycan Release

As a detailed structure of individual glycans cannot be achieved directly on the native glycoprotein, for a complete structural study the glycans must be released from the glycoprotein either by chemical or enzymatic hydrolysis.

4.2.1 Chemical Release of Glycans

As previously mentioned, glycans can be covalently linked to the peptidic backbone through *O*- or *N*-glycosidic linkages. These two fundamental types of linkages present different susceptibility to alkaline hydrolysis. The O-linkage is easily cleaved in diluted sodium hydroxide solutions (0.05–0.1 M) at 37 °C for 48 h[143,144] in the so-called β-elimination reaction. Alkaline treatment is generally achieved in the presence of a reductive agent such as sodium borohydride ($NaBH_4$) in order to avoid further degradation by a peeling reaction of released oligosaccharides. For that reason, *O*-glycans are released from the glycoprotein as oligosaccharide-alditols.

The N-linkage is quite stable during the β-elimination procedure and *N*-glycans are released as glycopeptides. Cleavage of *N*-glycans requires a strong alkaline agent such as hydrazine.[145] The hydrazinolysis procedure is commonly used for releasing *N*-glycans.[146–149] This is then a three-step procedure including the hydrazine treatment, the *N*-re-acetylation of hexosamines and sialic acids, and the cleavage of resulting hydrazones. *N*-Glycans are released as free oligosaccharides, having the di-*N*-acetyl-chitobiose unit at their reducing end. *O*-Glycans can be partially released using mild reaction conditions.[149] It should be mentioned that both procedures lead to the destruction of the polypeptide backbone.

4.2.2 Enzymatic Release of Glycans

Release of glycans from *N*-glycosylproteins can be achieved by two kinds of enzymes, namely *endo-N*-acetyl-β-D-glucosaminidases and peptide-*N*-glycosidases.

4.2.2.1 Endo-N-*Acetyl-β-D-Glucosaminidases (EC 3.2.1.96; Mannosyl-glycoprotein 1,4-*N*-Acetamido-deoxy-β-D-glycohydrolases)* This family of enzymes hydrolyses the di-*N*-acetylchitobiose unit common to all the *N*-glycans[150] according to Equation (1):

$$\begin{aligned}&(R)\text{-GlcNAc}(\beta1-4)\text{GlcNAc}(\beta1-\text{N})\text{Asn-}(\text{X})_n\\&\quad\longrightarrow (\text{R})\text{-GlcNAc} + \text{GlcNAc-Asn-}(\text{X})_n\\&\quad\text{oligosaccharide} + \text{protein}\end{aligned} \qquad (1)$$

- Endo H, isolated from *Streptomyces plicatus*,[151] and endo F1, from *Flavobacterium meningosepticum*,[152] hydrolyze both high-mannose-type and hybrid-type glycans. These enzymes represent useful tools to discriminate between high-mannose and lactosaminic-type glycoproteins.
- Endo B, isolated from *Sporotrichum dimorphosporum*,[153] and endo F2, from *Flavobacterium meningosepticum*,[152] act both on high-mannose and complex glycans.

Generally glycoproteins are denatured prior to the enzymatic hydrolysis; nevertheless *endo*-β-*N*-acetyl-hexosaminidases are able to deglycosylate native glycoproteins and may be employed for the study of glycan functions inside an individual glycoprotein.

*4.2.2.2 Peptide-*N*-glycosidase (EC 3.5.1.52; Peptide-*N*^{4}-(N-acetyl-β-glucosaminyl) Asparagine Amidase)* Peptide-*N*-glycosidase, also called *N*-glycanase® is able to split the monosaccharide–peptide linkage removing the complete glycan moiety of the molecule (Equation 2):

$$N\text{-glycosyl-Asn-protein} \xrightarrow{\text{Glycanase}} \text{glycan} + \text{Asp-protein} + \text{NH}_3 \qquad (2)$$

The enzyme cleaves all types of glycans and the reaction proceeds unless the peptide part is restricted to the sole asparagine residue. The enzyme is isolated from *Flavobacterium meningosepticum*[152] (peptide-*N*-glycosidase F). The enzyme is now cloned and

expressed in *Escherichia coli*.[154] Another peptide-*N*-glycosidase (peptide-*N*-glycosidase A) has been isolated from almond emulsin.[155] This enzyme is able to cleave glycans bearing an α-1,3-fucose attached to the first *N*-acetylglucosamine linked to asparagine as found in plant *N*-glycosylproteins.

4.3 Fractionation of Glycopeptides and Glycans

4.3.1 Use of Immobilized Lectins for Fractionation of Glycopeptides and Oligosaccharides

Even though the purification and fractionation of glycoproteins by lectin affinity chromatography is encumbered with some limitations, this method represents a powerful tool for the fractionation of glycopeptides and glycans. Nonspecific interactions often observed between glycoproteins and immobilized lectins, or strong interactions resulting from the presence on a glycoprotein of a high density of a particular saccharide determinant, are absent or very rare in the case of glycans or glycopeptides. These compounds are thus fractionated on the basis of a true affinity chromatography. Consequently, this implies, even more than in glycoprotein fractionations, a very precise knowledge of the exact specificities of the immobilized lectins to be used. Under these circumstances, it often becomes possible to predict the primary structure of lectin-reactive glycopeptides or glycans.

Many fractionation schemes of glycopeptides or glycans, obtained from soluble or membrane-bound glycoproteins, have been proposed.[95–105,156–159]

4.3.1.1 General Comments Before using a new batch of immobilized lectin or after repeated use of a lectin column which may lead to a decrease in its binding capacity, calibration of the lectin column with well-defined glycopeptides or glycans should be performed. Standard glycopeptides or mixtures of glycopeptides to be fractionated can be *N*-[^{14}C]- or [^{3}H]-acetylated[160] to help follow the lectin column performance. Standard oligosaccharides or mixtures of glycans to be fractionated, released from glycoproteins or glycopeptides either by chemical or by enzymatic cleavages, can be labeled at the terminal reducing GlcNAc residue by reduction with [^{3}H]-sodium or potassium borohydride.[161]

Most lectin columns equilibrated in PBS containing 0.02% sodium azide can be run at room temperature (20 °C) but should be stored at 4 °C between consecutive runs. Most lectins are less efficient at 37 °C than at room temperature or even at 4 °C.[101,115] The degree of separation of weakly reactive glycopeptides or glycans may depend on the geometry of the lectin column. A better separation of these compounds is generally obtained with a long and thin column rather than with a wider column containing the same amount of immobilized lectins.

When the exact capacity of a lectin column is not known, the unretained fraction must be recycled on the regenerated column. For example, commercially available Con A–Sepharose™ 4B from Pharmacia Biotech is able to bind 50–75 μg per milliliter of gel of biantennary *N*-glycosylpeptides.

Most immobilized lectins have the same affinity for *N*-glycosylpeptides and for glycans released from them either by chemical or enzymatic cleavages. However, *N*-*N*′-diacetylchitobiose-Asn inner core is an important determinant for binding to immobilized LCA or WGA, and glycans without the asparagine residue or a small peptidic sequence do not interact with these immobilized lectins. This is an important point to consider in a fractionation procedure to be used because the discriminating power of an immobilized LCA–or WGA–Sepharose™ column is considerable and cannot be neglected.

As previously described for glycoproteins, sequential use of immobilized lectins with different and well-defined specificities can fractionate complex mixtures of glycopeptides or glycans into homogeneous families. Generally, lectins with affinity for internal saccharide sequences are used first in the fractionation, whereas immobilized lectins with a specificity towards terminal oligosaccharide-dominant sequences or terminal nonreducing monosaccharides are used later.

4.3.1.2 Fractionation of N*-Glycosylpeptides or Related Glycans on Immobilized Lectins with Affinity for Internal Oligosaccharidic Sequences*

Affinity Chromatography on Immobilized Concanavalin A and Lens culinaris Agglutinin. Figure 8 presents the structures of *N*-glycosylpeptides or related glycans which can be found in the three fractions recovered from an immobilized Con A column as well as in the six fractions resulting from a subsequent affinity chromatography of *N*-glycosylpeptides on an LCA–Sepharose™ column. From this figure, it clearly appears that the α-1,6-linked fucose residue is an important determinant for the binding of *N*-glycosylpeptides and that affinity chromatography on immobilized LCA is a valuable tool with which to fractionate glycopeptides into α-1,6-fucosylated and nonfucosylated types. Only *N*-glycosylpeptides can be bound on immobilized LCA. So, glycans released from *N*-glycosylpeptides either by hydrazinolysis or by enzymatic cleavage are eluted at the void volume of the lectin column.

Immobilized *Vicia faba* and *Pisum sativum* lectins interact also with some α-1,6-fucosylated *N*-glycosylpeptides.

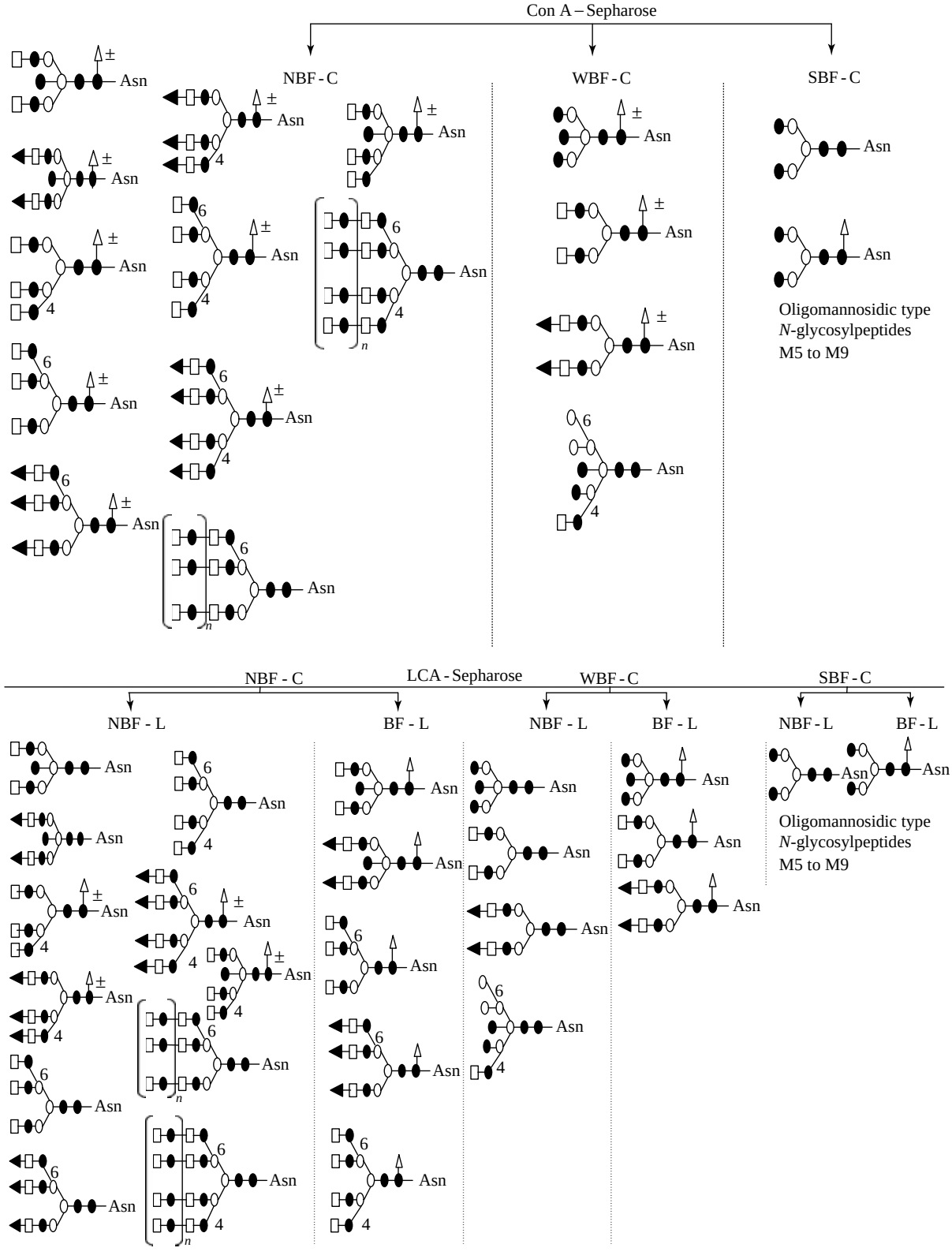

Figure 8 Interaction of *N*-glycosylpeptides with immobilized Con A and LCA (NBF = nonbound fraction, WBF = weakly bound fractions, BF = bound fractions, SBF = strongly bound fractions).

However, their affinity is weaker, so that the fractionation into the two classes is obtained with only the equilibration buffer.

Affinity Chromatography on Other Lectins with a Specificity for Internal Saccharidic Sequences. Other immobilized lectins with a specificity for internal saccharide sequences can also be used to subfractionate *N*-glycosylpeptide families obtained by affinity chromatography on immobilized Con A and LCA.

As shown in Table 3, *Phaseolus vulgaris* isolectin E4 (E4-PHA) interacts with bisected bi- and triantennary *N*-acetyllactosamine-type glycopeptides or related glycans, and *Phaseolus vulgaris* isolectin L4 (L4-PHA) interacts with asialotri′- and tetraantennary *N*-acetyllactosamine-type glycopeptides or related glycans, containing the dominant pentasaccharidic sequence Gal(β1–4)GlcNAc-(β1–2)[Gal(β1–4)GlcNAc(β1–6)]Man. The two immobilized lectins can be used to subfractionate glycopeptide fractions containing these oligosaccharidic determinants previously isolated by a sequential affinity chromatography on immobilized Con A–and LCA–sepharose™.

Immobilized WGA can be used to fractionate some hybrid-type *N*-glycosylpeptides which are bound on a WGA–Sepharose™ column, and to a lesser extent bisected biantennary *N*-acetyllactosaminic-type glycopeptides which are only retarded on such a column.[162,163] Linear poly(*N*-acetyllactosamine) as well as branched poly(*N*-acetyllactosamine)-type glycopeptides or glycans are also recognized with high affinity by immobilized WGA.[164–166] Immobilized WGA also interacts with glycopeptides having a high density of *O*-glycosidically linked sialyl oligosaccharides.[167,168]

4.3.1.3 Fractionation of N-*Glycosylpeptides or Related Glycans on Immobilized Lectins with Affinity Directed Towards Terminal Oligosaccharide-dominant Sequences*

*Lectins with a Specificity for Poly(*N*-acetyllactosamine)-type Glycans.* Immobilized pokeweed mitogens,[169] immobilized *Lycopersicon esculentum* agglutinin (tomato lectin),[95,170] or immobilized DSA[171] represent useful tools for the fractionation of *N*-acetyllactosamine-type glycans with repeating *N*-acetyllactosamine sequences.

Lectins with a Specificity for Terminal N-*Acetyllactosamine Sequences.* Immobilized RCA I and RCA II (*Ricinus communis* agglutinin II) present distinct oligosaccharide specificities and the combined use of these two lectins allows the fractionation of *N*-acetyllactosamine-type glycans according to the nature and the linkage of their terminal sugar moieties.[90] Like immobilized RCA I, immobilized RCA II strongly interacts with *N*-acetyllactosamine-type oligosaccharides possessing terminal β-1,4-linked galactose residues. However, immobilized RCA II can also strongly interact with *N*-acetyllactosamine-type oligosaccharides containing nonreducing β-1,4-linked GalNAc residues. Moreover, whereas immobilized RCA I has a weak affinity for *N*-acetyllactosamine-type oligosaccharides with one or two terminal β-1,3-linked galactose residues, immobilized RCA II strongly binds these glycans which can be eluted with a GalNAc solution.

The carbohydrate-binding specificity of four *Erythrina* (*E. cristagalli, E. latissima, E. corallodendron, E. lysistemon*) lectins is also directed towards unmasked terminal *N*-acetyllactosamine sequences.[172] As for immobilized RCA I, the interaction of *N*-acetyllactosamine-type glycans with immobilized *Erythrina* lectins depends on the number of terminal *N*-acetyllactosamine sequences.

Lectins with a Specificity for Sialylated N-*Acetyllactosamine Sequences.* The immobilized lectin *Sambucus nigra* agglutinin I (SNA I), isolated from the bark of elder, recognizes oligosaccharides, glycopeptides, or glycoproteins with terminal NeuAc(α2–6) Gal/GalNAc sequences which are bound and can be eluted from the lectin column with a lactose solution. However, asialo derivatives as well as those containing NeuAc(α2–3)Gal/GalNAc sequences do not interact with immobilized SNA I.[173]

Immobilized *Maackia amurensis* leukoagglutinin (MAL) binds with high-affinity *N*-acetyllactosamine-type glycopeptides containing terminal NeuAc(α2–3)Gal-(β1–4)GlcNAc sequences. Conversely, asialoglycopeptides as well as glycans with terminal NeuAc(α2–6)Gal-(β1–4)GlcNAc sequences do not interact with the immobilized leukoagglutinin.[174]

Therefore, immobilized MAL and SNA I are two interesting lectins, which can be sequentially used to fractionate sialylated glycans on the basis of their sialic acid linkages and according to the number of sialic acid residues they contain.

4.3.1.4 Subfractionation of N-*Glycosylpeptides on Immobilized Lectins with a Specificity Directed Towards Terminal Nonreducing Dominant Monosaccharides*

Lectins Recognizing Nonreducing Mannose Residues. The immobilized *Galanthus nivalis* agglutinin (GNA),[175] *Narcissus pseudonarcissus* agglutinin (NPA),[176] *Hippeastrum hybrida* agglutinin (HHA)[176] isolated from snowdrop, daffodil or amaryllis bulbs, as well as

Bowringia milbraedi agglutinin (BMA)[177] isolated from the seeds of the Nigerian legume *Bowringia milbraedi* provide interesting tools for the analysis and fractionation of glycoproteins or glycopeptides containing oligomannosidic-type glycans.

Lectins Recognizing Terminal Nonreducing N-*Acetylgalactosamine and Galactose Residues.* Seeds of *Griffonia simplicifolia* contain five tetrameric isolectins composed of two subunits A and B.[178] The carbohydrate-binding specificity of the two subunits differs significantly. The A4 tetrameric isolectin GSA I-A4 recognizes terminal nonreducing α-linked *N*-acetylgalactosamine residues, whereas the B4 isolectin (GSA I-B4) has strict specificity towards terminal nonreducing α-linked galactose residues.

Immobilized GSA I isolectins can be used to isolate either α-Gal-containing glycopeptides[179] or α-GalNAc-containing glycopeptides.[180]

Glycopeptides containing the GalNAc(β1–4)GlcNAc sequence in their outer branch moieties bind with high affinity to immobilized *Wisteria floribunda* agglutinin.[181]

Lectins Recognizing Terminal Nonreducing N-*Acetyl Glucosamine Residues.* Immobilized *Griffonia simplicifolia* agglutinin II (GSA II)[182] as well as immobilized *Psathyrella velutina* agglutinin, a lectin isolated from the fruiting bodies of the mushroom *Psathyrella velutina*[183] represent useful tools for the detection and fractionation of glycoproteins, glycopeptides or glycans with terminal nonreducing GlcNAc residues.

Fucose-binding Lectins. Immobilized *Lotus tetragonolobus* agglutinin (LTA) displays strong affinity for poly-*N*-acetyllactosamine-type *N*-glycosylpeptides containing high amounts of fucose linked α-1,3 to *N*-acetyllactosamine residues within the linear repeating disaccharide 3-Gal(β1–4)GlcNAc(β1–3)$_n$ (Lewis^x antigenic blood group). These glycopeptides are bound to an immobilized LTA column and can be eluted with a fucose solution.[184]

Another immobilized fucose-binding lectin, isolated from fruiting bodies of the mushroom *Aleuria aurantia*, represents a very valuable tool for the resolution of glycopeptides and glycans embellished with different Fuc substituents.[185,186]

4.3.1.5 Fractionation of O-*Glycosylpeptides Using Immobilized Lectins* A large panel of lectins with well-defined specificities towards oligosaccharidic sequences belonging to *N*-glycosylproteins is now available for the study of glycan microheterogeneity or for fractionation of *N*-glycoproteins, glycopeptides, or related glycans. However, very few applications can be found in the literature of similar studies using lectins specifically recognizing *O*-glycosidically linked glycans.

Nevertheless, by inhibition of hemagglutination or by quantitative precipitation, many lectins were shown to bind to mucin-type oligosaccharides.[75,187]

The carbohydrate-binding specificities of five of these lectins (ABA I (*Agaricus bisporus* agglutinin I), PNA (*Arachis hypogaea* agglutinin), BPA (*Bauhinia purpurea* agglutinin), SBA (soybean agglutinin) from *Glycine max*, and VVA B4 (*Vicia villosa* agglutinin B4)) has been studied by affinity chromatography of mucin-type glycopeptides and related oligosaccharides of known structures.[188]

Another lectin, jacalin, isolated from jackfruit (*Artocarpus integrifolia*) with a high affinity for the T-antigen Gal(β1–3)GalNAc[189] represents a useful tool for the analysis and fractionation of glycoproteins containing *O*-glycosidically linked glycans[190] as well as for the fractionation of *O*-glycosylpeptides or related oligosaccharide-alditols.[191,192]

4.3.2 High-performance Liquid Chromatography Separation of Oligosaccharides and Glycopeptides

4.3.2.1 Reversed-phase High-performance Liquid Chromatography of Glycopeptides Fractionation of glycopeptides released by proteolytic cleavage may be performed by hydrophobic binding and reversed-phase chromatography.

Separation of glycopeptides is generally achieved on C_{18} columns (octadecylsilyl silica) or C_8 columns (octasilyl silica). The essential criterion responsible for separation is the interaction of the packing with polar materials. Reversed-phase high-performance liquid chromatography (RPHPLC) will allow separation on the basis of peptide hydrophobicity. Modification of the amino terminus with tyrosine increases this hydrophobicity.[193] Under some elution conditions the degree of sialylation of glycopeptides can influence the separation. In the general procedure using aqueous solutions or solvents of medium polarity (acetonitrile), the more polar species elute first and then, as the polarity decreases, the tightly bound more hydrophobic species are eluted. When analyzing glycosylation site heterogeneity, an initial separation of glycopeptides according to their peptide structure, but not according to the oligosaccharide(s) structure, is advantageous.[194–196] The site-specific glycopeptides can then be fractionated on the basis of their carbohydrate structure. Glycopeptides separated by RPHPLC may be further studied by MS.

4.3.2.2 High-performance Liquid Chromatography Separation of Oligosaccharides As a result of their extreme structural diversity, oligosaccharides isolated

from glycoproteins are complex mixtures often of closely related isomeric compounds. For this reason HPLC has proved to be of great value in oligosaccharide purification because of the wide range of absorbents and solvent systems available as well as the speed of separation. Glycans can be separated according to different structural parameters such as their size, charge, or composition.[197–199] Some chromatographic systems such as HPAEC (high-pH anion-exchange chromatography) will also be selective for the ring substitution and branching configuration.[200]

Reversed-phase High-performance Liquid Chromatography of Oligosaccharides. Separation of neutral oligosaccharides may be achieved at room temperature on C_{18} octadecyl silane (ODS) RPHPLC columns using either pure water as an eluent or a low acetonitrile concentration (1–5%). As oligosaccharides occur in solution as a mixture of anomers (α, β) a pure oligosaccharide may provide two peaks during the HPLC separation or more generally poorly resolved broad peaks are obtained.

Free nonreduced oligosaccharides released from glycoproteins either by hydrazinolysis or endoglycosidases can be derivatized on their reducing group with an amino group containing fluorescent or ultraviolet (UV) dyes (Scheme 1) by reductive amination reaction.[201,202] This increases the sensitivity of detection and quantities in the picomole range of oligosaccharides can be detected. The resolution by RPHPLC of the derivatized oligosaccharides is also improved. Furthermore, the derivatized oligosaccharides can be directly analyzed by fast atom bombardment mass spectrometry (FABMS) or matrix-assisted laser desorption/ionization mass spectrometry (MALDIMS), the fluorescent or UV probes enhancing the ion intensity. A number of procedures have been described for the derivatization of oligosaccharides, the most popular methods being pyridylamination[203,204] and benzoylation.[205]

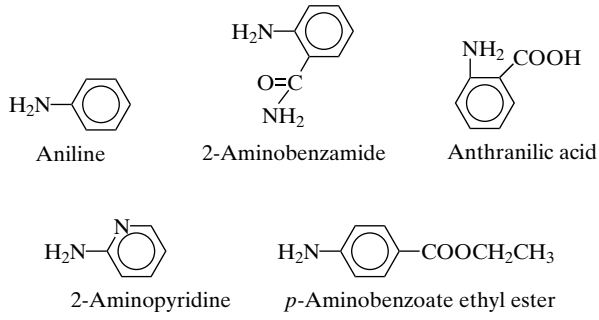

Scheme 1 Principal dyes used for tagging oligosaccharides.

Sialylated oligosaccharides are not retained on the C_{18} columns. Separation of sialylated oligosaccharides may be obtained by neutralizing their charges with triethylamine. Elution is made isocratically with triethylamine acetic acid buffer (10 mM, pH 5.0).[206] UV detection is then impossible and carbohydrates are detected using a refractometer. This method has been used for the separation of human milk sialyl oligosaccharides.[206–208] Separation of sialylated compounds is also possible in the derivatized forms.[209]

Elution of oligomers on ODS columns gives a separation with an increased retention time for oligomers with higher M_r values.[210] When oligosaccharides contain both neutral and acetamido sugars, oligosaccharides having a higher acetamido/neutral sugar ratio have a longer retention time because of the relative hydrophobicity of acetamido groups compared with hydroxyl groups. The same hydrophobicity is observed for fucosylated oligosaccharides, which are generally more retained on the column compared with the nonfucosylated form.[211]

If RPHPLC of neutral oligosaccharides is rather limited, separation of derivatized oligosaccharides (pyridylamino, benzoylated) gives a better separation of these compounds according to their molecular size or the nature of constituent monosaccharides (hexosamines, fucose). For these reasons, RPHPLC of derivatized oligosaccharides is used in two-dimensional mapping strategies. These methods are based on parameterization of the contribution of the individual component sugars to the chromatographic behavior of the oligosaccharides. Thus the oligosaccharides are chromatographed on two HPLC columns and their elution volumes are expressed relative to those of standard glucose oligomers.[212–219]

RPHPLC separation of perbenzoylated high-mannose-type oligosaccharides[205,220] allows the separation of positional isomers.

Ion-exchange High-performance Liquid Chromatography.

- Anion-exchange chromatography. Oligosaccharides may be differently charged according to the presence of acidic monosaccharides such as sialic acid, GlcA, or the presence of anionic substituents (sulfate or phosphate groups). Therefore, glycans can be subjected to ion exchange chromatography and separated into neutral, mono- di-, tri-, or tetracharged compounds. Separation of oligosaccharides according to the number of charges generally constitutes the first step in glycan separation. Either strong or weak anion exchangers can be used for the separation of oligosaccharides.[221]

  Sialylated and sulfated oligosaccharides can be separated on a Micropak® AXS (30 cm × 4.6 mm) column (Varian)[222] and acidic oligosaccharides have

been separated on a Mono-Q® HR 5/5 (50 cm × 5 mm) column (Pharmacia).[223]

- HPAEC of oligosaccharides. The recent introduction of HPLC using pellicular ion-exchange resins under high-pH conditions and detection of sugars with pulsed amperometric detection (PAD) has simplified the separation and analysis of both mono- and oligosaccharides.[200,224] The method may be used for analytical (i.e. for mapping oligosaccharides released from glycoproteins) or for preparative purposes, when it constitutes the final step in oligosaccharide purification. HPAEC has been applied with success in the separation of glycans originating both from *N*-glycosylproteins and mucins as well as from glycolipids.[224,225] The high resolving power of HPAEC and the high sensitivity of pulsed amperometry detection has allowed for a simple one-step procedure for profiling the total N-linked oligosaccharide released from a glycoprotein by either endoglycosidase digestion or hydrazinolysis. The advantages of HPAEC/PAD are its ability to separate underivatized oligosaccharides and its sensitivity (10 pmol). The high resolution obtained with commercially available columns makes possible the profiling of any given glycoprotein. Isomeric forms differing in only a single linkage or in branch position of one residue may be resolved by HPAEC.

  The principle of the method lies in the fact that carbohydrates – which are polyhydric compounds – may be considered to be weak acids, thus in basic solution oligosaccharides form negatively charged oxyanions that will interact on strong anion-exchange columns.

  Predicting the elution order of oligosaccharides is largely empirical because different factors affect the formation of oxyanions, such as the number and nature of the monosaccharides, linkage, branch location, hydrophobicity, or steric conformation. However, the following observations hold:

  - The retention volume increases with the degree of polymerization.
  - Oligosaccharide-alditols have shorter retention times in comparison to nonreduced oligosaccharides.
  - Oligosaccharides with a fucose (α1–6)-linked to the chitobiosyl core elute earlier than nonfucosylated oligosaccharides.
  - Sialylated oligosaccharides with (2–3)-linked NeuAc are more retained than those with (α2–6)-linked NeuAc.
  - Elongation with an additive Gal(β1–4)GlcNAc sequence decreases the retention time.
  - Oligosaccharides with a Gal(β1–3)GlcNAc unit elute later than their counterparts with a Gal(β1–4)GlcNAc sequence.
  - Branching tends to shorten the retention time.

Adsorption High-performance Liquid Chromatography of Oligosaccharides.

- *Silica adsorption HPLC*. Adsorption or normal-phase chromatography relies on the surface hydroxyl groups of silica which can interact with solutes, thereby allowing separation according to the different strengths of interactions. This method is generally used for the separation of derivatized oligosaccharides with a low degree of polymerization using nonaqueous eluents.
- *Hydrophobic HPLC on porous graphitized carbon*. These types of columns are now commercially available from Oxford Glycosystems (Glycosep® H). The columns are run in the same eluents as RPHPLC (aqueous/organic solvents in the presence of 0.05% trifluoroacetic acid (TFA)). Separation is based on the interactions of the hydrophobic areas of glycopeptides and oligosaccharides with the carbon surface.[226] The method may be used for the separation of derivatized oligosaccharides and provides separation of both neutral and sialyl oligosaccharides.

Size-partition High-performance Liquid Chromatography of Oligosaccharides.

- *Ion-suppression Amine-adsorption HPLC*. This method provides for the separation of neutral or negatively charged oligosaccharides, according to their size, on amine columns. Whereas neutral oligosaccharides can be eluted from the column by a liner gradient of acetonitrile/H_2O by increasing the proportion of water,[227] anionic oligosaccharides interact with the solid phase and cannot be eluted.

  The ionic interactions of such negatively charged oligosaccharides can be suppressed by including 3% (v/v) acetic acid titrated to pH 5.5 with triethylamine in the mobile phase. This modification of the mobile phase allows oligosaccharides to be eluted in order of size. Nevertheless, it should be mentioned that separations by ion-suppression amine-adsorption HPLC are sensitive to the identity of the anionic moiety, in that oligosaccharides containing (α2–3)-linked sialic acid elute earlier than those containing (α2–6)-linked sialic acid,[228] and to the branching pattern.[229]
- *Bonded-phase HPLC*. The most frequently used systems for the separation of oligosaccharides are those using chemically bonded phases which fractionate

material on the basis of their relative affinities for the mobile phase and the bonded phase. Supports containing cyano, diol or aminopropyl[230–233] phases may be used. This type of HPLC not only provides size separation of the oligomers, but also of isomers differing in the anomericity or their branching pattern.

- *Gel-permeation HPLC*. Gel-filtration medium-pressure HPLC of oligosaccharides may be realized on noncompressible matrices such as hydroxylated polyether-based matrices (TSK™-PW or TSK™-HW, from Merck). Nevertheless, the separations are inferior to those currently obtainable by other HPLC techniques and consequently little emphasis is placed here on gel-permeation HPLC of oligosaccharides. This method is generally limited to the desalting of oligosaccharide fractions. However, it should be mentioned that Oxford Glycosystems have developed a method of oligosaccharide analysis based on the sequential degradation by exoglycosidases followed by gel-permeation chromatography of the enzymatic hydrolysate. Each oligosaccharide gives a specific pattern calibrated in glucose units obtained by the reagent array analytical method (RAAM).[234]

High-performance Affinity Chromatography of Oligosaccharides. The introduction of macroporous silica as a solid support for affinity columns has improved their resolution and shortened their analysis time.[235,236] This method is based on weak ligand–ligate interactions at high concentrations of bound ligand on rigid particles approximately 10 μm in diameter. Use of monoclonal antibodies[237] or lectins as ligands[238] allows excellent chromatographic separation of oligosaccharides.

4.3.3 Electrophoretic Separation of Glycans and Glycoforms

4.3.3.1 Electrophoretic Separation of Glycans and Glycoforms by Capillary Electrophoresis During the last years capillary zone electrophoresis (CZE) has become a good alternative and rapid procedure for profiling both glycans and glycoforms.[239–242] The procedure is based on the differential migration of solutes through a narrow-bore fused-silica capillary. Charged compounds including sialylated, sulfated, and phosphorylated oligosaccharides migrate in an applied electric field at neutral pH. Neutral sugars can be given a charge by simply adjusting the pH of their environment (high pH), complexation with borate-containing buffers, or conjugation with charged species.[243–246]

A major limitation of using CZE for carbohydrate analysis has been the absence of sensitive detection methods. UV detection has been used for both underivatized and derivatized carbohydrates but it has limited detection sensitivity.[247] In contrast, fluorescence detection is easily adapted after conjugation of carbohydrates to different fluorophores, and charged fluorescent tags such as 7-amino-1,3-naphthalene disulfonic acid are commonly used.[248] Laser sources are capable of focusing light in small capillaries and thus give increased sensitivity of detection. Detection can also be achieved by PAD.[249]

In addition to oligosaccharides, CZE provides efficient separation of glycoforms of a single glycoprotein, according to the occupancy of the different glycosylation sites. This method is routinely used for control of the glycosylation pattern of recombinant glycoproteins.[250–253]

4.3.3.2 Fluorophore-assisted Carbohydrate Electrophoresis Acidic oligosaccharides can be separated by polyacrylamide gel electrophoresis (20–40% slab gels) according to their charge. For that reason, only charged compounds (e.g. those containing sialic acid, uronic acids or charged *O*-glycans such as sulfate or phosphate groups) can be analyzed by electrophoresis. To overcome this limitation of the FACE technique, reducing oligosaccharides are coupled to negatively charged fluorescent dyes such as 8-amino-naphthalene-1,3,6-trisulfonate, 4-amino-naphthalene sulfonate, or 7-amino-1,3-naphthalene disulfonate.[254–258]

Once a carbohydrate has been derivatized with one of these charged fluorescent tags, the carbohydrate molecule will possess an overall net negative charge and therefore will migrate in an electric field. Moreover, the fluorophore provides a high level of detection sensitivity and carbohydrate bands can be vizualized under UV (360 nm) or using a commercial imaging system, such as a charge-coupled device (CCD), to quantify the amount of oligosaccharide present in each band. This method permits separation of charged or uncharged oligosaccharides with high resolution and can detect single hydroxyl anomeric differences between mono- and oligosaccharides of sugars with otherwise identical molecular weight charge and sequence. The method can be combined with sequential hydrolysis using specific exoglycosidases in order to assess the structure of isolated glycans.[234]

The FACE method is a convenient and rapid procedure largely used in the quality control of recombinant glycoprotein glycosylation.[259,260]

4.4 Structural Characterization of Glycans

Complete structural analysis of purified glycans requires determination of: (a) the type and number of constituent monosaccharides including their ring form (D or L) and their anomeric linkages (α or β); (b) the positions of the linkages between individual sugars and their sequence;

and (c) the type, number, and position of any noncarbohydrate substituents. Several different methods are required to provide all the information. Commonly used methods combined GLC and HPLC chromatographies with physicochemical methods such as MS or NMR.

Monosaccharides commonly found in glycoproteins may be divided into neutral (Man, Gal or Fuc) hexosamines (GalNAc or GlcNAc) and acidic compounds (Neu5Ac or GlcA).

Additive heterogeneity comes from the possible substitution with aglycone residues such as sulfate, phosphate, or acetate groups. Prior to their analysis, monosaccharides have to be released from the oligosaccharide chain by acidic hydrolysis. Different methods are available for the analysis of monosaccharides depending mostly on the amount of material available. Several techniques, such as GLC or HPLC, allow both quantitative and qualitative analysis of monosaccharide mixtures. Other chromatographic or electrophoretic procedures just allow a rapid qualitative analysis of samples. Single separated monosaccharides may be further identified by physicochemical methods such as MS or NMR.

4.4.1 Release of Monosaccharides

Prior to their analysis, individual monosaccharides have to be released. The way in which individual monosaccharides are released and processed depends on the methods, which will be further used for their analysis. Different hydrolysis procedures using either chlorhydric, sulfuric, trifluoroacetic, or acetic acid may be used. It has to be kept in mind that the susceptibility to acid hydrolysis as well as the stability of the released units is different for each monosaccharide. Mild hydrolysis conditions may be employed to specifically release some sugars as fucose or sialic acids:

- Fucosyl residues are typically present on glycans in nonreducing external terminal positions; it is therefore susceptible to acid hydrolysis under mild conditions using diluted (0.05 M) HCl or TFA solutions.[261]
- Mild acid hydrolysis using acetic acid and/or enzymatic release with sialidase is recommended for sialic acid analysis.[262] Indeed, sialic acids represent a family of different compounds, differing by the presence of various substituents on the hydroxyl groups, which are rather labile and may be lost during the hydrolysis. The modifications have been shown to affect a wide spectrum of biological phenomena.[263] When a GLC approach is taken, methanolysis[264] is the method of choice because cleavage of all glycosidic linkages is very effective and less destruction of the monosaccharides is observed than during acid hydrolysis. Monosaccharides are converted to their methylglycosides. It should be mentioned that *N*-acetyl groups of the main sugars are completely cleaved, and a re-*N*-acetylation step should be introduced to reincorporate acetyl groups in hexosamines and sialic acids.

4.4.2 Gas–Liquid Chromatographic Analysis of Monosaccharides

GLC is a sensitive analytical technique for monosaccharide composition analysis, allowing detection of subnanomole amounts of carbohydrates. The method provides information on both the nature and the amount of the different monosaccharides.[262,265] The most commonly used GLC columns are fused-silica wall-coated capillary columns. Detection may be carried out by flame ionization detection (FID) or the gas chromatograph may be coupled to MS (i.e. GLC/MS). This last method provides a further refinement in identification.

Monosaccharides can also be identified by their specific fragmentation patterns.[266] Prior to their analysis, monosaccharides have to be derivatized. Trimethylsilylation of hydroxyl and carboxyl groups is the most widely used method.[267,268] If only neutral monosaccharides need to be measured, acid hydrolysis followed by reduction and peracetylation (alditol acetate derivatives) is preferred.[269,270]

4.4.3 Methylation Analysis

The use of methylation analysis in the structural determination of oligosaccharides represents the most classical method for defining both the substitution pattern of a given compound (in terms of linkage position, terminal sugars, branching points) and the ring size of the different monosaccharides. Methylation procedures differing in the nature of the basic reagent used for proton release have been described.[271–274] The steps of the methylation procedure are described in Scheme 2. The sequence is: (a) substitution of the different hydroxyl groups not engaged in the glycosidic linkages by methyl groups in alkaline medium; (b) hydrolysis of the methylated oligosaccharides; (c) substitution of the newly generated hydroxyl groups (previously engaged in the linkage) by another substituent such as acetyl groups; and (d) GLC separation of the different methylated ethers. The positions of methyl groups reflect free hydroxyl groups in the original monosaccharides of the native oligosaccharide, whereas acetyl groups give the position of hydroxyl groups initially engaged in the glycosidic linkages. Partially methylated and acetylated monosaccharides can be identified by GLC according to their retention time and in comparison with standard compounds.[275,276] Methylation analysis can be combined with GLC/MS analysis,

and the different methyl ethers identified on the basis of their specific fragmentation patterns obtained by electron ionization mass spectrometry (EIMS) or by chemical ionization mass spectrometry (CIMS).

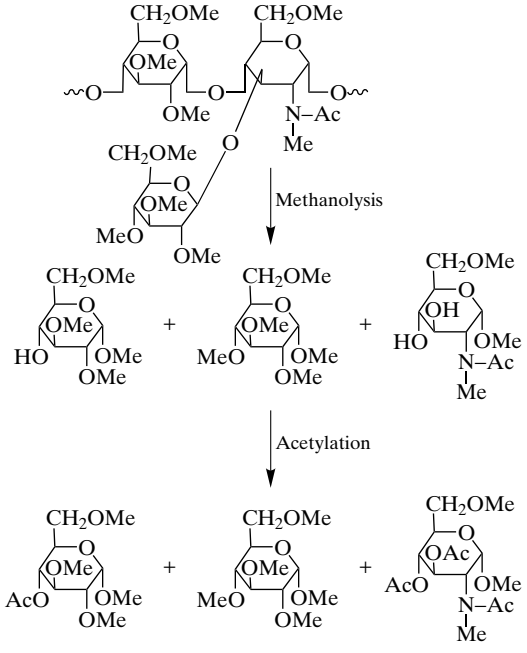

Scheme 2 Steps in the methylation procedure.

4.4.4 Nuclear Magnetic Resonance Spectroscopy in the Study of Glycans

NMR spectroscopy provides a powerful and non-destructive method for structural study of carbohydrates. It can provide almost complete structural information on the glycan including: (a) the nature of the monosaccharides including the ring size; (b) the linkage anomericity; (c) the linkage position; and (d) the nature and degree of substitution by aglycon groups (e.g. sulfate, phosphate, acetyl). Generally a single ^{1}H-NMR analysis is sufficient to assign all the parameters of a known oligosaccharide according to the structural reporter group method developed by Vliegenthart et al.[277,278] The ^{1}H-NMR spectra may be divided into three distinct regions, the region between 6 and 4.5 ppm being the most important for establishing the number of monosaccharide units. The coupling constant values enable discrimination between the α and β anomericity. The bulk region (3.6–3.2 ppm) is generated by other protons (H-2 to H-5) and cannot generally be interpreted in a conventional one-dimensional analysis. The third region (1–3 ppm) is also very useful for structural determination because it corresponds to axial and equatorial protons of sialic acids as well as acetamido protons from hexosamines. Several bidimensional homo- or heteronuclear correlated spectroscopy procedures provide other types of information concerning the oligosaccharide sequence.[279] The more useful are: (a) the correlation spectroscopy (COSY) experiments which measure the coupling constant J between the different monosaccharides; (b) heteronuclear multiquantum coherence (HMQC) which allows clear identification of the carbon substitutions; (c) heteronuclear multiple-bond correlation (HMBC) which gives direct access to the monosaccharide sequence; and (d) rotating-frame Overhauser effect spectroscopy (ROESY), which allows identification of short-distance proton interactions. The nuclear Overhauser effect intensities may be easily translated into conformational information.

4.4.5 Mass Spectrometry Methods for Structural Analysis of Glycans

The introduction of new techniques such as MALDIMS and electrospray ionization mass spectrometry (ESIMS) has had a considerable impact on the application of MS to carbohydrate research. An MS experiment provides two types of structural information, namely the masses of the intact molecules (molecular ions) and the masses of fragment ions. For oligosaccharides, the structural information from MS is essentially limited to monosaccharide sequence and molecular weight, and only in exceptional cases can glycosidic linkage positions be obtained.

Derivatization of glycans, such as by methylation or coupling to fluorescent dyes, generally increases the signals or may provide specific fragmentation patterns necessary to determine glycosidic linkage positions.[280–283]

4.4.5.1 Fast Atom Bombardment Mass Spectrometry Although FABMS can be applied to native oligosaccharides, the derivatized compounds (permethylated or peracetylated) are preferred because the sensitivity of the analysis is significantly increased.[284–287] The most important fragmentation observed concerns the cleavage of the glycosidic linkages. In particular, fragment ions are predominantly formed by cleavage of the glycosidic bonds engaged with amino sugars, such as GlcNAc or GalNAc.[288] Even though FABMS is less sensitive than other more modern MS techniques, it remains the method of choice for the determination of the primary structure of glycans.

4.4.5.2 Matrix-assisted Laser Desorption/Ionization Mass Spectrometry The use of lasers for ionizing carbohydrates has become popular for analysis of both glycans and their glycoconjugates.[280,289,290] In this technique, samples are mixed with a UV-absorbing matrix such as 2,5-dihydroxybenzoic acid (DHB)[291] or 2′,4′,6′-trihydroxyacetophenone (THAP)[289] and are ionized within the mass spectrometer by irradiation with a UV

laser. Spectra are usually examined with a linear time-of-flight (TOF) mass spectrometer, but greatly increased resolution can be obtained with a reflectron-type mass analyzer. The MALDIMS technique can be used for analysis of large molecules such as glycopeptides or intact glycoproteins. The technique is particularly useful for determining the presence or absence of glycosylation in a glycoprotein and may be combined with deglycosylation experiments using exo- or endoglycosidases.[292–294]

4.4.5.3 Electrospray Ionization Mass Spectrometry

The major application of ESIMS is in the analysis of glycopeptides and glycoproteins which possess multiple charge sites. The method is of particular interest for glycoform study[295,296] and for determination of the glycosylation site occupancy. Glycopeptides are obtained by proteolytic digestion of the native glycoproteins, separated by HPLC and analyzed by ESIMS. A number of methods have been devised to identify the glycopeptides in such analyses using either specific endoglycosidases or specific daughter ions characteristics of the carbohydrate (e.g. m/z 204 from *N*-acetylhexosamine) to highlight the glycopeptides peaks.[296,297] The method can be coupled to microbore liquid chromatography (LC), permitting on-line LC/ESIMS analysis, or to capillary electrophoresis which is useful when complex mixtures of peptides and glycopeptides are being examined (e.g. after proteolytic digestion of a glycoprotein).[298–300] Free neutral carbohydrates lack a natural charge size and are more difficult to ionize by ESI. However, due to the propensity of carbohydrates to attract sodium, sodium salts can be added to the spray to enhance ionization.[301] Increased sensitivity can be achieved by introduction of a constituent charge, after derivatization of free oligosaccharides with UV or fluorescent dyes by the reductive amination method or other procedures.[302–305]

5 CONCLUSIONS

The study of the glycosylation pattern of glycoprotein remains an enormous task due to the extreme heterogeneity of the glycans present at each glycosylation site. Physicochemical methods such as MS analyses associated with powerful separation procedures such as HPAEC, CZE, or FACE have made structural studies feasible on very small quantities of material. Improvements to these new methods have enabled the characterization of new types of glycosylation. Structural glycobiology has become part of the ''post-genome'' program, and the study of glycosylation can no longer be considered independent of protein characterization. New strategies, such as the proteomic approach,[306,307] propose to study both glycans and peptidic sequences in a single step.

ACKNOWLEDGMENTS

The authors are deeply indebted to Mrs Catherine Alonso for her skilful assistance in typing this manuscript.

ABBREVIATIONS AND ACRONYMS

ABA I	*Agaricus bisporus* Agglutinin I
Asn	L-Asparagine
BMA	*Bowringia milbraedi* Agglutinin
BPA	*Bauhinia purpurea* Agglutinin
CAIE	Crossed Affinoimmuno-electrophoresis
CCD	Charge-coupled Device
CIMS	Chemical Ionization Mass Spectrometry
CN	Cellulose Nitrate
Con A	Concanavalin A
COSY	Correlation Spectroscopy
Cys	Cysteine
CZE	Capillary Zone Electrophoresis
DEAE	Diethylaminoethyl
DHB	2,5-Dihydroxybenzoic Acid
DIG	Digoxigenin
DSA	*Datura stramonium* Agglutinin
EIMS	Electron Ionization Mass Spectrometry
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization Mass Spectrometry
FABMS	Fast Atom Bombardment Mass Spectrometry
FACE	Fluorophore-assisted Carbohydrate Electrophoresis
f-Ara	L-Furanoarabinose
FID	Flame Ionization Detection
FPLC	Fast Protein Liquid Chromatography
Fuc	L-Fucose
Gal	D-Galactose
GalNAc	*N*-Acetyl-D-galactosamine
GLC	Gas–Liquid Chromatography
Glc	D-Glucose
GlcA	Glucuronic Acid
GlcNAc	*N*-Acetyl-D-glucosamine
GNA	*Galanthus nivalis* Agglutinin
GPI	Glycosylphosphatidylinositol
GSA I	*Griffonia simplicifolia* Agglutinin I
GSA II	*Griffonia simplicifolia* Agglutinin II

HHA	*Hippeastrum hybrida* Agglutinin
HMBC	Heteronuclear Multiple-bond Correlation
HMQC	Heteronuclear Multiquantum Coherence
HPAEC	High-pH Anion-exchange Chromatography
HPLC	High-performance Liquid Chromatography
ITC	Isothermal Titration Microcalorimetry
IUPAC–IUB	International Union of Pure and Applied Chemistry–International Union of Biochemistry
Kdn	2-Keto-3-deoxynononic Acid
LC	Liquid Chromatography
LCA	*Lens culinaris* Agglutinin
LTA	*Lotus tetragonolobus* Agglutinin
MAL	*Maackia amurensis* Leukoagglutinin
MALDI	Matrix-assisted Laser Desorption/ Ionization
MALDIMS	Matrix-assisted Laser Desorption/ Ionization Mass Spectrometry
Man	D-Mannose
MS	Mass Spectrometry
Neu5Ac	*N*-Acetyl-neuraminic Acid
Neu5Gc	*N*-Glycolylneuraminic Acid
NMR	Nuclear Magnetic Resonance
NPA	*Narcissus pseudonarcissus* Agglutinin
ODS	Octadecyl Silane
PAD	Pulsed Amperometric Detection
PBS	Phosphate-buffered Saline
PDTRAPGS	Pro-Asp-Thr-Arg-Ala-Pro-Gly-Ser
PHA	*Phaseolus vulgaris* Phytohemagglutinin
PNA	*Arachis hypogaea* Agglutinin
PVDF	Polyvinylidene Difluoride
RAAM	Reagent Array Analytical Method
RCA I	*Ricinus communis* Agglutinin I
RCA II	*Ricinus communis* Agglutinin II
RIP	Ribosome-inactivating Protein
ROESY	Rotating-frame Overhauser Effect Spectroscopy
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SBA	Soybean Agglutinin
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate/ Polyacrylamide Gel Electrophoresis
Ser	Serine
SNA I	*Sambucus nigra* Agglutinin I
TBS	Tris-buffered Saline
TFA	Trifluoroacetic Acid
THAP	2′,4′,6′-Trihydroxyacetophenone
Thr	Threonine
TOF	Time-of-flight
UV	Ultraviolet
VVA	*Vicia villosa* Agglutinin
WGA	Wheat Germ Agglutinin
Xyl	D-Xylose

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Glycolipid Analysis • Glycoprotein Analysis: Using Nuclear Magnetic Resonance • Monosaccharides and Sugar Alcohol Analysis • Proteoglycan and Acidic Polysaccharide Analysis

REFERENCES

1. IUPAC–IUB Joint Commission on Biochemical Nomenclature (JCBN), 'Nomenclature of Glycoproteins, Glycopeptides and Peptidoglycans', *Eur. J. Biochem.*, **159**, 1–6 (1986).
2. J.E. Silbert, M. Bernfeld, R. Kokenyesi, 'Proteoglycans: a Special Class of Glycoproteins', in *Glycoproteins II*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 1–31, Vol. 29b, 1997.
3. R.J. Jackson, S.J. Bush, A.D. Cardin, 'Glycosaminoglycans: Molecular Properties, Protein Interactions, and Role in Physiological Processes', *Physiol. Rev.*, **71**, 481–539 (1993).
4. N.K. Karamanos, A. Hjerpe, 'A Survey of Methodological Challenges for Glycosaminoglycan/Proteoglycan Analysis and Structural Characterization by Capillary Electrophoresis', *Electrophoresis*, **19**, 2561–2571 (1998).
5. T.W. Rademacher, R.B. Parekh, R.A. Dwek, 'Glycobiology', *Ann. Rev. Biochem.*, **57**, 785–838 (1988).
6. H.J. Allen, E.C. Kisailus, *Glycoconjugates. Composition, Structure and Function*, M. Dekker Inc., New York, 1992.
7. A. Kobata, 'Structures and Functions of the Sugar Chains of Glycoproteins', *Eur. J. Biochem.*, **209**, 483–501 (1992).
8. H. Lis, N. Sharon, 'Protein Glycosylation. Structural and Functional Aspects', *Eur. J. Biochem.*, **218**, 1–27 (1993).
9. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, *Glycoproteins, New Comprehensive Biochemistry*, Elsevier, Amsterdam, Vol. 29a, 1995.

10. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, *Glycoproteins and Disease, New Comprehensive Biochemistry*, Elsevier, Amsterdam, Vol. 30, 1996.
11. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, *Glycoproteins II, New Comprehensive Biochemistry*, Elsevier, Amsterdam, Vol. 29b, 1997.
12. C.F. Goochee, M.J. Gramer, D.C. Andersen, J.B. Bahr, J.R. Rasmussen, 'The Oligosaccharides of Glycoproteins: Bioprocess Factors Affecting Oligosaccharide Structure and their Effect on Glycoprotein Properties', *Biotechnology*, **9**, 1347–1355 (1991).
13. J.F.G. Vliegenthart, J. Montreuil, 'Primary Structure of Glycoprotein Glycans', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 13–28, Vol. 29a, 1995.
14. A. Gooley, K.L. Williams, 'Towards Characterizing *O*-Glycans: the Relative Merits of in vivo and in vitro Approaches in Seeking Peptide Motifs Specifying *O*-Glycosylation Sites', *Glycobiology*, **4**, 413–417 (1994).
15. T. De Beer, J.F.G. Vliegenthart, A. Löffler, J. Hofsteenge, 'The Hexopyranosyl Residue that is C-glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase U_s is α-Mannopyranose', *Biochemistry*, **34**, 11 785–11 789 (1995).
16. M.A.J. Ferguson, A.F. Williams, 'Cell Surface Anchoring of Proteins via Glycosylphosphatidylinositol Structures', *Annu. Rev. Biochem.*, **57**, 285–320 (1988).
17. R. Schauer, J.P. Kamerling, 'Chemistry, Biochemistry and Biology of Sialic Acids', in *Glycoproteins II*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 243–402, Vol. 29b, 1997.
18. A. Verbert, 'Biosynthesis from $Glc_3Man_9GlcNAc_2$-Protein to $Man_5GlcNAc_2$-Protein: Transfer 'en bloc' and Processing', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 145–152, Vol. 29a, 1995.
19. J. Montreuil, 'Primary Structure of Glycoprotein Glycans: Basis for the Molecular Biology of Glycoproteins', *Adv. Carbohydr. Chem. Biochem.*, **37**, 157–223 (1980).
20. J. Montreuil, 'Spatial Conformation of Glycans and Glycoproteins', *Biol. Cell*, **51**, 115–131 (1984).
21. P.M. Rudd, R.A. Dwek, 'Glycosylation: Heterogeneity and the 3D Structure of Proteins', *Curr. Opin. Biotechnol.*, **8**, 488–497 (1997).
22. R. Kornfeld, S. Kornfeld, 'Assembly of Asparagine-linked Oligosaccharides', *Annu. Rev. Biochem.*, **54**, 631–664 (1985).
23. H. Schachter, 'Biosynthesis. Glycosyltransferases Involved in the Synthesis of *N*-Glycan Antennae', in *Glycoproteins*, eds. J. Montreuil, H. Schachter, J.F.G. Vliegenthart, Elsevier, Amsterdam, 153–200, Vol. 29a, 1995.
24. J. Montreuil, G. Spik, J. Mazurier, 'Transferrin Superfamily. An Outstanding Model for Studying Biochemical Evolution', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 203–242, Vol. 29b, 1997.
25. P. Stanley, 'Glycosylation Engineering', *Glycobiology*, **2**, 99–107 (1992).
26. M. Sumper, F.T. Wieland, 'Bacterial Glycoproteins', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 455–473, Vol. 29a, 1995.
27. L. März, F. Altmann, E. Staudacher, V. Kubelka, 'Protein Glycosylation in Insects', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 543–564, Vol. 29a, 1995.
28. A. Sturm, '*N*-Glycosylation of Plants Proteins', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 521–541, Vol. 29a, 1995.
29. P. Lerouge, M. Cabanes-Macheteau, C. Rayon, A.C. Fischette-Lainé, V. Gomord, L. Faye, '*N*-Glycoprotein Biosynthesis in Plants: Recent Development and Future Trends', *Plant Mol. Biol.*, **38**, 31–48 (1998).
30. L. Lehle, W. Tanner, 'Protein Glycosylation in Yeast', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 475–509, Vol. 29a, 1995.
31. D. Léger, B. Campion, J.P. Decottignies, J. Montreuil, G. Spik, 'Physiological Significance of the Marked Increased Branching of the Glycans of Human Serotransferrin During Pregnancy', *Biochem. J.*, **257**, 231–238 (1989).
32. A. Mackiewicz, 'Acute Phase Proteins and Transformed Cells', *Int. Rev. Cytol.*, **170**, 225–300 (1997).
33. H.J. Gabius, S. Gabius, *Glycosciences, Status and Perspectives*, Chapman & Hall, Weinheim, 133–162, 1997.
34. A. Varki, 'Biological Roles of Oligosaccharides: All of the Theories are Correct', *Glycobiology*, **3**, 97–130 (1993).
35. R.A. Dwek, 'Glycobiology: Toward Understanding the Function of Sugars', *Biochem. Soc. Trans.*, **23**, 1–25 (1995).
36. M. Fukuda, O. Hingaul, 'Molecular Biology', Oxford University Press, Oxford, 1994.
37. G. Ashwell, J. Harford, 'Carbohydrate-specific Receptor of the Liver', *Annu. Rev. Biochem.*, **51**, 531–554 (1982).
38. N. Sharon, 'Carbohydrate-lectin Interactions in Infectious Disease', *Adv. Exp. Med. Biol.*, **408**, 1–8 (1996).
39. S. Kornfeld, 'Structure and Function of the Mannose 6-Phosphate Insulin-like Growth Factor II Receptors', *Annu. Rev. Biochem.*, **61**, 307–330 (1992).
40. I. Brockhausen, J. Schtuzbach, W. Kuhns, 'Glycoproteins and their Relationship to Human Disease', *Acta Anat.*, **161**, 36–78 (1998).
41. N.R. Thotakura, M.W. Szkudlinski, B.D. Weintraub, 'Structure–Function Studies of Oligosaccharides of Recombinant Human Thyrotrophin by Sequential

Deglycosylation and Resialylation', *Glycobiology*, **4**, 525–533 (1994).
42. A. Pierce-Cretel, J.P. Decottignies, G. Strecker, J. Montreuil, G. Spik, 'Primary Structure of 23 Neutral and Monosialylated *O*-Glycosidically Linked Glycans to the sIgA Hinge Region Determined by a Combination of Permethylated Analysis and 400 MHz ^{1}H-NMR Spectroscopy', *Eur. J. Biochem.*, **182**, 457–476 (1989).
43. H. Krotkiewski, M. Duk, M. Syper, H. Lis, N. Sharon, E. Lisowska, 'Blood Group MN-dependent in Degree of Galactosylation of *O*-Glycans of Glycophorin A is Restricted to the GalNAc Residues Located on Amino Acid Residues 2–4 of the Glycophorin Polypeptide Chain', *FEBS Lett.*, **406**, 296–300 (1997).
44. P. Roussel, G. Lamblin, 'Human Mucosal Mucins in Diseases', in *Glycoproteins and Disease*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 351–383, Vol. 30, 1996.
45. T. Sasaki, H. Yamada, K. Matsumura, T. Shimizu, A. Kobata, T. Endo, 'Detection of *O*-Mannosyl Glycans in Rabbit Skeletal Muscle α-Dystroglycan', *Biochim. Biophys. Acta*, **1425**, 599–606 (1998).
46. G. Strecker, 'Amphibian Glycoproteins', in *Glycoproteins II*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 163–172, 1997.
47. F. Piller, V. Piller, R.I. Fox, M. Fukuda, 'Human T-Lymphocyte Activation is Associated with Changes in *O*-Glycan Biosynthesis', *J. Biol. Chem.*, **263**, 15 146–15 150 (1988).
48. J.M. Lo-Guidice, J.M. Wieruszeski, J. Lemoine, A. Verbert, P. Roussel, G. Lamblin, 'Sialylation and Sulfation of the Carbohydrate Chains in Respiratory Mucins from a Patient with Cystic Fibrosis', *J. Biol. Chem.*, **269**, 18 794–18 813 (1994).
49. P. Van den Steen, P.M. Rudd, R.A. Dwek, G. Opdenakher, 'Concepts and Principles of O-linked Glycosylation', *Crit. Rev. Biochem. Mol. Biol.*, **33**, 151–208 (1998).
50. I. Brockhausen, 'Biosynthesis of *O*-Glycans of the *N*-Acetylgalactosamine-α-Ser/Thr Linkage Type', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 201–259, Vol. 29a, 1995.
51. E.F. Hounsell, M.J. Davies, D.V. Renouf, 'O-Linked Protein Glycosylation Structure and Function', *Glycoconj. J.*, **13**, 19–25 (1996).
52. W.M. Watkins, 'Biosynthesis. Molecular Basis of Antigenic Specificity in the ABO, H and Lewis Blood-group Systems', in *Glycoproteins*, eds. J. Montreuil, J.F.G. Vliegenthart, Elsevier, Amsterdam, 313–376, Vol. 29a, 1995.
53. M.W. Nicholson, A.N. Barclay, M.S. Singer, S.D. Rosen, P.A. van der Merve, 'Affinity and Kinetic Analysis of L-Selectin (CD62L) Binding to Glycosylation-dependent Cell – Adhesion Molecule-1', *J. Biol. Chem.*, **273**, 763–770 (1998).
54. J.G. Geng, T.J. Raub, C.A. Baker, G.A. Sawada, L. Ma, A.P. Elhammer, 'Expression of a P-Selectin Ligand in Zona Pellucida of Porcine Oocytes and P-Selectin on Acrosomal Membrane of Porcine Sperm Cells. Potential Implications for their Involvement in Sperm–Egg Interactions', *J. Cell Biol.*, **137**, 743–754 (1997).
55. S.I. Hakomori, 'Tumor-associated Carbohydrate Antigens and Modified Blood Group Antigens', in *Glycoproteins and Disease*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 243–276, Vol. 30, 1996.
56. R.N. Cole, G.W. Hart, 'Glycosyl-Phosphatidylinositol Anchors: Structure, Biosynthesis and Function', in *Glycoproteins II*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 69–88, Vol. 29b, 1997.
57. D.J. Thornton, M. Howard, N. Khan, J.K. Sheehan, 'Identification of Two Glycoforms of the MUC5B Mucin in Human Respiratory Mucus – Evidence for a Cysteine-rich Sequence Repeated within the Molecule', *J. Biol. Chem.*, **272**, 9561–9566 (1997).
58. D.J. Thornton, N. Khan, R. Mehotra, M. Howard, E. Veerman, N.H. Packer, J.K. Sheehan, 'Salivary Mucin MG1 is Comprised Almost Entirely of Different Glycosylated Forms of the MUC5B Gene Product', *Glycobiology*, **9**, 293–302 (1999).
59. C. Wickström, J. Davies, G. Eriksen, E.C.I. Veerman, I. Carlstedt, 'MUC5B is the Major Gel-forming, Oligomeric Mucin from Salivary Gland, Respiratory Tract and Endocervix: Identification of Glycoforms and C-Terminal Cleavage', *Biochem. J.*, **334**, 685–693 (1998).
60. S.C. March, I. Parikh, P. Cuatrecasas, 'A Simplified Method for Cyanogen Bromide Activation of Agarose for Affinity Chromatography', *Anal. Biochem.*, **60**, 149–152 (1974).
61. J. Kohn, M. Wilcheck, 'A New Approach (Cyano-Transfer) for Cyanogen Bromide Activation Resins, Free of Interferring Mitogen Derivatives', *Biochem. Biophys. Res. Commun.*, **107**, 878–884 (1982).
62. M. Wilcheck, T. Miron, 'Stable, High Capacity and Non Charged Agarose Derivatives for Immobilization of Biologically Active Compounds and for Affinity Chromatography', *Mol. Cell. Biochem.*, **4**, 181–187 (1974).
63. M. Wilcheck, T. Miron, 'Limitations of *N*-Hydroxysuccinimide Esters in Affinity Chromatography and Protein Immobilization', *Biochemistry*, **26**, 2155–2161 (1987).
64. J. Turkova, 'Oriented Immobilization of Biologically Active Proteins as a Tool for Revealing Protein Interactions and Function', *J. Chromatogr., B*, **722**, 11–31 (1999).
65. G.W. Jack, 'Immunoaffinity Chromatography', *Mol. Biotechnol.*, **1**, 59–86 (1994).

66. D.S. Hage, 'Survey of Recent Advances in Analytical Applications of Immunoaffinity Chromatography', *J. Chromatogr., B*, **715**, 3–28 (1998).
67. A. Gokana, J.J. Winchenne, A. Ben-Ghanem, A. Ahaded, J.P. Cartron, P. Lambin, 'Chromatographic Separation of Recombinant Human Erythropoietin Isoforms', *J. Chromatogr., A*, **791**, 109–118 (1997).
68. D.J. Thornton, I. Carlstedt, M. Howard, P.L. Devine, M.R. Price, J.K. Sheehan, 'Respiratory Mucins: Identification of Core Proteins and Glycoforms', *Biochem. J.*, **316**, 967–975 (1996).
69. P.M. Jacquinot, D. Léger, J.M. Wieruszeski, B. Coddeville, J. Montreuil, G. Spik, 'Change in Glycosylation of Chicken Transferrin Glycans Biosynthesized During Embryogenesis and Primary Culture of Embryo Hepatocytes', *Glycobiology*, **4**, 617–624 (1994).
70. I.J. Goldstein, R.C. Hughes, M. Monsigny, T. Osawa, N. Sharon, 'What Should be Called a Lectin?', *Nature*, **285**, 66 (1980).
71. H. Lis, N. Sharon, 'Lectins: Carbohydrate-specific Proteins that Mediate Cellular Recognition', *Chem. Rev.*, **98**, 637–674 (1998).
72. R. Loris, T. Hamelryck, J. Bouckaert, L. Wyns, 'Legume Lectin Structure', *Biochim. Biophys. Acta*, **1383**, 9–36 (1998).
73. I.J. Goldstein, H.C. Winter, R.D. Poretz, 'Plant Lectins: Tools for The Study of Complex Carbohydrates', in *Glycoproteins II*, eds. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, Elsevier, Amsterdam, 403–474, 1997.
74. H. Rudiger, P. Rougé, 'Structure and Function of Plant Lectins', *Carbohydr. Eur.*, **23**, 18–22 (1998).
75. W.J. Peumans, E.J.M. Van Damme, 'Plant Lectins: Specific Tools for the Identification, Isolation and Characterization of O-Linked Glycans', *Crit. Rev. Biochem. Mol. Biol.*, **33**, 209–258 (1998).
76. E.J.M. Van Damme, W.J. Peumans, A. Barre, P. Rougé, 'Plant Lectins: a Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles', *Crit. Rev. Plant. Sci.*, **17**, 575–692 (1998).
77. J.M. Rini, 'Lectin Structure', *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 551–577 (1995).
78. W.I. Weis, K. Drickamer, 'Structural Basis of Lectin–Carbohydrate Recognition', *Annu. Rev. Biochem.*, **65**, 441–473 (1996).
79. R. Kornfeld, C. Ferris, 'Interactions of Immunoglobulin Glycopeptides with Concanavalin A', *J. Biol. Chem.*, **250**, 2614–2619 (1975).
80. H. Debray, D. Decout, G. Strecker, G. Spik, J. Montreuil, 'Specificity of Twelve Lectins Toward Oligosaccharides and Glycopeptides Related to *N*-Glycosylproteins', *Eur. J. Biochem.*, **117**, 41–55 (1981).
81. Y. Bourne, J. Mazurier, D. Legrand, P. Rougé, J. Montreuil, G. Spik, C. Cambillau, 'Structures of a Legume Lectin Complexed with the Human Lactoferrin N2 Fragment, and with an Isolated Biantennary Glycopeptide: Role of the Fucose Moiety', *Structure*, **2**, 209–219 (1994).
82. W.I. Weis, 'Cell Surface Carbohydrate Recognition by Animal and Viral Lectins', *Curr. Opin. Struct. Biol.*, **7**, 624–630 (1997).
83. A.M. Wu, S.J. Sugii, A. Herp, 'A Guide for Carbohydrate Specificities of Lectins', *Adv. Exp. Med. Biol.*, **228**, 819–847 (1988).
84. J.U. Baenziger, D. Fiete, 'Structural Determinants of Concanavalin A Specificity for Oligosaccharides', *J. Biol. Chem.*, **254**, 2400–2407 (1979).
85. H. Debray, J. Montreuil, 'Structural Basis for the Affinity of Four Insolubilized Lectins, with a Specificity for α-D-Mannose, Towards Various Glycopeptides with the *N*-Glycosylamine Linkage and Related Oligosaccharides', *J. Biosci.*, **5**(suppl. 1), 93–100 (1983).
86. T.K. Dam, S. Oscarson, C.F. Brewer, 'Thermodynamics of Binding of the Core Trimannoside of Asparagine-linked Carbohydrates and Deoxy Analogs to *Dioclea grandiflora* Lectin', *J. Biol. Chem.*, **273**, 32 812–32 817 (1998).
87. Y. Shinohara, H. Sota, F. Kim, M. Shimizu, M. Gotoh, M. Tosu, Y. Hasegawa, 'Use of a Biosensor Based on Surface Plasmon Resonance and Biotinyl Glycans for Analysis of Sugar Binding Specificities of Lectins', *J. Biochem.*, **117**, 1076–1082 (1995).
88. K. Kornfeld, M.L. Reitman, R. Kornfeld, 'The Carbohydrate-binding Specificity of Pea and Lentil Lectins. Fucose is an Important Determinant', *J. Biol. Chem.*, **256**, 6633–6640 (1981).
89. J. Montreuil, 'Glycoprotein Structure and Conformation: an Overview', in *Methods on Glycoconjugates, a Laboratory Manual*, ed. A. Verbert, Harwood Academic, Switzerland, 1–23, 1995.
90. E.D. Green, R.M. Brodbeck, J.U. Baenziger, 'Lectin Affinity High-performance Liquid Chromatography. Interactions of *N*-Glycanase-released Oligosaccharides with *Ricinus communis* Agglutinin I and *Ricinus communis* Agglutinin II', *J. Biol. Chem.*, **262**, 12 030–12 039 (1987).
91. E.D. Green, J.U. Baenziger, 'Oligosaccharide Specificities of *Phaseolus vulgaris* Leukoagglutinating and Erythroagglutinating Phytohemagglutinins. Interactions with *N*-Glycanase®-released Oligosaccharides', *J. Biol. Chem.*, **262**, 12 018–12 029 (1987).
92. R.D. Cummings, S. Kornfeld, 'Characterization of the Structural Determinants Required for the High Affinity Interaction of Asparagine-linked Oligosaccharides with Immobilized *Phaseolus vulgaris* Leukoagglutinating and Erythroagglutinating Lectins', *J. Biol. Chem.*, **257**, 11 230–11 234 (1982).
93. J.T. Gallagher, 'Carbohydrate-binding Properties of Lectins: a Possible Approach to Lectin Nomenclature and Classification', *Biosci. Rep.*, **4**, 621–632 (1984).

94. J.T. Gallagher, 'Affinity Chromatography of Complex Carbohydrates Using Lectins', in *HPLC of Macromolecules, a Practical Approach*, ed. R.W.A. Oliver, IRL Press, Oxford, 209–228, 1989.
95. R.K. Merkle, R.D. Cummings, 'Lectin Affinity Chromatography of Glycopeptides', *Methods Enzymol.*, **138**, 232–259 (1987).
96. T. Osawa, T. Tsuji, 'Fractionation and Structural Assessment of Oligosaccharides and Glycopeptides by Use of Immobilized Lectins', *Annu. Rev. Biochem.*, **56**, 21–42 (1987).
97. H. Debray, J. Montreuil, 'Lectin Affinity Chromatography of Glycoconjugates', *Adv. Lectin. Res.*, **4**, 51–96 (1991).
98. H. Debray, J. Montreuil, 'Specificity of Lectins toward Oligosaccharidic Sequences Belonging to *N*- and *O*-Glycosylproteins', in *Affinity Electrophoresis: Principles & Application*, eds. J. Breborowicz, A. Mackiewicz, CRC Press, Boca Raton, FL, 23–57, 1992.
99. T. Tsuji, K. Yamamoto, T. Osawa, 'Affinity Chromatography of Oligosaccharides and Glycopeptides with Immobilized Lectins', in *Molecular Interactions in Bioseparations*, ed. T.T. Ngo, Plenum Press, New York, 113–126, 1993.
100. K. Yamamoto, T. Tsuji, T. Osawa, 'Analysis of Asparagine-linked Oligosaccharides by Sequential Lectin-affinity Chromatography', in *Glycoprotein Analysis in Biomedicine, Methods in Molecular Biology*, ed. E.F. Hounsell, Humana Press, Totowa, NJ, 17–34, Vol. 14, 1993.
101. A. Kobata, K. Yamashita, 'Fractionation of Oligosaccharides by Serial Affinity Chromatography with Use of Immobilized Lectin Columns', in *Glycobiology: A Practical Approach*, eds. M. Fukuda, A. Kobata, IRL Press, 103–125, 1993.
102. R.D. Cummings, 'Use of Lectins in Analysis of Glycoconjugates', *Methods Enzymol.*, **230**, 66–86 (1994).
103. R.D. Cummings, 'Lectins as Tools for Glycoconjugate Purification and Characterization', in *Glycosciences*, eds. H.J. Gabius, S. Gabius, Chapman & Hall, New York, 191–199, 1997.
104. W.M. Lakhtin, 'Use of Lectins in the Analysis of Carbohydrate Moieties of Glycoproteins and Other Natural Glycoconjugates', *Biochemistry (Moscow)*, **60**, 131–153 (1995).
105. H. Debray, 'Use of Lectins', in *Methods on Glycoconjugates. A Laboratory Manual*, ed. A. Verbert, Harwood Academic, Switzerland, 103–133, 1995.
106. E.D. Green, R.M. Brodbeck, J.U. Baenziger, 'Lectin Affinity High-performance Liquid Chromatography: Interactions of *N*-Glycanase-released Oligosaccharides with Leukoagglutinating Phytohemagglutinin, Concanavalin A, *Datura stramonium* Agglutinin and *Vicia villosa* Agglutinin', *Anal. Biochem.*, **167**, 62–75 (1987).
107. H. Harada, M. Kamei, Y. Tokumoto, S. Yui, F. Koyama, N. Kochibe, T. Endo, A. Kobata, 'Systematic Fractionation of Oligosaccharides of Human Immunoglobulin G by Serial Affinity Chromatography on Immobilized Lectin Columns', *Anal. Biochem.*, **164**, 374–381 (1987).
108. J.T. Dulaney, 'Binding Interactions of Glycoproteins with Lectins', *Mol. Cell. Biochem.*, **21**, 43–63 (1979).
109. R. Lotan, G.L. Nicolson, 'Purification of Cell Membrane Glycoproteins by Lectin Affinity Chromatography', *Biochim. Biophys. Acta*, **559**, 329–376 (1979).
110. J.A. Hedo, 'Lectins as Tools for the Purification of Membrane Receptors', in *Receptor Purification Procedures*, eds. C. Venter, L. Harrison, Liss, New York, 45–60, Vol. 2, 1984.
111. H. Lis, N. Sharon, 'Lectins: Properties and Applications to the Study of Complex Carbohydrates in Solution and on Cell Surfaces', in *Biology of Carbohydrates*, eds. V. Ginsburg, P.W. Robbins, J. Wiley & Sons, New York, 2–85, Vol. 2, 1984.
112. H. Lis, N. Sharon, 'Applications of Lectins', in *The Lectins, Properties, Functions and Applications in Biology and Medicine*, eds. I.E. Liener, N. Sharon, I.J. Goldstein, Academic Press, New York, 265–291, 1986.
113. H. Lis, N. Sharon, 'Lectins as Molecules and as Tools', *Annu. Rev. Biochem.*, **55**, 35–67 (1986).
114. B. Bayard, J.P. Kerckaert, 'Uniformity of Carbohydrate Chains within Molecular Variants of Rat α_1-Fetoprotein with Distinct Affinity for Concanavalin A', *Eur. J. Biochem.*, **113**, 405–414 (1981).
115. R. Lotan, G. Beattie, W. Hubbell, G.L. Nicolson, 'Activities of Lectins and their Immobilized Derivatives in Detergent Solutions. Implications on the Use of Lectin Affinity Chromatography for the Purification of Membrane Glycoproteins', *Biochemistry*, **16**, 1787–1794 (1977).
116. I. Kahane, H. Furthmayr, V.T. Marchesi, 'Isolation of Membrane Glycoproteins by Affinity Chromatography in the Presence of Detergents', *Biochim. Biophys. Acta*, **426**, 464–476 (1976).
117. J.A. Hedo, L.C. Harrison, J. Roth, 'Binding of Insulin Receptors to Lectins: Evidence for Common Carbohydrate Determinants on Several Membrane Receptors', *Biochemistry*, **20**, 3385–3393 (1981).
118. S. Shibata, B.P. Peters, D.D. Roberts, I.J. Goldstein, L.A. Liotta, 'Isolation of Laminin by Affinity Chromatography on Immobilized *Griffonia simplicifolia* I Lectin', *FEBS Lett.*, **142**, 194–198 (1982).
119. M.C. Roque-Barreira, A. Campos-Neto, 'Jacalin: an IgA-binding Lectin', *J. Immunol.*, **134**, 1740–1743 (1985).
120. H. Kondoh, K. Kobayashi, K. Hagiwara, 'A Simple Procedure for the Isolation of Human Secretory IgA of IgA_1 and IgA_2 Subclass by a Jackfruit Lectin, Jacalin, Affinity Chromatography', *Mol. Immun.*, **24**, 1219–1222 (1987).

121. P. Aucouturier, E. Mihaesco, C. Mihaesco, J.L. Preud'homme, 'Characterization of Jacalin, the Human IgA and IgD Binding Lectin from Jackfruit', *Mol. Immunol.*, **24**, 503–511 (1987).
122. B.D. Zehr, S.D. Litwin, 'Human IgD and IgA_1 Compete for D-Galactose-related Binding Sites on the Lectin Jacalin', *Scand. J. Immunol.*, **26**, 229–236 (1987).
123. N. Shibuya, J.E. Berry, I.J. Goldstein, 'One-step Purification of Murine IgM and Human α_2-Macroglobulin by Affinity Chromatography on Immobilized Snowdrop Bulb Lectin', *Arch. Biochem. Biophys.*, **267**, 676–680 (1988).
124. M.F. Bierhuizen, M. De Wit, C.A. Govers, W. Ferwerda, C. Koeleman, O. Pos, W. Van Dijk, 'Glycosylation of Three Molecular Forms of Human α_1-Acid Glycoprotein Having Different Interactions with Concanavalin A. Variations in the Occurrence of Di-, Tri- and Tetraantennary Glycans and the Degree of Sialylation', *Eur. J. Biochem.*, **175**, 387–394 (1988).
125. E.A. Bayer, H. Ben-Hur, M. Wilchek, 'Enzyme-based Detection of Glycoproteins on Blot Transfers Using Avidin–Biotin Technology', *Anal. Biochem.*, **161**, 123–161 (1987).
126. D.J. O'Shannessy, P.J. Voorstad, R.H. Quarles, 'Quantitation of Glycoproteins on Electroblots Using the Biotin–Streptavidin Complex', *Anal. Biochem.*, **163**, 204–209 (1987).
127. A. Haselbeck, E. Schickaneder, H. Von der Eltz, W. Hösel, 'Structural Characterization of Glycoproteins Carbohydrate Chains by Using Digoxygenin-labeled Lectins on Blots', *Anal. Biochem.*, **191**, 25–30 (1990).
128. A. Haselbeck, W. Hösel, 'Immunological Detection of Glycoproteins on Blots Based on Labeling with Digoxygenin', in *Glycoprotein Analysis in Biomedicine*, ed. E.F. Hounsell, Humana Press, Totowa, NJ, 161–173, Vol. 14, 1993.
129. U.K. Laemmli, 'Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4', *Nature*, **227**, 680–685 (1970).
130. P.H. O'Farrell, 'High Resolution Two-dimensional Electrophoresis of Protein', *J. Biol. Chem.*, **250**, 4007–4021 (1975).
131. K. Burridge, 'Direct Identification of Specific Glycoproteins and Antigens in Sodium Dodecyl Sulfate Gels', *Methods Enzymol.*, **28**, 54–64 (1978).
132. J.G. Wood, F.O. Sarinana, 'The Staining of Sciatic Nerve Glycoproteins on Polyacrylamide Gels with Concanavalin A Peroxidase', *Anal. Biochem.*, **69**, 320–322 (1975).
133. J. Renart, J. Reiser, G.R. Stark, 'Transfer of Proteins from Gels to Diazobenzyloxymethyl-paper and Detection with Antisera: a Method for Studying Antibody Specificity and Antigen Structure', *Proc. Natl. Acad. Sci. USA*, **76**, 3116–3120 (1979).
134. H. Towbin, T. Staehelin, J. Gordon, 'Electrophoretic Transfer of Proteins from Polyacrylamide Gels on Nitrocellulose Sheets: Procedure and some Applications', *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354 (1979).
135. J.M. Gershoni, 'Protein Blotting: Developments and Perspectives', *Trends Biochem. Sci.*, **10**, 103–106 (1985).
136. U. Beisiegel, 'Protein Blotting', *Electrophoresis*, **7**, 1–18 (1986).
137. J.R. Bartles, A.L. Hubbard, '^{125}I-Wheat Germ Agglutinin Blotting: Increased Sensitivity with Polyvinylpyrolidone Quenching and Periodate Oxidation/Reductive Phenylamination', *Anal. Biochem.*, **140**, 284–292 (1984).
138. A. Haselbeck, E. Schickaneder, H. Von Der Eltz, W. Hösel, 'Structural Characterization of Glycoprotein Carbohydrate Chains by Using Digoxigenin-labelled Lectins on Blots', *Anal. Biochem.*, **191**, 25–30 (1990).
139. T.C. Bog-Hansen, 'Crossed Immuno-affinoelectrophoresis. An Analytical Method to Predict the Result of Affinity Chromatography', *Anal. Biochem.*, **56**, 480–488 (1973).
140. O.J. Bjerrum, T.C. Bog-Hansen, 'The Immunochemical Approach to the Characterization of Membrane Proteins. Human Erythrocyte Membrane Proteins Analyzed as a Model System', *Biochim. Biophys. Acta*, **455**, 66–89 (1976).
141. I. Hagen, O.J. Bjerrum, N.O. Solum, 'Characterization of Human Platelet Proteins Solubilized with Triton X-100 and Examined by Crossed Immunoelectrophoresis', *Eur. J. Biochem.*, **99**, 9–22 (1979).
142. P.M.H. Heegaard, N.H.H. Heegaard, T.C. Bog-Hansen, 'Affinity Electrophoresis for the Characterization of Glycoproteins. The Use of Lectins in Combination with Immunoelectrophoresis', in *Affinity Electrophoresis: Principles and Application*, eds. J. Breborowicz, A. Mackiewicz, CRC Press, Boca Raton, FL, 3–21, 1992.
143. K. Tanaka, W. Pigman, 'Improvements of Hydrogenation Procedures for Demonstration of O-Threonine Glycosidic Linkages in Bovine Submaxillary Mucin', *J. Biol. Chem.*, **240**, 1487–1489 (1965).
144. D.M. Carlson, 'Structures and Immunological Properties of Oligosaccharides Isolated from Pig Submaxillary Mucins', *J. Biol. Chem.*, **243**, 616–626 (1968).
145. B. Bayard, D. Roux, 'Hydrazinolysis and Nitrous Deamination of Glycoproteins. Evidence for a Common Inner Core in Carbohydrate Moiety', *FEBS Lett.*, **55**, 206–211 (1975).
146. S. Takasaki, T. Mizuochi, A. Kobata, 'Hydrazinolysis of Asparagine-linked Sugars Chains to Produce Free Oligosaccharides', *Methods Enzymol.*, **83**, 263–268 (1982).
147. B. Bendiak, D.A. Cumming, 'Purification of Oligosaccharides having a Free Reducing-end from Glycopeptide Sources', *Carbohydr. Res.*, **151**, 89–103 (1986).
148. J.C. Michalski, J. Peter-Katalinic, H. Egge, J. Paz-Parente, J. Montreuil, G. Strecker, 'Behavior of the 2-Acetamido-2-Deoxy-α-D-Glucopyranosyl Residue

during Sequential Hydrazinolysis, *N*-Reacetylation, Reduction and Methylation of Glycoasparagines', *Carbohydr. Res.*, **134**, 177–189 (1984).
149. T. Patel, J. Bruce, A. Merry, C. Bigge, M. Wormald, M. Jaques, R. Parekh, 'Use of Hydrazine to Release in Intact and Unreduced Form Both N- and O-linked Oligosaccharides from Glycoproteins', *Biochemistry*, **32**, 679–693 (1993).
150. S. Bouquelet, 'Enzymatic Cleavage', in *Methods on Glycoconjugates. A Laboratory Manual*, ed. A. Verbert, Harwood Academic, Switzerland, 51–59, 1995.
151. A.L. Tarentino, F. Maley, 'Purification and Properties of an *Endo*-β-*N*-Acetylglucosaminidase from the *Streptomyces griseus*', *J. Biol. Chem.*, **249**, 811–817 (1974).
152. T.H. Plummer, A.L. Tarentino, 'Purification of the Oligosaccharide Cleaving Enzymes of *Flavobacterium meningosepticum*', *Glycobiology*, **1**, 257–263 (1991).
153. O. Kol, D. Brassart, G. Spik, J. Montreuil, S. Bouquelet, 'Specificity towards Oligomannoside and Hybrid Type Glycans of an *Endo*-β-*N*-Acetylglucosaminidase from the Basidiomycete *Sporotrichum dimorphosporum*', *Glycoconj. J.*, **6**, 333–348 (1989).
154. D. Lemp, A. Haselbeck, F. Klebl, 'Molecular Cloning and Heterologous Expression of *N*-Glycosidase F from *Flavobacterium meningosepticum*', *J. Biol. Chem.*, **265**, 15 606–15 610 (1990).
155. N. Takahashi, N. Nishibe, 'Some Characteristics of a New Glycopeptidase Acting on Aspartylglycosylamine Linkages', *J. Biochem.*, **84**, 1467–1473 (1978).
156. J. Finne, T. Krusius, J. Jarnefelt, 'Fractionation of Glycopeptides', in *27th Int. Congress of Pure and Applied Chemistry*, ed. A. Varmavuory, Pergamon Press, New York, 147–159, 1980.
157. J. Montreuil, S. Bouquelet, H. Debray, J. Lemoine, J.C. Michalski, G. Spik, G. Strecker, 'Glycoproteins', in *Carbohydrate Analysis: A Practical Approach*, eds. M.F. Chaplin, J.F. Kennedy, IRL Press, Oxford, 181–293, 1994.
158. R.D. Cummings, R.K. Merkle, N.L. Stults, 'Separation and Analysis of Glycoprotein Oligosaccharides', *Methods Cell. Biol.*, **32**, 141–183 (1989).
159. A. Kobata, T. Endo, 'Immobilized Lectin Columns: Useful Tools for the Fractionation and Structural Analysis of Oligosaccharides', *J. Chromatogr.*, **597**, 111–122 (1992).
160. N. Koide, T. Muramatsu, '*Endo*-β-D-Acetylglucosaminidase Acting on Carbohydrate Moieties of Glycoproteins. Purification and Properties of the Enzyme from *Diplococcus pneumonia*', *J. Biol. Chem.*, **249**, 4897–4904 (1974).
161. S. Takasaki, A. Kobata, 'Microdetermination of Individual Neutral and Amino Sugars and *N*-Acetylneuraminic Acid in Complex Saccharides', *J. Biochem. (Tokyo)*, **76**, 783–789 (1974).
162. K. Yamamoto, T. Tsuji, I. Matsumoto, T. Osawa, 'Structural Requirements for the Binding of Oligosaccharides and Glycopeptides to Immobilized Wheat Germ Agglutinin', *Biochemistry*, **20**, 5894–5899 (1981).
163. H. Debray, A. Pierce-Cretel, G. Spik, J. Montreuil, 'Affinity of Ten Insolubilized Lectins towards Various Glycopeptides with the *N*-Glycosylamine Linkage and Related Oligosaccharides', in *Lectins: Biology, Biochemistry, Clinical Biochemistry*, eds. T.C. Bog-Hansen, G.A. Spengler, De Gruyter, Berlin, 335–350, Vol. 3, 1983.
164. J.T. Gallagher, A. Morris, T.M. Dexter, 'Identification of Two Binding Sites for Wheat Germ Agglutinin on Polylactosamine-type Oligosaccharides', *Biochem. J.*, **231**, 115–122 (1985).
165. R.J. Ivatt, P.B. Harnett, J.W. Reeder, 'Isolated Erythroglycans have a High-affinity Interaction with Wheat Germ Agglutinin but are Poorly Accessible *in situ*', *Biochim. Biophys. Acta*, **884**, 124–134 (1986).
166. R.J. Ivatt, J.W. Reeder, G.F. Clark, 'Structural and Conformational Features that Affect the Interaction of Polylactosamino-glycans with Immobilized Wheat Germ Agglutinin', *Biochim. Biophys. Acta*, **883**, 253–264 (1986).
167. V.P. Bhavanandan, J. Umemoto, J.R. Banks, E.A. Davidson, 'Isolation and Partial Characterization of Sialoglycopeptides Produced by a Murine Melanoma', *Biochemistry*, **16**, 4426–4437 (1977).
168. K. Furukawa, J.E. Minor, J.D. Hegarty, V.P. Bhavanandan, 'Interaction of Sialoglycoproteins with Wheat Germ Agglutinin-Sepharose of Varying Ratio of Lectins to Sepharose. Use for the Purification of Mucin Glycoproteins from Membrane Extracts', *J. Biol. Chem.*, **261**, 7755–7761 (1986).
169. T. Irimura, G.L. Nicolson, 'Interaction of Pokeweed Mitogen with Poly-(*N*-Acetyllactosamine)-type Carbohydrate Chains', *Carbohydr. Res.*, **120**, 187–195 (1983).
170. H. Kawashima, S. Sueyoshi, H. Li, K. Yamamoto, T. Osawa, 'Carbohydrate Binding Specificities of Several Poly-*N*-Acetyllactosamine-binding Lectins', *Glycoconj. J.*, **7**, 323–334 (1990).
171. K. Yamashita, K. Totani, T. Ohkura, S. Takasaki, I.J. Goldstein, A. Kobata, 'Carbohydrate Binding Properties of Complex-type Oligosaccharides on Immobilized *Datura stramonium* Lectin', *J. Biol. Chem.*, **262**, 1602–1607 (1987).
172. H. Debray, J. Montreuil, H. Lis, N. Sharon, 'Affinity of Four Immobilized *Erythrina* Lectins Toward Various N-linked Glycopeptides and Related Oligosaccharides', *Carbohydr. Res.*, **151**, 359–370 (1986).
173. N. Shibuya, I.J. Goldstein, W.F. Broekaert, M. Nsimba-Lubaki, B. Peeters, W.J. Peumans, 'Fractionation of Sialylated Oligosaccharides, Glycopeptides and Glycoproteins on Immobilized Elderberry (*Sambucus nigra* L.) Bark Lectin', *Arch. Biochem. Biophys.*, **254**, 1–8 (1987).
174. W.C. Wang, R.D. Cummings, 'The Immobilized Leukoagglutinin from the Seeds of *Maackia amurensis* Binds with High Affinity to Complex-type Asn-Linked

Oligosaccharides Containing Terminal Sialic Acid-linked α-2,3 to Penultimate Galactose Residues', *J. Biol. Chem.*, **263**, 4576–4585 (1988).
175. N. Shibuya, I.J. Goldstein, E.J.M. Van Damme, W.J. Peumans, 'Binding Properties of a Mannose-specific Lectin from the Snowdrop (*Galanthus nivalis*) Bulb', *J. Biol. Chem.*, **263**, 728–734 (1988).
176. H. Kaku, E.J.M. Van Damme, W.J. Peumans, I.J. Goldstein, 'Carbohydrate-binding Specificity of the Daffodil (*Narcissus pseudonarcissus*) and Amaryllis (*Hippeastrum hyb.*) Bulb Lectins', *Arch. Biochem. Biophys.*, **279**, 298–304 (1990).
177. T. Animashaun, R.C. Hughes, '*Bowringia milbraedii* Agglutinin: Specificity of Binding to Early Processing Intermediates of Asparagine-linked Oligosaccharide and Use as a Marker of Endoplasmic Reticulum Glycoproteins', *J. Biol. Chem.*, **264**, 4657–4663 (1989).
178. L.A. Murphy, I.J. Goldstein, 'Five α-D-Galactopyranosyl-binding Isolectins from *Bandeiraea simplicifolia* Seeds', *J. Biol. Chem.*, **252**, 4739–4742 (1977).
179. R.G. Spiro, V.D. Bhoyroo, 'Occurrence of α-D-Galactosyl Residues in the Thyroglobulins from Several Species. Localization in the Saccharide Chains of the Complex Carbohydrate Units', *J. Biol. Chem.*, **259**, 9858–9866 (1984).
180. J. Finne, T. Krusius, H. Rauvala, 'Alkali-stable Blood Group A- and B-active Poly(glycosyl)-peptides from Human Erythrocyte Membrane', *FEBS Lett.*, **89**, 111–115 (1978).
181. K. Nyame, D.F. Smith, R.T. Damian, R.D. Cummings, 'Complex-type Asparagine-linked Oligosaccharides in Glycoproteins Synthesized by *Schistosoma mansoni* Adult Males Contain Terminal β-Linked *N*-Acetylgalactosamine', *J. Biol. Chem.*, **264**, 3235–3243 (1989).
182. S. Ebisu, P.N. Iyer, I.J. Goldstein, 'Equilibrium Dialysis and Carbohydrate-binding Studies on the 2-Acetamido-2-deoxy-D-Glucopyranosyl Binding Lectin from *Bandeiraea simplicifolia* Seeds', *Carbohydr. Res.*, **61**, 129–138 (1978).
183. T. Endo, H. Ohbayashi, K. Kanazawa, N. Kochibe, A. Kobata, 'Carbohydrate Binding Specificity of Immobilized *Psathyrella velutina* Lectin', *J. Biol. Chem.*, **267**, 707–713 (1992).
184. T. Srivatsan, D.F. Smith, R.D. Cummings, 'The Human Blood Fluke *Schistosoma mansoni* Synthesizes Glycoproteins Containing the Lewis X Antigen', *J. Biol. Chem.*, **267**, 20 196–20 203 (1992).
185. K. Yamashita, N. Kochibe, T. Ohkura, I. Ueda, A. Kobata, 'Fractionation of L-Fucose-containing Oligosaccharides on Immobilized *Aleuria aurantia* Lectin', *J. Biol. Chem.*, **260**, 4688–4693 (1985).
186. H. Debray, J. Montreuil, '*Aleuria aurantia* Agglutinin: a New Isolation Procedure and Further Study of its Specificity Towards Various Glycopeptides and Oligosaccharides', *Carbohydr. Res.*, **185**, 15–26 (1989).
187. A.M. Wu, 'Differential Binding Characteristics and Applications of D-Galβ1, 3D-GalNAc Specific Lectins', *Mol. Cell. Biochem.*, **61**, 131–141 (1984).
188. S. Sueyoshi, T. Tsuji, T. Osawa, 'Carbohydrate-binding Specificities of Five Lectins that Bind to *O*-Glycosyl-linked Carbohydrate Chains. Quantitative Analysis by Frontal-affinity Chromatography', *Carbohydr. Res.*, **178**, 213–224 (1988).
189. M.V.K. Sastry, P. Banarjee, S.R. Patanjali, M.J. Swamy, G.V. Swarnalatha, A. Surolia, 'Analysis of Saccharide Binding to *Artocarpus integrifolia* Lectin Reveals Specific Recognition of T-Antigen (β-D-Gal1 → 3D-GalNAc)', *J. Biol. Chem.*, **261**, 11 726–11 733 (1986).
190. G.L. Hortin, B.L. Trimpe, 'Lectin Affinity Chromatography of Proteins Bearing O-Linked Oligosaccharides: Application of Jacalin-Sepharose', *Anal. Biochem.*, **188**, 271–277 (1990).
191. G.L. Hortin, 'Isolation of Glycopeptides Containing O-Linked Oligosaccharides by Lectin Affinity Chromatography on Jacalin-Agarose', *Anal. Biochem.*, **191**, 262–267 (1990).
192. K. Maemura, M. Fukuda, 'Poly-*N*-Acetyllactosaminyl O-Glycans Attached to Leukosialin. The Presence of Sialyl Le^x Structures in O-Glycans', *J. Biol. Chem.*, **267**, 24 379–24 386 (1992).
193. R.R. Towsend, M.R. Hardy, T.C. Wong, Y.C. Lee, 'Binding of N-linked Bovine Fetuin Glycopeptides to Isolated Rabbit Hepatocytes: Gal/GalNAc Hepatic Lectin Discrimination Between Gal(β1–4)GlcNAc and Gal(β1–3)GlcNAc in a Triantennary Structure', *Biochemistry*, **25**, 5716–5725 (1986).
194. I.S. Rohrer, G.A. Cooper, R.R. Towsend, 'Identification Quantification and Characterization of Glycopeptides in Reversed-phase HPLC Separation of Glycoprotein Proteolytic Digests', *Anal. Biochem.*, **212**, 7–16 (1993).
195. K.L. Hsi, L. Chen, D.H. Hawke, L.R. Zieske, P.M. Yuan, 'A General Approach for Characterizing Glycosylation Sites of Glycoproteins', *Anal. Biochem.*, **198**, 238–245 (1991).
196. J.R. Barr, K.R. Anumula, M.B. Vettese, P.B. Taylor, S.A. Carr, 'Structural Classification of Carbohydrates in Glycoproteins by Mass Spectrometry and High-performance Anion-exchange Chromatography', *Anal. Biochem.*, **192**, 181–192 (1991).
197. J.C. Michalski, C. Alonso, 'HPLC of Oligosaccharides and Glycopeptides', in *HPLC of Macromolecules. A Practical Approach*, 2nd edition, ed. R.W.A. Oliver, IRL Press, Oxford, 171–202, 1998.
198. M.J. Davies, E.F. Hounsell, 'Comparison of Separation Modes of High-performance Liquid Chromatography for the Analysis of Glycoprotein and Proteoglycan Derived Oligosaccharides', *J. Chromatogr., A*, **720**, 227–233 (1996).

199. J.M. Lo-Guidice, M. Lhermitte, 'HPLC of Oligosaccharides in Glycobiology', *Biomed. Chromatogr.*, **10**, 290–296 (1996).
200. Y.C. Lee, 'High-performance Anion-exchange Chromatography for Carbohydrate Analysis', *Anal. Biochem.*, **189**, 151 (1990).
201. R.F. Borch, M.D. Bernstein, H.D. Durst, 'The Cyanohydridoborate Anion as a Selective Reducing Agent', *J. Am. Chem. Soc.*, **93**, 2897–2904 (1971).
202. S. Hase, T. Ikenaka, Y. Matsushima, 'Structure Analysis of Oligosaccharides by Tagging of the Reducing End Sugars with a Fluorescent Compound', *Biochem. Biophys. Res. Commun.*, **85**, 257–263 (1978).
203. S. Hase, 'High-performance Liquid Chromatography of Pyridylaminated Saccharides', *Methods Enzymol.*, **230**, 225–237 (1994).
204. S. Hase, S. Hara, Y. Matsushima, 'Tagging of Sugars with a Fluorescent Compound, 2-Aminopyridine', *J. Biochem. (Tokyo)*, **85**, 217–220 (1979).
205. P.F. Daniel, D.F. Defeudis, I.T. Lott, R.H. McCluer, 'Quantitative Microanalysis of Oligosaccharides by High-performance Liquid Chromatography', *Carbohydr. Res.*, **97**, 161–180 (1981).
206. G. Grönberg, P. Lipniunias, T. Lundgren, K. Erlansson, F. Lindh, B. Nilsson, 'Isolation of Monosialylated Oligosaccharides from Human Milk and Structural Analysis of Three New Compounds', *Carbohydr. Res.*, **191**, 261–278 (1989).
207. G. Grönberg, P. Lipniunias, T. Lundgren, F. Lindh, B. Nilsson, 'Isolation and Structural Analysis of Three New Disialylated Oligosaccharides from Human Milk', *Arch. Biochem., Biophys.*, **278**, 297–311 (1990).
208. S. Fievre, J.M. Wieruszeski, J.C. Michalski, J. Lemoine, J. Montreuil, G. Strecker, 'Primary Structure of a Trisialylated Oligosaccharide from Human Milk', *Biochem. Biophys. Res. Commun.*, **177**, 720–725 (1991).
209. S. Yamamoto, S. Hase, H. Yamauchi, T. Tanimoto, T. Ikenaka, 'Studies on the Sugar Chain of Interferon Gamma from Human Peripheral Blood Lymphocytes', *J. Biochem. (Tokyo)*, **105**, 1034–1039 (1989).
210. C. Capon, P. Cache, Y. Leroy, G. Strecker, J. Montreuil, B. Fournet, 'Isolation of the Major O-Glycosidically Linked Oligosaccharides Obtained by Alkaline Borohydride Degradation of Human Meconium Glycoproteins', *J. Chromatogr.*, **425**, 35–45 (1988).
211. J.C. Michalski, J.M. Wieruszeski, C. Alonso, P. Cache, J. Montreuil, G. Strecker, 'Characterization and 400 MHz ^{1}H NMR Analysis of Urinary Fucosyl Glycoasparagines in Fucosidosis', *Eur. J. Biochem.*, **201**, 439–458 (1991).
212. H. Nakagawa, Y. Kawamura, K. Kato, I. Shimada, Y. Arata, N. Takahashi, 'Identification of Neutral and Sialyl N-linked Oligosaccharide Structures from Human Serum Glycoproteins Using Three Kinds of High-performance Liquid Chromatography', *Anal. Biochem.*, **226**, 130–138 (1995).
213. N. Takahashi, 'Three Dimensional Mapping of N-Linked Oligosaccharides Using Anion-exchange Hydrophobic and Hydrophilic Interaction Modes of High-performance Liquid Chromatography', *J. Chromatogr., A*, **720**, 217–225 (1996).
214. G.R. Guile, P.M. Rudd, D.R. Wing, S.B. Prime, R.A. Dwek, 'A Rapid High-resolution High-performance Liquid Chromatographic Method for Separating Glycan Mixtures and Analyzing Oligosaccharide Profiles', *Anal. Biochem.*, **240**, 210–226 (1996).
215. R.R. Towsend, P.H. Lipnunias, C. Bigge, A. Ventom, R. Parekh, 'Multimode High-performance Liquid Chromatography of Fluorescently Labeled Oligosaccharides from Glycoproteins', *Anal. Biochem.*, **239**, 200–207 (1996).
216. K.R. Anumula, S.T. Dhume, 'High Resolution and High Sensitivity Methods for Oligosaccharides Mapping and Characterization by Normal Phase High-performance Liquid Chromatography Following Derivatization with Highly Fluorescent Anthranilic Acid', *Glycobiology*, **8**, 685–694 (1998).
217. N. Tomiya, J. Awaya, M. Kurono, S. Endo, Y. Arata, N. Takahashi, 'Analyses of N-linked Oligosaccharides Using a Two-dimensional Mapping Technique', *Anal. Biochem.*, **171**, 73–90 (1988).
218. H. Higashi, M. Ito, N. Fukaya, S. Yamagata, T. Yamagata, 'Two-dimensional Mapping by High-performance Liquid Chromatography of Oligosaccharides released from Glycosphingolipids by Endoglucoceramidase', *Anal. Biochem.*, **186**, 355–362 (1990).
219. K. Ohara, M. Sano, A. Kondo, I. Kato, 'Two-dimensional Mapping by the High-performance Liquid Chromatography of Pyridylamino Oligosaccharides from Various Glycosphingolipids', *J. Chromatogr.*, **586**, 35–41 (1991).
220. P.F. Daniel, 'Separation of Benzoylated Oligosaccharides by Reversed Phase High Pressure Liquid Chromatography: Application to High Mannose Type Oligosaccharides', *Methods Enzymol.*, **138**, 94–116 (1987).
221. M.J. Davies, E.F. Hounsell, 'HPLC and HPAEC of Oligosaccharides and Glycopeptides', *Methods Mol. Biol.*, **76**, 79–100 (1998).
222. J.U. Baenziger, 'High-performance Liquid Chromatography of Oligosaccharides', *Methods Enzymol.*, **230**, 237–249 (1994).
223. J. Van Pelt, J.B.C. Damm, J.P. Kamerling, J.F.G. Vliegenthart, 'Separation of Sialyl Oligosaccharides by Medium Pressure Anion Exchange Chromatography on Mono Q', *Carbohydr. Res.*, **169**, 43–51 (1987).
224. R.R. Towsend, M.R. Hardy, 'Analysis of Glycoprotein Oligosaccharides Using High pH Anion Exchange Chromatography', *Glycobiology*, **1**, 139–147 (1991).
225. J.M. Lo-Guidice, J.M. Wieruszeski, J. Lemoine, A. Verbert, P. Roussel, G. Lamblin, 'Sialylation and Sulfation of the Carbohydrate Chains in Respiratory Mucins from

a Patient with Cystic Fibrosis', *J. Biol. Chem.*, **269**, 18 794–18 813 (1994).
226. M.J. Davies, K.D. Smith, R.A. Carruthers, W. Chai, A.M. Lawson, E.F. Hounsell, 'Use of a Porous Graphitized Carbon Column for the High-performance Liquid Chromatography of Oligosaccharides, Alditols and Glycopeptides with Subsequent Mass Spectrometry Analysis', *J. Chromatogr.*, **646**, 317–326 (1993).
227. J.U. Baenziger, M. Natowicz, 'Rapid Separation of Anionic Oligosaccharide Species by High-performance Liquid Chromatography', *Anal. Biochem.*, **112**, 357–361 (1981).
228. E.D. Green, G. Adelt, J.U. Baenziger, S. Wilson, H. van Halbeek, 'The Asparagine-linked Oligosaccharides on Bovine Fetuin. Structural Analysis of *N*-Glycanase-released Oligosaccharides by 500 MHz ^{1}H NMR Spectroscopy', *J. Biol. Chem.*, **263**, 18 253–18 268 (1988).
229. S.J. Mellis, J.U. Baenziger, 'Separation of Anionic Oligosaccharides by High-performance Liquid Chromatography', *Anal. Biochem.*, **158**, 42–49 (1981).
230. G.P. Reddy, C.A. Bush, 'High-performance Anion Exchange Chromatography of Neutral Milk Oligosaccharides and Oligosaccharide-alditols derived from Mucin Glycoproteins', *Anal. Biochem.*, **198**, 278–284 (1991).
231. S.J. Turco, 'Rapid Separation of High Mannose Type Oligosaccharides by High-pressure Liquid Chromatography', *Anal. Biochem.*, **118**, 278–283 (1981).
232. M.L.E. Bergh, P.L. Koppen, D. Van Den Eijnden, 'High-pressure Liquid Chromatography of Sialic Acid Containing Oligosaccharides', *Carbohydr. Res.*, **94**, 225–229 (1981).
233. M.L. Bergh, P. Koppen, D.H. Van Den Eijnden, 'High-pressure Liquid Chromatography of Isomeric Oligosaccharides that Form Part of the Complex-type Carbohydrate Chains of Glycoproteins', *Carbohydr. Res.*, **117**, 275–278 (1983).
234. C.J. Edge, T.W. Rademacher, M.R. Wormald, R.B. Parekh, T.D. Butters, D.R. Wing, R.A. Dwek, 'Fast Sequencing of Oligosaccharides: the Reagent Array Analysis Method', *Proc. Natl. Acad. Sci. USA*, **89**, 6338–6342 (1992).
235. S. Ohlson, L. Hansson, P.O. Larsson, K. Mosbach, 'High-performance Liquid Affinity Chromatography (HPLAC) and its Application to the Separation of Enzymes and Antigens', *FEBS Lett.*, **93**, 5–9 (1978).
236. E.D. Green, R.M. Brodbeck, J.U. Baenziger, 'High-pressure Liquid Chromatography of Isomeric Oligosaccharides that Form Part of the Complex Type Carbohydrate Chains of Glycoproteins', *Carbohydr. Res.*, **117**, 275–278 (1985).
237. D. Zopf, S. Ohlson, J. Dakour, W. Wang, A. Lundblad, 'Analysis and Purification of Oligosaccharides by High-performance Liquid Affinity Chromatography', *Methods Enzymol.*, **179**, 55–64 (1989).
238. J.T. Gallagher, 'Affinity Chromatography of Complex Carbohydrates Using Lectins', in *HPLC of Macromolecules. A Practical Approach*, ed. R.W.A. Oliver, IRL Press, Oxford, 209–227, 1989.
239. S. Suzuki, S. Honda, 'A Tabulated Review of Capillary Electrophoresis', *Electrophoresis*, **19**, 2539–2560 (1998).
240. Z. El Rassi, 'Recent Developments in Capillary Electrophoresis of Carbohydrate Species', *Electrophoresis*, **18**, 2400–2407 (1997).
241. R.J. Linhardt, 'Capillary Electrophoresis of Oligosaccharides', *Methods Enzymol.*, **230**, 265–280 (1994).
242. M.V. Novotny, J. Sudor, 'High-performance Capillary Electrophoresis of Glycoconjugates', *Electrophoresis*, **14**, 373–389 (1993).
243. K. Kakehi, S. Honda, 'Analysis of Glycoproteins, Glycopeptides and Glycoprotein-derived Oligosaccharides by High-performance Capillary Electrophoresis', *J. Chromatogr., A*, **720**, 377–393 (1996).
244. F.T. Chen, R.A. Evangelista, 'Profiling Glycoprotein N-linked Oligosaccharides by Capillary Electrophoresis', *Electrophoresis*, **19**, 2639–2644 (1998).
245. H. Kakehi, S. Honda, 'Analysis of Carbohydrates in Glycoproteins by High-performance Liquid Chromatography and High-performance Capillary Electrophoresis', *Appl. Biochem. Biotechnol.*, **43**, 55–71 (1993).
246. P. Hermentin, R. Witzel, R. Doenges, R. Bauer, H. Haupt, T. Patel, R.B. Parekh, D. Brazel, 'The Mapping by High-pH Anion-exchange Chromatography with Pulsed Amperometric Detection and Capillary Electrophoresis of the Carbohydrate Moieties of Human Plasma Alpha1-acid Glycoprotein', *Anal. Biochem.*, **206**, 419–429 (1992).
247. K. Kakehi, A. Susami, A. Taga, S. Suzuki, S. Honda, 'High-performance Capillary Electrophoresis of *O*-Glycosidically Linked Sialic Acid-containing Oligosaccharides in Glycoproteins as their Alditol Derivatives with Low-wavelength UV Monitoring', *J. Chromatogr., A*, **680**, 209–215 (1994).
248. A. Guttman, C. Starr, 'Capillary and Slab Gel Electrophoresis Profiling of Oligosaccharides', *Electrophoresis*, **16**, 993–997 (1995).
249. P.L. Weber, S.M. Lunte, 'Capillary Electrophoresis with Pulsed Amperometric Detection of Carbohydrates and Glycopeptides', *Electrophoresis*, **17**, 302–309 (1996).
250. E. Watson, F. Yao, 'Capillary Electrophoretic Separation of Human Recombinant Erythropoietin (r-HuEPO) Glycoforms', *Anal. Biochem.*, **210**, 389–393 (1993).
251. N.K. Klausen, T. Kornfelt, 'Analysis of the Glycoforms of Human Recombinant Factor VIIa by Capillary Electrophoresis and High-performance Liquid Chromatography', *J. Chromatogr., A*, **718**, 195–202 (1995).
252. P.J. Oefner, C. Chiesa, 'Capillary Electrophoresis of Carbohydrates', *Glycobiology*, **4**, 397–412 (1994).
253. K.A. Denton, S.A. Tate, 'Capillary Electrophoresis of Recombinant Proteins', *J. Chromatogr., B*, **697**, 111–121 (1997).

254. P. Jackson, 'The Use of Polyacrylamide Gel Electrophoresis for the High Resolution Separation of Reducing Sugars Labeled with the Fluorophore 8-Amino-naphthalene-1,3,6-trisulfonic Acid', *Biochem. J.*, **270**, 705–713 (1990).
255. P. Jackson, G.R. Williams, 'Polyacrylamide Gel Electrophoresis of Reducing Saccharides Labelled with the Fluorophore 8-Amino-naphthalene-1,3,6-trisulfonic Acid: Application to Enzymatic and Structural Analysis of Oligosaccharides', *Electrophoresis*, **12**, 94–96 (1991).
256. P. Jackson, 'The Analysis of Fluorophore Labelled Glycans by High Resolution Polyacrylamide Gel Electrophoresis', *Anal. Biochem.*, **216**, 243–252 (1994).
257. C.M. Starr, R.I. Masada, C. Hague, E. Skop, J.C. Klock, 'Fluorophore-assisted Carbohydrate Electrophoresis, FACE® in the Separation, Analysis, and Sequencing of Carbohydrates', *J. Chromatogr., A*, **720**, 295–321 (1996).
258. G.G. Guile, P.M. Rudd, D.R. Wing, S.B. Prime, R.A. Dwek, 'A Rapid High Resolution Method for Separating Oligosaccharide Mixtures and Analyzing Sugar Prints', *Anal. Biochem.*, **240**, 210–226 (1996).
259. R.I. Masada, C. Hague, R. Seid, S. Ho, S. McAlister, V. Pigiet, C. Starr, 'Fluorophore Assisted Carbohydrate Electrophoresis, FACE® for Determining the Nature and Consistency of Recombinant Glycoprotein Glycosylation', *Trends Glycosci. Glycotech.*, **7**, 133–147 (1995).
260. E. Higgins, Y. Friedman, 'A Method for Monitoring the Glycosylation of Recombinant Glycoproteins from Conditioned Medium Using Fluorophore Assisted Carbohydrate Electrophoresis', *Anal. Biochem.*, **228**, 221–225 (1995).
261. N.K. Kotchekov, V.A. Derevitskaya, N.P. Arbatsky, 'The Structure of Pentasaccharides and Hexasaccharides from Blood Group Substance H', *Eur. J. Biochem.*, **67**, 129–136 (1976).
262. A.E. Manzi, A. Varki, 'Compositional Analysis of Glycoproteins', in *Glycobiology: A Practical Approach*, ed. M. Fukuda, IRL Press, Oxford, 27–77, 1994.
263. R. Schauer, A.P. Corfield, in *Sialic Acids: Chemistry, Metabolism and Function*, ed. R. Schauer, Springer Verlag, New York, 51–57, 1982.
264. Z.P. Zanetta, W.C. Breckenridge, G. Vincendon, 'Analysis of Monosaccharides by Gas–Liquid Chromatography of the O-Methyl Glycosides as Trifuoroacetate Derivatives. Application to Glycoproteins and Glycolipids', *J. Chromatogr.*, **69**, 291–304 (1972).
265. M.F. Chaplin, 'Monosaccharides', in *Carbohydrate Analysis – A Practical Approach*, 2nd edition, eds. M.F. Chaplin, J.F. Kennedy, IRL Press, Oxford, 1–42, 1994.
266. R.K. Merkle, I. Poppe, 'Carbohydrate Composition Analysis of Glycoconjugates by Gas–Liquid Chromatography/Mass Spectrometry', *Methods Enzymol.*, **230**, 1–15 (1994).
267. J.R. Clamp, T. Bhatti, R.E. Chambers, 'The Determination of Carbohydrate in Biological Materials by Gas–Liquid Chromatography', *Methods Biochem. Anal.*, **19**, 229–244 (1971).
268. M.F. Chaplin, 'A Rapid and Sensitive Method for the Analysis of Carbohydrate Components in Glycoproteins Using Gas–Liquid Chromatography', *Anal. Biochem.*, **123**(2), 336–341 (1982).
269. R.J. Henry, A.B. Blakeney, P.J. Harris, B.A. Stone, 'Detection of Neutral and Aminosugars from Glycoproteins and Polysaccharides as their Alditol Acetates', *J. Chromatogr.*, **256**, 419–427 (1983).
270. L. Kenne, S. Stromberg, 'A Method for the Microanalysis of Hexoses in Glycoproteins', *Carbohydr. Res.*, **198**, 173–179 (1990).
271. S. Hakomori, 'A Rapid Permethylation of Glycolipid and Polysaccharide Catalyzed by Methyl Sulfinyl Carbanion in Dimethyl Sulfoxide', *J. Biochem. (Tokyo)*, **55**, 205–208 (1964).
272. J. Finne, T. Krusius, H. Rauvala, 'Use of Potassium *Tert*-butoxide in the Methylation of Carbohydrates', *Carbohydr. Res.*, **80**, 336–339 (1980).
273. I. Ciucanu, F. Kerek, 'A Simple and Rapid Method for the Permethylation of Carbohydrates', *Carbohydr. Res.*, **131**, 209–217 (1984).
274. K.R. Anumula, P.B. Taylor, 'A Comprehensive Procedure for Preparation of Partially Methylated Alditol Acetates from Glycoprotein Carbohydrates', *Anal. Biochem.*, **203**, 101–108 (1992).
275. R. Geyer, H. Geyer, S. Kuhnhardt, W. Mink, S. Stirm, 'Capillary Gas Chromatography of Methylhexitol Acetates Obtained upon Methylation of *N*-Glycosidically Linked Glycoprotein Oligosaccharides', *Anal. Biochem.*, **121**, 263–274 (1982).
276. B. Fournet, G. Strecker, Y. Leroy, J. Montreuil, 'Gas–Liquid Chromatography and Mass Spectrometry of Methylated and Acetylated Methyl-glycosides. Application to the Structure Analysis of Glycoprotein Glycans', *Anal. Biochem.*, **116**, 489–502 (1981).
277. J.F.G. Vliegenthart, L. Dorland, H. Van Halbeek, 'High-resolution ^{1}H-NMR Spectroscopy as a Tool in the Structural Analysis of Carbohydrates Related to Glycoproteins', *Adv. Carbohydr. Chem. Biochem.*, **41**, 209–374 (1983) .
278. J.F.G. Vliegenthart, H. Van Halbeek, L. Dorland, 'The Applicability of 500 MHz High Resolution ^{1}H-NMR Spectroscopy for the Structure Determination of Carbohydrates Derived from Glycoproteins', *Pure Appl. Chem.*, **53**, 45–77 (1981).
279. H. Van Halbeek, '^{1}H-NMR Spectroscopy of Carbohydrate Chains of Glycoproteins', *Methods Enzymol.*, **230**, 132–168 (1994).
280. D.J. Harvey, T.J. Naven, B. Kuster, 'Identification of Oligosaccharides by Matrix-assisted Laser Desorption Ionization and Electrospray MS', *Biochem. Soc. Trans.*, **24**, 905–912 (1996).

281. A. Dell, A.J. Reason, K.H. Khoo, M. Panico, R.A. McDowell, H.R. Morris, 'Mass Spectrometry of Carbohydrate Containing Biopolymers', *Methods Enzymol.*, **230**, 108–132 (1994).
282. H. Geyer, R. Geyer, 'Strategies for Glycoconjugates Analysis', *Acta. Anat.*, **161**, 18–35 (1998).
283. B. Nilsson, 'Analysis of Protein Glycosylation by Mass Spectrometry', *Mol. Biotechnol.*, **2**, 243–280 (1994).
284. G.J. Rademaker, J. Thomas-Oates, 'Analysis of Glycoproteins and Glycopeptides Using Fast-atom Bombardment', *Methods Mol. Biol.*, **61**, 231–241 (1996).
285. F.M. Greer, H.R. Morris, 'Fast Atom Bombardment and Electrospray Mass Spectrometry of Peptides, Proteins and Glycoproteins', *Methods Mol. Biol.*, **64**, 147–163 (1997).
286. A. Dell, 'FAB-Mass Spectrometry of Carbohydrates', *Adv. Carbohydr. Chem. Biochem.*, **45**, 19–72 (1987).
287. H. Egge, J. Peter-Katalinic, 'Fast Atom Bombardment Mass Spectrometry for Structural Elucidation of Glycoconjugates', *Mass Spectrom. Rev.*, **6**, 331–393 (1987).
288. A. Dell, 'Preparation and Desorption Mass Spectrometry of Permethyl and Peracetyl Derivatives of Oligosaccharides', *Methods Enzymol.*, **193**, 647–660 (1990).
289. D.I. Papac, A. Wong, A.J. Jones, 'Analysis of Acidic Oligosaccharides and Glycopeptides by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **68**, 3215–3223 (1996).
290. Y. Mechref, M.V. Novotny, 'Mass Spectrometric Mapping and Sequencing of N-Linked Oligosaccharides Derived from Submicrogram amounts of Glycoproteins', *Anal. Chem.*, **70**, 455–463 (1998).
291. B. Stahl, S. Thurl, J. Zeng, M. Karas, F. Hillenkamp, M. Steup, G. Sawatzki, 'Oligosaccharides from Human Milk as Revealed by Matrix-assisted Laser Desorption Ionization Mass Spectrometry', *Anal. Biochem.*, **223**, 218–226 (1994).
292. C.W. Sutton, J.A. O'Neill, J.S. Cottrell, 'Site-specific Characterization of Glycoprotein Carbohydrates by Endoglycosidase Digestion and Laser Desorption Mass Spectrometry', *Anal. Biochem.*, **281**, 34–46 (1994).
293. Y. Yang, R. Orlando, 'Identifying the Glycosylation Sites and Site-specific Carbohydrate Heterogeneity of Glycoproteins by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 932–936 (1996).
294. C.W. Sutton, J.A. O'Neill, 'Identification of Sites of Glycosylation', *Methods Mol. Biol.*, **64**, 307–315 (1997).
295. R.B. Parekh, 'Glycoform Analysis of Glycoproteins', *Methods Enzymol.*, **230**, 340–348 (1994).
296. R.A. Dwek, C.J. Edge, D.J. Harvey, M.R. Wormald, R.B. Parekh, 'Analysis of Glycoprotein Associated Oligosaccharides', *Annu. Rev. Biochem.*, **62**, 65–100 (1993).
297. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', *Anal. Chem.*, **66**, 634R–683R (1994).
298. M.J. Huddleston, M.F. Bean, S.A. Carr, 'Collisional Fragmentation of Glycopeptides by Electrospray Ionization LC/MS and LC/MS/MS: Methods for Selective Detection of Glycopeptides in Protein Digests', *Anal. Chem.*, **65**, 877–884 (1993).
299. K.F. Medzihradszky, M.J. Besman, A.L. Burlingame, 'Reverse Phase Capillary High-performance Liquid Chromatography/High-performance Electrospray Ionization Mass Spectrometry: an Essential tool for the Characterization of Complex Glycoprotein Digests', *Rapid Commun. Mass Spectrom.*, **12**, 472–478 (1998).
300. I. Mazsaroff, W. Yu, B.D. Kelley, J.E. Vath, 'Quantitative Comparison of Global Carbohydrate Structures of Glycoproteins Using LC/MS and In Source Fragmentation', *Anal. Chem.*, **69**, 2517–2524 (1997).
301. B. Yeung, T.J. Porter, J.E. Vath, 'Direct Isoform Analysis of High Mannose Containing Glycoproteins by On Line Capillary Electrophoresis Electrospray Mass Spectrometry', *Anal. Chem.*, **69**, 2510–2516 (1997).
302. K.L. Duffin, J.K. Welply, E. Huang, J.D. Henion, 'Characterization of N-Linked Oligosaccharides by Electrospray and Tandem Mass Spectrometry', *Anal. Chem.*, **64**, 1440–1448 (1992).
303. Y.K.E. Hogelend, M.L. Deinzer, 'Mass Spectrometric Studies of the N-Linked Oligosaccharides by Baculovirus Mouse Interleukin-3', *Biol. Mass Spectrom.*, **23**, 218–224 (1994).
304. A.L. Burlingame, 'Characterization of Protein Glycosylation by Mass Spectrometry', *Curr. Opin. Biotechnol.*, **7**, 4–10 (1996).
305. Y. Zhao, S.B. Kent, B.T. Chait, 'Rapid, Sensitive Structure Analysis of Oligosaccharides', *Proc. Natl. Acad. Sci. USA*, **94**, 1629–1633 (1997).
306. N.H. Packer, M.A. Lawson, D.R. Jardine, J.C. Sanchez, A.A. Gooley, 'Analysis of Glycoproteins Separated by Two-dimensional Gel Electrophoresis', *Electrophoresis*, **19**, 981–988 (1998).
307. N.H. Packer, A. Pawlak, W.C. Kett, A.A. Gooley, J.W. Redmond, K.L. Williams, 'Proteome Analysis of Glycoforms: a Review of Strategies for the Microcharacterization of Glycoproteins Separated by Two-dimensional Polyacrylamide Gel Electrophoresis', *Electrophoresis*, **18**, 452–460 (1997).

Glycoprotein Analysis: Using Nuclear Magnetic Resonance

Bas R. Leeflang and J.F.G. Vliegenthart
Utrecht University, Utrecht, The Netherlands

1 Introduction 1
2 Sample Preparation 2
2.1 Solvent 2
2.2 Purification and Desalting 2
2.3 Sample Amounts and Sensitivity 2
3 Primary Structure 3
3.1 Glycan Characterization 3
3.2 Structural Reporter Groups 3
3.3 Protein Characterization 4
3.4 Glycopeptides 5
4 Spatial Structure 5
4.1 Spectrum Assignment 5
4.2 From Nuclear Overhauser Effect Towards Spatial Structure 6
5 Molecular Mobility 6
5.1 Conformation and Flexibility 7
5.2 Interaction and Recognition 7
6 Overview of Nuclear Magnetic Resonance Experiments 8
6.1 Correlation Spectroscopy 8
6.2 Double Quantum Filtered Correlation Spectroscopy 9
6.3 Triple Quantum Filtered Correlation Spectroscopy 9
6.4 Total Correlation Spectroscopy 9
6.5 Heteronuclear Single Quantum Coherence Spectroscopy 10
6.6 Heteronuclear Multiple Quantum Coherence Spectroscopy 10
6.7 Heteronuclear Multiple Bond Coherence Spectroscopy 10
6.8 Nuclear Overhauser Enhancement Spectroscopy 11
6.9 Rotating Frame Nuclear Overhauser Enhancement Spectroscopy 11
Acknowledgments 11
Abbreviations and Acronyms 11
Related Articles 12
References 12

The majority of proteins are decorated with one or more covalently linked carbohydrate chains, and are thus glycoproteins. These carbohydrate chains may consist of mono-, oligo- or polysaccharides, which can affect the physiological behavior and biological functioning of the glycoprotein. Insight into the structural aspects and characteristics of glycoprotein glycosylation is an important step in unraveling the function of carbohydrate chains in a glycoprotein. In most cases the glycan and protein moieties are cleaved and analyzed separately. Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool in this respect. A plethora of NMR techniques are available to obtain NMR parameters such as chemical shift, scalar coupling constants and relaxation and cross-relaxation properties. From these, both primary structural, three-dimensional structural and dynamical features can be determined for these molecules. In this article, several aspects of glycoprotein structural studies will be discussed, and a selection of NMR experiments, including experimental conditions, will be presented.

1 INTRODUCTION

Proteins are among the most important biomolecules. More than 50% of all proteins are decorated with one or more covalently linked carbohydrate chains, and are thus glycoproteins. These carbohydrate chains may consist of mono-, oligo- or polysaccharides, which can affect the physiological behavior and biological functioning of the glycoprotein. Insight into the structural aspects and characteristics of glycoproteins glycosylation is an important step in unraveling the function of carbohydrate chains in a glycoprotein.[1–3]

Until recently, the NMR analysis of intact glycoproteins was too elaborate and/or too difficult to do. Only in recent years have some glycoprotein NMR studies been presented. In most cases, the structural analysis is subdivided into smaller problems. Initially, the glycan and the protein moieties are cleaved, and analyzed separately. In some cases this is followed by a study of the intact glycoprotein. The cleavage can be achieved enzymatically or chemically. The peptide sequence has to be analyzed also, in cases where it is not yet known. The monosaccharide composition and branching information can be established through chemical means. NMR spectroscopy has proven to be a suitable tool to determine the primary sequence of the glycan in combination with the monosaccharide data. In many studies the analysis was considered complete when the primary glycan structure had been determined. The combination of modern equipment and the desire to gain a fundamental understanding of the structure–function

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

relation resulted in more advanced studies. These studies include conformational analysis and spatial structure determination of both the glycan and the protein, as well as of the intact glycoprotein. Insight into the molecular mobility and flexibility is important in order to understand complex functions and interactions, such as molecular recognition processes. Again, NMR spectroscopy is a suitable tool to obtain this information. In this article, several aspects of glycoprotein structural studies will be discussed.

2 SAMPLE PREPARATION

In order to obtain the most information from NMR spectra, the preparation of the samples is of vital importance. Obviously much attention has to be given to the choice of solvent, to solvent conditions and to sample purity. In particular, contamination by protonated molecules should be avoided as much as possible. Depending on the type of NMR equipment, various sample holders are available, ranging from conventional NMR tubes to dedicated low-concentration, low-volume sample holders.

2.1 Solvent

For ^{1}H-NMR experiments on carbohydrates, the samples are normally dissolved in deuterium oxide (D_2O or ^2H_2O), a solvent that closely resembles the natural aqueous solution, however, without the characteristic huge amount of protons. In this way, the skeleton carbohydrate protons can readily be observed. In D_2O solution all exchangeable protons (−OH and −NH) are nonobservable. Often the exchange of these protons with deuterium atoms is done in advance in one or more D_2O-dissolution/lyophilization cycles. However, in some cases one is particularly interested in these labile protons, for instance when glycopeptides or glycoproteins are studied. The amide signals are needed to obtain a complete sequential assignment of the peptide moiety. Furthermore, the observation of carbohydrate hydroxyl signals or amine and amide resonances has definite applications in the field of conformational analysis. In these cases the sample is generally dissolved in an H_2O/D_2O mixture with a D_2O content between 5% and 10%, needed for the deuterium lock system. It should be noted here that for the latter type of experiments the pH plays a crucial role. The chemical exchange of these labile protons is readily catalyzed at a pH above 7. In many cases the pH is adjusted between pH 5.5 and pH 6.5 to ensure a slow exchange. The exchange rate of hydroxyl protons is larger than that of NH-protons. The presence of metal ions (or salts in general) will have a catalytic effect on the labile proton exchange. The occurrence of paramagnetic impurities is harmful to most NMR experiments. Metal ions are conveniently removed from the sample by passing it through a small column of Calix (Bio-Rad™).

2.2 Purification and Desalting

Next to the choice of solvent, a second condition for obtaining good NMR spectra is careful sample preparation. Next to a functional fractionation of the samples using several chromatographic techniques, all proton-containing contaminating substances, such as buffer compounds, must be removed from the samples. Anionic carbohydrates and large oligosaccharides are easily desalted on small (1 cm × 20 cm) gel filtration columns (e.g. Bio-Gel™ P-2, Bio-Rad™) eluted with water. The sample is recovered in the void volume and is ready for D_2O exchange after lyophilization. If the sample contains high amounts of acetate, as can be the case after high-pH anion exchange chromatography, the use of 5 mM NH_4HCO_3 as eluent instead of pure water facilitates the removal of the acetate. Small neutral oligosaccharides can be desalted on a mixed-bed of cation (e.g. AG 50 W, Bio-Rad™) and anion (e.g. AG 1, Bio-Rad™) exchange resins. A recent advance in sample desalting techniques is the introduction of the Hi-trap column (Pharmacia).

2.3 Sample Amounts and Sensitivity

The amount of information that can be obtained by NMR spectroscopy is immense. The technique, however, is rather insensitive compared with other spectroscopic or spectrometric techniques. The first solution is to put a large amount of sample in the NMR instrument. This can be achieved by using high sample concentrations and/or using NMR tubes with a large diameter. Therefore, sensitivity is not a problem in cases where ample amounts of material are available and are soluble. However, in the analysis of released glycoprotein glycans, the total available amounts are often small (nanomolar and micromolar ranges). Using the standard NMR hardware, the best choice is to use special NMR tubes with a thick glass bottom and a glass insert (plunger) on top of the sample. These kind of NMR tubes are often referred to as ''Shigemi-tubes'' , named after the main manufacturer. The glass has a magnetic susceptibility close to that of water. Consequently, the magnetic field lines are not refracted at the sample–glass interface. The result is that the effective sample volume can be reduced from 500 to 250–300 μL. Therefore, the NMR spectra can be recorded at increased sample concentrations. To reduce the sensitivity problem further, the manufacturers of NMR instruments offer ''micro-probes'' or ''nano-probes''. In comparison with conventional NMR measurements, the

sample volumes are drastically reduced from typically 500 to 150–50 μL, respectively. In this kind of NMR probe the sensitivity gain is achieved by placing the receiver coils much closer to the (smaller) sample. Especially with the nanoprobe technology, NMR spectra of released glycoprotein glycans can be recorded with sample amounts as small as 10 nmol.

3 PRIMARY STRUCTURE

To characterize the primary structure of a glycoprotein, several structural parameters have to be determined: amino acid sequence, type of linkage of the carbohydrate chain to the polypeptide backbone, position of the glycosylation site, and the primary structure of the carbohydrate chain. The amino acid sequence can be obtained in several ways. Thus far, several different glycan–protein linkage types have been identified. *N*-glycosyl linkages between *N*-acetylglucosamine and asparagine residues, *O*-glycosylation of serine and threonine by *N*-acetylgalactosamine, *N*-acetylglucosamine, or fucose residues all occur widely. In the linkage region of proteoglycans,[4] *O*-glycosylation of serine by xylose is found. Less well known are *C*-glycosidic linkages that were discovered in recent years.[5,6] In this article the focus will be mainly on the analysis of *N*- and *O*-glycosidic linkages.

3.1 Glycan Characterization

Glycoproteins are rarely single compounds. Even when they are homogeneous in the amino acid sequence, glycoproteins may consist of a family of compounds. This is due to (micro)heterogeneity that may occur in each glycan at each glycosylation site. This feature forms a serious complication in the structure determination of the glycans. Therefore, characterization of a glycoprotein sample often starts with the cleavage of the carbohydrate chains from the intact glycoprotein along chemical or enzymatic routes. *N*-linked glycans can be split off through hydrazinolysis. Although this technique has often been applied, in general the use of enzymes has to be preferred, because fewer artifacts are introduced. Several different types of enzymes can be taken into consideration for this purpose: first the so-called endo-enzymes that cleave within the *N*,*N*-diacetylchitobiose unit. The enzymes belonging to this category have a rather strict specificity, dependent on the glycan structure. The peptide-*N*-(*N*-acetyl-β-glucosaminyl)asparagine amidase F cleaves the linkage between the first *N*-acetylglucosaminyl residue and the amide group of asparagine. Since it is an amidase, it has a very broad specificity. However, PNGase F is almost inactive when fucose is (α 1–3)linked to the first *N*-acetylglucosamine residue. In that case, PNGase A has to be preferred. Following the enzyme digestion, the cleaved oligosaccharides are purified using several chromatographic steps.

For the *O*-linked carbohydrate chains, there are no enzymes available that are widely applicable for cleavage of the whole chain. The most generally applicable cleavage reaction is based on the alkaline borohydride treatment. The resulting oligosaccharide alditols are fractionated through chromatography. Instead of splitting the carbohydrate chains from the glycoprotein, glycopeptides can be prepared through digestion of the glycoprotein with proteolytic enzymes. The structure determination of glycan fragments comprises invariably multi-step processes. A first indication of the character of the glycans with regard to molecular size and charge is obtained from the fractionation by different high-performance liquid chromatography (HPLC) techniques. The qualitative and quantitative monosaccharide composition is established after methanolysis followed by gas chromatography (GC). The absolute configuration of the monosaccharides is determined after solvolysis of the methyl glycosides with 2-butanol by GC.[7,8] The molecular masses of oligosaccharides are derived by mass spectrometric techniques such as fast atom bombardment mass spectrometry or matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

NMR spectroscopy is by far the most informative method for the determination of the glycan structures. For the primary structure, it is the only method that provides all details, comprising types of constituting monosaccharides including ring size and anomeric configuration, position of glycosidic linkages as well as type and position of non-carbohydrate substituents. Since NMR spectroscopy is a nondestructive technique, it is the most obvious first approach in the analysis of unknown glycans. Prior to NMR spectroscopic analysis, the dry samples are dissolved in D_2O and lyophilized in order to exchange the hydroxyl and amide protons for deuterium and dissolved finally in 500 μL D_2O with a high isotopic purity.

3.2 Structural Reporter Groups

The one-dimensional ^{1}H-NMR spectra of glycoprotein-derived glycans in ^2H_2O solutions show that two groups of signals can be distinguished: the so-called bulk signal and the structural reporter group signals. The former contain mainly the nonanomeric protons, present in a rather narrow spectral area, between 3.2 and 3.9 ppm. Owing to overlap this group of resonances can hardly be interpreted in terms of individual protons. The structural reporter group signals are found outside the bulk region. It appears that it is not necessary to assign all ^{1}H-NMR signals for a full characterization in terms of a carbohydrate structure.

Table 1 Structural reporter groups for glycoprotein derived glycans

- Anomeric protons
- Alkyl and acyl substituents like methyl, acetyl, glycolyl
- Sialic acid H3 atoms
- GalNAc-ol H2, H3, H4 and H5, Man-H2 and Gal-H4 atoms
- Protons shifted out of the bulk-region due to glycosylation shifts or under the influence of substituents, such as sulfate, phosphate, and acyl groups
- 6-Deoxy sugar H6 atoms

The exact location of a proton resonance in the spectrum is determined by the average magnetic field experienced by this proton. This local magnetic field depends on the primary structure (i.e. electron distribution), the spatial structure, and the mobility of the structure. The relation between the chemical shift and the structural parameters is very complex, and not easily accessible for theoretical approaches. Prominent structural reporter groups are the signals as summarized in Table 1.

The various chemical-shift patterns comprising chemical shifts and couplings are translated into structural information based on a comparison with patterns in a library of relevant reference compounds.[9–11] The comparison of NMR data for many closely related glycans resulted in empirical rules to correlate chemical shift values with carbohydrate structures. Obviously, the application of this approach requires accurate calibration of the experimental conditions like sample temperature, solvent, and pH. It is remarkable that the reporter group signals are such excellent probes of local structural elements. They are rather insensitive for alterations in the structure distant from the corresponding locus. Although the concept has proved its usefulness in the structure elucidation of numerous compounds, it should be applied with some care in the investigation of completely new structures. For definitive conclusions on the identity of novel compounds, additional experimental data from independent techniques should be collected. The application of the ^{1}H-NMR structural reporter group approach to determine the structure of *N*-glycans derived from glycoproteins was very successful.[9] A groundbreaking study was the elucidation of the structure of the glycans of α-acid glycoprotein.[12] For *O*-linked glycans of glycoproteins released as oligosaccharide alditols, we demonstrated that analogous structural reporter group signals can be identified.[9,11] Since the peripheral parts of the *N*- and *O*-linked glycans have several elements in common, the same type of structural reporter signals can be used in both cases.

The large amount of NMR data collected for quite an array of compounds enabled us to develop an NMR database for the computer-assisted identification of carbohydrates.[13] The database, designated SUGABASE, combines 1411 ^{1}H- and 590 ^{13}C-NMR chemical shifts of carbohydrate with 1756 structures and literature references as present in the complex carbohydrate structure database (CCSD).[14] The CCSD currently comprises almost 50 000 records. Starting from an input of chemical shift values, the search program produces a list of partially or completely matching structures. The SUGABASE program is available for UNIX and Microsoft Windows computers, and there is an interface to the database on the Internet at http://www.boc.chem.uu.nl/sugabase/databases.html.

The chemical shifts of these signals, together with their coupling constants, give valuable information on the structure of the glycans.[10] Also, the program is useful for the study of novel compounds, because structural elements can be traced that are occurring in other carbohydrates, already present in the database. The program can be applied for characterizing glycoconjugate-derived glycans and polysaccharides.

3.3 Protein Characterization

In many cases the amino acid sequence of the glycoprotein is known at the start of the characterization thereof. Therefore, and because it goes beyond the scope of an article on the analysis of glycoproteins by using NMR spectroscopy, this aspect of primary structure analysis will not be discussed here. The glycosylation sites, however, are not always known. Although characterization of the glycans is most conveniently done on free oligosaccharides, for the determination of glycosylation sites another approach must be followed. Instead of splitting the carbohydrate chains from the glycoprotein, glycopeptides can be prepared through digestion of the glycoprotein with proteolytic enzymes.

Depending on the specificity of the enzymes, the peptide parts can vary in length and the mixture of peptides and glycopeptides can be rather complex. Often for one glycosylation site more than one glycopeptide is found. Peptides are characterized by using N-terminus analysis in conjunction with comparison to the complete amino acid sequence. Structure analysis of the glycopeptides provides information on the identity of the glycans and on the position of the glycosylation sites in the polypeptide backbone.[15,16] The recent finding of a novel glycosylation type, C-glycosylation, complicates the analysis, because this linkage can thus far not be cleaved without destroying the entire glycoprotein.[5]

NMR spectroscopy is nowadays widely used in protein analysis. The incorporation of ^{13}C, ^{15}N and ^{2}H isotopes in proteins allows very powerful NMR experiments to be performed. Isotope labeling for glycoproteins, however,

is not frequently used as yet. The reason for this is that the common bacterial cell lines used to prepare isotope-labeled proteins do not have the glycosylation machinery that is essential to glycosylate the proteins. This means that the protein may be obtained but without most of all of its glycans, or with a different glycosylation pattern. Furthermore, the native protein folding frequently depends on the presence of glycans. Therefore one cannot be sure whether the labeled protein has adopted the same three-dimensional shape.

3.4 Glycopeptides

Glycopeptides can often be studied as intact molecules. The primary structure characterization of the glycan part is basically identical to that of free glycans. The structural reporter group method can be applied here as well. Although the chemical shifts of the monosaccharide attached to the peptide will be affected, the overall chemical shift profile will be very close to that of the free glycans. Two-dimensional NMR methods, as described in section 4 on spatial structure, can also be applied to determine the final details of the primary structure. Especially in the case of glycopeptides, this approach is very effective, because the structural reporter group approach is only applicable to the glycan moiety and the peptide moiety. Generally, the primary structures of the peptides are known from chemical amino-acid determinations. This means that NMR spectroscopy is merely used here for confirmation, and in the preparation of three-dimensional structure determinations. In most cases, however, glycopeptides are too small to have a defined spatial structure.

4 SPATIAL STRUCTURE

The properties of nuclear spins depend heavily on molecular structure and dynamics. As a consequence, NMR spectra are intrinsically rich in information on these topics. The NMR spectroscopist needs to assign the NMR resonances to individual spins in the molecule. This is a prerequisite for further analysis and extraction of valuable structural information about the molecule under study. The NMR observables used in structure analysis are primarily the nuclear Overhauser effects (NOEs) and three-bond J-coupling constants. Each NOE cross-peak corresponds with a short distance between two protons. Each three-bond coupling provides information on a dihedral angle. When the assignment of many peaks in the NMR spectra has been completed, lists of NOE cross-peaks and/or three-bond coupling constants can be obtained, providing information on interatomic distances and dihedral angles, respectively. This information can be converted into a three-dimensional model, using methods that frequently include force-field calculations.

4.1 Spectrum Assignment

Signal assignment of complex oligosaccharides can partly be done by comparing occurring chemical shifts with data stored in a library of known compounds. This technique uses the structural reporter group method. The comparison can be performed manually or with the aid of computer database programs. Though this method is valuable for characterization purposes, it does not provide sufficient signals assigned for more advanced analysis of the NMR spectra. Furthermore, the application of the structural reporter group method may fail, when a molecule with novel structural properties or with a large complex structure is studied.

Typically, homonuclear correlation type spectra, such as various correlation spectroscopy (COSY) or total correlation spectroscopy (TOCSY) experiments, are needed to assign signals in the bulk area. This kind of correlation spectrum yields spin systems that correspond to monosaccharides or amino acids (or parts thereof). The glycan assignment procedure generally starts from the anomeric proton signal, or any other well-resolved NMR signal. The two-dimensional spectra can be traversed from diagonal peak to diagonal peak via cross-peaks. In this way the resonances in each spin system can be identified.

Homonuclear correlation experiments identify spin systems. They are essential in this respect, but they do not provide sequence information. Mostly nuclear Overhauser effect spectroscopy (NOESY) or rotating frame nuclear Overhauser spectroscopy (ROESY) spectra are used for this purpose. Often the most intense NOESY peak identifies the linkage location. There are, however, several examples known where the most prominent NOESY peak represents a short distance to another proton. Care should be taken here, because NOESY interactions are through-space interactions and the shortest inter-proton distance may be a long-range interaction with respect to the primary structure. An unambiguous method to resolve this problem is the heteronuclear multiple bond coherence spectroscopy (HMBC) experiment, which yields through-bond correlations. This yields cross-peaks between the anomeric carbon and the ring proton of the adjacent sugar residue, and between the anomeric proton and the ring carbon of the adjacent sugar residue. In order to make use of HMBC spectra, the carbon resonances should be assigned as well as the proton resonances. Heteronuclear single quantum coherence spectroscopy (HSQC) or heteronuclear multiple quantum coherence spectroscopy (HMQC) spectra provide one-bond $^{1}H-^{13}C$ correlations. In other words they identify chemical bonds between ^{1}H and ^{13}C, and thus

provide the link between the assigned proton resonances and the carbon resonances. The HMBC experiment is basically a HMQC experiment tuned at small (1–10 Hz) heteronuclear couplings. Often the one-bond correlations are explicitly filtered out of the spectrum. However, in our laboratory this filter is hardly used, because omitting the filter gives many one-bond correlations that are useful in the assignment process.

4.2 From Nuclear Overhauser Effect Towards Spatial Structure

NOE cross-peak intensities depend strongly, among other parameters, upon inter-proton distances. Often the NOE cross-peak intensities are assumed to be proportional to r^{-6}, where r is the inter-proton distance of the protons to which the cross-peaks were assigned. Although this is generally incorrect, a slight modification of this approach provides a decent set of intramolecular distances which are only marginally affected by the spin-diffusion process or multiple spin relaxation effects. This method is called the ''initial rate approach''. For the initial rate approach, several NOESY experiments need to be recorded at various mixing times. The cross-peaks need to be integrated. By fitting a polynomial function to these integrals as a function of the mixing time, τ_m, the slope (initial rate) at $\tau_m = 0$ can be estimated.

The experimental NOE cross-peaks can be either negative or positive. This depends on the magnetic field (i.e. the Larmor frequency) and the molecular tumbling, expressed in the rotational-correlation time τ_c. For molecules having a τ_c near the critical value of $\sqrt{5/(2w)}$ no NOEs can be measured, and NOESY is not applicable. Small oligosaccharides fall in this category. NOE-like information, however, can also be obtained from ROESY data. Quantification of inter-proton distances from rotating-frame nuclear Overhauser effect (ROE) intensities is more complicated than from NOESY data.

For small molecules, knowledge of a few distances may be enough to build an initial molecular model, although in many cases, one could find contradicting short distances. In these cases, the molecule in solution may occur in two or more significantly different conformations (Figure 1). For larger molecules, such as (glyco)proteins, several hundreds of relevant short distances may be obtained, making it virtually impossible to build a molecular model from scratch manually. Just as with smaller molecules, conflicting short distances may occur due to different conformations, making the problem even more difficult. Several computational approaches have been developed to handle this problem, such as distance geometry and simulated annealing methods, or combinations of both. These methods yield many first approximation molecular models that may represent genuinely different conformations. These models are optimized using a short molecular mechanics or molecular dynamics run.

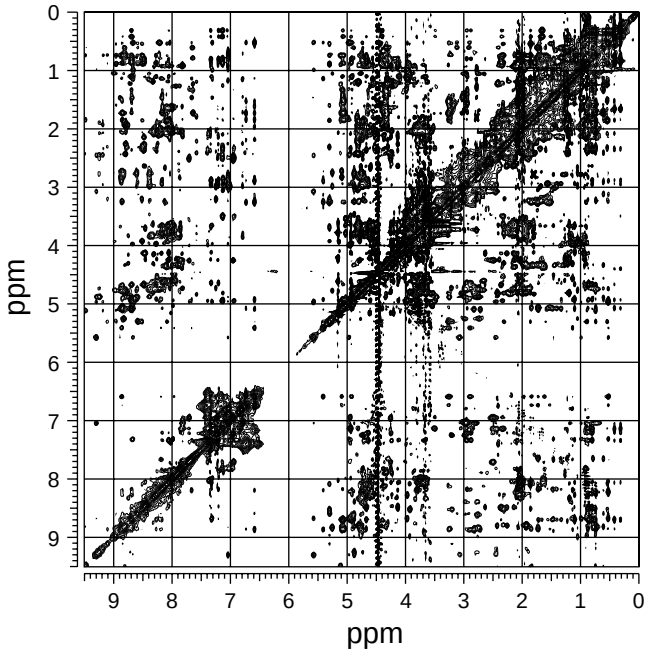

Figure 1 600 MHz NOESY spectrum of fully glycosylated alpha human chorionic gonadotrophin (αhCG) glycoprotein hormone in aqueous solution. The spectrum was recorded on a Bruker AMX-600 spectrometer at an elevated temperature of 333 K, with a NOE mixing time of 125 ms.

5 MOLECULAR MOBILITY

To gain insight into the function of oligosaccharides, glycoproteins, peptides or any other biomolecule, the determination of the spatial structure is a prerequisite. X-ray diffraction, NMR spectroscopy and computational methods are available to obtain molecular structures. Knowledge of the structure alone, however, is often not enough, since molecular flexibility is in most cases essential to their function in biochemical processes. NMR not only provides the tools to determine the spatial structure, but the internal molecular flexibility can also be studied. In NMR spectroscopy, molecular motions are reflected in the relaxation behavior of various NMR observables. The way in which NMR parameters, such as chemical shifts and coupling constants, and dipolar interactions, are affected depends on the relative timescale of the molecular motions. Atomic motion results in reduced NOE cross-peak intensities. The theory behind the NMR relaxation processes is well understood and can be used to determine both the molecular conformation and the molecular flexibility.

5.1 Conformation and Flexibility

Valuable NMR parameters for conformational analysis of biomolecules are the proton–proton NOEs or ROEs. They depend strongly on inter-proton distances and on the mobility of the inter-proton vectors. In the standard approach the mobility of all proton–proton vectors is considered to be constant. This is the so-called isotropic approximation. Now the cross-peak intensities are only a function of all inter-proton distances, and the ''overall'' tumbling rate (τ_c). This approximation is most useful to determine single conformational molecular models. In practice, however, various distinctly different conformations may occur. How these conformations affect the NMR spectra depends upon the timescale on which the conformations interconvert. In the very slow limit, the NMR spectrum will contain duplicate signals for the same nucleus. In this case, the spectrum is like a spectrum recorded for a mixture of molecules. A typical example of this kind of very slow interconversion is the α/β anomerization of reducing sugars in neutral, aqueous solution. At the other end of the timescale (faster than 1 ps) conformations change so rapidly that from a NMR point of view only the time-averaged structure can be observed. The rotation of a (hydroxy-)methyl group is a characteristic example of a very fast conformational change. It is the timescale region in between the extremes where most relevant conformational changes occur. NOEs and ROEs are especially sensitive for mobility in the microsecond to nanosecond timescale. In this regime the faster conformational changes happen, such as conformational changes in glycosidic linkages and glycan ring conversion. The NOESY spectra are now best analyzed, taking into account local mobility within the molecule that is faster than the overall tumbling rate. This is generally done with dedicated software, such as the CROSREL program.[17] Two application examples are given.[18,19] Conformational changes, however, occur at any timescale. Mostly, the effects can be observed by using T_1, T_2 and $T_{1\rho}$ relaxation times. For exchange rates between 1 μs and 1 ns, NMR spectroscopy does not have suitable parameters to judge the mobility. For the remaining exchange rates various NMR techniques are available.

As an example of the use of relaxation rates we would also like to mention the study on the mobilities of the glycan residues attached to Asn-78 of αhCG.[20–22] (Figure 2).

Figure 2 NMR structure of αhCG glycoprotein hormone obtained from NOESY data using the XPLOR program. Two well-defined protein regions are displayed on an atomic backbone level. The mobile protein region is represented by a ribbon. The glycan moieties are reduced to the first GlcNAc residues at Asn-52 and Asn-78, and are displayed on an all-atom level.

5.2 Interaction and Recognition

For the study of the conformation of oligosaccharides in oligosaccharide–protein complexes, transferred NOE (TRNOE) experiments[23,24] can be informative. Weakly bound ligands ($K_d \geq 10^{-6}$) in fast exchange (100–1000 Hz) between free and bound states can transfer NOEs obtained in the bound state to the free state. Since the line width of the ligand in the bound state is generally larger than that in the free state, and since the ligand must be present in excess, the signals from the free state ligand will dominate the spectrum. The NOE obtained in the bound state, however, is much stronger than that obtained in the free state. This causes the observed NOE peaks to be dominated by the conformation in the bound state. The TRNOE technique has been applied in a study of the binding of oligosaccharides to antibodies,[25,26] and to lectins.[26–28]

Laser photochemically induced dynamic nuclear polarization[29,30] is an advanced NMR technique that allows the detection in proteins and glycoproteins of surface exposed side chains of the aromatic amino acid residues tyrosine, tryptophan, and histidine. The

suitability of the method was demonstrated to study carbohydrate–protein interactions intramolecularly within glycoproteins[31] and intermolecularly for carbohydrate–lectin interactions.[32,33]

6 OVERVIEW OF NUCLEAR MAGNETIC RESONANCE EXPERIMENTS

NMR spectroscopy is renowned for its use of acronyms to identify the various available spectroscopic experiments that are mostly called pulse sequences. The large number of available pulse sequences and names obscures the fact that many multidimensional NMR experiments differ only in details. Pulse sequences can be categorized in different ways. Here we will classify them based on the magnetization transfer mechanism, which is linked to the information obtained from the resulting spectra. Two different pathways of magnetization transfer exist. All multidimensional NMR experiments use either one of these pathways, or combinations thereof.

Apart from the huge amount of experimental techniques available, NMR spectroscopy is often treated with caution because of the supposed need of quantum mechanics to understand the behavior of the nuclear spins. Fortunately, most experiments can be simplistically but effectively described in pictorial ways. For more comprehensive reviews of NMR theory, see Harris,[34] Ernst et al.,[35] Derome[36] and van der Ven,[37] and for a detailed practical overview see Braun et al.[38]

NMR experiments can be categorized based on the magnetization used or information transfer pathway. First, magnetization can be transferred through chemical bonds via the so-called scalar or *J*-coupling. This is the effect that causes the fine-splitting of NMR signals of individual nuclei. NMR spectra based on this route are often used for the assignment of NMR signals, because of the chemical relation between interacting nuclei. Second, magnetization can be transferred through space via cross-relaxation. Cross-relaxation is the effect when relaxation of the spin-state of a nucleus is stimulated by the close presence of one or more other nuclei. NMR spectra based on this route provide information on the spatial structure of the molecule.

6.1 Correlation Spectroscopy[38,39]

The COSY experiment was the first two-dimensional NMR experiment to be proposed and is still used today, although it has several drawbacks. Its strength, however, is that it provides information on proton pairs that are visibly coupled to each other in the one-dimensional NMR spectrum (fine splitting of the signal). With COSY there is no net magnetization transfer, which means that the cross-peaks have a zero integral. Half the multiplet signals are positive, whereas the other half is negative. As long as the *J*-coupling is larger than the line width this causes no problems. When the *J*-coupling becomes small compared to the line width, the cross-peak magnitude is reduced by the cancellation of the positive and negative multiplet

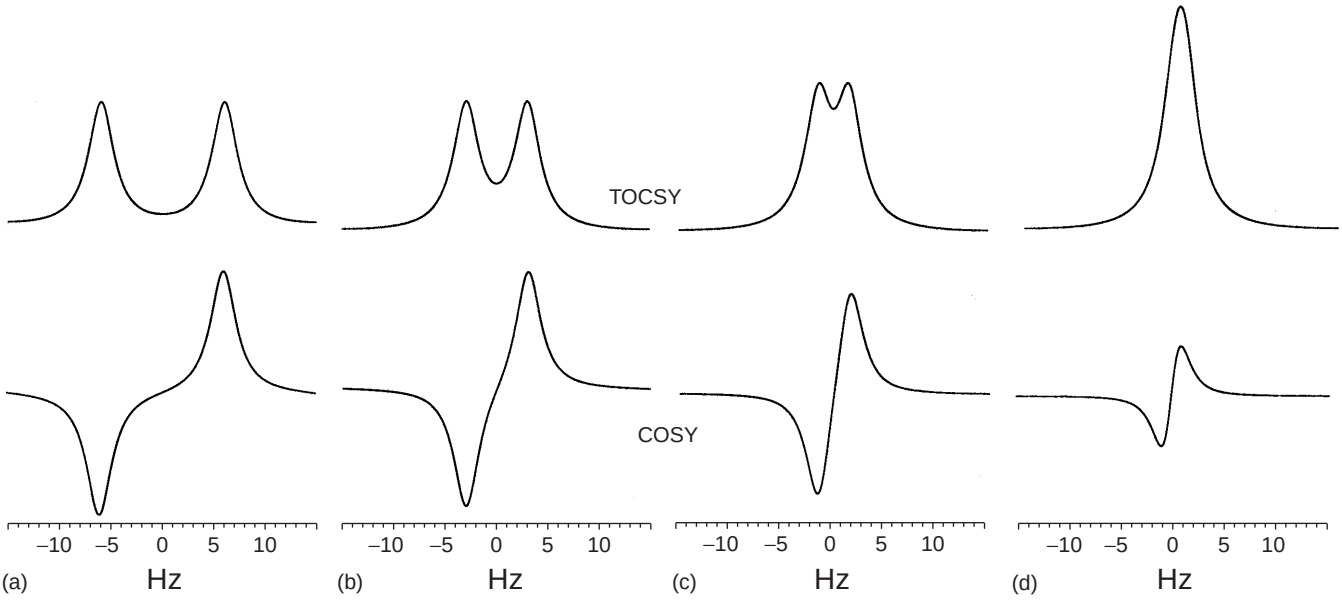

Figure 3 Demonstration of the signal canceling effects in simulated cross-peak cross-sections in COSY type spectra. Canceling occurs in cases where the *J*-coupling constant is smaller than the line width. A comparison is made between COSY and TOCSY peaks. In cases (a) through (d) the line width was set to 3 Hz, with varying *J*-coupling values: (a) $J = 12$ Hz; (b) $J = 6$ Hz; (c) $J = 3$ Hz; (d) $J = 1$ Hz.

signals (Figure 3). A second drawback, especially in the field of carbohydrate NMR, is the dispersive character of the diagonal peaks. Because of their long tails, the dispersive signals severely hamper the analysis of cross-peaks close to the diagonal. Preferred alternatives for COSY are double quantum filtered correlation spectroscopy (DQF-COSY) and TOCSY with a short mixing time.

6.1.1 Practical Information

Only hard 90° pulses are needed.

6.2 Double Quantum Filtered Correlation Spectroscopy[38,40]

The dispersive character of the diagonal peaks hampers the analysis of cross-peaks close to the diagonal using conventional COSY. Implementation of a double quantum filter, which allows only peaks involving two different spins, cancels the diagonal signals, including the solvent signal, to a very large extent. Now cross-peaks in the bulk area can also be analyzed. The problem of signal cancelling within the cross-peak multiplet still exists.

6.2.1 Practical Information

Only hard 90° pulses are needed.

6.3 Triple Quantum Filtered Correlation Spectroscopy[38,40]

In DQF-COSY a filtering technique is used that allows for cross-peaks between coupled protons that share at least a third proton to which they are mutually coupled. Triple quantum filtered correlation spectroscopy (TQF-COSY) is a COSY experiment with a triple quantum filter, which yields cross-peaks only when at least three spins are directly coupled. In practice this technique is often used to obtain COSY spectra containing H5/H6/H6′ cross-peaks in hexapyranosides, or H4/H5e/H5a cross-peaks for xylose, or similar triply coupled signals. This may be needed to overcome signal overlap in DQF-COSY spectra. TQF-COSY is not often used. It may, however, be useful in the case of Gal or GalNAc residues to assign the H5, H6 and H6′ atoms that cannot be found in TOCSY spectra.

6.3.1 Practical Information

Only hard 90° pulses are needed.

6.4 Total Correlation Spectroscopy[38,40,41]

In our laboratory, TOCSY, formerly also called HOHAHA (Homonuclear Hartmann Hahn spectroscopy) is the method of choice to assign as many protons as possible. TOCSY with a short mixing time, i.e. 8–16 ms, yielding spectra with approximately the same information content as DQF-COSY spectra would do, although with increased sensitivity. These experiments mostly have intense cross-peaks between the directly coupled protons. Often low-intensity cross-peaks are found, due to the relay of magnetization to the next coupling partner. This serves as an aid in the assignment. Not only the directly coupled proton is found on one track, but also those that are one or more coupling steps further. TOCSY with longer mixing times, typically up to 120 ms, may yield the complete subspectrum of one monosaccharide. The intensity of a TOCSY cross-peak depends on the magnitude of the J-coupling constants and on the mixing time. For Gal and GalNAc, where the coupling constants between H4 and H5 are small (<1.5 Hz), it is virtually impossible to obtain a full subspectrum. In these cases the magnetization transfer from the well-resolved H1 atom reaches the H4 but not the H5. H5, H6 and H6′ could be identified using TQF-COSY in combination with NOESY or ROESY experiments.

TOCSY does not suffer from the cross-peak cancellation when the line width is greater than the J-coupling (Figure 3). Occasionally the fact that TOCSY relays the magnetization over the whole spin system is a drawback. One is not always sure whether a cross-peak is a genuine 'COSY' type cross-peak or a RELAY-COSY peak. In these cases we still use a DQF-COSY experiment.

The occurrence of cross-peaks due to NOEs in the rotating frame (ROESY type cross-peaks) is a common artifact in TOCSY spectra. They can be distinguished from the intended TOCSY peaks by their invert sign. This effect is a smaller problem for free oligosaccharides than for larger molecules such as proteins or glycoproteins. The reason for this is that free oligosaccharide ROESY type cross-peaks build up slowly. For large molecules clean-TOCSY is often used, where the ROESY type relaxation is cancelled by NOESY type relaxation.

A second potential problem is the sample heating that may be caused by using intensive radiofrequency power, especially in cases when long mixing times are used. Heating can be detected by a changing lock signal level and will result in t_1 noise. In practice, a warming-up period of some 5 min suffices to overcome the effects of sample heating. At the end of this period the most critical shim(s) can be adjusted.

6.4.1 Practical Information

Mixing sequence: MLEV-17. Excitation using a hard 90° pulse and softer pulses (60, 90 and 180° pulses) to obtain a spin-lock field strength of 8–10 kHz. Effective 90° pulse: 25–30 μs. For TOCSY the effective 90° pulse corresponds to the actual 90° pulse. For clean TOCSY the length of the

clean delay should be included in the calculation. In our laboratory the clean delay is set to the same value as the length of the 90° pulse. Thus a 90° pulse of 16.66–20 μs is used. Mixing time: 8–150 ms for small molecules; 8–80 ms for (glyco-)proteins.

6.5 Heteronuclear Single Quantum Coherence Spectroscopy[38,41]

The HSQC experiment exploits the high sensitivity of protons compared to other nuclei such as ^{13}C and ^{15}N. This experiment yields a correlation spectrum for ^{1}H, ^{13}C and ^{1}H, ^{15}N, respectively, by a double insensitive nuclei enhanced by polarization transfer (INEPT) transfer step. For many years it was difficult to acquire heteronuclear correlation spectra for compounds having a natural abundance of ^{13}C isotopes. The emerging magnetic field gradient techniques have reduced these limitations dramatically. For HSQC the proton signal is detected. With a natural abundance of about 1.1% for ^{13}C, 98.9% of the protons are bound to a ^{12}C atom. This means that 98.9% of the detected signal must be eliminated. Without the use of magnetic field gradients, this could only be achieved by phase-cycling. Phase-cycling is a technique that cancels the unwanted signal by subtraction of free induction decays (FIDs) recorded with different phases of the radiofrequency pulses. Each FID, however, contains the unwanted signal for which the receiver has to be put at lower sensitivity. Furthermore, subtraction cancels the unwanted signal but not the associated noise. Using gradient techniques, the unwanted signal can be eliminated before the acquisition, resulting in a much higher sensitivity and preventing dynamical range problems. A further development in recent years is ''sensitivity enhancement'', mostly together with gradient coherence selection. To obtain narrow cross-peaks, normally the absorption mode signal is recorded and the dispersive signal is cancelled by phase cycling. Sensitivity-enhanced spectra allow both the absorption and dispersive mode signals to appear in the FID. In the processing stage these are separated, processed in such a way that an absorption mode spectrum is obtained with a sensitivity enhancement up to 1.41 ($\sqrt{2}$). The pulse sequence is longer, which allows for some relaxation and consequently loss of intensity. Typical sensitivity enhancements are about 1.3.

6.5.1 Practical Information

Hard 90 and 180° ^{1}H pulses and hard 90 and 180° ^{13}C pulses inversely detected. In order to obtain the shortest ^{13}C pulses the spectrometer is wired such that the output of the amplifier goes directly into the probehead. This wiring scheme prevents radiofrequency power loss in the pre-amplifier circuit, but at the same time prevents detection of the ^{13}C-NMR signals.

Depending on the available hardware, ^{13}C decoupling is possible during acquisition. A composite pulse decoupling scheme such as WALTZ or GARP is mostly used here.

To obtain a higher digital resolution in the indirect dimension, the carrier frequency and the sweep width in the ^{13}C domain may be adjusted, such that some signals are folded back into the spectrum in an empty region.

6.6 Heteronuclear Multiple Quantum Coherence Spectroscopy[38,41,42]

The information content of HMQC and HSQC spectra is the same. Both provide a correlation between ^{1}H and an X nucleus, mostly ^{13}C. The differences are hidden in the pulse-sequence. HSQC has more pulses and delays than HMQC and the period between the excitation pulse and the acquisition is longer in the case of HSQC. Therefore, HMQC is the preferred technique in cases where the accuracy of radiofrequency pulse lengths and radiofrequency pulse phases is not considered optimal. This may apply for the older hardware. Also, HMQC may be preferred in cases where T_2 relaxation is fast. A disadvantage of HMQC is that the cross-peak line widths are generally larger than those in the HSQC, because there is no proton decoupling in t_1.

6.6.1 Practical Information

See HSQC.

6.7 Heteronuclear Multiple Bond Coherence Spectroscopy[38,41,42]

The HMBC experiment is very important in unambiguous sequential assignment of carbohydrate residues, through the inter-glycosidic heteronuclear long-range coupling. Experimentally, the HMBC experiment is very similar to the HMQC experiment, where magnetization transfer is tuned to small heteronuclear couplings and the decoupling during acquisition is turned off. In many HMBC experiments the residual one-bond interactions are filtered out. In our laboratory this filter is purposely omitted. The presence of single bond correlations is useful for assignment purposes, because both HMBC and HMQC information are present in the same spectrum.

6.7.1 Practical Information

See HSQC.

6.8 Nuclear Overhauser Enhancement Spectroscopy[38,41,43]

NOESY spectra provide information related to inter-proton distances, and is therefore a valuable tool for structural analysis of carbohydrates, proteins and glycoproteins. The first application is determination of the primary structure of oligosaccharides and/or proteins. The linkage information cannot be found out unambiguously with NOESY and the observed NOE may be long range. This should be evaluated very carefully. HMBC provides this kind of information unambiguously, but at much lower sensitivity. The second application, by further analysis of NOESY spectra, is the determination of the (average) three-dimensional solution state structure of the molecules under study. NOEs can be used in a modeling study, either directly or by converting them into distance constraints. The sign and magnitude of the NOE do not only depend upon the inter-proton distances, but also on the overall (and local) mobility expressed in the rotational correlation time (τ_c). For large molecules (small τ_c) the NOESY cross-peaks have the same sign as the diagonal peaks, whereas for very small molecules the cross-peaks have the opposite sign. Di- and trisaccharides are generally in the area where the sign of the cross-peaks is changing, and can therefore not be analyzed with NOESY. In these cases ROESY experiments are the only alternative.

6.8.1 Practical Information

Only hard 90° proton pulses are needed. Solvent presaturation is normally done, using a low power radiofrequency pulse of 0.75 to 1.5 s. During the NOESY mixing time the solvent can be saturated as well. When the spectrometer is equipped with magnetic field gradients, a purging gradient may be given in the NOE mixing period, otherwise a homo-spoil pulse can be used.

6.9 Rotating Frame Nuclear Overhauser Enhancement Spectroscopy[38,44]

ROESY experiments yield spectra with cross-peaks of closely spaced protons. In this respect it is pretty much comparable to NOESY, although the experimental conditions differ. In contrast to NOESY, this experiment does not have a sign change depending on the size or mobility of the molecule. The cross-peaks are always negative relative to a positive diagonal. Positive cross-peaks do occur in ROESY spectra. Three possible causes for these phenomena exist. First, with chemical exchange, positive cross-peaks are obtained. Second, spin diffusion can be the cause. The first ROESY transfer inverts the sign of the cross-peak. Propagation via a second ROESY transfer within the same mixing period inverts the sign again, yielding a positive spin diffusion peak. Recording both ROESY and NOESY spectra is a method to analyze spin diffusion. Third, with ROESY being a spin-locking technique like TOCSY, it may suffer from TOCSY transfer, causing positive TOCSY peaks. TOCSY transfer may also give spurious peaks due to TOCSY transfer of cross-peak information to a third proton. Minimization of TOCSY transfer in the experimental phase is important for which several techniques are available, such as transverse ROESY (T-ROESY)[45] and off-resonance ROESY.[46] Equally important is coping with TOCSY transfer in the analysis phase. The CROSREL program[17–19] corrects for decreased ROESY cross-peak intensities caused by TOCSY transfer between the protons involved. Relay effects are not considered in this program.

6.9.1 Practical Information

Normally a hard 90° proton pulse is used for excitation. Solvent presaturation is normally done using a low-power radiofrequency pulse of 0.75 to 1.5 s. Spin-locking for conventional (not T-ROESY[45] or off-resonance ROESY[46]) is achieved at a field strength of 2–3 kHz (90° pulse corresponding to 125–90 µs). T-ROESY and off-resonance ROESY can be performed at higher field strengths, comparable to TOCSY experiments (10 kHz). For conventional ROESY the carrier frequency during spin-locking needs to be carefully chosen. This is important to reduce TOCSY artifacts on the one hand, and to obtain meaningful ROESY intensities on the other hand. The tilt angle θ of the most distant signal in the spectrum should be larger than 60°. Equation (1) shows the relation between θ, the spin-lock frequency ν_0 and the frequency of the most distant signal, ν_1:

$$\tan(\theta) = \frac{|\nu_1 - \nu_0|}{\gamma B_1} = \frac{|\nu_1 - \nu_0|}{4P_{90}} \qquad (1)$$

where P_{90} is the length of the 90° pulse at the spin-locking power in microseconds and B_1 is the magnetic field strength of the radiofrequency pulse.

ACKNOWLEDGMENTS

The authors acknowledge Prof. Dr J.P. Kamerling, Dr K. Hård, Dr P. Erbel and Dr J.A. van Kuik for their valuable contributions.

ABBREVIATIONS AND ACRONYMS

CCSD	Complex Carbohydrate Structure Database

COSY	Correlation Spectroscopy
DQF-COSY	Double Quantum Filtered Correlation Spectroscopy
FID	Free Induction Decay
GC	Gas Chromatography
HMBC	Heteronuclear Multiple Bond Coherence Spectroscopy
HMQC	Heteronuclear Multiple Quantum Coherence Spectroscopy
HPLC	High-performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
ROE	Rotating-frame Nuclear Overhauser Effect
ROESY	Rotating Frame Nuclear Overhauser Spectroscopy
TOCSY	Total Correlation Spectroscopy
TQF-COSY	Triple Quantum Filtered Correlation Spectroscopy
TRNOE	Transferred NOE
T-ROESY	Transverse ROESY
αhCG	Alpha Human Chorionic Gonadotrophin

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Circular Dichroism in Analysis of Biomolecules • High-performance Liquid Chromatography of Biological Macromolecules • Infrared Spectroscopy of Biological Applications • Nuclear Magnetic Resonance of Biomolecules

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance • High-performance Liquid Chromatography Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, 'New Comprehensive Glycoproteins I', in *Biochemistry*, Elsevier, Amsterdam, Vol. 29a, 1995.
2. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, 'Glycoproteins and Disease', in *New Comprehensive Biochemistry*, Elsevier, Amsterdam, Vol. 30, 1996.
3. J. Montreuil, J.F.G. Vliegenthart, H. Schachter, 'Glycoproteins II', in *New Comprehensive Biochemistry*, Elsevier, Amsterdam, Vol. 29b, 1997.
4. H. Lis, N. Sharon, 'Protein Glycosylation, Structural and Functional Aspects', *Eur. J. Biochem.*, **218**, 1–27 (1993).
5. J. Hofsteenge, D.R. Müller, T. de Beer, A. Löffler, W.J. Richter, J.F.G. Vliegenthart, 'New Type of Linkage Between a Carbohydrate and a Protein: C-Glycosylation of a Specific Tryptophan Residue in Human RNase Us', *Biochemistry*, **33**, 13 524–13 530 (1994).
6. T. de Beer, J.F.G. Vliegenthart, A. Löffler, J. Hofsteenge, 'The Hexopyranosyl Residue that is C-Glycosidically Linked to the Sidechain of Tryptophan-7 in Human RNase Us is Alpha-mannopyranose', *Biochemistry*, **34**, 11 785–11 789 (1995).
7. G.J. Gerwig, J.P. Kamerling, J.F.G. Vliegenthart, 'Determination of the D and L Configuration of Neutral/Monosaccharides by High-resolution Capillary g.l.c.', *Carbohydr. Res.*, **62**, 349–357 (1978).
8. G.J. Gerwig, J.P. Kamerling, J.F.G. Vliegenthart, 'Determination of the Absolute Configuration of Monosaccharides in Complex Carbohydrates by Capillary g.l.c.', *Carbohydr. Res.*, **77**, 1–7 (1979).
9. J.F.G. Vliegenthart, H. van Halbeek, L. Dorland, 'The Applicability of 500-MHz High-resolution ^{1}H-NMR Spectroscopy for the Structure Determination of Carbohydrates Derived from Glycoproteins', *Pure Appl. Chem.*, **53**, 45–77 (1981).
10. J.F.G. Vliegenthart, L. Dorland, H. van Halbeek, 'High-resolution, ^{1}H-Nuclear Magnetic Resonance Spectroscopy as a Tool in the Structural Analysis of Carbohydrates Related to Glycoproteins', *Adv. Carbohydr. Chem. Biochem.*, **41**, 209–374 (1983).
11. J.P. Kamerling, J.F.G. Vliegenthart, 'High-resolution H-1-Nuclear Magnetic Resonance Spectroscopy

of Oligosaccharide–Alditols Released from Mucin-type *O*-glycoproteins', *Biol. Magn. Reson.*, **10**, 1–194 (1992).
12. B. Fournet, J. Montreuil, G. Strecker, L. Dorland, J. Haverkamp, J.F.G. Vliegenthart, J.P. Binette, K. Schmid, 'Determination of the Primary Structures of 16 Asialo-Carbohydrate Units Derived from Human Plasma Alpha 1-acid Glycoprotein by 360-MHz H-1-NMR Spectroscopy and Permethylation Analysis', *Biochemistry*, **17**, 5206–5214 (1978).
13. J.A. van Kuik, J.F.G. Vliegenthart, 'Databases of Complex Carbohydrates', *Trends Food Sci. Technol.*, **4**, 73–77 (1993) .
14. S. Doubet, K. Bock, D. Smith, A. Darvill, P. Albersheim, 'The Complex Carbohydrate Structure Database', *Trends Biochem. Sci.*, **14**, 475–477 (1989).
15. J.J.M. van Rooijen, A.F. Voskamp, J.P. Kamerling, J.F.G. Vliegenthart, 'Glycosylation Sites and Site-speicifc Glycosylation in Human Tamm–Horsfall Glycoprotein', *Glycobiology*, **9**, 21–30 (1999).
16. G.J. Gerwig, J.F.G. Vliegenthart, 'Analysis of Glycoproteins-derived Glycopeptides', in *Proteomics in Functional Genomics*, eds. P. Jollès, H. Jörnval, Birkhäuser Publishers, Basel, 2000.
17. B.R. Leeflang, L.M.J. Kroon-Batenburg, 'CROSREL: Full Relaxation Matrix Analysis for NOESY and ROESY NMR Spectroscopy', *J. Biomol. NMR*, **2**, 495–518 (1992).
18. L.M.J. Kroon-Batenburg, J. Kroon, B.R. Leeflang, J.F.G. Vliegenthart, 'Conformational Analysis of Methyl Beta-cellobioside by ROESY NMR spectroscopy and MD Simulations in Combination with the CROSREL Method', *Carbohydr. Res.*, **245**, 3787–3797 (1993).
19. J.P.M. Lommerse, L.M.J. Kroon-Batenburg, J. Kroon, J.P. Kamerling, J.F.G. Vliegenthart, 'Conformations and Internal Mobility of a Glycopeptide Derived from Bromelain Using Molecular Dynamics Simulations and NOESY Analysis', *J. Biomol. NMR*, **6**, 79–94 (1995).
20. T. De Beer, C.W.E.M. van Zuylen, B.R. Leeflang, K. Hård, R. Boelens, R. Kaptein, J.P. Kamerling, J.F.G. Vliegenthart, 'NMR Studies of the Free Alpha-subunit of Human Chorionic Gonadotropin; Structural Influences of *N*-glycosylation and the Beta-subunit on the Conformation of the Alpha-subunit', *Eur. J. Biochem.*, **241**, 229–242 (1996).
21. C.W.E.M. van Zuylen, J.P. Kamerling, J.F.G. Vliegenthart, 'Glycosylation Beyond the Asn78-linked GlcNAc Residues has a Significant Enhancing Effect on the Stability of the Alpha Subunit of Human Chorionic Gonadotropin', *Biochem. Biophys. Res. Commun.*, **232**, 117–120 (1997).
22. C.W.E.M. Thijssen-van Zuijlen, T. De Beer, B.R. Leeflang, R. Boelens, R. Kaptein, J.P. Kamerling, J.F.G. Vliegenthart, 'Mobilities of the Inner Three Core Residues and the Man (alpha 1–6) branch of the Glycan at Asn78 of the Alpha-subunit of Human Chorionic Gonadotropin are Restricted by the Protein', *Biochemistry*, **37**, 1933–1940 (1998).
23. G.M. Clore, A.M. Gronenborn, 'Theory of the Time-dependent Transferred Nuclear Overhauser Effect: Application to Structural Analysis of Ligand–Protein Complexes in Solution', *J. Magn. Reson.*, **53**, 423–442 (1983).
24. F. Ni, 'Recent Developments in Transferred NOE Methods', *Prog. NMR Spectrosc.*, **26**, 517–606 (1994).
25. C.P.J. Glaudemans, L. Lerner, G.D. Daves, Jr, P. Kovàc, R. Venable, A. Bax, 'Significant Conformational Changes in an Antigenic Carbohydrate Epitope upon Binding to a Monoclonal Antibody', *Biochemistry*, **29**, 10 906–10 911 (1990).
26. D.R. Bundle, H. Baumann, J.-R. Brisson, S.M. Gagné, A. Zdanov, M. Cygler, 'Solution Structure of a Trisaccharide–Antibody Complex – Comparison of NMR Measurements with a Crystal Structure', *Biochemistry*, **33**, 5183–5192 (1994).
27. V.L. Bevilacqua, D.S. Thomson, J.H. Prestegård, 'Conformational Aspects of Methyl Lactoside Bound to the Risin B-Chain Interpretation of Transferred Nuclear Overhauser Effects Facilitated by Spin Simulation and Selective Deuteration', *Biochemistry*, **29**, 5529–5537 (1990).
28. V.L. Bevilacqua, Y. Kim, J.H. Prestegård, 'Conformation of Methylmebioside Bound to the Risin B-Chain, as Determined from Transferred Nuclear Overhauser Effects', *Biochemistry*, **31**, 9339–9349 (1992).
29. R. Kaptein, 'Photo-CIDNP of Proteins', *Biol. Magn. Reson.*, **4**, 145–191 (1982).
30. P.J. Hore, R.W. Broadhurst, 'Photo-CIDNP of Biopolymers', *Prog. NMR Spectrosc.*, **25**, 326–345 (1993).
31. K. Hård, J.P. Kamerling, J.F.G. Vliegenthart, 'Application of Laser Photo-CIDNP for an Intact Glycoprotein in Solution', *Carbohydr. Res.*, **236**, 315–320 (1992).
32. H.-C. Siebert, S. André, G. Reuter, H.-J. Gabius, R. Kaptein, J.F.G. Vliegenthart, 'Effect of Enzymatic Desialylation of Human Serum Amyloid P Component on Surface Exposure of Laser Photo CIDNP (Chemically Induced Dynamic Nuclear Polarization)–Reactive Histidine, Tryptophan and Residues Tyrosine', *FEBS Lett.*, **371**, 13–16 (1995).
33. H.-C. Siebert, C.-W. Von der Lieth, R. Kaptein, J.F.G. Vliegenthart, C.S. Wright, H.-J. Gabius, 'Knowledge-based Homology Modeling and Experimental Determination of Amino Acid Side Chain Accessibility by the Laser Photo CIDNP. Approach in Solution: Lessons from the Small Sialidase of *Chlostridium perfringens*', *J. Mol. Modeling*, **2**, 351–353 (1996).
34. R.K. Harris, *Nuclear Magnetic Resonance Spectroscopy: A Physicochemical View*, Longman, Oxford, 1986.
35. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.

36. A.A. Derome, *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, 1987.
37. F.J.M. van der Ven, *Multidimensional NMR in Liquids, Basic Principles and Experimental Methods*, VCH Publishers, New York, 1995.
38. S. Braun, H.-O. Kalinowski, S. Berger, *100 and More Basic NMR Experiments; A Practical Course*, VCH Publishers, New York, 1996.
39. H. Kessler, M. Gehrke, C. Griesinger, 'Two-dimensional NMR Spectroscopy: Background and Overview of the Experiments', *Angew. Chem., Int. Ed. Engl.*, **27**, 490 (1989).
40. S.W. Homans, 'Oligosaccharide Conformation: Application of NMR and Energy Calculations', *Prog. NMR Spectrosc.*, **22**, 55 (1990).
41. A. Bax, S.W. Sparks, D.A. Torchia, 'Homonuclear Hartman–Hahn Experiments', in *Methods in Enzymology*, eds. N.J. Oppenheimer, T.L. James, Academic Press, San Diego, 151–168, Vol. 176, 1989.
42. H. Van Halbeek, 'NMR of Complex Carbohydrates', in *Frontiers of NMR in Molecular Biology*, eds. D. Live, I.M. Armitrage, D. Patel, Wiley-Liss, New York, 195–213, 1990.
43. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, New York, 1989.
44. L.R. Brown, B.T. Farmer, II, 'Rotating-frame Nuclear Overhauser Effect', in *Methods in Enzymology*, eds. N.J. Oppenheimer, T.L. James, Academic Press, San Diego, 199–216, Vol. 176, 1989.
45. T.-L. Hwang, A.J. Shaka, 'Cross Relaxation Without TOCSY: Transverse Rotating-frame Overhauser Effect Spectroscopy', *J. Am. Chem. Soc.*, **114**, 3157–3159 (1992).
46. H. Desvaux, P. Berthault, N. Birlirakis, M. Goldman, M. Piotto, 'Improved Versions of Off-resonance ROESY', *J. Magn. Reson., Ser. A*, **113**, 47–52 (1995).

Monosaccharides and Sugar Alcohol Analysis

Robert J. Sturgeon
Department of Biological Sciences, Heriot-Watt University, Edinburgh, UK

1 Introduction **1**
1.1 Monosaccharides 1
1.2 Sugar Alcohols 2
1.3 Nucleosides, Nucleotides and Nucleotide Sugars 3
2 Occurrence **4**
2.1 Composition 4
2.2 Depolymerization 4
3 Spectrophotometric Analysis (Colorimetric Assays) **5**
3.1 Reducing Sugars 5
3.2 Enzyme Assays 5
4 Chromatographic Separation of Monosaccharides, Sugar Alcohols, and Nucleotide Sugars **7**
4.1 Paper Chromatography and Thin-layer Chromatography 8
4.2 Gas–Liquid Chromatography 8
4.3 Liquid Chromatography 10
4.4 Electrophoresis 15
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Carbohydrates, including monosaccharides and sugar alcohols, are widely distributed in nature, being found in plants, animals and micro-organisms. They occur free, or in glycosidic linkage as polysaccharides or glycoconjugates (glycoproteins and glycolipids). It is important that suitable and sensitive methods are available for the analysis of these sugars, as in many cases only very small (microgram) amounts of material may be available. Polysaccharides and glycoconjugates must first be depolymerized to produce the monosaccharides quantitatively. The total carbohydrate released in these reactions may be determined by spectrophotometric or enzymic procedures.

Numerous separation and detection methods are available to measure the individual sugar components. Paper and thin-layer chromatography were two of the earliest techniques available, giving mainly qualitative values for individual sugar components. Major advances were made first in gas–liquid chromatography (GLC) and then with various types of liquid chromatography (LC) in the separation of individual sugars.

High-performance liquid chromatography (HPLC) and capillary electrophoresis (CE) are now two of the most commonly used techniques in sugar chromatography. Highly sophisticated detection systems, such as mass spectrometry (MS) for gas–liquid chromatographic separations, and spectrophotometric, fluorometric and electrochemical detection systems for LC of carbohydrates with suitable chromophores or fluorophores have been developed.

Carbohydrates have been derivatized for use either in precolumn or postcolumn chromatography, resulting in the detection limit in the most sensitive systems approaching yoctomole levels for individual sugars.

1 INTRODUCTION

1.1 Monosaccharides

Monosaccharides are classified into triose, tetrose, pentose, hexose, etc. according to the number of carbon atoms in the molecule. Thus xylose ($C_5H_{10}O_5$) is a pentose, glucose ($C_6H_{12}O_6$) is a hexose, and rhamnose ($C_6H_{12}O_5$) is a deoxyhexose.

Monosaccharides are individual monomeric carbohydrate units of polysaccharides or glycoconjugates (glycolipids or glycoproteins) and cannot be hydrolyzed to simpler compounds.

Monosaccharides are polyhydroxy aldehydes (aldoses) or polyhydroxy ketones (ketoses) and, as reducing sugars, exist virtually exclusively as cyclic hemiacetals. Hemiacetals are formed in reactions between carbonyl compounds (aldehydes or ketones) with alcohols, as illustrated in Scheme 1.

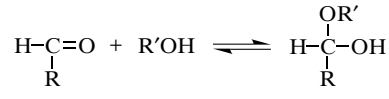

Scheme 1 Formation of a hemiacetal from an aldehyde and an alcohol.

Crystalline sugars exist in a single form, but in solution, sugars are present in both five-membered (furanose) and six-membered (pyranose) forms. These furanose and pyranose forms themselves each exist in two anomeric forms, α and β (epimeric at C-1 in the case of aldoses and at C-2 for ketoses). The epimeric forms arise because of the rapid equilibrium between themselves and the intermediate acyclic form. This is illustrated in Scheme 2 for D-glucose (Glc), for which sugar only

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

extremely small amounts of the furanose forms exist in aqueous solution. These structures can be described in the Haworth projection form. The pyranose ring is formed by reaction of the aldehyde group (carbon atom 1) with the hydroxyl group on carbon atom 5 of the same sugar molecule to form a hemiacetal. The corresponding furanose rings are formed between C-1 and the hydroxyl group on C-4.

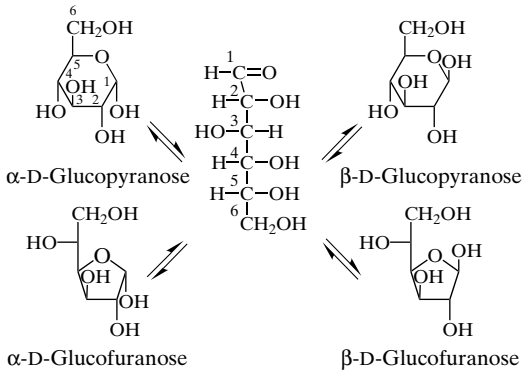

Scheme 2 Cyclic and acyclic forms of Glc.

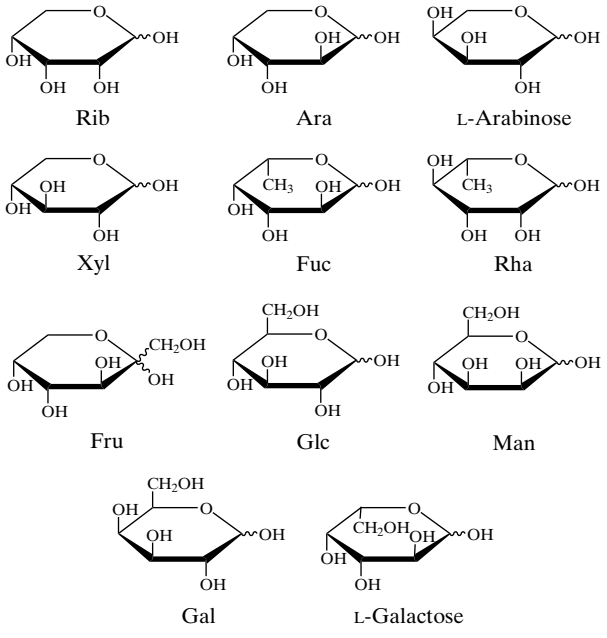

Scheme 3 Structures of commonly occurring neutral monosaccharides. Ara, D-arabinose; Fru, D-fructose; Fuc, L-fucose; Gal, D-galactose; Man, D-mannose; Rha, L-rhamnose; Rib, D-ribose; Xyl, D-xylose.

Proof of the structure of Glc, together with the elaboration of the configurational basis for the classification of aldose sugars, arises from the work of Emil Fischer over 100 years ago. The structures of

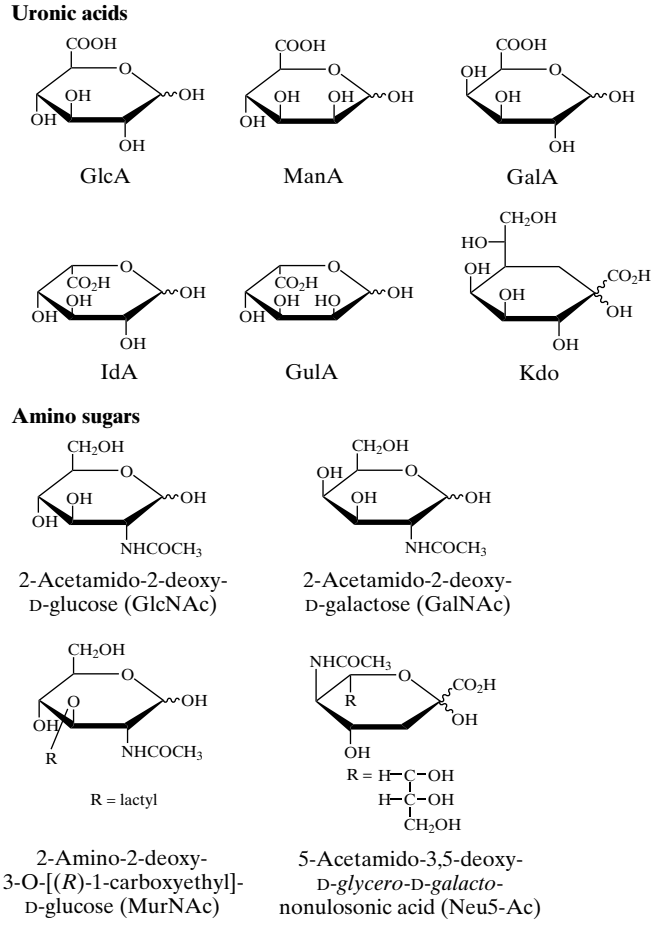

Scheme 4 Structures of uronic acids and amino sugars. GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; Neu5-Ac, *N*-acetylneuraminic acid; Kdo, 3-deoxy-D-*manno*-oct-2-ulosonic acid; GalA, D-galacturonic acid; GlcA, D-glucuronic acid; GulA, L-guluronic acid; IdA, L-iduronic acid; ManA, D-mannuronic acid; MurNAc, muramic acid.

commonly found monosaccharides and the established trivial names, including recommended three-letter abbreviations, of these neutral aldose sugars, are shown in Scheme 3. Sugars may occur with modifications, which may include the introduction of functionalities such as carboxyl (uronic acids), phosphate (sugar phosphates) or acetamido groups, as shown in Scheme 4. The current universal nomenclature for carbohydrates has been published.[1]

1.2 Sugar Alcohols

Sugar alcohols are those compounds obtained when the aldo or keto group of a monosaccharide is reduced to the corresponding hydroxy group. Alternative names for these compounds are polyols, polyalcohols, or polyhydric alcohols. Those sugar alcohols derived from aldoses are referred to as alditols. Alditols are named by changing the

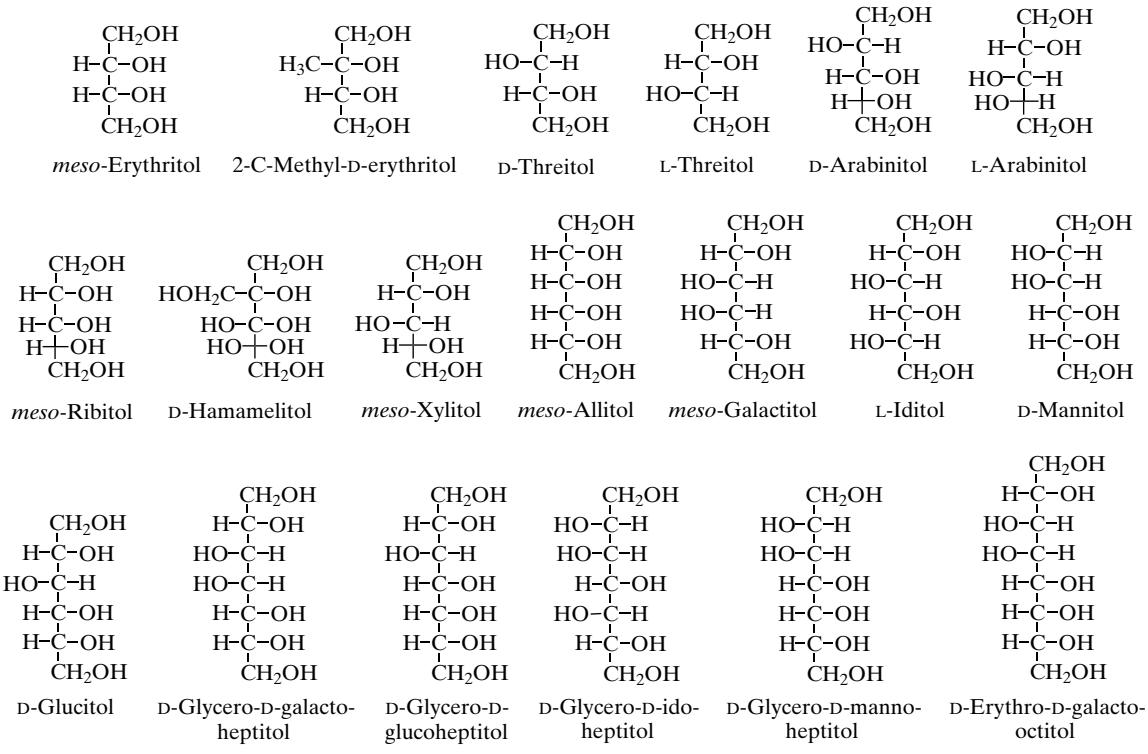

Scheme 5 Structures of commonly occurring sugar alcohols.

suffix '-ose' in the name of the corresponding aldose into '-itol'. Problems are encountered in the naming of sugar alcohols, due to the continuing use in some cases of trivial names which were given to some of these molecules when they were first isolated and characterized in the early history of natural products. In the time-honoured way, their names reflect their original source. Thus sorbitol was isolated from *Sorbus*, mannitol from manna, adonitol from *Adonis*, and perseitol from *Persea*, the avocado. In using current systematic chemistry and nomenclature, sorbitol is D-glucitol, mannitol is D-mannitol, adonitol is ribitol and dulcitol is galactitol. In examining different possible stereoisomeric configurations of various sugar alcohols, some have a plane of symmetry and as a result are meso (or internally compensated) compounds and are therefore optically inactive. These alditols are designated by the prefix *meso*-. Examples are *meso*-erythritol, *meso*-ribitol and *meso*-galactitol. However, the prefix D- or L- is given when the derivative of a meso form has become asymmetric by substitution, or in cases where there are more than four contiguous asymmetric carbon atoms present in the molecule. In Scheme 5, examples are given of commonly found sugar alcohols.

1.3 Nucleosides, Nucleotides and Nucleotide Sugars

Nucleosides are naturally occurring compounds of the glycosylamine type, derived from Rib or 2-deoxy-Rib and a pyrimidine (uracil, cytosine or thymine) or purine base (adenine or guanine); two examples are given in Scheme 6. Those nucleosides containing ribose are derived from ribonucleic acid, and those containing deoxyribose are derived from deoxyribonucleic acid. Nucleotides are phosphate esters of nucleosides, usually containing up to three phosphate groups (nucleotide mono-, di-, and triphosphates).

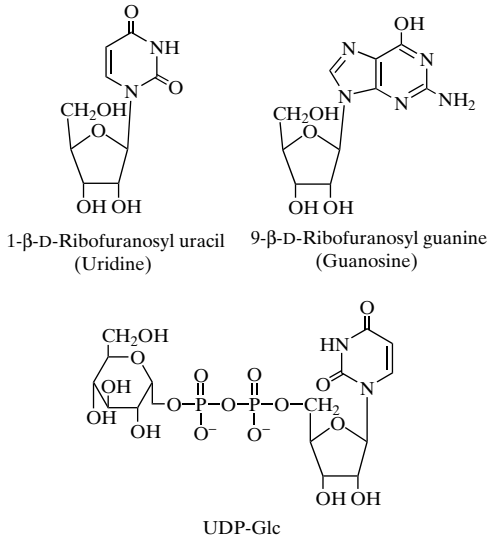

Scheme 6 Structures of the nucleosides uridine and guanosine, and UDP-Glucose.

Nucleotide sugars are formed in enzyme-catalyzed reactions from nucleotide triphosphates and sugar phosphates. In a typical reaction, uridine 5′-(α-D-glucopyranosyl diphosphate) (UDP-Glc) is generated from uridine triphosphate (UTP) and glucose 1-phosphate (Glc 1-P) (Equation 1):

$$\text{UTP} + \text{Glc 1-P} \longrightarrow \text{UDP-Glc} + \text{pyrophosphate} \quad (1)$$

The nucleotide sugars in turn are involved in carbohydrate biosynthetic reactions, when the monosaccharide is transferred to an acceptor molecule which may be a growing chain of a polysaccharide or a growing chain of a glycoprotein. A variety of nucleotide sugars, such as UDP-Glc, adenosine diphosphate D-glucose (ADP-Glc), guanosine diphosphate D-mannose (GDP-Man), thymidine diphosphate L-rhamnose (TDP-Rha), and cytidine diphosphate (CDP)-neuraminic acid are involved in highly specific reactions. For example, in the biosynthesis of sucrose, the glucose residue arises from UDP-Glc, transferred from fructose 6-phosphate (Fru 6-P), but in the biosynthesis of starch the glucose residue arises from ADP-Glc which is transferred to a growing linear chain of glucose residues. Examples of nucleosides, nucleotides and nucleotide sugars are given in Scheme 6.

2 OCCURRENCE

2.1 Composition

Monosaccharides are not normally found in significant amounts in the free form in either plants, animals or micro-organisms. Instead, they are found in vast abundance in the bound form as components of glycoconjugates (glycoproteins and glycolipids) and as polysaccharides. The most common building blocks for these polymers are the hexoses, hexuronic acids, and pentoses. Hirst[2] proposed that the hexose → uronic acid → pentose transformation occurs at the monosaccharide level. Subsequent incorporation of these sugars into polysaccharides was considered to take place by an unspecified mechanism. The involvement of nucleotide sugars in these transformations was first reported by Cardini et al.,[3] who proposed that a cofactor in the conversion of galactose 1-phosphate to Glc 1-P was UDP-Glc. The suggestion that compounds of the UDP-Glc type, i.e. other nucleotide sugars, could be involved in the formation of sugars and their subsequent incorporation into polysaccharides was then made by Buchanan et al.[4,5]

The polysaccharides are found widely in plant, animal and microbial sources, mainly as either storage materials which can be readily metabolized during growth of the organism (starch, seed galactomannans, glycogen) or as structural materials, such as cellulose in plants, chitin in *Crustaceae*. The classification of polysaccharides is now accepted to include examples such as proteoglycans and peptidoglycans, where polysaccharide chains are covalently linked to form sugar–amino acid linkages to proteins and peptides, respectively. Where polysaccharide chains are linked to lipid units, the molecules are referred to as lipopolysaccharides. The number of different monosaccharide residues found in these polysaccharides usually varies from one to three or four. In glycoproteins the protein backbone may contain a variable number of carbohydrate chains, each containing up to 20 sugar residues. In glycolipids, oligosaccharide units are linked to lipid moieties. Glycolipids and glycoproteins are frequently located bound in membranes of organisms, although water-soluble glycoproteins form the major components of animal sera.

The distribution of sugar alcohols covers plant, animal, and microbial sources, but those from plants have probably been studied in greater depth.[6,7] The discovery of sugar alcohols arose from the desire, over a century ago, to find sweetening agents other than honey or sugar cane. Among the earliest sugar alcohols isolated and characterized are glucitol (sorbitol), galactitol, and mannitol. In higher plants, the main site of sugar alcohol formation is the leaf where they arise as end products of photosynthesis. From the leaf, some of these products are translocated to the root, stem, fruits, and seeds.

2.2 Depolymerization

2.2.1 Acid Hydrolysis

In order to determine the composition of any polysaccharide, or glycoconjugate, it is necessary to obtain a quantitative estimation of the sugar constituents after hydrolysis of all glycosidic linkages. These polymers contain a range of sugars having a variety of glycosidic linkages, and the conditions required to achieve total depolymerization vary from one molecule to another. Glycosidic bonds differ in their susceptibility to acid hydrolysis. The rate of release of sugars is dependent on the position and anomeric configuration of the glycosidic linkage as well as on the identity of the monosaccharide. No acid hydrolysis procedure is capable of cleaving every linkage to give a quantitative yield of each monosaccharide. Some prior knowledge of the nature of the monosaccharides and the types of linkages involved help to make a choice.

For instance, the glycosidic linkages of polysaccharides containing monosaccharides in the furanose conformation and in addition keto sugars, are readily hydrolyzed under extremely mild conditions. Fructans are completely depolymerized with 0.1 M oxalic acid at 70 °C for 1 h.[8]

Neuraminic acid (sialic acid), a common monosaccharide constituent in animal glycoconjugates, is released quantitatively with 0.05 M sulfuric acid at 80 °C for 1 h, but is completed destroyed on heating in 1 M acid.[9] Successful release of neuraminic acid from serum glycoproteins has also been achieved using 0.025 M trifluoroacetic acid at 80 °C for 2 h,[10] whereas Kdo, a component of many bacterial lipopolysaccharides, is released by 1 M acetic acid at 40 °C for 6 h.[11] For most plant noncellulosic aldose-containing polysaccharides, including glucans, xylans, and galactomannans, conditions of 1 M sulfuric acid at 100 °C for 4–6 h will give complete hydrolysis with minimum loss of sugar,[12] or with 1 M trifluoroacetic acid at 120 °C for 1 h.[13] Cellulose can be hydrolyzed quantitatively, after first dispersing the polysaccharide in 72% sulfuric acid at room temperature followed by dilution to 2 M and heating at 100 °C. Carbohydrates containing uronic acid residues pose different problems, due to the greater resistance to hydrolysis of the uronosyl glycosidic bond. Incomplete hydrolysis results with the production of aldobiouronic acids and poor yields of the monosaccharide components. Improved yields are obtained by the incorporation of a prior hydrolysis step using formic acid.[14]

Amino sugars, such as *N*-acetyl derivatives, 2-acetamido-2-deoxy-Glc, and 2-acetamido-2-deoxy-Gal, are frequently found as components of glycoconjugates, especially glycoproteins containing *N*-glycan chains. During acid hydrolysis, *N*-deacetylation results, and the 2-amino-2-deoxyglycosides are resistant to hydrolysis. For complete depolymerization, treatment with 4 M hydrochloric acid at 100 °C for 6 h is necessary.[15] Minimal degradation occurs if sugars and amino sugars are released from polymeric material as methyl glycosides by heating in methanolic hydrogen chloride.[16]

2.2.2 Enzymic Hydrolysis

Hydrolysis of polysaccharides by enzymes is an attractive alternative to acid hydrolysis, because mild conditions of pH and temperature may liberate monosaccharide units in quantitative yield. Normally at least two or more enzymes are necessary to accomplish this. Knowledge of the specificity of the enzymes to be used is necessary in order to ensure complete depolymerization and the absence of any oligosaccharide material. Enzymes hydrolyzing polysaccharide material are classified into two types, the endoglycanases, which hydrolyze internal linkages and cause a rapid decrease in molecular size, and exoglycanases, whose action pattern is to release monosaccharide from the products of the endoglycanase reaction. α-Amylase is a well-characterised *endo*-α-(1 → 4)-D-glucanase which acts on starch and glycogen in a random manner to release oligosaccharides containing both α-(1 → 4)- and α-(1 → 6)-glucosyl residues.[17] Glucoamylase is an *exo*-α-(1 → 4)-D-glucosidase which acts by liberating Glc from the terminal nonreducing ends on hydrolysis of both α-(1 → 4) and α-(1 → 6) linkages.[18] In a similar way, inulin has been completely hydrolyzed to fructose and glucose on treatment with a fungal inulinase,[19] cereal β-glucans yield glucose only on treatment with a cellulase preparation,[20] and chitin can be quantitatively converted to 2-acetamido-2-deoxy-Glc with an *endo*-chitinase and *N*-acetyl-β-D-glucosaminidase.[21]

3 SPECTROPHOTOMETRIC ANALYSIS (COLORIMETRIC ASSAYS)

Many methods have been developed for the measurement of total carbohydrate in polysaccharides and glycoconjugates. These methods are usually nonspecific in relation to the measurement of individual monosaccharides, but generally will measure classes of sugars, such as pentoses, hexoses, uronic acids, deoxy sugars, and amino sugars. Frequently, the reactions are carried out in the presence of strong acid, which results in the hydrolysis of all glycosidic bonds, followed by the dehydration of the respective monosaccharides to furans, or pyrrole derivatives, which then condense with specific reagents to give colored products.[22] A number of methods useful for the measurement of total carbohydrate and alditols are listed in Table 1.[23–40] The methods are empirical and require calibration with known monosaccharides. It is important to appreciate the specificities of the methods which must be checked for interference by other classes of sugars, noncarbohydrate reducing agents, and protein. Several of the methods have proved suitable for the monitoring of carbohydrate eluates from LC columns.

3.1 Reducing Sugars

The amounts of reducing groups are frequently measured in aqueous extracts of carbohydrates, either before or after acid or enzymic hydrolysis. These assays are not specific for monosaccharides in the presence of reducing oligo or some low-molecular-weight carbohydrates. The most commonly used ones are based on modifications of the alkaline copper sulfate (Somogyi–Nelson reagent),[41] alkaline 3,5-dinitrosalicylic acid,[42] and alkaline 2,2′-bicinchoninate.[43]

3.2 Enzyme Assays

With the general availability of highly purified enzymes of known specificity, many monosaccharides and alditols can

Table 1 Chemical methods for the determination of monosaccharides and sugar alcohols

Method	Sugars determined	Refs.
Phenol–H_2SO_4	aldoses, ketoses, hexuronic acids, 6-deoxyhexoses	23
Orcinol–H_2SO_4	aldoses	24, 25
L-cysteine–H_2SO_4	6-deoxyhexoses	26
Orcinol–HCl	pentoses	27
L-cysteine–H_2SO_4	ketoses	28
Anthrone–H_2SO_4	aldoses, ketoses, hexuronic acids, 6-deoxyhexoses	29, 30
Resorcinol–H_2SO_4	aldoses	31
Carbazole–H_2SO_4	uronic acids	32
m-Phenylphenol–H_2SO_4	uronic acids	33
Pentane-2,4-dione	2-amino-2-deoxy-hexoses	34
4-(*N*,*N*′-dimethylamino) benzaldehyde	2-acetamido-2-de-oxyhexoses	35
Periodic acid–thiobarbituric acid	neuraminic acids	36
	polyols	37–40

be measured (Table 2).[44–59] The usefulness of enzymic techniques arises from the fact that, normally no prior separation from other carbohydrates is required, the reactions are carried out at pH values and temperatures unlikely to cause degradation of other products, and they convert the sugars to stoichiometric amounts of product. Most of the enzymes used in these procedures are commercially available, and those which are not may be readily isolated and purified. The experimental procedures for most of these methods have appeared in published collections.[60–62] Oxidases and dehydrogenases (oxidoreductases) are used in most of the procedures. Oxidases, which are usually highly specific, acting on only one substrate, require the presence of molecular oxygen and the monosaccharide is converted to the corresponding lactone and hydrogen peroxide. Hydrogen peroxide inhibits the forward reaction going to completion. In a coupled reaction, a second enzyme, usually peroxidase in the presence of an acceptor molecule, is used to destroy the hydrogen peroxide and produce a chromogen (Scheme 7). Whereas Glc oxidase oxidizes

Table 2 Enzymic estimation of monosaccharides and alditols

Sugar	Enzyme	EC number	Cofactor requirement	Ref.
L-Arabinose	β-Gal dehydrogenase	1.1.1.48	NAD^+	44
Xyl	Glc dehydrogenase	1.1.1.14	NAD^+	45
Gal	Gal oxidase	1.1.3.9	–	46
	β-Gal dehydrogenase	1.1.1.48	NAD^+	47
Gal	Fuc dehydrogenase	1.1.1.122	NAD^+	48
Glc	HK	2.7.1.1	ATP, $NADP^+$	49
	G6-PDH	1.1.1.49	ATP, $NADP^+$	49
	Glc dehydrogenase	1.1.1.147	NAD^+	50
	glucokinase	2.7.1.2	ATP, $NADP^+$	51
	G6-PDH	1.1.1.49	ATP, $NADP^+$	
	Glucose oxidase	1.1.3.4	–	52
Man	HK	2.7.1.1	ATP, $NADP^+$	49
	G6-PDH	1.1.1.49	ATP, $NADP^+$	49
	PGI	5.3.1.9	ATP, $NADP^+$	49
	PMI	5.3.1.8	ATP, $NADP^+$	49
Fru	HK	2.7.1.1	ATP, $NADP^+$	
	G6-PDH	1.1.1.49	ATP, $NADP^+$	49
	PGI	5.3.1.9	ATP, $NADP^+$	49
	fructose dehydrogenase	1.1.99.11	–	53
Fuc	Fuc dehydrogenase	1.1.1.122	NAD^+	54
Rha	Rha dehydrogenase		NAD^+	55
Neuraminic acid	*N*-acylneuraminate lyase	4.1.3.3		36
	lactate dehydrogenase	1.1.1.27	NADH	
Glycerol	glycerokinase		NAD^+	
	glycerokinase	2.7.1.30	NAD^+	56
	glycerol 3-phosphate dehydrogenase		ATP	57
Xylitol	glucitol dehydrogenase	?	NAD^+	58
Glucitol	glucitol dehydrogenase		NAD^+	58
Mannitol	mannitol dehydrogenase	1.1.1.138	NADPH	59

ATP, adenosine triphosphate; G6-PDH, glucose 6-phosphate dehydrogenase; HK, hexokinase; PGI, phosphoglucoisomerase; PMI, phosphomannoisomerase; NADH, reduced nicotinamide adenine dinucleotide; NADPH, reduced nicotinamide adenine dinucleotide phosphate.

Glc at C-1 to produce gluconolactone, Gal oxidase oxidizes Gal at C-6 to produce a C-6 aldehyde. Dehydrogenases require the addition of a cofactor, either nicotinamide adenine dinucleotide (NAD^+) or nicotinamide adenine dinucleotide phosphate ($NADP^+$). On oxidation of the substrate, for example, Fuc by Fuc dehydrogenase, an equimolar amount of cofactor is reduced and from its absorbance at 340 nm, the amount of sugar can be directly determined (Scheme 8).

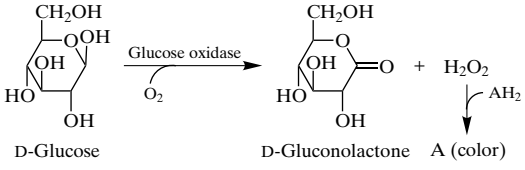

Scheme 7 The oxidation of Glc by Glc oxidase.

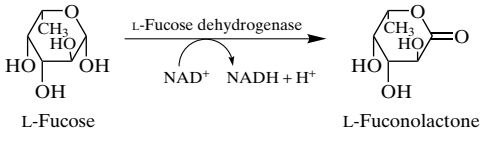

Scheme 8 The oxidation of Fuc by Fuc dehydrogenase.

Dehydrogenases frequently do not demonstrate absolute specificity for a substrate. Instead a number of structurally related monosaccharides may act as substrates. The enzyme Fuc dehydrogenase, purified from pig liver, acts not only on Fuc but also on Ara, L-lyxose, L-xylose, and L-Gal.[63] However, the latter sugars are not found commonly in biological extracts in the presence of Fuc. Similarly, Gal, D-Fuc and L-arabinopyranose are oxidized by Gal dehydrogenase.[44,64] When Gal and L-arabinose are present in the same sample, for instance in hydrolysates of polysaccharides such as L-arabino-D-galactans, the amount of Gal can be obtained from action of Gal oxidase, an enzyme with specificity only for Gal, and the amount of Gal plus L-arabinose can be obtained from the action of Gal dehydrogenase. Mixtures of the monosaccharides Glc, Fru and Man have been measured following the sequential addition of enzymes (Scheme 9).[49] All three monosaccharides are converted to the corresponding monosaccharide 6-phosphates by HK in the presence of ATP. G6-PDH is absolutely specific for the conversion of glucose 6-phosphate (Glc 6-P) to 6-phosphogluconate and NADPH. Sequential additions of PGI and PMI convert the corresponding Fru 6-P and mannose 6-phosphate (Man 6-P) to Glc 6-P, then to additional amounts of NADPH.

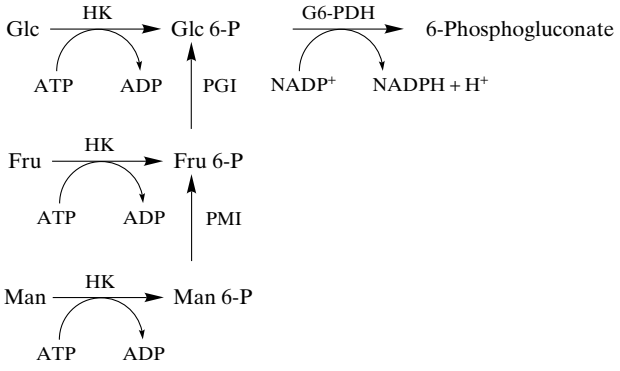

Scheme 9 The sequential enzymic determination of Glc, Fru, and Man.

4 CHROMATOGRAPHIC SEPARATION OF MONOSACCHARIDES, SUGAR ALCOHOLS, AND NUCLEOTIDE SUGARS

Much of carbohydrate analysis depends on the chromatographic separation and identification of these simple sugars, which themselves may have been isolated from polysaccharides or glycoconjugates.[65–68] From the earliest reports on the separation of sugars by paper chromatography (PC), the investigator now has the choice of a large number of techniques which, in chronological order, include thin-layer chromatography (TLC), GLC, LC, including HPLC, and electrophoretic techniques such as paper electrophoresis (PE) and CE.

The PC and TLC techniques, which are generally used as qualitative measures, are the least sensitive methods for the detection of carbohydrates, and require little outlay in equipment costs. The HPLC and CE techniques, when used with detection systems such as laser-induced fluorescence (LIF) of fluorescein derivatives, give the most sensitive detection, but require considerable outlay in capital equipment and running costs. The choice of technique to be used will depend greatly, among other things, on the number of samples routinely required for analysis and the amount of material available for the individual analyses. In the analysis of complex carbohydrates associated with glycoconjugates, 50–300 ng of a mixture of oligosaccharides containing up to six different monosaccharides may be available on release from a typical glycoprotein.[69] Currently, the most sensitive system used for the analysis of carbohydrates will use either HPLC or CE. Most carbohydrates are nonchromogenic, and this results in decreased sensitivity in ultraviolet (UV) or refractive index detection. Since the late 1980s, significant advances have been made both in separation and detection methods. This has been achieved by derivatization of the sugars to allow spectrophotometric analysis to be used in the detection systems.[70] These derivatives,

after separation, may now allow detection of sugars down to as little as 100 ymol (1×10^{-24} mol) in the case of LIF of fluorescein derivatives.[71]

4.1 Paper Chromatography and Thin-layer Chromatography

The procedures to be used for the separation of sugars by PC and TLC have been well-documented and reviewed.[60,67,72] Sample preparation is crucial, because inorganic ions, proteins and lipid materials cause interference with many of the organic solvent mixtures used in the separations. The procedures for TLC are generally superior to PC, because of faster separation times, a greater variety of supports, and increased sensitivity of detection. Microcrystalline cellulose, which has a similar capacity for resolution of sugars as does PC, is commonly used as the supporting material, although alternative systems using inorganic supports, such as silica gel, are used. Resolution, in this latter case, depends on both partition and adsorption effects. The addition of inorganic salts has been shown to improve the chromatographic behavior in many of the sugar separations on silica gel.[73–75]

For the resolution of mixtures of nucleotides and nucleotide sugars, an anion-exchange form of TLC using thin layers of poly(ethyleneimine) cellulose has been used.[76] Sugars form borate complexes, and these sugar complexes have been used in the separation of glucose, mannose and galactose-containing nucleotides of the same purine or pyrimidine base.[77]

Most of the solvent systems used in PC and TLC for the separation of monosaccharides can be used for the separation of sugar alcohols,[67] but distinction between these two groups may be difficult due to the similar R_f values, although Kremer[78] has developed a solvent system for the complete separation of the common monosaccharides and sugar alcohols by TLC, and Han and Robyt[79] have reported the separation of seven aldoses and their corresponding alditols. Alditols and some aldoses have been separated on *O*-(carboxymethyl) cellulose paper in the lanthanum, calcium and barium forms.[80] The TLC methodology lends itself readily to multiple development on chromatographic plates, and new supports have allowed the automated multiple development of high-performance thin-layer chromatography (HPTLC) analysis of sugars on hydrophilic layers.[81] A variety of monosaccharides, including pentoses, hexoses, heptuloses, deoxy-sugars, and acetamidodeoxyhexose, have been resolved on aminopropyl bonded silica plates in 70% acetonitrile.[82]

After chromatographic separation of sugars or sugar alcohols, it is customary to dip papers in reagents such as silver nitrate, anisidine–phthalate or periodate–pentane-2,4-dione.[60] Similar reagents, usually as sprays, are used for detection of these compounds on TLC plates. Some extremely sensitive detection methods have been developed. Reaction of aldoses with 4′-*N*,*N*-dimethyl-amino-4-azobenzene 1-sulfonyl hydrazine produces strongly chromophoric dabsylhydrazone derivatives which allow detection of 100 pmol of sugar.[83] Most aldoses, ketoses, uronic acids and amino sugars have been detected at levels as low as 250 pmol by fluorometric detection after treatment with a malonamide spray.[84] Amino sugars have been converted to the corresponding fluorescent dansyl derivatives before TLC separation.[85] Separated amino sugars on treatment with fluorescamine at pH 9.0 give highly fluorescent pyrrolines,[86] and nanogram amounts of sugars, sugar acids, lactones and polyols have been quantified after separation on HPTLC plates following treatment with 2,7-dichlorofluorescein.[87]

Nucleotides and nucleotide sugars do not normally require reagents for detection as they are conveniently detected after chromatography by examination of plates or papers under UV light.

4.2 Gas–Liquid Chromatography

GLC is widely used for the quantitative determination of monosaccharides, sugar alcohols and those components that have been derived by hydrolysis of polysaccharides and glycoconjugates. For polysaccharides and glycoconjugates depolymerization to the corresponding monomers is necessary prior to GLC (section 2.2). Total depolymerization frequently requires treatments with mineral acids. A major difficulty is the decomposition of the liberated monomers, if drastic conditions are necessary in order to hydrolyze the more resistant glycosidic bonds. Complete hydrolysis of polymeric carbohydrates with methanolic hydrochloric acid has been shown to give higher recoveries of both neutral sugars and uronic acids.[88]

Table 3 Derivatives used for the GLC of carbohydrates

Derivative[a]	Stationary phase	Detector[b]	Refs.
Alditol acetate		FID	89–91
TMS ethers		FTIR	92, 93
TMS oxime	SP 2250	FID	94
Trifluoroacetate		ECD	95
TMS methyl-glycoside	DB5 poly(5–95% methylsiloxane)	MS	96, 97
TMS methyloxime	SP 2100	MS	98
TFA alkyloxime	OV 225	MS	99–101
Aldononitrile acetate		MS	102–104

[a] TMS, trimethylsilyl; TFA, trifluoroacetate.
[b] FTIR, Fourier-transform infrared; FID, flame ionization detection; ECD, electron capture detection.

Simple sugars are polar and nonvolatile. Therefore, for use in monosaccharide analysis, derivatization is the necessary first step, in order to increase the volatility of the sugars. The most popular methods are the formation of alditol acetates, and trimethylsilylethers.[60,67] Trifluoroacetates and aldononitrile acetates have also been used. Some of the common derivatives of sugars used in GLC are listed in Table 3.[89–104] A single monosaccharide may produce several products on derivatization because in aqueous solution aldoses and ketoses exist in a dynamic equilibrium between the two ring (pyranose and furanose) and acyclic forms.[105] Unless the anomeric center, at C-1 for aldoses and C-2 for ketoses, is destroyed, for example by reduction to the corresponding alditols, up to four multiple peaks (two from the α- and β-pyranose forms and two from the corresponding furanose forms) may appear on the chromatograms. Although the production of multiple peaks on GLC will result in a complex chromatogram, the high efficiency provided by capillary columns usually gives satisfactory separations. Reduction of aldoses or ketoses with sodium borohydride or sodium borodeuteride eliminates the anomeric centers to produce alditols which are readily converted to stable acetates.[89,90] One of the limitations in using this system for the production of alditol acetates is that carboxyl groups, as in hexuronic acids, are not derivatized, and therefore these sugars do not produce chromatographic peaks. For polysaccharides, such as red-algal galactans, containing quantities of 3,6-anhydrogalactosyl residues, which are rapidly destroyed under harsh acidic conditions, a procedure has been developed to generate stable 3,6-anhydrogalactitol acetate, as an acid-stable 4-methylmorpholine-borane.[91,106] Several methodological changes, aimed at a simplification of the original alditol acetate derivatization method, have been reviewed.[90,107]

Other common derivatives, such as trimethylsilyl and trimethylacetyl, have simple work-up procedures, but both derivatives are unstable in the presence of moisture, making decomposition on storage a major problem. In the aldononitrile procedure, aldoses are reacted with hydroxylamine to produce the corresponding oximes. Acetylation of the aldoximes results in the dehydration of the oxime to a nitrile, and acetylation of all of the hydroxyl groups. Although the anomeric centers of the aldoses have been eliminated, thus giving simpler chromatograms, contaminating peaks arising from interaction between the hydroxylamine and acetylating reagents have been reported.[102,108] In the production of *O*-methyloxime acetates from aldoses, the original anomeric center is destroyed by the reaction with *O*-methyl hydroxylamine, but a new one is generated with the production of two products (the syn and anti isomers); this results in the production of multiple peaks on chromatograms.[109]

In addition to the GLC of simple monosaccharides and alditols, as described above, there is a major interest in the separation of partially methylated monosaccharides and their derivatives. This arises from well-established techniques for sugar linkage analysis of glycosidically-linked carbohydrates using methylation and other techniques.[16,110,111] Methylation analysis is the most commonly used procedure for linkage analysis, being based on the complete etherification of all free hydroxyl groups on carbohydrate oligomers or polymers. Those hydroxyl groups involved in inter-sugar glycosidic linkages, or which bear other constituents, are not involved in the methylation process. On subsequent hydrolysis of the carbohydrates, partially methylated aldoses are converted, after reduction and peracetylation, to a partially methylated alditol acetate (PMAA), prior to separation by GLC. In early reports, using this technique, monosaccharide derivatives were identified on the basis of their retention times, against standard PMAA derivatives.[112] With the resolving power of GLC coupled with the use of MS, it is possible to determine the positions of glycosidic attachment of monosaccharide units in a carbohydrate polymer, as well as identifying the ring form (pyranose or furanose) of each of the component sugars.[113,114] The mass spectra of these derivatives give characteristic fragmentation patterns, thus allowing identification of the original monomers and the nature of the glycosidic linkages. Under electron-impact/mass spectrometry (EI/MS), the fragmentation of PMAA is orderly, predictable, and easy to interpret.[115,116] It is not possible to distinguish stereoisomers by differences in their mass spectra while using EI/MS, but in most instances they can be distinguished by their GLC retention times.[117,118] Certain pairs of methylated sugars will yield alditols with the same substitution pattern (for example 2,3- and 3,4-*O*-methylpentose). They will be inseparable by GLC, but if the reduction at C-1 is carried out with sodium borodeuteride, the resulting MS patterns for these sugars will be distinguishable by relevant isotopic shifts in primary fragment ions.[119,120] Deuterium labeling has also been used in the identification of uronic acid residues in polymers. In order to overcome the well-known problem of the resistance to hydrolysis of glycosiduronic acids, reduction of the carboxyl group with sodium borodeuteride allows incorporation of deuterium atoms into the molecule ($-CO_2H \rightarrow -CD_2OH$) to give the corresponding labeled aldose. Those hexose derivatives formed from the uronic acid may then be identified by MS.

Chemical ionization/mass spectrometry (CI/MS) has also been used to identify PMAAs after GLC.[110–112] A number of gases, including methane, isobutane, and helium have been used in CI/MS procedures, but probably

the most popular one is ammonia.[110] In general, a superiority in sensitivity of chemical ionization (CI) detection over electron impact (EI) detection is observed. An ion-trap type of detector has been recommended for PMAA analysis.[113,116] By using micro-scale methylation analysis,[121] as little as 5–10 μg of glycoconjugate may allow complete analysis when using a multiple selected-ion monitoring system, which is reported to be about seven times more sensitive than standard GLC/MS.

4.3 Liquid Chromatography

For many years, until about 1980, LC for the analytical separation of carbohydrates was conducted on low-pressure systems, usually with an anionic resin as a support. However, advances in the design of supports, pumping systems and detection systems have resulted in a variety of new methods, including HPLC,[122–124] high-performance anion-exchange chromatography (HPAEC),[125] ligand-exchange chromatography (LEC),[126] and hydrophobic interaction chromatography (HIC).[127] The major differences achieved with these methods have been that the length of time required to separate simple monosaccharides has been reduced from several hours to several minutes. The improvements in the design of either precolumn or postcolumn derivatization of the carbohydrates with the corresponding detector systems has resulted in the original detection limits of nanomoles of monosaccharides being decreased to femtomoles or even lower. A general strategy for the chromatographic separation and monosaccharide composition of polysaccharides and glycoconjugates may include LEC and HPAEC in combination with pulsed amperometric detection (PAD), refractive index, or UV monitoring.[128] The identification of monosaccharides can be achieved on-line by coupling liquid chromatography/mass spectrometry (LC/MS).

4.3.1 Anion-exchange Chromatography

Using low-pressure systems, separations of monosaccharides as their borate complexes were achieved by Khym and Zyll,[129] and modifications to the system, involving fully automatic analysis of neutral monosaccharides and oligosaccharides on quaternary ammonium ion-exchange resins have been reported.[130,131] Hough et al. used a borate–chloride gradient to separate monosaccharides[132] and polyols.[133] As uronic acids are already charged, they can be separated on anion-exchange resins, in the absence of borate ions.[134] A wide variety of aldoses, ketoses, alditols, cyclitols and oligosaccharides have been measured in the effluents from borate anion-exchange columns after periodate oxidation.[135] Mixtures of monosaccharides and alditols have been successfully separated on anion-exchange resins in the molybdate form,[136] and postcolumn detection using cysteine–H_2SO_4 and periodate oxidation detected reducing sugars and alditols, respectively.

By far the most common method currently used for the analytical separation of monosaccharides, alditols and nucleotide sugars is HPAEC. This became possible with the development of quaternary-ammonium-bonded pellicular resins stable to operating conditions of high pH. The spherical pellicular resins contribute to the high resolution of the technique by locating all of the ionic groups on the surface, which results in low diffusion. Carbohydrates, in the form of polyhydric molecules, behave as weak acids, having pK_a values of 12–14.[137] Thus in basic solutions all monosaccharides, and even oligosaccharides, are negatively charged oxy anions. With increasing hydroxide concentration, several of the hydroxyl groups become ionized, with the C-2 being the most acidic. As the mutarotation of sugars is very rapid at high pH, individual anomeric forms of monosaccharides are no longer resolved. The resulting oxy anions can be chromatographed as anions, without prederivatization, with each monosaccharide yielding a single peak. The elution order of monosaccharides correlates with their pK_a values, and excellent separations have been obtained for neutral, amino-, and acidic monosaccharides derived from hydrolysis of many glycoconjugates,[138–141] as well as for the determination of polyols in food products.[142,143] Detection of the eluted sugars may be made by PAD, where sensitivities in the low picomole range can be achieved without derivatization. In this system, the carbohydrates are oxidized at the surface of a gold electrode. The resulting gold oxide layer is converted back to native gold on reversing the voltage on the electrode, and the potential is cycled back to the analytical voltage. The method is rapid and highly sensitive, with less than 1 nmol required for a typical monosaccharide separation.[144] Amino acids and peptides arising from the hydrolysis of glycoproteins can cause fouling at the working electrode surface.[145] A guard column and a newly developed separator have been designed to eliminate most of this interference.

Alternative electrodes, based on nickel–copper, nickel–chromium,[146] and silver,[147] have been developed for the detection of monosaccharides and sugar alcohols following HPAEC. The separation and PAD of carbohydrates and alditols are improved by the addition to the alkaline mobile phase of barium and strontium ions.[148] Peak symmetry and column efficiency are improved and significant enhancement of the response of the analytes in PAD at the gold working electrode are observed. These findings have been explained in terms of a rapid formation of adducts between polyhydroxy compounds and these divalent nonelectroactive

cations.[149] Aldoses and alditols, in their borate form, may also be separated on HPLC anion-exchange resins, although retention times tend to be long.[40] Detection is usually achieved by postcolumn derivatization.

4.3.2 Ligand-exchange Chromatography

LEC on microparticulate cation-exchange resins in metal ion forms has been applied to the separation of carbohydrates. Although mainly applicable for the separation of oligosaccharides, the technique will allow separations of monosaccharides and alditols.[123,124,150] The ability of neutral sugars to form complexes with metal ions coordinated to cation-exchange resins was demonstrated by Jones et al.[151] For LEC, the metal ions (usually Ca^{2+}, Ag^{2+}, or Pb^{2+}) are electrostatically immobilized on a strong cation exchanger. Water molecules bind to the metal ion, in relation to its coordination number and, when a mixture of carbohydrates is introduced into the mobile phase, they will displace some of the bound water molecules, forming a donor–acceptor complex.[152] Parameters such as carbohydrate structure, choice of metal ion, temperature, and additives to the mobile phase will affect the stability constants of carbohydrates and thus separation. Separations of pentoses, hexoses and their corresponding alditols have been reported in a study using seven different cations.[153] Good resolution of alditols had originally been achieved on cation-exchange resins in the Ba,[154] Ca,[155] and Li[156] forms. In a detailed study on the chromatography of 14 different alditols, effective separations were obtained with cation-exchange resin in its La form, using water as the eluant.[157] On sulfonated polystyrene residues in the europium(III) form with sodium hydroxide as the mobile phase, monosaccharides and alditols have longer retention times than oligosaccharides, whereas monosaccharides and alditols are selectively retained by the resin in the iron(III) form.[158]

4.3.3 Reversed-phase Liquid Chromatography

Reversed-phase liquid chromatography (RPLC) has been applied successfully to both small and large carbohydrate molecules. In this type of chromatography, nonpolar moieties of the solute and the stationary phase undergo noncovalent association in hydro-organic media. The stationary phase is a highly nonpolar surface of a solid support having covalently bound alkyl or aryl chains at the surface. Retention of carbohydrate species requires an interaction between two nonpolar groups, the nonpolar ligand bound to the surface of the solid support of the column, and the hydrophobic moiety of the solute. Basic theoretical concepts of RPLC have been the subject of several reviews.[127,159] Nonpolar stationary phases are used, usually consisting of microparticulate silica supports with covalently bound alkyl or aryl functional groups at the surface. The most widely used stationary phases include octadecyl (C_{18})-, octyl (C_8)-, and butyl (C_4)-coated sorbents. Underivatized carbohydrates are usually chromatographed using water as the mobile phase, but in general, monosaccharides which elute first are not resolved from one another.[160] Precolumn derivatization schemes have been widely exploited for the RPLC analysis of carbohydrates. The choice of type of derivative is usually made to introduce a chromophoric group onto the saccharides to aid detection of the separated molecules. Many of the chromophores are themselves hydrophobic and the conjugates are then readily separated by RPLC. Several reviews on precolumn derivatization have been published.[70,127] Chromophoric groups are covalently attached either to the carbonyl function of the monosaccharide or, less frequently, to hydroxyl or amino groups. In modifications involving the carbonyl groups, coupling is usually with amines to form imines (Schiff bases) or their more stable products by reductive amination using sodium cyanoborohydride to give substituted amines (Scheme 10). The derivatization reactions that have been developed are well-defined, quantitative, and nondestructive (when used on oligosaccharides). Excess reagents and byproducts produced during the reactions have to be removed prior to chromatography, and the introduced chemical groups must be stable.

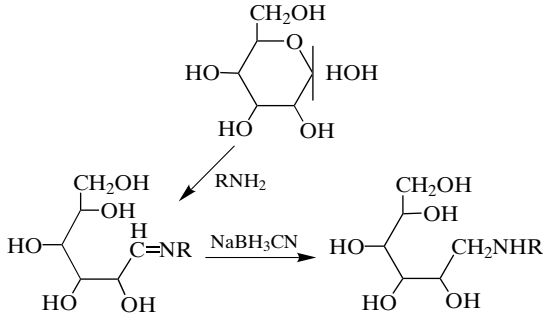

Scheme 10 Derivatization of aldoses by reductive amination.

The analysis of carbohydrates is hampered by the lack of chromophoric or fluorophoric functional groups. Many high-resolution separative techniques, such as HPLC, CE, and fluorophore-assisted carbohydrate electrophoresis (FACE), are not feasible without pre- or postcolumn derivatization of the carbohydrate. Introduction of a simple chromophore or fluorophore as a label at the reducing end, using the reductive amination technique, allows quantitation on a molar basis. The structures of a variety of primary amines and other selected molecules used for these applications are shown in Scheme 11.

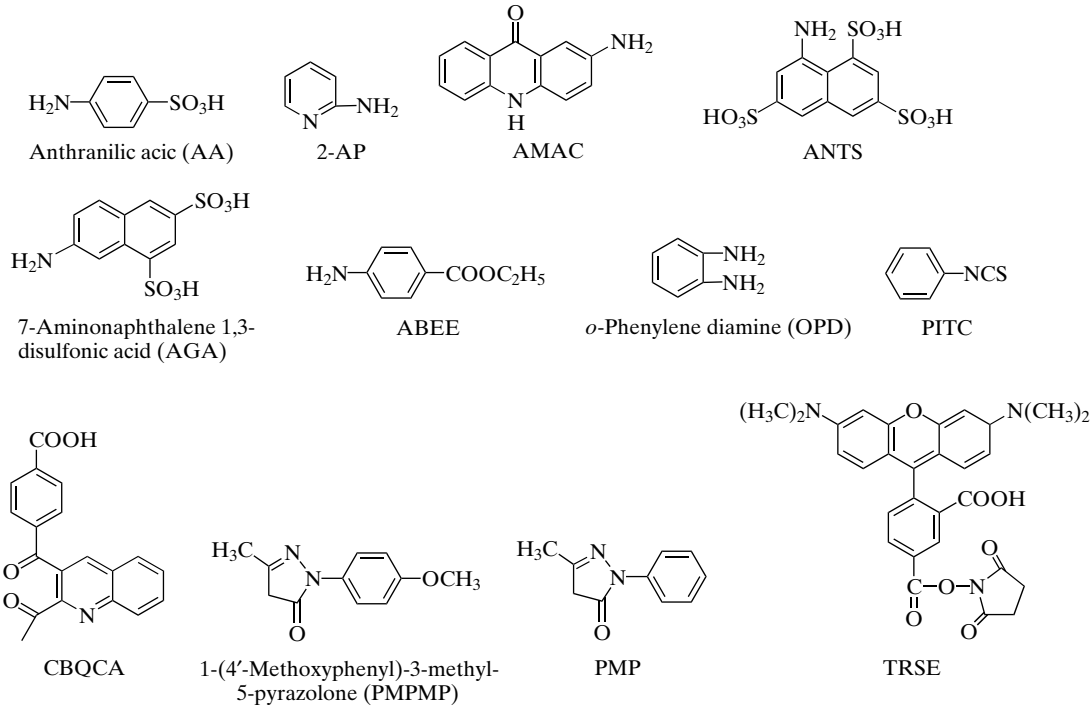

Scheme 11 Structures of molecules used for the introduction of chromophores and fluorophores into monosaccharides. AMAC, 2-aminoacridone; ABEE, aminobenzoic acid ethyl ester; ANTS, 8-aminonaphthalene-1,3,6-trisulfonic acid; 2-AP, 2-aminopyridine; CBQCA, 3-(4-carboxybenzoyl)-2-quinolone carboxaldehyde; TRSE, 5-carboxymethylrhodamine succinimidyl ester; PITC, phenyl isothiocyanate; PMP, 1-phenyl-3-methyl-5-pyrazolone.

Table 4 Precolumn derivatization and detection of carbohydrates

Carbohydrate detected	Reagent	Wavelength (nm)	Sensitivity	Refs.
Photometric detection				
Aldoses, amino sugars	2-AP	240		161, 162
Aldoses	PMP	245		163
Aldoses	1-(4′-methoxyphenyl)-3-methyl-5-pyrazolone (PMPMP)	249		164
Aldoses	ABEE	254	10–200 nmol	165
Aldoses	9-fluorenylmethylcarbonyl (FMOC) hydrazide	263	1–3 pmol	166
Aldoses	4′-*N*,*N*′-dimethylamino-4-azobenzene-1-sulfonyl hydrazine	460	10 pmol	167
Alditols	*per-O*-benzoyl	231	10 pmol	168
Amino sugars, amino sugar alcohols	*o*-phthaldehyde	254		169
Alditols, amino sugar alcohols	2-naphthoylimidazole	234		170
Fluorimetric detection				
Aldoses, amino sugars	2-aminobenzoic acid	230/425[a]	0.1 pmol	171
Aldoses, amino sugars	2-AP	320/400[a]	0.01–10 nmol	162
Aldoses, amino sugars	9-fluorenylmethylcarbonyl (FMOC) hydrazide	270/320[a]	0.05–0.4 pmol	166
Neuraminic acids, α-keto acids	1,2-diamino-4,5-methylenedioxybenzene	373/448[a]		172
Neuraminic acids, α-keto acids	*o*-phenylenediamine	232/420[a]	<2 pmol	173
Glycamines, amino sugars, amino sugar alcohols	PITC	340/450[a]	1 pmol	174
Alditols, amino sugar alcohols	2-naphthoylimidazole	234/374[a]		170

[a] Excitation/emission wavelengths shown.

4.3.3.1 Precolumn Derivatization Sensitive detection and high resolution of monosaccharides can be obtained by a judicious choice of derivatization to introduce stable chemical groups. Strict chemical techniques are required for the fulfilment of reaction conditions. The derivatization reaction should be quantitative and excess reagents and byproducts, formed during the reactions must be easily removed. Precolumn derivatizations by reductive amination are widely used in HPLC and CE for the separation of sugars. A selection of some of the more popular derivatives and their methods of detection is given in Table 4.[161–174]

In a two-step reaction (Scheme 12) an intermediate glycamine or 1-amino-1-deoxyalditol is first formed, on interaction of the reducing monosaccharide with the ammonium ion and sodium cyanoborohydride. The glycamine is subsequently derivatized with, for example, PITC to produce a phenylthiocarbamyl derivative,[174] or with *o*-phthalaldehyde.[169]

4.3.4 Ion-pair Reversed-phase Liquid Chromatography

Ion-pair RPLC has been widely used for the separation of a number of acidic oligosaccharides but, under suitable conditions, the technique may be applied to the separation of simple sugars, sugar acids and amino sugars. Carbohydrates that are weakly acidic and exist as anions in strongly alkaline mobile phases can be separated as ion pairs on addition of hydrophobic quaternary ammonium counter-ions to the mobile phases and using polymeric stationary phases.[175] Nucleotides and nucleotide sugars have also been separated using this technique.[176] Tetrabutylammonium hydrogen sulfate, which is very hydrophobic, was used as a counter-ion. The method allows the direct separation and detection of the 4′-epimeric-uridine 5′-diphosphate sugars (UDP-Gal and UDP-Glc) down to 40 pmol.

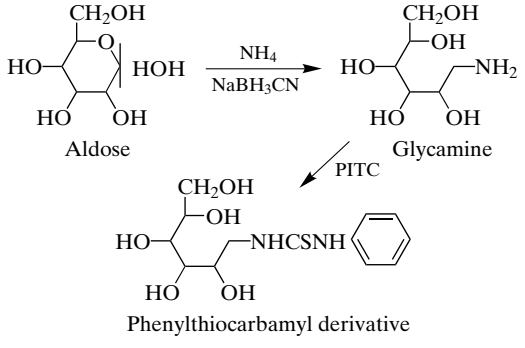

Scheme 12 Derivatization of aldoses in a two-step reaction via the glycamine (1-amino-1-deoxy) derivative.

4.3.5 Hydrophilic Liquid-interaction Chromatography

Chromatography, involving the use of a polar stationary phase with a less polar mobile phase was originally termed normal-phase partition chromatography. Alpert[177] has proposed the term hydrophilic interaction chromatography (HILIC), when applied to chromatography driven by polar forces. This then differentiates normal-phase partition chromatography from reversed-phase (hydrophobic) chromatography. Recent progress in HILIC has been published.[178] The stationary phases commonly used in HILIC are of the microparticulate and macroporous silica type, having average particle diameters of 3–5 μm, and also sorbents in which polar phases are chemically bonded to silica gel. A few synthetic polymer supports containing chemically bonded polar phases are also available. Some examples of these stationary phases and the types of sugars capable of separation on them are given in Table 5.[170,179–183]

Detection systems used in earlier work on HILIC involved the use of refractive index detectors, which were relatively insensitive and their sensitivity to changes in solvent composition, in the case of gradient elution with organic solvents, limited their use.

Although in the case of isocratic elution the detection limit is below 25 ng for sugars,[123] evaporative light-scattering detectors give superior sensitivity.[179,180] Precolumn derivatization to introduce chromophoric or fluorescent groups into monosaccharides or alditols (Table 4) is now widely used to increase the sensitivity of detection. Some postcolumn derivatization methods suitable for HILIC are described in the following section.

4.3.5.1 Postcolumn Derivatization Methods In LC, post-column derivatization plays an important role in the automation of carbohydrate analysis. The mobile phase is constantly eluted from the solid-phase column, on which the separation is achieved. By continuously mixing a flow of reagent solution into this eluant, followed by reaction

Table 5 Silica, silica-based and polymer-based packings used in HILIC of monosaccharides

Packing	Sugars separated	Ref.
Silica gel	aldoses, alditols	179
Silica gel	alditol *per*naphthoates	170
Diol-modified silica	aldoses, alditols	180
Aminopropyl silica	aldoses, alditols	180
Amine-bonded vinyl alcohol copolymer	monosaccharides, oligosaccharides derivatized with ABEE	181
Cyclodextrin-bonded silica	aldoses, alditols	182
Vinylpyridinium polymer	aldoses	183

for an appropriate length of time, and at an appropriate temperature, the separated components of the sample are converted to derivatives that are successively monitored at a detector.

Some of the earliest methods to be used in the post-column detection of carbohydrates include those methods in which carbohydrates including mono-, oligo-, and polysaccharides, are converted to furfurals in the presence of strong acids, which are condensed with chromogenic reagents, such as phenol, orcinol, cysteine, anthrone, resorcinol, and carbazole. Those methods listed in Table 1 that were originally designed for manual operations in detecting carbohydrates,[23–32] have been modified for automatic analyzers.[184] The major drawback in the use of such reagents is that they require the use of corrosive strong acids.

In order to eliminate methods requiring the use of corrosive acids, a large number of photometric, fluorimetric and electrochemical methods have been developed for postcolumn detection of monosaccharides and alditols following LC.[184] A wide range of reagents and methods, some of which are listed in Table 6,[40,185–194] are now available for the detection of sugars. A well-recognized property of reducing sugars is their ability to reduce solutions of salts of copper(II) to copper(I).[41] Addition of chelating agents such as 2,2′-bicinchoninate or bis(phenanthroline) form chelates with the copper(I) ion, which in the former case can be measured photometrically[185] and in the latter case electrochemically.[186]

Alditols cannot be detected with these reagents because they possess no reducing properties. The usual procedure for the detection of alditols involves periodate oxidation

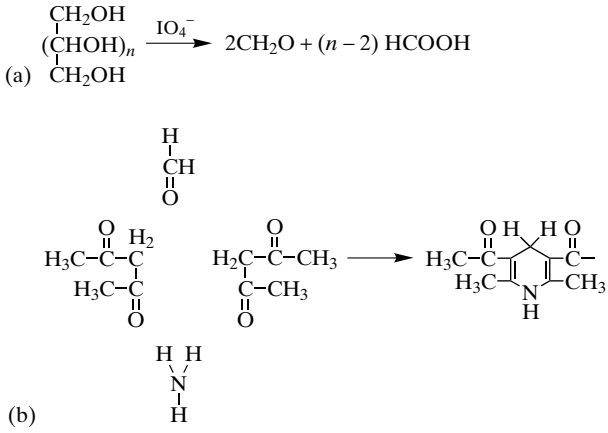

Scheme 13 (a) Periodate oxidation and (b) pentane-2,4-dione reaction for the detection of polyols.

of the samples to release formaldehyde, followed by the Hantzsch reaction to produce pyridine derivatives for photometric or fluorimetric detection. The two-step reaction involves, first the quantitative liberation of formaldehyde on periodate oxidation of the $-CHOH-CH_2OH$ or $-CO-CH_2OH$ groups (Scheme 13a) followed by the condensation of the formaldehyde with pentane-2,4-dione and ammonia to give the pyridine derivative (Scheme 13b). Detection is achieved by monitoring the absorption of the substituted pyridine.[40] The method is ideally suited for the determination of alditols separated in the anion-exchange mode as borate complexes. Fluorimetric monitoring gives a sensitivity of 0.5 nmol of alditol. As the initial oxidation step with periodate is carried out at high pH, aldoses or ketoses will also

Table 6 Postcolumn derivatization and detection of carbohydrates

Carbohydrate detected	Reagent	Wavelength (nm)	Sensitivity (nmol)	Ref.
Photometric detection				
Aldoses	Cu^{2+}-2,2′-bicinchoninate	562	0.1	185
Aldoses	anisyltetrazolium chloride	520	0.1	186
Aldoses	Purpald	550	100	187
Aldoses	2-cyanoacetamide	280		188
Glycamines, amino sugars	ninhydrin	550	500	189
Alditols	periodate–pentane-2,4-dione	410		40
Fluorimetric detection				
Aldoses	1,2-diaminoethane	390/470[a]	1	190
Uronic acids	1,2-diaminoethane	390/470[a]	1	191
Aldoses	2-cyanoacetamide	330/380[a]	0.01	192
Alditols	periodate–pentane-2,4-dione	410/503[a]	0.5	40
Electrochemical detection				
Aldoses	Cu^{2+}-bis(phenanthroline)		0.001	193
Aldoses	1,2-diaminoethane		0.001	194

[a] Excitation/emission wavelengths shown.
Purpald, 4-amino-3-hydrazino-5-mercapto-1,2,4-triazole.

release formaldehyde, but only 1 mol per mole instead of 2 mol per mole for alditols. In one of the earliest reports on the use of fluorogenic reagents being used in the detection of reducing sugars, solutions of carbohydrates heated with 1,2-diaminoethane (ethylenediamine) produced a strong fluorescence at 470 nm, with an excitation maximum at 390 nm. This has been developed as a selective fluorimetric method for the determination of mono- and oligosaccharides[190] and uronic acids.[191] Increased sensitivity, down to 1 pmol of monosaccharide, could be achieved when the reaction products were electrochemically oxidized on a glassy carbon electrode.[194] 2-Cyanoacetamide has proved to be a very versatile fluorogenic molecule, because aldoses heated in weakly alkaline solutions with this reagent produce an intense fluorescence at 380 nm.[192] The reaction products also absorb strongly at 280 nm, and can also be detected by electrolysis at a glass carbon electrode, thus allowing any one of three detection methods to be used. The reaction conditions for the photometric and fluorimetric detection differ and have been optimized separately. For separations on HPLC columns with acetonitrile as eluant, only UV monitoring is possible with this derivative.

Reducing sugars when heated with strong alkali and aromatic acid hydrazides produce yellow colors, which have been used as the basis of detection systems. The reagent Purpald, which was originally developed as a spray for the detection of reducing sugars on TLC plates, can be used for the postcolumn detection of aldoses.[187] The reaction is carried out in the presence of sodium hydroxide and hydrogen peroxide to produce a tetrazine–triazole fused ring condensate (Scheme 14).

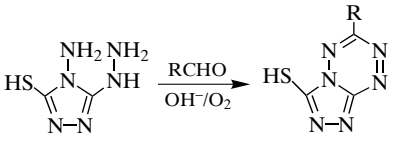

Scheme 14 Reaction of Purpald with an aldose.

Amino sugars are present in the total acid hydrolysates of most glycoproteins and proteoglycans. These sugars exist in macromolecules such as *N*-acetyl hexosamines, but the *N*-acetyl group is hydrolyzed under the acid conditions. The amino sugars can be chromatographed and detected in addition to the amino acids, on conventional amino acid analyzers with ninhydrin reagents. The same reagent may be used for the determination of monosaccharides which have been converted to glycamines (Scheme 12) prior to ion-exchange chromatography.[189]

4.4 Electrophoresis

4.4.1 Capillary Electrophoresis

CE, which had earlier been shown to give fast, automated and highly efficient separations of peptides, proteins and nucleic acids, has been developed into a technique for the analysis of mono- and oligosaccharides and glycoproteins. The CE method now provides a powerful alternative to HPLC for the detection of carbohydrates. Its popularity is reflected in the number of recent review articles dealing with current methods and general principles.[195–199] Carbohydrates are separated in CE, under the influence of an applied electric field (5–30 kV), by migrating in capillary columns (0.5–1.5 m long, 50–100 μm internal diameter), which span two buffer reservoirs and which are filled with buffer solutions. Total capillary volumes can be 10 μL, and injection volumes can be in the range of a few nanoliters. Factors influencing the mobility of carbohydrates include the applied voltage, the charge on the sugar molecule (positive, negative, or neutral), the pH and ionic strength of the buffer, the presence of buffer additives (surfactants, ion-pairing agents), and the operating temperature inside the capillary, as well as its length and diameter.[200] Capillary columns are generally constructed from fused silica and carry negative charges on the inner surface due to the presence of silanol groups, which are ionized above pH 3.0. Two forces – electro-osmotic flow and electrophoresis – govern the mobility of the carbohydrates. The electro-osmotic flow drives all species, irrespective of charge, from the injector, through the capillary, to the detector. This flow results from the cations in the buffer being attracted to the negatively charged silanol groups on the surface of the capillary. Under an applied high voltage, the cations migrate from the anode to the cathode, resulting in a bulk flow of electrolytes (electro-osmotic flow). The electrophoretic force, which also influences the separation of the species, is considered to be weaker.

The separation of both neutral and charged sugars may be achieved by either of two modes. The most common mode uses normal polarity (positive polarity) in which a neutral or basic buffer is used, sample application is at the anode, and detection is at the cathode. Therefore sugars bearing a positive charge will elute first, followed by neutral sugars. Negatively charged sugars are retarded because of electro-osmotic flow. In the second system, reversed polarity (negative polarity) acidic buffers are used. Electro-osmotic flow is reduced or eliminated, because of loss of charge on the silanol groups. Electrophoresis is then the dominant force, and only the anionic species migrate towards the anode. Electro-osmotic flow may also be eliminated by coating the capillaries with poly(vinyl alcohol). As only a relatively small number of monosaccharides (amino sugars, uronic

acids, sugar phosphates)[201] contain either anionic or cationic functional groups, separate methods have to be used to ensure electromigration of neutral monosaccharides. These methods include ionization of hydroxyl groups at high pH, complexation of vicinal or alternative hydroxyl groups with borate or alkaline earth metal ions, derivatization with reagents containing ionizable functions and partitioning into micelles in a pseudo-stationary phase (micellar electrokinetic chromatography). Similarly to HPLC, the detection of separated carbohydrates may use refractive index detectors. As most carbohydrates lack chromophoric or fluorophoric functions, precolumn derivatization is necessary to introduce these functions in order to achieve the greatest level of sensitivity. Depending on the choice of reagent, the derivatives may themselves be charged.

Nonderivatized monosaccharides can be monitored by direct UV detection after CE, following on-column complexation with borate at high pH, although the sensitivity is relatively low.[202] Indirect detection methods, which are available for nonchromophoric sugars, are based on the displacement of the chromophore in the background electrolyte by the analyte molecule, resulting in negative peaks. Indirect UV detection using sorbate as the background electrolyte is frequently used.[203] Of a number of other background electrolytes examined, 1-naphthyl acetic acid has been shown to be a suitable carrier buffer at pH 12, for detection of sugars at 222 nm, giving both an increased separation efficiency and improvement in detection limit, compared to sorbate.[204] Tryptophan and *N*-benzyl cinchonidium chloride have also been used successfully as background chromophores.[205] The PAD of nonderivatized sugars has been developed for CE.[206]

Carbohydrates are readily detectable after precolumn derivatization, giving products suitable for UV and fluorescence detection. This is the preferred approach for the electrophoresis and detection of monosaccharides. General methods to achieve derivatization include reductive amination,[207] condensation of carboxylic groups on saccharides with amines,[208] or condensation with PMP.[209] The chemistries involved are shown in Schemes 10, 15 and 16.

The use of 2-AP as a label for monosaccharides, after reductive amination, was originally developed for the precolumn labeling of monosaccharides for HPLC (Table 4). From these *N*-2-pyridyl-glycamines derivatives, monosaccharides give a rapid separation in borate buffer at pH 10.5.[207] 2-AP reacts satisfactorily with aldoses, alduronic acids, and *N*-acetamido-hexoses, but not with ketoses. An alternative number of carbohydrate labels based on aminobenzene, including 4-aminobenzoic acid,[210] ethyl 4-aminobenzoate,[211] and 4-aminobenzonitrile[212] give derivatives with ketoses. On-column detection of 6-aminoquinoline derivatives at 245 nm, using borate based electrolytes, has been used to give very sensitive detection of monosaccharides in wood-derived hemicelluloses.[213] The minimal detectable amount is reported to be a few femtomol with concentration limits of detection of the order of 10^{-6} M.

Other commonly used derivatives for precolumn labeling of monosaccharides, using reductive amination chemistry, include ANTS and 8-aminopyrene-1,3,6-trisulfonic acid (APTS). Two major advantages shown by these derivatives are that, as they exhibit very high extinction coefficients in the UV range, they can be detected photometrically at very low levels by LIF, and that the negative

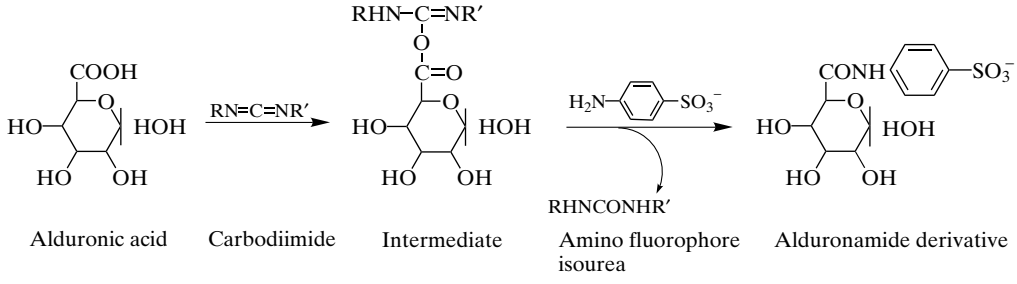

Scheme 15 Derivatization of uronic acids with carbodiimide coupling to amino compounds.

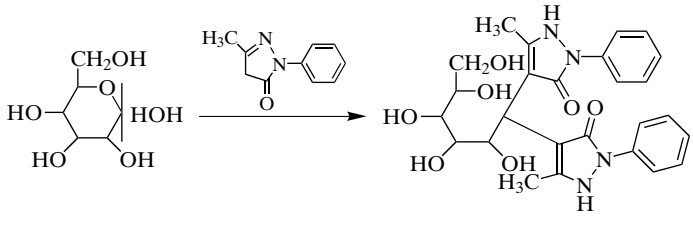

Scheme 16 Conversion of reducing sugars to PMP derivatives.

charges on the sulfonic acid groups are so strong that they are dissociated even in acidic media. In the analysis of the monosaccharide composition of glycoproteins, APTS has been used to derivatize the monosaccharide components. Neuraminic acid was separately derivatized with 9-aminoacridone.[214]

Isoindole derivatives of monosaccharides that give very sensitive labels are obtained from a two-step derivatization process. In the first step, sugars are converted to the corresponding glycamines by reductive amination (Scheme 17) prior to condensation with CBQCA in the presence of the cyanide ion.[215] The highly fluorescent isoindole derivatives of monosaccharides can be determined at attomole (10^{-18} mol) levels by LIF. The advantages and disadvantages of fluorescence and LIF have been examined in detail.[216]

In another two-step derivatization method for monosaccharides, neutral sugars are converted to glycamines by reductive amination (Scheme 12). Amino sugars are reduced to the corresponding 2-amino-2-deoxyalditols. These aminated sugars are then labeled with TRSE (Scheme 18). Difficulties have been reported with the stability of these derivatives, but, nevertheless, successful separations of hexoses derived from mammalian glycoproteins have been achieved with a concentration detection limit of 5×10^{-11} M.[217]

The formation of amide bonds from the condensation of carboxyl groups and amino groups in the presence of carbodiimide is a well-documented chemical process. Amide bond formation between alduronic acid residues and amines, such as 4-aminobenzene sulfonic acid (sulfanilic acid), and 7-aminonaphthalene-1,3-disulfonic acid (ANDSA) have been used to produce labeled carbohydrates (Scheme 15). The labeled products are strongly anionic at low or high pH levels. UV detection at 247 nm results in mass detection limits of 30 fmol for sulfanilic acid and 15 fmol for the ANDSA derivatives.[208] Under mildly basic conditions, PMP reacts with aldoses to form bis-PMP derivatives (Scheme 16). These derivatives, originally developed for HPLC, are used in capillary electrophoretic separations of mono- and oligosaccharides.[209]

4.4.2 *Fluorophore-assisted Carbohydrate Electrophoresis*

The FACE method is based, first, on the attachment of fluorescent tags to the reducing end of carbohydrates using standard reductive amination techniques prior to the separation of the derivatives on polyacrylamide (PA) gel slabs.[218–220] This technique has been used mainly for routine oligosaccharide profiling and sequencing of a variety of carbohydrates, as well as for monosaccharide analysis. The fluorophore most often used is ANTS.[221] Neutral and acidic monosaccharide derivatives exhibit net negative changes, imparted from the three sulfonic acid groups, and can be separated by electrophoresis and detected by their fluorescence. The alternative fluorophore, AMAC, which confers no charge on monosaccharide derivatives, is also used. In this case, borate ions, which complex with vicinal hydroxyl groups of monosaccharides, impart a negative

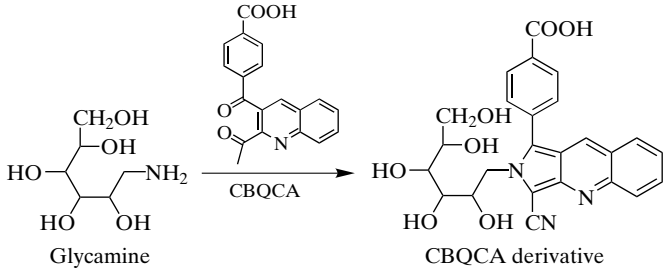

Scheme 17 Conversion of aldoses to isoindole derivatives.

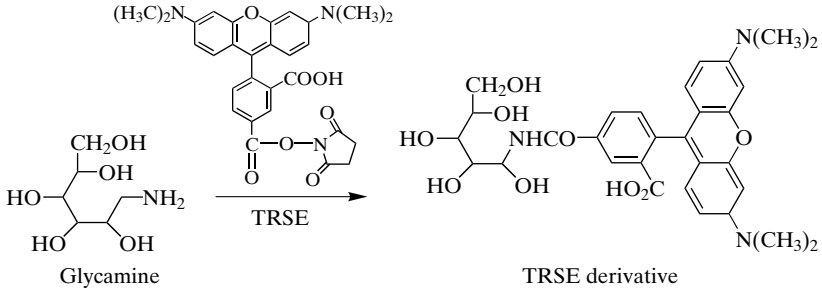

Scheme 18 Conversion of glycamines into TRSE derivatives.

charge for electrophoresis.[202] These derivatives are analyzed by polyacrylamide gel electrophoresis (PAGE). High-resolution separations are achieved by virtue of the electrophoretic buffer systems used. These are so-called stacking systems that are also called discontinuous, moving boundary, multizone, or multiphase. The PA gel inhibits convection. The resultant electrofluorograms are viewed on a graphics display unit after imaging with a suitable electronics camera.

Some ANTS–fluorophore derivatives of monosaccharides, oligosaccharides and low-molecular-weight polysaccharides have been separated by PAGE. For monosaccharides, high-resolution separations of some pairs of isomers, anomers and epimers have been achieved.

ABBREVIATIONS AND ACRONYMS

ABEE	Aminobenzoic Acid Ethyl Ester
ADP-Glc	Adenosine Diphosphate D-glucose
AMAC	2-Aminoacridone
ANDSA	7-Aminonaphthalene-1,3-disulfonic Acid
ANTS	8-Aminonaphthalene-1,3,6-trisulfonic Acid
APTS	8-Aminopyrene-1,3,6-trisulfonic Acid
Ara	D-Arabinose
ATP	Adenosine Triphosphate
CBQCA	3-(4-Carboxybenzoyl)-2-quinolone Carboxaldehyde
CDP	Cytidine Diphosphate
CE	Capillary Electrophoresis
CI	Chemical Ionization
CI/MS	Chemical Ionization/Mass Spectrometry
ECD	Electron Capture Detection
EI	Electron Impact
EI/MS	Electron-impact/Mass Spectrometry
FACE	Fluorophore-assisted Carbohydrate Electrophoresis
FID	Flame Ionization Detection
Fru	D-Fructose
Fru 6-P	Fructose 6-Phosphate
FTIR	Fourier-transform Infrared
Fuc	L-Fucose
Gal	D-Galactose
GalA	D-Galacturonic Acid
GalNAc	*N*-Acetylgalactosamine
GDP-Man	Guanosine Diphosphate D-Mannose
Glc	D-Glucose
GLC	Gas–Liquid Chromatography
GlcA	D-Glucuronic Acid
GlcNAc	*N*-Acetylglucosamine
Glc 1-P	Glucose 1-Phosphate
Glc 6-P	Glucose 6-Phosphate
GulA	L-Guluronic Acid
G6-PDH	Glucose 6-Phosphate Dehydrogenase
HIC	Hydrophobic Interaction Chromatography
HILIC	Hydrophilic Interaction Chromatography
HK	Hexokinase
HPAEC	High-performance Anion-exchange Chromatography
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
IdA	L-Iduronic Acid
Kdo	3-Deoxy-D-*manno*-oct-2-ulosonic Acid
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LEC	Ligand-exchange Chromatography
LIF	Laser-induced Fluorescence
Man	D-Mannose
ManA	D-Mannuronic Acid
Man 6-P	Mannose 6-Phosphate
MS	Mass Spectrometry
MurNAc	Muramic Acid
NAD^+	Nicotinamide Adenine Dinucleotide
NADH	Reduced Nicotinamide Adenine Dinucleotide
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	Reduced Nicotinamide Adenine Dinucleotide Phosphate
Neu5-Ac	*N*-Acetylneuraminic Acid
PA	Polyacrylamide
PAD	Pulsed Amperometric Detection
PAGE	Polyacrylamide Gel Electrophoresis
PC	Paper Chromatography
PE	Paper Electrophoresis
PGI	Phosphoglucoisomerase
PITC	Phenyl Isothiocyanate
PMAA	Partially Methylated Alditol Acetate
PMI	Phosphomannoisomerase
PMP	1-Phenyl-3-methyl-5-pyrazolone
Purpald	4-Amino-3-hydrazino-5-mercapto-1,2,4-triazole
Rha	L-Rhamnose
Rib	D-Ribose
RPLC	Reversed-phase Liquid Chromatography
TDP-Rha	Thymidine Diphosphate L-Rhamnose
TLC	Thin-layer Chromatography
TRSE	5-Carboxymethylrhodamine Succinimidyl Ester
UDP-Glc	Uridine 5′-(α-D-Glucopyranosyl Diphosphate)
UTP	Uridine Triphosphate
UV	Ultraviolet
Xyl	D-Xylose
2-AP	2-Aminopyridine

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis • Glycolipid Analysis • Glycoprotein Analysis: General Methods • Glycoprotein Analysis: Using Nuclear Magnetic Resonance • Proteoglycan and Acidic Polysaccharide Analysis

Clinical Chemistry **(Volume 2)**
Glucose Measurement

Food **(Volume 5)**
Dietary Fiber Analysis as Non-starch Polysaccharides • Enzyme Analysis and Bioassays in Food Analysis • Starch Analysis in Food

Pulp and Paper **(Volume 9)**
Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis

REFERENCES

1. 'Nomenclature of Carbohydrates', *Carbohydr. Res.*, **297**, 1–92 (1997).
2. E.L. Hirst, 'Recent Progress in the Chemistry of Pectic Materials and Plant Gums', *J. Chem. Soc.*, **208**, 70–78 (1942).
3. C.E. Cardini, R. Caputto, A.C. Paladini, L.F. Leloir, 'Uridine Diphosphate Glucose: The Coenzyme of Galactose–Glucose Isomerization', *Nature (London)*, **165**, 191–192 (1950).
4. J.G. Buchanan, J.A. Bassham, A.A. Benson, D.F. Bradley, M. Calvin, L.C. Dans, M. Goodman, P.M. Hayes, V.H. Lynch, L.T. Norris, A.T. Wilson, 'The Path of Carbon in Photosynthesis XVII. Phosphorus Compounds as Intermediates in Photosynthesis', in *Phosphorus Metabolism*, eds. W.D. McElroy, B. Glass, John Hopkins Press, Baltimore, 440–466, Vol. II, 1952.
5. J.G. Buchanan, V.H. Lynch, A.A. Benson, D.F. Bradley, M.F. Calvin, 'The Path of Carbon in Photosynthesis, XVIII. The Identification of Nucleotide Coenzymes', *J. Biol. Chem.*, **203**, 935–945 (1953).
6. E. Beck, H. Hopf, 'Branched-chain Sugars and Sugar Alcohols', in *Methods in Plant Biochemistry*, eds. P.M. Dey, J.B. Harborne, Academic Press, London, 235–289, Vol. 2, 1990.
7. R.L. Bieleski, 'Sugar Alcohols', in *Plant Carbohydrates I. Intracellular Carbohydrates*, eds. F.A. Loewus, W. Tanner, Springer-Verlag, Berlin, 158–192, 1982.
8. G.O. Aspinall, E.L. Hirst, E.G.V. Percival, R.G.J. Telfer, 'Studies on Fructosans, Part IV. A Fructosan from *Dactylis glomerata*', *J. Chem. Soc.*, 337–342 (1953).
9. J.F. Codington, K.B. Linsley, C. Silber, 'Removal of Sialic Acid from Glycoproteins by Chemical Methods and Determination of Sialic Acids', *Methods Carbohydr. Chem.*, **VII**, 226–232 (1976).
10. T. Tsegenidis, N.K. Karamanos, 'Sample Clean-up and Analysis of *N*-Acetyl and *N*-Glycolylneuraminic Acids in Blood Serum and Tissue Specimen by HPLC', *J. Liquid Chromatogr.*, **21**, 793–802 (1998).
11. W.S. York, A.G. Darvill, M. McNeil, P. Albersheim, '3-Deoxy-D-*manno*-2-octulosonic Acid (KDO) is a Component of Rhamnogalacturonan II, a Pectic Polysaccharide in the Primary Cell Wall of Plants', *Carbohydr. Res.*, **138**, 109–126 (1985).
12. R.R. Selevendran, J.F. March, S.C. Ring, 'Determination of Aldoses and Uronic Acid Content of Vegetable Fiber', *Anal. Biochem.*, **96**, 282–292 (1979).
13. P. Albersheim, D.J. Nevins, P.D. English, A. Karr, 'A Method for the Analysis of Sugars in Plant Cell-wall Polysaccharides by Gas–Liquid Chromatography', *Carbohydr. Res.*, **5**, 340–345 (1967).
14. B. Radhakrishnamurthy, G.S. Berenson, 'Identification of Uronic Acids in Mucopolysaccharides', *Arch. Biochem. Biophys.*, **101**, 360–362 (1963).
15. R.G. Spiro, 'Study of the Carbohydrates of Glycoproteins', *Methods Enzymol.*, **28**, 3–43 (1973).
16. R.L. Merkle, I. Poppe, 'Carbohydrate Composition Analysis of Glycoconjugates by Gas–Liquid Chromatography – Mass Spectrometry', *Methods Enzymol.*, **230**, 1–15 (1994).
17. F.R. Dintzis, C.C. Harris, 'Starch Determination in Some Dietary Fiber Sources', *Cereal Chem.*, **58**, 467–470 (1981).
18. B.J. Catley, 'Isolation and Enzymic Determination of Pullulan', *Methods Carbohydr. Chem.*, **X**, 105–110 (1994).
19. T. Nakamura, S. Maruki, S. Nakatsu, S. Ueda, 'Studies on Microbial Inulase', *J. Agric. Chem. Soc. J.*, **52**, 581–587 (1978).
20. C.W. Bamforth, '*Penicillium funiculosum* as a Source of β-Glucanase for the Estimation of Barley β-Glucan', *J. Inst. Brew.*, **89**, 391–392 (1983).
21. G.W. Gooday, 'Enzymic Determination of Chitin', *Methods Carbohydr. Chem.*, **X**, 81–84 (1994).
22. Z. Dische, 'Colorimetric Methods', *Methods Carbohydr. Chem.*, **I**, 477–514 (1962).
23. M. Dubois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, 'Colorimetric Method for Determination of Sugars and Related Substances', *Anal. Chem.*, **28**, 350–356 (1956).
24. R.J. Winzler, 'Determination of Serum Glycoproteins', *Methods Biochem. Anal.*, **2**, 279–311 (1955).

25. L. Svennerholm, 'The Quantitative Estimation of Cerebrosides in Nervous Tissue', *J. Neurochem.*, **1**, 42–53 (1956).
26. Z. Dische, L.B. Shettles, 'A Specific Color Reaction for Methyl Pentoses and a Spectrophotometric Micromethod for their Determination', *J. Biol. Chem.*, **175**, 595–603 (1948).
27. A.H. Brown, 'Determination of Pentose in the Presence of Glucose', *Arch. Biochem.*, **11**, 269–278 (1946).
28. Z. Dische, A. Devi, 'A New Colorimetric Method for the Determination of Ketohexose in the Presence of Aldoses, Ketoheptoses, and Ketopentoses', *Biochim. Biophys. Acta*, **39**, 140–144 (1960).
29. E. Van Handel, 'Determination of Fructose and Fructose-yielding Carbohydrates with Cold Anthrone', *Anal. Biochem.*, **19**, 193–194 (1967).
30. B.L. Somani, J. Khanade, R. Sinha, 'A Modified Anthrone–Sulphuric Acid Method for the Determination of Fructose in the Presence of Certain Proteins', *Anal. Biochem.*, **167**, 327–330 (1987).
31. S.A. Barker, B.W. Hatt, P.J. Somers, 'The Effect of Areneboronic Acids on the Alkaline Conversion of D-glucose into D-fructose', *Carbohydr. Res.*, **26**, 41–53 (1973).
32. T. Bitter, H.M. Muir, 'A Modified Uronic Acid–Carbazole Reaction', *Anal. Biochem.*, **4**, 330–334 (1962).
33. N. Blumenkrantz, G. Asboe-Hansen, 'A New Method for Quantitative Determination of Uronic Acids', *Anal. Biochem.*, **54**, 484–489 (1973).
34. C.J.M. Rondle, W.T.J. Morgan, 'Determination of Glucosamine and Galactosamine', *Biochem. J.*, **61**, 586–589 (1955).
35. J.L. Reissig, J.L. Strominger, L.F. Leloir, 'A Modified Colorimetric Method for the Estimation of *N*-Acetylamino Sugars', *J. Biol. Chem.*, **217**, 959–966 (1955).
36. G. Reuter, R. Schauer, 'Enzymic Methods for Sialic Acid Analysis', *Methods Carbohydr. Chem.*, **X**, 29–39 (1994).
37. J. Sanchez, 'Colorimetric Analysis of Alditols in Complex Biological Samples', *J. Agric. Food Chem.*, **46**, 157–160 (1998).
38. J.C. Speck, 'Periodate Oxidation', *Methods Carbohydr. Chem.*, **I**, 441–445 (1962).
39. O. Samuelson, H. Stromberg, 'Separation of Alditols and Aldoses by Partition Chromatography on Ion-exchange Resins', *Carbohydr. Res.*, **3**, 89–96 (1966).
40. S. Honda, M. Takahashi, S. Shimada, K. Kakehi, S. Ganno, 'Automated Analysis of Alditols by Anion-exchange Chromatography with Photometric and Fluorometric Post-column Derivatization', *Anal. Biochem.*, **128**, 429–437 (1983).
41. J.F. Robyt, W.J. Whelan, *Starch and its Derivatives*, 4th edition, Chapman and Hall, London, 432–470, 1968.
42. P. Bernfeld, 'Amylases, α and β', *Methods Enzymol.*, **1**, 149–158 (1955).
43. S. Waffenschmidt, L. Jaenicke, 'Assay of Reducing Sugars in the Nanomole Range with 2,2′-Bicinchoninate', *Anal. Biochem.*, **165**, 337–340 (1987).
44. J. Melrose, R.J. Sturgeon, 'An Enzymic Assay for L-arabinose Using β-D-galactose Dehydrogenase; its Application in the Assay of α-L-arabinofuranosidase', *Carbohydr. Res.*, **118**, 247–253 (1983).
45. C.P. Price, K. Burling, 'Enzymic Determination of D-xylose Using Glucose Dehydrogenase', *Methods Carbohydr. Chem.*, **X**, 41–44 (1994).
46. D. Amaral, F. Kelly-Falcoz, B.L. Horecker, 'Galactose Oxidase of *Polyporus circinatus*', *Methods Enzymol.*, **9**, 87–92 (1966).
47. R.J. Sturgeon, 'Enzymic Determination of D-galactose', *Methods Carbohydr. Chem.*, **VIII**, 131–133 (1980).
48. R.J. Sturgeon, 'Enzymic Determination of L-galactose', *Methods Carbohydr. Chem.*, **X**, 17–19 (1994).
49. R.J. Sturgeon, 'Enzymic Determination of D-glucose, D-fructose and D-mannose', *Methods Carbohydr. Chem.*, **VIII**, 135–137 (1980).
50. C.P. Price, 'Enzymic Determination of D-glucose Using Glucose Dehydrogenase', *Methods Carbohydr. Chem.*, **X**, 21–24 (1994).
51. D.A. Scott, C.R. Goward, P.M. Hammond, T. Atkinson, M.D. Scawen, C.P. Price, 'Enzymic Determination of D-glucose Using Glucose Kinase', *Methods Carbohydr. Chem.*, **X**, 25–28 (1994).
52. J. Okuda, I. Miwa, K. Maeda, K. Tokui, 'Rapid and Sensitive Colorimetric Determination of the Anomers of D-glucose with D-glucose Oxidase, Peroxidase and Mutarotase', *Carbohydr. Res.*, **58**, 267–270 (1977).
53. M. Ameyama, K. Matsushita, E. Shinagawa, O. Adachi, 'Enzymic Determination of D-fructose', *Methods Carbohydr. Chem.*, **X**, 7–10 (1994).
54. H. Schachter, 'Enzymic Determination of L-fucose', *Methods Carbohydr. Chem.*, **X**, 11–13 (1994).
55. F. Pittner, P.L. Turecek, 'The Application of Immobilised alpha-L-rhamnosidase and L-rhamnose Dehydrogenase in the Analysis of L-rhamnose and alpha-Rhamnosides', *Appl. Biochem. Biotechnol.*, **16**, 15–24 (1987).
56. R.J. Sturgeon, 'Enzymic Method for the Determination of Glycerol: Chain Lengths of Hexans', *Methods Carbohydr. Chem.*, **VIII**, 73–76 (1980).
57. M. Ameyama, E. Shinagawa, K. Matsushita, O. Adachi, 'Enzymic Determination of Glycerol', *Methods Carbohydr. Chem.*, **X**, 53–57 (1994).
58. R.J. Sturgeon, 'Enzymic Method for the Determination of the Degree of Polymerisation of Glucans and Xylans', *Methods Carbohydr. Chem.*, **VIII**, 77–80 (1980).
59. R.J. Sturgeon, 'Enzymic Determination of D-mannitol', *Methods Carbohydr. Chem.*, **X**, 59–61 (1994).
60. R.J. Sturgeon, 'Monosaccharides', in *Methods in Plant Biochemistry*, eds. P.M. Dey, J.B. Harborne, Academic Press, London, 1–37, Vol. 2, 1990.

61. *Methods Carbohydr. Chem.*, **X**, 1–39 (1994).
62. *Methods Enzymatic Anal., Metabolites I: Carbohydrates*, 3rd edition, eds. J. Bergmeyer, M. Grassl, Verlag Chemie, Weinheim, Vol. 6, 1984.
63. H. Schachter, J. Sarney, E.J. McGuire, S. Roseman, 'Isolation of Diphosphopyridine Nucleotide-dependent Fucose Dehydrogenase from Pork Liver', *J. Biol. Chem.*, **244**, 4785–4792 (1969).
64. R.J. Sturgeon, 'Enzymic Determination of L-arabinose', in *Methods of Enzymatic Analysis*, 3rd edition, eds. J. Bergmeyer, M. Grassl, Verlag Chemie, Weinheim, 427–431, Vol. 6, 1984.
65. M. Ghebregzabher, S. Rufini, B. Monaldi, M. Lato, 'Thin Layer Chromatography of Carbohydrates', *J. Chromatogr.*, **127**, 133–162 (1976).
66. L.A.T. Verhaar, B.F.M. Kuster, 'Liquid Chromatography of Sugars on Silica-based Stationary Phases', *J. Chromatogr.*, **220**, 313–328 (1981).
67. S.C. Churms, 'The Carbohydrates', in *CRC Handbook of Chromatography*, eds. G. Zweig, J. Sherma, CRC Press, Boca Raton, 1–272, 1982.
68. S.C. Churms, 'Recent Developments in the Chromatographic Analysis of Carbohydrates', *J. Chromatogr.*, **500**, 555–583 (1990).
69. E.F. Hounsell, 'Glycoprotein Analysis in Biomedicine', in *Methods in Molecular Biology*, ed. E.F. Hounsell, Humana Press. Totowa, New Jersey, Vol. 14, 1993.
70. S. Hase, 'Precolumn Derivatization of Chromatographic and Electrophoretic Analysis of Carbohydrates', *J. Chromatogr. A*, **720**, 173–182 (1996).
71. M.J. Davies, E.F. Hounsell, 'Carbohydrate Chromatography – Towards Yoctomole Sensitivity', *Biomed. Chromatogr.*, **10**, 285–289 (1996).
72. L. Hough, J.K.N. Jones, 'Chromatography on Paper', *Methods Carbohydr. Chem.*, **I**, 21–31 (1962).
73. Y.S. Ovodov, E.V. Evtushenko, V.E. Voskovsky, R.G. Ovodova, T.F. Soloveva, 'Thin Layer Chromatography of Carbohydrates', *J. Chromatogr.*, **26**, 111–115 (1967).
74. M. Ghebregzabher, S. Rufini, G. Ciuffini, M. Lato, 'A Two-dimensional Thin Layer Chromatographic Method for Screening Carbohydrate Anomalies', *J. Chromatogr.*, **95**, 51–58 (1974).
75. T. Mezzetti, M. Lato, S. Rufini, G. Ciuffini, 'Thin Layer Chromatography of Oligosaccharides with Tungstic or Molybdic Acid as Impregnant', *J. Chromatogr.*, **63**, 329–342 (1971).
76. E. Randerath, K. Randerath, 'Resolution of Complex Nucleotide Mixtures by Two-dimensional Anion-exchange Thin Layer Chromatography', *J. Chromatogr.*, **16**, 126–129 (1964).
77. H. Carminatti, S. Passeron, M. Dankert, E. Recondo, 'Separation of Sugar Nucleotides, Phosphoric Esters and Free Sugars with Solvents Containing Borates of Organic Bases', *J. Chromatogr.*, **18**, 432–438 (1965).
78. B.P. Kremer, 'Improved Method for the Thin Layer Chromatographic Identification of Alditols', *J. Chromatogr.*, **166**, 335–337 (1978).
79. N.S. Han, J.F. Robyt, 'Separation and Detection of Sugars and Alditols on Thin Layer Chromatograms', *Carbohydr. Res.*, **313**, 135–137 (1998).
80. L. Bilisics, L. Petruš, 'Chromatographic Separation of Alditols and Some Aldoses on *O*-(Carboxymethyl)-cellulose Paper in the Lanthanum, Calcium and Barium Forms', *Carbohydr. Res.*, **146**, 141–146 (1986).
81. G. Lodi, C. Bighi, V. Brandolini, E. Menziani, B. Tosi, 'Automated Multiple Development HPTLC Analysis of Sugars on Hydrophilic Layers. 2. Diol Layers', *J. Planar Chromatogr. – Mod. TLC*, **10**, 31–37 (1997).
82. L.W. Doner, L.M. Biller, 'High Performance Thin-layer Chromatographic Separation of Sugars – Preparation and Application of Amino-propyl Bonded-phase Silica Plates Impregnated with Monosodium Phosphate', *J. Chromatogr.*, **287**, 391–398 (1984).
83. J.-K. Lin, S.-S. Wu, 'Synthesis of Dabsylhydrazine and its use in the Chromatographic Determination of Monosaccharides by Thin-layer and High-performance Liquid Chromatography', *Anal. Chem.*, **59**, 1320–1326 (1987).
84. S. Honda, Y. Matsuda, K. Kakehi, 'The use of Malonamide as a General Spray Reagent for the Fluorometric Detection of Reducing Sugars on Filter Papers and Thin Layer Plates', *J. Chromatogr.*, **176**, 433–434 (1979).
85. D.C. Farwell, A.S. Dion, 'A Fluorimetric Assay for the Quantification of Amino Sugars', *Anal. Biochem.*, **95**, 533–539 (1979).
86. M.H. Jimenez, C.E. Weill, 'A Sensitive Fluorimetric Analysis of Amino Sugars', *Carbohydr. Res.*, **101**, 133–137 (1982).
87. R. Klaus, W. Fischer, 'Quantitative Thin Layer Chromatography of Sugar Acids and Polyalcohols', *Methods Enzymol.*, **160**, 159–175 (1988).
88. C.J. Biermann, 'Hydrolysis and Other Cleavages of Glycosidic Linkages in Polysaccharides', *Adv. Carbohydr. Chem. Biochem.*, **46**, 251–271 (1988).
89. A.B. Blakeney, P.J. Harris, R.J. Henry, B.A. Stone, 'A Simple and Rapid Preparation of Alditol Acetates for Monosaccharide Analysis', *Carbohydr. Res.*, **113**, 291–299 (1983).
90. A. Fox, S.L. Morgan, J. Gilbert, 'Preparation of Alditol Acetates and Their Analysis by Gas Chromatography (GC) and Mass Spectrometry (MS)', in *Analysis of Carbohydrates by GLC and MS*, eds. C.J. Biermann, G.D. McGinnis, CRC Press, Boca Raton, Florida, 87–117, 1989.
91. T.T. Stevenson, R.H. Furneaux, 'Chemical Methods for the Analysis of Sulfated Galactans from Red Algae', *Carbohydr. Chem.*, **210**, 277–298 (1991).

92. R.G. Veness, C.S. Evans, 'Identification of Monosaccharides and Related Compounds by Gas Chromatography Fourier Transform Infrared Spectroscopy of their Trimethylsilyl Ethers', *J. Chromatogr. A*, **721**, 165–172 (1996).
93. G.L. Cowie, J.I. Hedges, 'Determination of Neutral Sugars in Plankton Sediments and Wood by Capillary Gas-chromatography of Equilibrated Isomeric Mixtures', *Anal. Chem.*, **56**, 497–504 (1984).
94. K.J. Schaffer, P.G. Morel Du Boil, 'Quantitative Gas Chromatographic Analysis of Sucrose in the Presence of Sugar Oximes Using Buffered Oximation Reagent and Glass Capillary Columns', *J. Chromatogr.*, **207**, 221–229 (1981).
95. G. Eklund, B. Josefsson, C. Roos, 'Gas–Liquid Chromatography of Monosaccharides at the Picogram Level Using Glass Capillary Columns, Trifluoroacetyl Derivatives and Electron-capture Detection', *J. Chromatogr.*, **142**, 575–585 (1977).
96. J. Bleton, P. Mejanelle, J. Sansoulet, S. Goursaud, A. Tehapla, 'Characterisation of Neutral Sugars and Uronic Acids After Methanolysis and Trimethylsilylation for Recognised Plant Gums', *J. Chromatogr. A*, **720**, 27–49 (1996).
97. M.F. Chaplin, 'A Rapid and Sensitive Method for the Analysis of Carbohydrate Components in Glycoproteins Using Gas–Liquid Chromatography', *Anal. Biochem.*, **123**, 336–341 (1982).
98. O. Pelletier, C. Cadieux, 'Glass Capillary or Fused Silica Gas-chromatography of Several Monosaccharides and Related Sugars – Improved Resolution', *J. Chromatogr.*, **231**, 225–235 (1982).
99. H. Schweer, 'Gas Chromatography–Mass Spectrometry of Aldoses as *O*-Methoxime, *O*-2-Methyl-2-propoxime and *O*-*n*-Butoxime *per*Trifluoroacetyl Derivatives on OV225 with Methyl Propane as Ionisation Agent. I – Pentoses', *J. Chromatogr.*, **236**, 355–360 (1982).
100. P. Decker, H. Schweer, 'Gas Liquid Chromatography on OV-225 of Tetroses and Aldopentoses as their *O*-Methoxime and *O*-*n*-Butoxime *per*Trifluoracetyl Derivative and of C_3–C_6 Alditol *per*Trifluoroacetates', *J. Chromatogr.*, **236**, 369–373 (1982).
101. H. Schweer, 'Gas Chromatography–Mass Spectrometry of Aldoses as *O*-Methoxime, *O*-2-Methyl-2-propoxime and *O*-*n*-Butoxime *per*Trifluoroacetyl Derivatives on OV225 with Methyl Propane as Ionisation Agent. II – Hexoses', *J. Chromatogr.*, **236**, 361–367 (1982).
102. L.W. Eudy, M.D. Walla, S.L. Morgan, A. Fox, 'Gas Chromatographic–Mass Spectrometric Determination of the Muramic Acid Content and Pyrolysis Profiles for a Group of Gram-positive and Gram-negative Bacteria', *Analyst*, **110**, 381–385 (1985).
103. F.R. Seymour, R.D. Plattner, M.E. Slodki, 'Gas Liquid Chromatography–Mass Spectrometry of Methylated and Deuteriomethylated per-*O*-Acetylaldononitriles from D-mannose', *Carbohydr. Res.*, **44**, 181–198 (1975).
104. G.O. Guerrant, C.W. Moss, 'Determination of Monosaccharides as Aldononitrile, *O*-Methyloxime, Alditol and Cyclitol Acetate Derivatives by Gas-chromatography', *Anal. Chem.*, **56**, 633–638 (1984).
105. I.M. Morrison, M.B. Perry, 'The Analysis of Neutral Glycoses in Biological Materials by Gas–Liquid Partition Chromatography', *Canad. J. Biochem.*, **44**, 1115–1126 (1966).
106. P.J. Garegg, B. Lindberg, P. Konradsson, I. Kvarnstrom, 'Hydrolysis of Glycosides Under Reducing Conditions', *Carbohydr. Res.*, **176**, 145–148 (1988).
107. G.E. Black, A. Fox, 'Recent Progress in the Analysis of Sugar Monomers from Complex Matrices Using Chromatography in Conjunction with Mass Spectrometry or Stand-alone Tandem Mass Spectrometry', *J. Chromatogr. A.*, **720**, 51–60 (1996).
108. G.D. McGinnis, C.J. Biermann, 'Analysis of Monosaccharides as per-*O*-Acetylated Aldononitrile (PAAN) Derivatives by Gas–Liquid Chromatography (GLC)', in *Analysis of Carbohydrates by GLC and MS*, eds. C.J. Biermann, G.D. McGinnis, CRC Press, Boca Raton, 119–125, 1989.
109. J.R. Neeser, T.F. Schweizer, 'Analysis of Carbohydrates as *O*-Alkyloxime Derivatives by Gas–Liquid Chromatography (GLC)', in *Analysis of Carbohydrates by GLC and MS*, eds. C.J. Biermann, G.D. McGinnis, CRC Press, Boca Raton, 143–155, 1989.
110. R. Geyer, H. Geyer, 'Saccharide Linkage Analysis Using Methylation and Other Techniques', *Methods Enzymol.*, **230**, 86–108 (1994).
111. S.B. Levery, 'Use of Permethylation with GC–MS for Linkage and Sequence Analysis of Oligosaccharides: Historical Perspectives and Recent Developments', in *Glycopeptides and Related Compounds*, eds. D.G. Large, C.D. Warren, Marcel Dekker, NY, 541–592, 1997.
112. B. Lindberg, 'Methylation Analysis of Polysaccharides', *Methods Enzymol*, **28**, 178–195 (1972).
113. G.G.S. Dutton, 'Applications of Gas Liquid Chromatography', *Adv. Carbohydr. Chem.*, **28**, 11–160 (1973).
114. G.G.S. Dutton, 'Applications of Gas Liquid Chromatography to Carbohydrates: Part II', *Adv. Carbohydr. Chem.*, **30**, 9–110 (1974).
115. N.C. Carpita, E.M. Shea, 'Linkage Structure of Carbohydrates by Gas Chromatography – Mass Spectrometry (GC–MS) of Partially Methylated Alditol Acetates', in *Analysis of Carbohydrates by GLC and MS*, eds. C.J. Biermann, G.D. McGinnis, CRC Press, Boca Raton, 157–216, 1989.
116. C.G. Hellerqvist, 'Linkage Analysis Using Lindberg Method', *Methods Enzymol.*, **193**, 554–573 (1990).
117. T. Tai, K. Yamashita, A. Kobata, 'Synthesis and Mass Fragmentation Analysis of Partially *O*-Methylated 2-*N*-Methylglucosamines', *J. Biochem. (Tokyo)*, **78**, 679–686 (1975).

118. K. Stellner, H. Saito, S. Hakomori, 'Determination of Amino Sugar Linkages in Glycolipids by Methylation: Amino Sugar Linkages of Ceramide Pentasaccharides of Rabbit Erythrocytes and of Forssman Antigen', *Arch. Biochem. Biophys.*, **155**, 464–472 (1973).
119. R.A. Laine, W.J. Esselman, C.C. Sweeley, 'Gas Chromatography of Carbohydrates', *Methods Enzymol.*, **28**, 159–178 (1972).
120. G.R. Gray, 'Reductive Cleavage of *per*Methylated Polysaccharides', *Methods Enzymol.*, **138**, 26–38 (1987).
121. T.J. Waeghe, A.G. Darvill, M. McNeil, P. Albersheim, 'Determination by Methylation Analysis of the Glycosyl Linkage Compositions of Microgram Quantities of Complex Carbohydrates', *Carbohydr. Res.*, **123**, 281–304 (1983).
122. S.C. Churms, 'Recent Developments in the Chromatographic Analyses of Carbohydrates', *J. Chromatogr.*, **500**, 555–583 (1990).
123. K.B. Hicks, 'High Performance Liquid Chromatography of Carbohydrates', *Adv. Carbohydr. Chem. Biochem.*, **46**, 17–72 (1988).
124. M. Verzele, G. Simeons, F. van Damme, 'A Critical Review of Some Liquid-chromatography Systems for the Separation of Sugars', *Chromatographia*, **23**, 292–300 (1987).
125. M.R. Hardy, R.R. Townsend, 'High pH Anion-exchange Chromatography of Glycoprotein Derived Carbohydrates', *Methods Enzymol.*, **230**, 208–225 (1994).
126. M. Stefansson, D. Westerlund, 'Ligand-exchange Chromatography of Carbohydrates and Glycoconjugates', *J. Chromatogr. A*, **720**, 127–136 (1996).
127. Z. El Rassi, 'Recent Progress in Reversed-phase and Hydrophobic Interaction Chromatography of Carbohydrates Species', *J. Chromatogr. A*, **720**, 93–118 (1996).
128. M.H. Gey, K.K. Unger, 'A Strategy for Chromatographic and Structural Analysis of Monosaccharide Species from Glycoproteins', *Fresenius J. Anal. Chem.*, **356**, 488–494 (1996).
129. J.X. Khym, L.P. Zyll, 'Separation of Monosaccharides by Ion-exchange', *J. Amer. Chem. Soc.*, **73**, 2399–2400 (1951).
130. J.F. Kennedy, J.E. Fox, 'Fully Automatic Ion-exchange Chromatographic Analysis of Neutral Monosaccharides and Oligosaccharides', *Methods Carbohydr. Chem.*, **VIII**, 3–12 (1980).
131. J.F. Kennedy, 'Methodology and Instrumentation in the Chromatography of Carbohydrates on Ion-exchange Resins', *Biochem. Soc. Trans.*, **2**, 54–64 (1977).
132. L. Hough, J.V.S. Jones, P. Wusteman, 'On the Automated Analysis of Neutral Monosaccharides in Glycoproteins and Polysaccharides', *Carbohydr. Res.*, **21**, 9–17 (1972).
133. L. Hough, A.M.Y. Ko, P. Wusteman, 'The Simultaneous Separation and Analysis of Aminodeoxyhexitols and Alditols by Automated Ion-exchange Chromatography', *Carbohydr. Res.*, **44**, 97–100 (1975).
134. K. Mopper, 'Improved Chromatographic Separations on Anion-exchange Resins', *Anal. Biochem.*, **86**, 597–601 (1978).
135. P. Nordin, 'Monitoring of Carbohydrates with Periodate in Effluents from HPLC Columns', *Anal. Biochem.*, **86**, 597–601 (1978).
136. S.A. Barker, M.J. How, P.J. Peplow, P.J. Somers, 'Separation of Isomeric Polyols', *Anal. Biochem.*, **26**, 219–230 (1968).
137. J.A. Rendleman, 'Ionization of Carbohydrates in the Presence of Metal Hydroxides and Oxides', *Adv. Chem. Ser.*, **117**, 51–69 (1973).
138. Y.C. Lee, 'High Performance Anion-exchange Chromatography for Carbohydrate Analysis', *Anal. Biochem.*, **189**, 151–162 (1990).
139. M.R. Hardy, R.R. Townsend, 'Analysis of Glycoprotein Oligosaccharides Using High-pH Anion Exchange Chromatography', *Glycobiology*, **1**, 139–147 (1991).
140. N.G. Karlsson, G.C. Hansson, 'Analysis of Monosaccharide Composition of Mucin Oligosaccharide-alditols by High Performance Anion-exchange Chromatography', *Anal. Biochem.*, **224**, 538–541 (1995).
141. J.B. Jahnel, P. Ilieva, F.H. Frimmel, 'HPAE-PAD – a Sensitive Method for the Determination of Carbohydrates', *Fresenius J. Anal. Chem.*, **360**, 827–829 (1998).
142. C. Corradini, G. Canali, E. Cogliandro, I. Nicoletti, 'Separation of Alditols of Interest in Food Products by High-performance Anion-exchange Chromatography with Pulsed Amperometric Detection', *J. Chromatogr. A*, **791**, 343–349 (1997).
143. J.E. Hallsworth, N. Magan, 'A Rapid HPLC Protocol for the Detection of Polyols and Trehalose', *J. Microbiol. Meth.*, **29**, 7–13 (1997).
144. M.R. Hardy, R.R. Townsend, Y.C. Lee, 'Monosaccharide Analysis of Glycoconjugates by Anion-exchange Chromatography with Pulsed Amperometric Detection', *Anal. Biochem.*, **170**, 54–62 (1988).
145. M. Weitzhandler, C. Pohl, J. Rohrer, L. Narayanan, R. Slingsby, N. Avdalovic, 'Eliminating Amino-acid and Peptide Interference in High Performance Anion-exchange Pulsed Amperometric Detection in Glycoprotein Monosaccharide Analysis', *Anal. Biochem.*, **241**, 128–134 (1996).
146. J.M. Marioli, T. Kuwana, 'Electrochemical Detection of Carbohydrates at Nickel–Copper and Nickel–Chromium–Iron Alloy Electrodes', *Electroanalysis*, **5**, 11–15 (1993).
147. T.P. Tougas, M.J. Debenedetto, J.M. Demott, 'Post-chromatographic Electrochemical Detection of Carbohydrates at a Silver Oxide Electrode', *Electroanalysis*, **5**, 669–675 (1993).
148. T.R.I. Cataldi, D. Centonze, G. Margiotta, 'Separation and Pulsed Amperometric Detection of Alditols and Carbohydrates by Anion Exchange Chromatography Using Alkaline Mobile Phases Modified with Ba(II),

Sr(II) and Ca(II) Ions', *Anal. Chem.*, **69**, 4842–4848 (1997).
149. T.R.I. Cataldi, I.G. Casella, D. Centonze, 'Effect of Ca(II), Sr(II) and Ba(II) on Pulsed Amperometric Detection of Alditols and Carbohydrates at a Gold Electrode in Alkaline Solutions', *Anal. Chem.*, **69**, 4849–4855 (1997).
150. M. Stefansson, D. Westerlund, 'Ligand-exchange Chromatography of Carbohydrates and Glycoconjugates', *J. Chromatogr. A*, **720**, 127–136 (1996).
151. J.K.N. Jones, R.A. Wall, A.O. Pittet, 'The Separation of Sugars on Ion-exchange Resins. Part I. Separation of Oligosaccharides', *Can. J. Chem.*, **38**, 2285–2289 (1960).
152. R.W. Goulding, 'Liquid Chromatography of Sugars and Related Polyhydric Alcohols on Cation Exchangers', *J. Chromatogr.*, **103**, 229–239 (1975).
153. H. Caruel, L. Rigal, A. Gaset, 'Carbohydrate Separation by Ligand-exchange Liquid Chromatography. Correlation Between the Formation of Sugar Cation Complexes and the Elution Order', *J. Chromatogr.*, **558**, 89–104 (1991).
154. J.K.N. Jones, R.A. Wall, 'The Separation of Sugars on Ion-exchange Resins. Part II. Separation of Monosaccharides', *Can. J. Chem.*, **38**, 2290–2294 (1960).
155. S.J. Angyal, G.S. Bethell, R.J. Beveridge, 'The Separation of Sugars and of Polyols on Cation-exchange Resins in the Calcium Form', *Carbohydr. Res.*, **73**, 9–18 (1979).
156. O. Samuelson, H. Strömberg, 'Partition Chromatography of Mixtures Containing Polyols and Carbonyl Compounds (Including Sugars) on Ion-exchange Resins', *Acta Chem. Scand.*, **22**, 1252–1258 (1968).
157. L. Petrus, V. Bilnik, L. Kuniak, L. Stankovic, 'Chromatographic Separation of Alditols on a Cation-exchange Resin in the Lanthanum Form', *Chem. Zvestt*, **34**, 530–536 (1980).
158. M. Stefansson, 'Selective Complex-formation of Saccharides with Europium(III) and Iron(III) Ions at Alkaline pH Studied by Liquid-exchange Chromatography', *J. Chromatogr.*, **630**, 123–128 (1993).
159. C.S. Horvath, W.R. Melander, I. Molnar, 'Solvophobic Interactions in Liquid Chromatography with Non-polar Stationary Phases', *J. Chromatogr.*, **125**, 129–156 (1976).
160. N.W.H. Cheetham, P. Sirsimanne, W.R. Day, 'High Performance Liquid Chromatographic Separation of Carbohydrate Oligomers', *J. Chromatogr.*, **207**, 439–444 (1981).
161. Y.C. Lee, B.I. Lee, N. Tomiya, N. Takahashi, 'Parameterization of Contribution of Sugar Units to Elution Volumes in Reverse-phase HPLC of 2-Pyridylaminated Oligosaccharides', *Anal. Biochem.*, **188**, 259–266 (1990).
162. H. Takemoto, S. Hase, T. Ikenaka, 'Microquantitative Analysis of Neutral and Amino Sugars as Fluorescent Derivatives by HPLC', *Anal. Biochem.*, **145**, 245–250 (1985).
163. S. Honda, E. Akao, S. Susuki, M. Okuda, K. Kakehi, J. Nakamura, 'High-performance Liquid Chromatography of Reducing Carbohydrates as Strongly UV-absorbing and Electrochemically Sensitive 1-Phenyl-3-methyl-5-pyrazolone Derivatives', *Anal. Biochem.*, **180**, 351–357 (1989).
164. K. Kakehi, S. Suzuki, S. Honda, Y.C. Lee, 'Pre-column Labelling of Reducing Carbohydrates with (1-*p*-Methoxy)-3-methyl-5-pyrazolone. Analysis of Neutral and Sialic Acid-containing Oligosaccharides Found in Glycoproteins', *Anal. Biochem.*, **199**, 256–268 (1991).
165. H. Kwon, J. Kim, 'Determination of Monosaccharides in Glycoproteins by Reverse-phase High-performance Liquid Chromatography', *Anal. Biochem.*, **215**, 243–252 (1993).
166. R.N. Zhang, Z.D. Zhang, G.Q. Liu, Y.J. Hidaka, Y. Shimonishi, 'High-performance Liquid Chromatographic Determination of Neutral and Amino Monosaccharides by Ultraviolet and Fluorescence Detection of Sugar 9-Fluorenylmethoxycarbonyl Hydrazones and 9-Fluorenylmethoxycarbonyl Amino Sugars at Picomole and Sub-picomole Levels', *J. Chromatogr.*, **646**, 45–52 (1993).
167. K. Muramoto, R. Goto, H. Kamiya, 'Analysis of Reducing Sugars as their Chromophoric Hydrazones by High Performance Liquid Chromatography', *Anal. Biochem.*, **162**, 435–442 (1987).
168. N.K. Karamanos, 'High-performance Liquid Chromatographic Determination of Xylitol and Hexosaminitols Present in the Reduced Terminal of Glycosaminoglycans', *J. Liq. Chromatogr.*, **16**, 2639–2652 (1993).
169. F. Altmann, 'Determination of Amino-sugars and Amino-acids in Glycoconjugates Using Precolumn Derivatization with *ortho*-Phthaldehyde', *Anal. Biochem.*, **204**, 215–219 (1992).
170. N. Ikemoto, L.C. Lo, K. Nakanishi, 'Detection of Subpicomole Levels of Compounds Containing Hydroxyl and Amino-groups with the Fluorogenic Reagent, 2-Naphtholylimidazole', *Angew. Chem., Int. Ed. Engl.*, **31**, 890–891 (1992).
171. K.R. Anumula, 'Quantitative Determination of Monosaccharides in Glycoproteins by High-performance Liquid Chromatography with Highly Sensitive Fluorescence Detection', *Anal. Biochem*, **220**, 275–283 (1994).
172. A. Lagana, A. Marino, G. Fago, B.P. Martinez, 'A Hydrolysis Method Using Microwaves – Determination of *N*-Acetyl and *N*-Glycolyl-neuraminic Acids in Biological Systems by Fluorometric High-performance Liquid Chromatography', *Anal. Biochem.*, **215**, 266–272 (1993).
173. K.R. Anumula, 'Rapid Quantitative Determination of Sialic Acids in Glycoproteins by High-performance Liquid-chromatography with a Sensitive Fluorescence Detection', *Anal. Biochem.*, **230**, 24–30 (1995).

174. M.J. Spiro, R.G. Spiro, 'Monosaccharide Determination of Glycoconjugates by Reverse-phase High-performance Liquid Chromatography of their Phenylthiocarbamyl Derivatives', *Anal. Biochem.*, **204**, 152–157 (1992).
175. M. Stefansson, B. Lu, 'Ion-pair LC of Sugars at Alkaline pH and Pulsed Electrochemical Detection', *Chromatographia*, **35**, 61–66 (1993).
176. I. Meynial, V. Paquet, D. Combes, 'Simultaneous Separation of Nucleotides and Nucleotide Sugars Using an Ion-pair Reversed-phase HPLC Application for Assaying Glycosyltransferase Activity', *Anal. Chem.*, **67**, 1627–1631 (1995).
177. A.J. Alpert, 'Hydrophilic-interaction Chromatography for the Separation of Peptides, Nucleic-acids and Other Polar Compounds', *J. Chromatogr.*, **499**, 177–196 (1990).
178. S.C. Churms, 'Recent Progress in Carbohydrate Separation by High-performance Liquid Chromatography Based on Hydrophilic Interaction', *J. Chromatogr. A*, **720**, 75–91 (1996).
179. B. Herbreteau, M. Lafosse, L. Morin-Alloey, M. Dreux, 'Separation of Sugars and Polyols by Liquid Chromatography Using Bare Silica-gels and Evaporative Light-scattering Detection', *Anal. Chim. Acta*, **267**, 147–156 (1992).
180. B. Herbreteau, M. Lafosse, L. Morin-Alloey, M. Dreux, 'High-performance Liquid Chromatography of Raw Sugars and Polyols Using Bonded Silica Gels', *Chromatographia*, **33**, 325–330 (1992).
181. T. Akiyama, 'Separation of Neutral Monosaccharides and Oligosaccharides Derivatized with Ethyl *p*-Aminobenzoate by High Performance Liquid Chromatography on Amine-bonded Vinyl Alcohol Copolymer Column', *J. Chromatogr.*, **588**, 53–59 (1991).
182. D.W. Armstrong, H.L. Jin, 'Evaluation of Liquid Chromatographic Separation of Monosaccharides, Disaccharides, Tetrasaccharides, Deoxysaccharides and Sugar Alcohols with Stable Cyclodextrin Bonded Phase Columns', *J. Chromatogr.*, **462**, 219–232 (1989).
183. A. Sugii, K. Harada, Y. Tomita, 'Separation of Carbohydrates by High-performance Liquid-chromatography on Porous Pyridinium Polymer Columns', *J. Chromatogr.*, **366**, 412–416 (1986).
184. S. Honda, 'Post-column Derivatization for Chromatographic Analysis of Carbohydrates', *J. Chromatogr. A*, **720**, 183–199 (1996).
185. K. Mopper, 'Improved Chromatographic Separations on Anion-exchange Resins. III – Sugars in Borate Medium', *Anal. Biochem.*, **87**, 162–168 (1978).
186. K. Mopper, E.T. Dengens, 'A New Chromatographic Sugar Analyser with a Sensitivity of 10^{-10} Moles', *Anal. Biochem.*, **45**, 147–153 (1972).
187. M.J. Del Noazal, J.L. Bernal, V. Hernandex, L. Toribio, R. Mendez, 'Purpald (4-Amino-3-hydrazino-5-mercapto-1,2,4-triazole) as a Reagent for Post-column Derivatization of Neutral Monosaccharides in High Pressure Liquid Chromatography', *J. Liq. Chromatogr.*, **16**, 1105–1116 (1993).
188. S. Honda, M. Takahashi, Y. Nishimura, K. Kakehi, S. Ganno, 'Sensitive Ultraviolet Monitoring of Aldoses in Automated Borate Complex Anion-exchange Chromatography with 2-Cyanoacetamide', *Anal. Biochem.*, **118**, 162–167 (1981).
189. S. Hara, H. Ikegami, A. Shono, T. Mega, T. Ikenaka, Y. Matsushima, 'Analysis of Neutral Sugars as Glycamines Using an Amino Acid Analyser', *Anal. Biochem.*, **97**, 166–172 (1979).
190. K. Mopper, R. Dawson, G. Liebezeit, H.-P. Hansen, 'Borate Complex Ion-exchange Chromatography with Fluorometric Detection for Determination of Saccharides', *Anal. Chem.*, **52**, 2018–2022 (1980).
191. S. Honda, K. Sudo, K. Kakimoto, K. Kakehi, H. Yuki, K. Kakiura, 'Selective Micro-determination of Unconjugated Uronic Acids by Fluorescence Reactions with Ethylene Diamine Sulfate', *Anal. Chim. Acta*, **77**, 199–203 (1975).
192. S. Honda, Y. Matsuda, M. Takahashi, K. Kakehi, S. Ganno, 'Fluorometric Determination of Reducing Carbohydrates with 2-Cyanoacetamide and Application to Automated Analysis of Carbohydrates with Borate Complexes', *Anal. Chem.*, **52**, 1079–1082 (1980).
193. N. Watanabe, M. Inoue, 'Amperometric Detection of Reducing Carbohydrates in Liquid Chromatography', *Anal. Chem.*, **55**, 1016–1019 (1983).
194. S. Honda, K. Enami, T. Konishi, S. Suzuki, K. Kakehi, 'Use of Ethylene Diamine Sulfate for Post-column Derivatization of Reducing Carbohydrates to Electrochemically Oxidizable Compounds in High Performance Liquid Chromatography', *J. Chromatogr.*, **361**, 321–329 (1986).
195. P.J. Oefner, C. Chiesa, 'Capillary Electrophoresis of Carbohydrates', *Glycobiology*, **4**, 397–412 (1994).
196. S. Honda, 'Separation of Neutral Carbohydrates by Capillary Electrophoresis', *J. Chromatogr. A*, **720**, 337–351 (1996).
197. M.V. Novotny, J. Sudor, 'High-performance Capillary Electrophoresis of Glycoconjugates', *Electrophoresis*, **14**, 373–389 (1993).
198. A. Paulus, A. Klockow, 'Detection of Carbohydrates by Capillary Electrophoresis', *J. Chromatogr. A*, **720**, 353–376 (1996).
199. Z. El Rassi, 'Recent Developments in Capillary Electrophoresis of Carbohydrate Species', *Electrophoresis*, **18**, 2400–2407 (1997).
200. R.J. Linhardt, A. Pervin, 'Separation of Negatively Charged Carbohydrates by Capillary Electrophoresis', *J. Chromatogr. A*, **720**, 323–335 (1996).
201. Y. Mechref, G.K. Ostrander, Z. El Rassi, 'Capillary Electrophoresis of Carboxylated Carbohydrates – IV. Adjusting the Separation Selectivity of Derivatized Carboxylated Carbohydrates by Controlling the Electrolyte

Ionic Strength at Sub-ambient Temperature and in the Absence of Electro-osmotic Flow', *J. Chromatogr. A*, **792**, 75–82 (1997).
202. S. Hoffstetter-Kuhn, A. Paulus, E. Gassmann, H.M. Widmer, 'Influence of Borate Complexation on the Electrophoretic Behaviour of Carbohydrates in Capillary Electrophoresis', *Anal. Chem.*, **63**, 1541–1547 (1991).
203. A. Klockow, A. Paulus, V. Figueiredo, R. Amado, H.M. Widmer, 'Determination of Carbohydrates in Fruit Juices by Capillary Electrophoresis and High Performance Liquid Chromatography', *J. Chromatogr. A*, **680**, 187–200 (1994).
204. Y.H. Li, T.I. Lin, 'Determination of Carbohydrates by High Performance Capillary Electrophoresis with Indirect Absorbance Detection', *J. Chromatogr. B*, **681**, 87–97 (1996).
205. B. Lu, D. Westerlund, 'Indirect UV Detection of Carbohydrates in Capillary Zone Electrophoresis by Using Tryptophan as a Marker', *Electrophoresis*, **17**, 325–332 (1996).
206. L.A. Colon, R. Dadoo, R.N. Zarc, 'Determination of Carbohydrates by Capillary Zone Electrophoresis with Amperometric Detection at a Copper Microelectrode', *Anal. Chem.*, **65**, 476–481 (1993).
207. S. Honda, S. Iwase, A. Makino, S. Fujiwara, 'Simultaneous Determination of Reducing Monosaccharides by Capillary Zone Electrophoresis as the Borate Complexes of *N*-2-Pyridylglycamines', *Anal. Biochem.*, **176**, 72–77 (1989).
208. Y. Mechref, Z. El Rassi, 'Capillary Zone Electrophoresis of Derivatized Acidic Monosaccharides', *Electrophoresis*, **15**, 627–634 (1994).
209. S. Honda, S. Suzuki, A. Nose, K. Yamamoto, K. Kakehi, 'Capillary Zone Electrophoresis of Reducing Mono- and Oligo-saccharides as the Borate Complexes of their 3-Methyl-1-phenyl-2-pyrazolin-5-one Derivatives', *Carbohydr. Res.*, **215**, 193–198 (1991).
210. C. Huber, E. Grill, P.J. Oefner, O. Bobleter, 'Capillary Electrophoretic Determination of Component Monosaccharides in Hemicelluloses', *Fresenius J. Anal. Chem.*, **348**, 825–831 (1994).
211. A.E. Vorndran, E. Grill, C. Huber, P.J. Oefner, G.K. Bonn, 'Capillary Zone Electrophoresis of Aldoses, Ketoses and Uronic Acids Derivatized with Ethyl *para*-Aminobenzoate', *Chromatographia*, **34**, 109–114 (1992).
212. H. Schwaiger, P.J. Oefner, C. Huber, E. Grill, G.K. Bonn, 'Capillary Zone Electrophoresis and Micellar Electrokinetic Chromatography of 4-Aminobenzonitrile Carbohydrate Derivatives', *Electrophoresis*, **15**, 941–952 (1994).
213. A. Rylund, O. Dahlman, 'Efficient Capillary Zone Electrophoretic Separation of Wood-derived Neutral and Acidic Mono- and Oligosaccharides', *J. Chromatogr. A*, **738**, 129–140 (1996).
214. A. Guttmann, 'Analysis of Monosaccharide Composition by Capillary Electrophoresis', *J. Chromatogr. A*, **763**, 271–277 (1997).
215. J. Liu, O. Shirota, D. Wiesler, M. Novotny, 'Ultrasensitive Fluorometric Detection of Carbohydrates as Derivatives in Mixtures Separated by Capillary Electrophoresis', *Proc. Nat. Acad. Sci. USA*, **88**, 2302–2306 (1991).
216. J.P. Liu, O. Shirota, M. Novotny, 'Capillary Electrophoresis of Amino Sugars with Laser-induced Fluorescence Detection', *Anal. Chem.*, **63**, 413–417 (1991).
217. J.Y. Zhao, P. Diedrich, Y. Zhang, G. Hindogault, N.J. Dorichi, 'Separation of Aminated Monosaccharides by Capillary Zone Electrophoresis with Laser-induced Fluorescence Detection', *J. Chromatogr. B*, **657**, 307–313 (1994).
218. C.M. Starr, R.I. Masada, C. Hague, E. Skop, J.C. Klock, 'Fluorophore-assisted Carbohydrate Electrophoresis in the Separation, Analysis and Sequencing of Carbohydrates', *J. Chromatogr. A*, **720**, 295–321 (1996).
219. P. Jackson, 'The Analysis of Fluorophore-labeled Carbohydrates by Polyacrylamide Gel Electrophoresis', *Molec. Biotechnol.*, **5**, 101–123 (1996).
220. P. Jackson, 'High-resolution Polyacrylamide Gel Electrophoresis of Fluorophore-labeled Reducing Saccharides', *Methods Enzymol.*, **230**, 250–265 (1994).
221. P. Jackson, 'The use of Polyacrylamide Gel Electrophoresis for the High Resolution Separation of Reducing Saccharides Labeled with the Fluorophore 8-Aminonaphthalene 1,3,6-Trisulfonic Acid', *Biochem. J.*, **270**, 705–713 (1990).

Proteoglycan and Acidic Polysaccharide Analysis

Robert M. Lauder
Lancaster University, Lancaster, UK

1 Introduction **1**
2 Proteoglycans **1**
2.1 Is my Protein Glycosylated? 2
2.2 Inhibition of Glycosylation 2
2.3 Metabolic Labeling During Synthesis 2
2.4 Extraction 2
2.5 Isolation 3
2.6 Analysis of Intact Proteoglycan 5
2.7 Removal of Glycosaminoglycans to Yield an Intact Protein Core 6
2.8 Amino Acid Analysis 8
2.9 Where are the Chains Attached? 9
3 Acidic Polysaccharides **9**
3.1 Release of Intact Carbohydrates from the Protein Core 9
3.2 Isolation of Glycosaminoglycans from Other Glycosaminoglycan and Nonglycosaminoglycan Contaminants 10
3.3 Estimation of Glycosaminoglycan Abundance 11
3.4 Compositional Analysis 12
3.5 Characterization of Intact Carbohydrate Chains 13
3.6 General Techniques for Oligosaccharide Analysis 14
4 Hyaluronan **15**
4.1 Estimation of Hyaluronan Abundance 15
4.2 Analysis of Enzymatic Digestion Products 16
5 Keratan Sulfate **16**
5.1 Characterization 17
5.2 Depolymerizing Keratan Sulfate 17
5.3 Disaccharide Analysis 17
5.4 Fingerprinting Keratan Sulfate 18
6 Chondroitin Sulfate/Dermatan Sulfate **19**
6.1 Characterization 21
6.2 Specificity of Enzymes Acting Upon Chondroitin Sulfate/Dermatan Sulfate 21
6.3 Disaccharide Analysis 22
6.4 Linkage Region Analysis 27
6.5 Analysis of Chain Caps 27
7 Heparan Sulfate/Heparin **28**
7.1 Characterization 28
8 Suppliers **29**
9 Internet Glycoscience Resources **29**
Acknowledgments **29**
Abbreviations and Acronyms **29**
Related Articles **30**
References **30**

The isolation and analysis of proteoglycans (PGs) and the acidic polysaccharide glycosaminoglycan (GAG) chains attached to them is an area of increasing importance. PGs are a structurally diverse group of complex macromolecules comprising a core protein, varying in size from 11 to 400 kDa, to which are attached one or more GAG chains of the same or a different type. These chains carry a negative charge derived mainly from sulfate groups and are composed of a repeating disaccharide, a hexosamine and either a hexuronic acid or, in keratan sulfate (KS), galactose.

1 INTRODUCTION

The isolation and analysis of PGs and the acidic polysaccharide GAG chain attached to them is an area of increasing importance. PGs are a structurally diverse group of complex macromolecules comprising a core protein, varying in size from 11 kDa[1] to 400 kDa,[2] to which are attached one or more GAG chains of the same or a different type (Table 1). These chains carry a negative charge derived mainly from sulfate groups and are composed of a repeating disaccharide, a hexosamine and either a hexuronic acid or, in KS, galactose.

2 PROTEOGLYCANS

PGs are found on mammalian cell surfaces and in extracellular matrixes,[3] in plants,[4] and in invertebrates[5] (Table 1). Several large PGs share features and properties with the prototypical member of the class, the large aggregating PG aggrecan (Table 1). Aggrecan monomers comprise a protein core of ca. 250 kDa to which are attached ca. 100 CS and ca. 50 KS chains. Up to 100 of these monomers form large aggregates of up to 5×10^6 kDa through interaction with hyaluronate (HA). Small leucine-rich PGs[3] have a protein core comprising leucine-rich tandem repeats of ca. 24 amino acids

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Structures and properties of a selection of PGs. (This table is not designed to be a comprehensive list of PGs)

PG	Core (kDa)	Location	Glycosylation (number)	Notes
Aggrecan	220	Cartilage	CS ca. 100, KS ca. 50	
Versican	265	Fibroblasts	CS	
Neurocan	150	Brain	CS	
Fibromodulin	59	Cartilage	KS up to 4, N-linked oligos balance to 4	Inhibits collagen fibril formation
Decorin		Cartilage, skin, tendon	CS 1, N-linked oligos 3?	Inhibits collagen fibril formation. Binds to TGF-β
Biglycan		Cartilage, skin, tendon	CS 3, N-linked oligos 3	
Lumican (corneal)		Cornea	KS 1	
Lumican (arterial)		Artery	Lactosaminoglycan	Normal noncorneal glycosylation
Prelp	58	Cartilage	KS ? 3?	Binds to chondrocytes[137]
Keratocan		Cornea	KS 3	
Chondroadherin		Cartilage		Binds to chondrocytes[137]
Perlecan	30–400	Basement membrane	HS	
Serglycin	17–19	Hemopoietic cell vesicles	Heparin, CS, DS, HS	
Syndecan	31	Many cell types	CS, HS	
Betaglycan	110	Many cell types	CS, HS	
Synaptoglycan	82–83	Synaptic vesicles	KS	
CD44	90	Many cell types	CS, KS	
Glypican	62	Fibroblasts, brain	HS	
Cerebroglycan	58–59	Developing nervous system	CS 1	

CS, chondroitin sulfate; DS, dermatan sulfate; HS, heparan sulfate; TGF-β, transforming growth factor beta.

repeated up to 12 times, a motif which may be involved in binding to other macromolecules.[6]

Fibromodulin is glycosylated by four short N-linked KS chains[7–10] attached within, or close to, the leucine-rich repeat region. This is in contrast to both decorin and biglycan in which the CS/DS chains are located close to the N-terminus. There is evidence that fibromodulin, along with decorin, has a role in the maintenance of the collagen fibrils which make up the extracellular matrix of cartilage. The core protein of fibromodulin binds to collagen monomers[11,12] and the KS chains apparently control fibril diameter and interfibrillar spacing.[13]

2.1 Is my Protein Glycosylated?

If it is suspected that an isolated macromolecule is a PG, the presence of glycosylation should be confirmed; later, the nature of the carbohydrate may be investigated. This is most easily done using a commercial carbohydrate detection kit, which relies upon the oxidation of adjacent hydroxyl groups in carbohydrates to aldehydes. The aldehydes react with either digoxigenin (Roche Molecular Biochemicals, http://193.197.95.199 Glycan Detection Kit) or biotin hydrazide (Glyko, http://www.glyko.com) and are detected using peroxidase or alkaline phosphatase conjugated to an antibody or to streptavidin, respectively.

2.2 Inhibition of Glycosylation

An unglycosylated PG core protein may be expressed in culture if a glycosylation inhibitor is present. Inhibitors include deoxynojirimycin and swainsonine, which block the processing of high-mannose oligosaccharides to complex oligosaccharides; however, swainsonine does not affect hybrid chains. In contrast, tunicamycin inhibits the formation of dolichol pyrophosphate *N*-acetyl-glucosamine. All of these inhibitors will inhibit N- but not O-linked chain synthesis.

2.3 Metabolic Labeling During Synthesis

A radiolabel may be introduced by addition of ^{35}S- or ^{3}H-containing substrates to the growth media during culture. Radiolabeled [^{35}S]sulfate is incorporated into GAGs to examine sulfation, [^{3}H]glucosamine will label most GAGs, [^{3}H]glucose is converted to many sugar precursors and becomes incorporated into most carbohydrates, while [^{3}H]mannose will label only N-linked chains. If [^{3}H]leucine is used then the protein core, along with all other proteins, will be labeled. Radiolabeling can provide a measure of synthesis and sulfation under modified growth conditions, as only newly synthesized molecules are labeled, while pulse chase experiments can provide information on turnover.

2.4 Extraction

A tissue or cell layer can contain many different PGs and their structure and extractability may differ widely, so no single protocol is appropriate for the extraction, isolation and analysis of all PGs. In a tissue such as cartilage they may be trapped by the collagen meshwork

or bound by noncovalent interactions to other molecules, while brain PGs are very readily extracted. Cell surface PGs may be attached via a hydrophobic domain or phosphatidylinositol anchor while other PGs are found intracellularly.

Unless an extraction protocol has already been established, both 4 M guanidine hydrochloride (GuHCl) and 4% 3-([3-cholamidopropyl]dimethylammonio)-1-propanesulfonate (CHAPS) should be included in an extraction buffer. The zwitterionic detergent CHAPS is easier to remove subsequently than the micelle-forming Triton™ X-100, and its inclusion when extracting cells reduces artifactual hydrophobic interactions induced by GuHCl.[14]

Extraction of cartilage PGs[15] is performed for 24 h at 4 °C into 7–10 mL g^{-1} of tissue with 4 M GuHCl, 50 mM sodium acetate (pH 6.8), 100 mM 6-aminohexanoic acid, 20 mM benzamidine hydrochloride, 0.5 mM phenylmethylsulfonyl fluoride, 10 mM ethylenediamine tetraacetic acid (EDTA) and 5 mM *N*-ethylmaleimide, to prevent degradation. Extraction efficiency is increased if the tissue pieces are small; ideally 20-μm cryostat sections or chopped or powdered tissue should be used. Powdered tissue is obtained by pulverizing the tissue after freezing in liquid nitrogen, although there is some evidence of damage to PGs using this method.[16] Connective tissues contain insoluble collagen and inextractable PGs and GAGs which are only accessed by digestion of the collagen network or release of the GAGs (see section 3.1).

2.5 Isolation

Methods used in the separation of PGs from each other, and from nonPGs, often focus on the negative charge generated by the attached GAGs. However, other properties can also be utilized; PGs have larger hydrodynamic volumes and buoyant densities than the size of their protein cores would suggest, and cell surface PGs often have a hydrophobic transmembrane region.

2.5.1 Centrifugal Techniques

The high buoyant density of PGs allows the almost pure single-step isolation of large PGs in a self-forming CsCl density gradient. Smaller PGs also benefit, becoming separated from larger PGs and smaller proteins. If >0.5 M GuHCl is present the gradient is dissociative and may thus be used to isolate PGs from other extracted macromolecules with which they interact; at lower GuHCl concentrations the gradient is associative. CsCl is added to an extract to achieve a density of 1.6 g mL^{-1} using $X = V(1.347\rho - 0.038M - 1.347)$, where X = weight of CsCl in grams, V = final volume, M = molarity of GuHCl, and ρ = density. The density should be confirmed by weighing a 1-mL sample before ultracentrifugation at 88 500 g_{av} for 48 h at 12 °C. If cell surface PGs are being isolated, or the PG is believed to be hydrophobic, 0.5% CHAPS should also be included.[15]

Following centrifugation the gradient can be fractionated by sucking the solution out from the bottom of the tube with a peristaltic pump attached to a stiff, hollow tube, or by piercing the bottom of the tube and allowing the gradient to drip out. The fractions should be assayed for density and for GAGs or PGs of interest, then pooled appropriately. Large PGs such as aggrecan are found at the bottom (>ca. 1.5 g mL^{-1}, A1), while small PGs such as fibromodulin, decorin and biglycan are present towards the top (<ca. 1.45 g mL^{-1}, A3/A4).[17]

2.5.2 Chromatographic Analysis

Before their first use for chromatography of PGs or GAGs it is necessary to condition all column media with CS and BSA (bovine serum albumin) (both ca. 1 mg mL^{-1}) to prevent nonspecific binding of the samples.

2.5.2.1 Ion-exchange Chromatography The negative charge of the GAGs attached to PGs brings about their binding to anion-exchange matrices and their separation from both uncharged macromolecules and from each other in an appropriate elution gradient. The method is not sensitive to sample volume or to chaotropic agents, and so is often used as the first chromatographic step during purification. Nonspecific adsorption of nonPGs and less than 100% recovery of bound PGs may be minimized by careful attention to the capacity of the matrix and column size, or the amount of resin when used in batch mode, together with a low, nonzero, salt concentration in the starting buffer.

For example fibromodulin is isolated in an almost pure form by ion-exchange chromatography (IEC) following CsCl density gradient centrifugation, where it bands at A3/A4. It is eluted from a Mono-Q® HR 10/10 column with a linear gradient of 0.15–1 M NaCl in 6 M urea/50 mM TRIS (tris(hydroxymethyl)aminomethane) · HCl, pH 7.3, and is found in the fractions between 0.4 and 0.8 M NaCl. The related small PGs decorin and biglycan elute at higher salt concentrations (Figure 1). The eluent is monitored online at 280 nm and fractions are assayed for GAG and PGs. Nucleic acids, however, which are common in cell extracts, may bind to anion-exchange resins, and the 1,9-dimethylmethylene blue (DMB) assay (see section 3.3.5) which will respond to GAGs but not nucleic acids should be used to reliably estimate GAG abundance.

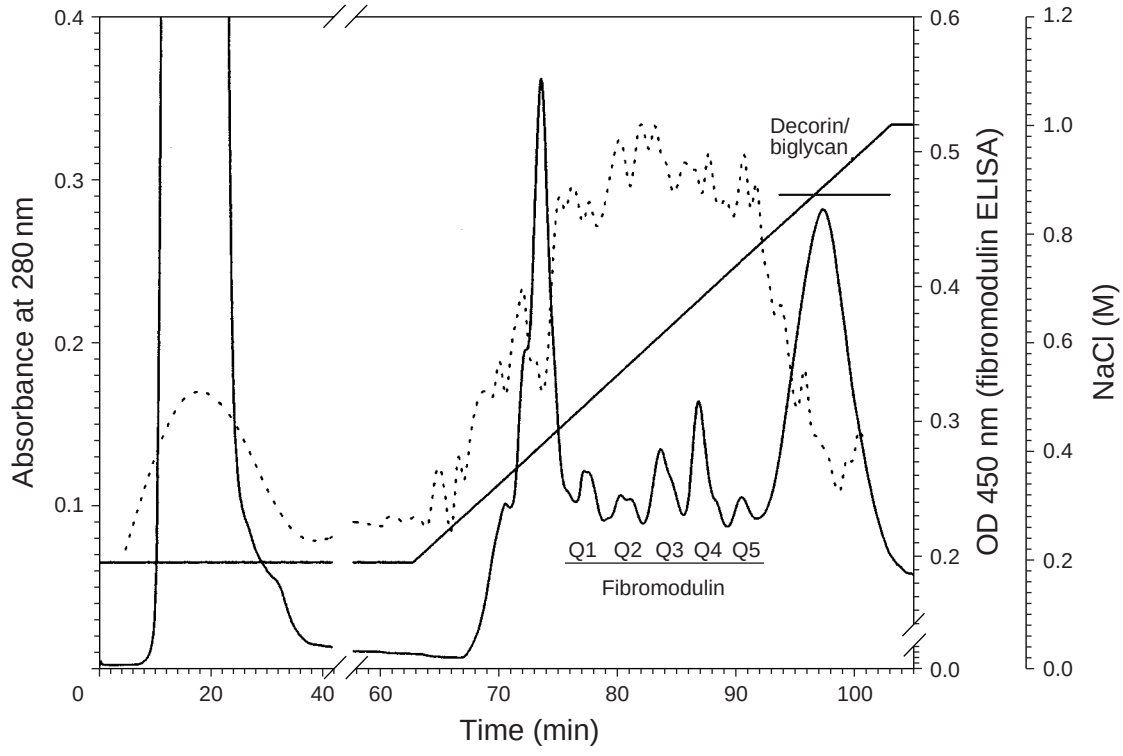

Figure 1 IEC on a Mono-Q® column of A3/A4 density gradient fraction. An aliquot of an A3/4 fraction recovered from a CsCl gradient of a GuHCl extract of bovine articular cartilage (ca. 20 mg) was loaded onto a Mono-Q® 10/10 column. The bound material was eluted in a 0.15–1 M NaCl gradient in 6 M urea, 50 mM TRIS · HCl at pH 7.5. The elution positions of fibromodulin, which elutes as five distinct functions (Q1–Q5), decorin and biglycan are shown. ——, optical density (OD) at 280 nm; - - - - -, fibromodulin enzyme-linked immunosorbent assay (ELISA).

2.5.2.2 Size-exclusion Chromatography Size-exclusion chromatography (SEC) is widely used for the isolation and analysis of molecules of all sizes. Because of the large apparent hydrodynamic volume of PGs, and GAGs, quoted ranges for globular proteins may not be a good measure for the appropriateness of a column. Chromatography will usually be performed using a salt-containing eluent to inhibit interactions with the resin. One exception is Bio-Gel® P2, which may be used at 50 °C in distilled water (dH_2O) for desalting. A volatile salt such as ammonium acetate may obviate the need to desalt before subsequent analysis, but during lyophilization significant drops in the pH can occur which have been found to cause the loss of labile groups such as *N*-acetylneuraminic acid (NeuAc). For large PGs, such as aggrecan, CL-2B resin allows isolation of the aggregated PG under associative conditions and isolation of the monomers under dissociative conditions. Tryptic digest products of aggrecan can be isolated using a CL-6B column while smaller PGs may be isolated using CL-6B, Superose 6, S-300 or HW-55 s. The final choice of medium will depend upon the structure of the PG under investigation and the nature of the material it is being separated from.

2.5.2.3 Hydrophobic Interaction Chromatography Hydrophobic interaction chromatography (HIC) relies upon the hydrophobicity of a molecule and separation is performed in a gradient of decreasing salt or increasing chaotropic agent or CHAPS. In common with IEC, it is not sensitive to loading volume but is not able to support the high sample loading of IEC. Factors including temperature and pH affect elution, with increasing temperature retarding elution.

Decorin and biglycan, which co-elute from a Mono-Q® column (Figure 1), are separated on a 300-mL Octyl-Sepharose® HIC[18] column. The PG mixture (40 mg) should be applied in 4 M or 2 M GuHCl and eluted with a gradient of 0–1% CHAPS or 2–6 M GuHCl respectively.

2.5.2.4 Reversed-phase Chromatography Reversed-phase chromatography (RPC) is not widely used in the isolation of PGs because of its low capacity and the fact that the nonpolar stationary phase may denature a PG. However, the high resolution attainable makes it appropriate for confirming purity[7] as elution is dictated almost exclusively by the peptide. It is also used in the separation of PG fragments generated by tryptic

digestion for the isolation of a GAG attached to a specific glycosylation site.[10]

2.5.3 *Immunopurification*

If an antibody to the PG or the attached GAG chains is available it may be conjugated to a medium such as Dyna-Beads™ or to a column support matrix that allows isolation by immunopurification.

2.6 Analysis of Intact Proteoglycan

2.6.1 *Electrophoresis of Proteoglycan or Protein Core*

The resolution of PGs by SDS/PAGE (sodium dodecyl sulfate/polyacrylamide gel electrophoresis) is often poor because of the polydispersity of chain sulfation and size. Small PGs have been successfully separated using electrophoresis on 8–20% gradient gels and although large PGs will not enter a conventional gel they may be electrophoretically separated on 1.2% polyacrylamide–0.6% agarose.[17,19] For this procedure an acrylamide plug is made by mixing 10 mL of 30% acrylamide (comprising 29.1% acrylamide–0.9% bisacrylamide) with 18.47 mL of distilled water, 30 µL of *N*,*N*,*N*′,*N*′-tetramethylethylenediamine and 1.5 mL of fresh 3% ammonium persulfate. A ca. 3-cm plug is poured and then overlaid with dH_2O saturated butanol. Then 0.24 g of agarose is added to 22.5 mL of solution A (40 mM TRIS · acetate–1 mM sodium sulfate, pH 6.8), and brought to the boil. Acrylamide (0.456 g) and 0.024 g of bisacrylamide are dissolved in 9.7 mL of solution A, then 4.8 mL of fresh 6.4% β-dimethylamino-propionitrile is added and the mixture is warmed. The plug is blotted dry and heated to 60 °C. When the boiled agarose solution has cooled to 50 °C, 3 mL of 3% ammonium persulfate is added to the acrylamide solution, and mixed and added to the agarose. The mixture is poured onto the plug, and a well former is inserted and overlaid with dH_2O saturated butanol.

Aggrecan monomers (5 µL at 1 mg mL^{-1}) in 25% solution A containing 20% sucrose and 0.001% bromophenol blue are then applied to each well. Electrophoresis is carried out in 25% solution A at 60 V until the dye enters the gel, then at 120 V; the PGs should not approach the supporting plug too closely. Following electrophoresis the gel is stained in 200 mg toluidine blue–100 mL 0.1 M acetic acid and destained in 3% acetic acid. Radiolabeled PGs are visualized by fluorography of a fixed gel (15 mM sodium acetate in ethanol containing 0.4% PPO (2,5-diphenyloxazole)). The dH_2O washed and dried gel is sealed in a light-tight box in contact with pre-flashed X-ray film at −70 °C. Exposure time depends upon activity, but as a general guide 1000 dpm (disintegrations per minute) ^{35}S requires ca. 48 h of exposure.

2.6.2 *Enzyme-linked Immunosorbent Assay*

ELISA is a well-established technique for the detection and quantification of structures for which an antibody is available. There are two methods, namely direct and inhibition. The former is appropriate for rapid determinations of the presence of a PG, and the latter more complex method is used for quantitative analyses. In a direct ELISA a sample is coated onto the well of a microtiter plate in 20 mM sodium carbonate at pH 9.6, and after washing in phosphate-buffered saline (PBS) with 0.5% Tween 20 to remove unbound material (performed after each step) an antibody to the antigen of interest is added to each well. This is followed after 1 h by an enzyme-conjugated second antibody which binds to the first. A substrate for the enzyme which generates a colored product is added, and after a fixed time the abundance of this product is estimated by measuring its absorbance, which is proportional to the concentration of the original antigen of interest. In an inhibition ELISA the sample and primary antibody are pre-incubated and then transferred to a microtiter plate, the wells of which are coated in a standard containing an example of the antigen being examined. Excess primary antibody binds to this coated material and may be estimated as described previously.

2.6.3 *Nuclear Magnetic Resonance and Mass Spectroscopy*

Nuclear magnetic resonance (NMR) is a powerful, nondestructive technique, allowing the analysis of an intact PG and its recovery for further analysis. Because a PG is large, signals deriving from the amino acid backbone will be complex, but if aromatic residues are present they will have signals in the otherwise clear area at ca. 7–8 ppm. It can be difficult to get GAG-related information from spectra acquired with an intact PG, although there are some key regions from which information can be derived that are generally free of protein signals. Samples need not be completely pure, although it is typically necessary to have ca. 0.5–1 mg of sample. However, complete ^{1}H characterization of an octasaccharide has been reported where less than 125 µg of sample was available.[20]

Mass spectroscopy (MS) techniques allow the accurate estimation of the mass of an intact PG as the heterogeneity of peaks reflects the heterogeneity of post-translational processing, including glycosylation. Clearly the estimation of mass before and after removal of any GAG chains will allow a determination of the mass of GAGs removed, providing what is potentially a very powerful method for the determination of the nature and size of a GAG along with its enzymatic susceptibility.

2.7 Removal of Glycosaminoglycans to Yield an Intact Protein Core

Peptide-*N*4-(acetyl-glucosaminyl)-asparagine amidase (EC 3.5.1.52, PNGase F) will release many N-linked KS chains by breaking the GlcNAc–Asn bond and completely removing the GAG chain leaving both it and the protein core intact, except for the deamination of Asn to aspartic acid. It acts upon Asn-linked high mannose, hybrid and complex oligosaccharides, but will not act upon chains which contain fucose α(1,3)-linked to the terminal GlcNAc, although α(1,6)-linked fucose has no effect; the Asn must have at least one amino acid residue on both sides (Figure 2).

The PG should be heat denatured with a minimum volume of 1% SDS (sodium dodecyl sulfate) and 1 M β-mercaptoethanol by heating at 100 °C for 5 min. Following cooling Triton™ X-100 is added at 5× excess relative to SDS. PNGase F is then added at 1 U per 20 mg of PG and allowed to digest for 12 h at 37 °C. This can be extended to 24 h if the PG is not denatured or the effect of SDS is not reduced by Triton™. The enzyme is extremely robust and will remain active in the digestion buffer for over 24 h.

There are other endoglycosidases that act upon N-linked chains, although none release the intact chain. Endoglycosidase A (PNGase A, EC 3.5.1.52) cleaves chains with fucose α(1,3)-linked to the terminal GlcNAc; chains upon which PNGase F will not act. However cleavage occurs between the two linkage region GlcNAc residues leaving a single GlcNAc attached to the protein core. There are several endoglycosidase Fs (EC 3.2.1.96) that have significantly different substrate specificities, so it is crucial to confirm the specificity of the enzyme purchased. Each endoglycosidase F cleaves between the two GlcNAc residues of the linkage region: F_1 cleaves hybrid and oligomannose (but not complex) oligosaccharides; F_2 cleaves high mannose and biantennary complex oligosaccharides, showing a significant preference for the complex type, and no activity towards hybrid type or tri- and tetraantennary oligosaccharides; F_3 cleaves only bi- and triantennary complex type oligosaccharides, showing no activity towards hybrid or high mannose type.

No single enzyme will remove all O-linked GAGs, and chains must be depolymerized to leave the linkage disaccharide Galβ(1-3)GalNAc, which may be removed using *endo-O*-glycosidase (*O*-glycopeptide *endo*-D-galactosyl-*N*-acetyl-α-galactosaminohydrolase, EC 3.2.1.97) only following the removal of the rest of the chain (Figure 2).

CS/DS chains may be depolymerized by chondroitin ABC lyase (EC 4.2.2.4) (Figure 3), and the terminal

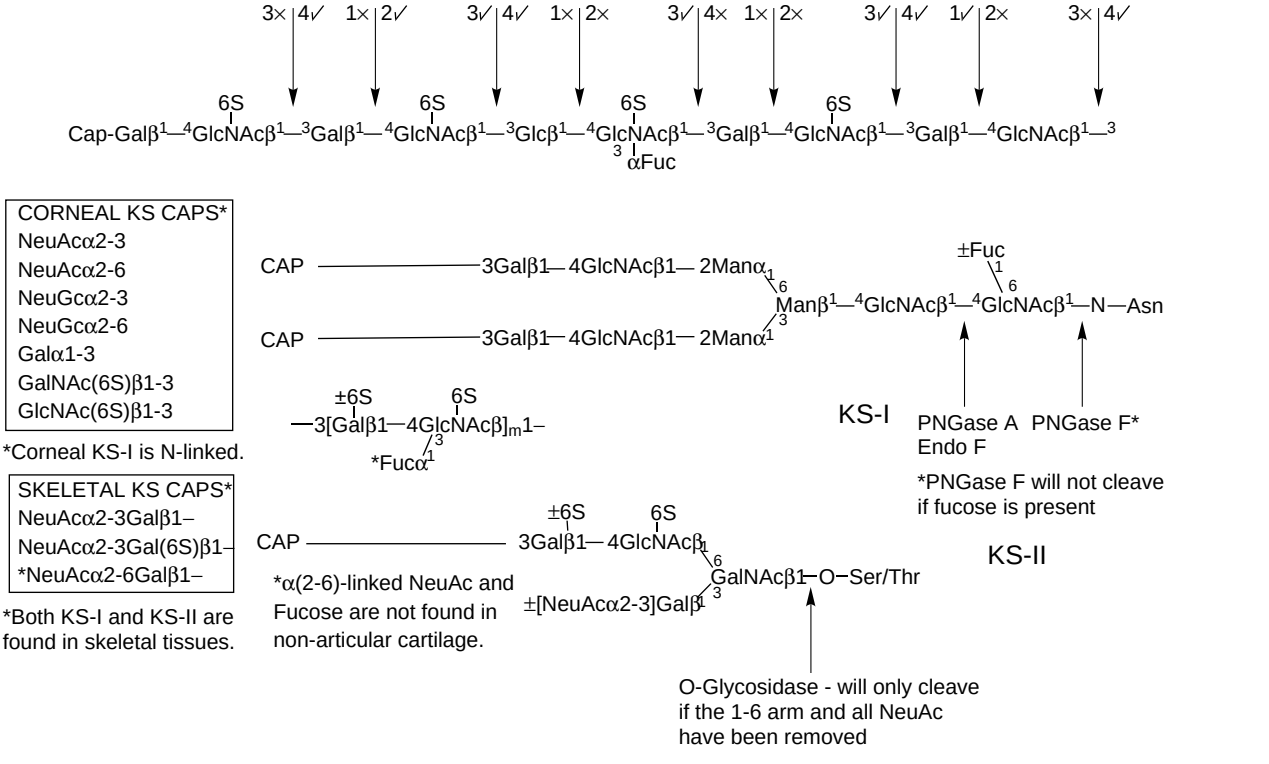

Figure 2 Structural models of KS and specificity of chemical and enzymatic methods for depolymerization. Structural models for corneal and skeletal KS are shown; note the absence of α(2-6)-linked NeuAc and fucose in both O- and N-linked KSs from nonarticular cartilage. The specificity of methods of depolymerizing KS are also shown. For the repeat region 1 = *endo*-β-galactosidase, 2 = keratanase, 3 = keratanase II and 4 = hydrazinolysis/nitrous acid. ✓ = can cleave, × = cannot cleave, ? = unknown.

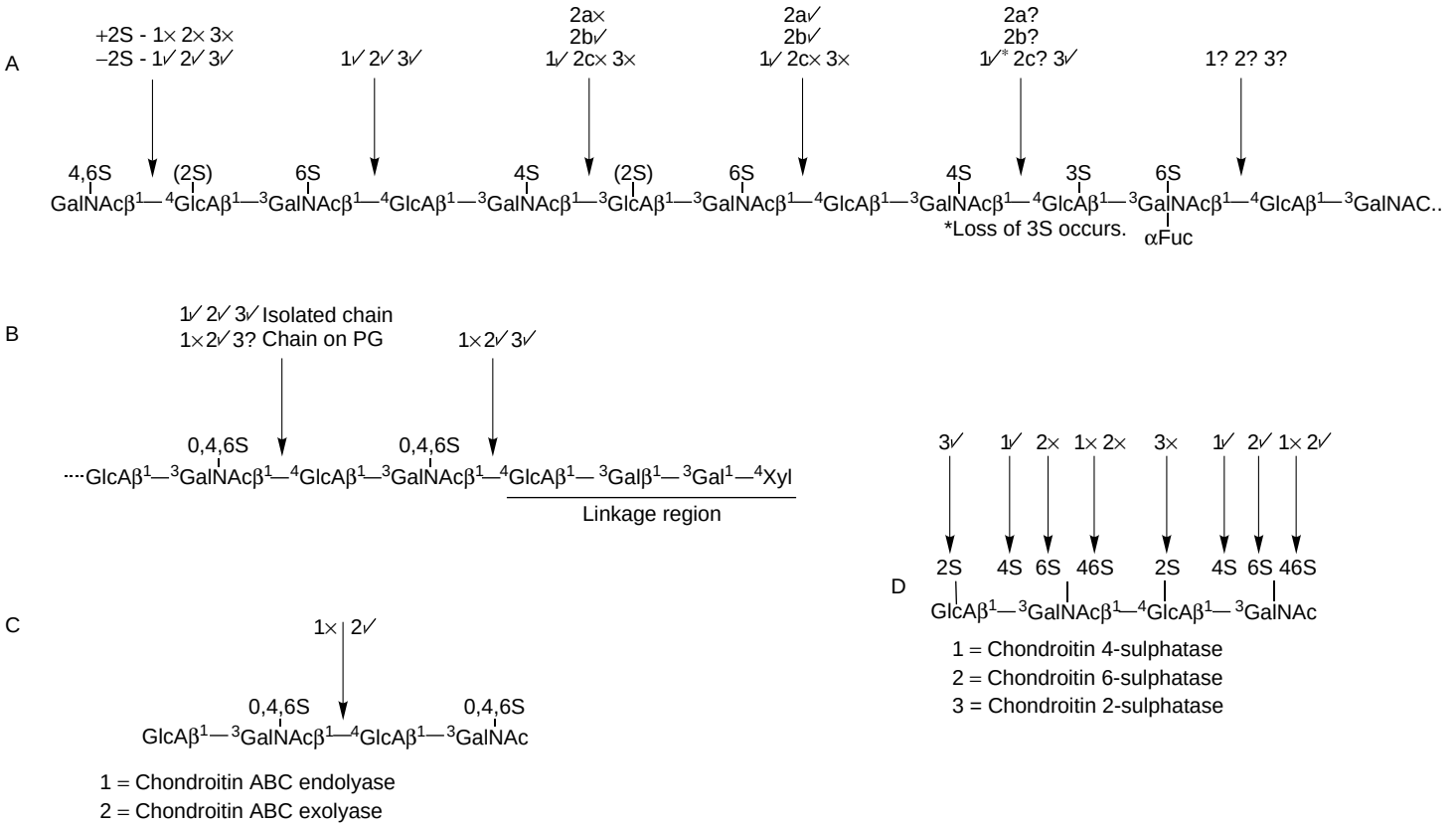

Figure 3 Specificity of CS degrading procedures. The specificity of enzymatic and chemical methods of depolymerizing CS are shown. A and B refer to enzyme action on intact chains while C and D are isolated tetrasaccharides; 1 = chondroitin ABC endo/exo lyase; 2 = chondroitin; a = ACI; b = ACII; c = ACIII (if (a), (b) and (c) are not shown separately all can cleave) and 3 = chondroitin C lyase. Chondroitin B lyase acts solely upon DS. ✓ = can cleave, × = cannot cleave, ? = unknown.

ΔUA that the enzyme generates can be removed using 35 mM mercuric acetate, pH 5, at ambient temperature for 30 min,[21] excess Hg being removed by cation-exchange chromatography. The Gal residues may then be cleaved by β-galactosidase (EC 3.2.1.23), which acts only upon a terminal residue. The remaining xylose requires a xylosidase for its cleavage. Such enzymes have been isolated and used to remove the intact chain, but are not commercially available at present; their widespread availability will represent a significant advance in the isolation of O-linked GAG chains.

Heparin/HS may be depolymerized and the remaining linkage region removed as described for CS/DS, but there is no single enzyme that will depolymerize all chains leaving the linkage region (see section 7.2).

Enzymatic methods are specific but they may not always be appropriate, e.g. if there is a spectrum of GAGs of known or unknown type, or if their structure is such that enzymatic deglycosylation fails. In contrast chemical deglycosylation methods have a broad specificity and may be used to deglycosylate a PG with GAG chains of an unknown type. Techniques used include anhydrous HF at 23 °C for 3 h, which cleaves neutral and acidic sugars along with *O*-glycosidic links but not *N*-glycosidic links,[22] 70% HF in pyridine with 10% anisole at ambient temperature for 8 h,[23] and Smith degradation[23] and methanolysis for 17 h at 80 °C.[24] Note however that there are reports of variable removal of carbohydrates and solubility problems following deglycosylation by HF methods.[24] Gerken et al.,[25] in a modification of the method of Edge et al.,[26] used trifluoromethanesulfonic acid (TFMS). Anisole and TFMS are mixed 1 : 2 in a glass tube, cooled in a dry ice–ethanol slurry and 1 mL was added to 5–10 mg of dry sample. Nitrogen was bubbled through the solution for 30 s followed by stirring at 0 °C for 6 h after which 1 volume of cold anhydrous diethyl ether was added and the mixture slowly added to 12.5 mL of a frozen slush of 60% pyridine. The mixture was allowed to warm to ambient temperature and extracted with ether, the aqueous phase being recovered. Subsequent periodate oxidation followed by an alkaline dialysis was required to remove the linkage residue remaining attached to the core protein. This method has been used to deglycosylate O- and N-linked KS;[27] no studies have examined its action on CS/DS and heparin/HS, although it would be expected to deglycosylate these GAGs.

The removal of a GAG chain may modify the structure or properties of the intact protein core by exposure of previously masked regions, or through disruption of intramolecular interactions.

2.8 Amino Acid Analysis

Amino acid analysis can be very important in the identification of a PG. A pure sample, ca. 20–100 μg, should be subjected to vapor phase acid hydrolysis with

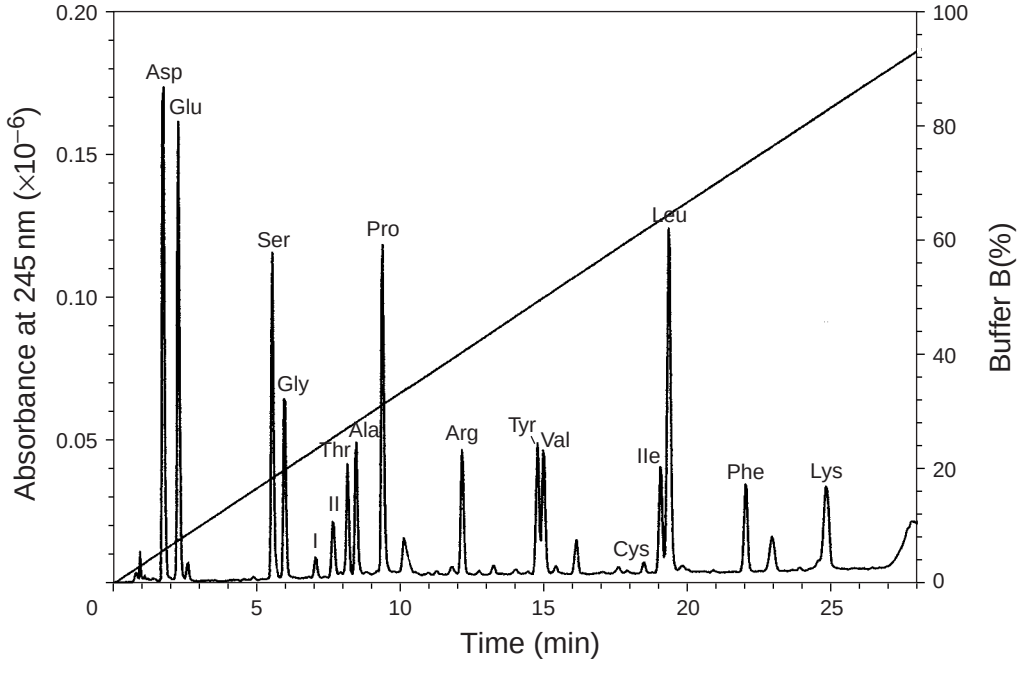

Figure 4 Amino acid analysis of fibromodulin. An aliquot (1 μL) of a PITC derivatized hydrolysate was applied to a Spherisorb® ODS (3 μm 150 × 4.6 mm) column and the amino acids were eluted at 1.5 mL min^{-1} in 0–100% buffer B (50% acetonitrile–50% 24 mM sodium phosphate, pH 6.4). Buffer A is 12 mM sodium phosphate, pH 6.4.

6 M HCl at 105 °C for 24 h, followed by lyophilization and a redrying step in 40% ethanol–20% triethylamine–40% water. Under basic conditions amino groups are coupled to phenylisothiocyanate (PITC) in 70% ethanol–19% water–10% triethylamine–1% PITC at 25 °C for 20 min. Free PITC will poison the column (Spherisorb® ODS 2–3 μm 150 × 4.6 mm) and must be removed by lyophilization prior to chromatography. Labeled amino acids are unstable at ambient temperature and should be stored at −80 °C. Aliquots are resuspended in a minimum volume of buffer A (12 mM sodium phosphate, pH 6.4) and chromatographed at 1.5 mL min^{-1} with a gradient of 0–100% buffer B (50% acetonitrile–50% 24 mM sodium phosphate, pH 6.4) with detection by absorbance at 245 nm (Figure 4).

A control protein should be examined to estimate losses, and the column calibrated with commercial amino acids labeled with PITC. When the composition has been determined, searching the Swiss-Prot database (http://www.expasy.ch/tools/aacomp/) for entries with the appropriate composition may yield the identity of an unknown. If the sequence is available there are tools for the analysis of this data (http://www.expasy.ch/tools/). The NetOGlyc site (http://www.cbs.dtu.dk/services/NetOGlyc/) predicts *O*-glycosylation sites in mammalian PGs where the sequence is known.

2.9 Where are the Chains Attached?

The primary amino acid requirements for glycosylation are well established, but not all potential sites are necessarily glycosylated. Experimentally, a blank cycle at a potential glycosylation site during amino acid sequencing points to the presence of a glycosubstitution. An elegant approach was taken by Plaas et al.[10] who used trypsin to digest a PG and isolated the peptido-KS chains by IEC. Then they used RPC to separate these into pools with the same amino acid sequence. The GAG chains of each pool were characterized and the peptide sequenced, identifying its position in the parent PG and thus revealing the nature of the glycosylation at a specific site.

3 ACIDIC POLYSACCHARIDES

In mammalian tissues there are four GAGs: HA, KS, CS/DS and HS/heparin (Figure 5) which, with the exception of HA, are synthesized attached to PGs and comprise a chain cap, a linkage region which connects the chain to the protein core, and an intervening repeat region. The repeat regions consist of a repeating disaccharide comprising a hexosamine and either a hexuronic acid or, in KS, galactose (Figure 5). Each residue, apart from those of HA which exist as an invariant linear sequence without sulfation or other modifications, may be sulfated, in some cases by more than one sulfate group, rendering GAGs highly negatively charged.

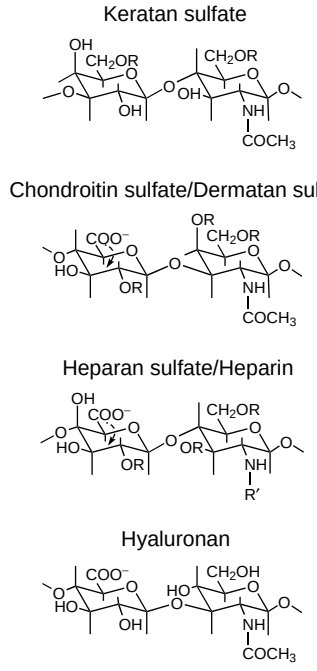

Figure 5 GAG disaccharide structure. The basic disaccharide structures of each GAG are shown. Note the epimerization of C5 in CS/DS and in HS/heparin.

The detailed microstructure of GAGs is crucial to their function, e.g. 4-sulfated GalNAc in CS has a role in the adherence of the malaria parasite to human placenta,[28,29] iduronic acid (IdoA) is required for binding DS to the growth promoter spermine[30] and *N*- and 2-*O*-sulfate groups on the HS of syndecan-1 are important in collagen binding.[31] Other important GAG structures include that of the antithrombin III binding site in heparin,[32] the heparin cofactor II binding site in DS[33] and the bFGF binding sequences in HS.[34]

3.1 Release of Intact Carbohydrates from the Protein Core

Enzymatic procedures for the removal of N-linked GAGs yield the intact chain (see section 2.7), but as discussed previously no such method exists for O-linked chains, and alternative approaches are required.

3.1.1 β-Elimination

Chains that are O-linked are alkali labile and incubation in 0.1 M NaOH at 45 °C[35] or 0.05 M LiOH at 4 °C for 24 h[36] will release them. The former method

requires 1 M $NaBH_4$ to reduce the released chains and prevent degradation due to the high pH, while the latter allows the isolation of intact chains terminated with a reducing sugar. This residue has been used to fluorescently label the CS linkage region for high pH anion-exchange chromatography (HPAEC) analysis[37] and sequencing.[38] Although not readily alkali-labile, N-linked KS chains may be liberated by strongly alkaline conditions such as 1 M NaOH–1 M $NaBH_4$ at 100 °C for 4–6 h.[39]

3.1.2 Hydrazinolysis

Chains with either N- or O-linkages may be released unreduced by hydrazinolysis.[40] Lyophilized samples (5–10 mg mL^{-1}) are resuspended in anhydrous hydrazine and heated to 60 °C for 5 h to release O-linked chains and to 95 °C for 5 h to release N- and O-linked chains. Excess hydrazine, made safe by the addition of one volume of dH_2O, is removed by evaporation under vacuum at 25 °C. Primary amino groups are re-*N*-acetylated by an excess of 0.5 M acetic anhydride in saturated $NaHCO_3$ at 4 °C for 20 min and then at ambient temperature for 30 min. The procedure may also be performed by automated hydrazinolysis, which is reported to require a minimum of ca. 50 μg of sample for efficient release and recovery.[41]

3.1.3 Digestion/Destruction of Protein

If the amino acid on the protein core to which the GAG chain is attached is to be retained, or if complete release is not essential, the protein core may be extensively digested by a broad specificity protease such as papain. It is used at a level of 1 U per 100 mg of tissue in 0.1 M sodium acetate at pH 6.8 with 2.4 mM EDTA and 10 mM cystein HCl, added just prior to digestion, at 60 °C overnight. The mixture is boiled for 30 min or iodoacetamide is added to 10 mM to inactivate. The elevated temperature at which papain digestion is performed gives it a significant advantage over the use of other proteases as unwanted enzymatic activity will be prevented. The remaining peptide may then be used to conjugate biotin, or any other label or probe, via the amino terminus.

3.2 Isolation of Glycosaminoglycans from Other Glycosaminoglycan and Nonglycosaminoglycan Contaminants

GAGs which have been released from their protein core need to be separated from peptides and may need to be fractionated if more than one GAG is present. The choice of method depends upon the amount of GAG expected, what it is being separated from, and the required purity.

3.2.1 Precipitation

The GAGs have different solubility in organic solvents, and are precipitated by the careful addition of ethanol while stirring to the following total number of volumes: heparin 0.1–0.4, DS 0.5–0.8, CS 2, KS 3. After 15 h at 4 °C the GAGs appear as a white precipitate or a sticky mass which is easily recovered. Higher purity will be achieved by precipitating all GAGs, resuspending them in an appropriate volume of 0.1 M sodium acetate, and performing another series of ethanol precipitations. Volpi[42] used the following volumes of propanol for precipitation: heparin 0.3–0.4 volumes, DS 0.5–0.6 volumes and CS > 0.8 volumes.

Cetyl pyridinium chloride (CPC) can also precipitate GAGs. Following papain digestion, 1% CPC is added and the mixture is centrifuged for 1 h, then the pellet is washed with NaOAc saturated ethanol.

3.2.2 Chromatographic Separation Methods

GAGs may be separated from nonGAG material by IEC, essentially as discussed for intact PGs. KS chains have been successfully isolated from nonGAGs and from contaminating CS using Mono-Q® or Q-Sepharose® IEC in a gradient of 2–500 mM $LiClO_4$–10 mM piperazine. Many IEC methods suffer from less than 100% recovery and use of a column with appropriate dimensions is important. Other strategies appropriate for the isolation of labeled GAGs involve the use of unlabeled carrier. Alternatively, intact GAG chains can be separated from smaller oligosaccharides by SEC.

3.2.3 Dialysis

Dialysis and centrifugal partition utilize a membrane through which small molecules such as salt and depolymerized oligosaccharides will pass, but larger molecules will not. Dialysis is appropriate if the GAG is abundant and in a large volume. However, if the amount of GAG and the volume it is in are small, then centrifugal partition will allow its separation from salt and smaller oligosaccharides, and permit it to be recovered in a small volume. The use of a high-quality dialysis membrane with a closely defined pore size is crucial. There are many tricks for the dialysis of small volumes. These include placing a sample in a centrifuge tube with a hole in the cap covered with dialysis tubing; the tube is inverted and the sample dialyzed before being recovered in the tube by gentle centrifugation. The large volume of a dialyzate makes its analysis impractical, while centrifugal partition allows examination of both retentate and filtrate to confirm that there is neither loss of desired material nor retention of undesired material.

3.3 Estimation of Glycosaminoglycan Abundance

GAG abundance may be estimated, following isolation or while still attached to a PG, by an assay relying upon the interaction of the GAGs and a dye. The GAG–dye complex may be precipitated and the amount of dye present estimated, or a change in absorbance of the complex quantified. It is important to note that with the exception of the method of Bjornson[43] the response from each GAG is different. Following depolymerization GAGs show no response and the abundance of each may be determined in sequence by the reduction in the response following digestion with an appropriate enzyme. The presence of common eluents such as GuHCl will modify the response, but if appropriate controls are used responses will remain linear.

3.3.1 Alcian Blue Solution Method[44]

In this method 100 μL of sample is mixed with 1.2 mL of fresh 1.4 mg mL^{-1} alcian blue in 0.5 M sodium acetate and the absorbance is determined at 480 nm after 10 min. This assay is not very sensitive and suffers from interference from high levels of salt, but it is very rapid and is linear from 10 to 100 μg mL^{-1}.

3.3.2 Alcian Blue Precipitation Assay I[45]

This assay involves the precipitation of an alcian blue–GAG complex and subsequent dissociation of the dye using surfactants. It is possible to use $MgCl_2$ in a critical electrolyte concentration technique to prevent the binding of certain GAGs,[46] although at low concentrations all GAGs will bind.

To 1 mL of freshly prepared centrifuged dye solution (0.05% alcian blue in 50 mM sodium acetate–50 mM $MgCl_2$, pH 5.8) is added 20 μL of sample (1–10 μg). After 2 h at ambient temperature the mixture is centrifuged at 2000 g for 10 min, then the pellet is washed in 2 mL of ethanol and resuspended in 1 mL of 10% SDS in 50 mM sodium acetate (pH 5.8) and the absorbance is measured at 620 nm.

3.3.3 Alcian Blue Precipitation Assay of Bjornsson[43]

Bjornsson has published an alcian blue precipitation assay with increased sensitivity (10–800 ng in 10 μL) which, as interference from proteins is minimized, is suitable for the analysis of PGs and GAGs in urine and blood. Significantly this method has the important feature that a single, parabolic, standard curve will allow the estimation of all sulfated GAGs.

Samples (10 μL) are added to a microtiter plate well, 20 μL of 1:1 8 M GuHCl : reagent A (54 mM H_2SO_4–0.75% Triton™ X-100) is added and the mixture shaken for 15 min. To each well is added 200 μL of reagent B (5% alcian blue stock (stock is alcian blue dissolved in 18 mM H_2SO_4–0.4 M GuHCl such that a 1% solution has an absorbance of ca. 1.4 at 600 nm) in 18 mM H_2SO_4–0.25% Triton™ X-100; stable for 1 month). After shaking for 1 h, 200 μL of the mix is transferred to a dot-blot apparatus and drawn through a washed and blocked PVDF (poly(vinylidene difluoride)) membrane. Each well is washed twice with 200 μL of 50% ethanol in 5 mM $MgCl_2$ and the membrane is air dried. The blot is quantified with image analysis software and a conventional document scanner using RGB (red–green–blue) mode with two pieces of clear plastic below the blot. Note that several standards are required as differences in density are recorded at different positions across the scanning arm.

3.3.4 Safranin O[47]

A 50-μL aliquot containing ca. 10–100 ng of sample is placed on the side of a dot-blot aparatus well and 400 μL of 0.02% safranin O in 50 mM sodium acetate at pH 4.75 is added with the vacuum applied. The mixture is drawn through a washed PVDF membrane and following two washes with water and methanol the OD of the precipitate is determined at 490 nm.

3.3.5 1,9-Dimethylmethylene Blue (Taylor's Blue) Assay[48]

The dye is prepared by adding 16 mg of DMB to 5 mL of ethanol and subsequently adding this to 995 mL of water containing 2 g of sodium formate and 2 mL of 90% formic acid; the low pH (3.5) prevents interference by nucleic acids. If protected from light the solution is stable for months. Samples (40 μL containing ca. 1–5 μg mL^{-1}) are added to a microtiter plate well and 250 μL of dye is added. GAGs cause a change in the dye color from blue to purple and the absorbance of the solution may be determined at either 600 nm or 535 nm. The former is more sensitive although it exhibits decreasing absorbance with increasing amounts of GAG.

3.3.6 Reducing Sugar

Following depolymerization of a GAG each oligosaccharide possesses a reducing sugar in equilibrium between ring-closed and ring-open forms, and its abundance may be estimated using this sensitive assay (0–50 μM). Samples (25 μL) are mixed with 50 μL of 0.38 M sodium carbonate, 0.21 M glycine, 1.8 mM copper sulfate pentahydrate and 6.1 mM neocuproine HCl, then heated to 100 °C for 12 min. The solutions are cooled, diluted with 125 μL of water, centrifuged and the absorbance at 450 nm measured.

3.4 Compositional Analysis

3.4.1 Monosaccharide Analysis

Monosaccharide analysis is an important step in the preliminary identification of an unknown GAG. Samples are hydrolyzed in 0.2 mL of 4 M TFA (trifluoroacetic acid) at 95 °C for 4 h[49] then lyophilized. The spectrum of monosaccharides found in GAGs is relatively limited and so most general methods for monosaccharide analysis are appropriate.[50] The analysis of unmodified sugars by HPAEC is a standard Dionex CarboPac™ PA1 application with elution in 16 mM NaOH and pulsed amperometric detection (PAD).

3.4.2 Total Sugar Assay (Orcinol)[51]

This assay uses H_2SO_4 to hydrolyze glycosidic linkages and dehydrate the monosaccharides generated, which react with orcinol to give a colored product with a detection range to ca. 500 μM. Samples (200 μL) are ice cooled and 800 μL of fresh ice-cold 14 mM orcinol in conc. H_2SO_4 is added. Following mixing, the solution is heated to 80 °C for 15 min and the absorbance of the cooled solution is determined at 510 nm (as used by the original authors) or 420 nm[52] (which shows greater sensitivity). Each monosaccharide generates a different absorbance, so unknown or mixed samples must be interpreted carefully.

3.4.3 Sialic Acid by Warren Assay[53] as Modified by Chaplin[50]

Sialic acid is present on KS as a cap, although asialo chain caps, rare in skeletal KS, are abundant in corneal KS.[54] Nucleic acids interfere with this assay, which requires three reagents. Reagent A is prepared by adding 58 mL of orthophosphoric acid to 4.278 g of sodium metaperiodate dissolved in 4.0 mL of water, and making the solution up to 100 mL with water. Reagent B is 0.77 M sodium arsenite–0.5 M sodium sulfate in 0.1 M H_2SO_4, while reagent C is 42 mM 2-thiobarbituric acid–0.5 M sodium sulfate. Reagent A (40 μL) is added to 80-μL samples (ca. 5–500 μM) and the solution is mixed, as it is after the addition of each reagent. After 20 min at ambient temperature 400 μL of reagent B is added and after a further 5 min 1.2 mL of reagent C is added. The tubes are heated to 100 °C for 15 min, then cooled and the chromophore is extracted into 1 mL of cyclohexanone. The solution is centrifuged before the absorbance of the upper layer is determined at 549 nm.

3.4.4 Uronic Acid[55]

The repeat region of each GAG, except for KS, contains alternating uronic acid residues which may be estimated by reaction of carbazole with the hydrolyzed uronic acid. Apart from the heating step the assay should be carried out on ice; 250-μL aliquots are added, with mixing, to 1.5 mL of 950 mg of sodium tetraborate decahydrate in 100 mL of cold H_2SO_4 and the tubes are heated to 100 °C for 10 min. Then 50 μL of 125 mg of carbazole in 100 mL of ethanol (keep cold and protect from light and air) is added to the cooled tubes which are mixed and heated to 100 °C for 15 min after which the absorbance of the cooled samples is determined at 525 nm. Several compounds interfere with this assay, including CHAPS, neutral carbohydrates and protein. Heparin and HS also give anomalously high values.

3.4.5 Anthrone for Neutral Sugars[56]

The assay of neutral sugars is important in the determination of KS which contains Gal but no uronic acid. Samples (1 mL containing up to 100 μg) are layered on top of 5 mL of 14.4 mM anthrone in dilute H_2SO_4 (500 mL H_2SO_4–200 mL dH_2O) and then mixed after 10 min, heated to 100 °C for 10 min and the absorbance at 620 nm determined.

3.4.6 Elson–Morgan Assay for Hexosamine[57]

The original method[57] has been simplified by Reissig et al.[58] and Chaplin[50] to provide a sensitive (1–100 μM) assay. To 50 μL of 0.2 M di-potassium tetraborate tetrahydrate is added 250 μL of sample (hydrolyzed and resuspended in NaOH, pH 10) and the mixture is heated to 100 °C for 3 min. After cooling 1.5 mL of reagent B is added. Reagent B is prepared by adding 1.5 mL of water to 11 mL of HCl, then adding 87.5 mL of glacial acetic acid and 10 g of 4-(*N*,*N*-dimethylamino)benzaldehyde, the working solution being a 1:10 dilution of this stock in glacial acetic acid. The sample mix is kept at 37 °C for 20 min and the absorbance determined at 585 nm.

3.4.7 Assay for Sulfate Groups[59]

An aliquot of a hydrolyzed sample is resuspended in 200 μL of dH_2O and added to 3.8 mL of 0.245 M trichloroacetic acid and 1 mL of reagent A (2 g of gelatine is dissolved with heating in 400 mL of dH_2O and then held at 4 °C for 6 h before 20 g of BaCl is added; the reagent is stable for only 1 week at 4 °C). After 15 min the absorbance is determined at 500 nm. The assay, with which phosphate interferes, is linear from 20 to 200 μg of SO_4^{2-}, and the standard is 104 mM potassium sulfate (10 mg mL^{-1} SO_4^{2-}).

3.4.8 Assay for N-*Sulfate Groups*

Heparin and HS contain *N*-sulfated hexosamines which may be assayed by a determination of 2,5-anhydromannose, found at the reducing terminus following hydrolysis by nitrous acid. Residues lacking *N*-sulfation are stable, therefore each oligosaccharide generated has a formerly *N*-sulfated residue at the reducing terminus. The principle behind this assay is also made use of for the depolymerization of GAGs (see section 3.6.1).

Reagent A is a fresh 1 : 20 mix of butylnitrite in ethanol; B is 2 M HCl, C is 0.11 M ammonium sulfamate, D is 5% HCl and E is 4.3 mM indole in ethanol. A sample of $<50\,\mu g\,mL^{-1}$ uronic acid in 200 µL is added to 200 µL of reagents A and B and after 1 h at ambient temperature 200 µL of reagent C is added. After 1 h 1 mL of reagent D and 200 µL of reagent E are added, and the mixture is heated to 100 °C for 5 min. After cooling 1.2 mL of ethanol is added and the absorbance at 520 nm determined.

3.5 Characterization of Intact Carbohydrate Chains

3.5.1 Nuclear Magnetic Resonance

NMR is a powerful, nondestructive technique, allowing the analysis of intact GAGs and subsequent recovery for further analysis. Because a GAG pool represents a heterogeneous population many detailed aspects of structure, particularly those from the repeat regions, will not be discernible. However information on the presence of some sugars and sulfation information may be obtained from an intact polymer. Significantly more data may be derived from a variety of two-dimensional (2-D) experiments which are essential for a full characterization.

In KS from noncorneal tissues the predominant nonreducing chain caps are α(2-3)- and α(2-6)-linked sialic acid, along with Gal in fetal tissues. The abundance of these caps may be determined by distinctive H3ax and H3eq doublets at ca. 1.7 and 2.7 ppm (relative to 3-trimethylsilyl[$^2H_4$]propionate) for α(2-6)-linked and at ca. 1.8 and 2.8 ppm for α(2-3)-linked caps. Nonreducing terminal Gal has an H1 resonance at ca. 4.55–4.56 ppm and the presence of fucose can be determined by doublets at ca. 1.17 and 5.12 ppm which represent CH_3 and H1. A key signal at ca. 3.97 ppm is derived from a sulfated Gal, which along with GlcNAc signals at ca. 4.75 ppm provides a measure of sulfation in KS.

In CS the ratio of 4- and 6-sulfation may be seen from the distinctive positions of *N*-acetyl CH_3, i.e. ca. 2.04 and 2.025 ppm respectively. If the chain has been depolymerized by an eliminase such as chondroitin ABC lyase, then the nonreducing terminal uronic acids in each oligosaccharide will have characteristic signals which report upon the sulfation of that residue and the adjacent GalNAc. The signals from H1 and H4 fall at ca. 5.19 and 5.89 ppm if the adjacent GalNAc is 0- or 6-sulfated but at 5.27 and 5.97 ppm if it is 4-sulfated. If the ΔUA is 2-sulfated the GalNAc will be 6-sulfated and H1 and H4 fall at the distinctive positions of ca. 5.53 and 6.03 ppm respectively.

Samples need not be completely pure, although it is necessary to have significant amounts of material. However, complete ^1H characterization of an octasaccharide linkage region of CS has been reported where less than 125 µg of sample was available.[20]

3.5.2 Lectins

Lectins can be very useful in studies of GAGs and PGs; they do, however, often exhibit distinctive binding characteristics, so failure to bind must be carefully interpreted. A variety of lectins are commercially available and may be used in the same way as antibodies in affinity isolation of GAGs or PGs and ELISA type experiments (see section 2.6.2).

The elderberry-derived lectin *Sambucus nigra* agglutinin (SNA) has been shown to bind a single α(2-6)-linked NeuAc chain cap strongly while having only a low affinity for the α(2-3)-linked form[60–62] of the chain cap, which is bound by *Maackia amurensis II* lectin. SNA conjugated to CL-4B has been used to isolate KS and PGs; samples are loaded in PBS and bound material is eluted with 20 mM ethylenediamine in water.

Concanavalin A, which binds α-mannose residues, has been used to isolated free N-linked KS chains from a large pool of other O-linked GAGs. Materials are loaded onto an affinity column in PBS containing 5 mM Mn and Ca, both of which are required for binding, and bound chains are eluted with 200 mM α-methyl mannoside.

3.5.3 Enzyme-linked Immunosorbent Assay

Free GAG chains do not bind to microtiter plate wells and so, if unmodified, must be estimated by competitive ELISA (see section 2.6.2). If an unreduced oligosaccharide is available then it may be linked to poly-L-lysine by reductive amination and this conjugate coated onto a microtiter plate well. Controls should be used to confirm that the lysine is not interacting with the GAG and masking epitopes. An example of ELISA in GAG analysis is found in the work of Karamanos et al.[5] in which the HNK-l epitope (3-sulfated GlcA) was examined using inhibition ELISA.

3.5.4 Electrophoresis

Electrophoretic separation of GAGs enables the analysis of several samples simultaneously and is increasing in both resolution and sensitivity. Cellulose acetate electrophoresis separates GAGs at $0.5\,mA\,cm^{-1}$ for 1 h

in 100 mM pyridine–150 mM formic acid at pH 3.1.[63] Volpi[64] has a longer two-stage procedure using agarose gels for electrophoresis of GAG (1 mg mL^{-1} with a sensitivity to ca. 1 μg) samples in 40 mM barium acetate at pH 5.8 for 1 h at 60 mA and then in 50 mM 1,2-diaminopropane with acetic acid added to pH 9 for 2 h. Following electrophoresis the gel is placed in 1% cetyltrimethylammonium bromide and stained. Volpi[65] has also described a nitrocellulose method which allows quantification of GAGs in the 0.2–2.5 μg range. Samples are loaded onto a nitrocellulose plate which has been soaked in 0.1 M barium acetate at pH 5 for 5 min and electrophoresis is performed in the same buffer for 40 min at 20 mA. The plates are stained in 0.08% Azure A in water for 5 min and washed in water before scanning at 600 nm. Fast and slow heparin, DS, CS, and KS can be separated, although HA runs very close to DS.

3.6 General Techniques for Oligosaccharide Analysis

The general approach for the analysis of GAGs has conventionally involved depolymerization of the parent chain to generate oligosaccharides of manageable size and number which may be characterized. Techniques specific to a type of GAG are examined in sections 4 to 7 but there are several which have general applicability to all GAGs.

3.6.1 Nitrous Acid Depolymerization

Nitrous acid depolymerizes all GAGs containing susceptible bonds by deaminative hydrolysis. *N*-Acetylhexosamines are not cleaved and must be de-*N*-acetylated[66] by hydrazine, a reagent which is both toxic and flammable. To anhydrous KS (10 mg) and 70 mg of hydrazine sulfate catalyst is added 1 mL of anhydrous hydrazine and the mixture is heated to 98 °C. After 4 h the reaction is stopped by the addition of 1 mL of dH_2O and the de-*N*-acetylated KS is recovered by dialysis and lyophilization. To 10 mg of de-*N*-acetylated KS is added 1 mL of 3.9 M $NaNO_2$–0.28 M CH_3COOH at pH 4.1, and after 2 h at ambient temperature the reaction is terminated by cooling on ice and raising to pH 7.5 with 2 M NH_4OH. Fucose prevents de-*N*-acetylation of the GlcNAc to which it is attached leaving it resistant to nitrous acid cleavage and generating a pentasaccharide, or a longer oligomer if several adjacent residues are fucosylated; linkage regions are not depolymerized, and chain caps are found as trisaccharides.

Heparin and HS are de-*N*-acetylated in 70% hydrazine–1% hydrazine sulfate at 96 °C for 6 h, which minimizes the conversion of uronic acids to hydrazides, and then depolymerized in 5.5 M sodium nitrate–0.5 M H_2SO_4 for 15 h.

This method may also be used to depolymerize CS/DS chains in which case the reducing terminal residue is 2,5-anhydro-D-talose. Heparin and HS are cleaved[67] at *N*-sulfated glucosamine sites by nitrous acid at pH 1.5, or following de-*N*-acetylation, at de-*N*-acetylated glucosamine at pH 4. It should be recalled that this is the basis of an assay for *N*-sulfated groups (see section 3.4.8). If de-*N*-acetylation is performed then both sites may be cleaved by treatment with nitrous acid at pH 3, generating mainly disaccharides.

During nitrous acid cleavage there is a competing ring contraction reaction, most common with heparin and HS, which prevents cleavage and generates tetrasaccharides. This is a random process which will reduce yield but have no effect upon the relative composition.

In each case the *N*-substituent is lost which removes the strong ultraviolet (UV) 206 nm absorbance of KS, so detection must be by electrochemical means, or by the incorporation of a label into the reducing terminal residue generated by nitrous acid cleavage.

3.6.2 Labeling of Reducing Sugar

Oligosaccharide fragments generated by the action of an eliminase from CS, DS, HS or heparin have a strong absorbance at 232 nm, and KS oligosaccharides may be observed, nonspecifically, by a strong absorbance at 206 nm. However those oligosaccharides generated enzymatically by the action of animal hyaluronidases or nitrous acid depolymerization have no such chromophore.

In such cases, or to increase sensitivity, a label may be incorporated at the reducing sugar. The most commonly used methods involve the conjugation of fluorescent probes; with radiolabeling techniques there are additional hazards and requirements for monitoring.

Bigge et al.[68] demonstrated that 2-aminobenzoic acid (2-AA) may be used to label oligosaccharides nonselectively, with high efficiency, and without the loss of either sialic acid or fucose. Labeling can be done using a commercial kit, which introduces fewer contaminants visible upon analysis, or the following reagents: add 150 μL of acetic acid to 350 μL of DMSO (dimethyl sulfoxide, then 120 μL of this mixture to 8.8 mg of 2-AA and 100 μL of that mixture to 11.8 mg of sodium cyanoborohydride. Labeling is performed with an anhydrous sample to which is added 1–2 μL of reagent, and the mixture is heated to 65 °C. After 2 h 100 μL of dH_2O is added and the oligosaccharides are ready for analysis at $\lambda_{ex} = 315$ nm and $\lambda_{em} = 400$ nm.

The label modifies the chromatographic behavior of KS and CS oligosaccharides; during SEC samples are slightly retarded, although free label may elute at up to 2 column volumes beyond the total volume V_t. Further, as the label carries a negative charge each oligosaccharide will be bound more tightly during IEC than without the label.

Kuster and Harvey[69] reported that GAGs depolymerized in ammonium-ion-containing buffers (which are used because they are volatile) have a reduced labeling efficiency due to formation of a 1-amino sugar following cleavage which may not be labeled and was reported to occur at levels of up to 30%.

Toomre and Varki[70] have labeled oligosaccharides with biotinylated aminopyridine (BAP), a combined biotinylated fluorescent probe. While the fluorescent tag allows easy detection of the oligosaccharide, the biotin may be used to bind it to a (strept)avadin support during isolation, or to generate an affinity column for the isolation of an antibody or the examination of binding partners. The sample and 50× molar excess of BAP are lyophilized, 10 μL of coupling reagent (2 : 1 pyridine–acetic acid) is added and the mixture is heated to 80 °C. After 1 h 10 μL of coupling reagent containing 1.25 mg of borane dimethylamine is added and the mixture is reheated to 80 °C for 1 h. The authors report coupling efficiencies of 40–80%, with large, highly charged oligosaccharides having the lower efficiencies.

A radiolabel may be introduced by borotritide reduction with 10 mCi of NaB^3H_4 for 1 h prior to reduction of the remaining residues by borohydride reduction with 25 mM $NaBH_4$ overnight.

3.6.3 Gradient Polyacrylamide Gel Electrophoresis

Gradient polyacrylamide gel electrophoresis (PAGE) on long 20–30% gels with a 5% stacking gel in a discontinuous buffer system allows the sequence and structural analysis of large oligosaccharides in a technique very similar in approach to that involving the familiar nucleic acid sequencing gels.[71] As with all electrophoretic methods many samples can be examined simultaneously. Techniques of this sort have been used to examine the sequence of HS[71] and CS.[38]

A conventional resolving gel with a 20–30% gradient should be cast in 32-cm plates to give a ca. 26–28 cm resolving gel with a 5% stacking gel. Samples and standards in 10% glycerol should be loaded and a marker containing phenol red used. Using a stacking gel buffer of 25 mM TRIS · HCl at pH 6.8 and a resolving gel buffer of 192 mM glycine at pH 8.3 the samples should be run into the gel at ca. 150 V for 30 min and be resolved at 300 V until the marker approaches the bottom (ca. 15 h). Note that disaccharides may migrate slightly in advance of the marker which should be in a separate lane. Several detection methods have been reported, including Azure-A or silver staining, allowing the detection of ca. 1 ng of an oligosaccharide, or labeled oligosaccharide, to be examined. Conventional wet or semi-dry electroblotting is possible using a positively charged nylon membrane for subsequent detection by staining, fluorography or autoradiography. Alternatively the oligosaccharide may be recovered from the membrane using NaCl.

3.6.4 Fluorochrome-assisted Carbohydrate Electrophoresis

This kit-based method (http://www.glyco.com) allows the electrophoretic separation of labeled oligosaccharides of 300–3000 Da on a precast polyacrylamide gel by charge/mass ratio as well as hydrodynamic volume. Oligosaccharides are labeled with the fluorescent tag 8-aminonaphthalene-1,3,6-trisulfonic acid disodium salt (ANTS) by reductive amination, which is reported to have 95–100% labeling efficiency and to be independent of composition. Following electrophoresis the gels are imaged using either a UV light box or a charge-coupled device (CCD) imaging system. Methods based on fluorochrome-assisted carbohydrate electrophoresis (FACE) have been used for the analysis of small oligosaccharides as well as intact chains.

4 HYALURONAN

HA, discovered by Myer and Palmer in 1934,[72] is composed of the repeating disaccharide [−4)GlcAβ(1-3) GlcNAcβ(1−]$_n$ (Figure 5) and may be up to 2×10^6 Da. Although unsulfated it carries a negative charge via carboxyl groups. It is not synthesized attached to a core protein and there are no reports of any substitutions or variance in the linear sequence.

There are a variety of hyaluronidases derived from animal and microbial sources. The latter are eliminases generating oligosaccharides with a nonreducing terminal ΔUA; in contrast those from animal sources generate a saturated product. At present only the enzyme isolated from *Streptomyces hyaluronolyticus* is specific for HA,[73] others act upon CS while chondroitin lyases act upon HA. To digest HA to ~50 000 Da use 4 U mg^{-1} at 20 °C, to digest it completely use 100 U mg^{-1} at 37 °C, both in 0.1 M acetate–0.15 M NaCl at pH 5.0. If the size of the HA chain is to be estimated then EDTA and cystein should be excluded from all buffers during isolation, to prevent production of radicals which can cleave HA so reducing its size.

4.1 Estimation of Hyaluronan Abundance

Jordian et al.[73] have used the specific lyase discussed in the previous section in an assay for HA. Digestion in 200 μL was terminated by the addition of 250 μL of 0.04 M $NaIO_4$ in 0.08 N H_2SO_4. After oxidation for 1 h at 37 °C excess periodate was removed by addition of 0.5 mL of 3% $NaAsO_2$ in 0.5 N HCl. After 30 min at

Table 2 Elution positions of HA oligosaccharides following HPAEC

Oligosaccharide	Code	Elution time (min)
$\Delta UA\beta^{(1-3)}GlcNAc\text{-ol}$	Δdi-HA	22.9
$\Delta UA\beta^{(1-3)}GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)}GlcNAc\text{-ol}$	Δtetra-HA	28.4
$\Delta UA\beta^{(1-3)}(GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)})_2GlcNAc\text{-ol}$	Δhexa-HA	31.7
$\Delta UA\beta^{(1-3)}(GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)})_3GlcNAc\text{-ol}$	Δocta-HA	34.6
$\Delta UA\beta^{(1-3)}(GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)})_4GlcNAc\text{-ol}$	Δdeca-HA	36.5
$\Delta UA\beta^{(1-3)}(GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)})_5GlcNAc\text{-ol}$	Δduodeca-HA	38.3
$\Delta UA\beta^{(1-3)}(GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)})_6GlcNAc\text{-ol}$	Δtetradeca-HA	39.7
$\Delta UA\beta^{(1-3)}(GlcNAc\beta^{(1-4)}GlcA\beta^{(1-3)})_7GlcNAc\text{-ol}$	Δhexadeca-HA	40.8

ambient temperature 4 mL of 0.3% thiobarbituic acid in 0.012 N HCl was added and the mixture heated to 100 °C for 15 min. Upon cooling 4 mL of 5% HCl in butanol was added and following mixing the absorbance of the butanol layer was determined at 549 nm. The authors report that this method is more sensitive than a determination of absorbance at 232 nm, and they identified 5 μg of HA in a 100-μL tissue extract.

If the HA is pure an estimation of uronic acids (see section 3.4.4) is also appropriate, but as this method is not specific for HA other GAGs, except KS, will interfere.

4.2 Analysis of Enzymatic Digestion Products

Price et al.[74] determined the size of HA oligosaccharides up to 16 mers by SEC using a Bio-Gel® P-30 column eluted in 250 mM NH_4HCO_3. As HA has an invariant chain the method may be reliably used to determine oligosaccharide length. The same authors also report that oligosaccharides within the same size range yielded negative-ion electrospray ionization mass spectra following analysis of a 16 μg mL^{-1} sample in 50% methanol and 0.3% Et_3N with capillary and cone voltages of 2.5 kV and 25 V respectively.

Several methods for the analysis of unsaturated CS oligosaccharides have also been calibrated for the analysis of the unsaturated HA disaccharide[75] with UV and PAD detection. Karamanos et al.[76] have calibrated its elution following high-performance liquid chromatography (HPLC) using an Econosphere™ NH2 column eluted with 5 mM sodium dihydrogen orthophosphate at pH 2.55, and also following capillary zone electrophoresis (CZE).[77] Price et al.[74] examined HA oligosaccharides up to 16 mers by HPAEC using a PA1 column and acetate as eluent with PAD detection. However, the method relies on a series of isocratic elutions and a prior SEC step is required as differently sized oligosaccharides co-elute. Lauder et al.[78] also reported an HPAEC method using a PA1 column for the analysis of HA oligosaccharides up to 16 mers which uses the same elution conditions as the analysis of CS repeat region[78] and linkage region[79] oligosaccharides (see sections 6.3.1 and 6.4). A CarboPac™ PA1 column (4 mm × 250 mm) maintained at 30 °C is eluted at 1 mL min^{-1}; a 12-min isocratic period of 98% 0.1 M NaOH–2% 1.3 M NaCl in 0.1 M NaOH was followed by a linear gradient of 2–46% 1.3 M NaCl in 0.1 M NaOH over 50 min; 46–87% 1.3 M NaCl in 0.1 M NaOH over 8 min and subsequently 87–100% 1.3 M NaCl in 0.1 M NaOH over 6 min followed by a 4-min isocratic phase of 100% 1.3 M NaCl in 0.1 M NaOH. The eluted oligosaccharides were monitored on-line by absorbance at 232 nm, and the elution times are reported in Table 2.

Toffanin et al.[80] reported the complete assignment of ^1H- and ^{13}C-NMR spectra for HA di- and tetrasaccharides prepared by bovine testicular hyaluronidase.

5 KERATAN SULFATE

KS is based upon a repeating sequence of −3Galβ1-4 GlcNAcβ1− which is usually sulfated on C(6) of GlcNAc, and further sulfate groups may be present on C(6) of Gal[81] (Figure 2 and Figure 5). Classification is based upon the type of linkage to the protein and the presence of specific structures. N-linked KS, originally isolated from cornea although now known to be present in cartilage,[7] is termed KS-I, while that which is O-linked is KS-II.[82] A third type, which is O-linked from mannose to serine or threonine, has been isolated from brain tissue.[83] Types I and II are further spilt based upon the presence of α(2-6)-linked NeuAc chain caps and fucose α(1-3)-linked to GlcNAc, both of which are present in articular but not nonarticular cartilage.[8,9,84] Lauder et al.[85] and Brown et al.[86] demonstrated for N- and O-linked chains respectively that the abundance of these tissue-specific structures is age-dependent, being absent in fetal and young tissue but appearing during maturity. Where fucose is present the Gal on the reducing side of the fucosylated GlcNAc may or may not be sulfated, while that on the nonreducing side is never sulfated. The length of KS chains is highly variable, having been reported as up to 46 disaccharides in cornea[54,87] and as short as 6–10 disaccharides in fibromodulin.[7–10] Corneal KS has a

very complex capping structure with α-Gal, $NeuGc^{2-6}$, $NeuGc^{2-3}$, GlcNAc(S)β^{1-3}, and GalNAc(S)β^{1-3}, in addition to α(2-3)- and α(2-6)-linked sialic acid caps.[54]

5.1 Characterization

5.1.1 Size

The hydrodynamic size of an isolated KS chain, or derived oligosaccharide, may be determined by SEC on a TSK™-30 column eluted at 30 °C in 0.2 M NaCl at 0.5 mL min^{-1}.[88] The K_{av} is determined and Equation (1) used to determine M_r. The method is valid for linear KSs with a K_{av} of between 0.15 and 0.65, but sialic-acid-containing oligosaccharides behave anomalously.

$$\mathrm{Log}_{10} M_r = 4.588 - 2.128K_{av} \quad (1)$$

5.1.2 Enzyme-linked Immunosorbent Assay and Antibodies

There are several anti-KS antibodies available, of which 5D4 and others of a similar type are the most common. They recognize a highly sulfated region of KS,[89] which comprises a small but variable percentage of the total KS chains.[7–9] The KS from mice (in which these antibodies are raised) and also from brain has a low level of sulfation and is not highly immunogenic with respect to 5D4 types of antibody, therefore estimation of KS by this method is not always appropriate.

5.2 Depolymerizing Keratan Sulfate

There are several methods for the depolymerization of KS (Figure 2, Table 3) which have been reviewed in depth by Brown and Nieduszynski[87] (see also section 3.6.1).

5.2.1 Endo-β-*Galactosidase (EC 3.2.1.103)*

Endo-β-Galactosidases, from *Escherischia freundii*,[90] *Flavobacterium keratolyticus*[91] and *Bacteroidies fragalis*,[92] hydrolyze any poly-*N*-acetyllactosamine chain, not just that of KS, at β(1-4) glycosidic linkages in which both the Gal and GlcNAc are unsubstituted (Figure 2).[93] Most KSs have a high level of GlcNAc sulfation and thus there are few cleavage sites, so oligosaccharides generated may be large and heterogeneous. However, if the KS has significant levels of unsulfated residues, then smaller oligosaccharides will be observed. Depolymerization should be carried out in 0.1 M TRIS · HCl at pH 7.2 with 0.1 U mg^{-1} of KS at 37 °C for 24 h.

Table 3 Methodologies used for analysis of KS

Depolymerization method	Oligosaccharides examined[a]	Separation	Detection
Keratanase II[137]	C, L, R	HPLC	206 nm
Keratanase II[9,85]	C, L, R	HPAEC	PAD
Keratanase II[96]	C, R	HPAEC	Fluorescence
Keratanase[7]	C, R	HPLC	206 nm
Nitrous acid[b][95]	R	HPAEC	^3H

[a] C = cap; L = link; R = repeat region.
[b] With prior deacetylation.

5.2.2 Keratanase (EC 3.2.1.103)

Keratanase, an *endo*-β-galactosidase from *Pseudomonas sp.*,[94] will cleave β(1-4) glycosidic linkages in which the Gal is unsulfated and the GlcNAc is sulfated but unfucosylated (Figure 2). This enzyme differs from other *endo*-β-galactosidases in its requirement for a sulfated GlcNAc, resulting in the cleavage of most KS chains.

Following depolymerization in 0.2 M sodium acetate at pH 7.2 with 1 U per 2.3 mg of KS at 37 °C for 24 h, a range of oligosaccharides are generated each with an unsulfated Gal at the reducing terminus. As the enzyme is inhibited by Gal(6S) the internal Gal residues will, with the exception of an unsulfated disaccharide, be fully sulfated unless fucose is present. Fucose inhibits the enzyme (Figure 2) and the oligosaccharide generated will have an unsulfated Gal on the nonreducing side of the fucosylated GalNAc (Table 4).

A sialidase contaminant has been identified in keratanase and 5 mM 2,3-dehydro-2-deoxy-*N*-acetylneuraminic acid should be used to inhibit sialic acid loss; any asialo caps generated will have an odd number of residues.[7] Keratanase has been used to examine O- and N-linked KS,[7] the oligosaccharides generated being isolated by HPLC using a Nucleosil™ 5SB eluted in 0–0.5 M $LiClO_4$ at 2 mL min^{-1} and detection by absorbance at 206 nm. Work by Tai et al.[54] has resolved a series of corneal KS oligosaccharides of up to 46 disaccharides by this method.

5.2.3 Keratanase II (EC Not Defined)

Keratanase II from *Bacillus sp.*[93] cleaves KS at a sulfated GlcNAc, fucosylated or not, adjacent to a sulfated or unsulfated galactose (Figure 2) at 0.002 U mg^{-1} KS in 10 mM sodium acetate at pH 6.8 at 37 °C for 30 h. The enzyme, which cannot act on a tetrasaccharide[87] yielding di- and tetrasaccharides from the repeat region and tri- and pentasaccharide caps, has been used to characterize N-[7–10] and O-linked[87,95] KSs from various sources.

5.3 Disaccharide Analysis

The disaccharide composition of KS is best determined by analysis of the oligosaccharides generated by nitrous acid cleavage which principally generates disaccharides,

Table 4 Elution positions of KS oligosaccharides generated by keratanase II digestion. The elution positions of keratanase II derived oligosaccharides are shown following analysis using the Dionex-based AS4A-SC methods described and PAD or fluorescent detection. See Figure 8 for chromatography conditions

Peak	Code	Oligosaccharide	Reduced	Relative molar response	Fluorescent label
			Elution time (min)		Elution time (min)
1	F1	Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)–	2.7	1.31	4.6
		Galβ$^{1-4}$GlcNAc(6S)β$^{1-6}$(Gal$^{1-3}$)GalNAc-ol	4.1	–	
2	R1	Galβ$^{1-4}$GlcNAc(6S)–	4.3	1.69	9.6
		Galβ$^{1-4}$GlcNAc(6S)β$^{1-6}$(NeuAcα$^{2-3}$Galβ$^{1-3}$)GalNAc–	11.2	–	
3	F2	Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)–	11.4	–	13.1
4	F3	Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$ Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)–	12.4	2.85	15.2
5	F4	Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)–	12.9	1.47	17.0
		Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)β$^{1-6}$(Galβ$^{1-3}$)GalNAc–	13.6	–	
6	R2	Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)–	14.4	1.67	19.1
7	C1a	NeuAcα$^{2-6}$Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)–	17.2	–	21.9
8	C1b	NeuAcα$^{2-3}$ Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)–	17.7	1.09	22.2
		NeuAcα$^{2-3}$Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$ Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)–	15.7	–	
		Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-6}$(Galβ$^{1-3}$)GalNAc–	16.5	–	
9	R3	Gal(6S)β$^{1-4}$GlcNAc(6S)–	16.9	1.00	24.6
		Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$ Galβ$^{1-4}$GlcNAc(6S)β$^{1-6}$(NeuAcα$^{2-3}$Galβ$^{1-3}$)GalNAc–	17.2	–	
		Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-6}$(NeuAcα$^{2-3}$Gal$^{1-3}$)GalNAc–	20.3	–	
10	F5	Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)–	23.2	2.05	26.3
		Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-3}$ Galβ$^{1-4}$GlcNAc(6S)β$^{1-6}$(Galβ$^{1-3}$)GalNAc–	24.3		
11	R4	Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)–	26.2	1.55	30.9
12	F6	Gal(6S)β$^{1-4}$(Fucα$^{1-3}$)GlcNAc(6S)β$^{1-3}$Galβ$^{1-4}$GlcNAc(6S)–	24.9	2.35	34.1
13	R5	Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Gal(6S)β$^{1-4}$GlcNAc(6S)–	27.5	3.00	36.7
		Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-3}$ Galβ$^{1-4}$GlcNAc(6S)β$^{1-6}$(NeuAcα$^{2-3}$ Galβ$^{1-3}$)GalNAc–	27.5	–	
14	C2a	NeuAcα$^{2-6}$ Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Gal(6S)β$^{1-4}$GlcNAc(6S)–	29.9	2.64	38.1
15	C2b	NeuAcα$^{2-3}$Galβ$^{1-4}$GlcNAc(6S)β$^{1-3}$Gal(6S)β$^{1-4}$GlcNAc(6S)–	30.8	2.38	38.9
16	R6	Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-3}$Gal(6S)β$^{1-4}$GlcNAc(6S)–	45.6	2.24	51.1
17	C3	NeuAcα$^{2-3}$Gal(6S)β$^{1-4}$GlcNAc(6S)β$^{1-3}$Gal(6S)β$^{1-4}$GlcNAc(6S)–	48.2	3.71	52.3

larger oligosaccharides occurring only if fucose is present. However, it should be recalled that the oligosaccharides have no absorbance at 206 nm and must be detected by electrochemical means or through the introduction of a label. Brown et al.[95] examined KS from articular cartilage in this way, the oligosaccharides being radiolabeled by borotritide reduction and eluted from a Nucleosil™ 5SB column with a gradient of 2 mM to 0.5 M $LiClO_4$. Alternatively the sample may be applied to an AS4A-SC column and eluted at 1 mL min^{-1} with a 5-min isocratic period of 50 mM NaOH and then in a gradient of 0–0.75 M sodium acetate over 30 min with constant 50 mM NaOH.

5.4 Fingerprinting Keratan Sulfate

For the full analysis of KS structure, keratanase II digestion is the best method for depolymerization; caps, linkage regions and repeat regions are found in a small number of oligosaccharides from which fucosylation, sialylation and sulfation levels may be determined. Digestion of a PG with the KS still attached results in only cap and repeat region oligosaccharides being generated, the linkage region remaining attached to the core protein.

Oligosaccharides have been examined by IEC on a Spherisorb® 5S column with a gradient of 2–250 mM $LiClO_4$ and UV detection at 206 nm (Figure 6).[7] This technique is sensitive and has good enough resolution that an analytical column may be used to fingerprint the oligosaccharides. However, the method requires long run times and lacks the excellent reproducibility of the Dionex AS4A-SC column, which, allied with PAD detection, allows the sensitive detection of unlabelled oligosaccharides. A 20-μL aliquot of reduced oligosaccharides is applied to a Dionex AS4A-SC column (4 mm × 250 mm) at 30 °C and running at 1 mL min^{-1}. The eluent is monitored by PAD and oligosaccharides eluted by a 5-min isocratic period of 5% 1 M NaOH–95% water, followed by a linear gradient of 0–95% 1.5 M sodium acetate and constant 5% NaOH over 50 min. The PAD used program 1 (carbohydrates), and the following time/potential program: 0.00 s, +0.10 V; 0.50 s, +0.10 V; 0.51 s, +0.60 V; 0.59 s, +0.60 V; 0.60 s, −0.60 V; 0.65 s, −0.60 V, integration

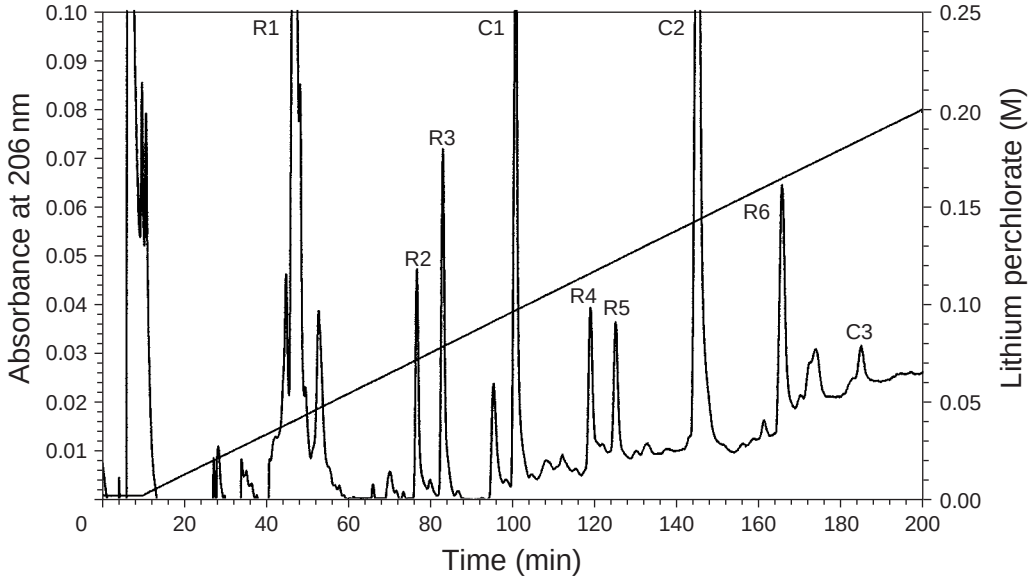

Figure 6 Strong anion exchange HPLC analysis of a keratanase II digest of KS. An aliquot (20 μL) of reduced oligosaccharides derived by keratanase II digestion of the KS attached to tracheal cartilage fibromodulin was applied to a Spherisorb® 5S column and eluted at 1 mL min^{-1} in a gradient of 2–250 mM $LiClO_4$ with UV detection at 206 nm. The KS was depolymerized while attached to the PG and hence no linkage regions are observed. For oligosaccharide codes see Table 4.

being 0.3–0.5 s. A sample chromatogram is shown in Figure 8 and the elution times are given in Table 4. The method allows the detection of ca. 10 ng of a single oligosaccharide and resolves α(2-3)- and α(2-6)-linked sialic acid caps, as well as fucosylated oligosaccharides. However, PAD suffers from a nonmolar response; for fully quantitative data, response factors must be determined for each oligosaccharide examined.[90]

Whitham et al.[96] have developed a fingerprinting method for keratanase II oligosaccharides labeled with the fluorescent probe 2-AA (see section 3.6.2). An aliquot is applied to a Dionex AS4A-SC column at 50 °C and eluted at 2 mL min^{-1}. After 5 min of constant 150 mM NaOH, bound oligosaccharides are eluted by a gradient of 0–600 mM NaCl with constant 150 mM NaOH over 60 min. A sample chromatogram is shown in Figure 8 and the elution times are given in Table 4. The same protocol has been employed by Lauder et al. for fingerprinting of CS linkage[37] and repeat regions[37] (see sections 6.3 and 6.4), allowing a coordinated approach to the analysis of KS and CS.

6 CHONDROITIN SULFATE/DERMATAN SULFATE

The repeat region of CS comprises a repeating disaccharide −4)GlcAβ(1-3)GalNAcβ(1−, while in DS D-GlcA is replaced by L-IdoA through epimerization at C(5) (Figures 3, 5 and 7). In mammals CS may be sulfated on C4 and/or C6 of GalNAc and C2 of GlcA. Variations found elsewhere include sulfation on C3 of GlcA residues[97] and fucosylation of GalNAc.[98] In DS the sulfation pattern is much simpler, with GalNAc being

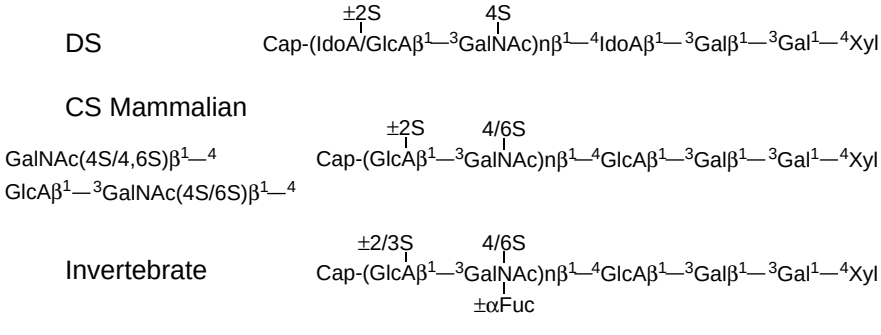

Figure 7 Structural models for CS/DS.

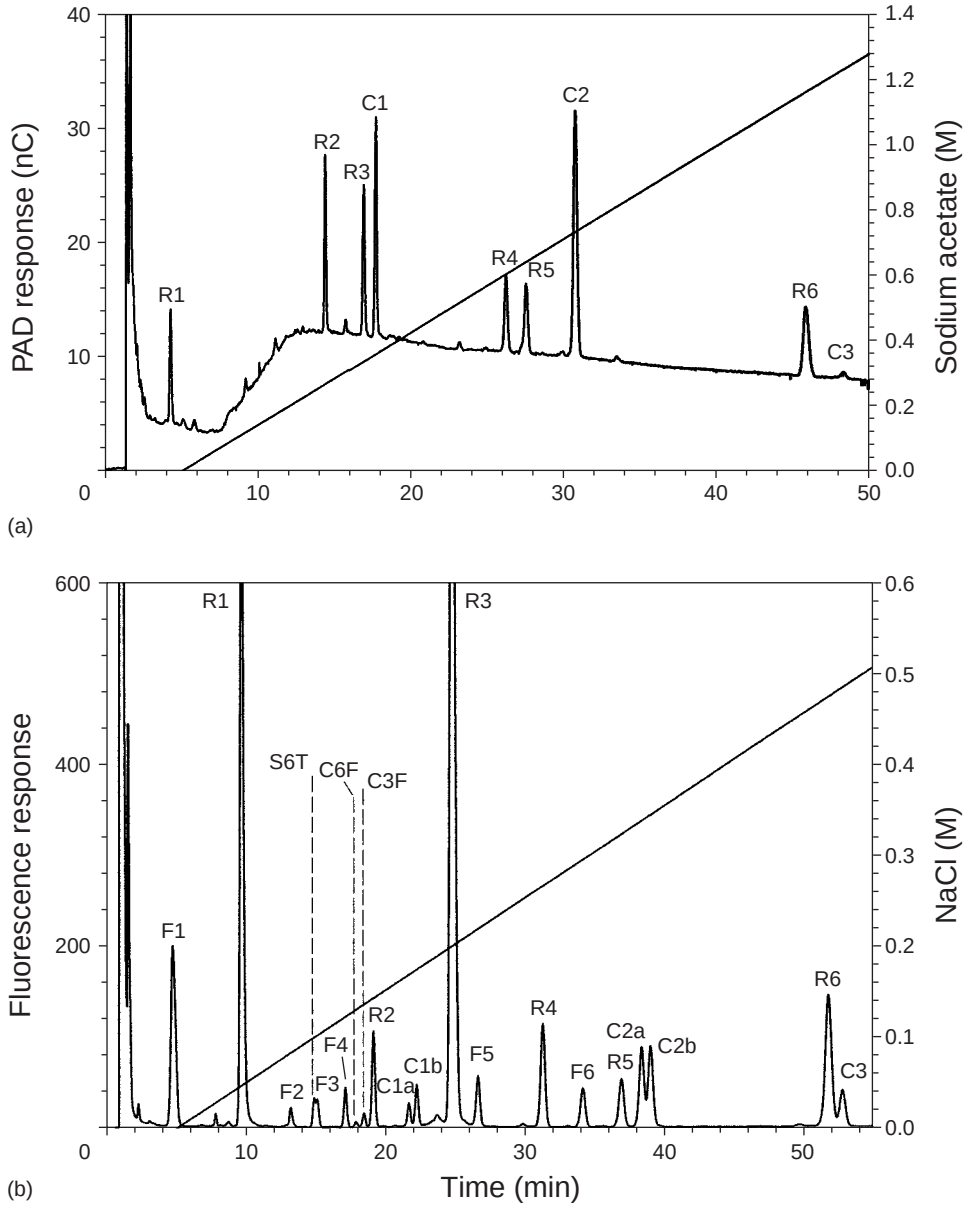

Figure 8 HPAEC analysis of a keratanase II digest of KS. Aliquots (20 μL) of oligosaccharides derived by keratanase II digestion of KS were applied to a Dionex AS4A-SC column and eluted (a) at 30 °C with a flow of 1 mL min^{-1} of 0%–95% 1.5 M sodium acetate with constant 50 mM NaOH and detected by PAD (N-linked KS from fibromodulin), or (b) at 50 °C with a flow of 2 mL min^{-1} of 0–600 mM NaCl with constant 150 mM NaOH and fluorescent detection of the 2-AA labeled oligosaccharides at $\lambda_{ex} = 315$ nm and $\lambda_{em} = 400$ nm (O-linked KS from aggrecan). Note the relatively abundant chain cap oligosaccharides in the shorter KS from fibromodulin. For oligosaccharide codes see Table 4.

almost 100% 4-sulfated. Although the functions of these GAGs are yet to be fully elucidated, it is known that CS has a role in articular cartilage where the osmotic swelling it causes helps load bearing. Other data suggest a more direct role in macromolecular interactions; it has been shown[99,100] that the expression of some CS epitopes in the rodent is developmentally regulated and chondroitin 4-sulfate has a role in the cytoadherence of malaria infected red blood cells,[101] the adherence of the malaria parasite to human placenta[28] and the regulation of neurite outgrowth.[102] There is evidence that a CS PG, appican, has a role in Alzheimer's disease[103] and DS has been shown to bind to the growth promoter spermine.[30]

A sulfated GalNAc has been shown to be the major nonreducing terminal residue of swarm rat chondrosarcoma CS:[104] 85–90% of chains isolated were capped by either GalNAc4S or GalNAc4,6S. However, there are very significant age and pathology related changes in the

abundance of GalNAc4,6S in human articular cartilage aggrecan: it is almost absent in fetal tissue while terminating ca. 60% of chains in the normal adult[105] but only ca. 30% in osteoarthritic cartilage;[106] GalNAc4,6S is very rare within the repeat region.

There are reports of Gal sulfation in the linkage region of CS isolated from aggrecan of shark cartilage,[107,108] bovine nasal septum,[109,110] articular cartilage[79] and swarm rat chondrosarcoma.[111] Moses et al.[112] determined that during the elongation of the linkage region there is transient xylose phosphorylation, as observed in CS from swarm rat chondrosarcoma,[113] but that this ester is cleaved after elongation of the linkage region to a trisaccharide but prior to elaboration of the complete CS chain. The repeat region close to the linkage region has a higher than average incidence of unsulfated and 4-sulfated GalNAc residues.[79] Sugahara et al.[114] identified a DS-specific linkage region tetrasaccharide containing IdoA rather than glucuronic acid, i.e. −Idoα(1-3)Galβ(1-3)Galβ(1-4)Xyl.

Given the range of sulfation possibilities there are a very large number of potential CS di- ($n = 8$), tetra- ($n = 64$) and hexasaccharides ($n = 512$). Including C(5) epimerization to IdoA raises these numbers to 16, 256 and 4096 respectively. However, not all are observed because (a) the C(4)/C(5) unsaturated bond of the ΔUA destroys the distinction between GlcA and IdoA, (b) the specificity of chondroitin lyases is such that cleavage occurs preferentially at 4-sulfated GalNAc residues, forcing them to the reducing terminus of oligosaccharides longer than disaccharides[115] and (c) the pattern of sulfation of CS is not random, 2-sulfated uronic acids being only found between a 4-sulfated GalNAc on the nonreducing side and a 6-sulfated GalNAc on the reducing side.[116,117]

6.1 Characterization

6.1.1 Iduronic Acid Levels

Karamanos et al.[118] devised a method for the determination of IdoA and GlcA levels in CS/DS chains, avoiding the serious losses normally associated with hydrolysis. The uronic acids are converted to their corresponding neutral sugars, glucose and 1,6-anhydroidose which are stable during hyrolysis. The GAG (1–250 µg in 50 µL) is mixed with 1 mg of *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide for 10 min and 5 µL of 0.04 M HCl is added; after 50 min a further 5 µL is added. After 1 h 100 µL of 2 M $NaBH_4$ is added and after a further 1 h at 50 °C excess borohydride is destroyed by 25 µL of glacial acetic acid and the GAG chain is recovered. Following hydrolysis per-*O*-benzoylated derivatives were examined by RPC using a Supelcosil™ LC-18 column eluted with 75% acetonitrile.[118] The derivatives were prepared by incubating the deionized monosaccharides with 500 µL of pyridine containing 10% benzoic anhydride and 5% 4-dimethylaminopyridine at 37 °C for 90 min. Following the addition of 4.5 mL of water the sample was de-salted on a Sep-Pak™ C-18 column washed with 10 mL of 10% pyridine in water and 5 mL of water and the sugars were eluted with 2.5 mL of acetonitrile. Lauder et al.[119] have quantified the unmodified Glc and 1,6-anhydroidose derived from the parent chain by hydrolysis, using HPAEC on a PA1 column with PAD under the same conditions used for monosaccharide analysis (see section 3.4).

6.2 Specificity of Enzymes Acting Upon Chondroitin Sulfate/Dermatan Sulfate

There are a number of enzymes which act upon CS/DS, either to depolymerize or to specifically remove sulfate esters (Figure 3). The naming of CS/DS lyase enzymes is partly based on a historical nomenclature whereby 4-sulfated CS was termed CS(A), 6-sulfate was termed CS(C) and DS was CS(B). Each lyase has a slightly different specificity but all fail to cleave a highly sulfated CS; specifically a bond between a 4- and 6-sulfated GalNAc and a 2-sulfated GlcA is resistant (Figure 3).[115] If highly sulfated CS is present SEC will allow the isolation of larger resistant regions of the chain for further analysis. Chondroitin ABC lyase will cleave both CS and DS; all oligosaccharides, except the cap, will have a nonreducing terminal ΔUA, removing the distinction between CS and DS at this residue. An IdoA will only be cleaved by a B lyase and not by A or C lyase. Oligosaccharides generated by the eliminase action of a chondroitin lyase have a strong chromaphore at 232 nm derived from the C(4) C(5) unsaturated bond of the ΔUA residue; molar extinction coefficient = 5500.[120] This value is similar, but not identical, for all oligosaccharides and allows sensitive detection without further modification.

6.2.1 Chondroitin ABC Lyase (EC 4.2.2.4)

Hamai et al.[121] identified two chondroitin ABC lyase activities, corresponding to an endo- and an exolyase, the endolyase being unable to cleave tetrasaccharides, whereas the exolyase can (Figure 3). These workers identified the Seikagaku Corporation protease free product as possessing solely the endolyase activity, enabling this enzyme to be used to generate tetrasaccharides from a limit digest.[37,78] Both enzymes can cleave chondroitin 4- and 6-sulfate as well as DS and unsulfated GalNAc. The activity towards the 4-sulfate is, however, significantly higher than that towards 6-sulfate, resulting in the preferential localization of 4-sulfated GalNAc residues at the reducing end of the resulting oligosaccharides.[37,78,115]

Sugahara et al.[97] determined that digestion with chondroitin ABC lyase will remove any GlcA 3-sulfate groups; to obtain oligosaccharides containing this structure testicular hyaluronidase must be used. Sugahara et al.[122] also determined that the enzyme will preferentially remove the nonreducing terminal disaccharide from a free hexasaccharide. To digest GAGs to completion use 0.01 U mg^{-1} CS at 37 °C for 24 h in 0.1 M TRIS · HCl at pH 8.

6.2.2 Chondroitin AC Lyase (EC 4.2.2.5)

There are three chondroitin AC lyases: ACI from *Flavobacterium heparinum*, ACII from *Arthrobacter aurescens* and ACIII from *Flavobacterium sp*. Hp102. Unlike chondroitin ABC endolyase these enzymes have no size restrictions on their activity. However, they differ in their ability to cleave bonds to different disaccharides: ACIII cannot cleave a bond to GlcA-GalNAc4,6S or GlcA2S-GalNAc6S, ACI can cleave the former but not the latter (Figure 3), and ACII is able to act upon both bonds, but not a very highly sulfated CS. Digestion should be performed using 0.1 U mg^{-1} in TRIS · HCl at pH 8 for 24 h at 37 °C.

6.2.3 Chondroitin 2-, 4- and 6-Sulfatase

These enzymes specifically remove a sulfate ester from C(2) of a uronic acid, and C(4) or C(6) of a GalNAc residue, respectively (Figure 3). The latter acts solely on the reducing terminal GalNAc[122] while chondroitin 4-sulfatase alone also acts upon the adjacent internal GalNAc residue. However, the 4-sulfatase will not act upon a GalNAc4S,6S residue, while chondroitin 6-sulfatase will if it is located at the reducing terminus of an oligosaccharide.[122] Digestion is often performed to examine the structure of an unknown oligosaccharide and may be performed at 1 mU $mmol^{-1}$ in a minimum volume (ca. 20 μL) of TRIS · HCl at pH 8 for 20 min at 37 °C.

6.2.4 Other Enzymes

Both CS B and C lyase are commercially available and cleave only DS and GalNac(6S) respectively. HA lyase also cleaves CS, but at a lower rate than HA.

6.3 Disaccharide Analysis

A determination of the disaccharide composition of CS/DS will show levels of GalNAc 4- and 6-sulfation along with uronic acid sulfation; rarer structures such as Δdi-4,6S and Δdi-0S will be observed, but there will be no information about their position within the chain.

There are many techniques for the analysis of the disaccharides generated following depolymerization of CS by chondroitin ACII or conventional ABC lyase, and several methods of detection, including the use of absorbance at 232 nm, fluorescent labeling and PAD (Table 5). Midura et al.[75] calibrated an HPAEC method using a CarboPac™ PA1 column with elution in TFA and detection by PAD. However, detection by PAD suffers from diminishing response with increasing sulfation, the trisulfated disaccharide Δdi-tris being undetectable.

Karamanos et al.[76] reported an extensive examination of 24 disaccharides, including 3-sulfated GlcA, from CS, DS and HA by HPLC on an Econosphere™ NH2 column with UV detection at 231 nm. Elution was by a series of isocratic eluents: 5 mM sodium dihydrogen orthophosphate, pH 2.55, for nonsulfated; 50 mM sodium dihydrogen orthophosphate, pH 2.50, for monosulfated; and 50 mM sodium sulfate – 10 mM sodium acetate, pH 5.0, for di- and trisulfated disaccharides.

Table 5 Methodologies used for analysis of CS/DS

Depolymerization method	Maximum size	Separation method	Detection method
Repeat region			
ABC endolyase[78]	Hexasaccharide	HPAEC	232 nm
ABC endolyase[97]	Hexasaccharide	HPAEC	Fluorescence
ACII[123]	Disaccharide	HPLC, ion pair	232 nm
ACII[76]	Disaccharide	HPLC, IEC	232 nm
ACII[77]	Disaccharide	HPCE	232 nm
ACII, hyaluronidase[124]	Tetrasaccharide	CZE	Fluorescence
ACII, hyaluronidase[124]	Tetrasaccharide	HPLC, IEC	Fluorescence
Caps			
ACII lyase[106,107]	Disaccharide	HPAEC	Fluorescence
Linkage region			
ABC endolyase[79]	Octasaccharide	HPAEC	232 nm
ABC endolyase[37]	Octasaccharide	HPAEC	Fluorescence

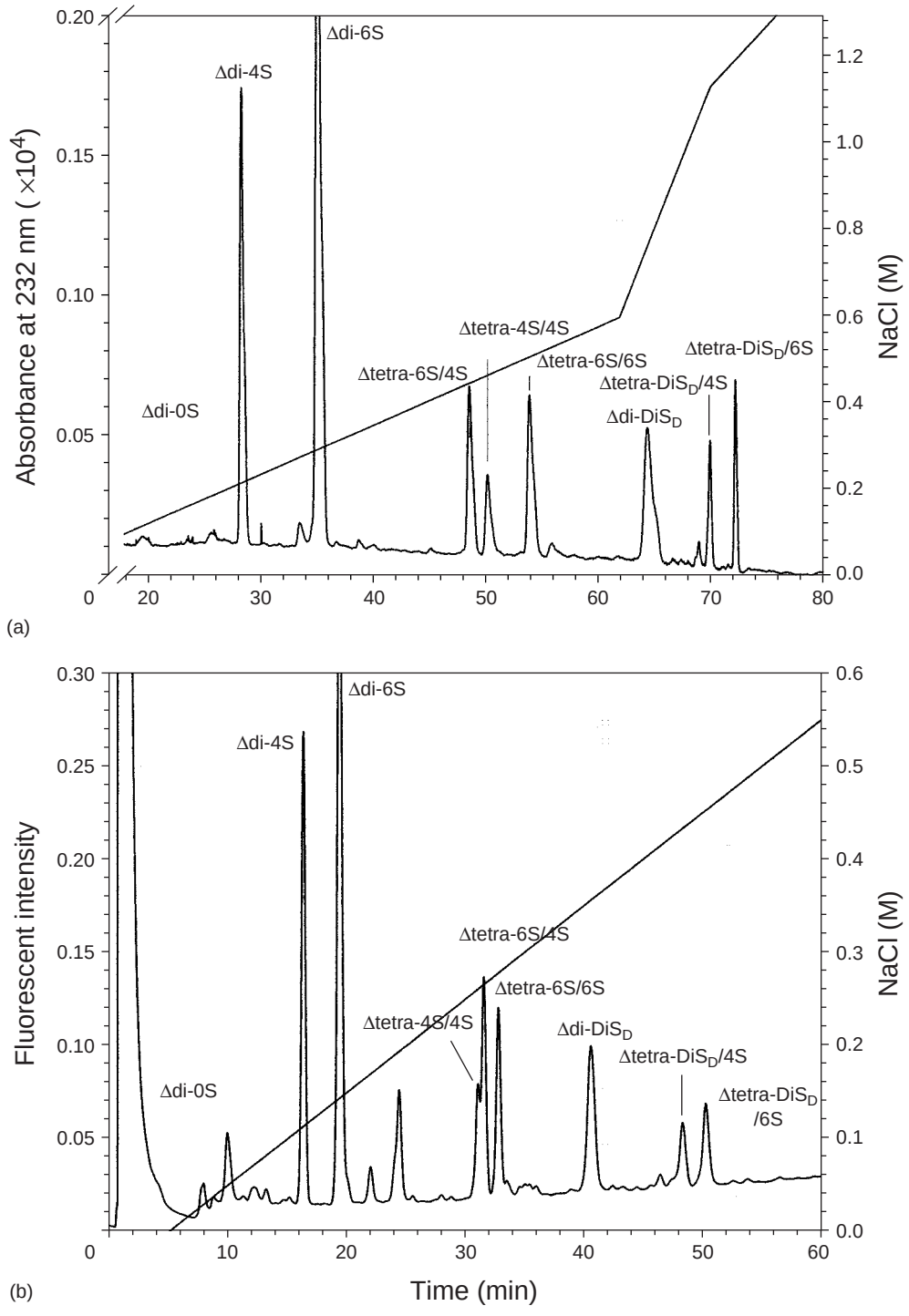

Figure 9 HPAEC analysis of CS oligosaccharides. An aliquot (20 μL) of chondroitin ABC endolyase digested CS was applied to (a) a Dionex CarboPac™ PA1 column (250 mm × 4 mm) maintained at 30 °C. Elution was at 1 mL min^{-1}; a 12-min isocratic period of 98% eluant A (0.1 M NaOH)–2% eluant B (1.3 M NaCl in 0.1 M NaOH) was followed by a linear gradient of 2–46% eluant B over 50 min, 46–87% eluant B over 8 min, and 87–100% eluant B over 6 min followed by a 4-min isocratic phase of 100% eluant B. The eluant was monitored by absorbance at 232 nm. (b) The aliquot was applied to a Dionex AS4A-SC column maintained at 50 °C running at 2 mL min^{-1} with 150 mM NaOH and after 5 min of isocratic elution a gradient of 0–600 mM NaCl over 60 min and fluorescent detection at $\lambda_{ex} = 315$ nm and $\lambda_{em} = 400$ nm.

Table 6 Elution positions of CS/DS oligosaccharides following HPAEC. The elution positions of CS/DS oligosaccharides are shown following analysis using the Dionex-based PA1 or AS4A-SC methods described with UV or fluorescent detection respectively. See Figure 9 for chromatography conditions

Oligosaccharide	Code	Oligosaccharide generated by the action of			Elution time (min)	
		ACII	ABC	C	PA1 reduced	AS4A-SC fluorescent label
ΔUAβ(1-3)GalNAc-ol	Δdi-0S	✓	✓		21.3	5.4
ΔUA2Sβ(1-3)GalNAc-ol	Δdi-UA2S	✓	✓		33.2	22.3
ΔUAβ(1-3)GalNAc4S-ol	Δdi-4S	✓	✓	×	29.7	17.2
ΔUAβ(1-3)GalNAc6S-ol	Δdi-6S	✓	✓	✓	36.3	20.1
ΔUA2Sβ(1-3)GalNAc4S-ol	Δdi-DiSB	✓	✓	×	57.1	35.3
ΔUA2Sβ(1-3)GalNAc6S-ol	Δdi-DiSD	✓	✓	✓	65.6	40.2
ΔUAβ(1-3)GalNAc4,6S-ol	Δdi-DiSE	✓	✓		56.4	34.2
ΔUA2Sβ(1-3)GalNAc4,6S-ol	Δdi-TRIS	✓	✓		75.4	54.5
ΔUAβ(1-3)GalNAc4Sβ(1-4)GlcA-ol	040	✓	✓	✓	37.9	
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcA-ol	060	✓	✓	✓	35.7	25.5
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)GalNAc4S-ol	Δtetra-6S/4S	×	✓	×	47.8	32.3
ΔUAβ(1-3)GalNAc4Sβ(1-4)GlcAβ(1-3)GalNAc4S-ol	Δtetra-4S/4S	×	✓	×	49.6	31.6
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)GalNAc6S-ol	Δtetra-6S/6S	×	✓	✓	53.4	33.6
ΔUAβ(1-3)GalNAc4Sα(1-4)IdoAα(1-3)GalNAc4S-ol	ΔtetraDS-4S/4S	×	✓	×	48.9	
ΔUA2Sβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)GalNAc4S-ol	Δtetra-DiSD/4S	×	✓	×	70.1	49
ΔUA2Sβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)GalNAc6S-ol	Δtetra-DiSD/6S	×	✓	✓	72.0	51
ΔUAβ(1-3)GalNAc4,6Sβ(1-4)GlcA2Sβ(1-3)GalNAc6S-ol	Δtetra-DiSE/DiSD	✓	×	✓	77.8	
ΔUA2Sβ(1-3)GalNAc4Sβ(1-4)IdoAα(1-3)GalNAc4S-ol	ΔtetraDS-DiSB/4S	×	✓	×	72.0	
ΔUAβ(1-3)(GalNAc6Sβ(1-4)GlcAβ(1-3))2GalNAc6S-ol	Δhexa-6S/6S/6S	×	✓	✓	62.7	
ΔUAβ(1-3)(GalNAc4Sβ(1-4)IdoAα(1-3))2GalNAc4S-ol	ΔtetraDS-4S/4S/4S	×	✓	×	62.8	

Oligosaccharides containing GlcA 3-sulfation and *N*-sulfated GalNAc were examined. A linear response was observed to 50 μg, with the detection limits being 58 ng for unsulfated, 8–11 ng for monosulfated, 12–15 ng for disulfated, and 25–30 ng for trisulfated disaccharides.

Karamanos et al.[77] reported a high-performance capillary electrophoresis method for the analysis of disaccharides from vertebrate CS, DS and HA, i.e. no 3-sulfated GlcA. The method allows a complete determination of the disaccharide profile within 14 min on a fused-silica capillary in 15 mM sodium dihydrogen orthophosphate, at pH 3, using reversed polarity at 20 kV with detection at 232 nm. Concentrations of 32 pmol L^{-1} (22 ng L^{-1}) were detectable and the method has been used to examine HA in human malignant mesothetioma.[125]

Karamanos et al.[135] also reported an ion pair method for the analysis of CS, DS and HA disaccharides. Using tetrabutylammonium as the ion-pair agent the oligosaccharides are analyzed using a conventional reversed-phase column with elution in acetonitrile. Many laboratories have such columns in use, whereas an ion-exchange column may not be available.

Lauder et al.[78] reported an HPAEC method for disaccharide analysis of CS/DS chains, using a Dionex CarboPac™ PA1 column and an elution protocol identical to that used for longer oligosaccharides, linkage regions (see section 6.4) and HA oligomers (see section 4.2) with detection at 232 nm. Two fluorescent-based HPAEC methods have been reported for disaccharides[126] and di- and tetrasaccharides.[37] The former labels with 2-aminopyridine and affects separation on an AS4A-SC column in a gradient of 0–500 mM sodium trifluroacetate. The latter labels with 2-AA (see section 3.6.2) and separates on an AS4A-SC column maintained at 50 °C running at 2 mL min^{-1} with 5 min of constant 150 mM NaOH after which bound oligosaccharides are eluted by a gradient of 0–600 mM NaCl with constant 150 mM NaOH over 60 min. A chromatogram is shown in Figure 9 and the elution times are reported in Table 6.

6.3.1 Longer Oligosaccharides

The methods reported above for disaccharide analysis can provide no information about larger blocks of structure. However, Lauder et al.[78] have reported an HPAEC method for the analysis of di-, tetra- and hexasaccharides derived from CS/DS chains. A CarboPac™ PA1 column (4 mm × 250 mm) maintained at 30 °C was eluted at 1 mL min^{-1}; a 12-min isocratic period of 98% 0.1 M

NaOH–2% 1.3 M NaCl in 0.1 M NaOH was followed by a linear gradient of 2–46% 1.3 M NaCl in 0.1 M NaOH over 50 min; 46–87% 1.3 M NaCl in 0.1 M NaOH over 8 min and subsequently 87–100% 1.3 M NaCl in 0.1 M NaOH over 6 min followed by a 4-min isocratic phase of 100% 1.3 M NaCl in 0.1 M NaOH; the eluted oligosaccharides are monitored by absorbance at 232 nm. A chromatogram is shown in Figure 9 and the elution times are reported in Table 6.

For fluorescently labeled oligosaccharides of tetrasaccharide and longer the method of Whitham et al.[96] has been adapted and the elution conditions are those

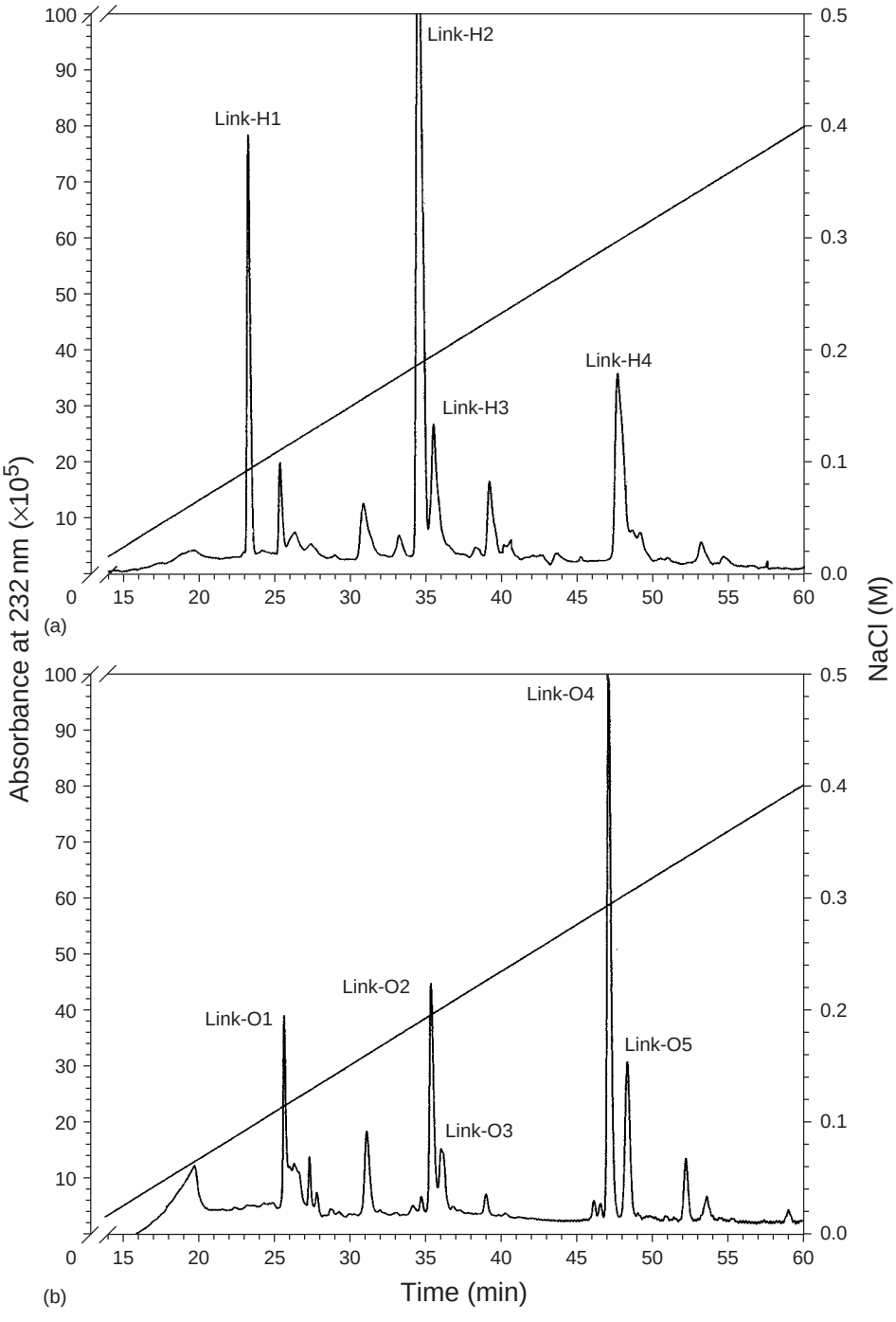

Figure 10 HPAEC analysis of CS linkage regions. An aliquot (20 µL) of CS linkage regions of (a) hexasaccharide and (b) octasaccharide size was applied to a Dionex CarboPac™ PA1 column and eluted as described for repeat region oligosaccharides (Figure 9). The oligosaccharide code is in Table 7.

Table 7 Elution positions of linkage region oligosaccharides following HPAEC. The elution positions of CS linkage region oligosaccharides are shown following analysis using the Dionex-based PA1 or AS4A-SC methods described with UV or fluorescent detection respectively. See Figure 10 for chromatography conditions

Oligosaccharide	Code	Oligosaccharide generated by the action of		Elution Time (min)	
		ACII	ABC	PA1 reduced	AS4A-SC fluorescent label
ΔUAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-T1	✓	×	20.4	10.8
ΔUAβ(1-3)Gal6Sβ(1-3)Galβ(1-4)Xyl-ol	Link-T3	✓	×	31.6	20.1
ΔUAβ(1-3)GalNAcβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-H1	×	✓	24.1	14.6
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-H2	×	✓	35.3	24.3
ΔUAβ(1-3)GalNAc4Sβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-H3	×	✓	36.1	23.1
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)Gal6Sβ(1-3)Galβ(1-4)Xyl-ol	Link-H4	×	✓	48.2	33.1
ΔUAβ(1-3)GalNAcβ(1-4)GlcAβ(1-3)GalNAcβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-O1	×	✓	26.0	
ΔUAβ(1-3)GalNAcβ(1-4)GlcAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-O2	×	✓	36.1	
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)GalNAcβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-O3	×	✓	35.6	
ΔUAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-O4	×	✓	47.7	
ΔUAβ(1-3)GalNAc4Sβ(1-4)GlcAβ(1-3)GalNAc6Sβ(1-4)GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xyl-ol	Link-O5	×	✓	49.3	

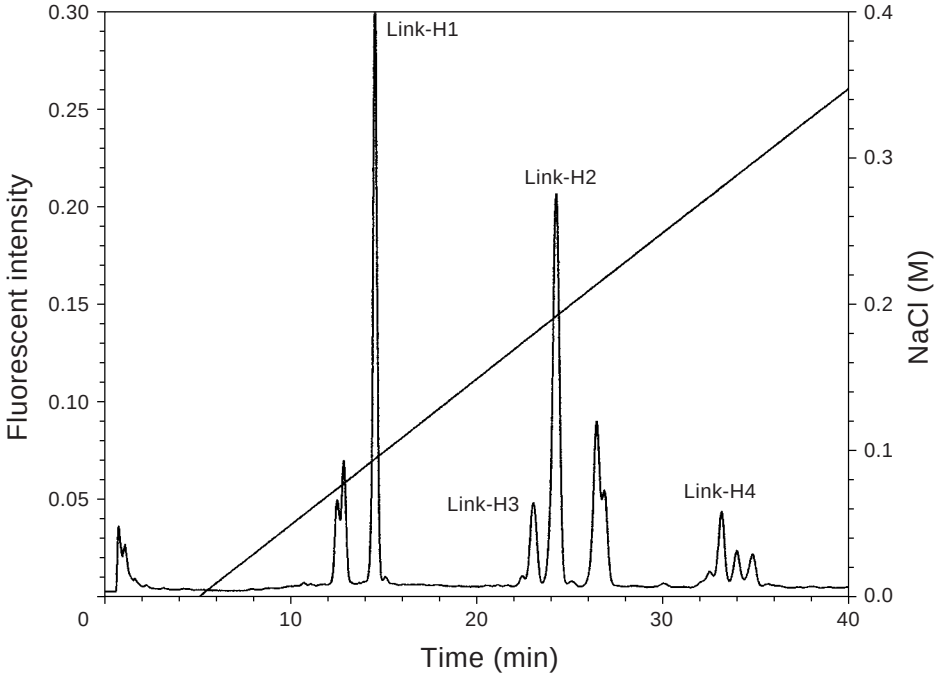

Figure 11 HPAEC analysis of fluorescently labeled CS linkage regions. An aliquot (1 μL) of hexasaccharide CS linkage regions released unreduced with LiOH and fluroescent labeling was applied to an AS4A-SC column eluted as described for fluorescently labeled repeat region oligosaccharides (Figure 9). The oligosaccharide code is in Table 7.

described previously for disaccharides, i.e. AS4A-SC column maintained at 50 °C running at 2 mL min^{-1} with 5 min of constant 150 mM NaOH after which bound oligosaccharides are eluted by a gradient of 0–600 mM NaCl with constant 150 mM NaOH over 60 min. A chromatogram is shown in Figure 9 and the elution times are reported in Table 6.

6.4 Linkage Region Analysis

The linkage regions of CS/DS have the general structure GlcAβ(1-3)Galβ(1-3)Galβ(1-4)Xylβ1-*O*-Ser,[127] although the GlcA may be replaced by IdoA in DS. This tetrasaccharide may be isolated, following chondroitin ACII lyase digestion, by SEC to separate it from repeat region disaccharides (Figure 3). Hexasaccharide linkage regions, retaining the first repeat region disaccharide, have been isolated following digestion with conventional chondroitin ABC lyase which is unable to cleave the bond between the disaccharide and the linkage region.[107] Other workers[79] have isolated octasaccharide linkage regions by digestion with chondroitin ABC endolyase. Although this enzyme cannot cleave an isolated tetrasaccharide[121] it is able to cleave an isolated octasaccharide linkage region to yield a di- and a hexasaccharide.[79] The isolation of octasaccharide linkage regions may rely upon steric hindrance as digestion is performed prior to the release of the CS linkage region.[79]

Following digestion and subsequent release by β-elimination the linkage region should be parsed by SEC to allow it to be separated from smaller oligosaccharides. This step is crucial as two octasaccharide and two hexasaccharide linkage regions, along with a disaccharide, all elute within 1 min upon HPAEC (Figure 10, Tables 6 and 7). Alternatively, following digestion and reduction the linkage regions may be released unreduced in 0.5 M LiOH[36] and can then be fluorescently labeled (Figure 11 and Table 7).

The methods used for fingerprinting the linkage regions, labeled or not, are those used for the analysis of the di- tetra- and hexasaccharides derived from the repeat region; a Dionex CarboPac™ PA1 column with UV detection at 232 nm or a AS4A-SC column and detection by fluorescence[37,78,79] (see sections 6.3 and 6.3.1).

6.5 Analysis of Chain Caps

Plaas et al. examined CS chain caps by HPAEC on a Dionex AS4A-SC column following fluorescent labeling of the chondroitin ACII and ABC lyase generated oligosaccharides.[105,106] Capping oligosaccharides are not modified by the action of these eliminases and therefore, unlike other oligosaccharides present, do not possess a strong absorbance at 232 nm. The 2-aminopyridine labeled oligosaccharides were examined by HPAEC on an AS4A-SC column eluted with a gradient

of 0–500 mM sodium trifluroacetate. The major caps, GalNAc4,6S and GalNAc4S, were found to occur as the monosaccharides and were eluted at unique positions. However, the more minor caps, GlcAβ(1-3)GalNAc4S and GlcAβ(1-3)GalNAc6S, co-eluted with Δdi-4S and Δdi-6S. Prior treatment with mercuric acetate[21] to remove the terminal Δ4,5-unsaturated uronic acids from these repeat region disaccharides, yielding two monosaccharides, leaves the saturated caps in their native state, allowing their subsequent chromatographic quantification.

7 HEPARAN SULFATE/HEPARIN

Of all the GAGs heparin and HS have the greatest structural complexity and they have received attention in several comprehensive reviews.[128–130] They have been extensively studied because of the anticoagulant functions of heparin and the interactions of HS with various growth factors. They comprise a repeating disaccharide $-4)$GlcNAcα(1-4)uronic acid$\alpha(1-$ in which the uronic acid may be GlcA or IdoA with the possibility of IdoA 2-sulfation and GlcNAc *N*- and 6-sulfation, although there is little GlcA 2-sulfation or GlcNAc 4-sulfation. HS consists primarily of unsulfated GlcA residues and GlcNAc 6- or *N*-sulfation. Heparin is structurally similar to HS but is more highly sulfated, the trisulfated disaccharide $-4)$GlcNS6Sα(1-4)IdoA2S$\alpha(1-$ often accounting for 80–90% of the chain, depending on tissue source. They are both linear GAGs, without fucosylation or other branching. In CS there is evidence that not all potential combinations of structure are found in vivo, but there is no such evidence for HS/heparin in which a wide variety of structures have been observed.

7.1 Characterization

7.1.1 Depolymerization of Heparin and Heparan Sulfate

Although chondroitin ACII lyase may be used to generate disaccharides from CS, no single enzyme is capable of this for HS/heparin, and even the use of all three enzymes described next does not always result in complete depolymerization. There is some confusion in the nomenclature used and care must be taken when sourcing these enzymes; full details of specificity and the temperature for maximal activity should be sought from the suppliers.

Heparinase I/Heparitinase III (EC 4.2.2.7) cleaves regions of high sulfation, acting on GlcNS($\pm$6S$\pm$3S)α(1-4)IdoA2S. Heparinase II/Heparitinase II (no EC number assigned) will cleave either GlcA or IdoA containing disaccharides GlcNR($\pm$6S)α(1-4)GlcA/IdoA, but not unsulfated disaccharides. Heparinase III/Heparitinase I (EC 4.2.2.8) shows specificity for regions of low sulfation and will cleave GlcA-containing disaccharides irrespective of the sulfation status of the glucosamine GlcNR($\pm$6S)α(1-4)GlcA.

7.1.2 Disaccharide Analysis

Nitrous acid depolymerization, with or without prior hydrazinolysis, yields mainly disaccharides. Following hydrazinolysis the entire chain may be depolymerized; however, without this treatment only *N*-sulfated regions will be susceptible to nitrous acid treatment. Importantly for these GAGs this method retains GluA/IdoA differentiation although *N*-sulfation is lost. The reverse is true of enzyme depolymerization in which both uronic acids are converted to a ΔUA, absorbing strongly at 232 nm, but *N*-sulfation is retained.

The requirement for all three enzymes to generate disaccharides and their differing stabilities and specificities means that they should be added sequentially to a digest. The following protocol is recommended by Lyon;[128] to a digest at 30 °C add 10 mU mL^{-1} heparinase I and repeat the addition at 1 h; at 3 and 4 h respectively raise the temperature to 37 °C and add 10 mU mL^{-1} of heparinase III and at 6 and 8 h add 10 mU mL^{-1} of heparinase II and digest overnight.

Following the depolymerization of heparin/HS to disaccharides there are several techniques for the analysis of the oligosaccharides, relying either upon the absorbance at 232 nm introduced by enzyme action or that of a label introduced at the reducing terminus. Yoshida et al.[131] used a LiChrosorb® NH2 column in a method also calibrated for CS/DS oligosaccharides, while Turnbull and Gallagher[132] report a strong anion exchange method with a 120-min 0.2–2 M or 0–1 M NaCl gradient for heparin or HS respectively. Karamanos et al.[133] reported a rapid, sensitive and accurate HPCE method for the analysis of heparin and HS. The analysis of disaccharides generated by a combination of all three enzymes is done with reversed polarity and 15 mM phosphate buffer, at pH 3.5, allowing the analysis of all 12 disaccharides by absorbance at 232 nm with detection of femtogram amounts and at the attomole level.

Karamanos et al.[134] have also reported an ion-pair HPLC method for the analysis of heparin and HS. The analysis of disaccharides generated by a combination of all three enzymes is performed with tetrabutylammonium as the ion pairing agent on a Supelcosil™ L18 column in an acetonitrile gradient with detection at 232 nm. This method allows the analysis of ca. 2–5 ng of a single disaccharide.

7.1.3 Analysis of Longer Oligosaccharides

While the lack of a single enzyme to generate heparin/HS disaccharides makes disaccharide analysis more complex, the different specificities of the three enzymes allow the isolation of oligosaccharides, hexasaccharide and longer, of defined disaccharide composition which varies with the enzyme used. Following the digestion of heparin/HS to generate larger oligosaccharides clearly a variety of these will be generated which, like CS/DS oligosaccharides, must be subjected to SEC to separate them into differing sizes. These can then be further examined or the individual oligosaccharides isolated.

Gradient PAGE has been widely used for the analysis of the structure of heparin and HS. The actual resolving gel gradient will depend upon the nature of the oligosaccharides which are under investigation. A 20–30% gradient with a 5% stacking gel will, however, often be used for general purpose analysis; the gradient may extend up to 40% for large oligosaccharides.[71,135] Samples should be loaded in 50% sucrose with bromophenol blue and phenol red as markers. Such gels provide greater resolution with oligosaccharides larger than tetrasaccharide.

Vives et al.[136] have reported an HPLC method for sequence analysis of heparin and HS oligosaccharides. Radiolabeled (^{3}H) oligosaccharides derived by partial depolymerization were isolated and subjected to specific chemical and enzymatic depolymerization procedures. The oligosaccharides generated by these further treatments were identified by HPLC and the structure of the parent oligosaccharide deduced. In principle this method can be used to sequence any of the GAGs.

8 SUPPLIERS

General laboratory reagents along with chromatographic and other materials are widely available. However, specialist carbohydrate materials, including enzymes and GAGs, may be sourced from several companies. A comprehensive list is available at the Glycoscience Network web site (http://www.vei.co.uk/tgn/).

9 INTERNET GLYCOSCIENCE RESOURCES

There is an increasing number of glycoscience resources accessible using the Internet. One is the Bionet.Glycoscience newsgroup (http://tantalum-e2.hgmp.mrc.ac.uk/hypermail/Glycosci/). Any newsreader will allow you to access this resource and many popular web browsers have a newsreader built in. The Glycoscience Network (http://www.vei.co.uk/tgn/) is a gateway to the most useful glycoscience resources available.

ACKNOWLEDGMENTS

The Arthritis Research Campaign is thanked for support and Tom Huckerby, Ian Nieduszynski, Gavin Brown and Haydn Morris are thanked for helpful comments and critical reading of the manuscript.

ABBREVIATIONS AND ACRONYMS

ANTS	8-Aminonaphthalene-1,3,6-Trisulfonic Acid Disodium Salt
BAP	Biotinylated Aminopyridine
BSA	Bovine Serum Albumin
CCD	Charge-coupled Device
CHAPS	3-([3-Cholamidopropyl]dimethyl-ammonio)-1-propanesulfonate
CPC	Cetyl Pyridinium Chloride
CS	Chondroitin Sulfate
CZE	Capillary Zone Electrophoresis
dH_2O	Distilled Water
DMB	1,9-Dimethylmethylene Blue
DMSO	Dimethyl Sulfoxide
dpm	disintegrations per minute
DS	Dermatan Sulfate
EDTA	Ethylenediamine Tetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Assay
FACE	Fluorochrome-assisted Carbohydrate Electrophoresis
GAG	Glycosaminoglycan
GuHCl	Guanidine Hydrochloride
HA	Hyaluronan
HIC	Hydrophobic Interaction Chromatography
HPAEC	High pH Anion-exchange Chromatography
HPLC	High-performance Liquid Chromatography
HS	Heparan Sulfate
IdoA	Iduronic Acid
IEC	Ion-exchange Chromatography
KS	Keratan Sulfate
MS	Mass Spectroscopy
NeuAc	*N*-Acetylneuraminic Acid
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PAD	Pulsed Amperometric Detection
PAGE	Polyacrylamide Gel Electrophoresis

PBS	Phosphate-buffered Saline
PG	Proteoglycan
PITC	Phenylisothiocyanate
PPO	2,5-Diphenyloxazole
PVDF	Poly(vinylidene difluoride)
RGB	Red–Green–Blue
RPC	Reversed-phase Chromatography
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate/ Polyacrylamide Gel Electrophoresis
SEC	Size-exclusion Chromatography
SNA	*Sambucus nigra* Agglutinin
TFA	Trifluoroacetic Acid
TFMS	Trifluoromethanesulfonic Acid
TGF-β	Transforming Growth Factor Beta
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
2-AA	2-Aminobenzoic Acid
2-D	Two-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis • Glycoprotein Analysis: General Methods • Glycoprotein Analysis: Using Nuclear Magnetic Resonance • Monosaccharides and Sugar Alcohol Analysis

Peptides and Proteins **(Volume 7)**
Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

REFERENCES

1. M.A. Bourden, A. Oldberg, M. Pierschbacher, E. Ruoslahti, 'Molecular Cloning and Sequence Analysis of a Chondroitin Sulfate Proteoglycan cDNA', *Proc. Natl. Acad. Sci.*, **82**, 1321–1325 (1985).
2. M. Paulsson, P.D. Yurchenco, G.C. Ruben, J. Engel, R. Timpl, 'Structure of Low-density Heparan Sulphate Proteoglycan Isolated from a Mouse-tumor Basement-membrane', *J. Mol. Biol.*, **197**, 297–313 (1987).
3. R.V. Iozzo, 'Matrix Proteoglycans: From Molecular Design to Cellular Function', *Annu. Rev. Biochem.*, **67**, 609–652 (1998).
4. A.M. Kidwai, R.K. Upreti, 'A Novel Plant Membrane Proteoglycan Which Causes Anorexia in Animals', *Mol. Cell. Biochem.*, **120**, 111–117 (1993).
5. N.K. Karamanos, A.J. Aletras, C. A. Antonopoulos, A. Hjerpe, 'Determination of the HNK-l Epitope (3-Sulphated Glucuronic-acid) in Intact Chondroitin Sulphates by ELISA – Application to Squid Skin Proteoglycans and their Oversulphated Carbohydrate Structures', *Biochimie*, **76**, 79–82 (1994).
6. B. Kobe, J. Deisenhofer, 'The Leucine-rich Repeat – a Versatile Binding Motif', *Trends Biochem. Sci.*, **19**, 415–421 (1994).
7. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'Structure of the Keratan Sulphate Chains Attached to Fibromodulin Isolated from Bovine Tracheal Cartilage – Oligosaccharides Generated by Keratanase Digestion', *Biochem. J.*, **302**, 417–423 (1994).
8. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'The Structure of the Keratan Sulphate Chains Attached to Fibromodulin Isolated from Bovine Tracheal Cartilage – Oligosaccharides Generated by Keratanase-II Digestion', *Glycoconj. J.*, **12**, 651–659 (1995).
9. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'The Structure of the Keratan Sulphate Chains Attached to Fibromodulin Isolated from Articular Cartilage', *Eur. J. Biochem.*, **242**, 402–409 (1996).
10. A.H.K. Plaas, P.J. Neame, C.M. Nivens, L. Reiss, 'Identification of the Keratan Sulphate Attachment Sites on Bovine Fibromodulin', *J. Biol. Chem.*, **265**, 20 634–20 640 (1990).
11. E. Hedbom, D. Heinegård, 'Interaction of a 59-kDa Connective-tissue Matrix Protein with Collagen-I and Collagen-II', *J. Biol. Chem.*, **264**, 6898–6905 (1989).
12. K. Noyori, H.E. Jasin, 'Inhibition of Human Fibroblast Adhesion by Cartilage Surface Proteoglycans', *Arth. Rheum.*, **37**, 1656–1663 (1994).
13. J.E. Scott, D.A.D. Parry, 'Control of Collagen Fibril Diameters in Tissue', *Int. J. Biol. Macromol.*, **14**, 292–293 (1992).
14. M. Yanagishita, V.C. Hascall, 'Proteoglycans Synthesized by Rat Ovarian Granulosa-cells in Culture – Isolation, Fractionation, and Characterization of Proteoglycans Associated with the Cell Layer', *J. Biol. Chem.*, **259**, 260–269 (1984).
15. V.C. Hascall, J.H. Kimura, in *Methods in Enzymology*, eds. L.W. Cunningham, D.W. Frederiksen, Academic Press, New York, 769, Vol. 82, 1982.

16. M.T. Bayliss, M.F. Venn, in *Studies in Joint Disease*, ed. A. Maroudas, et al., Pitman, Tunbridge Wells, 2–58, Vol. 1, 1980.
17. S.L. Carney, 'Proteoglycans', in *Carbohydrate Analysis: A Practical Approach*, eds. M.F. Chaplin, J.F. Kennedy, IRL Press, Oxford, 97–141, 1986.
18. H.U. Choi, T.L. Johnson, S. Pal, L.H. Tang, L. Rosenberg, P.J. Neame, 'Characterization of the Dermatan Sulphate Proteoglycans, DS-PGI and DS-PGII, from Bovine Articular Cartilage and Skin Isolated by Octyl-Sepharose® Chromatography', *J. Biol. Chem.*, **264**, 2876–2884 (1989).
19. C.A. McDevitt, H. Muir, 'Gel Electrophoresis of Proteoglycans and Glycosaminoglycans on Large-pore Composite Polyacrylamide–Agarose Gels', *Anal. Biochem.*, **44**, 612–622 (1971).
20. T.N. Huckerby, R.M. Lauder, I.A. Nieduszynski, 'Structure Determination for Octa-saccharides Derived from the Carbohydrate–Protein Linkage Region of Chondroitin Sulphate Chains in the Proteoglycan Aggrecan from Bovine Articular Cartilage', *Eur. J. Biochem.*, **258**, 669–676 (1998).
21. U. Ludwigs, A. Elgavish, J.D. Esko, E. Meezan, L. Roden, 'Reaction of Unsaturated Uronic Acid Residues with Mercuric Salts. Cleavage of the Hyaluronic Acid Disaccharide 2-Acetamido-2-deoxy-3-*O*-(beta-D-gluco-4-enepyranosyluronic acid)-D-glucose', *Biochem. J.*, **245**, 795–804 (1987).
22. A.J. Mort, D.T. Lamport, 'Anhydrous Hydrogen Fluoride Deglycosylates Glycoproteins', *Anal. Biochem.*, **82**, 289–309 (1977).
23. C. Coudron, K. Ellis, L. Philipson, N. Swartz, 'Preliminary Characterization of a Xylose Acceptor Prepared by Hydrogen Fluoride Treatment of Proteoglycan Core Protein', *Biochem. Biophys. Res. Commun.*, **92**, 618–623 (1980).
24. J.N. Glassman, C.W. Todd, J.E. Shively, 'Chemical Deglycosylation of Carcinoembryonic Antigen for Amino Acid Sequence Studies', *Biochem. Biophys. Res. Commun.*, **85**, 209–216 (1978).
25. T.A. Gerken, R. Gupta, N. Jentoft, 'A Novel-approach for Chemically Deglycosylating O-Linked Glycoproteins – the Deglycosylation of Submaxillary and Respiratory Mucins', *Biochemistry*, **31**, 639–648 (1992).
26. A.S. Edge, C.R. Faltynek, L. Hof, L.E. Reichert, P. Weber, 'Deglycosylation of Glycoproteins by Trifluoromethanesulphonic Acid', *Anal. Biochem.*, **118**, 131–137 (1981).
27. H. Watanabe, S.C. Cheung, N. Itano, K. Kimata, Y. Yamada, 'Identification of Hyaluronan-binding Domains of Aggrecan', *J. Biol. Chem.*, **272**, 28 057–28 065 (1997).
28. M. Fried, P.E. Duffy, 'Maternal Malaria and Parasite Adhesion', *J. Mol. Med.*, **76**, 162–171 (1998).
29. M. Fried, R.M. Lauder, P.E. Duffy, 'Minimal Chondroitin Sulphate Structures Required to Support *Plasmodium falciparum* Adhesion', *Glycobiology*, **8**, 1021 (1998).
30. M. Belting, L.A. Fransson, 'The Growth Promoter Spermine Interacts Specifically with Dermatan Sulphate Which is Rich in Iduronic Acid', in *Dermatan Sulphate Proteoglycans*, ed. J. Scott, Portland Press, London, 259–266, 1993.
31. R.D. Sanderson, J.E. Turnbull, J.T. Gallagher, A.D. Lander, 'Fine Structure of Heparan Sulphate Regulates Syndecan-1 Function and Cell Behavior', *J. Biol. Chem.*, **269**, 13 100–13 106 (1994).
32. U. Lindahl, G. Backstrom, M. Hook, L. Thunberg, L.A. Fransson, A. Linker, 'Structure of the Antithrombin-binding Site in Heparin', *Proc. Natl. Acad. Sci.*, **76**, 3198–3202 (1979).
33. M.M. Maimone, D.M. Tollefsen, 'Structure of a Dermatan Sulphate Hexasaccharide that Binds to Heparin Cofactor II with High Affinity', *J. Biol. Chem.*, **265**, 18 263–18 271 (1990).
34. J.E. Turnbull, D.G. Fernig, Y. Ke, M.C. Wilkinson, J.T. Gallagher, 'Identification of the Basic Fibroblast Growth Factor Binding Sequence in Fibroblast Heparan Sulphate', *J. Biol. Chem.*, **267**, 10 337–10 341 (1992).
35. D.M. Carlsson, 'Structures and Immunological Properties of Oligosaccharides Isolated from Pig Submaxillary Mucin', *J. Biol. Chem.*, **243**, 616–620 (1968).
36. D. Heinegard, 'Extraction, Fractionation and Characterization of Proteoglycans from Bovine Tracheal Cartilage', *Biochim. Biophys. Acta*, **285**, 181–192 (1972).
37. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'A Single Method for the Complete Analysis of Chondroitin Sulphates by High pH Anion Exchange Chromatography with Fluorescent Detection', *Glycobiology*, in press (2000).
38. F. Cheng, K. Yoshida, D. Heinegard, L.A. Fransson, 'A New Method for Sequence Analysis of Glycosaminoglycans from Heavily Substituted Proteoglycans Reveals Nonrandom Positioning of 4-*O*-Sulphated and 6-*O*-Sulphated *N*-Acetylgalactosamine in Aggrecan-derived Chondroitin Sulphate', *Glycobiology*, **2**, 553–561 (1992).
39. C.Y. Lee, J.R. Scocca, 'A Common Structural Unit in Asparagine-oligosaccharides of Several Glycoproteins from Different Sources', *J. Biol. Chem.*, **247**, 5753–5759 (1972).
40. T. Patel, J. Bruce, A. Merry, C. Bigge, M. Wormald, A. Jaques, R. Parekh, 'Use of Hydrazine to Release in Intact and Unreduced Form Both *N*- and *O*-Linked Oligosaccharides from Glycoproteins', *Biochemistry*, **32**, 679–693 (1993).
41. J. Charlwood, H. Birrell, P. Camilleri, 'Efficient Carbohydrate Release, Purification, and Derivatization', *Anal. Biochem.*, **262**, 197–200 (1998).
42. N. Volpi, 'Purification of Heparin, Dermatan Sulphate and Chondroitin Sulphate from Mixtures by Sequential

Precipitation with Various Organic Solvents', *J. Chrom., B Biomed. App.*, **11**, 27–34 (1996).
43. S. Bjornsson, 'Quantitation of Proteoglycans as Glycosaminoglycans in Biological Fluids Using an Alcian Blue Dot Blot Analysis', *Anal. Biochem.*, **256**, 229–237 (1998).
44. E.W. Gold, 'A Simple Spectrophotometric Method for Estimating Glycosaminoglycan Concentrations', *Anal. Biochem.*, **99**, 183–188 (1979).
45. P. Whiteman, 'The Quantitative Measurement of Alcian Blue–Glycosaminoglycan Complexes', *Biochem. J.*, **131**, 343–350 (1973).
46. J.E. Scott, J. Dorling, 'Differential Staining of Acid Glycosaminoglycans (Mucopolysaccharides) by Alcian Blue in Salt Solutions', *Histochemie*, **5**, 221–233 (1965).
47. M. Lammi, M. Tammi, 'Densitometric Assay of Nanogram Quantities of Proteoglycans Precipitated on Nitrocellulose Membrane with Safranin-O', *Anal. Biochem.*, **168**, 352–357 (1988).
48. R.W. Farndale, C.A. Sayers, A.J. Barrett, 'A Direct Spectrophotometric Micro-assay for Sulphated Glycosaminoglycans in Cartilage Cultures', *Connective Tissue Res.*, **9**, 247–248 (1982).
49. G.H. Tai, H.G. Morris, G.M. Brown, T.N. Huckerby, I.A. Nieduszynski, 'A Subpopulation of Keratan Sulphates Derived from Bovine Articular-cartilage is Capped with Alpha(2-6)-linked *N*-Acetylneuraminic Acid Residues – Affinity-chromatography Using Immobilized *Sambucus-Nigra* Lectin and Characterization Using H-NMR Spectroscopy', *Biochem. J.*, **286**, 231–234 (1992).
50. M.F. Chaplin, 'Monosaccharides', in *Carbohydrate Analysis. A Practical Approach*, eds. M.F Chaplin, J.F. Kennedy, IRL Press, Oxford, 1986.
51. M. Dubois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, 'Colorimetric Method for Determination of Sugars and Related Substances', *Anal. Chem.*, **28**, 350–356 (1956).
52. C.A. White, J.F. Kennedy, 'Oligosaccharides', in *Carbohydrate Analysis. A Practical Approach*, eds. M.F. Chaplin, J.F. Kennedy, IRL Press, Oxford, 1986.
53. L. Warren, 'The Thiobarbituric Acid Assay of Sialic Acids', *J. Biol. Chem.*, **234**, 1971–1975 (1959).
54. G.H. Tai, T.N. Huckerby, I.A. Nieduszynski, 'Multiple Non-reducing Chain Termini Isolated from Bovine Corneal Keratan Sulphates', *J. Biol. Chem.*, **271**, 23 535–23 546 (1996).
55. T. Bitter, H. Muir, 'A Modified Uronic Acid Carbazole Reaction', *Anal. Biochem.*, **4**, 330–334 (1962).
56. W.E. Trevelyan, J.E. Harrison, 'Studies on Yeast Metabolism I. Fractionation and Microdetermination of Cell Carbohydrates', *Biochem. J.*, **51**, 298–303 (1952).
57. L.A. Elson, W.T.J. Morgan, 'A Colorimetric Method for the Determination of Glucosamine and Chondrosamine', *Biochem. J.*, **23**, 1824–1828 (1933).
58. J.L. Reissig, J.L. Strominger, L.F. Leloir, 'A Modified Colorimetric Method for the Estimation of Acetylamino Sugars', *J. Biol. Chem.*, **217**, 959–966 (1955).
59. K.S. Dodgson, 'Determination of Inorganic Sulphate in Studies on the Enzymic and Non-enzymic Hydrolysis of Carbohydrate and Other Sulphate Esters', *Biochem. J.*, **78**, 312–319 (1961).
60. W.F. Broekaert, M. Nsimba-Lubaki, B. Peeters, W.J. Peumans, 'A Lectin from Elder (*Sambucus Nigra* L.) Bark', *Biochem. J.*, **221**, 163–169 (1984).
61. N. Shibuya, I.J. Goldstein, W.F. Broekart, M. Nsimba-Lubaki, B. Peeters, W.J. Peumans, 'The Elderberry (*Sambucus Nigra* L.) Bark Lectin Recognizes the Neu5Ac(α-2-6)GalGalNac Sequence', *J. Biol. Chem.*, **262**, 1596–1601 (1987).
62. N. Shibuya, I.J. Goldstein, W.F. Broekart, M. Nsimba-Lubaki, B. Peeters, W.J. Peumans, 'Fractionation of Sialylated Oligosaccharides, Glycopeptides, and Glycoproteins on Immobilized Elderberry (*Sambucus-nigra* L.) Bark Lectin', *Arch. Biochem. Biophys.*, **254**, 1–8 (1987).
63. R. Hata, T. Nagai, 'A Rapid and Micromethod for Separation of Acidic Glycosaminoglycans by Two-dimensional Electrophoresis', *Anal. Biochem.*, **45**, 462–468 (1972).
64. N. Volpi, 'Fast Moving and Slow Moving Heparins, Dermatan Sulphate, and Chondroitin Sulphate: Qualitative and Quantitative Analysis by Agarose-gel Electrophoresis', *Carbohydr. Res.*, **247**, 263–278 (1993).
65. N. Volpi, 'Electrophoresis Separation of Glycosaminoglycans on Nitrocellulose Membranes', *Anal. Biochem.*, **240**, 114–118 (1996).
66. J.J. Hopwood, H. Elliott, 'Selective Depolymerization of Keratan Sulphate–Production of Radiolabeled Substrates for 6-*O*-Sulphogalactose Sulphatase and Beta-D-Galactosidase', *Carbohydr. Res.*, **117**, 263–274 (1983).
67. Y. Guo, H.E. Conrad, 'The Disaccharide Composition of Heparins and Heparan Sulphates', *Anal. Biochem.*, **176**, 96–104 (1989).
68. J.C. Bigge, T.P. Patel, J.A. Bruce, P.N. Goulding, S.M. Charles, R.B. Parekh, 'Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-Amino Benzamide and Anthranilic Acid', *Anal. Biochem.*, **230**, 229–238 (1995).
69. B. Kuster, D.J. Harvey, 'Ammonium Containing Buffers Should Be Avoided During Enzymatic Release of Glycans from Glycoproteins When Followed by Reducing Terminal Derivatization', *Glycobiology*, **7**, vii–ix (1997).
70. D.K. Toomre, A. Varki, 'Simultaneous Fluorescent Labeling and Biotinylation of Oligosaccharides', in *A Laboratory Guide to Glycoconjugate Analysis*, eds. P. Jackson, J. Gallagher, Birkhausser, Basel, 1998.
71. J.E. Turnbull, J. Gallagher, 'Molecular Organization of Heparan Sulphate from Human Skin Fibroblasts', *Biochem. J.*, **265**, 715–724 (1990).

72. K. Myer, J.W. Palmer, 'The Polysaccharide of Vitreous Humor', *J. Biol. Chem.*, **107**, 629–634 (1934).
73. G.W. Jourdian, M. Wolfman, R. Sarber, J. Distler, 'A Specific, Sensitive Method for the Determination of Hyaluronate', *Anal. Biochem.*, **96**, 474–480 (1979).
74. K.N. Price, A. Tuinman, D.C. Baker, C. Chisena, R.L. Cysyk, 'Isolation and Characterization by Electrospray-ionization Mass Spectrometry and High-performance Anion-exchange Chromatography of Oligosaccharides Derived from Hyaluronic Acid by Hyaluronate Lyase Digestion. Observation of Some Heretofore Unobserved Oligosaccharides that Contain an Odd-number of Units', *Carbohydr. Res.*, **303**, 303–311 (1997).
75. R.J. Midura, A. Salustri, A. Calabro, M. Yanagishita, V.C. Hascall, 'High-resolution Separation of Disaccharide and Oligosaccharide-alditols from Chondroitin Sulphate, Dermatan Sulphate and Hyaluronan Using CarboPac™ PA1 Chromatography', *Glycobiology*, **4**, 333–342 (1994).
76. N.K. Karamanos, A. Syrokou, P. Vanky, M. Nurrinen, A. Hjerpe, 'Determination of 24 Variously Sulphated Galacosaminoglycan-derived and Hyaluronan-derived Disaccharides by High-performance Liquid Chromatography', *Anal. Biochem.*, **221**, 189–199 (1994).
77. N.K. Karamanos, S. Axelsson, P. Vanky, G.N. Ttanakakis, A. Hjerpe, 'Determination of Hyaluronan and Galactosaminoglycan Disaccharides by High-performance Capillary Electrophoresis at the Attomole Level–Applications to Analyses of Tissue and Cell-culture Proteoglycans', *J. Chrom. A.*, **696**, 295–305 (1995).
78. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'A High pH Anion Exchange Chromatography Fingerprinting Method for Chondroitin/Dermatan Sulphate and Hyaluronan Oligosaccharides', *Glycobiology*, **10**, 393–401 (2000).
79. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'Increased Incidence of Unsulphated and 4-Sulphated Residues in the Chondroitin Sulphate Linkage Region Observed by High pH Anion Exchange Chromatography', *Biochem. J.*, **347**, 339–348 (1998).
80. R. Toffanin, B.J. Kvam, A. Flaibani, M. Atzori, F. Bivi-ano, S. Paoletti, 'NMR-studies of Oligosaccharides Derived from Hyaluronate – Complete Assignment of H-1 and C-13 NMR-spectra of Aqueous Disaccharides and Tetra-saccharides, and Comparison of Chemical-shifts for Oligosaccharides of Increasing Degree of Polymerization', *Carbohydr. Res.*, **245**, 113–128 (1993).
81. V.P. Bhavanandan, K. Meyer, 'Studies on Keratosulphates. Methylation, Desulphation, and Acid Hydrolysis Studies on Old Human Rib Cartilage Keratosulphate', *J. Biol. Chem.*, **243**, 1052–1059 (1968).
82. B.A. Bray, R. Lieberman, K. Meyer, 'Structure of Human Skeletal Keratosulphate. The Linkage Region', *J. Biol. Chem.*, **242**, 3373–3380 (1967).
83. T. Krusius, J. Finne, R.K. Margolis, R.U. Margolis, 'Identification of an *O*-Glycosidic Mannose-linked Sialylated Tetrasaccharide and Keratan Sulphate Oligosaccharides in the Chondroitin Sulphate Proteoglycan of Brain', *J. Biol. Chem.*, **261**, 8237–8242 (1986).
84. I.A. Nieduszynski, T.N. Huckerby, J.M. Dickenson, G.M. Brown, G.H. Tai, H.G. Morris, S. Eady, 'There are 2 Major Types of Skeletal Keratan Sulphates', *Biochem. J.*, **271**, 243–245 (1990).
85. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, A.H.K. Plaas, 'Age-related Changes in the Structure of the Keratan Sulphate Chains Attached to Fibromodulin Isolated from Articular Cartilage', *Biochem. J.*, **330**, 753–757 (1998).
86. G.M. Brown, T.N. Huckerby, I.A. Nieduszynski, 'Human Aggrecan Keratan Sulphate Undergoes Structural Changes During Adolescent Development', *J. Biol. Chem.*, **273**, 26 408–26 414 (1998).
87. G.M. Brown, I.A. Nieduszynski, 'Oligosaccharide Profiling of Keratan Sulphate', in *A Laboratory Guide to Glycoconjugate Analysis*, eds. P. Jackson, J. Gallagher, Birkhausser, Basel, 1998.
88. J.M. Dickenson, T.N. Huckerby, I.A. Nieduszynski, 'A Non-reducing Terminal Fragment from Tracheal Cartilage Keratan Sulphate Chains Contains Alpha (2-3)-linked *N*-Acetylneuraminic Acid', *Biochem. J.*, **278**, 779–785 (1991).
89. H. Mehmet, P. Scudder, P.W. Tang, E.F. Hounsell, B. Caterson, T. Feizi, 'The Antigenic Determinants Recognized by Three Monoclonal Antibodies to Keratan Sulphate Involve Sulphated Hepta- or Larger Oligosaccharides of the Poly(*N*-acetyllactosamine) Series', *Eur. J. Biochem.*, **157**, 385–391 (1986).
90. M. Kitamikado, R. Ueno, T. Nakamura, 'Enzymatic Degradation of Whale Cartilage Keratosulphate. I. Isolation of a Keratosulphate-degrading Bacterium', *Bull. Jpn. Soc. Sci. Fish*, **36**, 592–596 (1970).
91. M. Kitamikado, M. Ito, Y.T. Li, 'Isolation and Characterization of a Keratan Sulphate-degrading *endo*-beta-Galactosidase from *Flavobacterium keratolyticus*', *J. Biol. Chem.*, **256**, 3906–3909 (1981).
92. P. Scudder, K. Uemura, J. Dolby, M.N. Fukuda, T. Feizi, 'Isolation and Characterization of an *endo*-beta-Galactosidase from *Bacteroides fragilis*', *Biochem. J.*, **213**, 485–494 (1983).
93. K. Nakazawa, M. Ito, T. Yamagata, S. Suzuki, 'Substrate Specificity of Keratan Sulphate-degrading Enzymes from Microorganisms', in *Keratan Sulphate*, eds. H. Greiling, J. Scott, The Biochemical Society, Portland Press, London, 1989.
94. K. Nakazawa, S. Suzuki, 'Purification of Keratan Sulphate-endogalactosidase and its Action on Keratan Sulphates of Different Origin', *J. Biol. Chem.*, **250**, 912–917 (1975).
95. G.M. Brown, T.N. Huckerby, H.G. Morris, I.A. Nieduszynski, 'Degradation of Articular Cartilage Keratan

Sulphates Using Hydrazinolysis and Nitrous Acid. Environment of Fucose Residues', *Biochem. J.*, **286**, 235–241 (1992).
96. K.M. Whitham, J.L. Hadley, H.G. Morris, S.M. Andrew, I.A. Nieduszynski, G.M. Brown, 'An Improved Method for the Structural Profiling of Keratan Sulphates; Analysis of Keratan Sulphates from Brain and Ovarian Tumours', *Glycobiology*, **9**, 285–291 (1999).
97. K. Sugahara, Y. Tanaka, S. Yamada, N. Seno, H. Kitagawa, S.M. Haslam, H.R. Morris, A. Dell, 'Novel Sulphated Oligosaccharides Containing 3-*O*-Sulphated Glucuronic- acid from King Crab Cartilage Chondroitin Sulphate-K–Unexpected Degradation by Chondroitinase ABC', *J. Biol. Chem.*, **271**, 26 745–26 754 (1996).
98. P.A.S. Mourao, M.S. Pereira, M.S.G. Pavao, B. Mulloy, D.M. Tollefsen, M.C. Mowinckel, U. Abildgaard, 'Structure and Anticoagulant Activity of a Fucosylated Chondroitin Sulphate from Echinoderm–Sulphated Fucose Branches on the Polysaccharide Account for its High Anticoagulant Action', *J. Biol. Chem.*, **271**, 23 973–23 984 (1996).
99. M.P. Mark, J.R. Baker, K. Kimata, J. Roth, 'Regulated Changes in Chondroitin Sulphation During Embryogenesis: An Immunohistochemical Approach', *Int. J. Dev. Biol.*, **34**, 191–204 (1990).
100. M.E. Herndon, A.D. Lander, 'A Diverse Set of Developmentally Regulated Proteoglycans is Expressed in the Rat Central-nervous-system', *Neuron*, **4**, 949–961 (1990).
101. B.M. Cooke, S.J. Rogerson, G.V. Brown, R.L. Coppel, 'Adhesion of Malaria-infected Red-blood-cells to Chondroitin Sulphate-A Under Flow Conditions', *Blood*, **88**, 4040–4044 (1996).
102. R.U. Margolis, R.K. Margolis, 'Chondroitin Sulphate Proteoglycans as Mediators of Axon Growth and Pathfinding', *Cell Tissue Res.*, **290**, 343–348 (1997).
103. M.N. Pangalos, J. Shioi, S. Efthimiopoulos, A.F. Wu, N.K. Robakis, 'Characterization of Appican, the Chondroitin Sulphate Proteoglycan Form of the Alzheimer Amyloid Precursor Protein', *Neurodegeneration*, **5**, 445–451 (1996).
104. R.J. Midura, A. Calabro, M. Yanagishita, V.C. Hascall, 'Nonreducing End Structures of Chondroitin Sulphate Chains on Aggrecan Isolated from Swarm Rat Chondrosarcoma Cultures', *J. Biol. Chem.*, **270**, 8009–8015 (1995).
105. A.H.K. Plaas, S. Wong-Palms, P.J. Roughley, R.J. Midura, V.C. Hascall, 'Chemical and Immunological Assay of the Nonreducing Terminal Residues of Chondroitin Sulphate from Human Aggrecan', *J. Biol. Chem.*, **272**, 20 603–20 610 (1997).
106. A.H.K. Plaas, L.A. West, S. Wong-Palms, F.R.T. Nelson, 'Glycosaminoglycan Sulphation in Human Osteoarthritis – Disease-related Alterations at the Non-reducing Termini of Chondroitin and Dermatan Sulphate', *J. Biol. Chem.*, **273**, 12 642–12 649 (1998).
107. P. de Waard, J.F.G. Vliegenthart, T. Harada, K. Sugahara, 'Structural Studies on Sulphated Oligosaccharides Derived from the Carbohydrate–Protein Linkage Region of Chondroitin-6-sulphate Proteoglycans of Shark Cartilage. II. Seven Compounds Containing 2 or 3 Sulphate Residues', *J. Biol. Chem.*, **267**, 6036–6043 (1992).
108. K. Sugahara, Y. Ohi, T. Harada, P. de Waard, J.F.G. Vliegenthart, 'Structural Studies on Sulphated Oligosaccharides Derived from the Carbohydrate–Protein Linkage Region of Chondroitin-6-sulphate Proteoglycans of Shark Cartilage. I. Six Compounds Containing 0 or 1 Sulphate and/or Phosphate Residues', *J. Biol. Chem.*, **276**, 6027–6025 (1992).
109. T. de Beer, A. Inui, H. Tsuda, K. Sugahara, J.F.G. Vliegenthart, 'Polydispersity in Sulphation Profile of Oligosaccharide-alditols Isolated from the Protein-linkage Region and the Repeating Disaccharide Region of Chondroitin-4-sulphate of Bovine Nasal Septal Cartilage', *Eur. J. Biochem.*, **240**, 789–797 (1996).
110. F. Cheng, D. Heinegård, L.A. Fransson, M. Bayliss, J. Bielicki, J. Hopwood, K. Yoshida, 'Variations in the Chondroitin Sulphate–Protein Linkage Region of Aggrecans from Bovine Nasal and Human Articular Cartilages', *J. Biol. Chem.*, **271**, 28 572–28 580 (1996).
111. K. Sugahara, I. Yamashina, P. de Waard, H. Van Halbeek, J.F.G. Vliegenthart, 'Structural Studies on Sulphated Glycopeptides from the Carbohydrate–Protein Linkage Region of Chondroitin 4-sulphate Proteoglycans of Swarm Rat Chondrosarcoma. Demonstration of the Structure Gal(4-*O*-sulphate)β1-3Galβ1-4Xylβ1-*O*-Ser', *J. Biol. Chem.*, **263**, 10 168–10 174 (1988).
112. J. Moses, A. Oldberg, E. Eklund, L.A. Fransson, 'Biosynthesis of the Proteoglycan Decorin–Identification of Intermediates in Galactosaminoglycan Assembly', *Eur. J. Biochem.*, **248**, 767–774 (1997).
113. S. Shibata, R.J. Midura, V.C. Hascall, 'Structural Analysis of the Linkage Region Oligosaccharides and Unsaturated Disaccharides from Chondroitin Sulphate Using CarboPac™ PA1', *J. Biol. Chem.*, **267**, 6548–6555 (1992).
114. K. Sugahara, Y. Ohkita, S. Shibata, K. Yoshida, A. Ikegami, 'Structural Studies on the Hexasaccharide Alditols Isolated from the Carbohydrate–Protein Linkage Region of Dermatan Sulphate Proteoglycans of Bovine Aorta–Demonstration of Iduronic Acid-containing Components', *J. Biol. Chem.*, **270**, 7204–7212 (1995).
115. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'Highly Sulphated Chondroitin Sulphate is Resistant to Enzyme Digestion', *FEBS Lett.*, submitted (2000).
116. S. Nadanaka, K. Sugahara, 'The Unusual Tetrasaccharide Sequence GlcAbeta1-3GalNAc(4-Sulphate)beta1-4GlcA(2-Sulphate)beta1-3GalNAc(6-Sulphate) Found in the Hexasaccharides Prepared by Testicular Hyaluronidase Digestion of Shark Cartilage Chondroitin Sulphate D', *Glycobiology*, **7**, 253–263 (1997).
117. W.G. Chai, A.M. Lawson, M.J. Gradwell, H. Kogelberg, 'Structural Characterization of Two Hexasaccharides

and an Octasaccharide from Chondroitin Sulphate C Containing the Unusual Sequence (4-Sulpho)-*N*-acetylgalactosamine beta 1-4-(2-Sulpho)-Glucuronic Acid beta 1-3-(6-Sulpho)-*N*-acetylgalactosamine', *Eur. J. Biochem.*, **251**, 114–121 (1997).

118. N.K. Karamanos, A. Hjerpe, T. Tsegenidis, B. Engfeldt, C.A. Antonopoulos, 'Determination of Iduronic Acid and Glucuronic-acid in Glycosaminoglycans After Stoichiometric Reduction and Depolymerization Using High-performance Liquid-chromatography and Ultraviolet Detection', *Anal. Biochem.*, **172**, 410–419 (1988).
119. R.M. Lauder, T.N. Huckerby, I.A. Nieduszynski, 'A High pH Anion Exchange Chromatographic Method for the Determination of Iduronic Acids in Chondroitin/Dermatan Sulphates', *Anal. Biochem.*, in press (2000).
120. T. Yamagata, H. Saito, O. Habuchi, S. Suzuki, 'Purification and Properties of Bacterial Chondroitinases and Chondrosulphatases', *J. Biol. Chem.*, **243**, 1523–1535 (1968).
121. A. Hamai, N. Hashimoto, H. Mochizuki, F. Kato, Y. Makiguchi, K. Horie, S. Suzuki, 'Two Distinct Chondroitin Sulphate ABC Lyases – An Endoeliminase Yielding Tetrasaccharides and an Exoeliminase Preferentially Acting on Oligosaccharides', *J. Biol. Chem.*, **272**, 9123–9130 (1997).
122. K. Sugahara, S. Nadanaka, K. Takeda, T. Kojima, 'Structural-analysis of Unsaturated Hexasaccharides Isolated from Shark Cartilage Chondroitin Sulphate-D that are Substrates for the Exolytic Action of Chondroitin ABC Lyase', *Eur. J. Biochem.*, **239**, 871–880 (1996).
123. N.K. Karamanos, P. Vanky, A. Syrokou, A. Hjerpe, 'Identity of Dermatan and Chondroitin Sequences in Dermatan Sulfate Chains Determined by Using Fragmentation with Chondroitinases and Ion-pair High-performance Liquid Chromatography', *Anal. Biochem.*, **225**, 220–230 (1995).
124. H. Kitagawa, A. Kinoshita, K. Sugahara, 'Microanalysis of Glycosaminoglycan-derived Disaccharides Labeled with the Fluorophore 2-Aminoacridone by Capillary Electrophoresis and High-performance Liquid Chromatography', *Anal. Biochem.*, **232**, 114–121 (1995).
125. N.K. Karamanos, A. Hjerpe, 'High Performance Capillary Electrophoretic Analysis of Hyaluronan in Effusions from Human Malignant Mesothelioma', *J. Chrom. B*, **697**, 277–281 (1997).
126. A.H.K. Plaas, V.C. Hascall, R.J. Midura, 'Ion Exchange HPLC Microanalysis of Chondroitin Sulphate: Quantitative Derivitization of Chondroitin Lyase Digestion Products with 2-Aminopyridine', *Glycobiology*, **6**, 823–829 (1996).
127. L. Roden, *The Biochemistry of Glycoproteins and Proteroglycans*, Plenum, New York, 1980.
128. M. Lyon, 'Determination of the Disaccharide Composition of Glycosaminoglycans: Comparison of Chemical and Enzymatic Scission', in *A Laboratory Guide to Glycoconjugate Analysis*, eds. P. Jackson, J. Gallagher, Birkhausser, Basel, 1998.
129. U. Lindahl, M. Kusche-Gullberg, L. Kjellen, 'Regulated Diversity of Heparan Sulphate', *J. Biol. Chem.*, **273**, 24 979–24 982 (1998).
130. U. Lindahl, K. Lidholt, D. Spillmann, L. Kjellen, 'More to ''Heparin'' than Anticoagulation', *Thromb. Res.*, **75**, 1–32 (1994).
131. K. Yoshida, S. Miyauchi, H. Kikuchi, A. Tawada, K. Tokuyasu, 'Analysis of Unsaturated Disaccharides from Glycosaminoglycuronan by High-performance Liquid Chromatography', *Anal. Biochem.*, **177**, 327–332 (1989).
132. J. Turnbull, J.T. Gallagher, 'Molecular Organization of Heparan Sulphate from Human Skin Fibroblasts', *Biochem. J.*, **265**, 715–724 (1990).
133. N.K. Karamanos, P. Vanky, G.N. Tzanakakis, A. Hjerpe, 'A High Performance Capillary Electrophoresis Method to Characterize Heparin and Heparan Sulphate Disaccharides', *Electrophoresis*, **17**, 391–395 (1996).
134. N.K. Karamanos, P. Vanky, G.N. Tzanakakis, T. Tsegenidis, A. Hjerpe, 'Ion-pair High-performance Liquid Chromatography for Determining Disaccharide Composition in Heparin and Heparan Sulphate', *J. Chrom. A*, **765**, 169–179 (1997).
135. J.E. Turnbull, J.T. Gallagher, 'Oligosaccharide Mapping of Heparan Sulphate by Polyacrylamide-gradient-gel Electrophoresis and Electrotransfer to Nylon Membrane', *Biochem. J.*, **251**, 597–608 (1988).
136. R.R. Vives, D.A. Pye, M. Salmivirta, J.J. Hopwood, U. Lindahl, J.T. Gallagher, 'Sequence Analysis of Heparan Sulphate and Heparin Oligosaccharides', *Biochem. J.*, **339**, 767–773 (1999).
137. Y. Sommarin, T. Larsson, D. Heinegard, 'Chondrocyte–Matrix Interactions. Attachment to Proteins Isolated from Cartilage', *Exp. Cell Res.*, 181–192 (1989).

Verification of Chemicals Related to the Chemical Weapons Convention

Markku Mesilaakso and Marjatta Rautio
Finnish Institute for Verification of the Chemical Weapons Convention, Helsinki, Finland

1 Introduction **1**
2 The Chemical Weapons Convention **2**
2.1 History 2
2.2 Obligations 2
2.3 Organisation for the Prohibition of Chemical Weapons 2
2.4 Definition of Chemical Weapons 2
2.5 Schedules of Chemicals 2
3 Detection and Verification of Chemical Weapons **6**
3.1 Recommended Operating Procedures 6
3.2 Organisation for the Prohibition of Chemical Weapons Standard Operating Procedures 7
3.3 Analysis Strategy in Off-site Laboratory 8
3.4 International Interlaboratory Comparison and Proficiency Tests 10
Abbreviations and Acronyms **10**
References **11**

This article gives a short historical review of the Chemical Weapons Convention (CWC), discusses the tasks of the Organisation for the Prohibition of Chemical Weapons (OPCW) related to chemical analysis, and lists the chemicals scheduled in the CWC. The Recommended Operating Procedures (ROPs) proposed originally by Finland and, subsequently, developed further in international cooperation, as well as the Standard Operating Procedures (SOPs) of the OPCW are briefly discussed. Strategy for analysis of CWC-related chemicals in an off-site laboratory is presented. The international interlaboratory comparison (round-robin) and proficiency tests are discussed.

1 INTRODUCTION

The section Chemical Weapons Chemicals Analysis consists of articles on sample preparation and analytical methods that have been used for analysis of chemicals related to the CWC[1] in laboratories worldwide in the international interlaboratory comparison (round-robin) and proficiency tests.[2–8] The analytical techniques are mass spectrometry (MS) (see **Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention; Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention; Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention**), nuclear magnetic resonance (NMR) spectroscopy (see **Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention**), Fourier transform infrared spectroscopy (FTIRS) (see **Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention**), gas chromatography (GC) (see **Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention**), and capillary electrophoresis (CE) (see **Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention**). Sample preparation for on-site and off-site analysis is presented in a separate article (**Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention**). On-site methods for analysis of CWC-related chemicals using FTIRS (**Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention**) and gas chromatography/mass spectrometry (GC/MS) (**Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention**) are described, the former in the article on off-site analysis. Two articles deal with detection and screening of CWC-related chemicals on-site (**Detection and Screening of Chemicals Related to the Chemical Weapons Convention; Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention**), while one also discusses the sampling (**Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention**). Techniques for real time detection of chemical warfare (CW) agents in the battlefield are not discussed.

The methods included in this section provide the best off-site performance for unambiguous identification of CWC-related chemicals. The success of the off-site analysis techniques has been unequivocally confirmed in international proficiency tests. The identification in on-site analysis relies on GC/EIMS (electron ionization mass spectrometry).

This article gives a short historical review of the CWC, discusses the tasks of the OPCW related to chemical analysis, and lists the chemicals scheduled in the CWC. The ROPs proposed originally by Finland and, subsequently,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

developed further in international cooperation, as well as the SOPs of the OPCW are briefly discussed. Strategy for analysis of CWC-related chemicals in an off-site laboratory is presented. The international interlaboratory comparison (round-robin) and proficiency tests are discussed.

2 THE CHEMICAL WEAPONS CONVENTION

2.1 History

After World War I an international convention was signed on June 17, 1925, for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare (the Geneva Protocol of 1925). The Convention entered into force on February 8, 1928. This Convention, which prohibits the use, but not the possession, development, storage and production of chemical weapons, was not respected and was openly violated. However, it prevented the massive use of chemical weapons in World War II. Negotiations for a new, comprehensive convention started in Geneva in 1968. The Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (the CWC) was signed on January 13, 1993, and entered into force on 29 April, 1997.[1]

2.2 Obligations

By October 1999, 126 countries have ratified the Convention and a further 44 countries have signed it. Each State Party to the Convention undertakes never, under any circumstances, to use, develop, produce, otherwise acquire, stockpile, or retain chemical weapons, or transfer, directly or indirectly, chemical weapons to anyone. Further, each State Party undertakes to destroy chemical weapons it owns or possesses and undertakes to submit declarations as required by the CWC. The State Parties also recognize that the new convention reaffirms principles and objectives of and obligations assumed under the Geneva Protocol of 1925.

2.3 Organisation for the Prohibition of Chemical Weapons

The OPCW was established in The Hague, the Netherlands, to achieve the objectives and purposes of the CWC and to ensure the implementation of its provisions, including those for international verification of compliance with it, and to provide a forum for consultation and cooperation among State Parties. The Convention requires State Parties to declare activities concerning chemical weapons and relevant parts of their civilian chemical industry. Declarations of existing chemical weapons and chemical industry are to be confirmed through on-site inspection. It is the responsibility of the inspectors, inter alia, to verify the quantity and identity of chemicals and the types and number of munitions, devices, and other equipment. Where possible, samples should be analyzed on-site, but if necessary they can be transferred for analysis off-site at designated laboratories.

The Director-General of the Technical Secretariat of the OPCW is to establish a stringent regime governing the collection, handling, transport, and analysis of samples, to certify the laboratories designated to perform different types of analysis, to oversee the standardization of equipment and procedures at these designated laboratories and the mobile analytical equipment and procedures, and to monitor quality control (QC) and overall standards in relation to the certification of the designated laboratories, mobile equipment and procedures. Selection will be made among the designated laboratories to perform analytical or other functions in relation to specific investigations.

In November 1998, the Director-General designated laboratories from seven countries (one from each country): China, Finland, the Netherlands, South Korea, Sweden, Switzerland, and the US. In June 1999, five other laboratories were designated (one from each country): Czech Republic, France, Germany, Poland, and the UK. In order to maintain or obtain designation, a laboratory has to participate annually in one OPCW proficiency test and perform successfully (see section 3.4). Another OPCW requirement for designation is national accreditation of the laboratory.

2.4 Definition of Chemical Weapons

Chemical weapons are defined as: (a) toxic chemicals and their precursors, except where intended for purposes not prohibited under the CWC, as long as the types and quantities are consistent with such purposes; (b) munitions and devices, specifically designed to cause death or other harm through the toxic properties of those toxic chemicals specified in (a), which would be released as a result of the employment of such munitions and devices; (c) any equipment specifically designed for use directly in connection with the employment of munitions and devices specified in (b).[1] A toxic chemical means any chemical which through its chemical action on life processes can cause death, temporary incapacitation, or permanent harm to humans or animals. This general-purpose criterion was included to protect the CWC from circumvention. CW agent is a synonym for a chemical that acts as a chemical weapon.

2.5 Schedules of Chemicals

The toxic chemicals and their precursors that have been identified for the application of verification measures are listed in Schedules 1–3 (Table 1) contained in the

Table 1 Schedules 1–3

Schedule 1

	A. Toxic chemicals (CAS reg. no.)	Structure
(1)	*O*-Alkyl (R_1: ≤C_{10}, incl. cycloalkyl) alkyl (R_2: Me, Et, *n*-Pr or *i*-Pr)-phosphonofluoridates e.g. Sarin: *O*-Isopropyl methylphosphonofluoridate (107-44-8); Soman: *O*-Pinacolyl methylphosphonofluoridate (96-64-0)	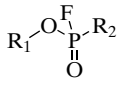
(2)	*O*-Alkyl (R_1: ≤C_{10}, incl. cycloalkyl) *N*,*N*-dialkyl (R_2: Me, Et, *n*-Pr or *i*-Pr) phosphoramidocyanidates e.g. Tabun: *O*-Ethyl *N*,*N*-dimethylphosphoramidocyanidate (77-81-6)	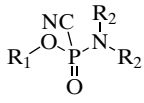
(3)	*O*-Alkyl (R_1: H or ≤C_{10}, incl. cycloalkyl) *S*-2-dialkyl (R_2: Me, Et, *n*-Pr or *i*-Pr)-aminoethyl alkyl (R_3: Me, Et, *n*-Pr or *i*-Pr) phosphonothiolates and corresponding alkylated or protonated salts (R_4: alkyl or H; X: halogen) e.g. VX: *O*-Ethyl *S*-2-diisopropylaminoethyl methyl phosphonothiolate (50782-69-9)	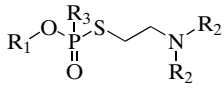 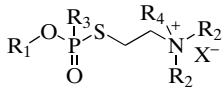
(4)	Sulfur mustards: 2-Chloroethylchloromethylsulfide (2625-76-5)	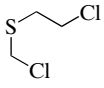
	Mustard gas: Bis(2-chloroethyl)sulfide (505-60-2)	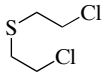
	Bis(2-chloroethylthio)methane (63869-13-6)	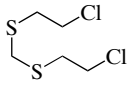
	Sesquimustard: 1,2-Bis(2-chloroethylthio)ethane (3563-36-8)	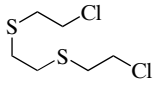
	1,3-Bis(2-chloroethylthio)-*n*-propane (63905-10-2)	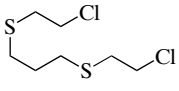
	1,4-Bis(2-chloroethylthio)-*n*-butane (142868-93-7)	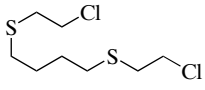
	1,5-Bis(2-chloroethylthio)-*n*-pentane (142868-94-8)	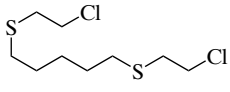
	Bis(2-chloroethylthiomethyl) ether (63918-90-1)	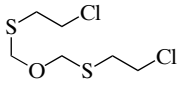
	O-Mustard: Bis(2-chloroethylthioethyl) ether (63918-89-8)	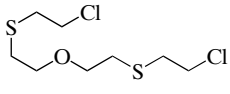
(5)	Lewisites: Lewisite 1: 2-Chlorovinyldichloroarsine (541-25-3)	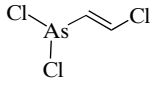
	Lewisite 2: Bis(2-chlorovinyl)chloroarsine (40334-69-8)	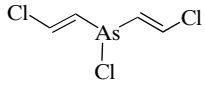

Table 1 (*continued*)

	Lewisite 3: Tris(2-chlorovinyl)arsine (40334-70-1)	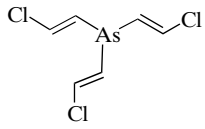
(6)	Nitrogen mustards: HN1: Bis(2-chloroethyl)ethylamine (538-07-8)	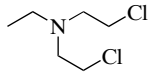
	HN2: Bis(2-chloroethyl)methylamine (51-75-2)	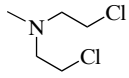
	HN3: Tris(2-chloroethyl)amine (555-77-1)	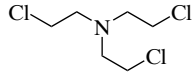
(7)	Saxitoxin (35523-89-8)	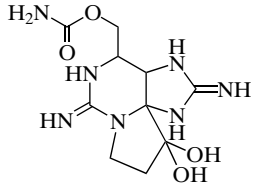
(8)	Ricin (9009-86-3)	Structure not specified
	Precursors	
B.		
(9)	Alkyl (R_1: Me, Et, *n*-Pr or *i*-Pr) phosphonyldifluorides e.g. DF: Methylphosphonyldifluoride (676-99-3)	
(10)	*O*-Alkyl (R_1: H or ≤C_{10}, incl. cycloalkyl) *O*-2-dialkyl (R_2: Me, Et, *n*-Pr or *i*-Pr)-aminoethyl alkyl (R_3: Me, Et, *n*-Pr or *i*-Pr)-phosphonites and corresponding alkylated or protonated salts (R_4: alkyl or H; X: halogen) e.g. QL: *O*-Ethyl *O*-2-diisopropylaminoethyl methylphosphonite (57856-11-8)	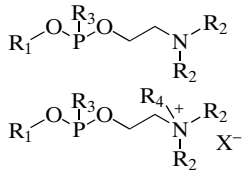
(11)	Chlorosarin: *O*-isopropyl methylphosphonochloridate (1445-76-7)	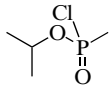
(12)	Chlorosoman: *O*-Pinacolyl methylphosphonochloridate (7040-57-5)	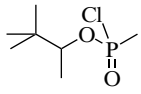

Schedule 2

A.	Toxic chemicals (CAS reg. no.)	Structure
(1)	Amiton: *O*,*O*-Diethyl *S*-[2-(diethylamino)ethyl] phosphorothiolate (78-53-5) and corresponding alkylated or protonated salts (R_1: alkyl or H; X: halogen)	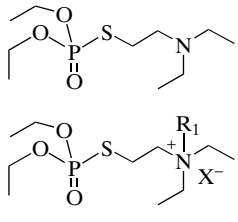

(*continued overleaf*)

Table 1 (*continued*)

(2) PFIB: 1,1,3,3,3-Pentafluoro-2-(trifluoromethyl)-1-propene (382-21-8)	
(3) BZ: 3-Quinuclidinyl benzilate (6581-06-2)[a]	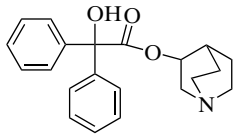
B. Precursors	
(4) Chemicals, except for those listed in Schedule 1, containing a phosphorus atom to which is bonded one methyl, ethyl or propyl (normal or iso) group but not further carbon atoms, e.g. Methylphosphonyl dichloride (676-97-1); Dimethyl methylphosphonate (756-79-6) Exemption: Fonofos: *O*-Ethyl *S*-phenyl ethylphosphonothiolothionate (944-22-9)	
(5) *N*,*N*-Dialkyl (R_1: Me, Et, *n*-Pr or *i*-Pr) phosphoramidic dihalides (X: halogen)	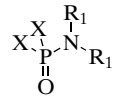
(6) Dialkyl (R_1: Me, Et, *n*-Pr or *i*-Pr) *N*,*N*-dialkyl (R_2: Me, Et, *n*-Pr or *i*-Pr)-phosphoramidates	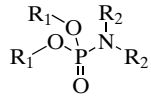
(7) Arsenic trichloride (7784-34-1)	$AsCl_3$
(8) 2,2-Diphenyl-2-hydroxyacetic acid (76-93-7)	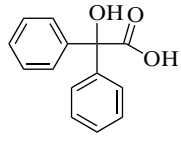
(9) Quinuclidin-3-ol (1619-34-7)	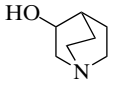
(10) *N*,*N*-Dialkyl (R_1: Me, Et, *n*-Pr or *i*-Pr) aminoethyl-2-chlorides and corresponding protonated salts (X: halogen)	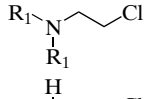 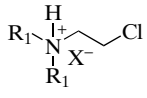
(11) *N*,*N*-Dialkyl (R_1: Me, Et, *n*-Pr or *i*-Pr)-aminoethan-2-ols and corresponding protonated salts (X: halogen) Exemptions: *N*,*N*-Dimethylaminoethanol (108-01-0) and corresponding protonated salts *N*,*N*-Diethylaminoethanol (100-37-8) and corresponding protonated salts	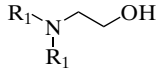 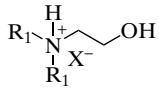
(12) *N*,*N*-Dialkyl (R_1: Me, Et, *n*-Pr or *i*-Pr)-aminoethane-2-thiols and corresponding protonated salts (X: halogen)	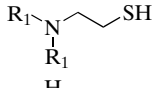 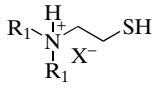

Table 1 (*continued*)

(13)	Thiodiglycol: Bis(2-hydroxyethyl)sulfide (111-48-8)	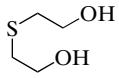
(14)	Xinacolyl alcohol: 3,3-Dimethylbutan-2-ol (464-07-3)	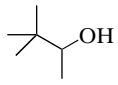

Schedule 3

A.	Toxic chemicals (CAS reg. no.)	Structure
(1)	Phosgene: Carbonyl dichloride (75-44-5)	$COCl_2$
(2)	Cyanogen chloride (506-77-4)	ClCN
(3)	Hydrogen cyanide (74-90-8)	HCN
(4)	Chloropicrin: Trichloronitromethane (76-06-2)	Cl_3CNO_2
B.	Precursors	
(5)	Phosphorus oxychloride (10025-87-3)	$POCl_2$
(6)	Phosphorus trichloride (7719-12-2)	PCl_3
(7)	Phosphorus pentachloride (10026-13-8)	PCl_5
(8)	Trimethyl phosphite (121-45-9)	$P(OCH_3)_3$
(9)	Triethyl phosphite (122-52-1)	$P(OCH_2CH_3)_3$
(10)	Dimethyl phosphite (868-85-9)	$(CH_3O)_2POH$
(11)	Diethyl phosphite (762-04-9)	$(CH_3CH_2O)_2POH$
(12)	Sulfur monochloride (10025-67-9)	SCl
(13)	Sulfur dichloride (10545-99-0)	SCl_2
(14)	Thionyl chloride (7719-09-7)	$SOCl_2$
(15)	Ethyldiethanolamine (139-87-7)	$CH_3CH_2N(CH_2CH_2OH)_2$
(16)	Methyldiethanolamine (105-59-9)	$CH_3N(CH_2CH_2OH)_2$
(17)	Triethanolamine (102-71-6)	$N(CH_2CH_2OH)_3$

[a] Subject to special thresholds for declaration and verification.
CAS, Chemical Abstracts Service.

Annex on Chemicals of the CWC.[1] Schedule 1 includes chemicals developed, produced, stockpiled, or used as a chemical weapon as defined above, and chemicals structurally close to them. Schedule 2 lists three toxic chemicals not included in Schedule 1 and degradation products and precursors of these toxic chemicals as well as of those of Schedule 1. Schedule 3 lists four toxic chemicals and precursors not listed in the other Schedules. Altogether tens of thousands of chemicals are included in the Schedules.

The Schedules contain mainly organic chemicals with different chemical and physical properties, being neutral chemicals, acids, bases, volatiles, and nonvolatiles, where phosphorus, fluorine, sulfur, chlorine, nitrogen, and oxygen occur frequently. Riot control agents are not included in the Schedules, but their use as a method of warfare is prohibited in the General Obligations of the CWC.

3 DETECTION AND VERIFICATION OF CHEMICAL WEAPONS

3.1 Recommended Operating Procedures

After more than 10 years of methods development for the identification of CW agents and their precursors and

degradation products,[9,10] the Finnish Research Project on the Verification of Chemical Disarmament (Finnish Institute for Verification of the Chemical Weapons Convention, VERIFIN) in 1988 and 1989 published SOPs for the Verification of Chemical Disarmament.[11] In 1989, the first international interlaboratory comparison (round-robin) test for the verification of chemical disarmament was carried out.[2] According to the proposal of the experts meeting, the SOPs were renamed ROPs. The first test and three subsequent tests[2] were aimed at testing and developing the existing procedures, the ROPs. Through sustained international collaboration, the ROPs for Sampling and Analysis in the Verification of Chemical Disarmament were upgraded annually.[2,12] In all these tests and in later proficiency tests the ROPs have been widely and successfully applied.

The ROPs[12] give instructions for sampling, sample preparation, and instrumental analysis. Sampling will be the responsibility of the OPCW inspectors and sample preparation and subsequent analysis of the samples is carried out by the off-site laboratory in a case where further analysis is required. Rather than being the best choice for a single analyte or matrix type, ROPs provide an optimal approach for all in an unknown situation. The ROPs related to the sample preparation are listed in a separate article (**Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention**); the ROPs for the analysis techniques are the following:

- ROP for the Analysis of Treaty-related Compounds by Gas Chromatography;
- ROP for the Identification of Degradation Products of Treaty-related Compounds by Micro-liquid Chromatography with Flame Photometric Detection;
- ROP for the Identification of Methylphosphonic Acid and its Alkyl Derivatives by Ion Chromatography with Conductivity Detection;
- ROP for Capillary Zone Electrophoretic Analysis of the Alkylphosphonic Acids and Related Monoesters Using Indirect UV Detection;
- ROP for Identification of Treaty-related Compounds by Gas Chromatography/Mass Spectrometry;
- ROP for the Identification of Treaty-related Compounds and their Degradation Products by Liquid Chromatography/Thermospray/Mass Spectrometry;
- ROP for the Identification of Treaty-related Compounds and their Degradation Products by Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry;
- ROP for the Identification of Treaty-related Compounds by NMR Spectrometry;
- ROP for the Identification of Treaty-related Chemicals by Gas Chromatography/Fourier Transform Infrared Spectrometry.

3.2 Organisation for the Prohibition of Chemical Weapons Standard Operating Procedures

During the Preparatory Commission of the OPCW the Task Force on Analytical Issues discussed the analytical procedures required for on-site analysis. In June 1995, Finland, France, and the US presented proposals for on-site sample preparation methods. In October 1995, the French proposal was withdrawn and in February 1996, a collaboration program between Finland and the US was agreed. In January 1997, improved methods were accepted for training of the analyst inspectors of the OPCW. After this, the methods have been continuously developed and the results have been forwarded to the OPCW.

The OPCW has adopted the procedures and modified them. On 27 January, 1999 the OPCW established the first version and on 8 March, 1999 the second version of the SOP for On-site Analysis. This SOP sets a general OPCW strategy for chemical analysis on-site applicable to different inspection sites and types. The detailed procedures for protection, decontamination, collection, splitting, sample preparation, analysis, equipment, and chain-of-custody and documentation on-site are to be specified in

- SOP for the Donning and Doffing of Individual Protective Equipment;
- SOP for Set-up and Operation of the Contamination Control Station for the Decontamination of Personnel and Equipment;
- SOP for Preparation, Storage and Use of Decontamination Solutions I and II;
- SOP for Collecting and Splitting Samples under Toxic and Hazardous Conditions On-site;
- SOP for the Splitting of Samples On-site;
- SOP for Handling of Inspection Samples;
- SOP for the Chain-of-custody and Documentation During Inspection Period;
- SOP for the Collection of Wipe Samples On-site;
- SOP for the Collection of Liquid Samples On-site;
- SOP for the Collection of Solid Samples On-site;
- SOP for Preparation of Wipe Samples On-site for GC/MS Analysis;
- SOP for Preparation of Aqueous Samples On-Site for GC/MS Analysis;
- SOP for Preparation of Soil Samples On-site for GC/MS Analysis;
- SOP for Bruker EM 640S Portable GC/MS with AMDIS Data Analysis Software (Blinded Mode) for

the Qualitative Analyses of Liquid Samples Prepared by On-site Methods;
- SOP for Pierce Reacti-therm Heating Module and Reacti-vap Evaporator;
- SOP for Jouan Centrifugal Evaporator RC 10-22;
- SOP for Buchi Vacuum System B-178;
- SOP for the Bransonic Ultrasonic Cleaner.

The main analytical techniques in the SOP for On-site Analysis are GC/EIMS and FTIRS. The procedures for fourier transform infrared (FTIR) are to be established in the future. GC/EIMS is the analytical technique used for chemical identification on-site. The off-site analyses performed by at least two different designated laboratories provide unambiguous results. In addition to the techniques mentioned here the inspection team may use, in certain conditions, the analytical techniques made available to it by the inspected State Party.

3.3 Analysis Strategy in Off-site Laboratory

In verification of presence or absence of chemicals related to the CWC (Table 1), the analysis of a specific sample in an off-site designated laboratory includes preparation of samples and their analysis by different chromatographic and spectrometric techniques. In general, the chromatographic techniques applying specific detectors are used in screening while the spectrometric techniques are used in a complementary manner to obtain unambiguous identification of Scheduled chemicals. According to the rules of the OPCW, identification is considered unambiguous if the results with two analytical, preferably spectrometric, techniques confirm the presence of the same chemical. Analyses are mostly qualitative (identification). Work is conducted according to the ROPs or other documented procedures of demonstrated performance. The analysis strategy employed at VERIFIN is described in the following (Figure 1). Laboratories worldwide may have different instrumentation and perhaps differing sample preparation procedures, which means that each laboratory actually requires its own analysis strategy.

3.3.1 Sample Preparation

After arrival, check-up, and coding, samples and blanks (if available) are divided into portions allowing multiple sample preparation procedures to meet the requirements of different analytical techniques and chemicals. QC samples may be prepared. By way of example, a blind sample containing only the solvents/reagents used in preparation of each type of sample is always prepared. One set of samples is prepared for the analysis techniques relying on GC separation. Samples from water and aqueous extracts can be prepared in a straightforward way for analysis by liquid chromatography/mass spectrometry (LC/MS) and CE. Samples for NMR are normally prepared so as to obtain 5- to 10-fold concentration. In the next stage, screening of the samples for the analytes is done by different analysis techniques.

3.3.2 Screening

The samples are always screened by GC. Results from screening by GC may be useful in analyses carried out by techniques relying on it. The screening is done with element (nitrogen, phosphorus, sulfur)-specific detectors (nitrogen–phosphorus detector (NPD) and flame photometric detector (FPD)) together with retention index (RI) monitoring. This is particularly advantageous, because almost all of the scheduled chemicals are then detectable without interference from hydrocarbons; the other chemicals not including phosphorus, nitrogen, and sulfur are screened using GC/EIMS by monitoring ions specific to them. GC/EIMS is also useful for screening of the family members of nerve agents (see Schedule 1.A.1–1.A.3) and their derivatized degradation products. Analogously, LC/MS is useful for screening of the degradation products, and there, no derivatization is required. Other screening techniques include ^{19}F- and/or ^{31}P{^{1}H}-NMR spectroscopy and CE. Screening by NMR can reveal phosphorus- and fluorine-containing chemicals, and screening by CE, for instance, presence of nonvolatile alkylphosphonic acids and their monoesters. One advantage of CE and LC/MS over GC techniques is the ability to separate and detect polar, nonvolatile analytes without sample pretreatment and derivatization. The boundary between screening and the identification is not always clear-cut.

3.3.3 Analysis, Data Evaluation, and Identification

Analysis for identification of known, volatile chemicals, which are fully resolved from the matrix, is carried out with GC/low-resolution MS: GC/EIMS provides a spectrum characteristic of the chemical, while GC/CIMS (chemical ionization mass spectrometry) provides molecular weight information. Problems caused by strong matrix interference are solved using GC/MS/MS (tandem mass spectrometry) or GC/HRMS (high-resolution mass spectrometry) in selective ion monitoring (SIM) mode. Polar and nonvolatile chemicals in aqueous solutions or extracts can be analyzed with LC/MS and LC/MS/MS by various ionization techniques (atmospheric pressure chemical ionization, APCI; electrospray ionization, ESI; or ionspray ionization, ISI). GC/FTIRS using cryodeposition offers pure and characteristic (solid-phase) spectra useful for identification. ^{1}H-, ^{13}C{^{1}H}-, ^{19}F-, ^{31}P{^{1}H}-, and ^{31}P-NMR spectra are, as well, characteristic for a chemical, but in ^{1}H-NMR,

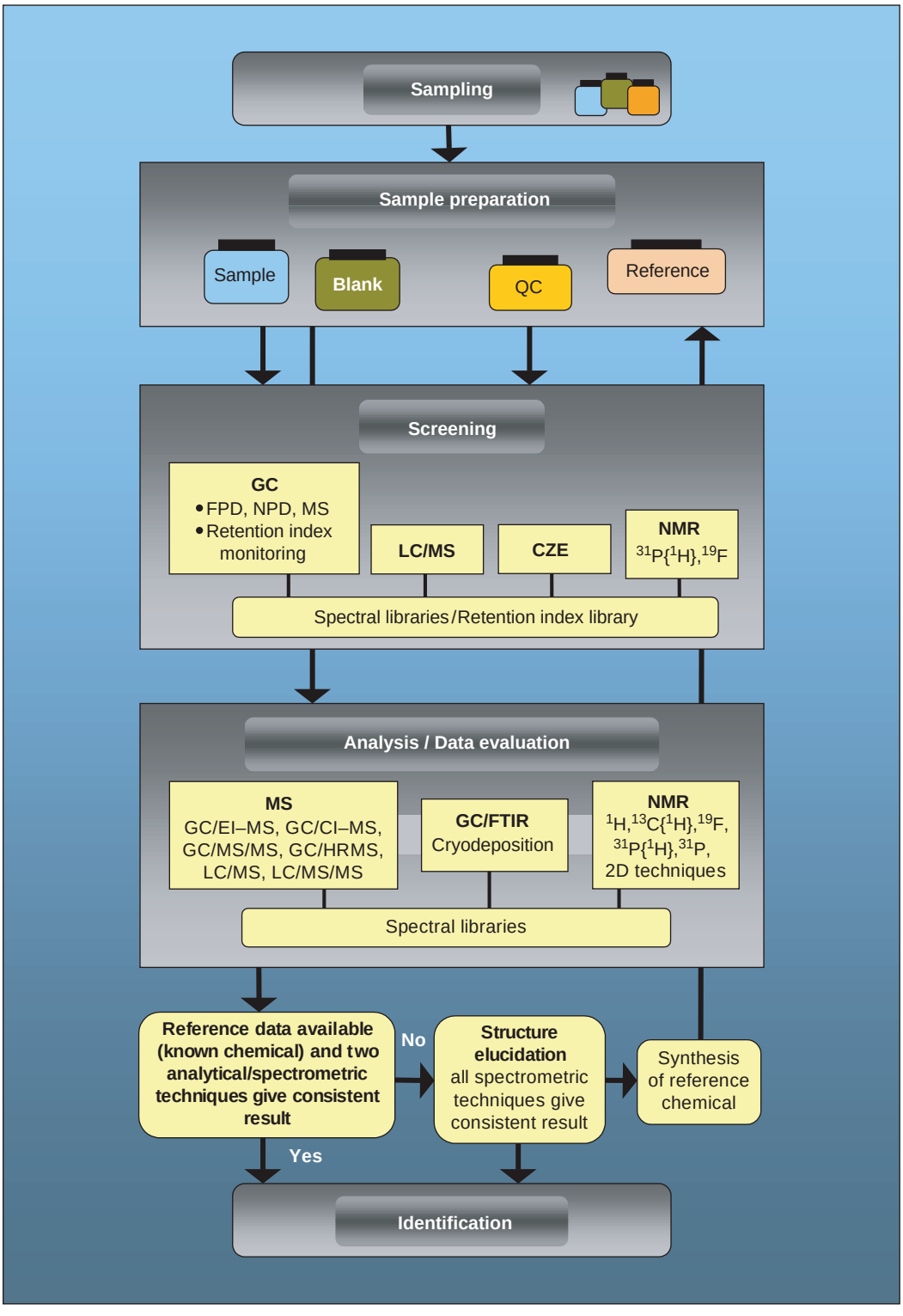

Figure 1 Analysis strategy employed at VERIFIN for verification of CWC-related chemicals. (Reproduced by permission from Mesilaakso.[4] © John Wiley & Sons Inc., 1998.)

sample background may cause problems of resonance overlapping.

In the case of MS, FTIRS, and NMR, the spectrum recorded from a sample is compared with library spectra, and upon fulfillment of certain criteria identification is obtained. In VERIFIN's view, unambiguous identification of a chemical is obtained if at least two spectrometric techniques give consistent results. Identical identifications obtained by GC/EIMS and GC/CIMS are considered to constitute an unambiguous result.

If preliminary results indicate a Scheduled chemical which is not in the spectral libraries, its structure elucidation is carried out. The GC/EIMS, GC/CIMS, and LC/MS spectra are interpreted. Molecular weight and presence of heteroatoms like sulfur, chlorine, and phosphorus are determined. GC/MS/MS and LC/MS/MS assist in interpretation of the fragmentation of ions of interest. Elemental composition of the molecular and fragment ions can be determined with GC/HRMS. Interpretation of GC/FTIR spectrum provides information on functional groups, and the different types of one- and two-dimensional NMR spectra information on type and number of nuclei, and their location in the molecule, i.e. the molecular structure. A consistent result from all spectrometric techniques used is needed for unambiguous identification. The final confirmation is obtained when the suspected chemical is synthesized as a reference chemical, and the spectra recorded from the sample and the synthesized chemical are identical.

The analysis report must include all analytical data (chromatographic and spectrometric) supporting the identifications made, describe the sample preparation and analytical methods in detail (or make reference to ROPs, SOPs, or other procedures), and give information on the identified chemical (CAS registry number, structural formula, and International Union of Pure and Applied Chemistry or CWC name).

3.4 International Interlaboratory Comparison and Proficiency Tests

The first international interlaboratory comparison (round-robin) test for the verification of chemical disarmament took place in 1989 as a Finnish initiative. This test and three subsequent tests were coordinated by Finland.[2] The Provisional Technical Secretariat (PTS) of the Preparatory Commission of the OPCW coordinated the first international OPCW/PTS interlaboratory comparison test in 1994[3] and two other tests in 1995. The tests of 1995 were also named ''trial proficiency tests''. In 1996, a series of official proficiency tests was started and two tests were arranged. One test was arranged in 1997, two tests in 1998 and one test in 1999. In the future, the purpose is to arrange two proficiency tests annually, one in spring and one in autumn.

The aim of the OPCW proficiency tests is to establish and maintain a recognized and transparent methodology for the continued assessment of the technical competence of the participating laboratories. For this purpose, the Preparatory Commission of the OPCW established protocols to cover sample preparation and evaluation of the results of the OPCW proficiency tests: SOP for Preparation of Test Samples for OPCW/PTS Proficiency Tests and SOP for Evaluation of Results of OPCW Proficiency Tests. The scoring rules have been described in Criteria for Acceptable Performance of Laboratories in Proficiency Testing. In addition, the OPCW has made decisions on the criteria for the conduct of OPCW/PTS proficiency testing and for acceptable performance of laboratories in proficiency testing, and has given recommendations for the future on the basis of the Fourth and Fifth Official Tests.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
CAS	Chemical Abstracts Service
CE	Capillary Electrophoresis
CIMS	Chemical Ionization Mass Spectrometry
CW	Chemical Warfare
CWC	Chemical Weapons Convention
EIMS	Electron Ionization Mass Spectrometry
ESI	Electrospray Ionization
FPD	Flame Photometric Detector
FTIR	Fourier Transform Infrared
FTIRS	Fourier Transform Infrared Spectroscopy
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HRMS	High-resolution Mass Spectrometry
ISI	Ionspray Ionization
LC/MS	Liquid Chromatography/Mass Spectrometry
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NPD	Nitrogen–Phosphorus Detector
OPCW	Organisation for the Prohibition of Chemical Weapons
PTS	Provisional Technical Secretariat
QC	Quality Control
RI	Retention Index

ROP	Recommended Operating Procedure
SIM	Selective Ion Monitoring
SOP	Standard Operating Procedure
VERIFIN	Finnish Institute for Verification of the Chemical Weapons Convention

REFERENCES

1. Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction. Signed in January 1993. Printed and Distributed by the OPCW/PTS. The Depositary of this Convention is the Secretary-General of the United Nations, from whom a Certified True Copy can be Obtained.
2. M. Rautio (ed.), 'F. International Inter-laboratory Comparison (Round-robin) Test for the Verification of Chemical Disarmament. F.1 Testing of Existing Procedures (1990). F.2 Testing of Procedures on Simulated Industry Samples (1991). F.3 Testing of Procedures on Simulated Military Facility Samples (1992). F.4 Validating of Procedures for Water and Soil Samples (1993). Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', The Ministry for Foreign Affairs of Finland, Helsinki, 1990–1993.
3. M. Rautio (ed.), 'H. Inter-laboratory Comparison Test Coordinated by the Provisional Technical Secretariat for the Preparatory Commission for the Organisation for the Prohibition of Chemical Weapons H.1 First Inter-laboratory Comparison Test. Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', The Ministry for Foreign Affairs of Finland, Helsinki, 1994.
4. M. Mesilaakso, 'Chemical Weapons Convention, NMR Analysis', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 949–971, 1998.
5. E.R.J. Wils, A.G. Hulst, P.E.J. Verwiel, S.H. van Krimpen, A. Niederhauser, 'Identification of an Octyl Methylphosphonofluoridate Mixture in Chemical Industry Samples', *Fresenius' J. Anal. Chem.*, **343**, 297–303 (1992).
6. M. Mesilaakso, E.-L. Tolppa, 'Detection of Trace Amounts of Chemical Warfare Agents and Related Compounds in Rubber, Paint, and Soil Samples by ^{1}H and $^{31}P\{^{1}H\}$ NMR Spectroscopy', *Anal. Chem.*, **68**, 2313–2318 (1996).
7. M.T. Söderström, H. Björk, V.M.A. Häkkinen, O. Kostiainen, M.-L. Kuitunen, M. Rautio, 'Identification of Compounds Relevant to the Chemical Weapons Convention Using Selective Gas Chromatography Detectors, Gas Chromatography–Mass Spectrometry and Gas Chromatography–Fourier Transform Infrared Spectroscopy in an International Proficiency Test', *J. Chromatogr.*, **A742**, 191–203 (1996).
8. C. Albaret, D. Lœillet, P. Augé, P.-L. Fortier, 'Application of Two-dimensional $^{1}H-^{31}P$ Inverse NMR Spectroscopy to the Detection of Trace Amounts of Organophosphorus Compounds Related to the Chemical Weapons Convention', *Anal. Chem.*, **69**, 2694–2700 (1997).
9. J. Enqvist, M. Rautio (eds.), 'A. General Reports. A.1 Chemical and Instrumental Verification of Organophosphorus Warfare Agents. A.2 Technical Evaluation of Selected Scientific Methods for the Verification of Chemical Disarmament. Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', The Ministry for Foreign Affairs of Finland, Helsinki, 1977, 1984.
10. J. Enqvist, M. Rautio (eds.), 'B. Systematic Identification of Chemical Warfare Agents. B.1 Identification of Organophosphorus Warfare Agents. An Approach for the Standardization of Techniques and Reference Data (1979). B.2 Identification of Degradation Products of Organophosphorus Warfare Agents (1980). B.3 Identification of Non-phosphorus Warfare Agents (1982). B.4 Identification of Precursors of Warfare Agents, Degradation Products of Non-phosphorus Agents, and Some Potential Agents (1983). B.5 Identification of Selected Mycotoxins: Trichothecenes, Aflatoxins, and Related Mycotoxins (1986). Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', The Ministry for Foreign Affairs of Finland, Helsinki, 1979–1986.
11. M. Rautio (ed.), 'D. Standard Operating Procedures for the Verification of Chemical Disarmament. D.1 A Proposal for Procedures Supporting the Reference Database (1988). D.2 Second Proposal for Procedures Supporting the Reference Database (1989). Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', The Ministry for Foreign Affairs of Finland, Helsinki, 1988–1989.
12. M. Rautio (ed.), 'Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament', The Ministry for Foreign Affairs of Finland, Helsinki, 1993, 1994.

Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Camille A. Boulet
Defence Research Establishment Suffield, Alberta, Canada

1 Introduction 1
2 Capillary Electrophoresis 2
2.1 Capillary Zone Electrophoresis 2
2.2 Electrophoretic Mobility 2
2.3 Capillary Electrophoresis Separation, Efficiency, and Resolution 3
2.4 Limits of Detection 3
3 Chemical Weapons Convention and Related Compounds 3
3.1 Scheduled Chemicals 3
3.2 Degradation Products of Chemical Warfare Agents 4
4 Capillary Electrophoretic Analysis of Chemical Weapons Convention Related Compounds 4
4.1 Chromatographic Analysis of Chemical Weapons Convention Related Compounds 4
4.2 Capillary Electrophoresis Analysis of Organophosphates 5
4.3 Sulfur-containing Compounds 7
4.4 Toxins 8
5 Applications 9
5.1 Environmental Analysis 9
5.2 Quality Assurance 9
6 Conclusion 10
Acknowledgments 10
Abbreviations and Acronyms 11
Related Articles 11
References 11

This article describes the application of capillary electrophoresis (CE) to the analysis of compounds relevant to the Chemical Weapons Convention (CWC). CE is an analytical technique that employs narrow-bore, fused-silica capillaries to perform high-efficiency separations of analytes based on their mobilities in an electric field. In CE, analytes are dissolved in a buffer solution and placed in a capillary to which an electric field is applied. The analytes then migrate at a rate determined by their charge and size and are detected as they migrate past a detector. CE can analyze a broad range of compounds and is particularly applicable to the analysis of water-soluble degradation products of scheduled compounds under the CWC. Compounds typically analyzed by CE include the hydrolysis products of nerve agents such as the alkylphosphonates and hydrolytic or oxidative products of the sulfur mustards such as thiodiglycol. The alkyphosphonates require the use of indirect ultraviolet (UV) detection or other detection methods as they do not possess a suitable UV chromophore. Degradation products of the sulfur mustards have been analyzed by micellar electrokinetic chromatography (MEKC) and direct UV detection. The ability of CE to analyze anionic, cationic, zwitterionic, and neutral CWC-related compounds, whether they be chemical warfare (CW) degradation products or scheduled starting materials, without the requirement for elaborate sample processing or labeling procedures, is an important demonstration of the utility of CE for the analysis of CWC-related compounds. CE can be an important screening method for rapid sample processing and further analysis for unambiguous identification. The relative ease of analysis should also increase its utility in field and on-site analysis where minimizing the logistic burden is attractive.

1 INTRODUCTION

CE is an analytical technique that employs narrow-bore, fused-silica capillaries to perform high-efficiency separations of analytes based on their mobilities in an electric field.[1] The separation efficiency, which may be greater than 1 000 000 theoretical plates, is superior to that of many other chromatographic methods and is an important reason for the rapid growth of this technique.[2–4] In CE, analytes are dissolved in a buffer solution and placed in a capillary to which an electric field is applied. The analytes then migrate at a rate determined by their charge and size and are detected as they migrate past a detector. CE can analyze a broad range of compounds, has a simple instrument design which can be fully automated, and has very low volume requirements (nanoliter injection volumes are typical).[5]

The *Convention on the Prohibition of the Development, Production, Stockpiling, and Use of Chemical Weapons and their Destruction*, otherwise known as the CWC,[6] is an international treaty which will potentially eliminate the worldwide threat of CW.[7] CW agent degradation

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

products or impurities can provide an indication of CW agent use, act as tracers to follow synthetic routes, or identify sources of precursors. For instance, analysis of samples taken during an allegation of use investigation were found to contain mustard (H) as well as other longer chain sulfur vesicants.[8] The identification of isopropyl methylphosphonic acid and methylphosphonic acid, decomposition products of the nerve agent isopropyl methylphosphonofluoridate (sarin (GB)), has also been reported in soil samples taken from bomb craters in Iraq years after the alleged attack.[9] More recently, isopropyl methylphosphonate has been detected in the urine[10] and serum[11] of persons exposed to GB during a terrorist attack in Japan.[12]

The degradation products of CW agents are often polar, water-soluble compounds. Thus analytical techniques which can directly analyze the target compounds without the need for derivatization are valuable for both on-site screening of samples and laboratory-based identification. CE is an important analytical method for the detection of these analytes because of the efficiency of separation, speed of analysis, and ease of operation. Degradation products of CW nerve agents are primary targets for analysis. However, these organophosphates are non-UV-absorbing and so are generally analyzed by indirect detection methods. Other compounds, such as the decomposition products of sulfur mustards, can be analyzed by direct UV detection but are non-ionizable and require the use of special CE experimental conditions. However, both classes of compounds have been readily analyzed, demonstrating the versatility of CE for the analysis of CWC-related compounds. The development of instruments for atmospheric pressure ionization mass spectrometry (MS) and the advent of commercial instruments for CE combined with electrospray ionization mass spectrometry (ESIMS) will greatly increase the importance of CE for the identification of CW agents and their degradation products.

2 CAPILLARY ELECTROPHORESIS

CE has been the subject of extensive reviews[13–16] and there are several books and publications describing the technique.[17,18] Key aspects of CE are discussed here to introduce CE and to review important fundamental issues for the analysis of CWC-related compounds. These issues can affect the quality of CE separations, detection of target compounds, and method development. Specialized CE methods, such as capillary gel electrophoresis (CGE), capillary isoelectric focusing (CIEF), capillary isotachophoresis (CITP), and capillary electrochromatography (CEC) are beyond the scope of this article.

2.1 Capillary Zone Electrophoresis

Capillary zone electrophoresis (CZE) has been used for the analysis of amino acids, peptides, ions, enantiomers, and many other ionic species. CZE is the simplest form of CE, employing a narrow-bore, bare fused-silica capillary (10–200 μm i.d., 370 μm o.d., polyimide coated) filled only with buffer. CZE achieves high-efficiency separations of both large and small molecules based on their mobilities in an electric field. The instrument design is fairly simple, consisting of two buffer reservoirs connected via a fused-silica capillary, a high-voltage power supply, and a detection system.[19]

Surface silanol groups of the bare fused-silica capillaries are partly ionized (SiO^-) in the presence of the electrophoretic medium, a run buffer in the pH range 2.5–10. This leads to the formation of a double layer of immobile cations tightly bound to the negative charges on the silica and a potential difference (zeta potential). Adjacent to this layer is a layer of weakly bound cations and their sphere of hydration. When a potential is applied across the capillary, these solvated cations migrate towards the cathode, resulting in a net flow of liquid through the capillary known as electroosmotic flow (EOF). The EOF is greater in magnitude than the mobility of the ions. Therefore, all analytes are swept in the same direction, regardless of charge. The zeta potential is also dependent on the ionic strength of the buffer. Increasing ionic strength decreases the zeta potential and thus the EOF. Selectivity in CZE is affected by the buffer pH and buffer additives (e.g. surfactants, EOF modifiers, or chiral selectors).

Since the EOF can be significantly greater than the electrophoretic mobilities of the individual ions in the sample, both anions and cations can be separated in the same run. Cations with the smallest mass-to-charge ratio are the fastest migrating species followed by cations with higher mass to charge ratios, then neutral species, and lastly anions. Anions migrate because they are swept to the cathode by the EOF. Control of the EOF is thus an important experimental consideration in method development and reproducibility of the analysis.[20–22] The conditions used during a CE separation can greatly affect the selectivity of the separation and so the run buffer selection and composition are extremely important to the success of any CE separation.

2.2 Electrophoretic Mobility

Identification of an analyte can be made on the basis of migration time or relative migration times, or by measurement of an analyte's electrophoretic mobility. Measurement of the electrophoretic mobility can compensate for variability in the migration time due to unstable electric

field, changes in buffer capacity or composition, temperature fluctuations, and changes to the capillary surface conditions.[23–25] The apparent electrophoretic mobility (μ_{app}) is characteristic for a given species at a certain pH. The apparent mobility results from the sum of the true electrophoretic mobility of the ion due to the electric field (μ_e) and the mobility due to the EOF (μ_{EOF}).

Apparent mobility (μ_{app}) can be calculated from the experimental data using Equation (1):

$$\mu_{app} = \frac{v}{E} = \frac{L_d}{t}\frac{L_t}{V} \quad (1)$$

where v is the velocity of an ion, E is the applied field strength ($V\,cm^{-1}$), L_d is the length from capillary to detector, L_t is the total length of capillary, t is the migration time, and V is the applied voltage.

In order to determine the electrophoretic mobility due solely to the attraction to the cathode (μ_e), a neutral EOF marker must be used to determine the mobility due to the EOF (μ_{EOF}), Equation (2):

$$\mu_e = \mu_{app} - \mu_{EOF} \quad (2)$$

Acetone, methanol, benzyl alcohol, and mesityl oxide can be used as EOF markers.

2.3 Capillary Electrophoresis Separation, Efficiency, and Resolution

Separations performed by CE are directly dependent on differences in the analyte mobilities and inversely dependent on the band broadening of the analytes as they migrate through the column. An important advantage of CE is that owing to EOF, the flow profile has a plug-shape as opposed to the parabolic flow profile of high-performance liquid chromatography (HPLC). This plug-shaped flow profile increases the resolution because the EOF is uniform throughout the capillary, resulting in much less band broadening due to frictional forces along the walls and pressure gradients across the column as is seen with pressure-driven systems.

Any factors contributing to band broadening will decrease the efficiency and resolution of the separation. A major source of band broadening in CE separations is the heat generated by the passage of electrical current through the fused-silica capillary. This is called Joule heat. The increase in temperature within the CE column depends on the power generated (power = voltage × current) and is determined by the capillary dimensions, conductivity of the buffer, and applied voltage. Excessive heat causes a nonuniform temperature gradient which changes the buffer viscosity across the capillary diameter. This results in different ion mobilities across the column, leading to band broadening. Thus an inherent advantage of performing CE in narrow-bore capillaries as compared with traditional electrophoresis is the rapid heat dissipation and reduction in the effects of heating. The efficiency expressed in theoretical plates (N) is given by Equation (3):

$$N = \mu_e \frac{EL_t}{D^2} \quad (3)$$

where μ_e is electrophoretic mobility, E is electric field, L_t is total length of the capillary, and D is the diffusion coefficient.[26]

It is evident from Equation (3) that the efficiency can be increased by increasing the electric field. However, this also results in increased Joule heating. Therefore, a balance between these two parameters must be achieved. Experimental parameters which can be optimized to limit Joule heating include decreasing the electric field (reducing the applied voltage), reducing the capillary internal diameter, decreasing the buffer ionic strength or concentration, and active temperature control.[27]

2.4 Limits of Detection

The limits of detection (LODs) of CE are comparable to those of other chromatographic methods. CE, because of its very low volume requirements, has very good mass sensitivity. Typical injection volumes are in the range 1–50 nL. Thus the often reported impressive picomole to femtomole mass detection limits are actually only in the low milligrams to high micrograms per liter (parts per million to parts per billion) range. Such concentrations are easily achieved by HPLC. The concentration limit of detection (CLOD), typically in the micromolar range, is higher than for HPLC where typical injection volumes of 1–10 μL allow for higher column loading.

Approaches to improvements in sensitivity include both precolumn and on-column sample concentration, derivatization with laser-induced fluorescence (LIF), and modification to the detection system.[28] On-column methods offer the advantage for trace analysis of minimizing the sample handling and possible loss of analyte. Three approaches are commonly used for sample concentration: sample stacking, sample focusing, and isotachophoresis. More recently, the development of on-line sample extraction devices such as solid-phase preconcentration phase CE has permitted injection volumes as high as 100 μL.[29]

3 CHEMICAL WEAPONS CONVENTION AND RELATED COMPOUNDS

3.1 Scheduled Chemicals

The CWC contains schedules of chemicals with specific verification and reporting requirements. The ''Annex on

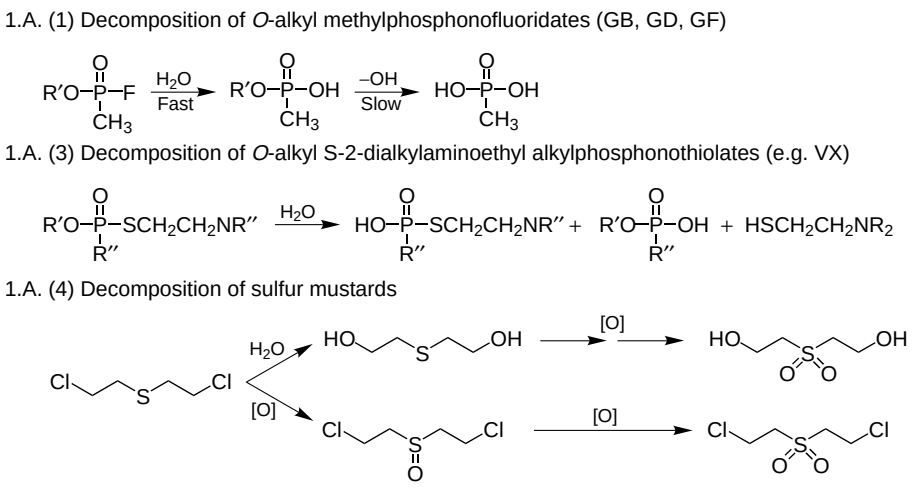

Figure 1 Major degradation pathways of Schedule 1 CW agents. GD, soman; GF, cyclohexyl sarin.

Chemicals'' describes the guidelines for inclusion of a chemical on each of the three schedules. The Annex also lists specific classes, toxic chemicals, and their precursors for which specific reporting and verification measures are applied under the Verification Annex of the CWC. Each schedule is divided into toxic chemicals and precursors. Schedule 1.A lists those chemicals which pose a high risk to the CWC and which have been developed, produced, stockpiled, or used as chemical weapons. In addition to the toxic chemicals, certain chemicals and classes of chemicals are also scheduled because they can be used as immediate precursors for the production of Schedule 1 chemicals. These chemicals are seen to have little or no use for purposes not prohibited by the CWC.

The principal classes and chemicals under Schedule 1 of the CWC include classical CW agents such as the *O*-alkyl alkylphosphonofluoridate nerve agents (e.g. GB, CAS 107-44-8), the *N*,*N*-dialkylphosphoramidocyanidates (e.g. tabun (GA), CAS 77-81-6), and the *S*-dialkylaminoethyl alkylphosphonothiolates (e.g. *O*-ethyl *S*-2-diisopropylaminoethyl-methyl phosphonothiolate (VX), CAS 50782-69-9). Three vesicant (blister agent) classes are scheduled: the sulfur mustards (e.g. mustard gas (HD), CAS 505-60-2), the lewisites (e.g. Lewisite 1, CAS 541-25-3); and the nitrogen mustards (e.g. ethyl nitrogen mustard (HN-1), CAS 538-07-8). Two special compounds are also included in the schedule for which CE analysis has been applied in particular, the paralytic shellfish toxin saxitoxin (STX, CAS 35523-89-8) and 1.A.(8) the glycoprotein ricin (CAS 9009-86-3).

3.2 Degradation Products of Chemical Warfare Agents

The degradation and decontamination chemistry of CW agents has been reviewed previously.[30,31] Degradation can occur from natural weathering, as a result of decontamination, or by deliberate attempts to destroy or neutralize CW agents as might occur in a clandestine production facility. Figure 1 shows the major degradation pathways of Schedule 1 chemicals which lead to water-soluble products which can be analyzed by CE. *O*-Alkyl alkylphosphonofluoridate nerve agents such as GB undergo a two-step degradation in water or under alkaline conditions. The initial degradation product is isopropyl methylphosphonate[32] followed by a slower hydrolysis to methylphosphonic acid. Hydrolysis of *S*-dialkylaminoethyl alkylphosphonothiolates such as VX leads to the formation of *S*-2-diisopropylaminoethyl methylphosphonothiolate and ethyl methylphosphonate. Sulfur mustards can undergo both hydrolytic and oxidative decomposition. The primary hydrolysis product is thiodiglycol. Thiodiglycol sulfoxide and sulfone can arise directly from the hydrolysis of the oxidation products of mustard or the oxidation of thiodiglycol.

4 CAPILLARY ELECTROPHORETIC ANALYSIS OF CHEMICAL WEAPONS CONVENTION RELATED COMPOUNDS

4.1 Chromatographic Analysis of Chemical Weapons Convention Related Compounds

Unfortunately, as organophosphorus nerve agents readily undergo hydrolysis, identification of the intact CW agent is not always possible and the detection or identification of degradation products including alkyl methylphosphonic acids and methylphosphonic acid is more likely. Chromatographic methods, including ion chromatography (IC),[33,34] have been developed

for the tentative identification of these hydrolysis products. However, unambiguous identification requires the acquisition of spectrometric data from two different spectrometric techniques.[35,36] Proton and phosphorus nuclear magnetic resonance (NMR) spectroscopy have been applied to the trace analysis of CW agents and related compounds in rubber, paint, and soil.[37] MS and NMR has been used for the unambiguous identification of novel CW decomposition products relevant to verification.[38]

Hyphenated MS techniques such as with gas chromatography (GC), HPLC, or CE are preferred for multicomponent environmental samples. Nerve agent hydrolysis products, including a number of alkyl methylphosphonic acids and methylphosphonic acid, have been analyzed by gas chromatography/mass spectrometry (GC/MS) (see below). The polarity of these products required derivatization of the hydroxyl groups prior to GC separation and a number of derivatizing agents have been employed to enhance volatility. GC/MS methods have been reported for alkyl methylphosphonic acids, methylphosphonic acid, and related degradation compounds following *tert*-butyldimethylsilylation,[39] methylation with trimethylphenylammonium hydroxide,[40] trimethylsilylation,[34] and derivatization with pentafluorobenzyl bromide.[41] Derivatization methods allow the analyst to use the same GC/MS instrumentation as used for the CW agent analysis but suffer from the need to modify the intact scheduled compound prior to identification.

HPLC coupled with MS offers the opportunity to analyze the polar, ionic organophosphorus nerve agent hydrolysis products without the need for derivatization. Thermospray MS and tandem mass spectrometry (MS/MS) have been demonstrated for the identification of organophosphorus nerve agents and their hydrolysis products.[42–44] Since the mid-1990s, this ionization technique has been largely superseded by atmospheric pressure ionization methods, including electrospray (ES) and ionspray (IS), both of which have been applied to the analysis of organophosphorus nerve agent hydrolysis products.[45,46] CE offers an important alternative chromatographic technique for the separation of target compounds prior to MS analysis.

4.2 Capillary Electrophoresis Analysis of Organophosphates

The primary degradation products of the alkylphosphonofluoridate CW agents [Schedule 1.A.(1)],[1] the alkyl alkyphosphonates, do not have a UV chromophore or fluorophore suitable for direct detection. The most common detection mode for these compounds is indirect UV detection using a high UV-absorbing background electrolyte (BGE). Analysis of low-molecular-weight organic ions, such as methylphosphonic acid, has also been described as capillary ion electrophoresis (CIE). In these CIE methods EOF modifiers are used to reverse the EOF. Under these conditions, the anionic organophosphates comigrate with the EOF, a condition referred to as co-EOF. The EOF is modified using run buffer additives or by modifying the capillary surface.[47] Table 1 lists the organophosphate compounds analyzed by CE, including all modes of detection.

4.2.1 Indirect Ultraviolet Detection

In indirect detection, the analyte displaces a UV-absorbing BGE in the separation buffer. Detection is due to a decrease in the UV signal as the analyte, which has displaced the BGE, passes through the detector window.[48]

Detection is in effect done in a high-noise system owing to the strongly absorbing BGE. Thus detection limits are generally poorer than for direct UV methods. As shown in Equation (4):

$$\mathrm{CLOD} = \frac{C_{\mathrm{BGE}}}{RDr} \tag{4}$$

the CLOD is directly related to the concentration of the BGE (C_{BGE}) and inversely proportional to the displacement ratio (R) of the analyte and BGE and the dynamic reserve (Dr). The dynamic reserve is the ability of the detector to detect a small change in signal on top of a large signal.[49] Optimization of detection limits involves careful selection of BGE and the lowest C_{BGE} should be employed. The detector must provide a large dynamic reserve and the displacement ratio should approach unity. The BGE should also have an electrophoretic mobility close to that of the analytes to maximize sensitivity and minimize dispersion effects that can arise from the electrophoretic mobility differences between the analyte and the BGE. [50]

CZE with indirect UV detection at 254 nm was first used for the detection of a series of alkylphosphonic acids.[51] Four BGEs were investigated, benzoic acid, phthalic acid, sorbic acid, and phenylphosphonic acid. Phenylphosphonic acid, which has a mobility close to that of the analyzed compounds, was shown to be the most suitable BGE. The influence of several parameters, such as the concentration of the BGE and the borate buffer on both sensitivity and efficiency, was investigated. An increase in the borate concentration produced an improvement of the signal-to-noise ratio (S/N) while the sensitivity decreased with increasing concentration of the phenylphosphonic acid. At pH 6, 200 mM sodium borate buffer containing 10 mM phenylphosphonic acid was selected and the separation of MPA, EPA, PPA, and BPA was demonstrated. The reproducibility of the μ_e for

Table 1 CWC-related organophosphorus compounds analyzed by CE

Structure	R_1	R_2	Name	CAS No.	Refs.
Alkylphosphonic acids [CH_3, CH_3CH_2, CH_3CH_2, $(CH_3)_2CH$]:					
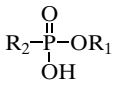	H	CH_3	MPA	363-06-7	51, 52, 54–57, 60, 76, 79
	H	CH_3CH_2	EPA	1782-26-4	51, 52, 55, 60, 76
	H	$CH_3CH_2CH_2$	nPPA	4672-38-2	51, 55, 56, 60
	H	$(CH_3)_2CH$	iPPA	4721-37-3	77
O-Alkyl (≤C_{10} including cycloalkyl) alkylphosphonates:					
	CH_3	CH_3	MMPA	1066-53-1	53, 55
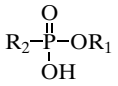	CH_3CH_2	CH_3	EMPA	1832-53-7	52–57, 60, 76
	$(CH_3)_2CH$	CH_3	IMPA	1832-54-8	52–57, 60, 76, 79
	Cyclohexyl	CH_3	CHMPA	1932-60-1	52, 53, 56, 60, 76
	$(CH_3)_2CHCH_2$	CH_3	iBMPA	1604-38-2	76
	Pinacolyl	CH_3	PMPA	616-52-4	52–58, 76, 79
	$(CH_3)_2CH(CH_3)CH$	CH_3	DMPA	151299-67-1	52, 53
	Cyclopentyl	CH_3	CPMPA	73207-99-5	56
	2-Ethylhexyl	CH_3	EHMPA	13688-82-9	76
	CH_3CH_2	CH_3CH_2	EEPA	7305-61-5	52, 53
	$(CH_3)_2CH$	CH_3CH_2	iPEPA	170135-50-9	53, 76
	$(CH_3)_2CH(CH_3)CH$	CH_3CH_2	DEPA	195158-11-3	53
	Cyclohexyl	CH_3CH_2	CEPA	170424-87-0	52, 53
Methylphosphonothioic acids:					
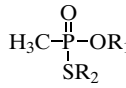	CH_3CH_2	H	EMPSA	18005-40-8	56, 57, 60, 61
	$(CH_3)_2CHCH_2$	H	IBMPSA	20626-99-7	60, 61
	H	$(iPr)_2CH_2CH_2N$	EA 2192	73207-98-4	60

MPA was 1.65% (relative standard deviation) and the LOD was <10 μM (0.21 pmol injected).

Sorbic acid has also been shown to be a suitable BGE for the analysis of the alkylphosphonic acids and their monoester derivatives.[52] A run buffer consisting of 5 mM sorbic acid and 0.1 M decamethonium bromide, adjusted to pH 6, was used to separate a mixture of eight alkyl alkylphosphonates. The use of an EOF modifier, decamethonium bromide, reduced the EOF at pH 6 and gave complete separation of the alkylphosphonic acids and their monoester derivatives (e.g. MPA and IMPA) in <15 min (Figure 2).

To reduce run times while maintaining high peak efficiencies, the capillary can be dynamically coated with a cationic polymer to reverse the EOF. In the procedure the column is flushed with a polymer solution prior to each analysis. This approach has been demonstrated with a 0.2% aqueous solution of polybren.[53] Sorbic acid was used as the BGE and hydroxylamine was added to improve the buffer capacity of the run buffer. Using an electrolyte composed of 5 mM sorbate anion (pH 6) and 1.6 mM hydroxylamine allowed the separation of alkylphosphonic acid monoesters in <6 min. Chromate

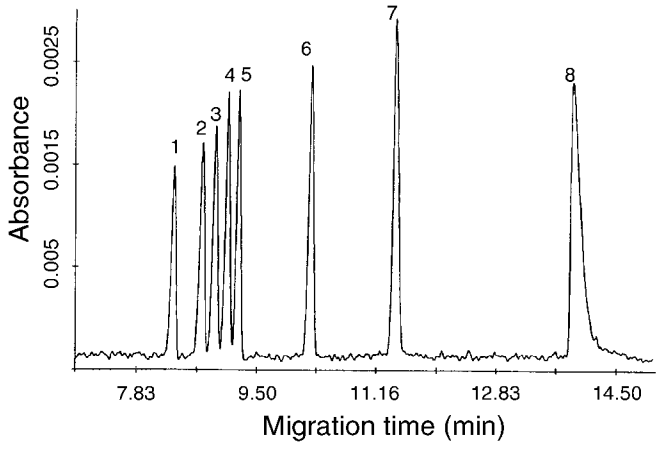

Figure 2 CZE of a standard mixture (solute concentration 20 mg L^{-1}) of eight alkylphosphonic acid anions: CEPA (1), PMPA (2), CMPA (3), DMPA (4), EEPA (5), IMPA (6), EMPA (7), and MPA (8). CZE analysis conditions: 67 cm L_t (60 cm L_d) × 50 μm i.d.; V = 30 kV; 5 mM sorbic acid, 0.1 mM decamethonium bromide, pH 6; indirect UV detection at 254 nm; 5 s hydrodynamic injection. (Reprinted from J.-P. Mercier, Ph. Morin, M. Dreux, A. Tambute, *J. Chromatogr. A*, **741**, 279–285 (1996), Copyright © 1996, with permission from Elsevier Science.)

has also been used as a BGE in the presence of a proprietary EOF modifier for the separation of MPA, EMPA, IMPA, and PMPA.[54]

4.2.2 Direct Ultraviolet Detection

The direct UV detection of alkylphosphonic acids through the formation of sodium borate complexes has been reported.[55] The buffer system consisted of 10 mM sodium tetraborate with complexation occurring in solution at pH 7 and higher. The reported detection limits were of the order of nanograms. The UV, Infrared (IR), NMR, and mass spectra of various borate–phosphonic acid esters were also reported.

4.2.3 Flame Photometric Detection

The separation performance and sensitivity of a CE system coupled on-line with a flame photometric detector (FPD), originally developed for GC, have been reported for the detection of alkylphosphonic acids.[56] The liquid junction used to decouple the electric field from the FPD showed a negligible influence on the performance of the system as compared with online UV detection. Use of an on-column sample stacking pre-concentration technique allowed for injection of 900 nL. With the large injection, the detection limits for the alkylphosphonic acids in water were 0.1–0.5 $\mu g\,mL^{-1}$.

4.2.4 Mass Spectrometric Detection

MS is the only technique which to date has been coupled with CE to provide the structural information required for unambiguous identification of CW degradation products. Capillary zone electrophoresis/ionspray mass spectrometry (CZE/ISMS) combines the high separation efficiency of CE and the high sensitivity of MS for the analysis of these polar degradation products. CZE/ISMS, in the negative-ion mode, was applied to the identification of five organophosphonic acids which are the primary hydrolysis products of nerve agents.[57] The spectra of EMPA, IMPA, and PMPA exhibited very abundant $[M - H]^-$ ions of m/z 123, 137, and 179, respectively. The fragmentation was minimal and common fragmentation ions at m/z 95 and 79 were observed due to the loss of the alkyl and alkoxy groups, respectively. Sensitivity in the range 10–30 pg (5 $ng\,\mu L^{-1}$) was achieved using selected ion monitoring (SIM).

4.3 Sulfur-containing Compounds

4.3.1 Micellar Electrokinetic Chromatography

MEKC is a combination of electrophoresis and chromatography.[58,59] The main advantage of MEKC is that it can be used for the separation of both neutral and charged solutes. This is done by adding a surfactant such as sodium dodecylsulfate (SDS) at concentrations above the critical micelle concentration to the running buffer to form micelles. Neutral species partition in and out of the micelle. Since the micelles are charged, this partitioning imparts a mobility to the neutral species. The more hydrophobic the analyte, the more it is retained within the micelle and the more it is affected by the charge of the micelle. Since MEKC can be used for both charged and uncharged species, its applications include analyses of amino acids, nucleotides, vitamins, pharmaceuticals, aromatic hydrocarbons, and explosive constituents.

The sulfur-containing, neutral degradation products of mustard and related vesicants can be analyzed by direct UV detection using MEKC. MEKC has been used for the analysis of thiodiglycol (TDG), 2,2′-sulfinyldiethanol (TDGO), 1,4-dithiane, 1,4-thioxane, *O*-isobutyl methylphosphonothioic acid (IBMPSA) and *O*-ethyl methylphosphonothioic acid (EMPSA). These were separated in under 10 min with a run buffer of 10 mM borate and 100 mM SDS at pH 9 (Figure 3).[60,61] Detection limits range from 1 to 10 $\mu g\,mL^{-1}$. Sulfur-containing compounds related to the decomposition of sulfur mustard and other materials analyzed by CE with direct UV detection are listed in Table 2.

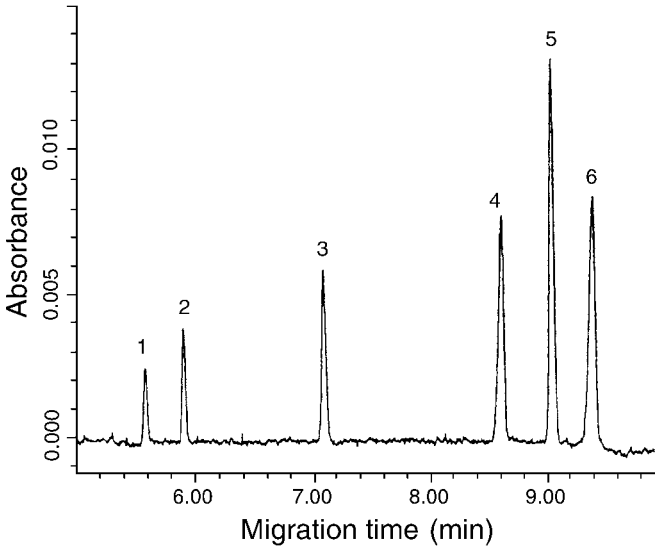

Figure 3 MEKC of six sulfur-containing compounds related to the degradation of mustard and *O*-alkyl alkylphosphonothiolates (solute concentration 50 $mg\,L^{-1}$): TDGO (1), TDG (2), thioxane (3), iBMPSA (4), dithiane (5), and EMPSA (6). MEKC analysis conditions: 57 cm L_t (50 cm L_d) × 75 μm i.d.; $V = 25$ kV; 10 mM borate, 100 mM SDS, pH 9; direct UV detection at 200 nm; 3 s hydrodynamic injection. (Reprinted from J.-P. Mercier, Ph. Morin, M. Dreux, A. Tambute, *J. Chromatogr. A*, **741**, 279–285 (1996), Copyright 1996, with permission from Elsevier Science.)

Table 2 CWC-related degradation products of sulfur mustard, lewisite and other compounds analyzed by CE

Structure	Name	CAS No.	Refs.
Sulfur mustard-related compounds:			
$HOCH_2CH_2SCH_2CH_2OH$	TDG	111-48-8	60, 61
$HOCH_2CH_2S(O)CH_2CH_2OH$	TDGO	3085-45-8	60, 61
$HOCH_2CH_2S(O)_2CH_2CH_2OH$	$TDGO_2$	2580-77-0	60
	1,4-Thioxane	15980-15-1	60, 61
	1,4-Dithiane	505-29-3	60, 61
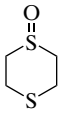	1,4-Dithiane-1-oxide	19087-70-8	60
Lewisite-related compounds:			
ClCHCHAs(O)	CVAOA	3088-37-7	60
Schedule 2B precursors			
$(CH_3CH_2)_2NCH_2CH_2OH$	DEAE	100-37-8	60
$[(CH_3)_2CH]_2NCH_2CH_2OH$	DIPAE	96-80-0	60
$[(CH_3)_2CH]_2NCH_2CH_2SH$	DIPAESA	1439-97-5	60

4.4 Toxins

4.4.1 Saxitoxin

STX, Schedule 1.A.(7), is a potent neurotoxin which can cause paralytic shellfish poisoning (PSP). STX is produced by certain strains of dinoflagellates, which leads to the contamination of commercial shellfish and causes severe outbreaks of seafood poisoning. The public health problems caused by these outbreaks have led to significant interest in the development of analytical methods for the determination of STX in environmental and biological samples. STX is also one of several closely related compounds and separation and so unambiguous identification of the target compound is a significant analytical requirement for its analysis.

STX is one of the only CWC-related compounds which has been analyzed by CZE with LIF detection. Peptides and proteins containing lysine residues or natural products which contain primary amino groups (NH_2) can be derivatized with a number of fluorescent labeling reagents.[62] LIF is an extremely sensitive detection method and can provide CLODs in the 10^{-11}–10^{-13} M range.[63] STX has been labeled with fluorescamine, *o*-phthaldialdehyde (OPA) and dansyl chloride and detection limits as low as 0.1 amol were reported for the OPA derivative of STX.[64] However, the precolumn labeling methods required micromolar concentrations of analyte, limiting the utility in trace analysis. Labeling, separation, and analysis of STX were best accomplished using the fluorescamine derivative, which produces ionic derivatives that can be separated from other fluorescently labeled marine toxins such as tetrodotoxin and microcystin.

CE analysis with direct UV absorbance detection at 200 nm has been described for the separation and detection of underivatized toxins, including STX, associated with PSP.[65] Confirmation of the electrophoretic peaks was made by capillary electrophoresis/electrospray ionization mass spectrometry (CE/ESIMS). STX and neosaxitoxin (NEO) were separated using 20 mM sodium citrate buffer (pH 2.1) with a mass detection limit of 15 pg (5 μM) for STX. This method was able to detect STX and NEO algal and mollusc extracts. Another reported UV method uses 208 nm and a 0.06 M sodium borate buffer for the analysis of STX in molluscs.[66]

The application of on-column sample preconcentration CITP and a discontinuous buffer system prior to CE separation has been investigated for the analysis of PSP toxins. Using a linear polyacrylamide-coated capillary to eliminate EOF during the CITP step, 35 mM morpholine (pH 5) as the leading electrolyte and 10 mM formic acid as the terminating electrolyte, the improvement of the CLOD is at least two orders of magnitude over that obtainable using the conventional CZE format.[67] Analyses performed using SIM ESIMS provided a CLOD of 16 nM for STX. This preconcentration technique was compatible with ESIMS and permitted the analysis of biological extracts containing PSP toxin levels as low as

100 nM. In situations where higher levels of selectivity are required for unambiguous identification of individual PSP toxins, CITP or CZE with MS/MS can be used. Further improvements in sample focusing prior to CE separation can be obtained using high-ionic-strength buffers and on-column sample stacking.[68]

4.4.2 Ricin

CE has found particular application in the analysis of peptides[69,70] and proteins.[71,72] The resolution and ease of operation combined with the various CE modes such as CGE have led to a tremendous number of applications for CE analysis of amino acids, peptides, and proteins. Proteins are readily analyzed by direct UV detection at 200 nm. CE analysis of proteins can present additional challenges such as sample microheterogeneity due to variations in carbohydrate structure, impurities from the isolation or preparation of the protein sample, or surface adhesion to the capillary walls. This can result in multiple peaks or significant peak broadening, peak tailing, and even irreversible loss of protein.

Ricin is a heterodimeric glycoprotein (MW = 66 000) produced by the castor bean plant *Ricinus communis* with the two protein subunits linked by a single disulfide bond. Conditions for the analysis of ricin with CE, using uncoated and coated columns, and the influence of buffer systems (run buffer composition, pH, ionic strength, and additives) have been investigated. Uncoated columns used with either zwitterionic salts or putrescine gave the best results.[73,74] Multiple peaks were resolved with these conditions. Analysis using a bare fused-silica column with a 0.03 M sodium borate run buffer gave a single, broad, tailing peak which could be due to protein–column interactions, additional components, or both. The use of 0.03 M putrescine (1,4-diaminobutane) in 0.02 M K_2SO_4 buffer (pH 7) gave better resolution and peak shape and multiple peaks were resolved, indicating possible impurities or microheterogeneity in the ricin sample.

5 APPLICATIONS

5.1 Environmental Analysis

The development of CE analytical methods for CW-related compounds has been directed primarily towards the analysis of these compounds in environmental matrices such as soil and water. These methods have been used in the analysis of samples distributed as part of interlaboratory comparisons exercises, the so-called Finnish Round Robin Exercises[38] and more recently the Organization for the Prohibition of Chemical Weapons (OPCW) Laboratory Proficiency Testing.[75]

The use of CE with indirect UV detection in environmental samples leads to matrix effects from inorganic ions that comigrate. Surface or groundwater samples can contain carbonate levels that interfere with alkylphosphonate analysis. Ion-exchange cartridges such as the Milli-Trap H^+ (Waters) can be used to reduce carbonate levels and Ag^+ resins can be used to remove chloride ions.[59] Using decamethonium bromide to reduce the EOF, CEPA- and EPA-spiked soil samples, extracted with water, could be analyzed without significant interferences.[57] However, differences in the migration times were noted owing to matrix effects in the soil extracts and required the use of a standard to achieve quantitative results. Environmental water samples, spiked with standard mustard degradation products, have been analyzed by MEKC with direct UV detection with no apparent interferences or requirement for sample pretreatment.[66]

As part of an Interlaboratory Proficiency Test, CE with FPD was used to screen water samples and aqueous extracts of soil, rubber, and paint for spiked alkylphosphonic acids.[61] The water sample had been spiked with PPA at a level of $8\,\mu g\,mL^{-1}$ and one soil sample was spiked with CEPA and EPA at 15 and $8\,\mu g\,g^{-1}$, respectively. Although the recovery efficiency was only approximately 20%, CEPA and EPA could still be detected using longer injection volumes and higher on-column concentrations. Subsequent NMR and *tert*-butyldimethylsilyl derivatization and GC/MS analysis confirmed the identity of the phosphorus-containing compounds in the extract.

Reversal of EOF was used in the analysis of samples in the second and third OPCW Laboratory Proficiency Tests.[76] The cationic surfactants didodecyldimethylammonium hydroxide and cetyltrimethylammonium hydroxide were used to reverse the EOF and remove interferences from common anions and give complete and rapid separation of alkylphosphonic acids. EPA and PEPA were detected in water samples as part of the second OPCW test and PPA and EHMPA were detected in the water samples as part of the third OPCW test. Identification of spiked compounds was first made on the basis of mobility data and confirmed by MS. The method was applied to aqueous leachates of soil, wipes of surfaces, and vegetation sampled from a field site known to have been exposed to nerve agents and subsequently remediated. MPA, IMPA, and PMPA were detected in these samples, indicating that the method can be useful for environmental monitoring.

5.2 Quality Assurance

Under the terms of the CWC, the Director General of the OPCW shall establish stringent regimes governing the collection, handling, transport, and analysis of

Table 3 Representative CE methods for the analysis of CW degradation products

Class	Mode	Buffer	pH	Detection : λ (nm)	E ($V\,cm^{-1}$)	Ref.
Alkylphosphonates	CZE	200 mM borate 10 mM phenylphosphonic acid[a]	6	Indirect UV (254)	465	51
Alkylphosphonates	CZE	5 mM sorbic acid[a] 0.1 mM decamethonium bromide	6	Indirect UV (254)	448	52
Alkylphosphonates	CZE	200 mM boric acid 10 mM phenylphosphonic acid[a] 0.03 wt% Triton X 0.35% didodecyldimethylammonium (DDAOH)	4	Indirect UV (210)	536	79
Alkylphosphonates	CZE	20 mM ammonium acetate	9	ISMS	375	57
Methylphosphonothioates	CZE	250 mM borate	7	Direct UV (200)	470	60
S-Mustard-related compounds	MEKC	10 mM borate 100 mM SDS	9	Direct UV (200)	440	61

[a] BGE.

samples. The Director General shall also certify laboratories designated for the analysis of samples, oversee standardization of procedures, and monitor quality control standards at certified laboratories. To be certified as a ''Designated Laboratory'' by the OPCW under the terms of the CWC, a laboratory must have established an internationally recognized quality assurance system, obtained accreditation by an internationally recognized accreditation body for which they are seeking accreditation, and regularly participate and perform successfully interlaboratory proficiency tests.[77] A requirement for accreditation of testing laboratories under such standards as the ''General Criteria for the Operation of Testing Laboratories'', European Standard 45 001: European Standards Institution, Brussels EN 45 001, is that the laboratory follow written standard operating procedures and implement a quality assurance/quality control program.

In addition to demonstrating the role of analytical methods in verification of the CWC, the round-robin methods have led to the publication of recommended operating procedures which can be used for the analysis of the target compounds. The published method of Pianetti et al.[51] has been put forward as a recommended operating procedure for the analysis of the alkylphosphonates.[78] Other representative methods for the analysis of various CWC-related compounds which could form the basis of a standard or recommended operating procedure in a designated laboratory are summarized in Table 3.

A validated quantitative analysis for CW degradation products has been developed.[79]The specificity, linearity, precision, and accuracy were examined for the analysis of IMPA and PMPA using two buffer systems and detection methods. The method gave inter- and intra-day reproducibility of 1–3% (relative standard deviation). The method requires little sample preparation except for dilution with deionized water and is not prone to interference from carbonate or fluoride ions.

6 CONCLUSION

The ability of CE to analyze anionic, cationic, zwitterionic, and neutral CWC-related compounds, whether they be CW degradation products or scheduled starting materials, without the requirement for elaborate sample processing or labeling procedures is an important demonstration of the utility of CE for the analysis of CWC-related compounds. CE can be an important screening method for rapid sample processing and further analysis for unambiguous identification. The relative ease of analysis should also increase its utility in field and on-site analysis where minimizing the logistic burden is attractive. The increased use of MS coupled techniques and the advent of CE coupled with NMR[80] will also serve to establish CE as an important method for the unambiguous identification of CWC compounds.

ACKNOWLEDGMENTS

The author would like to thank Mr Robert Poirier, Dr Paul A. D'Agostino, Mr James Hancock, and Dr Charles A. Lucy for their assistance and support in preparation of the manuscript.

ABBREVIATIONS AND ACRONYMS

BGE	Background Electrolyte
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CE/ESIMS	Capillary Electrophoresis/Electrospray Ionization Mass Spectrometry
CGE	Capillary Gel Electrophoresis
CIE	Capillary Ion Electrophoresis
CIEF	Capillary Isoelectric Focusing
CITP	Capillary Isotachophoresis
CLOD	Concentration Limit of Detection
CW	Chemical Warfare
CWC	Chemical Weapons Convention
CZE	Capillary Zone Electrophoresis
CZE/ISMS	Capillary Zone Electrophoresis/ Ionspray Mass Spectrometry
EOF	Electroosmotic Flow
ES	Electrospray
ESIMS	Electrospray Ionization Mass Spectrometry
FPD	Flame Photometric Detector
GA	Tabun
GB	Sarin
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GD	Soman
GF	Cyclohexyl Sarin
HD	Mustard Gas
HN-1	Ethyl Nitrogen Mustard
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
IR	Infrared
IS	Ionspray
LIF	Laser-induced Fluorescence
LOD	Limit of Detection
MEKC	Micellar Electrokinetic Chromatography
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NEO	Neosaxitoxin
NMR	Nuclear Magnetic Resonance
OPA	*o*-Phthaldialdehyde
OPCW	Organization for the Prohibition of Chemical Weapons
PSP	Paralytic Shellfish Poisoning
SDS	Sodium Dodecylsulfate
SIM	Selected Ion Monitoring
S/N	Signal-to-noise Ratio
STX	Saxitoxin
UV	Ultraviolet
VX	*O*-Ethyl *S*-2-diisopropylaminoethyl-methyl Phosphonothiolate

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Chromatography of Membrane Proteins and Lipoproteins

Pesticides **(Volume 7)**
Organophosphorus Pesticides in Water and Food, Analysis of

Electronic Absorption and Luminescence **(Volume 12)**
Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis

REFERENCES

1. J.W. Jorgenson, K.D. Lukacs, 'Capillary Zone Electrophoresis', *Science*, **222**, 266–272 (1983).
2. M. Albin, P.D. Grossman, S.E. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 489A–497A (1993).
3. H.-J. Gaus, A.G. Beck-Sickinger, E. Bayert, 'Optimization of Capillary Electrophoresis of Mixtures with Basic Peptides and Comparison with HPLC', *Anal. Chem.*, **65**, 1399–1405 (1993).
4. Y. Xu, 'Capillary Electrophoresis', *Anal. Chem.*, **65**, 425R–433R (1993).
5. N.A. Guzman, L. Hernandez, S. Terabe, 'High Resolution Nanotechnique for Separation, Characterization and Quantitation of Micro and Macro Molecules', *ACS Symp. Ser.*, **434**, 1–35 (1990).
6. *Convention on the Prohibition of the Development, Production, Stockpiling, and Use of Chemical Weapons and Their Destruction*, signed January 1993 and which entered into force April 29, 1997, following the 65th State Party Ratification on October 31, 1996. Printed and Distributed by the Organization for the Prohibition of Chemical Weapons, The Hague. The Depository of this Convention is the Secretary-General of United Nations, from which a certified true copy can be obtained.
7. E. Croddy, 'Putting the Lid Back on the Chemical Box', *Jane's Intell. Rev.*, 41–45 (1998).

8. P.A. D'Agostino, L.R. Provost, 'Capillary Column Isobutane Chemical Ionization Mass Spectrometry of Mustard and Related Compounds', *Biomed. Environ. Mass Spectrom.*, **15**, 553–564 (1988).
9. R.M. Black, R.J. Clark, R.W. Read, M.T.J. Reid, 'Application of Gas Chromatography–Mass Spectrometry to the Analysis of Chemical Warfare Samples Found to Contain Residues of the Nerve Agent Sarin, Sulfur Mustard and Their Degradation Products', *J. Chromatogr. A.*, **662**, 301–321 (1994).
10. M. Minami, D.-M. Hui, M. Katsumata, H. Inagaki, C.A. Boulet, 'Method for the Analysis of the Methylphosphonic Acid Metabolites of Sarin and its Ethanol-substituted Analog in Urine as Applied to the Victims of the Tokyo Sarin Disaster', *J. Chromatogr. B.*, **695**, 237–244 (1997).
11. M. Polhuijs, J.P. Langenberg, H.P. Benschop, 'New Method for Retrospective Detection of Exposure to Organophosphorus Anticholinesterases: Application to Alleged Sarin Victims of Japanese Terrorists', *Toxicol. Appl. Pharmacol.*, **145**, 156–161 (1997).
12. E. Croddy, 'Urban Terrorism – Chemical Warfare in Japan', *Jane's Intell. Rev.*, 520–523 (1995).
13. W.G. Kuhr, C.A. Monnig, 'Capillary Electrophoresis', *Anal. Chem.*, **64**, 389R–407R (1992).
14. C.A. Monnig, R.T. Kennedy, 'Capillary Electrophoresis', *Anal. Chem.*, **66**, 280R–314R (1994).
15. R.L. St.Claire III, 'Capillary Electrophoresis', *Anal. Chem.*, **68**, 569R–586R (1996).
16. S.C. Beale, 'Capillary Electrophoresis', *Anal. Chem.*, **70**, 279R–300R (1998).
17. J.P. Landers, (ed.), *Handbook of Capillary Electrophoresis*, 2nd edition, CRC Press, Boca Raton, FL, 1997.
18. P. Camilleri, (ed.), *Capillary Electrophoresis: Theory and Practice*, CRC Press, Boca Raton, FL, 1993.
19. N.J. Dovichi, 'General Instrumentation and Detection Systems', in *Capillary Electrophoresis: Theory and Practice*, ed. P. Camilleri, CRC Press, Boca Raton, FL, 25–64, 1993.
20. M.A. Hayes, I. Kheterpal, A.G. Ewing, 'Electroosmotic Flow and Surface Conductance in Capillary Zone Electrophoresis', *Anal. Chem.*, **65**, 2010–2013 (1993).
21. M. Huang, G. Yi, J.S. Bradshaw, M.L. Lee, 'Charged Surface Coatings for Capillary Electrophoresis', *J. Microcol. Sep.*, **5**, 199–205 (1993).
22. N. Cohen, E. Grushka, 'Controlling Electroosmotic Flow in Capillary Zone Electrophoresis', *J. Chromatogr. A.*, **678**, 167–175 (1994).
23. Y.K. Zhang, N. Chen, L. Wang, 'The Effect of Column Temperatures on the Migration Time of Peptides in Free-solution Capillary Electrophoresis', *J. Liq. Chromatogr.*, **16**, 3689–3697 (1993).
24. N. Chen, L. Wang, Y. Zhang, 'Influence of Column Temperature and Physico-chemical Properties on the Electrophoretic Behavior of Polyglycine Peptides in Free-solution Capillary Electrophoresis', *J. Chromatogr.*, **644**, 175–182 (1993).
25. T.L. Lee, R. Dadoo, R.N. Zare, ' Real-time Measurement of Electroosmotic Flow in Capillary Electrophoresis', *Anal. Chem.*, **66**, 2694–2700 (1994).
26. D.N. Heiger, *High Performance Capillary Electrophoresis – An Introduction*, Hewlett-Packard, Waldbronn, 1997.
27. H.T. Rasmussen, H.M. McNair, 'Influence of Buffer Concentration, Capillary Internal Diameter and Forced Convection on Resolution in Capillary Zone Electrophoresis', *J. Chromatogr.*, **516**, 223–231 (1990).
28. M. Albin, P.D. Grossman, S.E. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 489A–497A (1993).
29. A.J. Tomlinson, N.A.Guzman, S. Naylor, 'Enhancement of Concentration Limits of Detection in CE and CE–MS: a Review of On-line Sample Extraction, Cleanup, Analyte Preconcentration, and Microreactor Technology', *J. Capillary Electrophoresis*, **6**, 247–266 (1995).
30. Y.-C. Yang, J.A. Baker, J.R. Ward, 'Decontamination of Chemical Warfare Agents', *Chem. Rev.*, **92**, 1729–1743 (1992).
31. Y.-C. Yang, 'Chemical Reactions for Neutralising Chemical Warfare Agents', *Chem. Ind. (London)*, 334–337 (1995).
32. J. Epstein, 'Rate of Decomposition of GB in Seawater', *Science*, **170**, 1396–1398 (1970).
33. P.C. Bossle, D.J. Reutter, E.S. Sarver, 'Analysis of Alkyl Methylphosphonic Acids in Aqueous Matrices by Ion-pair Reversed-phase Ion Chromatography', *J.Chromatogr.*, **407**, 399–404 (1987).
34. A.F. Kingery, H.E. Allen, 'Ion Chromatographic Separation of Closely Related Nerve Agent Degradation Products Using an Organic Modifier to Provide Selectivity', *Anal. Chem.*, **66**, 155–159 (1994).
35. M.T. Soderstrom, H. Bjork, V.M.A. Hakkinen, O. Kostiainen, M.-L. Kuitunen, M. Rautio, 'Identification of Compounds Relevant to the Chemical Weapons Convention Using Selective Gas Chromatography Detectors, Gas Chromatography–Mass Spectrometry and Gas Chromatography–Fourier Transform Infrared Spectroscopy in an International Trial Proficiency Test', *J. Chromatogr. A.*, **742**, 191–203 (1996).
36. (a) *International Inter-laboratory Comparison (Round-robin) Test for the Verification of Chemical Disarmament. F.1. Testing of Existing Procedures*, Ministry of Foreign Affairs of Finland, Helsinki, 1990, ISBN 951-47-3872-1; (b) *International Inter-laboratory Comparison (Round-robin) Test for the Verification of Chemical Disarmament. F.2. Testing of Procedures on Simulated Industry Samples*, Ministry of Foreign Affairs of Finland, Helsinki, 1991, ISBN 951-47-4957-X; (c) *International Inter-laboratory Comparison (Round-robin) Test for the Verification of Chemical Disarmament. F.3. Testing of Procedures on Simulated Military Facility Samples*, Ministry of Foreign Affairs of Finland, Helsinki, 1992, ISBN 951-47-6272-X; (d) International Inter-laboratory Comparison (Round-robin) Test for the Verification of Chemical

Disarmament. F.4. Validating of Procedures for Water and Soil Samples, Ministry of Foreign Affairs of Finland, Helsinki, 1994, ISBN 951-47-8163-5.
37. M. Mesilaakso, E.-L. Tolppa, 'Detection of Trace Amounts of Chemical Warfare Agents and Related Compounds in Rubber, Paint, and Soil Samples by ^{1}H and $^{31}P\{^{1}H\}$ NMR Spectroscopy', *Anal. Chem.*, **68**, 2313–2318 (1996).
38. C.A. Boulet, P.A. D'Agostino, 'Analysis of Dimethylpyrophosphonate Decomposition Products of VX by GC/MS/MS and ^{31}P NMR', *Phosphorus Sulfur Silicon*, **104**, 93–101 (1995).
39. J.G. Purdon, J.G. Pagotto, R.K. Miller, 'Preparation, Stability, and Quantitative Analysis by Gas Chromatography and Gas Chromatography–Electron Impact Mass Spectrometry of *tert*-Butyldimethylsilyl Derivatives of some Alkylphosphonic and Alkyl Methylphosphonic Acids', *J. Chromatogr.*, **475**, 261–272 (1989).
40. J.Aa. Tornes, B.A. Johnsen, 'Gas Chromatographic Determination of Methylphosphonic Acids by Methylation with Trimethylphenylammonium Hydroxide', *J. Chromatogr.*, **467**, 129–138 (1989).
41. S.-A. Fredriksson, L.-G. Hammarstroem, L. Henriksson, H.-A. Lakso, 'Trace Determination of Alkyl Methylphosphonates in Environmental and Biological Samples Using Gas Chromatography/Negative-ion Chemical Ionization Mass Spectrometry and Tandem Mass Spectrometry', *J. Mass Spectrom.*, **30**, 1133–1143 (1995).
42. E.R.J. Wils, A.G. Hulst, 'Determination of Organophosphorus Acids by Thermospray Liquid Chromatography/Mass Spectrometry', *J. Chromatogr.*, **454**, 261–272 (1988).
43. E.R.J. Wils, A.G. Hulst, ' The Use of Thermospray–Liquid Chromatography/Mass Spectrometry for the Verification of Chemical Warfare Agents', *Fresenius' J. Anal. Chem.*, **342**, 749–758 (1992).
44. J.A. Tornes, 'Identification of Some Alkyl Methylphosphonic Acids by Thermospray Tandem Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 878–882 (1996).
45. V.T. Borrett, R. Colton, J.C. Traeger, 'The Electrospray Mass Spectra of Phosphonic Acid, Methylphosphonic Acid and Its Alkyl Esters, and Their Complexes with Alkali and Alkali Earth Metal Ions', *Eur. Mass Spectrom.*, **1**, 131–140 (1995).
46. V.T. Borrett, R.J. Matthews, R. Colton, J.C. Traeger, 'Verification of the United Nations Chemical Weapons Convention: the Application of Electrospray Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 114–118 (1996).
47. W.R. Jones, 'Capillary Ion Electrophoresis', in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 155–187, 1997.
48. E.S. Yeung, W.G. Kuhr, 'Indirect Detection Methods for Capillary Electrophoresis', *Anal. Chem.*, **63**, 275A–282A (1991).
49. Y. Mah, R. Zhang, 'Optimization of Indirect Photometric Detection of Anions in High-performance Capillary Electrophoresis', *J. Chromatogr.*, **625**, 341–348 (1992).
50. F. Foret, S. Fanali, L. Ossicini, P. Bocek, 'Indirect Photometric Detection in Capillary Zone Electrophoresis', *J. Chromatogr.*, **470**, 299–308 (1989).
51. G.A. Pianetti, M. Taverna, A. Baillet, G. Mahuzier, D. Baylocq-Ferrier, 'Determination of Alkylphosphonic Acids by Capillary Zone Electrophoresis Using Indirect UV Detection', *J. Chromatogr.*, **630**, 371–377 (1992).
52. J.-P. Mercier, Ph. Morin, M. Dreux, A. Tambute, 'Capillary Electrophoresis Analysis of Chemical Warfare Agent Breakdown Products. I. Counterelectroosmotic Separation of Alkylphosphonic Acids and Their Monoester Derivatives', *J. Chromatogr. A.*, **741**, 279–285 (1996).
53. J.-P. Mercier, Ph. Morin, M. Dreux, A. Tambute, 'Capillary Electrophoresis Separation of Alkylphosphonic Acid Monoesters with Indirect Ultraviolet Detection', *J. Chromatogr. A.*, **779**, 245–252 (1997).
54. S.A. Oehrle, P.C. Bossle, 'Analysis of Nerve Agent Degradation Products Using Capillary Ion Electrophoresis', *J. Chromatogr. A.*, **692**, 247–252 (1995).
55. W.H. Robins, B.W. Wright, 'Capillary Electrophoretic Separation of Organophosphonic Acids using Borate Esterification and Direct UV Detection', *J. Chromatogr. A.*, **680**, 667–673 (1994).
56. C.E. Kientz, E.W. Hooijschuur, U.A.Th. Brinkman, 'Capillary Electrophoresis Coupled Online with Flame Photometric Detection: Determination of Alkylphosphonic Acids.', *J. Microcol. Sep.*, **9**, 253–259 (1997).
57. R. Kostiainen, A.P. Bruins, V.M.A. Hakkinen, 'Identification of Degradation Products of Some Chemical Warfare Agents by Capillary Electrophoresis/Ionspray Mass Spectrometry', *J. Chromatogr.*, **634**, 113–118 (1993).
58. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open-tubular Capillaries', *Anal. Chem.*, **56**, 111–113 (1984).
59. S. Terabe, K. Otsuka, T. Ando, 'Electrokinetic Chromatography with Micellar Solution and Open-tubular Capillary', *Anal. Chem.*, **57**, 834–841 (1985).
60. R.L. Cheicante, J.R. Stuff, H.D. Durst, 'Analysis of Chemical Weapons Degradation Products by Capillary Electrophoresis with UV Detection', *J. Capillary Electrophoresis*, **2**, 157–163 (1995).
61. R.L. Cheicante, J.R. Stuff, H.D. Durst, 'Separation of Sulfur Containing Chemical Warfare Related Compounds in Aqueous Samples by Micellar Electrokinetic Chromatography', *J. Chromatogr. A.*, **711**, 347–352 (1995).
62. M.E. Szulc, I.S. Krull, 'Improved Detection and Derivatization in Capillary Electrophoresis', *J.Chromatogr. A.*, **659**, 231–245 (1994).
63. L. Hernandez, N. Joshi, P. Verdeguer, N.A. Guzman, 'Laser-induced Fluorescence Detection for Capillary

Electrophoresis: a Powerful Analytical Tool for the Separation and Detection of Trace Amounts of Analytes', *Chromatogr. Sci. Ser. (Capillary Electrophoresis Technol.)*, **64**, 605–614 (1993).

64. B.W. Wright, G.A. Ross, R.D. Smith, 'Capillary Zone Electrophoresis with Laser Fluorescence Detection of Marine Toxins', *J. Microcol. Sep.*, **1**, 85–89 (1989).
65. P. Thibault, S. Pleasance, M.V. Laycock, 'Analysis of Paralytic Shellfish Poisons by Capillary Electrophoresis', *J. Chromatogr.*, **542**, 483–501 (1991).
66. C.J. Sciacchitano, B. Mopper, 'Analysis of Paralytic Shellfish Toxin (Saxitoxin) in Mollusks by Capillary Zone Electrophoresis', *J. Liq. Chromatogr.*, **16**, 2081–2088 (1993).
67. S.J. Locke, P. Thibault, 'Improvement in Detection Limits for the Determination of Paralytic Shellfish Poisoning Toxins in Shellfish Tissues Using Capillary Electrophoresis/Electrospray Mass Spectrometry and Discontinuous Buffer Systems', *Anal. Chem.*, **66**, 3436–3446 (1994).
68. A. Buzy, P.Thibault, M.V. Laycock, ' Development of a Capillary Electrophoresis Method for the Characterization of Enzymatic Products Arising from the Carbamoylase Digestion of Paralytic Shellfish Poisoning Toxins', *J. Chromatogr. A.*, **688**, 301–316 (1994).
69. T. van de Goor, A. Apffel, J. Chakel, W. Hancock, 'Capillary Electrophoresis of Peptides', in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 213–258, 1997.
70. I. Messana, D.V. Rossetti, L. Cassiano, F. Misiti, B. Giardina, M. Castagnola, 'Peptide Analysis by Capillary (Zone) Electrophoresis', *J. Chromatogr. B.*, **699**, 149–171 (1997).
71. T. Pritchett, F.A. Robey, 'Capillary Electrophoresis of Proteins', in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 259–295, 1997.
72. C. Schoneich, A.F.R. Huhmer, S.R. Rabel, J.F. Stobaugh, S.D.S. Jois, C.K. Larive, T.J. Siahaan, T.C. Squier, D.J. Bigelow, T.D. Williams, 'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **67**, 155R–181R (1995).
73. H.B. Hines, E.E. Brueggemann, 'Factors Affecting the Capillary Electrophoresis of Ricin, a Toxic Glycoprotein', *J. Chromatogr. A.*, **670**, 199–208 (1994).
74. H.B. Hines, E.E. Brueggemann, 'Behavior of Ricin on Untreated and Treated Capillary Electrophoresis Columns', *Chromatogr. Sci. Ser. (Capillary Electrophoresis Technol.)*, **64**, 819–841 (1993).
75. Organization for the Prohibition of Chemical Weapons, *Decision: Proficiency Testing Leading to Certification of Designated Laboratories*, C-I/DEC.60, 22 May 1997.
76. A.-E.F. Nassar, S.V. Lucas, W.R. Jones, L.D. Hoffland, 'Separation of Chemical Warfare Agent Degradation Products by the Reversal of Electroosmotic Flow in Capillary Electrophoresis', *Anal. Chem.*, **70**, 1085–1091 (1998).
77. Organization for the Prohibition of Chemical Weapons, *Decision:Criteria for the Designation of Laboratories by the OPCW*, C-I/DEC.61, 22 May 1997.
78. *Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament*, Ministry of Foreign Affairs of Finland, Helsinki, 1994, ISBN 951-724-008-2.
79. A.-E.F. Nassar, S.V. Lucas, C.A. Myler, W.R. Jones, M. Campisano, L.D. Hoffland, 'Quantitative Analysis of Chemical Warfare Agent Degradation Products in Reaction Masses Using Capillary Electrophoresis', *Anal. Chem.*, **70**, 3598–3604 (1998).
80. D.L. Olson, M.E. Lacey, J.V. Sweedler, 'The Nanoliter Niche: NMR Detection for Trace Analysis and Capillary Separations', *Anal. Chem.*, **70**, 257A–264A (1998).

Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Maarten S. Nieuwenhuizen
TNO Prins Maurits Laboratory, Rijswijk, The Netherlands

1 Introduction **1**
2 Roles for Detection and Concomitant Requirements **2**
2.1 Roles for Detection 2
2.2 Specific and General Requirements 3
3 Detection of Chemical Warfare Agent Vapors **3**
3.1 Wet-chemical Detection Systems for Chemical Warfare Agent Vapors 4
3.2 Automated Chemical or Physical-electronic Detection of Chemical Warfare Agent Vapors 6
4 Detection of Chemical Warfare Agent Liquids **9**
4.1 Detection of Free-standing Chemical Warfare Agent Liquids 9
4.2 Nondestructive Evaluation of Chemical Warfare Agent Liquids 10
5 Detection of Solids Contaminated with Chemical Warfare Agents **14**
6 Detection of Industrial Gases or Vapors **14**
6.1 Glass Detection Tubes 14
6.2 Detection of Explosive Vapor Mixtures, Flammability or Lack of Oxygen 15
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

The Organization for the Prohibition of Chemical Weapons (OPCW), which was established to achieve the object and purpose of the Chemical Weapons Convention (CWC), conducts verification activities and as a result numerous inspections take place. Detection plays a major role during many of the possible inspection scenarios and is a most important component of the health and safety policy of the Organization. This article deals with detection equipment for vapors, liquids, contaminated solids and industrial chemicals as well as explosive or flammable vapor mixtures or lack of oxygen.

1 INTRODUCTION

Verification procedures dealing with chemical warfare agents (CWAs) and other chemicals scheduled in the text of the CWC[1] may consist of:

1. Location of the area of interest for inspection scenarios dealing with CWA storage, production and destruction facilities and industrial facilities and old and abandoned chemical weapons sites, as well as investigations to obtain information relevant to alleged use.
2. Sampling and transportation, which involves procedures to send materials from the location area to a receiving area, which can be on the site itself or a chemical analysis laboratory elsewhere.

The OPCW, which was established to achieve the object and purpose of the CWC, conducts verification activities and as a result numerous inspections take place. Detection plays a major role during many of the possible inspection scenarios and is a most important component of the health and safety policy of the OPCW. These health and safety aspects are laid down in the CWC[1] in Part II of the Verification Annex under paragraph 43 of Conduct of Inspections. The hazard from CWA and other very toxic compounds may originate from vapors or aerosols, liquids (neat liquids as well as solutions) or solids (neat as well as contaminated solid materials) which may be converted into each other, e.g. liquids may evaporate, vapors may adsorb to solids. In a number of the scenarios mentioned these compounds may be encountered by an inspector, who is a member of an OPCW inspection team. As a result an inspector could become contaminated with the toxic materials through a number of mechanisms. Those mechanisms may involve breathing, touching, eating or drinking which are all related to the main entry routes of toxic materials into his body: the respiratory tract, the digestive tract or the skin.

All operations involving personnel of the OPCW will be carried out in order to minimize exposure as far as possible, afford reasonable safety to personnel, minimize operational risks and encourage good health. Therefore, for the sake of his own safety, i.e. to prevent himself from getting ill or even dying, an inspector needs to protect himself against the hazards he may encounter by taking

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the right countermeasures, such as wearing protective clothing and a gas mask.

The chemicals of interest to the OPCW are listed in three so-called schedule lists, which are laid down in the Annex on Chemicals of the CWC.[1] These lists not only include CWAs but also many CWA-related compounds.

Within the framework of verification procedures, detection is somewhat similar to detection in the framework of chemical defense, i.e. the detection of CWAs. Many examples exist nowadays of commercially available military equipment[2] that is capable of performing different detection tasks. The various types of equipment include manually operated so-called wet-chemical detection kits, as well as advanced automatic equipment for the detection of certain CWAs. Military detection equipment is usually and increasingly designed to be quick and easy to operate, of limited size and weight as well as robust.

At this point, the use of military detection equipment for OPCW purposes should be considered. Taking into account the scheduled chemicals of the CWC as well as specific requirements for verification purposes as laid down in OPCW's list of equipment[3] the following conclusions can be drawn:

- Military detection equipment can be used to fulfill a number of required tasks especially in the case of production, storage and destruction of CWAs, during inspections for alleged use as well as at old and abandoned chemical weapon sites.
- Military detection equipment is only suited for the detection of a limited list of CWAs and certainly not all chemicals of interest to the CWC, such as CWA precursors or during industrial inspections.
- Military detection equipment is designed to operate against accepted war-time risks. During peace-time operations, such as during OPCW inspections, the acceptance of casualties resulting from those operations will be much lower. As a result, protection from the use of detectors should be much better than in the case of military use and therefore basic requirements such as response time and sensitivity are more stringent.
- A new generation of detection equipment should be developed in order to detect all CWAs listed in the Schedule 1 list of the CWC.
- Because there are many hazards not related to CWAs, especially during industrial inspection scenarios, the introduction of additional detection methods is necessary. This includes equipment to safeguard an inspector from all kinds of industrial chemicals as well as from explosive or flammable vapor mixtures or low oxygen content of the air.

During OPCW's preparative period many discussions took place with respect to the choice of the inspection equipment including all kinds of detectors. Eventually, a list of inspection equipment was approved[3] containing operational requirements (general and specific), common evaluation criteria and technical specifications. These equipment items were purchased and are applied during inspections.

In this article, the detection equipment as agreed by the OPCW includes: detectors for CWAs themselves in the form of vapors (section 3), both based on wet-chemical detection principles or physical-electronic techniques; in the form of liquids (section 4), both free-standing liquids as well as liquids present in closed vessels; or in the form of contaminated solids (section 5). According to the OPCW list of equipment for detectors for industrial chemicals (section 6) only detection tubes will be discussed, because the CWC limits industrial chemical detectors to equipment that does not collect samples, i.e. throw-away devices. Section 6 will include detectors for explosive or flammable vapor mixtures or lack of oxygen.

2 ROLES FOR DETECTION AND CONCOMITANT REQUIREMENTS

2.1 Roles for Detection

The first and most important role for detection refers to safety. In many cases an inspector's sensing organs will not be able or sensitive enough to tell him when to take protective countermeasures. This is clearly expressed in Table 1, which lists so-called maximum allowable concentrations (MACs) for short-term exposure of CWAs. The data were derived from the occupational health literature for this type of agent or approximately from the toxic properties of the agents involved. Furthermore Table 1 lists the LS of the agents which have become known from either systematic studies or accidents which have taken place in the past. As can be seen in Table 1 in most cases the LS is higher than the MAC.

Table 1 Comparison of MACs of CWAs and LS

Agent	CAS-number	MAC ($mg\,m^{-3}$)	LS ($mg\,m^{-3}$)
Phosgene	(75-44-5)	0.4	3
Hydrogen cyanide	(74-90-8)	11	1
Cyanogen chloride	(506-77-4)	0.6	3
L	(541-25-3)	0.5	15
HD	(505-60-2)	0.05	1
GA	(77-81-6)	0.02	10
GB	(107-44-8)	0.01	5

LS, limit of smell; L, lewisite-1; HD, sulfur mustard gas; GA, tabun; GB, sarin.

Because an inspector cannot smell the danger he would have to wear protective gear all the time, which is a physical burden and is of course not an option. The only solution is to obtain proper information about the hazardous status of the environment by employing technical means other than using his own nose. This is called detection.

Detection may play an additional role: it relates not only to safety, but also to the location of the area or the object that is most interesting from an inspector's point of view, in the event that sampling and analysis have to take place during an inspection. Indiscriminate sampling for the different inspection scenarios obviously leads to an excess of samples. Large numbers of samples take a considerable amount of time to analyze, slowing down the inspection process considerably, which conflicts with the, often limited, duration of the inspection. The inspection team must await the results of the analyses before it can act on its findings or discard part of the samples. Therefore the use of detection equipment during an inspection may accelerate the process by acting as a pre-selector to find ''hot spots''. The number of samples will be limited or the quality of the samples may increase.

An additional role of detection refers to the screening of samples. One can imagine that during an inspection not all samples are taken as a result of a previous detection, because (rarely) the detector may not be present or the detector may not be easily operable near the sampling site (e.g. because of the location) or when samples are taken in a random way (e.g. when the amount of time is limited). In that case the samples, when collected at a central place on the inspection site, may be screened by using detection equipment in order to be classified in accordance with the detection result. This may be important if a selection of samples has to be made or if priorities have to be applied.

2.2 Specific and General Requirements

In view of the extreme toxicity of the CWAs, it should be obvious that the sensitivity of detectors as well as their response rates should be concomitantly high. This holds especially for detectors that detect CWAs in open contact with the environment. In fact, a dose-related short response time is an important requirement: the detector is allowed to respond a little bit slower at low concentration, because the dose an inspector may obtain in the period during which he is unprotected (response time of the detector plus time needed for putting on protective gear) may be acceptably low. For example, for vapors of nerve and blister agents in air the lower detection limits (DLs) are in the range 5–20 $\mu g\,m^{-3}$ and 200 $\mu g\,m^{-3}$, respectively, with concomitant response times of less than 2 min and a clear-down time, which is the time needed for the signal to arrive at its original value, of less than 10 min, preferably less than 2 min.

Assuming a detector is sensitive enough and the response is fast, it should also be very selective. On the one hand, when detectors are used for verification purposes, they should be able to detect broad categories of chemicals relevant to the purpose of the inspection. On the other hand they are supposed to show a minimum of false indications. False indications can mean two things. One is called a false positive signal: a signal occurs when there should not be one. This will cause a lot of irrelevant protective measures to be taken. Although this actually leads to some kind of over-protection, in the long run the results may still be dangerous. Through some kind of psychological mechanism people tend to ignore signals when the false signal rate is too high and as a result no protective measures may be taken when a real alarm occurs. The second false indication is called a false negative signal: no signal appears when there should be one. This may be caused by the presence of one or more other chemicals in combination with CWAs or other compounds to be detected. It is quite obvious that the false negative alarm is even more dangerous than the false positive alarm.

On top of the requirements for sensitivity, speed of response and selectivity, there are many general requirements for transportable devices to be employed in the field in toxic environments. These are all listed in the OPCW's list of equipment[3] and include issues such as set-up time, portability, operability by personnel in full protective gear, power supply (preferably battery operable), reliability, low pressure (transport by aircraft), temperature (−25 °C to 45 °C) and humidity (up to 95% relative humidity (RH)). Chemical hardening, including the possibility for proper decontamination is also important in this respect. Furthermore, data storage by this type of equipment is not allowed, nor may it collect samples in any way nor bring the sampled material outside the inspected area.

3 DETECTION OF CHEMICAL WARFARE AGENT VAPORS

Before World War II the presence of vapors of CWAs or other toxic chemicals was detected with biological indicators such as man's own nose or the use of small animals (e.g. birds were taken to the battle-field). The days have passed that these methods can be employed to prevent the users from getting hurt.

With respect to detection in the vapor phase, two approaches have been identified. One approach is called the chemical approach, i.e. the application of detection kits. The other, more modern, approach employs physical-electronic devices. Detection kits differ from the physical-electronic devices in their operation: the former are

usually rather laborious to operate. Furthermore, in the case of detection kits, the extensive use of (bio)chemicals may cause problems from the point of view of logistics as well as shelf-life. On the other hand, detection kits are often more selective, although considerable progress is being made with the various physical-electronic systems in that respect. The latter systems are usually faster and more flexible as they may easily be adapted to detect other compounds or classes of compounds.

3.1 Wet-chemical Detection Systems for Chemical Warfare Agent Vapors

According to the OPCW,[3] detection kits employing wet-chemical detection schemes must be able to detect blister agents, nerve agents and so-called blood agents (agents causing blocking of the hemoglobin in the blood) as a minimum. The sensitivity, expressed as a lower DL, should be 0.01 mg m^{-3} for nerve agents, 0.07 mg m^{-3} for mustard gases, 0.7 mg m^{-3} for L, 1 mg m^{-3} for cyanogen chloride and 6 mg m^{-3} for hydrogen cyanide with a response time of 5–15 min per class of agent.

When using a wet-chemical vapor detection kit, a rather complicated process yields a positive or negative detection result. First of all, a sample should be taken. Air is drawn through an adsorbing material using a manual pump. The adsorbing material may be present in a tube of which both ends need to be broken prior to use or on a little pad, both of which can be placed in front of the pump. Afterwards, one or more reagents are added to the sample according to a fixed procedure. In the case of detection tubes one or more of the chemicals are also present in the tube contained in small ampoules which can be broken successively after sampling has taken place. After some time, a color develops which tells the operator whether the detection result is positive or negative.

Below, in order to illustrate some of the chemistry involved, a number of detection schemes is treated in more detail. Two reaction schemes are shown dealing with the detection of organophosphorus compounds (nerve agents) including the Schoenemann reaction and the enzymatic detection principle. A third detection scheme deals with the detection of alkylating compounds such as mustard gases.

3.1.1 The Schoenemann Reaction

The Schoenemann reaction, originating from 1944,[4] is based on the fact that peroxophosphonates oxidize amines much easier than other peroxy ions, Scheme 1. First, the organophosphorus compound reacts with hydrogen peroxide or sodium perborate in an alkaline solution (pH 9–10). Then the resulting peroxophosphonate reacts with a leuco-dye such as benzidine or *o*-dianisidine to form an orange-brown colored reaction product. In certain modified reactions the amine is replaced by a precursor of a chemiluminescent compound (e.g. luminol)[5] or a fluorescent compound[6] (e.g. indole). The colorimetric variant of the Schoenemann reaction has been applied to many kits using detection tubes or detection pads.

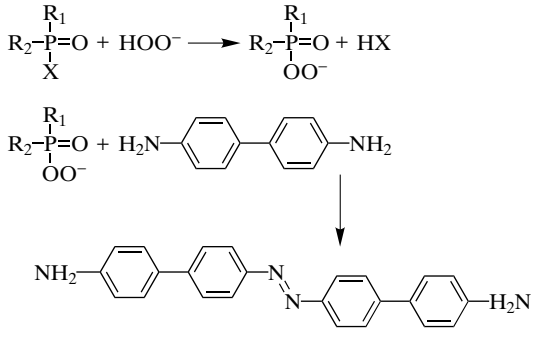

Scheme 1 The Schoenemann reaction for the detection of organophosphorus compounds.

3.1.2 Enzymatic Detection

The use of enzymes in the diagnosis of diseases is an important benefit from biochemical research and clinical analysis since the 1940s. The real impact of enzymes in chemical analysis was not felt until the early 1960s. In that period the use of enzymes for the detection of all kinds of compounds including so-called anticholinesterases was initiated.[7] During the impact of nerve agents on the human body the active center of the enzyme acetylcholinesterase is phosphorylated. Once the enzyme is deactivated it can no longer fulfill its function of hydrolyzing acetylcholine, which plays an important role in the transfer of nerve stimuli. Consequently, the acetylcholine concentration is increased and the specific symptoms of poisoning appear. The enzymatic detection method employs the enzyme-inhibiting properties of certain organophosphorus compounds. This kind of detection method can be regarded as a first example of a biosensor. In this detection method the naturally involved acetylcholine is replaced by a chromogenic substrate, for example 2,6-dichloroindophenyl acetate (Scheme 2), for that purpose. Normally, the enzyme will catalyze the hydrolysis of the ester group of the substrate, which causes a distinct color change from orange red to blue. As soon as the enzyme is inhibited this hydrolysis reaction will not proceed and the color will remain red. This difference in color between the active and inhibited state of the enzyme can be observed visually but also spectrophotometrically. The latter technique made automated detection of nerve agents possible.

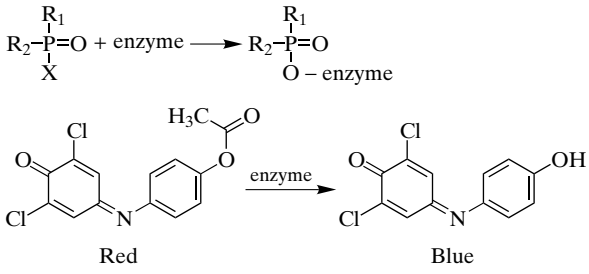

Scheme 2 The colorimetric version of the enzyme inhibition method for the detection of organophosphorus compounds.

3.1.3 Detection of Alkylating Compounds

Many CWAs other than organophosphorus compounds have in common their strongly alkylating properties, i.e. a good so-called leaving group such as the chlorine atoms of mustard gas, cyanogen chloride, phosgene and L. A detection method employing these alkylating properties especially for the detection of mustard gas was postulated in the 1950s[8] although the related chemistry had been known in the 1920s. As an example the reaction of HD with (4-nitrobenzyl)-pyridine, often called DB3, is shown in Scheme 3. Upon reaction with mustard gas a change from colorless to blue is observed after treatment with caustic soda, whereas reaction with cyanogen chloride yields a yellow-orange color.

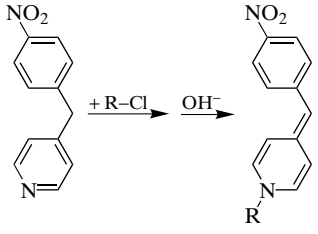

Scheme 3 The reaction of HD with (4-nitrobenzyl)-pyridine.

3.1.4 Detection Kits

The detection reactions described above are being applied nowadays in military detection kits.[2] Within the military context, the content of a specific detection kit is strongly dependent on the operational use that is foreseen as well as correlation with the results of other detection techniques, varying from individual devices to automated detectors, reconnaissance vehicles, remote sensing devices and even field laboratories. As a result the kits differ from each other with respect to the kind of CWA they are designed for, the kind of detection chemistry that is performed, the operating procedures, the sensitivities, the number of detections available per CWA and the lay-out of the kit as well as size, weight, shelf-life and price.

In the introduction it has already been stated that in the context of the OPCW, military detection kits are not always suitable, especially because their DLs are directly related to acceptable risks. Some military detection kits, however, already almost match the sensitivity requirement. In Table 2 the DLs of the current vapor reconnaissance kit in use by the Netherlands Armed Forces are listed as an example. In addition, the table lists the MAC already listed in Table 1 indicating that as a result of compromises related to the military use, the DLs are not always sufficient from a nonmilitary point of view. In order to lower the DLs of this type of wet-chemistry based detection kits one solution is simply to increase the sampling time. As a result more air is blown over the sampling pad or through the tube and more air contaminants are allowed to adsorb. The disadvantage is the increased amount of time needed for sampling as well as the increased number of false positives.

Apart from the rather complicated detection kits, individual hand-held detectors have also been developed, employing the same kind of chemistry. In the military context these are mostly used as end-of-alarm indicators. As an example of an individual nerve-agent detector, the so-called ''button'' detector was developed by the TNO Prins Maurits Laboratory and has been in use by the Netherlands Armed Forces since the mid 1980s. This cone-shaped detector consists of a plastic holder containing two separated air-permeable reagent papers. One paper (the enzyme paper) is impregnated with the enzyme butyryl cholinesterase, which acts in a similar way to the enzyme acetyl cholinesterase, and silica. The other paper (the substrate paper) contains the enzyme substrate 2,6-dichloroindophenyl acetate (see section 3.1.2). Also, the detector contains a reservoir that releases the reagent solution when punctured. When air is drawn through the enzyme paper, the organophosphorus compounds are adsorbed onto the silica. After some time the ''button'' is pressed. The reagent solution is released, the reagent papers are wetted and at the same time pressed together initiating the enzymatic detection reaction. After two minutes the blue color of the decomposed substrate

Table 2 DLs of the Netherlands vapor reconnaissance kit for a number of CWAs including MAC values for the agents listed in Table 1

Agent	MAC (mg m^{-3})	DL (mg m^{-3})
Phosgene	0.4	5
Hydrogen cyanide	11	5
Cyanogen chloride	0.6	1
L	0.5	3
HD	0.05	0.2
GA	0.02	0.002
GB	0.01	0.002

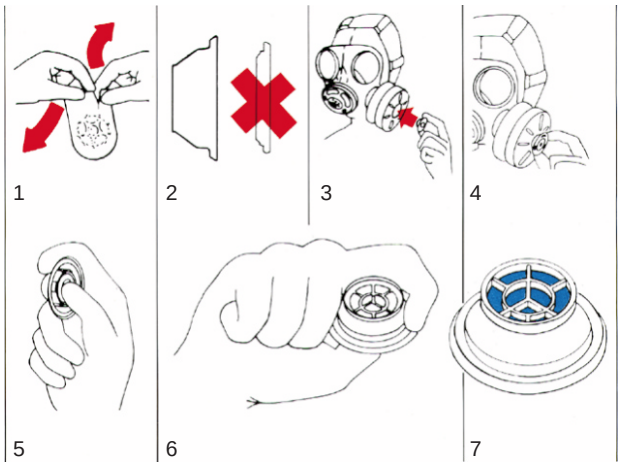

Figure 1 The Netherlands individual enzymatic nerve-agent detector (Steps: 1 = tear open, 2 = discard lid, 3 = place in inlet of gas-mask, 4 = breathe 20 times, 5 = take detector out and press firmly, 6 = wait 2 min, 7 = observe color).

can be clearly observed, or not in the case of a positive detection result. Figure 1 shows the very simple manual device of the Netherlands individual nerve-agent detector. Currently the DL for this device is in the range 0.005–0.020 mg m^{-3} depending upon the kind of nerve gas.

3.2 Automated Chemical or Physical-electronic Detection of Chemical Warfare Agent Vapors

Since 1980, several automated chemical or physical-electronic chemical vapor detectors have been developed for military,[2] industrial or environmental use. Although the systems are usually rather complicated inside, they are easy to handle. The systems have been designed as chemical vapor warning devices in order to tell persons when to take protective measures. These systems usually operate in an automated, continuous, stand-alone way. Other types of equipment often apply the same detection technologies in another way, i.e. as chemical vapor monitors, which are usually hand-held devices. From the point of view of the various CWC verification scenarios, monitors are much more relevant than warning devices. Military warning devices are meant to be operational for a relatively long period of time in a stand-alone mode in order to protect the troops against an attack with CWAs. Military monitoring equipment is used in close vicinity to the user and handled by the user. From the point of view of OPCW's inspection scenarios this way of using detectors shows more resemblance to warning equipment than the detection devices.

The most important vapor detection techniques currently available for the detection of CWA are based on electrochemical detection, flame photometry and ion mobility spectrometry (IMS) and examples of these techniques are actually used by the OPCW. These techniques are treated in more detail in the section below.

According to the OPCW[3] the dose-related detection requirements for hand-held vapor detectors are 5 mg s m^{-3} for nerve agents and 10 mg s m^{-3} for blister agents in less than 5 s and 0.5–2.4 mg s m^{-3} for nerve agents and 24 mg s m^{-3} for blister agents in less than 120 s.

3.2.1 Electrochemical Detection

An electrochemical method of detection is based on the fact that some organophosphorus compounds react with certain oximes such as isonitrosobenzoyl acetone (IBA) to liberate cyanide ions under the influence of a hydroxide ion and a good leaving group. In the case of V-type nerve agents a reactive leaving group is introduced upon reaction with a silver fluoride conversion filter prior to the reaction with IBA. The reaction intermediates initially yield one cyanide ion per molecule of CWA. Subsequently, a cyanide ion also reacts with IBA under the influence of hydroxide ions to yield two cyanide ions in return. Thus the initial level of cyanide ions is doubled which can be seen as chemical amplification. The cyanide can then be detected via a colorimetric reaction with *p*-nitrobenzaldehyde or can be determined electrochemically.

On the basis of this electrochemical method of detection a miniature CWA detector called Individual Chemical Agent Detector (ICAD) has been developed in the USA.[2] The detector consists of a reusable electronic module (processor, audible alarm and warning light), and a disposable sensor module containing the battery power source and the sensor cells.

Another approach to electrochemical detection is the use of ion-selective electrodes (ISEs). In the case of GB and soman, hydrolysis of the nerve agents yields a fluoride ion that can be detected with a fluoride-selective electrode whereas GA yields a cyanide ion that can be detected with a cyanide-selective electrode.

The enzyme inhibition approach can also be combined with electrochemical detection. When using acetyl or butyryl thiocholine as the substrate, a potentiometric detector can detect the activity of the enzyme. The concentration of converted substrate is a measure of the presence of the enzyme inhibitor.

3.2.2 Flame Photometry

The application of flame photometric detection (FPD) was first reported in 1966 as a selective gas-chromatographic detector.[9] Normally the concentration of organophosphorus compounds other than CWAs in the atmosphere will be very low. Therefore a more or less

selective method could be developed based on the mere detection of the phosphorus atom using flame emission. When phosphorus-containing compounds are burnt in a hydrogen-rich flame, excited HPO species are formed, whereas sulfur-containing compounds form excited S_2 species. When these species fall back to their ground state light is emitted in the range of 400–550 nm. In particular, the intensive flame emission near 526 nm is attributable to the HPO species and therefore characteristic for phosphorus, whereas flame emission from the S_2 molecule can be measured at 394 nm. Flame photometric devices only detect compounds containing phosphorus and/or sulfur and therefore do not cover the whole range of CWAs and related compounds. In the case of the nerve agent *O*-ethyl *S*-2-diisopropylaminoethylmethylphosphonothiolate (VX), which contains both the element phosphorus and the element sulfur, a response is shown in both the ''P-channel'' and the ''S-channel''.

In Figure 2 the basic scheme of an FPD device is shown. Air is drawn into the detection chamber using an air pump. The air contaminants are burnt in a hydrogen-rich flame. In the chamber a photometric cell in combination with optical filters is looking at the light emission phenomena at the relevant wavelengths.

Since the 1970s CWA detectors for military purposes were developed employing this method in a number of countries. The most advanced system nowadays is a hand-held system developed in France, called AP2C,[2] which can be used in a pleasant ergonomic way. The main problem, however, for this type of detection devices is the hydrogen supply, which requires an extra logistic pathway in addition to electrical power, which is used by any kind of physical-electronic system.

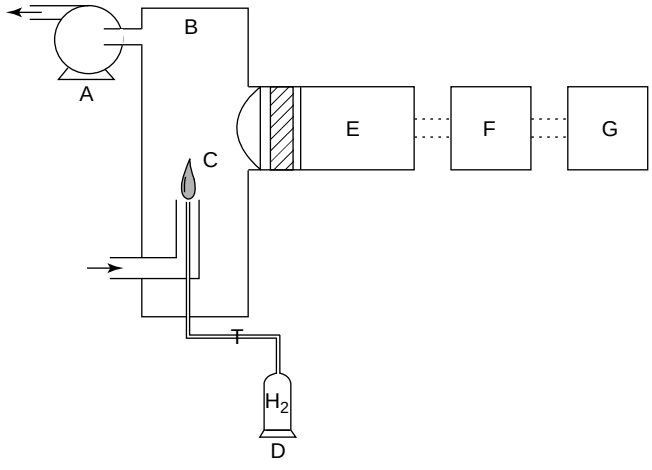

Figure 2 Schematic representation of an FPD device (A = air pump, B = reaction chamber, C = flame, D = hydrogen supply, E = photometric cell, F = electronics, G = display).

DLs for military flame photometry-based detectors are approximately 0.01 mg m^{-3} for nerve agents and approximately 0.1 mg m^{-3} for HD with almost real-time responses.

3.2.3 Ion Mobility Spectrometry

IMS refers to the principles, practice and instrumentation for characterizing chemical substances through the measurement of gas-phase ion mobilities. Although mobility measurements of gas-phase ions have been studied by physicists since the beginning of the twentieth century, no attempt was made to employ these phenomena for chemical detection purposes. In the 1960s numerous devices employing relatively simple ion separations (i.e. without the use of high vacuum and magnetic fields as in mass spectrometry) were described, mostly in a military context, until in 1970 the development of IMS began.[10] In modern analytical IMS methods,[11] ion mobilities are determined from ion velocities measured in a drift tube with supporting electronics. All processes occur commonly at ambient pressure in air or in nitrogen. Ion mobilities are characteristic of substances and provide a means for detecting and more or less identifying vapors.

In this paragraph a general description of IMS is given. In Figure 3 a schematic lay-out of an IMS device is shown indicating the reaction region and the drift region. This IMS device is the Chemical Agent Monitor (CAM)[2] used by many armed forces in the world, at our own TNO Prins Maurits Laboratory as well as by the OPCW.

Outside ambient air is drawn into the inlet of the detector by a pump. This air with its possible contaminants first passes through a dust filter and, subsequently, comes into contact with a silicone membrane. The membrane allows some of the target material to permeate whilst excluding (to a higher degree) a number of other contaminants

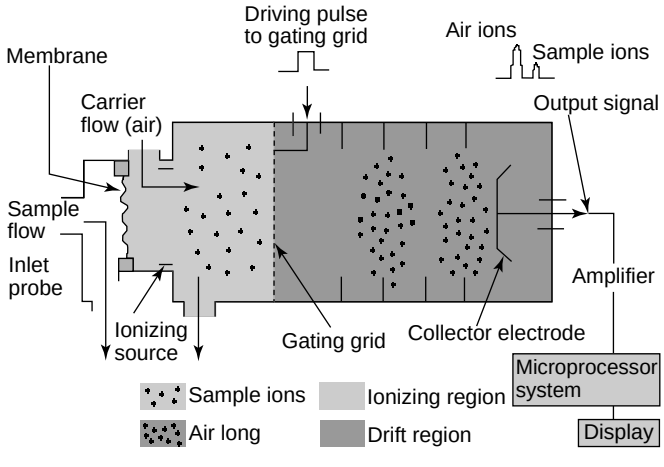

Figure 3 Schematic representation of an IMS detection device. (Illustration obtained from Graseby Dynamics, UK.)

found in air such as water vapor molecules. Subsequently, ions are generated by subjecting the permeated air to a ^{63}Ni radioactive source emitting β-radiation. The β-particles ionize the air molecules to form a population of highly mobile positive and negative air ions, the so-called reactant ions, among which are $[(N_2)_x \cdot H(H_2O)_n]^+$ and $[(N_2)_x \cdot (H_2O)_n \cdot O_2]^-$ ($x = 1–3$; $n = 1–3$). Charge transfers may happen by multiple collisions of these reactant ions with molecules of contaminants if present in the air. These complicated reactions result in the formation of relatively stable ionic clusters. These so-called product ions are usually relatively large compared to the (air) reactant ions. The ions are swept through the reaction region by an electrical field.

Figure 4 represents the formation of a positively charged ion cluster in air containing water (Equation a), and the formation of negatively charged ion clusters in air containing water as a result of ion transfer (Equation b), charge transfer (Equation c) or dissociative charge transfer (Equation d).

Entrance of the formed cluster ions into the IMS drift tube is restricted by a metal grid which acts as a shutter. The shutter grid opens periodically (typically for 0.3 ms in every 30 ms) allowing the ionic clusters to enter the drift region. The shutter grid provides control over the length of the period and the interval between them. An electric field (commonly 200 V cm^{-1}) imparts a constant force to the ion clusters leading to separation in accordance with their respective mobilities (commonly 1–2 cm^2 V^{-1} s^{-1}). Ion velocity is inversely dependent upon the effective collisional cross-section of an ion. This makes IMS an ion-size analyzer. Ions with the highest mobility travel faster and advance in front of those with lower mobilities. At the end of the drift tube the ions collide with a collector giving rise to current pulses. Plotting this current, which is a measure of the intensity of the ions, versus the drift time results in an ion mobility ''spectrum'', in the past called a ''plasma chromatogram''.[9] Although the latter term better represents the separation of the cluster ions, the technique is nevertheless called spectrometry, which may be confusing from a physical point of view.

a Proton transfer:

$$M + [(N_2)_x \cdot H(H_2O)_n]^+ \longrightarrow [(N_2)_x \cdot MH]^+ + n\,H_2O$$

b Ion transfer:

$$M + [(N_2)_x \cdot (H_2O)_n \cdot O_2]^- \longrightarrow [(N_2)_x \cdot M \cdot O_2]^- + n\,H_2O$$

c Charge transfer:

$$M + [(N_2)_x \cdot (H_2O)_n \cdot O_2]^- \longrightarrow [(N_2)_x \cdot M]^- + O_2 + n\,H_2O$$

d Dissociative charge transfer:

$$XY + [(N_2)_x \cdot (H_2O)_n \cdot O_2]^- \longrightarrow [(N_2)_x \cdot X]^- + Y + O_2 + n\,H_2O$$

Figure 4 Reaction equation for the formation of positively charged (a) and negatively charged (b, c and d) ion clusters in air with water (M = analyte molecule, X and Y = analyte molecule fragments; $x = 1–3$; $n = 1–3$).

The IMS plot will not always show one single peak. First of all, not all reactant ions are converted into clusters and, therefore, a residual reactant ion peak (RIP) is present, which can be used for internal reference purposes. Secondly, one single compound may generate a number of equally stable clusters with different mobilities as a result of their composition, i.e. so-called monomer clusters containing one analyte molecule and some reactant species as well as larger dimer clusters resulting from the interaction between the monomer cluster and one more analyte molecule resulting in a lower mobility (larger drift time). The relative peak height of the various peaks is usually concentration dependent, whereas peak shapes are affected by the complexity of the cluster mixture as well as by the kinetics of the reactions between the clusters as compared to the IMS timescale.

Two IMS modes may be employed: a positive mode detecting positively charged clusters originating from compounds with a high proton affinity such as organophosphorus compounds or a negative mode detecting negatively charged clusters originating from compounds with a high electron affinity such as organochlorine compounds (mustard gases, L).

Figure 5 shows a number of IMS plots recorded by the TNO Prins Maurits Laboratory with a CAM. The IMS plots could be revealed by using dedicated data-acquisition software. Normally, only qualitative and semi-quantitative information (bar level) is provided on a liquid crystal display screen. In Figure 5(a), a positive mode IMS plot of the nerve gas GB (0.2 mg m^{-3} at 45% RH and 22 °C) is depicted showing both the monomer and the dimer peak along with the RIP. In Figure 5(b), the negative mode IMS plot of the blister agent HD is depicted (0.6 mg m^{-3} at 45% RH and 21 °C). In this plot also a cluster is shown with a mobility larger than RIP. This is the result of HD breaking apart into smaller ions such as Cl^-. Figure 5(c) shows the IMS plot of the blister agent L (0.3 mg m^{-3} at 60% RH and 21 °C) indicating both a monomer and a dimer cluster as well as a cluster related to the decomposition product.

Owing to the short opening time of the shutter grid, only about 1% of the total ion current is collected. This affects the DL, a disadvantage which is claimed to be more than compensated for by the high selectivity of this IMS technology. This allows the manufacturer to set the alarm levels relatively low, bringing the DL to the required level. The disadvantage of setting the DL low is that in general more interferences may occur. Therefore, a major disadvantage of IMS-based detectors is the poor selectivity. In practice both false positives (signal without analyte present) and false negatives (no signal when analyte is present) tend to occur, especially when the

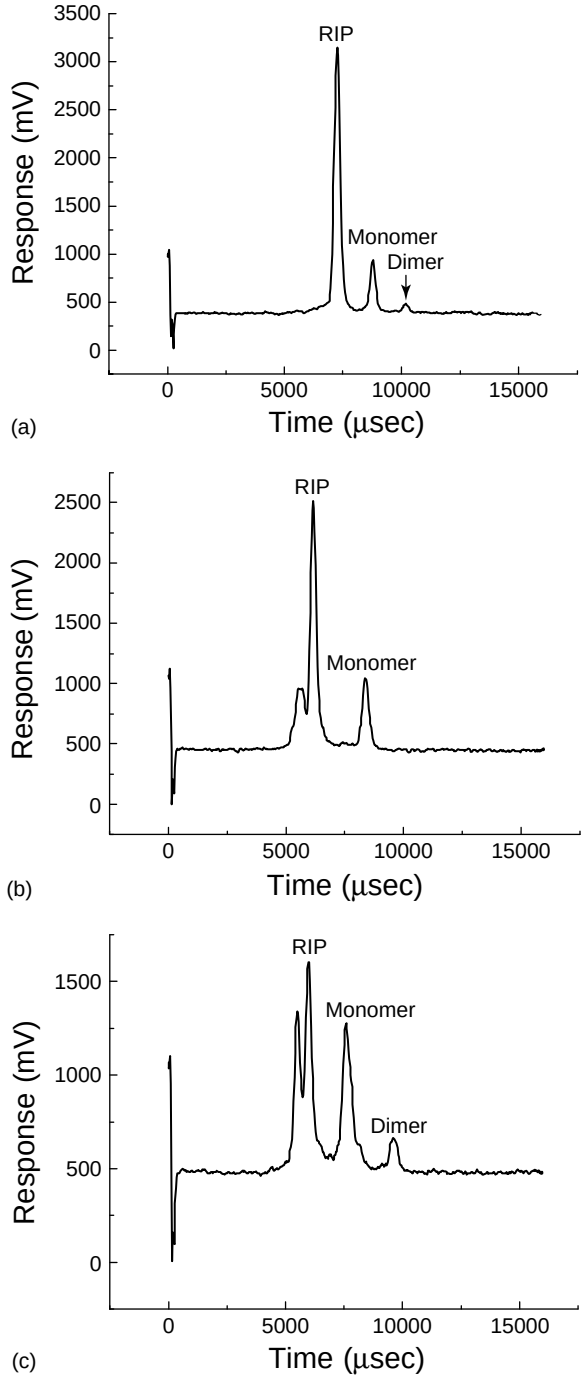

Figure 5 IMS spectra of (a) GB (positive mode; $0.2\,mg\,m^{-3}$), (b) HD (negative mode; $0.6\,mg\,m^{-3}$) and (c) L (negative mode; $0.3\,mg\,m^{-3}$). (Data recorded at the TNO Prins Maurits Laboratory.)

sensitivity is set relatively high. Training of operators may be very helpful in dealing with these phenomena. In addition the use of different physical-electronic techniques together with IMS may yield a large synergetic effect in the area of false signal rejection. During many UNSCOM missions in Iraq, for example, CAM and the flame photometric detector AP2C, operated "hand in hand" in a very satisfactory way. The DLs for military IMS-based detectors are $0.01–0.04\,mg\,m^{-3}$ for nerve agents and $0.2\,mg\,m^{-3}$ for HD with response time in the order of 1 min.

With respect to IMS it should be mentioned that all over the world developments are under way aiming at improved IMS, exploring nonradioactive ion sources, improved data-handling systems and miniaturization as well as coupling of IMS to gas-separation techniques such as gas chromatography. In particular, the search for nonradioactive ion sources is important from the point of view of the OPCW because the presence of such a source could potentially cause transport problems on international deployment. Furthermore, dedicated IMS systems for other compounds than CWA are already available, whereas flexibility is increased by using writable memories by which one can easily change from one target analyte to the other.

4 DETECTION OF CHEMICAL WARFARE AGENT LIQUIDS

This paragraph deals with the detection of neat CWA liquids (detection of CWAs dissolved in water is not considered here) including the detection of free-standing CWA liquids as well as so-called nondestructive evaluation (NDE) techniques for the detection of CWA liquids contained in various kinds of vessels. NDE is a screening technique which aims at obtaining more information (including some information about the chemical nature) about the liquid inside the vessel in order to determine if it contains a CWA or not. Depending on the kind of scenario this information is relevant in many ways. For instance the information may contribute to the decision whether sampling, i.e. opening of the vessel and taking a sample, is needed.

4.1 Detection of Free-standing Chemical Warfare Agent Liquids

According to the OPCW,[3] neat liquids of G-type nerve agents, V-type nerve agents and mustard gas should be determined quickly and separately. To detect neat liquids, several types of liquid detection paper, as well as powders or chalks employing colorimetric reactions have been developed for military use. The reactions should not be compared with the relatively complicated reaction schemes in the case of wet-chemical detection of vapors (see section 3.1). In general, the matrix material (paper, chalk, powder) is impregnated with one or more

reactive dyes which develop a color change upon reaction with a CWA. These methods may be especially useful within the context of the verification of alleged use and during inspections of military production, chemical weapon stockpiles or CWA destruction facilities. Liquid detection paper is easily used as an initial screening method for droplets or larger amounts of liquid that are suspected to consist of CWAs. Nerve agents and blister agents are detected quickly in a more-or-less selective way. Some liquid-detection methods are able to detect the three different categories of CWA as mentioned in the requirements of the OPCW[3] with one and the same device, because of the presence of three different dyes impregnated in the matrix.

In contrast to the advanced vapor-detection equipment which exists at the moment there is a lack of direct physical-electronic equipment for the detection of free-standing CWA liquids, especially when small amounts of liquids are lying on the ground or on material or even absorbed in the pores, cracks and crevices of material. Some systems have been proposed to fill this gap, most of them employing active laser-based detection methods, sometimes even preceded by chemical treatment of the surface to generate chemiluminescent compounds as in the case of the Schoenemann reaction mentioned in section 3.1.

It should be clear that from an operational point of view the simple methods for the detection of free-standing liquids or contaminated surfaces do not always suffice. In order to overcome this problem an indirect approach is often used. Assuming any liquid is accompanied by a certain amount of vapor, one could try to reveal the presence of a liquid by tracking down the vapor all the way to the source. This approach only works if the vapors are not evenly distributed in the atmosphere. Only in the case of vapor concentration gradients (plumes) may one track the vapor in the direction of higher concentrations ending up at the source.

At this point it should be stressed firmly that this indirect approach is limited in the case of CWAs dissolved in liquids because the vapor pressure will be much lower in that case allowing only a very small amount of agent vapor to be released into the atmosphere. Also at low temperatures and with agents which possess a low volatility (e.g. the nerve agent VX) or with CWAs that are very strongly adsorbed to the surface, the release of vapor into the atmosphere may be very limited whereas the so-called contact hazard may still be significant. CWAs will still be transferred from the liquid or the contaminated surface through the skin of a relatively warm bare hand.

The OPCW requirements for vapor detectors being able to detect chemicals in a liquid form are 0.01 mg cm^{-2} for nerve agents and 0.08 mg cm^{-2} for mustard gas.[3] Military vapor-detection equipment[2] can be used for this indirect approach. After the proper amount of training, which is even more important in this case than in the case of vapor detection as such, one can use a flame photometric device or an IMS-based device to screen surfaces in a very controlled way in order to follow concentration gradients and hit the liquid spots or the leaking tanks or ammunition.

One can also try to transfer the liquid to the detector by picking it up with an absorbing material and releasing the vapor by heating the material in front of the inlet of a vapor-detection device. Although the number of manipulations is increased and special tools are needed, the chances of locating the liquid or contaminated surfaces in a sensitive and selective way are raised significantly. One of the commercially available FPD devices, the previously mentioned AP2C,[2] employs a small battery-powered accessory to absorb liquids and subsequently release them into the inlet of the vapor detector.

4.2 Nondestructive Evaluation of Chemical Warfare Agent Liquids

As stated in the previous paragraph there are several ways to monitor the environment for the presence of CWA liquids. In certain types of inspections, for example at declared or suspected chemical weapons stockpiles, in the case of unexploded munitions or at destruction facilities, an inspection team may be faced with the issue of verifying the contents of a piece of munition, a reactor, a tubing system or a container. Especially in the case of chemical munitions, these scenarios are difficult to deal with. The objects are both chemically and explosively hazardous and therefore safety concerns are very important. Some of the problems an inspector is confronted with are: many chemical munitions appear identical to conventional weapons; agent containers or storage tanks are not unique; munitions can be multi-packed and containerized. Nevertheless, the verification of the contents both qualitatively and quantitatively forms a major task for an inspector. One could try to open the object by drilling or punching a hole to create access to the liquid or facilitate the release of vapor but one can imagine that this may be impossible, very dangerous, at least very time-consuming and require special skilled people (explosive ordnance personnel). Therefore, several techniques have been developed to obtain information about the content of an object without opening or damaging the object. A convenient suite of NDE equipment[12] is available for the characterization of munition fills. Acoustic, X-ray and radiometric techniques will be treated in more detail.

Decisions about inspection alternatives are in general complex because of the numbers and types of objects, storage configurations, inspection objectives and the

Table 3 Relationship between information derived (increasing information content from top-down), the method employed and the applicable CWA items for the method

Increasing information content	Detection method	Munition configuration
Solid	Human ear	Free standing
Solid versus liquid	Acoustic or X-ray	Free standing
Physical properties	Acoustic	Free standing
Integral fill signature	Acoustic	Free standing
Inventory of elements in the fill	Neutron activation	Any configuration
Molecular structure determination	Sampling and analysis	Any configuration

level of confidence required. Table 3 summarizes the relationship between information that may be derived, the method employed and the applicable CWA items for the method.

4.2.1 Acoustic and Ultrasonic Detection Methods

As a minimum the human ear can act as an NDE device to determine whether free-standing liquid is present or only solid materials. When tapped with a metal object such items will give a bell-like sound if they are empty or liquid filled. It should be noted that the human ear approach may be somewhat difficult when wearing protective clothing including a gas-mask. Acoustic and ultrasonic systems with their enhanced analysis ability over the human ear, can detect liquid-filled items, determine the fill levels and obtain signatures that can lead to distinguishing the CWA fill type. When other than free-standing munitions are present and when a more direct indication of the fill type is desired, the neutron interrogation systems can be applied.

The major limitation of the acoustic and ultrasonic techniques is that access to the primary wall of the container is required. Shipping containers and packaging effectively preclude the use of acoustic probing techniques.

In the case of acoustic resonance spectroscopy (ARS)[13] objects respond with resonant sound when mechanically excited. Two piezoelectric transducers are used: a transmitter and a receiver. By employing the piezoelectric effect, which is a long-known phenomenon,[14] electrical energy is converted into mechanical energy in the form of acoustical waves by the transmitter. The properties such as frequency (in the 1–30 kHz range) and amplitude of a sweep of acoustic waves with different wavelengths are modulated by the physical properties and geometry of the object the waves meet on their way to the receiver, where the mechanical energy is converted back to electrical energy employing the same piezoelectric effect. In the output spectrum thus obtained signals only become distinguishable (up to two orders of magnitude above the noise level) when a resonance condition is reached, which is determined by the fill and the fill level. As a result the resonance spectrum provides a unique signature for objects and their contents. Because the acoustic signature is based on whole-body vibrations of the object, the exact placement of the transducers on the object is not very critical.

According to the OPCW,[3] ARS must, apart from general requirements, be able to discriminate between solid and liquid-filled munitions with an acceptable degree of confidence and be able to determine the fill level of a container.

Figure 6 shows the response of the nerve agents GB and VX in the same type of projectiles. It should be stressed that this does not mean that some degree of absolute identification can be performed with this kind of equipment. Since the total system is responding, the orientation of the item is not critical and even stand-off excitation is possible. ARS can be best used to measure the degree of similarity between two or more filled items. Thus, if the resonance spectrum of a certain type of shell filled with, for instance, HD is known, the spectra of other shells of the same type with an unknown fill can be compared with it. Furthermore, this method can be employed to determine the thickness of the walls of a container. Especially, in the case of relatively old munitions or storage tanks this will be a very useful approach.

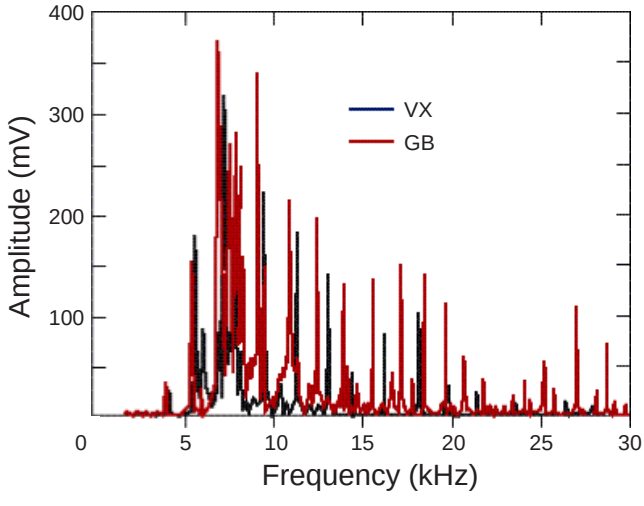

Figure 6 Acoustic resonance spectra of two 155-mm shells, one filled with the nerve agent VX and one filled with the nerve agent GB. (Reproduced from Sinha and Apt[13] by permission of the University of California, Lawrence Livermore National Laboratory (LLNL), and the US Department of Energy (DOE).)

Ultrasonic pulse echo (UPE) interrogation[15] of physical structures is a very mature technology. The basic principle of this technique is similar to that of sonar. A high-voltage transducer, coupled to the object to be investigated, provides a square-wave pulse of ultrasonic energy that travels through the wall and into the contents. A search unit made from a piezoelectric material is used to convert the reflected sound waves back into electrical signals. Pulse reflections and times of echo returns are a function of the physical properties of the object. The frequencies used are in the megahertz range. Key characteristics are that pulses are reflected from any material boundary, echoes are returned sooner in the case of solids, and powder fills damp out multiple echoes. All interactions provide indications of the nature of the fill. The simple presence or absence of an echo is a clear indication of whether a fill is a liquid or a solid. The times of the echo returns can indicate structural or fill variations. The level can be determined by sweeping the device alongside the object.

According to the OPCW,[3] UPE equipment must, apart from general requirements, be able to sort like items by comparison of sound propagation time with an acceptable degree of confidence and be able to determine the liquid fill level of a container.

In Table 4 the differences in speed of sound, expressed as their transit times, of the nerve agents GB and VX as well as a surrogate munition (ethylene glycol and water) in 155 mm shells are clearly indicated.[15]

4.2.2 X-ray Photography

Another way of performing NDE is the use of low-power portable X-ray analysis. Commercial X-ray equipment is available for this purpose. The density of the material which an X-ray, emitted from a source, meets on its way from the source to a detector, determines the amount of radiation that actually hits the detector. This affects the degree of whitening of a photosensitive material. In fact this technique operates in the same way as X-ray photography for medical purposes.

One can employ both X-ray photography or X-ray video. X-ray photography is cheaper and yields better resolution, whereas X-ray videos are more expensive but the operation is real-time and the resolution is adequate.

Table 4 Different transit times for different fills of a 155 mm shell

Fill type	Mean transit time (ms)
GB	66.7
VX	47.9
Ethylene glycol/ water surrogate	58.3

X-ray photography is routinely used to locate items in overpack containers in order to visualize the presence of steel or at least high-density objects in containers or boxes of plastic or wood. Furthermore, X-ray photography can be used to image the internal structure of munitions (X-rays of 150–300 kV can penetrate steel casings up to 1.5 cm thick) and to assess the presence or absence of buster charges.

X-ray interrogation can also be used to discriminate between solid and liquid contents by looking at the solid–gas or solid–liquid interface. By moving an object such as a shell with respect to the earth's gravitational field one can easily distinguish a solid from a liquid because the solid–gas interface will hardly move whereas the liquid–gas interface will move as a result of gravitation. This effect is schematically visualized in Figure 7. The information obtained i.e. whether the shell contains a solid or a liquid is an important part of the tiered inspection process as shown in Table 3.

4.2.3 Radiometric Detection Methods

Neutron activation analysis (NAA)[16–18] is a more powerful NDE technique than acoustic spectroscopy and X-ray interrogation although more time-consuming, costly and subjected to safety regulations. Instrumentation, including portable systems such as portable isotropic neutron spectrometry (PINS), is available for identifying the presence of individual elements indicative of CWAs inside thick-walled steel vessels such as artillery shells and even through secondary containers. Both CWAs and high explosives are rich in the elements carbon, hydrogen and oxygen, but CWAs also contain rather unique combinations of the elements arsenic, chlorine, phosphorus and sulfur. By determining the atomic ratios

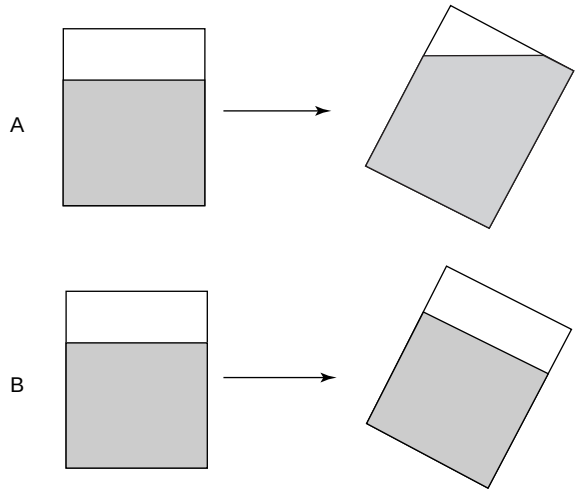

Figure 7 X-ray photographs of a container in a vertical and a tilted position with a liquid content (A) and a solid content (B).

present, phosphorus smoke munition and high explosives can easily and automatically be distinguished from G-type or V-type nerve agent munitions. The detector operates in a near real-time fashion and can be readily operated.

The object to be investigated is irradiated typically with low-energy neutrons (binding energies inside the nucleus are 7–10 MeV, except for hydrogen where it is 2.2 MeV) produced by an isotopic neutron source such as ^{252}Cf. The source is placed near the munition being surveyed. Each second, approximately one million neutrons are emitted by the source, some of which penetrate the casing and interact with the contents: nuclei are excited and decay promptly (in less than a picosecond) to the ground state with the emission of characteristic γ-rays. With a high-purity germanium γ-ray detector, requiring constant low temperature (77 K), the energies and intensities of γ-rays released by neutron interactions are measured. The energy of the γ-radiation is characteristic of elements present in the object. In Table 5 a number of characteristic energies are shown as well as the relative elemental composition in a number of CWAs. These energies have been measured and cataloged since the mid 1950s.[19]

Figure 8 illustrates the difference between the γ-ray spectrum of mustard gas in a ton container and that of a high explosive in a 155 mm shell. Some difficulty in easy access to randomly chosen items, such as shells at the back of a bunker or in the middle of a stack can be envisaged. Furthermore, the presence of a radioactive source could potentially cause transport problems on international deployment. According to the OPCW,[3]

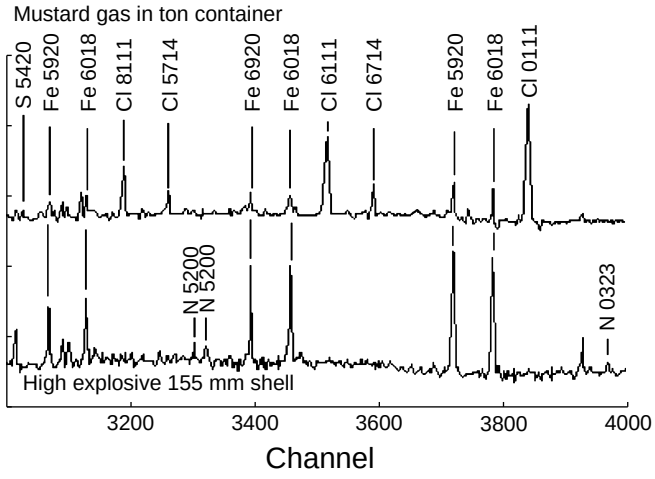

Figure 8 γ-Ray spectrum of mustard gas in a ton container (top) and high explosives in a 155-mm shell (bottom). (Reproduced from Caffrey et al.[19] by permission of the University of California Lawrence Livermore National Library and the US Department of Energy.)

neutron interrogation NDE equipment must, apart from fulfilling the general requirements, have levels of emitted radiation such that extensive shielding is not required to limit exposure to safe levels at a reasonable stand-off, be capable of detecting chemical elements and/or ratios of chemical elements in chemicals related to the CWC, including phosphorus, sulfur, arsenic, nitrogen, chlorine and, to the extent possible, carbon, hydrogen, fluorine, bromine and iodine.

Table 5 Characteristic γ-ray energies of some key elements for NAA detection as well as the relative elemental composition in a number of CWAs and the high explosive TNT

Energy (keV)	Element	Reaction	GB	VX	HD	L	TNT	Remark
279.5	As	n,n′γ				36.1		L
983.5	Ti	n,n′γ						Smokes
1266.1	P	n,n′γ	22.1	11.6				Nerve agents
2233.4		n,n′γ						
3900.3		n,n′γ						
1950.9	Cl	n,γ			44.7	51.3		L, HD, smokes, bleach
6110.9		n,γ						
2211.8	Al	n,γ						Casings
197.1	F	n,n′γ	13.6					GB
582.1		n,γ						
3683.9	C	n,γ	34.3	49.4	30.2	11.4	37.0	Many agents and materials
4945.3		n,γ						
2223.3	H	n,γ	7.1	9.7	5.0	1.0	2.2	Calibration line
2230.2	S	n,n′γ		12.0	20.1			HD, VX, smokes
5240.5		n,γ						
7631.1	Fe	n,γ						Steel containers
7645.5		n,γ						
10318.3	N	n,γ		5.2			18.5	HN_3, VX, explosives

For fast and large-scale screening of objects an important NAA method is based on hydrogen concentration measurement (HCM). In this method fast neutrons are emitted from a neutron source (^{252}Cf). Upon collision with protons, the neutrons slow down into so-called thermal neutrons which are in thermal equilibrium with the environment. The neutrons can leave the object and be measured by a detector, usually consisting of a scintillator converting neutron energy into light and a photomultiplier tube for amplification of the light emission. Under controlled circumstances the number of neutrons detected per unit of time relates to the number of hydrogen atoms per unit of volume. Therefore, calibration for each type of object and fill is required.

Without HCM the ratio of elements such as the P/S ratio cannot be accurately determined with a single measurement, for hydrogen significantly diminishes the flow of fast neutrons. A typical measuring time is less than one minute compared to the 30-min measuring time of other NAA methods. With HCM even discrimination between, for instance, different nerve agents or high explosives can take place provided the contents are relatively pure. On top of that, because HCM is so fast, the fill level may also be measured in this way, especially when UPE fails in the case of double-walled containers where the required contact between the object under investigation and the transducer is not possible.

One major drawback of the HCM technique is the fact that L can not be detected very well because of its very low hydrogen content.

According to the OPCW,[3] the HCM must be capable of determining hydrogen concentration within the fill of munition items or other containers at an accuracy of better than 10%. Furthermore, it must be capable of allowing discrimination between explosives and CWAs such as nerve gases or mustard as well as discriminating between representatives of the latter class of agents.

5 DETECTION OF SOLIDS CONTAMINATED WITH CHEMICAL WARFARE AGENTS

At present, there are hardly any direct detection methods to determine whether solid materials are contaminated. Relatively large amounts of suspect liquids can at least be observed visually, but a hazard resulting from contaminated solid materials cannot always be observed in the same way because the CWA may be absorbed or adsorbed, whereas the contact hazard is still considerably high. Therefore, for solid materials, the level of protection should be raised and gloves must be used at all times.

An indirect approach is the monitoring of the surrounding air with a vapor-detection system. However, this presents the same limitations as mentioned in section 4.1 for the indirect detection of liquid CWAs. The reader is referred to section 4.1 for further information about the use of vapor detectors for the detection of contaminated solids or for the checking of decontaminated solids.

6 DETECTION OF INDUSTRIAL GASES OR VAPORS

In the previous paragraphs the emphasis was put on the detection of CWAs. However a number of inspection scenarios are not CWA related, but they may also take place in a toxic environment generating safety problems. Although in general toxicity may not be as high as in the case of CWAs, the scope of chemical hazards which may be encountered by an inspector will be very wide. In the chemical industry safety of personnel is a very important issue. Within the field of occupational safety and health many toxic chemicals or classes of chemicals have been identified and a lot of toxicological information has been generated in order to set exposure limits for people working in environments containing those chemicals. As in the case of protection against CWA one has to detect a hazard in order to protect oneself against it. Therefore detection equipment for many different chemical hazards has been developed and is being sold commercially and used on a large scale. The same general problems apply, such as selectivity, sensitivity, false-alarm rates, stand-alone compared with manually operated equipment, that were discussed for CWA vapor detection and equipment.

A basic subdivision of the field is made into direct-reading colorimetric indicators and direct-reading instruments. From the point of view of the OPCW currently two classes of equipment are very important: direct-reading colorimeters, especially glass detection tubes, and equipment for the detection of flammability and of explosive vapor mixtures or lack of oxygen as well.

6.1 Glass Detection Tubes

According to the OPCW,[3] glass detection tubes are to be used to test the environment for the presence of toxic gases or vapors. The detection should be indicated by the length or shade of color change of the detector tube after sampling and shown some kind of concentration indication in ppm, $mg\,m^{-3}$ or % volume. After being used the tubes are thrown away.

Glass detection tubes containing solid and/or liquid chemicals are convenient and compact detectors. During the past decades there has been a great expansion in the development and use of these tubes. Several companies are employing a number of basic reaction schemes for

the detection of many organic and inorganic compounds. Some reactions are amongst others:

- reduction of chromate or dichromate to chromous ions for the detection of alcohols, aldehydes, ketones, esters, hydrocarbons or sulfur dioxide;
- reduction of iodine pentoxide with fuming sulfuric acid to iodine for the detection of aromatic hydrocarbons, acetylene, carbon monoxide or vinyl chloride;
- reduction of ammonium molybdate with palladium sulfate to molybdenum blue for the detection of ethylene, butadiene, hydrogen sulfide and sulfur dioxide;
- reaction with potassium palladosulfite for the detection of carbon monoxide or hydrogen cyanide;
- color change of pH indicators for the detection of acidic and alkaline compounds;
- reaction with *o*-toluidine for the detection of chlorine or chlorinated organic compounds, bromine or nitrogen dioxide;
- reaction with tetraphenylbenzidine for the detection of halogens, halogenated organic compounds, nitric oxide, or acetonitrile.

The use of these detection tubes is extremely simple. Before use, both ends are broken and the glass tube is placed in a dedicated holder which is fitted with a squeeze bulb or piston pump. Subsequently, the recommended air volume is drawn through the tube by the operator, closely following the manufacturer's instructions. After this the operator reads the concentration in air by examining the exposed tube. The length of the colored stain is a semi-quantitative measure of the concentration.

In some cases direct sampling may be a problem. Then the tube may be placed near the sampling point and the pump is operated at some distance. A rubber tube extension with the same inside diameter as the sampling tube may be inserted between the detector tube and the pump. When hot air samples need to be taken, cooling may be essential. In that case a special probe of glass or metal may be attached to the inlet of the tube. When employing special accessories it is very important that the air components are not being adsorbed on the materials used. Therefore the previously mentioned rubber extension is placed between the tube and the pump and not in front of the tube.

The selectivity of the tubes is a major consideration with respect to the applicability and the interpretation. In some cases the lack of selectivity permits detection of a class of chemical compounds rather than a single compound.

Although detection tubes are usually advertised as being capable of use by unskilled people and the operating procedures are indeed very simple, many limitations and potential errors are inherent in this method. The results may be dangerously misleading and sometimes need to be supervised by a well-trained person.

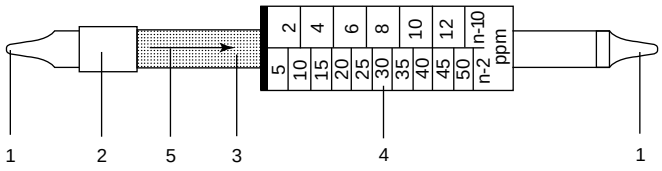

Figure 9 Lay-out of a detection tube for hydrogen cyanide (1 = sealed tube tips, 2 = marking area, 3 = white protective layer, 4 = blue indicating layer in ppm, one scale for two pump strokes and the other scale for 10 pump strokes, 5 = flow direction arrow).

Figure 9 shows an example of the lay-out of a detection tube for hydrogen cyanide, a toxic industrial chemical also placed on the CWC schedule list. The detection range is 5–50 ppm for two pump strokes and 2–12 ppm for 10 pump strokes. The analyte is reacting with mercury chloride to yield hydrogen chloride which causes a pH-sensitive dye to change from blue to yellow. The type of reaction employed is not sensitive for hydrogen, hydrocarbons and carbon monoxide up to 50 vol.%, carbon dioxide up to 15 vol.%, halogenated hydrocarbons, nitriles, carbon disulfide and acetic acid up to 1 vol.%, ammonia and sulfur dioxide up to 1000 ppm and hydrogen sulfide and hydrogen chloride up to 300 ppm.

6.2 Detection of Explosive Vapor Mixtures, Flammability or Lack of Oxygen

According to the OPCW,[3] the purpose of a flammability/explosive/air-quality monitor is to allow safe entry into and stay in confined spaces with potentially flammable/explosive atmospheres or insufficient air quality.

A general chemical hazard which does not relate to a specific chemical compound is the presence of combustible gases, i.e. the hazard that a mixture of oxygen and one or more chemical compounds (hydrogen or volatile hydrocarbons) might explode upon some kind of ignition mechanism. Explosive mixtures sometimes behave very subtly with respect to ignition. A very simple human action taking place inside or near the gas mixture such as a scratch or a spark resulting from static electricity may cause the mixture to explode.

A somewhat related issue is the lack of oxygen. As a result of a combination of the lack of fresh air flowing through a building or a similar construction and the release of gases or vapors driving out the oxygen or smoldering fires or biological activities consuming the oxygen, the amount of oxygen may no longer be sufficient

for inspectors to operate normally, causing an important safety and health risk.

Within the framework of a number of CWC-related inspection scenarios it is possible for an inspector to be confronted with explosive mixtures or oxygen deficiency. Inside closed or more-or-less closed buildings or bunkers, near storage sites, abandoned sites or chemical plants, all of which may be partly damaged, the hazards described may be present. Again, as in the cases described in the previous sections, one should be capable of protecting oneself by way of, amongst others, leaving the site, flushing the atmosphere with fresh air or consuming pressurized air. Specific detection equipment is to be used to tell the operator if and when the hazard exists.

Many companies sell different explosive-vapor monitors, but they employ more-or-less the same operating principle. Most combustible- or explosive-gas detectors are based on the principle that gases will react with oxygen in the presence of a heated catalyst mounted on a pellistor. Oxidation of a gas further raises the temperature of the pellistor causing, in turn, an increase in the resistance, which is detected by a Wheatstone bridge and indicated on the display as a gas concentration. This detection principle has some limitations as some specific gas mixtures may poison the catalyst. Leaded gasoline and certain compounds containing silicone are suspect in this respect. For this reason and because of the fact that the lower explosion limit (LEL) may differ from one gas to another, explosive-gas detectors have to be calibrated frequently and tuned in on the LEL of the worst case to be expected. New developments in the field of catalysis have improved the catalysts with respect to the risks of being poisoned. In the commercial literature the following vapors which can be detected employing this technique, are mentioned: aldehydes, alcohols, esters, saturated and unsaturated hydrocarbons, ketones.

Oxygen in the air is usually detected by the electrochemical operating principle in the 0–25% oxygen range. A number of instruments combine explosive-vapor detection and oxygen detection in a general-purpose detector.

ACKNOWLEDGMENTS

Mr E.R.J. Wils of the TNO Prins Maurits Laboratory is acknowledged for critically reading the manuscript. Mr A.L. De Jong is acknowledged for providing a number of figures. Mr R.C.M. Olivier is acknowledged for recording the IMS spectra as presented in Figure 8.

Figures 6 and 8 are based on a paper prepared at LLNL for US DOE. Neither LLNL nor US Government makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, product or process disclosed in this paper. Credit must be given to the University of California, LLNL and the DOE under whose auspices the work was performed, when this information or a reproduction of it is used.

ABBREVIATIONS AND ACRONYMS

ARS	Acoustic Resonance Spectroscopy
CAM	Chemical Agent Monitor
CWA	Chemical Warfare Agent
CWC	Chemical Weapons Convention
DL	Detection Limit
DOE	Department of Energy
FPD	Flame Photometric Detection
GA	Tabun
GB	Sarin
HCM	Hydrogen Concentration Measurement
HD	Sulfur Mustard Gas
IBA	Isonitrosobenzoyl Acetone
ICAD	Individual Chemical Agent Detector
IMS	Ion Mobility Spectrometry
ISE	Ion-selective Electrode
L	Lewisite-1
LEL	Lower Explosion Limit
LLNL	Lawrence Livermore National Laboratory
LS	Limit of Smell
MAC	Maximum Allowable Concentration
NAA	Neutron Activation Analysis
NDE	Nondestructive Evaluation
OPCW	Organization for the Prohibition of Chemical Weapons
PINS	Portable Isotropic Neutron Spectrometry
RH	Relative Humidity
RIP	Reactant Ion Peak
UPE	Ultrasonic Pulse Echo
VX	*O*-Ethyl *S*-2-diisopropylamino-ethylmethylphosphonothiolate

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)** Verification of Chemicals Related to the Chemical Weapons Convention • Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)** Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons

Convention • Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Optical Gas Sensors in Analytical Chemistry: Applications and Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Neutron Activation in Environmental Analysis

Forensic Science **(Volume 5)**
Ion Mobility Spectrometry in Forensic Science

Industrial Hygiene **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Nuclear Methods **(Volume 14)**
Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

REFERENCES

1. Anonymous, *Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction*, 13 January, 1993.
2. *Jane's NBC Protection Equipment 1996–1997*, ed. T.J. Gander, Jane's Data Division, Coulsdon, UK, 139–174, 1996.
3. OPCW First Conference of the State Parties, *Decision: List of Approved Equipment with Operational Requirements*, C-I/DEC.71, 23 May, 1997.
4. R.B.R. Schoenemann, *New Reaction for the Detection of Metalloid-nonmetal Linkages*, DB 119877, Office of Publication Board US Dept. of Commerce, 1944.
5. J. Goldenson, 'Detection of Nerve Gases by Chemiluminescence', *Anal. Chem.*, **29**, 877–879 (1957).
6. B. Gehauf, J. Goldenson, 'Detection and Estimation of Nerve Gases by Fluorescence Reaction', *Anal. Chem.*, **29**, 276–278 (1957).
7. C. Gelman, D.N. Kramer, *Enzymatic Methods for Detection of Anticholinesterases*, US Patent 3049411, 1962.
8. J. Epstein, R.W. Rosenthal, R.J. Ess, 'Use of γ-(4-Nitrobenzyl)pyridine as Analytical Reagent for Ethyleneimines and Alkylating Agents', *Anal. Chem.*, **27**, 1435–1439 (1955) and references cited herein.
9. J. Brody, J. Chaney, 'Flame Photometric Detection. The Application of a Specific Detector for Phosphorus and for Sulfur Compounds – Sensitivities to Subnanogram Quantities', *J. Gas Chromatogr.*, **4**, 42–46 (1966).
10. M.J. Cohen, F.W. Karasek, 'Plasma Chromatography. A New Dimension for Gas Chromatography and Mass Spectrometry', *J. Chromatogr. Sci.*, **8**, 330–337 (1970).
11. *Ion Mobility Spectrometry*, ed. G.A. Eiceman, Z. Karpas, CRC Press, Boca Raton, FL, 1994.
12. J.K. Hartwell, A.J. Caffrey, 'Non-destructive Evaluation Techniques for Chemical Weapons Destruction' in *Analytical Chemistry Associated with the Destruction of Chemical Weapons*, ed. M. Heyl, R. McGuire, Kluwer Academic Publishers, The Netherlands, 305–312, 1997.
13. D.N. Sinha, K.E. Apt, 'Acoustic Resonance Spectroscopy', *Verification Technologies*, First/Second Quarters, 52–56 (1992).
14. J. Curie, P. Curie, *Bulletin de la Societé Minerales de Paris*, **3**, 90 (1880).
15. T. Taylor, 'Ultrasonic Pulse-echo Measurements for CW Verification', *Verification Technologies*, First/Second Quarters, 57–64 (1992).
16. A.J. Caffrey, J.D. Cole, R.J. Gehrke, R.C. Greenwood, 'Chemical Warfare Agent and High Explosive Identification by Spectroscopy of Neutron-induced Gamma Rays', *IEEE Trans. Nucl. Sci.*, **39**, 1422–1426 (1992).
17. A.J. Caffrey, R.J. Gehrke, R.C. Greenwood, J.K. Hartwell, K.M. Krebs, G.D. McLaughlin, A.L. Siedenstrang, K.D. Watts, *US Army Experience with the PINS Chemical Assay System*, Report EGG-NRP-11443 by Idaho National Engineering Laboratory, 1994.
18. M.A. Lone, R.A. Leavit, D.A. Harrison, *Prompt Gamma Rays from Thermal-neutron Capture*, Atomic Data and Nuclear Data Tables 26, 511, 1981.
19. A.J. Caffrey, J.D. Cole, R.J. Gehrke, R.C. Greenwood, K.M. Krebs, 'Portable Isotopic Neutron Spectroscopy for Nondestructive Evaluation of CW', *Verification Technologies*, First/Second Quarters, 35–43 (1992).

Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Martin T. Söderström
Finnish Institute for Verification of the Chemical Weapons Convention, University of Helsinki, Finland

1 Introduction 1
2 Sample Preparation 2
2.1 KBr Pellets and Liquid Cells 3
2.2 Attenuated Total Reflectance 5
2.3 Gas Cell 5
2.4 Gas Chromatography 5
2.5 Other Techniques 7
3 Data Evaluation 7
3.1 Gram–Schmidt Chromatogram 7
3.2 Functional Group Monitoring 8
3.3 Reference Spectra and Spectral Libraries 8
3.4 Library Search 10
3.5 Interpretation 11
4 Applications 15
4.1 International Tests 15
4.2 Attenuated Total Reflection Applications 15
4.3 Gas Chromatography/Fourier Transform Infrared Applications 16
5 Quality Control 16
6 Comparison with Other Analytical Methods 17
Abbreviations and Acronyms 18
Related Articles 18
References 19

Fourier transform infrared spectroscopy (FTIR) is an analytical technique, which gives characteristic spectra. The spectra can be used like fingerprints in the identification of chemicals. The chemicals can be identified by comparison with reference data or, in some cases, by interpretation, which can be followed by synthesis for confirmation. When FTIR is connected to a gas chromatograph (GC) and the results are used together with those from mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy, very complex environmental samples containing unknown chemicals can be analyzed reliably.

1 INTRODUCTION

Infrared (IR) spectroscopy is one of the oldest instrumental analytical techniques but its value in structural analysis has decreased somewhat with the rise of NMR spectroscopy and MS. Compared to the traditional dispersive IR techniques, FTIR offers more sampling techniques.

In the analysis of environmental samples, all kinds of chromatographic techniques are necessary because the sample matrices can be very complicated. So far, the best commercially available IR technique for environmental analysis is gas chromatography/Fourier transform infrared spectroscopy (GC/FTIR). The association of GC/FTIR with different gas chromatography/mass spectrometry (GC/MS) methods is a valuable tool in analysis of environmental samples. The power of FTIR is that it gives very characteristic spectra, which can be used almost like fingerprints.

The Chemical Weapons Convention (CWC), which came into force in April 1997, contains provisions for chemical analysis of samples in both on-site and off-site laboratories. On-site laboratories will be manned by the inspectors of the Organization for the Prohibition of the Chemical Weapons (OPCW) and equipped with the analytical equipment approved by the states that are party to the OPCW. Off-site laboratories will analyze ambiguous samples which on-site analysis cannot resolve. These research laboratories are equipped with sensitive instruments and are designated by the OPCW on the basis of performance in proficiency tests.

Chemical warfare agents (CWA) are classified into several categories, e.g. nerve agents and vesicants. Tear gases are forbidden as a method of warfare but are allowed for riot control purposes. The CWC lists chemicals in three Schedules, which have been constructed on the basis of the toxicity of the chemicals, their industrial use and historical usage as warfare agents. Schedule 1 consists of families of toxic chemicals, which have been developed, produced, stockpiled or used as chemical weapons, e.g. sarin and mustard gas. These chemicals have very little industrial use. Schedule 3, on the other hand, consists of 17 single chemicals with very large industrial use, e.g. phosgene and triethanolamine.

There are challenges in the CWC-related analysis. Even though the Schedules restrict the number of possible target chemicals, the number of chemicals is still more

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

than 10 000. The properties of the chemicals vary from solids to gases and from volatile to nonvolatile chemicals. In IR the absorptivities vary considerably.

One challenge in the analysis of the CWA is the analysis of precursors and degradation products, which are often nonvolatile. The CWA degrade, e.g. hydrolyze or oxidize, quite easily. Traditional IR sampling techniques, like KBr pellets and liquid cells, are well suited to analysis of neat or quite concentrated nonvolatile chemicals. Environmental samples containing these kinds of chemical, however, normally require derivatization before GC/FTIR analysis.

In the field of CWC-related analysis, chemicals can be identified by comparing their spectra with spectra in spectral libraries or with spectra measured from authentic chemicals. Spectral interpretation is not enough for unambiguous identification but it is an important tool in structural elucidation of unknown chemicals.

As usual in analytical chemistry, a combination of the results of several different analytical techniques gives the most reliable results. In a well-equipped off-site laboratory, results from FTIR, MS and NMR, together with other analytical data, can be combined to produce reliable, unambiguous analytical proof to support chemical disarmament.

Good books are available on some of the subjects discussed in the article, for example, on the following subjects: chromatography FTIR,[1] FTIR spectroscopy in general[2,3] and IR spectra of phosphorus-containing chemicals.[4]

2 SAMPLE PREPARATION

Many of the chemicals related to the CWC are very toxic, and therefore care must be taken while handling samples that may contain these chemicals.

Different sample preparation or introduction methods in IR can be divided into two categories: methods where no sample separation takes place and methods where the sample undergoes separation prior to FTIR measurement. The former class (e.g. KBr pellets), liquid or gas cells, and attenuated total reflectance (ATR), can be used only for relatively pure and concentrated samples, e.g. bulk materials. The latter class (e.g. GC) is required when analyzing environmental samples or mixtures where the target chemical is not concentrated. The American Society of Testing and Materials (ASTM) has published a standard practice describing different general IR spectroscopy techniques.[5]

There are restrictions in the selection of the sample preparation or introduction method when the laboratory is taken on-site. Methods requiring supplies difficult to provide in field conditions, typically liquid nitrogen, cannot be applied on-site. Thus, GC-based and ATR methods are not normally applicable on-site. A suggestion for a basic on-site sample preparation kit is shown in Table 1.

The physical state of the chemical to be analyzed determines which sample preparation method is selected. The majority of the chemicals listed in the Schedules of the CWC are liquids in room temperature. There are also chemicals which are gases or solids, in each schedule. Some chemicals are borderline cases, which may be solids at room temperature but melt in the IR beam or may be liquids that are too volatile in the IR beam. Table 2 summarizes some typical chemicals of each type in the Schedules.

Sample introduction using GC solves many problems related to the analysis of chemicals related to the CWC. The amount of chemical needed is quite low, which almost removes the danger of handling these chemicals. A drawback in the GC is that it can only be used to analyze volatile chemicals. Volatile derivatives can be made from most of the nonvolatile Schedules chemicals.

Table 1 A suggestion for a basic on-site FTIR kit

General equipment		
Sealed and desiccated FTIR spectrometer		
Evacuable desiccator with desiccant		
Transport container for KBr pellets/powder with desiccant		
Small pump for evacuating desiccator and gas cells		
Pasteur pipettes and bulbs		
Forceps with replaceable rubber tubing around the tips		
Polishing kit: glass plate, felt pad, abrasive		
Solvents: acetone, dry ethanol		
Gloves (nonpowdered)		
Decontamination solution		
Containers for decontamination (for samples, tissue, and glass)		
Decontaminatable trays for sample preparation		
Tissue and filter paper		
Small vials (e.g. 4 or 8 mL) with caps		
Chemical agent detector		
Glove bag/box or fume cupboard		
Outer container for liquid or solid samples		
Equipment for liquid samples	**Equipment for solid samples**	**Equipment for gas samples**
Liquid cell	KBr powder	Gas cell
KBr windows for liquid cell	Hydraulic or hand press	KBr disk for gas cell
Holder for liquid cell	KBr pressing kit (with spare steel pellets if possible)	Holder for gas cell
Disposable sample cards (optional)	Mortar and pestle	Tubing and connectors as required
	Spatulas	Pressure sensor
	Mulling agent if required	Bottle for vaporization

Table 2 Summary of physical states of some typical scheduled chemicals in the CWC. (Chemical Abstracts Service (CAS) numbers and melting/boiling points for borderline chemicals are shown in parentheses.)

Physical state	Schedule 1	Schedule 2	Schedule 3
Solid	Salts of VX Saxitoxin (35523-89-8) Ricin (9009-86-3)	BZ (6581-06-2) salts of nitrogen compounds	Phosphorus pentachloride (10026-13-8)
Solid/liquid	bis(2-Chloroethylthio)-methane (63869-13-6, mp 6 °C)	Methylphosphonyl dichloride (676-97-1, mp 33 °C)	Triethanol amine (102-71-6, mp 21 °C)
Liquid	Sarin (107-44-8) Tabun (77-81-6) VX (50782-69-9) Mustard gas (505-60-2)	Dimethyl methylphosphonate (756-79-6) Thiodiglycol (111-48-8)	Phosphorus oxychloride (10025-87-3) Trimethyl phosphite (121-45-9)
Liquid/gas	Chlorosarin (1445-76-7, bp 51 °C) Chlorosoman (7040-57-5, bp 65 °C)	–	Hydrogen cyanide (74-90-8, bp 26 °C)
Gas	1,1-Dimethylethyl methylphosphono-fluoridate (13273-12-6)	PFIB (382-21-8)	Phosgene (75-44-5) Cyanogen chloride (506-77-4)

VX, *O*-ethyl *S*-2-(diisopropylamino)ethyl methylphosphonothioate; BZ, 3-quinuclidyl benzilate; PFIB, 1,1,3,3,3-pentafluoro-2-(trifluoromethyl)-1-propene.

2.1 KBr Pellets and Liquid Cells

Most solid or liquid chemicals related to the CWC can be analyzed using KBr pellets or liquid cells.

The preparation of KBr pellets is a standard procedure in laboratory conditions since the hydraulic press used for pressing is standard equipment. In an on-site laboratory, the press is not a very practical tool because it weighs about 50 kg. If a press is a necessity on-site, the possibility of using a hand-held KBr press should be considered.

2.1.1 Solid Samples

There are several techniques for preparing solid samples for IR analysis: solid KBr pellets, melts, solutions and mulls. While analyzing chemicals related to the CWC, introduction of additional absorbance bands in the spectra should be avoided. Therefore, the last two methods should not be used if other methods are available, since they always have peaks present due to solvent or oil.

If no equipment for preparing the KBr disks is available on-site, other methods that do not use solid KBr disks must be chosen. Solid chemicals form a minority in CWC-related chemicals. These chemicals could be prepared on-site, e.g. by dissolving them first in a suitable solvent, depositing a couple of droplets on a KBr disk, and then evaporating the solvent off. The spectra obtained in this way are not of the best quality and it is not normally a recommended procedure under laboratory conditions, but is often sufficient for identification as shown in Figure 1(c).

Thorough instructions for preparing solid KBr samples are presented by Dent.[6] The article also includes trouble shooting for the preparation of KBr disks.

When measuring spectra of very hard crystals, sometimes, distorted bands, almost like those found in derivative spectra, can be seen with solid KBr or a mulling technique. This is caused by the Christiansen effect[7] which appears if crystals that are too large are left in the sample after grinding. The IR beam is reflected from the crystals, and as the refractive index changes over an IR band the peak shape is distorted. If the crystals are ground more so that the particle size becomes smaller than the wavelength of the IR radiation (2.5–25 μm in the mid-IR region), no distortion appears. Figure 1 shows the Christiansen effect in the spectrum of tear gas CR (CAS 257-07-8). First (Figure 1a) the sample grinding was deficient and the Christiansen effect was clearly visible. After this, the sample was extensively reground, resulting in a nondistorted spectrum (Figure 1b). Atmospheric water had been introduced to the sample during the regrinding. For comparison, spectra obtained by a dissolution–deposition–evaporation method (Figure 1c) and by GC/FTIR (Figure 1d) are shown.

2.1.2 Liquid Samples

Generally, the analysis of pure liquid chemicals is very straightforward. Liquid chemicals with a boiling point below 100 °C should generally be analyzed in a sealed cell (preferable for CWC-related analysis) or in solution.[2] Other chemicals (bp >100 °C) can be easily analyzed between KBr pellets. CWC-related chemicals should be considered very toxic, by default, and therefore extra care should be taken in their analysis, to avoid the vapor hazard.

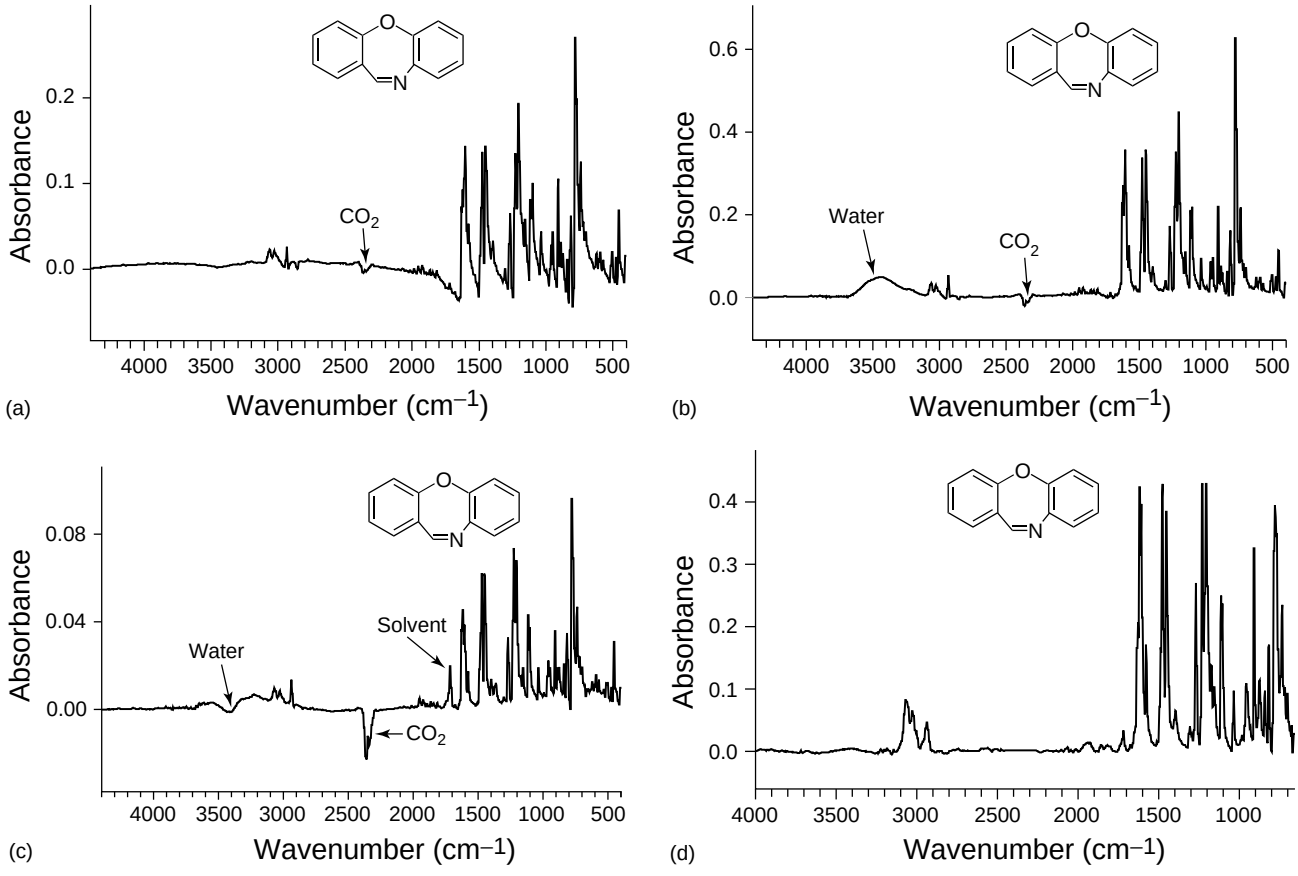

Figure 1 Spectra of CR (dibenz[*b*,*f*][1,4]oxazepine) using different sample preparation methods. The spectra have been measured (a) as solid KBr disks with deficient grinding of CR (Christiansen effect visible), (b) as solid KBr disks with CR well reground, (c) dissolving CR into acetone and deposition on a KBr disk and (d) by cryodeposition GC/FTIR. (See Söderström.[8])

Measuring the IR spectra of liquid chemicals is relatively simple, if commercial KBr windows for liquid cells are available. These windows can be reused if cleaned and decontaminated effectively. Decontamination solution containing potassium hydroxide and ethanol works effectively, but it also corrodes the windows extensively. The decontaminated windows often need polishing to make them fully transparent and smooth again.

In the case where the windows must be guaranteed to be clean, new windows should be used each time. However, this can become quite expensive if a large number of samples is to be analyzed. In such a case, the use of a press for preparing KBr pellets should be considered.

2.1.3 Safety

Always when handling CWC-related chemicals, safety must come first. Sample preparation must always be carried out in a fume cupboard. Protective gloves should always be used and a respirator should be kept available for emergencies. Work on live agents should not be performed alone; there should always be somebody who oversees the person handling the toxic chemicals.

An airtight sampler holder or at least an outer container for the sample should be used when possible. The spectrometer can be placed in a fume cupboard to avoid vapor hazard, but good ventilation of the sample compartment can be enough. It should be remembered that sample preparation should not be carried out in the same fume cupboard, to avoid contamination of the instrument. Decontamination of a FTIR spectrometer is very difficult, if not impossible.

The work area should be kept as tidy as possible to enable safe sample preparation. It is a good idea to use trays in the fume cupboard. These are easy to clean and decontaminate and also they contain any spillage of chemicals.

A good practice is to monitor the workspace inside the fume cupboard with a continuous hand-held or table-top chemical agent detector, if such a device is available. There are several models commercially available, based on ion mobility, flame photometric, enzymatic or photoacoustic detection.[9]

2.2 Attenuated Total Reflectance

ATR is a technique based on total internal reflections at the crystal surface. In a way, the IR spectrum is measured from a very thin film surrounding the IR transparent ATR crystal. This technique is one of the best IR sampling techniques suitable for analyses of chemicals in water. The detection limit is less than 1 mg mL^{-1} for nerve agents.[10]

The sampling cell with the crystal can be also a flow through cell, which makes the sample handling easier.

The use of an ATR accessory normally requires a mercury–cadmium–telluride (MCT) detector to achieve enough sensitivity. The detector requires liquid nitrogen cooling, which limits the on-site usage of this method.

2.3 Gas Cell

Gas cells can be used for gaseous and volatile liquid samples. The use of the gas cell is simple, but care should be taken in emptying the cell, since the sample may be very toxic. A gas sample can be sucked into an evacuated cell from the atmosphere or from a container containing gas, if a suitable connection is available. Volatile liquid samples can be placed as droplets into the cell, which is closed, and the chemical can be allowed to vaporize. The gas phase spectrum can then be measured. The droplets can be also placed in a separate bottle for vaporization, and the vapor can then be transferred into the evacuated gas cell.

The difficulty in gas sampling is that the spectra differ from condensed phase spectra. Therefore, reference spectra of gas phase spectra are required for accurate identification.

In principle, the cell should not need cleaning, other than emptying it, after the analysis, since only the gaseous phase is measured. However, one should be certain that all possible volatilizable chemicals are removed from the surfaces before the next measurement. The KBr windows of a gas cell may require polishing after use, especially when used in humid conditions.

2.4 Gas Chromatography

In general, GC is very well suited to analysis of very toxic volatile chemicals since the analyzed quantities can be very low. Normal injection volumes with capillary columns are 1–2 μL. Most of the CWA are volatile, and their nonvolatile degradation products can be made volatile through derivatization (e.g. methylation or silylation).

There are three different types of GC/FTIR interface: light-pipe, matrix isolation (MI), and cryodeposition (also direct-deposition, cryotrapping). All three have been used for the analysis of CWC-related chemicals. The light-pipe interface has been the most popular, and it is available from several manufacturers. At present, there are no GC/MI/FTIR (gas chromatography/matrix isolation/Fourier transform infrared spectroscopy) instruments available commercially but there used to be the Cryolect® system from Mattson. The usage of cryodeposition in this type of analysis has been increasing over the years, even though it is available from only one manufacturer (Bio-Rad Tracer®).

ASTM has published a standard practice for GC/FTIR analysis, covering basic features of each type of interface.[11]

2.4.1 Light-pipe

The light-pipe interface is the easiest method of connecting a GC and an FTIR spectrometer. The light-pipe is a long flow cell with reflective inner coating and IR transparent windows at both ends of the cell. It is very easy to operate and to maintain. The problem with the light-pipe interface is that it is not very sensitive compared to a normal benchtop mass spectrometer, and this makes the analysis of same samples difficult. The IR spectra of the chemicals are measured in the vapor phase, as the GC effluent is hot. One of the problems associated with the spectrum measurement in the gas phase is that a gas phase spectrum of a chemical differs considerably from those measured in other phases. See section 3.3 for further discussion of this subject.

2.4.2 Cryodeposition

Cryodeposition is the newest interface type for a GC/FTIR instrument. In this system, the eluents in the GC effluent are frozen on an IR transparent slide which is cooled using liquid nitrogen. The carrier gas evaporates in the process so that the chemicals are directly deposited on to the slide surface. Transmission spectra are then measured through the slide. These spectra are like normal condensed phase spectra, with rare exceptions. The sensitivity is five times better than in a light-pipe and the same or even slightly better than in GC/MI/FTIR.

Since the principle of the cryodeposition instrument is based on depositing the eluting compounds in very small spots (ca. 100 × 100 μm) on the depositing surface, all large amounts deposited on the surfaces will interfere with the operation of the system by spreading too widely on the surface. For this reason, the eluent exiting the deposition tip must be directed to a solvent sink (i.e. away from the slide) until the solvent has eluted. The wait time for solvent is controlled by a parameter, the ''solvent delay''. This necessary delay in starting the data collection means that all the chemicals eluting before the solvent or any other large peak has eluted are missed.

2.4.3 Comparison of Interfaces

Using either GC/MI/FTIR or cryodeposition, GC/FTIR spectra can normally be obtained from the same environmental samples (concentration, e.g. 1–10 ppm) as in GC/MS. The chromatograms produced by GC/FTIR instruments differ somewhat from those in GC/MS because in IR the absorbance varies greatly depending on the chemicals. However, the same chromatographic peaks can be located in the different chromatograms.

GC resolution degrades slightly in all of the available GC/FTIR interfaces. In the light-pipe type interface some broadening of the chromatographic peaks occurs because the diameter of the light-pipe (often 1 mm) is larger that of a capillary column (0.25–0.5 mm). Thus the GC resolution degrades by an average factor of 1.2.[1] In the MI and cryodeposition interfaces there is spreading of the chromatographic peaks owing to the deposition of the sample on a surface causing the GC resolution to decrease by an average factor of 1.8.[1]

It is also possible to combine GC/MS and GC/FTIR instruments so that two analysis results can be obtained with one injection. This type of system is cost effective, but from the analytical point of view not always recommended. Neither of the instruments can be used fully. If the GC/FTIR interface is of the light-pipe type, the concentration requirements are not met with most of the chemicals. In some instruments, the GC effluent is split 1 to 10 between MS and FTIR. The result can be that the sensitivity of the MS is lowered, but still FTIR is not sensitive enough.

The speed and sensitivity of the measurement needs a MCT detector to be used in a GC/FTIR instrument. Additionally, MI and cryodeposition techniques use at least liquid nitrogen to enable the trapping of the eluents. The need for cold liquids renders the GC/FTIR methods logistically difficult.

2.4.4 Derivatization

The most common derivatization methods in the analysis of chemicals related to the CWC are methylation with diazomethane (CAS 334-88-3) and silylation with either *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA; CAS 25561-30-2) or *N*-methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA; CAS 77377-52-7). The former method changes hydroxyls of acid groups to methoxy groups and the latter method changes hydrogens of hydroxyls to trimethylsilyl (TMS) or *tert*-butyldimethylsilyl (TBDMS) groups.[12] Figure 2 demonstrates the effect of the two derivatization methods on methylphosphonic acid (CAS 993-13-5).

The advantages of methylation are that the sizes of the derivatized chemicals do not increase very much

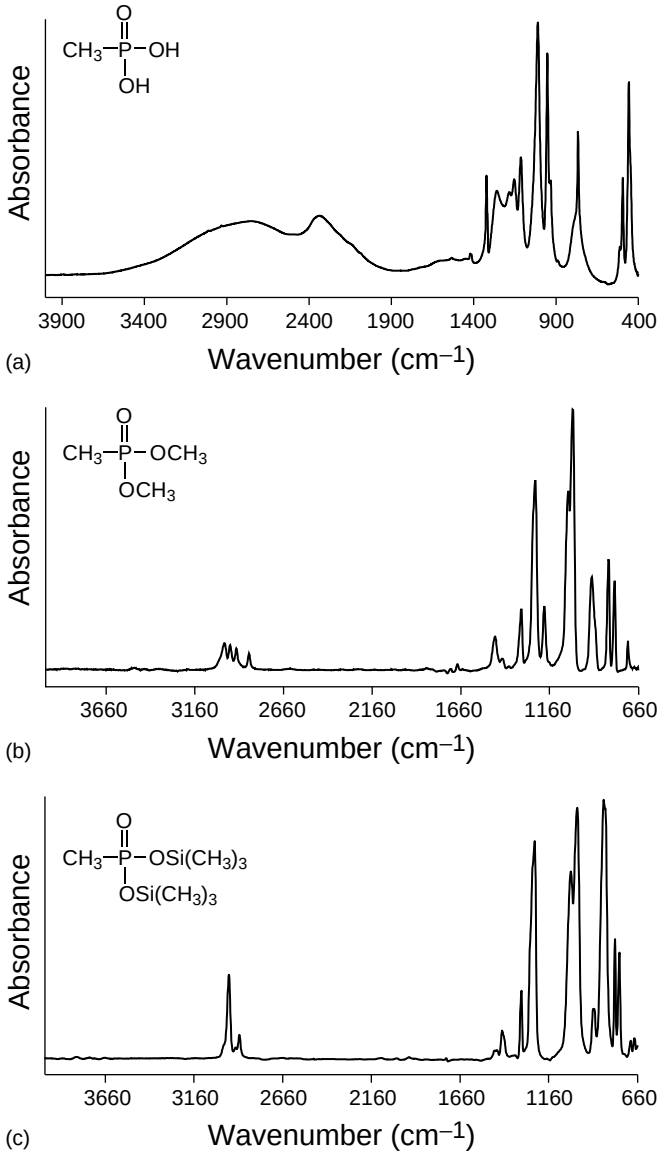

Figure 2 The effect of derivatization on methylphosphonic acid. (a) Methyl phosphonic acid analyzed as solid KBr pellets, (b) dimethyl methylphosphonate analyzed with cryodeposition-FTIR and (c) bis(trimethylsilyl) methylphosphonate (CAS 18279-83-9) analyzed with cryodeposition-FTIR. (See Söderström.[8])

and the dissolved derivatization reagent does not change the chromatogram very much. One of the disadvantages is that this method does not derivatize alcohols, which sometimes are difficult to analyze with nonpolar columns like SE-54 (5%-diphenyl–95%-dimethylsiloxane), which is commonly used in this type of analysis. Additionally, the group added to the molecule, methyl, is a naturally occurring group. It is impossible to know from the methylated sample alone if the chemical in it originally contained the hydroxyl or methoxy version of the

chemical. For confirmation of this, analysis of the underivatized sample is also required.

Silylation results in a more complicated sample than methylation, but it also shifts the derivatized chemicals to longer retention times. As the TMS group does not come from natural sources, the original chemical in a sample can be deduced without comparison with an untreated sample. Silylation enhances retention characteristics of chemicals much better than methylation. It also silylates active sites of the column material. However, when starting to analyze silylated samples with a cryodeposition system, the solvent delay time must be increased to avoid deposition of a large amount of early eluting residue of derivatization reagent present in the sample after the reaction. This results in loss of chemicals eluting simultaneously with the reagent.

Before an injection of a derivatized sample, care must be taken to ensure that the injection system is cleaned of derivatizable chemicals from previous injections. These might be derivatized by the derivatization reagent and elute giving false positive results. A good way to check the cleanliness of the system is to inject 1–2 μL of the derivatization reagent solution.

2.5 Other Techniques

There are some liquid chromatography/Fourier transform infrared spectroscopy (LC/FTIR) instruments commercially available, but there are so far no published applications on the analysis of chemicals related to the CWC. The analysis of water-soluble degradation products would be a suitable area for use of this method. It would not be logistically viable to take a LC/FTIR instrument on-site.

IR microscopy, diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy and photoacoustic spectroscopy (PAS) techniques may be suitable for some types of sample but the use of these methods for analyses of CWA has not been reported.

Promising studies have been published on the use of remote detection of CWA, but they are outside the scope of this article.

3 DATA EVALUATION

When IR is used for identification of chemicals from CWC-related samples, it is important to verify or at least compare the results with those from different analytical techniques. Especially, when environmental samples are analyzed using GC/FTIR, results should be compared with results from GC with selective detectors and GC/MS. These techniques are very powerful in chemical identification when used together in a collaborative manner. This requires the use of similar GC conditions and columns. It has been a standard practice to require consistent results from at least two different analytical methods to confirm unequivocal identification. These methods can include among others GC retention index monitoring, gas chromatography/electron ionization/mass spectrometry (GC/EI/MS), gas chromatography/electron ionization/high-resolution mass spectrometry (GC/EI/HRMS), gas chromatography/chemical ionization/mass spectrometry (GC/CI/MS), gas chromatography/chemical ionization/tandem mass spectrometry (GC/CI/MS/MS), GC/FTIR and NMR (^{1}H (proton), ^{13}C{^{1}H} (proton decoupled carbon-13), ^{19}F, ^{31}P, and ^{31}P{^{1}H} (proton decoupled phosphorus-31)).

One question regarding the data measured during an on-site analysis is the confidentiality of commercial and defense data. It is not in the interests of the chemical industry to allow analytical information on their processes to be made public. Such problems can be resolved by including only spectra relevant to the inspection in the instrument's database. However, if spectra are measured so that the analyst can see the spectra and print them, there always is a possibility that an experienced analyst can identify the chemical, even if its spectrum is not included in the instrument's database. A so-called blinded-software is used in the mobile mass spectrometer used by the OPCW to overcome this problem. This software has been developed in the National Institute of Standards and Technology (NIST) in the USA.[13] However, no such software has been developed for FTIR, so far.

After the analysis itself, there are several steps to the final result. In GC/FTIR the peaks of the analytes of interest must be located from the chromatogram. The chemicals must then be identified by comparison with existing data or if this is not possible the spectra should be interpreted so that reference material can be obtained or synthesized.

3.1 Gram–Schmidt Chromatogram

In GC/FTIR, the chromatogram representing the absorbance of each scan set versus retention time is called the Gram–Schmidt chromatogram (GSC). The GSC resembles the total ion chromatogram (TIC) in MS and a flame ionization detection (FID) chromatogram. It is not a simple sum of all absorbances of a spectrum, but calculated using interferograms as vectors.[14] This technique is faster than first performing a Fourier transform and then integrating the spectrum. Normally in GC/FTIR, the interval between consecutive data points has to be about one second to achieve the necessary chromatographic resolution. This one-second left for the calculation of a data point in a chromatogram used to be beyond the capabilities of the computer.

The GSC differs from a TIC in MS or a FID chromatogram. The relative intensities of the chromatographic peaks are sometimes very different in IR, due to differences in absorptivities between chemicals. For example, chemicals containing phosphorus–oxygen bonds have high absorptivities. Additionally, as the chromatographic resolution is lower in GC/FTIR, closely eluting peaks may overlap in IR, but not in MS or GC.

The software of some instruments uses a threshold for storing GC/FTIR data. This means that only data on peaks above a set threshold are stored. This approach is not generally recommended for analyses of environmental samples when the concentration of the target analytes is not known in advance. It is important also to be able to examine the spectra of minor components.

3.2 Functional Group Monitoring

A useful practical procedure in GC/FTIR is functional group monitoring. During the data collection or, in some systems, just after it, absorbance of certain selected spectral regions in each spectrum is plotted against the retention time to produce new chromatograms, functional group chromatograms (FGC). This enables the chromatographic peaks of interest to be selected in an easier way than by using the GSC.

Many chemicals related to the CWC contain functional groups which give rise to strong, characteristic absorption bands. These are summarized in Table 3.

The use of functional group monitoring will make it possible or at least easier to find the target chemicals in the midst of background chemicals. For example, likely background chemicals in samples from a battlefield or industry could be those from gasoline, diesel fuel, rubber and paint or even humic acids. These chemicals can mask the target chemicals by overlapping and make the sample too complex to be able to identify each peak in the chromatogram. Figure 3 shows an example of an extract of a rubber sample containing two mustard agents. In the functional group chromatogram 750–680 cm^{-1}, the peaks of the two agents can be seen clearly.

3.3 Reference Spectra and Spectral Libraries

In the FTIR analysis of the chemicals related to the CWC, the spectral libraries are essential. For the purposes of the CWC, there has to be a reference spectrum for identification. If there is no IR reference spectrum available, the identification cannot be accepted unless a reference compound can be synthesized and a reference spectrum measured.

3.3.1 Available Reference Data

There are no commercial libraries available which contain more than some occasional CWC-related chemicals. So far, each laboratory has been forced to create its own IR libraries but, since, 1994 the OPCW has collected

Table 3 Summary of the most characteristic functional groups of the chemicals related to the CWC. Many of these groups can be used as a basis for FGC. Wavenumber regions are based on data from Söderström[8] and Bellamy[15]

Chemical group	Functional group	Region (cm^{-1})	Strength
Nerve agents and their degradation products	−P=O and −P−O−C−	1350–990	s to vs
Sulfur and nitrogen mustards, aminoethylchlorides	−C−Cl	720–660	s
Degradation products of sulfur and nitrogen mustards	−C−OH	3400–3200[a]	vs
		3643–3630[b]	m
Lewisites, BZ and its degradation products, tear gases	=C−H	3150–3000	
BZ and its degradation products, some tear gases	−C=O	1750–1700	vs
TMS derivatives of hydroxyl-containing degradation products	$-C-O-Si(CH_3)_3$	1110–1080	vs
	$-P-O-Si(CH_3)_3$	1156–975	vs
		852–847	
Ethers and esters	−C−O−	1300–1000	vs
All	−C−H	3000–2800	s
Many of precursor chemicals and degradation products	−O−H	3400–3200[a]	m
	−N−H	3643–3600[b]	m
		3398–3381	w
		3344–3324	w
Aminoethylthiols	−S−H	2950–2550	w
BZ and its degradation products, tear gases	=C−H	3100–3000	w
	ring	2000–1660	w
		1600–1450	w
		1225–1000	w
		770–735	s to vs

[a] Intermolecular hydrogen bonded.
[b] Free OH (i.e. very low concentration or vapor phase). s = strong, vs = very strong, m = medium, w = weak.

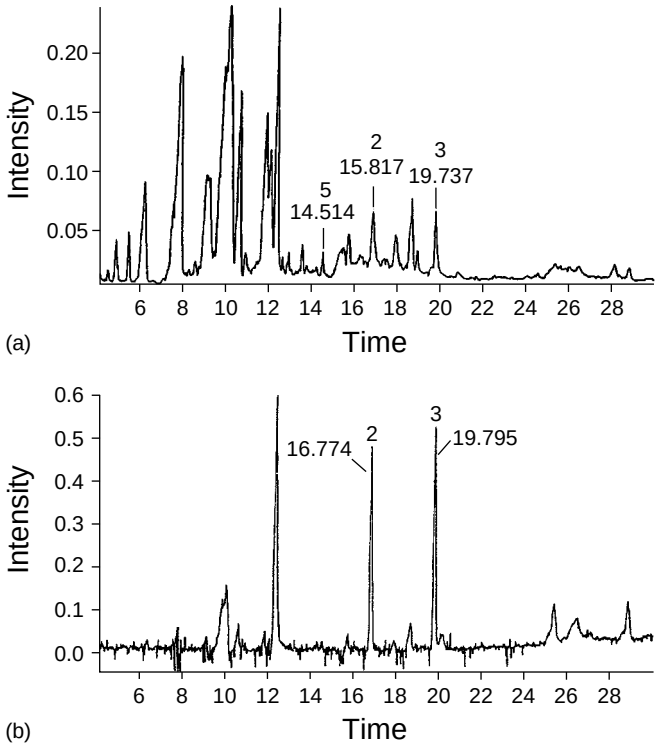

Figure 3 Chromatograms of an extract of a rubber sample containing two mustard agents (peaks marked with 2 and 3): (a) GSC and (b) functional group chromatogram 750–680 cm^{-1}. (Reproduced by permission of the Finnish Ministry for Foreign Affairs from (1992) Rautio,[16] Figure 16, p. 312.)

reference IR as well as mass and NMR spectra. These spectra are to be made available for states that are party to the Convention, i.e. countries which have ratified the CWC.

Some IR spectra of chemicals related to the CWC can be found in open literature. The largest collection of spectra can be found in two of the so-called Blue Books of the Finnish Institute for Verification of the Chemical Weapons Convention (VERIFIN) published in 1977 and 1982.[17,18] Blue Books describing the results of international interlaboratory comparison tests also contain some example spectra.[16,19–22] Shagidullin et al.[23] have published an IR atlas of organophosphorus chemicals which contains spectra of some relevant chemicals. Several articles contain IR spectra or spectral data for nerve agents and related chemicals[24–31] and vesicant related chemicals.[32–34]

3.3.2 Spectral Differences Caused by Sample Introduction

The phases of the chemicals measured in light-pipe and cryodeposition instruments are different: vapor and condensed phase, respectively. The intermolecular interactions are different in the two phases (i.e. missing in the vapor phase) and therefore, for example, all hydrogen bond-related vibrations are missing or different. Also, the vibration bands are narrower in the gas phase. Thus, the spectra cannot be compared with each other. Traditional IR spectra measured using salt pellets or windows also produce condensed phase spectra, which are therefore comparable with cryodeposition spectra (see example in Figure 4). There are other differences caused by factors of more practical nature: lower sensitivity and resolution. Owing to all these differences, two sets of reference spectra have to be measured.

The light-pipe spectra are usually measured at the resolution of 8 cm^{-1} with 4 co-added scans, whereas the cryodeposition are often rescanned to a resolution of 4 cm^{-1} with 64 to 512 scans. Figure 5 shows light-pipe

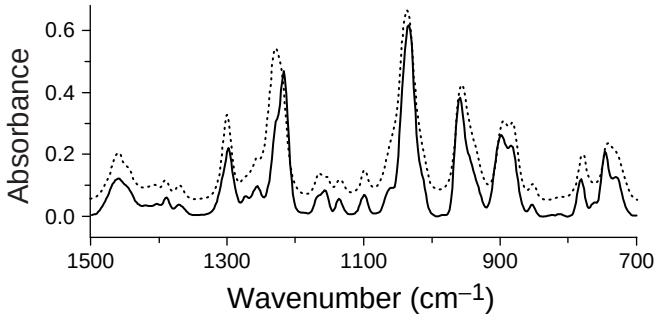

Figure 4 Part of cryodeposition (solid line) and liquid phase (dotted line) spectra of *O*-ethyl *S*-(2-(dimethylamino)ethyl)methylphosphonate. (Reproduced by permission of American Institute of Physics from M.T. Söderström, 'Identification of VX-type Nerve Agents Using Cryodeposition GC-FTIR', in *Fourier Transform Spectroscopy*, 11th International Conference, ed. J.A de Haseth, American Institute of Physics, New York, USA, 457–460, 1998.[28])

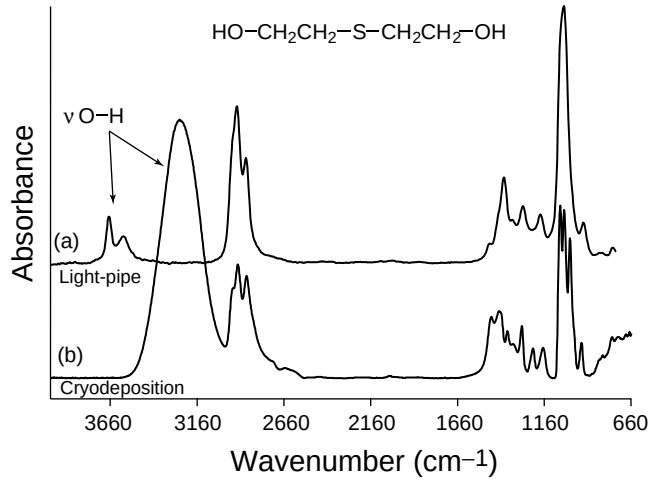

Figure 5 Spectra of thiodiglycol. (a) Light-pipe spectrum at 8 cm^{-1}, (b) cryodeposition spectrum at 4 cm^{-1}. (See Söderström.[8])

and cryodeposition spectra of thiodiglycol. MI spectra differ considerably from both light-pipe and cryodeposition spectra since the molecules are normally isolated molecules in noble gas matrix at very low temperatures. Thus, the spectral peaks are very sharp due to loss of intermolecular interactions and low rotation in the molecules.

More than 90% of the IR spectra collected by the OPCW are in the condensed phase. The largest commercial condensed phase IR library (Sadtler Condensed Phase IR Standards library) has over 75 000 spectra, while the largest vapor phase library (Sadtler Vapor Phase IR Standards library) has less than 10 000 spectra. There are no commercial MI libraries available.

3.3.3 Quality of Reference Spectra

To make the identification of chemicals unambiguous, some kind of criteria for acceptable reference spectra are needed. All reference spectra should be evaluated using acceptance criteria. The OPCW rules require that at least three evaluators agree on the spectrum before it is accepted. The evaluation criteria defined by the OPCW are the following:

- The spectrum must be consistent with molecular structure of the chemical.
- The spectrum must show the absence of extraneous spectral features attributable to impurities or contaminants.
- The sample preparation method for the measurement must be stated.
- The minimum resolution must be $4\,cm^{-1}$ for condensed phase and normal gas phase spectra, and $8\,cm^{-1}$ for GC light-pipe spectra.
- The minimum spectral range must be at least $3700-700\,cm^{-1}$.
- The signal-to-noise ratio must be adequate to detect all relevant peaks.
- The largest absorbing peak must not be saturated.
- The spectra must be adequately compensated for atmospheric carbon dioxide and water.
- The intensity of the bands arising from water contained in the sample should be less than 5% of the highest intensity absorption band of the sample.
- For inclusion in the central OPCW Analytical Database, a spectrum must fulfil at least one of the following criteria:

  (a) It is consistent with IR data of the same chemical from another source.
  (b) It is supported by accepted NMR or MS spectra obtained from the same sample with consistent results.

For the submitted reference spectra the OPCW requires that the following data is attached to the spectra:

- OPCW code for the spectrum (assigned by the OPCW),
- contributor's name and address,
- contributor's signature,
- chemical information,
  - chemical name
  - CAS registry number (if available)
  - chemical structure
  - molecular formula (optional)
- sample information,
  - sample purity (optional)
  - source (optional)
  - sample phase
- instrument information,
  - type (Fourier, grating or prism)
  - manufacturer
  - model
  - data system (optional)
  - Software version (optional)
- experimental information,
  - sampling mode (liquid, solid, solution, gas, light-pipe, cryodeposition, bulk, micro)
  - measurement mode (transmission, absorbance, reflectance)
  - baseline correction (manual, automatic, none)
  - matrix
  - detector (optional)
  - wavenumber range
  - indication of the ordinate scale
  - resolution
  - number of scans (optional)
  - date of experiment (optional).

3.4 Library Search

Searching the spectrum of an unknown chemical against a spectral library is routine method used to identify chemicals. Most of the commercial IR instruments include library search software which has several search algorithms to choose from. The search algorithm can sometimes have a very strong effect on the library search result. This is because of the different ways in which the actual comparison between the spectra is done. Particularly when the library and the unknown spectra have been measured differently (e.g. using solid KBr disk and cryodeposition GC/FTIR), the IR band shapes and positions may vary slightly. In these cases, the choice of proper search algorithm is essential.

Two of these search algorithms are least-squares metric (also called Euclidean) and derivative least-squares metric (also called derivative Euclidean).[35] In both algorithms the unknown and library spectra are normalized before the comparison. In the least-squares metric algorithm, the intensity difference between the corresponding data points in the spectra is raised to the second power. The sum of these differences is the hit quality index (HQI). The derivative least-squares metric algorithm is based on differences in the intensity change between two consecutive data points instead of the intensity of each point. There are many ways to apply the algorithms and to present the HQI, so that it is impossible to give any estimates for a value for a good library match. The derivative least-squares metric algorithm is much more sensitive to changes in band position, which means that it should be used when the phase of the unknown and the library spectra are the same.

Figure 6 demonstrates an IR library search. The unknown chemical was dimethylethylphosphonate (DMEP) measured using cryodeposition GC/FTIR. The library was a cryodeposition spectrum library. If the derivative least-squares metric algorithm (HQI can be 0–999; 999 is perfect match) is used the result is DMEP (HQI = 558), dimethylpropylphosphonate (DMPP) (HQI = 275), dimethylisopropylphosphonate (DMIP) (HQI = 246) and dimethylmethylphosphonate (DMMP) (HQI = 216) as shown in Figure 6. The least-squares metric search gives a slightly different result: DMEP (HQI = 861), DMPP (HQI = 744), DMMP (HQI = 632) and DMIP (HQI = 607). Note that the first case gives a much better indication of the correctness of the search result, by giving a bigger difference between the first and the second hit.

The search result should never be accepted based just on the result listing, but should always be verified visually. It is therefore necessary for the analyst to have at least a basic knowledge of which spectral features to look for. When the analyst is required to differentiate between very closely related chemicals, e.g. nerve agent homologs, more knowledge of the particular chemicals and the origin of the spectral bands is a prerequisite.

Identification in spectra of mixtures is always somewhat difficult. Spectral subtraction can sometimes be used to reduce the complexity of the spectrum by removing the features of the subtracted spectrum. Subtraction seldom gives clean spectra but it can at least help the final identification.

3.5 Interpretation

For the purposes of the CWC, results obtained with two spectrometric techniques are required for identification

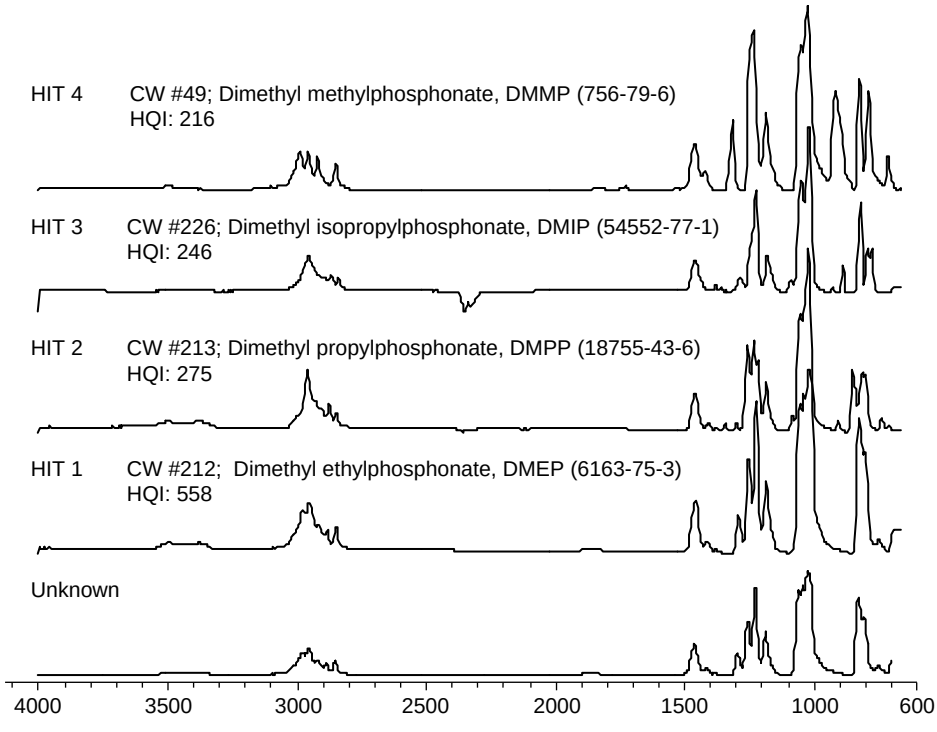

Figure 6 Example of library search. An unknown chemical was searched against a library containing chemicals related to the CWC. The first hit, DMEP, was clearly the best candidate and matches the unknown. The search algorithm was a square derivative and the maximum HQI value is 999. (See Söderström.[8]) CAS numbers are given in parentheses.

of a compound. Although very useful as such, the interpretation of IR spectra does not identify the chemical but may give tentative structures. Interpretation gives excellent results in some cases, but it should always be remembered that the accuracy of interpretation depends on the type of the chemical and the experience of the scientist. In the case when an unknown compound is encountered, spectral interpretation (MS, IR, and NMR) is required so that the reference chemical can then be synthesized.

In the interpretation of IR spectra of chemicals related to the CWC, it should be remembered that for some chemicals, e.g. nerve agents, a structure could be proposed very easily and accurately. For some other types of chemicals, e.g. mustards and other vesicants, only the chemical class can be proposed. This is because some functional groups do not have good structure-specific group frequencies. Such spectra can only be used as a fingerprint for the chemical.

3.5.1 Phosphorus-containing Chemicals

L.C. Thomas and R.A. Chittenden[4,36–42] have carried out a thorough series of studies on the identification of organophosphorus compounds. The group frequency tables in these studies enable interpretation of many of the characteristic features in the spectra of nerve agents and related chemicals. Several structure–spectrum relationships give specific information on the molecule. One very valuable relationship discovered by Thomas is the dependence of the position of the P=O bond stretching vibration (in wavenumbers), $\nu_{P=O}$, on the substituents on the phosphorus, represented by π constants (Equation 1):

$$\nu_{P=O} = 930 + 40 \sum \pi \qquad (1)$$

The π constants have been determined experimentally from over 900 compounds for a wide variety of substituents (91 substituents). Table 4 summarizes the most important groups that are present in the chemicals relevant to the CWC and their π-values. Most of the calculated values are within $\pm 12\,cm^{-1}$ from the experimental values. For example, for sarin the calculated value is $1280\,cm^{-1}$, while the experimental values are $1277\,cm^{-1}$ and $1272\,cm^{-1}$ for liquid and cryodeposition spectra, respectively.

Table 4 Summary of values of π constant (see Equation 1) for the most important substituents present in the chemicals relevant to the CWC (see Thomas[4])

Group	π	Group	π
CH_3	2.1	$O-SiR_3$	3.0
CH_2	2.0	C≡N	3.5
CH	1.8	S–alkyl	2.4
C	2.1	NR_2	2.4
OCH_3	2.9	Cl	3.4
OCH_2	2.85	F	3.9
OCH	2.75		

The alkyl groups connected to phosphorus or to oxygen next to phosphorus can be identified in many cases. One of the most characteristic bands is the P–Me deformation near $1300\,cm^{-1}$.[4] There is also the P–Me rocking band near $900\,cm^{-1}$.[4] The bands of the other P–alkyl group do not give characteristic bands but can be identified by comparison with spectra of similar chemicals.[28,43] Bands for P–O–Me (near $1180\,cm^{-1}$) and P–O–iPr (three bands, $1100–1200\,cm^{-1}$) are characteristic.[4] P–O–alkyl vibrations can also be identified by comparing patterns produced by known groups to those in the unknown spectrum.[43] A summary of some of these P–O–alkyl group patterns is shown in Figure 7.

In phosphonofluoridates, there is always the P–F stretching band. The range given by Thomas[4] for this vibration in sarin analogs is $858–833\,cm^{-1}$.

Some of the chemicals related to the CWC contain tertiary amine groups. The type of alkyl group (methyl, ethyl, propyl or isopropyl) can be deduced in some cases from the stretching of the C–H next to the nitrogen atom. The value of the asymmetric stretching is lowered due to the interaction with lone pair electrons of the nitrogen atom,[15] so that the bands of this vibration are separated from the normal C–H stretching bands. Figure 8 shows some typical peaks of this kind in the VX-type chemicals.[28]

One very characteristic peak for tabun-related chemicals is the stretching of the C≡N bond. A sharp band due to this vibration can be seen at $2200 \pm 2\,cm^{-1}$.[43] There is always a band in the spectra of tabun homologs, due to the P–N–Me group near $1320\,cm^{-1}$.[4] This band occurs very close to the P–Me deformation band in, for example, sarin, homologs. In the tabun homologs, however, the peak near $900\,cm^{-1}$ due to P–Me rocking is missing.

Often, IR spectra can be only partially interpreted. An example of interpretation of the spectrum of sarin is shown in Figure 9. In the interpretation of a spectrum of organophosphorus chemical, it is very important first to deduce the type of chemical. In the spectrum presented, the position of the P=O stretching band ($1272\,cm^{-1}$) is typical of sarin-type chemicals. A very strong band typical for P–O–C is present at $1020\,cm^{-1}$. When there is also a peak matching the P–F stretching at $838\,cm^{-1}$ as well as the P–Me peaks at $1322\,cm^{-1}$ and $927\,cm^{-1}$, the chemical seems to be an alkyl methylphosphonofluoridate. As the peak pattern in

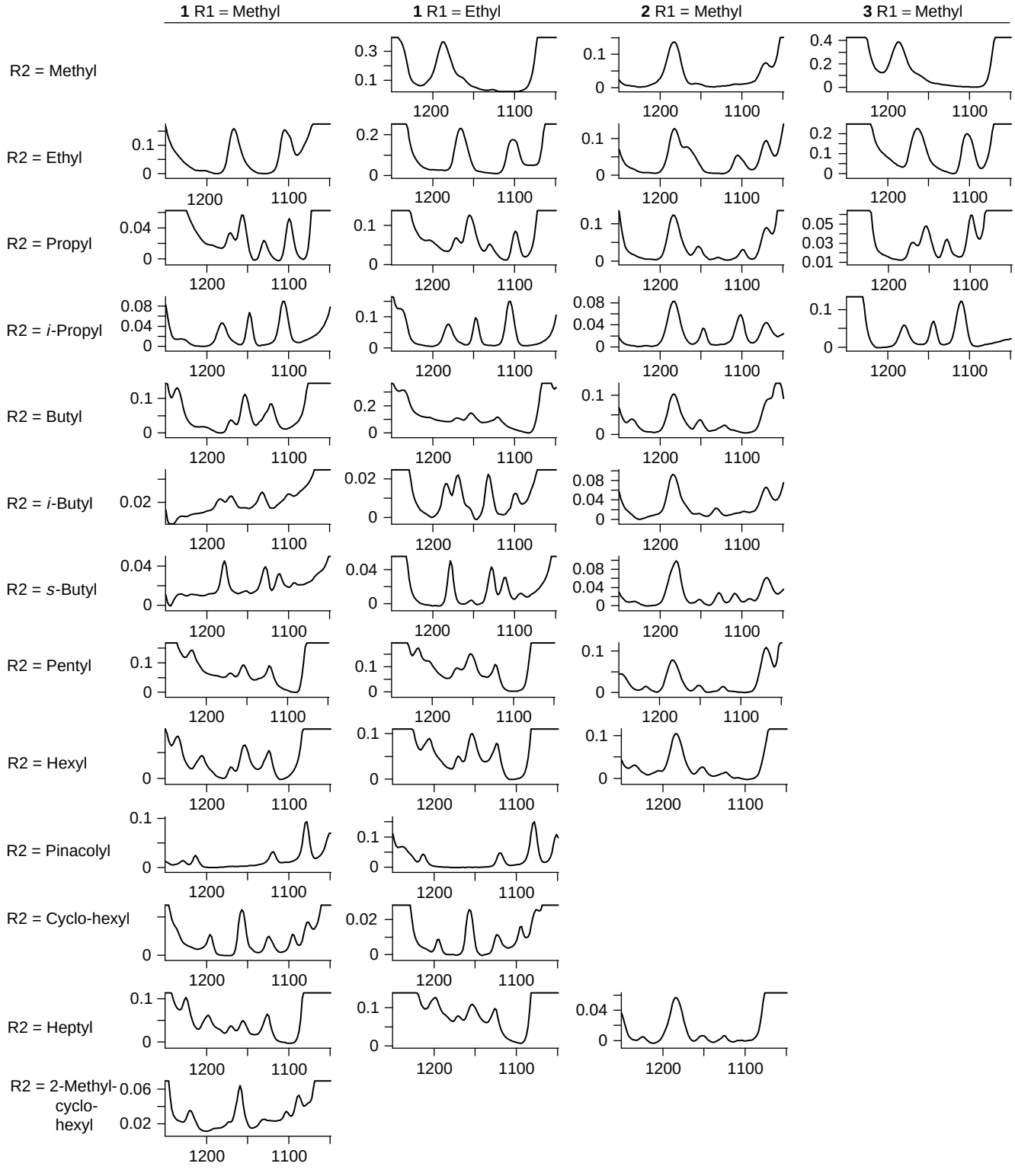

Figure 7 A summary of some P−O−alkyl groups. R1 is P−alkyl and R2 is P−O−alkyl). The compared compounds belong to the sarin (**1**; alkyl alkylphosphonofluoridates), the tabun (**2**; dialkylamino alkyl phosphoramidocyanidates) and the dialkyl methylphosphonate (**4**) families. There are two peaks (1185–1180 cm^{-1} and 1072–1070 cm^{-1}) in the tabun family (**2**) spectra in addition to the C−O peaks. These peaks are due to the C−N bond. The dialkyl methylphosphonate compounds (**4**) listed here are symmetric (R^2=$R^{2'}$). (Reproduced by permission of Springer-Verlag from M.T. Söderström, R.A. Ketola, *Fresenius' J. Anal. Chem.*, **350**, 162–167 (1994).[43])

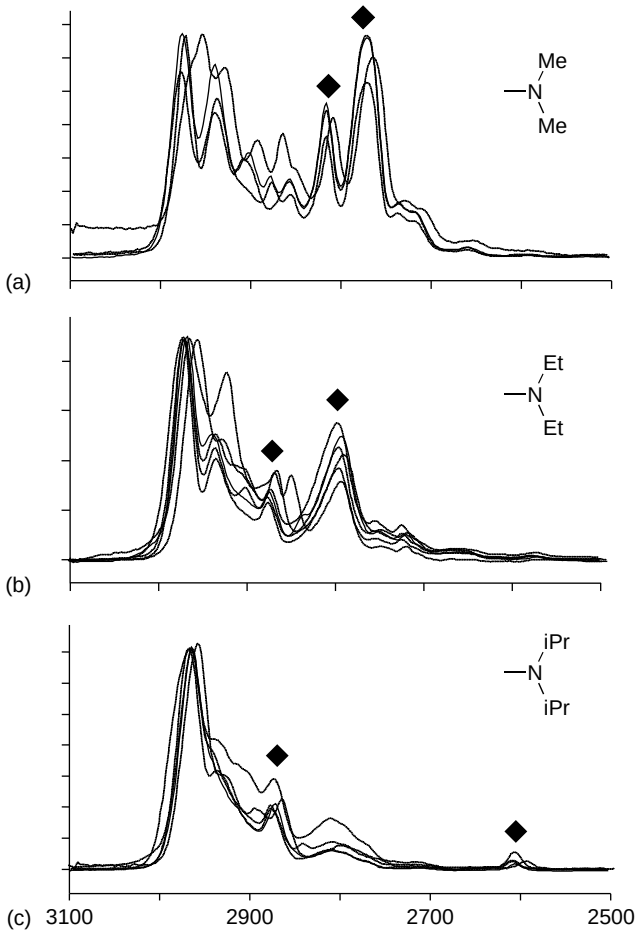

Figure 8 Typical bands (marked with ◆) due to the asymmetric C–H next to a nitrogen atom in some VX-type chemicals: chemicals containing group (a) *S*-[2-(dimethylamino)ethyl], (b) *S*-[2-(diethylamino)ethyl], and (c) *S*-[2-(diisopropylamino)-ethyl]. Several VX-type chemicals are shown overlaid. (Reproduced by permission of American Institute of Physics from M.T. Söderström, 'Identification of VX-type Nerve Agents Using Cryodeposition GC-FTIR', in *Fourier Transform Spectroscopy*, 11th International Conference, ed. J.A. de Haseth, American Institute of Physics, New York, USA, 457–460, 1998.[28])

the region 1250–1050 cm^{-1} matches the P–O-isopropyl patterns shown in Figure 7, it is proposed that the structure is isopropyl methylphosphonofluoridate, i.e. sarin.

In the analysis of chemicals related to the CWC, spectral interpretation is not done with only one type of spectra. Often MS or NMR spectra are used together with IR spectra in the elucidation of the structure of an unknown chemical. There are some examples of this type of elucidation available in the literature.[4,25,26,44]

One must remember that there are important spectral bands, which do not fit in the narrower detector range (e.g. 4000–700 cm^{-1}) of a GC/FTIR instrument with a MCT

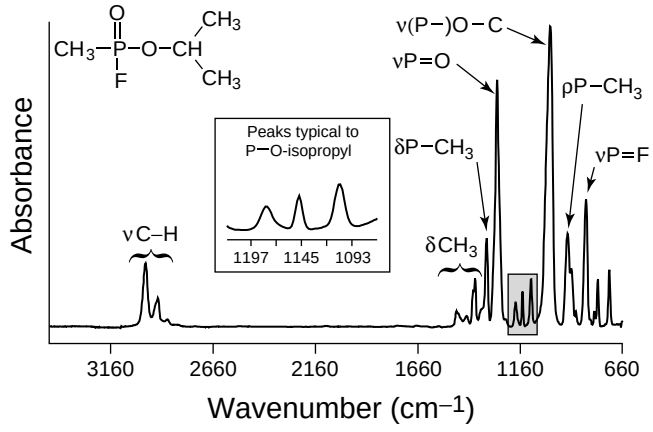

Figure 9 A partially interpreted cryodeposition spectrum of sarin. (See Söderström.[8])

detector. In CWC-related organophosphorus chemicals, e.g. the bands due to P–S–C and P–Cl bonds, are at too low frequencies to be seen in the spectra. The P–S–(C) frequency is 573–520 cm^{-1} in VX-type chemicals and the general P–Cl frequency is 587–420 cm^{-1}.[4]

3.5.2 Nonphosphorus Chemicals

Nonphosphorus chemicals related to the CWC are more or less difficult to interpret. There are functional groups with characteristic vibrations in many chemicals, but the structure cannot be elucidated fully by interpreting the IR spectra as is the case with e.g. nerve agents. Table 3 contains some of the characteristic vibrations in nonphosphorus chemicals related to the CWC.

A partially interpreted cryodeposition spectrum of 2-(diisopropylamino)ethanol (CAS 96-80-0) is presented in Figure 10. In this case, the full structure cannot be deduced from the spectrum. There are two

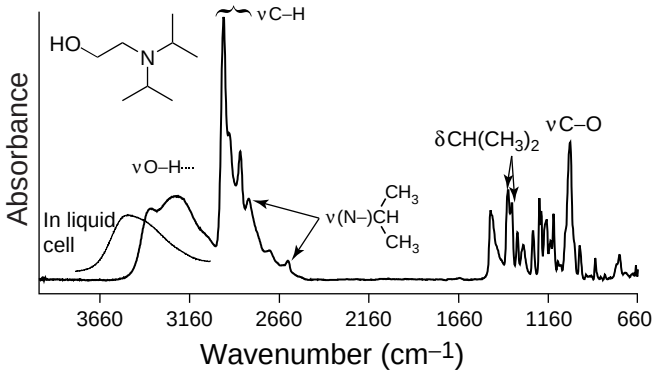

Figure 10 A partially interpreted cryodeposition spectrum of 2-(diisopropylamino)ethanol. Additionally the O–H stretching band in a spectrum measured using a liquid cell is shown. (See Söderström.[8])

bands indicating the presence of C−OH group in the molecule: stretching of C−O bond at 1035 cm^{-1} and the double peak of the hydrogen-bonded O−H stretching at 3375 and 3230 cm^{-1}. In the tracer spectra, the hydroxyl stretching band seems to appear as two bands in the symmetric dialkylaminoethanols.[8] In asymmetric dialkylaminoethanols and in liquid phase spectra (band shown in Figure 10) the vibration appears as one band. The presence of isopropyl can be detected from the presence of two bands at 2875 and 2607 cm^{-1} and two bands at 1382 and 1361 cm^{-1}. The former peaks are due to the (N−)C−H stretching, as in VX-type chemicals (shown in Figure 8) and the latter peaks to the methyl stretchings in the $-CH(CH_3)_2$ group. The lack of other characteristic bands indicates the presence of a scheduled chemical, 2-(diisopropylamino)ethanol. Further confirmation for this is, however, required.

4 APPLICATIONS

There are relatively few recent articles published on the FTIR analysis of CWC-related chemicals. Most of the research is old or unpublished.

4.1 International Tests

One of the best ways of testing methods and competence in the field of CWC-related analysis have been the international tests organized first by VERIFIN and later by the OPCW. Eleven tests were carried out during 1989–1998.[16,19–22,45–48] FTIR has not been used by many laboratories in the tests, as the laboratories with fewer resources have been concentrating on GC/MS analysis, e.g. because of the lack of GC/FTIR instruments. The maximum number of GC/FTIR instruments used in the test was eight out of 26 laboratories. On average, laboratories using cryodeposition GC/FTIR instruments have performed better GC/FTIR analysis. In the beginning, some laboratories used nonhyphenated FTIR instruments but the results were poor when analyzing environmental samples. It should be noted that in the interlaboratory comparison test in 1991,[20] FTIR was used successfully without GC separation. The test samples were not environmental but organic solution samples, for which FTIR is very suitable.

International tests have shown that a GC/FTIR instrument is not a necessity for a successful analysis because two other techniques (GC/EI/MS, GC/CI/MS, NMR or even GC retention index monitoring) can be used as the required alternative techniques. The major advantages of GC/FTIR are that it gives information on different characteristics of a chemical than, for example, MS and NMR and it can make the overall spectral elucidation much easier and faster, and the first ''guesses'' more accurate (see Söderström et al.[25] for examples). The latter advantage is emphasized when analyzing phosphorus chemicals.

In international tests false identifications have occurred on the basis of misinterpretation of MS or NMR spectra only, e.g. isopropyl methyl ethylphosphonate (CAS 141968-53-8) has been identified instead of methyl propyl ethylphosphonate (CAS 170082-59-4),[46] and methyl propyl ethylphosphonate instead of the methylation product of butylphosphonic acid (i.e. dimethyl butylphosphonate; CAS 24475-23-8).[22] These mistakes were due to matching molecular weight, similarity of the spectra and the lack of reference data, and could have been avoided by using additional techniques, such as GC/FTIR.

There are of course limitations to interpretation capability with some chemical types. For example, spectra of mustard-gas-type chemicals cannot be interpreted to produce an unambiguous analysis result. In the first official OPCW proficiency test,[45] there was a sample where three mustard-type chemicals (spectra shown in Figure 11) were present; sesquimustard (CAS 3563-36-8), 1,3-bis[(2-chloroethyl)thio]propane (CAS 63905-10-2) and 1,5-bis[(2-chloroethyl)thio]butane (CAS 142868-94-8). In this case, reference spectra existed only for the first two compounds. The identity of the third could not be proven (only inferred) because a reference spectrum was not available. The spectrum and retention time definitely indicated the correct answer but they are not enough for an unambiguous identification.

4.2 Attenuated Total Reflection Applications

Braue and Pannella[10] used a flow-through ATR cell for quantitative measurement of tabun, sarin and soman in water. The sample change and measurement could be performed safely, owing to quite dilute samples and airtight fittings. The method they developed has the useful range of 0.5–2.0 mg mL^{-1}. Quantitation is based on P=O stretching near 1240 cm^{-1}.

Parchment et al.[49] used a diamond ATR probe to monitor decontamination of mustard gas, sarin, soman and VX. They conclude that because of the relatively low sensitivity of the method, it can only be used for analysis of CWA in realistic concentrations in real time and in situ.

The breakthrough of mustard gas and mustard ether (CAS 63918-89-8) in different types of rubbers[50] and polypropylenes[51] has been studied using ATR as one of the analysis techniques. One side of the material was exposed to the agents while the unexposed side was against the ATR crystal.

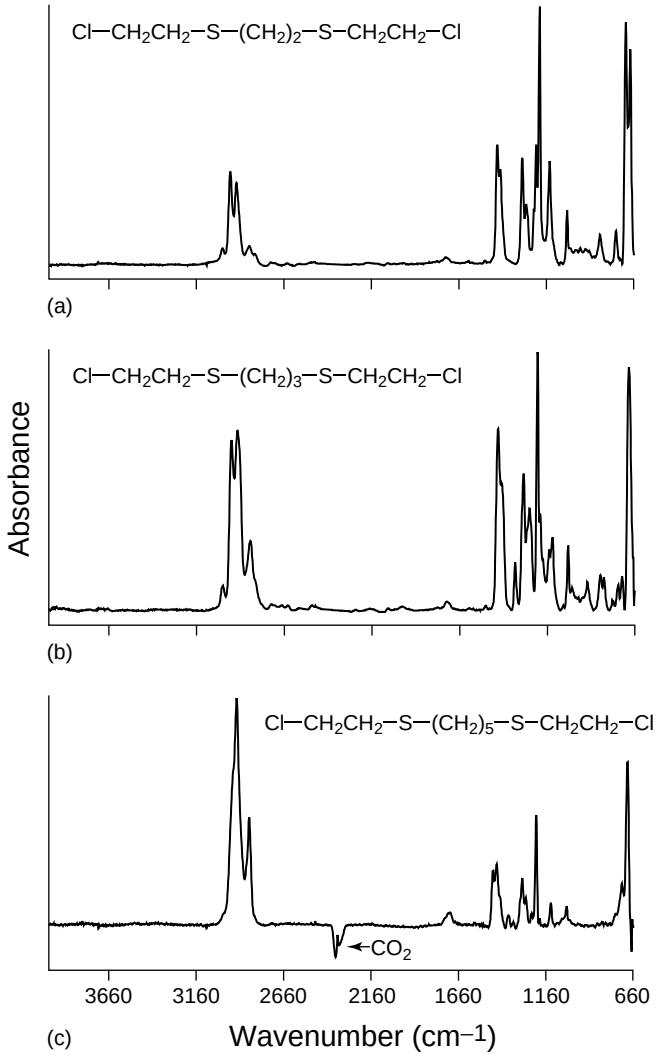

Figure 11 Cryodeposition-FTIR spectra of (a) sesquimustard, (b) 1,3-bis[(2-chloroethyl)thio]propane and (c) 1,5-bis[(2-chloroethyl)thio]butane. (See Söderström.[8])

4.3 Gas Chromatography/Fourier Transform Infrared Applications

Both Söderström et al.[25] and Creasy et al.[26] have applied GC/FTIR along with several other hyphenated chromatographic techniques to analyze chemical weapons-related chemicals in complex matrices. The results from the different techniques have been combined to identify unequivocally the relevant chemicals in low concentrations. In both studies, IR and MS spectra have been used together in spectral interpretation. Weimaster et al.[52] studied samples collected in Iraq, using a wide variety of instrumentation, but they could not find any scheduled chemicals.

Danian et al.[27] report structural identification of a VX-type chemical using GC/MS, NMR and GC/FTIR. They concluded that the identification of the chemical based only on MS data is very difficult. The use of NMR and FTIR data makes the identification more reliable.

Durst et al.[29] performed a microscale synthesis of several sarin analog autosampler vials and analyzed them directly with a gas chromatography/infrared spectroscopy/mass spectrometer (GC/IR/MS) instrument. They presented a couple of light-pipe IR spectra and a table of typical gas phase frequencies for 12 sarin-type chemicals.

Sokołowski and Szymańska[30] analyzed tabun, which had been stored at least 20 years, with a GC/IR/MS instrument. The sample was analyzed first in the underivatized form. Then portions of it were methylated or silylated and analyzed. They reported nine light-pipe IR spectra of derivatized degradation products.

Goeringer and Elizy[31] used GC/MI/FTIR to analyze alkylphosphonic acids from simulated river water samples. The chemicals were silylated before the analysis.

Smith et al.[34] studied three isomers (*cis*-, *trans*- and *geminal*) of lewisite using GC/MS, GC/FTIR and NMR.

Söderström et al.[43,53] studied interpretation of spectra of some organophosphorus chemicals as well as their IR and retention index searches. Some example spectra, chromatograms and interpretation are given.

5 QUALITY CONTROL

One aspect in the work of the OPCW is the use of designated laboratories where samples collected during an inspection could be sent for analysis. To be able to rely on the results provided by the designated laboratories, their performance must be rigorously tested. This is the purpose of the OPCW proficiency tests. In order to become a designated laboratory the laboratory must have a national accreditation (ISO/IEC Guide 25, EN 45001 or equivalent). If FTIR is to be used in accredited analysis, the laboratory must have written procedures for testing the performance of the instrumentation and for the analysis. The purpose of these tests is to guarantee high quality and reliability of the results.

The procedures can be adapted from standards or from the instrument manufacturer, and they can be defined by the laboratory.

For example, ASTM has described a two-level procedure for testing the performance of an FTIR spectrometer.[54] The routine check of instrument performance, the zero-level test, includes the following tasks:

- measurement and comparison of background against a previously measured background,
- measurement of a hundred per cent line (HPL) and

- measurement and comparison of the spectrum of a polystyrene sample against a previously measured spectrum.

The more extensive tests, level-one tests, include six tests which facilitate the diagnostics of the instrument:

- energy spectrum test
- HPL
- stability test
- signal-averaging test
- polystyrene sample test and
- photometric jitter test.

The performance of the GC must be tested regularly, e.g. daily, to ensure that the column retains no previously analyzed chemicals. When samples containing free acids are analyzed, e.g. soil samples containing humic acids, the column should be tested before other types of samples. When the column performance degrades, the column should be changed, or at least part of the column from the beginning should be replaced with a new piece of column. In order to lengthen the life of the columns, especially because of acidic injections, an uncoated inactivated precolumn can be used before the actual analytical column. There is a published Recommended Operating Procedure (ROP) for testing the performance of a GC system in the analysis of chemicals related to the CWC.[12]

All the solvents used must be checked before the analysis, and during the analyses relevant solvent blanks must be used to test the cleanliness of the system. When analyzing derivatized samples, it is not enough to use just the solvent; the derivatizing chemicals should also be added to the solvent to ensure that there are no derivatizable chemicals in the system before the actual analysis.

If a blank is supplied for each sample, as often has been the case in international tests so far, they should be analyzed before the samples. It should be remembered that in realistic cases no blank samples are available or that the blank and the sample have totally different backgrounds, e.g. water coming in to a factory (blank) and wastewater (sample with possible chemicals).

There should also be some criteria for the quality of spectra. OPCW has used criteria for the reference spectra (see section 3.3.3). These criteria can also be used for sample spectra where applicable.

6 COMPARISON WITH OTHER ANALYTICAL METHODS

In an analysis of a complicated sample, GC/FTIR alone cannot fulfill the identification criteria, just like any other analytical technique. Results from one technique can be correct but consistent results from other techniques are required for the final proof of the correctness.

Compared with GC/FTIR, GC/EI/MS gives comparable chromatograms, but is generally a more sensitive technique. The structural information provided by GC/FTIR and GC/EI/MS differs, so that they complement each other well. For example, there are cases when EI/MS (electron ionization/mass spectrometry) cannot differentiate between isomers but FTIR can, and cases where FTIR cannot give the exact size of an alkyl group but EI/MS can. These techniques also give overlapping structural information, thus increasing the reliability of the identification. Normally, commercial EI/MS spectral libraries are much cheaper and larger than FTIR spectral libraries. Therefore, it is easier routinely to identify chemicals present in the samples by GC/EI/MS and disregard the uninteresting chemicals after a match has been found from a library.

GC/CI/MS usually gives information on molecular mass. It can also give more information on some substituents if CI/MS/MS (chemical ionization/tandem mass spectrometry) is used. The molecular weight information becomes more important as the size of the molecule increases, since FTIR cannot easily distinguish between the sizes of longer alkyl groups.

Other mass spectral techniques which use liquid chromatography (LC) and capillary electrophoresis (CE) as the sample introduction method make it possible to analyze chemicals which should otherwise be derivatized for GC analysis and also those nonvolatile and nonderivatizable chemicals which cannot be analyzed at all with GC. Many of these chemicals could be analyzed with FTIR without GC separation, but, in the environment, they may be in, e.g. water or soil samples (which possibly have to be extracted with water). Water samples are difficult to analyze with FTIR, since water is quite a poor solvent for FTIR due to very high molar absorptivity.

NMR is one of the most important tools in structural elucidation, but since the technique cannot be hyphenated with GC like MS and FTIR, and its sensitivity is poorer, it cannot always identify chemicals from high background samples. The unchallenged capability of NMR to give information on connectivity and neighboring atoms makes it a very desirable method in the analysis of chemicals related to the CWC. NMR can for instance normally find out the type of the carbon (methyl, ethyl, propyl, or isopropyl) directly connected to phosphorus (^{1}H and ^{31}P spectra) and the presence of the fluorine–phosphorus link (^{1}H, ^{19}F, and ^{31}P spectra). Different two-dimensional-NMR experiments give additional information on connected atoms.

FTIR is a valuable addition to the instrumentation for a laboratory which is expanding its capabilities beyond just the basic necessities (GC and GC/MS). The effective

and complementary combination of analytical techniques based on different principles, e.g. as in IR, MS, and NMR, gives a solid base for any type of instrumental analysis. It is essential for the analysis related to the CWC that the analysis results are reliable and indisputable.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society of Testing and Materials
ATR	Attenuated Total Reflectance
BSTFA	*N*,*O*-Bis(trimethylsilyl)trifluoroacetamide
BZ	3-Quinuclidyl Benzilate
CAS	Chemical Abstracts Service
CE	Capillary Electrophoresis
CI/MS/MS	Chemical Ionization/Tandem Mass Spectrometry
CR	Dibenz[*b*,*f*][1,4]oxazepine
CWA	Chemical Warfare Agents
CWC	Chemical Weapons Convention
DMEP	Dimethylethylphosphonate
DMIP	Dimethylisopropylphosphonate
DMMP	Dimethylmethylphosphonate
DMPP	Dimethylpropylphosphonate
DRIFT	Diffuse Reflectance Infrared Fourier Transform
EI/MS	Electron Ionization/Mass Spectrometry
FGC	Functional Group Chromatograms
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared Spectroscopy
GC	Gas Chromatograph
GC/CI/MS	Gas Chromatography/Chemical Ionization/Mass Spectrometry
GC/CI/MS/MS	Gas Chromatography/Chemical Ionization/Tandem Mass Spectrometry
GC/EI/HRMS	Gas Chromatography/Electron Ionization/High-resolution Mass Spectrometry
GC/EI/MS	Gas Chromatography/Electron Ionization/Mass Spectrometry
GC/FTIR	Gas Chromatography/Fourier Transform Infrared Spectroscopy
GC/IR/MS	Gas Chromatography/Infrared Spectroscopy/Mass Spectrometer
GC/MI/FTIR	Gas Chromatography/Matrix Isolation/Fourier Transform Infrared Spectroscopy
GC/MS	Gas Chromatography/Mass Spectrometry
GSC	Gram–Schmidt Chromatogram
HPL	Hundred Per Cent Line
HQI	Hit Quality Index
IR	Infrared
LC	Liquid Chromatography
LC/FTIR	Liquid Chromatography/Fourier Transform Infrared Spectroscopy
MCT	Mercury–Cadmium–Telluride
MI	Matrix Isolation
MS	Mass Spectrometry
MTBSTFA	*N*-Methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
OPCW	Organization for the Prohibition of the Chemical Weapons
PAS	Photoacoustic Spectroscopy
PFIB	1,1,3,3,3-Pentafluoro-2-(trifluoromethyl)-1-propene
ROP	Recommended Operating Procedure
TBDMS	*tert*-Butyldimethylsilyl
TIC	Total Ion Chromatogram
TMS	Trimethylsilyl
VERIFIN	Finnish Institute for Verification of the Chemical Weapons Convention
VX	*O*-Ethyl *S*-2-(diisopropylamino)-ethyl Methylphosphonothioate

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)** Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)** Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Pesticides **(Volume 7)**
Organophosphorus Pesticides in Water and Food, Analysis of

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Gas Chromatography/Infrared Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Liquid Chromatography/Infrared Spectroscopy • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. R. White, *Chromatography/Fourier Transform Infrared Spectroscopy and Its Applications*, Marcel Dekker, New York, USA, 1990.
2. B. Stuart, *Modern Infrared Spectroscopy*, John Wiley & Sons, Chichester, UK, 1996.
3. P.R. Griffiths, J.A. deHaseth, *Fourier Transform Infrared Spectroscopy*, John Wiley & Sons, New York, USA, 1986.
4. L.C. Thomas, *Interpretation of the Infrared Spectra of Organophosphorus Compounds*, Heyden & Sons, London, UK, 1974.
5. ASTM Standard E 1252-94, 'Standard Practice for General Techniques for Qualitative Infrared Analysis', American Society for Testing and Materials, Philadelphia, PA, USA, February 1995.
6. G. Dent, 'Preparation of Samples for IR Spectroscopy as KBr Disks', *Internet J. Vibrat. Spectrosc.*, 1 edition, **1**, 2–5 (1997). (http://www.ijvs.com/volume1/edition1/section1.html#page2.)
7. W. Maddams, 'The Background to Sample Preparation for Infrared Transmission Measurements on Solids', *Internet J. Vibrat. Spectrosc.*, 1 edition, **1**, 6–11 (1997). (http://www.ijvs.com/volume1/edition1/section1.html #page6.)
8. M. Söderström, unpublished results.
9. Jane's NBC Protection Equipment, 10th edition, ed. T.J. Gander, Jane's Information Group Ltd., Coulsdon, UK, 141–179, 1997.
10. E.H. Braue, Jr, M.G. Pannella, 'CIRCLE CELL® FTIR Analysis of Chemical Warfare Agents in Aqueous Solutions', *Appl. Spectrosc.*, **44**(9), 1513–1520 (1990).
11. ASTM Standard E 1642-94, 'Standard Practice for General Techniques of Gas Chromatography Infrared (GC/IR) Analysis', American Society for Testing and Material, Philadelphia, PA, USA, February 1995.
12. M. Rautio (ed.), Recommended Operating Procedures for Sampling and Analysis in the Verification of the Chemical Disarmament, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1994.
13. Automated Mass Spectral Deconvolution and Identification System (AMDIS) (version 1, 0, 49, 0), National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA, 1998.
14. J.A. de Haseth, T.L. Isenhour, 'Reconstruction of Gas Chromatograms from Interferometric Gas Chromatography/Infrared Spectrometry Data', *Anal. Chem.*, **49**(13), 1977–1981 (1977).
15. L.J. Bellamy, *The Infrared Spectra of Complex Molecules*, Chapman and Hall, London, UK, Vol. 1, 1975.
16. M. Rautio (ed.), International Interlaboratory Comparison (Round-Robin) Test for the Verification of the Chemical Disarmament, F.3. Testing of Procedures on Simulated Military Facility Samples, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1992.
17. Chemical and Instrumental Verification of the Organophosphorus Warfare Agents, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1977.
18. Systematic Identification of Chemical Warfare Agents. B.3. Identification of Non-phosphorus Warfare Agents, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1982.
19. International Interlaboratory Comparison (Round-Robin) Test for the Verification of the Chemical Disarmament, F.1. Testing of Existing Procedures, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1990.
20. International Interlaboratory Comparison (Round-Robin) Test for the Verification of the Chemical Disarmament, F.2. Testing of Procedures on Simulated Industry Samples, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1991.
21. M. Rautio (ed.), International Interlaboratory Comparison (Round-Robin) Test for the Verification of the Chemical Disarmament, F.4. Validating of Procedures for Water and Soil Samples, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1993.
22. M. Rautio (ed.), Interlaboratory Comparison Test Coordinated by the Provisional Technical Secretariat for the Preparatory Commission for the Organization for the Prohibition of Chemical Weapons, H.1, First Interlaboratory Comparison Test, The Ministry for Foreign Affairs of Finland, Helsinki, Finland, 1994.
23. R.R. Shagidullin, A.V. Chernova, V.S. Vinogradova, F.S. Mukhametov, Atlas of IR Spectra of Organophosphorus

Compounds (Interpreted Spectrograms), Kluwer, Dordrecht, The Netherlands, 1990.
24. L. Fagerlind, I. Lindgren, J. Santesson, K.-E. Stensjö, G. Wallenberg, B. Östman Nervgaser och Besläktade Fosforföreningar. Syntes, Kromatografi, IR-, Mass- och NMR-spektrometrisk Analys, FOA 1 Rapport A 1541-C2(C4), Stockholm, Sweden, 1971.
25. M.T. Söderström, H. Björk, V.M.A. Häkkinen, O. Kostiainen, M.-L. Kuitunen, M. Rautio, 'Identification of Compounds Relevant to the Chemical Weapons Convention Using Selective GC Detectors, CC/MS and GC/FTIR in an International Trial Proficiency Test', *J. Chromatogr., A.*, **742**, 191–203 (1996).
26. W.R. Creasy, J.R. Stuff, B. Williams, K. Morrissey, J. Mays, R. Duevel, H.D. Durst, 'Identification of Chemical Weapons-related Compounds in Contamination Solutions and Other Matrices by Multiple Chromatographic Techniques', *J. Chromatogr., A*, **774**, 253–263 (1997).
27. X. Danian, C. Zhisheng, L. Jingquan, Y. Jiufeng, Z. Yuanzheng, P. Lixin, 'Structural Identification of *O*-alkyl *S*-2-dialkylaminoethyl Alkylphosphonothiolates by GC/MS, NMR and GC/FTIR', Proceedings from the 6th CBW Protection Symposium, Stockholm, Sweden, May 10–15 1998, FOA-R-98-00749-862-SE, 149–153, May 1998.
28. M.T. Söderström, 'Identification of VX-type Nerve Agents Using Cryodeposition GC-FTIR', in *Fourier Transform Spectroscopy*, 11th International Conference, ed. J.A de Haseth, American Institute of Physics, New York, USA, 457–460, 1998.
29. D. Durst, J.R. Mays, J.L. Ruth, B.R. Williams, R.V. Duevel, 'Micro-scale Synthesis and In-situ Spectroscopic Characterization of Some Chemical Weapons Related Organophosphonate Compounds', *Anal. Lett.*, **31**(8), 1429–1444 (1998).
30. M. Sokołowski, A. Szymańska, 'GC/MS Investigation of Ethyl *N*,*N*-dimethylphosphoramidocyanidate (Tabun) Age Decomposition Products', Proceedings 1995 ERDEC Science Conference Chemical Biological Defense Research, ed. D.A. Berg, National Technical Information Service, Springfield, VA, USA, 149–155, 1996.
31. K.E. Goeringer, M.W. Elizy, 'Analysis of Alkylphosphonic Acids in Environmental Waters by Gas Chromatography/Ion Trap Mass Spectrometry and Gas Chromatography/Matrix Isolated Fourier Transform Infrared Spectroscopy', Proceedings. 1994 ERDEC Science Conference Chemical Biological Defense Research (1996), ed. D.A. Berg, National Technical Information Service, Springfield, VA, USA, 719–725.
32. P.A. D'Agostino, L.R. Provost, A.S. Hansen, G.A. Luoma, 'Identification of Mustard Related Compounds in Aqueous Samples by Gas Chromatography/Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **18**, 484–491 (1989).
33. S.D. Christesen, 'Vibrational Spectra and Assignments of Diethyl Sulfide, 2-Chlorodiethyl Sulfide and 2,2′-dichlorodiethyl Sulfide', *J. Raman Spectrosc.*, **22**, 459–465 (1991).
34. J.R. Smith, T.P. Logan, L.L. Szafraniec, E.M. Jakubowski, 'Spectroscopic Characterization of the Geminal Isomer of Lewisite', *Anal. Lett.*, **28**(8), 1541–1554 (1995).
35. S.R. Lowry, D.A. Huppler, 'Infrared Spectral Search Systems for Gas Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chem.*, **53**(6), 889–893 (1981).
36. L.C. Thomas, R.A. Chittenden, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–I. Phosphorus–Oxygen Bonds', *Spectrochim. Acta*, **20**, 467–487 (1964).
37. L.C. Thomas, R.A. Chittenden, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–II. P–O–(X) bonds', *Spectrochim. Acta*, **20**, 489–502 (1964).
38. R.A. Chittenden, L.C. Thomas, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–III. Phosphorus–Sulfur and Phosphorus–Selenium Bonds', *Spectrochim. Acta*, **20**, 1679–1696 (1964).
39. R.A. Chittenden, L.C. Thomas, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–IV. P–X Bonds', *Spectrochim. Acta*, **21**, 861–876 (1965).
40. L.C. Thomas, R.A. Chittenden, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–V. Phosphorus–Carbon Bonds', *Spectrochim. Acta*, **21**, 1905–1914 (1965).
41. R.A. Chittenden, L.C. Thomas, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–VI. Bonds Between Phosphorus and Nitrogen', *Spectrochim. Acta*, **22**, 1449–1463 (1966).
42. L.C. Thomas, R.A. Chittenden, 'Characteristic Infrared Absorption Frequencies of Organophosphorus Compounds–VII. Phosphorus Ions', *Spectrochim. Acta*, Part A, **26**, 781–800 (1970).
43. M.T. Söderström, R.A. Ketola, 'Identification of Nerve Agents and their Homologues and Dialkyl Methylphosphonates by Gas Chromatography/Fourier Transform Infrared Spectrometry (GC/FTIR), Part I: Spectral Identification', *Fresenius' J. Anal. Chem.*, **350**, 162–167 (1994).
44. M. Mesilaakso, E.-L. Tolppa, 'Detection of Trace Amounts of Chemical Warfare Agents and Related Compounds in Rubber, Paint, and Soil Samples by ^{1}H and $^{31}P\{^{1}H\}$ NMR Spectroscopy', *Anal. Chem.*, **68**, 2313–2318 (1996).
45. Evaluation of Results, First Official OPCW/PTS Interlaboratory Comparison Test: Proficiency Test, Provisional Technical Secretariat for the Preparatory Commission for the Organization for the Prohibition of Chemical Weapons, 14 January 1997.
46. OPCW S/10/97, Evaluation of Results, Second Official Inter-laboratory Proficiency Test, Version 2, 26 August 1997.
47. OPCW S/23/97, Evaluation of Results, Third Official Proficiency Test, Version 2, 21 November 1997.

48. Evaluation of Results, Fourth Official Proficiency Test, Version 2, 9 December 1998.
49. K.E. Parchment, N. Govan, D.W. Anderson, 'Monitoring the Decontamination Reactions of Chemical Warfare Agents by FTIR/ATR Spectroscopy', Proceedings 1995 ERDEC Science Conference Chemical Biological Defense Research, ed. D.A. Berg, National Technical Information Service, Springfield, VA, USA, 863–869, 1996.
50. S. Banerjee, R.P. Semwal, S. Agarwal, 'Measurement of Diffusivities of Sulfur Mustard (SM) and its Analog Oxygen Mustard (OM) in Cured Butyl, Nitrile, and Natural Rubber Sheets by Weight Gain and FTIR/ATR Methods', *J. Appl. Polym. Sci.*, **57**, 1483–1490 (1995).
51. R.P. Semwal, S. Banerjee, L.R. Chauhan, A. Bhattacharya, N.B.S.N. Rao, 'Study of Diffusion and Sorption of bis(2-chloroethyl)sulfide (SM) and bis-(2-chloroethyl-ether (OM) through Polypropylene (PP) and Biaxial-oriented Polypropylene (BOPP) Films by the FTIR/ATR Spectroscopic Method', *J. Appl. Polym. Sci.*, **60**, 29–35 (1996).
52. J.F. Weimaster, W.T. Beaudry, P.C. Bossle, M.W. Ellzy, L.G. Janes, D.W. Johnson, J.M. Lochner, S.G. Pleva, J.H. Reeder, D.K. Rohrbaugh, T.E. Rosso, L.J. Szafraniec, L.L. Szafraniec, T.G. Albro, W.R. Creasy, J.R. Stuff, P.B. Smith, I.R. Stewart, 'Chemical Analysis of Environmental Samples Collected in Iraq: Analysis for the Presence of Chemical Warfare Agents', *J. Chem. Tech. Biotechnol.*, **64**, 115–128 (1995).
53. M.T. Söderström, R.A. Ketola, O. Kostiainen, 'Identification of Nerve Agents and their Homologues and Dialkyl Methylphosphonates by Gas Chromatography/Fourier Transform Infrared Spectrometry (GC/FTIR), Part II: Spectral Search with Help of Retention Indexes', *Fresenius' J. Anal. Chem.*, **352**, 550–556 (1995).
54. ASTM Standard E 1421-94, 'Standard Practice for Describing and Measurement Performance of Fourier Transform Infrared (FT/IR) Spectrometers: Level Zero and Level One Tests', American Society for Testing and Materials, Philadelphia, PA, USA, published February, 1995.

Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention

Olli Kostiainen
Finnish Institute for Verification of the Chemicals Weapons Treaty, Helsinki, Finland

1 Introduction 1
2 Instrumentation 1
2.1 Sample Introduction 1
2.2 Run Conditions 2
2.3 Columns 2
2.4 Detectors 3
3 Identification Methods 5
3.1 Absolute Retention Times 5
3.2 Relative Retention Times 5
3.3 Retention Indices 5
4 Quality Control 13
Abbreviations and Acronyms 13
Related Articles 13
References 14

Various gas chromatographic methods, which are useful for the screening of Chemical Weapons Convention (CWC)-related chemicals, are described in this section. Proper choice of columns, detectors, and injection techniques are discussed, because they are crucial to the optimization of gas chromatography (GC) as a routine tool. Also various identification methods are described. Retention index monitoring (RIM) is described in more detail than other identification methods because it is the most reliable GC technique to be used in the identification using a retention index (RI) library without the need for an authentic chemical.

1 INTRODUCTION

GC is the most widely used single analytical technique for the analysis of volatile organic compounds. The advantages of GC are good resolution, inertness, high sensitivity and possibilities for selective detection. Although the use of GC with conventional GC detectors has decreased due to the decreased prices of benchtop GC/mass spectrometry (MS) instruments, GC together with selective detectors still has an important role in the analysis of chemicals related to the CWC. Most of these chemicals have sufficiently high vapor pressure and thermal stability for GC analysis or they can be easily derivatized to a volatile form using methylation,[1] trimethylsilylation,[2] *tert*-butyldimethylsilylation[3] or pentafluorobenzylation.[4] Witkiewicz et al.[5] have published a comprehensive review concerning various chromatographic applications for the analysis of CWC-related chemicals.

GC has been the main separation technique in the international interlaboratory comparison and proficiency tests.[6–13] All participating laboratories used GC, either to determine the retention behavior of analyzed compounds or as part of the hyphenated techniques. It has been agreed based on these tests that the unambiguous identification of CWC-related chemicals has to be based on at least two different analysis techniques, preferably by two different spectrometric analysis techniques, when available, giving consistent results.[14] The minimum acceptable data are the correct electron impact mass spectrometry (EIMS) spectrum together with the correct retention data. GC/EIMS spectra and retention parameters obtained with the same instrument are not regarded as two different techniques.[14]

In verifying the absence or presence of CWC-related chemicals, the analysis of a specified sample includes analyzing samples by different chromatographic and spectrometric techniques. In general, the chromatographic techniques together with conventional detectors are useful for screening and tentative identification and analyses are in most cases rather qualitative than quantitative. GC methods used for screening purposes should be sensitive and selective and should detect as many CWC-related chemicals as possible in one GC analysis. Sensitivity can be obtained by specific detectors and selectivity by including a database in the GC computer containing identification data of relevant chemicals. GC monitoring before gas chromatography/mass spectrometry (GC/MS) identification, especially when done with specific detectors, is important in producing useful information from complex mixtures. When a lot of samples must be analyzed, GC can be used either to pick out the most valuable samples for more thorough analysis or to optimize the concentration of the samples for GC/MS identification.

2 INSTRUMENTATION

2.1 Sample Introduction

Injection techniques play an important role in chromatographic separation. Split, splitless, and on-column

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

injection methods have been studied[15] in order to optimize peak shapes, detection limits, and detector response reproducibility for CWC-related chemicals. Splitless injection at 250 °C with a splitless time of 45–60 s and using the hot-needle technique with a solvent flush was the most suitable injection method giving satisfactory results even for samples with difficult matrices. However, on-column injection has some advantages over other techniques, especially in the analysis of thermolabile compounds, high boiling compounds or compounds with adsorptive properties. The molecules will enter the column relatively slowly at a temperature that is lower than their boiling points. Detection of VX (CAS 50782-69-9) and its homologues and most of the methylated degradation products of nerve agents give better results by on-column injection: the detection limits for VX were about 50 pg with on-column injection and 2 ng with splitless injection when the same OV-1 column was used.[16] However, for environmental samples such as soil and plant extracts on-column injection may lead to rapid contamination of the analytical column. A changeable precolumn coupled with a press-fit connector can be used to protect the analytical column from contamination. In addition, an autosampler cannot always be used with the on-column technique. The split technique provides good separation efficiency for most volatile compounds such as hydrogen cyanide (CAS 74-90-8), phosgene (CAS 75-44-5), and cyanogen chloride (CAS 506-77-4) when the concentrations of these are high enough. However, discrimination against high boiling components is evident.

The large-volume injection technique together with on-column injection[17,18] and thermal vaporization injection[19,20] has been used to improve the sensitivity of GC analysis. Several hundred microliters of the sample can be injected by this technique. Degenhardt-Langelaan et al.[21] determined selected nerve agents at parts per trillion levels in organic extracts of water samples using large-volume injection. Although this method is very attractive, the optimization of all instrument parameters is quite time-consuming.

Cold trap injection is normally used in the analysis of air samples. During the thermal desorption from Tenax® resin[22] the sample chemicals are collected on a cold trap, after which the temperature of the cold trap is quickly raised to about 250–300 °C. The chemicals move as a narrow band to the capillary column for a normal gas chromatographic run without an interfering solvent peak. The usefulness of the thermal desorption method for the analysis of CWC-related chemicals has been tested.[23] The recoveries at low nanogram levels were between 81% and 103% for sarin (CAS 107-44-8), soman (CAS 96-64-0), tabun (CAS 77-81-6) and mustard gas (CAS 505-60-2) when the volume of sampled urban air was between 5 L and 30 L. VX (CAS 50782-69-9) could not be detected due to its high adsorptivity nature. The breakthrough volume of sarin, one of the most volatile nerve agents, was 60 L at the sampling temperature of 23 °C on 100 mg Tenax® adsorbent. Tenax® tubes loaded with some organophosphates and mustard gas from humidified atmospheres have been stored for up to 2 months without a decrease in recovery.[24] Hancock et al.[25] developed an automated air sampling analysis system for the monitoring of CWC-related chemicals in air. Air samples were collected on minitubes packed with Tenax® TA and thermally desorbed into GC without cryofocusing. A RIM method together with thermal desorption has been reported.[23,26,27] RIs were determined by loading a Tenax® tube with the monitored chemicals and subsequently with a series of index standards.

2.2 Run Conditions

According to the recommended operating procedure for the analysis of CWC-related chemicals by GC[22] when using SE-54 or OV-1701 columns (25 m × 0.32 mm inner diameter (ID), 0.25 μm film thickness) the following conditions are recommended: injector temperature, 250 °C; detector temperature, 280 °C; temperature program, initial temperature 40 °C for 1 min, temperature programming rate 10 °C min^{-1} to 280 °C for 10 min. However, the GC conditions depend on the injection techniques used, columns, sample types and the analytical demands: when GC is used for rapid sample screening or when maximum resolution is needed, the GC conditions are different.

2.3 Columns

The most critical parameters of a capillary column are resolution, support inertness, retention reproducibility, thermal stability, and bleed. To provide fast, reliable and accurate analysis, it is important that the stationary phase, ID of column, film thickness and length of column be chosen with a view to the particular application. CWC-related chemicals differ greatly from each other in their chemical and physical properties and thus the selection of the stationary phase is always a compromise between resolution and analysis time. The most suitable stationary phases for the separation of chemicals related to the CWC are listed in Table 1, along with their structures, and polarities.[22]

The nonpolar phases tend to have better characteristics in terms of resistance to oxygen, higher efficiencies, and greater maximum temperatures. Laboratories should have in their use at least two columns of different polarity. The SE-54 phase can be regarded as a better choice than OV-1 for the nonpolar phase because it provides better peak shapes for fairly polar compounds such as thiodiglycol and dialkyl alkylphosphonates. Preliminary

Table 1 Stationary phases suitable for the analysis of CWC-related chemicals

Polarity	Name	Structure
Low	OV-1, SE-30, SE-54	100% Methylpolysiloxane 5% Phenyl, 1% vinyl methylpolysiloxane
Intermediate	OV-1701	7% Phenyl, 7% cyanopropyl methylpolysiloxane
High	CW-20M	Poly(ethyleneglycol)

identification cannot always be achieved using retention data obtained on one column. When two columns of different polarities are needed, OV-1701 and SE-54 are a good pair, being sufficiently different in selectivity. An OV-1701 column is recommended for use with a sulfur-selective detector, because mustards and their corresponding primary degradation products elute close to each other on an SE-54 column, and identification may be difficult.[22] A polar CW-20M column can be used in only limited applications for the analysis of most polar compounds. Some of the monitored CWC-related chemicals do not elute through this column because of the strong retention and quite low maximum temperature of the phase material. Huber et al.[28] found that theoretically a maximum of about 1 million chemicals can be separated by using a combination of three columns with the stationary phases methyl silicone, 50% phenylmethyl silicone, and methylcyanopropyl silicone.

The ID has a direct influence on retention, efficiency, and capacity of the column. The on-column injection technique requires an ID of at least 0.30 mm. A narrow-bore column with ID of 0.20 mm provides good resolving power with a minimum bleed. It is a good choice for MS analysis as it facilitates a proper adjustment of the carrier gas flow. Narrow-bore columns of limited capacity, however, may be a disadvantage for identification. Columns of ID between 0.25 and 0.33 mm can be considered equal for the separation of treaty-related compounds. Columns of ID 0.53 mm are useful if the sample contains a limited number of chemicals in widely different concentrations.

Thin-film (below 0.20 μm) columns should be avoided because these tend to give poor peak shapes for strongly adsorptive compounds such as thiodiglycol (CAS 111-48-8), dialkyl alkylphosphonates, and aminoalcohols, especially at a low concentration level. Thick-film columns are ideal for the separation of very volatile compounds such as hydrogen cyanide (CAS 74-90-8), phosgene (CAS 75-44-5), and cyanogen chloride (CAS 506-77-4). They are unsuitable, however, for higher molecular-weight compounds such as BZ (CAS 6581-06-2). Standard-film (0.25 – 0.33 μm) thickness offers the best compromise between resolution and sample capacity.

The selection of column length depends on the required resolution and analysis time. Short columns (10 – 25 m) are useful for samples containing a relatively small number of compounds and for screening purposes. To keep the analysis time short and to minimize the adsorption, a 50-m column is useful only for very complex separations requiring maximum resolution. Intermediate column lengths of 25–30 m, providing sufficient separation power simultaneously with reasonable analysis time, are in most cases used for the separation of CWC-related chemicals.

2.4 Detectors

The critical properties of detectors are sensitivity, selectivity, linearity of response, reproducibility and reliability of operation. Spectrometric detectors, MS and Fourier transform infrared (FTIR), are described in more detail in other sections. The following detectors are useful for the detection of CWC-related chemicals.

The flame ionization detector (FID) is currently the most popular universal detector, having both a wide linear range and great reliability. This detector has been widely used for the detection of CWC-related chemicals.[24,29,30] The detection limits for CWC-related chemicals are about 0.1 ng, but the chromatography has a significant influence on the sensitivity: when baseline resolution can be obtained, high sensitivity is the result. In most cases chromatographic interferences are the main problem with the FID. The sensitivity is poor for some compounds such as phosgene (CAS 75-44-5) and hydrogen cyanide (CAS 74-90-8). In addition it is the only detector besides the atomic emission detector (AED) to detect Schedule 2 chemicals pinacolyl alcohol (CAS 464-07-3) and benzilic acid (CAS 76-93-7).

The photoionization detector (PID) is both a universal and a selective detector, with the observed response being dependent on the ionization efficiency of the detected compounds. The PID has a wide linear range of seven orders of magnitude. Any chemical with an ionization potential below 12 eV will give a response. The PID/FID response ratio shows PID to be the more sensitive detector for sulfur-containing CWC-related chemicals such as mustard gas (CAS 505-60-2) and its homologues, for tabun (CAS 77-81-6), and for compounds containing double bonds such as Lewisites.[26] Because no detector gases are needed, the PID may be useful for some on-site applications. However, it is not selective enough for some environmental samples (e.g. air) because of the high response to aromatic compounds.

Despite the great separation power of capillary columns, the analysis at trace level in some matrices may be difficult if only universal detectors are used.

Many selective detectors have been used for monitoring CWC-related chemicals because these chemicals are rich in heteroatoms. In most cases they contain from two to four such as P, S, O, N, Cl, As, and F. Chromatograms are simplified with these detectors and the sensitivity of analysis is increased.

Phosphorus-selective detectors are of greatest importance for verification analysis. The nitrogen–phosphorus detector (NPD) is very sensitive and selective towards compounds containing phosphorus and/or nitrogen, reaching detection limits of about 0.5–5 pg and 10–100 pg respectively. Arsenic-containing compounds such as Lewisites are also detectable with the NPD, but the sensitivity is worse than with the FID. Sensitivity of the NPD depends on the temperature of the NPD bead and can change significantly over time, which must be taken into account in the quantitation. In addition some models of the NPD are not compatible with chlorinated solvents and silylation reagents.[31] The NPD is normally used in the analysis of nerve agents at trace level.[32–34]

The flame photometric detector (FPD) is the most widely used detector for sulfur-containing compounds. Besides sulfur it is selective for phosphorus, which makes it a very suitable detector for both nerve agents and sulfur-containing mustards.[35–37] For sulfur, however, the detector is not linear and suffers from quenching by co-eluting hydrocarbons. Detection limits are about 5–50 pg of phosphorus s^{-1} and about 50–500 pg of sulfur s^{-1}. Two-channel detectors operating simultaneously in S- and P-mode are also commercially available. The highest operating temperature for some models is 250 °C. This prevents the detection of higher boiling compounds and leads to contamination of the detector especially when high concentrations of trimethylsilyl (TMS) derivatizing reagents are used.[31]

An alternative to FPD in sulfur mode is the sulfur chemiluminescence detector (SCD).[38] This detector works by forming sulfur monoxide in a reducing flame. Sulfur monoxide is detected by its chemiluminescent reaction with ozone. The SCD is at least one order of magnitude more sensitive than most FPDs. It provides a linear response with high selectivity and does not suffer considerably from quenching.

The electron capture detector (ECD) responds to compounds capable of reacting with thermal electrons to form negative ions from halogen-containing chemicals (the order of increasing response is F < Cl < Br < I) and also nitro and α-dicarbonyl groups.[39] In principle the ECD is useful for the analysis of some CWC-related chemicals. Unfortunately, environmental samples usually contain a lot of other ECD active material, which seriously interferes with the identification. The very poor sensitivity of the ECD for fluoridate nerve agents suggests the use of NPD or FPD for these compounds.[40] Also the relatively narrow linear range of the ECD tends to reduce the usefulness of this detector. Several CWC-related chemicals containing chlorine atoms, e.g. vesicants, are detectable with the ECD. Their degradation products are not detected by ECD, so the use of sulfur-selective detectors such as FPD or SCD is recommended for the detection of these compounds due to their better sensitivity and selectivity.

One of the most attractive detectors for screening of CWC-related chemicals is the AED, which is capable of detecting selectively any element below nanogram level. The sensitivity and other properties of the AED are compared to other detectors for a variety of analytes in a number of papers.[31,41,42] With this detector it is also possible to determine an approximate empirical formula for unknown compounds.[43,44] The AED relies on a microwave-induced helium plasma contained in a novel re-entrant cavity with a photodiode array-based optical emission spectrometer.[45] Very-high-purity helium is required as carrier and plasma gas. Leaks are insidious sources of error, especially when oxygen or nitrogen is being measured. The information obtained from the AED is very informative for the unknown chemicals, when the information is collected from various techniques for structure elucidation. Nowadays the AED has been used widely for the detection of CWC-related chemicals.[46–50]

The usefulness of selective detectors is well demonstrated with the spiked sample analyzed with the FID and selective detectors. Diesel fuel solution (0.03%) containing three CWC-related chemicals at low parts per million level is first analyzed with a FID (Figure 1). It is clear that

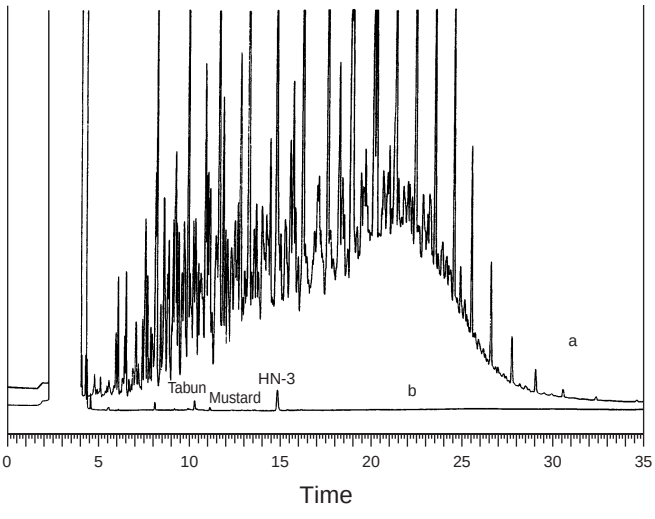

Figure 1 Chromatogram recorded with FID (a) from an ethyl acetate solution containing diesel fuel (0.03%), tabun (5 ng), mustard (3 ng), and HN-3 (20 ng); (b) is from a similar solution but with no diesel fuel.

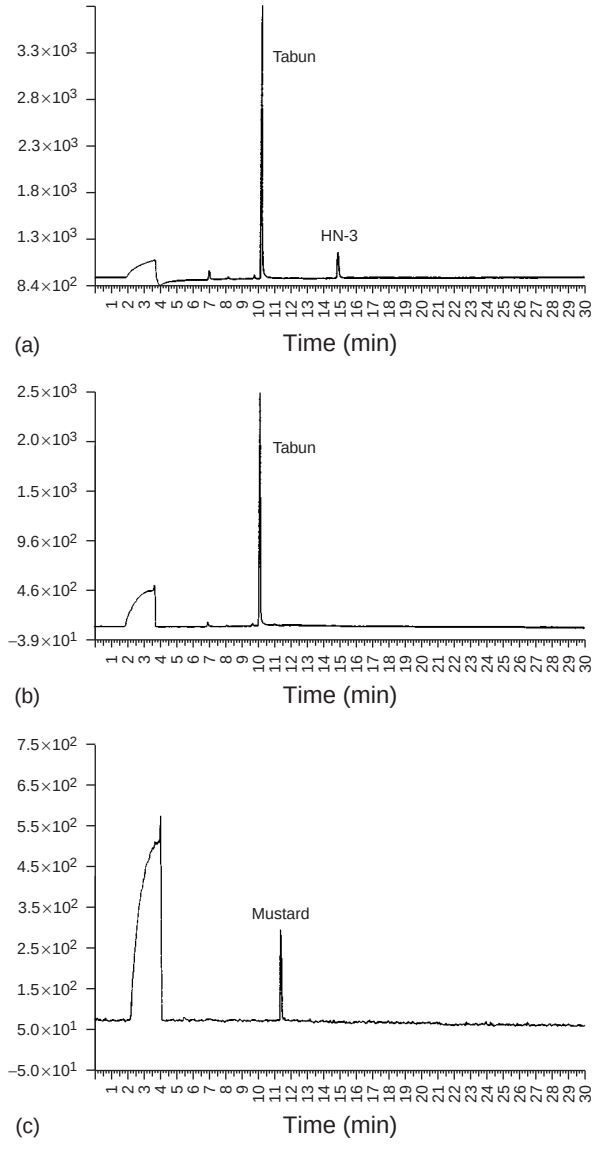

Figure 2 (a) NPD, (b) FPD-P and (c) FPD-S chromatograms from the same solution as Figure 1(a).

the identification is not possible in this case. The chromatograms are clearly simplified, when the same sample is analyzed with the two-channel instrument (Figure 2). The SE-54 column is split between the NPD and the FPD, which is simultaneously operating in S- and P-mode.

3 IDENTIFICATION METHODS

3.1 Absolute Retention Times

Absolute retention times are, in principle, useful for the identification, but in practice their reproducibility is not very good. Modern gas chromatographs have high precision and retention times within 0.1% or better can be obtained in subsequent runs during the same day. When the analysis is repeated by another instrument the deviation may be several per cent. The reproducibility of absolute retention times is strongly dependent on the proper adjustment of all chromatographic parameters. In addition, the column properties are not exactly the same even when similar columns from the same manufacturer are used. On the other hand, absolute retention times are the most useful identification parameter if only a few compounds are to be monitored and the background is low. This method requires frequent calibration because even small changes in chromatographic conditions will influence the absolute retention times.

3.2 Relative Retention Times

For better reproducibility of retention parameters, relative retention times referenced to a standard compound have been used in several GC applications. The added reference chemical has to be chosen so that it elutes in the middle of the chromatogram. There is always some degree of error in measuring, so that the relative retention times of compounds having short or long retention times might not be very accurate. To obtain better reproducibility it is recommended that several different reference compounds are used for the calculation of relative retention times. However, in this case the use of RIs becomes more attractive. It is a much more reliable approach than the use of absolute or relative retention times, because in RI measurements the retention is measured relative to a homologue series.

3.3 Retention Indices

The reproducibility and reliability of RIs makes it possible to create RI libraries. The reliability and simplicity of index monitoring is increased significantly by using a computer program that searches for the RI pattern, calculates the RIs for all peaks in the chromatogram and then compares the indices with the library data. In addition to the identification of target chemicals, RIs can also be used to locate the interesting peaks between different kinds of GC-based analytical techniques.[51] In this way it is possible to ensure that all GC based techniques used for identification focus on the same peaks even in samples with a complex mixture of chemicals.

The RI system proposed by Kovats[52] was determined at constant temperature using a homologue series of *n*-alkanes (C-series) for calibration. With mixtures that embrace a wide range of boiling points, several isothermal runs are needed. In these cases, the use of linear RIs is more useful, because this method allows determination of indices for rapidly and slowly eluting compounds in a single analysis. The column temperature is programmed

linearly and the RI of an unknown peak in the chromatogram is calculated from the retention times of the two alkanes eluted on both sides of the compound.[53] The distance between each pair of standard peaks is taken as equal, and the linear RIs are calculated by a linear polygon method as shown in Equation (1).

$$RI_C = 100C_n + 100(C_{n+i} - C_n)\frac{t_{R(x)} - t_{R(n)}}{t_{R(n+i)} - t_{R(n)}} \quad (1)$$

where C_n and C_{n+i} are the carbon number of the RI standard eluted before and after the unknown compound, $t_{R(x)}$ is the retention time of the unknown compound, and $t_{R(n)}$ and $t_{R(n+i)}$ are the retention times of C_n and C_{n+i}, respectively.

The C-series are the most commonly used RI standards. The RIs of nonpolar substances on nonpolar stationary phases show an almost linear dependence on column temperature. Investigations with other phases and substances suggest, however, that in general the dependence on temperature is not linear but hyperbolic.[54] The reproducibility of RIs is better if the index standards are structurally similar to the compounds to be analyzed. In the case of polycyclic aromatic hydrocarbons (PAHs) used as standards for other PAHs, the effects of film thickness and temperature programming rate were found to be less than ±0.16 index units. The effect of film thickness using *n*-alkanes as standard compounds was under the same conditions between 2.5 and 19.6 index units.[55] Nonpolar alkanes are not the most useful as reference standards for most polar CWC-related chemicals. Alkanes, moreover, are not detectable with selective detectors that have to be used for trace analysis. For the specific purpose of identifying CWC-related chemicals, several series of phosphorus-containing retention standards have been synthesized.[56,57] These standards contain *n*-alkane moieties but also phosphorus, sulfur and in most cases fluorine, for easy detection with selective detectors. The most stable and suitable for the analysis of CWC-related chemicals is the M-series, alkyl bis(trifluoromethyl)phosphine sulfides, $(CF_3)_2P(S)(CH_2)_nCH_3$ (where $n = 2, \ldots, 19$).[58]

3.3.1 Reproducibility of Retention Indices

The reproducibility of the GC RIs is essential to reliable screening and identification procedure for CWC-related chemicals. The effects of the following parameters on the reproducibility have been evaluated:[59,60] (1) carrier gas flow rate; (2) starting point of the temperature program; (3) temperature programming rate; (4) injection volume; (5) injection mode; (6) different operator; (7) one- and two-index standard series in the same run; (8) reduced number of index standards; (9) use of external standards; (10) sample solvent; (11) column from the same and different manufacturers; (12) column length; (13) ID of the column; (14) relative peak area compared with the nearest index standard; (15) repeated use of the same column; (16) concentration of the sample; (17) background; (18) multistep temperature program; (19) model of the instrument; and (20) different laboratories. The C- and M-series were used as index standards and the test compounds studied were sarin (CAS 107-44-8), soman (CAS 96-64-0), tabun (CAS 77-81-6), VX (CAS 50782-69-9), mustard gas (CAS 505-60-2), CS (CAS 2698-41-1), CR (CAS 257-07-8), and CN (CAS 532-27-4). A fused-silica capillary column with SE-54 stationary phase was used.

The results of the evaluation showed the RIs to vary only slightly with the operator, different instrument of

Table 2 RI values for various CWC-related chemicals on various columns determined against C- and M-series

Compounds	CAS number	C-series					M-series		
		I SE-54	II OV-1701	III OV-1	IV SE-54	V OV-1701	VI SE-54	VII SE-54	VIII OV-1701
Methyl methylphosphonofluoridate	353-88-8	586.6[a]	841.8				238.8[a]		279.2[a]
Pinacolyl alcohol	464-07-3	634.2[a]	828.6						
Trimethyl phosphite	121-45-9	688.8[a]	804.7				253.8[a]		242.4[a]
Ethyl methylphosphonofluoridate	673-97-2	754.9[a]	913.2				285.6[a]		353.2
Methyl ethylphosphonofluoridate	665-03-2	802.7	935.5				306.3		377.5
O-Isopropyl methylphosphonofluoridate (sarin)	107-44-8	820.2	953.4	791.8	823.8	965.9	329.3		400.0
Ethyl ethylphosphonofluoridate	650-20-4	865.7	1012.6				388.9		454.5
Propyl methylphosphonofluoridate	763-14-4	868.9	1022.1				392.6		464.8
1-Oxa-4-thiacyclohexane	15980-15-1	877.2	969.7	850.7	880.0	964.1	400.0		409.5
Dimethyl methylphosphonate	756-79-6	880.9	1048.4				403.9	398.8	492.8
2-Chloroethyl vinyl sulfide				873.7	903.0	992.1			
Isopropyl ethylphosphonofluoridate	1189-87-3	906.9	1046.9				432.2		491.6
sec-Butyl methylphosphonofluoridate[b]	352-52-3	915.4	1058.9				438.2		504.6
		917.9	1061.5				440.4		507.4
(2-Vinylthio)ethanol	3090-56-0			889.6	923.2	1089.2			

Table 2 (*continued*)

Compounds	CAS number	C-series					M-series		
		I SE-54	II OV-1701	III OV-1	IV SE-54	V OV-1701	VI SE-54	VII SE-54	VIII OV-1701
2-Butyl methylphosphonofluoridate								443.9	
Triethyl phosphite	122-52-1	921.9	985.3				446.9		427.2
Isobutyl methylphosphonofluoridate	2053-81-8							460.5	
3-Hydroxy-1-methylpiperidine	3554-74-3	947.5	1059.8				471.1		505.4
Ethyl methyl methylphosphonate	18755-36-7	951.5	1112.1				474.9		561.6
Diethyl phosphonate	762-04-9	952.9	1115.0				477.6		565.5
Propyl ethylphosphonofluoridate	2992-95-2	967.1	1112.1				491.3		562.3
Butyl methylphosphonofluoridate	352-63-6	971.2	1121.7				495.0		572.7
Neopentyl methylphosphonofluoridate								499.1	
Bis(2-chloroethyl) ether				948.3	984.0	1107.5			
4-Hydroxy-1-methylpiperidine	106-52-5	987.5	1123.7				510.3		574.7
Isopropyl methyl methylphosphonate	690-64-2	988.6	1137.1				512.1		588.6
Diethyl methylphosphonate	867-17-4			974.9	1006.5	1186.9			
sec-Butyl ethylphosphonofluoridate	162085-83-8	1015.0	1150.1				540.5		603.4
Diethyl methylphosphonate	683-08-9	1015.1	1169.6				544.5	546.9	625.5
Isobutyl ethylphosphonofluoridate	2261-83-8	1027.7	1168.0				554.3		623.1
O-Pinacolyl methylphosphonofluoridate (soman)[b]	96-64-0	1043.7 1048.1	1183.5 1189.4	1008.1 1012.9	1045.2 1049.3	1187.5 1193.3	571.9 576.8		640.3 646.7
Diethyl phosphorocyanidate	2942-58-7	1046.4	1236.8				574.1		695.9
4-Methyl-2-pentyl methylphosphonofluoridate[b]							580.4 588.8		
Dimethyl ethylphosphonate	6163-76-3							582.3	
Butyl ethylphosphonofluoridate	18358-34-4	1067.2	1210.0				597.8		668.6
1,4-Dithiacyclohexane	505-29-3	1067.6	1169.2	1018.7	1060.2	1156.7	596.8		623.9
Methyl *N*,*N*-dimethylphosphoramidocyanidate	63815-56-5	1070.2	1281.7				600.0		740.7
Diisopropyl methylphosphonate	1445-75-6	1072.7	1207.2				607.4	601.0	666.7
Pentyl methylphosphonofluoridate	13454-59-6	1073.4	1223.3				604.4		682.1
2-Chlorovinyldichloroarsine (Lewisite 1)	541-25-3	1082.9	[c]				610.0		[c]
Bis(2-chloroethyl)methylamine (HN-2)	51-75-2	1087.2	1204.3				618.3		661.9
2-(Diisopropylamino)ethanethiol	5842-07-9			1098.2	1113.4	1175.5			
Dimethyl isopropylphosphonate								623.3	
Cyclopentyl methylphosphonofluoridate	7284-82-4							648.4	
Isopropyl trimethylsilyl methylphosphonate		1115.8	1244.6				649.1		701.6
O-Ethyl *S*-methyl methylphosphonothiolate	51865-09-9	1120.9	1289.8				655.1		749.1
2-Chlorobenzaldehyde	89-98-5	1131.1	1273.1				667.3		731.8
O-Ethyl *N*,*N*-dimethylphosphoramidocyanidate (tabun)	77-81-6	1133.4	1342.4	1077.9	1131.6	1339.8	668.1		802.8
Diethyl *N*,*N*-dimethylphosphoramidate	2404-03-7	1133.5	1277.5				668.6		737.9
Triethyl phosphate	78-40-0			1090.5	1137.2	1308.5			
O-Ethyl *O*-trimethylsilyl methylphosphonothiolate		1137.7	1239.6				672.6		697.9
Pinacolyl ethylphosphonofluoridate	97931-20-9	1141.4	1279.7				677.4		739.4
Diethyl dimethylphosphoramidate	2404-03-7			1096.6	1145.2	1295.6			
Methyl neopentyl methylphosphonate								680.4	
Bis(trimethylsilyl) methylphosphonate	18279-83-9	1154.1	1276.1				689.8		733.9
Bis(2-chloroethyl)ethylamine (HN-1)	538-07-8	1156.1	1273.6				692.2		732.5
Dimethyl propylphosphonate	18755-43-6							696.7	
3-Methyl-1-butyl methyl methylphosphonate								696.9	
Isopropyl *N*,*N*-dimethylphosphoramidocyanidate	63815-55-4	1161.0	1360.9				697.3		823.4
Pentyl ethylphosphonofluoridate	162085-84-9	1165.2	1313.5				702.7		773.9
Ethyl isopropyl *N*,*N*-dimethylphosphoramidate	99520-56-6			1121.4	1166.6	1308.5			

(*continued overleaf*)

Table 2 (*continued*)

Compounds	CAS number	C-series					M-series		
		I SE-54	II OV-1701	III OV-1	IV SE-54	V OV-1701	VI SE-54	VII SE-54	VIII OV-1701
2-(Diisopropylamino)ethyl trimethylsilyl ether		1170.6	1186.8				708.0		644.7
Hexyl methylphosphonofluoridate	113548-89-3	1171.5	1326.5				709.5		787.1
1,4,5-Oxadithiepane	3886-40-6	1172.0	1306.3	1117.8	1162.7	1289.9	709.6		766.1
Diethyl ethylphosphonate	78-38-6							709.7	
O,*S*-Diethyl methylphosphonothiolate	2511-10-6	1174.3	1334.5				716.0		796.0
Bis(2-chloroethyl)sulfide (mustard gas)	505-60-2	1177.9	1337.4	1123.8	1172.7	1325.8	715.5		797.3
2-Pentyl methyl methylphosphonate								716.5	
Hemisulfur mustard	693-30-1			1132.3	1177.5	1400.0			
Thiodiglycol	111-48-8	1184.4	1467.6	1130.9	1181.5	1458.2	723.4		934.0
(2-Chloroethylthio)ethyl ethyl ether				1169.2	1200.0				
Dipropyl methylphosphonate	6410-56-6							738.5	
Methyl pinacolyl methylphosphonate[b]	7040-59-7	1200.0	1350.4				740.9		811.8
		1208.3	1360.2				748.7		822.7
Phenyl methylphosphonofluoridate	133826-40-1	1207.3	1407.5				747.7		874.5
Diisopropyl ethylphosphonate	1067-69-2							748.3	
Diethyl isopropylphosphonate								748.5	
4-Methyl-1-butyl methyl methylphosphonate								754.6	
2-Chloroethyl 3-chloropropyl sulfide				1168.8	1212.6	1357.9			
Ethyl tetramethylphosphorodiamidate	2404-65-1			1158.7	1216.6	1392.8			
4-Ethyl-2,6,7-trioxa-1-phospha-bicyclo[2.2.2]octane	824-11-3	1221.1	1415.7				761.2		882.7
Propyl *N*,*N*-dimethylphosphoramidocyanidate	162085-86-1	1222.1	1432.7				761.4		899.4
Tetramethylphosphorodiamidic cyanide				1179.5	1245.7	1514.7			
(2-Chloroethylthio)ethyl vinyl ether				1185.3	1222.8	1354.1			
Diisopropyl isopropylphosphonate								787.1	
Diethyl propylphosphonate	18812-51-6							787.3	
Cyclohexyl methylphosphonothionofluoridate	4241-34-3	1249.6	1365.8				788.2		828.4
sec-Butyl *N*,*N*-dimethylphosphoramidocyanidate[b]		1256.1	1457.8				794.9		924.4
		1259.3	1461.3				798.3		927.9
Neopentyl trimethylsilyl methylphosphonate								795.9	
2-Methylcyclohexyl methylphosphonofluoridate[b]	85473-32-1	1260.2	1422.5	1210.2	1256.7	1421.1	798.0		889.7
		1261.7	1425.5	1211.7	1258.5	1423.3	800.0		892.6
Hexyl ethylphosphonofluoridate	35445-19-1	1265.3	1413.7				805.0		881.0
O-Trimethylsilyl-3-quinuclidinol		1268.0	1332.1				808.2		793.2
3-Methyl-1-butyl trimethylsilyl methylphosphonate								808.3	
Heptyl methylphosphonofluoridate	162085-82-7	1272.4	1427.8				812.8	817.8	895.2
Isobutyl *N*,*N*-dimethylphosphoramidocyanidate	162085-88-3	1272.9	1480.9				812.2		947.4
2-Pentyl trimethylsilyl methylphosphonate								822.6	
2-(Diisopropylamino)ethyl ethyl sulfide	110501-54-7			1263.9	1277.6	1333.9			
Bis(2-chloropropyl)sulfide	22535-54-2			1235.8	1275.6	1420.5			
Chloroacetophenone (CN)	532-27-4	1290.1	1484.5	1230.2	1283.0	1471.0	833.2		951.1
Diisopropyl propylphosphonate								821.3	
Lewisite 2	40334-69-8	1289.5	[c]				829.9		[c]
Diisobutyl methylphosphonate								842.6	
Phenyl ethylphosphonofluoridate		1303.6	1499.3				845.9		965.9
Cyclohexyl ethylphosphonofluoridate	7284-84-6	1307.1	1471.9				849.9		939.0
Methyl cyclopentyl methylphosphonate								825.0	
Methylbis[2-(trimethylsiloxy)ethyl]amine	76710-52-6	1314.2	1344.7				857.6		807.0
Butyl *N*,*N*-dimethylphosphoramidocyanidate	162085-87-2	1317.0	1530.1				859.7		996.5

Table 2 (*continued*)

Compounds	CAS number	C-series					M-series		
		I SE-54	II OV-1701	III OV-1	IV SE-54	V OV-1701	VI SE-54	VII SE-54	VIII OV-1701
2-Chloroethyl 4-chlorobutyl sulfide				1269.3	1312.3	1457.1			
2-Ethanol 3-propanol sulfide	5323-60-4	1319.6	1593.5				862.0		1063.0
Pinacolyl trimethylsilyl methylphosphonate		1319.7	1448.6				861.7		914.8
4-Methyl-2-pentyl trimethylsilyl methylphosphonate								864.3	
4-Isopropyl-2,6,7-trioxa-1-phosphabicyclo[2.2.2]octane	51486-55-6	1327.2	1530.5				871.2		997.2
Dipropyl ethylphosphonate	6163-76-4							900.1	
Dimethyl morpholinophosphoramidate	597-25-1			1296.2	1353.6	1562.1			
Ethylbis[2-(trimethylsiloxy)ethyl]amine		1364.0	1388.3				910.3		855.2
Heptyl ethylphosphonofluoridate	162085-85-0	1364.2	1518.2				910.8		985.2
Dipropyl isopropylphosphonate	192698-90-1							921.4	
1,2,5-Trithiepane	6576-93-8	1374.8	1537.0	1300.0	1361.7	1513.3	921.6		1000.0
Diethyl dimethylpyrophosphonate				1339.7	1390.5	1669.7			
Dibutyl methylphosphonate	2404-73-1							933.7	
Cyclopentyl trimethylsilyl methylphosphonate								938.1	
Bis(2-chloroethyl) disulfide	1002-41-1	1400.0	1578.2	1336.3	1391.2	1561.8	948.3		1049.1
Tris(2-chloroethyl)amine (HN-3)	555-77-1	1411.0	1612.0				959.3		1084.1
Pentyl *N*,*N*-dimethyl-phosphoramidocyanidate	148461-87-4	1411.8	1630.0				960.0		1102.1
(2-Chloroethylthio)ethyl vinyl sulfide				1366.1	1413.4	1576.9			
1,4,7-Oxadithionane	40474-73-5	1420.9	1578.0	1358.0	1409.3	1558.6	968.8		1048.6
(2-Hydroxyethylthio)ethyl vinyl sulfide				1370.9	1418.8	1650.3			
Bis(2-hydroxyethyl) disulfide	1892-29-1	1425.1	1755.9				973.5		1228.8
Bis(2-trimethylsiloxyethyl) sulfide	20486-03-7	1426.7	1490.6				974.6		958.2
Dipropyl propylphosphonate								978.5	
Mustard gas sulfone	471-03-4	1433.3	1782.7				980.4		1258.4
O-Ethyl *S*-2-(dimethylamino)ethyl methylphosphonothiolate	20820-80-8	1442.5	1620.9				989.6		1093.2
Mustard gas sulfoxide	5819-08-9	1455.3	1782.7	1367.0	1443.2	1768.0	1000.0		1400.0
Diisobutyl ethylphosphonate	7242-55-9							1000.1	
Diisobutyl isopropylphosphonate	83218-20-6							1008.3	
Tris(2-chlorovinyl)arsine (Lewisite 3)	40334-70-1	1464.6	1614.3				1011.5		1086.6
Nonyl methylphosphonofluoridate								1025.3	
2-Chloroethyl (2-chloroethoxy)ethyl sulfide				1418.3	1472.6	1658.1			
Diisoamyl methylphosphonate								1044.0	
Bis[2(vinylthio)ethyl]ether	114811-39-1			1425.6	1466.4	1600.0			
Hexyl *N*,*N*-dimethyl-phosphoramidocyanidate	162085-90-7	1512.9	1733.9				1062.9		1205.1
Dipinacolyl methylphosphonate	7040-58-6							1065.0	
O-Ethyl *S*-2-(dimethylamino)ethyl ethylphosphonothiolate	98543-25-0	1521.7	1691.6				1072.1		1162.8
Diisobutyl propylphosphonate								1075.0	
Dibutyl ethylphosphonate	2404-58-2							1083.3	
O-Isopropyl *S*-2-(dimethylamino)ethyl ethylphosphonothiolate	162085-92-9	1532.4	1689.2				1083.6		1160.6
Dibutyl isopropylphosphonate	919-21-1							1108.3	
O-Ethyl *S*-3-(dimethylamino)propyl methylphosphonothiolate		1557.5	1744.9				1110.2		1217.0
2-Chlorobenzalmalononitrile (CS)	2698-41-1	1564.0	1823.9	1489.6	1555.1	1805.6	1116.6		1299.3
Bis(2-chloroethylthio)methane	63869-13-6	1569.2	1789.6				1121.9		1264.6
Bis(2-hydroxyethylthio)methane	44860-68-6	1573.9	1921.7				1127.8		1395.8
Di-*n*-amyl methylphosphonate								1132.4	

(*continued overleaf*)

Table 2 (*continued*)

Compounds	CAS number	C-series					M-series		
		I SE-54	II OV-1701	III OV-1	IV SE-54	V OV-1701	VI SE-54	VII SE-54	VIII OV-1701
O-Ethyl *S*-2-(diethylamino)ethyl methylphosphonothiolate	21770-86-5	1594.5	1767.9				1148.9		1241.8
Heptyl *N*,*N*-dimethyl-phosphoramidocyanidate	162085-91-8	1610.9	1837.3				1165.4		1313.0
Dibutyl propylphosphonate								1175.0	
O-Methyl *S*-2-(diethylamino)ethyl ethylphosphonothiolate	170800-77-8	1621.7	1790.1				1175.5		1266.0
Bis(2-chloroethyl) trisulfide	505-60-2			1561.7	1626.3	1805.4			
4-nitro-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide								1184.8	
Bis(2-trimethylsiloxyethylthio) disulfide		1630.7	1714.8				1184.9		1187.2
O-Ethyl *S*-2-(ethylthio)ethyl methylphosphonothiolate	556-75-2	1632.1	1848.2				1187.3		1325.2
Benzophenone	119-61-9	1644.5	1827.2				1200.0		1303.0
Tris[2-(trimethylsiloxy)ethyl]amine	20836-42-4	1645.1	1672.0				1200.0		1144.7
Diphenylmethanol	91-01-0	1651.7	1868.3				1207.4		1343.5
O-Ethyl *S*-2-(diethylamino)ethyl ethylphosphonothiolate	21738-25-0	1670.9	1832.3				1224.7		1308.1
O-Isopropyl *S*-2-(diethylamino)ethyl ethylphosphonothiolate		1680.0	1826.9				1234.0		1302.9
1,2-Bis(2-hydroxyethylthio)ethane	5244-34-8	1702.8	2066.3				1258.6		1543.7
1,2-Bis(2-chloroethylthio)ethane (sesquimustard)	3563-36-8	1703.3	1945.3	1622.7	1688.8	1922.7	1257.6		1419.5
4-Methyl-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide								1268.9	
O-Ethyl *S*-2-(diisopropylamino)ethyl methylphosphonothiolate (VX)	50782-69-9	1713.0	1882.2	1664.1	1710.1	1881.1	1271.1		1359.4
(2-Chloroethylthio)ethyl (vinylthio)ethyl ether				1660.6	1714.1	1907.6			
(2-Hydroxyethylthio)ethyl (vinylthio)ethyl ether				1660.5	1716.6	1971.5			
O-Methyl *S*-2-(diisopropylamino)ethyl ethylphosphonothioate	162085-94-1	1741.0	1905.8				1298.0		1380.2
1,7-Dioxa-4,10-dithia-cyclododecane	294-95-1	1757.0	1965.4	1681.0	1742.9	1939.1	1314.7		1441.4
Bis(2-trimethylsiloxyethylthio)methane		1767.1	1870.2				1325.4		1347.2
Di-*n*-hexyl methylphosphonate								1330.9	
O-Isopropyl *S*-2-(diisopropylamino)ethyl ethylphosphonothioate	162085-95-1	1789.3	1933.7				1348.6		1408.4
S-Ethyl *S*-[2-(diisopropylamino) ethyl]methylphosphonothiolate	110501-55-8			1759.8	1793.1	1934.1			
1,3-Bis(2-hydroxyethylthio)propane	16260-48-3	1809.2	2170.8				1368.8		1649.9
1,3-Bis(2-chloroethylthio)propane	63905-10-2	1810.9	2053.7				1369.4		1530.7
Dibenz[*b*,*f*]1,4-oxazepin (CR)	257-07-8	1810.9	2017.3	1736.0	1800.0	1990.8	1369.5		1495.1
4-Ethyl-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide	1005-93-2	1827.9	2486.6				1388.3	1404.1	1971.0
Bis[2-(diisopropylamino)ethyl] sulfide	6006-58-2			1820.4	1836.1	1902.6			
Methyl benzilate	76-89-1	1853.2	2058.5				1410.6		1536.6
O-Ethyl *S*-nonyl methyl-phosphonothiolate	13088-89-6	1871.9	2047.9				1431.9		1527.7
1,2-Bis(2-trimethylsiloxyethylthio)ethane		1885.1	2000.0				1444.6		1480.3
Dicyclohexyl methylphosphonate	7040-53-1							1441.6	
1,4-Bis(2-hydroxyethylthio)butane	7425-93-6	1928.6	2295.6				1498.9		1774.3
1,4-Bis(2-chloroethylthio)butane	142868-93-7	1932.7	2182.8				1492.8		1661.7
4-Propyl-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide								1509.3	

Table 2 (*continued*)

Compounds	CAS number	C-series					M-series		
		I SE-54	II OV-1701	III OV-1	IV SE-54	V OV-1701	VI SE-54	VII SE-54	VIII OV-1701
4-Isopropyl-2,6,7-trioxa-1-phospha-bicyclo[2.2.2]octane-1-oxide	51052-72-3	1947.4	2609.8				1517.1	1520.2	2090.3
1-Oxa-4,7,10-trithiacyclododecane	40254-02-2			1884.9	1959.9	2190.8			
Bis[(2-chloroethylthio)ethyl] ether (*O*-mustard)	63918-89-8	1990.3	2263.4	1909.9	1982.6	2241.3	1552.7		1742.3
1,3-Bis(2-trimethylsiloxyethylthio)propane		1990.5	2111.6				1552.9		1588.3
4-*tert*-Butyl-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide								1608.6	
4-Butyl-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide								1612.9	
Bis[2-(diisopropylamino)ethyl] disulfide				2036.0	2057.9	2146.0			
1,4-Bis(2-trimethylsiloxyethylthio)butane		2103.2	2228.2				1664.9		1709.2
4-Pentyl-2,6,7-trioxa-1-phospha-bicyclo[2,2,2]octane-1-oxide								1726.9	
Bis[(2-chloroethylthio)ethyl] sufide				2140.7	2222.0	2543.2			
3,9,12-Trithia-6-oxa-14-chloro-1-tetradecane				2160.6	2234.3	2517.3			
1,2-Bis[2-(diisopropylamino)ethylthio] ethane	110501-58-1			2309.8	2338.7	2460.9			
1,7-Dioxa-4,10,13-trithiacyclo-pentadecane	52559-77-0			2171.7	2257.1	2535.5			
1-Methyl-3-piperidyl cyclo-pentylphenylglycolate	7121-51-9	2279.8	2456.9[a]				1845.1		1938.1
1,14-Dichloro-3,9-dithia-6,12-dioxatetradecane				2200.0	2280.4	2577.8			
1-Methyl-4-piperidyl cyclo-pentylphenylglycolate	37830-21-0	2310.4	2490.3[a]				1876.8		1971.7
4-Piperidyl cyclopentyl-phenylglycolate		2350.9	2560.0[a]				1918.7		2060.5
1-Methyl-3-piperidyl cyclo-hexylphenylglycolate	4354-45-4	2374.0	2556.8[a]				1942.3		2035.5
1,9-Bis(diisopropylamino)-3,4,7-trithianonane				2532.9	2567.5	2717.5			
1-Methyl-3-piperidyl benzilate	3321-80-0	2400.0	2633.3[a]				1969.3		2113.5
1-Methyl-4-piperidyl cyclo-hexylphenylglycolate	33445-17-9	2408.4[a]	2593.8[a]				1976.9		2071.3
1-Methyl-4-piperidyl benzilate	3608-67-1	2431.4[a]	2664.1[a]				2000.0		2147.6
4-Piperidyl cyclohexylphenylglycolate		2445.1[a]	2653.8[a]				2017.8		2166.0
4-Piperidyl benzilate	25811-48-7	2466.5[a]	2729.5[a]				2043.9		2237.0[a]
1,14-Dichloro-3,16,12-trithia-9-oxatetradecane				2418.8	2514.5	2864.5			
1,16-Dichloro-3,9,15-trithia-6,12-dioxaheptadecane				2711.8	2813.8	3200.0			
3-Quinuclidinyl benzilate (BZ)	6581-06-2	2628.4[a]	2948.9[a]				2221.6[a]		2416.7[a]
1,10-Dioxa-4,7,13,16-tetrathia-cyclooctadecane	296-40-2	2833.2[a]	3389.6[a]				2394.8[a]		2780.1[a]

I and VI: column SE-54, 25 m × 0.32 mm × 0.25 μm; carrier gas He 1.5 mL min^{-1}; temperature program 40 °C (1 min), heating rate 10 °C min^{-1} to 280 °C for 10 min.[22]

II and VIII: column OV-1701, 25 m × 0.32 mm × 0.25 μm; carrier gas He 1.5 mL min^{-1}; temperature program 40 °C (1 min), heating rate 10 °C min^{-1} to 280 °C for 10 min.[22]

III: column OV-1, 15 m × 0.32 mm × 0.25 μm; carrier gas He 35 cm s^{-1}; temperature program 50 °C (2 min), heating rate 10 °C min^{-1} to 280 °C for 10 min.[27,30,63–65]

IV: column SE-54, 15 m × 0.32 mm × 0.25 μm; carrier gas He 35 cm s^{-1}; temperature program 50 °C (2 min), heating rate 10 °C min^{-1} to 280 °C for 10 min.[27,30,63–65]

V: column OV-1701, 15 m × 0.32 mm × 0.25 μm; carrier gas He 35 cm s^{-1}; temperature program 50 °C (2 min), heating rate 10 °C min^{-1} to 280 °C for 10 min.[27,30,63–65]

VII: column SE-54, 30 m × 0.25 mm × 0.25 μm; carrier gas He 35 cm s^{-1}; temperature program 60 °C (3 min), heating rate 10 °C min^{-1} to 280 °C.[66–71]

[a] Extrapolated.
[b] A diastereometric compound gives two peaks.
[c] Compound does not elute through column.

the same manufacturer with the same column in one or several laboratories, column individuals of the same manufacturer, solvents (diethyl ether, acetone, and ethyl acetate), use of C- and M-standards alone or together and use of a multistep temperature program. There were no significant differences in the reproducibility of RIs determined using different index standard series. M-standards are relatively nonpolar and differ little in their behavior from C-standards on a nonpolar SE-54 phase. The values of indices also differed when the index standard components were reduced in number. However, the reproducibility remained good. According to this study the external standard method, where retention times of the standards are measured in a separate calibration run, may be used, but this requires very good reproducibility of chromatographic conditions and frequent calibration. The most critical parameters affecting the indices were the carrier gas flow rate, the temperature programming rate, and the properties of the column (e.g. length, ID, and film thickness). The influence was greater at the end of the temperature program than at the beginning. Of the model compounds, VX (CAS 50782-69-9) and the weakly volatile CR (CAS 257-07-8) were more affected by nearly every investigated parameter.

3.3.2 Retention Index Libraries

The application of RIs in the identification of organic compounds has been extensively investigated in two comprehensive reviews.[61,62] The RIs of some CWC-related chemicals determined against the C-series using on-column injection have been reported.[27,30,63–65] Sokolowski et al.[66–71] have reported RIs for selected CWC-related chemicals determined against M-series.

A comprehensive RI library was first published by the Finnish Institute for Verification of the CWC.[72] Nowadays the library contains RIs for 145 CWC-related chemicals.[22] RIs have been determined against both M-series and C-series using two different column types (SE-54 and OV-1701: 25 m, 0.32 mm ID and 0.25 μm film thickness) with the following chromatographic conditions: carrier gas helium 1.5 mL min^{-1}, splitless injection with splitless time 45 s, temperature program 40 °C (1 min), and heating rate 10 °C min^{-1} to 280 °C for 10 min. The index values added to the library were the mean values of three measurements. A typical standard deviation for RI values was typically less than 0.5 index units. The correlation between RIs determined using the M-series and the C-series is good and the C-series indices (RI_C) can be calculated from the M-series (RI_M) indices using[71] $RI_C = 0.9666RI_M + 487.6$.

Analysis of all library values for the SE-54 column within M_4–M_{20} and C_8–C_{24} (100 compounds) using linear regression gave correlation 0.99991, and the standard deviation of the error between measured and calculated values was 3.5 index units. A collection of RI values of CWC-related chemicals is given in Table 2.

Because the number of CWC-related chemicals is large, a lot of work is needed before all of them are synthesized and reference data recorded. Some indices may be predicted from the previously determined index values. The RI increases for a homolog about 100 index units when a straight carbon chain is lengthened by one CH_2 unit. Thus, for these kinds of homolog the RIs of the long-chain (six or more carbons) family members can be predicted quite accurately from the RI of the family members having three to five or six carbons. The determined and predicted RI values for three different homolog series differed only by 1.3 index units at maximum.[73] In the case of branched-chain esters of alkyl phosphonofluoridates and their derivatized degradation products the prediction of RIs is more difficult.

For reliable creation of the RI library, all RIs of authentic compounds determined by the Finnish Institute for Verification of the CWC (see Table 2, columns I, II, VI and VIII) are checked against those of special test compounds obtained under the same operating conditions. The RIs of library compounds determined at the beginning and at the end of the library creation must be comparable and consistent. The RI values of test compounds in Table 3 are the mean values recorded during the creation of the RI library by running this test solution after every sixth run. This practice ensures the quality of the library and the analyst can estimate the reliability of the recorded RI values, by running first the test solution and comparing the RI values with those presented in Table 3.

The analysis should be carried out under conditions identical to those used for creating the RI libraries. However, it has been reported that the analytical and column parameters can be changed without altering RIs as long as the initial oven temperature, phase ratio, and the rt_0/β ratio (where r, t_0 and β are the programming rate, dead time, and phase ratio respectively) are kept unchanged. The RI values can be reproduced within few

Table 3 RIs of test compounds

	M-series		C-series	
	SE-54	OV-1701	SE-54	OV-1701
d_9-Dimethyl methylphosphonate	400.0	488.8	879.9	1044.4
2,6-Dimethylphenol	644.7	748.4	1111.8	1288.8
5-Chloro-2-methylaniline	850.3	974.5	1308.2	1508.0
Tri-*n*-butylphosphonate	1213.3	1301.5	1659.3	1824.9
Dibenzothiophene	1330.9	1426.3	1772.9	1951.1
Malathion	1551.1	1722.5	1988.3	2241.7
Methyl stearate	1696.5	1689.1	2132.3	2207.5

index units for two columns of different length and ID working under different heating rates with the same or different carrier gas at different gas flow rates.[74]

4 QUALITY CONTROL

According to the recommended operating procedure for the analysis of treaty-related compounds by GC[22] the solvent blank, sensitivity, and column performance tests should be carried out to check for working and sensitivity of instruments, accuracy of identification parameters, and cross-contamination.

A solvent blank is used to eliminate the possibility of contamination arising from outside the sample, e.g. from the syringe or from the equipment memory effect. It is recommended to use the sample solvent as solvent blank and run this test at the beginning and the end of each sample series and whenever contamination is suspected. The memory effect is possible in the analysis of alkyl phosphoric acids. If the derivatization of these compounds is not quantitative, the nonderivatized acids are adsorbed on the injector glass. The silylation reagent may after this run react in situ on the liner with the adsorbed acids yielding false positive results.

A sensitivity test solution is used to check the sensitivity of the detectors. The test compounds should not adsorb on the column material at low concentration level. It is recommended that test solutions supplied from the instrument manufacturer be used. If these are not available the test compounds presented in Table 4 are suitable for testing the sensitivity of detectors, which are used most often for the detection of CWC-related chemicals.[22]

The sensitivity test is run after the installation of detectors and columns. It is recommended that the test solution be included in the sample series at least every twentieth sample. The criteria for acceptable detector sensitivity are met if each test compound has a signal-to-noise ratio of 10.

A column performance test is done to ensure the condition of columns and the stability of retention parameters. Because CWC-related chemicals differ greatly from each other, the test compounds have been selected so that their physical, chemical and retention properties on the column are different and so that they elute evenly over the whole chromatogram. Use of the following compounds is recommended: d_9-dimethyl methylphosphonate, 2,6-dimethylphenol, 5-chloro-2-methylaniline, tri-*n*-butylphosphate, dibenzothiophene, malathion, and methyl stearate (Table 3). The concentration of test compounds depends on the sensitivity of detectors. The columns should be tested at quite low concentration level, about 10–50 times above the detection limit. The peak shape of d_9-dimethyl methylphosphonate can be used to estimate the column activity. The acidic and basic properties of the column can be estimated by comparing the peak heights and areas of phenol and aniline peaks. The same mixture can be used also for quality control in GC/MS to check the tuning and the performance of the instrument by checking the correctness of the mass spectra and the isotope ratios.

Table 4 Test compounds for checking the sensitivity of FID, NPD and FPD

Detector	Compound	Concentration ($ng\,\mu L^{-1}$)
FID	C_{15}-alkane	5
NPD-N	Azobenzene	1
NPD-P, FPD-P	Malathion	0.3
FPD-S	Dibenzothiophene	2

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detector
CWC	Chemical Weapons Convention
ECD	Electron Capture Detector
EIMS	Electron Impact Mass Spectrometry
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
ID	Inner Diameter
MS	Mass Spectrometry
NPD	Nitrogen–Phosphorus Detector
PAHs	Polycyclic Aromatic Hydrocarbons
PID	Photoionization Detector
RI	Retention Index
RIM	Retention Index Monitoring
SCD	Sulfur Chemiluminescence Detector
TMS	Trimethylsilyl

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Selective Detectors

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis

Gas Chromatography **(Volume 12)**
Instrumentation of Gas Chromatography

REFERENCES

1. A. Verweij, C.E.A.M. Degenhardt, H.L. Boter, 'The Occurrence and Determination of PCH_3-containing Compounds in Surface Water', *Chemosphere*, **8**, 115–124 (1979).
2. D.J. Harvey, M.G. Horning, 'Derivatives for the Characterization of Alkyl- and Aminoalkylphosphonates by Gas Chromatography and Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **79**, 65–74 (1973).
3. J.G. Purdon, J.G. Pagatto, R.K. Miller, 'Preparation, Stability and Quantitative Analysis by Gas Chromatography and Gas Chromatography/Electron Impact Mass Spectrometry of *tert*-Butyldimethylsilyl Derivatives of some Alkylphosphonic and Alkyl Methylphosphonic Acids', *J. Chromatogr.*, **475**, 261–272 (1989).
4. M.L. Shih, J.R. Smith, J.D. McMonagle, T.W. Dolzine, V.C. Gresham, 'Detection of Metabolites of Toxic Alkylmethylphosphonates in Biological Samples', *Biol. Mass. Spectrom.*, **20**, 717–723 (1991).
5. Z. Witkiewicz, M. Mazurek, J. Szulc, 'Chromatographic Analysis of Chemical Warfare Agents', *J. Chromatogr.*, **503**, 293–357 (1990).
6. M. Rautio (ed.), *International Interlaboratory Comparison (Round-robin) Test for the Verification of the Chemical Disarmament, F.1. Testing of Existing Procedures*, The Ministry for Foreign Affairs of Finland, Helsinki, 1990.
7. M. Rautio (ed.), *International Interlaboratory Comparison (Round-robin) Test for the Verification of the Chemical Disarmament, F.2. Testing of Procedures on Simulated Industry Samples*, The Ministry for Foreign Affairs of Finland, Helsinki, 1991.
8. M. Rautio (ed.), *International Interlaboratory Comparison (Round-robin) Test for the Verification of the Chemical Disarmament, F.3. Testing of Procedures on Simulated Military Facility Samples*, The Ministry for Foreign Affairs of Finland, Helsinki, 1992.
9. M. Rautio (ed.), *International Interlaboratory Comparison (Round-robin) Test for the Verification of the Chemical Disarmament, F.4. Validating of Procedures for Water and Soil Samples*, The Ministry for Foreign Affairs of Finland, Helsinki, 1993.
10. M. Rautio (ed.), *Interlaboratory Comparison Test Coordinated by the Provisional Technical Secretariat for the Preparatory Commission for the Organization for the Prohibition of Chemical Weapons, H.1. First Interlaboratory Comparison Test*, The Ministry for Foreign Affairs of Finland, Helsinki, 1994.
11. Evaluation of Results, First Official Inter-laboratory Comparison Test: Proficiency Test, Provisional Technical Secretariat for the Preparatory Commission for the Organization for the Prohibition of Chemical Weapons, The Hague, 14 January 1997.
12. Evaluation of Results, Second Official Inter-laboratory Proficiency Test, OPCW S/10/97, Version 2, 281, Organization for the Prohibition of Chemical Weapons, The Hague, 26 August 1997.
13. Evaluation of Results, Third Official Proficiency Test, OPCW S/23/97, Version 2, Organization for the Prohibition of Chemical Weapons, The Hague, 21 November 1997.
14. 'Standard Operating Procedure (SOP) for Evaluation of Results of OPCW Proficiency Tests', OPCW S/46/98, Version 2, Organization for the Prohibition of Chemical Weapons, The Hague, 31 March 1998.
15. A. Manninen (ed.), *Systematic Identification of Chemical Warfare Agents, Identification of Non-phosphorus Warfare Agents, B.3. Identification of Non-phosphorus Warfare Agents*, The Ministry for Foreign Affairs of Finland, Helsinki, 1982.
16. M. Rautio (ed.), *Air Monitoring as a Means for the Verification of Chemical Disarmament, C.4. Further Development and Testing of Methods, Part III*, The Ministry for Foreign Affairs of Finland, Helsinki, 95–96, 1987.
17. K. Grob, B. Schilling, 'Maximum Column Temperature during On-column Injections of Large Sample Volumes in Capillary Gas Chromatography', *J. Chromatogr.*, **299**, 415–419 (1984).
18. E. Fogelqvist, M. Larsson, 'Large On-column Injections for the Gas Chromatographic Determination of Low-molecular-weight Halocarbons', *J. Chromatogr.*, **279**, 297–306 (1983).
19. K. Grob, Z. Li, 'PTV Splitless Injection of Sample Volumes up to 20 µl', *J. High Resolut. Chromatogr.*, **11**, 626–632 (1988).
20. H.G.J. Mol, H.-G. Janssen, C.A. Cramers, 'Large Volumes Sample Introduction Using Temperature Programmable Injectors: Implication of Linear Diameter', *J. High Resolut. Chromatogr.*, **18**, 19–27 (1995).
21. C.E.A.M. Degenhardt-Langelaan, C.E. Kienz, 'Capillary Gas Chromatography Analysis of Nerve Agents Using Large Volume Injection', *J. Chromatogr. A*, **723**, 210–219 (1996).
22. M. Rautio (ed.), *Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament*, The Ministry for Foreign Affairs of Finland, Helsinki, 1994.
23. M. Rautio (ed.), *Air Monitoring as a Means for the Verification of Chemical Disarmament, C.4. Further Development and Testing of Methods, Part III*, The Ministry for Foreign Affairs of Finland, Helsinki, 15–25, 1987.
24. J. Steinhanses, K. Schoene, 'Thermal Desorption-Gas Chromatography of Some Organophosphates and

S-Mustard after Trapping on Tenax', *J. Chromatogr.*, **514**, 273–278 (1990).
25. J.R. Hancock, J.M. McAndless, R.P. Hicken, 'A Solid Adsorbent Based System for the Sampling and Analysis of Organic Compounds in Air: An Application to Compounds of Chemical Defense Interest', *J. Chromatogr. Sci.*, **29**, 40–45 (1991).
26. A. Kaipainen, O. Kostiainen, M.-L. Riekkola, 'Identification of Chemical Warfare Agents in Air Samples Using Capillary Column Gas Chromatography with Three Simultaneous Detectors', *J. Microcol. Sep.*, **4**, 245–251 (1992).
27. J.R. Hancock, G.R. Peters, 'Retention Index Monitoring of Compounds of Chemical Defense Interest Using Thermal Desorption Gas Chromatography', *J. Chromatogr.*, **538**, 249–257 (1991).
28. J.F.K. Huber, E. Kenndler, G. Reich, W. Hack, J. Wolf, 'Optimal Selection of Gas Chromatographic Columns for the Analytical Control of Chemical Warfare Agents by Application of Information Theory to Retention Data', *Anal. Chem.*, **65**, 2903–2906 (1993).
29. M.L. Shih, R.I. Ellin, 'Determination of Toxic Organophosphorus Compounds by Specific and Nonspecific Detectors', *Anal. Lett.*, **19**, 2197–2205 (1986).
30. P.A. D'Agostino, L.R. Provost, 'Gas Chromatographic Retention Indices of Chemical Warfare Agents and Simulants', *J. Chromatogr.*, **331**, 47–54 (1985).
31. T.G. Albro, J. Lippert, 'Use of the Atomic Emission Detector for Screening and Detection of Chemical Warfare Agents and their Breakdown Products', in *Proceedings of the 1994 ERDEC Scientific Conference on Chemical and Biological Defense Research*, ed. D.A. Berg, National Technical Information Service, Springfield, 171–176, 1996.
32. V.M.A. Häkkinen, 'Analysis of Chemical Warfare Agents in Water by Solid Phase Extraction and Two-channel Capillary Gas Chromatography', *J. High Resol. Chromatogr.*, **14**, 811–815 (1991).
33. H.-Å. Lakso, W.F. Ng, 'Determination of Chemical Warfare Agents in Natural Water Samples by Solid-phase Microextraction', *Anal. Chem.*, **69**, 1866–1872 (1997).
34. M.-L. Kuitunen, K. Hartonen, M.-L. Riekkola, 'Analysis of Chemical Warfare Agents in Soil Samples by Off-line Supercritical Fluid Extraction and Capillary Gas Chromatography', *J. Microcol. Sep.*, **3**, 505–512 (1991).
35. G.A. Sega, B.A. Tomkins, W.H. Griest, 'Analysis of Methylphosphonic Acid, Ethyl Methylphosphonic Acid and Isopropyl Methylphosphonic Acid at Low Microgram per Liter Levels in Groundwater', *J. Chromatogr. A*, **790**, 143–152 (1997).
36. S. Sass, G.A. Parker, 'Structure – Response Relationship of Gas Chromatography/Flame Photometric Detection to Some Organophosphorus Compounds', *J. Chromatogr.*, **189**, 331–349 (1980).
37. M. Minami, D.M. Hui, M. Kasumata, H. Inagaki, C.A. Boulet, 'Method for the Analysis of the Methylphosphonic Acid Metabolites of Sarin and its Ethanol-substituted Analogue in Urine as Applied to the Victims of the Tokyo Sarin Disaster', *J. Chromatogr. B*, **695**, 237–244 (1997).
38. R.L. Shearer, D.L. O'Neal, R. Rios, D.M. Baker, 'Analysis of Sulfur Compounds by Capillary Column Gas Chromatography with Sulfur Chemiluminescence Detection', *J. Chromatogr. Sci.*, **28**, 24–28 (1990).
39. J. Sevci'k, *Detectors in Gas Chromatography*, Elsevier, Amsterdam, 105, 1976.
40. J. Enqvist (ed.), *Chemical and Instrumental Verification of Organophosphorus Warfare Agents, A.1. Chemical and Instrumental Verification of Organophosphorus Warfare Agents*, The Ministry for Foreign Affairs of Finland, Helsinki, 1977.
41. S.E. Eckert-Tilotta, S.B. Hawthorne, D.J. Miller, 'Comparison of Commercially Available Atomic Emission and Chemiluminescence Detectors for Sulfur-selective Gas Chromatographic Detection', *J. Chromatogr.*, **591**, 313–323 (1992).
42. 'Comparison of the Sensitivity of the Flame Photometric Detector and the Atomic Emission Detector for the Analysis of Thiazone', Hewlett-Packard Application Note 228-136, Publication No. (43) 5091-1933E, August 1991.
43. P.L. Wylie, J.J. Sullivan, B.D. Quimby, 'An Investigation of Gas Chromatography with Atomic Emission Detection for the Determination of Empirical Formulas', *J. High. Res. Chromatogr.*, **13**, 499–506 (1990).
44. P.L. Wylie, R. Oguchi, 'Pesticide Analysis by Gas Chromatography with a Novel Atomic Emission Detector', *J. Chromatogr.*, **517**, 131–142 (1990).
45. P.L. Wylie, B.D. Quimby, 'Application of Gas Chromatography with an Atomic Emission Detector', *J. High. Res. Chromatogr.*, **12**, 813–818 (1989).
46. K. Schoene, H.-J. Bruckert, J. Steinhanses, 'Derivatization of Acylating Gases and Vapours on the Sorbent Tube and Gas Chromatographic Analysis of the Products by Atomic Emission and Mass Spectrometric Detection', *Fresenius' J. Anal. Chem.*, **345**, 688–694 (1993).
47. W.R. Creasy, A.A. Rodriquez, J.R. Stuff, R.W. Warren, 'Atomic Emission Detection for the Quantitation of Trimethylsilyl Derivatives of Chemical-warfare-agent Related Compounds in Environmental Samples', *J. Chromatogr. A*, **709**, 333–344 (1995).
48. K. Schoene, J. Steinhanses, H.-J. Bruckert, A. König, 'Speciation of Arsenic-containing Chemical Warfare Agents by Gas Chromatographic Analysis after Derivatization with Thioglycolic Acid Methyl Ester', *J. Chromatogr.*, **605**, 257–262 (1992).
49. J.F. Weinmaster, W.T. Beaudry, P.C. Bossle, M.W. Ellzy, L.G. Janes, D.W. Johnson, J.M. Lochner, S.G. Pleva, J.H. Reeder, D.K. Rohrbaugh, T.E. Rosso, L.J. Szafraniec, L.L. Szafraniec, T.G. Albro, W.R. Creasy, J.R. Stuff, P.B. Smith, I.R. Stewart, 'Chemical Analysis of Environmental Samples Collected in Iraq: Analysis for

the Presence of Chemical Warfare Agents', *Chem. Tech. Biotechnol.*, **64**, 115–128 (1995).
50. W.R. Creasy, J.R. Stuff, B. Williams, K. Morrissey, J. Mays, R. Duevel, H.D. Durst, 'Identification of Chemical-weapons-related Compounds in Decontamination Solutions and other Matrices by Multiple Chromatographic Techniques', *J. Chromatogr. A*, **774**, 253–263 (1997).
51. M.T. Söderström, H. Björk, V.M.A. Häkkinen, O. Kostiainen, M.-L. Kuitunen, M. Rautio, 'Identification of Compounds Relevant to the Chemical Weapons Convention Using Selective Gas Chromatography Detectors, Gas Chromatography/Mass Spectrometry and Gas Chromatography/Fourier Transform Infrared Spectroscopy in an International Trial Proficiency Test', *J. Chromatogr. A*, **742**, 191–203 (1996).
52. E. Kovats, 'Gas-chromatographische Characterisierung Organischer Verbindungen. Teil 1: Retentionsindeces Aliphatischer Halogenide, Alkohole, Aldehyde and Ketone', *Helv. Chim. Acta*, **41**, 1915–1932 (1958).
53. H. Van den Dool, P.D. Kratz, 'A Generalization of the Retention Index System Including Linear Temperature Programmed Gas–Liquid Partition Chromatography', *J. Chromatogr.*, **11**, 463–471 (1963).
54. J. Takacs, M. Rockenbauer, I. Olacsi, 'Determination of the Relationship between Retention Index and Column Temperature in Gas Chromatography through the Temperature-dependence of the Net Retention Volume', *J. Chromatogr.*, **42**, 19–28 (1969).
55. M.L. Lee, D. Vassilaros, C.M. White, M. Novotny, 'Retention Indices for Programmed-temperature Capillary-column Gas Chromatography of Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **51**, 768–774 (1979).
56. J. Enqvist (ed.), *Systematic Identification of Chemical Warfare Agents, B.1. Identification of Organophosphorus Warfare Agents. An Approach for the Standardization of Techniques and Reference Data*, The Ministry for Foreign Affairs of Finland, Helsinki, 1979.
57. J. Enqvist (ed.), *Trace Analysis of Chemical Warfare Agents, C.1. An Approach to the Environmental Monitoring of Nerve Agents*, The Ministry for Foreign Affairs of Finland, Helsinki, 1981.
58. A. Manninen, M.-L. Kuitunen, L. Julin, 'Gas Chromatographic Properties of the M-series of Universal Retention Index Standards and Their Application to Pesticide Analysis', *J. Chromatogr.*, **394**, 465–471 (1979).
59. M. Rautio (ed.), *Air Monitoring as a Means for the Verification of Chemical Disarmament, C.2. Development and Evaluation of Basic Techniques, Part I*, The Ministry for Foreign Affairs of Finland, Helsinki, 1985.
60. M. Kokko, 'Effect of Variations in Gas Chromatographic Conditions on the Linear Retention Indices of Selected Chemical Warfare Agents', *J. Chromatogr.*, **630**, 231–249 (1993).
61. M.V. Budahegyi, E.R. Lombosi, T.S. Lombosi, G. Tarjan, I. Timar, J.M. Takacs, 'Twenty-fifth Anniversary of the Retention Index System in Gas–Liquid Chromatography', *J. Chromatogr.*, **271**, 213–307 (1983).
62. M.B. Evans, J.K. Haken, 'Recent Developments in the Gas Chromatographic Retention Index Scheme', *J. Chromatogr.*, **472**, 93–127 (1989).
63. P.A. D'Agostino, A.S. Hansen, P.A. Lockwood, L.R. Provost, 'Capillary Column Gas Chromatography – Mass Spectrometry of Tabun', *J. Chromatogr.*, **347**, 257–266 (1985).
64. P.A. D'Agostino, L.R. Provost, J. Visenti, 'Analysis of *O*-Ethyl *S*-[2-(diisopropylamino)ethyl] methyl phosphonothiolate (VX) by Capillary Column Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **402**, 221–232 (1987).
65. P.A. D'Agostino, L.R. Provost, 'Gas Chromatographic Retention Indices of Sulfur Vesicants and Related Compounds', *J. Chromatogr.*, **436**, 399–411 (1988).
66. M. Sokolowski, A. Szymanska, 'GC/MS Investigation of Ethyl *N*, *N*-Dimethylphosphoramidocyanidate (Tabun) Age Decomposition Products', in *Proceedings of the 1995 ERDEC Scientific Conference on Chemical and Biological Defense Research*, ed. D.A. Berg, National Technical Information Service, Springfield, 149–155, 1996.
67. M. Sokolowski, 'GC Separation, MS and FTIR Spectra of Series of 4-Alkyl-2,6,7-Trioxa-1-Phosphabicyclo[2,2,2]-Octanes', in *Proceedings of the 1995 ERDEC Scientific Conference on Chemical and Biological Defense Research*, ed. D.A. Berg, National Technical Information Service, Springfield, 703–710, 1996.
68. M. Sokolowski, 'Analytical Data of Some *O*-Alkyl Methylphosphonofluoridates', NATO PfP-SICA Meeting Fabrica Nacional 'La Maranosa', Madrid, Spain, 6–10 May 1996.
69. M.S. Sokolowski, M. Sliwakowski, 'Retention Parameters and Ion-trap Mass Spectra of Symmetrical *O*, *O*-Dialkylmethylphosphonates', in *Proceedings of the 1994 ERDEC Scientific Conference on Chemical and Biological Defense Research*, ed. D.A. Berg, National Technical Information Service, Springfield, 561–567, 1996.
70. M. Sokololowski, 'Mass Spectra and Retention Parameters of some *O*-Alkyl Alkylphosphonic Acids', in *Proceedings of the NATO Advanced Research Workshop on Analytical Chemistry Associated with Destruction of Chemical Weapons*, Brno, Czech Republic, 12–15 May 1996, Kluwer Academic Press, Dordrecht, 99–104, 1997.
71. M. Sliwakowski, 'Ion-trap Mass Spectra and Retention Parameters of Alkylphosphonic Acids Low Aliphatic Alcohols Esters', in *Proceedings of the 1995 ERDEC Scientific Conference on Chemical and Biological Defense Research*, ed. D.A. Berg, National Technical Information Service, Springfield, 711–717, 1996.
72. M. Rautio (ed.), *Standard Operating Procedures for the Verification of Chemical Disarmament, D.2. Second Proposal for Procedures Supporting the Reference Database*,

The Ministry for Foreign Affairs of Finland, Helsinki, 1989.

73. M.T. Söderström, R.A. Ketola, O. Kostiainen, 'Identification of Some Nerve Agents Homologues and Dialkyl Methylphosphonates by Gas Chromatography/Fourier Transform Infrared Spectrometry, Part II: Spectral Search with the Help of Retention Indeces', *Fresenius' J. Anal. Chem.*, **352**, 550–556 (1995).
74. H.L. Yin, Y.L. Sun, 'The Achievement of Reproducible Temperature Programmed Retention Indices in Gas Chromatography when Using Different Columns and Detectors', *Chromatographia*, **29**, 39–43 (1990).

Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Eric R.J. Wils
TNO Prins Maurits Laboratory (TNO-PML), Rijswijk, The Netherlands

1 Introduction 1

2 Electron Impact Mass Spectrometry of Chemical Weapons Convention Scheduled Compounds 3
2.1 General 3
2.2 Alkyl/Cycloalkyl Alkylphosphonofluoridates 4
2.3 Alkylphosphonochloridates and Alkylphosphonate Diesters 5
2.4 *O*-Alkyl/Cycloalkyl *N*,*N*-Dialkylphosphoramidocyanidates 7
2.5 *O*-Alkyl/Cycloalkyl *S*-2-Dialkylaminoethyl Alkylphosphonothiolates 7
2.6 Vesicants 8
2.7 Other Chemical Weapons Convention Scheduled Compounds 9
2.8 Reproducibility of Mass Spectra 10

3 Chemical Ionization Mass Spectrometry of Chemical Weapons Convention Scheduled Compounds 10

4 Special Gas Chromatography/Mass Spectrometry Techniques 12

5 Analytical Derivatization for Gas Chromatography/Mass Spectrometry Analysis 12

6 Gas Chromatography/Mass Spectrometry Analysis Procedures for Environmental and Synthetic Material Samples 15
6.1 General 15
6.2 Air Samples 16
6.3 Synthetic Materials 16
6.4 Preparation of Water Samples 16
6.5 Preparation of Soil and Vegetation Samples 18

7 Gas Chromatography/Mass Spectrometry Analysis Procedures for Biomedical Samples 18

8 On-site Gas Chromatography/Mass Spectrometry Analysis 19

9 Quality Assurance 20

Acknowledgments 20

Abbreviations and Acronyms 20

Related Articles 21

References 21

Mass spectrometry in combination with gas chromatography (GC/MS) is at present the most suitable technique for the analysis of chemicals related to the Chemical Weapons Convention (CWC), as GC/MS is capable of providing the required analytical evidence needed to sustain any claim of noncompliance under the Convention. Chemical analysis will be carried out on-site, during an inspection using mobile GC/MS equipment, or off-site, in at least two designated laboratories selected by the Organization for the Prohibition of Chemical Weapons (OPCW). GC/MS analysis under the Convention is focused primarily on qualitative analysis (unambiguous identification) rather than on quantitative analysis. Moreover, GC/MS analysis has to be established under a strict quality assurance/quality control (QA/QC) program. The two most applied techniques in verification analysis are low-resolution electron impact (EI) and chemical ionization (CI) GC/MS under full scan conditions. EI is the oldest and still most used ionization technique for the analysis of CWC related chemicals. Therefore, special attention is paid in this article to the fragmentation under EI conditions of a number of chemicals belonging to the CWC Schedule list. The chemicals placed on this list are the target for the verification analysis, especially the Schedule 1 chemicals, which encompass the well-known chemical warfare (CW) agents such as the nerve agents sarin (GB), soman (GD), tabun (GA) and O-ethyl S-2-diisopropylaminoethyl methylphosphonothiolate (VX) and the vesicants mustard gas (HD) and lewisite (L). Sample preparation methods for these chemicals and their degradation products in environmental, synthetic material, and biological sample matrices are described in this article.

1 INTRODUCTION

Analytical measurements play a role in underpinning the elements of the verification regime of the CWC. It is difficult to conceive of circumstances where any claim of noncompliance under the Convention could be sustained without soundly based analytical evidence. GC/MS is at present the most suitable technique for the analysis of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

chemicals related to the CWC, as it is capable of providing the required analytical evidence. The provisions for the collection, handling and analysis of samples during an inspection are laid down in paragraphs 52–58 of Part II of the Verification Annex of the Convention text.[1] The Convention addresses the matter of security, integrity, and preservation of samples as well as the confidentiality aspects of the analysis. Special attention is also paid to QA. The collection of samples and the analysis of those samples should be focused on the inspection aims, that is, to gain information relevant to the verification of compliance with the Convention. During inspections at facilities where activities not prohibited by the Convention take place, a balance in analysis must be kept between on the one hand, the need to verify the presence or absence of scheduled chemicals and, on the other hand, the requirement that information not relevant to the CWC should not be revealed. Depending on the inspection scenario, samples shall be taken either by representatives of the inspected facility in the presence of inspectors or by the inspection team itself. Chemical analysis can be carried out on-site, during an inspection, or off-site, in at least two designated laboratories selected by the Technical Secretariat of the OPCW. It is clearly stated in the Convention text that, where possible, the analysis of inspection samples shall be performed on-site. Generally, an on-site laboratory will be small and equipped only with the most essential means, such as a sample preparation kit and a mobile GC/MS instrument. In the event that ambiguities arise during an inspection, the Convention offers the possibility to perform the analysis off-site in designated laboratories equipped with more sophisticated GC/MS instrumentation. In view of the possible political impact of verification analysis, the results reported by an off-site laboratory should be absolutely certain and false positive results are unacceptable. Therefore, a set of identification criteria has to be fulfilled. In principle, these criteria are based on the fact that at least two independent analytical techniques have to be used leading to consistent positive identification results. The two most applied techniques in verification analysis are low resolution EI and CI GC/MS under full scan conditions.

The chemicals placed on the CWC Schedule list are the target for the verification analysis, especially the Schedule 1 chemicals, which encompass the well-known CW agents such as the nerve agents GB, GD, GA, and VX and the vesicants HD and L (see Table 1). Incapacitants such as *w*-chloroacetophenone (code CN, CAS 532-27-4) and *o*-chlorobenzylidenemalononitrile (code CS, CAS 2698-41-1), which have been developed in the past as military tear gases, are not considered in this contribution. The toxic Schedule 1 chemicals, grouped under Schedule 1.A, consist of a specific number of vesicants and toxins, and several families of compounds. The first family (Schedule 1.A.1) encompasses homologous series of alkyl and cycloalkyl (alkyl/cycloalkyl $\leq C_{10}$) alkyl (methyl, ethyl, isopropyl, or propyl)phosphonofluoridates, of which the family of alkyl and cycloalkyl methylphosphonofluoridates is the most important. GB (alkyl : isopropyl) and GD (alkyl : 1,2,2-trimethylpropyl) belong to this family. The trivial name sarin is sometimes used as the reference compound name, for example GF instead of cyclohexyl methylphosphonofluoridate (see Table 1). The second family concerns *O*-alkyl/cycloalkyl *N*,*N*-dialkyl(methyl, ethyl, isopropyl, or propyl)phosphoramidocyanidates (Schedule 1.A.2). These compounds are related to GA (see Table 1). The third family concerns *O*-alkyl/cycloalkyl (H, $\leq C_{10}$) *S*-2-dialkyl(methyl, ethyl, isopropyl, propyl)aminoethyl alkyl(methyl, ethyl, isopropyl, propyl)phosphonothiolates (Schedule 1.A.3). These compounds are related to VX (see Table 1). In principle, the number of substituent permutations in this family is even greater than in the case of the GB and GA families.

Analysis under the Convention is focused primarily on qualitative analysis (unambiguous identification) rather than on quantitative analysis. For the presence or absence of Schedule 1 chemicals no quantitative limits are set in the Convention. Sample preparation methods for GC/MS proceed with a certain efficiency and for accurate quantitative analysis a determination of the recovery efficiency will be necessary. This may be performed by spiking a blank background sample and carrying out the same sample preparation and GC/MS analysis method on the spiked sample. However, background samples and spiking chemicals are not always available. Moreover, accurate quantitative results are only valid when stable analytes have to be determined, but some CW agents are certainly not stable in environmental samples. Although quantitative analysis may not be the prime aim, sample preparation and GC/MS analysis methods should proceed with sufficient efficiency to allow the determination of the compounds of CW interest at a reasonably low level (e.g. $<1\,mg\,kg^{-1}$) in environmental samples.

The analysis of scheduled compounds causes several problems owing to the large variety of these compounds. On the CWC Schedule list, very volatile chemicals such as the blood gases hydrogen cyanide (CAS 74-90-8) and phosgene (CAS 75-44-5) are placed along the very polar nonvolatile toxins saxitoxin (CAS 35523-89-8) and ricin (CAS 9009-86-3). In addition to the common organic CHNO-containing compounds, a great variety of other compounds needs to be considered. These encompass organic phosphorus-, sulfur-, arsenic-, chlorine- and fluorine-containing chemicals, some of which are quite reactive and, hence, unstable in environmental matrices.

Table 1 Chemical information and GC/MS literature references for some important CW agents and related compounds placed on the CWC Schedule list

Systematic name	Code/trivial name	CAS no.	Schedule no.	MS refs.
Isopropyl methylphosphonofluoridate	GB/sarin	107-44-8	1.A.1	2, 5
1,2,2-Trimethylpropyl methylphosphonofluoridate	GD/soman	96-64-0	1.A.1	2, 5, 10
Cyclohexyl methylphosphonofluoridate	GF/cyclohexyl sarin	329-99-7	1.A.1	5
Ethyl *N*,*N*-dimethylphosphoramidocyanidate	GA/tabun	77-81-6	1.A.2	2, 5, 13
O-Ethyl *S*-2-diisopropylaminoethyl methylphosphonothiolate	VX	50782-69-9	1.A.3	2, 3, 5
Bis(2-chloroethyl)sulfide	HD/mustard gas	505-60-2	1.A.4	6, 9, 11
1,2-Bis(2-chloroethylthio)ethane	Q/sesquimustard	3563-36-8	1.A.4	6, 9
2-Chlorovinyldichloroarsine	L 1/lewisite 1	541-25-3	1.A.5	6, 9
Bis(2-chloroethyl)methylamine	HN-2/methyl nitrogen mustard	51-75-2	1.A.6	6
Bis(2-chloroethyl)ethylamine	HN-1/ethyl nitrogen mustard	538-07-8	1.A.6	6
Tris(2-chloroethyl)amine	HN-3/nitrogen mustard	555-77-1	1.A.6	6, 9
Methylphosphonyldifluoride	DF	676-99-3	1.B.9	4, 8
O-Ethyl *O*-2-diisopropylaminoethylphosphonite	QL	57856-11-8	1.B.10	4, 8
Isopropyl methylphosphonochloridate	Chlorosarin	1445-76-7	1.B.11	4
1,2,2-Trimethylpropyl methylphosphonochloridate	Chlorosoman	7040-57-5	1.B.12	4, 10
O,*O*-Diethyl *S*-2-diethylaminoethylphosphorothiolate	Amiton	78-53-5	2.A.1	
1,1,3,3,3-Pentafluoro-2-(trifluoromethyl)-1-propene	PFIB	382-21-8	2.A.2	
3-Quinuclidinyl benzilate	BZ	6581-06-2	2.A.3	6, 12
Dimethyl methylphosphonate	DMMP	756-79-6	2.B.4	2, 4, 7
Diisopropyl methylphosphonate	DIMP	1445-75-6	2.B.4	2, 4
Methylphosphonyl dichloride		676-97-1	2.B.4	4, 8
Methylthiophosphonyl dichloride		676-98-2	2.B.4	4
N,*N*-Dimethylphosphoramidic dichloride		677-43-0	2.B.5	4
Diethyl *N*,*N*-dimethylphosphoramidate		2404-03-7	2.B.6	4, 13
Arsenic trichloride		7784-34-1	2.B.7	4
2,2-Diphenyl-2-hydroxyacetic acid	Benzilic acid	76-93-7	2.B.8	4, 8
3-Quinuclidinol		1619-34-7	2.B.9	4, 8
N,*N*-Diisopropylaminoethyl-2-chloride		96-79-7	2.B.10	4
N,*N*-Diisopropylaminoethane-2-ol		96-80-0	2.B.11	4
N,*N*-Diisopropylaminoethane-2-thiol		5842-07-9	2.B.12	3, 4
Bis(2-hydroxyethyl)sulfide	Thiodiglycol	111-48-8	2.B.13	4, 8, 11
3,3-Dimethylbutane-2-ol	Pinacolyl alcohol	464-07-3	2.B.14	4, 8

GF, cyclohexyl sarin; Q, sesquimustard; HN-2, methyl nitrogen mustard; HN-1, ethyl nitrogen mustard; HN-3, nitrogen mustard; DF, methylphosphonyldifluoride; QL, *O*-ethyl *O*-2-diisopropylaminoethylphosphonite; PFIB, 1,1,3,3,3-pentafluoro-2-(trifluoromethyl)-1-propene; BZ, 3-quinuclidinyl benzilate; DMMP, dimethyl methylphosphonate; DIMP, diisopropyl methylphosphonate.

It is virtually impossible for all these compounds of interest to the Convention to be determined with the highest efficiency by a few general sample preparation and GC/MS analysis methods. Moreover, saxitoxin and ricin cannot be determined by GC/MS, but require analysis by liquid chromatography (LC) and other mass spectrometric techniques (e.g. electrospray ionization, LC/ES/MS). Depending on the type of inspection, analyses may have to be directed towards the declaration and/or the type of facility. This will be the case, in particular, when biomedical samples have to be considered, resulting from an investigation of alleged use of chemical weapons.

2 ELECTRON IMPACT MASS SPECTROMETRY OF CHEMICAL WEAPONS CONVENTION SCHEDULED COMPOUNDS

2.1 General

EI is the oldest and still most used ionization technique for the analysis of CWC-related chemicals. All major mass spectral data collections consist of EI mass spectra, mostly recorded under accepted standardized conditions such as an ionization voltage of 70 eV, an emission current of 100–200 μA and an ion source temperature

of 150–200 °C. Several types of GC/MS system may be applied, for instance, magnetic sector, quadrupole, or ion trap analyzers. Ion trap systems are considered less applicable, when data comparison is required with spectra from a reference library. In particular, basic compounds related to VX or HN-3s tend to produce protonated molecular ions by self-protonation. Magnetic sector and quadrupole mass spectrometers suffer less interference from self-protonation, and spectra produced with these types of instrument are generally reproducible.

Various classes of CW agents and related compounds have been studied by EI/MS. Research on CW agents normally takes place in specialized chemical defense institutes capable of handling the extremely toxic materials. However, work in this area is sometimes restricted and not all information has been submitted for publication, particularly in the past. EI mass spectra of the most well-known nerve agents were published for the first time in 1979,[2] and a vast number of papers on GC/MS analysis of CW agents and related compounds has appeared in the scientific literature since then. Table 1 refers to the main papers dealing with EI mass spectral data and description of the fragmentation processes of CWC-related chemicals. MS of organophosphorus CW agents was reviewed briefly in 1996 as part of a general review of the chemistry of organophosphorus CW agents.[14] EI mass spectral data on 50 possible precursors for the production of CW agents have been presented in the form of an eight-peak index.[4] Several reports (so-called Blue Books) were published between 1979 and 1986, containing mass spectrometric and gas chromatographic data of compounds of CW interest, by the Finnish Institute for Verification of the CWC (VERIFIN). The compilations, in the form of atlases (available upon request), encompass EI and CI mass spectra of organophosphorus[5] and nonphosphorus CW agents[6] and of some degradation products and precursors of these agents.[7,8] The fragmentation of organophosphorus esters, including a number of possible nerve agent precursors, was extensively reviewed in 1976.[15] EI mass spectra of some simple scheduled chemicals used for the production of CW agents, such as pinacolyl alcohol and thiodiglycol (see Table 1), are found in commercially available mass spectral databases.

The major source for reference data is the OPCW Analytical Database. In principle, this database will contain all essential data for the identification of chemicals relevant to the Convention. Reference data obtained by various analytical techniques need to be compiled. In practice this database will encompass GC retention indices (RIs), EI mass spectra, infrared spectra (IR) and nuclear magnetic resonance (NMR) spectral data in addition to the chemical information data (trivial and International Union of Pure and Applied Chemistry (IUPAC) name, CAS number, Schedule number, structural composition, elemental composition, molecular weight MW). It will be an enormous task to collect analytical data on all chemicals relevant to the Convention as represented in the families containing all possible members of the alkyl/cycloalkyl moiety with up to 10 carbon atoms. Although the database is not complete, it contains the EI mass spectra of most well-known CW agents and their degradation products. The fragmentation under EI conditions of a number of chemicals belonging to the CWC Schedule list will be discussed briefly below.

2.2 Alkyl/Cycloalkyl Alkylphosphonofluoridates

Alkyl/cycloalkyl alkylphosphonofluoridates are generally obtained from an organophosphorus precursor, typically an alkylphosphonic difluoride, and an alcohol. In 1992, over 300 aliphatic and cyclic alcohols were found to be commercially available as fine chemicals, of which inexpensive primary and secondary alcohols are considered suitable for chemical weapons production. By this criterion, approximately 70 alcohols remain of prime interest for the preparation of Schedule 1.A.1 chemicals. Only pinacolyl alcohol, the precursor of GD, is contained in the CWC Schedule list. EI mass spectra have been recorded of the 70 corresponding alkyl/cycloalkyl methylphosphonofluoridates, but not for the corresponding alkyl(ethyl, isopropyl, or propyl)phosphonofluoridates. Most of the mass spectra have been recorded in the 1990s as a contribution to the OPCW Analytical Database.

The EI mass spectra of the alkyl and cycloalkyl esters of methylphosphonofluoridates show particular trends. Apart from the methyl esters, all compounds eliminate an alkene/cycloalkene from the ester group. The main fragmentation of, for instance, methyl methylphosphonofluoridate (CAS 353-88-8) leads towards the base peak at mass to electric charge ratio m/z 82, due to the loss of formaldehyde from the molecular ion (m/z 112). Methyl methylphosphonofluoridate is the only compound in the series of the 70 recorded alkyl/cycloalkyl methylphosphonofluoridates, which produces a molecular ion (int. ca. 5%). All other compounds give rise to the phosphorus-containing ion at m/z 99 ($[CH_5OF_2P]^+$), which gives either the base peak or is highly abundant.[2,10,16] This ion results from the loss of an alkenyl/cycloalkenyl radical from the molecular ion and is characteristic of the EI mass spectra of GB homologues. In addition to that particular ion, the EI mass spectra show signals due to the alkyl/cycloalkyl moiety or to phosphorus-containing fragments. Typical characteristic low mass ions are m/z 81 ($[CH_3FOP]^+$), m/z 65 ($[H_2O_2P]^+$), m/z 63 ($[O_2P]^+$) and m/z 47 ($[OP]^+$). Unfortunately, as the number of carbon atoms in the alkyl/cycloalkyl moiety exceeds six, several of the higher phosphorus-containing and hydrocarbon fragments may

have the same nominal mass (e.g. *m/z* 99, 113, 127) and high resolution is required to separate them. As an example, the fragmentation patterns of the nerve agents GB and GD are depicted in Schemes 1 and 2, respectively.

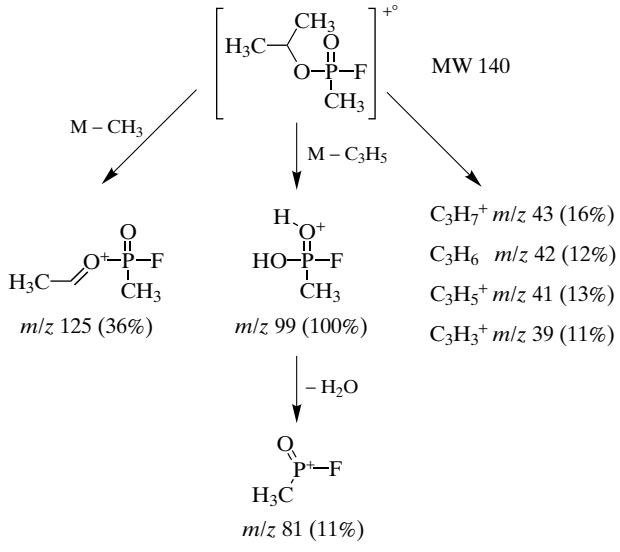

Scheme 1 Fragmentation of GB.

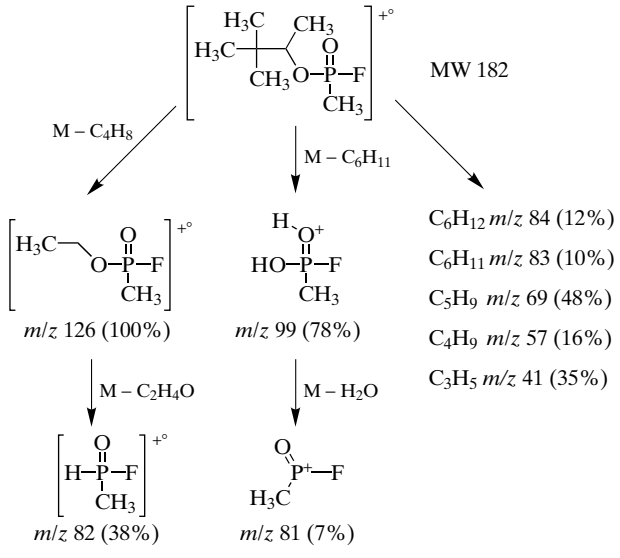

Scheme 2 Fragmentation of GD.

When comparing isomeric *O*-alkyl groups, branching has a large effect on the mass spectra.[16] Branched-chain compounds frequently display high-mass fragments in their mass spectra, due to the loss of alkyl radicals (typically: *m/z* 111, 125, 139, etc.) or due to the loss of alkenes (typically: *m/z* 112, 126, 140, etc.). This is illustrated in Figure 1 by the mass spectra of three different hexyl methylphosphonofluoridates (MW 182), of which the bottom one is from the nerve agent GD. The loss of alkene, in particular, produces information about the branching of the *O*-alkyl chain. Branching at the β-position in the *O*-alkyl chain leads to an intensive peak at *m/z* 112 (see Figure 1a). If the *O*-alkyl chain also contains a methyl group at position one, a strong peak at *m/z* 126 results (see Figure 1c). When the point of branching is further away from the phosphorus atom, the spectra (see Figure 1b) look more like the spectra of the unbranched homologues. Generally, spectra become less informative when branching is further away from phosphorus, because the phosphorus-containing high-mass fragments have a low abundance (<5%).

The high-mass ions in the EI mass spectra of the cycloalkyl homologues have generally low intensities, because fragmentation of the ester chain cannot occur easily. When the cycloalkyl ring is larger than cycloheptyl or substituted, the mass spectra correspond more to those of the corresponding cycloalkenes. For instance, the mass spectrum of 4-*t*-butylcyclohexyl GB corresponds quite well, except for *m/z* 99, to that of 4-*t*-butylcyclohexene. For an unambiguous identification by GC/MS of these substituted cycloalkyl compounds, additional information in the form of GC retention times or RI and CI data is essential. The fragmentation of the other alkyl and cycloalkyl alkyl(ethyl, isopropyl, or propyl)phosphonofluoridates proceeds, in analogy to that of the GB homologues, with the strong characteristic ions at *m/z* 113 (ethylphosphonofluoridates) and *m/z* 127 (isopropyl or propyl)phosphonofluoridates instead of *m/z* 99. Distinction of the propyl group linked to phosphorus can be made, because the *n*-propyl isomer is capable of undergoing the McLafferty rearrangement, whereas the isopropyl isomer is incapable of doing so.

2.3 Alkylphosphonochloridates and Alkylphosphonate Diesters

Alkylphosphonochloridates may be used as starting materials for the synthesis of the corresponding alkylphosphonofluoridates. Two of these compounds, chlorosarin and chlorosoman (see Table 1) are placed on the CWC Schedule 1 list. The main peaks in the EI mass spectrum of chlorosarin (MW 156) at *m/z* 115/117 (loss of C_3H_5) and *m/z* 141 (loss of CH_3) are in accordance with the fragmentation pattern of GB. The same analogy applies to chlorosoman (MW 198) with abundant ions at *m/z* 142/144 (loss of C_4H_8) and *m/z* 115/117 (loss of C_6H_{11}), and GD.[10]

Related to the Schedule 1 family of alkyl/cycloalkyl alkylphosphonofluoridates, are two important classes of

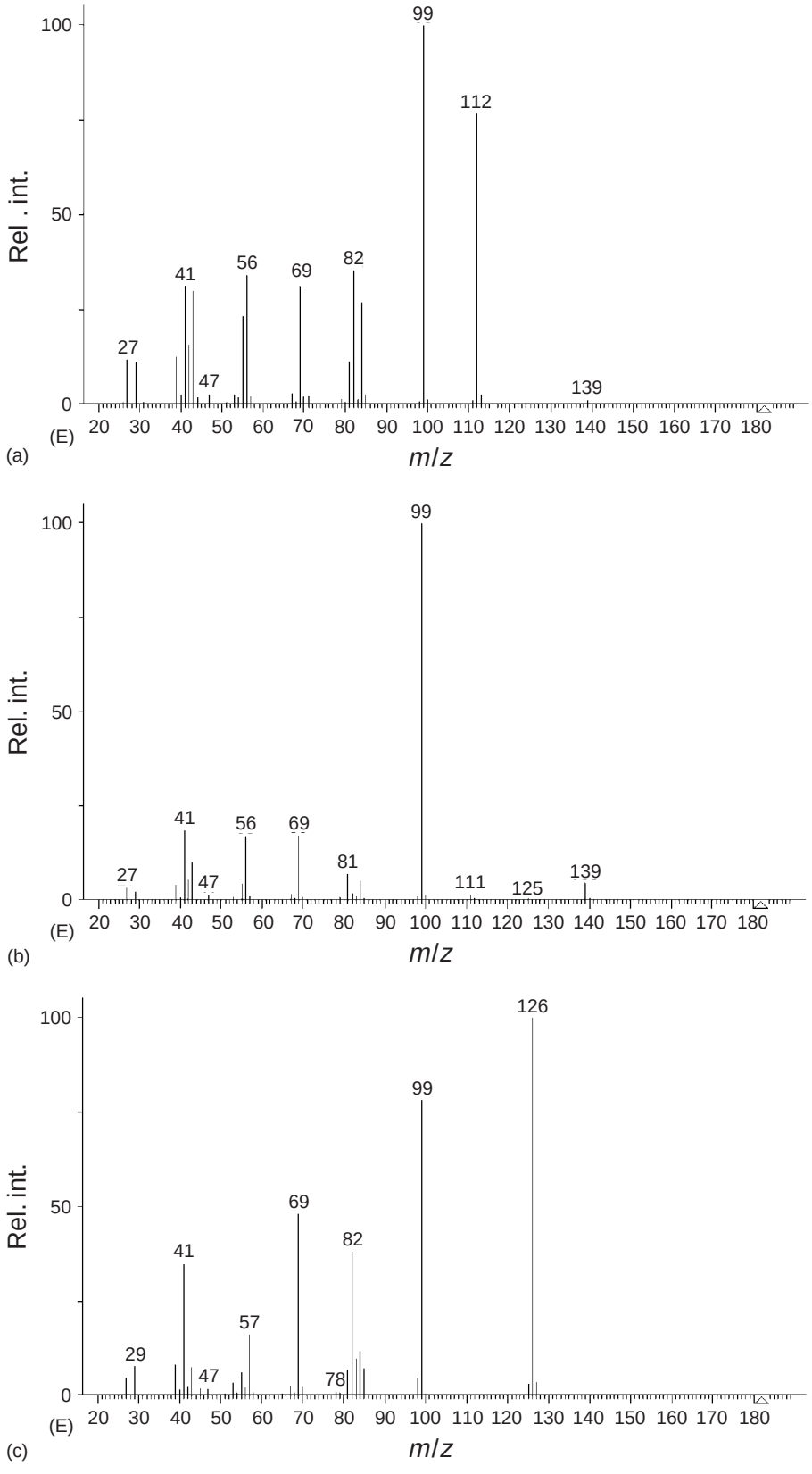

Figure 1 EI mass spectra of (a) 2-methylpentyl, (b) 4-methylpentyl, and (c) 1,2,2-trimethylpropyl methylphosphonofluoridate (GD) recorded at TNO-PML (TNO Prins Maurits Laboratory) on a VG 70-250S GC/MS instrument (Micromass, UK).

compounds. These are the dialkyl/dicycloalkyl alkylphosphonates and the methyl alkyl/cycloalkyl alkylphosphonates, both belonging to the Schedule 2.B.4 chemicals. The first class consists of known impurities of nerve agents and the second class consists of the methyl esters of primarily formed hydrolysis products of nerve agents. DMMP and DIMP (see Table 1) are examples of the first class of compounds and they are frequently used as simulants for nerve agents. These simple esters of methylphosphonic acid were amongst the first phosphorus compounds studied by EI/MS in the early 1960s.[17] The dialkyl esters, including the methyl alkyl esters, fragment by eliminating the alkyl/cycloalkyl group, analogous to alkylphosphonofluoridates. This gives rise to the following characteristic base peaks in the mass spectra:

- *m/z* 97 dialkyl/dicycloalkyl methylphosphonates
- *m/z* 111 dialkyl/dicycloalkyl ethylphosphonates
- *m/z* 125 dialkyl/dicycloalkyl (*n*- or iso)propylphosphonates
- *m/z* 111 methyl alkyl/cycloalkyl methylphosphonates
- *m/z* 125 methyl alkyl/cycloalkyl ethylphosphonates
- *m/z* 139 methyl alkyl/cycloalkyl (*n*- or iso)propylphosphonates.

These base peaks are normally accompanied by abundant signals from the subsequent loss of water.

The primary elimination of an alkenyl/cycloalkenyl radical from the molecular ion in the EI mass spectra of dialkyl/dicycloalkyl alkylphosphonates generally gives rise to the most significant high-mass ion, which in turn will lose the second alkene/cycloalkene group towards the above-mentioned base peak ions. The EI mass spectrum of dicyclohexyl methylphosphonate (MW 260, CAS 7040-53-1), for example, produces a peak at *m/z* 179 ($[M - C_6H_9]^+$) and the base peak at *m/z* 97. As with the EI mass spectra of the alkylphosphonofluoridates, information on branching of the *O*-alkyl chain may be derived from other high-mass fragments.

2.4 *O*-Alkyl/Cycloalkyl *N*,*N*-Dialkylphosphoramidocyanidates

GA is the best known example of the family of *O*-alkyl/cycloalkyl *N*,*N*-dialkylphosphoramidocyanidates. Its fragmentation is rather different, compared with that of the alkyl/cycloalkyl methylphosphonofluoridates. Considerable information on GA and its impurities has become available through the GC/MS analysis of munitions-grade GA.[13,18] The smaller phosphoramidocyanidates produce a molecular ion of moderate abundance (*m/z* 162 in GA is ca. 30%). However, when the *O*-alkyl chain becomes longer, the intensity of the molecular ion signal decreases or vanishes altogether. The EI mass spectrum of GA is dominated by the dimethylamino group (producing *m/z* 42, 43, and 44) and a characteristic rearrangement ion produces an intense peak at *m/z* 70 (see Scheme 3). The rearrangement loss of ethene or ethenyl from the ethoxy group does not occur, but the ethyl group is eliminated instead to produce the *m/z* 133 ion. In contrast, in the mass spectra of *O*-alkyl *N*,*N*-dimethylphosphoramidocyanidates with longer alkyl chains, the loss of the alkenyl radical from the molecular ion leads to a prominent ion or even to the base peak at *m/z* 135 ($[C_3H_8N_2O_2P]^+$). This ion is as characteristic as is *m/z* 99 for the alkyl/cycloalkyl methylphosphonofluoridates.

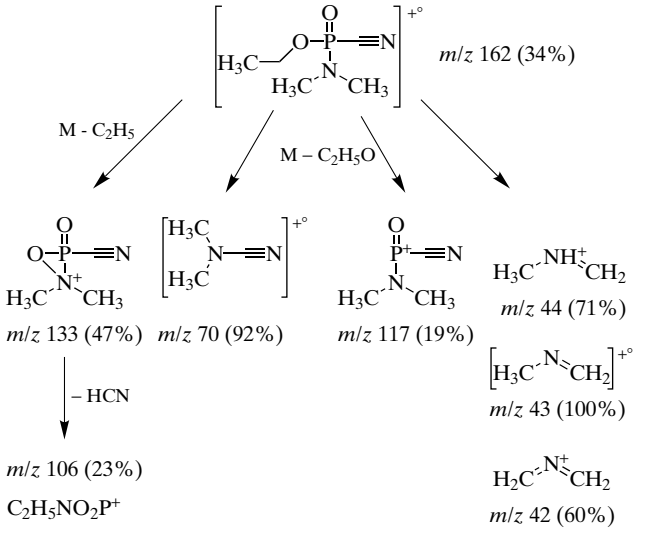

Scheme 3 Fragmentation of GA.

Isomerism of the *O*-alkyl chain does not influence the spectral intensities as much as with the alkyl/cycloalkyl alkylphosphonofluoridates (see section 2.2). Therefore, only mass spectra of a limited number of compounds need to be compiled to identify faithfully any member of the GA family by comparison with database spectra. There is at the time of writing little or no information available on EI mass spectra of higher *N*,*N*-dialkyl(ethyl, isopropyl, or propyl)phosphoramidocyanidates.

2.5 *O*-Alkyl/Cycloalkyl *S*-2-Dialkylaminoethyl Alkylphosphonothiolates

The EI mass spectra of the *O*-alkyl/cycloalkyl *S*-2-dialkylaminoethyl alkylphosphonothiolates are dominated by the *S*-2-dialkyl(methyl, ethyl, isopropyl, propyl)aminoethyl moieties. This leads for VX to extensive ions at *m/z* 127, 114, 72, and 30 (see Scheme 4). As for GA, the GC/MS analysis of munitions-grade VX provided a valid source of information.[14] The phosphorus-containing fragments are characteristic (*m/z* 252, 167, 139), but low in abundance. In the mass spectra of the

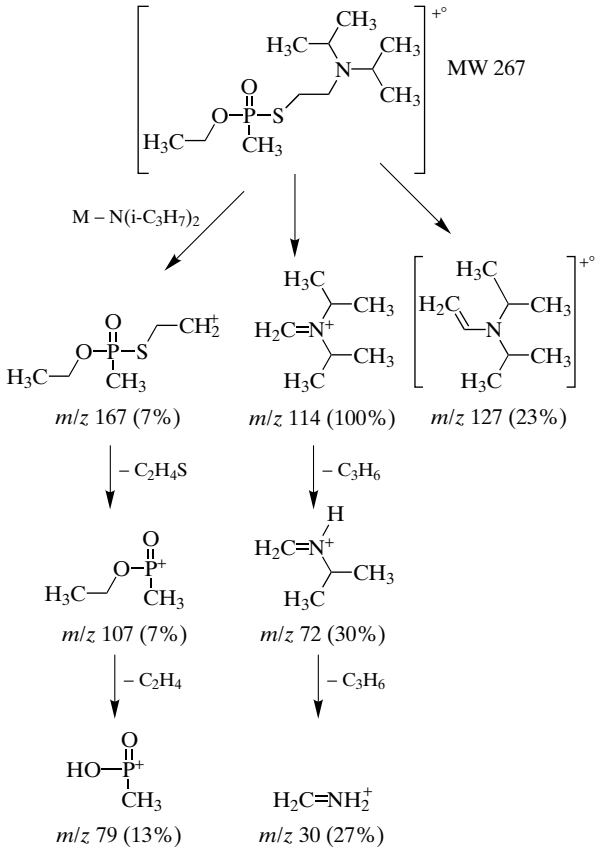

Scheme 4 Fragmentation of VX.

S-2-diethylaminoethyl (producing *m/z* 99, 86, 58) and *S*-2-dimethylaminoethyl (producing *m/z* 71, 58) homologues, corresponding fragmentations occur. In an analogous manner to the behavior of the GA family, branching of the alkyl/cycloalkyl chain does not have a large influence and mass spectra of a few compounds are generally sufficient for the primary identification of VX family members.

2.6 Vesicants

EI mass spectra of HD (see Table 1) and several derivatives, including its main precursor and hydrolysis product thiodiglycol, have been described.[9,11] The fragmentation of HD follows the same pattern as with other dialkyl sulfides. Cleavage of the bond β to the sulfur with charge retention at that atom (thioether cleavage) and cleavage of the C–S bond with charge retention at the chloroalkyl fragment are two major fragmentation routes. With HD (MW 158), this gives rise to the base peak at *m/z* 109/111 and an intensive signal at *m/z* 63/65 (see Scheme 5). MW information can be obtained for HD and Q (see Table 1, MW 218). However, with an increase in the length in the series of HD homologues,

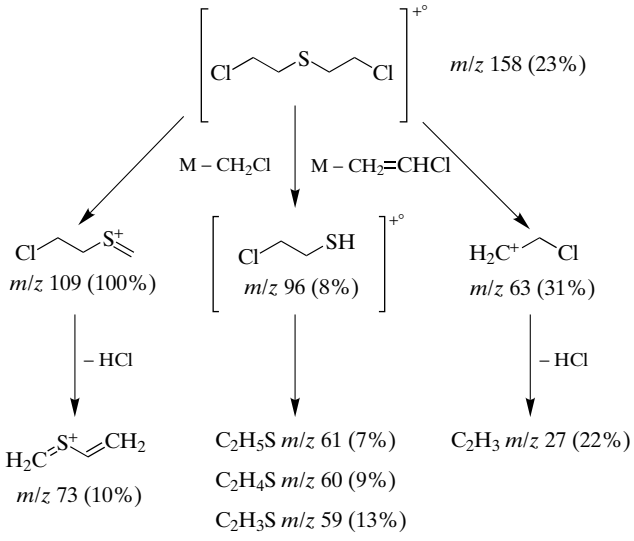

Scheme 5 Fragmentation of HD.

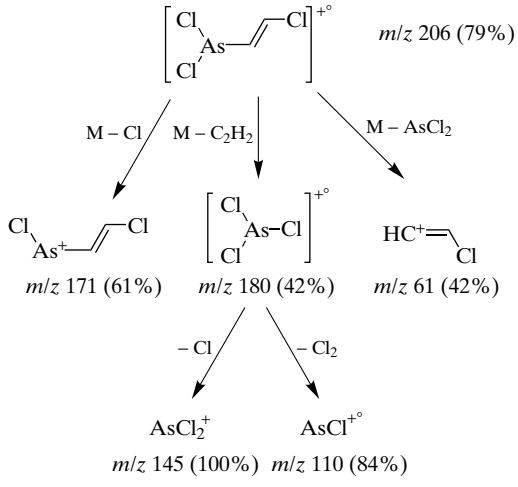

Scheme 6 Fragmentation of L 1.

the intensity of the molecular ion decreases. Cleavage of the C–S bond with elimination of the S-CH_2CH_2Cl moiety produces the base peak at *m/z* 123/125 in the EI mass spectrum of Q. Other strong peaks are at *m/z* 63/65 and *m/z* 109/111.

The fragmentation of L 1 (see Table 1) is rather straightforward.[9] A strong molecular ion cluster at *m/z* 206 is formed under EI conditions and prominent peaks at high mass values are observed, including a rearrangement towards arsenic trichloride at *m/z* 180/182/184 and *m/z* 145/147 (see Scheme 6). The *cis/trans* forms of L 1 have identical EI mass spectra. Technical L consists mainly of the *trans* isomer. The fragmentation of the three HN-3 gases proceeds in accordance with the basic fragmentation rules for amines (α-cleavage). This results in EI mass spectra with the following base peaks:

- *m/z* 106/108 for HN-2 (Cl_1, see Table 1, MW 155);
- *m/z* 120/122 for HN-1 (Cl_1, see Table 1, MW 169);
- *m/z* 154/156/158 for HN-3 (Cl_2, see Table 1, MW 203).

The molecular ion signals in the mass spectra of the HN-3 gases are weak (<5%) and the peak originating from the 2-chloroethyl group (*m/z* 63/65) is rather prominent (>20%).

2.7 Other Chemical Weapons Convention Scheduled Compounds

In addition to chlorosarin and chlorosoman (see section 2.3), there are two families of key precursors placed on the Schedule 1 list. DF (see Table 1) is the best known example of the Schedule 1.B.9 chemicals. The molecular ion, at *m/z* 100, gives the base peak in its EI mass spectrum.[4] The fragmentation is somewhat complicated, but gives abundant peaks for the loss of CH_3O (*m/z* 69) and HF (*m/z* 80). The trivalent alkyl/cycloalkyl 2-dialkylaminoethyl alkylphosphonites (Schedule 1.B.10), which are related to the VX family of compounds, have QL (see Table 1) as their main representative because this is the key precursor for the production of VX by a binary process. All main fragments in the mass spectrum of QL can be attributed to the diisopropylaminoethyl group, and none of the eight strongest peaks are phosphorus-containing fragments.[4]

In contrast to the vast number of Schedule 1.A chemicals, there are only three toxic chemicals placed under Schedule 2.A of the CWC Schedule list, viz. Amiton, PFIB, and BZ (see Table 1). Amiton (MW 269) produces an EI fragmentation pattern analogous to that of the structurally related VX, with strong peaks originating from the *S*-2-diethylaminoalkyl moiety (*m/z* 99, 86, 58). The molecular ion signal at *m/z* 269 is absent; the highest peak of any significance is at *m/z* 197 (ca. 1–2%). PFIB (MW 200) produces a molecular ion and peaks typical of fluorine-containing ions (*m/z* 181, base peak; *m/z* 69, $[CF_3]^+$). The compound is extremely volatile and requires special GC conditions (subambient temperature; special columns) in order to be analyzed by GC/MS. The EI mass spectrum of BZ has been studied along with the spectra of a series of other psychoactive glycolates.[12] The spectrum displays a weak molecular ion signal at *m/z* 337, and strong peaks arising from the diphenylglycolic (*m/z* 183 and 105) and from the quinuclidinyl part (*m/z* 126) of the molecule. The same strong peaks at *m/z* 105 (benzoyl cation) and *m/z* 183 (loss of COOH) are found in the EI mass spectrum of benzilic acid (MW 228, see Table 1), the key precursor for the production of BZ.[4] Peaks from the second BZ precursor 3-quinuclidinol (MW 127, see Table 1) are observed in the EI mass spectrum of BZ. Although BZ and 3-quinuclidinol can be determined as such by GC/MS, the GC performance is improved significantly after derivatization of the hydroxy group. Benzilic acid cannot be determined as such by GC/MS and conversion into a volatile nonpolar analytical derivative is required.

The Schedule 2 list contains two classes of precursors for the production of GA and its homologues. *N*,*N*-Dimethylphosphoramidic dichloride (see Table 1) can be converted to GA with ethanol and a cyanide. Its EI mass spectrum is rather straightforward, with signals due to the molecular ion cluster (*m/z* 161/163/165), to the loss of a hydrogen (*m/z* 160/162/164) and a chlorine radical (*m/z* 126/128), and to the dimethylamino group (*m/z* 42, 43, 44). Diethyl *N*,*N*-dimethylphosphoramidate (see Table 1) is the best-known example of the second class of possible GA precursors. The compound was also detected as an impurity in munitions-grade GA.[13] The molecular ion (*m/z* 181, 15%) fragments by losing an ethyl radical (to *m/z* 162), analogous to GA and subsequent elimination of ethene (to *m/z* 124). The spectrum base peak, at *m/z* 44, is due to the dimethylamino group.

The Schedule 2 list contains three classes of dialkyl-(methyl, ethyl, propyl, and isopropyl)aminoethane compounds, viz. 2-chlorides, 2-ols, and 2-thiols. All three classes are possible precursors of VX and its homologues. In an analogous manner to other simple scheduled chemicals, EI mass spectra are found in commercially available mass spectral databases. Table 1 contains information on the diisopropylaminoethyl compounds. The EI mass spectra of the alcohol (MW 145) and the thiol (MW 161) show the α-cleavage signal at *m/z* 114 (base peak), and typical signals for the subsequent elimination of alkenes at *m/z* 72 and *m/z* 30. In the EI mass spectra of the chloride (MW 163) cleavage takes place preferentially at the *N*-isopropyl group, giving *m/z* 148, and is followed by a loss of propene, resulting in the base peak at *m/z* 106.

In addition to the dialkylaminoethane compounds, organophosphorus precursors are needed for the production of Schedule 1.A.3 chemicals. These precursors, such as methylthiophosphonyl dichloride (see Table 1), are not specifically mentioned in the Convention text, but they belong to Schedule 2.B.4. The EI mass spectra of the closely related and important precursors, methylphosphonyl dichloride (MW 132) and methylthiophosphonyl dichloride (MW 148), are dominated by the molecular ion cluster and display the base peak signal for $[M - Cl]^+$. The relatively stronger molecular ion signal in the mass spectrum of the thio compound reflects the larger stability of the ionized molecule. The higher molecular ion signal in EI mass spectra of organothiophosphorus compounds, compared with the corresponding mass spectra of their oxygen analogues, is a general feature.[15]

Table 2 Major fragments in the EI mass spectra of three nerve agents taken from three different sources

GB				GD				GA			
m/z	Int. (%)[a]	Int. (%)[b]	Int. (%)[c]	*m/z*	Int. (%)[a]	Int. (%)[b]	Int. (%)[c]	*m/z*	Int. (%)[a]	Int. (%)[b]	Int. (%)[c]
41	12	7	13	41	18	37	35	42	29	43	60
42	10	5	12	43	10	15	8	43	72	100	100
43	20	8	16	82	40	50	38	44	43	55	71
79	10	2	3	84	25	10	12	70	100	85	92
81	20	10	11	99	100	85	78	106	29	21	23
99	100	100	100	126	88	100	100	117	29	14	19
125	25	33	36	127	5	4	4	133	43	42	47
								162	32	29	34

[a] Recorded on a Hewlett-Packard Dodecapol Model 5982A (USA), 70 eV, 150 °C source temperature.[2] (Reproduced by permission from S. Sass, T.L. Fisher, *Org. Mass Spectrom.*, **14**, 257–264 (1979) Copyright J. Wiley & Sons.)
[b] Recorded on a Jeol JMS-01-SG-2 magnetic sector instrument (Japan), 75 eV, 260 °C source temperature.[5] (Reproduced by permission of VERIFIN from "Identification of Potential Organophosphorus Warfare Agents", Ministry for Foreign Affairs of Finland, Helsinki, 1979.)
[c] Recorded at TNO-PML on a Micromass VG 70-250S magnetic sector instrument (UK), 70 eV, 200 °C source temperature.

2.8 Reproducibility of Mass Spectra

Reproducibility of the relative intensities in EI mass spectra is a concern, despite the fact that spectra are recorded under standardized conditions. Different mass spectrometers produce spectra with sometimes large spreads in the relative intensities. An impression of the possible variation is given in Table 2. In this table, the relative intensities of major fragments in the EI mass spectra of three nerve agents, taken from three different sources, are presented. It is clear that there are no reference mass spectra for these Schedule 1 chemicals that can be considered as true physical constants. A compilation of more than one spectrum of the same compound in the OPCW Analytical Database gives an indication of the possible spread.

3 CHEMICAL IONIZATION MASS SPECTROMETRY OF CHEMICAL WEAPONS CONVENTION SCHEDULED COMPOUNDS

Although EI is the most widely used ionization method, there is a clear need for CI data to identify CWC-related chemicals by GC/MS. Extensive fragmentation with resultant loss of MW information is particularly observed in EI mass spectra of the organophosphorus compounds of Schedules 1.A.1 (GB family) and 1.A.3 (VX family). As an example, Figure 2(a) shows the EI mass spectrum of the nerve agent VX. The molecular ion at *m/z* 267 is absent. Although the pattern of the peaks is characteristic, complete identification is only possible after observation of the pseudomolecular ion at *m/z* 268 by CI (see Figure 2b). Fragmentation of VX takes place under CI conditions, but by a different mechanism when compared with EI/MS.

Several investigations have been performed on the use of CI in the field of CW agents and related compounds and the influence of the various reagent gases on the selectivity and sensitivity of GC/CI/MS has been described.[2,19–21] Methane, isobutane, ammonia, ethylene, and methanol have been applied in the formation of CI reagent ions. The first three gases are commonly used for GC/CI/MS of scheduled chemicals and methane CI mass spectra of nerve agents were published in 1979.[2] Among these three gases, ammonia has the lowest proton affinity (PA) and methane has the highest, whereas isobutene (neutral formed from isobutane reactant ions) has an intermediate PA. Therefore, CI with methane generally results in more extensive fragmentation than CI with isobutane or ammonia. In addition ammonia often forms adduct ions $[M + NH_4]^+$ with many compounds. The ion source temperature may have an effect on the intensity of the pseudomolecular ions and lower source temperatures are generally favored. The information content of CI mass spectra is typically limited to MW information which is easily accessible from simple calculations. In addition, CI mass spectra vary strongly with CI conditions (choice of gas, etc.) and are more strongly influenced by the instrument used than EI mass spectra. Therefore, compilations of CI/MS data are not as widely used as EI mass spectral databases. Ammonia CI proved to be very useful for the organophosphorus CW agents leading to abundant $[M + H]^+$ and/or $[M + NH_4]^+$ pseudomolecular ions.[19] Owing to their relatively low PA, ubiquitously present hydrocarbons will not be ionized, opening up the possibility for more selective detection of other compounds. The difference between using isobutane or ammonia CI for HD and related compounds has been investigated.[20,21] Both gases provide MW information in the form of either $[M + H]^+$ or $[M + NH_4]^+$ peaks, but isobutane lacks the selectivity of ammonia. This was demonstrated by the detection of the HD impurity

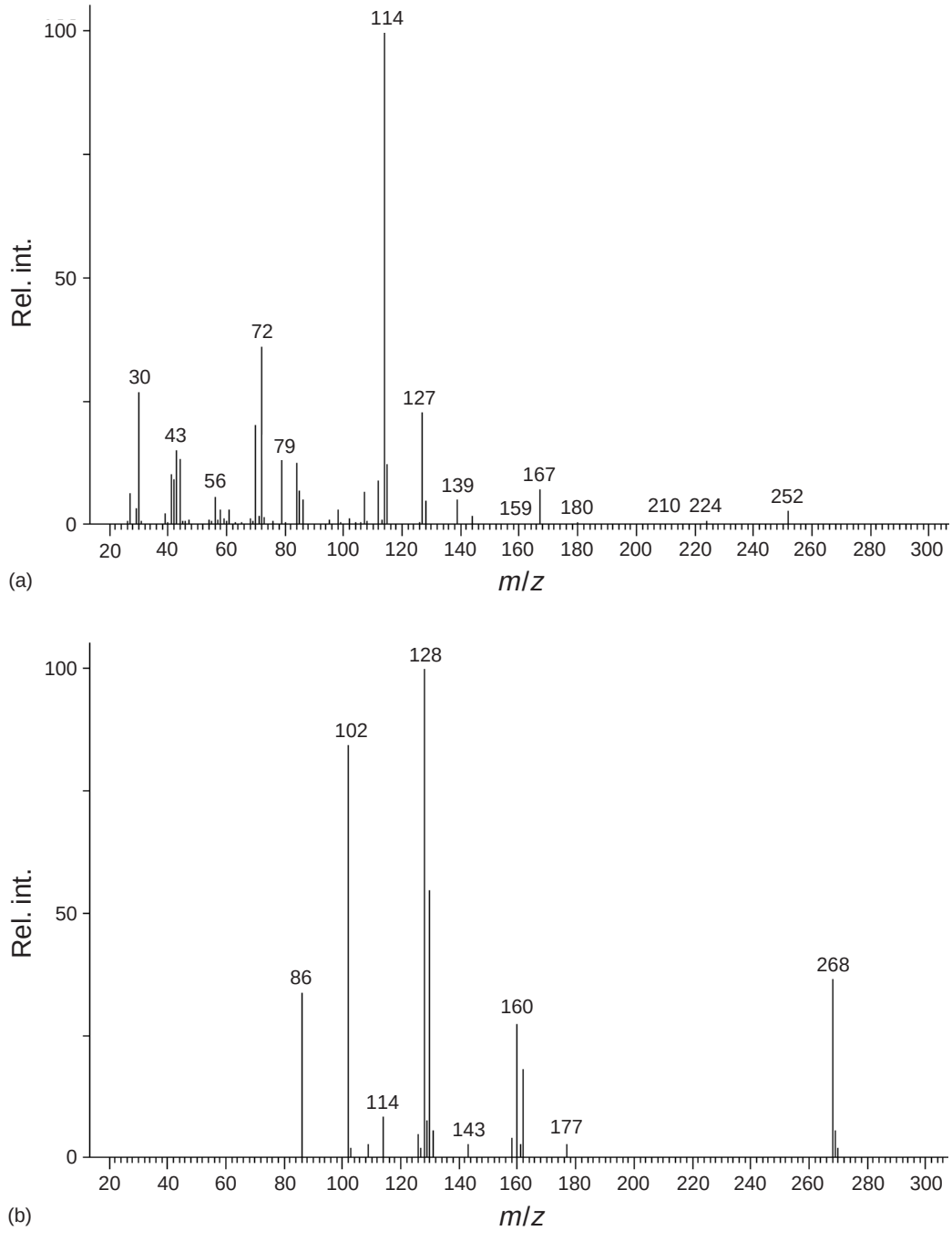

Figure 2 EI (a) and ammonia CI (b) mass spectra of VX recorded at TNO-PML on a VG 70-250S GC/MS instrument (Micromass, UK).

2-chloroethyl (2-chloroethoxy)ethyl sulfide (CAS 114811-38-0) under ammonia CI conditions in the presence of co-eluting hydrocarbons, achieved by ammonia CI but impossible by isobutane CI.[21]

The relatively soft ionization conditions of isobutane and ammonia CI often prevent the observation of fragment ions for identification purposes. This lack of information can be overcome by the use of tandem mass analysis (MS/MS), in which fragmentation can be obtained by collision induced dissociation (CID) of the pseudomolecular ions. Scheduled chemicals have mostly been investigated under positive ion CI conditions, but negative ion CI can be used in principle. However, of the compounds described in section 2 only PFIB

generates negative ions, by electron capture; in this particular case the negative ion CI detection limit lies two orders of magnitude below that observed with EI.

4 SPECIAL GAS CHROMATOGRAPHY/ MASS SPECTROMETRY TECHNIQUES

In cases where low detection limits are required, single ion monitoring (SIM) or multiple ion monitoring (MIM) can be applied. Although less specific than full scan MS, a factor of 100–1000 times greater sensitivity may be reached, to allow the detection of scheduled chemicals at low picogram levels per injection. However, for scheduled chemicals with a strong adsorptive nature, such as the nerve agent VX, this low level of detection may not be feasible. The selection of ion masses for SIM or MIM will depend on the compound and the background of the sample matrix, as well as on the GC column, but in general, high and even masses will be more specific. This may not be the case for the alkyl/cycloalkyl alkylphosphonofluoridates, as the masses arising from the organophosphorus moiety under EI conditions coincide with those of ubiquitously present hydrocarbons. In order to be useful for identification, GC/MS(MIM) has to be performed on at least three individual ions of a significant mass spectral (>10% of the base peak) intensity. The GC peaks obtained during the analysis must have coincident maxima, the same peak-width-at-half-height and a signal-to-noise ratio exceeding three. Furthermore, the retention time must agree with that of an authentic reference compound, and the intensity ratio of the selected ion signals should fall within 10% of that of the reference compound. The comparison of the retention time and spectral intensity ratio should be conducted with the same equipment and under the same experimental conditions. By applying high resolution GC/MS with a magnetic sector mass spectrometer, the exact mass of a fragment can be measured, allowing the calculation of the corresponding elemental composition. This provides a valuable tool for the identification of unknowns.

Some fragments in the EI mass spectra of Schedule 1 chemicals are extremely specific, for example, $[CH_5OF_2P]^+$ (m/z 99.0011) in the spectra of the GB homologues. This ion can easily be distinguished from a possible interfering hydrocarbon background ion $[C_7H_{15}]^+$ (m/z 99.1174), even by the moderate resolution of 5000 (10% valley). High-resolution SIM improves the selectivity and may even allow a higher sensitivity compared with low-resolution SIM, despite the fact that the ion throughput of the instrument is reduced at higher resolution. Detection levels of even less than 1 pg per injection can be reached under optimal conditions.

GC/MS/MS also allows the determination of scheduled chemicals in complex sample matrices with great selectivity and at low detection levels. Information by MS/MS can be obtained either by recording product (parent–daughter) ion spectra of a selected mass or by performing single reaction monitoring (SRM) or multiple reaction monitoring (MRM). In the latter case, a specific fragmentation is monitored, for example by selecting a specific parent ion mass in the first mass analysis step and by transmitting one (SRM) or a few (MRM) formed fragments in the second mass analysis step. Compared with SIM, the selectivity is greatly enhanced.

Several applications of this method have been described in the field of CWC-related chemicals, for instance the determination of small quantities of the nerve agents GB and GD in air contaminated with a huge amount of diesel fuel and sampled on a respirator canister. Reported detection limits for GC/CI/MS/MS(SRM) of ammonium adduct ions were 15 pg for GB (transition m/z 158 to m/z 99) and 80 pg for GD (transition m/z 200 to m/z 183) per injection.[22] This shows that the extra dimension of MS/MS adds to the power of GC/MS in the analysis of complex mixtures.

The application of several of the above described GC/MS techniques to real-world samples (clothing, grave debris, soil, and munitions fragments) collected from a Kurdish village in northern Iraq after alleged CW agents attacks in 1992 demonstrates that unequivocal identification can be carried out on different samples with analyte concentrations ranging from low microgram per kilogram to milligram per kilogram level.[23] As an illustrative example, the result of a GC/MS/MS(MRM) analysis of a metal fragment extracted with dichloromethane is depicted in Figure 3. The transition monitored was m/z 125 to m/z 99 and 81 proving the presence of the nerve agent GB.

5 ANALYTICAL DERIVATIZATION FOR GAS CHROMATOGRAPHY/MASS SPECTROMETRY ANALYSIS

Derivatization of analytes may be necessary in order to make them amenable to GC/MS analysis. In most cases of verification analysis, derivatization involves the conversion of polar nonvolatile degradation products of scheduled chemicals into more volatile nonpolar compounds. As polar compounds are normally in an aqueous matrix, this requires the removal of water. This can be accomplished by evaporation to dryness and redissolution of the residue in a suitable organic solvent. Care

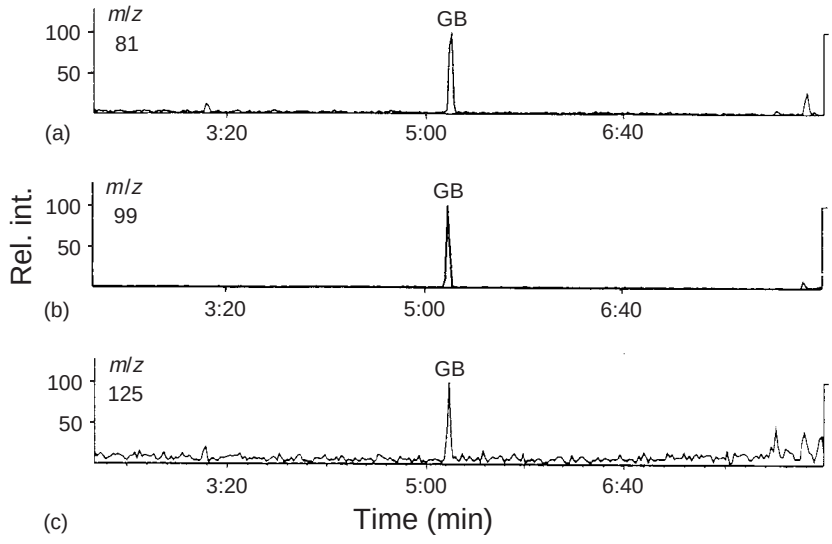

Figure 3 Result of the GC/MS/MS(MRM) analysis of a metal fragment extracted with dichloromethane proving the presence of GB. (c) m/z 125 (unfragmented), (b) m/z 125 → m/z 99, (a) m/z 125 → m/z 81. Data recorded at CBD, Porton Down (UK) on a Finnigan MAT TSQ 700 GC/MS instrument (Finnigan MAT, UK).[23] (Reproduced from R.M. Black, R.J. Clarke, R.W. Read, M.T.J. Reid, *J. Chromatogr. A*, **662**, 301–331 (1994) by permission of Elsevier Science.)

should be taken that polar but nonionic analytes, such as thiodiglycol, do not evaporate. A problem may occur with divalent organophosphorus acids, because they form insoluble salts with metal ions (e.g. Na^+, K^+, Mg^{2+}, Ca^{2+}) present in environmental samples. Removal of the cations with a cation-exchange cartridge before the evaporation to dryness may solve this problem. Contamination of the GC/MS system (injection system, column) may occur, especially with divalent acids such as methylphosphonic acid, in cases where derivatization is incomplete.

There are several derivatization reagents. Most frequently used in verification analysis are the following general reagents:

- Diazomethane: this reagent converts acids into methyl esters and is used for the conversion of alkylphosphonic acids into methyl alkylphosphonates. It is one of the most potent methylation reagents. However, it does not methylate the polar alcohols of the CWC Schedule list, for example thiodiglycol.
- Bis(trimethylsilyl)trifluoroacetamide (BSTFA): among the many commercially available trimethylsilylation reagents BSTFA is a popular one. It converts acids into trimethylsilyl (TMS) esters and polar alcohols into TMS ethers. However, the TMS derivatives are easily hydrolyzed by traces of water and measurements should proceed directly after derivatization.
- *N*-Methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA): this reagent has several advantages above its corresponding TMS analogue, BSTFA. The derivatization occurs readily at room temperature and the derivatives obtained are more stable towards hydrolysis.[24]

Methyl esters of alkylphosphonic acids may also be obtained by heating these acids together with trimethylphenylammonium hydroxide in the hot (e.g. 250 °C) injection port of the GC/MS system.

As an example, EI mass spectra of the different derivatives of isopropyl methylphosphonate (CAS 1832-54-8), the primary hydrolysis product of the nerve agent GB, are presented in Figure 4. The EI mass spectra of the methyl alkylphosphonates produce rather characteristic organophosphorus-containing fragments (see section 2.3), which is an advantage over the silylated alkylphosphonates. In contrast, the behavior of ionized silylated alkylphosphonates is dominated by the TMS or TBDMS moieties.[24] An ion at high mass, $[M - CH_3]^+$ and $[M - C_4H_9]^+$, respectively, is usually observed in the EI mass spectra of the TMS and TBDMS esters. Both esters produce the same base peak, at m/z 153, which reflects the elimination of the alkene group (see Figure 4). The base peak or the $[M - CH_3]^+/[M - C_4H_9]^+$ ion could be selected for SIM or daughter scan by GC/MS/MS. The three derivatization reactions are routinely carried out in most laboratories, with a slight preference for trimethylsilylation. This preference is reflected in the fact that trimethylsilylation was selected for the OPCW on-site sample preparation methods for GC/MS.

In addition to an increase in the volatility of the analytes, derivatization may also bring about an enhancement in the sensitivity of detection. This is particularly relevant

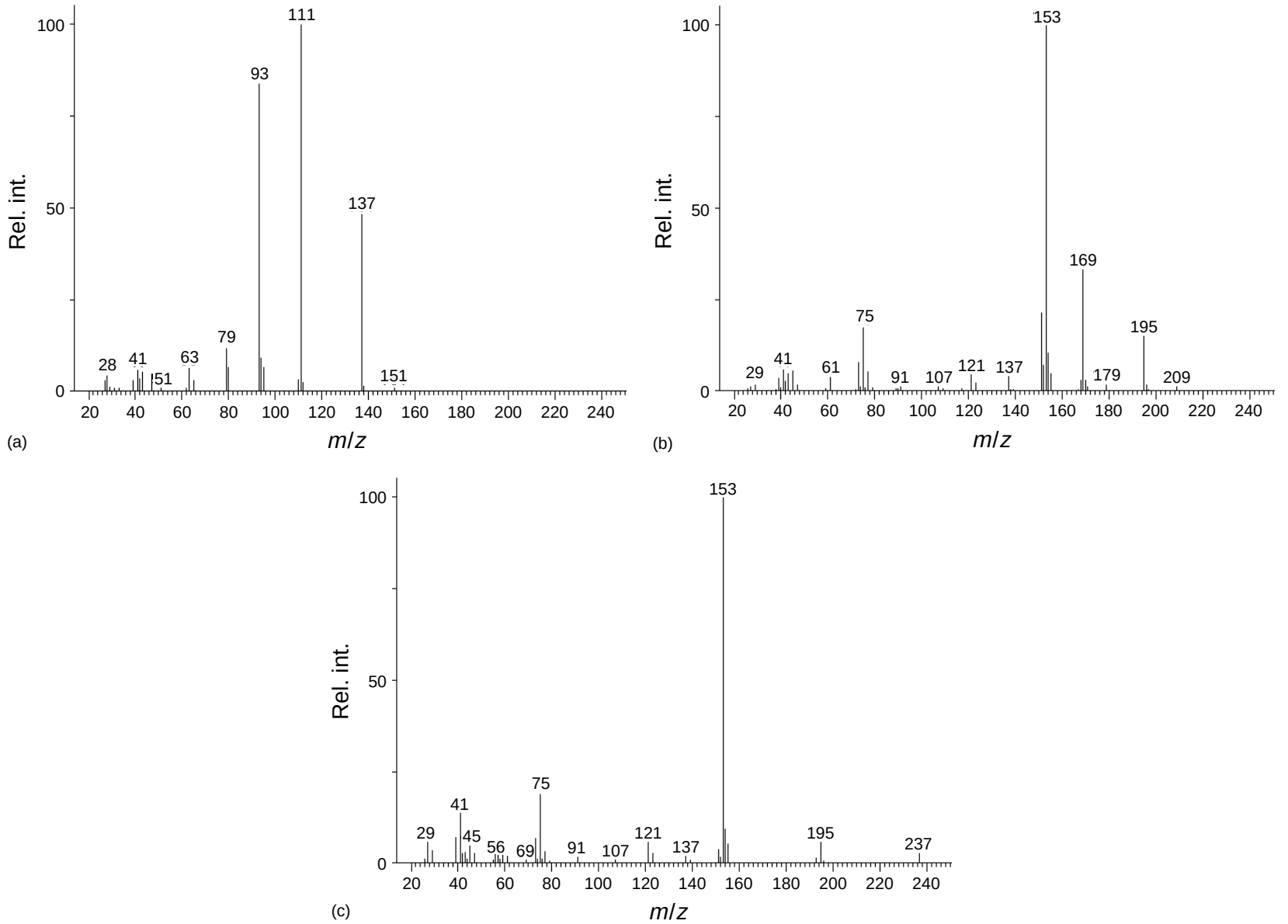

Figure 4 EI mass spectra of (a) isopropyl methyl methylphosphonate (MW 152, CAS 690-64-2), (b) isopropyl TMS methylphosphonate (MW 210), and (c) isopropyl *tert*-butyldimethylsilyl (TBDMS) methylphosphonate (MW 252, CAS 126281-76-3) recorded at TNO-PML on a VG 70-250S GC/MS instrument (Micromass, UK).

with fluorine-containing reagents (perfluorobenzylation, trifluoroacetylation, heptafluorobutylation), followed by GC/MS analysis under electron capture negative ion CI conditions.

A special derivatization reaction is required for L 1, which is so reactive that it cannot be determined by GC/MS in low quantities (e.g. below 10 ng per injection). It has been known for a long time that L 1 reacts with compounds having an α, β-dithiol structure, such as 2,3-dithiopropanol-1 (British-Anti-L; also used for medical treatment). The derivatization reaction can be performed at the analytical level and several examples have been described.[25] The reaction product of L with 3,4-dithiotoluene, 2-(2-chlorovinyl)-5-methyl-1,3,2-benzodithiarsole (see **(1)**), is a useful derivative for GC/MS analysis.

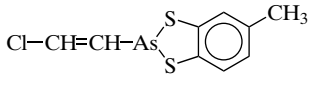

(1)

The reaction product of arsenic trichloride (see Table 1) with 3,4-dithiotoluene, 2-chloro-5-methyl-1,3,2-benzodithiarsole, still contains an active chlorine atom, rendering its determination by GC/MS difficult. The derivatization reaction can also be carried out with 2-chlorovinylarsenic oxide (L oxide, CAS 3088-37-7), which is one of the degradation products of L 1. Thus, the highly reactive arsenous compounds can be detected as less reactive derivatives amenable to GC/MS.

6 GAS CHROMATOGRAPHY/MASS SPECTROMETRY ANALYSIS PROCEDURES FOR ENVIRONMENTAL AND SYNTHETIC MATERIAL SAMPLES

6.1 General

GC/MS is the technique best suited to the identification of CW-related compounds. Nanogram amounts of these compounds are conveniently separated by GC and are then used to produce a full mass spectrum. Complex mixtures can be analyzed within hours. The use of GC for the analysis of CW agents in several matrices has been covered, up until 1990, in an extensive review.[26] Although this review focuses on chromatography, frequent mention of GC/MS underlines the popularity of the method.

Several types of GC column with different stationary phases can be used for separation. The most popular columns in verification analysis are fused silica columns (typically 0.2–0.3 mm inner dimension (ID), 25–50 m length) coated with the following phases:

- 100% methyl polysiloxanes (e.g. OV-1™, SE-30™, CPSil5™ or similar commercial phases);
- 95% methyl and 5% phenyl polysiloxane (e.g. SE-54™, DB-5™, CPSil8™ or similar commercial phases);
- 86% methyl, 7% phenyl, 7% cyanopropyl polysiloxane (e.g. OV-1701™, CPSil19™ or similar commercial phases);
- polyethyleneglycol (Carbowax™).

Most fused silica columns nowadays show little or no bleeding, resulting in a low MS background. Only the Carbowax™ phase may produce severe bleeding at high temperatures. Normally, fused silica columns are directly inserted into the ion source of the GC/MS instrument, producing a reduced pressure at the end of the column. The temperature of the transfer line is typically maintained at 250 °C, and helium is the preferred mobile phase. The nonpolar stationary phases, in particular SE-54™, are recommended for analyzing samples containing unknown scheduled chemicals. When unknowns have to be analyzed, temperature programming is the preferred option, with a typical program running from 50 to 250 °C at 10 °C min^{-1}. All common GC injectors, such as on-column or splitless injectors, may be used for GC/MS. A retention gap is sometimes used with on-column injections to prevent contamination of the analytical column with material from the sample matrix. Branching of the *O*-alkyl ester chain of organophosphorus compounds may introduce an asymmetric center, which together with an asymmetric substituted phosphorus atom creates a number of stereoisomers. Diastereoisomers may even be separated on a conventional capillary GC column. This is, for instance, the case with the nerve gas GD, which usually produces two peaks in a gas chromatogram. Although this is characteristic for identifying GD, it also increases the GC/MS detection limit by a factor of two.

Below, sample preparation methods of various types of sample will be briefly discussed. As a result of international round-robin exercises in the field of verification analysis a number of recommended operating procedures (ROPs) has been published and tested.[27] Further testing of these procedures has taken place since 1995 during proficiency testing under the auspices of the Technical Secretariat of the OPCW. These procedures were developed for the analysis of samples with unknown analytes, in order to allow the identification of as many different chemicals of CWC interest as possible. This approach implies that these procedures can be further optimized in cases where specific target compounds need to be determined by GC/MS.

Apart from thermodesorption, all other sample preparation techniques are directed towards production of a liquid extract that is subsequently injected into a GC/MS

system. For the simplest of the samples, neat organic liquids or concentrated solutions, the sample preparation method will consist of diluting the sample with a suitable, clean solvent (most often dichloromethane).

6.2 Air Samples

Air samples are normally collected on polymeric adsorbents such as Tenax™ or XAD™. Tenax™ tubes are analyzed in the most efficient way by on-line thermodesorption combined with GC/MS. The technique performs very well for thermostable relatively volatile scheduled compounds such as GB, GD, and HD.[28] High boiling nerve agents such as VX may cause problems due to their highly adsorptive nature. Typical desorption conditions imply heating the adsorption tube to 200 °C for 10 min in a specially designed thermodesorption injector in a helium flow and collecting the desorbed volatile compounds in a preconcentrator (a cold trap or the GC column at subambient temperature).

To overcome problems with VX, a reaction with silver fluoride impregnated on a filter in front of the adsorption tube has been applied. The fluoride converts VX to the more volatile ethyl methylphosphonofluoridate (CAS 673-97-2), which is more amenable to GC.[29] In an alternative approach, adsorption tubes can be extracted with a suitable solvent such as ethyl acetate or acetone (chlorinated solvents may dissolve the polymeric adsorption material). To compensate for the loss of sensitivity, compared with thermodesorption, on-column injection of large volumes, up to 0.5 mL, could be considered. This requires a GC/MS instrument equipped with a solvent vapor exit system, in order to maintain the mass spectrometer vacuum and to prevent contamination of the ion source.

6.3 Synthetic Materials

Scheduled compounds contaminating materials such as rubber, polymers, paint, clothing, charcoal, and concrete are normally extracted with two solvents of different polarity. In the course of an international interlaboratory comparison test, acetone and dichloromethane have been evaluated for the extraction of nonpolar scheduled compounds and water has been evaluated for the extraction of polar degradation products.[27] Sonication for 5–15 min is usually sufficient to obtain good extraction efficiencies. Increasing the surface by cutting or crushing the sample into smaller pieces will enhance the efficiency. The amount of solvent used will depend on the sample, because some samples may absorb relatively large quantities of solvent. As a rule, 1 g of a material needs to be extracted with a few milliliters of a solvent. In general, hydrophobic polymeric materials such as rubber and paint will preserve nerve agents and vesicants to a certain extent. Concrete contains alkaline salts that will hydrolyze nerve agents and vesicants. Therefore, analysis of concrete requires extraction and determination of polar degradation products. This will imply a derivatization step before the GC/MS analysis.

Polymeric material samples contaminated with relatively volatile scheduled compounds such as GB and HD may also be subjected directly to thermodesorption analysis like that employed for air samples.[30] In Figure 5(a), part of the resulting total ion current (TIC) chromatogram of a GC/MS analysis is shown for a rubber material contaminated with diesel fuel and the vesicant HD and its impurity mustard disulfide (CAS 1002-41-1). Despite the complexity of the mixture, the vesicants could be positively identified by extracting pieces of rubber with dichloromethane and subsequent GC/MS analysis of the extract. Polymeric materials, especially rubbers, may contain relatively high (>10%) quantities of additives (e.g. plasticizers), which may interfere with the GC/MS analysis. Better results were obtained after thermodesorption (see Figure 5b), because the coelution of the large amount of high boiling hydrocarbons is prevented, with concomitant improvement in the GC performance.

6.4 Preparation of Water Samples

Of the CW agents on the Schedule 1 list, only the nerve agent VX has some stability in water at low pH. Nerve agents of the GB and GA families, and especially vesicants such as HD and L, hydrolyze relatively quickly. Therefore, water samples must be analyzed not only for the scheduled compounds themselves but also for their polar decomposition products (e.g. methylphosphonic acids, thiodiglycol). Some of the hydrolysis products of well-known CW agents, such as GA and L, are unfortunately not placed on the CWC schedule list. Of the possible degradation products of HD, only thiodiglycol is scheduled, whereas 19 sulfur-containing compounds have been identified by GC/MS upon decomposition of munitions-grade HD in water.[31] Although most decomposition products have not been scheduled, the determination of their presence in water provides valuable information about the original presence of scheduled compounds. Water samples can be collected from the environment or from waste tanks at chemical facilities. Measurement and adjustment of the pH with either ammonium hydroxide or HCl will be a necessary step during the sample preparation, especially in the case of waste tank samples.

GC/MS analysis of water samples is frequently preceded by the following sample preparation approaches:

- liquid–liquid extraction with dichloromethane at three different pH values (low, neutral and high)

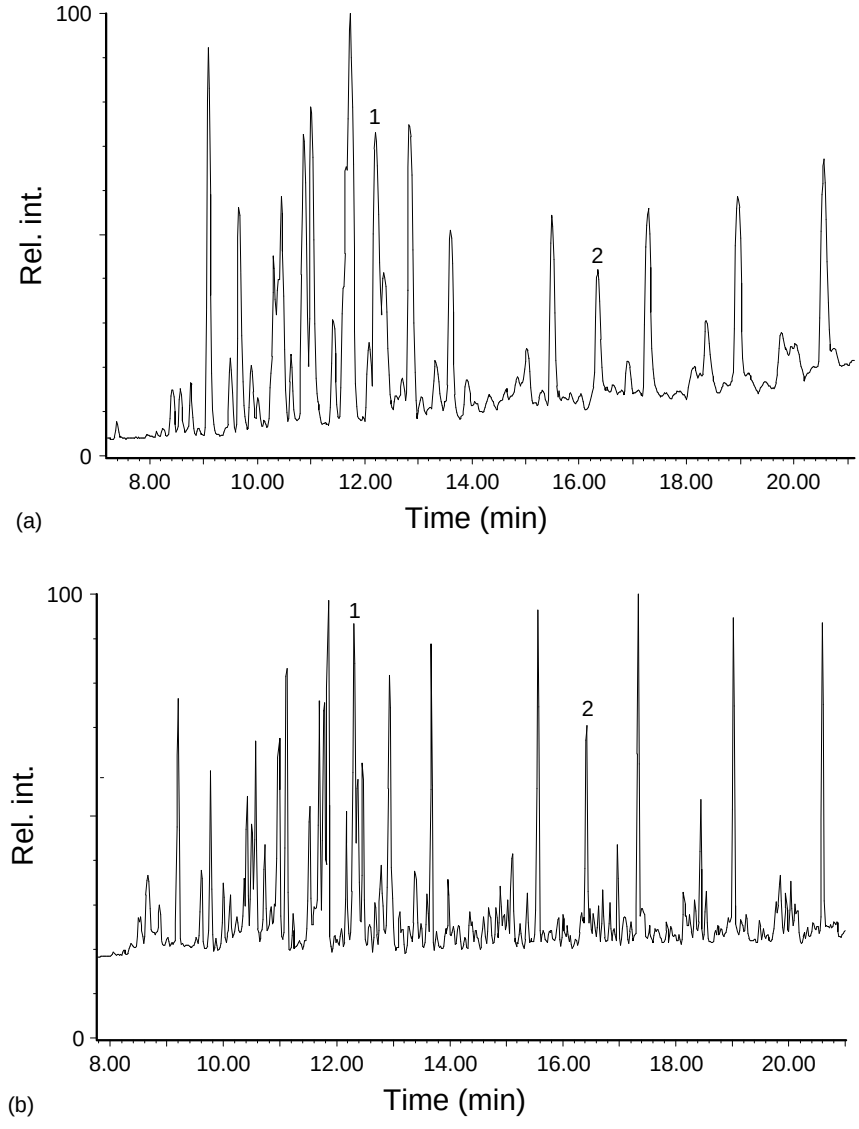

Figure 5 Part of a TIC chromatogram obtained after the GC/MS analysis of rubber contaminated with diesel fuel and sulfur vesicants. (a) Dichloromethane extract, (b) thermodesorption at 120 °C. 1, HD; 2, mustard disulfide. Data recorded at TNO-PML on a VG 70-250S GC/MS instrument (Micromass, UK).[30] (Reproduced from E.R.J. Wils, A.G. Hulst, A.L. de Jong, *J. Chromatogr.*, **625**, 382–386 (1992) by permission of Elsevier Science.)

and subsequently either direct analysis of the organic phase or analytical derivatization of the extract;

- isolation of nonpolar compounds using solid phase extraction (SPE), on a cartridge containing a suitable adsorbent (e.g. C_8-, C_{18}-silica or XAD-4™), or solid phase microextraction (SPME);
- isolation of alkylphosphonic acids by an ion-exchange column (e.g. SAX™ cartridge) followed by extraction with methanol and, subsequently, derivatization of the extract;
- evaporation to dryness followed by derivatization, with subsequent redissolution of the derivatives.

The sample size will be typically 5–10 mL, although applying SPE or ion exchange sometimes allows the use of larger volumes. For instance, the breakthrough volume of nonpolar analytes such as GD or VX on a conventional C_{18}-silica cartridge is over 100 mL. Among the above-mentioned methods, SPME is relatively new. It combines sample preparation and injection of the sample into one step. Analytes are adsorbed on a polymeric fiber coated with a stationary phase such as polydimethylsiloxane, which is thermally desorbed in the injection port at 250 °C. The successful use of the technique for the GC/MS analysis of the nerve agents in water has been described.[32] Levels of less than 60 pg mL^{-1} could be

detected under SIM conditions. This shows that SPME provides a valuable addition to the established extraction methods.

6.5 Preparation of Soil and Vegetation Samples

For the purpose of extraction, soil can be considered to be a solid material that may contain relatively large amounts of water (e.g. clay). The sample preparation methods described for solid synthetic materials and water can often be applied prior to the GC/MS analysis. Soil samples are extracted with a nonpolar solvent (e.g. dichloromethane) and a polar solvent (e.g. demineralized water at pH 7). For nitrogen-containing compounds, such as the decomposition products of the nerve agent VX and HN-3s, extraction with water/methanol at pH > 11 is recommended. Alternatively, methanol containing 1% triethylamine may be used as basic extractant. If the soil contains large quantities of water, nonpolar solvents cannot easily penetrate the pores. In those cases, extraction with a solvent that is more miscible with water (e.g. acetonitrile) is often more efficient. Centrifugation and/or filtration to remove small particles will normally be required before the GC/MS analysis is carried out. It is also recommended that the extract be dried (e.g. with sodium sulfate) and that concentration of the extracts be considered when scheduled compounds are not detected. GC/MS analysis of nonpolar scheduled compounds from soil, is typically preceded by a double extraction of 10 g of the soil sample with 10 mL of dichloromethane (shaking or sonication, during 10 min) and filtration of the combined liquid extracts over a microfilter (0.5–1.0 μm).

Extraction from vegetation may produce complicated sample extracts, because organic plant material (fats, waxes, resins, etc.) will be dissolved as well. Before GC/MS analysis of these samples, a clean-up step is usually performed to remove compounds originating from the matrix. Injection of resins, typically organic acids, could degrade the GC performance dramatically.

7 GAS CHROMATOGRAPHY/MASS SPECTROMETRY ANALYSIS PROCEDURES FOR BIOMEDICAL SAMPLES

During investigations of alleged use of CW agents, samples of biomedical origin may be collected for further analysis in at least two OPCW designated laboratories. The analysis of biomedical samples, such as blood, urine, tissues, and skin, for the presence of scheduled compounds (or their degradation products) generally poses large problems. These problems are due to a lack of knowledge of the fate of most scheduled compounds in biological matrices, usually low concentrations of the scheduled compounds or their metabolites, and interference of the biological matrix. Analysis of biomedical samples still has to be considered as a research, rather than a routine application of GC/MS. In contrast to the situation for environmental samples and synthetic materials, no internationally accepted analysis procedures have been established, neither have the procedures described in literature been validated in international collaborative studies nor has analysis of biomedical samples been the subject of proficiency testing within the context of the CWC. The development of procedures in this field poses the greatest challenge to the analytical chemist. Although GC/MS(/MS) still plays a major role in the analysis of biomedical samples, the role of LC/ES/MS(/MS) is increasing, because this latter method allows direct determination of the mostly polar and, sometimes, high MW metabolites.

Determination of the intact CW agents in urine or blood may proceed by the methods commonly applied to water samples. Extraction with an organic solvent and subsequent clean-up with a Florisil column is a well-established procedure. Rather volatile scheduled compounds can often be successfully recovered and purified from biological materials by means of dynamic headspace stripping and subsequent adsorption on Tenax™ tubes; these tubes are then subjected to GC/MS analysis.

After the use of HD in the Iran–Iraq conflict in the 1980s, the fate of HD in animals has been thoroughly investigated and procedures for the analysis of its metabolic decomposition products have been established.[33] Up to 10 metabolites were identified in the urine of the rats, among which were the compounds thiodiglycol and its sulfoxide. The main metabolites 1,1-sulfonylbis[2-(methylsulfinyl)ethane] [$O_2S(CH_2CH_2SOCH_3)_2$] and 1-methylsulfinyl-2-[2-(methylthio)ethylsulfonyl]ethane [$CH_3SOCH_2CH_2SO_2CH_2CH_2SCH_3$], were due to the reaction of HD with glutathione and subsequent conversion of the glutathione adduct by the enzyme β-lyase. Using GC/MS/MS(SRM) a detection limit of 0.1 ng mL^{-1} urine could be reached, when both compounds were reduced with titanium trichloride to the single analyte 1,1-sulfonylbis[2-(methylthio)ethane] [$O_2S(CH_2CH_2SCH_3)_2$]. The hydrolysis products as well as the β-lyase metabolites could be detected in the urine of human victims of HD attacks.[34]

An alternative approach for the analysis of blood samples from the same group of Iranian HD victims has been described.[35] As HD alkylates amino acids in hemoglobin, adducts will be formed which remain in the bloodstream for some time. Selective cleavage of the alkylated N-terminal valine of the α-chain of hemoglobin was carried out by using the modified Edman reagent pentafluorophenyl isothiocyanate. After derivatization of

the adduct-derived pentafluorophenylthiohydantoin with heptafluorobutyric anhydride, the product is analyzed by GC/MS under negative ion CI conditions.

Intact nerve agents can only be measured shortly after exposure. In blood, they bind primarily and extremely rapidly to the enzymes acetyl- and butyrylcholinesterase. In principle, organophosphate-inhibited butyrylcholinesterase in human plasma is the most persistent and abundant source for biomonitoring of exposure. Treatment of GB-inhibited esterase with fluoride ions (at pH 4), converts the organophosphorus moiety back into GB. The regenerated GB can be isolated by SPE on a C_{18}-cartridge and subsequently be determined by GC/MS. This procedure was successfully applied to serum samples from victims of the terrorist attack in the Tokyo subway (1995) and GB could be detected in concentrations in the range of 0.2–4.1 ng mL^{-1}.[36] An alternative approach to the detection of GB bonded to acetylcholinesterase consists of solubilization of the inhibited enzyme from erythrocyte membranes, digestion with trypsin, and release of the hydrolysis products of GB by alkaline phosphatase digestion. The two hydrolysis products, isopropyl methylphosphonate and methylphosphonic acid, were derivatized to TMS esters, and subsequently measured by GC/MS (EI and CI). In four victims of the terrorist attack in the Tokyo subway, both GB hydrolysis products, were thus identified.[37]

Alkyl methylphosphonic acids are prime targets for analyzing metabolites from nerve agents in serum and/or urine. An extremely sensitive procedure involving a number of steps has been developed allowing the detection of femtogram amounts of these acids in serum/urine samples as well in environmental (soil and water) samples. The procedure is based on the isolation and enrichment of the methylphosphonic acids on an ion-exchange cartridge, the conversion to their pentafluorobenzyl esters and CID of the derivatized methylphosphonate anions obtained under negative ion CI conditions.[38] The procedure has not been applied to samples from victims, as yet.

8 ON-SITE GAS CHROMATOGRAPHY/MASS SPECTROMETRY ANALYSIS

On-site GC/MS analysis at a chemical industry or military facility is limited compared with analysis in a well-equipped off-site laboratory. As instruments have to be sent to and installed at an inspected facility, there are requirements with respect to ruggedness, transportability in the field, power consumption, supportability in the field, and ease of operation. Highly sophisticated instruments such as double focusing magnetic sector or triple quadrupole mass spectrometers will certainly not fulfill these requirements. Additional factors for consideration when using on-site analytical techniques are confidentiality of the recorded data, security against possible tampering, compliance with appropriate safety standards, and relatively short setup and analysis times.

GC and MS can fulfill most of the requirements mentioned above. In fact, both methods have been used for years in mobile laboratories for environmental analysis. Relatively small, robust mass spectrometers are commercially available. A well-known example of such an instrument is the so-called MM1 (Bruker-Franzen, Germany). The MM1 instrument has been built into an armored vehicle that is used as a nuclear biological chemical (NBC) reconnaissance system by defense forces of several nations (e.g. Germany, USA). However, a mobile GC/MS combination system equipped with a suitable capillary GC column is required for the on-site analysis of mixtures. Only this technique can provide sufficient information for the different types of sample to be analyzed on-site. Owing to the present state of technology, ruggedized transportable GC/MS instruments have been developed for use in the field with comparable performance to small or bench-top laboratory instruments. The instruments selected in 1996 (EM640S, Bruker-Franzen, Germany and Viking SpectraTrak, Tradeways, USA)[39] for field work during OPCW inspections have only EI capability, limiting the identification possibilities.

The use of GC/MS during an inspection might be conflicting with the confidentiality aspects of the Convention. Therefore, special software and protocols have been developed by OPCW. In order to ensure that sensitive data does not leave an inspection site, the equipment should be configured in such a way that the stored analytical data as well the results of the database comparison can be separated from the instrument and left on-site. Furthermore, the analysis must not allow the identity of compounds which are not subject to the Convention (blinded detection). In practice, this means that no information (gas chromatogram or mass spectrum) obtained from authentic samples will be displayed. A specially designed software package is used to inspect recorded GC/MS data. Library searches (with forward and reverse fitting) are performed on the raw data, using a library restricted to scheduled compounds or to compounds relevant to a particular inspection aim. If necessary, the software can deconvolute GC peaks attributed to suspect compounds from peaks of co-eluting chemicals. A compound name is only displayed when spectrum and RI match factors exceed a preset value. Various controls have been built into the software to reassure proper functioning of the instrument, because analytical data cannot be made visible. Performance checking is carried out

with regular injection of control samples, to measure RIs (*n*-alkane series) of analytes and unscheduled reference compounds, and by co-injections of hexachlorobenzene with each extract. In this way, ease of operation, ''blinded analysis'', and QA are largely covered by software.

The contents of the OPCW Analytical Database is a key factor in the blinded analysis. The OPCW will base analysis for the presence or absence of chemicals relevant to the Convention on this analytical database. For on-site GC/MS use, a combination of GC and EI data will be required in order to reduce the possibility of false positive identifications.

9 QUALITY ASSURANCE

To guarantee the quality and the security of the samples and the information derived therefrom by analysis, it will be necessary to establish a strict QA/QC program for the GC/MS analysis of chemicals related to the CWC. This program has to be in accordance with an internationally recognized QA system (ISO Guide 25 or equivalent). QA requirements equally apply to on-site analysis in the field and to off-site analysis in OPCW designated laboratories. They encompass the whole chain of events, from taking samples to reporting the analysis of results, with special emphasis on sample preparation and GC/MS analytical methods. Proper coding and sealing of samples and extracts, together with the use of chain of custody forms, has to guarantee that a precise history of the samples can be presented. Storage of the samples and extracts in locked refrigerators helps to preserve the samples and to prevent any tampering. Furthermore, frequent checks of the performance of the GC/MS instrumentation and analyses of blank and control samples are required. As the amount of material to be identified decreases, the number of possible errors during the identification process increases. Particularly, when analysis at nanogram or a lower level has to be performed, cross-contamination could be prevented by the application of a strict QA/QC.

For the GC/MS analysis of unknown chemicals, more or less universal procedures have to be used, because of the large variety in matrices and scheduled compounds. This requires the use of ROPs rather than of standard operating procedures (SOPs), because ROPs allow modification of procedures in the process of analysis (validation on the job). However, SOPs have to be followed for instrument calibration and validation.

A key factor in the QA program is the performance control of the instrumentation. A number of test compounds have been recommended to check the performance of a GC/MS instrument for the analysis of scheduled chemicals.[27] These include DMMP-d_9, trimethylphosphate, 2,6-dimethylphenol, 5-chloro-2-methylaniline, tri-*n*-butylphosphate, dibenzothiophene, malathion, and methyl stearate. DMMP is a moderately polar compound and is considered to be a good test compound for checking the GC performance. The deuterated form is recommended because it does not give cross-contamination in the analysis of authentic samples. However, because no scheduled chemicals can be brought on-site during an inspection, the use of DMMP-d_9 has been replaced by trimethylphosphate for on-site analysis. The correctness of the intensity ratios in the EI mass spectra can be verified by means of the test compounds with different isotopic peaks: 5-chloro-2-methylaniline (isotope ratio m/z 141/143 should be $33 \pm 5\%$) and dibenzothiophene (isotope ratio m/z 184/186 should be $5.9 \pm 1\%$). In this way sensitivity, GC performance, and mass spectrum quality of the described mixtures (m/z and relative intensity values) can be checked.

ACKNOWLEDGMENTS

The author would like to thank staff members of TNO-PML, Albert Hulst and Ad de Jong for recording the mass spectra presented in this paper, and Dr. Ben van Baar and Dr. Maarten Nieuwenhuizen for providing valuable comments on the manuscript. Dr. Robin Black of the Chemical and Biological Defense Section, Porton Down, UK is acknowledged for providing the results of the GC/MS/MS(MRM) analysis of a metal fragment depicted in Figure 3.

ABBREVIATIONS AND ACRONYMS

BSTFA	Bis(trimethylsilyl)trifluoroacetamide
BZ	3-Quinuclidinyl Benzilate
CI	Chemical Ionization
CID	Collision Induced Dissociation
CN	*w*-Chloroacetophenone
CS	*o*-Chlorobenzylidenemalononitrile
CW	Chemical Warfare
CWC	Chemical Weapons Convention
DF	Methylphosphonyldifluoride
DIMP	Diisopropyl Methylphosphonate
DMMP	Dimethyl Methylphosphonate
EI	Electron Impact
ES	Electrospray Ionization
GA	Tabun
GB	Sarin
GC	Gas Chromatography
GD	Soman
GF	Cyclohexyl Sarin

HD	Mustard Gas
HN-1	Ethyl Nitrogen Mustard
HN-2	Methyl Nitrogen Mustard
HN-3	Nitrogen Mustard
ID	Inner Dimension
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
L	Lewisite
LC	Liquid Chromatography
MIM	Multiple Ion Monitoring
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem Mass Analysis
MTBSTFA	*N*-Methyl-*N*-(*tert*-butyldimethylsilyl)-trifluoroacetamide
MW	Molecular Weight
NBC	Nuclear Biological Chemical
NMR	Nuclear Magnetic Resonance
OPCW	Organization for the Prohibition of Chemical Weapons
PA	Proton Affinity
PFIB	1,1,3,3,3-Pentafluoro-2-(trifluoromethyl)-1-propene
Q	Sesquimustard
QA	Quality Assurance
QC	Quality Control
QL	*O*-Ethyl *O*-2-diisopropylamino-ethylphosphonite
RI	Retention Index
ROP	Recommended Operating Procedure
SIM	Single Ion Monitoring
SOP	Standard Operating Procedure
SPE	Solid Phase Extraction
SPME	Solid Phase Microextraction
SRM	Single Reaction Monitoring
TBDMS	*tert*-Butyldimethylsilyl
TIC	Total Ion Current
TMS	Trimethylsilyl
TNO-PML	TNO Prins Maurits Laboratory
VERIFIN	Finnish Institute for Verification
VX	*O*-Ethyl *S*-2-diisopropylaminoethyl Methylphosphonothiolate

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention

Mass Spectrometry **(Volume 13)**
Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry

REFERENCES

1. 'Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction', The Technical Secretariat of the Organization for the Prohibition of Chemical Weapons, The Hague, 1993.
2. S. Sass, T.L. Fisher, 'Chemical Ionization and Electron Impact Mass Spectrometry of Some Organophosphonate Compounds', *Org. Mass Spectrom.*, **14**, 257–264 (1979).
3. P.A. D'Agostino, L.R. Provost, J. Visentini, 'Analysis of *O*-Ethyl *S*-[2-(diisopropylamino)ethyl] methylphosphonothiolate (VX) by Capillary Column Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **402**, 221–233 (1987).
4. E.R.J. Wils, 'Mass Spectral Data of Precursors of Chemical Warfare Agents', *Fresenius' J. Anal. Chem.*, **338**, 22–27 (1990).
5. 'Identification of Potential Organophosphorus Warfare Agents. An Approach for the Standardization of Techniques and Reference Data', Ministry for Foreign Affairs of Finland, Helsinki, 1979.
6. 'Systematic Identification of Chemical Warfare Agents. B.3. Identification of Non-phosphorus Warfare Agents', Ministry for Foreign Affairs of Finland, Helsinki, 1982.
7. 'Identification of Degradation Products of Potential Organophosphorus Warfare Agents, An Approach for the Standardization of Techniques and Reference Data', Ministry for Foreign Affairs of Finland, Helsinki, 1980.
8. 'Systematic Identification of Chemical Warfare Agents. B.4. Identification of Precursors of Warfare Agents, Degradation Products of Non-phosphorus Warfare Agents, and Some Potential Agents', Ministry for Foreign Affairs of Finland, Helsinki, 1983.
9. E. Ali-Mattila, K. Siivinen, H. Kenttämaa, P. Savolahti, 'Mass Spectrometric Methods in Structural Analysis of Some Vesicants', *Int. J. Mass Spectrom. Ion Phys.*, **47**, 371–374 (1983).
10. E.R.J. Wils, A.G. Hulst, 'Mass Spectra of Some Pinacolyl Containing Organophosphorus Compounds', *Org. Mass Spectrom.*, **22**, 763–765 (1986).
11. E.R.J. Wils, A.G. Hulst, 'Mass Spectra of Some Derivatives of 2,2′-Dichlorodiethyl Sulfide (Mustard Gas)', *Fresenius' J. Anal. Chem.*, **331**, 471–474 (1985).
12. A. Vincze, L. Gefen, A. Fisher, P. Bel, 'Low Resolution Electron Impact Mass Spectra of Some Quinuclidine and *N*-Methylpiperidine Glycolates', *J. Forensic Sciences*, **25**(3), 655–665 (1980).
13. P.A. D'Agostino, A.S. Hansen, P.A. Lockwood, L.R. Provost, 'Capillary Column Gas Chromatography/Mass

Spectrometry of Tabun', *J. Chromatogr.*, **347**, 257–266 (1985).
14. R.M. Black, J.M. Harrison, 'The Chemistry of Organophosphorus Chemical Warfare Agents', in *The Chemistry of Organophosphorus Compounds*, ed. F.R. Hartley, John Wiley & Sons, Chichester, 781–840, Vol. 4, 1996.
15. J.M. Desmarcelier, D.A. Wustner, T.R. Fukuto, 'Mass Spectra of Organophosphorus Esters and Their Alteration Products', in *Pesticides Reviews*, ed. F.A. Gunther, Springer-Verlag, New York, 77–185, Vol. 63, and references cited therein, 1976.
16. V.T. Borrett, R.J. Mathews, E.R. Mattsson, 'Verification of the Chemical Weapons Convention: Mass Spectrometry of Alkyl Methylphosphonofluoridates', *Aust. J. Chem.*, **47**, 2065–2074 (1994).
17. R.G. Gillis, J.L. Occolowitz, 'The Mass Spectrometry of Phosphorous Compounds' in *The Analytical Chemistry of Phosphorus Compounds*, ed. M. Halman, Interscience, New York, 295–331, and references cited therein, 1972.
18. P.A. D'Agostino, L.R. Provost, K.M. Looye, 'Identification of Tabun Impurities by Combined Capillary Column Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **465**, 271–283 (1989).
19. P.A. D'Agostino, L.R. Provost, 'Capillary Column Ammonia Chemical Ionization Mass Spectrometry of Organophosphorus Chemical Warfare Agents and Simulants', *Biomed. Environ. Mass Spectrom.*, **13**, 231–236 (1986).
20. P.A. D'Agostino, L.R. Provost, 'Capillary Column Isobutane Chemical Ionization Mass Spectrometry of Mustard and Related Compounds', *Biomed. Environ. Mass Spectrom.*, **15**, 553–564 (1989).
21. P.A. D'Agostino, L.R. Provost, 'Capillary Column Gas Chromatography Ammonia and Deuterated Ammonia Chemical Ionization Mass Spectrometry of Sulfur Vesicants', *J. Chromatogr.*, **600**, 267–272 (1992).
22. P.A. D'Agostino, L.R. Provost, 'Detection of Sarin and Soman in a Complex Airborne Matrix by Capillary Column Ammonia Chemical Ionization Gas Chromatography/Mass Spectrometry and Gas Chromatography/Tandem Mass Spectrometry', *J. Chromatogr.*, **541**, 121–130 (1991).
23. R.M. Black, R.J. Clarke, R.W. Read, M.T.J. Reid, 'Application of Gas Chromatography/Mass Spectrometry and Gas Chromatography/Tandem Mass Spectrometry to the Analysis of Chemical Warfare Samples, Found to Contain Residues of the Nerve Agent Sarin, Sulfur Mustard and Their Degradation Products', *J. Chromatogr., A*, **662**, 301–331 (1994).
24. J. G. Purdon, J.G. Pagotto, R.K. Miller, 'Preparation, Stability and Quantitative Analysis by Gas Chromatography and Gas Chromatography/Electron Impact Mass Spectrometry of *tert*-Butyldimethylsilyl Derivatives of Some Alkylphosphonic and Alkyl Methylphosphonic Acids', *J. Chromatogr.*, **475**, 261–272 (1989).
25. W.K. Fowler, D.C. Stewart, D.S. Weinberg, E.W. Sarver, 'Gas Chromatographic Determination of the Lewisite Hydrolysate, 2-Chlorovinylarsonous Acid, after Derivatization with 1,2-Ethanedithiol', *J. Chromatogr.*, **558**, 235–246 (1991).
26. Z. Witkiewicz, M. Mazurek, J. Szulc, 'Chromatographic Analysis of Chemical Warfare Agents', *J. Chromatogr.*, **503**, 293–357 (1990).
27. 'Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament', ed. M. Rautio, The Ministry for Foreign Affairs of Finland, Helsinki, 1994.
28. J. Steinhanses, H. Schoene, 'Thermal Desorption/Gas Chromatography of Some Organophosphates and *S*-mustard after Trapping on Tenax™', *J. Chromatogr.*, **414**, 273–278 (1990).
29. W.K. Fowler, J.E. Smith, Jr, 'Indirect Determination of *O*-Ethyl *S*-(2-diisopropylaminoethyl) methylphosphonothioate in Air at Low Concentrations', *J. Chromatogr.*, **478**, 51–61 (1989).
30. E.R.J. Wils, A.G. Hulst, A.L. de Jong, 'Determination of Mustard Gas and Related Vesicants in Rubber and Paint by Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **625**, 382–386 (1992).
31. P.A. D'Agostino, L.R. Provost, A.S. Hansen, G.A. Luoma, 'Identification of Mustard Related Compounds in Aqueous Samples by Gas Chromatography/Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **18**, 484–491 (1989).
32. H.-Å. Lakso, W.F. Ng, 'Determination of Chemical Warfare Agents in Natural Water Samples by Solid-phase Microextraction', *Anal. Chem.*, **69**, 1866–1872 (1997).
33. R.M. Black, K. Brewster, R.J. Clarke, J.L. Hambrook, J.M. Harrison, D.J. Howells, 'Biological Fate of Sulfur Mustard, 1,1′-Thiobis(2-chloroethane): Isolation and Identification of Urinary Metabolites Following Intraperitoneal Administration to Rat', *Xenobiotica*, **22**, 405–418 (1992).
34. R.M. Black, R.W. Read, 'Improved Methodology for the Detection and Quantitation of Urinary Metabolites of Sulfur Mustard using Gas Chromatography/Tandem Mass Spectrometry', *J. Chromatogr., B*, **665**, 97–105 (1995).
35. H.P. Benschop, G.P. van der Schans, D. Noort, A. Fidder, R.H. Mars-Groenendijk, L.P.A. de Jong, 'Verification of Exposure to Sulfur Mustard in Two Casualties of the Iran-Iraq Conflict', *J. Anal. Toxicol.*, **21**, 249–251 (1997).
36. M. Polhuijs, J.P. Langenberg, H.P. Benschop, 'New Method for Retrospective Detection of Exposure to Organophosphorus Anticholinesterases: Application to Alleged Sarin Victims of Japanese Terrorists', *Toxicol. Appl. Pharmacol.*, **146**, 156–161 (1997).
37. M. Nagao, T. Takatori, Y. Matsuda, M. Nakajima, H. Iwase, K. Iwadate, 'Definitive Evidence for the Acute

Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention

Vesa M.A. Häkkinen
Finnish Institute for Verification of the Chemical Weapons Convention, Helsinki, Finland

1 Introduction 1
2 Gas Chromatography/Mass Spectrometry in On-site Inspections 1
2.1 Sample Preparation 1
2.2 Instrumentation 2
2.3 Software for Data Handling 3
2.4 Proposed Standard Operating Procedure for Gas Chromatography/Mass Spectrometry Analysis 3
3 Quality Control 4
3.1 Control of Gas Chromatography/Mass Spectrometry Instrument Performance 4
3.2 Criteria for Accepting Gas Chromatography/Mass Spectrometry Instrument Performance 4
3.3 Criteria for Acceptance of Solvent Blank Data 5
3.4 Criteria for Acceptance of Sample Data 5
3.5 Criteria for Acceptance of Positive Match with Organization for the Prohibition of Chemical Weapons Analytical Database 5
4 Comparison to Gas Chromatography/Mass Spectrometry Performed in an Off-site Laboratory 5
5 Discussion 5
Abbreviations and Acronyms 6
Related Articles 6
References 6

This article describes the current procedures for on-site gas chromatographic/mass spectrometric analysis of samples collected during inspections related to the Chemical Weapons Convention (CWC). Such inspections are made by an inspection team compiled by the Organization for the Prohibition of Chemical Weapons (OPCW). The analytical procedures are described in a standard operating procedure covering gas chromatographic analysis, mass spectrometric analysis, data analysis, and quality control. The approved analytical instrumentation for on-site inspections is also described.

1 INTRODUCTION

The convention on the prohibition of the development, production, stockpiling and use of chemical weapons and on their destruction[1] defines strict rules for implementation and verification of it (so called ''Verification annex''). The most important means of verification are on-site inspections to various sites defined in the convention.

The convention allows inspection teams during routine inspections to take samples from locations defined in the facility agreement of each inspected site. During challenge inspections and investigations in cases of alleged use of chemical weapons the inspectors may take samples from any location they consider relevant.

Samples taken during inspections are preferably analyzed on-site in laboratories set up by the inspection teams. The equipment and methods of analysis used in the on-site laboratories are tested and approved by the OPCW. The samples taken during on-site inspections are prepared according to standard operating procedures (SOPs) and analyzed by gas chromatography/mass spectrometry (GC/MS). The GC/MS instrument is a specially designed portable system with sophisticated data evaluation software, which is operated in blinded mode in most inspections.

2 GAS CHROMATOGRAPHY/MASS SPECTROMETRY IN ON-SITE INSPECTIONS

2.1 Sample Preparation

Samples taken during inspections are prepared according to SOPs, which are methods specially designed and tested for on-site laboratories. The sample preparation methods, although trying to be as simple as possible, are rather time-consuming requiring several extractions and derivatizations to allow reliable GC/MS analysis of a wide range of chemicals related to the CWC. The detailed descriptions of the sample preparation methods are given elsewhere in this encyclopedia.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2.2 Instrumentation

Several expert groups have thoroughly evaluated all inspection equipment for OPCW, including GC/MS instruments. The demand of portability has been a great challenge for instrument manufacturers, since most existing commercial benchtop instruments were according to the OPCW specifications nonportable. Only two instruments, the American SpectraTrack 672 and the German Bruker EM 640S, fulfilled the specifications of portability. Since the evaluation period the American SpectraTrack 672 instrument has been retired from the market. The German Bruker EM 640S instrument uses a quadrupole mass analyzer and electron impact (EI) ionization. The gas chromatograph is a specially designed, miniaturized device. The Bruker EM640 has a separate gas chromatography (GC) module which is easily mounted on the mass spectrometer. Instead of changing columns, which would be difficult because of the very small size of the equipment, the whole gas chromatograph is changed. Several GC modules may be used for inspection where a large number of samples is expected for analysis. The instrument is operated with a separate ruggedized personal computer (PC). Experts from the OPCW selected the Bruker EM 640S to be used in on-site inspection work.

2.2.1 Bruker EM 640S Mobile Gas Chromatography/ Mass Spectrometry Instrument

The Bruker EM 640S is a ruggedized compact analytical system consisting of the mass spectrometer/GC peripheral modules, a ruggedized PC, and an external power supply.

The mass spectrometer is based on a capillary direct inlet system with a 70 $L\,s^{-1}$ turbomolecular drag pump and an integrated membrane prepump. The only external connections needed are electrical power and helium at a maximum of 2 bar. The ion source is a variable EI source, with 70 eV/200 mA as standard settings, with two changeable filaments. The mass analyzer is an all-glass one-piece quadrupole filter with a dual detector system. The mass range is 1–640 amu with a line width variable from 0.1 amu to 4 amu, the standard line width being 1 amu.

The GC peripherals concept is based on various mass spectrometer inlet and injection modules. The GC modules are directly coupled to the mass spectrometer and modules with different analytical capillary columns can be interchanged with each other. Also the injection system's split/splitless injector, desorber, automatic air sampler and dual port injector are easily changeable (Figure 1).

The ruggedized PC is designed for use in on-site inspections. It is an Intel-based PC with effective dust protection. The keyboard has an integrated trackball and a touch-screen monitor. The OPCW requirements

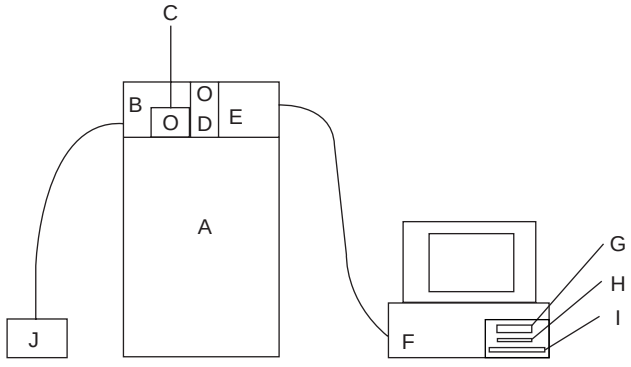

Figure 1 The schematic structure of the Bruker EM640S Portable GC/MS instrument: A, mass spectrometer; B, injector module; C, GC separator module; D, gas supply module; E, electronics module for peripherals; F, PC; G, removable PC card hard disk; H, floppy disk drive; I, optical disk drive; J, power supply.

for data handling and data storage are met by using a removable and exchangeable PC card hard disk and data storage medium. Both storage media may be kept at the inspection site or with the inspection team. After inspection neither the spectrometer nor the data system contains any inspection-related data.

The instrument is packed in four light transport cases with a total weight less than 70 kg. The first case contains the mass spectrometer EM 640S, which is operable in or out of the case. The power input for the instrument is 24 VDC/600 VA. The second case contains the GC periphery and the consumables, including split/splitless injector, GC and gas supply modules mounted on a module plate, electronic module with mounted set of cables, programmable air sampling pump, adsorbent tubes, automatic air sampler module, thermal desorber module, combined injector and thermal desorber module, and a tool kit. The third case contains an external power supply for the EM 640S with a power input of 110–240 VAC, 50–60 Hz/1.5 kVA and a power output of 24 VDC/1 kVA. The fourth case contains the PC, with exchangeable PC card hard disk, magneto-optical drive with write once, read many (WORM) medium, 10-inch touch-screen monitor, and a keyboard with integrated trackball. The power requirement for the PC is 115–230 VAC, 50–60 Hz/3.1 A.

2.2.2 On-site Operational Requirements

The EM 640S GC/MS system is designed for on-site operation and can tolerate a large temperature range, dust, and humidity. Nevertheless, some options/minimum requirements must be taken into consideration. The instrument should be set up in a moderately aired and dry place. A well-ventilated tent in a hot region or a container in a cold region would be sufficient. Direct rain on to the

instrument should be avoided. Dust and humidity are generally no problem for the instrument because of its sealed housing, but dust or sand can damage the fittings of the gas chromatograph. For working place ergonomic reasons, the GC/MS instrument should preferably be set up on, say, a table with enough room for the instrument, analytical tools, and PC.

The power supply for the instrument is 22–30 VDC at an average power consumption of 500 VA, which makes the system operable by batteries. The 24 VDC instrument input power is generated from regular net current voltages by the external power supply. Its power input is variable between 110 and 230 VAC/50–60 Hz at a maximum of 1.4 kVA. The power output is 24 VDC at a maximum of 1 kVA. The PC requires an additional 200 W at 110 or 220 VAC.

The whole instrument configuration requires an overall maximum power of approximately 1 kVA and a generator with 1.2–1.4 kVA electrical power output is sufficient for the on-site operation.

Helium is used as carrier gas with a maximum inlet pressure of 2 bar for the injectors. The average helium consumption per working day is approximately 50 L. This means a 5-L helium bottle at 150 bar keeps the system running for approximately 2 working weeks.

2.3 Software for Data Handling

On-site analysis sets special requirements for the data handling. The fear that confidential business information might be revealed when samples are analyzed on-site leads to very special arrangements when evaluating the analytical results obtained during inspections.

The inspectors are not allowed to see the actual raw data to visually investigate the obtained chromatograms and the corresponding mass spectra of a sample taken and analyzed during an inspection. Each set of raw data (a sample run) is analyzed automatically against the library which is preselected for each on-site mission according to the inspection scenario (e.g. to investigate the presence of sarin, its precursors or degradation products on the inspected site).

This so-called blinded software was developed by the National Institute of Standards and Technology (NIST) in cooperation with the OPCW according to the requirements and rules for on-site inspections agreed in the Conference of the State Parties in The Hague. The current, also commercially available, version of this software is called AMDIS.

The AMDIS software looks for ion patterns in the total ion chromatogram, which are similar to the spectra included in the current library. The software is also capable of finding compounds which co-elute with another, strongly interfering compound. The identification of compounds has been thoroughly tested with difficult cases, when the target compound has been covered by a large peak of co-eluting compound. These tests have shown that the software is capable of reliable identification in such an extreme situation. The reliability of the identification result can be increased further by simultaneous retention index monitoring.

The AMDIS software produces a standard report from each GC/MS run. The report shows the names of the identified compounds, their retention parameters, and the spectral match factors calculated against the current library. The compounds used for the quality control of the instrument must be included in the library to ensure the correct performance of the instrument during the on-site analysis.

2.4 Proposed Standard Operating Procedure for Gas Chromatography/Mass Spectrometry Analysis

The on-site analysis procedure follows the so-called SOP[2] which was specially designed to establish a GC/MS method to provide reliable analytical results in field conditions using blinded software. The analytical procedure has been thoroughly tested with several chemicals related to the CWC to produce reliable and repeatable results at concentration level of 1 mg kg^{-1} or higher.

2.4.1 Standard Gas Chromatography/Mass Spectrometry Conditions

The conditions and operating parameters for the Bruker EM640 instrument are recommended by the instrument manufacturer. Typical EI source conditions are 70 eV electron energy and a source temperature of 150–200 °C. The recommended scanning values (scan range/scan speed) are 40–500 amu/1.0 s.

The standard gas chromatograph operating parameters are as follows:

- column: 95%-dimethyl-5%-diphenyl siloxane (SE-54) stationary phase, 25 m column length, 0.25 mm inner diameter and 0.25 μm film thickness;
- injector port temperature: 250 °C;
- temperature program: 40 °C (1 min) 10 °C min^{-1} to 280 °C (10 min);
- carrier gas: helium;
- carrier gas flow rate: 40 cm s^{-1};
- injection mode: splitless injection, splitless time 1 min;
- injection volume: maximum 1 μL.

Changes in the GC operation parameters can be made if the samples to be analyzed so require. All changes made in the parameters must be carefully documented in the analysis report.

2.4.2 The Sequence for Gas Chromatography/Mass Spectrometry Analysis

The sequence of steps for GC/MS analysis of the samples, each performed under the control of the GC/MS data software is as follows:

1. tune and calibrate the GC/MS instrument (see section 3.1);
2. determine the performance of the GC/MS instrument by analyzing the test mixture by co-injecting it with alkane mixture (see section 3.1);
3. analyze the solvent blank by co-injecting it with 10 ng each of hexachlorobenzene (HCB), 4-bromofluorobenzene (BFB), and pyrene (a blank control mixture) (see section 3.3);
4. analyze the sample by co-injecting it with 30 ng each of HCB, BFB, and pyrene (see section 3.4);
5. repeat step 3 if:
   - there is a positive mass spectral match with the OPCW analytical database for a chemical other than from the blank control mixture (see section 3.5);
   - there is a high chemical background (dilute the sample before the new injection);
   - the next sample has a different solvent;
6. repeat step 2 as the last GC/MS measurement or if criteria in section 3.4 are not met.

3 QUALITY CONTROL

3.1 Control of Gas Chromatography/Mass Spectrometry Instrument Performance

Checking instrument performance is essential for the reliability of GC/MS analysis. The relevant aspects to be considered are sensitivity of the method (detection limits), reproducibility of retention indices, and quality of the produced mass spectrometry (MS) data.

The procedure for controlling the performance is divided into two stages:

1. The first stage is automatic tuning and calibration of the GC/MS instrument by using perfluorotributylamine according to the instructions of the instrument manufacturer.
2. The second stage is evaluation of the GC/MS instrument performance by injection of a mixture containing 10 ng of each test chemical (test mixture). The test chemicals are listed in Table 1. As stated, 10 ng of each *n*-alkane (C_8-C_{24}, even numbers, alkane mixture) is co-injected with this test mixture.

Table 1 Recommended test mixtures for determining the performance of an on-site GC/MS instrument

Chemical	Function
Performance test mixture	
Trimethylphosphate	GC performance
2,6-Dimethylphenol	GC performance
5-Chloro-2-methylaniline	MS performance
Tri-*n*-butylphosphate	GC performance
Dibenzothiophene	MS performance
Malathion	GC performance
Methyl stearate	GC performance
Sample control mixture	
HCB	Sample control
BFB	Sample control
Pyrene	Sample control
Alkane mixture	
n-Alkanes, C_8-C_{24}, even numbers	Retention index calibration

The absolute retention times of the *n*-alkanes are used to calculate retention indices.

Step 2 must be carried out before and after the analysis of a series of samples or if a malfunction of the GC/MS instrument is indicated.

The chemicals of the test mixture have been selected according to the following criteria:

- to contain no scheduled chemicals of CWC;
- to allow the checking of the adsorption behavior of a GC column for chemicals with different chemical properties; and
- to include chemicals containing the elements chlorine and sulfur, both having a distinctive isotope for checking the correct ionization process of the instrument.

The performance of the GC/MS instrument during the analysis of samples is checked by comparing the measured values of retention index and isotope pattern with reference values for the sample control mixture which is co-injected together with each sample.

3.2 Criteria for Accepting Gas Chromatography/Mass Spectrometry Instrument Performance

The GC/MS instrument is accepted for use if the gas chromatographic and mass spectrometric data produced from at least six of the seven test chemicals determined by the GC/MS data analysis software meet the following requirements:

- The mass spectra of the detected test chemicals have a match value greater than 85 (maximum 100).

- The intensity ratios of the isotopic ions are correct. The value is expressed as a percentage of the less abundant isotopic ions relative to the more abundant one ± the acceptable error limits. This means: *m*/*z* 143/141 ($^{37}Cl/^{35}Cl$) for 5-chloro-2-methylaniline is 33% ± 5% relative intensity.
- The retention indices of the detected test chemicals are within ±20 retention index units of the reference values (the retention indices of the test chemicals are calculated by the GC/MS data software using the absolute retention times of the co-injected *n*-alkanes).
- The signal-to-noise ratio of the chromatographic peak of the detected test chemicals in the total ion chromatogram is greater than five.

3.3 Criteria for Acceptance of Solvent Blank Data

The solvent blank data are accepted if the data for all three co-injected blank control chemicals meet the following requirements:

- mass spectral match factor is greater than 85;
- retention indices differ less than ±20 index units from the reference values;
- no positive mass spectral match of other chemicals of the OPCW Analytical Database is made.

If these criteria are not met, repeat section 2.4.2, step 3, and continue from there.

3.4 Criteria for Acceptance of Sample Data

The sample data are accepted if the data for at least two of the three co-injected control chemicals meet the following requirements:

- mass spectral match factor is greater than 80;
- when the retention index is available the measured index is within ±40 index units compared with the reference index (actual number to be determined by instrument testing).

If these criteria are not met, repeat section 2.4.2, step 2, and continue from there.

3.5 Criteria for Acceptance of Positive Match with Organization for the Prohibition of Chemical Weapons Analytical Database

Positive match for other than test, control, and retention index chemicals is accepted if all of the following requirements are met:

- sample data are accepted according to section 3.4;
- mass spectral match factor is greater than 80;
- when the retention index is available the measured index is within ±40 index units compared with the reference index (actual number to be determined by instrument testing).

4 COMPARISON TO GAS CHROMATOGRAPHY/MASS SPECTROMETRY PERFORMED IN AN OFF-SITE LABORATORY

The only principal difference between the on-site and off-site GC/MS analysis is the access to the raw data. During an on-site inspection the analyst has to rely on the report printed automatically by the AMDIS software. The extensive quality control system created to maintain the correct performance of the on-site instrument should guarantee the correctness and reliability of the analytical result even without access to the actual raw data.

In an off-site laboratory the analyst has the possibility to use his/her own expertise to evaluate the raw data. The obtained total ion or selected ion chromatograms can be viewed to look for co-eluting compounds. The spectra can be printed out and compared visually to library spectra. All relevant data can be annexed in the analytical reports to give the real scientific proof of the existence or nonexistence of any chemical included in each inspection scenario.

Technically the on-site instrument is of high quality. The sensitivity it provides is not far from that of large off-site laboratory instruments. The drawback is the lack of ionization methods other than EI ionization. In many cases the use of chemical ionization is required for measuring the molecular mass of a chemical to improve the reliability of the identification.

5 DISCUSSION

The methods for GC/MS in on-site analysis of chemicals related to the CWC were developed by the OPCW in cooperation with NIST and the Finnish Institute for Verification of the CWC (VERIFIN). The instrument manufacturer (Bruker-Franzen Analytik GmbH) also actively participated in this cooperation by developing their instrument to meet the requirements set by the OPCW to solve the difficult analytical task of blinded reliable identification of numerous chemicals with variable chemical and physical properties.

At this time it can be concluded that the laborious project has been successful. The complicated analytical procedure with extensive quality control ensures correct and sensitive instrument performance during the analyses, the NIST AMDIS software provides automatically reliable blinded results from each analyzed sample, and the Bruker EM640S GC/MS instrument has shown its working capabilities in real on-site conditions.

ABBREVIATIONS AND ACRONYMS

BFB	4-Bromofluorobenzene
CWC	Chemical Weapons Convention
EI	Electron Impact
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HCB	Hexachlorobenzene
MS	Mass Spectrometry
NIST	National Institute of Standards and Technology
OPCW	Organization for the Prohibition of Chemical Weapons
PC	Personal Computer
SOP	Standard Operating Procedures

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Mass Spectrometry **(Volume 13)**
Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. 'Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction', The Technical Secretariat of the Organization for the Prohibition of Chemical Weapons, The Hague, 1993.
2. 'Standard Operating Procedure (SOP) SOP-LAB-EQP-005, Bruker EM640S Portable GC/MS with AMDIS Data Analysis Software (Blinded Mode) for Qualitative Analyses of Liquid Samples Prepared by On-site Methods', version 3, The Technical Secretariat of the Organization for the Prohibition of Chemical Weapons, The Hague, 1999.

Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Robin M. Black and Robert W. Read
Defence Evaluation and Research Agency, Salisbury, UK

1 Introduction **1**
2 History **2**
2.1 Liquid Chromatography/Mass Spectrometry Interfaces/ Ionization Methods 2
2.2 Liquid Chromatography/Mass Spectrometry Analysis of Chemical Warfare Agents 3
2.3 Degradation Products of Chemical Warfare Agents 4
2.4 Analysis of Toxins of Chemical Warfare Interest 4
3 Sample Preparation **4**
3.1 Water Samples 4
3.2 Soil Samples 5
3.3 Biomedical Samples 5
4 Environmental Applications using Liquid Chromatography/Mass Spectrometry and Liquid Chromatography/Tandem Mass Spectrometry **5**
4.1 Analysis of Chemical Warfare Agents 6
4.2 Analysis of Chemical Warfare Agent Degradation Products 7
4.3 Dialkyl Alkylphosphonates 11
4.4 Toxins 12
5 Biomedical Applications **12**
5.1 Biological Markers of Sulfur Mustard Poisoning 13
6 Method Development **14**
6.1 Liquid Chromatography Aspects 14
6.2 Mass Spectrometry Aspects 14
6.3 Quality Control 15
7 Comparison with Other Hyphenated Techniques **15**
7.1 Gas Chromatography/Mass Spectrometry 15
7.2 Capillary Electrophoresis/Mass Spectrometry 16
Abbreviations and Acronyms **16**
Related Articles **17**
References **17**

Liquid chromatography/mass spectrometry (LC/MS) is the direct coupling of a chromatographic separation of analytes dissolved in a liquid phase with mass spectrometric detection/identification. Analytes are separated by one or more mechanisms, usually involving hydrophobic, polar or ionic interactions. The components eluting from the LC are introduced directly into the mass spectrometer via an interface, which removes the solvent and ionizes the analyte molecules to produce positive or negative ions. These ions are separated by the mass spectrometer on the basis of their mass-to-charge (m/z) ratios. The recent introduction of versatile and robust interfaces, which use atmospheric pressure ionization (API) (atmospheric pressure chemical ionization (APCI) and electrospray (ES) ionization), has lead to increasing application of LC/MS to the analysis of chemical warfare (CW)-related analytes.

Although LC/MS can be used for the direct analysis of CW agents in aqueous samples, its major application in this field is for the analysis of hydrolysis products of CW agents and other polar degradation products. Reversed phase LC/MS provides a rapid screening procedure for hydrolysis products such as phosphonic acids and thiodiglycol, avoiding the tedious isolation and derivatization required for gas chromatography/mass spectrometry (GC/MS). Increased structural information for unequivocal identification, or improved limits of detection in trace analysis, can be obtained using liquid chromatography/tandem mass spectrometry (LC/MS/MS). LC/MS/MS is also useful for detecting biological markers of poisoning by CW agents in cases where GC/MS is unsuitable.

1 INTRODUCTION

LC/MS is the direct on-line coupling of a chromatographic separation in the liquid phase with mass spectrometric detection/identification. Analytes in a solvent system (the mobile phase) are passed through a chromatography column, packed usually with a solid support coated with a bonded liquid stationary phase. The components are separated by one or more mechanisms, such as hydrophobic, polar or ionic interactions, depending on the nature of the analyte, and mobile and stationary phases. The components eluting from the LC are introduced directly into the mass spectrometer via

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

an interface, which removes the liquid solvent and ionizes the analyte molecules to produce positive or negative ions. These ions are separated by the mass spectrometer on the basis of their mass to charge (m/z) ratios. A key component of an LC/MS system is the interface between the chromatograph and mass spectrometer, the conditions under which the two operate being mutually incompatible. LC is performed using a large mass flow at or above atmospheric pressure at flow rates in the range 0.1–2 mL min^{-1}, commonly using solvent mixtures such as water–methanol or water–acetonitrile. Mass spectrometric separation of ions occurs under high vacuum. Although LC separation can be performed using normal or reversed-phase mechanisms, as a generalization, those analytes of CW interest that might be separated using normal phase LC are usually analyzed by gas chromatographic methods. LC/MS analysis is most commonly applied to the more polar analytes using reversed-phase conditions.

2 HISTORY

2.1 Liquid Chromatography/Mass Spectrometry Interfaces/Ionization Methods

The four LC/MS interface/ionization systems that have been most widely investigated for the analysis of low molecular mass analytes, such as pesticides and drugs, are thermospray (TS), particle beam (PB) with electron ionization (EI) or chemical ionization (CI), APCI and the related techniques ES and ionspray (IS).[1–4] A detailed description of these interfaces is given by Niessen and Tinke.[1]

In the TS interface, nebulization and near complete desolvation occur by transport of the LC eluent through a heated capillary. Ionization occurs by solvent-mediated CI and evaporation processes, the reagent ions being generated from added volatile buffers, such as ammonium acetate, and/or from an auxiliary ionization source such as a filament or discharge electrode. Liquid chromatography/thermospray mass spectrometry (LC/TSMS) operates at typical flow rates of 0.5–1.5 mL min^{-1}. TS is well suited to most low-molecular-weight polar compounds but detection limits can be quite variable; thermally labile compounds may be degraded in the hot capillary and ion currents may fluctuate. Occasional blockage of the TS probe tip by involatile salts may cause operational problems. Because of its commercial availability since the early 1980s, LC/TSMS has been widely used in pesticide analysis,[2–4] and it was the first LC/MS technique to be applied to the analysis of CW agents[5] and their degradation products.[6]

In the PB interface, an almost complete removal of mobile phase constituents occurs by pneumatic or thermal nebulization into a desolvation chamber. A beam of dry particles is then introduced via a momentum separator into the ionization source. PB using EI has the advantage that mass spectra resembling reference EI spectra (as obtained using GC/MS) are produced, but sensitivity for polar analytes is generally lower than with the other techniques and there are limitations to the percentage of aqueous phase that can be accommodated at normal LC flow rates. PB is not the best technique for water-soluble polar compounds.[7] It has been used rarely in CW analysis although a modification, the eluent jet interface, has been applied to CW-related analytes.[8]

The API techniques APCI, ES and IS have become accepted as the most robust, sensitive and versatile for the LC/MS analysis of low- and high-molecular-mass polar compounds;[1] they are being increasingly applied to the analysis of polar analytes such as pesticides, drug metabolites, toxins and CW agent degradation products. As the name suggests, these techniques rely on the formation of ions at atmospheric pressure in a source region separated from the high-vacuum mass analyzer, providing practical benefits for coupling to the MS. Both techniques involve nebulization of the liquid eluent from the LC column. With APCI, the LC eluent is introduced into the ion source region via a heated nebulizer assisted by an additional gas flow; ionization is induced usually from a corona discharge and relies on charge transfer in the gas phase, resembling normal CI. ES/IS rely more on liquid-phase chemistry in which an aerosol of highly charged droplets is produced by nebulization in an electric field, followed by desorption of solvated analyte ions from the droplets after vaporization of the solvent. With the IS modification, ES nebulization is combined with pneumatic nebulization, allowing larger LC flow rates to be accommodated. Ionized analyte is transferred from atmospheric pressure to vacuum through a heated capillary and/or skimmer. The literature suggests that APCI is best suited to low-molecular-mass nonionic analytes and ES/IS to ionic and high-molecular-mass materials. APCI can accommodate normal flow rates (up to $\sim$2 mL min^{-1}) with conventional bore reversed-phase LC systems; modern ES interfaces are usually operated at lower flow rates, $<$10 μL min^{-1}, and IS at 0.05–0.2 mL min^{-1}. Thermal degradation in the heated nebulizer appears to be much less of a problem with APCI than it is with TS.

TS, APCI and ES/IS are soft ionization techniques, i.e. they impart little excess energy to the molecule in the process of ionization. Under conditions providing optimum sensitivity they usually produce abundant protonated molecules or other adduct ions such as $[M + NH_4]^+$ or $[M + Na]^+$ (quasimolecular ions) with little or no fragmentation. They are therefore good techniques for selective detection and screening, but may

not provide the structural information or fingerprint required for unequivocal identification. One solution is to vary the source conditions to encourage preanalyzer collision-induced dissociation (CID). With TS, greater fragmentation of the molecular ion adducts can be induced by putting more energy into the ions using a discharge ionization or ''filament on'' technique, and/or by varying the repeller voltages to prolong the residence time of the ions within the source region. With APCI and ES, the ion–molecule collisions can be enhanced by applying a small potential difference between the spray nozzle and skimmer or between the skimmer and an octapole ion guide. With preanalyzer CID, all ions within the source region may be subject to fragmentation. Alternatively, and for optimum signal-to-noise ratio (S/N), CID is obtained using conventional (but instrumentally more expensive) tandem mass spectrometry (MS/MS). In addition to providing structural information, LC/MS/MS is advantageous for improving limit of detections (LODs) in trace analysis (as is GC/MS/MS). Selected reaction monitoring (SRM) of a structurally specific fragmentation is much more selective (in relation to chemical background) than simply monitoring the mass-to-charge ratio for a quasimolecular ion.

The combinations of API sources with benchtop quadrupole mass spectrometers, and more recently with quadrupole ion traps, are now offering LC/MS and LC/MS/MS systems at moderate prices. LC–ion trap instruments offer relatively simple robust LC/$(MS)^n$ capabilities and they have the added advantage that full scan MS/MS can be performed without sacrificing sensitivity in comparison with an SRM experiment.[9] Advances in API sources have been reviewed by Niessen.[10]

2.2 Liquid Chromatography/Mass Spectrometry Analysis of Chemical Warfare Agents

The major CW agents, included in Schedule 1 of the Annex on Chemicals in the text of the Chemical Weapons Convention (CWC),[11] are the nerve agents, e.g. sarin (GB) **(1)** [107-44-8], soman (GD) **(2)** [96-64-0], cyclohexyl sarin (GF) **(3)** [329-99-7], *O*-ethyl *S*-2-diisopropylaminoethyl methylphosphonothiolate (VX) **(4)** [50782-69-9] and tabun (GA) **(5)** [77-81-6], the vesicants, e.g. sulfur mustard **(6)** [505-60-2], nitrogen mustards **(7)** [51-75-2] and **(8)** [555-7-1], and Lewisite-1 **(9)** [541-25-4]; the incapacitant 3-quinuclidinyl benzilate (BZ) **(10)** [6581-06-2] is included in Schedule 2. Saxitoxin **(11)** [35523-89-8] and ricin [9009-86-3] are the only examples of toxins included in Schedule 1.

In addition to the established CW agents, Schedule 1 of the CWC includes some analogs of sulfur mustard and defines generic structures **(12–14)** that encompass

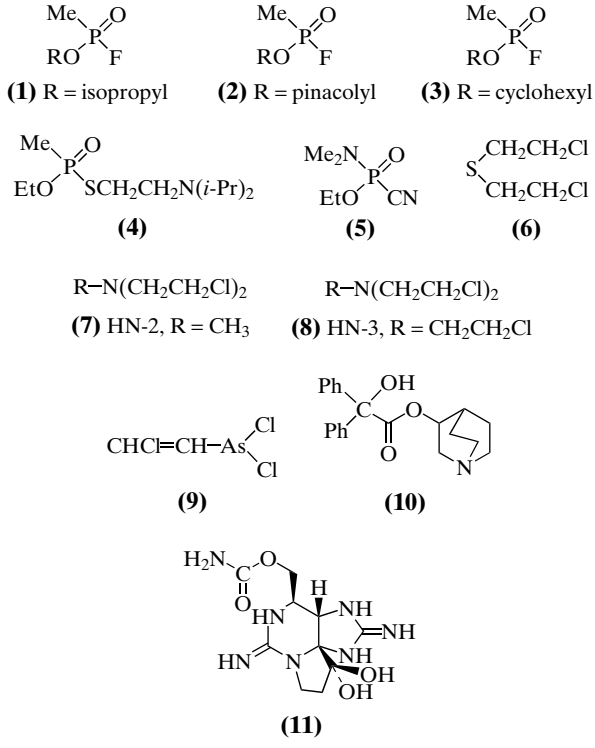

(1) R = isopropyl **(2)** R = pinacolyl **(3)** R = cyclohexyl

(4) **(5)** **(6)**

(7) HN-2, R = CH_3 **(8)** HN-3, R = CH_2CH_2Cl

(9) **(10)**

(11)

thousands of possible nerve agent analogs, where the alkyl substituents on phosphorus or nitrogen (R^1, R^3) can be C_1–C_3, and on oxygen (R^2) C_1–C_{10}.

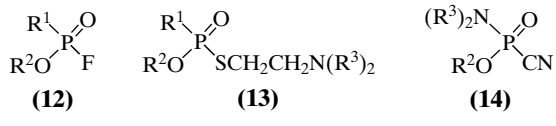

(12) **(13)** **(14)**

The nerve agents, mustards and BZ are sufficiently volatile for analysis by GC/MS, although S/Ns with VX and BZ can deteriorate owing to thermal degradation, particularly with dirty injection ports. Capillary GC/MS methodology for CW agents has been well developed since the late 1970s[12] and is usually the method of choice for these compounds. LC/MS analysis of intact CW agents was first reported in 1988[5] but has generally been used only when the rapid analysis of aqueous samples or extracts is required. An additional factor is that many CW agents react to varying degrees with nucleophilic solvents such as water and methanol, and are usually isolated from aqueous media prior to analysis. The half-lives for GB, GD and VX in water, at typical ambient temperatures and neutral pH, range from 2–3 days to greater than 1 week;[13] sulfur mustard is much less stable, with a half-life of a few minutes, and reversed-phase LC/MS is inappropriate.

As described in section 4.1, LC/MS has been reported for the direct analysis of aqueous solutions of the established nerve agents in water.[14] Nitrogen mustards are

hydrolyzed more slowly than sulfur mustard and analysis by reversed-phase LC/MS has been demonstrated.[14] An example of the use of LC/MS for the identification of BZ in soil is reported in section 4.1.

2.3 Degradation Products of Chemical Warfare Agents

The most important application of LC/MS in CW analysis is for the hydrolysis products of CW agents and other polar degradation products e.g. derived from oxidative and/or hydrolytic decontamination reactions. Polar degradation products are formed in environments such as soil[13] and in biological systems. Important examples are alkyl alkylphosphonic acids such as **(15–18)** and methylphosphonic acid (MPA) **(19)** [993-13-5] from the hydrolysis of nerve agents, thiodiglycol **(20)** [11-48-8] and thiodiglycol sulfoxide **(21)** [3085-45-8] derived from sulfur mustard, ethanolamines such as triethanolamine **(22)** [102-71-6] from nitrogen mustards, and benzilic acid **(23)** [76-93-7] and 3-quinuclidinol **(24)** [1619-34-7] from the hydrolysis of BZ.

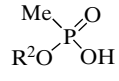
(15) R^2 = Et

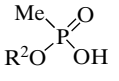
(16) R^2 = *i*-Pr

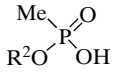
(17) R^2 = pinacolyl

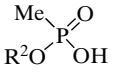
(18) R^2 = cyclohexyl

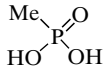
(19)

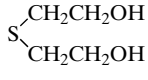
(20)

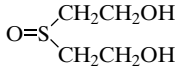
(21)

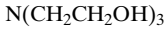
(22)

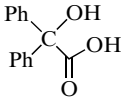
(23)

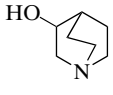
(24)

These degradation products are considerably more polar and less volatile than their parent CW agents. For GC/MS analysis they are usually isolated from aqueous media and derivatized to improve their gas chromatographic properties. In most cases they are efficiently extracted from environmental residues with water following a Recommended Operating Procedure (ROP).[15] Reversed-phase LC/MS allows the direct analysis of these aqueous extracts or water samples with little or no sample preparation, depending on the concentration of analytes and extraneous materials present. LC/MS methods have been reported for the analysis of all of the above hydrolysis products, plus some biological markers of sulfur mustard poisoning, using TS, ES or APCI. As described in section 1, API interfaces have distinct advantages over TS and are superseding it as new instruments and accessories are purchased. LC/MS is being included increasingly in strategies for CW analysis as its advantages as a rapid screening technique are demonstrated in proficiency tests and as instrument costs fall. Degradation products of the arsenical Lewisite are a special case and the only method based on LC/MS to have been reported uses inductively coupled plasma mass spectrometry (ICPMS) for detection.[16]

2.4 Analysis of Toxins of Chemical Warfare Interest

Schedule 1 of the CWC includes only two toxins. These are saxitoxin **(11)**, which has a relatively low molecular mass of 299 (in its nonprotonated form), and ricin which is a complex glycoprotein with a molecular mass of around 62 000 Da. Saxitoxin is of greater concern as a health hazard in seafood (paralytic shellfish poisoning) rather than as a CW agent. It is produced by cold-water dinoflagellates such as *Gonyaulax* spp. and accumulates in shellfish which feed on these microorganisms. Saxitoxin is a very polar toxin which is protonated at neutral pH and cannot be analyzed by GC/MS. Physicochemical methods for the analysis of saxitoxin have been based mainly on LC or capillary electrophoresis (CE), and methods based on LC/ES/ISMS have appeared.[17] LC/MS has not been reported for ricin although preliminary studies of its ES spectra are being undertaken. LC/ESMS has been reported for another proteinaceous toxin of possible CW interest, staphylococcal enterotoxin B (SEB).[18]

3 SAMPLE PREPARATION

3.1 Water Samples

A major advantage of LC/MS is that aqueous samples can be analyzed with no or minimum sample preparation (other than filtration if necessary), provided that the analyte is present at sufficient concentration. The need for concentration or cleanup of the sample depends on the LOD required and on the presence of any interferents. With concentrations of analyte > ca. 100 ng mL^{-1} direct analysis using selected ion monitoring (SIM) without pretreatment is possible for most analytes and using full mass spectral scanning in favorable cases. In the case of the very polar analytes MPA and ethylphosphonic acid (EPA), which are poorly retained on reversed-phase columns and tend to co-elute with inorganic salts, it may be advantageous to pass the water through a cation-exchange resin, particularly if high concentrations

of calcium or magnesium ions are present (these bind strongly to these acids). However, this is not recommended as part of the initial screening procedure because other analytes of interest, such as ethanolamines, may be removed.

Simple concentration by evaporation can be used to increase the S/N for the analyte if chemical background is not the limiting factor. Lower LODs can in principle be achieved using various off-line or on-line preconcentration procedures like those used in pesticide analysis.[19–21] The majority of the CW analytes of interest with respect to LC/MS are difficult to extract using liquid–liquid extraction. The general trend in the 1990s has been towards the more efficient solid-phase extraction (SPE) on disposable cartridges. SPE cartridges can be used to extract water sample volumes up to about 1 L and the analytes eluted with a small volume of organic solvent (e.g. methanol, acetonitrile). Wils and Hulst[5] concentrated VX from 50 mL volumes of river water prior to LC/MS, using a C_{18} SPE cartridge, with subsequent elution with methanol. GB and GD can be similarly concentrated from water samples using C_{18} cartridges.

For polar analytes with a degree of hydrophobicity, such as alkyl alkylphosphonic acids, C_{18}, C_8 or C_2 SPE cartridges can be used for extraction provided that the pH is adjusted to around 1 to ensure that the acids are predominantly nonionized.[22] An alternative is to add a cationic ion-pairing agent such as a tetrabutylammonium salt to enhance extraction onto the lipophilic C_{18} phase.[6] Anion exchange provides a second mechanism for concentrating the acids from aqueous extracts and matrices such as urine and serum,[23,24] although variable success has been reported.

The most difficult analytes to concentrate from aqueous solution are the small very polar molecules such as MPA, thiodiglycol sulfoxide and to a lesser degree thiodiglycol. In the analysis of polar pesticides in water there has been increasing use of polymeric materials such as polystyrene divinylbenzene and porous graphitic carbon.[20] These are more retentive for the more polar analytes; polymeric cartridges are showing some promise for the extraction of thiodiglycol and alkyl alkylphosphonic acids. The most recent technique being applied to the preconcentration of pesticides and CW-related analytes is solid-phase microextraction (SPME).[25–27] Analytes are partitioned into an immobilized polymeric stationary phase coated onto fused silica fibers. Small volumes (e.g. 10–15 mL) of water sample are required and desorption from the fibers can be achieved directly into a gas chromatograph by thermal desorption or a liquid chromatograph by solvent desorption, using modified inlet systems.

3.2 Soil Samples

The ROP for the analysis of soil samples for CW agents and their degradation products involves extraction with dichloromethane, water and methanol–1% triethylamine.[15] Any further treatment of the aqueous extract depends on factors similar to those discussed above for water samples. The presence of high concentrations of inorganic cations (see above) is the most common problem.

3.3 Biomedical Samples

Biomedical samples are generally more complex than environmental ones, and the LODs required to detect biological markers of poisoning (e.g. urinary metabolites) are generally in the region of 1 ppb or lower. LC/MS analysis has rarely been used for the trace level detection of metabolites of CW agents in urine, primarily because lower LODs can be achieved in most cases using GC/MS. The cleanup required depends very much on the particular matrix and the pK_a of the analyte. Most of the methods described above for the concentration of phosphonic acids in water have been applied to urine prior to GC/MS analysis.[22,23] Blood is a more difficult matrix and the plasma fraction is usually separated prior to concentration. For the LC/MS analysis of 2-hydroxyethylthioethylhistidine in hemoglobin (see section 5.1.2), a biological marker of sulfur mustard poisoning, the proteinaceous fraction from hemoglobin was digested with hydrochloric acid and the liberated alkylated amino acid concentrated on a cation exchange resin prior to formation of the fluorenylmethoxycarbonyl (Fmoc) derivative.[28]

4 ENVIRONMENTAL APPLICATIONS USING LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY AND LIQUID CHROMATOGRAPHY/TANDEM MASS SPECTROMETRY

The need to analyze environmental samples for CW-related chemicals may arise from allegations of CW use, inspections by the Organization for the Prohibition of Chemical Weapons (OPCW) of suspect CW facilities, activities such as the United Nations Special Commission (UNSCOM) missions to Iraq and demilitarization or decontamination activities. ROPs for environmental analysis have been compiled and revised by technical experts following a series of interlaboratory comparison exercises (round robins).[15] The latter have now been replaced by official proficiency tests organized by the OPCW for those laboratories who wish to become designated for investigations conducted by the OPCW. To

date, all OPCW proficiency tests have involved environmental samples. Many of the more recent developments in analytical methodology have arisen during the course of these round robins and proficiency tests. The focus has been on unequivocal identification of a wide range of scheduled compounds and their degradation products at levels typically around 10 ppm, rather than on the trace analysis of selected analytes at parts per billion levels. Typical environmental samples include soil, water, paint, rubber, plastic, concrete and residues from decontamination processes. Water, soil and decontamination residues are the most likely to contain degradation products and these have been the samples to which LC/MS analysis has most often been applied. The majority of LC/MS applications have employed quadrupole mass spectrometers although the use of ion trap systems is likely to increase.

4.1 Analysis of Chemical Warfare Agents

4.1.1 Liquid Chromatography/Thermospray Mass Spectrometry

LC/TSMS was reported by Wils and Hulst for the analysis of VX,[5] GB, GD and GA[14] in samples such as river water and aqueous soil extracts. The analytes were chromatographed on a C_{18} column using a mobile phase of methanol or acetonitrile–0.1 M aqueous ammonium acetate. VX gave a good peak shape when chromatographed at a pH close to its pK_a value (8.5), using a high percentage of organic modifier to obtain an acceptable retention time (ca. 6 min). A problem sometimes encountered when strong solvents are used to elute the more hydrophobic materials under full scanning conditions is co-eluting chemical background. The common plasticizer tri-*n*-butyl phosphate co-eluted with VX using methanol–0.1 M ammonium acetate (80 : 20) but resolution was achieved using acetonitrile–methanol–0.25 M ammonium acetate (70 : 20 : 10). Because of the range of polarities, GD being considerably more hydrophobic than GB and GA, gradient elution is required to chromatograph the range of nerve agents in a single run.

The TS spectra are dominated by abundant quasi-molecular ions, $[M + H]^+$ or $[M + NH_4]^+$, depending on whether the proton affinity is less than or greater than that of ammonia. VX has a strongly basic side chain and the TS mass spectrum exhibits an abundant protonated molecular ion; GB, GA and GD possess lower proton affinities than ammonia and their spectra exhibit strong ammonium adduct ions in addition to protonated molecules. The degree of fragmentation observed depends on the mobile phase, the discharge ionization (if any) and repeller voltages. Using a mobile phase of acetonitrile–methanol–0.089 M ammonium acetate (50 : 14 : 46), the spectrum of VX exhibited an abundant $[M + H]^+$ at *m*/*z* 268 with very weak fragment ions at *m*/*z* 188 and 146/145. The latter were assigned structures $[AcOCH_2CH_2N(i\text{-}Pr)_2 + H]^+$, $[HOCH_2CH_2N(i\text{-}Pr)_2 + H]^+$ and $[H_2NCH_2CH_2N(i\text{-}Pr)_2 + H]^+$, presumably formed by Michael-type additions of solvent/buffer molecules to the fragment $[CH_2{=}CHN(i\text{-}Pr)_2 + H]^+$. Increased fragmentation was observed in a mobile phase of 0.1 M ammonium acetate. It is therefore important that similar conditions are used when comparing spectra for confirmatory purposes.

The methodology successfully identified VX in an aqueous extract of soil spiked at 100 ppm as part of a round robin exercise. Preconcentration using a C_{18} SPE cartridge and SIM enable lower concentrations of VX to be detected. A S/N of 10 was observed using SIM of the protonated molecule after forcing 50 mL of river water spiked at 0.1 ng mL^{-1} (0.1 ppb) through a C_{18} cartridge and eluting with 1 mL of methanol.

Wils and Hulst[14] demonstrated that nitrogen mustard, HN-2 could be analyzed in water using LC/TSMS. However, the mustard was converted to a diacetoxy derivative by reaction with ammonium acetate in the heated interface. A procedure was developed for the determination of HN-2 in the atmosphere after sampling the air onto a silica tube (LOD 1 μg m^{-3}, sampling 10 L of air).

4.1.2 Liquid Chromatography/Atmospheric Pressure Chemical Ionization/Mass Spectrometry and Liquid Chromatography/Electrospray/Mass Spectrometry

Few applications of LC/APCI/ESMS have yet been reported for the analysis of intact CW agents, but both techniques should be suitable for the applications described above. APCI and ES can be useful for confirming the molecular mass of V-type agents in proficiency tests as an alternative to GC/CIMS (GC/EIMS gives weak or nonexistent molecular ions, the spectra being dominated by the fragment $[CH_2 = NR_2]^+$ derived from the side chain). Creasy et al.[29] reported the application of LC/APCIMS/MS with SRM to the simultaneous detection of VX and its hydrolysis product EA 2192, **(25)** (Scheme 1) [73207-98-4] in decontamination solutions. The latter is one of several hydrolysis products of VX but is of particular importance because it retains a high degree of toxicity. The two compounds were separated on a C_{18} column (octadecylsilane (ODS) Hypersil) using gradient elution with an acetonitrile–0.05 M aqueous ammonium acetate solvent system. The fragmentations *m*/*z* 268 ⟶ 128 and *m*/*z* 240 ⟶ 128 were monitored for VX and EA 2192, respectively (Scheme 1).

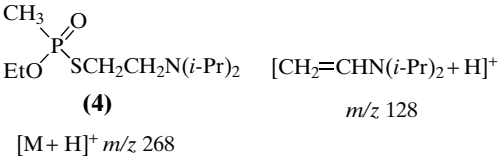

(4)

$[M+H]^+$ *m/z* 268

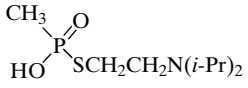

(25)

$[M+H]^+$ *m/z* 240

Scheme 1

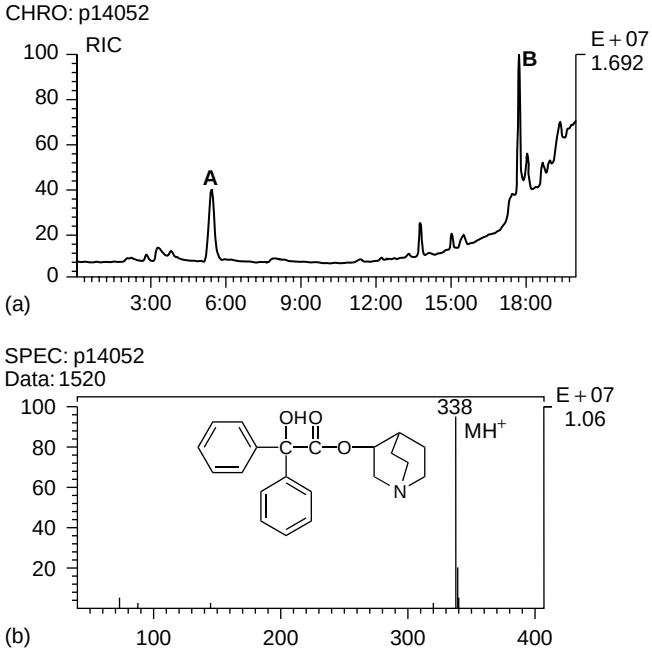

Figure 1 Liquid chromatography/atmospheric pressure chemical ionization/mass spectrometry (LC/APCIMS) total ion chromatogram showing the detection of **A** thiodiglycol and **B** BZ in a water extract of a spiked soil sample (a), and APCI mass spectrum of BZ (b).

In a recent OPCW proficiency test, APCI proved useful in detecting and identifying BZ spiked into a soil sample at 10 ppm. Figure 1 shows the chromatogram of the water extract of the soil showing the simultaneous detection of the spiked compounds BZ and thiodiglycol.

4.2 Analysis of Chemical Warfare Agent Degradation Products

4.2.1 Liquid Chromatography/Thermospray Mass Spectrometry

LC/TSMS in CW analysis was first applied to the degradation products of nerve agents and to thiodiglycol.[6,14] Thirteen organophosphorus acids consisting of phosphonic acids **(26)** (hydrolysis products of G and V agents), a thiophosphonic acid **(27)**, plus some phosphoric and thiophosphoric acids **(28)** (hydrolysis products of pesticides) were resolved or partially resolved on a C_{18} column. The most polar acids could be chromatographed satisfactorily with 0.1 M aqueous ammonium acetate alone, but the more lipophilic acids (e.g. cyclohexyl methylphosphonic acid (cHexMPA) and pinacolyl methylphosphonic acid (PinMPA) [1932-60-1, 616-52-4]) required methanol–0.1 M aqueous ammonium acetate (30 : 70). Resolution of thiophosphonic acids from phosphonic acids was improved by lowering the pH of the eluting solvent (from pH 6.8 to pH 5.0), exploiting the lower pK_a values of the former. The use of tetra-*n*-butylammonium hydroxide as ion-pairing reagent was explored to improve the capacity factors which were poor for the more polar acids, but problems were experienced with column degradation, blockage of the TS interface and memory effects.

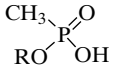

(26) R = H, Et, *i*-Pr, *c*-Pent, *c*-Hex, Pinacolyl

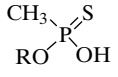

(27) R = Et

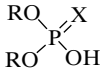

(28) R = Me, Et, *i*-Pr, X = O, S

In positive-ion mode both protonated and ammonium adduct ions are observed in TS spectra, ammonium adducts being the more abundant. The ratio of the quasimolecular ions varies, protonated molecules being weaker with the more polar acids; an increase in vaporizer temperature reduces the $[M+NH_4]^+/[M+H]^+$ ratio. PinMPA shows additional ions corresponding to $[CH_3P(O)(OH)_2+H]^+$ at *m/z* 97 and $[CH_3P(O)(OH)_2+NH_4]^+$ at *m/z* 114; these increase with source temperature. Negative ion spectra are much simpler giving predominantly $[M-H]^-$ ions. LODs were reported in the range 100 pg to 1 ng injected, S/Ns being larger for the acids with smaller alkyl groups R. With the more hydrophobic acids, such as PinMPA and cHexMPA, peak broadening occurs at the longer retention times and the increased methanol content of the eluent reduces the efficiency of the ionization. Water spiked with PinMPA at 50 ng mL^{-1} was at the LOD for this analyte, concentration being required to detect lower concentrations.

Similar methodology successfully detected PinMPA and the isomeric diisopropyl methylphosphonate in a soil sample analyzed in a round robin exercise.[30] The compounds could be distinguished by retention time and by the $[M+NH_4]^+/[M+H]^+$ ratio, the ammonium adduct being the more intense in the acid. LC/TSMS was among a battery of chromatographic and spectroscopic techniques used to analyze environmental samples collected during the UNSCOM 65 mission to southeast Iraq.[31] No

phosphonic acids were detected but diethyl phosphoric acid and tentatively ethyl phosphoric acid were identified at concentrations in the range 0.1–2 ppm in some water samples. Although these acids could be derived from the hydrolysis of GA, they are also the hydrolysis products of several pesticides.

Tørnes[32] reported the application of LC/TSMS/MS (using a hybrid electrostatic sector magnetic sector quadrupole quadrupole (EBQQ) system) to the analysis of alkyl MPAs **(26)** (R = H, Et, *n*-Pr, *i*-Pr, pinacolyl). CID of the protonated molecules gave a major product ion at *m/z* 97, $[MeP(O)(OH)_2 + H]^+$, formed from C–O bond cleavage with loss of alkene, plus *m/z* 79, $[MeP(O)OH]^+$, from further loss of H_2O. Fragment ions originating from the *O*-alkyl groups were also observed, e.g. *m/z* 85, 57 and 43 from pinacolyl, *m/z* 43 and 27 from propyl, providing useful additional structural information. *n*-Propyl and isopropyl MPAs gave almost identical MS/MS spectra and could only be reliably distinguished by their different retention times. As a general rule, isomers with *n*-alkyl groups are retained more than those with branched chain alkyl groups by the lipophilic C_{18} stationary phase. Mass chromatograms derived from full scan daughter ion spectra showed a 20–50 times improvement in S/N compared to full scanning LC/TSMS; detection limits were in the region of 0.1 μg mL^{-1} using this mode of operation.

Thiodiglycol, thiodiglycol sulfoxide and thiodiglycol sulfone have been analyzed using LC/TSMS.[14,30,33] Fragmentation pathways are discussed by Munavalli et al.[33] LODs of ~50 ng mL^{-1} were quoted for standards in water, using SIM of the $[M + H]^+$ and $[M + NH_4]^+$ ions. The importance of the sulfoxide in the environment was demonstrated by the apparent complete oxidation of thiodiglycol spiked into a soil sample in a round robin exercise.[30]

4.2.2 *Liquid Chromatography/Atmospheric Pressure Chemical Ionization/Mass Spectrometry and Liquid Chromatography/Electrospray/Mass Spectrometry*

LC/APCIMS and LC/ESMS methods are now superseding those based on LC/TSMS. Black and Read[34] developed a qualitative screening procedure for the simultaneous detection of 19 hydrolysis products derived from nerve agents, sulfur and nitrogen mustards, and BZ, using APCI on a triple stage quadrupole LC/MS/MS instrument. Because of the range of physicochemical properties, i.e. acidic, neutral and basic analytes, varying from very polar (MPA and thiodiglycol sulfoxide) to moderately hydrophobic (PinMPA, benzilic acid), the choice of LC conditions was of necessity a compromise. All 19 analytes chromatographed satisfactorily on a mixed C_8/C_{18} reversed-phase column, using a

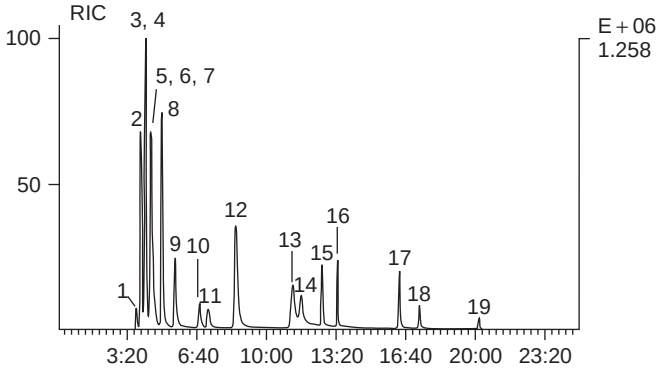

Figure 2 LC/APCIMS total ion chromatogram of a standard mixture of degradation products (10 μg mL^{-1} each, 5 μL injected): 1, MPA; 2, thiodiglycol sulfoxide; 3, triethanolamine; 4, *N*-methyldiethanolamine; 5, EPA; 6, *N*-ethyldiethanolamine; 7, thiodiglycol sulfone; 8, 3-quinuclidinol; 9, ethyl methylphosphonic acid (EMPA); 10, thiodiglycol; 11, *n*-propylphosphonic acid; 12, diisopropylaminoethanol; 13, ethyl EPA; 14, isopropyl MPA; 15, *t*-butylphosphonic acid; 16, *n*-butylphosphonic acid; 17, cHexMPA; 18, PinMPA; 19, benzilic acid. (Reproduced by permission of Elsevier Science from R.M. Black, R.W. Read, *J. Chromatogr., A*, **759**, 79–92 (1997).[34])

water–acetonitrile–0.05% trifluoroacetic acid (TFA) mobile phase and gradient elution. Full or close to baseline resolution was achieved for 14 of the analytes; triethanolamine and *N*-methyldiethanolamine co-eluted, as did EPA, *N*-ethyldiethanolamine and thiodiglycol sulfone. However, these can readily be distinguished using mass chromatograms or SIM. A total ion current chromatogram is shown in Figure 2.

Simple alkylphosphonic acids $R^1P(O)(OH)_2$ give abundant $[M + H]^+$ ions plus adduct ions $[M + CH_3CN + H]^+$. The alkyl alkylphosphonic acids $R^1P(O)(OR^2)OH$ with smaller R^2 groups give $[M + H]^+$ ions as the base peak plus a strong fragment ion corresponding to $[R^1P(O)(OH)_2 + H]^+$; concentration-dependent dimeric ions $[2M + H]^+$ are also observed. The $[M + H]^+$ ions are weaker with cHexMPA and PinMPA. The intensities of the $[R^1P(O)(OH)_2 + H]^+$ ions increase with the capillary temperature and Q0 offset (source octapole offset voltage). These ions can be utilized for the generic screening of phosphonic acids, by monitoring the selected ions *m/z* 97, 111, 125 and so on for alkyl MPA, EPA, and propylphosphonic acid and so on. Benzilic acid does not give a protonated molecular ion under the conditions used but an abundant ion at *m/z* 224, tentatively assigned to $[Ph_2C{=}OH + CH_3CN]^+$. Under the conditions employed, all but one of the neutral and basic analytes give APCI spectra containing abundant $[M + H]^+$ ions with few or no significant fragment ions. Thiodiglycol is the exception, giving an $[M + H - H_2O]^+$ ion as the base peak with a rather weak (~7% relative abundance) $[M + H]^+$. As is the case for TS and ES,

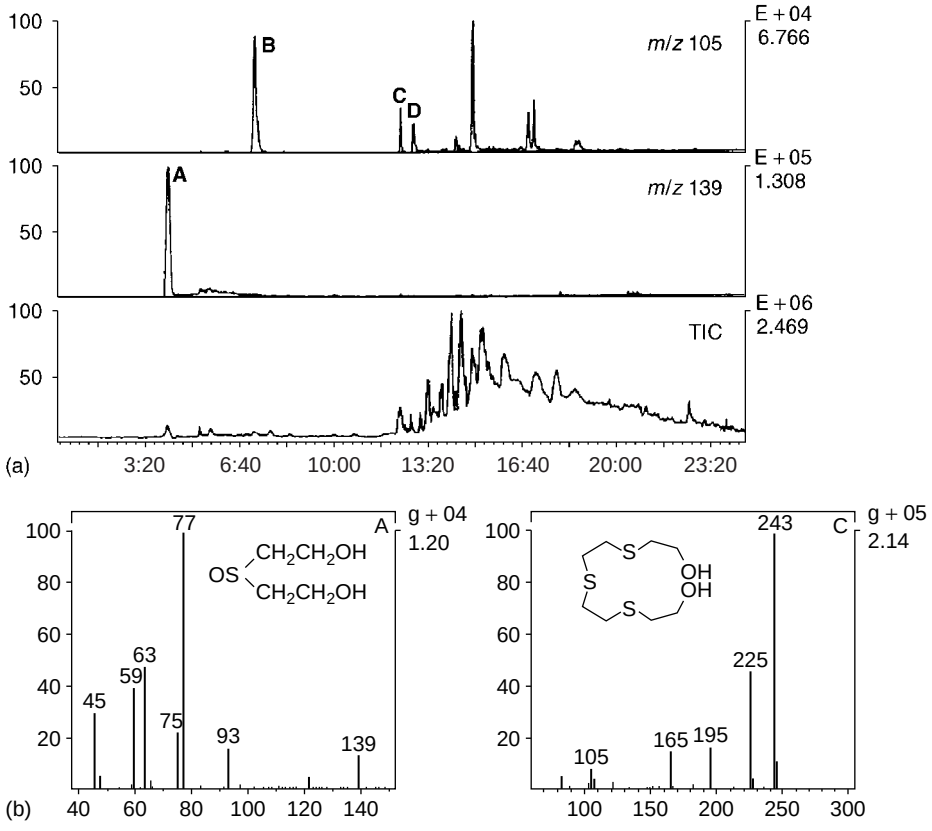

Figure 3 LC/APCIMS mass and total ion chromatograms showing the detection of **A** thiodiglycol sulfoxide, **B** thiodiglycol and **C**,**D** additional degradation products derived from sulfur mustard, in a water extract of a soil sample from a bomb crater (a), and CID spectra of thiodiglycol sulfoxide and component **C** (b). (Reproduced by permission of Elsevier Science from R.M. Black, R.W. Read, *J. Chromatogr., A*, **759**, 79–92 (1997).[34])

the relative intensities of adduct ions in the spectra are dependent on the mobile phase composition at the time of elution and spectra should be compared with standards run under the same conditions.

LC/MS/MS spectra are used for confirmation of identification. CID spectra obtained from the protonated molecules of phosphonic acids are similar to those obtained using LC/TSMS, although under the collision cell conditions used (triple quadrupole instrument, collision offset −25 V) either none or only weak alkyl fragments are observed with most alkyl alkylphosphonic acids. Hydrolysis products derived from mustards give very good fingerprints under CID conditions. The protonated molecule of 3-quinuclidinol gave very little fragmentation at an increased offset of −35 V, and LC/MS/MS is therefore not appropriate for confirmation.

For lower LODs, a selected ion screening procedure monitoring a total of 14 ions covers the entire range of 19 hydrolysis products. LODs for pure standards were <0.2 ng injected for the hydrolysis products of sulfur and nitrogen mustards, thiodiglycol sulfoxide, diisopropylaminoethanol and 3-quinuclidinol, and in the range 0.2–8 ng injected for a series of alkylphosphonic acids and alkyl alkylphosphonic acids, and for benzilic acid. The highest LODs were obtained for MPA and EPA. The methodology was applied to the analysis of aqueous extracts of soil samples collected from bomb craters associated with a CW attack in Iraq. These samples had previously been analyzed by GC/MS and found to contain sulfur mustard, thiodiglycol, isopropyl MPA, MPA and a trace of GB. By searching the mass chromatograms of *m/z* 105 and 139, derived from full scan LC/APCIMS data, three sulfur-containing compounds in addition to thiodiglycol were detected in an aqueous soil extract (Figure 3). These compounds were identified by LC/MS/MS as thiodiglycol sulfoxide and tentatively as the products $HOCH_2CH_2(SCH_2CH_2)_3OH$ **(29)** 1,11-dihydroxy-3,6,9-trithiaundecane ($C_8H_{18}O_2S_3$) and $HOCH_2CH_2SOCH_2CH_2SCH_2CH_2OH$ **(30)** 1,8-dihydroxy-3-oxo-3, 6-dithiaoctane ($C_6H_{14}O_3S_{12}$).

LC/ESMS has also been studied for this range of analytes.[35] Using similar LC conditions to those described above, ES was generally more sensitive than APCI for phosphonic acids by a factor of approximately

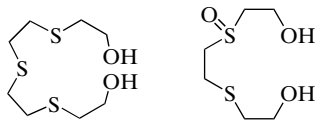

(29) (30)

two. A more substantial improvement in sensitivity, up to 10-fold, was obtained by substituting 0.1% formic acid for TFA as the acidic modifier following Zhou and Hamburger's[36] demonstration that TFA suppresses and formic acid enhances the formation of $[M + H]^+$ ions for a range of natural products using ES. Using SIM, LODs with 5 µL injections were $\leq 0.20\,ng\,mL^{-1}$, MPA and EPA being at the upper limit. Positive ES was less sensitive than APCI for thiodiglycol sulfoxide.

The ES spectra, like APCI, generally show abundant $[M + H]^+$ ions for the lower alkyl alkylphosphonic acids plus $[RP(O)(OH)_2 + H]^+$ fragment ions. In comparison to APCI spectra, a greater number of adduct ions are observed, particularly $[M + Na]^+$ and $[M + H + MeOH]^+$. Under the conditions used,[35] the fragment ions $[RP(O)(OH)_2 + H]^+$ were generally more intense than with APCI, probably due to the higher capillary temperature (200 °C) required to obtain robust ES in comparison to APCI (150 °C). The ready formation of adduct ions with metal ions can be a disadvantage of ES, because the ion current is shared over several quasimolecular ions.

Borrett et al.[37] reported a detailed study of metal adduct ions formed with phosphonic acids in positive and negative ion ES. These authors applied ES to the direct analysis of aqueous solutions (without LC).[38] The positive ion ES spectrum of a water solution containing MPA, thiodiglycol and diisopropyl methylphosphonate was complicated by the formation of protonated molecular ions and metal adducts; in contrast, the negative ES gave a much cleaner spectrum dominated by the deprotonated MPA molecule. Negative ion LC/ESMS and LC/APCIMS have been investigated in the authors' laboratory for the selective detection of phosphonic acids and benzilic acid. Using water–methanol–ammonium formate as mobile phase, total ion current traces are significantly cleaner than with positive ion ES and APCI. S/Ns are much improved for cHexMPA and PinMPA, which give very abundant $[M - H]^-$ ions in negative mode in contrast to the much weaker protonated molecules in positive mode. The positive and negative APCI spectra of cHexMPA using ammonium formate as modifier are shown in Figure 4.

The advantage of negative ion LC/ESMS for phosphonic acid analysis has been exploited by Siegenthaler.[39] Some phosphonic and thiophosphonic acids $R^1P(X)(OR^2)OH$ (X = O,S) were separated on a C_{18} column using gradient elution with methanol–0.02 M aqueous ammonium acetate buffer. To enhance the structural information, fragmentation was induced by in-source CID, by accelerating the ions in the octapole region of the analyzer, producing intense $[R^1P(X)(OH)O]^-$ product ions. Screening for these fragment ions using mass chromatography provided a selective screening procedure, enabling the acids to be detected against a heavy chemical background. The mass of the fragment ions in combination with $[M - H]^-$ tentatively identifies the R^1 and R^2 alkyl groups present, but does not distinguish isomeric alkyl groups. LODs under full scanning conditions were in the range $20–100\,ng\,mL^{-1}$, and $5–25\,ng\,mL^{-1}$ using SIM.

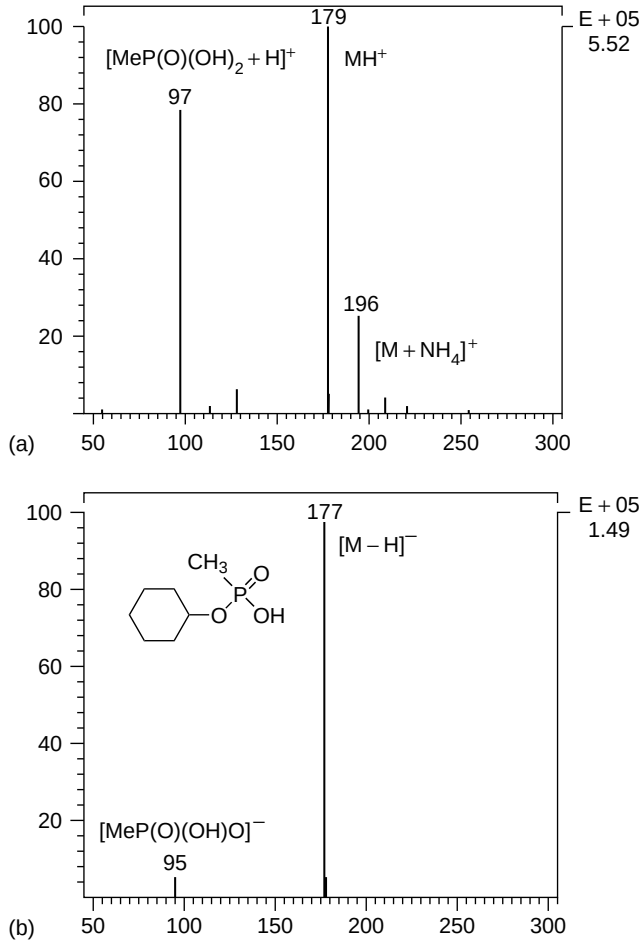

Figure 4 Positive (a) and negative (b) LC/APCIMS spectra of cHexMPA.

Rohrbaugh and Yang[40] used LC/ESMS to investigate the sulfonium ions formed on prolonged storage and hydrolysis of sulfur mustard. 1-(2-Chloroethyl)-1,4-dithianium ion was identified as the major component in solid mustard heels from large containers.

4.2.3 *Atmospheric Pressure Chemical Ionization Versus Electrospray*

Work in the authors' laboratory has been performed on a Finnigan TSQ 700 system and the relative merits of APCI and ES/IS may differ according to the source design. ES was initially regarded as an improvement over APCI for the analysis of phosphonic acids, but experience has shown more variability in comparison to APCI. Greater background has generally been observed using positive ion ES and ion intensities and S/Ns can be quite variable, presumably owing to differing degrees of adduct ion formation according to the background levels of metal ions in the sample. Negative ion APCI also affords cleaner traces than negative ES, with improved S/Ns and limits of detectability in comparison with negative ES. A negative APCI total ion chromatogram (using full scanning) derived from a solution containing a mixture of acidic, basic and neutral hydrolysis products at concentrations of $10\,\mu g\,mL^{-1}$, showing the selective detection of phosphonic acids and benzilic acid, is shown in Figure 5. Using SIM, all of the phosphonic acids were detected at $100\,ng\,mL^{-1}$ using $5\,\mu L$ injections except for MPA; the latter could be detected at this concentration by using $20\,\mu L$ injections. As found with negative ES, S/Ns were particularly good for cHexMPA and PinMPA.

The screening procedure currently employed at CBD uses both positive and negative APCI and a common mobile phase of methanol–water–0.02 M ammonium formate (under gradient conditions). The neutral and basic hydrolysis products are detected at levels similar to those described above in section 4.2.2, although the chromatographic resolution using ammonium formate modifier is not as good as that obtained with TFA. LC/APCIMS provides a rapid and robust screening procedure for aqueous samples and extracts, and has been particularly useful in proficiency tests where time is limited and spiking levels are around 10 ppm. Degradation products that have been identified in these tests by LC/APCIMS and LC/APCIMS/MS include MPA, cyclopentyl MPA [73207-99-5], cyclohexyl EPA [170424-87-0], thiodiglycol sulfoxide and *N*-ethyldiethanolamine [139-87-7].

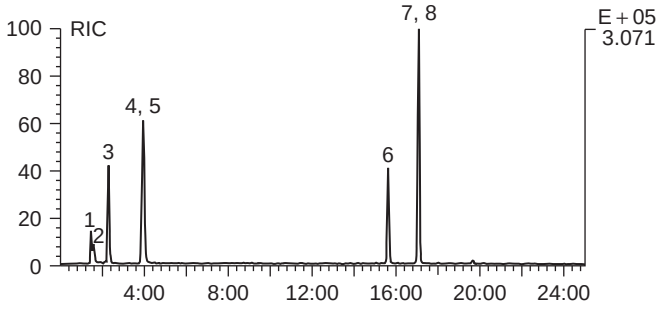

Figure 5 Negative LC/APCIMS total ion chromatogram showing the selective detection of acidic analytes in a standard mixture of acidic, neutral and basic degradation products: 1, MPA; 2, EPA; 3, EMPA; 4, isopropyl MPA; 5, ethyl EPA; 6, cHexMPA; 7, PinMPA; 8 benzilic acid.

4.2.4 *Liquid Chromatography/Particle Beam/Mass Spectrometry*

No applications of commercial LC/PB/MS systems to CW analysis have been reported but Kientz et al.[8] have described preliminary results using a related eluent jet interface for combination with micro-LC. The formation of the eluent jet is based on radiofrequency inductive heating. Using flow injection analysis, the EI spectrum of thiodiglycol was found to be comparable to reference EI spectra. The system provided EI-type spectra at levels in the range 1–10 ng injected and showed much better linearity of response in comparison to most PB systems. An improved version was reported for obtaining EI and CI spectra.[41] Good results were obtained for 2-diethylaminoethanol [100-37-8], a V agent precursor. Using SIM, 50 pg of 2-diethylaminoethanol could be detected in CI mode.

4.2.5 *Liquid Chromatography/Inductively Coupled Plasma Mass Spectrometry (for Lewisite Degradation Products)*

Lewisite is hydrolyzed to chlorovinyl arsenous acid [85090-33-1] in environments such as soil, and may be oxidized to chlorovinyl arsenic oxide [3088-37-7]. These analytes are difficult to analyze by LC/MS. LC/ICPMS enables the compounds to be detected as As(III) or As(V), but does not provide a mass spectrum for positive identification.[16] The inductively coupled plasma (ICP) detector is very sensitive, detecting As levels down to 0.1 pg; LODs were <1 ppb.

4.3 Dialkyl Alkylphosphonates

Dialkyl alkylphosphonates are often found as impurities in nerve agents, e.g. diisopropyl methylphosphonate **(31)** [1445-75-6] in GB and diethyl methylphosphonate **(32)** [683-08-9] in VX. These diesters are usually more stable in the environment than the CW agents. Dialkyl alkylphosphonates may also be formed in decontamination reactions, e.g. by reactions with ethanol or methoxyethanol present in some decontaminants. They may be present in the aqueous extracts of soil samples and can be identified by LC/MS. The positive APCI and ES spectra of dialkyl alkylphosphonates give intense $[M + H]^+$ ions, plus metal or solvent adducts and one or

two low abundance fragment ions, depending on the conditions. With the exception of methyl esters, CID spectra of the protonated molecules are dominated by the product ions arising from sequential neutral losses of alkene, with hydrogen transfer to P–O, to give ions **(33)** and **(34)**,

(31) R = *i*-Pr **(32)** R = Et

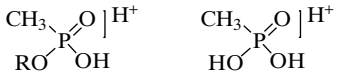

(33) **(34)**

e.g. $[M + H - 28]^+$ for ethyl esters, $[M + H - 42]^+$ for propyl esters. These product ions provide a facile means of obtaining a tentative structure for unknowns. Methyl esters fragment by a different pathway, apparently by loss of methanol.[42] Dialkyl alkylphosphonates are easily distinguished from isomeric alkyl alkylphosphonic acids because they do not give significant responses in negative ion mode.

4.4 Toxins

4.4.1 Saxitoxin

Saxitoxin **(11)** produces satisfactory spectra with TS, fast atom bombardment (FAB), ES and IS.[17] Positive IS is very sensitive for saxitoxin (down to 30 pg injected) and gives an abundant $[M + H]^+$ ion. Saxitoxin is usually analyzed as one of several structurally related paralytic shell poisons. The variability in charge states of these toxins creates problems in finding a mobile phase that gives good chromatography and allows efficient IS ionization. Pleasance et al.[43] reported an LC/ISMS method for the analysis of saxitoxin and related toxins, using a polymeric PRP-1 column with a mobile phase of acetonitrile–5 mM aqueous ammonium acetate and gradient elution. The toxins eluted in order of increasing positive charge, saxitoxin as a rather broad tailing peak with poor reproducibility. Using SIM, a LOD around 0.8–1 $\mu g\,mL^{-1}$ was estimated (1 µL injection). The highly charged toxins are ideally suited to CE separation and CE/MS looks more promising than LC/MS.

4.4.2 Ricin

No LC/MS method has yet been reported for ricin. ESMS of a sample isolated from *Ricinus communis* var *zanzibariensis*, purified by affinity chromatography,

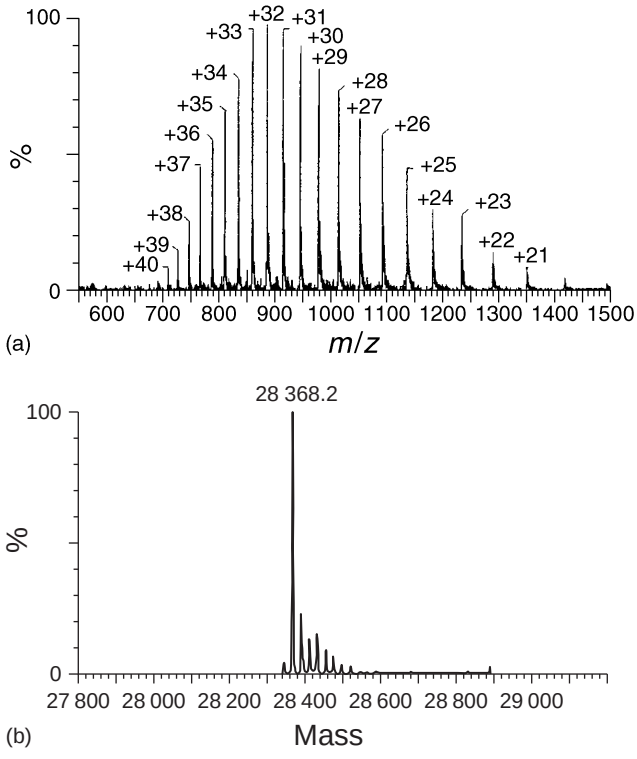

Figure 6 ES mass spectrum of SEB (20 pmol µL^{-1} in 1:1 CH_3CN-H_2O + 0.1% formic acid) (a) and deconvoluted spectrum (b).

contained at least three components with a molecular mass ranging from 62 360 to 63 380 Da.[44] These probably arise from variations in co- or post-translational glycosylation in the biosynthesis.

4.4.3 Staphylococcal Enterotoxin B

The characterization and LC/ESMS of SEB was reported by Kientz et al.[18] Chromatography was performed on a microcolumn (600 mm × 0.3 mm ID) with a TSK-gel Phenyl-5PW column packing. The best chromatographic results were obtained using water–acetonitrile gradients with added 0.2% TFA. TFA gave superior peak shape and retention but a four-fold reduction in sensitivity compared to formic acid. SEB could be determined at levels down to 3 pmol mL^{-1}. Additional evidence of identification was obtained using LC/MS/MS of tryptic digests. A typical ES spectrum of SEB is shown in Figure 6.[44]

5 BIOMEDICAL APPLICATIONS

In the 1990s there has been increased interest in developing forensic methods for the retrospective confirmation of poisoning in victims of chemical attacks. Methods

for the detection of sulfur mustard poisoning are now well developed and some of these rely on LC/MS rather than GC/MS. Methods for the retrospective identification of nerve agent poisoning are still being developed, particularly with regard to blood analysis.

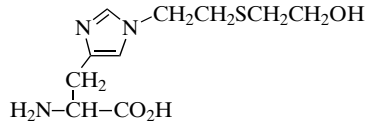

(37)

5.1 Biological Markers of Sulfur Mustard Poisoning

Biological markers of sulfur mustard poisoning include urinary metabolites, amino acid residues on the abundant blood protein hemoglobin alkylated with the 2-hydroxyethylthioethyl (HETE) moiety, and similarly alkylated DNA.

5.1.1 Urinary Metabolites

Important urinary metabolites include thiodiglycol sulfoxide (and to a much lesser extent thiodiglycol), conjugates of mustard sulfone with cysteine, e.g. **(35)**, and metabolites formed from the β-lyase pathway, e.g. **(36)**. These metabolites were initially identified by LC/TSMS/MS of fractions separated by semipreparative high-performance liquid chromatography (HPLC), obtained after rats were treated with ^{35}S, $^{13}C_4$-labelled sulfur mustard.[45] Concentrations in the urine of alleged victims of sulfur mustard poisoning are typically in the range 0.1–100 ppb and detection at the lower limits of this range is difficult to achieve using LC/MS. The bis-sulfone cysteine conjugate **(35)** was not amenable

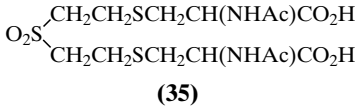

(35)

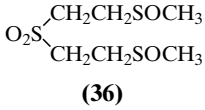

(36)

to GC/MS and an LC/TSMS/MS method was developed based on SRM of the bis-methyl ester, LOD 25 ppb.[46] Thermal instability is a problem with these conjugates and APCI or ES are likely to be superior to TS. The conjugate was not detected in urine samples obtained from casualties of sulfur mustard poisoning (but β-lyase metabolites and thiodiglycol sulfoxide were detected using more sensitive GC/MS methodology).

5.1.2 Alkylated Hemoglobin

Micro-LC/ESMS/MS[47] and LS/ESMS/MS[48] were used to identify the sites of alkylation of hemoglobin by sulfur mustard, by sequencing alkylated peptides detected in tryptic digests of hemoglobin isolated from human blood or hemoglobin treated with sulfur mustard. HETE alkylated N-terminal valine and HETE-histidine **(37)** were identified as important biological markers.

GC/MS methods were developed for detection of the alkylated valine but problems were experienced in finding thermally stable derivatives for the histidine adduct **(37)**. LC/MS of the underivatized alkylated histidine was possible but the chromatography was poor. Superior LODs were obtained by conversion to the Fmoc derivative (commonly used in liquid chromatography/ultraviolet (LC/UV) analysis of amino acids);[28,49] the ES spectrum exhibited the protonated molecule as the only significant ion; monitoring the fragmentation $[M+H]^+ \longrightarrow [CH_2CH_2SCH_2CH_2OH]^+ (m/z\ 482 \longrightarrow m/z\ 105)$ gave good sensitivity (10 pg injected). The method detected the biological marker in two human subjects who had been exposed to sulfur mustard (Figure 7).

5.1.3 Alkylated DNA

Sulfur mustard alkylates DNA, and N7-HETE-guanine **(38)** was identified as an alternative biological marker of sulfur mustard poisoning, initially using LC/TS/MS.[50] As found with HETE-histidine, GC/MS analysis was unsatisfactory and an analogous LC/ESMS/MS method was used based on monitoring the reaction $[M+H]^+ \longrightarrow [CH_2CH_2SCH_2CH_2OH]^+ (m/z\ 256 \longrightarrow m/z\ 105)$ for its

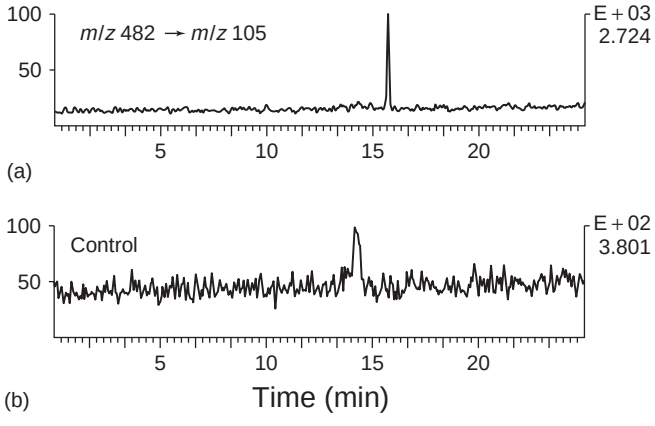

Figure 7 LC/ESMS/MS SRM chromatogram showing the detection of HETE-histidine in the hemoglobin of a casualty of sulfur mustard poisoning (a) and control (b). (Reproduced by permission of Taylor & Francis from R.M. Black, et al., *Xenobiotica*, **27**, 499–512 (1997).[28])

detection in urine.[51] The LOD was 0.2 ng mL^{-1} in urine or 8 pg injected (S/N = 5). The method is also suitable for the determination of the adduct in skin and blood. The adduct was readily detected in the urine of guinea pigs exposed to sulfur mustard.

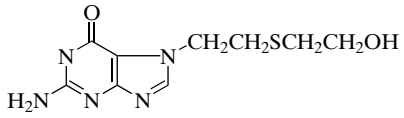

(38)

6 METHOD DEVELOPMENT

Development of LC/MS methods requires simultaneous consideration of LC and MS aspects, for example ionization efficiency is affected by mobile-phase composition and flow rate. The following discussion is based on the authors' experience with the Finnigan API source/interface, although it should be generally applicable to other manufacturers' interfaces, and to TS/MS. Zhou and Hamburger[36] have provided a detailed discussion of the parameters to consider when developing screening procedures based on LC/APIMS; Lacorte et al.[52] reported the effects of temperature and extraction voltages in LC/APCIMS of organophosphorus pesticides; optimization strategies for LC/TSMS have been discussed by Heeremans et al.[53]

6.1 Liquid Chromatography Aspects

The majority of LC separations of CWC-related chemicals are performed using reversed-phase conditions on C_8 or C_{18} columns. Unlike LC/UV methods, exhaustive optimization of LC conditions is unnecessary owing to the selectivity of MS detection; simple isocratic or, more commonly, linear gradient separations are usually adequate. Modern type B silica packings allow good chromatography of acidic, basic and neutral CW agent degradation products to be obtained using methanol–water or acetonitrile–water mobile phases containing low concentrations of acidic modifiers or volatile buffers. Poor retention of the more polar analytes, such as MPA, is problematic.

The effect on ionization of mobile phase constituents and modifiers needs some consideration. TFA at concentrations > ~0.05% suppresses ES ionization with consequent reduced sensitivity; formic acid can be used for greater sensitivity. TFA can give better separations, particularly of basic compounds and peptide or amino acid adducts. The use of ammonium acetate or formate buffer allows analyses to be performed in both positive and negative ion modes under similar LC conditions. This is particularly useful for the selective detection of acidic hydrolysis products in screening procedures.

Column dimensions are of lesser importance and internal diameters from 4.6 mm down to capillary have been used. Columns with 2 mm internal diameter appear to be a good compromise. Flow rates of 0.2 mL min^{-1} are compatible with pneumatically assisted ES or IS without resort to postcolumn splitting; injection volumes of at least 20 µL are possible (though some peak splitting is occasionally observed) and accurate gradients can be delivered by most modern LC pumps. Capillary columns are worth consideration if sample volumes are limited and suitable pumps are available (they are also far more economical with solvents). Column length is dependent on the chromatographic resolution required. Lengths of greater than 150 mm are seldom necessary except for the analysis of complex samples, e.g. protein or peptide digests, where 250 mm columns may be more appropriate.

6.2 Mass Spectrometry Aspects

The most important MS consideration is the ionization technique. TS is now of lesser importance than the more sensitive and robust API techniques. In our experience, APCI is more suitable for small molecules such as CW-agent hydrolysis products and ES is more suitable for the analysis of more complex biological markers. ES tends to suffer from higher background at low m/z values than APCI, although this may be dependent on interface design.

Once the ionization technique has been chosen, various parameters need to be optimized for the analysis in question. The main requirement is to achieve efficient ionization and ion transmission. Optimization requires a systematic approach, particularly as some parameters interact. The most important are spray voltage in ES, corona voltage and vaporizer temperature in APCI, capillary temperature and voltage, tube lens voltage and the mobile-phase flow rate and composition. Most of these parameters can be optimized using loop injections, or infusion of analyte into the LC mobile-phase flow, with the assistance of data system software procedures. The effect of capillary temperature on the intensities of the protonated molecules of EMPA, PinMPA and cHexMPA is shown in Figure 8. Optimum intensities are obtained at a temperature around 150 °C; as the temperature is raised towards 200 °C, $[M + H]^+$ decreases and the fragment $[MeP(O)(OH)_2 + H]^+$ increases.

When a single analyte is of interest, optimization is performed using a structurally significant ion from that compound, usually $[M + H]^+$. For general screening procedures a careful choice of analyte is required. For example, when screening for CW agent hydrolysis products we have found that optimization of interface

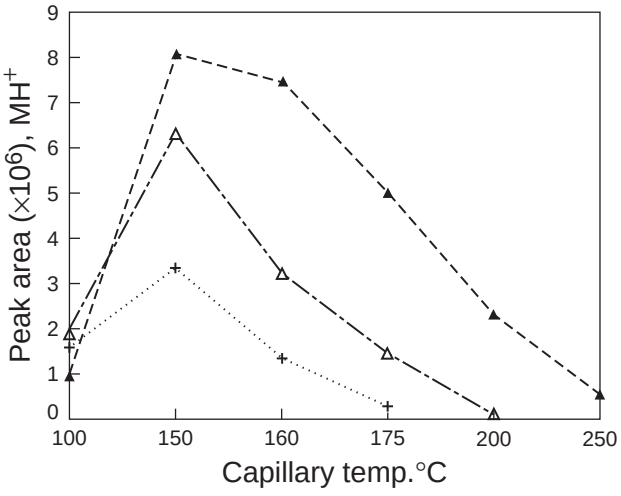

Figure 8 Effect of capillary temperature on the abundance of the protonated molecules in the APCI spectra of EMPA, ; PinMPA, +; and cHexMPA, △.

parameters to give maximum $[M + H]^+$ (*m*/*z* 181) for PinMPA results in conditions giving good sensitivity for all analytes. The main reason for this choice is that of the compounds in our screening procedure, PinMPA has one of the least intense $[M + H]^+$ ions. Similarly, the corresponding $[M - H]^-$ ion (*m*/*z* 179) is used for negative-ion mode. Optimization should be carried out at the appropriate mobile phase composition.

Solvent cluster and adduct ions can cause problems in optimization procedures. When using methanol–water mobile phases, particularly with ES at high methanol concentrations, a cluster ion $[(MeOH)_3 + H]^+$ is formed. This ion is isobaric with the $[MePO(OH)_2 + H]^+$ ion (*m*/*z* 97) present in the spectra of all alkyl MPAs and can cause high backgrounds in mass chromatograms of *m*/*z* 97. Similarly, when analyzing alkyl MPAs in mobile phases containing acetonitrile, an adduct ion $[MePO(OH)_2 + H + 2CH_3CN]^+$ is isobaric with $[M + H]^+$ for cHexMPA (*m*/*z* 179). Thus, when optimizing MS conditions it is necessary to ensure that a genuine analyte ion is being used and not an interfering cluster, adduct or other background ion that may result in spurious optima being obtained.

The intensity of solvent adduct ions may be reduced by use of a small source octapole offset voltage. Too high a voltage may result in prominent fragment ions due to source CID. Alkali metal adduct ions are not generally prominent in LC/MS, although they may be visible in ES spectra. Ammonium ion adducts are formed with alkyl alkylphosphonic acids when using ammonium acetate or formate as mobile phase modifiers, but are not of great intensity provided modifier concentrations are low. Ammonium adducts are more prominent in the spectra of MPA and EPA, and the neutral analytes thiodiglycol and thiodiglycol sulfone.

6.3 Quality Control

Quality control of screening procedures is performed using analyses of solutions containing a representative mixture of CW-agent hydrolysis products, allowing the quality of both chromatography and mass spectra to be assessed, and determination of LODs. Standard LC column test mixes are not suitable for CW-related analyses as the compounds in these mixes are neither chromatographically nor mass spectrometrically representative of the analytes of interest. In the case of assays for single biological markers, the analysis of standards is required to construct calibration curves and determine detection limits. The use of isotopically labeled internal standards for quantitative analyses is recommended wherever possible.

7 COMPARISON WITH OTHER HYPHENATED TECHNIQUES

7.1 Gas Chromatography/Mass Spectrometry

LC/MS is not used as a substitute for GC/MS in our laboratory but as a complementary technique, mainly for the analysis of polar degradation products. The clear advantage of LC/MS is in the simplicity of sample preparation and speed of analysis. For GC/MS screening procedures, the aqueous extract is concentrated to dryness and the hydrolysis products converted to trimethylsilyl (TMS), *tert*-butyldimethylsilyl (TBDMS) derivatives, or to methyl esters in the case of phosphonic acids and benzilic acid.[15] This is time-consuming and derivatization can result in a large variability in the apparent recovery, particularly in the case of silyl derivatives of alkylphosphonic acids. Problems that can seriously affect derivatization include the presence of large amounts of extraneous organic materials, residual traces of water, and the presence of divalent metal ions such as Ca^{2+} in the case of phosphonic acids. LC/MS allows the very rapid screening of aqueous samples and extracts, provided that analytes are present at concentrations above ca. 0.1–1 $\mu g\,mL^{-1}$. It is particularly useful in OPCW proficiency tests where a very large choice of analytes is available from the thousands of possible scheduled compounds and chemical synthesis within a few days may be required for confirmation. Even where samples do have to be concentrated prior to analysis, e.g. using SPE, LC/MS may have advantages because no precautions need to be taken to remove traces of water from the cartridge.

The range of compounds amenable to LC/MS analysis is much broader than with GC/MS. Thus LC/MS can be used to analyze moderately hydrophobic compounds such as VX, very polar compounds such as thiodiglycol sulfoxide, and large molecular mass toxins such as SEB. An additional advantage is that LC/MS and LC/MS/MS spectra are usually very simple, and tentative identification of an unknown may be much easier than for a silylated derivative, where fragmentations can be complex. An example is the additional compounds identified in the soil extract from bomb craters described in section 4.2.

Other useful applications of LC/MS include method development for sample preparation. Recoveries of analytes such as thiodiglycol or thiodiglycol sulfoxide can often be more reliably determined using LC/MS because potential variability due to derivatization is avoided.

On the negative side, the resolution obtainable with capillary gas chromatography columns is superior to that currently obtainable with LC. In terms of sensitivity, and LODs in most matrices, GC/MS is superior. Where trace level analysis is required, e.g. for metabolites such as alkyl alkylphosphonic acids in urine, GC/MS/MS remains the method of choice where suitable derivatives can be found. With more complex biological markers, such as alkylated histidine, thermal instability may prohibit analysis by GC/MS/MS and LC/MS/MS then becomes the method of choice.

7.2 Capillary Electrophoresis/Mass Spectrometry

Only one CE/MS method has been reported for CW-agent analysis at the time of writing, although further development is in progress. Kostiainen et al.[54] successfully combined capillary zone electrophoresis (CZE) with IS for the analysis of five nerve agent hydrolysis products. The technique provides higher inherent sensitivity and shorter analysis times than LC/MS, but the LODs obtainable are limited by the small injection volumes. CE in combination with less specific detection, e.g. indirect ultraviolet (UV),[55] has been reported for the analysis of phosphonic acids. CE is a good technique for molecules that are ionized in solution, such as phosphonic acids, toxins such as saxitoxin **(11)**, and peptides and proteins, but not for neutral molecules such as thiodiglycol. LC/MS remains the most versatile and broadly applicable method for the screening of aqueous solutions for polar compounds of relevance to the CWC.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
BZ	Quinuclidinyl Benzilate
CE	Capillary Electrophoresis
cHexMPA	Cyclohexyl Methylphosphonic Acid
CI	Chemical Ionization
CID	Collision-induced Dissociation
CW	Chemical Warfare
CWC	Chemical Weapons Convention
CZE	Capillary Zone Electrophoresis
EBQQ	Electrostatic Sector Magnetic Sector Quadrupole Quadrupole
EI	Electron Ionization
EMPA	Ethyl Methylphosphonic Acid
EPA	Ethylphosphonic Acid
ES	Electrospray
FAB	Fast Atom Bombardment
Fmoc	Fluorenylmethoxycarbonyl
GA	Tabun
GB	Sarin
GC/MS	Gas Chromatography/Mass Spectrometry
GD	Soman
GF	Cyclohexyl Sarin
HETE	2-Hydroxyethylthioethyl
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IS	Ionspray
LC	Liquid Chromatography
LC/APCIMS	Liquid Chromatography/Atmospheric Pressure Chemical Ionization/ Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LC/TSMS	Liquid Chromatography/Thermospray Mass Spectrometry
LC/UV	Liquid Chromatography/Ultraviolet
LOD	Limit of Detection
MPA	Methylphosphonic Acid
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
ODS	Octadecylsilane
OPCW	Organization for the Prohibition of Chemical Weapons
PB	Particle Beam
PinMPA	Pinacolyl Methylphosphonic Acid
ROP	Recommended Operating Procedure
SEB	Staphylococcal Enterotoxin B
SIM	Selected Ion Monitoring
S/N	Signal-to-noise Ratio
SPE	Solid-phase Extraction

SPME	Solid-phase Microextraction
SRM	Selected Reaction Monitoring
TBDMS	*tert*-Butyldimethylsilyl
TFA	Trifluoroacetic Acid
TMS	Trimethylsilyl
TS	Thermospray
UNSCOM	United Nations Special Commission
UV	Ultraviolet
VX	*O*-Ethyl *S*-2-Diisopropylaminoethyl Methylphosphonothiolate

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis

Pesticides **(Volume 7)**
High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. W.M.A. Niessen, A.P. Tinke, 'Liquid Chromatography–Mass Spectrometry General Principles and Instrumentation', *J. Chromatogr., A*, **703**, 37–57 (1995).
2. M.H. Lamoree, R.T. Ghijsen, U.A.Th. Brinkman, 'LC–MS Interfacing Systems in Environmental Analysis: Application to Polar Pesticides, in *Environmental Analysis Techniques, Applications and Quality Assurance*, ed. D. Barcelo, Elsevier, Amsterdam, 1993.
3. J. Slobodnik, B.L.M. Van Baar, U.A.Th. Brinkman, 'Column Liquid Chromatography–Mass Spectrometry: Selected Techniques in Environmental Applications for Polar Pesticides and Related Compounds', *J. Chromatogr., A*, **703**, 81–121 (1995).
4. M. Careri, A. Mangia, M. Musci, 'Applications of Liquid Chromatography–Mass Spectrometry Interfacing Systems in Food Analysis: Pesticide, Drug and Toxic Substance Residues', *J. Chromatogr., A*, **727**, 153–184 (1996).
5. E.R.J. Wils, A.G. Hulst, 'Determination of *O*-Ethyl *S*-2-Diisopropylaminoethyl Methylphosphonothiolate (VX) by Thermospray Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **523**, 151–161 (1990).
6. E.R.J. Wils, A.G. Hulst, 'Determination of Organophosphorus Acids by Thermospray Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **454**, 261–272 (1988).
7. S. Pleasance, J.F. Anacleto, M.R. Bailey, D.H. North, 'An Evaluation of Atmospheric Pressure Ionization Techniques for the Analysis of *N*-Methyl Carbamate Pesticides by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **3**, 378 (1992).
8. C.E. Kientz, A. Hulst, A.L. De Jong, E.R.J. Wils, 'Eluent Jet Interface for Combining Capillary Liquid Flows with Electron Impact Mass Spectrometry', *Anal. Chem.*, **68**, 675–681 (1996).
9. P.R. Tiller, J.B. Cunniff, A.P. Land, 'Advantages of Full Scan Tandem Mass Spectrometry for Quantitative Analysis on a Modern Ion Trap Mass Spectrometer: Increased Selectivity and Mixture Analysis', *Rapid Commun. Mass Spectrom.*, **11**, 1151–1153 (1997).
10. W.M.A. Niessen, 'Advances in Instrumentation in Liquid Chromatography–Mass Spectrometry and Related Liquid-introduction Techniques', *J. Chromatogr., A*, **794**, 407–435 (1998).
11. 'Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction', Organisation for the Prohibition of Chemical Weapons, The Hague, 1994.

12. Z. Witkiewicz, M. Mazurek, J. Szulc, 'Chromatographic Analysis of Chemical Warfare Agents', *J. Chromatogr.*, **503**, 293–357 (1990).
13. A.F. Kingery, H.E. Allen, 'The Environmental Fate of Organophosphorus Nerve Agents: A Review', *Toxicol. Environ. Chem.*, **47**, 155–184 (1995).
14. E.R.J. Wils, A.G. Hulst, 'The Use of Thermospray–Liquid Chromatography/ Mass Spectrometry for the Verification of Chemical Warfare Agents', *Fresenius' J. Anal. Chem.*, **342**, 749–758 (1992).
15. 'Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament, 1994 Edition', ed. M. Rautio, The Ministry for Foreign Affairs of Finland, Helsinki, 1994.
16. D.A. Bass, J.S. Yaeger, J.T. Kiely, J.S. Crain, L.M. Shem, H.J. O'Neill, M.J. Gowdy, 'Detecting and Quantifying Lewisite Degradation Products in Environmental Samples Using Arsenic Speciation', Proceedings of the 1995 Scientific Conference on Chemical and Biological Defense Research, US Army Edgewood Research Development and Engineering Center, 14–17 Nov 1995, 233–239, 1995.
17. M.A. Quilliam, 'Liquid Chromatography–Mass Spectrometry of Seafood Toxins', *J. Chromatogr. Library*, **59**, 415–444 (1996).
18. C.E. Kientz, A.G. Hulst, E.R.J. Wils, 'Determination of Staphylococcal Enterotoxin B by On-Line (Micro) Liquid Chromatography–Electrospray Mass Spectrometry', *J. Chromatogr., A*, **757**, 51–64 (1997).
19. I. Liska, 'On-line versus Off-line Solid-phase Extraction in the Determination of Organic Contaminants in Water', *J. Chromatogr., A*, **655**, 163–176 (1993).
20. J. Schulein, D. Martens, P. Spitzauer, A. Kettrup, 'Comparison of Different Solid Phase Extraction Materials and Techniques by Application of Multiresidue Methods for the Determination of Pesticides in Water by High-performance Liquid Chromatography (HPLC)', *Fresenius' J. Anal. Chem.*, **352**, 565–571 (1995).
21. M.T. Sng, F.K. Lee, H.A. Lakso, 'Solid-phase Microextraction of Organophosphorus Pesticides from Water', *J. Chromatogr., A*, **759**, 225–230 (1997).
22. M.L. Shih, J.R. Smith, J.D. McMonagle, T.W. Dolzine, V.C. Gresham, 'Detection of Metabolites of Toxic Alkylmethylphosphonates in Biological Samples', *Biol. Mass Spectrom.*, **20**, 717–723 (1991).
23. S-Å. Fredriksson, L-G. Hammarström, L. Henriksson, H-Å. Lakso, 'Trace Determination of Alkyl Methylphosphonic Acids in Environmental and Biological Samples Using Gas Chromatography/Negative-ion Chemical Ionization Mass Spectrometry and Tandem Mass Spectrometry', *J. Mass Spectrom.*, **30**, 1133–1143 (1995).
24. J.Aa. Tørnes, B.A. Johnsen, 'Gas Chromatographic Determination of Methylphosphonic Acids by Methylation with Trimethylphenylammonium Hydroxide', *J. Chromatogr.*, **467**, 129–138 (1989).
25. K. Jinno, T. Muramatsu, Y. Saito, Y. Kiso, S. Magdic, J. Pawliszyn, 'Analysis of Pesticides in Environmental Water Samples by Solid-phase Microextraction/High-performance Liquid Chromatography,' *J. Chromatogr., A*, **754**, 137–144 (1996).
26. H. Lakso, F.W. Ng, 'Determination of Chemical Warfare Agents in Natural Water Samples by Solid-phase Microextraction', *Anal. Chem.*, **69**, 1866–1872 (1997).
27. A. Alcaraz, S.S. Hulsey, R.E. Whipple, B.D. Andersen, 'On-site Sample Work-up Procedures to Isolate Chemical Warfare Related Compounds Using Solid Phase Extraction and Solid Phase Microextraction Technology', in *Analytical Chemistry Associated with the Destruction of Chemical Weapons*, eds. M. Heyl and R. McGuire, Kluwer Academic Publishers, Dordrecht, 65–76, 1997.
28. R.M. Black, R.J. Clarke, J.M. Harrison, R.W. Read, 'Biological Fate of Sulfur Mustard: Identification of Valine and Histidine Adducts in Hemoglobin from Casualties of Sulfur Mustard Poisoning', *Xenobiotica*, **27**, 499–512 (1997).
29. W.R. Creasy, J.R. Stuff, B. Williams, K. Morrissey, J. Mays, R. Duevel, H.D. Durst, 'Identification of Chemical-weapons-related Compounds in Decontamination Solutions and Other Matrices by Multiple Chromatographic Techniques', *J. Chromatogr., A*, **774**, 253–263 (1997).
30. W.R. Creasy, T.G. Albro, R. Cheicante, Jr., J.R. Stuff, 'Determining Nonvolatile, Polar Compounds in Environmental Samples', *Spectroscopy*, **9**, 42–47 (1994).
31. J.F. Weimaster et al., 'Chemical Analysis of Environmental Samples Collected in Iraq: Analysis for the Presence of Chemical Warfare Agents', *J. Chem. Technol. Biotechnol.*, **64**, 115–128 (1995).
32. J.A. Tørnes, 'Identification of Some Alkyl Methylphosphonic Acids by Thermospray Tandem Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 878–882 (1996).
33. S. Munavalli, E.M. Jakubowski, H.D. Durst, 'Liquid Chromatography/Thermospray Mass Spectrometry of Mustard and its Metabolites', *J. Mass Spectrom.*, **30**, 1716–1722 (1995).
34. R.M. Black, R.W. Read, 'Application of Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry, and Tandem Mass Spectrometry, to the Analysis and Identification of Degradation Products of Chemical Warfare Agents', *J. Chromatogr., A*, **759**, 79–92 (1997).
35. R.M. Black, R.W. Read, 'Analysis of Degradation Products of Organophosphorus Chemical Warfare Agents and Related Compounds by Liquid Chromatography–Mass Spectrometry Using Electrospray and Atmospheric Pressure Chemical Ionization', *J. Chromatogr., A*, **794**, 233–244 (1998).
36. S. Zhou, M. Hamburger, 'Application of Liquid Chromatography–Atmospheric Pressure Ionization Mass

Spectrometry in Natural Product Analysis. Evaluation and Optimization of Electrospray and Heated Nebulizer Interfaces', *J. Chromatogr., A*, **755**, 189–204 (1996).
37. V.T. Borrett, R. Colton, J.C. Traeger, 'The Electrospray Mass Spectra of Phosphoric Acid, Methylphosphonic Acid and its Alkyl Esters, and their Complexes with Alkali and Alkali Earth Metal Ions', *Eur. Mass Spectrom.*, **1**, 131–140 (1995).
38. V.T. Borrett, R.J. Mathews, R. Colton, J.C. Traeger, 'Verification of the United Nations Chemical Weapons Convention: the Application of Electrospray Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 114–118 (1996).
39. P. Siegenthaler, 'The Use of LC/MS Techniques for Screening and Identification of Degradation Products of Nerve Agents', Poster Presentation, Sixth International Symposium on Protection Against Chemical and Biological Warfare Agents, Stockholm, Sweden, 10–14 May, 1998.
40. D.K. Rohrbaugh, Y-C. Yang, 'Liquid Chromatography/Mass Spectrometry of Mustard-related Sulfonium Ions', *J. Mass Spectrom.*, **32**, 1247–1252 (1997).
41. R.J. Dijkstra, J. Reyer, B.L.M. Van Baar, C.E. Kientz, W.M.A. Niessen, U.A.Th. Brinkman, 'An Eluent-Jet Interface for Chemical Ionization Mass Spectrometry and Coupling of Microcolumn Liquid Chromatography with Electron Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**, 5–10 (1998).
42. A.J. Bell, D. Despeyroux, J. Murrell, P. Watts, 'Fragmentation and Reactions of Organophosphate Ions Produced by Electrospray Ionization', *Int. J. Mass Spectrom. Ion Processes*, **165/166**, 533–550 (1997).
43. S. Pleasance, S.W. Ayer, M.V. Laycock, P. Thibault, 'Ionspray Mass Spectrometry of Marine Toxins. III. Analysis of Paralytic Shellfish Poisoning Toxins by Flow Injection Analysis, Liquid Chromatography/Mass Spectrometry and Capillary Electrophoresis/Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 14–24 (1992).
44. D. Despeyroux, P. Watts, CBD unpublished results.
45. R.M. Black, K. Brewster, R.J. Clarke, J.L. Hambrook, J.M. Harrison, D.J. Howells, 'Biological Fate of Sulfur Mustard, 1,1′-Thiobis(2-chloroethane): Isolation and Identification of Urinary Metabolites Following Intraperitoneal Administration to Rat', *Xenobiotica*, **22**, 405–418 (1992).
46. R.M. Black, R.W. Read, 'Biological Fate of Sulfur Mustard, 1,1′-Thiobis(2-chloroethane): Identification of β-Lyase Metabolites and Hydrolysis Products in Human Urine', *Xenobiotica*, **25**, 167–173 (1995).
47. D. Noort, E.R. Verheij, A.G. Hulst, L.P.A. de Jong, H.P. Benschop, 'Characterization of Sulfur Mustard Induced Structural Modifications in Human Hemoglobin by Liquid Chromatography–Tandem Mass Spectrometry', *Chem. Res. Toxicol.*, **9**, 781–787 (1996).
48. R.M. Black, J.M. Harrison, R.W. Read, 'Biological Fate of Sulfur Mustard: In Vitro Alkylation of Human Hemoglobin by Sulfur Mustard', *Xenobiotica*, **27**, 11–32 (1997).
49. D. Noort, A. Fidder, A.G. Hulst, A.L. De Jong, L.P.A. De Jong, H.P. Benschop, 'Dosimetry of Exposure to Sulfur Mustard: Mass Spectrometric Detection of Sulphur Mustard Adducts to Proteins and DNA', Presentation to Colloque, Les Alkylants, CEB Le Bouchet, France, October 1995.
50. A. Fidder, G.W.H. Moes, A.G. Scheffer, G.P. van der Schans, R.A. Baan, L.P.A. de Jong, H.P. Benschop, 'Synthesis, Characterization, and Quantitation of the Major Adducts Formed Between Sulfur Mustard and DNA of Calf Thymus and Human Blood', *Chem. Res. Toxicol.*, **7**, 199–204 (1994).
51. A. Fidder, D. Noort, L.P.A. de Jong, H.P. Benschop, A.G. Hulst, 'N7-(2-Hydroxyethylthioethyl)-guanine: A Novel Urinary Metabolite Following Exposure to Sulfur Mustard', *Arch. Toxicol.*, **70**, 854–855 (1996).
52. S. Lacorte, C. Molina, D. Barceló, 'Temperature and Extraction Voltage Effect on Fragmentation of Organophosphorus Pesticides in Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr., A*, **795**, 13–26 (1998).
53. C.E.M. Heeremans, R.A.M. Van der Hoeven, W.M.A. Niessen, U.R. Tjaden, J. Van der Greef, 'Development of Optimization Strategies in Thermospray Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **474**, 149–162 (1989).
54. R. Kostiainen, A.P. Bruins, V.M.A. Hakkinen, 'Identification of Degradation Products of Some Chemical Warfare Agents by Capillary Electrophoresis–Ionspray Mass Spectrometry', *J. Chromatogr., A*, **634**, 113–118 (1993).
55. J-P. Mercier, Ph. Morin, M. Dreux, A. Tambute, 'Capillary Electrophoresis Separation of Alkylphosphonic Acid Monoesters with Indirect Ultraviolet Detection', *J. Chromatogr., A*, **779**, 245–252 (1997).

Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Markku Mesilaakso
Finnish Institute for Verification of the Chemical Weapons Convention, University of Helsinki, Finland

Andreas Niederhauser
AC-Laboratory, Spiez, Switzerland

1 Introduction **2**
2 Instruments and Equipment **3**
2.1 Spectrometer 3
2.2 Probe Heads 3
2.3 Automatic Operation 3
2.4 Data Systems 4
3 Experimental **4**
3.1 Solvents 4
3.2 Chemical Shift References 4
3.3 Nuclear Magnetic Resonance Sample Tube 5
3.4 Sample Preparation 5
3.5 Experiments and Data Processing 19
4 Quality Control **20**
4.1 Quality Management System 20
4.2 Data Quality 20
4.3 Spectrometer Performance 20
4.4 Tests and Test Samples 21
5 Applications **21**
5.1 Main One-dimensional Experiments 21
5.2 $^{1}H\{^{31}P\}$, $^{1}H\{^{19}F\}$, and $^{1}H\{^{1}H\}$ Nuclear Magnetic Resonance Experiments 22
5.3 Relevant Two-dimensional Techniques 22
5.4 Scheduled Chemicals 22
5.5 Detection Limits 23
6 Analysis **23**
6.1 Acquired Nuclear Magnetic Resonance Data 23
6.2 Identification Criteria 23
6.3 Interpretation 24
6.4 Data Evaluation 24
6.5 Identification 24
6.6 Nuclear Magnetic Resonance Spectra of Sarin and Tabun 25
6.7 Nuclear Magnetic Resonance Spectral Parameters of Selected Chemical Weapons Convention-related Chemicals 25
7 Databases **25**
7.1 Requirements for an In-house Library 26
7.2 The Organization for the Prohibition of Chemical Weapons Analytical Database 26
7.3 VERIFY Database 27
8 Perspectives and Future Developments **27**
8.1 Liquid Chromatography/Nuclear Magnetic Resonance 27
8.2 Ultrahigh-field Spectrometers 27
8.3 ''Nano'' Probe Technology 27
9 Comparison with Other Spectrometric Techniques **27**
Acknowledgments **28**
Abbreviations and Acronyms **28**
Related Articles **28**
References **29**

Nuclear magnetic resonance (NMR) spectroscopy is a valuable analytical technique for the identification of chemicals related to the Chemical Weapons Convention (CWC). One-dimensional (1-D) proton, carbon, fluorine, and phosphorus NMR spectra (^{1}H, $^{13}C\{^{1}H\}$, ^{19}F, $^{31}P\{^{1}H\}$, and ^{31}P spectra) (terminology: ^{1}H, nucleus hydrogen-1 or proton, experiment hydrogen-1 or proton observation; ^{13}C, nucleus carbon-13; $^{13}C\{^{1}H\}$, experiment carbon-13 observation with simultaneous ^{1}H decoupling; ^{19}F, nucleus fluorine-19, experiment fluorine-19 observation; ^{31}P, nucleus phosphorus-31, experiment phosphorus-31 observation; $^{31}P\{^{1}H\}$, experiment phosphorus-31 observation with simultaneous ^{1}H decoupling) are recorded from test samples for use in the identification of chemicals. The NMR spectrum of a chemical is like a fingerprint. The chemical is identified on the basis of similar sample and reference spectra. In the case of an unknown chemical two-dimensional (2-D) spectra can assist in its structure elucidation. 2-D NMR is also more suitable than 1-D NMR for the analysis of mixtures.

NMR sample preparation is simple. A liquid sample can be run as such or the liquid is partially or totally changed to a deuterated solvent. Solid samples are extracted with organic or aqueous solvents for the preparation of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

analytical samples. The main disadvantage of NMR is the lack of separation methods and the consequent problems of sample background causing occasional resonance overlapping, especially at very low concentrations.

1 INTRODUCTION

NMR spectroscopy is one of the most powerful analytical techniques in organic chemistry for elucidating the molecular structures of chemicals.[1,2] Moreover, an NMR spectrum may be used like a fingerprint to identify a chemical by comparison with its reference spectrum recorded from the authentic chemical under comparable conditions. The spectrum also reveals information on molecular conformation, isomerism, molecular dynamics, and diastereomers.[3–6]

During several international interlaboratory comparison (round-robin) and proficiency tests for the verification of chemical disarmament, NMR spectroscopy proved itself to be a useful complementary technique to mass spectrometry (MS) and infrared (IR) spectroscopy.[7–14] Test samples included aqueous and organic liquids, soils, sands, concrete, paints, and rubbers spiked with chemicals related to the CWC, usually at levels of about 10 ppm.[15] NMR has also been applied to related field sample analysis.[16,17]

NMR spectroscopy is applicable to all chemicals dissolving in sufficient amount in a deuterated solvent. Nondeuterated solvents can be used if the experiment is performed unlocked or if a small amount of the corresponding deuterated solvent is added to provide the lock conditions. This article describes the NMR spectroscopic methods used in the analysis of chemical warfare (CW) agents and related chemicals. The flow chart shown in Figure 1 demonstrates the application of NMR spectroscopy to the field of verification of CWC-related chemicals.

The most common nuclei observed are ^{1}H, ^{31}P, ^{19}F, and ^{13}C. For high concentrations (>1000 ppm) of CWC-related chemicals and low background levels, the ^{1}H, $^{13}C\{^{1}H\}$, ^{19}F, $^{31}P\{^{1}H\}$, and ^{31}P NMR experiments are all useful for identification. For the analysis of chemicals present in trace amounts (low-ppm level) in environmental samples, the more sensitive ^{1}H NMR experiment is the method of choice. A 400-MHz spectrometer will generally offer sufficient sensitivity for the detection of chemicals present in trace amounts, although unknown amounts of background may cause problems. The ^{19}F and $^{31}P\{^{1}H\}$ experiments are useful as screening techniques to check for the presence of these nuclei. The acquisition of proton-coupled ^{31}P spectra, which reveal more detailed structural information, is limited more by the poor sensitivity than by background problems. Recording of $^{13}C\{^{1}H\}$ NMR spectra may be futile in concentrations below 100 ppm. In the international tests, the low sensitivity of NMR experiments was compensated in part by concentrating the NMR samples and by running long-term experiments.

Compared with MS and IR spectroscopy, the main problems in NMR experiments are the lower sensitivity and the resonance overlapping caused by large amounts of background chemicals present in the sample.

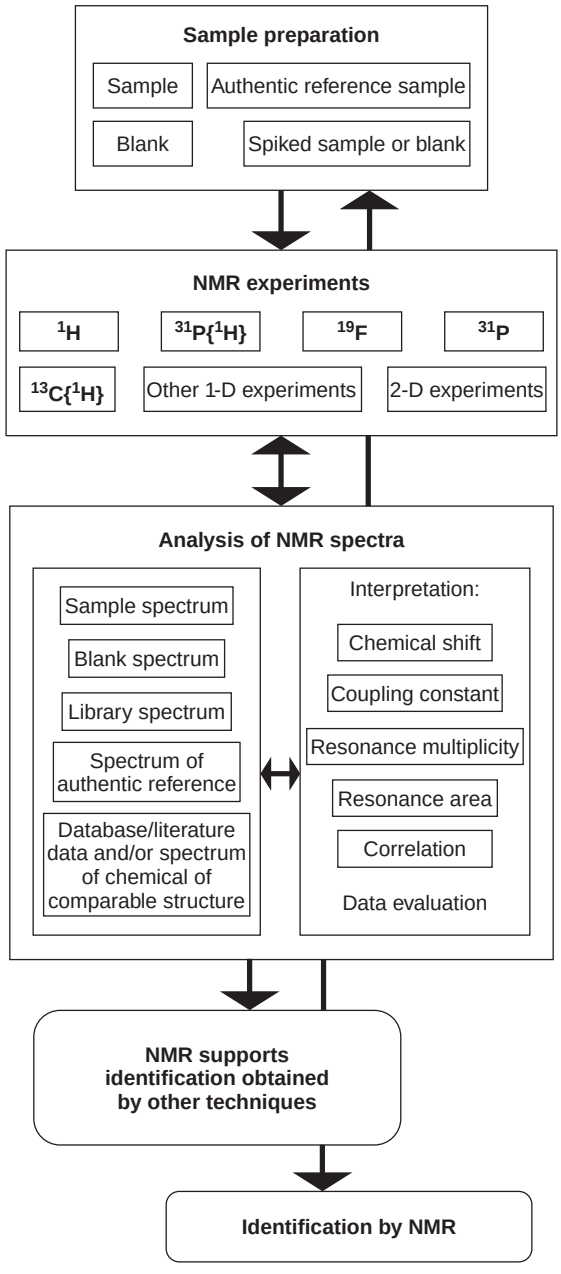

Figure 1 Verification of CWC-related chemicals by NMR spectroscopy.

Techniques other than NMR are usually hyphenated with gas chromatography (GC) or liquid chromatography (LC) to provide the superior analytical power achieved by combining selectivity and sensitivity. However, the structural specificity of NMR spectra, the capability for the observation of different nuclei, and the 2-D correlation experiments make NMR spectroscopy a valuable technique to be used alongside other techniques. With regard to the chemicals listed in the Schedules of the CWC,[15] NMR spectroscopy is particularly useful for the analysis of alkylphosphonates (e.g. precursors or degradation products of nerve agents), thioalcohols, and amino alcohols and their salts, but it is not limited to these.

2 INSTRUMENTS AND EQUIPMENT

2.1 Spectrometer

The following are recommendations for an NMR spectrometer suitable for the analysis of CWC-related chemicals:

- capability for observing ^{1}H, ^{13}C, ^{19}F, and ^{31}P nuclei (recording ^{1}H, $^{13}C\{^{1}H\}$, ^{19}F, $^{31}P\{^{1}H\}$, and ^{31}P NMR spectra);
- a proton resonance frequency preferably of 200 MHz or more;
- capability for control of the sample temperature to an accuracy of within $\pm 0.5\,^{\circ}C$;
- a data system adequate for acquisition, processing, printing, and storage of all necessary data.[18]

Modern computerized high-field NMR spectrometers normally fulfill these requirements and, in addition, are routinely capable of running the latest pulse sequences to record the multidimensional spectra useful for structure elucidation of unknown chemicals. The higher the resonance frequency is, the smaller the amount of a chemical needed for its identification. Higher-frequency instruments also provide larger resonance dispersion, which simplifies the spectra and helps in their analysis. In interlaboratory comparison/proficiency tests, the proton frequencies have ranged from 250 to 600 MHz, but even the lower-frequency instruments have produced useful spectra. Temperature control is recommended because line broadening may occur with a drift in temperature – an effect that may be particularly important during long-term accumulations.

2.2 Probe Heads

The capability of recording the spectra of different nuclei depends on the probe head. ^{1}H, ^{13}C, ^{19}F, and ^{31}P nuclei are observed in verifying CWC-related chemicals and the probe head is selected accordingly. The most common probe head diameters are 5 and 10 mm. For the same sample concentration a 10-mm probe head is more sensitive than a 5-mm probe head, but the former requires a larger sample volume and accordingly more analyte. The 5-mm probe heads were the most common choice in interlaboratory comparison/proficiency tests; in the selection of a probe head, its sensitivity for a particular nucleus is also important. For probe heads available, see catalogues of NMR spectrometer and probe head manufacturers.

Most common is a switchable probe head, which can be used to observe ^{1}H and all NMR active nuclei from the low-frequency limit up to the frequency of ^{31}P. The proton coil can be tuned for the observation of ^{19}F. The switchable probe head is designed for either direct or inverse observation. The direct observation probe head is most sensitive for 1-D experiments on ^{13}C and ^{31}P. The inverse probe head in turn is most sensitive for the direct observation of ^{1}H and indirect detection, for example of ^{31}P, in 2-D experiments taking advantage of polarization-transfer phenomena.

A four-nucleus probe head with which ^{1}H, ^{13}C, ^{19}F, and ^{31}P can be observed is also useful. Such a probe head can be pretuned for each nucleus, and the switching between nuclei is easy and can even be automated. Sensitivity is comparable to that of the switchable probe head with direct observation.

Selective probe heads are used for optimal sensitivity for a particular nucleus. Sensitivity of a selective ^{1}H-probe head is normally greater than that of a switchable probe head with indirect observation. With a selective X-nucleus (a nucleus other than proton) probe head, decoupling of protons is normally possible. Because of their limited usefulness, selective probe heads are rare in NMR laboratories. Other probe heads are also available, e.g. ones for triple resonance experiments and experiments utilizing pulsed-field gradients. In addition to their suitability for 2-D experiments, the gradients are particularly suitable for solvent suppression.[19]

2.3 Automatic Operation

An automatic sample changer (ASC) is often an option for a spectrometer. The operator sets the samples into the rack of the ASC and predefines the experiments, which are then performed with automatic homogeneity adjustment, processing, and plotting of data. When the experiments for one NMR sample are completed, the sample is changed. In this way, the spectrometer can work unattended during nights and weekends. The ASC works best for recording spectra from standard NMR samples in which the concentration is relatively high. Requirements for optimal experimental performance for measuring test samples, which may contain anything from

organic liquids to aqueous liquids of high salt content, may be too demanding for the ASC.

The main requirements for measuring test samples are (a) separate tuning of the probe head for each sample and nucleus in order to guarantee optimum sensitivity, (b) accurate homogeneity adjustment even for the most difficult samples, in order to achieve the narrowest lines and pure line shape, and (c) plotting of expanded regions for spectra where the resonances of interest may be less than 1% of the intensity of the background resonances. Manual operation by an experienced operator is mandatory if highest quality spectra are to be guaranteed.

Automatic operation is possible without a sample changer if ''macros'' are available or programmable spectrometer software. After manual tuning for the observed nucleus, manual homogeneity adjustment, and a manual check for an acceptable line shape, defined experiments can be run one after the other with a single sample. A series of experiments can be run overnight or longer, with the computer taking care of the measurements. The data are plotted manually for evaluation and analysis.

The spectral analysis is carried out manually because automatic interpretation and library programs are normally not available. Difficulties in NMR and automatic interpretation are (a) high spectral background in spectra recorded from environmental samples, often leading to resonance overlap, (b) solvent dependence of chemical shifts (δ), which affects the appearance of the spectrum, and (c) in the case of ^{1}H NMR spectra, the complexity. The other spectra, particularly $^{13}C\{^{1}H\}$, are simple, but low sensitivity is then a problem.

2.4 Data Systems

The spectrometer computer is equipped with a computer operating system (e.g. a version of UNIX) and the specific operating system for the spectrometer. After the experiments, data can be stored on different media, sent through the internet, processed in various ways, and plotted.

3 EXPERIMENTAL

3.1 Solvents

Deuterated solvents are preferred for NMR samples. The solvents should be of spectroscopic grade or corresponding quality and should have a high enrichment of deuterium, preferably >99.5%. Deuterium oxide (D_2O), chloroform-d_1 ($CDCl_3$), acetone-d_6 [$(CD_3)_2CO$], and dichloromethane-d_2 (CD_2Cl_2) are common in the analysis of CWC-related chemicals, but many others [e.g. acetonitrile-d_3 (CD_3CN), methanol-d_4 (CD_3OD)] are available and may be suitable. The solvent is selected according to the requirements of the sample preparation. Where evaporation of the (protonated) solvent is not desirable, a small portion of the corresponding deuterated solvent is added to the sample [e.g. 5–10% (v/v)] for field/frequency stabilization.

The impurities in some solvents may disturb the analysis of low-ppm concentrations of CWC-related chemicals through resonance overlap. Traces of water are common in organic solvents, but the water can be removed by using a suitable drying agent in contact with the sample solution before transferring it to an NMR tube. A representative blank sample, if available, or a blind sample (solvents and reagents that have gone through the sample preparation path without having been in touch with the test sample) serves for excluding the resonances of the background and solvent chemicals from the spectrum of the test sample.

3.2 Chemical Shift References

Chemical shift references for calibration of the spectrum scale may be internal or external. An internal chemical shift reference substance is dissolved directly in the NMR sample solution that is measured. In the external method, a separate NMR sample is prepared, containing the reference substance(s) dissolved in the same solvent as used for the test sample. Alternatively, the reference substance may be placed in an insert (capillary) tube, as such (e.g. 85% H_3PO_4) or dissolved in the same solvent as used for the test sample. The insert tube is placed into the NMR sample to be analyzed, or into a separate NMR tube containing only the solvent. The spectrum is then recorded for the reference position.

To avoid overlap of the resonances of the reference substance and chemicals of interest, and a possible overdose of the reference substance, and also from the point of view of optimum magnet homogeneity, we recommend the external chemical shift reference method with use of a separate rather than an insert sample tube.

The external reference method is common in ^{31}P NMR. The reference substance, 85% H_3PO_4 (phosphoric acid), is transferred in a cylindrical or spherical capillary tube to an NMR sample tube containing the same solvent as the test sample. Owing to its chemical reactivity, H_3PO_4 cannot be added as an internal reference substance. Although not an exact method,[20] the resonance of H_3PO_4 is set at 0.00 ppm when the coaxial capillary tube system is used.

The most common reference chemicals in ^{1}H and ^{13}C NMR are tetramethylsilane (TMS; dissolves in organic solvents, $\delta_H = \delta_C = 0.00$ ppm), 3-(trimethylsilyl)-3,3,2,2-tetradeuteropropionic acid sodium salt (TSPA-d_4;

dissolves in water/D_2O, $\delta_H = \delta_C = 0.0$ ppm), and 3-(trimethylsilyl)propane sulfonic acid sodium salt (TSPSA; dissolves in water/D_2O, $\delta_H = 0.015$ ppm, $\delta_C = 0.00$ ppm).[21] A common reference substance for ^{19}F NMR is trichlorofluoromethane ($CFCl_3$; dissolves in organic solvents, $\delta_F = 0.00$ ppm). The fluorine resonance of $CFCl_3$ consists of four lines with intensities (from high frequency down) 8 : 8 : 3 : 1; the second line from the high-frequency end is used as the scale origin.[22]

Other reference substances are available and can also be used. Sometimes the resonance of the actual solvent can serve as a reference. Preferably, the referencing method used for the sample should be the same as that used for the authentic reference sample (or library spectrum), or at least the scale position should be corrected by a known factor. Whatever referencing method is used it needs to be reported in sufficient detail.

3.3 Nuclear Magnetic Resonance Sample Tube

NMR tubes are usually made of high-quality Pyrex glass. The tube is selected according to the size of the probe head, normally 5 or 10 mm (o.d.). Different qualities of tubes are available. Provided the sample solution is of good quality, thin-walled extra-high-quality NMR tubes give the best resolution, line shape, and sensitivity. Routine quality tubes for daily use (e.g. Wilmad 507-PP) are normally adequate, however. Usually the tube is sealed by a plastic cap, but other cap materials are also available. For the sake of safety in working with toxic chemicals, sealing of the tube by melting the top of it by flame may be advisable (an air-tight ampule is formed). Microcylinders and coaxial capillary inserts of different shapes, volumes, and materials are available for small sample amounts, external locking, external referencing, and corrosive materials (see catalogues of NMR tube sellers).

A simple method to remove dust, which even new NMR tubes may contain, employs a 10–20-mL plastic syringe, a 0.45 μm high-performance liquid chromatography (HPLC) filter unit suitable for water, a needle, distilled water, and an oven (at 45–55 °C). Rinse the NMR tube with 1-mL portions of filtered distilled water by shaking while the cap is on, until no dust is seen in the rinsing solution in the tube. Empty the NMR tube and put it open end down in a decanter glass, and into the oven for drying.

3.4 Sample Preparation

In principle, preparation of an NMR sample is simple: take a suitable amount (1–100 mg) of chemical to be analyzed, dissolve it in a selected solvent, add the chemical shift reference substance, and then transfer the sample with simultaneous filtration into an NMR tube. Attach a code label to the top of the NMR tube. General reading on sample preparation is available.[23–25]

The quality of the results in NMR largely depends on the care with which the sample has been prepared. Not only should the NMR sample tube be of good (routine) quality, but also the sample solution in the tube should be free of nondissolved particles and dust.

NMR spectral parameters, i.e. chemical shift and coupling constant (J), may be considerably affected by the sample condition, i.e. solvent, pH, sample temperature, concentration, and choice of internal and/or external chemical shift references. Solvent and pH (in water/D_2O samples) have the greatest effect. The sample condition should therefore be the same as or comparable to that used for the authentic reference chemical (or library spectrum) and the blank sample.

A sample for NMR spectroscopy can be taken from several stages of the sample preparation path [see **Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention** and the literature[14,18]]. Preferably, this would be a 5–10 mL portion of extract or of aqueous or organic liquid. Common to all these solutions is the large molar excess of ^1H in the solvent compared with the amount of ^1H in the possible target chemicals. This yields an intense solvent (e.g. H_2O, CH_2Cl_2) resonance in the ^1H NMR spectrum, making the trace analysis difficult or even impossible. A usual procedure to avoid this problem is to replace the protonated solvent (e.g. H_2O) with the corresponding deuterated solvent (D_2O). Deuterated solvents are also used for the field-frequency lock of the spectrometer.

The solvent is replaced by evaporating the protonated solvent and adding the corresponding deuterated solvent in its place. The sample is concentrated during the evaporation; however, one must be alert to the danger of losing volatile chemicals at the same time. The sample solution is then transferred to an NMR tube and is ready for the experiments. Because NMR spectroscopy is a nondestructive technique, the sample can be used after the NMR experiments for other analyses.

Except for extraction, no separation methods have been used in the NMR sample preparations in interlaboratory comparison tests.[7–14] Thus, the chemicals of interest, impurities, and background present in the original sample and in the extract were all present in the NMR sample. The advantage of this is that, generally, not much or nothing is lost, while the disadvantage is overlapping of the resonances of interest with resonances of other chemicals.

3.4.1 Preparation of Nuclear Magnetic Resonance Samples from Aqueous or Organic Liquid or Extract

A portion of aqueous or organic liquid or extract reserved for NMR spectroscopy is filtered, if necessary,

Sarin

[107-44-8]

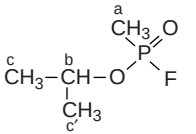

^{1}H–NMR

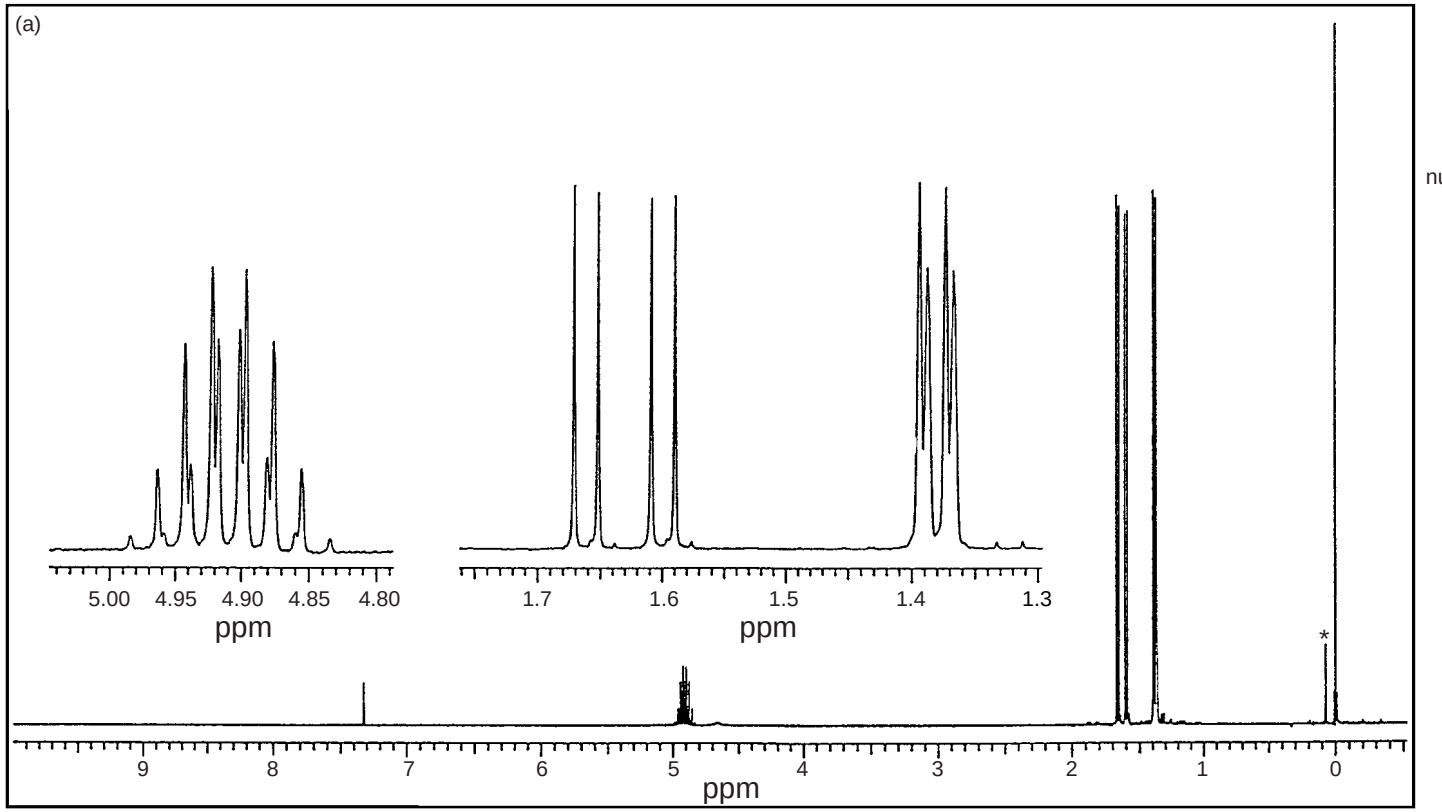

nuc.	δ (ppm)
a	1.63
b	4.91
c,c′	1.37, 1.38
J(a,P)	18.7 Hz
J(a,F)	5.7 Hz
J(b,P)	7.6 Hz
J(b,c)	6.2 Hz

Res. freq.	299.943 MHz	Spectral width:	4000 Hz	Rep. time:	4.20 s
Solvent:	$CDCl_3$	Data points (FID):	33600	Line broad.:	not used
Temperature:	24 °C	Data points (spec.):	32 K	Resolution:	0.39 Hz (TMS)
Concentration:	100 mg/ml	Pulse width:	7 µs (24°)	Instrument:	Varian VXR-300
Reference:	TMS (internal)	Number of scans:	12	Filename:	SA2511/94 Switzerland

Figure 2 (a) ^{1}H, (b) ^{31}P{^{1}H}, (c) ^{31}P, (d) ^{19}F, and (e) ^{13}C{^{1}H} NMR spectra of authentic sarin (GB) in $CDCl_3$ from the AC-Laboratory, Switzerland.

Sarin

[107-44-8]

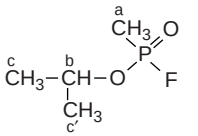

$^{31}P\{H\}$–NMR

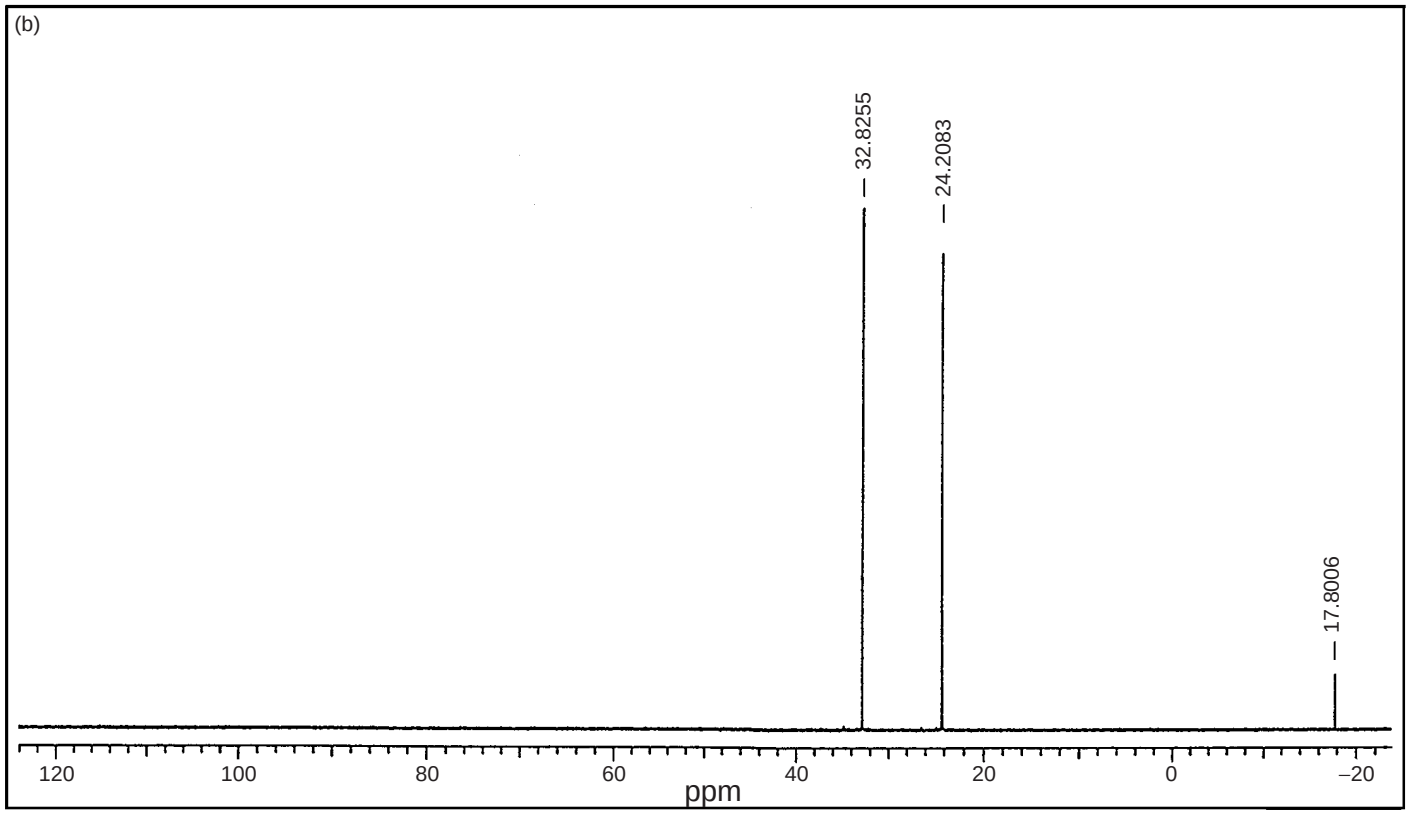

nuc.	δ (ppm)
P	28.52
J(P,F)	1046.2 Hz

Res. freq.	121.421 MHz	Spectral width:	17986 Hz	Rep. time:	1.82 s
Solvent:	$CDCl_3$	Data points (FID):	65472	Line broad.:	0.5 Hz
Temperature:	24 °C	Data points (spec.):	32 K	Resolution:	1.01 Hz (TPP), 1.75 Hz (P)
Concentration:	20 mg/ml	Pulse width:	5.5 µs (45°)	Instrument:	Varian VXR-300
Reference:	Triphenylphosphate (TPP) (–17.8 ppm, internal)	Number of scans:	84	Filename:	SA2611/94 Switzerland

Figure 2 (*Continued)*

Sarin

[107-44-8]

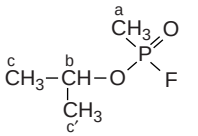

^{31}P–NMR

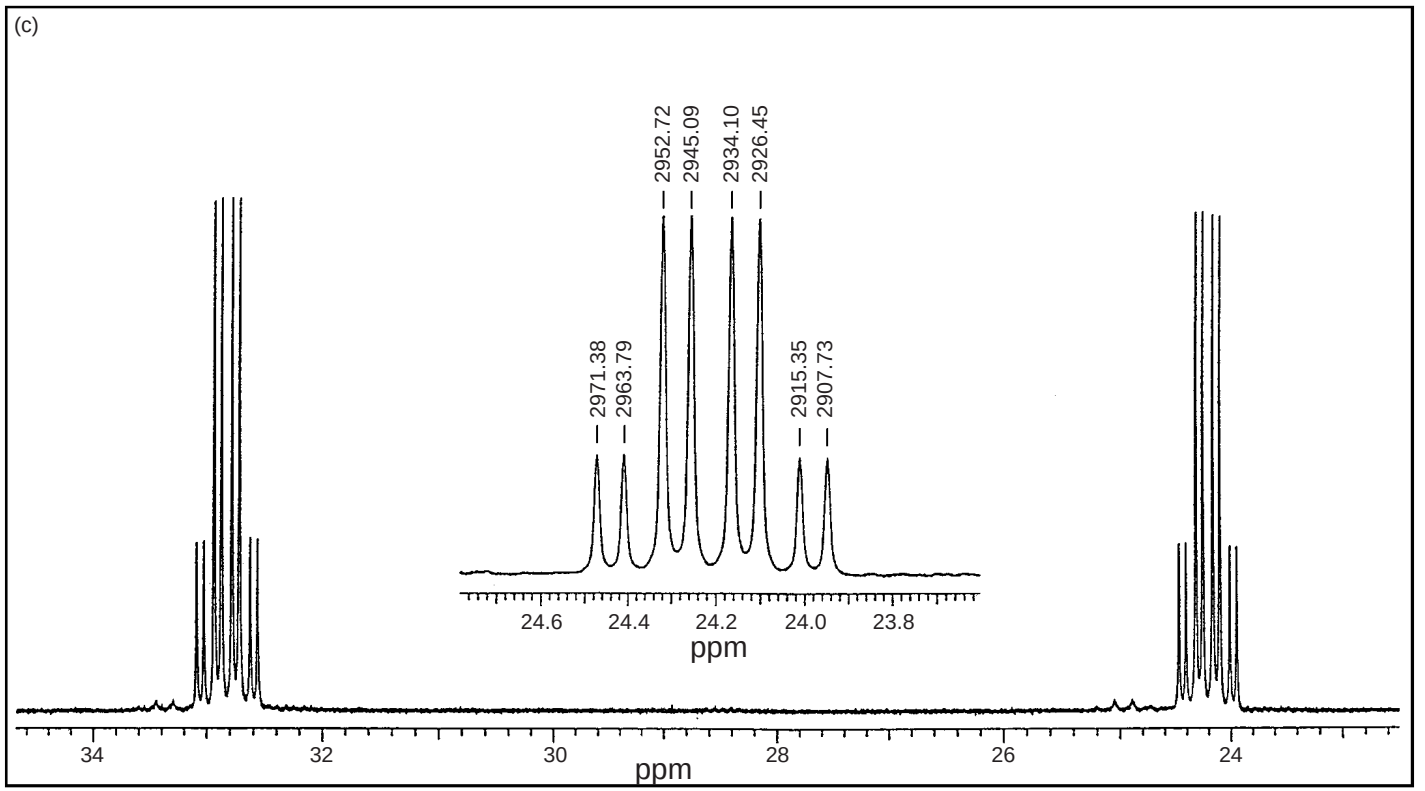

J(P,F)	1046.2 Hz
J(P,a)	18.7 Hz
J(P,b)	7.6 Hz

Res. freq.	121.421 MHz	Spectral width:	2564 Hz	Rep. time:	11.71 s
Solvent:	$CDCl_3$	Data points (FID):	60032	Line broad.:	not used
Temperature:	24 °C	Data points (spec.):	32 K	Resolution:	
Concentration:	100 mg/ml	Pulse width:	5.5 µs (45°)	Instrument:	Varian VXR-300
Reference:	Triphenylphosphate (–17.8 ppm, external)	Number of scans:	239	Filename:	SA2513/94 Switzerland

Figure 2 (*Continued)*

Sarin

[107-44-8]

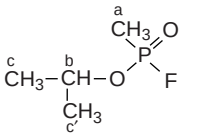

^{19}F–NMR

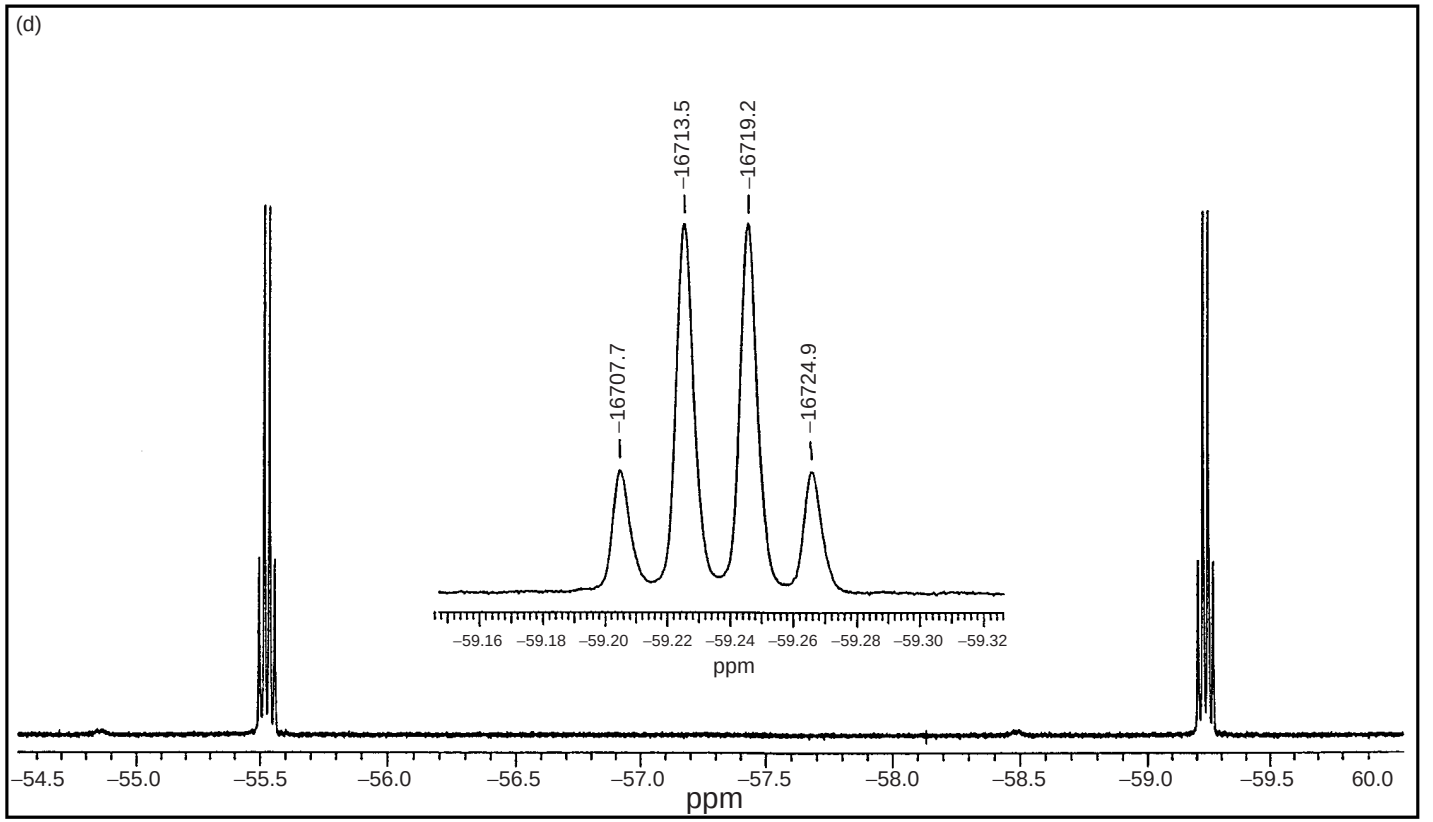

nuc.	δ [ppm]
F	–57.38

J(F,P)	1046.4 Hz
J(F,a)	5.7 Hz

Res. freq.	282.203 MHz	Spectral width:	3315 Hz	Rep. time:	9.06 s
Solvent:	$CDCl_3$	Data points (FID):	60032	Line broad.:	not used
Temperature:	24 °C	Data points (spec.):	32 K	Resolution:	
Concentration:	20 mg/ml	Pulse width:	11.0 µs (45°)	Instrument:	Varian VXR-300
Reference:	Fluorobenzene (–113.59 ppm, external)	Number of scans:	159	Filename:	SA2613/94 Switzerland

Figure 2 (*Continued)*

Sarin

[107-44-8]

$^{13}C\{H\}$–NMR

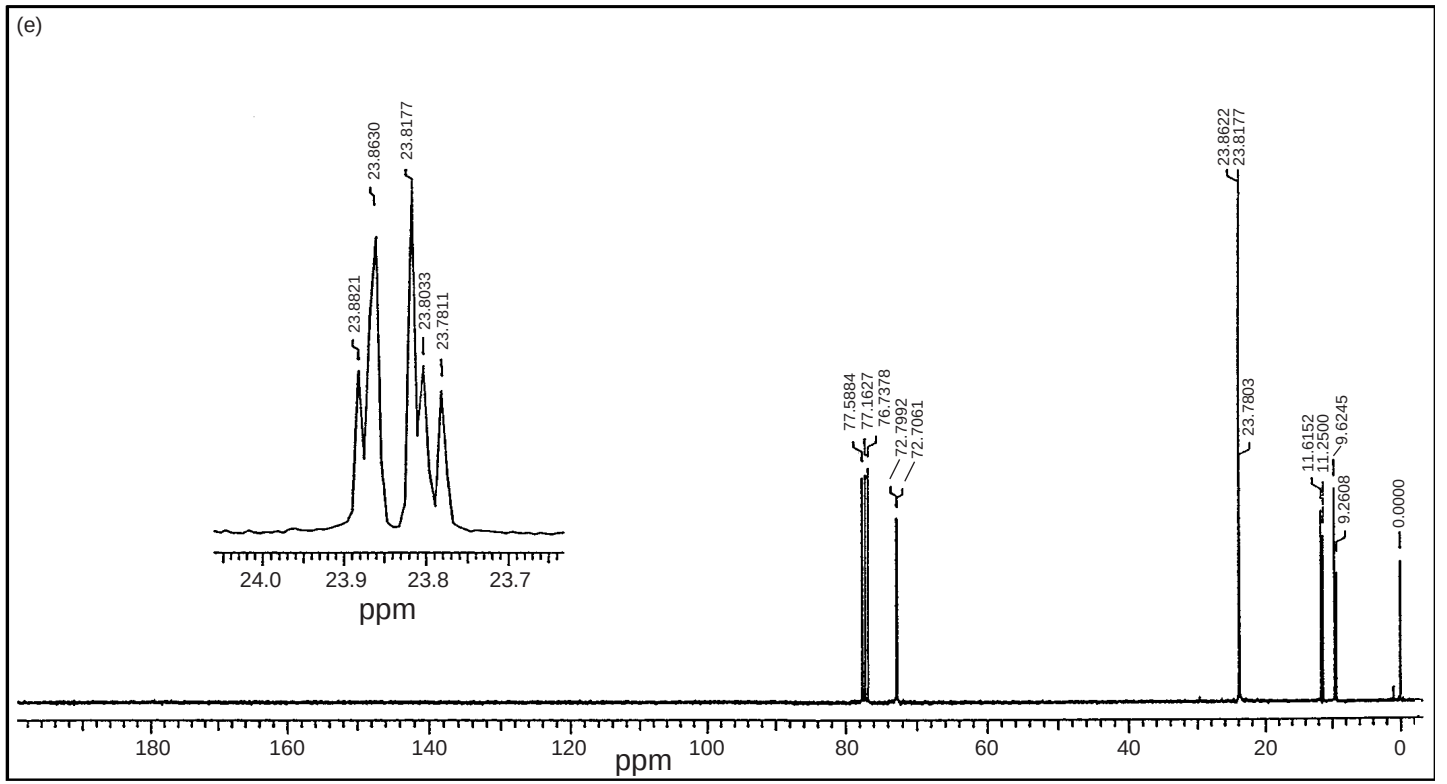

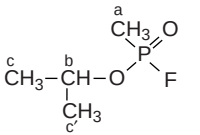

nuc.	δ (ppm)
a	10.44
b	72.73
c,c′	23.83 23.84

J(a,P)	150.2 Hz
J(a,F)	27.5 Hz
J(b,P)	7.0 Hz
J(c,P)	5.9 Hz
J(c′P)	3.6 Hz
J(c,F)	1.7 Hz

Res. freq.	75.429 MHz	Spectral width:	17699 Hz	Rep. time:	1.85 s
Solvent:	$CDCl_3$	Data points (FID):	64 K	Line broad.:	0.5 Hz
Temperature:	24 °C	Data points (spec.):	32 K	Resolution:	0.93 Hz (TMS)
Concentration:	100 mg/ml	Pulse width:	5.6 µs (38°)	Instrument:	Varian VXR-300
Reference:	TMS (internal)	Number of scans:	6806	Filename:	SA2512/94 Switzerland

Figure 2 (*Continued)*

^{1}H NMR

Ethyl N,N-dimethylphosphoramidocyanidate (Tabun)

[77-81-6]

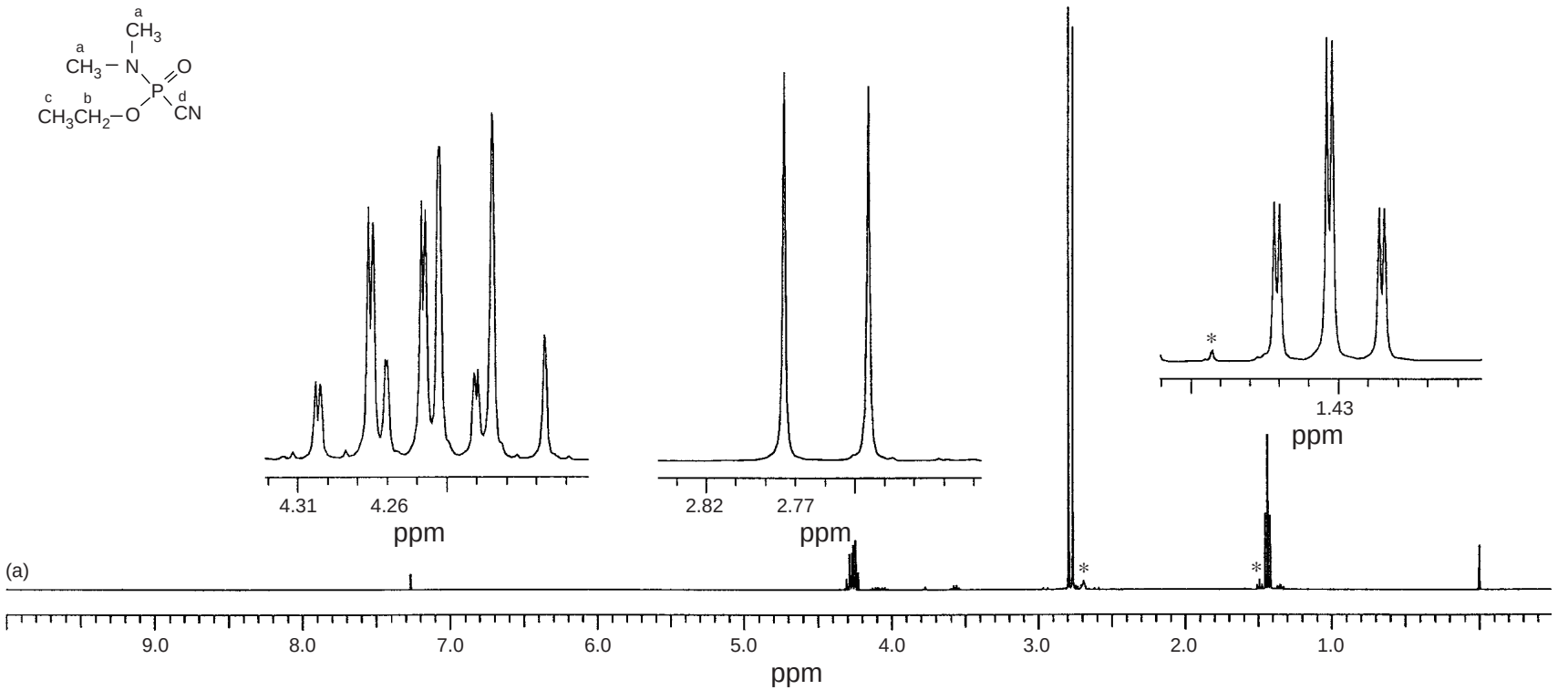

				nuc.	δ [ppm]	J [Hz]
Solvent:	$CDCl_3$	Flip angle; Pulse width:	45 °; 6.2 µs	a	2.779	11.3 (P)
Concentration:	16.4 mg/1.1 ml	No. of scans; Rep. time:	32; 19.5 s	b	4.268	–10.2 (b′); 7.1 (c); 9.8 (P)
Reference substance:	TMS	Weighting; Line broad:	exp.; 0.1 Hz	b′	4.261	7.0 (c); 8.6 (P)
Sample temperature:	29.5 °C	Spectral resolution:	0.3 Hz (TMS)	c	1.433	0.7 (P)
Resonance frequency:	400.136 MHz	Instrument:	Bruker AMX-400			
Spectral width:	4629.6 Hz	Source reference:	930209-429			
Data points (FID; spec.):	32 k; 32 k					

Figure 3 (a) ^{1}H, (b) ^{31}P{^{1}H}, (c) ^{31}P, and (d) ^{13}C{^{1}H} NMR spectra of authentic tabun (GA) in $CDCl_3$ from VERIFIN, Finland.

$^{31}P\{^{1}H\}$ NMR

Ethyl N,N-dimethylphosphoramidocyanidate (Tabun)

[77-81-6]

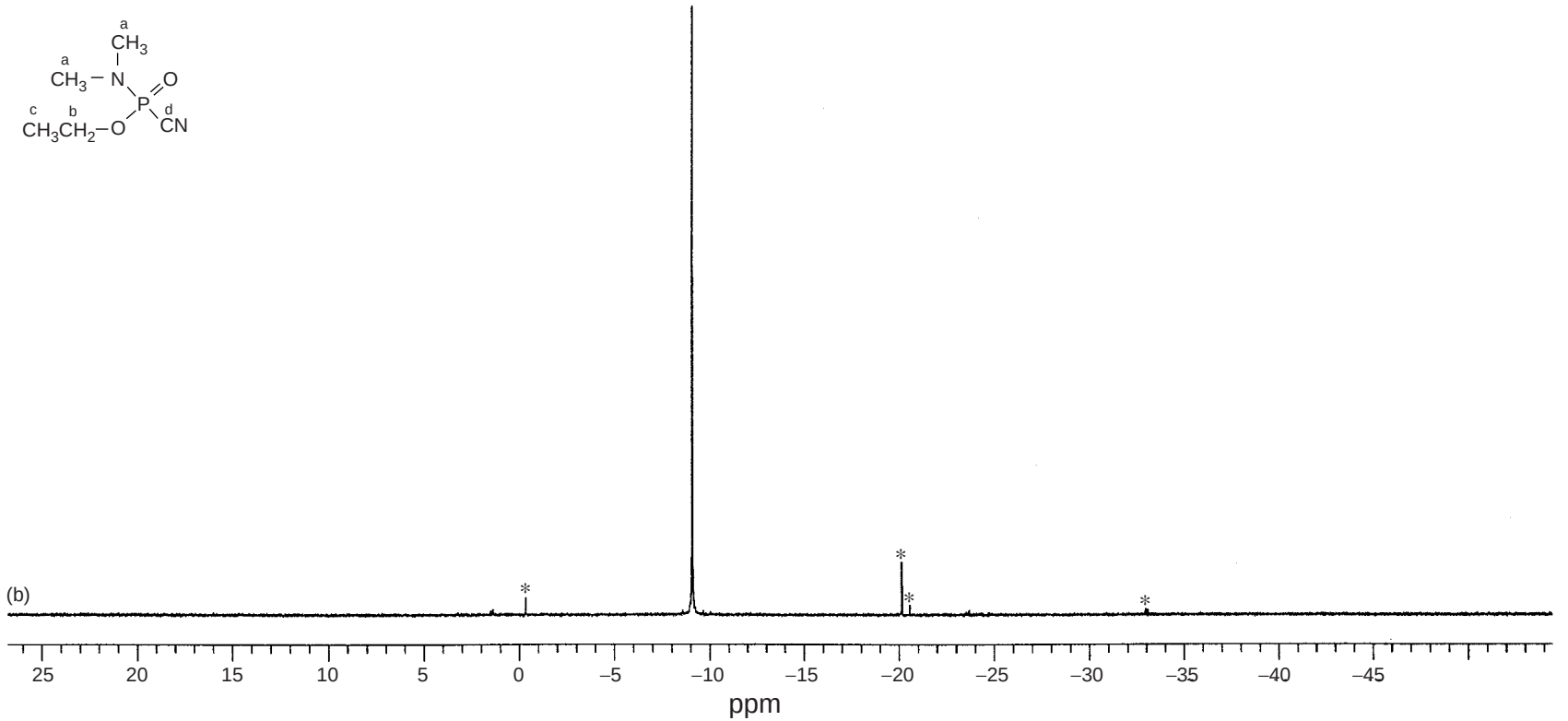

Solvent:	$CDCl_3$	Flip angle; Pulse width:	45 °; 4.6 µs	nuc.	δ [ppm]
Concentration:	16.4 mg/1.1 ml	No. of scans; Rep. time:	64; 18.5 s	P	–9.10
Reference substance:	ext. $P(OH)_4ClO_4$	Weighting; Line broad:	exp.; 1.0 Hz		
Sample temperature:	29.5 °C	Spectral resolution:	3.0 Hz (P)		
Resonance frequency:	161.975 MHz	Instrument:	Bruker AMX-400		
Spectral width:	13157.5 Hz	Source reference:	930209-427		
Data points (FID; spec.):	64 k; 64 k				

$\delta(P(OH)_4ClO_4) = 0.66$ vs. $\delta(H_3PO_4) = 0.00$

Figure 3 (*Continued)*

Ethyl N,N-dimethylphosphoramidocyanidate (Tabun)

[77-81-6]

^{31}P NMR

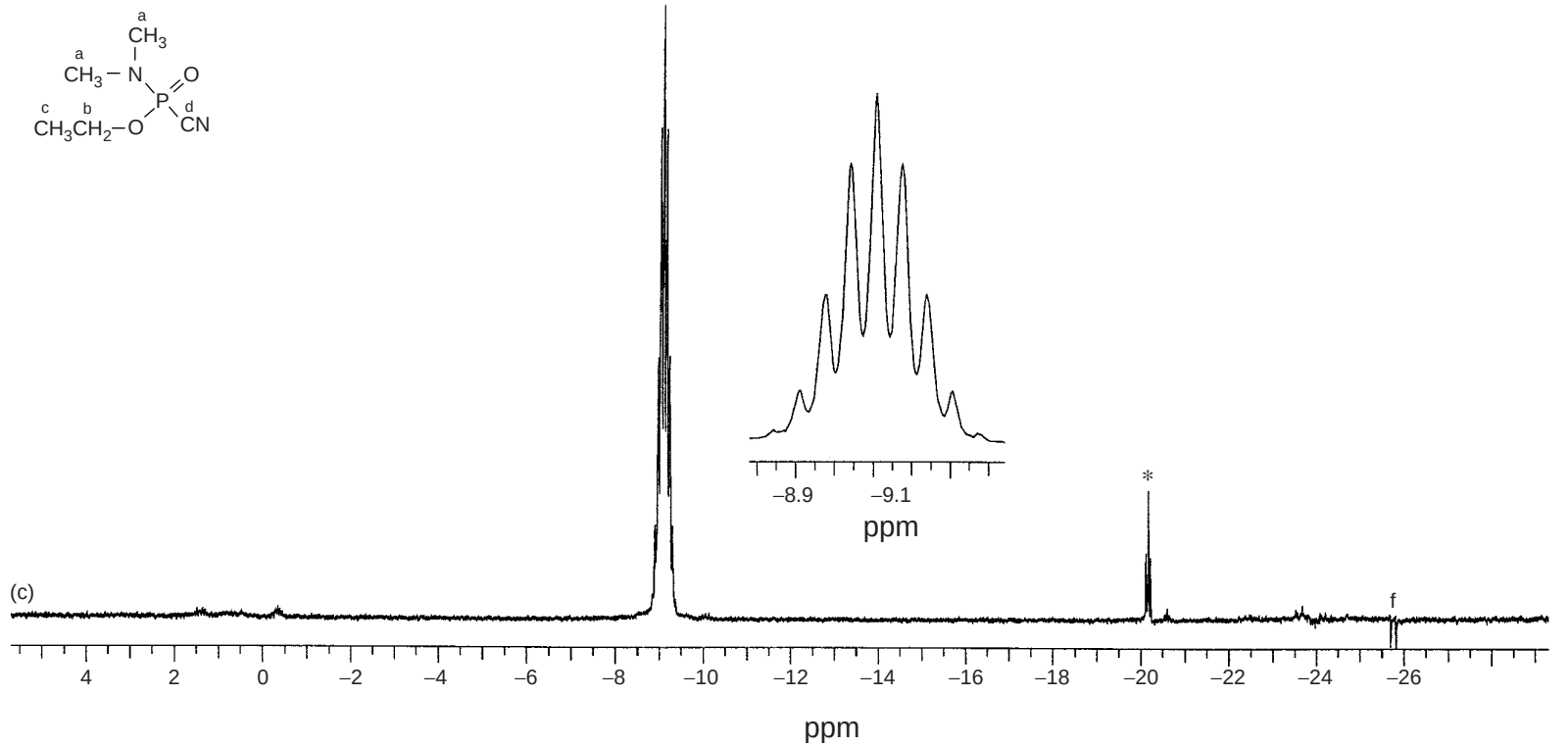

Solvent:	$CDCl_3$	Flip angle; Pulse width:	45 °; 4.6 µs	nuc. δ [ppm]	
Concentration:	16.4 mg/1.1 ml	No. of scans; Rep. time:	256; 7.8 s	P –9.10	For couplings see the ^{1}H spectrum
Reference substance:	ext. $P(OH)_4ClO_4$	Weighting; Line broad:	exp.; 0.5 Hz		
Sample temperature:	29.5 °C	Spectral resolution:	7.0 Hz (P)	f = folded impurity resonance	
Resonance frequency:	161.975 MHz	Instrument:	Bruker AMX-400	$\delta(P(OH)_4ClO_4) = 0.66$ vs. $\delta(H_3PO_4) = 0.00$	
Spectral width:	5681.8 Hz	Source reference:	930209-428		
Data points (FID; spec.):	64 k; 32 k				

Figure 3 (*Continued)*

$^{13}C\{^{1}H\}$ NMR

Ethyl N,N-dimethylphosphoramidocyanidate (Tabun)

[77-81-6]

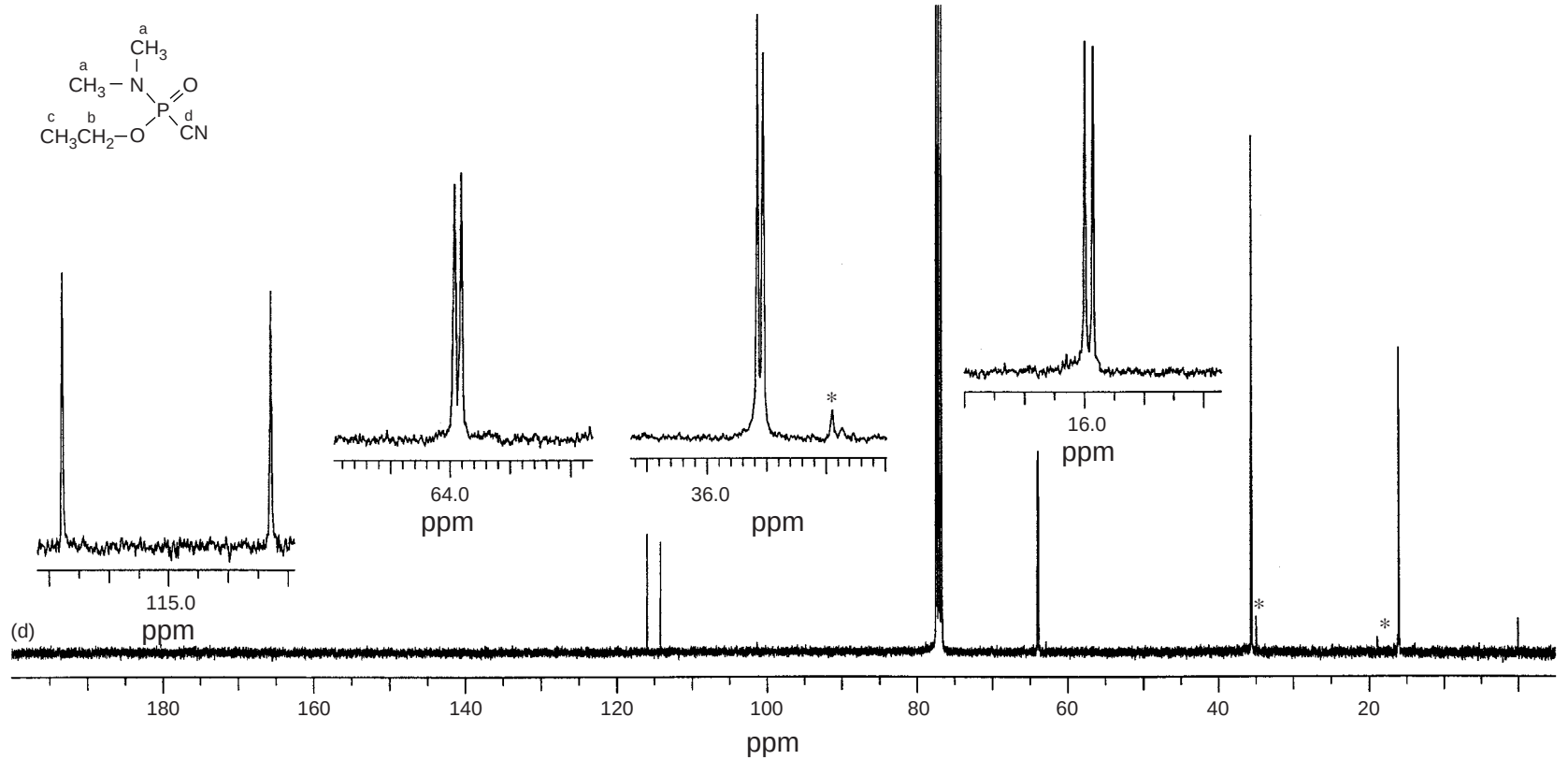

				nuc.	δ [ppm]	J [Hz}
Solvent:	$CDCl_3$	Flip angle; Pulse width:	33 °; 2.6 µs	a	35.55	4.9 (P)
Concentration:	16.4 mg/1.1 ml	No. of scans; Rep. time:	13612; 3.7 s	b	63.93	5.5 (P)
Reference substance:	TMS	Weighting; Line broad:	exp.; 0.5 Hz	c	15.96	6.8 (P)
Sample temperature:	29.5 °C	Spectral resolution:	1.0 Hz (TMS)	d	115.01	175.5 (P)
Resonance frequency:	100.625 MHz	Instrument:	Bruker AMX-400			
Spectral width:	23809.5 Hz	Source reference:	930620-710			
Data points (FID; spec.):	64 k; 64 k					

Figure 3 (*Continued)*

Table 1 NMR spectral parameters of selected CWC-related chemicals

Trivial name; NATO code[a]; IUPAC name [CAS][b] (Schedule)[c]	
Structure	Chemical shift δ (ppm)[d] (*multiplicity*)[e], coupling constant *J* (Hz)[f]; concentration and solvent, pH; chemical shift reference(s); remarks.
Sarin; GB; isopropyl methylphosphonofluoridate [107-44-8] (1.A.1)	
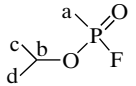	$\delta_{Ha} = 1.623(dd)$, $\delta_{Hb} = 4.908(dspd)$, $\delta_{Hc} = 1.375(d)$, $\delta_{Hd} = 1.382(d)$, $^2J_{Ha,P} = 18.7$, $^3J_{Ha,F} = 5.7$, $^3J_{Hb,(Hc,Hd)} = 6.2$, $^3J_{Hb,P} = 7.7$, $^4J_{Hb,F} = 0.4$, $\delta_{Ca} = 10.47(dd)$, $\delta_{Cb} = 72.73(d)$, $\delta_{Cc} = 23.840(dd)$, $\delta_{Cd} = 23.845(d)$, $^1J_{Ca,P} = 150.1$, $^2J_{Ca,F} = 27.5$, $^2J_{Cb,P} = 7.0$, $^3J_{Cc,P} = 5.9$, $^4J_{Cd,F} = 1.7$, $^3J_{Cd,P} = 3.5$, $\delta_F = -57.43(dq)$, $^1J_{F,P} = 1046.3$, $\delta_P = 29.15(ddq)$; 14.1 mg/1.1 mL $CDCl_3$; TMS, $CFCl_3$, ext. H_3PO_4
Soman; GD; pinacolyl methylphosphonofluoridate [96-64-0] (1.A.1)	
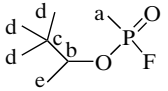	More abundant diastereomer: $\delta_{Ha} = 1.633(dd)$, $\delta_{Hb} = 4.487(ddq)$, $\delta_{Hd} = 0.936(s)$, $\delta_{He} = 1.317(d)$, $^2J_{Ha,P} = 18.7$, $^3J_{Ha,F} = 5.7$, $^3J_{Hb,P} = 7.9$, $^4J_{Hb,F} = 0.5$, $^3J_{Hb,He} = 6.4$, $\delta_{Ca} = 10.05(dd)$, $\delta_{Cb} = 83.23(d)$, $\delta_{Cc} = 34.89(dd)$, $\delta_{Cd} = 25.43(s)$, $\delta_{Ce} = 16.93(s)$, $^1J_{Ca,P} = 151.3$, $^2J_{Ca,F} = 28.2$, $^2J_{Cb,P} = 8.1$, $^3J_{Cc,P} = 7.2$, $^4J_{Cc,F} = 1.5$, $^3J_{Ce,P} = 0.9$, $\delta_F = -55.48(dq)$, $^1J_{F,P} = 1047.2$, $\delta_P = 29.75(ddq)$ Less abundant diastereomer: $\delta_{Ha} = 1.636(dd)$, $\delta_{Hb} = 4.438(ddq)$, $\delta_{Hd} = 0.936(s)$, $\delta_{He} = 1.328(d)$, $^2J_{Ha,P} = 18.7$, $^3J_{Ha,F} = 5.7$, $^3J_{Hb,P} = 8.9$, $^3J_{Hb,He} = 6.4$, $\delta_{Ca} = 10.63(dd)$, $\delta_{Cb} = 83.73(d)$, $\delta_{Cc} = 34.80(dd)$, $\delta_{Cd} = 25.43(s)$, $\delta_{Ce} = 17.06(s)$, $^1J_{Ca,P} = 150.7$, $^2J_{Ca,F} = 28.5$, $^2J_{Cb,P} = 7.9$, $^3J_{Cc,P} = 5.5$, $^3J_{Ce,P} = 2.0$, $^4J_{Ce,F} = 2.0$, $\delta_F = -58.33(dq)$, $^1J_{F,P} = 1046.4$, $\delta_P = 28.73(ddq)$; 75.5 mg/1.1 mL $CDCl_3$; TMS, $CFCl_3$, ext. H_3PO_4; molar ratio of diastereomers: approx. 55/45
Tabun; GA; *O*-ethyl *N*,*N*-dimethylphosphoramidocyanidate [77-81-6] (1.A.2)	
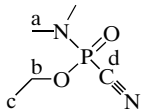	$\delta_{Ha} = 2.779(d)$, $\delta_{Hb} = 4.261(m)$, $\delta_{Hb'} = 4.269(m)$, $\delta_{Hc} = 1.433(dt)$, $^3J_{Ha,P} = 11.3$, $^2J_{Hb,Hb'} = -10.2$, $^3J_{(Hb,Hb'),Hc} = 7.1$, $^3J_{Hb,P} = 8.6$, $^3J_{Hb',P} = 9.8$, $^4J_{Hc,P} = 0.7$, $\delta_{Ca} = 35.55(d)$, $\delta_{Cb} = 63.93(d)$, $\delta_{Cc} = 15.96(d)$, $\delta_{Cd} = 115.01(d)$, $^2J_{Ca,P} = 4.9$, $^2J_{Cb,P} = 5.5$, $^3J_{Cc,P} = 6.8$, $\delta_P = -9.02$; 16.4 mg/1.1 mL $CDCl_3$; TMS, ext. H_3PO_4
VX; *O*-ethyl *S*-2-diisopropylaminoethyl methylphosphonothiolate [50782-69-9] (1.A.3)	
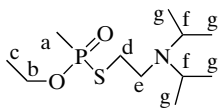	$\delta_{Ha} = 1.783(d)$, $\delta_{Hb} = 4.094(ddq)$, $\delta_{Hb'} = 4.187(ddq)$, $\delta_{Hc} = 1.347(t)$, $\delta_{Hd} = 2.799(m)$, $\delta_{Hd'} = 2.834(m)$, $\delta_{He} = 2.684(m)$, $\delta_{He'} = 2.703(m)$, $\delta_{Hf} = 3.011(sp)$, $\delta_{Hg} = 1.018(d)$, $^2J_{Ha,P} = 15.6$, $^3J_{Hb,P} = 9.3$, $^2J_{Hb,Hb'} = -10.2$, $^3J_{(Hb,Hb'),Hc} = 7.1$, $^3J_{Hb',P} = 8.4$, $^3J_{Hd,P} = 12.4$, $^2J_{Hd,Hd'} = -12.5$, $^3J_{Hd,He} = 10.0$, $^3J_{Hd,He'} = 5.6$, $^3J_{Hd',P} = 12.6$, $^3J_{Hd',He} = 5.7$, $^3J_{Hd',He'} = 10.0$, $^2J_{He,He'} = -14.1$, $^3J_{Hf,Hg'} = 6.6$, $\delta_{Ca} = 20.06(d)$, $\delta_{Cb} = 61.15(d)$, $\delta_{Cc} = 16.26(d)$, $\delta_{Cd} = 31.62(d)$, $\delta_{Ce} = 46.62(d)$, $\delta_{Cf} = 48.92(s)$, $\delta_{Cg} = 20.94(s)$, $20.99(s)$, $^1J_{Ca,P} = 110.4$, $^2J_{Cb,P} = 6.9$, $^3J_{Cc,P} = 7.1$, $^2J_{Cd,P} = 2.8$, $^4J_{Ce,P} = 4.6$, $\delta_P = 54.80(m)$; 50 µL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4; 2 µL of tributylamine was added
Mustard gas; HD; bis(2-chloroethyl) sulfide [505-60-2] (1.A.4)	
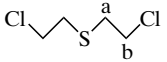	$\delta_{Ha} = 2.923(XX')$, $\delta_{Hb} = 3.645(AA')$, $\delta_{Hc} = 2.802(s)$, $^2J_{Ha,Ha'} = -10.9$, $^2J_{Hb,Hb'} = -13.9$, $^3J_{Ha,Hb} = 9.5$, $^3J_{Ha,Hb'} = 5.9$, $\delta_{Ca} = 34.61$, $\delta_{Cb} = 43.04$; 17.0 mg/1.0 mL $CDCl_3$; TMS
Sesquimustard; Q; bis(2-chloroethylthio)ethane [3563-36-8] (1.A.4)	
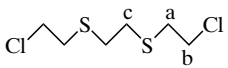	$\delta_{Ha} = 2.907(XX')$, $\delta_{Hb} = 3.645(AA')$, $^2J_{Ha,Ha'} = -10.9$, $^2J_{Hb,Hb'} = -13.8$, $^3J_{Ha,Hb} = 9.6$, $^3J_{Ha,Hb'} = 5.9$, $\delta_{Ca} = 34.50$, $\delta_{Cb} = 43.12$, $\delta_{Cc} = 32.65$; 21.3 mg/1.1 mL $CDCl_3$; TMS
***O*-Mustard; T; bis(2-chloroethylthioethyl) ether [63918-89-8] (1.A.4)**	
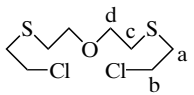	$\delta_{Ha} = 2.924(XX')$, $\delta_{Hb} = 3.653(AA')$, $\delta_{Hc} = 2.762(t)$, $\delta_{Hd} = 3.653(t)$, $^2J_{Ha,Ha'} = -10.8$, $^2J_{Hb,Hb'} = -13.8$, $^3J_{Ha,Hb} = 9.8$, $^3J_{Ha,Hb'} = 5.8$, $^3J_{Hc,Hd} = 6.4$, $\delta_{Ca} = 34.82$, $\delta_{Cb} = 43.16$, $\delta_{Cc} = 32.04$, $\delta_{Cd} = 71.08$; 26.2 mg/1.0 mL $CDCl_3$; TMS

Lewisite-1; L; 2-chlorovinyldichloroarsine [541-25-3] (1.A.5)

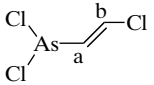

$\delta_{Ha} = 7.146(d)$, $\delta_{Hb} = 6.936(d)$, $^3J_{Ha,Hb} = 14.8$, $\delta_{Ca} = 139.27$, $\delta_{Cb} = 133.17$; 22.3 mg/1.1 mL $CDCl_3$; TMS

Nitrogen mustard one; HN-1; bis(2-chloroethyl)ethylamine [538-07-8] (1.A.6)

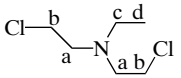

$\delta_{Ha} = 2.865(XX')$, $\delta_{Hb} = 3.507(XX')$, $\delta_{Hc} = 2.654(q)$, $\delta_{Hd} = 1.051(t)$, $^3J_{Hc,Hd} = 7.1$, $\delta_{Ca} = 55.97$, $\delta_{Cb} = 42.08$, $\delta_{Cc} = 48.52$, $\delta_{Cd} = 12.35$; 30 µL/0.8 mL $CDCl_3$; TMS

DF; methylphosphonyl difluoride [676-99-3] (1.B.9)

$\delta_{Ha} = 1.886(dt)$, $^2J_{Ha,P} = 19.4$, $^3J_{Ha,F} = 5.8$, $\delta_{Ca} = 8.49(dt)$, $^1J_{Ca,P} = 148.4$, $^2J_{Cb,F} = 22.5$, $\delta_F = -59.57(dq)$, $^1J_{F,P} = 1109.8$, $\delta_P = 24.47(tq)$; 20 µL/1.0 mL $CDCl_3$; TMS, $CFCl_3$, ext. H_3PO_4

Chlorosarin; isopropyl methylphosphonochloridate [1445-76-7] (1.B.11)

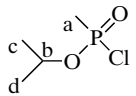

$\delta_{Ha} = 1.955(d)$, $\delta_{Hb} = 4.942(dsp)$, $\delta_{Hc/Hd} = 1.399(d)$, $\delta_{Hd/Hc} = 1.413(d)$, $^2J_{Ha,P} = 17.6$, $^3J_{Hb,(Hc,Hd)} = 6.2$, $^3J_{Hb,P} = 10.5$, $\delta_{Ca} = 20.71(d)$, $\delta_{Cb} = 73.35(d)$, $\delta_{Cc/Cd} = 23.39(d)$, $\delta_{Cd/Cc} = 24.01(d)$, $^1J_{Ca,P} = 130.8$, $^2J_{Cb,P} = 8.0$, $^3J_{Cc/Cd,P} = 5.0$, $^3J_{Cd/Cc,P} = 4.7$, $\delta_P = 39.21(dq)$; 50 µL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4

Isopropyl methylphosphonate [1832-54-8] (2.B.4)

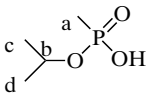

$\delta_{Ha} = 1.480(d)$, $\delta_{Hb} = 4.671(dsp)$, $\delta_{Hc} = \delta_{Hd} = 1.334(d)$, $^2J_{Ha,P} = 17.9$, $^3J_{Hb,(Hc,Hd)} = 6.2$, $^3J_{Hb,P} = 8.3$, $\delta_{Ca} = 12.37(d)$, $\delta_{Cb} = 70.08(d)$, $\delta_{Cc} = \delta_{Cd} = 23.98(d)$, $^1J_{Ca,P} = 148.2$, $^2J_{Cb,P} = 6.5$, $^3J_{(Cc,Cd),P} = 4.4$, $\delta_P = 33.40(dq)$; 5 µL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4

$\delta_{Ha} = 1.552(d)$, $\delta_{Hb} = 4.659(dsp)$, $\delta_{Hc} = \delta_{Hd} = 1.326(d)$, $^2J_{Ha,P} = 17.4$, $^3J_{Hb,(Hc,Hd)} = 6.2$, $^3J_{Hb,P} = 8.2$, $\delta_{Ca} = 13.79(d)$, $\delta_{Cb} = 74.41(d)$, $\delta_{Cc} = \delta_{Cd} = 25.83(d)$, $^1J_{Ca,P} = 139.0$, $^2J_{Cb,P} = 6.0$, $^3J_{(Cc,Cd),P} = 4.1$, $\delta_P = 32.21(dq)$; 10 µL/0.8 mL D_2O, pH 0; TSPSA, ext. H_3PO_4

$\delta_{Ha} = 1.275(d)$, $\delta_{Hb} = 4.442(dsp)$, $\delta_{Hc} = \delta_{Hd} = 1.252(d)$, $^2J_{Ha,P} = 16.4$, $^3J_{Hb,(Hc,Hd)} = 6.2$, $^3J_{Hb,P} = 8.6$, $\delta_{Ca} = 14.97(d)$, $\delta_{Cb} = 71.25(d)$, $\delta_{Cc} = \delta_{Cd} = 26.15(d)$, $^1J_{Ca,P} = 137.2$, $^2J_{Cb,P} = 5.3$, $^3J_{(Cc,Cd),P} = 3.7$, $\delta_P = 26.22(dq)$; 10 µL/0.8 mL D_2O, pH 7.6; TSPSA, ext. H_3PO_4

$\delta_{Ha} = 1.275(d)$, $\delta_{Hb} = 4.436(dsp)$, $\delta_{Hc} = \delta_{Hd} = 1.252(d)$, $^2J_{Ha,P} = 16.4$, $^3J_{Hb,(Hc,Hd)} = 6.2$, $^3J_{Hb,P} = 8.6$, $\delta_{Ca} = 14.96(d)$, $\delta_{Cb} = 71.24(d)$, $\delta_{Cc} = \delta_{Cd} = 26.13(d)$, $^1J_{Ca,P} = 136.9$, $^2J_{Cb,P} = 5.5$, $^3J_{(Cc,Cd),P} = 3.8$, $\delta_P = 26.22(dq)$; 10 µL/0.8 mL D_2O, pH 14; TSPSA, ext. H_3PO_4

Pinacolyl methylphosphonate [616-52-4] (2.B.4)

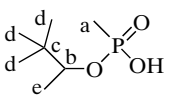

$\delta_{Ha} = 1.481(d)$, $\delta_{Hb} = 4.199(dq)$, $\delta_{Hd} = 0.919(s)$, $\delta_{He} = 1.286(d)$, $^2J_{Ha,P} = 17.9$, $^3J_{Hb,P} = 8.9$, $^3J_{Hb,He} = 6.4$, $^4J_{He,P} = 0.3$, $\delta_{Ca} = 12.36(d)$, $\delta_{Cb} = 80.89(d)$, $\delta_{Cc} = 34.88(d)$, $\delta_{Cd} = 25.58(s)$, $\delta_{Ce} = 16.88(s)$, $^1J_{Ca,P} = 149.6$, $^2J_{Cb,P} = 7.7$, $^3J_{Cc,P} = 6.3$, $^3J_{Ce,P} = 1.1$, $\delta_P = 33.52(dq)$; 5 µL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4

$\delta_{Ha} = 1.562(d)$, $\delta_{Hb} = 4.188(dq)$, $\delta_{Hd} = 0.918(s)$, $\delta_{He} = 1.277(d)$, $^2J_{Ha,P} = 17.4$, $^3J_{Hb,P} = 8.6$, $^3J_{Hb,He} = 6.4$, $\delta_{Ca} = 13.87(d)$, $\delta_{Cb} = 85.01(d)$, $\delta_{Cc} = 36.93(d)$, $\delta_{Cd} = 27.46(s)$, $\delta_{Ce} = 19.03(d)$, $^1J_{Ca,P} = 149.6$, $^2J_{Cb,P} = 7.3$, $^3J_{Cc,P} = 6.1$, $^3J_{Ce,P} = 1.2$, $\delta_P = 31.50(dq)$; 21 mg/0.7 mL D_2O, pH 0; TSPSA, ext. H_3PO_4

$\delta_{Ha} = 1.275(d)$, $\delta_{Hb} = 3.959(dq)$, $\delta_{Hd} = 0.895(s)$, $\delta_{He} = 1.198(d)$, $^2J_{Ha,P} = 16.4$, $^3J_{Hb,P} = 9.1$, $^3J_{Hb,He} = 6.4$, $\delta_{Ca} = 15.28(d)$, $\delta_{Cb} = 82.14(d)$, $\delta_{Cc} = 36.95(d)$, $\delta_{Cd} = 27.81(s)$, $\delta_{Ce} = 19.20(d)$, $^1J_{Ca,P} = 138.0$, $^2J_{Cb,P} = 6.8$, $^3J_{Cc,P} = 5.3$, $^3J_{Ce,P} = 1.0$, $\delta_P = 25.97(dq)$; 7.6 mg/0.7 mL D_2O, pH 7.6; TSPSA, ext. H_3PO_4

$\delta_{Ha} = 1.275(d)$, $\delta_{Hb} = 3.954(dq)$, $\delta_{Hd} = 0.894(s)$, $\delta_{He} = 1.197(d)$, $^2J_{Ha,P} = 16.4$, $^3J_{Hb,P} = 9.1$, $^3J_{Hb,He} = 6.4$, $\delta_{Ca} = 15.27(d)$, $\delta_{Cb} = 82.11(d)$, $\delta_{Cc} = 36.93(d)$, $\delta_{Cd} = 27.80(s)$, $\delta_{Ce} = 19.20(d)$, $^1J_{Ca,P} = 137.8$, $^2J_{Cb,P} = 6.8$, $^3J_{Cc,P} = 5.5$, $^3J_{Ce,P} = 1.1$, $\delta_P = 25.94(dq)$; 10.9 mg/0.8 mL D_2O, pH 14; TSPSA, ext. H_3PO_4

(*continued overleaf*)

Table 1 (*continued*)

Trivial name; NATO code[a]; IUPAC name [CAS][b] (Schedule)[c]	
Structure	Chemical shift δ (ppm)[d] (*multiplicity*)[e], coupling constant *J* (Hz)[f]; concentration and solvent, pH; chemical shift reference(s); remarks.
Methylphosphonic acid [993-13-5] (2.B.4)	
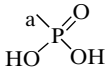	$\delta_{Ha} = 1.543(d)$, $^{2}J_{Ha,P} = 17.5$, $\delta_{Ca} = 14.39(d)$, $^{1}J_{Ca,P} = 136.6$, $\delta_{P} = 31.47(q)$; 9.6 mg/0.9 mL D_2O, pH 0; TSPSA, ext. H_3PO_4
	$\delta_{Ha} = 1.206(d)$, $^{2}J_{Ha,P} = 16.1$, $\delta_{Ca} = 16.20(d)$, $^{1}J_{Ca,P} = 132.9$, $\delta_{P} = 24.49(q)$; 10.4 mg/0.9 mL D_2O, pH 7.6; TSPSA, ext. H_3PO_4
	$\delta_{Ha} = 1.075(d)$, $^{2}J_{Ha,P} = 15.5$, $\delta_{Ca} = 16.86(d)$, $^{1}J_{Ca,P} = 130.5$, $\delta_{P} = 20.93(q)$; 11.6 mg/0.9 mL D_2O, pH 14; TSPSA, ext. H_3PO_4
***O*-Ethyl methylthiophosphonate (2.B.4)**	
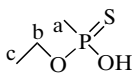	$\delta_{Ha} = 1.880(d)$, $\delta_{Hb'} = 4.201(ddq)$, $\delta_{Hb} = 4.142(ddq)$, $\delta_{Hc} = 1.348(dd)$, $^{2}J_{Ha,P} = 15.8$, $^{2}J_{Hb,Hb'} = -10.1$, $^{2}J_{Hb,P} = 9.0$, $^{2}J_{Hb',P} = 10.7$, $^{3}J_{Hb,Hc} = 7.0$, $^{3}J_{Hb',Hc} = 7.1$, $^{4}J_{Hc,P} = 0.2$, $\delta_{Ca} = 21.88(d)$, $\delta_{Cb} = 62.10(d)$, $\delta_{Cc} = 16.09(d)$, $^{1}J_{Ca,P} = 114.4$, $^{2}J_{Cb,P} = 6.8$, $^{3}J_{Cc,P} = 7.7$, $\delta_{P} = 89.95(m)$; 20 µL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4
***N*,*N*-Dimethylphosphoramidic dichloride [677-43-0] (2.B.5)**	
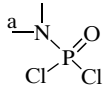	$\delta_{Ha} = 2.879(d)$, $^{3}J_{Ha,P} = 15.7$, $\delta_{Ca} = 37.02(d)$, $^{2}J_{Ca,P} = 3.6$, $\delta_{P} = 19.95(sp)$; 50 mL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4
Benzilic acid; 2,2-diphenyl-2-hydroxyacetic acid [76-93-7] (2.B.8)	
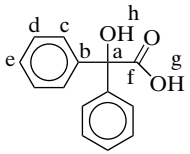	$\delta_{Hc} = 7.468(m)$, $\delta_{Hd} = 7.337(m)$, $\delta_{He} = 7.319(m)$, $\delta_{Hg,Hh} = 5.08(bs)$, $^{3}J_{Hc,Hd} = 7.95$, $^{4}J_{Hc,He} = 1.24$, $^{5}J_{Hc,Hd'} = 0.58$, $^{4}J_{Hc,Hc'} = 2.12$, $^{3}J_{Hd,He} = 7.45$, $^{4}J_{Hd,Hd'} = 1.48$, $\delta_{Ca} = 81.02$, $\delta_{Cb} = 141.55$, $\delta_{Cc} = 127.41$[g], $\delta_{Cd} = 128.21$[g], $\delta_{Ce} = 128.25$, $\delta_{Cf} = 176.93$; 24.7 mg/1.0 mL $CDCl_3$ + ca. 20 µL CD_3OD; TMS
***N*,*N*-Diisopropylaminoethan-2-ol [96-80-0] (2.B.11)**	
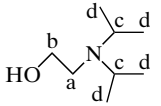	$\delta_{Ha} = 2.633(XX')$, $\delta_{Hb} = 3.460(AA')$, $\delta_{Hc} = 3.044(sp)$, $\delta_{Hd} = 1.029(d)$, $\delta_{OH} = 3.14(b)$, $^{3}J_{Hc,Hd} = 6.6$, $\delta_{Ca} = 45.26$, $\delta_{Cb} = 58.36$, $\delta_{Cc} = 47.61$, $\delta_{Cd} = 20.92$; 40 µL/0.8 mL $CDCl_3$; TMS
	$\delta_{Ha} = 3.288(XX')$, $\delta_{Hb} = 3.886(AA')$, $\delta_{Hc} = 3.776(sp)$, $\delta_{Hd} = 1.351, 1.364(d)$, $^{3}J_{Hc,Hd} = 6.6$, $\delta_{Ca} = 51.40$, $\delta_{Cb} = 60.22$, $\delta_{Cc} = 58.54$, $\delta_{Cd} = 19.28$, 20.99; 20 µL/0.8 mL D_2O, pH 0.5; TSP-d_4
	$\delta_{Ha} = 3.297(XX')$, $\delta_{Hb} = 3.899(AA')$, $\delta_{Hc} = 3.785(sp)$, $\delta_{Hd} = 1.371(d)$, $^{3}J_{Hc,Hd} = 6.6$, $\delta_{Ca} = 51.28$, $\delta_{Cb} = 60.15$, $\delta_{Cc} = 58.37$, $\delta_{Cd} = 20.02$; 23 µL/0.8 mL D_2O, pH 8.4; TSP-d_4
	$\delta_{Ha} = 2.590(XX')$, $\delta_{Hb} = 3.564(AA')$, $\delta_{Hc} = 3.020(sp)$, $\delta_{Hd} = 1.036(d)$, $^{3}J_{Hc,Hd} = 6.5$, $\delta_{Ca} = 50.13$, $\delta_{Cb} = 64.78$, $\delta_{Cc} = 52.71$, $\delta_{Cd} = 21.66$; 20 µL/0.8 mL D_2O, pH 12.8; TSP-d_4
Thiodiglycol; bis(2-hydroxyethyl)sulfide [111-48-8] (2.B.13)	
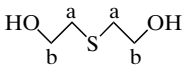	$\delta_{Ha} = 2.757(t)$, $\delta_{Hb} = 3.755(t)$, $^{3}J_{Ha,Hb} = 6.3$, $\delta_{Ca} = 36.10$, $\delta_{Cb} = 63.09$; 9.7 mg/0.6 mL D_2O, pH 6.5; TSPSA

Diethyl phosphite [762-04-9] (3.B.11)

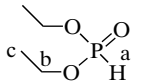

$\delta_{Ha} = 6.811(d)$, $\delta_{Hb'} = 4.157(ddq)$, $\delta_{Hb} = 4.148(ddq)$, $\delta_{Hc} = 1.368(t)$, $^{1}J_{Ha,P} = 692.4$, $^{2}J_{Hb,Hb'} = -10.2$, $^{2}J_{Hb,P} = 8.9$, $^{2}J_{Hb',P} = 9.3$, $^{3}J_{Hb,Hc} = 7.1$, $^{3}J_{Hb',Hc} = 7.1$, $\delta_{Cb} = 61.82(d)$, $\delta_{Cc} = 16.34(d)$, $^{2}J_{Cb,P} = 5.6$, $^{3}J_{Cc,P} = 6.2$, $\delta_{P} = 7.88(dqu)$; 20 µL/0.8 mL $CDCl_3$; TMS, ext. H_3PO_4

Ethyldiethanolamine [139-87-7] (3.B.15)

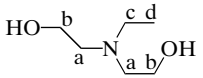

$\delta_{Ha} = 2.646(XX')$, $\delta_{Hb} = 3.613(AA')$, $\delta_{Hc} = 2.633(q)$, $\delta_{Hd} = 1.046(t)$, $\delta_{OH} = 3.29(b)$, $^{3}J_{Hc,Hd} = 7.1$, $\delta_{Ca} = 55.53$, $\delta_{Cb} = 59.66$, $\delta_{Cc} = 48.22$, $\delta_{Cd} = 11.62$; 40 µL/0.8 mL $CDCl_3$; TMS

$\delta_{Ha} = 3.38(XX')$, $\delta_{Hb} = 3.934(AA')$, $\delta_{Hc} = 3.370(q)$, $\delta_{Hd} = 1.327(t)$, $^{3}J_{Hc,Hd} = 7.3$, $\delta_{Ca} = 57.04$, $\delta_{Cb} = 58.20$, $\delta_{Cc} = 51.95$, $\delta_{Cd} = 10.89$; 40 µL/0.8 mL D_2O, pH 0.5; TSP-d_4

$\delta_{Ha} = 3.378(XX')$, $\delta_{Hb} = 3.942(AA')$, $\delta_{Hc} = 3.362(q)$, $\delta_{Hd} = 1.334(t)$, $^{3}J_{Hc,Hd} = 7.3$, $\delta_{Ca} = 56.94$, $\delta_{Cb} = 58.21$, $\delta_{Cc} = 51.81$, $\delta_{Cd} = 10.84$; 40 µL/0.8 mL D_2O, pH 7.7; TSP-d_4

$\delta_{Ha} = 2.678(XX')$, $\delta_{Hb} = 3.685(AA')$, $\delta_{Hc} = 2.617(q)$, $\delta_{Hd} = 1.039(t)$, $^{3}J_{Hc,Hd} = 7.2$, $\delta_{Ca} = 57.51$, $\delta_{Cb} = 61.54$, $\delta_{Cc} = 50.88$, $\delta_{Cd} = 13.05$; 40 µL/0.8 mL D_2O, pH 12.5; TSP-d_4

---

[a] Standard NATO agreement (STANAG) code.
[b] Chemical Abstracts Service registry number.
[c] Schedule number.
[d] Chemical shifts are given relative to $\delta_H(TMS) = \delta_H(TSP\text{-}\delta_4) = 0.00$, $\delta_H(TSPSA) = 0.015$, $\delta_C(TMS) = \delta_C(TSP\text{-}d_4) = \delta_C(TSPSA) = 0.00$, $\delta_F(CFCl_3) = 0.00$, and $\delta_P(H_3PO_4) = 0.00$ ppm.
[e] Multiplicity: b = broad, d = doublet, dd = doublet of doublets, ddq = doublet of doublet of quartets, dq = doublet of quartets, dqu = doublet of quintets, dspd = doublet of septet of doublets, dt = doublet of triplets, m = multiplet, q = quartet, s = singlet, sp = septet, t = triplet, AA′ = AA′ part of an AA′XX′ spin system, XX′ = XX′ part of an AA′XX′ spin system. If no multiplicity is given, the resonance appears as singlet.
[f] Only the couplings $^{2}J_{HH}$ are marked negative; all others are marked positive even though they may be negative.
[g] Tentative assignment.
GD, soman; Q, sesquimustard; T, *O*-mustard; L, lewisite-1; HN-1, nitrogen mustard one.

and concentrated by evaporation. The evaporation is carried out on a rotary evaporator (50 °C, 35 mmHg for evaporation of water), or by a gentle flow of nitrogen gas (for organic liquid) with warming (40–50 °C) or without. The small amount (ca. 1 mL) of solution left in the rotary evaporator is further evaporated nearly to dryness by a flow of nitrogen gas. To each concentrate, about 1 mL of the corresponding deuterated solvent (e.g. D_2O, CD_2Cl_2) is added and the solution is again evaporated nearly to dryness. These two steps are then repeated. The remaining solution is filtered through a cotton plug (or a plug of glass wool) placed in the neck of a Pasteur pipet into a washed, dust-free NMR tube.

Alternatively, an HPLC filter unit may be used for the filtration. The sample vessel is rinsed with a suitable amount of (deuterated) solvent and the rinsing solutions are added to the sample in the NMR tube. The pH of the aqueous liquid sample is determined, and can be adjusted with DCl and NaOD solution. For measurements carried out in a standard 5-mm (o.d.) NMR tube, about 0.6–0.8 mL of solvent is required, and for 10-mm tubes about 3–4 mL. If considered necessary, some of the extractions can be performed with deuterated solvents, which will simplify the sample-preparation procedure. Related practical sample preparations can be found in the literature.[10,12,14]

Because chemicals that distill easily or with water (e.g. pinacolyl alcohol) may be lost during rotary evaporation, an additional sample may be prepared for screening/analysis by taking a suitable amount of original aqueous liquid (e.g. 0.7 mL for a 5-mm o.d. tube), adding a small amount (e.g. 0.1 mL) of D_2O, determining the pH, and filtering the sample solution into an NMR tube. A similar non-evaporated NMR sample can be prepared from the organic extract.

If an internal chemical shift reference is to be used, a dilute solution of reference substance (e.g. TSPA-d_4) is prepared in the same solvent as used for the test sample (e.g. D_2O). A few microliters (or a suitable amount) of this reference solution are added to the evaporated sample solution before filtration into the NMR tube.

3.4.2 Preparation of Nuclear Magnetic Resonance Samples from Authentic Reference Chemicals

In a 2-mL vial, dissolve the selected amount of authentic reference chemical (e.g. 20 µL of mustard gas (HD)) in 0.8 mL of deuterated solvent (e.g. $CDCl_3$). [To ensure safety, always prepare NMR samples of toxic chemicals in a properly equipped fume cupboard or glovebox.] Using a Pasteur pipet, add as an internal chemical shift reference three to four drops of TMS solution, prepared by dissolving three to four drops of TMS in 0.5 mL of solvent ($CDCl_3$ in this case). Tightly place a plug of purified cotton for a filter in the neck of a Pasteur pipet and fix the pipet on a clamp stand. Rinse the cotton plug with the sample solution, then let the solution flow into the washed, dust-free NMR tube. With a suitable vacuum and welding apparatus, flame seal the tube near its open end in a low vacuum.

Alternatively, instead of cotton, glass wool or a suitable HPLC filter unit may be used for the filtration. Before filtering the authentic reference sample solution (in D_2O) into the NMR tube, adjust the pH to the same value as in the test sample.

See Figures 2 and 3 for NMR conditions for samples of authentic sarin (GB) and tabun (GA) and Table 1 for conditions for some other authentic samples.

3.5 Experiments and Data Processing

3.5.1 Preliminary Preparation

The probe head is separately tuned and matched[23] for each sample and experiment. The magnetic field homogeneity is carefully adjusted[23] and, always before final accumulation, line shape and line width are checked by recording trial 1-D spectra with a few scans. Every experiment run in the laboratory should be logged in an experiment logbook; a useful reference entry would contain a run number, the date, the file name, the sample code, and the initials of the operator.

3.5.2 Experimental Conditions

Unlike the sample condition, the experimental parameters have only a minor effect on the NMR spectral parameters. Experimental parameters such as spectral width, flip angle, repetition time, number of points in the free induction decay (FID) and in the real spectrum, number of scans, and processing parameters need to be comparable to those used for the acquisition of the database spectrum or spectrum of the authentic reference chemical. The experimental parameters determine whether quantitative information can be obtained from an NMR spectrum: under certain experimental conditions the resonance area is directly proportional to the number of nuclei.[26]

^1H and ^{19}F NMR spectra are recorded with a normal one-pulse sequence or, alternatively, the ^1H spectra are recorded with a sequence that allows simultaneous solvent suppression with presaturation[27] or a sequence that includes some other method of suppression; $^{13}C\{^1H\}$ and $^{31}P\{^1H\}$ spectra are recorded with proton broadband (composite pulse) decoupling,[28] and ^{31}P spectra with gated proton decoupling.[29]

The experimental conditions are the following: spectral ranges (all approximate) are from −0.5 to 11 ppm in ^1H NMR, from −40 to 120 ppm in $^{31}P\{^1H\}$ and ^{31}P NMR

(certain types of phosphorus-containing chemicals, e.g. phosphonites and phosphites, may have larger chemical shifts and their observation needs an extended spectral window), from −80 to 5 ppm in ^{19}F NMR, and from −5 to 230 ppm in $^{13}C\{^1H\}$ NMR, all relative to the respective references TMS, H_3PO_4, $CFCl_3$, and TMS ($\delta = 0$ ppm). The number of points in the FID and in the real spectrum normally ranges from 32 K to 128 K. The number of scans is selected so as to obtain a sufficient signal-to-noise ratio (S/N). Pulses of 45–90° flip angles are used and the repetition time (pulse interval) is selected to be 3–6 s. Note that because the relaxation times of the nuclei of the target chemicals are normally not known, these conditions are a compromise, and should be modified according to the case: shorter repetition time if the nuclei relax fast (this can be approximately checked from duration of the FID).

Experimental conditions for recording reference spectra from authentic chemicals are similar to those used for test samples. By way of example, Figure 2 shows the library NMR spectra of GB from the AC-Laboratory and Figure 3 those of GA from VERIFIN. The spectra are presented in their original form (size reduced from A4) to show layout and format of library spectra that have been acceptable to the Organisation for the Prohibition of Chemical Weapons (OPCW) Convention. The experimental conditions shown are the ''standard conditions'' used by the laboratory for the particular experiment.

3.5.3 Processing of Data

Usually, the FID is Fourier transformed after application of an exponential window function.[23] The line-broadening factor is selected to provide a good S/N. As a rule, the line broadening is selected according to the width of a narrow singlet resonance – in ^1H NMR normally to a value from 0.1 to 0.5 Hz (cf. Figures 2 and 3). The resolution of the spectrum may be enhanced, though with reduced sensitivity, if the FID is multiplied, for example by a suitable Gaussian shape function, before Fourier transformation. The resolution enhancement is applied to better resolve the details of resonances, as in the case of resonance overlapping. Zero filling, i.e. adding zero data points at the end of the FID prior to Fourier transformation, is sometimes necessary for proper representation of the line shape, particularly when resolution enhancement is performed. Baseline correction may be needed in cases where very-small-intensity resonances need to be presented as large expansions.

3.5.4 Plotting of Data

A spectrum is plotted in its full width and as expanded regions. The spectrum is plotted on paper, always with a title that assigns a link between the actual spectrum, the sample (sample code), and the entry in the experiment logbook (e.g. a source reference code). The expanded regions of the spectrum of the sample, the blank (if available), and the authentic reference chemical should be presented at the same scale and be sufficiently detailed. The experimental conditions are also printed out. Spectrum integration and peak picking are done where necessary.

4 QUALITY CONTROL

4.1 Quality Management System

Work performed in a laboratory must be traceable. This is achieved when all relevant information from sample preparation to experiments and reporting of data are recorded on paper or stored in a computer. In practice, this requires a quality management system of the laboratory. Continued effort toward quality in the laboratory and, if possible, national accreditation of the laboratory is recommended.

4.2 Data Quality

The quality of an NMR experiment can be assessed from the sample spectrum. In an optimal spectrum a singlet resonance line is both narrow and symmetric, the S/N is high, and all resonances of interest are seen. The spectrum quality is a measure known to the laboratory after work with the particular spectrometer/magnet and probe head on different samples; thus no limits can be suggested for the width of a singlet resonance. An experienced operator notices rather easily whether the problems with line shapes arise from a poor sample or inadequate homogeneity adjustment.

Resonance overlap is a real problem in NMR when the technique is applied for chemical identification: details of all resonances of the identified chemical should be revealed (see section 6.2). In particular, one should consider whether the outer lines of multiplets are being seen, e.g. those of the doublet of septets of the methine proton of GB (cf. Figure 2a); if the problem is poor S/N, more scans should be acquired in the experiment. Further homogeneity adjustment for a narrower line width and/or resolution enhancement may partially solve the problem of resonance overlap.

4.3 Spectrometer Performance

In accordance with the quality control (QC) regime of the laboratory, the performance of the spectrometer should be checked and documented regularly. The specifications

set by the spectrometer manufacturer are considered as the principal target to be obtained. Each experiment for the check of spectrometer performance is logged in the experiment logbook. The spectrometer-performance tests comprise checking of the ^{1}H line shape and sensitivity for a probe head and for a selected X-nucleus (a ^{1}H resolution test may be unnecessary if the ^{1}H line-shape test passes well). Other tests include checking of the radio-frequency field strengths of the observation and decoupling channels. A recommended practice for S/N determination in Fourier transform NMR is described in Standard Definitions of Terms, Symbols, Conventions, and References Relating to High-resolution NMR Spectroscopy (ASTM Standards, E 386-90). Spectrometer manufacturers provide instructions for S/N determination and the other tests [see also Braun et al.[22]].

4.4 Tests and Test Samples

Testing of the spectrometer performance requires a variety of test samples, usually provided by the spectrometer manufacturer. A set of tests and test samples may be as follows:

- ^{1}H line shape, resolution, and spinning side band test: 3% of chloroform in acetone-d_6;
- ^{1}H sensitivity: 0.1% ethylbenzene in $CDCl_3$;
- ^{31}P sensitivity: 0.0485 M triphenyl phosphate in $CDCl_3$;
- ^{19}F sensitivity: 0.05% trifluorotoluene in $CDCl_3$;
- ^{13}C sensitivity: 10% ethylbenzene in $CDCl_3$;
- ^{1}H resolution test: 5% of *o*-dichlorobenzene in acetone-d_6;
- temperature calibration: 80% of ethylene glycol in DMSO-d_6 and 4% of methanol in methanol-d_4.

Specifications for a probe head/field strength (magnet) combination are given by the spectrometer manufacturer for line shape, sensitivity, resolution, and radio-frequency field strengths. A specific test is considered "passed" when the specification value is met.

5 APPLICATIONS

5.1 Main One-dimensional Experiments

The most important experiments (see section 3.5) in the verification of CWC-related chemicals by NMR spectroscopy are ^{1}H, $^{31}P\{^{1}H\}$, ^{19}F, ^{31}P, and $^{13}C\{^{1}H\}$ (listed in the order of importance in interlaboratory comparison/proficiency tests; Figure 1). These 1-D experiments are useful both for the identification of chemicals and in structure elucidation. The ^{1}H, $^{13}C\{^{1}H\}$, ^{19}F, $^{31}P\{^{1}H\}$, and/or ^{31}P NMR spectra can be recorded one after the other from the same NMR sample. The ^{1}H NMR experiment is most sensitive and often will be run first. Specific couplings in the ^{1}H spectrum may indicate the presence of phosphorus or fluorine, which is then confirmed by $^{31}P\{^{1}H\}$ and/or ^{31}P and ^{19}F experiments. $^{13}C\{^{1}H\}$ experiments are performed if necessary and if the sample concentration is high enough. The literature can be consulted for general reading on NMR,[3,4] ^{31}P and multinuclear NMR,[30,31] and practical performance of experiments.[22]

5.1.1 Screening

The presence of phosphorus- or fluorine-containing chemicals above the detection limit can be checked by running $^{31}P\{^{1}H\}$ and ^{19}F NMR experiments. Normally, in environmental samples, no organic phosphorus- or fluorine-containing chemicals exist in the background, which makes these experiments useful as a screening method. Fluorine observation has an advantage because the sensitivity is nearly as high as for protons. Although ^{1}H NMR also offers a sensitive screening method, there could be interference from a high proton background. The presence of phosphorus-containing chemicals could then not be ruled out simply by recording the ^{1}H spectrum.

5.1.2 ^{1}H Nuclear Magnetic Resonance

The ^{1}H NMR spectrum (Figures 2a and 3a) provides information on the number and type of protons [number of resonances, areas of resonances, δ_H (ppm)] in the molecule and, in the form of couplings [J (Hz)], may reveal other magnetic atoms (e.g. ^{1}H, ^{31}P, ^{19}F) one to four bonds away from a hydrogen atom. (The maximum distance of four bonds between the coupling nuclei should be considered approximate.) Note that the experimental conditions required for quantitative work, as given in section 3.5.2, may not necessarily have been obtained, and the areas of the resonances should therefore be considered only approximate. In general, much longer repetition times are needed for quantitative conditions.

5.1.3 $^{31}P\{^{1}H\}$ and ^{31}P Nuclear Magnetic Resonance

The $^{31}P\{^{1}H\}$ NMR spectrum (Figures 2b and 3b) provides information on the number and type of phosphorus atoms (number of resonances, areas of resonances, δ_P) in the molecule and, in the form of couplings, may reveal other magnetic atoms (e.g. ^{31}P, ^{19}F) one to four bonds from a phosphorus atom. (The maximum distance of four bonds between the coupling nuclei should be considered approximate.) The notation $^{31}P\{^{1}H\}$ means that, when ^{31}P is observed, protons are irradiated so that the J-coupling

between protons and phosphorus disappears; however, the interaction with other magnetic nuclei remains.

The ^{31}P NMR spectrum (Figures 2c and 3c) provides information on the number and type of phosphorus atoms (number of resonances, areas of resonances, δ_P) in the molecule and, in the form of couplings, may reveal other magnetic atoms (e.g. ^{1}H, ^{31}P, ^{19}F) one to four bonds from a phosphorus atom. (The maximum distance of four bonds between the coupling nuclei should be considered approximate.) Note that the $^{31}P\{^{1}H\}$ and ^{31}P NMR spectra are not quantitative and the areas of the resonances should be considered only approximate. In addition, a quantitative $^{31}P\{^{1}H\}$ experiment utilizes an inverse gated pulse sequence.[32]

5.1.4 ^{19}F Nuclear Magnetic Resonance

The ^{19}F NMR spectrum (Figure 2d) provides information on the number and type of fluorine atoms (number of resonances, areas of resonances, δ_F) in the molecule and, in the form of couplings, may reveal other magnetic atoms (e.g. ^{1}H, ^{31}P, ^{19}F) one to four bonds from a fluorine atom. (The maximum distance of four bonds between the coupling nuclei should be considered approximate.) Note that the experimental conditions required for quantitative work, as given in section 3.5.2, may not necessarily have been obtained, and the areas of the resonances should therefore be considered only approximate. In general, much longer repetition times are needed for quantitative conditions.

5.1.5 $^{13}C\{^{1}H\}$ Nuclear Magnetic Resonance

The $^{13}C\{^{1}H\}$-NMR spectrum (Figures 2e and 3d) provides information on the number and type of carbons (number of resonances, areas of resonances, δ_C) in the molecule and, in the form of couplings, may reveal other magnetic atoms (e.g. ^{31}P, ^{19}F) one to four bonds from a carbon atom. (The maximum distance of four bonds between the coupling nuclei should be considered approximate.) Note that the $^{13}C\{^{1}H\}$ NMR spectra are not quantitative and the areas of the resonances should be considered only approximate. In addition, a quantitative $^{13}C\{^{1}H\}$ experiment utilizes an inverse gated pulse sequence.[32]

5.2 $^{1}H\{^{31}P\}$, $^{1}H\{^{19}F\}$, and $^{1}H\{^{1}H\}$ Nuclear Magnetic Resonance Experiments

In $^{1}H\{^{31}P\}$, $^{1}H\{^{19}F\}$, and $^{1}H\{^{1}H\}$ NMR experiments a proton is observed, while phosphorus, fluorine, or proton is decoupled. The phosphorus and fluorine are decoupled either selectively or with broad-band irradiation, while the proton is irradiated selectively. In each case a ^{1}H NMR spectrum simplified by the removal of the effect of the irradiated nucleus to the spectrum is obtained, i.e. all nuclei that are coupled to the decoupled nucleus/nuclei exhibit a simplified resonance pattern. Comparison of the normal ^{1}H NMR spectrum with the decoupled spectrum may help in interpretation of the former and in resolution of the molecular structure. Suitable 2-D experiments (homo- and heteronuclear correlation) provide comparable information; however, the 1-D experiments are quicker to perform and show the resonances in a more detailed manner.

5.3 Relevant Two-dimensional Techniques

2-D chemical shift correlation NMR techniques may be useful in supporting the verification analysis (Figure 1). In particular, they may be useful in the structure elucidation of an unknown CWC-related chemical. In 2-D spectra, the information is spread in the form of correlations over an *xy*-plane instead of being contained in a single spectrum curve as in 1-D spectra. Correlations connect resonances of the same or different types of nuclei of a molecule and give information on how close the nuclei yielding the resonances are. Most of the techniques make use of a *J*-coupling constant (a spin–spin coupling through chemical bonds). The rule that the size of the coupling constant (and the intensity of correlation) decreases as the distance between the coupled nuclei increases is roughly valid. Because of the two dimensions and the correlations, spectral analysis may be possible even in the presence of strong background. The analysis of mixtures is more straightforward with 2-D techniques because the correlations do not overlap as readily as do the resonances in a 1-D spectrum.

By way of example, useful 2-D techniques are correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), heteronuclear multiple quantum coherence (HMQC), and heteronuclear multiple quantum coherence/total correlation spectroscopy (HMQC/TOCSY).[1,2,22] An excellent technique is a proton detected proton–phosphorus correlation experiment with combined HSQC and TOCSY that has been used for the analysis of CWC-related chemicals during an international interlaboratory proficiency test.[13] In this technique, protons of an alkyl phosphonofluoridate, for example, show correlations to the phosphorus nucleus of the molecule, which is then used to assign from the 1-D ^{1}H spectrum the resonances that belong to this particular chemical. A 2-D NMR spectrum might also serve as a library or reference spectrum in the identification of a chemical.

5.4 Scheduled Chemicals

The suitability of NMR spectroscopy and the detection limits for the analysis of the Schedule 1–3 chemicals[15]

are summarized below. Riot control agents and vomiting agents are normally soluble in organic solvents, and are readily analyzed by NMR.[11,33]

5.4.1 Schedule 1 Chemicals

Almost all Schedule 1 chemicals are soluble in the organic NMR solvents used in verification (see section 3.1) and can be analyzed by NMR spectroscopy. Saxitoxin (1.A.7) and ricin (1.A.8) differ from the others. Both are derived from natural sources – the former is a paralytic shellfish poison and the latter a glycoprotein toxin.[34] Analytical methods [Recommended Operating Procedures (ROPs)] have not been established for either chemical. NMR data of saxitoxin in D_2O have been reported.[35,36] By way of example, Table 1 presents the NMR spectral parameters of GB, GD, GA, VX, HD, Q, T, L, HN-1, methyl phosphonyl difluoride, and chlorosarin.

5.4.2 Schedule 2 Chemicals

All Schedule 2 chemicals are analyzable by NMR, except arsenic trichloride, which does not possess a useful nucleus. Either an organic solvent or D_2O is used. Note that the hydrolysis products of alkylphosphonofluoridates (1.A.1), alkylphosphonothiolates (1.A.3), alkylphosphonyl difluorides (1.B.9), alkylphosphonites (1.B.10), chlorosarin (1.B.11), and chlorosoman (1.B.12), which have an alkyl group (Me, Et, *n*-Pr, or *i*-Pr) linked directly to a phosphorus, all belong to Schedule 2, group B.4, and NMR analysis for them is made either in D_2O or in an organic solvent. By way of example, NMR spectral parameters of isopropyl methylphosphonate, pinacolyl methylphosphonate, methylphosphonic acid, *O*-ethyl methylthiophosphonate, *N*,*N*-dimethylphosphoramidic dichloride, benzilic acid, *N*,*N*-diisopropylaminoethan-2-ol, and thiodiglycol are presented in Table 1.

5.4.3 Schedule 3 Chemicals

Most of the Schedule 3 chemicals are readily analyzed by NMR: trimethyl phosphite, triethyl phosphite, dimethyl phosphite, diethyl phosphite, ethyldiethanolamine, methyldiethanolamine, and triethanolamine. Organic solvents or D_2O are used. Although chloropicrine and the gases phosgene (3.A.1), cyanogen chloride (3.A.2), and hydrogen cyanide (3.A.3) can also be analyzed by NMR, their identification by other methods is more reliable. Phosphorus oxychloride (3.B.5), phosphorus trichloride (3.B.6), and phosphorus pentachloride (3.B.7) react readily with moisture, but in principle can be analyzed by NMR in a dry organic solvent. Sulfur monochloride (3.B.12), sulfur dichloride (3.B.13), and thionyl chloride (3.B.14) do not possess an NMR nucleus suitable for verifying them. By way of example, NMR spectral parameters of diethyl phosphite and ethyldiethanolamine are presented in Table 1.

5.5 Detection Limits

The detection limit of a chemical in NMR depends on its NMR properties (e.g. number of equivalent nuclei, presence and extent of couplings), the magnetic field strength of the spectrometer, the number of pulses, the flip angle used, the observed nucleus and its relaxation, the NMR tube diameter, the sample matrix, and the magnet shimming. Detection limits in ^1H, ^{13}C, ^{19}F, and ^{31}P observation should all be determined, remembering, however, that the detection limit for a reference chemical does not guarantee the same detection limit for the chemical in a real sample. Typically the detection limit of a chemical when ^1H is observed at 400 MHz or ^{19}F at 376 MHz may be 1 $\mu g\,mL^{-1}$ (1 ppm). Almost one order of magnitude more of the analyte is required for phosphorus observation (5 ppm at 162 MHz), and recording of $^{13}C\{^1H\}$ spectra may be futile in concentrations below 100 $\mu g\,mL^{-1}$ (100 ppm).

6 ANALYSIS

6.1 Acquired Nuclear Magnetic Resonance Data

The spectral data required to support an identification by NMR are the sample spectrum, the reference spectrum, and the spectrum of a blank sample (if available). All should be recorded under comparable conditions. The reference spectrum may be from a library or database, it may be recorded from the authentic reference chemical, or it may be a spectrum of a closely related chemical; in this last case spectral interpretation must be enclosed.[37] Other 1-D and 2-D NMR spectra (see sections 5.2 and 5.3) may have been recorded from the test sample to support the interpretation, and identification.

6.2 Identification Criteria

As the main rule in verification, unambiguous identification of a CWC-related chemical should be based on at least two analytical, preferably different spectrometric, techniques (e.g. gas chromatography/mass spectrometry (GC/MS) and NMR) giving consistent results.[37] The NMR spectral parameters for a chemical should agree within certain limits with those of a database or spectrum library, or with the parameters derived from the spectrum of an authentic reference chemical. According to the identification criteria,[18] the values of δ_H, δ_C, and δ_F should agree with the corresponding values in the database or library or the values of an authentic reference chemical

within limits ±0.1 ppm, and those of δ_P within ±1 ppm. The values of coupling constants $^nJ_{H,H}$, $^nJ_{H,F}$, $^nJ_{H,P}$, $^nJ_{C,P}$, $^nJ_{C,F}$, and $^nJ_{P,F}$ should agree within limits of ±0.5 Hz. In addition, the resonance patterns and their intensities must match. Normally, when the conditions of the test sample and the reference sample are comparable, the differences in the NMR parameters are much less than suggested above; sometimes, within reasonable limits and when credibly explained, greater differences may be acceptable.

Sample conditions (solvent, pH in water, chemical shift references, concentration) and experimental conditions should correspond in the sample and reference sample. Sample concentration and the experimental conditions are less critical for the NMR spectral parameters.

Other chemicals (matrix) present in the sample, especially at low analyte concentration, may affect the NMR spectral parameters and how the resonances are revealed. A general requirement for a spectrum acceptable for identification is that resonances of other chemicals do not overlap with resonances of the identified chemical. Partial overlapping may be acceptable if the resonance of the identified chemical can still be credibly explained. Where insufficient data are obtained, for example owing to severe overlapping, the resonances revealed may still be useful in supporting identifications based on other analytical techniques.

6.3 Interpretation

The sample spectra are compared with the corresponding spectra recorded from the blank sample (if available), and the differences are investigated. If the blank sample represents the background well, the resonances that are not from the background can be determined directly. Obtaining a representative blank sample from a contaminated environmental site will not always be a realistic possibility, however. Initial interpretation of the spectra may reveal a need for further sample preparation and experiments (Figure 1).

Indicative structural details may be revealed. For example methyl, ethyl, propyl, and isopropyl groups directly linked to phosphorus will show different resonance patterns in ^{1}H NMR spectra. The $^{31}P\{^1H\}$ NMR spectrum can reveal the presence of phosphorus-containing chemicals and, in the case of a large coupling constant (due to a P–F bond, see Figure 2b), also the presence of fluorine. The number of resonances and their integrated areas may give an idea of the number of chemicals present.

Note that nerve agent homologs that possess the same atoms one to three bonds apart from phosphorus may show similar phosphorus or fluorine resonances at comparable chemical shifts. Thus an analysis based solely on ^{31}P and/or ^{19}F NMR spectra may lead to an erroneous identification. Identification should not be based solely on $^{31}P\{^1H\}$ NMR spectra either. In any identification by NMR spectroscopy, ^{1}H NMR is recommended as the first method of choice.

In the case of an unknown chemical, or where resonance overlap occurs, it may be necessary to call upon the full arsenal of NMR methods. To confirm a heteronuclear coupling the normal ^{1}H NMR spectrum is compared with $^1H\{^{19}F\}$ and/or $^1H\{^{31}P\}$ NMR spectra. After this, and in particular where a strong background is present, the various 2-D NMR spectra are recorded. Homonuclear chemical shift correlation experiments such as COSY and TOCSY (or some of their variants) provide information on coupled protons, even networks of protons,[1] while the inverse detected heteronuclear correlation experiments such as HMQC and HMQC/TOCSY provide similar information but only for protons coupling to heteronuclei, e.g. the pairs ^{1}H–^{31}P and ^{1}H–^{13}C. Although interpretation of these data provides abundant information on the molecular structure, the results obtained with other analytical or spectrometric techniques must also be taken into account. The various methods of MS and gas chromatography/Fourier transform infrared (GC/FTIR) spectroscopy supply complementary information to fully resolve or confirm the structure. Unambiguous identification of an unknown chemical requires consistent results from all spectrometric techniques employed.

6.4 Data Evaluation

In the evaluation of acquired NMR data one must consider whether the quality of the spectra and other accompanying information are sufficient for identification, or whether the data should only be regarded as support to an identification made by other techniques. One must consider whether all resonances of the chemical were revealed (S/N; occurrence and significance of overlapping resonances of other chemicals), whether the resonance patterns in the sample and reference spectra match and whether they are correctly interpreted, whether the sample conditions were comparable to those of the reference sample, whether the reference data are of adequate quality, and, in the case of interpretation, how the related data supported the identification.[37]

6.5 Identification

Identification by NMR of a CWC-related chemical is made through reference to spectra (^{1}H, $^{13}C\{^1H\}$, ^{19}F, $^{31}P\{^1H\}$, and/or ^{31}P) included in a spectral library, through comparison with a spectrum of the authentic reference chemical, or through spiking the test sample with an authentic reference chemical. The spectrum of a

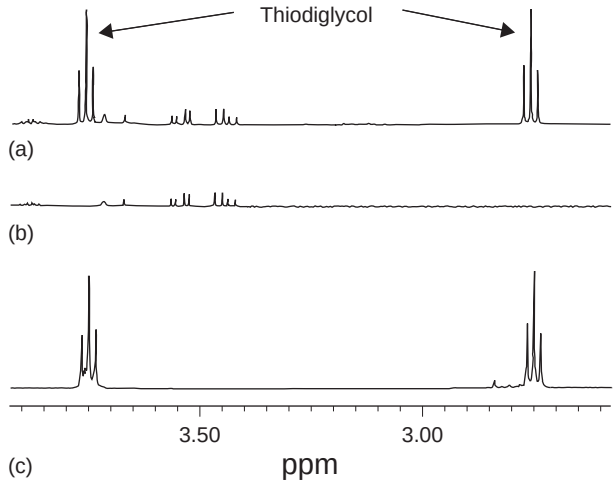

Figure 4 Expanded regions of ^{1}H NMR spectra recorded from (a) a soil sample (pH 8.2) and (b) a blank soil sample in the Fourth Official OPCW Proficiency Test (May 1998). The soils were extracted with water, the extract was concentrated, and the remaining water was replaced with D_2O. (c) Spectrum of authentic thiodiglycol (pH 6.3), which was identified in the soil sample. The ^{1}H chemical shifts were 2.763 and 3.758 ppm in (a) and 2.757 and 3.755 in (c); the coupling constant was 6.3 Hz in both spectra. The experimental conditions in all spectra were similar.

blank sample (if available) is examined for comparison. Identification of a chemical is obtained if all the identification criteria are met. Identification made by spiking the sample with the suspected chemical requires exact overlapping of the resonances. By way of example, the ^{1}H NMR spectra for the identification of thiodiglycol in a soil sample in the Fourth Official OPCW Proficiency Test are presented in Figure 4.

6.6 Nuclear Magnetic Resonance Spectra of Sarin and Tabun

The ^{1}H, $^{31}P\{^{1}H\}$, ^{31}P, $^{13}C\{^{1}H\}$, and ^{19}F NMR spectra of GB and GA are shown in Figures 2 and 3. The spectra of GB (Figure 2a–e) are from the Atlas of NMR Spectra of the AC-Laboratory and the spectra of GA (Figure 3a–d) from VERIFIN's NMR Spectrum Library. Interpretations of the spectra are given below. The chemical shifts of the nuclei are not discussed since they can be explained with use of the common chemical shift correlation rules found in textbooks.[3,4,30]

In the ^{1}H NMR spectrum of GB (Figure 2a), protons H-a yield a doublet of doublets due to couplings to phosphorus and fluorine. This is typical of all *O*-alkyl methylphosphonofluoridates. Proton H-b shows a doublet of septets due to coupling to phosphorus and the six methyl protons. Protons H-c and H-c′ are diastereotopic and resonate as separate doublets; the intensity of the former is lower because of the different long-range couplings to phosphorus and fluorine. The $^{31}P\{^{1}H\}$ spectrum (Figure 2b) shows a typical doublet due to a large $^{1}J_{P,F}$. The ^{31}P spectrum (Figure 2c) shows in addition doublets of quartets through couplings of phosphorus to H-b and H-a, respectively. In general, this kind of resonance is interpreted as indicating the structure $CH{-}O{-}P(=O)(CH_3)F$, and does not necessarily identify GB. The ^{19}F spectrum (Figure 2d) reveals a doublet of quartets due to couplings with the phosphorus and protons H-a, thus giving an indication of substructure $-O-P(=O)(CH_3)F$. Carbons C-a and C-c (Figure 2e) yield a doublet of doublets due to couplings to phosphorus and fluorine, and C-b and C-c′ doublets due to coupling to phosphorus.

In the ^{1}H NMR spectrum of GA (Figure 3a), protons H-a yield a typical doublet due to coupling with phosphorus. Protons H-b show a resonance that resembles a doublet of quartets; however, the protons are diastereotopic with a small chemical shift difference and reveal a higher-order resonance. The couplings and chemical shifts can be analyzed with use of a spectral analysis program. Protons H-c yield a doublet of triplets with a four-bond coupling to phosphorus and couplings to H-b and H-b′. The $^{31}P\{^{1}H\}$ spectrum (Figure 3b) shows a singlet at a chemical shift typical for alkyl *N*,*N*-dialkylphosphoramidocyanidates. The ^{31}P resonance (Figure 3c) resembles a nonet; however, the correct interpretation is a doublet (8.6 Hz) of doublets (9.8 Hz) of septets (11.3 Hz), but because of the broad lines the resonance is not seen in detail. All resonances in the $^{13}C\{^{1}H\}$ spectrum (Figure 3d) of GA are doublets due to phosphorus couplings.

6.7 Nuclear Magnetic Resonance Spectral Parameters of Selected Chemical Weapons Convention-related Chemicals

Molecular structures, NMR spectral parameters, and sample conditions for selected scheduled CW agents and some of their precursors and hydrolysis products are presented in Table 1. The NMR spectral parameters are given in $CDCl_3$ or D_2O, or both (sometimes at different pH values). Resonances of the chemicals can be assigned from the data, but the assignments are not alone sufficient for identifications.

7 DATABASES

In the following, we list the requirements for the construction of an in-house NMR spectrum library for verification of CWC-related chemicals and briefly describe the OPCW Analytical Database and the VERIFY database (VERIFIN's analytical reference database).

7.1 Requirements for an In-house Library

The main requirements for the construction of an in-house NMR spectral library for verification of CWC-related chemicals may be listed as follows:

- All ^{1}H, $^{13}C\{^{1}H\}$, ^{19}F, $^{31}P\{^{1}H\}$, and ^{31}P NMR spectra relevant to the molecular structure should be included (cf. Figures 2 and 3).
- Spectra should be recorded with the authentic chemical dissolved in a commonly used NMR solvent, e.g. $CDCl_3$ or D_2O. Since the solvent, and the pH in aqueous samples, may affect the appearance of the spectrum, particularly in ^{1}H NMR, it is advisable to record spectra in different solvents (e.g. CD_2Cl_2, acetone-d_6) and at different pH values (e.g. pH 0; pH ca. 7 or non-adjusted; pH 14).
- NMR spectral parameters that can be easily extracted from the spectrum should be included, and the spectrum quality should be evaluated. The authentic chemical must be relatively pure so that there is no disturbing overlap of the resonances of the authentic chemical. In evaluation of the spectra, the laboratory should follow the OPCW evaluation criteria (see section 7.2).
- MS and/or IR spectroscopic analyses of the same authentic chemical should have been done to confirm the identity.

The principal contents of an NMR spectrum library page are as follows (see Figures 2 and 3).

- spectrum type;
- full-width spectrum and sufficient expansions to reveal the spectral details;
- chemical information (name, CAS number, structure);
- NMR spectral parameters (chemical shifts and coupling constants);
- sample conditions, experimental conditions, and data processing conditions;
- spectrum quality (line width) and traceability (source reference).

7.2 The Organization for the Prohibition of Chemical Weapons Analytical Database

An important part of implementing the CWC is establishing reliable verification mechanisms. As the analysis of samples taken on the occasion of inspections is one part of this mechanism, the availability of collections of analytical data is essential. The OPCW, as the body responsible for the implementation of the CWC, wishes to have its own Analytical Database, which contains as much data on CWC-related chemicals as possible. Since on-site analysis by IR and in particular MS is planned for future verification activities, libraries of IR and mass spectra are mandatory. Although NMR spectroscopy is not suitable for on-site analysis, it is nevertheless considered an essential technique for laboratories specialized in the detection and identification of CW agents and related chemicals. Qualified laboratories from all parts of the world were therefore asked to submit their IR, mass, and NMR spectra to be included in the OPCW Analytical Database.

An atlas of NMR spectra of the OPCW Analytical Database has been compiled with the assistance of dedicated laboratories worldwide. The efforts of the laboratories and the OPCW have yielded a useful high-quality NMR spectral database. The only factor limiting its usefulness could be the difference of the instruments (in magnetic field strength and resonance frequency) that were used, because the resonance frequency may affect the spectrum appearance, in particular in ^{1}H NMR. However, this is not considered to be a serious problem because many of the spectra were recorded on 300–400 MHz instruments whose spectra do not differ much from those recorded at 200 or 500 MHz. The difference between the two extremes may be larger. The OPCW requires that all spectra to be included in the OPCW Analytical Database be evaluated.

On behalf of the OPCW, the submitted spectra were evaluated by groups of specialists. It was decided that only 1-D ^{1}H, $^{13}C\{^{1}H\}$, ^{19}F, ^{31}P, and $^{31}P\{^{1}H\}$ NMR spectra recorded on FT spectrometers with a proton frequency of 200 MHz or higher should be included in the database. The spectra were evaluated according to the following main criteria:

- the spectrum must be consistent with the assigned structure;
- the name, CAS number (where available), and chemical structure with numbering of atoms must be indicated on the spectra;
- resonances must be assigned where reasonably possible;
- coupling constants must be included where easily extractable;
- expansions must be displayed where relevant;
- resonances of impurities must be marked with asterisks;
- resonances of impurities must not overlap the resonances of the chemical of interest;
- the solvent and pH (if relevant) must be specified;
- the reference chemical must be indicated;
- if the reference chemicals are other than TMS or TSPA-d_4 for ^{1}H, TMS for $^{13}C\{^{1}H\}$, external H_3PO_4 for ^{31}P, and $CFCl_3$ for ^{19}F set at 0 ppm, their chemical shifts must be specified;

- the spectrum must have been recorded with adequate resolution to enable all required information to be derived;
- the spectra must be phased correctly;
- the S/N must be adequate for the detection of all relevant signals;
- the type and frequency of the spectrometer must be indicated;
- the spectral width must be indicated.

Even before the CWC entered into force, the group of specialists approved some 900 NMR spectra recorded from 205 chemicals. It was decided to compile the accepted spectra as paper copies. A first version of the resulting atlas of NMR spectra was certified in February 1998. The process of extending the OPCW Database, which is to be made available on request to member states, is in continuous progress.

7.3 VERIFY Database

VERIFY is a computerized database designed for the storage, management, and search of chemical and analytical reference data.[38] Properties and contents of the NMR part of VERIFY are as follows:

- NMR spectrum;
- chemical information;
- spectrum identification code and spectrum type;
- molecular structure and numbering of nuclei;
- NMR spectral parameters (chemical shift and coupling constant data) for nuclei;
- sample and experimental conditions;
- instrument information;
- capability for display, expansion, and printing of spectra; importing and exporting of spectra in JCAMP-DX format; no capability for spectral search.

The first versions of VERIFY were designed to work in the VAX environment,[38] but the new version will work on a personal computer (PC) platform. The advantages of the computerized version relative to the paper version of the NMR library are easier access to the spectra and data search routines.

8 PERSPECTIVES AND FUTURE DEVELOPMENTS

8.1 Liquid Chromatography/Nuclear Magnetic Resonance

LC and supercritical fluid chromatography (SFC) have been combined on-line for the separation of analytes prior to their detection by NMR [the literature can be consulted for LC/NMR[39–41] and SFC/NMR[42]]. Although none of these techniques has yet been used in the analysis of CWC-related chemicals, their application in the field of verification is likely to begin soon.

8.2 Ultrahigh-field Spectrometers

A 400 MHz spectrometer is today a mid-frequency instrument. With a higher-field spectrometer, the detection limit is lower and the resonances are better dispersed, thus solving in part the problem of resonance overlap. By way of example, a ^{1}H sensitivity of 1234 : 1 (0.1% ethylbenzene; cf. section 4.4) was obtained on an 800 MHz spectrometer.[43] This means that roughly one order of magnitude less chemical can be detected.

8.3 "Nano" Probe Technology

Less than 1-μg quantities of analytes in about 50–60 μL of solvent can be detected with a novel probe head known as the "nano" probe.[44] Although with this probe head the absolute amount of a chemical needed for a useful spectrum is low, the problem of resonance overlap remains.

9 COMPARISON WITH OTHER SPECTROMETRIC TECHNIQUES

The main advantages of NMR spectroscopy in verification of CWC-related chemicals are as follows:

- At sufficient concentration and in the absence of disturbing background resonances, NMR is the superior method both for the identification of known chemicals and for the structural elucidation of unknown chemicals. Its usefulness in identification is attributable to the fingerprint nature of spectra, while the usefulness in structural elucidation rests on the structural specificity of the spectra. The wide variety of routine 1-D and 2-D experiments available is of assistance in both identifying of chemicals and structure elucidation.
- Most of the scheduled chemicals are directly analyzable by NMR. The spectra can be recorded in aqueous or in organic liquids from acidic, neutral, or basic chemicals, and from alkylated and protonated salts.
- Because NMR is a nondestructive method, the NMR samples can be used for the preparation of samples for other analytical techniques. (Note that in some cases the deuterated solvent may deuterate the chemicals of interest affecting their detection, for example by GC/MS.)

The main disadvantages of NMR are that

- there may be background resonances and, therefore, possible overlap of resonances of the chemicals of interest;
- there is no separation method for NMR comparable to GC for MS and FTIR (see section 8.1, however);
- sensitivity is low: the detection limits are high compared with MS and FTIR;
- NMR is not suitable for on-site analysis.

ACKNOWLEDGMENTS

Most of the authentic chemicals used for NMR analysis in this work were provided by the AC-Laboratory, Spiez, Switzerland. We wish to thank the AC-Laboratory for this gift and we also warmly thank Dr Walter Aue of the same laboratory for his comments on the manuscript.

ABBREVIATIONS AND ACRONYMS

ASC	Automatic Sample Changer
COSY	Correlation Spectroscopy
CW	Chemical Warfare
CWC	Chemical Weapons Convention
FID	Free Induction Decay
GA	Tabun
GB	Sarin
GC	Gas Chromatography
GC/FTIR	Gas Chromatography/Fourier Transform Infrared
GC/MS	Gas Chromatography/Mass Spectrometry
GD	Soman
HD	Mustard Gas
HMQC	Heteronuclear Multiple Quantum Coherence
HMQC/TOCSY	Heteronuclear Multiple Quantum Coherence/Total Correlation Spectroscopy
HN-1	Nitrogen Mustard One
HPLC	High-performance Liquid Chromatography
IR	Infrared
L	Lewisite-1
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OPCW	Organisation for the Prohibition of Chemical Weapons
PC	Personal Computer
Q	Sesquimustard
QC	Quality Control
ROPs	Recommended Operating Procedures
SFC	Supercritical Fluid Chromatography
S/N	Signal-to-noise Ratio
T	*O*-Mustard
TMS	Tetramethylsilane
TOCSY	Total Correlation Spectroscopy
TSPA-d_4	3-(Trimethylsilyl)-3,3,2,2-tetra-deuteropropionic Acid Sodium Salt
TSPSA	3-(Trimethylsilyl)propane Sulfonic Acid Sodium Salt
VERIFY	VERIFIN's Analytical Reference Database
1-D	One-dimensional
2-D	Two-dimensional
^{13}C	Nucleus Carbon-13
$^{13}C\{^1H\}$	Experiment Carbon-13 Observation with Simultaneous ^1H Decoupling
^{19}F	Nucleus Fluorine-19, Experiment Fluorine-19 Observation
^1H	Nucleus Hydrogen-1 or Proton, Experiment Hydrogen-1 or Proton Observation
^{31}P	Nucleus Phosphorus-31, Experiment Phosphorus-31 Observation
$^{31}P\{^1H\}$	Experiment Phosphorus-31 Observation with Simultaneous ^1H Decoupling
δ	Chemical Shift

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance • High-performance Liquid Chromatography Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General

REFERENCES

1. W.R. Croasmun, R.M.K. Carlson (eds.), *Two-dimensional NMR Spectroscopy. Applications for Chemists and Biochemists*, VCH Publishers, New York, 1994.
2. G.E. Martin, A.S. Zektzer, *Two-dimensional NMR Methods for Establishing Molecular Connectivity: a Chemist's Guide to Experiment Selection, Performance, and Interpretation*, VCH Publishers, New York, 1988.
3. H. Günther, *NMR Spectroscopy. Basic Principles, Concepts, and Applications in Chemistry*, 2nd edition, John Wiley & Sons, Chichester, 1992.
4. H. Friebolin, *Basic One- and Two-dimensional NMR Spectroscopy*, 2nd edition, VCH, Weinheim, 1993.
5. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, Weinheim, 1989.
6. J.-J. Delpuech (ed.), *Dynamics of Solutions and Fluid Mixtures by NMR*, John Wiley & Sons, Chichester, 1995.
7. M. Rautio (ed.), 'F.1 Testing of Existing Procedures' (1990). 'F.2 Testing of Procedures on Simulated Industry Samples' (1991). 'F.3 Testing of Procedures on Simulated Military Facility Samples' (1992). 'F.4 Validating of Procedures for Water and Soil Samples' (1993). 'F. International Interlaboratory Comparison (Round-robin) Test for the Verification of Chemical Disarmament. Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', Ministry for Foreign Affairs of Finland, Helsinki, 1990–1993.
8. M. Rautio (ed.), 'H.1 First Interlaboratory Comparison Test'. 'H.Interlaboratory Comparison Test Coordinated by the Provisional Technical Secretariat for the Preparatory Commission for the Organisation for the Prohibition of Chemical Weapons. Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', Ministry for Foreign Affairs of Finland, Helsinki, 1994.
9. E.R.J. Wils, A.G. Hulst, P.E.J. Verwiel, S.H. van Krimpen, A. Niederhauser, 'Identification of an Octyl Methylphosphonofluoridate Mixture in Chemical Industry Samples', *Fresenius' J. Anal. Chem.*, **343**, 297–303 (1992).
10. M. Mesilaakso, E-L. Tolppa, 'Detection of Trace Amounts of Chemical Warfare Agents and Related Compounds in Rubber, Paint, and Soil Samples by ^{1}H and $^{31}P\{^{1}H\}$ NMR Spectroscopy', *Anal. Chem.*, **68**, 2313–2318 (1996).
11. M. Mesilaakso, 'Analysis of the ^{1}H and $^{13}C\{^{1}H\}$ NMR Spectral Parameters of the Tear Gases α-Chloroacetophenone, Dibenz[*b*,*f*][1,4]oxazepine, and 2-Chlorobenzylidene Malononitrile', *Magn. Reson. Chem.*, **34**, 989–994 (1996).
12. M. Mesilaakso, 'Application of NMR Spectroscopy to Environmental Analysis: Detection of Trace Amounts of Chemical Warfare Agents and Related Compounds in Organic Extract, Water, and Sand', *Environ. Sci. Technol.*, **31**, 518–522 (1997).
13. C. Albaret, D. Lœillet, P. Augé, P.-L. Fortier, 'Application of Two-Dimensional $^{1}H-^{31}P$ Inverse NMR Spectroscopy to the Detection of Trace Amounts of Organophosphorus Compounds Related to the Chemical Weapons Convention', *Anal. Chem.*, **69**, 2694–2700 (1997).
14. M. Mesilaakso, 'Chemical Weapons Convention, NMR Analysis', in *Encyclopedia of Environmental Analysis and Remediation*, John Wiley & Sons, Chichester, 949–971, 1998.
15. 'Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction'. Signed in January 1993. Printed and distributed by the OPCW/PTS. The Depositary of this Convention is the Secretary-General of the United Nations, from whom a certified true copy can be obtained.
16. G. Andersson, 'Analysis of Two Chemical Weapons Samples from the Iran–Iraq War', *NBC Defense Technol. Int.*, April, 62–65 (1986).
17. J.F. Weimaster, W.T. Beaudry, P.C. Bossle, M.W. Ellzy, L.G. Janes, D.W. Johnson, J.M. Lochner, S.G. Pleva, J.H. Reeder, D.K. Rohrbaugh, T.E. Rosso, L.J. Szafraniec, L.L. Szafraniec, T.G. Albro, W.R. Creasy, J.R. Stuff, P.B. Smith, I.R. Stewart, 'Chemical Analysis of Environmental Samples Collected in Iraq: Analysis for the Presence of Chemical Warfare Agents', *J. Chem. Technol. Biotechnol.*, **64**, 115–128 (1995).
18. M. Rautio (ed.), 'Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament', Ministry for Foreign Affairs of Finland, Helsinki, 1994.
19. W.R. Croasmun, R.M.K. Carlson (eds.), *Two-dimensional NMR Spectroscopy. Applications for Chemists and Biochemists*, VCH Publishers, New York, Sections 2.4.2, 2.4.5, 2.7.10, 2.8.12, 1994.
20. M. Batley, J.W. Redmond, '^{31}P NMR Reference Standards for Aqueous Samples', *J. Magn. Reson.*, **49**, 172–174 (1982).
21. M.L. Martin, G.J. Martin, J.-J. Delpuech, *Practical NMR Spectroscopy*, Heyden & Son, London, 182, 1980.
22. S. Braun, H.-O. Kalinowski, S. Berger, 'Heteronuclear NMR Spectroscopy', in *100 and More Basic NMR Experiments. A Practical Course*, VCH, Weinheim, 255, 1996.
23. A.E. Derome, 'Basic Experimental Methods', in *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, Chapter 3, 1987.
24. W.U. Primrose, in *NMR of Macromolecules. a Practical Approach*, ed. G.C.K. Roberts, Oxford University Press, Oxford, Chapter 2, 1993.

25. M.L. Martin, G.J. Martin, J.-J. Delpuech, *Practical NMR Spectroscopy*, Heyden & Son, London, Chapter 5, 1980.
26. M.L. Martin, G.J. Martin, J.-J. Delpuech, *Practical NMR Spectroscopy*, Heyden & Son, London, Chapter 9, 1980.
27. A.E. Derome, *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, 172–173, 1987.
28. A.E. Derome, *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, 162–173, 1987.
29. M.L. Martin, G.J. Martin, J.-J. Delpuech, 'Multiple Irradiation Experiments', in *Practical NMR Spectroscopy*, Heyden & Son, London, 235–236, 1980.
30. J.G. Verkade, L.D. Quin (eds.), *Phosphorus-31 NMR Spectroscopy in Stereochemical Analysis. Organic Compounds and Metal Complexes* (Methods in Stereochemical Analysis, Vol. 8), VCH Publishers, Deerfield Beach, FL, 1987.
31. J. Mason (ed.), *Multinuclear NMR*, Plenum Press, New York, 1987.
32. H. Friebolin, 'Double Resonance Experiments', in *Basic One- and Two-dimensional NMR Spectroscopy*, 2nd edition, VCH, Weinheim, 1993.
33. M. Mesilaakso, E.-L. Tolppa, P. Nousiainen, 'Analysis of ^{1}H and $^{13}C\{^{1}H\}$ NMR Spectral Parameters of Diphenylchloroarsine, Diphenylcyanoarsine, and 10-Chloro-5,10-dihydrophenarsazine. Identification of the Compounds Through Reference to Simulated Spectra', *Appl. Spectrosc.*, **51**, 733–737 (1997).
34. J.A.F. Compton, 'Biotoxins', in *Military Chemical and Biological Agents. Chemical and Toxicological Properties*, Telford Press, Caldwell, 343–347, 1987.
35. N. Niccolai, H.K. Schnoes, W.A. Gibbons, 'Study of the Stereochemistry, Relaxation Mechanisms, and Internal Motions of Natural Products Utilizing Proton Relaxation Parameters: Solution and Crystal Structures of Saxitoxin', *J. Am. Chem. Soc.*, **102**, 1513–1517 (1980).
36. Y. Shimizu, C.-p. Hsu, A. Genenah, 'Structure of Saxitoxin in Solutions and Stereochemistry of Dihydrosaxitoxins', *J. Am. Chem. Soc.*, **103**, 605–609 (1981).
37. 'Revised Standard Operating Procedure for Evaluation of the Results of OPCW Proficiency Tests', OPCW Technical Secretariat, The Hague, S/46/98, 21 April 1998.
38. J.K. Miettinen (ed.), 'Computer-Aided Techniques for the Verification of Chemical Disarmament. E.1. Verification Database. Methodology and Instrumentation for Sampling and Analysis in the Verification of Chemical Disarmament', Ministry for Foreign Affairs of Finland, Helsinki, 1988.
39. M. Godejohann, A. Preiss, C. Mügge, G. Wünsch, 'Application of On-Line HPLC-^{1}H NMR to Environmental Samples: Analysis of Groundwater Near Former Ammunition Plants', *Anal. Chem.*, **69**, 3832–3837 (1997).
40. U.G. Sidelmann, U. Braumann, M. Hofmann, M. Spraul, J.C. Lindon, J.K. Nicholson, S.H. Hansen, 'Directly Coupled 800 MHz HPLC–NMR Spectroscopy of Urine and Its Application to the Identification of the Major Phase II Metabolites of Tolfenamic Acid', *Anal. Chem.*, **69**, 607–612 (1997).
41. S.A. Korhammer, A. Bernreuther, 'Hyphenation of High-performance Liquid Chromatography (HPLC) and Other Chromatographic Techniques (SFC, GPC, GC, CE) with Nuclear Magnetic Resonance (NMR): a Review', *Fresenius' J. Anal. Chem.*, **354**, 131–135 (1996).
42. U. Braumann, H. Händel, K. Albert, R. Ecker, M. Spraul, 'On-Line Monitoring of the Supercritical Fluid Extraction Process with Proton Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **67**, 930–935 (1995).
43. 'The World's First 800 MHz Spectrometer Installed in a Customer's Laboratory', Bruker Rep., **142**, 28–29, 1996.
44. '^{1}H Nano.nmr Probe for 500 MHz NMR Systems', Brochure 814040-001 MAG-7029/874 5/93, Varian, Palo Alto, CA, 1993.

Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention

Marja-Leena Kuitunen
Finnish Institute for Verification of the Chemical Weapons Convention, University of Helsinki, Finland

1 Introduction 1
2 Preparation of Samples for Off-site Analysis 3
2.1 Air Samples 3
2.2 Aqueous Liquid Samples 3
2.3 Soil Samples 4
2.4 Wipe Samples 6
2.5 Active Charcoal Samples 6
2.6 Concrete Samples 7
2.7 Paint, Rubber, and other Polymeric Samples 8
3 Preparation of Samples for On-site Analysis 9
3.1 Aqueous Liquid Samples 9
3.2 Soil Samples 10
3.3 Wipe Samples 11
4 Quality Control 12
5 Safety 12
6 International Comparison and Proficiency Tests 12
Acknowledgments 14
Abbreviations and Acronymns 15
Related Articles 15
References 15

This article discusses the preparation of environmental samples to be analyzed for chemicals related to the chemical weapons convention (CWC), where analysis takes place on-site or off-site by gas chromatography (GC) or liquid chromatography (LC) as an introduction technique to spectrometers or by capillary zone electrophoresis (CZE). The specific sample preparation methods required for analysis by spectroscopic techniques such as Fourier transform infrared (FTIR) and nuclear magnetic resonance (NMR) spectroscopy are dealt with in separate articles. Preparations of analytical samples from air, aqueous liquid (water), soil, active charcoal, wipe, concrete, paint, rubber, and other polymeric samples for off-site analysis are presented in section 2, and preparations of analytical samples from aqueous liquid (water), soil, and wipe samples for on-site analysis, are covered in section 3. Preparations for on-site analysis are as described in the standard operating procedures (SOPs) of the international organization responsible for verifying the Convention (Technical Secretariat of the Organization for the Prohibition of Chemical Weapons) and those for off-site analysis are as described in the corresponding recommended operating procedures (ROPs). Section 4 examines the question of quality control (QC) in sample preparation and section 5 the important matter of safety in handling samples containing CWC-related chemicals. Section 6 reports on the use of the methods in international comparison and proficiency tests.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The reliable verification of chemicals related to the CWC depends essentially on the collection of good samples and well-planned, effective, and reasonably simple sample preparations suitable for the method of analysis. The collection of good samples is arguably the most critical part of a successful analysis and this is discussed in a separate article (see **Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention**). For its part, proper preparation of samples requires a thorough understanding of the behavior of the various types of chemical in different sample matrices, both before and during the sample preparation, and an awareness of the limitations of the chosen instrumental method of analysis.

Mass spectrometry (MS), infrared (IR) spectroscopy, and NMR spectroscopy with their numerous applications are the main instrumental techniques for the detection and identification of CWC-related chemicals. During the last few years, however, less laborious techniques such as LC and CZE have become attractive for the analysis of water samples and extracts where sample preparation is either not required or is relatively simple.

Sample preparation procedures and analytical techniques for the off-site laboratories of the CWC have been developed and tested in five international interlaboratory comparison (round-robin) tests,[1–5] in two trial proficiency tests, and in three official proficiency tests (see **Verification of Chemicals Related to the Chemical Weapons Convention**). The ROPs for sampling and sample preparation[6,7] were written and updated on the basis of the results of the round-robin tests. The ROPs (see Table 1) were designed to be comprehensive

Table 1 ROPs for sampling and sample preparation in the verification of CWC-related chemicals[7]

Procedure	Definition
	General techniques
GT 5	ROP for cleaning glassware for environmental sampling
	General sampling
GS 1	General procedures for sampling
GS 2	ROP for packing samples containing chemical agents
GS 3	ROP for packing environmental samples
GS 4	Recommended coding procedure for samples
GS 5	ROP for handling background and control samples
GS 6	Recommended procedure for guaranteeing sample integrity
	Sample collection
SC 1	ROP for sampling and analysis of low-volume Tenax air samples
SC 4	ROP for collection of low-volume XAD-2 air samples
SC 5	ROP for collection of soil samples
SC 6	ROP for taking aqueous liquid samples
SC 7	ROP for taking liquid samples
SC 8	ROP for analysis of solid material samples by thermodesorption
	Sample preparation
SP 3	ROP for preparation of low-volume XAD-2 air samples
SP 4	ROP for preparation of soil samples
SP 5	ROP for preparation of wipe samples
SP 6	ROP for preparation of active charcoal samples
SP 7	ROP for preparation of aqueous liquid samples
SP 8	ROP for preparation of concrete samples
SP 9	ROP for preparation of paint, rubber, and other polymeric samples
SP 10	Quality assurance and QC in sample preparation

enough to allow the analysis of all CWC-related chemicals. Accordingly, some of the procedures contain many sample preparation steps. It is also recommended that the ROPs should be used as first choice in the analysis, with other approaches not excluded.

Bulk samples of CWC-related chemicals are analyzed with minimum sample preparation by FTIR spectroscopy (see **Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention**). At present no on-site SOPs or off-site ROPs are available for bulk samples.

Air samples are usually collected to solid adsorbents such as Tenax, or XAD resins, graphitized carbons (e.g. Carbopak), active charcoal, or porous polymers (e.g. Chromosorb). The chemicals are eluted from the adsorbent to a liquid or gas phase by liquid–solid elution or extraction or by thermal desorption. Extraction is the most commonly used method. Thermal desorption can be applied when GC is used for analysis.

Air sampling for volatile CWC-related chemicals was intensively developed at our Institute during the 1980s.[8–10] A good review of the preparation of air samples was included in a review article presented in 1990.[11] Although other sample matrices have aroused increasing international interest during the 1990s, the methods presented in the review continue to be highly useful. Volatile nonpolar CWC-related chemicals can be analyzed directly after elution or thermal desorption of the adsorbent tube. The ROPs include sampling, sample preparation, and analysis instructions for low-volume Tenax and XAD-2 air samples.

Only the preparation of an XAD-2 low-volume air sample is presented in this article, while the thermal desorption of a Tenax tube is described in the context of gas chromatographic analysis (see **Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention**). Active charcoal is such a strong adsorbent that it requires more effective extraction methods than XAD-2 resin or Tenax tubes. Thus the recoveries of CWC-related chemicals tend to be lower than from other air sampling materials. Furthermore, active charcoal is not usually used for the collection of organophosphorus chemicals.The sample preparation methods for active charcoal samples have not been validated in interlaboratory or proficiency tests.

CWC-related chemicals in aqueous liquid samples (water samples) are usually recovered by extraction with organic solvent. Modern methods such as solid-phase extraction (SPE) and solid-phase microextraction (SPME) have also been presented.[12–18] Organic extractions and these modern methods mainly recover nonpolar CWC-related chemicals, but leave behind the water-soluble and nonvolatile chemicals. These must also be recovered, however, because the agents tend to decompose (hydrolyze) rapidly under conditions in the environment. Techniques such as CZE and LC relying on element specific or mass spectrometric detection have been developed to provide easy and effective ways of recovering these chemicals from water samples with only minor sample preparation.[19–25] For GC/MS analysis the water must be displaced and the analytes derivatized.

Soil samples have proved to be of critical importance in confirming the use of chemical warfare agents. Little has been published about the preparation of soil samples in the open literature, however.[11,26–30] Usually the soil is extracted with organic solvent to recover the nonpolar CWC-related chemicals such as nerve and mustard agents and, in view of the probable degradation (hydrolysis), it

Table 2 SOPs for sampling and sample preparation in the verification of CWC-related chemicals[31]

SOP	Sample preparation
SP-003	SOP for aqueous sample preparation on-site for GC/MS analysis
SP-004	SOP for soil sample preparation on-site for GC/MS analysis
SP-005	SOP for wipe sample preparation on-site for GC/MS analysis

is also extracted with water to recover the water-soluble polar CWC-related chemicals.

The procedures used in off-site laboratories had to be modified for on-site use in order to reduce the amount and weight of solvents, reagents, and equipment. At present, three different procedures are being developed, tested, and validated for on-site inspections (Table 2).[31] The existing SOPs by no means cover all analytical tasks required in on-site situations, however, and method development and improvement are continuing.

Existing methods for the on-site analysis of CWC-related chemicals in environmental samples are based on GC/MS. Methods based on liquid chromatographic separation are not yet available. Before GC analysis, the chemicals must be extracted from the sample matrix. Volatile analytes can be extracted with organic solvents, concentrated, and then analyzed directly, whereas non-volatile and water-soluble or more polar analytes must be extracted with water or other polar solvent, the solvent evaporated, and the analytes derivatized.

This article covers the off-site and on-site sample preparation methods for analytical techniques used in off-site and on-site laboratories. The procedures described are from ROPs[7] and from procedures followed at the Institute for the off-site methods and from the SOPs[31] for the on-site methods. The usefulness of the methods as demonstrated in international comparison and proficiency tests is noted.

2 PREPARATION OF SAMPLES FOR OFF-SITE ANALYSIS

ROPs have been developed and validated for air, aqueous liquid (water), soil, wipe, active charcoal, concrete, paint, rubber, and other polymeric samples for off-site analysis.

2.1 Air Samples

It is recommended that the control zone and the collection zone of a XAD-2 tube be analyzed separately if the breakthrough volume of the most volatile chemicals of interest in the sampling conditions is not known. The control zone is removed from the tube to a small glass vial and sonicated with 2 mL of ethyl acetate for 3 min. The resin is filtered rapidly and the supernatant is analyzed. The trapped chemicals are desorbed from the collection zone by eluting 2 mL of ethyl acetate to the back end of the zone. (It is important to note that the elution direction should be opposite to the collection direction.) The eluate is collected in a glass vial under gravity.

If the breakthrough volume is known and was taken into account during sample collection, the whole tube can be eluted as described above for the collection zone. If required, the sample solutions can be concentrated under mild nitrogen flow. Care must be taken that the sample solutions are never concentrated to dryness because certain CWC-related chemicals are adsorbed firmly on to glass surfaces from residues of organic extracts.

It should be noted that highly volatile chemicals such as hydrogen cyanide (CAS 74-90-8) and phosgene (CAS 75-44-5) as well as polar water-soluble chemicals such as alkylphosphonic and alkylthiophosphonic acids, thiodiglycol (CAS 111-48-8), and aminoalcohols are not quantitatively trapped by XAD-2 resin.

2.2 Aqueous Liquid Samples

The original sample is divided into four or more portions (1–10 mL), if possible, for different analytical purposes and transferred to screw-capped glass vials. The pH of the first portion (Figure 1, fraction 1A) is

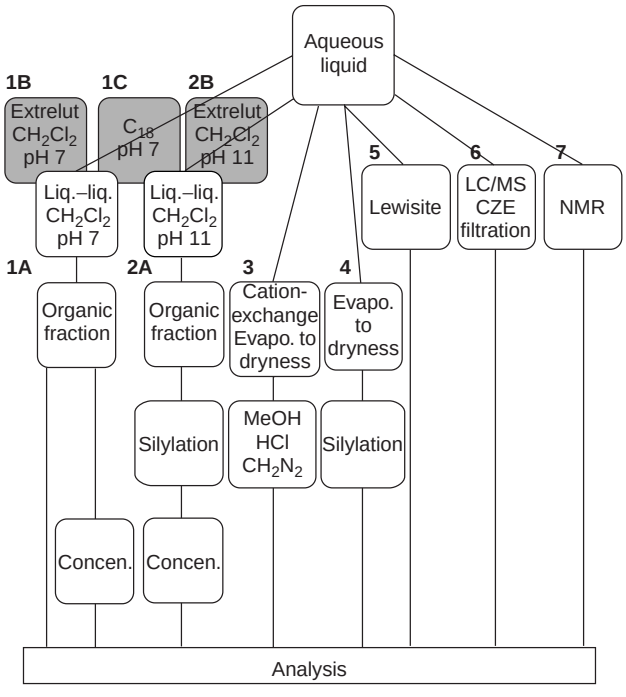

Figure 1 Schematic presentation of the ROP for aqueous liquid sample preparation.

determined with a universal pH paper and, if necessary, the sample is neutralized with ammonium hydroxide or hydrochloric acid. The neutralized sample is liquid–liquid extracted with two portions of water-immiscible organic solvent using a 1:2 volume ratio of solvent to sample. Usually dichloromethane is used as extraction solvent and liquid–liquid extraction times are 2–5 min. The extracts are combined and dried. The dried extract is analyzed as such or, where necessary, after approximately 10-fold concentration with mild nitrogen flow. Concentration to dryness must be avoided because certain CWC-related chemicals are firmly adsorbed to glass surfaces from residues of organic extracts. This dichloromethane sample is analyzed for nonpolar CWC-related chemicals.

The aqueous fraction from the first extraction or another portion of the original sample is made alkaline with ammonium hydroxide and the same liquid–liquid extraction procedure is repeated (Figure 1, fraction 2A). This extract is analyzed after derivatization, for example, with 100 μL of *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA, CAS 25561-30-2) for 1 mL extract. This sample is analyzed for alkaline CWC-related chemicals.

Alternatively (Figure 1, fractions 1B and 2B), both the neutralized and alkaline liquid–liquid extraction steps can be performed using Extrelut® (Merck) or another corresponding extraction cartridge under similar conditions to those of the liquid–liquid extraction. A portion of the sample is eluted into the cartridge and the analytes of interest are eluted from the sorbent with a suitable solvent (e.g. dichloromethane). The cartridges are used according to the instructions for use delivered by the cartridge manufacturer.

Further, a C_{18} SPE cartridge may be used (Figure 1, fraction 1C) instead of the neutralized liquid–liquid extraction step. A C_{18} cartridge (100 or 200 mg) is conditioned according to the instructions delivered by the manufacturer and a portion of the sample is eluted through the cartridge. Most of the water deposited in the cartridge is removed by sucking air through and, if necessary, interfering matrix components are removed by a wash solvent (e.g. water). Note that washing may also remove chemicals of interest. The cartridge is eluted with 2 mL of high-purity solvent (e.g. acetone). The eluate is dried, analyzed, and, if necessary, concentrated.

The aqueous fraction from the neutral or alkaline liquid–liquid extraction or another portion of the original sample is eluted slowly through a conditioned SCX (100 or 200 mg) cation-exchange cartridge (Figure 1, fraction 3). The pH of the eluate is determined to test the completeness of the cation exchange, which is repeated, if necessary. The eluate is evaporated to dryness on a rotary evaporator at 50 °C and 366 mPa. The residue is dissolved in dry acidic methanol and the solution is methylated with diazomethane. This sample is analyzed for polar alkylphosphonic and alkylthiophosphonic acids. Aminoalcohols remain in the SCX cartridge and cannot be recovered.

Another fraction from the neutral or alkaline liquid–liquid extraction or another portion of the original sample is evaporated to dryness, but without cation exchange, as described above (Figure 1, fraction 4). The residue is dissolved in acetonitrile or other solvent suitable for silylation. BSTFA is added and the silylation is completed by heating the sample at 60 °C for 30 min. The cooled sample is analyzed for polar CWC-related chemicals such as thiodiglycol and aminoalcohols. Polar alkylphosphonic and alkylthiophosphonic acids can also be recovered from this fraction, but since the procedure does not contain a cation-exchange step, a large number of inorganic cations in the sample may mean that the recoveries of smaller acids such as methylphosphonic acid (CAS 993-13-5) are low, or the chemicals may not be recovered at all.

Methyl-*N*-*tert*(butyldimethylsilyl)trifluoroacetamide (MTBSTFA, CAS 77377-52-7) is another common silylation reagent used for CWC-related chemicals. BSTFA silylates the aminoalcohols such as triethanolamine (CAS 102-71-6) giving higher recoveries than MTBSTFA, but the *tert*-butyldimethylsilyl ethers formed when MTBSTFA is used are more stable than the trimethylsilyl ethers formed with BSTFA.

The aqueous fraction from the neutral or alkaline liquid–liquid extraction or another portion of the original sample is prepared for analysis of the sample for Lewisite 1 (CAS 541-25-3) and Lewisite oxide (CAS 3088-37-7) (Figure 1, fraction 5). The pH of the sample is adjusted with hydrochloric acid to pH 2, the sample is filtered, and 0.1 mL of freshly prepared solution containing 3,4-dimercaptotoluene (DMT, CAS 496-74-2) 5 mg mL^{-1} in acetone is added. The sample vial is shaken vigorously and allowed to stand for 10 min. One milliliter of *n*-hexane is added and the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water fraction is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

In addition, an appropriate portion of the aqueous liquid sample is filtered through a high-performance liquid chromatography (HPLC) filter and the filtrate is analyzed by LC/MS and CZE (Figure 1, fraction 6) and another portion is prepared for NMR analysis (fraction 7).

2.3 Soil Samples

If possible, the original sample is divided into four or more portions (1–10 g) for different analytical purposes, each placed in a screw-capped glass vial. One portion

of the soil (Figure 2, fraction 1) is extracted twice with dichloromethane by sonication, shaking, tumbling, or agitation for 10 min, each time using the same volume in milliliters (mL) of the solvent as the amount of the sample in grams (g). The extracts are centrifuged, filtered, combined, and dried. The dried extract is analyzed as such or, where necessary, after approximately 10-fold concentration under mild nitrogen flow. Concentration to dryness must be avoided since certain CWC-related chemicals are firmly adsorbed to glass surfaces from residues of organic extracts. This sample is analyzed for nonpolar CWC-related chemicals.

A new portion of soil, or the soil fraction from the dichloromethane extraction (after drying in a fume cupboard for approximately 1 h), is extracted with distilled deionized water in the manner described above for the dichloromethane extract. Usually sonication is avoided because it may reduce recoveries of the polar CWC-related chemicals by creating new active sites for analytes in the soil. The extracts are centrifuged, filtered, and combined and the pH is determined.

An appropriate portion of the water extract (Figure 2, fraction 2) is eluted slowly through a conditioned SCX (100 or 200 mg) cation-exchange cartridge. The pH of the eluate is determined to test the completeness of the cation exchange and the eluate is evaporated to dryness on a rotary evaporator at 50 °C and 366 mPa. The residue is dissolved in dry acidic methanol and the solution is methylated with diazomethane. This sample is analyzed for polar alkylphosphonic and alkylthiophosphonic acids. Aminoalcohols remain in the SCX cartridge and cannot be recovered.

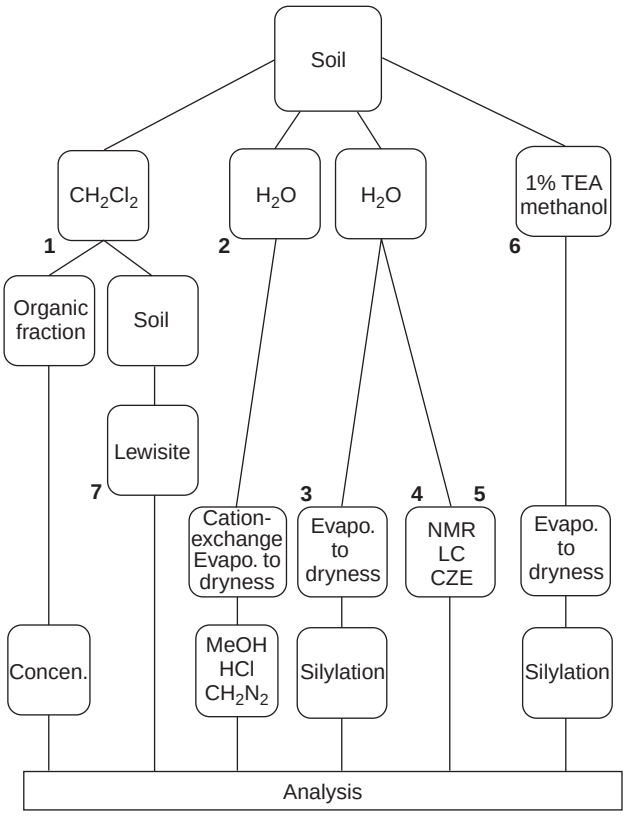

Figure 2 Schematic presentation of the ROP for soil sample preparation.

The extraction procedure with water is repeated, or another portion of a water extract is evaporated to dryness, but without cation exchange, as described above. The residue is dissolved in acetonitrile or other solvent suitable for silylation and silylated with BSTFA. The silylation is completed by heating the sample at 60 °C for 30 min (Figure 2, fraction 3). This sample is analyzed for polar CWC-related chemicals such as thiodiglycol and aminoalcohols. Polar alkylphosphonic and alkylthiophosphonic acids can also be recovered from this fraction, but since the procedure does not include a cation-exchange step, if there are large amounts of inorganic cations in the sample, the recoveries of smaller acids such as methylphosphonic acid may be low or they may not be recovered at all.

MTBSTFA is also commonly used as a silylation reagent for the CWC-related chemicals (see section 2.2).

Next, an appropriate portion of a water extract(s) is filtered through an HPLC filter and analyzed by LC/MS and CZE (Figure 2, fraction 4) and another portion is prepared for NMR analysis (fraction 5).

In the next step (Figure 2, fraction 6) a new portion of soil or the soil fraction from the water extraction step is extracted with 1% triethylamine (TEA, CAS 121-44-8) in methanol in the manner described above for dichloromethane. The centrifuged, filtered, and combined TEA/methanol extract is evaporated to dryness on a rotary evaporator at 50 °C and 31 Pa for ca. 10 min and then at 50 °C and 366 mPa for ca. 20 min. The evaporation residue is silylated as described above for the evaporation residue of the water extract. This sample is analyzed for alkaline CWC-related chemicals such as VX (CAS 50782-69-9).

Another sample portion, or the soil portion after dichloromethane extraction, is prepared for analysis of the sample for Lewisite 1 and Lewisite oxide (Figure 2, fraction 7). The soil is mixed with hydrochloric acid (e.g. 10 g of soil with ca. 15 mL of HCl). The pH of the water fraction is determined and adjusted to pH 2 if necessary. The pH of the water fraction must be maintained between 1.5 and 2.5 over a period of 30 min. The sample is then shaken vigorously for 30 s every 5 min for 60 min, allowed to settle for 1 h, centrifuged, and filtered. To the filtrate is added 0.1 mL of freshly prepared solution containing DMT 5 mg mL^{-1} in acetone. The sample vial is shaken vigorously and allowed to stand for 10 min, and then

1 mL of *n*-hexane is added and the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

2.4 Wipe Samples

The wipe sample is inserted into a glass vial or the original sample container is used as an extraction vessel. If blank sample is available, a small amount of this is tested for solubility in the selected solvent. Nonpolar organic solvent is poured over the sample. Acetone, ethyl acetate, dichloromethane, and deuterated chloroform (NMR analysis) are all good solvents. Care must be taken that the wipe sample is completely covered with the solvent. The mixture is sonicated for 3 min. The organic phase is quickly decanted and, if necessary, filtered and centrifuged. The extraction procedure is repeated with a second portion of the solvent. The extracts are combined and analyzed and, if necessary, an appropriate amount of the extract is concentrated approximately 10-fold with mild nitrogen flow. Concentration to dryness must be avoided since certain CWC-related chemicals are firmly adsorbed on to glass surfaces from residues of organic extracts (Figure 3, fraction 1). This extract is analyzed for nonpolar CWC-related chemicals. Since a wipe sample is not a very adsorptive sample matrix one can also expect to recover polar CWC-related chemicals and a 0.5–1 mL portion of the extract can be derivatized with 100 μL of BSTFA for their analysis.

The wipe sample from the organic solvent extraction is then extracted with polar solvent (e.g. water, methanol, or acetonitrile) in a manner described above for nonpolar solvent (Figure 3, fraction 2A). For analysis with GC and GC hyphenated techniques, the polar extract must be derivatized. A part of the acetonitrile extract can be silylated with BSTFA in the same way as the organic extract of the wipe sample, although methanol or water extracts cannot be silylated directly and must first be evaporated to dryness. An appropriate part of the extract is evaporated to dryness on a rotary evaporator (water extract at 50 °C and 366 mPa for ca. 30 min and methanol extract at 50 °C and 31 Pa for ca. 10 min followed at 50 °C and 366 mPa for ca. 20 min). The methanol extract can also be evaporated under mild nitrogen flow. Over the evaporation residue 0.5 mL of acetonitrile and 0.5 mL of BSTFA are added and the silylation mixture is heated at 60 °C for 30 min to complete the silylation. This sample is analyzed for polar CWC-related chemicals such as thiodiglycol, aminoalcohols, and polar alkylphosphonic and alkylthiophosphonic acids.

A suitable portion of a water, methanol, or acetonitrile extract can also be analyzed by LC/MS or CZE after

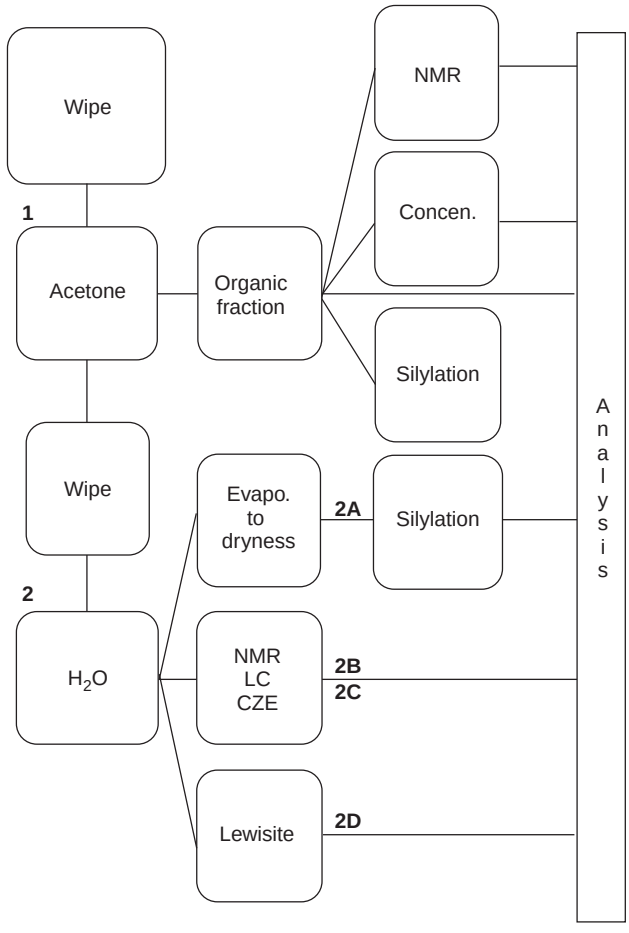

Figure 3 Schematic presentation of the ROP for wipe sample preparation.

filtration through a HPLC filter (Figure 3, fraction 2B), and another portion can be prepared for NMR analysis (fraction 2C).

An appropriate portion of the polar extract of the wipe sample is prepared for analysis of the sample for Lewisite 1 and Lewisite oxide (Figure 3, fraction 2D). The pH of the sample is adjusted with hydrochloric acid to pH 2, the sample is filtered, and 0.1 mL of freshly prepared solution containing DMT 5 mg mL^{-1} in acetone is added. The sample vial is shaken vigorously and allowed to stand for 10 min, 1 mL of *n*-hexane is added, and the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

2.5 Active Charcoal Samples

Active charcoal sometimes is doped with various chemicals to modify its properties for a specific purpose. This

may cause degradation of the sample chemicals and extraction problems. The matrix may also contain substantial amounts of water. For these reasons the origin and background of active charcoal samples should always be ascertained.

The active charcoal material is placed in a glass vial. The extraction solvent is added (twice the volume of the charcoal) and the mixture is sonicated for 3 min. Suitable solvents are acetone, dichloromethane, carbon disulfide, and deuterated chloroform (for NMR analysis). The organic phase is quickly decanted and, if necessary, filtered and centrifuged. The extracts are analyzed and, if necessary, an appropriate amount of the extract is concentrated approximately 10-fold with mild nitrogen flow.

The sample or a few grains of it can also be analyzed by thermal desorption (see **Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention**) if the origin and nature of the charcoal material and the sampling procedure are known.

2.6 Concrete Samples

Small pieces of a concrete sample may be extracted as such, but if the sample pieces are large, homogenizing by crushing will be necessary before extraction. A piece of concrete (ca. 10 g) is inserted into a glass bottle, 10 mL of acetone or dichloromethane is added, and the sample mixture is sonicated for 30 min (Figure 4, fraction 1). More solvent must be used if the sample is not covered by the solvent. To keep the bath at ambient temperature, the temperature of the water bath is monitored and water is added, if necessary. The organic phase is quickly decanted and, if necessary, centrifuged and filtered. The extraction procedure is repeated with a second portion of the solvent. The extracts are combined and analyzed and, if necessary, an appropriate amount of the extract is concentrated approximately 10-fold under mild nitrogen flow. Concentration to dryness must be avoided since certain CWC-related chemicals are firmly adsorbed to glass surfaces from residues of organic extracts. This extract is analyzed for nonpolar CWC-related chemicals.

The concrete sample from the organic extraction or a new portion of 10 g of the original sample is then extracted with distilled, deionized water in the manner described above for organic solvent (Figure 4, fraction 2). This sample is analyzed for polar CWC-related chemicals. For analysis with GC and GC hyphenated techniques the water extract must first be derivatized: an appropriate part of the extract is evaporated to dryness on a rotary evaporator at 50 °C and 366 mPa for ca. 30 min and then 0.5 mL of acetonitrile and 0.5 mL of BSTFA are poured over the evaporation residue and the silylation

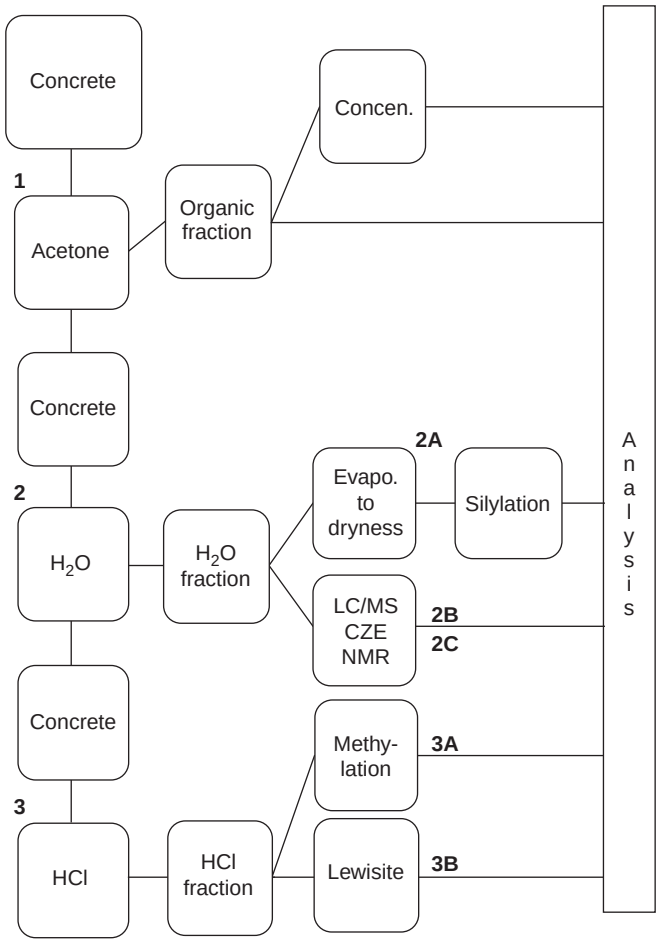

Figure 4 Schematic presentation of the ROP for concrete sample preparation.

mixture is heated at 60 °C for 30 min to complete the silylation (Figure 4, fraction 2A). This sample is analyzed for polar CWC-related chemicals such as thiodiglycol, aminoalcohols, and polar alkylphosphonic and alkylthiophosphonic acids. The evaporation residue may also be methylated instead: the residue is dissolved in dry acidic methanol and the solution is methylated with diazomethane. This sample is analyzed for polar alkylphosphonic and alkylthiophosphonic acids.

A suitable portion of the water extract can be analyzed by LC/MS or CZE after filtration through a HPLC filter (Figure 4, fraction 2B) and another portion can be prepared for NMR analysis (fraction 2C).

The concrete sample from the water extraction or a new portion of 10 g of the original sample is then extracted in a similar way with 1 M HCl solution (Figure 4, fraction 3A). This extract is handled in the same way as the water extract, for the same analytes. Note that silylation of acidic evaporation residues may be difficult and neutralization of the extract before evaporation to dryness is recommended

if the HCl extract is to be silylated to recover chemicals such as thiodiglycol and aminoalcohols.

An appropriate portion of the HCl extract of the concrete sample is prepared for analysis of the sample for Lewisite 1 and Lewisite oxide (Figure 4, fraction 3B). The pH of the sample is adjusted with hydrochloric acid to pH 2, the sample is filtered, and 0.1 mL of freshly prepared solution containing DMT 5 mg mL^{-1} in acetone is added. The sample vial is shaken vigorously and then allowed to stand for 10 min. After addition of 1 mL of *n*-hexane the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

2.7 Paint, Rubber, and other Polymeric Samples

The sample is crushed or cut into small pieces, if necessary, and placed into a glass bottle. Then 10 mL of acetone (dichloromethane, acetone-d_6, or $CDCl_3$) is added and the sample mixture is sonicated for 10 min (Figure 5, fraction 1). Solvent must be sufficient to cover the sample. Some paint, rubber, and polymeric matrices may, however, be soluble in or swell in dichloromethane

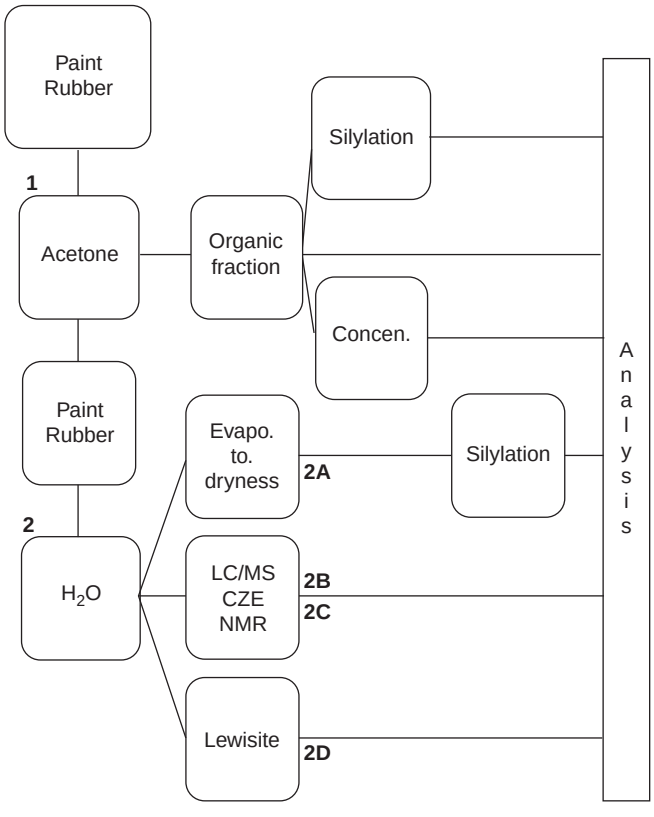

Figure 5 Schematic presentation of the ROP for paint, rubber, or other polymeric sample preparation.

or chloroform and the solubility of the sample in the extraction solvent should be tested before the solvent is used, especially if a blank sample is available. The organic phase is quickly decanted and, if necessary, centrifuged and filtered. The extraction procedure is repeated with an additional portion of the solvent. The extracts are combined and analyzed and, if necessary, an appropriate amount of the extract is concentrated approximately 10-fold under mild nitrogen flow. Concentration to dryness must be avoided since certain CWC-related chemicals are firmly adsorbed on to glass surfaces from residues of organic extracts. This extract is analyzed for nonpolar CWC-related chemicals. Paint, rubber, and other polymeric samples are not very adsorptive sample matrices. For this reason one could also expect to recover some polar CWC-related chemicals from this fraction. To find these, a portion of 0.5–1 mL of the extract can be derivatized with 100 μL of BSTFA.

The paint, rubber, or other polymeric sample from the organic extraction, or a new portion of the original sample is extracted with distilled deionized water in the manner described above for organic solvent (Figure 5, fraction 2). This sample is analyzed for polar CWC-related chemicals. For analysis with GC and GC hyphenated techniques the polar extract must first be derivatized: an appropriate part of the extract is evaporated to dryness on a rotary evaporator at 50 °C and 366 mPa for ca. 30 min. Then 0.5 mL of acetonitrile and 0.5 mL of BSTFA are poured over the evaporation residue and the silylation mixture is heated at 60 °C for 30 min to complete the silylation (fraction 2A). This sample is analyzed for polar CWC-related chemicals such as thiodiglycol, aminoalcohols, and polar alkylphosphonic and alkylthiophosphonic acids. The evaporation residue may also be methylated: the residue is dissolved in dry acidic methanol and the solution is methylated with diazomethane. This sample is analyzed for polar alkylphosphonic and alkylthiophosphonic acids.

A suitable portion of the water extract can be analyzed by LC/MS or CZE after filtration through a HPLC filter (Figure 5, fraction 2B) and another portion can be prepared for NMR analysis (fraction 2C).

Another suitable portion of the water extract is prepared for analysis of the sample for Lewisite 1 and Lewisite oxide (Figure 5, fraction 2D). The pH of the sample is adjusted with hydrochloric acid to pH 2, the sample is filtered, and 0.1 mL of freshly prepared solution containing DMT 5 mg mL^{-1} in acetone is added. The sample vial is shaken vigorously and allowed to stand for 10 min. After 1 mL of *n*-hexane is added the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

3 PREPARATION OF SAMPLES FOR ON-SITE ANALYSIS

To date, SOPs have been developed and validated for on-site analysis of aqueous liquid (water), soil, and wipe samples.[31] Method development for other matrices such as air and improvement of the existing methods are continuing.

3.1 Aqueous Liquid Samples

A 10 mL portion of aqueous liquid is taken into a screw-capped vial. The pH of the sample is determined with a universal pH paper and, if necessary, the sample is neutralized with ammonium hydroxide or hydrochloric acid. The sample is extracted twice with 5 mL of dichloromethane, with shaking. The dichloromethane extracts are combined and, after drying, 1 mL of the extract is set aside for GC/MS analysis as such, and the rest is concentrated approximately 10-fold with mild nitrogen flow and analyzed (Figure 6, fraction 1). These dichloromethane samples are analyzed for nonpolar CWC-related chemicals.

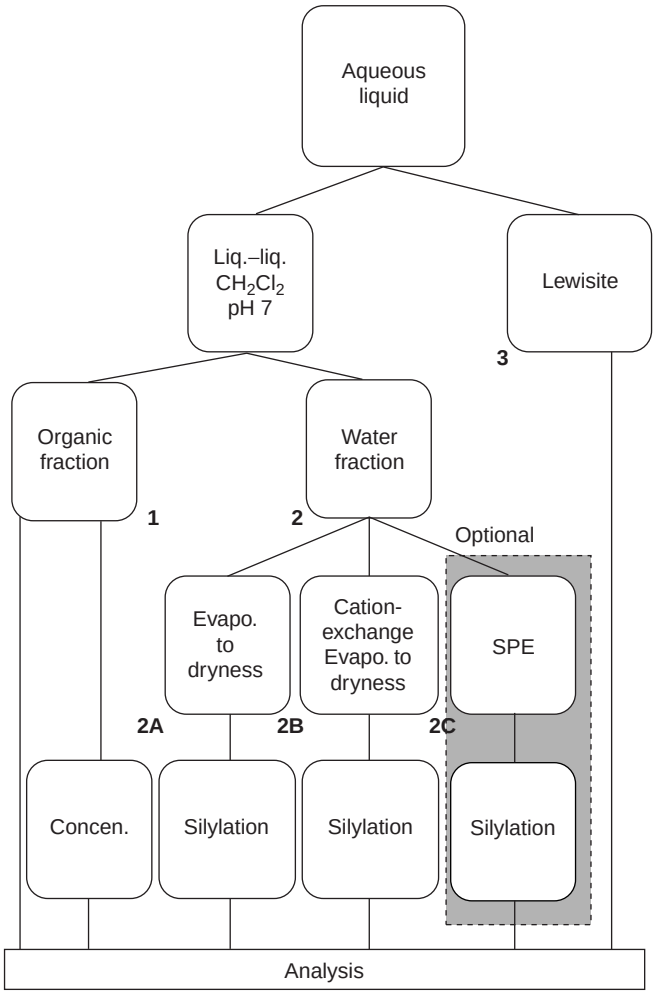

Figure 6 Schematic presentation of the SOP for aqueous liquid sample preparation.

The aqueous fraction remaining after dichloromethane extraction is divided into two equal portions. One portion (Figure 6, fraction 2A) is evaporated to dryness in a centrifugal evaporator, 0.5 mL of acetonitrile and 0.5 mL of BSTFA are added to the evaporation tube, and the silylation mixture is heated at 60 °C for 30 min to complete the silylation. This sample is analyzed for polar CWC-related chemicals such as thiodiglycol and aminoalcohols.

The other portion of the aqueous fraction (Figure 6, fraction 2B) is eluted slowly through a conditioned SCX (500 mg) cation-exchange cartridge. The pH of the eluate is determined to test the completeness of the cation exchange and, if necessary, the eluate is neutralized. The neutral eluate is evaporated and silylated as described above for the first aqueous portion. This sample is analyzed for polar alkylphosphonic and alkylthiophosphonic acids. Aminoalcohols remain in the SCX cartridge and cannot be recovered from this sample.

An optional SPE procedure (Figure 6, fraction 2C) for aqueous samples is included in the SOPs for on-site analysis. The pH of a 5 mL portion of the aqueous fraction is adjusted with hydrochloric acid and ammonium hydroxide to pH 4.5. The sample is eluted slowly through a conditioned SCX (500 mg) cation-exchange cartridge and after that through a conditioned NH_2 (100 mg) anion-exchange cartridge. The retained analytes are eluted slowly from the NH_2 cartridge with 2 mL of 2% hydrochloric acid in methanol. The eluate is adjusted to pH 8 with concentrated ammonia and then evaporated to dryness with mild nitrogen flow. The evaporation residue is silylated as described above for the first aqueous portion to provide the sample in a form suitable for GC/MS analysis. Aminoalcohols remain in the SCX cartridge and cannot be recovered from this sample. This method was developed for analysis of samples for polar alkylphosphonic and alkylthiophosphonic acids. The effectiveness of the method varies very much with the sample matrix, however. Depending, among other things, on the capacity and quality of the cation- and anion-exchange cartridges and the amount and character of impurities in the sample, the recovery values for alkylphosphonic and alkylthiophosphonic acids may be very low. Accordingly, this method should be considered as an optional technique only for sample matrices of known ion content.

A new sample portion of 10 mL is prepared for analysis of the sample for Lewisite 1 and Lewisite oxide (Figure 6, fraction 3). The pH of the sample is adjusted with hydrochloric acid to pH 2, the sample is filtered, and 0.1 mL of freshly prepared solution containing DMT 5 mg mL^{-1} in acetone is added. The sample vial is

shaken vigorously and allowed to stand for 10 min. One milliliter of *n*-hexane is added and the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

3.2 Soil Samples

A 5 g portion of soil is extracted with 5 mL of dichloromethane by shaking manually for 1 min. The extract is centrifuged and filtered and the extraction procedure is repeated with an additional 5 mL of dichloromethane. The extracts are combined and dried. One milliliter of the dried extract is set aside for GC/MS analysis as such and 3 mL is concentrated approximately 10-fold under mild nitrogen flow and analyzed (Figure 7, fraction 1). These dichloromethane samples are analyzed for nonpolar CWC-related chemicals.

The soil sample from the dichloromethane extraction is placed in a fume cupboard and allowed to dry for approximately 1 h. The sample is then extracted twice with 5 mL of distilled deionized water in the manner described above for dichloromethane. The centrifuged, filtered, and combined water extract is divided into two equal portions (Figure 7, fraction 2).

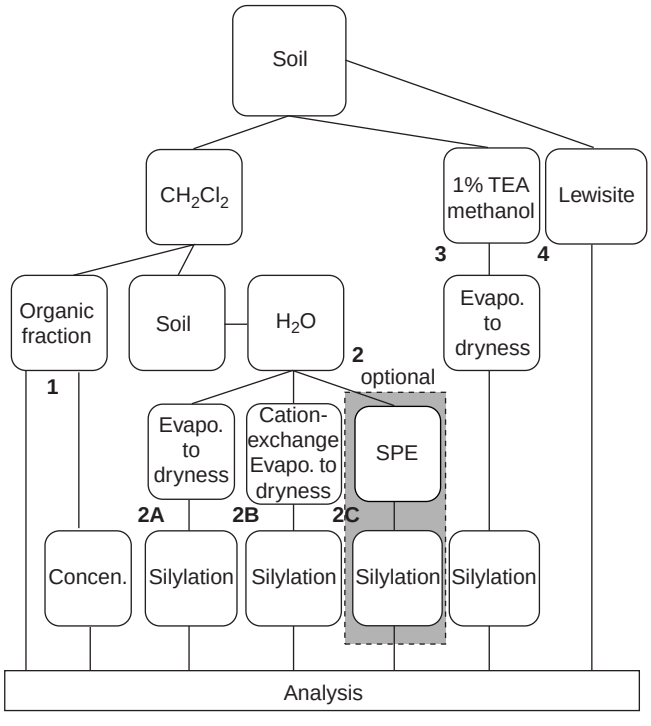

Figure 7 Schematic presentation of the SOP for soil sample preparation.

The first portion (Figure 7, fraction 2A) is evaporated to dryness in a centrifugal evaporator, 0.5 mL of acetonitrile and 0.5 mL of BSTFA are added to the evaporation tube, and the silylation mixture is heated at 60 °C for 30 min to complete the silylation. This sample is analyzed for polar CWC-related chemicals such as thiodiglycol and aminoalcohols.

The other portion (Figure 7, fraction 2B) of the water extract is eluted slowly through a conditioned SCX (500 mg) cation-exchange cartridge. The pH of the eluate is determined to test the completeness of the cation exchange and, if necessary, the eluate is neutralized. The neutral eluate is evaporated and silylated as described above for the first aqueous portion. This sample is analyzed for polar alkylphosphonic and alkylthiophosphonic acids. Aminoalcohols remain in the SCX cartridge and cannot be recovered.

An optional SPE procedure (Figure 7, fraction 2C) for the water extract can be applied. The pH of a 5 mL portion of the water extract is adjusted with hydrochloric acid and ammonium hydroxide to pH 4.5. The sample is eluted slowly through a conditioned SCX (500 mg) cation-exchange cartridge and after that through a conditioned NH_2 (100 mg) anion-exchange cartridge. The retained analytes are eluted slowly from the NH_2 cartridge with 2 mL of 2% hydrochloric acid in methanol. The eluate is adjusted to pH 8 with concentrated ammonia and then evaporated to dryness under mild nitrogen flow. The evaporation residue is silylated as described above for the first aqueous portion, to provide the sample in a form suitable for GC/MS analysis. Aminoalcohols remain in the SCX cartridge and cannot be recovered from this sample. This method was developed for analysis of samples for polar alkylphosphonic and alkylthiophosphonic acids. The effectiveness of the method varies very much with the sample matrix, however. Depending, among other things, on the capacity and quality of the cation- and anion-exchange cartridges and the amount and character of impurities in the sample, the recoveries of alkylphosphonic and alkylthiophosphonic acids may be very low. Accordingly, this method should be considered as an optional technique only for sample matrices of known ion content.

In the next step (Figure 7, fraction 3) a new 5 g portion of the soil is extracted with 5 mL of 1% TEA in methanol as described above for dichloromethane. Under mild nitrogen flow at 30 °C 4 mL of the centrifuged, filtered, and combined TEA/methanol extract is evaporated to dryness. The evaporation residue is silylated in the manner described above for the evaporation residue of the water extract. This sample is analyzed for alkaline CWC-related chemicals such as VX.

Another new sample portion of 5 g is prepared for analysis of the sample for Lewisite 1 and Lewisite oxide

(Figure 7, fraction 4). The soil is mixed with 15 mL of hydrochloric acid (pH 2). The pH of the water fraction is determined and, if necessary, adjusted to pH 2. The pH of the water fraction must be maintained between 1.5 and 2.5 over a period of 30 min. The sample is then shaken vigorously for 30 s every 5 min for 60 min, allowed to settle for 1 h, centrifuged, and filtered. To the filtered extract is added 0.1 mL of the freshly prepared solution containing DMT 5 mg mL^{-1} in acetone. The sample vial is shaken vigorously and allowed to stand for 10 min, 1 mL of *n*-hexane is added, and the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

3.3 Wipe Samples

The wipe sample is inserted into a glass vial and approximately 10 mL of dichloromethane is added. Care must be taken that the wipe is completely covered by the solvent. The sample is shaken for 1 min. The organic phase is quickly decanted and filtered, and the extraction procedure is repeated with an additional 10 mL of dichloromethane. The extracts are combined and dried. Under mild nitrogen flow, 2 mL of the dried extract is concentrated approximately 10-fold and analyzed (Figure 8, fraction 1). These dichloromethane extracts are analyzed for nonpolar CWC-related chemicals.

Next the wipe sample from the dichloromethane extraction is placed in a fume cupboard and allowed to dry for approximately 1 h. About 12 mL of distilled, deionized water is added, with care taken that the wipe is completely covered by the solvent. The sample is shaken for 1 min. The wipe is then extracted, in the manner described above for dichloromethane. The centrifuged, filtered, and combined water extract is divided into two portions of 3 mL each (Figure 8, fraction 2).

The first portion (Figure 8, fraction 2A) is evaporated to dryness in a centrifugal evaporator. Added to the evaporation tube are 0.5 mL of acetonitrile and 0.5 mL of BSTFA and the silylation mixture is heated at 60 °C for 30 min to complete the silylation. This sample is analyzed for polar CWC-related chemicals such as alkylphosphonic and alkylthiophosphonic acids, thiodiglycol, and aminoalcohols.

An optional SPE procedure for the water extract can be applied (Figure 8, fraction 2B). The pH of a 3 mL portion of the water extract is adjusted with hydrochloric acid and ammonium hydroxide to pH 4.5. The sample is eluted slowly through a conditioned SCX (500 mg) cation-exchange cartridge and after that through a conditioned NH_2 (100 mg) anion-exchange cartridge. The retained

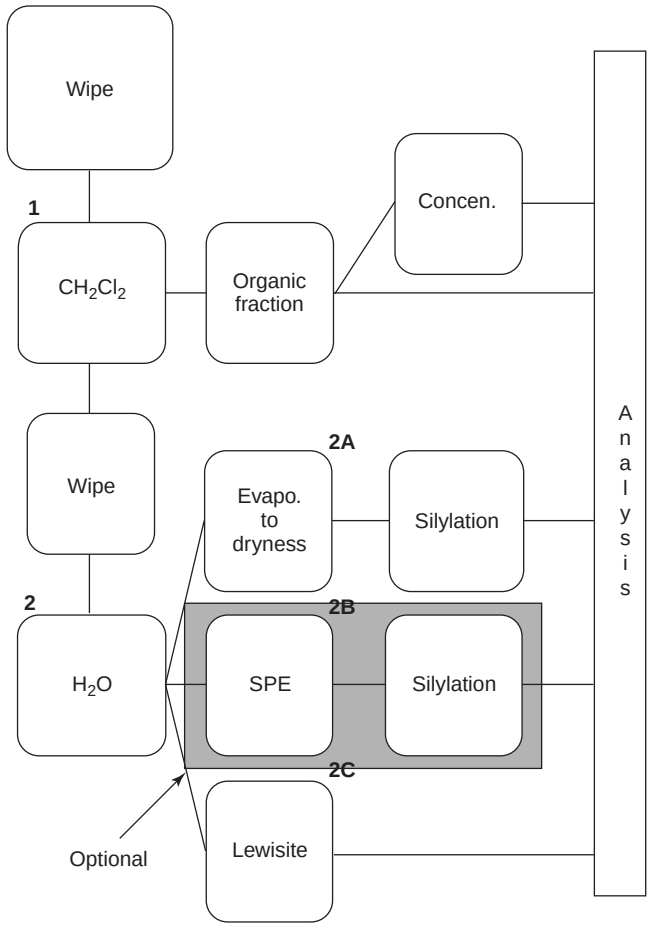

Figure 8 Schematic presentation of the SOP for wipe sample preparation.

analytes are eluted slowly from the NH_2 cartridge with 2 mL of 2% hydrochloric acid in methanol. The eluate is adjusted to pH 8 with concentrated ammonia and then evaporated to dryness under mild nitrogen flow. The evaporation residue is silylated as described above for the first aqueous portion to provide the sample in a form suitable for GC/MS analysis. Aminoalcohols remain in the SCX cartridge and cannot be recovered from this sample. This method was developed for analysis of samples for polar alkylphosphonic and alkylthiophosphonic acids. The effectiveness of the method varies very much with the sample matrix, however. Depending, among other things, on the capacity and quality of the cation- and anion-exchange cartridges and the amount and character of impurities in the sample, the recovery values for alkylphosphonic and alkylthiophosphonic acids may be very low. Accordingly, this method should be considered as an optional technique only for sample matrices of known ion content.

The second portion of 3 mL is prepared for analysis of the samples for Lewisite 1 and Lewisite oxide (Figure 8,

fraction 2C). The pH of the sample is adjusted with hydrochloric acid to pH 2, the sample is filtered, and 0.1 mL of freshly prepared solution containing DMT 5 mg mL^{-1} in acetone is added. The sample vial is shaken vigorously and allowed to stand for 10 min. Then 1 mL of *n*-hexane is added and the sample is shaken vigorously for 30 s every 10 min for 30 min. The hexane fraction is allowed to separate. The top hexane–water layer is transferred to a new vial and centrifuged. After centrifugation the hexane fraction is separated, dried, and submitted for analysis.

4 QUALITY CONTROL

Ensuring quality during sample preparation requires that the samples do not come in contact with one another, are not mixed with one another, nor are they contaminated by chemicals from the laboratory or other samples, and that the procedures are shown to be controlled.

Meeting the quality demands of sample preparation requires that the ROPs or SOPs are followed in all sample preparations. The purity of solvents and reagents is checked before sample preparation by the same methods used for analysis. The purity of 10-fold concentrated solvents is also tested beforehand. The background generated during the sample preparation procedure is demonstrated by preparing a blind sample, and by passing the same solvents, same reagents, and similar glassware used in sample preparation through the whole procedure. Similarly, where possible the background generated by the sample matrix is demonstrated by passing a blank sample through the entire procedure.

It cannot be emphasized too strongly that mistakes in sample preparation may cause failure of the whole analytical procedure, even with the most sophisticated instrumentation. This means that not only must the QC rules be followed but work must be done unhurriedly and with total concentration.

5 SAFETY

CWC-related chemicals must be handled with great care. Persons handling toxic chemicals must be specially trained for the work. When toxic samples are handled, decontamination solution, protective masks, and autoinjectors of nerve agent antidotes must always be available, and no-one must ever work alone. Individual protective gear such as a laboratory coat, chemically resistant protective gloves, and safety goggles are essential during sample preparation. Toxic samples must always be prepared in a fume cupboard.

The fume cupboard is cleaned directly after sample preparation. Any samples, organic solvent waste, chlorinated solvent waste, and aqueous wastes which do not require decontamination are collected into separate clearly marked waste containers. In the same way paper and consumable wastes which do not require decontamination are collected in a clearly marked waste box. Materials requiring decontamination must be treated with a proper decontamination solution and disposed of in designated waste containers. Glassware and accessories are flushed with decontamination solution and soaked in potassium hydroxide solution and, if not destroyed, they are washed with alkaline nonphosphorus detergent before further cleaning.

6 INTERNATIONAL COMPARISON AND PROFICIENCY TESTS

As described in the Introduction of this article, sample preparation procedures for off-site laboratories for the CWC have been developed and tested in five international interlaboratory comparison (round-robin) tests,[1–5] in two trial proficiency tests, and in four official proficiency tests (see **Verification of Chemicals Related to the Chemical Weapons Convention**). Table 3 lists the types of samples in these tests. The first three tests were arranged mainly for purposes of method development, the fourth and fifth also for testing and validating of methods. The reports describing these tests (the round-robin books) contain a thorough description of how each of the participating laboratories prepared their samples.[1–5]

Only in one test (the fourth round-robin) were the identified chemicals quantified,[4] and then only to facilitate the evaluation of sample preparation methods. The quantitative results varied widely. Weaknesses were revealed in the procedures because all the chemicals were polar and adsorptive. Comparison of the quantitative results as a means of deciding upon the best sample preparation procedures proved to be an unreasonable approach, however, as the total variance in the results was clearly the product of several factors: the sample preparation procedure, the quantitative method, and the instrumental technique. Nevertheless, the exercise proved to have considerable educational value. The test revealed weaknesses in the existing ROPs[6] and the methods were subsequently improved to cover the gaps.[7]

In both the trial and official proficiency tests, the ROPs for sample preparation validated in the round-robin tests proved to be useful methods for recovering CWC-related chemicals spiked at trace level. The participating laboratories had prepared the samples following the ROPs, although sometimes with slight modifications. Many of

Table 3 Sample types in the international interlaboratory comparison (round-robin) tests for verification of chemical disarmament held for method development and testing, and in trial and official proficiency tests held for selection of designated laboratories

Test	Samples	Comment
Round-Robin 1 (1989)[1]	Tenax air samples	ROPs were established
	XAD-2 air samples	
	Soil samples	
Round-Robin 2 (1990)[2]	Tenax air samples	Existing ROPs were tested for air samples
	XAD-2 air samples	
	Wipe samples	ROPs were established
	Charcoal samples	
	Aqueous liquid samples	
Round-Robin 3 (1991)[3]	Concrete samples	ROPs were established
	Paint samples	
	Rubber samples	
Round-Robin 4 (1993)[4]	Soil samples	Existing ROPs were validated
	Water samples	Quantitative analysis
	Water samples trapped in a C_{18} cartridge	
Round-Robin 5 (1994)[5]	Soil samples	Existing ROPs were tested
	Water samples	
	Decontamination solutions	
Trial Proficiency Test (1995)[a]	Rubber samples	Existing ROPs were tested
	Paint samples	
	Soil samples	
The Second Trial Proficiency Test (1995)[a]	Organic liquid samples	Existing ROPs were tested
	Water samples	
	Soil samples	
The First Official Proficiency Test (1996)[a]	Organic liquid samples	Existing ROPs were tested
	Water samples	
	Polymer samples	
The Second Official Proficiency Test (1996)[a]	Organic liquid samples	Existing ROPs were tested
	Water samples	
	Soil samples	
The Third Official Proficiency Test (1997)[a]	Organic liquid samples	Existing ROPs were tested
	Water samples	
	Paint samples	
The Fourth Official Proficiency Test (1998)[a]	Water samples	Existing ROPs were tested
	Soil samples	
	Wipe samples	

The year refers to the dates the tests were carried out.
[a] Reports are available from Organization for the Prohibition of Chemical Weapons (OPCW), Technical Secretariat, The Hague, The Netherlands.

the laboratories were also able to identify degradation products or impurities of the spiking chemicals present in only very low concentration. The results show that the ROPs work well in the recovery of chemicals.Where a laboratory had problems the most common difficulty was the unsuccessful identification of the spiking chemicals or lack of reference data, not improper sample preparation.

An essential requirement for achieving good analytical results is the experience and knowledge of CWC-related chemicals of those workers doing the sample preparation and instrumental analysis. Some examples of incorrect sample preparation revealed in the evaluation of results of the proficiency tests are noted in the following.

In the first trial proficiency test one laboratory did not find either of the spiking chemicals, sesquimustard (CAS 3563-36-8) and CR (CAS 257-07-8), in a paint sample. The laboratory scratched the paint away from the paint plate surface before extraction, which may have caused the chemicals of interest to escape. Furthermore, the matrix probably dissolved during the extraction, making the analysis of the extract, with so much interference from the sample matrix, more difficult.

In the second trial proficiency test one laboratory failed to find methylphosphonic acid and cyclopentyl methylphosphonate (CAS 73207-99-5) in a water sample. The laboratory evidently did not evaporate a portion of the water sample to dryness and derivatize the

evaporation residue, but only derivatized a portion of an alkaline dichloromethane extract. Since the spiking chemicals are polar and acidic they are not extractable into a nonpolar solvent under alkaline conditions.

The soil sample was even more problematic in this test. Only six of the 15 participating laboratories successfully identified both *N*,*N*-diethylaminoethanol (CAS 100-37-8) and *N*-ethyldiethanolamine (CAS 139-87-7). For these particular spiking chemicals, efficient TEA/methanol extraction followed by a silylation procedure was essential in the sample preparation. One laboratory missed both spiking chemicals, probably because the TEA/methanol extraction was not carried out. Another laboratory also missed them but for another reason. Instead of TEA, ethyldimethylamine (EDMA, CAS 598-56-1) was used as extractant, and although this should not have had a dramatic effect on the recoveries of the spiking chemicals since EDMA should be as good a modifier as TEA, the laboratory then derivatized the 1% EDMA/methanol extract with BSTFA, thus removing any chance of success. Chemicals in alcohol solutions cannot be silylated, and this must have been the reason why the spiking chemicals were not found.

In the first official proficiency test, two laboratories did not evaporate and derivatize the evaporation residue of a water sample and failed to identify *n*-propylphosphonic acid (PPA, CAS 4672-38-2). One laboratory employed a SPE method but used a strong cation-exchange cartridge SAX. Weak anion-exchange cartridges are recommended for the SPE method to ensure recovery of the acidic degradation products from water samples. Acids such as PPA are strongly retained on strong anion-exchange cartridges and hence difficult to elute from the cartridge.

3,3-Dimethyl-2-butanol (DMB, CAS 464-07-03) was another spiking chemical in the water sample in this test. One laboratory missed it because of insufficient sample preparation: the water sample was not extracted with dichloromethane in which the chemical is soluble, but was merely evaporated and derivatized. DMB was evaporated together with the water and in this way was lost.

In the third official proficiency test at least four laboratories failed to find (1-methylethyl)phosphonic acid (IPPA, CAS 4721-37-3) most probably because the water extract was not prepared properly. These laboratories proceeded to neutral and alkaline dichloromethane extraction and some even evaporated the water to dryness. However, recovery of IPPA requires direct analysis of the water sample by LC/MS (or CZE) or else cation exchange, evaporation of the sample to dryness, and derivatization (methylation) of the residue.

In the fourth official proficiency test two laboratories failed to find propyl propylphosphonate (CAS 21921-97-1) and *O*-propyl propylthiophosphonate (CAS unknown) in the water sample. The laboratories followed the ROPs but with slight modifications. One laboratory cation exchanged a portion of the water sample, and after evaporating it to dryness, silylated the residue with 1% *tert*-butyldimethylchlorosilane (CAS 18162-48-6)/MTBSTFA (10%) in toluene. The other laboratory extracted a portion of the water sample with dichloromethane, evaporated the water phase to dryness, and dissolved the evaporation residue in diazomethane/ether solution. Another portion of the water sample was extracted with ethyl acetate and evaporated to dryness and the evaporation residue was dissolved in acetone. The acetone solution was then itself evaporated to dryness and silylated with BSTFA/pyridine in acetonitrile. Yet a third portion of the water sample was adjusted to the same pH as the blank water sample and extracted with dichloromethane. After that the pH was reduced to ca. 2.5 and the dichloromethane extraction was repeated. This acidic dichloromethane was dried, filtered, and evaporated to dryness before silylation with BSTFA/pyridine in acetonitrile.

It should have been possible to identify the two spiking chemicals in the cation-exchanged and silylated water fraction. However, following cation exchange, higher recoveries are usually obtained for alkyl phosphonates and alkylthiophosphonates if methylation is used for derivatization. Moreover, the solubility of these chemicals in the silylation mixture is critical for their recovery and therefore either acetonitrile or tetrahydrofuran is typically used as a silylation solvent. The solubility of the spiking chemicals in toluene, which the first laboratory chose as silylation solvent, perhaps was not sufficient. The second laboratory dissolved the first evaporation residue in diazomethane/ether solution, but perhaps the solubility of the spiking chemicals in ether was not sufficient either, especially since the water was not cation exchanged before evaporation. Similarly, before silylation the second sample fraction was not cation exchanged, and the evaporation residue was dissolved in acetone in which the spiking chemicals are only sparingly soluble. Small amounts of these types of spiking chemicals should be recovered in acidic dichloromethane extracts that are derivatized as such. However, the laboratory first evaporated the dichloromethane extracts to dryness and then silylated the evaporation residue. This might have caused loss of the spiking chemicals.

ACKNOWLEDGMENTS

My thanks to Olli Kostiainen, Otto Ahonen, Markku Mesilaakso, Marja-Leena Rapinoja, and Professor Marjatta Rautio of the Institute for their helpful comments,

and to Harri Kiljunen for his assistance with computer programs during preparation of the manuscript. Seija Lemettinen and Kirsi Rosendahl provided excellent technical assistance during the round-robin and proficiency tests, and Marja-Leena Rapinoja during the validation of on-site sample preparation methods, and to them my thanks as well.

I am grateful to Dr. Kathleen Ahonen for improving the language of the manuscript.

ABBREVIATIONS AND ACRONYMNS

BSTFA	*N*,*O*-bis(Trimethylsilyl)-trifluoroacetamide
CWC	Chemical Weapons Convention
CZE	Capillary Zone Electrophoresis
DMB	3,3-Dimethyl-2-butanol
DMT	3,4-Dimercaptotoluene
EDMA	Ethyldimethylamine
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
IPPA	(1-Methylethyl)phosphonic Acid
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
MTBSTFA	Methyl-*N*-*tert*(butyldimethylsilyl)-trifluoroacetamide
NMR	Nuclear Magnetic Resonance
OPCW	Organization for the Prohibition of Chemical Weapons
PPA	*n*-Propylphosphonic Acid
QC	Quality Control
ROP	Recommended Operating Procedure
SOP	Standard Operating Procedure
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TEA	Triethylamine

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)** Verification of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)** Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

REFERENCES

1. M. Rautio (ed.), 'International Interlaboratory Comparison (Round-Robin) Test for the Verification of Chemical Disarmament, F.1. Testing of Existing Procedures', Ministry for Foreign Affairs of Finland, Helsinki, 1990.
2. M. Rautio (ed.), 'International Interlaboratory Comparison (Round-Robin) Test for the Verification of Chemical Disarmament, F.2. Testing of Procedures on Simulated Industry Samples', Ministry for Foreign Affairs of Finland, Helsinki, 1991.
3. M. Rautio (ed.), 'International Interlaboratory Comparison (Round-Robin) Test for the Verification of Chemical Disarmament, F.3. Testing of Procedures on Simulated Military Facility Samples', Ministry for Foreign Affairs of Finland, Helsinki, 1992.
4. M. Rautio (ed.), 'International Interlaboratory Comparison (Round-Robin) Test for the Verification of Chemical Disarmament, F.4. Validating of Procedures for Water and Soil Samples', Ministry for Foreign Affairs of Finland, Helsinki, 1993.
5. M. Rautio (ed.), 'Interlaboratory Comparison Test Coordinated by the Provisional Technical Secretariat for the Preparatory Commission for the Organization for the Prohibition of Chemical Weapons, H.1. First Interlaboratory Comparison Test', Ministry for Foreign Affairs of Finland, Helsinki, 1994.
6. M. Rautio (ed.), 'Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament', 1993 edition, Ministry for Foreign Affairs of Finland, Helsinki, 1993.
7. M. Rautio (ed.), 'Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament', 1994 edition, Ministry for Foreign Affairs of Finland, Helsinki, 1994.
8. M. Rautio (ed.), 'Air Monitoring as a Means for the Verification of Chemical Disarmament, Part I. Development and Evaluation of Basic Techniques', Ministry for Foreign Affairs of Finland, Helsinki, 1985.
9. M. Rautio (ed.), 'Air Monitoring as a Means for the Verification of Chemical Disarmament, Part II. Field Tests', Ministry for Foreign Affairs of Finland, Helsinki, 1986.
10. M. Rautio (ed.), 'Air Monitoring as a Means for the Verification of Chemical Disarmament, Part III. Further Development and Testing of Methods', Ministry for Foreign Affairs of Finland, Helsinki, 1987.
11. Z. Witkiewicz, M. Mazurek, J. Szulc, 'Chromatographic Analysis of Chemical Warfare Agents', *J. Chromatogr.*, **503**, 293–357 (1990).
12. V.M.A. Häkkinen, 'Analysis of Chemical Warfare Agents in Water by Solid-phase Extraction and Two-Channel

Capillary Gas Chromatography', *J. High Resolut Chromatogr.*, **14**, 811–815 (1991).
13. J. Aa. Tørnes, B.A. Johnsen, 'Gas Chromatographic Determination of Methylphosphonic Acids by Methylation with Trimethylammonium Hydroxide', *J. Chromatogr.*, **467**, 129–138 (1989).
14. J. Aa. Tørnes, A.M. Opstad, B.A. Johnsen, 'Use of Solid-phase Extraction in Determination of Chemical Warfare Agents. Part I Evaluation of the Solid-phase Extraction Technique', *Int. J. Environ. Anal. Chem.*, **44**, 209–225 (1991).
15. J. Aa. Tørnes, A.M. Opstad, B.A. Johnsen, 'Use of Solid-phase Extraction in Determination of Chemical Warfare Agents. Part II Determination of Chemical Warfare Agents in Samples from a Battlefield Environment', *Int. J. Environ. Anal. Chem.*, **44**, 227–232 (1991).
16. G.A. Sega, B.A. Tomkins, W.H. Griest, 'Analysis of Methylphosphonic Acid, Ethyl Methylphosphonic Acid and Isopropyl Methylphosphonic Acid at Low Microgram per Liter Levels in Groundwater', *J. Chromatogr., A*, **790**, 143–152 (1997).
17. H.-Å. Lakso, W.F. Ng, 'Determination of Chemical Warfare Agents in Natural Water Samples by Solid-phase Microextraction', *Anal. Chem.*, **69**(10), 1866–1872 (1997).
18. A. Alcaraz, S.S. Hulsey, R.E. Whipple, B.D. Andresen, 'On-Site Sample Work-up Procedures to Isolate Chemical Warfare Related Compounds Using Solid-phase Extraction and Solid-phase Microextraction Technology', NATO ASI Ser. Ser 1 **13**, Analytical Chemistry Associated with the Destruction of Chemical Weapons, 65–76, 1997.
19. R. Kostiainen, A.P. Bruins, V.M.A. Häkkinen, 'Identification of Degradation Products of Some Chemical Warfare Agents by Capillary Electrophoresis–Ionspray Mass Spectrometry', *J. Chromatogr.*, **634**, 113–118 (1993).
20. C.E. Kientz, E.W.J. Hooijschuur, U.A. Th. Brinkman, 'Capillary Electrophoresis Coupled On-line with Flame Photometric Detection: Determination of Alkylphosphonic Acids', *J. Microcolumn Sep.*, **9**(4), 253–259 (1997).
21. C.E. Kientz, A. Verweij, G.J. de Jong, U.A. Th. Brinkman, 'Verification of Nonproduction of Chemical Warfare Agents: I. Determination of Organophosphorus Compounds by Microcolumn Liquid Chromatography with Flame Photometric or Thermionic Detection', *J. Microcolumn Sep.*, **4**(6), 465–475 (1992).
22. C.E. Kientz, A. Verweij, G.J. de Jong, U.A. Th. Brinkman, 'Verification of Nonproduction of Chemical Warfare Agents: II. Large Volume Injections in Microcolumn Liquid Chromatography using Flame Photometric Detection', *J. Microcolumn Sep.*, **4**(6), 477–483 (1992).
23. E.R.J. Wils, A.G. Hulst, 'Determination of Organophosphorus Acids by Thermospray Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **454**, 261–272 (1988).
24. R.M. Black, R.W. Read, 'Application of Liquid Chromatography–Atmospheric Pressure Chemical Ionisation Mass Spectrometry and Tandem Mass Spectrometry to the Analysis and Identification of Degradation Products of Chemical Warfare Agents', *J. Chromatogr., A*, **759**, 79–92 (1997).
25. R.M. Black, R.W. Read, 'Analysis of Degradation Products of Organophosphorus Chemical Warfare Agents and Related Compounds by Liquid Chromatography–Mass Spectrometry Using Electrospray and Atmospheric Pressure Chemical Ionization', *J. Chromatogr., A*, **794**, 233–244 (1998).
26. P.A. D'Agostino, L.R. Provost, 'Determination of Chemical Warfare Agents, Their Hydrolysis Products and Related Compounds in Soil', *J. Chromatogr.*, **589**, 287–294 (1992).
27. P.A. D'Agostino, L.R. Provost, 'Gas Chromatographic Retention Indices of Chemical Warfare Agents and Simulants', *J. Chromatogr.*, **331**, 47–54 (1985).
28. M.-L. Kuitunen, K. Hartonen, M.-L. Riekkola, 'Analysis of Chemical Warfare Agents in Soil Samples by Off-line Supercritical Fluid Extraction and Capillary Gas Chromatography', *J. Microcolumn Sep.*, **3**(6), 505–512 (1991).
29. W.R. Creasy, A.A. Rodrigues, J.R. Stuff, R.W. Warren, 'Atomic Emission Detection for the Quantitation of Trimethylsilyl Derivatives of Chemical Warfare Agent Related Compounds in Environmental Samples', *J. Chromatogr., A*, **709**, 333–344 (1995).
30. V. Camel, M. Caude, A. Tambuté, 'SFE of an Organophosphorus Compound from Soils with Capillary GC Analysis', *J. Chromatogr. Sci.*, **33**, 123–132 (1995).
31. Standard Operating Procedures, Gas Chromatography/Mass Spectrometry (GC/MS) Sample Preparation, OPCW Inspection Manual. Available from OPCW, Technical Secretariat, The Hague, The Netherlands.

Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

David B. Cooper
DERA, Salisbury, UK

1 Introduction **1**
1.1 The Chemical Weapons Convention 1
1.2 Sampling, Detection and Screening 2
2 Sampling **2**
2.1 Elements of a Sample Plan 2
2.2 Factors Influencing a Sampling Plan 2
2.3 Sample Collection 3
2.4 Sample Management 4
2.5 Sample Splitting 4
2.6 Sample Transportation 6
2.7 Sample Collection Techniques 7
2.8 Sample Collection Team 7
2.9 Sample Types 8
3 Screening **10**
4 Detection **10**
4.1 Use of Detectors 10
4.2 Ion Mobility Spectrometry 10
4.3 Flame Photometric Detection 11
4.4 Detector Papers 11
4.5 Detector Tubes 11
4.6 On-site Analysis Equipment 11
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**

The Chemical Weapons Convention (CWC), which entered into force on April 29, 1997, prohibits the development, production, stockpiling and use of chemical weapons (CW). It is administered by the Organization for the Prohibition of Chemical Weapons (OPCW), which is based in The Hague.

During the course of inspections, whether routine or challenge, there may be a requirement for the inspection team (IT) to investigate unresolved ambiguities. The IT has the right to request clarification, and if ambiguities cannot be clarified by negotiation then samples may be taken for analysis.

Examples of sampling equipment, procedures and sample management techniques are described, together with a brief description of screening and detection. Some analyses may be performed on-site by the inspectors, using equipment they bring with them, but for more difficult questions it may be necessary to remove a sample for analysis on an anonymous basis in two or more of a global network of laboratories ''designated'' by the OPCW. The end result of any sampling and analysis operation must be able to stand up to international scrutiny, and therefore chain-of-custody or audit-trail procedures must be strictly followed. The process should allow the re-creation of the position or state of the sample at any time during the handling of the sample from cradle to grave.

1 INTRODUCTION

1.1 The Chemical Weapons Convention

The CWC, which entered into force on April 29, 1997, prohibits the development, production, stockpiling and use of CW.[1] It is administered by the OPCW, which is based in The Hague. The CWC centers on a general purpose criterion which stipulates that toxic chemicals and their precursors may only be developed, produced, acquired, retained, transferred or used for permitted purposes and in types and quantities consistent with such purposes. In addition to provisions for one State Party to call for a challenge inspection against another suspected of contravening the convention, there are declaration, routine inspection and transfer obligations for specified chemicals that pose particular risks. These chemicals are grouped in three schedules. Schedule 1 contains toxic chemicals which are a high risk to the convention; they are mainly nerve agents and mustards. Schedule 2 contains industrial chemicals which are a significant risk; these include chemicals that could be used in CW, such as the pesticide amiton, and also precursors to the chemicals in Schedule 1. Schedule 3 contains more widely used industrial chemicals; these include industrial chemicals such as phosgene, and precursors to some Schedule 1 or 2 chemicals. In addition to the routine inspections of facilities producing Schedule 1, 2 or 3 chemicals, there are provisions for inspecting old or abandoned CW sites, and production, storage and destruction facilities. This allows the OPCW to monitor the destruction of facilities and weapons by the small number of States Parties who are declared CWC-related chemical possessors. During the course of inspections, whether routine or challenge, there may be a requirement for the IT to investigate unresolved ambiguities. The IT has the right to request clarification and, if ambiguities cannot be clarified by negotiation, to take samples for analysis. Procedures followed by the IT

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

for sampling and analysis are laid down in the OPCW inspection manual. These procedures were developed by technical experts of the States Parties, working in expert and task force groups before entry into force (EIF) of the CWC. Some analyses may be performed on-site by the inspectors using equipment they take with them, but for more difficult questions it may be necessary to remove a sample for analysis on an anonymous basis in two or more of a global network of laboratories designated by the OPCW. The end result of any sampling and analysis operation must be able to stand up to international scrutiny, and therefore OPCW-designated laboratories must have achieved accreditation by a recognized body and must also pass proficiency tests organized regularly by the OPCW.

1.2 Sampling, Detection and Screening

There is an inter-relationship and synergism associated with the three elements of sampling, detection and screening in the investigation of sites which are either contaminated with CWC-related chemicals or thought to contain munitions or other materials of concern to the CWC. These three elements in combination are essential for the safe and accurate assessment of CWC-related chemical sites. Screening a site with the appropriate detection equipment is useful both for the safe conduct of a sampling mission and for verification purposes. In the safety role the use of detectors will greatly assist the safe progress of the sampling mission, by determining the level of protection to be used during both the reconnaissance and sampling processes. In the verification role detectors can be used to screen the areas of concern and pinpoint chemical contamination for subsequent sampling. Generally methods of detection should be sensitive and discriminating, and should support the sampling and analysis process but not supersede it.

2 SAMPLING

2.1 Elements of a Sample Plan

Few reference books offer guidelines for the collection of samples that have relevance to CW. Therefore, to aid the planning of missions to CW sites, one must look to the area of environmental sampling for assistance. The Royal Society of Chemistry and the American Chemical Society have produced guidelines for the sampling process which are equally relevant in the CW field.[2,3] The most important phase of sampling and analysis in the verification of the CWC is the sample collection process, because even with the most sophisticated analytical equipment an inappropriate sample will always produce meaningless analytical results. When carrying out any site assessment that will entail sampling and analysis the first requirement is to produce a sampling plan. A typical plan will include the following elements in a sequential consideration of the factors involved:

- definition of the objective/aims of the measurement;
- selection of constituents (analytes) and analytical methods;
- determination of sampling locations;
- determination of sample types and methods of sampling;
- selection of methods for sample preservation, transport and protection from the possibility of tampering;
- preparation of the final plan (protocol);
- revision in the light of experience.

The above summary of factors to be considered when designing a sampling plan is reproduced by permission of the Royal Society of Chemistry.

2.2 Factors Influencing a Sampling Plan

When assembling a sampling plan the following factors need to be considered.

- Preinspection planning should be carried out before embarking on any sampling mission. The following information can be useful in formulating an initial plan:
  - preinspection knowledge, e.g. geographical, meteorological and intelligence;
  - construction plans of the facility, if available;
  - local knowledge from present or former workers at the site;
  - intelligence reports from the area (this is especially important in an ''alleged use'' scenario);
  - historical documentation from government agencies;
  - aerial/satellite photographs.
- The availability and suitability of detection equipment for on-site safety and screening purposes should be considered.
- Safety/risk management, and formulation of a safety plan that takes into account chemical and physical hazards, are essential for the safe completion of the mission. The IT should be fully aware of the toxicological properties of all expected hazards, and be trained to operate in a contaminated environment.
- When defining a chemical hazard under the CWC a ''chemical weapon'' is stated in Article II of the convention as:

  - ''Toxic chemicals and their precursors, except where intended for purposes not prohibited under this convention, as long as the types and quantities are consistent with such purposes.
  - Munitions and devices, specifically designed to cause death or other harm through the toxic properties of those toxic chemicals specified in the above subparagraph, which would be released as a result of the employment of such munitions and devices.
  - Any equipment specifically designed for use directly in connection with the employment of munitions and devices specified in the above subparagraph.
  - 'Toxic chemical' means: any chemical which through its chemical action on life processes can cause death, temporary incapacitation or permanent harm to humans or animals. This includes all such chemicals, regardless of their origin or of their method of production, and regardless of whether they are produced in facilities, in munitions or elsewhere.
  - 'Precursor' means: any chemical reactant which takes part at any stage in the production, by whatever method, of a toxic chemical. This includes any key component of a binary or multi-component chemical system.''

- Physical hazards and specific site concerns which may cause problems to the samplers include the following:
  - the possibility of site contamination and the associated decontamination problems this presents;
  - the possibility of explosive or fire hazards from the use of electrical equipment in hazardous environments;
  - the problem of site safety in terms of unsafe structures, for example old or semidemolished buildings, underground facilities, complicated chemical plant, storage bunkers, flooding;
  - availability of resources, e.g. power, water, shelter, communications;
  - environment/terrain, accessibility;
  - military activity of a hostile nature, especially when dealing with alleged use incidents.
- Arrangements for medical backup should be as follows:
  - the IT must be sure adequate medical safety measures are in place;
  - medical facilities of an adequate nature should either be available on-site or a medical team should be present as a subunit of the IT;
  - real-time monitoring for toxic chemicals should be available if necessary (this would perhaps be dealt with in the site risk assessment and any precautionary monitoring initiated as soon as the IT reaches the site).
- Selection of sampling equipment will be determined by types and numbers of samples and will vary according to the inspection. Once on-site the team may be isolated and therefore initial planning and selection of sufficient and suitable sampling equipment will be essential to the successful completion of the mission.
- Availability of personal protective equipment (PPE); the type of equipment needed will be determined by the expected hazard. A preinspection assessment of the hazards to be encountered will be needed and will rely upon knowledge of the site gained prior to departure. Without this information it will be necessary to take the fullest protection possible, although this may severely affect the logistical requirements of the mission and hamper the sampling team in the execution of its remit. The IT should be fully trained and familiar with the operation and use of all equipment provided for the safe execution of their duties.
- Arrangements for decontamination and waste management may be available locally. If not, the team will need to cater for both expected and unexpected hazards, and suitable equipment will need to accompany the team. All sampling activities produce waste and therefore the sampling plan should cater for the decontamination and disposal of all waste created during the mission.
- On-site planning will involve adaptation of the sampling plan to suit the on-site conditions. For example, a statistical approach to sampling may be necessary to identify suitable sampling points where no obvious areas are immediately apparent.

2.3 Sample Collection

When considering the special requirements involved in the collection of samples from a site believed to be contaminated with CWC-related chemicals, there are very few reference textbooks to consult for advice. The equipment used and procedures followed by an OPCW IT for sampling and analysis have been laid down in the draft OPCW inspection manual. These recommended operating procedures (ROPs) and sampling kits were initially drawn up by technical experts of the States

Parties, working in expert and task force groups before EIF, and since then modifications and additions have been made in the light of experience. Similarly, there are procedures and equipment lists in the NATO (North Atlantic Treaty Organization) Handbook, AEP-10, which contains procedures to follow in the event of CW usage in the battlefield, and in the Finnish ''Blue Book'', which contains ROPs for sampling and analysis in the verification of chemical disarmament.[4,5]

Locations from which neat chemical samples may be collected include munition bunkers, intact munitions, chemical plant, reactors and process/filling lines. Other areas from which environmental samples will be taken include waste outflows, waste treatment or storage areas, burning/demolition grounds, trials areas, vehicle tyres, flooring in storage areas, reactor flanges, ventilation equipment, filters, door seals, and entrances and exits of buildings.

For chemical agents or degradation products the typical sample collection kit will contain equipment suitable for the following types of sample: neat chemical agent (solid or liquid), wastewater, air, concrete, rubber, paint, swabs/wipes, soil, and vegetation and debris. A list of equipment in a typical sampling kit is shown in Table 1.

The use of specialist equipment, for example power augers, sampling drills, crushers and mills, is not envisaged for the types of investigation and inspection carried out by the OPCW. However, it has been decided that for the sampling of intact munitions a safe system for automatically drilling and sampling the contents is required and a specification for such a piece of equipment is to be produced, but as yet the OPCW has not acquired such a system.

2.4 Sample Management

The CWC specifies that during the course of inspection activities there may be a requirement to resolve ambiguities by sampling and analysis. The result of this process must stand up to international scrutiny and be legally and scientifically defensible. The sample must therefore be collected and handled in such a way that the data obtained can be used in evidence when proving compliance or noncompliance with the CWC. The elements of a sample management system include unique sample identification, chain-of-custody or auditable procedures throughout the life of the sample, documentation or logging procedures, sample-splitting techniques, transportation, preservation, antitampering and storage. These are dealt with next in a logical sequence.

When a sample is collected it is given a unique identification number. This may be used to indicate not only a rolling number, but also a site code, the sampler's identity, and the sample matrix. A suffix may be added later to indicate other events.

When a sample is collected a sample log-sheet should be completed detailing all necessary information relating to that specific sample, such as time, date, position (site related or using global positioning system (GPS) data), sample identification number, why the sample was taken, detector responses, in fact everything pertinent to that sample. It may also be prudent, if safe to do so, to photograph the sample location and the sampling process. This document should then accompany the sample to the reception area, as all information about the sample is important in the analytical processes which follow.

The chain-of-custody procedure begins immediately a sample is taken and is maintained throughout the life of the sample. The process should allow the re-creation of the position or state of the sample at any time during the handling of the sample from cradle to grave. When a sample is taken, unique tamper-proof seals should be applied to the sample container. The number of the seal should be entered on the sample log-sheet and a ''traveler'' form prepared, which will accompany the sample.

A traveler form is a document on which all sample transfers are recorded. When the sample is passed between individuals this movement is recorded by the signature of both the donor and the receiver, and this is repeated whenever a transfer is made. When movement of the sample is to an off-site location the transport container will be packed and uniquely sealed in the presence of the accompanying inspector, who will take it into his/her custody and sign the traveler form. When the sample reaches its destination the receiver will sign it into their custody only on opening the transit container and confirming the integrity of the sample. Whenever a sample is subdivided, the subdivision is allocated a new number, usually an extension of the parent number with an appropriate suffix. This subsample will then be treated in exactly the same manner as the original and a new traveler form will be initiated.

Whenever a sample is taken for subsequent analysis, whether on- or off-site, it will be subdivided or split. The number of subdivisions will depend on the eventual disposition of the sample; for example, an initial distribution between the IT and the inspected State Party may be followed by an eight-fold split for distribution off-site. Each subsample will be put into a clean sample container and labelled and sealed with a security seal. Blank samples will be treated in exactly the same way.

2.5 Sample Splitting

The techniques used for splitting samples will vary according to sample type. The process for splitting liquid

Table 1 List of contents of a typical sampling kit

Item description	Number in kit	Comment
Transit case	2	Size commensurate with contents
Polaroid® camera	2	Automatic date and time stamp
Polaroid® film	5 packs	Packs of 10
Pens and permanent markers	6	3 of each
Log-sheets, carbonless, duplicate	2	Packs of 50
Clipboard	2	Dependent on number of subteams
Seals	100	Tamper indicating
Portable refrigerator	1	15–20 L capacity
On-site equipment box, disposable	8	Collapsible
On-site sample transport container	2	Air tight
Toxic Transport Container, small samples	2	Off-site transport
Protective Sample Container, large samples	2	Off-site transport
Groundsheet, chemically resistant	8	2 m × 2 m
Pegs, plastic	32	For groundsheet
Bottles, glass, wide neck, 500 mL	8	Soil samples
Bottle, glass, wide neck, 50 mL	72	Solid bulk or soil subsamples
Bottle, glass, narrow neck, 500 mL	8	Water samples
Bottle, glass, narrow neck, 50 mL	72	Neat liquid or water subsamples
Bottles, glass, narrow neck, 10 mL	128	Neat liquid or bulk solid subsamples
Vacutainer® tubes	8	Neat liquid samples
Vacutainer® accessories	8	Needle, Teflon® tube and weight
Bags, chemically resistant, 20 × 30 cm	50	Nylon/Mylar®
Swabs, clean cloth or cotton wool	8	Individually packed
Self-locking forceps	8	Hemostat
Pump, battery, 4L min^{-1}	2	Air samples, intrinsically safe
Tubing, butyl rubber	10 m	Compatible with pump and manifold
Manifold, 8-way	2	Air subsamples
Tubes, adsorbent, Tenax®/Carbosieve®	64	
PVC adhesive tape	3 rolls	Assorted colors
Pipette, plastic, 2 mL	8	Disposable
Syringe, plastic, 100 mL	8	Luer-lock®
Needle, 18 g	8	
Stopcock, 3-way, Luer-lock®	8	
PTFE tubing	50 m	Compatible with stopcock
Extending dipper, PTFE, 400 mL	1	Extendable to 3 m
Scoop, stainless steel	4	10 cm blade
Spatulas	10	Assorted
Sieve, 1.7 mm mesh	8	Sample splitting
Bowl, stainless steel, 1 L	8	Sample splitting
Aluminum foil, 30 cm	100 m roll	Sample splitting
Board, polythene, 25 × 25 cm	5	Sample splitting
Glove bag, Mylar®, 100 cm^3	8	Sample splitting
Scissors, large and small	4 of each	
Knife, multitool	2	Swiss army knife
Knife, retractable blade	2	Stanley knife®
Hacksaw, junior, with blades	2	
Head-torch, intrinsically safe	4	
Flashlight, intrinsically safe	2	
Charcoal granules	1 kg	Sample packaging
Fuller's earth	1 kg	Sample decontamination
Batteries, assorted	1 set	All battery operated equipment

PVC, poly(vinyl chloride); PTFE, poly(tetrafluoroethylene).

samples or swab/wipe samples is much simpler than that for solid samples. For liquids, a pipette is used to place equal volumes of well-mixed sample into eight separate containers. For swab/wipe samples, a solvent is first used to extract any soluble nonpolar analytes from the material. This procedure is carried out at least twice, then the resultant solutions are mixed and divided equally between eight containers. The swab/wipe is allowed to dry, so any remaining solvent evaporates, then it is extracted twice with water to remove any

polar analytes and the resultant solutions are mixed and divided equally as described for liquid samples. If aqueous samples are taken and then subsampled for analysis off-site a preservative is added to prevent degradation due to bacterial action. The addition of approximately 0.1% methylene chloride to the main sample followed by thorough mixing will ensure stabilization until analysis can be carried out. After standing for a few minutes the sample can be subdivided.

When splitting soil or bulk solid samples the process has to be more thorough to ensure homogeneity and give fully representative samples for analysis. The rationale behind subsampling or splitting for subsequent analysis is normally to provide one sample that is representative of the whole, but in this case it is to provide eight similar samples. The method chosen by the OPCW for carrying out this process is cone-and-quartering, the only requirement of the method being that the sample is dry and free-flowing. Cone-and-quartering is an internationally accepted procedure for subsampling nonhomogeneous soil and bulk samples. Because of the possibly toxic nature of the samples the procedure is carried out within the confines of a fumehood or an impermeable glovebag, or alternatively it may be carried out in the field wearing suitable protective clothing and a respirator. After sieving (1.7 mm mesh) and thorough mixing about 400 g of the sample is placed on a clean nonabsorbant surface (25 cm square board covered with aluminum foil) in a cone-shaped pile. The pile is then flattened with a second board until it is about a quarter of its original height and divided into four roughly equal parts by making right-angled cuts through the centre of the pile with a spatula. One pair of opposite corners is then transferred to another board and the quartering process is repeated for the two re-formed piles. The pairs of opposite corners from the two boards are then transferred to two more boards and the process is repeated for all four boards. After quartering, the eight pairs of opposite corners can be placed in clean sample jars using the spatula. The aluminum foil is removed from the boards for decontamination and disposal, and new foil is used for each additional sample. Other methods available for splitting samples include ''riffling'' or the use of rotary sample dividers, but the method described is convenient in that it uses few items of kit, enables easy decontamination and is widely applicable.

2.6 Sample Transportation

During the course of inspections samples may be taken which will be neat chemicals or highly contaminated environmental samples. If these samples are to be analyzed off-site then they must be transported in such a way as to preserve the sample integrity and also ensure the safety of the carrier. The integrity of the sample should be preserved by ensuring that measures are taken to eliminate or minimize degradation. Factors that can influence the stability of the sample include air, water, direct sunlight and temperature, and therefore exposure to these effects should be kept to a minimum. Generally it is best to pack the samples in packaging which excludes light, moisture and air, undertake analysis as soon as possible after sample collection and only open the sample container when ready to carry out analysis. When samples are transported off-site to the OPCW, or to designated laboratories, transportation time should be minimized and this means they will generally travel by air. The transportation of dangerous goods by commercial aircraft is regulated by the International Civil Aviation Organization (ICAO) and the system is based upon recommendations from the United Nations committee of experts on the transport of dangerous goods. A manual based on the ICAO instructions is produced by the International Air Transport Association (IATA) and this is rigorously followed by the commercial airline industry. Dangerous goods are categorised in hazard classes from 1 to 9, dependent upon their type and hazard. Class 6 covers poisonous and infectious substances. CW-relevant chemicals are therefore covered by hazard class 6.1, with class 6.2 covering infectious substances. Dangerous goods are also classified in three packing groups based on the associated hazards. For chemicals this is based on their toxicity and for CW-related compounds a full descriptor would be hazard class 6.1, packing group 1. Until recently, the IATA regulations stipulated that the transport of CW-related chemicals by air was strictly forbidden; however this has been addressed by the use of special containers specifically designed for the OPCW by the United Kingdom and the United States of America. These systems meet the criteria laid down by the ICAO for the air transport of CW-related chemicals and have been granted a Certificate of Packaging Performance (packing instruction 623). An IATA special provision (A106) applies to the movement of samples, and this states that the provision may only be used for samples of chemicals taken for analysis in connection with the implementation of the CWC. The UK Toxic Transport Container allows for the packaging of a total of 800 mg of neat material, or eight vapor adsorption tubes, or 80 g of dilute material, or a proportional mixture, contained in eight separate primary containers. The US Protective Sample Container allows for the packaging of a total of 350 mL of unspecified chemical sample, contained within seven primary containers.

When packing samples for transportation, mixing of different samples within the same secondary container should be avoided, although packing separate aliquots of the same sample is permissible. Once the secondary

containers have been packed into the transport system it should be sealed with a high-security seal, such as a fiber optic seal, and the seal record can then be checked for tampering on arrival at its destination. When samples are moved in this way a member of the IT will be assigned to accompany the container on its journey to the OPCW laboratory. The team member will travel with the container whenever possible, witnessing transfers, loading and unloading, to ensure the chain of custody is not broken.

The IT is responsible for ensuring that all the required documentation is completed whenever samples are shipped off-site. Although the bulk of this documentation will be prepared by the OPCW in advance, the IT should be familiar with, and capable of processing, all the necessary forms. The IT (shipper) will fill out the ''Shippers Declaration for Dangerous Goods''. This form contains all necessary information and must be filled out correctly, in English, and contain no alterations, unless signed by the shipper. The form must contain the proper name of the goods as listed in the IATA regulations. For CW-related materials this is ''chemical sample, toxic, liquid or solid'', the hazard class 6.1 and the packing instruction 623. An air waybill for dangerous goods is also required.

Once samples have been signed into the secure storage at the OPCW facility any access will be audited, to ensure the chain of custody is not broken, until the samples are despatched to designated laboratories for analysis.

2.7 Sample Collection Techniques

The sample collection process begins in earnest upon arrival at the inspection site. In most cases time will be at a premium, and therefore strict organization of all stages of the process is essential. The sampling and detection equipment should be unpacked and organized as soon as possible after arrival on-site. If a laboratory is also to be set up, then two separate but easily accessible areas should be established. This will allow for good communication and also prevent any contamination control problems. The first stage of any inspection is to carry out a ''windshield'' tour of the facility, and depending on the size of the site this may either be on foot or in vehicles. This familiarizes the inspectors with the site layout, gives a reasonable idea of sampling locations and facilitates the final planning of the inspection. This stage in the inspection process, before any equipment is exposed to contaminated areas, provides the first opportunity for collection of blanks. Because actual samples have yet to be decided upon it will be necessary to collect a variety of blanks, concentrating on air, water and soil. The selection of which blanks to take and which areas to sample will be at the discretion of the IT, who can only decide after careful consideration of the characteristics of the site. The use of detectors to show the absence of CW contamination will aid in the collection of these samples.

After the initial formalities of site briefings and negotiation and planning procedures have been dealt with, the inspection activities will begin. Depending upon the number of inspectors present on the mission, one or more subteams will carry out a thorough investigation of the site. The site safety assessment undertaken in the initial stages of planning will have determined the hazards present on-site and will dictate the level of protective clothing to be worn during any stage of the inspection process. The following description of the sampling process assumes that the IT will be taking the samples. If, on-site personnel carry out sample collection most of the procedures will still apply, as the IT will need to witness the act of sampling and the on-site personnel will be using equipment belonging to the IT. During the course of the inspection the sample collection teams will be subjected to the most arduous and dangerous aspects of the verification procedure, therefore they should be fully trained and appreciate all aspects of the chemical and physical hazards to which they could be exposed. They must be able to act appropriately in emergency situations brought about by physical injury or exposure to chemicals and therefore must be trained in all aspects of basic first aid.

2.8 Sample Collection Team

To carry out the process of sample collection, the size of a team should necessarily be a minimum of three persons. This is the required number to ensure safety and sample integrity. Before starting any sampling operation a contamination control line (CCL) must be established in an area known to be free of contamination and preferably upwind of the sampling site. It is normally best to indicate this line by means of tape, and a rigorous regime of controlled entry and exit to the sample collection site should be applied. The kit assembled at the CCL will include decontamination facilities, sampling kit, detection equipment and on-site transport containers. The individual team members are defined in terms of their potential for contamination, and the team is made up of a ''hot'', a ''warm'' and a ''cold'' person, ranked in order of the possibility of contact with contamination. In larger teams there may be more than one cold person. The duties of the hot person include surveying the suspected contaminated area with detectors to identify ''hot spots'' and safe entry/exit routes, selection of equipment and sample collection. The warm person acts as an assistant to the hot person by retrieving tools and equipment as requested, if necessary holding the sample container during sampling, and decontaminating the outside of the container before dropping it into a bag held by the cold person. The warm person also acts as a safety link between the CCL and the sampling point, transmits information to the cold person to record on the sample log-sheet and

may also photograph the sampling area if it is inaccessible to the cold members of the team. The cold person controls the sample collection process, and their duties include recording all relevant information as each sample is taken, photographing the area if feasible, passing sampling equipment to the warm person when requested, collecting, labelling and packing decontaminated sample bottles and bags, and checking personnel across the CCL after personal decontamination has taken place.

2.9 Sample Types

The number of samples taken may be strictly limited by the analysis time available. Some form of prioritization will therefore need to be addressed, so that the best samples are obtained for analysis. The type of inspection and the specific site concerns will always determine the sort of sample available. For instance, an ''alleged use'' investigation will provide different sampling opportunities compared with investigation of an industrial site. In general terms, the following decreasing order of priority should be observed: category one samples, then category two samples, and finally category three samples. Category one samples are likely to contain neat chemicals, and include munitions contents, residues or fragments from munitions, bulk storage containers, chemical reactors and process lines. Category two samples include environmental samples, such as soil or water taken from the point of a munitions impact; contaminated paints, plastics or rubber; swabs from munitions or chemical plant; air samples from enclosed areas; and equipment or clothing from affected personnel. Finally, category three samples include body fluids such as blood or urine, vegetation and dead animals or insects.

The sampling kit, as described in Table 1, provides an adequate selection of equipment to deal with any sample collection scenario. The ancillary equipment, such as detectors, tools and general inspection equipment, which will accompany all ITs is not listed.

Collection methods or ROPs for the various types of sample that will be taken for CWC verification are similar to those described in the field of industrial and environmental sampling.[2,3] ROPs for air, solid and liquid samples are also listed in AEP-10 and in the ''Blue Book''.[4,5] These publications describe the step-by-step process and list the equipment necessary to obtain each type of sample. General guidelines for the collection of various types of sample follow, and in all cases, because of the hazardous nature of the different types of sample, it is essential that the safe operating procedures listed previously are followed. It is also necessary to remember that sufficient material should be collected to allow for subsequent splitting of the sample.

2.9.1 Bulk Samples

Bulk samples from munitions, storage tanks, drums, reactors and process or filling lines are, in essence, neat chemicals or concentrated solutions and therefore only a relatively small amount of material is required for analysis. For solid material, a spatula should be used to remove approximately 10–20 g of material and place it in a glass wide-neck bottle with a Teflon®-lined cap. For liquid bulk samples in containers which are easily accessible several techniques are available and the sampler will need to decide upon the most appropriate method. Collection can be carried out using the Vacutainer® system, a syringe, a pipette or by simply utilizing container facilities such as fitted taps. The sampling kit does not allow for samples to be taken from intact munitions, except in those cases where munitions experts are on-hand to give advice and if possible remove filling caps when the munition has been declared safe. Samples from reactors and process or filling lines should only be taken when the specialist advice of a chemical engineer or plant operative is available. The main concern with chemical plants is that the contents may be under increased or decreased pressure and may also be at an elevated temperature. In this case it is sensible to allow an experienced operative to take the sample while the process is witnessed by the sampling team. The amount of sample required is again small and 10–20 mL should suffice. The exception is in those cases where the sample is known to be a dilute solution of chemicals in a solvent, when a larger amount (approximately 400 mL) should be taken. Because of the extreme hazard associated with neat chemicals it is essential that the most rigorous safety regime is adopted for this sample type.

2.9.2 Air Samples

Air or vapour samples are taken for several reasons during inspection activities and may be used for verification or safety. In a safety role, and provided an on-site laboratory is available, air monitoring can be carried out to establish background levels of chemical contamination, thus aiding the IT in the operation of its mission by enabling appropriate protective measures to be chosen. In a verification role air sampling is carried out if it is felt that other sample types are either unobtainable or will give inconclusive results. For example, a complicated sample matrix may give poor analytical results whereas vapour collected from off-gassing gives a clear result. It may also be impossible to collect material samples from a demolished building or munitions bunker, whereas a sample may be taken by inserting a probe into the area and sampling onto the adsorbent tube. Samples may be taken using either a mechanical or a hand pump, and a collection volume of approximately 10 L should be sufficient. The tubes supplied with the sampling kit include both Tenax® and

Table 2 Abbreviations, common names and chemical names for chemical agents

Abbreviation	Common name	Chemical name
GB	Sarin	Isopropyl methylphosphonofluoridate
GD	Soman	Pinacolyl methylphosphonofluoridate
GA	Tabun	Ethyl *N*,*N*-dimethylphosphoramidocyanidate
GF	Cyclohexyl sarin	Cyclohexyl methylphosphonofluoridate
VX	V agent	Ethyl *S*-2-diisopropylaminoethyl methylphosphonothiolate
AC		Hydrogen cyanide
CK		Cyanogen chloride
SA		Arsine
CX		Phosgene oxime
ED	Dick	Ethyldichloroarsine
MD		Methyldichloroarsine
PD	Pfiffikus	Phenyldichloroarsine
DA	Clark I	Diphenylchloroarsine
DC	Clark II	Diphenylcyanoarsine
DM	Adamsite	Phenarsazine chloride
H	Sulfur mustard	Bis(2-chloroethyl)sulfide
HD	Distilled mustard	Bis(2-chloroethyl)sulfide
Q	Sesquimustard	Bis(2-chloroethylthio)ethane
T	*O*-Mustard	Bis(2-chloroethylthioethyl)ether
HN1	Nitrogen mustard 1	Bis(2-chloroethyl)ethylamine
HN2	Nitrogen mustard 2	Bis(2-chloroethyl)methylamine
HN3	Nitrogen mustard 3	Tris(2-chloroethyl)amine
LI	Lewisite 1	2-Chlorovinyl dichloroarsine
LII	Lewisite 2	Bis(2-chlorovinyl)chloroarsine
LIII	Lewisite 3	Tris(2-chlorovinyl)arsine
CG	Phosgene	Carbonyl chloride
DP	Diphosgene	Trichloromethyl chloroformate
PS	Chloropicrin	Trichloronitromethane
CA/BBC	Camite	Bromobenzylcyanide
CN/CAP		Chloroacetophenone
CS	Tear gas	o-Chlorobenzylidenemalononitrile
CR		Dibenzoxazepine
BZ		3-Quinuclidinyl benzilate

Carbosieve® as sampling media. These tubes should be capable of dealing with a wide range of chemicals although exceptions include those with high boiling points, such as ethyl *S*-2-diisopropylaminoethyl methylphosphonothiolate (VX) and bis(2-chloroethylthioethyl)ether (T), which are difficult to thermally desorb from the tubes but may be eluted with solvents. Table 2 lists abbreviations, common names and chemical names of classical chemical agents.

2.9.3 Soil Samples

The collection of solid/soil samples will be associated mainly with the following types of inspection: alleged use, storage facilities and industrial sites. The retrospective identification of chemicals used on-site, or recently deployed in a military sense, is possible by analysis for intact CWC-related chemicals or more likely their degradation products. It may be possible to identify areas of interest for sampling in several ways. Color differences from the surroundings may indicate areas of contamination, detector responses will be obtained from highly contaminated ground, and areas of disturbance such as bomb craters or dumping grounds will be visible. Approximately 400 g of sample will be required for splitting and subsequent analysis; an area approximately 10 cm × 10 cm should be selected and soil removed to a depth of 1–2 cm. A spatula or stainless steel scoop is used and the sample is placed in a 400 mL wide-necked bottle with a Teflon® cap, or in a Mylar® bag which is then sealed with tape.

2.9.4 Water Samples

Aqueous liquid samples will be taken for similar scenarios to soil samples. In this case, however, detectors will be of limited use and sampling points will need to be identified from pointers on the ground and surroundings. Sampling opportunities to look for include pools of water near or in bomb craters; runoff water collected in depressions; and in industrial facilities, wastewater containers, drains and treatment plants. The selection of sampling kit will vary from sample to sample and common sense on the part

of the sampler will be needed. There should, however, always be a way of obtaining a sample by using a syringe, pipette, Vacutainer® or dipper, and in all cases the amount collected should be approximately 400 mL to allow for subdivision.

2.9.5 Swab Samples

The use of swabs or wipes to collect samples is a technique which allows samples to be taken from surfaces previously wetted or exposed to chemicals. In situations where munitions or chemical plants are found to be leaking this may be the only available technique for acquiring samples. Other collection situations include contaminated ventilation equipment, shelving or walls inside storage bunkers, loading trolleys, vehicles and other equipment exposed to chemicals. Swabs can be made of either pre-cleaned cloth or cotton wool. The technique is to hold the material with a pair of forceps, moisten the swab with a solvent (normally isopropanol), and then swab an area of approximately 15 cm × 15 cm. The swab is then placed into a clean screw-top bottle and subsequently extracted with further solvent before analysis.

2.9.6 Material Samples

Other samples may be taken if the sampling team considers it necessary. In some scenarios sampling opportunities will be few and the best sample may not necessarily be of a conventional type. Some examples include material gaskets or rubber flanges from chemical plants, door seals from storage bunkers, filter units from ventilation equipment, electrical cabling, vehicle tyres, plastic storage containers and protective clothing. These items are all made from plastic or rubber materials and as such they are extremely good sampling media for organic chemicals. Samples can be taken using any of the cutting implements in the sampling kit and then they should either be placed in glass wide-necked bottles or double-wrapped in Mylar® bags.

3 SCREENING

In this context, screening is the use of specialized equipment to carry out a survey of a site to determine areas of a hazardous nature to the sampling team. It also aims to provide an indication of the presence of chemical contamination and pinpoint the site of maximum contamination. This is true for many situations, but for trace levels of contamination, for instance in alleged use investigations, the use of detectors will not always be feasible. The use of detectors in the inspection of CW-related sites will assist the IT in several important ways. Three areas where screening with chemical detectors can help to expedite the mission are safety monitoring, contamination mapping and sample selection.

4 DETECTION

4.1 Use of Detectors

A number of systems can be used for the real-time monitoring of the hazards from volatile chemical materials present on-site. Other techniques that can determine the hazards from nonvolatile material are not real-time, but could be employed in an on-site laboratory to provide effective monitoring, allowing contaminated areas to be identified. Many of the field detectors in current use by armed forces actively sample air but are only set up to detect a limited number of volatile chemicals of CWC concern. The low-volatility agents are detected by modifications of these detectors or by different tests. A selection of current screening technologies is briefly discussed and reviewed in terms of their use on-site for screening for chemicals identified as presenting a hazard. A more thorough treatment is presented in the article **Detection and Screening of Chemicals Related to the Chemical Weapons Convention** in this publication.

4.2 Ion Mobility Spectrometry

Ion mobility spectrometry (IMS) is employed in hand-held, field-deployable instruments which are very useful in the real-time determination of hazards presented by certain CWC-related chemicals. It is recommended that these instruments are used in the field to aid identification of suitable sampling sites, as a safety system in the CCL, and in an air monitoring role. Two instruments that work on this principle are employed by the OPCW: the chemical agent monitor (CAM, Graseby Ionics, UK); and the German rapid alarm and identification detector (RAID-1, Bruker-Saxonia Analytik, Germany). In its normal set-up a CAM will detect H (sulfur mustard) and the nerve agents GA, GB, GD and VX, but it is not set up to detect the whole spectrum of chemicals listed in the CWC schedules. It is possible that the programming could be changed so that a CAM could detect volatile agents of concern to a particular site, but this is not a modification currently within the skills of an inspector. The RAID-1 unit has the capability of detecting both sulfur and nitrogen mustards (H, HN1, HN2, and HN3), lewisite, the nerve agents GA, GB, GD, GF and VX, and also the blood agents AC, CK and CG. It has a more sophisticated display than a CAM in that it will also display the identity of the chemical detected and its agent group, and give an audible alarm. Both detectors display

a measured concentration range and can be deployed in a stationary or mobile role. They have similar sensitivities, i.e. 0.1 mg m^{-3} for the blister agents (H, HN1, HN2, HN3, and LI) and VX and 0.03 mg m^{-3} for the G-agents.

4.3 Flame Photometric Detection

Flame photometric detection (FPD) technology is used in the hand-held AP2C field detection system developed by the French (PROENGIN, France). The detector can be used in the field in a similar role to the IMS systems. It is set up to detect volatile chemicals in both the sulfur and phosphorus modes. This limits the use of this system on-site to the detection of volatile sulfur- or phosphorus-containing CW-related compounds. The system differentiates between the two families of chemicals but does not have the selectivity demonstrated by IMS detectors. The sensitivity of the AP2C detector is 0.01 mg m^{-3} for most phosphorus compounds and 0.6 mg m^{-3} for sulfur mustard (H).

4.4 Detector Papers

Military detector papers are used on the battlefield to determine the presence of liquid agents. They employ the principle of microencapsulated dye solubility to give rise to color development at the paper's surface. These papers are of limited value for on-site use except where ground or surface contamination is suspected. They can be used in a site screening role by taping the paper to a roller fixed to a pole, which is then rolled over suspected ground contamination. There are two types of paper available: single color, which responds to all CW-related chemicals; and three color, which discriminates between mustard (red), G-agents (amber) and VX (green).

4.5 Detector Tubes

Detection tubes containing reagents immobilized on a solid adsorbent are used for sampling gases and vapors. Two systems are employed by the OPCW inspectors: the Czech military chemical detection kit ORI-217 (ORITEST, Prague), and the commercial Auergesellschaft/Mine Safety Appliances Gas Tester IIH kit (AUERGESELLSCHAFT, Germany). The detector tubes are either chemical-based in which the analyte reacts with a specific reagent giving a characteristic color change, or enzyme-based where the analyte of interest inhibits the enzyme. The ORI-217 kit specifically targets a range of classical CW-related compounds and contains four types of tube for the detection of the nerve agents GA, GB, GD, GF and VX, the choking and blood agents CG, DP, CK and AC, and the blister agents H and LI. The AUER/MSA kit is a general purpose gas-testing kit which utilizes a wide range of tubes mainly concentrating on the safety monitoring of industrial gases and vapors, for example chlorine, AC, CG, hydrogen sulfide and sulfur dioxide.

4.6 On-site Analysis Equipment

In addition to the detector systems mentioned previously, the use of on-site analysis equipment which can be used to rapidly screen samples in the field or in small portable laboratories will aid the verification process. For the present, OPCW inspectors have to rely on the use of detection equipment, sampling and on-site analysis to map the contamination of the site. In the future, especially when large areas are involved, other quick on-site analytical screening methods could be utilized. Two such items of equipment, one an off-the-shelf commercial product, the second a product in development, are briefly described below.

4.6.1 Element Specific Detection

Atomic fluorescence total arsenic determination is a quick screening technique for different arsenic compounds believed to be present on-site. It may, in certain investigations, be important to have an indication of the total levels of arsenic present in organic extracts and aqueous extracts of soil and other matrices. This can easily be achieved using an atomic fluorescence detector (e.g. The Excalibur Plus system, P.S. Analytical, UK). The sample is reacted with acidic borohydride to produce SA gas, which passes on through a flame and is detected by the atomic fluorescence detector. This is a very simple system which could be set up in an on-site laboratory to give a fast screening capability to aid in the pinpointing of contamination on site.

4.6.2 Specific Degradation Product Detection

An emerging technology for site screening is the chemiluminescence (CL) screen for thiodiglycol,[6] the major breakdown product of H. Porton Down has developed a fast, sensitive and specific screen for thiodiglycol in soil. The thiodiglycol in the sample is oxidized with sodium hypochlorite to produce a high-energy species. The energy from this species is transferred to a rhodamine B sensitizer. The CL produced from this reaction is detected and is directly related to the quantity of thiodiglycol present in the extract. The thiodiglycol is extracted by simply shaking with a small quantity of methanol. An aliquot of this extract is injected into the flow injection system and the result is determined within a minute. This system is applicable to field operations and could easily be set up in a mobile laboratory. The presence of mustard in the sample is also easily determined

by carrying out a simple hydrolysis step and re-measuring the thiodiglycol. Mustard, although a World War I chemical agent, is still one of the major CWC concerns and a quick and reliable screen will aid the CW inspectors in inspections of alleged use, production and storage.

ABBREVIATIONS AND ACRONYMS

AC	Hydrogen Cyanide
BZ	3-Quinuclidinyl Benzilate
CA/BBC	Bromobenzylcyanide
CAM	Chemical Agent Monitor
CCL	Contamination Control Line
CG	Carbonyl Chloride
CK	Cyanogen Chloride
CL	Chemiluminescence
CN/CAP	Chloroacetophenone
CR	Dibenzoxazepine
CS	o-Chlorobenzylidenemalononitrile
CW	Chemical Weapons
CWC	Chemical Weapons Convention
CX	Phosgene Oxime
DA	Diphenylchloroarsine
DC	Diphenylcyanoarsine
DM	Phenarsazine Chloride
DP	Trichloromethyl Chloroformate
ED	Ethyldichloroarsine
EIF	Entry Into Force
FPD	Flame Photometric Detection
GA	Ethyl *N*,*N*-dimethylphosphoramidocyanidate
GB	Isopropyl Methylphosphonofluoridate
GD	Pinacolyl Methylphosphonofluoridate
GF	Cyclohexyl Methylphosphonofluoridate
GPS	Global Positioning System
H	Bis(2-chloroethyl)sulfide
HD	Bis(2-chloroethyl)sulfide
HN1	Bis(2-chloroethyl)ethylamine
HN2	Bis(2-chloroethyl)methylamine
HN3	Tris(2-chloroethyl)amine
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
IMS	Ion Mobility Spectrometry
IT	Inspection Team
LI	2-Chlorovinyl Dichloroarsine
LII	Bis(2-chlorovinyl)chloroarsine
LIII	Tris(2-chlorovinyl)arsine
MD	Methyldichloroarsine
NATO	North Atlantic Treaty Organization
OPCW	Organization for the Prohibition of Chemical Weapons
PD	Phenyldichloroarsine
PPE	Personal Protective Equipment
PS	Trichloronitromethane
PTFE	Poly(tetrafluoroethylene)
PVC	Poly(vinyl chloride)
Q	Bis(2-chloroethylthio)ethane
RAID-1	Rapid Alarm and Identification Detector
ROP	Recommended Operating Procedure
SA	Arsine
T	Bis(2-chloroethylthioethyl)ether
VX	Ethyl *S*-2-diisopropylaminoethyl Methylphosphonothiolate

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements • Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Ion Mobility Spectrometry in Forensic Science

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction • Sampling and Sample Preparation in Process Analysis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. *Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction*, printed and distributed by the Technical Secretariat for the Organization for the Prohibition of Chemical Weapons, The Hague, 1997.
2. N.T. Crosby, I. Patel, *General Principles of Good Sampling Practice*, Royal Society of Chemistry, Cambridge, 1995.
3. *Principles of Environmental Sampling*, ed. L.H. Keith, American Chemical Society, Washington, DC, 1988.
4. *NATO Handbook for Sampling & Identification of Chemical Warfare Agents*, Allied Engineering Publication 10 (AEP-10), 4th edition, North Atlantic Treaty Organization, Brussels, 1995.
5. *Recommended Operating Procedures for Sampling and Analysis in the Verification of Chemical Disarmament*, The Ministry for Foreign Affairs of Finland, Helsinki, 1994.
6. T.J. Christie, 'The Development of a Liquid Phase Chemiluminescence System for the Detection of Organosulfur Compounds', PhD Thesis, University of Hull, 1996.

Clinical Chemistry: Introduction

Merle A. Evenson
University of Wisconsin, Madison, USA

1 Introduction 1
2 Past 2
3 Present 3
4 Future 4
Abbreviations and Acronyms 5

1 INTRODUCTION

The amount of chemical knowledge in the field of clinical chemistry is enormous, is growing rapidly and the task of restricting information presented to an encyclopedia-sized document is difficult. Nearly every topic included in undergraduate quantitative analysis and instrumental analytical chemistry courses is used in clinical chemistry laboratories as well as many methods from the fields of biochemistry and medicine.

An analytical chemistry mission statement for the clinical chemistry portion of clinical laboratories could be stated as: the application of analytical chemistry principles, methods and instruments to measurements of solids, fluids, tissues, cellular components and other biological materials for the measurement of more than 500 analytes to produce data to help in the diagnosis and prognosis of human diseases.

Organization of this section and the entire encyclopedia is a daunting task. There are more than 600 articles in total and more than 30 articles in the Clinical Chemistry section that will focus on the more common topics. It is an organizational dilemma as to what to include in the clinical chemistry section.

One way to view the field of clinical chemistry is to examine the ''organ based medical curriculum'' as is often used to arrange the medical school courses. This would allow us to organize clinical chemistry to include those laboratory tests that seek information about organ systems. The organ systems may include the brain and peripheral nervous system, lungs, kidneys, liver, muscles, bones, bone marrow and all the endocrine and exocrine gland systems. One could also organize the section by the classes of analytes frequently encountered in clinical chemistry and those classes of tests could include illicit and prescription drugs, lipids, electrolytes, blood gases, endocrine hormones, glucose, blood urea nitrogen, creatinine, DNA–RNA, proteins, peptides, enzyme activities, enzyme mass and therapeutic drug monitoring. Still another way to view the field of clinical chemistry would be to examine the laboratory tests of the various organ systems when and after those systems are challenged. For example, one would study the pancreas by examination of glucose concentration change after the subject ingests a large dose of glucose (glucose tolerance test). Kidney, nitrogen metabolism, endocrine function and liver function tests also may use additions of substrates or metabolites to glean information about the function of the organ system of interest. Heart, lungs and kidney function, however, are usually conducted by analysis of naturally occurring metabolites without an exogenous challenge. For organization of this section, a sampling of all three of the above approaches was used.

A goal of this section is to rely upon the other sections for most of the analytical chemistry principles, procedures, descriptions of the instruments and even borrow the exact methods for application in clinical chemistry cases. For example most of the information in DNA–RNA analyses, on nucleic acid sequencing, on the genome project and on genetic screening are in articles in those sections of this encyclopedia. Most of the information about gel electrophoresis, capillary electrophoresis (CE), point-of-care testing, field flow fractionation, protein and peptide analysis are likewise in those sections that describe the techniques. In a similar manner, ion-selective electrodes, voltametry, amperometry, kinetic measurements, ultraviolet/visible (UV/VIS) and infrared (IR) spectrophotometry, and pH measurements are found in other sections of this encyclopedia. Previously IR was used for kidney and gall stone analysis in clinical chemistry laboratories. Especially interesting are the new proposals concerning the use of near-IR for noninvasive reflective or absorption analysis for glucose monitoring and other metabolite analysis without removal of a sample from the patient. An article on this subject is found in this encyclopedia.

Until recently UV/VIS absorption spectrophotometry was the most common read-out for the analyses performed. With the increases in automation and the huge multichannel analytical instruments, a larger number of measurements are now made with ion-selective electrodes. Fluorescence, phosphorescence, chemiluminescence and the use of radioactive tracers have increased significantly in the past 15 years in support of the use of immunological methods of analysis. Molecules such as antibodies, antigens, proteins and/or protein-like reagents are used in competitive binding and/or isotope dilution-type analyses. Automated analytical instruments have blended computers into the instruments for both process

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

control and for the calculation of results such that computers are an integral part of the measurement package. Many such instruments use kinetic methods, either end-point or rate measurement techniques, to determine enzyme activities important for patient care.

Chromatography has become an indispensable technique in clinical chemistry laboratories. Thin-layer chromatography was widely used for determining the presence of drugs of abuse, for concentrations of therapeutic drugs and for lipid analyses until newer instrumentation became available. The use of high-performance liquid chromatography (HPLC), gas–liquid chromatography (GLC), CE, supercritical fluid chromatography, size exclusion chromatography and other separation instruments are directly interfaced to low resolution mass spectrometers. These types of instruments have become more common in clinical chemistry laboratories. These instruments are used both for forensic testing and for toxicological identifications in clinical laboratories. Modification of these instruments using thermospray, electrospray volatilization and matrix-assisted laser desorption/ionization (MALDI) has been achieved for proteins, peptides and hormones in the clinical chemistry laboratory.

Emission flame spectrophotometry for sodium, potassium and lithium and atomic absorption spectrometry (AAS) for calcium and magnesium in serum and urine were the standard techniques used for years. Flame atomic absorption for calcium, magnesium and in some cases lithium became the method of choice in the 1960s. More recently, ion-selective electrodes have replaced these techniques in the larger more modern clinical chemistry laboratories. Graphite furnace AAS and inductively coupled plasma (ICP) emission and mass spectrometry (MS) have become more routine in some clinical chemistry service laboratories. ICP multielement trace metal emission and MS instruments are now more widely used in the analysis of biomedical samples obtained from patients.

When we examine the subjects in the fundamental and applied reviews of *Analytical Chemistry*, published in alternate years, we will notice that nearly all the subject headings in these two reviews are used daily in clinical chemistry laboratories. Hence, the subject matter knowledge required in these types of service laboratories is indeed encyclopedic in breadth and in depth!

2 PAST

If we look historically at this field, we will notice that biochemists began working in hospital laboratories in the USA at about the turn of the century – mostly in large cities like Boston, New York and in other eastern cities where Ivy League colleges are located. Shortly thereafter, many of these biochemists sought more academic freedom, started independent departments, created new journals and began working on nonhuman biological problems. If one looks at the first issue of the *Biochemical Journal* published in 1906, one notices that in northern Europe and in the UK there were many studies on measuring components of blood and urine. In the USA a few of the original clinical biochemists remained closely connected to medical schools. Folin, often called the first clinical biochemist, and Wu of Harvard, Peters at Yale, Van Slyke of New York and Somogyi in St Louis became the founders of the new field of quantitative clinical chemistry. In the mid-1940s hospital biochemists in the New York City area began to organize a professional organization that would eventually become the American Association of Clinical Chemistry (AACC). In the late 1940s, led by Max Friedman, the journal *Clinical Chemistry* was started. In the early 1960s, *Analytical Chemistry* began to publish articles by academic chemists on kinetic measurement of enzymes, on enzyme electrodes, and on enzymes as analytical reagents. *Biochemistry*, first published in 1962, had several articles on the use of physicochemical methods to study enzymes and proteins in blood and urine. The earlier mathematical focus, on enzyme kinetic studies, borrowed from physical chemistry by Michaelis and Menten and adapted to a graphical presentation by Lineweaver and Burk, became the foundation for a whole field of study by the clinical biochemists. In the mid-1960s analytical chemistry divisions in mostly midwestern universities began to apply the ''Measurement Science'' component of analytical chemistry to clinical chemistry. These university research laboratories developed the early instruments using enzyme kinetics as the basis for their instrumentation. About this same time in other academic analytical chemistry research laboratories, some automated instruments were being built for the drug industry. Kolthoff, at the University of Minnesota, often considered one of the founders of modern analytical chemistry as a separate division in chemistry departments, noticed that analytical chemistry measurements had much to contribute to the pharmaceutical industry for their drug analysis needs and he promoted this application.

In the late 1950s Technicon of Ardsley, New York took the Auto Analyzer developments of Skeggs to commercial availability. This was the beginning of a flurry of mechanization, automation and instrument building that would soon become much larger than originally thought possible only a few years earlier. Other large companies, Du Pont, Beckman, Kodak, Xerox, Dow, Perkin-Elmer, Hewlett-Packard and others, created instruments and products for the centralized clinical chemistry laboratory. Since the early 1960s enormous growth has continued in the contributions of analytical chemistry to clinical chemistry. The application of absorption spectrophotometry, enzyme

kinetics, statistics, electrochemistry, chromatography, and computers for data collection, for data calculations and reporting of results, along with record keeping, and billing of charges have become subjects of study and contributions by analytical chemists.

It is easily evident that the two major roots for clinical chemistry evolved from the field of biochemistry (for the early biochemistry principles) and then later attracted interest from analytical chemistry (for the development of instrumentation, quantitation, and specific method development). More recently, instrument development has focused on hand-held instruments. ''Point-of-care'' testing instruments such as glucose meters for diabetics, home test kits and portable pH meters are examples. More sophisticated hand-held instruments such as the I-Stat, which can measure up to nine analytes in 90 s with a single drop of blood, are now widely available. These developments are directed at the decentralization of large clinical laboratories and the moving of testing closer to the patient, which will quicken the turn around time of analytical results for patient care.

3 PRESENT

The main objective for the section on clinical chemistry is to provide a broad overview of the field in the evolving health care industry as we enter the 21st century, and to illustrate how most of the fundamental principles, methods and instruments of analytical chemistry are being applied to this field.

The first objective was to select some of the more common methods of analysis and to ask the authors of these articles to illustrate how these analytical approaches, methods, and techniques are used in typical hospital clinical chemistry laboratories. These articles include automation, biosensors, CE, HPLC, gas chromatography/mass spectrometry (GC/MS), planar chromatography, electrochemistry, electrophoresis, IR spectrophotometry, enzymology, UV/VIS spectrophotometry, and DNA analysis in the areas of molecular diagnostics and molecular pathology. In addition, drug testing for clinical toxicology and forensic toxicology cases and the statistical quality control of laboratory results were included for the instrumental analysis areas of this section. Next, a few of the more common human diseases were selected. Topics related to analysis in endocrinology, hematology, heart disease, liver diseases, atherosclerosis, genetic diseases, and diabetes and then authors were asked to describe how measurements from the clinical chemistry laboratory contributed to the information that physicians use in managing the health care of patients with these medical conditions. Finally, authors were selected for some of the more common analytes such as lipids, nucleic acids, drugs, glucose, cardiac markers, electrolytes, blood pH, enzymes, proteins, peptides, and hormones and it was suggested that they illustrate the analytical chemistry used for these analytes and how the laboratory information contributes to patient care. Together, these three approaches provide a representative coverage of the vast range of analytical chemistry techniques used in the very large spectrum of different diseases found in the typical hospital clinical chemistry laboratory.

The large commercial clinical laboratories provide yet another use of instrumental analytical chemistry measurements in clinical chemistry. These companies use large, complex computer-controlled analytical instruments that produce thousands of results per hour at a cost of pennies per result. Another role for commercial clinical chemistry laboratories is to perform extremely difficult and expensive laboratory tests that are requested infrequently of routine hospital laboratories. These esoteric analyses support the management of more complex disease patients.

There are some excellent up-to-date textbooks that provide significant analytical chemistry fundamentals as well as human physiological chemistry principles. Two examples of these textbooks are the *Tietz Textbook of Clinical Chemistry*, edited by Burtis and Ashworth, and published by Saunders, and *Harrison's Principals of Internal Medicine*, edited by Fauci and published by McGraw-Hill. These textbooks are used by the physicians as resources for the interpretation of laboratory results generated by the clinical laboratory. Details of the Fundamentals of analytical chemistry and principles of the instrumental analysis methods are found in other textbooks mentioned elsewhere in this encyclopedia.

Another excellent source of up-to-date books, chapters and published articles on clinical chemistry is published in alternate years by *Analytical Chemistry* in its ''Application Reviews''. There are sections on clinical chemistry, on drug analysis, on forensic analyses and on the analysis of proteins and peptides. All these areas do have applications in clinical chemistry. The *Fundamental Reviews of Analytical Chemistry*, again published in alternate years, provide the principles and physicochemical foundations for the instrumentation discussed in the ''Application Reviews''.

There are a couple of areas of analytical chemistry often taken for granted or not considered when tests are requested. These two areas are sampling errors and the analytical process. One can keep in mind the general outline of the generalized ''analytical process'' listed below. Each time a physician requests a test, the entire list of tasks is utilized. In many situations, the different individuals performing the jobs are unaware of the importance of the whole list and are unaware of how their efforts fit into the overall procedure of producing a

laboratory result. An often-used list of tasks included in the ''analytical process'' is as follows:

1. blood collection (sampling)
2. sample transport
3. validation of sample
4. sample preparation steps
5. choice of methods to be used
6. use of standards, calibrators and controls
7. initial collection of analytical raw data
8. preliminary calculation of analytical results
9. re-calculation considering blank effects, interferences, matrix effects, enhancement or quenching of analytical signal
10. use of statistics to evaluate the preliminary results
11. method comparison of result to different physicochemical analysis of same material when questions arise
12. consideration of other information for analytical result
13. report final result using statistics including accuracy and precision
14. conversion of analytical result into clinical meaning or interpretation
15. re-evaluation of all the steps in the ''analytical process''
16. reporting the final result.

Next a few words about the potential for sampling error. In the clinical chemistry laboratory the most common sample analyzed is human blood. Let us for a moment examine the magnitude of the potential sampling error in that process. In general blood will account for 5–8% of the body weight of an individual. A 75-kg man has between 3.75 kg and 6.0 kg of blood. Let us use 5 kg or 5 L of blood as an example. Venous sampling of blood will usually remove 5 mL of blood. For a manual cell counting through a microscope or for a modern clinical instrument 10 μL of blood or serum (liquid portion of blood with cells removed) will be taken for analysis. For a whole blood cell count, a small drop of blood will be spread thinly on a glass slide and a total of 100 different types of blood cells are counted. From the percentages of each of the cell types of these 100 counted cells, the results will be multiplied to give the millions of cells present in 1 mL of blood. Just think of the potential sampling errors! Yet, this sampling procedure has worked well for almost 100 years.

A similar situation occurs when a clinical chemistry instrument picks up a 10-μL sample, does the analysis for one or more analytes and calculates the number of moles per liter (older common units were milligrams per 100 mL) and then reports these results to the physicians that made the request. Once again, one must realize that the serum is about 50% of whole blood, that perhaps 3–5 mL of blood were drawn from the patient, and the instrument for the analysis used 10 μL. One would expect the sampling error to be large but it has never been a significant problem for clinical chemists.

In summary, all the principles of analytical chemistry are critically important for the accurate and precise results obtained in the clinical chemistry laboratory. In addition, nearly every method and/or technique studied in analytical chemistry is used in hospital clinical laboratories. Hence, by this measure as well as by the financial size, this is a very large business area with plenty of technical challenges for all the analytical and/or clinical chemists that choose to contribute to this field.

4 FUTURE

In the next few years as the human genome sequencing project nears completion, most supporters of this project believe it will have major effects on what we now know about the human disease processes. Many believe that it will be possible for tightly regulated genetic diseases, along with laboratory results, to predict if an individual will acquire that specific disease. In addition, many believe that once all the DNA sequences are known, researchers will find that there is genetic susceptibility to environment challenges. These challenges also sometimes cause disease. It is hoped that the DNA sequences, measured in the clinical laboratory, will be useful in these medical situations. The tentative names for these projected fields of clinical chemistry are molecular diagnostics and/or molecular pathology.

Another area that is expected to become important is the detection of a single nucleic acid substitution in messenger RNA (mRNA). The demands for accuracy in the laboratory for such nucleotide polymorphism testing are unbelievably high. Consider the situation whereby a single nucleic acid substitution occurs in the mRNA, causing a single amino acid substitution in a critical protein or enzyme. Usually more than 99% of the sequences from one individual to another are identical. Therefore, a single nucleic acid substitution must be detected in hundreds of thousands of different mRNA sequences. In addition, the concentration of the essential mRNA is only one-millionth of that of other molecules in that sample. MALDI/time-of-flight MS has become the instrument of choice to obtain the necessary accuracy for these types of analyses. Much greater instrumental capabilities will be needed in the future because, at the current time, as the size of the mRNA increases the resolution decreases. Today sequences greater than 50 start to become difficult and inaccurate.

To date, there have been more problems in genetic analyses using DNA–RNA methodologies than expected.

In the case of cystic fibrosis, it was originally thought that the disease involved a single mutation on a single gene. Today, we know there are multiple mutations on multiple genes that relate to the variation and the severity of the disease. It has been estimated that 99% of the DNA is unchanged from individuals and that we only know the functions for about 5% of the DNA found in our bodies. In the case of mRNA, the mRNA may not be translated into a protein, so even if a mutation is accurately found it may not always produce the protein or the enzyme. More knowledge about the mechanisms, kinetics and control of the translation processes from mRNA to protein synthesis will be necessary even if the laboratory analysis is deemed accurate.

One cannot overemphasize the major ramifications that will occur to the medical profession, to the clinical laboratory, to the law profession, to the governmental finances of health care, to the ethics of our society and to individuals when the genetic predisposition to disease tests become accurate and widely used. There will be major perturbations throughout our society when these laboratory tests become available. Already, in the case of cystic fibrosis, in multiple sclerosis, and in other disease conditions, new laboratory tests and diagnostic imaging instrumentation, such as magnetic resonance imaging, are redefining the clinical symptoms that previously were the basis of diagnosis of these diseases. These new tests may cause greater uncertainty for physicians as they manage their patients with complex diseases.

We can conclude that the amount of progress in the field of clinical chemistry in the past 50 years is remarkable. We know that the rate of knowledge and capabilities of instruments in this field continues to expand at a logarithmic rate. We can expect many exciting contributions from the application of analytical chemistry to clinical chemistry in the health care field. We all will benefit as a result.

ABBREVIATIONS AND ACRONYMS

AACC	American Association of Clinical Chemistry
AAS	Atomic Absorption Spectrometry
CE	Capillary Electrophoresis
GC/MS	Gas Chromatography/Mass Spectrometry
GLC	Gas–Liquid Chromatography
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
IR	Infrared
MALDI	Matrix-assisted Laser Desorption/ Ionization
MS	Mass Spectrometry
UV/VIS	Ultraviolet/Visible

Atomic Spectrometry in Clinical Chemistry

Patrick J. Parsons
New York State Department of Health, Albany and The University at Albany, Albany, USA

1 Introduction **2**
2 Historical Background **6**
3 Specimen Collection and Preanalytical Issues **8**
3.1 Contamination Control 8
3.2 Standard Precautions 8
3.3 Specimen Selection 9
3.4 Tissues 9
4 Units for Element Concentrations in Clinical Samples **9**
5 Single Element Clinical Applications of Atomic Absorption Spectrometry **10**
5.1 Calcium in Serum and Urine 10
5.2 Magnesium in Serum 10
5.3 Zinc in Serum 10
5.4 Copper in Serum and Urine 11
5.5 Lead in Blood and Urine 11
5.6 Cadmium in Blood and Urine 12
5.7 Aluminum in Serum, Plasma and Dialysate 12
5.8 Selenium in Serum 13
5.9 Chromium in Urine 14
5.10 Arsenic in Urine 15
5.11 Mercury in Urine and Blood 15
6 Literature Review **16**
6.1 General Reviews of Trace Element Analysis of Biological Systems with Atomic Absorption Spectrometry Techniques 16
6.2 Clinical Applications of Inductively Coupled Plasma Atomic Emission Spectrometry 17
6.3 Clinical Applications of Inductively Coupled Plasma Mass Spectrometry 18
7 Clinical Laboratory Regulations **20**
8 Quality Assurance and Quality Control **22**
8.1 Reference Materials, Control Materials and Proficiency Testing 22
Acknowledgments **24**
Abbreviations and Acronyms **25**
Related Articles **25**
References **26**

Atomic absorption spectrometry (AAS) and atomic emission spectrometry (AES) are two closely related instrumental techniques that have many applications in many different fields. In clinical chemistry, these techniques are well established for the quantitative measurement of essential and nonessential elements found in body tissues and fluids. For most practical purposes, the elements that are routinely measured with modern commercial instrumentation include almost all of the metallic elements of the s, d and p blocks of the periodic table. The exceptions in the p block include the halogens, the noble gases and the elements C, H, N, O and S. Above atomic number 83 (Bi), only U is measured in absorption. With the emergence of commercial instrumentation for atomic mass spectrometry, e.g. inductively coupled plasma mass spectrometry (ICPMS) clinical chemists have another powerful tool with which to measure multielement composition of biological specimens at ultratrace concentrations, i.e. $<10\,\mu g\,L^{-1}$. With ICPMS, the halogens become detectable, along with B, C, N and Th. The coupling of ICPMS with separation technologies has opened new possibilities in clinical laboratory medicine in the field of speciation.

In some areas of clinical chemistry, older atomic spectroscopy techniques are still considered the reference method for selected clinical applications, e.g. the determination of serum K and serum Na by flame atomic emission spectrometry (FAES). In other areas, flame atomic absorption spectrometry (FAAS) remains the method of choice for serum Mg and serum Cu because of its sensitivity, rapid throughput and the relatively low cost of instrumentation. Electrothermal atomic absorption spectrometry (ETAAS), while more complex than flame atomization methods, is the method of choice for applications such as blood Pb, serum Al and serum Se, where physiological concentrations are too low to be measured in the flame without laborious preconcentration efforts. Robust procedures for quality assurance (QA) and quality control (QC) are now well established within the clinical laboratory and cover the entire spectrum of clinical laboratory activities. These include preanalytical issues, personnel training and education, and also the appropriate use of certified reference materials (CRMs) for method validation purposes. Participation in proficiency testing (PT) programs and external quality assessment schemes (EQAS) for trace elements in biological fluids have also contributed to improved clinical laboratory performance. Despite the advent of ever more sensitive techniques and robust procedures, the quality of results is still only as good as the specimen collected. Contamination can occur at the preanalytical stage of the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

analysis and also during the analysis. Selection of the appropriate specimen is critical to obtaining meaningful clinical information about nutritional trace element status or exposure to toxic elements.

1 INTRODUCTION

AAS and AES are two closely related instrumental techniques that are routinely used to determine many of the chemical elements in samples of both environmental and biological origin. Although it is possible to detect almost all naturally occurring elements in biological materials, only a handful of those that can be measured using atomic spectrometry are considered essential for normal healthy growth. Definitions of essentiality vary but the consensus among life scientists appears to be that an element is considered essential if an organism cannot complete its normal life cycle, or achieve normal healthy growth without it, or if the element is part of a molecule that is an essential constituent of normal healthy growth.[1] Thus, the ''essential'' elements that are of interest in clinical laboratory medicine, and that can be measured directly by atomic spectrometry, are shown in Figure 1 on a simplified version of the periodic table. Many elements are ubiquitous in the environment and are easily detected in human body tissues and fluids. Some are relatively benign, but others are quite toxic even at concentrations considered trace, i.e. below $10\,\mathrm{mg\,L^{-1}}$ in fluids or $<100\,\mathrm{mg\,kg^{-1}}$ in tissues. Those toxic, nonessential elements that are most frequently encountered in clinical situations are shown in Figure 1 against a shaded background. All elements, including

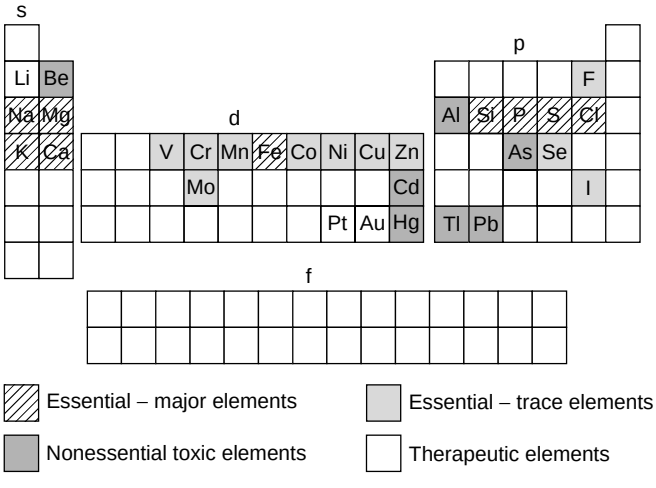

Figure 1 Periodic table showing selected elements of clinical significance. Essential elements are subdivided into major and trace ($0.01-100\,\mu\mathrm{g\,g^{-1}}$ or $10\,\mu\mathrm{g\,L^{-1}}-10\,\mathrm{mg\,L^{-1}}$), with the latter subdivided into those of particular toxic significance.

the essential elements, can exhibit toxic effects above a critical concentration threshold. However, a deficiency of essential elements, so that they are present at concentrations below those required for normal health growth, is also associated with adverse health effects. The toxicity of any element will depend on its concentration, duration and route of exposure, as well as the chemical form, i.e. chemical species, of the element. The extent to which some elements have been become clinically, occupationally, or therapeutically significant has been the subject of a review.[2] Some populations may be more at risk, or more sensitive than others, to toxic exposures. For example, patients on hemodialysis, or those on total parenteral nutrition, are at particular risk of exposure to elements such as Al and As. Young children are especially at risk to Pb exposure, because they are more likely to ingest Pb from normal hand-to-mouth activity, and they absorb Pb more efficiently in the gastrointestinal tract compared to adults. Assessment of trace element status is made by analysis of body fluids and tissues, and atomic spectrometry is often the technique of choice because of its sensitivity and selectivity for most of the naturally occurring elements in the periodic table.

In analytical AAS, quantitation is based on the amount of light absorbed at a specific wavelength (λ) by ground state atoms of the analyte in the gas phase. The wavelength selected is a resonance line that is generated by exciting atoms of the element being measured. For many elements, a hollow-cathode lamp (HCL) is an excellent line source. However, for several elements, e.g. Se, P and As, an electrodeless discharge lamp (EDL) is much better. The application of continuum-source AAS instrumentation for the analysis of clinical samples has also been described but the instrumentation is not commercially available and must be assembled in-house.[3]

In analytical AES, quantitation is based on the amount of light emitted by excited atoms as they return to the ground electronic state. A variety of thermal excitation sources are used to achieve atomic emission including flames and inductively coupled plasmas (ICPs). Figure 2 shows the various arrangements for analytical atomic spectrometry in a simplified schematic form. Differences between the technologies include cost and complexity, as well as analytical performance.

The distribution of atoms between the ground state and various excited states is described by Equation (1).

$$\frac{N_i}{N_0} = \frac{g_i}{g_0} e^{-\frac{\Delta E}{kT}} \qquad (1)$$

where N_i and N_0 are the number of atoms in the ith excited state and ground state respectively, g_i and g_0 are the statistical weights of the ith excited state

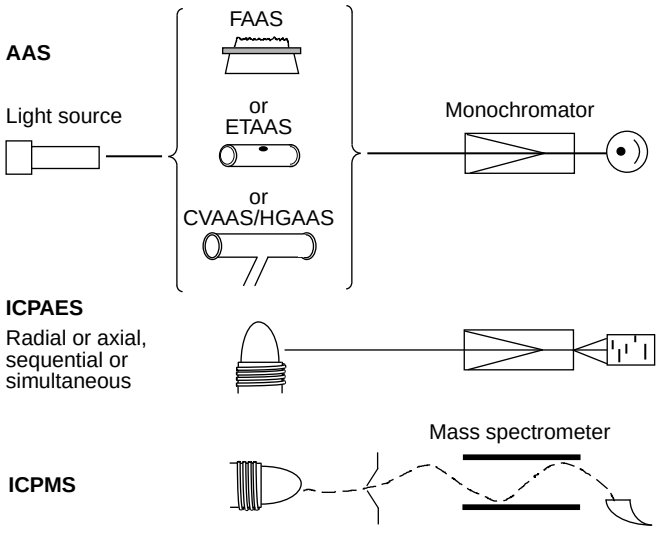

Figure 2 Instrumental configurations used in analytical atomic spectrometry. (Reproduced by permission from Perkin-Elmer LLC.[4])

and ground state, respectively, k is the Boltzmann constant ($1.38066 \times 10^{-23}\,\mathrm{J\,K^{-1}}$), and T is thermodynamic temperature in Kelvin. ΔE is the difference in energy between the ith and ground state, i.e. $E_i - E_0$, and is inversely related to wavelength as described in Equation (2).

$$\Delta E = \frac{hc}{\lambda} \tag{2}$$

where the terms h and c are the Planck constant ($6.62608 \times 10^{-34}\,\mathrm{J\,s}$) and the speed of light ($2.99792 \times 10^{8}\,\mathrm{m\,s^{-1}}$) respectively.

It is clear from the Boltzmann distribution in Equation (1) that, for both AAS and AES, the ratio of excited to unexcited atoms (N_i/N_0) of an element will depend on two key variables, ΔE and T. As ΔE increases, the ratio N_i/N_0 decreases. In practical terms, elements such as Zn and Mg, with resonance lines that are well into the ultraviolet (UV) region, i.e. large ΔE (<350 nm), will have N_i/N_0 ratios of the order of 10^{-7} to 10^{-11} when generated in an air–acetylene flame ($T = 2600\,\mathrm{K}$). In contrast, elements such as Na and K, with resonance lines in the visible region, i.e. small ΔE (400–750 nm), will have N_i/N_0 ratios of the orders 10^{-3}–10^{-4} in the same flame. As T increases, more atoms are excited and the ratio, N_i/N_0, also increases such that for Zn, in an argon ICP ($T = 8000\,\mathrm{K}$), it is of the order of 10^{-3}. Thus, the choice between AES and AAS will depend on the ratio N_i/N_0, which will depend on the element and the wavelength of its resonance line, and on the selection of the atom source, i.e. flame, furnace or ICP. Additionally, the concentration of the analyte will also dictate the selection of technique.

It is obvious that simple flame emission techniques, also called flame photometry, provide adequate sensitivity for Na, K and Li because a significant number of those atoms can be excited, even in an air–propane flame ($T = 2200\,\mathrm{K}$), but Zn and Mg would require an ICP to generate measurable signals in emission. Even in air–acetylene flames, Mg, Zn and Cu are much more sensitive in absorption because the vast majority of the analyte atoms are still in the ground state, where they are available for the absorption transition. Of course there are other factors that may influence the choice between AAS and AES. They include the limitation of AAS (in most commercial instruments) to measuring a single element at a time. While it is possible to determine more than one element simultaneously by AAS, the capability is limited to instruments equipped with multimonochromator detector systems or, more recently, to instruments equipped with an Echelle dispersion and diode array detector arrangement. Harnly[5] recently evaluated four commercially available multielement line source AAS instruments. Although AES provides the capability of simultaneous multielement analysis, spectral interferences continue to be troublesome, especially in complex biological matrices. Nonetheless, some laboratories do use inductively coupled plasma atomic emission spectrometry (ICPAES) for clinical applications, where information on several elements at microgram per liter concentrations is available within a very short time.

The application of AAS in clinical laboratory medicine is now so well established that, in many situations, it is the preferred analytical technique and, in some instances, it is regarded as the reference method. In recent years, however, ICPMS has emerged as a powerful alternative to AAS and AES. In ICPMS, an ICP is used as an ion source, and is interfaced to a mass spectrometer designed to separate and sort the ions according to their atomic mass (Figure 2). In most typical arrangements, an argon plasma is used with a quadrupole mass analyzer. The sensitivity of ICPMS is generally several orders of magnitude better than either AAS or AES, and this advantage can be exploited for clinical samples by using a higher dilution ratio. Thus, ICPMS can achieve similar method detection limits to AAS because the use of a higher dilution ratio minimizes interferences from the complex biological matrix (see Figure 3). Where several elements must be measured, ICPMS is much faster and its linear dynamic range is greater than either ETAAS or ICPAES.

Typical detection limits for ICPMS range from parts per billion ($\mu\mathrm{g\,L^{-1}}$) to parts per trillion ($\mathrm{ng\,L^{-1}}$), with precision on the order of 2–3%.[6] An important advantage of ICPMS is the opportunity to obtain information on individual isotopes. When used in isotope dilution (ID) analysis mode, the technique can, in experienced hands,

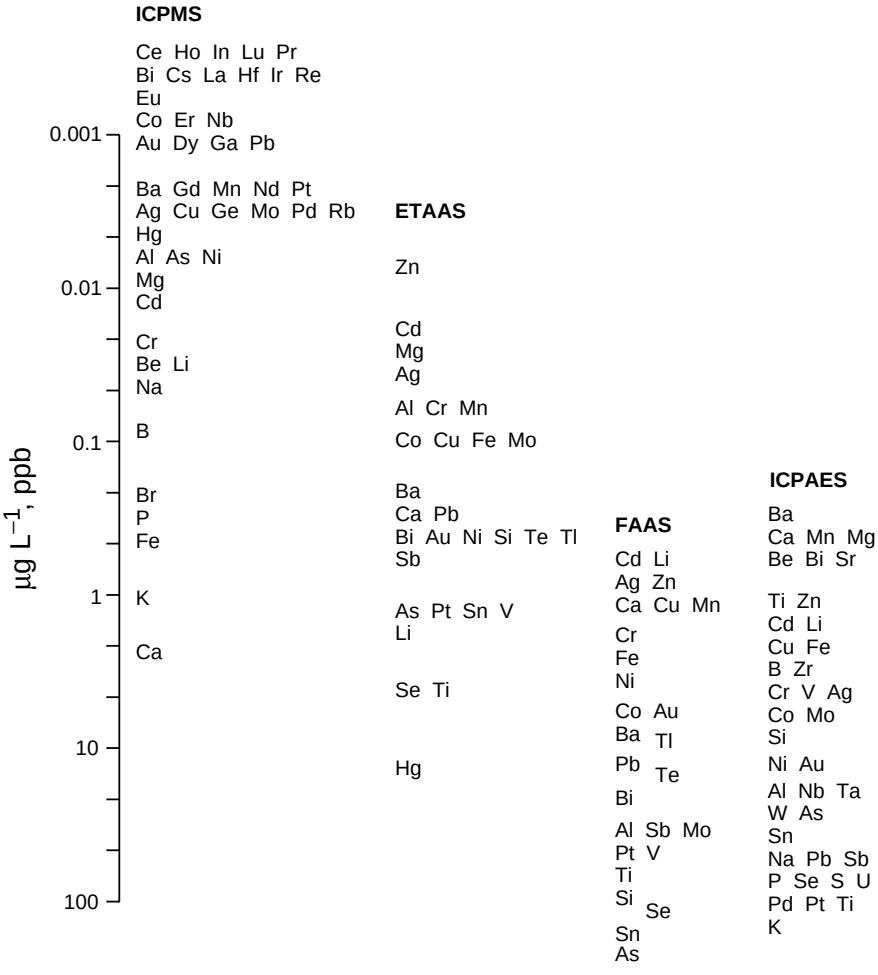

Figure 3 Detection limit comparison for ICPMS, ETAAS (20 µL sample size), FAAS and ICPAES. (Redrawn and adapted with permission.[6] Copyright by Advanstar Communications Inc. Advanstar Communications Inc. retains all rights to this reference.)

be considered a fundamental reference method because it is judged to be free of systematic error. However, the precision of isotope dilution inductively coupled plasma mass spectrometry (IDICPMS) is generally an order of magnitude worse than ID analysis performed with thermal ionization mass spectrometry (TIMS), which can reach levels of precision approaching 0.05% when magnetic sector mass spectrometry instrumentation is used. Still, the potential for ID analysis using the ICPMS is tremendous, and this has been exploited for clinical purposes. Depending on the element, access to isotopic ratios can also provide valuable information on source identification. The significance of ICPMS will continue to grow in specialized reference laboratories and, as the cost of modern instrumentation decreases, it is likely to become the preferred technique in the future.

Selecting an appropriate analytical technique for clinical purposes will depend on the number and nature of clinical specimens received for analysis, the element(s) requested for analysis and the skill of the laboratory personnel. The type of specimen selected for analysis depends on the element and its chemical species, e.g. inorganic or organometallic. Typically, whole blood, serum, plasma or urine are used to assess trace element status or monitor exposure to toxic elements. Specimens such as tissue, hair, saliva, nails, feces are less frequently encountered, and are complicated by the need for extensive sample preparation before instrumental analysis can proceed. The use of such specimens is reviewed in more detail below. For clinical purposes, specific applications may be categorized on the basis of analyte and concentration since this almost always governs the selection of analytical method. The major essential elements from group 1, Na and K, that, together with Li, are normally measured in emission, and the major essential elements from group 2, Ca and Mg, that are usually measured in absorption, are shown together in Table 1. The essential trace elements Zn, Cu, Se and Cr together with the toxic elements Pb, Cd, Hg and As are grouped together under Table 2.

Table 1 Determination of clinically significant elements in group 1 and group 2 by atomic spectrometry with flame atomization

Element	Atomic mass	Technique	Atomic line (nm)	Flame	Sample dilution	Reference range serum/plasma		Reference range urine		Refs.
						mg L^{-1}	mmol L^{-1}	mg L^{-1}	mmol L^{-1}	
Na	23.0	FAES	589.0	Air–C_3H_8	1 : 100	3130–3360[a]	136–146[a]	920–5060[a]	40–220[a]	7, 8
K	39.1	FAES	766.5	Air–C_3H_8	1 : 100	137–200[a,b]	3.5–5.1[a,b]	98–4900[a,b]	2.5–125[a,b]	8, 9
Li	6.94	FAAS	670.8	Air–C_2H_2	1 : 25	5.5–8.3[c]	0.8–1.2[c]	–	–	10
Ca	40.1	FAAS	422.7	Air–C_2H_2	1 : 50	84–100	2.1–2.5	<275	<6.87	11, 12
Mg	24.3	FAAS	285.2	Air–C_2H_2	1 : 50	17.0–26.7	0.6–1.1	24–255	1–11	13

[a] Taken from National Committee for Clinical Laboratory Standards (NCCLS) Document C29-A, March 1995.[14]
[b] Plasma K values may be 0.1–0.2 mmol L^{-1} lower than those in serum.[183]
[c] Therapeutic range following treatment with Li_2CO_3.

Table 2 Essential and toxic trace elements of clinical significance measured by AAS

Element	Atomic mass	Instrumental technique	Atomic line (nm)	Matrix	Sample dilution	Reference range[a]		Detection limit (μg L^{-1})	Ref.
						μg L^{-1}	μmol L^{-1}		
Al	27.0	ETAAS	309.3	Serum	1 + 4	<10	<0.370	1	15
As	74.9	FI/HGAAS	193.7	Urine	1 + 49	<120 μg dL^{-1}	<1.60 μmol dL^{-1}	0.07–0.69	16
Cd	112.4	ETAAS	228.8	Blood	1 + 2	<10	<0.089	0.146	17
Cr	52.0	ETAAS	357.9	Urine	none	<0.70	0.0135	0.5	18
Cu	63.5	FAAS	324.7	Serum	1 + 9	640–1600[b]	10.1–25.2[b]	1.27	19
Pb	207.2	ETAAS	283.3	Blood	1 + 9	<10 μg dL^{-1}	<0.48	1 μg dL^{-1}	20
Hg	200.6	CVAAS	253.6	Urine	1 + 1	<20 μg dL^{-1}	<0.100 μmol dL^{-1}	0.010	21
Se	80.0	ETAAS	196.0	Serum	1 + 9	109–181[c]	1.38–2.29[c]	27 pg	22
Zn	65.4	FAAS	213.9	Serum	1 + 9	700–1200	10.7–18.4	0.78	19

[a] Taken from NCCLS Document C38-A, September 1997.[23] Note that age, gender, estrogenic and/or dietary influences need to be considered for Cu and Se concentration ranges, and gestation-specific serum Zn ranges are required for pregnant women.
[b] For the age range 14–19 years (M/F).
[c] Adults.
CVAAS, cold vapor atomic absorption spectrometry; FI, flow injection; HGAAS, hydride generation atomic absorption spectrometry.

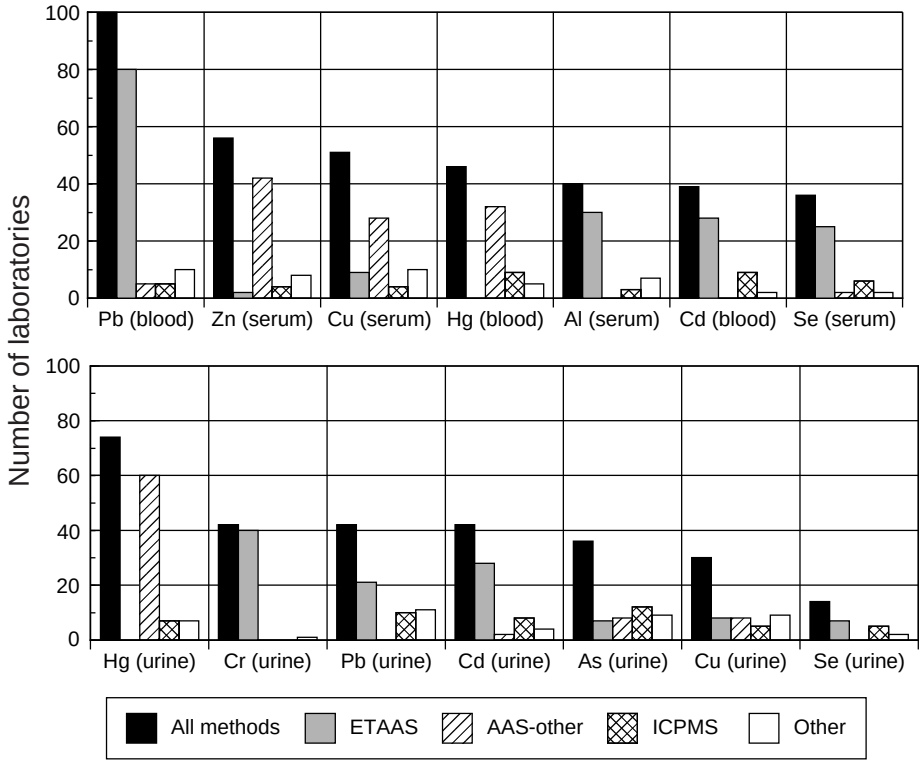

Figure 4 Number of laboratory participants in the Quebec Trace Elements Interlaboratory Comparison Program by element, matrix and analytical technique. (Data are taken from the first run of 1999.[24])

In this document, there is some background discussion of the group 1 elements Na and K from a historical perspective, but more detailed discussion is limited to elements that are considered clinically significant, and that are measured by AAS. Figure 4 shows the number of laboratories participating in the Interlaboratory Comparison Program operated by Le Centre de toxicologie du Québec during early 1999. The Québec program is an international external quality assessment scheme for selected trace elements in clinical matrices.[25] The number of participants varies according to the element and the matrix (blood, serum or urine) such that it is clear from Figure 4 that some analyses, e.g. blood lead, attract far more participants than do others, e.g. serum selenium. Using these data, it is possible to identify a panel of trace elements that are considered more clinically significant than others. Therefore, this document is primarily limited to the ''major'' elements of clinical and environmental significance (Ca, Mg, Zn, Cu, Pb, Cd, Al, Se, Cr, As and Hg) that are routinely measured by atomic spectrometry.

2 HISTORICAL BACKGROUND

Historically, FAES methods using air–propane were among the first to be used for routine clinical measurements, and the technique was referred to as 'flame photometry'. During the 1960s, flame photometry was used successfully for the determination of Na and K in serum. Today, Na and K are routinely measured in serum by automated analyzers based on ion-selective electrode (ISE) potentiometry. Nonetheless, FAES remains the reference method for Na[7] and K,[9] as described by the National Reference System for the Clinical Laboratory (NRSCL),[26,27] and is also used to standardize direct ISE systems for Na and K.[14] Apart from Na and K, Li has also been routinely measured with success by flame photometry, for therapeutic monitoring of the antipsychotic drug, $LiCO_3$. A reference method for serum Li is also available based on FAES.[10]

Development of clinical applications for the group 2 essential elements, Ca and Mg, was spurred by the advent of commercial FAAS instrumentation during the 1960s. The analytical conditions under which the major group 1 and group 2 elements are determined using flame atomization techniques are summarized in Table 1. Although the current trend in clinical laboratory medicine favors measurement of ionized Ca, as a better indicator of physiological status, measurements of total Ca are still regarded as clinically useful. Indeed, the application of FAAS for measuring total Ca in serum is considered a referee method.[11] Magnesium is routinely

measured in serum and urine by FAAS. The physiological concentration of Mg is high enough to permit direct determination following a sample dilution up to 50-fold. FAAS is also the preferred method for the determination of Zn and Cu in clinical specimens. For serum Cu and serum Zn, sample dilutions of $1+3$ to $1+9$ have been used successfully with viscosity-matched calibration standards. The application of FAAS to the determination of Ca, Mg, Zn, and Cu in serum is discussed in more detail below.

One of the most significant developments in clinical applications of FAAS was the introduction of a microsampling accessory that improved analytical sensitivity for some volatile elements by 100-fold. The approach pioneered by Delves for the determination of blood lead was particularly successful, and is known as the Delves' cup microsampling flame atomic absorption spectrometry (DCAAS) method.[28] Delves' approach is to deposit a relatively small amount of whole blood ($<100\,\mu L$) into a nickel cup, dry it and then insert it into an air–acetylene flame. A hollow quartz tube is mounted directly above the flame and along the optical axis such that the source beam passes directly through the sample cell. A small hole at the center of the quartz tube permits analyte atoms and blood matrix components, liberated in the air–acetylene flame, to move directly into the heated quartz tube, where they are momentarily trapped, thus increasing their residence time in the optical path. The resultant transient atomic absorption signal is separated in time from nonspecific absorption due to the blood matrix. Further improvements of the Delves' cup method have included integrating the absorbance signals, pretreating new nickel cups to improve precision and interfacing the instrumentation to a computer to permit semi-automated analyses.[29,30] A major advantage of the Delves' cup method is that only a very small blood sample volume is required for analysis. This made mass screening of pediatric populations possible by collecting capillary blood specimens using skin puncture techniques. Although DCAAS was the method of choice throughout the 1970s and 1980s, it has been largely superseded by ETAAS, which is more easily automated and which, at least with modern instruments, requires less skill for routine operation.

ETAAS technology has evolved over the last two decades into a sensitive, robust and reliable tool for trace element measurements. In ETAAS, the concept of characteristic mass (m_o) is used to define sensitivity as the mass of the analyte that produces an integrated absorbance equal to 0.0044 s. The term characteristic mass is normally quoted in picograms and is analogous to the concept of characteristic concentration (C_0) that is used in FAAS to define sensitivity as the concentration of an analyte producing an absorbance of 0.0044 (1% absorption). Modern ETAAS technology is designed around the Stabilized Temperature Platform Furnace (STPF), a concept developed by Slavin and others since the 1980s.[31–33] The key aspects of the STPF approach are summarized in Table 3. With few exceptions, they can be applied successfully to almost all matrices and to all elements that are measured by AAS, with exception of Hg, which is measured by CVAAS. The L'vov platform is a key aspect for successful analyses because it significantly delays analyte atomization until the furnace has more fully achieved thermal equilibrium, thereby minimizing chemical interferences that can plague wall atomization methods. The advent of fast digital electronics provides better data acquisition systems that more faithfully record the transient absorption signals that typify ETAAS. While some still debate the use of peak absorbance measurements, most accept that integrated absorbance (A_i) is much less likely to be influenced by chemical or matrix interferences, and because precision is better with A_i, detection limits are improved.

Table 3 The STPF

1.	The L'vov Platform
2.	Fast digital electronics, $<10-20$ ms per measurement
3.	Integrated absorbance (A_i) signals, not peak absorbance signals
4.	Modifiers to stabilize the analyte during the pyrolysis step
5.	High quality pyrolytically coated graphite tubes
6.	Fast heating of the furnace tube, $>1500\,^{\circ}C\,s^{-1}$
7.	Gas-stop during atomization to increase residence time
8.	Use of high-purity argon as the support gas, not nitrogen
9.	Zeeman-effect background correction

In clinical applications, the sample matrix can be quite troublesome, so methods must be carefully optimized to avoid problems. The use of modifiers, i.e. chemical compounds or other elements, that stabilize the analyte during either pyrolysis or atomization, or both, and can free the analyte from the matrix, thereby increasing production of ground state atoms, is considered mandatory for most clinical applications. High-quality pyrolytically coated graphite tubes are also considered an important requirement for successful analyses. Rapid heating of the furnace, and use of a gas stop feature that prolongs the residence time of atoms in the optical path, improve the sensitivity. Background correction is also mandatory and although Zeeman-effect systems are preferred for their simplicity and ruggedness, other systems also work well except in a few selected cases.

While ETAAS remains the technique of preference in most routine clinical laboratories, AES has re-emerged as a viable option in situations where rapid measurement of several elements is required, such as screening urine for toxic elements. Modern AES instrumentation typically

employs an ICP excitation source, which is viewed either radially or axially with respect to the optical bench (Figure 2). Advantages of ICPAES over FAAS include the capability for simultaneous multielement analysis and an extended linear range that obviates the need for re-dilution of samples that are above the calibration range. One of the limitations of ICP techniques is the difficulty associated with using nebulizers for introducing samples of high viscosity and high dissolved-solids content. This requires carefully matching the samples with the calibration standards and, for samples such as urine, a cross-flow nebulizer is used to accommodate the relatively high dissolved-solids content.[34]

In clinical applications, biological matrixes can be especially troublesome because of the increased opportunity for spectral interferences, a problem that plagues emission techniques. In some respects, ICPAES now competes with ICPMS as a multielement technique, but the latter has very much better detection limits. There are, however, some important limitations with ICPMS, especially instrumentation equipped with quadrupole mass analyzers that are unable to resolve isobaric or polyatomic interferences. Selected examples of isobaric or polyatomic species are reviewed below under clinical applications of ICPMS. For clinical laboratories that require the capability to measure several trace elements simultaneously, ICPMS is rapidly becoming the preferred technique. As more ICPMS instruments find their way into clinical laboratories, new applications are being developed. Significant among these is the coupling of separation techniques with ICPMS to permit trace element speciation. While speciation is still in its infancy, for some applications, determining inorganic and organic species in urine and blood is now well established.[35,36] The development and application of trace element speciation techniques in clinical laboratory medicine is reviewed in more detail below.

3 SPECIMEN COLLECTION AND PREANALYTICAL ISSUES

3.1 Contamination Control

In analytical chemistry, it is often stated that the quality of an analytical result is only as good as the sample collected. In trace element analysis, the concern with contamination errors is always present, although the depth of concern depends on the element measured. Contamination control begins with specimen collection. Selecting an appropriate specimen collection device is critical to success. Simply using standard blood collection tubes may appear convenient but contamination from uncertified tubes will render the results useless. Several manufacturers market blood collection tubes specifically for trace element analysis, and some certify tubes for specific element(s). Contamination may originate from the tube stoppers, e.g. Zn, the anticoagulant or even from the container material, e.g. Al in glass. Thus it is good practice to check that collection devices are free of significant contamination for the element(s) to be measured. Standard test protocols now exist for checking blood collection tubes for Pb contamination,[37] and they are easily applied to other elements. Delves[38] recommends using simple preparation procedures with minimum handling of body tissues, tissue fluids, excreta and diets for those trace element analyses for which contamination is the major problem.

Apart from contamination, other preanalytical variables may also affect the results of trace element analysis. Several authoritative reviews[23,38,39] of preanalytical variables in trace element determinations are available that provide detailed information on specimen collection procedures and transportation to the clinical laboratory for analysis. Guidelines for collecting clinical samples for trace element analysis are provided by the International Union of Pure and Applied Chemistry (IUPAC) Commission of Toxicology[39] and by the NCCLS.[23]

3.2 Standard Precautions

When clinical specimens of human origin are submitted for analysis, the analyst is faced with a very different situation compared to environmental samples. Because it is often impossible for the analyst to know which are infectious, all clinical specimens should be treated as if potentially infectious. In the US, the Centers for Disease Control and Prevention (CDC) provides ''Standard Precautions'' guidelines for preventing the transmission of infectious disease for health care workers.[40,41] Other organizations have also issued specific guidelines for laboratory workers exposed to infectious disease transmitted by blood, body fluids and tissues.[42] These guidelines cover use of protective gloves and biohazard hoods, decontamination and clean-up procedures. Recently, Moreton and Delves[43] discussed some specific biohazard concerns that arise when biological fluids are nebulized into ICPMS instrumentation. The generation of aerosols containing potentially infectious material is a particular concern with all nebulization techniques, including FAAS, ICPAES and ICPMS. These authors recommend adding a commercial virucidal disinfectant (Virkron) to all high-risk biological samples, and then subtracting the blank value for the 13 elements found in this product. They report the successful determination of Hg in urine by ICPMS, and Cu and Zn in serum by FAAS using this approach.

3.3 Specimen Selection

The specific body fluid, or tissue, that is sampled for analysis depends on the analyte that is to be measured, and the purpose for which the test result will be used. Typically, serum, plasma, whole blood or urine is collected for trace element analysis by atomic spectrometry. Whole blood specimens are preferred for those elements, e.g. Pb and Cd, that are concentrated within the erythrocytes or red blood cells (RBCs) and that, following exposure, remain within the blood compartment for some time, and are said to have biological residence times of the order of at least several weeks. Whole blood must be preserved with an anticoagulant to prevent clot formation and, of the various anticoagulants available, ethylenediamine tetraacetic acid (EDTA) salts or heparin are used. While heparinized blood can form fibrin clots more readily, trace element-free EDTA salts are difficult to locate.

The terms serum and plasma are often used interchangeably, and are sometimes mistakenly assumed to be the same specimen. While either serum or plasma can be used to assess trace elements status, they are biochemically different specimens. The differences between them can be analytically important (see aluminum below). When a serum or plasma specimen is required, it is important to separate it from the clot or RBCs as quickly as possible to prevent leakage of the RBC-bound element. Care must be taken to assure that exogenous contamination does not occur during the separation process. Several up-to-date documents are available that provide recommended sample/specimen collection guidelines for trace element analysis.[23,38,39] The IUPAC Commission on Toxicology's sample collection guidelines for trace elements in blood and urine[39] provide protocols for collecting whole blood, serum or plasma, packed cells or erythrocytes, and urine for the determination of Al, As, Cd, Cr, Co, Cu, Pb, Li, Mn, Hg, Ni, Se and Zn. Recommended guidelines for the control of preanalytical variation in trace element determinations are also available from the NCCLS in the United States. The NCCLS guidelines[23] include sample collection protocols for trace element analysis, which is defined as a body content of $0.01-100\,\mu g\,g^{-1}$, or $10\,\mu g\,L^{-1}-10\,mg\,L^{-1}$, and ultra trace element analysis, which is defined as a body content $<0.01\,\mu g\,g^{-1}$, or $<10\,\mu g\,L^{-1}$.

Urine is widely regarded as an inferior indicator of trace element status except for a few elements such as Hg, As and Cr, which are so rapidly cleared from the blood compartment into the urine that blood concentrations do not adequately reflect recent exposure. While it is clearly more convenient and less invasive to collect spot urine specimens, urine analyte concentrations do not always correlate as well as serum or whole blood measurements, even when urine is corrected with a creatinine measurement, or when urine is collected as a timed specimen, e.g. 24 h.

In some situations, capillary blood collected by skin puncture, i.e. via fingerstick or heel stick, is desirable for screening purposes especially in pediatric populations, where, in the absence of a trained phlebotomist, venipuncture is difficult to perform. The choice of capillary blood will affect the selection of analytical technique because of the relatively small blood volumes (150–300 µL) that are typical of such specimens. Contamination errors with capillary specimens are always a concern, but for childhood lead screening purposes, the frequency of contamination errors can be held to around 5% if a rigorous cleaning protocol is adopted.[44]

3.4 Tissues

It is uncommon for routine clinical laboratories to receive tissues for trace element analysis. In most situations, the specimen would be referred to a specialized reference laboratory for analysis. Tissue biopsies are difficult to perform without contaminating the specimen with the elements typically found in stainless steel. Liver biopsies are analyzed for copper in diagnosing Wilson's disease (see section 5.4). Use of hair and nails are generally considered unreliable indicators of trace element status[23,45] except, perhaps, in a few cases such as Hg and As poisoning. In the case of As, hair and nails have been used for forensic purposes to investigate arsenic intoxication. One of the major limitations of hair and nails is the extent to which the specimen is compromised by contamination.[46] Although elaborate procedures have been devised for sampling these specimens, it is difficult to distinguish between exogenous environmental contamination and removal of endogenously bound elements that may be leached during a preanalytical washing procedure. The use of saliva and sweat specimens for assessing trace elements status is even more dubious and little information exists to validate these specimens.

4 UNITS FOR ELEMENT CONCENTRATIONS IN CLINICAL SAMPLES

A variety of units is used to report element concentrations in clinical samples. While much of the clinical laboratory community throughout the world uses substance ($nmol\,L^{-1}-mmol\,L^{-1}$) concentrations, there is no such consensus within the US where a myriad of mass ($\mu g\,L^{-1}$, $mg\,L^{-1}$, $\mu g\,dL^{-1}$, $ng\,mL^{-1}$, ppm, ppb, etc.) concentrations are more frequently used. Mass concentration units are also frequently preferred within the analytical laboratory.

This situation invites confusion and misunderstandings. The International Federation of Clinical Chemistry (IFCC) and the IUPAC recommend the exclusive use of the liter as the unit of volume when reporting laboratory results using the Systeme International d'Unites, SI. Furthermore, the IUPAC/IFCC recommends units of substance concentration, and have published specific guidelines for trace elements in clinical samples[47,48] that also contain conversion factors for mass-based units.

In some areas of clinical chemistry, the use of mass concentrations is particularly pervasive. For example in the US, published values for blood lead are universally reported as mass concentration ($\mu g\,dL^{-1}$ or $\mu g\,100\,mL^{-1}$) rather than as substance concentration ($mol\,L^{-1}$). Urine lead measurements are also reported as mass concentration, but with different units, ($\mu g\,L^{-1}$). In this article, units for elemental concentrations are given in mass concentration. The IUPAC/IFCC-recommended units of volume (L) and substance (elemental) concentration ($mol\,L^{-1}$) are included in parentheses, where appropriate.

5 SINGLE ELEMENT CLINICAL APPLICATIONS OF ATOMIC ABSORPTION SPECTROMETRY

5.1 Calcium in Serum and Urine

Calcium is a major essential element from group 2 of the periodic table (alkaline earth elements) and is found at high concentrations in every compartment of the body. FAAS is considered a referee method for determination of serum Ca because of its excellent sensitivity, freedom from interferences, and its traceability to reference standards.[11] The FAAS technique measures ''total'' Ca in serum regardless of the element's chemical binding. However, most routine clinical laboratories measure total serum Ca by photometric methods, and some use potentiometric ISE technologies. Some ISE technologies are designed to measure only the ''free'' or unbound Ca fraction in serum or plasma. The unbound Ca is also referred to by some as ''ionized'' Ca, but all serum calcium exists as the Ca^{2+} ion. Approximately 46% of Ca in serum is unbound, the remainder is bound to plasma proteins and complexed with other ligands. Some[49] have argued that measurements of ''ionized'' Ca are physiologically more meaningful than total Ca, but a thorough discussion of this specific issue is beyond the scope of this text. The preferred specimen for total Ca measurement is either serum or heparinized plasma. Use of EDTA or oxalate anticoagulants is inappropriate since they bind Ca. In some situations, when urinary Ca is measured, hydrochloric acid is added to the collection container to prevent precipitation.

Calcium can be determined in an air–acetylene flame using any standard AAS instrumentation. Recommended analytical details[12] are given in Table 1, and are based on the original National Bureau of Standards method published by Cali et al.[11] Serum or plasma is separated from RBCs and diluted by a factor of 50 with diluent. Viscosity differences between serum or plasma and aqueous standard solutions may lead to a matrix effect with some nebulizer arrangements in FAAS. If this is a problem even with a dilution factor of 50, then matrix-matching the calibration standards is required. Typically, glycerol is added to the calibration standards to match viscosity, and a releasing agent is added to standards, controls and samples to prevent chemical interferences from phosphate, a natural component of biological materials. Phosphate binds Ca, forming a stable compound that does not dissociate in the air–acetylene flame, and causes a suppression in sensitivity. Lanthanum chloride is often then used as a releasing agent for Ca, although strontium chloride may also be used.

5.2 Magnesium in Serum

Magnesium is another group 2 essential element that is found throughout the body at relatively high concentrations. Measurement of serum magnesium is important clinically for assessing magnesium deficiency (hypomagnesemia), which often results in neuromuscular impairment, e.g. hyperirritability, tetany and convulsions.[13] In rare instances, increased serum Mg concentrations (hypermagnesemia) have been observed with dehydration, severe diabetic acidosis and Addison's disease. The reference method for serum Mg is FAAS because of its selectivity, freedom from interferences and sensitivity. Serum specimens are separated from RBCs as quickly as possible since the Mg content of erythrocytes is greater than that of serum. In some situations, urine specimens are also analyzed to monitor Mg clearance. The FAAS method for serum Mg is straightforward and well established.[13] Serum is diluted by a factor of 50 with a solution containing NaCl (2.8 mM), KCl (0.1 mM), to reduce ionization interferences, and $LaCl_2$ releasing agent to reduce interferences from phosphate. Viscosity effects may be corrected by adding glycerol to the calibration standards. Diluted samples are aspirated directly into an air–acetylene flame, and Mg is determined at the 285.2-nm line with a Mg HCL.

5.3 Zinc in Serum

Zinc is well established as an essential trace element for normal healthy growth. It is required by numerous metalloenzymes including carbonic anhydrase, alkaline phosphatase and carboxypeptidase. Zinc deficiency has been recognized in humans since the 1960s, and is

associated with growth retardation, impaired wound healing and dermatitis. Serum or plasma Zn measurements are considered the best single choice of test to assess deficiency. Serum must be collected in tubes certified for trace element analyses since Zn contamination is particularly troublesome, with collection tube stoppers being a known source of contamination.

The preferred method for serum zinc is FAAS, which has ample sensitivity and good precision (detection limit $\sim 1\,\mu g\,L^{-1}$), and rapid throughput. Some[50] have used ETAAS to measure Zn in plasma and peripheral blood cells. The application of FI or pulsed-nebulization techniques coupled with FAAS is particularly attractive since sample volumes as low as 10–100 µL can be accommodated.[51,52] Generally, samples and standards are matrix-matched for acid content and for viscosity. Serum is diluted by a factor of 5, and Zn is determined in an air–acetylene flame at the 213.9-nm line. Pizent and Telisman[19] describe a FAAS method for serum Zn in which samples are diluted 1 : 9 with 6% aqueous butan-1-ol. Calibration is linear up to $15\,mg\,L^{-1}$ and a detection limit of $0.78\,\mu g\,L^{-1}$ is reported.

5.4 Copper in Serum and Urine

Copper is an essential element and, as a d-block element (Figure 1), its biochemistry is dominated by Cu^{2+} (d^{10}) complexes with various proteins such as ceruloplasmin and superoxide dismutase. Approximately 90% of serum Cu is bound to ceruloplasmin, with the remainder bound to albumin or amino acids, or other copper enzymes. Serum Cu is normally used to diagnose cases of copper deficiency due to malnutrition or malabsorption, or due to Menkes' disease, a genetic disorder that is characterized by the appearance of ''steely hair'' and is often referred to as ''kinky hair syndrome''.

Serum Cu is also used to diagnose Cu toxicity due to occupational or environmental exposures, or due to Wilson's disease, a genetic defect in Cu metabolism that is characterized by Cu accumulation, i.e. Cu overload, in the liver, brain, kidneys and cornea. In cases of Wilson's disease, liver biopsies are analyzed for Cu content. When collecting liver biopsies for Cu determination, care should be taken to prevent contamination from metal instruments. Penicillamine, a chelation drug, is used to promote excretion of excess Cu into the urine, which is analyzed for Cu content. For serum (or plasma) Cu measurements, the blood collection tubes should be certified for trace element measurements.

The analytical method most commonly used for serum Cu measurements is FAAS, although ETAAS and ICPMS are also used. Most manufacturers of FAAS instrumentation provide standard protocols for serum Cu measurement based on matrix-matched standards, and determination in an air–acetylene flame at the 324.8-nm line. FI techniques coupled to FAAS have also been used to measure serum Cu.[53] Pizent and Telisman[19] describe a FAAS procedure for serum Cu in which samples are diluted 1 : 9 with 6% aqueous butan-1-ol. Calibration is linear up to $15\,mg\,L^{-1}$ and a detection limit of $1.27\,\mu g\,L^{-1}$ is reported. A detailed standard operating procedure for serum Cu by FAAS is also described by Alcock.[54] This method may also be used to determine Cu in urine at the relatively high concentrations found in copper overload cases, i.e. Wilson's disease, and following a penicillamine challenge. Where FAAS detection limits are a concern, such as in estimating urine Cu at normal concentrations, either ETAAS or ICPMS is preferred. Alcock[54] also described an ETAAS procedure for serum and urine Cu that is based on the STPF approach.

Other workers have also been successful in using ETAAS for Cu determinations in clinical samples. Lapointe and LeBlanc[55] described an ETAAS procedure for the determination of Cu in serum or plasma based on the STPF concept with Zeeman-effect background correction. Serum or plasma samples are diluted 1 : 50 with 0.4% Triton X-100, 1% $NH_4H_2PO_4$ and 0.06% $Mg(NO_3)_2$. Calibration with aqueous standards was linear up to $2.5\,mg\,L^{-1}$ and a detection limit of $60\,\mu g\,L^{-1}$ was reported, with a characteristic mass of 7 pg. Method validation using proficiency materials from the Quebec Intercomparison program for trace elements is described. Schmitt[50] also used ETAAS to determine Cu (and Zn) in plasma and peripheral blood cells, and reported finding elevated platelet Cu levels in renal insufficient patients.

It is generally agreed that there are age- and sex-specific reference intervals for serum Cu (Table 2). Thus, it is advisable to consult an authoritative source for interpreting serum Cu levels.[23]

5.5 Lead in Blood and Urine

Lead is a nonessential element that is widely recognized as a serious environmental health problem following many centuries of industrial use. It is a particular health problem for very young children, who are exposed to environmental lead through normal hand-to-mouth activities, and who absorb lead much more efficiently than adults via their gastrointestinal tracts. Since more than 90% of Pb in the blood compartment is bound to the RBCs, whole blood Pb measurements are preferred for diagnosing Pb poisoning cases. Adults are more likely to be exposed either in occupations that involve Pb or Pb products, or in hobbies and recreational activities that involve Pb, e.g. stained-glass art work, illegal moonshine production, etc. Capillary blood obtained via skin puncture is acceptable for screening purposes, but only venous blood should be used for diagnostic and treatment purposes.

Either heparin or EDTA anticoagulant may be used to preserve whole blood specimens, although the latter is preferred because it permits longer storage time. In the US, urine Pb measurements are used for long-term monitoring of workers occupationally exposed to lead because collection is less invasive, and it does not require a phlebotomist. In European countries, however, long-term monitoring of lead workers is based on blood lead measurements. Urine Pb measurements are also used in the $CaNa_2EDTA$ challenge test to assess lead mobilization, but this test is considered less useful today, at least in children.[56] Special attention to contamination errors is important because industrial uses of lead have resulted in significant environmental contamination. Serum or plasma Pb measurements have been attempted in epidemiological studies, but normal concentrations are generally below the method detection limit by ETAAS, so ICPMS with electrothermal vaporization (ETV) is appropriate.[57] Determination of Pb in bone is important in toxicology research studies, with most bone specimens obtained post-mortem.[58] Measurement of Pb in other samples, e.g. nails, is not considered reliable for routine clinical purposes because of contamination concerns and the lack of robust validation resources.[46]

Today, the preferred method for blood lead is ETAAS.[20,60,61] Modern methods, based on the STPF approach, have culminated in a consensus method[37] for blood lead that has been validated with various instruments with different background correction systems,[62] and is recommended for use in clinical laboratories. The method involves a simple $1 + 9$ dilution of whole blood, with a phosphate modifier, Triton X-100 and nitric acid, followed by platform atomization and integrated absorbance measurements at the 283.3-nm line.[20] Urine lead measurements are also possible using the same dilution and method, although the method detection limit does not support urine lead measurements at normal concentrations.[63]

The reference range for blood lead is population-dependent. In 1991,[64,65] the US geometric mean blood Pb level for the general population was reported to have fallen to $2.8\,\mu g\,dL^{-1}$ ($0.13\,mol\,L^{-1}$), and to $3.6\,\mu g\,dL^{-1}$ ($0.17\,mol\,L^{-1}$) in children aged 1–2 years. In 1994,[66] the geometric mean blood Pb fell even further to $2.3\,\mu g\,dL^{-1}$ ($0.11\,mol\,L^{-1}$) and to $3.1\,\mu g\,dL^{-1}$ ($0.15\,mol\,L^{-1}$) in children. Public health agencies consider a blood lead level above $10\,\mu g\,dL^{-1}$ ($0.48\,mol\,L^{-1}$) in children harmful, and above $40\,\mu g\,dL^{-1}$ ($1.93\,mol\,L^{-1}$), treatment with chelation drugs is indicated. Interestingly, occupational exposures up to $40\,\mu g\,dL^{-1}$ in adults are tolerated in the US in spite of evidence of biochemical effects on the hematopoietic system.

The reference range for urine lead in the US population[67] indicates the geometric mean is $2.08\,\mu g\,L^{-1}$ ($0.0.1\,mol\,L^{-1}$) with the 95th percentile at $6.07\,\mu g\,L^{-1}$ ($0.03\,mol\,L^{-1}$). This is considerably lower than mean urine Pb levels of $14.6–17\,\mu g\,L^{-1}$ ($0.07–0.08\,\mu mol\,L^{-1}$) reported for European populations.[68]

5.6 Cadmium in Blood and Urine

Cadmium is a nonessential toxic metal that has many industrial and commercial uses. It is taken up by plants, particularly vegetables grown in contaminated soils, resulting in exposures to human populations. Tobacco plants are particularly efficient accumulators of Cd, such that tobacco smokers are more exposed to Cd than are nonsmokers. Cadmium exposure is usually assessed by measuring its concentration in whole blood since it is largely taken up into RBCs. Urinary Cd is used for long-term monitoring of occupational exposure in adults. Contamination errors can occur through using supplies and reagents that have not been prechecked for Cd content. Colored plastic pipette tips are a known source of Cd contamination.

The method routinely selected for measuring Cd in blood or urine is ETAAS. Current experience indicates that ETAAS methods based on the STPF approach are successful, with Zeeman-effect background correction systems preferred. D'Haese et al.[69] report that it is possible to overcome the difficulties experienced with continuum-based correction systems by using a Mo-coated tube/platform. The excellent sensitivity of ETAAS ($m_o = 0.4–1.3\,pg$) provides detection limits on the order of $0.01–0.1\,\mu g\,L^{-1}$.[70] The sensitivity for Cd by AAS is so good that FAAS methods have been developed based on slotted tube atom trapping[71,72] and FI techniques.[73] An exhaustive review of AAS applications for determination of Cd in biological samples by AAS is available, and covers the literature up to 1993.[74] The consensus from the literature is that with modern ETAAS instrumentation equipped with Zeeman background connection, direct analysis is feasible using phosphate and Mg modifier, peak area measurements and platform atomization, i.e. the STPF approach. With older instrumentation, it may be necessary to perform additional sample pretreatment, such as protein precipitation and/or digestion with nitric acid.

Reference ranges for Cd in blood are different for nonsmokers and smokers such that for the latter, blood Cd is double that of the former. A blood Cd value in excess of $10\,\mu g\,L^{-1}$ ($0.089\,\mu mol\,L^{-1}$) indicates significant exposure.

5.7 Aluminum in Serum, Plasma and Dialysate

Aluminum is another nonessential metal that is ubiquitous in the environment, so it is hardly surprising

that it is found in appreciable amounts in almost every compartment in the human body. Its ubiquitous nature in the environment also presents a formidable challenge to analysts who must struggle with contamination control.

The significance of aluminum measurements in clinical situations became evident as a result of advances in the medical management of patients with end-stage renal disease through treatment with hemodialysis. The accumulation of Al in dialysis patients results in aluminum loading in bone and brain tissues, which leads to a fatal neurological syndrome called dialysis dementia. Historically, the most important source of aluminum in the development of dialysis dementia was tap water, treated with alum, and used for preparation of dialysis fluids without any further treatment. Aluminum used to treat hyperphosphatemia was a much smaller input although still important. Now that water supplies are treated to remove aluminum, the principal source of aluminum is from the administration of aluminum hydroxide, used to treat hyperphosphatemia in end-stage renal disease. Dialysis centers must be assured that the Al content of the water and dialysis fluid is within acceptable levels. For monitoring Al levels in dialysis patients, a serum (or plasma) specimen must be obtained for the analysis. In the US, urine Al is sometimes used to monitor exposed workers, but the clinical value of such measurements is debatable, and specimen contamination in occupational settings is much more serious.

In most situations, the preferred analytical method for routine serum Al determination is ETAAS, which has the required sensitivity and precision to meet the detection limits necessary to report concentrations above $50\,\mu g\,L^{-1}$ ($1.85\,\mu mol\,L^{-1}$). With modern instrumentation and the STPF approach, it is considered no longer necessary to remove plasma proteins by precipitation prior to analysis. Direct analysis, following simple dilution with a $Mg(NO_3)_2$ matrix modifier containing Triton X-100 plus nitric acid, is straightforward and does not require the method of standard additions. Aluminum can be determined either at the 309.3-nm line ($m_o = 11\,pg$) or at the 396.2-nm line ($m_o = 14\,pg$), although the latter has a greater linear dynamic range.[73] The use of $Ca(NO_3)_2$ modifier instead of $Mg(NO_3)_2$ is preferable when analyzing bone and tissue digests and can be used with serum and dialysate samples.[15] But addition of $Ca(NO_3)_2$ to plasma samples results in rapid protein precipitation, which makes the analysis troublesome. The general consensus is that it is preferable to use $Mg(NO_3)_2$ modifier for serum, plasma and dialysate specimens. Using nitrogen as the purge gas causes a loss of sensitivity for Al, so argon gas should be used.

The normal clinical range for serum Al is less than $10\,\mu g\,L^{-1}$ ($0.37\,\mu mol\,L^{-1}$), and concentrations up to $50\,\mu g\,L^{-1}$ ($1.85\,\mu mol\,L^{-1}$) are not considered toxic. Serum Al concentrations up to $100\,\mu g\,L^{-1}$ indicate an increased Al intake, and at concentrations above $100\,\mu g\,L^{-1}$ ($3.70\,\mu mol\,L^{-1}$) the potential for Al toxicity exists and medical investigation is warranted.[76]

5.8 Selenium in Serum

Selenium is an essential element that is a component of glutathione peroxidase, an enzyme that is involved in the antioxidant defense system. A group 16 (or group VI B) element, Se is classified as a nonmetal (Figure 1). Chronic Se deficiency is associated with cardiovascular disease in humans[77] and with Keshan disease, an endemic cardiomyopathy that affects women and children in certain areas of China.[13] Kashin-Beck disease is also found in China, and is linked to low Se status. Selenium toxicity is rare in humans but in animals (cattle and horses), it is characterized by a neurological dysfunction (''staggers''). Nutritional assessment of Se status is by measuring Se in serum, but toxicity is assessed via urinary Se.[77]

There are a variety of analytical techniques available for measurement of serum Se, including fluorimetry, AAS, ICPMS and neutron activation analysis (NAA). The latter technique is limited to facilities with access to neutron sources. Methods based on AAS can be categorized as either electrothermal atomization or HGAAS techniques. The latter, HGAAS, has been coupled to both ICPMS and AES instrumentation for determination of serum Se.[78]

Methods using ETAAS have been developed for a wide range of instrumentation. Some analysts favor using an EDL as the radiation source, with determination at the 196.0-nm line. However, while the EDL gives a more intense emission, these lamps are not convenient for use with many commercial instruments and HCLs are widely used without too great a problem. The optimum modifier is still a matter of debate among the experts with some recommending Ni,[79,80] others Pd,[81] sometimes with a reducing agent such as ascorbic acid[82] or as a Pd/Mg[22,83] mixed modifier, while others have used Cu and Fe.[84] Some experts[17] recommend air (oxygen) ashing in situ prior to atomization to reduce carbonaceous build up. Van Dael et al.[22] describe a standard additions method for serum Se using a transversely heated graphite furnace (THGA) with a longitudinal Zeeman background correction system. Serum ($100\,\mu L$) is diluted with $200\,\mu L$ 0.25% Triton X-100, $200\,\mu L$ water and $500\,\mu L$ Se standard, and $20\,\mu L$ of the diluted specimen is injected on to the platform, along with $10\,\mu L$ modifier ($15\,\mu g$ Pd + $10\,\mu g$ Mg). Calibration is linear up to $25\,\mu g\,L^{-1}$, and a detection limit of 27 pg is reported. A comprehensive critique of the various modifiers proposed for the serum Se assay covering over 200 papers up to the early 1990s

is very helpful.[74] A more recent review of modifiers and other parameters associated with measurement of selenium is also available.[85] The general picture is that Se is one of the more troublesome elements to determine, and analytical methods require optimization based on the STPF approach. Zeeman background correction is considered mandatory owing to a structured background around 196.0 nm that causes problems with continuum-based systems. The structured background is particularly troublesome with whole blood and urine, but less so with plasma/serum, owing to the presence of Fe and phosphate.

Methods based on HGAAS have a major advantage over ETAAS because they are much faster, and are reported to accommodate as many as 300 samples per hour. There are several published HGAAS methods for Se in serum[86] and urine[87] in the literature. The technique is based on the generation of H_2Se gas, which is easily decomposed into ground state atoms. First, all selenium in the serum must be reduced (or oxidized) to the Se(IV) oxidation state (selenite). This is generally accomplished using hydrochloric acid at 95–100 °C. The Se(IV) is reacted with $NaBH_4$ in solution to form gaseous H_2Se, which is then directed into the optical path of an AAS instrument equipped with a heated quartz tube. The quartz tube may be heated either electrically or by a flame. Upon entering the heated tube, H_2Se rapidly decomposes to form ground state atoms of Se and H_2 in the gas phase. The resultant transient signal is captured for quantitation. An IUPAC interlaboratory round-robin[88] conducted more than a decade ago for Se determination in serum and urine with acid digestion and HGAAS showed good interlaboratory agreement.

Van Dael et al.[22] compared their THGA method for serum Se (described above) with a FI/HGAAS method, based on work published by Weltz and Shubert-Jacobs[89] using direct calibration with peak height measurements. Their reported detection limit is 59 pg. Both methods gave acceptable values for CRMs. Weltz and Stauss[90] investigated interferences from hydride-forming elements on selenium in HGAAS with a heated quartz-tube atomizer. They compared two instrumental arrangements: a Perkin–Elmer model MHS-20 Hg/hydride batch system, and a Perkin–Elmer model FIAS-200 FI system. A deficiency of radicals was cited as the main reason for the low tolerance towards other hydride-forming elements in the batch system. However, there were no problems in the FI system. With conditions optimized, tolerance limits were one to two orders of magnitude higher than with the batch system. More detailed discussions of HGAAS methods for Se in biological samples are provided by Tsalev.[74]

Reference intervals for serum Se are age- and geography-dependent. Values for adults range from 109 to 181 $\mu g\,L^{-1}$ (1.38–2.29 $\mu mol\,L^{-1}$) for British Columbia[23] to 103.5±21.3 $\mu g\,L^{-1}$ (1.31±0.27 $\mu mol\,L^{-1}$) for the UK.[91]

5.9 Chromium in Urine

Chromium is a transition metal (d^4) that is now widely accepted as an essential trace element required for normal healthy growth. Its precise biochemical function remains a matter of debate, but in the +3 oxide state it is believed to be involved in the metabolism of carbohydrate and lipids. In the +6 oxidation state, i.e. as chromate, it is known to be highly toxic, such that exposure to chromate dust is known to cause lung cancer.

Measurement of Cr in clinical samples is usually for occupational monitoring purposes. In such cases, a spot urine specimen is typically obtained. Serum Cr is occasionally measured for research purposes, but because serum Cr concentrations are so low, ranging from 0.04 to 0.39 $\mu g\,L^{-1}$ (0.77–7.5 $nmol\,L^{-1}$), care should be taken to avoid contamination from stainless steel needles.[92] Plastic blood collection tubes certified for trace element analyses should be used. The method of choice for measuring urinary Cr is ETAAS, and the procedure described by Veillon et al.[93,94] was one of the first to document true normal levels. Chromium is usually determined at the 357.9-nm line, which can be troublesome for continuum-based correction systems both because the intensity of the D_2 source is low, and the background intensity from the urine and modifier matrix is large and structured at this wavelength. Use of high quality pyrolytically coated graphite is also particularly important with this analyte. Paschal and Bailey's method[18] is a classic STPF approach with Zeeman background correction. They used a simple dilution of urine with nitric acid and Triton X-100 as matrix modifier, and reported a detection limit of 0.5 $\mu g\,L^{-1}$. Burguera et al.[95] evaluated several modifiers and two ETAAS systems for the determination of Cr in urine. They compared a longitudinally heated furnace (HGA), with D_2 background correction, to a THGA with a longitudinal Zeeman background correction system. Urine (20 µL) and modifier (10 µL) were deposited on the platform sequentially. Results show that, of the five modifiers evaluated, including Eu, $Mg(NO_3)_2$, Pd, Eu–Pd and Ni, only Eu was effective with the HGA–D_2 arrangement. However, the THGA arrangement was much less troublesome in relation to background correction, and $Mg(NO_3)_2$ was selected for routine work. Both the HGA, with Eu, and the THGA, with $Mg(NO_3)_2$, methods were validated with Standard Reference Material (SRM) 2670 and Seronorm™ control materials. Detection limits reported for Cr in urine are 0.03 $\mu g\,L^{-1}$ (HGA–D_2) and 0.05 $\mu g\,L^{-1}$ (THGA–ZL).

Reference intervals for urine Cr are reported in the range 0.05–0.58 μg L^{-1} (0.962–11.15 nmol L^{-1}), and 0.04–0.39 μg L^{-1} (6.769–7.50 nmol L^{-1}) for Cr in serum.[23] The geometric mean for urine Cr in the US population is 0.13 μg L^{-1} (2.50 nmol L^{-1}) with the 95th percentile at 0.70 μg L^{-1} (13.46 nmol L^{-1}).[67]

5.10 Arsenic in Urine

Arsenic is a clinically significant element because of its distribution in the environment and its relative toxicity. The most common exposures are either occupational or via consumption of contaminated seafood, although intentional poisonings do receive abundant coverage in the news media. The specimen of choice for exposure assessment is a 24-h urine collection. Although As is largely excreted in urine as dimethylarsenic acid (DMA), measurement of total As is routinely performed for clinical diagnostic purposes. Determination of specific As species is discussed in more detail below.

For laboratories with considerable experience with HGAAS, this is the method of choice for urinary As determination. In HGAAS, arsine gas (AsH_3) is generated by the reduction of inorganic As with $NaBH_4$ solution. The arsine gas is swept into a heated quartz absorption tube, where it is reduced to ground state atoms of As that absorb light at the 193.7-nm resonance line. In recent years, HGAAS coupled with FI techniques have improved the determination of As. Hanna et al.[16] described a FI/HGAAS method for measurement of urinary As and its metabolites in which urine is applied to a 1-mL Bond-Elut cartridge with HNO_3–ethanol (1 : 10). The eluate is heated with $K_2Cr_2O_7$ and HNO_3, and then H_2SO_4. The mixture is cooled, H_2O_2 is added and then the mixture is reheated. After cooling, the mixture is treated with KI in HCl followed by mixing with HCl and $NaBH_4$ in NaOH solution. Detection limits of 0.1–0.2 μg L^{-1} are obtained.

The application of ETAAS with continuum background correction systems is known to be troublesome in urinary As determination because of spectral interferences at the 193.7-nm line, volatilization losses and species-specific behavior.[74] However, ETAAS methods based on the STPF concept with Zeeman instrumentation are much more successful. Paschal et al.[96] determined As in acidified urine by ETAAS with Zeeman background correction. They diluted specimens with a $Ni(NO_3)_2 + Mg(NO_3)_2$ modifier in HNO_3. The method detection limit was 8 μg L^{-1} and validation was documented. Subramanian[97] describes an ETAAS method for As in urine that requires a separation step in which As is first extracted from 9 mol L^{-1} HCl into toluene, and then back-extracted into 1% HNO_3 containing 0.05% Ni for deposition on the platform. Nixon et al.[98] compared palladium–persulfate with nickel as a modifier for urinary As by ETAAS and reported better recoveries for some As species using the Pd modifier.

Reference values for As in urine are reported separately, based on seafood consumption. For those with minimal seafood consumption, urinary As levels are generally between 5 and 50 μg 24 h^{-1} (0.07–0.67 μmol 24 h^{-1}).[13] In cases of heavy seafood consumption, As levels in the 200–1000 μg 24 h^{-1} (2.67–13.33 μmol 24 h^{-1}) are known.[23]

5.11 Mercury in Urine and Blood

Mercury is a nonessential, toxic element that has numerous industrial applications. Exposure to toxic levels of Hg is most frequently encountered in occupational settings. However, low-level exposures can also result from consuming contaminated seafood, which not only accumulates Hg, but enables its biotransformation into organomercury species. In some cultures, Hg-containing compounds are used for medicinal purposes, and this can lead to Hg poisoning.

The preference for assessing exposure to inorganic Hg (Hg^0) is a 24-h urine specimen because Hg^0 is cleared from the blood compartment quite rapidly. Whole blood is preferred when exposure to organomercury compounds is suspected. Determination of Hg in blood or urine by AAS is accomplished using the cold vapor technique, in which Hg is reduced to ground state atoms by a reagent such as $SnCl_2$, and the gaseous atoms are swept into a quartz cell mounted in the optical path of the AAS instrument. Measurement is at the 253.7-nm resonance line. In one of the earliest applications, Magos[99] described the selective determination of Hg^0 and methylmercury (Me–Hg) in undigested biological samples by AAS and, later, Magos and Clarkson[100] described the determination of total and inorganic Hg species in blood using CVAAS. They used L-cysteine to bind Hg^0, which was then subsequently treated with $SnCl_2$ to liberate Hg atoms. Total Hg was measured by directly treating the sample with a mixture of $SnCl_2$ and $CdCl_2$; organic Hg was estimated as the difference between the total and the inorganic fractions. Littlejohn et al.[101] modified Magos' CVAAS method further for measuring total and inorganic Hg in urine, and compared it against NAA. They reported a CVAAS detection limit of 0.82 μg L^{-1} for Hg in urine. Toffaletti and Savory[102] used $NaBH_4$ as the reducing agent for the determination of total Hg in urine by CVAAS with a detection limit of 2 μg L^{-1}. However, this procedure does not distinguish between inorganic and total Hg. Sharma and Davis[103] also used $NaBH_4$ reduction to measure total Hg in blood. A similar detection limit (1.15 μg L^{-1}) was reported. Compared to $SnCl_2$ reduction, the advantages of $NaBH_4$ are that it is simple (one

reagent) and rapid (2.5 min analysis time) and requires no sample pretreatment.

The introduction of on-line microwave-assisted digestion equipment coupled with AAS instrumentation has opened up new possibilities for both CVAAS and HGAAS. Several publications[21,104] explore the application of on-line microwave sample pretreatment for the determination of Hg in urine. Weltz et al.[21] described the use of such an arrangement with FI to measure urinary Hg. Urine is mixed off-line with 0.5% $K_2Cr_2O_7$ in HNO_3 (1 : 1) and, in the autosampler vessels, with 1 to 2% (v/v) of a bromination reagent (2.23% $KBrO_3$–8% KBr) before introduction into a carrier stream of 0.3% HCl and passage to the microwave digester. For the lower concentration range (ng L^{-1}), H_2O rather than HCl is used. The effluent from the digester (50–90 °C) is merged with a reductant containing $NaBH_4$ and NaOH plus an antifoaming agent, and passed through a hydride manifold and a gas–liquid separator before filtration and CVAAS. A detection limit of 10 ng L^{-1} is reported based on a 10-mL urine sample.

Guo and Baasner[105] describe a similar FI/CVAAS method for Hg in blood, in which 500 µL of whole blood is mixed off-line with Triton X-100, and mixed with a bromination reagent. The sample is mixed with HCl on-line, passed through the digester and further oxidized with $KMnO_4$, followed by reduction with $NaBH_4$–NaOH and Hg measured by CVAAS. Calibration was linear up to 20 µg L^{-1} with a detection limit of 1 µg L^{-1}.

Nixon et al.[106] described a CVAAS procedure for the determination of inorganic, organic and total Hg in blood and urine. The method is based on a commercial automated FI system with CVAAS. For Hg^0 measurement, either 500 µL whole blood or 1 mL urine are combined with HNO_3, H_2SO_4 and $KMnO_4$ off-line, vortexed and centrifuged. For total Hg, either 500 µL of whole blood or 1 mL of urine are combined with HNO_3, H_2SO_4 and $K_2S_2O_8$ off-line and heated at 95 °C. The samples are analyzed by FI/CVAAS with $SnCl_2$ reduction and HCl for the sample probe wash solution. Calibration was linear from 0.5 to 25 µg L^{-1}. A commercially available FI system with on-line digestion coupled to CVAAS for determination of total and inorganic Hg in whole blood was also described by Chen et al.[107] Inorganic Hg is measured following reduction with $SnCl_2$ and without using the microwave digestion system. Total Hg is determined using a modification of the bromination procedure described by Guo and Baasner,[105] including microwave-assisted digestion followed by $NaBH_4$ reduction. Organomercury concentrations are calculated as the difference between the two. They report detection limits of 0.45 µg L^{-1} and 0.14 µg L^{-1} for Hg^0 and total Hg, respectively.

Normal concentrations of Hg in blood and urine are complicated by the methodology used and whether Hg^0 or total Hg are being reported. For total Hg, recent reports[23] suggest normal levels are <5 µg L^{-1} (<25 nmol L^{-1}) for blood, and <20 µg 24 h^{-1} (<100 nmol L^{-1}) in urine.

6 LITERATURE REVIEW

6.1 General Reviews of Trace Element Analysis of Biological Systems with Atomic Absorption Spectrometry Techniques

Several reviews of clinical applications of analytical atomic spectrometry are available to assist clinical laboratory personnel with the selection of appropriate methods for trace element analysis. Perhaps some of the most definitive texts on atomic absorption methods in occupational and environmental health are the series published by Tsalev and Zaprianov[108] and Tsalev[109] in 1984, and an updated edition by Tsalev[74] in 1995 reporting progress over the 10-year interval. Volumes I[108] and II[109] focus on 34 selected elements that are of occupational and environmental health significance: Al, As, Au, Ag, Ba, Be, Bi, B, Cd, Cr, Co, Cu, Fe,Ga, Ge, Hg, In, Li, Mn, Mo, Ni, Pb, Pd, Pt, Ru, Sb, Se, Si, Sn, St, Te, Tl, V, Zn. In Volume III,[74] 21 new elements were added to bring the total number of elements covered up to 55. A wide variety of biological matrices are included for each element.

An excellent review published by Delves in 1987[110] assesses the role of AAS in clinical analyses and compares it to other spectroscopic techniques, including plasma source mass spectrometry and atomic emission techniques. In his review, Delves also discussed the physical and ionization interferences that can occur with clinical samples in FAAS, and the control and elimination of chemical and physical interferences in ETAAS. Various approaches to background correction in AAS are reviewed with respect to clinical matrices, and include continuum, Zeeman-effect and Smith–Hieftje arrangements. Pre-analytical issues are also discussed, especially contamination control.

Taylor[111] presented a thorough review of clinical applications of FAAS in a volume devoted to the principles and applications of AAS. Sample preparation techniques, including simple dilution, protein precipitation, chelation with solvent extraction and acid digestion are reviewed in some detail. The review also covers devices, such as the slotted quartz tube, that were previously shown to improve sensitivity in the flame for Pb and Cd in blood and Cu and Zn in serum.[72] In the same volume on principles and applications of AAS, Delves and Shuttler[17] critique ETAAS methods for analysis of body fluids and tissues. They address a number of problem areas with ETAAS methods for clinical samples

including physical, chemical and spectral interferences, and contamination errors, and they also address several specific problems that are well known with clinical samples. For example, the problem of carbonaceous build-up on the L'vov platform that occurs with blood matrices is discussed. Delves and Shuttler[17] recommend using an oxygen ashing phase as part of the furnace program, to overcome this problem. These authors also described special difficulties when analyzing urine samples, particularly irreproducible results that are evident with some specimens. Methods for the essential trace elements Zn, Cu, Fe, Se, Mn, Cr, Co, Mo, V and Al are reviewed individually, as are methods for therapeutic elements, Pt and Ag, and the nonessential toxic elements Pb, Cd, Ni, Be and Tl.

Extensive annual reviews on the analysis of clinical and biological materials for the major, trace and ultratrace elements have been published since 1986 as Atomic Spectrometry Updates.[112–124] These updates are an invaluable resource for the latest developments in the field since they include material presented at major conferences on analytical atomic spectrometry. The updates are structured to cover sample collection and pretreatment, including digestion procedures, and cover clinical applications of AAS, AES and ICPMS. In the most recent update,[124] some 35 elements of clinical significance were reviewed separately. Occasionally, clinical applications of atomic spectrometry are reviewed in other analytical journals. Paschal[125] reviewed the selection of spectroscopic method for the determination of toxic and essential elements, and discussed the collection and storage of samples with reference to two examples: Pb in blood and Tl in urine. Alcock[126] presented a concise update on atomic spectrometric methods for clinical samples, reviewing the status of FAAS, ETAAS, ICPAES and ICPMS. In that review, Alcock noted the emergence of ICPMS in clinical applications as a unique tool among the arsenal of spectroscopic methods in its ability to discriminate between different stable isotopes of the same element where they exist. Savory and Herman[127] discussed recent advances in instrumental methods for the measurement and speciation of trace metals in biological samples. Their review covers novel applications of ETAAS and the emergence of ICPMS for ultra-trace metal analysis, and methods for the intracellular localization of trace elements, which are beyond the scope of this text.

6.2 Clinical Applications of Inductively Coupled Plasma Atomic Emission Spectrometry

ICPAES is limited by its poorer sensitivity for some clinically significant elements, and by the need either to dilute blood and serum samples or to pretreat them to avoid physical interferences in the sample introduction arrangement. The performances of ICPAES and NAA were examined for their suitability in determining Al, Cd, Cr, Cu, Mg, Mn, V and Zn in human lung tissue.[128] For ICPAES, samples were pretreated by ashing and the residue dissolved in nitric acid. Both ICPAES and NAA gave reliable results, although Al, Cd and V were troublesome. Direct determination of Ca, Cu, Fe, Mg, Mn and Zn in amniotic fluid samples was accomplished using ICPAES,[129] without the need for sample pretreatment and/or dilution. A sheath gas device and a matrix-matching calibration technique, where standards were prepared in 0.5% bovine serum albumin and 0.76% NaCl were used, and the results validated against National Bureau of Standards human serum. However, amniotic fluid samples had to be centrifuged before being introduced into the nebulizer. In a comparison between ICPAES and AAS, serum samples were analyzed for Zn, Cu, Fe, Mg, Ca, Na and K.[130] Results of this study were that ICPAES yields slightly higher serum Cu, Mg, Ca and K, and lower serum Zn, Fe and Na values compared to AAS. Kimberly and Paschal[34] describe an ICPAES method for rapidly screening urine specimens for up to nine elements including Se, As, Cr, Zn, Cd, Pb, Ni, Mn and Cu. Their method is based on a sequential scanning ICPAES instrument that utilizes a cross-flow nebulizer to handle the high dissolved-solids content of urine specimens. They used Y as an internal standard to correct for viscosity differences between calibration standards and undiluted urine specimens. Although the method was validated, the authors report difficulties with Cd, Se and Zn. Detection limits range from $2\,\mu g\,L^{-1}$ for Cd to $81\,\mu g\,L^{-1}$ for As.

Several workers have optimized ICPAES instrumentation to measure a single element in clinical matrices. Perhaps it is not surprising that, where single-element methods have been developed for clinical matrices, the element has invariably been either Al or Pb. Lyon et al.[131] evaluated an ICPAES procedure for the determination of Al in serum, dialysate fluid and water, and validated it by comparison to an ETAAS method. They concluded that the technique was a suitable replacement for ETAAS. Hu et al.[132] also developed an ICPAES method for Al in biological materials, including human serum, but used an ETV device to deliver the analyte to the plasma. They used a polytetrafluoroethylene chemical modifier to produce a slurry, which was deposited in the furnace. The detection limit was $0.5\,ng\,mL^{-1}$ and the method was validated against CRMs. A method for blood lead was developed for a capacitively coupled microwave-induced plasma (MIP) AES,[133] in which a 5-μL blood sample is deposited, dried and ashed on a W electrode, and subsequently vaporized into the MIP. The method was validated with blood-based reference materials (RMs). Alvarado et al.[134] also developed a

method for the direct determination of lead in blood using an ETV/ICPAES arrangement. A detection limit of 0.7 μg dL^{-1} was reported. Recknagel et al.[135] devised a method for the determination of Se in serum by ICPAES, using an on-line wet-digestion procedure with hydride formation. The detection limit was 5.5 μg L^{-1} with a linear range up to 500 μg L^{-1}. In addition to Se, Fe, Cu and Zn can be determined in the microgram per liter concentration range. Method validation included analysis of reference controls and a comparison with ETAAS for human serum samples.

6.3 Clinical Applications of Inductively Coupled Plasma Mass Spectrometry

6.3.1 Multielement Inductively Coupled Plasma Mass Spectrometry Analysis of Biological Samples

In the late 1980s, Lyon et al.[136] evaluated ICPMS performance for multielement trace analysis in clinical chemistry. They conducted an interlaboratory round-robin in which ICPMS was compared with ICPAES and various AAS instrumentation for a plasma protein solution and a urine QC material. While polyatomic interferences occur with some elements by ICPMS, ICPAES has complex spectra and spectral overlap can be troublesome, while in ETAAS, under- or over-background correction can be problematic. Vaughan et al.[137] also described using ICPMS in a semiquantitative mode as a rapid survey method for profiling trace elements in body fluids. At a 10-fold dilution, up to 13 elements can be detected in serum, and up to 15 elements in whole blood. These authors also reported difficulties with polyatomic interferences, specifically with Cr, Mn and V.

Although clinical applications of ICPMS are less numerous than in environmental and other fields, the appearance of affordable and compact instrumentation based upon an argon plasma and a quadrupole mass analyzer has opened the door for clinical laboratories to explore novel uses. In a 1993 review, Barnes[138] discussed advances in ICPMS and its application in human nutrition and toxicology. The role that ICPMS can play in stable isotope tracer studies was examined and, in particular, its application to kinetic modelling studies. Barnes[138] also concludes that ICPMS was likely to contribute to a wide range of interdisciplinary areas of trace and toxic element biochemical interaction, especially in regard to speciation. This latter subject is reviewed below in more detail.

In a recent treatise on ICPMS, Taylor et al.[139] reviewed clinical applications of ICPMS. Clinical sample preparation is discussed and microwave digestion is highly recommended because the loss of volatile elements is minimized and contamination is under better control. Micronebulizers that operate with solution uptake rates of 85 μL min^{-1} are particularly attractive since they are unaffected by viscosity and surface tension. While tissue analyses are preceded by a digestion step, biological fluids such as serum, blood and urine can be analyzed directly following dilution by factors of up to 50 or more to reduce the amount of matrix introduced into the nebulizer and, thence, into the skimmer cones. Detection limits that are comparable with ETAAS are made possible by the greater sensitivity of ICPMS.

Pruszkowski et al.[140] presented a brief summary of clinical applications of ICPMS and discussed methods for correcting polyatomic interferences that can occur with those elements that are common constituents of biological matrices. For example, the determination of ^{75}As in urine is complicated by an interference from the $^{40}Ar^{35}Cl^+$ ion that also occurs at mass 75. An elemental correction equation is employed that subtracts the contribution from $^{40}Ar^{35}Cl^+$ by measuring its abundance at mass 77 ($^{40}Ar^{37}Cl^+$), which must first be corrected for the ^{77}Se contribution. Several other polyatomic interferences exist for elements of clinical significance including ^{27}Al (CN^+), ^{52}Cr (ArC^+, $OClH^+$), and ^{63}Cu ($ArNa^+$). While use of correction equations may be necessary with ICPMS instruments equipped with quadrupole mass analyzers, use of instrumentation equipped with magnetic and/or electric sector mass analyzers, enable high-resolution inductively coupled plasma mass spectrometry (HRICPMS) to overcome many isobaric and polyatomic interferences.

Nixon and Moyer[141] describe a simple method for the routine determination of four elements, Pb, As, Cd, and Tl, in human urine and whole blood using quadrupole ICPMS in the peak-hopping mode. Urine specimens are diluted by a factor of 10 with 1% nitric acid containing Ga, Rh and Bi as internal standards. Whole blood is treated in much the same way following centrifugation to remove cell debris. The interference for As at mass 75, caused by the $^{40}Ar^{35}Cl^+$ ion, is corrected by counting of $^{16}O^{35}Cl$. Alimonti et al.[142] determined Cr and Ni in human whole blood by ICPMS following microwave-assisted nitric-acid digestion. A Rh internal standard was used. The linear dynamic range was 0.03–30 μg L^{-1} for Cr and 0.04–40 μg L^{-1} for Ni in whole blood, with detection limits of 30 and 40 ng L^{-1} respectively.

Paschal et al.[67] used ICPMS to analyze a subset ($n = 496$) of urine specimens collected as part of the Third National Health and Nutrition Examination Survey (NHANES III) of the United States. Data for 13 elements (Mo, Pb, Sn, Tl, Sb, Mn, Cs, W, Pt, Ba, Be, Cr, Th and Co) were obtained at detection limits that surpassed what was previously possible with other techniques. Normal ranges for these analytes in human urine in the US population have now been established. Nuttal et al.[143] describe their

experience with ICPMS for the routine determination of Al, Cu and Zn in serum, As, Cd, Pb, Mn, Hg and Te in whole blood, and up to 10 elements in urine. They report that Cr and Fe are particularly troublesome to measure in clinical samples by ICPMS because of polyatomic interferences.

In a recent review of trace element analysis of biological materials, Moens[144] described applications of mass spectrometry, including quadrupole ICPMS, HRICPMS, TIMS and gas chromatography/mass spectrometry. The advantages and disadvantages of inorganic mass spectrometry were compared with those of other spectroscopic techniques. Moens[144] believes ICPMS, as a successor to AAS and ICPAES, has some major weaknesses with biological materials, including matrix effects and mass spectral interferences by polyatomic species. Some workers[145] have used on-line anion-exchange columns to remove interferences from digested blood and urine samples prior to analysis by ICPMS. Using this approach, Ge, As and Se could be determined at detection limits of 70, 130 and 300 ng L^{-1} respectively. Other workers have studied interferences from biological fluids, and have suggested use of RMs, and three internal standards rather than one in the determination of Br, Ca, Mg, Rb, Sr and Zn in serum. According to Moens,[142] however, although the multielement capability of ICPMS is one of its most often advertised advantages, more than half of the papers published on biomedical applications deal with single-element determinations. Using a magnetic sector mass spectrometer coupled to a quadrupole ICPMS (i.e. HRICPMS), Moens et al.[144] reported the determination of V, Fe, Cu, Zn and Ag in human serum. Interferences from polyatomic species were significantly reduced with this HRICPMS compared with quadrupole ICPMS.

6.3.2 *Single Element (Total) Determinations in Biological Materials by Inductively Coupled Plasma Mass Spectrometry*

As with ICPAES methods, the most common single element (total) determinations in biological samples by ICPMS are those involving Pb, although methods featuring Al, As or Se are also reported.

Delves and Campbell[147] determined total and isotopic blood Pb concentrations by ICPMS with external calibration and matrix-matched standards. Blood samples were diluted 1 + 24 with the same solution they routinely use as a modifier with ETAAS methods: $(NH_4)_2H_2EDTA$, $NH_4H_2PO_4$ and Triton X-100, and which they report does not block the Meinhard nebulizer. They report a within-run precision of 1.40% relative standard deviation (RSD) at 11.2 μg dL^{-1}, and a detection limit (3 SD) of 1.5 μg dL^{-1}. For Pb isotope ratio measurements, these authors report a between-run precision of 0.054% RSD for $^{208}Pb:^{206}Pb$ and 1.02% RSD for $^{206}Pb:^{207}Pb$. Lead isotope ratios can be useful in apportioning the source of lead exposure in humans by characterizing the specific isotopic signature associated with different Pb sources.[148] According to Delves and Campbell,[147] the $^{206}Pb:^{207}Pb$ ratio determined by ICPMS appears to give better discrimination between various exposure sources.

Paschal et al.[149] from the CDC used ICPMS in the ID mode to measure Pb concentrations in whole blood RMs. This approach gave values for blood Pb that agreed well with TIMS, and is currently used to establish certified target values for blood-based RMs in the CDC's Blood Lead Laboratory Reference System (BLLRS). Murphy and Paulsen,[150] from the US National Institute of Standards and Technology (NIST), independently described the application of IDICPMS to measure lead concentrations in NIST SRM 955a Lead in Blood. They achieved an overall precision on the order of 0.44% RSD around the 5 μg dL^{-1} level ($n = 10$) and 0.22% around 12 μg dL^{-1}.

Bowins and McNutt[57] developed a procedure for the determination of Pb in plasma based on ETV/ICPMS with ID analysis. Because the majority (>95%) of Pb in whole blood is bound to the erythrocytes, normal concentrations of plasma lead are below the detection limit for most ETAAS methods. The mean value estimated by Bowins and McNutt is 1.27 μg L^{-1} (6.1 nmol L^{-1}).

Brunk[151] described a simplified method for serum Al determinations by ICPMS. In this method, 500 μL serum is spiked with 500 μL Y internal standard, and diluted to 5 mL with water. The detection limit is given as 1 μg L^{-1} and the limit of quantitation is 3 μg L^{-1}. Huang et al.[152] describe a FI hydride-generation method with ICPMS for determination of As in biological samples, including urine, with a reported detection limit of 0.6 pg. Rayman et al.[153] measured serum Se by coupling a hydride generation system to ICPMS. This arrangement produced a detection limit for Se of 0.04 μg L^{-1}, but the authors also reported a negative bias when analyzing selected RMs.

6.3.3 *Speciation Analysis*

In most clinical situations, only total element concentrations are routinely reported. This is because total element concentrations are relatively easy to determine, and are interpreted against an established clinical database of reference ranges. The analytical methods are well established and are under reasonably good control, and so the test results are considered reliable. However, reporting total element concentrations implies that only a single chemical species is present. Trace elements, and especially the transition metals, may not be present as a

single species. In respect of oxidation state, coordination number, stereochemistry and binding ligands, trace elements exist as many different chemical species, and with very different biochemical properties. One often cited example is the element Cr, which in the Cr(III) oxidation state is considered an essential element, but in the Cr(VI) oxidation state is carcinogenic. Perhaps a better illustration is the nonessential element Pt. Platinum is successfully used to treat several different tumors as the complex cisplatin (*cis*-diamminedichloroplatinum II). However, as the trans isomer, (*trans*-diamminedichloroplatinum II), it is highly toxic and shows no antitumor activity, despite having the same oxidation state and relative molecular mass as the cis isomer. Thus, an entire area of analytical chemistry has now opened up in which the determination of specific element species is the goal.

There are several excellent texts available that review the current status of clinical applications of speciation.[35,36] Sanz-Medel[36] describes the analytical strategies developed to tackle speciation problems in biological materials. They include computational approaches, direct species-specific systems and hybrid techniques. It is the latter where most efforts are now directed, and involve coupling a separation technique with an element-specific detector. With the emergence of ICPMS as a sensitive multielement detector, its potential in trace element speciation is immense. Most speciation methods focus on a single element, with determination of clinically significant species. They can range from a simple separation of organo-bound species from inorganic species to separation and identification of specific compounds. The separation step can involve the selective use of an on-line microwave-assisted heating arrangement to measure total element concentrations compared to just the inorganic species, or sophisticated chromatographic separations followed by element-specific detection. Most applications of the latter involve coupling the output from liquid chromatography (LC) systems to either AAS, ICPAES or ICPMS systems. Most LC arrangements utilize reversed-phase separations, many with ion-pairing, to separate metal ion species. Yet others have explored size-exclusion, capillary electrophoresis and other chromatographic approaches. Makarov and Szpunar[154] review the use of size exclusion, or gel permeation, separations coupled with ICPMS for speciation in biological matrices. In an authoritative review, Cornelis[35] discusses the current status of speciation studies for a wide range of elements: Al, As, Cd, Cr, Co, Cu, Hg, Ni, Pb, Pt, Se, Sn and Zn.

Several groups have succeeded in coupling element-specific techniques with chromatographic separation for the determination of selenium species in human body fluids. Yang and Jiang[155] used ion-pair LC and ICPMS to measure selenite, SeO_3^{2-}, selenate, SeO_4^{4-} and trimethylselenomium $(CH_3)_3SeO_3^-$ in human urine with detection limits of 0.76, 0.53 and 0.17 $\mu g\,L^{-1}$, respectively. The LC eluant was fed into an ultrasonic nebulizer. Suzuki et al.[156] used size-exclusion chromatography and ICPMS to determine Se-containing compounds in plasma, urine and RBC extracts from rats.

Element-specific techniques have also been coupled with separation techniques for the on-line determination of As species in serum and urine. Zhang et al.[157] report using cation-exchange LC, UV photo-oxidation and continuous HGAAS to determine monomethylarsonic acid (MMA), DMA, arsenobetaine (AsB) and arsenocholine (AsC) in human serum. Detection limits for the As species are 1–1.5 $\mu g\,L^{-1}$. Serum from uremic patients was analyzed but only DMA and AsB were detected. The method was subsequently used to determine DMA and AsB in urine RMs.[158] This same group also evaluated three type of LC systems coupled to HGAAS to measure As species in human serum.[159] The LC systems included reversed-phase ion-pair, and anion-exchange columns of both the cross-linked polymer and silica-bonded types. The latter appeared best suited to this kind of separation in serum samples, and inorganic As(III) and As(IV), MMA and DMA were determined with detection limits of 0.49, 0.44, 0.92 and 0.40 $\mu g\,L^{-1}$ respectively. Ding et al.[160] measured four Se species in urine using LC/ICPMS, including MMA, DMA, As(III) and As(V), with detection limits from 90–300 pg, and a dynamic range of three orders of magnitude.

Aizpun et al.[161] described a method for the speciation of inorganic Hg^0 and methylmercury (MeHg) in urine using LC and Hg detection by CVAAS. The linear dynamic range for MeHg was 10–400 $\mu g\,L^{-1}$ and 16–400 $\mu g\,L^{-1}$ for Hg^0. Detection limits were 10 $\mu g\,L^{-1}$ and 16 $\mu g\,L^{-1}$ respectively. Yin et al.[162] reported the determination of MeHg, ethylmercury (EtHg), phenylmercury (PhHg) and Hg^0 in urine. Organomercury species are first preconcentrated on a C_{18} microcolumn, and then separated by reversed-phase LC with detection by CVAAS. Use of $NaBH_4$ as a reductant combined with a thermolysis step improved sensitivity, and detection limits for MeHg, EtHg, PhHg and Hg(II) were 9, 6, 10 and 5 ng L, respectively, based on a 58.5 mL sample volume.

7 CLINICAL LABORATORY REGULATIONS

In the United States, all clinical laboratories that test specimens of human origin are regulated under The Clinical Laboratory Improvement Amendments of 1988 (CLIA 88), using a complexity-based model. The federal regulations that govern clinical laboratories are the responsibility of the Health Care and Financing

Administration (HCFA), which is an arm of the US Department of Health and Human Services (DHHS). The final rules implementing CLIA 88 were published in 1992.[163] Other federal agencies also have responsibilities in clinical laboratory issues. For example, the Food and Drug Administration (FDA), which is also part of the DHHS, has responsibility for classifying clinical test devices marketed and sold in the US. The established analytical atomic spectrometric instrumentation that is typically used to determine elemental content in body tissues and fluids is classified as highly complex, which means such testing must be carried out in a duly licensed clinical laboratory facility. The Occupational Safety and Health Administration (OSHA), which is part of the US Department of Labor, issues regulations governing the biomonitoring of workers occupationally exposed to heavy metals. Clinical laboratories that test human blood or urine specimens for heavy metal exposure as part of an occupational monitoring program are required to be approved by OSHA too. In additional to federal agencies, many states also have laws governing clinical laboratory operations. Some states, e.g. New York, are ''CLIA-exempt'', which means that any clinical laboratory based in that state, and holding a valid state permit, is deemed to have satisfied federal requirements, i.e. CLIA 88.

Clinical laboratory regulations cover a variety of matters ranging from the qualifications and education of the director and technical staff to the operation of the laboratory, its analytical methods, and the competency of the laboratory to perform those methods. QA and QC requirements are specified in the regulations. As part of the QA/QC requirements, all clinical laboratories must

Table 4 PT programs and EQAS for trace elements in clinical samples

North American programs approved for blood lead

Program	Address	City	State/ Province	Zip code/ Post code	Country	HCFA	OSHA
College of American Pathologists	325 Waukegan Road	Northfield	IL	60093	USA	✓	✓
New York State Department of Health	Wadsworth Center PO Box 509	Albany	NY	12201	USA	✓	✓
Pennsylvania Department of Health	Bureau of Labs. Pickering Way & Welsh Rd	Lionville	PA	19353	USA	✓	
Wisconsin State Laboratory of Hygiene	2601 Agriculture Drive PO Box 7996	Madison	WI	53707	USA	✓	✓
Le Centre de Toxicologie du Québec	2705 Boul. Laurier	Sainte-Foy	Qc	G1V 4G2	Canada		✓

Programs available for trace elements in clinical samples

Program	Address	City	State/ Province	Zip code/ Post code	Country	Blood	Serum	Urine
College of American Pathologists	325 Waukegan Road	Northfield	IL	60093	USA	Pb	Se, Al, Cr, Zn	
Le Centre de Toxicologie du Québec	2705 Boul. Laurier	Sainte-Foy	Qc	G1V 4G2	Canada	Cd, Pb, Hg,	Al, Cu, Se, Zn	As, Cd, Cr, Cu, Pb, Hg, Se
Guildford Trace Elements EQA Scheme	Centre for Clinical Science and Measurement, School of Biological Sciences, University of Surrey	Guildford	Surrey	GU2 5XH	UK	Pb, Cd	Al, Au, Cu, Se, Zn	As, Cd, Hg
UK NEQAS	Wolfson EQA Laboratory PO Box 3909	Birmingham	Worcs.	B15 2UE	UK	Pb, Cd		
Worldwide Interlaboratory Aluminum QC	Lab of Biochemistry and Toxicololgy Jean Bernard Hospital, BP 577	Poitiers – Cédex		86021	France		Al	

participate in an external PT program that is approved by DHHS, and that reflects the specific laboratory specialty. For the trace element area, only blood lead is specifically identified for PT purposes under CLIA 88. Table 4 lists the currently approved PT programs for blood lead. The design of the PT program is specified under CLIA 88.[163] Five PT challenges (or samples) are provided three times a year. Target values are established by the program, and the acceptable range for scoring results is fixed at $\pm 4\,\mu g\,dL^{-1}$ or $\pm 15\%$, whichever is greater. Sanctions may be imposed on laboratories that fail to report correct responses for at least 80% (four out of five) of the PT samples for two out of three consecutive test events.

Outside the USA, clinical laboratory standards vary considerably, and are beyond the scope of this text. The approach in Europe relies more on a voluntary compliance standard, with participation in EQAS encouraged. Several schemes are available that offer challenges for a range of trace elements in serum, blood and urine (Table 4).

8 QUALITY ASSURANCE AND QUALITY CONTROL

QA is defined as the planned and systematic activities implemented to provide confidence that an entity will fulfill requirements for quality.[164] Within the analytical laboratory, QC is defined as a systematic process for evaluating and monitoring the accuracy and precision of any analytical process performed either repetitively in the same laboratory (internal quality control, IQC), or between laboratories (external quality control, EQC).[164] Therefore, in a broad sense, components of a QA program will include staff training, standard operating procedure manuals, preventative maintenance programs, and documentation of remediation, as well as the more familiar aspects of QC that are part of the analytical phase, such as analyzing routine QC materials, that assure the accuracy and validity of the analytical result.

8.1 Reference Materials, Control Materials and Proficiency Testing

Weber[25] discussed laboratory quality within the context of environmental toxicology measurements. Although use of CRMs is recommended to verify accuracy, they are usually treated (e.g. lyophilized) for long-term conservation, and so may behave differently from real patient samples. Thus, the analyst should be cautious in interpreting CRM results. Weber concluded that participation in an intercomparison program, such as the Québec Trace Elements Intercomparison Program, is desirable to demonstrate that results are not only reproducible but also accurate. Taylor and Briggs[165] discussed changes that have occurred over a 10-year period in the Trace Elements EQAS that originated at the Robens Institute.[166] Assays assessed by the scheme now number 23, and include measurements in serum, whole blood, urine, dialysis fluids and water. Issues such as sample preparation, assay range, performance assessment and shared IQC programs are also discussed. They conclude that laboratory performance has improved considerably and cite as an example the measurement of Al in serum.

Since blood lead is specifically covered in regulations under CLIA 88, the situation here is much better. Three PT schemes for blood lead are approved by HCFA in the US (Table 4), and improvements in laboratory performance have been documented.[167,168] Elsewhere, laboratory improvements for blood lead have been documented as a result of participation in an EQAS.[169]

Reference materials are generally provided by government-run laboratories with expertise in the measurement of trace elements, and who provide a certificate stating the origin of the materials, the procedures by which values are certified and a statement of the analytical uncertainty. A listing of RM providers is given in Table 5. The NIST, formerly known as the National Bureau of Standards, has been particularly active in the development of well-characterized biological SRMs.[170,171] A listing of NIST's current clinical SRMs, with either certified or recommended values for trace elements, is provided in Table 5. Veillon et al.[172,173] described in great detail the procedures by which a bovine serum RM was prepared and characterized for up to 15 elements, which was later made available as NBS SRM 8419.

Control materials from commercial sources may be less well characterized, and ''expected'' values may not always be reliable, or they may be accompanied by very large uncertainties.[19,174] A listing of commercial control materials for trace elements studies in biological materials is given Table 6. Külpmann et al.[63] analyzed NIST SRM 909 and several commercial control sera for Ca, Li and Mg using well-established FAAS reference methods, and compared the results with method-dependent target values. Their results confirm the reliability of FAAS as a reference method, but there was also considerable disagreement between expected Mg values and those actually found by FAAS. Parsons and Slavin[63] analyzed a number of commercial urine control materials for lead content using an ETAAS method that was validated with NIST SRM 2670, and New York State RMs for urine lead. They found considerable variability with some commercial materials. Pizent and Teilsman[19] analyzed

Table 5 Clinical RMs provided by government or international agencies with certified trace element concentrations

Source	Product	Description	Elements certified (info. only)
NIST Office of SRMs – Rm 205 Gaithersburg, MD 20899 USA	SRMs	909b Human serum 955b Lead in blood 966 Toxic elements in blood 2670-2 Toxic metals in urine 2672a Mercury in urine	Ca, Li, Mg, K, Na Pb (4 levels) (in prep) (Al), As, (Be), Cd, Ca, Cl, Cr, Cu, (Au), Pb, Mg, (Mn), Hg, (Ni), (Pt), (K), Se, Na (V) Hg
Commission of the European Communities Community Bureau of Reference Rue de la Loi 200, B-1049, Brussels, Belgium	BCR No. 194–196 BCR No. 184 BCR No. 185 BCR No. 186	Lead and cadmium in blood Trace elements in bovine muscle Trace elements in bovine liver Trace elements in pig kidney	Pb, Cd (three levels) Cd, Pb, Hg, Se, Cu, Zn, Fe, Mn Cd, Pb, Hg, As, Se, Cu, Zn, Fe, Mn Cd, Pb, Hg, As, Se, Cu, Zn, Fe, Mn
International Atomic Energy Agency PO Box 100, A-1400, Vienna, Austria	IAEA-A-13	Animal blood	Br, Ca, Cu, Fe, K, (Mg), Na, (P), Rb, S, Se, Zn (IAEA recommended values)
Centers for Disease Control BLLRS 4770 Buford Highway NE Atlanta, GA 30341-3724, USA	BLLRS	Lead in whole blood	Pb
New York State Department of Health Wadsworth Center, PO Box 509 Albany NY 12201-0509, USA	Blood Pb RMs Urine Pb RMs	Lead in whole blood RM Lead in human urine RM	Pb (three levels) Pb (four levels)
Guildford Trace Elements EQA Scheme Centre for Clinical Science and Measurement, School of Biological Sciences, University of Surrey, Guildford, GU2 5XH, UK	EQAS materials from previous distributions	Whole blood RMs Serum RMs Urine RMs	Pb, Cd Al, Au, Cu, Se, Zn As, Cd, Hg
Le Centre de Toxicologie du Québec, 2705 Boul. Laurier Sainte-Foy, Qc G1V 4G2, Canada	Materials from previous intercomparisons	Whole blood RMs Serum RMs Urine RMs	Cd, Pb, Hg Al, Cu, Se, Zn, As, Cd, Cr, Cu, Pb, Hg, Se
Office of CRMs National Research Centre for CRMs No. 7 District 11 Hepingjie Chaoyangqu Beijing 100013, People's Republic of China	GBW 09132-09134 GBW 09102-09103 GBW 09104-09105 GBW 09135 GBW 09139-09140 GBW 09141-09142	Whole bovine blood CRMs Lyophilized human urine Pb in freeze-dried human urine Inorganic elements in human serum Freeze-dried bovine blood Se in freeze-dried bovine blood	Pb, Cd As, Be, Cd, Cr, Cu, Mn, Ni, Pb, Se, Zn Pb Mg, Cu, Zn, Ca, Fe, K, Na, Cl, P, Pb Pb, Cd Se

commercial sera control materials for Cu and Zn using a FAAS method. Results obtained for three out of four controls were in good agreement but precision and accuracy for the other were poor. They suggest that nonhomogeneity of the lyophilized material and/or uncertified values were the cause. Veillon et al.[175] analyzed a commercial control material for Cr, Zn and Se using IDMS and/or AAS. They found Cr and Zn were seriously affected by contamination from the sample containers. Other potential contaminants included B, Ba, Sr, Mo and Cu.

The use of well-characterized RMs for IQC has proven to be very successful in improving laboratory performance for blood lead.[176] Delves[177] argued that RMs, rather than reference methods, are more likely to improve laboratory performance, based on the experience

Table 6 Commercial clinical control materials provided with certified trace element concentrations

Source	Product	Description	Elements certified (info. only)
Nycomed Pharma AS (Sero AS, PO Box 24 N-1361 Billingstad, Norway) (US distribution: Accurate Chemical and Scientific Corp., 300 Shames Drive, Westbury, NY 11590, USA)	Seronorm™	Trace elements whole blood	Cd, Cr, Co, Pb, Mn, Hg Ni, Se (Al, Sb, As, Be, Bi, Mo, Tl, V)
	Seronorm™	Trace elements serum	Al, Ca, Cr, Co, Cu, Fe, Mn, Mg, Ni, P, K, Se, S, Na, Zn
	Seronorm™	Trace elements urine	Al, Sb, As, Cd, Ca, Cr, Co, Cu, Pb, Mn, Mg, Hg, Ni, K, Se, Na, Tl, Sn Zn (Be, Bi, Au, Te, V)
Bio-Rad, ECS Division 1000 Alfred Nobel Drive Hercules, CA 94547, USA	Lyphocheck®	Whole blood control	Pb
	Lyphocheck®	Urine metals control	Al, Sb, As, Cr, Co, Cu, Pb, Mg, Mn, Hg, Ni, Se, Tl, Zn
	Lyphocheck®	Quantitative urine controls	Ca, K, Na, As, Cu, Fe, Pb, Mg, Hg, Zn
Utak Laboratories, Inc. 25020 Avenue Tibbitts Valencia, CA 91355, USA	TDM–Tox Controls	Urine metals control	Al, As, Cd, Cr, Co, Cu, Pb, Mn, Hg, Mo, Ni, Se, V, Zn
	TDM–Tox Controls	Serum trace elements	Al, Cd, Cr, Cu, Mg, Mn, Se, Zn
	Blood Lead Tri set	Whole blood lead controls	Pb
Kaulson Laboratories, Inc. 687-691 Bloomfield Ave. West Caldwell, NJ 07006, USA	Contox®	Heavy–trace metal urine	As, Cd, Hg, Pb, Cu, Fe, Zn, Al, Mg, Mn, Be, Co, Cr, Ni, Se, Sb, Bi, Ti, Sn
	Contox®	Trace metal serum control	Cu, Fe, Zn, Al, Mg, Mn, Be, Co, Cr, Ni, Se, Sb, Bi, Sn, As, Cd, Pb, Hg, Ti
	Contox®	Whole blood and urine lead	Pb
	Contox®	Heavy metal blood control	As, Cd, Hg, Pb
AMI/Referensmaterial AB Lobeliav. 6, S-52333 Ulricehamm, Sweden	AMI B1001-B1005	Pb in lyophilized human blood	Pb
	AMI B1701-B1703	Lyophilized human blood	Pb
Ciba-Corning Diagnostics Corporation 63 North Street, Medfield, MA 02052, USA	Urine control I, II	Urine metals control	As, Cu, Pb, Mg, Hg, Zn, Ca, K, Na
	BLD TOX I, II	Whole blood toxicology control	

of European laboratories conducting blood Pb testing. Extrapolating from the blood Pb experience, it is likely that with the increased development of reliable clinical RMs certified for other trace elements, laboratory performance for other elements will improve too.[178] This has certainly been the case for serum Al, which, like Pb, has been the focus of increased attention in recent times.[179,180]

Clinical applications of analytical atomic spectrometry continue to appear in literature with even lower detection limits, improved performance and simpler sample preparation procedures. One area that has received attention recently is the development of low-cost, portable ETAAS instrumentation based on W-filament atomization, where measurements of Pb and Cd in blood have been shown.[181–184] But as applications of ETAAS decline, ICPMS is emerging as a powerful alternative, and as the cost of ICPMS instrumentation becomes more affordable, more routine laboratories will likely adopt the technology. The limitations of current ICPMS technologies include isobaric and polyatomic interferences that originate from the sample matrix. In this respect, HRICPMS has some definite advantages but current instrumentation may be too costly for routine clinical laboratories. Development of atomic mass spectrometers that are more efficient at detecting atoms and use more of the sample will be required in the future. ICPMS instrumentation based on time-of-flight mass spectrometry will be a serious competitor to quadrupole-based systems. Perhaps the most interesting future developments will occur in speciation techniques, where information on specific metal species will become increasingly significant within a clinical context. However, there is still an urgent need for certified clinical RMs that contain endogenously bound trace elements certified at clinically relevant concentrations. Such materials will be needed to validate speciation methods.

ACKNOWLEDGMENTS

I am indebted to Walter Slavin for his many helpful suggestions and encouragement in preparing this article. I am also very grateful to the external reviewer who critiqued the material very carefully, and suggested helpful improvements.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
AsB	Arsenobetaine
AsC	Arsenocholine
BLLRS	Blood Lead Laboratory Reference System
CDC	Centers for Disease Control and Prevention
CRM	Certified Reference Material
CVAAS	Cold Vapor Atomic Absorption Spectrometry
DCAAS	Delves' Cup Microsampling Flame Atomic Absorption Spectrometry
DHHS	Department of Health and Human Services
DMA	Dimethylarsenic Acid
EDL	Electrodeless Discharge Lamp
EDTA	Ethylenediamine Tetraacetic Acid
EQAS	External Quality Assessment Schemes
EQC	External Quality Control
ETAAS	Electrothermal Atomic Absorption Spectrometry
ETV	Electrothermal Vaporization
FAAS	Flame Atomic Absorption Spectrometry
FAES	Flame Atomic Emission Spectrometry
FDA	Food and Drug Administration
FI	Flow Injection
HCFA	Health Care and Financing Administration
HCL	Hollow-cathode Lamp
HGA	Longitudinally Heated Furnace
HGAAS	Hydride Generation Atomic Absorption Spectrometry
HRICPMS	High-resolution Inductively Coupled Plasma Mass Spectrometry
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ID	Isotope Dilution
IDICPMS	Isotope Dilution Inductively Coupled Plasma Mass Spectrometry
IFCC	International Federation of Clinical Chemistry
IQC	Internal Quality Control
ISE	Ion-selective Electrode
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
MIP	Microwave-induced Plasma
MMA	Monomethylarsonic Acid
NAA	Neutron Activation Analysis
NCCLS	National Committee for Clinical Laboratory Standards
NIST	National Institute of Standards and Technology
NRSCL	National Reference System for the Clinical Laboratory
OSHA	Occupational Safety and Health Administration
PT	Proficiency Testing
QA	Quality Assurance
QC	Quality Control
RBC	Red Blood Cell
RM	Reference Material
RSD	Relative Standard Deviation
SRM	Standard Reference Material
STPF	Stabilized Temperature Platform Furnace
THGA	Transversely Heated Graphite Furnace
TIMS	Thermal Ionization Mass Spectrometry
UV	Ultraviolet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. E. Epstein, 'Mineral Metabolism', in *Plant Biochemistry*, eds. J. Bonner, J.E. Varner, Academic Press, London, 438–466, 1965.
2. M. Patriarca, A. Menditto, G. Difelice, F. Petrucci, S. Caroli, M. Merli, C. Valente, 'Recent Developments in Trace Element Analysis in the Prevention, Diagnosis, and Treatment of Diseases', *Microchem. J.*, **59**, 194–202 (1998).
3. S.A. Lewis, T.C. O'Haver, J.M. Harnly, 'Analysis of Blood Serum for Essential Metals by Simultaneous Multi-element Atomic-absorption Spectrometry with Flame Atomization', *Anal. Chem.*, **56**, 1066–1070 (1984).
4. Z.A. Grosser, C.A. Schneider, 'Flame Atomic Absorption, what is its Future?', *At. Spectrosc.*, **17**, 209–210 (1996).
5. J.M. Harnly, 'Instrumentation for Simultaneous Multielement Atomic-absorption Spectrometry with Graphite-furnace Atomization', *Fresenius' J. Anal. Chem.*, **355**, 501–509 (1996).
6. W. Slavin, 'Comparison of Atomic-spectroscopic Analytical Techniques', *Spectroscopy*, **6**, 16–21 (1991).
7. R.A. Velapoldi, R.C. Paule, R. Schaffer, J. Mandel, J.R. Moody, 'A Reference Method for the Determination of Sodium in Serum', *NBS Spec. Publ.*, **260-60**, 1–94 (1978).
8. W.J. Korzun, W.G. Miller, 'Sodium and Potassium', in *Methods in Clinical Chemistry*, eds. A.J. Pesce, L.A. Kaplan, The C.V. Mosby Company, St Louis, 86–91, 1987.
9. R.A. Velapoldi, R.C. Paule, R. Schaffer, J. Mandel, L.A. Machlan, J.W. Gramlich, 'A Reference Method for the Determination of Potassium in Serum', *NBS Spec. Publ.*, **260-63**, 1–89 (1979).
10. R.A. Velapoldi, R.C. Paule, R. Schaffer, J. Mandel, L.A. Machlan, E.L. Garner, T.C. Rains, 'Reference Method for the Determination of Lithium in Serum', *NBS Spec. Publ., 1980*, 114 (1980).
11. J.P. Cali, J. Mandel, L. Moore, D.S. Young, 'A Referee Method for the Determination of Calcium in Serum', *NBS Spec. Publ.*, **260-36**, 1–121 (1978).
12. E.C. Farrell Jr, 'Calcium', in *Methods in Clinical Chemistry*, eds. A.J. Pesce, L.A. Kaplan, The C.V. Mosby Company, St Louis, 1003–1009, 1987.
13. D.B. Milne, 'Trace Elements', in *Tietz Textbook of Clinical Chemistry*, 3rd edition, eds. C.A. Burtis, E.R. Ashwood, W.B. Saunders, Philadelphia, 1029–1055, 1999.
14. National Committee for Clinical Laboratory Standards, 'Standardization of Sodium and Potassium Ion-selective Electrode Systems to the Flame Photometric Reference Method; Approved Standard', Ed. 15 NCCLS Document C29-A, Villanova, PA, 1–22, 1995.
15. S. Tang, P.J. Parsons, W. Slavin, 'Rapid and Reliable Method for the Determination of Aluminium in Bone by Electrothermal Atomic Absorption Spectrometry', *Analyst*, **121**, 195–200 (1996).
16. C.P. Hanna, J.F. Tyson, S. McIntosh, 'Determination of Inorganic Arsenic and its Organic Metabolites in Urine by Flow-injection Hydride-generation Atomic-absorption Spectrometry', *Clin. Chem.*, **39**, 1662–1667 (1993).
17. H.T. Delves, I.L. Shuttler, 'Elemental Analysis of Body Fluids and Tissues by Electrothermal Atomization and Atomic Absorption Spectrometry', in *Atomic Absorption Spectrometry. Theory, Design and Applications*, ed. S.J. Haswell, Elsevier, Amsterdam, 381–438, 1991.
18. D.C. Paschal, G.G. Bailey, 'Determination of Chromium in Urine with Graphite-furnace Atomic-absorption Spectroscopy Using Zeeman Correction', *At. Spectrosc.*, **12**, 151–154 (1991).
19. A. Pizent, S. Telisman, 'Analysis of Reference Materials for Serum Copper and Zinc by Flame AAS', *At. Spectrosc.*, **17**, 88–91 (1996).
20. P.J. Parsons, W. Slavin, 'A Rapid Zeeman Graphite Furnace Atomic Absorption Spectrometric Method for the Determination of Lead in Blood', *Spectrochim. Acta, Part B*, **48**, 925–939 (1993).
21. B. Welz, D.L. Tsalev, M. Sperling, 'Online Microwave Sample Pre-treatment for the Determination of Mercury in Water and Urine by Flow-injection Cold-vapour Atomic-absorption Spectrometry', *Anal. Chim. Acta*, **261**, 91–103 (1992).
22. P. Van Dael, R. Van Cauwenbergh, H. Robberecht, H. Deelstra, M. Calomme, 'Determination of Selenium in Human Serum by AAS Using Electrothermal Atomization with Longitudinal Zeeman-effect Background Correction or Flow-injection Hydride Generation', *At. Spectrosc.*, **16**, 251–255 (1995).
23. National Committee for Clinical Laboratory Standards, 'Control of Pre-analytical Variation in Trace Element Determinations; Approved Guideline', Ed. 15 NCCLS Document C38-A, Wayne, PA, 1–37, 1997.
24. Interlaboratory Comparsion Program [Report]. Run #1', Le Centre de toxicologie du Québec, Sainte-Foy, 1999.
25. J.P. Weber, 'Quality in Environmental Toxicology Measurements', *Therapeutic Drug Monitoring*, **18**, 477–483 (1996).
26. National Committee for Clinical Laboratory Standards, 'Sodium; Proposed Summary of Methods and Materials

Credentialed by the NRSCL Council', Ed. 15 NCCLS Document RS7-P, Villanova, PA, 1988.
27. National Committee for Clinical Laboratory Standards, 'Potassium; Proposed Summary of Methods and Materials Credentialed by the NRSCL Council', Ed. 15 NCCLS Document RS8-P, Villanova, PA, 1988.
28. H.T. Delves, 'A Micro-sampling Method for the Rapid Determination of Lead in Blood by Atomic-absorption Spectrophotometry', *Analyst*, **95**, 431–438 (1970).
29. D.G. Mitchell, K.M. Aldous, F.J. Ryan, 'Automated Determination of Blood Lead Levels in Microsamples Using a Delves Cup, Photon-counting, On-line Computer, Atomic Absorption System', in *Environmental Health Aspects of Lead: Proceedings of an International Symposium*, eds. D. Barth, A. Berlin, R. Engel, P. Recht, J. Smeets, Commission of the European Communities, Luxembourg, 10813–11089, 1973.
30. K.M. Aldous, D.G. Mitchell, F.J. Ryan, 'Computer Controlled Atomic Absorption Spectrometer for Measurement of Transient Atom Populations', *Anal. Chem.*, **45**, 1990–1993 (1973).
31. W. Slavin, D.C. Manning, G.R. Carnrick, 'The Stabilized Temperature Platform Furnace', *At. Spectrosc.*, **2**(5), 137–145 (1981).
32. W. Slavin, *Graphite Furnace AAS: A Source Book*, The Perkin-Elmer Corporation, Norwalk, CT, 1–229, 1984.
33. W. Slavin, 'Graphite-furnace AAS for Biological Materials', *Sci. Total Environ.*, **71**, 17–35 (1988).
34. M.M. Kimberly, D.C. Paschal, 'Screening for Selected Toxic Elements in Urine by Sequential-scanning Inductively Coupled Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **174**, 203–210 (1985).
35. R. Cornelis, 'Involvement of Analytical Chemistry in Chemical Speciation of Metals in Clinical Samples', *Ann. Clin. Lab. Sci.*, **26**, 252–263 (1996).
36. A. Sanz-Medel, 'Trace Element Analytical Speciation in Biological Systems – Importance, Challenges and Trends', *Spectrochim. Acta, Part B*, **53**, 197–211 (1998).
37. National Committee for Clinical Laboratory Standards, 'Analytical Procedures for the Determination of Lead in Blood and Urine; Proposed Guideline', NCCLS Document C40-P, Wayne, PA, 1–74, 1998.
38. H.T. Delves, 'Sample Preparation and Handling', *Food Chem.*, **43**, 277–281 (1992).
39. R. Cornelis, B. Heinzow, R.F. Herber, J.M. Christensen, O.M. Poulsen, E. Sabbioni, D.M. Templeton, Y. Thomassen, M. Vahter, O. Vesterberg, 'Sample Collection Guidelines for Trace Elements in Blood and Urine. IUPAC Commission of Toxicology', *J. Trace Elem. Med Biol.*, **10**, 103–127 (1996).
40. 'Recommendations for Prevention of HIV Transmission in Health-care Settings', *MMWR Morb. Mortal. Wkly Rep.*, **36**(Suppl. 2), 1S–18S (1987).
41. J.S. Garner, 'Guideline for Isolation Precautions in Hospitals. The Hospital Infection Control Practices Committee', *Infect. Control Hosp. Epidemiol.*, **17**, 53–80 (1996).
42. National Committee for Clinical Laboratory Standards, 'Protection of Laboratory Workers from Instrument Biohazards and Infectious Disease Transmitted by Blood, Body Fluids, and Tissue; Approved Guideline', Ed. 15 NCCLS Document M29-A, Wayne, PA, 1, 1997.
43. J.A. Moreton, H.T. Delves, 'Use of Virkon as a Disinfectant for Clinical Samples Carrying a High Risk of Infection in Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **14**, 893–894 (1999).
44. P.J. Parsons, A.A. Reilly, D. Esernio-Jenssen, 'Screening Children Exposed to Lead: An Assessment of the Capillary Blood Lead Fingerstick Test', *Clin. Chem.*, **43**, 302–311 (1997).
45. A. Taylor, 'Usefulness of Measurements of Trace Elements in Hair', *Ann. Clin. Biochem.*, **23**, 364–378 (1986).
46. B.L. Gulson, 'Nails – Concern over their Use in Lead Exposure Assessment', *Sci. Total. Environ.*, **177**, 323–327 (1996).
47. R. Cornelis, X. Fuentes-Arderiu, I. Bruunshuus, D. Templeton, 'Properties and Units in the Clinical Laboratory Sciences. IX. Properties, Units in Trace Elements (IUPAC-IFCC Technical Report 1997). International Union of Pure and Applied Chemistry and International Federation of Clinical Chemistry', *Clin. Chim. Acta*, **268**, S75–S89 (1997).
48. R. Cornelis, X. Fuentes-Arderiu, I. Bruunshuus, D. Templeton, 'International Union of Pure and Applied Chemistry and International Federation of Clinical Chemistry, Committee on Nomenclature. Properties and Units (C-NPU): Properties and Units in the Clinical Laboratory Sciences. IX. Properties and Units in Trace Elements (IFCC-IUPAC Technical Report 1997)', *Eur. J. Clin. Chem. Clin. Biochem.*, **35**, 833–843 (1997).
49. National Committee for Clinical Laboratory Standards, 'A Designated Comparison Method for the Measurement of Ionized Calcium in Serum; Proposed Standard', NCCLS Document C39-P, Wayne, PA, 1–74, 1998.
50. Y. Schmitt, 'Copper and Zinc Determination in Plasma and Corpuscular Components of Peripheral Blood of Patients with Preterminal and Terminal Renal Failure', *J. Trace Elem. Med. Biol.*, **11**, 210–214 (1997).
51. A.S. Attiyat, G.D. Christian, 'Flow-injection Analysis – Atomic-absorption Determination of Serum Zinc', *Clin. Chim. Acta*, **137**, 151–157 (1984).
52. K.W. Simonsen, B. Nielsen, A. Jensen, J.R. Andersen, 'Direct Microcomputer-controlled Determination of Zinc in Human Serum by Flow-injection Atomic-absorption Spectrometry', *J. Anal. At. Spectrom.*, **1**, 453–456 (1986).
53. B.F. Rocks, R.A. Sherwood, L.M. Bayford, C. Riley, 'Zinc and Copper Determinations in Micro-samples of Serum by Flow Injection and Atomic-absorption Spectroscopy', *Ann. Clin. Biochem.*, **19**, 338–344 (1982).

54. N.W. Alcock, 'Copper', in *Methods in Clinical Chemistry*, eds. A.J. Pesce, L.A. Kaplan, The C.V. Mosby Company, St Louis, 527–538, 1987.
55. S. Lapointe, A. LeBlanc, 'Determination of Copper in Blood Serum or Plasma by Graphite-furnace AAS', *At. Spectrosc.*, **17**, 163–166 (1996).
56. American Academy of Pediatrics Committee on Drugs, 'Treatment Guidelines for Lead Exposure in Children', *Pediatrics.*, **96**, 155–160 (1995).
57. R.J. Bowins, R.H. McNutt, 'Electrothermal Isotope Dilution Inductively Coupled Plasma Mass Spectrometry Method for the Determination of Sub-ng mL^{-1} Levels of Lead in Human Plasma', *J. Anal. At. Spectrom.*, **9**, 1233–1236 (1994).
58. Y.Y. Zong, P.J. Parsons, W. Slavin, 'Accurate and precise Measurements of Lead in Bone Using Graphite Furnace Atomic Absorption Spectrometry with Zeeman-effect Background Correction', *J. Anal. At. Spectrom.*, **11**, 25–30 (1996).
59. E. Pruszkowska, G.R. Carnrick, W. Slavin, 'Blood Lead Determination with the Platform-furnace Technique', *At. Spectrosc.*, **4**, 59–61 (1983).
60. D.T. Miller, D.C. Paschal, E.W. Gunter, P.E. Stroud, J. D' Angelo, 'Determination of Lead in Blood Using Electrothermal Atomization Atomic-absorption Spectrometry with a L'vov Platform and Matrix Modifier', *Analyst*, **112**, 1701–1704 (1987).
61. K.S. Subramanian, 'Determination of Lead in Blood by Graphite Furnace Atomic Absorption Spectrometry – A Critique. [Review.]', *Sci. Total. Environ.*, **89**, 237–250 (1989).
62. H. Qiao, P.J. Parsons, W. Slavin, 'Transferability of Blood Lead Determinations by Furnace Atomic Absorption Spectrophotometry and Continuum Background Correction', *Clin. Chem.*, **41**, 1451–1454 (1995).
63. P.J. Parsons, W. Slavin, 'Electrothermal Atomization Atomic Absorption Spectrometry for the Determination of Lead in Urine: Results of an Interlaboratory Study', *Spectrochim. Acta, Part B*, **54**, 853–864 (1999).
64. D.J. Brody, J.L. Pirkle, R.A. Kramer, K.M. Flegal, T.D. Matte, E.W. Gunter, D.C. Paschal, 'Blood Lead Levels in the US Population. Phase 1 of the Third National Health and Nutrition Examination Survey (NHANES III, 1988 to 1991)', *J. Am. Med. Assoc.*, **272**, 277–283 (1994).
65. J.L. Pirkle, D.J. Brody, E.W. Gunter, R.A. Kramer, D.C. Paschal, K.M. Flegal, T.D. Matte, 'The Decline in Blood Lead Levels in the United States. The National Health and Nutrition Examination Surveys (NHANES)', *J. Am. Med. Assoc.*, **272**, 284–291 (1994).
66. Anonymous, 'Update: Blood Lead Levels – United States, 1991–1994', *Morbid. Mortal. Week. Rep.*, **46**, 141–146 (1997).
67. D.C. Paschal, B.G. Ting, J.C. Morrow, J.L. Pirkle, R.J. Jackson, E.J. Sampson, D.T. Miller, K.L. Caldwell, 'Trace Metals in Urine of United States Residents: Reference Range Concentrations', *Environ. Res.*, **76**, 53–59 (1998).
68. C. Minoia, E. Sabbioni, R. Pietra, A. Ronchi, F. Poggio, A. Salvadeo, *Applications of Zeeman Graphite Furnace Atomic Absorption Spectrometry in the Chemical Laboratory and in Toxicology*, Pergamon Press, New York, 627–646, 1992.
69. P.C. D'Haese, L.V. Lamberts, L. Liang, F.L. Van de Vyver, M.E. De Broe, 'Elimination of Matrix and Spectral Interferences in the Measurement of Lead and Cadmium in Urine and Blood by Electrothermal Atomic Absorption Spectrometry with Deuterium Background Correction', *Clin. Chem.*, **37**, 1583–1588 (1991).
70. B. Welz, G. Schlemmer, J.R. Mudakavi, 'Palladium Nitrate–Magnesium Nitrate Modifier for Electrothermal Atomic Absorption Spectrometry. Part 5. Performance for the Determination of 21 Elements', *J. Anal. At. Spectrom.*, **7**, 1257–1271 (1992).
71. A.A. Brown, B.A. Milner, A. Taylor, 'Use of a Slotted Quartz Tube to Enhance the Sensitivity of Conventional Flame Atomic-absorption Spectrometry', *Analyst*, **110**, 501–505 (1985).
72. A.A. Brown, A. Taylor, 'Applications of a Slotted Quartz Tube and Flame Atomic-absorption Spectrometry to the Analysis of Biological Samples', *Analyst*, **110**, 579–582 (1985).
73. J.L. Burguera, M. Burguera, 'Determination of Cadmium in Human Urine by Extraction with Dithizone in a Flow-injection System', *Anal. Chim. Acta*, **153**, 207–212 (1983).
74. D.L. Tsalev, *Atomic Absorption Spectroscopy in Occupational and Environmental Health Practice*, CRC Press, Boca Raton, FL, Vol. III, 1995.
75. S. Tang, P.J. Parsons, W. Slavin, 'Effect of Acids and Modifiers on the Atomization of Aluminum in Electrothermal Atomic Absorption Spectrometry (ETAAS)', *J. Anal. At. Spectrom.*, **10**, 521–526 (1995).
76. M.R. Wills, S. Brown, J. Savory, 'Aluminum', in *Methods in Clinical Chemistry*, eds. A.J. Pesce, L.A. Kaplan, The C.V. Mosby Company, St. Louis, 330–335, 1987.
77. R.J. Shamberger, 'Selenium', in *Methods in Clinical Chemistry*, eds. A.J. Pesce, L.A. Kaplan, The C.V. Mosby Company, St Louis, 543–550, 1987.
78. M. Zhuang, R.M. Barnes, 'Determination of Trace Elements in Serum [and Liver] Using Inductively Coupled Plasma Atomic Spectroscopy with Hydride Generation and Chelating Resin Pre-concentration', *Appl. Spectrosc.*, **38**, 635–644 (1984).
79. G.R. Carnrick, D.C. Manning, W. Slavin, 'Determination of Selenium in Biological Materials with Platform Furnace Atomic-absorption Spectroscopy and Zeeman Background Correction', *Analyst*, **108**, 1297–1312 (1983).
80. D.C. Paschal, M.M. Kimberly, 'Automated Direct Determination of Selenium in Serum by Electrothermal

Atomic-absorption Spectroscopy', *At. Spectrosc.*, **7**, 75–78 (1986).
81. B.E. Jacobson, G. Lockitch, 'Direct Determination of Selenium in Serum by Graphite-furnace Atomic-absorption Spectrometry with Deuterium Background Correction and a Reduced Palladium Modifier: Age-specific Reference Ranges', *Clin. Chem.*, **34**, 709–714 (1988).
82. M.B. Knowles, K.G. Brodie, 'Determination of Selenium in Blood by Zeeman Graphite-furnace Atomic-absorption Spectrometry Using a Palladium – Ascorbic Acid Chemical Modifier', *J. Anal. At. Spectrom.*, **3**, 511–516 (1988).
83. B. Welz, G. Schlemmer, J.R. Mudakavi, 'Palladium Nitrate – Magnesium Nitrate Modifier for Graphite-furnace Atomic-absorption Spectrometry. II. Determination of Arsenic, Cadmium, Copper, Manganese, Lead, Antimony, Selenium and Thallium in Water', *J. Anal. At. Spectrom.*, **3**, 695–701 (1988).
84. B. Welz, M. Melcher, G. Schlemmer, 'Determination of Selenium in Human Blood Serum. Comparison of Two Atomic-absorption Spectrometric Procedures', *Fresenius' J. Anal. Chem.*, **316**, 271–276 (1983).
85. T.M. Sheehan, D.J. Halls, 'Measurement of Selenium in Clinical Specimens', *Ann Clin Biochem*, **36**, 301–315 (1999).
86. B. Welz, M. Melcher, J. Neve, 'Determination of Selenium in Human Body Fluids by Hydride-generation Atomic-absorption Spectrometry. Optimization of Sample Decomposition', *Anal. Chim. Acta*, **165**, 131–140 (1984).
87. M. Navarro, H. Lopez, M.C. Lopez, V. Perez, 'Determination of Selenium in Urine by Hydride Generation Atomic-absorption Spectrometry', *J. AOAC. Int.*, **79**, 773–776 (1996).
88. International Union of Pure and Applied Chemistry, 'Inter-laboratory Trial on the Determination of Selenium in Lyophilized Human Serum, Blood and Urine Using Hydride-generation Atomic-absorption Spectrometry', *Pure Appl. Chem.*, **59**, 927–936 (1987).
89. B. Welz, M. Schubert-Jacobs, 'Evaluation of a Flow-injection System and Optimization of Parameters for Hydride-generation Atomic-absorption Spectrometry', *At. Spectrosc.*, **12**, 91–104 (1991).
90. B. Welz, P. Stauss, 'Interferences from Hydride-forming Elements on Selenium in Hydride-generation Atomic-absorption Spectrometry with a Heated-quartz-tube Atomizer', *Spectrochim. Acta, Part B*, **48**, 951–976 (1993).
91. B. Sampson, 'Determination of Selenium in Serum by Electrothermal-atomization Atomic-absorption Spectrometry with Deuterium-arc Background Correction', *J. Anal. At. Spectrom.*, **2**, 447–450 (1987).
92. C. Veillon, 'Trace Element Analysis of Biological Samples. Problems and Precautions', *Anal. Chem.*, **58**, 851A–852A, 854A (1986).
93. C. Veillon, K.Y. Patterson, N.A. Bryden, 'Chromium in Urine as Measured by Atomic-absorption Spectrometry', *Clin. Chem.*, **28**, 2309–2311 (1982).
94. C. Veillon, K.Y. Patterson, N.A. Bryden, 'Direct Determination of Chromium in Human Urine by Electrothermal Atomic-absorption Spectrometry', *Anal. Chim. Acta*, **136**, 233–241 (1982).
95. J.L. Burguera, M. Burguera, C. Rondon, L. Rodriguez, P. Carrero, Y. Petit de Peña, E. Burguera, 'Determination of Chromium in Urine by Electrothermal Atomic-absorption Spectrometry Using Different Modifiers', *J. Anal. At. Spectrom.*, **14**, 821–825 (1999).
96. D.C. Paschal, M.M. Kimberly, G.G. Bailey, 'Determination of Urinary Arsenic by Electrothermal Atomic-absorption Spectrometry with the L'vov Platform and Matrix Modification', *Anal. Chim. Acta*, **181**, 179–186 (1986).
97. K.S. Subramanian, 'Determination of Arsenic in Urine by Graphite-platform In-furnace Atomic-absorption Spectrometry', *Can. J. Spectrosc.*, **33**, 173–181 (1988).
98. D.E. Nixon, G.V. Mussmann, S.J. Eckdahl, T.P. Moyer, 'Total Arsenic in Urine: Palladium – Persulphate vs. Nickel as a Matrix Modifier for Graphite-furnace Atomic-absorption Spectrophotometry', *Clin. Chem.*, **37**, 1575–1579 (1991).
99. L. Magos, 'Selective Atomic-absorption Determination of Inorganic Mercury and Methylmercury in Undigested Biological Samples', *Analyst*, **96**, 847–853 (1971).
100. L. Magos, T.W. Clarkson, 'Atomic Absorption Determination of Total, Inorganic, and Organic Mercury in Blood', *J. Assoc. Off. Anal. Chem.*, **55**, 966–971 (1972).
101. D. Littlejohn, G.S. Fell, J.M. Ottaway, 'Modified Determination of Total and Inorganic Mercury in Urine by Cold Vapor Atomic Absorption Spectrometry', *Clin. Chem.*, **22**, 1719–1723 (1976).
102. J. Toffaletti, J. Savory, 'Use of Sodium Borohydride for Determination of Total Mercury in Urine by Atomic Absorption Spectrometry', *Anal. Chem.*, **47**, 2091–2095 (1975).
103. D.C. Sharma, P.S. Davis, 'Direct Determination of Mercury in Blood by Sodium Tetrahydroborate Reduction and Atomic-absorption Spectrophotometry', *Clin. Chem.*, **25**, 769–772 (1979).
104. D.L. Tsalev, M. Sperling, B. Welz, 'Online Microwave Sample Pre-treatment for Hydride-generation and Cold-vapour Atomic-absorption Spectrometry. II. Chemistry and Applications', *Analyst*, **117**, 1735–1741 (1992).
105. T. Guo, J. Baasner, 'Online Microwave Sample Pretreatment for the Determination of Mercury in Blood by Flow Injection Cold Vapour Atomic-absorption Spectrometry', *Talanta*, **40**, 1927–1936 (1993).
106. D.E. Nixon, G.V. Mussmann, T.P. Moyer, 'Inorganic, Organic, and Total Mercury in Blood and Urine: Cold Vapour Analysis with Automated Flow-injection Sample Delivery', *J. Anal. Toxicol.*, **20**, 17–22 (1996).

107. H.P. Chen, D.C. Paschal, D.T. Miller, J.C. Morrow, 'Determination of Total and Inorganic Mercury in Whole Blood by On-line Digestion with Flow Injection', *At. Spectrosc.*, **19**, 176–179 (1998).
108. D.L. Tsalev, Z.K. Zaprianov, *Atomic Absorption Spectrometry in Occupational and Environmental Health Practice*, CRC Press, Boca Raton, FL, 1–252, Vol. 1, 1984.
109. D.L. Tsalev, *Atomic Absorption Spectroscopy in Occupational and Environmental Health Practice*, CRC Press, Boca Raton, FL, 1–295, Vol. II, 1984.
110. H.T. Delves, 'Atomic Absorption Spectroscopy in Clinical Analysis. [Review.]', *Ann. Clin. Biochem.*, **24**, 529–551 (1987).
111. A. Taylor, 'Clinical Applications of Flame Techniques', in *Atomic Absorption Spectrometry. Theory, Design and Applications*, ed. S.J. Haswell, Elsevier, Amsterdam, 359–379, 1991.
112. A.A. Brown, D.J. Halls, A. Taylor, 'Atomic Spectrometry Update: Clinical Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **1**, 29R–44R (1986).
113. A.A. Brown, D.J. Halls, A. Taylor, 'Atomic Spectrometry Update: Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **2**, 43R–77R (1987).
114. A.A. Brown, D.J. Halls, A. Taylor, 'Atomic Spectrometry Update: Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **3**, 45R–77R (1988).
115. A.A. Brown, D.J. Halls, A. Taylor, 'Atomic Spectrometry Update – Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **4**, 47R–87R (1989).
116. H.M. Crews, D.J. Halls, A. Taylor, 'Atomic Spectrometry Update – Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **5**, 75R–149R (1990).
117. S. Branch, H.M. Crews, D.J. Halls, A. Taylor, 'Atomic Spectrometry Update – Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **6**, 69R–107R (1991).
118. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, 'Atomic Spectrometry Update – Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **7**, 67R–154R (1992).
119. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, 'Atomic Spectrometry Update – Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **8**, 79R–136R (1993).
120. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **9**, 87R–146R (1994).
121. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **10**, 61R–112R (1995).
122. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **11**, 103R–186R (1996).
123. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, L.M.W. Owen, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **12**, 119R–221R (1997).
124. A. Taylor, S. Branch, D.J. Halls, L.M.W. Owen, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **14**, 717–781 (1999).
125. D.C. Paschal, 'Biological Monitoring with Atomic Spectroscopy', *Spectrochim. Acta, Part B*, **44B**, 1229–1236 (1989).
126. N.W. Alcock, 'Flame, Flameless and Plasma Spectroscopy', *Anal. Chem.*, **67**, 503R–507R (1995).
127. J. Savory, M.M. Herman, 'Advances in Instrumental Methods for the Measurement and Speciation of Trace Metals', *Ann. Clin. & Lab. Sci.*, **29**, 118–126 (1999).
128. A. Alimonti, E. Coni, S. Caroli, E. Sabbioni, G.E. Nicolaou, R. Pietra, 'Critical Comparison of Performances of Inductively Coupled Plasma Atomic-emission Spectrometry and Neutron-activation Analysis for the Determination of Elements in Human Lungs', *J. Anal. At. Spectrom.*, **4**, 577–580 (1989).
129. L. Bussiere, J. Dumont, J. Hubert, 'Direct Determination of Calcium, Copper, Iron, Magnesium, Manganese and Zinc in Amniotic Fluid Samples Using Inductively Coupled Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **224**, 73–81 (1989).
130. C. Coker, B. Ersoz, S. Habif, N. Cetiner, Y. Gultekin, 'Elemental Analysis of Serum by Inductively Coupled Plasma Atomic Emission Spectroscopy in Comparison to Atomic Absorption Spectroscopy', *Turk. J. Med. Sci.*, **26**, 553–557 (1996).
131. T.D.B. Lyon, C. Cunningham, D.J. Halls, J. Gibbons, A. Keating, G.S. Fell, 'Determination of Aluminium in Serum, Dialysate Fluid and Water by Inductively Coupled Plasma Optical Emission Spectrometry', *Ann. Clin. Biochem.*, **32**, 160–166 (1995).
132. B. Hu, Z.C. Jiang, Y. Zeng, 'Direct Determination of Aluminium in Biological Materials by Electrothermal Vaporization Inductively Coupled Plasma Atomic-emission Spectrometry with Polytetrafluoroethylene as Chemical Modifier', *Anal. Chim. Acta*, **296**, 213–218 (1994).
133. M.W. Wensing, D.Y. Liu, B.W. Smith, J.D. Winefordner, 'Determination of Lead in Whole Blood Using a Capacitively Coupled Microwave Plasma Atomic-emission Spectrometer', *Anal. Chim. Acta*, **299**, 1–7 (1994).
134. J. Alvarado, P. Cavalli, N. Omenetto, G. Rossi, J.M. Ottaway, D. Littlejohn, 'Direct Determination of Lead in Whole Blood Using Electrothermal-vaporization Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Lett.*, **22**, 2975–2984 (1989).

135. S. Recknagel, P. Braetter, A. Tomiak, U. Roesick, 'Determination of Selenium in Blood Serum by ICP OES [AES] Including an Online Wet-digestion and Selenium Hydride Formation Procedure', *Fresenius' J. Anal. Chem.*, **346**, 833–836 (1993).
136. T.D.B. Lyon, G.S. Fell, R.C. Hutton, A.N. Eaton, 'Evaluation of Inductively Coupled Argon-plasma Mass Spectrometry (ICP MS) for Simultaneous Multi-element Trace Analysis in Clinical Chemistry', *J. Anal. At. Spectrom.*, **3**, 265–271 (1988).
137. M.A. Vaughan, A.D. Baines, D.M. Templeton, 'Multi-element Analysis of Biological Samples by Inductively Coupled Plasma-mass Spectrometry. II. Rapid Survey Method for Profiling Trace Elements in Body Fluids', *Clin. Chem.*, **37**, 210–215 (1991).
138. R.M. Barnes, 'Advances in Inductively Coupled Plasma Mass Spectrometry: Human Nutrition and Toxicology', *Anal. Chim. Acta*, **283**, 115–130 (1993).
139. H.E. Taylor, R.A. Huff, A. Montaser, 'Novel Applications of ICPMS', in *Inductively Coupled Plasma Mass Spectrometry*, ed. A. Montaser, Wiley-VCH, New York, 681–807, 1998.
140. E. Pruszkowski, K. Neubauer, R. Thomas, 'An Overview of Clinical Applications by Inductively Coupled Plasma Mass Spectrometry', *At. Spectrosc.*, **19**, 111–115 (1998).
141. D.E. Nixon, T.P. Moyer, 'Routine Clinical Determination of Lead, Arsenic, Cadmium and Thallium in Urine and Whole Blood by Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta, Part B*, **51B**, 13–25 (1996).
142. A. Alimonti, F. Petrucci, B. Santucci, A. Cristaudo, S. Caroli, 'Determination of Chromium and Nickel in Human Blood by Means of Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **306**, 35–41 (1995).
143. K. Nuttall, W. Gordon, K.O. Ash, 'Inductively Coupled Plasma Mass Spectrometry for Trace Element Analysis in the Clinical Laboratory', *Ann. Clin. Lab. Sci.*, **25**, 264–271 (1995).
144. L. Moens, 'Applications of Mass Spectrometry in the Trace Elements Analysis of Biological Materials', *Fresenius' J. Anal. Chem.*, **359**, 309–316 (1997).
145. F.H. Ko, M.H. Yang, 'Online Removal of Interferences via Anion-exchange Column Separation for the Determination of Germanium, Arsenic and Selenium in Biological Samples by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 413–420 (1996).
146. L. Moens, P. Verrept, R. Dams, U. Greb, G. Jung, B. Laser, 'New High-resolution Inductively Coupled Plasma Mass Spectrometry Technology Applied for the Determination of Vanadium, Iron, Copper, Zinc and Silver in Human Serum', *J. Anal. At. Spectrom.*, **9**, 1075–1078 (1994).
147. H.T. Delves, M.J. Campbell, 'Measurements of Total Lead Concentrations and of Lead Isotope Ratios in Whole Blood by Use of Inductively Coupled Plasma Source Mass Spectrometry', *J. Anal. At. Spectrom.*, **3**, 343–348 (1988).
148. M.J. Campbell, H.T. Delves, 'Accurate and Precise Determination of Lead Isotope Ratios in Clinical and Environmental Samples Using Inductively Coupled Plasma Source Mass Spectrometry', *J. Anal. At. Spectrom.*, **4**, 235–236 (1989).
149. D.C. Paschal, K.L. Caldwell, B.G. Ting, 'Determination of Lead in Whole Blood Using Inductively Coupled Argon Plasma Mass Spectrometry with Isotope Dilution', *J. Anal. At. Spectrom.*, **10**, 367–370 (1995).
150. K.E. Murphy, P.J. Paulsen, 'The Determination of Lead in Blood Using Isotope Dilution Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **352**, 203–208 (1995).
151. S. Brunk, 'Simplified Method for Serum Aluminium by ICP MS', *At. Spectrosc.*, **15**, 145–147 (1994).
152. M.F. Huang, S.J. Jiang, C.J. Hwang, 'Determination of Arsenic in Environmental and Biological Samples by Flow Injection Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **10**, 31–35 (1995).
153. M.P. Rayman, F.R. Abou Shakra, N.I. Ward, 'Determination of Selenium in Blood Serum by Hydride Generation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 61–68 (1996).
154. A. Makarov, J. Szpunar, 'The Coupling of Size-exclusion HPLC with ICPMS in Bio-inorganic Analysis', *Analusis*, **26**, M44–M48 (1998).
155. K.L. Yang, S.J. Jiang, 'Determination of Selenium Compounds in Urine Samples by Liquid Chromatography/Inductively Coupled Plasma Mass Spectrometry with an Ultrasonic Nebulizer', *Anal. Chim. Acta*, **307**, 109–115 (1995).
156. K.T. Suzuki, M. Itoh, M. Ohmichi, 'Detection of Selenium-containing Biological Constituents by High-performance Liquid Chromatography–Plasma Source Mass Spectrometry', *J. Chromatogr. B*, **666**, 13–19 (1995).
157. X. Zhang, R. Cornelis, J. de Kimpe, L. Mees, 'Arsenic Speciation on Serum of Uraemic Patients Based on Liquid Chromatography with Hydride Generation Atomic Absorption Spectrometry and On-line UV Photo-oxidation Digestion', *Anal. Chim. Acta*, **319**, 177–185 (1996).
158. R. Cornelis, X.R. Zhang, L. Mees, J.M. Christensen, K. Byrialsen, C. Dyrschel, 'Speciation Measurements by HPLC-hydride Generation AAS of Dimethylarsinic Acid and Arsenobetaine in Three Candidate Lyophilized Urine Reference Materials', *Analyst*, **123**, 2883–2886 (1998).
159. X. Zhang, R. Cornelis, J. de Kimpe, L. Mees, 'Speciation of Toxicologically Important Arsenic Species in Human Serum by Liquid Chromatography–Hydride Generation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1075–1079 (1996).

160. H. Ding, J. Wang, J.G. Dorsey, J.A. Caruso, 'Arsenic Speciation by Micellar Liquid Chromatography with Inductively Coupled Plasma Mass-spectrometric Detection', *J. Chromatogr.*, **694**, 425–431 (1995).
161. B. Aizpun, M.L. Fernandez, E. Blanco, A. Sanz Medel, 'Speciation of Inorganic Mercury(II) and Methylmercury by Vesicle-mediated High-performance Liquid Chromatography Coupled to Cold Vapour Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **9**, 1279–1284 (1994).
162. X.F. Yin, W. Frech, E. Hoffmann, C. Luedke, J. Skole, 'Mercury Speciation by Coupling Cold Vapor Atomic Absorption Spectrometry with Flow Injection Online Preconcentration and Liquid Chromatographic Separation', *Fresenius' J. Anal. Chem.*, **361**, 761–766 (1998).
163. US Department of Health and Human Services, 'Medicare, Medicaid, and CLIA Programs; Regulations Implementing the Clinical Laboratory Improvement Ammendments of 1988 (CLIA). Final Rule', *Fed. Regist.*, **57**, 7002–7186 (1992).
164. National Committee for Clinical Laboratory Standards, 'Terminology and Definitions for Use in NCCLS Documents – Third Edition; Proposed Standard', NCCLS Publication NRSCL8-P3, Wayne, PA, 1–82, 1996.
165. A. Taylor, R.J. Briggs, 'External Quality Assessment Schemes and Improvements in the Measurement of Trace Elements in Biological Fluids', *J. Anal. At. Spectrom.*, **10**, 1033–1037 (1995).
166. A. Taylor, R.J. Briggs, 'An External Quality Assessment Scheme for Trace Element in Biological Fluids', *J. Anal. At. Spectrom.*, **1**, 391–395 (1986).
167. P.J. Parsons, 'Monitoring Human Exposure to Lead: An Assessment of Current Laboratory Performance for the Determination of Blood Lead', *Environ. Res.*, **57**(2), 149–162 (1992).
168. N.V. Stanton, 'Blood Lead Proficiency Testing: Overview of the Federally Sponsored Program in the US', *J. Int. Fed. Clin. Chem.*, **5**, 158–161 (1998).
169. D.G. Bullock, N.J. Smith, T.P. Whitehead, 'External Quality Assessment of Assays of Lead in Blood', *Clin. Chem.*, **32**, 1884–1889 (1986).
170. S.D. Rasberry, 'Biological Reference Materials from the US National Bureau of Standards: An Update', *Fresenius' J. Anal. Chem.*, **326**, 609–612 (1987).
171. S.D. Rasberry, 'NBS Activities in Biological Reference Materials', *Fresenius' J. Anal. Chem.*, **332**, 528–532 (1988).
172. C. Veillon, S.A. Lewis, K.Y. Patterson, W.R. Wolf, J.M. Harnly, J. Versieck, L. Vanballenberghe, R. Cornelis, T.C. O'Haver, 'Characterization of a Bovine Serum Reference Material for Major, Minor and Trace Elements', *Anal. Chem.*, **57**, 2106–2109 (1985).
173. C. Veillon, S.A. Lewis, K.Y. Patterson, 'Bovine Serum Reference Material for Major, Minor and Trace Elements', *Acta Pharmacol. Toxicol.*, **59**, 573–576 (1986).
174. W.R. Külpmann, R. Buchholz, C. Dyrssen, D. Ruschke, P. Maibaum, 'Comparison of Reference Method Values for Calcium, Lithium and Magnesium with Method-dependent Assigned Values', *J. Clin. Chem. Clin. Biochem.*, **27**, 631–637 (1989).
175. C. Veillon, K.Y. Patterson, 'Trace Elements in a Commercial Freeze-dried Human Urine Reference Material', *Analyst*, **121**, 983–986 (1996).
176. A. Taylor, 'The Use of Internal Quality Control Materials for the Preparation and Maintenance of Reliable Methods for the Measurement of Lead in Blood', *Fresenius' J. Anal. Chem.*, **332**, 732–735 (1988).
177. H.T. Delves, 'Use Reference Samples Rather than Reference Methods', *Anal. Proc.*, **21**, 391–394 (1984).
178. A. Taylor, 'Reference Materials and Analytical Standards to Stimulate Improved Laboratory Performance: Experience from the External Quality Assessment Scheme for Trace Elements in Biological Samples', *Mikrochim. Acta*, **123**, 251–260 (1996).
179. A. Taylor, 'Reference Materials for Measurement of Al in Biological Samples', *Fresenius' J. Anal. Chem.*, **332**, 616–619 (1988).
180. H.T. Delves, B. Suchak, C.S. Fellows, 'The Determination of Aluminum in Foods and Biological Materials', in *Aluminum in Food and the Environment [Conference Proceedings]*, eds. R. Massey, D. Taylor, Royal Society of Chemistry, London, 52–67, 1989.
181. F.J. Krug, M.M. Silva, P.V. Oliveira, J.A. Nobrega, 'Determination of Lead in Blood by Tungsten-coil Electrothermal Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **50**, 1469–1474 (1995).
182. P.J. Parsons, H. Qiao, K.M. Aldous, E. Mills, W. Slavin, 'A Low-cost Tungsten-filament Atomizer for Measuring Lead in Blood by Atomic-absorption Spectrometry', *Spectrochim. Acta, Part B*, **50B**, 1475–1480 (1995).
183. C.L. Sanford, S.E. Thomas, B.T. Jones, 'Portable, Battery-powered, Tungsten-coil Atomic-absorption Spectrometer for Lead Determinations', *Appl. Spectrosc.*, **50**, 174–181 (1996).
184. J.D. Batchelor, S.E. Thomas, B.T. Jones, 'Determination of Cadmium with a Portable, Battery-powered Tungsten Coil Atomic-absorption Spectrometer', *Appl. Spectrosc.*, **52**, 1086–1091 (1998).

Automation in the Clinical Laboratory

Richard A. McPherson
Medical College of Virginia Hospitals, Virginia Commonwealth University, Richmond, USA

1 Introduction **1**
2 Clinical Laboratory Processes **2**
2.1 Preanalytical Processes 2
2.2 Analytical Processes 4
2.3 Postanalytical Processes 4
2.4 Turn-around-times 5
2.5 Likely Targets for Automation 6
3 Control of Laboratory Analyses **6**
3.1 Laboratory Information Systems 7
3.2 Laboratory Automation Systems 8
3.3 Processes in a Totally Automated Laboratory 8
3.4 Laboratory Workstations 11
4 Workstations and Automation **11**
4.1 Trends in the Consolidation of Workstations 11
4.2 Characteristics of Instruments that can be Used in Total Laboratory Automation 11
5 Accomplishments in Total Laboratory Automation **12**
6 Cost Justification of Laboratory Automation **13**
7 Expected Benefits from Total Laboratory Automation **13**
8 Problems and Difficulties from Total Laboratory Automation **14**
9 Final Assessment of Total Laboratory Automation **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **14**

Total laboratory automation (TLA) is an approach to laboratory instrumentation that replaces human operators with robotic devices for the performance of highly repetitive tasks in the preparation and transport of specimens. TLA will consolidate the control and oversight for multiple different analytical instruments to a smaller number of operators, thus reducing the costs of labor in laboratory testing. Benefits of TLA include reduction of human errors in specimen handling and improvements in overall process control, including faster turn-around-times (TAT) from specimen collection to test result reporting. Possible detriments include very high costs for equipment installation and disruption of workforces trained in existing technologies. Most clinical laboratories will soon face a decision about implementing TLA (or portions of it) to reduce operating costs in order to offset falling reimbursements for medical services.

1 INTRODUCTION

Clinical diagnostic laboratories are about to undergo a major change in which the performance of high-volume assays will be controlled completely automatically. The essence of this change is automation of several highly repetitive tasks that are now usually done manually in the preanalytical phase of processing specimens and presenting them to instruments.[1–3] Other aspects of the new automation will include tracking and storing specimens so that they may be retrieved immediately for repeat (e.g. at dilution) or additional testing and ultimately discarded in a timely manner. The preanalytical tasks present immediate targets for automation; they include centrifugation, aliquot formation, transport to analyzers, and various judgments about specimen quantity and quality. The assays most likely to be involved include those for which automated analyzers already exist, such as clinical chemistry, hematology, coagulation, immunochemistry, and perhaps urinalysis. Highly repetitive tasks of preparing and presenting specimens for these analyses may be performed by robotic devices without direct intervention by human operators. Instead, the role of the operator should become one of monitoring quality control for a wide range of different assays simultaneously and to trouble-shoot problems that arise on instruments. A move toward such wide control over assays is actually an extension of manufacturing trends from the last few decades in which multiple assays have been installed on individual analytical instruments. Thus, linking multiple different instruments through a robotic processing and transport system completes the trend of multichannel analyzer development; the end result is termed "TLA".

Implementation of TLA is expected to displace a large number of both highly skilled technologists and lesser skilled technicians, leading to disruption in the careers of many clinical laboratory scientists. Those who remain will either work in other disciplines that are not

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

yet automated, or adapt to demands for new skills and competencies. The expected benefits from TLA include reduction in operating costs (primarily labor), reduction in errors, faster TAT and overall improvement in process control. The detriments of installing TLA are very high costs of acquisition and maintenance, displacement of workers, and the need to retrain employees to work in a new environment.

This presentation begins with a discussion of clinical laboratory processes and how their TATs affect delivery of healthcare. Emphasis is given to how these processes potentially can be controlled in a TLA laboratory. This presentation is organized from the viewpoint of a laboratory director who will make decisions about how and when to implement TLA based on progress in the industry and the critical need to cut operating expenses.

2 CLINICAL LABORATORY PROCESSES

The sequence of processes in the analysis of clinical laboratory specimens is traditionally divided into preanalytical, analytical, and postanalytical phases,[4] each of which has multiple component activities (Figure 1). It is useful to include another earlier conceptual phase as well, because the requesting physician typically thinks of the entire process beginning when specific tests are first entertained to assist in diagnosis and medical management, even before test orders are formally placed. TATs are based on these individual phases; in practice, they are conveniently measured from the time that an order is entered into a hospital information system (HIS) or a laboratory information system (LIS) until the time when the results are verified and available to the physician through a computer. Time of receipt and acknowledgement of those results by the physician usually takes place at a later time that is not easily captured for calculating the total TAT.

2.1 Preanalytical Processes

In outpatient clinics, orders to collect specimens and perform specific tests are usually written on laboratory requisition forms by the requesting physician. Such orders are typically transcribed into a central computer of the medical center, which will assign unique specimen identification numbers, record specimen status as it reaches the laboratory, report test results to the physician, and keep those results in a file along with all other results on that patient for long-term information retrieval. Ordering of tests on hospitalized patients (inpatients) is typically done through direct entry into a computer system by the physician.

A work list of patients to be drawn by phlebotomists is generated by the computer system for scheduled (routine) and immediate (stat) collections from inpatients. Outpatients are generally handled individually in the order that they present in person to a phlebotomy station, thus obviating the need for comprehensive work lists in outpatient clinics.

Patient identification is accomplished by comparing the written order information with an inpatient's identification bracelet and with active verbal responses that inpatients and outpatients make to standard questions about their identity asked by the phlebotomist. As trivial or mundane as this task may seem, it is crucially important because all subsequent activities rely completely on

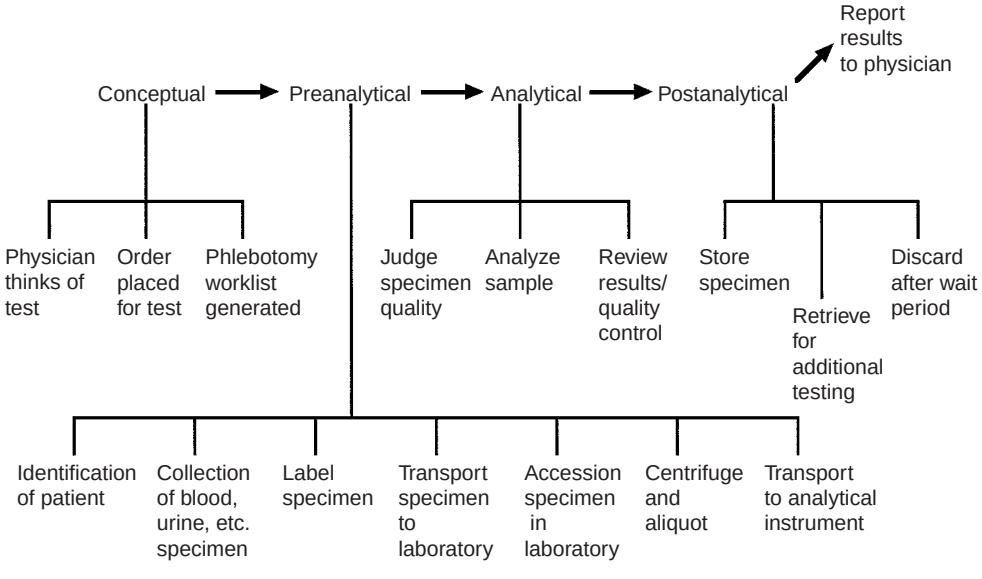

Figure 1 Components of activity in clinical laboratory processes.

having accurately labeled specimens. There is no other failsafe mechanism to establish the identity of a specimen after it has left the patient's side; in fact, most so-called ''laboratory errors'' can be traced to mislabeled specimens, usually collected by nonlaboratory staff.

After positive patient identification, the specimen is collected. Blood samples are collected into standard vacuum tubes containing various anticoagulants to provide whole blood and plasma or clot-promoting powder and a separating gel to yield serum. Urine samples are generally collected into cups privately by the patient (except in medical–legal cases for drugs of abuse). Other body fluids (cerebrospinal fluid, peritoneal/ascites fluid, pleural fluid, joint fluid, gastric aspirate, etc.) are almost always collected by a physician working in a sterile field, but labeled by another person, sometimes not even at the collection site. Other specimens consist of wound swabs, sputum, skin scrapings, and other sources for microbiological culture.

The label should contain the patient's name and identification number (e.g. medical record number, social security number, etc.). Initials of the person who collected the specimen and the time of collection are also hand written on the label. More modern labeling consists of encrypting all this information into a bar-coded label generated by a central computer system. Ideally, the entire identification process could be done with portable bar code readers interfaced to the computer. The phlebotomist would identify and automatically record:

- the patient by scanning the bar code on the patient's wrist identification band;
- the phlebotomist him/herself by scanning the bar code on an employee identification badge;
- the time of collection.

The label should be attached with the bar code parallel to the length of the tube (as opposed to being wrapped around the tube) for proper presentation to various bar code readers in subsequent steps.

Transportation of specimens to the laboratory is guided in part by the status of ordering. Stat specimens should be transported quickly (usually individually, as soon as they are collected); routine specimens may be held for batch transportation. Couriers can be used for both levels of these priorities; however, in a busy medical center the availability of couriers is not continuous and frequently is not sufficient for the need. Automated transportation systems are very useful for providing this service on demand around the clock. Such a system consists of pneumatic tubes (4–6 inches in diameter) that run between sending and receiving stations. Carriers fit flush within the tubes, allowing them to be propelled by air pressure from blowers. Complicated systems use switching stations as nodes to connect tens of stations within a building, and even extending to adjacent buildings up to a few thousand feet away. Transit times are usually 30–90 s or more; owing to acceleration and abrupt deceleration at switching nodes and upon landing at arrival, specimen integrity may be compromised and glass containers may break. Consequently, foam rubber padding and other protective wrap is needed. Pneumatic tube systems are controlled by their own computers, which can plan and track the course that a carrier follows; transit time from start to finish can be monitored for signs of system deterioration (e.g. air leak, increased resistance to movement, etc.) over long periods of time. One feature that is not incorporated yet into pneumatic systems is the identification of which specimens are transported in which particular carriers. Furthermore, emptying carriers is a manual task; some specimens may not be recognized by the human operator and so they are left in the carrier after it arrives in the laboratory. This lack of mechanical interface between pneumatic tube systems and analytical areas is a potential target for future automation.

Specimen accessioning is the process of confirming receipt of the correct specimen with label and test requisition in the laboratory. The activities usually include manual steps of matching specimen to requisition and entering data elements into an information system (either HIS or LIS) that uniquely identify that specimen for that patient. The next immediate step is to prepare the specimen for analysis. Whole-blood samples for hematological analysis (e.g. complete blood count (CBC)) are the simplest to prepare by inverting the closed collection tubes several times to mix the erythrocytes, leukocytes, and platelets thoroughly within the plasma. This preparation counters the settling of cellular elements that occurs when the sample sits in a fixed position. Most other common, high-volume specimens are intended for plasma or serum analysis; both are prepared by centrifugation. Plasma from blood collected into anticoagulants can be centrifuged as soon as the specimens are collected; common applications include the coagulation tests of prothrombin time (PT) and activated partial thromboplastin time (APTT) for samples collected into sodium citrate; also electrolytes and other chemical analytes such as glucose, urea, creatinine, etc. needed on a stat basis are collected into heparin.

Preparation of serum requires sufficient time for the blood to clot thoroughly prior to centrifugation. Usually the period from time of collection to receipt in the laboratory is sufficient for clotting to occur, unless the patient has been anticoagulated, in which case fibrin strands may continue to form in the supernatant after separation of serum from the clot. Fibrin strands from incompletely clotted serum can seriously interfere with the aspiration of a sample into analytical instruments,

which then may need to be cleaned before analyzing subsequent samples. This problem is usually prevented through visual inspection of all samples by an operator screening the specimens for problems. Collection tubes for serum samples commonly contain a separating gel that migrates to a position between the dense cellular clot and the lighter serum. These serum separator tubes help to preserve the integrity of many analytes in serum samples after centrifugation; when the serum is no longer in contact with cellular elements, constituents such as glucose, potassium, lactate dehydrogenase, etc. can remain stable for several hours while waiting analysis.

The original specimen tube may be placed directly onto an instrument for aspiration of the sample to analyze. Alternatively, an aliquot of serum is taken from the original draw tube and placed into a secondary tube that is placed onto the instrument; the secondary tube also must be labeled uniquely to identify it, usually with a bar code that is read by the instrument. When different types of tests are ordered on a single specimen, it is also common practice to use separate aliquots for analysis on different instruments.

Processed samples and their aliquots must be transported to the analytical instruments. Stat specimens may be carried individually as they arrive; specimens for routine testing may be taken to the instruments in batches arranged manually in test-tube racks or in analyzer-specific carousels for direct placement onto instruments. This transportation step is also one of the targets for automation of repetitive manual tasks.

2.2 Analytical Processes

Immediately prior to analysis, an operator typically judges the adequacy and quality of each specimen by visual inspection for volume, hemolysis, icterus, and lipemia. Specimens that are unacceptable or problematic for these reasons may be analyzed anyway, and the results reported along with a comment to let the physician interpret whether they are valid or not; or they may be tested by a different methodology that has less potential for interference. The latter alternative usually requires that a specimen be handled off-line along a nonautomated pathway. Sometimes this problem-searching step is done on an analyzer by measuring optical absorbance in the sample at wavelengths where hemoglobin and bilirubin have absorption maxima and where light scattering occurs.

Most modern laboratory equipment for high-volume testing has built-in mechanisms for aspirating samples from the specimen tubes placed onto an instrument's carousel. Thus, once the instrument has been primed for operation, it automatically measures analytes in the specimen as directed through the instrument's own computer. The desired tests may be entered manually into that computer by a technologist, or they may be transmitted directly from the LIS through an electronic interface to the instrument. The period of time over which a previously processed specimen resides at an instrument for sampling and analysis (i.e. instrument dwell time) can be as short as 30–90 s; test results then may be sent electronically from the instrument's computer to the LIS.

Completed test results typically must pass some level of review before they are verified, at which point they are made available for reporting to the physician. The first criterion for accepting patient test results is whether quality control results are within acceptable limits using repeatedly analyzed sample material (usually commercially prepared). The general framework for accepting quality control results has been formalized by Westgard in terms of how many standard deviations away from the mean the present value lies and what the last several results were. In this manner, both abrupt changes in analytical performance can be detected as well as trends that may develop slowly. If quality control results are acceptable, then the entire batch of patient specimens can be assumed to have been tested while the instrument was functioning properly.

The second criterion for accepting patient results is sometimes called a ''delta check''. In general terms, the delta check is a comparison of present results with previous results on the same patient. If analyte values have not changed by more than a defined percentage, then the specimen is judged to be from the correct (i.e. the same) patient, and the results are accepted for verification. These result review steps may be done by a human operator, or they may be done automatically by the LIS.

2.3 Postanalytical Processes

After verification, test results are immediately available to the ordering physician through the LIS or HIS if the physician has access to such a computer. Stat results are typically printed out by the computer at the patient's location in a hospital ward or emergency room as soon as they are verified; stat results also may be reported by telephone to patient care locations not supported by computer. Routine results are reported by some slower mechanism, such as paper copy delivered by a courier the next day. One other category of results is panic or action values. These results are so abnormal that the patient may be in danger of dying from a chemical imbalance in a short time (e.g. very high or very low serum potassium; very high levels of therapeutic drugs that are toxic); the laboratory is responsible for recognizing panic values and reporting them directly to the physician immediately, whether the test was ordered stat or routine. Information systems play a major role in keeping track of these specimen priorities,

in following the appropriate reporting pathway, and in documenting reporting actions.

Specimen tubes are stored after testing for a few days to allow physicians to request additional tests or for repeat testing to confirm previous results. Specimens must be covered or recapped and then stored at refrigerated temperature. They should be stored in a sequential arrangement within each day's specimens to allow timely retrieval when additional testing is requested. Specimens are stored for several days and then are discarded as biohazardous waste. Although these tasks of storage, retrieval, and discarding are simple actions in principle, large numbers of specimens can complicate the whole process enormously.

2.4 Turn-around-times

The "total TAT" is generally defined as the time from specimen collection (usually time of blood draw) to the time that test results are available through a computer system (Figure 2). This conventional definition of total TAT is commonly employed; however, physicians may perceive the TAT differently, namely from the time that the need for a diagnostic test is first considered on a particular patient until the physician actually receives a copy of the results in hand and can use them to make a medical management decision. The specific activities delineated in Figure 1 are conveniently denoted to belong to within-lab TAT, pre-lab time, or some other more distant actions perhaps relating to particular office or medical unit practice.

Clinical laboratories have greatest control of processes that occur within their own walls, and so naturally they can best monitor within-laboratory TATs. This period runs from receipt of the specimen at the laboratory with appropriate order through all processing and testing to the time that the result is electronically verified and is available on a computer system or is otherwise reported directly to a physician or medical team member (e.g. by telephone, facsimile transmission, paper copy sent by courier or pneumatic tube system, etc.).

Most laboratories promise within-lab TAT for stat tests of 60 min or less.[5] Pre-lab time consists of collecting and transporting specimens to the laboratory; this time may be from several minutes to 30 min or more. Within-lab TAT for routinely requested tests may be 3–6 h for assays that are done throughout a 24 h day.

Many factors can influence the within-lab TAT:

- number and complexity of steps in accessioning and processing specimens (may be substantially different from laboratory to laboratory, frequently according to procedures dictated by an institution's HIS, the laboratory LIS, interfaces between them, and how specimens are labeled);
- number of specimens arriving at the same time;
- number of resources (especially centrifuges and analytical instruments) available to handle the work load;
- number of personnel available to process specimens and perform testing;
- proportion of difficult specimens (e.g. small volumes collected from babies or geriatric patients; poor-quality specimens with hemolysis, icterus, lipemia, etc.) that require special handling;
- proportion of stat specimens (which require immediate attention) and routine specimens (which may be held for batch testing).

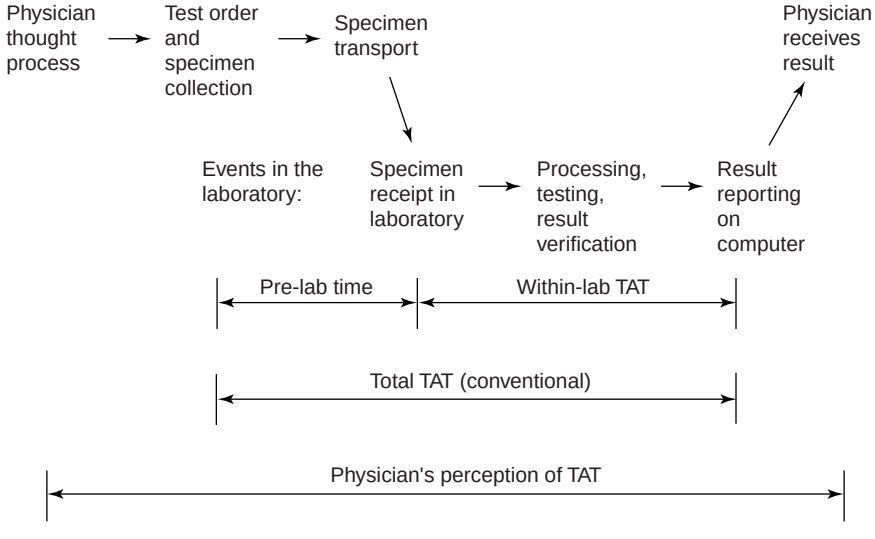

Figure 2 Schematic of TATs for clinical laboratory processes.

The goal of TLA is to develop new ways of performing these tasks in an effort to eliminate bottlenecks in the work flow of these within-lab activities. Other processes external to total TAT will not be altered directly by improving within-lab TAT. They may be improved more through use of better information systems and transportation devices that interface with laboratory automation systems (LASs).

As improvements in TAT occur, expectations for even faster TATs grow in the minds of physicians.[6] Not only does faster performance by central laboratories fuel this expectation but so also does the growth of point-of-care testing (POCT), which can be done at the patient's location with no requirement for transportation and essentially no waiting time for test results other than a brief period from sample collection to analytical completion. Many analytes such as glucose, coagulation times, arterial blood gases, and electrolytes can now be measured by POCT. It is expensive in terms of labor (one sample done at the time), reagent costs (generally prepackaged for single tests), and redundancy (many identical instruments sitting idle throughout a medical center). The challenge for future laboratory automation is to improve within-lab TAT so much, while keeping costs low and quality of results high, that centralized testing will remain preferable to distributed testing (POCT) in the minds of physicians who must rely on timeliness of results to treat their patients.

2.5 Likely Targets for Automation

Every place that a bottleneck now exists in clinical laboratory processes is a potential target for automation, especially if the tasks involved are highly repetitive and do not require a value judgment by a skilled operator. Within the laboratory, the possibilities include:

- *Accessioning* by optically scanning a bar-coded label that was placed on a specimen at the site of collection.[7] This label may be sufficient for a specimen's entire course through analysis and storage. Some laboratories now place a second label with bar code onto specimens after they arrive with another label (e.g. not bar coded, perhaps not specific for the specimen and requested tests, merely patient-specific with name and demographics). Double labeling is redundant, which makes it a clear target for improvement.
- *Centrifugation* of specimens to produce serum or plasma samples for analysis is generally done by an individual who inserts specimen tubes into holders and balances them in a centrifuge. After 5- to 10-min spins (according to specimen type), the operator removes the tubes manually for the next step in processing. The wait period between spins can vary considerably according to operator; furthermore, multiple specimens arriving at the same time may easily saturate the capacity of a centrifuge or the few that are assigned to processing. At the very least, automation of centrifugation can be expected to achieve uniformity in the timing of spins based on standard rules rather than on the arbitrary decisions of individual operators, who may have different approaches for waiting until a centrifuge is full or not before beginning a spin.
- *Aliquot* formation is done to allow a single specimen to be split for analysis on different instruments or even in different laboratory sections (e.g. chemistry and immunology). Different approaches to automation may change dramatically or even eliminate the need for aliquots. First, the specimen tube with bar code label from time of blood draw may be sampled directly, thus eliminating the need for an aliquot if only a single instrument is involved. Second, multiple smaller tubes may be collected from the patient, each for a different instrument. Third, a single large tube may be transported automatically from instrument to instrument for sampling without any aliquots. Fourth, aliquots may be made with a device that generates an additional bar-coded label for each aliquot tube to match the original label. Each of these strategies may be applied in different automation systems.
- *Transporting* specimens after processing to analyzers for sampling and analysis may be done automatically.
- *Storage* of specimens and *retrieval* for repeat testing or additional requests from a physician can both consume enormous effort (although probably few laboratories can define rigorously the actual work effort involved, due to extremely complex modes of storage according to date, time of analysis, which instrument was used, etc.).

Each of these activities consists of repetitive tasks that are done manually in most laboratories but each potentially could be automated without loss of critical operator interactions.

3 CONTROL OF LABORATORY ANALYSES

It is essential for the laboratory to have information about what analyses to perform on which specimens. The path of this information begins with the ordering physician, who may fill out a paper requisition or enter test requests into a computer; this information provides direction for subsequent laboratory activities. The path culminates

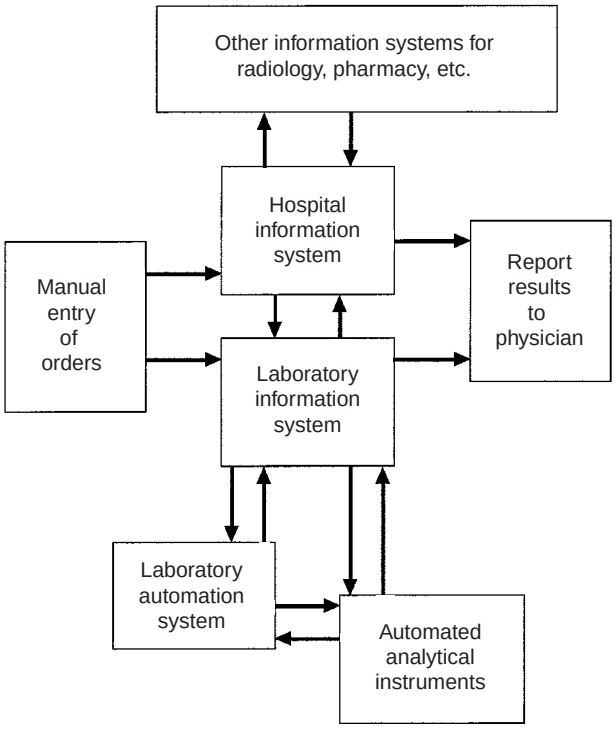

Figure 3 Communication links between hospital and laboratory systems for transmission of information, from ordering laboratory tests to reporting results.

with the reporting of results to the ordering physician. Segments of activity come directly under the control of the LIS and a LAS, which may include interlinked systems for processing, transportation, analysis, and storage (Figure 3).

3.1 Laboratory Information Systems

The LIS provides the highest level of control for directing activities in a clinical laboratory. It receives orders for specific tests from the hospital information system (HIS; also called clinical system or medical information system). A modern HIS generally connects multiple functional units of a medical center, such as laboratory, pharmacy, radiology, dietary services, etc., with patient care units. Physicians, nurses, and other healthcare providers can gather medical information and request tests or treatments through a single HIS workstation, which in turn transmits those orders to separate information systems for laboratory, radiology, pharmacy, etc. The latest generation of clinical systems further integrate the HIS with all these other systems into a single system in which common databases can be shared readily, thereby facilitating searches and electronic manipulation of information. One common example is to confirm whether a patient is receiving anticoagulant medication (information stored in a pharmacy system) when coagulation test results are prolonged beyond the normal range (information in a laboratory system). Many other examples occur in therapeutic drug monitoring for antiarrhythmic, anticonvulsant, and antibiotic medications: what is the dosing schedule with regard to time of specimen collection for peak level, trough level, or other time of blood draw? Thus, the LIS acts as the main conduit of information from the HIS regarding laboratory requests. The LIS may also be the route by which requests are entered for specimens originating in locations not served by the HIS; this situation may arise with specimens collected at physician offices or other sites that use paper requisition forms for ordering.

All events that happen to specimens in the laboratory, beginning with accessioning, are either directed or monitored by the LIS. Of particular note is the role that the LIS plays in assigning accession numbers to each specimen that uniquely identify it versus thousands of other specimens potentially in the laboratory at the same time. The LIS can direct printing of specimen labels with bar codes to track each specimen as it passes through different sections of the laboratory and to identify each specimen as it arrives at an instrument for analysis.

Specific processing (e.g. centrifugation and aliquot formation) is dictated by the LIS, as well as the routing of samples for analysis to individual instruments. These particular tasks in theory may be under the moment-to-moment control of either the LIS or a LAS if it also has a computer that can receive orders from the LIS and then direct those activities.

The LIS is also interfaced to each analytical instrument for receipt of verified results after testing is complete. The LIS participates in verification through checks on quality-control sample results plus delta checks performed on individual patient results (see section 2.2). The electronic interfaces for exchange of specimen test orders and test results between LIS and instruments constitute a significant portion of set-up efforts and cost whenever a laboratory introduces a new automated instrument; not only are the laboratory and instrument vendor involved, but also the LIS vendor for programming and validation of information transfer. Some efforts have tried to link multiple instruments through a single interface engine to the LIS in order to keep costs lower; however, this concept must extend to include instruments from different manufacturers if it is to be successful and widely adopted.

The LIS should store test results in electronic patient files so that all results on a given patient may be retrieved together. This function allows the LIS to print result report forms for physicians to receive. Alternatively, reporting can be done by the HIS, or by a composite HIS/LIS in an integrated clinical system. Although result reporting occurs outside the formal walls of a laboratory,

it is practically the only way in which physicians can judge the performance and quality of laboratory services on a daily basis. Accordingly, laboratories place a great deal of effort in the function of reporting and communicating results to healthcare providers (e.g. panic values by telephone or paging system; printer or facsimile transmission to clinic office or medical ward sites). The reporting capabilities are just as important as analytical capacity whenever a laboratory is considering whether or not to accept new work; the cost of installing these additional peripheral devices becomes a factor in such decisions.

3.2 Laboratory Automation Systems

The LAS can accomplish several different tasks of specimen handling, processing, transport, presentation to analyzer, storage, and discarding. Such systems are modular, which allows the laboratory to choose particular functions for automation according to the most significant bottlenecks in the flow of specimen analysis.

TLA is a concept in which essentially all manual processes in the laboratory are replaced by automated systems. By TLA, no operator need handle or otherwise directly interact with patient specimens after they have been collected, transported to the laboratory, and placed into the receiving device for specimen handling. Of course, technologists or other operators would be needed to monitor quality control and system stability, as well as to trouble-shoot problems that arise unexpectedly with equipment or with particular specimens that may not be acceptable for automated analysis. Laboratory assays that are not automated, due to relatively low volumes or highly complex methods, would not be handled by TLA; however, the bulk of requests for laboratory work consists of a relatively small number of procedures that are performed widely on automated instruments that can be adapted readily to a TLA environment.

3.3 Processes in a Totally Automated Laboratory

In a totally automated laboratory, tubes of blood already uniquely identified with bar-coded labels are placed into some type of specimen carrier that will transport the specimen throughout the laboratory. Because tubes are recognized by bar code readers solely through their labels, all tubes from the same blood draw on a patient must have different accession numbers; by this means, tubes with different anticoagulants and clot tubes will not be confused with each other. Most automation systems read the bar code on the specimen label at various locations throughout the laboratory in order to track tube progress and arrival at workstations. Alternatively, the tube may be placed into a carrier that has a bar code label that becomes the unique identifier linked to the specimen tube during the period when the specimen is actively being processed and analyzed in the TLA system.

Reading the bar code (Figure 4a) may be facilitated by lifting and/or rotating the tube with a robotic device to move the label past an optical scanner. Picking up tubes, moving them in a prescribed manner, and placing them back into the carrier or other defined space are actions that require mechanical precision and accuracy plus gentle forces that will not damage glass or plastic tubes. Fortunately, much of the technology that is necessary for these manipulations has been developed already for other applications, such as the semiconductor manufacturing industry, and can be translated to laboratory medicine.

The carrier should allow several different things to be done with a specimen tube. First, it should be able to handle the different tube sizes that are common in medical centers treating widely different patients, some of whom cannot tolerate frequent draws of large tubes of blood (e.g. neonatal, pediatrics, geriatrics, oncology, etc.). The carrier must travel on a transport system between laboratory locations while holding the specimen in an upright position and keeping it from disruption. The carrier should hold the specimen firmly but allow the bar code label to be visualized through ports or openings to the side. If a specimen tube does break and potentially biohazardous blood or body fluid leaks out, the carrier should be cleaned and decontaminated readily for continued operation; it should also contain the spill as much as possible, to minimize its spread through the laboratory.

Once a specimen has been accessioned, it must be directed toward the appropriate preanalytical processing. Hematology samples are relatively simple in this regard because they can be sent immediately to a cell-counting instrument that analyzes whole-blood samples that are thoroughly mixed. As in the case of essentially all high-volume equipment, most clinical laboratories generally have two or more hematology analyzers; they provide back-up for one another during both planned and unplanned downtime for repairs and to do preventive maintenance; they also afford expanded capacity for testing during peak workload hours. The TLA system chooses which instrument should receive each specimen according to available instrument time, and delivers it to that workstation by way of a transport or track system.

Other specimens that are intended for chemistry, coagulation, and immunochemistry testing must be processed first by centrifugation. The TLA system recognizes these specimens by their bar code label and, using information from the LIS about requested tests, sends them to one or another centrifuge. The centrifuge is automatically loaded by a device such as a robotic arm, which places tubes into the centrifuge head with some action that also balances the centrifuge, such as putting other patient specimen tubes or blank tubes into opposite

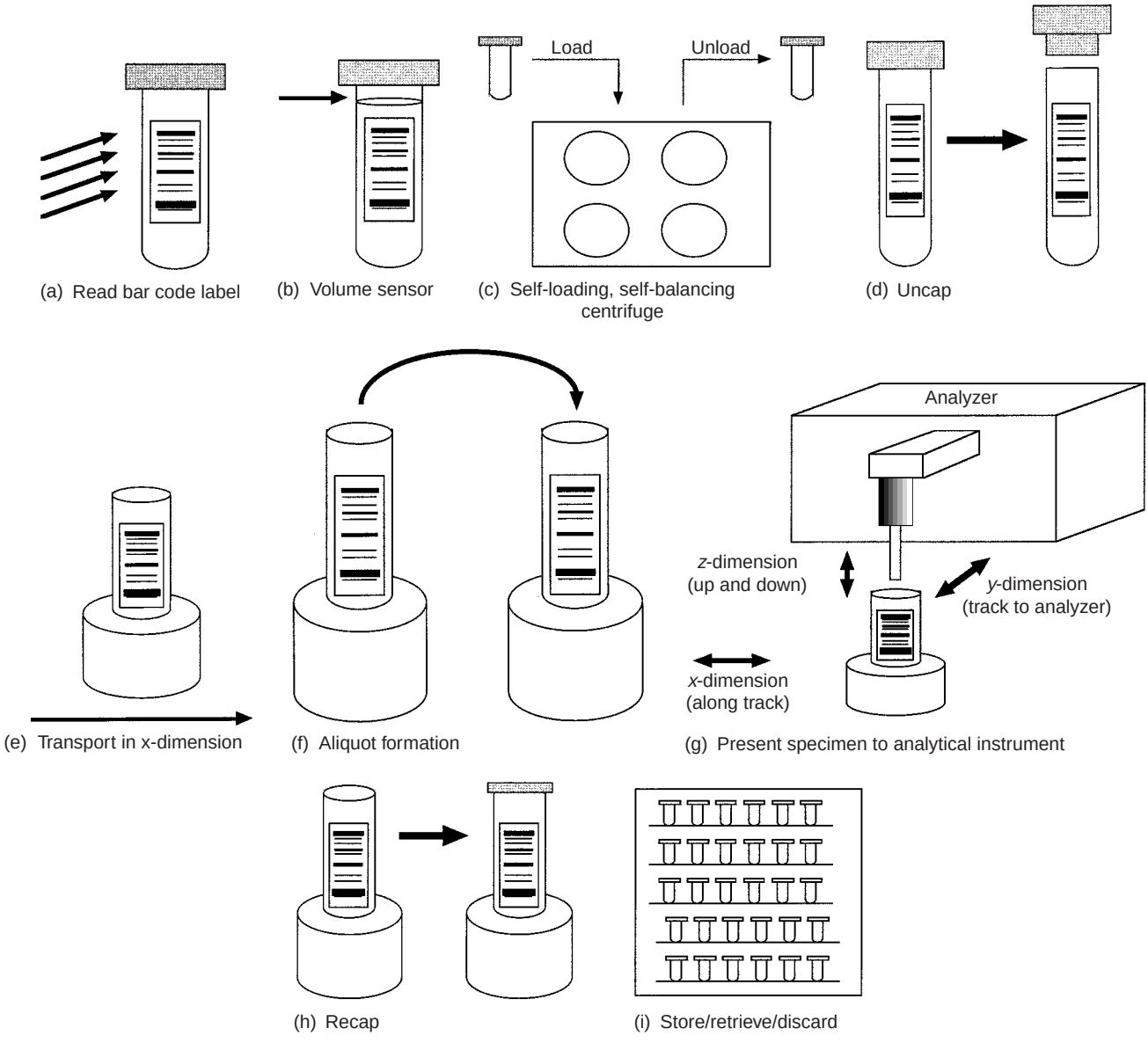

Figure 4 Clinical laboratory tasks for which automation can replace repetitive manual actions.

positions (Figure 4c). After some defined waiting period (for additional specimen tubes to arrive and also be loaded), the centrifuge is spun for a prescribed time (usually 5 or 10 min). The robotic arm then automatically removes tubes from the centrifuge and places them into carriers for further transport.

One alternative at this point is to have the tubes placed into instrument-specific carousels that will be carried manually by an operator to the appropriate analyzer. This configuration uses automation solely for centrifugation, and so is formally not TLA; however, it accomplishes the goal of automating an early bottleneck in specimen processing that consists of repetitive tasks for which operator interaction and oversight add little to no value.

In a TLA configuration, following centrifugation, serum and plasma tubes are automatically decapped (Figure 4d) and the rubber stoppers are discarded. Sample volume may be estimated by a sensor that detects the top of the serum/plasma fluid level using infrared light, which can read through a label (Figure 4b). Alternatively, sample volume may be measured on analytical instruments by an aspirating probe that detects a pressure change at the airfluid interface at the top of the tube and another one at the serumgel interface in serum separator tubes.

Serum or plasma sample quality may be judged at this stage by examination for hemolysis (pink to red from hemoglobin in solution), icterus (intensely yellow

from increases in circulating bilirubin), and lipemia (turbid sample due to large amount of chylomicrons or occasionally other lipids), plus any other unusual appearance. These common specimen abnormalities cause a large number of potential interferences with assays that depend on optical absorbance measurements. One approach may be to ignore such specimen problems unless they cause marked abnormalities in test results, and only then examine samples for visual explanation of discrepant findings. Another approach is to station an operator at a viewpoint in the movement of tubes to observe the color and appearance of each sample; questionable ones could be identified and removed manually. Neither of these approaches is consistent with TLA because they depend on having an operator constantly available for interaction as part of the process.

To have a TLA system perform the specimen quality check automatically is not a trivial matter because of a technical problem: primary tubes may have at least two paper labels on them (one from the manufacturer and the second to identify the patient and specimen with a bar code), which will obscure the tube contents and block light from being transmitted consistently through the serum or plasma to detect absorption at wavelengths specific for hemoglobin and bilirubin, or scattering by particulate chylomicrons. Some analytical instrument manufacturers have already addressed the problem of detecting these interferents by aspirating the sample and making the appropriate optical absorbance measurements on the instrument itself. This capability could transfer the responsibility of interference checking from the TLA system to an instrument; however, a potential weakness may occur for samples that are intended for testing only on instruments that do not have the checking capability. In that case, it might still be possible to route the specimen to the checking instrument first for that function, even though that instrument will perform no other testing on it; the specimen then would be routed to the instrument for its requested tests, and the LIS would have the responsibility to track whether any detected interferents would be important for the assay procedures performed.

If aliquots are to be made for analyses on different instruments, serum or plasma is aspirated from the primary tube and dispensed into secondary (aliquot) tubes labeled with bar coding that relates them to the original specimen (Figure 4f). Using aliquots is not necessarily a preferred practice unless a laboratory has an extremely good reason to do so (e.g. a laboratory design in which only centrifugation and aliquot formation are automated; processed tubes are placed into separate specimen carousels for operators to place manually onto different types of instruments). If the processed specimen can be transported automatically to several different instruments in the same laboratory, that primary tube should be sampled directly by each of the instruments for which testing is requested. Avoiding the need for aliquots removes a step that consumes time, requires additional nonanalytical equipment, and increases the total number of tubes that must be dealt with in the whole TLA system at one time.

Transport to analytical instruments is done along a track system that runs horizontally above the floor. Motion along the track is referred to as the *x*-dimension (Figure 4e). Motion horizontally perpendicular to the track is in the *y*-dimension. Motion vertically perpendicular to the track is in the *z*-dimension. Transport between various workstations is essentially in the *x*-dimension (with whatever allowances are necessary for track curvature). When a specimen tube arrives in its carrier at an instrument, the instrument is set back in the *y*-dimension from the track. Because the TLA recognizes that the specimen tube has arrived at its proper testing site, the instrument may extend an arm with a probe in the *y*-dimension followed by descent into the tube along the *z*-dimension, to aspirate the sample (Figure 4g). This ''direct track sampling'' is done without moving the tube further. It can be accommodated by many instrument manufacturers whose protocol for sample aspiration consists of moving an arm with probe back and forth over a carousel of samples placed onto the instrument. Modification and extension of the arm motion readily allows sample aspiration to occur with the same devices but from a different direction (e.g. the back or side of the instrument past which the track runs). Another manner in which specimens may reach the instrument is for a robotic arm to lift the tube from its carrier on the track, place it onto the instrument for sampling by whatever method is involved, and return the specimen tube to its carrier for further transport.

Both the direct track sampling and tube-moving methods require that the tube be delivered by the track to the instrument in a defined position that is referred to as the ''point of reference (POR)''. The TLA system has the responsibility to deliver the specimen tube to a target zone in which the instrument expects to find that tube. Communications and status reports are essential to achieve an orderly sequence of events in which the proper specimen is delivered to the correct instrument, a sample is aspirated, analysis is performed, and the specimen tube proceeds to other stations for further testing or for storage.

Once all tests are performed on a specimen tube, it is recapped (Figure 4h) and placed into a refrigerated chamber in a location from which it can be retrieved readily for repeat or additional testing (Figure 4i). In addition, the TLA should keep specimens organized so that when they are outdated they may be discarded in a timely manner.

3.4 Laboratory Workstations

One alternative to TLA is to use many of the same automated devices without using a track or other automated transport system. In this approach, automated centrifugation and other specimen processing would prepare tubes for operators to carry in batches to individual instruments for analysis. Following testing, the operator also removes the specimens and prepares them for storage. Although this configuration may appear to avoid the benefits of automation, it in fact maximizes the use of particular components for eliminating the traditional bottlenecks in which specimens entering the laboratory must await initial processing before they are taken to analytical instruments. Keeping operators involved also uses their judgment regarding specimen volumes and quality at a critical point before placing specimens onto analytical instruments.

4 WORKSTATIONS AND AUTOMATION

4.1 Trends in the Consolidation of Workstations

Over the last few decades, analytical instruments across many laboratory disciplines have followed similar lines of evolution. The first general trend has been to put more analyses onto a single analyzer.[4,8] Chemistry instruments provide an excellent example: high-volume analyzers in most clinical laboratories now have the ability to perform a wide array of assays based on different principles of measurement, such as spectrophotometry and ion-selective electrodes. Essentially, all common serum and even urine analytes can be quantitated on the same instrument, thereby allowing a single technologist operator today to perform at a single workstation what 30 years ago might have required multiple operators manning ten or more different instruments. The labor component for generating these laboratory results thus has fallen markedly as multichannel analyzers have developed.

Other common trends in instrument development are related to the presentation of specimens. Virtually all analyzers now have the capability of reading bar code labels to identify specimens; they confirm specimen identity through electronic interfaces with the LIS, which also directs specific test performance and receives verified results through standard electrical parts. Random access mode of operation allows the analyzer to perform only those tests requested but at any time, so that batch mode is no longer necessary. Thus, multiple different assays can be done from a single sampling of a specimen. Sampling occurs by aspiration of serum with a disposable tip (to prevent carryover between specimens) on a probe that is moved into position by an instrument-controlled armature.

In hematology testing the CBC, with five-part differential, is now fully automated on a single instrument, with other features (such as reticulocyte count, immaturity markers for granulocytes and platelets, and lymphocyte surface markers) being introduced on the newest models. Automated immunochemistry analyzers also have made impressive gains, largely due to the easy availability of monoclonal antibodies as reagents for endocrine markers (thyroxine, cortisol), therapeutic drug monitoring markers (digoxin, other antiarrhythmics, anticonvulsants, antibiotics), tumor markers (carcinoembryonic antigen, human chorionic gonadotropin, prostatic specific antigen), myocardial injury markers (myoglobin, creatine kinase MB, cardiac troponins I and T), and infectious disease markers (hepatitis viruses A, B, C). These immunoassays generally required 48 h for completion when they were introduced as radioimmunoassays many years ago; today they can be performed in minutes – a time frame that compares favorably with chemistry analyses. Accordingly, the next major consolidation is to link traditional chemistry analyses with immunoassays on the same instrument. This consolidation would potentially bring together two or three already large workstations into a single massive one. For a TLA environment, the consequence is greater simplicity for delivering specimen tubes to a single workstation for sampling into multiple channels of testing.

4.2 Characteristics of Instruments that can be Used in Total Laboratory Automation

Most high-volume instruments in use today have potential for incorporation into a TLA laboratory in which specimens are taken off a transport track or in which sampling occurs from a position with the specimen tube still on the track. How a particular instrument can be adapted to automated delivery of specimens depends to a great extent on how close the track can be mounted to the sampling device. In some instances, a vendor-supplied retrofit may allow sampling to occur from a side or rear location compared to normal front sampling. The task is made much more feasible when the instrument's own armature and probe can be adjusted to move to virtually any site around the instrument, especially to the rear, where a track may fit flush against the instrument's back. The electromechanical interface between instrument and track system may be quite simple when it relies on readily modified instrument capabilities.

Current instrumentation in chemistry, immunochemistry, hematology, coagulation, and possibly urinalysis appears to be reasonably adaptable to TLA operations (Figure 5). Most of the limitations probably are related to software incompatibilities among instruments, LASs, and LISs rather than on mechanical incompatibilities. Other

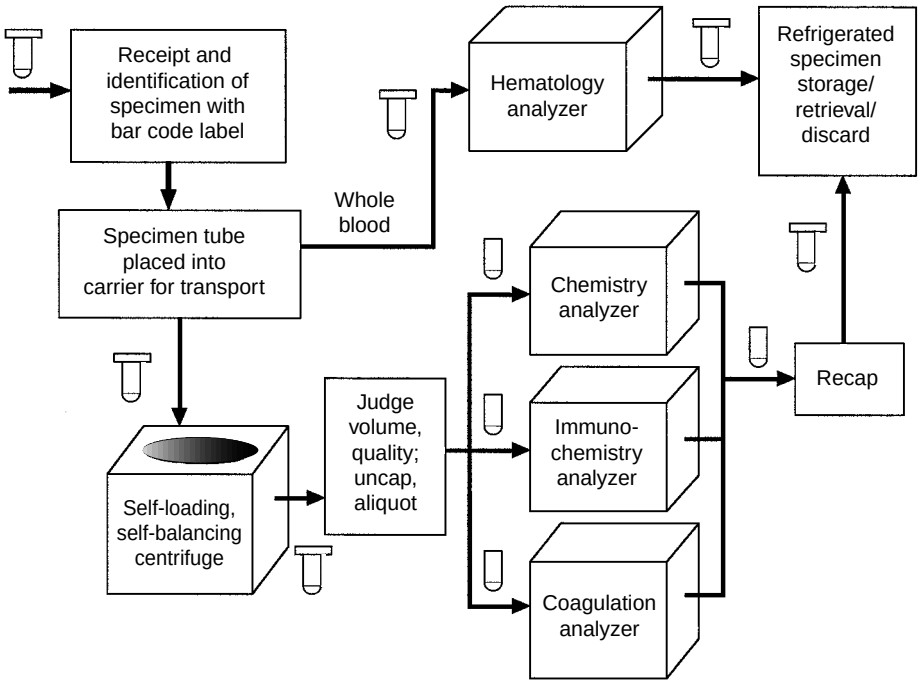

Figure 5 Scheme for processing, transporting, analyzing, and storing specimens in a TLA environment.

long-term applications are likely to arise with automation of processes in laboratory sections that are presently highly manual, such as microbiology, blood banking, cytology, flow cytometry, and eventually perhaps histology and molecular diagnostics.

Of greater concern to the future success of TLA systems is the front end portion, namely the specimen processing devices that automatically load and balance a centrifuge and then unload specimens and produce aliquots in accurately labeled tubes. Few laboratories have had experience with these automated front ends, so little actually is known about their reliability. Because specimen processing begins the laboratory activities, it must be robust and immune from serious breakdown. Any downtime would require replacement with a duplicate system or extensive labor. Front-end processing entails high-speed moving elements, and so may have significant problems during an operational lifetime of several years on a 24-h day, 7-day per week basis.

5 ACCOMPLISHMENTS IN TOTAL LABORATORY AUTOMATION

The earliest efforts in TLA were implemented successfully in Japan, first by Professor Masahide Sasaki[9] from the Kochi Medical School and then by companies such as A&T Corporation, Hitachi Instruments, Toshiba/IDS, and Sysmex Corporation. TLA systems have been implemented in many Japanese healthcare facilities. Because generally they have been interfaced with LISs that are not available in the USA, such TLA systems have not been directly transferable to other LISs in common use in North America. Accordingly, vendors are now making efforts for implementation in US laboratories with interfaces to different LISs. Current vendors of TLA systems or its components include Abbott Diagnostics (front-end automation and separate workcell automation at instruments), Auto Lab Systems, Bayer Diagnostics, Beckman Coulter, Inc., LAB-Interlink, Labotix, Roche Diagnostics/Boehringer Mannheim Corp., Sysmex Corporation (thermooptical absorbance; automated hematology system), and others possibly entering the market.

Because of the diversity of companies that may supply TLA systems and also of those manufacturing analytical instruments for clinical laboratories, the National Committee for Clinical Laboratory Standards (NCCLS), Wayne, PA has embarked on a program to set standards prospectively for developing TLA. The overall goal is to achieve standardization of equipment and processes so that instruments and automation systems will be highly compatible between manufacturers. This effort is intended to stimulate the overall growth of laboratory automation in the USA and elsewhere by encouraging standards that will minimize future problems of implementation at the individual laboratory level. As the culmination of this effort, the NCCLS has drafted five

standards that are currently or soon to be approved. They cover the topics:

- specimen container/specimen carrier;
- bar codes for specimen container identification;
- communications with automated clinical laboratory systems, instruments, devices, and information systems;
- systems operational requirements, characteristics, and information elements;
- electromechanical interfaces.

6 COST JUSTIFICATION OF LABORATORY AUTOMATION

The current financial climate is one in which reimbursement for laboratory and medical services in general has decreased and can be expected to decrease further. This reduction is true for all major classes of payers: private insurance; Medicare and Medicaid; indigent care (city, county, or state payments); and self-pay individuals. Managed care organizations negotiate low payment rates that are typically capitated on the basis of covered lives and not the number of tests performed; doing more laboratory tests leads to no more reimbursement; doing less testing can be undesirable if it does not provide at least a minimum of preventive measures. Consequently, laboratories are generally expected to maintain productivity while reducing their costs of operation.

Capital expenditures are used to acquire large pieces of equipment and to make changes in building structures and utilities for installation. Acquisition of equipment for TLA is in this budget category; most institutions will depreciate capital purchases at a flat rate over a period such as 5 years. The source of capital may be private investors in a for-profit corporation, or it may be the excess of revenues over operating expenses (direct costs plus overhead and depreciation) in a nonprofit entity. In either case, the justification for major capital expenditure is usually a promise of reduced operating expenses (i.e. costs of labor, reagents, other consumables, and various purchased services such as maintenance contracts, etc.).

The evolution of workstations into consolidated multichannel analyzers has favored a reduction in labor expense because the number of personnel needed to operate fewer instruments has fallen. Technologists are utilized to prepare and present specimens to analyzers and to monitor quality control; they trouble-shoot when instrument problems arise. The current staffing pattern in most laboratories is one operator per instrument per shift. In a TLA environment in which robotic devices prepare and present specimens to analyzers, a single operator could monitor the quality control of several analyzers simultaneously and still be available for trouble-shooting. Thus, conversion to TLA has the potential to eliminate a substantial number of the workforce now employed in chemistry, hematology, immunochemistry, and coagulation testing.[10]

To justify the cost of implementing TLA or some of its components with a total price tag in the range of one to several million dollars would require staff reduction with total salaries over 5 years or less equaling the purchase expense. Some laboratories claim to have made such staffing reductions by implementing TLA; other laboratories have achieved similar staff reductions by reorganizing work flow without using TLA equipment. Consequently, the real challenge is for TLA to reduce staffing in an already pared-down laboratory. This issue is paramount today when laboratory directors consider the prospects for TLA.

How is a laboratory to assess its current performance versus other laboratories? Some form of benchmarking should be used to compare productivity (e.g. tests per technologist) and costs (e.g. total expenses per test) in similar institutions. One such program is the College of American Pathologists' Laboratory Management Index Program (LMIP); using LMIP or another benchmarking system is essential for laboratories to assess whether improvements occur or not in response to changes that are instituted in operation. Reliable data regarding changes from TLA can be expected to appear in the next few years.

7 EXPECTED BENEFITS FROM TOTAL LABORATORY AUTOMATION

The major benefit expected to occur is a reduction or at least stabilization of labor costs by reducing the number of personnel involved in test performance. Financial benefit will be the major reason why institutions choose to invest in TLA. Other benefits also should be expected as a result of transferring repetitive tasks from humans (who are error prone in such work) to instruments and mechanical devices with extremely low failure rates. Thus, simple mistakes of labeling and specimen identification may be virtually eliminated by TLA as long as the patient is correctly identified at the time of specimen collection.

The whole process of specimen flow through a laboratory also should benefit from standard application of time sequences and decision rules for centrifugation, aliquot formation, analysis, storage, and retrieval/discarding of specimens. The within-laboratory time should improve with a high-capacity TLA system that can deliver specimens to instruments as rapidly as they can be analyzed. The distinction between stat and routine testing may also disappear as the throughput is optimized in a TLA laboratory.

One issue deserves special attention: what would be the consequences of making essentially all automated testing constantly available instead of batching some lower volume assays once per day (e.g. endocrine testing, therapeutic drug monitoring, hepatitis serologies)? Currently, most laboratories batch routine requests for these tests because the expense of reagents is high to perform quality control testing on each shift of operation when only a few patient specimens are analyzed. Thus, constant availability of testing could increase the fixed costs of an assay; however, these costs could be offset by lower labor costs associated with special handling for batched testing. The economics of this balance could be different for every laboratory and so must be resolved on an individual basis.

8 PROBLEMS AND DIFFICULTIES FROM TOTAL LABORATORY AUTOMATION

The overwhelming obstacle to initiating a course of TLA is the huge expense to acquire the equipment, although acquiring it gradually in stages may ameliorate the financial impact.[11] Additional expertise at a high level will be needed to maintain TLA operations, although probably not substantially different from that already required for standard laboratory instruments. The interfaces between LAS and LIS as well as between instruments and LAS will be new to laboratories; they will require efforts to install and may be quite expensive beyond the initial costs of acquisition.

Probably the most daunting aspect of TLA will be the effect that it has on workers who are displaced. Unless they learn new skills, they may not adapt to a TLA environment. These changes will affect laboratory culture and the way in which people value themselves; institutions should maintain programs of essential continuing education to ease change and to bring out the best from workers as part of an overall strategy for implementing TLA.

9 FINAL ASSESSMENT OF TOTAL LABORATORY AUTOMATION

How can a laboratory decide whether it has done the right thing by implementing TLA? Improvement in laboratory processes from TLA should be assessed in terms of the following parameters: increased productivity per technologist; decreased cost per test; shorter TATs; fewer errors; and a better ability to monitor the flow of specimens through a laboratory. Although changes in intralaboratory processes probably will not be known to them, physicians should be satisfied with shortened TATs. Success in the reduction of TATs from TLA may be gauged in part by the reduction in the amount of decentralized testing performed as POCT for the same analytes delivered more efficiently and at less cost in a centralized laboratory. Finally, and very importantly, improvement should be judged also by employee satisfaction and the degree to which workers adapt to this new technology through re-education and commitment to changes that improve medical care.

ABBREVIATIONS AND ACRONYMS

APTT	Activated Partial Thromboplastin Time
CBC	Complete Blood Count
HIS	Hospital Information System
LAS	Laboratory Automation System
LIS	Laboratory Information System
LMIP	Laboratory Management Index Program
NCCLS	National Committee for Clinical Laboratory Standards
POCT	Point-of-care Testing
POR	Point of Reference
PT	Prothrombin Time
TAT	Turn-around-time
TLA	Total Laboratory Automation

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Diagnostic Hematology • Laboratory Instruments in Clinical Chemistry, Principles of • Point-of-care Testing • Statistical Quality Control in Clinical Laboratories

Pharmaceuticals and Drugs **(Volume 8)**
Robotics and Laboratory Automation in Pharmaceuticals Analysis

REFERENCES

1. R.S. Markin, 'Implementing Automation in a Modern Clinical Laboratory', *Chemomet. Intelli. Lab. Syst. Lab. Inf. Manage.*, **21**, 169–179 (1993).
2. J.C. Boyd, R.A. Felder, J. Savory, 'Robotics and the Changing Face of the Clinical Laboratory', *Clin. Chem.*, **42**, 1901–1910 (1996).
3. R.S. Markin, 'Recent Trends in Clinical Laboratory Automation', *Clin. Lab. Manag. Rev.*, **12**, 176–180 (1998).
4. R.A. McPherson, 'Robotics, Automation, and the New Role of Process Control', *Clin. Lab. Manage. Rev.*, **12**, 339–346 (1998).

5. S.J. Steindel, D.A. Novis, 'Using Outlier Events to Monitor Test Turnaround Time. A College of America Pathologists Q-Probes Study in 496 Laboratories', *Arch. Pathol. Lab. Med.*, **123**, 607–614 (1999).
6. S.J. Steindel, 'Timeliness of Clinical Laboratory Tests: A Discussion Based on Five College of American Pathologists Q-Probe Studies', *Arch. Pathol. Lab. Med.*, **119**, 918–923 (1995).
7. L.A. Brzezicki, 'Bedside Bar Coding', *Adv. Admin. Lab.*, **7**, 46–50 (1998).
8. L.A. Brzezicki, 'Merging Chemistries Through Automation', *Adv. Admin. Lab.*, **7**, 47–49 (1998).
9. M. Sasaki, 'The Belt Line System – Completely Automatic Clinical Laboratory Using a Sample Transportation System', *Jpn. J. Clin. Pathol.*, **32**, 119–126 (1984).
10. L.A. Brzezicki, 'Lab Automation Risks, Opportunities', *Adv. Admin. Lab.*, **8**, 76–84 (1999).
11. R.A. Felder, S. Graves, D. Robertson, B. Hulette, D.A. Ferkany, 'Modular, Step-wise Approach to Automation', *Adv. Admin. Lab.*, **8**, 72–76 (1999).

Biochemical Markers of Acute Coronary Syndromes

Hassan M.E. Azzazy and Robert H. Christenson
University of Maryland School of Medicine, Baltimore, USA

1 Introduction: Where Do Biochemical Markers Fit In? 1
2 Acute Coronary Syndromes: Definition 2
3 Pathophysiology of Acute Coronary Syndromes 3
4 Diagnosis of Myocardial Infarction 4
5 Link Between Heart Disease and Infectious Diseases 4
6 Methods for Cardiac Marker Measurements 5
6.1 Electrophoresis 5
6.2 Immunoassay 5
6.3 Point-of-care Testing 5
7 Decision Limits for Markers 6
8 Specific Markers of Acute Coronary Syndromes 7
8.1 Characteristics of an Ideal Cardiac Marker 7
8.2 Myoglobin 7
8.3 Creatine Kinase-MB Isoenzyme 8
8.4 Creatine Kinase Isoforms 9
8.5 Lactate Dehydrogenase 9
8.6 Proteins of the Troponin Complex 10
8.7 Carbonic Anhydrase Isoenzyme III 10
8.8 Human Heart Fatty Acid-binding Protein 11
8.9 Markers of Platelet Activation 11
8.10 Acute Phase Proteins 11
8.11 Markers of Thrombosis 11
8.12 Markers of Early Myocardial Ischemia 11
9 Applications for Clinical Measurements 12
9.1 Diagnosis of Myocardial Infarction 12
9.2 Reinfarction 13
9.3 Perioperative Myocardial Infarction 13
9.4 Reperfusion Assessment 14
9.5 Risk Stratification/Therapeutic Intervention 14
9.6 Infarct Sizing 16
10 Economic Issues in Cardiac Markers 17
10.1 Is Rapid Testing Necessary? 17
10.2 Triage Strategies 19
10.3 Cost of Patient Hospitalization 20
11 Recommendations for the Use of Cardiac Markers in Coronary Artery Disease 21
11.1 Definition of Classification Scheme 21
Abbreviations and Acronyms 21
Related Articles 21
References 21

The acute coronary syndromes (ACSs) are a continuum of myocardial ischemia ranging from angina, indicating reversible tissue damage, through frank myocardial infarction (MI) with extensive tissue necrosis. The mortality from acute MI in the USA alone is 500 000 deaths per year. New predictors of cardiovascular events, that are easily measurable, are needed to improve diagnostic strategies for treating ACS patients. Cardiac markers, proteins released from the injured myocardium into the circulation, are useful for the diagnosis of MI according to the World Health Organization (WHO) criteria. Troponins I and T have emerged as sensitive, more cardiac-specific clinical indicators for MI diagnosis and for risk stratification. A new generation of biochemical markers for indicating plaque rupture, platelet reactivity, and early MI offers promise for better assessment of patient risk so that clinicians may intervene to avoid adverse outcomes. Other biochemical markers, including the acute phase reactants and markers of thrombosis, may become useful for identifying a patient's location on the ACS spectrum and consequently the risk of adverse events.

This article describes the spectrum of ACS current and future biomarkers that are closely linked to the progression of cardiac injury from ACS, and various applications of biochemical markers in assessment of reinfarction, reperfusion, perioperative MI, risk stratification, guiding therapeutic intervention, and infarct sizing. Finally, an assessment of the importance of rapid testing, different triage strategies for chest pain patients, and new trends to reduce the overall costs of patient hospitalization are discussed. The recommendations of the National Academy of Clinical Biochemistry for use of cardiac markers are also presented.

1 INTRODUCTION: WHERE DO BIOCHEMICAL MARKERS FIT IN?

The new advances in cardiology practice and the new knowledge in the pathophysiology of ACSs in addition

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

to the new diagnostic strategies in treating ACS patients have placed extensive demands on the clinical laboratory to develop and utilize new cardiac markers to answer these needs.

Utilization of biochemical markers of cardiac injury has evolved substantially over the past several decades. In the 1980s, cardiac markers were primarily used for diagnosis of MI. During the 1990s, it became clear that cardiac markers are elevated in cardiac ischemia patients who are at greatest risk of experiencing adverse events. Now at the start of the 21st century, it is reasonable to predict that cardiac markers will increasingly take on a new role for guidance of either medical and/or interventional therapy.

This article describes current and future markers for ACS and their applications in clinical setting for diagnosis of acute myocardial infarction (AMI), risk stratification, assessment of reperfusion, infarct sizing, and perioperative MI. Routine laboratory assays and point-of-care (POC) testing for the determination of cardiac markers are summarized. Finally, an assessment of the importance of rapid testing, different triage strategies for chest pain patients, and new trends to reduce overall cost of patient hospitalization are presented.

2 ACUTE CORONARY SYNDROMES: DEFINITION

For the past several decades, heart attack patients were largely regarded as having a binary condition, i.e. MI or non-MI, using the WHO recommendations that include at least two of the following three criteria: clinical symptoms suggestive of myocardial ischemia, characteristic changes on the electrocardiogram (ECG), and a rise and fall in serum biochemical markers. However, acute cardiac ischemia actually represents a spectrum of disease, termed the ACSs, which ranges from angina, reversible tissue injury → unstable angina, frequently associated with minor myocardial damage → MI and extensive tissue necrosis. The ACSs are a pathological continuum involving erosion and rupture of coronary artery plaques, activation of platelets, and typically mural or transmural thrombus. As indicated in Figure 1, the continuum of ACSs reflects the physiological process of acute myocardial ischemia and, most importantly from a clinical standpoint, a continuum of patient risk.

Much of the laboratory's role has traditionally been associated with biochemical marker testing for diagnosis

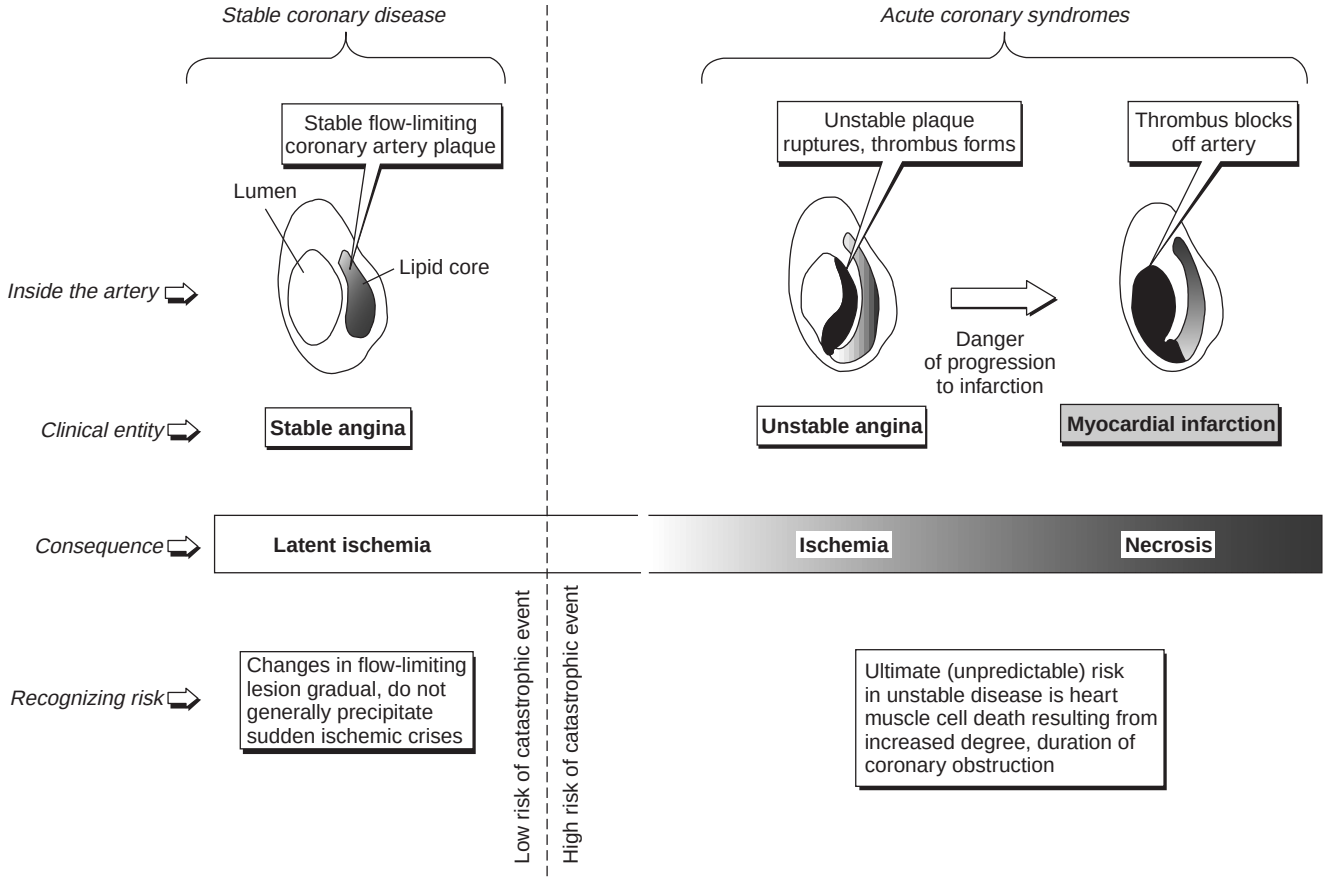

Figure 1 The continuum of coronary disease.

Table 1 Biochemical markers in the continuum of ACSs

Pathophysiology	Biochemical marker	Molecular weight (Da or g mol^{-1})	Cardiac specific?	Type of assay	Duration of elevation	Comments
Plaque rupture	CRP	120 000	NA	Latex photometric immunoassay or rate nephelometry	48–72 h	CRP and SAA are acute phase proteins, may indicate plaque disruption and be prognostic in unstable angina patients
	SAA	12 500	NA	Sandwich-type enzyme immunoassay	48–72 h	
Intracoronary thrombosis	Platelet activation	NA	No	Functional assays using platelet agonists	–	Must be performed soon after blood collection
	P-selectin	140 000	NA	Flow cytometric assay	–	Marker of angry platelets, may indicate risk for acute coronary events
	Soluble fibrin	–		ELISA[b]	–	Marker of procoagulant fibrinolytic activity; may predict patients at higher risk for MI-related complications
Myocardial ischemia	GPBB	177 000	+++	Immunoenzymometric assay	8 h	Marker of cardiac ischemia
Myocardial necrosis[a]	Myoglobin	18 000	No	Immunoassay	12–24 h	Markers of cardiac necrosis
	CK-MB, mass assays	85 000	++	Immunoassay	24–36 h	
	cTnT	37 000	++++	Immunoassay	10–14 d	
	cTnI	23 500	++++	Immunoassay	4–7 d	

[a] Times of first increase for the markers are 1–3 h for myoglobin, 3–4 h for CK-MB mass, 3–4 h for cTnT, and 4–6 h for cTnI.
[b] ELISA, enzyme-linked immunosorbent assay.
NA, not applicable. CRP, C-reactive protein; SAA, serum amyloid A; cTnI, cardiac troponin I; cTnT, cardiac troponin T; GPBB, glycogen phosphorylase-BB.

of AMI, assessment of reinfarction or infarct extension, and estimating the quantity of infarcted tissue (infarct sizing). Logically, these roles have been clearly correlated with the WHO's criteria for MI diagnosis. The laboratory's future role will be more closely linked to the continuum of the ACSs through risk stratification and monitoring of thrombolytic therapy, platelet inhibition strategies, other therapeutic strategies, and perhaps interventions such as coronary angioplasty.

The events depicted in Figure 1 represent a spectrum of risk for an adverse outcome. Patients with stable angina, i.e. reversible ischemia, or myocardial preconditioning are at lower risk for adverse events in both the short-term and long-term compared with patients with Q-wave infarction. Identifying where an individual patient's disease lies in the continuum of ACSs has biological implications regarding the reversibility of injury, quantity of ischemic cell injury, and the patient's relative risk of an adverse outcome.

Much of the focus of biochemical markers has included markers of necrosis, the benchmark being creatine kinase (CK)-MB. However, other markers and substances shown in Table 1 are released or activated prior to necrosis and may have an important role in identifying risk in the ACS patient.

3 PATHOPHYSIOLOGY OF ACUTE CORONARY SYNDROMES

Rupture of atherosclerotic plaque occurs in 50–90% of patients with ACSs. The sequence of events in the ACSs includes plaque disruption, thrombus formation, abnormal vasomotion, platelet activation, and aggregation, as illustrated in Figure 2. Rupture of the surface of an atherosclerotic plaque in a focal segment of an epicardial coronary artery exposes thrombogenic components to circulation, which greatly increases the potential for thrombus at the site (Figure 2). Platelet aggregation and activation of the coagulation systems result in the formation of a platelet-rich thrombus (white

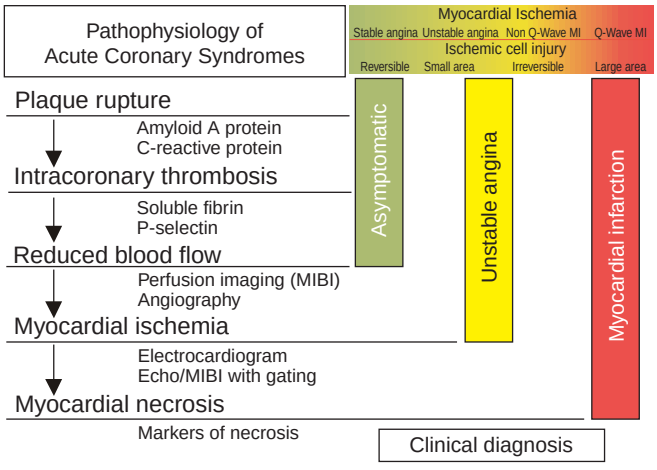

Figure 2 Pathophysiology of ACS. (MIBI = sestamibi.)

thrombus). A thrombus rich in fibrin and erythrocytes (red thrombus) may evolve and extend up- or downstream. Extensive local thrombosis will result in episodic flow-limiting coronary stenosis and myocardial ischemia associated symptomatically with unstable angina or with necrosis that characterizes non-Q-wave AMI. If the plaque rupture and thrombosis are extensive, the coronary artery could occlude completely and generally result in Q-wave AMI. In unstable angina patients, spontaneous ischemia results from reductions in coronary blood flow caused by atherosclerotic plaque combined with thrombosis and vasoconstriction. It is not clear, however, why disruption of a particular plaque would lead to occlusive thrombus in some cases but not others.

Circulating catecholamines have been shown to induce hypercoagulable state, vasoconstriction, platelet activation, and thrombin generation.[1] Therefore, states of sympathetic nervous system hyperactivity such as emotional stress, smoking, and extensive exercise may contribute to arterial thrombosis formation. High cholesterol levels are associated with hypercoagulability and platelet hyperactivity at the site of experimentally induced acute vascular injury.[2] Finally, diabetes mellitus may increase platelet reactivity and induce coagulation. Diabetics who are not treated intensively have a significantly higher incidence of MI.[3] Infectious diseases may also play a role (see section 5).

4 DIAGNOSIS OF MYOCARDIAL INFARCTION

Careful assessment of clinical symptoms is of paramount importance, but can be nonspecific in up to one-third of patients, particularly in diabetics and the elderly who most frequently present with atypical symptoms of ischemia.[4] ECG monitoring is also an important tool that should be performed quickly after presentation of the suspected MI patient because those having diagnostic changes are candidates for immediate reperfusion therapy.[5] The electrocardiograph, however, is not a perfect instrument because its clinical sensitivity for MI is only about 50%.[6] The third WHO criterion, monitoring a rise and fall in cardiac markers, is considered the benchmark for the diagnosis of MI.[7] In symptomatic patients presenting with diagnostic ECG changes, biochemical markers have a limited role for acute MI diagnosis, except for confirmation. On the other hand, biochemical markers are essential for assessment of patients who present with nonspecific or vague symptoms and a nondiagnostic ECG. These nondiagnostic ECG patients comprise the majority, about 55%, of individuals eventually diagnosed as having MI. It is of note that in 4–8% of MI patients the diagnosis is missed, and there is high mortality among this group.[8] Further, misdiagnosis of MI represents the highest outlay of malpractice costs among Emergency Medicine physicians.[9] To help deal with this high prevalence disease and its substantial associated costs, many institutions have initiated a Chest Pain Evaluation Center (CPEC), a specific protocol-driven treatment area intended for systematic and cost-effective care.[10] Rapid, real-time availability of cardiac markers has become an integral part of most CPEC protocols. Recent research has shown that the rate of missed MI was cut from 4.2% to 0.4% by establishment of CPEC protocols.[11]

5 LINK BETWEEN HEART DISEASE AND INFECTIOUS DISEASES

The potential link between infectious diseases, such as typhoid, scarlet fever, and sepsis, and atherosclerosis dates back to the late 1800s. In the 1940s, researchers noted a strong association between atherosclerosis and a herpes-type DNA virus known to cause T cell lymphomatosis.[12] Other diseases linked to infections include peptic ulcers caused by *Helicobacter pylori*, kaposi's sarcoma caused by herpes simplex virus, and the association of hepatocellular carcinoma with hepatitis B virus.

There has been substantial research investigating the association of heart disease with *Chlamydia*, an obligate intracellular bacterial pathogen of eukaryotic cells. Using immunocytochemistry, polymerase chain reaction, and culture techniques, several reports have demonstrated that *C. pneumoniae* can localize to blood vessels.[13]

Other research groups found *Chlamydia* proteins in 79% of plaque specimens obtained from the coronary arteries of heart disease patients, versus only 4% of heart artery walls of normal individuals.[14] Infecting rabbits with *Chlamydia* measurably thickened the arterial walls of

the animals, especially those fed high-fat diets. Treatment of infected animals with azithromycin, an antibiotic used to treat *Chlamydia* infection, resulted in normal arteries that appeared similar to those of uninfected animals.

Mechanistically, molecular mimicry may play a role. A peptide from the murine cardiac specific α myosin heavy chain with sequence homology to a cysteine-rich outer membrane protein of *Chlamydia pneumoniae*, *C. psittaci*, and *C. trachomatis* was demonstrated to induce autoimmune inflammatory heart disease in mice. Injecting the homologous *Chlamydia* peptides into mice was also shown to induce perivascular inflammation, fibrotic changes, and blood vessel occlusion in the heart, in addition to triggering T and B cell immune responses to the homologous endogenous heart muscle specific peptide. *Chlamydia* DNA functions as an immune booster ''adjuvant'' in the triggering of peptide-induced inflammatory heart disease. Infection with *C. trachomatis* led to the production of autoantibodies to heart muscle-specific epitopes. Therefore, it is suggested that *Chlamydia*-induced heart disease is mediated by antigenic mimicry of a heart muscle-specific protein. In other studies, introducing live bacteria by catheter and syringes into mice caused heart inflammation.[15] It is important to note, however, that the inflammation resulting from the mimicry seems confined to the heart muscle rather than extending to the arteries where it could trigger atherosclerotic plaques. One theory postulates an ''echo effect'' which may explain how microbes affect arteries that are some distance from the primary infection sites.[16] This theory suggests that infections release toxins into the bloodstream or display molecules on their surfaces that accelerate inflammatory responses at the blood vessel lining. This work can provide a molecular-based explanation for a link between infections and heart disease if similar inflammation can be shown in humans to play a role in the formation of atherosclerosis.

Many questions remain regarding the potential etiological role of infections in the formation and development of atherosclerosis. It is not clear whether bacteria are direct contributors to atherogenesis or simply innocent bystanders. Would antibacterial therapy be useful for the prevention and treatment of arterial disease? If so, may vaccination be an option for ultimate prevention? Clinical studies are necessary to clarify issues concerning infectious causes of atherosclerosis.

6 METHODS FOR CARDIAC MARKER MEASUREMENTS

6.1 Electrophoresis

Electrophoresis refers to the migration of charged particles in a liquid medium under the influence of an electrical field. Electrophoresis on agarose or cellulose acetate can separate the CK and lactate dehydrogenase (LD) isoenzymes. The isoenzyme bands can be visualized by incubating the support with a coupled reaction and the resulting nicotinamide adenine dinucleotide phosphate (reduced form) (NADPH) can be detected by observing fluorescence following excitation by long-wave (340 nm) ultraviolet (UV) light or by reacting the NADPH formed with a tetrazolium salt to form a colored formazan. Electrophoresis is usually considered a labor-intensive and time-consuming technique.

6.2 Immunoassay

6.2.1 Immunoinhibition

In this technique, anti-CK-M sera inhibit both M subunits of CK-MM and the single M subunit of CK-MB, allowing the determination of enzyme activity of the B subunit of CK-MB. A caveat with this method is that the B subunits of CK-BB, any macro or mitochondrial CK activity, or adenylate kinase (AK) may interfere. However, when CK-BB is present in a specimen, its activity is $<5\,\mathrm{U\,L^{-1}}$, which is below the detection limit of the immunoinhibition technique. However, other sources of interference may also give false-positive CK-MB results.

6.2.2 Two-site ''Sandwich'' Immunoassays

In contrast to immunoinhibition and electrophoresis techniques, which measure the isoenzyme by determination of enzyme activity, immunoassays measure the enzyme as a protein. Immunoassays for enzymes, e.g. CK-MB, are termed mass assays. ''Mass'' assays have no dependence on whether or not the enzyme is catalytically active. Several sandwich immunoassays are currently available for myoglobin, CK-MB, cTnI and cTnT.

6.3 Point-of-care Testing

The fundamental driver of POC testing is that caregivers in triage, treatment, and patient monitoring areas base time-sensitive decisions on the results. Cardiac marker testing has evolved from labor-intensive, rather esoteric assays, that were offered once per day, to systems with a turnaround as fast as 10–15 min once the front-end work of transport, centrifugation, serum–plasma separation, and instrument set-up is completed. The practice of batching these assays, even multiple times per day, may become an anachronism. POC testing may be defined as assays that are performed either directly on the patient, at the bedside, or at ''near-patient'' satellite locations. POC testing of myocardial injury markers may evolve into a standard of care for workup of acute cardiac ischemia. Table 2 lists criteria for POC testing of cardiac markers.

Table 2 Criteria for POC testing of cardiac markers

A strategy for specimen collection and transfer that minimizes risk of infectious disease
Rapid TAT
Low volume, preferably whole-blood sample
Direct application of a noncritical volume or placement of sample directly into instrument
Disposable device or minimal maintenance required
Minimal technical expertise required
Positive identification and specimen tracking strategy that eliminates specimen identification errors
Simple strategy for recording collection time and result reporting
Simple strategy for calibration and quality control
Transferability of data to the Laboratory Information System and/or Hospital Information System
Agreement of results with accepted ''gold standard'' assays
Affordable cost

TAT, turnaround time.

6.3.1 Trends and Limitations for Point-of-care Testing of Cardiac Markers

There is a clear need for better defined strategies that integrate ECG, clinical symptoms, biochemical markers, and other technologies for care of cardiac patients.[17] A model for achieving this integration has been implemented in many institutions through the establishment of areas that focus on care of patients suspected of having acute cardiac ischemia. The National Heart Attack Alert Program (NHAAP), an initiative of the National Institutes of Health (NIH) National Heart Lung and Blood Institute, has recently published document No. 93-3278, which endorses the rapid identification and treatment of AMI patients in emergency department (ED) areas, such as CPECs. The role of POC cardiac marker testing must be evidence-based; real-time or POC testing for cardiac markers may not be appropriate for all aspects of care in the context of cardiac ischemia.[17,18] Specifically, clinical and economic considerations may never warrant POC testing for the initial assessment of ACSs in patients for whom ECG changes are diagnostic of myocardial necrosis. POC testing may be important for noninvasive assessment of reperfusion and for guidance of therapy because TAT is vital if therapy has not been successful, and alternative patient care strategies must be considered in the early phase to ''open the artery''. POC measurement of biochemical markers may also be helpful for assessing infarct artery reocclusion and reinfarction.

7 DECISION LIMITS FOR MARKERS

In the context of cardiac markers or other chemical laboratory tests, interpretation consists of comparison with previous knowledge, ideally from studies conducted in an identical population. These data are best collected through a prospective cohort study design in which the reference standard is applied to every patient, and the results are examined in a blinded fashion. The size and characteristics of population studied, enrollment criteria, and clinical setting in which the studies are conducted should reflect closely the practice environment in which the test will be used. The data must then be analyzed using appropriate mathematical techniques. These techniques and their application in the context of biochemical markers have been reviewed recently.[19]

Calculations of clinical sensitivity, specificity, and predictive values have been the traditional means for assessing the efficacy of tests, although it is important to note that other more effective strategies are available. Clinical sensitivity, specificity, and predictive value calculations require that the outcome be expressed in binary terms as positive, i.e. indicating the presence of a specific outcome, or negative, indicating the absence of the outcome. Analysis of test performance requires designation of decision limits or ''cutoffs'' for examining whether the test data accurately reflect the patient's positive or negative condition. Thus, test results are classified as either true positive (TP), where the test is above the decision limit and the patient has the outcome; true negative (TN), where the test is below the decision limit and the patient does not have the outcome; false positive (FP), where the test is positive but the patient does not have the outcome; or false negative (FN), where the test is below the cutoff but the patient has the outcome. Calculations for clinical sensitivity, specificity, and positive and negative predictive values (NPV) are based on the following:

- sensitivity = TP/(TP + FN)
- specificity = TN/(TN + FP)
- positive predictive value (PPV) = TP/(TP + FP)
- NPV = TN/(TN + FN).

An important issue is that PPV and NPV are dependent on the prevalence of the specific positive outcome in the population tested.

Various methods for deriving cutoffs have been used, including use of the assay's minimum detectable concentration or the 97.5 percentile for either healthy individuals or ''rule-out'' patients. These cutoffs, however, are usually inappropriate because they do not consider performance of the test in the positive outcome population. Receiver operator characteristic (ROC) curves, logistic regression analysis, neural networks, and calculation of likelihood ratios are all better evaluation techniques for comparing a patient result to the previous knowledge.

For the construction of ROC curves, data are typically displayed for the outcome groups. The data are examined

at a continuum of different cutoffs, with determination of the TP, FP, TN and FN results at each; these results are used to calculate clinical sensitivity and specificity at each individual decision limit. The ROC curve is constructed by plotting sensitivity on the *y*-axis and (1 – specificity) on the *x*-axis. The ideal test has an area under the ROC curve of unity.

ROC curves are an effective means for both determining decision limits and for assessing the relative accuracy of tests.[20] The upper left corner of the ROC curve indicates the minimum tradeoff of FP and FN results; this point is frequently considered the optimum decision point.[20] It is important to note when examining cutoffs with ROC curves that there are often insufficient data in the area of the optimum cut off to provide a high degree of certainty.

ROC curves can also be used for comparing two or more tests examined in the same population. Areas under the ROC curves for each test can be calculated and compared for differences to show if one test is better than another for diagnostic performance.[20]

8 SPECIFIC MARKERS OF ACUTE CORONARY SYNDROMES

8.1 Characteristics of an Ideal Cardiac Marker

The ideal marker would have absolute cardiac specificity to allow reliable diagnosis of myocardial injury in the presence of skeletal muscle injury. An ideal marker should not exist in other tissues under normal or pathological conditions. It must have high diagnostic sensitivity: the marker should be enriched in cardiomyocytes to indicate even minor damage such as in unstable angina and myocarditis. It should be able to differentiate between reversible (ischemia) and irreversible (necrosis) damage. In AMI, the marker should permit monitoring of reperfusion therapy and the estimation of infarct size and prognosis and should allow both early and late diagnosis. The marker should be stable and appear in blood immediately following damage, but remain in circulation long enough to allow detection then be eliminated from circulation to allow detection of reinfarction. A rapid (whole blood is desirable), easy to perform, quantitative, and cost-effective assay should be available for this marker. It is important to note that no such marker is known at present. Finding a single marker with all these characteristics could be impossible, so the use of a panel of markers may be necessary.

8.2 Myoglobin

Myoglobin (17 800 Da) is a heme protein that is located in the cytoplasm of both cardiac and skeletal muscle

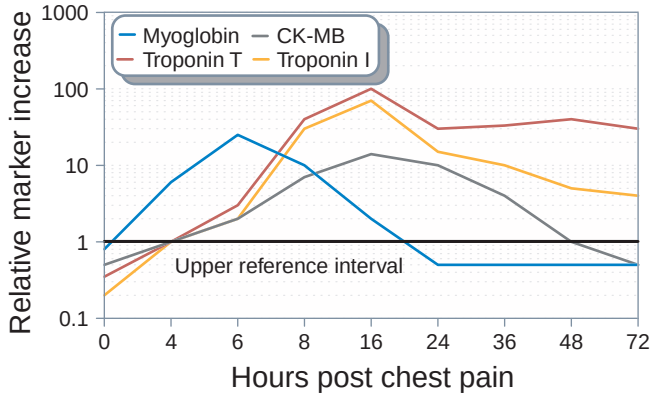

Figure 3 Relative marker increase after MI. Markers are expressed as multiples of the reference interval. Thus the relative increase will vary depending on the normal reference interval utilized. Note that the timescale on the *x*-axis is not linear and that the scale on the *y*-axis is logarithmic.

cells. As shown in Figure 3, myoglobin is among the earliest appearing biochemical markers routinely available for assessment of the ACS after myocardial damage. Myoglobin's characteristic early rise and rapid clearance are mainly due to its relatively small molecular size, high concentration in tissue, and renal clearance mechanism. Because myoglobin's amino acid sequence is the same in both cardiac and skeletal muscle, this marker has decreased clinical specificity due to positive results in populations having skeletal muscle injury. Also, patients will have elevated myoglobin levels due to decreased clearance from blood. Myoglobin elevation in noncardiac disease indicates why the clinical specificity of this marker depends on the population tested. The clinical sensitivity of myoglobin is accepted to exceed 90–95% within 6 h of presentation. The role for myoglobin in testing guidelines focuses on early clinical sensitivity, with rapid availability of results. As indicated in Table 3, there are a number of studies that have examined the NPV of myoglobin, finding satisfactory high values when chest pain onset is acute. Based on these data, testing strategies at many medical centers include myoglobin, in combination with more myocardial specific markers, because of its high NPV. On the other hand, many institutions believe that myoglobin adds very little information, and in fact the NHAAP report referred to above rated the clinical impact of myoglobin as "+", indicating only modest accuracy and small clinical impact.[21]

In summary, myoglobin has high NPV in the early hours after MI. Serial sampling with optimum timing at presentation and between 2 and 6 h later is necessary to optimize clinical utilization. Not all clinicians are convinced that myoglobin measurements contribute significantly to the assessment of ACS patients. To be useful, myoglobin

Table 3 Various studies that have examined the NPV of myoglobin measurements: a total of 972 patients were included in these studies, 218 (22.4%) being diagnosed as having MI

Study	No. of patients (MI)	NPV (%)
G.X. Brogan, Jr, S. Friedman, C. McCuskey, D.S. Cooling, L. Berrutti, H.C. Thode, J.L. Bock, 'Evaluation of a New Rapid Quantitative Immunoassay for Serum Myoglobin versus CK-MB for Ruling Out Acute Myocardial Infarction in the Emergency Department', *Ann. Emerg. Med.*, **24**, 665–671 (1994).	189 (22)	97
J.F. Tucker, R.A. Collins, A.J. Anderson, M. Hess, I.M. Farley, D.A. Hageman, H.J. Harkins, D. Zwicke, 'Value of Serial Myoglobin Levels in the Early Diagnosis of Patients Admitted for Acute Myocardial Infarction', *Ann. Emerg. Med.*, **24**, 704–708 (1994).	133 (22)	97
V. Bhayana, A.R. Henderson, 'Markers for Early Diagnosis of Myocardial Infarction', *Lancet*, **342**, 1554 (1993).	173 (81)	92
M.C. Kontos, P. Anderson, C.M. Hanbury, C.S. Roberts, W.G. Miller, R.L. Jesse, 'Use of the Combination of Myoglobin and CK-MB Mass for the Rapid Diagnosis of Acute Myocardial Infarction', *Am. J. Emerg. Med.*, **15**, 14–19 (1997).	101 (20)	97
M.A. Levitt, S.B. Promes, S. Bullock, M. Disano, G.P. Young, G. Gee, D. Peaslee, 'Combined Cardiac Marker Approach with Adjunct Two-dimensional Echocardiography to Diagnose Acute Myocardial Infarction in the Emergency Department', *Ann. Emerg. Med.*, **27**, 1–7 (1996).	190 (21)	99
G.J. Kost, J.D. Kirk, K. Omand, 'A Strategy for the Use of Cardiac Injury Markers (Troponin I and T, Creatine Kinase-MB Mass and Isoforms, and Myoglobin) in the Diagnosis of Acute Myocardial Infarction', *Arch. Pathol. Lab. Med.*, **122**, 245–251 (1998).	97 (27)	94
C. Montague, T. Kircher, 'Myoglobin in the Early Evaluation of Acute Chest Pain', *Am. J. Clin. Pathol.*, **104**, 472–476 (1995).	89 (25)	100

results must be available in real time, i.e. with an available testing TAT of approximately 0.5–1 h, in a time frame useful for decisions in the early assessment of patients. The major method used for myoglobin measurement is two-site immunoassay.

8.3 Creatine Kinase-MB Isoenzyme

CK (85 000 Da) is an enzyme located in the cytoplasm of all striated muscle cells where it catalyzes the phosphorylation of creatine to creatine phosphate. Cytoplasmic CK is dimeric, composed of M and/or B subunits that associate to form CK-MM (CK-3), CK-MB (CK-2), and CK-BB (CK-1) isoenzymes. At present, CK-MB is an important tool in the evaluation of ACSs because it has been established as a benchmark for biochemical markers of MI due to experience spanning more than two decades.[7] In patients having significant myocardial disease, the CK-MB isoenzyme comprises approximately 20% of the total CK in this tissue. CK-MB is diagnostically sensitive for myocardial injury, but skeletal muscle has both higher total CK activity per gram of tissue and may be comprised of up to 3% CK-MB. This potentiates compromised diagnostic specificity, particularly in patients with concomitant myocardial and skeletal muscle injury. To confer greater cardiac specificity to CK-MB measurements, a CK-MB ''index'' is frequently calculated according to Equation (1):

$$\text{CK-MB index} = \frac{\text{CK-MB}}{\text{CK-total}} \times 100\% \qquad (1)$$

Some authors suggest that CK-MB index values exceeding 2.5% are associated with a myocardial source of the MB isoenzyme; however, a recent review shows that the index is reportedly as low as 2% and as high as 5% depending on the variability of both the numerator and denominator.

As indicated in Figure 3, the first rise in CK-MB after MI requires 4–6 h after onset of symptoms, but serial sampling over a period of 8–12 h is required for diagnosis with high diagnostic sensitivity and specificity. CK-MB mass assays were examined in a meta-analysis for the retrospective diagnosis of MI within 12–48 h after onset of symptoms or admission.[22] This analysis showed that CK-MB had a clinical sensitivity of 96.8% [95% confidence interval (CI) = 95–98%] and a clinical specificity of 89.6% (CI 87–92%). Also, the NHAAP evaluated cardiac marker utilization focusing on the ED environment.[21] When used with multiple sampling, CK-MB testing was very accurate for the diagnosis of MI and was noted to have a large clinical impact.[21] Despite this excellent clinical performance, CK-MB is not an ideal marker because its rise requires 8–12 h after symptom onset for use in diagnosis and tissue specificity is an issue.

Methods for CK-MB measurement include electrophoresis, immunoinhibition, and so-called ''mass''

assays, which are two-site immunoassays. Most of the mass assays utilize the Conan antibody developed at the University of Washington, St. Louis, MO, USA.[23] Specific determinations of CK-MB require the application of the ''sandwich'' technique in which two antibodies (anti-M subunit, anti-B subunit, or anti-MB) with affinities for different epitopes of the CK-MB molecule are used sequentially. The first antibody (usually a monoclonal) is immobilized on a matrix, while the other antibody is conjugated with a label reporter molecule. CK-MB mass assays are automated, highly sensitive ($<1\,\mu g\,L^{-1}$), and specific, and have a rapid TAT (as short as 7 min), giving them characteristics that surpass other technologies for measurement. A number of mass assays, using various labels, are commercially available and used for routine and emergency measurement of CK-MB. These methods require the use of CK-MB calibrator standards, and there are no agreed-on values or materials. Recently, the American Association for Clinical Chemistry proposed a recombinant CK-MB material for standardization of CK-MB mass assays.[24]

8.4 Creatine Kinase Isoforms

CK isoforms are variants of CK-MM and CK-MB isozymes, which are generated upon irreversible cleavage of the terminal lysine of the M-subunit in vivo.[25] Because CK-MM is a dimeric M, there are three possible isoforms: MM3 (unmodified), MM2 (one subunit modified), and MM1 (both subunits modified), based on their relative migration toward the anode during electrophoresis. MM3 is the tissue isoform that exists in both skeletal muscle and heart. Following muscle damage, CK-MM3 is released into circulation where serum carboxypeptidase cleaves the carboxy terminal lysine residue of one of M-subunit to produce CK-MM2. Finally, the second M-subunit will lose its C-terminal lysine to generate the serum isoform CK-MM1.

CK-MB1 and CK-MB2 are isoforms of the isoenzyme CK-MB. CK-MB2 is released from cardiac tissue and CK-MB1 is the product formed when the terminal lysine of the M subunit of CK-MB2 is cleaved by carboxypeptidase B.[26] The resulting isoforms have different electrophoretic activities and can be assayed using high-voltage electrophoresis. The CK-MB2/CK-MB1 ratio becomes abnormal (>1.5) before CK-MB levels rise above normal. Studies have shown diagnostic sensitivities of up to 92% for detecting AMI at 4–6 h.[27] FP results may occur when CK-MB2 becomes elevated as a result of skeletal muscle injury.[28]

A recent diagnostic marker cooperative study included 995 chest pain patients and showed that CK isoforms were most sensitive and specific (91 and 89%) within 6 h of onset followed by myoglobin (78 and 89%).[29] In unstable angina patients, CK-MB isoforms were increased in 29.5%, myoglobin in 23.7%, troponin I (TnI) in 19.7%, and troponin T (TnT) in 14.8%. For early diagnosis of MI, CK-MB isoforms and myoglobin were concluded to have the best sensitivity.[29]

Helena Laboratories (Beaumont, TX, USA) have developed an automated Cardio REP system to quantify CK isoforms. This system can assay six specimens within ~18 min. This rapid TAT is achieved by automation of pipetting, electrophoresis, substrate spreading, and incubation, gel drying, and the use of high-voltage electrophoresis and computer-controlled densitometry. CK-MB isoforms migrate anodic to those of CK-MM; CK-MB1 is the fastest moving, most anodic band of the MB series.

8.5 Lactate Dehydrogenase

LD, an important enzyme of glucose metabolism, exists as a tetramer of approximately 135 kDa. Mammals have two different types of subunits, M and H, which together form five isoenzymes, H4, MH3, M2H2, M3H, and M4. According to their electrophoretic mobility (from anode to cathode), these isoenzymes are referred to as LD1, LD2, LD3, LD4 and LD5, respectively. In the aerobic heart tissue, the H4 tetramer predominates, but the M4 isoenzyme predominates in skeletal muscle. LD exists in almost every tissue and is abundant in skeletal muscle, heart, brain, kidneys, lungs, and erythrocytes.[30] LD isoenzymes differ in their stability: LD4 and LD5 are unstable at temperatures $> 45\,^{\circ}C$ or at $-20\,^{\circ}C$, whereas LD1 is stable up to $65\,^{\circ}C$. Electrophoresis followed by densitometry is the method of choice for the separation of LD isoenzymes and the detection of LD1–LD2 flip in AMI. Following onset of chest pain, LD activity starts to rise by 8–12 h, peaks at 24–72 h, and returns to the baseline by 8–14 days. Normally LD1 does not exceed 40% of total LD activity. Following myocardial damage, however, LD1 activity exceeds LD2 activity, and the LD1/LD2 ratio is fairly specific for myocardial injury. Determination of LD isoenzyme adds little to the diagnosis of AMI if elevated CK-MB levels have already been detected. Similarly, LD isoenzymes that have low tissue specificity are no longer the marker of choice for detection of late-presenting MI, because other cardio-specific markers such as troponins that have long half-lives in blood are better suited for this purpose.

LD1 has been used to assess infarct size in AMI patients by calculating the cumulative enzyme release.[31] A decrease of the enzyme activity that corresponds to a reduction of infarct size after thrombolysis was observed in patients with ST elevation in the initial ECG.[31] (ST segment is a component of the ECG complex.) Although

LD1 is enriched in erythrocytes, free hemoglobin levels can be used to correct for LD1 activities contributed by erythrocytes.

Immunoinhibition methods and electrophoresis followed by densitometry have been used for the separation of LD isoenzymes and the detection of LD1–LD2 flip in AMI patients. With the wide availability of troponin assays, LD measurements for MI diagnosis are considered an anachronism.

8.6 Proteins of the Troponin Complex

The troponin complex is located on the thin filament of the contractile apparatus in striated muscle and consists of three subunits: TnT, which binds to tropomyosin; troponin C (TnC), which is the calcium-binding subunit, and TnI, which binds to actin.[32] Specific isoforms of TnT and TnI are found in both cardiac and skeletal muscle. The isoforms found in cardiac muscle (cTnT and cTnI) are encoded by different genes to those found in skeletal muscle, resulting in slightly different amino acid sequences between the two forms. This makes it possible to produce antibodies that are specific for only the cardiac forms.[33,34] Because TnC has identical amino acid sequence in both skeletal and cardiac tissues, it has little potential as a cardiac marker.[35]

TnI (23 500 Da) is the inhibitor of the troponin–tropomyosin regulatory complex that confers calcium sensitivity to actomyosin adenosine triphosphatase (ATPase) activity in striated muscle, thus modulating the interaction of actin and myosin in these muscles.[36] cTnI has a unique amino acid sequence that shows 40% dissimilarity from the two skeletal muscle isoforms (slow-twitch and fast-twitch), making this protein a cardiac-specific marker for myocardial damage. cTnI is not expressed in skeletal muscle at any developmental stage or in response to any pathological stimuli.[37] Accordingly, elevations do not occur in plasma from patients with acute or chronic skeletal muscle disease unless acute myocardial injury is concomitant.

TnT (37 000 Da) is an essential component of the contractile apparatus that binds the troponin complex to the tropomyosin strand of the thin myosin filament of striated muscle. Antibodies that specifically recognize the amino acid sequence distinct for cTnT were developed and are the basis for quantitative measurement by immunoassay.

Both cTnT and cTnI rise above normal detectable limits about 3 h after symptoms begin. Depending on the method, cTnI may remain elevated for as long as 3–10 days and cTnT for as long as 2 weeks.[38] Owing to these prolonged elevation periods, both cTnI and cTnT allow for diagnosis many days after MI has occurred. Problems may arise when trying to diagnose reinfarction, however, as serum levels may not have had adequate time to decline. cTnT has also been shown to be useful in following patients with unstable angina. As minor cell necrosis occurs, cTnT is released into the circulation. In patients with unstable angina, cTnT is either borderline or increased, and a substantial number of all those with elevated levels later develop an MI.[39]

Although the release of the C–T–I troponin complex after myocardial injury is not fully characterized, several reports suggest that a large portion of TnI is liberated in the form of complex (predominantly with TnC) and only a small portion of TnI circulates in a free form.[40] Few data are available on the release of TnT subunits into circulation after AMI. The release kinetics of troponin proteins are crucial especially when comparing troponin results of different commercial assays for these proteins. For the generation of monoclonal or polyclonal antibodies, highly purified proteins are usually used. However, TnI tightly interacts with the other two proteins of the troponin complex and separation of these proteins requires a high concentration of urea (6–8 mol L^{-1}).[41] The exact structure of purified TnI in solution is different from that of TnI complexed with the other troponin proteins. Furthermore, TnC significantly alters the phosphorylation of TnI by different kinases. Finally, TnC and TnT affect the accessibility of the lysine residues of TnI to chemical modifications. All these factors indicate that the interaction of TnC and TnT with TnI will dramatically affect the overall structure of TnI. Consequently, antibodies raised against purified TnI may not be able to access epitopes on TnI that are modified in the whole troponin complex or in the binary complexes of TnI with TnC or TnT. Therefore, for developing a reliable TnI immunoassay, antibodies should be able to recognize both free TnI and TnI complexed with other troponin proteins.

8.7 Carbonic Anhydrase Isoenzyme III

Carbonic anhydrase (CA) exists in seven closely related isoforms in mammalian tissue.[42] Two of the most abundant of these isozymes are CA I and CA II; both are found in erythrocytes with CA I in highest concentrations. CA II is also found in the cytosol of cells from many organs, including brain, liver, kidney, and lungs. Each of these isozymes shares approximately 55% amino acid sequence homology with CA III.[43] This increases the importance of using a monospecific detection antibody in any assay system used for CA III quantification.

CA III is a cytosolic protein (28 000 Da) found almost exclusively in type I (slow-twitch) skeletal muscle fibers and is a useful marker for many different types of muscle damage.[44,45] Reports have stated that it is a more sensitive marker of muscle damage than CK.[46] It can be used to monitor and diagnose patients with Duchenne

muscular dystrophy and many other neuromuscular disorders.[44] Recently, there has been increased interest in monitoring CA III along with myoglobin levels to evaluate patients for the possibility of AMI. In the event of skeletal muscle damage, myoglobin and CA III are released into the plasma by damaged cells in a fixed ratio (~3 : 1).[45] This ratio is measurably increased in the case of an AMI. Because CA III is absent in cardiac muscle the ratio begins to rise as soon as myoglobin is released by the damaged cardiac cells. Recent studies have indicated that the use of the myoglobin/CA III ratio was more specific than CK-MB in diagnosing an AMI.[45,47] This is due to the early release of myoglobin from injured myocardial cells. Peak serum levels are reached as fast as 1–4 h after AMI. Commercial diagnostic CA III assays are not yet available.

8.8 Human Heart Fatty Acid-binding Protein

Fatty acid-binding proteins (FABPs) are low-molecular weight (15 000 Da) cytosolic proteins which were first identified in the intestinal mucosa of rats.[48] Since this discovery, several different forms have been found, including one which is abundant in the cytoplasm of human cardiomyocytes [human heart fatty acid-binding protein (HHFABP)]; HHFABP is also found in other body tissues.[49] As quickly as 3 h following AMI, plasma concentrations of HHFABP have been shown to increase significantly.[50] The initial problem with using HHFABP as a diagnostic indicator of AMI is that it is not specific to cardiac tissue. This problem may have been solved by recent work, which uses the ratio of myoglobin to HHFABP to distinguish between cardiac and skeletal muscle damage.[51] The study indicates that the ratio of myoglobin to HHFABP following cardiac injury is significantly different than the ratio following skeletal muscle damage. Another problem with measuring HHFABP occurs in patients with chronic renal failure. HHFABP is cleared by the kidney and, as a result, patients with chronic renal failure can have plasma levels much higher (20–25-fold) than those in normal patients.[52] Immunoassay methods are used for HHFABP determination.

8.9 Markers of Platelet Activation

Platelet activation is important in the mechanism of thrombus formation and the mechanism of ACS. Indicators of platelet activation such as platelet function assays or P-selectin measurement may help assess a patient's tendency for intracoronary thrombosis. Platelet activation can result from contact with exposed collagen, thrombin and/or other agonists induced by plaque disruption. P-selectin is an adhesion molecule expressed on the surface of activated platelets.[53] Expression of this protein is increased in patients with symptomatic coronary artery disease.[53] Thus P-selectin may be a marker of activated or ''angry'' platelets that indicates their tendency to adhere to leukocytes causing accumulation and consequent thrombotic complications in the ischemic myocardium.[53]

8.10 Acute Phase Proteins

Increased levels of the acute phase proteins CRP and SAA are known to be nonspecific, but may have a role in identifying patients having unstable coronary plaques. Studies have investigated the use of CRP for predicting unfavorable outcome and impairment of left ventricular function resulting from acute cardiac necrosis or previous MI.[54] The increase in acute phase proteins may indicate plaque disruption that causes release of cytokines from activated monocytes and macrophages at the disrupted site.[55] Among other systemic effects, cytokines including interleukin-6 promote hepatic synthesis of the acute phase proteins. In this way, patients having unstable coronary artery disease who are at increased risk have abnormal circulating concentrations of acute phase proteins.[56,57] A possible component of the observed association between acute phase proteins and increased risk is that these proteins may reflect infectious disease affecting the coronary vessels.[58] In any case, aspirin or other nonsteroidal anti-inflammatory agents may reduce risk in coronary artery disease patients presumably by inhibiting the inflammatory process.[54]

8.11 Markers of Thrombosis

Thrombus formation is fundamental to blockage of the infarct related artery. Therefore, markers of thrombosis, including soluble fibrin and fibrin degradation products, may also reveal a recent thrombotic process or risk of an impending event. Although not sensitive or specific enough to diagnose MI, soluble fibrin and cross-linked fibrin degradation are increased in patients who are at higher risk for complications.[59] Physiologically, these markers are thought to indicate increased fibrinolysis prior to development of MI.[59]

8.12 Markers of Early Myocardial Ischemia

A marker that is indicative of myocardial ischemia occurring prior to frank necrosis would be helpful for locating a patient's position on the ACS continuum. Glycogen phosphorylase b is a key enzyme of glycogenolysis. It has three major isoenzymes: BB (brain and heart), MM (muscle), and LL (liver). GPBB is the predominant in the heart and has been found to be a sensitive indicator of myocardial injury.[60] GPBB appears to increase upon ischemia without necrosis and may represent a

marker of ischemia rather that necrosis.[61] GPBB release is thought to be linked to the sudden burst of glycogenolysis which occurs in the injured myocardium following acute MI. A rise in GPBB levels can be detected as early as 3–6 h after the onset of symptoms. In a recent study, Rabitzsch et al. were able to develop an immunoenzymometric assay to measure GPBB.[61] In testing patients during the first 4 h after the onset of symptoms, they reported GPBB to be more sensitive than CK, CK-MB mass, myoglobin, and cTnT. Research on GPBB as a biochemical marker for MI is still ongoing but preliminary results are encouraging.

Taken together, the markers of plaque rupture (CRP and SAA), indicators of intracoronary thrombosis (P-selectin and soluble fibrin), myocardial ischemia (GPBB), and markers of necrosis could be combined with clinical indicators, ECG, echocardiogram, and imaging studies to form an integrated combined model for optimum assessment of patient risk.

9 APPLICATIONS FOR CLINICAL MEASUREMENTS

9.1 Diagnosis of Myocardial Infarction

Retrospective confirmation of MI is the classical role for biochemical markers of myocyte damage. Compared with CK-MB, troponin measurements offer substantial advantages to the detection of myocardial injury with respect to sensitivity and specificity. Feng et al. reported cTnI to be the most sensitive analyte in a swine model with induced severe coronary artery stenosis.[62] Electron microscopy showed frequent TnI release during early stage myocardial ischemia and even from reversibly damaged myocardium.[63]

The release kinetics of TnT and I are similar; both are released within 4–12 h after myocardial necrosis, with a peak value 12–48 h from symptom onset, affected by the duration of ischemia and reperfusion.[63,64]

Because of the known delay in release, troponins are not generally considered early markers of MI.[65–68] For greatest clinical sensitivity, serial sampling is recommended at least 12 h after onset of pain. Furthermore, troponin elevations due to minor myocardial necrosis, but with MI ruled out by WHO criteria, have been viewed as FP affecting diagnostic specificity.[69] However, even in the absence of typical ST-segment changes or clinical symptoms, troponins can reliably diagnose acute MI. Patients with elevated troponin levels on admission are more likely to experience complications during follow-up.[67,70] Comparison of TnT and I showed similar clinical sensitivity even when used with different cutoffs;[71] therefore, both have equal potential to rule out MI within an acceptable time interval and allow early and cost-effective triage decisions in chest pain patients.[72]

A temporal difference in clinical sensitivity between cTnI and CK-MB mass was demonstrated in several studies including a study based in the Emergency Medicine area.[71] Use of a cTnI decision point of 1.5 ng mL^{-1} showed significantly lower clinical sensitivity compared with CK-MB mass at both 2 h ($P = 0.0001$) and 6 h ($P = 0.00016$) after presentation. The use of a lower cutoff of 0.6 ng mL^{-1} also showed statistically significant differences between cTnI and CK-MB at 2 h ($P < 0.0238$) and 6 h ($P = 0.025$). In the 12–24 h time frame, the differences in comparative sensitivities did not achieve statistical significance. Clinical specificity for MI diagnosis exceeded 97% for cTnI in all time frames examined.[71] For diagnosis of MI, serial sampling is necessary unless results of the first measurement exceed the decision point. The data indicate that cTnI sampling to at least 12 h after onset of symptoms should be performed to optimize clinical sensitivity for MI diagnosis.

Recently, the specificity of TnT to myocardial damage has been questioned; however, clinical data from studies on TnT re-expression in regenerating human skeletal muscle confirm troponins as superior markers of myocardial injury.[73,74] Various researchers suggest that TnT or I should replace CK-MB for detecting acute MI (Table 4).[63,65–68,71,75–87]

A meta-analysis showed that cTnT had a clinical sensitivity of 98.2% (CI 97–99%) for diagnosis of MI 12 h following onset of symptoms.[22] This clinical sensitivity was not significantly different from results for CK-MB mass. However clinical specificity for cTnT was 68.8% (CI 66–72%), which was significantly different from CK-MB mass ($P < 0.001$). The reason for this difference was the inclusion of patients with minor myocardial injury associated with unstable angina. These unstable angina patients had cTnT results that exceeded the decision limits examined in this analysis. As will be presented later, these positive cTnT results are important because they identify unstable angina patients at increased risk for adverse cardiac events. Also, since this meta-analysis was performed, even more cardiac-specific second- and third-generation versions of the cTnT assay have become available.[34]

Utilization of cTnT for MI diagnosis with optimum sensitivity requires samples collected 8–12 h after onset of symptoms. Owing to cTnT's elevation in minor myocardial injury, clinical specificity in some studies was not as high for cTnT as CK-MB. However, subsequent studies using short-term mortality as the outcome have shown that these positive results are meaningful for the identification of high-risk patients. Very myocardial-specific cTnT immunoassays are available.[34]

Table 4 Detection of MI by cTnT or cTnI. (Adapted from B.U. Goldmann, R.H. Christenson, C.W. Hamm, M. Ohman, 'Can Troponin be the New Standard for Detection of Myocardial Damage? A Review of Recent Investigations', *Clin. Chem.*, in press.)

Study	Year	*n*	Marker	Entry Criteria	Sensitivity (%)	Specificity (%)
Katus et al.[63]	1991	388	cTnT	MI	100	93
Ravkilde et al.[65]	1993	298	cTnT	MI	100	74
De Winter et al.[66]	1995	309	cTnT	Chest pain	<4 h: 34	91
Stubbs et al.[67]	1996	240	cTnT	MI < 48 h	Admission: 45	NA
Mair et al.[68]	1995	37	cTnT	MI < 4 h	46	NA
			cTnI		49	NA
Antman et al.[81]	1995	100	cTnT	Chest pain on admission	33	95
Mach et al.[82]	1995	32	cTnT	MI	100	86
Panteghini et al.[83]	1998	101	cTnT	Chest pain	73	100
Gerhardt et al.[84]	1997	487	cTnT	Chest pain	87	100
REACCT[85]	1998	721	cTnT	Chest pain	70	100
			cTnT		80	100
Adams et al.[100]	1994	188	cTnI	MI + no MI	97	94
D'Costa et al.[86]	1997	316	cTnI	MI	97	98
Christenson et al.[76]	1998	208	cTnI	MI + no MI	88	91
Apple et al.[78]	1999	192	cTnI	Chest pain	98	100
Zaninotto et al.[87]	1999	26	cTnI	MI + no MI	<6 h: 81	90
Falahati et al.[80]	1999	327	cTnI	Chest pain	100	96.3
Apple et al.[77]	1999	633	cTnI	MI + no MI	77	94

MI = myocardial infarction (diagnosis according to WHO criteria); NA = not available; no MI = myocardial infarction excluded (WHO criteria).

9.2 Reinfarction

The magnitude of infarcted tissue is an important determinant of early and late morbidity and mortality.[88] Because early reinfarction increases the cumulative extent of injury, recurrent infarction is suggested to have deleterious effect on prognosis.[88] However, determination of the frequency and severity of reinfarction as well as patients at particularly high risk has been difficult. Clinical criteria commonly used to identify recurrent MI (also known as infarction extension) include severe prolonged chest pain and new ST-T wave electrocardiographic changes. (T wave is a component of the ECG complex.) However, these criteria have low specificity for reinfarction and may reflect other conditions such as pericarditis. Based on ST-T wave changes, the apparent incidence rate of reinfarction was reported to be as high as 86%.[89] However, the reported incidence rate is <20% as judged from morphological criteria among the subgroup of patients who succumb early after infarction.[90] Using biochemical markers, such as repetitive evaluation of plasma CK activity, the apparent rate of reinfarction ranged between 34 and 65%.[91,92] Such estimates may have been falsely elevated by frequent contributions to plasma CK from noncardiac sources.

Marmor et al. characterized the frequency and nature of recurrent infarction prospectively in 200 AMI patients.[93] In addition to clinical and electrocardiographic criteria, they evaluated serial changes in plasma CK-MB activity and plasma myoglobin concentrations in all patients and radioventriculograms in selected patients to delineate the incidence of early reinfarction and identify features of patients at high risk. Among the patients studied, the overall incidence of early recurrent infarction was 17%, based on confirmation of recurrence of plasma CK-MB and plasma myoglobin elevations after both had returned to baseline levels following the initial infarction. It was observed that recurrent infarction was often manifested by regional ventricular deterioration. Ventricular function, as assessed with radioventriculography, in patients who did not undergo reinfarction either remained the same or improved slightly during the hospital stay. Recurrent infarction was observed in 43% of patients with initial subendothelial MI compared with only 8% of those with a transmural MI.[93] The authors observed close correlations between (1) the magnitude of peak CK associated with the initial infarction and peak CK after early recurrent infarction and (2) sites of electrocardiographic and radioventriculographic changes with each of the two episodes. Therefore, the initial site of coronary arterial obstruction may be involved in both episodes and transitory improvement of flow may leave some myocardium at risk for a second wave of necrosis in proportion to the extent of initially jeopardized tissue.

9.3 Perioperative Myocardial Infarction

MI is the most common cause of morbidity and mortality in patients who have had noncardiac surgery.[94] In patients with perioperative MI, mortality ranges between 36 and 70%.[95,96] However, it is difficult to detect perioperative cardiac damage, because most episodes of

myocardial ischemia occur without changes in the heart rate or blood pressure.[97] Although CK-MB has been the marker of choice for diagnosis of MI, elevated CK-MB levels can sometimes occur after surgery in the absence of apparent cardiac injury.[98] This is because small amounts of CK-MB exist in healthy skeletal muscle, and therefore such FP elevations have been attributed to skeletal muscle injury occurring during surgery.[99]

Data from one study indicated that patients undergoing vascular ($n = 96$) or spinal ($n = 8$) surgery had serial measurements of cTnI that were accurate in detecting perioperative MI or excluding the diagnosis.[100] There was concordance between the development of abnormalities in segmental wall motion, as detected by echocardiography, and elevations in cTnI in 107 of the 108 patients included in this study. In contrast, CK-MB was elevated in 19% of patients without perioperative cardiac injury, including eight of the 12 patients undergoing spinal surgery. FP elevations of CK-MB were more common than TP elevations among patients undergoing vascular surgery. Many cases of perioperative elevations of CK-MB are suggested to result from skeletal muscle rather than a cardiac injury. Since elevations of cTnI persist for 5–7 days after MI, the use of this marker to diagnose perioperative MI may demand one of two diagnostic strategies: (1) to measure cTnI only if MI is suspected and CK-MB levels are elevated, or (2) to obtain a preoperative cTnI value for comparison with postoperative values.

9.4 Reperfusion Assessment

The Thrombolysis in Myocardial Infarction (TIMI) study group has established a grading scale for the extent of patency: TIMI 0, no anterograde flow beyond the point of occlusion; TIMI 1, penetration without perfusion; TIMI 2, partial perfusion; and TIMI 3, complete perfusion.[101] The TIMI grading scale is the ''gold standard'' in characterizing the success of biochemical markers and other indicators of coronary artery reperfusion status. The ability to evaluate intravenous thrombolytic therapy is important, with successful reperfusion defined as the establishment of full anterograde flow through the infarcted artery (TIMI grade 3 flow).[102] Although coronary angiography is the gold standard for assessing patency, occlusion or reocclusion, this invasive procedure has risks that limit its routine use in all patients receiving thrombolytic therapy.[103] Cardiac markers present a noninvasive tool to assess reperfusion. However, for reliable prediction of reperfusion, serial testing of cardiac markers to identify the initial rate of increase or the time to peak is needed.

Myoglobin shows an early and rapid increase after successful reperfusion compared with nonperfused patients. Abe et al. showed that the rate of increase of myoglobin (9.7-fold) is higher than that for total CK (2.8-fold) in the reperfused group of patients within 26 min of commencing thrombolysis.[104] An approach that combined myoglobin, CK-MB, and clinical indicators has demonstrated success in assessing successful thrombolysis.

TnT has characteristic biphasic release kinetics: an early peak containing the cytosolic form which is affected by the washout after reperfusion and a later peak containing the bound form which is not affected by reperfusion. One study showed that the probability of reperfusion is >95% when the ratio of peak cytosolic cTnT to the concentration at 38 h is >1.42 or the ratio of cTnT concentration at 14 to 38 h is >1.09.[105]

In successfully reperfused AMI patients, cTnI increased 30, 60, and 90 min following thrombolytic therapy compared with the baseline and this increase was significant compared with nonperfused patients.[106] Preliminary studies suggest that the calculation of the relative index of reperfusion at 90 min after thrombolytic therapy (cTnI at 90 min divided by cTnI at 0 min baseline) showed ~80% sensitivity for detection of reperfusion.[106]

9.5 Risk Stratification/Therapeutic Intervention

Risk stratification of ACS patients based on cardiac markers is an area of great potential. Also, recent studies suggest that cTnT and cTnI may be useful for therapy selection or guidance of intervention.[107,108]

cTnI was measured at enrollment for 1404 patients enrolled in the TIMI IIIb study using the Stratus method.[70] The TIMI IIIb population represented the unstable angina/non-Q-wave MI part of the ACS spectrum. Of this total population, 845 patients presented >6 h after onset of clinical symptoms, and were the focus of much of the analysis. The cTnI decision point used in this study was 0.4 ng mL^{-1}, the minimum detectable concentration of the assay. This study showed that for each increase of 1.0 ng mL^{-1} in cTnI, there was a significant increase in risk of 42-day mortality after correction for baseline variables of age and ST-segment depression.[70] Overall conclusions noted that a single cTnI measurement at enrollment was an independent risk factor for 42-day mortality and that the timing of cTnI measurement was critical. It is of note that there was no significant difference in mortality risk between patients with undetectable values, i.e. <0.4 ng mL^{-1} and those with values between 0.4 and 1.0 ng mL^{-1}.[70] An important finding was that among patients presenting >6 h after onset of symptoms, TnI was prognostic when the CK-MB mass was within the reference interval. This is presumably because CK-MB has insufficient sensitivity and specificity for microinfarction.[109]

cTnT was examined at enrollment in a GUSTO (global use of strategies to open occluded coronary arteries)

Table 5 Ability of various cTnT measures to predict 30-day and 1-year mortality

Variable(s) in logistic regression model	30-day mortality chi-squared	30-day mortality *P* value	1-year mortality chi-squared	1-year mortality *P* value
Baseline cTnT alone	8.96	0.0113	10.00	0.0067
What 8 h adds to baseline	3.08	0.0792	10.05	0.0015
What 16 h adds to baseline	4.56	0.0382	12.40	0.0004
Baseline + 8 h + 16 h	13.95	0.0075	22.42	0.0002
What 8 h adds to baseline + 16 h	0.43	0.5144	0.02	0.8802
What 16 h adds to baseline + 8 h	1.90	0.1679	2.37	0.1233

IIa substudy that included 801 ACS patients using a primary end-point of 30-day mortality.[110] It is of note that these patients represent a relatively high risk subset of the ACS continuum, as about 90% had MI within 18 h of enrollment.[110] In these patients, cTnT measurement at a median of 178 min after onset of symptoms was an independent risk factor demonstrating a larger contribution than either ECG category or CK-MB mass. Further, this study used logistic regression analysis to demonstrate that after adjustment for ECG category and baseline CK-MB mass concentration, cTnT made a highly significant contribution for prediction of 30-day mortality ($P = 0.027$). However, after adjustment for cTnT and ECG category, CK-MB did not add significantly to the ability of the model for predicting 30-day mortality ($P = 0.717$). Hence CK-MB mass is not useful for risk stratification when cTnT measurements are available,[111] as was found for cTnI in the TIMI IIIb study.[70]

The contribution of later cTnT measurements, at 8 and 16 h after presentation, for improving the ability of predicting 30-day and 1-year mortality in the GUSTO IIa population was reported.[112] As shown in Table 5, the 8-h cTnT measurement added significantly to a logistic regression model for predicting 1-year mortality ($P = 0.0015$). The 16-h cTnT measurement significantly improved the predictive ability of the model for both 30-day ($P = 0.0382$) and 1-year mortality ($P = 0.0004$). Only age and ST-segment elevation were stronger predictors of 30-day or 1-year mortality than cTnT measurement. The data in Table 4 also indicate that it is not necessary to collect both 8- and 16-h samples, as they do not add significant information to the predictive ability of cTnT. This study showed that serial sampling at the baseline and then at 8 h or preferably 16 h is useful for evaluating the risk of serious cardiac events.[112]

The GUSTO IIa cohort was also used to compare the performance of baseline measurements of cTnT and cTnI for predicting risk of 30-day mortality.[113] Both the cTnT and cTnI measurements were quantitative and were considered continuous variables in a combined logistic regression analysis in which 30-day mortality was the outcome. Table 6 shows that when each variable was considered alone, cTnT showed the highest significance for predicting 30-day mortality in this patient group. Further, combined logistic regression models were constructed to assess the relative contribution of cTnT, cTnI, and ECG for predicting 30-day cardiac death. When cTnI was added to a model developed with the ECG category and cTnT, the model's predictive ability was not improved significantly (Table 6). However, when cTnT was added to a model containing ECG category and cTnI, there was a significant improvement in the model's ability to predict 30-day mortality ($P = 0.045$). When ECG category was added to a model combining cTnT and cTnI, it added the most information ($P = 0.019$). It is of note that preliminary analysis indicates that after adding the 8- or 16-h samples there was no difference between the ability of cTnT and cTnI to predict 30-day mortality (data not shown).

Table 6 Relative value of cTnT, cTnI and the ECG in the prediction of 30-day mortality.[113] (this analysis included data from 770 ACS patients enrolled in the GUSTO IIa study)

Model	Variable	Chi-squared value	*P* Value
Univariate	cTnT	21.0	<0.001
Univariate	ECG	14.2	0.003
Univariate	cTnI	12.3	0.002
Model	**Added variable**	**Additional chi-squared value**	***P* Value**
cTnI + ECG	cTnT	8.03	0.045
cTnT + cTnI	ECG	9.96	0.019
cTnT + ECG	cTnI	0.84	0.675

A meta-analysis comparing risk stratification using cTnT and cTnI was conducted for studies having specimens collected within 12–24 h of presentation or chest pain, focusing on the unstable angina patients.[114] As indicated in Table 7, this meta-analysis revealed no significant difference between cTnT and cTnI for predicting nonfatal MI and death at a median of 30 days for cTnT and a median of 42 days for cTnI.[114]

Risk of nonfatal MI and mortality in the 30 days following the index event was also assessed for cTnT and

Table 7 Cumulative meta-analysis of cTnT and cTnI studies that included samples > 12 h after onset of symptoms or presentation

Marker	Total no. of patients	Odds ratio	95% CI	Chi-squared
cTnT	2847	2.7	2.1–3.4	66
cTnI	1901	4.2	2.7–6.4	42

Although the chi-squared value indicates highly significant associations between either cTnI or cTnT with short-term mortality, there was no significant difference between the odds ratios for cTnT and cTnI.[114]

cTnI in 773 acute chest pain patients, with measurements performed using POC devices from Roche Diagnostics and Spectral Diagnostics, respectively.[115] Patients with ST-segment elevation were excluded from the study and sampling was performed at presentation and at ≥6 h after onset of symptoms. This study showed that the cTnT device was positive in 94% of the MI patients and the cTnI device was positive in 100% of patients. The risk of nonfatal MI or cardiac death in patients with negative test results was extremely low, as only 1.1% of the negative cTnT patients and 0.3% of the cTnI patients had these outcomes within the 30-day follow-up period.[115] The study reiterated the need for serial sampling to assess risk and concluded that TnT or TnI results cannot replace clinical evaluation but that troponin results should be available for Emergency Medicine caregivers and chest pain units.[115]

9.5.1 Guidance of Therapy

There is strong evidence that TnT and TnI are indicators of increased risk in ACS patients and efforts have begun to examine if therapeutic intervention can reduce this risk. These activities are confluent with efforts in the therapeutics industry because drugs for ACS patients are frequently expensive, and definitive benefit is difficult to demonstrate. Although administration to all patients may be untenable based on costs, strategies using biochemical markers to identify patients at increased risk and who would benefit from therapies is being investigated. To gain insight into the future role of cardiac markers, and also other in vitro diagnostics, it is important to monitor the direction and evidence base of therapeutics and therapeutic trials. Two examples of the potential role for cardiac markers in the guidance of intervention are presented below.

The FRISC study examined whether TnT concentrations could be useful for identifying patients with unstable coronary artery disease who might benefit from therapeutic intervention.[112] This issue was investigated by measuring cTnT levels in serum from 971 patients who received either placebo or low molecular weight heparin in short-term (6-day) or long-term (5-week) regimens. Among patients having cTnT levels below a cutoff of $<0.1\ \mu g\ L^{-1}$, short-term treatment showed no significant effect in decreasing the incidence of death and/or MI. However, in patients having levels of cTnT $\geq 0.1\ \mu g\ L^{-1}$, differences in the incidence of death and/or MI were significant as the event rate was 6.0% in the placebo group versus 2.5% in short-term treatment patients ($P < 0.05$). With long-term treatment with low molecular weight heparin the results were more dramatic. In patients having cTnT $\geq 0.1\ \mu g\ L^{-1}$, death and/or MI occurred at a rate in the placebo group nearly double that in the group receiving low molecular weight heparin (14.2 vs 7.4%; $P < 0.01$). On the other hand, cTnT levels $< 0.1\ \mu g\ L^{-1}$ identified a low-risk group in whom death and MI showed no difference between the treated and placebo groups. This study indicated that elevated cTnT concentrations could potentially identify patients who would benefit from long-term treatment with low molecular weight heparin.

An important focus in cardiovascular therapeutics has been the development of agents that are antagonists for the platelet glycoprotein (GP) IIb/IIIa receptor. These agents are important because conformational changes in the GP IIb/IIIa receptor is the final step in platelet activation. The idea behind blockage of this receptor is to prevent platelets from becoming ''angry'' and contributing to the potential of thrombosis. The CAPTURE study included unstable angina patients who were randomized to receive either a GP IIb/IIIa receptor inhibitor or a placebo.[113] Prior to randomization, cTnT was measured in the patients and the subjects were classified as either positive or negative. In the 24–36 h after receiving either GP IIb/IIIa inhibitor or placebo, cardiac catheterization was performed with angioplasty if necessary. The patients for whom the cTnT was negative showed a low incidence of death or MI while awaiting cardiac catheterization. Likewise, the cTnT positive patients who received the GP IIb/IIIa inhibitor had a similarly low rate of death or MI. However, the patients receiving placebo who were cTnT positive had a significantly higher event rate of 4.1% compared with the cTnT group who received the platelet inhibitor ($P = 0.03$). Thus, in the CAPTURE trial cTnT measurements identified a group that benefited from administration of the GP IIb/IIIa inhibitor. cTnT measurement may help focus this treatment in ACS patients.

9.6 Infarct Sizing

Infarct size correlates closely with mortality[111] and also other prognostic indexes such as cardiac failure,[116] ventricular function, and arrhythmia.[117] Infarct size can be determined by electrocardiography, echocardiography, left ventriculography, and radionuclide methods or with cardiac markers. Although the ECG is convenient for estimating infarct size clinically, the estimation of infarct size

from the ECG at the very early stage of MI is unreliable because ST-segment elevation and Q-wave amplitudes change spontaneously within 24 h of the onset of AMI and change rapidly after reperfusion.[118]

Infarct size can be estimated by measuring myocardial proteins released from the injured myocardium. CK and CK-MB have been used for infarct sizing,[88] but they are not suitable for early estimation of infarct size because a period of 1–6 days is needed. On the other hand, myoglobin appears in the blood after infarction much earlier than CK owing to its low molecular weight. Serial measurements of myoglobin can be used for early estimation of infarct size. However, size from total myoglobin may be overestimated in patients with severe renal dysfunction, because myoglobin is rapidly excreted in urine and is not specific to the myocardium. The infarct size estimated from total myoglobin may be overestimated in patients receiving intramuscular injection, or those with shock, trauma, alcoholism, or muscle disorder. CK-MB, which has a better cardio-specificity than myoglobin, may be more suitable for estimating infarct size in patients with skeletal muscle damage. Infarct sizing can be estimated by plotting the release of a cardiac marker against time after the onset of chest pain and calculating the area under the curve. Calculation of infarct size using cardiac markers has correlated fairly well with estimates of necrotic zones made on cardiac tissue at autopsy.[119] Myocardial extension following an initial infarct reduces regional blood flow and delays the appearance of cardiac markers in blood, therefore affecting the use of cardiac markers for infarct sizing.

10 ECONOMIC ISSUES IN CARDIAC MARKERS

10.1 Is Rapid Testing Necessary?

The need for rapid testing must be evaluated within the context of the role of biochemical markers in the management of patients with suspected ACS. Although ST-segment elevation is suggested for identifying patients

Table 8 Recommendations for markers in the triage of chest pain patients

Recommendation	Strength/ Consensus
1. Members of the EDs, divisions of cardiology, hospital administrations, and clinical laboratories should work collectively to develop an accelerated protocol for the use of biochemical markers in the evaluation of patients with possible ACS.	Class I
For simplicity, this protocol should apply to either the facilitated diagnosis or the rule-out of AMI in the ED or to routine diagnosis from other areas of the hospital, should a patient develop symptoms consistent with ACSs while hospitalized.	Class II
2. For routine clinical practice, blood collections should be referenced relative to the time of presentation to the ED and (when available) the reported time of chest pain onset.	Class I
3. Two biochemical markers should be used for routine AMI diagnosis: an early marker (reliably increased in blood within 6 h after onset of symptoms) and a definitive marker (increased in blood after 6–9 h, but has high sensitivity and specificity for myocardial injury, remaining abnormal for several days after onset).	Class II
4. In patients with a diagnostic ECG, on presentation (ST-segment elevations, presence of Q-waves or left bundle branch block in two or more contiguous leads), the diagnosis of AMI can be made and acute treatment initiated without results of acute cardiac marker testing.	Class I
In AMI patients with diagnostic ECGs, biochemical marker testing at a reduced frequency of blood collection (e.g. twice per day) is valuable for confirmation of diagnosis, to qualitatively estimate the size of the infarction, and to detect the presence of complications such as reinfarction.	Class I
5. For detection of AMI by enzyme or protein markers, in the absence of definitive ECGs, the following sampling frequency is recommended:	Class II

Marker	Admission	2–4 h	6–9 h	12–24 h
Early (<6 h)	×	×	×	(×)
Late (>6 h)	×	×	×	(×)

(×) indicates optional determinations.

6. For those EDs in which patients' triage decisions are not made within the first few hours after ED presentation, the use of an early marker such as myoglobin may be unnecessary. In this case, only one definitive marker such as cardiac troponin is needed. The frequency of blood collection should also be reduced.	Class I

with AMI with a relatively high degree of specificity, the rate of misdiagnosis of AMI was found to be 20–24% even when strict clinical criteria and ECG are used to select for thrombolysis.[120,121] Therefore, there is a role for cardiac markers in patients who have a conclusive ECG diagnosis of AMI, whether they receive thrombolytic therapy or not, for audit and clinical quality monitoring. Even in patients presenting with ST-segment elevation on ECG, measurement of cardiac troponins (cTnT or cTnI) allows a further risk stratification.[67,70,110]

Recent technological advances in assay development have allowed short TATs and rapid results. The rationale behind rapid testing is the provision of a rapid accurate biochemical test result. Therefore, the evidence of rapid testing requires a number of questions to be answered. This area is best described by Collinson[122] with the following four questions.

10.1.1 Is Rapid Biochemical Diagnosis Possible?

Rapid rule-in and rule-out of AMI using biochemical markers can be achieved by serial measurement of CK,[123,124] and CK-MB alone or with myoglobin.[125] In addition to their role in diagnosing AMI, cardiac troponins have additional value in prognostic risk stratification. The presence of detectable cTnT on admission in patients who receive thrombolysis is associated with an adverse prognosis[70,110] and failure to achieve TIMI III flow on postthrombolysis angioplasty.[126] Such groups may benefit from primary angioplasty or aggressive thrombolytic treatment. The presence of cardiac troponin identifies a subgroup of unstable angina patients who will respond to low molecular weight heparin and GP IIb/IIIa antagonists. Several ongoing clinical trials are investigating the use of biochemical markers to guide therapy.

10.1.2 Is Rapid Biochemical Diagnosis Clinically Relevant?

Rapid diagnosis has a direct impact on patient management. Rapid diagnostic algorithm using 8-h CK measurement combined with calculation of rate of change was shown to reduce median Critical Care Unit stay by 1 day.[127] Rapid diagnosis can be used for early rule-out of AMI in the ED, allowing safe and early discharge, minimizing inappropriate admissions and misdiagnosis rate,[128] and reducing the length of hospital stay. Overall, rapid testing can result in considerable clinical benefit.

10.1.3 How Can Rapid Biochemical Testing be Delivered Accurately and Effectively?

There is a need for analytically accurate technology that can provide a rapid TAT. The laboratory can shorten TAT by the use of pneumatic tube sample delivery systems, fast analyzers, and electronic data systems. The limitation is therefore the sample processing time. Serum can be ready within 30 min at best, as the sample is allowed to clot before processing. Faster TAT can be achieved with plasma or whole blood. The alternative is the use of POC testing either by replicating the laboratory analyzers at the bedside or by using innovative technologies such as immuno-chromatographic devices, gold-labeled optically read immunoassay (GLORIA) and the use of whole-blood single or multiple marker assays. POC testing must be simple, robust, and analytically accurate. No prospective studies have been published where POC testing was used exclusively for management.

Table 9 Recommendations for markers in ACS

Recommendation	Strength/ Consensus
1. Two decision limits are needed for the optimum use of sensitive and specific cardiac markers such as cTnT or cTnI. A low abnormal value establishes the first presence of true myocardial injury, and a higher value is suggestive of injury to the extent that it qualifies as AMI, as defined previously by WHO.	Class II
2. Chest pain patients with laboratory results for cTnT and cTnI between the upper limit of the reference interval and the decision limit for AMI should be labeled as having ''myocardial injury''. These patients should be admitted and acutely treated to reduce the risks associated with this injury.	Class I
3. The WHO definition of AMI should be expanded to include the use of serial biochemical markers and not be limited to enzyme changes. It should be emphasized that rule-out of AMI cannot be made on the basis of data from a single blood collection. However, when very specific cardiac markers are used, the presence of an abnormal concentration from a single specimen can be highly diagnostic of myocardial injury.	Class I
4. At this time, there are no data available to recommend the use of cardiac markers such as cTnT or cTnI for screening asymptomatic patients for the presence of ACSs. The likelihood of detecting silent ischemia is extremely low and cannot justify the costs of screening programs. Additionally, there is no evidence that cardiac marker analysis of blood following stress testing can indicate the presence of coronary artery disease.	Class III (for use of cardiac markers for screening)

10.1.4 Are the Cost Economics of Rapid Testing Justified?

Cost economic analysis of rapid rule-out protocols shows significant cost reductions compared with conventional approaches. Rapid diagnosis is effective and cost effective only within the context of data-driven decision-making protocols. POC testing for cardiac markers has a definitive role in management when a very short TAT is required such as in the ED. If the laboratory TAT exceeds 25% of the decision time, then rapid testing will be required. For therapeutic decisions to be made, only rapid testing will be able to satisfy the TAT requirement.

10.2 Triage Strategies

In the ED, little impartial evidence is available to differentiate patients with coronary artery diseases from those presenting with other causes of chest pain. Therefore, the assessment of symptomatic patients presenting to the ED is a challenging task. To determine the probability that a particular patient having an acute ischemic coronary syndrome, ED clinicians rely on the patient's description of the chest discomfort, the presence or absence of nausea, vomiting, or diaphoresis, and the presence of risk factors for coronary artery disease.

Gibler et al. introduced a model that routinely incorporates serial testing for CK-MB on presentation to the ED and 3, 6, and 9 h later with continuous 12-lead ECGs/serial ST-segment trend monitoring for 9 h.[129] Such a model was suggested to be effective for evaluating low- to moderate-risk patients with possible acute ischemic coronary syndrome in the ED. The cost effectiveness of this approach compared with standards in hospital evaluation warrants further assessment, as others have argued that widespread adoption of the chest pain ED concept may not be cost-effective.[10]

Table 10 Recommendations for markers clinical applications other than AMI and research

Recommendation	Strength/Consensus
1. For assessment of reperfusion status following thrombolytic therapy, at least two blood samples are collected and marker concentration compared: time = 0, defined as just before initiation of therapy, and time = 1, defined as 90 min after the start. From these values, the determination of the (a) slope value [($\text{marker}_{t=90} - \text{marker}_{t=0}$)/90 min]; (b) absolute value of $\text{marker}_{t=90}$, in minutes; or (c) the ratio of $\text{marker}_{t=90}/\text{marker}_{t=0}$ can be used as the discriminating factor between successful and unsuccessful reperfusion. However, monitoring with biochemical marker strategies has not been successful in distinguishing between TIMI grade 3 and TIMI grade 2 flow patients, rendering the utility of these measurements clinically problematic for determining complete reperfusion.	Class II
2. cTnT or cTnI should be used for the detection of perioperative AMI in patients undergoing noncardiac surgical procedures. The same AMI decision limit should be used.	Class I
3. Cardiac markers should not be routinely used for infarct sizing because the existing markers are inaccurate in the presence of spontaneous, pharmacological, or surgical reperfusion.	Class III (for use of markers in infarct sizing)
4. Early in the process, manufacturers should seek assistance and provide support to professional organizations, such as the AACC or IFCC to develop committees for the standardization of new analytes. These organizations will determine the need for analyte standardization based on the potential clinical importance of the marker and gather the necessary scientific expertise for the formation of a standardization committee.	Class I
5. Reference intervals are established for each marker on a population of healthy individuals, using the 97.5 percentile (one-tail) of results. Separate cutoff concentrations for results indicative of AMI are also necessary for all cardiac markers. Standardized ROC curves should be used to establish AMI decision limits, using carefully selected and diagnosed populations.	Class I
6. For research studies involving the kinetics of release and appearance of new biochemical markers, the time course of release and appearance in blood must be defined relative to the onset of clinical symptoms.	Class I
The diagnostic accuracy of these new markers may be compromised if the diagnosis of AMI for study patients is based on standard enzyme markers that themselves have sensitivity and/or specificity limitations (e.g. total CK and CK-MB). Therefore AMI diagnosis should be defined by WHO criteria, but with the substitution of ''unequivocal serial changes of cTnT or cTnI'' as the principal biochemical marker, in place of the current WHO criteria of ''unequivocal serial enzyme changes''.	Class II

AACC, American Association for Clinical Chemistry; IFCC, International Federation of Clinical Chemistry.

10.3 Cost of Patient Hospitalization

Despite major progress in managing coronary artery disease, approximately half a million individuals die each year from its complications in the USA alone. This reality has urged efforts to diagnose AMI quickly in the ED. Tests used in the ED to distinguish between chest pain of ischemic origin and that of benign causes are often inconclusive, thus leaving a large group of patients with low but important risk of poor outcome. A risk-avoidance strategy has been adopted by many clinicians leading to conservative admission of over 3 million chest pain patients in the USA alone, costing 3–4 billion dollars yearly for those who were found to be disease free. However, AMI remains one of the leading causes of malpractice suits against physicians. In addition, the traditional models of heath care delivery ''inpatients'' versus ''outpatients'' may not match patient needs. This has inspired the development of alternative health care delivery models such as chest pain units[130] and accelerated diagnostic protocols (ADPs).

One study evaluated whether the use of an ED-based ADP can reduce hospital admission rate, total cost, and length of stay for patients needing admission for evaluation of chest pain.[131] This randomized trial included 165 patients and compared admission rate, total cost, and length of stay for patients treated using ADP versus inpatient controls. Control patients were admitted to the telemetry unit on the internal medicine service for standard management including two sets of cardiac marker studies, two ECGs, and 24 h of cardiac and clinical monitoring. Until a diagnostic end-point was met (positive CK-MB result, recurrent chest pain, or ECG findings consistent with myocardial ischemia) or the protocol was completed, the ADP patients received (1) 12 h of rhythm monitoring, (2) CK-MB level measurements at 0, 4, 8, and 12 h, (3) ECG at 0, 6, and 12 h, (4) clinical examination and review of test results at 0, 6, and 12 h or for any change in condition, (5) aspirin, (6) 2 L of oxygen by nasal cannula, and (7) intravenous line. During the first 12 h, patients with any positive test results were hospitalized, and patients with suggestive findings for myocardial ischemia were also hospitalized. If all clinical findings were negative, patients were subjected to ECG exercise stress test and classified as ADP discharges (negative stress test) or ADP admissions (positive ECG stress test). This study concluded that the hospital admission rate for ADP versus control patients was ~45% versus 100% ($P < 0.001$). The mean total cost per patient for ADP versus control patients was $1528 versus $2095 ($P < 0.001$). The mean length of stay was 33.1 h for ADP patients versus 44.8 h for controls ($P < 0.1$). In this trial, ADP strategy saved $567 in total hospital costs per patient treated. Therefore, use of ED-based ADP was shown to reduce hospitalization rates, length of stay, and total cost for evaluating low-risk patients with chest pain for possible AMI or acute cardiac ischemia. Locating space for ADPs in or near the ED is advantageous because it will reduce inpatient bed requirements while allowing optimum use of the high fixed costs of ED maintenance. Only hospitals already at full capacity may consider initiating a separate unit with new staff, equipment, and

Table 11 Recommendations for assay platforms and markers of acute MI

Recommendation	Strength/ Consensus
1. Cardiac troponin (T or I) is the new standard for diagnosis of MI and detection of myocardial cell damage, replacing CK-MB.	Class II
There is no longer a role for LD and its isoenzymes in the diagnosis of cardiac diseases.	Class I
2. The laboratory should perform STAT cardiac marker testing on a continuous random-access basis, with a target TAT of 1 h or less. The TAT is defined as the time from blood collection to the reporting of results.	Class II
3. Institutions that cannot consistently deliver cardiac marker TATs of ~1 h should implement POC testing devices. The cutoff concentration of these devices should be set at 97.5% upper reference limits so that the devices can detect the first presence of true myocardial injury.	Class I
4. Among other tasks, laboratory personnel must be involved in the selection of devices, the training of individuals to perform the analysis, the maintenance of POC equipment, the verification of the proficiency of operators on a regular basis, and the compliance of documentation with requirements by regulatory agencies such as the Health Care Finance administration and the Clinical Laboratory Improvement Act of 1988. In meeting these requirements, quality-assurance and quality-control programs must be instituted and fully documented on a regular basis.	Class I
5. Assays for cardiac markers should have an imprecision CV <10% at the AMI decision limits and an assay TAT of <30 min. Before launch, assays must be characterized with respect to potentially interfering substances, e.g. other related proteins, human anti-mouse antibodies, and other interference.	Class II
6. Plasma or anticoagulated whole blood is the specimen of choice for the STAT analysis of cardiac markers.	Class I

CV, coefficient of variation.

construction. There is a consensus that ED-based ADPs can maintain quality care in a cost-cutting environment. Maximum benefits of ADPs may be achieved via further research to determine optimum settings, patient selection, and protocol design.

11 RECOMMENDATIONS FOR THE USE OF CARDIAC MARKERS IN CORONARY ARTERY DISEASE

The National Academy of Clinical Biochemistry Standards of Laboratory Practice assembled a committee to write recommendations on the use of cardiac markers in coronary artery diseases.[132] The recommendations were divided into four areas: the use of markers in the triage of patients with chest pain, ACS, clinical applications other than AMI and research, and assay platforms and markers of AMI.[132]

11.1 Definition of Classification Scheme

The classes are defined as follows:

- Class I: a recommendation for which there is evidence and/or general agreement.
- Class II: a recommendation for which there is conflicting evidence and/or a divergence of opinion about its usefulness/efficacy, but where the weight of evidence/opinion is in its favor.
- Class III: a recommendation for which there is evidence and/or general agreement that a procedure is not useful or effective.

Details of the recommendations are given in Tables 8–11.

ABBREVIATIONS AND ACRONYMS

ACS	Acute Coronary Syndrome
ADP	Accelerated Diagnostic Protocol
AK	Adenylate Kinase
AMI	Acute Myocardial Infarction
ATPase	Adenosine Triphosphatase
CA	Carbonic Anhydrase
CI	Confidence Interval
CK	Creatine Kinase
CPEC	Chest Pain Evaluation Center
CRP	C-reactive Protein
cTnI	Cardiac Troponin I
cTnT	Cardiac Troponin T
CV	Coefficient of Variation
ECG	Electrocardiogram
ED	Emergency Department
ELISA	Enzyme-linked Immunosorbent Assay
FABP	Fatty Acid-binding Protein
FN	False Negative
FP	False Positive
GLORIA	Gold-labeled Optically Read Immunoassay
GP	Glycoprotein
GPBB	Glycogen Phosphorylase-BB
GUSTO	Global Use of Strategies to Open Occluded Coronary Arteries
HHFABP	Human Heart Fatty Acid-binding Protein
LD	Lactate Dehydrogenase
MI	Myocardial Infarction
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
NHAAP	National Heart Attack Alert Program
NIH	National Institutes of Health
NPV	Negative Predictive Value
POC	Point-of-care
PPV	Positive Predictive Value
ROC	Receiver Operator Characteristic
SAA	Serum Amyloid A
TAT	Turnaround Time
TIMI	Thrombolysis in Myocardial Infarction
TN	True Negative
TnC	Troponin C
TnI	Troponin I
TnT	Troponin T
TP	True Positive
UV	Ultraviolet
WHO	World Health Organization

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Immunochemistry • Point-of-care Testing • Product Development for the Clinical Laboratory • Serum Proteins • Statistical Quality Control in Clinical Laboratories

Peptides and Proteins **(Volume 7)**
Gel Electrophoresis in Protein and Peptide Analysis

REFERENCES

1. P.T. Larson, N.H. Wallen, P. Hjemdahl, 'Norepinephrine-induced Platelet Activation In Vivo is Only Partly

Counteracted by Aspirin', *Circulation*, **89**, 1951–1957 (1994).
2. J.J. Badimon, L. Badimon, V.T. Turitto, V. Fuster, 'Platelet Deposition at Higher Shear Rates is Enhanced by High Plasma Cholesterol Levels: In Vivo Study in a Rabbit Model', *Arterioscl. Thromb.*, **11**, 395–402 (1991).
3. R.M. Jacoby, R.W. Nesto, 'Acute Myocardial Infarction in the Diabetic Patient: Pathophysiology, Clinical Course and Prognosis', *J. Am. Coll. Cardiol.*, **20**, 736–744 (1992).
4. B.F. Uretsky, D.S. Farquhar, A.F. Berezin, W.B. Hood, Jr, 'Symptomatic Myocardial Infarction Without Chest Pain: Prevalence and Clinical Course', *Am. J. Cardiol.*, **40**, 498–503 (1977).
5. T.J. Ryan, J.L. Anderson, E.M. Antman, B.A. Braniff, N.H. Brooks, R.M. Califf, L.D. Hillis, L.F. Hiratzka, E. Rapaport, B.J. Riegel, R.O. Russell, E.E. Smith, III, W.D. Weaver, 'ACC/AHA Guidelines for the Management of Patients with Acute Myocardial Infarction: Executive Summary', *Circulation*, **94**, 2341–2350 (1996).
6. R.E. Rude, W.K. Poole, J.E. Muller, Z. Turi, J. Rutherford, C. Parker, R. Roberts, D.S. Raabe, Jr, H.K. Gold, P.H. Stone, 'Electrocardiographic and Clinical Criteria for the Recognition of Acute Myocardial Infarction Based on Analysis of 3,697 Patients', *Am. J. Cardiol.*, **52**, 936–942 (1983).
7. T.H. Lee, G.W. Rouan, M.C. Weisberg, D.A. Brand, E.F. Cook, D. Acampora, L. Goldman, 'Sensitivity of Routine Clinical Criteria for Diagnosing Myocardial Infarction within 24 hours of Hospitalization', *Ann. Intern. Med.*, **106**, 181–186 (1987).
8. B.D. McCarthy, J.R. Behansky, R.B. D'Agostino, H.P. Selker, 'Missed Diagnoses of Acute Myocardial Infarction in the Emergency Department: Results from a Multicenter Study', *Ann. Emerg. Med.*, **22**, 579–582 (1993).
9. R.A. Rusnak, T.O. Stair, K. Hansen, J.S. Fastow, 'Litigation Against the Emergency Physician: Common Features in Cases of Missed Myocardial Infarction', *Ann. Emerg. Med.*, **18**, 1029–1034 (1989).
10. R. Shesser, M. Smith, 'The Chest Pain Emergency Department and the Outpatient Chest Pain Evaluation Center: Revolution or Evolution?', *Ann. Emerg. Med.*, **23**, 334–341 (1994).
11. L.G. Graff, J. Dallara, M.A. Ross, A.J. Joseph, J. Itzcovitz, R.P. Andelman, C. Emerman, S. Turbiner, J.A. Espinosa, H. Severance, 'Impact of the Care of the Emergency Department Chest Pain Patient from the Chest Pain Registry (CHEPER) Study', *Am. J. Cardiol.*, **80**, 563–568 (1997).
12. J.C. Paterson, G.E. Cottral, 'Experimental Coronary Sclerosis: Lymphomatosis as a Cause of Coronary Sclerosis in Chickens', *Arch. Pathol.*, **49**, 699–709 (1950).
13. Y.K. Wong, P.J. Gallagher, M.E. Ward, '*Chlamydia pneumoniae* and Atherosclerosis', *Heart*, **81**, 232–238 (1999).
14. J.B. Muhlestein, E.H. Hammond, J.F. Carlquist, E. Radicke, M.J. Thomson, L.A. Karagounis, M.L. Woods, J.L. Anderson, 'Increased Incidence of Chlamydia Species Within the Coronary Arteries of Patients with Symptomatic Atherosclerosis Versus Other Forms of Cardiovascular Disease', *J. Am. Coll. Cardiol.*, **27**(7), 1555–1561 (1996).
15. K. Bachmaier, N. Neu, L.M. de la Maza, S. Pal, A. Hessel, J.M. Penninger, 'Chlamydia Infections and Heart Disease Linked Through Antigenic Mimicry', *Science*, **283**, 1335–1339 (1999).
16. P. Libby, D. Egan, S. Skarlatos, 'Roles of Infectious Agents in Atherosclerosis and Restenosis: an Assessment of the Evidence and Need for Future Research', *Circulation*, **96**(11), 4095–4103 (1997).
17. A.S. Jaffe, 'More Rapid Biochemical Diagnosis of Myocardial Infarction: Necessary? Prudent? Cost Effective?' [Editorial], *Clin. Chem.*, **39**, 1567–1569 (1993).
18. L.K. Newby, W.B. Gibler, E.M. Ohman, R.H. Christenson, 'Biochemical Markers in Suspected Acute Myocardial Infarction; the Need for Early Assessment' [Editorial], *Clin. Chem.*, **41**, 1263–1265 (1995).
19. J.C. Boyd, 'Mathematical Models for Demonstrating Clinical Usefulness of Biochemical Markers', *Scand. J. Clin. Lab. Invest.*, **57**(Suppl. 227), 46–63 (1997).
20. J.R. Beck, E.K. Shultz, 'The Use of Relative Operating Characteristic (ROC) Curves in Test Performance Evaluation', *Arch. Pathol. Lab. Med.*, **110**, 13–20 (1986).
21. H.P. Selker, R.J. Zelenski, E.M. Antman, T.P. Aufderheide, S.A. Bernard, R.O. Bonow, W.B. Gibler, M.D. Hagen, P. Johnson, J. Lau, R.A. McNutt, J. Ornato, J.S. Schwartz, J.D. Scott, P.A. Tunick, W.D. Weaver, 'An Evaluation of Technologies for Identifying Acute Cardiac Ischemia in the Emergency Department: a Report from a National Heart Attack Alert Program Working Group', *Ann. Emerg. Med.*, **29**, 13–87 (1997).
22. A.H.B. Wu, P.L. Lane, 'Metaanalysis in Clinical Chemistry: Validation of Cardiac Troponin T as a Marker for Ischemic Heart Diseases', *Clin. Chem.*, **41**, 1228–1233 (1995).
23. Y. Landt, H.C. Vaidya, S.E. Porter, D.N. Dietzler, J.H. Ladenson, 'Immunoaffinity Purification of Creatine Kinase-MB from Human, Dog, and Rabbit Heart with Use of a Monoclonal Antibody Specific for CK-MB', *Clin. Chem.*, **35**(6), 985–989 (1989).
24. R.H. Christenson, H. Vaidya, Y. Landt, R.S. Bauer, S.F. Green, F. Apple, A. Jacob, G.R. Magneson, S. Nag, A.H.B. Wu, H.M.E. Azzazy, 'Standardization of Creatine Kinase-MB (CK-MB) Mass Assays: the Use of Recombinant CK-MB as a Reference Material', *Clin. Chem.*, **45**, 1414–1423 (1999).
25. M. Panteghini, 'Serum Isoforms of Creatine Kinase Isoenzymes', *Clin. Biochem.*, **21**, 211–218 (1988).
26. A.H.B. Wu, 'Creatine Kinase Isoforms in Ischemic Heart Disease', *Clin. Chem.*, **35**, 7–13 (1989).

27. P.R. Puleo, P.A. Guadagno, R. Roberts, M.V. Schell, A.J. Marian, D. Churchill, M.B. Perryman, 'Early Diagnosis of Acute Myocardial Infarction Based on Assay for Subforms of Creatine Kinase-MB', *Circulation*, **82**, 1073–1075 (1990).
28. A.H.B. Wu, X.M. Wang, T.G. Gornet, J. Ordonez-Llanos, 'Creatine Kinase MB Isoforms in Patients with Skeletal Muscle Injury: Ramifications for Early Detection of Acute Myocardial Infarction', *Clin. Chem.*, **38**, 2396–2400 (1992).
29. J. Zimmerman, R. Fromm, D. Meyer, A. Boudreaux, C.C. Wun, R. Smalling, B. Davis, G. Habib, R. Roberts, 'Diagnostic Marker Cooperative Study for the Diagnosis of Myocardial Infarction', *Circulation*, **99**, 1671–1677 (1999).
30. J.A. Lott, E. Nemesanszky, 'Lactate Dehydrogenase (LD)', in *Clinical Enzymology. A Case-oriented Approach*, eds. J.A. Lott, P.L. Wolf, Field and Rich/Year Book, New York, 213–244, 1986.
31. F. Vermeer, A. Van der Laarse, 'Cumulative Enzyme Release as a Measure of Infarct Size in Patients with Acute Myocardial Infarction Receiving Thrombolytic Therapy', *Arch. Mal. Coeur. Vaiss.*, **4**, 25–28 (1993).
32. J.M. Wilkinson, R.J.A. Grand, 'Comparison of Amino Acid Sequences of Troponin I from Different Striated Muscles', *Nature (London)*, **271**, 31–35 (1978).
33. G.S. Bodor, S. Porter, Y. Landt, J.H. Ladenson, 'Development of Monoclonal Antibodies for an Assay of Cardiac Troponin I and Preliminary Results in Suspected Cases of Myocardial Infarction', *Clin. Chem.*, **38**, 2203–2214 (1992).
34. H. Baum, S. Braun, W. Gerhardt, G. Gilson, G. Hafner, M. Muller-Bardorff, W. Stein, G. Klein, C. Ebert, K. Hallermayer, H.A. Katus, 'Multicenter Evaluation of a Second-generation Assay for Cardiac Troponin T', *Clin. Chem.*, **43**, 1877–1884 (1997).
35. A.S. Zot, J.D. Potter, 'Structural Aspects of Troponin–Tropomyosin Regulation of Skeletal Muscle Contraction', *Annu. Rev. Biophys. Chem.*, **16**, 535–539 (1987).
36. S.V. Perry, 'The Regulation of Contractile Activity in Muscle', *Biochem. Soc. Trans.*, **7**, 593–617 (1979).
37. J.E. Adams, III, G.S. Bodor, V.G. Davilla-Roman, J.A. Delmez, F.S. Apple, J.H. Ladenson, A.S. Jaffe, 'Cardiac Troponin I: a Marker with High Specificity for Cardiac Injury', *Circulation*, **88**, 101–106 (1993).
38. J.H. Keffer, 'Myocardial Markers of Injury Evolution and Insights', *Am. J. Clin. Pathol.*, **105**, 305–320 (1996).
39. A.H.B. Wu, S.A. Abbas, S. Green, L. Pearsall, S. Dhakam, R. Azar, M. Onoroski, A. Senaie, R.G. Mckay, D. Waters, 'Prognostic Value of Cardiac Troponin T in Unstable Angina Pectoris', *Am. J. Cardiol.*, **76**, 970–972 (1995).
40. A.G. Katrukha, A.V. Bereznikova, T.V. Esakova, K. Pettersson, T. Lovegren, M.E. Severina, K. Pulkki, L.-M. Vuopio-Pulkki, N.B. Gusev, 'Troponin I is Released in Bloodstream of Patients with Acute Myocardial Infarction Not in Free Form but as Complex', *Clin. Chem.*, **43**(8), 1379–1385 (1997).
41. J.D. Potter, 'Preparation of Troponin and Its Subunits', *Methods Enzymol.*, **85**, 241–263 (1982).
42. S.J. Dodgson, 'The Carbonic Anhydrases: Overview of Their Importance in Cellular Physiology and in Molecular Genetics', in *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*, eds. S.J. Dodgson, R.E. Tashian, G. Gros, N.D. Carter, Plenum Press, New York, 3–32, 1991.
43. H.F. Deutsch, 'Carbonic Anhydrases', *Int. J. Biochem.*, **19**, 101–113 (1987).
44. M. Ohta, Y. Itagaki, N. Itoh, K. Hayashi, H. Nishitani, K. Ohta, 'Carbonic Anhydrase III in Serum in Muscular Dystrophy and Other Neurological Disorders: Relationship with Creatine Kinase', *Clin. Chem.*, **37**, 36–39 (1991).
45. H.K. Väänänen, H. Syrjälä, P. Rahkila, J. Vuori, L.M. Melamies, V. Myllylä, T.E.S. Takala, 'Serum Carbonic Anhydrase III and Myoglobin Concentrations in Acute Myocardial Infarction', *Clin. Chem.*, **36**, 635–638 (1990).
46. H.K. Väänänen, E.S.T. Timo, U. Tolonen, J. Vuori, V.V. Myllylä, 'Muscle-specific Carbonic Anhydrase III is a More Sensitive Marker of Muscle Damage than Creatine Kinase in Neuromuscular Disorders', *Arch. Neurol.*, **45**, 1254–1256 (1988).
47. G.X. Brogan, Jr, J. Vuori, S. Freidman, C.F. McCuskey, H.C. Thode, Jr, H.K. Väänänen, D.S. Cooling, J.L. Bock, 'Improved Specificity of Myoglobin Plus Carbonic Anhydrase Assay Versus that of Creatine Kinase-MB for Early Diagnosis of Acute Myocardial Infarction', *Ann. Emerg. Med.*, **27**, 22–28 (1995).
48. R.K. Ockner, J.A. Manning, R.B. Poppenhausen, W.K. Ho, 'A Binding Protein for Fatty Acids in Cytosol of Intestinal Mucosa, Liver, Myocardium, and Other Tissues', *Science*, **177**, 56–58 (1972).
49. R.K. Ockner, 'Historic Overview of Studies on Fatty Acid-binding Proteins', *Mol. Cell. Biochem.*, **98**, 3–9 (1990).
50. A.H. Kleine, J.F.C. Glatz, F.A. Van Nieuwenhoven, G.J. Van der Vusse, 'Release of Heart Fatty Acid-binding Protein into Plasma After Acute Myocardial Infarction in Man', *Mol. Cell. Biochem.*, **116**, 155–162 (1992).
51. F.A. Van Nieuwenhoven, A.H. Kleine, K.W.H. Wodzig, W.T. Hermens, H.A. Kragten, J.G. Maessen, C.D. Punt, M.P. Van Dieijen, G.J. Van der Vusse, J.F.C. Glatz, 'Discrimination Between Myocardial and Skeletal Muscle Injury by Assessment of the Plasma Ratio of Myoglobin Over Fatty Acid-binding Protein', *Circulation*, **92**, 2848–2854 (1995).
52. J. Gorski, W.T. Hermens, J. Borawski, M. Mysliwiec, J.F.C. Glatz, 'Increased Fatty Acid-binding Protein Concentration in Plasma of Patients with Chronic Renal Failure', *Clin. Chem.*, **43**, 193–195 (1997).

53. M. Gawaz, A. Reininger, F.J. Neumann, 'Platelet Function and Platelet–Leukocyte Adhesion in Symptomatic Coronary Heart Disease. Effects of Intravenous Magnesium', *Thromb. Res.*, **83**, 341–349 (1996).
54. A. Maseri, 'Inflammation, Atherosclerosis, and Ischemic Events – Exploring the Hidden Side of the Moon', *N. Engl. J. Med.*, **336**, 1014–1016 (1997).
55. G. Baroldi, M.D. Silver, F. Mariani, G. Giuliano, 'Correlation of Morphological Variables in the Coronary Atherosclerotic Plaque with Clinical Patterns of Ischemic Heart Disease', *Am. J. Cardiovasc. Pathol.*, **2**, 159–172 (1988).
56. G. Liuzzo, L.M. Biasucci, J.R. Gallimore, R.L. Grillo, A.G. Tebuzzi, M.B. Pepys, A. Maseri, 'The Prognostic Value of C-reactive Protein and Serum Amyloid A Protein in Severe Unstable Angina', *N. Engl. J. Med.*, **331**, 417–424 (1994).
57. P.M. Ridker, M. Cushman, M.J. Stampfer, R.P. Tracy, C.H. Hennekens, 'Inflammation, Aspirin, and the Risk of Cardiovascular Disease in Apparently Healthy Men', *N. Engl. J. Med.*, **336**, 973–979 (1997).
58. M.S. Nieminen, K. Mattila, V. Valtonen, 'Infection and Inflammation as Risk Factors for Myocardial Infarction', *Eur. Heart J.*, **14**(Suppl. K), 12–16 (1993).
59. L. Lee, G.A. Ewald, C.R. McKenzie, P.R. Eisenberg, 'The Relationship of Soluble Fibrin and Cross-linked Fibrin Degradation Products to the Clinical Course of Myocardial Infarction', *Arter. Thromb. Vasc. Biol.*, **17**, 628–633 (1997).
60. S. Hornykewycz, H. Gabriel, K. Huber, 'Biochemical Markers of Myocardial Necrosis in Acute Myocardial Infarction and Thrombolysis', *Ann. Hematol.*, **69**, 559–563 (1994).
61. G. Rabitzsch, J. Mair, P. Lechleitner, F. Noll, U. Hofmann, E. Krause, F. Dienstl, B. Puschendorf, 'Immunoenzymometric Assay of Human Glycogen Phosphorylase Isoenzyme BB in Diagnosis of Ischemic Myocardial Injury', *Clin. Chem.*, **41**, 966–978 (1995).
62. Y.J. Feng, C. Chen, J.T. Fallon, T. Lai, L. Chen, D.R. Knibbs, D.D. Waters, A.H. Wu, 'Comparison of Cardiac Troponin I, Creatine Kinase-MB, and Myoglobin for Detection of Acute Ischemic Myocardial Injury in a Swine Model', *Am. J. Clin. Pathol.*, **110**, 70–77 (1998).
63. H.A. Katus, A. Remppis, F.J. Neumann, T. Scheffold, K.W. Diederich, G. Vinar, A. Noe, G. Matern, W. Kuebler, 'Diagnostic Efficiency of Troponin T Measurements in Acute Myocardial Infarction', *Circulation*, **83**, 902–912 (1991).
64. J.P. Bertinchant, C. Larue, I. Pernel, B. Ledermann, P. Fabbro-Peray, L. Beck, C. Calzolari, S. Trinquier, J. Nigond, B. Pau, 'Release Kinetics of Serum Cardiac Troponin I in Ischemic Myocardial Injury', *Clin. Biochem.*, **29**, 587–594 (1996).
65. J. Ravkilde, M. Horder, W. Gerhardt, L. Ljungdahl, T. Pettersson, N. Tryding, B.H. Moller, A. Hamfelt, T. Graven, A. Asberg, 'Diagnostic Performance and Prognostic Value of Serum Troponin T in Suspected Acute Myocardial Infarction', *Scand. J. Clin. Lab. Invest.*, **53**, 677–685 (1993).
66. R.J. de Winter, R.W. Koster, A. Sturk, G.T. Sanders, 'Value of Myoglobin, Troponin T, and CK-MB Mass in Ruling Out an Acute Myocardial Infarction in the Emergency Room', *Circulation*, **92**, 3401–3407 (1995).
67. P. Stubbs, P. Collinson, D. Moseley, T. Greenwood, M. Noble, 'Prognostic Significance of Admission Troponin T Concentrations in Patients with Myocardial Infarction', *Circulation*, **94**, 1291–1297 (1996).
68. J. Mair, D. Morandell, N. Genser, P. Lechleitner, F. Dienstl, B. Puschendorf, 'Equivalent Early Sensitivities of Myoglobin, Creatine Kinase MB Mass, Creatine Kinase Isoform Ratios, and Cardiac Troponins I and T for Acute Myocardial Infarction', *Clin. Chem.*, **41**, 1266–1272 (1995).
69. Joint International Society and Federation of Cardiology/World Health Organization Task Force on Standardization of Clinical Nomenclature, 'Nomenclature and Criteria for Diagnosis of Ischemic Heart Disease', *Circulation*, **59**, 607–609 (1979).
70. E.M. Antman, M.J. Tanasijevic, B. Thompson, M. Schactman, C.H. McCabe, C.P. Cannon, G.A. Fischer, A.Y. Fung, C. Thompson, D. Wybenga, E. Braunwald, 'Cardiac-specific Troponin I Levels to Predict the Risk of Mortality in Patients with Acute Coronary Syndromes', *N. Engl. J. Med.*, **335**, 1342–1349 (1996).
71. J.F. Tucker, R.A. Collins, A.J. Anderson, J. Hauser, J. Kalas, F. Apple, 'Early Diagnostic Efficiency of Cardiac Troponin I and Troponin T for Acute Myocardial Infarction', *Acad. Emerg. Med.*, **4**, 13–21 (1997).
72. C. Heeschen, C.W. Hamm, B.U. Goldmann, R.H. Moeller, T. Meinertz, 'Cost-effectiveness of a Rapid Test for Troponin in Emergency Admissions (in German)', *Dtsch. Med. Wochenschr.*, **123**, 1229–1234 (1998).
73. G.S. Bodor, L. Survant, E.M. Voss, S. Smith, D. Porterfield, F. Apple, 'Cardiac Troponin T Composition in Normal and Regenerating Human Skeletal Muscle', *Clin. Chem.*, **43**, 476–484 (1997).
74. M.D. McLaurin, F. Apple, E.M. Voss, C.A. Herzog, S.W. Sharkey, 'Cardiac Troponin I, Cardiac Troponin T, and Creatine Kinase MB in Dialysis Patients Without Ischemic Heart Disease: Evidence of Cardiac Troponin T Expression in Skeletal Muscle', *Clin. Chem.*, **43**, 976–982 (1997).
75. J.E. Adams, 3rd, K.B. Schechtman, Y. Landt, J.H. Ladenson, A.S. Jaffe, 'Comparable Detection of Acute Myocardial Infarction by Creatine Kinase MB Isoenzyme and Cardiac Troponin I', *Clin. Chem.*, **40**, 1291–1295 (1994).
76. R.H. Christenson, F.S. Apple, D.L. Morgan, G.L. Alonsozana, K. Mascotti, M. Olson, R.T. McCormack, F.H. Wians, Jr, J.H. Keffer, S.H. Duh, 'Cardiac Troponin I Measurement with the ACCESS Immunoassay System:

Analytical and Clinical Performance Characteristics', *Clin. Chem.*, **44**, 52–60 (1998).
77. F.S. Apple, A.J. Maturen, R.E. Mullins, P.C. Painter, M.S. Pessin-Minsley, R.A. Webster, J. Spray Flores, R. DeCresce, D.J. Fink, P.M. Buckley, J. Marsh, V. Ricchiuti, R.H. Christenson, 'Multicenter Clinical and Analytical Evaluation of the AxSYM Troponin-I Immunoassay to Assist in the Diagnosis of Myocardial Infarction', *Clin. Chem.*, **45**, 206–212 (1999).
78. F.S. Apple, R.H. Christenson, R. Valdes, Jr, A.J. Andriak, A. Berg, S.H. Duh, Y.J. Feng, S.A. Jortani, N.A. Johnson, B. Koplen, K. Mascotti, A.H. Wu, 'Simultaneous Rapid Measurement of Whole Blood Myoglobin, Creatine Kinase MB, and Cardiac Troponin I by the Triage Cardiac Panel for Detection of Myocardial Infarction', *Clin. Chem.*, **45**, 199–205 (1999).
79. P.O. Collinson, W. Gerhardt, H.A. Katus, M. Muller-Bardorff, S. Braun, U. Schricke, W. Vogt, D. Nagel, M. Zander, R. Leinberger, D. Mangold, R. Zerback, 'Multicentre Evaluation of an Immunological Rapid Test for the Detection of Troponin T in Whole Blood Samples', *Eur. J. Clin. Chem. Clin. Biochem.*, **34**, 591–598 (1996).
80. A. Falahati, S.W. Sharkey, D. Christensen, M. McCoy, E.A. Miller, M.A. Murakami, F.S. Apple, 'Implementation of Serum Cardiac Troponin I as Marker for Detection of Acute Myocardial Infarction', *Am. Heart J.*, **137**, 332–337 (1999).
81. E.M. Antman, C. Grudzien, D.B. Sacks, 'Evaluation of a Rapid Bedside Assay for Detection of Serum Cardiac Troponin T', *J. Am. Med. Assoc.*, **273**, 1279–1282 (1995).
82. F. Mach, C. Lovis, J.C. Chevrolet, P. Urban, P.F. Unger, M. Bouillie, J.M. Gaspoz, 'Rapid Bedside Whole Blood Cardiospecific Troponin T Immunoassay for the Diagnosis of Acute Myocardial Infarction', *Am. J. Cardiol.*, **75**, 842–845 (1995).
83. M. Panteghini, C. Cuccia, F. Pagani, C. Turla, 'Comparison of the Diagnostic Performance of Two Rapid Bedside Biochemical Assays in the Early Detection of Acute Myocardial Infarction', *Clin. Cardiol.*, **21**, 394–398 (1998).
84. W. Gerhardt, L. Ljungdahl, P.O. Collinson, C. Lovis, F. Mach, C. Sylven, G. Rasmanis, R. Leinberger, R. Zerback, M. Muller-Bardorff, H.A. Katus, 'An Improved Rapid Troponin T Test with a Decreased Detection Limit: a Multicenter Study of the Analytical and Clinical Performance in Suspected Myocardial Damage', *Scand. J. Clin. Lab. Invest.*, **57**, 549–557 (1997).
85. Rapid Evaluation by Assay of Cardiac Troponin T (REACTT) Investigators Study Group, 'Evaluation of a Bedside Whole-blood Rapid Troponin T Assay in the Emergency Department', *Acad. Emerg. Med.*, **4**, 1018–1024 (1997).
86. M. D'Costa, E. Fleming, M.C. Patterson, 'Cardiac Troponin I for the Diagnosis of Acute Myocardial Infarction in the Emergency Department', *Am. J. Clin. Pathol.*, **108**, 550–555 (1997).
87. M. Zaninotto, S. Altinier, M. Lachin, L. Celegon, M. Plebani, 'Strategies for the Early Diagnosis of Acute Myocardial Infarction Using Biochemical Markers', *Am. J. Clin. Pathol.*, **111**, 399–405 (1999).
88. B.E. Sobel, G.E. Bresnahan, W.E. Shell, R.D. Yoder, 'Estimation of Infarct Size in Man and Its Relation to Prognosis', *Circulation*, **46**, 640–648 (1972).
89. P.R. Reld, D.R. Taylor, D.T. Kelly, M.L. Weisfeldt, J. O'Neal Humphries, R.S. Ross, B. Pitt, 'Myocardial-infarct Extension Detected by Precordial ST-segment Mapping', *N. Engl. J. Med.*, **290**, 123–128 (1974).
90. G.M. Hutchins, B.H. Bulkley, 'Infarct Expansion Versus Extension: Two Different Complications of Acute Myocardial Infarction', *Am. J. Cardiol.*, **41**, 1127–1132 (1978).
91. J.E. Madias, K. Venkataraman, W.B. Hood, 'Precordial ST Segment Mapping. I. Clinical Studies in the Coronary Care Unit', *Circulation*, **52**, 799–808 (1975).
92. D. Mathey, W. Bleifeld, H. Buss, P. Henrath, 'Creatine Kinase Release in Acute Myocardial Infarction: Correlation with Clinical Electrocardiographic and Pathological Findings', *Br. Heart J.*, **37**, 1161–1168 (1975).
93. A. Marmor, B.E. Sobel, R. Roberts, 'Factors Presaging Early Recurrent Myocardial Infarction ('Extension')', *Am J. Cardiol.*, **48**, 603–610 (1981).
94. D.T. Mangano, 'Perioperative Cardiac Morbidity', *Anesthesiology*, **72**, 153–184 (1990).
95. S.L. Roberts, J.H. Tinker, *Cardiovascular Disease, Risk, and Outcome in Anesthesia*, ed. J.B. Lippincott, Philadelphia, 33–49, 1988.
96. M.J. London, D.T. Mangano, 'Assessment of Perioperative Risk', in *Advances in Anesthesia*, ed. R.K. Stoelting, Year Book Medical, Chicago, 53–87, Vol. 5, 1988.
97. D.T. Mangano, M. Hollenberg, C. Fegert, M.L. Meyer, M.J. London, J.F. Tubau, W.C. Krupski, 'Perioperative Myocardial Ischemia in Patients Undergoing Noncardiac Surgery. I: Incidence and Severity During the 4 Day Perioperative Period: the Study of Perioperative Ischemia (SPI) Research Group', *J. Am. Coll. Cardiol.*, **17**, 843–850 (1991).
98. L.G. Lenke, K.H. Bridwell, A.S. Jaffe, 'Increase in Creatine Kinase MB Isoenzyme Levels After Spinal Injury', *J. Spinal Disord.*, **7**(1), 70–76 (1994).
99. J.S. Tsung, S.S. Tsung, 'Creatine Kinase Isoenzymes in Extracts of Various Human Skeletal Muscles', *Clin. Chem.*, **32**, 1568–1570 (1986).
100. J.E. Adams, III, G.A. Sicard, B.T. Allen, K.H. Bridwell, L.G. Lenke, V.G. Davila-Roman, G.S. Bodor, J.H. Ladenson, A.S. Jaffe, 'Diagnosis of Perioperative Myocardial Infarction with Measurement of Cardiac Troponin I', *N. Engl. J. Med.*, **330**, 670–674 (1994).
101. C.R. Conti, 'Myocardial Infarction: Summary of the TIMI Trials', *Clin. Cardiol.*, **21**(7), 459–461 (1998).

102. K.P. Rentrop, 'Restoration of Anterograde Flow in Acute Myocardial Infarction: the First 15 Years', *J. Am. Coll. Cardiol.*, **25**, 1S–2S (1995).
103. W.B. Burns, C.J. Davidson, 'Thrombolysis or Primary Angioplasty? Reviewing the Evidence', *Emerg. Med.*, **28**, 18–33 (1995).
104. J. Abe, T. Yamaguchi, T. Isshiki, H. Naka, J. Taguchi, N. Ishizaka, K. Kurokawa, F. Saeki, Y. Ishizaka, K. Ui, 'Myocardial Reperfusion Can be Predicted by Myoglobin/Creatine Kinase Ratio of a Single Blood Sample Obtained at the Time of Admission', *Am. Heart J.*, **126**, 279–285 (1993).
105. A. Remppis, T. Scheffold, O. Karrer, J. Zehelein, C. Hamm, E. Grunig, C. Bode, W. Kubler, H.A. Katus, 'Assessment of Reperfusion of the Infarct Zone After Acute Myocardial Infarction by Serial Cardiac Troponin T Measurement in Serum', *Br. Heart J.*, **71**, 242–248 (1994).
106. F.S. Apple, T.D. Henry, C.R. Berger, Y. Landt, 'Early Monitoring of Serum Cardiac Troponin I for Assessment of Coronary Reperfusion Following Thrombolytic Therapy', *Am. J. Clin. Pathol.*, **105**, 6–10 (1996).
107. B. Lindahl, P. Venge, L. Wallentin, 'Relation Between Troponin T and the Risk of Subsequent Cardiac Events in Unstable Coronary Artery Disease', *Circulation*, **93**, 1651–1657 (1996).
108. C.W. Hamm, C. Heeschen, B.U. Goldmann, E. Barnathan, M.L. Simons, 'Value of Troponins in Predicting Therapeutic Efficacy of Abciximab in Patients with Unstable Angina', *J. Am. Coll. Cardiol.*, **31**, 185A (1998).
109. C.A. Clyne, L.J. Medeiros, K.I. Marton, 'The Prognostic Significance of Immunoradiometric CK-MB Assay (IRMA) Diagnosis of Myocardial Infarction in Patients with Low Total CK and Elevated MB Isoenzymes', *Am. Heart J.*, **118**, 901–906 (1989).
110. E.M. Ohman, P.W. Armstrong, R.H. Christenson, C.B. Granger, H. Katus, C.W. Hamm, M.A. O'Hanesian, G.S. Wagner, N.S. Kleiman, F.E. Harrell, Jr, R.M. Califf, E.J. Topol, 'Cardiac Troponin T Levels for Risk Stratification in Acute Myocardial Ischemia. GUSTO IIA Investigators', *N. Engl. J. Med.*, **335**, 1333–1341 (1996).
111. E.M. Geltman, A.A. Ehsani, M.K. Campbell, K. Schechtman, R. Roberts, B.E. Sobel, 'The Influence of Location and Extent of Myocardial Infarction on Long-term Ventricular Dysrhythmia and Mortality', *Circulation*, **60**, 805–814 (1979).
112. L.K. Newby, R.H. Christenson, E.M. Ohman, P.W. Armstrong, T.D. Thompson, K.L. Lee, C.W. Hamm, H.A. Katus, C. Cianciolo, C.B. Granger, E.J. Topol, R.M. Califf, 'Value of Serial Troponin T Measures for Early and Late Risk Stratification in Patients with Acute Coronary Syndromes', *Circulation*, **98**, 1853–1859 (1998).
113. R.H. Christenson, S.H. Duh, L.K. Newby, E.M. Ohman, R.M. Califf, C.B. Granger, S. Peck, K.S. Pieper, P.W. Armstrong, H.A. Katus, E.J. Topol, 'Cardiac Troponin T and Cardiac Troponin I: Relative Values in Short-term Risk Stratification of Patients with Acute Coronary Syndromes', *Clin. Chem.*, **44**, 494–501 (1998).
114. A.G. Olatidoye, A.H.B. Wu, Y.J. Feng, D. Waters, 'Prognostic Role of Troponin T Versus Troponin I in Unstable Angina Pectoris for Cardiac Events with Meta-analysis Comparing Published Studies', *Am. J. Cardiol.*, **81**, 1405–1410 (1998).
115. C.W. Hamm, B.U. Goldmann, C. Heeschen, G. Kreymann, J. Berger, T. Meinertz, 'Emergency Room Triage of Patients with Acute Chest Pain by Means of Rapid Testing for Cardiac Troponin T or Troponin I', *N. Engl. J. Med.*, **337**, 1648–1653 (1997).
116. J.C. Kahn, P. Gueret, R. Menier, P. Giraudet, M.B. Farhat, J.P. Bourdarias, 'Prognostic Value of Enzymatic (CPK) Estimation of Infarct Size', *J. Mol. Med.*, **2**, 223–231 (1977).
117. R. Roberts, A. Husain, H.D. Ambos, G.C. Oliver, J.R. Cox, Jr, B.E. Sobel, 'Relation Between Infarct Size and Ventricular Arrhythmia', *Br. Heart J.*, **37**, 1169–1175 (1975).
118. R.V. Essen, W. Merx, S. Effert, 'Spontaneous Course of ST-segment Elevation in Acute Anterior Myocardial Infarction', *Circulation*, **59**, 105–112 (1979).
119. P. Grande, C. Christiansen, K. Alstrup, 'Comparison of ASAT, CK, CK-MB, and LD for the Estimation of Acute Myocardial Infarct Size in Man', *Clin. Chim. Acta*, **128**, 329–335 (1983).
120. T.H. Lee, M.C. Weisberg, D.A. Brand, G.W. Rouan, L. Goldman, 'Candidates for Thrombolysis Among Emergency Room Patients with Acute Chest Pain', *Ann. Intern. Med.*, **110**, 957–962 (1989).
121. J.N. Adams, M. Jamieson, R.J. Trent, J. Rawles, K.P. Jennings, 'The Risks of Thrombolysis in Patients Without Acute Myocardial Infarction', *Int. J. Cardiol.*, **51**, 177–181 (1995).
122. P.O. Collinson, 'The Need for Point of Care Testing: an Evidence-based Appraisal', *Scand. J. Clin. Lab. Invest.*, **59**(S 230), 67–73 (1999).
123. P.O. Collinson, E.M. Ramhamadamy, S.B. Rosalki, J. Joffe, D.H. Evans, R.S. Fink, T.W. Greenwood, I.M. Baird, 'Diagnosis of Acute Myocardial Infarction from Sequential Enzyme Measurements Obtained Within 12 hours of Admission to Hospital', *J. Clin. Pathol.*, **42**, 1126–1131 (1989).
124. P.O. Collinson, P.J. Stubbs, 'Four Hour ''rule-in'' Diagnosis of Acute Myocardial Infarction for Early Risk Stratification by Creatine Kinase Increment (CK Change)', *Ann. Clin. Biochem.*, **33**, 308–313 (1996).
125. B. Lindahl, P. Venge, L. Wallentin, 'Early Diagnosis and Exclusion of Acute Myocardial Infarction Using Biochemical Monitoring. The BIOMACS Study Group.

Biochemical Markers of Acute Coronary Syndromes', *Coron. Artery Dis.*, **6**(4), 321–328 (1995).
126. B. Lindahl, P. Venge, L. Wallentin, 'Troponin T Identifies Patients with Unstable Coronary Artery Disease Who Benefit from Long-term Anti-thrombotic Protection. Fragmin in Unstable Coronary Artery Disease (FRISC) Study Group', *J. Am. Coll. Cardiol.*, **29**(1), 43–48 (1997).
127. P.O. Collinson, E.M. Ramhamadamy, B.J. Stubbs, S.B. Rosalki, H.M. Garrat, D. Mosely, D.H. Evans, F.S. Fink, I.M. Baird, T.W. Greenwood, 'Rapid Enzyme Diagnosis of Patients with Acute Chest Pain Reduces Patient Stay in the Coronary Care Unit', *Ann. Clin. Biochem.*, **30**, 17–22 (1993).
128. J.R. Hedges, W.B. Gilber, G.P. Young, J.W. Hoekstra, C. Slovis, R. Aghababian, M. Smith, M. Rubinson, 'Multicenter Study of Creatine Kinase-MB Use: Effect on Chest Pain Clinical Decision Making', *Acad. Emerg. Med.*, **3**, 7–15 (1996).
129. W.B. Gibler, J.P. Runyon, R.C. Levy, M.R. Sayre, R. Kacich, C.R. Hattemer, C. Hamilton, J.W. Gerlach, R.A. Walsh, 'A Rapid Diagnostic and Treatment Center for Patients with Chest Pain in the Emergency Department', *Ann. Emerg. Med.*, **25**(1), 1–8 (1995).
130. R.J. Zalenski, R.J. Rydman, M. McCarren, R.R. Roberts, B. Jovanovic, K. Das, E.K. Mensah, L.M. Kampe, 'Feasibility of a Rapid Diagnostic Protocol for an Emergency Department Chest Pain Unit', *Ann. Emerg. Med.*, **29**, 99–108 (1997).
131. R.R. Roberts, R.J. Zalenski, E.K. Mensah, R.J. Rydman, G. Ciavarella, L. Gussow, K. Das, L.M. Kampe, B. Dickover, M.F. McDermott, A. Hart, H.E. Straus, D.G. Murphy, R. Rao, 'Costs of an Emergency Department-based Accelerated Diagnostic Protocol Versus Hospitalization in Patients with Chest Pain. A Randomized Controlled Trial', *J. Am. Med. Assoc.*, **278**(20), 1670–1676 (1997).
132. A.H.B. Wu, F.S. Apple, W.B. Gibler, R.L. Jesse, M.M. Warshaw, R. Valdes, Jr, 'National Academy of Clinical Biochemistry Standards of Laboratory Practice: Recommendations for the Use of Cardiac Markers in Coronary Artery Disease', *Clin. Chem.*, **45**(7), 1104–1121 (1999).

Biosensor Design and Fabrication

G.A. Urban
Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

1 Introduction **1**
2 Principles of Biosensors **2**
3 Clinical Diagnostics **2**
4 Biosensor Devices **3**
4.1 Optical Microtransducers for Biosensors 3
4.2 Biosensors Based on Acoustic Transducers 4
4.3 Biosensors Based on Calorimetric Transducers 5
4.4 Electrochemical Biosensors 5
5 Microanalytic Systems **10**
5.1 Electrolyte Sensing Systems 10
5.2 Sensing Systems for Metabolic Parameters 10
5.3 Cell-based Systems 12
5.4 Affinity-based Systems 13
6 Conclusions and Outlook **13**
Abbreviations and Acronyms **13**
Related Articles **14**
References **14**

This paper presents the use of biosensor devices for bioanalytical purposes. In contrast to common analytical tools using sophisticated and bulky devices, biosensors should be able to perform measurements in untreated samples in a simple and inexpensive manner. As a model creating biosensors serves nature which produces a variety of selective molecules during evolution that can be used in technical sensor devices. Combining such natural molecules with methods from microelectronics, mass fabrication of laboratories on chip seems feasible. Emerging markets for such tools can be seen in medicine, drug screening and biotechnology.

After explaining the principles of biosensors different transducing methods are explained showing advantages and disadvantages for creating biosensors. The greatest emphasis is laid on electrochemical and optical principles which exhibit an overwhelming experience and the largest impact on future developments. Despite such developments, acoustic and calorimetric transducers are also explained. As biological sensing agents enzymes, antibodies and DNA are described with appropriate immobilizing procedures.

Additionally, new aspects for creating a laboratory-on-chip using microanalytical systems (μTAS) including microfluidics and actuator modules are given for creating sensing systems for metabolic parameters and also cell-based and affinity-based systems for screening purposes. The purpose of this section is to give an overview and to explain the present possibilities for creating biosensor systems and ends up with an outlook in the near analytical future for new sophisticated and miniaturized analytical systems.

1 INTRODUCTION

Biosensors are fascinating analytical tools that attracted great attention in the past, promising to revolutionize the whole analytical and diagnostic world, causing the prediction of a huge market in the near future.[1] However, these general predictions have not yet been fulfilled due to technical and production problems and for market reasons. Unquestionably, the use of biological highly selective and sensitive sensing parts in the vicinity of a transducer, which is the original definition of a biosensor, still creates a lot of interest. Nature has developed similar systems during evolution; all biological systems inherently not in thermodynamic equilibrium need a variety of sensing components to maintain control and stability which might serve as a model for complex multibiosensor systems for industrial applications. However, despite this challenging analytical approach until now the only types of biosensors with market success are single-shot glucose sensors and biosensor modules for clinical analyzers.

In the 1960s biosensors were fabricated conventionally with macroscopic membranes and transducers, gluing wires together and dipping them in membrane solutions. For small-scale production, as required for clinical analyzers, this was a realistic approach. It is interesting to note that even though biosensors have a long history very few are able to perform measurements in undiluted biological fluids for long-term applications. Changing the daily optical-based glucose check for diabetes patients to electrochemical sensors, with millions of sensors sold per year, the method of production also changed, and screen printed sensors conquered the market.

However, for further diagnostic and analytical procedures such as continuous metabolic monitoring, high throughout screening (HTS), DNA, immunological diagnostics and point of care management, new ways for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

biosensor design and fabrication have to be developed. Multiple biosensor arrays able to perform measurements over an extended period in undiluted biological media with the possibility of mass production have to be developed.

Due to the problems related to biosensors in principle, finding appropriate immobilizing strategies, overcoming transducer-related problems and finding the right production method, only a few activities take place in research and development dealing with all the problems. Up to now, no device is commercially available for use in a daily routine.

Some activities are related to silicon-based technology that is not necessarily the best for the apparent user problems. A combination of microelectronic-compatible immobilizing strategies with robust electrochemical microtransducers and proper analyte diffusion and enzymatic pathway control leads to an interesting candidate for creating multianalyte-multitarget biosensor arrays. This is described in more detail in the following sections.

2 PRINCIPLES OF BIOSENSORS

Small, integrated and reliable sensing elements are necessary for metabolic monitoring, decentralized laboratories, and for drug screening. Such devices can be realized by ''sensors'', elements receiving physico-chemical information and transforming such information into an electric signal. To measure metabolic parameters selectively ''biosensors'' are required. A schematic drawing of a biosensor is given in Figure 1.

A biosensor is a device incorporating a biological sensing agent either intimately connected to or integrated into a transducer using the old definition of tuner and is capable of performing a variety of clinical analyses in principle.[2] The most advanced biosensors use enzymes with the glucose sensor as the main representative. The commonly used biochemical systems utilize the enzyme glucose oxidase (GOx) and measure the product H_2O_2 electrochemically. In this case, GOx converts the substrate glucose and oxygen to gluconolactone, H_2O_2 and generated heat ΔH (Equation 1).

$$\text{Glucose} + O_2 \xrightarrow{\text{GOx}} \text{gluconolactone} + H_2O_2 + \Delta H \quad (1)$$

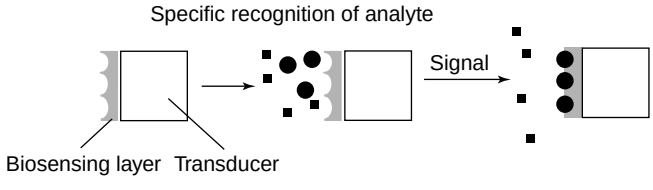

Figure 1 Schematic drawing of a biosensor.

Many different enzymes commercially available can be used for biosensors in principle. Also affinity-based devices with immobilized antibodies or receptors can be used to get highly sensitive biosensors.[3] Many biosensing structures are available ranging from enzymes, antibodies, receptors, tissue slices and whole cells.[3,13] Recent developments incorporate organ or tissue slices or whole cells on a transducing element to get a more biologically based environment, thus performing reliable measurements.[4]

There are also several transducing principles available: The systems best investigated use electrochemical sensors but there are also acoustic devices, calorimetric and optical sensors available. To obtain a general overview of biosensors there are some excellent reviews available.[2,3,5–7]

3 CLINICAL DIAGNOSTICS

Diagnostics has a high priority in medical practice, and the microelectronic revolution has played an important part in establishing radiological diagnostic tools. However, to obtain a deeper insight into metabolic functions of the body, clinically important metabolic parameters have to be measured continuously in different body compartments. Until now, this task could only be done with certain restrictions using radioactive tracers or nuclear magnetic resonance (NMR).[8] Such technologies are not able to perform continuous measurements of metabolic parameters required for monitoring in the intensive care unit or multianalyte-multitarget detection for rapid screening in pharmacology in the near future.

Metabolic parameters are accessible externally only by standard clinical laboratory methods. In recent years, analytical laboratory instrumentation has become highly sophisticated, with the ability to routinely process hundreds to thousands of samples a day. However, modern clinical and pharmacological practice urgently requires miniaturized monitoring tools and point-of-care systems that cannot be realized by standard clinical laboratory methods.[9–12] Therefore, miniaturized *and* integrated biosensor systems are required consisting of glucose-, lactate-, urea-, creatinine-, immuno-, cell-based and DNA-sensor arrays, which suggests the use of microsystem production technology.[13,14] The devices of greatest interest to be used for whole blood testing and monitoring in the intensive care unit, for emergencies, for satellite laboratories and in new generations of analytical instruments analyze glucose which tops the list of analytes that clinicians would like to measure continuously in vivo.[15] The discrete measurement of glucose for diabetics with a market volume of 3×10^9 US dollars in 1998 is of particular interest.[11]

4 BIOSENSOR DEVICES

The ultimate goal in producing biosensor microsystems is the realization of a complete microanalysis system or laboratory-on-chip, and therefore new combinations and integration of technologies have to be applied. Microsystem technology with the claim of interdisciplinarity could be a good way to produce several thousands to millions of devices per year in typical production plants.

Different production technologies, such as silicon, thin-film and screen printing technologies, are available with different features and are well suited for different market sizes and customer needs (Table 1).[6] Together with genetic engineering and material sciences, these technologies yield promising expectations of new routes to achieve broader use of biosensor systems.

For choosing the right production technology a user- and market-driven approach is recommended: application areas are clinical, pharmacological and biological fields as well as process control with several new and rapidly growing markets, e.g. point of care measurements and pharmaceutical screening topics. These cases require decentralized portable analytical devices for unskilled personnel with high reproducibility and reliability at low production costs. If there is no need for extreme miniaturization, and no advantage comes from a special technology (e.g. Si-monolithic integration), a low-tech production such as screen-printing or polymer-based technologies is recommended.

An often-cited argument is the benefit of using standard silicon technology. However, there are not too many biosensor applications with a high market volume to make use of the benefits of silicon technology, and also very few biosensors need the integration of microelectronic components in the vicinity of the transducer. Therefore, already marketed microelectronic biosensors make use of thin-film and screen printing technology!

The successful introduction of new technologies in emerging markets can be seen in the recent rapid progress of microelectronics. The tremendous advance of this technology was inherently connected to the completely new production methods allowing the reproducible production of a large quantity of identical products and devices. The benefits of the mentioned technology such as cheap mass production, high reproducibility and capability of integration of various functions on a miniaturized scale raise the expectation of success for microelectronic sensors. Up to now this expectation has not been entirely fulfilled; primarily physical sensors such as pressure, acceleration sensors or charge coupled dence (CCD) imaging chips have been successfully introduced into the market.

Microelectronics plays an important part in medical imaging techniques, data acquisition and processing but is less important in medical sensors. This will probably change by the introduction of new ways of designing analytical systems with microsystems. Additional to the required miniaturization and production of the transducer, sensor performance is of importance. It should be stable for a sufficiently long time, exhibiting full response without pretreatment of samples. Thus, designers of a biosensor microsystem have also to overcome the problems of choosing the right transducers, optimizing them and immobilizing biological sensing agents onto them on a large-scale level. In the realization of biosensor microsystems, the transducing element as the base device is of primary importance.

4.1 Optical Microtransducers for Biosensors

Optical sensors are interesting analytic tools capable of performing a variety of different analyses where most of the work has been done on measuring pH, P_{O_2} and P_{CO_2}.[16,17] The biosensing devices are mainly focusing on affinity principles such as e.g. antibody–antigen reactions and are based on surface plasmon resonance (SPR), grating couplers or interferometers.[18–20] It seems possible to obtain stable and highly sensitive devices based on these principles, and further investigations can lead to miniaturized sensor modules at reduced cost, size and complexity.[21] In diagnostics and rapid screening environments label-free in situ monitoring of molecular binding events are of outstanding interest.

Table 1 Overview about different important production technologies

	Si-technology	Thin-film technology	Screen printing technology
Plant price (rel.)	100	10–50	1
Degree of miniaturization	Sub µm	Sub µm	0.1 mm
Flexibility in developing special processes	Very low	Very high	Medium
Minimum pieces per year (approx.)	10^5	10^3	10^2
Cost per piece			
High production numbers	Very low	Low	Low
Low production numbers	Very high	High	Low
Reproducibility, reliability	Very high	High	Medium

The surface plasmon principle was commercialized as one of the first direct biosensor devices.[22] A SPR occurs when surface plasmon waves are excited at a thin metal/liquid interface, which is the sensor surface. A polarized light beam is directed at, and reflected from, the side of the surface not in contact with the sample. The SPR causes a reduction in the reflected light intensity at a specific combination of angle and wavelength. The intensity of reflected light exhibits a pronounced maximum at a distinct angle, which is a function of the refractive index of the metal/liquid interface that is seen by the evanescent incident wave. Molecular binding to the sensor surface causes changes in the refractive index close to the surface that can be detected as changes in the SPR signal. The principles of the optical set-up are seen in Figure 2; the reflected light beam I is changed during a surface binding event to beam II.

However, the detection of a direct binding assay is limited to a mass coverage of $10\,pg\,mm^{-2}$ in the case of SPR and is approximately 10 times better for interferometers.[20,22] Not only sensitivity is a limiting factor, but also unwanted unspecific adsorption of proteins can reduce selectivity.

Integrated optical devices combine microelectronic production technology with the inherent advantages of optical sensing. Many of these developments are at an early state of research but a variety of optical biosensors can be realized in principle. Apart from the described problems, direct binding approaches suffer from the drawback that a highly paralleled set-up is not easily possible. Instead, optical methods using labeled biorecognition molecules and chemoluminescence can be applied.[23] CCD cameras are able to perform reliable simultaneous optical detection of sensor array in a titer plate format with low detection limits. Additionally, microsystems using micro-optical and diffractive elements can be used to get an optical analysis system.[24,25] Integrated optical device manufacturing is commercially available today, and nearly all optical elements can be integrated and miniaturized on chip.[26] Sophisticated optical systems using the label approach are also of interest for DNA sensor arrays using chip technology for DNA immobilization and a confocal laser microscope for the signal read out (see section 5.4).

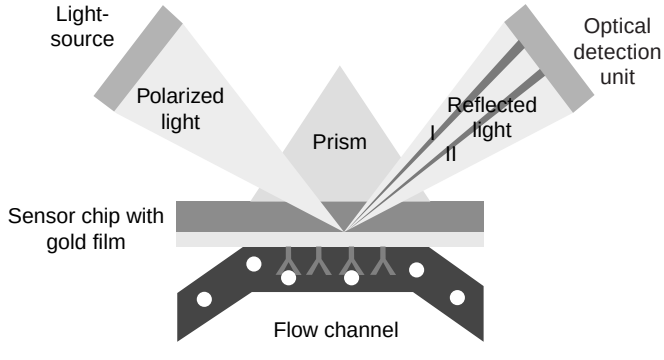

Figure 2 Principles of the optical set-up.

4.2 Biosensors Based on Acoustic Transducers

The principle of acoustic devices is the change of mass on a surface during an adsorption or binding event. Three types of transducer mechanisms are used: Bulk Acoustic Wave (BAW) using shear oscillations of piezoelectric crystals, Surface Acoustic Wave (SAW) using wavemode propagation, and Lamb Waves (LW) which are flexure acoustic waves of a free standing piezoelectric layer. Such detection methods are very sensitive, and submonolayers can be detected that are routinely used with BAW devices to control the evaporation process in microelectronic technology.

Many applications have been reported using piezoelectric sensors for gaseous components. Acoustic transducers such as piezoelectric microbalance and SAW devices that are in an early state of research have been developed primarily for volatile analytes and recently for affinity reactions.[27] The use of acoustic devices in liquids has long been inhibited by the problem of the high energy losses at the solid/liquid interface.

The commonly used devices utilize commercially available and robust BAW transducers. However, the sensitivity of acoustic devices is directly proportional to frequency and inversely proportional to the sensor area giving SAW devices an advantage.[7] A SAW device consists of a piezoelectric substrate and a pair of interdigitated electrodes that act as sensing and receiving transducers (Figure 3). By applying high frequency electrical waves on the electrodes the electric field generates acoustic waves at the piezoelectric surface (Rayleigh waves). The distribution of a SAW is a function of the surface conditions such as mass loading or changing the elastic properties of the material. All changes due to adsorbed species will be measured with high sensitivity. Therefore selectivity means a problem for all acoustic devices, because unwanted adsorption of interfering

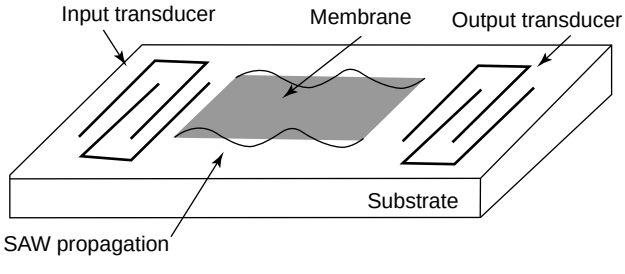

Figure 3 Basic structure of a SAW chemical sensor.

substances is detected in all cases, e.g. humidity is a general problem for gas sensing.

Another source of problems with acoustic devices for assaying binding events is the liquid analyte phase. All variations in density and viscosity of the liquid influence the acoustic signal. Additionally, if a binding event changes the volume of the used biological substance, e.g. receptors, the buoyancy force will also change. If used in biological samples the unwanted adsorption of proteins or tissue compartments remains a severe problem. In any case, SAW sensors are candidates for proper test systems with specified constant analyte matrix.

Recently encouraging results for immunosensing were presented using low-loss SAW filters based on horizontal shear waves with lithium tantalate.[28] Such devices exhibit a typical attenuation of 4 dB in water. Using an appropriate polyamide insulation with subsequent immobilization of dextrans an immunosensor could be produced with sufficient sensitivity and low unspecific response.

4.3 Biosensors Based on Calorimetric Transducers

Calorimetric biosensors are based on the detection of the heat of biological reactions caused by changes in enthalpy. Such a microcalorimetric sensing principle is very versatile because of the exothermic nature of nearly all enzymatic reactions and was introduced in conventionally constructed devices very early.[29,30] If a chemical reaction is occurring in an adiabatically isolated mass, which is defined as a calorimeter, then the temperature of the mass will change according to Equation (2)

$$\Delta T = \frac{\Delta Q}{MC_c} \tag{2}$$

where ΔT is the temperature change, ΔQ the energy difference, M the total mass of the calorimeter and C_c the total thermal capacity of the calorimeter.

Conventional calorimetric biosensors with thermistors as transducers were invented early by proposing a thermal biosensor in a flow stream.[31] So far the design of enzyme thermistors does not entirely match the market demand, but it seems well-suited for special applications utilizing discrete thermistors with immobilized enzymes or with separate enzyme columns.[30–32] A drawback for performing calorimetric measurements in a conventional way is the lack of well-matched thermistors. Recently a thin-film thermistor based on amorphous germanium has closed this gap.[33]

As an alternative, transducer thermopiles are well suited for performing differential measurements due to their underlying physical principle to measure temperature differences directly with high common-mode rejection.[34] Using conventional wire-based thermocouples, there are some difficulties related to size, sensitivity, and mechanical stability. If silicon technology is involved all thermal sensors suffer from the high thermal conductivity of silicon dramatically decreasing their sensitivity.[35]

By use of micromachining and integrated silicon technology powerful thermal biosensors can be realized. Using a thermopile integrated on a thin micromachined silicon membrane reduces thermal losses due to the substrate and allows excellent performance to be achieved.[34]

To realize glucose-, urea- and penicillin sensors in a flow injection mode appropriate enzymes can be immobilized though the measuring range is very limited in all reported cases due to enzyme kinetics, and a more sophisticated set-up is required.[34] For clinical applications the calorimetric method was extended for whole blood and in vivo measurements of glucose.[36,37] With microelectronic and micromachining tools, an array of thermistors covered by different enzymes can be easily produced leading to a multianalyte detection using the same transducing principle.[38] Such technologies allow the sample volume to be reduced to 1 μL and the response time is well below 1 min. A schematic drawing of a multianalyte device is shown in Figure 4.

4.4 Electrochemical Biosensors

Electrochemical biosensors are most commonly used and based on well-known transducing principles that have been investigated intensively during the last decades. Electrochemical transducers can be divided into conductometric, potentiometric, and amperometric measuring principles.

4.4.1 Conductometric Biosensors

A simple transducing principle is the measurement of impedance between planar electrodes changed by altering

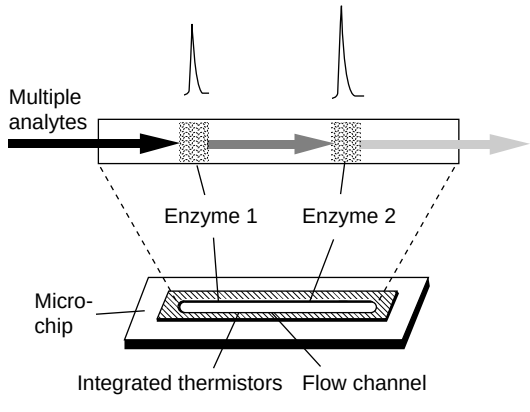

Figure 4 Integrated multienzyme thermistor device for the simultaneous assay of more than one analyte.

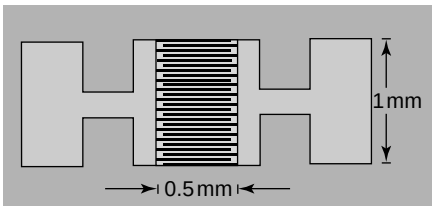

Figure 5 Planar interdigitated electrodes.

the ionic content of the measuring system. If biological sensing elements are immobilized on such electrodes, biosensors can be realized using changes of the conductivity in the electrolyte solution caused by biological reactions or changes of the double layer capacitance by binding (affinity) reactions of the biological molecules immobilized on the electrode surface.

Planar interdigitated electrodes for impedimetric measurements can be produced by means of microelectronic technology (Figure 5).[39] Although this measuring principle has been known for a long time further investigations are required to understand the transduction principle in more detail for biosensing.[40] One obstacle is the high concentration of the ionic background always present in biological samples, a second one is associated with electrode fouling.[41] One advantage of this general transducing principle is the possibility of creating biosensing arrays.[39]

Using highly miniaturized microelectrode arrays with interdistances in the 0.5 µm range affinity-based sensors can be realized to measure DNA hybridization.[42] Encouraging results recommend such a device for more investigations to get closer to practical applications and commercialization.

4.4.2 Potentiometric Transducers

Potentiometric measurements involve the determination of the electrical potential between two electrodes at zero current flow. A reference electrode has to establish a constant potential independent of the electrolyte used which is not easily performed even with standard electrochemical systems. For miniaturized reference electrodes various constructions were investigated using different electrolyte chambers and porous plugs.[43,44]

The working electrode has to respond directly to the analyte to be measured, mostly pH and ion concentration. Such devices respond to changes in the analyte ion activity by varying the electrical potential described by the Nernst Equation (3)

$$E = E_o + \frac{RT}{zF} \ln a_1 \qquad (3)$$

where E is the electromotive force, E_o the standard potential, R the gas constant, T the absolute temperature, z the exchanged electrons, F the Faraday constant and a_1 the activity of the species to be detected.

The first representative of a potentiometric sensor was the pH-glass electrode invented in 1906.[45] Many ion-selective electrodes have been developed including more recently those based on neutral carrier membranes and the microelectronic ion selective field effect transistor (ISFET).[46,47]

Such an ISFET was originally used for pH detection but by casting with ion-selective membranes a lot of different ion-selective sensors can be obtained.[48,49] Even a multiparameter electrolyte sensitive chip for clinical applications was developed.[49] To obtain a microelectronic biosensor, a biological substance has to be immobilized onto the gate insulator of the ISFET.

The enzyme-containing field effect transistor (ENFET), is fabricated from an ISFET by casting a thin membrane containing the enzyme over an ion-selective membrane or an inorganic gate insulator. The underlying mechanism of ISFETs and the enzyme reactions are described in Turner et al.,[2] Hall,[3] Scheller and Schubert[5] and Lundström et al.[50] A schematic set-up of an ENFET is shown in Figure 6 where the enzyme is immobilized in a thin membrane over an ion-selective membrane.

Local changes in the ion activity due to substrate conversion by the immobilized enzymes have been measured using pH, fluoride-sensitive field effect transistors or ammonia-sensitive field effect transistors.[26,51,52] Immunosensing with the field effect transistor gate by the

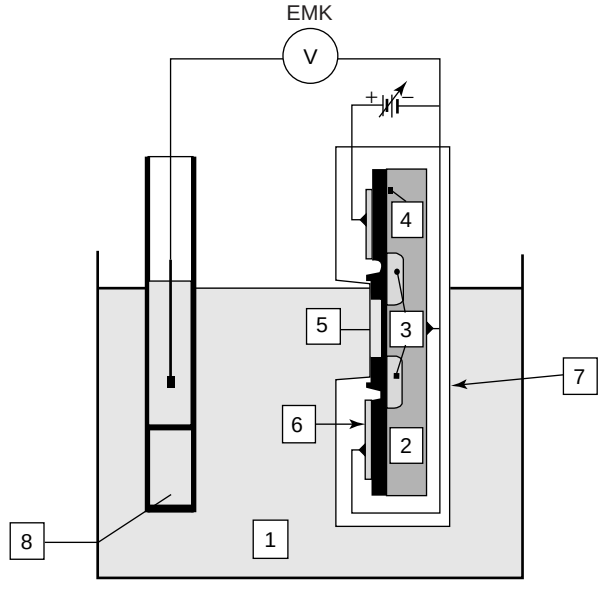

Figure 6 Schematic set-up of an ENFET with the enzyme immobilized in a thin membrane over an ion-selective membrane. 1, Analyte; 2, p-doped silicon; 3, n^+-doped silicon; 4, SiO_2; 5, enzyme-containing membrane; 6, contact pad; 7, encapsulated ENFET; 8, reference electrode.

direct detection of protein interaction has been described in the literature but direct detection of protein interaction with a pH-sensitive gate is limited by fundamental principles.[53]

Commercialization of ISFET devices is limited due to several problems. First, the ISFET device itself reveals some problems: drift, sensitivity to ambient light, dependence on buffer capacity in the sample, encapsulation problems and sensitivity against contamination.[50] The provisionally used gate materials for pH sensing, SiO_2 and Si_3N_4, exhibit a large drift and do not show the expected Nernstian response.[47,54] Special gate materials such as Al_2O_3 or Ta_2O_5 result in an enhancement of slope and better drift properties, though with large differences in sensor performance reported.[55] Additionally the described special gate materials cannot be deposited in a standard complementary metal oxide semiconductor production plant.[49,56]

Furthermore there are severe problems related to packaging, and an attempt was made to create suspended gates and sophisticated assembling procedures.[50,57] Another drawback for ISFET devices is the relatively large area required for a device leading to increased costs. These problems are extended for biosensor devices because of the additional integration of a biological component on a planar device surface.

A fundamental limitation of typical ENFET is the biochemical principle used as the sensing reactions have to be a pH change or a variation of surface potential. In the case of glucose sensing a double enzyme system has to be used to get optimum sensitivity which increases the complexity of the system.[58]

Different approaches were reported for enzyme immobilization and membrane deposition including drop-on techniques, ink-jet printing, and photolithographically patterned enzyme membranes.[51,59–62]

Integrated multibiosensors based on ISFET can be produced by using a photosensitive polymer that incorporates the enzyme and that is spun onto the gate (Figure 7).[7] However, the reported measuring range for glucose is rather weak and some cross-sensitivities due to problems regarded with buffer type and capacity are also reported. Another approach immobilizes enzymes by printing with an ink jet nozzle.[60] A photopolymer was used to create enzyme solution pools and to protect the ISFET surface from mechanical damage. To these pools, about 100 drops of enzyme solution were added. Using such a method, an integrated glucose and urea sensor can be produced. Again, however, the measuring range is very limited.

A possibility to extend the linear range and to perform measurements in whole blood is to use a further diffusion-limited membrane.[63] Unfortunately, such an approach slows down the sensor response; nevertheless, measurements in serum and biological fluid are possible.

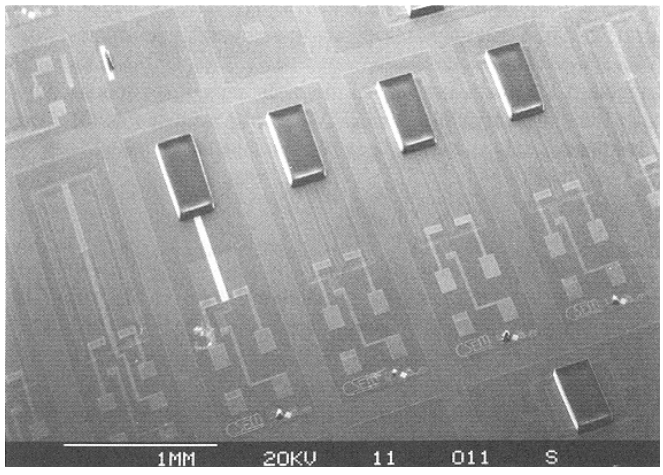

Figure 7 Microscopic view of polymerized membranes on ISFET insulating gates.

As nearly all ENFET approaches lead to insufficient or unsatisfactory sensor performance, no ISFET-based biosensor has been fully commercialized for wider applications to date. Although the ISFET exhibits a long history, a lot of work still remains to be done.

Another approach is the use of the potentiometric principle with planar thin film electrodes on a separate chip but in close vicinity to a field effect transistor input amplifier. Glucose and urea chips are now on the market commercialized by the i-stat company. These sensors are based on ion sensitive electrodes placed on silicon wafers. The problems of stability are circumvented by a simple on-chip calibration procedure and by the use of such microelectronic electrodes as disposable single-shot probes.[64,65]

4.4.3 Amperometric Biosensors

The most frequently used biosensors are based on amperometric principles, using the Faraday current derived from a redox reaction at an electrode. A distinct voltage controlled by a potentiostat is applied between a working electrode and the electrolyte where all redox species are electrochemically converted. This results in a stationary current following Equation (4) in the case of a diffusion-controlled measurement of redox species.

$$I = \frac{nFADc_o}{d} \tag{4}$$

where I is the diffusion limited current, n the exchanged electrons per reaction, F the Faraday equivalent, A the electrode area, D the diffusion constant of the analyte, c_o the analyte concentration and d the thickness of the diffusion layer.

In contrast to the potentiometric principle where nearly no transport of analyte to the sensor occurs,

the amperometric measurement requires the control of Faraday current at the electrodes and the diffusion of an analyte towards the electrode. This divides the problems of amperometric sensing devices in transducer- and membrane-related ones. Miniaturized enzymatic amperometric biosensors produced by means of thin and thick film technology have been investigated extensively in recent years.[66–71] Electrochemical analysis using planar thin film metal electrodes as transducer can be done in vitro with high performance.[72]

At present, nearly all of the cited amperometric microelectronic sensors intended for long-term application are not able to work well in undiluted blood, and only a few conventional macroscopic sensor systems are able to measure glucose and lactate concentration in undiluted media for an extended period.[73,74] Transducer- and biomembrane-related difficulties have to be overcome with electrochemical transducers for measurements in undiluted biological media:

4.4.3.1 Transducer The biosensor systems best investigated use H_2O_2 producing oxidases. The subsequent oxidation of H_2O_2 on the working electrodes at a potential of 400 mV to 650 mV versus an Ag/AgCl reference electrode serves as a transducing mechanism.

One of the problems with amperometric devices in biological samples is the degradation of electrochemical performance over time due to contamination of the electrode.[75] Therefore surface modifications or special electrode materials, e.g. carbon, are needed and the electrodes have to be covered with functional membranes.[76]

Another problem is the unwanted oxidation of other substances such as paracetamol, ascorbic acid or uric acid. The erroneous readings deriving from these substances can be overcome by differential methods using an enzyme-covered electrode and a blank electrode. Such systems are able to compensate for electrochemical interference but cannot hinder fouling of the electrodes.[77]

A further method is the use of mediated devices allowing reduced working potentials down to +200 mV where the interfering substances are not oxidized.[78] However, the drawback of mediator-based devices in long-term applications is the leakage of mediators.[69] Mediated devices are an excellent way to overcome the problems mentioned for short-term measurements.

The first commercially available disposable electrochemical device for measuring glucose in whole blood was a mediated system produced by screen printing.[79] Several companies are now following with similar approaches.[80,81]

For long-term operations, other approaches have to be used. One possibility is the direct ''wiring'' of the enzyme to the electrode. ''Wiring'' is accomplished by immobilizing the enzyme in an electron conducting redox hydrogel.[82,83] The hydrogel, being water rich, is permeable to both analyte and product. It conducts electrons through collisions between its mobile chain segments, the electrons being transferred from redox segment to redox segment by self-exchange. Because the redox functions are chemically bound to the polymer, they are not leached by the fluid. The redox potential of the enzyme ''wired'' polymer is sufficiently reduced to avoid unwanted oxidation of interferents.[84]

Another approach to protect the Pt working electrodes against fouling and to prevent erroneous reading due to electrochemical interference, an electropolymerized semi-permeable membrane can be utilized, which can be done on a wafer level.[85]

The reported semi-permeable membrane consists of an electropolymerized diaminobenzene polymerized in phosphate buffer (pH = 7). The polymerization was done by cycling the potential between 200 mV and 800 mV for a certain period. In principle, such an electrode modification hinders fouling in an excellent way.[86]

4.4.3.2 Biomembrane Membrane fouling caused by protein adsorption is another problem for all types of amperometric sensors that leads to a decrease in sensitivity. To overcome the described problems apart from the dilution method, the solution is to utilize complex membrane systems.[74,77] In addition, problems to immobilize biological substances and to establish proper diffusion limitation have to be overcome. In the case of the cited mediated disposable glucose sensor screen printing of free enzymes and mediator was used. Due to the solubility of both substances only single shot applications can be performed. For long-term applications more sophisticated methods have to be used.

A further problem for biosensor microsystems is the immobilization procedure of biosensing structures, e.g. enzymes. Again different approaches were tried including drop-on techniques, ink-jet printing, spray techniques, electropolymerization, lift-off techniques, and photolithographically patterned enzyme membranes.[66–71,87]

A miniaturized planar amperometric glucose sensor can be placed on sapphire substrates to overcome problems with conducting silicon substrates. Thin film metal electrodes are covered with an enzyme layer patterned by a lift-off technique.[71] This sensor exhibits a fast response time of 30 s but the linear measuring range is poor.

The continuous in vivo measurement of glucose by a sensor implanted in patients is of great importance. Such monitoring can lead to a better control of normoglycaemia, a better quality of life and a hypoglycemic alarm which is of outstanding importance.[88] For in vivo measurements, another planar glucose biosensor with Pt electrodes on a silicon substrate was developed (Figure 8).[72]

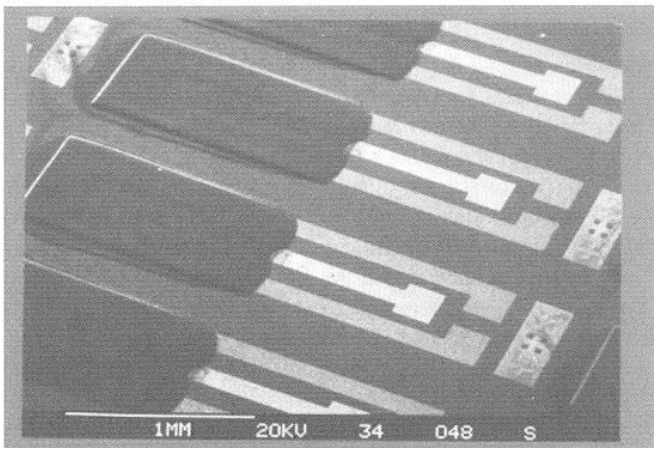

Figure 8 Three-electrode structures covered with an enzymatic GOx membrane.

The enzyme GOx was immobilized by the well-known glutaraldehyde–bovine serum albumin method and the complete sensor was covered subsequently by a polyurethane membrane. This silicon chip had to be sawed and assembled on a flexible carrier for application as an in vivo sensor and was successfully evaluated in rats.[89] This sensor gives encouraging results in aqueous solutions and subcutaneous applications. Drawbacks of this device are the complicated mounting and difficult and cumbersome assembling procedures.

Another approach places the Pt electrodes directly on a flexible polymer carrier.[90] The company Eli Lilly developed a three-electrode transducing system based on a polyimide carrier with electroadsorbed enzyme and a highly oxygen permeable membrane covering the sensor. Such a system was tested in vivo, and published results seem encouraging. Due to general problems with implanted sensors, company policies cancelled this project.

For microelectronic production and to obtain clinical reliable sensors the ultraviolet initiated free radical crosslinking of the polymer directly on the substrate is a great advantage by designing the physical-chemical properties of the membrane. An interesting development is the use of planar biosensors with immobilized enzymes incorporated in the photopatterned hydrogel poly(vinyl alcohol)-stilbene.[70] This leads to a convenient production process compatible with microelectronic technology. Such sensor devices are encouraging and interesting showing the possibility for creating microelectronic amperometric biosensors but suffer from insufficient measuring range and are not tested in whole blood required for clinical applications.

A combination of different technologies such as electropolymerization and photopatternable enzyme membranes can lead to reliable sensor systems. An easy to handle thin film process was developed for immobilizing different H_2O_2 producing enzymes.[86] The hydrogel layer, containing e.g. the enzyme GOx, can be patterned by photolithography and placed selectively on the individual working electrode.[77] An uppermost photopatternable membrane was introduced containing the enzyme catalase decomposing excess H_2O_2 into O_2 and water in order to prevent the release of the cytotoxic agent H_2O_2 into the biological environment. This membrane also prevents electrode fouling by blood components because of the low protein deposition characteristics of poly(hydroxy methacrylate).[91]

To increase the diffusion path and to separate the H_2O_2 source (GOx) and the H_2O_2 sink (catalase), an additional poly(hydroxy methacrylate) membrane was placed between the GOx and catalase membrane. A schematic drawing of this device is shown in Figure 9.

Multianalyte chips can be produced in this way using microelectronic technology. The sensors do not show any dependence on interferences. They exhibit extended linear ranges with high sensitivities and low residual currents. Due to the wafer processing a high reproducibility can be obtained.[86] The thin hydrogel membranes exhibit a fast response time of 25 s to 98% equilibrated signal and a fast hydration time of several minutes. The sensor chips can be stored dry for at least 3 months at 4 °C without changes in performance. The long term operational stability in undiluted bovine serum spiked with analyte is more than 1 month at 37 °C for the lactate sensor.

This is an example that multienzyme sensor devices can be designed to obtain precise, reliable, integrated biosensors with extended measuring range for clinical

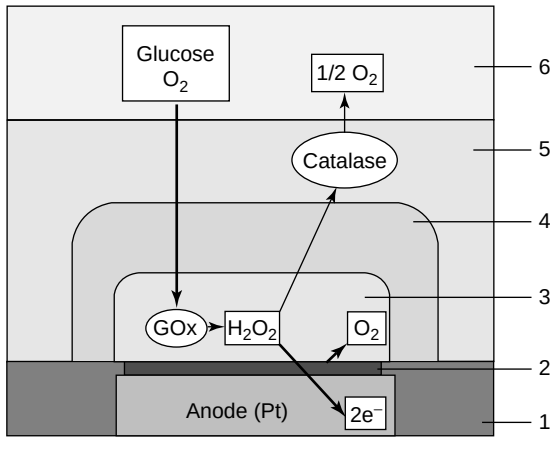

Figure 9 Schematic cross-section of the multienzyme biosensor. 1, Insulation layer; 2, semipermeable membrane; 3, enzyme membrane; 4, spacer membrane; 5, catalase containing hydrogel; 6, analyte.

Figure 10 Flexible in vivo sensor for glucose, lactate, and pH measurements.

use. They can be produced by means of microelectronic technology. Such multienzyme sensors were accomplished by immobilizing different enzymes into stacked membranes structured by photolithography.

This is an important step towards commercialization, and an additional feature is the possibility of integrating additional electrochemical sensors for measuring O_2, CO_2 and pH on one substrate. Such an integrated lab-on-chip seems to be a realistic vision and can revolutionize the point of care testing. Additionally the technology of photopatterned multienzyme sensors can be used for the creation of in vivo devices on a flexible polyimide strip (Figure 10).[92]

5 MICROANALYTIC SYSTEMS

The technology for integrating different biosensors on a chip is the basis for the lab-on-chip, which requires additional liquid handling and optionally active liquid treatment. Different approaches for a microanalytic system (μTAS) are published and are now an emerging field of research.[93,94]

For several reasons, most applications of chemo- and biosensors require flow-through devices. They allow analysis of different samples from multiple sources, to get rid of flow sensitivities, to avoid sterilization and biocompatibility problems, to enable recalibration, or to design complete bioanalytical assays. A classic approach for a flow-through device with sensors is the clinical-chemical analyzer using photometric techniques and a bulky fluid handling system.[95] In order to reduce system costs and allow miniaturization the classic optical approach is successively replaced by solid-state devices. In future, integration of the liquid handling environment with sensor elements will reduce assembling effort, minimize internal analyte volumes, and allow sample pretreatment and reagent addition.

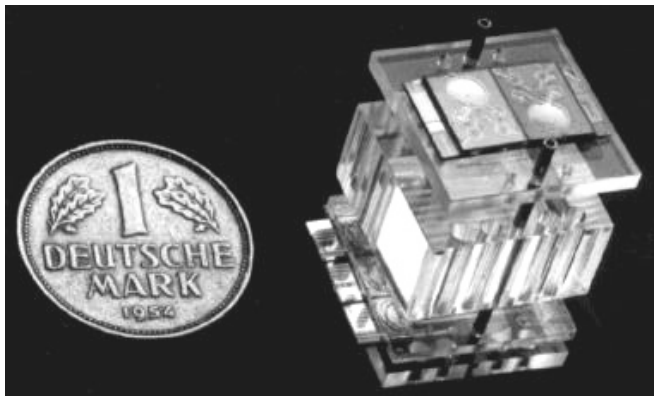

Figure 11 Electrochemical microsystem.

5.1 Electrolyte Sensing Systems

ISFET-based systems were developed at different laboratories using a planar or three-dimensional set-up.[96–98] Such systems equipped with micropumps and valves are able to detect ions and other important chemical parameters (Figure 11). Microarrays for measuring pressure, temperature, pH, O_2 and CO_2 were realized using silicon technology and a hybrid set-up of the different components.[99,100]

Hybrid microflow injection analysis systems produced in silicon technology using oxygen microelectrodes and microcavities were presented, also optic-based microanalytical systems were achieved.[101,102]

5.2 Sensing Systems for Metabolic Parameters

Multiple analyte biosensor arrays can also be realized using thin film and silicon technology. The so-called containment technology was applied to immobilize enzymes in three-dimensional cavities formed in silicon wafers to get fully process compatible biosensor devices.[103] Such systems were developed and used for microdialysis sampling in human serum.[104,105]

Another approach combines the described biosensor microsystems (section 3.4.3) with a low cost microfluidics by utilizing printed circuit board (PCB) technology which is capable of producing microsystems.[106,107] The PCB for the assembly of the thin film biosensor array comprises conducting pads for the thin film device and the plug connection to the potentiostat. Additionally a $5 \times 1\ mm^2$ gold counter electrode, drilled holes for liquid inlet and outlet, and a photopatterned spacer made from the dry film resist used for insulation are integrated onto the PCB (Figures 12 and 13).

To meet the demands for multianalyte monitoring in biotechnology where closed loop control of metabolites is expected to increase fermentation reliability and yield a glucose–lactate–glutamine–glutamate device

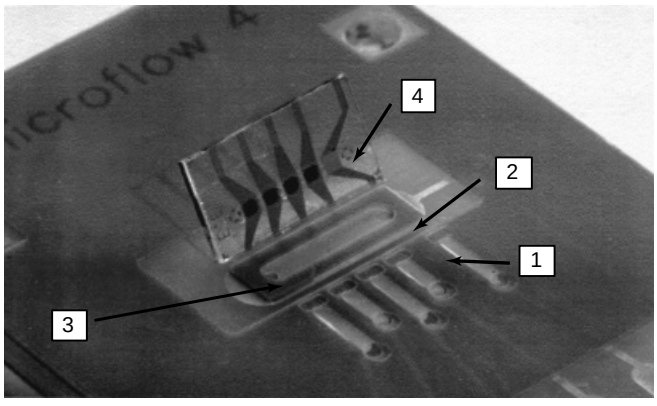

Figure 12 Photomicrograph illustrating the assembling of the thin film sensor array with the PCB. 1, Contact structure; 2, spacer; 3, counter electrode; 4, sensor array.

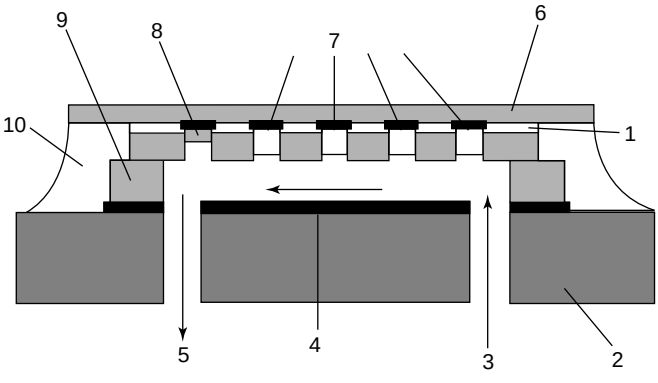

Figure 13 Schematic cross-section of a microflow system. 1, SiN_x insulation layer; 2, PCB; 3, inlet; 4, counter electrode; 5, outlet; 6, glass wafer; 7, individual enzyme sensors; 8, reference electrode; 9, spacer; 10, glue.

monitoring the concentrations of these four substances simultaneously, without any reagent, and with no need for sample pretreatment was realized by Moser et al.[108] Figure 14 shows the calibration graphs of such a device. The small apparatus effort needed for monitoring is very attractive.

Ex vivo blood monitoring experiments with human volunteers were performed by placing a device comprising two glucose and two lactate sensors into the sampling line of a double lumen catheter. Venous blood was continuously withdrawn and in line heparinized with a dilution of less than 5%.

These experiments are described in detail in Jobst et al.[109] Before measurement the glucose/lactate device is one point calibrated with a protein-free buffer solution. The measured sensitivities are applied to the sensor currents of the experiments for calculation of glucose and lactate levels and show the time course of glucose and lactate levels of a combined intravenous and oral

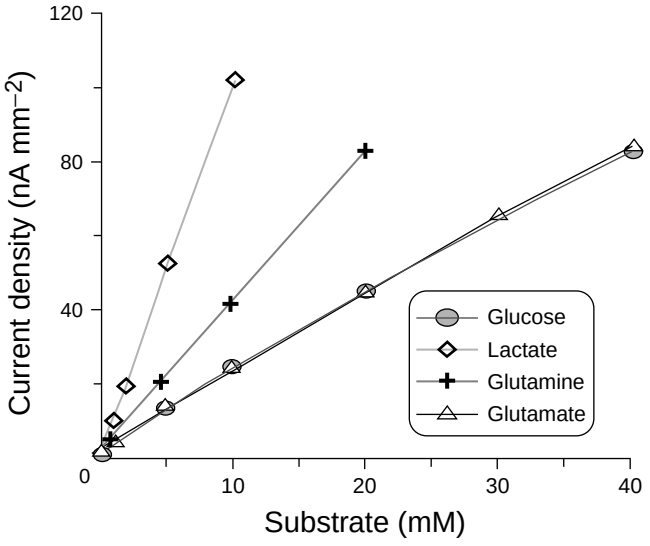

Figure 14 Calibration graphs.

glucose tolerance test together with the reference values (Figures 15 and 16).

The good correspondence between sensor and laboratory results proves the effectiveness of the precalibration procedure and sensor stability. The results suggest that these devices can be used for short-term glucose/lactate monitoring, e.g. in the intensive care unit, the operating theater, in sports medicine and rehabilitation, and in diabetology.

An advanced lab-on-chip equipped with biosensors and microfluidics can also be performed. On-line buffering and reagent addition can be also done integrating a

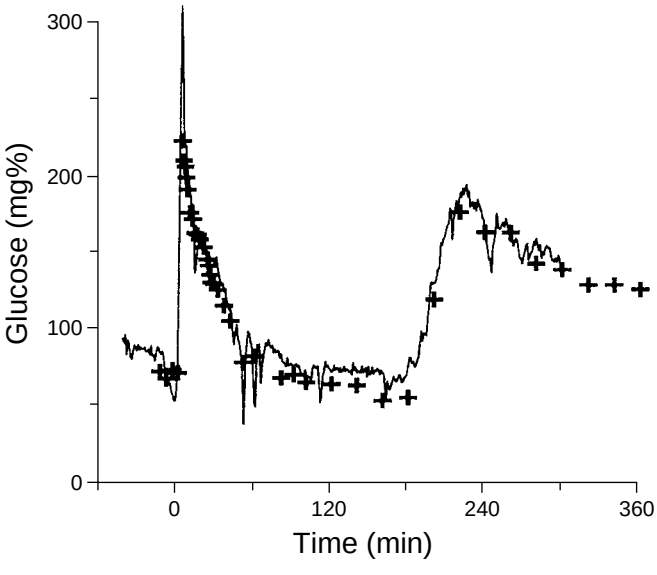

Figure 15 Glucose sensor reading (solid line) and reference values (crosses).

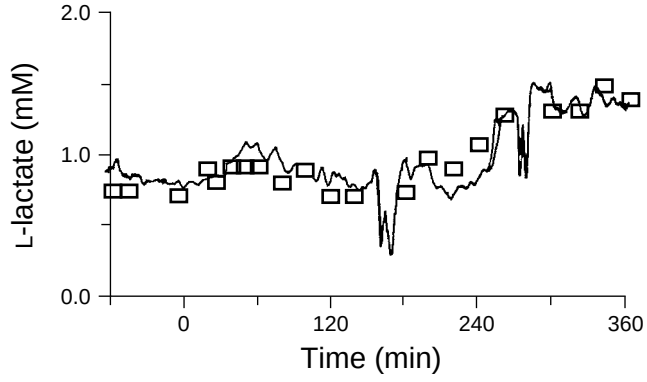

Figure 16 L-lactate sensor reading (solid line) and reference values (boxes).

mixing coil of $170 \times 0.6 \times 0.1\,mm^3$ (=10.2 μL) onto the PCB together with a further inlet hole. The entire processing was done with conventional PCB technology equipment. Assembling the thin film sensor array with this PCB similar to that shown in Figures 12 and 13 gives the analytical microflow system. The dimension of the flow chamber is $5 \times 1 \times 0.3\,mm^3$ (=1.5 μL).

A sensor array comprising two lactate and two lactose sensors is assembled with a PCB comprising a mixing coil (Figure 17) for yogurt fermentation monitoring.

The sample flow of $6\,\mu L\,min^{-1}$ is on-line buffered and diluted with $100\,\mu L\,min^{-1}$ of a phosphate-buffered saline pH 7.4 carrier solution. During yogurt fermentation, the pH of the sample changes drastically over several pH units. Reliable measurements with biosensors are not possible without sample pretreatment due to the pH sensitivity of the enzymes used. However, with an on-line buffering strategy, monitoring with biosensors can be done even in strongly changing matrices. Such strong changes in matrix composition occur during most bacterial and fungal fermentation.[110]

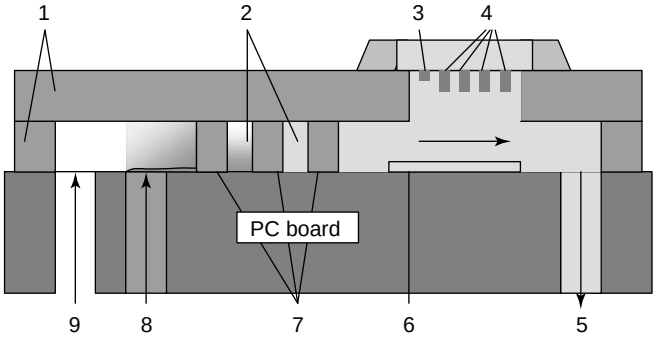

Figure 17 Schematic cross-section of the flow through device with integrated static mixer. 1, Dry film resist layers; 2, mixing coil; 3, reference electrode; 4, individual enzyme sensors; 5, outlet; 6, gold counter electrode.

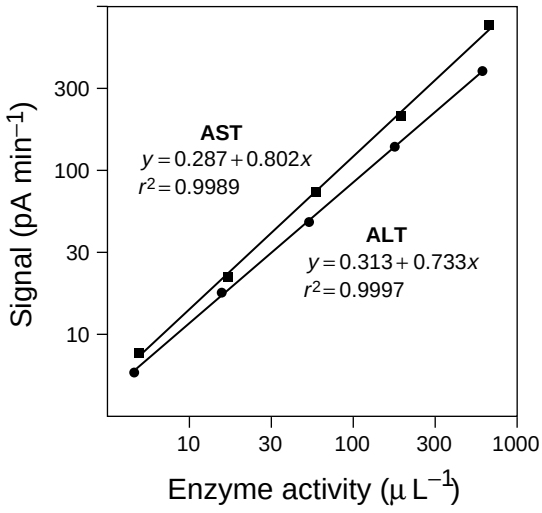

Figure 18 Calibration graph for AST and ALT assay.

Another application of such a lab-on-chip device is the reagent addition, demonstrated with an assay for the glutamate producing enzymes glutamic oxaloacetic transaminase (AST) and glutamic pyruvic transaminase (ALT).[111] A thin film biosensor array comprising two glutamate and two blank sensors is assembled with a PCB comprising a mixing coil (Figure 17). The sample containing the enzyme to be assayed and a reagent with the respective enzyme substrates are mixed in the mixing coil. After flow stop the increase in glutamate concentration is monitored by the difference signal from the glutamate and the blank sensors. The calibration graph of such a device for a AST and a ALT assay is shown in Figure 18.

Compared to conventional spectrophotometric assays for these enzymes, this assay offers the advantages of needing no enzymes to be added to the reagent and electrical readout that allows running this assay even in whole blood. Such an assay scheme with a bioanalytic microsystem can be applied to a variety of diagnostic important enzymes. Also attractive is the possibility to directly combine metabolite analysis, e.g. lactate and glucose, with enzyme assays in one device, giving a more versatile lab-on-chip.

The set-up of the whole device is performed using the HYMOS (HYbrid MOdular microSystem) approach which was reported in Urban.[112] This device consists of two components made by means of well established mass production technologies (thin film and PCB technology) and assembling of the parts is compatible to intensive care packaging techniques, so cost effective mass fabrication of this device seems realistic.

5.3 Cell-based Systems

Microphysiometry is the measurement of the metabolic activity of living cells in a miniaturized environment.

Considerable interest in such devices from the industry arises from their applicability for drug screening. Other applications are toxicity monitoring and cytotoxin response measurement of cancer cells.[113] Future applications could be strain prescreening in biotechnology, rapid sterility tests, and HTS in combinatorial chemistry.

According to the different ways in which cells alter their environment, e.g. produce heat, acidification, oxygen consumption or metabolite production, various sensors have to be used to measure these complex metabolic activities. To get a deeper insight into the cell metabolism, different sensors were integrated into one small test chamber.[113] Using microsystem technology even an entire fermenter can be realized.[114]

For high throughput applications, e.g. drug screening, massive parallelization and therefore miniaturization is required. Different companies are now entering the market with microphysiometer arrays based on light-addressable potentiometric sensors for pH sensing.[115] In addition, an electrode arrangement for amperometry can be realized in subnanoliter containments, and even single cell microphysiometry seems feasible with this method.[116] In an optical low cost approach, an ultra-thin pH responsive membrane is placed on replicated chirped grating couplers.[117,118] A resolution of 10^{-4} pH units was achieved but miniaturization of such a type of optical device is limited.

Probing oxygen consumption or metabolite production gives additional or complementary information about the cell metabolism.[119] In addition, the metabolic activity of single cells can be measured by means of microsystem technology.[120]

The challenge of establishing hundreds or thousands of electrical contacts if such amperometrically transduced microphysiometers undergo massive parallelization can be satisfied by multiplexing the individual sensors by complementary metal oxide semiconductor components integrated on a silicon wafer.[121]

A localized cocultivation of cells on a chip to get an artificial liver or diagnostic devices is also possible.[122] Some even more exciting developments can be expected from this combination of biosciences and microsystem technology in the near future.

5.4 Affinity-based Systems

Among biosensor microsystems using chemosensors and enzymes, affinity-based arrays as immunoassays and DNA assays are emerging fields of biosensor microsystems. Microchip arrays for parallel binding and detection of DNA or proteins for drug discovery and different clinical microsystems are under development.[123]

The most advanced system uses a DNA chip with up to 65 000 different exactly specified 20-mer polynucleotides immobilized on a silicon chip and an epifluoroscopic read out of fluorescence labeled prober nucleotides whereas point defects in DNA can be easily detected.[124]

Other strategies immobilize DNA selectively on complementary metal oxide semiconductor biochips using electrochemical methods. Hybridization can be controlled electrostatically, the readout is performed using fluorophore labeled target DNA and a CCD camera.[125] For arrays up to several hundred DNA spots this seems a feasible and the least expensive method for DNA screening.

6 CONCLUSIONS AND OUTLOOK

The enthusiasm and the high expectations in biosensor microsystems have not been fulfilled till now, and technological breakthroughs and further work are still needed to ultimately fulfil them. Up to now only the glucose sensor has been widely marketed and has conquered the market, otherwise only niche markets could be covered by microelectronic biosensor devices for decentralized clinical analyzers.

Increasing production volumes resulting in lower prices of devices will push the whole market and new technologies will open additional markets. The expectations of using small samples as well as small reagent volumes, minimization of time expenditure of skilled clinical staff, minimization of calibration fluids consumption and waste are still key advantages of such a microtechnology. Additionally performing highly parallel assays to process thousands of samples immediately will also push rapid drug screening.

Realizing highly reliable devices produced by the appropriate production technology means a breakthrough of microelectronic biosensor systems is still expected. For market driven applications such as HTS, gene-based analysis, and decentralized laboratories a distinct solution can still be expected in the near future.

ABBREVIATIONS AND ACRONYMS

ALT	Glutamic Pyruvic Transaminase
AST	Glutamic Oxaloacetic Transaminase
BAW	Bulk Acoustic Wave
CCD	Charge Coupled Dence
ENFET	Enzyme-containing Field Effect Transistor
GOx	Glucose Oxidase
HTS	High Throughout Screening
ISFET	Ion Selective Field Effect Transistor
LW	Lamb Waves
NMR	Nuclear Magnetic Resonance

PCB	Printed Circuit Board
SAW	Surface Acoustic Wave
SPR	Surface Plasmon Resonance

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Electroanalysis and Biosensors in Clinical Chemistry • Glucose Measurement • Micro Total Analytical Systems in Clinical Chemistry • Point-of-care Testing

REFERENCES

1. Frost & Sullivan, Report #A2139, Biosensor Markets, 1989.
2. A.P.F. Turner, I. Karube, G. Wilson, *Biosensors Fundamentals and Applications*, Oxford University Press, Oxford, 1987.
3. E.A.H. Hall, *Biosensors*, UK Redwood Press (Open University Press Biotechnology Series), Melsham, 1990.
4. A. Riehle, S. Grün, M. Diesmann, A. Aertsen, 'Spike Synchronization and Rate Modulation Differentially Involved in Motor Cortical Function', *Science*, **278**, 1950–1953 (1997).
5. F. Scheller, F. Schubert, *Biosensoren, Beiträge zur Forschungstechnologie*, 18, Akademie Verlag, Berlin, 1989.
6. E. Kress-Rogers (ed.), *Handbook of Biosensors and Electronic Noses: Medicine, Food, and the Environment*, CRC Press, Boca Raton, 1997.
7. A. Grisel, 'Microelectronic Devices', in *Handbook of Biosensors and Electronic Noses: Medicine, Food, and the Environment*, ed. E. Kress-Rogers, CRC Press, Boca Raton, 137–147, 1997.
8. T. Michaelis, K. Merboldt, W. Hänicke, M. Gyngell, H. Bruhn, J. Frahm, 'On the Identification of Cerebral Metabolites in Localized ^{1}H NMR Spectra of Human Brain in vivo', *NMR Biomed.*, **4**, 90 (1991).
9. J. Astrup, L. Symon, N. Branston, Na. Lassen, 'Cortical Evoked Potential and Extracellular K^{+} and H^{+} at Critical Levels of Brain Ischemia', *Stroke*, **8**, 51–57 (1977).
10. R.H.C. Strang, H.S. Bachelard, 'Rates of Cerebral Glucose Utilization in Rats Anaesthetized with Phenobarbitone', *J. Neurochem.*, **20**, 987–996 (1973).
11. Anonymous, *MedPro Month*, 44 (1995).
12. D. Huckle, *Changes in Diagnostic Practice*, Financial Times Managment Report, 1996.
13. R. Anderson, G. McGall, R. Lipshutz, 'Polynucleotide Arrays for Genetic Sequence Analysis', *Topics in Current Chemistry*, eds. A. Manz, H. Beckerl, Springer, Berlin, 117–129, Vol. 194, 1998.
14. P. Telleman, U.D. Larsen, J. Philip, G. Blankenstein, A. Wolff, 'Cell Sorting in Microfluidic Systems', in *Proceedings Micro Total Analysis Systems '98*, ed. J.Harrison, Banff, Canada, 39–44, 1998.
15. J.C. Pickup, 'In vivo Glucose Monitoring: Sense and Sensorbility', *Diabetes Care*, **16**, 535–539 (1993).
16. O.S. Wolfbeis (ed.), *Fiber Optical Chemical Sensors and Biosensors*, CRC Press, Boca Raton, Vols 1 and 2, 1991.
17. B. Weigl, A. Holobar, W. Trettnak, I. Klimant, H.Kraus, P. O'Leary, O.S. Wolfbeis, 'Optical Triple sensor for Measuring pH, Oxygen and Carbon Dioxide', *J. Biotechnol.*, **32**, 127–138 (1994).
18. R. Karlsson, A. Michaelson, L. Mattson, 'Kinetics Analysis of Monoclonal Antibody–Antigen Interactions with a new Biosensor Based Analytical System', *J. Immunol. Meth.*, **145**, 229–240 (1991).
19. Ph.M. Nellen, W. Lukosz, 'Integrated Optical Input Grating Couplers as Direct Affinity Sensor', *Biosens. Bioelectron.*, **8**, 129–147 (1993).
20. Ch. Stamm, W. Lukosz, 'Integrated Optical Difference Interferometer as Biochemical Sensor', *Sens. Actuators, B*, **18–19**, 183–187 (1994).
21. R.E. Kunz, G. Duveneck, M. Ehrat, *Proc. SPIE*, 2331 (1994).
22. Anonymous, Pharmacia Technical Digest, 'Biacore 3000', 1999.
23. M. Winklmair, A. Schuetz, M. Weller, R. Niessner, 'Immunochemical Array for the Identification of Cross-reacting Analytes', *Fresenius' J. Anal. Chem.*, **363**, 731–737 (1999).
24. J. Roulet, K. Fluri, E. Verpoorte, R. Völkl, H. Herzig, N. de Rooij, R. Dändliker, 'Micro-optical Systems for Fluorescence Detection in μTAS Applications', in *Proceedings Micro Total Analysis Systems '98*, ed. J. Harrison, Banff, Canada, 287–290, 1998.
25. A. Bruno, E. Baer, R. Völkl, C. Effenhauser, 'Microoptical Fluorescence Detection for Chip-based Multiplexed Analysis Systems', in *Proceedings Micro Total Analysis Systems '98*, ed. J. Harrison, Banff, Canada, 281–285, 1998.
26. L. Roß, 'Integrated Optical Components in Substrate Glasses', *Glastechn. Ber.*, **62**(8), 285 (1989).
27. A. Amico, C. Di Natale, E. Verone, 'Acoustic Devices', in *Handbook of Biosensors and Electronic Noses: Medicine, Food, and the Environment*, ed. E. Kress-Rogers, CRC Press, Boca Raton, 197–223, 1997.
28. N. Barie, M. Rapp, H. Sigrist, 'A New and Flexible Covalent Binding Concept for Surface Acoustic Wave Biosensors Using Dextrans as Intermediate Layer', in *Proceedings of Transducer 99*, Sendai, Japan, 1708–1711, 1999.

29. C.H. Spink, 'Analytical Calorimetry in Biochemical and Clinical Applications', *CRC Crit. Rev. Anal. Chem.*, (1980).
30. K. Mosbach, B. Danielsson, 'An Enzyme Thermistor', *Biochim. Biophys. Acta*, **364**, 140 (1974).
31. B. Danielsson, K. Mosbach, 'Theory and Application of Calorimetric Sensors', *J. Neurochem.*, **20**, 987 (1973).
32. S. Kröger, B. Danielsson, 'Calorimetric Biosensors', in *Handbook of Biosensors and Electronic Noses: Medicine, Food, and the Environment*, ed. E. Kress-Rogers, CRC Press, Boca Raton, 279–298, 1997.
33. G. Urban, H. Kamper, A. Jachimowicz, F. Kohl, H. Kuttner, F. Olcaytug, P. Goiser, F. Pittner, T. Schalkhammer, E. Mann-Buxbaum, 'The Construction of Microcalorimetric Biosensors by Use of High Resolution Thin-film Thermistors', *Biosens. Bioelectron.*, **6**, 275 (1991).
34. P. Bataillard, E. Steffgen, S. Haemmerli, A. Manz, H. Widmer, 'An Integrated Silicon Thermopile as Biosensor for the Thermal Monitoring of Glucose, Urea and Penicillin', *Biosen. Bioelectron.*, **8**, 89 (1993).
35. H. Muramatsu, J.M. Dicks, I. Karube, 'Integrated-circuit Bio-calorimetric Sensor for Glucose', *Anal. Chim. Acta*, **197**, 347 (1987).
36. B. Xie, U. Hedberg, M. Mecklenburg, B. Danielsson, 'Fast Determination of Whole Blood Glucose with a Calorimetric Micro-biosensor', *Sens. Actuators, B*, **141**, 15–16 (1993).
37. M. Muehlbauer, E. Guilbeau, B. Towe, 'Applications and Stability of a Thermoelectric Enzyme Sensor', *Sens. Actuators, B*, **2**, 223 (1990).
38. B. Xie, B. Danielsson, F. Winquist, 'Miniaturized Thermal Biosensors', *Sens. Actuators, B*, **15–16**, 443 (1993).
39. D.C. Cullen, R.S. Sethi, C.R. Lowe, 'Multi-analyte Miniature Conductance Biosensor', *Anal. Chim. Acta*, **231**, 33 (1990).
40. P. Jacobs, J. Suls, W. Sansen, 'Performance of Planar Differential-conductivity Sensor for Urea', *Sens. Actuators, B*, **20**, 193 (1994).
41. L.D. Watson, P. Maynard, D.C. Cullen, R.S. Sethi, J. Brettle, C.R.A. Lowe, 'A Microelectronic Conductometric Biosensor', *Biosensors*, **3**, 101 (1987/88).
42. P. Jacobs, 'From Conductometric Sensors to Molecular Devices', in *Proceedings of Transducer 99*, Sendai, Japan, 272–275, 1999.
43. D.J.G. Ives, G.J. Janz (eds.), *Reference Electrodes-Theory and Practice*, Academic Press, New York, 1961.
44. A. van den Berg, A. Grisel, H.H. van den Vlekkert, N.F. De Rooij, 'A Micro-volume Open Liquid-junction Reference Electrode for pH-ISFETs', *Sens. Actuators, B*, **1**, 425–432 (1990).
45. M. Cremer, 'Über die Ursachen der elektromotorischen Eigenschaften der Gewebe, zugleich ein Beitrag zur Lehre von den polyphasischen Elektrolytketten', *Z. Biol.*, **47**, 468–472 (1906).
46. D. Ammann, W.E. Morf, P. Anker, P.C. Meier, E. Pretsch, W. Simon, 'Neutral Carrier Based Ion-selective Electrodes', *Ion-selective Electrode Rev.*, **5**, 3 (1983).
47. P. Bergveld, 'Development Operation, and Application of the Ion-sensitive Field-effect Transistor as a Tool for Electrophysiology', *IEEE Trans. Biomed. Engin.*, **BME-19**(5), 342 (1972).
48. J.M. Thompson, 'Performance Evaluation of ISFETs and other ISE Sensors for Whole Blood Ion Assay', *Med. Biol. Eng. Comput.*, **28**, B29–B33 (1990).
49. A. Sibbald, A.K. Covington, R.F. Carter, 'Simultaneous On-line Measurement of Blood K^+, Ca^{2+}, Na^+, and pH with a Four-function chemFET Integrated-circuit Sensor', *Clin. Chem.*, **30**(1), 135 (1984).
50. I. Lundström, A. van den Berg, B. van der Schoot, H. van den Vlekkert, 'Field Effect Chemical Sensors', in *Sensors, A Comprehensive Survey*, eds. W. Göpel, J. Hesse, J. Zemel, VCH Publishers, Weinheim, Vol. 2, 1991.
51. R. Hintsche, I. Dransfeld, F. Scheller, M. Pham, W. Hoffmann, J. Hueller, W. Moritz, 'Integrated Differential Enzyme Sensor Using Hydrogen and Fluoride Ion Sensitive Multigate FETs', *Biosens. Bioelectron.*, **5**, 327 (1990).
52. F. Winquist, B. Danielsson, in *Semiconducting Field Effect Devices in Biosensors–A Practical Approach*, ed. A.E.G. Cass, IRL Press, Oxford, 171–175, 1990.
53. P. Bergveld, 'A Critical Evaluation of Direct Electrical Protein Detection Methods', *Biosens. Bioelectron.*, **6**, 55 (1991).
54. T. Matsuo, M. Esashi, 'Methods of ISFET Fabrication', *Sens. Actuators*, **1**, 77 (1981).
55. Y. Ito, 'Long Term Drift Mechanism of Ta_2O_5 Gate pH-ISFETS', in *Proceedings of Transducer 99*, Sendai, Japan, 648–651, 1999.
56. H. Abe, M. Esashi, M. Matsuo, 'ISFET's Using Inorganic Gate Thin Films', *IEEE Trans. Electron. Devices*, **ED-26**, 1939–1944 (1979).
57. R.E.G. van Hal, P. Bergveld, J.F.J. Engbersen, D.N. Reinhoudt, 'Characterization and Testing of Polymer–Oxide Adhesion to Improve the Packaging Reliability of ISFETs', *Sens. Actuators, B*, **23**, 17 (1995).
58. Y. Hanazato, K. Inatomi, M. Nakako, S. Shiono, M. Maeda, 'Glucose-sensitive Field-effect Transistor with a Membrane Containing Co-immobilized Gluconolactonase and Glucose Oxidase', *Anal. Chim. Acta*, **212**, 49 (1988).
59. J. Anzai, S. Tezuka, T. Osa, H. Nakajima, T. Matsuo, 'Urea Sensor Based on an Ion-sensitive Field Effect Transistor IV. Determination of Urea in Human Blood', *Chem. Pharm. Bull.*, **35**(2), 693 (1987).
60. J. Kimura, Y. Kawana, T. Kuriyama, 'An Immobilized Enzyme Membrane Fabrication Method Using an Ink Jet Nozzle', *Biosensors*, **4**, 41 (1988).
61. S. Shiono, Y. Hanazato, M. Nakako, 'Urea and Glucose Sensors Based on Ion Sensitivity Field Effect Transistor

with Photolithographically Patterned Enzyme Membrane', *Anal. Sci.*, **2**, 517–521 (1986).
62. Y. Hanazato, M. Nakako, S. Satorus, M. Mitsuo, 'Integrated Multi-biosensors Based on Ion-sensitive Field-effect Transistor Using Photolithographic Techniques', *IEEE Trans. Electron. Dev.*, **36**(7), 1303 (1989).
63. A. Saito, N. Ito, J. Kumura, T. Kuriyama, 'An ISFET Glucose Sensor with a Silicone Rubber Membrane for Undiluted Serum Monitoring', *Sens. Actuators, B*, **20**, 125 (1994).
64. K. Erickson, P. Wilding, 'Evaluation of a Novel Point-of-care System, the i-STAT Portable Clinical Analyzer', *Clin. Chem.*, **39**(2), 283 (1993).
65. E. Jacobs, E. Vadasdi, L. Sarkozi, N. Colman, 'Analytical Evaluation of i-STAT Portable Clinical Analyzer and Use by Nonlaboratory Health-care Professionals', *Clin. Chem.*, **39**(6), 1069 (1993).
66. M. Koudelka, S. Gernet, N.F. DeRooij, 'Planar Amperometric Enzyme-based Glucose Microelectrode', *Sens. Actuators*, **18**, 157 (1989).
67. K. Yokoyama, K. Sode, E. Tamiya, I. Karube, 'Integrated Biosensor for Glucose and Galactose', *Anal. Chim. Acta*, **218**, 137 (1989).
68. T. Murakami, S. Nakamoto, J. Kimura, T. Kuriyama, I. Karube, 'A Micro Planar Amperometric Glucose Sensor Using an Isfet as a Reference Electrode', *Anal. Lett.*, **19**(19&20), 1973 (1986).
69. E. Tamiya, I. Karube, S. Hattori, M. Suzuki, K. Yokoyama, 'Micro Glucose Sensor Using Electron Mediators Immobilized on a Polypyrrole-modified Electrode', *Sens. Actuators*, **18**, 297 (1989).
70. I. Takatsu, T. Moriizumi, 'Solid State Biosensors Using Thin-film Electrodes', *Sens. Actuators*, **11**, 309 (1987).
71. S. Nakamoto, N. Ito, T. Kuriyama, J. Kimura, 'A Lift-off Method for Patterning Enzyme-immobilized Membranes in Multi-biosensors', *Sens. Actuators*, **13**, 165 (1988).
72. C. Pifl, A. Jachimowicz, G. Urban, F. Kohl, P. Goiser, J. Theiner, G. Nauer, 'A New Type of Thin-layer Microelectrode for in vivo Voltammetry: Electro-oxidation of Dopamine, DOPAC and Ascorbic Acid', *Sens. Actuators*, **B1**, 468 (1990).
73. D. Pfeiffer, F. Scheller, K. Setz, 'Amperometric Enzyme Electrodes for Lactate and Glucose Determinations in Highly Diluted and Undiluted Media', *Anal. Chim. Acta*, **281**(3), 489 (1993).
74. G. Kost, D. Wiese, T. Bowen, 'New Whole Blood Methods and Instruments: Glucose Measurement and Test Menus for Critical Care', *J. Int. Fed. Clin. Chem.*, **3**(4), 160–165 (1991).
75. J.M. Elbicki, W.G. Stephen, 'Ultrafiltration of Human Serum to Determine the Size of Species that Poison Voltammetric Electrodes', *Biosensors*, **4**, 251 (1989).
76. S. Reddy, P. Vadgama, 'Membranes to Improve Amperometric Sensor Characteristics', in *Handbook of Biosensors and Electronic Noses Medicine, Food and the Environment*, ed. E. Kress-Rogers, CRC Press, Boca Raton, 111–136, 1997.
77. G. Urban, G. Jobst, F. Keplinger, E. Aschauer, O. Tilado, R. Fasching, F. Kohl, 'Miniaturized Multi-enzyme Biosensors Integrated with pH-sensors on Flexible Polymer Carriers for in vivo Applications', *Biosens. Bioelectron.*, **7**, 733 (1992).
78. D. Matthews, R. Holman, Brown, J. Sreemson, A. Watson, S. Hughes, 'Pen Sized Digital 30-second Blood Glucosemeter', *Lancet*, **i**, 778 (1987).
79. A.E.G. Cass, G. Davis, G.D. Francis, H.A.O. Hill, W.J. Aston, I.J. Higgins, E.V. Plotkin, L.D.L. Scott, A.P.F. Turner, 'Ferrocene-mediated Enzyme Electrode for Amperometric Determination of Glucose', *Anal. Chem.*, **56**, 667 (1984).
80. K.H. Pollmann, M. Gerber, K.M. Kost, M.L. Ochs, P.D. Walljng, J. Bateson, L.S. Kuhn, C.A. Han, 'Enzyme Electrode System', Boehringer Mannheim, *US Patent 5 288 636*, 1994.
81. T. Yoshioka, 'Biosensor Utilizing Enzyme and a Method for Producing Them', *US Patent 5 192 415*, 1993.
82. A. Aoki, A. Heller, 'Electron Diffusion Coefficients in Hydrogels Formed of Cross-linked Redox Polymers', *J. Phys. Chem.*, **97**, 11 014–11 019 (1993).
83. C. Yarnitzky, A.M. Caruana. D.W. Schmidtke, A. Heller, 'Linear Dependence of the Potential of "Wired" Glucose Oxidase Electrodes on the Concentration of Glucose', *J. Phys. Chem. B*, **102**, 10057–10061 (1998).
84. T. Ohara, R. Rajagopalan, A. Heller, '"Wired" Enzyme Electrodes for Amperometric Determination of Glucose or Lactate in the Presence of Interfering Substances', *Anal. Chem.*, **66**, 2451–2457 (1994).
85. R. Geise, J.M. Adams, N.J. Barone, A.M. Yacynych, 'Electropolymerized Films to Prevent Interferences and Electrode Fouling in Biosensors', *Biosens. Bioelectron.*, **6**, 151 (1991).
86. G. Urban, G. Jobst, E. Aschauer, O. Tilado, P. Svasek, M. Varahram, 'Performance of Integrated Glucose and Lactate Thin Film Microbiosensors for Clinical Analyzers', *Sens. Actuators, B*, **19**(1–3), 592 (1994).
87. J.D. Newman, A.P.F. Turner, G. Marazza, 'Ink-jet Printing for the Fabrication of Amperometric Glucose Biosensors', *Anal. Chim. Acta*, **262**, 13–17 (1992).
88. U. Fischer, K. Rebrin, T.v. Woedtke, P. Abel, 'Clinical Usefulness of the Glucose Concentration in the Subcutaneous Tissue – Properties and Pitfalls of Electrochemical Biosensors', *Horm. Metab. Res.*, **26**, 515 (1994).
89. E. Bobbioni-Harsch, F. Rohner-Jeanrenaud, M. Koudelka, N. deRooij, B. Jeanrenaud, 'Lifespan of Subcutaneous Glucose Sensors and their Performances during Dynamic Glycaemia Changes in Rats', *J. Biomed. Eng.*, **15**, 457 (1993).
90. J. Mastrototaro, K. Johnson, R. Morff, D. Lipson, C. Andrew, D. Allen, 'An Electroenzymatic Glucose Sensor Fabricated on a Flexible Substrate', *Sens. Actuators, B*, **5**(1–4), 139 (1991).

91. R. Jeyanthi, P.R. Rao, 'In vivo Biocompatibility of Collagen-poly(hydroxyethyl methacrylate) Hydrogels', *Biomaterials*, **11**, 238 (1990).
92. G. Urban, G. Jobst, F. Keplinger, E. Aschauer, R. Fasching, P. Svasek, 'Miniaturized Integrated Biosensors', *Technol. Health Care*, **1**, 215 (1994).
93. A. Manz, N. Graber, H.M. Widme, 'Miniaturized Total Chemical Analysis System: a Novel Concept for Chemical Sensing', *Sens. Actuators, B*, **1**, 244 (1990).
94. A. van den Berg, P. Bergveld (eds.), *Micrototal Analysis Systems*, Kluwer Academic Publishers, Norwell, 1995.
95. Ch. Sachs, P. Finetti, J. Mollard, C. Ritter, 'The AVL Omni Analyzer: New Concepts in Blood Gas/Electrolyte Analysis', in *Proceedings 14th International Symposium on Electrolyte/blood Gas*, Hawai, 428–433, 1996.
96. W. Hoffmann, R. Rapp, 'Integrated Microanalytic System with Electrochemical Detection', *Sens. Actuators, B*, **34**, 471–475 (1996).
97. B. van der Schoot, S. Jeanneret, A. van den Berg, N. de Rooij, 'Modular Set Up for a Miniaturized Chemical Analysis System', *Sens. Actuators, B*, **15–16**, 211–213 (1993).
98. S. Drost, W. Wörmann, B. Ross, G. Chemnitius, M. Rospert, W. Konz, F. Hartmann Frimmel, W. Schuhmann, R. Ferretti, L. Meixner, 'Microanalytic Systems for Environmental Control: Aims, Structure and Integration of μTAS-modules', in *Proceedings of the 2nd International Symposium on Miniaturized Total Analysis Systems*, Basel, 199–202, 1996.
99. F. Steenkiste, H. Grünkorn, L. Claesen, K. Baert, L. Hermans, D. Debruyker, M. DeCooman, V. Spiering, A. van den Berg, B. van der Schoot, P. Arquint, R. Born, K. Schuhmann, 'A Microsensorarray for Biomedical Sensing', in *Proceedings of the 2nd International Symposium on Miniaturized Total Analysis Systems*, Basel, 138–140, 1996.
100. Ph. Arquint, A. van den Berg, B.H. van der Schoot, H. Bühler, W.E. Morf, L. Dürselen, N.F. de Rooij, 'Integrated Blood Gas Sensors for pO_2, pCO_2 and pH', *Sens. Actuators, B*, **13–14**, 340–344 (1993).
101. H. Hinkers, N. Conrath, N. Czupor, H. Frebel, S. Hüwel, K. Köckemann, D. Trau, M. Wittkampf, G. Chemnitius, L. Haalck, M. Meusel, K. Cammann, M. Knoll, F. Spener, M. Rospert, R. Kakerow, O. Köster, T. Lerch, W. Mokwa, P. Woias, M. Richter, T. Abel, L. Meixner, 'Results and Development of Sensors and μTAS-modules', in *Proceedings of the 2nd International Symposium on Miniaturized Total Analysis Systems*, Basel, 110–112, 1996.
102. A. Van den Berg, T.S.J. Lammerink, 'Micro Total Analysis Systems: Microfluidic Aspects, Integration Concept and Applications', *Topics in Current Chemistry*, eds. A. Manz, H. Becker, Springer, 21–50, Vol. 194, 1998.
103. M. Knoll, *German Patent Application DE4 115 414 A1*.
104. R. Steinkuhl, C. Sundermeier, H. Hinkers, C. Dumschat, K. Cammann, M. Knoll, 'Microdialysis System for Continuous Glucose Monitoring', *Sens. Actuators, B*, **33**, 19–24 (1996).
105. S. Böhm, W. Olthuis, P. Bergveld, 'An Electrochemically Actuated Micropump for use in a 'Push–Pull' Microdialysis Based in-vivo Monitoring System', in *Proceedings of Transducer 99*, 880–881, 1999.
106. G. Urban, G. Jobst, P. Svasek, M. Varahram, I. Moser, E. Aschauer, 'Development of a Micro Flow-system with Integrated Biosensor Array', *MESA Monographs, Micro Total Analysis System*, University of Twente, 249–259, 1994.
107. G. Jobst, P. Svasek, I. Moser, M. Varahram, Z. Trajanoski, P. Wach, P. Kotanko, F. Skrabal, G. Urban, 'Mass Producible Miniaturized Flow Through Device with Biosensor Array', *Sens. Actuators, B*, **43**, 121–125 (1997).
108. I. Moser, G. Jobst, E. Aschauer, P. Svasek, M. Varahram, G. Urban, V. Zanin, G. Tjoutrina, A. Zharikova, T. Berezov, *Biosens. Bioelectron.*, **10**, 527 (1995).
109. G. Jobst, I. Moser, M. Varahram, P. Svasek, E. Aschauer, Z. Trajanoski, P. Wach, P. Kotanko, F. Skrabal, G. Urban, 'Thin-film Microbiosensors for Glucose-lactate Monitoring', *Anal. Chem.*, **68**(18), 3173 (1996).
110. G. Jobst, I. Moser, P. Svasek, M. Varahram, G. Urban, 'Application of Miniaturized Liquid Handling System with Integrated Biosensor Array for Milk Analysis', in *Transducers '95, Eurosensors, Digest of Technical Papers*, Stockholm, 353–354, Vol. 2, 1995.
111. I. Moser, G. Jobst, P. Svasek, E. Svasek, M. Varahram, G. Urban, 'Rapid Liver Enzyme Assay with Miniaturized Liquid Handling System Comprising Thin Film Biosensor Array', *Sens. Actuators, B*, **44**(1–3), 377–380 (1997).
112. G. Urban, 'Microstructuring of Organic Layers for Microsystems', *Sens. Actuators, A*, **74**, 219–224 (1999).
113. M. Brischwein, W. Baumann, R. Ehret, A. Schwinde, M. Kraus, B. Wolf, 'Mikrosensorische Systeme in der zellbiologischen Grundlagenforschung und medizinischen Diagnostik', *Naturwissenschaften*, **83**, 193 (1996).
114. I. Walther, B.H. van der Schoot, S. Jeanneret, Ph. Arquint, N.F. de Rooij, V. Gass, B. Bechler, G. Lorenzi, A. Cogoli, 'Development of a Miniature Bioreactor for Continuous Culture in a Space Laboratory', *J. Biotechnol.*, **38**, 21 (1994).
115. L. Bousse, R.J. McReynolds, G. Kirks, T. Dawes, P. Lam, W.R. Bemiss, J.W. Parce, 'Micromachined Multichannel Systems for the Measurement of Cellular Metabolism', *Sens. Actuators, B*, **20**(2–3), 145 (1994).
116. C.D.T. Bratten, P.H. Cobbold, J.M. Cooper, 'Micromachining Sensors for Electrochemical Measurement in Sub-nanolitre Volumes', *Anal. Chem.*, **69**, 253 (1997).
117. G. Jobst, I. Moser, G. Urban, J. Dübendorfer, R. Kunz, 'Integrated Optical pH Sensor Using Replicated Chirped Grating Coupler Sensor Chips', *Sens. Actuators, B*, **50**(3), 210–219 (1998).
118. R.E. Kunz, J. Edlinger, P. Sixt, M.T. Gale, 'Replicated Chirped Waveguide Gratings for Optical Sensing Applications', *Sens. Actuators, A*, **47**, 482 (1995).

119. G. Jobst, G. Urban, A. Jachimowicz, F. Kohl, F. Olcaytug, I. Lettenbichler, G. Nauer, 'Thin-film Clark-type Oxygen Sensor Based on Novel Polymer Membrane Systems for in vivo and Biosensor Applications', *Biosens. Bioelectron.*, **8**(3/4), 123 (1993).
120. G. Ocvirk, H. Salimi-Moosavi, R. Szarka, E. Arriaga, P. Andersson, R. Smith, N. Dovichi, J. Harrison, 'Single Cell Enzymatic Analysis on a Microchip: Lysing of Single Cells and Identification of Theri β-galactosidase Activity', in *Micro Total Analysis Systems 98*, eds. J. Harrison, A. van den Berg, Kluwer Academic Publishers, Norwell, 203–207, 1998.
121. H. Hinkers, T. Hermes, C. Sundermeier, M. Borchardt, C. Dumschat, K. Cammann, M. Knoll, S. Bucher, M. Buhner, 'An Amperometric Microsensor Array with 1024 Individually Addressable Elements for Two-dimensional Concentration Mapping', *Sens. Actuators, B*, **24**(1/3), 300 (1995).
122. A. Folch, M. Toner, 'Microfluidic Patterning of Biological Material', in *Proceedings of Transducer 99*, Sendai, Japan, 116–119, 1999.
123. Anonymous, 'Microminiature Tools Affecting Lab Products, Labmedica International', *Biomed. Bus. Int.*, **11/12**, 5–6 (1994).
124. R. Anderson, G. McGall, R. Lipshutz, 'Polynucleotide Arrays for Genetic Sequence Analysis', *Topics in Current Chemistry*, Springer, Berlin, 117–131, Vol. 194, 1998.
125. P. Swanson, R. Gelbart, E. Atlas, S. Graham, L. Yang, W. Butler, D.E. Ackley, E. Sheldon, 'A Fully Multiplexed CMOS Biochip for DNA Analysis', in *Proceedings of Transducer 99*, Sendai, Japan, 48–51, 1999.

Capillary Electrophoresis in Clinical Chemistry

Yan Xu
Cleveland State University, Cleveland, USA

1 Introduction **1**
2 Separation Principle **2**
2.1 Terminology 2
2.2 Modes of Capillary Electrophoresis Operation 3
3 Instrumentation **5**
3.1 Capillary Column 5
3.2 Sample Injection 6
3.3 Detection 6
3.4 Commercial Instrumentation 6
4 Practical Considerations **7**
4.1 Qualitative Analysis 7
4.2 Quantitative Analysis 7
4.3 Sample Matrix Effect 7
4.4 Sample Preparation 7
4.5 Linearity and Reproducibility 8
5 Clinical Applications **8**
5.1 Protein Analysis 10
5.2 Drug Monitoring 10
5.3 DNA and RNA Analyses 11
5.4 Diagnosis of Metabolic Disorders 11
5.5 Single-cell Analysis 11
6 Future Prospects **11**
Abbreviations and Acronyms **12**
Related Articles **12**
References **12**

Capillary electrophoresis (CE) is electrophoresis in capillary format. It is the most efficient and versatile separation technique nowadays for analysis of small and large molecules. Clinical chemistry is a science that applies the knowledge of chemistry to the understanding of the human in health and disease. Implementation of CE technique in clinical laboratories will enhance the capability of the current clinical diagnostic system, and improve the quality of clinical testing. Owing to the nature of clinical diagnostics, the adoption of CE methods in clinical laboratories has been slow. However, CE has begun to replace some older and antiquated methods in some pioneering clinical laboratories.

This review includes a brief introduction to the CE technique (i.e. the separation principle, modes of operation, instrumentation, and practical considerations), and a summary of CE's applications in clinical chemistry (i.e. protein analysis, drug monitoring, DNA and RNA analyses, diagnosis of metabolic disorders, and single-cell analysis), as well as the future prospects for the CE technique.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

CE is electrophoresis in capillary format. It is the most efficient and versatile separation technique nowadays for analysis of small and large molecules. Although many concepts of CE separation had been demonstrated in the 1960s,[1] research on CE did not gain momentum until the early 1980s.[2–5] The transformation of conventional electrophoresis to modern CE was facilitated by the production of inexpensive narrow-bore fused-silica capillary, and by the development of highly sensitive on-line detectors for microcolumn liquid chromatography (LC).

A CE instrument (Figure 1) usually consists of the following basic components: a high-voltage power supply (0–30 kV), a fused-silica (SiO_2) capillary, two electrodes, two buffer reservoirs, and an on-line detector. Sample injection is accomplished by temporarily replacing one of the buffer reservoirs with a sample vial. A desirable amount of sample can be introduced into the capillary by controlling either voltage or pressure.

CE analysis is usually very fast and efficient, uses little sample and reagents, and costs much less than chromatography or conventional electrophoresis. The unprecedented efficiency of CE is an inherent feature of the technique. Like that of LC, the separation efficiency of CE can be modeled by the Van Deemter equation, which relates the plate height, H, to the velocity, v_x, of the carrier fluid along the separation axis x (Equation 1).

$$H = A + \frac{B}{v_x} + Cv_x \tag{1}$$

where A, B, and C are constants. A lower value of H corresponds to a higher separation efficiency – when the plate height is reduced, more theoretical plates (N) can be packed into a given length (L) along the separation axis (Equation 2).

$$N = \frac{L}{H} \tag{2}$$

The resolution (R_s) of two components in a mixture is proportional to $\sqrt{N}$ (Equation 3):

$$R_s = \frac{\sqrt{N}}{4}\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k'_2}{1 + k'_2}\right) \tag{3}$$

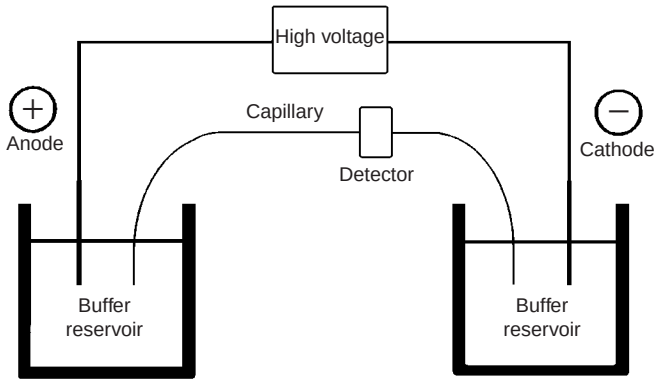

Figure 1 Basic set-up of a CE system.

where $\alpha = k'_2/k'_1$ is the selectivity of two adjacent peaks 1 and 2; and k'_1 and k'_2 are the capacity factors of these peaks. In CE, because the separation is carried out in a single phase of uniformly flowing carrier fluid, two of the three terms in Equation 1 (the multiple-path term, A, and the mass-transfer term, Cv_x) that contribute to the plate height are eliminated. Therefore, the only source of band broadening under ideal condition is the longitudinal diffusion term, B/v_x. Typically, a CE separation invokes 50 000 to 500 000 theoretical plates, which is an order-of-magnitude better than the competing LC method.

CE is the fastest growing analytical technique in the 1990s. Although CE is still in its early stage of development, it has demonstrated tremendous potential for the analysis of a wide range of biologically active molecules. In this article, the focus will be on the applications of CE in routine hospital laboratories and in specialized clinical settings. It is not intended to be comprehensive of all published papers; a selection has been made of those that are significant to clinical diagnosis of diseases.

2 SEPARATION PRINCIPLE

2.1 Terminology

2.1.1 Electrophoresis

Electrophoresis is defined as the migration of ions under the influence of an electric field. The force ($F_e = qE$) produced by the electrical field on an ion is directly proportional to its effective charge, q, and the electric field strength, E. The translational movement of the ion is opposed by a retarding frictional force ($F_f = fv_{ep}$) that is proportional to the velocity of the ion, v_{ep}, and the friction coefficient, f. The ion almost instantly reaches a steady-state velocity where the accelerating force equals the frictional force (Equation 4):

$$qE = fv_{ep} \tag{4}$$

Rearranging Equation (4) yields Equation (5):

$$v_{ep} = \frac{q}{f}E = \mu_{ep}E \tag{5}$$

where μ_{ep} is the electrophoretic mobility of the ion, a proportional constant between the velocity of the ion and the strength of the electric field. The electrophoretic mobility is directly proportional to the charge of the ion and inversely proportional to the friction coefficient.

The friction coefficient of the moving ion is related to the hydrodynamic radius of the ion, r, and the viscosity of the medium, η (Equation 6):

$$f = 6\pi\eta r \tag{6}$$

Because of $\mu_{ep} = q/f$, a larger hydrodynamic radius translates to a lower electrophoretic mobility.

2.1.2 Electroosmosis

Electroosmosis refers to the movement of a buffer solution in a capillary under the influence of an electric field. The inner surface of a fused-silica capillary is covered with silanol (Si−OH) groups, which are ionized to SiO^- at pH > 2. The negatively charged surface is counterbalanced by positive ions from the buffer, forming a so-called ''electrical double layer''. Under the influence of an electric field, the positive ions in the diffuse part of the double layer migrate towards the cathode and in doing so they entrain water of hydration and cause electroosmotic flow (EOF). The equations of EOF are identical to those developed for electrophoresis, as the two phenomena are complementary. The electroosmotic velocity, v_{eo}, is defined by Equation (7):

$$v_{eo} = \mu_{eo}E \tag{7}$$

where μ_{eo} is the electroosmotic mobility, a proportional constant between the velocity of the EOF and the strength of the electric field. Electroosmotic mobility is directly proportional to the dielectric constant of the medium, ε, and the zeta potential at the capillary-buffer interface, ζ, and inversely proportional to the viscosity of the medium, η (Equation 8):

$$\mu_{eo} = \frac{\varepsilon\zeta}{4\pi\eta} \tag{8}$$

The zeta potential is largely dependent on the electrostatic nature of the inner surface of the capillary, and to a small extent on the ionic strength of the buffer. In a fused-silica capillary, electroosmosis is diminished

at low pH because protons convert the charged SiO^- surface to SiOH, causing a decrease in zeta potential. Electroosmosis also decreases with increasing ionic strength owing to the collapse of the electrical double layer. Practically, EOF can be eliminated by derivatizing the silanol groups on the inner surface of a capillary with a polymer solution (e.g. polyacrylamide or methylcellulose).

2.1.3 Apparent Mobility

The apparent mobility of a solute, μ_{app}, is a vector sum of the electrophoretic mobility of the solute, μ_{ep}, and the electroosmotic mobility of the buffer, μ_{eo} (Equation 9):

$$\mu_{app} = \mu_{ep} + \mu_{eo} \tag{9}$$

The apparent velocity of a solute, v_{app}, is directly proportional to μ_{app} and the strength of an electric field across the capillary, E (Equation 10):

$$v_{app} = \mu_{app} E \tag{10}$$

In CE, the on-line detector is placed in the outlet side of the capillary, where the cathode is typically located. Neutral solutes migrate together under EOF, and are not separated. Cations and anions are separated by the differences in their apparent mobilities. Because cations move in the same direction as the EOF, their electrophoretic mobilities (μ_{ep}) have the same sign as the electroosmotic mobility (μ_{eo}); therefore, $\mu_{app} > \mu_{eo}$ and they reach the detector before neutral solutes. In the meantime, anions move in the opposite direction; thus, their electrophoretic mobilities (μ_{ep}) have the opposite sign compared to the electroosmotic mobility (μ_{eo}). At moderate pH values (pH > 3), μ_{eo} is usually greater than μ_{ep} and causes anions to migrate towards the detector following neutral solutes. At lower pH values, EOF is so weak that anions may never reach the detector unless the polarity of the voltage supply is reversed to change the location of the detector from the cathode to the anode side of the capillary.

2.2 Modes of Capillary Electrophoresis Operation

The main separation modes used in CE are listed in Table 1. They are:

- capillary zone electrophoresis (CZE)
- micellar electrokinetic capillary chromatography (MECC)
- capillary isotachophoresis (CITP)
- capillary gel electrophoresis (CGE)
- capillary isoelectric focusing (CIEF)
- capillary electrochromatography (CEC)

Each of these modes uses a high applied voltage to achieve efficient separation.

CZE is the simplest form of CE. In CZE, sample is applied as a narrow zone that is surrounded by a separation medium. During electrophoresis, each component in the sample zone migrates according to its own apparent mobility. Ideally, all sample components will eventually separate from each other to form individual zones of pure material. However, neutral molecules cannot be separated because they migrate at the velocity of EOF. Therefore, CZE is a separation technique only for the charged species. The separation of charged species is accomplished most efficiently when the differences among the apparent velocities of the components are

Table 1 Modes of CE operation, principles of separation, and applications

Modes	Principles	Applications
CZE	Differential migration (charge-to-mass ratio) in the medium	Ionic and polar species, lipids, peptides, proteins, RNA, DNA
MECC	*Neutral species*: differential partitioning between the aqueous buffer and the micellar phase *Charged species*: differential migration, and differential partitioning between the aqueous buffer and the micellar phase	Neutral and ionic species, lipids, peptides, RNA, DNA
CITP	Analytes with different mobilities are separated into different zones that migrate at the same velocity	Ionic and polar species, peptides, proteins
CGE	Analytes are separated according to their sizes in the nonconvective medium	Peptides, proteins, RNA, DNA
CIEF	Separation is based on the differences in isoelectric points	Peptides, proteins
CEC	*Neutral species*: differential partitioning between the mobile phase and the LC stationary phase *Charged species:* differential migration in the mobile phase	Neutral and ionic species

maximized and random dispersion of the individual zones is minimized.

The apparent migration time of a sample component, t_{app}, is the time required for the component to move from the beginning of the capillary to the detection window; it can be calculated from the distance between the sample inlet and the detection window, L_d, and the apparent velocity (Equation 11):

$$t_{app} = \frac{L_d}{v_{app}} = \frac{L_d}{\mu_{app} E} \tag{11}$$

MECC extends the application of CE to the separation of both neutral and charged species through the use of micelles in the separation buffer. Micelles are aggregates of amphiphilic monomers known as surfactants that possess a hydrophilic head and a hydrophobic tail. The hydrophobic tail of a surfactant is a straight or branched chain of hydrocarbon or a steroidal skeleton, whereas the hydrophilic head can be cationic, anionic, zwitterionic, or nonionic. When surfactant molecules exceed their critical micelle concentration (CMC), they are pushed together by the polar medium. Therefore, spherical micelles are formed with the hydrophobic tails pointing inward and the hydrophilic heads facing outward into the aqueous solution, resulting in a system of lower free energy. Typically, a micelle is composed of an average number of monomers (known as the aggregation number, (AN)) with a diameter of 3 to 6 nm. Both CMC and AN values are susceptible to the chemical composition of surfactant and the surrounding conditions such as temperature, pH, and ionic strength of the buffer solution.

In MECC, micelles serve as a pseudostationary phase that resembles the stationary phase in reversed-phase LC. Solutes partition between the aqueous phase and the micellar phase through various types of interactions, including hydrophobic, electrostatic, and hydrogen-bonding interactions. These interactions play a major role in the retention process of MECC.

Separation in MECC is a result of the combined effect of the differential partitioning of molecules between the aqueous buffer and the micellar phase, as well as the differential migration of ionic species. The apparent velocity of a solute in MECC can be written as Equation (12):

$$v_{app,MECC} = F_{aq} v_{app,aq} + F_{mc} v_{app,mc} \tag{12}$$

where $F_{aq} = n_{aq}/n_{total}$ and $F_{mc} = n_{mc}/n_{total}$, n_{total}, n_{aq}, and n_{mc} are the total number of moles of solute, the number of moles of solute in aqueous phase, and the number of moles of solute in micellar phase, respectively; $v_{app,aq} = v_{ep} + v_{eo}$, which is the apparent velocity of the solute in aqueous phase; $v_{app,mc} = v_{ep,mc} + v_{eo}$, which is the apparent velocity of the solute in micellar phase. Both $v_{app,aq}$ and $v_{app,mc}$ are vector sums. In most cases, $v_{eo} > v_{ep}$ and $v_{eo} > v_{ep,mc}$; therefore $v_{app,aq}$ and $v_{app,mc}$ usually have the same sign.

CITP is isotachophoresis performed in capillary format. Isotachophoresis means "same" (iso) + "velocity" (tacho) + "electrophoresis" (phoresis). It is also known as multizonal or displacement electrophoresis. In CITP, a sample is inserted between a leading electrolyte and a trailing electrolyte in the absence of electroosmosis. The leading electrolyte has a higher mobility and the trailing electrolyte has a lower mobility than the ions in the sample zone. Separation in CITP relies on the differences in the mobilities, $\mu_{ep,i}$, of analyte ions in the sample zone (Equation 13):

$$v_i = \mu_{ep,i} E_{sz} \tag{13}$$

where v_i is the velocity of species i, and E_{sz} is the electric field strength of the sample zone prior to separation. During the transient separation process, analyte ions are separated into consecutive bands under constant electric current and temperature. An ion band is defined as a homogeneous solution separated by moving or stationary boundaries. Because each ion band migrates at the same velocity, v, a steady-state band stacking is achieved after the separation (Equation 14):

$$v = \mu_{ep,i} E_i = \mu_{ep,j} E_j \tag{14}$$

where the subscripts i and j represent two different analyte ions; E_i and E_j are the electric field strengths of ion bands i and j. Since the electric field strength increases as the mobility of the ion band decreases, the order of the ion bands can be determined by Ohm's law.

The ion concentration in each band is adjusted to the ion concentration of the leading electrolyte according to the Kohlraugh regulating function (Equation 15):

$$C_i = C_l \frac{|\mu_i|(|\mu_l| + |\mu_c|)}{|\mu_l|(|\mu_i| + |\mu_c|)} \tag{15}$$

where C_i is the concentration of species i in ion band i, C_l is the concentration of the leading electrolyte, and μ_i, μ_l and μ_c are the mobilities of species i, the leading electrolyte ion, and the counter ion in the steady state, respectively. CITP can concentrate those analyte ions whose concentrations in the sample are lower than their steady-state concentrations defined by Equation (15).

Detection methods of CITP are based on conductivity, differential conductivity, or direct ultraviolet (UV) adsorption. In the latter case, non-UV-adsorbing spacer that has mobility between those of the two neighboring bands is used to separate two ion bands. Cations and anions can be detected in separated runs.

CGE is an adaptation of traditional slab gel electrophoresis to capillary format. It is CZE performed in a

polymeric gel medium. CGE is potentially useful for the separation of large biological molecules such as proteins and DNAs, which have similar electrophoretic migration rates in free solution due to their similar charge-to-mass ratios. CGE separates molecules according to their sizes in a nonconvective medium. Separation media include non-cross-linked polymers such as linear polyacrylamide, polyethylene glycol, and cellulose derivatives, as well as cross-linked polymers or gels such as polyacrylamide and agarose.

The entangled polymer network inside the capillary serves as a molecular sieve in which smaller molecules migrate faster than large ones. The polymer network reduces solute diffusion, and adsorption, as well as suppressing electroosmosis inside the capillary. These features increase the separation efficiency and permit the use of a shorter column. In comparison to non-cross-linked polymers, the resolution of cross-linked polymers can be easily optimized for a given range of molecular weights by varying the total monomer concentration and the degree of cross-linking. However, non-cross-linked polymers can be easily flushed out of the capillary when a problem develops (e.g. trapped air bubbles), and freshly loaded into the capillary for each individual separation.

CIEF is used for the separation of amphoteric substances such as proteins, peptides, amino acids, and pharmaceuticals in polymeric matrices, as well as in free solutions. Separation by CIEF is based on the differences in isoelectric points (pIs) of sample components rather than on the differences in their apparent velocities. In CIEF, a series of zwitterions known as ampholytes are used to form a pH gradient inside the focusing capillary. Ampholytes that are positively charged migrate towards the cathode while those that are negatively charged migrate towards the anode. After an ampholyte reaches its own pI, it is no longer charged and ceases migration. As a result, a stable pH gradient is formed, which has a high-value end at the cathode and a low-value end at the anode side of the capillary. During the separation, if an analyte possesses a net positive charge, it will migrate towards the cathode. Eventually, it will meet a pH at which it has zero net charge and cease migration. The pH gradient is smoother when a larger number of ampholytes are used. To prevent the migrations of buffers from the vials into the capillary, the buffer in the cathode side must have a pH value higher than the pIs of all basic ampholytes, and the buffer in the anode must have a pH value lower than the pIs of all acidic ampholytes.

CIEF is most effective when EOF and other convective forces are eliminated or greatly suppressed. However, it is possible to perform CIEF in the presence of electroosmosis, as long as the electroosmotic velocity of the solution inside the capillary does not exceed the electrophoretic velocities of the analytes.

CIEF is a true focusing technique. If a solute from a focused zone happens to diffuse away from the zone center, it immediately loses or gains protons, and thus acquires charge. In its charged state, the solute migrates back toward the zone center. Eventually, a steady state is reached where the zones are stationary and sharply focused. The width of solute zone can be characterized by the variance of Gaussian distribution σ^2 (Equation 16):

$$\sigma^2 = \frac{D}{E}\frac{\mathrm{d(pH)}}{\mathrm{d}\mu_{\mathrm{app}}}\frac{\mathrm{d}x}{\mathrm{pH}} \tag{16}$$

Small variance that results in a sharp zone is favored by a high field strength (E), a low diffusion coefficient (D), and high value of $\mathrm{d}\mu_{\mathrm{app}}/\mathrm{d(pH)}$ (the rate of change of mobility with pH), as well as high value of $\mathrm{pH}/\mathrm{d}x$ (the slope of the pH gradient). For a complete separation of two solutes, a difference between two pIs must be greater than 4σ.

CEC combines the selectivity of LC and the efficiency of CE for separation. In CEC, the capillary is packed with a stationary phase ($d_{\mathrm{p}} = 1.5–3\,\mu\mathrm{m}$) similar to those used in reversed-phase LC, and the mobile phase is driven by electroosmosis. CEC has high separation efficiency and selectivity for both neutral and charged species. Separation in CEC is based on the differential partitioning of neutral species between mobile phase and stationary phase, as well as the differential migration of ionic species in an electric field. In CEC, the stationary phase assumes dual roles as both selector and EOF generator.

3 INSTRUMENTATION

3.1 Capillary Column

The capillary column is the heart of CE technology. Owing to its large surface area to volume ratio, a capillary column can dissipate heat much more efficiently than the slab gel used in conventional electrophoresis. As a result, high voltage is used in CE for rapid and efficient separations. Fused silica is the most frequently used material for the capillary column, though other materials (e.g. Teflon and borosilicate glass) have been used. The widespread use of fused silica is attributable to its intrinsic properties such as transparency over a wide range of electromagnetic wavelengths and high thermal conductivity. Fused silica is also readily drawn into capillaries with diameters of a few micrometers. Furthermore, the inner surface of fused-silica capillaries can be modified by the covalent attachment of neutral or hydrophilic substituents of silane. Such coated capillary columns may be used to reduce EOF and prevent the adsorption of analytes, as well as to enhance the separation efficiency.

The preparation of an uncoated fused-silica capillary column for its first use is rather simple. The capillary is rinsed firstly with 10–15 column volumes of 0.1 M NaOH, then 10–15 column volumes of water and 5–10 column volumes of separation buffer. The preparation of a coated capillary is the same except that 0.1 M NaOH is replaced with methanol. In commercial instruments, solutions are forced through the capillary by either applying pressure to the inlet vial or reducing pressure at the outlet vial.

In an uncoated capillary, the direction of EOF can be reversed by the addition of a cationic surfactant such as cetyltrimethylammonium bromide (CTAB) to the separation buffer. The positively charged head of CTAB interacts with the negatively charged silanol group on the inner surface of the capillary, while the hydrophobic tail of CTAB points away from the surface. As the second layer of CTAB orients itself in the opposite direction, a bilayer of surfactant is formed on the surface, which effectively reverses the surface charge, from negative to positive, as well as EOF. Such procedure is known as dynamic coating.

3.2 Sample Injection

One particular advantage of CE is its ability to analyze extremely small volumes of sample. Typical injection volumes in CE range from picoliters to nanoliters. To perform such a task, two methods are commonly applied (i.e. hydrodynamic and electrokinetic injection).

Hydrodynamic injection is carried out by applying a pressure drop between the inlet and the outlet of a capillary column. The amount of sample injected can be calculated by the Poiseuille equation (Equation 17):

$$V_c = \frac{\Delta P \pi d^4 t}{128 \eta L_t} \tag{17}$$

where V_c is the calculated injection volume, ΔP is the pressure drop between the ends of the capillary, d is the inside diameter of the capillary, t is the time of injection, η is the viscosity of the separation medium, and L_t is the total length of the capillary.

Electrokinetic injection is accomplished by simply turning on the voltage for a short period of time. The number of moles of each analyte injected, Q_i, is determined by the apparent velocity of each analyte, v_{app}, the time of injection, t, and the conductivity ratio of separation medium and sample, k_m/k_s (Equation 18):

$$Q_i = v_{app} \left(\frac{k_m}{k_s} \right. \quad t \pi r^2 C_i \tag{18}$$

where r is the radius of the capillary and C_i is the molar concentration of species i. Since different analytes have their own mobilities, the concentrations of the injected analytes may be different from those in the original sample. Although electrokinetic injection may lead to biased results in quantitative analysis, it is usually not a problem in qualitative analysis. Moreover, electrokinetic injection is the sole choice of sample introduction in most CGE applications because the cross-linked polymers inside the capillary prevent the use of hydrodynamic injection.

3.3 Detection

With modifications of instrumentation and techniques originally developed for microcolumn LC, a wide variety of detection schemes is available for CE applications. Generally, spectrophotometric detection in CE is about an order of magnitude less sensitive compared to that in LC because the former detector has a shorter optical pathlength. However, the reduced sensitivity can be partially compensated by the high separation efficiency of CE, which allows for precise integration of peak areas (PAs). Moreover, other methods can be used to improve the limit of detection (LOD), which include the uses of low detection wavelengths (down to 185 nm) where many solutes have higher absorptivities, sample stacking technique[6] and packed-inlet capillary[7] for on-column concentration of samples, and new cell designs for extended optical pathlength.[8] As a general reference, Table 2 provides a list of CE detectors and their representative LODs.[9]

3.4 Commercial Instrumentation

Commercial CE systems specifically designed for research and high-throughput clinical analysis are available

Table 2 CE detectors and their representative LODs

Principles of measurement	LODs (mol)
Spectrophotometric detection	
Adsorption	$10^{-15} - 10^{-13}$
Fluorescence	
Pre-column derivatization	$10^{-20} - 10^{-17}$
On-column derivatization	8×10^{-16}
Post-column derivatization	2×10^{-17}
Indirect	
UV	$10^{-13} - 10^{-12}$
Fluorescence	5×10^{-17}
Thermal lens	4×10^{-17}
Raman	2×10^{-15}
Mass spectrometric detection	1×10^{-17}
Electrochemical detection	
Amperometric	7×10^{-19}
Conductivity	1×10^{-16}
Radiometric detection	1×10^{-19}

through several instrument manufacturers. The representatives of the state-of-the-art instruments include (not limited to) P/ACE™ 5000 series, P/ACE™ System MDQ and Paragon CZE™ 2000 of Beckman Coulter, HP^{3D} of Hewlett-Packard, Quanta 4000 series of Waters, and BioFocus® 2000 of Bio-Rad.

4 PRACTICAL CONSIDERATIONS

4.1 Qualitative Analysis

Qualitative analysis by CE provides identifications for the peaks in an electropherogram. The simplest way to identify a CE peak is to compare its migration time with that of a known compound. However, as in other separation techniques, the time of migration alone is not always sufficient to confirm peak identity and purity. It is necessary to obtain additional information about the peak for final confirmation. One way to achieve this is to compare the ratios of absorbances at different wavelengths between the unknown and a standard in spectrophotometric detection mode, or the ratios of currents in potential regions where most changes occur in amperometric detection mode.[10]

4.2 Quantitative Analysis

Quantitative analysis provides information about the amount or the concentration of a substance in a given sample. CE will become a widely accepted analytical technique only when it is truly quantitative. Currently, two main approaches (i.e. direct and internal calibration) are used in CE quantitation.

4.2.1 Direct Calibration

The concentration of a solute is directly proportional to the peak height, and residence time of the solution in a detector is directly proportional to the peak width. Therefore, the amount of solute is directly proportional to the PA, and quantitative information can be obtained by directly comparing the PA or the height with those of the calibrators.

4.2.2 Internal Calibration

Internal calibration usually results in better precision compared to direct calibration because neither the amount injected nor the detector response needs to remain constant. In internal calibration, a known amount of internal standard is added to each calibrator and unknown sample prior to the sample pretreatment procedure. After pretreatment, the solutions containing analyte and internal standard are subjected to electrophoresis. The calibration graph is constructed by plotting the ratios of PA (or height) of the calibrators to that of the internal standard against the concentrations of the calibrators. The concentration of unknown sample can be determined from the graph by knowing the ratio of PA (or height) between the unknown and internal standard.[11]

4.3 Sample Matrix Effect

The sample matrix strongly influences quantitative precision and accuracy in CE, especially when electrokinetic injection is used for sample introduction. The sample matrix is a major problem in analysis of samples of biological fluids, such as serum. As the sample size increases, the matrix effect becomes more apparent. The matrix effect can be accounted for by internal calibration using matrix-corrected peak areas (COPA) instead of PA for analyte and internal standard.[12] COPA can be calculated by dividing PA obtained in a given sample matrix by a matrix factor (F_M) that represents the ratio of PA of an analyte in the given sample matrix to PA of the analyte in the calibration matrix. This can be accomplished by either spiking blank solution with the analyte of interest at known concentrations or by standard addition of the analyte to the sample solution.

4.4 Sample Preparation

Many CE applications do not require sample pretreatment other than a simple dilution, whereas some other applications require the sample to be treated before injection, especially when the specificity or the detection limit of a method is our concern. Pretreatment methods used for CE include liquid–liquid extraction[13–25] (Table 3), solid–liquid extraction[26–34] (Table 4), and ultrafiltration and microdialysis[35–43] (Table 5), as well as acetonitrile deproteinization. The latter technique is well suited for

Table 3 Some CE applications that use liquid–liquid extraction

Analytes	Sample matrices	CE modes	Ref.
Antiepileptics	Serum	MECC	13
Cicletanine	Plasma	MECC	14
Cimetidine	Urine	CZE	15
Cocaine/morphine	Hair	CZE	16
Creatinine	Serum	CZE	17
Iohexol	Serum	CZE	18
Naproxen	Serum	CZE	19
Pentobarbital	Serum	CZE	20
Phenylalanine	Serum	CZE	21
Purine nucleotides	Cord plasma	CZE	22
Suramin	Serum	CZE	23
Theophylline	Plasma	CZE	24
Warfarin	Plasma	CZE	25

Table 4 Some CE applications that use solid–liquid extraction

Analytes	Sample matrices	CE modes	Ref.
γ-Aminobutyrate	Cerebrospinal fluid	CITP	26
Barbiturates	Serum	MECC	27
Benzodiazepines	Urine	MECC	28
Benzylpenicillin	Gastric contents	CZE	29
Morphine-3-glucuronide	Urine	CZE/MECC	30
Neurotensin	Tissues	CZE	31
Nitrazepam/metabolites	Urine	MECC	32
Racemethorphan/racemorphan isomers	Urine	MECC	33
Vasoactive intestinal peptide	Brain	CZE	34

Table 5 Some CE applications that use microdialysis or ultrafiltration

Analytes	Sample matrices	CE modes	Refs.
Amino acid	Frontoparietal cortex	CZE	35
Antineoplastic SR 4233/metabolite SR 4317	Jugular vein	MECC	36
Glutamate	Brain	CZE	37–39
Low-molecular-mass proteins	Seminal plasma	CZE	40[a]
Phenobarbital	Blood/brain	CZE	41
Phenobarbital/norepinephrine/amino acids	Tissues	CZE	42
Salicylate/acetaminophen/antiepileptics	Serum	CZE/MECC/CITP	43[a]

[a] Ultrafiltration.

analysis of small molecules by CE. It does not only remove proteins but also induces sample stacking in CE.[44,45]

4.5 Linearity and Reproducibility

The linear dynamic range of CE calibration can be as narrow as one order of magnitude[23] or as wide as six orders of magnitude,[46] depending on the analytes. The reproducibility (defined as the percent of coefficient of variation of replicate analyses) is typically in the range of 1–2% for PA, and 3–7% for peak height.

5 CLINICAL APPLICATIONS

Clinical chemistry is a science that applies the knowledge of chemistry to the understanding of the human in health

Table 6 Examples of protein analysis by CE

Proteins	Diseases	CE mode	Assay time (min)	Refs.
Cerebrospinal fluid protein				
Synaptotagmin	Alzheimer disease	CZE	30	49
Serum protein				
Alumin, α_1, α_2, β, γ	Dysproteinemia/paraproteinemia	CZE	2–4	50, 51
Hemoglobin variants (A_2, S, F, A_0, A_{1c})	Hemoglobinopathies/thalassemias	CIEF	18	52
Lactate dehydrogenase isoenzymes (LD_1, LD_2, LD_3, LD_4, LD_5)	Liver/heart disease	CZE	20	53
Lipoproteins (HDL, IDL, LDL, VLDL)	Lipoprotein metabolism disorders	CITP	7	54
Transferrin sialoforms	Carbohydrate-deficient glycoprotein syndrome	CZE	8	55
Urine protein				
Albumin, Bence Jones, intact IgG (Kappa), Tamm Horsfall	Multiple myeloma/glomerular proteinuria	CZE	15	56
Albumin, Bence Jones, γ, Tamm Horsfall, transferrin	Renal disorders/light chain disease	CZE	10	57

and disease. In modern medicine, the findings of the clinical laboratory play a crucial role in diagnosis of diseases and in management of patients. The nature of clinical testing has meant that the implementation of CE methods in clinical laboratories has been slow. However, CE has begun to replace some older and antiquated methods in some pioneering clinical laboratories. The following clinical applications show that CE methods enhance the capabilities of clinical diagnostics in such ways that are markedly faster, analytically superior, and

Table 7 Examples of drug monitoring by CE

Drugs	CE modes	Refs.
Abuse drugs		
Amphetamine, morphine, heroin, codeine, caffeine	MECC	58
Cocaine, morphine	MECC	59, 60
d-Lysergic acid diethylamide and its metabolites	CZE/MS CZE	61
Morphine 3-glucuronide	CZE	62
Triamterene, acebutolol, bendroflumethiazide	CZE	63
Analgesic agents		
Antipyrine	MECC	64
Ibuprofen and its metabolites	CZE	65, 66
Methadone and its metabolites	CZE	67
Naproxen, quinidine, salicylate and its metabolites	CZE	68
Anti-infective agent		
Cephalexin	CZE	69
Anticoagulant agents		
7-Hydroxycoumarin	CZE	70, 71
Mephenytoin and its metabolite	MECC	72
Anticonvulsant agents		
Gabapentin	CZE	73
Zopiclone enantiomers	CZE	74
Antiepileptic agents		
Phenobarbital, phenytoin, sodium valproate, carbamazepine	MECC/CZE	75
Antimalarial agents		
Proguanil, chloroquine, and their metabolites	CZE	76
Antineoplastic agents		
Prospidin	CZE	77
Methotrexate	CZE	78
Cardiovascular agents		
Amiloride, metoprolol, deacetylmetipranolol, labetalol, furosemide	CITP	79
Oxprenolol enantiomers and their metabolites	CZE	80
Diltiazem, desacetydiltiazem	CZE	81
Central nervous system stimulants		
Strychnine, brucine	CZE	82
Diuretic agent		
Theophylline and its metabolites	CZE/MECC	83, 84
H_2-receptor antagonists		
Cimetidine and its metabolites	CZE	85
Famotidine	CZE	86
Hypnotic agent		
Zolpidem and its metabolite	CZE	87
Sympathomimetic agents		
Clenbuterol	CZE	88
Ephedrine, pseudoephedrine, norephedrine	CZE	89

CZE/MS, capillary zone electrophoresis/mass spectrometry.

more cost effective compared with the current methods, or not attainable with other techniques.

5.1 Protein Analysis

A large number of clinical assays are based on the electrophoretic determination of proteins in serum, urine, and cerebrospinal and other biological fluids. To date, routine clinical protein analyses are mostly carried out in the slab-gel format, which is labor-intensive and difficult for total automation; therefore it is cost ineffective. The development of CE methods has shown undisputed advantages in respect to separation efficiency, system automation, and data acquisition over those of slab-gel methods, as well as higher detection sensitivity for small monoclonal proteins.[47,48] Hence, CE methods have a promising future in replacing slab-gel methods. Examples of protein analysis by CE[49–57] are given in Table 6.

5.2 Drug Monitoring

Clinical drug analysis includes several areas of interest: (a) therapeutic drug monitoring of specific drugs or metabolites in blood for assessment of organ function,

Table 8 Examples of DNA and RNA analyses by CE

Applications	Methods	Refs.
Genetic Diseases		
Androgen insensitivity syndrome	PCR/CAG triplet polymorphism	90
Angiotensin I-converting enzyme gene	Deletion/insertion polymorphism	91
Apolipoprotein B gene	PCR/VNTR	92
Apolipoprotein E gene	PCR/RFLP	93, 94
Cancer (microsatellite instability)	PCR/CA repeat	95
Congenital adrenal hyperplasia	PCR/ARMS	96, 97
Cystic fibrosis	Allele-specific PCR/deletion	98
Cystic fibrosis	PCR/ΔF508	99–101
Cystic fibrosis	PCR/GATT microsatellite	102
Cystic fibrosis	PCR/point mutant (TGCE)	103, 104
Duchenne/Becker muscular dystrophy	Multiplex PCR reactions	105
ERBB2 oncogene	RFLP	106
Factor V gene	Allele specific PCR	107
Kennedy's disease	CAG triplet	108
K-ras gene (lung cancer)	PCR/point mutant	109
Medium-chain Acyl-CoA dehydrogenase deficiency	Allele specific PCR	110
N-ras gene (human cancer)	SSCP/point mutant	111
P53	SSCP/DNA hybridization	112, 113
Thalassemia	Point mutant (TGCE)	114
TX gene	PCR	115
Von Willebrand factor gene	PCR/VNTR	116
Gene Dosage		
Basic fibroblast growth factor	Competitive PCR	117
Down's syndrome	Quantitative PCR	118
Follicular lymphomas	Competitive PCR	119
Rh D/d genotyping	Quantitative PCR	120
Microbiology/Virology		
Hepatitis B virus	RT-PCR	121
Hepatitis C virus	RT-PCR	122
HIV-1	RT-PCR	123–125
Mycobacterium tuberculosis	SSCP/dideoxy fingerprinting	126
Polio virus	RT-PCR	123
Forensic Analysis		
Mitochondrial DNA	PCR	127, 128
VNTRs at locus D1S80	PCR/VNTR	129–131
VNTRs at locus HUMTH01	PCR/VNTR	132, 133
X-Y homologous amelogenin gene	PCR	134

PCR, polymerase chain reaction; VNTR, variable number of tandem repeats; ARMS, amplification refractory mutation system; RT-PCR, reverse-transcription polymerase chain reaction; SSCP, single-strand conformation polymorphism; RFLP, restriction fragment length polymorphism; TGCE, thermal gradient capillary electrophoresis.

pharmacokinetics, intoxications, and metabolic patterns of pharmacogenetics, (b) screening of drugs of abuse and their metabolites in urine, and (c) athletic drug testing. Currently the high-throughput immunoassay autoanalyzer is the workhorse in clinical drug analysis. Approximately 90% of the drugs of clinical interest are assayed using various types of homogeneous immunoassays, and the remaining 10% of the drugs are analyzed using LC because of the lack of commercial immunoassays. The limited use of LC in clinical drug analysis is mainly attributed to several factors, including (a) labor intensive and difficult for total automation, (b) low sample throughput, and (c) requiring operation experience and periodic troubleshooting. The advent of the CE technique has provided an attractive alternative to LC because of its separation efficiency, choices of operation mode, and low operational cost. In CE not only the drug itself can be measured in the run, but its metabolites too. Also in drug screening, several drugs can be detected and identified in CE at the same time rather than only one general class of drugs being detected as in immunoassays. Examples of drug analysis by CE[58–89] are shown in Table 7.

5.3 DNA and RNA Analyses

DNA and RNA analyses play an important role in genetic research, clinical diagnosis of diseases, and forensic testing. Up to date, most separation methods developed in the slab-gel format for mutation and polymorphism analyses have been adopted in the capillary format. The advantages of CE include (a) total system automation (i.e. sample loading, separation, detection and report), (b) minute sample requirement, and (c) multiple usages of capillary. Table 8 provides an overview of CE applications in this field.[90–134]

5.4 Diagnosis of Metabolic Disorders

Metabolic disorders often result in accumulation of characteristic metabolites of physiological molecules in body fluids. A metabolic profile is required for the evaluation of organ function, or in searching for the cause of definite symptoms of an unknown disease. In profiling, a multicomponent analytical technique that is able to detect structurally or metabolically related analytes is needed. Because of its high separation efficiency and versatility, CE is a multicomponent analytical technique and suitable for the detection of important changes in metabolic profiles. Although CE has not yet become a routine technique used in clinical diagnosis of disease, it is clearly evident that it is an excellent alternative to chromatography and gel electrophoresis for profiling metabolites of physiological molecules in metabolic disorders[135–153] (Table 9).

Table 9 Examples of diagnosis of metabolic disorders by CE

Metabolic disorders	CE modes	Refs.
Arthritis	CZE	135, 136
Adenylosuccinate lyase deficiency	CZE	137
Cobalamin deficiency	CZE	138, 139
2,8-Dihydroxyladenine urolithiasis	CZE	140
Neurological disorders	CZE	141–143
Neuropsychiatric diseases	CZE	144
Organic acidurias	CZE	145, 146
Phenylketonuria	CZE	147
Serum iron	CZE	148
Testosterone metabolites	MECC	149
Uremia	CZE	150
Urinary estrogens	MECC	151
Vitamin A	CZE	152
Xanthinuria	CZE	153

Table 10 Examples of single-cell analysis by CE

Cells	CE modes	Refs.
Adrenal medullary cells	CZE	154
Erythrocytes	CZE	155–159
Leukemia cells	MECC	160
Lymphocytes	CZE	161–162
Pheochromocytoma cells	CZE	163

5.5 Single-cell Analysis

A cell is a unit of life. Chemical analysis of single cells provides us the knowledge of the chemical composition of individual cells that should lead to a better understanding of the chemistry of life. Information of the chemical composition of single cells promises to advance our knowledge of such diverse processes as healthy or diseased states to the molecular level, and our understanding of cellular differentiation, intracellular communication, and neurotransmission, as well as the physiological effects of internal and external stimuli.

CE is uniquely suited for single-cell analysis because it can handle small sample volumes, profile a variety of compounds at very low concentrations, and provide quality analytical information. Even though single-cell analysis is far from its clinical application, it will definitely play a significant role in biological and medical research. Table 10 shows some examples of single-cell analysis by CE.[154–163]

6 FUTURE PROSPECTS

Although CE has proved to be a powerful separation technique that often provides higher efficiency than LC and conventional gel electrophoresis, it must also

prove its ability to withstand the rigors of the clinical environment before it will be fully accepted in clinical laboratories. Practically, CE must prove to be reliable for routine clinical testing, be easy to use, low cost, and high throughput; be rugged and as maintenance-free as possible; and be extremely good at performing at least one important clinical analysis. This author believes that the CE technique expands the capability of clinical diagnostic systems and the implementation of CE in clinical laboratories is an inevitable trend for clinical chemistry.

ABBREVIATIONS AND ACRONYMS

AN	Aggregation Number
ARMS	Amplification Refractory Mutation System
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CGE	Capillary Gel Electrophoresis
CIEF	Capillary Isoelectric Focusing
CITP	Capillary Isotachophoresis
CMC	Critical Micelle Concentration
COPA	Matrix-corrected Peak Areas
CTAB	Cetyltrimethylammonium bromide
CZE	Capillary Zone Electrophoresis
CZE/MS	Capillary Zone Electrophoresis/ Mass Spectrometry
EOF	Electroosmotic Flow
LC	Liquid Chromatography
LOD	Limit of Detection
MECC	Micellar Electrokinetic Capillary Chromatography
PA	Peak Area
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse-transcription Polymerase Chain Reaction
SSCP	Single-strand Conformation Polymorphism
TGCE	Thermal Gradient Capillary Electrophoresis
UV	Ultraviolet
VNTR	Variable Number of Tandem Repeats

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Pulp and Paper **(Volume 9)**
Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis

Electronic Absorption and Luminescence **(Volume 12)**
Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis

REFERENCES

1. S. Hjerten, 'Free Zone Electrophoresis', *Chromatogr. Rev.*, **9**, 122–239 (1967).
2. J.W. Jorgenson, K.D. Lukacs, 'Zone Electrophoresis in Open-tubular Glass Capillaries', *Anal. Chem.*, **53**, 1298–1302 (1981).
3. J.W. Jorgenson, K.D. Lukacs, 'High-resolution Separations Based on Electrophoresis and Electroosmosis', *J. Chromatogr.*, **218**, 209–216 (1981).
4. J.W. Jorgenson, K.D. Lukacs, 'Capillary Zone Electrophoresis', *Science*, **222**, 266–272 (1983).
5. J.P. Landers, 'Clinical Capillary Electrophoresis', *Clin. Chem.*, **41**, 495–509 (1995).
6. R-L. Chien, D.S. Burgi, 'On-column Sample Concentration Using Field Amplification in CZE', *Anal. Chem.*, **64**, 489A–496A (1992).
7. M.E. Swartz, M. Merion, 'On-line Sample Preconcentration on Packed-inlet Capillary for Improving the Sensitivity of Capillary Electrophoretic Analysis of Pharmaceuticals', *J. Chromatogr.*, **632**, 209–213 (1993).
8. M. Albin, P.D. Grossman, S.E. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 489A–497A (1993).

9. A.G. Ewing, R.A. Wallingford, T.M. Olefirowicz, 'Capillary Electrophoresis', *Anal. Chem.*, **61**, 292A–303A (1989).
10. M.S. Lunte, T.J. O'Shea, 'Pharmaceutical and Biomedical Applications of Capillary Electrophoresis/Electrochemistry', *Electrophoresis*, **15**, 79–86 (1994).
11. X. Liu, Y. Xu, M.P.C. Ip, 'Capillary Electrophoretic Enzyme Immunoassay for Digoxin in Human Serum', *Anal. Chem.*, **67**, 3211–3218 (1995).
12. J. Leube, O. Roeckel, 'Quantification in Capillary Zone Electrophoresis for Samples Differing in Composition from the Electrophoresis Buffer', *Anal. Chem.*, **66**, 1090–1096 (1994).
13. A. Schmutz, W. Thormann, 'Factors Affecting the Determination of Drugs and Endogenous Low Molecular Mass Compounds in Human Serum by Micellar Electrokinetic Capillary Chromatography with Direct Sample Injection', *Electrophoresis*, **15**, 51–61 (1994).
14. J. Prunonosa, R. Obach, A. Diez-Cascon, L. Gouesclou, 'Comparison of High-performance Liquid Chromatography and High-performance Capillary Electrophoresis for the Determination of Cicletanine in Plasma', *J. Chromatogr.*, **581**, 219–226 (1992).
15. S. Arrowood, A.M. Hoyt, Jr, 'Determination of Cimetidine in Urine by Capillary Zone Electrophoresis', *Microchem. J.*, **47**, 90–95 (1993).
16. F. Tagliaro, C. Poiesi, R. Aiello, R. Dorizzi, S. Ghielmi, M. Marigo, 'Capillary Electrophoresis for the Investigation of Illicit Drugs in Hair: Determination of Cocaine and Morphine', *J. Chromatogr.*, **638**, 303–309 (1993).
17. Y. Ma, Z. Wu, H.C. Furr, C. Lammi-Keefe, N.E. Craft, 'Fast Minimicroassay of Serum Retinol (Vitamin A) by Capillary Zone Electrophoresis with Laser-Excited Fluorescence Detection', *J. Chromatogr., B*, **616**, 31–37 (1993).
18. Z.K. Shihabi, M.S. Constantinescu, 'Iohexol in Serum Determined by Capillary Electrophoresis', *Clin. Chem.*, **38**, 2117–2120 (1992).
19. H. Soini, M.V. Novotny, 'Determination of Naproxen in Serum by Capillary Electrophoresis with Ultraviolet Absorbance and Laser-induced Fluorescence Detection', *J. Microcolumn Sep.*, **4**, 313–318 (1993).
20. Z.K. Shihabi, 'Serum Pentobarbital Assay by Capillary Electrophoresis', *J. Liq. Chromatogr.*, **16**, 2059–2068 (1993).
21. F. Tagliaro, S. Moretto, R. Valentini, G. Gamboro, M. Moffa, L. Tato, 'Capillary Zone Electrophoresis Determination of Phenylalanine in Serum: a Rapid, Inexpensive and Simple Method for the Diagnosis of Phenylketonuria', *Electrophoresis*, **15**, 94–97 (1994).
22. T. Grune, G.A. Ross, H. Schmidt, W. Siems, D. Perrett, 'Optimized Separation of Purine Bases and Nucleotides in Human Cord Plasma by Capillary Zone Electrophoresis', *J. Chromatogr.*, **636**, 105–111 (1993).
23. L.L. Garcia, Z.K. Shihabi, 'Suramin Determination by Capillary Electrophoresis', *J. Liq. Chromatogr.*, **16**, 2049–2057 (1993).
24. I.M. Johansson, M-B. Groen-Rydberg, B. Schmekel, 'Determination of Theophylline in Plasma Using Different Capillary Electrophoretic Systems', *J. Chromatogr.*, **652**, 487–493 (1993).
25. P. Gareil, J.P. Gramond, F. Guyon, 'Separation and Determination of Warfarin Enantiomers in Human Plasma Samples by Capillary Zone Electrophoresis Using a Methylated β-Cyclodextrin-containing Electrolyte', *J. Chromatogr., B*, **615**, 317–325 (1993).
26. M. Hutta, D. Kaniansky, E. Simunicova, V. Zelenska, V. Madajova, A. Siskova, 'Solid-phase Extraction for Sample Preparation in Trace Analysis for Inorganic Compounds by Capillary Electrophoresis', *J. Radioanal. Nucl. Chem.*, **163**, 87–98 (1992).
27. M. Tomita, T. Okuyama, S. Sato, H. Ishizu, 'Determination of Barbiturates in Serum by Capillary Electrophoresis', *Hochudoku*, **10**, 120–121 (1992).
28. M. Schafroth, W. Thormann, D. Allemann, 'Micellar Electrokinetic Capillary Chromatography of Benzodiazepines in Human Urine', *Electrophoresis*, **15**, 72–78 (1994).
29. S. Arrowood, A.M. Hoyt, Jr, M.J. Sepaniak, 'Monitoring of Benzylpenicillin Decomposition in Gastric Contents by Capillary Zone Electrophoresis', *J. Chromatogr., B*, **583**, 105–110 (1992).
30. P. Wernly, W. Thormann, D. Bourquin, R. Brenneisen, 'Determination of Morphine-3-glucuronide in Human Urine by Capillary Zone Electrophoresis and Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr., B*, **616**, 305 (1993).
31. L. Denoroy, 'Detection of Neurotensin in Tissue by Capillary Zone Electrophoresis', *Electrophoresis*, **15**, 46–50 (1994).
32. M. Tomita, T. Okuyama, S. Sato, H. Ishizu, 'Determination of Nitrazepam and its Metabolites in Urine by Capillary Electrophoresis', *Hochudoku*, **10**, 122–123 (1992).
33. A. Aumatell, R.J. Wells, 'Chiral Differentiation of the Optical Isomers of Racemethorphan and Racemorphan in Urine by Capillary Zone Electrophoresis', *J. Chromatogr. Sci.*, **31**, 502–508 (1993).
34. J. Soucheleau, L. Denoroy, 'Determination of Vasoactive Intestinal Peptide in Rat Brain by High-performance Capillary Electrophoresis', *J. Chromatogr.*, **608**, 181–188 (1992).
35. T.J. O'Shea, P.L. Weber, B.P. Bammel, C.E. Lunte, S.M. Lunte, M.R. Smyth, 'Monitoring Excitatory Amino Acid Release In Vivo by Microdialysis with Capillary Electrophoresis-electrochemistry', *J. Chromatogr.*, **608**, 189–195 (1992).
36. B.L. Hogan, S.M. Lunte, J.F. Stobaugh, C.E. Lunte, 'On-line Coupling of In Vivo Microdialysis Sampling

with Capillary Electrophoresis', *Anal. Chem.*, **66**, 596–602 (1994).
37. L. Hernandez, S. Tucci, N. Guzman, X. Paez, 'In Vivo Monitoring of Glutamate in the Brain by Microdialysis and Capillary Electrophoresis with Laser-induced Fluorescence Detection', *J. Chromatogr.*, **652**, 393–398 (1993).
38. L. Hernandez, N. Joshi, E. Murzi, P. Verdeguer, J.C. Mifsud, N. Guzman, 'Collinear Laser-induced Fluorescence Detection for Capillary Electrophoresis Analysis of Glutamic Acid in Brain Dialyzates', *J. Chromatogr.*, **652**, 399–405 (1993).
39. L. Hernandez, J. Escalona, P. Verdeguer, N.A. Guzman, 'In Vivo Monitoring of Brain Glutamate by Microdialysis Coupled to Capillary Electrophoresis and Laser-induced Fluorescence Detection', *J. Liq. Chromatogr.*, **16**, 2149–2160 (1993).
40. J. Tie, S. Tsukamoto, S. Oshida, 'An Analysis of Low Side Molecular Weight Proteins in Human Seminal Plasma by Capillary Electrophoresis', *Nippon Hoigaku Zasshi*, **47**, 295–301 (1993).
41. S. Tellez, N. Forges, A. Roussin, L. Hernandez, 'Coupling of Microdialysis with Capillary Electrophoresis: a New Approach to the Study of Drug Transfer between Two Compartments of the Body in Freely Moving Rats', *J. Chromatogr.*, **581**, 257–266 (1992).
42. A. Roussin, P. Verdeguer, L. Hernandez, 'Coupling of Microdialysis and Capillary Electrophoresis for In Vivo Analysis', *Analusis*, **21**, M42–M45 (1993).
43. J. Caslavska, S. Lienhard, W. Thormann, 'Comparative Use of Three Electrokinetic Capillary Methods for the Determination of Drugs in Body Fluids', *J. Chromatogr.*, **638**, 335–342 (1993).
44. Z.K. Shihabi, 'Sample Stacking by Acetonitrile–Salt Mixtures', *J. Capil. Electrophor.*, **2**, 267–271 (1995).
45. M.A. Friedberg, M. Hinsdale, Z.K. Shihabi, 'Effect of pH and Ions in the Sample on Stacking in Capillary Electrophoresis', *J. Chromatogr. A*, **781**, 35–42 (1997).
46. K. Khrapko, J.S. Hanekamp, W.G. Thilly, A. Belenkii, F. Foret, B.L. Karger, 'Constant Denaturant Capillary Electrophoresis (CDCE): a High Resolution Approach to Mutational Analysis', *Nucleic Acids Res.*, **22**, 364–369 (1994).
47. R. Clark, J.A. Katzmann, R.A. Kyle, M. Fleisher, J.P. Landers, 'Differential Diagnosis of Gammopathies by Capillary Electrophoresis and Immunosubtraction: Analysis of Serum Samples Problematic by Agarose Gel Electrophoresis', *Electrophoresis*, **19**, 2479–2484 (1998).
48. J.A. Katzmann, R. Clark, E. Sanders, J.P. Landers, R.A. Kyle, 'Prospective Study of Serum Protein Capillary Zone Electrophoresis and Immunotyping of Monoclonal Proteins by Immunosubtraction', *Am. J. Clin. Pathol.*, **110**, 503–509 (1998).
49. P. Davidsson, R. Jahn, J. Bergquist, R. Ekman, K. Blennow, 'Synaptotagmin, a Synaptic Vesicle Protein, is Present in Human Cerebrospinal Fluid: a New Biochemical Marker for Synaptic Pathology in Alzheimer Disease?', *Mol. Chem. Neuropathol.*, **27**, 195–210 (1996).
50. J.A. Katzmann, R. Clark, E. Wiegert, E. Sanders, R.P. Oda, R.A. Kyle, C. Namyst-Goldberg, J.P. Landers, 'Identification of Monoclonal Proteins in Serum: a Quantitative Comparison of Acetate, Agarose Gel, and Capillary Electrophoresis', *Electrophoresis*, **18**, 1755–1780 (1997).
51. C.R. Jolliff, C.R. Blessum, 'Comparison of Serum Protein Electrophoresis by Agarose Gel and Capillary Zone Electrophoresis in a Clinical Setting', *Electrophoresis*, **18**, 1781–1784 (1997).
52. J.M. Hempe, J.N. Granger, R.D. Craver, 'Capillary Isoelectric Focusing of Hemoglobin Variants in the Pediatric Clinical Laboratory', *Electrophoresis*, **18**, 1785–1795 (1997).
53. G.L. Klein, C.R. Joliff, 'Capillary Electrophoresis for the Routine Clinical Laboratory', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 419–458, 1993.
54. G. Schmitz, C. Mollers, V. Richter, 'Analytical Capillary Isotachophoresis of Human Serum Lipoproteins', *Electrophoresis*, **18**, 1807–1813 (1997).
55. R.P. Oda, R. Prasad, R.L. Stout, D. Coffin, W.P. Patton, D.L. Kraft, J.F. O'Brien, J.P. Landers, 'Capillary Electrophoresis-based Separation of Transferrin Sialoforms in Patients with Carbohydrate-deficient Glycoprotein Syndrome', *Electrophoresis*, **18**, 1819–1826 (1997).
56. M.A. Jenkins, 'Clinical Application of Capillary Electrophoresis to Unconcentrated Human Urine Proteins', *Electrophoresis*, **18**, 1842–1846 (1997).
57. M.A. Friedberg, Z.K. Shihabi, 'Urine Protein Analysis by Capillary Electrophoresis', *Electrophoresis*, **18**, 1836–1841 (1997).
58. T. Hyoetylaeinen, H. Siren, M-L. Riekkola, 'Determination of Morphine Analogs, Caffeine and Amphetamine in Biological Fluids by Capillary Electrophoresis with the Marker Technique', *J. Chromatogr., A*, **735**, 439–447 (1996).
59. F. Tagliaro, C. Poiesi, R. Aiello, R. Dorizzi, S. Ghielmi, M. Marigo, 'Capillary Electrophoresis for the Investigation of Illicit Drugs in Hair: Determination of Cocaine and Morphine', *J. Chromatogr.*, **638**, 303–309 (1993).
60. F. Tagliaro, C. Antonioli, S. Moretto, S. Archetti, S. Ghielmi, M. Marigo, 'High-sensitivity Low-cost Methods for Determination of Cocaine in Hair: High-performance Liquid Chromatography and Capillary Electrophoresis', *Forensic Sci. Int.*, **63**, 227–238 (1993).
61. J. Cai, J. Henion, 'Elucidation of LSD In Vitro Metabolism by Liquid Chromatography and Capillary Electrophoresis Coupled with Tandem Mass Spectrometry', *J. Anal. Toxicol.*, **20**, 27–37 (1996).
62. P. Wernly, W. Thormann, D. Bourquin, R. Brenneisen, 'Determination of Morphine-3-glucuronide in Human Urine by Capillary Zone Electrophoresis and Micellar

Electrokinetic Capillary Chromatography', *J. Chromatogr., B*, **616**, 305–310 (1993).
63. E. Gonzalez, R. Montes, J.J. Laserna, 'Pulsed-laser-fluorescence Detection in Capillary Zone Electrophoresis of Some Banned Substances in Sport', *Anal. Chim. Acta*, **282**, 687–693 (1993).
64. D. Perrett, G.A. Ross, 'Rapid Determination of Drugs in Biofluids by Capillary Electrophoresis. Measurement of Antipyrine in Saliva for Pharmacokinetic Studies', *J. Chromatogr., A*, **700**, 179–186 (1995).
65. Z.K. Shihabi, M.E. Hinsdale, 'Analysis of Ibuprofen in Serum by Capillary Electrophoresis', *J. Chromatogr., B*, **683**, 115–118 (1996).
66. A.E. Ashcroft, H.J. Major, S. Lowes, I.D. Wilson, 'Identification of Non-steroidal Anti-inflammatory Drugs and Their Metabolites in Solid Phase Extracts of Human Urine Using Capillary Electrophoresis/Mass Spectrometry', *Anal. Proc.*, **32**, 459–462 (1995).
67. S. Molteni, J. Caslavska, D. Alleman, W. Thormann, 'Determination of Methadone and its Primary Metabolite in Human Urine by Capillary Electrophoretic Techniques', *J. Chromatogr., B*, **658**, 355–367 (1994).
68. J. Caslavska, E. Gassmann, W. Thormann, 'Modification of a Tunable UV-Visible Capillary Electrophoresis Detector for Simultaneous Absorbance and Fluorescence Detection: Profiling of Body Fluids for Drugs and Endogenous Compounds', *J. Chromatogr., A*, **709**, 147–156 (1995).
69. Z. Yu, Y. Chen, L. Chang, 'Capillary Electrophoretic Separation and Detection of Common Opiates', *Fenxi Huaxue*, **24**, 444–446 (1996).
70. D.P. Bogan, B. Deasy, R. O'Kennedy, M.R. Smyth, U. Fuhr, 'Determination of Free and Total 7-Hydroxycoumarin in Urine and Serum by Capillary Electrophoresis', *J. Chromatogr., B*, **663**, 371–378 (1995).
71. B. Deasy, D.P. Bogan, M.R. Smyth, R. O'Kennedy, U. Fuhr, 'Study of Coumarin Metabolism by Human Liver Microsomes Using Capillary Electrophoresis', *Anal. Chim. Acta*, **310**, 101–107 (1995).
72. C. Desiderio, S. Fanali, 'Stereoselective Analysis of Pharmaceutical Compounds in Biological Fluids Using Capillary Electrophoresis', *Boll. Chim. Farm.*, **134**, 541–546 (1995).
73. L.L. Garcia, Z.K. Shihabi, K. Oles, 'Determination of Gabapentin in Serum by Capillary Electrophoresis', *J. Chromatogr., B*, **669**, 157–162 (1995).
74. G. Hempel, G. Blaschke, 'Enantioselective Determination of Zopiclone and its Metabolites in Urine by Capillary Electrophoresis', *J. Chromatogr., B*, **675**, 139–146 (1996).
75. A. Hiraoka, I. Miura, J. Akai, I. Tominaga, M. Hattori, S. Machida, T. Arato, 'Capillary Electrophoresis as an Aid in Therapeutic Monitoring of the Serum Antiepileptic Drug Concentrations', *Seibutsu Shiryo Bunseki*, **17**, 274–279 (1994).
76. R.B. Taylor, R.G. Reid, 'Analysis of Basic Antimalarial Drugs by CZE. Part 2. Validation and Application to Bioanalysis', *J. Pharm. Biomed. Anal.*, **13**, 21–26 (1995).
77. V.M. Okun, O.V. Aak, V.Y. Kozlov, 'Determination of the Anticancer Drug Prospidin in Human Tissue by High-performance Capillary Electrophoresis', *J. Chromatogr., B*, **675**, 313–319 (1996).
78. G. Nouadje, H. Rubie, E. Chatelut, P. Canal, M. Nertz, P. Puig, F. Couderc, 'Child Cerebrospinal Fluid Analysis by Capillary Electrophoresis and Laser-induced Fluorescence Detection', *J. Chromatogr., A*, **717**, 293–298 (1995).
79. J. Sadecka, J. Polonsky, 'Determination of Some Cardiovascular Drugs in Serum and Urine by Capillary Electrophoresis', *J. Chromatogr., A*, **735**, 403–408 (1996).
80. F. Li, S.F. Cooper, S.R. Mikkelsen, 'Enantioselective Determination of Oxprenolol and its Metabolites in Human Urine by Cyclodextrin-modified Capillary Zone Electrophoresis', *J. Chromatogr., B*, **674**, 277–285 (1995).
81. C. Coors, H-G. Schulz, F. Stache, 'Development and Validation of a Bioanalytical Method for the Quantification of Diltiazem and Desacetyldiltiazem in Plasma by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **717**, 235–243 (1995).
82. Y-Y. Zong, C-T. Che, 'Determination of Strychnine and Brucine by Capillary Zone Electrophoresis', *Planta Med.*, **61**, 456–458 (1995).
83. Z-Y. Zhang, M.J. Fasco, L.S. Kaminsky, 'Determination of Theophylline and its Metabolites in Rat Liver Microsomes and Human Urine by Capillary Electrophoresis', *J. Chromatogr., B*, **665**, 201–208 (1995).
84. J. Duan, B. Yan, 'Determination of Theophylline in Plasma by High Performance Capillary Electrophoresis', *Zhongguo Yaoxue Zazhi*, **31**, 99–101 (1996).
85. J. Luksa, D. Josic, 'Determination of Cimetidine in Human Plasma by Free Capillary Zone Electrophoresis', *J. Chromatogr., B*, **667**, 321–327 (1995).
86. H. Soini, M-L. Riekkola, M.V. Novotny, 'Mixed Polymer Networks in the Direct Analysis of Pharmaceuticals in Urine by Capillary Electrophoresis', *J. Chromatogr., A*, **680**, 623–634 (1994).
87. G. Hempel, G. Blaschke, 'Direct Determination of Zolpidem and its Main Metabolites in Urine Using Capillary Electrophoresis with Laser-induced Fluorescence Detection', *J. Chromatogr., B*, **675**, 131–137 (1996).
88. S.P.D. Lalljie, J. Vindevogel, P. Sandra, 'Capillary Electrophoresis as a Fast Screening Method for the Determination of Clenbuterol in Urine', *J. Capil. Electrophor.*, **1**, 241–245 (1994).
89. M. Chicharro, A. Zapardiel, E. Bermejo, J.A.P. Lopez, L. Hernandez, 'Direct Determination of Sympathomimetic Drugs in Human Urine by Capillary Zone Electrophoresis', *Quim. Anal. (Barcelona)*, **13**, 134–137 (1994).

90. C. Gelfi, P.G. Righetti, F. Leoncini, V. Brunelli, P. Carrera, M. Ferrari, 'CAG Triplet Analysis in Families with Androgen Insensitivity Syndrome by Capillary Electrophoresis in Polymer Networks', *J. Chromatogr., A*, **706**, 463 (1995).
91. M. Koch, R. Lehmann, M. Pfohl, W. Voelter, H. Liebich, 'Analysis of the Deletion/Insertion Polymorphism of the Angiotensin I-converting Enzyme Gene by Capillary Electrophoresis', *Clin. Chim. Acta*, **248**, 197–203 (1996).
92. Y. Baba, R. Tomisaki, C. Sumita, I. Morimoto, S. Sugita, M. Tsuhako, T. Miki, T. Ogihara, 'Rapid Typing of Variable Number of Tandem Repeat Locus in the Human Apolipoprotein B Gene for DNA Diagnosis of Heart Disease by Polymerase Chain Reaction and Capillary Electrophoresis', *Electrophoresis*, **16**, 1437–1440 (1995).
93. Y. Baba, R. Tomisaki, C. Sumita, M. Tsuhako, T Miki, T. Ogihara, 'High-resolution Separation of PCR Product and Gene Diagnosis by Capillary Gel Electrophoresis', *Biomed. Chromatogr.*, **8**, 291–293 (1994).
94. A. Schlenck, S. Visvikis, M. O'Kane, G. Siest, 'High-resolution Separation of PCR Product and Gene Diagnosis by Capillary Gel Electrophoresis', *Biomed. Chromatogr.*, **10**, 48–50 (1996).
95. M. Oto, T. Suehiro, Y. Akiyama, Y. Yuasa, 'Microsatellite Instability in Cancer Identified by Non-gel Sieving Capillary Electrophoresis', *Clin. Chem.*, **41**, 482–483 (1995).
96. C. Gelfi, A. Orsi, P.G. Righetti, M. Zanussi, P. Carrera, M. Ferrari, 'Capillary Zone Electrophoresis in Polymer Networks of Polymerase Chain Reaction-amplified Oligonucleotides: the Case of Congenital Adrenal Hyperplasia', *J. Chromatogr., B*, **675**, 201–205 (1996).
97. P. Carrera, A. Barbieri, M. Ferrari, P.G. Righetti, M. Perego, C. Gelfi, 'Rapid Detection of 21-Hydroxylase Deficiency Mutations by Allele-specific In Vitro Amplification and Capillary Zone Electrophoresis', *Clin. Chem.*, **43**, 2121–2127 (1997).
98. C. Gelfi, P.G. Righetti, C. Magnani, L. Cremonesi, M. Ferrari, 'Simultaneous Detection of D/F508, G542X, N1303K and 1717-1G->A Mutations in Cystic Fibrosis by Capillary Electrophoresis in Polymer Networks', *Clin. Chim. Acta*, **229**, 181–189 (1994).
99. C. Gelfi, P.G. Righetti, V. Brancolini, L. Cremonesi, M. Ferrari, 'Capillary Electrophoresis in Polymer Networks for Analysis of PCR Products: Detection of ΔF508 Mutation in Cystic Fibrosis', *Clin. Chem.*, **40**, 1603–1605 (1994).
100. K.R. Mitchelson, J. Cheng, 'Point Mutation Screening by High-performance Capillary Electrophoresis', *J. Capil. Electrophor.*, **2**, 137–143 (1995).
101. C. Bory, C. Chantin, D. Bozon, 'Capillary Electrophoretic Analysis of DNA Restriction-fragments and PCR Products for Polymorphism and Mutation Studies in Cystic Fibrosis and Gaucher's Disease', *J. Pharm. Biomed. Analysis*, **13**, 511–514 (1995).
102. C. Gelfi, A. Orsi, P.G. Righetti, V. Brancolini, L. Cremonesi, M. Ferrari, 'Capillary Zone Electrophoresis of Polymerase Chain Reaction-amplified DNA Fragments in Polymer Networks: The Case of GATT Microsatellites in Cystic Fibrosis', *Electrophoresis*, **15**, 640–643 (1994).
103. C. Gelfi, P.G. Righetti, L. Cremonesi, M. Ferrari, 'Determination of Point Mutations by Capillary Electrophoresis in Liquid Polymers in Temporal Thermal Gradients', *Electrophoresis*, **15**, 1506–1511 (1994).
104. C. Gelfi, L. Cremonesi, M. Ferrari, P.G. Righetti, 'Temperature-programmed Capillary Electrophoresis for Detection of DNA Point Mutation', *BioTechniques*, **21**, 926–932 (1996).
105. C. Gelfi, A. Orsi, F. Leoncini, P.G. Righetti, I. Spiga, P. Carrera, M. Ferrari, 'Amplification of 18 Dystrophin Gene Exons in DMD/BMD Patients: Simultaneous Resolution by Capillary Electrophoresis in Sieving Liquid Polymers', *BioTechniques*, **19**, 254–256 (1995).
106. K.J. Ulfelder, H.E. Schwartz, J.M. Hall, F.I. Sunzery, 'Restriction Fragment Length Polymorphism Analysis of ERBB2 Oncogene by Capillary Electrophoresis', *Anal. Biochem.*, **200**, 260–267 (1992).
107. L.T.F. Van de Locht, A.W.H.M. Kupers, B.W. Verbruggen, P.C.M. Linssen, I.R.O. Novakova, E.J.B.M. Mensink, 'Semi-automated Detection of the Factor V Mutation by Allele Specific Amplification and Capillary Electrophoresis', *Thromb. Haemostasis*, **74**, 1276–1279 (1995).
108. M. Nesi, P.G. Righetti, M.C. Patrosso, A. Ferlini, M. Chiari, 'Capillary Electrophoresis of Polymerase Chain Reaction-amplified Products in Polymer Networks: The Case of Kennedy's Disease', *Electrophoresis*, **15**, 644–646 (1994).
109. C.E. Mitchell, S.A. Belinsky, J.F. Lechner, 'Detection and Quantitation of Mutant K-ras Codon 12 Restriction Fragments by Capillary Electrophoresis', *Anal. Biochem.*, **224**, 148–153 (1995).
110. H. Arakawa, K. Uetanaka, M. Maeda, A. Tsuji, Y. Matsubara, K. Narisawa, 'Analysis of Polymerase Chain Reaction-products by Capillary Electrophoresis with Laser-induced Fluorescence Detection and its Application to the Diagnosis of Medium-chain Acyl-coenzyme A Dehydrogenase Deficiency', *J. Chromatogr., A*, **680**, 517–523 (1994).
111. H. Arakawa, S. Nakashiro, M. Maeda, A. Tsuji, 'Analysis of Single-strand DNA Conformation Polymorphism by Capillary Electrophoresis', *J. Chromatogr., A*, **722**, 359–368 (1996).
112. A.W.H.M. Kuypers, P.M.W. Willem, M.J. Van Der Schrans, P.C.M. Linssen, H.M.C. Wessels, C.H.M.M. de Bruijn, F.M. Everaerts, E.J.B.M. Mensk, 'Determination of Point Mutations in DNA Using Capillary Electrophoresis in a Polymer Network', *J. Chromatogr., B*, **621**, 149–156 (1993).
113. K. Katsuragi, K. Kitagishi, W. Chiba, S. Ikeda, M. Kinoshita, 'Fluorescence-based Polymerase Chain

Reaction-single-strand Conformation Polymorphism Analysis of p53 Gene by Capillary Electrophoresis', *J. Chromatogr., A*, **744**, 311–320 (1996).

114. C. Gelfi, P.G. Righetti, M. Travi, S. Fattore, 'Temperature-programmed Capillary Electrophoresis for the Analysis of High-melting Point Mutants in Thalassemias', *Electrophoresis*, **18**, 724–731 (1997).
115. J. Ren, X. Deng, Y. Cao, K. Cao, 'Analysis of DNA Fragments and Polymerase Chain Reaction Products from the Tx Gene by Capillary Electrophoresis with a Laser-induced Fluorescence Detector Using No-gel Sieving Media', *Anal. Biochem.*, **233**, 246–249 (1996).
116. B.R. McCord, D.L. McLure, J.M. Jung, 'Capillary Electrophoresis of Polymerase Chain Reaction-amplified DNA Using Fluorescence Detection with an Intercalating Dye', *J. Chromatogr., A*, **652**, 75–82 (1993).
117. C. Gelfi, F. Leoncini, P.G. Righetti, L. Cremonesi, A.M. di Blasio, C. Carniti, M. Vignali, 'Separation and Quantitation of Reverse Transcriptase Polymerase Chain Reaction Fragments of Basic Fibroblast Growth Factor by Capillary Electrophoresis in Polymer Networks', *Electrophoresis*, **16**, 780–783 (1995).
118. C. Gelfi, G. Cossu, P. Carta, M. Serra, P.G. Righetti, 'Gene Dosage in Capillary Electrophoresis: Pre-natal Diagnosis of Down's Syndrome', *J. Chromatogr., A*, **718**, 405–412 (1995).
119. A.W.H.M. Kuypers, J.P.P. Meijrink, T.F.C.M. Smetsers, P.C.M. Linssen, E.J.B.M. Mensik, 'Quantitative Analysis of DNA Aberrations Amplified by Competitive Polymerase Chain Reaction Using Capillary Electrophoresis', *J. Chromatogr., B*, **660**, 271–277 (1994).
120. G. Cossu, A. Angius, C. Gelfi, P.G. Righetti, 'Rh D/d Genotyping by Quantitative Polymerase Chain Reaction and Capillary Zone Electrophoresis', *Electrophoresis*, **17**, 1911–1915 (1996).
121. B. Lin, H. Diao, Y. Zhang, X. Chu, 'Determination of Polymerase Chain Reaction-amplified Products of Hepatitis B Virus with Non-gel Sieving Capillary Electrophoresis', *Fenxi Huaxue*, **22**, 449–452 (1994).
122. E.F. Rossomando, L. White, K.J. Ulfelder, 'Capillary Electrophoresis: Separation and Quantification of Reverse Transcriptase Polymerase Chain Reaction Products from Polio Virus', *J. Chromatogr., B*, **656**, 159 (1994).
123. H.E. Schwartz, K. Ulfelder, F.J. Sunzeri, M.P. Busch, R.G. Brownlee, 'Analysis of DNA Restriction Fragments and Polymerase Chain Reaction Products Towards Detection of the AIDS (HIV-1) Virus in Blood', *J. Chromatogr.*, **559**, 267–283 (1991).
124. W. Lu, D-S. Han, J. Yuan, J-M. Andrieu, 'Multi-target PCR Analysis by Capillary Electrophoresis and Laser-induced Fluorescence', *Nature*, **368**, 269–271 (1994).
125. S.J. Williams, C. Schwer, A.S.M. Krishnarao, C. Heid, B.L. Karger, P.M. Williams, 'Quantitative Competitive Polymerase Chain Reaction: Analysis of Amplified Products of the HIV-1 gag Gene by Capillary Electrophoresis with Laser-induced Fluorescence Detection', *Anal. Biochem.*, **236**, 146–152 (1996).
126. T.A. Felmlee, P.S. Mitchell, K.J. Ulfelder, D.A. Persing, J.P. Landers, 'Capillary Electrophoresis for the Post-amplification Detection of a Hepatitis C Virus-specific DNA Product in Human Serum', *J. Capil. Electrophor.*, **2**, 125–130 (1995).
127. J.M. Butler, B.R. McCord, J.M. Jung, M.R. Wilson, B. Budowle, R.O. Allen, 'Quantitation of Polymerase Chain Reaction Products by Capillary Electrophoresis Using Laser Fluorescence', *J. Chromatogr., B*, **658**, 271–280 (1994).
128. K. Srinivasan, J.E. Girard, P. Williams, R.K. Roby, V.W. Weedn, S.C. Morris, M.C. Kline, D.J. Reeder, 'Electrophoretic Separations of Polymerase Chain Reaction-amplified DNA Fragments in DNA Typing Using a Capillary Electrophoresis/Laser induced Fluorescence System', *J. Chromatogr.*, **652**, 83–91 (1993).
129. B.R. McCord, J.M. Jung, E.A. Holleran, 'High Resolution Capillary Electrophoresis of Forensic DNA Using a Non-gel Sieving Buffer', *J. Liquid Chromatogr.*, **16**, 1963–1981 (1993).
130. A.R. Isenberg, B.R. McCord, B.W. Koons, B. Budowle, R.O. Allen, 'DNA Typing of a Polymerase Chain Reaction-amplified D1S80/Amelogenin Multiplex Using Capillary Electrophoresis and a Mixed Entangled Polymer Matrix', *Electrophoresis*, **17**, 1505–1511 (1996).
131. K. Srinivasan, S.C. Morris, J.E. Girard, M.C. Kline, D.J. Reeder, 'Enhanced Detection of PCR Products through Use of TOTO and YOYO Intercalating Dyes with Laser-induced Fluorescence/Capillary Electrophoresis', *Appl. Theor. Electrophor.*, **3**, 235–239 (1993).
132. J.M. Butler, B.R. McCord, J.M. Jung, R.O. Allen, 'Rapid Analysis of the Short Tandem Repeat HUMBH01 by Capillary Electrophoresis', *BioTechniques*, **17**, 1062–1070 (1994).
133. P.E. Williams, M.A. Marino, S.A. Del Rio, L.A. Turni, J.M. Devaney, 'Analysis of DNA Restriction Fragments and Polymerase Chain Reaction Products by Capillary Electrophoresis', *J. Chromatogr., A*, **680**, 525 (1994).
134. A. Kloosterman, M. Schans van der, H. Janssen, F. Everaerts, 'Rapid Sex-typing by Fluorescent Based PCR of the X-Y Homologous Amelogenin Gene and Analysis by CGE', *Adv. Forensic Haemogenet.*, **6**, 343–345 (1996).
135. J. Grimshaw, J. Trocha-Grimshaw, W. Fisher, A. Rice, S. Smith, P. Spedding, J. Duffy, R. Mollan, 'Quantitative Analysis of Hyaluronan in Human Synovial Fluid Using Capillary Electrophoresis', *Electrophoresis*, **17**, 396–400 (1996).
136. J. Grimshaw, J. Trocha-Grimshaw, W. Fisher, A. Kane, P. McCarron, A. Rice, P.L. Spedding, J. Duffy, R.A.B. Mollan, 'An Investigation of Human Synovial Fluid by Capillary Electrophoresis', *J. Capil. Electrophor.*, **2**, 40–45 (1995).

137. M. Gross, B.S. Gathof, P. Koelle, U. Gresser, 'Capillary Electrophoresis for Screening of Adenylosuccinate Lyase Deficiency', *Electrophoresis*, **16**, 1927–1929 (1995).
138. D.B. Marsh, K.L. Nuttall, 'Methylmalonic Acid in Clinical Urine Specimens by Capillary Zone Electrophoresis Using Indirect Photometric Detection', *J. Capil. Electrophor.*, **2**, 63–67 (1995).
139. J. Schneede, P.M. Ueland, 'Application of Capillary Electrophoresis with Laser-induced Fluorescence Detection for Routine Determination of Methylmalonic Acid in Human Serum', *Anal. Chem.*, **67**, 812–819 (1995).
140. J. Sevick, T. Adam, H. Mazacova, 'A Fast and Simple Screening Method for Detection of 2,8-Dihydroxyadenine Urolithiasis by Capillary Zone Electrophoresis', *Clin. Chim. Acta*, **245**, 85–92 (1996).
141. J. Bergquist, S.D. Gilman, A.G. Ewing, R. Ekman, 'Analysis of Human Cerobrospinal Fluid by Capillary Electrophoresis with Laser-induced Fluorescence Detection', *Anal. Chem.*, **66**, 3512–3518 (1994).
142. A. Hiraoka, J. Akai, I. Tominaga, M. Hattori, H. Sasaki, T. Arato, 'Capillary Zone Electrophoretic Determination of Organic Acids in Cerebrospinal Fluid from Patients with Central Nervous System Diseases', *J. Chromatogr., A*, **680**, 243–246 (1994).
143. A. Hiraoka, I. Miura, I. Tominaga, M. Hattori, H. Sasaki, 'Capillary-Zone Electrophoretic Determination of Lactate in Cerebrospinal Fluid from Patients with Central Nervous System Diseases', *Seibutsu Butsuri Kagaku*, **38**, 233–236 (1994).
144. A. Hiraoka, M. Kobayashi, T. Takabuchi, J. Akai, T. Arato, I. Tominaga, 'Capillary-Zone Electrophoretic Determination of Uric Acid in Cerebrospinal Fluid from Patients with Neuropsychiatric Diseases', *Seibutsu Shiryo Bunseki*, **18**, 90–95 (1993).
145. C.M. Jariego, A. Hernanz, 'Determination of Organic Acids by Capillary Electrophoresis in Screening of Organic Acidurias', *Clin. Chem.*, **42**, 477–478 (1996).
146. M. Shirao, R. Furuta, S. Suzuki, H. Nakazawa, S. Fujita, T. Maruyama, 'Determination of Organic Acids in Urine by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **680**, 247–251 (1994).
147. V. Dolnik, 'Capillary Zone Electrophoresis of Pathological Metabolites in Phenylketonuria', *J. Microcolumn Sep.*, **6**, 63–67 (1994).
148. P. Xu, J. Che, H. Shi, Y. Ma, 'Quantitative Determination of Serum Iron in Human Blood by High-performance Capillary Electrophoresis', *J. Chromatogr., B*, **669**, 45–51 (1995).
149. C. Fernandez, G. Egginger, I.W. Wainer, D.K. Lloyd, 'Separation of Testosterone Metabolites in Microsomal Incubates Using a New Capillary Electrophoresis Assay', *J. Chromatogr., B*, **677**, 363–368 (1996).
150. C.J. Petucci, H.L. Kantes, T.G. Strein, H. Veening, 'Capillary Electrophoresis as a Clinical Tool Determination of Organic Anions in Normal and Uremic Serum Using Photodiode-array Detection', *J. Chromatogr., B*, **668**, 241–251 (1995).
151. A.J. Ji, M.F. Nunez, D. Machacek, J.E. Ferguson, M.F. Iossi, P.C. Kao, J.P. Landers, 'Separation of Urinary Estrogens by Micellar Electrokinetic Chromatography', *J. Chromatogr., B*, **669**, 15–26 (1995).
152. H. Shi, Y. Ma, J.H. Humphrey, N.E. Craft, 'Determination of Vitamin A in Dried Human Blood Spots by High-performance Capillary Electrophoresis with Laser-excited Fluorescence Detection', *J. Chromatogr., B*, **665**, 89–96 (1995).
153. C. Bory, C. Chantin, R. Boulieu, 'Comparison of Capillary Electrophoretic and Liquid Chromatographic Determination of Hypoxanthine and Xanthine for the Diagnosis of Xanthinuria', *J. Chromatogr., A*, **730**, 329–331 (1996).
154. H-T. Chang, E.S. Yeung, 'Determination of Catecholamines in Single Adrenal Medullary Cells by Capillary Electrophoresis and Laser-induced Native Fluorescence', *Anal. Chem.*, **67**, 1079–1083 (1995).
155. G. Piccoli, M. Fiorani, B. Biagiarelli, F. Palma, L. Potenza, A. Amicucci, V. Stocchi, 'Simultaneous High-performance Capillary Electrophoretic Determination of Reduced and Oxidized Glutathione in Red Blood Cells in the Femtomole Range', *J. Chromatogr., A*, **676**, 239–246 (1994).
156. Q. Li, E.S. Yeung, 'Contamination Control in Capillary Electrophoresis and Quantitative Measurement of Potassium and Sodium in Single Human Erythrocytes', *J. Capil. Electrophor.*, **1**, 55–61 (1994).
157. S.R. Rabel, J.F. Stobaugh, R. Trueworthy, 'Determination of Intracellular Levels of 6-Mercaptopurine Metabolites in Erythrocytes Utilizing Capillary Electrophoresis with Laser-induced Fluorescence Detection', *Anal. Biochem.*, **224**, 315–322 (1995).
158. S.J. Lillard, E.S. Yeung, R.M.A. Lautamo, D.T. Mao, 'Separation of Hemoglobin Variants in Single Human Erythrocytes by Capillary Electrophoresis with Laser-induced Native Fluorescence Detection', *J. Chromatogr., A*, **718**, 397–404 (1995).
159. S.A. Hofstadler, J.C. Severs, R.D. Smith, F.D. Swanek, A.G. Ewing, 'Analysis of Single Cells with Capillary Electrophoresis Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 919–922 (1996).
160. Y. Takemura, H. Kobayashi, S. Sekiguchi, 'Separation of Methotrexate-polyglutamates by Capillary Electrophoresis and its Application to the Measurement of γ-Glutamyl Hydrolase Activity in Human Leukemia Cells in Culture', *Rinsho Byori*, **44**, 51–56 (1996).
161. J. Bergquist, A. Tarkowski, R. Ekman, A. Ewing, 'Discovery of Endogenous Catecholamines in Lymphocytes and Evidence for Catecholamine Regulation of Lymphocyte Function via an Autocrine Loop', *Proc. Natl. Acad. Sci. USA*, **91**, 12912–12916 (1994).

162. Q. Xue, E.S. Yeung, 'Determination of Lactate Dehydrogenase Isoenzymes in Single Lymphocytes from Normal and Leukemia Cell Lines', *J. Chromatogr., B*, **677**, 233–240 (1996).

163. S.D. Gilman, A.G. Ewing, 'Analysis of Single Cells by Capillary Electrophoresis with On-column Derivatization and Laser-induced Fluorescence Detection', *Anal. Chem.*, **67**, 58–64 (1995).

Diagnostic Hematology

J. Hubbard
Texas Technical University, Lubbock, USA

1 Introduction to Hematology and Hemostasis 1
2 Blood Coagulation 1
2.1 Initiating Reactions 1
2.2 Intermediate Reactions 2
2.3 Clot Formation 3
2.4 Coagulation Instrumentation 3
2.5 Laboratory Screening Tests of Coagulation 4
3 Overview of Diagnostic Red Cell Pathophysiology 4
3.1 Introduction to Anemias 4
3.2 Stem Cell Disorders 6
3.3 Heme Disorders 6
3.4 Globin Chain Disorders 7
3.5 Disorders of Impaired Production 9
3.6 Hemolytic Disorders 10
4 The Diagnosis of Leukemias 11
4.1 Common Characteristics of Leukemias 11
4.2 Differential Diagnosis of the Leukemia 11
4.3 Differentiation by Cytochemical Stains 14
4.4 Differentiation by Immunochemical Typing 15
5 Electronic Cell Counters 16
5.1 Principle of Impedance Counters 16
5.2 Measured Parameters and Methods 16
5.3 Flow Cytometry 18
Abbreviations and Acronyms 19
Related Articles 20
References 20

Hematology, in its pure form, is the study of the red blood cells (RBCs), white blood cells (WBCs), and platelets during normal and pathological conditions. Quantitative and qualitative evaluation and interpretation of the formed elements of the blood require a whole blood sample to be counted using automated cell counters based on electrical impedance or laser light scatter techniques. A wide range of disease types can be encountered, such as hereditary, nutritional, metabolic, traumatic, infections and inflammatory conditions, and hormonal, immunological and neoplastic disorders. All these disorders can result in changes to the blood cell profile, which can be used to aid in diagnosis. Most important to the diagnosis of hematological diseases is an in-depth understanding of the different causes of anemia and the pathogenesis of the leukemias.

1 INTRODUCTION TO HEMATOLOGY AND HEMOSTASIS

Although blood is a liquid, it can rightly be thought of as a tissue that is circulated throughout the body via the vascular system; it comes into contact with almost every cell of the body, except the central nervous system. If blood is drawn into a tube containing an anticoagulant such as ethylenediamine tetraacetate (EDTA) or citrate, clot formation is prevented. When anticoagulated blood separates into layers, the top yellow liquid is referred to as plasma rather than serum; plasma still contains all coagulation factors. Below the plasma is a thin whitish layer known as the ''buffy coat'' which contains the WBCs (leukocytes) and platelets (thrombocytes). The large portion at the bottom of the tube will be the RBCs. In a normal adult male, the proportions would be approximately 52% plasma, 1% buffy coat, and 47% the packed RBC volume [the hematocrit (HCT)].

Plasma is a complex mixture of water, proteins and antibodies, lipids, carbohydrates, electrolytes, clotting factors, enzymes, vitamins, hormones, and trace metals or drugs. The major source of blood production is the bone marrow, although the lymph nodes and spleen contribute some of the WBCs.

The blood has many diverse functions. Blood is very important for natural defense mechanisms, which involve the phagocytic cells, immunoglobulins (i.e. antibodies), and immunological cells (i.e. lymphocytes). In addition, the blood supplies nutrients and oxygen to all tissues, carries the end-products of metabolism to the liver and kidneys for disposal, is important for thermal regulation, and functions in processes of hemostasis. Hemostasis implies that a balance between bleeding, clotting, and cellular elements of blood is maintained and in equilibrium. An abnormal demand for one cell type following external loss, consumption, or destruction is accompanied by a compensatory hyperplasia (increased cell growth) of the marrow.

2 BLOOD COAGULATION

2.1 Initiating Reactions

The vascular system is self-maintained by an interaction of platelets and plasma coagulation factors. First, the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

formation of a platelet plug or ''white thrombus'' occurs rapidly within seconds of vascular damage. A ''red thrombus'' is formed when red cells become entangled in the plug of fibrin polymers. Fibrinolysis (i.e. clot dissolution) occurs when thrombin and fibrin activate clot dissolution mechanisms. These mechanisms confine clotting to a localized area. Plasma clotting factors are closely linked to platelet factors and function. Most plasma coagulation factors circulate as inactive precursor zymogens until activated by molecular rearrangement catalyzed by serine proteases at the site of vascular injury. The sequence of coagulation flows in a cascade or ''waterfall'' fashion (i.e. the activation of one factor leads to the activation of the next, and so on), as seen in Figure 1. There are 13 numbered factors designated by roman numerals (I–XIII) (Table 1). An activated form is indicated by adding the letter ''a'' to the number.

The intrinsic system refers to the path of the coagulation cascade in which factors prekallikrein, HMWK, XII, XI, X, IX, VIII, V, II (prothrombin), and I (fibrinogen) are involved in the formation of a fibrin clot.[1] The laboratory test for this pathway is the APTT. The intrinsic system coagulation cascade is initiated by the activation of factor XII (Hageman factor). Vascular damage will expose negatively charged subendothelial tissue. The inactive zymogen form of factor XII is attracted to the negative charge of the endothelial surface of the damaged blood vessel where its conformation changes to expose its active serine center. The activated form of XII is denoted as XIIa. Factor XIIa will enzymatically activate factor XI (plasma thromboplastin antecedent) to yield factor XIa, which continues the coagulation cascade.

The extrinsic system contact activation begins with the activation of factor VII. Factor III, known as tissue factor, is the primary activator of factor VII to VIIa, which is a potent serine protease. Tissue factor consists of lipoproteins, which are produced in most tissues. Minor activation of factor VII can occur by proteolytic attack from factors XIIa, Xa, IXa, or thrombin. In laboratory testing of the extrinsic system, lipoprotein-rich extracts are added to citrated plasma as the PT reagent to support the activation of factor VII by tissue.

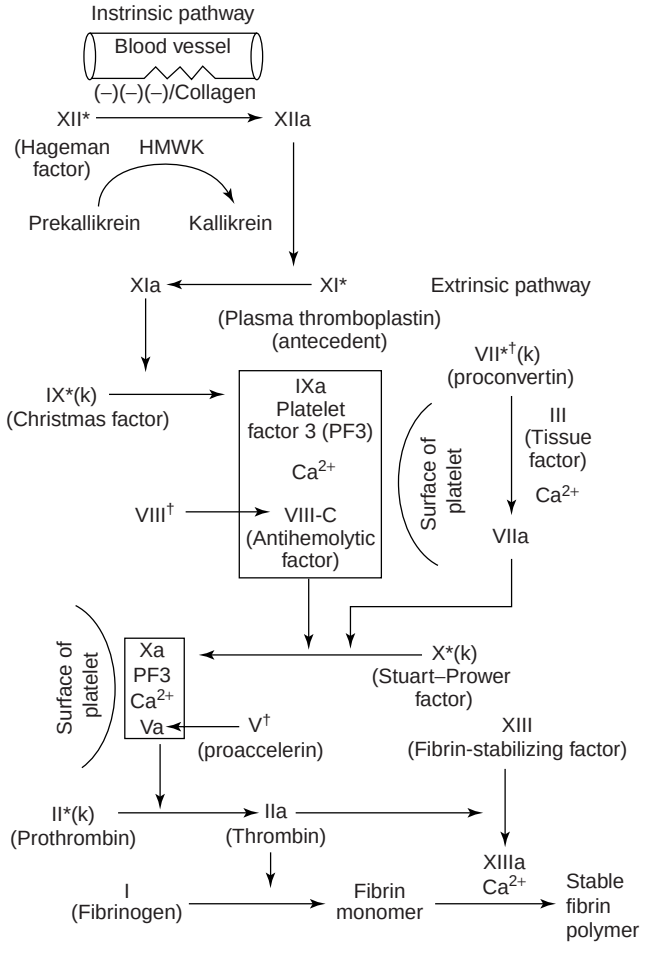

Figure 1 The coagulation cascade. The coagulation factors are represented by roman numerals. Each one is changed to an activated form designated by the letter a. The intrinsic pathway consists of high-molecular-weight kinogen (HMWK), kallikrein, and factors XII, XI, X, VIII, V, II, and I. It is screened by the activated partial thromboplastin time (APTT) test. The extrinsic pathway consists of factors VII, X, V, II and I and is screened by the prothrombin time (PT) test. *a serine protease; †activated by thrombin; (–) negative surface charge of exposed subendothelium k, vitamin K factor.

2.2 Intermediate Reactions

Factor VIIa in the extrinsic pathway will enzymatically alter factor X to yield Xa in the presence of factor III. Factor VIIa has limited ability to activate the conversion of factor IX to its activated form (IXa). Factor IX in the intrinsic pathway is most strongly activated by the direct enzymatic action of factor XIa. This reaction does not require tissue factor lipoprotein as does the extrinsic pathway activation. It does, however, require negatively charged membrane phospholipids and Ca^{2+}. Platelets are the main source in vivo of phospholipid surfaces. In laboratory testing of the intrinsic system, phospholipid extracts are added to citrated plasma as part of the APTT reagent to provide the activation for platelet-supported reactions.

Factor X (Stuart–Prower factor) is activated by two different pathways. In the extrinsic pathway, factor X can be enzymatically activated by factor VIIa, with factor III and Ca^{2+} as cofactors. In the intrinsic pathway, factor X is activated by factor IXa. Factor IXa forms a complex with a platelet phospholipid membrane surface and factor VIII (antihemolytic factor) in the presence of Ca^{2+}.

Factor VIII is a high-molecular-weight complex formed from two subunits. The first, factor VIII: C is also

Table 1 Blood coagulation factors

Factor	Name	Vitamin K dependence	Chemical form	Function
I	Fibrinogen	No	Structural protein	Clot formation
II	Prothrombin	Yes	Serine protease	Activates I
III	Tissue factor	No	Lipoprotein	Activates VII
V	Proaccelerin	No	Cofactor	Activates II
VII	Proconvertin	Yes	Serine protease	Activates X (Ex)
VIII: C	Antihemolytic factor	No	Cofactor	Activates X (In)
VIII: R	von Willebrand's factor (vWF)	No	Cofactor for platelets	Platelet aggregation
IX	Christmas factor	Yes	Serine protease	Activates X (In)
X	Stuart–Prower factor	Yes	Serine protease	Activates II
XI	Plasma thromboplastin antecedent	No	Serine protease	Activates IX
XII	Hageman factor	No	Serine protease	Activates XI
XIII	Fibrin stabilizing factor	No	Transamidase	Stabilizes clot

Factors I, II and VIII: R are usually referred to by their common names. In, intrinsic pathway; Ex, extrinsic pathway.

known as antihemolytic factor. Its formation is genetically controlled by a gene on chromosome X (sex-linked transmission). Factor VIII: C serves as a cofactor in the activation of factor X by IXa; its presence accelerates the reaction rate by 500- to 1000-fold. An inherited deficiency of this factor is known as hemophilia A. The other part of the complex is factor VIII: R, which is also known as von Willebrand's factor. This subunit is synthesized by endothelial cells, megakaryocytes, and platelets and demonstrates autosomal genetic expression. It functions to support the adhesion of platelets to the exposed subendothelial surface of the blood vessel and as a regulator of factor VIII: C. An inherited deficiency of factor VIII: R is known clinically as von Willebrand's disease.

Factor Xa activation begins the section of the cascades that is known as the common pathway because the subsequent enzymatic reactions are shared by both the intrinsic and the extrinsic pathways. Factor Xa enzymatically cleaves the zymogen prothrombin (factor II) to its activated form, thrombin. The combination of phospholipid membrane surface, factors Xa and Va, and Ca^{2+} form the receptor complex known as thrombomodulin-protein C/protein S, which supports the enzymatic conversion of prothrombin to the active enzyme thrombin. The activated form factor V (proaccelerin) is a cofactor for factor Xa activation of prothrombin. Factor V is converted to its active form by thrombin.

2.3 Clot Formation

Thrombin enzymatically converts fibrinogen to fibrin.[2] Fibrinogen has the highest plasma concentration of any clotting factor, with a normal range of 150–400 mg dL^{-1}. The molecule is produced in liver. The fibrinogen monomer consists of two identical subunits bound together to give a symmetrical structure. Three nodular domains in the fibrinogen molecule have been identified as two identical D-regions at the C-terminal ends, and a central E-domain at the N-terminus. Thrombin enzymatically activates the fibrinogen monomer by splitting off the fibrinopeptides Aα and Bβ from the N-terminus in the E-domain. The thrombin-exposed N-terminus peptides in the E-domain react noncovalently by electrostatic forces with polar D-domain regions of adjacent fibrin molecules to form a polymer structure. Formation of a fibrin polymer is the end-point detected in the majority of in vitro clotting time tests. Clot stabilization is achieved by factor XIIIa (fibrin stabilizing factor) catalyzed formation of covalent bonds between chains of adjacent fibrin molecules.

2.4 Coagulation Instrumentation

Most modern instruments are based upon the principles of photo-optics. A less-popular alternative is based upon mechanical clot detection by electrical impedance. All coagulation instrumentation is designed to detect the final fibrin clot formed as the result of fibrinogen's activation by fibrin. The time it takes for a clot to form after an in vitro coagulation activator is added to the plasma is measured as the clotting time. The second is the basic unit for reporting most coagulation tests.

The photo-optical method is based upon clot detection in a cuvette by measuring the degree of change in plasma optical density. A controlled beam of light passes through the plasma/reagent mixture in a reaction cuvette.[3] The light beam leaving the cuvette is detected by a photodetector that measures light intensity. As a solid clot forms in the cuvette, the optical density of the plasma mixture increases and reduces the intensity of the light detected on the far side of the cuvette. This increase in optical density indicates the end-point, and an electrical signal is sent to stop the timing mechanism, thus providing the exact number of seconds it takes for clot formation.

Mechanical semiautomatic clotting devices use electrical impedance to detect clot formation. These instruments have two probes that are suspended in a reaction cup. One probe moves to detect the clot while the other probe is stationary. A current is applied to the stationary probe during the timing sequence. Plasma and coagulation-activating reagents are added to the cup and the timer and rotating probe are activated. When a clot forms between the two probes, the current passes through both probes, completing the circuit and stopping the current.

2.5 Laboratory Screening Tests of Coagulation

All coagulation testing is dependent upon the quality and freshness of the plasma specimen obtained. Whole blood anticoagulated with sodium citrate is the specimen of choice. A blood to citrate ratio of 9 : 1 is very important to ensure accurate testing. Basic coagulation testing is designed to test for factor deficiencies in both the intrinsic and extrinsic coagulation pathways.[4]

The PT tests for extrinsic pathway deficiencies in factors VII, X, V, II, and I. A lipoprotein tissue extract from brain or lung tissue serves as the reagent source of tissue factor. An excess of calcium chloride is also added to the PT reagent. The calcium-enriched lipoprotein reagent is added to a sample of plasma, and the time for fibrin clot formation is measured. The reference range for the PT assay is approximately 11–13 s, but it is important for each laboratory to establish its own range. A value for the international reference ratio (INR) should be included for patients taking coumarin anticoagulants. By definition, the INR is the PT ratio that reflects the results that would have been obtained if the World Health Organization (WHO) international reference preparation (IRP) thromboplastin had been used to perform the test. The specific purpose of the INR is to report results for patients who are stabilized on oral coumarin anticoagulants. Patients who are not taking oral anticoagulants and have normal coagulation should have low PT values, in the range of the INR. When the INR is correctly used for monitoring a patient's oral anticoagulant therapy, the physician gives a standard dose of coumarin to achieve a target INR between 2.0 and 3.0.

The APTT tests for intrinsic pathway deficiencies in prekallikrein, HMWK, and factors XII, XI, IX, X, VIII, V, II, and I. A phospholipid-rich preparation is used as a platelet membrane substitute. An activator such as kaolin, ellagic acid, or celite is also added to the APTT reagent to provide the negative surface charge needed to activate factor XII and prekallikrein. Calcium chloride is used as an additional reagent to initiate the clotting. A citrated plasma sample is preincubated with the phospholipid/activator reagent to activate contact activation factors in the intrinsic pathway. Following incubation, calcium chloride reagent is added as a separate reagent to initiate the clotting cascade. The time it takes fibrin clot formation to occur is measured. The APPT assay has an approximate normal range of 20–35 s, but it is important for each laboratory to establish its own range.

The thrombin time (TT) tests for a deficiency or inhibition of fibrinogen. Commercially prepared thrombin reagent is added to citrated plasma and the time to clot formation is measured. Again each laboratory should establish its own reference range but generally this is 10–20 s.

Final confirmation of a factor deficiency and measurement of specific factor levels is carried out with specialized assays. These methods use a test plasma with a known deficiency to titrate and test against patient plasma with unknown factor deficiencies using a standardized curve. Factors can also be assayed with enzyme-linked immunoassay methodology.

3 OVERVIEW OF DIAGNOSTIC RED CELL PATHOPHYSIOLOGY

3.1 Introduction to Anemias

In general, pathophysiologic disorders affecting the formed blood elements can be divided into leukoproliferative disorders, lymphoproliferative disorders, and RBC disorders. Disorders of RBCs can generally be divided between the polycythemias and the anemias. Polycythemias are disorders with an increase in bone marrow and circulating RBCs. They give rise to a high HCT. Anemias are a collection of disorders that present with a decrease in the RBC count and HCT to below normal levels. They are characterized by hemoglobin concentrations reduced to less than 12 g dL^{-1} in males and less than 11 g dL^{-1} in females. Clinical symptoms result from the reduced oxygen-carrying capacity of the blood and give rise to tissue hypoxia. Anemia is caused by impaired RBC production in the marrow, ineffective production by the marrow, acute or chronic blood loss, and/or accelerated RBC destruction in the systemic circulation. Based on these causes, anemias can be classified into five categories: stem cell disorders, DNA disorders, heme and globin disorders, and RBC survival disorders. Figure 2 illustrates the relationship between the anemias in an algorithm for differential diagnosis. Anemias can also be classified according to their common RBC morphology. Morphological classifications include macrocytic, microcytic, and normocytic anemias.[5] RBC indices, provided by the majority of automated cell counters, are useful in the diagnosis, classification, and morphological differentiation of anemias (Table 2). They can also provide useful guidelines in blood smear RBC morphology.

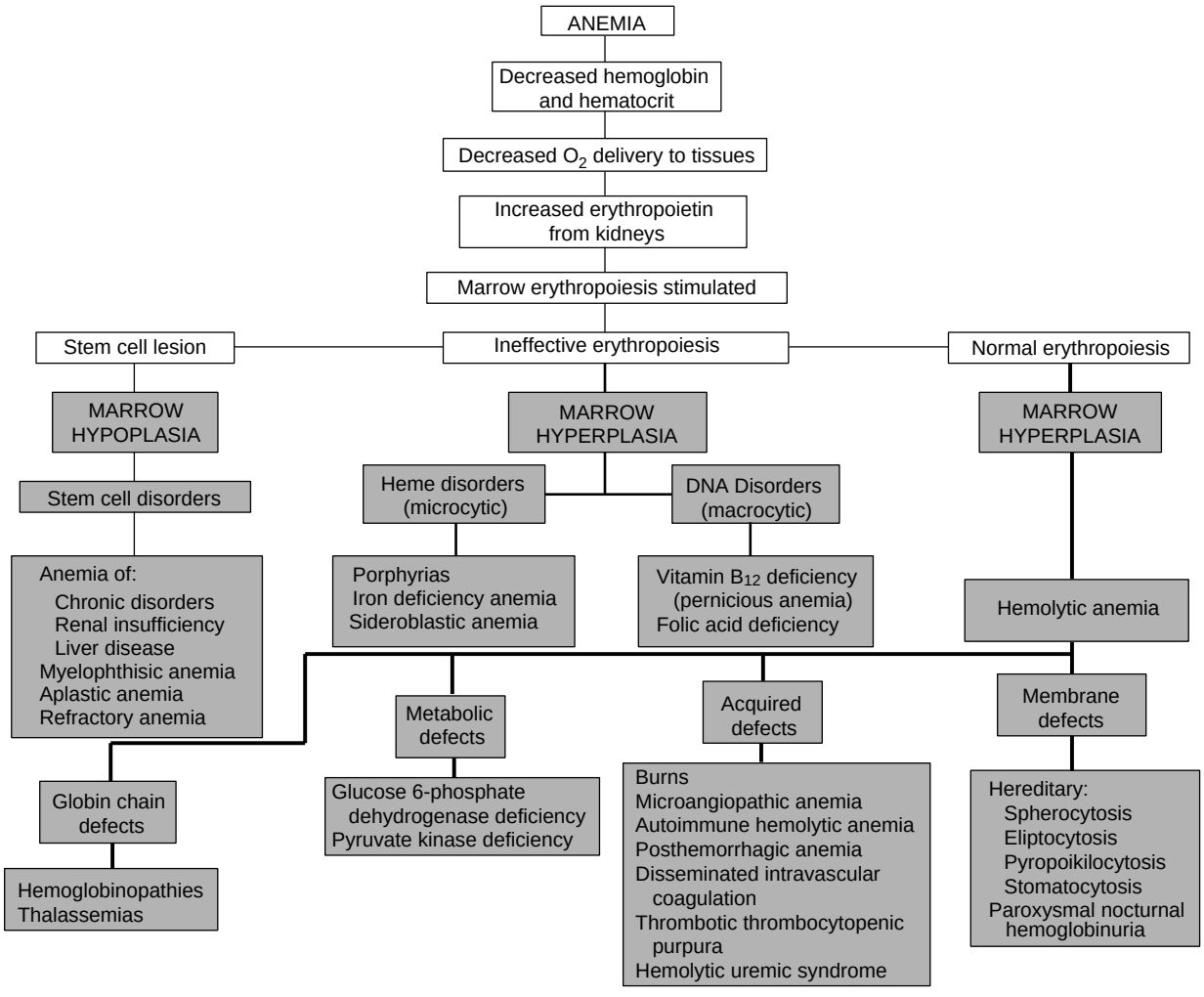

Figure 2 Algorithm summary of the anemias. The anemias are separated and differentiated on the basis of bone marrow erythropoiesis.

Table 2 Classification of anemias and RBC morphology by index range

RBC morphology	MCV (m^3)	MCH (pg)	MCHC (%)	Anemias
Normocytic/normochromic	80–100[a]	27–31[a]	32–36[a]	Acute blood loss Hemolytic anemias Aplastic anemia (early stage) Myelophthisic anemia Stem cell-related anemias
Macrocytic/normochromic	High	High	Normal	Megaloblastic anemia Anemia of liver disease Chronic aplastic anemia Acute hemolytic anemia (with shift reticulocytosis)
Microcytic/normochromic	Low	Normal	Normal	Anemia of chronic inflammation
Microcytic/hypochromic	Low	Low	Low	Iron-deficiency anemia Thalassemia Lead poisoning Porphyrias Sideroblastic anemia

[a] Indices shown as the normal range.
MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume.

3.2 Stem Cell Disorders

Stem cell disorders are RBC disorders that result from an imbalance in marrow multipotential or unipotential stem cell production and maturation. The imbalance is seen as either an increase (e.g. polycythemia) or a decrease (the anemias) in cell production.

Anemia of chronic disorders is usually a mild form of anemia occurring secondary to a chronic inflammatory disorder such as infection, rheumatoid arthritis, or neoplastic disease. RBC production by the marrow is normal but is insufficient to compensate for decreased cell survival. The marrow erythrocyte colony-forming unit is capable of responding to erythropoietin (EPO), but EPO production is low. This anemia results partially from a defect in iron metabolism caused by a block in the secondary iron-storage system. Monocytes and macrophages have a reduced ability to move stored intracellular iron to the erythroid marrow cells.

An anemia of renal insufficiency occurs in patients with end-stage renal disease. There is a general correlation between the severity of the anemia and the degree of elevation of the blood urea nitrogen. When this is greater than $100\,mg\,dL^{-1}$ the HCT is usually less than 30%. The primary cause of the anemia is a decreased production of EPO by the damaged kidney. Ineffective erythropoiesis also results because of an impaired ability of the erythroid stem cells to respond to EPO. Hemolysis is caused by the uremic plasma, which also impairs RBC survival.

Anemia of liver disease is characterized by shortened RBC survival and inadequate RBC production occurring secondary to a chronic liver disease. Red cell morphology is usually macrocytic or normocytic. The patient often has an accompanying splenomegaly with cirrhosis, which further decreases RBC survival. A differential diagnosis must distinguish this anemia from other macrocytic anemias. Anemia of liver disease usually presents with normal WBC count, serum vitamin B_{12} (cyanocobalamin), and folate compared with the megaloblastic anemias. Myelophthisic anemia is an anemia associated with bone marrow infiltration and hyperproliferation by nonerythroid cells. A leukoerythroblastic reaction commonly accompanies the anemia. The usual causes of myelophthisic anemia include metastatic carcinoma, multiple myeloma, leukemia, lymphoma, and lipidoses and storage diseases.

Aplastic anemia is a marrow disorder characterized by a reduction in the number or function of multipotential stem cells, with resulting pancytopenia. The marrow is hypocellular with patchy areas of normal cellularity and increased fat cell infiltration. The diagnosis of severe aplastic anemia is made in pancytopenic patients when at least two of the following three peripheral blood values are found: a WBC count of $<0.5 \times 10^9\,cells\,L^{-1}$; a platelet count of $<20 \times 10^9\,cells\,L^{-1}$; and a reticulocyte count of <1%. Aplastic anemia can be acute and rapidly fatal or it can have a slow onset and a chronic course. Common clinical characteristics are bleeding as a result of thrombocytopenia, increased susceptibility to infections owing to leukopenia, all the symptoms typical of anemia, no splenomegaly, and iron overload from repeated transfusions.

Refractory anemia is an ill-defined group of chronic anemias occurring in people over the age of 50 years. Laboratory features commonly include a normocytic or macrocytic anemia, decreased reticulocyte count, pancytopenia, and a hypercellular marrow with erythroid hypoplasia. Refractory anemia is now classified with the myeloproliferative disorders as one of five myelodysplastic syndromes. The anemia can develop into an acute leukemia with the presence of precursor blast cells in the peripheral circulation.

3.3 Heme Disorders

Heme disorders represent a group of anemias that result from a defect in the synthesis of the heme ring in the mitochondria of developing normoblasts. Disorders of heme synthesis result in a microcytic/hypochromic anemia. Microcytosis is accompanied by a MCV < 80 fL (normal range is 80–100 fL). Hypochromia is accompanied by a MCH < 25 pg (where the normal range is 27–33 pg) and a MCHC of $<32\,g\,dL^{-1}$ ($<320\,g\,L^{-1}$; where the normal range is $32–36\,g\,dL^{-1}$).

Disorders associated with iron metabolism most commonly occur as a result of an iron deficiency but can result from a block of the enzyme that inserts iron into the heme ring (ferrochelatase). Iron and its metabolism are vital to the body since the hemoglobin molecule cannot function without iron; two thirds or more of the total body iron is in the RBCs and their precursors. Each milliliter of RBCs contains 1 mg of iron. Storage iron is present in macrophages or normoblasts as ferritin (Fe^{3+} plus apoferritin) or as hemosiderin. Most of the iron in RBCs is released from hemoglobin as it is degraded in macrophages and it is transported to normoblasts by plasma transferrin. Only 0.9 to 1.3 mg iron per day is lost from the body. The small amount of iron lost from the body is replaced by dietary iron absorption, which amounts to about 1 mg daily. Men have a daily dietary iron intake of about 11 mg with intestinal absorption of 6%, in women these figures are 11 mg daily and 12% absorption. Dietary absorption is increased in an iron-deficiency state, but only to a maximum of 20%. Dietary iron is digested in the ferric form but is reduced to the ferrous form by hydrochloric acid in the stomach and by and other reducing agents such as foods. The ferrous form

is much more rapidly absorbed by the mucosal cells of the duodenum and upper jejunum than the ferric form. Once in the mucosal cell, the iron is oxidized back again to the ferric form and is coupled to the protein apoferritin to form ferritin. Iron enters the circulation bound to transferrin. The normal serum iron concentration is $0.5-1.5\,g\,L^{-1}$. The normal total iron-binding capacity of transferrin is $2.5-4.0\,g\,L^{-1}$. Transferrin is normally only 30% saturated. Serum ferritin concentration is normally $15-200\,\mu g\,L^{-1}$. Normoblasts in the marrow have transferrin receptors and the capacity to extract iron from plasma transferrin. The marrow receives only 5% of the total cardiac output of blood but it extracts 85% of the circulating iron.

Iron deficiency anemia (IDA) is one of the most common forms of anemia in the USA. IDA results when iron loss exceeds iron intake over a long period and the body's iron stores are depleted, (e.g. chronic blood loss); insufficient iron is then available for normal heme production. Iron deficiency can also develop when there is an increased need for iron such as in rapid growth in infancy, childhood, or in pregnancy. IDA is the most common cause of anemia in children aged 6 to 24 months. In the adult male, even with no dietary iron intake, body iron stores of 1000 mg would last for 3 or 4 years before iron depletion anemia would occur. Most cases of IDA in adult males, therefore, results from chronic blood loss. IDA presents with such clinical symptoms as numbness and tingling of extremities; atrophy of the epithelium of the tongue with soreness, cracks, or ulcers at the corners of the mouth; abnormal cravings for such things as dirt or ice (pica); and concave or spoon-shaped nails. IDA must be distinguished from other microcytic/hypochromic anemias such as thalassemia, anemia of chronic disease, and sideroblastic anemia. Table 3 illustrates the laboratory differentiation of these anemias using iron-related parameters.

Sideroblastic anemia is associated with an increase in tissue iron stores. Defective synthesis of heme occurs as the result of multiple enzyme defects, and consequently, there is iron overload in the mitochondria of normoblasts. Several variants have been classified. A hereditary, X-linked form, is found mostly in males and does not become apparent until adolescence. Acquired idiopathic sideroblastic anemia can occur in either sex and has its onset in later adult life. Some of the RBCs may be megaloblastic and 10% of patients develop acute leukemia. Pyridoxine-responsive sideroblastic anemia can be treated with high amounts of pyridoxine (vitamin B_6) to maintain normal hemoglobin synthesis. Drug-induced sideroblastic anemia has been known to occur in patients chronically exposed to antituberculosis drugs (e.g. isoniazid), lead, chloramphenicol, or ethanol.

3.4 Globin Chain Disorders

Globin dysfunction can occur from genetic mutations and gives rise to two types of anemia. Most abnormal hemoglobins result from hemoglobinopathies where single amino acid substitutions in one of the polypeptide chains ($\alpha, \beta, \gamma, \delta$) causes malfunction. In the thalassemias, the globin chains normally have an unaltered amino acid sequence but have diminished (or absence of) synthesis. They usually involve an abnormal long or short polypeptide resulting from gene termination errors, frameshift mutations, crossover in-phase, deletion of codons, or fused hybrid chains.

In the classification of hemoglobinopathies, the abnormal hemoglobin is most commonly represented by a letter abbreviation. For example, sickle cell anemia involves hemoglobin S (Hb-S) and methemoglobinemia involves hemoglobin M (Hb-M). Hemoglobin variants can also be described by their amino acid substitutions (e.g. Hb-S is $\alpha_2\beta_2$ val 6). There are over 100 identified abnormal hemoglobins, but the most common worldwide occur in hemoglobinopathies such as sickle cell trait, hemoglobin C disorder, and hemoglobin E disorder.

Sickle cell disease is a genetically homozygous condition (Hb-SS) that results in a serious chronic hemolytic anemia. The anemia is apparent in childhood and is often fatal by 30 to 40 years of age. It is found almost exclusively in people of African/Carribean origin. The genetic alteration results in the replacement of glutamic acid at position 6 of the β-chain by valine. This molecular alteration changes the overall charge and electrophoretic mobility of the hemoglobin molecule. Hb-S is freely soluble in its fully oxygenated form. In its deoxygenated form, the abnormal molecules polymerize to form intracellular crystals that deform the RBC to the "sickling"

Table 3 Differential diagnosis of the microcytic/hypochromic anemias

	Serum iron	Total iron binding capacity	Transferrin iron saturation (%)	Ferritin	Bone marrow iron
Iron-deficiency anemia	↓	↑	↓	↓	↓
Anemia of chronic disorders	↓	N to ↓	N to ↓	N to ↑	N to ↑
Sideroblastic anemia	↑	↓	↑	↑	↑
Thalassemia	N to ↑	N to ↓	N to ↑	N to ↑	N to ↑

N, normal; ↑, increased; ↓, decreased.

shape that gives the disease its name. With homozygous sickle cell disease, sickling occurs even at physiological oxygen tensions. The rigidity of the RBC in the sickle shape prevents the cells rolling and bending as they pass through narrow vessels. This results in intravascular hemolysis. In compensation, the bone marrow becomes hyperplastic early in childhood, expanding the marrow space. Clinical complications accompanying sickle cell disease are serious, resulting in systemic organ damage. Early in childhood, bilateral painful swelling of hands and feet occur as a result of RBC sickling and capillary plugging (hand–foot syndrome). Splenic complications occur through splenic blood sequestration and pooling of blood, causing rapid splenomegaly and systemic hypovolemia. Patients can develop functional asplenia, which will lead to an impaired immune system and render the patient more susceptible to infections. Anoxic damage to the kidneys results from renal vascular plugging. Patients cannot produce a concentrated urine and commonly present with hematuria (hemoglobin in the urine). Vaso-occlusive crises result from capillary occlusion by sickle cells and loss of circulation to tissues; these occur intermittently and cause joint and abdominal pain. Aplastic crises can occur in the marrow owing to systemic intravascular hemolysis. Leg ulcers are common because of tissue ischemia from capillary blocking. A definitive diagnosis of sickle cell disease can be made by hemoglobin electrophoresis at alkaline and acid pH. Homozygous Hb-SS will produce a typical pattern of 0% Hb-A, Hb-S up to 80%, Hb-F (hemoglobin F) 1–20%, and Hb-A_2 2–5%.

Sickle cell trait is a heterozygous β-chain defect (Hb-AS) and is the most common hemoglobinopathy worldwide. The heterozygous genetic trait in healthy individuals normally has no clinical signs or symptoms. With physiologic acidosis, hypoxia of high altitudes, respiratory infection, anesthesia, or congestive heart failure, sickling will be caused with the same symptoms as seen in the homozygous form. Sickle cell trait protects against the lethal effects of falciparum malaria. Heterozygous Hb-AS will produce a hemoglobin electrophoresis pattern showing 50–65% Hb-A, 35–45% Hb-S, and normal Hb-F and Hb-A_2.

Hemoglobin C disease (Hb-CC) is a genetically homozygous condition affecting mainly people of African/Carribean origin that presents as a moderate hemolytic anemia with splenomegaly, jaundice, and abdominal discomfort. The genetic alteration results in a β-chain amino acid substitution of the glutamate at position 6 with lysine. Hexagonal or rod-shaped crystals of Hb-C may be found in RBCs on the blood smear. These crystals result from cellular dehydration of older RBCs, which increases their rigidity, splenic trapping, and destruction.

Hemoglobin E disease and trait is found mainly in Orientals. Homozygous patients have a mild anemia with moderate microcytosis and target cells on their blood smear.

Doubly heterozygous β-hemoglobinopathies are disorders in which the individual inherits a different abnormal β-chain gene from each parent. Hemoglobin-SC disease has a frequency about the same as homozygous sickle cell disease in people of African/Carribean origin. The severity of the disease falls between that of the sickle cell trait and sickle cell disease. Onset is in childhood, but physical difficulties do not occur until teenage years. Basically, the symptoms are the same as for sickle cell disease, but splenomegaly is more common.

Thalassemias are disorders of hemoglobin synthesis that occur particularly in people of Mediterranean, African, and Asian ancestry. The dysfunction involves low levels of functional globin chains, the chain normally having an unchanged amino acid sequence. Total lack of α- or β-globin production is known as thalassemia major. Production at a decreased rate is known as thalassemia minor. α- and β-thalassemias are associated with a decrease in the production of α- and β-chains, respectively.

Homozygous β-thalassemia (i.e. thalassemia major) results from either low production or a failure in production of the β-chain by both gene alleles. With a decrease in β-chain production, γ-chain production is high, resulting in increased hemoglobin F (Hb-F; the fetal form of hemoglobin). In normal blood, α-chains and β-chains combine to form hemoglobin. Because of the absence of β-chains, free α-chains accumulate and form unstable tetramers (α_4) that precipitate out of solution in the normoblast or RBC. These precipitates adhere to the inner membrane of the RBC and damage the cell, resulting in ineffective erythropoiesis and a severe hemolytic anemia. Clinical symptoms of β-thalassemia major include jaundice and splenomegaly early in childhood, prominent frontal bones (i.e. cheek, jaws), marrow hyperplasia resulting in a thinned cortex of the long bones, stunted growth and delayed puberty, hemochromatosis from regular transfusions, and cardiac failure (a major cause of death) from myocardial siderosis (often by 30 years of age).

Heterozygous β-thalassemia (thalassemia minor) involves malfunction of only one of the two alleles. Consequently, some normal β-chains are produced and the degree of anemia may vary from a severe microcytic/hypochromic anemia to normal clinical findings. A moderate-to-severe anemia is more common in those of Mediterranean decent. More severely affected individuals will also have a slight hemolytic jaundice and splenomegaly.

The α-thalassemias result from a partial or total decrease in the production of α-chains. While there are two β-globin genes per diploid genotype, there are

four α-globin genes. These are normally designated as αα/αα. Mild α-thalassemia (α^+-thalassemia) is a deletion of only one gene and is represented as $-\alpha/\alpha\alpha$. Severe α-thalassemia (α^0-thalassemia) is missing two α-globin genes and is represented as $--/\alpha\alpha$. Hydrops fetalis with hemoglobin Barts is the most severe form of α-thalassemia because individuals with this disorder have a diploid genotype of $--/--$, which is the absence of all α-chains. The absence of all α-chains is incompatible with life, and infants with this disorder are stillborn or die soon after birth. In the absence of α-chains, γ-chains and β-chains form tetramers known as Hb-Barts and Hb-H, respectively. Both of these tetramer forms migrate faster than Hb-A on electrophoresis gel.

3.5 Disorders of Impaired Production

Anemias resulting from impaired DNA replication result in a macrocytic, megaloblastic anemia. The anemia is referred to as megaloblastic because the red cell precursors in the bone marrow are larger at all stages of development than normal. Megaloblastosis is caused by a block in thymidine production. Thymidine is one of the four DNA bases and its deficiency leads to a diminished capacity for DNA synthesis because precursors have a prolonged intermitotic resting phase and mitosis is slowed. Cytoplasmic growth and hemoglobin synthesis continue at a normal rate but nuclear maturation is slow or arrested, resulting in an enlargement of the cell. The marrow is characterized by enlargement of precursor cells. Macrocytic anemia with a megaloblastic marrow can result in ineffective hematopoiesis and patients typically also have lower WBC and platelet counts. The marrow is hypercellular owing to increased EPO stimulation, but erythropoiesis is ineffective. Increased intramedullary destruction of the blood cell precursors leads to increased heme catabolism and signs of hemolysis. Cells of the granulocyte and megakaryocytic series are also enlarged. Megalaloblastosis is nearly always caused by vitamin B_{12} or folic acid deficiencies.

Vitamin B_{12} absorption occurs in the small intestine. Dietary B_{12} is released from digestion of animal proteins and is bound by gastric intrinsic factor (IF), a glycoprotein produced by the parietal cells of the stomach. The B_{12}–IF complex adheres to specific receptor sites on mucosal epithelial cells of the terminal ileum of the small intestine. Absorption of vitamin B_{12} into the intestinal mucosal cell is favored by an acid pH less than 6.5. Only about 1% of dietary vitamin B_{12} is absorbed without IF binding. Normal serum B_{12} values are $200–900\ ng\ L^{-1}$. If vitamin B_{12} intake is stopped, total body stores of 2–5 mg will last for several years before a megaloblastic anemia will result. Vitamin B_{12} is a vital cofactor in the conversion of methyl tetrahydrofolate (folic acid; pteroyl monoglutamic acid) to tetrahydrofolate. This substrate is an important cofactor needed for the production of thymidine.

Inadequate dietary intake of vitamin B_{12} is extremely rare in the USA and is usually seen in strict vegetarians. Defective production of IF is the most common cause of vitamin B_{12} deficiency. Pernicious anemia is a ''conditioned'' nutritional deficiency of vitamin B_{12} caused by failure of the gastric mucosa to secrete IF. It is an inherited disorder, most commonly occurring in people after 40 years of age. People diagnosed with this anemia typically present with skin pallor, jaundice, a smooth, sore tongue, anemia, abdominal pain, constipation, and diarrhea. In addition, patients often demonstrate central nervous system symptoms such as numbness and tingling of the extremities, difficulty with balance and gait, loss of vibration sensation, weakness and spasticity, and personality changes in advanced disease. Many patients present with atrophic gastritis of varying degrees, since the gastric parietal cells secrete hydrochloric acid as well as IF. It is common to have an accompanying deficiency in stomach acid secretion.

Although pernicious anemia is the most common cause of B_{12}-related anemia, other causes are also seen. Gastrectomy (the surgical removal of part or all of the stomach) will remove the source of IF and result in megaloblastic anemia if vitamin B_{12} supplements are not provided. Malabsorption syndromes such as tropical sprue or inflammatory disease of the small intestine can result in defective absorption of vitamin B_{12}. Lack of availability of the vitamin in the small intestine can result from competition for dietary sources by other organisms such as a tapeworm or bacteria in blind-loop syndrome.

A deficiency of folate can also result in megaloblastic anemia. Folic acid is primarily acquired from the diet in such foods as eggs, milk, leafy vegetables, yeast, liver, and fruits. A small amount is formed by intestinal flora. The minimum daily dietary requirement has been set at 50 mg, which is just barely met by dietary sources. Folic acid is mainly stored in the liver. Its normal levels are $5–21\ \mu g\ L^{-1}$ in serum and $150–600\ \mu g\ L^{-1}$ in RBC.

Clinically, symptoms from inadequate dietary folate can occur within weeks, rather than the years that are required for vitamin B_{12} deficiency to become apparent. Leukopenia and thrombocytopenia are less common with folate deficiency but the symptoms and the hematologic picture are generally the same. Unlike vitamin B_{12} deficiency, a megaloblastic anemia caused by folate deficiency is most commonly a consequence of insufficient dietary intake. The usual diet contains just enough folate to meet minimum requirements, and body reserves will only last for 3 months. A female's demand for folate will increase in pregnancy, and pregnant women should receive daily folate supplements of about 500 mg. Dietary folate insufficiency is also common in the elderly on poor

diets and in alcoholics with liver disease. It is necessary to distinguish the latter from the anemia of liver disease with normal folate levels, in which the marrow will be macrocytic but normoblastic. In addition to dietary deficiencies, defective absorption of folate in the small intestine occurs in disorders such as celiac disease or nontropical sprue. Some drugs used in chemotherapy, such as methotrexate, are folic acid inhibitors and can produce a megatoblastic anemia.

Diagnosis of a folate deficiency is based upon the common clinical symptoms and on laboratory results. Serum folate is commonly decreased to $<3\,\mathrm{mg\,L^{-1}}$. RBC folate is a better indicator of tissue folate reserves and can fall to $<150\,\mathrm{mg\,L^{-1}}$ in late megaloblastic anemia.

3.6 Hemolytic Disorders

Increased RBC destruction and shortened lifespans give rise to the hemolytic anemias, which can be categorized as either intrinsic or extrinsic anemias. Intrinsic hemolytic anemias are usually hereditary and occur from defects in RBC membrane, RBC metabolism, or hemoglobin. Extrinsic hemolytic anemias form the RBC survival disorders that are acquired and occur secondary to a primary condition or stimulus. The RBC membrane disorders comprise a group of hereditary anemias that are caused by a molecular rearrangement of a RBC membrane structural protein known as spectrin. Hereditary spherocytosis is an autosomal dominant condition that presents as a moderate-to-severe anemia. It is characterized by spherocytic RBCs, which form the majority of the RBCs. These abnormal RBCs are intrinsically defective and have a smaller diameter and no central pallor. Splenomegaly is the cause of the shortened RBC survival and is often corrected by splenectomy. Hereditary eliptocytosis is an autosomal dominate condition resulting in an assembly defect of spectrin in the formation of the RBC skeletal lattice. Eliptocytes are abundant and often form more than 25% of the RBC population. In 90% of patients with hereditary eliptocytosis, there is no anemia and in the other 10% there is only a mild-to-moderate hemolytic anemia. Hereditary pyropoikilocytosis is an autosomal recessive severe hemolytic anemia that occurs rarely in people of African/Carribean origin. This condition is commonly characterized by RBC fragmentation. In the laboratory RBC fragmentation occurs at 45–46 °C instead of, at 49 °C as seen with normal RBC. Hereditary stomatocytosis results in a mild-to-moderate hemolytic anemia. Stomatocytes are RBCs that have a slit shape rather than a normal circular central pallor when viewed on an air-dried blood film. The stomatocyte are not as flexible as the normal biconcave disc-shaped RBC and have a shortened survival time. Paroxysmal nocturnal hemoglobinuria occurs more in young adults and is an acquired intrinsic defect of the RBC membrane. The defect renders the RBCs hypersensitive to complement C3 binding and, therefore, to hemolysis. The disease is characterized by chronic intravascular hemolysis, with or without hemoglobinuria.

RBC metabolic disorders resulting in hemolysis are genetic disorders leading to deficiencies of either glucose-6-phosphate dehydrogenase (G6PD) or pyruvase kinase. Deficiencies in G6PD are inherited as a sex-linked trait with full expression in the male hemizygote and partial expression in the heterozygous female. Anemia from a G6PD deficiency is found in all groups, but the highest incidence is among people of African/Carribean origin and those of Mediterranean origin. A deficiency of G6PD limits the regeneration of NADPH and, consequently, of glutathione. RBCs that lack the latter are vulnerable to the oxidative degeneration of hemoglobin. G6PD is normally highest in young RBCs and decreases as the cell ages; therefore, the older RBCs will be preferentially destroyed. Oxidized hemoglobin will denature and precipitate intracellularly as Heinz bodies, which adhere to the membrane causing rigidity, a tendency to lysis, and splenic trapping. A patient's susceptibility to hemolytic crisis can greatly increase with illness or exposure to drugs with oxidant properties. Pyruvate kinase deficiency is a rare autosomal recessive inherited condition. The enzyme catalyzes the conversion of phosphophenolpyruvate to pyruvate in the Embden–Meyerhof pathway, with the production of ATP. ATP is needed to maintain RBC membrane stability and flexibility, and this deficiency results in a mild-to-moderate hemolytic anemia. The RBC loses its flexibility and splenic trapping and removal of the RBC ensues. The patient may not have any observable RBC abnormalities until after splenectomy.

Acquired extrinsic hemolysis can be caused by physical forces that destroy the shape of the normal RBC and result in fragmentation. Heat from extensive burns will result in a hemolytic anemia with RBC fragments. Cardiovascular disease and insertion of prostheses can produce hemolysis through mechanical damage of RBCs. Microangiopathic hemolytic anemia with moderate RBC fragmentation can result as a secondary manifestation of chronic hypertension, thrombotic thrombocytopenic purpura, disseminated carcinoma, or disseminated intravascular coagulation.

Immune hemolytic anemias result from immunoglobulin binding to the RBC membrane and splenic removal of the cell. Autoimmune hemolytic anemia presents a hematologic picture of a hemolytic anemia with a positive direct antiglobulin test for the presence of bound antibodies. The disease can be associated with immunoglobulin G autoantibodies directed against rhesus

antigens. Isoimmune hemolytic anemia results from an immune response to a foreign antigen; for example, hemolytic disease of the newborn is caused by rhesus or ABO incompatibility between mother and fetus. Drug-induced immune hemolytic anemias are common and result from the adsorption of drug-stimulated immune complexes onto RBC membranes. Drugs such as penicillin or cephalosporin can act as haptens, inducing an immune response upon attaching to the RBC membrane.

4 THE DIAGNOSIS OF LEUKEMIAS

4.1 Common Characteristics of Leukemias

Leukemia can be defined as a generalized neoplastic proliferation or accumulation of hemopoietic cells in the bone marrow, organs, and the peripheral blood. Leukemias can be divided into three general clinical classifications based upon the course of the disorder and the prognosis.[6] Acute leukemias are refractory to remission, usually are fatal within 3 months, and present with a bone marrow packed with primitive cells (called blasts) of the cell type involved, with little differentiation. Subacute leukemias have a longer patient survival, of about 3–12 months, and usually have a clinical picture of an acute leukemia. Chronic leukemias are defined as having a survival of greater than 1 year if no remission occurs. Immature precursor blasts are elevated in marrow and blood but are usually less than 5% for most chronic leukemias. Maturation within a cell line still occurs. Two major cytologic groups of leukemia are myeloid or lymphoid. Each group has acute and chronic variations. The etiology and pathogenesis of leukemia is not totally known but can include basic causes such as ionizing radiation (most commonly related to myeloid leukemia), and benzene poisoning. There is a higher incidence of acute myelogenous leukemia in patients with a history of aplastic anemia that is drug-induced with chloramphenicol or melphalan. Many leukemias have strong evidence of a genetic influence. For example, a chromosomal translocation known as the Philadelphia chromosome is found in 85% of those diagnosed with chronic myelogenous leukemia (CML). Some leukemias demonstrate evidence of being virally induced, such as the link between Burkitt's leukemia and Epstein–Barr virus. Patients with most types of leukemia present with some or all of such common clinical symptoms as fatigue, low-grade fever, and repeated bleeding episodes. Many patients present with an enlargement of lymph nodes, liver, and spleen caused by hemapoietic cellular infiltration, resulting in abdominal discomfort. Bone pain, particularly the sternum, is common owing to the cellular hyperplasia of the bone marrow. Chronic headache and nausea secondary to cellular infiltration or hemorrhage around the central nervous system are seen in some acute leukemias. Most leukemic patients present with such laboratory results as anemia, extreme leukocytosis, thrombocytopenia, basophilia, an increase in immature precursor or blast cells in the peripheral blood and marrow, and nucleated RBC precursor cells in the peripheral blood.

4.2 Differential Diagnosis of the Leukemia

All leukemias are differentiated into acute versus chronic, and lymphocytic versus nonlymphocytic, based upon which cell type is found to be proliferating in the marrow. Nonlymphocytic leukemias include granulocytic, monocytic, megakaryocytic, and erythrocytic leukemias. Cell lineage is determined by morphological characteristics, cytochemical stains, and immunochemical typing with cluster of differentiation (CD) and other markers.

4.2.1 Myeloproliferative Disorders

Myeloproliferative disorders share some general characteristics. They are defined as a group of closely related diseases characterized by the spontaneous proliferation of erythroid, granulocyte, monocyte, or megakaryocyte precursors in the marrow. The spleen, liver, and, rarely, the lymph nodes may also be involved. All cell lines or only a single cell line may be involved. Myeloproliferative disorders are clonal in origin, having arisen from a single pluripotential hemapoietic stem cell. Cytogenetic abnormalities are common to most myeloproliferative disorders. Chronic myeloproliferative disorders include a group of leukemias that share a common stem cell lesion and are slow in their clinical course. These disorders consist of five clinical subtypes based on stem cell lesion and which cell lines are observed to hyperproliferate.

4.2.1.1 Chronic Disorders CML occurs mainly in middle-aged adults (e.g. 50 to 70 years old) with a slow and unrevealing onset of symptoms. Clinical symptoms are mainly caused by the body's increased load of myeloid cells and nutritional demands. CML has characteristic WBC differential count, showing complete maturation of granulocyte cells from myeloblasts to segmented neutrophils, and a bimodal distribution with myelocytes and neutrophils both exceeding other types in absolute numbers. Myeloblasts are usually less than 5% in the blood and marrow. Cytogenetic abnormalities found in CML are highly diagnostic. In direct bone marrow preparations, it can be found that 90% of patients with CML demonstrate the Philadelphia chromosome. The mutation results from a translocation from the long arm of chromosome 22 to an arm of chromosome 9.

Polycythemia vera is a clonal hyperproliferation characterized by excessive proliferation of erythroid, granulocytic, and megakaryocytic elements of the marrow, and an increase in circulating RBC mass (i.e. increase in HCT). The neoplasia is classified as an absolute primary polycythemia and is differentiated from secondary polycythemia by the hyperproliferation in all cell lines. There is a higher frequency in middle-aged men. Splenomegaly, thrombocytosis, functional platelet abnormalities (giant platelets), and hemorrhagic problems are found in about two thirds of reported cases. Most symptoms are caused by the increased RBC mass, increased blood viscosity, and hypermetabolism. In many patients polycythemia vera can evolve into myelofibrosis.

Myelofibrosis with myeloid metaplasia is characterized by a chronic, progressive panmyelosis with varying degrees of fibrosis of the marrow. Patients present with a massive splenomegaly owing to extramedullary hematopoiesis. There is a slow onset of the disorder and it occurs most often in people older than 50 years. Typical symptoms are weight loss, anemia, abdominal discomfort from the enlarged spleen, and an enlarged liver with jaundice.

Essential thrombocythemia is a clonal malignancy, most closely related to polycythemia vera and characterized by a predominance of megakaryocytic proliferation in the marrow. Clinically, patients often present with recurring, spontaneous hemorrhages (mostly gastrointestinal in origin). The peripheral blood shows a marked increase in platelets, up to 900 to 1400 $\times 10^9$ cells L^{-1}, with abnormal giant forms and fragments of megakaryocytes. A hypercellular marrow demonstrates a panmyelosis with increased megakaryocytes and increased reticulin, which may result in a ''dry'' aspiration. Thrombocythemia has a stable course for many years but may develop into other myeloproliferative disorders.

Myelodysplastic syndromes are a family of marrow disorders found mainly in people over 50 years of age and characterized by ineffective cellular production. They are believed to be caused by a defect in a member of the marrow stem cell pool, resulting in increased proliferation and inadequate maturation or an imbalance in one or more cell lines. This group of disorders has been termed preleukemias because most cases progress to acute leukemias. Chromosomal abnormalities have been demonstrated in 40–90% of affected patients. Mutations have been commonly found on chromosome 5 but also have been noted to involve chromosomes 7, 8, and, less commonly, 12 and 20. The myelodysplastic syndromes are classified into five subtypes as defined by the French–American–British Cooperative Pathology Group (FAB): refractory anemia, refractory anemia with ring sideroblasts, refractory anemia with excess blasts, refractory anemia with excess blasts in transformation, and chronic myelomonocytic leukemia.

4.2.1.2 Acute Disorders Acute myeloproliferative leukemia (AML) includes a group of leukemias that share a common stem cell lesion, are more refractory to treatment, and have a rapid onset in their clinical course. AML is a disease of adulthood or babies less than 1 year old. Many cases are believed to be viral induced or can be related to exposure to radiation, chemicals, or diseases with a long preleukemic phase. Many patients present with very high WBC counts but sone can have below normal counts. AML is resistant to treatment with chemotherapeutic agents. The disease has a rapid onset, with symptoms often resembling an acute infection. Acute leukemias consist of eight clinical and morphological subtypes, depending on marrow and cell lineage characteristics. The FAB classification of AML has seven subtypes: AML-M0 to AML-M7. Acute stem cell leukemia (AML-M0) is a leukemia in which the predominate hyperproliferating cells are early progenitor cells that cannot be classified using cytochemical or routine immunological techniques. AML-M1 and AML-M2 are known as acute myeloblastic leukemias. AML-M1 is defined by a predominance of myeloblasts in the marrow with very little maturation beyond the blast stage. AML-M2 differs from M1 in that a significant number of myeloid precursor cells are found in the marrow beyond the blast and promyelocyte stages (i.e. maturation is involved). AML-M3 is known as acute promyelocytic leukemia. Promyelocytes instead of myeloblasts are the major abnormal cell found in the marrow and blood. A hypergranular and hypogranular varient (M3v) of AML-M3 can be seen, based on the granular appearance of malignant promyelocytes. AML-M4, known as acute myelomonocytic leukemia, demonstrates a predominance of both monocyte and granulocyte precursor cells in the marrow and peripheral blood. A relatively small number of these malignancies are accompanied by a moderate eosinophilia. This variation is known as AML-M4E. By contrast, AML-M5 is a relatively pure monoblastic leukemia. Two variations can be found for the monoblastic type. AML-M5a is poorly differentiated monocytic leukemia where more than 80% of the monocytic cells are blasts. Patients with this subtype tend to be younger and have a poorer prognosis. AML-M5b is a differentiated monocytic leukemia where fewer than 80% of the monocytic cells are blasts. AML-M5 may be accompanied by lymph node enlargement. AML-M6, or erythroleukemia, was formally known as Di Gugliemo syndrome. This type of AML results from an abnormal proliferation of both erythroid and granulocytic precursors, with a predominance of erythroid precursors in the marrow, forming 50% or more of the marrow

precursor cell population. Often, AML-M6 progresses to AML-M1. Acute megakaryoblastic leukemia (AML-M7) is a newer FAB classification of acute nonlymphocytic leukemias. This leukemia presents with a predominance of megakaryoblasts and megakaryocytes in the marrow, with an increase in reticulum fibrosis.

4.2.2 Lymphocytic Disorders

Lymphocytic leukemias represent a group of clonal disorders originating from cells of the lymphoreticular system of the bone marrow and lymph nodes. When neoplastic cells involve mainly the bone marrow and blood, the disorder is known as a leukemia. When the disease is limited mainly to lymph nodes and/or organs, the disease is known as a lymphoma. Occasionally, a lymphoma will develop into leukemia. Chronic as well as acute lymphocytic variations of lymphoma exist. There are four clinical variations of lymphocytic leukemias demonstrating a chronic course. In addition, three morphological and clinical variations of acute lymphocytic leukemia (ALL) have been identified by the FAB.

4.2.2.1 Chronic Disorders Chronic lymphocytic leukemia (CLL) is a slowly progressing clonal malignancy of lymphocytes in an arrested stage of maturation. CLL most commonly is seen in adults, with a mean age of occurrence at 55 years. The disorder is twice as common in men as women. Onset is slow, unrevealing, and is commonly discovered by accident or only in late stages of the disease. Patients have a leukocytosis of $10–150 \times 10^9$ cells L^{-1}, with 80–90% lymphocytes persistent over a period of weeks to months. Patients with CLL do not usually present with anemia or thrombocytopenia in early stages. As lymphoblast proliferation replaces the marrow with leukemic cells, production of other cell lines may suffer and these symptoms will appear. Hypogammaglobulinemia may also be present as a result of defective production in the leukemic lymphocytes. A marrow aspiration commonly shows an increase in morphologically mature lymphocytes. CLL can be divided into several types immunologically; 95% of patients have a B cell leukemia and 5% have a T cell form; there is a third rare form, hairy cell leukemia, which is found four times more in men than in women and has a mean age of occurrence at 50 years. The onset of the disease is slow and is characterized by proliferation of abnormal lymphocytes in the secondary lymphoid organs. Splenomegaly is a common physical finding. The laboratory profile of a patient with hairy cell leukemia includes pancytopenia or a depression of both B and T cells and the appearance of variable numbers of ''hairy cell'' lymphocytes, so-called because they have less cytoplasm than normal and numerous ''hair-like'' projections and frayed borders. The ''hairy'' cells are usually B cells in origin. Lymphosarcoma cell leukemia is a lymphocytic lymphoma that has transformed into a leukemic phase with the invasion of the marrow and blood with leukemic lymphocytes. The clinical course is more aggressive than in common B cell CLL. The lymphocytes in this disorder are B cells that are morphologically different to those seen in regular B cell CLL. Prolymphocytic leukemia is a variation of CLL in which there is a high number of morphologically immature larger lymphocytes, of B cell type, appearing as prolymphocytes. This variation is typically seen usually in older males. The leukemia is characterized by a high lymphocytosis, up to 350×10^9 cells L^{-1}. The disease is usually subacute and more resistant to treatment than is common CLL.

4.2.2.2 Acute Disorders ALL is a rapidly progressing clonal malignancy of early immature lymphocytes in an arrested stage of maturation. ALL is mainly a disease of childhood with a peak incidence at 4 years. A second peak of incidence occurs in young adults between 20 and 40 years of age (i.e. there is a bimodal distribution). Onset of leukemia is sudden, with symptoms of anemia, bleeding, fever, and fatigue. The spleen, liver, and lymph nodes are commonly enlarged. The predominate cell type is the lymphoblast or immature lymphocyte. A marrow aspirate commonly shows diffuse infiltration of lymphoblasts. ALLs are classified by the FAB as ALL-L1–ALL-L3. In childhood, 70% have ALL-L1 and 27% have ALL-L2; in adults, ALL-L2 is most common. ALL-L3 is also known as B cell Burkitt lymphoma and has vacuolated lymphoblasts.

The diagnosis of ALL cannot be made with complete certainty until cytochemical staining procedures have been performed to distinguish the lymphoblasts and lymphocytes from positively reacting cells of AML. There are several functional subclasses of ALL based upon immunological membrane surface markers. Common ALL antigen (CALLA), which is also CD10, is the most common marker and is seen in lymphocytes that do not show surface features of either B or T cells. This marker is the most common surface marker for ALL in children and, morphologically, the lymphocytes carrying it are ALL-L1. Null cell ALL is a proliferation of lymphocytes that do not carry markers for T or B cells or CALLA. It occurs in a small proportion of children but more in adults. Null cell ALL is thought to be a pre-B cell leukemia. T cell ALL accounts for 10–20% of all cases of ALL. This functional subtype occurs mainly in boys. Patients usually show a high leukocyte count pattern and have a poorer prognosis. T cell ALL is characterized by a high frequency of mediastinal tumor, skin, and central nervous system involvement. B cell ALL is the rarest of subtypes and corresponds to ALL-L3. Patients generally have a poor prognosis and can be diagnosed by demonstration of surface immunoglobulin receptors typical for B cells. It is

thought to be a memory cell leukemia. Better prognosis of ALL is seen in females, patients with lower WBC counts, and patients diagnosed with ALL-L1 subtype.

4.3 Differentiation by Cytochemical Stains

Cytochemical stains provide a useful tool in the diagnosis of leukemia by helping to differentiate the leukemic blasts of AML from leukemias of lymphoid origin.[7] It is difficult to distinguish between leukemias with just a blood smear stain, such as the Wright's stain. The general principle of most cytochemical stains is to incubate cells on a blood smear with a substrate that will react with an intracellular marker (e.g. lipid, lysozyme, glycogen, enzyme, etc.). In enzymatic procedures, the method detects an enzyme within the cell (e.g. acid phosphatase, alkaline phosphatase, or esterases) by coupling the reaction product of the enzyme to a diazonium salt (dye) to produce a visible reaction product. Some of the more common cytochemical stains used in the diagnosis of the cellular origin of leukemias are Sudan black, peroxidase, specific and nonspecific esterases, periodic acid–Schiff base, acid phosphatase, and the leukocyte alkaline phosphatase stain.

Sudan black and peroxidase (myeloperoxidase) share a common staining pattern. Sudan black stains phospholipids and sterols (lipids) in cytoplasmic lysosomes and mitochondrial membranes. Peroxidase stains the cytoplasmic primary azurophilic granules. Cytoplasmic granules stain faintly in myeloblast cells and strongly in mature neutrophils. As the cell matures from the promyelocyte stage to the mature segmented neutrophil, each stage is generally more positive for peroxidase stain. Leukemic myeloblasts will demonstrate stronger staining than normal blasts. Monocytes are generally unstained or weakly positive, with a few scattered stained cytoplasmic granules. Lymphoblasts and lymphocytes are negative. Peroxidase and Sudan black will strongly stain myeloid leukemias such as AML-M1–AML-M3 but will generally be weak or negative for monocytic and lymphocytic leukemias.

Esterase stains include either specific or nonspecific procedures. AS-D chloroacetate esterase stain (a specific esterase) is more useful in separating monocyte precursors from granulocyte precursors than Sudan black or peroxidase. Neutrophils and neutrophil precursors stain most strongly of all cell lines with AS-D chloroacetate esterase, with monocytes, lymphocytes, and their precursors not staining. α-Naphthyl acetate esterase stain (a nonspecific esterase) is also useful to differentiate between neutrophil precursors from monocytes and their precursors in acute leukemias. Monocytes stain strongly at all stages of maturity. The addition of sodium fluoride to the incubation solution inhibits the staining reaction in monocytes but not in granulocytes. Granulocytes at all stages of maturation do not stain or only stain weakly. Basophils, plasma cells, and T lymphocytes also stain. The esterases are most useful to differentiate a relatively pure myeloid leukemia or to differentiate between monoblastic leukemias. Blasts from acute leukemias AML-M1–AML-M3 will stain strongly with the specific esterase but only weakly or not at all with the nonspecific esterase. By contrast, monoblasts of AML-M4 and AML-M5 will stain strongly with the nonspecific esterase but will not stain with the specific esterase.

Periodic acid–Schiff base is often useful to identify some lymphocytic leukemias. Periodic acid is an oxidizing agent that converts hydroxyl groups on adjacent carbon atoms to aldehydes (carbohydrates). A positive reaction is seen in the presence of polysaccharides, mucopolysaccharides, glycoproteins, and other carbohydrates. Neutrophils are stained at most stages of development but most strongly in the mature stage. Myeloblasts usually are weakly stained or unstained. Lymphocytes may contain a few positive granules; however, in lymphocytic leukemias, the malignant lymphocytes of ALL-L1 and ALL-L2 may have an increased number of staining granules in a coarse or block-like pattern. Nucleated RBCs do not stain except in the abnormal erythroid precursors of AML-M6 (erythroleukemia).

Acid phosphatase staining is useful to confirm a diagnosis of hairy-cell leukemia and to separate T cell leukemia from B cell leukemia. The T cells stain whereas B cells usually do not. Monocytes also stain strongly. Neutrophils and their precursors stain but less intensely than monocytes. A variation of this staining procedure calls for the addition of L-tartaric acid to the staining reagent: tartaric acid-resistant acid phosphatase stain. The addition of L-tartaric acid inhibits the normal isoenzymes of acid phosphatase and, therefore, prevents staining in cells that would normally be positively stained. However, the acid phosphatase isoenzyme in the malignant lymphocytes of hairy cell leukemia is not inhibited by the addition of L-tartarate and the cells will still stain.

The leukocyte alkaline phosphatase staining procedure is primarily useful for distinguishing a leukemic neutrophilia from a nonleukemic neutrophilia seen with chronic inflammation and infection. This enzyme is found in neutrophils from the myelocyte stage to the mature segmented neutrophil stage. The enzyme is detected by its reaction with a naphthol phosphate in the presence of a diazonium salt (e.g. Fast blue or Fast violet) at pH 9.5. After staining, a Kaplow count is performed. A total of 100 mature neutrophils are scored from 0–4 based upon the intensity of the staining reaction (where 0 is negative and 4 is strongly positive). Adding the scores will result in a final score of 0 to 400. Reference values are usually

in the range 20 to 200, although each hospital laboratory usually establishes its own normal range. Increased staining occurs in infections, polycythemia vera, Hodgkin's disease, and in myelofibrosis with myeloid metaplasia. Decreased activity is seen with CML, AML, paroxysmal nocturnal hemoglobinuria, aplastic anemia, and in some viral infections such as infectious mononucleosis.

4.4 Differentiation by Immunochemical Typing

The science of immunochemical typing involves the use of monoclonal antibodies to identify unique antigens on the surface of different types of blood cell. Visualization of the antigen–antibody complex on the cell surface membrane is achieved by tagging the antibody with an indicator such as a fluorochrome or an enzyme that can be seen under the microscope or analyzed by its emitted wavelength with a cytoflow instrument. The antigens expressed on hemopoietic cells are now studied with the use of purified commercially prepared monoclonal antibodies that are chemically linked to different dyes for identification and quantification. These specific antibodies have been assigned numbers in the CD system developed by an international workshop group. Both the cellular antigen and the corresponding monoclonal antibody are given the same CD number. Antibodies directed to nuclear terminal deoxynucleotydyl transferase (TdT) have not, as yet, been assigned a number. If more than 20% of cells of interest react with a monoclonal antibody, they are considered positive for the corresponding antigen. Immunotyping of cells with CD markers is now recognized as useful to characterize the immunological characteristics of leukemias and lymphomas and to identify the exact origin of the malignant cell line.

The enzyme TdT is an intracellular DNA polymerase that is present in T and B lymphoblasts. TdT is weakly detected in pre-B cells and strongly in 60–85% of T cell lymphoblasts. Mature peripheral T cells and mature B cells are TdT negative. This assay is used in the diagnosis of ALL, lymphoblastic lymphoma, and the blast phase of CML. Lymphoblasts in 90% of patients with ALL are TdT positive, but only 5% of blasts in nonlymphocytic leukemias such as CML stain positive for TdT.

The majority of myeloproliferative disorders would type positive with a panel of CD11b or CD33. The marker CD11b is specific for the surface receptor for complement C3. Monocytes and granulocytes at most stages of development will be positive for this marker. Myelocytes and monocytes of all stages of maturation will be positive for CD33, which is known as an early myeloid marker. Monocytic leukemias such as AML-M4 and AML-M5 will also demonstrate positivity for CD14. This marker is expressed on 90% of monocytes of all stages of maturation and is not expressed on cells of myeloid origin.

Immunochemical typing is especially useful in the lymphocytic leukemias and lymphomas. CD markers can identify a leukemia as a B or T cell leukemia and can identify functional types of T cell leukemia (Table 4). Ninety-five percent of CLL immunologically types as a B cell leukemia and 50% as T cell leukemia. B lymphocyte CLLs and some B cell lymphomas will be positive with a pan-lymphocyte marker such as CD5. Late B cells and lymphocytes of hairy cell leukemia will be CD22

Table 4 Differentiation of the malignant lymphoproliferative disorders

Disorder	Cell type	CD marker	Cytogenetic abnormality
CLL			
B cell type	B cell; 95%	CD5, CD19, CD20, CD22, CD24	+12; t(11:14)
T cell type	T cell; 5%	CD2, CD3, CD8	
Hairy cell leukemia	B cell	CD5, CD19, CD20, CD22, CD25 (also sIg)	
Prolymphocytic leukemia	B cell	CD19, CD20, CD22, CD24 (also sIg)	
ALL			t(4:11)
CALLA	Early B; 70%	TdT, CD10, CD19, CD20	
Null cell	Pre B	TdT (also cIg)	
T Cell	Early-T; 15–20%	TdT, CD1, CD2	
Burkitt's type (ALL-L3)	Late-B	CD19, CD20, CD22, CD24 (also sIg)	
Lymphoma			
Nodular small cleaved lymphoma	B cell	CD19, CD20, CD24 (also sIg)	t(14:18)
Burkitt's	Late-B cell	CD19, CD20, CD24 (also sIg)	t(8 or 2, or 22:14)
Small cell lymphocytic and diffuse large cell lymphoma	B cell	CD19, CD20, CD24 (also sIg)	t(11:14)
T cell lymphoma	Late-T cell	CD2, CD3, CD4, CD5	t(7:14; or 11 or 9)

cIg and sIg, cytoplasmic and surface immunoglobulin, respectively; t, chromosomal transformation; T, trisomy.

positive; CD24 will react with peripheral B lymphocytes. Immunological markers for T cell leukemia include CD2, CD3, and CD8. CD2 is a general marker for all functional types of T cell. T cell CLL and T cell prolymphocytic leukemia express positivity for CD3, CD19, CD20, CD22, and CD24. A T cell CLL will present with positivity with CD8. Lymphocytes in hairy cell leukemia are usually the B type, demonstrating strong surface receptors for immunoglobulin and a positive reaction for common B cell markers such as CD19, CD20, CD22, CD24, and CD25 (the interleukin 2 receptor); CD25 is unique to hairy cell leukemia among the leukemias. There are several functional subclasses of ALL based upon immunological membrane surface markers. The most common type of ALL is a non-B, non-T leukemia, positive for CALLA. This immunological subtype is most common in children. Null cell ALL is a proliferation of lymphocytes that do not carry T cell, B cell, and CALLA surface antigens. T cell ALL accounts for 10–20% of all cases of ALL. This functional subtype occurs mainly in boys and typically demonstrates a high leukocyte count pattern and has a poorer prognosis. Malignant lymphoblasts demonstrate T cell markers such as CD2 (a pan-T marker), TdT, and CD1. B cell ALL is the rarest of subtypes and corresponds to ALL-L3. Patients with ALL-L3 generally have a poor prognosis. It can be diagnosed by demonstration of B cell typical surface immunoglobulin receptors as well as typical surface membrane markers such as CD19, CD20, CD22, and CD24.

5 ELECTRONIC CELL COUNTERS

5.1 Principle of Impedance Counters

The principle of electrical impedance was developed by Wallace Coulter in the 1950s. His discovery was based on the detection and measurement of changes in electrical resistance produced by cells passing through a small aperture.[8] Particles, which are more commonly cells, are made to flow through small openings between two electrodes. Cells are suspended in an electrically conductive diluent, physiological saline, and pulled through the aperture by a vacuum applied inside the aperture tube. A current is applied between an external electrode suspended in the cell solution and an internal electrode inside the glass aperture assembly. As each cell passes through the opening, the electrical resistance between the two electrodes changes because cells are poor conductors of electricity. Each cell that passes through the aperture increases resistance (R) between the electrodes. As R increases, voltage (V) increases and a voltage pulse of short duration is produced. The number of pulses is directly proportional to the number of cells being pulled through the aperture and the magnitude of the voltage pulse is proportional to the volume of the particle passing through the aperture. This principle is described by Ohm's law (Equation 1):

$$V = IR \tag{1}$$

where I is current. Using this methodology, RBCs and WBCs can be counted. It is necessary to eliminate interference by the RBCs for WBC counting; therefore, the RBCs are first lysed by a chemical additive. Multiparameter counters include a totally automatic diluting system, count RBCs and WBCs at the same time, and provide ten or more parameters (Table 5).

Modern blood cell counters can analyze whole blood directly without predilution, aspirating as little as 100 μL of blood. Anticoagulated blood is aspirated and samples are taken to prepare two separate dilutions. To the dilution prepared for WBC counting, a RBC-lysing agent is added to release hemoglobin from the RBCs. In addition, a hemoglobin stabilization reagent is added to the mixture. The WBC dilution is directed into the WBC aperture where the WBCs are counted and the hemoglobin is measured. The RBC dilution is directed to the RBC aperture bath where the RBCs and platelets are counted. Particles in the RBC aperture path greater than 50 fL in size are sorted by the analyzer as red cells. Particles ranging from 2 to 20 fL in the RBC bath are counted as platelets. The RBCs, WBCs, and platelets are each counted in triplicate. The voltage signals from each aperture are sent to pre-amplifier analyzers where the signals are amplified and sent to what is known as the voting analyzers. In each voting card, the three output voltages from the three cell counts are received and averaged. Each of the three signals must agree within 3 standard deviation units. If one does not agree, it is rejected and the other two are averaged. If all three are beyond the range, the analyzer will reject the entire count.

5.2 Measured Parameters and Methods

Most modern cell counters provide values for ten standard parameters, which is known as a hemogram and displays values for a differential representation of each common cell population as a histogram. Cell counts such as the RBC, WBC, and platelet counts are obtained by direct counting methods in the aperture baths. The hemoglobin is measured spectrophotometrically from the WBC sample after it has drained into the hemoglobin cuvette. The absorbance of the cell diluent is taken as the zero absorbance. The MCV is determined directly from the voltage pulse heights of the RBC count. The HCT

Table 5 Standard hemogram parameters produced by modern automated cell counters

Parameter	Normal range[a]		How obtained
	Conventional units	SI units	
WBC	$4.8–10.0 \times 10^3\,mm^{-3}$	$4.8–10.0 \times 10^9\,L^{-1}$	Directly from WBC apertures
RBC			
Males	$4.7–6.1 \times 10^6\,mm^{-3}$	$4.7–6.1 \times 10^{12}\,L^{-1}$	Directly from RBC apertures
Females	$4.2–5.4 \times 10^6\,mm^{-3}$	$4.2–5.4 \times 10^{12}\,L^{-1}$	
Hemoglobin			
Males	$14–18\,g\,dL^{-1}$	$8.8–11.3\,mmol\,L^{-1}$	Directly from WBC apertures
Females	$12–16\,g\,dL^{-1}$	$7.5–10.0\,mmol\,L^{-1}$	
HCT			
Males	42–52%	42–52%	Calculated from MCV and RBC
Females	37–47%	37–47%	
MCV			
Males	$80–94\,\mu m^3$	80–94 fL	Derived from RBC histogram
Females	81–99	81–99	
MCH	27–31 pg	27–31 pg	Calculated from RBC and hemoglobin
MCHC	$32–36\,g\,dL^{-1}$	$320–360\,g\,L^{-1}$	Calculated from hemoglobin and HCT
RDW	11.5–14.5%	11.5–14.5%	Derived from histogram
Platelets	$140–440 \times 10^3\,mm^{-3}$	$140–440 \times 10^9\,L^{-1}$	Directly from RBC apertures
Mean platelet volume	$7.4–10.4\,\mu m^3$	7.4–10.4 fL	Derived from platelet histogram

[a] A range given as $A–B \times 10^3$ should be read as $A \times 10^3$ to $B \times 10^3$.
RDW, red cell distribution width.

is computed from the RBC count and the MCV using Equation (2):

$$\text{HCT} = \frac{\text{RBC} \times \text{MCV}}{10}\% \qquad (2)$$

The MCHC is computed from hemoglobin concentration and the HCT using Equation (3):

$$\text{MCHC} = \frac{\text{hemoglobin}}{\text{HCT}} \times 100\% \qquad (3)$$

The MCH is computed from the MCV and the MCHC using Equation (4):

$$\text{MCH} = \frac{\text{MCV} \times \text{MCHC}}{100}\% \qquad (4)$$

The RDW provides an estimate of RBC anisocytosis, which refers to RBC morphological variation in cell size. The RDW is obtained from Equation (5):

$$\text{RDW} = \frac{\text{A} - \text{B}}{k(\text{A} + \text{B})} \qquad (5)$$

where A is the MCV at which 20% of the RBCs are larger than the rest of the RBC population and B is the MCV at which 80% of the RBCs are larger than the rest of the RBC population. A constant k denotes the number that is required to give a mean value of 10. The normal range is 8.5–14.5%. Samples showing values greater than 14.5 would indicate a increased variation for RBC size. When considered with the MCV, the RDW can be a useful combination in the differential diagnosis of anemias.

Many impedance cell counters are equipped to provide a graphical analysis of blood cell populations as a histogram on an x–y plotter or matrix printer. The histogram describes a cell population by comparing cell volume on the x axis (as femtoliters, where $1\,fL = 10^{-15}\,L = 1\,\mu m^3$) with cell frequency on the y axis (as cells per cubic millimeter of blood, where $1\,mm^3 = 1 \times 10^{-6}\,L$). Cell distributions are displayed for platelets, WBCs, and RBCs. Cells in the WBC sample that have a volume larger than 45 fL are counted and plotted as WBCs. The WBC histogram has three peaks. The first peak is at 45–90 fL and represents the lymphocyte population, or the small mononuclear population of cells. The second population of cells peaks at 90–160 fL and represents the large mononuclear cell population. This is a minor population in normal blood and represents the monocytes. It can also contain abnormal cell types such as the immature precursor cell types found in leukemia.

The major population is found at the third peak, 160–450 fL; this wider volume span represents a normal population of mature granulocytes, such as the segmented neutrophil. The relative percentage values for lymphocytes, monocytes, and granulocytes are calculated by combining the distribution values and the total WBC distribution spread. In many types of cell counter the percentages of eosinophils and basophils are also provided. The absolute count of each cell fraction is also given as the product of the relative percentage of each cell population multiplied by the total WBC count. Instrument differential values estimations are useful as a screen

or as a system to check for abnormal WBC differentials. Instrument differentials are surprisingly accurate when compared with the slide manual differential.

Cells in the RBC sample with volumes of 36–360 fL are counted as RBCs and cell frequency is plotted on the histogram RBC scale as a normal single peak between 70 and 110 fL. The peak of the curve coincides with the MCV given on the hemogram. Cells in the RBC sample with volumes of 2–20 fL are counted as platelets and their frequency is plotted on the platelet histogram area. An example of a normal blood cell histogram is given in Figure 3.

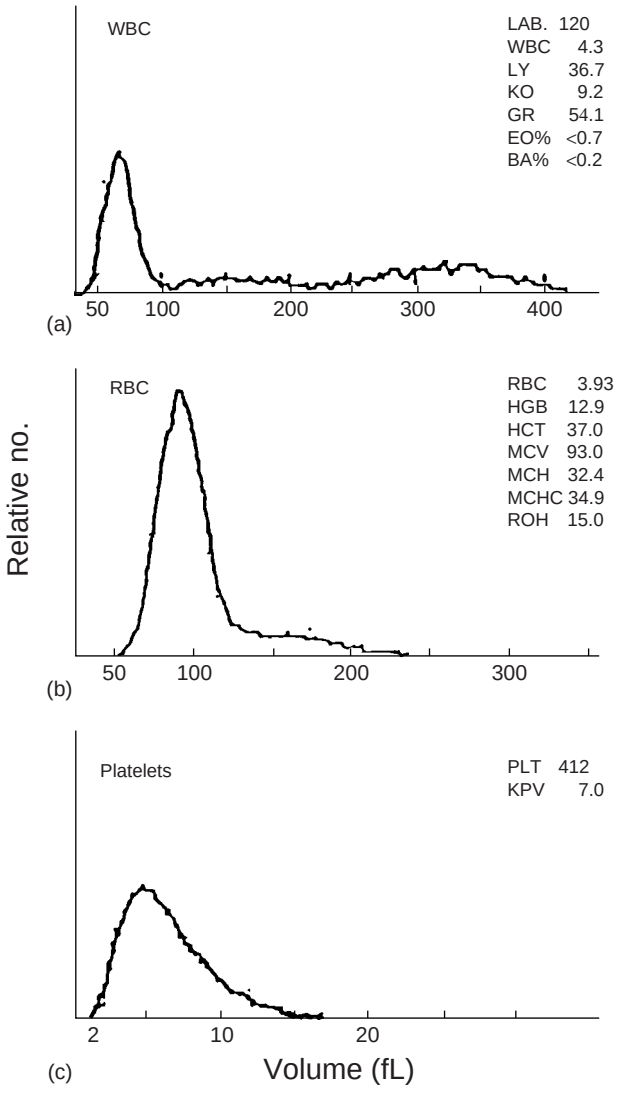

Figure 3 Normal blood cell histogram from a Coulter STKR cell counter. (a) WBCs are counted as cells larger than 45 fL; the peaks at increasing volumes represent lymphocytes, monocytes, and granulocytes. (b) Cells in the RBC counting chamber between 36 and 360 fL are counted and plotted as RBCs. (c) Cells in the RBC channel between 2 and 20 fL are plotted as platelets. Typical hemogram patient values are given on the right (see Table 5).

Abnormal WBC histogram patterns provide important information to aid in the diagnosis and differentiation of hematological disorders.[9] Instruments are programmed to warn the operator of any cell population abnormalities using a system of error "flags". These flags are commonly denoted as region error 1 through 4 and are represented as R1, R2, etc. They appear next to the differential parameters affected. An R1 flag warns of interference in the valley to the left of the lymphocyte peak, measuring about 35 fL. The lower threshold of the WBC histogram is 45 fL, but the histogram will extend lower to pick up abnormalities that show up left of the small mononuclear peak. An R1 error warns the operator of the possible presence of sickled RBCs, nucleated RBCs, or clumped or giant platelets. Any and all of these findings are associated with an abnormal blood cell population and may artificially raise the WBC count. An R2 error may indicate excessive overlap of cell populations at the lymphocyte/mononuclear cell boundary. An increase in the number of WBC counted around 90 fL will activate this error. Possible causes of an R2 warning indicate the presence of abnormal cell types such as reactive lymphocytes, immature blast cells, or plasma cells. An R3 warns of excessive overlap of cell populations at the mononuclear/granulocyte boundary around 160 fL. This may indicate the increased presence of immature granulocytes and is commonly referred to as a maturation "shift to the left". An R4 flag warns of an extension of the cell distribution at the upper end of the WBC threshold at or around 450 fL. This can most commonly be seen when the segmented granulocyte population is very high. The fifth region code, represented as RM, indicates possible region code violations.

5.3 Flow Cytometry

Cell counting instruments using laser-based optical light scatter were developed in the 1970s by Ortho Diagnostics Systems. Since its development, flow cytometry has been found to be beneficial in several applications. Cytoflow technology can be used to perform basic cell counting and to produce basic hemogram results. Cytoflow instrumentation, however, has its greatest advantages in being able to perform cell surface marker analysis for cell differentiation, counting lymphocyte subpopulations, and in performing DNA analysis of malignant tissues.[10]

To be distinguished and differentiated in flow analysis, white cells must first be tagged with a fluorochrome or chromophore in order to absorb the light the laser emits. Then the tagged cell emits light back at a wavelength sufficiently longer than the excitation light that the two wavelengths can be optically separated with selective filters. Fluorescein isothiocyanate (FITC) is the most

Table 6 Example of CD markers

Cell type	CD positivity
Stem cell	CD34
Myeloid	CD13, CD33, CD66
Monocytoid	CD33, CD11b, CD14
T Lymphocyte	CD2, CD7, CD5
B Lymphocyte	CD22, CD24, CD45, CD72
Megakaryocyte	CD31, CD36, CD41, CD61
Erythroid	CD45, CD71

common fluorochrome used in cytoflow. If multiple fluorochromes are used to identify more than one cell population, their emission spectra must have minimal overlap so that they can be separated and quantified. Commercially available fluorochromes used in cytoflow are covalently attached to antibodies. The antibodies are directed against surface antigens on the blood cells, the CD antigens.[11] All antibodies directed against a specific surface molecule, therefore, will have the same CD number. The CD designation of a cell can define both its cell type and its level of maturation within a cell line. Presently, there are over 100 commercially available CD markers. By using a panel of antibodies that covers a spectrum of CDs, a heterogeneous cell population such as that of the bone marrow or a lymph node can be accurately characterized and quantified. A sample list of commonly used CD markers can be found in Table 6.

The principle of flow cytometry is based on the channeling of the fluorochrome-tagged cells in a fluid stream to pass single file through a beam of laser light. Lasers emit monochromatic light (light of a single

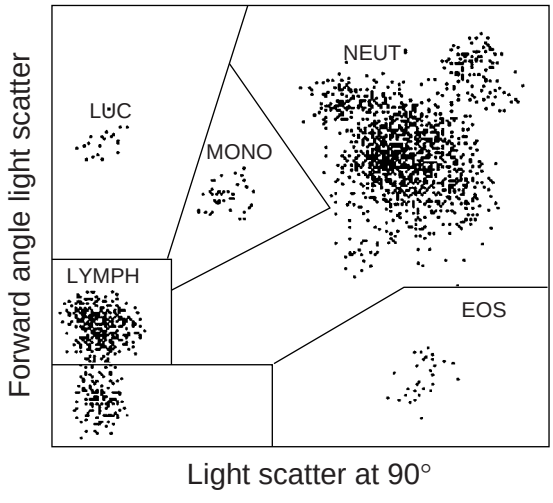

Figure 4 Example of cytoflow scattergram. Individual cells are represented by dots. Cell types are separated based on the degree of foward angle and 90° angle light scatter each cell type demonstrates. NEUT, neutrophils; MONO, monocytes; EOS, eosinophils; LUC, large unstained cells; LYMPH, lymphocytes.

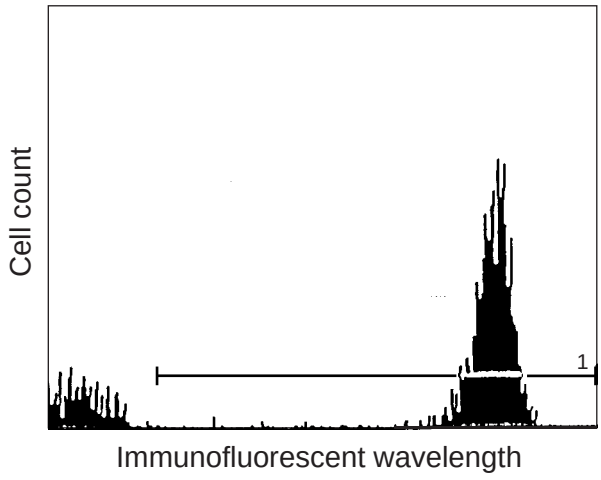

Figure 5 An example of an immunofluorescence histogram for a population of T helper cells tagged with fluorescent CD4.

wavelength). As each cell passes through and breaks the laser beam, photons are scattered and emitted by the cells and then separated into the resulting wavelengths by a series of filters and mirrors known as a photomultiplier tube. The separated light is then passed to individual detectors that will generate electrical impulses proportional to the amount of light striking the detector. The intrinsic cellular characteristics of each cell can be analyzed by the angles of light scatter. Intracellular characteristics are detected by forward-angle light scatter (FALS) and by side scatter (SS). FALS is related to the light collected by the photomultiplier tube along the axis of the laser and is proportional to cell size. The light deflected off the cells at 90° to the laser beam is SS. The reflected light reveals information about cell density, nuclear complexity, and cell granularity. Deflected light is converted to a scattergram plotting FALS against SS (Figure 4). Each dot will represent a single cell of given size and density. Emitted light from tagged fluorochromes is converted to a histogram that plots increasing intensity of cellular fluorescence against the cell frequency (Figure 5).

ABBREVIATIONS AND ACRONYMS

ALL	Acute Lymphocytic Leukemia
AML	Acute Myeloproliferative Leukemia
APTT	Activated Partial Thromboplastin Time
CALLA	Common ALL Antigen
CD	Cluster of Differentiation
CLL	Chronic Lymphocytic Leukemia
CML	Chronic Myelogenous Leukemia
EDTA	Ethylenediamine Tetraacetate

EPO	Erythropoietin
FAB	French–American–British Cooperative Pathology Group
FALS	Forward-angle Light Scatter
FITC	Fluorescein Isothiocyanate
G6PD	Glucose-6-phosphate Dehydrogenase
Hb-F	Hemoglobin F
Hb-M	Hemoglobin M
Hb-S	Hemoglobin S
HCT	Hematocrit
HMWK	High-molecular-weight Kinogen
IDA	Iron Deficiency Anemia
IF	Intrinsic Factor
INR	International Reference Ratio
IRP	International Reference Preparation
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MCV	Mean Corpuscular Volume
PT	Prothrombin Time
RBC	Red Blood Cell
RDW	Red Cell Distribution Width
SS	Side Scatter
TdT	Terminal Deoxynucleotydyl Transferase
TT	Thrombin Time
WBC	White Blood Cell
WHO	World Health Organization

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory • Phospholipids of Plasma Lipoproteins, Red Blood Cells and Atheroma, Analysis of

REFERENCES

1. R.G. McFarland, 'An Enzyme Cascade in Blood Clotting Mechanisms and its Function as a Biochemical Amplifier', *Nature*, **202**, 498 (1964).
2. B. Furie, B.C. Furie, 'Molecular and Cellular Biology of Blood Coagulation', *N. Engl. J. Med.*, **326**, 800–806 (1992).
3. *MLA 1000 Instrument Manual*, Medical Laboratory Automation, Pleasantville, NY, 1995.
4. A.R. Thompson, L.A. Harker, *Manual of Hemostasis and Thrombosis*, 3rd edition, Davis, Philadelphia, PA, 1983.
5. A.J. Erslev, 'Clinical Manifestations and Classification of Erythrocyte Disorders', in *Hematology*, 4th edition, eds. W.F. Williams, E. Beutler, A.J. Erslev, M.A. Lichtman, McGraw Hill, New York, 423–429, 1990.
6. D. Puti, R. Hromas, 'Treatment of Leukocyte Neoplasia', in *Diagnostic Hematology*, ed. B.F. Rodak, Saunders, Philadelphia, PA, 401–410, 1995.
7. T. Sun, C. Li, L.T. Yam, *Atlas of Cytochemistry and Immunochemistry of Hematologic Neoplasms*, American Society of Clinical Pathologists Press, Chicago, IL, 1985.
8. Coulter Electronics, *Significant Advances in Hematology: Hematology Education Series*, PN426115A, Coulter Electronics, Hialeah, FL, 1983.
9. Coulter Electronics, *Coulter Counter Model S-Plus IV with Three-Population Differential: Product Reference Manual*, PN423560B, Coulter Electronics, Hialeah, FL, 1983.
10. J.P. McCoy, E.J. Lovett, 'Basic Principles in Flow Cytometry', in *Flow Cytometry in Clinical Diagnosis*, ed. D.F. Keren, ASCP Press, Chicago, IL, 12–40, 1989.
11. W. Knapp, B. Dorker, W.R. Gilks, E.P. Rieber, R.E. Schmidt, H. Stein, A.E.G.Kr. Von dem Borne (eds.), *Leukocyte Typing IV: White Cell Differentiation Antigens*, Oxford University Press, New York, 1989.

DNA Arrays: Preparation and Application

Tom J. Whitaker and K. Bruce Jacobson
Atom Sciences Inc., Oak Ridge, USA

Mitchel J. Doktycz
Oak Ridge National Laboratory, Oak Ridge, USA

1 Introduction **1**
2 Applications **2**
2.1 Existence of Specific Sequences 2
2.2 Gene Expression 2
2.3 Genetic Research 4
2.4 Comparative Genomics 4
3 DNA Chip Construction **4**
3.1 Probe Types and Design 4
3.2 Chip Materials and Structures 6
3.3 Array Construction 7
3.4 Probe Immobilization 9
4 Hybridization **10**
4.1 Hybridization Conditions 10
4.2 Stability Factors 10
4.3 Probe Density 11
4.4 Hybridization Rate Enhancement 11
5 Detection **11**
5.1 Radioactive Labels 11
5.2 Fluorescent Labels 12
5.3 Mass Spectrometry 13
5.4 Electronic Detection 14
5.5 Other Detection Methods 14
6 Bioinformatics **15**
7 Summary **15**
Acknowledgments **16**
Tradenames **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

DNA arrays, which may also be called DNA chips, biochips, genosensors, GeneChips®, microarrays, or other names, are considered by many researchers and entrepreneurs to be the most promising technology for economically determining whether a specific sequence of bases exists in an unknown DNA sample. The sequence of bases in DNA is the source of genetic information; consequently, DNA arrays can be used for a number of applications, including diagnosis of genetic diseases or mutations, detection of polymorphisms, and identification of microorganisms. In addition, DNA arrays can also establish the amount of one or more particular messenger RNA (mRNA) in a cell or tissue to reveal gene activity.

DNA arrays take on many forms and there are exceptions to almost any statement describing them. However, a typical array will have multiple segments of single-stranded DNA (ssDNA) attached to the surface of a substrate. These probes have a known sequence and, under the right conditions, an unknown DNA sequence will bind or ''hybridize'' to one or more of them only if the unknown DNA has sequence that is related to the probe's sequence. This selective hybridization allows DNA arrays to provide sequence information about the unknown DNA.

This article describes some of the primary considerations in the design, construction, and implementation of DNA arrays. Items discussed are probe selection and design, selection of a suitable substrate, probe immobilization, array construction, hybridization, hybrid detection, how the application affects each of these factors, and how decisions made at one point of the process may affect other decisions.

1 INTRODUCTION

All genetic information is contained within the sequence of four bases, adenine (A), guanine (G), thymine (T), and cytosine (C), in DNA. As we learn more about the function of genes in humans, the ability to determine rapidly and inexpensively whether a specific sequence exists in a patient's DNA is becoming increasingly important for diagnosis of genetic diseases, mutations, and polymorphisms. This process should cover a wide range of diseases because, as F. Collins, Director of the US National Human Genome Research Institute in Bethesda, MD, has said, ''We now believe that the only disease not having some genetic component is trauma''.[1] DNA arrays (also called DNA chips, biochips, genosensors, GeneChips®, microarrays, etc.) are considered by many researchers and entrepreneurs to be the most promising technology for economically determining whether a specific DNA sequence exists in an unknown DNA sample.

DNA arrays typically consist of a thin solid substrate on which segments of DNA (called probes) of known sequence are attached in a pattern. The probes can be short oligonucleotides or amplified (duplicated) DNA products. Information about the unknown (target) DNA's sequence is obtained by allowing target ssDNA to bind, or ''hybridize'', to the probes. Under proper conditions, the resulting hybrid will be

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

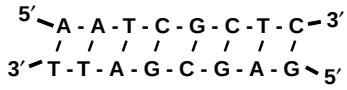

Figure 1 Example of a short (8-mer) DNA hybrid. The dashed lines indicate hydrogen bonds between the bases in the two DNA strands. The DNA backbone on which these bases are attached has directionality, with one end designated as the 3′ end and the other the 5′ end. Hybridization occurs between strands of DNA that are aligned opposite, or antiparallel to, each other.

stable only if the sequence in a section of the target DNA is complementary to the sequence in the probe according to Watson–Crick rules for base pairing: A to T and G to C. For example, if a probe site has the sequence AATCGCTC, the existence of a stable hybrid (shown in Figure 1) at this site would indicate that a section of the target has the sequence GAGCGATT. Note that the convention for writing ssDNA sequences is to present the sequence in the 5′ to 3′ direction, a nomenclature indicating the structural directionality of the DNA backbone. Two strands of DNA hybridize in an antiparallel sense, as seen in Figure 1.

Early concepts for DNA arrays and their applications were developed in the late 1980s at several laboratories.[2–6] One of the first meetings on the topic was a *Sequencing by Hybridization* workshop in November 1991 in Moscow, Russia.[7] As the title of the workshop suggests, DNA arrays were initially intended to determine the sequence of genomic DNA. Genomic mapping and diagnosis of genetic diseases were also emphasized as applications. Two formats were developed for sequencing by hybridization. In format I, DNA samples were immobilized on a surface and a solution of labeled oligonucleotides of known sequence was added and allowed to hybridize. In format II, oligonucleotides of known sequences were immobilized on the surface and labeled DNA or DNA fragments were allowed to hybridize. In either case, the sites at which hybridization occurred were identified by the presence of the label at probe sites on the surface. Hybridization at a specific site provided direct evidence of a sequence in the unknown DNA because it had to be complementary to the probe sequence at that site. These basic features, established in early work, are still the cornerstone of modern DNA arrays. Figure 2 illustrates these features for a format II array (the most common format). To economize on space in the illustration, the probes in Figure 2 have only three nucleotides (3-mers) (in practice, probes have eight or more nucleotides). There are $4^3 = 64$ possible combinations for these 3-mers, and there would be 4^n possible combinations for an *n*-mer. Because of the directionality of the DNA backbone, sequences that are mirror images of each other, for example AGT and TGA, are unique and both must be included in order for a set of sequences to be complete.

The early aspirations of using DNA arrays to determine the sequence of genomic DNA have not been realized. Initially, it was believed that the comprehensive manner in which the sequence data are obtained in a DNA array would provide adequate redundancy to correct for minor irregularities in the base pairing stabilities. For example, on any DNA array with a complete set of probe sequences for an *n*-mer, *n* probe sites will provide evidence of the identity of a single nucleotide in the target (see shaded column in Figure 2a). Although this redundancy was reassuring, the short length of the oligonucleotide probes proved to be a limitation because hybridization techniques cannot uniquely identify target sequences with three or more repeating internal sequences if the length of these repeating sections is one less than the length of the probes or greater. This is illustrated in Figure 2(a), where exchange of the first guanine with the first cytosine in the target DNA would give exactly the same result on the DNA array because of the repeating TT sequence around these bases. Of course, Figure 2(a) is based upon 3-mer probes, which are not used in practice, and longer probes could be used to uniquely identify this sequence. However, even with long probes, longer repeating sequences in genomic DNA occur sufficiently often to eliminate essentially the use of DNA arrays for sequencing totally unknown targets.

2 APPLICATIONS

2.1 Existence of Specific Sequences

The major uses for DNA arrays at the end of the 1990s are for determining if specific sequences exist in a DNA sample (Figure 2b) and for establishing the amount of one or more particular mRNAs in a cell or tissue to reveal gene activity (Figure 2c). Determining the existence of a specific sequence can identify genes and gene mutations, diagnose genetic diseases, and identify microorganisms. Because a specific sequence is being sought, the number of probes is greatly reduced. Figure 2(b) shows how a few short oligonucleotide probes can be designed so that their sequences overlap to establish a longer sequence in the target.

2.2 Gene Expression

Within the human and other genome projects that are currently underway, not only is the sequence of a gene important but also how that gene is expressed is of

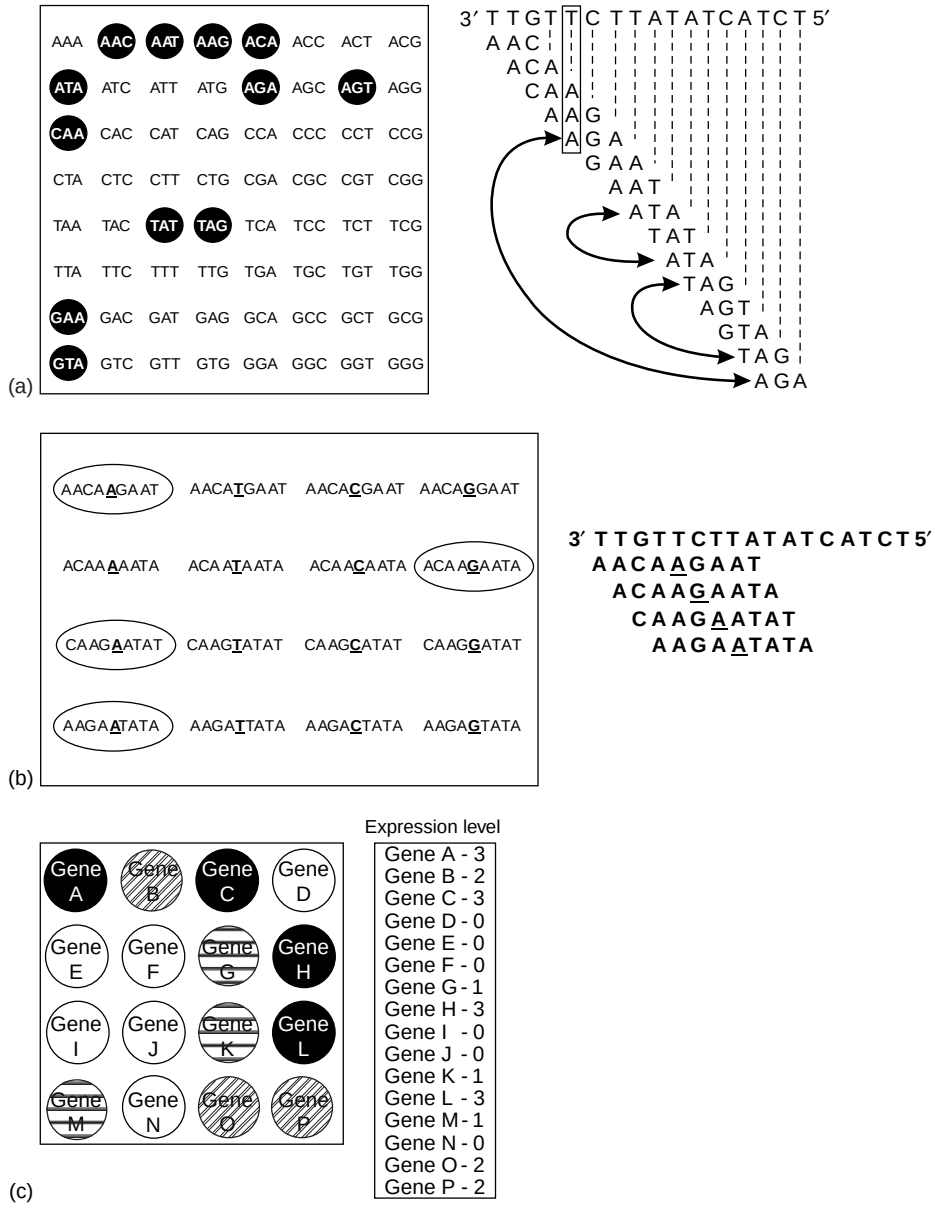

Figure 2 Arrays used in sequence analysis. (a) An array designed for sequencing. Here the array shows all possible 3-mer probes because there are only $4^3 = 64$ possibilities. In practice, probes are typically much larger (see text). The probe sites that would hybridize to the depicted 17-mer DNA target are highlighted with black circles. In general, the number of n-mer probes that will hybridize to an m-mer target is $m - n + 1 - r$, where r is the number of repeats. In the illustration, there are three repeats (shown by connecting arrows) and so there are $17 - 3 + 1 - 3 = 12$ hybridized probe sites. (b) An array designed to detect a specific sequence. The four circled probes are ''tiled'' complements of the desired 12-mer target, and all four would hybridize if the illustrated target DNA is present. The probes in each row that are not circled are single-base variations of the circled probe. This acts as a control and allows detection of single-nucleotide polymorphisms at four locations in the specified target sequence. (c) An array designed for gene expression. In this case, long probes hybridize to complementary DNAs (cDNAs) produced by reverse transcriptase from mRNA. The level of hybridization indicates the degree of gene expression.

vital interest. When the complete sequences of certain organisms become available, previously unknown genes will be revealed. The role that these, and previously known, genes play in cell function and development can be explored with DNA arrays by monitoring the rate of production of the mRNAs, which are the immediate gene products created when a gene is functioning to produce a protein (see Figure 2c). Gene expression is measured by the amount of cDNA produced by reverse transcriptase from the mRNA isolated from the biological source.

For this application, DNA arrays are usually designed to detect the cDNAs, but they can also be created for direct detection of RNA. In addition to genes that specify proteins through the mRNAs, there are also those that produce functional RNAs, such as transfer RNAs, ribosomal RNAs and others, which can also be probed by arrays.

2.3 Genetic Research

DNA varies in size among bacterial plasmids and different organisms from about 5000 base pairs (bp) to over 1×10^{10} bp and a single gene may have approximately 1000–10 000 bp. In the DNA of mice and humans, there are approximately 50 000 to 100 000 genes, each of which is regulated by one or more other sequences within the DNA. Even though the base pairs involved in protein production and regulation of gene activity account for only a small fraction of the total base pairs in DNA, an understanding of all the human gene sequences is a goal. For this purpose, a framework analogous to the periodic table of the chemical elements would be desirable[8] so that properties of the genes can be viewed in a systematic manner and their features compared. DNA arrays offer the opportunity to gather this information quickly and organize it for further consideration and study.

2.4 Comparative Genomics

Comparative genomics can utilize the DNA arrays in a particularly useful way, either in terms of exploring genomic structures or for detecting the presence of foreign DNA. A gene may have two or more alleles that result from modification of the DNA sequence within the gene's boundaries. Alleles arise as mutations, which can be transmitted to offspring in subsequent generations. Inherited diseases result from alleles that contain altered DNA sequences; the diagnosis of such diseases and characterization of various alleles is well suited to DNA arrays. Infectious diseases are caused by viruses or bacteria and these entities contain DNA or RNA that is characteristic of the infectious species. Similarly, soil farming, hydroponic farming, and fermentation processes benefit from knowing which microbial populations are present. Biological warfare agents are often bacteria or viruses and rapid identification is desirable. Again, application of DNA arrays would provide a rapid method of differentiating among the variety of infectious agents or microbial species.

Distinguishing individuals can be accomplished by comparing ''DNA fingerprints''. Distinct patterns of DNA fragments seen on electrophoresis gels are produced by cutting genomic DNA with restriction endonucleases, enzymes that cut the DNA only at specific sequences, to produce fragments known as restriction fragment length polymorphisms. DNA arrays can be designed not only to recognize the various structures that give rise to these polymorphisms but also to provide much richer patterns that can distinguish different DNAs.

3 DNA CHIP CONSTRUCTION

Constructing DNA arrays was one of the early obstructions to performing array-based analyses. Chip construction requires a library of test probes, and methods for constructing arrays either in situ or by positioning and attaching previously prepared probes to solid surfaces. Numerous laboratories have developed methods for chip construction, resulting in a variety of approaches. Many of the methods have specific attributes that make them suitable for particular substrates, applications, or detection methodologies. The primary steps in chip construction are probe selection and design, selection of a suitable chip substrate, probe immobilization chemistry, and array construction. All steps are strongly interrelated, and the benefits bestowed by any particular choice in options may dictate the overall strategy that an investigator pursues.

3.1 Probe Types and Design

The selection and preparation of a useful set of probe sequences is a challenging aspect of array technology and one that critically affects the success of an array application. One of the first considerations in selecting an immobilized probe is the probe length. Immobilized DNA probes are generally classified according to length. Long DNA probes are prepared by molecular biology techniques such as polymerase chain reaction (PCR) or cloning. Short oligonucleotide probes (up to about 25 bases) are usually synthesized by chemical means.

3.1.1 Long Probes

PCR is becoming the standard method for preparing long probes, although some applications, such as sequencing by hybridization using format I, use large insert clone arrays that are evaluated with short synthetic probes. In general, long probes (1) form stable hybrids with target sequences, (2) are easily prepared from devices supplied by a mature industry devoted to high-throughput PCR product generation, (3) do not require modification for immobilization, and (4) tend to average out sequence-dependent variations in stability. Because of their size, the instability caused by a single-base mismatch is less significant in a long probe, making allele-specific hybridization difficult and preventing the use of long

probes in applications directed at polymorphism characterization. However, long probes are well suited for applications that monitor gene expression. The sequences of long probes generated by PCR for expression monitoring are often limited to those that can be reliably amplified. Beyond this factor, probe selection is dependent on availability of relevant gene sequences and/or clone libraries. Complete sequence information and effective gene-finding programs enable creation of expression probes by directly amplifying appropriate regions of genomic DNA. The value of this approach has been demonstrated in expression studies on yeast, where more than 6000 genes have been identified.[9] Although uncharacterized cDNA libraries can be useful for identifying new genes of related function,[10] their use places greater demands on array size, information handling, and associated cost. Several databanks, such as GenBank, dbEST, and UniGene, help to meet the need for genome and cDNA sequence information.[11–13] In addition, various species-specific and tissue-specific cDNA libraries are becoming commercially available, but caution is advised as some currently fall short of being sequence verified and they do not supply complete coverage for all genes.[10]

3.1.2 Short Probes

The primary advantage of short probes is that a single-base mismatch destabilizes a short hybrid more than it does a long hybrid. This property can be exploited for applications that require allele-specific hybridization, such as determining single-nucleotide polymorphisms. Even though very short probes would provide greater discrimination between single-base mismatches, overall hybrid stability decreases at shorter lengths. Because of this, and the fact that longer probes assist in melting out (dissociating) intramolecular structures that may exist in the target, probes are typically eight bases in length or longer.

In short probes, a variety of interactions contribute to the stability and specificity of the DNA hybrids. For example, mismatches located near the center of the probe are significantly more destabilizing than those residing in the end positions.[14] For this reason, allele-specific probes are designed to have the diagnostic nucleotides (those where polymorphisms are expected to occur) at the interior of the probe. Other contributions to hybridization stability in short probes include nearest-neighbor effects, number of GC base pairs, dangling end effects owing to unpaired bases at the end of a duplex region (sticky ends), and probe concentration. The sequence-dependent stability of short DNA hybrids has been thoroughly evaluated and reliable parameters for sequence-dependent stability are available.[15–17] Likewise, the kinetics of DNA association and dissociation have also been characterized.[18] This literature serves as a basis for selecting effective probe sequences and hybridization conditions. However, despite the predictive powers resulting from such information, empirical rules are still used[19] because of the lack of information related to other structural and environmental factors that affect hybridization to probes on chips. These factors include the local concentration of the probe and target, the effect of higher order structures (intramolecular foldings), and the effect of surface chemistry and proximity to the surface. But even with an understanding of these influences, it is often advisable to build redundancy into the probe design. For example, in gene expression monitoring and gene diagnostics, where a single 20-mer probe could potentially identify the target, several probes are typically used for a single gene target.[19,20] This use of closely related probes is often a principal feature of the DNA array. Photolithographic in situ synthesis techniques (see section 3.3) are particularly well suited to the creation of large numbers of systematically varying probes.

3.1.3 Other Design Factors: Hybridization Stability

General considerations for selection of probe sequences include: maintaining similar hybridization stability among all probes on the chip; targeting a unique sequence; avoiding homonucleotide runs, particularly runs of guanine, which can form various intramolecular structures; and avoiding targeting regions in the target DNA that are prone to intramolecular folding. Alternatively, arbitrary sequences could be used in place of gene-specific sequences.[21] Such an approach may be useful for applications where there is little knowledge of the target sequence.

3.1.4 Modified Probes

The use of simple hybridization techniques with standard oligonucleotide probes may not provide sufficient selectivity for certain allele-specific applications. Modifications to the hybridization protocol or to the oligonucleotide probes themselves have been developed to enhance selectivity or hybrid stability in these cases. For example, peptide nucleic acids have an ethylenediamine glycyl backbone instead of the usual sugar phosphate background[22] and, therefore, do not have an ionic charge. This eliminates the need for high ionic strength buffers for hybridization and allows the probes to form more stable hybrids with DNA. Modified nucleic acid bases have also been employed, such as deoxyinosine,[23] 3-nitropyrrole,[24] or 5-nitroindole.[25] These bases can be used to enhance hybridization discrimination.

Enzymes have also been used with immobilized hybridization probes to add specificity, as illustrated in

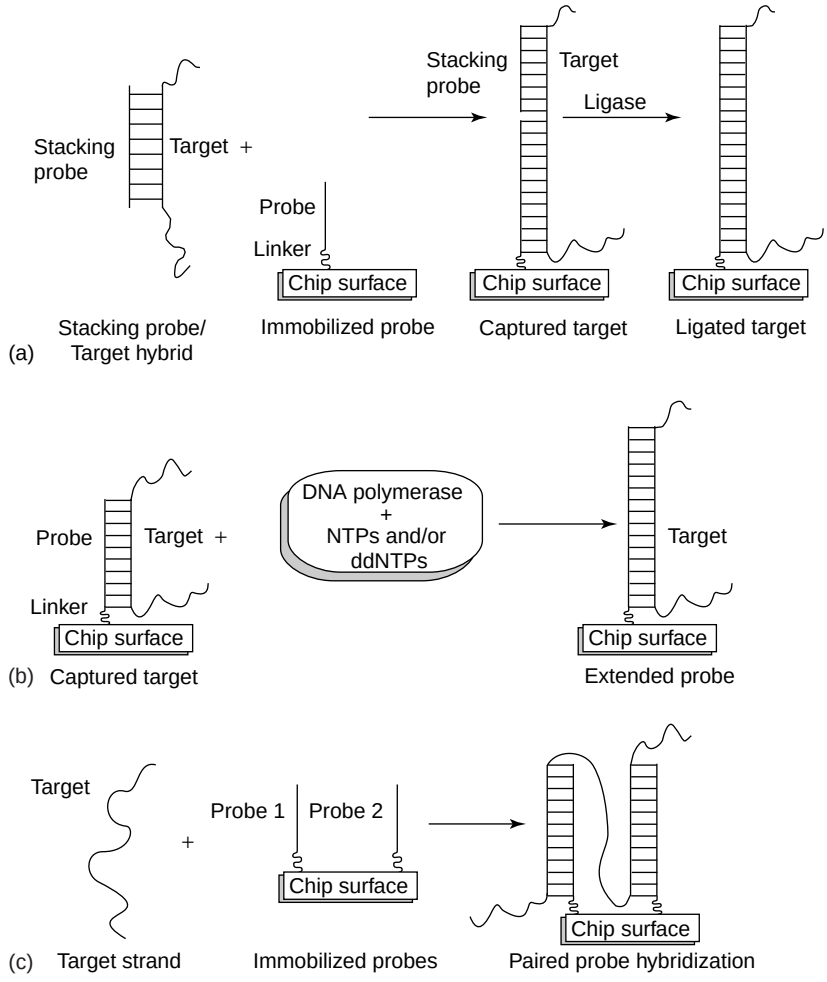

Figure 3 Modification of probes to improve specificity. (a) A second "stacking" probe can be attached to the hybrid using ligase. (b) The length of the hybrid can be extended using DNA polymerase with the deoxynucleoside triphosphates (NTPs) of adenine, guanine, cytosine, and thymine to form a sequence governed by hydrogen bonding with the template; dideoxynucleoside triphosphates (ddNTPs) are used when a single addition is desired. (c) Another technique uses two separate probes, one to capture the target and one to provide diagnostic information.

Figure 3. Ligase can be used to attach a second hybridization probe to the immobilized probe (Figure 3a)[26–28] and DNA polymerase can be used to extend the immobilized hybridization probes (Figure 3b).[29–31] In both instances, the added specificity of the enzyme has been used to enhance the allele-specific analysis. Unlike direct hybridization techniques, the diagnostic nucleotides are placed at the end of the immobilized probe, where the enzyme reaction occurs, rather than in the center.

Increased hybridization stability and specificity can also be accomplished by use of additional complementary probes. For example, various contiguous stacking procedures have been described where a second, solution-phase, hybridization probe hybridizes to the region adjacent to the first immobilized probe.[32,33] This is similar to the ligase-based approach shown in Figure 3(a) but relies solely on the hybridization stability of contiguously stacked oligonucleotides. In addition, the use of long stacking probes prevents intramolecular hybridization regions, such as hairpins. Another approach, shown in Figure 3(c), uses a paired probe array where two hybridization probes are immobilized in the same location.[34] One probe serves to assist in the capture of the target, while the second provides the diagnostic information.

3.2 Chip Materials and Structures

The substrate material serves as the base for immobilizing previously synthesized DNA probes or for building in situ probes. Features that should be considered when selecting a chip substrate include cost, ease of handling, and compatibility with attachment methods, hybridization labeling, and detection.

3.2.1 Glass

Of the various substrates, glass is the most commonly used. It has a reactive surface and can be obtained in an inexpensive form (microscope slides). Glass is compatible with commonly used fluorescent detection schemes, and it can be etched or molded to modify the structure. Some glass structures, such as microchannel arrays, can lead to advantages in terms of detection sensitivity and binding capacity (see section 4.7). In another form of glass substrate, probes are immobilized on the ends of glass fibers, which can be only a few hundred micrometers in diameter.[35] Individual fibers may each contain a different probe, and after hybridization to fluorescing target DNAs, light-channeling effects of the fiber allow sensitive optical detection of individual hybrids, even when the fibers are bundled closely together. However, until high-throughput, high-density construction methods are available for glass fiber arrays, their use will likely be limited to applications requiring only a few dozen probes.

3.2.2 Silicon

Silicon substrates for DNA arrays are becoming more common. Silicon possesses many of the same attributes as glass: a reactive surface when oxidized, and the ability to be constructed in different forms. This last attribute is significant, as many of the tools developed for the construction of integrated circuits can be applied. The application of microfabricated structures and the integration of microelectronics devices in DNA arrays present an exciting approach to DNA-based analyses. For example, DNA probes have been directly attached to individual pixels of a charge-coupled device (CCD) array and also to individual photodiodes of a photodiode array.[36,37] Direct attachment to the detection element enables efficient signal collection from radioactive or fluorescent labels and simultaneous acquisition of data from multiple probes. Some of the novel silicon-based techniques require design and microfabrication of devices, which is an expensive undertaking. The initial cost limits these structures to specialized laboratories. However, once in production phase, the costs of microfabricated silicon structures will drop significantly and they should be viable candidates for high-volume DNA array markets.

3.2.3 Gold

Gold surfaces have also been used as chip substrates. Much of the work has been associated with the self-assembly of thiol compounds on gold surfaces, and this provides several avenues for development of chemical sensors and biosensors.[38] The coverage of this insulating monolayer is very complete and uniform, and, at neutral pH, long-chain alkylthiol monolayers are stable over several days.[39] The alkyl chain can be terminated with a carboxyl group and then activated with *N*-hydroxysuccinimide so that oligonucleotide probes synthesized with a terminal amine can be attached.[40,41] Gold surfaces are especially useful for DNA chips using electronic means of detection or surface plasmon resonance detection (see section 5.5).

3.2.4 Polymers

Several polymer-based materials have been used as chip substrates. Nitrocellulose or nylon membranes[42–44] have a high binding capacity but usually require derivatization to enable immobilization of probes. Procedures and equipment for handling polymer membranes are common to most molecular biology laboratories. However, these membranes generally possess a high fluorescence background and are most commonly used with radioisotope labels.

The modified polystyrene surface of microtiter plates is also employed for DNA hybridization assays. DNA probes arrayed in these microtiter plate wells take advantage of a mature industry that uses plate-based immunological screening.[45] The plates are commercially available with derivatized surfaces for immobilizing DNA probes and are typically produced in a particular size and with a specific number of wells. Relatively high-density plates of 1536 wells and greater, plate readers, and plate washers are readily available.

Other polymeric substrates include polyacrylamide and polypropylene.[46,47] Small pads of polyacrylamide can be created on glass surfaces and derivatized for immobilizing DNA probes. These pads are a three-dimensional network of fibers providing a high binding capacity. Probe concentration can be varied to enable hybridization discrimination. Polypropylene is an inexpensive material, moldable in a variety of forms, and quite rugged. It is generally unreactive and requires specialized equipment for derivatization. The derivatization methods are not common to molecular biology laboratories and, again, appear to be pursued primarily by specialized laboratories.

3.3 Array Construction

Array construction techniques can be grouped into two basic approaches: in situ synthesis where synthetic oligonucleotide probes are prepared in place, or ''spotting'' procedures where previously prepared probes are mechanically positioned to create the array.

3.3.1 In Situ Synthesis

In general, the in situ synthesis techniques are capable of producing extremely large arrays of probes but are

limited to the production of short oligonucleotides. The synthesis of oligonucleotides in situ uses either standard phosphoramidite chemistry or a modified procedure where photolabile-protecting groups are used in place of the standard acid-labile protecting group.

3.3.1.1 Phosphoramidite Chemistry With standard phosphoramidite chemistry, oligonucleotides can be synthesized using a physical mask that constrains the flow of synthesis reagents between a conventional DNA synthesizer and a defined probe position on the flat substrate.[48,49] The substrate is usually first derivatized with a linker. Translation of a single mask, or of multiple masks, can be used to build up the array of oligonucleotide probes. Alternatively, the phosphoramidite reagents can be delivered by ink jet printing techniques to precise locations that define the probe sites.[50] In this case, oligonucleotides are built up by the delivery and removal of sub-nanoliter volumes of the synthesis reagents. In both cases, the arrayed oligonucleotides are prepared in parallel, allowing construction of the entire set of probes in just a few hours. The physical masking approach is useful for the construction of tiled oligonucleotide arrays to be used for sequence checking purposes. One benefit is that the yield of the full-length phosphoramidite product can be quite high because of the high coupling efficiency (>99%). However, the primary drawback is that the physical mask limits the density of the oligonucleotide probes. Currently, the devices for constructing arrays by these approaches are custom-built and are not available commercially.

3.3.1.2 Photomasking The highest density probe arrays are prepared using photomasking techniques, which have been adapted from methods commonly used for construction of electronic integrated circuits.[51,52] In this approach, the monomer units are prepared with a photolabile-protecting group. This protecting group prevents coupling of subsequent nucleotides until it is removed by exposure to light of the appropriate wavelength and intensity. This technique is being used to construct commercial high-density DNA arrays. One drawback to light-based deprotection synthesis is that the yield of full-length product is generally lower than that obtained with conventional phosphoramidite chemistry as a result of lower coupling efficiency. An alternative technique, using photoresistors as physical masks combined with conventional chemistry, has been described.[53]

3.3.2 Spotting Techniques

High-density oligonucleotide arrays can also be produced by first synthesizing the individual probes and then applying them to spots (spotting) on the substrate. This approach is suitable for arraying any sort of reagent and is the method most widely used for preparing cDNA arrays. The DNA obtained from a single 100-μL PCR amplification (long probes) can be used to prepare hundreds of arrays. Similarly, the amount of material resulting from a conventional 0.2-μmol chemical synthesis (short probes) can be used to prepare thousands of arrays. When the probes for the array are defined and available, this approach can be economical and easily set up in any laboratory.

Once the probes are synthesized, they must be spotted onto the substrate. Several commercial supplies of array instrumentation are available for this purpose, and there are detailed plans for building arrayers.[10] The various spotting methods are all based on a ''pick and place'' routine, similar to commercial automated liquid-handling systems. In this approach, the sample is picked up from a source plate, typically a 96- or 384-well microtiter plate, and placed onto the target substrate. Transfer of the probe may be accomplished via a sharp pin, a fine capillary, or a droplet ejection technique.

3.3.2.1 Pin-based Techniques Pin-based spotting techniques are becoming a standard method for constructing high-density arrays on microscope slides or membranes,[54–56] and several instruments that use this technique are available commercially.[10] In this approach, the sample is withdrawn by dipping a pin tip into a well of the microtiter plate. A small amount of sample (of the order of several hundred nanoliters) is retained by capillary action in a small groove cut into the sharply pointed pin. This is typically enough material to deposit a spot of the probe on several dozen arrays. Delivery of the sample is achieved by physically touching the pin tip to the substrate surface. The diameter of the resulting probe spot depends on the dimensions of the pin and can be as small as 100 μm with current technology. The pins are commercially available and can be fitted onto conventional automated liquid-handling systems. However, many of these systems do not possess the positional accuracy needed to prepare high-density arrays.

3.3.2.2 Capillary-based Techniques Another approach to creating microarrays uses fine glass or metal capillaries to aspirate and dispense samples.[57,58] Typically, the inner diameter of the capillary may be of the order of 100 μm. By using low-volume, precisely actuated syringe pumps to aspirate and dispense the sample, volumes of the order of a few nanoliters can be applied to the surface. Because of surface tension, these low volumes must be brought in contact with the surface to ''touch-off'' the droplet. In contrast to pin-based spotting, this approach allows for metering of the reagents.

3.3.2.3 Droplet Ejection Techniques Droplet ejection techniques borrow features from ink jet printers and use either pressurized solenoid valves or piezoelectric or thermal pumps to eject reagent droplets.[59–61] In operation, the sample is aspirated into a narrow-bore dispensing tip, which is then brought to the array and a pressure pulse ejects the droplets. In the solenoid ink jet, the pressure pulse results from rapid opening and closing of a pressurized valve. The valves can be actuated in a sub-millisecond time frame and, by altering the actuation time, pressure head, or nozzle dimensions, precise volumes can be metered. In piezoelectric ink jets, electromechanical actuation exerts pressure on the fluid in the passageway, causing a droplet to eject, while thermally based jets superheat the liquid and the expanding vapor ejects the probe. Actuation can be performed in tens of microseconds, resulting in droplet volumes ranging from tens to hundreds of picoliters. Reagent metering is typically accomplished by dispensing a preset volume repeatedly. The piezoelectric and thermal approaches are amenable to greater miniaturization than the solenoid valve approach, as demonstrated by their use to deliver ink in desktop printers. Commercial systems based on solenoid or piezoelectric reagent jets are available.

3.3.2.4 Speed and Density in Spotting Methods With all spotting methods the speed of array construction is directly related to the number of dispensing tips. On current devices that use multiple tips with the touch-off technique, dispensing tips are fixed relative to each other at a spacing equivalent to that of the wells in the source plate, which is 9 or 4.5 mm, respectively, for standard 96- or 384-well microtiter plates. Because all the tips in these devices contact a single substrate simultaneously, the physical size of the array is closely related to the number of tips and their spacing. High probe density is achieved by offsetting the contact position on each delivery. In current instruments, as many as 32 dispensing tips are used for the construction of dozens of arrays at a rate of approximately one set of spots every second. In droplet ejection devices, liquid can be ejected from each tip individually, allowing probes to be closely spaced by moving the set of tips between each ejection. Operating in this manner allows printing on small formats but requires additional movements. The ability to print on small formats is especially important when using patterned structures on the substrate. Such structures can be created when using micromachined, molded, or printed substrates, or they may exist inherently, as when probes are deposited directly onto detector elements. The pattern defines the location of the immobilized probe and confines the delivered droplet to prevent spreading. The use of these patterned structures requires greater positional accuracy and registration between the dispensing tips and target substrate, and their fragility necessitates gentle deposition of the probe droplets.

3.4 Probe Immobilization

A final consideration in constructing DNA microarrays is the procedure for fixing the probe to the substrate surface. Stable attachment of the probe to the surface is usually accomplished through covalent bonds. This procedure will depend on the probe and surface but in most cases it involves three steps: surface derivatization, probe modification, and postattachment processing. The processes we describe here will be on glass or SiO_2 surfaces where relevant experimental characterization has been performed.

3.4.1 Long Probes

For larger probes containing several hundred base pairs, the general approach has been to crosslink the DNA probe to polylysine-derivatized glass surfaces.[9] Larger probes offer a greater number of inherent sites for crosslinking and generally do not require modification before attachment. After spotting the probes to the polylysine-derivatized glass and briefly rehydrating the surface, crosslinking is activated by exposure to ultraviolet light. After treatment, the surface is typically ''blocked'' to prevent background association of other substances to the surface. This approach is simple and easily applied to the simultaneous production of many arrays.

3.4.2 Short Probes

For low-molecular-weight DNA probes, such as those produced by synthetic means, some derivatization of the probe is necessary. Typically a primary amine, thiol, or carboxylic acid functionality is attached to the oligonucleotide during the initial or final round of chemical synthesis to yield, respectively, either a 3′ or 5′ modification. These groups are then either directly attached or crosslinked through an intermediate linker molecule to a modified substrate surface. Surface modification usually involves the use of silane reagents to derivatize the surface. For example, epoxysilane treatment has been used to yield an epoxide group that is reactive with primary amines or hydrolyzed for direct synthesis of probes by phosphoramidite chemistry.[62,63] After spotting the probes, coupling occurs within a few hours at slightly above room temperature. Amine-derivatized oligonucleotides have been coupled directly to isothiocyanate-derivatized surfaces[64] and other surface preparations, such as aminophenyl and aminopropyl silane treatments, have been described.[65]

Direct coupling of 3′-propanolamine-derivatized oligonucleotides to unmodified glass surfaces has also been reported.[14]

3.4.3 Other Factors

Other issues related to attachment chemistry are the rate of coupling and the processing steps required after derivatization. For spotting volumes of approximately 1 nL, complete evaporation of the solvent can be quite rapid. When preparing arrays with a large number of probes, individual spots will have evaporated before complete construction of the array. This may necessitate the use of a humidified enclosure around the arraying apparatus or methods for rehydration to achieve effective attachment. Processing after the spotting stage is often required to remove unbound reagent and block unreacted surface functionalities. Unbound reagents are usually removed by simple washing procedures, and blocking is achieved with a small organic molecule that reacts with the remaining reactive sites. For all of the attachment schemes, further blocking is often necessary to reduce the background associated with certain detection schemes.[57] This may involve treatment with reagents such as bovine serum albumin, various polymers, or nonspecifically binding DNA.

4 HYBRIDIZATION

Hybridization is the process that provides specificity to DNA arrays. It is the specific pairing of a sequence in a single-stranded probe with the complementary sequence in target ssDNA molecules to form a double-stranded structure. The pairing results from hydrogen bonds in the base pairs AT and GC. This simple and selective pairing is complicated by sequence-dependent and solution-dependent conditions that affect the stability and specificity of hybridization. These complications can be addressed by effective probe design (discussed earlier), modification of hybridization conditions, and the manner by which probes are immobilized on the surface.

4.1 Hybridization Conditions

The hybridization conditions describe the time, temperature, and buffer conditions for forming the double-stranded complex. Because formation of this complex is reversible, proper choice of conditions is dependent on the stability of the complex. Duplex stability is often characterized by a melting temperature, which is defined as the temperature at which 50% of potential base pairs are formed under specified conditions.

4.2 Stability Factors

4.2.1 Base Pair Composition

One major factor that influences stability values is the ratio of AT to GC base pairs. Hybrids with a large fraction of GC base pairs are more stable than those with mostly AT base pairs because the latter contain only two hydrogen bonds whereas the GC base pair has three. In addition, the GC base pairs apparently form stronger van der Waal bonding to adjacent base pairs (stacking interactions). In long hybrids, the ratio of GC to AT does not vary far from an average value. But in short hybrids, such as the 8–20 nucleotide length commonly used for probes on DNA arrays, the stability of the hybrid varies with its base pair content. This inherent difference in stability between fully complementary hybrids owing to base pair content can be minimized by using tetramethylammonium chloride, *N*,*N*,*N*-trimethylglycine, or certain chaotropic salts.[66–68] Although these salts reduce the dependency of the melting temperature on GC content, results with this approach have been variable.

4.2.2 Salt Concentration

Higher salt concentration tends to stabilize the hybrid by shielding the negative charges in the DNA backbone and reducing the repulsion between strands. Typically, the ionic strength is adjusted by adding simple salts such as NaCl or $MgCl_2$ to the buffer solution. Arrays made with peptide nucleic acids have no charge on the probe and, consequently, the salt concentration can be very low. Low salt concentration will disrupt intramolecular hybridization in the target DNA, which can prevent hybridization to the probes.

4.2.3 Temperature

Another major factor that influences hybrid stability is the temperature. Ideally, the probes would be designed so that all complementary hybrids would have similar thermostabilities. This is desirable so that interaction between bases that are not complementary can be discriminated against. This discrimination is usually accomplished by adjusting the temperature, salt conditions, or time for the hybridization and/or post-hybridization washing steps. Usually the conditions are determined empirically.

4.2.4 Other Factors

A set of rules for determining optimal hybridization conditions is highly desirable; however, this is complicated by the variety of interactions that affect duplex

stability. Other features that affect hybridization include the nearest-neighbor stacking interactions, the length of the probe–target hybrid, the position of mispaired bases within the duplex, the length and sequence of unpaired bases at the end of the duplexed regions, and the concentration of the probe and target. Basic thermodynamic studies based on evaluation of melting temperatures have revealed the importance of some of these interactions and their relative significance.[14,15,69] For example, centrally located mismatches in 8-mer oligonucleotide probes are significantly more destabilizing than similar mismatches residing in end positions. The sequence and length of a dangling end will affect the stability of a duplex. Certain sequences can stabilize a duplex by an amount equivalent to an additional base pair. Probe and target lengths are also important. Longer probes form more stable hybrids; however, as the length increases, the ability to discriminate mispaired bases decreases. Target length is also critical. Hybridization to short probes is generally reduced when long target sequences (>100 bases) are used. This is presumably because intramolecular structures are formed in the target or there is restricted access to the surface. Some of these difficulties can be ignored when DNA arrays are used in a comparative manner, as is common when comparing gene expression profiles. In these experiments, the design of the array allows probe-dependent features to be normalized so that the patterns of hybridization between the two different samples can be compared without attention to hybridization yield.

4.3 Probe Density

Even at low temperatures, where strand dissociation (melting) occurs very slowly,[70,71] large variations are observed in hybridization yield, indicating that there are factors other than stability of the hybrid which affect the yield. For example, in both long and short probes, the fact that the probes are attached to a surface limits access to the bases nearest the surface, especially if the probes are tightly packed. Reducing the surface density of probes can reduce these steric problems and improve the yield. Studies with short probes have shown that optimum hybridization is achieved when a surface density of 1.4×10^{13} probes cm^{-2} is used,[64] which is a factor of about 40 less than the highest density achievable on some materials. In addition, inserting spacers between the surface and the first base of the probe can reduce steric problems and dramatically increase the hybridization yield. The effect is dependent on the length of the spacer. The yield increases with increased length of the spacers until a maximum is reached, and then slowly decreases as the length is further increased.

4.4 Hybridization Rate Enhancement

Hybridization is normally a slow process requiring several hours and ways to increase the rate are being explored. One technique takes advantage of the fact that DNA has a strong negative charge and will be attracted toward a positively charged body. Electronically charging a probe site attracts target DNAs to the site and increases the concentration of target at the probe. This decreases the time required for hybridization from hours to minutes.[72,73] Concentrating the target in this manner also allows lower concentrations of the target to be used, thus potentially reducing the number of amplification steps required for preparation of the target DNA. After hybridization, the polarity can be reversed to enable ''electronic stringency'', a technique for making sure that unbound or weakly bound species are removed, leaving only fully complementary strands of DNA at the probe site. In addition, the same electronic charging process can be used to attach probes to a specific probe site, using one solution of probes at a time and concentrating each at the desired site.

Another method of increasing the rate of hybridization is to attach probes to the inside of microchannel glass structures and force the target DNA to flow through the channels. Microchannel glass is essentially a bundle of microcapillary tubes, each about 0.5 mm in length with a bore diameter of only a few micrometers. This not only provides over 100 times the surface area of the same cross-section on a flat surface, but also the confinement of the target DNA to narrow channels reduces diffusion time and speeds up the hybridization process. Further, microchannel glass provides light channeling effects that are useful for optically based detection. Other advantages include the reduced spreading and slow evaporation of deposited reagents.

5 DETECTION

After hybridization, the next step is to determine at which sites the target DNA has hybridized. Generally, a label is attached to the target DNA before hybridization. After hybridization and washing, detection of the label at a site is evidence that the probe at that site is complementary to a section of the target (with due consideration of potential binding of mismatches).

5.1 Radioactive Labels

Early developmental work on DNA chips utilized radioactive labels. In these studies, ^{32}P was most commonly used to replace the analogous stable isotopes in DNA or RNA. The presence and concentration of these radioactively labeled oligonucleotides can be sensitively

and quantitatively determined using a Geiger counter or scintillator, or by the exposure of photographic film (radiography), although the last technique easily saturates and is quantitative only at low concentrations. The sensitivity and accuracy of these techniques led to early acceptance among researchers. Advances in imaging detectors and the use of isotopes with longer half-lives such as ^{33}P now provide even crisper images. These attributes, together with the consistent hybridization of radiolabeled oligonucleotides and the familiarity of researchers with established protocols, have maintained the popularity of radiolabels in many biological applications. However, radiolabels pose long-term health risks, including increased danger of cancer and reproductive problems. These hazards have led to governmental regulations and concomitant, time-consuming paperwork. The inherent lack of spatial resolution in their radiation measurement has necessitated the development of other label and detection schemes for the high-density DNA chips. As both the density of probes on DNA chips and the cost of handling and disposing of radioactive waste have increased, radioactive labels have given way to fluorescence labels in popularity.

5.2 Fluorescent Labels

The most frequently used nonradioactive method to detect hybridization is fluorescence. In this technique, a fluorescent label is attached to the target DNA before hybridization. After hybridization, the chip is washed to remove any unbound target and a light source sequentially or simultaneously illuminates all the probe sites on the array. Where the target has hybridized to the probes, the fluorescent labels absorb photons from the light source and then emit photons at a longer wavelength. Theoretically, for a given excitation intensity, the measured fluorescence intensity is linearly related to the concentration of labels and so the technique should be quantitative. However, the actual fluorescence intensity is related to the microenvironment of the fluorophore, and careful control and calibration must be exercised to achieve quantitative results.

When broadband sources such as Hg or Xe lamps are used, excitation filters must be used to limit the bandwidth of the excitation light. This means that the amount of light available for excitation is considerably less than the total output of the source. The emitted (fluorescence) photons are monitored through wavelength-dependent filters that block photons from the excitation source but pass the longer-wavelength fluorescence photons. The most critical feature in filtering is to assure that the wavelength of photons transmitted through the excitation filter does not overlap with the wavelength of photons passed by the detection filter.

Lasers are often used as light sources because they provide an intense source of photons in a narrow bandwidth, eliminating the need for an excitation filter and allowing the entire output of the laser to be available for excitation. Other advantages of laser sources over lamp sources include a simpler alignment procedure and longer time before failure. However, unlike lamps, which, in combination with excitation filters, can produce outputs over a wide range of wavelengths, lasers operate at only one or, at most, a few specific wavelengths. In order to select the appropriate combination of laser, dye, and detection filter, it is necessary to know the wavelength of the laser output and both the absorption and emission spectrum of the dye. Dye absorption bands may be quite broad and it is not necessary to have the laser wavelength exactly matching the peak of the absorption. Some frequently used dyes are shown in Table 1 together with the wavelength of their absorption and emission

Table 1 Characteristics of some dyes used for labeling DNA

Dye	Chemical nature	Absorption peak (nm)	Emission peak (nm)	Excitation laser (wavelength, nm)
Alexa® 488	Sulfonated rhodamine derivatives	491	515	Ar^+ (488)
Bodipy® 630/650	Substituted 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene derivative	625	640	He : Ne (633)
Cy3™	Cyanine	552	565	He : Ne (543)
Cy5™	Cyanine	650	667	He : Ne (633)
FAM™	Carboxyfluorescein	495	535	Ar^+ (488)
FITC	Fluorescein isothiocyanate	490	520	Ar^+ (488)
Fluor X™	–	494	520	Ar^+ (488)
JOE™	Carboxy-4′,5′-dichloro-2′,7′-dimethoxyfluorescein	526	548	He : Ne (543)
ROX™	Carboxy-X-rhodamine	580	605	He : Ne (594)
TAMRA™	Carboxytetramethylrhodamine	559	578	He : Ne (543)
Tetramethylrhodamine	–	546	572	He : Ne (543)
Texas Red®	Rhodamine-based dye	578	602	He : Ne (594)

Ar^+, argon ion laser; He : Ne, helium : neon laser.

maxima. The lasers most commonly used to excite these dyes in commercial instruments are the helium : neon laser and the argon ion laser. Table 1 shows which of these lasers and which of their wavelengths can be used to excite each dye. This table does not present an exhaustive list; several other dyes and other types of laser can be used.

The above discussion has assumed a continuous wave laser, i.e. a laser that operates continuously. It is also possible to use pulsed lasers with time-resolved fluorescence labels to separate the excitation photons from the fluorescence photons. These labels have long half-lives (several tenths of a nanosecond or longer) and are excited using a pulsed light source such as a Q-switched laser or a continuous wave laser with fast optical switches. Light detection is time gated to occur after the pulsed excitation has disappeared and after any background from short-lived fluorescent species has decayed. This gives excellent isolation of the desired signal without the use of wavelength-dependent filters but is somewhat more complex and less sensitive than the continuous wave technique. Rare earth chelates such as europium cryptate[74] are commonly used for time-resolved fluorescent labels.

5.2.1 Multiplexing

The use of multiple DNA targets in a single hybridization experiment is called multiplexing. When fluorescence detection is used, multiplexing requires a different dye to label each DNA target. It is necessary that each dye emits in a different spectral region that can be isolated with appropriate filters. For example, a fluorescein dye such as FAM™ or FITC emits in the green and is common for single-dye applications. A red fluorescing rhodamine such as ROX™ or Texas Red® might be selected as the second dye. If third or fourth dyes are desired, there are blue fluorescing dyes, such as Cascade blue, and deep red or infrared fluorescing dyes. Beyond a multiplicity of four, combinatorially or ratiolabeled probes are required.[75–78]

5.2.2 Image Collection

Because of the small dimensions of probes on DNA chips and the low concentration of hybridized targets on these probes, a high-power lens is needed to collect the fluorescence efficiently. Typically, a confocal microscope is used to focus the excitation light and collect the emitted light from an extremely small spot on the DNA chip. After light is collected by the microscope, it is passed through the detection filter and detected by a photomultiplier or avalanche photodiode. In order to create an image of the chip with information about all the probe sites, it is necessary to translate (scan) the DNA chip, relative to the microscope objective, in two dimensions. This technique is used in many commercial devices.

It is also possible to use a lower-power magnification lens and image an entire array onto a CCD detector. CCD detectors have a good response at longer wavelengths and may be especially useful for dyes emitting in the far-red and near-infrared spectral regions. Most of the dyes that emit in this region can be excited with red light, opening the possibility of using inexpensive diode lasers for excitation. The main problem associated with such imaging techniques is that light-collection efficiency is lower. To overcome this problem, large-area CCD arrays can be located in very close proximity to the detector. This essentially uses each pixel in the CCD array as a detector for a specific area on the chip. A similar technique, using microelectronic techniques to incorporate multiple phototransistors into an array detector, has been described at low densities.[79]

5.3 Mass Spectrometry

In mass spectrometric analysis of DNA chips, the chip is mounted in a vacuum chamber after hybridization and a portion of the material at a probe site is subjected to bombardment from a pulsed laser beam or an ion beam. This process fragments the surface molecules, desorbs them from the surface, and ionizes a small fraction of them. The ions are separated according to their mass using a time-of-flight magnetic sector or quadrupole mass filter and are then detected by an ion detector. This provides mass information about the material at the probe site and can be used to determine if hybridization has occurred.

5.3.1 Matrix-assisted Laser Desorption/Ionization

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry uses a short laser pulse to desorb and ionize a portion of the sample at a particular site. Molecular fragmentation, which normally accompanies laser desorption ionization, can be greatly reduced and desorption efficiency enhanced by proper selection of the matrix, or substrate, on which the sample is mounted. In particular, the discovery of 3-hydroxypicolinic acid as the matrix compound in MALDI[80] allows intact nucleotides to be analyzed. As mass spectrometers can easily resolve the mass difference between nucleotides, which is at least 9 atomic mass units, this has led to a number of sequencing methods and assays for point mutations and polymorphisms.[81] In experiments on DNA arrays, it has been shown that laser desorption removes only the hybridized target and leaves the bound probe attached to the surface.[82,83] This technique can be more sensitive than fluorescence and it does not require attaching a label to the target DNA.

5.3.2 Secondary Ion Mass Spectrometry

Secondary ion mass spectrometry (SIMS) uses a short pulse of ions to sputter a portion of the sample from the sample surface. SIMS is somewhat more destructive than MALDI mass spectrometry and considerable fragmentation occurs. Nevertheless, nucleotides can be identified in most cases. One technique takes advantage of the fragmentation observed in SIMS and the chemical difference between peptide nucleic acid and DNA to provide detection of DNA targets to peptide nucleic acid probes (T.J. Whitaker et al., unpublished data).[84] This technique measures the PO_2^- or PO_3^- fragments from the phosphate in the DNA backbone. As peptide nucleic acids do not contain phosphorus, these fragments only arise if DNA has hybridized to the probe. This technique has the advantage of potentially being simpler than other mass spectrometry methods because it measures only one or two masses and these are low masses requiring modest mass resolution.

SIMS and MALDI techniques both analyze ions created during the desorption process. It is also possible to use a laser to ionize the neutral component of the desorbed material. This gives a larger signal because most of the desorbed material is neutral, and it essentially eliminates the matrix dependence of the signal, which can be important if quantitative information is needed. A selective form of this process called sputter-initiated resonance ionization spectroscopy (SIRIS) has been used to analyze Sn-labeled DNA hybridized to oligonucleotides on a DNA chip.[84] Multiplexing was achieved by using different enriched stable isotopes of Sn to label different DNAs. Tin has eight affordable, stable isotopes that can be incorporated in an organic substance that is readily attached to an oligonucleotide.[85]

5.4 Electronic Detection

DNA may be an insulator or a conductor depending on whether electron donors are present and whether the DNA is single or double stranded. Direct electronic detection of a nucleic acid hybrid has been demonstrated using an electron donor, ferrocene, attached to a ''signaling probe'' that binds to target DNA.[86] Capture probes are attached to arrayable electrodes via phenylacetylene (ethynylbenzene) ''molecular wires'', providing electrical conductance between the electrode and the probe. When the target DNA hybridizes to a complementary probe, a small voltage applied to the electrode causes the ferrocene molecules present on the signaling probe as part of this sandwich assay to release electrons, which are transduced to the electrode and are detected. The insulating layer on the electrode eliminates nonspecific signal; only those ferrocenes brought into proximity of the electrode by virtue of specific hybridization are able to generate detectable signal. The current generated is detected with standard electronic circuits.

Another electronic detection scheme utilizes the insulating properties of DNA. This technique measures the capacitance change when target DNA hybridizes to a probe attached to an insulating monolayer on an electrode (T.J. Whitaker, private communication regarding research at Atom Sciences, Inc.). The electrode and insulating monolayer are submerged in a conducting solution, which acts as the second electrode in the capacitance circuit. As target DNA binds to the probe, the effective thickness of the dielectric increases and the capacitance decreases. A similar technique has been used successfully to detect albumin as it formed a complex with its antibody.[40] The maximum change in capacitance from the antigen–antibody binding was approximately $100\,nF\,cm^{-2}$, which is easily measured on 1-mm diameter electrodes.

5.5 Other Detection Methods

Several other techniques have been used to detect the formation of hybrids. Colorimetric and chemiluminescent techniques are common and the appropriate labels and reagents are commercially available. Some newer optical techniques do not require a label. For example, surface plasmon resonance is a sensitive technique based upon the reflectivity of a laser beam from the interface between a metallic surface and a bulk solution. Coupling of plasmons into the metallic surface reduces the reflectance and a minimum will be observed at a particular incidence angle. This angle is a sensitive function of the index of refraction of materials at the interface. Either the angle of minimum reflectance or the reflectance itself can be used to monitor changes at the interface. In a DNA array, hybridization can be monitored in real-time by this technique.[87] Commercial instruments are available.

A related technique brings in the light from the substrate. The substrate is first coated with a layer that has a low index of refraction and then with a layer having a high index. At a particular angle of incidence, the light couples through these layers and an evanescent wave interacts with molecules on the surface. This resonant angle is sensitive to the refractive index and thickness of molecules on the surface and has been used to detect DNA hybridization.[88] Again, commercial instruments based on this technique are available.

Biotin-labeled DNA and RNA have been bound to streptavidin molecules attached to magnetic particles.[89] These particle-labeled nucleic acids will hybridize to complementary targets in the normal way. After hybridization, detection, quantification, and molecular sizing of the targets are accomplished by adapting techniques

developed in the disk-drive magnetic recording industry. Inexpensive detection is the key advantage of the magnetic particle labeling technology.

6 BIOINFORMATICS

Over 30 000 of the approximately 100 000 human genes have been mapped.[90] New data are being gathered, not just about the sequence of genes and proteins, but also about their function, the distribution of polymorphisms in the population, and other details. This information deluge is stretching our ability to make practical use of all the data. Bioinformatics refers to the creation and use of computer databases to facilitate biological analyses. The term is sometimes extended to the use of algorithmic tools, an activity more appropriately called ''computational biology''. Obviously, the design of a DNA chip will determine the degree to which it requires sophisticated access to databases to interpret the results it generates. High-density chips with hundreds of thousands or millions of probes representing every possible sequence will require such sophisticated tools and access to large databases. There are literally hundreds of bioinformatics tools available over the internet, often free-of-charge, but few are applicable to chip-based analyses.

At the other extreme, low-density chips aimed at determining one or only a few specific genes may contain only a few hundred probes and will be much easier to interpret. All the data necessary for interpreting these chips can be contained on a desktop computer.

7 SUMMARY

The utilization of DNA arrays for sequence checking, medical diagnosis, identification of microorganisms, and for DNA fingerprinting is a rapidly developing field and we presume that much of the information presented above will be outdated in a very few years. A number of companies have established areas in which they will participate in this new field and their internet web sites are shown in Table 2. It may be useful to investigate the progress that these companies and other laboratories

Table 2 Companies supplying DNA array-related products or services

Company	Arrays	Detector	Software	Service	Internet URL
Affinity Sensors		X			affinity-sensors.com
Affymetrix	X	X	X		affymetrix.com
Alpha Innotech		X			alphainnotech.com
Alphagene	X				alphagene.com
Amersham Pharmacia Biotech	X	X			www.apbiotech.com
Atom Sciences Inc.		X			atom-sci.com
Axon Instruments Inc.		X	X		axon.com
Biacore	X	X	X		biacore.com
Biodiscovery			X		biodiscovery.com
BioDot, Inc.	X				biodot.com
Biomedical Photometrics Inc.		X			confocal.com
Biorobotics Ltd	X				biorobotics.com
Cartesian Technologies Inc.		X			cartesiantech.com
Clinical Micro Sensors Inc.	X	X			microsensor.com
Clontech Laboratories Inc.	X	X			clontech.com/
Display Systems Biotech			X		displaysystems.com
Gene Logic Inc.	X		X		oncormed.com/accdd.htm
Gene Machines	X				genemachines.com
GeneData			X		genedata.com
GeneFocus™		X			genefocus.com
Genosys	X				genosys.com
Engineering Services Inc.	X				esit.com
Genetic Microsystems™	X	X	X		geneticmicro.com
Genisphere	X				genisphere.com
Genome Systems	X	X			genomesystems.com
Genomic Solutions™	X	X			genomicsolutions.com
Genometrix Inc.	X	X	X		genometrix.com
Genpak Ltd	X				genpakdna.com
GSI Lumonics		X			gsilumonics.com
Hyseq	X		X		hyseq.com

(*continued overleaf*)

Table 2 (*continued*)

Company	Arrays	Detector	Software	Service	Internet URL
Incyte Inc.	X	X	X		incyte.com
Interactiva	X			X	interactiva.de
IRIS Biotechnologies Inc.	X				irisbiotech.com
MicroFab Technologies Inc.	X				microfab.com
Molecular Probes Inc.	X				probes.com
Nanogen Inc.	X	X	X		nanogen.com
NEN™ Life Science Products	X				nenlifesci.com
Pangea Corp.			X		pangea.com
Phoretix International			X		phoretix.com
Protogene Inc.			X		protogene.com
Radius Biosciences	X				ultranet.com/~radius
Research Genetics Inc.				X	resgen.com/
Scanalytics Inc.		X			scanalytics.com
Silicon Genetics			X		sigenetics.com
Technology Mentors Inc.				X	technologymentors.com
TeleChem International	X	X	X		arrayit.com
Universal Imaging Corp.		X			image1.com
Vysis Inc.	X	X			vysis.com

have made before making choices on the strategy of a particular use or manner of preparation of DNA arrays. The basic approaches outlined in this article are based on well-established chemical properties of DNA hybrids, but the technologies to construct DNA arrays and to locate sites of hybridization can be and are being pursued through many avenues.

ACKNOWLEDGMENTS

The authors wish to thank Dr Ken Willey for helpful discussions. One of the co-authors, Mitchel Doktycz, is a contractor of the US Government under contract no. DE-AC05-96OR22464. Accordingly, the US Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes. Oak Ridge National Laboratory is managed by Lockheed Martin Energy Research Corp. for the US Department of Energy under contract number DE-AC05-96OR22464.

TRADENAMES

GeneChip® is a registered trademark of Affymetrix Inc., Santa Clara, CA, USA. Alexa® and Bodipy® are registered trademarks of Molecular Probes Inc., Eugene, OR, USA and Leiden, The Netherlands. Cy3™, Cy5™, Fluor X™, and Texas Red® are trademarks of Amersham Pharmacia Biotech, Piscataway, NJ, USA and Uppsala, Sweden. FAM™, JOE™, ROX™, and TAMRA™ are trademarks of Perkin-Elmer, LLC, Norwalk, CT, USA.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
cDNA	Complementary DNA
ddNTP	Dideoxynucleoside Triphosphate
FITC	Fluorescein Isothiocyanate
MALDI	Matrix-assisted Laser Desorption/Ionization
mRNA	Messenger RNA
NTP	Deoxynucleoside Triphosphate
PCR	Polymerase Chain Reaction
SIMS	Secondary Ion Mass Spectrometry
SIRIS	Sputter-initiated Resonance Ionization Spectroscopy
ssDNA	Single-stranded DNA

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry • Pharmacogenetic Testing

Forensic Science **(Volume 5)**
DNA Extraction Methods in Forensic Analysis • Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Fluorescence In Situ Hybridization • Mass Spectrometry of Nucleic Acids • PNA and Its Applications • Polymerase Chain Reaction and Other Amplification Systems • RNA Tertiary Structure • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Sequencing Strategies and Tactics in DNA and RNA Analysis

REFERENCES

1. D.J. Graves, 'Powerful Tools for Genetic Analysis Come of Age', *Trends Biotech.*, **17**, 127–134 (1999).
2. R. Drmanac, R. Crkvenjakov, Yugoslav Patent Application, 570, 1987.
3. W. Bains, G.C.J. Smith, 'A Novel Method for Nucleic-acid Sequence Determination', *J. Theor. Biol.*, **135**(3), 303–307 (1988).
4. Y.P. Lysov, V.L. Florentiev, A.A. Khorlin, K.R. Khrapko, V.V. Shik, A.D. Mirzabekov, 'A New Method for Determining the DNA Nucleotide Sequence by Hybridization with Oligonucleotides', *Dokl. Akad. Nauk*, **303**, 1508–1511 (1988).
5. E. Southern, United Kingdom Patent Application GB8810400, 1988.
6. S.C. Macevicz, International Patent Application PS US89 04741, 1989.
7. C.R. Cantor, A.D. Mirzabekov, E. Southern, 'Report on the Sequencing by Hybridization Workshop', *Genomics*, **13**, 1378–1383 (1992).
8. E.S. Lander, 'Array of Hope', *Nat. Genet.*, **21**, 3–4 (1999).
9. J.L. DeRisi, V.R. Iyer, P.O. Brown, 'Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale', *Science*, **278**, 680–686 (1997).
10. D.L. Bowtell, 'Options Available from Start to Finish for Obtaining Expression Data by Microarray', *Nat. Genet.*, **21**, 25–32 (1999).
11. D.A. Benson, M.S. Boguski, D.J. Lipman, J. Ostell, 'GenBank', *Nucl. Acids Res.*, **25**, 1–6 (1997).
12. M.S. Boguski, T.M. Lowe, C.M. Tolstoshev, 'dbEST-database for Expressed Sequence Tags', *Nat. Genet.*, **4**, 332–333 (1993).
13. G.D. Schuler, M.S. Boguski, E.A. Stewart, 'A Gene Map of the Human Genome', *Science*, **274**, 540–546 (1996).
14. M.J. Doktycz, K.L. Beattie, 'Genosensors and Model Hybridization Studies', in *Automation Technologies for Genome Characterization*, ed. T.J. Beugelsdijk, John Wiley & Sons, New York, 205–225, 1997.
15. M.J. Doktycz, M.D. Morris, S.J. Dormady, K.L. Beattie, K.B. Jacobson, 'Optical Melting of 128 Octamer DNA Duplexes', *J. Biol. Chem.*, **270**(15), 8439–8445 (1995).
16. R. Owczarzy, P.M. Vallone, F.J. Gallo, T.M. Paner, M.J. Lane, A.S. Benight, 'Predicting Sequence-dependent Melting Stability of Short Duplex DNA Oligomers', *Biopolymers*, 217–239 (1998).
17. J. Santa Lucia, Jr, 'A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics', *Proc. Natl. Acad. Sci. USA*, **95**, 1460–1465 (1998).
18. J.G. Wetmur, 'DNA Probes: Applications of the Principles of Nucleic Acid Hybridization', *Crit. Rev. Biochem. Mol. Biol.*, **26**, 227–259 (1991).
19. D.J. Lockhart, H. Dong, M.C. Byrne, M.T. Folletie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, E.L. Brown, 'Expression Monitoring by Hybridization to High-density Oligonucleotide Arrays', *Nat. Biotech.*, **14**, 1675–1680 (1996).
20. L. Wodicka, H. Dong, M. Mittmann, M.-H. Ho, D.J. Lockhart, 'Genome-wide Expression Monitoring in *Saccharomyces cerevisiae*', *Nat. Biotech.*, **15**, 1359–1367 (1997).
21. K.L. Beattie, 'Genomic Fingerprinting Using Oligonucleotide Arrays', in *DNA Markers Protocols, Applications, and Overviews*, eds. G. Caetano-Anolls, P.M. Gresshoff, Wiley-VCH, New York, Chapter 14, 213–224, 1997.
22. J. Weiler, H. Gausepohl, N. Hauser, O.N. Jensen, J.D. Hoheisel, 'Hybridisation Based DNA Screening on Peptide Nucleic Acid (PNA) Oligomer Arrays', *Nucl. Acids Res.*, **25**(14), 2792–2799 (1997).
23. S.C. Case-Green, E.M. Southern, 'Studies on the Base Pairing Properties of Deoxyinosine by Solid Phase Hybridisation to Oligonucleotides', *Nucl. Acids Res.*, **22**(2), 131–136 (1994).
24. Z. Guo, Q. Liu, L.M. Smith, 'Enhanced Discrimination of Single Nucleotide Polymorphisms by Artificial Mismatch Hybridization', *Nat. Biotech.*, **15**, 331–335 (1997).
25. D. Loakes, D.M. Brown, '5-Nitroindole as an Universal Base Analog', *Nucl. Acids Res.*, **22**(20), 4039–4043 (1994).
26. S. Dubiley, E. Kirillov, Y. Lysov, A. Mirzabekov, 'Fractionation, Phosphorylation and Ligation on Oligonucleotide Microchips to Enhance Sequencing by Hybridization', *Nucl. Acids Res.*, **25**(12), 2259–2265 (1997).
27. K.L. Gunderson, X.C. Huang, M.S. Morris, R.J. Lipshutz, D.J. Lockhart, M.S. Chee, 'Mutation Detection by Ligation to Complete *n*-mer DNA Arrays', *Genome Res.*, **8**, 1142–1153 (1998).
28. N.E. Broude, T. Sano, C.L. Smith, C.R. Cantor, 'Enhanced DNA Sequencing by Hybridization', *Proc. Natl. Acad. Sci. USA*, **91**, 3072–3076 (1994).
29. T.T. Nikiforov, R.B. Rendle, P. Goelet, Y.-H. Rogers, M.L. Kotewicz, S. Anderson, G.T. Trainor, M.R. Knapp, 'Genetic Bit Analysis: a Solid Phase Method for Typing Single Nucleotide Polymorphisms', *Nucl. Acids Res.*, **22**, 4167–4175 (1994).

30. T. Pastinen, A. Kurg, A. Metspalu, L. Peltonen, A.-C. Syvänen, 'Minisequencing: a Specific Tool for DNA Analysis and Diagnostics on Oligonucleotide Arrays', *Genome Res.*, **7**, 606–614 (1997).
31. J.M. Shumaker, A. Metspalu, C.T. Caskey, 'Mutation Detection by Solid-phase Primer Extension', *Hum. Mutat.*, **7**, 346–354 (1996).
32. S. Parinov, V. Barsky, G. Yershov, E. Kirillov, E. Timofeev, A. Belgovskiy, A. Mirzabekov, 'DNA Sequencing by Hybridization to Microchip Octa- and Decanucleotides Extended by Stacked Pentanucleotides', *Nucl. Acids Res.*, **24**(15), 2998–3004 (1996).
33. R. Maldonado-Rodriguez, M. Espinosa-Lara, P. Loyola-Abitia, W.G. Beattie, K.L. Beattie, 'Mutation Detection by Stacking Hybridization on Genosensor Arrays', *Mol. Biotech.*, **11**, 13–25 (1999).
34. E. Gentalen, M. Chee, 'A Novel Method for Determining Linkage between DNA Sequences: Hybridization to Paired Probe Arrays', *Nucl. Acids Res.*, **27**(6), 1485–1491 (1999).
35. J.A. Ferguson, T.C. Boles, C.P. Adams, D.R. Walt, 'A Fiber-optic DNA Biosensor Microarray for the Analysis of Gene Expression', *Nat. Biotech.*, **14**, 1681–1684 (1996).
36. J.B. Lamture, K.L. Beattie, B.E. Burke, M.D. Eggers, D.J. Ehrlich, R. Fowler, M.A. Hollis, B.B. Kosicki, R.K. Reich, S.R. Smith, R.S. Varma, M.E. Hogan, 'Direct Detection of Nucleic Acid Hybridization on the Surface of a Charge Coupled Device', *Nucl. Acids Res.*, **22**(11), 2121–2125 (1994).
37. T. Vo Dinh, 'Development of a DNA Biochip: Principle and Applications', *Sens. Actuators B*, **51**, 52–59 (1998).
38. R.G. Nuzzo, D.L. Allara, 'Adsorption of Bifunctional Organic Disulfides on Gold Surfaces', *J. Am. Chem. Soc.*, **105**(13), 4481–4483 (1983).
39. C. Krause, V.M. Mirsky, K.D. Heckmann, 'Capacitive Detection of Surfactant Adsorption on Hydrophobized Gold Electrodes', *Langmuir*, **12**(25), 6059–6064 (1996).
40. V.M. Mirsky, M. Riepl, O.S. Wolfbeis, 'Capacitive Monitoring of Protein Immobilization and Antigen–Antibody Reactions on Monomolecular Alkylthiol Films on Gold Electrodes', *Biosens. Bioelectron.*, **12**(9–10), 977–989 (1997).
41. M. Lyn, 'Inorganic Support Intermediates: Covalent Coupling of Enzymes, Antigens, Antibodies and Peptides', in *Immobilized Enzymes, Antigens, Antibodies and Peptides*, ed. H.H. Weetal, Marcel Dekker, New York, 12–39, 1978.
42. Y. Zhang, M.Y. Coyne, S.G. Will, C.H. Levenson, E.S. Kawasaki, 'Single-base Mutational Analysis of Cancer and Genetic Diseases Using Membrane Bound Modified Oligonucleotides', *Nucl. Acids Res.*, **19**(14), 3929–3933 (1991).
43. J. van Ness, S. Kalbfleisch, C.R. Petrie, M.W. Reed, J.C. Tabone, N.M.J. Vermeulen, 'A Versatile Solid Support System for Oligodeoxynucleotide Probe-based Hybridization Assays', *Nucl. Acids Res.*, **19**(12), 3345–3350 (1991).
44. R. Drmanac, Z. Strezoska, I. Labat, S. Drmanac, R. Crkvenjakov, 'Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides', *DNA Cell Biol.*, **9**, 527–534 (1990).
45. H. Hirayama, J. Tamoaka, K. Horikoshi, 'Improved Immobilization of DNA to Microwell Plates for DNA–DNA Hybridization', *Nucl. Acids Res.*, **24**(20), 4098–4099 (1996).
46. K.R. Khrapko, Y.P. Lysov, A.A. Khorlin, I.B. Ivanov, G.M. Yershov, S.K. Vasilenko, V.L. Florentiev, A.D. Mirzabekov, 'A Method for DNA Sequencing by Hybridization with Oligonucleotide Matrix', *DNA Seq.*, **1**, 375–388 (1991).
47. R.S. Matson, J.B. Rampal, P.J. Coassin, 'Biopolymer Synthesis on Polypropylene Supports', *Anal. Biochem.*, **217**, 306–310 (1994).
48. E.M. Southern, U. Maskos, J.K. Elder, 'Analyzing and Comparing Nucleic Acid Sequences by Hybridization to Arrays of Oligonucleotides: Evaluation Using Experimental Models', *Genomics*, **13**, 1008–1017 (1992).
49. E.M. Southern, S.C. Case-Green, J.K. Elder, M. Johnson, K.U. Mir, L. Wang, J.C. Williams, 'Arrays of Complementary Oligonucleotides for Analyzing the Hybridisation Behaviour of Nucleic Acids', *Nucl. Acids Res.*, **22**(8), 1368–1373 (1994).
50. A.P. Blanchard, R.J. Kaiser, L.E. Hood, 'High-density Oligonucleotide Arrays', *Biosens. Bioelectron.*, **11**(6/7), 687–690 (1996).
51. S.P.A. Fodor, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, D. Solas, 'Light-directed, Spatially Addressable Parallel Chemical Synthesis', *Science*, **251**, 767–773 (1991).
52. A.C. Pease, D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, S. Fodor, 'Light-generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis', *Proc. Natl. Acad. Sci. USA*, **91**, 5022–5026 (1994).
53. G. McGall, J. Labadie, P. Brock, G. Wallraff, T. Nguyen, W. Hinsberg, 'Light-directed Synthesis of High-density Oligonucleotide Arrays Using Semiconductor Photoresists', *Proc. Natl. Acad. Sci. USA*, **93**, 13 555–13 560 (1996).
54. M. Schena, D. Shalon, R.W. Davis, P.O. Brown, 'Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray', *Science*, **270**, 467–470 (1995).
55. J. Macas, M. Nouzova, D.W. Galbraith, 'Adapting the Biomek® 2000 Laboratory Automation Workstation for Printing DNA Microarrays', *BioTechniques*, **25**, 106–110 (1998).
56. V.G. Cheung, M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, G. Childs, 'Making and Reading Microarrays', *Nat. Genet.*, **21**, 15–19 (1998).
57. W.G. Beattie, L. Meng, S.L. Turner, R.S. Varma, D.D. Dao, K.L. Beattie, 'Hybridization of DNA Targets to Glass-tethered Oligonucleotide Probes', *Mol. Biotech.*, **4**, 213–225 (1995).

58. D.J. Graves, H.-J. Su, S.E. McKenzie, S. Surrey, P. Fortina, 'System for Preparing Microhybridization Arrays on Glass Slides', *Anal. Chem.*, **70**, 5085–5092 (1998).
59. D.I. Stimpson, P.W. Cooley, S.M. Knepper, D.B. Wallace, 'Parallel Production of Oligonucleotide Arrays Using Membranes and Reagent Jet Printing', *BioTechniques*, **25**, 886–890 (1998).
60. R. Papen, K. Croker, A. Kolb, 'Nanoliter Dispensing Technology–Piezoelectric Pipetting Allows 0.2 µl-per-second Delivery', *Genet. Eng. News*, **18**(4), 16–17 (1998).
61. A.V. Lemmo, J.T. Fisher, H.M. Geysen, D.J. Rose, 'Characterization of an Inkjet Chemical Microdispenser for Combinatorial Library Synthesis', *Anal. Chem.*, **69**, 543–551 (1997).
62. K.L. Beattie, W.G. Beattie, L. Meng, S.L. Turner, R. Coral-Vazques, D.D. Smith, P.M. McIntyre, D.D. Dao, 'Advances in Genosensor Research', *Clin. Chem.*, **41**(5), 700–706 (1995).
63. U. Maskos, E.M. Southern, 'Oligonucleotide Hybridisations on Glass Supports: a Novel Linker for Oligonucleotide Synthesis and Hybridisation Properties of Oligonucleotides Synthesised In Situ', *Nucl. Acids Res.*, **20**(7), 1679–1684 (1992).
64. Z. Guo, R.A. Guilfoyle, A.J. Thiel, R. Wang, L.M. Smith, 'Direct Fluorescence Analysis of Genetic Polymorphisms by Hybridization with Oligonucleotide Arrays on Glass Supports', *Nucl. Acids Res.*, **22**(24), 5456–5465 (1994).
65. B. Joos, H. Kuster, R. Cone, 'Covalent Attachment of Hybridizable Oligonucleotides to Glass Supports', *Anal. Biochem.*, **247**, 96–101 (1997).
66. W.L. Wood, J. Gitschier, L.A. Lasky, R.M. Lawn, 'Base Composition-independent Hybridization in Tetramethylammonium Chloride: a Method for Oligonucleotide Screening of Highly Complex Gene Libraries', *Proc. Natl. Acad. Sci. USA*, **82**, 1585–1588 (1985).
67. W.A. Rees, T.D. Yagfer, J. Korte, P.H. von Hippel, 'Betaine can Eliminate the Base Pair Composition Dependence of DNA Melting', *Biochemistry*, **32**, 137–144 (1993).
68. T. van Ness, L. Chen, 'The Use of Oligonucleotide Probes in Chaotrope-based Hybridization Solutions', *Nucl. Acids Res.*, **19**, 5143–5151 (1991).
69. M.J. Doktycz, K.B. Jacobson, K.L. Beattie, R.S. Foote, 'Optical Melting as a Tool for Optimizing SBH Analysis of DNA', in *Proceedings of SPIE: Ultrasensitive Instrumentation for DNA Sequencing and Biochemical Diagnostics*, eds. G.E. Cohn, J.M. Lerner, K.J. Liddane, A. Scheeline, S.A. Soper, Proc. SPIE, San Jose, CA, 2386, 30–34, 1995.
70. U. Maskos, E.M. Southern, 'Parallel Analysis of Oligodeoxyribonucleotide (Oligonucleotide) Interactions, I. Analysis of Factors Influencing Oligonucleotide Duplex Formation', *Nucl. Acids Res.*, **20**, 1675–1678 (1992).
71. K.A. Jacobs, R. Rudersdorf, S.D. Neill, J.P. Dougherty, J.P. Brown, E.L. Brown, E.F. Fritsch, 'The Thermal Stability of Oligonucleotide Duplexes is Sequence Independent in Tetramethylammonium Salt Solutions: Application to Identifying Recombinant DNA Clones', *Nucl. Acids Res.*, **16**, 4637–4650 (1988).
72. J. Cheng, E.L. Sheldon, L. Wu, A. Uribe, L.O. Gerrue, J. Carrino, M. Heller, J. O'Connell, 'Electric Field Controlled Preparation and Hybridization Analysis of DNA/RNA from *E. coli* on Microfabricated Bioelectronic Chips', *Nat. Biotech.*, **16**, 541–546 (1998).
73. C. Edman, D. Raymond, D. Wu, E. Tu, R. Sosnowski, W. Butler, M. Nerenberg, M. Heller, 'Electric Field Directed Nucleic Acid Hybridization on Microchips', *Nucl. Acids Res.*, **25**, 4907–4914 (1998).
74. O. Prat, E. Lopez, G. Mathis, 'Europium(III) Cryptate: a Fluorescent Label for the Detection of DNA Hybrids on Solid Support', *Anal. Biochem.*, **195**(2), 283–289 (1991).
75. P.M. Nederlof, S. van der Flier, J. Wiegant, A.K. Raap, H.J. Tanke, J.S. Ploem, M. van der Ploeg, 'Multiple Fluorescence In Situ Hybridization', *Cytometry*, **11**, 126–131 (1990).
76. P.M. Nederlof, S. van der Flier, J. Vrolijk, H.J. Tanke, A.K. Raap, 'Fluorescence Ratio Measurements of Double-labeled Probes for Multiple In Situ Hybridization by Digital Imaging Technology', *Cytometry*, **13**(8), 839–845 (1992).
77. J.G. Dauwerse, J. Wiegant, A.K. Raap, M.H. Breuning, G.J.B. van Ommen, 'Multiple Colors by Fluorescence In Situ Hybridization Using Ratio-labelled DNA Probes Create a Molecular Karyotype', *Hum. Mol. Genet.*, **1**, 593–598 (1992).
78. T. Ried, A. Baldini, T. Rand, D.C. Ward, 'Simultaneous Visualization of Seven Different DNA Probes by In Situ Hybridization Using Combinatorial Fluorescence and Digital Imaging Microscopy', *Proc. Natl. Acad. Sci. USA*, **89**, 1388–1392 (1992).
79. T. Vo-Dinh, J.P. Alarie, N. Isola, D. Landis, A.L. Wintenberg, M.N. Erikson, 'DNA Biochip Using a Phototransistor Integrated Circuit', *Anal. Chem.*, **71**(2), 358–363 (1999).
80. K.J. Wu, A. Steding, C.H. Becker, 'Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry of Oligonucleotides Using 3-Hydroxypicolinic Acid as an Ultraviolet Sensitive Matrix', *Rapid. Commun. Mass Spectrom.*, **9**, 1792–1796 (1993).
81. P. Ross, L. Hall, I. Smirnov, L. Haff, 'High Level Multiplex Genotyping by MALDI/TOF Mass Spectrometry', *Nat. Biotech.*, **16**, 1347–1351 (1998) [and references therein].
82. D.P. Little, T.J. Cornish, M.J. O'Donnell, A. Braun, R.J. Cotter, H. Koster, 'MALDI on a Chip: Analysis of Arrays of Low- to Sub-femtomole Quantities of Synthetic Oligonucleotides and DNA Diagnostic Products Dispensed by a Piezo-electric Pipette', *Anal. Chem.*, **69**, 4540–4546 (1997).
83. M.J. O'Donnell, K. Tang, H. Koster, C.L. Smith, C.R. Cantor, 'High Density, Covalent Attachment of DNA

to Silicon Wafers for Analysis by MALDI/TOF Mass Spectrometry', *Anal. Chem.*, **69**(13), 2438–2443 (1997).
84. H.F. Arlinghaus, X.-Q. Guo, M.N. Kwoka, K.B. Jacobson, 'Multiplexed DNA Sequencing and Diagnostics by Hybridization with Enriched Stable Isotope Labels', *Anal. Chem.*, **69**(8), 1510–1517 (1997).
85. F.V. Sloop, G.M. Brown, R.S. Foote, K.B. Jacobson, R.A. Sachleben, 'Synthesis of 3-(Triethylstannyl)propanoic Acid: an Organotin Mass Label for DNA', *Bioconjugate Chem.*, **4**, 406–409 (1993).
86. D.H. Farkus, 'Clinical Micro Sensors Developing Point-of-care Molecular Diagnostics', *IVD Technol.*, **5**(1), 42–43 (1999).
87. C.E. Jordan, A.G. Frutos, A.J. Thiel, R.M. Corn, 'Surface Plasmon Resonance Imaging Measurements of DNA Hybridization Adsorption and Streptavidin/DNA Multilayer Formation at Chemically Modified Gold Surfaces', *Anal. Chem.*, **69**, 4939–4947 (1997).
88. H.J. Watts, D. Yeung, H. Parkes, 'Real-time Detection and Quantification of DNA Hybridization by an Optical Biosensor', *Anal. Chem.*, **67**, 4283–4289 (1995).
89. L. Adelman, 'Magnetic Labeling Technology–Computer Drive Technology Enhances Biotech Assays', *Genet. Eng. News*, **18**(18), 36–37 (1998).
90. P. Deloukas, G.D. Schuler, G. Gyapay, 'A Physical Map of 30 000 Human Genes', *Science*, **282**, 744–746 (1998).

Drugs of Abuse, Analysis of

Amitava Dasgupta
University of Texas-Houston Medical School, Houston, USA

1 Introduction **2**
1.1 Alcohol Use and Abuse 2
1.2 Major Illegal Drugs Abused in the USA 3
2 Drug Testing **3**
2.1 Specimen Types for Abused Drug Testing 3
2.2 Federal Guidelines for Drugs of Abuse Testing: Cutoff Limits and Other Issues 3
2.3 Forensic Drug Testing 4
2.4 Immunoassays for Screening of Drugs in Urine 4
2.5 Common Adulterants used to Mask Drug Testing 4
2.6 Window of Detection and Detection Limits (Cutoffs) for Abused Drugs 5
2.7 True Positives in Drugs of Abuse Testing 5
3 Methods of Testing **6**
3.1 Thin-layer Chromatography in Testing for Abused Drugs 6
3.2 Gas Chromatography/Mass Spectrometry Confirmation and Quantification of Abused Drugs 6
3.3 Protocols for Gas Chromatography/Mass Spectrometry Analysis of Amphetamines 6
3.4 Gas Chromatography/Mass Spectrometry Analysis of Cocaine Metabolites 10
3.5 Gas Chromatography/Mass Spectrometry Confirmation of Opiates 12
3.6 Gas Chromatography/Mass Spectrometry Confirmation of Tetrahydrocannabinol (Marijuana) 13
3.7 Gas Chromatography/Mass Spectrometry Confirmation of Urinary Phencyclidine 14
4 Analysis of Drugs **14**
4.1 Analysis of Benzodiazepines 14
4.2 Date Rape Drug Rohypnol 14
4.3 Analysis of Barbiturates 14
4.4 Date Rape Drug γ-Hydroxybutyric Acid 15
4.5 Confirmation of Lysergic Acid Diethylamide 15
4.6 Hair Analysis for Drugs of Abuse 16
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

Alcohol is the most widely abused legal drug in the USA. Other widely abused illegal drugs are amphetamines, cocaine, cannabinoids, opiates, barbiturates, benzodiazepines, methadone, methoqualone and phencyclidine (PCP). Lysergic acid diethylamide (LSD), a popular drug in 1960, is coming back to the illegal market. Drug testing in the USA is either mandatory or nonmandatory. In the first group, an employer is required by federal regulation to test. Urine is the preferred specimen, although blood and gastric content are useful in detecting drugs in the case of an acute recent overdose. Screening of abused drugs by immunoassay has certain limitations. For example, several legal drugs used in over-the-counter cold medications contain phenylpropanolamine, ephedrine or pseudoephedrine. These drugs can cause positive results in amphetamine immunoassays because of their structural similarity to methamphetamine. Moreover, in the gas chromatography/mass spectrometry (GC/MS) confirmation of amphetamine and methamphetamine, if the temperature of the injector port is over 185 °C, ephedrine and pseudoephedrine can be dehydrated to methamphetamine, resulting in a false positive result. Alcohol levels in blood can be measured by either immunoassay or gas chromatography (GC). Many clinical toxicology laboratories use immunoassay for measuring blood alcohol because it is automated, speedy and a simple technique. False positive alcohol concentration may result from high concentrations of lactate dehydrogenase (LDH) (lactic acidosis) and lactate in a specimen.

However, this interference can be eliminated by measuring alcohol concentration in the protein-free ultrafiltrate. Alcohol is not bound to serum protein and appears in the ultrafiltrate. On the other hand, LDH has a molecular weight of 180 000 and is absent in the ultrafiltrate. GC is the method of choice for measuring legal blood alcohol.

Immunoassay screening tests provide only preliminary results. Moreover, a negative immunoassay result does not mean that no drug was present. It may indicate that the concentration of drug was below the cutoff limit. Guidelines have been developed by the Department of Health and Human Services (DHHS) for drug testing.

People try to beat drug testing by adding a variety of compounds such as salt, lemon juice, household

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

bleach and Visine eye drops to their urine. Therefore it is important to check sample integrity (temperature, pH, specific gravity). True positive results from ingestion of poppy seed cake (contains opiates), health inca tea (contains cocaine) or hemp seed oil (contains tetrahydrocannabinol (THC)) in drug testing have been reported. Confirmation of the abused drug is essential to follow up screening tests and should be done using a different analytical technique. GC/MS is considered the gold standard for the confirmation of abused drugs in biological fluids and tissue specimens. The pre-analytical steps include extraction of the drug from urine or serum matrix using an organic solvent followed by derivatization. The common derivatizing agents are trifluoroacetic anhydride, pentafluoropropionic anhydride, heptafluorobutyric anhydride, methanolic sulfuric acid, N-O-*bis(trimethylsilyl) trifluoroacetamide (BSTFA) etc. New derivatization techniques have appeared in the literature for GC/MS confirmation of abused drugs. The perfluorooctanoyl and 4-carbethoxyhexafluorobutyryl derivatives of amphetamines are less volatile than traditional fluoro derivatives, have higher molecular weights and are more useful for unambiguous identification. Propyl chloroformate and very recently 2,2,2-trichloroethyl chloroformate have been introduced as new derivatization reagents for amphetamines.*

Another challenge in testing drugs of abuse is the identification of designer drugs, which are usually derivatives of known drugs. Usually, screening test results using immunoassays are negative if designer drugs are present. For example, phenmetrazine at concentrations up to 6000 ng mL^{-1} in urine would be tested negative by the EMIT (Enzyme Multiplied Immunoassay Technique, Behring Diagnostics, San Jose, CA). Recently γ-hydroxybutyric acid (GABA) has been reported as a date rape drug. This drug can be identified in blood and urine using high-performance liquid chromatography/mass spectrometry (HPLC/MS).

Hair analysis is becoming popular owing to the long-term deposition of many abused drugs in hair. However, the analytical challenges are greater as the concentrations are low. This is especially true for Δ^9-THC and LSD where the sensitivity of a bench model GC/MS is not sufficient for analysis.

1 INTRODUCTION

Abuse of drugs is a critical problem in health care in the USA and also in the rest of the world. Like caffeine and nicotine, alcohol (ethanol) is a social drug and is rarely used in therapy (to treat ethylene glycol and methanol overdose). Of all the illegal drugs in the USA, cocaine is considered the largest producer of illicit income with the street sales of cocaine reaching an estimated $30 billion in the USA alone (*Time Magazine* ''High on cocaine, a drug with status and menace'', July 6, 1981). Cocaine is an alkaloid of the plant *Erythroxylon coca*, which grows extensively in the Andes Mountains. It is now estimated that 10 million Americans abuse cocaine with some regularity. Cocaine can also be lethal. For centuries marijuana has been widely used as a recreational drug. This drug is very popular with young adults. Marijuana is the principal active ingredient of *Cannabis sativa*, a hemp plant. Marijuana in general refers to tobacco-like preparations of the leaves and flowers. Hashish, the resin extract from the tops of the flowering plants may have a Δ^9-THC concentration over 10%. Opiate poisoning occurs most commonly following intravenous (IV) or oral methadone overdose by addicts or new abusers. Fortunately, opiate overdose can be treated effectively with naloxone, an opiate antagonist. Amphetamine, a popular drug of abuse, was first synthesized in 1887 and became available in 1937 in the USA in tablet form for the treatment of narcolepsy. During World War II this drug became very popular to overcome battle fatigue among soldiers. Following World War II, there was an epidemic of amphetamine abuse. The drug is still abused widely today. PCP, also known as ''angel dust'', is also a popular drug on the street. The chemical name of PCP is 1-(1-phencyclohexyl)piperidine. The drug was discovered by pharmacologists at Parke-Davis in 1956, marketed as an IV anesthetic called Sernyl, but was discontinued later owing to adverse reactions. In 1967, the drug was reintroduced in veterinary medicine as an anesthetic but the production of the drug was discontinued in 1979. In the 1970s the drug was introduced in the street and by the mid 1970s, there was an epidemic of abuse of PCP in the USA, after which its use declined. Eleven benzodiazepines are currently available in the USA and abuse is common. For many years diazepam was the most prescribed drug in the USA. Flunitrazepam (rohypnol) has recently gained publicity in date rape situations.

1.1 Alcohol Use and Abuse

Alcohol is found in 3–6% by volume in naturally fermented beers, 10–12% in wine and 20–60% in hard drinks which are often distilled drinks. Ethanol is also found in mouthwash and medicinal liquids as solvent. Blood ethanol levels varied between 180 mg dL^{-1} to 600 mg dL^{-1} in 94 acute fetal toxicity cases.[1] Ethanol can be formed or destroyed in vitro. Loss of ethanol from a specimen may occur by volatilization, or ethanol can be destroyed in vitro by microorganisms. Formation of alcohol in post-mortem blood by microorganisms can be inhibited by fluoride. Blood alcohol production

produced by post-mortem decomposition rarely exceeds 50 mg dL^{-1}.[2] Legal blood alcohol analysis should be carried out by GC. The flame ionization gas chromatographic technique usually involves direct injection of the sample after dilution with a solvent containing the internal standard.[3,4] Headspace analysis of the specimen is also useful for determination of serum alcohol. Penton described a headspace measurement of ethanol in blood by GC where *n*-propanol was used as an internal standard.[5] A blood specimen containing ethanol was placed in a 2-mL vial followed by addition of an internal standard solution saturated with sodium chloride. The sample was allowed to sit in the autosampler for 30 min prior to analysis. Pereira et al. described a GC/MS technique for the determination of ethanol concentration in blood and urine.[6] The authors used a deuterated analog of ethanol as the internal standard. The authors monitored ions with *m*/*z* 31 and 45 for ethanol and ions with *m*/*z* 33 and 49 for the internal standard. Recently Tangerman reported a highly sensitive gas chromatographic analysis of ethanol in whole blood, serum, urine and fecal supernatant. The author used direct injection of the biological specimen into the gas chromatograph without any pretreatment. Contamination of the gas chromatographic column with nonvolatile material was prevented by the use of a glass liner in the injector. The liner, which acted as a precolumn, was partly filled with glass beads. Injections were made between the glass beads.[7]

In many clinical laboratories, serum ethanol concentrations are determined using automated assays. In this assay the ability of alcohol dehydrogenase to convert alcohol to acetaldehyde is used and nicotinamide adenine dinucleotide (NAD) as a cofactor is reduced to reduced nicotinamide adenine dinucleotide (NADH). The EMIT serum alcohol assay (Behring, San Jose, CA) has been reported to produce false positive results in post-mortem samples because of increased concentrations of lactate and LDH.[8] This phenomenon was also observed in living subjects with high concentrations of lactate and LDH. A patient with lactic acidosis may have high serum LDH concentration caused by cellular breakdown. However, taking advantage of the high molecular weight of LDH and low molecular weight of ethanol, this interference can be eliminated by measuring alcohol concentrations in protein-free ultrafiltrate. EMIT measured alcohol concentrations in serum which were unaffected by the ultrafiltration technique because alcohol is not bound to serum protein.[9]

1.2 Major Illegal Drugs Abused in the USA

- Narcotics: opiates (morphine, codeine, heroin, meperidine, methadone)
- Stimulants: cocaine, amphetamines, methamphetamines, 3,4-methylenedioxymethamphetamine (MDMA)
- Cannabis: marijuana
- Sedative-hyponotic: barbiturates
- Benzodiazepines: diazepam, oxazepam, flurazepam
- 3,4-Methylenedioxyamphetamine (MDA) and MDMA cross-react with the amphetamine immunoassay producing a positive screen.
- 3-Methyl fentanyl and related compounds
- LSD
- PCP
- GABA, and Rohypnol® (Flunitrazepam). Both drugs have been reported in date rape situations.

2 DRUG TESTING

2.1 Specimen Types for Abused Drug Testing

In the USA, urine is the preferred sample for testing of abused drugs, as collecting a urine specimen is noninvasive. Moreover, metabolites of drugs remain in urine longer than in serum thus lengthening the window of detection. Immunoassays are available for rapid automated screening of the common drugs of abuse in urine. Blood and gastric contents are useful to detect recent overdose in hospital emergency units. Measurement of ethanol in vitreous humor is useful in forensic cases where the body has decomposed and blood is grossly hemolyzed. Other drugs are also known to diffuse into vitreous humor but further research will be required before a clear guideline can be established for forensic detection by this method.

2.2 Federal Guidelines for Drugs of Abuse Testing: Cutoff Limits and Other Issues

The drug testing program in the USA can be classified as mandatory or nonmandatory. In the first group (for example Department of Transportation) a regulated employer is required by federal regulation to test the employees. In the second category, employers choose to test for reasons other than the federal requirements. Private employers who are not mandated to test under federal authority have instituted employee drug testing in order to create a drug-free workplace. These programs also formalized the role of a specialist physician, designated as Medical Review Officer (MRO). An MRO is an integral part of a drug testing program, who can determine the cause of a positive result in drug testing (interference, other prescription drugs, etc.) and counsel the employee.

On September 15, 1986, President Reagan issued Executive Order No 12564 directing federal agencies

to achieve a drug-free work environment. Each agency was directed to develop criteria for the testing of drugs of abuse, and the use of illegal drugs for recreational purpose was banned. Then the DHHS, USA, (formerly NIDA) developed guidelines for drugs of abuse testing. The overall testing process under mandatory testing consists of proper collection of the specimen, initiation of a chain of custody and finally analysis of the specimen (screening and GC/MS confirmation if needed) by a laboratory certified by the Substance Abuse and Mental Health Services Administration (SAMHSA). The screening by immunoassay should be performed using a FDA approved method. The confirmation is by GC/MS. The cutoff values for drugs of abuse testing are given in Table 1 (section 2.6).

2.3 Forensic Drug Testing

Forensic drug testing can be more complex than workplace drug testing because, in order to investigate the cause of death, the specimen can be tested for the presence of a wide variety of drugs including commonly abused drugs. The toxicology section of the American Academy of Forensic Sciences has issued the following policy statement: ''Confirmation of results is essential in forensic toxicology. Positive results of toxicological screening tests, regardless of the method used, and positive toxicological analysis results obtained by immunoassay methods should be adequately confirmed before the results are used for forensic purposes, or clearly be designated as unconfirmed results. Analysis methods used for attempted confirmation of presumptive results must be appropriately sensitive and specific or unequivocally selective for the analyte in question, and must be based upon different chemical or physical principles than the initial method.'' GC/MS is widely used in a forensic laboratory for confirmation of variety of drugs. In addition, HPLC/MS is a very useful technique in a forensic toxicology laboratory. Both polar and nonpolar drugs can be analyzed using HPLC/MS. Spratt and Vallaro have published a general protocol to analyze over 100 drugs using HPLC/MS. The authors used simple liquid–liquid extraction (LLE) for the analysis of acidic, basic and neutral drugs by their protocol.[10] Several of these are polar and cannot be analyzed by conventional GC/MS.

2.4 Immunoassays for Screening of Drugs in Urine

EMIT kits are available for 10 commonly abused drugs. These drugs are amphetamines, opiates, benzodiazepines, cocaine, methadone, barbiturates, methaqualone, PCP, Δ^9-THC and propoxyphene. These kits are marketed by Behring Diagnostics (San Jose, CA, USA). Fluorescence polarization immunoassay kits for the screening of drugs of abuse are available from Abbott Laboratories (Abbott Park, IL). Particle immunoaggregation assay is also available (Roche Abuscreen and Biosite Diagnostics Triage assay) commercially. Although immunoassays are popular in the clinical laboratory for screening of abused drugs because of the speed of analysis and total automation, there are some serious problems with immunoassays regarding specificity. Adulterants can cause false negative results in the immunoassay for drug testing.

2.5 Common Adulterants Used to Mask Drug Testing

Immunoassay results are very important in testing drugs of abuse. In most laboratories, if the immunoassay result is negative, no further testing is performed. Because GC/MS confirmation and quantification of drugs of abuse is an intensive process, this protocol is applied only to confirm the presence of an abused drug if the initial immunoassay result is positive. GC/MS confirmation of

Table 1 Window of detection and detection limit (DL) of abused drugs

Drug	Window	Screening cutoff ($ng\,mL^{-1}$)	GC/MS confirmation cutoff ($ng\,mL^{-1}$)
Amphetamine	2–3 days	1000	500
Methamphetamine	2–3 days	1000	500
Cocaine as BE	2–3 days	300, 100 for Military	150
Marijuana metabolites	2 days–3 weeks	20, 50–100	15
Opiate metabolites	2–3 days	2000[a]	2000[a]
6-monoacetylmorphine			10
PCP	8 days–3 weeks	25	25
Benzodiazepines	3 or more days	300	300
Methadone	3 days	300	300
Methaqualone	2 weeks	75	75

[a] DHHS has increased the cutoff for both screening and confirmation of opiates to 2000 $ng\,mL^{-1}$ from 300 $ng\,mL^{-1}$ in order to avoid false positives caused by the ingestion of food containing poppy seeds.
BE, benzoylecgonine.

drugs is required in legal drug testing. People try to beat drug testing by adding adulterants to urine specimens. Several adulterants can cause false negative results in drug testing by immunoassays. Although these commonly used adulterants have no effect on the GC/MS analysis, their ability to mask an immunoassay screening step has a profound effect on testing drugs of abuse. Common adulterants used to mask drug testing are:[11]

- table salt
- household vinegar
- liquid laundry bleach
- concentrated lemon juice
- golden seal tea (produces dark urine)
- diluted urine (creatinine below 15 mg dL^{-1})
- Visine eye drop (masks drug tests by EMIT).

Household vinegar and concentrated lemon juice make urine acidic and can easily be detected by checking the pH of urine. Table salt increases the specific gravity of urine and the diluted urine specimen should have a low specific gravity which can easily be identified by checking the pH of the specimen. The presence of Visine eye drops in urine cannot be detected by specimen integrity test. Unfortunately, Visine eye drops can mask drug testing by EMIT, by oxidizing NADH back to NAD in the final step of the reaction so that the signal is lost. Both the collection site and the laboratory have a number of mechanisms available to detect potentially invalid specimens. The temperature, for instance, should be within 90.5–98.9 °F. The specific gravity should be 1.005–1.030 and the pH should be 4–10. The creatinine concentration should be 20–400 mg dL^{-1}, although some drug testing laboratories consider a creatinine concentration of 15 mg dL^{-1} as the lower end of the cutoff concentration.

2.6 Window of Detection and Detection Limits (Cutoffs) for Abused Drugs

A negative immunoassay result does not rule out the presence of any abused drug in the urine. If the drug concentrations were to be below the DL of the assay, the immunoassay screening result would be negative. Moreover, depending on the metabolism and elimination kinetics of the drug, different drugs have different windows of detection. Cocaine can be detected as BE, the major metabolite, up to three days after last use. The DLs and approximate windows of detection of major abused drugs are given in Table 1.

2.7 True Positives in Drugs of Abuse Testing

There are some situations where consuming a particular food may cause a positive immunoassay result because a drug is present. Poppy seeds contain opiates. Consuming poppy seed cakes or muffins can produce a positive opiate screen and eventually confirmation of the presence of morphine and possibly codeine in urine by GC/MS. The concentration of morphine and codeine were 2797 ng mL^{-1} and 214 ng mL^{-1} respectively in one healthy volunteer who ingested three poppy seed bagels (5 mg opiate). Opiate was present in the urine up to 25 h post ingestion.[12] No opiate was detected 45 h after ingestion of poppy seed bagels. To circumvent this problem, DHHS increased the cutoff of opiate immunoassay from 300 ng mL^{-1} to 2000 ng mL^{-1}.

Drinking health inca tea may cause a positive test for cocaine. Although USA custom regulations require that no cocaine should be present in any herbal tea, literature references indicate that some health inca tea sold in the USA contains cocaine. Jackson et al. reported urinary concentration of BE, the major metabolite of cocaine after ingestion of one cup of health inca tea by volunteers. BE was present even 26 h after ingestion. Maximum urinary BE concentration of 14–2.8 mg L^{-1} was observed 4–11 h after ingestion of health inca tea.[13]

A positive cannabinoids workplace drug testing following ingestion of commercially available hemp seed oil preparation has been reported.[14] The first specimen testing negative was 53 h after ingestion.

Positive inhalation of marijuana is a popular defense strategy. However, literature reference indicates that passive inhalation of marijuana does not produce enough urinary concentration of metabolite to test positive in an immunoassay.[15] Recently, cocaine has been isolated from USA currency.[15,16] Oyler et al. examined 10 single dollar bills from several big cities in the USA for the presence of cocaine.[16] They extracted dollar bills using methanol, and subsequently cocaine was purified using solid-phase extraction (SPE). The presence of cocaine was confirmed by GC/MS. The authors found cocaine in 74% of the bills in an amount above 0.1 μg and in 54% of the currency with concentrations above 1.0 μg. The highest amount observed was 1327 μg in one dollar bill. Negrusz et al. analyzed ten $20 bills collected from Rockford, IL and four $1 bills collected from Chicago.[17] The concentrations of cocaine varied from 10.02 μg to 0.14 μg in $20 bills, and from 2.99 μg to none detected in $1 bills. Overall 92.8% of all bills analyzed were contaminated with cocaine.

ElSohly investigated whether individuals who handled cocaine-contaminated money would test positive by urinalysis. Two dollar bills were immersed in dry, powdered cocaine, and then shaken free of loose cocaine. One individual then handled the money several times during the course of the day. Analysis of urine samples collected over a period of approximately 24 h after handling the contaminated money revealed that the

individual excreted a maximum of 72 ng mL^{-1} of BE 12.5 h after handling the money. It was concluded that casual handling of articles contaminated with cocaine would not result in a positive test for cocaine at the recommended cutoff of 100 ng mL^{-1}.[18]

3 METHODS OF TESTING

3.1 Thin-layer Chromatography in Testing for Abused Drugs

Thin-layer chromatography (TLC) is a useful technique in screening for abused drugs. Currently, automated immunoassays are available for the ten most commonly abused drugs. However, more drugs can be detected by TLC. Another advantage of the TLC technique is that drug metabolites can also be detected. The TOXI-LAB AB drug detection system is a TLC system modified for quick identification of a broad spectrum of drugs. TOXI-LAB A is used for the detection of basic and neutral drugs including propoxyphene, methadone, meperidine, codeine, morphine, diazepam, cocaine, acetaminophen, caffeine, nicotine, amphetamine, methamphetamine, pseudoephedrine, PCP, trimeprazine, trifluoperazine, quinine, methaqualone, meprobamate, amitriptyline, doxepin, nortriptyline and strychnine. TOXI-LAB B is used for the detection of acidic and neutral drugs including secobarbital, phenytoin, phenobarbital, gluthimide, pentobarbital, ethinamate, amobarbital, butabarbital and barbital.

3.2 Gas Chromatography/Mass Spectrometry Confirmation and Quantification of Abused Drugs

GC/MS confirmation of abused drugs is considered to be the gold standard for drugs of abuse analysis. This analysis is essential in order to defend a legal challenge of a positive drug testing result. Mass spectrometry produces a fingerprint of the drug molecule and positive identification can be made by comparing the fragmentation pattern of the unknown peak with the known fragmentation pattern of the drug standard. A high matching quality is required to identify a drug positively. Sometimes, visual comparison of the fragmentation patterns by an experienced toxicologist is needed for positive identification.

GC is not capable of analyzing polar molecules. Sometimes derivatization of polar drug metabolites is needed before GC/MS analysis. Some polar drug metabolites, for example morphine, are conjugated with glucuronic acid or sulfate. In this case, acid or enzyme hydrolysis of very polar drug metabolites is needed to break the conjugate prior to extraction of the drug metabolite from urine using an appropriate organic solvent. Therefore, GC/MS analysis of abused drugs involves multiple steps:

1. acid or enzyme hydrolysis of drug conjugate if necessary;
2. making urine acidic or alkaline depending on the chemical structure of the drug to make it more soluble in an organic solvent;
3. extraction of the drug in an organic phase;
4. concentration of the organic phase and derivatization if necessary.

3.3 Protocols for Gas Chromatography/Mass Spectrometry Analysis of Amphetamines

3.3.1 Derivatization Reactions for Amphetamines

Fluoro derivatives of amphetamines are commonly used. These derivatives can easily be prepared by using the commercially available reagents, trifluoroacetic anhydride, pentafluoropropionyl anhydride and heptafluorobutyl anhydride.[19] The trichloroacetyl derivative of amphetamine and methamphetamine have also been reported.[20] Amphetamine should be extracted from urine prior to derivatization. Derivatization reactions require heating of specimens in a heating bath for different lengths of time depending on the particular type of derivative. Since 1989, 4-carbethoxyhexafluorobutyryl and perfluorooctanoyl derivatives of amphetamines have been described in the literature.[21,22] 4-carbethoxyhexafluorobutyryl chloride ($Cl-CO-CF_2CF_2COOCH_2CH_3$) can be used for the preparation of the 4-carbethoxy derivative while perfluorooctanoyl chloride ($Cl-COCF_2CF_2CF_2CF_2CF_2CF_2CF_3$) can be used for the preparation of the perfluorooctanoyl derivative. Both are available from PCR Incorporated (P.O. Box 1466, Gainesville, FL 32602, USA).

The conventional trifluoroacetyl derivative of amphetamine is very volatile and has a low molecular weight. The pentafluoropropionyl and heptafluorobutyryl derivatives are less volatile but the molecular weights are not high. Usually if the molecular weight of the derivative is high, then more mass fragments are observed in the higher mass range thus aiding in unambiguous identification of the molecule. Moreover, a derivative with a higher molecular weight is usually less volatile than the derivative of the same compound with a lower molecular weight. Therefore, one advantage of using a derivative with a high molecular weight is that the assay is free from interference from volatile components of the urine or serum matrix. The addition of a 4-carbethoxyhexafluorobutyryl group to amphetamine increases the molecular weight by 251 amu while the introduction of a perfluorooctanoyl group increases the molecular weight by 397 amu.

In contrast, following trifluoroacetyl derivatization of amphetamine, the molecular weight is increased only by 97 amu.

Derivatization reactions can be carried out rapidly using microwave technology. Different extraction protocols for amphetamines have been described in various publications. However, the basic principle is to make urine alkaline in order to make amphetamine more soluble in the organic phase, followed by extraction using a relatively nonpolar solvent such as methylene chloride or 1-chlorobutane. Extraction of amphetamines using a SPE column has also been reported by Wu et al.[23] The authors extracted amphetamines from 3 mL urine using Bond-Elute Extraction column. The extraction column was washed with 1 M acetic acid and methanol. Amphetamines were extracted using ethyl acetate (2% ammonium hydroxide). Derivatization was carried out using heptafluorobutyric anhydride. Quantitation was done using *m*/*z* 261 for the d8 methamphetamine (internal standard), *m*/*z* 240 for derivatized amphetamine and *m*/*z* 254 for the derivatized methamphetamine.

3.3.2 *Use of Microwave Irradiation for Rapid Preparation of Derivatives for Amphetamines*

Domestic microwave ovens can be used for rapid preparation of trifluoroacetyl, pentafluoropropyl, heptafluorobutyl, perfluorooctanoyl and 4-carbethoxyhexafluorobutyryl derivatives of amphetamine, methamphetamine, and MDMA.[24] The concentrations of amphetamine and methamphetamine obtained by the microwave derivatization technique were similar to the concentrations obtained by the conventional pentafluoropropionic anhydride derivatization method in both urine controls and EMIT positive urine specimens. Conventional techniques require heating of the reaction mixture at 40 °C for 15 min for the preparation of the trifluoroacetyl derivative, 75 °C for 15 min for the preparation of the pentafluoropropionyl derivative and 60 °C for 40 min for the preparation of the heptafluorobutyl derivative. Using microwave irradiation, the trifluoroacetyl derivative can be formed in 45 s, the pentafluoropropyl derivative in 1 min and the heptafluorobutyl derivative in 6 min.

The reaction conditions for derivatization of amphetamine using microwave irradiation are given in Table 2.

Table 2 Reaction conditions for microwave-induced derivatization of amphetamine

Derivative	Reaction time	Microwave setting	Microwave power (W)
Trifluoroacetyl	45 s	3	240
Pentafluoropropyl	1 min	5	400
Heptafluorobutyl	6 min	4	320
Perfluorooctanoyl	1 min	8	640
4-Carbethoxy	6 min	3	240

Domestic Microwave oven used: 800 W (Model MW 5510-T; Samsung, Seoul, South Korea).

3.3.3 *Carbamate Derivatives of Amphetamines*

Recently, carbamate derivatives of amphetamines have been introduced. Amphetamine and methamphetamine can be converted to their corresponding carbamate derivatives using any of the following chloroformate reagents available from Aldrich Chemical company (Milwaukee, WI):

Methyl chloroformate: $Cl-COO-CH_3$

Propyl chloroformate: $Cl-COO-CH_2CH_2CH_3$

2,2,2-trichloroethyl chloroformate: $Cl-COO-CH_2CCl_3$

3.3.4 *Derivatization of Amphetamines using Methyl Chloroformate, Propyl Chloroformate and 2,2,2-Trichloroethyl Chloroformate*

Although carbamate derivatives of amphetamine and methamphetamine have relatively low molecular weight, one advantage of this derivatization process is that no heating is necessary. Derivatives can be formed at room temperature even in an aqueous medium.[25] The authors extracted amphetamines from only 0.2 mL of urine specimen after adjusting the pH to 9 by adding 1 M sodium carbonate buffer. Internal standard (1000 ng mL^{-1} of ethylamphetamine) and 0.5 mL of isooctane were added. Then 50 μL of methyl chloroformate was added and the mixture was shaken at room temperature for 10 min. Then the organic phase was separated and transferred to a new tube. 0.2 mL of methanol saturated with potassium hydroxide was added to destroy any excess derivatizing agent. Then 0.5 mL of 2 M potassium hydroxide was added and the isooctane layer was transferred to a micro vial and 1 μL was injected into GC/MS. Because the derivative had a relatively low molecular weight, the initial oven temperature was 85 °C. The oven temperature was held for 1 min and then increased at a rate of 28 °C min^{-1} to 180 °C. Then the oven temperature was again increased at 35 °C min^{-1} to 280 °C. The injector port temperature was 210 °C and the mass spectrometer interface was 300 °C. The column was a 30 mHP-5 MS. Splitless injection was used. The derivatization reaction was complete in 10 min for all analytes except for phentermine. An explanation might be that the second methyl group in the alpha position causes steric hindrance.

Like methyl chloroformate, propyl chloroformate derivatives also can be formed at room temperature in an aqueous medium.[26,27] However, the authors used

a higher volume of urine specimen for extraction of amphetamines. Again, the derivatives can be formed at room temperature in the aqueous phase. Another alternative is to extract amphetamine in the organic phase using 1-chlorobutane or methylene chloride and then add propyl chloroformate in the organic phase.[27] Because amphetamine propyl carbamate has a higher molecular weight than the corresponding methyl carbamate derivative, a higher initial oven temperature of 130 °C was used by the authors. Another derivatization of amphetamines using 2,2,2-trichloroethyl chloroformate has recently been described.[28] This derivative can also be formed at room temperature in 10 min. 2,2,2-trichloroethyl derivatives are less volatile and elute at a relatively higher temperature. A chromatogram showing separation between derivatized amphetamine and methamphetamine in a clinical urine sample is given in Figure 1. Because of the isotopic effect of chlorine (major isotopes, 35 and 37), a cluster of peaks instead of a single peak was observed, which aids unambiguous confirmation. Moreover, a mass unit of 175 was added to amphetamine instead of 87 for propyl chloroformate (Figures 2, 3). Because the 2,2,2-trichloroethyl carbamate derivative of amphetamine is even less volatile than the corresponding propyl carbamate derivative, a relatively high initial oven temperature of 175 °C can be used.

3.3.5 Problem of Misidentification of Ephedrine/Pseudoephedrine as Methamphetamine in the Gas Chromatography/Mass Spectrometry Confirmation

Misidentification of ephedrine/pseudoephedrine as methamphetamine has been reported.[29] The license of two NIDA certified laboratories was suspended owing to misidentification of ephedrine as methamphetamine in survey samples. The causes of misidentification are:

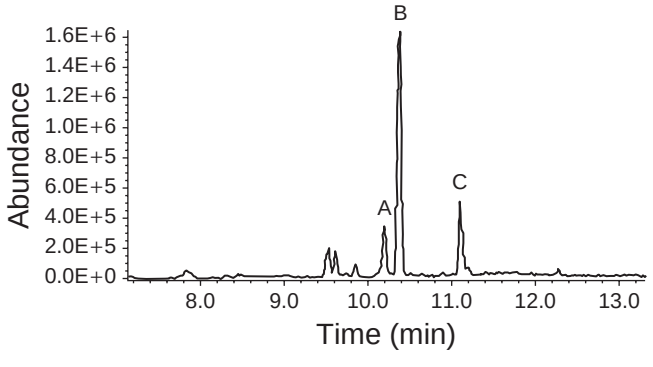

Figure 1 Total ion chromatogram showing separation between amphetamine, methamphetamine and *N*-propyl amphetamine (internal standard) after derivatization with 2,2,2-trichloroethyl chloroformate. Concentration of amphetamine (peak A) was 709 ng mL^{-1}, while the concentration of methamphetamine (peak B) was 3303 ng mL^{-1} in the urine specimen. The concentration of internal standard (peak C) was 1000 ng mL^{-1}.

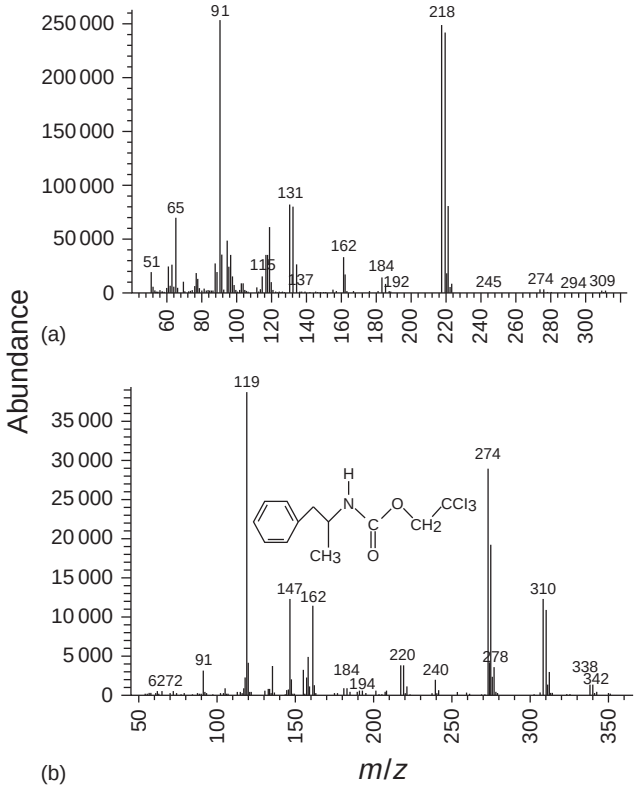

Figure 2 2,2,2-Trichloroethyl amphetamine carbamate. (a) Electron ionization. (b) Chemical ionization.

1. Ephedrine and pseudoephedrine can be thermally dehydrated to methamphetamine at an injector port temperature over 185 °C.[29] The authors reported a GC/MS artifact peak as methamphetamine after derivatization of ephedrine/pseudoephedrine with heptafluoro butyric anhydride, 4-carbethoxyhexafluorobutyryl chloride and *N*-trifluoroacetyl-1-propyl chloride. The artifact peaks were produced when the injector port temperature of GC was 300 °C.
2. Electron impact mass spectra of trifluoroacetyl, pentafluoropropyl, heptafluorobutyl, 4-carbethoxyhexafluorobutyryl and various carbamate derivatives of methamphetamine are similar to the corresponding derivatized ephedrine/pseudoephedrine. Because ephedrine and pseudoephedrine elute just after the methamphetamine peak, if the mass spectral analysis is not done carefully, ephedrine/pseudoephedrine can be misidentified as methamphetamine. Because both ephedrine and pseudoephedrine are legal drugs and are found in over-the-counter medications, misidentification of ephedrine/pseudoephedrine as methamphetamine has serious medical and legal implications.

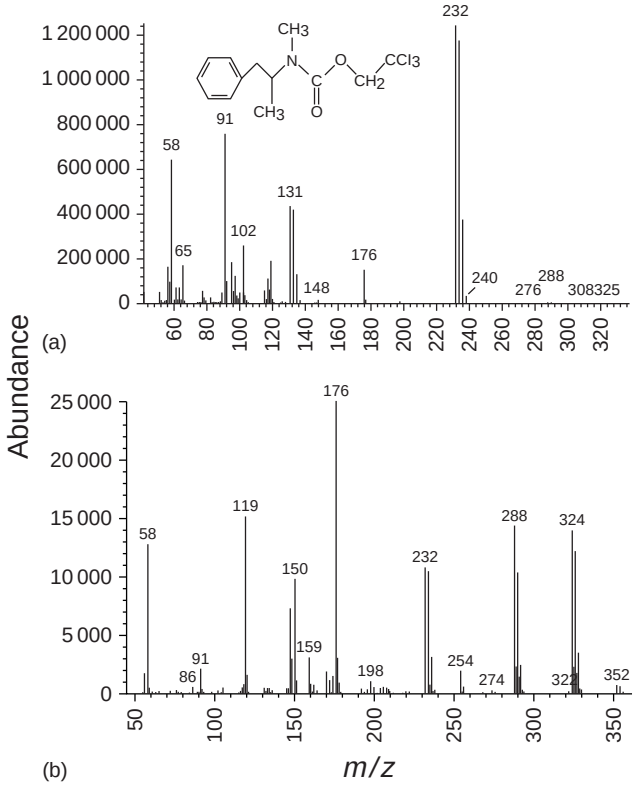

Figure 3 2,2,2-Trichloroethyl methamphetamine carbamate. (a) Electron ionization. (b) Chemical ionization.

3.3.6 *Elimination of Misidentification by Using Chemical Ionization Mass Spectrometry*

The chemical ionization mass spectrum of derivatized methamphetamine is very different from that of derivatized ephedrine/pseudoephedrine. A misidentification problem can be completely eliminated.[30] In the electron ionization mode, the molecular ion of derivatized amphetamine is usually present as a very weak peak. In contrast, the protonated molecular ion is the base peak (100% abundance) in the chemical ionization mass spectra of derivatized amphetamine and methamphetamine. The major mass spectral fragmentation pattern differences are given in Table 3. This misidentification problem also exists with recently described propyl carbamate derivatives of methamphetamine. In the electron ionization mode, both methamphetamine propyl carbamate and ephedrine propyl carbamate showed almost identical mass spectral fragmentation patterns (Figure 4). Again this problem can be circumvented by using chemical ionization where methamphetamine propyl carbamate showed a base peak at m/z 236 and ephedrine propyl carbamate showed a base peak at m/z 192 (Figure 5).

3.3.7 *Analysis of Designer Drugs*

MDA (Ecstasy) and MDMA (Love) are also widely abused. These drugs cross-react with most amphetamine screening assays. These designer drugs can be analyzed using any extraction and derivatization protocol for GC/MS confirmation of amphetamines. Usually both drugs show much longer retention times than amphetamine and methamphetamine.

3.3.8 *Analysis of Phenmetrazine*

Phenmetrazine (Preludin) is the *N*-desmethyl analog of phendimetrazine and also is a primary metabolite of phendimetrazine. The drug is used clinically as an anorectic agent and is available for oral use as a hydrochloride salt in a single dose of 25 mg. The drug has a high abuse potential and several deaths have been reported from phenmetrazine overdoses. This drug cross-reacts with the EMIT screen for amphetamine at much higher concentrations (6600 ng mL^{-1} to record a positive). Phenmetrazine can be confirmed by GC/MS using *N*-ethylamphetamine as an internal standard.[31]

Table 3 Electron ionization and chemical ionization mass spectral features of derivatized amphetamine and methamphetamine

Compound	Electron Impact (m/z)	Chemical Ionization (m/z)
Methamphetamine propyl carbamate	Base: 144 Other peaks: 102, 91, 58	Base: 236 $(M+1)^+$ Other peaks: 176, 144, 119, 58
Ephedrine propyl carbamate	Base: 144 Other peaks: 102, 77, 58	Base: 192 Other peaks: 220, 148, 58
Methamphetamine trifluoroacetyl	Base: 154 Other peaks: 118, 58	Base: 246 $(M+1)^+$ Other peaks: 154, 119
Ephedrine trifluoroacetyl	Base: 154 Other peaks: 118, 58	Base: 244 Other peaks: 276, 158
Methamphetamine 4-carbethoxy	Base: 315 Other peaks: 118, 91	Base: 400 $(M+1)^+$ Other peaks: 308, 119
Ephedrine 4-carbethoxy	Base: 308 Other peaks: 118, 91	Base: 398 Other peaks: 121

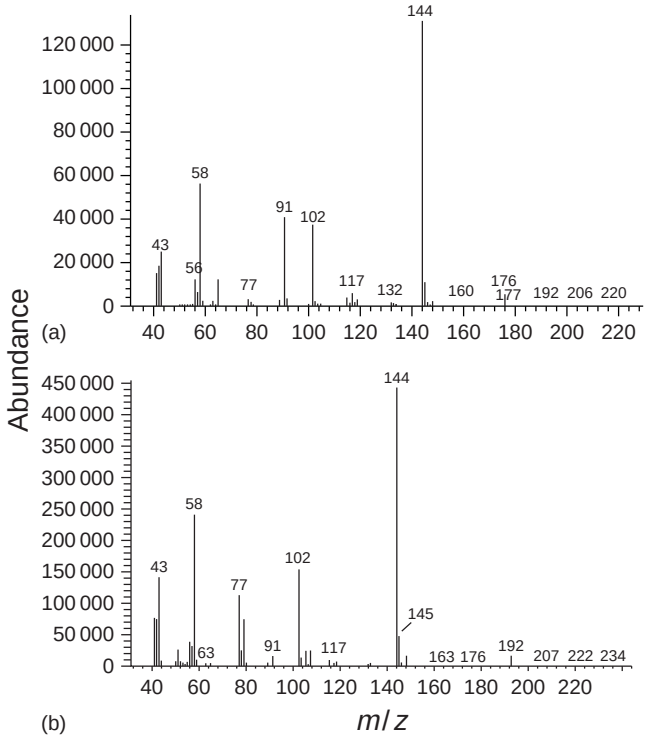

Figure 4 Electron ionization mass spectrum of (a) methamphetamine propyl carbamate, (b) ephedrine propyl carbamate.

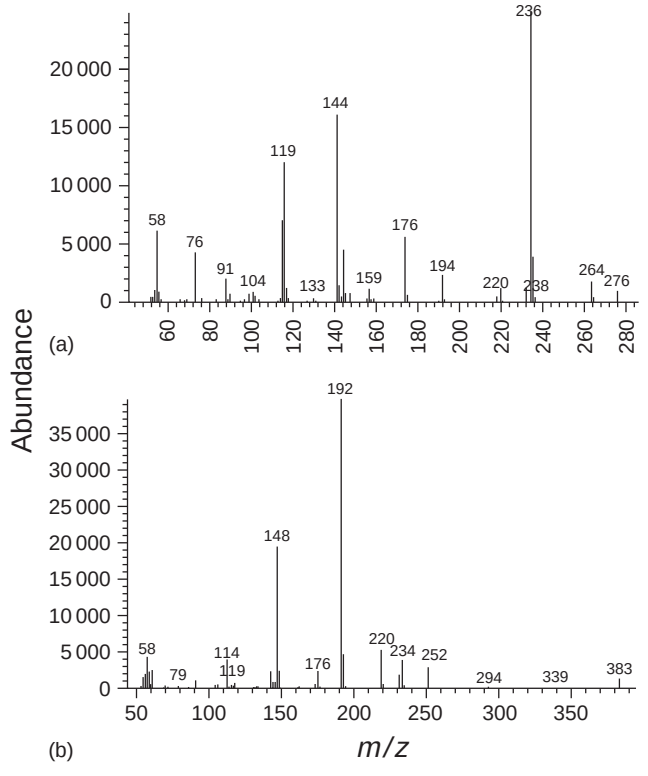

Figure 5 Chemical ionization mass spectrum of (a) methamphetamine propyl carbamate, (b) ephedrine propyl carbamate.

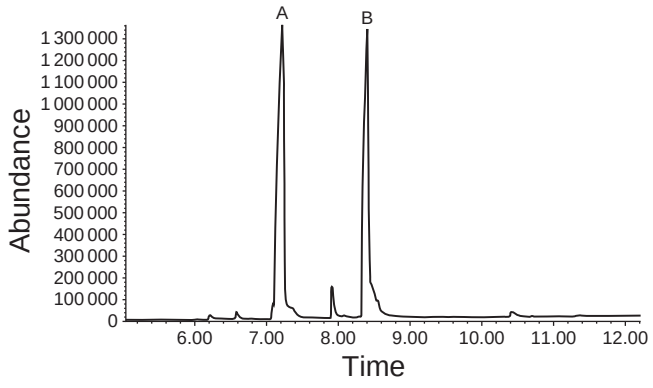

Figure 6 Total ion chromatogram showing separation between phenmetrazine and *N*-propyl amphetamine, the internal standard after derivatization with 4-carbethoxyhexafluoro butyryl chloride. Peak A is the derivatized internal standard (*N*-ethylamphetamine) and peak B is the derivatized phenmetrazine.

Excellent separation between derivatized phenmetrazine and the derivatized internal standard was observed (Figure 6).

3.4 Gas Chromatography/Mass Spectrometry Analysis of Cocaine Metabolites

The major metabolite of cocaine is BE and most GC/MS methods are directed towards confirmation and quantification of BE. Another important metabolite of cocaine is ecgonine methyl ester (EME). Other metabolites of cocaine, such as norbenzoylecgonine (NBE), norcocaine, and ecgonine have also been described.[32,33] Jatlow et al. reported the presence of cocaethylene (CE) in seven post-mortem blood samples from individuals who had ingested both cocaine and ethanol.[34] In contrast to BE, which is an inactive metabolite, CE is an active metabolite and contributes significantly to the toxicity of cocaine.[34] Significant amounts of CE have been reported in the urine of patients hospitalized after cocaine overdose.[35]

Although cocaine and CE can be analyzed by GC/MS, without derivatization, BE and EME require derivatization prior to GC/MS analysis. Several derivatives of BE have been reported in the literature. Methylation of BE to convert it back to cocaine can be used for GC/MS confirmation. BE can be extracted from urine using chloroform–ethanol (80:20 by vol) followed by derivatization using sulfuric acid in methanol.[36] Joren described a protocol by which BE can be simultaneously extracted and derivatized using an extractive alkylation solvent solution of methylene chloride–1-iodopropane (99:1 by vol). The author used butylbenzoylecgonine as an internal standard. This method is very sensitive with a DL of 35 ng mL^{-1} of BE.[37]

SPE of cocaine metabolites from urine is common. Deuterated metabolites were also widely used as internal standards. Jenkins et al. recently described a SPE of cocaine metabolites using Clean Screen SPE column (United Chemical Technologies, Horsham, PA). Deuterated cocaine, d_3-BE, and d_3-EME were used as internal standards. Cocaine metabolites were eluted from the solid-phase column with an elution solvent (methylene chloride–isopropyl alcohol–ammonium hydroxide: 80 : 20 : 2). Trimethylsilyl derivatives were prepared using BSTFA by heating the organic extract at 60 °C.[38] Taylor et al. also described a SPE protocol for BE using an Amberlite XAD-2 extraction column. BE was eluted with *n*-butylchloride–acetonitrile–methanol (40 : 50 : 10 by vol) and analyzed as the trifluorosilyl derivative.[39]

The *t*-butyldimethylsilyl derivative of BE has also been described in the literature.[40] The trimethylsilyl derivative of BE is sensitive to moisture and can slowly be hydrolyzed back to BE. On the other hand the *t*-butyldimethylsilyl derivative is stable in the presence of moisture. These derivatives can be prepared by heating BE with the appropriate silylating reagent. Another alternative is to use microwave irradiation for the rapid preparation of these derivatives.[41] The mass spectra of derivatized compounds obtained by microwave irradiation were identical to those obtained by conventional heating. Therefore the derivatized molecules produced by microwave irradiation have the same chemical identity as the derivatized molecules produced by conventional heating. Mass spectral fragmentation patterns of the *t*-butyldimethylsilyl derivative of BE are given in Figure 7. The authors obtained similar concentrations of cocaine metabolites in EMIT positive urine specimens by both microwave irradiation and conventional heating. Therefore, yields of derivatives by the microwave technique were similar to those obtained by conventional heating. Moreover, concentrations of cocaine and CE obtained by the microwave technique were similar to those obtained by a conventional heating method.

3.4.1 Fluconazole Interference in Gas Chromatography/ Mass Spectrometry Confirmation of Benzoylecgonine and Elimination of this Interference

The trimethylsilyl derivatives of fluconazole and BE elute at the same time and produces false negative results.[42] Fluconazole does not interfere in the EMIT screening assay for cocaine metabolite. Since a specimen positive by the screening assay and negative by the GC/MS confirmatory assay is considered negative for the presence of cocaine metabolite, negative interference of fluconazole in the GC/MS assay is a serious problem.

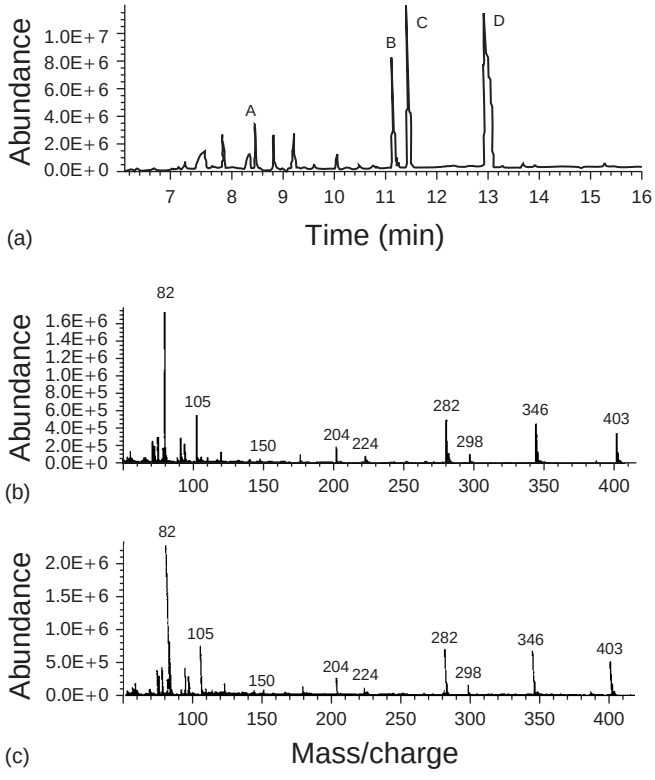

Figure 7 (a) Total ion chromatogram showing separation between cocaine, CE and the metabolites. Peak A: *t*-butyldimethylsilyl derivative of EME; B: cocaine; C: CE; D: *t*-butylmethylsilyl derivative of BE. Mass spectral fragmentation pattern of *t*-butyldimethylsilyl derivative of BE prepared by (b) microwave irradiation, (c) conventional heating.

However, this problem can be completely eliminated by preparing a pentafluoropropionyl derivative of BE. Derivatized BE eluted earlier (retention time: 14.7 min) than derivatized fluconazole (retention time: 15.6 min).[43] The chemical structures of the conventionally used trimethylsilyl derivative of BE and the other derivative are given in Figure 8.

In order to prepare the pentafluoropropionic derivative of BE, pentafluoropropionic anhydride and 25 µL of 2,2,3,3,3-pentafluoropropanol (both available from Aldrich Chemical Company, Milwaukee, WI) were added to the dried organic extract of BE. The reaction mixture was incubated at 85 °C for 20 min.[44] After cooling, the excess derivatizing agents were evaporated and the dry

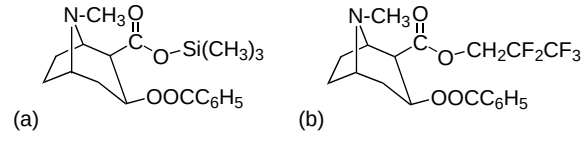

Figure 8 Chemical structures of (a) BE-trimethylsilyl ester and (b) BE-pentafluoropropionic ester.

extract was reconstituted with 50 µL of ethyl acetate. 2 µL was injected into the GC/MS system.

The free carboxylic group of BE was esterified by 2,2,3,3,3-pentafluoropropanol in the presence of pentafluoropropionic anhydride as the catalyst. The hydroxyl group of fluconazole was converted to the corresponding pentafluoropropionyl derivative. Derivatized BE was well separated from the derivatized fluconazole. Moreover, the mass spectral fragmentation patterns were distinctively different. Derivatized BE showed a base peak at *m*/*z* 300 and a strong molecular ion at *m*/*z* 421 in the positive electron ionization mass spectrum. In contrast, derivatized fluconazole showed a base peak at *m*/*z* 224 and a strong peak at *m*/*z* 127 (Figures 9, 10).

3.5 Gas Chromatography/Mass Spectrometry Confirmation of Opiates

Heroin, morphine and codeine are widely abused drugs. Morphine is used medically for pain control and codeine is also found in pain and cold medications. Hydromorphone, hydrocodone and oxycodone also have medical use. Most of these drugs achieve low plasma concentrations but tend to concentrate in urine most often as glucuronide conjugate. Acid hydrolysis or enzyme hydrolysis of this conjugate is essential to generate free drug. Acid hydrolysis for 1 h in a boiling water bath using concentrated hydrochloric acid is essential to generate

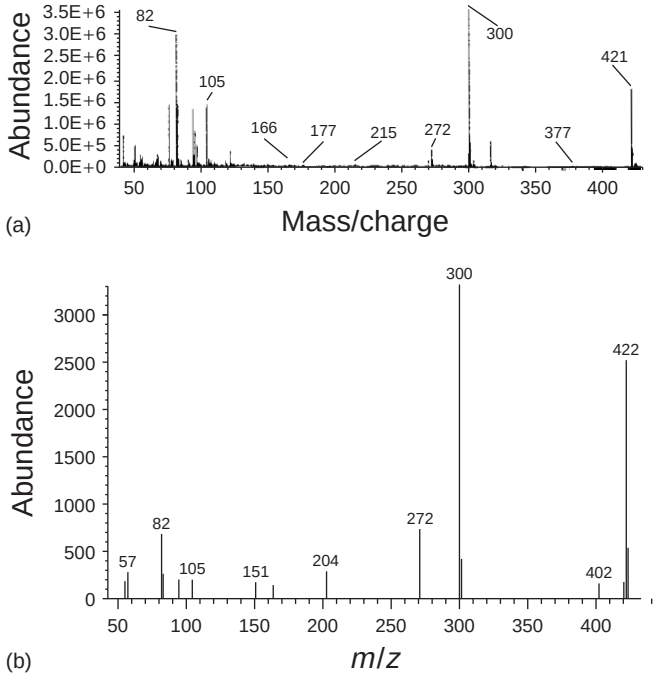

Figure 9 Mass spectral fragmentation patterns of BE: (a) electron ionization, (b) chemical ionization after derivatization with pentafluoropropionic anhydride/pentafluoropropanol.

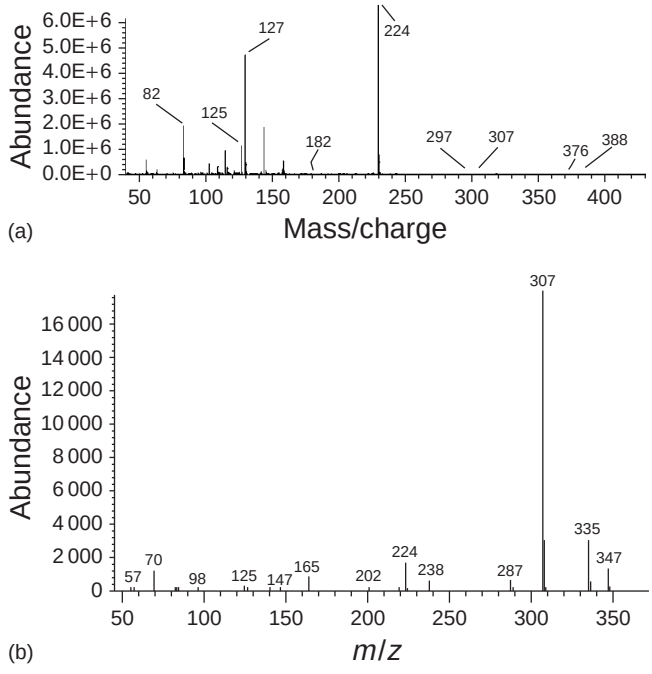

Figure 10 Mass spectral fragmentation patterns of fluconazole: (a) electron ionization, (b) chemical ionization after derivatization with pentafluoropropionic anhydride/pentafluoropropanol.

free morphine. Enzyme hydrolysis using glucuronidase requires a longer incubation of up to 24 h depending on preparation. Following acid or enzyme hydrolysis, the urine pH should be adjusted with a base in order to make morphine more soluble in the organic phase. Morphine can be extracted using ethyl acetate–isopropanol (9:1 by vol). Because morphine has two free hydroxyl groups, derivatization is essential prior to GC/MS confirmation. The acetyl derivative of morphine can easily be prepared by incubating dried organic extract with acetic anhydride–pyridine (1:1 by vol) at 80 °C for 30 min.[45] The trimethylsilyl derivatives of morphine and codeine can also be used for GC/MS confirmation of opiates.[46]

Acetic anhydride derivatization has one major limitation. In this technique, both morphine and 6-monoacetyl morphine are converted back to heroin. Pentafluoropropionyl derivatives are useful for codeine and morphine, but one author suggests that hydromorphone interferes with morphine quantitation.[47] Another author claims that oxycodone co-elutes with codeine.[48] Trimethylsilyl derivatives often form mono-, di- and triderivatives for the synthetic opiates, making quantitation a major problem.[49]

One approach to circumvent this problem is to use hydroxylamine to convert keto opiates to an oxime derivative and then further derivatizing the oxime product using BSTFA to form trimethylsilyl derivatives.[50]

Recently Cremese et al. described this approach in detail for GC/MS analysis of morphine, hydrocodone, hydromorphone, oxycodone and oxymorphone.[51] The authors used deuterated codeine as an internal standard. To a urine sample, internal standard solution was added and then samples were simultaneously reacted with 200 μL of 10 g dL^{-1} hydroxylamine and hydrolyzed with 5000 U mL^{-1} of β-glucuronidase enzyme (Sigma Chemical Company, St. Louis, MO) by incubation at 65 °C for 2 h. After cooling, the pH of the specimen was adjusted to 6 with 0.1 mol L^{-1} sodium hydroxide followed by SPE of opiates using Clean Screen DAU (C8 200 mg. Worldwide Monitoring). Opiates were eluted with methylene chloride–isopropanol–ammonium hydroxide (78 : 20 : 2 by vol) in centrifuge tubes. The elutes were evaporated to dryness and derivatized with 75 μL BSTFA containing 1% trimethylsilylchlorosilane (TMCS) and 75 μL ethyl acetate. Samples were reacted for 20 min at 65 °C. Selected ions monitored were those at *m*/*z* 371 (quantitation ion); 334 and 235 for codeine; 374 and 345 for deuterated codeine; 429 (quantitation ion); 414, 432 and 404 for morphine.[51] Wu et al. also described a SPE protocol for opiates using Bond Elute columns. Codeine and morphine conjugates were hydrolyzed and extracted from 2 mL urine. Deuterated dihydromorphine and dihydrocodeine were used as internal standards. Urine pH was adjusted to 3.8 using acetate buffer and hydrolysis of conjugates were carried out with β-glucuronidase at 60 °C for 3 h. Then pH of the urine was adjusted to 8.9 using tris buffer (tris(hydroxymethyl)aminomethane). Opiates were finally eluted from the column using 2 mL methylene chloride–isopropanol (80 : 20 by vol) and derivatized with pentafluoropropionic anhydride. Quantitation was done using *m*/*z* 414 and 445 for derivatized morphine and codeine and *m*/*z* 418 and 449 for the corresponding internal standards.[23]

3.5.1 Gas Chromatography/Mass Spectrometry Confirmation of Heroin Abuse

Heroin (diacetylmorphine) was synthesized in 1874. This drug currently has no medical use. The average street heroin is 2–6% pure and each dosage packet may contain 3–16 mg of heroin. Heroin is rapidly deacetylated to 6-monoacetylmorphine (half life: 9 min) and then 6-monoacetylmorphine is metabolized at a slower rate to morphine (half life: 38 min). The first reaction is catalyzed by blood esterase while the second reaction takes place in the liver. Although found in a small amount (depending on the time of abuse and time of sample collection), 6-monoacetylmorphine is a specific marker of heroin abuse. Several GC/MS methods have been described in the literature to confirm and quantitate the amount of 6-monoacetylmorphine along with morphine in biological fluids.[52]

Moriya and Hashimato studied the distribution of free and conjugated morphine in body fluids and tissue in fatal heroin overdoses.[53] Blood levels of free and total morphine were very site dependent. The proportion of free morphine in various fluids and tissue ranged from 23.0% to 98.8% of total morphine and less than 30% in the stomach and urine. Free and total morphine ratio is helpful for determining the cause of death from heroin abuse. Staub et al. reported that free morphine in blood represented 50–100% in 44 individuals who died rapidly after heroin abuse while free morphine represented 20–40% of total morphine in eight individuals who survived for a time.[54]

LLE has also been used to extract heroin metabolites. Moriya and Hashimoto added 1 mL of saturated acetate buffer (pH 9.0) in 2 mL of urine and extracted heroin metabolites using chloroform–isopropanol (3 : 1 by vol). Trifluoroacetic anhydride was used for derivatization.[53]

3.6 Gas Chromatography/Mass Spectrometry Confirmation of Tetrahydrocannabinol (Marijuana)

THC is the most active constituent of marijuana. It is administered either orally or by smoking. THC is metabolized to two monohydroxy compounds, 11-hydroxy-THC, and 8-β-hydroxy-THC. The major metabolite is 11-carboxy-THC (THC-acid). Unchanged THC is present only in trace amount in urine and 11-hydroxy-THC as a conjugate account for 25% of THC metabolite.[55] Therefore, GC/MS confirmation of THC abuse is targeted towards identification and quantification of 11-carboxy-THC. The metabolite is polar, and the carboxy and phenolic hydroxyl groups of the molecule should be derivatized prior to GC/MS analysis. Derivatization can be achieved by using pentafluoropropionic anhydride and pentafluoropropanol. The derivatization reaction can be achieved by heating the reaction mixture at 75 °C for 15 min.[56] Alternatively, silyl derivatives of THC-acid using BSTFA as a derivatizing agent can also be used. Costantino et al. used a SPE and derivatization with BSTFA for confirmation of THC metabolite.[57] The authors used SPE (United Technology, Bristol, PA) for extraction using 3 mL of specimen. The extracts were derivatized using BSTFA containing 1% TMCS. The authors used selected ion monitoring (SIM), using ions with *m*/*z* 371, 473 and 488 for THC and *m*/*z* 374, 476 and 491 for d_3-THC, the internal standard. Baker et al. also used BSTFA for derivatization of THC-acid but used a LLE instead of SPE. The authors carried out base hydrolysis using 1.0 M potassium hydroxide. Then the specimen was acidified and LLE was carried out using hexane–ethyl acetate (70 : 30 by vol).[58] McCurdy et al. also used a SPE using a C_{18} bonded phase column. Following base hydrolysis, specimen was acidified and then

applied to a SPE column. The authors used the propyl derivative for GC/MS analysis.[59]

3.7 Gas Chromatography/Mass Spectrometry Confirmation of Urinary Phencyclidine

PCP is extensively metabolized and less than 10% is excreted unchanged in urine. The GC/MS confirmation procedure for PCP is a method to show conclusively the present or absence of the parent drug. No derivatization is necessary and D5-PCP can be used as an internal standard. Wu et al. described a SPE of PCP using a Bond Elute column. The authors extracted PCP from 5 mL of urine and used difluorophencyclidine as an internal standard. The SPE column was washed with 1 M acetic acid and methanol prior to extraction. PCP was eluted from the column using 2% ammonium hydroxide in methanol. After evaporation of solvent, the dry residue was reconstituted in 50 μL of ethyl acetate. Quantitation was achieved using m/z 242 for the PCP and m/z 278 for the internal standard.[23]

A number of PCP analogs with similar pharmacological properties are also found in street drugs. These drugs include phencyclohexamine, thienylcyclohexylpiperidine, phencyclohexyl pyrrolodine, and phencyclopentylpiperidine. These drugs also can be analyzed by GC/MS.

4 ANALYSIS OF DRUGS

4.1 Analysis of Benzodiazepines

Benzodiazepines are a group of structurally similar compounds used as sedative hypnotic and anticonvulsant drugs. Analysis of benzodiazepines is not mandated by Federal Guidelines. Benzodiazepines are all weak bases and produce low plasma concentrations in therapeutic use. In contrast to previous drugs discussed, plasma or serum estimation of benzodiazepine concentrations is more common. Benzodiazepines can be analyzed as a class using high-performance liquid chromatography (HPLC), or GC. GC/MS analysis is less common in clinical laboratories. GC/MS analysis is important in the forensic toxicology laboratory if benzodiazepine is involved as a part of a drug overdose. Benzodiazepines can be extracted from basic urine using 1-chlorobutane. Alternatively, a mixture of toluene–heptane–isoamyl alcohol (76 : 20 : 4) can be used for extraction.[60] Usually benzodiazepines can be analyzed by GC or GC/MS without derivatization.

He et al. have described a HPLC determination of benzodiazepines in forensic samples using photo-diode array detection.[61] This method is useful in confirming the chemical identity of eluting peaks by obtaining additional spectral information. The authors added 2 μg of triazolam as an internal standard to 1 mL of whole blood, plasma or another biological matrix. Then 0.5 mL of 1 M potassium carbonate was added to make the specimen alkaline. Benzodiazepines were extracted using 8 mL of hexane–ethyl acetate (7 : 3 by vol). After separating the organic layer from the aqueous phase, the organic layer was evaporated to dryness and then reconstituted with 100 μL of methanol–water (50 : 50 by vol) for injection into the chromatographic system. For urine samples, acid hydrolysis was performed using 37% hydrochloric acid.

4.2 Date Rape Drug Rohypnol®

Rohypnol® is the trade name of flunitrazepam, a benzodiazepine. The chemical structure of the drug is given in Figure 11(a). The drug is available in some Western European countries and South America for use as a hypnotic and anesthetic agent and can be administered orally or by IV injection in 2 mg dose. The blood concentration after 2 h of an oral dose was 6 ng mL^{-1} and declined to 3 ng mL^{-1} after 4 h. Maximal plasma concentration in a 2 mg chronic dose ranged from 10 to 20 ng mL^{-1}.

In response to reported cases of sexual assaults involving flunitrazepam, Hoffman–LaRoche marketed a new formulation which takes about 20 min to dissolve and turns bright blue in lightly colored liquid.

Flunitrazepam cross-reacts with commercially available immunoassays for the screening of benzodiazepines in urine. These includes enzyme immunoassays and fluorescence polarization immunoassays. Flunitrazepam can be analyzed using a HPLC method for benzodiazepines.[62] A GC method for the simultaneous determination of clonazepam and flunitrazepam has been described by Arthur et al.[63]

4.3 Analysis of Barbiturates

Barbiturate concentrations can be determined by ultraviolet spectrophotometry. Barbiturates are extracted from serum using chloroform and then back-extracted into sodium hydroxide. Aliquots of alkaline fraction are

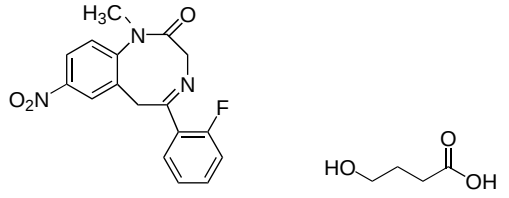

(a) Flunitrazepam (Rohypnol®) (b) γ-Hydroxy butyric acid (GABA)

Figure 11 Chemical structures of date rape drugs (a) Rohypnol® and (b) GABA.

adjusted to pH 10 and pH 13 and a differential spectrum is recorded from 280 nm to 226 nm. The absorption difference between 240 and 260 nm is proportional to the amount of barbiturates present. However, the most accurate way to estimate barbiturate concentration in serum is by HPLC. Immunoassays are also available for rapid screening of barbiturates.

4.4 Date Rape Drug γ-Hydroxybutyric Acid

GABA is an endogenous constituent of mammalian brain and acts as a neurotransmitter. Until 1990 it was sold as an unapproved food additive in health food stores and became popular among athletes as an alternative to steroids. Recent reports indicate that it is used in sexual assault cases. The structure of the drug is given in Figure 11(b).

Illicit use of GABA often involves oral doses of one teaspoon (2.5 g, or 35 mg kg^{-1} in a 70 kg man). A 25 mg kg^{-1} oral dose caused dizziness or drowsiness in adult subjects and the average plasma concentration was 80 μg mL^{-1} at 0.5 h. Blood GABA concentrations > 260 μg mL^{-1} caused deep sleep, levels of 156–260 μg mL^{-1} caused moderate sleep and levels of 52–156 μg mL^{-1} caused light sleep.[64]

4.4.1 Analysis of γ-Hydroxybutyric Acid

GABA can be analyzed using HPLC with ultraviolet/visible (UV/VIS) spectrophotometric detection or thermospray mass spectrometry.[65] The authors analyzed solid or liquid samples seized from the black market for analysis of GABA or γ-butyrolactone (GBL). Solid samples were prepared by dissolving about 20 mg specimen in 4 mL of 50% methanol in water, while liquid samples were prepared by diluting them 50-fold with 50% methanol in water. The HPLC column used was C_{18} μ Bondapak (3.9 mm × 30 cm, 10 μm particle size) from Waters Chromatography Division, Millipore, Bedford, MA. The binary mobile phase consisted of 70% buffer, 30% methanol at a flow rate of 1.0 mL min^{-1}. The buffer was 10 mM KH_2PO_4, with pH adjusted to 3 using phosphoric acid. The elution of GABA and GBL were measured at 215 nm. GABA eluted before (5 min) GBL (6.8 min).

4.4.2 High-performance Liquid Chromatography Thermospray Mass Spectrometry

Because phosphate buffer is incompatible with the mass spectrometric system, the authors used 85% acetate buffer and 15% methanol at a flow rate of 0.75 mL min^{-1} for HPLC thermospray mass spectrometric analysis. The acetate buffer was 25 mM ammonium acetate with pH adjusted to 4 with acetic acid. The ion at m/z 122 in GABA mass spectrum was due to $[M + NH_4]^+$.

4.4.3 Gas Chromatography/Mass Spectrometry Analysis of γ-Hydroxybutyric Acid

GABA can also be analyzed after derivatization using GC/MS.[66] The authors used δ-hydroxyvaleric acid (DHV) as an internal standard. The free carboxylic group was esterified first using pentafluorobenzyl bromide. Then the free hydroxyl group was derivatized using BSTFA. The authors used an OV-17 column with an oven temperature of 185 °C. Carrier gas was helium with a linear flow rate of 20 mL min^{-1}. The injector port temperature was 230 °C. The authors used electron impact with SIM.

In the mass spectrum of derivatized GABA, intense peaks were observed at m/z 341, 255 and 181. The corresponding peaks in the derivatized internal standard were at m/z 355, 255, and 181.

4.5 Confirmation of Lysergic Acid Diethylamide

Increased rates of abuse of LSD in recent years have been reported in the literature.[67] The concentration of LSD in urine may rapidly decrease to 1 ng mL^{-1} owing to rapid elimination (half-life: 3.6 h). The low concentration of LSD in biosamples is an analytical challenge for detection and quantification. Several methods using GC/MS, gas chromatography/tandem mass spectrometry (GC/MS/MS), liquid chromatography/mass spectrometry (LC/MS) or liquid chromatography/tandem mass spectrometry (LC/MS/MS) have been described in the literature.[68–71] Some methods do not achieve the desired sensitivity. Musshoff and Daldrup have described a GC/MS protocol for confirmation and quantification of LSD in serum. The authors used LLE and trimethylsilyl derivatization prior to GC/MS analysis. The linearity of the assay was 0.1–10 ng mL^{-1} of serum LSD concentration.[72] Clarkson et al. described GC/MS analysis of iso-LSD from urine. After extraction iso-LSD was converted into LSD by sodium ethoxide in ethanol. The authors also used the trimethylsilyl derivative of LSD for GC/MS analysis.[73]

Recently, De Kanel et al. described a method for analysis of LSD in blood, serum, plasma and urine using LC/MS/MS.[74] The authors used LSD-d_3 as the internal standard. The authors used an automated extraction system using Zymark Rapid Trace modules (Hopkinton, MA). This system uses a SPE method. The HPLC column used was Zorbax-SB-phenyl 4.6 mm × 7.5 cm column (Mac-mod Analytical, Chadds Ford, PA). Positive electron spray ionization tandem mass spectrometry was used for identification of LSD. No derivatization was

necessary. The limit of quantification was 0.5 ng mL^{-1}. White et al. used isotope dilution mass spectrometry for forensic confirmation of LSD in urine using LC/MS. The authors also used LSD-d_3 as the internal standard. The DL can be achieved as low as 0.1 ng mL^{-1} using SIM.[75] Hoja et al. described a method for confirmation of LSD and *N*-demethyl-LSD using LC/MS with electrospray ionization. The authors used SPE and a Nucleosil C_{18} column for liquid chromatographic separation. A mixture of 2 mM ammonium formate buffer (pH 3.0) and acetonitrile (70 : 30 by vol) was used as mobile phase. Limits of quantitation were 0.05 and 0.10 ng mL^{-1} for LSD and *N*-demethyl-LSD respectively.[76]

4.6 Hair Analysis for Drugs of Abuse

In recent years, significant numbers of papers have appeared in the literature describing the usefulness of hair as a specimen for drugs of abuse testing. Deposition of drugs in skin, nails and hair had been demonstrated by dermatologists for a long time. However, only recently has hair been used as a biological specimen for drugs of abuse testing. A number of drugs have been reported to be detectable in the hair of chronic drug abusers. Amphetamines, cocaine, PCP, opiates and nicotine are known to deposit in hair.[77–83] LSD poses a particular problem to the analyst as only low concentrations are found in urine, and it shows rapid metabolism. The methods for analysis of hair include immunological methods, chromatographic methods, and electrophoretic/electrokinetic methods.

Hair analysis requires a relatively complex extraction step. The washing step is essential to remove nonendogenous nonspecific materials like sweat, hair treatment chemicals etc. The washing step is usually followed by a digestion step where the drug is solubilized. This can be achieved by digestion of the hair with alkali or enzyme. Radioimmunoassay was adopted for analysis of drugs isolated from hair. Commercially available EMIT assay for testing of drugs of abuse do not usually have enough sensitivity.

Chromatographic methods are useful for analysis of drugs of abuse in hair specimens. TLC with fluorescence detection was among the early methods to be applied for detection of morphine. GC with traditional flame ionization detection has very limited sensitivity and is not useful. GC/MS is by far the most important analytical technique for the analysis of hair specimens. Usually the sensitivity achieved by table-top GC/MS is sufficient for routine determination of many drugs present in hair down to a concentration of 0.5–1.0 ng mg^{-1} of hair. However, for THC and LSD the sensitivity of GC/MS is a problem. THC and THC-COOH have been determined by GC/MS/MS.[83] HPLC coupled with electrospray ionization mass spectrometry is very useful in the analysis of hair specimens in terms of both selectivity and specificity.

ABBREVIATIONS AND ACRONYMS

BE	Benzoylecgonine
BSTFA	*N*-*O*-bis(trimethylsilyl) Trifluoroacetamide
CE	Cocaethylene
DHHS	Department of Health and Human Services
DHV	δ-Hydroxyvaleric Acid
DL	Detection Limit
EME	Ecgonine Methyl Ester
EMIT	Enzyme Multiplied Immunoassay Technique
GABA	γ-Hydroxybutyric Acid
GBL	γ-Butyrolactone
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GC/MS/MS	Gas Chromatography/Tandem Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
IV	Intravenous
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LDH	Lactate Dehydrogenase
LLE	Liquid–Liquid Extraction
LSD	Lysergic Acid Diethylamide
MDA	3,4-Methylenedioxyamphetamine
MDMA	3,4-Methylenedioxymethamphetamine
MRO	Medical Review Officer
NAD	Nicotinamide Adenine Dinucleotide
NADH	Reduced Nicotinamide Adenine Dinucleotide
NBE	Norbenzoylecgonine
PCP	Phencyclidine
SAMHSA	Substance Abuse and Mental Health Services Administration
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
THC	Tetrahydrocannabinol
TLC	Thin-layer Chromatography
TMCS	Trimethylsilylchlorosilane
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. S. Kaye, H.B. Haah, 'Terminal Blood Alcohol Concentrations in Ninety Four Fatal Cases of Alcoholism', *J. Am. Med. Assoc.*, **165**, 451–452 (1957).
2. C.L. Winek, 'Reliability of 22 Hour Postmortem Blood and Gastric Alcohol Samples', *J. Am. Med. Assoc.*, **233**, 912 (1975).
3. J.D.H. Cooper, 'Determination of Blood Alcohol by Gas Chromatography', *Clin. Chim. Acta.*, **33**, 483–485 (1971).
4. B.R. Manno, J.E. Manno, 'A Simple Approach to Gas Chromatographic Microanalysis of Alcohol in Blood and Urine by a Direct Injection Method', *J. Anal. Toxicol.*, **2**, 257–261 (1978).
5. Z. Penton, 'Head Space Measurement of Ethanol in Blood by Gas Chromatography with a Modified Autosampler', *Clin. Chem.*, **31**, 439–441 (1985).
6. W.E. Pereira, T.C. Summons, T.C. Rindfleisch, A.M. Duffield, 'The Determination of Ethanol in Blood and Urine by Mass Fragmentography', *Clin. Chem. Acta*, **51**, 109–112 (1974).
7. A. Tangerman, 'Highly Sensitive Gas Chromatographic Analysis of Ethanol in Whole Blood, Serum, Urine and Fecal Supernatants by Direct Injection Method', *Clin. Chem.*, **43**, 1003–1009 (1997).
8. N.R. Badcock, D.A. O'Reilly, 'False Positive EMIT Ethanol Screen with Postmortem Infant Plasma', *Clin. Chem.*, **38**, 434 (1992).
9. W. Thompson, D. Malhotra, D. Schammel, W. Blackwell, M. Ward, A. Dasgupta, 'False Positive Ethanol Determination by Enzymatic Assay: Elimination of Interference by Measuring Alcohol in Protein Free Ultrafiltrates', *Clin. Chem.*, **40**, 1594–1595 (1994).
10. E. Spratt, G. Vallaro, 'LC/MS with a Particle Beam Interference in Forensic Toxicology', in *Clinics in Laboratory Medicine*, ed. B. Gerson, W.B Saunders Company, Philadelphia, PA, 651–663, 1998.
11. A. Warner, 'Interference of Common Household Chemicals in Immunoassay Methods of Drugs of Abuse Testing', *Clin. Chem.*, **35**, 648–651 (1989).
12. R.E. Struempler, 'Excretion of Codeine and Morphine Following Injestion of Poppy Seeds', *J. Anal. Toxicol.*, **11**, 97–99 (1991).
13. G.F. Jackson, J.J. Saady, A. Poklis, 'Urinary Excretion of Benzoylecgonine Following Ingestion of Health Inca Tea', *Forensic Sci. Int.*, **49**, 57–64 (1991).
14. R.E. Struempler, G. Nelson, T.M. Urry, 'A Positive Cannabinoids Workplace Drug Test Following the Ingestion of Hemp Seed Oil', *J. Anal. Toxicol.*, **21**, 283–285 (1997).
15. S.J. Mule, P. Lomax, S.J. Gross, 'Active and Realistic Passive Marijuana Exposure Tested by Three Immunoassays and GC/MS', *J. Anal. Toxicol.*, **12**, 113–116 (1988).
16. J. Oyler, W.D. Darwin, E.J. Cone, 'Cocaine Contamination of United States Paper Currency', *J. Anal. Toxicol.*, **20**, 213–216 (1996).
17. A. Negrusz, J.L. Perry, C. Moore, 'Detection of Cocaine on Various Denominations of United States Currency', *J. Forensic Sci.*, **43**, 626–629 (1998).
18. M.A. ElSohly, 'Urinalysis and Casual Handling of Marijuana and Cocaine', *J. Anal. Toxicol.*, **14**, 46 (1991).
19. R.W. Taylor, S.D. Le, 'Simultaneous Identification of Amphetamine and Methamphetamine Using Solid-phase Extraction and Gas Chromatography/Nitrogen Phosphorus Detection or Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **13**, 293–295 (1989).
20. C.L. Hornbeck, R.J. Czarny, 'Quantitation of Methamphetamine and Amphetamine in Urine by Capillary GC/MS. Part I. Advantages of Trichloroacetyl Derivatization', *J. Anal. Toxicol.*, **13**, 144–149 (1989).
21. C.L. Hornbeck, R.J. Czarny, 'Quantitation of Methamphetamine and Amphetamine in Urine by Capillary GC/MS. Part II. Derivatization with 4-Carbethoxyhexafluoro Butyryl Chloride', *J. Anal. Toxicol.*, **13**, 257–262 (1989).
22. H. Gjerde, I. Hasvold, G. Pettersen, A.S. Christophersen, 'Determination of Amphetamine and Methamphetamine in Blood by Derivatization with Perfluorooctanoyl Chloride and Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **17**, 65–68 (1993).
23. A. Wu, T.A. Onigbinde, S. Wong, K. Johnson, 'Evaluation of Full Scanning GC/Ion Trap MS Analysis of NIDA Drugs of Abuse Urine Testing in Urine', *J. Anal. Toxicol.*, **16**, 202–206 (1992).
24. W.C. Thompson, A. Dasgupta, 'Microwave-induced Rapid Preparation of Fluoro Derivatives of Amphetamine, Methamphetamine, and 3,4-Methylenedioxymethamphetamine for GC/MS Confirmation Assay', *Clin. Chem.*, **40**, 1703–1706 (1994).
25. J. Jonsson, R. Kronstrand, M. Hatanapaa, 'A Convenient Derivatization Method for the Determination of Amphetamine and Related Drugs in Urine', *J. Forensic Sci.*, **41**, 148–151 (1996).
26. R. Meatherall, 'Rapid GC/MS Confirmation of Urinary Amphetamine and Methamphetamine as Their Propylchloroformate Derivatives', *J. Anal. Toxicol.*, **19**, 316–322 (1995).
27. A. Dasgupta, A. Hart, 'Distinguishing Amphetamine, Methamphetamine and 3,4-Methylenedioxymethamphetamine from Other Sympathomimetic Amines After

Rapid Derivatization with Propyl Chloroformate and Analysis by Gas Chromatography/Chemical Ionization Mass Spectrometry', *J. Forensic Sci.*, **42**, 106–110 (1997).
28. A. Dasgupta, J. Spies, 'A Rapid Novel Derivatization of Amphetamine and Methamphetamine Using 2,2,2-Trichloroethyl Chloroformate for Gas Chromatography Electron Ionization and Chemical Ionization Mass Spectrometry', *Am. J. Clin. Pathol.*, **109**, 527–532 (1998).
29. C.L. Hornbeck, J.E. Carrig, R.J. Czarny, 'Detection of GC/MS Artifact Peak as Methamphetamine', *J. Anal. Toxicol.*, **17**, 257–263 (1993).
30. A. Dasgupta, C. Gardner, 'Distinguishing Amphetamine and Methamphetamine from Other Interfering Sympathomimetic Amines After Various Fluoro Derivatization and Analysis by Gas Chromatography/Chemical Ionization Mass Spectrometry', *J. Forensic Sci.*, **40**, 1077–1081 (1995).
31. A. Dasgupta, A. Hart, P. Humphrey, W. Blackwell, 'Gas Chromatography/Electron Ionization and Chemical Ionization Mass Spectrometric Analysis of Urinary Phenmetrazine After Derivatization with 4-Carbethoxyhexafluorobutyryl Chloride – a Novel Derivative', *J. Forensic Sci.*, **43**, 636–640 (1998).
32. P.I. Jatlow, D.N. Bailey, 'Gas Chromatographic Analysis for Cocaine in Human Plasma with Use of a Nitrogen Detector', *Clin. Chem.*, **21**, 1918–1921 (1975).
33. J.I. Javaid, H. Dekiremenjian, E.G. Brunngraber, J.M. Davis, 'Quantitative Determination of Cocaine and its Metabolites Benzoylecgonine and Ecgonine by Gas–Liquid Chromatography', *J. Chromatogr.*, **110**, 141–149 (1975).
34. P.I. Jatlow, J.D. Elsworth, C.W. Bradberry, G. Winger, J.R. Taylor, R. Russel et al., 'Cocaethylene: A Neuropharmacologically Active Metabolite Associated with Concurrent Cocaine–Ethanol Ingestion', *Life Sci.*, **48**, 787–794 (1991).
35. D.N. Bailey, 'Plasma Cocaethylene Concentrations in Patients Treated in the Emergency Room or Trauma Unit', *Am. J. Clin. Pathol.*, **99**, 123–127 (1993).
36. R.C. Baselt, in *Analytical Procedures for Therapeutic Drug Monitoring and Emergency Toxicology*, Biomedical Publications, Davis, CA, 89–91, 1980.
37. W. Joern, 'Routine Detection of Benzoylecgonine in Urine at a Sensitivity of 35 ng mL^{-1} by a Combination of EMIT and Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **11**, 110–112 (1987).
38. A. Jenkins, B. Goldberger, 'Identification of Unique Cocaine Metabolites and Smoking By-products in Postmortem Blood and Urine Specimens', *J. Forensic Sci.*, **42**, 824–827 (1997).
39. R.W. Taylor, N.C. Jain, M.P. George, 'Simultaneous Identification of Cocaine and Benzoylecgonine Using Solid-phase Extraction and Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **11**, 233–234 (1987).
40. J. Gerlits, 'GC/MS Quantitation of Benzoylecgonine Following Liquid–Liquid Extraction', *J. Forensic Sci.*, **38**, 1210–1214 (1993).
41. W. Thompson, A. Dasgupta, 'Confirmation and Quantitation of Cocaine, Benzoylecgonine, Ecgonine Methyl Ester and Cocaethylene by Gas Chromatography/Mass Spectrometry: Use of Microwave Irradiation for Rapid Preparation of Trimethylsilyl and Tert-Butyldimethyl Silyl Derivatives', *Am. J. Clin. Pathol.*, **104**, 187–192 (1995).
42. A. Wu, D. Ostheimer, M. Cremese, E. Forte, D. Hill, 'Characterization of Drug Interferences in Gas Chromatography/Mass Spectrometric Confirmation of Targeted Drugs in Full Scan and Selected Ion Monitoring', *Clin. Chem.*, **40**, 216–220 (1994).
43. A. Dasgupta, C. Mahle, J. McLemore, 'Elimination of Fluconazole Interference in Gas Chromatography/Mass Spectrometric Confirmation of Benzoylecgonine, the Major Metabolite of Cocaine Using Pentafluoropropionyl Derivative', *J. Forensic Sci.*, **41**, 511–513 (1996).
44. S.J. Mule, G.A. Casella, 'Confirmation and Quantitation of Cocaine, Benzoylecgonine, Ecgonine Methyl Ester in Human Urine by GC/MS', *J. Anal. Toxicol.*, **12**, 153–155 (1988).
45. R.C. Baselt, in *Analytical Procedures for Therapeutic Drug Monitoring and Emergency Toxicology*, Biomedical Publications, Davis, CA, 186–189, 1980.
46. L.A. Broussard, L.C. Presley, T. Pittman, R. Clouette, G.H. Wimbish, 'Simultaneous Identification and Quantitation of Codeine, Morphine, Hydrocodone and Hydromorphine in Urine as Trimethylsilyl and Oxime Derivatives by Gas Chromatography/Mass Spectrometry', *Clin. Chem.*, **43**, 1029–1032 (1997).
47. B.H. Chen, E.H. Taylor, A.H. Pappa, 'Comparison of Derivatives for Determination of Codeine and Morphine by Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **14**, 12–17 (1990).
48. G.F. Grinstead, 'A Closer Look at Acetyl and Pentafluoropropionyl Derivatives for Quantitative Analysis of Morphine and Codeine by Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **15**, 293–298 (1991).
49. J. Fenton, J. Mummert, M. Childers, 'Hydromorphone and Hydrocodone Interference in GC/MS Assays for Morphine and Codeine', *J. Anal. Toxicol.*, **18**, 159–164 (1994).
50. D.R. Knapp, *Handbook of Analytical Derivatization Reactions*, Wiley Interscience, New York, 342–343, 1979.
51. M. Cremese, A. Wu, G. Cassella, E. O'Corner, K. Rymut, D.W. Hill, 'Improved GC/MS Analysis of Opiates with Use of Oxime-TMS Derivatives', *J. Forensic Sci.*, **43**, 1220–1224 (1998).
52. B.A. Goldberger, R.J. Cone, T.M. Grant, Y.H. Caplan, B.S. Levine, J.E. Smialek, 'Disposition of Heroin and its Metabolites in Heroin Related Deaths', *J. Anal. Toxicol.*, **18**, 22–28 (1994).

53. F. Moriya, Y. Hashimoto, 'Distribution of Free and Conjugated Morphine in Body Fluids and Tissue in a Fatal Heroin Overdose: Is Conjugated Morphine Stable in Postmortem Specimens?', *J. Forensic Sci.*, **42**, 736–740 (1997).
54. C. Staub, R. Jeanmondo, O. Fryc, 'Morphine in Postmortem Blood: Its Importance for the Diagnosis of Deaths Associated with Opiate Addiction', *Int. J. Legal Med.*, **104**, 39–42 (1990).
55. L. Lemberger, A. Rubin, 'The Physiological Disposition of Marijuana in Man', *Life Sci.*, **17**, 1637–1642 (1975).
56. W.A. Joern, 'Detection of Past and Recurrent Marijuana Use by a Modified GC/MS Procedure', *J. Anal. Toxicol.*, **11**, 49–52 (1987).
57. A. Costantino, R.H. Schwartz, P. Kaplan, 'Hemp Oil Ingestion Causes Positive Urine Tests for Tetrahydrocannabinol Carboxylic Acid', *J. Anal. Toxicol.*, (1997).
58. T.S. Baker, J.V. Harry, J.W. Russell, R. Lee Myers, 'Rapid Method for GC/MS Confirmation of 11-nor-9-Carboxy-Δ^9-Tetrahydrocannabinol in Urine', *J. Anal. Toxicol.*, **8**, 255–259 (1984).
59. H.H. McCurdy, L.S. Lewellen, L. Callahan, P.S. Childs, 'Evaluation of the Ion Trap Detector for the Detection of 11-nor Δ^9 THC Carboxylic Acid in Urine After Extraction by Bonded Phase Absorption', *J. Anal. Toxicol.*, **10**, 175–177 (1986).
60. A. Sioufi, J.P. Dubois, 'Chromatography of Benzodiazepines', *J. Chromatogr.*, **531**, 459–480 (1990).
61. W. He, N. Parissis, T. Kiratzidis, 'Determination of Benzodiazepines in Forensic Samples by HPLC with Photo-diode Array Detection', *J. Forensic Sci.*, **43**, 1061–1067 (1998).
62. W. He, N. Parissis, 'Simultaneous Determination of Flunitrazepam and its Metabolites in Plasma and Urine by HPLC/DAD After Solid-phase Extraction', *J. Pharm. Biomed. Anal.*, **16**, 707–714 (1997).
63. J. Arthur, D. Silva, I. Bekersky, 'Determination of Clonazepam and Flunitrazepam in Blood by Electron Capture Gas Liquid Chromatography', *J. Chromatogr.*, **99**, 447–560 (1974).
64. M. Helrich, T.C. McAslan, S. Skolnick, S.P. Bessman, 'Correlation of Blood Levels of 4-Hydroxybutarate with State of Consciousness', *Anesthesiology*, **25**, 771–775 (1965).
65. M.Z. Mesmer, R.D. Satzger, 'Determination of γ-Hydroxybutyrate (GABA) and γ-Butyrolactone (GBL) by HPLC/UV/VIS Spectrometry and HPLC/Thermospray Mass Spectrometry', *J. Forensic Sci.*, **43**, 489–492 (1988).
66. M. Eli, F. Cattabeni, 'Endogenous γ-hydroxybutyrate in Rat Brain Areas: Postmortem Changes and Effects of Drug Interfering with Gamma Aminobutyric Acid Metabolism', *J. Neuro. Chem.*, **41**, 524–530 (1983).
67. B.D. Paul, J.M. Mitchell, R. Burbage, M. Moy, R. Sroka, 'Gas Chromatographic–Electron Impact Mass Fragmentometric Determination of Lysergic Acid Diethylamide in Urine', *J. Chromatogr.*, **120**, 103–112 (1992).
68. J. Cai, J.D. Henion, 'Elucidation of LSD In Vitro Metabolism by Liquid Chromatography and Capillary Electrophoresis Coupled with Tandem Mass Spectrometry', *J. Anal. Toxicol.*, **20**, 27–37 (1996).
69. C.C. Nelson, R.L. Foltz, 'Chromatographic and Mass Spectrometric Methods for Determination of Lysergic Acid Diethylamide (LSD) and Metabolites in Body Fluid', *J. Chromatogr.*, **118**, 97–109 (1992).
70. C.C. Nelson, R.L. Foltz, 'Determination of Lysergic Acid Diethylamide (LSD), iso-LSD, and *N*-Demethyl LSD in Body Fluids by Gas Chromatography/Tandem Mass Spectrometry', *Anal. Chem.*, **64**, 1578–1585 (1992).
71. H.K. Lim, D. Andrenyak, P. Francom, R.L. Foltz, R.T. Jones, 'Quantification of LSD (Lysergide) and *N*-Demethyl-LSD in Urine by Gas Chromatography/Resonance Electron Capture Ionization Mass Spectrometry', *Anal. Chem.*, **60**, 1420–1425 (1988).
72. F. Musshoff, T. Daldrup, 'Gas Chromatography/Mass Spectrometric Determination of Lysergic Acid Diethylamide (LSD) in Serum Samples', *Forensic Sci. Int.*, **88**, 133–140 (1997).
73. E.D. Clarkson, D. Lesser, B.D. Paul, 'Effective GC/MS Procedure for Detecting *iso*-LSD in Urine After Base Catalyzed Conversion to LSD', *Clin. Chem.*, **44**, 287–292 (1998).
74. J. De Kanal, W.E. Vickery, B.A. Waldner, R.M. Monahan, F.X. Diamond, 'Automated Extraction of Lysergic Acid Diethylamide (LSD) and *N*-Demethyl-LSD from Blood, Serum, Plasma, and Urine Samples Using the Zymark Rapid Trace with LC/MS/MS Confirmation', *J. Forensic Sci.*, **43**, 622–625 (1998).
75. S.A. White, A.S. Kidd, K.S. Webb, 'The Determination of Lysergide (LSD) in Urine by High Performance Liquid Chromatography/Isotope Dilution Mass Spectrometry', *J. Forensic Sci.*, **44**, 375–379 (1999).
76. H. Hoja, P. Marquet, B. Verneuil, H. Lofti, J.L. Dupuy, G. Lachatre, 'Determination of LSD in Urine by Liquid Chromatography Coupled to Electrospray Ionization Mass Spectrometry', *J. Chromatogr.*, **692**, 329–335 (1997).
77. S. Suzuki, H. Inoue, H. Hori, S. Inayama, 'Analysis of Methamphetamine in Hair, Nail, Sweat and Saliva by Mass Fragmentometry', *J. Anal. Toxicol.*, **13**, 176–178 (1989).
78. D.A. Kidwell, 'Analysis of Phencyclidine and Cocaine in Human Hair by Tandem Mass Spectrometry', *J. Forensic Sci.*, **38**, 272–284 (1993).
79. F. Tagliaro, F.P. Smith, Z. Battisti, G. Manetto, M. Marigo, 'Hair Analysis, a Novel Tool in Forensic and Biomedical Sciences: New Chromatographic and Electrophoretic/Electrokinetic Analytical Strategies', *J. Chromatogr.*, **689**, 261–271 (1997).
80. G.L. Henderson, M.R. Harkey, C. Zhou, R.T. Jones, P. Jacob, 'Incorporation of Isotopically Labeled Cocaine and Metabolites into Human Hair 1: Dose Response Relationships', *J. Anal. Toxicol.*, **20**, 1–12 (1996).

81. A.M. Baumgaetner, P.F. Jones, C.T. Black, 'Detection of Phencyclidine in Hair', *J. Forensic Sci.*, **26**, 576–581 (1981).
82. N.J. Haley, D. Hoffman, 'Analysis for Nicotine and Cotinine in Hair to Determine Cigarette Smoker Status', *Clin. Chem.*, **31**, 1598–1600 (1985).
83. T. Mieczkowski, 'A Research Note: The Outcome of GC/MS/MS Confirmation of Hair Assays on 93 Cannabinoid(t) Cases', *Forensic Sci. Int.*, **70**, 83–91 (1995).

Electroanalysis and Biosensors in Clinical Chemistry

Anthony J. Killard, Margaret Sequeira, Dermot Diamond, and Malcolm R. Smyth
Dublin City University, Dublin, Ireland

1 Introduction **1**
1.1 Introduction to Electrochemical Terms 1
1.2 Potentiometric Methods Using Ion-selective Electrodes 2
1.3 Amperometry 3
1.4 Voltammetry 4
1.5 Biosensors 6
2 Techniques Associated with Clinical Analysis **7**
2.1 Microdialysis 7
2.2 Flow Injection Analysis and Chromatographic Separation Systems 7
2.3 In Vivo Analysis 9
2.4 Automated Clinical Chemistry Analyzers 9
3 Nature of Biological Matrixes and Sample Preparation **10**
3.1 Blood 10
3.2 Urine 12
3.3 Other Body Fluids 12
3.4 Solid Tissues 13
4 Clinical Analysis **13**
4.1 Blood Gas Analysis 13
4.2 Electrolytes 17
4.3 Carbohydrates 19
4.4 Nonprotein Nitrogen 22
4.5 Vitamins 23
4.6 Lipids 25
4.7 Trace Metals 25
4.8 Amino Acids 25
4.9 Blood Group Antigens 25
4.10 DNA Biosensors 27
4.11 Hormones, Pharmaceuticals and Antibiotics 27
5 Conclusion **28**
Acknowledgments **29**
Abbreviations and Acronyms **29**
Related Articles **29**
References **29**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Electroanalysis involves the measurement or determination of a chemical or biochemical species, using a wide range of electrochemical methods. The species being measured can either bring about a direct electrochemical response that can be detected and measured, or can be induced to bring about such a response, resulting in a process that can be detected electrochemically. In both cases an interface is required between the chemical or biochemical species being measured, and the electrochemical system performing the measurement. This interface is referred to as an ''electrode''. It is the transfer of electrons or ions between the electrode and the environment being sampled which is of concern in electroanalysis.

The application of electrochemistry in clinical analysis is concerned with the measurement of chemical and biochemical species present in biological fluids (primarily blood but also urine and other body fluids and tissues). It is applied in several areas of critical care where its rapidity, sensitivity, precision, and accuracy at biologically relevant concentrations make it an ideal choice over other, typically slower, biochemical techniques. Indeed, it is the speed of electrochemical analysis that principally differentiates it from other methods. It is not universally applicable to clinical chemistry, however, and has only been applied in situations where such analyses are relevant. This has principally been in the analysis of blood gases and pH and blood electrolytes. Electrochemical methods are, however, gradually being extended to the analysis of small biochemical components such as sugars, amino acids and other metabolites, with the advent of enzyme-based ''biosensors''. There is also little doubt that the application of electrochemistry in clinical analysis will be further exploited by the use of such biosensor devices based on other forms of biorecognition, such as antibodies, cell ligands, membranes, tissues and whole cells. These biosensor devices are also amenable to other forms of detection such as acoustic wave and optical measurements. Together, these techniques offer a wide range of advantages over traditional analytical methodologies, and will undoubtedly have a rapidly increasing impact on many types of clinical analysis. This article focuses on the use of electrochemical techniques and electrochemical biosensors and their application in clinical chemistry.

1 INTRODUCTION

1.1 Introduction to Electrochemical Terms

The measurement of a chemical species in a biological fluid using electroanalytical methods is as a result of either a spontaneous oxidation/reduction (redox) reaction between the species and the electrode (termed the

''indicator'' electrode in potentiometry or the ''working'' electrode in voltammetry), or a redox reaction induced as a result of the application of an electrical potential (E) or the movement of charged species across boundaries which creates an electrical potential. When oxidation/reduction reactions occur at the electrode surface and electron transfer occurs, the electrochemical techniques are known as ''voltammetry'' (potential scan) or ''amperometry'' (constant potential). When movement of ions occurs, the technique is known as ''potentiometry''. To induce such changes and make such measurements at the indicator or working electrodes, however, one requires a complete electrical circuit. As the potential across a single electrode interface cannot be measured, it must be measured relative to a constant potential at a second interface, known as the ''reference'' electrode. The electrical circuit is commonly completed by a ''salt bridge'' between the working and reference electrodes that allows a controlled movement of ions, and by an external electrical circuit connecting the working and reference electrodes via which an external potential may be applied or measurements may be made via a voltmeter (Figure 1). This is the basis of the electrochemical cell.

The reference electrode is designed to give a stable and constant potential. Various electrodes have fulfilled this role. The standard hydrogen electrode (SHE) is the primary reference. However, it is difficult to set-up and maintain. Thus, the standard calomel electrode (SCE) was introduced, and now the silver/silver chloride (Ag/AgCl) electrode is the most common.

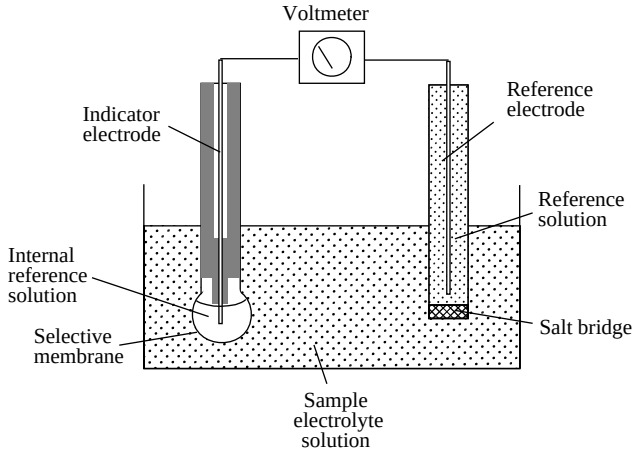

Figure 1 The electrochemical cell. Two electrodes, the indicator electrode and the reference electrode are immersed in the sample electrolyte solution. The indicator electrode contains a solution of a salt of the ion being measured and the reference electrode contains a solution of silver/silver chloride. The circuit is completed by a voltmeter to allow the external flow of electrons, and a salt bridge to allow the movement of ions.

1.2 Potentiometric Methods Using Ion-selective Electrodes

The design of the indicator electrode varies greatly depending on the particular application. Most of the electrodes in clinical chemical applications are membrane-based electrodes. These contain a selective membrane for the species being measured which

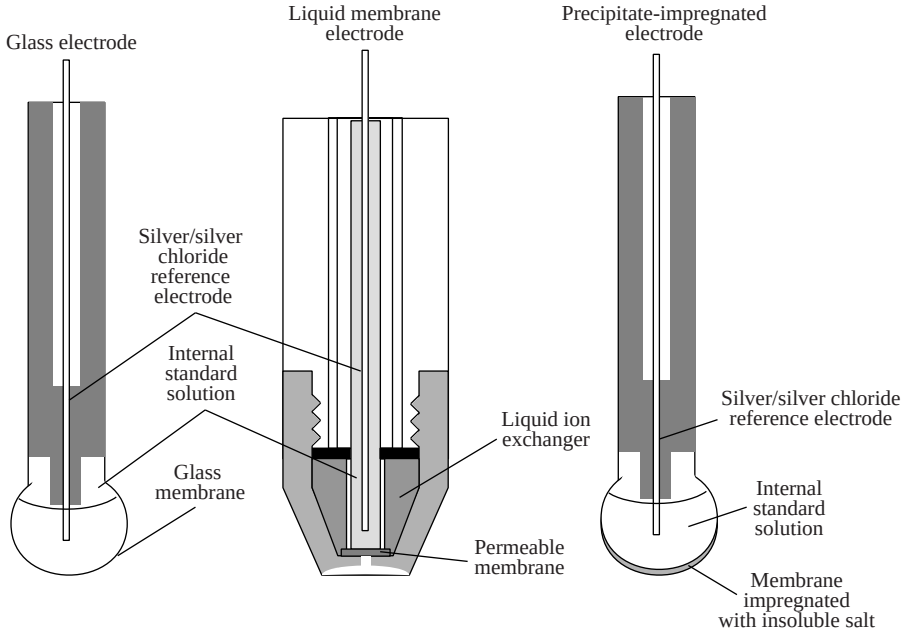

Figure 2 The principle ISEs used in clinical chemistry. The glass electrode is permeable to the movement of ions, such as H^+. The liquid membrane electrode contains a buffer or an ion exchanger specific to the ion of interest. The precipitate-impregnated electrode contains an insoluble salt of the ion to be measured. (Reproduced by permission of Orion Research Inc.)

separates the external sample solution, e.g. blood, from the inner electrolyte, which is typically composed of a salt of the ion being measured, e.g. NaCl in the case of the sodium electrode. These are referred to as ion-selective electrodes (ISEs). Several types of membrane electrode have been found application: glass, liquid membrane, precipitate-impregnated (Figure 2), and poly(vinyl chloride) (PVC) membranes containing perm-selective ionophores (Figure 3). The design and properties of these membranes will be discussed in greater depth in section 4. Other potentiometric electrodes of some note are solid-state electrodes, gas-sensing electrodes, enzyme electrodes and immunosensors. The latter incorporate immunochemical reagents and often possess additional membranes to impart a selective response to the target species.

In the two-electrode system shown in Figure 1, the potential at the reference electrode remains constant.

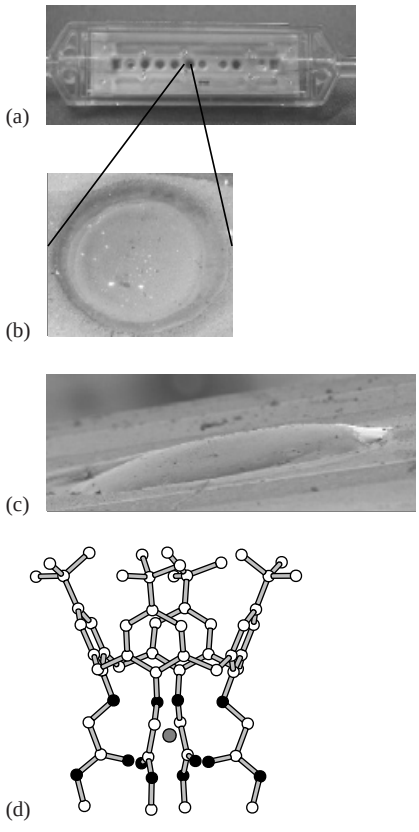

Figure 3 Microelectrode arrays based on PVC membranes doped with selective ionophores. The SenDx Medical Inc. array (a) contains miniature PVC electrodes. (b) Shown at an angle by scanning electron microscopy (SEM), (c) the membrane can be seen to bulge upwards due to the presence of buffer components beneath. These membranes contain ion-selective carriers, such as the calix[4]tetramethoxyethyl ester illustrated. (d) Carbon atoms are shown in white, oxygen in black and the sodium ion is gray.

The only variation in potential that takes place occurs at the outer membrane boundary of the indicator electrode in contact with the sample solution. This results from differences in concentration of the measured ion across the electrode membrane boundary. Because of a difference in analyte concentration on the inside and outside of the membrane boundary and the presence of a receptor capable of selectively binding this ion, the ions diffuse across the selective interface. This movement of ions across the membrane results in the formation of a potential which can be measured by the voltmeter. The concentration of ions and the potential it produces are governed by the Nernst equation, as shown in Equation (1)

$$E = E^0 + \frac{RT}{z_i F} \ln a_i \tag{1}$$

where E is the electrode potential, E^0 is the standard cell potential, R is the gas constant, T is the temperature in Kelvin, z_i is the charge of the primary ion (analyte), F is the Faraday constant, and $\ln a_i$ is the natural logarithm of the activity of analyte i. To base 10 at body temperature (37 °C), this equates to Equation (2).

$$E = E^0 + \frac{0.06155}{z_i} \log_{10} a_i \tag{2}$$

Since all potentials in the cell are constant except at the indicator electrode outer membrane boundary, then measurements of potential are directly related to measurements of activity. This is the basis of potentiometry. Activity can then be related to concentration using Equation (3)

$$a_i = f_i c_i \tag{3}$$

where c_i is the concentration of analyte and f_i is the activity coefficient of the ion i.

Potentiometry is a widely used technique in clinical chemistry, responsible for the analysis of pH and CO_2 levels, and of Na^+, Ca^+, Li^+ and K^+ electrolytes. It is also the basis of many enzyme electrodes. Problems that exist with ISEs include electrode selectivity and stability. Initially, electrodes will become coated with biological materials such as proteins. New electrodes may take a period of weeks to settle to an ideal slope. Electrodes can also be affected by other surface contaminants, loss of membrane components, membrane poisoning and electrical leakage pathways.

1.3 Amperometry

Not all species to be measured are charged in nature at all times, and those that aren't charged cannot be analyzed by potentiometric techniques. They may, however, undergo

oxidation and reduction (redox) reactions. Such reactions can occur spontaneously, but can also be induced to occur by the application of an electrical potential. Amperometry involves the application of a fixed potential to bring about either oxidation or reduction of an analyte species. These reactions will result either in the loss or gain of electrons and the current generated is proportional to the amount of analyte present. Amperometry is used as the basis for the measurement of oxygen levels in blood, and is also used in many enzyme-based electrodes that utilize oxygen or yield redox active species as the product of enzymatic reactions. Such techniques utilize a three-electrode system like that shown in Figure 4. The third electrode, called the ''auxiliary'' electrode, is required to cope with the passage of current between it and the working electrode. This auxiliary electrode is most often composed of a high-surface-area platinum mesh. The reference electrode still supplies a reference potential for this system.

A related technique is that of coulometry. Again, a fixed potential is applied to an electrochemical system, which results in all the species being measured becoming oxidized or reduced. Instead of measuring current flow, the total amount of electricity produced (coulombs) is measured according to Equation (4)

$$c = \frac{It}{nF} \tag{4}$$

where c is the molar quantity of analyte reacted, I is the current, t is the time taken for all the analyte to be titrated, n is the number of electrons involved in the reaction and F is the Faraday constant.

This technique forms the basis for the measurement of chloride concentration using coulometric titration. This technique uses a fixed current to complex silver ions with chloride in solution to convert it to insoluble AgCl (see section 4.2.5).

1.4 Voltammetry

Voltammetry covers a group of electrochemical techniques in which an applied potential in an electrochemical cell is varied in some way. The resulting relationship

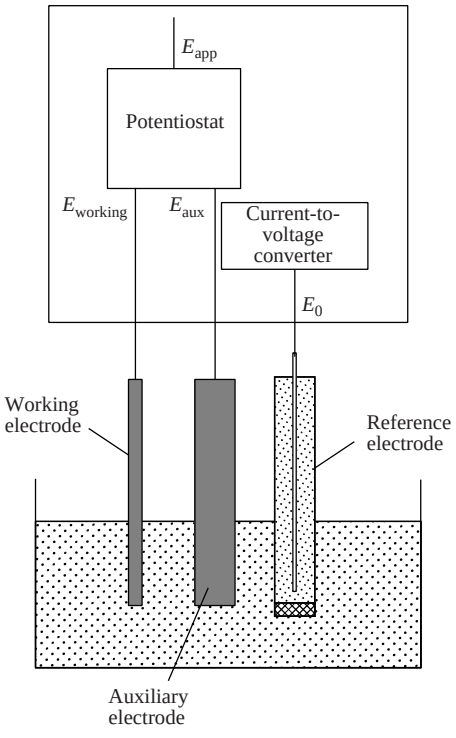

Figure 4 The amperometric electrochemical cell. In addition to the reference and working electrodes, a high-surface-area platinum auxiliary electrode is required due to the flow of current in this system. The working and auxiliary electrodes are connected to a potentiostat which maintains the applied potential, relative to the reference electrode.

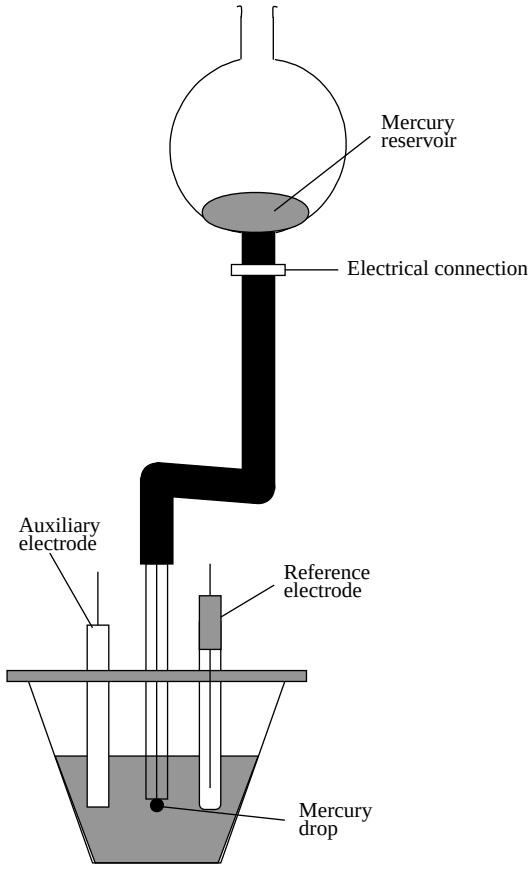

Figure 5 The DME. A reservoir of mercury supplies droplets under gravity to the end of a glass capillary. This reservoir acts as the cathode. The test solution contains the redox species to be measured. At some applied potential, the analyte undergoes reduction at the cathode, yielding an increase in current proportional to the activity of the ion.

between current and potential is known as a voltammogram. Many voltammetric techniques exist, depending on the type of potential waveform applied and the type of electrode system used. The original voltammetric technique, called polarography, utilizes a self-renewing electrode composed of mercury drops emanating from a glass capillary. A typical electrochemical cell arrangement using the dropping mercury electrode (DME) is shown in Figure 5. The DME acts as the cathode, and a pool of liquid mercury as the anode. The capillary produces extremely reproducible droplets of mercury with a defined surface area and drop rate. The electrode is immersed in the analyte solution with a supporting electrolyte. To detect the analyte, a linear voltage sweep is applied to the cell. At some applied potential, the analyte species undergoes reduction at the cathode according to the generalized reaction scheme for a metallic ion A^{n+} shown in Equation (5).

$$A^{n+} + ne^- + Hg \rightleftharpoons A^0(Hg) \qquad (5)$$

This results in an increase in current, which is proportional to the diffusion rate of the analyte to the electrode, and thus is proportional to the concentration of the analyte in solution. Figure 6 shows a polarogram for the reduction of five metallic ions in $1\,mol\,L^{-1}$ NH_3/$1\,mol\,L^{-1}$ NH_4Cl/0.002% (v/v) Triton® X-100. The fluctuating nature of the resulting signal (plot A) should be noted. This is a result of the change in the size of the electrode surface as the droplet grows and then drops.

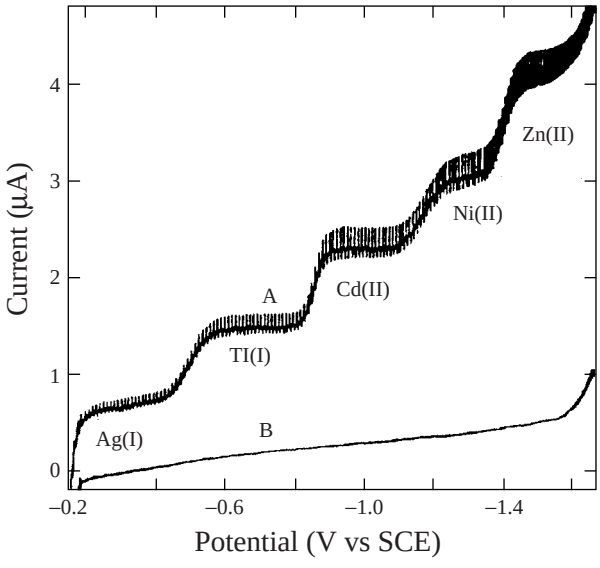

Figure 6 Polarogram of the reduction of Ag^+, Tl^+, Cd^{2+}, Ni^{2+} and Zn^{2+} at a DME in supporting electrolyte (plot A) and supporting electrolyte ($1\,mol\,L^{-1}$ NH_3/$1\,mol\,L^{-1}$ NH_4Cl/0.002% (v/v) Triton® X-100) alone (plot B). (Reprinted by permission of John Wiley & Sons, Inc., from L. Meites, *Polarographic Techniques*. Copyright © 1967.)

Measurements can be performed either by taking the average current, or confining sampling to a narrow time period in the last 5–20 ms of each drop (current-sampled voltammetry). Both of these techniques have detection limits as low as $10^{-6}\,mol\,L^{-1}$. A residual current (plot B) also results from a non-Faradaic current formed as a result of the charging of the droplets of mercury. The half-wave potential $E_{1/2}$ of the analyte is a function of its formal potential, as defined by Equation (6)

$$E_{1/2} = E^0 + \frac{0.059}{n} \log \left(\frac{D_R}{D_O} \right)^{1/2} \qquad (6)$$

where E^0 is the formal potential, n is the number of electrons taking part in the reaction and D_R and D_O are the diffusion coefficients of the reduced and oxidized forms of the analyte species, respectively.

Thus, the half-wave potential also gives information as to the nature of the analyte being measured and is independent of its concentration. Several analytes can be measured simultaneously in this manner, providing that they have sufficiently different half-wave potentials: 0.3 V and 0.2 V for two- and one-electron reductions, respectively. Suitable complexing agents may be used to modify the half-wave potentials of the measured species, making them amenable to simultaneous measurement. However, it is not very convenient and there is little inherent scope for multi-analyte measurements on some electrodes due to the broad nature of the features.

Other important forms of voltammetry with clinical relevance also exist. These use either stepped or cyclic variations in the applied potential. These variations are principally employed to discriminate between Faradaic and non-Faradaic currents, so increasing the limit of detection (LOD) to $10^{-7}\,mol\,L^{-1}$. The principle types of nonlinear techniques are illustrated in Figure 7. These are ''normal pulse'', ''differential pulse'', and ''alternating current'' voltammetry. Both linear and nonlinear techniques are used in clinical chemistry, principally for the analysis of vitamins, drugs, hormones and metals. However, in many situations, the mercury electrode is not suitable due to its limited anodic potential range. In this instance, many solid electrodes are used, such as platinum and carbon. These have detection limits of the order of $10^{-5}\,mol\,L^{-1}$, but suffer from problems of stability and fouling of the electrode surface, which limits their lifetime.

One further voltammetric technique is ''stripping voltammetry''. This system is employed in the analysis of trace metals and is often sensitive to approximately $10^{-9}\,mol\,L^{-1}$. It achieves this as a result of a pre-analytical concentration step. The apparatus for anodic stripping voltammetry is illustrated in Figure 8. It employs a hanging mercury drop electrode (HMDE). Cathodic stripping can be performed at the surface of a metal or

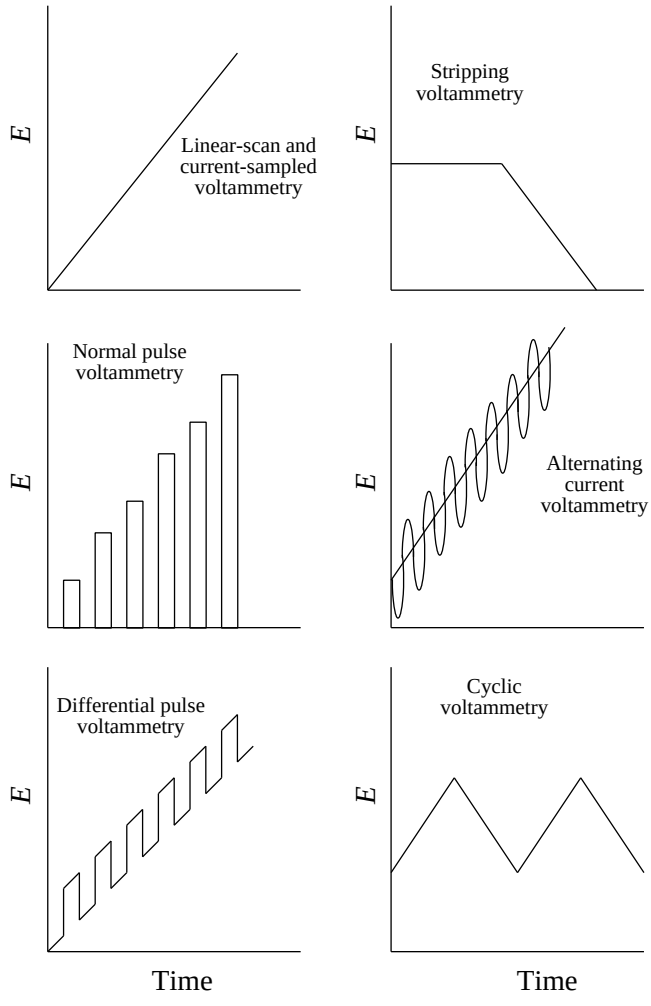

Figure 7 Principle voltammetric techniques.

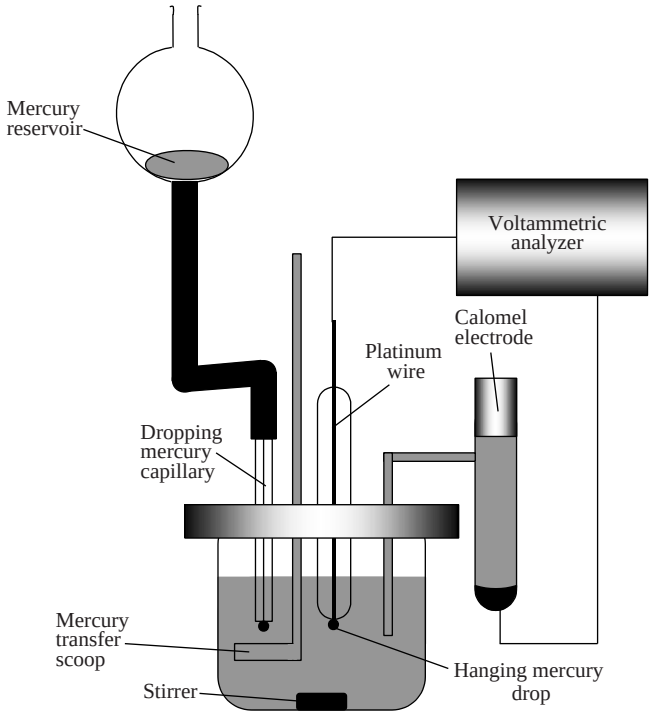

Figure 8 Voltammetric stripping analysis. A hanging mercury drop is taken from a mercury reservoir and suspended beneath a carbon wire. The analyte solution is stirred at a rate that does not disturb the hanging drop. Preconcentration of the analyte takes place at a fixed potential for a fixed time period. Stripping analysis then follows, caused by the application of an increasingly negative potential. Ions are stripped from the mercury in order of their reduction potential.

carbon electrode. Initially, the analyte is deposited on the mercury drop by applying a potential a few tenths of a volt below the half-wave potential of the ion. This is done over a carefully defined period of time, and can be adjusted to measure much lower concentrations (as low as 10^{-9} mol L^{-1} over 60 min). The deposited ion can now be measured in a number of ways. With an anodic linear scan, the applied potential is decreased. As the ion reaches its formal potential it becomes oxidized, generating an increase in current as it is stripped from the electrode surface. Stripping analysis is extremely sensitive to oxygen. Also, difficulties arise as a result of the formation of metal complexes, and as a result of the presence of inhibitors of deposition. Care should also be taken to avoid contamination of the sample with metals due to the very low concentrations being measured. For analytes that cannot be accumulated by electrolysis, techniques are available that allow for their adsorptive deposition to the electrode surface. This is referred to as ''adsorptive stripping voltammetry'', and this has been applied in the analysis of hormones, drugs and antibiotics.

1.5 Biosensors

Many varied definitions have been put forward for biosensors. Essentially, however, they have two important components: a biologically selective layer and a transduction layer (Figure 9).[1] The bioselective layer is most commonly composed of enzymes, which bring about the conversion of a substrate to its products. However, other bioselective interactions can be used, such as those involving antibodies, ligands and nucleic acids or combinations of these. Techniques using biological membranes, tissues and whole cells are also used. These interactions bring about physical or chemical changes that are converted by the transduction layer into an electrical signal. These physical or chemical changes can be in the form of the generation or consumption of electrons in the case of electrochemical transducers, changes in mass (which can be detected optically or by the use of acoustic techniques), or the generation of heat or light or gases. Biosensors have become an important element

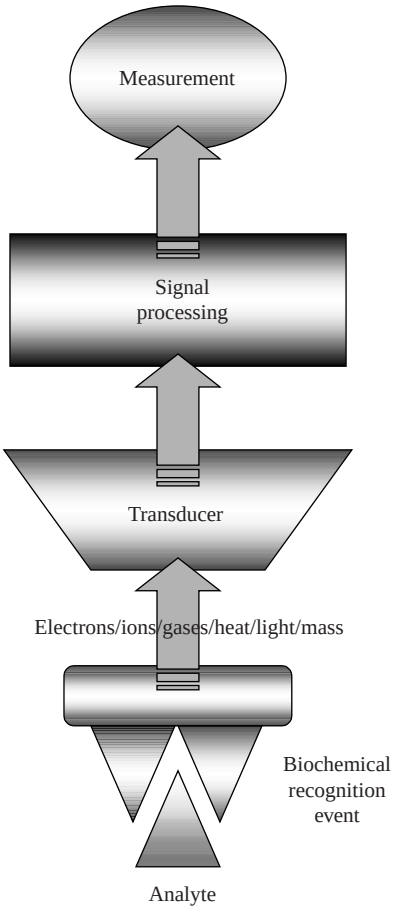

Figure 9 Fundamental elements of a biosensor. A biospecific recognition event takes place at the sensor surface. This may be the catalysis of a substrate by an enzyme, the recognition of an antibody for an antigen, nucleic acid hybridization, receptor–ligand interactions, or transport across membranes. These interactions lead to a signal which can be transduced by the sensor to an electrical signal. This in turn is processed and output as a measure of the extent of the interaction at the sensor surface.

of clinical chemical analysis, mainly based on enzyme biosensors for components such as lactate and glucose. Many other enzyme-based sensors are now available, and there has also been a sudden increase in the number of sensors based on optical techniques, which will become important elements of clinical analysis in the near future.

2 TECHNIQUES ASSOCIATED WITH CLINICAL ANALYSIS

2.1 Microdialysis

Microdialysis is a modification of the hemodialysis technology developed for renal patients.[2] In its original form, semipermeable membranes were used to remove the build up of toxic waste products from the blood. In microdialysis, however, the principle is the removal of fractions from body fluids for analysis. Only small volumes (0.1–5 μL) are used, and this reduces the potential harm that may result from the loss of analyte, and the effect the removal of a constituent will have on subsequent measurements. This also reduces the amount of sample available for analysis, and so the analytical methods employed must be of very high sensitivity. In essence, body fluids such as interstitial fluids or blood are allowed to come into contact with a fine dialysis membrane. These permeable fibers are composed of cellulose, polyacrylonitrile and polycarbonate. Within the fiber is a flowing stream of a suitable buffer. Low-molecular-weight analytical components in the sample present at a higher concentration diffuse across the membrane into the dialysis buffer. The concentration gradient is maintained as a result of the flow of the dialysis buffer. Factors important in the mass transport across the membrane are the nature and flow rate of the dialysis buffer, the nature and size of the membrane, and local conditions at the probe site. From here the samples flow to the appropriate analytical system for measurement. Such methodologies lend themselves well to continuous in vivo monitoring, and work excellently when in combination with flow injection analysis (FIA) systems. A problem with the microdialysis technique is the fouling of the dialysis membrane. This is a particular problem when blood sampling is performed because of the biocompatibility problems associated with implanted devices. As an alternative, the sampling of interstitial fluids may be performed to prevent this. When used to estimate blood constituents, care should be taken when calculating values in blood from this source. The lack of widespread use of this type of in vivo monitoring is mostly due to problems with the electrode or biosensing system used for detection, such as stability, repeatability and reproducibility.[3]

The most commonly used type of microdialysis probe is illustrated in Figure 10. The probes can be as small as 0.1 mm in diameter, and consist of a membrane tip, an inlet and an outlet. Dialysis buffer flows down a central inlet tube. Dialysis takes place across the membrane and the analytes are carried away from the membrane by the outer tube. An electrode can be inserted into the outlet flowing stream. This technique is being most widely used in the measurement of neurochemical species in the brain.[4–6]

2.2 Flow Injection Analysis and Chromatographic Separation Systems

Flow injection is a form of instrumentation that allows for greater control in analytical procedures. Fundamentally, it allows the application of reagents to, and removal

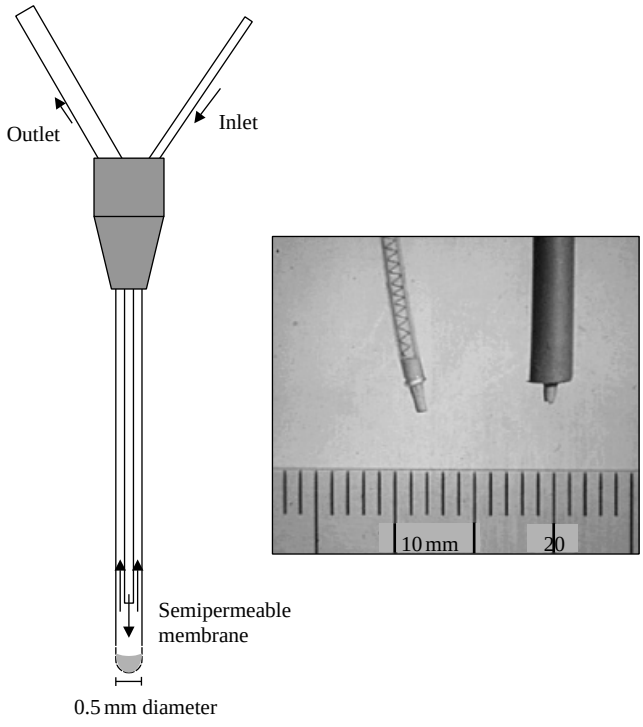

Figure 10 The microdialysis catheter. This extremely narrow (0.5 mm) fiber consists of a central inlet tube entering a tip with a semipermeable membrane. The flow of dialysis buffer carries dialysate from the tip to an electrode or sensor placed in the sample stream.

from, a test surface (an electrode or sensor in the cases being discussed here). The volume of these reagents and the time they spend at the test surface can be precisely controlled. This allows an increased level of automation and control of the assay procedure leading to improved precision, reproducibility and higher sample throughputs. It also allows other features to be introduced into the assay system, such as calibration, regeneration and multianalyte measurements. Flow injection is an integral part of microdialysis systems; it is also a feature of automated chemical analyzers and is being integrated into electrochemical systems more frequently.[7]

Flow injection is also necessary when chromatographic separation methods, such as liquid chromatography, are used as a feature of the analytical procedure. The combination of electrochemical techniques with high-performance liquid chromatography (HPLC) separation is important in the analysis of many biologically important substances, as chromatographic separation provides the required selectivity and the electrochemical methods contribute their excellent sensitivity. Such techniques are referred to as liquid chromatography–electrochemical (LCEC) detection systems. Techniques incorporating capillary electrophoresis are also proving useful due to the very small sample volumes required in this technique.[8]

In flow analysis, samples are successively aspirated or injected into a moving carrier stream. The sample volume becomes part of the continuously moving stream into which, at established time intervals, reagents are added at fixed flow rates. The processed stream then flows through the detector where the quantitative measurement is carried out.

Cell design is an important factor affecting the type of electrochemical response obtained from a detector. Four types of cell design are illustrated in Figure 11. Electrochemical processes rely on transport of an electroactive species toward an electrochemically active surface. If the particular analyte of interest is not effectively carried from the bulk solution to the diffusion layer and across the diffusion layer to the electroactive surface, it cannot be detected.[9]

Signal resolution, sensitivity and selectivity of an electrochemical cell are dependent on three important factors: (a) cell volume, (b) electrode area and (c) electrode material. An increase in cell volume can lead to an increase in dilution and hence poor signal resolution. On the other hand a decrease in cell volume results in a lower dilution due to a higher mean linear velocity of the flow, and this in turn yields a more intense signal. An increase in electrode area will enhance the sensitivity but may reduce the LOD, as the noise is also proportional to the electrode area. Therefore cells should incorporate tubular or disk electrodes with small radii and lengths.

Over the years many cell designs have been investigated.[10–13] Miniaturization of the flow-through cells has been the most important development. The simplest flow-through cell with ISEs was formed by fixing plastic caps to the ends of the ISE and the reference electrode, the caps being connected by a short plastic tube (Figure 11a). Commercial analyzers (Figure 11b) usually employ relatively large plastic cells with internal volumes of a few milliliters.

Coated wire ISEs have been useful for flow measurements, even though they often display rather unstable and poorly defined potentials, because they are readily miniaturized.[14–19] Figure 11(c) depicts a flow-through cell with separate channels for the coated wire ISE and the reference electrode so that the reference electrode is free from contamination. This cell has been successfully used for the determination of calcium ions in serum.[20]

Flow-through cells have also been successfully used as detectors in HPLC. Recent methodologies have described the pulsed amperometric detection of cephalosporins following chromatographic separation on a C_{18} reverse-phase column[21] and the simultaneous determination of L-dopa and carbidopa with ion-pair HPLC using electrochemical detection at a glassy carbon electrode (GCE) at +800 mV vs Ag/AgCl.[22] Figure 11(d) illustrates a detector for a conventional column, with a relatively large internal volume.

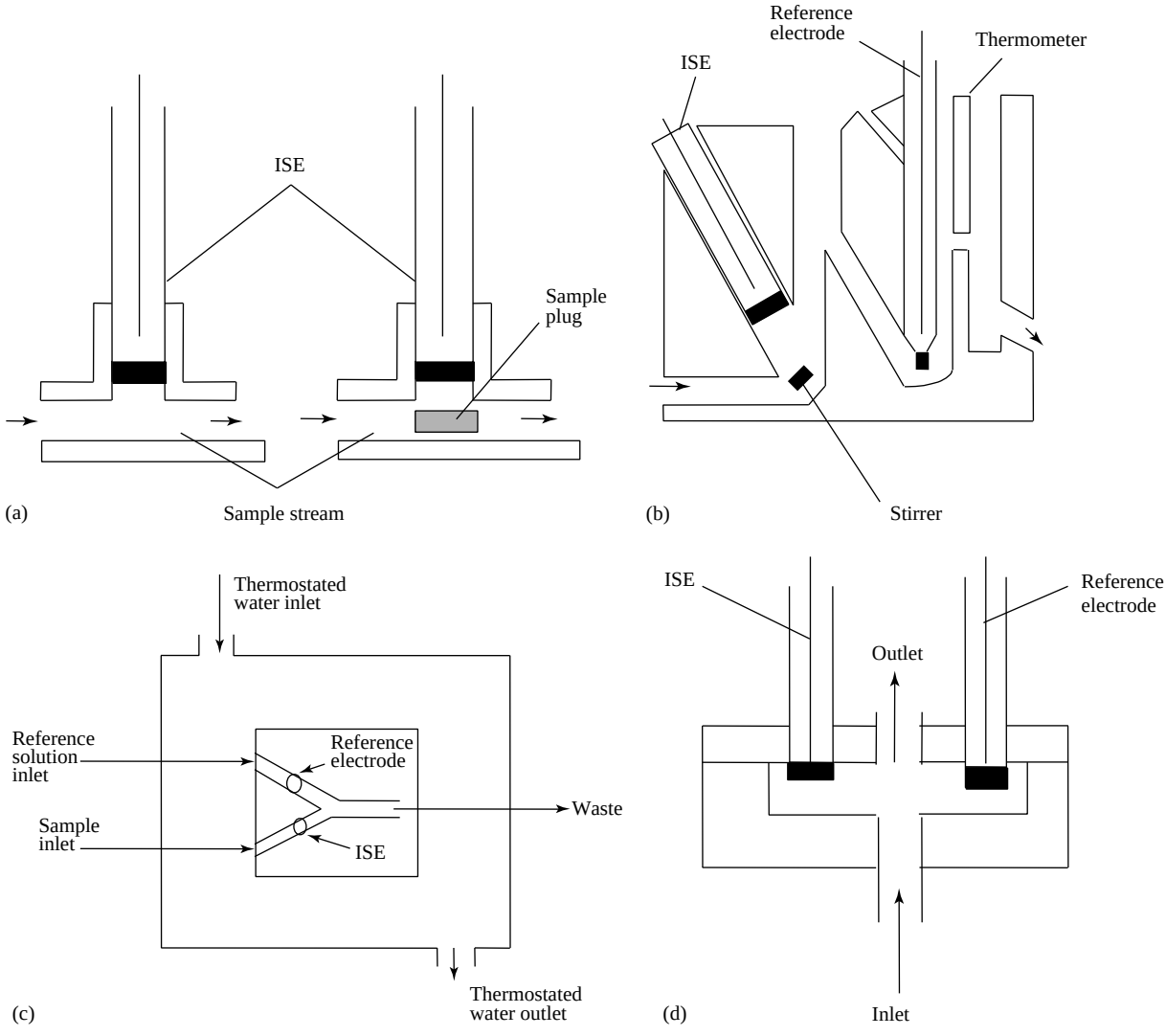

Figure 11 Flow injection cell configurations.

2.3 In Vivo Analysis

At present, most clinical chemical analysis is still performed ex vivo. This requires periodic sampling and off-line analysis by a traditional clinical chemical analyzer. While this is adequate for the majority of analyses in the clinical setting, continuous monitoring would be substantially more useful for the monitoring of critically ill patients or during surgery. To this end, electrodes suitable for in vivo monitoring are being developed and have already been applied in the clinical monitoring of blood gases, electrolytes and glucose.[23] To be suitable for in vivo use, the electrodes must possess all the features of in vitro sensors: namely stability, sensitivity and selectivity, with a rapid response while remaining unaffected by other changes in other variables. However, they must also be biocompatible, sterilizable and of suitable dimensions to be applied to the monitoring site. Many types of in vivo analysis combine the techniques of microdialysis and FIA described under sections 2.1 and 2.2.[4,8] The transdermal delivery of nicotine has been monitored using an implanted catheter connected to a capillary electrophoretic system which in turn is coupled to an electrochemical detector.[24]

2.4 Automated Clinical Chemistry Analyzers

The clinical laboratory environment places great demands on analytical systems. The methods must be standardized, reproducible, and have high throughputs. This level of experimental rigor is best achieved by automation of the analysis. Clinical chemistry analysis is routinely carried out using commercial analyzers. These systems can integrate many analyses into a single system that also uses liquid handling and reagent systems. Such systems can also perform regular self-calibration.[25]

Clinical chemical analyzers have gone through several evolutions from the manual instruments of 30 years past. Originally, machines performed a single analysis, and this normally required a large amount of operator intervention. Following this, devices that could automate the addition of samples and reagents, called sequential batch systems, were introduced. This reduced the level of operator involvement during the analysis. Fixed profile systems followed that could automate a number of analyses in the same sequential batch manner. Then came centrifugal systems which could time the addition of many reagents. In these systems, batches of similar tests still had to be performed together. Faster throughput batch analyzers were also available, and these systems were extremely useful in instances where the result of many identical tests was the requirement. However, the major requirement of the clinical laboratory now is to obtain a range of results for different tests pertaining to a single individual. This was achieved with the advent of stat systems, which allowed a panel of analyses to be performed on a single sample. Today's clinical laboratories are equipped with the most up-to-date random access analyzers, which can perform a range of selected tests on many samples. Most instruments achieve this by utilizing a single analytical pathway for all tests. Other systems achieve optimum efficiency by performing batches of the same experiment on many samples if this is desired. Either way, it means that detailed, comprehensive data as to the condition of a patient are provided to health care workers in a short period of time. The Beckman Synchron CX®3 Delta flow cell assembly is illustrated in Figure 12. Other systems have been adapted to bring the analytical systems closer to the patient, such as smaller bedside analyzers for use in operating theaters and trauma centers. Table 1 shows some currently available clinical chemical analyzers that utilize electrochemical and biosensor methodologies.

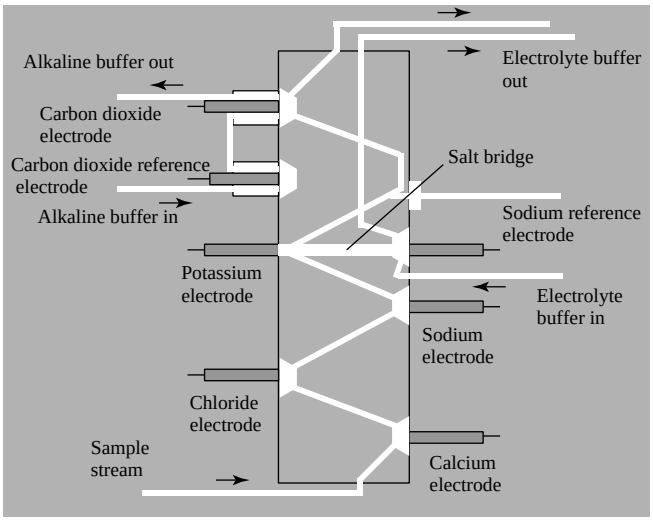

Figure 12 Flow cell assembly of the Beckman Synchron CX®3 Delta. A single sample stream carries analyte across the surface of all electrodes. Calibration solutions are also carried to the CO_2 and ISE electrodes. (Reproduced by permission of Beckman Coulter Inc.)

Although these large automated systems remain extremely important in the hospital environment, they are gradually being replaced by 'point-of-care' systems which allow many parameters to be monitored at the patient's bedside. These systems are portable, hand-held devices which incorporate miniaturized PVC ISEs and biosensor arrays (Figure 3). They can routinely perform most of the same analyses as bigger devices using one-use, disposable biosensor strips or cartridges containing microelectrodes. One such device is the i-STAT system (i-STAT Corporation, recently purchased by Abbott), which is capable of potentiometric and amperometric analyses of pH, P_{CO_2} (partial pressure of carbon dioxide), P_{O_2} (partial pressure of oxygen), glucose, creatinine, sodium, potassium, chloride, bicarbonate, calcium and urea nitrogen. Similar devices are also manufactured by Diametrics Medical Inc., Biologix International Inc. and Sendx Medical Inc. Other major developments include home-monitoring devices, the direct and continuous monitoring of patients through the application of in vivo monitoring using implanted biosensors, and also the growing development of "noninvasive" systems. Most of these developments have occurred in association with glucose monitoring and will be discussed in section 4.3.1.

3 NATURE OF BIOLOGICAL MATRIXES AND SAMPLE PREPARATION

3.1 Blood

For clinical chemical analysis, blood is quite obviously the most important source of information on the condition of a patient. Principally, it carries oxygen and nutrients to all the components of the body, and removes waste products such as carbon dioxide and urea to the lungs and kidneys respectively. It is also involved in other areas of homeostasis, such as the regulation of pH. Blood is pumped by the heart to the lungs via the pulmonary artery where it becomes oxygenated. Oxygenated blood passes throughout the body through the arterial network where it eventually reaches individual cells via capillaries. Here it becomes mixed with interstitial fluid and products of metabolism. Blood is returned to the heart via the venous network. Blood samples from these three sources (arterial, capillary and venous) are of distinct compositions, and the choice of source from which to derive blood

Table 1 Clinical chemical and biochemical analyzers employing electrochemical measurement

Manufacturer/Series	Model	Description
Abbott Laboratories		
Cell-Dyn®	3000, 3500, 4000	Automated blood gas analyzers
	1700	Small/medium-sized hospitals
Spectrum™		Low volume bloods plus glucose, cholesterol and triglycerides
CCx™		Low volume bloods plus glucose, cholesterol and triglycerides
Aeroset™		Glucose, Mg^{2+}, Ca^{2+}, Cl^-, creatinine, Na^+, cholesterol, K^+, CO_2
Alcyon™		
Bayer Diagnostics		
DAX®	24, 48, 72, 96	Na^+, K^+, Cl^-, disposable cartridges
OpeRA		Na^+, K^+, Cl^-
Axon®		Na^+, K^+, Cl^-
Chem1®		Na^+, K^+, Cl^-
RA-50™		Na^+, K^+, Cl^-
Technicon	RA-100	Na^+, K^+, Cl^-
	RA-XT	Na^+, K^+, Cl^-
Beckman-Coulter		
Synchron	CX®3 Delta CX®4 Delta CX®5 Delta CX®9 ALX LX™	Glucose, BUN, creatinine, Na^+, K^+, Cl^-, CO_2, Ca^{2+}
IL Series		pH, P_{O_2}, P_{CO_2}, Na^+, K^+, Cl^-, Ca^{2+}
ILyte		Na^+, K^+, Ca^{2+}, pH
Boehringer Mannheim (Roche, Hitachi)		
COBAS® Integra		Random, continuous access analyzer: Na^+, K^+, Cl^-, Li^+
BM/Hitachi	902, 912, 917, 747	Na^+, K^+, Cl^-
Terumo Medical Corporation		
CDI™ Blood Parameter Extracorpoeal Monitoring System	500	Continually monitored pH, P_{CO_2}, P_{O_2}, K^+, updated every 6 s
	400	As above: pH, P_{CO_2}, P_{O_2}
Radiometer		
ABL	700, 70	
YSI Inc.		
YSI 2300™ STAT Plus YSI 2700™		Glucose, lactate, glutamine, glutamate, ethanol, lactose, sucrose, galactose, hydrogen peroxide

BUN, blood urea nitrogen.

samples is greatly influenced by the type of analysis being performed. For the accurate measurement of blood gases and pH arterial whole blood is required, whereas for the analysis of blood electrolytes, venous or capillary blood may suffice as long as the site of sampling is noted and recognized in later interpretation of clinical data.

The type of analysis to be performed will also influence the blood sample collection route and the way in which the sample is subsequently handled and processed. Although many clinical tests may be performed using blood samples, only those analytes that involve clinically applied electrochemical or biosensor systems will be considered here. Standard clinical precautions, which are also beyond the scope of this work, should always be taken when obtaining blood samples.

3.1.1 Arterial Whole Blood

Arterial whole blood sampling is required for the measurement of pH, P_{CO_2}, and P_{O_2}, as blood from the arteries is considered uniform around the body. Any changes in these measured parameters thus reflect the metabolic condition of that individual.

Sampling of arterial whole blood should only be performed by those who have been instructed in the proper techniques. The blood is usually obtained from radial, brachial, femoral or temporal arteries, and will normally be standardized by the hospital, or depend on the condition of the patient.

Blood gas measurements are also significantly affected by errors brought about by sample handling.[26] Many

precautions are required when taking samples for blood gas analysis, the principle one being avoiding the exchange of gases between the sample and the external environment. Upon sampling, thorough mixing with anticoagulants and rapid transportation on ice are essential for accurate estimations.

3.1.2 Venous Blood, Serum and Plasma

Except for blood gas analysis, venous blood is the normal source of specimens for electrochemical analyses. Venous blood is obtained by venipuncture, normally at the antecubital fossa, from either the cephalic vein, basilic vein, or median cubital. Treatment of the specimen following venipuncture is dependent on the analysis to be performed. Whole blood is rarely required for such analyses, and clotting is allowed to occur removing clotting factors, resulting in serum. For stat chemistries, thrombin is added to allow clot formation in 5 min. For plasma chemistries, heparin may be added. Care should be taken as to which salt of heparin is used, e.g. sodium heparin will affect subsequent sodium measurements. In other analyses, certain other additives or precautions may be required. In glucose or lactate analysis and alcohol measurements, sodium fluoride or iodacetate is added to inhibit glycolysis and clotting is prevented using potassium oxalate. For trace metal analysis, sample tubes must be chemically cleaned and the use of low-trace-metal rubber caps employed. Potassium determinations also require a gel separator and a clot inhibitor to prevent cellular contamination of the resulting plasma.

3.1.3 Capillary Blood

Capillary blood is a mixture of venous and arterial blood and interstitial fluids. It is obtained as a result of skin puncture, normally using the tip of the finger in adults and children, and the side of the heel of the foot in babies less than 15 months old. In this way, blood is brought to the surface of the skin and dripped into a sample tube. The type of analysis that can be performed with blood taken from this source is limited. However, it is normally the blood specimen of choice for babies and children and for diabetics where regular sampling is often necessary.

3.2 Urine

Urine is the preferred sample for the electrochemical analysis of several components relating to renal function, and can also be used in the measurement of many analytes such as pH, carbohydrates, electrolytes, etc., that can all be indicative of a pathologic state.[27]

The volume and concentration of urine samples is not critical. For most electrochemical procedures, refrigeration is the only key sampling requirement. However, for chloride determinations the addition of acetic acid is required when timed urine collections are made.

Urine specimens may be taken at random or be timed collections, such as at 12- or 24-h periods to show diurnal variations or the effects of food consumption, etc. This is certainly a requirement for certain glucose determinations.

Refrigerated urine should be analyzed within 4 h to prevent degradation of the sample. Upon storage, the pH will increase due to bacterial decomposition of urea to ammonia. Normal urine should have a pH between 4.5 and 8.0. Glucose concentration will also decrease as a result of glycolytic processes, either bacterial or cellular.

3.3 Other Body Fluids

Many other body fluids are analyzed in clinical chemistry. However, in electrochemical methods glucose measurements are by far the most important in all these other matrixes.

3.3.1 Cerebrospinal Fluid

Cerebrospinal fluid (CSF) is a plasma ultrafiltrate that surrounds the brain and spinal cord. Samples are obtained by lumbar puncture. For electrochemical analysis glucose is the most important constituent, its decrease being indicative of microbial infection, inflammation, or the presence of tumors. For such measurements, a venous blood sample taken at the same time (4 h post-prandial) should be used to compare levels between blood and CSF. pH, CO_2 and bicarbonate measurements may also be made, in which case the sample should be processed as if it were an arterial whole blood specimen for blood gas analysis. An arterial blood gas sample should also be taken at the same time.

3.3.2 Synovial Fluid

Synovial fluid is essentially an ultrafiltrate of plasma with the addition of hyaluronic acid, and is a viscous fluid that lubricates the joints and supplies nutrients. It is present in very small volumes, and clinical analysis of its glucose content can be used to confirm joint infection. Taken in conjunction with serum glucose, it should contain a glucose concentration approximately 100 mg L^{-1} lower than that of blood. Any significant decrease is probably due to infection of the joint.

3.3.3 Serous or Paracentesis Fluid

Serous fluids occupy the internal cavities of the body, namely the peritoneal, pleural and pericardial cavities. These fluids also have the same characteristics as serum, in that they lubricate and cushion body organs. Volumes

of these fluids are small and effusions (increases in serous fluid) may be indicative of a diseased state. Sampling is performed by paracentesis. Electrochemically, the analytes of importance are glucose and lactate dehydrogenase (LDH). Decreases in the former are typically indicative of infection, inflammation or malignancy. Raised LDH levels are also indicators for these conditions, but in addition are symptomatic of congestive heart failure, liver disease and hypertension.

3.3.4 Amniotic Fluid

This fluid, which is removed by the process of amniocentesis, bathes the growing fetus. Electrochemically, the only test of significance involving amniotic fluid is that for creatinine levels, which acts as an indicator of fetal maturity.

3.3.5 Seminal Fluid

The composition of seminal fluid affects the production of healthy sperm. The sugar, fructose, and pH levels of the fluid are particularly important in this regard. Variations in pH below 7.2–7.8 may indicate abnormalities of the seminal vesicles, epididymis, or vas deferens. A decrease in fructose levels below $1.5–6.0\,g\,L^{-1}$ may indicate azoospermia (no sperm production) or obstruction of the ejaculatory duct.

3.3.6 Saliva

Human saliva contains many constituents, and although not used in present-day clinical electrochemical analysis, it is included for completeness as it has the potential to serve as a noninvasive source of analytical material when suitably sensitive analytical systems are developed in the future.

3.3.7 Sweat

Sweat secretions are also extremely complex, and the composition of sweat is indeed related to many diseased states. Sweat sampling would also be noninvasive and could be constantly monitored by suitable skin-mounted sensors. Most recently, the diagnosis of cystic fibrosis by monitoring electrolyte levels in sweat has been investigated.[28]

3.4 Solid Tissues

The vast majority of clinical analysis is performed on liquid biological matrixes, as is clear from the previous sections. However, analyses can also be performed on solid and semi-solid tissues of the body such as hair, nails, teeth, feces, bones and body organs. This is particularly relevant for the analysis of inorganic ions. For accurate measurements organic material must be removed by oxidation. This is most readily performed by thermal methods such as ashing, or by chemical methods such as wet digestion. Both processes yield a carbon-free inorganic product suitable for this type of analysis. It does, of course, lead to the destruction of the tissue, and this may preclude its use in many tests.[29]

The different biological matrixes used in clinical electroanalysis and their sample handling requirements are summarized in Table 2.

4 CLINICAL ANALYSIS

The sections that follow discuss, in detail, the principles and methodologies of the major electroanalytical techniques used in the clinical setting. The information is summarized in Table 3.

4.1 Blood Gas Analysis

As well as performing many other vital functions, blood controls the exchange of gases in the body. It removes CO_2, and carries O_2 to the various tissues. It performs these tasks in combination with another function of the blood: acid–base homeostasis.[30] The principle buffering system employed in the blood is the carbonate–carbonic acid system. In combination with the blood, the lungs, the kidneys, and the respiratory center, this system regulates the concentration of H^+ and CO_2 in the body. In this regard the red blood cells are extremely important, as they bind O_2 and carry it throughout the body, and when it has been released they carry H^+ back to the lungs.[31]

Metabolic activities in the body result in the production of hydrogen ions. Extracellular body fluids typically contain between 36 and $44\,nmol\,L^{-1}$ H^+. An alteration in the concentration of H^+ leads to alterations in the rates of chemical reactions within cells and also affects other metabolic processes throughout the body. The pH of arterial blood is typically 7.4 and this corresponds to a H^+ concentration of $40\,nmol\,L^{-1}$. A pH level below this is termed ''acidosis'' and above this is termed ''alkalosis''.

Carbon dioxide is produced as a result of metabolic processes in the tissues and diffuses to the blood plasma where it is picked up mostly by red blood cells that have released oxygen to the tissues. The CO_2 becomes carbonic acid, which then dissociates to give H^+ and HCO_3^-. This increase in bicarbonate levels causes the bicarbonate to diffuse into the plasma and chloride ions then diffuse into the red blood cells. Having released O_2, hemoglobin (Hb) picks up a proton to form deoxyhemoglobin (HHb) and carries it to the lungs where it forms carbonic acid. This

Table 2 Sampling requirements for electroanalytical methods

Source	Sampling	Sample handling	Measurements
Arterial whole blood	Radial, brachial, femoral or temporal arteries	For blood gas analysis, storage on ice immediately after sampling	pH, P_{CO_2} and P_{O_2}
Venous blood, serum, plasma	Venipuncture at the antecubital fossa	Removal of clotting factors for serum production Addition of thrombin for stat chemistries Addition of heparin/citrate for plasma production Addition of iodacetate/sodium fluoride for glycolysis inhibition and addition of potassium oxalate for clotting Low-trace-metal caps and chemically-cleaned tubes for trace metal analysis For analysis of electrolytes, salts composed of cations being analyzed should not be used for storage, e.g. sodium heparin in the case of Na^+ analysis	Analyses other than blood gas
Capillary blood	Skin puncture	(see venous blood, serum, plasma)	Analyses other than blood gas
Urine	Random or timed sampling over 12- or 24-h period	Refrigeration Addition of acetic acid for chloride analysis of timed collections	pH, glucose and other carbohydrates, electrolytes
CSF	Lumbar puncture	For gas analysis, process as for arterial whole blood	pH, CO_2, glucose
Synovial fluid		Inhibition of glycolysis	Glucose
Serous fluid	Paracentesis	Inhibition of glycolysis	Glucose
Amniotic fluid	Amniocentesis		Creatinine
Seminal fluid	Ejaculation		pH, fructose
Saliva	Collection of saliva		Glucose and others
Sweat	Sweat collection. Sweating can be accelerated by applying dimethylsulfoxide to the skin surface		Electrolytes

quickly dissociates to CO_2 and H_2O, which are eliminated in exhaled air.

The kidneys, which aid in its control, also affect the H^+ concentration by the selective excretion or reabsorption of H^+, Na^+, Cl^-, PO_4^-, K^+, NH_4^+ and HCO_3^- to restore equilibrium.

The pH and gas content of other body fluids are also often of clinical relevance. Measurement techniques for blood gas analysis are appropriate for other body fluids when appropriate sample handling measures have been taken (see section 3).

4.1.1 Determination of pH

Measurement of pH is based on a potentiometric electrochemical measurement.[32] The indicator electrode is normally a glass electrode. This is composed of a silver wire with a coating of silver chloride, immersed in dilute HCl and contained in a glass chamber. The surface of the glass is made selective for hydrogen ions by manipulating the oxide content of the glass. The normal composition for this application is typically 72.2% SiO_2, 6.4% CaO and 21.4% Na_2O. Compositions of lithium oxide and barium and lanthanum silicates or oxides are also found, and these are less prone to interference from Na^+ ions at elevated pH. The dimensions of the electrode are made to suit the application of small volumes of sample. The electrode surface is flat and is normally part of a flow injection system for ease of automation. The reference electrode is a silver/silver chloride electrode immersed in 3 M chloride, normally KCl or NaCl. When the electrode is exposed to a solution containing hydrogen ions, these hydrogen ions combine with oxide sites in the glass membrane, exchanging other cations. This results in a difference in hydrogen ion hydration on the outer surface of the membrane compared with the inner surface and this concentration difference generates a potential, which at 37 °C is proportional to the H^+ concentration according to the Nernst equation (Equation 7)

$$E = E^0 + 0.0613\,\mathrm{pH} \tag{7}$$

where E is the potential change, and E^0 is the standard cell potential for hydrogen. This results in a slope of 61.3 mV per pH unit, allowing H^+ measurement over several orders of magnitude.

Table 3 Principle electroanalytical methods in clinical chemistry

Analyte	Electrode type	Electrochemical technique	Matrixes	Clinical range	Analytical range	Interferences	Other methods
pH	Glass electrode	Potentiometry	Arterial blood	36–44 nmol L^{-1} H^+ (pH 7.4)	0.5–10	Na^+ 0.5 > pH > 10	
P_{O_2}	Oxygen glass electrode	Amperometry	Arterial blood Venous blood	11.97 kPa 5.32 kPa		Ascorbic acid	Oximeter (spectrophotometric)
P_{CO_2}	Glass electrode	Potentiometry	Arterial blood Venous blood	5.32 kPa 6.12 kPa			
Sodium (Na^+)	ISE glass	Potentiometry	Serum Urine (24 h) CSF	136–146 mmol L^{-1} 40–220 mmol L^{-1} 138–150 mmol L^{-1}	10^{-1}–10^{-6} mol L^{-1}	H^+	FES
Potassium (K^+)	ISE glass	Potentiometry	Serum Urine (24 h)	3.5–5.1 mmol L^{-1} 25–125 mmol L^{-1}	10^{-1}–10^{-6} mol L^{-1}	Na^+	FES
Calcium (Ca^{2+})	ISE liquid	Potentiometry	Serum	46–53 mg L^{-1}	$>5 \times 10^{-7}$ mol L^{-1}		AAS/turbidimetry
Lithium (Li^{2+})	ISE	Potentiometry				Na^+	AAS
Magnesium (Mg^{2+})	ISE liquid	Potentiometry	Serum Urine (24 h)	18–30 mg L^{-1} 73–104 mg L^{-1}		Ca^{2+}, Na^+	AAS/FES
Lactate	Enzyme	Amperometry	Arterial blood Venous blood	30–70 mg L^{-1} 50–200 mg L^{-1}		Ascorbic acid	
Chloride			Serum Urine (24 h)	98–106 mmol L^{-1} 110–250 mmol L^{-1}	$1–1 \times 10^{-5}$ mol L^{-1}	Halides, cyanides and sulfides	Spectrophotometric
Glucose	Enzyme	Amperometry	Serum Serum (postprandial)	4.5–5.5 mmol L^{-1} 7.2–8.8 mmol L^{-1}	10^{-2}–10^{-7} mol L^{-1}	Ascorbic acid	Colorimetric
Galactose	Enzyme	Amperometry					
Fructose	Enzyme	Amperometry					
Sucrose	Enzyme	Amperometry					
Starch	Enzyme	Amperometry					
Urea	ISE Glass electrode Voltaic cell	Potentiometry Potentiometry Conductimetry	Serum/plasma	171–556 mg L^{-1}	$1–10^{-6}$ mol L^{-1}	Na^+, K^+ Electrolytes	Colorimetric/Spectrophotometric
Uric acid	Enzyme CO_2 electrode	Amperometry	Serum/plasma	0.15–0.42 mmol L^{-1}	2.5×10^{-3}–1×10^{-4} mmol L^{-1}	Ascorbic acid	Spectrophotometric
Creatine/creatinine	Enzyme	Amperometry	Serum/plasma	44–106 mmol L^{-1} (creatinine)	7.6–760 µmol L^{-1}		Colorimetric
Ammonia	Ammonia gas electrode	Potentiometry	Serum/plasma	11–35 µmol L^{-1}	$\geq 10^{-6}$ mol L^{-1}		Spectrophotometric
Cholesterol	Enzyme	Amperometry Adsorptive stripping voltammetry	Serum	1.44–2.75 g L^{-1}	0.2–0.8 µmol L^{-1}	Ascorbic acid	Spectrophotometric
Vitamins		Voltammetry					
Trace metals		Voltammetry					
Amino acids	Enzyme Ammonia gas electrode	Amperometry Potentiometry					
Blood group antigens		Amperometry					

FES, flame emission spectrometry; AAS, atomic absorption spectrometry.

The glass electrode is the standard method for the determination of pH. It has been mostly replaced with PVC-based membranes in automated systems, new point-of-care systems and for in vivo measurements. Problems occur when extremes of pH are encountered. Inaccurate results are obtained below pH 0.5, and above pH 10 the electrode is affected by Na^+, but as has been mentioned this can be minimized if non-sodium-based oxides are used. However, such extremes are not encountered in the clinical setting. pH electrodes also require regular calibration. Two buffers of known pH are required to correct for slope.[33]

4.1.2 Measurement of Carbon Dioxide

The CO_2 electrode is based on the glass electrode, or PVC H^+ electrode.[34] In addition, a gas permeable membrane of polypropylene or Teflon® is placed over the selective glass surface. Between the glass and polypropylene is placed a bicarbonate buffer solution. The electrode design is in a similar configuration to the pH electrode, being flat and exposed to a stream of sample in a flow cell device. Samples are kept anaerobic to prevent changes in the CO_2 content. Gases, including CO_2, diffuse across the membrane where carbonic acid (H_2CO_3) is formed, which dissociates into a hydrogen ion and bicarbonate (HCO_3^-), as depicted in Equation (8)

$$CO_{2(aq)} + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+ \quad (8)$$

The concentration of hydrogen ions is then measured potentiometrically as is the case with the pH electrode. The change in pH is proportional to the rate of diffusion of CO_2 into the membrane, which can be related back to PCO_2 using Henry's law. This states that if the concentration of the internal carbonate solution is sufficiently high, then the potential of the electrode is dependent on the boundary potentials of the electrode and the concentration of carbon dioxide in the external solution.

4.1.3 Calculation of Carbonate Concentration

For an acid–base equilibrium, the pH is defined by the Henderson–Hasselbach equation (Equation 9)

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \quad (9)$$

where pK_a is the pH at which there are equal concentrations of both proton donor [HA] and proton acceptor [A^-], respectively. The pK_a of blood plasma at 37 °C is 6.1.[30] The ratio of carbonic acid to carbon dioxide in the blood is normally 1 : 800. Therefore, the concentration of carbonic acid can be calculated from the measurement of PCO_2. To calculate carbonic acid concentration, the conversion factor used is $PCO_2 \times 0.031\ mmol\ L^{-1}$.[31] Knowing the two parameters pH and PCO_2, carbonate concentration can be calculated as shown in Equation (10).

$$pH = pK + \log \frac{[HCO_3^-]}{0.031 PCO_2} \quad (10)$$

In normal healthy individuals the ratio of carbonate to carbonic acid is 20 : 1, yielding a pH of 7.4.

ISEs for the direct determination of bicarbonate are not yet commonly used, but are becoming more widespread.

4.1.4 Calculation of Base Excess/Base Deficit

The base excess/base deficit is a theoretical unit of acid or base required to return blood plasma to a pH of 7.40 at a PCO_2 of 5.32 kPa and 37 °C. Three parameters are required to calculate this value: PCO_2, pH and the Hb count. The Hb count is normally measured by blood gas analyzers, but the measurement is not performed electrochemically. A positive value indicates a base (HCO_3^-) excess, or a deficit of noncarbonic acid and is indicative of alkalosis. A negative value is indicative of a base deficit or an excess of noncarbonic acid and is indicative of acidosis. The figure allows the estimation of the amount of bicarbonate or ammonium chloride required to return a patient's blood to normal pH.

4.1.5 Measurement of Oxygen

Oxygen is carried in the blood by Hb. In normal, healthy individuals, 95% of Hb will be bound to oxygen. An increase in oxygen concentration will saturate Hb and eventually cause the accumulation of dissolved oxygen in the blood. Several parameters are associated with blood oxygen measurement.[35] sO_2 is the percentage oxygen saturation of Hb and is a ratio of the total amount of Hb to the amount of HHb. This parameter can be measured directly, but can also be calculated from electrochemical blood gas analyzers using another quantity, PO_2, and the total Hb count.

Blood PO_2 is measured amperometrically using an oxygen electrode. This device uses a semipermeable membrane selective for oxygen enclosing a supporting electrolyte of phosphate within which is suspended a platinum cathodic working electrode and a reference electrode. The gas-permeable membrane is composed of polypropylene or Teflon®. The electrode configuration allows the flow of a small volume of blood or other sample material across the electrode surface. Oxygen diffuses to the surface of the electrode and is reduced at the electrode surface by the application of a potential of −0.65 V. Though other gases can also diffuse across the membrane, only O_2 is reduced at this potential.

The reduction at the cathode results in the flow of electrons and a measurable current that is quantified by the application of an ammeter. Four electrons are required to reduce a molecule of oxygen, as shown in Equation (11).

$$O_2 + 2H_2O + 4e^- \longrightarrow 4OH^- \qquad (11)$$

Diffusion of oxygen to the sensor surface is rate limiting and so the level of current produced is proportional to the oxygen concentration and is an indicator of PO_2.

Oxygen measurements can be made in infants using transcutaneous electrodes. Here, skin thickness is an obstacle to gas diffusion and warming the skin surface to dilate capillaries can improve signals.[36]

Errors in blood gas measurements can occur as a result of changes in the sensor geometry. It is essential that the surface of the membrane be kept free of surface defects including folds and bubbles.[37] The rate of diffusion of gases to the sensor is also dependent on the characteristics of the membranes and the electrolytes. Contamination of test equipment with bacteria, for example, or the accumulation of debris, can cause changes in the diffusion rates of oxygen and other gases, or have the potential to consume oxygen. Further errors can also be encountered with the infiltration of room-temperature air mixing with calibration gases. The temperature of the patient should be noted and all readings should be corrected for the partial pressures of all gases at that temperature.

Other nonelectrochemical oxygen measurements are performed using an oximeter, which measures oxygen saturation of Hb using a spectrophotometer.[38] Such data are very useful, but can also be calculated from amperometric blood gas measurements in combination with total Hb, which is often an integral component measurement of blood gas instruments.

4.2 Electrolytes

The body possesses many charged ionic species, which take part in important metabolic processes. The concentrations of these species are extremely important for correct functioning of many body processes. The body thus monitors and regulates the levels of these species. Changes in the metabolic condition of the body brought about by disease or injury may bring about changes in the concentrations of these species. Thus, monitoring of electrolytes is useful clinically.

Most electrolytes are measured using ISEs.[39] These devices utilize the electrochemical technique of potentiometry, in which the movement of ions into and out of a perm-selective membrane is monitored at zero current by the resulting potential that the movement of these ions induces (as discussed in section 1.1). Electrolyte analysis is now a standard function of most clinical chemistry analyzers. Analyzers such as the Beckman Synchron CX®3 Delta have a series of ISEs for the measurement of the electrolytes sodium, potassium and calcium. The ISEs also measure the most significant anion, chloride.

4.2.1 Sodium (Na^+)

The most common cation in the body is sodium (Na^+) which accounts for approximately 90% of extracellular cations with a normal serum concentration of 136–146 mmol L^{-1}.[40] The traditional sodium glass electrode was made selective by adjusting the composition of the electrode glass. Two formulations were widely used: 11% Na_2O–18% Al_2O_3–71% SiO_2 or 11% Li_2O–18% Al_2O_3–71% SiO_2. The composition of the glass is very highly selective for Na^+. This, and the relatively low concentrations of other alkali metal ions in blood, plasma or serum, means that measurements are unaffected by other cations.[41] Sodium ISEs show Nernstian behavior across a very wide range of ion concentration (10^{-1}–10^{-6} mol L^{-1}). However, sodium measurements are strongly affected by the concentration of H^+, but this normally remains within narrow physiological ranges.

Sodium electrodes are now typically composed of PVC membranes impregnated with perm-selective agents, such as calix[4]arene esters and amides, crown ethers and cyclodextrins.[42]

Discrepancies still exist between measurements of electrolytes derived from flame photometry and those obtained from ISEs, and calculations are also affected depending on whether values are derived from diluted or undiluted samples. Electrochemical methods measure the molality of the sample[43] (the activity of the ion in a volume of water), whereas flame photometers take measurements relating to the solution volume, which contains significant levels of proteins, lipids and other macromolecules.[44–46] Typically, however, the discrepancy is of the order of 2–4%. Complexation of Na^+ to bicarbonate may also cause decreases in values when undiluted samples are used for analysis.

4.2.2 Potassium (K^+)

Potassium is the major intracellular cation, but is not a major component of extracellular fluids. It is essential for regulating neuromuscular activity throughout the body, including the heart.[47] Potassium ISEs are prepared using a valinomycin-based ion-selective PVC membrane.[48] Valinomycin is a cyclic dodecadepsipeptide, the twelve residues alternating between ester and peptide linkages. The pocket formed at the center is electron rich, and due to its radius of 1.33 Å the goodness of fit for K^+ ions makes it selective for K^+ over Na^+ at a ratio of

5000 : 1 and for K^+ over H^+ at a ratio of 18 000 : 1. The response of the K^+ electrode is also Nernstian across a very broad concentration range (10^{-1}–10^{-6} mol L^{-1}), and is used for measurements in blood and serum. Adult serum levels of K^+ range between 3.5 and 5.1 mmol L^{-1}. In urine, 25–125 mmol of K^+ are excreted in a 24-h period. For direct measurements of urine, silicone rubber membranes are more appropriate as the PVC membrane becomes poisoned by anions present in the sample.[49]

Analytical discrepancies between the traditional method of flame photometry and ISE measurement exist for potassium ion determination also.

4.2.3 Calcium (Ca^{2+})

The calcium ion is also important in the regulation of neuromuscular activity and is essential for the proper contraction of the heart. Adult levels of Ca^{2+} are in the region of 1.20–1.38 mmol L^{-1}, approximately half the total calcium level. Calcium ion determination was previously performed using an ISE incorporating a liquid membrane containing a calcium ion exchanger. This was composed of a mixture of calcium didecyl phosphate and di-*n*-octylphenyl phosphonate held in a porous membrane.[50] Exchange of Ca^{2+} ions at the interface forms an equilibrium with a Nernstian response down to a Ca^{2+} concentration of 5×10^{-7} mol L^{-1}. No significant interference results from either sodium or magnesium ions, to which it is 3000 and 200 times more selective, respectively. The calcium electrode is also unaffected by changes in pH at physiological levels. The accuracy of calcium ion levels obtained by ISE as compared with data from flame photometry is still difficult to validate. At low concentrations, calcium ion activity is also affected by biological macromolecules, both from changing water activity and with the complexation of calcium ions by proteins. This type of calcium ISE had a reduced lifetime due to the loss of ion exchanger from the membrane layer. Although good precision exists, values obtained from different electrodes may vary due to differences in electrode design. Calcium ISE liquid membrane electrodes have also been superseded by PVC membranes based on ionophores such as the ETH series (Fluka Chemie, AG, Switzerland) (ETH 227) and lipophilic hexapeptides that act as perm-selective agents for Ca^{2+}.[51]

4.2.4 Lithium (Li^+)

Lithium ISEs are available, and are gradually overcoming the significant problem of sodium sensitivity. Many ion-selective carriers have been investigated, including crown ethers and polyethers, which are gradually showing greater promise, particularly derivatives of 14-crown-4. Alone this has a selectivity ratio of Li^+ to Na^+ of 1000 : 1.[52] Derivatives such as tetradecyldecalino-14-crown-4 and 2,3-([(*R*)-(1α, 2α, 3α, 5α)]-2,6,6-trimethylbicyclo[3.1.1]heptano)-9,9,10, 10-tetramethyl-1,4,8,11-tetraoxacyclotetradecane (PTM 14-crown-4) exhibit greatly improved selectivities of 4000 : 1 and 10 000 : 1, respectively.[53]

4.2.5 Chloride (Cl^-)

Chloride is the major anion in the human body and is important for the maintenance of electrical neutrality, both in the kidneys, and in red blood cells using the chloride shift.[40,43] Chloride measurement is also of concern in cystic fibrosis, where it can be used as a diagnostic aid and as a gold standard in sweat tests. Its serum concentration is between 98 and 106 mmol L^{-1}, while in urine the variation is from 110 to 250 mmol L^{-1} over 24 h.[40]

Chloride can be measured using two electrochemical techniques: controlled current coulometry, and ISE potentiometry.[48] In the chloridometer, titration of Cl^- is based on the coulometric generation of Ag^+ ions from a silver electrode, which then combine with and titrate Cl^-, as shown in Equation (12).

$$Ag^+ + Cl^- \longrightarrow AgCl \qquad (12)$$

When all the Cl^- has been complexed, the excess Ag^+ end-point can be monitored amperometrically, and the elapsed time taken to reach the titration point is a measure of the chloride ion concentration (according to Equation 3). Other anions cause interference with this system, including halides, cyanides and sulfides.

The ion-selective approach utilizes a silicone polymer impregnated with silver chloride particles.[54] This forms the selective membrane, with potassium chloride as the supporting electrolyte. PVC membranes doped with perm-selective agents such as porphyrins are also used.[55] Direct chloride concentrations in the 1–1×10^{-5} mol L^{-1} range are possible with these approaches. Halides and sulfides also cause interferences with these systems. Liquid membrane ISEs for chloride have also been produced, but suffer from serious interferences.

Nonelectrochemical methods include mercurimetric titration. This is similar to silver titration, where complexation of chloride is measured by a colored end-point. This method also suffers from interferences of other anionic species. Most commercial systems rely on coulometric techniques.

4.2.6 Magnesium (Mg^{2+})

While magnesium is still preferably measured by AAS, ISEs have been developed using liquid membranes

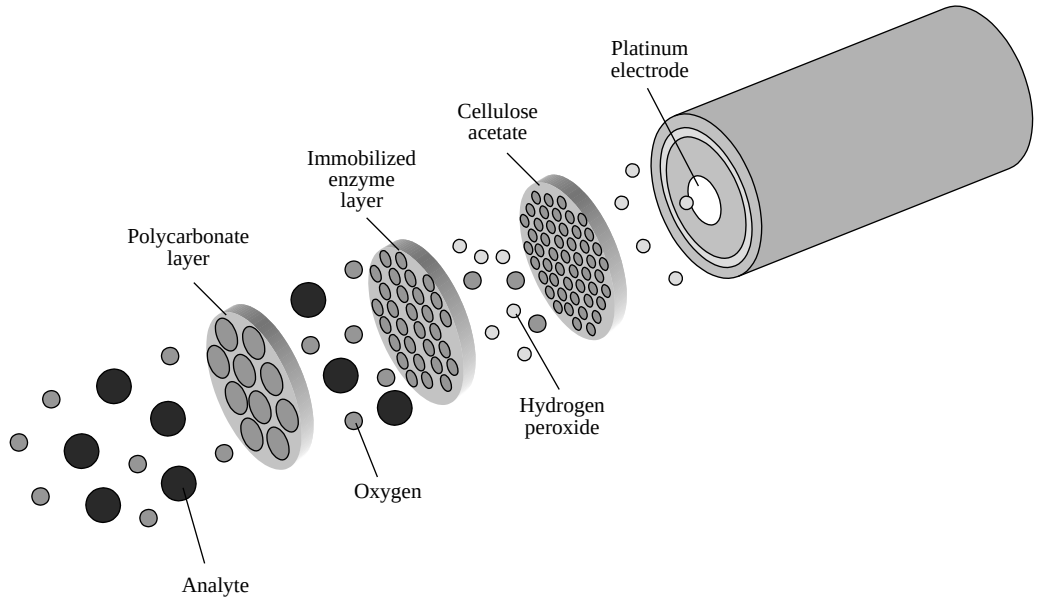

Figure 13 The YSI Inc. biosensor. Initially, a polycarbonate membrane limits diffusion of the analyte to the immobilized enzyme layer, where it undergoes oxidation. Hydrogen peroxide diffuses to the platinum electrode surface via a cellulose acetate membrane which excludes interferences such as those due to ascorbate. (Reproduced by permission of YSI Incorporated.)

incorporating cyclodecapeptides and PVC membranes incorporating ETH 7025 (Fluka Chemie AG, Switzerland).[56,57] However, calcium and sodium ion sensitivities preclude their use as accurate determinants of Mg^{2+} concentration, and simultaneous subtractive calculations would be required for these interferents.

4.2.7 Lactate ($C_3H_6O_3^-$)

Lactate is of major importance in clinical chemistry as it is a major determinant of the survival of patients during critical care.[58] This makes its stat analysis extremely valuable, and explains why it has been a major target for the development of rapid electrochemical techniques for its measurement.

The standard determination of lactate is now based on amperometric enzyme biosensors.[59] These use enzymes to bring about the biochemical conversion of lactate, which results in the consumption or generation of species that can be measured amperometrically. Initial systems were based on the conversion of lactate to pyruvate by LDH in the presence of the oxidized form of nicotinamide adenine dinucleotide (NAD^+), which is subsequently reduced to the reduced form of nicotinamide adenine dinucleotide (NADH) according to Equation (13).

$$\underset{\text{L-lactate}}{CH_3HCOHCOOH} + NAD^+ \overset{\text{LDH}}{\rightleftharpoons} \underset{\text{Pyruvate}}{CH_3COCOOH} + NADH + H^+ \qquad (13)$$

Electrocatalytic conversion of NAD^+ to NADH is monitored amperometrically and is proportional to the concentration of lactate.

YSI Inc. have developed an enzyme electrode for lactate (Figure 13) based on the conversion of lactate to pyruvate by lactate oxidase according to Equation (14).

$$CH_3HCOHCOOH + O_2 \overset{\text{Lactate oxidase}}{\rightleftharpoons} CH_3COCOOH + H_2O_2 \qquad (14)$$

The consumption of O_2 or the production of H_2O_2 can both be monitored for the conversion of lactate to pyruvate. In this system, H_2O_2 is oxidized to O_2 by the platinum anode.

Glucose may be anaerobically metabolized to lactate upon sampling. Inhibition of glycolysis can be achieved by the addition of iodoacetate.

4.3 Carbohydrates

4.3.1 Glucose

Glucose is the primary source of all energy in metabolism, and is released in the process of glycolysis. Its metabolism and regulation is a major feature of human homeostasis. It is controlled by several hormones such as insulin, glucagon, epinephrine, growth hormone, adenocorticotropic hormone, glucocorticoids and thyroid hormones. These all maintain glucose levels throughout the body, which results in blood glucose levels between

4.5 and 5.5 mmol L^{-1}, although this may rise significantly post-prandially to between 7.2 and 8.8 mmol L^{-1}.

The major medical problem associated with glucose homeostasis is diabetes. Although there are two primary forms, they both result in the decreased regulation of glucose availability, which can be either debilitating or fatal if not treated with insulin.

The ability to monitor glucose levels in the blood, urine and other body fluids is crucial for the treatment of diabetes and for the diagnosis of many other disorders outlined in section 3. For this reason, the electrochemical measurement of glucose has become the largest sector in the electrochemical sensor industry worldwide. Many electrochemical procedures, with a myriad of modifications, exist today for the determination of glucose, as researchers and companies attempt to capture a fraction of this extremely lucrative market. For this reason, more examples exist of glucose-sensing systems than of systems for detecting any other analyte. Most glucose electrodes are enzyme-based amperometric devices.[60] The principle system of glucose measurement is based on the conversion of glucose to gluconic acid by glucose oxidase, as shown in Equation (15).

$$\text{D-Glucose} + O_2 \overset{\text{Glucose oxidase}}{\rightleftharpoons} \text{Gluconic acid} + H_2O_2 \quad (15)$$

In these enzyme-based sensors the biocatalyst is immobilized on the surface of the electrode. Many methodologies have been employed to accomplish the stable and reproducible application of enzymes to electrode surfaces. These include covalent attachment, adsorption, entrapment and encapsulation. The consumption of O_2 or the production of H_2O_2 can be monitored amperometrically at a platinum electrode. Oxygen uptake using a Clark-type electrode can be used. However, because of the presence of oxygen in samples the monitoring of hydrogen peroxide production is preferred. Oxidation of H_2O_2 can be monitored at +800 mV vs an SCE. The selectivity of the electrode can be further improved by the incorporation of a suitable membrane such as cellulose acetate to prevent the diffusion of other oxidizable products to the electrode surface. Glucose sensors have an operational range in the order of 10^{-2}–10^{-7} mol L^{-1}, with rapid response times in the order of seconds. They are amenable to the direct quantification of glucose from undiluted whole blood.

The difficulties encountered with interferences at high electrode potentials can be circumvented by the use of mediators other than oxygen, such as redox polymers. These act as direct mediators between the enzyme and electrode surface and reduce the potential necessary for oxidation to inhibit interferences.[61]

Many glucose systems are available for home monitoring as well as clinical use. Companies producing glucose monitoring devices have undergone regular rearrangements due to mergers and acquisitions. A period of consolidation has occurred in recent years, with the large biopharmaceutical companies purchasing smaller, specialized firms that were responsible for the initial developments of these products. Currently, the market is dominated by four companies. Johnson & Johnson, who purchased LifeScan Inc., hold some 45% of market share with a mixture of electrochemical and color reflectance monitoring systems. Boehringer Mannheim holds some 25%, since their acquisition of Roche, and again manufacture a mixture of color reflectance and electrochemical glucose meters. Bayer (integrated with Chiron Corporation, Miles Inc. and Ciba-Corning) and Abbott (who acquired MediSense) each hold between 10 and 15%, with smaller companies making up the rest.

Point-of-care glucose analysis involves the use of disposable electrode strips, such as the ExacTech® biosensor illustrated in Figure 14. This has a rigid PVC support onto which are screen-printed the various components of the electrode, such as the conducting silver and carbon tracks and the reference and working electrodes. The working electrode is a mixture of carbon paste and glucose oxidase enzyme. The strip is inserted into the portable glucose monitor. A drop of blood is placed upon the working/reference electrode spot and the appropriate reading is given by the monitor after the specified time period.

Table 4 lists currently available electrochemical glucose sensors. Presently, one of the most important features of glucose sensor design is the amount of blood required for analysis, as this determines the amount of discomfort that occurs in sampling and also the frequency with which sampling can be performed at particular sites. The most popular electrochemical glucose sensor at present is the FastTake from LifeScan Inc., which requires only 2.5 μL

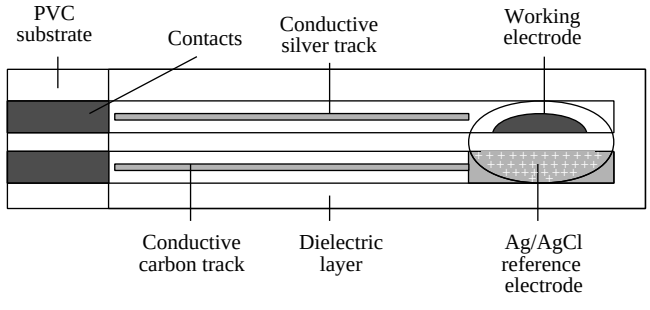

Figure 14 The ExacTech® disposable biosensor electrode strip. Two electrodes are screen-printed onto a PVC support with conductive tracks and protective dielectrics. The working electrode contains glucose oxidase enzyme. (Reproduced by permission of the Royal Society of Chemistry.)

Table 4 Characteristics of electrochemical glucose biosensors

Company/product	Data port	Calibration	Response time (s)	Sample volume (μL)	Other information
Abbott Laboratories (including MediSense)					
Precision Q.I.D.®	Y	Plasma	20	3.5	
Precision Q.I.D. Pen Sensor®	Y	Plasma	20	3.5	
MediSense Card Sensor	N	Whole blood/plasma	20	10	Sold in Europe
MediSense Pen Sensor	N	Whole blood/plasma	20	10	Sold in Europe
ExacTech®	N	Whole blood	30	10–50	
ExacTech® R.S.G.™	N	Whole blood	30	10–50	
Bayer Corp. (including Chiron and Miles Inc.)					
Glucometer Elite XL®	Y	Plasma	30	3	Hospital usage
Glucometer Elite®	N	Plasma	30	3	Hospital usage
Glucometer Esprit®	Y	Whole blood	30	3–4	
Cascade Medical Inc.					
CheckMate Plus	Y	Whole blood/plasma	15–70	3	
Chronimed Inc.					
Assure™	Y	Whole blood/plasma	30	10	
Select GT®	N	Whole blood/plasma	45	9	
Supreme II®	N	Whole blood/plasma	45–50	9	
LifeScan Inc. (Johnson & Johnson)					
FastTake Compact®	Y	Plasma	15	2.5	
Boehringer Mannheim (Roche Group)					
Accu-Chek™ Complete	Y	Plasma	40	4	Fax/modem
Accu-Chek™ Advantage®	Y	Plasma	40	4	Voice synthesizer

of blood and yields results in only 15 s. Other important issues are the time taken for analysis, the presence of a data port to connect to diabetes monitoring systems and whether the device is whole-blood or plasma calibrated. Presently, there is a shift to plasma calibration as this allows closer comparison of off-site data with clinical results that generally use plasma data.

A move to ''less invasive'' glucose monitoring is fuelling the development of a new range of products. Many purport to be noninvasive, but are not so. They do, however, attempt to reduce the level of discomfort experienced when giving blood samples. The GlucoWatch® by Cygnus Inc. utilizes a low-level electrical field to extract glucose transdermally before detection by a biosensor pad worn like a wristwatch. SpectRx Inc. also markets a transdermal system that creates micropores in the outer layers of skin through which interstitial fluids may pass for glucose measurement. Other systems such as the Kumetrix system and the TheraSense system attempt to significantly reduce the amount of blood required for sampling by using a micro-needle that draws a fraction of a microliter of blood. In this way it can be used on parts of the body not as richly supplied with blood vessels or nerve endings. Implantable glucose biosensors are also being developed that use the cannula-type design described in section 2.1, such as those produced by MiniMed and Synthetic Blood International. Many of these devices were at the early stages of FDA (Food and Drug Administration) approval at the time of writing.

4.3.2 Other Carbohydrates

Many other mono-, di- and polysaccahrides can also be measured using enzyme biosensors. Sugars of some clinical relevance are lactose, galactose and fructose, which are mostly associated with various forms of childhood intolerance and, in the case of galactose, as a measure of liver function. Measurements have been made of these sugars using both oxidases and dehydrogenases.[62–64] However, these assays are not performed as standard clinical tests in these forms as yet. These assays also suffer from the same interferences as the glucose electrode, as well as nonspecificity of some of the enzymes, e.g. galactose oxidase.

The disaccharide sucrose can be measured using enzyme electrodes. YSI produce a sensor with a series of three immobilized enzymes: invertase, mutarotase and glucose oxidase. In three steps, these convert sucrose to alpha-D-glucose and fructose, beta-D-glucose and finally gluconic acid, with the release of hydrogen peroxide being monitored amperometrically. Subtraction of glucose values will yield sucrose concentration.

Multi-layer electrode devices have been developed for the determination of polysaccharides, such as starch, while excluding interferences brought about by the presence of glucose.[65] The devices use an anti-interference layer of glucose oxidase to first remove glucose. In the second layer, amylase converts the starch to maltose and dextrins. The maltose is then converted to glucose by glucoamylase, which is in turn converted to gluconic acid by glucose oxidase, the production of hydrogen peroxide being monitored amperometrically as before.

4.4 Nonprotein Nitrogen

The measurement of nitrogenous substances in the blood is important clinically for the monitoring of renal functions. Several relevant molecules (urea, uric acid, creatine, creatinine and ammonia) can be analyzed by suitable electrochemical methodologies.

4.4.1 Urea

Urea is the major breakdown product of protein metabolism, and is produced in the liver from CO_2 and ammonia. It is carried in the blood to the kidneys where it is excreted. Urea is also excreted via the skin and the gastrointestinal tract. Increases in blood urea levels are associated with many disorders, most notably renal dysfunction but also congestive heart failure, shock, dehydration and hemorrhage.[66] Decreases can be associated with liver disease. The typical concentration of urea nitrogen in plasma and serum is $80–260\,mg\,L^{-1}$, which corresponds to $171–556\,mg\,L^{-1}$ urea.[40]

Urea can be measured electrochemically using enzyme-based ammonium ISEs, glass pH electrodes or conductimetric techniques. The first two methods are based on the production of ammonium ions by the decomposition of urea with urea aminohydrolase (urease) as shown in Equation (16).

$$(NH_2)_2CO + 2H_2O + H^+ \overset{\text{Urease}}{\rightleftharpoons} HCO_3^- + 2NH_4^+ \qquad (16)$$

Ammonium ions are detected by the ammonium ISE.[67] However, problems exist with respect to ionic interferences from K^+ and Na^+. In the gas electrode, a glass H^+ electrode is used which contains an electrolyte of basic ammonium chloride. At pH 10.5 or greater, all the ammonia generated by the breakdown of urea is in the ionized form. The resulting production of hydroxyl ions leads to a decrease in the H^+ activity. In this way, a detection range for ammonia of $1–10^{-6}\,mol\,L^{-1}$ can be achieved.[68]

The enzymatic conversion of urea to carbonic acid and ammonia can also be monitored conductimetrically, as is the case with the Beckman Astra 8 system. This very simple electrochemical system utilizes the conversion of the nonionic urea to the ionic species NH_4^+ and HCO_3^-. The increase in the ionic concentration of the sample can be monitored by the application of an electrical potential. There are, however, associated difficulties due to the presence of background electrolytic species that will reduce the dynamic range of this system. The presence of ammonia in the sample, however, does not affect this measurement. Conductimetric measurements have also been performed using miniaturized interdigitated electrode arrays.[69]

4.4.2 Uric Acid

Uric acid is the final breakdown product of purine metabolism in human beings, and has implications in renal disease, gout and disorders of nucleic acid metabolism. Its normal physiological range is $0.15–0.42\,mmol\,L^{-1}$.[40] It can be monitored electrochemically using biosensors based on uricase. This results in two products that may be monitored electrochemically (hydrogen peroxide[70] and carbon dioxide), and also the consumption of O_2[71] according to Equation (17).

$$\text{Uric acid} + O_2 + 2H_2O \overset{\text{Uricase}}{\rightleftharpoons} \text{Allantoin} + CO_2 + H_2O_2 \qquad (17)$$

The monitoring of carbon dioxide using a gas electrode is preferable, as there is less risk of interference. This is performed potentiometrically in a similar fashion to that already described for the measurement of P_{CO_2}. Sensors of this type are linear over a physiologically useful range ($2.5 \times 10^{-3}–1 \times 10^{-4}\,mmol\,L^{-1}$).

4.4.3 Creatine and Creatinine

Creatine is an amino acid metabolite and is used as an energy source by muscle tissue. It is carried in the blood as its dehydrogenated form, creatinine. Both components have clinical relevance. Elevated levels of creatinine are well correlated with renal dysfunction and reduced glomerular filtration rates, and can be used in combination with blood urea measurements to determine whether elevated urea levels are due to renal or nonrenal disorders.[72] In several disorders relating to muscle disease, such as muscular dystrophy, hyperthyroidism and poliomyelitis, plasma creatine and urinary creatinine levels are elevated. The typical physiological range is $44–106\,\mu mol\,L^{-1}$ for creatinine in plasma.[40]

Both molecules can be measured amperometrically[73] according to the three-stage enzymatic reaction shown in

Equations (18), (19) and (20).

$$\text{Creatinine} + H_2O \xrightleftharpoons{\text{Creatinine amidohydrolase}} \text{Creatine} \quad (18)$$

$$\text{Creatine} + H_2O \xrightarrow{\text{Creatine amidinohydrolase}} \text{Sarcosine} + \text{Urea} \quad (19)$$

$$\text{Sarcosine} + O_2 \xrightarrow{\text{Sarcosine oxidase}} \text{Glycine} + \text{Formaldehyde} + H_2O_2 \quad (20)$$

Creatinine analysis is available on the i-STAT point-of-care system.

4.4.4 Ammonia

Ammonia is a breakdown product of protein metabolism and is combined with carbon dioxide in the liver to produce urea. Its concentration in plasma is very low ($11-35\,\mu mol\,L^{-1}$),[40] and it is primarily used as a diagnostic for liver disease, and for the liver disorder Reye's syndrome.

The most established system for the detection of ammonia is based on the glass pH electrode, which contains a basic solution of ammonium chloride. The production of hydroxide ions decreases the H^+ activity, bringing about a potential difference. With a LOD as low as $10^{-6}\,mol\,L^{-1}$, this lies within the physiological range of ammonia measurement.[74]

4.5 Vitamins

Electrochemistry can be applied to the analysis of most vitamins. Various techniques have been used, the most favorable combining separation techniques such as liquid chromatography with electrochemical detection. These methods are summarized in Table 5.

4.5.1 Vitamin A

The A vitamins are a group of compounds, the most important one biologically being all-*trans*-retinol, which predominates in serum. Several electrochemical methodologies have been investigated for their ability to characterize and quantify vitamin A. The classical methodology utilized the DME.[75] Due to its insolubility in aqueous buffers, the supporting electrolyte was composed of quaternary ammonium compounds in a high concentration organic solvent buffer.

The levels of all-*trans*-retinol in human serum have been measured by LCEC.[76] Separation and detection was performed in a supporting electrolyte composed of 95% (v/v) methanol, 0.075 M acetate buffer, pH 5.0. Detection was via a GCE at +1.1 V vs Ag/AgCl in a wall-jet type electrochemical cell. Circulatory levels

Table 5 Electrochemical techniques for the analysis of vitamins

Vitamin	Electrochemical technique	Clinical range	Analytical range
Vitamin A (all-*trans*-retinol)	Polarography (DME)		
	LCEC	$676\,ng\,mL^{-1}$ (serum)	
Vitamin B_1 (thiamin)	Polarography (DME)		
	Differential pulse polarography (DME)		
Vitamin B_2 (riboflavin)	Differential pulse polarography		
	Adsorptive stripping voltammetry	$2.5 \times 10^{-11}\,mol\,L^{-1}$	
Vitamin B_6 (pyridoxine)	LCEC	$60.4-69.8\,nmol\,L^{-1}$	$5-80\,nmol\,L^{-1}$
Folic acid	Adsorptive stripping voltammetry	$5 \times 10^{-9}\,mol\,L^{-1}$ (serum)	
		$1 \times 10^{-10}\,mol\,L^{-1}$ (urine)	
	LCEC		$2 \times 10^{-9}\,mol\,L^{-1}$
Nicotinamide	Differential pulse polarography (DME)		$1.0-0.02\,\mu g\,mL^{-1}$
Vitamin C (L-ascorbic acid)	Amperometric enzyme electrode		$5 \times 10^{-4}-5 \times 10^{-5}\,mol\,L^{-1}$
	LCEC		
Vitamin D_2 (ergocalciferol)	Polarography (DME)		
Vitamin D_3 (cholicalciferol)	Polarography (DME)		
	Differential pulse voltammetry		$2 \times 10^{-4}-2 \times 10^{-6}\,mol\,L^{-1}$
Vitamin E (tocopherol)	Polarography	$6.1-13.6\,\mu g\,mL^{-1}$ (serum)	
	LCEC		$>0.1\,\mu g\,mL^{-1}$
Vitamin K_1 (phylloquinone)	Differential pulse polarography	$0.08-1.24\,ng\,mL^{-1}$ (serum)	
	LCEC		

of all-*trans*-retinol were found to be in the region of 676 ng mL^{-1}, with 100% recovery.

4.5.2 Vitamin B_1 (Thiamin)

Thiamin has been shown to undergo oxidation using direct current polarography at a DME.[77] In Britton–Robinson buffer, at pH 9.0–9.3, it yields a half-wave potential of −0.4 V vs an SCE due to the reduction of its thiol group. Differential pulse polarography at −0.42 V has also been performed at a DME in NaOH.[78]

4.5.3 Vitamin B_2 (Riboflavin)

Riboflavin has been measured using differential pulse polarography[79] and adsorptive stripping voltammetry.[80] Differential pulse polarography is performed in phosphate buffer, at pH 7.2 and can detect levels as low as 100 ng mL^{-1}. Adsorptive stripping voltammetry has been performed at a static mercury electrode with adsorption at −0.2 V. Stripping using a negative scan differential pulse waveform in 0.001 M NaOH results in a LOD of 2.5×10^{-11} mol L^{-1}.

4.5.4 Vitamin B_6 (Pyridoxine)

Pyridoxine is predominately found in the human circulation as pyridoxal 5′-phosphate. It has a plasma concentration between 60.4 and 69.8 nmol L^{-1}. It is a coenzyme in several metabolic reactions involving calalysis by tyrosine apodecarboxylase. It is most readily measured using enzyme-based sensors utilizing this enzyme system. It has been measured in an LCEC system using reverse-phase chromatography followed by amperometric detection at an enzyme electrode containing tyrosine apodecarboxylase. L-Tyrosine is converted to tyramine in an amount proportional to the concentration of pyridoxine 5′-phosphate.[81] Tyramine is measured at a GCE at +0.85 V vs Ag/AgCl. The analytical range of this method is between 5 and 80 nmol L^{-1}.

4.5.5 Folic Acid

Folic acid undergoes cathodic reduction, and this has been used as the basis of several electrochemical methods.[82,83] Methods utilizing adsorptive stripping voltammetry have been applied for the analysis of folic acid in serum and urine. For serum, separation on a reverse-phase C_{18} column was required. Preconcentration was performed at a static HMDE in 0.1 M acetate buffer, at pH 5.0 for 70 s. Using phase sensitive alternating current voltammetry at 75 Hz and 10 mV s^{-1}, scanning between −0.35 and −0.70 V results in a peak at −0.65 V and a LOD of 5×10^{-9} mol L^{-1}. In urine, detection levels as low as 1×10^{-10} mol L^{-1} were achieved in a supporting electrolyte of 0.1 M H_2SO_4 with preconcentration for 2 min at −0.3 V vs Ag/AgCl.

LCEC methods on plasma and CSF have been performed with limits of detection of 2×10^{-9} mol L^{-1}. Sample pretreatment was required in the form of protein precipitation and two stages of concentration/separation by reverse-phase and anion-exchange chromatography. Amperometric detection was performed at +0.3 V vs Ag/AgCl.[84]

4.5.6 Nicotinamide

Nicotinamide has been measured using differential pulse polarography at a DME.[79] Detection ranges of 1.0–0.02 μg mL^{-1} have been reported in LiOH supporting electrolyte buffer.

4.5.7 Vitamin C (L-Ascorbic Acid)

Ascorbic acid is readily oxidized at mercury and carbon electrodes with the resulting irreversible loss of two electrons and two protons to form dehydroascorbic acid. It has been measured with several electrochemical techniques, notably polarography, voltammetry, LCEC and amperometric enzyme electrodes.[85] An amperometric enzyme biosensor utilizing ascorbate oxidase has been used, following the reaction scheme shown in Equation (21).

$$\text{L-Ascorbic acid} + \tfrac{1}{2}O_2 \xrightleftharpoons{\text{Ascorbate oxidase}} \text{Dehydroascorbic acid} + H_2O_2 \quad (21)$$

This system had a linear range of 4.5×10^{-4}–8×10^{-6} mol L^{-1}.

4.5.8 Vitamin D

Vitamin D is a mixture of two major vitamin forms: D_2 (ergocalciferol) and D_3 (cholecalciferol). Both undergo reduction at a DME under the same conditions as those described for vitamin A, yielding the half-wave potentials −2.01 and −2.25 V respectively.[86] LCEC detections of D_2 and D_3 have been performed at a GCE in 95% (v/v) methanol/0.05 mol L^{-1} acetate buffer, pH 6. Detection was performed at +1.3 V vs Ag/AgCl and yielded linear ranges of 10–100 ng for D_3 and 20–200 ng for D_2.[87]

4.5.9 Vitamin E

Vitamin E is a group of tocopherols all characterized by an oxidizable phenolic hydroxy group. α-Tocopherol is the most easily oxidizable by polarography.[88] They have a combined serum concentration of 6.1–13.6 μg mL^{-1}.

Tocopheronolactone is a vitamin E metabolite and has been measured in urine using polarography following extraction in ethanolic acetate buffer, at pH 5.3, yielding a half-wave potential of −0.1 V.[89]

The tocopherols have also been measured in serum using LCEC, following extraction into ethanol and hexane. Chromatography was performed on a C_{18} column in 96% (v/v) methanol : water, 40 mM sodium perchlorate mobile phase. Tocopherol was detected coulometrically at +0.4 V, with a detection limit of 50 pg.[90]

4.5.10 Vitamin K

The K vitamins are another structurally similar group characterized by the presence of a 1,4-naphthoquinone moiety. Levels of vitamin K_1 (phylloquinone) in blood are in the range 0.08–1.24 ng mL^{-1}. Vitamin K can undergo reduction at a mercury electrode. Differential pulse polarography has been used for its determination in plasma, following extraction in methanol/chloroform, with a peak observed at −0.58 V vs SCE, and was capable of the measurement of concentrations as low as 200 ng mL^{-1}.[91]

Vitamin K_1 was analyzed in human plasma using LCEC. Vitamin K_1 required extraction into ethanol–hexane and concentration on a normal phase HPLC column. Chromatographic separation was performed using reverse-phase HPLC on a C_8 column in methanol acetate (95 : 5), at pH 3.0. Electrochemical detection was performed in a thin-layer cell at a GCE at a potential of −1.0 V vs Ag/AgCl. The LOD was in the order of 500 pg.[92]

4.6 Lipids

The main lipid to gain major clinical interest has been cholesterol. This is due to its links with atherosclerosis and myocardial infarction. Cholesterol has been measured using adsorptive stripping voltammetry and also amperometrically at enzyme electrodes.

Adsorptive stripping voltammetry has been performed at a static mercury electrode, utilizing a 10 min preconcentration period. Stripping was performed by differential pulse polarography. A stripping current was obtained at −1.0 V vs Ag/AgCl, and was linear between 0.2 and 0.8 μmol L^{-1} with a LOD of 20 nmol L^{-1}.[93]

Enzyme electrodes have also been used for cholesterol measurement. Amperometric detection of cholesterol in the range 6×10^{-5}–5×10^{-3} was shown to be possible with a carbon paste electrode incorporating cholesterol oxidase, by monitoring the production of hydrogen peroxide.[94] Cholesterol in serum has also been measured using electrodes which entrap cholesterol oxidase in a conductive polymer.[95] The Biologix Inc., EPOC hand-held point-of-care blood analysis system incorporates cholesterol measurement.

4.7 Trace Metals

The sensitivity of electrochemical methodologies has made them an excellent technique for the detection and measurement of trace metals of clinical importance. Metals such as lead, copper, zinc and many others have traditionally been analyzed by AAS. However, they are also suited to analysis by electrochemical methods, mostly using voltammetric methods. The electrochemical techniques, sample matrixes and sample treatments required for analysis have been excellently reviewed by Wang[96] according to Table 6.

4.8 Amino Acids

Amperometric and potentiometric detection of amino acids is based on the enzyme L-amino acid oxidase, according to the reaction in Equation (22).

$$RCHNH_3^+COO^- + H_2O + O_2 \underset{}{\overset{\text{L-Amino acid oxidase}}{\rightleftharpoons}} RCOCOO^- + NH_4^+ + H_2O_2 \quad (22)$$

Amperometric measurement is performed by monitoring the oxidation of hydrogen peroxide at a platinum or carbon electrode.[127] This reaction can also be monitored potentiometrically at an ammonia-selective electrode.[128] Fiber optic biosensors for D-amino acids in blood have also been developed which generate highly sensitive fluorescent signals. These have detection limits of the order of 1×10^{-7}.[129]

In vivo amino acid measurement has also been performed by combining the techniques of microdialysis, capillary electrophoresis and electrochemical detection. Detection limits of aspartate and glutamate were in the order of 10^{-7} mol L^{-1}.[130]

4.9 Blood Group Antigens

Several electrochemical and biosensor-based techniques have been applied to the analysis of blood group antigens and their antibodies. Erythrocytes containing the rhesus D antigen (RhD^+) have been immobilized within the electroactive polymer polypyrrole and the binding of rhesus anti-D antibodies has been monitored by the resistivity of the polymer while cycling between +0.35 and −0.7 V. Decreases in resistivity in the order of 1.1 Ω were observed for an antibody concentration of 250 μg mL^{-1}.[131]

An amperometric and an optical method have both been applied to the analysis of erythrocytes according to

Table 6 Stripping voltammetric techniques for the analysis of trace metals. (Reproduced from J. Wang, *Electroanalytical Techniques in Clinical Chemistry and Laboratory Medicine*. Copyright © 1988 John Wiley & Sons, Inc. Translated by permission of John Wiley & Sons, Inc. All rights reserved.)

Trace metal	Matrix	Electrochemical technique	Sample handling	Refs.
Antimony (Sb)	Blood	Staircase (MFE)	Wet digestion	97
	Urine	Staircase (MFE)	Wet digestion	97
Arsenic (As)	Blood	Staircase (MFE)	Wet digestion	98
	Urine	Staircase (MFE)	Wet digestion	98
Cadmium (Cd)	Blood	Differential pulse (HMDE)	Ashing	99
		Differential pulse (HMDE)	Wet and pressure	100
		Differential pulse (MFE)	Dilution	101
		Potentiometric (MFE)	Dilution	102
	Urine	Differential pulse (MFE)	Wet digestion	103, 104
		Differential pulse (HMDE)		105
		Differential pulse (MFE)	Dilution	101
	Teeth	Differential pulse (HMDE)	Wet digestion	106, 107
	Hair	Differential pulse (HMDE)	Wet digestion	108, 109
		Staircase (MFE)	Wet digestion	110
	Nails	Staircase (MFE)	Wet digestion	110
	Feces	Differential pulse (HMDE)	Wet digestion	109
	Bone	Derivative pulse (HMDE)	Wet digestion	111
Cobalt (Co)	All biological matrixes	Differential pulse (HMDE)	Dry ashing and chelate	112
Copper (Cu)	Blood	Differential pulse (HMDE)	Ashing	99
		Differential pulse (HMDE)	Wet and pressure	100
		Potentiometric (MFE)	Dilution	102
	Urine	Coulostatic (MFE)	Dilution	113
	Teeth	Differential pulse (HMDE)	Wet digestion	106, 107
	Hair	Staircase (MFE)	Wet digestion	110
		Differential pulse (HMDE)	Wet digestion	108, 109
	Nails	Staircase (MFE)	Wet digestion	110
	Feces	Differential pulse (HMDE)	Wet digestion	109
Gold (Au)	Blood	Chronopotentiometric (carbon paste)		114
Lead (Pb)	Blood	Differential pulse (HMDE)	Wet digestion	115
		Differential pulse (MFE)	Wet digestion	116
		Differential pulse (HMDE)	Ashing	99
		Differential pulse (HMDE)	Wet and pressure	100
		Differential pulse (MFE)	Dilution	101
		Potentiometric (MFE)	Dilution	102
		Staircase (MFE)	Wet digestion	117
		Staircase (MFE)	Ion exchange	118
	Urine	Differential pulse (MFE)	Wet digestion	103, 104
		Differential pulse (HMDE)		105
		Differential pulse (MFE)	Dilution	101
		Potentiometric (MFE)		119
		Linear scan (HMDE)	Acidification	120
		Coulostatic	Dilution	113
	Teeth	Differential pulse (HMDE)	Wet digestion	106, 107
	Hair	Differential pulse (HMDE)	Wet digestion	108, 109
		Staircase (MFE)	Wet digestion	110
	Nails	Staircase (MFE)	Wet digestion	110
	Feces	Differential pulse (HMDE)	Wet digestion	109
	Bone	Derivative pulse (HMDE)	Wet digestion	111
Mercury (Hg)	Urine	Potentiometric	Acidification	121
Nickel (Ni)	All biological samples	Differential pulse (HMDE)	Dry ashing and chelation	112
	Nails	Differential pulse (HMDE)	Wet digestion and chelation	122
Selenium (Se)	All biological samples	Cathodic stripping (HMDE)		123
Thallium (Tl)	Urine	Differential pulse (HMDE, MFE)		124
		AC voltammetry	Acidification	125

Table 6 (*continued*)

Trace metal	Matrix	Electrochemical technique	Sample handling	Refs.
Zinc (Zn)	Teeth	Differential pulse (HMDE)	Wet digestion	106, 107
	Hair	Differential pulse (HMDE)	Wet digestion	108, 109
		Staircase (MFE)	Wet digestion	110
	Nails	Staircase (MFE)	Wet digestion	110
	Feces	Differential pulse (HMDE)	Wet digestion	109
	Eye tissue	Differential pulse (HMDE)	Wet digestion	126

MFE, mercury film electrode.

their ABO grouping. The amperometric system used flow injection across a GCE on which anti-A IgM antibodies had been covalently immobilized. A sandwich immunoassay was performed following binding of the antigen to the electrode surface. This required addition of an anti-A antibody labelled with peroxidase. Detection was via the reduction of hexacyanoferrate at -0.25 V vs Ag/AgCl. A linear relationship between electrode response and cell numbers was achieved between 1 and 30×10^8 cells mL^{-1}.[132] An optical system using the surface plasmon resonance-based BIAcore biosensor was also applied to blood group antigen detection. This system was able to differentiate between A^+ and B^+ red blood cells, over a range of $0.33–2.33 \times 10^9$ cells mL^{-1}. However, clogging of flow channels with cell materials was a problem. More optimal flow systems would be required for better flow dynamics of cellular materials.[133]

4.10 DNA Biosensors

The DNA revolution has begun and is continuing. Vast amounts of data are being generated by the Human Genome Project which will result in great insights, central to which will be the better understanding, diagnosis and treatment of genetic disease. Presently, many molecular biological techniques are based on slow amplification and electrophoretic separation techniques. Biosensors offer the potential of great sensitivity, specificity and speed that could allow them to replace such methodologies. Most of these technologies involve DNA hybridization, which gives the analytical system its selectivity. Many sensors are being developed that can detect the hybridization of many DNA strands to known single-stranded DNA (ssDNA) sequences. These ''gene chips'' can hold up to 10 000 different oligomers. Various transduction methodologies are being used such as optical, piezoelectric and electrochemical devices.[134] The electrochemical devices are based on the selective intercalation of certain redox complexes with double-stranded DNA (dsDNA) over ssDNA. When this occurs, changes in the intercalating agent can be monitored using voltammetry, chronoamperometry and other techniques such as constant-current chronopotentiometric stripping analysis.

Millan et al. have used the covalent immobilization of ssDNA sequences to both glassy carbon and carbon paste electrodes and studied the interaction of various complexes with resulting dsDNA hybrids.[135–137] These included reducible cobalt(II) complexes, tris(1,10-phenanthroline)cobalt(III) perchlorate and tris(2-2′-bipyridyl)cobalt(III) perchlorate, and the osmium complex Os(2,2′-bipyridine)$_3$ chloride. These form reversible complexes with dsDNA that concentrate in the DNA layer close to the surface of the electrode, and yield modified voltammetric currents compared with those observed when uncomplexed with ssDNA. Wang et al. have used daunomycin in a similar manner on DNA-modified carbon paste electrodes.[138] Intercalation of daunamycin can be visualized by cyclic voltammetry which shows a decrease and a shift in the daunamycin redox coupled peaks. Chronoamperometric stripping yielded well-defined peaks after a concentration time of 60 s at $+0.5$ V vs Ag/AgCl.

More research is needed if electrochemical DNA biosensors are to compete effectively with other technologies, but they have the advantages of simplicity and low cost. Improvements in sensitivity and specificity may allow them to compete with the relatively costly optical techniques available.

4.11 Hormones, Pharmaceuticals and Antibiotics

The need to rapidly and continuously monitor many molecules such as drugs, hormones and antibiotics is ever increasing. This allows real-time information to be obtained as to the behavior, metabolism, and distribution of these substances in the body. In this regard, other techniques such as microdialysis and microelectrodes are proving important allies in the development of this technology. Table 7 gives a summary of the varied, if not exhaustive applications that electrochemical and other biosensors are being put to in the monitoring and analysis of pharmaceutical drugs, hormones and antibiotics. These serve to illustrate the wide range of analytes that are amenable to these types of measurement. Of note is the expansion of DNA-based technologies that are exploring drug–nucleic acid interactions, probing the effects these drugs have at the nuclear level.

Table 7 Electrochemical techniques for the analysis of pharmaceutical drugs, antibiotics, hormones and proteins

Analyte	Electrochemical technique	Electrode type	Fluidic system	Matrix	Range/LOD	Ref.
Acetaminophen	Square wave voltammetry					139
Acetylcholine	Amperometry		Microdialysis			140
Adriamycin	Adsorptive stripping voltammetry	Carbon paste		Urine		141
Bromazepam	Differential pulse polarography	DME		Plasma		142
Carbimazole	Voltammetry	GCE			5×10^{-5}–1×10^{-3} mol L^{-1}	143
Cephalosporin	Cathodic stripping voltammetry	DME		Biological samples	1×10^{-5}–2.5×10^{-8} mol L^{-1}	144
Chloramphenicol			LCEC	Serum	3–1000 μg mL^{-1}	145
Chlordeazepoxide	Differential pulse polarography	DME		Plasma		146
Chlorpromazine	Differential pulse voltammetry	Carbon paste		Urine		147
Cimetidine	Square wave polarography	DME		Urine		148
Cortisol	Amperometry	Antibody			1×10^{-7}–1×10^{-5} mol L^{-1}	149
Cyclophosphamide	Potentiometry	PVC–carboxylate membrane			10^{-3}–10^{-5} mol L^{-1}	150
Dontrolene	Differential pulse polarography	DME		Plasma, urine		151
Follicle stimulating hormone	Amperometry	Antibody			2.1 U L^{-1}	152
Ifosphamide	Potentiometry	PVC–carboxylate membrane			10^{-3}–10^{-5} mol L^{-1}	150
Ibuprofen	Potentiometry	PVC-feroin membrane			1×10^{-2}–2×10^{-5} mol L^{-1}	151
Luteinising hormone						152
Methimazole	Voltammetry	GCE			5×10^{-5}–1×10^{-3} mol L^{-1}	143
Methylthiouracil	Voltammetry	GCE			5×10^{-5}–1×10^{-3} mol L^{-1}	143
Mitomycin C	Coulometry		LCEC			154
Oestriol			LCEC			155
Penicillin	Potentiometry	Glass electrode				156
Phenothiazine	Chronopotentiometry	DNA			>5 nmol L^{-1}	157
Progesterone						158
Propylthiouracil	Voltammetry	GCE			5×10^{-5}–1×10^{-3} mol L^{-1}	143
Salycilic acid	Amperometry	Enzyme	LCEC	Urine	$>3.5 \times 10^{-6}$ mol L^{-1}	159
Testosteroids			LCEC			160
Theophylline						161
Thyroid stimulating hormone						162
Trimethoprin	Differential pulse polarography	DME		Blood, urine		163
Vinblastine	Differential pulse polarography	DME		Plasma, urine		164
Warfarin	Potentiometry	PVC–feroin membrane			1×10^{-2}–2×10^{-5} mol L^{-1}	153

5 CONCLUSION

Electrochemical techniques are widely applicable to the clinical environment. They can compete with traditional analytical methodologies in terms of analytical range and limits of detection. They excel in the speed at which assays can be performed. This has brought the possibility of continuous monitoring, and with it progress in in vivo systems. Integral to this technique are others such as flow analysis and microdialysis which make these systems feasible for human monitoring.

The range of analytes that can be monitored using biosensors has also expanded beyond simple electroactive molecules. Antibody and DNA biosensors are making possible the analysis of many more complex biomolecules and pharmaceutical drugs. It is likely that in future all clinical analyses will be performed simply and quickly at the patient's bedside, bringing the clinical laboratory to the patient. It is certain that electrochemical electrodes and biosensors will also be there.

ACKNOWLEDGMENTS

The authors would like to thank Andrea Hiney at the Diabetes Day Care Centre in St James's Hospital, Dublin, and Jemma Farrell in the Clinical Chemistry Unit at Temple Street Children's Hospital, Dublin, for their assistance.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
BUN	Blood Urea Nitrogen
CSF	Cerebrospinal Fluid
DME	Dropping Mercury Electrode
dsDNA	Double-stranded DNA
FDA	Food and Drug Administration
FES	Flame Emission Spectrometry
FIA	Flow Injection Analysis
GCE	Glassy Carbon Electrode
Hb	Hemoglobin
HHb	Deoxyhemoglobin
HMDE	Hanging Mercury Drop Electrode
HPLC	High-performance Liquid Chromatography
ISE	Ion-selective Electrode
LCEC	Liquid Chromatography–Electrochemical
LDH	Lactate Dehydrogenase
LOD	Limit of Detection
MFE	Mercury Film Electrode
NAD^+	Oxidized Form of Nicotinamide Adenine Dinucleotide
NADH	Reduced Form of Nicotinamide Adenine Dinucleotide
PVC	Poly(vinyl Chloride)
SCE	Standard Calomel Electrode
SEM	Scanning Electron Microscopy
SHE	Standard Hydrogen Electrode
ssDNA	Single-stranded DNA

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Automation in the Clinical Laboratory • Biosensor Design and Fabrication • Capillary Electrophoresis in Clinical Chemistry • DNA Arrays: Preparation and Application • Electrolytes, Blood Gases, and Blood pH • Glucose Measurement • Point-of-care Testing • Urinalysis and Other Bodily Fluids

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Ion-selective Electrodes: Fundamentals • Pulse Voltammetry

Kinetic Determinations **(Volume 12)**
Electrocatalysis-based Kinetics Determinations

REFERENCES

1. C. Ziegler, W. Göpel, 'Biosensor Development', *Curr. Opin. Chem. Biol.*, **2**, 585–591 (1998).
2. S.B. Tuwiner, *Diffusion and Membrane Technology*, Reinhold, New York, 1962.
3. R. Ballerstadt, J.S. Schultz, 'Sensor Methods for Use with Microdialysis and Ultrafiltration', *Adv. Drug Delivery Rev.*, **21**, 225–238 (1996).
4. R.P. O'Neill, J.P. Lowry, M. Mas, 'Monitoring of Brain Chemistry In Vivo: Volumetric Techniques, Sensors and Behavioral Applications', *Crit. Rev. Neurobiol.*, **12**(1–2), 69–127 (1998).
5. R. John, M.J. John, 'Evaluation of the Microdialysis Electrode Concept as a Generic Biosensing Technology', *Electroanalysis*, **10**(16), 1125–1129 (1998).
6. C.J. Cook, C.E. Devine, 'Antibody-based Electrodes for Hormonal and Neurotransmitter Measurements In Vivo', *Electroanalysis*, **10**(16), 1108–1111 (1998).
7. E.H. Hansen, 'Principles and Applications of Flow Injection Analysis in Biosensors', *J. Mol. Recognit*, **9**, 316–325 (1996).
8. L. Denroy, L. Bert, S. Parrot, F. Robert, B. Renaud, 'Assessment of Pharmacodynamic and Pharmacokinetic Characteristics of Drugs Using Microdialysis Sampling and Capillary Electrophoresis', *Electrophoresis*, **19**, 2841–2847 (1998).

9. A.K. Covington, *Ion Selective Electrode Methodology*, CRC Press, Boca Raton, Vol. 2, 1979.
10. M. Valcarcel, M.D. Luque de Castro, *Flow-injection Analysis, Principles and Applications*, John Wiley and Sons, New York, 1987.
11. A.G. Fogg, N.K. Bsebsu, 'Flow Injection Voltammetric Determination of Phosphate: Direct Injection of Phosphate into Molybdate Reagent', *Analyst*, **107**, 566–570 (1982).
12. A.G. Fogg, N.K. Bsebsu, 'Differential-pulse Voltammetric Determination of Phosphate as Molybdovando Phosphate at a Glassy Carbon Electrode and Assessment of Eluents for Flow Injection Voltammetric Determination of Phosphate, Silicate, Arsenate and Germanate', *Analyst*, **106**, 1288–1295 (1981).
13. J. Ruzicka, E.H. Hansen, 'Integrated Microconduits in Flow Injection Analysis', *Anal. Chim. Acta*, **161**, 1–25 (1984).
14. R.W. Cattrall, I.C. Hamilton, 'Coated Wire Ion-selective Electrodes', *Ion-Selec. Elect. Rev.*, **6**(2), 125–171 (1984).
15. R.W. Cattrall, H. Freiser, 'Coated Wire Ion-selective Electrodes', *Anal. Chem.*, **43**(13), 1905–1906 (1971).
16. C. Sánchez-Pedreño, J.A. Ortuño, A.L. Ros, 'Kinetic Potentiometric Determination of Iodide with an Antimony (V) Coated-wire Ion-selective Electrode', *Analyst*, **117**, 1619–1621 (1992).
17. R. Matesic-Puac, M. Dimitrijevic, M. Sak-Bosnar, 'Influence of Plasticizer, and Metallic and Graphite Conducting Substrate on the Characteristics of a Poly(Vinyl Chloride) Membrane Coated Wire Calcium Ion-selective Electrode', *Analyst*, **117**, 1323–1327 (1992).
18. H. Hara, Y. Kondoho, Mitani, S. Okazaki, 'Solid Solvent Membrane Coated Wire Electrodes and their Response to Nitrate, *p*-Toluenesulfonate, Chloride, Calcium and Potassium', *Anal. Chem.*, **62**, 1139–1143 (1990).
19. J.F. Pearson, J.M. Slater, V. Jovanovic, 'Coated-wire and Composite Ion-selective Electrodes Based on Doped Poly(Pyrrole)', *Analyst*, **117**, 1885–1890 (1992).
20. R.W. Cattrall, K.T. Fong, 'Serum Calcium and Potassium Determination by the Split Membrane Ion-selective Electrode Technique', *Talanta*, **25**, 541–543 (1978).
21. E.K. Yun, A.J. Prince, J.E. McMillin, L.E. Welch, 'High-performance Liquid Chromatographic Separation and Electrochemical Detection of Cephalosporins', *J. Chromatogr., B*, **712**, 145–152 (1998).
22. A. Tolokán, I. Klebovich, K. Balogh-Nemes, G. Horvai, 'Automated Determination of Levodopa and Carbidopa in Plasma by High-performance Liquid Chromatography–Electrochemical Detection Using an On-line Flow Injection Analysis Sample Pretreatment Unit', *J. Chromatogr., B*, **698**, 201–207 (1997).
23. G.J. Kost, C. Hague, 'The Current and Future-status of Critical Care Testing and Patient Monitoring', *Am. J. Clin. Pathol.*, **104**(4), S2–S17 (1995).
24. J. Zhou, D.M. Heckert, H. Zuo, C.E. Lunte, S.M. Lunte, 'On-line Coupling of In Vivo Microdialysis with Capillary Electrophoresis/Electrochemistry', *Anal. Chim. Acta*, **379**, 307–317 (1999).
25. R.P. De Cresce, M.S. Lifshitz, 'Clinical Chemistry Automation', in *Clinical Laboratory Instrumentation and Automation: Principles, Applications and Selection*, eds. K.M. Ward, C.A. Lehmann, A.M. Leiken, W.B. Saunders Co., Philadelphia, 1994.
26. R.A. Bageant, 'Variations in Arterial Blood Gas Measurements Due to Sampling Techniques', *Respir. Care*, **20**(6), 565–570 (1975).
27. G.B. Schumann, S.C. Schweitzer, 'Examination of Urine', in *Clinical Chemistry: Theory, Analysis and Correlation*, 2nd edition, eds. L.A. Kaplan, A.J. Pesce, Mosby, St. Louis, 820–849, 1989.
28. A. Lynch, D. Diamond, P. Lemoine, J. McLaughlin, M. Leader, 'Solid-state Ion-selective Electrode Arrays', *Electroanalysis*, **10**, 1096–1100 (1998).
29. M.R. Smyth, *Chemical Analysis in Complex Matrices*, Ellis Horwood–Prentice Hall, New York, 1992.
30. O. Siggaard-Anderson, *The Acid–Base Status of the Blood*, 4th edition, Williams and Wilkins, Baltimore, 1974.
31. H.F. Weisberg, *Water, Electrolytes and Acid–Base Balance*, 2nd edition, Williams and Wilkins, Baltimore, 1962.
32. A. White, P. Handler, E.L. Smith, *Principles of Biochemistry*, 5th edition, McGraw-Hill, New York, 1973.
33. G. Eisenman, *Glass Electrodes for Hydrogen and Other Cations*, Marcel Dekker, New York, 1976.
34. J.D. Czaban, 'Electrochemical Sensors in Clinical Chemistry: Yesterday, Today, Tomorrow', *Anal. Chem.*, **57**, 345A–356A (1985).
35. 'Hemoglobin Oxygen Saturation and Related Quantities in Blood: Terminology, Measurement and Reporting', National Committee for Clinical Laboratory Standards, Philadelphia, 1979.
36. D. Beyerl, 'Non-invasive Measurement of Blood Oxygen Levels', *Am. J. Med. Tech.*, **48**(5), 355–359 (1982).
37. R.G. Buckles, H. Heitmann, M.B. Laver, 'A Theoretical and Practical Analysis of pO_2 Microelectrode Behavior: The Three Shell Model', The National Bureau of Standards (USA) Special Publication 450, 207–225, June 1977.
38. V.F. Fairbanks, 'Hemoglobin, Hemoglobin Derivatives and Myoglobin', in *Fundamentals of Clinical Chemistry*, 2nd edition, ed. N. Tietz, W.B. Saunders Co., Philadelphia, 1976.
39. J. Vesely, D. Weiss, K. Stulik, *Analysis with Ion-selective Electrodes*, John Wiley and Sons, New York, 1979.
40. N.W. Tietz (ed.), *Clinical Guide to Laboratory Tests*, W.B. Saunders, Philadelphia, 1983.
41. P. Bijster, K.L.J. Vink, 'Direct Potentiometric Determination of Sodium Ion in Blood. II. Influence of Cations', *Clin. Chem.*, **30**, 865–870 (1984).
42. R.J. Forster, D. Diamond, 'Nonlinear Calibration of Ion-selective Electrode Arrays for Flow-injection Analysis', *Anal. Chem.*, **64**, 1721–1726 (1992).

43. J.F. Zilva, P.R. Pannall, *Clinical Chemistry in Diagnosis and Treatment*, 3rd edition, Year Book Medical Publishers, Chicago, 1979.
44. F.S. Apple, D. Koch, S. Graves, J.H. Landenson, 'Relationship Between Direct-potentiometric and Flame-photometric Measurement of Sodium in Blood', *Clin. Chem.*, **28**, 1931–1935 (1982).
45. J.D. Czaban, A.D. Cormier, K.D. Legg, 'Establishing the Direct-potentiometric "Normal" Range for Na/K: Residual Liquid Junction Potential and Activity Coefficient Effects', *Clin. Chem.*, **28**, 1936–1945 (1982).
46. M.M. Christopher, E.B. Belknap, D.J. Meyer, M.N. Lackey, L.M. Vap, 'Comparison of Methods for Sodium and Potassium Determination in Llama Urine', *Am. J. Vet. Res.*, **57**(1), 25–30 (1996).
47. M.L. Bishop, J.L. Duben-Von Laufen, E.P. Fody, *Clinical Chemistry: Principles, Procedures and Correlations*, J.J. Lippincott, Philadelphia, 1985.
48. L.W. Lee, *Elementary Principles of Laboratory Instruments*, 4th edition, Mosby, St. Louis, 1978.
49. P. Anker, H.B. Jenny, U. Wuthier, R. Asper, D. Ammann, W. Simon, 'Determination of $[K^+]$ in Blood Serum with Valinomycin-based Silicone Rubber Membrane of Universal Applicability to Body Fluids', *Clin. Chem.*, **29**, 1447–1448 (1983).
50. J. Ruzicka, E.H. Hansen, J.C. Tjell, 'Selectrode–The Universal Ion-selective Electrode. Part VI. The Calcium (II) Selectrode Employing a New Ion Exchanger in a Nonporous Membrane and a Solid-state Reference System', *Anal. Chim. Acta*, **67**, 155–178 (1973).
51. D. Diamond, F.J. Sáez de Viteri, 'Ion-selective Electrodes and Optodes', in *Principles of Chemical and Biological Sensors*, ed. D. Diamond, John Wiley and Sons, New York, 19–72, 1998.
52. K. Kobiro, Y. Tobe, K. Watanabe, H. Yamada, K. Suzuki, 'Highly Selective Lithium Ion Electrode Based on Decalino-14-crown-4', *Anal. Lett.*, **26**, 49–54 (1993).
53. G.D. Christian, 'Analytical Strategies for Measurement of Lithium', *Biosens. Bioelectron.*, **12**, vii–viii (1997).
54. G.J. Moody, J.D.R. Thomas, 'The Role of Ion-selective Electrodes in Cystic Fibrosis Studies', *Ion-Sel. Elec. Rev.*, **2**, 73–115 (1980).
55. S.B. Park, W. Matuszewski, M.E. Meyerhoff, Y.H. Liu, K.M. Kadish, 'Potentiometric Anion Selectivities of Polymer Membranes Doped with Indium(III)-porphyrins', *Electroanalysis*, **3**, 909–916 (1991).
56. F. Behm, D. Ammann, W. Simon, K. Brunfeldt, J. Halstrom, 'Cyclic Octa- and Decapeptides as Ionophores for Magnesium', *Helv. Chim. Acta*, **68**, 110–118 (1985).
57. E. Malinowska, A. Manzoni, M.E. Meyerhoff, 'Potentiometric Response of Magnesium-selective Membrane Electrode in the Presence of Nonionic Surfactants', *Anal. Chim. Acta*, **382**, 265–275 (1999).
58. M.C. Rashkin, C. Bosken, R.P. Baughman, 'Oxygen Delivery in Critically Ill Patients: Relationship to Blood Lactate and Survival', *Chest*, **87**, 580–584 (1985).
59. M.H. de Keijzer, R.W. Brandts, P.G.W. Brans, 'Evaluation of a Biosensor for the Measurement of Lactate in Whole Blood', *Clin. Biochem.*, **32**(2), 109–112 (1999).
60. E. Wilkins, P. Atanasov, 'Glucose Monitoring: State of the Art and Future Possibilities', *Med. Eng. Phys.*, **18**(4), 273–288 (1996).
61. T.M. Park, E.I. Iwuoha, M.R. Smyth, B.D. MacCraith, 'Sol–Gel-based Amperometric Glucose Biosensor Incorporating an Osmium Redox Polymer as Mediator', *Anal. Comm.*, **33**(8), 271–273 (1996).
62. H.Y. Liu, H.H. Li, T.L. Ying, K. Sun, Y.Q. Qin, D.Y. Qi, 'Amperometric Biosensor Sensitive to Glucose and Lactose Based on Co-immobilization of Ferrocene, Glucose Oxidase, Beta-galactosidase and Mutarotase in Beta-cyclodextrin Polymer', *Anal. Chim. Acta*, **358**(2), 137–144 (1998).
63. E.E. Szabo, N. Adanyi, M. Varadi, 'Application of Biosensor for Monitoring Galactose Content', *Biosens. Bioelectron.*, **11**(10), 1051–1058 (1996).
64. P.A. Paredes, J. Parellada, V.M. Fernandez, I. Katakis, E. Dominguez, 'Amperometric Mediated Carbon Paste Biosensor Based on D-Fructose Dehydrogenase for the Determination of Fructose in Food Analysis', *Biosens. Bioelectron.*, **12**(12), 1233–1243 (1997).
65. F.W. Scheller, D. Pfeifer, F. Schubert, R. Renneberg, D. Kirsten, 'Application of Enzyme-based Amperometric Biosensors to the Analysis of 'Real' Samples', in *Biosensors: Fundamentals and Applications*, eds. A.P.E. Turner, I. Karube, G.S. Wilson, Oxford University Press, Oxford, 315–346, 1987.
66. D.B. Morgan, M.E. Carver, R.B. Payne, 'Plasma Creatinine and Urea: Creatinine Ration in Patients with Raised Plasma Urea', *Br. J. Med.*, **2**, 929–932 (1977).
67. M. Senda, Y. Yamamoto, 'Urea Biosensor Based on Amperometric Ammonium Ion Electrode', *Electroanalysis*, **5**(9–10), 775–779 (1993).
68. D.S. Papastathopoulos, G.A. Rechnitz, 'A Urea-sensing Membrane Electrode for Whole Blood Measurements', *Anal. Chim. Acta*, **79**, 17–26 (1975).
69. N.F. Sheppard, D.J. Mears, A. Guiseppi-Elie, 'Model of an Immobilized Enzyme Conductimetric Urea Biosensor', *Biosens. Bioelectron.*, **11**(10), 967–979 (1996).
70. M.A.T. Gilmartin, J.P. Hart, 'Novel, Reagentless, Amperometric Biosensor for Uric Acid Based on a Chemically-modified Screen-printed Carbon Electrode Coated with Cellulose Acetate and Uricase', *Analyst*, **119**(5), 833–840 (1994).
71. Y.Q. Zhang, R.A. Gu, J. Zhu, P.G. Cao, 'The Characterization of Uricase-immobilized Silk Fibroin Membrane and its Application to Biosensor for Uric Acid', *Prog. Biochem. Biophys.*, **25**(3), 275–278 (1998).

72. D.B. Morgan, S. Dillon, R.B. Payne, 'The Assessment of Glomerular Function: Creatinine Clearance or Plasma Creatinine?', *Postgrad. Med.*, **54**, 302–310 (1978).
73. T.M. Nguyen, H.G. Monbouquette, 'An Amperometric Creatine/Creatinine Biosensor Based on Immobilized Creatininase, Creatinase, and Sarcosine Dehydrogenase', *Abstr. Pap. Am. Chem. Soc.*, **7**, U194 (1999).
74. G.G. Guilbault, J.P. Czarnecki, M.A. Nabi Rahmi, 'Performance Improvements of Gas-diffusion Ion-selective and Enzyme Electrodes', *Anal. Chem.*, **57**, 2110–2116 (1985).
75. R. Takahashi, I. Tachi, 'Polarographic Studies on Fat-soluble Vitamins', *Agric. Biol. Chem.*, **26**, 771–776 (1962).
76. S.A. Wring, J.P. Hart, D.W. Knight, 'Voltammetric Behavior of All-*trans*-retinol (Vitamin A_1) at a Glassy Carbon Electrode and its Determination in Human Serum Using High-performance Liquid Chromatography with Electrochemical Detection', *Analyst*, **113**, 1785–1789 (1988).
77. T. Vergara, D. Martin, J. Vera, 'Polarographic Determination of Thiamine and its Monophosphate and Pyrophosphate Esters', *Anal. Chim. Acta*, **120**, 347–351 (1980).
78. C. Van Kerchove, R. Bontemps, 'La Vitamine B_1 dans des Solutions Injectables per Polarographic Impulsionelle Différentielle', *J. Pharm. Belg.*, **37**, 169–176 (1982).
79. J. Lindquist, S.M. Farroha, 'Application of Differential Pulse Polarography to the Assay of Vitamins', *Analyst*, **100**, 375–377 (1975).
80. J. Wang, D. Luo, P.A.M. Farias, J.S. Mahmoud, 'Adsorptive Stripping Voltammetry of Riboflavin and Other Flavin Analogues at the Static Mercury Drop Electrode', *Anal. Chem.*, **57**, 158–162 (1985).
81. B. Lequeu, J.C. Guilland, J. Keppling, 'Measurement of Plasma Pyridoxal 5′-Phosphate by Combination of an Enzymatic Assay with High-performance Liquid Chromatography/Electrochemistry', *Anal. Biochem.*, **149**, 296–300 (1985).
82. E. Jacobsen, M.W. Bjornsen, 'Polarographic Determination of Folic Acid in Pharmaceutical Preparations', *Anal. Chim. Acta*, **96**, 345–351 (1978).
83. S. Kwee, 'Electrochemistry of the C-(9)–N(10) Bond in Folic Acid and Related Compounds', *Bioelectrochem. Bioenerg.*, **11**, 467–475 (1983).
84. J. Lankelma, E. van der Kleijn, M.J.Th. Jansen, 'Determination of 5-Methyltetrahydrofolic Acid in Plasma and Spinal Fluid by High-performance Liquid Chromatography Using On-column Concentration and Electrochemical Detection', *J. Chromatogr., B*, **182**, 35–45 (1980).
85. J.C.B. Fernandes, L.T. Kubota, G.D. Neto, 'Potentiometric Biosensor for L-Ascorbic Acid Based on Ascorbate Oxidase of Natural Source Immobilized on Ethylene–Vinylacetate Membrane', *Anal. Chim. Acta*, **385**(1–3), 3–12 (1999).
86. E.J. Kuta, 'Polarographic Investigation of Conjugated Fat-soluble Vitamins', *Science*, **144**, 1130–1131 (1964).
87. J.P. Hart, M.D. Norman, C.J. Lacey, 'Voltammetric Behavior of Vitamin-D_2 and Vitamin-D_3 at a Glassy-carbon Electrode and their Determination in Pharmaceutical Products by using Liquid-chromatography with Amperometric Detection', *Analyst*, **117**(9), 1441–1445 (1992).
88. W.F. Smyth (ed.), *Polarography of Molecules of Biological Significance*, Academic Press, London, 1979.
89. H. Smandke, H. Crohlke, 'Die Polarographische und Photometrische Bestimmung von Tocopheronolecton in Wässriger Lösung unter Verwendung der Dünnschichtchromatographic', *Clin. Chim. Acta*, **11**, 491–494 (1965).
90. H. Takeda, T. Shibuya, K. Yanagawa, H. Kanoh, M. Takasaki, 'Simultaneous Determination of Alpha-tocopherol and Alpha-tocopherolquinone by High-performance Liquid Chromatography and Coulometric Detection in the Redox Mode', *J. Chromatogr., A*, **722**(1–2), 287–294 (1996).
91. J.P. Hart, A.M. Nahir, J. Chayen, A. Caterall, 'Determination of Vitamin K_1 in Plasma by Differential Pulse Polarography and its Possible Chemical Applications', *Anal. Chim. Acta*, **144**, 267–271 (1982).
92. J.P. Hart, M.J. Shearer, P.T. McCarthy, S. Rahim, 'Voltammetric Behaviour of Phylloquinone (Vitamin K_1) at a Glassy-Carbon Electrode and Determination of the Vitamin in Plasma using High-performance Liquid Chromatography with Electrochemical Detection', *Analyst*, **109**, 477–481 (1984).
93. T.Z. Peng, H.P. Li, R.S. Lu, 'Cathodic Stripping Voltammetric Measurements of Cholesterol Following Adsorptive Preconcentration at the Static Mercury Drop Electrode', *Electroanalysis*, **5**(2), 177–181 (1993).
94. M.A.T. Gilmartin, J.P. Hart, 'Fabrication and Characterization of a Screen-printed, Disposable, Amperometric Cholesterol Biosensor', *Analyst*, **119**(11), 2331–2336 (1994).
95. J.C. Vidal, E. Garcia, J.R. Castillo, 'In Situ Preparation of a Cholesterol Biosensor: Entrapment of Cholesterol Oxidase in an Overoxidized Polypyrrole Film Electrodeposited in a Flow System–Determination of Total Cholesterol in Serum', *Anal. Chim. Acta*, **385**(1–3), 213–222 (1999).
96. J. Wang, *Electroanalytical Techniques in Clinical Chemistry and Laboratory Medicine*, VCH Publishers Inc., New York, 1988.
97. S. Constantine, R. Giordano, M. Rizzica, F. Benedetti, 'Applicability of Anodic–Stripping Voltammetry and Graphite Furnace Atomic Absorption Spectrometry for the Determination of Antimony in Biological Matrices: A Comparative Study', *Analyst*, **110**, 1355–1359 (1985).
98. P.H. Davis, F.J. Berlandi, G.R. Dulude, R.M. Griffin, W.R. Matson, E.W. Zink, 'Analysis of Total Arsenic in Urine and Blood by High Speed Anodic Stripping

Voltammetry', *Am. Ind. Hyg. Assoc. J.*, **6**, 480–490 (1978).
99. P. Valenta, H. Rutzel, H.W. Nürnberg, M. Stoeppler, 'Toxic Metals in Biomatrices in Voltammetric Determination of the Trace Content of Cadmium and other Toxic Metals in Human Blood', *Z. Anal. Chem.*, **284**, 25–34 (1977).
100. M. Oehme, W.E. Lund, 'Comparison of Digestion Procedures for the Determination of Heavy Metals (Cd, Cu, Pb) in Blood by Anodic Stripping Voltammetry', *Z. Anal. Chem.*, **298**, 260–268 (1979).
101. T.R. Copeland, J.H. Christie, R.A. Osteryoung, R.K. Skogerboe, 'Analytical Applications of Pulsed Voltammetric Stripping at Thin Film Mercury Electrodes', *Anal. Chem.*, **45**, 2171–2174 (1973).
102. D. Jagner, M. Josefson, S. Westerlund, K. Aren, 'Simultaneous Determination of Cadmium and Lead in Whole Blood and Serum by Computerized Potentiometric Stripping Analysis', *Anal. Chem.*, **53**, 1406–1410 (1981).
103. J. Golimowski, P. Valenta, M. Stoeppler, H.W. Nürnberg, 'A Rapid High-performance Analytical Procedure with Simultaneous Voltammetric Determination of Toxic Trace Metals in Urine', *Talanta*, **26**, 649–656 (1979).
104. J. Golimowski, P. Valenta, M. Stoeppler, H.W. Nürnberg, 'Ein Verfahren zur raschen voltammetrieschen Simultabestimmung der toxischen Spurenmetalle Cd und Pb un Harn', *Z. Anal. Chem.*, **290**, 107–108 (1977).
105. W. Lund, R. Erikson, 'The Determination of Cadmium, Lead and Copper in Urine by Differential Pulse Anodic Stripping Voltammetry', *Anal. Chim. Acta*, **107**, 37–46 (1979).
106. M. Oehme, W. Lunde, J. Jonsen, 'The Determination of Copper, Lead, Cadmium and Zinc in Human Teeth by Anodic Stripping Voltammetry', *Anal. Chim Acta*, **100**, 389–398 (1978).
107. A. Attramadal, J. Jonsen, 'The Contect of Lead, Cadmium, Zinc and Copper in Deciduous and Permanent Human Teeth', *Acta Odontol. Scand.*, **34**, 127–131 (1976).
108. G. Chittleborough, B.J. Steel, 'The Determination of Zinc, Cadmium, Lead and Copper in Human Hair by Differential Pulse Anodic Stripping Voltammetry at a Hanging Mercury Drop Electrode After Nitrate Fusion', *Anal. Chim. Acta*, **119**, 235–241 (1980).
109. J.T. Kinard, 'Diagnosis of Metal Poisoning and Evaluation of Chelation Therapy by Differential Pulse Anodic Stripping Voltammetry Coupled to a Novel Digestion Procedure', *Anal. Lett.*, **10**, 1147–1161 (1977).
110. W.P. Ferren, 'Analyses of Environmental Samples by Means of Anodic Stripping Voltammetry', *Am. Lab.*, **7**, 52–58 (1978).
111. J. Simon, T. Liese, 'Anwendung der Electrothermoleu AAS zur Spurenbestimmung, von Blei und Cadmium Knocheu', *Z. Anal. Chem.*, **314**, 483–486 (1983).
112. S.B. Adelojou, A.M. Bond, M.H. Briggs, 'Assessment of Differential-pulse Adsorption Voltammetry for the Simultaneous Determination of Nickel and Cobalt in Biological Materials', *Anal. Chim. Acta*, **164**, 181–194 (1984).
113. M.A. Schreiber, T.A. Last, 'Computer-controlled Coulastic Stripping Analysis', *Anal. Chem.*, **53**, 2095–2100 (1981).
114. G.M. Schmid, G.W. Bolger, 'Determination of Gold in Drugs and Serum by Use of Anodic Stripping Voltammetry', *Clin. Chem.*, **19**, 1002–1005 (1973).
115. L. Duic, S. Szechter, S. Srinivasan, 'Microdetermination of Lead in Blood: A Derivative Pulse Stripping Voltammetric Method', *J. Electroanal. Chem.*, **41**, 83–93 (1973).
116. T.P. DeAngelis, R.E. Bond, E.E. Brooks, W.R. Heineman, 'Thin-layer Differential Pulse Voltammetry', *Anal. Chem.*, **49**, 1792–1797 (1977).
117. F. Peter, R.G. Reynolds, 'Perchloric Acid-free Digestion of Blood in Microdetermination of Lead by Anodic Stripping Voltammetry', *Anal. Chem.*, **48**, 2041–2042 (1976).
118. G. Morrell, G. Giridhar, 'Rapid Macromethod for Blood Lead Analysis by Anodic Stripping Voltammetry', *Clin. Chem.*, **22**, 221–223 (1976).
119. D. Jagner, L.G. Danielsson, K. Aren, 'Potentiometric Stripping Analysis for Lead in Urine', *Anal. Chim. Acta*, **106**, 15–21 (1979).
120. W. Kemula, Z. Kublik, 'Use of the Mercury Drop Electrode Method for Direct Evaluation of Minute Amounts of Lead in Urine', *Nature (London)*, **189**, 57–58 (1961).
121. D. Jagner, K. Aren, 'Flow Potentiometric Stripping Analysis for Mercury(II) in Urine, Sediment and Acid Digest of Biological Material', *Anal. Chim. Acta*, **141**, 157–162 (1982).
122. J.W. Munson, H. Abdine, 'Determination of Theophylline in Plasma by Oxidation at the Stationary Carbon-paste Electrode', *Talanta*, **25**, 221–222 (1978).
123. S.B. Adelojou, A.M. Bond, H.C. Hughes, 'Determination of Selenium, Copper, Lead and Cadmium in Biological Materials by Differential Pulse Stripping Voltammetry', *Anal. Chim. Acta*, **148**, 59–69 (1983).
124. J.P. Franke, P.M.J. Coengrach, R.A. de Zeeuw, 'A Rapid and Specific Method for the Determination of Traces of Thallium Directly in Urine by Differential Pulse Anodic Stripping Voltammetry', *Arch. Toxicol.*, **34**, 137–143 (1975).
125. D.I. Levit, 'Determination of Traces of Thallium in Urine by Anodic Stripping AC Voltammetry', *Anal. Chem.*, **45**, 1291–1292 (1973).
126. T.R. Williams, D.R. Foy, C. Benson, 'The Determination of Zinc in Human Eye Tissues by Anodic Stripping Voltammetry', *Anal. Chim. Acta*, **75**, 250–252 (1975).
127. J.J. Kulys, M.V. Pesliakiene, A.S. Samalius, 'The Development of Bienzyme Glucose Electrodes', *Bioelectrochem. Bioenerg.*, **8**, 81–88 (1981).

128. R.M. Ianniello, A.M. Yacynych, 'Immobilized Enzyme Chemically Modified Electrode as an Amperometric Sensor', *Anal. Chem.*, **53**, 2090–2095 (1981).
129. Z.J. Zhang, Z.L. Gong, W.B. Ma, 'An Enzyme-based Fiber Optic Biosensor for determination of D-Amino Acid in Serum', *Microchem. J.*, **52**(2), 131–138 (1995).
130. T.J. O'Shea, P.L. Weber, B.P. Bammel, C.E. Lunte, S.M. Lunte, M.R. Smyth, 'Monitoring Excitatory Amino-acid Release In Vivo by Microdialysis with Capillary Electrophoresis Electrochemistry', *J. Chromatogr.*, **608**(1–2), 189–195 (1992).
131. T.E. Campbell, A.J. Hodgson, G.G. Wallace, 'Incorporation of Erythrocytes into Polypyrrole to form the Basis of a Biosensor to Screen for Rhesus (D) Blood Groups and Rhesus (D) Antibodies', *Electroanalysis*, **11**(4), 215–222 (1999).
132. B. Lu, M.R. Smyth, R. O'Kennedy, J. Moulds, T. Frame, 'Development of an Amperometric Immunosensor Based on Flow Injection Analysis for the Detection of Red Blood Cells', *Anal. Chim. Acta*, **340**(1–3), 175–180 (1997).
133. J.G. Quinn, R. O'Kennedy, M.R. Smyth, J. Moulds, T. Frame, 'Detection of Blood Group Antigens Utilising Immobilised Antibodies and Surface Plasmon Resonance', *J. Immunol. Methods*, **206**(1–2), 87–96 (1997).
134. Z. Junhui, C. Hong, Y. Ruifu, 'DNA Based Biosensors', *Biotech. Advances*, **15**, 43–58 (1997).
135. K.M. Millan, A. Saraullo, S.R. Mikkelsen, 'Voltammetric DNA Biosensor for Cystic Fibrosis Based on a Modified Carbon Paste Electrode', *Anal. Chem.*, **66**, 2943–2948 (1994).
136. K.M. Millan, S.R. Mikkelsen, 'Sequence Selection Biosensor for DNA Based on Electroactive Hybridization Indicators', *Anal. Chem.*, **65**, 2317–2324 (1993).
137. K.M. Millan, A.J. Spurmanis, S.R. Mikkelsen, 'Covalent Immobilization of DNA onto Glassy Carbon Paste Electrodes', *Electroanalysis*, **4**, 929–932 (1992).
138. J. Wang, M. Ozsoz, X. Cai, G. Rivas, H. Shiraishi, D.H. Grant, M. Chicharro, J. Fernandes, E. Paleček, 'Interactions of Antitumor Drug Daunomycin with DNA in Solution and at the Surface', *Bioelec. Bioenerg.*, **45**, 33–40 (1998).
139. J.M. Zen, Y.S. Ting, 'Simultaneous Determination of Caffeine and Acetaminophen in Drug Formulations by Square-wave Voltammetry using a Chemically Modified Electrode', *Anal. Chim. Acta*, **342**(2–3), 175–180 (1997).
140. N. Larsson, T. Ruzgas, L. Gorton, M. Kokaia, P. Kissinger, E. Csoregi, 'Design and Development of an Amperometric Biosensor for Acetylcholine Determination in Brain Microdialysates', *Electrochim. Acta*, **43**(23), 3541–3554 (1998).
141. E.N. Claney, R.P. Baldwin, 'Electrochemical Determination of Adriamycin Compounds in Urine by Preconcentration at Carbon Paste Electrodes', *Anal. Chem.*, **54**, 2556–2560 (1982).
142. M.A. Brooks, M.R. Hackman, 'Trace Level Determination of 1,4-Benzodiazepines in Blood by Differential Pulse Polarography', *Anal. Chem.*, **47**, 2059–2062 (1975).
143. K. Sarna, Z. Fijalek, 'Voltammetric and Electrochemical Quartz Crystal Microbalance Study of Antithyroid Drugs', *Chemia Analityczna*, **42**(3), 425–433 (1997).
144. N.A. Elmaali, A.M.M. Ali, M. Khodari, M.A. Ghandour, 'Cathodic Stripping Voltammetric Determination of the Cephalosporin Antibiotic Ceftriaxone at the Mercury-electrode in Aqueous and Biological Media', *Bioelectrochem. Bioenerg.*, **26**(3), 485–492 (1991).
145. S. McNiven, M. Kato, R. Levi, K. Yano, I. Karube, 'Chloramphenicol Sensor Based on an In Situ Imprinted Polymer', *Anal. Chim. Acta*, **365**(1–3), 69–74 (1998).
146. M.A. Brooks, J.A.F. de Silva, T.S. Ma, 'Determination of 2,4-Diamino-5-(3,4,5-trimethoxybenzyl)-pyrimidine (Trimethoprim) in Blood and Urine by Differential Pulse Polarography', *Anal. Chem.*, **45**, 263–266 (1973).
147. T.B. Jarbawi, W.R. Heineman, G.J. Patriarche, 'Thin-layer Differential Pulse Voltammetric Determination of Chlorpomazine in Biological Fluids', *Anal. Chim. Acta*, **126**, 57–64 (1981).
148. A. Webber, M. Shah, J. Osteryoung, 'Electrochemical Reduction and Determination of Cimetidine at Nanomolar to Micromolar Levels of Concentration', *Anal. Chim. Acta*, **154**, 105–119 (1983).
149. Y.H. Xu, A.A. Suleiman, 'Reusable Amperometric Immunosensor for the Determination of Cortisol', *Anal. Lett.*, **30**(15), 2675–2689 (1997).
150. S.S. Hassan, M.M. Amer, S.A. Abd El-Fatah, A.M. El-Kossay, 'Microcoated Wire Sensors for the Determination of Anticancer Drugs Cyclophosphamide and Ifosphamide in the Presence of their Degradates', *Talanta*, **46**(6), 1395–1402 (1998).
151. S.A. Akman, F. Kusus, K.T. Takamura, R. Chlebowski, J. Block, 'Differential Pulse Polarographic Determination of Plasma Menadione', *Anal. Biochem.*, **141**, 488–493 (1984).
152. D.J. Pritchard, H. Morgan, J.M. Cooper, 'Simultaneous Determination of Follicle-stimulating Hormone and Luteinizing Hormone using a Multianalyte Immunosensor', *Anal. Chim. Acta*, **310**(2), 251–256 (1995).
153. S.S.S.M. Hassan, W.H. Mahmoud, M.S. Abdel-Samad, 'Direct Potentiometry and Potentiotitrimetry of Warfarin and Ibuprofen in Pharmaceutical Preparations using PVC Ferroin-based Membrane Sensors', *Mikrochim. Acta*, **129**(3–4), 251–257 (1998).
154. T. Ohkubo, T. Nambara, 'Determination of Mitomycin-C in Serum by High-performance Liquid-chromatography with Dual-electrode Coulometric Detection', *J. Chromatogr., B*, **527**(2), 441–446 (1990).
155. S.S. Hu, M.D. Guo, G.Z. Hu, M. Jiang, 'Electrochemical Oxidation and Analytical Applications of Estrogens at

Modified Carbon-paste', *Anal. Lett.*, **28**(11), 1993–2003 (1995).
156. E. Leszczynska, R. Koncki, S. Glab, 'Biosensors with Immobilized Penicillin Amidohydrolase and Penicillinase for Determination of Beta-lactam Antibiotics', *Chemia Analityczna*, **41**(5), 839–844 (1996).
157. J. Wang, G. Rivas, X.H. Cai, H. Shiraishi, P.A.M. Farias, N. Dontha, D.B. Luo, 'Accumulation and Trace Measurements of Phenothiazine Drugs at DNA-modified Electrodes', *Anal. Chim. Acta*, **332**(2–3), 139–144 (1996).
158. R.W. Claycomb, M.J. Delwiche, 'Biosensor for On-line Measurement of Bovine Progesterone During Milking', *Biosens. Bioelectron.*, **13**(11), 1173–1180 (1998).
159. C. Martin, E. Dominguez, 'A New Enzyme Electrode for Quantification of Salicylic Acid in a FIA System', *J. Pharm. Biomed. Anal.*, **19**(1–2), 107–113 (1999).
160. R. Wintersteiger, M.J. Sepulveda, 'Electrochemical Detection of Anabolics in Human Plasma and Urine', *Anal. Chim. Acta*, **273**(1–2), 383–390 (1993).
161. J. Wang, E. Dempsey, M. Ozsoz, M.R. Smyth, 'Amperometric Enzyme Electrode for Theophylline', *Analyst*, **116**(10), 997–999 (1991).
162. Z.H. Liu, G.L. Shen, Q. Miao, R.Q. Yu, 'A Thyroid-stimulating Hormone Immuno-electrode', *Anal. Chim. Acta*, **325**(1–2), 87–92 (1996).
163. M.A. Brooks, J.A.F. de Silva, L.M. D'Arconte, 'Determination of 2,4-Diamino-5-(3,4,5-trimethoxybenzyl)-pyrimidine (Trimethoprim) in Blood and Urine by Differential Pulse Polarography', *Anal. Chem.*, **45**, 263–266 (1973).
164. A. Temizer, 'Electroanalytical Determination of Vinca Alkaloids Used in Cancer Immunotherapy', *Talanta*, **33**, 791–794 (1986).

Electroanalytical Chemistry in Clinical Analysis

Evangeline Y. Su and Andrew B. Bocarsly
Princeton University, Princeton, USA

1 Introduction **1**
1.1 Current Electrochemistry in Clinical Application 1
1.2 Electrochemical Methods 2
2 Selected Clinical Applications in Development Categorized by Mechanism **3**
2.1 Enzyme Electrodes 4
2.2 Mediators or Mediated Sensors 4
2.3 Ion-selective Electrodes 5
2.4 Chemically Modified Electrodes 5
3 State of the Art: Systems Available for Clinical Usage **5**
Abbreviations and Acronyms **5**
Related Articles **6**
Further Reading **6**
General Information 6
Immunoassays and Immunosensors 6
Glucose 7
References **7**

Clinical electrochemistry has historically been limited to the detection of blood gases, physiological pH, and alkali metal cations using ion-selective electrodes (ISEs). In recent years, current flow techniques have been introduced allowing for the detection and quantification of a variety of proteins and metabolites. Key to this type of analysis is the development of an electrode interface which is both catalytic and selective for the analyte of interest. To this end, a wide variety of chemically modified electrode (CME) interfaces have been developed. Such interfaces are often based on the interaction of a redox enzyme with the analyte of interest. In addition, interfaces composed of immunospecific elements or size specific elements have been examined as selective electrode surfaces.

1 INTRODUCTION

Electrochemistry is inherently an applied part of biological studies, since all living organisms operate on a very basic level electrochemically. Using electrodes, the clinician can mimic the charge transfers between molecules within the human body. Theoretically, this mimicry can be used in two ways. The first capacity is potentially to fix what *should* be happening within the human body, but is not, as in the case of a diabetic, whose pancreas is not monitoring and compensating for blood sugar (glucose) within levels which do not shock the body. The second, slightly less ambitious, usage is to monitor what is there, for example, to monitor whether a certain level of potassium ions is present in blood. Many sensors designed for the latter purpose are used today in the clinical setting to test samples in vitro.

In the past 45 years, the major areas for electrochemical analysis were blood gas detection[1,2] and blood and urine electrolytes. Existing state of the art electrochemistry deals with blood metabolites and blood enzymes. Preclinical work for the next generation of electrochemical analysis devices focuses on blood hormones and proteins. With the introduction of the Clark electrode in 1956 and the subsequent improvement in the area of blood gas detection with the Severinghaus electrode in 1958, the analysis of blood gases matured into a now standard clinical laboratory test. Blood and urine electrolytes, however, continue to be refined. From the early 1970s, ISEs have been commercialized for use in hospitals, both in their laboratories and at the patient's bedside. In fact, one in vivo usage of such electrochemical devices is a miniature pH electrode that is used to monitor fetal scalp pH during delivery in hospitals. This area has been most lucrative, since the potentiometric analysis can be achieved at a relatively low cost and is easy to run with little electrochemical knowledge. Blood metabolites, i.e. glucose, urea and creatinine, have followed as the next advancing field.[3,4] Much of current research in this area is devoted to creating enzyme electrodes to detect blood metabolites, though it is mostly in vitro processes. Recent research, however, has moved towards working with in vivo applications for electrochemistry. Much of this new technology is not yet ready for widespread clinical use, as many researchers have yet to successfully address the difficulty of creating an electrode that is dependably functional for extended periods of time inside the human body. Beyond that, the field of immunoassays for detecting hormones, drugs, and proteins has also expanded. Since electrodes designed for immunoassays are relatively far from commercialization for the clinical setting, we will only provide a short reading list for those who would like to cultivate a beginner's interest.

1.1 Current Electrochemistry in Clinical Application

Electrochemical analysis has enjoyed the reputation of being highly sensitive, highly selective, and in many cases where it has been commercialized, fairly inexpensive as well as fairly easy to run. An ideal electrochemical system

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

for clinical applications requires portability, low cost, ease of operation, and both high accuracy and high sensitivity. The ideal working electrode is designed to maximize selectivity. For commercialization, it should be made of cheap, stable and nontoxic materials, be easy to mass produce, and be long-lived.

Electrodes for in vitro measurements in general are made of electrochemically inert materials, i.e. platinum, carbon paste, and glassy carbon. Recent technology finds that these materials are useful also in the microfabrication of electrodes for clinical applications. (We define clinical as for use with and within humans in the medical setting, rather than encompassing drug research studies that have also been conducted with electrochemical methods and electrodes.) The standard use of electrochemistry in the clinical setting is for potentiometric measurements, associated with ISE. In contrast to the current state of clinical electrochemistry, the field is moving towards the development of reliable in vivo electrodes, the further development of cyclic voltammetry and amperometry in clinical applications, and the development of electrodes that can handle multiple analyses. With the advent of new technologies in other fields, electrochemical applications have been further enhanced for reliable clinical use. For example, in several cases, silicon-based microfabrication has been used to produce and easily reproduce a higher quality electrochemical system of miniature electrodes and other associated components, as will be discussed later.

Much has been written on pre-clinical electrochemical sensors. J. Wang wrote one of the most comprehensive recent journal reviews with useful references.[5] In recent years, the focus has been on glucose sensors, some of which are now commercially available. The majority of glucose sensors work on the following premise. The sensor is an amperometric electrode, which is in electrical communication through various kinds of media – solutions of mediating electrolytes and the like – with the glucose oxidase (GOX). Each iteration of the glucose interacting with the GOX cycles another oxidation–reduction between the enzyme and either the sensor surface or a mediating substance between the GOX and the sensor. The commercially available system developed by Lucisano et al. is hooked up to an intravenous (IV) tubing system or an intra-arterial catheter, which withdraws the blood, analyzes it and then reinfuses the blood into the patient.[6] Similar systems have been developed.[7] (See the glucose papers cited under further reading and their references for more literature specifically on glucose sensors.)

Much recent interest has been directed towards the creation of an electrochemical biosensor-based artificial pancreas. The idea of in vivo electrodes sensors that can detect, correct, and strictly control the rise and drop of blood sugar is a tantalizing goal. Adam Heller's group published a continuing series of papers culminating to date in 1998 with a manuscript on the ''continuous amperometric monitoring of glucose in a brittle diabetic chimpanzee with a miniature subcutaneous electrode''.[8] The Heller group continues to forge on with their ''wiring'' of enzyme redox centers to electrodes through electron-conducting selectively permeable hydrogels to develop a system for continuous glucose monitoring. Furthermore, Heller's group studied and reported their findings on the ''measurement and modeling of the transient difference between blood and subcutaneous glucose concentrations in the rat after injection of insulin'',[9] in order to further hone their electrodes, which they envision inserted in the subcutaneous layer of skin rather than somewhere directly in the path of blood flow. Many other groups are exploring different means in this lucrative area of research.[10] For example, an artificial pancreas would require something to trigger an injection of insulin, such as the hypoglycemic alarm connected to a subcutaneous glucose sensor.[11] For a more comprehensive look at the advances being made, see the literature cited below in further reading.

1.2 Electrochemical Methods

The most common electrochemical method for clinical use is potentiometry. It stands to reason that this would be naturally obvious as potentiometry simply requires a voltmeter to measure the voltage across the cell and does not require a power supply to apply a voltage, so it is easily portable. That portability makes a potentiometric device easy to incorporate into on-the-spot clinical examinations. The pre-clinical electrodes use amperometry, potentiometry, and in a limited number of cases voltammetry and conductometry. Amperometry, the measurement of electrical current over a range of electrode potential, has quickly become the next electrochemical method of choice for the development of new electrodes for clinical use. Though amperometry requires a power source to supply a potential to the system, it still is easy to run and is the most common electrochemical method applied for enzyme electrodes. Conductometric studies, which involve the direct measurement of an analyte's conductance, are not common, but have been used, for example, in the determination of hematocrit concentrations within a blood sample. The conductivity is measured, corrected for the base electrolyte concentration, and can be related to hematocrit, HCO_3, and hemoglobin, among other things. (According to the product literature, the i-STAT system incorporates one such electrode into its array of electrodes for the study of hematocrit.) Various types of voltammetry, e.g. pulse voltammetry, cyclic voltammetry, adsorptive stripping voltammetry, are employed for assays designed

for in vitro applications within a laboratory setting rather than in the clinic.[12] Because voltammetry is a more complex technique, fewer pre-clinical devices have been designed for voltammetric use. It should be noted carefully though that some of the literature describes amperometric and potentiometric biosensors together as amperometric electrodes.

In a specific case, glucose electrodes are perhaps by far the most common of the metabolite and enzyme electrodes. In general, glucose measurements are done amperometrically or voltammetrically. The electrode system selectively oxidizes glucose using GOX, which is either in solution or directly attached via sol–gel, organic compounds, or in inorganic matrices to the electrode. This oxidation produces hydrogen peroxide in systems containing O_2, while in anaerobic systems a mediator is used to carry the charge to the sensing electrode. The H_2O_2 or mediator is oxidized at the working electrode to measure a current that is proportional to the amount of glucose. With any given amperometric analysis of this type, the anodic or cathodic current is related in a complex manner, subject to artifacts to the concentration of the analyte. Pragmatically, the current is proportional to the amount of analyte so long as the H_2O_2 and reduced mediating substrates are produced through a reaction with the glucose and glucose enzyme, which is also dependent upon the partial pressure of O_2 in the system. All the resulting complications with such analytical systems as the glucose sensors have been the subject of research involving CMEs, as well as selective membrane and sol–gel electrode systems (among other types) for the past 20 years.

ISEs, the most widely used electrode type in current clinical applications, are potentiometric in nature. Glass pH electrodes and electrodes that detect several different kinds of analytes, including alkali metal ions and sodium, ammonium, and potassium ion activity, employ potentiometry, which relies on the measurement of a potential induced by a chemical gradient. The calculations for metabolite concentrations are done by solving the Nernst equation (Equation 1) based on the potential measured (E_R), the known standard redox potential (E_R^o), and the stoichiometric number of electrons (n):

$$E_R = E_R^o + \frac{2.303RT}{nF} \log \left(\frac{a_{ox}}{a_{red}} \right) = E_R^o + \frac{0.059}{n} \log \left(\frac{a_{ox}}{a_{red}} \right) \quad \text{at (room temp.)} \quad (1)$$

Typically, activities for the oxidized and reduced analyte (a_{ox} and a_{red}) are assumed to be equal to the concentrations of these species. This technique is very sensitive. It does, however, tend to provide a slow instrument response. Additionally, the necessity of a membrane system (to produce a selective chemical gradient) makes the potentiometric sensor subject to poisoning and mechanical failure. The measurement is also relatively sensitive to the temperature of the system, as indicated by Equation (1), and thus the temperature must be closely regulated or compensated for.

Impedance-based sensors have been suggested as an alternative to potentiometric detection. Though not yet in clinical use, such a sensor has been produced for the determination of urea and leaves open possibilities for other metabolites. Ball reported in 1995, ''electrochemical sensors based on impedance measurement of enzyme-catalyzed polymer dissolution: theory and applications''.[13]

Classically, different types of electrodes have been used with these different methods, for the most part almost all of those types have appeared at some point or another in the literature written about pre-clinical electrode prototypes. Dropping mercury electrodes and static mercury drop electrodes are used in some clinical laboratory settings, though not in clinics themselves.[14] Rotating disk electrodes, CMEs, semiconductor electrodes, thin-film membrane electrode systems, glassy carbon electrodes, carbon paste electrodes, and platinum and gold plate electrodes have all been reported in the literature from the past 15 years.

2 SELECTED CLINICAL APPLICATIONS IN DEVELOPMENT CATEGORIZED BY MECHANISM

Since the two major areas of clinical electrochemical analysis are focused on blood metabolites using enzyme detection and developing better systems for blood electrolyte detection, this review focuses on those topics. However, as immunoassays for proteins, drugs and hormones have truly become the future of electrochemistry, a set of readings is provided under Further Reading. Electrochemical sensors take advantage of the following natural and artificially stimulated redox reactions (Equations 2 and 3)

$$\text{Substrate} + \text{enzyme} \longrightarrow \text{enzyme}^* + \text{Product A} + \text{Product B} + \cdots \quad (2)$$

$$\text{enzyme}^* + \text{Reagent} \longrightarrow \text{enzyme} + \text{Reagent}^* \quad (3)$$

where enzyme* is the oxidized (or reduced) state of the enzyme and Reagent is an added redox mediator which can be reversibly oxidized (or reduced) to Reagent*.

An electrochemical sensor used in clinical settings generally is a two-part probe. One part is the actual working electrode that conducts the current from the

analytes. The other part is the chemical mechanism that provides the system's selectivity. This second part is often a biological component such as an enzyme, DNA, or hemoglobin. When the sensor has a biological component involved in its system, it is called a biosensor or a bioelectrode. The word biosensor is often alternatively used in the field to describe anything that senses biological materials or ions. This can often lead to some confusion about electrodes being reported in the literature. For the purposes of this article, however, biosensor will refer only to electrodes or sensors in which an enzyme or some other biological material is employed.

Enzyme electrodes work with the chemical mechanisms shown in Equations (2) and (3). There are two types. One type of enzyme electrode system is designed to detect one of the products in Equation (2); in the case of glucose sensors product 1 is H_2O_2. For that first type of electrode the chemical mechanism ends there. The other type of enzyme electrode works in the same way as mediated sensors do, by detecting the reduced or oxidized reagent from step 2, but it is a combination of step 2 followed by step 3. ISEs, on the other hand, are potentiometric in nature, with the electrode simply sensing a chemical gradient established by solution ions across a permselective membrane.

2.1 Enzyme Electrodes

Enzyme electrodes are used mostly to detect and quantify metabolites, such as glucose, sarcosine oxidase, lactate, and cholesterol. They are made in several different ways. For the most part, though, the enzyme is attached to the electrode in various ways (i.e. chemically modified surfaces, thin poly(vinyl chloride) (PVC) films, sol–gels),[15–21] so that as soon as the enzyme oxidizes its target, the electrode can detect the electroactive product of that interaction. The current that is created is measured amperometrically for most of these biosensors. Some biosensors have been developed to analyze more than one electroactive species:[22,23] three-enzyme amperometric biosensors and dual analyte devices for creatinine and urea. These include the disposable glucose sensor plate and the disposable multi-ion sensor plate from above also.[24] In fact, most of the research done with enzyme electrodes has focused on GOX detection and monitoring.[25–27]

David Hage's review article in *Analytical Chemistry*[28] cites many useful articles for investigating electrochemical applications for immunoassays, which use electrochemical detection techniques with biosensors. Though immunoassays cover a wide range of electrochemical possibilities,[29] several labeled experiments are of particular interest. In the immunoassays cited by Hage, enzymes are used as labels to generate an electroactive product. Such electrode combinations can be used to detect a prostate-specific antigen[30] and *Helicobacter pylori*-specific IgG antibodies.[31] Several other related methods use hemoglobin,[32] copper ions[33,34] and iodinated erythrosin B[35] as the catalyst or the catalytic label in their immunoassays.

One of the major problems with enzyme electrodes, however, is their instability and 'poisoning' of the electrode after a limited time period. Some researchers have chosen to address this issue by creating sensors that are disposable after one use in the clinic. Unfortunately, for something like an artificial pancreas that would be internal, such problems can not be so readily addressed. Another one of the major difficulties in using the current enzyme electrodes is that often other electrolytes in solution will interfere with the electrode and produce an error in the reading. A lot of work has been done to research the use of sol–gels, PVC membrane, and cellulose acetate (CA) membranes to either reduce interference and selectively take up the desired analytes.[36–38]

2.2 Mediators or Mediated Sensors

Mediated sensors employ a chemical and an electrochemical mechanism, though not always in that order, to detect their targeted analytes. Often the mediating compounds act as the go-between from the electrode to the targeted analyte, without ever being a part of the overall process as a product or a reactant. Mediators can either be in the electrolyte solution or on the surface of a CME. In either case, they act as a redox shuttle between the electrode and a redox component. The electrocatalytic charge-transfer process works as indicated in Equation (4).

$$\text{Analyte}_{\text{reduced}} + \text{Mediator}_{\text{oxidized}} \longrightarrow \text{Analyte}_{\text{oxidized}} + \text{Mediator}_{\text{reduced}} \qquad (4)$$

$\text{Mediator}_{\text{reduced}}$ gives up the e^- at the electrode and cycles back out as $\text{Mediator}_{\text{oxidized}}$.

In some cases there is a parallel enzyme cycle added in between the analyte and mediator. Note however that the catalytic mediator is regenerated and ready to react with another molecule.[39]

Organometallic and organic mediators, including but not limited to ferrocenes, phthalocyanines, metalloporphyrins, phenoxazines, phenathiazines, phenazines, quinones, tetrathiofulvalene (TTF) and tetracyanoquinodimethane (TCNQ), have been reviewed as interfacial species for chemically modified, carbon-based electrodes to be used in the analysis of biologically active compounds.[42]

Some of the mediator sensors also employ a combination approach and use enzymes in addition to the

mediator,[40] like Belanger, Nadreau, and Fortier's rotating ring disk electrode.[41] This electrode is made of a platinum disk coated with a polypyrrole–glucose oxidase (Ppy–GOX) film, leaving the ring bare. The polypyrrole acts as the mediator and the GOX is the enzyme. Using an amperometric current, the GOX oxidizes the glucose that is present, is in turn oxidized by the polypyrrole film, which then transfers the electron to the platinum electrode, where the current is measured and calculated in proportion to the concentration of glucose. Since these electrodes are employed like enzyme electrodes, indeed the difference is simply that mediated electrode systems encompass a greater number of analytes than enzyme electrodes. Such electrodes are also used in immunoassays[42] and in DNA biosensors[43] The latter is not, however, ready for use in clinical settings. In general, mediated electrodes tend to be amperometric sensors.

2.3 Ion-selective Electrodes

ISEs are most commonly used in the clinical setting, they are used to monitor electrolytes vital to human survival, i.e. potassium, calcium, sodium, chloride. Two electrolytes in particular have received more recent attention, since most of the common electrode systems for detecting the same electrolytes have already become commercialized. Sensors for detecting blood magnesium and lithium require a higher degree of selectivity than the sensors previously developed, because of interference from competing ions like calcium. Huijgen et al. critiqued three commercial magnesium-selective electrodes.[44] Many of the ion-selective sensors are potentiometric sensors. In fact, most of the miniature sensors in the i-STAT system are potentiometric. Though one notable forerunner of these sensors, the Clark electrode, is an amperometric membrane electrode system that uses a gas permeable membrane to analyze PO_2 levels.[2–4] Present research is directed at the employment of polymer membranes,[45,46] nitric-oxide-releasing films,[47] and sol–gel membranes,[48] in conjunction with ISEs.

2.4 Chemically Modified Electrodes

Recent research has involved the design of CMEs, which are capable of being both mediating electrodes and ISEs. In fact, in one case, the same chemically modified surface was used for both an ion detection in human whole blood and for the electroanalytical detection of glucose.[49,50] The former application was as an ISE and the latter as a mediating electrode. CMEs are used with both the amperometric and the potentiometric technique in order to detect metabolites, such as glucose and urea.[51] In fact, the example of a voltammetric electrode system for detecting glucose (cited above) uses a CME. Chang and Bae designed their CME without a mediating matrix chemically attached to it. They use a thin film in conjunction with their CME to successfully detect glucose by voltammetry.[16] CMEs have been found to be versatile and are also being designed for immunoassays to detect proteins, sugars, hormones and the like.[52–55]

3 STATE OF THE ART: SYSTEMS AVAILABLE FOR CLINICAL USAGE

The i-STAT system, one of a few commercially available, electrochemical blood analysis units, is a portable, multi-electrode unit that uses miniature thin film electrodes. The i-STAT blood sample is placed in a cartridge by capillary action, where it is held in a reservoir for analysis. The system applies a variety of electrochemical methods to detect pH, PCO_2, PO_2, glucose, creatinine, sodium, potassium, chloride, bicarbonate, calcium, urea nitrogen, hematocrit, and various calculated parameters with a series of miniature electrodes. This battery-driven, hand-held device takes 90 seconds to give a health care provider the blood analysis information, without him/her ever having to leave the patient's bedside. The i-STAT system involves a microfluidic cartridge, much like, but not the same as, the technology reported by Boyd et al. in their patent application,[56] which operates with a sample delivery system by capillary action similar to the one described by Foster et al. in their patent application.[57]

The i-STAT system is not the only one of its kind, as the patent applications and the recent publications reveal a pattern that other similar systems are certainly on the market or soon to be arriving there,[58–61] like the ExacTech disposable glucose sensor for the personal monitoring of whole blood glucose concentrations in diabetics.[62–64] Commercial analysis systems that use electrochemical detection techniques are available for sensing analytes in whole blood, cholesterol, carbohydrates, Po_2, body fat, glucose, creatinine, transaminase, potassium, sodium, calcium, and other nutritional minerals in human body fluids. Because many of these systems are considered trade secrets, it is difficult to identify the actual electrochemical designs for many of the electrochemical instruments currently in clinical use.

ABBREVIATIONS AND ACRONYMS

CA	Cellulose Acetate
CME	Chemically Modified Electrode
GOX	Glucose Oxidase
ISE	Ion-selective Electrode
IV	Intravenous

Ppy–GOX	Polypyrrole–Glucose Oxidase
PVC	Poly(vinyl chloride)
TCNQ	Tetracyanoquinodimethane
TTF	Tetrathiofulvalene

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors • Voltammetry In Vivo for Chemical Analysis of the Living Brain • Voltammetry In Vivo for Chemical Analysis of the Nervous System

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Automation in the Clinical Laboratory • Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electrolytes, Blood Gases, and Blood pH • Glucose Measurement • Immunochemistry • Laboratory Instruments in Clinical Chemistry, Principles of • Micro Total Analytical Systems in Clinical Chemistry

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Ion-selective Electrodes: Fundamentals • Liquid/Liquid Interfaces, Electrochemistry at • Pulse Voltammetry • Selective Electrode Coatings for Electroanalysis

FURTHER READING

General Information

Annals of Biomedical Engineering – Special Issue: Bioelectrodes, **20**(3), Blackwell Science, Inc., Malden, 1992.

C.M.A. Brett, A.M.O. Brett, *Electrochemistry: Principles, Methods, and Applications*, Oxford University Press, New York, 1993.

V.V. Cosofret, R.P. Buck, 'Recent Advances in Pharmaceutical Analysis with Potentiometric Membrane Sensors', *Crit. Rev. Anal. Chem.*, **24**(1), 1–58 (1993).

J.R. Flores, E. Lorenzo, 'Amperometric Biosensors', *Wilson and Wilson's Comprehensive Analytical Chemistry*, Elsevier, New York, Chapter 8: 531–570, Vol. 27, 1982.

J.E. Frew, M.J. Green, H.A.O. Hill, 'New Electrochemical Techniques Applied to Medicine and Biology', *J. Chem. Technol. Biotechnol.*, **36**(8), 357–363 (1986).

J.P. Hart, 'Applications of Modern Voltammetric Techniques in Clinical Chemistry', *Anal. Chem. Symp., Ser. 25 (Electrochem., Sens. Anal.)*, 355–366 (1986).

A. Heller, 'Amperometric Biosensors', *Curr. Opin. Biotech.*, **7**(1), 50–54 (1996).

E.I. Iwuoha, M.R. Smyth, 'Polymer-based Amperometric Biosensors', in *Electroactive Polymer Electrochemistry Vol. 2*, ed. M.E.G. Lyons, Plenum Press, New York (Chapter 11), 297–328, 1996.

J. Koryta, M. Brezina, J. Pradac, J. Pradacova, 'Methods for Electroanalysis In Vivo', in *Electroanalytical Chemistry*, 85–140, Vol. 11, 1979.

I. Kostov, F. Filipov, 'Bioelectronics – Status and Prospectives', *Biotechnology*, **EQ13**(1), 81–90 (1999).

U.J. Krull, D.P. Nikolelis, 'Biosensors: A State of Art in Chemistry', *Chem. Chron.*, **22**(2), 91–107 (1993).

M. Lambrechts, W. Sansen, *Biosensors: Microelectrochemical Devices*, IOP, Bristol, UK, 1992.

M. Mascini, G.G. Guilbault, 'Clinical Uses of Enzyme Electrode Probes', *Biosensors*, **2**(3), 147–172 (1986).

M.E. Meyerhoff, B. Fu, E. Bakker, J.H. Yun, V.C. Yang, 'Poly Ion-sensitive Membrane Electrodes for Biomedical Analysis', *Anal. Chem.*, **68**(5), A168–A175 (1996).

D. Moscone, M. Mascini, *Optimized Biosensors in Clinical Applications*, Handb. Biosens. Electron. Notes, Conference, 409–434, 1997.

G. Ramsay (ed.), *Commercial Biosensors*, Wiley, New York, 1998.

K. Rogers, A. Mulchandani, W. Zhou (eds.), *Biosensor and Chemical Sensor Technology: Process Monitoring and Control*, ACS Symposium Series 613, 1995.

R.L. Solsky, 'Ion-selective Electrodes', *Anal. Chem.*, **60**(12), 106R–113R (1988).

G. Urgan, *Microelectronic Biosensors for Clinical Applications*, Handb. Biosens. Electron. Notes, Conference, 257–277, 1997.

J. Wang, 'Electroanalysis and Biosensors', *Anal. Chem.*, **71**(12), 328R–332R (1999).

O.S. Wolfbeis, R. Reisfeld, I. Oehme, 'Sol–gels and Chemical Sensors', *Struct. Bonding (Berlin)* **85** (Optical and Electronic Phenomena in Sol–Gel Glasses and Modern Application), 51–98 (1996).

Immunoassays and Immunosensors

See also references in this paper

C.M. Duan, M.E. Meyerhoff, 'Separation-free Sandwich Enzyme Immunoassays using Microporous Gold Electrodes and Self-assembled Monolayer Immobilized Capture Antibodies', *Anal. Chem.*, **66**(9), 1369–1377 (1994).

A. La Gal La Salle, B. Limoges, C. Degrand, P. Brossier, 'Enzyme Immunoassays with an Electrochemical Detection Method using Alkaline Phosphatase and a Perfluorosulfonated Ionomer-modified Electrode. Application to Pheytoin Assays', *J. Anal.Chem.*, **67**(7), 1245–1253 (1995).

H. Maupas, A.P. Soldatkin, C. Martelet, N. Jaffrezic-Renault, B. Mandrand, 'Direct Immunosensing Using Differential Measurements of Impedimetric Variations', *J. Electroanal. Chem.*, **421**(1–2), 165–171 (1997).

P. Skladal, 'Advances in Electrochemical Immunosensors', *Electroanalysis*, **9**(10), 737–745 (1997).

J. Wang, B. Tian, K.R. Rogers, 'Thick-film Electrochemical Immunosensor Based on Stripping Potentiometric Detection of a Metal Ion Label', *Anal. Chem.*, **70**(9), 1682–1685 (1998).

Glucose

E. Csoregi, C.P. Quinn, D.W. Schmidtke et al., 'Design, Characterization, and One-point in vivo Calibration of a Subcutaneously Implanted Glucose Electrode', *Anal. Chem.*, **66**(19), 3131–3138 (1994).

G. Kenausis, C. Taylor, A. Heller, 'Glucose and Lactate Biosensors Based on Oxidases ''Wired'' Within Hydrogels Made With Polyvinyl Pyridine and Polyvinyl Imidazole Complexed with [(Os-4,4′-dimethoxy 2,2′-bipyridine)Cl](+/2+)', *Abstr. Pap. Am. Chem. Soc.*, **213**(308-PMSE Part 2), (1997).

U. Fischer, K. Rebrin, T. Woedtke, P. Abel, 'Clinical Usefulness of the Glucose Concentration in Subcutaneous Tissue – Properties and Pitfalls of Electrochemical Biosensors', *Horm. Metab. Res.*, **26**(11), 515–522 (1994).

M. Mascini, S. Selleri, M. Riviello, 'Glucose Electrochemical Probe with Extended Linearity for Whole Blood', *J. Pharm. Biomed. Anal.*, **7**(12), 1507–1512 (1989).

T.J. Ohara, R. Ranjagopalan, A. Heller, 'Wired Enzyme Electrodes for Amperometric Determination of Glucose or Lactate in the Presence of Interfering Substances', *Anal. Chem.*, **66**(15), 2451–2457 (1994).

C.P. Quinn, R.E. Connor, A. Heller, 'Biocompatible, Glucose-permeable Hydrogel for in situ Coating of Implantable Biosensors', *Biomaterials*, **18**(24), 1665–1670 (1997).

P. Yarnoz, J.R. Castillo, J.C. Vidal, E. Garcia, S. Mendez, 'Three Approaches to the Development of Selective Bilayer Amperometric Biosensors for Glucose By in situ Electropolymerization', *Analyst*, **124**(3), 319–324 (1999).

REFERENCES

1. K. O'Connell, 'Electrochemical Sensors: Progress in the Clinical Laboratory', Proc. Symp. Electrochemical Sensors for Biomedical Applications, *Proc. Electrochem. Soc.*, **86**(14), 7–15 (1986).
2. C.E.W. Hahn, 'Electrochemical Analysis of Clinical Blood-gases, Gases and Vapours', *Analyst*, **123**(6), 57R–86R (1998).
3. V. Razumas, J. Kanapieniene, T. Nylander, S. Engstrom, K. Larsson, 'Electrochemical Biosensors for Glucose, Lactate, Urea, and Creatinine Based on Enzymes Entrapped in a Cubic Liquid Crystalline Phase', *Anal. Chim. Acta*, **289**(2), 155–162 (1994).
4. E. Brandsteterova, P. Kubalec, I. Skacani, I. Balazovjech, 'HPLC–ED Determination of Catecholamines and their Metabolites in Urine', *Neoplasma*, **41**(4), 205–211 (1994).
5. J. Wang, 'Electroanalysis and Biosensors', *Anal. Chem.*, **71**(12), 328R–332R (1999).
6. J. Lucisano, S.V. Edelman, B.D. Quinto, D.K. Wong, 'Development of a Biosensor-based, Patient Attached Blood Glucose Monitoring System', *Polym. Mater. Sci. Eng.*, **76**, 564–565 (1997).
7. W. Lager, I.V. Lucadou, H. Nischik, T. Nowak, W. Preidel, L. Ruprecht, M.J. Stanzel, V. Tegeder, 'Implantable Electrocatalytic Glucose Sensor', *Horm. Metab. Res.*, **26**(11), 520–530 (1994).
8. J.G. Wagner, D.W. Schmidtke, P. Quinn, T.F. Fleming, B. Bernacky, A. Heller, 'Continuous Amperometric Monitoring of Glucose in a Brittle Diabetic Chimpanzee with a Miniature Subcutaneous Electrode', *Proc. Natl Acad. Sci.*, **95**, 6379–6382 (1998).
9. D.W. Schmidtke, A.C. Freeland, A. Heller, R.T. Bonnecaze, 'Measurement and Modeling of the Transient Difference between Blood and Subcutaneous Glucose Concentrations in the Rat after Injection of Insulin', *Proc. Natl Acad. Sci.*, **95**, 294–299 (1998).
10. C. Henry, 'Getting Under the Skin: Implantable Glucose Sensors', *Anal. Chem.*, **70**(17), 594A–598A (1998).
11. B. Aussedat, V. Thome-Duret, G. Reach, J. Lemmonier, J. Klein, Y. Hu, G.S. Wilson, 'A User-friendly Method for Calibrating a Subcutaneous Glucose Sensor-based Hypoglycemic Alarm', *Biosens. Bioelectron.*, **12**(11), 1061–1071 (1997).
12. H.Y. Chang, Z.U. Bae, 'Adenosine-modified Electrodes for the Determination of Glucose Without Using an Immobilized Redox Mediator', *Anal. Lett.*, **30**(11), 1981–1992 (1997).
13. C.J. McNeil, D. Athey, M. Ball, W.O. Ho, S. Krause, R.D. Armstrong, J. Des Wright, K. Rawson, 'Electrochemical Sensors Based on Impedance Measurement of Enzyme-catalyzed Polymer Dissolution: Theory and Applications', *Anal. Chem.*, **67**(21), 3928–3935 (1995).
14. I.N. Papadoyannis, *HPLC in Clinical Chemistry*, Marcel Dekker, New York, 1990. (Citation on p. 76, but chapter 6 (pp. 69–81) is useful reading.)
15. A. Heller, 'Electrical Wiring of Redox Enzymes', *Acc. Chem. Res.*, **23**(5), 128–134 (1990).
16. A. Heller, 'Electrical Connection of Enzyme Redox Centers to Electrodes', *J. Phys. Chem.*, **96**(9), 3579–3587 (1992).
17. T. Katsu, T. Hashimoto, S. Hiramaki, 'Potentiometric Determination of NAD(P)H Coenzymes Using a Nitrite-selective Membrane Electrode and a Nitrate Reductase Enzyme and the Application to Glucose Assay in Human Serum', *Anal. Sci.*, **15**(2), 135–139 (1999).
18. M. Hedenmo, A. Narvaez, E. Dominguez, I. Katakis, 'Reagentless Amperometric Glucose Dehydrogenase Biosensor Based on Electrocatalytic Oxidation of NADH by Osmium Phenathrolinedione Mediator', *Analyst*, **121**(12), 1891–1895 (1996).
19. E.J. Calvo, C. Danilowocz, 'Amperometric Enzyme Electrodes', *J. Brazilian Chem. Soc.*, **8**(6), 563–574 (1997).
20. M. Alvarezicaza, H.M. Kalisz, H.J. Hecht, K.D. Aumann, D. Schomburg, R.D. Schmid, 'The Design of Enzyme Biosensors Based on the Enzyme Structure', *Biosens. Bioelectron.*, **10**(8), 735–742 (1995).
21. S. Cosnier, J.L. Decout, M. Fontecave, C. Frier, C. Innocent, 'A Reagentless Biosensor for the Amperometric Determination of NADH', *Electroanalysis*, **10**(8), 521–525 (1998).

22. G. Khan, W. Wernet, 'A Highly Sensitive Amperometric Creatinine Sensor', *Anal. Chim. Acta.*, **351**(1–3), 151–158 (1997).
23. M. Jurkiewicz, S. Algeret, J. Amirall, M. Garcia, E. Fabregas, 'Development of a Bioanalyser for Creatinine and Urea. Validation of the Determination of Biochemical Parameters Associated with Hemodialysis', *Analyst*, **123**(6), 1321–1327 (1998).
24. A. Muochizuki, S. Hirakuni, H. Iida, 'Disposable Glucose Sensor Plate and Disposable Multi-ion Sensor Plate', *Sens. Actuators, B*, **B1**(1–6), 510–513 (1990).
25. M.S. Vreeke, P. Rocca, 'Biosensors Based on Crosslinking of Biotinylated Glucose Oxidase by Avidin', *Electroanalysis*, **8**(1), 55–60 (1996).
26. J. Kulys, H.E. Hansen, 'Long-term Response of an Integrated Carbon Paste Based Glucose Biosensor', *Anal. Chim. Acta.*, **303**(2–3), 285–294 (1995).
27. C. Saby, F. Mizutani, S. Yabuki, 'Glucose Sensor Based on Carbon Paste Electrode Incorporating Poly(ethylene glycol)-modified Glucose Oxidase and Various Mediators', *Anal. Chim. Acta*, **304**(1), 33–39 (1995).
28. D. Hage, 'Immunoassays', *Anal. Chem.*, **71**(12), 294R–304R (1999).
29. P. Treloar, J. Kane, P. Vadgama, in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Stockton Press: New York, 481–509, 1997.
30. S.-F. Chen, Y. Xu, M.P.-C. Ip, 'Electrochemical Enzyme Immunoassay for Serum Prostate Specific Antigen at Low Concentrations', *Clin. Chem.*, **43**(8), 1459–1561 (1997).
31. Y.-N. He, H.-Y. Chen, J.-J. Zheng, G.-Y. Zhang, Z.-L. Chen, 'Differential Pulse Voltammetric Enzyme-linked Immunoassay for the Determination of *Helicobacter pylori* Specific Immunoglobulin G (IgG) Antibody', *Talanta*, **44**(5), 823–830 (1997).
32. J.F. Song, J.Q. Xu, W. Guo, 'Polarographic Immunoassay for Alpha-feto Protein (AFP) with Erythrocyte Sensitized by AFP Antibody', *Anal. Lett.*, **31**(1), 27–39 (1998).
33. W. Guo, J.-F. Song, M.-R. Zhao, J.-X. Wang, 'Electrochemical Immunoassay Based on Catalytic Conversion of Substrate by Labeled Metal Ion and Polarographic Detection of the Product Generated', *Anal. Biochem.*, **259**(1), 74–79 (1998).
34. J. Song, M. Zhao, W. Guo, J. Wang, J.F. Song, M.R. Zhao, W. Guo, J. Wang, 'Polarographic Immunoassay Coupled with Catalytic Amplification of Labeled Copper Ions', *Sci. China, Ser. B: Chem.*, **40**(6), 561–567 (1997).
35. J.F. Song, Y.L. Bai, W. Guo, X. Jia, 'Polarographic Immunoassay Coupled with Catalysis of Non-radioactive Multiple Iodine Label', *Sci. China, Ser. B: Chem.*, **40**(5), 457–463 (1997).
36. J. Havas, E. Porjesz, E. Rohonczy-Boksay, 'Urea Selective Enzyme Probe and its Analytical/clinical Application', *Bioelectroanal.*, **1**, (Symp., 1st, 1986), 225–235 (1987).
37. S.M. Murphy, C.J. Hamilton, M.L. Davies, B.J. Tighe, 'Polymer Membranes in Clinical Sensor Applications. II. The Design and Fabrication of Permselective Hydrogels for Electrochemical Devices', *Biomaterials*, **13**(14), 979–990 (1992).
38. S.M. Reddy, S.P.J. Higson, I.M. Christie, P.M. Vadgama, 'Selective Membranes for the Construction and Optimization of an Amperometric Oxalate Enzyme Electrode', *Analyst*, **119**(5), 949–952 (1994).
39. S.A. Wring, J.P. Hart, 'Chemically Modified, Carbon-based Electrodes and their Application as Electrochemical Sensors for the Analysis of Biologically Important Compounds: A Review', *Analyst*, **117**(8), 1215–1229 (1992).
40. A. Mulchandani, C.L. Wang, H.H. Weetall, 'Amperometric Detection of Peroxides with Poly(anilinomethylferrocene)-modified Enzyme Electrodes', *Anal. Chem.*, **67**(1), 94–100 (1995).
41. D. Belanger, J. Nadreau, G. Fortier, 'Rotating Ring Disk Electrode Studies of Polypyrrole–glucose Oxidase Biosensors', *Electroanalysis*, **4**(10), 933–940 (1992).
42. F. Wendinski, B. Grundig, R. Renneberg, F. Spencer, 'Highly Sensitive Determination of Hydrogen Peroxidase with Tetrathiafulvalene-based Electrodes and the Application in Immunosensing', *Biosens. Bioelectron.*, **12**(1), 43–52 (1997).
43. J. Wang, G. Rivas, X. Cai, N. Dontha, H. Shiraishi, D. Luo, F. Valera, 'Sequence-specific Electrochemical Biosensing of *M. tuberculosis* DNA', *Anal. Chim. Acta*, **337**(1), 41–48 (1997).
44. H.J. Huijgen, R. Sanders, S.A. Cecco, N.N. Rehak, G.T. Sanders, R.J. Elin, 'Comparison of Three Commercially Available Ion-selective Electrodes for Ionized Magnesium Determination in Serum: a Two Center Study', *Clin. Chem.*, **44**(480), A111 (1998).
45. N. Baliga, V. Yang, M.E. Meyerhoff, 'Protamine-sensitive Polymer Membrane Electrodes for the Determination of Heparin and Low-molecular Weight Heparin in Whole Blood', *Clin.Chem.*, **44**(231), A52–A53 (1998).
46. M. Nagele, Y. Mi, E. Bakker, E. Pretsch, 'Influence of Lipophilic Inert Electrolytes on the Selectivity of Polymer Membrane Electrodes', *Anal. Chem.*, **70**(9), 1686–1691 (1998).
47. C. Espadas-Torre, V. Oklejas, K. Mowery, M.E. Meyerhoff, 'Thromboresistant Chemical Sensors Using Combined Nitric Oxide Release/Ion Sensing Polymeric Films', *J. Am. Chem. Soc.*, **119**(9), 2321–2322 (1997).
48. K. Kimura, T. Sunagawa, S. Yajima, S. Miyake, M. Yokoyama, 'Neutral Carrier Type Ion Sensors Based on Sol–Gel-derived Membranes Incorporating a bis (Crown Ether) Derivative by Covalent Bonding', *Anal. Chem.*, **70**(20), 4309–4313 (1998).
49. D.R. Coon, L.J. Amos, A.B. Bocarsly, P.A.F. Bocarsly, 'Analytical Applications of Cooperative Interactions Associated with Charge Transfer in Cyanometalate Electrodes: Analysis of Sodium and Potassium in Human Whole Blood', *Anal. Chem.*, **70**(15), 3137–3145 (1998).

50. G.Q. Du, C. Lin, A.B. Bocarsly, 'Electroanalytical Detection of Glucose using a Cyanometalate Modified Electrode: Requirements for the Oxidation of Buried Redox Sites in Glucose Oxidase', *Anal. Chem.*, **68**(5), 796–806 (1996).
51. S. Milardovic, I. Kruhak, B.S. Grabaric, 'Urea Determination in FIA Mode by a Newly Designed Urease-based Biosensor', *Laboratory Robotics and Automation*, **11**(5), 266–271 (1999).
52. R.M. Nakamura, Y. Kasahara, G.A. Rechnitz (eds.), 'Immunochemical Assays and Biosensor Technology for the 1990s', *Am. Soc. Microbiology*, Washington, D.C., 1992.
53. J.A. Cox, M.E. Tess, T.E. Cummings, 'Electroanalytical Methods Based on Modified Electrodes: A Review of Recent Advances', *Rev. Anal. Chem.*, **15**(3), 173–233 (1996).
54. M.D.B. Souza, 'Chemically Modified Electrodes Applied to Electroanalysis: A Brief Presentation', *Quimica Nova*, **20**(2), 191–195 (1997).
55. T. Ikeda, 'Enzyme-modified Electrodes with Bioelectrocatalytic Function (Review)', *Bunseki Kagaku*, **44**(5), 333–354 (1995).
56. D.E. Boyd, J.B. Yates, R.K. Coleman, 'Electroanalytical Cartridge', Patent application: US 98-79034 980514.
57. R.A. Foster, J. Cassidy, N. O'Sullivan, 'Electrochemical Diagnostic Detection Device', Patent application: WO 98-IE2 980112.
58. A. Hiratsuka, S. Sasaki, I. Karube, 'A Self-contained Glucose Sensor Chip with Arrayed Capillaries', *Electroanalysis*, **10**(4), 231–235 (1998).
59. H. Zier, F.S. Keck, C. Meyerhoff, W. Kerner, F. Bischof, E.F. Pfeiffer, 'The Continuous Measurement of Metabolic Parameters with Electrochemical Sensors', *GBF Monogr., Biosens.: Fundam., Technol. Appl.*), **17**, 369–373 (1992).
60. A. Muochizuki, S. Hirakuni, H. Iida, 'Disposable Glucose Sensor Plate and Disposable Multi-ion Sensor Plate', *Sens. Actuators, B*, **B1**(1–6), 510–513 (1990).
61. B.F.Y. Yon Hin, R.S. Sethi, C.R. Lowe, 'Multi-analyte Microelectronic Biosensors', *Sens. Actuators, B*, **B1**(1–6), 550–554 (1990).
62. M.J. Green, P.I. Hilditch, 'Disposable Single-use Sensors', *Anal. Proc.*, **28**(11), 374–376 (1991).
63. P.I. Hilditch, M.J. Green, 'Disposable Electrochemical Biosensors', *Analyst*, **116**(12), 1217–1220 (1991).
64. D.R. Matthews, R.R. Holman, E. Bown, J. Steemson, A. Watson, S. Hughes, D. Scott, 'Pen-sized Digital 30-second Blood-Glucose Meter', *Lancet*, **1**(8536), 778–779 (1987).

Electrolytes, Blood Gases, and Blood pH

Robert C. Elser
York Hospital, York, USA

1 Introduction 1
2 History 1
3 Electrolytes (Sodium, Potassium, Chloride, and Bicarbonate) 3
3.1 Measurement Principles for Sodium and Potassium 3
3.2 Measurement Principles for Chloride 4
3.3 Measurement Principles for Carbon Dioxide/Bicarbonate 5
4 Blood Gases and pH 6
4.1 pH Measurement Principles 6
4.2 PCO_2 Measurement Principles 6
4.3 PO_2 Measurement Principles 7
4.4 Optodes 8
5 Specimens 10
5.1 Whole Blood 10
5.2 Plasma 10
5.3 Anticoagulants 10
5.4 Blood Collection 10
5.5 Sample Preparation 10
5.6 Sample Storage 11
5.7 Effects of Hemolysis 11
6 Calibration 12
6.1 Electrolytes 12
6.2 pH 12
6.3 Blood Gases 12
7 Quality Control 13
Acknowledgments 14
Abbreviations and Acronyms 14
Related Articles 14
References 14

The term electrolytes, in the clinical sense, refers to the principal ionic constituents of blood: sodium (Na^+), potassium (K^+), chloride (Cl^-), and bicarbonate (HCO_3^-). Technology for the measurement of these ions has evolved from colorimetric and spectrophotometric methods to ion selective sensors. With the evolution of the technology has come an important reduction in the quantity of sample required to make the determination of concentrations of the ions in blood.

The assessment of respiratory function and the acid–base balance of the blood is accomplished by the measurement of pH, PCO_2, and PO_2. These measurements are usually not made singly, but rather as a combination of determinations made at the same time on the same blood sample.

In critically ill persons, these seven analytes are of major diagnostic importance.

1 INTRODUCTION

Electrolytes, pH, and blood gases comprise the biological constituents of blood that are frequently measured in critically ill patients. The commonly used term ''electrolytes'' refers to a tetrad of components that includes sodium ion (Na^+), potassium ion (K^+), chloride ion (Cl^-), and bicarbonate ion (HCO_3^-). The measurement of these substances in blood or blood components has been important in the diagnosis of illness for more than half a century. These ions have been measured using flame photometry, electrochemistry, and ion-selective electrodes (ISEs). Blood concentration of the electrolytes is in the millimolar range, mmol L^{-1}; the measurement methods are capable of imprecision within 1–2%.

Estimation of the acid–base balance of the blood is accomplished by the measurement of pH (an estimate of the concentration of hydrogen ion) and its relation to the concentration of HCO_3^-. The assessment of respiratory function, on the other hand, requires the measurement of the blood gases, O_2 and CO_2. In the routine clinical laboratory, the partial pressure of the gases, rather than their concentration, is measured. Gas-sensitive electrodes are used to measure PO_2 and PCO_2. In biological fluids, oxygen can be present at pressures ranging from 0 to 760 mm (101 kPa) mercury while carbon dioxide can be present at pressures ranging from 10 to 120 mm mercury (1.3–16.0 kPa). The gas electrodes are capable of measurement imprecision of about 2–5%.

2 HISTORY

The historical development of the measurement of electrolytes progressed from methods involving precipitation and gravimetric determinations in the 1920s to flame photometric measurements in the 1950s and finally to ISEs in the 1970s. Measurements of both sodium and potassium employed chemical reactions which led to insoluble precipitates of the desired cation. Although gravimetry, weighing the precipitated salt, was favored in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

early methods, subsequent development of photometric devices permitted optical measurements of the dissolved precipitates. The evolution of methods was successful in dramatically reducing the requirements of analytical sample volumes and testing time. The sodium method proposed by Albanese in 1948[1] required 500 μL of serum and took about 2 h to complete. Similar developments occurred for the measurement of potassium. Precipitation methods required 200 μL and took about 2–3 h to complete. Currently, the measurement of sodium and potassium using an ISE can be completed in less than 1 min using a sample size of less than 100 μL of whole blood or serum. The history of chloride measurements in biological fluids harks back to Volhard[1] who, in 1874, proposed an argentimetric analysis using silver nitrate to precipitate chloride. Excess silver was titrated using standard thiocyanate with ferric ion as an indicator. Following the evolution of the titration method, photometric methods, some of which used a Volhard-like reaction, were developed in the 1950s. Coulometric titrations (discussed later) marked the introduction of electrochemical techniques to the measurement of chloride in the latter part of the decade of the 1950s. Subsequently, chloride electrodes were developed in the 1970s and serve as the mainstay of chloride measurements today.

The modern history of pH, PCO_2, and PO_2 measurements on blood dates from the 1950s. Much of the development work occurred in Scandinavia, in the hospitals and laboratories of Denmark. The names of Astrup, Severinghaus, and Clark are associated with the developments that have led to the modern generation of blood-gas analyzers. Their contributions to the theoretical and practical development of measurement systems cannot be overestimated. Astrup developed the first practical system to measure blood pH[2] and to estimate PCO_2. His accomplishment was based on the relationship between pH and PCO_2 which had been well established by the mid-1950s. Using a gas equilibration apparatus and two gas mixtures having different PCO_2 values, Astrup was able to equilibrate tiny volumes of blood with the gas mixtures. By measuring the pH of the equilibrated blood and plotting the PCO_2 versus the measured pH, a graph was constructed that could be used to interpolate the actual PCO_2 of the unequilibrated blood by measuring its pH (Figure 1). This development led to significant advances in the field of respiratory therapy.

Subsequently, Stow[3] reported the concept of a direct PCO_2 measurement using a pH electrode separated from the measured solution by a gas-permeable membrane. Severinghaus elaborated on Stow's concept and developed a working PCO_2 electrode which was eventually commercialized in 1958.[4] This electrode, in its basic form, has endured for nearly 40 years, being improved only by miniaturization, and continues to be used in nearly all commercial blood-gas instrumentation.

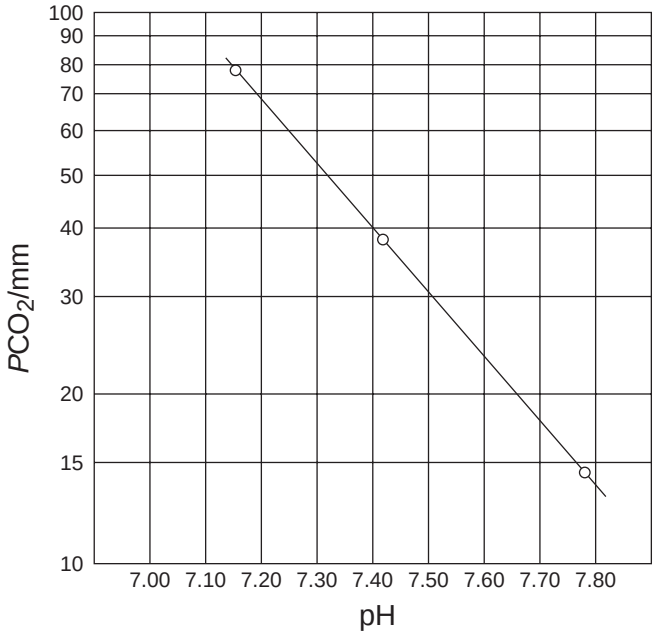

Figure 1 Astrup's pH–log PCO_2 diagram for CO_2 titration of blood or plasma with known PCO_2 values. The pH of the original sample defines its PCO_2. (Reproduced by permission of RADIOMETER from Chapter XVIII "AHA!" in *The History of Blood Gases, Acids and Bases* by P. Astrup and J.W. Severinghaus, 1st edition, Radiometer A/S, Copenhagen 1986.[6])

The development of the concept of the measurement of PO_2 in liquids is attributed to Clark[5] who serendipitously discovered a way to measure oxygen while doing research on the chemistry of steroids in blood using polarography. Before any polarographic measurements could be made on the steroids, the solutions had to be purged of oxygen. Clark used the disappearance of the polarographic signal for oxygen to determine when he could begin to study the steroids. He soon realized that he could use a polarographic electrode to measure oxygen in blood. Subsequent experiments led him to isolate the electrode from the blood solution by using a polyethylene membrane that was permeable to oxygen. Just as the Stow/Severinghaus electrode has survived for the past four decades, so has the Clark electrode which is incorporated into nearly every blood-gas instrument currently used. Improvements have been incremental and generally in the direction of miniaturization.

The history of this fascinating branch of clinical chemistry and physiology is told by two of the principal participants, Astrup and Severinghaus, in *The History of Blood Gases, Acids, and Bases*.[6]

3 ELECTROLYTES (SODIUM, POTASSIUM, CHLORIDE, AND BICARBONATE)

The most abundant cations present in biological samples, sodium (Na^+) and potassium (K^+), commonly are measured simultaneously using one of several common analytical techniques. For several decades, beginning in the 1950s, Na^+ and K^+ were measured using flame photometry. Although, in theory, any of the atomic measurement systems of atomic absorption, atomic emission, or atomic fluorescence could have been used for the measurement of sodium and potassium in biological specimens, flame emission photometry became the method of choice. As electrochemical technology evolved and ISEs became available, flame photometry was largely abandoned in favor of an electrochemical sensor technique which was inherently safer, not requiring inflammable gases or compressed air, and which allowed measurements to be made directly in undiluted whole blood. In addition to the flame and electrochemical techniques, colorimetric methods utilizing dyes which had some inherent ion specificity for either Na^+ or K^+ were also developed. Although these methods never achieved widespread use, they had the potential to be used in nonlaboratory settings, e.g. in physicians' offices.

The most abundant anions in biological samples are chloride (Cl^-) and bicarbonate (HCO_3^-). Common terminology used in the clinical field frequently misnames HCO_3^- by calling it "total CO_2 content" when, in fact, total CO_2 content includes dissolved CO_2 species, as well. In practice, the difference between total CO_2 content and HCO_3^- is less than 5%, usually within the imprecision of the analysis. Measurement of these anions almost always occurs in conjunction with the measurement of Na^+ and K^+. The development of measurement technology for these anions has progressed from largely manual and nonelectrochemical technology to automated ion selective methods.

3.1 Measurement Principles for Sodium and Potassium

3.1.1 Flame Atomic Emission Spectroscopy

Flame atomic emission spectroscopy (FAES) suffers, as do all flame atomic measurement techniques, from the variability of the number of atoms capable of being measured at any instant. The variability is due to changes in flame temperature, aspiration rate of the sample solution into the flame, and the presence of variable concentrations of other nonanalytical flame molecular species. For this reason, virtually all of the methods developed for measuring Na^+ and K^+ by FAES have incorporated an internal standard (IS) into the measurement system. All biological samples require dilution prior to FAES and the diluent provides the vehicle which introduces the IS. The IS is chosen to have similar atomic physical properties to the analyte atoms. In the case of sodium and potassium measurements, either lithium (Li) or cesium (Cs), other members of the alkali metal group, serve as the IS. The diluent contains IS at a constant concentration; each dilution of analytical specimen therefore contains, to all intents and purposes, an unvarying and identical concentration of IS. When the solution containing analyte Na^+ and/or K^+ at variable concentrations and IS at known and invariant concentration is aspirated into a flame, the atomic emission intensity of each of the analytes and the IS may vary slightly due to the effects of the flame. However, the ratio of the emission intensity of Na^+/IS and K^+/IS will be nearly constant irrespective of small variations in flame temperature or sample aspiration rate. As a result, FAES possesses the requisite imprecision, less than 1–2%, to be clinically useful.

Although nearly any flame having a temperature in excess of 1900 °C could be used for FAES, the most common gases used in clinical FAES are propane for fuel and compressed air for oxidant. The temperature of a lean propane/air flame is about 1925 °C. Lean flames should be used to avoid the deposition of soot on optical surfaces used to view the flame. The most significant danger in the use of these flames is the possibility of propane, which is heavier than air, pooling in a low spot prior to ignition and subsequently causing an explosion.

Virtually all FAES instruments utilize filter photometers to measure the atomic emission of Na^+, K^+, and either Li^+ or Cs^+. Interference filters have the advantage of inexpensively providing appropriately narrow bandwidths, 1.5 nm, to view atomic line emission. However, wavelength accuracy and bandwidth may be lost when the filters delaminate. Filters must be checked periodically for wavelength accuracy and bandwidth using precision spectrophotometry, preferably with a double-beam spectrophotometer. Sodium, potassium, lithium, and cesium emit radiation at 589 nm, 768 nm, 671 nm, and 852 nm respectively.

3.1.2 Ion-selective Electrodes

All electrode measurements require two electrodes: a measurement electrode and a reference electrode. The pair of electrodes constitutes an electrochemical cell. The principle underlying the ISE measurements of Na^+ and K^+ is potentiometry: the measurement of the potential developed by the electrodes constituting the electrochemical cell. In contrast to FAES, ISEs measure the activity of the ions instead of their concentrations. The relationship between activity and concentration depends upon the activity coefficient of the ion in the biological

fluid, the degree of dissociation for the ion, and the mass concentration of water in the sample as shown by Equation (1)

$$\text{Factor} = \frac{\rho\, H_2O}{\gamma\, Na^+ \times \beta\, Na^+} \quad (1)$$

where ρ is the mass concentration of water, γ is the activity coefficient of the ion, and β is the degree of dissociation of the ion. Concentration may be calculated, following measurements of activity, using Equation (2)

$$\text{Concentration} = \text{Factor} \times \text{Activity} \quad (2)$$

The theoretical values of the factors for Na^+ and K^+ in whole blood, plasma, or serum are 1.25×10^3 mmol L^{-1} and 1.31×10^3 mmol L^{-1}, respectively, when the mass concentration of water in the sample is normal, about 0.93 kg L^{-1}.

The reference electrodes that may be used in ISE cells are:

1. a silver/silver chloride electrode
2. a calomel electrode using a potassium chloride (KCl) bridge solution, where the KCl is either saturated (4.52 mol L^{-1}) or concentrated (3.5 mol L^{-1}) or
3. any other indicator electrode immersed in a solution containing a constant concentration of the ion to which it is sensitive.

Samples that have been tested using a reference electrode with a flowing bridge solution cannot be retested owing to contamination of the sample by the bridge solution.

3.1.2.1 Sodium Ion-selective Electrodes Glass electrodes constructed of specially formulated glass containing silicon dioxide, sodium oxide, and aluminum oxide in the ratio of 71 : 11 : 18 have good selectivity for Na^+ over K^+ ions, greater than 1000-fold, and are insensitive to H^+ in the pH range from 6 to 10. Sodium ISEs have been constructed in a variety of shapes to permit samples to flow through or across the sensor surface. A drawback of the sodium glass electrode is the requirement for etching the glass surface on a daily basis using a bifluoride (HF_2^-) solution.

3.1.2.2 Potassium Ion-selective Electrodes Glass electrodes have not been particularly useful for the measurement of K^+ in the presence of sodium because the selectivity ratio of currently available glass is only about 20 and the concentration ratio of Na^+/K^+ in plasma, serum, and whole blood ranges from approximately 20–50. Instead, liquid ion-carrier membranes containing valinomycin have been used. Valinomycin is a heterocyclic molecule that will bind K^+. The selectivity ratio of valinomycin for K^+ in the presence of sodium is 4×10^5. Valinomycin electrodes are commonly used for the measurement of K^+ in biological fluids. A major advantage in the use of this electrode is its freedom from interference by anticoagulants or native biological constituents of the sample.

3.2 Measurement Principles for Chloride

Methods for the measurement of chloride may be divided into titrimetric and ISE measurements. The titrimetric methods may be further divided into coulometric titrations[7] and chromogenic titrations. Virtually all chloride methods will respond to other halide anions such as bromide and iodide.

3.2.1 Coulometric Titration

Coulometry relies upon the extremely accurate and precise generation of silver ions electrochemically, at a constant rate, from a pure silver electrode. Coulometric titrations require three electrodes: generation, current detection, and reference electrodes. Although the details of the process cannot be described here, the principle is quite simple; silver ions are generated to form insoluble silver chloride with free chloride ions in the sample. During the titration process, before the end-point, the current detection electrode measures an essentially constant current flow between the current detection and reference electrodes. After all chloride ions have been removed by formation of insoluble silver chloride from the generated silver ions, excess silver ions accumulate and the current flow at the detection electrode increases. At the beginning of the titration, at the instant that the silver generator is turned on, a clock is started. The clock continues to count time until the current detection electrode senses the increase in current caused by the accumulation of excess silver ions, at which time it is stopped. The clock time is directly proportional to the amount of chloride ions that were present in the titrated sample as long as the rate of the silver generating process is constant and reproducible. Coulometric devices are capable of extremely precise measurements. Typical imprecision for a coulometric titration using 10-μL samples is 1–2%. Because coulometry relies upon the electrical generation of the titrant and because electrical current can be measured so accurately, coulometry has been called an ''absolute'' analytical method, a method which does not require calibration. However, in practice, the accuracy of the method depends upon the accuracy of the pipetting tool used to measure the analytical sample. Therefore, calibration of the assay eliminates any inherent inaccuracy of the pipettor from causing inaccuracy in the test result. With a calibrator, the test results of the unknown analytical samples are calculated as a ratio

(Equation 3):

$$\text{Chloride} = \frac{\text{Clocktime for Unknown}}{\text{Clocktime for Calibrator}} \times \text{Calibrator Concentration} \quad (3)$$

3.2.2 *Chromogenic Titration*

Chromogenic titrations have been devised using soluble mercuric salts such as mercuric thiocyanate ($Hg(SCN)_2$) and mercuric nitrate ($Hg(NO_3)_2$).

Mercuric thiocyanate is used in conjunction with ferric nitrate. In the presence of chloride, insoluble mercuric chloride and soluble ferric thiocyanate are formed. Ferric thiocyanate is colored and absorbs radiation at 480 nm. The technique has been automated, but has a limited linear range of approximately 80–125 mmol L^{-1}.

Mercuric nitrate ($Hg(NO_3)_2$) has been used for the manual titration of chloride using a visual end-point when excess Hg^{2+} forms a blue–violet complex with the indicator dye, *s*-diphenylhydrazine.[8] As with any visual end-point titration, the subjective nature of the end-point introduces variability between observers and the imprecision of the assay is poor, by comparison to automated techniques. Of the chromogenic techniques, the former method still enjoys some use on automated chemistry analyzers while the latter has fallen into nearly complete disuse.

3.2.3 *Ion-selective Electrodes*

Chloride-sensitive electrodes include a silver/silver chloride wire immersed in an electrolyte solution having a fixed concentration of chloride ions. The solution is contained by an ion-exchange membrane which has selectivity for chloride ions over other ions that may be present in the sample. When the sample comes in contact with the membrane, chloride ion exchange occurs at the membrane surface, creating a membrane potential. Comparison of the membrane potential with the reference electrode potential results in a signal that is proportional to the logarithm of the chloride ion concentration in the sample. Figure 2 illustrates the design of a chloride-sensitive electrode sensor.

3.3 Measurement Principles for Carbon Dioxide/ Bicarbonate

3.3.1 *Gasometry*

Total CO_2 is frequently measured by converting all CO_2 species to gas by acidification of the sample. The analytical system is designed so that following acidification of the sample, the solution, now containing dissolved CO_2 gas, passes across the sensor before the solution is exposed

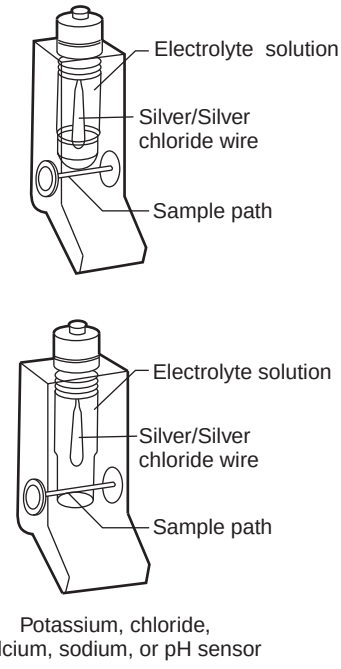

Figure 2 Flow through chloride ISE and pH sensor. (Reproduced by permission of Chiron Diagnostics from 800 Series Operator's Manual, p. 1–14, 1998.)

to the atmosphere. The most common CO_2 sensors are gas-sensitive electrodes which are specific for CO_2 and other acidic gases. These electrodes are essentially pH electrodes immersed in a weakly buffered bicarbonate solution enclosed in a chamber covered with a CO_2 gas permeable membrane.

The electrode shown in Figure 3 is incorporated into a flowing sample stream downstream from the point at which acid is introduced. In this version, the weak bicarbonate buffer is replenished between samples. The typical sample volume for these measurements is 100–150 μL.

3.3.2 *Enzymatic Reaction*

Bicarbonate is usually measured using an enzymatic reaction. The reaction involves a primary (Equation 4) and an indicator reaction (Equation 5). The primary reaction employs phosphoenolpyruvate carboxylase (PEPC) while the indicator reaction uses malate dehydrogenase (MDH)

$$\text{Bicarbonate} + \text{Phosphoenolpyruvate} \xrightarrow{\text{PEPC}} \text{Oxaloacetate} + \text{Inorganic Phosphate} \quad (4)$$

$$\text{Oxaloacetate} + NADH_2 \xrightarrow{\text{MDH}} \text{Malate} + NAD \quad (5)$$

The reaction is followed spectrophotometrically at 340 nm. The decrease in absorbance is proportional to

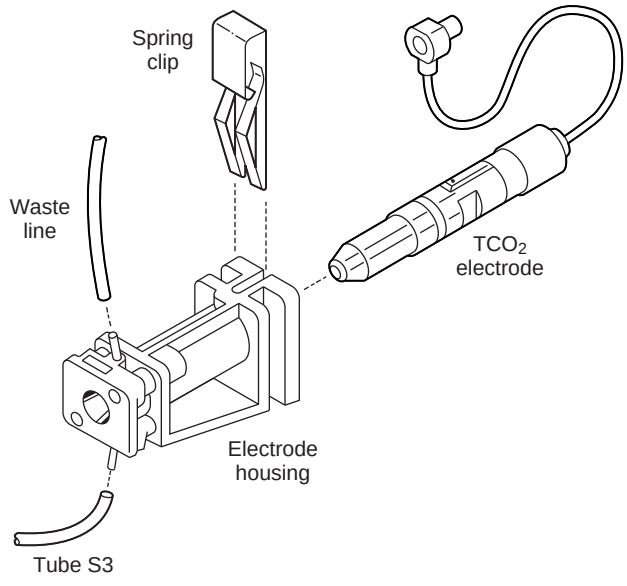

Figure 3 Total CO_2 electrode used in Dade Dimension analyzer. (Reproduced by permission of Dade Behring. Inc. from Dimension RxL Operator's Manual, p. 3–104, 1997.)

the quantity of bicarbonate present in the sample. If an estimate, rather than an accurate measurement, of bicarbonate is desired, the Henderson–Hasselbalch equation may be used following the measurement of pH and PCO_2 (see section 4).

4 BLOOD GASES AND pH

Clinical jargon associates the partial pressure measurements of CO_2 and O_2 with the term "blood gases". The abbreviations for the partial pressures sometimes employ capital "*P*" and sometimes lower case "*p*". The National Committee for Clinical Laboratory Standards (NCCLS) of the USA assigns "*P*" as the correct usage.[9]

The use of pH measurements for clinical diagnosis gained widespread usage in the 1950s and 1960s with the commercialization of clinical sensors capable of measuring pH on small volumes of blood. The commercial development of PCO_2 and PO_2 sensors in the 1960s made possible the routine measurement of blood gases. Prior to their development, the blood gases were measured either manometrically or volumetrically following release of the gases from the samples. Manometric methods involved measurement of the pressure of the released gas at a constant volume; volumetric methods measured the volume of released gas at constant pressure. Acidification of the blood sample was required to release CO_2 and addition of potassium ferricyanide caused release of oxygen. These techniques required large quantities of blood, 5–10 mL, and a manually operated apparatus incorporating liquid mercury, the Van Slyke apparatus. Electrode sensors eliminated the tedious manual procedure, avoided the use of liquid mercury, and required substantially less sample volume, 100–200 µL. As commercialization progressed, these three sensors, pH, PCO_2, and PO_2, became incorporated into analyzers which made simultaneous measurement of pH and blood gases routinely possible on extremely small volumes of blood, 80–120 µL. Current instrumentation incorporates computers that can derive additional parameters, e.g. bicarbonate and O_2 saturation, on a theoretical basis from these three measurements: pH, PCO_2, and PO_2. The relationship between pH and PCO_2 is given by the Henderson–Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K + \frac{\log[HCO_3^-]}{PCO_2 \times \text{Solubility Coefficient}} \qquad (6)$$

Equation (6) can be used to calculate the concentration of HCO_3^- when the pH and PCO_2 have been measured.

4.1 pH Measurement Principles

Virtually all blood pH sensors are glass electrodes that are designed to permit the sample to flow past the sensor surface of the electrode. By incorporating the pH-sensitive glass into the sample flow path, the electrode sensor can be miniaturized to require an extremely small sample. Since the pH electrode requires a reference electrode for comparison of its electromotive potential, the reference electrode must also be designed to permit electrolytic contact with the flowing sample stream at a liquid junction. Because the reference cell solution contains potassium chloride, the reference electrode contaminates the sample with potassium and chloride ions and must be located downstream from other measurement sensors. Figure 2 shows an example of a pH sensor.

4.2 PCO_2 Measurement Principles

Electrodes that are sensitive to the dissolved CO_2 in blood are called PCO_2 electrodes. A dissolved gas exerts a pressure at the surface of the solution equal to the pressure of gas that would be required to cause it to be dissolved at the same concentration. The pressure is called "partial pressure" because it is only part of the total pressure of the atmosphere that is exerted on the surface of the solution. Atmospheric pressure is equal to the sum of the partial pressures of all of the gases included in its composition. The PCO_2 of the atmosphere, at sea level, is only about 0.3 mmHg (0.04 kPa) because the atmosphere contains only 0.039% CO_2. However, the PCO_2 of blood is approximately 40 mmHg (5.32 kPa), much higher than

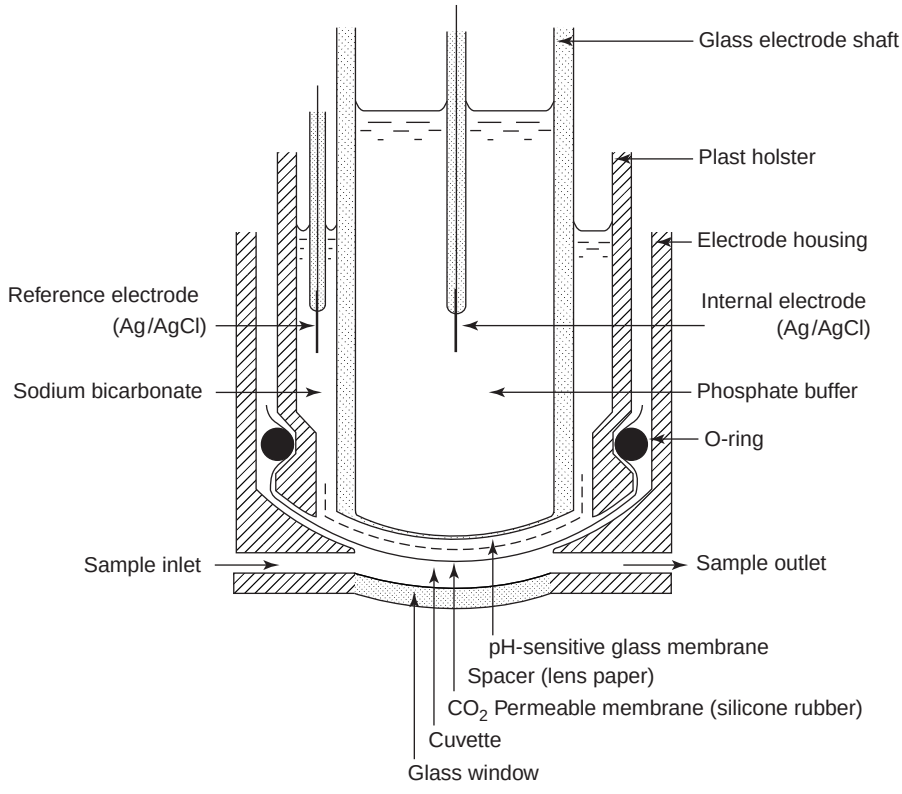

Figure 4 Schematic illustration of a PCO_2 electrode. (Reproduced by permission of Lippincott Williams and Wilkins from O. Siggaard-Andersen, *The Acid–Base Status of the Blood*, 4th edition, Williams and Wilkins, Baltimore, p. 172, 1974.)

atmospheric PCO_2, because CO_2 is produced as a waste product by the body.

The original electrode was described by Severinghaus in 1958.[4] The construction of a PCO_2 electrode is quite similar to the construction of a CO_2 sensor (see above). These electrodes behave as pH electrodes. Figure 4 shows a typical sensor with its silicone rubber membrane and buffer reservoir. Some electrode sensor designs require periodic replacement of the membrane while others are maintenance-free for the lifetime of the electrode, usually about 1 year. Figure 4 shows the electrode as originally described, while Figure 5 shows a modern miniaturized version of it as it is incorporated into a blood gas analyzer.

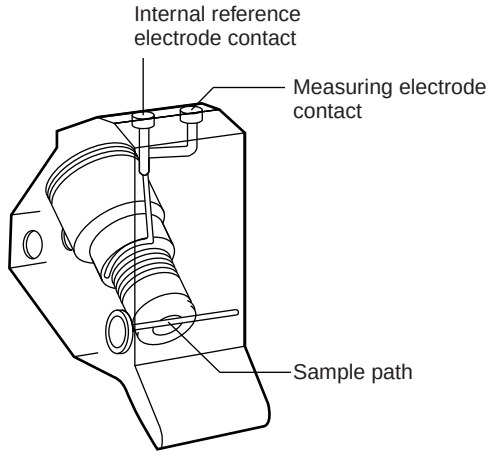

Figure 5 Flow through PCO_2 electrode. (Reproduced by permission of Chiron Diagnostics from 800 Series Operator's Manual, p. 1–14, 1998.)

4.3 PO_2 Measurement Principles

Oxygen-sensitive electrodes are called PO_2 electrodes. The original sensor was described by Clark in 1956.[5] This electrode uses a platinum cathode and a silver/silver chloride anode polarized at −650 mV. At this potential, molecular oxygen is oxidized with a concomitant current flow in the electrode sensor circuit. An amperometric circuit measures the current flow in picoamperes. In practice, the anode and cathode are covered with a thin film of buffered electrolyte held in place by a thin polypropylene membrane. As oxygen from the sample diffuses through the membrane into the electrolyte, it is oxidized and current flows in the external measuring circuit. Figure 6 shows the schematic construction of a PO_2 electrode.

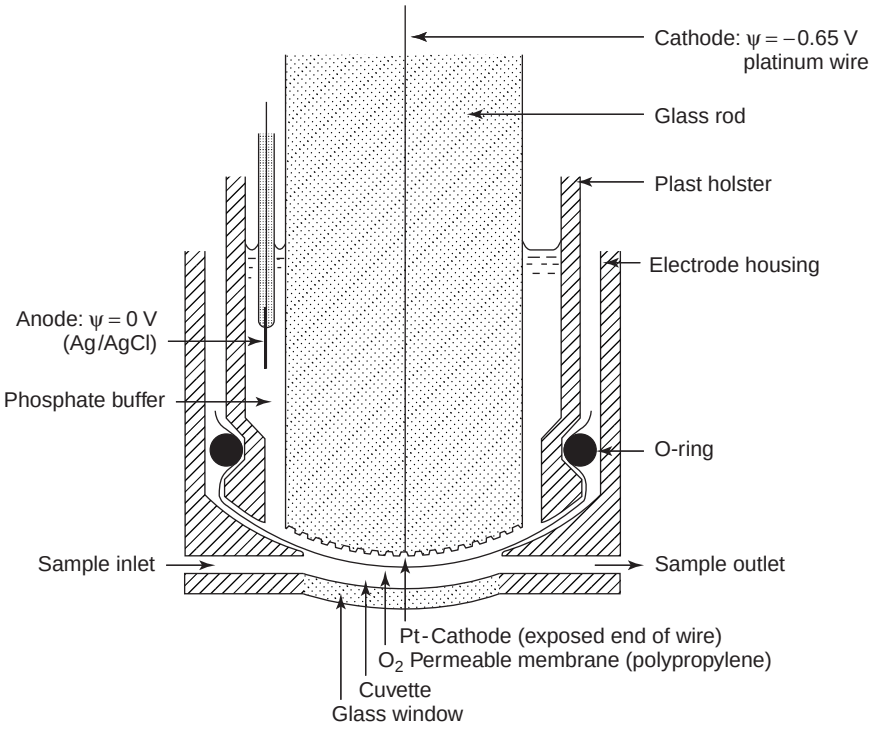

Figure 6 Schematic illustration of a PO_2 electrode. (Reproduced by permission of Lippincott Williams and Wilkins from O. Siggaard-Andersen, *The Acid–Base Status of the Blood*, 4th edition, Williams and Wilkins, Baltimore, p. 178, 1974.)

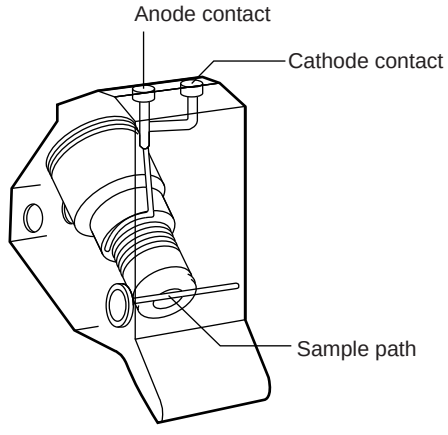

Figure 7 Flow through PO_2 electrode. (Reproduced by permission of Chiron Diagnostics from 800 Series Operator's Manual, p. 1–14, 1998.)

A miniaturized flow-through version of the electrode is shown in Figure 7. This variety of electrode is incorporated into modern automated blood-gas analyzers.

The imprecision of oxygen electrodes is about ±3–4 mmHg (0.4–0.5 kPa) at physiological partial pressures of 80–100 mmHg (10.6–13.3 kPa).

The useful lifetime of the membrane and buffered electrolyte is dependent upon electrode design. Some manufacturers choose to produce ''maintenance-free'' electrodes which are discarded after the response begins to decline. Some of these electrodes are usable for periods up to 1 year. Other manufacturers produce electrodes which have removable membranes which must be replaced on a regular basis, usually about every 6–8 weeks.

Current models of blood-gas analyzers incorporate the above-mentioned electrodes in a general design format that permits a blood sample to pass over or through them sequentially. Figure 8 shows an example of a blood-gas analyzer in which the electrodes are arranged horizontally.

4.4 Optodes

The electrodes that previously have been described represent the traditional approach for measurement of pH, PCO_2, and PO_2, using in vitro systems: systems which require withdrawal and removal of a sample, usually of arterial blood, from the human body. A new class of sensors, called optodes, has been developed to permit in vivo or ex vivo measurements of these same three analytes. While optode measurement principles have been known for some time, it is only since the mid-1980s[10,11] that they have been used for medical applications.

Optodes are photochemical sensors which incorporate optical measurements of light intensity in miniature cells.

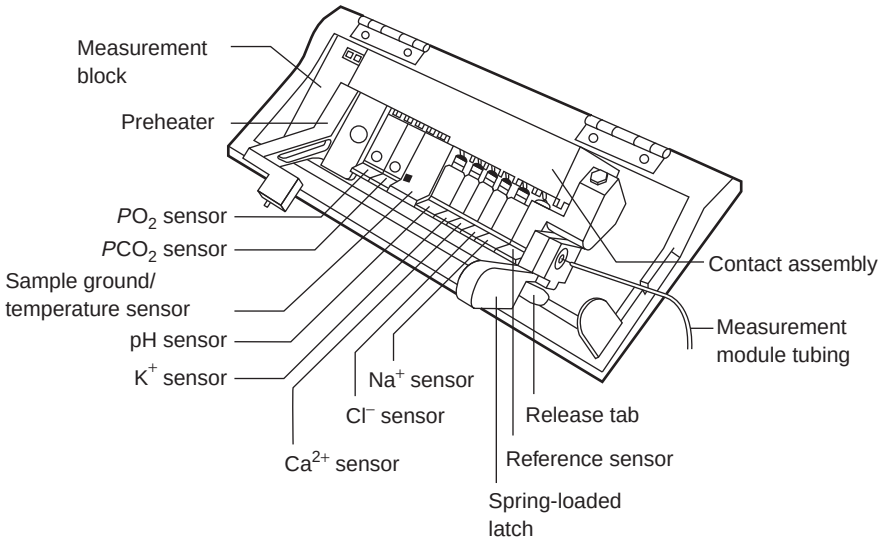

Figure 8 Bayer Blood Gas Analyzer model 850 layout. (Reproduced by permission of Chiron Diagnostics from 800 Series Operator's Manual, p. 1–14, 1998.)

These devices act as transducers between an analyte to be measured and an optical fiber. In vivo optodes are placed intravascularly and continuously measure the pH, PCO_2, and PO_2 of the blood flowing past them. Ex vivo optodes are placed outside of the body but within a catheter that gives direct vascular access as required. Blood may be withdrawn through the catheter, passed across the sensors for measurement, and returned to the circulation via the catheter.

Optodes are constructed of fiber-optic cables, a light source, and a photodetector as the common components. The measurement cells of the optodes contain a dye appropriate for the measurement to be made. Figure 9 shows the general construction of an optode with the incident transmission of light contained within the optical fiber and the fluorophore or dye altering the wavelength or intensity of radiation transmitted back to the detector.

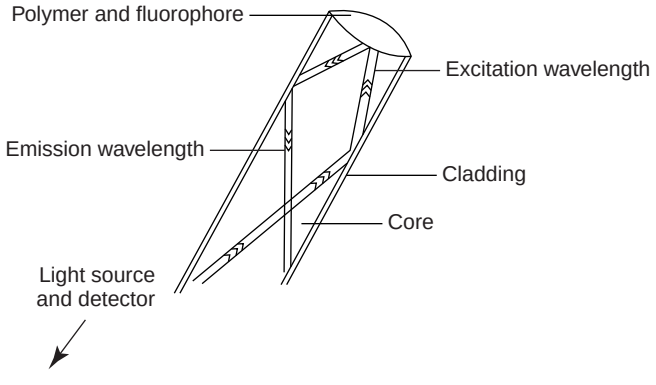

Figure 9 Schematic diagram of optode fiber showing the fluorophore and radiation traveling along the fiber. (Reproduced by permission of Optical Sensors Incorporated.)

For pH measurements, the dye characteristically exhibits optical characteristics which vary according to the pH to which it is exposed. Both absorbance and fluorescence dyes have been used for pH optodes.

PCO_2 optodes are constructed in a manner exactly analogous to PCO_2 electrodes: a pH-sensitive dye and an appropriate buffer are enclosed by a CO_2 permeable membrane. As CO_2 diffuses into the buffered solution, a change in pH is sensed and related to the PCO_2 of the sample.

PO_2 optodes utilize fluorescent dyes and capitalize on the physicochemical property of O_2 to quench, or reduce the intensity of the fluorescence emission of the radiationally excited molecules of dye.

Optodes in general suffer from several problems. The first is the photoinstability of the dyes used to generate the analytical signals. These dyes are susceptible to photobleaching and, therefore, have a limited useful lifetime. Because optodes are incorporated into the vascular system either inside the body (in vivo) or outside (ex vivo) they present the possibility of sepsis. The sensor systems must, therefore, be rugged enough to withstand sterilization techniques prior to their use. They may also suffer from coagulation of proteins on their sensor surface. Calibration of in vivo sensors must be completed prior to their insertion into the vasculature. After insertion into the vasculature, recalibration is not possible. Calibration of ex vivo sensors is accomplished prior to insertion into the arterial pressure monitoring catheter. These devices are stable for up to six days of use.

5 SPECIMENS

While the measurement of electrolytes and blood gases might be made on virtually any body fluid, the most common specimen types are whole blood, plasma, or serum. Whole blood represents the fluid that circulates in the vasculature and is comprised of red blood cells (RBCs), white blood cells (WBCs), platelets, and plasma. Plasma differs from serum only in that it contains fibrinogen while serum does not.

5.1 Whole Blood

Whole blood may be withdrawn from either veins or arteries and is termed venous or arterial blood, respectively. Either type of specimen may be used for any of the measurements of electrolytes, pH, or blood gases. The choice of using arterial specimens is usually based on the necessity to measure one of the oxygenated components of blood, PO_2 or oxygen saturation of hemoglobin.

5.2 Plasma

When blood is prevented from clotting by the addition of an anticoagulant, it can be separated by centrifugation into the solid formed cellular elements and the liquid plasma. In the absence of an anticoagulant, whole blood will clot soon after it has been withdrawn from the body. Clot formation is caused by lysis of platelets which initiate the clotting process and promote the conversion of fibrinogen to fibrin. Centrifugal separation of clotted blood results in a gelatinous clot containing the formed cellular elements held in a fibrin matrix and the supernatant liquid serum.

5.3 Anticoagulants

Anticoagulants most commonly used for electrolyte, pH, and blood gas specimens are heparins having either sodium, lithium, or ammonium ions as the cation of the heparinate salt. Both liquid and dried heparin preparations have been used clinically. The choice of dried heparinate prevents any dilution of the sample by the added volume of anticoagulant and also eliminates the contamination of specimens for blood gases, particularly PO_2, from dissolved gases in the volume of liquid anticoagulant.

The choice to use whole blood, plasma, or serum depends upon the available sensors and the requirement to make measurements of other analytes in the same sample.

For the purposes of measuring the analytes described in this chapter, relatively few anticoagulants are available that will prevent clot formation and avoid interference with any of the measurements. Only the heparin salts meet these criteria. These are available as the sodium, lithium, and ammonium heparinates and are effective at concentrations of $15-50\,\mathrm{IU\,mL^{-1}}$.

5.4 Blood Collection

Blood is collected by venipuncture into sterile, evacuated, and sealed glass or plastic tubes. Rubber stoppers are used to seal the evacuated tubes. Adapters permit a double ended needle to be used to puncture a vein with one end of the needle while the other end punctures the rubber stopper of the tube. The vacuum of the tubes is set by the manufacturer to permit a specific volume of blood to be aspirated into the tube under the combined influence of venous blood pressure and the vacuum of the tube. The tubes may contain an anticoagulant, if plasma samples are desired, or no additive, if serum is desired. Following blood collection, the blood and anticoagulant, if any, are mixed and the tube is taken to the laboratory for further processing.

5.5 Sample Preparation

5.5.1 Electrolytes

To measure electrolytes in plasma or whole blood, the choice of the cation component of the anticoagulant is important. To avoid sodium contamination, either the lithium or ammonium heparinate salts must be used. Although the measurement of lithium has not been discussed in this chapter, it is a relatively commonly measured analyte in certain clinical situations. For this reason, lithium cannot be measured on samples collected using lithium heparinate. Ammonium heparinate presents a similar problem when ammonia is likely to be measured.

When either plasma or serum are used, the sample should be prepared by centrifuging the blood collection tube, with its stopper in place, at a time and speed combination sufficient to sediment all of the cellular material completely. The tube should be stored upright, with its stopper in place, until it is tested. Storage of the sample tube without removal of the stopper keeps the sample anaerobic. Maintenance of the specimen in an anaerobic state is important to preserve the concentration of bicarbonate. Because the PCO_2 of blood is much higher than the PCO_2 of the atmosphere, CO_2 will be lost from the specimen when serum or plasma is exposed to the air. The magnitude of the loss is time-dependent (see section 5.6.3).

Plasma and whole-blood potassium concentrations are usually lower than the potassium concentration of the corresponding serum. There is a positive bias of $0.1-0.7\,\mathrm{mmol\,L^{-1}}$ in serum which is caused by the release of intracellular potassium from platelets lysed

during the clotting process. For accurate measurements of potassium, anticoagulated plasma or whole blood are the specimens of choice.

5.5.2 pH and Blood Gases

To measure pH and blood gases, arterial blood is the most common specimen. It is usually collected into a syringe containing heparin and a closure is placed over the syringe inlet to maintain the specimen in an anaerobic state. Although blood gases and pH can be measured on plasma, a whole-blood specimen is usually mixed to homogeneity prior to testing.

5.6 Sample Storage

5.6.1 Effects of Temperature

5.6.1.1 Electrolytes Specimens to be tested for electrolytes should be maintained at room temperature prior to testing. If whole blood or plasma in contact with RBCs is refrigerated, the concentration of potassium will increase. This effect is due to the slowing of the energy-dependent process that maintains the high intracellular potassium concentration. The gradient between low extracellular potassium concentration ($3–5\,mmol\,L^{-1}$) and high intracellular potassium concentration ($100–150\,mmol\,L^{-1}$) is mediated by an ATP (Adenosine 5′-triphosphate)-dependent enzyme. Since enzyme activity is proportional to temperature, cooling reduces the effective enzyme activity. When the activity is insufficient to maintain the normal gradient, intracellular potassium leaks into the plasma or serum. The increase in extracellular K^+ concentration may be as great as $0.5\,mmol\,L^{-1}$ in 90 min at 4 °C. The concentrations of other electrolytes are unaffected by storage at room or refrigerator temperatures.

5.6.1.2 pH and Blood Gases Specimens to be tested for pH and blood gases may be maintained at room temperature if they can be tested within 10 min of collection. Otherwise, they should be cooled in an ice water slurry until they are tested. Cooling reduces cellular respiration and protects the concentration of dissolved oxygen which is reflected in measured PO_2. Neither pH nor PCO_2 are affected by the temperature of storage for periods of at least 60 min.

5.6.2 Effect of Container Material on Stability of PO_2

For electrolyte measurements, the composition of the container tube is irrelevant. Glass is the preferred choice for blood-gas measurements. However, in practice, plastic syringes are widely used for the collection of specimens for blood-gas testing.

Most plastic syringes in current use are constructed of polypropylene or polyethylene. These plastics are permeable to oxygen and, in fact, contain oxygen introduced during the manufacturing process and during storage. Because of this property, blood contained in these syringes is not truly in an anaerobic state and is not protected against changes induced by its surroundings.[12,13] The PO_2 of whole-blood samples stored in plastic syringes tends to change in the direction of the PO_2 of the ambient medium surrounding the syringe. The PO_2 of room air is approximately 150 mmHg because the composition of air is 20.99% O_2. The PO_2 of ice water is approximately 250 mmHg because the solubility of oxygen in water is inversely proportional to temperature. Thus, cold water can have a higher PO_2 than the atmosphere. Samples that have PO_2 lower than that of the surroundings will show an increase in measured PO_2 while samples that have PO_2 higher than that of the surroundings will show the opposite effect.

5.6.3 Effect of Time of Storage

Serum or plasma may be stored refrigerated or frozen for extended periods (up to 72 h) without effect on concentration of sodium or chloride. Potassium is stable in separated serum for 48 h refrigerated or frozen. However, potassium is unstable in plasma or in contact with RBCs and should not be tested after more than 4 h have elapsed since collection. Bicarbonate is stable while stored anaerobically but declines rapidly after the stopper has been removed from the blood collection tube. It is stable for no longer than 8 h in aerobic storage.

PCO_2 and pH are stable for at least 60 min at room or ice water temperature. PO_2, on the other hand, is inherently unstable as soon as the blood sample has been withdrawn from the body. For practical purposes, in patients with normal leukocyte counts, PO_2 is stable at room temperature for 10 min and at ice water temperature for 30 min. In patients who have extremely high WBC counts, in excess of 100 000,[14] the apparent PO_2 may be extremely low even when it has been measured within 10 min of collection because of the excessively high consumption of O_2 by leukocytes.

5.7 Effects of Hemolysis

When RBCs rupture, the process is termed hemolysis. Hemolysis chiefly affects the measurement of potassium. Even a small percentage of RBCs that become hemolyzed during storage can cause a significant increase in the measured potassium concentration. Hemolysis to the extent of 0.5% will elevate plasma potassium by as much as $0.7\,mmol\,L^{-1}$. When plasma samples are used, hemolysis can be visibly detected by the appearance of the

red protein, hemoglobin. A reddish tinge of color in the plasma of a sample is evidence that RBCs have lysed and released their contents into the surrounding sample. The degree of redness is a measure, albeit semiquantitative, of the degree of hemolysis and, by extension, of the amount of cellular contamination of the sample. However, when whole-blood samples are used, the presence of hemolysis is likely to go undetected unless plasma is separated and examined for the visible presence of hemoglobin. Other electrolyte and blood-gas measurements are virtually unaffected by hemolysis.

6 CALIBRATION

6.1 Electrolytes

The calibration of flame photometers for the measurement of sodium and potassium is accomplished by using two calibrator solutions containing sodium and potassium, as the chloride salts. The concentrations of sodium and potassium are chosen so that one point is within the range normally encountered in blood samples and the other is outside the normal range, usually higher. The concentrations most frequently chosen are 100 and 160 mmol L^{-1} for sodium and 2.0 and 8.0 mmol L^{-1} for potassium. Because the concentration of sodium can influence the degree of ionization of potassium in the flame, potassium standards are prepared to contain a physiological amount of sodium, usually 140 mmol L^{-1}.

For calibration of chloride titrimetric methods, a single standard is chosen having a concentration of 100 mmol L^{-1}. The standard is prepared from sodium chloride. For calibration of photometric methods, two standards are usually used where one is around 80 mmol L^{-1} and the other approximates 110 mmol L^{-1}.

Calibrators for total CO_2 content or bicarbonate concentration may be made by dissolving dry sodium carbonate or sodium bicarbonate in saline to give final concentrations of 10–40 mmol L^{-1} equivalents of CO_2 or bicarbonate.

The preparation of calibrators for ISEs requires that the ionic strength of the calibrator solutions be equivalent to the ionic strength (I) of blood, 160 mmol kg^{-1}.[15] Calibrators are usually controlled to an ionic strength of 160 mmol kg^{-1} for two reasons. First, ionic strength, I, is the primary variable influencing the activity coefficient of an ion. At equal I, the activity coefficients of ions in the calibrators will approximate those of the sample, allowing concentration units to be reported, instead of activity. Second, equivalent ionic strengths and ionic composition between calibrator and sample will minimize the measurement of error produced by the residual liquid junction potential, which is the difference in the liquid junction potential at the reference electrode between the calibrator and the sample. Usually, one calibration point is chosen in the middle of the normally expected range for each ion and the second point is outside of the normal range. These points are used to construct a calibration line where the signal generated by the sensor electrode, in millivolts, is proportional to the logarithm of the ion concentration. Calibrators are usually pH buffered with a buffer species that does not complex any ions. Zwitterionic buffers like HEPES (*N*-(2-hydroxyethyl)piperazine-*N*′-ethanesulfonic acid) and MOPS (3-(*N*-Morpholino)propanesulfonic acid) are the most popular. The use of these buffers allows simultaneous calibration of pH, as well.

6.2 pH

Calibration of pH is accomplished by the use of buffers of well-defined composition. For clinical purposes, two buffers certified by the National Institute of Standards and Technology (NIST) have been widely used. These are phosphate buffers having pH of 6.85 and 7.375 at 37 °C. The International Federation of Clinical Chemistry (IFCC) has defined a reference method for the measurement of pH in blood.[16] Buffers that are supplied by commercial vendors of measuring instruments are traceable to the NIST or IFCC buffers. All buffers must be kept sterile during use because of the possibility of contamination with mold or bacteria that could change the pH of the buffer.

6.3 Blood Gases

Traditionally, mixtures of gases have been used to calibrate the PCO_2 and PO_2 sensors. These mixtures of nitrogen, oxygen, and carbon dioxide should be traceable to NIST Standard Reference Materials (SRMs) of gas mixtures for clinical use. The most common mixtures are 10% CO_2, 0% O_2, 90% nitrogen and 5% CO_2, 12% O_2, 83% nitrogen. These correspond to SRMs 1700 and 1701, respectively. More recently, aqueous solutions equilibrated with these common mixtures have been used for calibration. Nearly all blood-gas analyzers use two gas mixtures or solutions to provide calibration for the gas electrodes.

The process of calibration of electrodes involves the introduction of the two calibrators, usually the nonzero O_2 calibrator first followed by the zero O_2 calibrator. Prior to the introduction of calibration gases into the sensor measurement compartment, they are warmed and humidified by bubbling them through a water-filled chamber. During initial calibration of a new electrode, or one that has been recently maintained, several cycles of calibration may be required before the electrode responses are repeatable.

In general, pH and gas sensors are stable for periods up to 8 h. Because of federal regulations in the USA, calibration frequency is probably higher than necessary. Regulations require a two-point calibration once every 4–8 h and a one-point calibration every 30 min. The one-point calibration is made using the 5% CO_2, 12% O_2 gas mixture. The single-point calibration provides intermediate correction for electrode drift and is faster than the two-point calibration. In busy clinical laboratories, two-point calibrations are usually scheduled at slack work time.

The accurate calibration of a blood-gas analyzer may be affected by the changing response of the electrodes due to aging of the sensor surfaces or the formation of a protein coating over the surface. Calibration of gas sensors, using fixed-percentage gas mixtures, is based on the accurate calculation of the partial pressure of the measured gas in the mixture. Equation (7) shows how the *P*Gas is calculated:

$$P\text{Gas} = (\text{Barometric Pressure} - 47) \times \text{Gas Percentage} \quad (7)$$

The barometric pressure must be corrected for the partial pressure of water contained in the humidified gas. At 37 °C, the partial pressure of water is 47 mmHg (6.3 kPa). Errors in the measurement of barometric pressure translate into errors in the calibration value assigned to the gas calibrators.

7 QUALITY CONTROL

The measurement of electrolytes, pH, and blood gases assumes critical importance in the care and assessment of seriously ill patients. The most widely used method to assure quality of results is to employ a quality control (QC) program. The purpose of a QC program is to allow testing of analytical instruments, in a statistical manner, so that their performance may be monitored to some desired confidence level.

The purpose of testing QC materials is to assess the accuracy and precision of the measurements made by analytical instrumentation. The concepts of accuracy and precision can be visualized by an analogy to target shooting. In this analogy, the bull's eye represents the "true" result. The closeness of any hit to the bull's eye represents the accuracy. Precision is represented by the pattern of hits. The more closely spaced the pattern, the higher the degree of precision. Of course, a method could be accurate, on average, and extremely imprecise or, alternatively, inaccurate but extremely precise.

QC measurements are made on materials that mimic the characteristics of authentic specimens; the ideal QC specimen would contain proteins, cellular components, and inorganic salts that would contribute ionic strength, viscosity, solubility, and oxygen transport characteristics similar to that of human plasma, serum, or whole blood. Materials that mimic serum and plasma are commercially available, Although there are no materials that contain cellular elements, solutions containing stabilized bovine hemoglobin mimic whole-blood oxygen dissociation characteristics.

Alternative QC materials include freeze-dried (lyophilized) human serum, human serum to which ethylene glycol has been added, buffered hemoglobin-containing solutions, aqueous buffer solutions, and fluorocarbon emulsions. Lyophilized and glycol-supplemented serum are commonly used as QC specimens for electrolytes. These materials have physical characteristics that are similar to human serum specimens and behave similarly to human serum in techniques which do not require whole blood as a sample. However, at this time, there are no commercially available materials which adequately imitate whole blood.

None of the commercially available pH and blood gas QC materials behave identically to human specimens. The characteristics which have been difficult to reproduce are the protein and cellular composition and the oxygen buffering capacity of hemoglobin. Commercial aqueous buffers which have been equilibrated with O_2 and CO_2 to produce appropriate PO_2 and PCO_2 values are widely used. These buffers are quite adequate for pH and PCO_2 QC but are only marginally suitable for PO_2. Because these buffers lack hemoglobin, the concentration of dissolved oxygen cannot be buffered against changes due to loss of O_2 into the atmosphere. In whole blood, the equilibrium between dissolved O_2 and hemoglobin-bound O_2 protects the PO_2 values from rapid change due to diffusional loss of oxygen from the sample to the atmosphere. The PO_2 values of these materials change after aerobic exposure and the changes are unacceptably large within 1–2 min of exposure to room air.

In addition to PO_2 changes observed on exposure to ambient air, the PO_2 of these materials is also sensitive to the temperature at which the QC materials have been stored and re-equilibrated. Because the aqueous solubility of oxygen is inversely proportional to temperature, the PO_2 of the airspace above the liquid in sealed QC ampoules changes when the temperature of the ampoule is different from the temperature at which it was originally produced. Re-equilibration of the QC solution, by shaking the ampoule, may cause a change in the measured PO_2 values if the temperature is more than several degrees different from the original equilibration temperature. Aqueous controls equilibrated at 26 °C and opened at either 26 °C or 21 °C, temperatures that might be experienced at seasonal extremes in a single laboratory, showed PO_2 values that differed by up to

10 mmHg (1.3 kPa) for[17] the same material at the two temperatures.

Perfluorinated compounds, such as perfluorodecalin and perfluorotributylamine, have been prepared as oxygen-equilibrated emulsions in Tris (tris hydroxymethylamino methane) buffers. These materials have oxygen solubility that more closely approximates that of human blood and have been used for as QC materials for PO_2. They have yielded better imprecision values for PO_2 than have aqueous QC materials.

Whole blood and aqueous buffers may be equilibrated with gases of known composition in devices called tonometers. These devices are thermostatically controlled to maintain the temperature of the contents at 37 °C. Equilibration of the liquid is accomplished by maximizing the surface area of the liquid sample. Some of these devices achieve this by bubbling the gas through the liquid while others, called ''spinning basket'' tonometers, increase surface area by causing the liquid to form a thin film on the sides of the tonometer while the liquid container is spinning at a high speed.

Tonometry of liquid samples has some advantages over the use of commercially prepared QC samples: gas tensions can be selected according to local needs, when whole blood is used as the liquid.

One of the paradoxes encountered in the QC of blood gases is that the QC testing results may better reflect the care taken in handling the QC material itself prior to testing, than the quality or adequacy of instrument performance.

ACKNOWLEDGMENTS

The author acknowledges his colleagues Paul A. D'Orazio PhD, and Alan D. Cormier PhD for their assistance and guidance in completing this article.

ABBREVIATIONS AND ACRONYMS

ATP	Adenosine 5′-Triphosphate
FAES	Flame Atomic Emission Spectroscopy
HEPES	*N*-(2-Hydroxyethyl)piperazine-*N*′-ethanesulfonic acid
IFCC	International Federation of Clinical Chemistry
IS	Internal Standard
ISE	Ion-selective Electrode
MDH	Malate Dehydrogenase
NCCLS	National Committee for Clinical Laboratory Standards
NIST	National Institute of Standards and Technology
PEPC	Phosphoenolpyruvate Carboxylase
QC	Quality Control
RBC	Red Blood Cell
SRM	Standard Reference Material
WBC	White Blood Cell

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Automation in the Clinical Laboratory • Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Glucose Measurement • Point-of-care Testing • Statistical Quality Control in Clinical Laboratories

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals

REFERENCES

1. R.J. Henry (ed.), *Clinical Chemistry: Principles and Technics*, Hoeber Medical Division, Harper and Row, New York, 347, 1964.
2. P. Astrup, S. Schrøder, 'Apparatus for Anaerobic Determination of pH in Blood', *Scand. J. Clin. Invest.*, **8**, 30–32 (1956).
3. R.W. Stow, B.F. Randall, 'Electrical Measurement of the PCO_2 of Blood', *Am. J. Physiol.*, **179**, 678 (1954).
4. J.W. Severinghaus, A.F. Bradley, 'Electrodes for Blood PO_2 and PCO_2 Determination', *J. Appl. Physiol.*, **13**, 515–520 (1958).
5. L.C. Clark, Jr, 'Monitoring and Control of Blood and Tissue O_2 Tensions', *Trans. Am. Soc. Artif. Intern. Organs*, **2**, 41–48 (1956).
6. P. Astrup, J.W. Severinghaus, *The History of Blood Gases, Acids and Bases*, 1st edition, Radiometer A/S, Copenhagen, 1986.
7. E. Cotlove, H.V. Trantham, R.L. Bowman, 'An Instrument and Method for Automatic, Rapid, Accurate, and Sensitive Titration of Chloride in Biologic Samples', *J. Lab. Clin. Med.*, **51**(3), 461–468 (1958).
8. O. Schales, S.S. Schales, 'Simple and Accurate Method for Determination of Chloride in Biological Fluids', *J. Biol. Chem.*, **140**, 879–884 (1941).

9. National Committee for Clinical Laboratory Standards, 'Definitions of Quantities and Conventions Related to Blood pH and Gas Analysis', 2nd edition, NCCLS Document C12-T2 (ISBN 1-56238-242-X), NCCLS, Villanova, PA, 1991.
10. N. Opitz, D.W. Lubbers, 'Increased Resolution Power in PO_2 Analysis at Lower PO_2 Levels Via Sensitivity Enhanced Optical PO_2 Sensors (PO_2 optodes) Using Fluorescence Dyes', *Adv. Exp. Med. Biol.*, **180**, 261–267 (1984).
11. N. Opitz, D.W. Lubbers, 'Compact CO_2 Gas Analyzer with Favourable Signal-to-noise Ratio and Resolution Using Special Fluorescence Sensors (Optodes) Illuminated by Blue LEDs', *Adv. Exp. Med. Biol.*, **180**, 757–762 (1984).
12. J.J. Mahoney, J.A. Harvey, R.J. Wong, A.L. Van Kessel, 'Changes in Oxygen Measurements When Whole Blood is Stored in Iced Plastic or Glass Syringes', *Clin. Chem.*, **37**, 1244–1248 (1991).
13. O. Muller-Plathe, S. Heyduck, 'Stability of Blood Gases, Electrolytes, and Haemoglobin in Heparinized Whole Blood Samples: Influence of the Type of Syringe', *Eur. J. Clin. Chem. Clin. Biochem.*, **30**, 349–355 (1992).
14. M.J. Fox, J.S. Brody, L.R. Weintraub, J. Szymanski, C. O'Donnell, 'Leukocyte Larceny: A Cause of Spurious Hypoxemia', *Am. J. Med.*, **67**, 742–746 (1979).
15. H.F. Osswald, H.R. Wuhrmann, 'Calibration Standards for Multi Ion Analysis in Whole Blood Samples', in *Progress in Enzyme and Ion-selective Electrodes*, eds., D.W. Lubbers, H. Acker, R.P. Buck, G. Eisenman, M. Kessler, W. Simon, Springer Verlag, Berlin, 74–78, 1980.
16. A.H.J. Maas, H.F. Weisberg, R.W. Burnett, O. Muller-Plathe, P.D. Wimberly, W.G. Zijlstra, R.A. Durst, O. Siggaard-Andersen, 'Approved IFCC methods. Reference Method (1986) for pH Measurement in Blood', *J. Clin. Chem. Clin. Biochem.*, **25**, 281–289 (1987).
17. S.T. Ong, D. David, M. Snow, J.E. Hansen, 'Effect of Variations in Room Temperature on Measured Values of Blood Gas Quality-Control Materials', *Clin. Chem.*, **29**, 502–505 (1983).

Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Roger L. Bertholf
University of Florida Health Science Center, Jacksonville, USA

1 Introduction 1
2 Instrument Design and Theory 2
2.1 Gas Chromatography 2
2.2 Mass Spectrometry 3
2.3 Combined Gas Chromatography and Mass Spectrometry 4
3 Specimen Preparation 4
3.1 Considerations for Biological Specimens 5
3.2 Preparation of Volatile Derivatives for Gas Chromatography 5
3.3 Solid Sampling Techniques for Mass Spectrometry 5
4 Applications to Drug Measurements 6
4.1 Drug Screening by Gas Chromatography 6
4.2 Quantitative Applications of Gas Chromatography/Mass Spectrometry 8
5 Clinical Applications of Mass Spectrometry 13
5.1 Screening for Inborn Errors of Metabolism 13
5.2 Protein Measurement 14
5.3 Drug Measurements 16
5.4 Elemental Analytes 17
5.5 Characterization of Protein Epitopes 17
5.6 Detection of Genetic Mutations 17
Abbreviations and Acronyms 18
Related Articles 18
References 19

Automated spectrophotometric, electrochemical, and immunochemical analyses have become the mainstays of clinical chemistry and toxicology laboratory services, but their scope is limited. A wide array of clinically relevant analytes demand more sophisticated analytical techniques to provide sensitive and specific assays for diagnostic purposes. Gas chromatography (GC) and mass spectrometry (MS) are becoming cost-effective alternatives for many of these challenging assays. GC is a robust technique that offers the ability to resolve volatile components of a complex mixture. MS provides structural information that can unambiguously identify a compound. In combination, these two techniques provide qualitative and quantitative answers to many difficult analytical problems. GC and MS applications have been developed for a variety of clinical analytes, and the use of these methods has created new and exciting frontiers for clinical laboratory medicine. Several clinical applications of these analytical techniques will be discussed in this chapter.

1 INTRODUCTION

GC and MS are complementary techniques that together create a powerful and versatile analytical method. Separation of the volatile components of a mixture by GC is a technology that was first described in 1952,[1] and it was immediately recognized as an indispensable tool for the analysis of organic compounds. Of particular importance in the evolution of GC toward modern instruments was the introduction of capillary chromatographic columns, which improved the resolution of GC separations by several orders of magnitude. There are two significant limitations of GC as a qualitative and quantitative analytical technique, however. The first limitation is the necessity for analytes to be sufficiently volatile and thermally stable to vaporize at practical temperatures. A second limitation is the specificity of GC detectors, which can range from very nonspecific (e.g. thermal conductivity, flame ionization detectors (FIDs)), to highly specific (mass spectrometer). GC/MS combines the resolving capabilities of GC with the unique structural information from MS, making it the hybrid analytical method of choice for qualitative analysis of suitably volatile organic compounds. Quantitative applications of GC/MS are more complicated, and typically require internal standards. The ability to resolve the components of complex mixtures, and provide qualitative information about organic molecules, makes GC/MS an attractive technique for biomedical applications. Several definitive methods for bioanalytes are based on GC/MS applications. A common application of GC/MS in clinical laboratory operations is for toxicological purposes, including identification and measurement of drugs of abuse in urine.

As an independent analytical method, GC is also useful in clinical chemistry and toxicology laboratories. The method can be used to screen for a wide variety of drugs and metabolites in urine, where chromatographic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

retention times are used for presumptive identification of the compounds detected by flame ionization or nitrogen–phosphorus techniques.

MS has limited standalone applications, since specimen purity is essential. MS methods for measuring low-boiling compounds require a procedure that will volatilize enough molecules to be detected. There are several approaches to MS measurement of nonvolatile compounds, including liquid chromatography/MS interfaces, fast atom bombardment (FAB), electrospray, thermospray, and matrix-assisted laser desorption/ionization (MALDI). All of these methods incorporate techniques that ultimately produce vapor-phase molecules that are subsequently fragmented in the mass spectrometer's ion source.

2 INSTRUMENT DESIGN AND THEORY

2.1 Gas Chromatography

A typical gas chromatograph comprises three fundamental components: an injection system, a chromatographic column, and a detector. In most cases, specimens for GC analyses are dissolved in a volatile solvent, although neat or gaseous specimens can also be used. Most GC injection systems are designed to vaporize liquid specimens, and they accomplish this by heating the injector body to a temperature above the boiling point of the solvent and analyte. In older GC designs, the sample was injected directly into the chromatographic column, which was preheated. However, the introduction of capillary chromatographic columns, which have bores half a millimeter or less in diameter, required innovative injector designs. The challenge was to avoid peak broadening due to leakage of residual sample into the capillary column over an extended period of time. One microliter of specimen, when volatized, occupies a considerable volume within the injector body, and the small caliber of the capillary column cannot accommodate the large volume of vapor. One approach to minimizing the injection bandwidth is to constantly purge the injector body so that only a small amount of the vapor has the opportunity to enter the capillary column – this technique is called *split* injection. The split ratio (amount of specimen entering the column versus the amount purged) typically varies from 1 : 10 to 1 : 99. A limitation of split injection is the loss of analytical sensitivity, since a smaller amount of specimen enters the column and detector. In some cases, the loss of analytical sensitivity is not problematic, and may even be beneficial, especially when analyte concentration is high and the detector's range of linear response is limited.

Another approach to capillary column injectors is *splitless*, or *Gröb*, injection, in honor of the technique's inventor. In a splitless injection, the injector body is kept hot enough to vaporize the specimen and solvent, but the column temperature remains below the boiling point of the solvent. As the vaporized specimen enters the capillary column, it condenses and therefore the bandwidth is minimized. After a sufficient period of time (usually about 60 s), the injector body is purged and the column is warmed up to re-vaporize the specimen and begin the chromatography. Splitless injections are technically more complex and involve more variables than split injections, but a significantly greater amount of specimen is delivered to the capillary column, resulting in better analytical sensitivity.

On-column injections with capillary columns are also possible, and require specially designed syringes fitted with needles that terminate with a length of very small capillary, which fits inside the chromatographic column. Because of the fine capillary point, the syringes are delicate, and generally not compatible with autosampler mechanisms.

For sufficiently volatile compounds, vapor may be injected into the gas chromatograph using an airtight syringe. Raoult's law states that the mole fractions contained in the vapor phase above a liquid are determined by the respective vapor pressures of the constituents of the liquid, which in turn are proportional to their relative concentrations. Therefore, the vapor in equilibrium with a liquid can be used to quantify volatile constituents in the liquid – this technique is called *headspace analysis*. Headspace sampling offers several advantages over conventional liquid injections: the vapor is substantially free of nonvolatile constituents that may form residue inside the injector; the injection bandwidth is considerably reduced; and specimen delivery is more nearly quantitative. Headspace analysis is only useful for highly volatile compounds such as low-molecular-weight alcohols.

GC column performance improved dramatically with the introduction of fused-silica capillary columns, a technology derived from fiber optics. Resolution equivalent to several hundred thousand theoretical plates is commonly achievable with capillary GC columns. Microprocessor control of the GC oven temperature has enhanced the ability to program temperature changes, improving both the resolution and speed of GC analyses. In most GC columns the stationary phase is a liquid and the analytical method is therefore gas–liquid chromatography, following the widely used convention of specifying the state of both stationary and mobile phases in the names of chromatographic applications. Gas–solid chromatography applications also exist, but are less common. The liquid stationary phase may be coated on a solid support or chemically bonded to the inner wall of a fused silica capillary column (''bonded phase'' columns).

The choice of GC detector depends on the type of compound that is to be measured, the sensitivity that is required, and the degree of selectivity necessary to avoid significant interference. Thermal conductivity detectors have moderate sensitivity, but are not selective. FIDs have better sensitivity, and respond mostly to hydrocarbon compounds. Nitrogen–phosphorus detectors are specific for nitrogen- and phosphorus-containing compounds, and are very sensitive. Electron capture detectors can measure chlorine-containing compounds in subpicogram amounts. The properties and performance characteristics of various GC detectors are summarized in Table 1.

The versatility and ruggedness of GC makes this analytical method an attractive choice for the measurement of easily vaporized compounds, which include many drugs, steroid hormones, vitamins, and metabolic products. The factor that has most limited the incorporation of GC into routine clinical laboratory services is the low throughput and difficulty in automating GC applications.

2.2 Mass Spectrometry

Several instrumental techniques have been devised to separate and measure charged particles based on their mass. A typical mass spectrometer consists of four components: an inlet system, an ion source, a mass analyzer, and a detector. The first of these components, the inlet system, must ensure that a pure compound is delivered to the mass analyzer. For this reason, chromatographic systems are a popular choice for a mass spectrometer inlet system. The ion source is where the compound is ionized, a process that is ordinarily followed by decomposition of the analyte into unique, charged fragments. The mass analyzer sorts the charged fragments and the detector measures the number of charged fragments of any given mass.

Since a mass spectrum (sometimes called a mass fragmentogram) uniquely identifies a compound based on its fragmentation pattern, superimposition of the fragments from a second compound in the ion source would make the spectrum ambiguous. Therefore, the inlet system for a mass spectrometer must deliver pure compound to the ion source in order for the mass spectrometer to be useful for qualitative analysis. Inlet systems for MS include GC, liquid chromatographs, and several methods for vaporization and ionization of nonvolatile compounds. The ion source in a mass spectrometer usually operates under a vacuum – the presence of oxygen and nitrogen may affect ionization and contribute interfering fragments to the mass spectrum – so a pressure differential exists between the ion source and inlet system. This pressure differential is difficult to maintain when the inlet system is pressurized, as are gas and liquid chromatographs. Several devices have been created to remove the mobile phase as it elutes from the chromatographic system so that only analyte enters the ion source; examples are vacuum jet separators (for packed-column GC systems), and moving-belt solvent evaporators (for high-performance liquid chromatographs). Capillary GC columns can usually terminate at the entrance to the ion source since the minimal carrier gas flow can be removed efficiently by the mass spectrometer's vacuum system.

Table 1 Performance characteristics of common GC detectors

Detector	Detection limit	Linear range	Application
Thermal conductivity	0.5 ng	10^5	Universal
Flame ionization	10 pg	10^7	Hydrocarbons
Electron capture	0.05 pg	10^4	Halides
Thermionic (nitrogen–phosphorus)	0.1 pg	10^3	N, P
Mass spectrometer	10 pg	10^6	Universal

When solid sampling systems for nonvolatile analytes are used, the pressure differential is less of a concern because the sampling system can operate under vacuum. Solid sampling inlet systems include MALDI, FAB, thermospray, and electrospray. In a MALDI system, the analyte is embedded into a pure crystalline matrix. When a laser is directed at the crystal, analyte and crystal molecules are ejected. FAB is a similar technique, except that high-energy beams of inert atoms, such as argon, are used to initiate molecular ejection. In electrospray ionization, the analyte is dissolved in an organic solvent, and passed through an electrically charged capillary. Small clusters of analyte/solvent form in the capillary, and become charged. As the clusters are accelerated through a series of lenses, the solvent is gradually removed, resulting in smaller and smaller clusters. When the clusters reach a certain size, coulombic forces cause them to explode, and the resulting fragments are measured in the mass analyzer. Thermospray ionization is a similar technique, except that the capillary is heated, and solvent evaporates quickly after the analyte/solvent aerosol exits the capillary. In both electrospray and thermospray applications, nonvolatile analytes are stranded in the vapor phase as solvent is removed, and can therefore enter the mass analyzer and be measured. These solid sampling techniques are particularly useful for high-molecular-weight compounds, which include proteins and nucleic acids.

The ion source of a mass spectrometer shatters the analyte molecules so that their fragments can be separated and measured. Most mass spectrometers use a high-energy flux of electrons to ionize molecules – the method is called electron impact ionization. Most reference mass spectra are generated by electron impact ionization. There are circumstances, though, when electron impact

ionization does not produce satisfactory spectral uniqueness or analytical sensitivity, and other ionization methods may be preferable. One alternative method is chemical ionization, in which the ion source is pressurized with a reagent gas. The electron flux ionizes the reagent gas, which in turn interacts with the analyte to produce charged species. This approach is particularly useful for generating negatively charged ions. Fragments may also be produced by collisional dissociation, where analyte molecules (or fragments) are accelerated and collide with inert gas molecules to produce fragments. This technique is often used in mass spectrometers that have multiple mass analyzers, and the collisionally induced fragments are therefore called daughter ions since they are produced after initial ionization and passage through the first-stage mass analyzer.

There are several types of mass analyzers, and some instruments combine multiple mass analyzers. Time-of-flight mass spectrometers incorporate a simple design in which fragments are separated based on their velocities. Magnetic sector mass spectrometers separate fragments based on the degree to which they are deflected in a magnetic field. Magnetic sector instruments are very sensitive, but cost and complexity have limited their widespread use in clinical laboratories. Instruments that incorporate two magnetic sector mass analyzers (double focusing MS) can achieve very high resolution, and are useful for making accurate mass measurements. Mass measurements with accuracy to 0.0001 amu are usually sufficient to determine the exact empirical formula of a parent ion or fragment.

The most popular mass analyzer is the quadrupole, which uses a combination of static and oscillating (radio frequency) electromagnetic fields to separate the ions produced in the ion source. Quadrupole instruments are relatively inexpensive, have <1.0 amu resolution, and have detection limits for most compounds in the picogram range. Multiple quadrupole instruments have also been designed, their principal advantage being the ability to analyze mixtures of compounds.

A variation on the quadrupole mass analyzer is the ion trap mass spectrometer. The principal difference between a quadrupole analyzer and an ion trap is that the former filters ions by creating an oscillating electromagnetic path through which the ions travel, whereas an ion trap confines the ions with the oscillating electromagnetic field. An advantage of the ion trap mass spectrometer is its sensitivity, since ions of a particular mass can be accumulated, then released to the detector – the ion yield is greater than that achievable by the quadrupole design. Ion trap instruments cost about the same as quadrupole instruments, and are more sensitive, but also have two disadvantages: mass spectra obtained on ion trap instruments do not always correspond closely with reference spectra generated by quadrupole or magnetic sector instruments; and ion trap instruments are, generally, less precise for quantitative analysis than are quadrupole instruments. Nevertheless, ion trap mass spectrometers are used in many of the same applications as quadrupole instruments. Multiple mass analyzer instruments using ion traps have also been designed; usually the ion trap accumulates a particular ion, and a quadrupole is used to subsequently measure the daughter ions.

Most mass spectrometers use an electron multiplier tube as the detector, although the design may be modified with dynodes in order to measure both positive and negative ions.

2.3 Combined Gas Chromatography and Mass Spectrometry

The combination of GC and MS is one of the most useful and versatile analytical configurations available for measuring organic molecules. Although in principle any gas chromatograph and mass spectrometer could be combined, the most popular configuration nowadays is a capillary gas chromatograph with a split/splitless injector and a quadrupole mass spectrometer using electron impact ionization. This configuration is especially useful for measuring drugs in body fluids, when unequivocal identification is necessary. Addition of isotope-labeled internal standards (isotope dilution MS) is an accurate method for quantitative analysis as well.

GC/MS applications are limited to analytes with sufficient volatility to be vaporized in the GC injector and oven; temperatures in these devices do not ordinarily exceed 300 °C. A wide variety of clinically important biomolecules do not meet this criterion, including proteins (enzymes, hemoglobin, immunoglobulins, most peptide hormones), lipids, and some steroid compounds. In some cases involving polar analytes, perfluorinated or trimethylsilyl (TMS) reagents can be used to generate volatile derivatives of these compounds, which can be separated on a gas chromatograph.

3 SPECIMEN PREPARATION

GC and MS applications both have specific specimen requirements; GC requires volatile specimens and MS requires pure specimens. As a consequence, biological specimens are rarely suited to analysis by GC or MS without some degree of preanalytical manipulation. Specimen preparation may involve extraction of the analyte from a complex matrix, separation of components by a chromatographic procedure, generation of volatile

derivatives, or dissolution into a special matrix prior to using a solid sampling technique.

3.1 Considerations for Biological Specimens

Most biological specimens are complex matrices with thousands of constituents, many of which may not be known. Blood is the most commonly analyzed biological specimen, and it contains both soluble and cellular components. Even within the noncellular fraction of blood (serum or plasma), there exist many highly lipophilic components suspended in micellar carriers. Water-soluble components of blood include small ions (e.g. sodium, potassium and chloride) as well as immunoglobulins with molecular weights exceeding 500 kDa. Many constituents in blood are complexed with other components; for example, about half of the calcium in blood is bound to albumin. Many drugs are bound to protein carriers, and some ions, such as iron and copper, exist almost exclusively in a complexed form. Finally, the concentration of blood components can vary from millimolar quantities (glucose, electrolytes, phosphate) to pico and femtomolar quantities (insulin, some toxins). Urine is also commonly analyzed in clinical laboratories, and contains many waste products of metabolism (such as urea), excess dietary components (sodium, phosphate), and other purposefully excreted bodily constituents, such as drugs or nonuseful dietary components. The concentration of the constituents in urine vary based on the rate at which they are filtered and excreted by the kidney, as well as the total volume of urine that is produced, which is affected by many factors. Hence, concentrations of normal components found in urine are generally more variable than the common constituents of blood, which are under a higher degree of metabolic control.

An additional consideration, when interpreting the concentration of constituents in biological matrices, is that the body is a heterogeneous matrix, so the concentration of a particular constituent in the blood may not necessarily reflect its concentration in other tissues. This is an important consideration for many clinically relevant compounds, such as therapeutic drugs, because they exert their principal effects at receptors in extravascular tissues. Hence, the ability to measure a constituent in the blood does not guarantee that the information will be clinically useful. It is also important to consider the way a compound is metabolized. This is especially true for drugs, many of which are quickly modified in vivo by hydrolytic or glucuronidation reactions.

Other biological matrices include extravascular fluids (peritoneal, pericardial, synovial fluids), cerebrospinal fluid, saliva, feces, sweat, and in some cases keratinized matrices such as hair or fingernails. Although these specimens are less commonly used for clinical analyses, each can be useful for a specific analytical purpose. Enzyme measurements in peritoneal fluid help distinguish between transudative and exudative processes, sweat chloride measurements are used to diagnose cystic fibrosis, abnormally high protein concentration in the cerebrospinal fluid is consistent with meningitis, and analyses of hair and fingernails can reveal historical exposure to drugs or certain heavy metals.

3.2 Preparation of Volatile Derivatives for Gas Chromatography

Organic compounds often contain carboxylic, hydroxyl, or amino functional groups that contribute to noncovalent intermolecular bonds, reducing the vapor pressure of these compounds. The low vapor pressure and, in some instances, thermal lability of such compounds restricts the application of methods such as GC, which require vapor-phase analytes. One approach to adapting compounds with low volatility to GC analysis is to synthesize derivatives that are more volatile. A diverse array of reagents is available for generating volatile derivatives. The most popular of these reagents are polymethylated silyl and perfluorinated compounds.

In a typical strategy, a derivatizing reagent is chosen to replace the hydrogen in a carboxylic acid or hydroxy functional group, generating the corresponding ester or ether derivative. Polymethylated silyl moieties impart considerable nonpolar bulk to otherwise polar species, thereby limiting intermolecular associations and increasing volatility and thermal stability. Perfluorinated derivatives produce much the same effect, although the increase in molecular size is less dramatic. The choice of derivatizing reagent usually depends on the facility of the derivatization reaction and also the chemical properties of the resulting complex. Electron capture GC detectors respond most sensitively to halide-containing compounds, whereas high-molecular-weight fragments are most useful for mass spectral identification. Acidic perfluorinated derivatizing reagents may react with the solid support of the chromatographic column, accelerating the deterioration of column performance. In GC applications, high-molecular-weight silyl derivatives may lengthen the retention time unacceptably. Overall, the most appropriate derivative will strike the best compromise between sensitivity and selectivity of the chromatographic method. When internal standards are involved, it is also important that the derivatization reaction does not selectively modify the analyte at the expense of standard, or vice versa.

3.3 Solid Sampling Techniques for Mass Spectrometry

The sampling requirements for MS are more demanding, since specimen purity is essential. However, the ion source

of a mass spectrometer is operated under vacuum, so compounds with low volatility are more accessible than in pressurized systems like GC. Many biological compounds, however, are thermally labile, so techniques must be devised to vaporize these analytes at low temperatures. Solid sampling methods are designed to produce vapor state molecules without sacrificing the integrity of the analyte, upon which mass spectral identification depends.

Most solid sampling techniques for MS rely on physical, rather than thermal, means to produce vapor state molecules. Electrospray ionization uses a concentrated coulombic charge to disrupt solvent/analyte droplets, leaving vapor-phase analyte molecules for mass analysis. Thermospray ionization is a similar technique, but evaporates the solvent. FAB methods rely on the kinetic energy of bombarding particles to eject analyte molecules from the solid matrix, and MALDI uses a laser to vaporize the analyte-embedded crystalline matrix.

4 APPLICATIONS TO DRUG MEASUREMENTS

Detection and measurement of therapeutic and illicit drugs in blood and urine is an important component of any clinical toxicology service. Although testing for illicit drug use is a common practice associated with drug-free workplace initiatives, clinical drug testing is distinct from forensic drug testing in several respects. Forensic drug testing is a highly regulated practice, in which the analytical specifications are set forth in guidelines published by the specific agency (federal, state, independent) that licenses or accredits the laboratory to perform workplace drug testing. Clinical toxicology services, however, adhere to standards established by agencies that license or accredit clinical laboratories. Whereas the regulation of forensic laboratories is intended to standardize practice across all such laboratories, clinical toxicology laboratories have more flexibility to configure their services to meet the specific needs of the patient population they serve. As an example, a truck driver who must submit to a drug test that is required by the Department of Transportation will have his urine tested in a federally certified laboratory for five different drug classifications. Federal regulations require that any positive result be confirmed by GC/MS analysis, and further specify the quantitative amount of each drug above which the test is to be reported as positive. In contrast, a physician evaluating a patient in an emergency department may only be interested in whether opiates or barbiturates are involved. In this clinical setting, confirmatory testing is usually unnecessary since the test result is parallel with the clinical condition of the patient. In this setting, the quantitative threshold for a positive result should be established based on the need for clinical sensitivity.

There are also situations in which toxicological tests ordered for medical reasons are used in legal proceedings, and in these cases the distinction between clinical and forensic drug testing is less clear. It would be routine for an emergency physician to request a blood alcohol on a trauma victim, in order to evaluate the risk of using anesthesia. However, if the trauma occurred as a result of an automobile accident, then the results of blood alcohol measurements are pertinent to civil or criminal charges of driving while intoxicated. Another common situation in which medically relevant drug tests can be used for legal proceedings is testing neonates for intrauterine drug exposure.

The vast majority of clinical drug tests, for both illicit and therapeutic drugs, are performed by immunochemical analyses, which are economical and easily automated. Commercial immunoassay reagents are not available for all clinically relevant drugs, though, and GC or GC/MS are attractive choices of analytical methods for measuring drugs when automated methods do not exist. GC and GC/MS methods have an additional advantage over immunochemical screening methods, in that they can be configured to screen for multiple drugs; immunoassays typically screen for a single drug or class of drug (benzodiazepines, amphetamines, etc.).

4.1 Drug Screening by Gas Chromatography

4.1.1 Alcohols

Gas chromatographic analysis requires volatile analytes, and ethanol is a common example of a highly volatile drug found in biological specimens. Ethanol is the most widely used alcohol, but clinical situations arise in which analyses for methanol, isopropanol, or ethylene glycol are also important.

Numerous methods have been published for measuring ethanol in blood.[2,3] Typical features of a GC ethanol method are summarized in Table 2. Since alcohol is a volatile constituent in biological matrices, interference from other matrix components is not a serious problem. GC methods for alcohol can ordinarily be used for whole blood, serum or plasma, or urine. In many states, the legal statutes that establish alcohol concentrations above which a person is considered intoxicated (per se laws) specify whole-blood alcohol. Because whole blood has a higher lipid content, its alcohol content is 10–15% less than serum or plasma, from which cellular components have been removed. Urinary alcohol measurements should be used for screening purposes only, since the concentration is rarely of clinical relevance.

In some GC methods for ethanol, the specimen is injected directly into the chromatographic column. This

Table 2 Gas chromatographic measurement of ethanol

Specimen type	Whole blood, serum or plasma, urine; whole blood is preferable, since per se laws typically refer to whole-blood alcohol (see text)
Injection	Can be injected directly, or headspace sampling
Column	Either packed or fused-silica capillary column; stationary phase may be polar (e.g. Carbowax) or nonpolar
Detector	FID is most common
Standardization	Internal or external standards may be used; internal standard should be nonconsumed alcohol, such as 1-propanol (methanol or isopropanol would not be appropriate internal standards)

is a simple procedure since it requires no sample preparation, but nonvolatile constituents in the specimen may foul the chromatographic column prematurely. The preferred method involves headspace analysis. Headspace analysis for alcohol typically involves incubating a specimen (60 °C is common) in a closed container until equilibrium is reached and sampling (with an airtight syringe) the vapor above the liquid. The headspace vapor is introduced into the chromatographic column, and analysis proceeds the same as if liquid had been injected. Headspace analysis is particularly useful for highly volatile solutes, since higher vapor concentrations increase the sensitivity of the measurement. Among the toxicologically relevant alcohols, ethylene glycol is the only toxicant that is insufficiently volatile for headspace GC analysis.

Alcohols may be separated with a variety of stationary phases, and with packed- or fused-silica capillary GC columns. Polar stationary phases such as Carbowax (poly(ethylene glycol)) are often used to separate alcohols, but relatively nonpolar phenylsilicone or dimethylsiloxane stationary phases can also be used. Isothermal chromatographic analysis of ethanol is usually performed between 50 and 100 °C, or the oven temperature can be programmed to increase, possibly to expedite the analytical cycle time. Helium is a common choice of carrier gas. FID provides the best combination of sensitivity and economy for GC measurement of hydrocarbons.

Although external calibration of a GC alcohol method is possible, and may even be preferable in certain circumstances (e.g. heavy workload volume), the matrix in which standards are prepared should match, as closely as possible, the specimen matrix. This requirement can be difficult to meet in biological specimens such as blood, since the exact composition is not known, and it can vary from one specimen to another. A better method is the addition of an internal standard (1-propanol is a common choice).

A GC method has been described for measuring ethanol, methanol, acetaldehyde, and acetone in blood and urine;[4] the method also measures ethanol in fecal specimens. Whole blood, urine, or serum was injected directly, without any pretreatment, on to a polar stationary phase in a packed column. Buildup of nonvolatile material in the chromatographic column was minimized by using an 8-cm pre-column packed with glass beads and stoppered with dimethylchlorosilane treated glass wool. Nitrogen was used as the carrier gas, and a FID measured the compounds. The method was calibrated against external aqueous standards. Measurement of ethanol in feces involved homogenization and centrifugation to recover a liquid supernatant, which was analyzed in the same manner as above.

Ethylene glycol is widely available as the principal ingredient in antifreeze, and can be ingested accidentally by children or pets, or intentionally by adults for suicidal or alcoholic substitution purposes. Ethylene glycol has a variety of toxic effects in humans, most notably renal failure caused by precipitation of oxalate and hippurate crystals from the urine.[5] The methods applied to GC measurement of ethanol, methanol, and isopropyl alcohol are not applicable to ethylene glycol because of its low volatility.[6] One published method for GC measurement of ethylene glycol[7] in serum involved pretreating the specimen with acetonitrile to precipitate proteins. An internal standard, 3-bromo-1-propanol was added to the specimen. Ethylene glycol and its metabolite, glycolic acid, were derivatized with *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) before analysis. Chromatographic separation was performed on a 5% phenylmethylsilicone coated fused silica capillary column, and the oven temperature was increased from 90 to 170 °C at a rate of 15 °C min^{-1}. Split injection (1 : 20 split ratio) was used, and ethylene glycol and glycolic acid were measured with a FID.

4.1.2 Multidrug Profiles

Another useful application of GC in a clinical toxicology service is for multidrug profiling.[8] When patients present with altered mental status, particularly in an emergency room setting, clinicians must include drug intoxication in their differential diagnosis. Conventional automated methods for drug screening have a limited scope. Immunoassays are available for most abused drugs, but are not always sensitive for every drug within a particular classification. Thin-layer chromatography is sometimes used for drug screening, but it is strictly a qualitative method and considerable experience is required to interpret the results. High-performance liquid chromatography (HPLC) with ultraviolet detection has also been applied to multidrug screening, but its use is not widespread.

The performance of flame ionization and nitrogen–phosphorus detectors have been compared, using reference standards, with respect to their suitability for drug screening by GC.[9] For nitrogen-containing compounds, the nitrogen–phosphorus detector increases sensitivity by one or two orders of magnitude, compared to the FID. Retention times of drugs separated on packed columns versus fused-silica capillary columns have also been evaluated (also with reference standards), and the results were determined to be comparable.[10] Choice of a detection system may depend primarily on availability and cost; a nitrogen–phosphorus detector is more expensive, but eliminates interferences from non-nitrogen-containing substances in the specimen. Fused-silica capillary columns offer important advantages, including convenience and higher resolution.

Numerous methods have been described for screening blood or urine specimens for drugs by GC.[11–13] Typical features of these methods include extraction of drugs into an organic solvent, separation on a moderately polar chromatographic column, and measurement with a FID or nitrogen–phosphorus detector. Retention indices are used to identify the different drugs.

Extraction of the drugs prior to analysis is necessary due to the many nonvolatile constituents of blood or urine, which can quickly result in a buildup of residue in the small bore of a fused silica capillary column. Extraction procedures for basic drugs are usually designed to generate the free base (uncharged) form of the drug, which is soluble in a nonpolar organic solvent. Adjusting the pH of the specimen to a point above the pK_a of the drug is necessary; the converse is true for extracting acidic drugs. Neutral drugs should be extracted in either procedure. Liquid–liquid extraction steps can be performed as many times as necessary to generate a sufficiently pure sample, but the extraction inefficiency will ultimately compromise the sensitivity of the method. Liquid–solid extractions are also useful, and will be discussed later in this article. An attractive feature of extraction steps, in addition to separation of analyte from other matrix constituents, is that the solvent can be evaporated, thereby concentrating the analyte and improving sensitivity. Because analytical sensitivity is an important consideration in drug screening, splitless injections are preferred, but not essential.

Temperature programming is usually necessary in order to elute analytes with a wide range of retention indices within a reasonable period of time. One approach to improving the chromatographic characteristics of the detectable drugs is to generate methylated derivatives, which reduce the overall polarity of individual analytes. A method for generating methylated derivatives in the chromatographic column has been described.[12] On-column derivatization involves mixing the specimen with a derivatizing reagent – tetramethyl ammonium hydride is a common choice – so that the heat of the injector will cause the derivatizing reaction to take place. This method is particularly useful for acidic drugs, which have poor chromatographic characteristics when the carboxylic acid is left exposed.

4.1.3 Gas Chromatography as a Reference Method

GC is a sophisticated analytical technique that ordinarily requires some sample manipulation and a considerable degree of user training and experience in order to produce reliable results. Most clinical laboratories favor automated methods, due to their higher throughput. For this reason, GC applications in clinical laboratories – even clinical toxicology services – are not widespread, but can be useful adjuncts when personnel and clinical needs are sufficient to justify the investment. Perhaps a more appropriate use of GC in a clinical laboratory is to verify the accuracy of routine automated methods. An example taken from the literature[14] involves comparison of four automated methods for measuring lactate with gas chromatographic quantitation. Lactic acid was extracted from serum and the butyl derivative was synthesized before GC analysis on a capillary column coated with methylsilicone as the stationary phase. The lactate results obtained by GC analysis, which was regarded as the reference method, were compared to four commercial automated lactate assays; regression slopes were between 0.99 and 1.06, with standard errors of the estimate between 0.12 and 0.23.

4.2 Quantitative Applications of Gas Chromatography/Mass Spectrometry

The most common application of GC/MS in a clinical setting is for toxicological analyses, most notably the confirmation and quantitation of drugs in specimens that had previously had positive immunoassay screening results.[15] Confirmatory analysis by GC/MS is required in forensic drug testing services, and can also be an optional enhancement to clinical toxicology services. The value of confirmatory analyses for drugs is not as great in a clinical setting, though. For example, when a positive test for benzodiazepines (a class of sedative/hypnotic drugs) is obtained in the urine of a job applicant, it is important to know which particular drug is present, in the event that the applicant may have a physician's prescription for one of these drugs. However, an emergency physician who is concerned whether a patient's depressed mental state might be due to oversedation may not require information about which specific benzodiazepine is present, since it is not likely to influence the treatment of the patient. Numerous applications have been described for measuring drugs

Table 3 Selected GC/MS reference methods

Analyte	GC/MS method features	Clinical importance
Creatinine[18]	Isotope dilution; filtration to remove proteins; purification with HPLC; MBDSTFA derivative; quadrupole MS	Creatinine is a waste product of muscle catabolism that is commonly measured to evaluate renal function
Cortisol[20]	Isotope dilution; extraction into dichloromethane; purification by column chromatography; HBFA derivative; quadrupole MS	Cortisol is a mineralocorticoid emanating from the adrenal gland; it is measured to evaluate adrenal function
Triglycerides[25]	Radiolabeled standard; extensive specimen preparation including alkalinization, liquid–liquid extraction, solid-phase extraction; BSTFA derivative; magnetic sector MS; chemical ionization	Serum triglyceride concentrations have been correlated with risk for coronary artery disease
Cholesterol[23]	Internal standard; hydrolysis; extraction; BSTFA derivative; quadrupole MS	Sterol derivative found in all animals; associated with risk for coronary artery disease

by GC/MS, but the vast majority of these are used for forensic drug testing. In this article, the use of GC/MS as a reference quantitative method will be emphasized, as well as its application to esoteric analytes.

Table 3 summarizes the features of several recently published methods using GC/MS as a reference technique. Reference methods are not devised with expediency or convenience as a goal, and some of these methods are quite complex. Reference methods are intended to provide a means to validate routine methods, not for use as routine methods. Therefore, accuracy is the principal goal in the design of a reference method.

4.2.1 Creatinine

Creatinine is a byproduct of muscle catabolism, and its concentration in blood is roughly proportional to muscle mass. Creatinine has no physiological function, and is continually excreted by the kidneys. Therefore, it is a convenient indicator of renal function, since loss of kidney viability will result in increased concentrations of creatinine in serum. Measurement of urinary creatinine concentration can also be useful in assessing renal function. In the late nineteenth century, Jaffe discovered that creatinine reacted with alkaline picric acid to form a bright yellow complex, and this reaction formed the basis of an analytical method for measuring creatinine that is still used today. The Jaffe reaction is notoriously nonspecific; picric acid reacts with a wide array of normal and abnormal constituents of urine and serum.[16] Hence, a reference creatinine method is especially helpful in assessing the accuracy of various methods based on the Jaffe reaction. HPLC has been applied to reference measurements of creatinine.[17] In one method combining HPLC with GC/MS,[18] creatinine was isolated from serum by deproteinization followed by HPLC, and the *tert*-butyldimethylsilyl derivative was synthesized before GC/MS analysis. The special purification procedure was necessary to remove any creatine, which is the hydrated form of creatinine. An isotopically labeled internal standard was used to calibrate the method.

4.2.2 Cortisol

Cortisol is a steroid hormone produced by the adrenal cortex that has mineralocorticoid and glucocorticoid activity. The principal function of cortisol is the regulation of carbohydrate metabolism. Excess cortisol production, which can result from several causes, produces a set of clinical symptoms known collectively as Cushing's syndrome. Subnormal cortisol excretion is consistent with adrenal hypofunction, a clinical condition called Addison's disease, but more sensitive indicators exist for this condition. Measurement of cortisol in serum is an important factor in evaluating patients with suspected adrenal abnormalities.[19] A variety of immunochemical methods are available for cortisol measurement, and several of them were compared with quantitative cortisol results obtained by GC/MS.[20] Cortisol was extracted into dichloromethane after alkalinization, and an isotopically labeled internal standard was added for quantitation. The extract was purified by liquid chromatography, and derivatized with heptafluorobutyric anhydride prior to GC/MS analysis. The comparison revealed significant discrepancies between various commercial immunoassays for cortisol, underscoring the importance of a reliable reference method for this analyte.

4.2.3 Cholesterol and Triglycerides

Coronary artery disease is the leading cause of death in most industrialized societies, and considerable effort has been directed toward identifying individuals susceptible to coronary disease. The most widely used biochemical indicators of risk for coronary artery disease are serum cholesterol and triglycerides. Since these compounds are so frequently measured, and the results used to

approximate risk for disease, it is important that the methods used to quantify cholesterol and triglycerides be as accurate as possible. Reference methods are helpful in standardizing results across different analytical methods and laboratories.

The consensus reference method for cholesterol is based on a method described by Abell et al.[21] in 1952, which involves alkaline hydrolysis of cholesterol esters, extraction into hexane, and reaction with sulfuric acid to produce a colored adduct. Reference standards created by the Centers for Disease Control are based on this method.[22] However, several excellent GC/MS procedures for cholesterol measurement have been described, and it is likely that GC/MS will eventually supplant the manual Abell technique as the preferred reference method. In a recently described procedure,[23] isotopically labeled cholesterol was added to a serum specimen before hydrolysis with methanolic potassium hydroxide at 70 °C. Free cholesterol was extracted into hexane, and derivatized with *N*-trimethylsilyl-*N*-methyltrifluoroacetamide before GC/MS analysis. The accuracy of the method was verified using a standard reference material obtained from the National Institute for Standards and Technology. Regression slopes for correlation of the GC/MS procedure with four automated cholesterol methods were between 0.97 and 1.03.

Triglycerides present a more difficult analytical challenge, since they comprise a heterogeneous population of acylglycerol species. The Centers for Disease Control recognize a reference method for triglycerides, based on extraction, alkaline hydrolysis to generate glycerol, and oxidation of glycerol to formaldehyde; formaldehyde reacts with chromotropic acid to form a colored adduct.[24] As with cholesterol, a GC/MS procedure may eventually supersede the wet chemical method. A recent example of a GC/MS method[25] involves extraction of triglycerides into a chloroform–methanol solvent, followed by solid-phase extraction. Specimens were hydrolyzed with ethanolic potassium hydroxide at 70 °C for 2 h, and isotopically labeled tripalmitin was used as the internal standard for calibration and quantitation. *N*-Methyl-*N*-trimethylsilyltrifluoroacetamide was used to generate the TMS derivative of glycerol before GC/MS analysis. The identity of the derivative was verified by negative ion/chemical ionization MS, which produced a protonated molecular ion. For triglyceride standards, the GC/MS method produced results that deviated less that 0.3% from expected values.

4.2.4 Therapeutic and Abused Drugs

GC/MS confirmation of positive screening results for abused drugs is commonplace, whereas immunochemical measurements of therapeutic drugs are rarely confirmed. There are, however, GC/MS applications for less commonly abused drugs and therapeutic agents for which immunochemical methods are not available. Three examples will be considered here: lysergic acid diethylamide (LSD), busulfan, and mannitol/sorbitol.

LSD is a semisynthetic hallucinogenic drug that is derived from ergot alkaloids. Detection and measurement of LSD is difficult because of the potency of the drug; a typical dose may be between 40 and 120 μg, and the parent compound is extensively metabolized.[26,27] A method has been described for GC/MS measurement of *iso*-LSD,[28] which is a co-product in the hydrolysis reaction that produces LSD from isomeric ergot precursors. Prior to analysis, *iso*-LSD was extracted into 1-chlorobutane after alkalization, and the *iso*-LSD was converted to LSD by a hydrolytic isomerization reaction using sodium ethoxide. The product was purified by solid-phase extraction, then liquid–liquid extraction prior to derivatization with BSTFA. GC/MS analysis included separation on a 5% phenylmethyl siloxane capillary column, and electron impact ionization. Quantitation of *iso*-LSD was based on a nonisotopic internal standard, lysergic acid methylpropylamide, and the method was capable of detecting *iso*-LSD at a concentration of 50 ng per liter of urine.

Busulfan (1,4-butanediol dimethanesulfonate) is a chemotherapeutic agent that is used for marrow ablation prior to bone marrow transplant, but in high doses it can be toxic to the liver and central nervous system.[29] These adverse effects are relatively common, and frequently fatal.[30] In an effort to correlate plasma concentrations of busulfan with clinical outcomes, a GC/MS method was developed for measuring busulfan in the plasma of patients being prepared for bone marrow transplant.[31] Pusulfan (1,5-pentanediol dimethanesulfonate) was added to the plasma specimens as internal standard before extracting the drugs into ethyl acetate. Busulfan and pusulfan were iodinated prior to GC/MS analysis. The chromatographic separation was on a nonpolar capillary column (methyl silicon), and electron impact ionization was used for mass spectral analysis. Quantitative analysis of busulfan had a between-day imprecision of approximately 8%, and the limit of detection corresponded to 2 μg per liter of plasma.

Mannitol and sorbitol are often added to irrigation fluids used in endoscopic procedures to adjust their osmolality.[32,33] Absorption of these compounds can lead to complications, so it may be important to monitor mannitol and sorbitol concentrations in patients undergoing certain endoscopic procedures.[34] A method has been described[35] for GC/MS measurement of mannitol and sorbitol using dulcitol as internal standard. Specimens were deproteinized with ethanol, and the *n*-butyldiboronate derivatives were generated by reaction with *n*-butyl boronic acid. Chromatographic separation

was sufficient to resolve mannitol and sorbitol peaks from several potentially interfering compounds, including D-galactose, D-arabinose, D-ribose, and D-xylose. The mass spectrometer was set for electron impact ionization, and identification of mannitol and sorbitol derivatives was based on the fragment $m/z = 253$.

4.2.5 Fatty Acids

Fatty acids are important components of metabolic reactions, and concentrations in biological specimens can be affected by the presence of metabolic disorders or dietary deficiencies. The general structure of fatty acids is R−COOH, where R is an alkyl chain. Because fatty acid structures are analogous, measurement of individual compounds in this group by chemical analysis is difficult without first separating the different fatty acids. GC/MS is well suited to fatty acid analysis, because it involves both separation and specific detection based on the mass of the analyte.

Oxidation of fatty acids occurs in the mitochondria, and the ultimate product of fatty acid catabolism is acetyl-coenzyme A (CoA), which enters the Krebs cycle. Fatty acid β-oxidation is an important source of energy when dietary sources are unavailable or insufficient to meet physiological demands. There are several genetic disorders that result in compromised to nonexistent mitochondrial β-oxidative capacity, and these disorders often go unnoticed at birth since fatty acid utilization is negligible except under circumstances of dietary restriction or unusual energy expenditure.[36] Measurement of fatty acids in plasma or urine can help reveal metabolic deficiencies in the oxidation of these compounds.

In one recently described method,[37] C-6 through C-18 fatty acids were measured by a GC/MS procedure in plasma obtained from patients with various mitochondrial fatty acid β-oxidation defects. Hydroxylated fatty acid analogs (3-hydroxyhexanoic acid and 3-hydroxyoctanoic acid) were used as internal standards. Plasma specimens were acidified before extracting the fatty acids into ethyl acetate. The reactive carboxylic acid ends of the analytes, and free hydroxy groups of the internal standards, were derivatized with *N*-(*tert*-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (MTBSTFA) and *N*-methylbis(trifluoroacetamide) (MBTFA) in pyridine, and the mass spectra were generated with electron impact ionization. This method was a modified version of a previously described procedure involving bis-trimethyl derivatives.[38]

In another procedure,[39] C-12 through C-22 fatty acids were measured in plasma from patients with three types of disorders: medium-chain acyl-CoA dehydrogenase deficiency; very long-chain acyl-CoA dehydrogenase deficiency; and mild-type multiple acyl-CoA dehydrogenase deficiency. Nonanoic and nonadecanoic acids were used as internal standards. The fatty acids were methylated[40] and extracted into hexane prior to GC/MS analysis on a nonpolar column with electron impact ionization.

A method has also been described[41] for GC/MS measurement of fatty acid ethyl esters in order to quantify the activity fatty acid ethyl ester synthase, which is thought to be released in alcoholic liver and pancreatic disease.[42] Although enzymes can be measured directly, most assays quantify enzyme activity by measuring the rate at which substrate is consumed or product is generated. Enzymes are important biological markers, because their release into the blood can be a sensitive indicator of tissue damage. Ethanol abuse is common in western societies, but the mechanisms by which it causes organ damage are not well understood. Fatty acid ethyl ester synthase activity was quantified by measuring its product by GC/MS. Ethanol and palmitic, stearic, oleic, and arachidonic acids were added to serum specimens and incubated for 4 h at 37 °C. Following incubation, the fatty acid ethyl esters were extracted into a chloroform–methanol phase, and ethyl heptadecanoic acid was added as internal standard. Fatty acid esters were purified by thin-layer chromatography before GC/MS analysis.

4.2.6 Urinary Organic Acids

There are numerous enzymes involved in metabolic pathways using amino acids to synthesize compounds of biological importance, and genetic deficiencies in these enzymes result in the accumulation, and excretion, of large amounts of the precursor compounds. Collectively, many of these inherited disorders are known as organic acidurias, since the appearance of abnormal concentrations of organic acids in the urine is used to detect the disease. GC/MS is an attractive technique for urinary organic acid profiling, owing to its resolution, specificity, and sensitivity. Several methods for measuring urinary organic acids have been described; a few of these are summarized in Table 4.

Methylmalonic aciduria can result from inhereted methylmalonyl-CoA mutase deficiency or defects in cobalamin (vitamin B_{12}) metabolism.[43] Vitamin B_{12} can be measured in patients with suspected genetic defects in cobalamin metabolism, but the test is not specific.[44] Measurement of methylmalonic acid in blood or urine is a more sensitive test for disorders associated with vitamin B_{12} metabolism. A method has been described for measurement of methylmalonic acid in urine collected and dried on filter paper.[45] Measured sections of filter paper were rehydrated, and organic acids were partitioned into ethyl acetate after acidification of the aqueous filter paper extract. TMS derivatives were synthesized

Table 4 GC/MS methods for detecting organic acidurias

Analyte	Derivative	Internal standard	Reference
Methyl-malonic acid	BSTFA/TMCS	Isotopic	McCann et al.[45]
Orotic acid	BSTFA/TMCS	Isotopic	McCann et al.[50]
Profile	Ethoxylamine HCl, BSTFA and perdeuterated BSTFA	External calibration	Shaw et al.[53]
Profile	Hydroxylamine, BSTFA/TMCS	External calibration	Duez et al.[52]

TMCS, trimethylchlorosilane.

prior to GC/MS analysis on a 5% phenylmethylsilicone capillary column and electron impact ionization. Quantitative measurements of urinary constituents can be difficult to interpret since their concentrations can vary depending on the total volume of urine produced. Because the amount of excreted creatinine is relatively constant, its concentration is sometimes used to compensate for variations in total urinary output. In this method for urinary methylmalonic acid, creatinine was measured in the aqueous filter paper extract, and the acid concentration was expressed relative to creatinine concentration. Trideuterated methylmalonic acid was added as internal standard.

Inherited deficiencies in urea cycle enzymes usually result in an accumulation of intermediary compounds. One of these intermediates, orotic acid, is increased in several urea cycle disorders, the most common being ornithine transcarbamylase deficiency.[46] Several methods for chromatographic or GC/MS measurement of orotic acid in urine have been described.[47–49] In a procedure involving collection of dried urine specimens,[50] soluble urine components were eluted from measured sections of filter paper by shaking with 5 mL of water, and creatinine was measured in the eluate for reasons explained above. An isotopically labeled internal standard was added to the extract, and organic acids were extracted into ethyl acetate after acidification. TMS derivatives were synthesized using BSTFA/TMCS, and the analytes were separated on a 5% phenylmethylsilicone capillary column. Mass spectra were obtained by electron impact ionization.

More general approaches to diagnosing inherited metabolic disorders involve methods similar to drug screens, where the intent is to detect multiple constituents and reveal unusual concentrations when they exist. HPLC has been applied to screening urine for inherited metabolic disorders,[51] and more recently GC/MS analysis of volatile derivatives of organic acids in urine has been used for this purpose. The difficulties associated with designing a useful analytical procedure increase in proportion to the number of analytes it is intended to measure, and the challenge of diagnosing inherited metabolic disorders is considerable. A recent paper described a typical approach to quantitative profiling of urinary organic acids.[52] After addition of nonphysiological internal standards (2-ketocaproic and *dl*-tropic acid) urine specimens were incubated with hydroxylamine for 30 min, then acidified with HCl before extracting organic acids into ethyl acetate. The organic acids were further purified by two cycles involving addition of ammonia–ethanol and re-extraction. TMS derivatives were generated using BSTFA/TMCS–pyridine before GC/MS analysis. Twelve organic acids were quantitated by comparison of the ion intensity for characteristic fragments with the internal standards. Coefficients of variation ranged from 2.8% to 20.8% (within-run), and recoveries from standard solutions varied from 78% to 251%; most of the organic acids, however, had acceptable recoveries between 90% and 110%. A few of the organic acids exhibited a nonlinear response, which may explain the highly variable recovery data.

In another urinary organic acid profiling procedure,[53] ethoxime derivatives were synthesized by addition of ethoxylamine HCl prior to acid extraction and derivatization with BSTFA/TMCS. A novel feature of this method was the use of perdeuterated BSTFA in order to identify unknown constituents that appeared in the chromatogram based on the ion shift observed in the characteristic fragment. Using this procedure, the Krebs cycle intermediates, citramalic, tartaric, and 3-oxyglutaric acids, were identified in urine from two patients diagnosed with autism. Arabitol and arabinose were also identified.

4.2.7 Blood Lead

GC/MS is most useful for qualitative and quantitative analysis of organic molecules, but this versatile technique can be applied to nonorganic analytes as well. An example of clinical and toxicological interest is lead. Lead is a potent biological toxin, exposure to which can occur from many sources, such as paints, ceramics, plumbing, and environmental contamination. Lead exposure is particularly damaging in childhood, since it may result in central nervous system toxicity, developmental delays, and learning disabilities.[54] In adults, lead poisoning is manifested primarily in hematological disorders, including anemia, and hypertenstion. Screening of school-age children for lead toxicity has been suggested in order to identify children at risk for lead-related problems.

Lead inhibits the activity of amino levulinic acid hydratase, an enzyme involved in the synthesis of heme from porphyrin, and the accumulation of protoporphyrin

in red blood cells is an indication of lead exposure, although it is less sensitive than direct measurements of lead in whole blood. The most common methods for measuring lead in blood involve electrothermal atomic absorption spectrometry, or electrochemical methods.[55] However, a GC/MS method has been described[56] for measuring lead in blood. A low-abundance isotope of lead (^{204}Pb; natural abundance 1.48%) was used as internal standard, and the method was calibrated against pure and matrix-matched standards. Whole-blood specimens were digested with concentrated nitric acid, and the free lead ions were complexed with ammonium pyrrolidine dithiocarbamate. The lead complexes were extracted into toluene, and 4-fluorophenylmagnesium bromide (Grignard reagent) was added to generate the $Pb(FC_6H_4)_4$ complex. The complex was measured by negative ion/chemical ionization MS using methane as the reagent gas. Typical within-run imprecision for the method was 0.5%, and quantitative results were possible at concentrations as low as 1 μg per liter of blood. The general approach, involving complexation, extraction, and GC/MS analysis with an isotopic internal standard is adaptable to other heavy metals.[57]

5 CLINICAL APPLICATIONS OF MASS SPECTROMETRY

It has been previously noted in this article that mass spectrometric analysis requires pure specimens, and this limitation complicates its application to bioanalyses, where complex matrices are commonplace. There are strategies for overcoming this limitation, however, and these techniques have been applied successfully to detection and quantitation of biomolecules in complicated matrices. One strategy involves the use of high-resolution inlet systems; GC is but one example that was discussed in the previous section. Another analytical strategy involves the use of tandem mass spectrometers, the first stage of which selects a pertinent fragment and the second stage measures daughter ion fragments produced therefrom. Also, mass spectrometric methods designed to measure large molecules using soft, or low-energy, ionization techniques are less affected by impurities in the specimen matrix since large fragments with unique masses are often used for identification and quantitation. The potential of mass spectrometric methods for clinical applications have been stressed by several reviewers,[58] and a few representative examples of this promising technology will be discussed here.

5.1 Screening for Inborn Errors of Metabolism

Like the GC/MS applications reviewed in the previous section, MS alone has been applied to screening for inborn errors of metabolism.[59] These screening methods ordinarily involve detection of accumulated metabolic intermediates that result from deficiencies in enzymes necessary for biosynthesis of essential biochemicals. One versatile profiling technique utilizing MS with electrospray ionization[60] involved spotting blood specimens on a piece of filter paper for stable transport to the laboratory. A disk was cut from the filter paper, which was immersed in a methanol solution containing isotopically labeled standards for 13 organic acids. After cold incubation, the methanolic extract was dried and reconstituted in butanolic HCl and incubated at 65 °C for 20 min. The derivatized extract was washed with hexane and dried prior to preparation for dissolution in the electrospray solvent, consisting of an acetonitrile–water mixture (80 : 20). A triple-quadrupole mass spectrometer was used to sequentially select characteristic masses in the first sector, to allow argon-collisional fragmentation in the middle sector, and to allow daughter ion analysis in the final sector. A computer algorithm was developed to tabulate the mass spectral data, perform quantitative calculations, and flag abnormal results. This method facilitated the detection of a wide array of metabolic disorders, including phenylketonuria (PKU), maple syrup urine disease, and methioninopathies. A significant advantage of quantitative techniques such is this one is a notable increase in the specificity of the procedure – falsely positive results are substantially reduced.

Quantitative methods for urinary constituents are inherently problematic because urine volume varies with such factors as liquid intake and renal function. The variability in concentrations of urinary constituents can be compensated, to some degree, by measuring the ratio of one constituent to another, which is less affected by changes in total urinary output. This approach has been used in mass spectrometric procedures designed to detect PKU. PKU is an inherited deficiency in amino acid metabolism that can result in mental retardation unless dietary restrictions are implemented at a very early age. Because PKU is comparatively common among inborn metabolic errors, although its incidence is still less than 0.01%, newborn urine screening for this disease has been widely adopted. As with any screening test for a rare disease, though, a slight underperformance in specificity results in a large number of falsely positive results, and low predictive value.[61] Falsely positive results of screening tests for heritable diseases can result in unnecessary anxiety and expensive follow-up testing. A recent report[62] describes a mass spectrometric method for detecting PKU. The clinical sensitivity and specificity of the screening test was improved by using the ratio of phenylalanine to tyrosine. Tandem MS (MS/MS) was used to quantify phenylalanine and tyrosine

following initial screening for hyperphenylalaninemia by a fluorometric method. The MS/MS procedure confirmed PKU in less than 2% of specimens classified as positive by fluorometry. Other studies have produced similar results.[63]

MS/MS has been applied to the detection of maple syrup urine disease, a rare disorder that traces its name to the sweet odor of accumulated branched-chain amino acids in the urine. The disease is caused by a deficiency in branched-chain α-keto acid dehydrogenase activity, which is responsible for the oxidative decarboxylation of leucine, isoleucine, and valine. Measurement of blood leucine can reveal maple syrup urine disease at an early age when dietary restrictions are helpful in preventing the severe mental and physical complications that accompany the disease. In the MS/MS procedure,[64] butyl derivatives of amino acids extracted from filter paper blood spots were prepared by reaction with acidic 1-butanol. Deuterated alanine, valine, leucine, methionine, phenylalanine, tyrosine, and glutamine, as well as $[^{15}N_1,^{13}C_1]$glycine, were added as internal standards. The derivatized specimens were reconstituted in a methanol–glycerol solution prior to MS/MS analysis. In the tandem mass spectrometer, a cesium gun was used to vaporize and ionize analyte molecules, and relevant positively charged fragments were selected in the first mass analyzer region. Selected fragments were exposed to argon in the second quadrupole region, and collision-induced decomposition fragments were measured in the third mass analyzer. Predictable neutral-fragment losses in the collision sector yielded high-abundance ions that were used for quantitation. Concentrations of leucine, isoleucine, and valine were expressed relative to phenylalanine concentration. Analytical recoveries of leucine and valine ranged from 86% to 105%, and the method imprecision was less than 11.5% (between-run).

Two inherited metabolic disorders can result in accumulation of methionine in blood: homocystinuria and isolated hypermethioninemia. Homocystinuria results from cystathionine β-synthase deficiency, and hypermethioninemia is most often caused by a deficiency in methionine adenosyltransferase enzyme. Both disorders can result in neurological and skeletal abnormalities. Dietary restrictions, and sometimes pyridoxine supplementation, can improve the prognosis when affected individuals are identified at an early age.[65] An MS/MS technique has been described[66] for measuring methionine in filter paper blood spots, and the method can be used to screen newborns for the associated genetic disorders. In this procedure, amino acids are eluted with a methanol solvent containing deuterated internal standards. The method is substantially the same as the procedure for the branched-chain amino acids leucine, isoleucine, and valine, described above, including derivatization with acidic 1-butanol and use of deuterated internal standards. FAB (cesium ion gun) was used to vaporize and ionize the analytes from a methanol/glycerol solvent. The specimens used to develop this method had been previously screened by a bacterial inhibition assay (Guthrie test). The best precision and correlation with bacterial inhibition results was obtained when quantitative results for methionine quantitation were expressed relative to the sum of leucine and isoleucine concentrations.

Smith–Lemli–Opitz (SLO) syndrome is an autosomal recessive disorder that is characterized by mental retardation, facial dismorphism, and microcephaly, as well as cardiovascular and urogenital malformations.[67] It has been suggested that SLO syndrome results from a deficiency in sterol Δ^7-reductase, which catalyzes the conversion of 7-dehydrocholesterol to cholesterol.[68] As a result, high concentrations of 7-dehydrocholesterol are found in SLO syndrome; the structure of this metabolic intermediate differs from cholesterol only by two hydrogen atoms. A mass-spectrometric method has been described for measuring 7-dehydrocholesterol,[69] using a time-of-flight MS instrument. Neutral sterols were extracted from whole blood into hexane, and applied directly to the MS target platform. An argon beam volatilized the analytes, which were separated in the time-of-flight mass analyzer. Stigmasterol was added as internal standard for quantitation. The method was able to detect 7-dehydrocholesterol in 3 patients diagnosed with SLO syndrome, but not in 10 normal control subjects.

5.2 Protein Measurement

Measurement of proteins involves special analytical challenges, because the chemical composition of proteins consists of only 20 amino acids. Therefore, all proteins have inherently similar chemical compositions. The physical properties of proteins are largely determined by secondary, tertiary, and quaternary structures, which establish the spatial orientation of the primary amino acid chain. Purification techniques that separate components based on size have been used for protein measurement, but these methods are cumbersome and usually imprecise. Certain proteins have unique properties that can be exploited in an analytical method – certain dyes, for example, react with anionic proteins. Most specific proteins are measured by immunochemical methods, but antibody specificity is not absolute, and the chemical similarities between proteins exacerbate the problem of cross-reactivity. For these reasons, reference methods for specific protein quantitation are not widely available, and MS has been proposed as a method to fill that void.

5.2.1 Measurement of Trypsin Cleavage Products

Since measurement of high-molecular-weight species requires specially adapted MS instruments, most approaches to MS protein identification and quantitation involve measurement of smaller protein fragments produced by digestion prior to analysis.[70,71] A general method has been described[72] for quantifying specific proteins by isotope-dilution MS, and a prototype application was designed to measure apolipoprotein A-1 in a standard reference material. In this procedure, the protein is digested with trypsin, which reproducibly cleaves two fragments containing 7 and 11 amino acid residues that are unique to the apolipoprotein A-1 parent molecule. Analogous deuterated oligopeptides were synthesized and used as internal standards. Trypsin hydrolysis was allowed to proceed for 24 h at 37 °C, after which acetic acid was added to quench the reaction. The digestion products were purified by HPLC before measurement on a tandem magnetic sector MS, using FAB ionization (cesium ion gun). The fragments used for MS analysis were characterized by peptide mapping and sequencing by HPLC and nitrogen analysis.

5.2.2 Plasma Renin Activity

Renin is produced in the juxtaglomerular cells in the kidney, and its function is to catalyze the cleavage of angiotensinogen to produce the decapeptide angiotensin 1. Angiotensin-converting enzyme cleaves two amino acid residues from angiotensin 1 to produce angiotensin 2, which is the physiologically active vasopressive form. Angiotensin 2 is rarely measured, because its concentration is very low and it is rapidly degraded into inactive oligopeptides. Therefore, renin activity is customarily assessed by measuring angiotensin 1 concentration. In hypertensive patients, plasma renin activity may help identify whether the cause is renovascular or adrenocortical in origin. Immunoassays have been developed for measuring angiotensin 1,[73] but antibodies may react nonspecifically with other endogenous angiotensins. HPLC has also been used for angiotensin 1 measurement,[74] but most detection systems for HPLC lack the sensitivity to measure clinically relevant angiotensin 1 concentrations.

An electrospray ionization/mass spectrometry (EI/MS) method has been described for measuring plasma renin activity by quantitating angiotensin 1.[75] This method involves incubation of plasma specimens for 18 h at 37 °C to allow the renin-catalyzed conversion of angiotensinogen to angiotensin 1 to occur. A Leu → Val substituted angiotensin 1 analog was used as internal standard; the valine substitution increased the molecular weight by 14 amu. Analytes were extracted from plasma on a solid-phase C_{18} column, and then purified by reverse-phase HPLC. MS analysis was performed on a tandem mass analyzer with electrospray ionizer operating in the positive-ion mode. The limit of detection for the procedure was 0.14 ng of angiotensin 1 per milliliter of blood per hour, and coefficients of variation were less than 15%.

5.2.3 Hemoglobin

Hemoglobin is a tetrameric 64-kDa protein that is the principal oxygen carrier in blood. The hemoglobin tetramer is usually composed of two α-subunits, and two non-α-subunits, which may be pairs of β-, γ-, δ-, or ε-chains. The combinations of paired subunits give rise to several forms of hemoglobin, most of which disappear before or shortly after birth. The predominant form (approximately 96%) of hemoglobin in adults is $\alpha_2\beta_2$, and is designated HbA. Hemoglobins can be post-translationally modified by nonenzymatic covalent addition of sugar molecules at the N-terminal valine residues of the β-chains. The modified form of hemoglobin has been called glycated hemoglobin, and when the sugar is glucose, the complex is designated HbA_{1c}. HbA_{1c} measurement is useful because its formation is regulated by blood glucose concentration, and therefore abnormally high concentrations of HbA_{1c} reflect persistent hyperglycemia.[76] A variety of methods exist for measurement of HbA_{1c}, including affinity chromatography, immunoassay, electrophoresis, and HPLC. There is considerable variation, though, in quantitative HbA_{1c} results obtained by these methods,[77] possibly due to ambiguities in the definition of what HbA_{1c} really is. Unlike chromatographic and immunochemical methods, MS is capable of identifying specific molecular species, and therefore its application to HbA_{1c} measurement and standardization has been suggested.

A proposed candidate reference method for HbA_{1c}[78] involves isolation of red blood cells by centrifugation, followed by lysis in hypotonic solution. Hemolysates were mixed with buffer and cellular debris was removed by centrifugation. Hemoglobin was enzymatically cleaved by endoproteinase Glu-C, generating an N-terminal hexapeptide that was unique to the hemoglobin molecule and contained the sugar residue from glycated hemoglobin species. Endoproteinase cleavage products were separated by HPLC before mass spectral measurement on a single-stage quadrupole instrument with electrospray ionization. The method was calibrated with external standards purified by a combination of cation exchange and affinity chromatographic procedures. Results of the MS method were compared to HPLC measurement of HbA_{1c}, and the two methods produces nearly equivalent results.

A procedure similar to the method described above has been used to detect a genetic hemoglobin variant that is more facile toward glycation.[79] The hemoglobin

Rambam variant, which is characterized by a Gly → Asp substitution at position 69 in the β-subunit, was detected by measuring the mass shift of 58 Da that resulted from the substitution. The single amino acid substitution apparently promotes glycation of β-chain lysine residues, which alters the results of conventional HbA_{1c} measurements.

5.2.4 Albumin

Albumin is the predominant protein in blood, and serves many functions including regulation of plasma osmotic pressure and as a carrier protein for a variety of blood constituents. Numerous genetic variants of albumin have been reported, but interest in these variants is more academic than clinical since they do not ordinarily result in clinical disease. Nevertheless, precise measurement of albumin and its genetic variants is difficult for the reasons described above: its composition is comparable to other proteins, it is not significantly modified post-translationally with unique compounds that might be easily detectable, and genetic variants differ only slightly, often by the substitution of a single amino acid residue. Mass spectrometric analysis of albumin affords the opportunity to resolve the many variants of albumin.

EI/MS has been used to quantify albumin and some of its variants.[80–83] In one procedure,[83] albumin was purified by dialysis and size exclusion gel chromatography on a DEAE–Sephadex column. Eluates were collected and concentrated by filtration and redialyzed. Agarose gel electrophoresis indicated greater than 95% purity of the albumin specimens purified in this manner. Purified specimens were dissolved in the electrospray buffer, consisting of acetonitrile–formic acid (500 : 1) before analysis. The procedure was calibrated with standards prepared from human α-globin and hemoglobin. Using this method, multiple forms of albumin were demonstrated in patients with heterozygous albumin variants. The method was able to resolve mass differences associated with the point mutations 177Cys → Phe (44 Da), 1Asp → Val (20 Da), and Arg-albumin (160 Da), as well as more significant mutations involving carbohydrate and oligosaccharide modifications.

5.3 Drug Measurements

5.3.1 Tacrolimus

MS using non-GC inlet systems has also been applied to drug analyses. One reported example is the use of liquid chromatography and electrospray ionization to detect an immunosuppressive drug, tacrolimus (FK506), in blood.[84] Immunosuppressive therapy has become standard practice in organ transplantation programs, and the availability of potent immunosuppressive agents may be largely responsible for the success these programs have achieved. Immunosuppressive drugs, however, are characterized by a narrow therapeutic index, and the potential for toxicity from these drugs is correspondingly high.[85] Therefore, frequent monitoring of blood levels of immunosuppressive drugs is an important component of this type of therapy.[86] Immunochemical methods for tacrolimus are problematic, because of cross-reactivity of the antibody with metabolites of the drug, many of which may be inactive.[87] Therefore, quantitative results of immunoassays can give misleading results when the concentrations of inactive metabolites are high. HPLC methods can overcome this limitation, but may involve complicated and lengthy specimen preparation procedures.[88] In an MS procedure, blood specimens were mixed with an acetonitrile–water solution to precipitate proteinaceous material. Drugs were purified with a solid-phase extraction column, and reconstituted in the liquid chromatograph mobile phase. Mass spectral analysis was performed on a triple quadrupole instrument with an ion-spray interface to the liquid chromatograph. Collisional dissociation in the middle quadrupole was induced by argon gas. The method was calibrated with a series of external standards.

5.3.2 Anabolic Steroids

Drug-use in athletics has received much attention, and sports drug-testing laboratories have responded with increasingly sophisticated methods for detecting the presence of performance-enhancing drugs. Anabolic steroids are particularly difficult to detect, since many of the abused compounds are endogenous. Metabolic modification of steroid compounds also complicates efforts to detect illicit use. Finally, detection of many anabolic agents requires highly sensitive analytical methods.[89] Ion trap mass spectrometers have the ability to sequester individual fragments, and coupling an ion trap to a conventional quadrupole mass analyzer allows measurements that are both highly selective and very sensitive. The ion trap MS approach has been applied to workplace drug testing,[90] and to measuring anabolic agents in urine.[91] Common features of these applications include solid-phase extraction to isolate the drugs from urine matrix and addition of β-glucuronidase enzyme to hydrolyze glucruonyl conjugates. The drugs were derivatized with BSTFA before injection in the ion trap mass spectrometer.

5.3.3 β-Blockers

Along with anabolic steroids, another class of drugs that has been banned by the International Olympic Committee and International Sports Federation is β-blockers,

which have β-adrenoceptor antagonistic activity. A combined liquid chromatography/electrospray mass spectrometric method has been described for measuring several of the commonly used β-blocking drugs.[92] The drugs were extracted from serum or urine with a solid-phase C_{18} column prior to liquid chromatography. The mass spectrometer was interfaced to the liquid chromatograph with an electrospray ionizer, and selected ion monitoring was used to measure fragments unique to the β-blocker drugs and their metabolites.

5.4 Elemental Analytes

Because of the high temperatures required to volatilize metallic elements, spectrometric methods for these species typically involve flame or electrothermal atomization. Higher temperatures can be achieved using inductively coupled plasma as the ion source, and MS has been used as a detector for this type of instrument. Urinary iodine,[93] and titanium in spleen tissue[94] have been measured using inductively coupled plasma mass spectrometry (ICPMS).

Iodine is an essential element that is incorporated into thyroid hormones. Iodine deficiency is a well-described clinical condition that is prevalent in many underdeveloped areas of the world.[95] Monitoring urinary iodine is one method for detecting iodine deficiency, and an ICPMS procedure for measuring urinary iodine has been described.[93] Urine specimens were mixed with radioactive ^{129}I as internal standard, and mathematical corrections were made for the conversion of ^{129}I to stable ^{127}I. The accuracy of the method was verified with standard reference materials. An advantage of ICPMS is that minimal sample preparation is necessary, since all organic material in the specimen is destroyed at the 10 000 °C temperature of the plasma discharge.

Another application of ICPMS was developed to detect titanium in spleen tissue, which was suspected to emanate from a titanium-containing articular prosthesis.[94] Paraffin-embedded tissue was removed by heating, then rinsing with organic solvent. Delipidated tissue was digested with hot nitric acid before ICPMS analysis. Although results were mostly qualitative, the ICPMS method confirmed the presence of titanium in tissues, correlating with abnormal histological findings that raised the suspicion.

5.5 Characterization of Protein Epitopes

An epitope is the region of an antigen molecule that is recognized by an antibody and is responsible for the noncovalent interactions that result in the formation of an antigen–antibody complex. Epitopes may comprise a continuous sequence of amino acids in the protein molecule, or may include segments brought in close proximity of each other by the secondary or tertiary protein structure. Knowledge of these specific molecular regions that bond to antibodies is essential for understanding the selectivity of antibody-based analytical methods, and for developing effective synthetic vaccines. X-ray crystallographic data have been used to characterize the antigen–antibody complex and reveal complementary regions, but this technique requires isolation of the complex in crystalline form; this is not always possible. Other approaches utilize chemical modification of candidate epitopes, structural variants, and synthetic peptides to provide indirect information about the nature of the antibody–antigen complex. These methods are time-consuming and costly, and the results can be difficult to interpret. An innovative method has been described[96] that uses MS to identify antigen fragments in the presence and absence of antibody by MALDI.

Alkylated human and hen lysozyme were prepared by incubation with 4-vinylpyridine, guanidine hydrochloride, ethylene diamine tetraacetic acid (EDTA), and dithiothreitol in Tris buffer at 37 °C overnight. The alkylated proteins were purified by HPLC, and then digested with endoproteinase for 15 h at 37 °C. The mixture of proteolytic products was incubated with anti-lysozyme IgG for 2 h at 4 °C, and a portion of the resulting solution was prepared for MALDI/MS analysis. A control digest, without antibody, was processed in parallel. The MALDI matrix consisted of α-cyano-4-hydroxycinnamic acid in acetonitrile and water. Mass spectral analysis was performed in the positive-ion mode with a time-of-flight mass analyzer.

The key to epitope mapping by this approach is comparison of antibody-incubated antigen fragments with control fragments in the absence of antibody. Since MALDI is a soft ionization technique, antibody-bound fragments are less susceptible to vaporization. Antigen fragments that appear in control analyses but are absent in antibody-incubated preparations are presumed to be bound to antibody. Mass analysis provides molecular-weight resolution sufficient to identify the peptides bound to the antibody. In the reported experiment, a 13-residue fragment was identified as containing the epitope. Epitope mapping by MALDI/MS is a promising technique that offers the advantages of simplicity, economy and speed.

5.6 Detection of Genetic Mutations

Although both proteins and DNA are polymeric biomolecules constructed from a few available subunits, analysis of DNA material is much more difficult because the phosphodiester-linked strands are less stable than

polypeptides and are prone to the formation of nonvolatile salts. In addition, the amount of DNA available for analysis is usually very small, although amplification techniques can be used to generate sufficient material for analytical studies. DNA analysis has revolutionized several areas of clinical laboratory science, including microbiology, histology, and genetic screening. Most methods for DNA analysis involve hybridization with radioactive, fluorescent, or enzyme-labeled complementary strands.[97] MS has been applied to identification and measurement of DNA probes, using low-energy ionization methods and mass determinations to unambiguously identify relevant oligonucleotides.[98,99]

Cystic fibrosis is an autosomal recessive genetic disorder caused by a mutation in a gene that has been mapped to chromosome 7. The gene, which codes for a protein called cystic fibrosis transmembrane regulator (CFTR), is over 250-kbp in length, and over 500 specific mutations have been identified that cause disease. This high degree of allelic diversity complicates the task of screening individuals for genetic evidence of disease, since the specific mutations exhibit different ethnic incidences. Whereas testing for one group of mutations may identify a high percentage of cystic fibrosis carriers in one population, its sensitivity in another population may be quite low. Analytical methods capable of economically screening for large numbers of mutations would be beneficial. MS is a promising technique for this type of analysis, and methods have been described for detecting mutations associated with cystic fibrosis using MALDI and time-of-flight MS.[100–102]

In one of these procedures,[102] oligonucleotides complementary to several CFTR intron sequences were synthesized and amplified by polymerase chain reaction; the amplification products were purified and used as detection primers. Using a technique called primer oligo base extension (PROBE), primer oligonucleotides were annealed to native DNA, and extension was promoted by polymerase enzyme so that the oligonucleotide was lengthened to include the mutated region. The resulting oligonucleotides were measured by time-of-flight MS, which was able to identify nucleic acid substitutions based on the mass of the fragment produced in the MALDI ionizer. Advantages of this MALDI/MS method are the specificity of oligonucleotide detection, and elimination of the requirement to label probe fragments.

ABBREVIATIONS AND ACRONYMS

BSTFA	*N*,*O*-Bis(trimethylsilyl)trifluoroacetamide
CFTR	Cystic Fibrosis Transmembrane Regulator
CoA	Coenzyme A
EDTA	Ethylene Diamine Tetraacetic Acid
EI/MS	Electrospray Ionization/Mass Spectrometry
FAB	Fast Atom Bombardment
FID	Flame Ionization Detector
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LSD	Lysergic Acid Diethylamide
MALDI	Matrix-assisted Laser Desorption/Ionization
MBTFA	*N*-Methylbis(trifluoroacetamide)
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MTBSTFA	*N*-(*tert*-Butyldimethylsilyl)-*N*-methyltrifluoroacetamide
PKU	Phenylketonuria
SLO	Smith–Lemli–Opitz
TMCS	Trimethylchlorosilane
TMS	Trimethylsilyl

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Atomic Spectrometry in Clinical Chemistry • Drugs of Abuse, Analysis of • Pharmacogenetic Testing • Serum Proteins

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Hyphenated Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. A.T. James, A.J.P. Martin, 'Separation and Identification of Methyl Esters of Saturated and Unsaturated Fatty Acids from *n*-pentanoic to *n*-octadecanoic Acids', *Analyst*, **77**, 915 (1952).
2. N.C. Jain, R.H. Cravey, 'Analysis of Alcohol. II. A Review of Gas Chromatographic Methods', *J. Chromatogr. Sci.*, **10**, 263–267 (1972).
3. F. Tagliaro, G. Lubli, S. Ghielmi, D. Franchi, M. Marigo, 'Chromatographic Methods for Blood Alcohol Determination', *J. Chromatogr.*, **580**, 161–190 (1992).
4. A. Tangerman, 'Highly Sensitive Gas Chromatographic Analysis of Ethanol in Whole Blood, Serum, Urine, and Fecal Supernatants by the Direct Injection Method', *Clin. Chem.*, **43**(6), 1003–1009 (1997).
5. T.P. Hewlett, K.E. McMartin, 'Ethylene Glycol Poisoning. The Value of Glycolic Acid Determinations for Diagnosis and Treatment', *Clin. Toxicol.*, **24**, 389–402 (1986).
6. L.E. Edinboro, C.R. Nanco, D.M. Soghioan, A. Poklis, 'Determination of Ethylene Glycol in Serum Utilizing Direct Injection on a Wide-bore Capillary Column', *Ther. Drug Monit.*, **15**, 220–223 (1993).
7. H.H. Yao, W.H. Porter, 'Simultaneous Determination of Ethylene Glycol and its Major Toxic Metabolite, Glycolic Acid, in Serum by Gas Chromatography', *Clin. Chem.*, **42**(2), 292–297 (1996).
8. M. Bogusz, J. Bailka, J. Gierz, M. Klys, 'Use of Short Wide-bore Capillary Columns in GC Toxicological Screening', *J. Anal. Toxicol.*, **10**, 135–138 (1986).
9. D.W. Christ, P. Noomano, M. Rosas, D. Rhone, 'Retention Indices by Wide-bore Capillary Gas Chromatography with Nitrogen–Phosphorus Detection', *J. Anal. Toxicol.*, **12**, 84–88 (1988).
10. J.P. Franke, R.A. de Zeeuw, J. Wijsbeek, 'Potentials of Wide-bore Fused Silica Capillary Columns for Substance Identification by Means of Retention Indices', *J. Anal. Toxicol.*, **10**, 132–134 (1986).
11. R.W. Taylor, C. Greutink, N.C. Jain, 'Identification of Underivatized Basic Drugs in Urine by Capillary Column Gas Chromatography', *J. Anal. Toxicol.*, **10**, 205–208 (1986).
12. M.E. Sharp, 'A Rapid Screening Procedure for Acidic and Neutral Drugs in Blood by High Resolution Gas Chromatography', *J. Anal. Toxicol.*, **11**, 8–11 (1987).
13. D.J. Ehresman, S.M. Price, D.J. Lakatua, 'Screening Biological Samples for Underivatized Drugs Using a Splitless Injection Technique on Fused Silica Capillary Column Gas Chromatography', *J. Anal. Toxicol.*, **9**, 55–60 (1985).
14. K.E. Brooks, N.B. Smith, 'Four Automated Methods for Plasma Lactate Assessed by Comparison with Capillary Gas Chromatography', *Clin. Chem.*, **42**(7), 1111–1113 (1996).
15. R.L. Bertholf, 'Confirmatory Testing for Drugs of Abuse in Urine by Gas Chromatography/Mass Spectrometry', *The Chemist*, **67**, 7–20 (1990).
16. K. Spencer, 'Analytical Reviews in Clinical Biochemistry: The Estimation of Creatinine', *Ann. Clin. Biochem.*, **23**, 1–25 (1986).
17. B. Kagedal, B. Olsson, 'Determination of Creatinine in Serum by High Performance Liquid Chromatography: A Comparison of Three Ion-exchange Methods', *J. Chromatogr. Biomed. Appl.*, **527**, 21–30 (1990).
18. L.M. Thienpont, K.G. Van Landuyt, D. Stockl, A.P. De Leenheer, 'Candidate Reference Method for Determining Serum Creatinine by Isocratic HPLC: Validation with Isotope Dilution Gas Chromatography/Mass Spectrometry and Application for Accuracy Assessment of Routine Test Kits [see Comments]', *Clin. Chem.*, **41**(7), 995–1003 (1995).
19. A. Moore, R. Aitken, C. Burke, S. Gaskell, H. Holder, C. Selby, P. Wood, 'Cortisol Assays: Guidelines for the Provision of a Clinical Biochemistry Service', *Ann. Clin. Biochem.*, **22**, 435–454 (1985).
20. V.I. De Brabandere, L.M. Thienpont, D. Stockl, A.P. De Leenheer, 'Three Routine Methods for Serum Cortisol Evaluated by Comparison with an Isotope Dilution Gas Chromatography/Mass Spectrometry Method', *Clin. Chem.*, **41**(12), 1781–1783 (1995).
21. L.L. Abell, B.B. Levy, B.B. Brodie, F.E. Kendall, 'A Simplified Method for the Estimation of Total Cholesterol in Serum and Demonstration of its Specificity', *J. Biol. Chem.*, **195**, 357–366 (1952).
22. G.L. Myers, G.R. Cooper, C.L. Winn, S.J. Smith, 'The Centers for Disease Control – National Heart, Lung and Blood Institute Lipid Standardization Program. An Approach to Accurate and Precise Lipid Measurements', *Clin. Lab. Med.*, **9**, 105–135 (1989).
23. L.M. Thienpont, K.G. Van Landuyt, D. Stockl, A.P. De Leenheer, 'Four Frequently Used Test Systems for Serum Cholesterol Evaluated by Isotope Dilution Gas Chromatography/Mass Spectrometry Candidate Reference Method', *Clin. Chem.*, **42**(4), 531–535 (1996).
24. N. Rifai, P.S. Bachorik, J.J. Albers, 'Lipids, Lipoproteins, and Apolipoproteins', in *Tietz Textbook of Clinical*

Chemistry, 3rd edition, eds. C.A. Burtis, E.R. Ashwood, W.B. Saunders, Philadelphia, 809–861, 1999.
25. P. Ellerbe, L.T. Sniegoski, M.J. Welch, 'Isotope Dilution Mass Spectrometry as a Candidate Definitive Method for Determining Total Glycerides and Triglycerides in Serum', *Clin. Chem.*, **41**(3), 397–404 (1995).
26. J. Axelrod, R.O. Brady, B. Witkop, E.V. Evarts, 'The Distribution and Metabolism of Lysergic Acid Diethylamide', *Ann. N. Y. Acad. Sci.*, **66**, 435–444 (1957).
27. B.D. Paul, J.M. Mitchell, R. Burbage, M. Moy, R. Sroka, 'Gas Chromatography/Electron Impact Mass Fragmentometric Determination of Lysergic Acid Diethylamide in Urine', *J. Chromatogr.*, **529**, 103–112 (1990).
28. E.D. Clarkson, D. Lesser, B.D. Paul, 'Effective GC/MS Procedure for Detecting *iso*-LSD in Urine After Base-catalyzed Conversion to LSD', *Clin. Chem.*, **44**(2), 287–292 (1998).
29. O.P. Peters, W.D. Henner, L.B. Grochow, 'Clinical and Pharmacologic Effects of High Dose Single Agent Busulfan with Autologous Marrow Support in the Treatment of Solid Tumors', *Cancer Res.*, **47**, 6402–6406 (1987).
30. G.B. McDonald, M.S. Hinds, L.D. Fisher, 'Veno-occlusive Disease of the Liver and Multiorgan Failure After Bone Marrow Transplantation: A Cohort of 355 Patients', *Ann. Intern. Med.*, **118**, 225–267 (1993)
31. W.K. Lai, C.P. Pang, L.K. Law, R. Wong, C.K. Li, P.M. Yuen, 'Routine Analysis of Plasma Busulfan by Gas Chromatography/Mass Fragmentography', *Clin. Chem.*, **44**(12), 2506–2510 (1998).
32. H. Norlén, L.G. Allgén, B. Wicksell, 'Mannitol Concentrations in Blood Plasma in Connection with Transurethral Resection of the Prostate Using Mannitol Solution as an Irrigating Fluid', *Scand. J. Urol. Nephrol.*, **20**, 119–126 (1986).
33. H. Norlén, L.G. Allgén, B. Wicksell, 'Sorbitol Concentrations in Blood Plasma in Connection with Transurethral Resection of the Prostate Using Mannitol Solution as an Irrigating Fluid', *Scand. J. Urol. Nephrol.*, **20**, 9–17 (1986).
34. D. Gravenstein, 'Transurethral Resection of the Prostate (TURP) Syndrome: A Review of the Pathophysiology and the Management', *Anesth. Analg.*, **84**, 438–446 (1997).
35. F. Renner, A. Schmitz, H. Gehring, 'Rapid and Sensitive Gas Chromatography/Mass Spectroscopy Method for the Detection of Mannitol and Sorbitol in Serum Samples', *Clin. Chem.*, **44**(4), 886–887 (1998).
36. D.E. Hale, M.J. Bennett, 'Fatty Acid Oxidation Disorders: A New Class of Metabolic Diseases', *J. Pediatr.*, **121**, 1–11 (1992).
37. C.G. Costa, L. Dorland, U. Holwerda, I.T. de Almeida, B.T. Poll-The, C. Jakobs, M. Duran, 'Simultaneous Analysis of Plasma Free Fatty Acids and Their 3-hydroxy Analogs in Fatty Acid Beta-oxidation Disorders', *Clin. Chem.*, **44**(3), 463–471 (1998).
38. L. Dorland, D. Ketting, L. Bruinvis, M. Duran, 'Medium and Long-chain 3-hydroxymonocarboxylic Acids: Analysis by Gas Chromatography Combined with Mass Spectrometry', *Biomed. Chromatogr.*, **5**, 161–164 (1991).
39. W. Onkenhout, V. Venizelos, P.F. van der Poel, M.P. van den Heuvel, B.J. Poorthuis, 'Identification and Quantification of Intermediates of Unsaturated Fatty Acid Metabolism in Plasma of Patients with Fatty Acid Oxidation Disorders', *Clin. Chem.*, **41**(10), 1467–1474 (1995).
40. G. Lepage, C.C. Roy, 'Specific Methylation of Plasma Nonesterified Fatty Acids in a One-step Reaction', *J. Lipid. Res.*, **29**, 227–235 (1988).
41. S. Aleryani, A. Kabakibi, J. Cluette-Brown, M. Laposata, 'Fatty Acid Ethyl Ester Synthase, an Enzyme for Nonoxidative Ethanol Metabolism, is Present in Serum After Liver and Pancreatic Injury', *Clin. Chem.*, **42**(1), 24–27 (1996).
42. E.A. Laposata, L.G. Lange, 'Presence of Nonoxidative Ethanol Metabolism in Human Organs Commonly Damaged by Ethanol Abuse', *Science*, **231**, 497–499 (1986).
43. R. Carmel, 'Subtle and Atypical Cobalamin Deficiency States', *Am. J. Hematol.*, **34**, 108–114 (1990).
44. L.P. Stabler, R.H. Allen, D.G. Savage, J. Lindenbaum, 'Clinical Spectrum and Diagnosis of Cobalamine Deficiency', *Blood*, **76**, 871–881 (1990).
45. M.T. McCann, M.M. Thompson, I.C. Gueron, B. Lemieux, R. Giguere, M. Tuchman, 'Methylmalonic Acid Quantification by Stable Isotope Dilution Gas Chromatography/Mass Spectrometry from Filter Paper Urine Samples', *Clin. Chem.*, **42**(6), 910–914 (1996).
46. C. Bachmann, J.P. Columbo, 'Diagnostic Value of Orotic Acid Excretion in Heritable Disorders of the Ures Cycle and in Hyperammonemia due to Organic Acidurias', *Eur. J. Pediatr.*, **134**, 109–113 (1980).
47. S.W. Brusilow, E. Hauser, 'Simple Method of Measurement of Orotic Acid and Orotidine in Urine', *J. Chromatogr.*, **493**, 388–391 (1989).
48. M. Rimoldi, P. Bergomi, A. Romeo, 'A New Stable Isotope Dilution Method for the Measurement of Orotic Acid Utilizing Solvent-extracted Urine', *J. Inherit. Metab. Dis.*, **17**, 243–244 (1994).
49. W.L. Hwu, S.P. Chou, T.R. Wang, 'Measurement of Urinary Orotic Acid by Gas Chromatography/Mass Spectrometry', *Acta. Paediatr. Sin.*, **33**, 176–179 (1992).
50. M.T. McCann, M.M. Thompson, I.C. Gueron, M. Tuchman, 'Quantification of Orotic Acid in Dried Filter-paper Urine Samples by Stable Isotope Dilution', *Clin. Chem.*, **41**(5), 739–743 (1995).
51. E. Schmidt-Sommerfeld, D. Penn, M. Duran, M.J. Bennett, R. Santer, C.A. Stanley, 'Detection of Inborn Errors of Fatty Acid Oxidation from Acylcarnitine Analysis of Plasma and Blood Spots with the Radioisotopic

Exchange-high-performance Liquid Chromatographic Method', *J. Pediatr.*, **122**, 708–713 (1993).
52. P. Duez, A. Kumps, Y. Mardens, 'GC/MS Profiling of Urinary Organic Acids Evaluated as a Quantitative Method', *Clin. Chem.*, **42**(10), 1609–1615 (1996).
53. W. Shaw, E. Kassen, E. Chaves, 'Increased Urinary Excretion of Analogs of Krebs Cycle Metabolites and Arabinose in Two Brothers with Autistic Features', *Clin. Chem.*, **41**(8), 1094–1104 (1995).
54. H.L. Needleman, A. Schell, D. Bellinger, A. Leviton, E.N. Alfred, 'The Long-term Effects of Exposure to Low Doses of Lead in Childhood', *N. Engl. J. Med.*, **263**, 83–88 (1990).
55. D. Jagner, L. Renman, Y.D. Wang, 'Determination of Lead in Microliter Amounts of Whole Blood by Stripping Potentiometry', *Electroanalysis*, **6**, 285–291 (1994).
56. G.S. Baird, R.L. Fitzgerald, S.K. Aggarwal, D.A. Herold, 'Determination of Blood Lead by Electron-capture Negative Chemical Ionization Gas Chromatography/Mass Spectrometry', *Clin. Chem.*, **42**(2), 286–291 (1996).
57. D.A. Herold, M. Kinter, S.K. Aggarwal, 'Analysis by Gas Chromatography/Mass Spectrometry', in *Handbook on Metals in Clinical Chemistry*, eds. H.G. Seiler, A. Sigel, H. Sigel, Marcel Dekker, New York, 35–87, 1994.
58. R.L. Fitzgerald, D.A. Herold, A.L. Yergey, 'Trade-offs in Mass Spectrometry [Editorial; Comment]', *Clin. Chem.*, **43**(6), 915 (1997).
59. M.S. Rashed, P.T. Ozand, M.E. Harrison, P.J.F. Watkins, S. Evans, 'Electrospray Tandem Mass Spectrometry in the Diagnosis of Organic Acidemias', *Rapid Commun. Mass Spectrom.*, **8**, 129–133 (1994).
60. M.S. Rashed, M.P. Bucknall, D. Little, A. Awad, M. Jacob, M. Alamoudi, M. Alwattar, P.T. Ozand, 'Screening Blood Spots for Inborn Errors of Metabolism by Electrospray Tandem Mass Spectrometry with a Microplate Batch Process and a Computer Algorithm for Automated Flagging of Abnormal Profiles', *Clin. Chem.*, **43**(7), 1129–1141 (1997).
61. E.R.B. McCabe, L. McCabe, G.A. Mosher, R.J. Allen, J.L. Berman, 'Newborn Screening for Phenylketonuria – Predictive Validity as a Function of Age', *Pediatrics*, **72**, 390–398 (1983).
62. D.H. Chace, J.E. Sherwin, S.L. Hillman, F. Lorey, G.C. Cunningham, 'Use of Phenylalanine-to-tyrosine Ratio Determined by Tandem Mass Spectrometry to Improve Newborn Screening for Phenylketonuria of Early Discharge Specimens Collected in the First 24 Hours [see Comments]', *Clin. Chem.*, **44**(12), 2405–2409 (1998).
63. D.S. Millington, N. Kodo, N. Terada, D. Roe, D.H. Chace, 'The Analysis of Diagnostic Markers of Genetic Disorders in Human Blood and Urine Using Tandem Mass Spectrometry with Liquid Secondary Ion Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **111**, 211–228 (1991).
64. D.H. Chace, S.L. Hillman, D.S. Millington, S.G. Kahler, C.R. Roe, E.W. Naylor, 'Rapid Diagnosis of Maple Syrup Urine Disease in Blood Spots from Newborns by Tandem Mass Spectrometry', *Clin. Chem.*, **41**(1), 62–68 (1995).
65. S.H. Mudd, H.L. Levy, F. Skovby, 'Disorders of Transsulfuration', *Electroanalysis*, **7**, 1279–1327 (1995).
66. D.H. Chace, S.L. Hillman, D.S. Millington, S.G. Kahler, B.W. Adam, H.L. Levy, 'Rapid Diagnosis of Homocystinuria and Other Hypermethioninemias from Newborns' Blood Spots by Tandem Mass Spectrometry [see Comments]', *Clin. Chem.*, **42**(3), 349–355 (1996).
67. J.M. Opitz, V.B. Penchaszadeh, M.C. Holt, L.M. Spano, V.L. Smith, 'Smith–Lemli–Opitz (RSH) Syndrome Bibliography: 1964–1993', *Am. J. Med. Genet.*, **50**, 339–343 (1994).
68. G.S. Tint, M. Irons, E.R. Elias, A.K. Batta, R. Frieden, T.S. Chen, G. Salen, 'Defective Cholesterol Biosynthesis Associated with the Smith–Lemli–Opitz Syndrome', *N. Engl. J. Med.*, **330**, 107–113 (1994).
69. U. Seedorf, M. Fobker, R. Voss, K. Meyer, F. Kannenberg, D. Meschede, K. Ullrich, J. Horst, A. Benninghoven, G. Assmann, 'Smith–Lemli–Opitz Syndrome Diagnosed by Using Time-of-flight Secondary-ion Mass Spectrometry', *Clin. Chem.*, **41**(4), 548–552 (1995).
70. R.E. Byrne, A.M. Scanu, 'Soluble and Immobilized Trypsin as Structural Probes of Human Plasma High-density Lipoproteins: Enzyme Properties and Kinetics of Proteolysis', *Biochemistry*, **22**, 2894–2903 (1983).
71. S.K. Chowdhury, V. Katta, B.T. Chait, 'Electrospray Ionization Mass Spectrometric Peptide Mapping: A Rapid Sensitive Technique for Protein Structure Analysis', *Biochem. Biophys. Res. Commun.*, **167**, 676–692 (1990).
72. J.R. Barr, V.L. Maggio, D.G. Patterson, Jr, G.R. Cooper, L.O. Henderson, W.E. Turner, S.J. Smith, W.H. Hannon, L.L. Needham, E.J. Sampson, 'Isotope Dilution–Mass Spectrometric Quantification of Specific Proteins: Model Application with Apolipoprotein A-I', *Clin. Chem.*, **42**(10), 1676–1682 (1996).
73. K. Sugimoto, H. Shionoiri, K. Minamisawa, Y. Abe, S. Ueda, H. Himeno, 'Measurement of Plasma Total Renin by the Anti-human Renin Monoclonal Antibodies', *Am. J. Med. Sci.*, **306**, 342–346 (1991).
74. T. Miyazaki, M. Kai, Y. Ohkura, 'Determination of Renin Activity in Human Plasma by Column-switching High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **490**, 43–51 (1989).
75. V.F. Fredline, E.M. Kovacs, P.J. Taylor, A.G. Johnson, 'Measurement of Plasma Renin Activity with Use of HPLC/Electrospray/Tandem Mass Spectrometry', *Clin. Chem.*, **45**(5), 659–664 (1999).
76. The Diabetes Control and Complications Trial Research Group, 'The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-term

Complications in Insulin-dependent Diabetes Mellitus', *N. Engl. J. Med.*, **329**, 977–986 (1993).
77. K. Shima, J. Endo, M. Oimomi, I. Oshima, Y. Omori, Y. Katayama, 'Interlaboratory Difference in HbA_{1c} Measurement in Japan. An Interim Report of the Committee on an Interlaboratory Standardization of HbA_{1c}. Determination, the Japan Diabetes Society', *J. Jpn. Diabetes. Soc.*, **37**, 233–243 (1994).
78. U. Kobold, J.O. Jeppsson, T. Dulffer, A. Finke, W. Hoelzel, K. Miedema, 'Candidate Reference Methods for Hemoglobin A1c Based on Peptide Mapping [see Comments]', *Clin. Chem.*, **43**(10), 1944–1951 (1997).
79. E. Bisse, N. Zorn, A. Eigel, M. Lizama, P. Huaman-Guillen, W. Marz, A. Van Dorsselaer, H. Wieland, 'Hemoglobin Rambam (beta69[E13]Gly → Asp), a Pitfall in the Assessment of Diabetic Control: Characterization by Electrospray Mass Spectrometry and HPLC', *Clin. Chem.*, **44**(10), 2172–2177 (1998).
80. H.E. Witkowska, F. Bitsch, C.H.L. Shackleton, 'Expediting Rare Hemoglobin Variant Identification by Combination HPLC/Electrospray Mass Spectrometry', *Hemoglobin*, **17**, 227–247 (1993).
81. S.O. Brennan, J.R.D. Mathews, 'Auckland (α87His → Asn): A New Mutation of the Proximal Histidine Identified by Electrospray Mass Spectrometry', *Hemoglobin*, **21**, 393–403 (1997).
82. A.P. Fellowes, S.O. Brennan, H.J. Ridgway, D.C. Heaton, P.M. George, 'Electrospray Ionization Mass Spectrometry Identification of Fibrinogen Banks Peninsula (γ280Tyr → Cys): A New Variant with Defective Polymerization', *Br. J. Haematol.*, **101**, 24–31 (1998).
83. S.O. Brennan, 'Electrospray Ionization Mass Analysis of Normal and Genetic Variants of Human Serum Albumin', *Clin. Chem.*, **44**(11), 2264–2269 (1998).
84. P.J. Taylor, A. Jones, G.A. Balderson, S.V. Lynch, R.L. Norris, S.M. Pond, 'Sensitive, Specific Quantitative Analysis of Tacrolimus (FK506) in Blood by Liquid Chromatography/Electrospray Tandem Mass Spectrometry', *Clin. Chem.*, **42**(2), 279–285 (1996).
85. D.H. Peters, A. Fitton, G.L. Plosker, D. Faulds, 'Tacrolimus. A Review of its Pharmacology and Therapeutic Potential in Hepatic and Renal Transplantation', *Drugs*, **46**, 746–794 (1993).
86. W.J. Jusko, R. D'Ambrosio, 'Monitoring FK-506 Concentrations in Plasma and Whole Blood', *Transplant Proc.*, **23**, 2732–2735 (1991).
87. K. Iwasaki, T. Shiraga, H. Matsuda, K. Nagase, Y. Tokuma, T. Hata, 'Further Metabolism of FK506 (tacrolimus). Identification and Biological Activities of the Metabolites Oxidised at Multiple Sites of FK506', *Drug Metab. Dispos.*, **23**, 28–34 (1995).
88. V. Warty, S. Zuckerman, R. Venkataramanaan, J. Lever, J. Fung, T. Starzi, 'FK506 Measurement: Comparison of Different Analytical Methods', *Ther. Drug Monit.*, **15**, 204–208 (1993).
89. D. de Boer, E.G. de Jong, R.A.A. Maes, 'Mass Spectrometric Characterization of Different Norandrosterone Derivatives by Low-cost Mass Spectrometric Detectors Using Electron Ionization and Chemical Ionization', *Rapid Commun. Mass Spectrom.*, **4**, 181–185 (1990).
90. A.H. Wu, T.A. Onigbinde, S.S. Wong, K.G. Johnson, 'Evaluation of Full-scanning GC/Ion Trap MS Analysis of NIDA Drugs-of-abuse Testing in Urine', *J. Anal. Toxicol.*, **16**, 202–206 (1992).
91. L.D. Bowers, D.J. Borts, 'Evaluation of Selected-ion Storage Ion-trap Mass Spectrometry for Detecting Urinary Anabolic Agents [see Comments]', *Clin. Chem.*, **43**(6), 1033–1039 (1997).
92. H. Kataoka, S. Narimatsu, H.L. Lord, J. Pawliszyn, 'Automated In-tube Solid-phase Microextraction Coupled with Liquid Chromatography/Electrospray Ionization Mass Spectrometry for the Determination of β-blockers and Metabolites in Urine and Serum Samples', *Anal. Chem.*, **71**, 4237–4244 (1999).
93. M. Haldimann, B. Zimmerli, C. Als, H. Gerber, 'Direct Determination of Urinary Iodine by Inductively Coupled Plasma Mass Spectrometry Using Isotope Dilution with Iodine-129', *Clin. Chem.*, **44**(4), 817–824 (1998).
94. V. Ducros, M. Peoc'h, C. Moulin, D. Ruffieux, J. Amosse, A. Favier, B. Pasquier, 'Titanium Identification and Measurement by ICP/MS and ICP/OES in a Human Spleen Fragment', *Clin. Chem.*, **42**(11), 1875–1877 (1996).
95. E. Gaitan, J.T. Dunn, 'Epidemiology of Iodine Deficiency', *Trends Endocrinol. Metab.*, **3**, 170–175 (1992).
96. J.G. Kiselar, K.M. Downard, 'Direct Identification of Protein Epitopes by Mass Spectrometry without Immobilization of Antibody and Isolation of Antibody–Peptide Complexes', *Anal. Chem.*, **71**(9), 1792–1801 (1999).
97. R. Drmanac, S. Drmanac, Z. Stezoska, T. Paunesku, I. Labat, M. Zeremski, 'DNA Sequence Determination by Hybridization: A Strategy for Efficient Large-scale Sequencing', *Science*, **260**, 1649–1652 (1993).
98. D.P. Little, T.W. Thannhauser, F.W. McLafferty, 'Verification of 50- to 100-mer DNA and RNA Sequences with High-resolution Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **82**, 2318–2322 (1995).
99. M.C. Fitzgerald, L.M. Smith, 'MS of Nucleic Acids: The Promise of MALDI MS', *Ann. Rev. Biophys. Biomol. Struct.*, **24**, 117–140 (1995).
100. Y.-H. Liu, J. Bai, Y. Zhu, X. Liang, D. Siemieniak, P.J. Ventra, 'Rapid Screening of Genetic Polymorphisms Using Buccal Cell DNA with Detection by Matrix-assisted Laser Desorption/Ionization Mass

Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 735–743 (1995).

101. L.-Y. Chang, K. Tang, M. Schell, C. Ringelberg, K.J. Matteson, S.L. Allman, 'Detection of ΔF508 Mutation of the Cystic Fibrosis Gene by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 772–774 (1995).

102. A. Braun, D.P. Little, H. Koster, 'Detecting CFTR Gene Mutations by Using Primer Oligo Base Extension and Mass Spectrometry [see Comments]', *Clin. Chem.*, **43**(7), 1151–1158 (1997).

Glucose Measurement

Marco F. Cardosi
University of Paisley, Paisley, UK

1 Introduction 1
1.1 Definition and Classification 1
1.2 (+)-Glucose: an Aldohexose 2
1.3 Osazone Formation 2
1.4 Tautomerism and Mutarotation of Glucose 2
2 The Importance of Measuring Glucose 2
3 Measurement Methods for Glucose 3
3.1 Chemical Methods: the *o*-Toluidine Approach 3
3.2 Enzymatic Methods for Measuring Glucose 3
4 Electrochemical Detection of Glucose 6
4.1 Pulsed Amperometric Detection of Glucose 6
4.2 Enzyme Electrodes 7
5 Conclusions 11
Abbreviations and Acronyms 11
Related Articles 12
References 12

Because of the pivotal role of glucose in physiological processes, much effort has been devoted to the development of methods for measuring glucose in food, microbiological and clinical matrices. The development of direct methods to measure glucose has been hindered by the lack of a suitable chromophore or indeed an electrophore. Consequently, more complex indirect methods have been developed based either on the chemical reactivity of the sugar, a particular enzymatic reaction or an indirect electrochemical measurement based upon the measurement of tensametric peaks or waves.

Although accurate and offering good sensitivity, chemical methods for the analysis of glucose such as the o*-toluidine approach are not specific for glucose. Other hexoses such as mannose and galactose, some of their derivatives, aldopentoses and some disaccharides may also react, thereby complicating the measurement.*

Because of the lack of specificity associated with chemical methods of measuring glucose, enzymatic methods have gained in popularity. Enzymes are very selective biological catalysts which carry out the conversion of a particular substrate into a product under mild operating conditions, i.e. room temperature, low ionic strength and near neutral pH. An enzyme will usually catalyze a single chemical reaction or a set of closely related chemical reactions. Side reactions leading to wasteful formation of by-products rarely occur. Thus, quantitative assays may be done on crude materials with little or no sample preparation. The enzymes which are currently used for the measurement of glucose include hexokinase (EC 2.7.1.1; adenosine triphosphate (ATP): D-hexose-6-phosphotransferase), glucose oxidase (EC 1.1.3.4; β-D-glucose: oxygen 1-oxidoreductase) and glucose dehydrogenase (EC 1.1.1.47; β-D-glucose: (nicotinamide adenine dinucleotide phosphate (oxidized form)) NAD(P)$^+$ 1-oxidoreductase). The mode of detection that is used to measure the activity and hence concentration of glucose, ultimately depends upon the nature of the enzymatic reaction itself and is usually colorimetric or electrochemical.

More recently, new analytical methodologies based upon the direct electrochemical detection of glucose using pulsed amperometric detection (PAD) have been developed. Coupled with liquid chromatography (LC), this technique is proving to be a highly selective and sensitive method for the measurement of glucose in a wide variety of sample matrices.

1 INTRODUCTION

In the leaf of a plant, the simple compounds carbon dioxide and water are condensed to form the monosaccharide sugar (+)-*glucose*. This process, known as photosynthesis, requires the catalytic action of the green pigment chlorophyll and requires energy in the form of sunlight. Thousands of (+)-glucose molecules can then be combined to form polymers such as cellulose, which constitutes the supporting framework of the plant, and starch which is stored in the seeds to form a food reserve for the growing plant.

When eaten by an animal, the starch (and in the case of ruminants the cellulose) is broken down once more into (+)-glucose. The sugar is then carried by the bloodstream to tissue where it is ultimately oxidized to carbon dioxide and water with the concomitant release of energy originally supplied by sunlight. Some of the glucose is converted into fats, some reacts with nitrogenous compounds to form amino acids and some in turn is converted into the storage polysaccharide glycogen.

1.1 Definition and Classification

Glucose belongs to the class of organic compounds known as carbohydrates. Carbohydrates are polyhydroxy aldehydes, polyhydroxy ketones or compounds that can be hydrolyzed to them. A carbohydrate such as glucose

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

that cannot be hydrolyzed to a simpler compound is known as a monosaccharide.

A monosaccharide may be further classified depending upon the presence of certain chemical groups. If it contains an aldehyde group, it is known as an aldose; if it contains a ketone group, it is known as a ketose. Furthermore, depending upon the number of carbon atoms that it contains, a monosaccharide can be classified as a triose, tetrose, pentose, hexose and so on.

1.2 (+)-Glucose: an Aldohexose

(+)-Glucose has the molecular formula $C_6H_{12}O_6$ as shown by elemental analysis and molecular weight determination. The straight chain structure (Fisher projection) of (+)-glucose is shown as Structure **(1)**. It is a colorless crystalline substance with a sweet taste, though not as sweet as cane sugar. It is readily soluble in water but only sparingly soluble in alcohol. Like other monosaccharides, glucose is optically active and the naturally occurring form is dextro-rotatory (D-(+)-glucose). On the whole, glucose displays the properties of an aldose, i.e. it can be oxidized in four important ways: (1) by Fehling's or Tollens' reagent, (2) by bromine water, (3) by nitric acid and (4) by periodic acid. It does however give a negative Schiff test and does not form a bisulfite addition product.

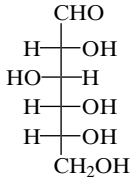

(1)

1.3 Osazone Formation

Because glucose is an aldose sugar it will react with phenylhydrazine to form phenylhydrazones. If an excess of phenylhydrazine is used, the reaction proceeds further to yield products known as osazones which contain two phenylhydrazine residues per molecule at positions 1 and 2.

1.4 Tautomerism and Mutarotation of Glucose

Glucose exhibits tautomerism of a cyclic type made possible by the fact that the chain of carbon atoms in the molecule is not actually straight but is in fact bent around the periphery of a ring. Consequently, the −CHO group is very close to the C−OH group at position 5 and reacts forming an intramolecular hemiacetal to give two anomers known as α- and β-D-(+)-glucose, Structures **(2)** and **(3)**, respectively. The two isomers differ by having opposite configurations for the C1-hydroxy.

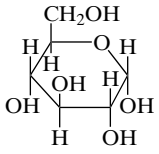

(2)

α-D-(+)-Glucose

m.p 146°C, [α] = + 112°

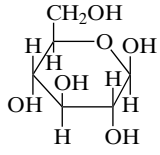

(3)

β-D-(+)-Glucose

m.p 150°C, [α] = + 19°

When crystals of D-(+)-glucose, which have been obtained from alcohol or water, of m.p. 146 °C are dissolved in water, the specific rotation gradually drops from an initial +112° to +52.7°. This change in optical rotation on standing at room temperature is referred to as mutarotation and is caused by tautomeric changes in solution in the original α- and β-glucose to an equilibrium mixture which contains 37% of the α form and 63% of the β form.

2 THE IMPORTANCE OF MEASURING GLUCOSE

Because of the pivotal role of glucose in physiological processes, much effort has been devoted to the development of methods for measuring glucose in food, microbiological and clinical matrices. In addition, the accurate estimation of blood glucose levels plays a very important role in the management of diabetes mellitus, a worldwide health problem involving an insulin deficiency that results in an inability to control blood glucose levels. About 2% of the world's population suffer from insulin-dependent diabetes. The long-term complications of diabetes involve vascular problems such as nerve damage, kidney disease, retinal damage and accelerated arteriosclerosis. Finally, increased levels of plasma glucose have been shown to be associated with hyperactivity of the thyroid pituitary or adrenal glands while decreased levels are observed in cases of insulin overdosage, insulin-secreting tumours, myxedema, hypopituitarism, hypoadrenalism and conditions interfering with glucose absorption. Expected values for normal glucose levels in whole blood or plasma are summarized in Table 1.

Table 1 Expected glucose values in healthy adult subjects

Normal values for fasting adults	
Whole blood	50–90 mg dL^{-1} (2.8–5.0 mM)
Serum or plasma	70–105 mg dL^{-1} (3.7–5.8 mM)

The development of direct methods to measure glucose has been hindered by the lack of a suitable chromophore or indeed an electrophore. Consequently, more complex indirect methods have been developed based either on the chemical reactivity of the sugar, a particular enzymatic reaction or an indirect electrochemical measurement based upon the measurement of tensametric peaks or waves. This article will critically review the current methodologies available for the measurement of glucose.

3 MEASUREMENT METHODS FOR GLUCOSE

3.1 Chemical Methods: the *o*-Toluidine Approach

The usage of *o*-toluidine for the detection of glucose has an established history. In 1959, Hultman and coworkers[1] reported the use of *o*-toluidine **(4)** occupational exposure limit (OEL) 9 mg m^{-3}; LD_{50}, 670 mg Kg^{-1} (oral rat) as a reagent for the colorimetric determination of glucose in body fluids. The color reagent, which was essentially *o*-toluidine in glacial acetic acid (6% v/v), was found to be specific for aldose sugars. In 1962, Dubowski[2] repeated this work and extended the study to include an investigation of the spectral characteristics of the reaction, reaction composition, sensitivity, reproducibility, recovery and specificity. Based on his findings, Dubowski[2] concluded that the *o*-toluidine method fulfilled the requirements of a practical analytical assay for body fluid glucose. In the same year, the assay method was modified by adding thiourea to the *o*-toluidine reagent resulting in improved reagent stability and appreciable lowering of the blank reading.

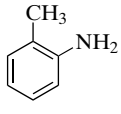

(4)

o-Toluidine is toxic by inhalation, toxic in contact with skin, toxic if swallowed.

The principle of the assay procedure is that in the presence of heat and acid, *o*-toluidine reacts with glucose to form a blue-green colored complex, the intensity of which is directly proportional to the concentration of glucose (Equation 1).

$$o\text{-toluidine} + \text{glucose} \xrightarrow{\text{acid, heat}} \text{colored complex} \quad (1)$$
$$\text{(measured at 620–650 nm)}$$

Table 2 Substances that interfere with the *o*-toluidine determination of glucose. (Adapted from Sigma Technical Bulletin, Procedure 635 with permission)

Substance	Degree of reaction relative to glucose
Galactose	100%
Mannose	100%
Lactose	33%
Glucuronic acid	20%
Xylose	19%
Maltose	5%

Other hexoses such as mannose and galactose, some of their derivatives, aldopentoses and some disaccharides may also react and form a colored complex with *o*-toluidine, see Table 2. These substances are normally present in low concentrations in body fluids of healthy individuals and do not interfere. Because this procedure does not distinguish between glucose and galactose, the test cannot be used for glucose determination from patients suffering from galactosemia or from those undergoing galactose tolerance tests. A kit for the analysis of glucose using the *o*-toluidine reagent can be purchased from Sigma-Aldrich Co. Ltd, product number 635.

3.2 Enzymatic Methods for Measuring Glucose

Because of the lack of specificity associated with chemical methods of measuring glucose, enzymatic methods have gained in popularity. Enzymes are very selective biological catalysts which carry out the conversion of a particular substrate into a product under mild operating conditions, i.e. room temperature, low ionic strength and near neutral pH. An enzyme will usually catalyze a single chemical reaction or a set of closely related chemical reactions. Side reactions leading to wasteful formation of by-products rarely occur. Thus, quantitative assays may be done on crude materials with little or no sample preparation.

Much of the catalytic power of enzymes comes from lowering the activation energy for reaction through stabilization of the transition state when bound to the enzyme. The substrates are bound to a specific region of the enzyme known as the active site. Most enzymes are highly specific in the binding of their substrates and, indeed, the catalytic specificity of an enzyme is dependent at least in part on the specificity of this binding.

For many enzymes, the rate of catalysis (v_0), defined as the number of moles of product formed per second varies with substrate concentration [S] in a manner shown in Figure 1.

At a fixed concentration of enzyme, v_0 is almost linearly proportional to [S] when the concentration is small. At high substrate concentrations, the rate of reaction

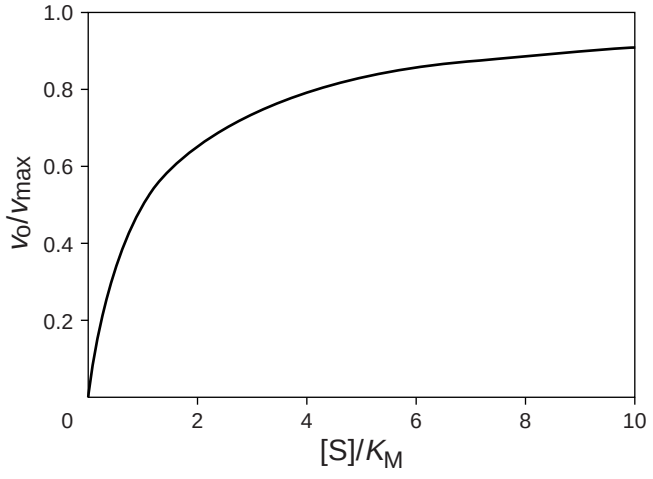

Figure 1 Dimensionless plot of the Michaelis–Menten equation.

becomes independent of [S] and reaches a maximum value. The model proposed, which is the simplest one to account for these kinetic properties, is shown in Equation (2).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \tag{2}$$

According to this model, an enzyme molecule E combines with substrate to form an ES complex with a rate constant k_1. The ES complex then has two possible fates; it can dissociate back to E and S, with a rate constant k_{-1} or it can proceed in an irreversible manner to form a product P, with a rate constant k_2. Using this model, it is possible to define a rate equation (Michaelis–Menten equation) for an enzyme-catalyzed reaction which states (Equation 3) that

$$v_0 = v_{max}\frac{[S]}{[S] + K_M} \tag{3}$$

where K_M (the Michaelis constant) is defined as $(k_{-1} + k_2)/k_1$ and v_{max} is the product of the total enzyme concentration and k_3. Thus, when $[S] \ll K_M$ the rate of reaction v is defined as Equation (4):

$$v_0 = \frac{v_{max}[S]}{K_M} \tag{4}$$

that is the rate is directly proportional to [S].

The enzymes that are currently used for the measurement of glucose include hexokinase (EC 2.7.1.1; ATP: D-hexose-6-phosphotransferase), glucose oxidase (EC 1.1.3.4; β-D-glucose: oxygen 1-oxidoreductase) and glucose dehydrogenase (EC 1.1.1.47; β-D-glucose: $NAD(P)^+$ 1-oxidoreductase). The mode of detection that is used to measure the activity, and hence concentration of glucose, ultimately depends upon the nature of the enzymatic reaction itself.

3.2.1 Hexokinase Method (Ultraviolet Measurement)

The methodology for the hexokinase method of quantifying glucose is summarized in Equations (5) and (6).

$$\text{glucose} + \text{ATP} \xrightarrow{\text{hexokinase}} \text{glucose-6-phosphate} + \text{ADP} \tag{5}$$

$$\text{glucose-6-phosphate} + NAD^+ \xrightarrow{\text{glucose-6-phosphate dehydrogenase}} \text{6-phosphogluconate} + \text{NADH} \tag{6}$$

Glucose is phosphorylated in the presence of ATP in a reaction catalyzed by hexokinase. Glucose-6-phosphate is then oxidized to 6-phosphogluconate in the presence of NAD^+ (nicotinamide adenine dinucleotide, oxidized form) in a reaction catalyzed by the second enzyme (known as the indicator enzyme) glucose-6-phosphate dehydrogenase (EC 1.1.1.49; glucose-6-phosphate: $NAD(P)^+$ oxidoreductase). During the oxidation, an equimolar amount of NAD^+ is reduced to NADH (nicotinamide adenine dinucleotide, reduced form) which has a chromophore at 340 nm ($\varepsilon = 6200\,M^{-1}\,cm^{-1}$) and can therefore be measured conveniently using a spectrophotometer. The resultant increase in absorbance at 340 nm is directly proportional to the glucose concentration provided that the reaction is allowed to go to completion. (This is an example of an "end-point" assay procedure). Typically, solutions ranging from 0.5–50 μg of glucose per milliliter should give an $A_{340\,nm}$ between 0.03 and 1.6 (data adapted from Sigma Technical Bulletin GHKB-2). Normal procedures should be carried out with regard to sample preparation. Cloudy samples should be either filtered or deproteinized, solutions which are strongly colored should be decolorized and carbonated or fermented samples should be degassed. A typical enzymatic assay solution contains 1.5 mM NAD^+, 1.0 mM ATP, $1\,U\,mL^{-1}$ hexokinase and $1\,U\,mL^{-1}$ glucose-6-phosphate dehydrogenase (where U is a unit of enzyme activity). Using this solution, assay times at room temperature should not exceed 15 min. An assay kit for determining glucose by this method is available from Sigma-Aldrich Co. Ltd, product number 16-100P. It is important to note that hexokinase is not totally specific for glucose. Indeed, it will catalyze the phosphorylation of other hexoses including D-mannose, D-fructose, 2-deoxy-D-glucose and D-glucosamine. This apparent lack of specificity is unimportant in the assay of glucose since the indicator enzyme is specific for glucose-6-phosphate and will not react with either mannose-6-phosphate or

fructose-6-phosphate without the incorporation of phosphoglucose isomerase (EC 5.3.1.9; D-glucose-6-phosphate ketol isomerase).

3.2.2 Hexokinase Method (Colorimetric Measurement)

An interesting modification of the above method makes use of the fact that reduced nicotinamide coenzyme can give rise to a colored product in the presence of 2-(4-iodophenyl)-3-(4-nitrophenyl)-5-phenyltetrazolium chloride (INT) **(5)** and phenazine methosulfate (PMS) **(6)**, which has a chromophore at 520 nm. In this approach, the reduced coenzyme nicotinamide adenine dinucleotide phosphate (reduced form) (NADPH) is oxidized back to $NADP^+$ by PMS. The reduced PMSH is then reoxidized by INT resulting in the colored form of the dye INT-Formazan. The sequence of reactions that constitute this assay method is summarized in Equations (7–10).

$$\text{glucose} + \text{ATP} \xrightarrow{\text{hexokinase}} \text{glucose-6-phosphate} + \text{ADP} \quad (7)$$

$$\text{glucose-6-phosphate} + \text{NADP}^+ \xrightarrow{\text{glucose-6-phosphate dehydrogenase}} \text{6-phosphogluconate} + \text{NADPH} \quad (8)$$

$$\text{NADPH} + \text{PMS} \longrightarrow \text{NADP}^+ + \text{PMSH} \quad (9)$$

$$\text{PMSH} + \text{INT} \longrightarrow \text{PMS} + \text{INT-Formazan} \quad (10)$$

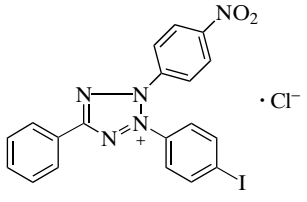

(5)

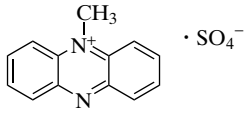

(6)

An assay kit based upon this methodology is available from Sigma-Aldrich Co. Ltd., product number 115-A. The total test time by this method is 5–10 min. The dynamic range of the test is 0.5–300 mg dL^{-1} (0.028–16.7 mol L^{-1}) and so is useful for plasma glucose determinations where the expected values lie in the range 70–105 mg dL^{-1}. Although more sensitive than the ultraviolet (UV) method, this technique is more prone to interference, particularly if any reducing agents are present in the sample.

3.2.3 Glucose Oxidase Method (Colorimetric Measurement)

Glucose oxidase is a glycosylated flavoprotein which catalyzes the oxidation of β-D-glucose by atmospheric oxygen to produce D-gluconolactone, which is rapidly converted to gluconic acid with the production of hydrogen peroxide, Equation (11)

$$\text{glucose} + \text{O}_2 + 2\text{H}_2\text{O} \xrightarrow{\text{glucose oxidase}} \text{gluconic acid} + \text{H}_2\text{O}_2 \quad (11)$$

The enzyme shows a very high degree of specificity for β-D-glucose[3,4] although 2-deoxy-D-glucose, D-mannose and D-fructose are also oxidized albeit at a much reduced rate.[5]

In 1956, Keston[6] proposed the simultaneous use of glucose oxidase and peroxidase (EC 1.11.1.7; donor : hydrogen peroxide oxidoreductase) coupled with a chromogenic electron donor (e.g. *o*-dianisidine, *o*-toluidine) for the colorimetric determination of glucose in body fluids. Since then, numerous workers have adapted the original method (e.g. variations in incubation time, temperature, buffer composition and choice of chromogen) and applied it to the determination of glucose in blood, urine and cerebrospinal fluid. The sequence of reactions that constitute this assay method are summarized in Equations (12) and (13).

$$\text{glucose} + \text{O}_2 + 2\text{H}_2\text{O} \xrightarrow{\text{glucose oxidase}} \text{gluconic acid} + \text{H}_2\text{O}_2 \quad (12)$$

$$\underset{\text{(colorless)}}{\text{H}_2\text{O}_2 + \textit{o}\text{-dianisidine}} \xrightarrow{\text{peroxidase}} \underset{\text{(brown)}}{\text{oxidized } \textit{o}\text{-dianisidine}} \quad (13)$$

The intensity of the brown color, which is stoichiometrically proportional to the amount of glucose present in the sample, is measured at 425–475 nm.[6] The chemical structure of *o*-dianisidine (3,3′-dimethoxybenzidine) is shown in Structure **(7)**.

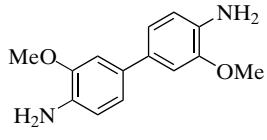

(7)

It has been reported in the literature that several substances may interfere with this procedure. Included amongst these are ascorbic acid, catechols, cysteine, glutathione, acetylsalicylic acid, L-dopamine (L-DOPA), mercurial diuretics and tetracycline. All these tend

to depress the glucose value.[7–14] The presence of exogenous peroxide, on the other hand, may increase the glucose value by oxidizing the chromogen.

Commercial kits based upon this method (e.g. Sigma-Aldrich Co. Ltd, product number 510) consist of 5 U of glucose oxidase and 1 U of peroxidase per milliliter of enzyme solution and *o*-dianisidine as the chromogen. (**Caution:** *o*-dianisidine is a possible carcinogen.) The limit of detection by this method is reported to be 25 mg dL^{-1}. Correlation between this method and the *o*-toluidine method described above has been reported. The correlation coefficient was 0.9857 (data adapted from Sigma Technical Bulletin, Procedure No. 510).

To perform the above assays, conventional instrumentation is required. Any photoelectric colorimeter that transmits light in the UV to visible range can be used. For measurements involving the detection of reduced nicotinamide coenzyme, hexokinase method (UV measurement), UV transparent quartz cuvets should be used.

An interesting development with colorimetric methods of glucose determination has been the development of glucose test strips. In their simplest forms, the sample is placed on the test strip (which contains the necessary freeze-dried enzymes and buffer salts for the test) and a color develops. After wiping, the amount of color is checked against a reference chart and the concentration of glucose deduced. Alternatively, the amount of color formed can be measured using a hand-held spectrophotometer. This idea is used in the One Touch® glucose meter developed by Lifescan Inc. California, USA. In this device, once a drop of blood is applied to the target area of the test strip, a blue color is developed which is proportional to the concentration of glucose in the blood sample. After 45 s, the meter gives an accurate reading of the blood glucose levels. No wiping or timing is required thereby making the whole monitoring procedure easy to carry out. Calibration and quality control routines are supplied by the manufacturers thereby ensuring correct operation of the meter.

4 ELECTROCHEMICAL DETECTION OF GLUCOSE

4.1 Pulsed Amperometric Detection of Glucose

The electrochemical detection of organic molecules using a fixed voltage is most successful when the product of the electron-transfer reaction at the interface is stabilized by delocalization through a conjugated π system. Inspection of the glucose molecule reveals that such stabilization is not possible. This means that any free radicals formed by oxidation or reduction are high-energy species, that is they posses a high activation energy. Consequently, large overvoltages are required for the electrochemical conversion, which is a disadvantage in electroanalysis since it introduces serious interference problems.

One solution to this problem is to carry out the oxidation at a catalytic surface such as the surface of a platinum or other noble electrode. This ensures that the oxidation takes place at a moderate potential through a mechanism that involves adsorption of the glucose to a clean platinum or gold electrode. Unfortunately, the products of the catalytic oxidation can remain adsorbed to the electrode thus blocking the catalytic sites and preventing further oxidation (poisoning of the surface). Consequently, the response of such a system is characterized by a time-dependent diminution of the response. The deleterious effects of product adsorption can be reversed and the catalytic activity of the electrode surface restored by a two-step cleaning process. The triple-pulse waveform incorporates the amperometric measurement with a potentiostatic cleaning and reactivation of the surface within a single waveform (see Figure 2). The first pulse is at a potential where the electrocatalytic reaction takes place. Since this step passivates the electrode, the next pulse is at a more positive voltage. This results in the oxidative removal of passivating species. Concomitant with this is the formation of a surface oxide layer. The electrode is reactivated by removal of the oxide layer using a negative potential. The electrode is now ready for the next detection step. Because the mechanism of detection involves prior adsorption, the shape of the overall calibration plot reflects in some way the corresponding adsorption isotherm. To examine a general example of analysis by this method consider the steps involved in

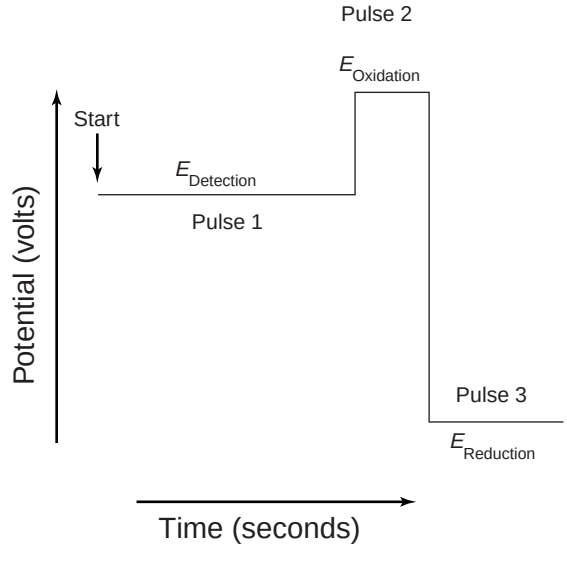

Figure 2 Triple-pulse amperometric waveform for PAD of carbohydrates.

the oxidation of glucose in 0.1 M NaOH at a gold electrode. As the voltage is swept positive (oxidizing) glucose is oxidized at the surface of the electrode. At the same time, the electrode surface is oxidized to gold oxide which inhibits any further oxidation of glucose. On the reverse scan, the gold oxide is reduced and a fresh gold surface is formed. Concomitantly, any adsorbed by-products from the oxidation process are removed from the electrode and the sequential oxidation of glucose can recommence.

The integration of triple-pulse amperometry with ion chromatography systems has resulted in a very powerful set of analytical tools especially when CHOH-bearing compounds are measured. Indeed, PAD at gold electrodes is a reproducible and sensitive method for the detection of almost all carbohydrates of molecular weight up to 10 000. Carbohydrates can only be detected by PAD however in solution of high pH (>11). This concept of detection has been successfully commercialized by the Dionex Corporation. The effectiveness of this technique is best illustrated in Figure 3 which compares the reproducibility of chromatograms obtained using PAD (trace a) and a conventional fixed voltage detection methodology (trace b).

Alternatively, catalytic oxidation of glucose has also been reported using electrodes modified with cupric salts or fabricated from the metal itself. It has been proposed that these electrodes function via a mechanism involving the copper(III) oxidation state.[15] One major advantage of using copper electrodes is that the products of the oxidation do not adsorb to the surface. Therefore, no complicated cleaning regime is required, implying that these electrodes can be operated at a constant voltage. As with the noble electrodes, however, the oxidation process is only specific for the alcohol or aldehyde group. Therefore, there is no discrimination between glucose and other sugars or alcohol and chromatographic separation is required (see Figure 4).

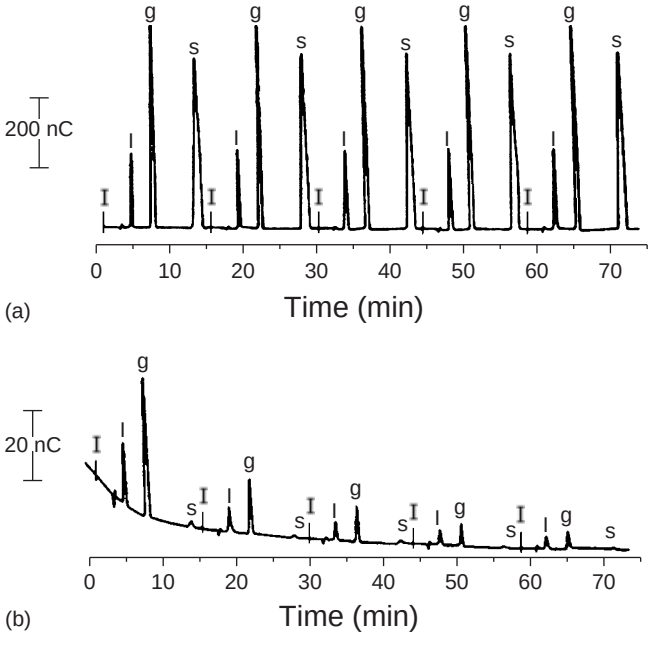

Figure 3 Comparison of (a) PAD and (b) constant potential detection of a solution containing (l) lysine (30 ppm), (g) glucose (10 ppm) and (s) sucrose (40 ppm). *I* is the injection of sample.

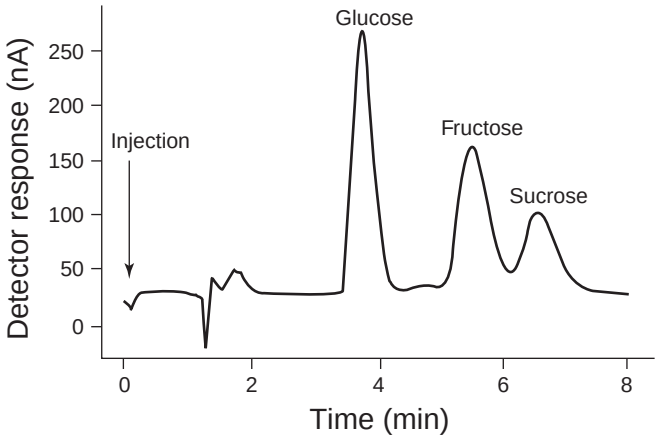

Figure 4 Chromatogram of grape juice recorded using a BAS LC-4C amperometric detector with a BAS copper electrode. (Data adapted from Bioanalytical Systems inc., West Lafayette, Indiana, USA.)

4.2 Enzyme Electrodes

The enzyme electrode is a combination of any type of electrochemical probe with a thin layer (10–200 μm) of immobilized (insolubilized) enzyme. Typically, the progress of the enzymatic reaction (which is related to the concentration of analyte) is measured by monitoring the rate of formation of a product or the disappearance of a reactant. If either the product or reactant is electroactive, the reaction may be monitored directly using amperometry.

Because enzyme electrodes generally have the enzyme immobilized in a thin layer of some finite thickness on the surface of the probe, any modeling of the system response must consider the diffusional and partitioning effects of the enzyme layer. It is interesting to note that unlike solution reactions, where the rate is dependent upon a steady-state population of encounter complexes, with immobilized enzymes diffusion of the substrate to the enzyme is often, but not necessarily, the rate-determining step. Mass transfer by diffusion is a first-order reaction with respect to substrate concentration. Imposing a diffusional barrier thus has the effect of extending the linear range of initial reaction velocity beyond the K_M value of the soluble enzyme (cf. Figure 1). Because of this linear relationship, however, the observed rate of

reaction, and therefore analytical signal, is lower than would have been in a kinetically controlled enzyme reaction conforming to the rectangular hyperbola of Michaelis–Menten kinetics.

The proper functioning of an enzyme-based sensor is, therefore, heavily dependent on both the chemical and physical properties of the immobilized enzyme layer. To this end, there are many possible methods for immobilizing an enzyme at the surface of an electrode. For convenience, these can be divided into physical methods and chemical methods. An important consideration when developing an immobilization procedure is that the process should be applicable to a range of surfaces. This allows the choice of support to be as wide as possible and ensures that no refabrication of the support is required. Other advantages sought in the immobilization method include: (1) an ability of the biological component to operate at a wider pH range than in solution; (2) attainment of greater stability resulting from the immobilization and (3) the generation of a defined diffusion region on the surface of the electrode.

The enzyme most widely used for glucose detection in this context is the flavoprotein glucose oxidase. At the active site of this enzyme is a flavin moiety (flavin adenine dinucleotide) which exists either in the oxidized form FAD or in the reduced form flavin adenine dinucleotide (reduced form) ($FADH_2$). FAD oxidizes glucose to gluconic acid and the $FADH_2$ formed in the process is reoxidized by molecular oxygen (or an artificial electron acceptor) yielding hydrogen peroxide or the reduced form of the artificial electron acceptor.

In order to be used as a glucose sensor, glucose oxidase activity must be converted into an electrical signal. The earliest electrochemical glucose sensors were based upon the electrochemical monitoring of the decrease in oxygen tension (disappearance of one of the reactants) or the appearance of hydrogen peroxide (product formation). The accuracy of the first method is limited by the natural fluctuations in oxygen tension that result from changes in pH, temperature, ionic strength or partial pressure which means that the preferred method of choice is based upon peroxide determination.

The electrochemical detection of hydrogen peroxide is normally achieved by its oxidation at a platinum anode polarized at a potential of 0.5–0.6 V vs the silver/silver chloride reference electrode. The requirements for such a high operating voltage mean that the measurement at a naked platinum electrode is subject to interference by naturally occurring opportunistic species such as ascorbic acid, uric acid, glutathione, cysteine and paracetamol. Use of suitable membrane technology (cf. the Yellow Springs Instrument Glucose Analyser) means, however, that this approach has been successful for the detection of glucose.

4.2.1 Glucose Oxidase Method (Electrochemical Measurement)

4.2.1.1 Measurements using the Oxygen Electrode An oxygen electrode basically consists of a platinum cathode separated from the sample solution by a membrane which is selectively permeable to oxygen, e.g. Teflon™ or polyethylene. A second, pseudo-electrode consisting of silver/silver chloride in saturated potassium chloride is used to complete the electrochemical cell. When a voltage of −0.6 V vs the Ag/AgCl reference electrode is applied the amount of current that flows in the external circuit is related to the partial pressure of oxygen in the sample via the relationship, Equation (14)

$$i = nF\alpha_m pP_s \left(\frac{D_m}{b} \right. \tag{14}$$

where n, the number of electrons involved in the reduction step = 4, F is the Faraday constant, α_m is the solubility of oxygen in the membrane phase, pP_s is the partial pressure of oxygen in the sample, D_m is the diffusion coefficient of oxygen in the membrane phase and b is the membrane thickness. Since diffusion through the membrane is likely to be the slowest contribution to the overall process, the response time, t_r, of an oxygen electrode can be approximated to (Equation 15)

$$t_r \approx \left(\frac{b^2}{D_m} \right. \tag{15}$$

The instrumentation required for this assay method is a commercially available oxygen electrode (Rank Brothers, Cambridge, UK).

4.2.1.2 Measurements based on Hydrogen Peroxide Detection An alternative approach to monitoring the glucose oxidase reaction electrochemically is to measure the production of hydrogen peroxide. Hydrogen peroxide is electroactive and can be conveniently measured by oxidation at a platinum anode according to Equation (16).

$$H_2O_2 \longrightarrow 2H^+ + O_2 + 2e^- \tag{16}$$

This approach forms the basis of instruments often referred to as glucose analyzers. Several are commercially available, e.g. the Yellow Springs Instrument Company Sensor,[16] and although the design varies from one manufacturer to another, a common feature is that the glucose oxidase is immobilized either as a membrane or in a packed bed thus allowing multiple measurements. In order to prevent interference from the oxidation of extraneous substances present in the sample, e.g. ascorbate, urease, tyrosine, paracetamol, glutathione and so on the anode is usually covered by a thin Nucleopore™ polycarbonate/cellulose acetate combination membrane. In

this configuration, the porous polycarbonate membrane limits the diffusion of glucose into the enzyme layer thus preventing the reaction from becoming enzyme-limited. Cellulose acetate permits only small molecules such as hydrogen peroxide to reach the electrode thereby eliminating many electrochemically active compounds that would interfere with the measurement. Alternatively, the response can be compensated by a nonenzymatic detector and the difference current used to calculate the true concentration of glucose. A summary of some of the commercially available glucose analyzers is given in Table 3.

Unlike peroxide measurement which starts from a very low basal level (no peroxide present) and permits a sensitive detector (detection limit ca. 1 μM), in the case of oxygen reduction a decrease in signal is monitored in response to glucose. Consequently, the sensitivity to glucose by this method is lower by two or three orders of magnitude.

4.2.1.3 Mediated Electrodes An alternative approach is to use oxidants other than molecular oxygen to regenerate the catalytically active flavin coenzyme. The ideal method would be direct electron exchange from the active site of glucose oxidase to the electrode surface but this is not possible since the active site is embedded in the protein and the distances involved are hence too great. Thus, electron transfer must be achieved through the use of mediators. Because of many desirable chemical and electrochemical properties, ferrocenes have been the most successful class of mediators to date. The simulated cyclic voltammograms of a glucose sensor based upon the combination of glucose oxidase with ferrocene is shown in Figure 5. In the absence of glucose (a), the reversible cyclic voltammogram of ferrocene is observed. When glucose is present (b), the response changes. The voltammogram displays an increase in oxidation current and the disappearance of the reduction wave, which is characteristic of catalytic reduction of ferrocenium by glucose oxidase.

This approach has been commercialized to make disposable glucose sensors for self-monitoring for diabetic patients. The Exac-Tech pocket-sized meter (MediSense, Abingdon, UK, and Bedford, MA, USA) is designed for mediated amperometric determination of glucose in capillary whole blood. It consists of a pen-shaped barrel, housing a custom-built single chip microprocessor, sealed power source liquid crystal display and operating button. The sensors are ferrocene-mediated glucose oxidase electrodes screen printed onto a plastic substrate. Several print runs are involved laying down silver impregnated carbon ink for electrical connection, and using a silver/silver chloride reference electrode, a carbon working electrode containing glucose oxidase and a ferrocene derivative. The construction is completed by a coarse gauze spreading layer covering the two electrodes. The enzyme electrodes respond to capillary blood glucose concentrations in the range 2.2–25 mM.

Because the spreading layer does not act as an anti-interference layer, the electrodes are prone to interference. A urate concentration of 700 μM, for example, causes an increase in the apparent glucose reading of 3 mM which could be clinically misleading to a diabetic. Capillary blood samples containing 700 μM urate are, however, rare. Paracetamol also interferes with the measurement but only at levels associated with overdose. Salicylate, acetone and acetoacetate also interfere but the effects are not clinically significant.[17] Typical responses to glucose utilizing these electrodes in this laboratory are shown in Figure 6. A more recent design of the mediated-glucose sensor is Medisense Card Sensor which has the approximate dimensions of a credit card. The operation

Table 3 Commercially available glucose analyzers based on glucose oxidase

Company	Enzyme used	Measuring range (mM)	Sample volume (μL)	Measuring frequency (sample h^{-1})	Serial precision (%)	Stability (samples)
Yellow Springs Instruments (USA)	Glucose oxidase (glucox)	1–45	25	40	2.0	300
ZWG, Academy of Sciences, Germany	Glucox	0.5–50	20–25	60–90	1.5	1000
Radelkis (Hungary)	Glucox	1.7–2.0	100	40	5.0–10	–
Fuji Electric (Japan)	Glucox	0–27	20	80–90	1.7	500
Seres (France)	Glucox	1.0–22.0	200	60	–	500
Analytical Instruments (Japan)	Glucox	0–55.5	20–40	120–150	2.0	–

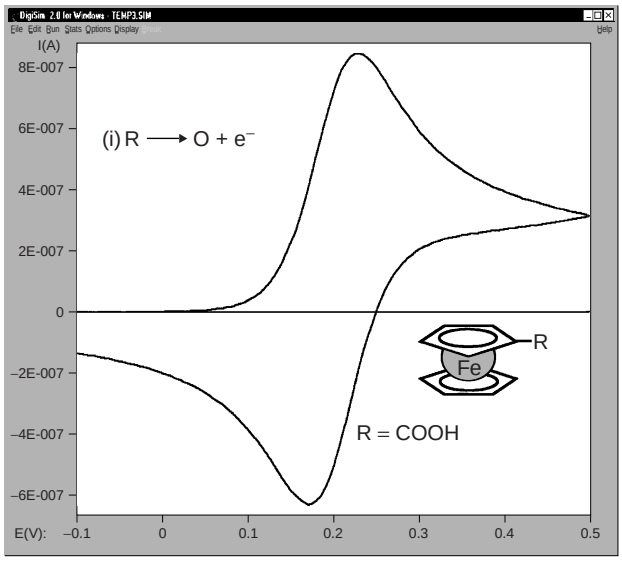

E(init) = –0.1 V; E(rev) = +0.5 V; v = 0.01 V s^{-1}

(a) [R] = 10^{-6} M

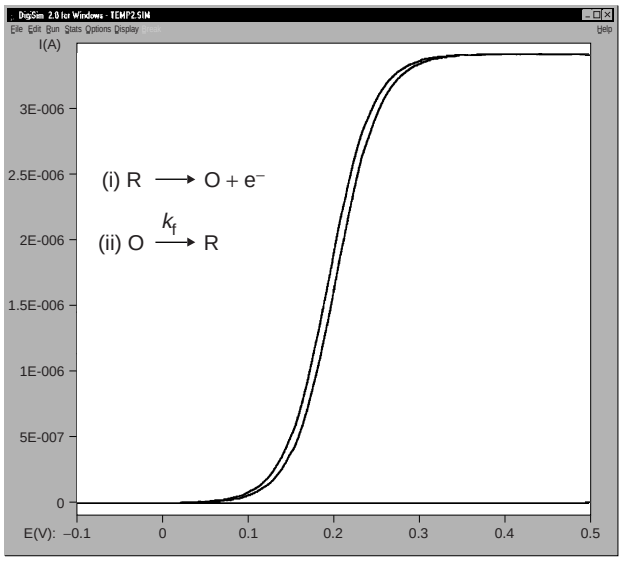

E(init) = –0.1 V; E(rev) = +0.5 V; v = 0.01 V s^{-1}

(b) [R] = 10^{-6} M; k_f = 10^{-3} s^{-1}

Figure 5 Simulated hydrodynamic cyclic voltammograms of a ferrocene-mediated glucose oxidase electrode in the absence (a) and presence of saturating levels of glucose (b). The voltammograms were simulated using DigiSim (version 2.0). Bioanalytical Systems Inc., USA. Simulation parameters are indicated on the right hand side of the voltammograms. The generic structure of ferrocene is shown in (a).

of this device is similar to the Exac-Tech Pen described above with the exception that once the drop of blood is applied to the target area of the glucose electrode, the blood glucose test starts automatically.

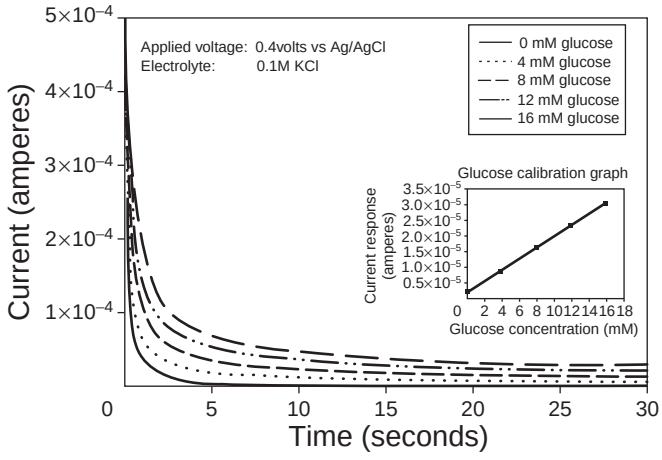

Figure 6 Transient responses to increasing concentrations of glucose obtained using commercially available glucose/ferrocene strip electrodes. The inset shows the calibration graph. Current values represent the response of the sensor after 30 s. The correlation coefficient using these electrodes was 0.999. (Data courtesy of Mr Kenneth McKeegan, Ph.D student, University of Paisley.)

Another successful approach involving mediators has been the ''wired'' enzyme electrode. Here, the mediators tend to be osmium bipyridine complexes which are cationic in nature and hence bind electrostatically to the anionic glucose oxidase molecule. This allows efficient exchange of electrons between the osmium centers of the complex and the active site of the enzyme via the mechanism of electron hopping. If one of the coordination sites of the osmium complex is occupied by the N-atom of an imidazole or pyridine moiety of a polyvinylimidazole or polyvinylpyridine polymer, then the mediator can be retained at an electrode surface in a three-dimensional polymer network. If the glucose oxidase is also entrapped within this network, electron hopping still occurs and the catalytic activity can be measured electrochemically. The advantage of this approach over the ferrocene-mediated case described above, is that the sensor is more stable because the mediator is physically retained at the electrode surface.

As mentioned above, one method of detecting glucose using the enzyme glucose oxidase is to measure the hydrogen peroxide produced at a platinum electrode. An alternative approach to peroxide detection is to use an enzyme electrode based on ''wired'' horseradish peroxidase. Here, coordinated osmium complexes are again used as the mediator. The wired peroxidase electrode can be used for the detection of glucose following separation by LC. As an illustration, its performance is compared with a traditional platinum electrode in Figure 7. In a typical experimental set-up, the sample containing glucose, ascorbic

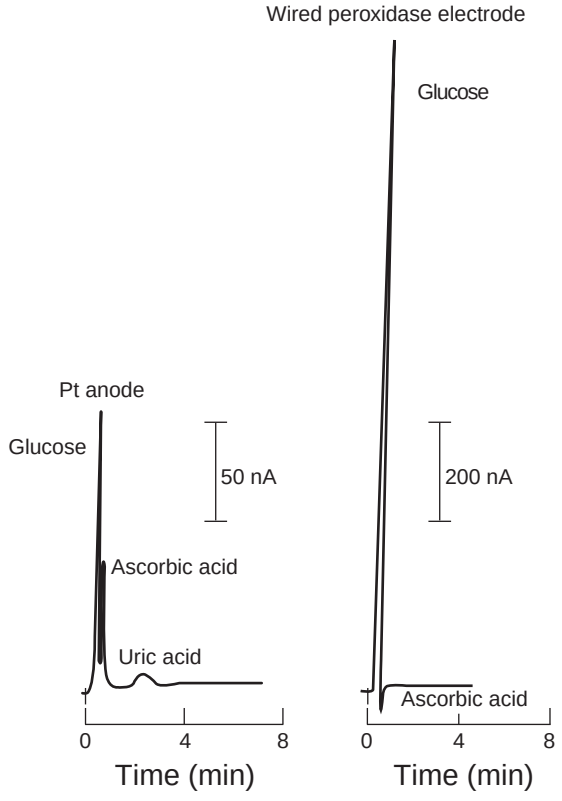

Figure 7 Chromatograms for 1 nmol glucose, 0.23 nmol ascorbic acid and 0.24 nmol uric acid using a platinum anode detector and a wired horseradish peroxidase enzyme electrode detector.

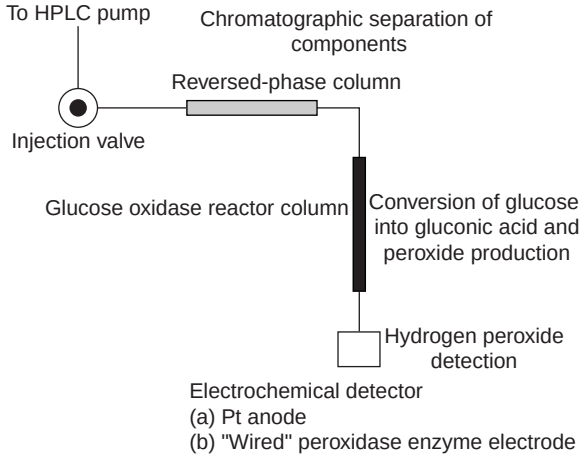

Figure 8 Components for the chromatographic separation and detection of glucose using either a platinum anode or a ''wired'' peroxidase enzyme electrode detector.

acid and uric acid is first passed through a reversed-phase chromatographic column. The eluant is then passed through a reactor column containing immobilized glucose oxidase which oxidizes the glucose and produces hydrogen peroxide. The peroxide is detected downstream either at a wired peroxidase electrode or a conventional platinum anode. This arrangement is summarized in Figure 8. Points to note with the wired electrode are (1) the sensitivity and detection limits for glucose are better than with the naked platinum detector, (2) the electrode is more stable than the platinum anode, (3) because of the lower operating voltage, the wired enzyme detector is less prone to interference and (4) the equilibration time for the peroxidase electrode is reportedly shorter than that required for the platinum electrode.[18] Instrumentation as well as the wired peroxidase enzyme and the packed glucose oxidase reactor column are available from Bioanalytical Systems Inc., West Lafayette, USA.

5 CONCLUSIONS

This article has discussed various analytical methodologies for the quantification of glucose. For convenience, these have been classified as traditional colorimetric methods involving the use of spectrophotometers or electrochemical approaches. I have, however, limited myself to only those methods which are commercially available. Where possible, reference has been made to technical information supplied by the manufacturer of the analytical procedure.

ABBREVIATIONS AND ACRONYMS

ATP	Adenosine Triphosphate
FAD	Flavin Adenine Dinucleotide (Oxidized Form)
$FADH_2$	Flavin Adenine Dinucleotide (Reduced Form)
INT	Iodophenyl-3-(4-nitrophenyl)-5-phenyl-tetrazolium Chloride
LC	Liquid Chromatography
L-DOPA	L-Dopamine
NAD^+	Nicotinamide Adenine Dinucleotide (Oxidized Form)
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
$NAD(P)^+$	Nicotinamide Adenine Dinucleotide Phosphate (Oxidized Form)
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
OEL	Occupational Exposure Limit
PAD	Pulsed Amperometric Detection
PMS	Phenazine Methosulfate
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Point-of-care Testing

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup

Food **(Volume 5)**
Enzyme Analysis and Bioassays in Food Analysis

Pulp and Paper **(Volume 9)**
Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Pulse Voltammetry

Kinetic Determinations **(Volume 12)**
Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics

REFERENCES

1. E. Hultman, 'Rapid Specific Method for Determination of Aldosaccharides in Body Fluids', *Nature*, **183**, 108–112 (1959).
2. K.M. Dubowski, 'An *o*-Toluidine Method for Body-fluid Glucose Determination', *Clin. Chem.*, **7**, 215–222 (1962).
3. D. Keilin, E.F. Hartree, 'Properties of Glucose Oxidase (Notatin)', *Biochem. J.*, **42**, 221–226 (1948).
4. D. Keilin, E.F. Hartree, 'Specificity of Glucose Oxidase (Notatin)', *Biochem. J.*, **50**, 331–338 (1952).
5. R.B. McComb, W.D. Yushok, W.G. Batt, '2-Deoxy-D-glucose, a New Substrate for Glucose Oxidase (Glucose Aerodehydrogenase)', *J. Franklin Inst.*, **263**, 161–170 (1957).
6. A.S. Keston, 'Specific Colorimetric Enzymatic Analytical Reagents for Glucose', Abstract of Papers, 129th Meeting, ACS, Dallas (Texas), April, 310, 1956.
7. L.P. Cawley, F.E. Spear, R. Kendall, 'Ultramicro Chemical Analysis of Blood Glucose with Glucose Oxidase', *Am. J. Clin. Pathol.*, **32**, 195–205 (1959).
8. R.J. Henry, *Clinical Chemistry–Principles and Technics*, Harper & Row, New York, 632–635, 1964.
9. J.M. Feldman, W.N. Kelly, H.E. Lebovitz, 'Inhibition of Glucose Oxidase Paper Tests by Reducing Metabolites', *Diabetes*, **19**, 337–342 (1970).
10. P.D. Hansten, *Drug Interactions*, Lea and Febiger, Philadelphia, 1971.
11. F.H. Meyers, E. Jawetz, A. Goldfien, *A Review of Medical Pharmacology*, Lange, Los Altos, CA, 1968.
12. M.P. Elking, H.F. Kabat, 'Drug Induced Modification of Laboratory Test Values', *Am. J. Hosp. Pharm.*, **25**, 485–493 (1968).
13. F.W. Fales, J.A. Russel, J.N. Fain, 'Some Applications and Limitations of the Enzymatic Reducing (Somogyi) and Anthrone Methods for Estimating Sugars', *Clin.Chem.*, **7**, 289–296 (1963).
14. E. Raabo, T.C. Terkildsen, 'On the Enzymatic Determination of Blood Glucose', *Scand. J. Clin. Lab. Invest.*, **12**, 402–418 (1960).
15. J.M. Zadeii, J. Marioli, T. Kuwana, 'Electrochemical Detector for Liquid Chromatographic Determination of Carbohydrates', *Anal. Chem.*, **63**, 649–653 (1991).
16. YSI 2700 Select, Biochemistry Analyzer, Technical Bulletin, YSI incorporated, Yellow Springs, OH.
17. M.F. Cardosi, A.P.F. Turner, 'Mediated Electrochemistry: A Practical Approach to Biosensing', *Advances in Biosensors*, **1**, 125–170 (1991).
18. A.W. Bott, 'Electrochemical Methods for the Determination of Glucose', *Current Separations*, **17**, 25–30 (1998).

Immunochemistry

Christopher P. Price
St. Bartholomew's and the Royal London School of Medicine and Dentistry, London, UK

1 Introduction 1
2 The Antigen–Antibody Reaction 2
2.1 Molecular Chemistry 2
2.2 Thermodynamics 3
2.3 Kinetics 3
2.4 Factors that Influence the Antigen–Antibody Reaction 4
3 Preparation of Antibody Reagents 5
3.1 Immunogen 5
3.2 Polyclonal Antisera 5
3.3 Monoclonal Antibodies 5
3.4 Antibody Engineering 6
3.5 Molecular Imprinting 6
3.6 Aptamers 6
4 Labels Used in Immunochemistry 7
4.1 Radioisotopes 7
4.2 Enzymes 7
4.3 Fluorophores 9
4.4 Luminescence 9
4.5 Substrate/Inhibitors/Enhancers 10
4.6 Particles and Metal Sols 10
5 Recognition and Characterization 10
5.1 Immunoelectrophoresis 10
5.2 Immunofixation 10
5.3 Western Blotting 11
5.4 Immunocytochemistry 11
5.5 Flow Cytometry 11
5.6 Immunosensors 12
6 Quantitation 13
6.1 Heterogeneous Immunoassays 13
6.2 Homogeneous Immunoassays 15
7 Microfabrication and Encapsulation 17
7.1 Disposable Reaction Cells 17
7.2 Flow-through Cells 18
7.3 Lateral Flow Devices 18
7.4 Miniaturization of Devices 19
8 Applications 19
8.1 Calibration and Quality Assurance 19
8.2 Automation 20
8.3 Hapten Assays 20
8.4 Macromolecule Assays 21
8.5 Limitations and Interferences 21
9 Concluding Remarks 22
Abbreviations and Acronyms 22
Related Articles 22
References 23

Immunochemistry is a generic term to describe all analytical methods that employ the reaction between an antigen and its complementary antibody for the recognition and/or quantitation of an analyte. The analyte may be the antigen or antibody and in each case the complementary species (antibody or antigen, respectively) will be the primary reagent in the analytical system. The analyte may be present in free solution (e.g. in blood, urine, fermentation broth, effluent, etc.) or may be structurally incorporated in a cell (e.g. red cell, bacterium, virus particle) or solid tissue. Our knowledge of the chemical structure and reactivity of the combining sites of antigens and antibodies (epitopes and paratopes, respectively) has meant that antigen- and antibody-like species can now be synthesized. In the former case this may be undertaken to provide an appropriate immunogen or calibration material whilst several antibody mimics have been demonstrated.

The immunoassay format will depend on the purpose for which the assay is to be used. In that the antigen–antibody reaction is unique to two complementary partners, in terms of cells and tissues, there is no other technique that will provide the same specificity – with the exception of DNA. Quantitative assays are usually required for the analysis of molecules in complex fluids. In the research phase, chromatographic techniques, e.g. high-performance liquid chromatography (HPLC), gas chromatography (GC), or capillary zone electrophoresis (CZE) can provide the characterization and quantitation required; however, these techniques are not always applicable when the need for the assay is established and a large workload is envisaged. An immunoassay then provides the necessary specificity and workload capability. In addition, when assays are used outside the laboratory, e.g. at the bedside, then only immunoassays are applicable because they can be encapsulated in microfabricated devices.

This article describes the underlying principles of immunoassay, the design of reagents and assay formats, and illustrates their application to a wide range of analytes.

1 INTRODUCTION

Immunochemistry is a term that embraces all aspects of the reaction between an antibody and its complementary antigen. The production of an antibody is a biological phenomenon and represents the product of a response of a higher vertebrate (the immune response) to the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

presence of a foreign species (the immunogen or antigen). The antibody that is produced binds tightly to the foreign species, thereby hastening its removal from the circulation. The tightness of this binding is a result of the complementarity between the antigen and the antibody molecules – in terms of chemistry (i.e. thermodynamics) and spatial orientation (i.e. topography of molecular surfaces).

Immunoanalysis describes any technique that utilizes an antibody or antigen in order to recognize (and quantitate) its complementary partner (namely the antigen or antibody, respectively). Whether the purpose is characterization or quantitation of a species (the analyte of interest), the quality of assay performance is primarily dependent on the nature of the antibody or antigen employed as the primary capture or recognition reagent.

The reaction between antigen and antibody is the result of an array of chemical interactions which are easily modified by the nature of the reaction milieu, and consequently reagent design and optimization are a vital feature in ensuring good analytical performance.

The following discussion will focus on a description of current knowledge of the antigen–antibody reaction, on the production of an antibody preparation, and on the design and optimization of immunoanalytical systems. Finally, examples will be given to illustrate the range and quality of assay performance that can be expected.

2 THE ANTIGEN–ANTIBODY REACTION

The product of the immune response in higher vertebrates is a family of antibodies that bind to varying degrees (in terms of both avidity and site on the surface of the molecule) to the antigen. Antibodies comprise a family of proteins known as immunoglobulins; the immunoglobulins are heterodimers of a similar architecture. The basic unit comprises two heavy chains of about 450–600 amino acid residues and two light chains of about 220 residues; the chains are linked by disulfide bridges (Figure 1). In mammals there are five immunoglobulin classes, IgG, IgA, IgM, IgD, and IgE; they differ in the amino acid sequence and carbohydrate content of the heavy chains. There are a total of nine heavy-chain isotypes; however, there are only two types of light chain, kappa and lambda. In the case of IgA and IgM the molecules are dimeric and pentameric versions of the basic immunoglobulin molecule with extra of additional residues, secretory component with J chain and J chain, respectively. IgA generally exists in monomeric form in the circulation.

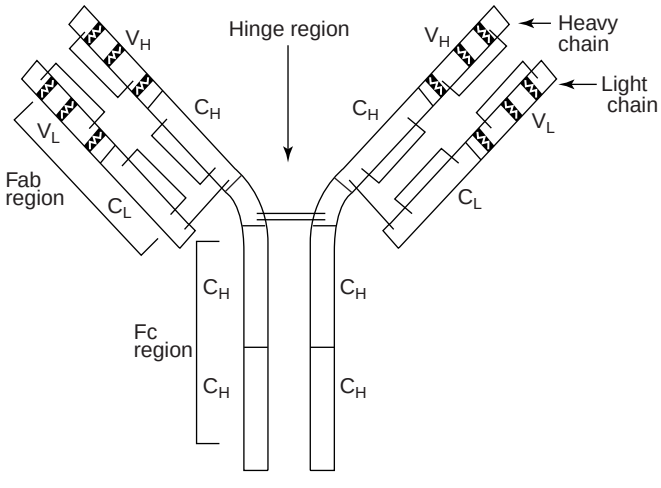

Figure 1 Representation of the immunoglobulin G molecule indicating heavy and light chains, variable and constant domains (V and C), and hypervariable regions (▩). Solid lines denote disulfide bridges.

The antigen-binding region of the immunoglobulin, known as the paratope region, is located in the amino-terminal domain of the heavy and light chains. In these regions there are so-called variable regions, and within each of these variable regions there are three regions of hypervariability which constitute the complementarity-determining regions (CDRs) of the immunoglobulin.[1] It is the variability of amino acid residues in the CDRs that imbue the antibody with unique specificity for an antigen or antigenic region of a molecule.

2.1 Molecular Chemistry

The basic structure of the immunoglobulin molecule is in the form of two β-pleated sheets of strands in antiparallel directions linked by a disulfide bridge. In the paratope region of the immunoglobulin the variable element of the heavy and light chains associate noncovalently to form a β-barrel structure that brings the six CDR close together; it is this colocation that creates the antigen-binding pocket or paratope region. Both the length of the CDRs and the amino acid sequences are variable.

Detailed analysis of the chemistry and structure of immunoglobulins and analysis of antigen–antibody interactions using techniques such as X-ray diffraction suggest that the region of the antigen (the epitope) that is in contact with the immunoglobulin is about 700–900 Å in area made up of 15–22 amino acid residues within the immunoglobulin. The paratope comprises a limited number of residues originating from most or all of the CDRs; about half of the CDR residues are aromatic.[2] It is suggested about one third of the CDR residues may contribute to a unique paratope and it has been shown that more than one unique paratope may exist within one antigen-binding pocket; more commonly, epitope sites may overlap, although concurrent binding of respective paratopes (i.e. antibodies) will not happen in practice owing to steric hindrance.

The region of the antigen that binds with the antibody is known as the antigenic-determining region or epitope. In the case of a macromolecule such as a protein, it will constitute an area and number of residues similar to that of the paratope. The epitope may comprise a series of either continuous or discontinuous residues, more commonly the latter. Thus the reaction between antigen and antibody will depend not solely on the chemical structure of the residues but also on the spatial configuration.[3,4] This is an important observation because purification of a macromolecule in order to use it as an immunogen may destroy the unique epitope and result in no functional antibody being produced; it is also an important consideration in the use of a synthetic peptide as an immunogen. It has been suggested that in a macromolecule the epitope involves only a few residues and equates to a hapten size of the order of 400–500 Da; these data are supported by studies of affinity constants for haptens of different sizes in which little increase is seen for haptens of about 500 Da.[5]

2.2 Thermodynamics

In aqueous solutions molecules exist as discrete entities by virtue of a surrounding hydration shell of water molecules; thus in the case of proteins the loss of the influence of the hydration shell leads to aggregation and precipitation of the protein molecules. The repulsion between protein molecules in order to keep them apart (i.e. in solution) operates over a distance of 20–30 Å and the primary forces of attraction between the epitopic region of an antigen and the complementary paratope region of the antibody must overcome this barrier. Thus the energy of attraction at a distance of 30 Å must be sufficient to overcome the hydrophilic forces of repulsion.[6]

There are four types of attractive forces involved in the antigen–antibody reaction. The major forces in the initial or primary reaction are due to hydrophobic and electrostatic interactions. The hydrophobic force is the result of the tendency of nonpolar residues to aggregate when immersed in water. The strong hydrophobic forces of the CDRs are therefore complementary to the predominantly hydrophilic pattern of residues within the epitope region. The electrostatic interactions result from the ionized species of the epitope and the oppositely charged species of the paratope. The force of the interaction is inversely proportional to the distance between the charged groups and also depends on the ionic strength of the surrounding milieu. Thus in more dilute salt solutions electrostatic forces operate over greater distances than in a solution of high ionic strength.[7]

The hydrophobic and electrostatic forces of the primary reaction will draw the antigen and antibody surfaces together and then the secondary van der Waals interactions and hydrogen bonding come into play. Thus whilst the primary forces can influence over distances of 30–100 Å, the secondary attractive forces are only effective over distances of 2–10 Å, with hydrogen bonding being over the shorter distances. Hydrophobic interactions are also important in the secondary phase of the antigen–antibody reaction involving the ''squeezing out'' of water molecules that surround the surface of both epitope and paratope regions.[8,9]

In the latter stages of the reaction and, in part, depending on the size of the antigen there may be further interactions involving part of the molecules outside of the complementary epitope : paratope regions. This can result in aggregation of the molecules, attested to by the fact that it can become increasingly difficult to dissociate the complexes.

2.3 Kinetics

Whilst the reaction between antigen and antibody may be complex, the complexity is dependent, in part, on the size of the antigen molecule. It can be described in a simple equation derived from the law of mass action. Thus the reaction can be described as Equation (1):

$$\mathrm{Ag} + \mathrm{Ab} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{Ag:Ab\ complex} \tag{1}$$

where k_a and k_d are the association and dissociation constants, respectively. The association (K_a) and dissociation (K_d) constants at equilibrium are then described by Equations (2) and (3):

$$K_a = \frac{k_a}{k_d} = \frac{(\mathrm{Ag:Ab})}{[\mathrm{Ag}][\mathrm{Ab}]}\,\mathrm{L\,mol^{-1}} \tag{2}$$

$$K_d = \frac{k_d}{k_a} = \frac{[\mathrm{Ag}][\mathrm{Ab}]}{(\mathrm{Ag:Ab})}\,\mathrm{mol\,L^{-1}} \tag{3}$$

This is a simplified view because, whilst it is important to recognize that the antigen–antibody interactions are reversible, the energy of the dissociation step is usually higher than the energy of association, particularly when the secondary bonding formation has occurred. This observation is known as hysteresis[6] and is more applicable to complex antigens reacting with their complementary antibody molecules, i.e. more applicable to proteins than haptens.

The association and dissociation constants of the reaction give a measure of the affinity constant, equivalent to the equilibrium constant, and will have an important influence on the design of an immunoassay. Strictly the affinity constant of an antibody should be equated to the association constant of the primary reaction, but in practice the contribution of the secondary association reaction and the influence of dissociation will contribute to the observed value. In the case of a polyclonal

antiserum the situation is further complicated because the observed affinity constant will be a composite measure of the primary association and dissociation constants of the range of individual antibodies. Thus in a heterogeneous immunoassay, using a polyclonal antiserum, the presence of antibodies with a low association constant and/or high dissociation constant will have little impact on the observed reaction. Antibodies with a high dissociation constant will be of little value in a heterogeneous immunoassay because the antigen–antibody complex will dissociate during the washing step.[10] When considering the design of an immunoassay, antibody affinities up to 10^{12} L mol^{-1} are achievable, although in practice the majority of antibodies that are employed possess affinities within the range of 10^7–10^9 L mol^{-1}.

A range of techniques have been employed to study antigen–antibody interactions; they all depend on a characteristic of the reactants or products of the reaction and include spectroscopic, chromatographic, sedimentation, and calorimetric methods. The most commonly used methods are enzyme-linked immunosorbent assay (ELISA), calorimetry, and optical biosensing.

In the case of ELISA, antibody at a fixed concentration is incubated with various concentrations of antigen; at equilibrium the mixture is then transferred to another tube coated with antigen. The free antibody binds to solid-phase antigen and after washing is determined with the aid of a labeled second antibody. In this method labeling of the primary reactants is avoided.[11] This technique cannot be used for kinetic analysis.

The heat of reaction is used to measure the amount of complex in calorimetry, with titration by antigen at a fixed antibody concentration or vice versa. The method is not sensitive enough for affinities higher than 10^8 L mol^{-1} and cannot be employed for kinetic analysis.[12]

The major development in this field is the introduction of surface plasmon resonance techniques.[13] The technique permits real-time monitoring of binding, facilitating the assessment of both kinetics and concentration of analyte. The method depends on the change in refractive index at the optical surface, and so is more suited to binding of antibody or high-molecular-weight antigen. However, the binding of a hapten from solution can be assessed by using a high-molecular-weight label, e.g. a latex particle.[14] There are important issues to consider in this approach to kinetic analysis, including (1) the effect of coupling one of the species (antigen or antibody) to the optical surface on the conformation of the bound species, (2) the influence of the proximity of the surface on binding, (3) the effects of mass transfer, namely the diffusion effects, and (4) the influence of sample depletion immediately surrounding the immobilized reactants.[15] These factors can be taken into account and there is no doubt that this type of real-time reaction monitoring has contributed a great deal to our knowledge of antigen–antibody interactions.[16]

2.4 Factors that Influence the Antigen–Antibody Reaction

The rate and strength of the reaction between antigen and antibody are determined by the complementarity of the epitope and paratope regions, respectively, being modified to various degrees by other constituents of the reaction milieu. Bearing in mind the earlier discussion on molecular chemistry, the reaction characteristics are determined by the sequence of chemical substituents, the nature of the intermolecular forces of attraction and repulsion, and the spatial disposition of the reactive substituents. The basic reaction takes place in two phases; the first is rapid, specific and reversible and involves overcoming the repulsive forces between antigen and antibody. The reactions involve hydrophobic and electrostatic interactions; as an example, the reaction can be inhibited by increasing the pH beyond 10 in most instances. The secondary reaction involves hydrophobic interactions again, in addition to van der Waals force and hydrogen bonding. Thus, conditions that remove the water shield around molecules will enhance the rate of reaction. However, in considering all of the forces of interaction there may be little apparent difference between the specific antigen–antibody reaction and the nonspecific interaction between molecules; thus in many cases reaction conditions that promote the rate of reaction between protein and antibody will also tend to promote nonspecific protein precipitation, e.g. the use of ammonium sulfate and poly(ethylene glycol).

The charge on the surface of paratope and epitope regions clearly plays an important role in the initial primary binding reaction and thus the pH of the reaction environment, the nature of the ionic species, and the ionic strength are all important. It should also be recognized that these characteristics will also have an effect on the conformation and stability of the proteins present. In addition, the concentration of the constituent antigen and antibody species will also have an important influence on the speed of the reaction and on the colloidal stability of the complex formed. Hence most immunoassay reactions are performed around pH 7; at lower pHs there is a tendency for nonspecific association between protein molecules to occur with little specific binding. At higher pHs the nonspecific effects are less evident although the specific effects are also less than at pH 7.[17]

The nature of the ionic species plays an important role in both specific and nonspecific reactions and whilst considerable variation is seen there are general observations that can be made. There is a chaotropic series of anions (the Hofmeister series) in which phosphate is

known to be antichaotropic and thus promote association whilst certain amino acids and amino alcohols, borate, and thiocyanate are chaotropic.[18] Phosphate buffers can therefore be used successfully whereas buffers containing glycine or borate, whilst good for storing reagents, yield poor reaction characteristics. These observations are particularly true for immunoagglutination reactions where the concentration of constituents is often higher than in labeled heterogeneous assays. Polymers such as the poly(ethylene glycol)s can increase the rate of reaction in many systems, again particularly in the case of the agglutination reactions. It is thought that they act by removing water molecules from the immediate environment surrounding the epitope and paratope regions, promoting binding.[19]

Finally, although often not considered to be important (say compared with the effect on enzyme activity), temperature can have a major influence on the rate of the antigen–antibody reaction. The major effect will be on the rate of diffusion, and thence collision, of reactants.

3 PREPARATION OF ANTIBODY REAGENTS

Immunochemistry is founded on a natural biological phenomenon and the early developments were all based on the use of polyclonal antisera: such preparations containing a population of antibodies varying in their affinity and thence specificity for the antigenic partner. The development of monoclonal antibodies by Kohler and Milstein in 1975[20] revolutionized the whole field of immunoreagents, and the advent of antibody engineering has taken this process of refinement a stage further. It should be evident from the earlier discussion that considerable research has been devoted to the molecular chemistry of the antigen–antibody reaction and it is therefore not surprising that synthetic chemists have developed a range of options for producing antibody mimics. However, today the majority of reagents are still based on polyclonal or monoclonal antibody preparations.

3.1 Immunogen

Generally a molecule needs to be greater than 3000 Da in size to produce an antibody response, and in order to produce a polyclonal antiserum a purified preparation of antigen is required. In the case of low-molecular-weight antigens a carrier molecule is required in order to achieve a response; typical carriers include bovine serum albumin, keyhole limpet hemocyanin and thyroglobulin.[21] In the case of large macromolecules, where a unique specificity is required, synthetic peptides have been employed to achieve the desired complementarity. When producing a monoclonal antibody the purity of the immunogen is less important theoretically, because the specificity can be selected at the time of subcloning.

3.2 Polyclonal Antisera

It is worth considering the immune response for a moment when a ''foreign antigen'' enters the bloodstream of a higher mammal. The initial response is the expansion of antigen-specific T and B lymphocytes with the production of a small amount of immunoglobulin M followed by the production of immunoglobulin G (the primary response). The introduction of a second bolus of immunogen leads to a secondary response with the differentiation of the B cells into plasma cells and the production of larger amounts of immunoglobulin G. The result will be a mixed population of antibodies reflecting specificity for different parts of the antigen molecule.

A range of mammals have been used for the production of antisera including mice, rabbits, goats, and sheep. A wide range of immunization protocols have been described based on the injection of a preparation of antigen dissolved in an adjuvant (e.g. complete Freund's adjuvant). A second injection may be given after several days (e.g. 14) and a third after a similar period. The animal may be bled after 21 days (e.g. in the case of a mouse) to determine the presence and titer of antibody response with blood collection for antisera after about 5 weeks.[22]

The wide range of protocols encompass a host of conditions that research workers have found optimal for their applications, including the exact nature of the immunogen, the type of adjuvant, the type and strain of animal, the route of immunization, and the timing of injections. The choice of animal will often depend on the amount of antiserum required.

3.3 Monoclonal Antibodies

The core of the production of monoclonal antibodies is the fusion of the immune spleen cell with a myeloma cell.[20] In order to achieve this the spleen is removed from the immunized animal (typically a mouse and about 35 days after the initial immunization) and teased into a single-cell suspension. The spleen cells are gradually mixed with the myeloma cells in the presence of poly(ethylene glycol) to promote fusion. The suspension is then diluted in first a serum-free and then a serum-containing medium. The fused cells are then cultured in a large number of wells containing growth medium incorporating hypoxanthine, aminopterin, and thymidine (HAT), which induces apoptosis in the non-fused cells, which die. The fusion preparation is then fed every few days for about 3 weeks until confluent

growth is achieved.[23] The supernatant from the fused hybridoma preparation can then be used to screen for the presence of antibody. A range of techniques can be used for screening, depending on the antigen, including radioimmunoassay, ELISA, antibody capture, and Western blotting.

Once the antibody-producing cells have been identified, it will be necessary to employ subcloning, to ensure monoclonality, and cryopreservation. The subcloning also helps to stabilize cell growth and prevent the overgrowth of nonsecreting cells; it is most commonly achieved by limiting dilution. Large-scale production of antibody can then be undertaken either in vivo by growth in the peritoneal cavity of the animal (most usually when the mouse is employed for immunization) or in vitro using cell culture. Many techniques have been described for the production of large amounts of antibody by cell culture with downstream processing to achieve purified material.[24]

3.4 Antibody Engineering

As knowledge of the structure and genetics of the immunoglobulin molecule has increased and with the ability to manipulate DNA, it has become possible to engineer recombinant molecules. A number of genetic modifications of the immunoglobulin have been described, the best known being the humanizing of the molecule to facilitate therapies employing antibodies.

One of the techniques that has been described for the modification of antibody molecules used for diagnostic purposes employs the production of a single-chain antibody.[25] This is based on the isolation of the Fv fragment of the antibody which can then be cloned by standard recombinant techniques with expression of the genes coding for the V_H and V_L domains stabilized with a peptide bridge sequence. Modification of the nucleotide sequence by, for example, site-directed mutagenesis, can then be used to change the CDR, thereby altering the specificity of the recombinant antibody.[26]

Bivalent or bispecific antibodies have also been produced using genetic engineering, one option being the linkage of two single-chain antibodies. An antibody fragment can also be used to create a fusion protein with another protein or fragment; the fusion of antibody fragments with enzymes has been described by several research groups.[27]

An alternative approach to engineering of an antibody from an existing hybridoma is to obtain novel antibody specificities from a gene library using phage display technology. In addition to the opportunity to obtain novel specificities, this approach offers an alternative to polyclonal or hybridoma technology.[28]

3.5 Molecular Imprinting

Knowledge of the chemical substituents and reaction mechanisms harnessed in the complementary epitope and paratope regions has facilitated the creation of a paratope pocket in a polymeric matrix. Initially monomers are chosen that are capable of interacting with specific residues of the molecule of interest. A second cross-linking monomer is then added that fixes the ''recognition monomers'' that are bound to the imprint molecule of interest. If this imprint molecule is then removed (e.g. by use of a solvent), an imprint cavity is left which is then capable of participating in a competitive reaction similar to an antibody molecule.[29] A heterogeneous population of binding pockets is created in the polymeric structure and affinity constants of $10^9\,L\,mol^{-1}$ are achieved with a limit of detection of hapten of $6\,nmol\,L^{-1}$ being demonstrated. Assays have been described for a range of hapten molecules such as morphine and theophylline, giving comparable results with established methods.[30]

3.6 Aptamers

Aptamers are nucleic acid molecules that bind ligands and are derived from the selection of random sequence pools of oligonucleotides of DNA or RNA molecules.[31] Modifications to this approach, known as systematic evolution of ligands by exponential enrichment (SELEX), has further expanded the opportunities for exploitation; these have included the use of partial randomized pools as the starting point for defining functional sequences, the introduction of modified oligonucleotides to stabilize the aptamer molecule, and the ability to link them to signal molecules.[32] However, the major advantage is the ability to build combinatorial libraries of compounds that can then be evaluated for their binding properties, and then the preferred sequence enriched using the polymerase chain reaction. Examples include aptamers that inhibit enzymes such as human neutrophil elastase and human thrombin. Antiprotein kinase C aptamers have been demonstrated in the equivalent of a ligand binding assay,[33] whilst an aptamer that was bound to vascular endothelial growth factor was synthesized with a fluorescein molecule coupled to its 5′ end; the antigen was subsequently detected by use of antifluorescein antibody labeled with alkaline phosphatase.[34] Fluorescein-labeled aptamers directed against human neutrophil elastase have also been used in flow cytometry.[35]

Experience to date indicates that aptamers possess binding affinities similar to those of antibodies. They may therefore provide an alternative means of producing the paratope pocket which can be produced in bulk and is applicable to toxic molecules and molecules that are not strongly immunogenic.[36]

4 LABELS USED IN IMMUNOCHEMISTRY

Immunochemical reactions are employed for the recognition and (possibly) quantitation of the substance of interest. In both of these applications it is necessary to detect the binding of antibody to antigen; the majority of immunochemical techniques that have been developed have required the use of a label of either antigen or antibody to detect the presence and extent of the reaction.

4.1 Radioisotopes

The first label described for an immunoassay was iodine-131 in a radioimmunoassay for insulin. The most commonly used labels in quantitative immunoassays have been iodine (^{125}I), carbon (^{14}C), cobalt (^{57}Co), selenium (^{70}Se) and tritium (^{3}H). The carbon and tritium isotopes emit β charged particles (radioactivity); the charged particles arising from the radioactive disintegrations are absorbed by a phosphor and converted into light which is then detected by a photomultiplier tube. The phosphors are organic compounds with examples being anthracene and naphthalene; it is common for mixtures of phosphors to be used. As the amount of energy produced is low the isotope and phosphor need to be in close proximity and the sample is therefore dissolved in a solution containing the phosphor (liquid scintillation counting).[37]

The cobalt, iodine, and selenium isotopes emit λ charged particles and this radiation is detected using inorganic phosphors such as thallium-activated sodium iodide crystals which are optically coupled to a photomultiplier tube. A feature of the use of λ radiation has been the design of multihead γ counters which greatly speed up the process of counting batches of tubes; subsequently a few fully automated immunoassay analyzers incorporating isotope counting were developed.

The choice of label depends in part on the sensitivity required for the assay and the nature of the labeled species. Thus the larger iodine molecule can be used for labeling proteins with attachment well away from the epitope or paratope region whereas carbon or tritium labels are required for smaller hapten molecules in order to preserve the epitope region. The iodine label is used because it can be readily incorporated into many molecules through the use of the cationic iodine species produced by the oxidation of iodide; a typical oxidant used is chloramine T, lactoperoxidase, or iodogen. In the iodination of proteins it is usually effected through the tyrosine residues present. An alternative approach, if there are no tyrosine residues present, is to conjugate a molecule containing the radioiodine to the compound of interest – pure antigen in most instances. A typical conjugation molecule is that described by Bolton and Hunter, i.e. radioiodinated *N*-succinicimidyl-3-(4-hydroxyphenyl) propionate, which reacts with primary amine groups.[38] Tritium labeled compounds are produced either by direct chemical synthesis or isotope exchange reactions.

The production and use of radioisotopes are heavily regulated and it is for this reason that alternative labels have been sought.

4.2 Enzymes

The choice of an enzyme as a label is logical because of its role as a biological catalyst. However, an enzyme provides the additional flexibility because of the alternative substrates available that enable the use of different detector technologies.

4.2.1 Choice of Enzyme

The choice of enzyme is primarily determined by the availability of a purified preparation, the stability of the enzyme in both a purified and diluted form and also during the conjugation process and when conjugated, together with the type of substrates available and the ease of detection of product.[39] In the case of a homogeneous assay (discussed later), the enzyme of choice must not be present in the sample, in order to avoid interference from endogenous enzyme during measurement. A list of some of the enzymes employed in immunoassays is given in Table 1.

4.2.2 Conjugation of Enzymes

A range of protein conjugation techniques have been described for coupling the enzyme to the antigen or antibody molecule; the final choice is based on what is shown to work but preservation of the enzyme activity and protection of the epitope and paratope

Table 1 A few examples of the enzymes commonly employed in immunoassays

Enzyme	Product	Detection
Alkaline phosphatase	Nitrophenol	Absorbance
	Methylumbelliferone	Fluorescence
	Dioxetane derivative	Luminescence
	Aminophenol	Electrochemical
β-Galactosidase	Nitrophenol	Absorbance
	Methylumbelliferone	Fluorescence
	Dioxetane derivative	Luminescence
Glucose oxidase	Oxidized dye	Absorbance
Glucose-6-phosphate dehydrogenase	NADH	Absorbance
	NADH	Fluorescence
	NADH	Luminescence
Peroxidase	Oxidized dye	Absorbance
	Oxidized luminol	Luminescence

NADH, reduced nicotinamide adenine dinucleotide.

regions are obviously critical. The most commonly used methods employ either glutaraldehyde or a range of heterobifunctional compounds. In the former the five-carbon dialdehyde homofunctional molecule links protein molecules through lysine ε-amino and N-terminal groups. The obvious disadvantage of this approach is the opportunity for cross-linking between enzyme molecules, thereby reducing the yield of conjugate.

The heterobifunctional linker molecules offer a wider range of residues for coupling but also the choice of two reaction conditions such that only one group is activated at a time, reducing the potential for cross-linking of enzyme molecules. Compounds that have been described for enzyme protein conjugation include *N*-(4-aminobenzoyl)-*N*′-(pyridyldithiopropionyl)hydrazine (ABDP), *N*-[β-(4-diazophenyl)ethyl]maleimide (DPEM), *N*-hydroxysuccinimidyl 4-azidobenzoate (HSAB), and *m*-maleimidobenzoyl *N*-hydroxysuccinimide (MBS); a larger list is given by Gosling.[40]

In the case of hapten molecules, conjugation has also been employed through phenolic, imidazole, and carboxyl groups. In addition, a spacer molecule may be incorporated to reduce the shielding of the epitope site; however, the length of this spacer can be critical. The most commonly used conjugation schemes employ carbodiimide, which is used to form an active *N*-succinicimidyl ester from the carboxyl-containing hapten and *N*-succinimide.

4.2.3 Recombinant Molecules

The fusion of enzymes with antibody or other acceptor molecules has been described, although not for many immunoassays.[41]

4.2.4 Signal Technologies

Many of the enzymes described earlier can be measured using a range of substrates, the choice being determined by the environment in which the assay might be used, the type of assay, the concentration range of detection required, and the nature of the sample, together with the practical issues of reagent costs, reagent stability, and instrumentation available.

Thus if the detection takes place in or on the surface of a solid matrix (e.g. gel, biopsy section) then it is important to choose a substrate that will permit the generation of a high color yield and insoluble product that remains associated with the matrix after formation.[42] Colorimetric detection will depend to some extent on the instrumentation involved but the molar absorptivity of the product and the absorbing wavelength will be the major guiding criteria. The most commonly used substrates have been nitrophenol derivatives with an absorbance maximum of the product at 405 nm or redox dyes with absorbance maxima in the 500–600 nm region. The latter is more applicable for visual detection by the naked eye.

Fluorescent and luminescent products are also popular, albeit requiring more specialist equipment for detection. Methylumbelliferyl derivatives have often been employed with alkaline phosphatase and β-galactosidase labels.[43] Luminescent detection systems have been described for alkaline phosphatase, β-galactosidase and horseradish peroxidase, in the former case with the synthesis of a wide range of compounds based on adamantyl derivatives.[44] Enhanced luminescence has been very effective in improving the light yield, particularly in the case of the luminol peroxide reaction, with molecules such as *p*-indophenol and *p*-phenylphenol increasing the light yield (by over 1000-fold) in addition to prolonging the period of light emission.[45] The luminescent detection systems have proved to be extremely valuable in immunocharacterization techniques (immunoblotting) with the position of the label being detected using X-ray film.

In the situation where a relatively opaque reaction medium is encountered, e.g. when using a whole blood sample, it may be appropriate to use a substrate capable of generating an electrochemically active product. These products include phenol, *p*-aminophenol, and NADH.[46]

4.2.5 Enzyme Amplification

Whilst it is possible to obtain improved sensitivity of detection by the choice of substrate and detector technology, this can also be achieved by the use of an amplification technique. A typical example is that described by Self,[47] in which the alkaline phosphatase label is employed with nicotinamide adenine dinucleotide phosphate (NADP) as substrate; the product nicotinamide adenine dinucleotide (NAD) is then used as the coenzyme in a secondary reaction with alcohol dehydrogenase, the NADH produced being detected with a tetrazolium dye reaction, the NAD being recycled. A similar approach has been described by Obzansky et al. in which phosphorylated flavin adenine dinucleotide (FAD) is split by alkaline phosphatase, the cofactor activating the apoenzyme D-amino acid oxidase, which then acts on proline to release hydrogen peroxide.[48]

4.2.6 Relative Detection Limits

It is difficult to make accurate comparisons between the relative detection limits of the various substrates that can be used, because they have all been described in different analytical systems incorporating other variables. However, an attempt is made in Table 2 to compare

Table 2 Comparison of the detection limits for immunometric assays of thyroid stimulating hormone (TSH) employing alkaline phosphatase as the label. Detection limit expressed as two standard deviations above the zero standard signal [data extracted by permission from Bronstein et al.[44]]

Detection reaction	Detection limit (TSH mIU L^{-1})
Nitrophenol/absorbance	0.02
NADP/amplification/absorbance	0.024
Luminol/luminescence	0.06
Terbium chelate/fluorescence	0.003
AMPPD/luminescence	0.0045

AMPPD, 3-(2′-spiroadamantane)-4-methoxy-4-(3″-phosphoryloxy)-phenyl-1,2-dioxetane.

performances using the enzyme label alkaline phosphatase in an immunometric assay format.

4.3 Fluorophores

The choice of fluorophore label will depend on the quantum yield, the separation between the emission and excitation wavelength peaks (Stokes shift) and the potential quenching effects which may result from the conjugation procedure or from any endogenous substances in the sample or reaction mixture; the latter will be particularly important in the case of a homogeneous immunoassay format. Fluorophores with the highest quantum yield include phycobiliprotein, fluorescein, and rhodamine derivatives; other molecules that have been used include umbelliferone derivatives and Lucifer Yellow.[49] The fluorophores are conjugated to the antigen or antibody depending on the type of assay using conventional techniques, usually with the aid of a small linker molecule in the case of hapten molecules.

The limitations of quantum yield, background fluorescence, and quenching have been overcome by the development of time-resolved fluorescence probes. The fluorescent molecules employed in this technique have longer decay times than conventional fluorophores, such as fluorescein and endogenous fluorophores in biological samples (60–100 μs compared with 5–100 ns). Thus if such a fluorophore is excited the emission decays more slowly such that, by delaying the collection of light emitted, any endogenous interfering fluorescence can be excluded.[50] In the applications to quantitative immunoassay this property is linked to pulsed excitation using a laser light source in such a way that amplification of the signal is achieved by the use of multiple excitation/emission cycles. The longer decay times are exhibited by the lanthanide chelates such as europium (^{37}Eu), samarium (^{37}Sm) and terbium (^{37}Tb) coupled with organic ligands, such as β-naphthyltrifluoroacetone, pivaloytrifluoroacetone, and benzoyltrifluoroacetone. The most common application described has been for dissociation-enhanced lanthanide fluorescence immunoassay (DELFIA), in which the antigen or antibody is labeled with a europium chelate, the europium exhibiting only weak fluorescence in this state.[51] On completion of the immunochemical reaction the europium is released from the hydrophilic chelation with the use of a highly lipophilic environment containing a β-diketone chelator, enhancing the fluorescence of the europium ion. The use of alternative lanthanide ions enables more than one assay to be performed in a single tube.

4.4 Luminescence

Luminescence reactions are generally divided into chemi- and bioluminescent reactions, the former depending on a chemical reaction and the latter the result of an enzyme-mediated reaction. The light emitted from such reactions is collected on a photomultiplier tube. The most commonly employed molecules in chemiluminescent reactions are luminol, acridinium esters, adamantyldioxetanes, and oxalate esters. As indicated earlier, some of these luminescent species have been incorporated in enzyme substrates. The oxidation of the molecules with hydrogen peroxide leads to the creation of an excited-state decay, leading to the release of light.[45,52,53]

The most commonly used bioluminescent reactions employ either firefly luciferase which catalyzes the oxidation of luciferin requiring adenosine triphosphate (ATP), or a bacterial luciferase which catalyzes the oxidation of a long-chain aldehyde requiring NADH and an oxidoreductase. In most immunoassay applications of bioluminescence the enzyme has been employed as the label.[54]

There have been a number of calcium-dependent photoproteins described in which the luminescent emission is triggered by the presence of calcium. The most commonly used of these photoproteins are the aquourin, and recombinant engineered forms of these proteins have been developed to facilitate large-scale production.[55]

The most recent addition to the family of luminescent reactions harnessed for immunoassay has been electrochemiluminescence. In one example the ruthenium(II) tris(bipyridyl) ion $Ru(bpy)_3^{2+}$ (employed as the covalently coupled label) is oxidized at the surface of an electrode forming the strong oxidant $Ru(bpy)_3^{3+}$. At the same time tripropylamine (TPA) which is present in excess in solution is oxidized at the electrode to form the cation radical $TPA^{\bullet+}$, which spontaneously loses a proton to form the radical $TPA^{\bullet}$. The two species react to form the excited state of the ruthenium complex – the energy derived from the large difference in electrochemical potentials of the two species. The excited-state ruthenium complex decays, emitting light at 620 nm. The

ruthenium compound returns to its original state ready to be oxidized again and re-enter the cycle.[56] Alternative electrochemiluminescence systems and a wide range of applications have also been described.[57]

4.5 Substrate/Inhibitors/Enhancers

It is obvious that in a two- or multicomponent reaction it is possible to consider any one of the components as a label, given that it can be coupled to the antigen or antibody without compromising the unique binding properties of either. Indeed, one of the components may be more suited to the production of a conjugate, e.g. substrates linked to haptens. However, in addition, the complementarity of two reactants has been exploited in a range of immunoassays, in particular those using a homogeneous format.

In addition to conventional enzyme or substrate labels the use of inhibitors and cofactors as labels has also been described, e.g. the use of a phosphonate as a label which inhibits the enzyme acetylcholinesterase[58] and FAD and apoglucose oxidase.[48,59]

4.6 Particles and Metal Sols

The incorporation of particles into an immunoprecipitate and the accumulation of particles (without agglutination) have been used for the detection and/or quantitation of antigens (and antibodies). One of the first particles used as a label was the erythrocyte, taking advantage of the size and color to demonstrate the presence of antigen using labeled antibodies.[60] In more recent years latex particles have become the particle of choice.[61] The quantitation of antigens (or antibodies) has been described using turbidimetric or nephelometric detection of the immunoaggregate formed (see later description), the presence of the particle allowing the production of a multivalent antigen either to facilitate the design of the competitive assay for haptens or to increase the sensitivity of an immunoagglutination reaction. In the majority of turbidimetric and nephelometric assays, particles in the size range 40–200 nm are employed, although larger species are also used, particularly when using a near-infrared light source for nephelometric detection. The use of carbon, silver, and gold sols has been described, although methods are mainly confined to assays for urine analytes because the sols appear to suffer nonspecific protein adsorption from serum samples.[62]

The use of colored latex particles, typically blue, and metal sols, has become extremely popular for labeling of immunoreagents employed in lateral flow, liquidic circuit, and immunochromatographic disposable devices. These methods depend on the visibility of the concentrated latex particles at the site of capture of antigen by a covalently bound antibody (see later discussion).

5 RECOGNITION AND CHARACTERIZATION

Antibodies are an important part of the armamentarium of the diagnostician in the recognition of proteins and other macromolecules, including cells. Similarly, an antigen molecule is vital to the recognition of the presence of specific antibodies. The use of the antibody reaction is usually complementary to some form of separation technique.

5.1 Immunoelectrophoresis

One of the earliest techniques for characterizing a mixture of antigens of interest in a solution, such as serum or urine, was immunoelectrophoresis performed in a thin agarose film. The technique involves the electrophoresis of sample from a cylindrical well parallel to a narrow trough cut in the gel; a single gel may comprise several sample wells and troughs. In the process of electrophoresis, the proteins are separated according to their electrophoretic mobility. After electrophoresis, antiserum against the protein(s) of interest is placed in the trough. Diffusion of antigen from the original electrophoresis track and antibody from the trough will led to the formation of a precipitin arc when the antigen is present. If a general polyclonal antiserum is used (i.e. raised against a range of antigens), then a series of arcs will form, the mobility of the original sample analyte giving an idea of the identity of the antigen causing arc formation.[63]

An alternative approach, termed two-dimensional immunoelectrophoresis, has been employed to obtain a higher resolution of mixtures of proteins. In this technique the sample is again electrophoresed into an agarose gel from a cylindrical well. The gel track in which the electrophoresis has taken place is then cut out and butted against a second square agarose gel containing antibody or a mixture of antibodies. Electrophoresis is performed into this second gel and at 90° to the dimension of the first electrophoresis cycle. Immunoprecipitation occurs at the position of equivalence for each antigen (see later discussions on immunoprecipitation reactions); the height of the precipitin peak will give an indication of the amount of antigen present in the sample. When using a single antibody reagent in the second gel, the technique is particularly useful for detecting variants, fragments, or complexes of the antigen of interest.[64]

5.2 Immunofixation

This technique has replaced immunoelectrophoresis as the method for detection of individual proteins, such as paraproteins, in serum or urine from patients with myeloma. Electrophoresis is performed in an agarose

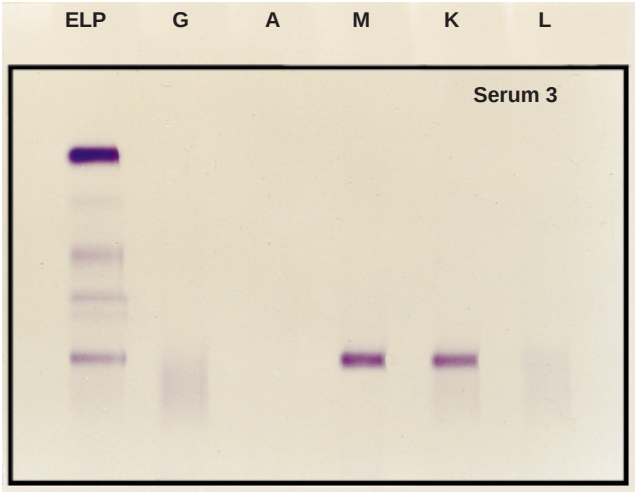

Figure 2 Results from an immunofixation of a patient's serum sample containing a paraprotein subsequently shown to be an immunoglobulin M kappa.

gel film. The film is then covered with a layer of antiserum containing antibodies against the protein of interest, usually in the form of an impregnated strip of cellulose acetate. The antibodies diffuse into the agarose gel forming an immunoprecipitate; the unreacted proteins can then be washed out of the gel and the immunoprecipitate visualized with a protein stain. An example is illustrated in Figure 2.

5.3 Western Blotting

This technique extends the use of immunoblotting to other separation matrices and does not depend upon the formation of an immunoprecipitate. It is used particularly with polyacrylamide gel matrices and permits the detection of much lower concentrations of protein. The technique depends on the transfer of the proteins at the completion of the electrophoretic step from the gel, with immobilization on to a nitrocellulose or nylon membrane. The transfer is commonly achieved by electrophoresis, trapping the gel and nylon membranes between buffer-wetted filter paper. After capture of the proteins on the membrane they can be detected using labeled antibody probes; typical labels include horseradish peroxidase with calorimetric or luminescence detection, or gold sols. The label may be attached to an antispecies second antibody as a more generic reagent which can be used with a reagent of primary detection antibodies.

This immunoblotting technique can be used to detect a single protein (or epitope) following separation according to molecular weight, size, and charge or isoelectric point. The technique can be used for epitope mapping, structural domain analysis, and ligand binding studies as well as then providing the material for antibody production or amino acid analysis and protein sequencing.[65]

The detection limit of this technique is between 100 and 1000 times lower than that using immunofixation, with a limit of detection of 10 pg quoted for colorimetric staining and 1 pg or less claimed for luminescence, gold sol, and radioisotope labeling.[66]

5.4 Immunocytochemistry

Labeled antibody probes have been used for many years in the study of the viability, morphology, and lineage of cells. The technique has been applied to biopsy material, to blood and bone marrow cells, and to microbiological organisms. The critical first step in any cytochemical application is the preservation of the cellular morphology, many of the conventional tissue-processing techniques incorporating extreme conditions that may destroy the structure of the epitopes of interest. Fixation can be achieved in paraffin or acrylic blocks whilst maintaining antigenicity, whilst cryofixation techniques have also been described.[67]

Antibody probes can either be labeled directly or with a second labeled antispecies antibody. The most commonly used labels are enzymes such as horseradish peroxidase, alkaline phosphatase, and gold sols. Fluorescent labels are not commonly used because of the autofluorescence of the sample. The enzyme labels produce a permanent record and for electron microscopy can produce the density of product required; gold sols are particularly useful for electron microscopic examination.[68]

One of the critical issues both for the preparation of the sample and for the choice of reagent is the penetration of the antibody probe. Excessive labeling can produce an extremely large immunoreagent complex that will limit the sensitivity of detection. The use of an antibody fragment with a sensitive signal reaction may help to increase penetration and lower detection limits.

5.5 Flow Cytometry

The technique combines the light-scattering properties to determine the size and granularity of cells together with the attachment of fluorescent-labeled antibody probes to give further differentiation to the cell types, as well as the detection of specific antigens. It is also possible to assess the membrane potential, pH, enzyme activity, and DNA content of the cells.[69]

The basic design of the flow cytometer involves a means of projecting a stream of cells through a flow cell that is illuminated by a laser light source. The cells enter the flow cell under pressure through a small aperture which is surrounded by a sheath fluid. The sheath fluid creates a hydrodynamic focusing effect and draws the sample into

a stream; careful optimization of sample and sheath fluids ensures that the cells pass singly through the detection zone.

Light scattered in the forward angle (which is proportional to the radius of a sphere) provides information on the size of the cell, whilst the light scattered at 90° (proportional to the refractive index of the cell) provides information on the granularity of the cell.

Fluorescence emitted from suitable probes covalently coupled to specific antibodies that have reacted with their complementary antigen is detected at 90°. Selection of light scattering and fluorescence is usually achieved by the use of appropriate filters.

The third component of the flow cytometer is the computer which collects the light scattering and fluorescence data and manipulates them according to established algorithms, providing a wide range of parameters with respect to type and number of cells present in a given sample population.

The advent of monoclonal antibodies led to the creation of a library of antibodies directed against cellular antigens. These antibodies have been grouped into clusters based on their cellular reactivity, and termed a cluster of differentiation (CD). Considerable research effort has been focused on the clinical application of the study of cellular antigens.[70]

Whilst the initial application of flow cytometry was in the field of differentiation of white blood cells, the unique capabilities of the instrumentation to differentiate a mixed population of fluorescently labeled particles have led to the description of its use for multianalyte immunoassays using fluorescent-dye-labeled latex particles. In one configuration, differential dyeing of identically sized particles with two different dyes, with emissions at two different wavelengths, allows the recognition of aggregates and permits the discrimination of at least 64 different sets of particles. In fact, three fluorescent dyes are used, orange and red for classification of the particles and green for analyte measurement.[71]

5.5.1 Cell Sorting

Cell sorting is an important adjunct to the diagnostic capabilities of flow cytometry. It is achieved by employing one of two different techniques. The first utilizes the change in direction of flow of individual cells by electromechanical means. As the technique does not require passing cells through a narrow orifice, fragile and large cells and aggregates can be separated successfully. The second technique involves the packing of cells, after analysis, into tiny droplets, charging the droplets that contain cells and then deflecting these droplets in an electrical field. The former technique can sort about 10^3 particles per second whilst the latter can accommodate about 10^4 particles per second.

5.6 Immunosensors

The requirement to label one component of the antigen–antibody reaction, and in many cases to "develop" or visualize that label, had introduced complexity in the use of the reaction and made interrogation of the reaction kinetics both difficult and in many instances unreliable. The advent of the immunosensor therefore allowed real-time or kinetic analysis of the antigen–antibody reaction.[13,14]

An immunosensor is a form of biosensor in which the biorecognition molecule is an antibody (or antigen in the case of the detection of an antibody) which is directly coupled to a transducer that detects the binding event and converts it into an electronic signal. A wide range of examples have been described including those detecting the binding directly whilst others employ detection of a label. The direct detection of binding can potentially be achieved by the change in electrochemical characteristics (potential, current, conductance), heat released (calorimetry), mass, or refractive index. It is beyond the scope of this article to describe all of the principles[72] and further discussion will therefore be limited to those that have been commercialized and are in routine use.

The greatest success in immunosensors has been achieved with optical sensors which employ the phenomenon of the evanescent wave. There are several different approaches to detection including surface plasmon resonance, employing a thin gold film on the sensor surface,[73] a resonant mirror approach with high and low refractive films on the surface of a prism,[74] and waveguiding material in combination with a diffraction grating.[75] The surfaces are coated with a dextran layer or some other means of limiting nonspecific binding whilst also providing a means of coupling the antibody (or antigen) to the surface.

The sensor techniques measure the proximity of higher-molecular-weight species close to the surface, and consequently reflect the specific binding interaction (assuming that nonspecific binding to the surface has been negated). A signal will be produced due to the presence of the species of interest, and other high-molecular-weight species in the medium (the bulk effect) which will contribute a signal and must be discounted; the bulk effect is seen as a rapid increase in signal which remains constant (Figure 3). The true binding is obtained after subtraction of the bulk effect signal, and is proportional to the mass of bound analyte. The rate of binding will also be determined by the speed with which the species of interest approaches the surface of the sensor; rapid binding will deplete the local concentration and thus influence the progress of the binding curve. This can be overcome by stirring the mixture or flowing

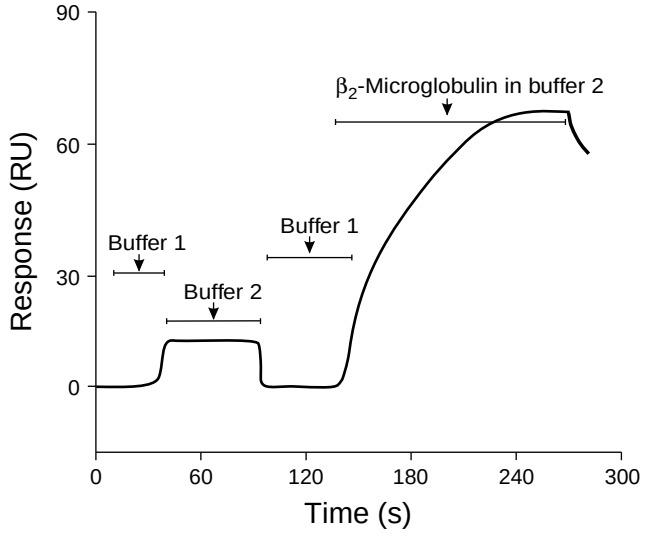

Figure 3 Response observed for a surface plasmon resonance optical sensor to demonstrate the bulk effect response on addition of buffer 2 to a surface on which had been covalently coupled a monoclonal antibody against β_2-microglobulin followed by addition of antigen in buffer 2. The response is expressed as relative to the baseline established with buffer 1.

the sample over the sensor surface; failure to do this will introduce a mass-transfer effect on the binding characteristics.

The real-time monitoring of the association and, by removal of the sample and replacement with buffer, dissociation will enable the characteristics of the antigen–antibody reaction to be determined and thus the binding characteristics of antibodies to be compared. The technique can also be used for epitope mapping and for the quantitation of both low- and high-molecular-weight analytes; in the case of a low-molecular-weight analyte a competitive technique has to be used with a high-molecular-weight multivalent analyte complex being created (e.g. coupling hapten molecules to a latex particle or gold particle core).[76]

6 QUANTITATION

It is clear that several of the techniques already described can be used for semi- and full quantitation of the species of interest. In some situations a qualitative answer is all that is required but it should be noted that in many cases a strict cutoff between negative and positive results is required and therefore the rigor of analytical performance required is the same as for a quantitative assay.

There are many different analytical approaches to quantitative immunoassays that would be impossible to encompass in a short review. The following discussion will therefore be limited to the quantitation of antigen; however, the reader interested in the detection of antibodies will readily be able to reconfigure the sequence and choice of reagents for his or her needs. In certain examples the antigen may have a unique property which either negates the need for additional immunoreagents and makes use of other properties, such as enzyme activity, as in the case of an immunocapture assay for bone alkaline phosphatase,[77] or binding to an alternative affinity ligand (e.g. a lectin).

6.1 Heterogeneous Immunoassays

The basic requirement for any quantitative immunoassay is the differentiation between bound and unbound (unreacted) species. The two basic assay designs embodied in the heterogeneous format are those of competitive and noncompetitive immunoassay.

6.1.1 Competitive Immunoassay

In the predominant example of this format, labeled antigen competes with the sample antigen for a limited number of antibody binding sites. The antibody molecules are immobilized on a solid phase which permits the separation of bound and free labeled antigen; the amount of bound labeled antigen is then measured according to the properties of the label and the amount of sample antigen is quantified by comparison of the signal with that obtained from a calibration curve. A schematic of the assay principle and the relationship between bound label and sample analyte concentration are depicted in Figure 4(a) and (b). The addition of sample and labeled analyte can be performed simultaneously or sequentially, the latter offering a 2–4-fold improvement in the detection limit. An alternative approach for quantitation of a hapten would be to couple a limited amount of

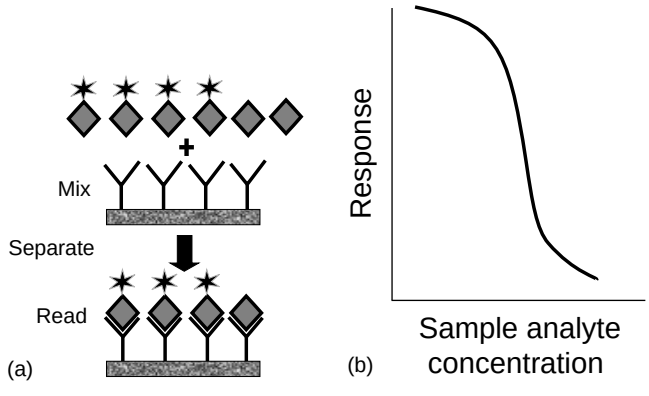

Figure 4 (a) Schematic diagram of a heterogeneous competitive immunoassay employing an immobilized antibody (Y) with labeled antigen (✶◆) and sample antigen (◆) competing for binding sites. (b) A typical response curve.

antigen to the solid phase and then use a labeled antibody with competition of sample analyte with the immobilized antigen. The bound antibody could alternatively be quantitated with a labeled second antispecies antibody.

The optimization of reactant conditions will depend on the concentration range of analyte in the sample and the equilibrium (affinity) constant of the antibody. Assays requiring higher levels of sensitivity require antibodies with higher affinity constants. Several optimization strategies have been developed with assessment being based on the precision profile (coefficient of variation of duplicates plotted against analyte concentration) which gives a practical demonstration of the analytical range of the assay.[78] In broad terms, greater sensitivity is achieved at a lower concentration of antibody whilst better precision is achieved at a higher concentration.

The competitive form of assay is applicable to an analyte where the concentration range is either relatively narrow (i.e. where good discrimination is required over a narrow concentration range) and/or where the sample concentration range may be very large and where the hook effect in the immunometric assay could be a problem (see the next section); this is exemplified in the case of an assay for albumin in urine used for the early detection of diabetic nephropathy but where a sample with a very high level may also be encountered, as in the case of nephrotic syndrome.[79] The competitive format employs a limited amount of antibody, especially when seeking higher sensitivity, and therefore the addition of an accurately defined amount of antibody reagent is critical to the reproducible performance of the assay. Finally, the competitive assay format may be the only option for the quantitation of small molecules (haptens) which effectively possess only a single epitope.

6.1.2 Immunometric Assay

In the case of larger molecules which possess more than one discrete epitope region, it is possible to employ an immunometric or sandwich format. An antibody coupled to a solid phase is employed to capture the analyte of interest, the captured antigen then being detected with a second labeled antibody. Whilst this technique was first described in the late 1960s,[80] it has become more popular with the advent of monoclonal antibodies capable of providing binding pairs that recognize different epitope regions of the antigen.

The addition of sample analyte and labeled second antibody can be undertaken simultaneously or sequentially. In the former case there is a risk, at very high concentrations of antigen, of the analyte saturating both capture and second antibodies, producing the so-called ''hook effect'', after which point the assay response falls at higher analyte concentrations, effectively as a consequence of the failure to form ''sandwich complexes'' and

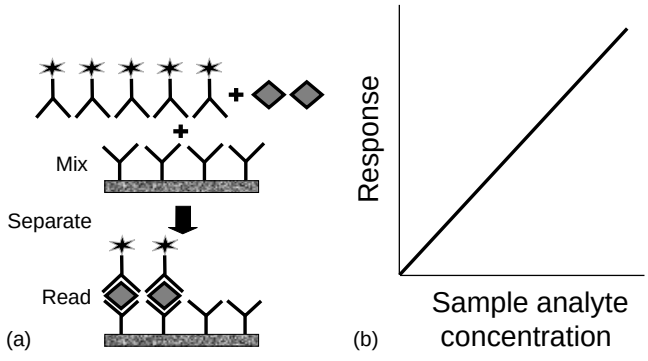

Figure 5 (a) Schematic diagram of a heterogeneous immunometric assay employing an immobilized first antibody (Y), sample analyte (◆), and labeled second antibody (Y*). (b) A typical response curve.

thereby giving a falsely low result. This effect can be overcome by using a sequential approach and by ensuring there are sufficient amounts of capture and second (detector) antibody available; the alternative is to employ a competitive format. The format of the immunometric assay and a typical calibration curve are shown in Figure 5(a) and (b). The sensitivity of the immunometric format can be enhanced by the choice of the label: the amount of signal increasing in proportion to the amount of analyte present. However, a limitation on sensitivity, as in the case of all immunoassays, is the level of nonspecific binding, and it is important to minimize this effect as much as possible.

Whilst the competitive format is required for most hapten assays, the option of the immunometric approach has been demonstrated with the use of anti-idiotypic antibodies.[81,82]

6.1.3 Separation Methods

The quality of the separation of bound from free label is vital to the performance of an immunoassay, being a major influence on the nonspecific binding component of the response. The solid phase must therefore be capable of rapid separation and effective washing to remove nonimmunochemically bound labeled reactant. It goes without saying that the antibody must have a sufficiently high affinity constant for the antigen to ensure that the immunocomplex is not disrupted during the washing stage. The solid phase should also ideally have a large surface area to ensure minimization of diffusion distances, whilst in the case of the immunometric assay enabling a high concentration of capture antibody to be achieved. However, it is also important to ensure that the possible adsorption of labeled reactants is minimized by blocking of any ''bare'' surface of the solid phase.

Early separation methods were based on the use of protein precipitation techniques, e.g. poly(ethylene glycol)

or ammonium sulfate, the immunochemical reactants all being in "free" solution. Alternative separation techniques have also been described, including electrophoresis, gel filtration and ion-exchange chromatography.

However, today the majority of separation methods are based on the use of antibodies (or antigens) either adsorbed or covalently coupled to a solid phase. The choice of solid-phase matrix may depend on the equipment available in the laboratory, but comprises the inside surface of a tube or microtiter plate well, plastic beads, paddles, porous disks or magnetizable particles. Reaction tubes offer the lowest surface area available, followed by microtiter plate wells and large plastic beads, whilst small beads (e.g. latex particles) and magnetizable particles offer the largest surface areas;[83] the latter can also offer the shortest diffusion distances. The porous disk approach, with disks made from several materials including chromatography paper, has been extremely popular in disposable immunoassay devices (see later discussion).

In some cases another binding protein has been coupled to the solid phase; this can provide a universal solid-phase reagent which can then be employed with a range of immunoassays. This reagent may be an antispecies antibody or protein A.[84] An alternative which also provides opportunities for signal amplification is to use one of the biotin–avidin or biotin–streptavidin pair, the other being coupled to the reagent which is intended to be the primary reagent attached to the solid phase.[85]

6.2 Homogeneous Immunoassays

As the name implies, the homogeneous immunoassay does not require a separation step, the delineation between the free and bound fraction being achieved by modulation of a property of one of the reactants by the antigen–antibody binding reaction. Thus in the context of the earlier discussion the immunosensor systems described could be considered as homogeneous immunoassays. However, there have been a range of free-solution homogeneous immunoassays described, with many being commercialized very successfully.

6.2.1 Light Scattering Immunoassay

One of the first examples was the use of the immunoprecipitation reaction for the quantitation of specific proteins. The principle, described in Figure 6(a–c), is based on antibody molecules binding first to one protein antigen molecule to form a dimer and then to a second antigen, thereby forming a complex which increases in size, and ultimately forming an immunoprecipitate. As the complex forms there is an increase in the light scattering which can be measured either turbidimetrically or nephelometrically (utilizing right- or forward-angle scatter, the latter giving better sensitivity for larger complexes). The majority of methods described employ a polyclonal antiserum with the opportunity of more epitopes available for binding, although mixtures of monoclonal antibodies have been used. The rate of immunoprecipitate formation is usually enhanced by the inclusion of a polymer, typically poly(ethylene glycol) 6000.[86]

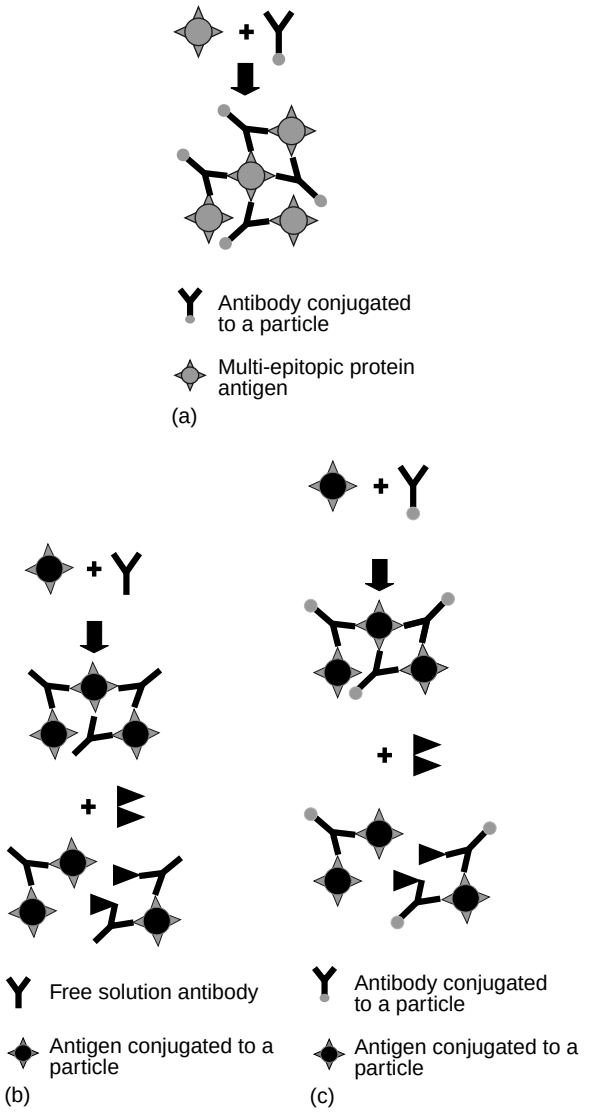

Figure 6 Various configurations for light scattering assays in which direct agglutination (a) can be employed with (Y) or without (Y) the use of a latex particle-labeled antibody; (b) inhibition of agglutination by a small sample analyte; (c) an inhibition assay with both reactants labeled with particles.

The applicability of this approach has been greatly enhanced by the use of latex particles either to enhance the amount of light scattering or to act as a core for the production of a multivalent hapten molecule which extends the use of the light scattering assays to small

molecules in addition to avoiding the ''hook effect'' by analogy with the competitive assay format described earlier. The particle size employed has ranged from 40 to over 500 nm; the smaller particles, however, provide more surface area and will remain in suspension.[87] The choice of particle may be determined by the coupling chemistry for attachment of the antibody (or antigen in the case of a competitive assay); many examples of latex-labeled assays have, however, depended on adsorption without any obvious deleterious effects.

An alternative approach to turbidimetry and nephelometry when using a latex particle-labeled assay employs particle counting, which has been shown to achieve lower detectable levels and also permits the quantitation of polypeptide hormones and other proteins not measured by the conventional light scattering techniques.[88]

6.2.2 *Enzyme-labeled Immunoassay*

The enzyme multiplied immunoassay technique (EMIT®) is a competitive technique in which the antigen, typically hapten (although an application for proteins has been described), is coupled to an enzyme in such a way that binding to the antibody results in inhibition or activation of the activity.[89] The main enzymes used have been lysozyme, malate dehydrogenase, and glucose-6-phosphate dehydrogenase, with the latter being the most commonly described in a range of commercial assays for drugs. A schematic diagram illustrating the principle of the method is shown in Figure 7(a) and (b).

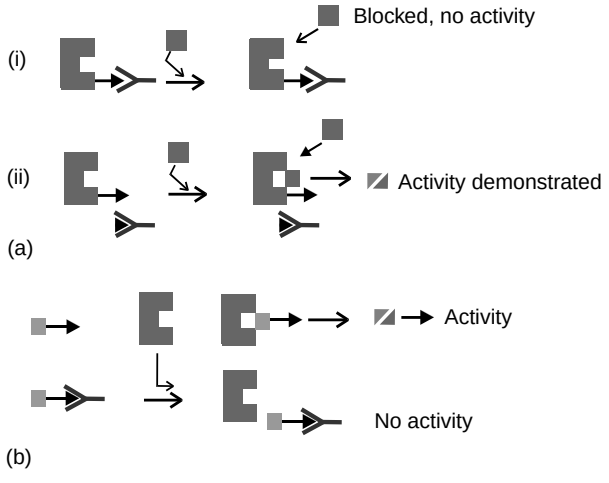

Figure 7 Schematic diagram of a homogeneous enzyme immunoassay which in (a) shows that the active site of the enzyme linked to hapten (→) is blocked by the addition of antibody (≻). In the presence of sample analyte (ii) there is competition for antibody and the antibody binds to hapten enzyme conjugate such that activity is present leading to breakdown of substrate (→). In the case of a substrate-(or inhibitor)-labeled assay (b) the enzyme is employed as the reagent with access of the labeled hapten limited upon binding of antibody.

Subsequently both substrate and cofactor labeled techniques were described, the basic principle being similar in that coupling of substrate (or cofactor) to enzyme was inhibited by the proximity of antibody binding to the labeled antigen. In one example FAD was employed as the label, with the apoenzyme of glucose oxidase employed as a reagent.[59]

Cloned enzyme donor immunoassay (CEDIA®) employs recombinant DNA technology to produce new strains of *E. coli* synthesizing inactive forms of the tetrameric enzyme β-galactosidase. These fragments comprise a protein of approximate molecular weight 113 000 Da, called the acceptor protein, and a small polypeptide (of approximately 90 amino acid residues), called the donor polypeptide. Association of the acceptor and donor fragments has been shown to form monomer which can then aggregate to form active β-galactosidase. Hapten coupled to the donor polypeptide does not impair the formation of monomer and thus generation of enzyme activity; however, antibody bound to the hapten inhibits monomer formation. This competition of sample analyte (or hapten) for antibody binding sites results in more donor polypeptide-labeled hapten being available to form monomer with acceptor protein and thence enzyme activity.[90]

6.2.3 *Fluorophore-labeled Immunoassays*

The most commonly used homogeneous fluorescence immunoassay employs a polarized excitation light source. If a polarized light source is used to excite the fluorophore molecules, the energy is absorbed more efficiently by molecules whose oscillating dipoles are parallel to the plane of the polarized light. Furthermore, the degree of polarization of the emitted light is dependent on the lifetime of the excited state and the rotational motion of the molecule. All molecules in solution are in a constant state of motion (Brownian motion) which is dependent on their size and shape as well as solvent viscosity and temperature.

Thus a fluorophore coupled to a hapten molecule will exhibit greater Brownian motion when free in solution than when the hapten fluorophore complex is coupled to its complementary antibody. The rotational time of the immuno complex will be about 10–100 ns compared with 1 ns in the case of the fluorophore–hapten molecule. Experience has shown that the antigen needs to be less than 20 000 Da in order to detect the change in polarization; furthermore, the fluorescence lifetime of the fluorophore must be greater than the rotational time of the labeled antigen (hapten) and fluorescein, with a lifetime of 4.5 ns, has been one of the most

successful labels used. The degree of polarization of the fluorophore hapten which is inversely proportional to the concentration of the sample analyte is measured in the most commonly used application of this technique.[91] The main limitation of the method is the effect of serum proteins, and a high dilution of sample is required, which limits the sensitivity of the assays.[92]

The energy emitted by a fluorophore may be influenced by many factors, including the close proximity of energy-donating or energy-accepting groups. Examples have been described of direct quenching of the fluorophore-labeled antigen upon binding of the antibody.[93] An alternative approach, termed fluorescence excitation transfer immunoassay (FETI), involves the labeling of antigen with fluorophore and the labeling of antibody with an energy-accepting (or -quenching) molecule. When labeled antigen binds to antibody the fluorescence emission energy of the antigen label excites the antibody label, itself therefore being quenched. The transfer of energy requires the wavelength of emission of the first molecule to match that of the excitation of the second and the close proximity of the two molecules (approximately 10 nm or less) because the rate of transfer is inversely proportional to the sixth power of the distance between the molecules.[94] The approach was first described using fluorescein as the donor and rhodamine as the acceptor molecule.

The method effectively extends the Stokes shift, although in most applications the quenching of the donor molecule is usually measured rather than the excitation of the donor.[95] A number of alternative fluorophore pairs have been described and methods have been reported for both hapten and protein antigens. A schematic diagram illustrating the principle of the assay is shown in Figure 8.

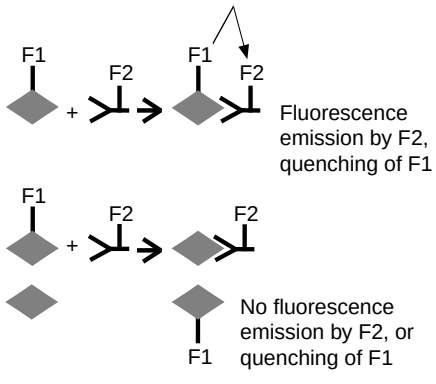

Figure 8 Diagrammatic representation of a FETI in which the proximity of fluorophore-labeled antigen (^{F1}) to fluorophore-labeled antibody (^{F2}) leads to excitation of F2 by emission of F1; the presence of sample analyte (◆) blocks excitation transfer.

A number of energy-enhanced and energy-transfer luminescence immunoassays have also been described, although none has been commercialized to date; all are based on the interaction between two molecules in close proximity.[96,97] In one configuration recently described, a luminescent oxygen channeling assay (LOCI™),[97] the assay depends on the proximity of two particles. A photosensitive compound is dissolved in one particle which upon exposure to light produces singlet oxygen. The second particle species contains an olefin which reacts with the singlet oxygen to produce a chemiluminescent emission. This reaction only takes place when the particles are brought together as a pair. Haptens are measured using competition of sample with hapten and antibody-labeled particles – photosensitive molecule and olefin containing, respectively. In the case of macromolecular antigens one antibody is coupled to the olefin-containing particles; the photosensitive particle is coupled to streptavidin, which is subsequently bound to the initial antibody complex with the aid of a biotin-labeled second antibody. This homogeneous format was demonstrated for thyrotropin and other macromolecules.[97]

7 MICROFABRICATION AND ENCAPSULATION

An important development in the field of immunoassay technology has been the encapsulation of, initially some and now all, of the reaction components of the quantitative immunoassay into a device that enables the technology to be used by operators with little or no technical training. This development has been achieved through the successful combination of several technologies, including liquid handling, printing, porous matrices, and plastic moulding. The innovation today is now moving toward miniaturization of devices, hence microfabrication technologies with the additional benefits associated with miniaturized devices, e.g. rapidity of response, portability of device, and economy of manufacture, are being exploited.

7.1 Disposable Reaction Cells

These devices typically comprise plastic cuvettes with several compartments that contain the reagents necessary for the reaction, channels to facilitate mixing, an entry port for the sample, and a compartment, where the reactants are mixed, that forms the optical cell. An example is the Bayer DCA 2000® (Bayer, Tarrytown, NY, USA) for the measurement of HbA_{1C} in which the reagents for the immunoassay comprise latex particles to which are coupled peptide molecules that have been used as the

immunogen to produce a monoclonal antibody that recognizes the HbA_{1C}. Mixing of these two reagents leads to the formation of a light-scattering aggregate. The addition of sample HbA_{1C} will inhibit the amount of aggregate formation (a competitive light-scattering assay). The cartridge also contains reagents to hemolyze the erythrocytes, necessary for the immunoassay, and for quantitation of the total hemoglobin to enable the percentage of glycated hemoglobin (HbA_{1C}) to be calculated.[98]

In another configuration, the same manufacturer has adapted the same cassette format for the quantitation of albumin in urine by a direct agglutination reaction. In this instance the cassette also contains reagents for the quantitation of creatinine so as to allow calculation of the albumin : creatinine ratio. An alternative configuration has been produced by another manufacturer for the measurement of drugs, such as theophylline. In this case the measurement of hemoglobin is used to correct for the hematocrit of the sample.[99]

7.2 Flow-through Cells

A very simple device that employs a porous matrix to which an antibody is coupled to provide a capture solid phase has been described for several analytes and for which a number of variants are available.[100] The assay format is usually employed for high-molecular-weight analytes. Addition of sample to the porous matrix pad leads to capture of the antigen; addition of a second labeled antibody results in the formation of a sandwich. If the antibody is labeled with colored latex or gold sol particles the presence of antigen is immediately evident. Alternatively, if the label requires visualization, e.g. an enzyme requiring substrate, then another reagent is added. The porous matrix is usually attached to an absorbent reservoir to collect excess liquid, and washing steps are included to improve discrimination at low analyte concentrations.

This type of device is a sophisticated form of the dot blot that is used in many research laboratories to detect the presence of antigen or antibody species and may be considered as a variation of the immunofixation principle described earlier. A diagrammatic representation is shown in Figure 9.

7.3 Lateral Flow Devices

An alternative approach to the reactants (and solvent) flowing through a porous matrix is to employ the porous matrix as the total reaction cell by flowing reactants along the device, allowing the separation of free and bound labeled antigen or antibody within the confines of the porous matrix. These devices have been given a variety of names, including lateral flow devices – often also called stick tests because of their apparent simplicity.

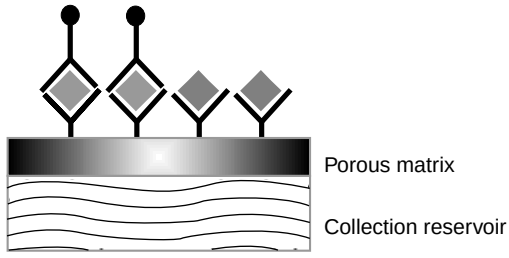

Figure 9 Schematic diagram of a flow-through device with analyte of interest (◆) captured by an immobilized antibody (Y) and detected with a second antibody (Y) labeled with either colored particles or enzymes.

In its simplest configuration, the lateral flow device comprises a porous matrix, often a type of chromatography paper which at one end, the sample addition zone, an amount of labeled antibody reagent (in the case of an immunometric assay for a large antigen) is deposited (dried down). Further along the matrix a defined amount of capture antibody is covalently coupled. At the farther end from the sample addition zone, an antispecies antibody is covalently coupled to the matrix. The sample when added dissolves the first labeled antibody reagent which binds to any antigen present and migrates along the strip of matrix until captured by the immobilized antibody. The excess of labeled antibody migrates further until captured by the antispecies antibody providing proof of migration (an internal quality control check). The first antibody is usually labeled with color latex particles or gold sol particles to visualize the sandwich, indicating the presence of antigen.[101]

It is also possible to configure the assay with labeled antigen in order to detect a small molecule. The principle of both assays is depicted in Figure 10. The ratio of the antigen and antibody reagents in the competitive style of assay is critical to the cutoff value that determines a negative from a positive result in most of the configurations of this type of assay.[102] It is possible for the flow-through and lateral flow devices to achieve quantitation with the aid of a reflectance meter. The type of assay can also be configured to detect more than one

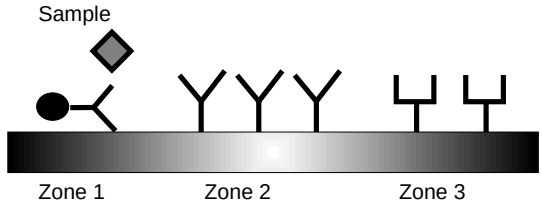

Figure 10 Schematic diagram of a lateral flow device in which a sample (◆) added to zone 1 dissolves the detector antibody (•—<), the complex migrating until captured in zone 2. Excess labeled antibody is captured in zone 3.

analyte in the same sample by immobilizing more than one antibody specificity, in a series of discrete bands, together with the appropriate mixture of reactants at the sample application area.[103] In addition, more complex configurations can be achieved by the design of liquidic circuits which can be used to delay the mixing sequence of reactants.[104]

The flow-through approach can also be used to produce an analog signal such that quantitation is achieved without the need for a reader (i.e. similar to a thermometer output), termed an immunochromatographic device. In one configuration the analyte binds to labeled antibody at the application zone and this complex migrates along a porous matrix to which capture antibody is immobilized. The distance of migration is proportional to the amount of antigen, i.e. until all of the antigen has been captured.[105]

A critical feature of the flow-through and lateral flow devices is the choice of the porous matrix. Much of this intellectual property is either closely guarded in the patent literature or not disclosed at all, with few papers describing the development of this technology. However, it is clear that the surface charge on the matrix, as is the case for all solid surfaces, e.g. latex particles, plays an important role, as do the surface tension properties and the ionic species of the solvent. The porous matrix is particularly critical when the whole blood provides the solvent for solution of reactants and the mobile phase.

7.4 Miniaturization of Devices

The miniaturization of immunochemical reactions has a number of benefits, including reduction in diffusion distances, thereby increasing the rates of reaction, and reduction in the reagent requirements. The technology takes advantage of modern microfabrication techniques that enable large numbers of devices to be produced with high levels of accuracy of manufacture. Such microfabrication has permitted the machining of narrow channels in materials such as silicon chips, glass, and chromes and the construction of valves and pumps.[106] There are several examples in the literature where small numbers of micromachined devices have been produced for encapsulating an element of the antigen–antibody reactions for quantitation of an antigen. Thus capillary electrophoresis has been employed to provide the separation in a cortisol assay with a 30-s reaction time.[107] A latex agglutination assay with fluorescent-labeled beads has been described for a protein in a reaction cell etched in Pyrex glass. An alternating current was employed to enhance the rate of agglutination, the end-point being monitored with a fluorescence microscope and image analyzer; the reaction volume employed was 0.4 μL.[108]

8 APPLICATIONS

Immunochemical techniques now dominate the analytical methods in the biological sciences and in particular in the application to diagnostic techniques in medicine. It would be impossible to document all of the applications, qualitative or quantitative, that have been described, and therefore the discussion will be limited to providing an overview of the limits of detection of the methods described, some comments on the accuracy and precision of methods, and then some specific examples of method performance. The range of analytes that can be measured by immunoassay and the applicability of the techniques described in this article are illustrated in Figure 11.

8.1 Calibration and Quality Assurance

It will be clear from the earlier discussion that the immunogen plays an important role in determining the specificity of the antibodies produced. Differences between the immunogen and the analyte of interest either inherent or produced during purification (and conjugation in terms of a hapten) can result in unsatisfactory antibody reagents; similarly, differences in the nature of the calibration material antigen can compromise the accuracy of an assay. However, probably the greatest influence on the accuracy of an immunoassay is the incorrect assignment of the analyte value to the calibrator – in the case of a primary calibrator due to use of impure analyte and incorrect value assignment by the reference method. Errors may also arise due to instability of the

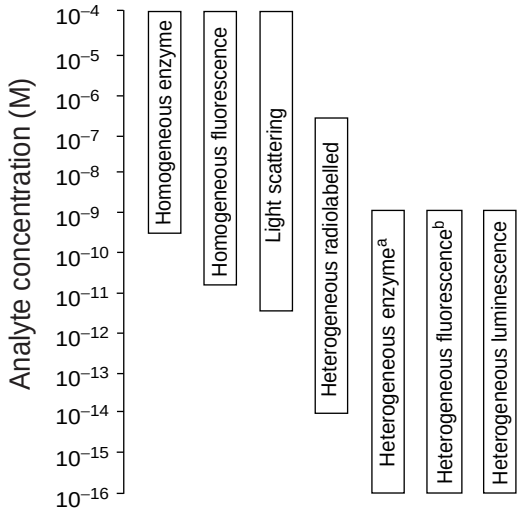

Figure 11 Diagram to indicate the concentration ranges over which the various quantitative immunoassay technologies can be used; it provides only a very rough guide. [a]Heterogeneous enzyme immunoassay includes amplified systems. [b]Heterogeneous fluorescence immunoassay includes time-resolved fluorescence systems.

analyte, particularly in the case of a macromolecule in a dilute solution, or differences in the effect of the matrix between calibrator and sample material.

In the case of large molecules there is also the possibility of considerable antigen heterogeneity. Thus many hormones may be present as prohormones and fragments in addition to the hormone molecule itself; examples include the existence of the prohormone of adrenocorticotropic hormone (ACTH), fragments of parathyroid hormone, and macromolecular forms of prolactin and gastrin.[109] In other cases the antigen of interest may exist in both free and complexed forms, as in the case of prostate-specific antigen (PSA),[110] and oxidized and reduced, phosphorylated, and dephosphorylated forms, as in the case of troponin I.[109,111] Genetic variants will also exist in many instances and this may influence attempts to quantitate all molecular forms of the antigen. Glycation may also influence the accuracy of an immunoassay, sometimes producing subtle differences in specificity.[112]

Quality assurance (QA) is designed to ensure that the correct result is produced on the correct specimen at the correct time; the major concern in terms of the analytical phase focuses on producing the correct result and embraces both internal quality control and external quality assessment. In setting up protocols to monitor or control performance, the closer the analyte species and matrix parallel those of the sample the more reliable is the scheme; particular attention in the production of QA materials must be paid to the effects of lyophilization or stabilization techniques on the analyte of interest, together with the effects of freezing and thawing.

External quality assessment schemes allow comparisons of performances to be made between methods with respect to both accuracy and precision. Such schemes have been valuable in improving the performance of quantitative immunoassays, particularly for larger biological molecules.[113,114]

8.2 Automation

A quantitative assay will benefit from automation of the various steps, especially when there are multiple reagents, additional mixing, and separation and washing steps involved. The homogeneous assay format has been configured for routine automated analytical instruments to achieve highly reproducible performance.

Whilst the heterogeneous format provided a greater challenge, in particular the separation and washing steps, full automation has now been achieved in many examples that have been commercialized.[115] In terms of the research laboratory, automation may be limited to the use of a multihead pipet, a microtiter plate, and an associated washing device. This has proved successful although care must be taken with (1) reproducible adsorption of reagent species to each reaction well and (2) avoidance of temperature gradients across the plate. An alternative semiautomated approach suitable for the research laboratory is to use a pipeting stage with a magnetizable solid phase, it being the most readily maintained in a homogeneous mixture for pipeting purposes.

8.3 Hapten Assays

A wide range of immunoassays for haptens have been described, including therapeutic drugs, drugs of abuse, and steroid hormones. Many of the assays for drugs have employed homogeneous techniques, their concentration ranges lying within the range 10^{-6}–10^{-9} M. The lower limit of detection is best exemplified by the assay for digoxin which approaches the limits of the majority of homogeneous formats.[116] Furthermore, the homogeneous formats, in many instances the subject of patented technology, have not been described for the more esoteric assays with limited commercial requirements.

Heterogeneous immunoassay formats have been described for a range of analytes but are generally confined in common usage to analytes of a lower concentration range. The heterogeneous approach is more readily set up where a limited usage is envisaged and consequently most of the more esoteric assays have only been described in this format.

Some examples of the performance that can be expected for haptens are illustrated in Table 3. A number of disposable devices for the detection of haptens have also been developed; some examples are given in Table 4.

Table 3 Some examples of immunoassays for haptens

Analyte	Technique	Between-batch precision (CV, %)	Analytical range	Ref.
Theophylline	EMIT®	4.2–5.2	0–40 mg L^{-1}	117
Cortisol	FETI[a]	3.9–6.8	400–1400 nmol L^{-1}	118
Digoxin	Immunometric (enzyme label)	1.0–8.0	0–5 μg L^{-1}	119
Digoxin	CEDIA®	3.5–12.0	0–4 μg L^{-1}	120
Cyclosporin	Heterogeneous chemiluminescence	2.0–11.4	25–1600 μg L^{-1}	121

[a] FETI, fluorescence excitation transfer immunoassay.

Table 4 Some examples of disposable devices for detection of haptens

Analyte	Technology	Sample	Ref.
Drugs of abuse	Lateral flow	Urine	102
Estrone 3-glucuronide	Lateral flow	Urine	122

8.4 Macromolecule Assays

The main homogeneous format applied to the quantitation of large molecules is that of the light scattering technique, it being the most popular approach for the quantitation of individual proteins. Sensitivity of detection has been enhanced by the use of latex particle-labeled antibodies. The method has been demonstrated over a sample concentration range of $10\,\mu g\,L^{-1}$–$50\,g\,L^{-1}$, which equates to a molar range of approximately 0.56 nm–0.75 mM.[61]

The immunometric format has been described, using a wide range of labels referred to earlier and capable of detection down to 10^{-19} M. Selection of antibodies is critical to the accuracy of the assays, as has already been mentioned. Some examples of the assays developed for macromolecules are described in Table 5. Disposable devices for a range of macromolecules have been developed; some examples are given in Table 6.

8.5 Limitations and Interferences

Reference has already been made to some of the limitations of the techniques described, particularly in the case of the quantitative assays, with reference to the design of immunogen, selection of antibody reagent, and the choice of solid phase.

8.5.1 Reagents

Dilute solutions of proteins are invariably unstable and therefore storage of reagents is critical; freezing and thawing may also harm immunoreagents. Similarly, calibrators may not be present in an identical matrix to the unknown samples and thus stability characteristics may differ. Batches of solid phase may differ and blocking conditions to minimize nonspecific binding may require reoptimization. Finally, different batches of reagents may change in relation to their potency, and careful optimization is important, as is validation with respect to accuracy and calibration range, prior to routine use.

8.5.2 Sample

In addition to the question of analyte stability, the most important concern with any sample is the potential presence of interferents. This is particularly important in the case of homogeneous assays; examples include endogenous enzymes that mimic the enzyme label

Table 5 Some examples of immunoassays for large molecules

Analyte	Technique	Between-batch precision (%)	Analytical range	Ref.
Thyrotropin	Enzyme-amplified microtiter plate, manual	5.1–9.8	0.6–$15.0\,mIU\,L^{-1}$	123
Creatine kinase MB	Enzyme label, bead, manual	6.3–11.0	0–$120\,\mu g\,L^{-1}$	124
Albumin (urine)	Latex particle immunoinhibition	1.4–8.8	0–$250\,mg\,L^{-1}$	79
Bone alkaline phosphatase	Immunocapture microtiter plate	4.3–6.7	0–$210\,IU\,L^{-1}$ at 37 °C	112
Bone alkaline phosphatase	Immunometric ^{125}I label, bead	7.1–8.3	0–$120\,\mu g\,L^{-1}$	112
Cystatin C	Latex particle nephelometry	3.2–4.4	0.23–$7.25\,mg\,L^{-1}$	125
Thyrotropin	Enzyme, chemiluminescent, manual	2.5–4.2	0.005–$60\,mIU\,L^{-1}$	44
Troponin T	Peroxidase, absorbance, tube, automated	2.0–5.8	0–$20\,\mu g\,L^{-1}$	126

Table 6 Some examples of disposable devices for the detection and/or quantitation of large molecules

Analyte	Technology	Sample	Lowest detectable amount	Ref.
Troponin I	Lateral flow, gold label	Whole blood	$0.1\,\mu g\,L^{-1}$	127
Troponin T	Lateral flow, gold label	Whole blood	$0.2\,\mu g\,L^{-1}$	128
Luteinizing hormone	Lateral flow, latex particle blue	Urine	–	122
Albumin	Direct agglutination, cassette	Urine	$1\,mg\,L^{-1}$	129
HbA_{1C}	Latex immunoinhibition, cassette	Whole blood	–	98
C-reactive protein	Lateral flow, gold label	Whole blood	$12\,mg\,L^{-1}$	130
Chorionic gonadotrophin	Lateral flow, latex particles, blue	Urine	$1.9\,\mu g\,L^{-1}$	131
Chorionic gonadotrophin	Flow through	Serum	$5\,\mu g\,L^{-1}$	100

reaction (e.g. an NAD → NADH conversion in the presence of pyruvate and lactate dehydrogenase) and the interference in fluorescence assays due to endogenous light scattering, endogenous fluorescence, or the presence of quenching agents. It should also be remembered that samples may sometimes contain exogenous compounds that may interfere (e.g. drugs). Fewer problems are likely in the case of the heterogeneous assay, although the problem of binding proteins, whether normally present or abnormal, may cause problems in any assays; the presence of antispecies antibodies, and heterophilic antibodies can cause particular problems, particularly in the case of heterogeneous immunoassays.[132]

9 CONCLUDING REMARKS

Readers will see from this article that immunochemistry is a huge field in terms of basic chemistry, biochemistry, analytical chemistry, and applied science. The discussion has attempted to illuminate readers' understanding of the basic aspects of the antigen–antibody reaction and in so doing illustrate how it can be exploited for the recognition and quantitation of biomolecules. Some examples have been given to indicate the ingenuity of research scientists in this field.

ABBREVIATIONS AND ACRONYMS

ABDP	*N*-(4-Aminobenzoyl)-*N*′-(pyridyldithiopropionyl)hydrazine
ACTH	Adrenocorticotropic Hormone
AMPPD	3-(2′-Spiroadamantane)-4-methoxy-4-(3″-phosphoryloxy)-phenyl-1,2-dioxetane
ATP	Adenosine Triphosphate
CD	Cluster of Differentiation
CDR	Complementarity-determining Region
CEDIA	Cloned Enzyme Donor Immunoassay
CZE	Capillary Zone Electrophoresis
DELFIA	Dissociation-enhanced Lanthanide Fluorescence Immunoassay
DPEM	*N*-[β-(4-Diazophenyl)ethyl]maleimide
ELISA	Enzyme-linked Immunosorbent Assay
EMIT	Enzyme-multiplied Immunoassay Technique
FAD	Flavin Adenine Dinucleotide
FETI	Fluorescence Excitation Transfer Immunoassay
GC	Gas Chromatography
HAT	Hypoxanthine, Aminopterin, and Thymidine
HPLC	High-performance Liquid Chromatography
HSAB	*N*-Hydroxysuccinimidyl 4-Azidobenzoate
LOCI	Luminescent Oxygen Channeling Assay
MBS	*m*-Maleimidobenzoyl *N*-Hydroxysuccinimide
NAD	Nicotinamide Adenine Dinucleotide
NADH	Reduced Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
PSA	Prostate-specific Antigen
QA	Quality Assurance
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
TPA	Tripropylamine
TSH	Thyroid-stimulating Hormone

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors • High-performance Liquid Chromatography of Biological Macromolecules

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Automation in the Clinical Laboratory • Biochemical Markers of Acute Coronary Syndromes • Biosensor Design and Fabrication • Capillary Electrophoresis in Clinical Chemistry • Drugs of Abuse, Analysis of • Laboratory Instruments in Clinical Chemistry, Principles of • Point-of-care Testing • Serum Proteins

Environment: Water and Waste **(Volume 3)**
Immunoassay Techniques in Environmental Analyses

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Forensic Science **(Volume 5)**
Immunoassays in Forensic Toxicology

Nucleic Acids Structure and Mapping **(Volume 6)**
Aptamers

Particle Size Analysis **(Volume 6)**
Optical Particle Counting • Turbidimetry in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Proteins and Glycoproteins

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Antibiotics, Pharmaceutical Analysis of

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. B.J. Sutton, 'Molecular Basis of Antibody–Antigen Reactions: Structural Aspects', in *Methods of Immunological Analysis*, eds. R.F. Masseyeff, W.H. Albert, N.A. Stainess, VCH, Weinheim, 66–79, 1993.
2. E.A. Padlan, 'Anatomy of the Antibody Molecule', *Mol. Immunol.*, **31**, 169–217 (1994).
3. M.H.V. Van Regenmortel, 'Molecular Dissection of Protein Antigens', in *Structure of Antigens*, ed. M.H.V. Van Regenmortel, CRC Press, Boca Raton, FL, 1–27, Vol. 1, 1992.
4. L. Jin, J.A. Wells, 'Mutational Analysis of Antibody Binding Sites', in *Structure of Antigens*, ed. M.H.V. Van Regenmortel, CRC Press, Boca Raton, FL, 21–36, Vol. 3, 1995.
5. O. Chappey, M. Debray, E. Niel, J.M. Scherrmann, 'Association Constants of Monoclonal Antibodies for Hapten: Heterogeneity of Frequency Distribution and Possible Relationship with Hapten Molecular Weight', *J. Immunol. Methods*, **172**, 219–225 (1994).
6. C.J. van Oss, 'Nature of Specific Ligand–Receptor Bonds, in Particular the Antibody–Antigen Bond', in *Immunochemistry*, eds. C.J. van Oss, M.H.V. Van Regenmortel, Marcel Dekker, New York, 581–614, 1994.
7. C.J. van Oss, 'Hydrophobic, Hydrophilic and Other Interactions in Epitope–Paratope Binding', *Mol. Immunol.*, **32**, 199–211 (1995).
8. T.N. Bhat, G.A. Bentley, G. Boulot, M.I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F.P. Schwarz, R.A. Mariuzza, R.J. Poljak, 'Bound Water Molecules and Conformational Stabilization Help Mediate an Antigen–Antibody Association', *Proc. Natl. Acad. Sci. USA*, **91**, 1089–1093 (1994).
9. D.R. Davies, E.A. Padlan, S. Sheriff, 'Antigen–Antibody Complexes', *Annu. Rev. Biochem.*, **59**, 439–473 (1990).
10. M.H.V. Van Regenmortel, 'The Antigen–Antibody Reaction', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 13–34, 1997.
11. B. Friguiet, A.F. Chafotte, L. Djavadi-Ohaniance, M.E. Goldberg, 'Measurement of the True Affinity Constant in Solution of Antigen–Antibody Complexes by Enzyme-linked Immunosorbent Assay', *J. Immunol. Methods*, **77**, 305–319 (1985).
12. M.L. Doyle, 'Characterization of Binding Interactions by Isothermal Titration Calorimetry', *Curr. Opin. Biotechnol.*, **8**, 31–35 (1997).
13. U. Jonsson, M. Malmqvist, 'Real-time Biospecific Interaction Analysis. The Integration of Surface Plasmon Resonance Detection, General Biospecific Interface Chemistry and Microfluidics into One Analytical System', *Adv. Biosens.*, **2**, 291–336 (1992).
14. R. Karlsson, 'Real Time Competitive Kinetic Analysis of Interactions Between Low Molecular Weight Ligands in Solution and Surface Immobilized Receptors', *Anal. Biochem.*, **221**, 142–151 (1994).
15. S. Terrettaz, T. Stora, C. Duschl, H. Vogel, 'Protein Binding to Supported Lipid Membranes – Investigation of the Cholera Toxin Ganglioside Interaction by Simultaneous Impedance Spectroscopy and Surface Plasmon Resonance', *Langmuir*, **9**, 1361–1369 (1993).
16. E.M. Phizicky, S. Fields, 'Protein–Protein Interactions: Methods for Detection and Analysis', *Microbiol. Rev.*, **59**, 94–123 (1995).
17. C.P. Price, D.J. Newman, 'Light-scattering Immunoassay', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 443–480, 1997.
18. S.A. Levinson, F. Kierszenbaum, W.B. Dandliker, 'Salt Effects on Antigen–Antibody Kinetics', *Biochemistry*, **9**, 322–331 (1970).
19. J. Israelachvili, H. Wennerstrom, 'Role of Hydration and Water Structure in Biological and Colloidal Interactions', *Nature (London)*, **379**, 219–225 (1996).
20. G. Kohler, C. Milstein, 'Continuous Cultures of Fused Cells Secreting Antibody of Predefined Specificity', *Nature (London)*, **256**, 495–497 (1975).
21. O. Makela, I.J.T. Seppala, 'Haptens and Carriers', in *Handbook of Experimental Immunology, Immunochemistry*, ed. D.M. Weir, Blackwell, Oxford, Vol. 1, 3.1–3.31, 1986.
22. E. Harlow, D. Lane, 'Immunizations', in *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 53–138, 1988.
23. A.M. Campbell, *Monoclonal Antibody and Immunoassay Technology*, Elsevier, Amsterdam, 1991.
24. J.R. Birch, R. Boraston, L. Wood, 'Bulk Production of Antibody in Fermenters', *Trends Biotechnol.*, **3**, 162–165 (1985).
25. S.J. Searle, J.T. Pederen, A.H. Henry, 'Antibody Structure and Function', in *Antibody Engineering: A Practical Approach*, 2nd edition, ed. C A.K. Borrebaeck, Oxford University Press, Oxford, 3–51, 1995.
26. E.S. Ward, 'V_H Shuffling Can Be Used to Convert an Fv Fragment of Anti-hen Egg Lysozyme Specificity to One that Recognizes a T-cell Receptor V Alpha', *Mol. Immunol.*, **32**, 147–156 (1995).

27. R. Ducancel, D. Gillet, A. Carrier, E. Lajeunesse, A. Menez, J.-C. Boulain, 'Recombinant Colorimetric Antibodies: Construction and Characterization of a Bifunctional $F(ab)_2$/Alkaline Phosphatase Conjugate Produced in *Escherichia coli*', *Bio/Technology*, **11**, 601–605 (1993).
28. G. Winter, A.D. Griffiths, R.E. Hawkins, H.R. Hoogenboom, 'Making Antibodies by Phage Display Technology', *Anu. Rev. Immunol.*, **12**, 433–455 (1994).
29. L.I. Andersson, I.A. Nichols, K. Mosbach, 'Molecular Imprinting: A Versatile Technique for the Preparation of Separations Materials of Predetermined Selectivity', in *Highly Selective Separations in Biotechnology*, ed. G. Street, Blackie, Glasgow, 206–224, 1994.
30. G. Vlatakis, L.I. Andersson, R. Muller, K. Mosbach, 'Drug Assay Using Antibody Mimics Made by Molecular Imprinting', *Nature (London)*, **361**, 645–647 (1993).
31. A.D. Ellington, J.W. Szostak, 'In Vitro Selection of RNA Molecules that Bind Specific Ligands', *Nature (London)*, **346**, 818–822 (1990).
32. Y. Lin, Q. Qiu, S.C. Hill, S.D. Jayasena, 'Modified RNA Sequence Pools for In Vitro Selection', *Nucleic Acids Res.*, **22**, 5229–5234 (1994).
33. R. Conrad, A.D. Ellington, 'Detecting Immobilized Protein Kinase C Isozymes with RNA Aptamers', *Anal. Biochem.*, **242**, 261–265 (1996).
34. D.W. Drolet, L. Moon-McDermott, T.S. Romig, 'An Enzyme Linked Oligonucleotide Assay', *Nature Biotechnol.*, **14**, 1021–1025 (1996).
35. K.A. Davis, B. Abrams, Y. Lin, S.D. Jayasena, 'Use of a High Affinity DNA Ligand in Flow Cytometry', *Nucleic Acids Res.*, **24**, 702–706 (1996).
36. S.E. Osborne, I. Matsumura, A.D. Ellington, 'Aptamers as Therapeutic and Diagnostic Reagents: Problems and Prospects', *Curr. Opin. Chem. Biol.*, **1**, 5–9 (1997).
37. R. Edwards, 'Radiolabelled Immunoassays', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 327–348, 1997.
38. A.E. Bolton, W.M. Hunter, 'Radioimmunoassay and Related Methods', in *Handbook of Experimental Immunology*, ed. D.M. Weir, Blackwell, Oxford, 26.1–26.56, Vol. 1, 1986.
39. J.P. Gosling, 'A Decade of Development in Immunoassay Methodology', *Clin. Chem.*, **36**, 1408–1427 (1990).
40. J.P. Gosling, 'Enzyme Immunoassay: With and Without Separation', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 351–388, 1997.
41. S. Sun, A.M. Lew, 'Chimaeric Protein A/Protein G and Protein G/Alkaline Phosphatase as Reporter Molecules', *J. Immunol. Methods*, **152**, 43–48 (1992).
42. G. Michal, H. Mollering, J. Siedel, 'Chemical Design of Indicator Reactions for the Visible Range', in *Methods of Enzymatic Analysis*, eds. H.U. Bergmeyer, J. Bergmeyer, M. Grassl, Verlag Chemie, Wienheim, 197–232, 1983.
43. E. Ishikawa, 'Development and Clinical Application of Sensitive Enzyme Immunoassay for Macromolecular Antigens – A Review', *Clin. Biochem.*, **20**, 375–385 (1987).
44. I. Bronstein, J.C. Voyta, G.H.G. Thorpe, L.J. Kricka, G. Armstrong, 'Chemiluminescent Assay of Alkaline Phosphatase Applied in an Ultrasensitive Enzyme Immunoassay of Thyrotropin', *Clin. Chem.*, **35**, 1441–1446 (1989).
45. G.H.G. Thorpe, L.J. Kricka, S.B. Mosely, 'Phenols as Enhancers of the Chemiluminescent Horseradish Peroxidase–Luminol–Hydrogen Peroxide Reaction: Application in Luminescence Monitored Enzyme Immunoassays', *Clin. Chem.*, **31**, 1335–1341 (1985).
46. I.M. Christie, P.H. Treloar, Z.B. Koochaki, P.M. Vadgama, G.N. Smith, 'Simplified Measurement of Serum Alkaline Phosphatase Utilizing Electrochemical Detection of 4-Aminophenol', *Anal. Chim. Acta*, **257**, 21–28 (1992).
47. C.H. Self, 'Enzyme Amplification: A General Method Applied to Provide an Immunoassisted Assay for Placental Alkaline Phosphatase', *J. Immunol. Methods.*, **76**, 389–393 (1985).
48. D.M. Obzansky, B.R. Rabin, D.M. Simons, S.Y. Tseng, 'Sensitive, Calorimetric Enzyme Amplification Cascade for Determination of Alkaline Phosphatase and Application of the Method to an Immunoassay of Thyrotropin', *Clin. Chem.*, **37**, 1513–1518 (1991).
49. I. Hemmila, *Applications of Fluorescence in Immunoassays*, John Wiley & Sons, New York, 1991.
50. I. Hemmila, 'Fluoroimmunoassays and Immunofluorometric assays', *Clin. Chem.*, **31**, 359–370 (1985).
51. E. Soini, H. Kojola, 'Time-resolved Fluorometer for Lanthanide Chelates: A New Generation of Non-isotopic Immunoassays', *Clin. Chem.*, **29**, 65–68 (1983).
52. F. McCapra, D. Watmore, F. Sumun, A. Patel, I. Beheshti, K. Ramakrishnan, J. Branson, 'Luminescent Labels for Immunoassay – From Concept to Practice', *J. Biolumin. Chemilumin.*, **4**, 51–58 (1989).
53. L.J. Kricka, 'Chemiluminescent and Bioluminescent Techniques', *Clin. Chem.*, **37**(9), 1472–1481 (1991).
54. E. Jablonski, 'The Preparation of Bacterial Luciferase Conjugates for Immunoassay and Application to Rubella Antibody Detection', *Anal. Biochem.*, **148**, 199–206 (1985).
55. D.F. Smith, N.L. Stults, H. Rivera, W.D. Gehle, R.D. Cummings, M.J. Cormier, 'Applications of Recombinant Bioluminescent Proteins in Diagnostic Assays', in *Bioluminescence and Chemiluminescence: Current Status*, eds. P.E. Stanley, L.J. Kricka, John Wiley & Sons, Chichester, 529–532, 1991.
56. G.F. Blackburn, H.P. Shah, J.H. Kenten, J. Leland, R.A. Kamin, J. Link, J. Peterman, M.J. Powell, A. Shah, D.B. Talley, S.K. Tyagi, E. Wilkins, T.-G. Wu,

R.J. Massey, 'Electrochemiluminescence Detection for Development of Immunoassays and DNA Probe Assays for Clinical Diagnostics', *Clin. Chem.*, **37**(9), 1534–1539 (1991).
57. N.R. Hoyle, 'The Application of Electrochemiluminescence to Immunoassay-based Analyte Measurement, in *Bioluminescence and Chemiluminescence*, eds. A.K. Campbell, L.J. Kricka, P.E. Stanley, John Wiley & Sons, Chichester, 28–31, 1994.
58. P.R. Finley, R.J. Williams, D.A. Lichti, 'Evaluation of a New Homogeneous Enzyme Inhibitor Immunoassay of Serum Thyroxine with Use of a Bichromatic Analyzer', *Clin. Chem.*, **26**, 1723–1726 (1980).
59. R. Sommer, C. Nelson, A.C. Greenquist, 'Dry-reagent Strips for Measuring Phenytoin in Serum', *Clin. Chem.*, **32**, 1770–1774 (1985).
60. R.R.A. Coombs, M.L. Scott, M.P. Cranage, 'Assays Using Red Cell-labelled Antibodies', *J. Immunol. Methods*, **101**, 1–14 (1987).
61. D.J. Newman, H. Henneberry, C.P. Price, 'Particle Enhanced Light Scattering Immunoassay', *Ann. Clin. Biochem.*, **29**, 22–42 (1992).
62. H.H. Weetall, A.K. Gaigalas, 'Studies on Antigen–Antibody Reactions Using Light Scattering from Antigen Coated Colloidal Particles', *Anal. Lett.*, **25**, 1039–1053 (1992).
63. D.F. Keren, *High Resolution Electrophoresis and Immunofixation Techniques*, Butterworth–Heinemann, Boston, 1994.
64. C.B. Laurell, 'Antigen–Antibody Crossed Electrophoresis', *Anal. Biochem.*, **10**, 358–361 (1965).
65. D.M. Bollag, S.J. Edelstein (eds.), *Protein Methods*, Wiley-Liss, New York, 1991.
66. L.J. Kricka, 'Principles of Immunochemical Techniques', in *Tietz Textbook of Clinical Chemistry*, 3rd edition, eds. C.A. Burtis, E.R. Ashwood, W.B. Saunders, Philadelphia, 205–225, 1998.
67. H.Y. Elder, 'Cryofixation', in *Techniques in Immunocytochemistry*, eds. G.R. Bullock, P. Petrusz, Academic Press, London, 1–28, Vol. 4, 1989.
68. M. de Waele, J.E. Beesley, 'Immunocytochemistry of Blood and Bone Marrow Cells', in *Techniques in Immunocytochemistry*, eds. G.R. Bullock, P. Petrusz, Academic Press, London, 95–126, Vol. 4, 1989.
69. M.G. Macey (ed.), *Flow Cytometry: Clinical Applications*, Blackwell, Oxford, 1994.
70. C.H. Geisler, J.K. Larsen, N.E. Hansen, B.E. Christensen, B. Lund, H. Nielsen, T. Plesner, K. Thorling, E. Andersen, P.K. Andersen, 'Prognostic Importance of Flow Cytometric Immunophenotyping of 540 Consecutive Patients with B-Cell Chronic Lymphocytic Leukemia', *Blood*, **78**, 1795–1802 (1991).
71. R.J. Fulton, R.L. McDade, P.L. Smith, L.J. Kienker, J.R. Kettman, Jr, 'Advanced Multiplexed Analysis with the FlowMetrix™ System', *Clin. Chem.*, **43**, 1749–1756 (1997).
72. C.L. Morgan, D.J. Newman, C.P. Price, 'Immunosensors: Technology and Opportunities in Laboratory Medicine', *Clin. Chem.*, **42**, 193–209 (1996).
73. L.G. Fagerstam, A. Frostellkarlsson, R. Karlsson, B. Persson, I. Ronnberg, 'Biospecific Interaction Analysis Using Surface-plasmon Resonance Detection Applied to Kinetic, Binding-site and Concentration Analysis', *J. Chromatogr.*, **597**, 397–410 (1992).
74. R.J. Davies, P.R. Edwards, H.J. Watts, 'The Resonant Mirror: A Versatile Tool for the Study of Biomolecular Interactions', *Tech. Protein Chem.*, **5**, 285–292 (1994).
75. C.R. Lawrence, N.J. Geddes, D.N. Furlong, 'Surface Plasmon Resonance Studies of Immunoreactions Utilizing Disposable Diffraction Gratings', *Biosens. Bioelectron.*, **11**, 389–400 (1996).
76. P.E. Buckle, R.J. Davies, T. Kinning, D. Yeung, P.R. Edwards, D. Pollardknight, C.R. Lowe, 'The Resonant Mirror: A Novel Optical Sensor for Direct Sensing of Biomolecular Interactions. Part II: Applications', *Biosens. Bioelectron.*, **8**, 355–363 (1993).
77. T.P. Milligan, H.R. Park, K. Noonan, C.P. Price, 'Assessment of the Performance of a Capture Immunoassay for the Bone Isoform of Alkaline Phosphatase in Serum', *Clin. Chim. Acta*, **263**, 165–175 (1997).
78. R.P. Ekins, 'Immunoassay Design and Optimization', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 175–207, 1997.
79. H. Thakkar, D.J. Newman, P. Holownia, C.L. Davey, C.C. Wang, J. Lloyd, A.R. Craig, C.P. Price, 'Development and Validation of a Particle Turbidimetric Inhibition Assay for Urine Albumin on the DuPont aca® Analyser', *Clin. Chem.*, **43**, 109–113 (1997).
80. L.E.H. Miles, C.N. Hales, 'Labelled Antibodies and Immunological Assay Systems', *Nature (London)*, **219**, 186–189 (1968).
81. G. Barnard, F. Kohen, 'Idiometric Assay: Noncompetitive Immunoassay for Small Molecules Typified by the Measurement of Estradiol in Serum', *Clin. Chem.*, **36**, 1945–1950 (1990).
82. C.H. Self, J.L. Dessi, L.A. Winger, 'High-performance Assays of Small Molecules: Enhanced Sensitivity, Rapidity, and Convenience Demonstrated with a Noncompetitive Immunometric Anti-immune Complex Assay System for Digoxin', *Clin. Chem.*, **40**, 2035–2041 (1994).
83. D.M. Kemeny, S.J. Challcombe (eds.), *ELISA and Other Solid Phase Immunoassays: Theoretical and Practical Aspects*, Wiley, Chichester, 1988.
84. M.D.P. Boyle (ed.), *Bacterial Immunoglobulin-binding Proteins, Vol. II Applications in Immunotechnology*, Academic Press, New York, 1990.
85. E.P. Diamandis, T.K. Christopoulos, 'The Biotin–(Strept)avidin System: Principles and Applications in Biotechnology', *Clin. Chem.*, **37**, 625–636 (1991).

86. K. Hellsing, 'Enhancing Effects of Nonionic Polymers on Immunochemical Reactions', in *Automated Immunoanalysis*, ed. R.F. Ritchie, Marcel Dekker, New York, 67–112, Vol. 1, 1978.
87. W.J. Litchfield, A.R. Craig, W.A. Frey, C.C. Leflar, C.E. Looney, M.A. Luddy, 'Novel Shell/Core Particles for Automated Immunoassays', *Clin. Chem.*, **30**, 1489–1493 (1984).
88. M.J. Benecky, D.R. Post, S.M. Schmitt, M.S. Kochar, 'Detection of Hepatitis B Surface Antigen in Whole Blood by Coupled Particle Light Scattering (Copalis™)', *Clin. Chem.*, **43**(9), 1764–1770 (1997).
89. K.E. Rubenstein, R.S. Schneider, E.F. Ullman, 'Homogeneous Enzyme Immunoassay: A New Immunochemical Technique', *Biochem. Biophys. Res. Commun.*, **47**, 846–851 (1972).
90. W.D. Engel, P.L. Khanna, 'CEDIA® In Vitro Diagnostics with a Novel Homogeneous Immunoassay Technique: Current Status and Future Prospects', *J. Immunol. Methods*, **150**, 99–102 (1992).
91. M.E. Jolley, S.J. Stroupe, K.S. Schwenzer, C.J. Wang, M. Lu-Steffes, H.D. Hill, S.R. Popelka, J.T. Holen, D.M. Kelso, 'Fluorescence Polarization Immunoassay III. An Automated System for Therapeutic Drug Determination', *Clin. Chem.*, **27**, 1575–1579 (1981).
92. W.H. Porter, V.M. Hanver, B.A. Bush, 'Effect of Protein Concentration on the Determination of Digoxin in Serum by Fluorescence Polarization Immunoassay', *Clin. Chem.*, **30**, 1826–1829 (1984).
93. E.J. Shaw, R.A.A. Watson, J. Landon, D.S. Smith, 'Estimation of Serum Gentamicin by Quenching Fluoroimmunoassay', *J. Clin. Pathol.*, **30**, 526–531 (1977).
94. E.F. Ullman, M. Schwarzberg, K.E. Rubenstein, 'Fluorescence Excitation Transfer Immunoassays. A General Method for Determination of Antigens', *J. Biol. Chem.*, **251**, 4172–4178 (1976).
95. J. Calvin, K. Burling, R.S. Campbell, S.A.P. Chubb, C.P. Price, 'Evaluation of Fluorescence Excitation Transfer Immunoassay for the Measurement of Plasma Cortisol', *J. Autom. Chem.*, **8**, 80–84 (1986).
96. A.K. Campbell, A. Patel, 'A Homogeneous Immunoassay for Cyclic Nucleotides Based on Chemiluminescence Energy Transfer', *Biochem. J.*, **216**, 185–194 (1983).
97. E.F. Ullman, H. Kirakossia, A.C. Switchenko, J. Ishkanian, M. Ericson, C.A. Wartchow, M. Pirio, J. Pease, B.R. Irvin, S. Singh, R. Singh, R. Patel, A. Dafforn, D. Davalian, C. Skold, N. Kurn, D.B. Wagner, 'Luminescent Oxygen Channeling Assay (LOCI™): Sensitive, Broadly Applicable Homogeneous Immunoassay Method', *Clin. Chem.*, **42**(9), 1518–1526 (1996).
98. R.M. Pope, J.M. Apps, M.D. Page, K. Allen, H.J. Bodansky, 'A Novel Device for the Rapid In-clinic Measurement of Haemoglobin A1c', *Diab. Med.*, **10**, 260–263 (1993).
99. U. Klotz, 'Comparison of Theophylline Blood Measured by the Standard TD_X Assay and a New Patient-side Immunoassay Cartridge System', *Ther. Drug Monit.*, **15**, 462–464 (1993).
100. G.E. Valkirs, R. Barton, 'ImmunoConcentration™: A New Format for Solid Phase Immunoassays', *Clin. Chem.*, **31**, 1427–1431 (1985).
101. K. May, 'Unipath ClearBlue One Step™, Clearplan One Step™ and Clearview™', in *The Immunoassay Handbook*, ed. D. Wild, Macmillan, London, 233–235, 1994.
102. K.F. Buechler, S. Moi, B. Noar, D. McGrath, J. Villela, M. Clancy, A. Shenhav, A. Colleymore, G. Valkirs, T. Lee et al., 'Simultaneous Detection of Seven Drugs of Abuse by the Triage™ Panel for Drugs of Abuse', *Clin. Chem.*, **38**, 1678–1684 (1992).
103. Y. Schouten, R.J. de Winter, J.P.M.C. Gorgels, R.W. Koster, R. Adams, G.T. Sanders, 'Clinical Evaluation of the CARDIAC STATus™, a Rapid Immuno-chromatographic Assay for Simultaneous Detection of Elevated Concentrations of CK-MB and Myoglobin in Whole Blood', *Clin. Chem. Lab. Med.*, **36**, 469–473 (1998).
104. R.A. Bunce, G.H. Thorpe, L. Keen, 'Disposable Analytical Devices Permitting Automatic, Timed Sequential Delivery of Multiple Reagents', *Anal. Chim. Acta*, **249**, 263–269 (1991).
105. R.F. Zuk, V.K. Ginsberg, T. Houts, J. Rabbis, H. Merrick, E.F. Ullman, M.M. Fischer, C.C. Sizto, S.N. Stiso, D.J. Litman, 'Enzyme Immunochromatography: A Quantitative Immunoassay Requiring No Instrumentation', *Clin. Chem.*, **31**, 1144–1150 (1985).
106. W.K. Schomburg, J. Fehrenberg, D. Maas, R. Rapp, 'Active Valves and Pumps for Microfluidics', *J. Micromech. Microeng.*, **3**, 216–218 (1993).
107. L.B. Koutny, D. Schmalzing, T.A. Taylor, M. Fuchs, 'Microchip Electrophoretic Immunoassay for Serum Cortisol', *Anal. Chem.*, **68**, 18–22 (1996).
108. I.M. Song, K. Iwata, M. Yamada, K. Yokoyama, T. Takeuchi, E. Tamiya, I. Karube, 'Multisample Analysis Using an Array of Microreactors for an Alternating-current Field-enhanced Latex Immunoassay', *Anal. Chem.*, **66**, 778–781 (1994).
109. U.-H. Stenman, 'Standardization of Immunoassays', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 245–268, 1997.
110. U.-H. Stenman, J. Leinonen, H. Alfthan, S. Rannikko, K. Tuhkanen, O. Alfthan, 'A Complex Between Prostate-specific Antigen and Alpha 1-Antichymotrypsin is the Major Form of Prostate-specific Antigen in Serum of Patients with Prostatic Cancer: Assay of the Complex Improves Clinical Sensitivity for Cancer', *Cancer Res.*, **51**, 222–226 (1991).
111. K. Jaquet, R. Thieleczek, L.M.G. Heilmeyer, 'Pattern Formation on Cardiac Troponin I by Consecutive Phosphorylation and Dephosphorylation', *Eur. J. Biochem.*, **231**, 486–490 (1995).
112. C.P. Price, T.P. Milligan, C. Darte, 'Direct Comparison of Performance Characteristics of Two Immunoassays

for Bone Isoform of Alkaline Phosphatase in Serum', *Clin. Chem.*, **43**, 2051–2057 (1997).
113. J. Seth, 'Quality Assurance', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Macmillan, London, 211–241, 1997.
114. C.M. Sturgeon, J. Seth, 'Why do Immunoassays for Tumour Markers Give Differing Results? – A View from the UK National External Quality Assessment Schemes', *Eur. J. Clin. Chem. Clin. Biochem.*, **34**, 755–759 (1996).
115. D.W. Chan, 'Immunoassay Automation: From Concept to System Performance', in *Clinical Automation, Robotics, and Optimization*, ed. G.J. Kost, John Wiley & Sons, New York, 419–441, 1996.
116. C.L. Davey, H. Thakkar, D.J. Newman, C.P. Price, 'A Simple, Automated, Nonextraction Assay for Serum Digoxin Using Particle-enhanced Immunoassay', *Ther. Drug Monit.*, **19**, 108–111 (1997).
117. E. Sung, W.E. Neeley, 'A Cost-effective System for Performing Therapeutic Drug Assays. I. Optimization of the Theophylline Assay', *Clin. Chem.*, **31**, 1210–1215 (1985).
118. J. Calvin, K. Burling, R.S. Campbell, S.A.P. Chubb, C.P. Price, 'Evaluation of Fluorescence Excitation Transfer Immunoassay for the Measurement of Plasma Cortisol', *J. Autom. Chem.*, **8**, 80–84 (1986).
119. C.H. Self, J.L. Dessi, L.A. Winger, 'Ultra-specific Immunoassays for Small Molecules: Roles of Wash Steps and Multiple Binding Formats', *Clin Chem.*, **42**, 1527–1531 (1996).
120. V.V. Murthy, A. Karmen, 'A Study of the CEDIA™ Digoxin Immunoassay', *Clin. Chem.*, **36**, 559–561 (1990).
121. T.V. Stabler, A.L. Siegel, 'Chemiluminescence Immunoassay of Cyclosporine in Whole Blood', *Clin. Chem.*, **36**, 906–908 (1990).
122. K. May, 'The Unipath Personal Contraceptive System', in *Natural Contraception Through Personal Hormone Monitoring*, ed. J. Bonnar, Parthenon, London, 35–44, 1996.
123. P.M.S. Clark, C.P. Price, 'Enzyme-amplified Immunoassays: A New Ultrasensitive Assay of Thyrotropin Evaluated', *Clin. Chem.*, **32**, 88–92 (1986).
124. J.E.B. Youens, J. Calvin, C.P. Price, 'Clinical and Analytical Validation of an Enzymometric Assay for Creatinine Kinase-MB Isoenzyme', *Ann. Clin. Biochem.*, **23**, 463–469 (1986).
125. H. Finney, D.J. Newman, W. Gruber, P. Merle, C.P. Price, 'Initial Evaluation of Cystatin C Measurement by Particle-enhanced Immunonephelometry on the Behring Nephelometer Systems (BNA, BN II)', *Clin. Chem.*, **43**, 1016–1022 (1997).
126. M. Muller-Bardorff, K. Hallermayer, A. Schroder, C. Ebert, A. Borgya, W. Gerhardt, A. Remppis, J. Zehelein, H.A. Katus, 'Improved Troponin T ELISA Specific for Cardiac Troponin T Isoform: Assay Development and Analytical and Clinical Validation', *Clin. Chem.*, **43**, 458–466 (1997).
127. C. Heeschen, B.U. Goldmann, R.H. Moeller, C.W. Hamm, 'Analytical Performance and Clinical Application of a New Rapid Bedside Assay for the Detection of Serum Cardiac Troponin I', *Clin. Chem.*, **44**, 1925–1930 (1998).
128. E.M. Antman, C. Grudzien, D.B. Sacks, 'Evaluation of a Rapid Bedside Assay for Detection of Serum Cardiac Troponin T', *J. Am. Med. Assoc.*, **273**, 1279–1282 (1995).
129. M.P. Parsons, D.J. Newman, R.G. Newall, C.P. Price, 'Validation of a Point-of-care Assay for the Urinary Albumin : Creatinine Ratio', *Clin. Chem.*, **45**, 414–417 (1999).
130. P. Urdal, S.M. Borch, S. Landaas, M.B. Krutnes, G.O. Gogstad, P. Hjortdahl, 'Rapid Immunometric Measurement of C-Reactive Protein in Whole Blood', *Clin. Chem.*, **38**, 580–584 (1992).
131. J.C.P. Kingdom, T. Kelly, A.B. MacLean, 'Rapid One-step Urine Test for Human Chorionic Gonadotrophin in Evaluating Suspected Complications of Early Pregnancy', *Br. Med. J.*, **302**, 1308–1311 (1991).
132. J.J. Miller, S.S. Levinson, 'Interferences in Immunoassays', in *Immunoassay*, eds. E.P. Diamandis, T.K. Christopoulos, Academic Press, San Diego, CA, 165–190, 1996.

Infrared Spectroscopy in Clinical Chemistry

Lily M. Ng and Reiko Simmons
Cleveland State University, Cleveland, USA

1 Introduction **1**
2 Specialized Instrumentation for Clinical Chemistry Studies **3**
2.1 Transmission Spectroscopy 4
2.2 Reflective Techniques: Attenuated Total Internal Reflection 5
2.3 Infrared Microspectroscopy 5
3 Data-analysis Techniques **6**
3.1 Spectral Subtraction 6
3.2 Spectral Derivatives 8
3.3 Deconvolution 8
3.4 Multivariate Analysis 10
4 Clinical Applications **13**
4.1 Diagnostic Application 13
4.2 Laboratory Analyses of Pathological Samples 14
4.3 Noninvasive In Vivo Monitoring 14
4.4 Monitoring of Drug/Medicine Metabolism 15
5 Future Prospects **15**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

Infrared (IR) spectroscopy measures the absorption of IR radiation by chemical bonds in a material. Chemical structural fragments of molecules, known as functional groups, tend to absorb IR radiation in the same frequency range regardless of the structure of the rest of the molecule in which the functional group is located. This correlation between the structure of a molecule and the frequencies at which it absorbs IR radiation allows the structures of unknown molecules to be identified and structural or chemical changes of the molecule to be followed. Traditionally, IR spectroscopy has been one of the most important physical methods in the chemical laboratory as it plays an important role in the elucidation of structures and the identification of organic and inorganic compounds. The quantitative analysis of samples is straightforward and is not affected by the physical state of the sample. Gaseous, liquid, and homogeneous and inhomogeneous solid samples all can be conveniently studied.

Applications of IR spectroscopy in clinical chemistry have progressed rapidly in recent years. IR pathological analysis is becoming increasingly important in the detection of biochemical changes in body fluids and tissues. Information obtained from their IR spectra can be used to identify diseases. Advanced investigations are carried out using different data-analysis techniques such as multivariate analysis. Using a pattern-recognition approach, IR spectroscopic data provide a nonsubjective aid in the diagnosis of the disease state and the staging of the disease. Unlike most colorimetric or electrochemical/enzymatic clinical chemistry assays, IR spectroscopic analysis is applicable to a variety of body fluids and tissues: whole blood, plasma or serum, synovial fluid, saliva, urine, cells, and membranes. The methodology is fast and readily automated. Sample preparation is simple and no special reagents or other consumables such as electrodes are required, making IR spectroscopy a low-cost operation. The most attractive advantage of the method is the potential for a rapid multicomponent analysis to be carried out from a single spectrum.

1 INTRODUCTION

IR spectroscopy measures the absorption of IR radiation by chemical bonds in a material. Chemical structural fragments of molecules, known as functional groups, tend to absorb IR radiation in the same frequency range regardless of the structure of the rest of the molecule in which the functional group is located. This correlation between the structure of a molecule and the frequencies at which it absorbs IR radiation allows the structures of unknown molecules to be identified and structural or chemical changes of the molecule to be followed.

The energy of a light wave is directly proportional to its frequency. The frequency is in turn related to the velocity of light by a parameter called the wavenumber, which is defined as the reciprocal of the wavelength of the light wave (Equations 1 and 2):

$$\text{energy} = hcW \tag{1}$$

$$\text{frequency} = cW \tag{2}$$

where h = Planck's constant (6.63×10^{-34} J s) c = velocity of light (3×10^{8} m s^{-1}), and W = wavenumber (cm^{-1}). Wavenumbers are the units typically used in IR spectroscopy to denote different energies.

Qualitatively, the chemical bonds present are identified by the frequencies at which they absorb IR radiation and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

vibrate. For example, the C=O stretching vibrational mode of a carbonyl group occurs at $\sim 1700\,cm^{-1}$ in carboxylic acids, ketones, and aldehydes. Quantitative IR spectroscopic analysis determines the concentrations of molecules in a sample. The molecule of interest is called the analyte. For absorption spectroscopy, the intensity of a spectral band can be represented by the height or area of a peak in an absorbance spectrum. The law relating band intensity to concentration of analyte in the IR beam is the well-known Bouguer–Beer–Lambert law, which we shall abbreviate to Beer's law. For a single analyte at low concentration, in a nonabsorbing solvent, Beer's law gives the absorbance at any wavenumber according to Equation (3):

$$A = \log\left(\frac{1}{T}\right. \quad = alc \tag{3}$$

where A = absorbance, T = transmittance at the wavenumber, a = absorptivity at the wavenumber, l = pathlength, and c = concentration of the analyte. Transmittance is defined according to Equation (4):

$$T = \frac{I}{I_0} \tag{4}$$

where I = light intensity with a sample in the beam and I_0 = light intensity with no sample in the beam.

For a mixture of N analytes, the absorbance at a specific wavenumber is the sum of the absorbances of all analytes which absorb IR radiation at that wavenumber. Hence the quantitative analysis of samples is straightforward and can be reliable down to the picogram level for molecules with a high molar absorptivity. Furthermore, IR spectroscopic analysis is not affected by the physical state of the sample: gaseous, liquid, and homogeneous or inhomogeneous solid samples all can be conveniently studied.

The ''routine'' IR spectroscopy in use at the present is Fourier transform infrared (FTIR) spectroscopy. There are many reviews of FTIR spectroscopy which describe the theory, hardware, techniques, and applications.[1–7] Briefly, the development of FTIR spectroscopy began with the invention of the Michelson interferometer. The purpose of an interferometer is to take a beam of light, split it into two beams, and make one of the light beams travel a different distance than the other. The difference in distance traveled by these two light beams is called the optical path difference. A diagram of a Michelson interferometer is shown in Figure 1. In the Michelson interferometer, an optical path difference is introduced between the two light beams by translating the moving mirror away from the beamsplitter. The light that reflects off the moving mirror will travel further than the light that reflects off the fixed mirror. For monochromatic light, different optical path differences create constructive and destructive interference and the light-beam intensity varies between very bright and very weak. A plot of light intensity versus optical path difference is called an interferogram. A broad-band IR source gives off light at a continuum of wavelengths and each frequency component gives rise to a different wave interferogram. The total interferogram measured by the detector is the summation of all the interferograms from all the different IR frequencies. A simplistic way to think about the process is that the interferometer ''encodes'' the intensity and wavelength information so that all the data can be measured at once, and the Fourier transform decodes the information to obtain the absorbance versus wavenumber spectrum.

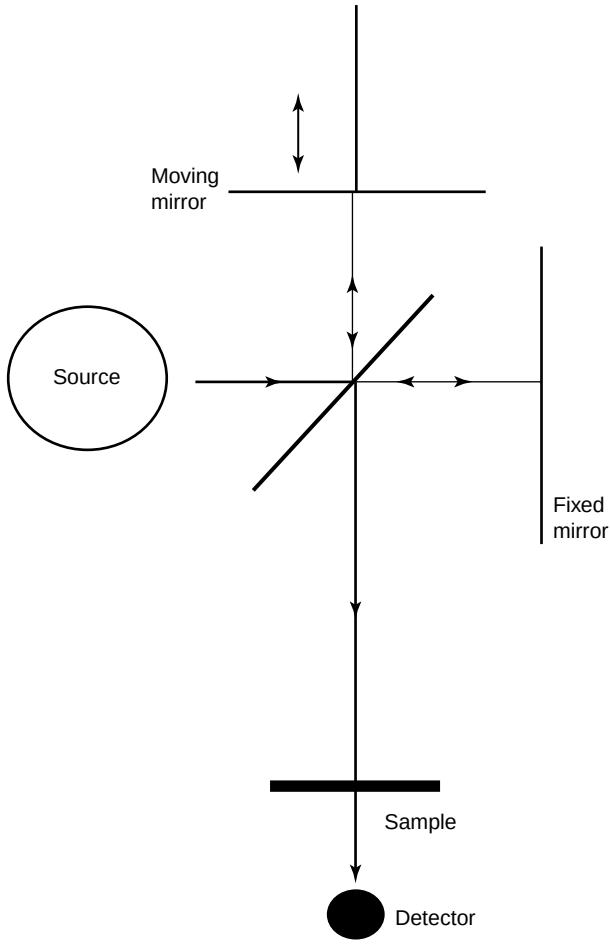

Figure 1 Schematic diagram of an FTIR spectrometer.

The ultimate performance of any spectrometer is determined by measuring its signal-to-noise ratio (S/N). The S/N is calculated by measuring the peak height of a feature in an IR spectrum (such as a sample absorbance peak) and ratioing it to the level of noise at some baseline point nearby in the spectrum. Noise is usually observed as random fluctuations in the spectrum

above and below the baseline. There are two reasons why FTIR spectrometers are capable of significantly higher S/Ns than the old dispersive instruments. The first is called the throughput advantage of FTIR spectroscopy. It is based on the fact that all the IR radiation passes through the sample and strikes the detector at once in an FTIR spectrometer. Hence, the detector sees the maximum amount of light at all points during a scan. The second S/N advantage of FTIR spectroscopy is called the multiplex advantage. It is based on the fact that in an FTIR spectrometer all wavenumbers of light are detected at once. The noise at a specific wavenumber is inversely proportional to the square root of the time spent observing that wavenumber. Because of the multiplex advantage, acquiring data for 10 min on an FTIR spectrometer means that all wavenumbers are observed for a full 10 min. The practical advantage of the multiplexing is that an FTIR spectrometer can acquire a spectrum much faster than a dispersive instrument. This allows multiple scans of the same sample to be added together. For a constant resolution, the S/N is proportional to the square root of the number of scans added together. Thus, a spectrum consisting of 100 added spectra would have an S/N 10 times better than a spectrum comprised of just one scan. Table 1 lists the S/N improvement for various number of scans compared with a single scan.

Despite the many advantages of FTIR spectroscopy, there are limitations on what is achievable with IR spectroscopy in general, and there is one specific limitation of the FTIR technique in particular. A general limitation of FTIR spectroscopy is that it cannot detect atoms or monoatomic ions since these species do not contain chemical bonds, do not posses vibrational motion, and thus do not absorb IR radiation. Homonuclear diatomic molecules do not give IR spectra owing to their symmetry. Another limitation of IR spectroscopy is its use in analyzing complex mixtures. These samples give rise to complex spectra which are difficult to interpret. Fortunately, special data-analysis techniques are now available to aid the interpretation of these complicated spectra. Some of these methods will be discuss in detail in section 3.

Table 1 Number of scans and S/N improvement

Number of scans	S/N improvement
1	–
4	2
16	4
64	8
100	10
400	20
1024	32

It is also difficult to analyze aqueous solutions using IR spectroscopy because water dissolves many IR-transparent materials and water is also a strong IR absorber. Special window and cell materials are available for aqueous solutions and special techniques are available to obtain spectra of aqueous solutions. These will be discussed in section 2.

A specific limitation of FTIR spectroscopy is that most FTIR spectrometers are single-beam instruments. The background spectrum which measures the contribution of the instrument and the environment to the spectrum is measured at a different time than the spectrum of the sample. The sample spectrum is then ratioed to the background spectrum to eliminate the contributions of the instrument and environment. If there are changes in the instrument or the environment between the times when the sample and background spectra are obtained, artifacts can appear in the sample spectrum. This can be overcome by using purging boxes.

There are many general reviews of FTIR spectroscopy which describe the hardware, techniques, and applications and some representative ones are cited here.[1–7]

2 SPECIALIZED INSTRUMENTATION FOR CLINICAL CHEMISTRY STUDIES

Obtaining the IR spectra of biological samples is a new and rapidly expanding field. One difficulty of this application is that biological samples are complex and usually in aqueous solutions. Complex mixtures give complex spectra that require special detection methods and special data-analysis techniques. On the other hand, IR spectroscopy has high sensitivity, is easy to perform, requires little or no sample preparation, is generally reproducible, and is rapid and inexpensive, being able to identify many components all at once.

IR spectroscopy offers a number of advantages for the analysis of body fluids. Unlike most colorimetric or electrochemical/enzymatic clinical-chemistry assays, IR spectroscopic analysis can be applied to a variety of biofluids such as whole blood, plasma, serum, synovial fluid, saliva, and urine. As an in vitro diagnostic and pathological analytical technique, the methodology is fast and readily automated, making it suitable for use as a routine technique in the clinical-chemistry laboratory. As an in vivo monitoring technique, it is noninvasive and has great promise as an accurate, rapid tool to monitor progress and response to therapy in patients. IR spectrometers can be made compact, rugged, relatively inexpensive, and user independent, making the technology capable of point-of-care operation in the emergency room, critical care unit, and general practitioner's office, and potentially

even as a home monitoring device. Since no reagents or other consumables are required, IR spectroscopy has a low operational cost. The most attractive advantage of the method is the potential for a rapid multicomponent analysis to be carried out from a single measurement.

IR spectroscopy for biological samples is usually performed in two frequency ranges: mid-IR from 4000 to 500 cm^{-1} and near-IR from 12 800 to 4000 cm^{-1}. The near-IR range is particularly useful for biological samples because water does not absorb in this range. IR bands in the near-IR region are combination and overtone vibrations arising from the C–H, O–H and N–H moieties of molecules whereas the bands which dominate the mid-IR region are fundamental molecular vibrations of all functional groups. Thus, the near-IR and the mid-IR regions carry different information contents. The near-IR assays of substrates such as total protein, total cholesterol, and triglycerides compare well in performance with those which use the mid-IR range.

2.1 Transmission Spectroscopy

In this technique, the sample is placed in the light beam of the spectrometer and the intensity of the incident beam is compared with the intensity transmitted by the sample (Figure 2). The concentration of the absorbing sample is calculated directly using Beer's law. The transmission method is widely used, primarily because it produces spectra with high S/N and requires only simple accessories. Liquid samples have generally been measured in fixed-path-length transmission cells. The transmission path length is usually dictated by the absorption characteristics of the liquid sample matrix. Owing to the high water absorptivities in the mid-IR region, aqueous samples are limited to path lengths of less than 0.03 mm. Another requirement for quantitative analysis in the mid-IR range is the use of an IR-transparent window material that is compatible with the water matrix. Table 2 lists commonly used window materials for transmission spectroscopy of aqueous solutions.

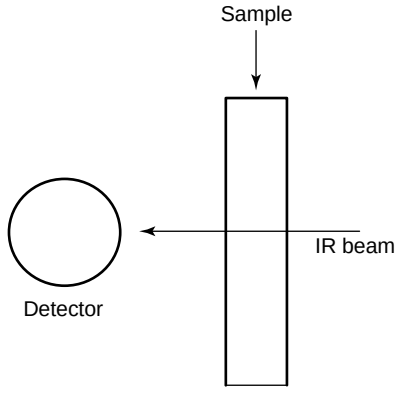

Figure 2 Transmission IR spectroscopy.

2.1.1 Transmission Difference Spectroscopy

Transmission difference spectroscopy uses the spectral subtraction technique. A ''reference'' spectrum is subtracted from a ''sample'' spectrum to obtain the ''difference'' spectrum. The difference spectrum indicates the changes in concentrations of chemical species present in the original reference spectrum or the appearance of new species in the sample spectrum. Transmission difference spectroscopy is ideal for tracking reactions or detecting small changes in the original species as a function of time,

Table 2 Optical materials for IR spectroscopy of aqueous solutions

Window material	Useful range (cm^{-1})	Refractive index at 2000 cm^{-1}	Properties
LiF	83 000–1400	1.33	Slightly soluble in water, suitable for near-IR
CaF_2	77 000–900	1.40	Insoluble in water, resists most acids and alkalis
BaF_2	66 666–800	1.45	Insoluble in water, soluble in acids and NH_4Cl, good resistance to F and F^-
AMTIR	11 000–725	2.5	Hard, brittle, insoluble in water, attacked by alkalis, good for ATR
AgBr	22 000–300	2.3	Insoluble in water, soluble in acids and NH_4Cl, corrosive to metals and alloys
KRS-5	17 000–250	2.38	Slightly soluble in water, soluble in bases, not soluble in acids, good transmission, ideal for ATR, highly toxic
Fused silica	40 000–3000	1.46	Unaffected by most solvents, good for near-IR
ZnS	50 000–770	2.25	Insoluble in water, normal acids and bases, and virtually all organic solvents, reacts with strong oxidizing agents
ZnSe	20 000–500	2.43	Insoluble in water, high resistance to chemical attack, ideal for ATR
Ge	5000–550	4.01	Insoluble in water, soluble in hot sulfuric acid and aqua regia, suitable for ATR
Si	8333–33	3.42	Very hard and inert crystal, suitable for ATR in selective spectral range

ATR, attenuated total reflection.

chemical environment, or reaction. Spectral subtraction of spectra of biological molecules will be discussed in more detail in section 3.

2.1.2 Near-infrared Spectroscopy

Measurement of aqueous fluids in the near-IR region is almost exclusively carried out in the transmission mode. Unlike in the mid-IR region, near-IR transmission cells can be made from conventional glass or quartz window material. The cell path length in the short-wavelength near-IR range (6000–12 500 cm^{-1}) can be about a factor of 10 greater for physiological fluids than in the longer wavelength range. This is because of the reduced water absorptivities observed in this range. The longer path-length requirement is more convenient, but it also dictates a larger sample volume which may not be readily available for certain biofluids.

2.2 Reflective Techniques: Attenuated Total Internal Reflection

For ATR spectroscopy, the sample is brought near a crystal of IR-transparent material of high refractive index where the light is totally internally reflected and where the sample interacts with the evanescent wave (Figure 3). The effective path length for this interaction (the depth of penetration of the IR beam into the sample) is typically a fraction of a wavelength, and depends on many parameters, making ATR a very powerful technique. Recently, the ATR technique has experienced a resurgence in popularity in the analysis of biological samples, primarily owing to its suitability for studies of aqueous solutions. Internal reflection is also nondestructive. For example, to record the spectrum of the skin on one's finger via transmission spectroscopy, one must cut a thin slice of the skin, whereas via ATR spectroscopy one need only place the finger in contact with the optical element.

ATR is the IR spectroscopic method of choice for aqueous solutions. ATR crystals are impervious to water and the short path lengths mean that water will not be totally absorbing. This allows water to be subtracted from sample spectra so that the spectra of solutes in aqueous solutions can be obtained. A disadvantage of liquid ATR spectroscopy is that it is not very sensitive because the path lengths are small. Special efforts have to be made to detect solutes at concentrations less that 0.1%. For quantitative work, it is important to have a way to position the crystal in the sampling accessory reproducibly such that the depth of penetration is controlled. The same ATR crystal must be used for standards and samples when performing quantitative analyses. The crystal must be kept clean and scratch free. Scratches on an ATR crystal affect sample–crystal contact and thus influence the depth of penetration. The most commonly used ATR crystal materials are listed in Table 2.

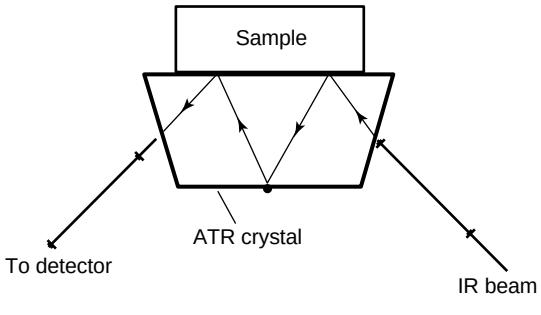

Figure 3 ATR spectroscopy.

2.3 Infrared Microspectroscopy

IR microspectroscopy is the coupling of optical microscopy and IR spectroscopy. IR microspectroscopy is attractive because it combines IR spectroscopy's ability to identify subtle differences in the chemical components and optical microscopy's capability to resolve spatially the location of sample features. It allows the visual and IR examination of different parts of the sample on the microscopic scale. This technique is particularly attractive because it is highly sensitive. The IR microscope can be used in four ways: visual transmission, visual reflection, IR transmission, and IR reflection. Detailed descriptions of microscope designs can be found in the literature.[8–12] In IR microscopy, it is imperative that the sample and background spectra be run using the same settings. This is as important as using the same number of scans and the same resolution for these spectra. The sample spectrum is usually taken before the background spectrum and the two spectra are ratioed at a later time to obtain the absorbance or transmittance spectrum. IR spectra of microscopic samples can also be obtained in reflectance. The sample is typically mounted on a gold mirror and the IR beam is reflected off the sample. The background spectrum is obtained by reflecting the IR beam off a clean portion of the gold mirror.

IR microspectroscopy has great potential as a biochemical and biomedical probe. In contrast to conventional histological techniques, IR spectral images are derived from the intrinsic vibrational properties of the biological components. The IR spectra reflect changes in biochemical and biophysical properties resulting from disease or toxic injuries. Using IR microspectroscopy, these intrinsic changes across a sample can be visualized.

Kidney stones, arteries, and individual cells have been studied using this technique. One of the most exciting developments in the field is the discovery that healthy and cancerous cells have different IR spectral features.[10–14]

Chemically induced DNA damage that prompt the progression of cancers to the metastatic state can also be detected.[10]

2.3.1 Mapping

To define the area of the sample to be analyzed, physical apertures are placed in the beam path such that radiation is sampled only from a well-specified region of the sample. To acquire spatially resolved spectral images using this technique, a sampling area is first defined by the apertures. Spectral data are then obtained from this region, the sample is moved so that the apertures are focused on an adjacent portion of the sample, and spectral data are again obtained. The whole process is repeated until the desired sampling region is mapped. A high S/N spectrum with 4 cm^{-1} resolution in the 700–4000 cm^{-1} range can be obtained in ~30 s in a typical case. The smallest regions that can be probed are on the order of 15 μm. The major drawbacks of using mapping for the collection of spatially resolved images are the long collection times and the limited spatial resolution.

2.3.2 Imaging

FTIR spectral imaging can be accomplished by a new instrument that incorporates a step-scan Fourier transform interferometer, a microscope, and a focal plane array. This instrument simultaneously obtains high-resolution IR spectra for each pixel of a two-dimensional imaging array. This new technique has been applied to the identification of leaking silicone from breast implants, metastatic cells and tissues, and neurotoxicity.[1,12] It shows great promise for diagnosis of pathologies resulting from disease and toxic insult.

3 DATA-ANALYSIS TECHNIQUES

The raw data obtained from an FTIR spectrometer are often put through spectral manipulations. The main goal of such spectral manipulations is to extract more information, which can supply possible hints for further interpretations. Understanding and working with the IR data is crucial because the raw measurements often appear simple yet contain much qualitative and quantitative information. As researchers struggle to glean insights into the studies of bioanalytical and clinical samples with FTIR spectroscopy, various approaches are developed and utilized. Currently, there are many types of software which will allow quick access to these methods. The following are a few generally well-accepted methods especially applicable to clinical-chemistry research. In a chapter on the proper use of spectral manipulations, Smith stresses the importance of protecting the integrity of the data.[1] He recommends two very practical rules: (i) always retain a copy of the unaltered spectrum and (ii) always note what kind of spectral manipulations have been performed. One needs to have a clear understanding of both the principles and the logical path involved in the techniques to be able to interpret the resulting data fully. The section on each method will contain brief discussions of the principles, applicability, advantages, and limitations.

3.1 Spectral Subtraction

As early as the late 1960s, researchers studying proteins, DNA and peptides attempted to address the overwhelming, broad water absorption band in the mid-1600 cm^{-1} wavenumber range. The importance of the amide I, amide II, and amide III bands in studies of the biochemical and clinical samples is well established. Unfortunately, amide I bands (due to the carbonyl absorption of amides), amide II bands (due to N–H bending) and amide III bands (due to C–N stretching) associated with these samples invariably cluster in the same general range as that of the water band mentioned above. It was necessary for researchers to find a way to get past the opaque, large absorption of the water in order to ''see'' the information contained in the spectra of the sample in aqueous solution. Spectral subtraction is a method well suited for this purpose.

The principle behind spectral subtraction is easy to understand. A spectrum of a mixture such as a protein being studied in solution is called a ''sample'' spectrum. On the other hand, a ''reference'' spectrum is one component of the mixture such as water for the above example. A set of subtractions can be performed, point-by-point, to subtract away the absorbance values of the reference from the sample spectra to obtain the difference. When performed following certain stipulations, the results can show peaks attributable to the nonreference material, which were previously hidden.[1]

A closer examination of the above principle reveals that the validity of Beer's law must hold for both spectra in question. Specifically, the absorbance measured by the spectra must be linearly proportional to concentration. Usually, the concentrations of the sample and the reference will not be exactly the same. If the linearity relationship holds for the two spectra, however, the reference spectrum can be multiplied by a factor to resolve this problem.

When computer software is used, the following procedural steps may be helpful:

1. Place the absorbance spectra of the sample and the reference on the same screen and compare the two. The peaks that are common to both belong

to the reference material. Identify these common features. If the reference spectrum is one which is well established (e.g. water, saline, methanol), one of these common peaks may be easily identified as the characteristic peak to be normalized.

2. Pick a common peak and use the ''interactive'' feature of the software to experiment with subtraction factors. Keep in mind that the factor chosen will fit into the subtraction Equation (5):

$$(\text{sample}) - (\text{factor})\,(\text{reference}) = \text{results} \qquad (5)$$

The goal is to subtract out this common peak and produce a flat section. When the factor is too small, the reference peak will still remain. On the other hand, a large factor will produce negative peaks pointing downwards. It is usually possible to expand the wavelength range near the peak in question to achieve successively better results. A subtraction factor slightly above or below 1000 gives the best results. This is an indication that the concentrations of the two spectra are similar. Sometimes, an internal reference may be used. An internal reference is a molecule which absorbs at a wavelength different from the reference and the sample. The absorbance of the internal reference is then used to adjust the subtraction factor.

3. Go back and examine visually the three spectra: sample, reference, and results. Evaluate for two types of possible artifacts. When linear proportionality does not hold for a certain section of the wavelength, jagged remnants of the subtraction may show as artifacts. The second type of artifacts appear as relatively sharp peaks above and below the baseline resembling a derivative curve. These artifacts are often caused by shifts in the reference peaks due to the chemical interactions occurring in the mixture. When these artifacts are apparent but are located in the nonessential wavenumber range, one may be able to disregard these artifacts and focus on the crucial section of the spectrum. In addition, it will be wise to check to see if in fact additional information can be gained from the overall subtraction results.

Once the approach to the water subtraction is perfected, application for the general method of spectral subtraction is greatly expanded. Using the label, ''difference spectra'', one may subtract any known component spectra from the mixture. In addition, it is reasonable then to obtain the spectra of the same sample under different conditions and perform subtraction. In fact, Griffiths and de Haseth call attention to the fact that ''Much of the work described since 1980 has made use of spectral subtraction techniques to focus on reactions or interactions taking place on specific sites of large molecules''.[7] In their studies of the metal–DNA interaction, Tajmir-Riahi et al. utilized the difference spectra algorithm [(DNA solution + metal cation solution) − (DNA solution)].[15,16] They mentioned the use of certain DNA bands as an internal reference. They reasoned that the band at 893 cm^{-1} was useful because its absorbance is due to deoxyribose–phosphate vibrations and exhibited no change upon complexation of DNA with cation. In other words, this band will cancel out completely when subtraction is performed.

Criteria for water subtraction can be spelled out as part of the experimental methods. For an example, Olinger et al. specified the following: (i) a straight baseline from 2000 to 1725 cm^{-1} and (ii) no lobes with a negative absorbance.[17]

The main data-analysis method used by Rothschild et al., in their studies of the light-activated membrane protein bacteriorhodopsin, was also difference spectrometry.[18] By performing subtraction, they were successful in pinpointing the absorption changes due to the alterations in the protein. Bagley et al.[19] also studied rhodopsin but took the FTIR subtraction method a step further. They were successful in setting up an innovative experimental design whereby three different products of ''photosequence'', rhodopsin, bathorhodopsin and isorhodopsin, were compared. Owing to the light absorption, rhodopsin is known to go through a sequence of changes, which triggers electrical conversion of the photoreceptor cell membrane. This conversion of electrical properties, in turn, results in the neural signal. By pairing the three products (bathorhodopsin–rhodopsin, bathorhodopsin–isorhodopsin, and isorhodopsin–rhodopsin) and calculating the difference spectra for each pair, they were able to obtain information regarding ''the protein–chromophore linkage, isomeric state, and opsin conformation''.[19] Griffiths and de Haseth cited a biocompatibility study of synthetic polymers as one of the example of applications of FTIR spectroscopy to biomedicine.[7] To study the changes in blood proteins as they are adsorbed on a polymer surface, they utilized two types of subtractions. For each spectrum, water subtraction was conducted to obtain the spectrum of the adsorbed protein. As these researchers were interested in the complex changes that occur with time, several scans representing certain time intervals were paired and subtracted from each other. Hence, the study was able to focus on the detailed differences occurring during the timescale of the experiment.[19]

Advantages and some practical considerations regarding the use of ATR in studies of whole blood and plasma were carefully outlined by Wang et al.[5] They caution against achieving a single water subtraction factor

when various thickness and concentrations of biofluids are studied. Water, blood, and plasma can have different refractive indices and penetration depths of the evanescent waves. The resulting spectra may show high variability in intensities. Therefore, it is especially important to be mindful of the conditions affecting the subtrahend in the clinical analysis of biofluids.

In summary, spectrum subtraction is a relatively simple, yet useful data-analysis technique, which has been applied in several resourceful ways by various researchers. Smith reminds any user of this method that "very few subtractions are perfect".[1] Most of all, one needs to keep in mind the linearity stipulation regarding the concentration of the sample and the reference (Beer's law).

3.2 Spectral Derivatives

Many of the articles and reviews on spectral derivatives mention the original study of the second-derivative spectra of proteins in ^2H_2O by Susi and Byler.[20] In 1983, they explained the use of the second-derivative Fourier transform spectra in relation to the resolution of inherently broad, overlapping protein bands. They presented a comparison of the original spectrum and its second derivative (in graphical form) in order to pinpoint the additional information gained by the process. They stated that, "the second derivative of the original spectra offers a direct way to identify the peak frequencies of characteristic components and thus permits much more detailed qualitative and, eventually, quantitative studies".[20]

Smith points out that "an IR spectrum is a mathematical function".[1] Therefore, one can calculate and determine the slope of such function by taking the *derivative*. Currently, most computer software will furnish several methods, such as "point difference" and "Savitsky–Golay", which will calculate several orders (i.e. first through ninth) of derivatives. In most cases, the analyst need only enter the desired number for the order (first = 1, second = 2, etc.) of the derivative. There are no complicated parameters to be chosen or considered for this process!

Actually, the main purpose of obtaining the derivatives is *not* to obtain the value of the "slope", but to use the information contained in the resulting derivative spectrum. As implied in the previous paragraph, the second derivative is most useful because it contains three features corresponding to each peak in the original spectra. In particular, "the bottom of the downward-pointing feature"[1] pinpoints the exact wavenumber of the maximum absorbance of the peak. These wavenumbers, then, can be used in turn to estimate the number of overlapping bands in the composite and to locate their possible peak positions. A further explanation of the use of the derivatives is given in the section 3.3.

When calculating and utilizing the second derivative, care must be taken to keep "the increase in noise and/or the appearance of side lobes at a minimum".[21] The absorbance interferogram will be multiplied by a parabolic weighting function of the form $F(x) = ax^n$ $(n = 2)$ in the process of the second-derivative calculation. The effect of this multiplication, unfortunately, is to amplify the noise at high retardations.[7] In the regions of high retardation (regions further out from the base of the peak), generally, the signal level is already lower and the noise tends to be constant and predominant even before the multiplication. At the same time, since it can be proven that "best resolution = maximum retardation",[7] it will be helpful to experiment to find the appropriate level of "resolution" *prior* to taking the derivatives. Often, a resolution of $2\,cm^{-1}$ is used for the original spectra. Perhaps this consideration should be taken into account when the original spectrum is actually measured, long before the analysis stage. Users should keep in mind the following rule of thumb: "each time the order of the derivative is increased by two, the S/N of the spectrum will degrade by approximately an order of magnitude".[7]

In summary, the computational procedure of taking the spectral derivatives is commonly utilized to estimate the number of smaller bands contained in a large composite band and to locate specific peak positions. Computational execution for the process is simple using, for example, the "point difference" method in suitable software. Since the number and the peak positions are needed in the data-analysis method referred to as "deconvolution", the application of the spectral-derivative method will be discussed in the next section.

3.3 Deconvolution

Deconvolution can be one of the most useful and powerful methods of interpreting spectra. At the same time, its use can easily be challenged. Therefore, the literature abounds in reviews of the method, evaluations of its merits, applications, etc.

Biochemical analysis of protein secondary structure continues to be a popular topic of investigation, originating in the late 1970s to early 1980s. As mentioned previously, the amide I band is often studied in relation to the changes in the protein structure. It is generally known that the amide I band consists of a number of overlapping component bands. Many researchers have attempted to establish the correspondence between the protein structural changes and the changes in the band components. This approach led to various attempts to resolve the large composite band into separate narrower bands for the purpose of identification.

In 1981, Kauppinen et al. developed a computational method that they entitled, "a digital self-deconvolution method".[22] It is known today as Fourier self-deconvolution (FSD) and is widely used as the most effective procedure for narrowing IR bands.[23] Again, some fundamental understanding of the procedure will be helpful when using the computer software.

An absorbance band can be expressed mathematically according to Equation (6):

$$A(\nu) = \frac{A_0\gamma^2}{\gamma^2 + (\nu - \nu_0)^2} \quad (6)$$

where A_0 = maximum absorbance, γ = half-width at half-height (HWHH) and ν_0 = wavenumber for maximum absorbance.[23] When this band is Fourier transformed, the resulting interferogram is expressed by Equation (7):

$$I(x) = F\{A(\nu)\} = 0.5A_0\gamma\cos(2\pi\nu_0 x)\exp(-2\pi\gamma x) \quad (7)$$

where x = a spatial frequency with units of centimeters. In this equation, the γ value determines the rate of the exponential decay. Therefore, the greater the width of the absorbance band, the quicker the transform will decay. If this process is taken in reverse, one could decrease the decay of the exponential term and thereby reduce the bandwidths of the spectrum. To achieve the decrease in the exponential decay, $I(x)$ needs to be multiplied by an exponentially increasing function. When the reverse Fourier transform is performed on the product, $I'(x)$ [i.e. $F^{-1}\{I'(x)\} = A'(\nu)$], a narrower absorbance band is obtained. The above process does, however, increase the noise in the same manner, requiring a smoothing function.

In a paper on the general theory of FSD, Kauppinen et al. listed three instrumental factors which limit the degree of resolution enhancement: in addition to the S/N of the spectrum, they include the dynamic range of the digitization and computational processes and the instrumental resolution used when the spectrum is recorded.[22] They continued to state that all three factors result in the truncation of the interferogram. This truncation, in turn, results in the introduction of artifacts called side-lobes into the spectrum. The multiplication of the interferogram by an apodization function, prior to the performance of the reverse Fourier transform will eliminate these side-lobes. Instruction regarding apodization will be given in the next section.

The following suggestions may be useful with most computer software on deconvolution:

1. Identify a band (or a group of bands) which most likely is composed of several smaller bands. These bands are often broad, asymmetric, and may contain a shoulder or two, giving suggestions of hidden bands. (Follow the software directions to define this region graphically.) The narrowest bandwidth of the group, measured at the half-height of the peak, must be broader than the instrumental resolution used for the original spectrum.
2. Smith states that "the most important user adjustable parameter in deconvolution is the amount of resolution enhancement".[1] This parameter, called the "resolution enhancement factor", is often denoted by γ or K. It is the ratio of the widths measured before and after deconvolution. In other words, K or $\gamma_{(\text{factor})} = \gamma_{(\text{before})}/\gamma_{(\text{after})}$, where γ = full width at half-height (FWHH). Several guidelines are outlined below which can be utilized to choose the appropriate value of the resolution enhancement factor.
   (a) Check the S/N of the spectra and use Table 3 to estimate the maximum.
   (b) Take the second derivative of the bands in the region to determine how many real bands can be expected and where they may be located (see section 3.2).
   (c) Experiment with the factor value until the resulting spectra shows the approximate number of expected bands in the defined wavenumber region. Overdeconvolution may show hundreds of new features with the possible appearance of high, regular, periodic noise features and large negative side-lobes (see the explanation regarding apodization for the side-lobes). Underdeconvolution will result in the bands retaining the shapes similar to the original spectrum. Bands may be only partially separated and/or not separated as expected from the derivative spectrum. In addition, the value of the spectral resolution may be lower than expected. According to Surewicz and Mantsch, "the condition that gives the maximum band narrowing while keeping the increase in noise and/or the appearance of the side-lobes at a minimum"[23] is the key. Apply the parameter for the resolution enhancement.
   (d) To eliminate the side-lobes, there are usually choices of apodization functions such as boxcar, triangular, and Bessel by which the interferogram can be multiplied. Again, experiment to

Table 3 Maximum resolution enhancement vs S/N

S/N	Maximum resolution enhancement factor
100 : 1	2
1000 : 1	3
10 000 : 1	4

achieve a good balance between the width of the bands and the amount of reduction of the side-lobes. Determine which function and what percentage of smoothing is desirable. Apply these parameters and let the software do the work!

Applications of deconvolution methodology provide the possibility of detailed protein secondary structural analysis.[23] The spectra resolved by FSD can be helpful in studying subtle spectral changes brought about because of an external parameter.[22] Once the correlation between resolved components in amide band areas and specific polypeptide structures have been established, researchers can design novel approaches to observe structural changes. Using various experimental designs, parameters such as temperature, pressure, concentration, pH, and solvent can be varied and the resulting detailed structural changes can be observed. Although applicable to all types of protein, Surewicz and Mantsch pointed out that the method is especially helpful in studies of membrane proteins in the hydrophobic environment of a lipid bilayer.[23] The IR study of this particular species of protein tends not to be hampered by the problems encountered in the X-ray diffraction or circular dichroism (CD) spectra (e.g. unavailability of quality crystals, distortions of CD spectra by large membrane fragments). A review by Surewicz and Mantsch includes a large section regarding the applications of resolution enhancement for studies of protein secondary structure.[23] They list studies of Ca^{2+} transport ATPase from sarcoplasmic reticulum, bacteriorhodopsin, myelin basic protein, β-lactoglobulin, methionine aporepressor from *Escherichia coli* and two peptide hormones. It is remarkable to see the variety represented in the experimental designs.

Investigations of protein structural changes using FSD frequently make comparisons of percentages of areas under the component bands with percentages of the conformational structures (α-helix and β-turns, etc.) referenced by other experimental techniques. An assumption is made in the above instance, that the effective absorptivities of the observed bands are ''very similar'' for the corresponding, yet different, conformational structures.[23] According to Surewicz and Mantsch, studies showed a good correlation ''for about 20 proteins between the secondary structure estimates obtained from X-ray data and IR analysis''.[20,23] They concluded that although the general validity remains to be tested, the assumption appears to be reasonable.[23] Nevertheless, they warn against the use of these estimated percentages as absolute values of the structural contents.

Jakobsen and Wasacz conducted an extensive study of IR spectra–structure correlations of albumin and managed to assign a vibrational assignment for most of the observed bands.[24] Along with the band assignments, they studied the conformational changes upon and during adsorption. Deconvolution was the method they utilized for all spectra, but they based their results on several approaches: past assignments, their own experimental justifications, Raman spectra, and a special software program.[24] This software enabled the researchers to calculate ''the percentage contribution of each amino acid to the intensity at any frequency''. The IR spectra of the 20 amino acids that make up the primary sequence of a protein were used to come up with ''a composite spectrum''. This composite spectrum was based on ''the weight percentage of each amino acid present in the protein''.

There are three main points to be made as summary statements regarding deconvolution. First and foremost, all authors caution against the use of peak heights for quantitative analysis. The process of deconvolution retains peak positions but change the areas and the shapes of the peaks.[1] Therefore, peak heights must not be used for measures of proportions or as indicators of changes.[23] Secondly, although FSD is generally applicable to any spectrum [IR, ultraviolet/visible (UV/VIS), fluorescence, Raman, nuclear magnetic resonance (NMR), etc.], the method must be applied only to bands whose widths are inherently broader than the instrumental resolution.[1,22] Finally, one must take special care in determining the ''resolution enhancement factor''. One needs to choose spectra with high S/N to begin the process and verify to make sure that all the features in the deconvoluted spectra are real. The best way to check is to remeasure the spectrum, if possible, at higher resolution and S/N. If a spectral feature at a lower resolution disappears at higher resolution and S/N, it was most likely an artifact! In other words, it would be wise to confirm the real features by using higher factor values and different apodization functions.

3.4 Multivariate Analysis

In bioanalytical and clinical studies, a researcher is rarely faced with the analysis of a single component. The complexity of biological systems demands multicomponent quantitative analysis. At the same time, the numbers of spectra taken in a typical near-IR or mid-IR experiment could easily number in the hundreds. When combined with the range of the wavelengths, the data points involved quickly become large. However, despite the potentially high information content of FTIR measurements, without a sound analysis technique the qualitative and quantitative information could remain buried. This is why bio/clinical studies using FTIR spectroscopy call for multivariate data analysis. In recent years, mathematical, statistical, database, and expert-system techniques have

come together to be known as *chemometrics*. Multivariate data analysis is a good example of chemometrics applied to the spectroscopic data. At the time it was published (1990), Davies stated that, ''a considerable number of multivariate data analysis methods exist'', and new methods are constantly being developed.[25] In other words, the choice of the method for one's experimental data analysis depends on the software and the particular software parameter selected. An attempt will be made to summarize a few major categories with emphasis on the basic understanding and practical information.

The term ''multivariate analysis'' is often used for analysis dealing with multicomponent data. Most multivariate analysis methods applied to chemical data are based on least squares (LS) techniques.[26] At the same time, all multicomponent spectral analyses are based on the additivity of Beer's law. Therefore, it will be advantageous to grasp a basic understanding of the two ideas, LS techniques and Beer's law.

The additivity of Beer's law may be expressed in the following sequential steps:

1. for a series of components at *a single absorbance* (Equation 8):

$$A(\nu) = \sum a_i(\nu) l c_i \tag{8}$$

   at the i th component of the mixture;
2. for *a range of absorbance* (a spectrum) of a mixture (Equation 9):

$$\mathbf{A}_j = \sum \mathbf{a}_i \mathbf{l} \mathbf{c}_{ji} \tag{9}$$

   where **A**, **a**, **l** and **c** are vector;
3. for *a series of absorbance spectra* with components of linearly independent concentrations (Equation 10):

$$\mathbf{A} = \mathbf{ELC} \tag{10}$$

   where the equation is expressed in a matrix form; **A** = the vector of absorbances, **E** = the matrix of absorptivities, **L** = the vector of pathlengths, and **C** = the vector of concentrations. All vectors and matrices are in bold type.

This expression in matrix form is important because various methods in multicomponent analysis begin with this form. In fact, various quantitative analyses differ only in the ways in which they perform matrix algebra upon the data represented in the above format.

The initial step for the multivariate analysis is the calibration step to obtain the standard spectra. The experimental design must provide for standards with known concentrations of the components of interest. Hence the information regarding **A**, **L**, and **C** is provided by the standards and the values of **E** will be supplied by the calibration. After the calibration has been performed, the concentrations of unknowns can be predicted by measuring the sample absorbances and entering the appropriate values into the calculations.

3.4.1 Least Squares Methods

The use of linear regression is a fundamental and classical approach. When dedicated microcomputers became more readily available, its applications were expanded. Once the basic application of LS techniques to IR spectroscopy was established, both expansion and improvements continued, which resulted in many sophisticated variations. Two approaches involving LS methods will be summarized.

3.4.1.1 Classical Least Squares or **K** *Matrix Method* In the classical least squares (CLS) approach, pathlength and absorptivity matrices are combined into a single matrix called **K**. Therefore, the basic equations are $\mathbf{K} = \mathbf{EL}$ and $\mathbf{A} = \mathbf{KC}$, with the absorbance expressed as a function of the concentration. Taking the **K** matrix, a matrix operation called ''LS fit'' is performed to produce, by definition, ''the best available model of the data''.[1] When **K** has been determined, the unknown concentrations comprising the matrix **C** can be calculated through another matrix operation. The advantage of the **K** matrix method is the relatively straightforward mathematics involved. Also, if so desired, use of this method permits ''an averaging effect'' by use of many standards and wavenumbers in the calibration. There are two basic limitations: (i) its sensitivity to the presence of impurities or unexpected components and (ii) its intolerance of interactions between the sample components. These limitations require, first of all, that the standards do not have any impurities. Also, all components and their concentrations must be known prior to analysis. If impurities are present in the samples, since they must be included in the **K** matrix presentation, researchers must be fully aware of their existence. In other words, one must know the complete chemical composition prior to analysis. This can be a tall order in the case of a complex system such as bio/clinical samples. For these reasons, applications of the **K** matrix work best where the sample compositions are predictable and interactions between components are nonexistent.

3.4.1.2 Inverse Least Squares or **P** *Matrix Method* In the inverse least squares (ILS) method, the original matrix expression $\mathbf{A} = \mathbf{ELC}$ can be rewritten as $\mathbf{C} = \mathbf{PA}$ where $\mathbf{P} = (\mathbf{EL})^{-1}$. In this format, the concentration is a function of absorbance. Therefore, even if the concentration of an impurity is not taken into account, the **P** matrix can be obtained as long as there are as many absorbance measures as there are concentrations. Consequently, with

P matrix operation, the same LS matrix manipulation can be performed successfully with only the concentrations of the known components.

According to Griffiths et al., the **P** matrix can be very time-consuming.[2] One must ensure that there are as many absorbance measurements as components, and these absorbance measurements must be sensitive to changes in concentration. Generally, this means preparing and running many standards. However, not every wavenumber in an entire spectrum can be selected to perform the analysis, and therefore it makes inefficient use of the data.[1] However, in a practical situation, accounting for every component in a sample is difficult. For this reason, the **P** matrix method is often chosen.

3.4.2 Factor Analysis

The aim of the factor-analysis approach is to develop a meaningful mathematical model of the chemical system and use it to predict the properties of test samples.[25] Its major strength is the ability to aid the analyst in isolating the desired information that may be hidden in relatively large amounts of data.[27] Factor analysis is a general technique which has been applied to a wide variety of problems. Apparently, Antoon et al. first applied the factor-analysis approach to IR spectroscopy.[2] As a technique, in addition to factor analysis, the method employs some form of linear regression to reach the final analytical results. The following reasons were listed by Smith for the recent popularity of factor analysis.[1] Factor analysis:

- is insensitive to the presence of impurities;
- makes full use of all spectral data;
- does not require a standard for every wavenumber used;
- helps in identifying the number of components in a mixture;
- can reproduce the spectrum of each component;
- can predict the component concentrations in unknown samples;
- is tolerant of low-quality data.

There are several factor-analysis methods and, in addition, many variations have evolved. Two major types will be introduced: principal component regression (PCR) and partial least squares (PLS) methods. Since these methods are intended for routine, rapid, and nondestructive determination of chemical composition, they are especially suited for clinical analysis.[28] When these methods are employed using computer software, the analysis is performed on the entire spectra with no special requirements for selections of peaks, baselines, etc. As alluded to in the previous section, they may be used when the spectral bands are overlapped or affected by baseline variations, spectral noise, purge variations, sample impurities, and nonlinear response.

3.4.2.1 Principal Component Regression In PCR, prior to the application of the ''regression'' technique, the information content is distilled into a smaller number of *principal components*. Those principal components which correlate strongly with the calibration set of concentration values are used to develop the model system, while those with low correlations are dropped. Hence the method can indicate the errors associated with the prediction and the validity of applying the model.[25]

When the data from the standard spectra are entered, the program will compute the *average* of these spectra. Each standard spectrum is then compared with the *average* spectrum and a new spectrum called a *factor* is calculated. The *factor* describes the difference between the standard and the *average*. The amount of the *factor* in each standard spectrum is called a *score*. Each *score* is multiplied by the *factor*, and the product is subtracted from each standard spectrum. The result of the subtraction is called a *residual*. The *residuals* in turn are used in the next iteration to calculate a new *average* spectrum, a new *factor*, and a new set of *scores*. This is how, each time, the iteration process will strip the spectral contribution of some components from the standard spectra. Using the method described above, factor analysis accounts for the variability in the standard spectra.

The following steps may be incorporated when using PCR as a part of a software program:

1. obtain standard spectra and enter the necessary information regarding the standards into the software;
2. choose the part of the spectrum to be analyzed;
3. determine the optimum number of *factors*;
4. use a process called ''cross-validation'' to test for the optimum number of *factors*;
5. let the software conduct the calibration;
6. measure unknown samples and apply the factor analysis.

It will be wise to keep in mind that the computer program does not know how many *factors* are needed to best describe the data. The iteration will continue until the user stops the process. In fact, several authors point to the determination of the optimum number of *factors* as the key to the use of factor analysis.[1,7,25,27] Some experimentation and practice, in addition to the use of ''cross-validation'', will be required for the best results.

3.4.2.2 Partial Least Squares Method According to McClure et al., the PLS method was developed by Wold et al.[27] PLS differs from PCR in that information contained in both the dependent and independent

variables (absorbance and concentration) plays a role in the development of the calibration model.[1,25,27] Davis also stated that PLS tends to show relevance toward chemical values, since the chemical data are used to find a pattern in the data which correlates with factors.[25]

There are two ways of performing PLS. PLS-1 calculates a separate set of factors and scores for each component whereas PLS-2 calculates one set for the entire data set. Although good software will allow the choice of different factor-analysis methods, it will be advantageous to have some comparative information regarding these methods. The following list was compiled from information noted by several authors:[1,25,27]

- PCR is better known and the easiest of the factor-analysis methods;
- PLS is more accurate than PCR;
- PLS has fewer dimensions, generally, and is easier to interpret;
- PLS-1 is probably more accurate than PLS-2 and works better where the concentrations vary widely;
- PLS-2 takes much less computer time than PLS-1 but gives equivalent results if the concentrations do not vary much.

In a chapter describing the software Computerized Infrared Calculations on Materials (CIRCOM), McClure and Lehmann gave an example of a study of lipids.[27] They analyzed the IR spectra of 21 synthetic mixtures of 14 lipids found in serum. They concluded that the method, which is an offshoot of PCR, made successful predictions for triglycerides, lecithin, sphingomyelin, free cholesterol, and esterified cholesterol. The results for fatty acids were usable for the differentiation of normal and potentially pathological types. They did not feel their results were useful for other lipids included in the study. They remarked that the study was a beginning of ''a more extensive investigation of multiple lipid analysis''.[27] Closer analysis of this example may shed some light regarding the details of PCR method.

Heise and Bittner studied blood substrates in human plasma using PLS.[29] The substrates they chose to study were total protein, glucose, total cholesterol, triglycerides, and urea. Near-IR ATR spectroscopy was utilized. They employed a ''cross-validation'' method to verify the calibration results. This paper with its many references gives a useful description of the development of a multicomponent assay.

The topic of the study conducted by Malins et al.[10] was the linkage of hydroxyl-radical-induced DNA damage with the metastatic state of human breast cancers. The experimental design used the IR spectra in an innovative manner. A three-dimensional plot of principal components from factor analysis was constructed to help visualize the spectral relationship of the cancer and noncancer DNA groups. The authors stated that, ''future research may confirm that principal component analysis (PCA) of FTIR spectra is a promising technique with high structural specificity for discriminating DNA phenotypes in cancer diagnosis and prediction''.[10]

In summary, with the aid of good software, multivariate analysis can be a useful method in handling a large set of complex data. Smith states that ''as long as a physical or chemical property of a sample causes the sample's spectrum to change, the spectral changes and measurements of the property can be correlated''.[1] Furthermore, by simply obtaining their spectra, these properties can be predicted for various experimental samples. Using the above-mentioned multicomponent-analysis methods, one is able to obtain useful predictions from complex bio/clinical data. In general, then, the use of IR spectra can save time and expense in terms of analysis as well as sample preparations.

4 CLINICAL APPLICATIONS

To date, the emergence of IR spectroscopy in clinical chemistry is still in its infancy. The application of this analytical tool to clinical studies and diagnosis has generated a lot of enthusiasm as well as speculation.[30–34] Recently, IR pathology has become increasingly important, since diseases are manifested by changes in the composition of body fluids and tissues, and these changes can be elucidated from the chemical information contained in their IR spectra. Diagnostic applications of IR spectroscopy provide physicians with an objective aid in the identification of the disease state or for staging the disease. This can be achieved with a pattern-recognition approach. Quantitative analysis provides clinically relevant parameters in body fluids which may be used to indicate the metabolic status of the patient. The following applications show that IR spectroscopy has great promise as an accurate and rapid multicomponent analytical technique in clinical chemistry.

4.1 Diagnostic Application

IR spectroscopy has been increasingly applied to the diagnosis of diseases. The initial development was for convenient ways for routine screening and fast diagnosis, e.g. a ^{13}C-labeled urea breath test for screening for *Helicobacter pylori* infection and the diagnosis of liver disease by the detection of $^{13}CO_2$ in exhaled air. More recent developments have involved the combination of visible microscopy and IR spectroscopy in the identification of cancer cells and the elucidation of the possible

Table 4 Examples of diagnostic application of IR spectroscopy

Disease/condition	Analyte	Technique[a]	Refs.
Breast cancer	DNA	Microspectroscopy	10, 35
Breast cancer	Tissues	FEWFTIR	36, 37
Astrocytomas	Protein	Microspectroscopy	38
Cervical lesions/cancer	Tissues, cells	High-pressure FTIR	39, 40
		Microspectroscopy	41
		FTIR	42
Lymphoid tumors	RNA/DNA	Microspectroscopy	43, 44
Drug resistance	Cells	FTIR	45
Biocatalytic activity	Model systems	DRASTIC	46
Metabolite overproduction	Model systems	DRASTIC	46
Skin cancer	Tissues	Fiber-optic ATR	47
Single-cell screening	DNA, proteins, lipids	Synchrotron microspectrometry	48
Bone diseases	Bone samples	Microspectroscopy	49

[a] FEWFTIR, fiber-optic evanescent wave Fourier transform infrared; DRASTIC, diffuse reflectance absorbance spectroscopy taking in chemometrics.

mechanism of the development of cell cancer and the progression of cancer to the metastatic state. Some examples of diagnostic applications are given in Table 4.

4.2 Laboratory Analyses of Pathological Samples

FTIR spectroscopy has been used for fast multicomponent analyses of human blood and other pathological samples. Improvements in analysis are obtained by the adoption of multivariate calibration techniques as described above. For routine blood-serum analysis, it is possible to use a PLS algorithm to calculate the concentrations of cholesterol, triglycerides, glucose, total protein, lipids, urea, and uric acid. In addition to blood serum, other pathological samples have also been analyzed by IR spectroscopy. Fecal fat and fecal carbohydrates can be studied by near-IR reflectance analysis. IR analysis of stool samples is fast and does not require unpopular sample handling. Results indicate that near-IR spectroscopy may be a new, reliable, and accurate test in the diagnosis of carbohydrates and fat malabsorption. Other samples such as gallstones, urinary stones, tissues, bones, and urine have also been analyzed. Selective examples of IR pathological analyses are listed in Table 5.

4.3 Noninvasive In Vivo Monitoring

In vitro monitoring is inherently invasive and the results are often delayed by 1 h or more when the analyses are performed in the central laboratory. The delay may be greatly reduced if the analyses are performed near the patient. In vivo monitoring may be noninvasive and may provide continuous real-time data, but the accuracy sometimes does not match that of in vitro measurements. In vivo monitoring is best applied in the detection of trends of change, and it is used for quantities that change rapidly

Table 5 Examples of pathological samples analyzed by IR spectroscopy

Sample	Analyte	Technique	Refs.
Gallstones	Cholesterol, bilirubinate salt, elements	Microspectroscopy	50–53
Fecal fat	Fat contents	Near-IR	54, 55
Calcified tissues	Minerals	Photoacoustic IR	56
Human adipose tissue	*trans*-Fatty acids	ATR	57
Synovial fluid	Synovial fluid	Near-IR	58
Serum, blood	Multicomponents	Near-IR	58–60
Amniotic fluid	Amniotic fluid	Near-IR	58
Lipase	Lipase	ATR	61
Serum	Multicomponents	Laser excitation IR	62
Biomembranes	Vesicles	Surface-enhanced IR	63
Tissues	Hemoglobin, oxyhemoglobin, cytochrome aa_3, indocyanine green	Near-IR	64
Urine	Urease	FTIR	65
Iliac crest biopsies	Mineral contents	Microspectroscopy	66
Cortical bones	Minerals and matrix	Microspectroscopy	67
Blood	Glucose	FEWFTIR	68
Hair	Drugs	Microscopy	69

Table 6 Examples of non-invasive in vivo monitoring by IR spectroscopy

Sample	Analyte	Refs.
Blood (through skin)	Glucose	70–77
Tissue, muscle	Oxygen	78–86
Blood (through skin)	Hemoglobin	74–76, 79, 87–90
Tissue, muscle	Hemoglobin, oxyhemoglobin	74–76, 85, 87, 90, 91
Brain/cerebral	Hemoglobin, oxyhemoglobin	88, 89, 92, 93
Blood (through skin)	Urea, cholesterol	76, 89
Blood (through skin)	Albumin, proteins	76, 89, 94
Muscle	Phosphocreatine	82
Muscle	Glycolytic metabolites	95
Mammalian fluids	Nitric oxide	96
Tissue, blood	pH	97, 98

and unpredictably and where a suitable therapeutic action is available. Near-IR spectroscopy is normally used for monitoring. Examples of noninvasive in vivo monitoring include oxy- and deoxyhemoglobin in the brain, blood glucose, bilirubin, albumin, urea, cholesterol, tissue oxygenation, cerebral circulation, and oxygen metabolism during surgery, cerebral blood volume, and cytochrome oxidase. Examples of these are listed in Table 6.

4.4 Monitoring of Drug/Medicine Metabolism

IR spectroscopy is used routinely for the identification and structural determination of new pharmaceutical compounds (Table 7). A new application of IR spectroscopy in clinical chemistry is for the monitoring of drug/medicine metabolism in the body. It can be used to determine the concentration of drugs and the metabolites in body fluids or body parts, e.g. hair. A breath test has also been used with an isotopically labeled drug to determine the concentration of a metabolite of the drug. The concentration of the metabolite is then used to determined the rate of metabolism of the drug. The effects of drugs in vitro can also be monitored by the changes in cell viability after treatment with drug using FTIR microspectroscopy. Ex vivo drug response, sensitivity, or resistance to drugs can be predicted from the IR spectra of the diseased cells before and after treatment of the patients with the drugs.

5 FUTURE PROSPECTS

Future prospects for application of various FTIR experimental techniques in bioanalytical/clinical studies appear to have at least two main possibilities.

The first possibility is the study of protein and DNA secondary structural analysis. Until recently, although IR spectroscopy was recognized as a potentially useful method, several major limitations interfered with its

Table 7 Examples of monitoring of drug/medicine metabolism by IR spectroscopy

Drug/medicine	Methodology	Refs.
Various medicaments	Noninvasive IR detection of the medicine, metabolites, or markers (patent)	99
Drugs administered by transdermal delivery	Noninvasive quantification by IR of diffusing species (e.g. salicylic acid) through artificial neural network (research model)	100
Drugs detectable in hair	Probe of longitudinal sections of hair along growth line for time measurement; probe of cross-sections to review hydrophobic/hydrophilic characteristics of drug by IR microspectroscopy (clinical)	101
Isotopically labeled drugs	Detection of isotopically labelled carbon dioxide or other metabolites in the breath for rate of metabolism (clinical)	102
Tetracycline	Tracking of rate of release from local drug delivery system consisting of poly(lactide-co-glycolide) films (research)	103
Bryostatin 1,2-chlorodeoxyadenosine	Detection of cell membrane changes as basis of predicting ex vivo drug response (sensitivity/resistance) (preclinical)	104–106
Indomethacin, ibuprofen	Detection of effects on cerebral perfusion and oxygenation of intravenous ibuprofen and indomethacin treatment in preterm infants by noninvasive near-IR spectroscopy (clinical)	107
Propofol (an anesthetic drug)	Study of magnesium–propofol interaction in vitro as related to fluctuations of Mg in serum after clinical administration of propofol (research)	108
Adamantyl maleimide	Study of conformational changes in gastric carcinoma cell membrane protein after treatment by microspectroscopy (research)	109

application to studies of protein conformation. With the introduction of the powerful, computerized FTIR instrumentation, the improvement of the S/N and the availability of extensive data manipulation became a reality.[110] When these new developments are coupled with sophisticated methods of analysis, the use of FTIR spectroscopy may present several definite advantages. Unfortunately, so-called classical methods of protein and DNA structural analysis have known limitations. Even when crystals of high quality are available, the crystallographic analysis gives static information. CD spectroscopy used in connection with the study of secondary structure addresses proteins in aqueous solution but may not be suitable for the lipid environment.[111] To understand better the conformational changes, studies of protein must involve investigations in different environments, under physiological and nonphysiological conditions. FTIR spectroscopy allows the easy detection of structural changes with samples in water, organic solvents, lipids, and solid films with reasonable requirements for sample preparation. Furthermore, the results obtained will be dynamic in nature. As these changes are documented quantitatively, use of the spectral manipulations such as derivatives and deconvolution may permit further investigations of the breakdown of components (e.g. α-helix and β-turns). In addition, FTIR difference spectroscopy and the use of FTIR with ATR can increase detection capabilities. Hand in hand with the new growth in the information content regarding proteins and DNA, it is reasonable to expect further expansion in the use of FTIR methods in the area of structural analysis.

The second area of possible future growth is clinical/medical diagnostics and prediction research. Admittedly, use of FTIR spectroscopy in this area is a new venture. Recently, however, many examples can be found in various professional journals. They utilize a full complement of multicomponent analysis with computer software. Many researchers skillfully manipulate the experimental designs to analyze chemical changes under various environmental and bio/medical conditions. Studies involving substrates in various body fluids are one of these groups. Effective comparisons can be made, for example, of blood samples in dried film with that of the liquid form. The second group of research involves work on the analysis of diseased bio-samples. The pioneering work of Malins et al.,[10] for instance, shows interesting future possibilities.

The ease and speed of analysis and the sample preparation and measurements using FTIR spectroscopy have not been fully utilized. With proper consideration of innovative experimental designs, the possibilities for the use of FTIR spectroscopy plus computerized analysis are wide open. The challenge awaiting researchers is to continue to work to build on the groundwork, which will extend the powerful physical method of FTIR spectroscopy from the chemistry laboratory into the biochemical and clinical settings.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflection
CD	Circular Dichroism
CIRCOM	Computerized Infrared Calculations on Materials
CLS	Classical Least Squares
DRASTIC	Diffuse Reflectance Absorbance Spectroscopy Taking in Chemometrics
FEWFTIR	Fiber-optic Evanescent Wave Fourier Transform Infrared
FSD	Fourier Self-deconvolution
FTIR	Fourier Transform Infrared
ILS	Inverse Least Squares
IR	Infrared
LS	Least Squares
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
S/N	Signal-to-noise Ratio
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy in Microbiology • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by • Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules

Chemical Weapons Chemicals Analysis **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Diagnostic Hematology • Drugs of Abuse, Analysis of • Electrolytes, Blood Gases, and Blood pH • Glucose Measurement • Laboratory Instruments in Clinical Chemistry, Principles of • Molecular Biological Analyses

and Molecular Pathology in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry • Point-of-care Testing

Forensic Science **(Volume 5)**
Microspectrophotometry in Forensic Science

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. B.C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton, FL, 1996.
2. J.R. Ferraro, L.S. Basile (eds.), *Fourier Transform Infrared Spectroscopy, Application to Chemical Systems*, Academic Press, New York, 1978.
3. N.J. Harrick, *Internal Reflection Spectroscopy*, Interscience, New York, 1967.
4. H.H. Mantsch, D. Chapman (eds.), *Biomedical Infrared Spectroscopy in Infrared Spectroscopy of Biomolecules*, Wiley-Liss, New York, 1996.
5. J. Wang, M. Sowa, H.H. Mantsch, A. Bittner, H.M. Heise, 'Comparison of Different Infrared Measurement Techniques in the Clinical Analysis of Biofluids', *Trends Anal. Chem.*, **15**, 286–296 (1996).
6. H.A. Havel (ed.), *Spectroscopic Methods for Determination of Protein Structures in Solution*, VCH, New York, 1996.
7. P.R. Griffiths, J.A. De Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York, 1986.
8. R.G. Messerschmidt, M.A. Harthcock (eds.), *Infrared Microspectroscopy, Theory and Applications*, Marcel Dekker, New York, 1988.
9. H. Humecki (ed.), *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York, 1995.
10. D.C. Malins, N.L. Polissor, S.J. Gunselman, 'Progression of Human Breast Cancers to the Metastatic State is Linked to Hydroxyl Radical-induced DNA Damage', *Proc. Natl. Acad. Sci. USA*, **93**, 2557–2563 (1996).
11. E.W. Ciurczak, '3-D Hyperspectral Infrared Biomedical Imaging', *Spectroscopy*, **14**(1), 12–15 (1999), and references cited therein.
12. H.H. Mantsch, M. Jackson (eds.), *Progress in Biomedical Optics, Proceedings of SPIE*, SPIE, Billingham, WA, Vol. 3257, 1998.
13. J.L. Knenig, C.M. Snively, 'Fast FTIR Imaging: Theory and Applications', *Spectroscopy*, **13**(11), 22–28 (1998), and references cited therein.
14. M. Jackson, M.G. Sowa, H.H. Mantsch, 'Infrared Spectroscopy: A New Frontier in Medicine', *Biophys. Chem.*, **68**(1–3), 109–125 (1997).
15. H.A. Tajmir-Riahi, R. Ahmad, M. Naoui, S. Diamantoglou, 'The Effects of HCl on the Solution Structure of Calf Thymus DNA: a Comparative Study of DNA Denaturation by Proton and Metal Cations Using FTIR Difference Spectroscopy', *Biopolymers*, **35**, 493–501 (1995).
16. S. Alex, P. Dupuis, 'FTIR and Raman Investigation of Cadmium Binding by DNA', *Inorg. Chim. Acta*, **157**, 271–281 (1989).
17. J.M. Olinger, D.M. Hill, R.J. Jakobsen, R.S. Brody, 'Fourier Transform Infrared Studies of Ribonuclease in H_2O and ^2H_2O Solutions', *Biochim. Biophys. Acta*, **869**, 89–98 (1986).
18. K.J. Rothschild, H. Marrero, 'Infrared Evidence that the Schiff Base of Bacteriorhodopsin is Protonated: bR570 and k Intermediates', *Proc. Natl. Acad. Sci. USA*, **79**, 4045–4049 (1982).
19. K.A. Bagley, V. Balogh-Nair, A.A. Croteau, G. Dollinger, T.G. Ebrey, L. Eisenstein, M.K. Hong, K. Nakanishi, J. Vittitow, 'Fourier-transform Infrared Difference Spectroscopy of Rhodopsin and Its Photoproducts at Low Temperature', *Biochemistry*, **24**, 6055–6071 (1985).
20. H. Susi, D.M. Byler, 'Protein Structure by Fourier Transform Infrared Spectroscopy: Second Derivative Spectra', *Biochem. Biophys. Res. Commun.*, **115**(1), 391–397 (1983).
21. *GRAMS/32 Software Manual*, Galactic Industries Corporation, Saleur, NH, 1991.
22. J.K. Kauppinen, D.J. Moffatt, H.H. Mantsch, 'Fourier Self-deconvolution: a Method for Resolving Intrinsically Overlapped Bands', *Appl. Spectrosc.*, **35**(3), 271–276 (1981).
23. W.K. Surewicz, H.H. Mantsch, 'New Insight into Protein Secondary Structure from Resolution-enhanced Infrared Spectra', *Biochim. Biophys. Acta*, **952**, 115–130 (1988).
24. R.J. Jakobsen, F.M. Wasacz, 'Infrared Spectra–Structure Correlations and Absorption Behavior for Helix Proteins', *Appl. Spectrosc.*, **44**(9), 1478–1490 (1990).
25. B. Davies, in *Computer Methods in UV, Visible and IR Spectroscopy*, eds. W.O. George, H.A. Willis, Royal Society of Chemistry, Cambridge, 53–67, 1990.
26. Y. Liang, O.M. Kvalheim, 'Robust Methods for Multivariate Analysis – A Tutorial Review', *Chemom. Intell. Lab. Syst.*, **32**, 1–10 (1996).
27. G.L. McClure, C.A. Lehmann, in *Computer-enhanced Analytical Spectroscopy*, ed. H.L.C. Meuzelaar, Plenum Press, New York, 277–312, Vol. 2, 1990.

28. P.S. Wilson, in *Computer Methods in UV, Visible and IR Spectroscopy*, eds. W.O. George, H.A. Willis, Royal Society of Chemistry, Cambridge, 129–146, 1990.
29. H.M. Heise, A. Bittner, 'Multivariate Calibration for Near-infrared Spectroscopic Assays of Blood Substrates in Human Plasma Based on Variable Selection Using PLS-regression Vector Choices', *Fresenius' J. Anal. Chem.*, **362**, 141–147 (1998).
30. P. Franck, P. Nabet, B. Dousset, 'Applications of Infrared Spectroscopy to Medical Biology', *Cell Mol. Biol.*, **44**(2), 273–275 (1998).
31. L. Rintoul, H. Panayiotou, S. Kokot, G. George, G. Cash, R. Frost, T. Bui, P. Fredericks, 'Fourier Transform Infrared Spectrometry: a Versatile Technique for Real World Samples', *Analyst*, **123**(4), 571–577 (1998).
32. H. Georg, K. Hauser, C. Rodig, O. Weidlich, F. Siebert, 'Time-resolved Infrared Spectroscopy of Biomolecules', *NATO ASI Ser., Ser. E*, **342**, 243–261 (1997).
33. J.L.R. Arrondo, F.M. Goni, 'Infrared Spectroscopic Studies of Membrane Lipids', *NATO ASI Ser., Ser. E*, **342**, 229–242 (1997).
34. M.J. Nowak, L. Lapinski, J.S. Kwiatkowski, J. Leszczynski, 'Molecular Structure and Infrared Spectra of the DNA Bases and Their Derivatives: Theory and Experiment', *Comput. Chem. (Singapore)*, **2**, 140–216 (1997).
35. D.C. Malins, 'Methods for Determining Oxidative Injury to DNA Using IR Absorption Spectroscopy for Cancer Diagnosis', PCT Int. Appl., W09526502, A1 951005, 1995.
36. N.I. Afanasyeva, S.F. Kolyakov, S.G. Artjushenko, V.V. Sokolov, G.A. Frank, 'Minimally Invasive and Ex Vivo Diagnostics of Breast Cancer Tissues by Fiber Optic Evanescent Wave Fourier Transform IR (FEW-FTIR) Spectroscopy', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3250**, 140–147 (1998).
37. J. Spievogel, L. Lobik, I. Nissenkorn, R. Hibst, Y. Gotshal, A. Katzir, 'Cancer Diagnostic Using Fourier Transform Fiberoptic Infrared Evanescent Wave Spectroscopy (FTIR-FEWS)', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3262**, 185–191 (1998).
38. L.-S. Lee, C.-W. Chi, H.-C. Liu, C.-L. Cheng, M.-J. Li, S.-Y. Lin, 'Assessment of Protein Conformation in Human Benign and Malignant Astrocytomas by Reflectance Fourier Transform Infrared Microspectroscopy', *Oncol. Res.*, **10**(1), 23–27 (1998).
39. P.T.T. Wong, 'High Pressure FT-IR Spectroscopy for Biomedical and Cancer Research', in *AIP Conference Proceedings 309 (High-pressure Science and Technology–1993, Pt. 2)*, 1431–1434, 1994.
40. P.T.T. Wong, 'High Pressure FT-IR Spectroscopy for Biomedical and Cancer Research', in *AIP Conference Proceedings 309 (High-pressure Science and Technology–1993, Pt. 2)*, 1431–1434, 1994.
41. B.R. Wood, M.A. Quinn, B. Tait, M. Ashdown, T. Hislop, M. Romeo, D. McNaughton, 'FTIR Microspectroscopic Study of Cell Types and Potential Confounding Variables in Screening for Cervical Malignancies', *Biospectroscopy*, **4**(2), 75–91 (1998).
42. J. Liu, W. Han, Y. Peng, R. Feng, X. Su, 'Determination of Canceration Cells with Fourier Transform Infrared Spectrometry', *Fenxi Huaxue*, **26**(7), 883–885 (1998).
43. P.G.L. Andrus, R.D. Strickland, 'Cancer Grading by Fourier Transform Infrared Spectroscopy', *Biospectroscopy*, **4**(1), 37–46 (1998).
44. P.G.L. Andrus, R.D. Strickland, 'Cancer Grading by Fourier Transform Infrared Spectroscopy', *Biospectroscopy*, **4**(1), 37–46 (1998).
45. J.M. Le Gal, H. Morjani, M. Manfait, 'Application of Fourier Transform Infrared Spectroscopy to the Characterization of the Biomolecular Mechanisms Underlying Resistance to Drugs in Cancer Cell Lines', *Proceedings of 5th International Conference on Spectroscopy of Biological Molecules*, eds. T. Theophanides, J. Anastassopoulou, N. Fotopoulos, Kluwer, Dordrecht, 399–400, 1993.
46. D.B. Kell, M.K. Winson, R. Goodacre, A.M. Woodward, B.L. Alsberg, A. Jones, E.M. Timmins, J.J. Rowland, 'DRASTIC (Diffuse Reflectance Absorbance Spectroscopy Taking in Chemometrics). A Novel, Rapid, Hyperspectral, FTIR-based Approach to Screening for Biocatalytic Activity and Metabolite Overproduction', *Stud. Org. Chem. (Amsterdam)*, **53**, 61–75 (1998).
47. N.I. Afanas'eva, V.G. Artyushenko, S.F. Kol'yakov, D.Y. Pakhomov, V.V. Sokolov, G.A. Frank, V.S. Letokhov, 'Spectral Diagnosis of Tumor Tissues by Fiberoptic IR Spectroscopy', *Dokl. Akad. Nauk*, **356**(1), 118–121 (1997).
48. N. Jamin, P. Dumas, J. Moncuit, W.H. Fridman, J.L. Teillaud, G.L. Carr, G.P. Williams, 'Chemical Imaging of Nucleic Acids, Proteins and Lipids of a Single Living Cell. Application of Synchrotron Infrared Microspectrometry in Cell Biology', *Cell. Mol. Biol.*, **44**(1), 9–13 (1998).
49. L.M. Miller, C.S. Carlson, G.L. Carr, M.R. Chance, 'A Method for Examing the Chemical Basis for Bone Disease: Synchrotron Infrared Microspectroscopy', *Cell. Mol. Biol.*, **44**(1), 117–127 (1998).
50. C. Paluszkiewicz, W.M. Kwiatek, M. Galka, D. Sobieraj, E. Wentrup-Byrne, 'FT-Raman, FT-IR Spectroscopy and PIXE Analysis Applied to Gallstones Specimens', *Cell. Mol. Biol.*, **44**(1), 65–73 (1998).
51. E. Wentrup-Byrne, L. Rintoul, 'Vibrational Spectroscopy of Human Gallstones', *Proceedings of 6th European Conference on Spectroscopy of Biological Molecules*, eds. J.C. Merlin, S. Turrell, J.P. Huvenne, Kluwer, Dordrecht, 481–482, 1995.
52. I.I. Sheleketina, L.P. Averyanova, U. Dnepropetrovsk, 'Ratio of Bile Components Causing its Precipitation and Solubilization of Cholesterol as a Gallstone Differentiation Factor in Cholelithiasis', *Dopov. Nats. Akad. Nauk Ukr.*, **2**, 176–183 (1997).

53. B. Shabsoug, M. Samrah, N. Bashir, F. Ammari, 'Qualitative Analysis of Gallstones by Infrared and Wet Chemical Methods', *J. Chem. Soc. Pak.*, **19**(2), 149–152 (1997).
54. O. Bekers, C. Postma, A.J.P.F. Lombarts, 'Determination of Fecal Fat by Near-infrared Spectroscopy', *Eur. J. Clin. Chem. Clin. Biochem.*, **33**(2), 83–6 (1995).
55. V. Neumeister, J. Henker, C. Sproessig, W. Jaross, 'Near-infrared Spectroscopic Determination of Fecal Fat', *Klin. Labor*, **40**(6), 523–7 (1994).
56. M. di Renzo, E. Sacher, T.H. Ellis, I. Stangel, 'Investigation of Calcified Tissues by Photoacoustic IR Spectroscopy', *J. Phys. IV*, **7**(C6), 221–228 (1997).
57. J. Fritsche, H. Steinhart, M.M. Mossoba, M.P. Yurawecz, N. Sehat, Y. Ku, 'Rapid Determination of *trans*-Fatty Acids in Human Adipose Tissue. Comparison of Attenuated Total Reflection Infrared Spectroscopy and Gas Chromatography', *J. Chromatogr. B*, **705**(2), 177–182 (1998).
58. R.A. Shaw, H.H. Eysel, K.-Z. Liu, H.H. Mantsch, 'Infrared Spectroscopic Analysis of Biomedical Specimens Using Glass Substrates', *Anal. Biochem.*, **259**(2), 181–186 (1998).
59. H.M. Heise, A. Bittner, 'Multivariate Calibration for Near-infrared Spectroscopic Assays of Blood Substrates in Human Plasma Based on Variable Selection Using PLS-regression Vector Choices', *Fresenius' J. Anal. Chem.*, **362**(1), 141–147 (1998).
60. K. Murayama, K. Yamada, R. Tsenkova, Y. Wang, Y. Ozaki, 'Determination of Human Serum Albumin and γ-Globulin in a Control Serum Solution by Near-infrared Spectroscopy and Partial Least Squares Regression', *Fresenius' J. Anal. Chem.*, **362**(1), 155–161 (1998).
61. H. Yang, X. Gao, S. Cao, K. Zhang, 'Conformational Studies of Lipase in Nonaqueous Solvents by Fourier Transform Infrared Spectroscopy', *Wuxi Qinggong Daxue Xuebao*, **16**(2), 42–46 (1997).
62. D.A. Zarling, M.J. Rossi, N.A. Peppers, J. Kane, G.W. Faris, M.J. Dyer, S.Y. Ng, L.V. Schneider, 'Up-converting Reporters for Biological and Other Assays Using Laser Excitation Techniques', *US Pat.*, 5736410, A 980407, 1998.
63. C. Kuhne, G. Steiner, W.B. Fischer, R. Salzer, 'Surface Enhanced FTIR Spectroscopy on Membranes', *Fresenius' J. Anal. Chem.*, **360**(7–8), 750–754 (1998).
64. S. Van Huffel, 'Improved Quantification of Chromophore Concentrations in Near Infrared Spectroscopy Using Prior Knowledge', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 485–494 (1998).
65. T. Komura, 'Infection Urolithiasis. Urease Induced Crystallization in Synthetic Urine', *Nippon Hinyokika Gakkai Zasshi*, **89**(4), 477–483 (1998).
66. E.P. Paschalis, F. Betts, E. DiCarlo, R. Mendelsohn, A.L. Boskey, 'FTIR Microspectroscopic Analysis of Human Iiac Crest Biopsies From Untreated Osteoporotic Bone', *Calcif. Tissue Int.*, **61**(6), 487–492 (1997).
67. E.P. Paschalis, F. Betts, E. Dicarlo, R. Mendelsohn, A.L. Boskey, 'FTIR Microspectroscopic Analysis of Normal Human Cortical and Trabecular Bone', *Calcif. Tissue Int.*, **61**(6), 480–486 (1997).
68. Y. Gotshal, I. Adam, A. Katzir, 'Glucose Measurements in Solutions Using Fiberoptic Evanescent Wave Spectroscopy and Tunable CO_2 Laser', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3262**, 192–196 (1998).
69. S. Frisch, T. Keweloh, H.-H. Belz, H. Kijewski, 'Study of the Infiltration of Organic Molecules into Hairs and Hair Analysis', *Rechtsmed. Forschungsergeb.*, **16**, 13–29 (1997).
70. H.J. Kim, Y.A. Woo, S.H. Chang, C.H. Cho, K. Cantrell, E.H. Piepmeier, 'Fundamental Investigation of Noninvasive Determination of Glucose by Near Infrared Spectrometry', *Anal. Sci. Technol.*, **11**(1), 47–53 (1998).
71. J.J. Burmeister, H. Chung, M.A. Arnold, 'Phantoms for Noninvasive Blood Glucose Sensing with Near Infrared Transmission Spectroscopy', *Photochem. Photobiol.*, **67**(1), 50–55 (1998).
72. M. Takashima, Y. Suzuki, T. Ookura, 'Apparatus for Blood Sugar Determination', *Jpn. Kokai Tokkyo Koho*, JP 07043295, A2 950217 Heisei, 1995.
73. K. Maruo, K. Shimizu, M. Oka, 'Device for Non-invasive Determination of Glucose Concentration in Blood', *Eur. Pat. Appl.*, EP 843986, A2 980527, 1998.
74. H.H. Mantsch, M. Jackson (eds.), *Proceedings of Infrared Spectroscopy: New Tool in Medicine, Proceedings of SPIE*, SPIE, Bellingham, WA, Vol. 3257, 1998.
75. Y. Ozaki, H. Maeda, 'Nondestructive Analysis by Near-infrared Spectroscopy', *Nippon Sekigaisen Gakkaishi*, **5**(2), 78–88 (1995).
76. L. Papenkordt, U.A. Mueller, J. Schueler, K. Danzer, K.-U. Jagemann, C. Fischbacher, 'Method and Apparatus for Noninvasive Transcutaneous Chemical Analysis in Human Body Fluids or Tissues', *Ger. Offen.*, DE 4339067, A1 950518, 1995.
77. M.A. Arnold, J.J. Burmeister, G.W. Small, 'Phantom Glucose Calibration Models from Simulated Noninvasive Human Near-infrared Spectra', *Anal. Chem.*, **70**(9), 1773–1781 (1998).
78. H. Schollermann, 'Spectrophotometric Method for Noninvasive Determination of Oxygen Saturation During Blood Circulation in Tissues', *PCT Int. Appl.*, WO 9804903, A1 980205, 1998.
79. R. Niesner, G. Spanner, 'Method and Device for Measuring an Analyte in a Sample', *Ger. Pat.*, DE 4446390, C1 960704, 1996.
80. S. Homma, A. Kagaya, 'Detection of Oxygen Consumption in Different Forearm Muscles During Handgrip Exercise by Spatially Resolved NIR Spectroscopy', *Adv. Exp. Med. Biol.*, **428**, 327–332l (1997).
81. P.G. Cooper, G.J. Wilson, D.T.A. Hardman, O. Kawaguchi, Y.-F. Huang, A. Martinez-Coll, R.A. Carrington, E. Puchert, R. Crameri, C. Horamand, S.N. Hunyor, 'Blood Oxygen Desaturation Heterogeneity During

Muscle Contraction Recorded by Near Infrared Spectroscopy', *Adv. Exp. Med. Biol.*, **428**, 285–292 (1997).
82. T. Hamaoka, M. Mizuno, T. Osada, A. Ratkevicius, A.N. Nielsen, Y. Nakagawa, T. Katsumura, T. Shimomitsu, B. Quistorff, 'Change in Oxygenation and Phosphocreatine During Exercise and Recovery in Relation to Fiber Types and Capillary Supply in Human Skeletal Muscle', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 478–484 (1998).
83. M.C. Van Der Sluijs, W.N.J.M. Colier, R.J.F. Houston, B. Oeseburg, 'A New and Highly Sensitive Continuous Wave Near Infrared Spectrophotometer with Multiple Detectors', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 63–72 (1998).
84. E. Chihara, T. Shiga, K. Tanabe, Y. Tanaka, 'Higher Temporal Resolution is Necessary for Continuous Wave Near Infrared Spectrophotometric Monitor in Both Cerebral and Muscular Tissue Oximetry', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 83–91 (1998).
85. M. Pagani, R. Ansjon, F. Lind, C. Jonsson, J. Uusijarvi, G. Sumen, H. Jacobsson, S.A. Larsson, 'Comparison Between Near Infrared Oximetry and ^{99m}Tc-HMPAO Uptake in the Resting Peripheral Muscle Under Normobaric Normoxia and Hypbaric Hypoxia', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 151–158 (1998).
86. W. Bank, J. Park, G. Lech, B. Chance, 'Near-infrared Spectroscopy in the Diagnosis of Mitochondrial Disorders', *BioFactors*, **7**(3), 243–245 (1998).
87. F. Steinberg, H.J. Rohrborn, T. Otto, K.M. Scheufler, C. Streffer, 'NIR Reflection Measurements of Hemoglobin and Cytochrome aa_3 in Healthy Tissue and Tumors. Correlations to Oxygen Consumption: Preclinical and Clinical Data', *Adv. Exp. Med. Biol.*, **428**, 69–77 (1997).
88. T. Kusaka, K. Isobe, K. Kawafa, Y. Ohtaki, S. Itoh, K. Hirao, S. Onishi, 'Postnatal Changes in the Cerebral Oxygenation in Normal and Asphyxiated Neonates', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 92–102 (1998).
89. S.L. Schwartz, P. Witorsch, R.R. Bies, S.W. Collier, 'Transdermal Chemical Monitoring Device', *PCT Int. Appl.*, WO 9814768, A2 980409, 1998.
90. L. Salganicoff, 'Near-infrared Spectrophotometry', *BioFactors*, **7**(3), 239–242 (1998).
91. T. Shiga, E. Chihara, K. Tanabe, Y. Tanaka, K. Yamamoto, 'Real Time Monitoring of Pulsatile Change in Hemoglobin Concentration of Cerebral Tissue by a Portable Tissue Oximeter with a 10 Hz Sampling Rate', *Proc. SPIE–Int. Soc. Opt. Eng.*, **3194**, 435–442 (1998).
92. P. Madsen, F. Pott, S.B. Olsen, H.B. Nielsen, I. Burcev, N.H. Secher, 'Near-infrared Spectrophotometry Determined Brain Oxygenation During Fainting', *Acta Physiol. Scand.*, **162**(4), 501–507 (1998).
93. K.A. Oberg, A.L. Fink, 'Methods for Collecting and Analyzing Attenuated Total Reflectance FTIR Spectra of Proteins in Solution', in *Techniques in Protein Chemistry VI, Pap. Symp. Protein Soc., 8th, 1994*, 475–484, 1995.
94. Y. Hoshi, S. Kosaka, Y. Xie, S. Kohri, M. Tamura, 'Relationship Between Fluctuations in the Cerebral Hemoglobin Oxygenation State and Neuronal Activity Under Resting Conditions in Man', *Neurosci. Lett.*, **245**(3), 147–150 (1998).
95. T. Binzoni, V. Quaresima, E. Hiltbrand, L. Gurke, P. Cerretelli, M. Ferrari, 'Influence of Repeated Ischemia/Reperfusion Cycles (Ischemic Preconditioning) on Human Calf Energy Metabolism by Simultaneous Near Infrared Spectroscopy, ^{31}P-NMR and ^{23}Na-NMR Measurements', *Adv. Exp. Med. Biol.*, **428**, 533–538 (1997).
96. C.-S. Lai, 'Methods and Articles for the Detection of Nitric Oxide in Fluid Media Using Semipermeable Membrane Bags Containing Nitric Oxide-trapping Agents', *PCT Int. Appl.*, WO 9820336, A1 980514, 1998.
97. S. Zhang, B.R. Soller, R.H. Micheels, 'Partial Least-squares Modeling of Near-infrared Reflectance Data for Noninvasive In Vivo Determination of Deep-tissue pH', *Appl. Spectrosc.*, **52**(3), 401–406 (1998).
98. M.K. Alam, J.E. Franke, T.M. Niemczyk, J.D. Maynard, M.R. Rohrscheib, M.R. Robinson, R.P. Eaton, 'Characterization of pH Variation in Lysed Blood by Near-infrared Spectroscopy', *Appl. Spectrosc.*, **52**(3), 393–399 (1998).
99. H.-G. Batz, 'Monitoring System for the Regular Intake of a Medicament', *PCT Int. Appl.*, WO 9807364, A1 980226, 1998.
100. N.G. Nerella, J.K. Drennen, 'Depth-resolved Near-infrared Spectroscopy', *Appl. Spectrosc.*, **50**(2), 285–91 (1996).
101. K.S. Kalasinsky, 'Drug Distribution in Human Hair By Infrared Microscopy', *Cell. Mol. Biol.*, **44**(1), 81–87 (1998).
102. D.E. Katzman, 'Breath Test for Detection of Drug Metabolism', *PCT Int. Appl.*, WO 9829728, A1 980709, 1998.
103. W.L. Webber, F. Lago, C. Thanos, E. Mathiowitz, 'Characterization of Soluble Salt-loaded, Degradable PLGA Films and Their Release of Tetracycline', *J. Biomed. Mater. Res.*, **41**(1), 18–29 (1998).
104. K.-Z. Liu, C.P. Schultz, R.M. Mohammad, A.M. Al-Katib, J.B. Johnston, H.H. Mantsch, 'Similarities Between the Sensitivity to 2-Chlorodeoxyadenosine of Lymphocytes from CLL Patients and Bryostatin 1-treated WSU-CLL cells: an Infrared Spectroscopic Study', *Cancer Lett. (Shannon, Irel.)*, **127**(1,2), 185–193 (1998).
105. R.M. Mohammad, K. Katato, V.P. Almatchy, N. Wall, K.-Z. Liu, C.P. Schultz, H.H. Mantsch, M. Varterasian, A.M. Al-Katib, 'Sequential Treatment of Human Chronic Lymphocytic Leukemia with Bryostatin 1 Followed by 2-Chlorodexyadenosine: Preclinical Studies', *Clin. Cancer Res.*, **4**(2), 445–453 (1998).
106. K.Z. Liu, C.P. Schultz, J.B. Johnston, K. Lee, H.H. Mantsch, 'Comparison of Infrared Spectra of CCL Cells with Their Ex Vivo Sensitivity (MTT assay) to Chlorambucil

and Cladribine', *Leuk. Res.*, **21**(11/12), 1125–1133 (1997).

107. F. Mosca, M. Bray, M. Lattanzio, M. Fumagalli, C. Tosetto, 'Comparative Evaluation of the Effects of Indomethacin and Ibuprofen on Cerebral Perfusion and Oxygenation in Preterm Infants with Patient Ductus Arteriosus', *J. Pediatr.*, **131**(4), 549–554 (1997).
108. J. Anastassopoulou, V. Korbaki, H. Nastou, M. Palta, V. Sarrou, 'FT-IR Studies of Propofol–Magnesium Interactions', in *Magnesium: Curr. Status New Dev., Proceedings of 8th International Symposium on Magnesium*, eds. T. Theophanides, J. Anastassopoulou, Kluwer, Dordrecht, 41–42, 1997.
109. J.J. Wang, C.-W. Chi, S.-Y Lin, Y.-T. Chern, 'Conformational Changes in Gastric Carcinoma Cell Membrane Protein Correlated to Cell Viability After Treatment with Adamantyl Maleimide', *Anticancer Res.*, **17**(5A), 3473–3477 (1997).
110. H.C. Shin, B. Stuart, E.F. McFarlane, 'Conformation of an Antigenic Determinant for Experimental Autoimmune Neuritis', *Biophys. Res. Commun.*, **224**, 5–9 (1996).
111. J.N. Whitaker, M.A. Moscarello, P.K. Herman, R.M. Epland, W.K. Surewicz, 'Conformational Correlates of the Epitopes of Human Myelin Basic Protein Peptide 80–89', *J. Neurochem.*, **55**(2), 568–576 (1990).

Laboratory Instruments in Clinical Chemistry, Principles of

Hemant B. Karnik
Howard University, Washington D.C., USA

1 Introduction 1
2 Historical Perspective 2
2.1 Spectral Instruments 2
2.2 History of Electrochemical Methods 3
2.3 History of Separation Techniques 3
2.4 History of Automation 4
3 Sample Requirement and Preparation for Analysis 4
4 Discussion of Principles of Instruments and their Major Components 5
4.1 Instruments Based on Molecular Absorption Spectroscopy 5
4.2 Performance Parameters for Spectrophotometric Techniques and Wavelength Calibration 6
4.3 Reflectance Photometry 6
4.4 Nephelometry 6
4.5 Atomic Absorption Spectrophotometry 6
4.6 Emission Flame Photometry 7
4.7 Use of Internal Standard in Flame Photometry 7
4.8 Molecular Emission Spectroscopy–Fluorimetry 8
4.9 Electrochemistry 8
4.10 Electrophoresis 9
4.11 Capillary Electrophoresis 10
4.12 Isoelectric Focusing 10
4.13 High-performance Liquid Chromatography 10
4.14 Gas Chromatography 10
4.15 Osmometry 11
4.16 Solid Scintillation Counters 11
5 Advantages and/or Limitations of Instruments and/or Methodologies 11
6 Calibration/Standardization of Instruments 12
7 Automation in Clinical Chemistry Laboratory 12
8 Quality Control and Quality Assurance 13
9 Future Developments of Instruments 14
Acknowledgments 14
Abbreviations and Acronyms 14
Related Articles 14
References 15

In the last 40–45 years, instruments have evolved from simple analyzers that analyze few analytes in an hour to auto analyzers that analyze hundreds of tests in a fraction of an hour. A historical perspective on the evolution of the instruments, with discussion on the basic principles of their operation, along with the advantages and limitations of methodologies, is provided. Information on standardization, quality control and quality assurance in the clinical laboratory is also included. The article concludes with a brief discussion on the future developments in instrumentation. The material presented in this article is meant as a reference. For detailed information the reader is referred to other books and publications.

1 INTRODUCTION

In the last 40–45 years there has been tremendous evolution in clinical laboratory instruments. Instruments have evolved from simple analyzers that would analyze one or two analytes in an hour to autoanalyzers that analyze hundreds of tests in a fraction of an hour. This evolution has been brought about by greater utilization and dependence on clinical laboratories in the treatment of disease. Clinical laboratory instruments have evolved to keep pace with greater demand and reliance on laboratory data. Not only has there been an increase in the request for routine tests, but with the advance of medicine there has been an increase in the menu of clinical diagnostic tests available. Many of these changes have been brought about by changes in the health care industry. With managed care and cost reduction the face of clinical laboratory is changing rapidly. Rapid changes are also taking place in industry that manufacture these instruments and autoanalyzers. Acquisitions and mergers of companies have become common and therefore names of instruments and companies that manufacture them may have changed from the ones that laboratory professionals have been familiar with. Much of the work done in the clinical laboratory has been in the area of clinical chemistry; therefore the clinical chemistry section of the laboratory has seen the most rapid growth in instrumentation and automation.

An overall perspective on the principles of laboratory instruments is presented here. The article begins with

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

a historical perspective. It is interesting to note that although a major evolution of laboratory instruments has taken place in the last 40–45 years, rudimentary instrumentation in some form has existed for over a century. The historical perspective section ends with the current status of instruments in the laboratory. The next section covers the basic principles and the major components of instruments. The instruments discussed range from spectral methods to those that employ nuclear radiation, with advantages and or limitations discussed in the following section. The other topics that follow are standardization, quality control and assurance. The article ends with a discussion on future developments in instrumentation.

It is hoped that the information provided here on the principles of clinical laboratory instrument in clinical chemistry will be found useful by all professionals working, not only in the clinical laboratory setting but also by academics in the teaching profession and other professionals in physicians' office laboratories. The information is given in a simple, concise manner without resorting to details which would be beyond the intended scope. The material presented here is meant as a reference. For detailed information the reader is referred to other books and publications.

2 HISTORICAL PERSPECTIVE

The field of clinical chemistry involves analysis of chemicals present in body fluids and tissues to diagnose disease. The field may date back more than 2000 years when the Hindu physician Sasruta reported the presence of ants around urine, an indication of the presence of sugar in urine. During that period diagnosis was based on inferences from observation.

This field of clinical chemistry, a part of diagnostic laboratory medicine, has evolved not only from various branches of chemistry, especially analytical chemistry; but also from other areas of sciences, such as physics, mathematics, physiology and biology. Therefore a comprehensive historical perspective would truly involve historical landmarks of various branches of science, which would be beyond the scope of this section. Nevertheless certain salient contributions specific to the field are worth noting. Historical perspective, provides us with an insight of the past and the future directions we are headed in.

2.1 Spectral Instruments

The earliest clinical laboratory instruments were based on the measurement of color formed when a body fluid under examination was made to react with a reagent. The intensity of color formed was proportional to the quantity of chemical in the examination fluid. These instruments were named colorimeters based on their ability to compare or measure color with that of similar solution of known strength. A working manual of clinical pathology by Todd named *Clinical Diagnosis by Laboratory Methods* was first published in 1908 and underwent several revisions. It describes the colorimeteric methods and the early colorimeter instruments in the July 1925 edition.[1]

During that period the Duboscq type of colorimeter, the Kober and certain others manufactured in the USA were considered satisfactory. Color match between the unknown solution and the standard solution was achieved by either raising or lowering the glass bottomed cups of solution in which glass plungers were immersed to observe the color match viewed through a single field. Other types of colorimeter that were used during that period and are mentioned in the July 1925 edition are the Bock–Benedict, the Hellige and the Denison colorimeters working on general principles similar to that of the Duboscq. All of these instruments required good eyesight for observation.

The colorimeters of the late nineteenth century underwent several evolutions and modifications brought about by availability of new tests and techniques. During the 1940–1950 period many of these tests were conducted at the laboratories of Otto Folin, Michael Somogyi and D.D. Van Slyke using simple photometers. These simple photometers are today's sophisticated spectrophotometers that employ the principle of light in the measurement of the body fluid analytes. In 1940 Arnold Beckman contributed to the development of the DU spectrophotometer that was introduced at that time. The Beckman model DU-2 ultraviolet (UV) spectrophotometer was an extensively revised version of the pioneer model DU. This instrument was regarded over the years as the authoritative instrument in the rapidly expanding field of UV spectrophotometry. It facilitated such studies as fluorescence, reflectance, flame emission, reaction rate, enzyme kinetics, microanalysis and column chromatography. This was made possible by a wide selection of accessories for the instrument.[2] Other clinical laboratory spectrophotometers were developed between late 1950 and 1960. The Gilford Stasar III evolved from Gilford 300N that was introduced in the late 1960s. The Bausch and Lomb Spectronic 100 was introduced in 1970. The earlier model was Spectronic 21.[3] Nephelometers were developed later in the 1960s.

The atomic absorption (AA) spectrophotometer, used in the analysis of trace metal, was introduced in the middle to late 1950. Its application has grown steadily since then. Perkin-Elmer Corp. was among the first to offer microprocessor-based analytical instruments. In 1975, Perkin-Elmer introduced the model 460 AA

spectrophotometer. In 1978 Perkin-Elmer pioneered the infrared Data Station concept, and in 1981 the Laboratory Information Management Systems LIMS/2000 series was introduced.[4]

In the middle to late 1950s other instruments were developed by companies based on emission of radiant energy. The Coleman flame photometer model 21 developed by the Coleman Division of the Perkin-Elmer Corporation was common in many clinical laboratories. This model was later replaced by model 51, with automatic igniter. Instrumentation Laboratories introduced model 143 that was later replaced by an improved version model 343. Other flame photometers were the Turner model 510 and the Kina Flame System by Beckman Instruments Incorporation. Although the phenomenon of fluorescence, phosphorescence and chemiluminescence was discovered earlier, fluorometers were only introduced in the middle to late 1950s.The earlier fluorometer models were Coleman 12C by Perkin-Elmer, and Turner model 111 by G.K. Turner Associates. Many of the earlier versions were bulky and had some limitations and were later modified or replaced with more compact models.

2.2 History of Electrochemical Methods

The need for more compact and maintenance-free instruments that were also very sensitive and specific for analysis of microanalytes revolutionized the field of clinical laboratory investigation. Advances in the field of microchemistry are credited to Samuel Natelson. Potentiometric measurements using glass electrodes were applied to pH determination prior to 1920.[5] In 1905 Cremer discovered that two solutions of different hydrogen concentration separated by a thin glass membrane produced a potential that was proportional to the difference in hydrogen ion concentration on either side of the membrane. Cremer's discovery led to the development of the first glass electrode by Haber and Klemensiewicz in 1909.[6] One of the first pH meters developed for laboratory use was the Beckman model G pH meter. Other companies on the market were Orion Research Inc., Radiometer, Instrument Laboratories, Corning Scientific Instruments, and Coleman Instruments Division, Perkin-Elmer. The development of ion-selective electrode (ISE) technology is due mainly to advancements made in the area of potentiometry and electronics. With the development of microprocessors, and further developments in potentiometry, ISE technology became an established procedure in the clinical laboratory for electrolyte testing. The focus shifted from measuring ion concentration to measuring ion activity and many different electrodes began to be developed for measuring various analytes including blood gases and enzymes. Stow and Randall[7] were the first to record measurement of the partial pressure of carbon dioxide, P_{CO_2}, by P_{CO_2} electrode. The present-day P_{CO_2} electrode is developed from the earlier work of Severinghaus and Bradley.[8]

In the late 1950s, Poul Astrup contributed to the measurement of blood gases and pH using capillary electrodes. Later, Beckman Instruments introduced a blood pH measuring system. In the 1970s measurements of sodium, potassium, chloride, and ionized calcium by ISE became routine laboratory procedures. ISE methodology was applied to the determination of many electrolytes and other analytes. In fact the field of electrochemistry had began much earlier in the nineteenth century. Michael Faraday's contribution – Faraday's laws of electrolysis – has a direct application to coulometry. The measurement of chlorides in biological samples is an extension of the electrolyte measurement methods developed by Zall et al.[9] Later, Cotlove[10] applied the principle of coulometry to the determination of chloride, using a silver electrode. In the 1960s the principles of polarography and amperometry were applied to developing electrodes that could sense oxygen consumption. The Beckman Automated Stat/Routine Analyzer (ASTRA) Systems employed the oxygen sensing electrode for glucose determination. This P_{O_2} polarographic electrode, also known as the Clark electrode, is used in the recent Beckman ASTRA CX 7 and 9 systems.

2.3 History of Separation Techniques

The separation techniques of chromatography and electrophoresis date back to the early twentieth century when the Russian botanist, Tswett, in 1906 was working on separating plant pigments, thus yielding the name chromatography for the separation of the pigments. In 1952 James and Martin developed gas–liquid chromatography (GLC). In 1965 Cawley,[11] published the ''Principles of Chromatography'' in a manual for a workshop in chromatography. In the early 1970s advances in chromatography and microprocessor-aided computer technology have led to the more efficient technique of high-performance liquid chromatography (HPLC). Several HPLC systems and detectors are available.[12] Although liquid chromatography (LC) was discovered before gas chromatography (GC), its use in the clinical laboratory was limited. With the development of small (10 μm) totally porous particles in the early 1970s, LC's applications for clinical laboratory use gained popularity over GC.

Tiselius in 1937 first devised the free solution or moving boundary method for the study of proteins by electrophoresis. Although the technique is still used in research, it is not used in clinical laboratories for routine work. Because of the limited resolution of serum proteins by paper, cellulose acetate and agarose gel, polyacrylamide gel electrophoresis (PAGE) is a

popular choice for the separation of serum proteins. Applications of electrophoresis for better resolution and sharper identification of protein patterns, have been seen in techniques including isoelectric focusing (IEF), and capillary electrophoresis (CE). In the 1970s Corning and Helena electrophoresis systems were quite popular. Later, Helena Laboratories developed the rapid electrophoresis system for automated electrophoresis. Based on the pioneering work of Hjerten, Jorgensen and Lukas, Mikkers, Terabe, and Cohen and Karger, different modes of capillary electrophoretic separations were developed.[13] In contrast to traditional electrophoresis, CE is well suited for automation.

Instruments based on the measurement of nuclear radiation date back to the period when Geiger-Müller counters were the first to be used in the clinical nuclear medicine laboratory in the late 1930s. Most of the instruments in a clinical chemistry laboratory employ scintillation detectors to measure γ-radiation, emitted by radioactive substances.

2.4 History of Automation

Automation in the clinical laboratory was the result of an increase in demand for diagnostic tests and as a result of development of microprocessors and computer-aided technology. During the 1950s, Leonard Skeggs of Technicon Inc. developed and introduced the continuous flow system for the analysis of routine chemistries. Initially these were single channel autoanalyzers that would perform one single test. By the mid 1960s multichannel autoanalyzers were available that would perform a number of profiles. Technicon introduced SMA 6/60 and SMA 12/60 sequential multiple analyzers that would perform six or 12 parameters on one sample and would process a number of samples in 1 hour. The evolution of this continuous flow technology from single and dual channel analyzers had a significant impact on clinical laboratory operations and became the yardstick for measuring the performance of any new instrumentation for the clinical laboratory. In 1981 Technicon introduced SMAC II, its top-of-the-range, high-volume chemistry analyzer, which succeeded the earlier model SMAC (sequential multiple analyzer with computer). The SMAC II instrument had the capability of analyzing 20 chemistries at the rate of 150 samples per hour. In the late 1960s due to rapid advancement in electronic solid-state transistors, printed circuit boards replaced vacuum tubes in instruments, and this revolutionized troubleshooting of instruments. At the first Oak Ridge conference in 1968, instruments based on centrifugal analysis were presented. In the 1970s, centrifugal analyzers assumed a pivotal role in the automated clinical laboratories. With improvements in on-board microprocessors and minicomputers, coupled with the development of new assay procedures, these instruments became increasingly more flexible in handling the growing volume and diversity of laboratory determinations.[14] Roto Chem by American Instruments, GEMSAEC by Electro-Nucleonics and CentrifiChem by Union Carbide were the early centrifugal analyzers on the market. Many of the earlier centrifugal analyzers were batch analyzers, that is they ran the same test or tests on the same sample in one run. Random access analyzers that were soon to follow in 1980s, provided more flexibility in sample testing from a menu of over 20 tests. Random access analyzers, however, utilized more time in programming the analyzer for particular tests requested for analysis. Later on, bar coding of the samples alleviated the problems. These analyzers had a very high output of tests per hour. Eastman Kodak introduced in 1981 the Kodak Ektachem system that had the capability of random access as well as discrete analysis using dry slide technology. From a menu of tests available on a cathode ray tube (CRT) monitor, tests were selected. Many hybrid autoanalyzers of random access and discrete analysis have been on the market since the 1980s: the Coulter DACOS, the Cooper DEMAND, the Hitachi 705, the Beckman ASTRA, the Du Pont Aca, the Roche Cobas Bio, Technicon RA-1000 are just a few examples.

3 SAMPLE REQUIREMENT AND PREPARATION FOR ANALYSIS

Collection and preparation of samples for analysis require careful attention to various factors. An important factor for consideration is the diagnostic laboratory test for which the sample is to be collected. The diagnostic test will determine the nature of the specimen, collection time, fasting or nonfasting state, and the nutritional status of the patient. Depending on the diagnostic test, the type of vacutainer tube or the collection system required should be considered. The diagnostic test may also determine length of time before a sample is analyzed. Preanalytical and analytical temperature of the specimen, exposure of specimen to light or air immediately after collection, the volume of specimen needed, disinfectant for venipuncture site, and the site of collection are some other factors that depend on the type of diagnostic test. For detailed reviews of biological and preanalytical variables refer to Young[15,16] and Ladenson.[17] Blood is the most common specimen analyzed in the clinical chemistry laboratory. Certain chemicals may be drastically affected by hemolysis. It is important that plasma or serum be separated from cells as soon as possible but certainly within 2 h.[18] Urine samples although common are not as frequently analyzed for chemistries as blood. At times

other body fluids, such as spinal fluids and gastric contents, may also be analyzed. Proper identification of the sample is crucial.

Winsten[19] describes the methods for collection and preservation of specimens for clinical chemistry. With the exception of glucose, triglycerides and inorganic phosphorus most blood constituents reveal no significant change after a standard breakfast, so it is not essential for the patient to be in an absolute fasting state prior to blood specimen collection.[20] Lipemia caused by chylomicrons from transient increase in triglycerides from meals containing fat however can interfere with many chemical analyses because of turbidity. Therefore, blood should be collected in the postabsorptive state, usually after an overnight fast of 12–14 h. This is critical for lipid analysis. The National Committee for Clinical Laboratory Standards (NCCLS) gives procedures for the collection of diagnostic blood specimens by venipuncture and provides further information for the collection of specimens.[21] The proper procedure for collecting blood from infants is described in the NCCLS standard H4-A3.[22] The proper technique for arterial puncture is described in NCCLS standard H11-A2.[23] For further reading on specimen collection and processing refer to Young and Bermes.[24]

4 DISCUSSION OF PRINCIPLES OF INSTRUMENTS AND THEIR MAJOR COMPONENTS

4.1 Instruments Based on Molecular Absorption Spectroscopy

Instruments that employ radiant energy in the measurement of analyte are based on either the absorbance, emission, reflectance or scattering of the radiant energy. These instruments employ visible light or UV light of the electromagnetic spectrum in the quantitation of the analyte. Instruments that employ the absorbance of radiant energy are colorimeters, photometers, spectrophotometers and UV spectrophotometers.

The term colorimeter refers to the measurement of colors. The term and the instrument are obsolete in the modern clinical laboratory. These are simple instruments that lack specificity and sensitivity in measuring analyte. The instruments employ light from a tungsten filament lamp usually passing through a simple glass filter, for a chosen wavelength. The glass filters remove all other wavelengths from the white light of the tungsten filament lamp, except one particular band that is the color of the glass filter that it passes through. This band passes through the cuvette which holds the colored solution. Depending on the intensity of the colored solution, the band of light incident on the cuvette is absorbed and the rest is transmitted. The transmitted light falls on a photodetector, which converts the light into electrical energy. The output of the electrical energy is in proportion to the amount of light transmitted after absorbance by the colored solution in the cuvette. The electrical energy passes through a meter which makes a needle move on a scale of absorbance or transmittance readings, or the electrical energy may be converted to a digital readout of transmittance, absorbance or concentration.

The term photometer literally means measurement of light. In the clinical laboratory the term is used in a generic sense as an instrument that measures light. The instrument is similar to a colorimeter and works on the same principle.

Spectrophotometers are instruments that utilize a single narrow band of the spectrum of white light to pass through a solution in a cuvette. The solution in the cuvette is the one in which the analyte from the sample to be analyzed has reacted with a reagent. The main components of a spectrophotometer are a tungsten filament lamp as the light source, entrance and exit slits with associated optics, a monochromator, the detector, and the meter or readout device and the printer or the recorder. The quality of a spectrophotometer depends chiefly on the monochromator and the detector. The monochromator is a device that isolates a single narrow band of wavelength from the spectrum of white light of the tungsten filament lamp. The narrower the band of wavelength the better is the quality of the instrument. In the specifications of instruments this is given as the bandwidth or bandpass. The quality of the spectrophotometer, and therefore its cost, depends on the type of monochromator. Monochromators are usually prisms or gratings. Simple glass filters or interference glass filters are usually found in photometers or colorimeters. High-quality prisms are used in very sophisticated models of spectrophotometers. The quality of prisms adds to the specificity and the sensitivity of the instrument and therefore to its cost. Grating monochromators provide better resolution of the wavelength than prisms, but this depends on the blaze angle of the grating. The detectors are phototubes or photomultiplier tubes that convert the light energy into electrical energy. A phototube consists of a photosensitive cathode on which the transmitted light from the cuvette falls. When the transmitted light from the cuvette strikes the photosensitive cathode, electrons are ejected which are then captured on the anode. Thus, light is converted into electrical energy. Spectrophotometers are designed as either single or double beam instruments. Single beam spectrophotometers have a single optical path from the light source to the detector. In the double beam instruments, light passes through the sample as well as a reagent or water blank (reference beam). This compensates for line voltage fluctuation or any other variation in light source intensity which may otherwise affect

the readings. The double beam spectrophotometers are either with single detectors or double detectors. The ones with single detectors (double-beam-in-time) are equipped with either a rotating chopper or a mirror that alternately focuses the sample and the reference beam on the detector. The ones with two detectors (double-beam-in-space) have the isolated beam from the monochromator split by half mirrors. The double beam spectrophotometers are equipped to continuously monitor absorbance against time (kinetic measurements) or absorbance over a range of wavelengths. The absorbance of the blank is automatically subtracted. Advances in technology have provided truly monochromatic light beams by means of tunable lasers, ultrafast scanning spectrophotometers and extremely sensitive photodiodes. The Beckman DU-70 scanning spectrophotometer is an example of a highly sophisticated model. It has a UV range as low as 200 nm with a fixed 2 nm bandpass. This computerized model has a high resolution graphic video display. The spectral data are displayed in transmittance and absorbance. In UV spectrophotometry, hydrogen and deuterium lamps are used to provide continuous spectra in the UV range. As UV light is strongly absorbed by lime glass, quartz cuvettes must be used.

4.2 Performance Parameters for Spectrophotometric Techniques and Wavelength Calibration

In quantitating the analyte, the absorbance of the unknown is usually compared to the absorbance of a standard (calibrator). Thus minor errors in wavelength calibration, variation in spectral bandwidths, presence of *stray light*, etc. do not usually contribute to serious errors. Using a series of standards (calibrators over a wide range) also provides a check for linearity, and therefore agreement with Beer's law for a particular procedure and instrument. When calculating the analyte, with previously determined or published values for molar absorptivities or absorption coefficients, e.g. in enzyme assays using NAD–NADH (nicotinamide adenine dinucleotide and nicotinamide adenine dinucleotide reduced form) reactions, the instrument needs to be checked more thoroughly for photolinearity and wavelength calibrations. Performance verification of spectrophotometers on a periodic basis also improves reliability of routine analysis. When wavelength calibrations are performed on narrow spectral bandwidth instruments, holmium oxide may be used as calibrator over a range of 280–650 nm. With broader-band pass instruments, a didymium filter may be used. A solution of potassium salt may also be used in performing wavelength calibration and photometric accuracy checks. For details regarding standards in absorption spectrometry see Burgess and Knowles.[25] Standard Reference Materials (SRM) available from the National Institute of Standards and Technology (NIST) for calibrating and verifying spectrophotometric and fluorometric measurements are published in the NIST SRM catalog.[26]

4.3 Reflectance Photometry

Reflectance photometry involves measurement of the reflected radiant energy. This technique is applied by Eastman Kodak to its Kodak Ektachem 700 series analyzer. Dry reagents are impregnated on to thin layers of plastic film and mounted in a slide. A drop of the specimen applied on the slide allows the analyte to react with the reagent to produce a color. This color can be measured by reflectance photometry. As with spectrophotometric methods, the color produced is proportional to the analyte present in the specimen. The autoanalyzer section details the instrumentation and slide technology.

4.4 Nephelometry

Nephelometry involves measurement of light scattered by particles. The amount of light measured is proportional to the amount of particles in solution. Simply, the instrument involves a lamp, optics to include a lens, the cuvette containing the sample and detector. The source of radiation may be either the helium–neon laser, quartz–halide lamp or the xenon lamp. Nephelometry is widely used in the measurement of antigen–antibody complexes. Another application is in the measurement of enzymes that act on particulate substrate. The Coleman amylase/lipase analyzer is a nephelometer that is used in the measurement of the enzymes amylase and lipase. When the enzymes act on the substrate the loss in turbidity can be measured. The loss in turbidity is in proportion to the enzyme present. Nephelometry has been applied in immunological work. For comparisons of features of various immunonephelometers see Nalesnik et al.[27]

4.5 Atomic Absorption Spectrophotometry

AA spectrophotometers are used in clinical laboratories for the measurement of metals such as calcium, magnesium, copper, lead, zinc, mercury and several other trace metals. The principle of this instrument involves maintaining the atoms of the element being analyzed in the ground state, so that these atoms are able to absorb light of a specific wavelength. This specific wavelength that the atoms can absorb corresponds to the wavelength the atoms would emit if they were excited. The amount of light absorbed is proportional to the analyte in question. The atoms of the analyte, however, need to be dissociated from their chemical bonds. This dissociation is achieved by heat of flame or furnace. A hollow cathode lamp made

from the metal in question is employed to produce a very specific wavelength that the atom of the element to be analyzed will absorb. The lamp is a neon or argon lamp with the cathode made from the metal in question. In disassociating of the atom from its chemical bonds, some emission occurs, but the percentage is small. With an efficient atomization technique, it is possible to maintain a large population of atoms in the ground state. The components of the instrument are a hollow cathode lamp, a chopper, an atomizer burner, a monochromator, a detector and a readout device. A hollow cathode lamp is available for many metallic elements. A hollow cathode lamp is adequate for most elements. For volatile elements, electrodeless discharge lamps are available. For determination of alkali metals (rubidium, cesium, sodium, potassium) and other elements (zinc, cadmium, mercury) a vapor discharge lamp is preferred.[28] For most metals, an air/acetylene flame which reaches a temperature of 2400 °C, is adequate for atomization. However, certain metals (aluminum, tin) require higher flame temperature which can be achieved by a nitrous oxide/acetylene flame. The atomization of elements that require high temperature can be achieved by raising the temperature of the flame, but it is likely that the population of neutral atoms available for absorption will be reduced because of ionization. Atomization of the element also depends on the design of the nebulizer/burner assembly. There are two types of burners, the premix and the direct consumption burner. Each has its advantages and disadvantages. One of the disadvantages of the premix nebulizer/burner design is the large sample size that is needed to compensate for dilution before aspiration and the fact that a small percentage (about 10%) of the aspirated sample reaches the flame. This limits the sensitivity of trace elements in biological fluids.[28] In this respect the direct consumption burner has an advantage in that it eliminates sample waste.

Although the direct consumption burner design eliminates the sample waste problems, the samples are still subject to dilution by flame gases thereby affecting the sensitivity of determination. The flameless systems avoid the dilution effects of the flame and also eliminate the separate sample preparation steps required in flame systems. Graphite systems have been widely used in place of flame systems. In the graphite system high temperatures can be attained to volatilize matrix constituents of the sample with minimum volatilization of the sample. As in spectrophotometers, both single beam and double beam designs are available. Matrix interference due to sample viscosity problems can be overcome by making the standards in the same matrix as that of the sample. Chemical interference occurs when thermally stable compounds do not readily disassociate into neutral atoms, as in the case of calcium and magnesium forming stable complexes with phosphate. In this case, excess lanthanum or strontium is added to tie up the phosphate and release calcium and magnesium for measurement. Analysis of serum or plasma is done after appropriate dilution with buffer in the flame system or without pretreatment in the graphite system. Metals bound to proteins can be released by treating the sample with trichloroacetic acid. The metal is then chelated in a protein-free filtrate by a chelating agent, and after adjustment of pH, extracted in an organic solvent such as methyl isobutyl ketone before measurement. AA has wide applications in the measurement of lead, cadmium, arsenic in toxicology and selenium, manganese, zinc and other trace elements in nutrition.

4.6 Emission Flame Photometry

The flame photometry technique of measuring sodium, potassium, calcium and lithium is being rapidly replaced by ISE technology. Many laboratories still use flame photometry for electrolyte determination. Flame photometry involves measurement of emission spectra produced when a metal is ignited in a flame. The spectra are characteristic of the element being measured. When an atom of an element absorbs energy, the outer electrons move to a higher orbit (unsteady state); the electrons return to the ground state and in so doing emit the absorbed energy as light. The light that is emitted is characteristic of the element in question. Flame photometry can only be applied to measure easily excitable atoms. The components of the instrument are the burner assembly, optics including the lens and exit slit, the monochromator and the detector. There is a monochromator and a detector for each analyte, i.e sodium has its own filter and detector, as does potassium. The detector is usually a photomultiplier tube to detect even the weak signals from the flame. The burner assembly, as seen in AA spectrophotometry, is the direct consumption or the premix assembly. The advantages or disadvantages of the two types of burner assembly are similar to the ones seen in AA spectrophotometry. The fuel content of the gas and the pressure and composition of the gas mixture are critical for the proper configuration of the flame. The blue inner cone portion of the flame is the hottest and atoms absorb energy in this zone. Light from the element is emitted in the outer portion of the flame. Compressed air used to support combustion may contain moisture and can pose a serious problem in the element being ignited.

4.7 Use of Internal Standard in Flame Photometry

In order to compensate for flame fluctuation, an internal standard, usually lithium, is aspirated along with the sample. The emission peak of the internal standard should be very different from the emission peak of the sample. This enables the two spectra to be distinguished readily by the detector. In cases where lithium is present in

the biological sample, another internal standard such as cesium should be used. The criteria for the internal standard are that its concentration should be the same in the sample being analyzed as in the calibration standard. The excitation potential of the internal standard should be similar to that of the analyte being measured. The emission spectra of the internal standard should be distinct from the emission spectra of the sample being analyzed. The internal standard should not be present in the biological fluid of the patient.

4.8 Molecular Emission Spectroscopy–Fluorimetry

Certain substances have the ability to absorb short waves (UV light) and convert them to longer wavelengths, or visible light; these substances are said to fluoresce and the phenomenon is known as fluorescence. Fluorimetric techniques not only provide increased analytical sensitivity but also have wider applications in flow cytometry and fluorescence microscopy. There are many compounds of biomedical importance that can be determined by fluorescent spectroscopy. As with spectrophotometers, fluorimeters are designed as single beam or double beam instruments. The main component of a single beam spectrophotometer is the light source which produces UV light. The optics include the lens, the primary filter and the secondary filter. The secondary filter is at right angles to the sample cuvette. The detector that is in line with the secondary filter is a photomultiplier tube. The light source is usually a mercury or xenon lamp; both lamps carry high intensity. Mercury lamps are not effective at wavelengths below 250 nm. Xenon lamps, although expensive, provide a wider range of wavelengths in the UV and visible regions. As UV light is absorbed by glass, the lamp envelope and other optics are of quartz. The primary filter allows only excitation energy (high energy or short waves) to fall on the sample cuvette. This enhances the specificity for the analyte in question by avoiding any other interfering substances from excitation. It also avoids photodecomposition of the analyte due to unwanted radiation falling on the sample. The secondary filter allows only the fluorescent or emitted, longer waves to fall on the detector. The detector serves the same purpose as in spectrophotometers. Fluorescent units are read off the dial or from a digital readout. The double beam fluorimeters have the same advantages as with other double beam design instruments.

Although fluorescent spectrophotometry is a very sensitive technique and can detect concentrations up to $1\,\mu g\,L^{-1}$ and is ideal to measure low concentrations of analytes, several factors need to be taken into consideration when working at very low concentrations: glassware can adsorb the analyte onto its surface and pose problems; quenching is another problem; and trace contaminants in solutions can suppress fluorescence. Also, it is important to adjust the pH of the solutions exactly, as slight changes in pH can affect the readings.

4.9 Electrochemistry

Electrochemistry has a wide application in industry. In a diagnostic clinical laboratory, electrochemistry involves the application of chemistry and electricity principles to the measurement of analytes. Based on a chemical reaction of an analyte an electrical signal is produced. This electrical signal can be determined, and can then be related to the concentration of the analyte. When a metal electrode is placed in a solution of its ions, a potential exists at the interface junction of the metal electrode and its solution. The solution containing the metal electrode may be considered as a half-cell. If this half-cell is connected by an electrical conductor to another half-cell of a constant potential, then the potential difference between the two half-cells can be measured and related to the concentration of the ions in the first cell. *Potentiometry*, that is the measurement of potential between two cells in which a chemical reaction takes place, utilizes this principle in measuring the concentrations of analytes. In practice, the electrode potential of one half-cell is kept at constant potential. The electrode potential at the other half-cell varies based on the concentration of ions in the cell. The electrode that has constant potential in one half-cell is called the reference electrode and the electrode in the other half-cell is known as the measuring or indicator electrode. The potential of a cell is equal to the potential of the reference electrode plus the potential of the indicator electrode plus the junction potential. The reference potential and the junction potential are constant, therefore the potential of the indicator electrode can be determined, and this can be related to the concentration of ions. This is described by the Nernst equation, which when simplified relates the cell potential to the concentration of ions. An electrode potential is also produced when varying concentrations of ions are separated by semipermeable membrane to the particular ion. Different indicator and reference electrodes are available. Reference electrodes are compared against the normal hydrogen electrode (NHE) with a potential of 0.00 V, and their relative potential is determined. One application of the potentiometric principle is the measurement of pH, i.e. the negative log of hydrogen ion concentration. Standardization of pH measurements has been reported by Wu et al.[29] Based on the principle of potentiometry, ISEs have also been developed.[30] These indicator electrodes at the tip have a thin glass membrane with a special composition that makes the electrode uniquely sensitive to the particular ion. The potential developed across the glass membrane (from the ions of the

analyte in the sample) is proportional to the concentration of the analyte. These indicator electrodes are now available for various elements such as sodium, potassium, calcium, ammonium and lithium. Indicator electrodes are also available to measure blood gases such as PO_2 and PCO_2. There are electrodes with tips composed of a silicone rubber matrix immobilized with a slightly soluble salt of the anion to be measured. These precipitate-impregnated membrane electrodes are developed to measure anions such as chlorides. On a similar principle, electrodes have been developed for bromine, cadmium, fluoride, etc. Besides being able to measure cations and anions by electrochemical means, certain nonion molecules are also being measured. Application of ISE is extended to measure such compounds, e.g glucose and blood urea nitrogen (BUN), using enzyme electrodes. Enzymes are impregnated in membranes that catalyze compounds to produce ions that are directly proportional to the analyte being determined. For example, enzyme urease is impregnated to act on urea from BUN. Urea is broken down to ammonium ions that are measured by a similar principle to that of the pH meter (change in hydrogen ion concentration). Besides potentiometric measurements, polarographic and amperometric measurements are also employed in electrochemistry. *Polarography* involves measurement of current flowing through an electrochemical cell and the potential between the two electrodes, when a constant electrical external voltage is applied to the cell. Application of ISE technology and Polarography are seen in Beckman ASTRA instruments. The Beckman ASTRA analyzer employs a polarographic electrode that responds to oxygen concentration. The rate of oxygen consumption as oxygen reacts with glucose, in the presence of enzyme glucose oxidase, is determined by the electrode. The rate of oxygen consumption is directly proportional to the glucose in the sample. The oxygen electrode is a polarographic electrode in that it measures a current limited by diffusion of oxygen through a Teflon membrane to the rhodium cathode, while the anode is held at a constant potential in reference to the cathode.

Amperometry involves measurement of current flowing through an electrochemical cell when a constant electrical potential is applied to the electrodes. Applications of amperometry are seen in the measurement of partial pressure of oxygen by the PO_2 electrode (Clark electrode). The oxygen that enters through a permeable membrane is reduced at the cathode. The current flow that results is directly proportional to the amount of oxygen. *Coulometry* is the measurement of coulombs or the amount of electricity flowing between the two electrodes in an electrochemical cell. The amount of electricity is directly proportional to the amount of substance produced or consumed by the oxidation/reduction reactions at the electrodes. Application of this technique has been seen in the Cotlove titrator for the measurement of chloride ions.[10]

4.10 Electrophoresis

A simple definition of electrophoresis is separation by electrical means. Charged particles are separated by electrical means. Although the term electrophoresis refers to separation of all charged particles, specific terms such as zone electrophoresis refer to separation of macromolecules such as proteins and lipoproteins for laboratory investigation. The macromolecules are separated using various supporting media such as cellulose acetate paper, polyacrylamide, agarose or starch gel. The choice of supporting media or matrix helps in the resolution of the separation. The equipment involves a power source to supply the electrical charge, a chamber to carry out the electrophoresis, and a densitometer to quantitate the separated molecules. The power supply is of two types, with either constant current or constant voltage. There may be advantages with the power supply with constant current, especially if the heat generated during electrophoresis increases, which causes the resistance to decrease, the voltage also decreases as seen by Ohm's law $V = I \times R$ (where V is voltage, I is current and R is resistance). Heat generated can affect the macromolecules by denaturing them. The resistance decreased due to heat may also affect the migration rate of the molecules. The power supply is connected to the electrophoretic chamber that contains the electrodes. The chamber is filled with the buffer, of appropriate pH and ionic concentration for conductance of electrical current. The level of buffer in the chamber is enough to cover the electrodes so as to complete the circuit after the supporting media with the sample applied is placed in the chamber. Parameters for current, voltage and time for electrophoresis are set according to the procedure and the procedure is completed. The separated molecules may be stained for proper identification of zones and quantitated using a densitometer. Densitometers are basically photometers with mechanisms for mounting the electrophoretic pattern that is made to move past a light beam on the opposite side of which is a detector. The recorder, if attached to the densitometer, converts the absorbances into 'peaks and valleys' or patterns which correspond to the density of the separated zones. Computerized systems are available that compute the results by converting the patterns into concentrations of the macromolecules present in the sample. With the advancing technology of microprocessors, the electrophoresis process has been automated by some manufacturers. Helena Laboratories has manufactured the REP system. This fully automated system performs electrophoresis from application of the sample to the

quantitation of results. Other leading manufacturers are Beckman and Ciba-Corning.

4.11 Capillary Electrophoresis

CE employs narrow bore (10–100 μm diameter) and 20–200 cm in length fused silica capillaries to separate, with greater efficiency, both large and small molecules for identification and studies. The high voltage (greater than 500 V) used in the technique generates electro-osmotic and electrophoretic flow of buffer solution and of ionic species in the capillary. This technique resembles a mix between PAGE and high-performance electrophoresis. Several advantages of this technique are noted: small sample volume; better resolution; easy automation and precise quantitation applicable to wider selection of analytes.

4.12 Isoelectric Focusing

In IEF, proteins are separated based not only on the charge they carry but also because they move through a medium that has a stable pH gradient. During the migration when the protein reaches a gradient that has pH equal to its isoelectric point, there is no net charge on the molecule. At this point the protein stops migrating and is separated out. The power source used in IEF is a high-voltage source, to enable the protein migration through the medium of high concentration of carrier ampholytes. Some modifications of this technique involve using various other media such as polyacrylamide gel that provides better resolution of proteins.

4.13 High-performance Liquid Chromatography

By definition chromatography is the separation of colored compounds. In the clinical laboratory, HPLC and GC have wider applications in toxicology and in therapeutic drug monitoring (TDM). HPLC is a type of LC, in which a mobile phase (liquid) is pressured to flow over a column (stationary phase). The mobile phase carries the sample through the column, during which components of the sample separate out based on their interaction with the stationary column packing. Each component is eluted out of the column and quantitated by the detection system. The components of HPLC are reservoir of solvent (mobile phase), pump, injection port, column, detector and readout device. The solvent reservoir is usually a bottle or a flask that contains the mobile phase. A solvent reservoir may be used if the pump is operated in the isocratic mode, or more than one reservoir used, if gradient mode is used in which the mobile-phase composition is changed over a period continuously by two or more solvents or in a stepwise fashion. The pump used is usually a reciprocating pump that provides a continuous flow of the solvent through the injector, over the stationary phase, the column and through the detection system. During the fill stroke the piston of the pump moves up the chamber. A certain volume of the mobile phase is drawn in the chamber through the inlet valve. In the pumping stroke, the inlet valve closes and the outlet valve opens and the piston moves down the chamber forcing the mobile phase or the solvent through the inlet port into the column. An aliquot of the sample is injected into the injection port with a microsyringe. Depending on the system, the sample may be held in a sample loop. In the inject mode the loop is rotated, causing the sample to be carried with the mobile phase into the column. A wide choice of columns is available. The column is usually made of stainless steel. There is a wide choice of packing materials. Silica is commonly used as a packing material. The detection system ranges from photometers, spectrophotometers and UV spectrophotometers to fluorometers and electrochemical systems. Various factors such as the composition of the mobile phase, the flow rate, the nature of packing material (stationary phase), the length of the column, etc. affect the resolution of components.

4.14 Gas Chromatography

Essentially the workings of a gas chromatograph are similar to a liquid chromatograph. In GC, the compound to be identified and quantitated is volatilized and carried by the mobile phase that is an inert carrier gas, through a column. In the column, constituents of the compound separate out and are identified or quantitated by a detection system similar to the one in LC. GC is of two major types: gas–solid chromatography (GSC) in which the packing material of the column is solid; and GLC where a nonvolatile liquid is coated on a solid column support.

The instrumentation involves the following:

- an oven with a temperature control;
- a chromatography column;
- inert gas, which is pressure regulated;
- a detection system and a recording system.

The prepared sample injected with a microsyringe through an injection port gets immediately volatilized due to the high temperature of the injection port and of the oven. The temperature of the oven is controlled so that constituents of the compound with the highest boiling point can be vaporized. The volatilized compound is carried by the inert gas through the column. The choice of inert gas depends on the detection system. Several detectors are available.[31,32] Nitrogen is usually used as the carrier gas with thermal conductivity and flame ionization detectors. Other gases used are argon,

helium and hydrogen. The columns may be made either of glass or stainless steel and are of various diameters and lengths. Usually 1/8 in. diameter columns are used. Capillary columns are about 0.8 mm in diameter and have no packing materials. The columns are of glass or fused silica. The inner walls of the columns act as the stationary phase. The narrower the diameter of the column the better the resolution of the constituents. Resolution of the constituents also depends on several other factors such as the nature of the packing material, the flow rate of the carrier gas and retention time. Diatomaceous earth is often used as packing material. Silica, alumina, and activated carbon are also used. The detectors in the clinical laboratories usually are thermal conductivity, flame ionization, or mass spectrometer type detectors. For details of theory and practice of chromatography refer to the textbook by Kaplan and Pesce.[33] For further reading refer to the book by Bender.[34]

4.15 Osmometry

Osmometry is the measurement of osmolality, that is the number of moles of particles of solute per kilogram of solvent. Osmolality is related to the osmotic pressure of the solution that regulates the movement of fluids between the cells and extracellular fluids. The instrument that measures osmolality is called an osmometer. Instruments that measure osmolality can theoretically be built on any of the colligative properties of solutions, which are increase in osmotic pressure, increase in boiling point, lowering of vapor pressure, and decrease in freezing point. The decrease in freezing point property has commonly been used to measure the osmolality of a solution as it is more practical. This property is based on the fact that the freezing point of water is lowered by 1.86 °C per molal of particle. The instrument involves an insulated coolant bath that holds an antifreeze solution such as ethylene glycol at −6 °C by a refrigeration unit. The sample (about 1 mL) such as serum or urine is placed in a small glass tube. The tube is placed in the instrument such that the stirrer and a thermistor dip in the sample. The thermistor is attached through a galvanometer to a read-out device. When the tube holding the sample is lowered into the coolant, the temperature of the sample starts to fall below the freezing point until it is a few degrees below the freezing point. At this point, the stirrer starts to vigorously stir the sample raising the temperature of the sample so that the sample forms a slush when the freezing point temperature is reached. The difference between 0 °C and the temperature when the slush is formed is the depression of the freezing point by the particle of the solute. This difference in temperature is related through the galvanometer to a read-out device that gives the reading in osmol kg^{-1}.

4.16 Solid Scintillation Counters

Solid scintillation counters (γ-ray counters) are instruments that are employed in the measurement of nuclear radiation. Nuclear radiation is emitted by an isotope of an atom when the ratio of neutrons to protons in its nucleus is unstable. In the process of stabilization of the nucleus the atom emits energy in the form of radiation. The isotope commonly employed in the clinical chemistry laboratory is iodine-125 and at times cobalt-57. Iodine-125 is in an excited state due to its unstable nucleus. A γ-ray is emitted when iodine-125 reverts to its ground state. Scintillation counters measure the γ-rays emitted by iodine-125 or cobalt-57.

Scintillation counters comprise of a crystal (sodium iodide–thallium) that is encased in a thin aluminum foil. This crystal, also known as phosphor, is flushed to a photomultiplier tube. Other components of the scintillation counter include a discriminator or a pulse height analyzer and a scaler. Processed samples that contain iodine-125 when placed in the well of the scintillation counter, either manually or by automation, emit γ-rays. The γ-rays strike the crystal that produces a flash (scintillation). This radiant energy is captured by the photomultiplier tube. The photomultiplier tube converts the light energy into electrical energy that may be amplified in proportion by an amplifier. The pulse height analyzer or the discriminator allows only energy of certain magnitude of the isotope to be measured; it sorts out pulses of certain heights. All voltage pulses below or above a certain magnitude are eliminated from reaching the scaler and thus from counting. Pulses are counted as counts per minute by the scaler.

5 ADVANTAGES AND/OR LIMITATIONS OF INSTRUMENTS AND/OR METHODOLOGIES

Numerous techniques, methods and instruments are available for the determination of an analyte. Therefore the decision on the choice of method, technique or instrumentation plays a critical role in the quality of the results. Sensitivity and specificity are two important criteria in the quantitation of an analyte. Sensitivity refers to the smallest concentration of an analyte that can be determined by a particular methodology or an instrument. Specificity refers to a particular analyte in question. Both sensitivity and specificity should not be compromised, and other factors such as turnaround time for the results of analysis and cost per analysis play an important role. Most methodologies, techniques, and instrumentation have certain advantages and limitations. The objective therefore in choice of instrumentation or methodology

is to utilize one with minimum limitations and optimum advantages.

Much of the chemical analysis in a clinical diagnostic laboratory is performed by spectral methods. In fact "spectrophotometry is the corner stone of analytical instrumentation".[35] With rapid advances in spectrophotometer designs, including truly monochromatic light beams by means of tunable lasers, fiber optics, ultra fast scanning spectrophotometers and extremely sensitive photodiodes, resolution has improved tremendously.[35] However, due to high sensitivity and accuracy there is a great need for performance checks and quality control.

Fluorimetric measurements have provided increased analytical sensitivity along with specificity. A major advantage of fluorimetry is that it is 100 to 1000 times more sensitive than absorption measurements. Much progress has been made in the development of instruments and methods.[36] Time-resolved fluorimeters have eliminated background interference and, by averaging signals, have improved precision. Detection limits of approximately $10-13\,\mathrm{mol\,L^{-1}}$ can be achieved. Fluorimetric measurements have their limitations. Fluorescence is affected by many variables such as pH, temperature, presence of quenching compounds and intensity of incident light.

In AA spectrophotometry, the two instrumental performance characteristics are sensitivity and detection limits. Sensitivity is a convention for defining the slope of absorbance versus concentration calibration for each element. For a flame AA it is expressed as the concentration of an element in micrograms per milliliter that will produce an absorbance of 1% absorption. In absorption, unit sensitivity is the micrograms of element per milliliter that will give an absorbance of 0.0044. For a graphite system, sensitivity is stated in terms of weight in picagrams of element that will produce 1% absorption. The detection limit is the lowest concentration that can be measured from the baseline noise. For a flame system it is the concentration that will produce a signal-to-noise ratio of 2 to 1. For a graphite system it is expressed in terms of weight of an element that will produce the same ratio.[28]

In the measurements of electrolytes by potentiometric methods, the results are reported not as activity of ions but activity multiplied by a certain factor. In this case the results reported by flame photometry and by potentiometric methods are not significantly different from each other in normal samples. However, major differences between the two methodologies can occur in certain pathological conditions when the mass water concentration in the sample is abnormal, and also for severe hyperlipemia or high protein concentrations.

6 CALIBRATION/STANDARDIZATION OF INSTRUMENTS

Calibration and standardization of clinical laboratory instruments are essential for validity of results. Calibration involves adjustment or manipulation of a component or components for correct output of the results. Standardization refers to analyzing a pure substance of a known concentration or a series of concentrations that can provide a calibration curve, against which an analyte of unknown concentration will later be compared. In practice, a standard may be analyzed, and as the recorder displays the results, the component of the instrument can be calibrated so that the recorder displays the exact concentration of the standard. In spectrophotometry, and other instrumentation, where absorbance of the unknown is compared with the absorbance of the standard for calculations of the results, minor errors in calibration of wavelength or other parameters are compensated since the unknown sample is being compared with a standard of known concentration. Using a series of standards over a wide range of concentrations, rather than a single standard, in addition provides a measure of photolinearity and agreement with Beer's law. However, when calculations of the results are based on certain constant values such as molar absorptivities or absorption coefficients then calibration of the instrument should be conducted with utmost care. In spectrophotometry, wavelength calibration is performed by use of solid filters such as holmium oxide or didymium. In verifying the band width of a spectrophotometer, the range of wavelength at one-half the peak transmission is taken into consideration. All spectral instruments are adjusted for baseline correction, either with a water blank or a reagent blank. In electrochemical methods, reference electrodes are calibrated against the hydrogen electrode. Measuring electrodes are calibrated against two different concentrations of standards to provide a calibration curve. The slope of the curve, along with the intercept, is then used by microprocessor to calculate the concentrations of the samples. In chromatography, standards are injected as pure analyte and, from the standard curve, the peak area of the unknown is compared with the standard to determine the unknown concentration.

7 AUTOMATION IN CLINICAL CHEMISTRY LABORATORY

Automation in the clinical chemistry laboratory began with the introduction of continuous flow single channel autoanalyzers by Technicon Corp. Skeggs introduced this system in the 1950s. These analyzers were called continuous flow because they employed peristaltic pumps and

plastic tubing to advance the samples and the reagents. Air bubbles separated the specimens and also cleaned the inner tube walls to minimize carry over of specimens. The incubation system included a mineral oil bath with a glass coil through which sample mixed with reagents advanced at a slow speed that allowed for incubation and color formation. The color formed during incubation was measured by a flow cell after air was debubbled. This system went through several modifications; SMAC (Sequential Multiple Analyzer Computerized) and CHEM 1+ were the modified versions. As in single channel autoanalyzers, automation mimics the steps of routine analysis, such as aspiration of the sample, addition and mixing of reagents, incubation of sample and reagent and measurement of color formation. It minimizes or even replaces human involvement in the process. Specimen identification and analysis of specimens are aided by bar codes. Autoanalyzers are classified based on their mode of operation. Continuous flow analyzers process samples sequentially, and each in the same manner. Each sample gets analyzed for all the tests available. Continuous flow analyzers have an advantage in a high-volume laboratory. The disadvantages of this system are that reagents are wasted if all the tests available are not requested. An example of continuous flow analyzer is the CHEM 1+ analyzer by Technicon. Batch analyzers as the term indicates analyze specimens sequentially in a run for a particular test at a time. Specimen analysis is performed from a carousel in discrete or individual cells or cuvettes. An example of batch analysis is COBAS-BIO that works on the principle of centrifugal analysis. Random access analyzers allow analysis of specimens in any order at any time. Tests may be performed as profiles or individually. Random access analyzers are selective and discrete. Discrete analysis allows each specimen to be analyzed in its own separate space. The main advantage of these analyzers is that they save reagents, facilitate Statin analysis and prevent any carry over of samples. Random access analyzers are very popular. Today's random access analyzers make the analytical process more integrated and efficient. These analyzers are built to increase productivity. Beckman's Synchron LX Clinical Systems, developed for a high-volume laboratory, offer random access, Statin and batch profiling of more than 80 analytes, including critical care tests, general chemistries, TDM, and drugs of abuse, thyroids and other esoterics (DATs). System operation is so automated that it requires an absolute minimum of a human interface performing up to 2880 tests per hour. Beckman's Synchron CX 9 ALX is also a random access analyzer, and offers a wide menu for analysis from general chemistry and urine chemistries to TDM and DATs. The random access analyzers provide access for a selectable number of tests. Reagents are available in a variety of ways: a pack in an automated clinical analyzer has a binary code; Paramax 720Z has different tablets; or reagent impregnated slides such as in VITROS. A wide choice of autoanalyzers is available from high volume to desktop. All of these analyzers now are interfaced and integrated with computers that control the process and handle the data. Standards for transmission of information between clinical instruments and Laboratory Information Systems have been developed.[37,38]

8 QUALITY CONTROL AND QUALITY ASSURANCE

Quality control and quality assurance refers to total quality management of the clinical laboratory. It refers to analyzing the overall performance of the laboratory in terms of laboratory results, consumer satisfaction (physicians and nurses), patient satisfaction, and laboratory staff satisfaction. Quality control implies a more limited aspect of the overall quality management. It refers to technical aspects, such as the methodology and the quality of laboratory data output based on statistical applications. US clinical laboratory regulations[39] require quality control procedures to be set for each clinical laboratory. Quality assurance implies a broader term guaranteeing overall quality of the laboratory performance. Quality control involves pre-analytical, analytical and post-analytical controls. Pre-analytical controls include such things as proper identification, collection and processing of patient samples. Analytical controls include selection and execution of methodologies, incubation temperature checks, calibrators, calibration/standardization of instruments and reference controls used in the procedures. Post-analytical checks are based on statistics and quality control charts. Selection of the correct methodology with accuracy and precision in its execution plays a very crucial role in quality control. Methods may be subject to random or systematic errors. Barnett and Youden[40] in 1970 first introduced a scheme for selecting and evaluating clinical laboratory methods. General discussion of the philosophy of method evaluation in relation to the overall quality assurance process has been developed by the International Federation of Clinical Chemistry (IFCC). Several articles on the topic of method evaluation have been published in the *Archives of Pathology and Laboratory Medicine* in 1992.[41] A textbook on the selection of methods has also been published.[40] Accuracy in a procedure is defined by IFCC as the closeness of the agreement between the measured value of an analyte and its true value. The true value is determined by using reference methods. Precision of a method refers to obtaining the same laboratory value on replicate determinations. When reference controls are

run, along with the patient specimens, using selected methodology, reference values are obtained. When these obtained reference values are compared against the range of values provided with the reference controls for that particular analyte, and if the values fall within the expected range, this attests to the validity of the methodology employed. A common method is to plot the obtained reference values on a chart each day against the day the values were determined. This chart has a mean known value for the analyte, with the upper and lower limits (range) set at certain standard deviations from the mean. When the obtained reference values fall within the range of standard deviation set, the ''run'' is said to be ''in control''. Such control charts are called Levey–Jennings charts. The interpretation of the control data in the laboratories is by implementing Westgard Multirule Chart. Several books and publications by Westgard are available in the area of Total Quality Management for additional reading.[42]

9 FUTURE DEVELOPMENTS OF INSTRUMENTS

At the December 1998 American Association for Clinical Chemistry meeting entitled Preparing for the Millennium: Laboratory Medicine in the 21st Century, topics such as what technologies will emerge? dominate? and other topics relating to the future of the clinical laboratory were discussed. From this and other information, it is evident that robotics will play a very important role in the future development of medium to high volume instruments. Interface and communication standards will be needed for controlling laboratory instruments by remote interface. There will be a trend towards modular automation, and automation will be guided by highly sophisticated software in the forms of CD ROM. Workstations to monitor and control several different clusters of autoanalyzers by a single operator will be available that will reduce the cost. The configuration of various workstations is described by Boyd et al.[43] Many of the autoanalyzers with bidirectional interfaces will be monitored, and controlled by the mainframe computer. Rapid growth in point of care testing (POCT), will demand smaller, operator friendly, and more easily portable instruments that employ dry chemistries or reagent cartridges. With rapid advances in microprocessor technology and miniaturization of electrodes, instruments employing electrochemical methods and maintenance-free electrodes will find greater applications in POCT. Hand-held analyzers, that utilize whole blood for glucose and cholesterol testing will be commonly used in POCT and in the home healthcare industry.

ACKNOWLEDGMENTS

I would like to thank Adam Holyoake, Jenny Cossham and the rest of the staff at John Wiley & Sons for giving me this opportunity and helping me to prepare this manuscript. My sincere thanks go also to Dr Robert Meyers, editor-in-chief and the rest of the editorial staff for their comments and suggestions.

My thanks also to Albert Akiyama of Beckman Coulter for directing me to the website of Beckman Coulter and providing me with some literature and brochures about the company's clinical analyzers. I also appreciate the efforts of Dr Ralph Ito of Olympus America in helping me to obtain information about the company's analyzers.

Last but not the least I would like to thank my wife Smita Karnik for typing and formatting the manuscript.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
ASTRA	Automated Stat/Routine Analyzer
BUN	Blood Urea Nitrogen
CE	Capillary Electrophoresis
CRT	Cathode Ray Tube
DATs	Drugs of Abuse, Thyroids and Other Esoterics
GC	Gas Chromatography
GLC	Gas–Liquid Chromatography
GSC	Gas–Solid Chromatography
HPLC	High-performance Liquid Chromatography
IEF	Isoelectric Focusing
IFCC	International Federation of Clinical Chemistry
ISE	Ion-selective Electrode
LC	Liquid Chromatography
NCCLS	National Committee for Clinical Laboratory Standards
NHE	Normal Hydrogen Electrode
NIST	National Institute of Standards and Technology
PAGE	Polyacrylamide Gel Electrophoresis
POCT	Point of Care Testing
SRM	Standard Reference Materials
TDM	Therapeutic Drug Monitoring
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Atomic Spectrometry in Clinical Chemistry • Automation in the Clinical Laboratory • Capillary Electrophoresis in Clinical Chemistry • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Electrolytes, Blood Gases, and Blood pH • Glucose Measurement • Point-of-care Testing • Statistical Quality Control in Clinical Laboratories • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis • Electrophoresis and Isoelectric Focusing in Food Analysis • Fluorescence Spectroscopy in Food Analysis • Liquid Chromatography in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Proteins and Glycoproteins • Chromatography of Membrane Proteins and Lipoproteins • Fluorescence Spectroscopy in Peptide and Protein Analysis • Gel Electrophoresis in Protein and Peptide Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Robotics and Laboratory Automation in Pharmaceuticals Analysis

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in

Atomic Spectroscopy **(Volume 11)**
Graphite Furnace Atomic Absorption Spectrometry

Electroanalytical Methods **(Volume 11)**
Liquid/Liquid Interfaces, Electrochemistry at

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography • Normal-phase Liquid Chromatography • Thin-layer Chromatography

REFERENCES

1. J.C. Todd, *Clinical Diagnosis by Laboratory Methods. A Working Manual of Clinical Pathology*, W.B. Saunders, Philadelphia, 114–119, 1925.
2. Model DU-2UV Spectrophotometer, Beckman Instrument Inc. Manual, Fullerton, 1964.
3. M.E. Parker, A.F. Krieg, L.M. Demers, 'Analytical Systems for Microchemistry II Instrumentation', *J. Clin. Lab. Automat.*, **1**(2), 101–112 (1981).
4. G.N. Kelly, 'Computer Aided Chemistry and the Role of Perkin Elmer', *Instrument News*, **31**(1), 2 (1981).
5. S. Ishii, O. Flores, T. Belisle, 'The Electrolyte 2 Analyzer', *J. Clin. Lab. Automat.*, **2**(5), 338 (1982).
6. G. Widdowson, 'Electrolyte Technology in Clinical Chemistry Today and Tomorrow', *Lab. Med.*, **19**(8), 483 (1988).
7. R.W. Stow, B.F. Randall, *Am. J. Physiol.*, **179**, 678 (1954).
8. J.W. Severinghaus, A.F. Bradley, *J. Appl. Physiol.*, **13**, 515 (1985).
9. D.M. Zall, D. Fisher, M.O. Garner, *Anal. Chem.*, **28**, 1665 (1956).
10. E. Cotlove, 'Determination of Chloride in Biological Materials', Beckman ASTRA Chloride Module, 9–1, 1981.
11. L.P. Cawley, 'Principles of Chromatography', in *Clinical Diagnosis by Laboratory Methods*, 15th edition, W.B. Saunders, Philadelphia, 547, 1974.
12. L.E. Schoeff, R.H. Williams, *Principles of Laboratory Instruments*, C.V. Mosby, St Louis, 198–199, 1993.
13. P.R. Brown, E. Grushka (eds.), 'High Performance Capillary Electrophoresis (CE) of Human Serum and Plasma Proteins', in *Advances in Chromatography*, Marcel Dekker Inc., New York, Vol. 34, 1994.
14. P.P. Sher, 'Instrument Features Comparison Chart. Centrifugal Analyzers', *J. Clin. Lab. Automat.*, **1**(1), 60 (1981).

15. D.S. Young, 'Biological Variability', in *Chemical Diagnosis of Disease*, eds. S.S. Brown, F.L. Mitchell, D.S. Young, Elsevier Science, Amsterdam, 113, 1979.
16. D.S. Young, *Effects of Preanalytical Variables on Clinical Laboratory Tests*, 2nd edition, AACC Press, Washington, DC, 1997.
17. J.H. Ladenson, 'Nonanalytical Sources of Variation in Clinical Chemistry Results', in *Gradwohl's Clinical Laboratory Methods and Diagnosis*, 8th edition, eds. A.C. Sonnenwirth, L. Jarett, C.V. Mosby, St Louis, 149–192, 1980.
18. R.H. Laessig, A.A. Indriksons, D.J. Hassemer, T.A. Paskey, T.H. Schwarz, 'Changes in Serum Chemical Values as a Result of Prolonged Contact with the Clot', *Am. J. Clin. Pathol.*, **66**, 598–604 (1976).
19. S. Winsten, 'Collection and Preservation of Specimens', in *Standard Methods of Clinical Chemistry*, ed. S. Meites, Academic Press, New York, Vol. 5, 1965.
20. J.S. Annino, A.S. Relman, 'The Effect of Eating on Some of the Clinically Important Chemical Constituents of Blood', *Am. J. Clin. Path.*, **31**, 155 (1959).
21. National Committee for Clinical Laboratory Standards (NCCLS), Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture, Approved Standard Document H3-A3, 3rd edition, NCCLS, Wayne, PA, 1991.
22. National Committee for Clinical Laboratory Standards (NCCLS), Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture, Approved Standard Document H4-A3, 3rd edition, NCCLS, Wayne, PA, 1991.
23. National Committee for Clinical Laboratory Standards (NCCLS), Procedures for the Collection of Diagnostic Blood Specimens by Venipuncture, Approved Standard Document H11-A2, 3rd edition, NCCLS, Wayne, PA, 1991.
24. D.S. Young, E.W. Bermes. 'Specimen Collection and Processing: Sources of Biological Variation', in *Tietz Textbook of Clinical Chemistry*, 3rd edition, eds. C.A. Burtis, E.R. Ashwood, W.B. Saunders, Philadelphia, 42–72, 1999.
25. C. Burgess, A. Knowles, (eds.), *Standards in Absorption Spectrophotometry*, Chapman and Hall, New York, 1981.
26. NIST Special Publication 260, US Department of Commerce, National Institute of Standards and Technology, Washington, DC, 104–106, 1995.
27. M. Nalesnik, C. Frisch, T. Whiteside, 'Instrument Features, Comparison Chart. Immunonephelometers', *J. Clin. Lab. Automat.*, **2**(3), 218–219 (1982).
28. S. Narayanan, 'Atomic Absorption Spectroscopy', *J. Clin. Lab. Automat.*, **2**(4), 243–249 (1982).
29. Y.C. Wu, W.F. Koch, R.A. Durst, 'Standard Reference Materials: Standardization of pH Measurements', Special Publication 260-53, US Government Printing Office, Washington, DC, 1988.
30. M.S. Mohan, R.G. Bates, 'Standardization of Ion Selective Electrodes for Serum Analysis', National Bureau of Standards Special Publication, Gaithersburg, MD, **450**, 293–299, 1977.
31. D.J. David, *Gas Chromatographic Detectors*, Wiley & Sons, New York, 1974.
32. J. Sevcik, *Detectors in Gas Chromatography*, Elsevier, New York, 1975.
33. L.A. Kaplan, A.J. Pesce, *Clinical Chemistry, Theory, Analysis and Correlation*, The C.V. Mosby Company, St Louis, 1984.
34. G.T. Bender, *Principles of Chemical Instrumentation*, W.B. Saunders, Philadelphia, 1987.
35. S. Narayanan, 'Spectrophotometry: Fundamental Concepts', *J. Clin. Lab. Automat.*, **1**(1), 13–16 (1981).
36. E.L. Wehry, *Modern Fluorescence Spectroscopy*, Plenum Press, New York, Vols 3 and 4, 1981.
37. American Society for Testing and Materials, Transferring Information Between Clinical Instruments and Computer Systems, ASTM, Philadelphia, PA, E 1394–91, 1991.
38. American Society for Testing and Materials, Transferring Information Between Clinical Instruments and Computer Systems, ASTM, Philadelphia, PA, E 1381–95, 1995.
39. US Department of Health and Human Services, Clinical Laboratory Improvement Amendments of 1988, Final Rules and Notice 42 CFR Part 493, The Federal Register, Rockville, MD, **57**, 7188–7288, 1992.
40. R.N. Barnett, W.J. Youden, 'A Revised Scheme for the Comparison of Quantitative Methods', *Am. J. Clin. Pathol.*, **54**, 454–462 (1970).
41. A.E. Hartmann, 'The Concept of Establishing vs Verifying Analytic Claims in Method Evaluations', *Arch. Pathol. Lab. Med.*, **116**, 777–780 (1992).
42. J. Westgard, G. Klee, 'Quality Management', in *Tietz Textbook of Clinical Chemistry*, 3rd edition, eds. C.A. Burtis, E.R. Ashwood, W.B. Saunders, Philadelphia, 1996.
43. J.C. Boyd, R.A. Felder, J. Savory, 'Robotics and the Changing Face of the Clinical Laboratory', *Clin. Chem.*, **42**, 1901–1910 (1996).

Lipid Analysis for Important Clinical Conditions

Elizabeth Teng Leary and Timothy H. Carlson
Pacific Biometrics, Inc., Seattle, USA

1 Introduction **2**
2 Lipoprotein Structure and Basic Biochemistry **3**
2.1 Classification 3
2.2 Structure 3
2.3 Classical Techniques for Separation of Lipoproteins 4
3 Important Clinical Conditions **6**
4 Lipid and Lipoprotein Methods **7**
4.1 Overview 7
4.2 Preanalytical Considerations 7
4.3 Cholesterol 8
4.4 Low-density Lipoprotein Cholesterol 10
4.5 High-density Lipoprotein Cholesterol 11
4.6 Triglycerides 13
5 Apolipoproteins **14**
5.1 Apolipoprotein A-I 14
5.2 Apolipoprotein B 15
5.3 Apolipoprotein C-II and Apolipoprotein C-III 15
5.4 Apolipoprotein E 16
6 Lipoprotein Particle Size **16**
6.1 Methods for Assessment of Low-density Lipoprotein Subclasses 16
6.2 Methods for Assessment of High-density Lipoprotein Subclasses 17
6.3 Methods for Simultaneous Assessment of Low-density Lipoprotein and High-density Lipoprotein Subclasses 17
7 Other **17**
7.1 Lipoprotein(a) 17
7.2 Triglyceride-rich Remnant Lipoproteins 18
7.3 Oxidized Low-density Lipoproteins and Lipid Markers of Oxidative Stress 18
7.4 Nonesterified Fatty Acids 18
Abbreviations and Acronyms **19**
Related Articles **19**
References **20**

Lipid analyses are primarily used for determining the risk of diseases resulting from atherosclerosis, including coronary heart disease (CHD), ischemic cerebrovascular disease, and peripheral arterial disease. These disorders are often marked by hyperlipidemia, with increased serum low-density lipoprotein cholesterol (LDL-C) and apolipoprotein (Apo) B, but decreased high-density lipoprotein cholesterol (HDL-C) and Apo A-I. For use in risk assessment, serum cholesterol, LDL-C and HDL-C cut-points have been derived from epidemiological data from several nations. The acceptance of these cut-points, coupled with the remarkable heterogeneity of lipoproteins, has made standardization of cholesterol analytical procedures essential. An international network of laboratories (the Cholesterol Reference Method Laboratory Network (CRMLN)) coordinated by the US Centers for Disease Control (CDC) was formed to address this issue. Thus, several reference methods have been adopted and are used to standardize and certify the commercial reagents for lipid assays for use in the routine clinical laboratory. Other standardization efforts include those by the International Federation of Clinical Chemistry, World Health Organization (WHO) and CDC on Apos or lipoprotein(a) (Lp(a)). In addition to analytical standardization, it is important to standardize preanalytical procedures, and to be aware of potential biological variability.

In the 1980s and 1990s, significant advances have been made in the procedures for routine analysis of lipids and lipoproteins. The analyses have become much more precise and efficient, frequently performed in automated analyzers using small sample volumes. A variety of enzymatic methods for cholesterol and triglyceride (TG) measurements are in current use. HDL-C is generally determined after precipitation of low-density lipoprotein (LDL) and very low-density lipoprotein (VLDL). LDL-C is calculated from the difference between total cholesterol (TC), and the sum of HDL-C and a total TG-derived estimate of the very low-density lipoprotein cholesterol (VLDL-C) (the Friedewald equation). Several direct methods for measurement of LDL-C and HDL-C have been introduced with improvement in performance and labor efficiency, Apo B which is associated with LDL in equimolar ratio and Apo A-I which is associated with high-density lipoprotein (HDL) may be measured in conjunction with LDL-C and HDL-C or used in their place for clinical assessment. Other important lipid measurements include the C and E Apos, Lp(a), TG-rich remnant lipoproteins, lipoprotein particle size and markers of oxidative stress.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Several of the many lipids found in humans have been identified as biomarkers of clinical conditions. By far the most important of these conditions is CHD, which is the number one cause of death in most first-world nations. In 1984, the Lipids Research Clinics Coronary Primary Prevention Trial provided definitive evidence that lowering serum cholesterol reduces the risk of contracting coronary artery disease.[1,2] In the United States, the US Public Health Service responded to these findings. Thus, the National Cholesterol Education Program (NCEP), a cardiovascular risk factor education program administered by the Office of Prevention, Education, and Control at the National Heart Lung and Blood Institute of the National Institutes of Health (NHLBI), was established the following year.

The coordinating committee of the NCEP now directs expert panels and working groups to develop lipid diagnostic and treatment guidelines. These panels include the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (The Adult Treatment Panel), the Expert Panel on Population Strategies for Blood Cholesterol Reduction (The Population Panel), the Expert Panel on Blood Cholesterol Levels in Children and Adolescents, the Laboratory Standardization Panel, and the Working Group on Lipoprotein Measurement. Reports of the NCEP Panels and Working Group are available from the United States Public Health Service National Institutes of Health (NIH) or on the NHLBI web site (www.nhlbi.nih.gov).

In 1993, the second report of the Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults was released, taking precedence over the first (1988) report by the panel. It, like the first report of the panel, continued to consider LDL (specifically LDL-C) to be the primary target of lipid lowering therapy (see Table 1).[3,4] However, it drew additional attention to HDL-C as a "negative" risk factor for CHD risk, and recommended that it be part of initial cholesterol testing. Diagnostic and treatment recommendations were provided for the primary prevention of CHD in adults without evidence of disease, and secondary prevention for patients with established CHD. These include epidemiologically based cut-points for cholesterol for use in diagnosis and treatment.

The challenge facing a campaign to institute universal treatment cut-points is that all laboratories performing the analysis must be standardized to the same accuracy base. This is different from traditional laboratory medicine where different testing methods often yield results that are numerically different. Under these circumstances, a reference range is developed by each institution or for each method. On the other hand, when the same treatment cut-points are universally applied, the same accuracy base for testing, the same sample types (serum, plasma or whole blood) and specimen collection conditions (posture, fasting status, and so forth) must be carefully maintained.

In 1995, the Working Group on Lipoprotein Measurement established the current analytical goals for measurement of lipid markers in the screening and monitoring of hyperlipidemias (Table 2). In addition, the limits for routine methods were formalized, guidelines for specimen collection, types and storage provided, and the need to obtain serial measurements on each patient was stressed. Reference methods recommendations were also made, along with recommendations to continue to link routinely used methods with epidemiological data obtained using the same methods.[5–8]

Naturally, the art and science of lipid testing run in advance of its standardization, so the standardization of many newer tests has not yet been addressed. This may be because of a lack of reference materials, reference methods or epidemiologically derived treatment cut-points. The evolution of lipid testing nearly always far precedes the recognition of how such parameters as biological matrix affect analyte behavior. Once these effects are known, how to assimilate this knowledge into

Table 1 NCEP Adult Treatment Panel II categories for total and LDL-C concentrations for patients with and without CHD

TC category	TC mg dL^{-1} (mmol L^{-1})	LDL-C category	LDL-C mg dL^{-1} (mmol L^{-1})
Patients without CHD			
Desirable	<200 (5.2)	Desirable	<130 (3.4)
Borderline-high	200–<240 (5.2–6.2)	Borderline	130–<160 (3.6–4.1)
High	≥240 (6.2)	High-risk	≥160 (4.1)
Patients with CHD			
		Optimal	≤100 (2.6)
		Higher than optimal	>100 (2.6)

Table 2 NCEP performance criteria for lipid and lipoprotein measurements

Analyte	Bias	Imprecision	Total error (%)
TC	≤3% of RV	CV ≤ 3%	≤9
TG	≤5% of RV	CV ≤ 5%	≤15
HDL-C	≤5% of RV	SD ≤ 1.7 at <42 mg dL^{-1}	≤13
		CV ≤ 4% at ≥42 mg dL^{-1}	
LDL-C	≤4% of RV	CV ≤ 4%	≤12

CV, coefficients of variation; RV, reference value determined by CDC reference methods; SD, standard deviation.

the development of assay standards and calibrators must be researched.

The standardization of lipids and lipoproteins presents a particular challenge because of their heterogeneous nature. To enable laboratories and manufacturers of lipid reagents to achieve the analytical goals set by the NCEP, the CDC created the CRMLN.[9,10] Using reference methods and fresh serum samples, analytical accuracy is passed from this laboratory network to the ''field''. CRMLN has an international presence, currently with six North American (five US, one Canadian), three European (United Kingdom, The Netherlands and Italy) and one Asian (Japan) laboratories. Each of these laboratories has instigated or participated in standardization efforts in their respective countries. Other programs include the National Initiative on Cholesterol Accuracy, Methods and Standardization (NICAMS) in the United Kingdom, the cholesterol target assignment of the External Quality Assessment Scheme (EQAS) control and calibrator materials conducted by the Rotterdam University Hospital Network Laboratory in The Netherlands, the Programa de Evaluacion Externa de Calidad in Argentina, as well as several lipid proficiency surveys using fresh or frozen human sera that are traceable to a reference method accuracy base.

As discussed, the NCEP has focused on standardization of lipoprotein cholesterol measurements. Advances in standardization of lipid methods that are outside the scope of the NCEP program are being addressed by other organizations. For example, the International Federation for Clinical Chemistry (IFCC), the WHO and the CDC are collaborating on Apo and Lp(a) standardization efforts to derive international reference materials.[11–13]

It should be stressed that in addition to being important risk factors for CHD, lipid markers are useful in the diagnosis and monitoring of several other disease states. These include primary and secondary hyperlipoproteinemia associated with obesity, disturbed glucose tolerance, diabetes mellitus, diseases of the kidney, liver, thyroid, pancreas, and gastrointestinal track, cancer, and perhaps neurodegenerative diseases.

2 LIPOPROTEIN STRUCTURE AND BASIC BIOCHEMISTRY

2.1 Classification

Lipids are a diverse class of compounds. Many have in common only their uniform solubility in organic solvents like ethers, carbon disulfide and carbon tetrachloride. Of this diverse class of compounds, only a few have been found to be of use in the study or diagnosis of important clinical conditions. The primary clinically significant lipids are members of the most abundant categories, including triacylglycerols (usually referred to as TGs by clinicians), and sterol compounds. TGs and the primary sterol, cholesterol and its esters, are associated in blood in lipoprotein particles. These particles contain lipid, as well as several proteins that facilitate the transport and tissue distribution of the hydrophobic lipids within the primarily aqueous environment of the blood and lymph.

2.2 Structure

2.2.1 Triacylglycerols (Triglycerides)

The simplest description of a TG is the esterification product of glycerol and three fatty acids. These neutral lipids are biosynthetic products of L-glycerol and fatty acids. The specificity of glycerol kinase in the first step in the synthesis of acylglycerols maintains the stereointegrity of TGs, which may contain three identical fatty acid esters (a simple TG) or contain a mixture of fatty acids (a mixed TG). Generally, mono- and diacylglycerols are present only as partial breakdown products of triacylglycerols.

Since mixed fatty acids are common, it is necessary to differentiate the three positions on the glycerol backbone. The three positions are designated 1, 2 and 3, or α, β and α'. A TG with two different fatty acids can exhibit up to six distinct molecular structures. Although the TGs are structurally diverse, the possible structures are limited both by species-specific dietary intake and enzymatic specificity.

Human triacylglycerols primarily contain 16- (C_{16}) and 18-carbon (C_{18}) fatty acids with smaller amounts of shorter (usually C_{12} or C_{14}) or longer fatty acids (commonly C_{20} or occasionally C_{22}). The fatty acids may have from one to several double bonds with the first double bond often occurring between the C_9 and C_{10} positions. Certain $w3$ and $w6$ fatty acids (fatty acids with the double bond closest to alkyl end are designated by an w or n) are required for normal growth and

development. These include α-linolenic acid ($C_{18}:3w3$), docosahexanoic acid ($22:6w3$), linoleic acid ($18:2w6$), and arachidonic acid ($20:4w6$). Some fatty acids from the $w3$ and some from the $w6$ series are required for normal skin, growth, vision, learning, reproduction, liver function and hydration.

The unsaturated fatty acids have a nonlinear carbon sequence. This means the van der Waals forces that cause attraction between the aliphatic tails of saturated fatty acids are weaker for unsaturated fatty acids. Hence, unsaturated fats have a lower melting point than saturated fats of the same chain length.

Usually the double bonds of fatty acids are found in the *cis* conformation, but *trans* fatty acids are present in TGs that have undergone industrial hydrogenation for use in the manufacture of margarine and shortening. They are also present in a portion of the fatty acids in fats of ruminant mammals (produced by bacterial action within the rumen).

Fatty acids with multiple double bonds (polyunsaturates) may be classified on the basis of the first double bond from the aliphatic end of the molecule. Those with the first double bond between the 3rd and 4th carbons are called *n*-3 (or *w*-3) fatty acids. Another common class of unsaturated fatty acids is the *n*-6 (or *w*-6). One (monounsaturated) to five double bonds occur most often.

2.2.2 *Cholesterol and Cholesterol Esters*

Cholesterol **(1)** is a derivative of the fused, reduced polycyclic structure perhydrocyclopentanophenanthrene **(2)**. For cholesterol, two methyl groups (C_{18} and C_{19}), a hydroxy group (at C_3), an eight carbon branched aliphatic chain (at C_{17}), and a double bond (at C_5-C_6) are added to this basic structure. Unlike triacylglycerols, cholesterol is not saponifiable. However, cholesteryl esters, containing unsaturated fatty acids, are synthesized in, for example, HDL particles. This occurs as the result of the action of lecithin–cholesterol acyltransferase (LCAT) that is present in HDL. LCAT converts the so-called free cholesterol present at extravascular sites to esterified cholesterol for "reverse-cholesterol transport" by transfer of the fatty acid at the β position of lecithin (phosphatidylcholine) to the C_3 hydroxy site of cholesterol. Thus, a concentration gradient is set-up that encourages further diffusion of cholesterol from the extravascular space.

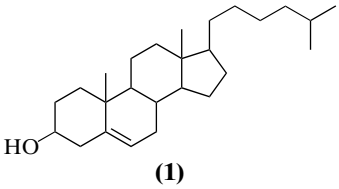

(1)

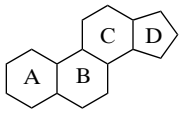

(2)

2.2.3 *Basic Lipoprotein Structure*

Lipoproteins and the closely related chylomicrons are assemblages of proteins, called Apos, with TGs, cholesterol, cholesterol esters, phospholipids, and other compounds (see Figure 1). Free cholesterol, the hydrophobic heads of phospholipids, and the hydrophobic portions of Apos occupy the monolayer at the lipoprotein surface. The core of the particle contains cholesteryl esters and TGs. The resulting aggregates vary in size, density and function.

The largest of the particles are chylomicrons and VLDLs. These contain large amounts of TGs and considerable cholesterol that are destined for transport from the intestine and liver, respectively. The TGs in these particles are hydrolyzed primarily at the capillary surface of adipose and muscle tissue by the enzyme lipoprotein lipase. The resulting fatty acids are transported into the cell. This decreases the size of the particles and increases their density, as the fraction of Apo, cholesterol and cholesteryl esters that are present increases. A fraction of VLDL is successively converted by this process into intermediate density lipoproteins (IDLs) and LDLs, whereas a equal or larger portion are converted to remnant particles.[14] Likewise, chylomicrons are converted to chylomicron remnants. Both of these remnants are called TG-rich lipoprotein remnants. The remnants of VLDL and chylomicrons, and LDL particles are atherogenic, and are relatively slowly removed from the circulation by binding to hepatic receptors.

2.2.4 *Lipoprotein(a)*

Lp(a) is an atherogenic lipoprotein particle whose function is not currently clear. Structurally, Lp(a) is an LDL particle with its single copy of Apo B-100, plus one copy of a protein designated Apo(a) that is covalently linked by a disulfide bond to the Apo B-100. Apo(a) is a glycoprotein with a variable number of peptide subunits that have a high degree of homology to the highly conserved kringle subunits of plasminogen, the central protein in the fibrinolytic system. Variability in the number of kringle repeats occurs from person to person, and is responsible for Lp(a) sizes ranging from 187 to 662 kD.

2.3 Classical Techniques for Separation of Lipoproteins

2.3.1 *Ultracentrifugation*

Historically, lipoproteins have been separated by methods based on their different physical properties. While ultracentrifugation is now rarely done for clinical purposes, it is the method that gives us much of the current lipoprotein nomenclature and it is still widely used in research.

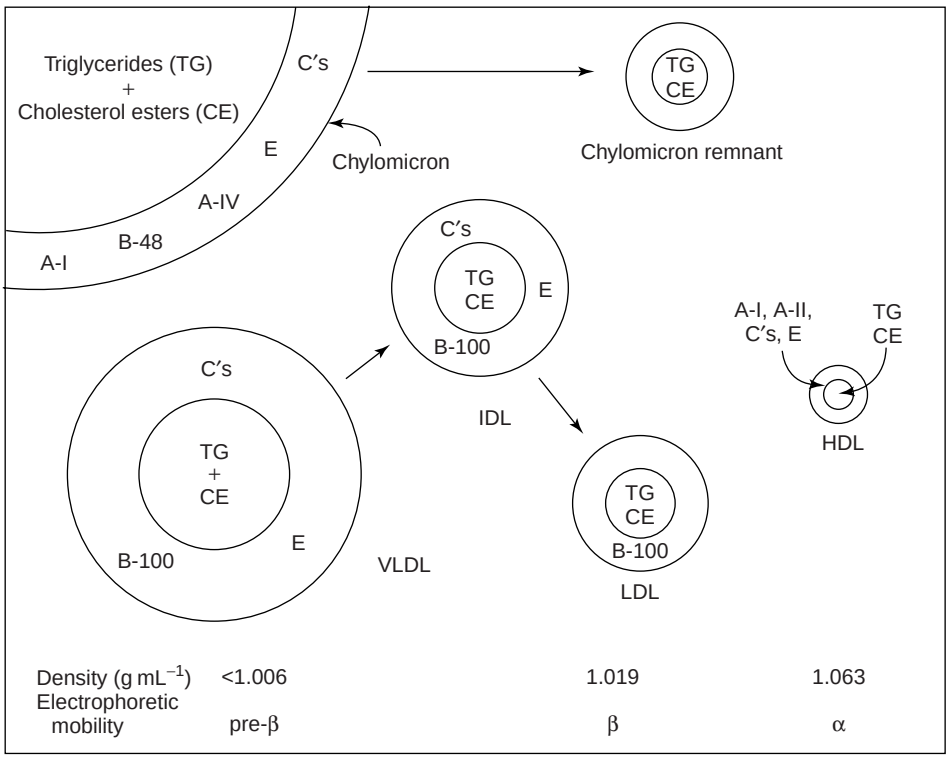

Figure 1 A simplified overview of lipoprotein characteristics. The various particles are represented by double rings to indicate that each is made up of an outer layer that contains phospholipid, free cholesterol, and Apos (A-I, B-48, A-IV, E, C-I, C-II, C-III, B-100, A-II). The inner ring represents the particle core, which contains TGs and cholesterol esters in varying proportions. The density (*d*) and electrophoretic mobility of the various particles are indicated at the bottom of the figure. Each lipoprotein illustration represents a family of particles with a distribution of different sizes and densities.

The principal method for isolation of lipoproteins is the sequential flotation technique.[15] In this method, the density of the plasma is adjusted to upper limit of the class of lipoprotein particle to be isolated (see Table 3) before centrifugation. The lipoprotein floats at a specific density and can be removed from the solution. Thus, after removal of the chylomicron by a short centrifugation, and a longer one for removal of VLDL, the density of the plasma can be adjusted to 1.019, 1.063 and 1.210 with 18-h centrifugation between adjustments. The density of salt solutions prepared from NaCl, KBr or NaBr can be determined from the refractive index of the solution or from a density meter, a glass tubing shaped like a tuning fork that produces pitches that are dependent on the density of the solution with which it is filled.[16]

Distinct classes of VLDL, IDL, LDL can be isolated by using density gradient ultracentrifugation,[17] or rate zonal ultracentrifugation.[18] These techniques have been refined, allowing the separation of three to seven LDL subfractions.[19,20] In addition, several HDL subfractions may be isolated.

Flotation ultracentrifugation to remove chylomicrons and VLDL can be combined with precipitation of LDL from the LDL–HDL mixture, and chemical or enzymatic assay of cholesterol to determine the cholesterol in each

Table 3 Physical properties of lipoprotein classes

Lipoprotein class	Particle density ($g\,mL^{-1}$)	Molecular wt ($\times 10^{-6}$ Da)	Diameter (Å)	Electrophoretic mobility
Chylomicron	<0.94	50–1000	750–12 000	Origin
VLDL	0.94–1.006	10–80	300–800	Prebeta
IDL	1.006–1.019	5–10	290–350	Slow prebeta/beta
LDL	1.019–1.063	2–3	230–290	Beta
HDL	1.063–1.210	0.07–0.40	50–120	Alpha
Lp(a)	1.05–1.08	1.9–6.6		Fast prebeta

of the lipoprotein fractions, in a technique that is called β-quantification (see below).[21]

2.3.2 *Electrophoresis*

As indicated in Table 3, lipoprotein particles can be separated by electrophoresis. Historically, this led to the classification of six different hyperlipoproteinemic phenotypes: type I, type II-a, type II-b, type III, type IV and type V (see review by Schaefer and Levy[22]). This classification system results in large part from the relative amounts of lipoproteins present in each of four distinct lipid-staining bands. The relative migration of the lipoproteins compared to plasma proteins is in the so-called α- and β-regions (Table 3). These are strictly qualitative classifications, and as cholesterol quantification became a routine procedure, electrophoretic systems came to be used only in classification of rare hyperlipoproteinemias. However, advances in automation, scanning, and enzymatic lipid staining have led to commercial development of automated instruments that electrophorese specimens on agarose, stain the gel, scan the electrophoretograms and calculate the specimen lipoprotein fraction cholesterol concentration.[23,65]

2.3.3 *Chromatography*

Lipoproteins can be separated by chromatography using size discrimination as the criterion. Gel permeation (size exclusion) high-performance liquid chromatography (HPLC)[24] or fast-performance liquid chromatography (FPLC)[25] columns can be used to separate the primary lipoprotein classes and subclasses of VLDL, LDL, and HDL. A commercially available system has been placed on the market that is dedicated to the separation of lipoproteins and set up to allow on-line enzymatic measurement of cholesterol in each fraction. Automated assays that correlate well with those obtained by sequential ultracentrifugation are available.[26]

3 IMPORTANT CLINICAL CONDITIONS

NCEP guidelines indicate that serum cholesterol levels $\geq 200\,mg\,dL^{-1}$ ($5.2\,mmol\,L^{-1}$) are associated with increased risk of CHD.[3,4] In point of fact, serum cholesterol levels between $150\,mg\,dL^{-1}$ ($3.88\,mmol\,L^{-1}$) and $200\,mg\,dL^{-1}$ ($5.2\,mmol\,L^{-1}$) are associated with a graded increase in CHD risk, especially when other risk factors are present.[27,28] These other risk factors include, but are not limited to, genetics, age, gender, smoking history, blood pressure, and serum TGs, Lp(a), homocysteine, LDL-C and HDL-C.[28,29]

Over 95% of individuals with hyperlipidemia (dyslipidemia), elevated blood lipid concentration, suffer from hyperlipoproteinemia, elevated serum lipoprotein concentration. As already discussed, LDL-C is a primary focus of the NCEP guidelines and is a better risk marker than TC, and HDL-C is a negative risk marker. On the other hand, the ratio of TC or LDL-C to HDL-C may even more faithfully predict risk.[30] The importance of detection of hypertriglyceridemia for risk assessment is not as clear as that of hypercholesterolemia, but evidence from cellular, genetic, and molecular, as well as epidemiological and clinical studies, suggests that it is associated with increased risk.[31]

Hypocholesterolemia is relatively frequent in individuals with affective disorders (depression, bipolar disorder, and schizoaffective disorder).[32] Low cholesterol levels are also associated with some forms of cancer,[33,34] but this may be an effect of the disease rather than a cause.[35] Extremely low concentration is a prognostic indicator of patient survival in hospitalized patients,[36] and is associated with protein–calorie malnutrition.[37] Low levels of cholesterol are prevalent in human immunodeficiency virus-1 (HIV-1) infection even before progression to acquired immune deficiency syndrome (AIDS), and marked hypocholesterolemia, as a sequel to malabsorption, is associated with poor prognosis in AIDS patients.[38] Chronic hypotriglyceridemia is not generally a diagnostic or prognostic marker. However, like hypocholesterolemia, it is associated with malabsorption syndromes. For example, AIDS patients with fat malabsorption, and chronic diarrhea and weight loss are thus affected.[39]

Some individuals with total serum cholesterol concentrations that are deemed low risk by the NCEP develop CHD, just as some people with serum cholesterol concentrations associated with markedly increased risk do not. To begin with, this occurs because the cholesterol associated with the various lipoproteins that comprise total serum cholesterol may be either positively or inversely linked to the development of atherosclerosis, the underlying cause of CHD. Furthermore, the atherosclerotic lesion initiation and development are multifactorial processes. There seem to be several factors that are agents in or influencers on the development of atherosclerosis. These include low-grade coagulation,[40] inflammatory processes,[40,41] increased oxidative stress or reduced antioxidant levels,[41] and factors contributing to one or more of these conditions, such as excess plasma homocysteine.[42]

Obesity is associated with changes in serum lipid concentrations that are associated with increased risk of CHD. This is apparently in large part the result of insulin resistance, an eventual consequence of excess adiposity, especially in the abdominal region.[43] Insulin resistance is

associated with decreased adipose and muscular uptake of glucose and fatty acids because of downregulation of insulin receptors on the cellular membrane. These processes lead to increases in serum TG levels, decreased HDL-C, and changes in the composition of LDL-C. Overproduction of TG-rich VLDL, fewer HDL particles, and more small dense LDL particles occur.[44]

Insulin resistance leads to type II diabetes when insulin production is less than that needed to maintain cellular glucose uptake. Besides obesity, type II diabetes and hyperlipidemia, hypertension occurs as a consequence of persistent insulin resistance. This constellation of clinical outcomes, sometimes called syndrome X,[45] is associated with extremely high risk for the development of atherosclerosis and CHD.

Early dietary studies suggested that fat and total calorific intake is related to the risk of colon, prostate and breast cancer (see Wu et al.[46]). However, except for low levels of HDL cholesterol, serum lipids do not correlate with the cancer incidence.[47] It appears that some other factors associated with increased fat intake, e.g. obesity or red meat consumption, and not dietary fat intake or serum lipids alone correlate with the risk of cancer.[46]

4 LIPID AND LIPOPROTEIN METHODS

4.1 Overview

Procedures for measurement of clinically important lipids have evolved over the last half of the twentieth century. Some of the earliest techniques for measurement of serum lipids are still in use as reference methods. In the clinical laboratory, lipids are usually measured by quantification of colorimetric products of enzymatic reactions. Apos are assayed by a variety of immunoassay techniques. Table 4 summarizes methods for assay of the lipid components of lipoproteins.

Table 4 Methods for analysis of the lipid components of lipoproteins

Method type	Principle	Refs.
	Cholesterol	
Definitive	Isotope dilution/mass spectrometry (NIST)	48–50
Reference	Abell–Kendall (CDC)	51, 52
Routine	Enzymic (esterase, oxidase, peroxidase)	53–56
	HDL Cholesterol	
Definitive	None	
Reference	Ultracentrifugation/heparin-Mn^{2+} precipitation/Abell–Kendall cholesterol (CDC)	57, 58
Routine	Chemical precipitation/enzymic cholesterol	59–62
	Homogeneous – direct	63, 64
	Agarose electrophoresis/cholesterol-specific staining/densitometry	23, 65
	LDL Cholesterol	
Definitive	None	
Reference	β-Quantification (ultracentrifugation/heparin-Mn^{2+} precipitation/Abell–Kendall cholesterol) (CDC)	15, 66
Routine	β-Quantification (ultracentrifugation/precipitation/enzymatic cholesterol)	
	Friedewald estimation	67
	Homogeneous – direct	68
	Agarose electrophoresis/cholesterol-specific staining/densitometry	23, 65
	Triglycerides	
Definitive	Isotope dilution/mass spectrometry (NIST)	69
Reference	Chromotropic acid (CDC)	70, 71
Routine	Enzymatic	72–74
	Phospholipids	
Definitive	None	
Reference	None	
Routine	Enzymatic (choline-containing), chemical, TLC, and HPLC	75–79

NIST, National Institute for Standards and Technology; TLC, thin-layer chromatography.

4.2 Preanalytical Considerations

In order to achieve accurate and precise lipid determinations, sources of preanalytical variation should be carefully monitored and controlled. Preanalytical variation encompasses biological variation (gender, age, race, physiological state), variation due to underlying pathological or physiological conditions, and those associated with specimen collection and handling. Numerous reports have underscored the significance of intraindividual biological variability in lipids.[80–82] For most lipid analytes, the mean biological CV range is 6–10%. The exceptions are TGs, which have a much higher biological variation with an average CV of 28%.[81–83] With the availability of increasingly more accurate and precise analytical methods, the preanalytical variability may have a larger contribution to measurement error than analytical variability.

Controllable factors which may have a significant impact on lipid values include body posture during blood collection,[84,85] fasting status,[86,87] capillary or venous blood sample,[88,89] sample type (serum or plasma), anticoagulant or preservative utilized, and specimen

processing and storage conditions. In addition, usual lipid status should be assessed under stable metabolic and lifestyle conditions. Blood specimens should be collected with minimal stasis after the subject has remained seated for at least 5 min.[85] A 12 h or longer fast is preferred for most lipid testing, although nonfasting samples are acceptable for TC, HDL-C, some LDL-C methods and some Apo assays.

Serum is the preferred sample type for routine analysis since the NCEP treatment guidelines are based on serum values.[3] Plasma samples, especially with ethylenediaminetetraacetic acid (EDTA) as the anticoagulant, are also used for lipid analysis. However, one must be aware of possible latent fibrin clot formation in plasma, which can be especially troublesome after freezing. In addition to venous serum and plasma, capillary whole blood may be used in portable ''point-of-care'' analyzers for lipid testing. Owing to the wide variety of point-of-care technology, the comparability of results between capillary blood and venous blood must be verified for each system.[88,89]

In addition to standardizing patient preparation and sample collection carefully, serial patient testing, at least at one week intervals, is recommended to decrease intraindividual variation by averaging several results.[81] Clinical sources of variation such as medications and underlining clinical conditions including pregnancy[90,91] should be considered when measuring lipids. Antihypertensive drugs, immunosuppressive drugs and sex hormones used in, for example, oral contraceptives and hormone replacement therapy, are known to increase lipid values.

4.3 Cholesterol

Lipoproteins have traditionally been quantified on the basis of their cholesterol content.[92] TC as well as cholesterol associated with each class of lipoprotein, i.e. cholesterol in VLDL, IDL, LDL, and HDL, can be determined either by chemical means or enzymatically. The measurement of cholesterol is performed directly on serum or plasma or after separation of the desired lipoprotein particle from other lipoproteins or biological components.

The definitive method for cholesterol determination is one employing isotope dilution/mass spectrometry (ID/MS).[93,94] This procedure was developed at the NIST, but is rarely if ever used in the clinical laboratory. The currently accepted reference procedure for cholesterol is the modification by the CDC of the original cholesterol assay of Abell, Levy, Brody and Kendall.[51,52] The Abell–Kendall reference method is based on the alcoholic hydrolysis of cholesterol ester followed by hexane extraction. The purified cholesterol is quantified by reacting with the Liebermann–Burchard (L–B) color reagent.[95,96] It has been demonstrated that the Abell–Kendall method has a positive bias of 1.6% compared to the definitive method.[97,98] The Abell–Kendall method, as prescribed by the CDC, serves as the accuracy base for the NCEP performance guidelines[5–8] used in the support of both the adult and pediatric NCEP treatment guidelines.[99,100] The CDC also serves to provide the accuracy targets for the CRMLN and the NHLBI Lipid Standardization Programs.[10]

As the Abell–Kendall procedure is labor intensive, requiring a meticulous care for detail beyond that practiced in the normal clinical laboratory setting, it is not recommended for routine use. The cholesterol esterase–cholesterol oxidase enzymatic methods are used almost universally for routine cholesterol determination, both in clinical and research laboratories. To help ensure the transfer of accuracy to the cholesterol reagent manufacturers and clinical laboratories, the CRMLN offers a certification program based on fresh serum comparisons. Thus, these samples are analyzed in the field by a specific enzymatic method and in the CRMLN laboratory by the reference Abell–Kendall method following a carefully defined protocol.

4.3.1 Abell–Kendall Reference Method

The modified Abell–Kendall assay performed according to CDC[21,51,52] protocol is linear to 400 mg dL^{-1} (10.36 mmol L^{-1}). One limitation of the method is the lack of sensitivity in the low concentration range resulting in poor precision at very low concentrations such as those encountered in some lipoprotein fractions. When performed under standardized conditions, the CRMLN laboratories routinely have CV of <1% and bias <1% of the CDC target values.

4.3.1.1 Reagents Exact protocol conditions must be followed in all steps of the procedure. Only reagent grade reagent, certified hexane and reagent grade water may be used:

- *Aqueous potassium hydroxide (33%).*
- *Alcoholic KOH (approximately 0.36 mol L^{-1})* – Prepare immediately prior to use by mixing the 33% KOH with absolute ethanol.
- *L–B reagent* – Prepare fresh by carefully combining acetic anhydride and sulfuric acid at 5 °C followed by acetic acid.
- *Cholesterol standard solutions* – Use dry cholesterol SRM911b to prepare working solutions in absolute ethanol.

4.3.1.2 Procedure Add 5 mL of alcoholic KOH to 0.5 mL of each standard, control and sample. Incubate at 50 °C for 60 min. Equilibrate at 25 °C for 10 min

with 5 mL of water, and then extract with 10 mL of hexane. Remove 2 mL of the organic phase and deliver it with 3 mL of hexane to ensure quantitative transfer. Evaporate hexane to dryness with a vacuum oven or under nitrogen. Add 3.25 mL of the L–B reagent to each of the dried extract at timed intervals. After a 30-min incubation, take absorbance reading at 620 nm at the same time intervals. The concentrations of the samples are determined from the linear regression curve for data from standards analyzed in duplicate.

4.3.2 Enzymatic Methods

For routine measurements, enzymatic methods have replaced the older chemical methods such as those employing the L–B reagents because they are easy to use, do not utilize caustic reagents and are easily automated. Most of the cholesterol measurement methods in use are based on the cholesterol esterase–cholesterol oxidase coupled reactions. These are followed by a peroxidase-catalyzed reaction resulting in a colored end-point. Unesterified (free) cholesterol can be quantified by omitting the esterase from the reagent system:

$$\text{cholesterol esters} + H_2O \xrightarrow{\text{cholesterol esterase}} \text{cholesterol} + \text{fatty acids} \quad (1)$$

$$\text{cholesterol} + O_2 \xrightarrow{\text{cholesterol oxidase}} \Delta^4 - \text{cholestenone} + H_2O_2 \quad (2)$$

$$2H_2O_2 + \text{phenolic} + 4 - \text{aminoantipyrine} \xrightarrow{\text{peroxidase}} \text{red chromagen} + 4H_2O \quad (3)$$

A wide variety of reagents are commercially available in kit format. Many are adapted to specific instruments for optimal performance and convenience. A typical cholesterol test is performed in an automated analyzer using a few microliters of sample and requires 5–10 min to complete. Throughput ranges from a few dozen tests to several hundred tests per hour. The cholesterol results are quantified from standard curves that may be stored on-line for several weeks or even months. Final results are frequently uploaded to a central laboratory computer for consolidated reporting with other test results.

While convenient and practical, the enzymatic cholesterol methods can suffer from specificity and interference problems[101] not encountered in the chemical methods. Each reagent and instrument combination including standardization must be evaluated for acceptable performance. Potential sources of variation include the source of the enzymes as well as the final component concentrations in the reagent kit. Common interfering substances in biological samples include ascorbic acid, bilirubin, hemolysis and turbidity caused by hypertriglyceridemia (lipemia). Ascorbic acid interferes in the peroxidase final step by competing with the chromagen substrate for generated peroxide. Although the normal amount of ascorbate present in clinical samples may be spontaneously oxidized before the samples are scheduled for analysis, mega doses of ascorbic acid and other antioxidants may cause difficulties.[102] Point-of-care analyzers will need to incorporate ascorbate oxidase in the reagent format as the samples (usually whole blood) are analyzed immediately after collection where atmospheric oxidation of the ascorbate will not have taken place before analysis. A comprehensive review of enzyme cholesterol methods has been published.[103]

Although the cholesterol reagent is usually purchased in commercial assay kits and measured by an automated instrument, it may be prepared in-house and analyzed manually. Nonesterified cholesterol reagent, which is not readily available in kits, may be prepared by omitting cholesterol esterase.

4.3.2.1 Enzymatic Cholesterol Reagent

- Cholesterol color reagent: an aqueous solution containing 200-mM phosphate buffer (pH 6.7), 1.6-mM 4-aminoantipyrine, 5560 U L^{-1} (enzyme units per liter) peroxidase, 400 U L^{-1} cholesterol esterase, 400 U L^{-1} cholesterol oxidase, a preservative, and a stabilizer.
- Cholesterol phenol reagent: an aqueous solution containing 40-mM phenol, a surfactant, and a stabilizer.

4.3.2.2 Procedure Add 1.0 mL of the reagent to 3.0 µL of sample. Incubate at 37 °C for 10 min. Measure the absorbance at 510 nm against a reagent blank. Lower cholesterol concentration, which requires greater analytical sensitivity, may be measured by increasing the sample-to-reagent ratio using the same esterase/oxidase reagent, or by use of an ultrasensitive reagent system, such as that used for remnant lipoprotein cholesterol measurement.[87]

4.3.2.3 Standardization Standardization of the enzymatic cholesterol reagent requires special care because the serum matrix may affect enzyme function. The goal is to assign a target value to a standard material, which usually has a different matrix from patient samples, so that the cholesterol concentrations obtained for the patient samples measured enzymatically are in agreement with those obtained by the reference methods. The CRMLN certification program was created for this purpose.[10,104]

In the 1990s, several portable cholesterol-measuring devices designed for ''near patient'' testing were introduced. These self-contained instruments for screening or monitoring generally use capillary blood obtained by

''fingerstick''. The chemistry is most often based on the cholesterol esterase–cholesterol oxidase reagent system. The accuracy traceability of these analyzers is achieved by collecting venous serum at the same time as the capillary blood is obtained. The venous serum is standardized to the reference method and provides an accuracy target for the capillary blood samples. Because of the unique design of each of these devices, careful evaluation of the performance of each device must be done before use.

4.4 Low-density Lipoprotein Cholesterol

In serum, approximately two-thirds of TC normally associates with LDL, with the remaining cholesterol residing in the two other major lipoproteins, VLDL and HDL. IDL and lipoprotein ''little a'' [lipoprotein (a) or Lp(a)] are minor lipoproteins containing relatively small amounts of cholesterol. By mass, cholesterol esters constitute approximately 38% and unesterified cholesterol constitutes 8% of LDL.[105] As for all lipoproteins, each of which consists of a complex family of rather distinct particles, LDL is generally assayed by cholesterol content. Because LDL-C is the key NCEP determinant for treatment decisions,[3,4] it is imperative that its determination be both accurate, in order to adopt the global treatment cutpoints, and precise, so that patient progress following therapy may be monitored.

To measure LDL-C, the LDL particles must be first separated from the other lipoproteins before cholesterol determination. LDL particles can be separated on the basis of such physical properties as size, charge, density or precipitation characteristics, or by Apo composition. Historically lipoprotein particles have been separated by sequential ultracentrifugation[105] and equilibrium density gradient ultracentrifugation.[106–108] These methods are labor intensive and tedious and not suitable for the routine clinical or research laboratories. Other methods of LDL-C determination include electrophoresis and chemical precipitation. Several direct measurement methods for LDL-C have been introduced. These methods incorporate the automated LDL separation in one of a series of reactions. It is important to note that most of these LDL separation methods do not distinguish IDL and Lp(a) from LDL, so LDL-C is somewhat overestimated by these measurements. However, the current cut-points for LDL-C include the contributions of the atherogenic IDL and Lp(a).

4.4.1 Beta Quantification

The term beta quantification[15,21,66] commonly refers to a procedure that combines ultracentrifugation and polyanion-mediated lipoprotein precipitation. In addition to LDL-C concentration, beta quantification also provides quantitative VLDL-C and HDL-C values. It is based on the physical separation of the VLDL particles ($d < 1.006\,g\,mL^{-1}$) from non-VLDL lipoproteins ($d > 1.006\,g\,mL^{-1}$) by ultracentrifugation of serum (density $d \approx 1.006\,g\,mL^{-1}$). VLDL and chylomicrons (if present) float to the top of the tube, producing a clear zone between the VLDL and the more dense LDL and HDL fractions at the bottom of the tube. The top and bottom fractions are physically separated by slicing the centrifuge tube with a tube slicer, followed by volumetric recovery of the fractions. Apo B-containing lipoproteins (primarily LDL) are removed from the bottom fraction by chemical precipitation. HDL-C is quantified by measuring the cholesterol content of the HDL remaining in the supernatant. By measuring the cholesterol in the $d > 1.006\,g\,mL^{-1}$ fraction (''bottom fraction''), the LDL-C can be calculated. Thus Equation (4),

$$\text{LDL-C} = \text{bottom fraction-C} - \text{HDL-C} \qquad (4)$$

VLDL-C is calculated by the difference between TC and that in the bottom fraction, Equation (5):

$$\text{VLDL-C} = \text{TC} - \text{bottom fraction-C} \qquad (5)$$

The above calculation produces a more accurate value for VLDL-C than direct measurement of VLDL-C in the top fraction as total recovery of cholesterol in the $<1.006\,g\,mL^{-1}$ fraction may be difficult in samples with high TG concentrations. The LDL-C recovered in the beta quantification method includes intermediate density lipoprotein cholesterol (IDL-C) and Lp(a) cholesterol.

Because of the heterogeneity of lipoproteins, there are many possible variations in the beta quantification procedure, each of which may affect the results. These include ultracentrifugal conditions such as temperature, relative centrifugal force and total *g*-minutes, the specific centrifuge rotor used, as well as the sample volume and the fraction recovery procedure.[109] The cholesterol analysis may be by enzymatic method or the reference Abell–Kendall method. The HDL separation may be by direct precipitation of the unfractionated sample or the $d > 1.006\,g\,mL^{-1}$ fraction using one of several available precipitation reagents. Consequently, the reference beta quantification method must be carefully standardized in order to provide consistent target values.

4.4.2 United States Centers for Disease Control and Prevention Reference Beta Quantification Method

The procedure used by the CDC has been recommended by the NCEP Lipoprotein Measurement Working Group as the reference method for measuring LDL-C. This method has served as the accuracy base for the CRMLN LDL-C certification program. To prepare the $d > 1.006\,g\,mL^{-1}$ fraction, a 5-mL aliquot of

serum or plasma is centrifuged with a 0.195-mol L^{-1} NaCl overlay at an average of 115 000 × g (40 000 rpm (revolutions per minute)) using a Beckman-type 50.3 fixed angle (20° angle) rotor. After an 18.5-h centrifugation at 18 °C, the centrifuge tube is sliced 37 mm from the bottom of the tube. The bottom fraction is then transferred to a 5-mL volumetric flask and brought to volume with 0.15 mol L^{-1} NaCl. Cholesterol in the bottom fraction is quantified by the Abell–Kendall reference cholesterol method.[21,51,52] For HDL-C measurement, the Apo B-containing lipoproteins in the bottom fraction are precipitated with heparin/manganese chloride followed by Abell–Kendall cholesterol analysis of the HDL-containg supernatant. LDL-C is calculated by the difference between the bottom fraction cholesterol and HDL-C.

The CDC reference method differs slightly from the beta quantification method used by the Lipid Research Clinics (LRC) Program,[21] which served as the accuracy base for the LRC Population Studies, a major program of the NHLBI.[110] The LRC method, which specifies fresh sera, a centrifugation temperature of 10 °C, and 0.15 mol L^{-1} NaCl overlay, yields a slightly higher LDL-C value (2–3% in samples with normal TG concentrations) than the CDC reference method.[111] Details of the CDC beta quantification procedure may be obtained by contacting CDC directly.

4.4.3 Friedewald Estimation

The majority of clinical laboratories derive LDL-C by the Friedewald estimation from TC, TGs and HDL-C. The Friedewald equation is based on the assumption that the average VLDL-C is equal to one-fifth of total TGs measured in fasting samples. Therefore, Equation (6)

$$\text{LDL-C} = \text{TC} - \text{HDL-C} - \frac{\text{TG}}{5} \qquad (6)$$

Although acceptable for routine use in samples with TGs up to 400 mg dL^{-1} (4.52 mmol L^{-1}), LDL-C obtained by the Friedewald equation may begin to deviate from the results obtained by beta quantification when the fasting TGs are above 250 mg dL^{-1} (2.82 mmol L^{-1}).[112,113] As a rule, the Friedewald equation is not suitable for use in samples containing chylomicrons (e.g. nonfasting samples, type I and type V hyperlipidemia) and in type III hyperlipidemia because the assumption that VLDL-C = TG/5 does not apply. The Friedewald estimation is convenient and economical for routine laboratory use when TC, HDL-C and TG values are available. However, it includes the additive analytical error of the above tests and assumes that the patient has truly fasted.

4.4.4 Direct Measurements

Numerous attempts were made to measure LDL-C directly bypassing physical separation steps, such as ultracentrifugation. However, the various attempts including chemical precipitation[114] and the immunoseparation methods[115,116] have not always been entirely satisfactory. Several direct homogeneous LDL-C methods have been introduced as commercial kits.[117] These methods do not involve a pretreatment step, and are generally based on selective removal of the non-LDL particles by antibodies that specifically recognize Apos, or are based on the selective blocking of non-LDL lipoproteins followed by a LDL-selective second reagent. These methods may be automated on routine chemistry analyzers. The advantages of the homogeneous assay include the acceptance of nonfasting samples,[118] improved precision and significant labor savings. However, the homogeneous methods present a greater challenge to consistent standardization because LDL-C standards must be employed rather than simply cholesterol standards. Early results are encouraging. Long-term performance, as well as the performance in samples of unusual lipidemic and disease states remains to be established.

4.4.5 Electrophoresis

Serum lipoproteins may be separated into VLDL, LDL and HDL by polyacrylamide or agarose gel electrophoresis followed by quantification of the cholesterol with an enzymatic substrate system.[23,65] The gel composition of some products has been modified to allow separation and possible quantification of Lp(a). The electrophoretic resolution and accurate quantification of LDL and VLDL lipoprotein by electrophoresis depends on prompt analysis of the serum sample after collection. Samples with greatly elevated TGs may have poor resolution because the VLDL band trails into the LDL region. Currently the combination of ultracentrifugation and electrophoresis is the ''reference method'' for β-VLDL determination. β-VLDL particles, which are present in the <1.006 g mL^{-1} fraction and found in type III hyperlipidemia, have a β-migrating electrophoretic mobility.

4.5 High-density Lipoprotein Cholesterol

Like other major lipoproteins, HDL consists of a complex family of particles with varying lipid and protein compositions. HDL is the smallest and most dense lipoprotein, and has the highest protein-to-lipid ratio. This lipoprotein is measured by its cholesterol content, or less often by its major Apo, Apo A-I. NCEP performance guidelines for HDL-C are summarized in Warnick and Wood,[119] and in Table 2.

4.5.1 Ultracentrifugation

HDL can be separated from other lipoproteins by ultracentrifugation following adjustment of the serum density with NaBr or KBr salts to $1.063\,g\,mL^{-1}$.[120] The HDL-containing bottom fraction is quantified by its cholesterol content. The physical separation of HDL by ultracentrifugation is used primarily in research settings.

4.5.2 Chemical Precipitation Methods

The most commonly used HDL-C method in the routine laboratory has been selective chemical precipitation using a mixture of polyanion and divalent cation to remove apo B-containing lipoproteins. For this procedure, the treated serum or plasma is centrifuged and the clear supernatant is analyzed for cholesterol using a cholesterol method (usually enzymatic) optimized for the HDL-C concentration range. The most common precipitation procedures include heparin–manganese chloride,[121] dextran sulfate–magnesium chloride,[59] phosphotungstate–magnesium chloride,[60] and polyethylene glycol.[61] A number of variations of each reagent system have been developed to improve assay performance. Because of the heterogeneous nature of HDL particles, total concordance among methods is difficult if not impossible to achieve. When a precipitation procedure is employed, it is important that turbid supernatants, which indicate incomplete sedimentation of LDL and VLDL, be clarified by dilution, or by ultrafiltration using a 0.22-μm filter before cholesterol analysis in order to avoid overestimation.[59] Alternatively, the serum may be ultracentrifuged at serum denstiy to remove the VLDL and the $>1.006\,g\,mL^{-1}$ bottom fraction is processed as described above.

4.5.3 United States Centers for Disease Control and Prevention Reference Method

As with other lipoprotein measurements, the reliability of HDL-C measurement depends on both the performance of the lipoprotein separation and cholesterol quantification steps. The CDC reference method stipulates the exact conditions for both of these steps. It has served as an accuracy target for numerous population and epidemiological studies as well as the CRMLN. The HDL-C reference method, which combines ultracentrifugation, heparin–manganese chloride precipitation of the ultracentrifuge bottom fraction, and Abell–Kendall cholesterol analysis of the supernatant, is included in the CDC reference beta quantification method outlined above (see section 4.4).

4.5.4 Designated Comparison Method

A designated comparison method (DCM)[62] for HDL-C was developed by the CRMLN to provide a more practical accuracy base than the ultracentrifugation reference method. The latter requires large sample volumes, an ultracentrifuge, and is labor intensive and costly. The DCM requires a 1-mL sample and employs direct precipitation of the serum by dextran sulfate–magnesium chloride without prior separation of the HDL. However, the method is only recommended for samples with TGs less than $200\,mg\,dL^{-1}$ ($5.18\,mmol\,L^{-1}$). Samples with elevated TGs may produce incomplete precipitation of non-HDL lipoproteins, which necessitates an ultrafiltration step. The DCM is standardized to the CDC reference method. It is utilized as a reference method by the CRMLN in its HDL-C certification program.

Briefly, the dextran sulfate–magnesium chloride precipitating reagent is prepared from 50-kD dextran sulfate (Dextralip 50; Sochibo, France), reagent grade magnesium chloride ($MgCl_2 \cdot 6H_2O$) and sodium azide (NaN_3). Analysis begins with addition of 0.1 mL of the working reagent to 1.0 mL of serum. After incubation at room temperature for 10–30 min and centrifugation at $1500 \times g$ for 30 min at 4 °C, the cholesterol in the clear supernatant is measured by the reference Abell–Kendall method.

4.5.5 Homogeneous Methods

Several direct methods for HDL-C measurement have become available as commercial kits and have gained increasing acceptance.[64,122,123] These direct homogeneous methods use a small sample (10–40 μL), do not require pretreatment or prior separation steps, and can be fully automated in routine clinical chemistry analyzers. Precision values are generally excellent. Long-term accuracy, however, is dependent on the stability and correct value assignment of the HDL-C calibrators provided by the manufacturers (in contrast to the much simpler cholesterol calibrators used in the chemical precipitation methods). The homogeneous methods usually involve a first reagent which blocks the measurement of cholesterol in non-HDL lipoproteins, followed by a second reagent for HDL-C measurement.

Although extremely promising and acceptable as routine methods, the new homogeneous methods may not yield accurate results in some disease states and in samples with greatly elevated TGs.[123–125] These new methods will undoubtedly be improved with time and their overall limitations remain to be established.

4.5.6 Electrophoresis

HDL may also be separated by cellulose acetate, agarose or polyacrylamide gel electrophoresis and stained with

lipid stains such as Oil Red O or Sudan Black. The lipid bands including HDL can also be quantified by cholesterol content with an enzymatic cholesterol reagent system.[23,65]

4.6 Triglycerides

Fatty acids are the most concentrated source of biochemical energy, yielding approximately 9 kcal g^{-1} (38 kJ g^{-1}), more than twice that obtained from the metabolism of carbohydrates or proteins. As described above, fatty acids are stored and transported in the form of triacylglycerols (TGs) primarily in association with chylomicrons and VLDL.

The general approach for measurement of TGs for clinical purposes is as follows:

- TGs $\longrightarrow$ glycerol + fatty acids
- Glycerol $\longrightarrow$ intermediate
- Intermediate $\longrightarrow$ colored or ultraviolet reporter.

Although there is a definitive method established at the NIST and a recommended reference method performed at the CDC, generally, enzymatic assays are considered to be the current state-of-the-art. Enzymatic hydrolysis by a lipase to give glycerol must be quantitative, and the fatty acid composition of the TGs must be estimated. The chromotropic acid reference method used at the CDC employs synthetic standards containing a 2 : 1 weight ratio of triolein and tripalmitin. Because TG results for clinical purposes are reported in the United States in milligrams per deciliter, a difference of ~10% results if the molecular weight of tripalmitin were substituted for the weight average molecular weight of the standard triolein/tripalmitin mixture. Similarly, if the true weight ratios of the fatty acids in an actual specimen are different from those in the synthetic standards, errors result. This problem is avoided when TG concentrations are reported in millimols per liter, the Système International (SI) units.

Because up to about 3% of the acylglycerols (glycerides) in plasma are mono- and diglycerides, a small variable error occurs that is tacitly accepted. On the other hand, a variable amount of glycerol is present in serum which is also being included in the ''total TGs'' estimation in the routine laboratory. In the average serum sample, endogenous glycerol amounts to the equivalent of 5–20 mg dL^{-1} (0.06–0.22 mmol L^{-1}) TGs. Specimens from patients with certain conditions, such as in diabetes, during emotional stress or during administration of some drugs may have elevated endogenous glycerol. It is advisable to measure endogenous glycerol in such samples by omitting lipase from the reagent system and subtracting the free glycerol thus obtained from the total TG to arrive at the glycerol-blanked TG value.

Because of the wide clinical reference range of TGs [NCEP defines <200 mg dL^{-1} (2.26 mmol L^{-1}) as normal, 200–400 mg dL^{-1} (2.26–4.52 mmol L^{-1}) as borderline high and >400 mg dL^{-1} (4.52 mmol L^{-1}) as high], accurate TG determination does not have significant impact on patient care except for its role in the LDL-C estimation in the Friedewald equation.[67] However, with the advent of direct LDL assays, the usage of the Friedewald equation will decline with time. Thus, the need for correcting for endogenous free glycerol in the average patient population becomes less crucial.

4.6.1 Analytical Methodologies[126]

4.6.1.1 Definitive Method The definitive method for TG measurement is the ID/MS procedure developed by NIST.[69] This method is actually two ID/MS methods; one for total glycerides (sum of TGs, diglycerides, monoglycerides and free glycerol) and another for TGs. Currently, the vast majority of clinical laboratories measure total glycerides.

4.6.1.2 Reference Method The current recommended reference method is the chromotropic acid method preformed by the CDC. This procedure measures TGs as well as a small fraction of the mono- and diglycerides, but the mono- and diglycerides amounts to less than 1–3% of the total glyceride value. Free glycerol and phospholipids are removed in the silicic acid step of this procedure. The reaction sequence is summarized below:[70,71]

1. extraction of TG with organic solvent (e.g. chloroform);
2. removal of phospholipids by absorption on absorbent material (e.g. silicic acid);
3. saponification with alcoholic KOH to fatty acids and glycerol;
4. oxidation of glycerol, Equation (7):

$$\text{glycerol} + IO_4^- \longrightarrow \underset{\text{formaldehyde}}{HCHO + HCOOH} + H_2O + IO_3^- \qquad (7)$$

5. detection:
   (a) Hantzsch reaction, Scheme 1
   (b) Chromotropic acid (Eegriwe) reaction, Scheme 2

4.6.1.3 Routine (Enzymatic) Methods[72–74] In the clinical laboratory, serum TGs are determined using enzymatic (lipase) hydrolysis of the acylglycerols, and coupled enzymatic reactions to produce a chromophore from glycerol. The majority of clinical laboratories

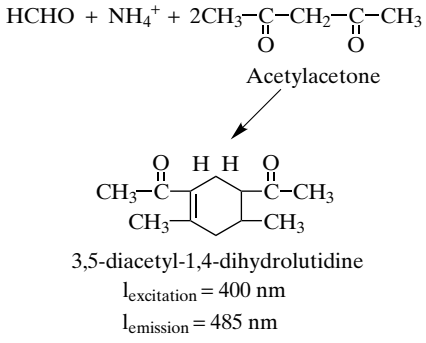

Scheme 1

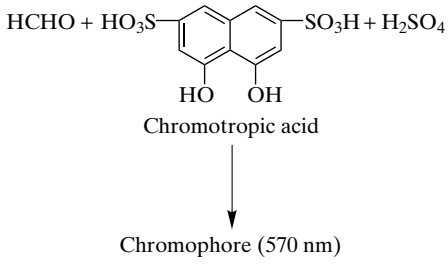

Scheme 2

measure all forms of glycerides (tri, di and mono) as well as endogenous free glycerol. Important examples are summarized, Equations (8–19):

1. $\text{acylglycerides} \xrightarrow{\text{lipase}} \text{glycerol} + \text{fatty acid}$ (8)

2. chromogenic step:

   (a) NADH (nicotinamide adenine dinucleotide (reduced form)) disappearance

$$\text{glycerol} + \text{ATP} \xrightarrow{\text{glycerol kinase}} \text{glycerol-3-phosphate} + \text{ADP} \quad (9)$$

$$\text{ADP} + \text{phosphoenolpyruvate} \xrightarrow{\text{pyruvate kinase}} \text{ATP} + \text{pyruvate} \quad (10)$$

$$\text{pyruvate} + \text{NADH} + \text{H}^+ \xrightarrow{\text{lactate dehydrogenase}} \text{lactate} + \text{NAD}^+ \quad (11)$$

   where NAD^+ is nicotinamide adenine dinucleotide (oxidized form)

   (b) NADH appearance

$$\text{glycerol} + \text{ATP} \xrightarrow{\text{glycerol kinase}} \text{glycerol-3-phosphate} + \text{ADP} \quad (12)$$

$$\text{glycerol-3-phosphate} + \text{NAD}^+ \xrightarrow{\text{glycerol-P dehydrogenase}} \text{dihydroxyacetone phosphate} + \text{NADH} + \text{H}^+ \quad (13)$$

   (c) quinone dye formation

$$\text{glycerol} + \text{ATP} \xrightarrow{\text{glycerol kinase}} \text{glycerol-3-phosphate} + \text{ADP} \quad (14)$$

$$\text{glycerol-3-phosphate} + \text{O}_2 \xrightarrow{\text{glycerol-P oxidase}} \text{dihydroxyacetone} + \text{H}_2\text{O}_2 \quad (15)$$

$$\text{H}_2\text{O}_2 + \text{3,5-dichloro-2-hydroxybenzene sulfonate} + \text{4-aminophenazone} \xrightarrow{\text{peroxidase}} \text{quinone dye (chromophore 510 nm)} \quad (16)$$

   (d) formazan dye formation

$$\text{glycerol} + \text{ATP} \xrightarrow{\text{glycerol kinase}} \text{glycerol-3-phosphate} + \text{ADP} \quad (17)$$

$$\text{glycerol-3-phosphate} + \text{NAD}^+ \xrightarrow{\text{glycerol-P dehydrogenase}} \text{dihydroxyacetone phosphate} + \text{NADH} + \text{H}^+ \quad (18)$$

$$\text{NADH} + \text{H}^+ + \text{oxidize tetrazolium} \xrightarrow{\text{diaphorase}} \text{reduced tetrazolium (chromophore 500–590 nm)} \quad (19)$$

Specimens for TG analysis may be serum, EDTA plasma or heparin plasma. Specimens should be refrigerated at 2–8 °C if analyzed within five days. For longer storage, freezing below −15 °C in a tightly closed container is necessary. To eliminate chylomicrons from the specimen, the patient should be fasting, ≥12 h. If postprandial specimens must be analyzed they should be well mixed immediately before analysis to accommodate expected chylomicron flotation. It should be noted that specimens from heparin-treated patients are expected to contain high levels of glycerol resulting from heparin activation of vascular lipoprotein lipase.

5 APOLIPOPROTEINS

5.1 Apolipoprotein A-I

Apo A-I is synthesized in both the gut and liver, and is secreted in the phospholipid-rich nascent HDL particles.

It performs a structural role, is the major activator of LCAT and, therefore, has a major role in "reverse cholesterol transfer" (see Kwiterovich[127] for review).

5.1.1 Standard Reference Method and Material

Apo A-I assay accuracy and precision are dependent on the availability of a pure stable reference material. Apo A-I is the first Apo for which a primary reference method has been certified. The CDC has validated an ID/MS characterization for certifying a primary reference material. A purified and lyophilized Apo A-I from the European Community Bureau of Reference, BCR-CRM 393, has been certified.[128] The WHO, based on collaborative work done under the auspices of the IFCC, and traceable to the certified reference material, selected a well-characterized lyophilized serum pool, SP1-01,[129] for the Apo A-I secondary (International) reference material. The standardization of all commercially available assays should be traceable to this material.

5.1.2 Routine Assays for Apolipoprotein A-I

Historically significant immunoassay methods that have been developed for Apo A-I include rocket electroimmunoassay, radial immunodiffusion (RID), and radioimmunoassay (RIA). Most routine laboratories use immunoturbidimetric or nephelometric assays, although enzyme-linked immunosorbent assays (ELISAs) are also available. The light-scattering assays can be performed precisely and accurately when calibrated against the WHO–IFCC international reference material. These assays, however, are susceptible to interference from sample turbidity (e.g. lipemia). This may be in part overcome by the use of detergents or lipase.[130] An additional problem with these assays is immune complex instability. This difficulty can be largely overcome by using techniques like latex particle enhancement.[131] Finally, it should be noted that production of antisera for use in Apo A-I immunoassay has been hampered by the lipophilic character of most of this protein. This leads to concealment of many potential immunoreactive sites from immunological response.[130]

5.2 Apolipoprotein B

Apo B is required for hepatic and enteric secretion of TG-rich lipoproteins. It plays a central role in chylomicron, VLDL, and LDL particle metabolism, each of which has one copy of Apo B. The Apo B in the intestinally secreted chylomicron is a truncated form of the protein, Apo B-48, while VLDL and LDL contain the larger Apo B-100. Approximately 90% of the latter are in LDL. Thus, it is closely associated with the development of atherosclerotic lesions and serves as a surrogate for measurement of LDL cholesterol.[132] It is a large molecule (550 kD), and is more hydrophobic than other Apos. During the removal of TG from its core, VLDL is progressively converted to IDL and LDL, and at the same time, a single LDL receptor-binding site becomes exposed on the lipoprotein particle surface.

5.2.1 Standard Reference Method and Material

Production of an Apo B primary reference material has been problematic because it is difficult to produce a stable lyophilized material.[133,134] However, a fresh frozen secondary reference material has been developed by comparison with Apo B isolated from freshly prepared LDL. This material, SP3-07, was evaluated by an IFCC-administered program and accepted by the WHO as the international reference material.[13] A validated ID/MS spectrometric characterization for certifying this material has not been done.

5.2.2 Routine Assays for Apolipoprotein B

As with Apo A-I, Apo B has been analyzed by immunoassay methods that include rocket electroimmunoassay, RID, RIA, and ELISA. Most routine laboratories employ immunoturbidimetric or nephelometric assays. These assays can be performed precisely and accurately, when calibrated against the WHO–IFCC international reference material. As already mentioned for Apo A-I, assays that measure immune complex formation are susceptible to interference from sample turbidity (e.g. lipemia). This may be in part overcome by the use of detergents or lipase.[130] Complex instability problems can be largely overcome by using techniques like latex particle enhancement.[131]

Apo B-48 is of interest in conditions related to postprandial chylomicronemia. This truncated form of the Apo has all of the same epitopes of the N-terminal 48% of Apo B-100. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE), staining and densitometry may be used to estimate it.[135] However, for routine determination, an ELISA has been developed.[136] An indirect approach to the measurement of Apo B-48 is the measurement of chylomicron remnants (see section 7.2 on remnant lipoproteins).

5.3 Apolipoprotein C-II and Apolipoprotein C-III

These two small Apos are found in chylomicrons, VLDL and HDL and play roles in regulation of the extrahepatic lipoprotein lipase. Apo C-II activates the lipase enhancing the transfer of TGs to adipose and muscle tissue. On the other hand, Apo C-III seems to inhibit the hepatic uptake of TG-rich particles.

5.3.1 Standard Reference Material

Although no official reference material is available for use in standardizing Apo C-II and Apo C-III assays, these proteins have been isolated by chromatography and electrofocusing from VLDL prepared by ultracentrifugation.[137] There does not seem to be enough interest among clinicians in the measurement of these Apos to warrant an international effort to develop primary and secondary reference materials.

5.3.2 Assays for Apolipoprotein C-II and Apolipoprotein C-III

These proteins are less hydrophobic than other Apos, and therefore antisera and monoclonal antibodies to them are somewhat easier to prepare. RID[138] is a good method for use in research studies or clinical diagnosis when relatively few samples need to be assayed. When larger numbers of specimens must be routinely analyzed, RIAs[139] or ELISAs[140–142] can be used. In addition, commercially available immunoturbidimetric assays have been developed.[143] The drawback to the latter assay is the potential for overestimation of results for turbid (lipemic) samples. This may be overcome, as described for the immunoturbidimetric assay of the other Apos, by use of detergents or lipase enzymes.

5.4 Apolipoprotein E

Apo E is a relatively small Apo that is secreted from the liver in VLDL particles. It is also found in chylomicrons, chylomicron remnants, and some subclasses of HDL, but only a very small amount is present in LDL. In addition to the hepatic synthesis, Apo E is synthesized in the central and peripheral nervous systems, and by macrophages. It is a ligand for the hepatic LDL receptor (the Apo B/E receptor) and probably for a second receptor (the LDL receptor related protein).

Apo E is found as a wild-type isoform, Apo E3, and as other two rather common isoforms, Apo E2 and Apo E4 (gene frequencies of 10% and 15%, respectively). The latter arise from two separate amino acid substitutions, causing these proteins to have altered affinity for Apo E receptors. Individuals with the E2/2 homozygous pattern of Apo E expression may have high levels of TC and TGs resulting from the accumulation of chylomicron and VLDL remnants. However, a relatively small fraction of them actually develop this type of hyperlipidemia.

The other condition associated with a common Apo E polymorphism is Alzheimer's disease. A high fraction of patients with Alzheimer's disease carry at least one Apo E4 allele.

5.4.1 Standard Reference Material

There is no official reference material for use in standardizing Apo E immunoassay. Purified Apo E for preparation of antigen for immunoassay development can be prepared by ultracentrifugation and electrofocusing.[142]

5.4.2 Assays for Apolipoprotein E

Apo E can be measured by RID, but 2–4 days is generally required.[138] RIA can be used, and, of course, it is more sensitive and rapid.[144] Enzyme immunoassays have also been developed, and are very sensitive and can be done quickly.[145] Finally, immunoturbidimetric assays have been developed and are commercially available.[146] As with other immunoturbidimetric assays, the results of this assay are compromised by lipemia. This interference is commonly eliminated by the use of detergents or lipase treatment.

The common Apo E phenotypes can be determined by various approaches to isoelectric focusing.[147–149] Genotyping can be done to give a more definite description of the genetics of Apo E isoforms. As with any genetic analyses, the reliability of Apo E genotyping requires successful DNA amplification. This can be difficult because of the very high guanine–cytosine content of the Apo E gene DNA. Once an acceptable polymerase chain reaction (PCR) product is obtained, the DNA may be analyzed by restriction fragment-length polymorphism analysis.[150] This technique usually employs polyacrylamide gel electrophoresis and visualization by ethidium bromide, but this time-consuming procedure can be replaced by automated capillary electrophoresis.[151]

6 LIPOPROTEIN PARTICLE SIZE

There is a growing body of evidence indicating that lipoprotein particle size, density and chemical composition are related to the risk of ischemic cardiovascular disease. The presence of subclasses of LDL and HDL particles has been known for some time, but with the development of semiroutine measurement techniques has come awareness of their clinical significance. It is now clear the small dense LDL is a risk factor for CHD (see Krauss[152] for a review). In addition, some HDL subclasses are correlated with the risk of CHD.[153]

6.1 Methods for Assessment of Low-density Lipoprotein Subclasses

Ultracentrifugation was first used to identify lipoprotein subclasses and this method may still be considered the

reference. One often-used approach employs a vertical rotor in a preparative ultracentrifuge.[154] One method for identifying LDL and other Apo B-containing lipoproteins uses a discontinuous salt gradient.[155]

Nondenaturing polyacrylamide gradient gel electrophoresis is the most commonly used method for determining LDL subclass size and concentration.[156] Gradients of approximately 2–14% polyacrylamide are used with the Pharmacia GE-4 or equivalent electrophoresis system.[157,158] Protein and synthetic polymer molecular weight standards are used as reference size markers. After staining, gels may be scanned and the size and amount of each subclass determined. Several LDL subclass classification schemes have been proposed. CHD risk is generally associated with subclasses below 256 Å. This LDL mobility is associated with ''small-dense LDL'' (''pattern B''), while subclasses associated with reduced mobility, the so-called ''large LDL'' (diameter > 260 Å; ''pattern A''), are associated with decreased CHD risk.[156,158]

6.2 Methods for Assessment of High-density Lipoprotein Subclasses

Sequential density ultracentrifugation is the classical method for determination of the primary HDL subclasses, HDL_2 and HDL_3.[105] This procedure has been modified to include a single centrifugation step before measurement of HDL_3-C. The HDL_3-C result combined with total HDL-C determination by standard precipitation techniques is used to determine HDL_2-C from the difference, i.e. HDL-C − HDL_3-C.[159]

Density gradient ultracentrifugation can also be used to separate the HDL subclasses in a single spin. This approach gives the major subclasses as well as some of the minor ones.[160–162] Careful attention to detail is required to obtain reproducible data, so this technique is usually reserved for use in small studies done in research laboratories.

Another approach to identifying and quantifying HDL subclasses is polyacrylamide gradient gel electrophoresis.[163] This technique identifies several minor subclasses as well as the major ones. Chromatographic methods can also be used to determine the major and some minor HDL subclasses.[164]

Routine identification and quantification of major HDL subclasses are done by chemical precipitation. Selective precipitation procedures take advantage of the HDL-subclass charge and size differences. One common method uses varying amounts of dextran sulfate and divalent cation to precipitate all the HDL in a serum aliquot, and to precipitate HDL_3 selectively from another aliquot. HDL_2 concentration is determined from the difference.[165]

6.3 Methods for Simultaneous Assessment of Low-density Lipoprotein and High-density Lipoprotein Subclasses

Both LDL and HDL subclasses may be determined simultaneously by gradient gel electrophoresis using a 4–30% polyacrylamide gradient gel that has been called a ''composite gel''.[166]

Another method for simultaneous quantification of lipoprotein classes employs a single short (<1 h) vertical ultracentrifugation step. A nonsegmented continuous flow (controlled-dispersion flow) analyzer is then used for enzymatic analysis of cholesterol in the lipoproteins present in the various fractions.[167]

A purely physical method of lipoprotein subclass analysis involves analysis of a nuclear magnetic resonance spectral fingerprint of plasma.[168] No separation of lipoprotein subclasses is employed. Rather, the methyl resonance spectrum produced by phospholipid in the outer shell of the lipoprotein in whole plasma is deconvoluted by a computerized algorithm that is based on a series of VLDL, IDL, LDL, and HDL reference spectra. This enables determination of the relative amounts of six VLDL, one IDL, four LDL, and five HDL subclasses, and allows calculation of plasma cholesterol and TG and LDL-C and HDL-C concentrations.

Specimens for lipoprotein subclass analysis should be drawn in EDTA after a 12-h fast. Plasma can be obtained by centrifugation for 20 min at 1500 *g*. Any chylomicrons must be removed by high-speed centrifugation before storage. Specimens may be stored at 4 °C for up to one week or −70 °C for longer term.

7 OTHER

7.1 Lipoprotein(a)

As described above, the atherogenic Lp(a) contains a single copy of Apo(a). Lp(a) can be quantified by measuring the total lipoprotein mass or by separating it from other lipoprotein particles and measuring its cholesterol content. The first approach requires that attention be paid to the molecular heterogeneity of Apo(a). Thus, measurement based on total Lp(a) mass, as opposed to stoichiometric measurement, may misrepresent the actual quantity of Lp(a) particles in persons carrying different isoforms of the protein.

7.1.1 Standard Reference Material

Apo(a) isolated for use in immunization animals for production of antisera for use in developing immunoassays is immunochemically different from Apo(a) in Lp(a).[169] Hence, immunoassays for Lp(a) are often developed

using the intact lipoprotein. There is currently no standardized reference material available for Apo(a), but an IFCC working group on Lp(a) is working to develop and characterize one.

7.1.2 Assays for Lipoprotein(a)

Several types of immunoassay procedure have been developed for measurement of Lp(a), including RIA,[170] nephelometric,[171] ELISA,[172,173] and immunoturbidimetric assay.[174] The latter two assays are most commonly performed. Heterogeneity in Apo(a) size can affect each of these assays. This problem is overcome in ELISA assays that use a monoclonal antibody that specifically recognizes an epitope in the region of Apo(a) that is not part of the repeat sequence found in most human Apo(a) isoforms. On the other hand, methods that rely strictly on polyclonal antibodies for recognition of Apo(a) cannot be used to assess the stoichiometric concentration of Lp(a) in serum samples.

Two assays have been developed that are based on separation of Lp(a) from other lipoproteins and estimation of the cholesterol in the isolated material. The first of these uses the binding of an agarose-bound lectin, wheat germ agglutinin, to Apo(a) to separate Lp(a) from other lipoproteins.[175] Whether the specificity of the all wheat germ agglutinin preparations for Apo(a) are equivalent, and if it binds all Apo(a) isoforms equally is not entirely clear. The assay does correlate very well with a common commercially available molar-based ELISA.

The second approach to specific measurement of Lp(a) cholesterol is more or less independent of Apo(a) immunochemistry. Thus, this procedure uses electrophoretic separation of serum components, followed by enzymatic cholesterol staining and densitometric scanning.[23,65] The method is highly automated, but separation of Lp(a) from other lipoprotein bands may be incomplete and proper selection of band area for quantification is subjective. Correlation of this method with existing commercial methods is fair.[23,65]

7.2 Triglyceride-rich Remnant Lipoproteins

TGs are primarily associated with VLDL and chylomicrons in plasma. These triglyceride-rich lipoproteins (TRLs) have been linked to pathogenesis of atherosclerosis.[176,177] However, the smaller partially catabolized TRL remnants are believed to be more atherogenic and therefore provide a better assessment of coronary artery disease risk than that provided by the measurement of total plasma TGs. Plasma TRLs of both intestinal and hepatic origins have been isolated using a number of biochemical methods and procedures based on TRL density, charge, size, lipid components and Apo composition.[178] However, because of the difficulties in isolating, characterizing and quantifying these TRLs, these methods have been primarily studied in research laboratories.

The methods most commonly used employ quantification of IDL-C in the $1.019\,g\,mL^{-1}$ ultracentrifugation-density fraction, or the measurement of β-VLDL by agarose gel electrophoresis of whole serum for comparison with that obtained from $d > 1.006\,g\,mL^{-1}$ and $d < 1.006\,g\,mL^{-1}$ ultracentrifugation-density fractions. Samples containing β-VLDL have a VLDL band with LDL-like electrophoretic mobility in the $d < 1.006\,g\,mL^{-1}$ fraction. A commercial reagent kit has been introduced that quantifies the remnant-like particles by their cholesterol content. This method is based on immunoseparation using an anti-Apo A-I and a specific Apo B-100 monoclonal antibody.[179,180] The method lends itself to automation and can be used in the routine laboratory.

7.3 Oxidized Low-density Lipoproteins and Lipid Markers of Oxidative Stress

Free radical-mediated oxidation of LDL is thought to be a major mediator in the development of atherosclerotic lesions. The presence of oxidized LDL particles in plasma has been reported to be measurable using an ELISA based on a monoclonal antibody to oxidized LDL.[181] However, it is not clear that this a viable approach for evaluating the oxidized LDL burden of a subject. A more reliable approach may be to measure the presence of autoantibodies that arise from exposure of Apo epitopes that are usually hidden from immune detection (see Craig et al.,[182] for example). Serum concentration of these antibodies has been linked to the development of coronary artery disease.[183]

The oxidative stress that an individual is exposed to depends on many factors, including nutritional, environmental, physiological and pathological conditions. Many approaches to the assessment of oxidative stress have been reported.[184,185] A promising approach for determining the overall oxidative stress of an individual is the measurement of a class of prostanoids that are formed from arachidonic acid as nonregulated, nonenzymatic free-radical reaction products. These compounds are called isoprostanes and they can be measured in plasma and urine by gas chromatography/mass spectrometry (GC/MS),[186] or immunoassay.[187] One group of isoprostanes, the F_2 isoprostanes, has been found to be elevated in individuals with hypercholesterolemia.[188]

7.4 Nonesterified Fatty Acids

Plasma nonesterified fatty acids (NEFAs or free fatty acids) have been linked to type 2 diabetes and obesity,

and indirectly with coronary artery disease. An increase in plasma NEFA is associated with insulin resistance and apparently results from failure of fatty acid transfer into adipose, muscle and other cells after hydrolysis by lipoprotein lipase.

NEFAs are analyzed by direct chemical titration,[189] HPLC,[190] gas chromatography,[191] and enzymatic analysis.[192] The latter procedure is simple and can be done on large or small scales, but it is somewhat less sensitive than other techniques and does not give information about NEFA species.

For NEFA assays to give useful information, care must be taken to prevent release of fatty acids as the result of TG hydrolysis after specimen collections. Blood specimens may be immediately transferred to ice, centrifuged at 4 °C, and stored at −70 °C until assay. Alternatively, specimens may be mixed with paraoxon, a cholinesterase inhibitor, immediately after collection.[192] However, even when the inhibitor is used, specimens should be rapidly cooled and stored at −70 °C until analysis.

ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immune Deficiency Syndrome
Apo	Apolipoprotein
CDC	Centers for Disease Control
CHD	Coronary Heart Disease
CRMLN	Cholesterol Reference Method Laboratory Network
CV	Coefficients of Variation
DCM	Designated Comparison Method
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Assay
EQAS	External Quality Assessment Scheme
FPLC	Fast-performance Liquid Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HDL	High-density Lipoprotein
HDL-C	High-density Lipoprotein Cholesterol
HIV-1	Human Immunodeficiency Virus-1
HPLC	High-performance Liquid Chromatography
IDL	Intermediate Density Lipoprotein
IDL-C	Intermediate Density Lipoprotein Cholesterol
ID/MS	Isotope Dilution/Mass Spectrometry
IFCC	International Federation for Clinical Chemistry
L–B	Liebermann–Burchard
LCAT	Lecithin–Cholesterol Acyltransferase
LDL	Low-density Lipoprotein
LDL-C	Low-density Lipoprotein Cholesterol
Lp(a)	Lipoprotein(a)
LRC	Lipid Research Clinics
NAD^+	Nicotinamide Adenine Dinucleotide (oxidized form)
NADH	Nicotinamide Adenine Dinucleotide (reduced form)
NCEP	National Cholesterol Education Program
NEFA	Nonesterified Fatty Acid
NHLBI	National Heart Lung and Blood Institute of the National Institutes of Health
NICAMS	National Initiative on Cholesterol Accuracy, Methods and Standardization
NIH	National Institutes of Health
NIST	National Institute for Standards and Technology
PCR	Polymerase Chain Reaction
RIA	Radioimmunoassay
RID	Radial Immunodiffusion
rpm	Revolutions Per Minute
RV	Reference Value
SD	Standard Deviation
SDS/PAGE	Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis
SI	Système International
TC	Total Cholesterol
TG	Triglyceride
TLC	Thin-layer Chromatography
TRL	Triglyceride-rich Lipoprotein
VLDL	Very Low-density Lipoprotein
VLDL-C	Very Low-density Lipoprotein Cholesterol
WHO	World Health Organization

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Phospholipids of Plasma Lipoproteins, Red Blood Cells and Atheroma, Analysis of • Product Development for the Clinical Laboratory • Statistical Quality Control in Clinical Laboratories

Food **(Volume 5)**
Lipid Analyses in Food

Particle Size Analysis **(Volume 6)**
Sedimentation in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
Chromatography of Membrane Proteins and Lipoproteins

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry

REFERENCES

1. 'The Lipid Research Clinics Coronary Primary Prevention Trial Results. I. Reduction in Incidence of Coronary Heart Disease', *JAMA*, **251**(3), 351–364 (1984).
2. 'The Lipid Research Clinics Coronary Primary Prevention Trial Results. II. The Relationship of Reduction in Incidence of Coronary Heart Disease to Cholesterol Lowering', *JAMA*, **251**(3), 365–374 (1984).
3. 'Summary of the Second Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults', *JAMA*, **269**(23), 3015–3023 (1993).
4. 'National Cholesterol Education Program. Second Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II)', *Circulation*, **89**(3), 1333–1445 (1994).
5. P.S. Bachorik, J.W. Ross, 'National Cholesterol Education Program Recommendations for Measurement of Low-density Lipoprotein Cholesterol: Executive Summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement', *Clin. Chem.*, **41**(10), 1414–1420 (1995).
6. E.A. Stein, G.L. Myers, 'National Cholesterol Education Program Recommendations for Triglyceride Measurement: Executive Summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement', *Clin. Chem.*, **41**(10), 1421–1426 (1995).
7. G.R. Warnick, P.D. Wood, 'National Cholesterol Education Program Recommendations for Measurement of High-density Lipoprotein Cholesterol: Executive Summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement', *Clin. Chem.*, **41**(10), 1427–1433 (1995).
8. National Institutes of Health, National Heart, Lung and Blood Institute Working Group on Lipoprotein Measurement, 'Recommendations on Lipoprotein Measurement', in *NIH Publication No. 95-3044*, US Government Printing Office, 1995.
9. G.L. Myers, G.R. Cooper, C.L. Winn, S.J. Smith, 'The Centers for Disease Control–National Heart, Lung and Blood Institute Lipid Standardization Program. An Approach to Accurate and Precise Lipid Measurements', *Clin. Lab. Med.*, **9**(1), 105–135 (1989).
10. G.L. Myers, G.R. Cooper, E.J. Sampson, 'Traditional Lipoprotein Profile: Clinical Utility, Performance Requirement, and Standardization', *Atherosclerosis*, **108** (Suppl), S157–S169 (1994).
11. S. Marcovina, L.K. Curtiss, R. Milne, J.J. Albers, 'International Federation of Clinical Chemistry (IFCC) Scientific Division, Committee on Apolipoproteins, Working Group on Antibody Reagents. Selection and Characterization of Monoclonal Antibodies for Measuring Plasma Levels of Apolipoproteins A-I and B', *Ann. Biol. Clin. (Paris)*, **48**(8), 597–600 (1990).
12. S.M. Marcovina, J.L. Adolphson, M. Parlavecchia, J.J. Albers, 'Effects of Lyophilization of Serum on the Measurement of Apolipoproteins A-I and B', *Clin. Chem.*, **36**(2), 366–369 (1990).
13. S.M. Marcovina, J.J. Albers, H. Kennedy, J.V. Mei, L.O. Henderson, W.H. Hannon, 'International Federation of Clinical Chemistry Standardization Project for Measurements of Apolipoproteins A-I and B. IV. Comparability of Apolipoprotein B Values by Use of International Reference Material', *Clin. Chem.*, **40**(4), 586–592 (1994).
14. R.J. Havel, 'The Formation of LDL: Mechanisms and Regulation', *J. Lipid Res.*, **25**(13), 1570–1576 (1984).
15. M.J. Caslake, C.J. Packard, 'The Use of Ultracentrifugation for Separation of Lipoproteins', in *Handbook of Lipoprotein Testing*, eds. N. Rifai, G.R. Warnick, M.H. Dominiczak, AACC Press, Washington, 509–529, 1997.
16. G.L. Mills, P.A. Lane, P.K. Weech, 'The Isolation and Purification of Plasma Lipoproteins', in *Laboratory Techniques in Biochemistry and Molecular Biology: A Guidebook To Lipoprotein Technique*, eds. R.H. Burdon, P.H. Knippenberg, Elsevier, Amsterdam, 18–116, 1984.
17. F.T. Lindgren, L.C. Jensen, F.T. Hatch, 'The Isolation and Quantitative Analysis of Serum Lipoprotein', in *Blood Lipids and Lipoproteins: Quantitation, Composition and Metabolism*, ed. G.J. Nelson, Wiley Interscience, NY, 181–274, 1972.
18. W. Patsch, J.R. Patsch, G.M. Kostner, S. Sailer, H. Bransteiner, 'Isolation of Subfractions of Human Very Low Density Lipoproteins by Zonal Ultracentrifugation', *J. Biol. Chem.*, **49**, 11–15 (1978).
19. R.M. Krauss, D.H. Burke, 'Identification of Multiple Subclasses of Plasma Low Density Lipoproteins', *J. Lipid Res.*, **23**, 97–104 (1982).
20. B.A. Griffin, M.J. Caslake, B. Yip, G.W. Tait, C.J. Packard, J. Shepherd, 'Rapid Isolation of Low Density Lipoprotein (LDL) Subfractions from Plasma by Density Gradient Ultracentrifugation', *Atherosclerosis*, **82**, 59–67 (1990).
21. Lipid Research Clinics Program, *Manual of Laboratory Operations. Vol. 1: Lipid and Lipoprotein Analysis*, National Institutes of Health Publication No. 75-628, Bethesda, US DHEW (NIH), 51–59, 1975.

22. E.H. Schaefer, R.I. Levy, 'Pathogenesis and Management of Lipoprotein Disorders', *N. Engl. J. Med.*, **312**, 1300–1310 (1985).
23. J.H. Contois, R.G. Gillmor, R.E. Moore, L.R. Contois, J.L. Macer, A.H. Wu, 'Quantitative Determination of Cholesterol in Lipoprotein Fractions by Electrophoresis', *Clin. Chim. Acta*, **282**(1–2), 1–14 (1999).
24. K.A. Kieft, T.M.A. Bocan, B.R. Krause, 'Rapid On-line Determination of Cholesterol Distribution Among Plasma Lipoproteins after High-performance Gel Filtration Chromatography', *J. Lipid Res.*, **32**, 859–866 (1991).
25. W. März, R. Siekmeier, H. Scharnagl, U.B. Seiffert, W. Gross, 'Fast Lipoprotein Chromatography: New Method of Analysis for Plasma Lipoproteins', *Clin. Chem.*, **39**, 2276–2281 (1993).
26. T. Kitamura, S. Ito, H. Moriyama, Y. Kato, K. Sasamoto, M. Okazaki, 'Quantitative Analysis of Serum Lipoproteins (CM, VLDL, LDL and HDL) by High-performance Gel Filtration Chromatography', *Chromatography*, **17**, 33–37 (1996).
27. D.S. Freedman, J.B. Croft, A.J. Anderson, T. Byers, S.J. Jacobsen, H.W. Gruchow, J.A. Walker, J.J. Barboriak, 'The Relation of Documented Coronary Artery Disease to Levels of Total Cholesterol and High-density Lipoprotein Cholesterol', *Epidemiology*, **5**(1), 80–87 (1994).
28. K.J. Harjai, 'Potential New Cardiovascular Risk Factors: Left Ventricular Hypertrophy, Homocysteine, Lipoprotein(a), Triglycerides, Oxidative Stress, and Fibrinogen', *Ann. Intern. Med.*, **131**(5), 376–386 (1999).
29. L.L. Wu, 'Review of Risk Factors for Cardiovascular Diseases', *Ann. Clin. Lab. Sci.*, **29**(2), 127–133 (1999).
30. B. Kinosian, H. Glick, L. Preiss, K.L. Puder, 'Cholesterol and Coronary Heart Disease: Predicting Risks in Men by Changes in Levels and Ratios', *J. Investig. Med.*, **43**(5), 443–450 (1995).
31. D.L. Sprecher, 'Triglycerides as a Risk Factor for Coronary Artery Disease', *Am. J. Cardiol.*, **82**(12a), 49u–56u (1998).
32. C.J. Glueck, M. Tieger, R. Kunkel, T. Hamer, T. Tracy, J. Speirs, 'Hypocholesterolemia and Affective Disorders', *Am. J. Med. Sci.*, **308**(4), 218–225 (1994).
33. A.M. Nomura, G.N. Stemmermann, P.H. Chyou, 'Prospective Study of Serum Cholesterol Levels and Large-bowel Cancer', *J. Natl. Cancer Inst.*, **83**(19), 1403–1407 (1991).
34. B.E. Kreger, K.M. Anderson, A. Schatzkin, G.L. Splansky, 'Serum Cholesterol Level, Body Mass Index, and the Risk of Colon Cancer. The Framingham Study', *Cancer*, **70**(5), 1038–1043 (1992).
35. N.J. Wald, S.G. Thompson, M.R. Law, J.W. Densem, A. Bailey, 'Serum Cholesterol and Subsequent Risk of Cancer: Results from the BUPA Study', *Br. J. Cancer*, **59**(6), 936–938 (1989).
36. E. Windler, U. Ewers-Grabow, J. Thiery, A. Walli, D. Seidel, H. Greten, 'The Prognostic Value of Hypocholesterolemia in Hospitalized Patients', *Clin. Invest.*, **72**(12), 939–943 (1994).
37. R.A. Breslow, J. Hallfrisch, A.P. Goldberg, 'Malnutrition in Tubefed Nursing Home Patients with Pressure Sores', *J. Parenteral. Enteral. Nutr.*, **15**(6), 663–668 (1991).
38. R.T. Chlebowski, M. Grosvenor, L. Lillington, J. Sayre, G. Beall, 'Dietary Intake and Counseling, Weight Maintenance, and the Course of HIV Infection', *J. Am. Diet. Assoc.*, **95**(4), 428–432 (1995).
39. Y. Benhamou, I. Hilmarsdottir, I. Desportes-Livage, C. Hoang, A. Datry, M. Danis, M. Gentilini, P. Opolon, 'Association of Lipid Accumulation in Small Intestinal Mucosa with Decreased Serum Triglyceride and Cholesterol Levels in AIDS', *Dig. Dis. Sci.*, **39**(10), 2163–2169 (1994).
40. P. Libby, 'The Interface of Atherosclerosis and Thrombosis: Basic Mechanisms', *Vasc. Med.*, **3**(3), 225–229 (1998).
41. J.W. Heinecke, 'Mechanisms of Oxidative Damage by Myeloperoxidase in Atherosclerosis and Other Inflammatory Disorders', *J. Lab. Clin. Med.*, **133**(4), 321–325 (1999).
42. S. Guthikonda, W.G. Haynes, 'Homocysteine as a Novel Risk Factor for Atherosclerosis', *Curr. Opin. Cardiol.*, **14**(4), 283–291 (1999).
43. J.D. Brunzell, J.E. Hokanson, 'Dyslipidemia of Central Obesity and Insulin Resistance', *Diabetes Care*, **22** (Suppl 3), C10–C13 (1999).
44. B.V. Howard, 'Insulin Resistance and Lipid Metabolism', *Am. J. Cardiol.*, **84**(1a), 28j–32j (1999).
45. D.J. Chisholm, L.V. Campbell, E.W. Kraegen, 'Pathogenesis of the Insulin Resistance Syndrome (Syndrome X)', *Clin. Exp. Pharmacol. Physiol.*, **24**(9–10), 782–784 (1997).
46. A.H. Wu, M.C. Pike, D.O. Stram, 'Meta-Analysis: Dietary Fat Intake, Serum Estrogen Levels, and the Risk of Breast Cancer', *J. Natl. Cancer Inst.*, **91**(6), 529–534 (1999).
47. S.L. Rywik, T.A. Manolio, A. Pajak, W. Piotrowski, C.E. Davis, G.B. Broda, E. Kawalec, 'Association of Lipids and Lipoprotein Level with Total Mortality and Mortality Caused by Cardiovascular and Cancer Diseases (Poland and United States Collaborative Study on Cardiovascular Epidemiology)', *Am. J. Cardiol.*, **84**(5), 540–548 (1999).
48. A. Cohen, H.S. Hertz, J. Mandel, R.C. Paule, R. Schaffer, L.T. Sniegoski, T. Sun, M.J. Welch, E. White, 5th, 'Total Serum Cholesterol by Isotope Dilution/Mass Spectrometry: A Candidate Definitive Method', *Clin. Chem.*, **26**(7), 854–860 (1980).
49. R. Schaffer, L.T. Sniegoski, M.U. Welch, V.E. White, A. Cohen, H.S. Hertz, J. Mandel, R.C. Paule, L. Svensson, I. Bjorkhem, R. Blomstrand, 'Comparison of Two

Isotope Dilution/Mass Spectrometric Methods for Determination of Total Serum Cholesterol', *Clin. Chem.*, **28**(1), 5–8 (1982).
50. R. Kock, B. Delvoux, H. Greiling, 'Determination of Total Cholesterol in Serum by Liquid Chromatography–Isotope Dilution Mass Spectrometry', *Clin. Chem.*, **43**(10), 1896–1903 (1997).
51. L.L. Abell, B.B. Levy, B.B. Brody, R.C. Kendall, 'A Simplified Method for the Estimation of Total Cholesterol in Serum and Demonstration of its Specificity', *J. Biol. Chem.*, **195**, 357–366 (1952).
52. G.R. Cooper, S.J. Smith, I.W. Duncan, A. Mather, W.D. Fellows, T. Foley, I.D. Frantz, Jr, J.B. Gill, T.A. Grooms, I. Hynie, et al., 'Interlaboratory Testing of the Transferability of a Candidate Reference Method for Total Cholesterol in Serum', *Clin. Chem.*, **32**(6), 921–929 (1986).
53. W. Richmond, 'Preparation and Properties of a Cholesterol Oxidase from *Nocardia* Sp. and its Application to the Enzymatic Assay of Total Cholesterol in Serum', *Clin. Chem.*, **119**(12), 1350–1356 (1973).
54. C.C. Allain, L.S. Poon, C.S. Chan, W. Richmond, P.C. Fu, 'Enzymatic Determination of Total Serum Cholesterol', *Clin. Chem.*, **20**(4), 470–475 (1974).
55. J. Siedel, E.O. Hagele, J. Ziegenhorn, A.W. Wahlefeld, 'Reagent for the Enzymatic Determination of Serum Total Cholesterol with Improved Lipolytic Efficiency', *Clin. Chem.*, **29**(6), 1075–1080 (1983).
56. K.W. Weingand, C.L. Hudson, L. Odioso, 'Enzymatic Reagents for Quantifying Total Plasma Cholesterol from Dogs, Rats, and Hamsters', *Clin. Chem.*, **36**(11), 1995 (1990).
57. D.A. Wiebe, G.R. Warnick, 'Measurement of High-density Lipoprotein Cholesterol', in *Handbook of Lipoprotein Testing*, eds. N. Rifai, G.R. Warnick, M.H. Dominiczak, AACC Press, Washington, 127–144, 1997.
58. G.L. Meyers, D.J. Hassemer, 'Standardization of Lipid and Lipoprotein Measurements', in *Methods for Clinical Laboratory Measurement of Lipid and Lipoprotein Risk Factors*, eds. N. Rifai, G.R. Warnick, AACC Press, Washington, 101, 1991.
59. G.R. Warnick, J. Benderson, J.J. Albers, 'Dextran Sulfate–Mg^{2+} Precipitation Procedure for Quantitation of High-density Lipoprotein Cholesterol', *Clin. Chem.*, **28**(6), 1379–1388 (1982).
60. M.F. Lopes-Virella, P. Stone, S. Ellis, J.A. Colwell, 'Cholesterol Determination in High-density Lipoproteins Separated by Three Different Methods', *Clin. Chem.*, **23**(5), 882–884 (1977).
61. C.J. Briggs, D. Anderson, P. Johnson, T. Deegan, 'Evaluation of the Polyethylene Glycol Precipitation Method for the Estimation of High-density Lipoprotein Cholesterol', *Ann. Clin. Biochem.*, **18**(Pt 3), 177–181 (1981).
62. M.M. Kimberly, E.T. Leary, T.G. Cole, P.P. Waymack, 'Selection, Validation, Standardization, and Performance of a Designated Comparison Method for HDL-Cholesterol for Use in the Cholesterol Reference Method Laboratory Network', *Clin. Chem.*, **45**(10), 1803–1812 (1999).
63. H. Sugiuchi, Y. Uji, H. Okabe, T. Irie, K. Uekama, N. Kayahara, K. Miyauchi, 'Direct Measurement of High-density Lipoprotein Cholesterol in Serum with Polyethylene Glycol-modified Enzymes and Sulfated Alpha-Cyclodextrin', *Clin. Chem.*, **41**(5), 717–723 (1995).
64. M. Nauck, W. Marz, B. Haas, H. Wieland, 'Homogeneous Assay for Direct Determination of High-density Lipoprotein Cholesterol Evaluated', *Clin. Chem.*, **42**(3), 424–429 (1996).
65. M. Nauck, K. Winkler, W. Marz, H. Wieland, 'Quantitative Determination of High-, Low-, and Very-low-density Lipoproteins and Lipoprotein(a) by Agarose Gel Electrophoresis and Enzymatic Cholesterol Staining', *Clin. Chem.*, **41**, 1761–1767 (1995).
66. P.S. Bachorik, 'Measurement of Low-density Lipoprotein Cholesterol', in *Handbook of Lipoprotein Testing*, eds. N. Rifai, G.R. Warnick, M.H. Dominiczak, AACC Press, Washington, 145–160, 1997.
67. W.T. Friedewald, R.I. Levy, D.S. Fredrickson, 'Estimation of the Concentration of Low-density Lipoprotein Cholesterol in Plasma, without Use of the Preparative Ultracentrifuge', *Clin. Chem.*, **18**(6), 499–502 (1972).
68. H. Sugiuchi, T. Irie, Y. Uji, T. Ueno, T. Chaen, K. Uekama, H. Okabe, 'Homogeneous Assay for Measuring Low-density Lipoprotein Cholesterol in Serum with Triblock Copolymer and Alpha-Cyclodextrin Sulfate', *Clin. Chem.*, **44**(3), 522–531 (1998).
69. P. Ellerbe, L.T. Sniegoski, M.J. Welch, 'Isotope Dilution Mass Spectrometry as a Candidate Definitive Method for Determining Total Glycerides and Triglycerides in Serum', *Clin. Chem.*, **41**(3), 397–404 (1995).
70. L.A. Carlson, L.B. Wadström, 'Determination of Glycerides in Blood Serum', *Clin. Chim. Acta.*, **4**, 197–205 (1959).
71. H.B. Lofland, Jr, 'A Semiautomated Procedure for the Determination of Triglycerides in Serum', *Anal. Biochem.*, **9**, 393–400 (1964).
72. G. Bucolo, H. David, 'Quantitative Determination of Serum Triglycerides by the Use of Enzymes', *Clin. Chem.*, **19**(5), 476–482 (1973).
73. R.E. Megraw, D.E. Dunn, H.G. Biggs, 'Manual and Continuous-flow Colorimetry of Triacylglycerols by a Fully Enzymic Method', *Clin. Chem.*, **25**(2), 273–278 (1979).
74. P. Fossati, L. Prencipe, 'Serum Triglycerides Determined Colorimetrically with an Enzyme that Produces Hydrogen Peroxide', *Clin. Chem.*, **28**(10), 2077–2080 (1982).
75. Y. Nie, J.L. He, S.L. Hsia, 'A Micro Enzymic Method for Determination of Choline-containing Phospholipids in Serum and High Density Lipoproteins', *Lipids*, **28**(10), 949–951 (1993).

76. P.V. Subbaiah, 'Determination and Clinical Significance of Phospholipids', in *Handbook of Lipoprotein Testing*, eds. N. Rifai, G.R. Warnick, M.H. Dominiczak, AACC Press, Washington, 415–429, 1997.
77. I.J. Cartwright, 'Separation and Analysis of Phospholipids by Thin Layer Chromatography', *Methods Mol. Biol.*, **19**, 153–167 (1993).
78. W.W. Christie, 'Rapid Separation and Quantification of Lipid Classes by High Performance Liquid Chromatography and Mass (Light-Scattering) Detection', *J. Lipid Res.*, **26**(4), 507–512 (1985).
79. R. Homan, M.K. Anderson, 'Rapid Separation and Quantitation of Combined Neutral and Polar Lipid Classes by High-performance Liquid Chromatography and Evaporative Light-scattering Mass Detection', *J. Chromatogr. B, Biomed. Sci. Appl.*, **708**(1–2), 21–26 (1998).
80. S.J. Smith, G.R. Cooper, G.L. Myers, E.J. Sampson, 'Biological Variability in Concentrations of Serum Lipids: Sources of Variation Among Results from Published Studies and Composite Predicted Values', *Clin. Chem.*, **39**(6), 1012–1022 (1993).
81. G.R. Cooper, S.J. Smith, G.L. Myers, E.J. Sampson, E. Magid, 'Estimating and Minimizing Effects of Biologic Sources of Variation by Relative Range when Measuring the Mean of Serum Lipids and Lipoproteins', *Clin. Chem.*, **40**(2), 227–332 (1994).
82. S.M. Marcovina, V.P. Gaur, J.J. Albers, 'Biological Variability of Cholesterol, Triglyceride, Low- and High-density Lipoprotein Cholesterol, Lipoprotein(a), and Apolipoproteins A-I and B', *Clin. Chem.*, **40**(4), 574–578 (1994).
83. G.R. Warnick, J.J. Albers, 'Physiological and Analytical Variation in Cholesterol and Triglycerides', *Lipids*, **11**(3), 203–208 (1976).
84. M.H. Tan, E.G. Wilmshurst, R.E. Gleason, J.S. Soeldner, 'Effect of Posture on Serum Lipids', *N. Engl. J. Med.*, **289**(8), 416–419 (1973).
85. M. Miller, P.S. Bachorik, T.A. Cloey, 'Normal Variation of Plasma Lipoproteins: Postural Effects on Plasma Concentrations of Lipids, Lipoproteins, and Apolipoproteins', *Clin. Chem.*, **38**(4), 569–574 (1992).
86. J.S. Cohn, J.R. McNamara, E.J. Schaefer, 'Lipoprotein Cholesterol Concentrations in the Plasma of Human Subjects as Measured in the Fed and Fasted States', *Clin. Chem.*, **34**(12), 2456–2459 (1988).
87. E.T. Leary, T. Wang, D.J. Baker, D.D. Cilla, J. Zhong, G.R. Warnick, K. Nakajima, R.J. Havel, 'Evaluation of an Immunoseparation Method for Quantitative Measurement of Remnant-like Particle-Cholesterol in Serum and Plasma', *Clin. Chem.*, **44**(12), 2490–2498 (1998).
88. G.R. Warnick, E.T. Leary, E.B. Ammirati, M.P. Allen, 'Cholesterol in Fingerstick Capillary Specimens Can be Equivalent to Conventional Venous Measurements', *Arch. Pathol. Lab. Med.*, **118**(11), 1110–1114 (1994).
89. M.J. Naughton, R.V. Luepker, D. Strickland, 'The Accuracy of Portable Cholesterol Analyzers in Public Screening Programs', *JAMA*, **263**(9), 1213–1217 (1990).
90. N. Rifai, Q. Pham, R.G. McMurray, 'Serum Lipid and Apolipoprotein Changes in Normal Pregnancy', *Clin. Chem.*, **35**, 1066 (1989).
91. M. Panteghini, F. Pagani, 'Serum Concentrations of Lipoprotein(a) During Normal Pregnancy and Postpartum', *Clin. Chem.*, **37**(11), 2009–2010 (1991).
92. B. Zak, 'Cholesterol Methodologies: A Review', *Clin. Chem.*, **23**, 1201–1214 (1977).
93. A. Cohen, H.S. Hertz, J. Mandel, R.C. Paule, R. Schaffer, L.T. Sniegoski, T. Sun, M.J. Welch, E. White, 5th, 'Total Serum Cholesterol by Isotope Dilution/Mass Spectrometry: A Candidate Definitive Method', *Clin. Chem.*, **26**, 854–860 (1980).
94. P.A. Villanova, 'Development of Definitive Methods for the National Reference System for the Clinical Laboratory, NRSCL 1-A', National Committee for Clinical Laboratory Standards, PA, USA, 1991.
95. C. Liebermann, 'Ueber des Oxychinoterpen', *Ber. Dtsch. Chem. Ges.*, **18**, 1803–1809 (1885).
96. H. Burchard, 'Beitrage zur Kenntnis des Cholesterins', *Chem. Zentralbl.*, **61**, 25–27 (1890).
97. P. Ellerbe, G.L. Myers, G.R. Cooper, H.S. Hertz, L.T. Sniegoski, M.J. Welch, E. White, 5th, 'A Comparison of Results for Cholesterol in Human Serum Obtained by the Reference Method and by the Definitive Method of the National Reference System for Cholesterol', *Clin. Chem.*, **36**, 370–375 (1990).
98. J.T. Bernert, Jr, J.R. Akins, G.R. Cooper, A.K. Poulose, G.L. Myers, E.J. Sampson, 'Factors Influencing the Accuracy of the National Reference System Total Cholesterol Reference Method', *Clin. Chem.*, **37**, 2053–2061 (1991).
99. National Institutes of Health, National Heart, Lung and Blood Institute, 'Second Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults', in *NIH Publication No. 93-3095*, US Government Printing Office, 1993.
100. National Institutes of Health, National Heart, Lung and Blood Institute, 'National Cholesterol Education Program Report of the Expert Panel on Blood Cholesterol levels in Children and Adolescents', in *NIH Publication No. 91-2732*, US Government Printing Office, 1991.
101. J.E. Bateson, J.D. Artiss, B. Zak, 'The Development of an Enzymic Reference Equivalent Method for Total Cholesterol', *Clin. Chem.*, **35**, 1071–1072 (1979).
102. C.H. Peddicord, W.A. Barnes, 'Ascorbic Acid Interferogram for Cholesterol in the Demand and TDX', *Clin. Chem.*, **34**(4), 773–774 (1988).
103. J. Chapman, 'Comparative Analysis of Mammalian Plasma Lipoproteins', *Methods Enzymol.*, **128**, 70–143 (1986).
104. J.T. Bernert, Jr, J.R. Akins, G.R. Cooper, A.K. Poulose, G.L. Myers, E.J. Sampson, 'Factors Influencing the

Accuracy of the National Reference System Total Cholesterol Reference Method', *Clin. Chem.*, **37**, 2053–2061 (1991).
105. R.J. Havel, H.A. Eder, J.H. Bragdon, 'The Distribution and Chemical Composition of Ultracentrifugally Separated Lipoproteins in Human Serum', *J. Clin. Invest.*, **34**, 1345–1353 (1955).
106. W.J. Lossow, F.T. Lindgren, J.C. Murchio, G.R. Stevens, L.C. Jensen, 'Particle Size and Protein Content of Six Fractions of the $S_f > 20$ Plasma Lipoproteins Isolated by Density Gradient Centrifugation', *J. Lipid Res.*, **10**, 68–76 (1969).
107. J.D. Belcher, J.O. Egan, G. Bridgmen, R. Baker, J. Flack, 'A Micro-enzymatic Method to Measure Cholesterol and Triglyceride in Lipoprotein Subfractions Separated by Density Gradient Ultracentrifugation from 200 Microliters of Plasma or Serum', *J. Lipid Res.*, **32**, 359–370 (1991).
108. J.T. Cone, J.P. Segrest, B.H. Chung, J.B. Ragland, S.M. Sabesin, A. Glasscock, 'Computerized Rapid High Resolution Quantitative Analysis of Plasma Lipoprotein Based Upon Single Vertical Spin Centrifugation', *J. Lipid Res.*, **23**, 923–935 (1982).
109. L.L. Wu, G.R. Warnick, J.T. Wu, R.R. Williams, J.M. Lalouel, 'A Rapid Micro-scale Procedure for Determination of the Total Lipid Profile', *Clin. Chem.*, **35**(7), 1486–1491 (1989).
110. National Institutes of Health, National Heart, Lung and Blood Institute, 'Aggregate Distribution of Lipids, Lipoproteins and Selected Variables in 11 North American Populations', in *The Lipid Research Clinics Population Study Book: Volume 1 The Prevalence Study, NIH Publication No. 80-1527*, US Government Printing Office, 1995.
111. P.P. Waymack, S.F. Ethridge, G.L. Meyers, 'The Effect of Temperature on the Betaquantification Procedure for Low-density Lipoprotein Cholesterol in Serum', *Clin. Chem.*, **45**(S6), A18 (1999).
112. J.R. McNamara, J.S. Cohn, P.W. Wilson, E.J. Schaefer, 'Calculated Values for Low-density Lipoprotein Cholesterol in the Assessment of Lipid Abnormalities and Coronary Disease Risk', *Clin. Chem.*, **36**(1), 36–42 (1990).
113. N. Rifai, G.R. Warnick, J.R. McNamara, J.D. Belcher, G.F. Grinstead, I.D. Frantz, Jr, 'Measurement of Low-density Lipoprotein Cholesterol In Serum: A Status Report', *Clin. Chem.*, **38**(1), 150–160 (1992).
114. P.N. Demacker, A.G. Hijmans, B.J. Brenninkmeijer, A.P. Jansen, A. van 't Laar, 'Five Methods for Determining Low-density Lipoprotein Cholesterol Compared', *Clin. Chem.*, **30**(11), 1797–1800 (1984).
115. T. Pisani, C.P. Gebski, E.T. Leary, G.R. Warnick, J.F. Ollington, 'Accurate Direct Determination of Low-density Lipoprotein Cholesterol Using an Immunoseparation Reagent and Enzymatic Cholesterol Assay', *Arch. Pathol. Lab. Med.*, **119**(12), 1127–1135 (1995).
116. J.R. McNamara, T.G. Cole, J.H. Contois, C.A. Ferguson, J.M. Ordovas, E.J. Schaefer, 'Immunoseparation Method for Measuring Low-density Lipoprotein Cholesterol Directly from Serum Evaluated', *Clin. Chem.*, **41**(2), 232–240 (1995).
117. M. Nakamura, S. Saito, M. Iida, 'Certification of Five Japanese Manufacturers who have Developed or are Developing Homogeneous Methods for LDL-Cholesterol', *Clin. Chem.*, **45**(6), A13 (1999).
118. K. Slickers, D. Burton, A. Foster, M. Macke, T. Cole, 'Statistical Analysis of Total, HDL and LDL Cholesterol Values Determined on Paired Fasting and Non-fasting Samples', *Clin. Chem.*, **45**(6), A16 (1999).
119. G.R. Warnick, P.D. Wood, 'National Cholesterol Education Program Recommendations for Measurement of High-density Lipoprotein Cholesterol: Executive Summary', *Clin. Chem.*, **41**, 1427–1433 (1995).
120. F.T. Lindgren, L.C. Jensen, F.T. Hatch, 'The Isolation and Quantitative Analysis of Serum Lipoproteins', in *Blood Lipids and Lipoproteins: Quantitation, Composition, and Metabolism*, ed. G.F. Nelson, Wiley Interscience, New York, 181–274, 1972.
121. G.R. Warnick, J.J. Albers, 'A Comprehensive Evaluation of the Heparin–Manganese Precipitation Procedure for Estimating High Density Lipoprotein Cholesterol', *J. Lipid Res.*, **19**(1), 65–76 (1978).
122. C. Cobbaert, L. Zwang, F. Ceriotti, A. Modenese, P. Cremer, W. Herrmann, G. Hoss, J. Jarausch, R. Turk, W. Marz, M. Nauck, 'Reference Standardization and Triglyceride Interference of a New Homogeneous HDL–Cholesterol Assay Compared with a Former Chemical Precipitation Assay', *Clin. Chem.*, **44**(4), 779–789 (1998).
123. N. Rifai, T.G. Cole, E. Iannotti, T. Law, M. Macke, R. Miller, D. Dowd, D.A. Wiebe, 'Assessment of Interlaboratory Performance in External Proficiency Testing Programs with a Direct HDL–Cholesterol Assay', *Clin. Chem.*, **44**(7), 1452–1458 (1998).
124. J. Camps, J.M. Simo, S. Guaita, N. Ferre, J. Joven, 'Altered Composition of Lipoproteins in Liver Cirrhosis Compromises Three Homogeneous Methods for HDL–Cholesterol', *Clin. Chem.*, **45**, 685–688 (1999).
125. K.J. Lackner, G. Schmitz, 'Beta VLDL of Patients with Type III Hyperlipoproteinemia Interferes with Homogeneous Determination of HDL–Cholesterol Based on Polyethylene Glycol-modified Enzymes', *Clin. Chem.*, **44**, 2546–2548 (1998).
126. S.G. Klotzsch, J.T. McNamara, 'Triglyceride Measurements: A Review of Methods and Interferences', *Clin. Chem.*, **36**(9), 1605–1613 (1990).
127. P.O. Kwiterovich, Jr, 'The Antiatherogenic Role of High-density Lipoprotein Cholesterol', *Am. J. Cardiol.*, **82**(9A), 13Q–21Q (1998).
128. J.R. Barr, V.L. Maggio, D.G. Patterson, Jr, G.R. Cooper, L.O. Henderson, W.E. Turner, S.J. Smith, W.H.

Hannon, L.L. Needham, E.J. Sampson, 'Isotope Dilution–Mass Spectrometric Quantification of Specific Proteins: Model Application with Apolipoprotein A-I', *Clin. Chem.*, **42**(10), 1676–1682 (1996).

129. S.M. Marcovina, J.J. Albers, L.O. Henderson, W.H. Hannon, 'International Federation of Clinical Chemistry Standardization Project for Measurements of Apolipoproteins A-I and B. III. Comparability of Apolipoprotein A-I Values by Use of International Reference Material', *Clin. Chem.*, **39**(5), 773–781 (1993).
130. P. Dacol, G.M. Kostner, 'Immunoquantification of Total Apolipoprotein B in Serum by Nephelometry: Influence of Lipase Treatment and Detergents', *Clin. Chem.*, **29**(6), 1045–1050 (1983).
131. H. Thakkar, C.L. Davey, E.A. Medcalf, L. Skingle, A.R. Craig, D.J. Newman, C.P. Price, 'Stabilization of Turbidimetric Immunoassay by Covalent Coupling of Antibody to Latex Particles', *Clin. Chem.*, **37**(7), 1248–1251 (1991).
132. G. Sigurdsson, A. Baldursdottir, H. Sigvaldason, U. Agnarsson, G. Thorgeirsson, N. Sigfusson, 'Predictive Value of Apolipoproteins in a Prospective Survey of Coronary Artery Disease in Men', *Am. J. Cardiol.*, **69**(16), 1251–1254 (1992).
133. S.M. Marcovina, J.L. Adolphson, M. Parlavecchia, J.J. Albers, 'Effects of Lyophilization of Serum on the Measurement of Apolipoproteins A-I and B', *Clin. Chem.*, **36**(2), 366–369 (1990).
134. J.V. Mei, M.K. Powell, L.O. Henderson, S.J. Smith, G.R. Cooper, S.M. Marcovina, W.H. Hannon, 'Method-dependent Variations in the Stability of Apolipoprotein B in a Stabilized Liquid Reference Material', *Clin. Chem.*, **40**(5), 716–722 (1994).
135. F. Karpe, A. Hamsten, K. Uffelman, G. Steiner, 'Apolipoprotein B-48', *Methods Enzymol.*, **263**, 95–104 (1996).
136. Y. Uchida, Y. Kurano, S. Ito, 'Establishment of Monoclonal Antibody Against Human Apo B-48 and Measurement of Apo B-48 in Serum by ELISA Method', *J. Clin. Lab. Anal.*, **12**(5), 289–292 (1998).
137. P.W. Connelly, C. Vezina, G.F. Maguire, 'Quantification of Apolipoprotein C-II by Immunochemical and Chromatographic Methods', *Methods Enzymol.*, **263**, 188–208 (1996).
138. K. Itakura, T. Matsudate, T. Sakurai, S. Hashimoto, K. Ito, H. Kanno, M. Hirata, K. Nakamura, 'Single Radial Immunodiffusion of Serum Apolipoproteins C-II, C-III and E – Pretreatment of Samples with Surfactant', *Clin. Chim. Acta*, **161**(3), 275–282 (1986).
139. N.D. Bren, A. Rastogi, B.A. Kottke, 'Quantification of Human Plasma Apolipoproteins C-I, C-II, and C-III by Radioimmunoassays', *Mayo Clin. Proc.*, **68**(7), 657–664 (1993).
140. W.F. Riesen, E. Sturzenegger, 'Quantitation of Apolipoprotein C-III in Normal and in Hyperlipaemic Serum Samples by Enzyme-linked Immunosorbent Assay', *Ann. Clin. Biochem.*, **24**(Pt 1), 66–72 (1987).
141. J. Bury, G. Michiels, M. Rosseneu, 'Human Apolipoprotein C-II Quantitation by Sandwich Enzyme-linked Immunosorbent Assay', *J. Clin. Chem. Clin. Biochem.*, **24**(7), 457–463 (1986).
142. N. Alsayed, R. Rebourcet, J. Chapman, 'Concentrations of Apoprotein CII, CIII, and E in Total Serum and in the Apoprotein B-containing Lipoproteins, Determined by a New Enzyme-linked Immunosorbent Assay', *Clin. Chem.*, **36**(12), 2047–2052 (1990).
143. N. Rifai, L.M. Silverman, 'Immunoturbidimetric Techniques for Quantifying Apolipoproteins CII and CIII', *Clin. Chem.*, **32**(10), 1969–1972 (1986).
144. S. Gustafson, B. Vessby, A.M. Ostlund-Lindqvist, 'Serum and Interstitial Fluid Apolipoprotein E Levels in the Healthy and in Hyperlipoproteinemia Type III as Studied by Radioimmunoassay', *Clin. Chim. Acta*, **151**(1), 49–59 (1985).
145. K. Yamauchi, M. Tozuka, T. Nakabayashi, M. Sugano, H. Hidaka, Y. Kondo, T. Katsuyama, 'Apolipoprotein E in Cerebrospinal Fluid: Relation to Phenotype and Plasma Apolipoprotein E Concentrations', *Clin. Chem.*, **45**(4), 497–504 (1999).
146. N. Rifai, L.M. Silverman, 'A Simple Immunotechnique for the Determination of Serum Concentration of Apolipoprotein E', *Clin. Chim. Acta.*, **163**(2), 207–213 (1987).
147. M. Eto, K. Watanabe, K. Ishii, 'A Rapid Flat Gel Isoelectric Focusing Method for the Determination of Apolipoprotein E Phenotypes and Its Application', *Clin. Chim. Acta*, **149**(1), 21–28 (1985).
148. W. Marz, S. Cezanne, W. Gross, 'Phenotyping of Apolipoprotein E by Immunoblotting in Immobilized pH Gradients', *Electrophoresis*, **12**(1), 59–63 (1991).
149. C.H. Luley, B. Haas, B. Buhrer, H. Wieland, 'Improvement of Apolipoprotein E Phenotyping by Isoelectric Focusing/Immunofixation', *Clin. Chem.*, **38**(1), 168 (1992).
150. P. Richard, G. Thomas, M.P. De Zulueta, J.L. De Gennes, M. Thomas, A. Cassaigne, G. Bereziat, A. Iron, 'Common and Rare Genotypes of Human Apolipoprotein E Determined by Specific Restriction Profiles of Polymerase Chain Reaction-Amplified DNA', *Clin. Chem.*, **40**(1), 24–29 (1994).
151. S.M. Sell, K. Ren, 'Automated Capillary Electrophoresis in the Genotyping of Apolipoprotein E', *Genomics*, **46**(1), 163–164 (1997).
152. R.M. Krauss, 'Dense Low Density Lipoproteins and Coronary Artery Disease', *Am. J. Cardiol.*, **75**(6), 53B–57B (1995).
153. A. Tailleux, J.C. Fruchart, 'HDL Heterogeneity and Atherosclerosis', *Crit. Rev. Clin. Lab. Sci.*, **33**(3), 163–201 (1996).
154. B.H. Chung, J.P. Segrest, J.T. Cone, J. Pfau, J.C. Geer, L.A. Duncan, 'High Resolution Plasma Lipoprotein Cholesterol Profiles by a Rapid, High Volume Semi-automated Method', *J. Lipid Res.*, **22**(6), 1003–1014 (1981).

155. J.H. Auwerx, C.A. Marzetta, J.E. Hokanson, J.D. Brunzell, 'Large Buoyant LDL-like Particles in Hepatic Lipase Deficiency', *Arteriosclerosis*, **9**(3), 319–325 (1989).
156. R.M. Krauss, D.J. Burke, 'Identification of Multiple Subclasses of Plasma Low Density Lipoproteins in Normal Humans', *J. Lipid Res.*, **23**(1), 97–104 (1982).
157. D.L. Rainwater, D.W. Andres, A.L. Ford, F. Lowe, P.J. Blanche, R.M. Krauss, 'Production of Polyacrylamide Gradient Gels for the Electrophoretic Resolution of Lipoproteins', *J. Lipid Res.*, **33**(12), 1876–1881 (1992).
158. M.A. Austin, L. Mykkanen, L.J. Kuusisto, K.L. Edwards, C. Nelson, S.M. Haffner, K. Pyorala, M. Laakso, 'Prospective Study of Small LDLs as a Risk Factor for Non-insulin Dependent Diabetes Mellitus in Elderly Men and Women', *Circulation*, **92**(7), 1770–1778 (1995).
159. P.N. Demacker, D.F. Van Sommeren-Zondag, A.F. Stalenhoef, P.M. Stuyt, A. Van't Laar, 'Ultracentrifugation in Swinging-bucket and Fixed-angle Rotors Evaluated for Isolation and Determination of High-density Lipoprotein Subfractions HDL_2 and HDL_3', *Clin. Chem.*, **29**(4), 656–663 (1983).
160. P.H. Groot, L.M. Scheek, L. Havekes, W.L. Van Noort, F.M. Van't Hooft, 'A One-step Separation of Human Serum High Density Lipoproteins 2 and 3 by Rate-zonal Density Gradient Ultracentrifugation in a Swinging Bucket Rotor', *J. Lipid Res.*, **23**(9), 1342–1353 (1982).
161. M.C. Cheung, J.P. Segrest, J.J. Albers, J.T. Cone, C.G. Brouillette, B.H. Chung, M. Kashyap, M.A. Glasscock, G.M. Anantharamaiah, 'Characterization of High Density Lipoprotein Subspecies: Structural Studies by Single Vertical Spin Ultracentrifugation and Immunoaffinity Chromatography', *J. Lipid Res.*, **28**(8), 913–929 (1987).
162. J.R. Patsch, W. Patsch, 'Zonal Ultracentrifugation', *Methods Enzymol.*, **129**, 3–26 (1986).
163. P.T. Williams, R.M. Krauss, K.M. Vranizan, M.L. Stefanick, P.D. Wood, F.T. Lindgren, 'Associations of Lipoproteins and Apolipoproteins with Gradient Gel Electrophoresis Estimates of High Density Lipoprotein Subfractions in Men and Women', *Arterioscler. Thromb.*, **12**(3), 332–340 (1992).
164. P.M. Clifton, A.M. Mackinnon, P.J. Barter, 'Separation and Characterization of High-density Lipoprotein Subpopulations by Gel Permeation Chromatography', *J. Chromatogr.*, **414**(1), 25–34 (1987).
165. P.S. Bachorik, J.J. Albers, 'Precipitation Methods for Quantification of Lipoproteins', *Methods Enzymol.*, **129**, 78–100 (1986).
166. D.L. Rainwater, P.H. Moore, Jr, W.R. Shelledy, T.D. Dyer, S.H. Slifer, 'Characterization of a Composite Gradient Gel for the Electrophoretic Separation of Lipoproteins', *J. Lipid Res.*, **38**(6), 1261–1266 (1997).
167. K.R. Kulkarni, D.W. Garber, S.M. Marcovina, J.P. Segrest, 'Quantification of Cholesterol in all Lipoprotein Classes by the VAP-II Method', *J. Lipid Res.*, **35**(1), 159–168 (1994).
168. J.D. Otvos, E.J. Jeyarajah, D.W. Bennett, R.M. Krauss, 'Development of a Proton Nuclear Magnetic Resonance Spectroscopic Method for Determining Plasma Lipoprotein Concentrations and Subspecies Distributions from a Single, Rapid Measurement', *Clin. Chem.*, **38**(9), 1632–1638 (1992).
169. H.C. Guo, V.W. Armstrong, G. Luc, C. Billardon, S. Goulinet, R. Nustede, D. Seidel, M.J. Chapman, 'Characterization of Five Mouse Monoclonal Antibodies to Apolipoprotein(a) from Human Lp(a): Evidence for Weak Plasminogen Reactivity', *J. Lipid Res.*, **30**(1), 23–37 (1989).
170. J.J. Albers, J.L. Adolphson, W.R. Hazzard, 'Radioimmunoassay of Human Plasma Lp(a) Lipoprotein', *J. Lipid Res.*, **18**(3), 331–338 (1977).
171. G. Cazzolato, G. Prakasch, S. Green, G.M. Kostner, 'The Determination of Lipoprotein Lp(a) by Rate and End-point Nephelometry', *Clin. Chim. Acta*, **135**(2), 203–208 (1983).
172. W.C. Taddei-Peters, B.T. Butman, G.R. Jones, T.M. Venetta, P.F. Macomber, J.H. Ransom, 'Quantification of Lipoprotein(a) Particles Containing Various Apolipoprotein(a) Isoforms by a Monoclonal Anti-Apo(a) Capture Antibody and a Polyclonal Anti-Apolipoprotein B Detection Antibody Sandwich Enzyme Immunoassay', *Clin. Chem.*, **39**(7), 1382–1389 (1993).
173. G.M. Fless, M.L. Snyder, A.M. Scanu, 'Enzyme-linked Immunoassay for Lp(a)', *J. Lipid Res.*, **30**(5), 651–662 (1989).
174. D.M. Levine, B.J. Sloan, J.E. Donner, J.D. Lorenz, R.H. Heinzerling, 'Automated Measurement of Lipoprotein(a) by Immunoturbidimetric Analysis', *Int. J. Clin. Lab. Res.*, **22**(3), 173–178 (1992).
175. L.J. Seman, J.L. Jenner, J.R. McNamara, E.J. Schaefer, 'Quantification of Lipoprotein(a) in Plasma by Assaying Cholesterol in Lectin-Bound Plasma Fraction', *Clin. Chem.*, **40**(3), 400–403 (1994).
176. R. Havel, 'McCollum Award Lecture, 1993: Triglyceride-rich Lipoproteins and Atherosclerosis – New Perspectives', *Am. J. Clin. Nutr.*, **59**(4), 795–796 (1994).
177. R.M. Krauss, 'Triglycerides and Atherogenic Lipoproteins: Rationale for Lipid Management', *Am. J. Med.*, **105**(1A), 58S–62S (1998).
178. J.S. Cohn, C. Marcoux, J. Davignon, 'Detection, Quantification, and Characterization of Potentially Atherogenic Triglyceride-rich Remnant Lipoproteins', *Arterioscler. Thromb. Vasc. Biol.*, **19**(10), 2474–2486 (1999).
179. E.T. Leary, T. Wang, D.J. Baker, D.D. Cilla, J. Zhong, G.R. Warnick, K. Nakajima, R.J. Havel, 'Evaluation of an Immunoseparation Method for Quantitative Measurement of Remnant-like Particle-cholesterol in Serum and Plasma', *Clin. Chem.*, **44**(12), 2490–2498 (1998).
180. T. Wang, K. Nakajima, E.T. Leary, G.R. Warnick, J.S. Cohn, P.N. Hopkins, L.L. Wu, D.D. Cilla, J. Zhong,

R.J. Havel, 'Ratio of Remnant-like Particle-cholesterol to Serum Total Triglycerides is an Effective Alternative to Ultracentrifugal and Electrophoretic Methods in the Diagnosis of Familial Type III Hyperlipoproteinemia', *Clin. Chem.*, **45**(11), 1981–1987 (1999).

181. H. Itabe, H. Yamamoto, T. Imanaka, K. Shimamura, H. Uchiyama, J. Kimura, T. Sanaka, Y. Hata, T. Takano, 'Sensitive Detection of Oxidatively Modified Low Density Lipoprotein Using a Monoclonal Antibody', *J. Lipid Res.*, **37**(1), 45–53 (1996).
182. W.Y. Craig, S.E. Poulin, C.P. Nelson, R.F. Ritchie, 'ELISA of IgG Antibody to Oxidized Low-density Lipoprotein: Effects of Blocking Buffer and Method of Data Expression', *Clin. Chem.*, **40**(6), 882–888 (1994).
183. M. Puurunen, M. Manttari, V. Manninen, L. Tenkanen, G. Alfthan, C. Ehnholm, O. Vaarala, K. Aho, T. Palosuo, 'Antibody Against Oxidized Low-density Lipoprotein Predicting Myocardial Infarction', *Arch. Intern. Med.*, **154**(22), 2605–2609 (1994).
184. J.M. Gutteridge, 'Lipid Peroxidation and Antioxidants as Biomarkers of Tissue Damage', *Clin. Chem.*, **41**(12 Pt 2), 1819–1828 (1995).
185. J.M. Onorato, S.R. Thorpe, J.W. Baynes, 'Immunohistochemical and ELISA Assays for Biomarkers of Oxidative Stress in Aging and Disease', *Ann. N.Y. Acad. Sci.*, **854**, 277–290 (1998).
186. T.A. Mori, K.D. Croft, I.B. Puddey, L.J. Beilin, 'An Improved Method for the Measurement of Urinary and Plasma F_2-Isoprostanes Using Gas Chromatography–Mass Spectrometry', *Anal. Biochem.*, **268**(1), 117–125 (1999).
187. S. Basu, 'Radioimmunoassay of 8-Iso-Prostaglandin $F_{2\alpha}$: An Index for Oxidative Injury Via Free Radical Catalysed Lipid Peroxidation', *Prostaglandins Leukot. Essent. Fatty Acids*, **58**(4), 319–325 (1998).
188. M.P. Reilly, D. Pratico, N. Delanty, G. Diminno, E. Tremoli, D. Rader, S. Kapoor, J. Rokach, J. Lawson, G.A. Fitzgerald, 'Increased Formation of Distinct F_2 Isoprostanes in Hypercholesterolemia', *Circulation*, **98**(25), 2822–2828 (1998).
189. V.P. Dole, 'A Relation Between Non-esterified Fatty Acids in Plasma and the Metabolism of Glucose', *J. Clin. Invest.*, **35**, 105–154 (1956).
190. M. Puttmann, H. Krug, E. Von Ochsenstein, R. Kattermann, 'Fast HPLC Determination of Serum Free Fatty Acids in the Picomole Range', *Clin. Chem.*, **39**(5), 825–832 (1993).
191. G. Lepage, C.C. Roy, 'Specific Methylation of Plasma Nonesterified Fatty Acids in a One-step Reaction', *J. Lipid Res.*, **29**(2), 227–235 (1988).
192. A.J. Degen, J. Van Der Vies, 'Enzymatic Microdetermination of Free Fatty Acids in Plasma of Animals Using Paraoxon to Prevent Lipolysis', *Scand. J. Clin. Lab. Invest.*, **45**(3), 283–285 (1985).

Micro Total Analytical Systems in Clinical Chemistry

J. Hendrikse and A. van den Berg
MESA Research Institute, University of Twente, Enschede, The Netherlands

1 Introduction 1
2 Theoretical Considerations on Miniaturization 3
2.1 Advantages 3
2.2 Modeling and Simulation for Miniaturized Total Analytical Systems 4
3 Materials 4
3.1 Silicon 5
3.2 Glass and Quartz 5
3.3 Polymers 5
4 Micromachining Techniques 6
4.1 Pattern Transfer 6
4.2 Wet Etching 6
4.3 Dry Etching 7
4.4 Polymer Molding 7
4.5 The Lithographie Galvanoformung Abformung Process 9
4.6 Thick Photoresists 9
4.7 Sealing of Channels and Reaction Chambers 9
5 Basic Hardware Components 9
5.1 Channels, Junctions and Connectors 10
5.2 Valves 10
5.3 Pumps 11
5.4 Filters 12
5.5 Mixers and Reactors 13
5.6 Dispensers 14
6 Separation Techniques 14
6.1 Dialysis, Filtration and Centrifugation 15
6.2 Chromatography 15
6.3 Electrophoresis 15
6.4 Sorting and Manipulation of Large Biomolecules, Particles and Cells 16
7 Detection 17
7.1 Optical Detection 17
7.2 Electrochemical Detection 18
7.3 Mass Spectrometry 18
8 Applications 18
8.1 Continuous Blood Monitoring 18
8.2 Environmental Monitoring 19
8.3 Intermittent Bedside Blood Monitoring 20
8.4 On-chip DNA Characterization 21
Abbreviations and Acronyms 21
Related Articles 22
References 22

A micro total analytical system (μTAS) is a system that enhances the performance of a complete chemical analysis by minimizing the scale on which it is performed. The main advantages are automated analyses, higher speed and better separation performance as well as the ability to analyze extremely small (picoliter) sample volumes and reactions involving minute amounts of reagents. A complete analysis may involve taking a sample, its pretreatment and separation into its different components and their detection. In order to perform this sequence of functions, the sample is transported through microchannels that connect the various system parts performing these tasks, while the various (feedback) signals are transported through a separate microelectronic network. Examples of system parts in silicon, glass and polymers that perform tasks such as hydrodynamic and electroosmotic pumping, laminar and turbulent reagent mixing and (di)electrophoretic separations of DNA fragments and particles and cells are presented. Currently, μTAS are mainly being developed and used for remote environmental and process monitoring, patient monitoring and the analysis of nucleic acids. An integrated system in each of these areas is discussed.

1 INTRODUCTION

A μTAS is a system that enhances the performance of a complete chemical analysis by minimizing the scale on which it is performed. The complete analysis may involve sampling, transport, sample pretreatment, separation into its different components and their detection. Whether stages such as sampling are part of a complete analysis will depend mainly on the practical use of the μTAS and is not very important here. To distinguish a μTAS from a chemical sensor, the requirement that at least some stages have to be integrated is useful. Moreover, the enhanced performance of the system has to be seen in its broadest sense. Not only are lower detection limits or a better sensitivity considered for better performance, but also faster or parallel analysis, less use of reagents, the possibility of continuous monitoring, remote operation or automation can be incentives for miniaturization.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Many aspects of the development of μTAS can probably best be understood by a comparison with integrated microelectronic circuits. At first sight this analogy may seem to be based on a number of more or less superficial similarities: until recently, both used silicon as their most important construction material and used lithographic methods for pattern transfer. Moreover, fluids behave in many ways like electric currents and the importance of both technologies rises exponentially. The last similarity has even prompted the extension of Moore's law to the number density and speed of on-chip DNA analyses.[1]

A somewhat closer inspection shows that the differences are manifold, however. Plastic and glass are taking the place of silicon in some μTAS applications. Whereas electronic circuits dealt until recently only with "clean" information, μTAS are in many cases supposed to deal with the dirty reality of complex samples.

Again, on a conceptual level, the similarities between the miniaturization and integration of analytical systems and electronic circuits are striking. The comparison is simplified by Table 1. It should be noted that the different approaches presented here do not mean to replace each other, they have not even been developed sequentially, and the development of some of them still goes on.

Chemical sensors form the basic μTAS components that started the development and are still important components of any μTAS. They have been in development for several decades, but never became the expected commercial success. In the early 1990s, it became apparent that for many applications, problems regarding drift, nonlinearity, calibration, but most of all selectivity would not be solved by improvements on stand-alone sensors.

A first attempt at solving these problems was made when feedback control of chemical sensors was introduced by Bergveld's group.[2] The use of feedback is a standard method in electronic circuit design to enhance the stability and linearize the response of components, as illustrated by the following example. In ion-sensitive field effect transistor (ISFET)-based urease sensors, the pH changes due to an enzymatic reaction are measured.

Table 1 Integration and miniaturization in analytical systems compared with the same trends in electronics

Electronics		μTAS
Components		Chemical sensors
Operational amplifiers	Feedback	Sensor–actuator systems
PCBs	Modular approach	MCBs
Integrated circuits	Integration	On-chip reaction, separation and detection

PCBs, printed circuit boards; MCBs, mixed circuit boards.

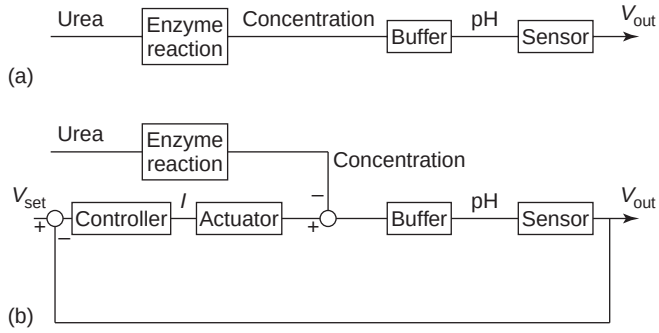

Figure 1 Block diagrams of (a) a conventional chemical sensor and (b) a chemical sensor–actuator system employing feedback.

Obviously, the pH changes depend on the buffer concentration of the sample, which in turn depends on the pH. Since the enzyme kinetics also depend on the pH, a very complicated nonlinear response is to be expected. By placement of a proton/hydroxyl ion actuator around the sensor, the pH can be kept at a stable value and the sensor signal used for feedback as illustrated in Figure 1(a) and (b).[3] The pH-static enzyme sensor–actuator system that was thus developed showed a linear response and had a signal that was independent of the buffer capacity of the sample. However, not all problems were solved by this approach: the sensor signal was still influenced by a deteriorating enzyme activity with time. Moreover, it turned out to be difficult to find chemical actuators besides the one mentioned above, and the cross-sensitivity remained a problem.

Traditionally, these problems were solved by a human analyst. The connection of the different steps performed by the analyst to build a system that needs no human intervention led to the development of a total analytical system (TAS), a rather bulky machine in which all the steps mentioned earlier were present.[4] The demand for portability of such machines led to the first miniaturization efforts, of which the Stanford gas chromatograph is the earliest example.[5] At Ciba-Geigy, the advantages of miniaturization for the analytical performance were systematically investigated for the first time by Manz et al.[6] At the same time that this key idea was formulated for the first time, the term μTAS was introduced.

The first systems consisted of modules performing different tasks, connected by polymer tubing, as illustrated in Figure 2. The modular approach of these systems allows the use of different materials and fabrication methods and permits separate testing of the various components, advantages that still make it a sound approach to many analytical challenges. Especially for applications such as remote sensing and environmental monitoring, where all components, including pumps and detection modules, are miniaturized, this approach is useful.

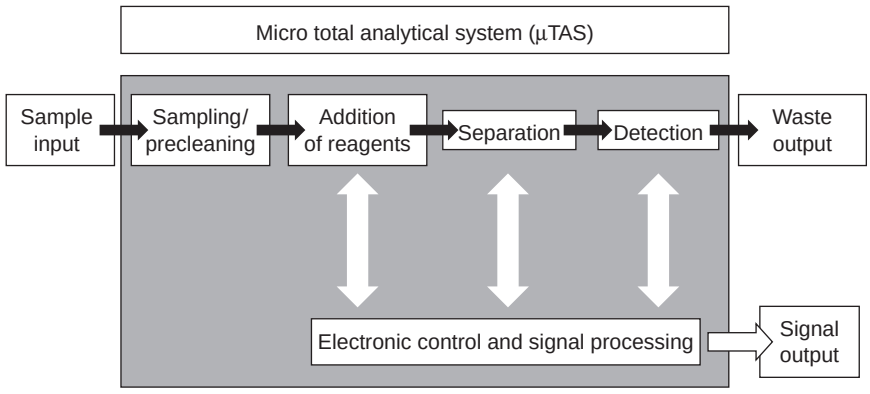

Figure 2 Block diagram of a μTAS.

The miniaturization of these systems led directly to their integration: dead volumes can be minimized and new components, only efficient at this small scale, have been developed. In contrast to the modular approach, in the integrated approach not all parts of the system are miniaturized: conventional high-voltage sources, mass spectrometers and laser fluorescence detectors are commonly used. This is because the main applications of these systems are found in laboratories for drug discovery and DNA characterization.

Two remarks on the future potential of μTAS are in place within the context of the comparison given here. First, in microelectronics, the applications surpass the original ideas of the inventors and a new engineering field has been born. Second, the use of some kind of software allows the use of the same circuit for a large number of tasks. While this step has been made in electronics, its counterpart in analytical chemistry is not apparent.

2 THEORETICAL CONSIDERATIONS ON MINIATURIZATION

2.1 Advantages

The μTAS field can be defined as exploiting the advantages of miniaturization on the analytical performance of chemical analysis systems. The effects of miniaturization on a number of key parameters in chemical analysis systems are summarized in Table 2.

As is indicated in Table 2, the Reynolds number (Re) is proportional to the channel diameter. In microchannels, Re is usually around 1500 and flow patterns are laminar under most circumstances. However, at high enough flow rates, flow can become turbulent in channels having diameters of micrometers[7] to several hundreds of micrometers.[8] On the one hand, laminar flow leads to a reproducible analytical performance that moreover can be predicted accurately by numerical calculations. On the other hand, special measures have to be taken in some cases to induce sufficiently fast mixing.

Table 2 Key parameters in the miniaturization of chemical analysis systems

Heat and mass transfer		
Reynolds number (Re)	$\frac{\text{Inertia}}{\text{Viscosity}}$	$\frac{\rho v d}{\eta}$
Weber number (We)	$\frac{\text{Inertia}}{\text{Surface tension}}$	$\frac{\rho v^2 d}{\sigma}$
Fourier number (F_o) mass	$\frac{\text{Process time}}{\text{Effective diffusion time}}$	$\frac{Dt}{d^2}$
Fourier number (F_o) heat	$\frac{\text{Process time}}{\text{Effective diffusion time}}$	$\frac{at}{d^2}$
Capillary flow	Capillary pressure	$\Delta p \propto d^{-1}$
Separation efficiency		
Electrophoresis, chromatography	Number of theoretical plates, separation time	$N \propto \frac{L}{d}, t \propto Ld$
Centrifugation	Acceleration forces	$a \propto w^2 d \propto d^{-1}$
Detection		
Volume	Fluorescence	$S \propto d^3$
Through surface	Amperometry	$S \propto d^2$
Across surface	Voltammetry	$S \propto d^0$

The Weber number (We) indicates the relative importance of surface tension effects in fluid flow. For a given liquid, We decreases with d, indicating that surface tension becomes more important.[9] On the one hand this means that narrow hydrophilic channels will fill spontaneously by capillary filling, but on the other hand that hydrophobic channels cannot be filled and (hydrophobic) gas bubbles are difficult to remove from channels.

The Fourier number (F_o) indicates the time that is needed for a concentration or heat gradient to even out. The effect of miniaturization is extremely strong here; the time needed decreases with the square of the channel

diameter. This effect makes mixing by diffusion a solution to many mixing problems and allows fast heating and cooling cycles.

The efficiencies of both electrophoresis and chromatography are most effectively described by two terms, the maximum number of plates, indicating the resolution of the separation, and the time needed to perform a separation. Both indicators are shown to improve as d becomes smaller. A good example of a fast separation was demonstrated by Jacobson et al.,[10] who performed the electrophoretic separation of rhodamine B and dichlorofluorescein in approximately 1 ms by applying a field strength of 53 kV cm^{-1}.

The efficiency of centrifuges is determined by the accelerations that they can achieve. On the one hand, the acceleration is proportional to the centrifuge radius, and miniaturization seems counterproductive. However, since the maximum allowed rotation speed of a ball bearing is proportional to its size, smaller centrifuges can rotate faster, leading to improved overall performance.

In one of their field-defining papers,[6] Manz et al. pointed out that the detection signal for a fluorescence detector is proportional to d^3, whereas amperometric signals are proportional to d^2, and refractive index and potentiometric detectors are more or less insensitive to d, where d is some characteristic dimension of the system. Obviously, as miniaturization proceeds, potentiometric detectors will become superior at some point. It remains to be seen, however, if this point will be reached in the near future; until then this is possibly the only remark in the paper mentioned above that has not been prophetic. Apart from the problems with detection that are to be expected upon further miniaturization, a sampling problem will occur when low-concentration analytes are to be determined. Obviously, at least one molecule needs to be present in the sample for detection, and in order to estimate its concentration a few tens of molecules are needed.

2.2 Modeling and Simulation for Miniaturized Total Analytical Systems

The use of modeling tools plays an important role in the design, understanding, prediction and optimization of the performance of parts of μTAS. Ideally, modeling tools should be able to predict the system performance well enough to fabricate a working device in the first process run, thus reducing the development time. Such modeling tools are not limited to the prediction of device performance, but are also used to determine the fabrication steps needed to produce a certain device.

Within the field of modeling programs for a microelectromechanical system (MEMS), recently reviewed by Senturia,[11] microfluidics is considered a ''difficult'' field because of the energy dissipation that is inherent to viscous flow and diffusion as well as the nonlinearity of chemical reactions. An extra complication, commonly occurring in other MEMS applications also, is the combination of the electromagnetic, mechanical and fluidic domain in a single device. In principle, all domains could be included in a single, larger model to be used for simulations that are very economical with regard to computation time (the Newton method). However, commonly available programs combine the effects from the various domains by solving within the domains first and then model the influence of the domains by an iterative procedure. Because of its modular set-up, this approach is easier to implement and allows the addition of new domain solvers at a later stage. However, the method is not computationally efficient, which means that only relatively simple structures can be modeled. Programs based on this method have proved to be effective for the determination of the dispersion of sample plugs in channels under hydrodynamic[12,13] and electroosmotic pumping. A number of commonly used computational fluid dynamics programs are Flowtran by ANSYS (http://www.ansys.com), Flow-3D (http://www.flow3d.com) and FlumeCAD by MEMCAD (http://www.memcad.com). Each of these has its characteristic advantages for use in a microfluidic environment. For example, Flowtran can be coupled to mechanic and electromagnetic ANSYS modules and allows the simulation of fluids having a nonNewtonian viscosity, Flow-3D allows the simulation of chemical reactions taking place in the flowing fluid and in FlumeCAD electrophoresis and electroosmotic flow (EOF) can be simulated.

The modeling of more complicated structures such as complete micropumps takes too much computation time to be modeled effectively in this way. Therefore, a hybrid approach is often more fruitful. In this approach, the various parts of the system are described by lumped elements, the characteristics of which are determined by numerical simulations or from measurements, as required. The various lumped elements are then brought together in a network and the characteristics of the network are determined by a dedicated program such as SPICE or Simulink.[14–16] An extremely simple example of this approach is the modeling of EOF in fluidic networks by electrical networks containing resistors only.

3 MATERIALS

Fabrication of μTAS is dominated by three material classes: silicon, being the most expensive and allowing the smallest structures to be made; polymers, being cheapest; and glass and quartz, playing an intermediate role. Historically, silicon is the most important material

in microsystems engineering, and many of the first microfluidic systems were made of the material.[17] However, for commercial applications, its role as a construction material is being replaced more and more by glass and especially polymers because they offer lower material and machining costs when large series have to be produced.

3.1 Silicon

Silicon is the most tried and trusted material in MEMS in general, and also in μTAS applications.[18] Because of its use in the microelectronics industry, the mechanical and electrical properties and also the etching behavior of silicon are thoroughly known. Moreover, a large number of fabrication techniques are available and the easy metallization of silicon allows on-chip electrochemical detection and the integration of a variety of optical sensors. The on-chip integration of signal processing electronics used to be an often cited reason, but this argument has moved somewhat to the background recently. Moreover, the coupling chemistry between silicon oxide and biomolecules is well developed. This advantage is of importance in the creation of enzyme microreactors,[19] which will be discussed in more detail in section 5.5. Other advantages related to microreactors are the possibility of fabricating silicon walls having a well-defined porosity by anodic etching, as described in section 4.2.1, and the high thermal conductivity of silicon, allowing the fast temperature cycles desirable for on-chip polymerase chain reaction (PCR).[20]

Silicon also has its drawbacks. For commercialization, the fact that wafers are expensive per unit area when compared with glass or plastic is very important. This high cost per unit area can be offset in principle when the devices can be miniaturized far enough, as in microelectronics. These miniaturized devices need to be fabricated by photolithography in a clean room, however, which means very high initial investments in clean room area and mask making if the devices are to be produced in-house. Alternatively, structures can be made by one of the dedicated foundries, a number of which are mentioned at the start of section 4. It should be noted that silicon can be wet-etched outside of a clean room fairly well and that the small feature dimensions, rather than the use of silicon, call for an expensive clean room and lithography equipment. Silicon templates for impression molding of microchannels having a smallest width of 20 μm have been fabricated by isotropic etching and using overhead sheets as lithographic masks.[21]

Its high electrical conductivity makes silicon less fit than plastic or glass for systems that employ high-voltage applications such as electroosmotic pumping or electrophoresis. In applications requiring smaller applied voltages, such as the free flow electrophoresis device described in section 6.4,[22] and in early electrophoretic devices,[23] a thin silicon oxide–silicon nitride film is sufficient to obtain a breakdown voltage of 100–200 V. This breakdown voltage can be increased further when the silicon at the other side of the insulating layer is removed in a subsequent etch step.[24]

Despite its shortcomings, silicon may be expected to keep its important role, especially at the research front and for the demonstration of novel concepts. Once a concept has been proven in silicon, optimization for large product series needed for commercial applications can be performed in glass and, more importantly, in polymers.

3.2 Glass and Quartz

Nowadays, glass and quartz are very extensively used in electrophoretic devices because of their good insulating properties. This allows the high electric fields (over 5500 V cm^{-1}) needed for high electroosmotic pumping rates and fast separations. Moreover, glass and quartz are chemically inert, can be sealed to silicon and are transparent, which allows the use of the optical detection methods discussed in section 7.1.

Because glass has, by definition, no crystallographic orientation, wet etching methods in glass are always isotropic. As with silicon, metallization of these materials is no problem, allowing the precise placement of electrodes for chemical detection at the end of the separation channel. Through holes for inlets and outlets can be made by ultrasonic drilling. Other advantages of glass include its chemical inertness, which makes etching slightly more complicated than the etching of silicon, and the lower cost of glass compared with silicon.

3.3 Polymers

When it comes to physical properties, polymers form easily the broadest group of materials used in μTAS. The properties that most polymers in use share and that are the key to their success in commercially oriented applications are their low cost, ease of machinability and weldability. For the construction of devices, the flexibility of the polymer is important. On the one hand, inflexible polymers poly(methyl methacrylate) (PMMA) and acrylic resins are suited for the definition of complete structures. On the other hand, flexible polymers such as poly(dimethylsiloxane) (PDMS) are generally used in combination with an inflexible polymer or glass backplate to provide the necessary support. However, difficulties have been reported with the accurate definition of channels in PDMS, in addition to problems arising from its permeability toward oxygen.[25]

4 MICROMACHINING TECHNIQUES

Over the last decade, the use of micromachining techniques in microsystems technology (MST), also called MEMS, has moved from the laboratory to the production stage. Miniaturized microphones, pressure sensors and accelerometers for automotive applications are routinely used. While μTAS fabrication technologies have a distinct character, they are commonly rooted in MST. While in MST bulk micromachining (making structures by etching into a wafer) is partly being replaced by surface micromachining (making structures by adding layers to a wafer), in μTAS there is a trend towards using new materials, most importantly plastic.

Because materials and machining are so closely related, the construction material for a μTAS has always to be chosen in combination with an appropriate micromachining technique. When a process is too expensive to be performed in-house, foundries form an interesting alternative. An example of a typical MEMS oriented foundry is MCNC, where under the MUMPs program, both micromachined and Lithographie Galvanoformung Abformung (LIGA)-processed wafer areas can be bought. A number of μTAS-dedicated foundries routinely manufacture glass chips that have customer-designed etched channels, etc. Some of the companies that are reported to perform micromachining for μTAS on a contract basis include the Alberta Microelectronics Center (http://www.amc.ab.ca) in Canada and Twente Micro Products (http://www.microproducts.nl) in The Netherlands.

4.1 Pattern Transfer

With occasional exceptions such as direct laser or electron writing, all micromachining techniques described in the following sections will etch the surface of any part of the substrate that is not protected. Photolithography permits the transfer of a pattern on to a substrate by selective protection of parts of its surface in the following manner. Photoresist [a thin, ultraviolet (UV)-sensitive, polymer film] is spin-coated on top of the substrate to be patterned, and then selectively exposed to UV radiation through a quartz slide having the desired pattern in chromium. Exposed areas of the photoresist are either made soluble (positive photoresist) or insoluble (negative photoresist) in this way. The soluble part of the resist can subsequently be removed by the appropriate developing solution. The patterned photoresist acts as an etch mask, protecting the underlying film during the subsequent etching process. Once the etching of the thin film has been completed, the photoresist mask is removed from the surface in the appropriate solvent or in an oxygen plasma. Obviously, the protecting layer has to remain intact during the etching process, a requirement that is not always easy to fulfill, especially in the more aggressive wet etchants. Therefore, the photoresist layer is often used to define a pattern in a more inert layer. Silicon oxide and nitride, chromium and gold are used to this end. The different materials used for masking will be mentioned separately for every procedure.

4.2 Wet Etching

4.2.1 Silicon

Wet chemical etching is probably the micromachining method for silicon that involves the smallest initial investment. Because it is also one of the earliest developed methods, a large number of different etch systems are well documented. Only a short outline and some examples important for microfluidic devices are given here; the interested reader is referred to the excellent book by Madou.[26]

The most important classification of these systems is by the directional dependence of the etch rate relative to the Si crystal planes: systems that have the same etch rate in all directions are called isotropic, while systems that have different etch rates for different crystal planes are called anisotropic. Examples of the latter systems are KOH, NaOH, hydrazine, ethylenediamine–pyrocathechol and tetramethylammonium hydroxide solutions. The most popular and best documented etching system is the KOH–water mixture. The reasons for the extreme difference between the slow etching {111} planes and the much faster {110} and {100} planes of Si are still unclear. However, etch rates for these planes at various KOH : H_2O ratios and solution temperatures, and also the etch selectivity relative to SiO_2 and Si_3N_4 etch masks, are well documented.[27] When these systems are used in an intelligent way, complicated structures can be realized,[28] an example of which is shown in Figure 3.

Channels that are etched by anisotropic systems usually display a characteristic and well-controlled trapezoid cross-section. The etching of channel junctions is more difficult to control and too much etching easily leads to high dead volumes in the junction. As was mentioned in section 2, some simulation tools are able to predict shapes etched in this way.[11]

For the fabrication of channels that have a (semi)circular cross-section, wet isotropic etching is more suitable. This is typically done in solutions composed of HNO_3, HF, CH_3COOH and/or water. The well-accepted etching mechanism consists in the oxidation of the Si by nitric acid, followed by its dissolution by hydrofluoric acid. The etching proceeds at the same rate in all crystallographic directions, and thus gives rise to circular etch profiles characteristic of isotropic etching. Silicon nitride etch masks are appropriate.

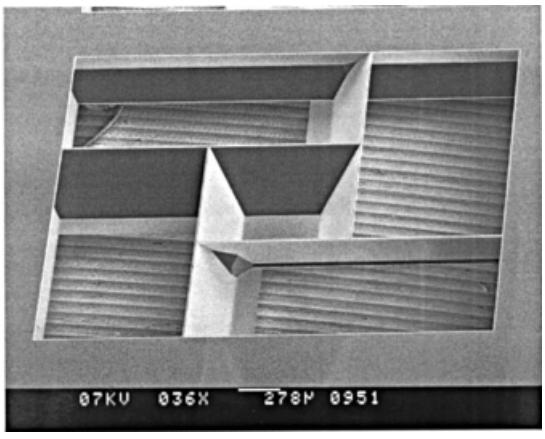

Figure 3 Structure showing the possibilities of anisotropic etching of silicon.

In the anodic etching process, the oxidation of the Si is achieved by application of an anodic current while the material is immersed in an HF–H_2O or HF–ethanol system. At high applied current density, the isotropic etching is complete, whereas at lower current densities porous, oxide-covered Si remains. Using this method, microreactors having porous silicon walls have been made.[19] These can be used to enhance the surface area of the reactor, as discussed in section 6.4. When the current is varied between the high and low magnitudes, empty spaces separated by porous Si can be made and concentric channels can be fabricated,[29] as shown in Figure 4.

4.2.2 Glass and Quartz

Isotropic etching of channels in glass and quartz substrates can be accomplished in stirred concentrated HF–HNO_3–H_2O (20 : 14 : 66)[30] or dilute HF–NH_4F at 50 °C.[31] Channel depths of 5.2 µm have been reported after 20 min of etching using the latter system.[32] Alternatively, buffered oxide etch (BOE) is available from Olin Hunt. While the chemical inertness of glass and quartz is a distinct advantage once the structures are finished, the etchants mentioned above have to be extremely aggressive to achieve reasonable etch rates. This means that special mask materials are needed. Masks are typically made of chromium–gold; polysilicon, which can be used in a subsequent anodic bonding step, has also been proposed.[33]

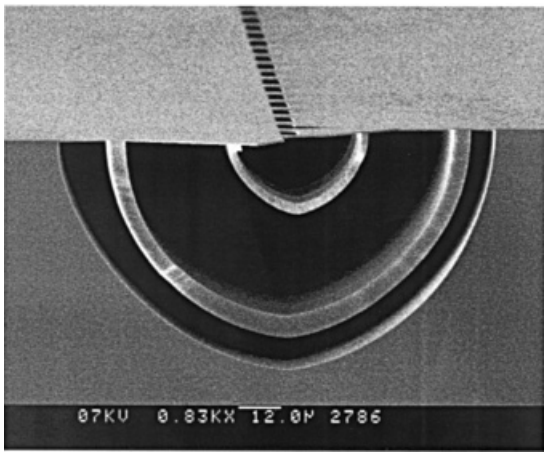

Figure 4 Concentric microchannels fabricated by anodic etching of silicon. The porous walls between the channels can be closed by an extra deposition step if needed.

4.3 Dry Etching

One of the great advantages of dry (plasma) etching is the high degree of anisotropy that can be achieved. Under the right conditions, it is possible to etch almost straight down from the substrate surface. In a standard oxygen plasma, the etching takes place due to physical milling by the oxygen ions. The etch rate can be enhanced by the addition of reactive species such as CF_3. Several micrometers of Si can readily be engraved by this method, called reactive ion etching (RIE). Recent plasma etch systems featuring higher plasma densities have been devised for the production of high aspect ratio features in silicon and in quartz.[34,35] This etching technique, known as deep reactive ion etching (DRIE), offers aspect ratios of at least 15 : 1 and a good selectivity relative to photoresist masking. Because of these high aspect ratios, DRIE reduces the substrate area needed and offers a higher design flexibility than anisotropic etching. When a combined oxygen–fluorine-based plasma is used, the balance between the etching fluorine radicals and the passivating oxygen radicals determines the degree of anisotropy of the process. In this way, vertical walls and positively and negatively tapered channel walls can be etched.[36]

4.4 Polymer Molding

Polymer molding is a technique that has relatively high start-up costs because a master template has to be microfabricated. However, once such a template is available, large numbers of chips can be fabricated, so that the variable cost per chip is acceptable. Therefore, this is a fabrication method that appears most popular for commercial applications.

4.4.1 Injection Molding

In thermoplastic injection molding, a polymer is simply heated to a temperature above its melting point, injected into the mold and cooled. This method is most suitable when large series have to be produced.

For prototyping, however, reaction injection molding is more convenient. Typically, a polymer and its curing agent are poured or injected in a mold where they are left to cure in the desired shape, which may take up to several hours. When the parts are ejected too early, sharp-edged features can become rounded, leading to a different channel cross-section than was designed. Using this method, PDMS (Sylgard 184™, Dow Corning) was cast molded in metal[37] and silicon molds,[38] both made using photolithography. The minimum dimensions that can be molded are determined by the fabrication method that is used for the master.

In order to obtain as many plastic chips as possible from the expensive silicon master template, a number of daughter templates can be fabricated from the master by nickel electroforming. Each nickel daughter can then be used for injection molding of the resin.[39]

4.4.2 Hot Embossing/Impression Molding

The hot embossing process, which is also known as ''impression molding'' or ''relief printing'' process, has been used for decades and features with sizes smaller than a micron are routinely fabricated in thermoplastic polymers by impression with a micromachined template. A well-known example is the fabrication of compact disks in polycarbonate (PC). In its most simple form, channels can be made by pressing a thin wire in a polymer chip. When two such chips are mounted perpendicularly, a very simple electrophoresis device complete with injection cross can be fabricated.[21] The quality of channels and cross-sections that are made in this way is far from perfect, however. Therefore, silicon or metal micromachined templates for complete channel structures are more useful. The fabrication of these templates is discussed in the next subsection.

Solvent-assisted molding is an interesting variation on this theme.[40] In this technique, the polymer is softened not by heating but by solvent absorption. This can be achieved by wetting a PDMS mold with an appropriate solvent and bringing it in contact with the surface to be impressed. The solvent dissolves or swells the surface and the PDMS structure is imprinted on the surface. The polymer solidifies as the solvent evaporates. Acrylonitrile–butadiene–styrene (ABS), polystyrene and cellulose acetate surfaces have been shaped in this way, using acetone as a solvent. Free-standing structures can be fabricated using this technique by shaping a thin film of photoresist and subsequently etching away the thinner part of the film.

The structures that are made by the methods described above can be used in turn as stamps for pattern transfer. This method is especially fit for the transfer of patterns having features of the order of tens of nanometers that cannot be transferred by the lithographic methods described in section 4.1. A PDMS stamp was wrapped around a cylinder to facilitate the printing process and was used for the application of masks for wet etching.[41] A PMMA film was patterned by impression molding and the impressed parts were removed by RIE and the resulting part was used for lift-off.[42]

4.4.3 Template Fabrication

The large number of different ways of fabricating templates is illustrated in Figure 5. For large structures, conventionally computer numerical control (CNC)-machined templates may be made in stainless steel. Templates having smaller structures can be made by etching of silicon. In order to facilitate the release of the polymer chips, a thin fluorocarbon coating may be applied to the silicon.[43] Anisotropically wet-etched silicon templates have been used to emboss PMMA at 135 °C for 5 min, yielding channels having a trapezoid cross-section 61.5 μm deep and 15.5 μm wide at the bottom.[21] Silicon templates fabricated by dry etching were used to emboss structures only 800 nm wide and 5 μm deep in PMMA and PC.[44]

The mold material needed depends strongly on the material that has to be molded. For example, molding of flexible polymers such as PDMS can be achieved using a photoresist template.[45] Metal molds having small features can be made by electroplating of silicon structures or thick photoresist,[43] or by LIGA, as will be discussed in the next section.

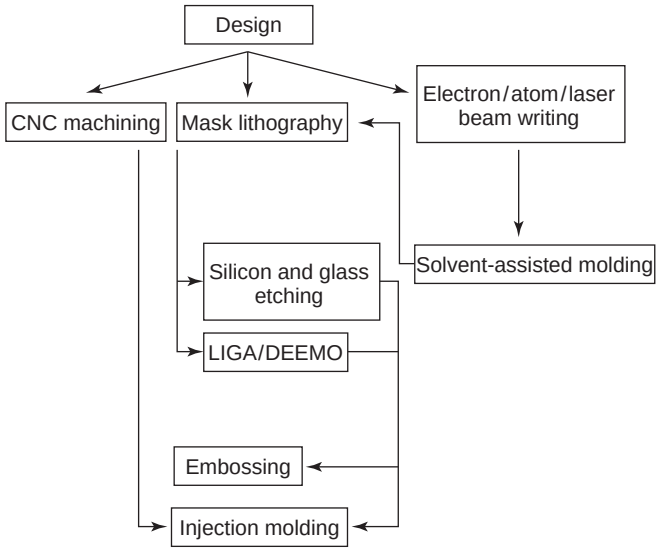

Figure 5 Micromachining techniques. Top layer: expensive single products. Middle layer: small batches of finished products and templates for the bottom layer. Bottom layer: mass production of inexpensive structures. (DEEMO: Dry etching, electroplating, molding.)

Molds do not have to be rigid, and in fact flexible templates may be at an advantage for the release of small structures and for the fabrication of structures on a nonflat substrate.[46] PDMS templates were fabricated by pouring the polymer over a chromium master. These templates were used for the fabrication of photochemically curable polyurethane (PU) structures only 8 nm wide.[47]

4.5 The Lithographie Galvanoformung Abformung Process

The LIGA process combines lithographic and molding techniques.[48,49] The lithography part is performed by X-rays making the irradiated part of a thick (micrometers to centimeters) PMMA film soluble. The X-rays are needed to provide the high energies that are needed to break the chemical bonds in the cross-links of the polymer. In the next step, the PMMA mold is filled by electrodeposition, resulting in a free-standing metal part. This part can be the end product, but may also be used in turn as a mold for the injection molding and hot embossing processes described in the previous section.

The main advantages of the method are the high resolution ($<0.2\,\mu m$) and aspect ratios, and also the smooth surfaces that can be obtained.[50] The main disadvantage is the need for a synchrotron to produce the X-rays.

4.6 Thick Photoresists

When submicrometer resolution is not a necessity, thick photoresists form an interesting alternative.[51,52] This new technology is based on an epoxy-based, thick negative-tone photoresist that can readily be patterned by classical UV sources. It is primarily composed of Shell SU-8™ Epon resin, photosensitized with a triarylsulfonium salt. Its low optical absorption in the near-UV spectrum favors good aspect ratio structures. Patterning of features up to 1200 μm in size with an aspect ratio of 18 has been demonstrated. Applications will be similar to those of the LIGA process, with a reduced resolution and aspect ratio, but at far lower cost. A number of SU-8™ applications have been reported.[53,54]

4.7 Sealing of Channels and Reaction Chambers

The fabrication methods discussed above typically provide only three walls of a channel or reaction chamber. The fourth wall has to be provided by fixing the micromachined part to a second, usually flat surface. The methods that are available to achieve this depend strongly on the materials used.

Silicon can be bonded to glass by anodic bonding.[55] Corning #7740 borosilicate glass (Pyrex®) is widely used for this purpose as its thermal expansion coefficient matches almost perfectly that of silicon. Both substrates are brought into contact and heated at a temperature in the range 350–500 °C so that the Pyrex® glass becomes slightly conductive. If a voltage between 200 and 2000 V is applied across the wafers, with the glass being biased negatively, an electrostatic force will pull both substrates into intimate contact. It is assumed that the oxygen ions leave the glass and chemically bind with Si through the formation of SiO_2. Note that anodic bonding can be used on bare silicon wafers as well as on wafers covered with thermal SiO_2, sputtered glass, polysilicon or LPCVD (low-pressure chemical vapor deposition) Si_3N_4.

By fusion bonding, Si can be bonded to Si, Si to glass, glass to glass and quartz to quartz.[56–58] The surfaces of the materials to be bonded are first hydrolyzed to create a high density of hydroxyl groups on their surfaces. Then the two parts are joined and annealed at high temperature for a long period (1100 °C for 5 h). The anneal step strengthens the bonding through a series of chemical reactions that presumably are the water formation from the OH groups, followed by its dissociation and the reaction of oxygen with Si.

Polymer channel structures can be sealed by lamination. Injection-molded acrylic chips have been laminated using Mylar® sheet coated with a thermally activated adhesive at 105 °C for 5 min,[39] while the application of poly(ethylene terephthalate) (PET)–polyethylene (PE) films at 125 °C has also been reported.[59] Alternatively, two molded plastic parts can be sealed together within less than 1 s by ultrasonic welding.[60] In this method, ultrasonic energy is brought to the contact area between the two plastic chips. When the two plastic chips to be sealed only touch in specifically designed places, energy can be delivered very efficiently to these spots only, and local welds can be made. Using this method, ABS and polystyrene have been reported for the fabrication of blood testing devices discussed in section 8.3.

Leak-free sealing of channels made of flexible polymers such as PDMS has been reported by simply placing the polymer on a glass support.[38] The resulting structures have been used for electrophoresis. Not only does this type of bonding take place immediately at room temperature, it is also reversible: the PDMS can be removed from the glass without too much trouble.

5 BASIC HARDWARE COMPONENTS

In this section, examples of a number of the basic components that can be used in microfluidic environments are discussed. For a more in-depth treatment, the reader is referred to the reviews by van den Berg et al.,[61] Elwenspoek et al.[62] and by Gravesen et al.[63] In the

literature, the parts described in this section are often given the prefix micro-. Since we deal exclusively with miniaturized components here, the prefix is omitted.

5.1 Channels, Junctions and Connectors

Channels form the backbone of any microfluidic system. In many cases, the channel is not only used for sample transport but also for separation as in electrophoresis, or as a reaction chamber. Cross-sections having a semicircular shape are routinely made by isotropic wet etching in glass, while anisotropic etching in silicon generally results in trapezoid cross-sections. The influence of the channel cross-section on the performance of the device seems to be limited, however, possibly because other problems concerning dead volume, etc., are still dominant at this point. An interesting idea that goes in an entirely different direction is the confinement of fluid streams by hydrophilic stripes on a hydrophobic substrate.[64] The fabrication of such stripes by microcontact printing, described in section 4.4, is extremely simple. However, it is far from obvious how, for example, pumping through such wall-less channels can be achieved, and no applications have yet been demonstrated.

While the fabrication of channels is straightforward, junctions may lead to dead volumes if they are not carefully designed and fabricated. Especially the connection of the μTAS device to the outside world by zero dead volume connectors is an issue that receives considerable attention. In the laboratory, connectors are often made by gluing capillary tube perpendicular to the substrate into holes connecting to the channels of the μTAS. However, such connections are difficult to align and insert, and are unable to withstand high pressures or voltages. In contrast, when the connections are properly designed and fabricated, as shown in Figure 6,[65] connectors can be made having dead volumes of less than 0.5 nl that can withstand powers up to 120 bar. The dead volume may be minimized further by etching a sleeve in the silicon connector and connection without glue is possible through the use of injection-molded couplers.[66]

Alternatively, connectors can be made parallel to the substrate.[67] This way of connecting allows a greater contact area between the substrate and the connection and is self-aligning. Moreover, concentric tubes can be connected with great ease, allowing connection to standard hollow fibers used in microdialysis.

5.2 Valves

One of the main requirements for the functioning of a μTAS is the ability to move fluid plugs over a predetermined trajectory. When electroosmotic pumping is used, the flow direction at junctions can simply be directed by application of a voltage to the appropriate channels. In contrast, pressure-driven systems need active valves that use a force perpendicular to the flow direction to push a restriction in the shape of a diaphragm or cantilever into the channel. These valves are usually made by bulk micromachining in silicon,[68] while other techniques such as LIGA[69] and polymer injection molding[70,71] are sometimes used. An important requirement in any design is whether the valve should be open or closed when the power goes down (some valves have more than one stable position). From this requirement, the valve types ''normally open'' and ''normally closed'' can be distinguished.

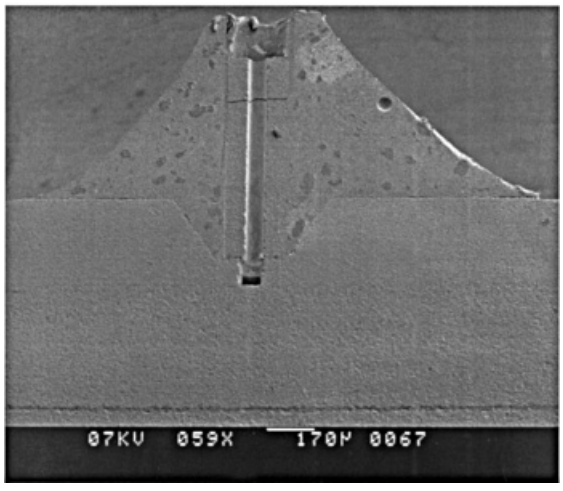

Figure 6 Low dead volume connection between capillary and chip. (Reproduced by permission from V.L. Spiering, J.N. van der Moolen, G.-J. Burger, A. van den Berg, 'Novel Microstructures and Technologies Applied in Chemical Analysis Techniques', *Tech. Dig. Transducers*, **1**, 511–514 (1997).)

Most active valves are micromechanics based and are miniaturized versions of macroscale valves. In these active valves, the fluid flow can be directed by piezoelectric,[72] electrostatic[73] or electromagnetic[74] actuation. Recently, two different ways of electrochemical actuation of a valve diaphragm have been proposed. The valve shown in Figure 7[75] employs the pressure that is generated when a $CuSO_4$ solution is electrolyzed, at two planar electrodes patterned by lift-off. The $1 \times 1\,mm^2$ diaphragms are made of 1-μm thick low-stress LPCVD silicon nitride or of 8-μm thick polyimide. The Cu electrode has been protected against oxygen gas with a layer of Nafion® (gel). To improve the adhesion of Nafion® on a Cu electrode, a polyimide mesh is used. The principal advantage of this system is that power is only needed to open or close the valve, but not to keep it in a stable position. This is of importance for remote devices that carry their own batteries. A second method makes use of the volume changes that occur in conducting

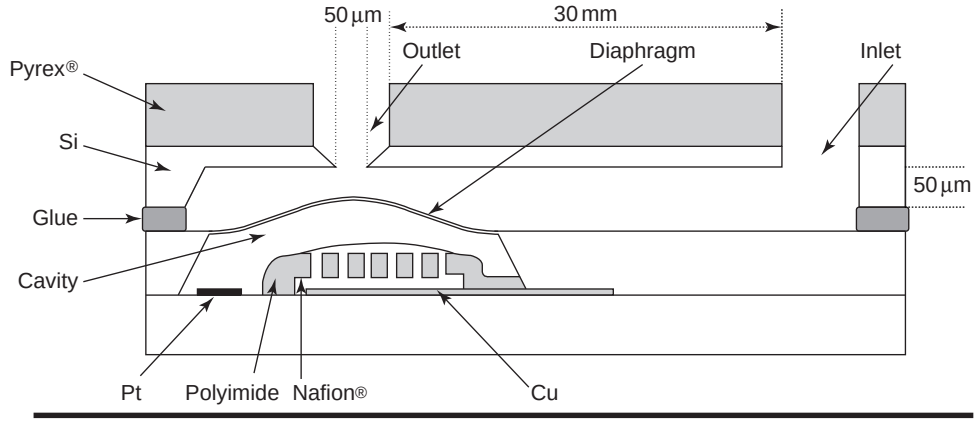

Figure 7 Cross-section of an electrochemically actuated active valve. The valve, which is made by bonding of micromachined silicon and Pyrex® wafers, can be used for flow regulation and uses energy only when it is adjusted.

polymer films upon cation insertion. When such a film is deposited on a thin, free-standing gold film, volume changes result in bending of the polymer–gold bilayer, as was demonstrated for polypyrrole.[76] Valves and pumps based on this principle have been proposed.[77]

An entirely different class of valves is based on properties that are characteristic for microfluidic systems only. For example, a new type of valve is formed by freezing and melting of a part (plug) of a fluid passage. This type is also called the microelectrothermofluidic (METF) valve.[78] The main advantage of this type of valve is the completely zero dead volume and absence of moving parts. A similar effect can be achieved by the use of plug of a ferrofluid that can be actuated by a magnet.[79]

An interesting example of the use of the characteristic properties of microfluidics, the simple but very effective capillary-driven valves, is worth mentioning. These valves have been used for precise positioning of sample plugs by exploiting the relative importance of surface tension in microfluidics. A change in the diameter of a channel will work as a fluid stop, and unidirectional valves can be made by giving the channel the proper shape.[80] Moreover, if the fluid is pushed through a valve by an external pressure compensating the surface tension, it will keep flowing until a next valve is met. Using this technology, complete systems can be made.[81] Because of its high surface tension, all reported devices are machined in polymers.

While active valves are mainly used to direct fluid flows at channel junctions, passive valves find their main application as check valves in micropumps, allowing fluid to flow in one direction only. Some examples are given in the next section.

5.3 Pumps

Moving fluids in the µTAS requires very precise pumping and dosing in the nanoliter range. While high pump rates may be of importance in some cases, often the reproducibility and stability of the pump rate, combined with the capability of the pump to prime itself, is of more importance. A large number of different pumps have been demonstrated over the years.

5.3.1 Micromechanical Pumps

One class employs miniaturized classical concepts using a membrane and check valves. As an example, a pump based on piezoelectric actuation of the membrane is shown in Figure 8.[82] After the initial work at the University of Twente, other groups optimized[83,84] or modified[85] the concept, and probably the most advanced micropump to date is the electrostatic bidirectional pump presented by Zengerle et al.[86] Pumps actuating the membrane by electromagnetic,[87] thermopneumatic[88] and piezoelectric[89,90] actuators have also been reported. These miniaturized ''classical'' pumps all suffer from the same problems: they have moving parts that limit their lifetime and are not very tolerant to particles in the

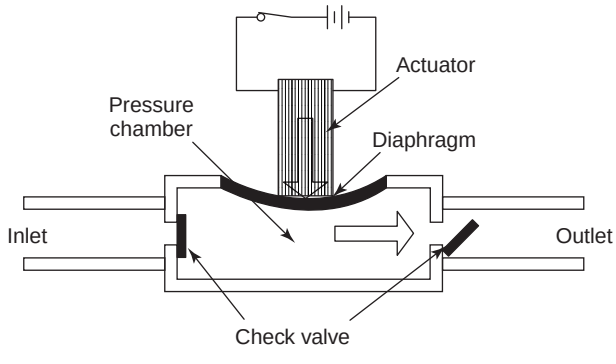

Figure 8 Diagram of a check-valve micropump with piezoelectric actuation.

sample stream. Moreover, they tend to produce pulsed flow patterns rather than a constant flow rate.

The problems with moving parts in valves have led to the development of pumps employing nozzles instead of check valves.[91] A pumping action is achieved because the flow restriction imposed by the nozzle depends on the flow direction, as indicated in Figure 9(a) and (b). These valves are called dynamic valves because their working principle is based on the (turbulent) fluid dynamics in the pump chamber. A similar effect can be achieved when the viscosity of the fluid is changed at the pump inlets and outlets by heating.[92] A second class of pumps relies on mechanisms that are effective in miniaturized systems only. The main advantage of these pumps, which are described below, is that they have no moving parts.

5.3.2 Electroosmotic Pumping

The most widely used pumping mechanism, EOF, relies on the presence of (immobile) surface charges on the channel wall. When a high voltage is applied the mobile counterions start to move with respect to the charged wall and drag the rest of the fluid column with them. This phenomenon is called EOF. It should be noted that this type of pump is of the ''current–source'' type. This means that the pressures that can be obtained depend on the internal flow resistance; in wide glass tubing with little resistance, very little pressure can be built up and a very small hydrostatically induced differential inlet/outlet pressure immediately overrules the electroosmotic pumping. However, in the narrow (<100 μm diameter) capillaries commonly used on chips, relatively high pressures can be obtained (up to tens of bars). Moreover, when the capillary is packed with silica beads, the surface to volume ratio of the channel increases further and pressures in excess of 8000 psi can be achieved.[93] Samples can be manipulated across junctions by application of different voltages to the ends of the respective channels, as will be discussed in section 6.3 on electrophoresis. Disadvantages of EOF pumping are the need for high voltages to achieve reasonable pumping speeds, the occurrence of electrolysis at the electrodes and electrophoretic side effects.

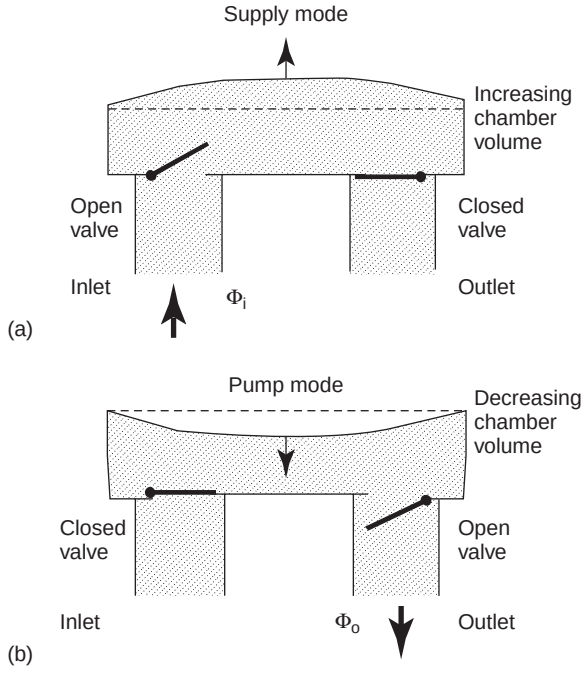

Figure 9 Cross-section of a diffuser nozzle pump. (Reproduced by permission from E. Stemme, G. Stemme, 'A Valveless Diffuser/Nozzle-based Fluid Pump', *Sens. Actuators A*, **39**, 159–167 (1993).)

5.3.3 Capillary Pumping

An elegant way of pumping in microchannels is by capillary action.[94] Fluids are simply drawn at a reproducible speed into channels having hydrophilic walls. Valves for capillary pumping can be constructed in several ways, as was discussed in the previous subsection.

5.3.4 Thermocapillary Pumping

Thermocapillary pumping is used to move discrete drops rather than continuous streams through microchannels. It is based on the temperature dependence of the surface tension between the droplet and the channel wall. As one end of the droplet is heated by 20–40 °C, the surface tension at that end decreases so that the internal pressure in the droplet increases enough to move the droplet towards the cooler end.[95]

5.3.5 Electrohydrodynamic Pumping

Electrohydrodynamic pumping is of interest because it allows the pumping of fluids having a low conductivity, such as organic solvents and low ionic strength solutions.[96] The pumping activity is achieved by an inhomogeneously induced polarization of the fluid, an effect that has been known for a long time. The exact mechanism(s) involved appears to be open to discussion, however. A similar, but better understood, actuation principle is used for transport of particles and cells, where polarization differences may be expected since the fluid is not homogeneous. Advantages are the absence of moving parts and electrolysis, leading to a good sample integrity, and the absence of bubbles.

5.4 Filters

Nowadays, most current applications of μTAS are in drug discovery and genetic engineering, where sample solutions are relatively clean and well defined. In contrast,

the use of a μTAS for the on-line monitoring of industrial processes and in biofluids or waste water requires a microfilter to separate the system from the dirty ''outside world'' to prevent the blocking of channels, valves and pumps by dust particles, sand grains and living cells.

Sampling modules using conventional polymer microfilters (e.g. a Millipore™ filter) are an option. However, the anticipated problems with the large-scale integration of conventional membranes have prompted several researchers to make micromachined filters. The stream to be filtered can either flow perpendicular to the silicon substrate through etched holes,[97] as shown in Figure 10, or parallel to it through holes formed by the removal of a sacrificial layer,[98,99] as shown in Figure 11. Apart from the ease of integration, perpendicular micromachined filters can be made to have a large variety of well-defined pore sizes and shapes, extremely low pressure loss and very small internal volume. Pore diameters down to 100 nm and fluxes of 4000 $L\,m^{-2}\,h^{-1}$ have been realized.[100] On the other hand, the pore size of parallel filters is determined by the thickness of the spacer layer, which can have very small and uniform dimensions. The pressure loss across the filter can be kept low by placing large numbers of parallel filters in parallel, albeit at the expense of the increase of the dead volume.[101]

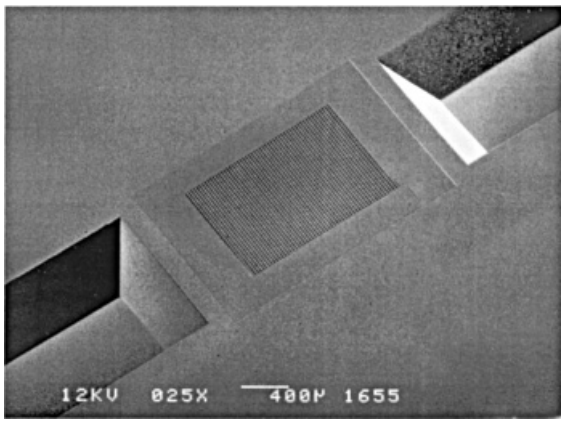

Figure 10 Perpendicular silicon microfilter, which has also been used as a mixer.

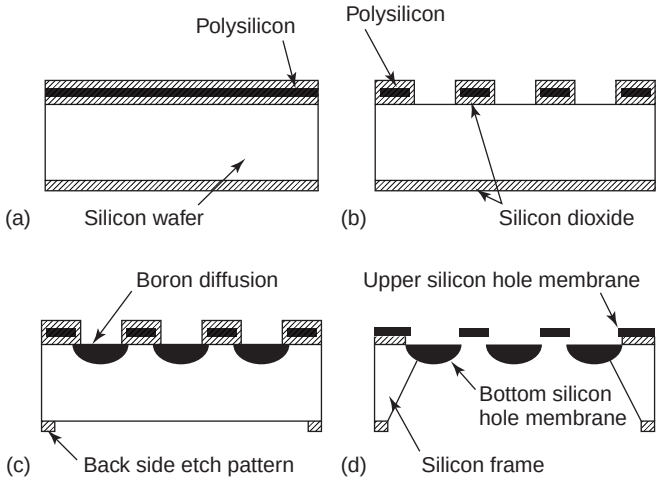

Figure 11 Planar filter made on silicon. Pore size is defined by the thickness of the separator film. (Reproduced by permission from G. Stemme, G. Kittilsland, 'New Fluid Structure in Silicon Fabricated Using a Self-aligning Technique', *Appl. Phys. Lett.*, **53**(16), 1566–1568 (1988).)

5.5 Mixers and Reactors

In a μTAS, the sample stream may have to be mixed with reagents to produce fluorescent products,[102] with enzymes or with cell lysates. In many cases this can be effectively done by a simple channel junction, but even then mixing becomes important and the channel may be referred to as a microreactor. When electroosmotic pumping is used, different mixing ratios can be realized to a certain extent by manipulation of the channel voltages.[7] When the sample streams to be mixed differ greatly in size, mixing in several stages tends to be more accurate, which can readily be achieved on a single chip.[103]

Microreactors form a fast expanding field of research, not only as μTAS building blocks, but also for the synthesis of small amounts of compounds using extremely rare ingredients, as is the case in drug discovery, or for the on-site and on-demand production of toxic or explosive chemicals.[104] Miniaturized reactors are different from macroscale reactors for a number of reasons, as described below.

5.5.1 Mass Transport

Stirring in the conventional sense is hardly possible because of the dominance of viscous effects in flow, as was discussed in section 2. Therefore, mixing in addition to chemical reaction tends to be dominated by diffusion if no special measures are taken. In many cases, this is no problem whatsoever because the distances within a microreactor can be covered easily by diffusion. If needed, there are two strategies to speed up mixing, by reduction of the diffusion distance or by induction of turbulent mass transport.

In its simplest form, the diffusion distance can be reduced by a decrease in channel width.[105] A more sophisticated way to reduce the diffusion distance is by perpendicular mixing. The experiments using this method were first performed using the perpendicular filter that was shown in Figure 10,[106] while the concept was later refined by other researchers.[107,108]

Mixing can also be sped up by inducing nondiffusional mass transport. First, at high enough flow rates, turbulence can still be induced. Extremely fast mixing, having a dead time of only 0.1 ms, has been demonstrated in ±300-μm diameter channels at flow rates of a few milliliters per second.[8] Another way to induce turbulence is by ultrasonic actuation from a piezoelectric transducer, a method that is as yet used only sporadically in μTAS,[109] but has gained much attention in other fields of chemistry.[110] Apart from a number of other effects such as the formation of hot spots, mass transfer is an important phenomenon induced by ultrasonic actuation. Acoustic streaming is induced by the sound field entering the fluid while other forms of mass transport are connected to the oscillation and collapse of bubbles that are formed by the ultrasound.[111] A second way to induce nondiffusional mass transport is by moving the fluid back and forth through a nozzle/diffuser. As was discussed in section 2.1, turbulent flow can be achieved even at these small scales. A third way to induce nonlaminar flow is by moving around discontinuous fluid plugs rather than continuous streams. The recirculation patterns developing in such plugs provide an elegant way of mixing.

5.5.2 Surface to Volume Ratio

Microreactors have a far higher surface to volume ratio than do normal reactors. Therefore, heat exchange with the surroundings is much more effective, allowing fast heating and cooling. These short temperature cycles are useful in, for example, fast PCR reactions, as will be discussed in section 8.4. Moreover, the large surface area can be used for the immobilization of catalysts, such as enzymes. For example, the oxidation of glucose by glucose oxidase (GOD) was enhanced when the enzyme was attached to 100–250-μm deep lamella or V-grooves wet etched in silicon wafers, rather than to the flat wafer surface.[112] The reactor surface area can be increased further by the anodic etching procedure described in section 3.1. GOD attached to the porous silicon resulting from this process showed an activity increase by a factor of 100.[19]

5.6 Dispensers

Although the manipulation of small volumes within the μTAS is relatively easy, bringing these volumes into the system is generally impossible using normal pipets. Therefore, structures that are able to dispense small, well-defined droplets of analyte are being developed. The design of most dispensers is based on that of commercially available ink-jet printer nozzles. The analyte droplets can be applied to an array to be analyzed later[113] or directly be accelerated into the analysis chamber of a mass spectrometer, as discussed in section 7.3. An especially interesting dispenser structure that can be inserted in the flow line of a channel structure is shown in Figure 12(a) and (b). It was fabricated by assembling a micromachined silicon nozzle to a membrane made by the same technique. Droplet volumes between 30 and 200 pL could be ejected at a maximum frequency of 500 Hz.

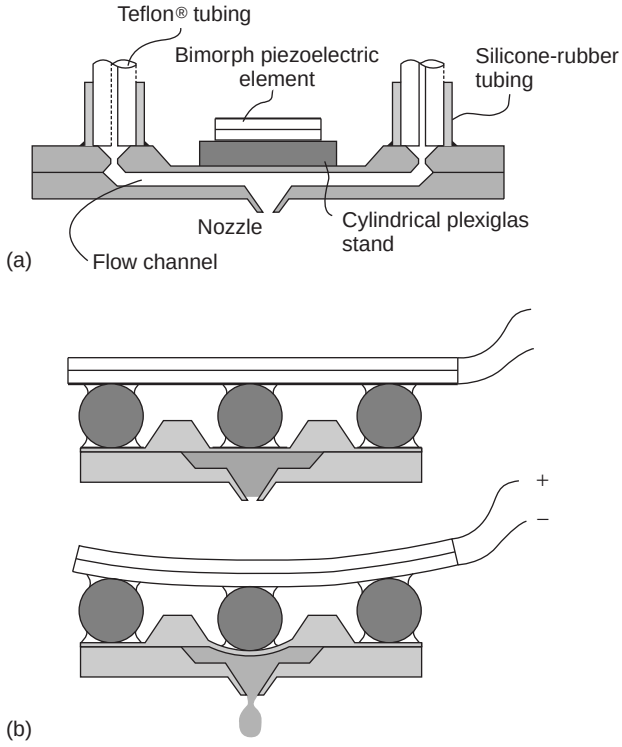

Figure 12 Drop dispenser that can be placed in a microchannel. In a typical application, only a small part of the fluid passing through the channel will be dispensed.

6 SEPARATION TECHNIQUES

Separations in μTAS modules have different applications, influencing the choice of the technique to be used. The dialysis and filtration techniques discussed in section 6.1 are mainly used to prevent the malfunctioning of the μTAS device owing to obstruction by particles, cells and dirt. In order to fulfil this task, at least one side of such a filtration or dialysis module needs to be able to withstand the conditions it has to keep out of the system. Other separation techniques are focused on the separation of a sample into its different components, allowing subsequent analysis. The requirements of devices performing these functions are totally different from those performing the first.

6.1 Dialysis, Filtration and Centrifugation

The rationale behind dialysis and filtration as a first step in a μTAS was explained in section 5.4. In microdialysis sampling, polymeric hollow-fiber membranes are used to separate large biomolecules and cells from the analyte. This technique, introduced about 10 years ago, has now become a standard technique in neurochemical laboratories.[114] These hollow fibers have been integrated with biosensors[115,116] and serve as sample cleanup modules for mass spectrometry (MS).[117] Because the smallest fiber diameters available today are rather large (200 μm), a counterflow dialysis module was micromachined by sandwiching a dialysis membrane between two polymer chips that had laser-ablated channels in their surface.[118]

Whereas many filters have been proposed and developed, actual ultrafiltration modules have not, to our knowledge, been applied in actual systems.

As was discussed in section 5.5, the strictly laminar flow in microfabricated channels makes diffusion play an important and reproducible role in mass transport. The diffusion coefficient increases with increase in molecule size, permitting separation methods based upon the diffusion speed, and thus on molecule size, without the use of a membrane.[119] Obviously, the method is most effective when the fractions to be separated have very different diffusion coefficients. The principle has been demonstrated, but accounts of actual separations are still to be published. Both sides of this separation unit are connected to microchannels and therefore the method is probably less suited for precleaning stages.

Moreover, obstacle courses for the trapping to branched DNA molecules[120] have been developed on-chip.

Centrifugation is a very effective way to separate particles and cells from fluids as a precleaning step, especially for DNA analysis, where the cells contain the sample. At first sight, centrifuges may not seem likely candidates for miniaturization; in Table 2 it was shown that the separation efficiency of a centrifuge is proportional to its diameter. However, the speed rating of a bearing is proportional to its size, so that smaller centrifuges can be run at higher speeds. Using conventional machining techniques, an 8.6-mm diameter microcentrifuge was built that allows 40 times faster separations than conventional centrifuges.[121] The centrifuge is of the flow-through type and can be used in arrays compatible with standard 96-well microtiter plates.

6.2 Chromatography

Gas chromatography (GC) is the first analytical procedure that was performed on a chip, at least 10 years before the μTAS concept began to take shape.[5] The advantage of miniaturization of the chromatograph was believed to be its reduced size and weight, rather than its improved analytical performance. Recently, however, a 10 μm × 100 μm × 2 m on-chip separation column was demonstrated that does exhibit a separation performance that is superior to that of a conventional column.[122]

In contrast, when Manz et al. fabricated their first on-chip open-channel liquid chromatograph, the μTAS concept had just started to develop.[17] Their chromatograph employed a normal off-chip high-performance liquid chromatography (HPLC) pump, a 6 μm × 2 μm × 15 cm channel, anisotropically etched in silicon, and on-chip conductometric detection. Later, the reversed-phase separation of coumarin 440, 450 and 460 was demonstrated using an open channel etched in a glass chip with on-chip electroosmotic pumping and off-chip laser-induced fluorescence (LIF) detection and 4–5-μm plate heights were realized.[123] The use of conventionally packed columns is complicated by the nonuniformity of the packing at the channel walls (the increase in surface to volume ratio upon miniaturization is a disadvantage in this application). This problem can be solved by micromachining the ''column packing'' in addition to the channel walls. Recently, RIE has been used for the fabrication in quartz of the collocated monolith support structures (COMOSS) shown in Figure 13(a) and (b).[124] The separation efficiency of rhodamine 123 in reversed-phase coated structures of this type was 35 000 plates ($h = 1.29\,\mu m$).

6.3 Electrophoresis

Capillary electrophoresis (CE) is the most commonly used separation technique in μTAS today. The great separation efficiency, the absence of moving parts to induce fluid flow and the availability of electrophoretic schemes for on-chip sample injection and addition of reactants make it an exceptionally potent technique. There are many excellent books introducing the broad field of electrophoresis, and one of the most recent has a separate chapter discussing on-chip electrophoresis.[125] The interested reader is referred to these for a more in-depth treatment of this subject. Compared with CE, on-chip electrophoresis can be faster and more efficient when short separation distances and high separation fields are combined with very short injection plugs and integrated detection systems.

Briefly, electrophoretic separations are achieved by application of a high voltage between the two ends of a liquid-filled channel, so that each individual molecule inside the channel migrates in the electric field according to its own mobility. The mobility of the ions decreases with the ratio of their mass to charge (m/z) so that small, highly charged molecules reach the end of the channel before larger, singly charged molecules. When

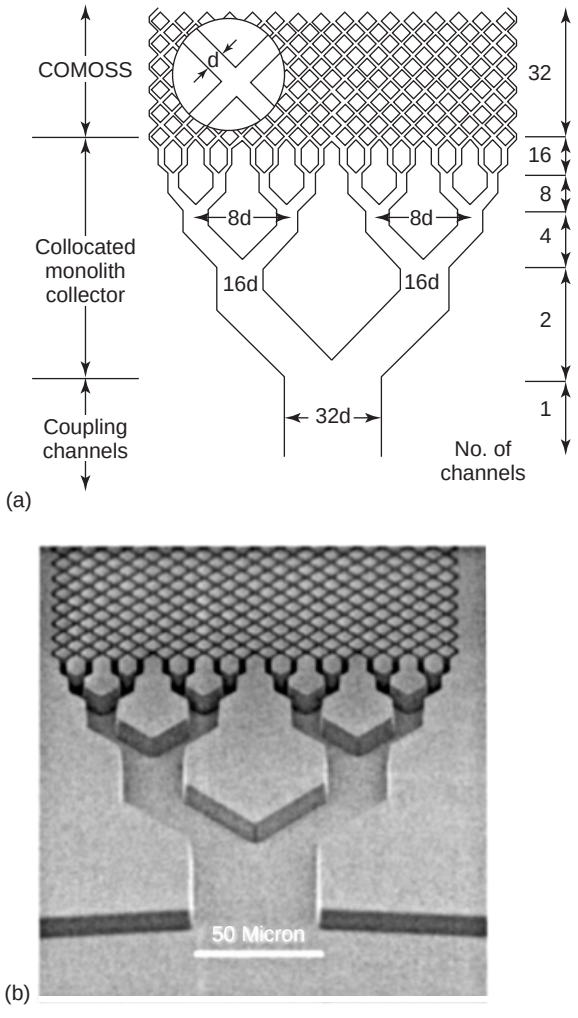

Figure 13 COMOSS, part of a micromachined liquid chromatography (LC) column. (Reproduced with permission from B. He, N. Tait, F. Regnier, 'Fabrication of Nanocolumns for Liquid Chromatography', *Anal. Chem.*, **70**(18), 3790–3797 (1998).)

the separation is efficient, concentrations of the different molecules are detected as Gaussian peaks having a standard deviation σ. When the m/z values of the different molecules are known, a single, nonspecific detector can be used to distinguish between the different molecules present. As an example, on a capillary gel electrophoresis chip that was optimized with respect to separation efficiency, separations having efficiencies of $N = 200\,000$ ($h = 0.24\,\mu m$ at $L = 70\,mm$) were realized within 5 min.[126] The separation efficiency is determined by the following factors:

- the length of the injected sample plug;
- the length of the area sampled by the detector;
- the diffusion of the ions during the separation, which is increased by Joule heating.

Because on-chip separations usually involve relatively short separation times, the diffusion is relatively unimportant. The detector length is only of some importance in ultrafast (<150 ms) separations in extremely short (900 μm) channels.[127] In all applications, the length of the injected sample plug plays the most important role and should be kept as narrow as possible. There has been a steady development in injection schemes. The first injections, into a simple T-junction where a voltage is applied to two channel ends while the third is floating,[128] were reproducible to within 3%,[129] but suffered from mobility bias (highly mobile ions are over-represented in the sample), from leakage from the floating channel and from peak broadening.[30] Various solutions to the problem have been demonstrated. More sophisticated junction layouts, such as double T-junctions,[130] and more sophisticated voltage control of all side channels[131] have enabled the separation efficiencies mentioned above to be achieved.

CE has many different specialized varieties, and a number of these have been demonstrated on chip, including isotachophoresis (ITP),[132–134] isoelectric focusing (IEF),[135] gel electrophoresis[136] and micellar electrokinetic chromatography (MEKC).[137] Of these methods, gel electrophoresis is the most extensively used because of its usefulness for the separation of DNA fragments.

The detector measures a sequence of separated sample components during a certain time window, which makes the method unsuited for on-line monitoring. This problem has been addressed by moving the output of the electrophoretic channel by a stepper motor along the 15-mm side of a 21 μm × 15 mm × 48 mm channel.[138] At the other end of the channel, a 100-electrode-wide amperometric sensor array served as a detection unit. The electrode number provided a measure of the position along the channel side, which could be translated to a certain sampling moment.

6.4 Sorting and Manipulation of Large Biomolecules, Particles and Cells

One of the frequently cited advantages of μTAS is the avoidance of contamination of analyses by on-chip sorting of the relevant cells from a culture, cell lysis, DNA amplification and analysis. A number of these devices are discussed in section 8.4.[139] However, the on-line sorting and manipulation of cells received little attention until a few years ago.

The behavior of cells in constrained environments is complicated. However, the assumption that they behave as particles of the same size is good enough for most of the applications discussed here. The electrophoretic motion and manipulation schemes used for high-molecular-weight samples can be used for whole cells[140] and large

molecules or particles are often used as model systems for cell manipulation.

One way of sorting large biomolecules is by free-flow electrophoresis: a sample stream is fed into the structure. A voltage is applied between the left- and right-hand sides of the cell so that, as the stream flows from the top to the bottom, the sample is separated according to its electrophoretic mobility. A certain fraction of the sample can be led to a unit for continuous detection. Such an on-chip system was demonstrated to be competitive with a commercial instrument for the separation of large biomolecules and protein tryptic digests and for the purification of whole rat plasma.[22]

Cells have also been separated by exploiting their differences in polarizability.[141] A mixture of cultured human breast cancer cells and normal human peripheral blood cells was led over an interdigitated electrode array that actuated an upward dielectrophoretic force to the cells. The cells moved upward until the dielectrophoretic force, which decreases with increasing distance from the electrodes, was balanced by gravity. Because of their different polarizabilities, different cells experience different upward forces and will have different equilibrium heights with respect to the electrode array. Because the horizontal flow rate depends on the distance from the channel wall, the height difference is translated to a mobility difference and separations of different types of cells can be achieved.

In order to minimize the sticking of cells to channel walls, dielectrophoretic forces have also been used to manipulate cells without the use of channel walls. Funnels and also channels and switches based on dielectrophoretic forces have been developed.[142] Alternatively, a checkerboard pattern of electrodes that allowed the actuation of fields in different patterns was used for cell separation. In order to minimize contact between the cells and electrodes, a gel coating was applied to the electrodes in this experiment.[143] The same actuation method was combined with optical radiation pressure manipulators for the noncontact manipulation of *E. coli* cells.[144]

7 DETECTION

In drug- and DNA-related applications, detection is one of the least integrated parts of μTAS. Optical detection schemes such as LIF are compatible with the disposable chip format without integration. Therefore, at the moment LIF detection is by far the most popular detection system in this field. However, in applications where portability is an issue, integration of the detection on the chip is an important requirement. Therefore, alternative detectors have also been investigated. Moreover, as was indicated in section 2, as the trend of miniaturization proceeds, electrochemical sensors become more sensitive than optical detection schemes. Finally, a few examples of detection by MS are given to illustrate its capability to identify large biomolecules.

7.1 Optical Detection

Optical detection schemes, especially LIF, are dominant as far as chip-based systems are concerned. In the channel, a 50–200-μm spot is exited by a focused laser beam and emitted light is collected by a lens system and detected by a photomultiplier tube. Although numerous detection schemes are available,[145] most optical detection systems in μTAS applications simply measure changes in fluorescence intensity. An interesting exception is an integrated system developed by Evotec Biosystem (http://www.evotec.de) that allows the use of a number of different techniques. For example, fluorescence correlation spectroscopy (FCS) is employed to determine the diffusion coefficients of fluorescent molecules. Since the diffusion coefficient depends on the size of the molecule, molecular binding events and also the activity of restriction enzymes[146] and conformational changes of molecules[147] can be detected. Other optical detection schemes that have been used for on-chip detection include UV absorption[125] and Raman spectroscopy,[148] which is interesting because of the compound identification capabilities it shares with MS.

As an alternative to the laser and photomultiplier set-up, less expensive light-emitting diode (LED) and photodiode systems have been used to validate theoretical dispersion models.[12,13] Similar systems have been integrated in a system for the detection of orthophosphate, where fiber optics were used to direct the light,[149] and in the MCB demonstrator described in section 8.2.[61]

When optical detection is performed perpendicularly to the channel, the fluorescent signal decreases quickly with decreasing channel diameter and scattering at the channel walls decreases the signal-to-noise ratio. These problems have been reduced by fabrication of specially designed detector cells. In its simplest form this can be a loop in the channel to which optical fibers are fitted.[150] The sensitivity of the system was reported to be 10 times better than that of perpendicular detection because of the increased path length of the light in the channel and the decrease of scattering phenomena. Alternatively, the optical path length is increased when the light beam is introduced at an angle, so that reflections at the channel walls allow it to travel a number of times back and forth before detection. A silicon detection cell employing this method was constructed, but the losses occurring at each reflection led to disappointing efficiencies.[151] Problems of this kind can be avoided by integrated optics, a good example of which is the channel wall integrated

evanescent wave detector.[152] In this detector, a large number of reflections occurs and the evanescent part of the light wave is influenced by molecules adsorbed at the channel wall at each reflection. The sensor signal is expected to be independent of the channel diameter for channels wider than approximately 2 μm.

For the observation of flow patterns in channels and at junctions, charge-coupled device (CCD) cameras are used, because these systems are especially suited for the simultaneous optical detection at different positions in a system. This advantage is exploited in an interesting way in a device called the T-cell. These structures are deliberately designed so that mixing across the channel diameter is incomplete. This allows the measurement of background and calibration signals at the same time as the sample signal detection.[153]

7.2 Electrochemical Detection

Whereas optical techniques are commonly used for the off-chip detection of on-chip separations, electrochemical methods are more promising for integrated detection. Using standard lithographic and integrated circuit compatible methods for the fabrication of microelectrodes[154] and field effect transistor (FET) structures,[155] a wealth of inexpensive continuous detection schemes are available. Especially amperometric detection schemes have been developed for combination with LC and CE[156] because of the potential of their superior sensitivity. Methods for the wafer-scale fabrication of the membranes that are needed for the fabrication of ISFETs[157] and enzyme electrodes[158] are available. The range of compounds that can be detected is broadened even further when electrochemically active reagents are added. For example, DNA fragments separated by CE were detected using on-chip amperometry by addition of electrochemically active intercalation agents.[159] Moreover, an amperometric electrode array was described that permits time-resolved detection from continuous electrophoretic separations.[160] Although such a set-up might also be realized using a CCD camera, the strength of electrochemistry for parallel data acquisition is demonstrated by this method.

When electrochemical detection schemes are to be combined with CE, special measures are needed to minimize the interference of the detection module by the electrophoretic high-voltage part of the system. This was achieved by placing the detector electrodes directly behind the separation electrode.[159] In order to decrease the electric field in the detection region, the channel was widened at the detector position.

Electrochemical detection has proven extremely useful in portable applications for environmental[161] and point of care (POC) clinical monitoring. A μTAS for bedside monitoring of patients has been developed that offers a good example of the possibilities of electrochemical detection.[162] It employs potentiometric detection to measure sodium, potassium, chloride, ionized calcium, pH and pCO_2, while urea is detected by potentiometry after an enzymatic reaction, glucose and pO_2 are determined by amperometry and hematocrit is detected conductometrically. This μTAS will be described in more detail in section 8.3. In another system for continuous monitoring of patients, pH and pCO_2 detectors based on ISFETs were used in combination with an amperometric O_2 sensor.[163] Alternative possibilities demonstrated for on-chip electrochemical detection include stripping voltammetry,[164] coulometric titration[25] and redox-sensitive FETs.[165] The last devices are especially interesting because they contact the sample in only one spot,[166] minimizing noise and eliminating the need for a reference electrode.

7.3 Mass Spectrometry

In recent years, MS of biomolecules has received a lot of attention[167] because of its enormous fingerprinting capabilities. In electrospray ionization mass spectrometry (ESIMS), the most commonly miniaturized format, the analytes that are leaving the on-chip separation channel are directly ionized and accelerated towards the mass spectrometer.[168] When the detection and identification power of MS is combined with the separation efficiency of electrophoresis, a system can be built that allows the selective determination and identification of all compounds in a complex biological sample. Direct coupling of the mass spectrometer to a CE-based μTAS has been reported by several researchers.[169,170] Absolute detection limits as low as 1 fmol and concentrations at the 1 nM level combined with identification of the compounds have been reported.[171] Multichannel chips were used to analyze sequences of up to five samples from one chip.[170]

8 APPLICATIONS

In this section, four examples of a μTAS are given that are to perform completely different analyses in completely different environments. As a result, the functionality of the devices is also very different, thus giving a good overview of the scope of μTAS. In Table 3, the similarities and differences are summarized.

8.1 Continuous Blood Monitoring

The continuous monitoring of basic parameters such as pH, pO_2, pCO_2, potassium and sodium in blood was

Table 3 Requirements in different fields of μTAS

Field	Size	Robustness	Maintenance	Integration
Continuous blood monitoring	Portable by patient	High	Hours/days	Zero/low
Environmental and process monitoring	Portable	High	Weeks/months	Intermediate
Bedside blood monitoring	Portable	Intermediate	N/A	Intermediate
On-chip DNA characterization	Laboratory	Low	Minutes	High

one of the first challenges that was met by μTAS. The requirements of such systems are severe: since the system has to be carried on the patient, miniaturization of the complete system is needed. Moreover, frequent calibrations are needed to correct for signal drift due to the fouling of the sensor surface that is common with biological fluids.

A flow-through cell containing a pH ISFET and reference electrode, four miniaturized Clarke-type amperometric dissolved oxygen sensors and two ISFET-based Severinghaus-type potentiometric CO_2 sensors was fabricated at the University of Neuchâtel.[163] All sensors were made on a single 22×6 mm silicon chip that served as the bottom of the flow cell. Al_2O_3 was used as a pH-sensitive gate material on the ISFETs, that showed a sensitivity of 51 mV pH^{-1}. It would be inappropriate here to describe the fabrication of the ISFET sensors, and the interested reader is referred to an extensive description by Bergveld and Sibbald.[172] The Ag/AgCl reference electrodes were fabricated by galvanostatic growth of a 20-μm-thick Ag layer on a thin-film Pt electrode and subsequent chlorination in 0.1 M KCl. The Ag layer was made this thick in order to improve the lifetime of the electrode. The oxygen sensor had an amperometric three-electrode set-up consisting of a Pt microelectrode array, Pt counter electrodes and a reference electrode that was made as described above. A sensitivity of 0.36 nA $Torr^{-1}$ (1 Torr = 133.3 Pa) was achieved.

The CO_2 sensor used an Al_2O_3 ISFET to measure the pH change in a pocket of internal electrolyte that is induced by the dissociation of CO_2 into carbonic acid. A sensitivity of -39 mV $(pCO_2)^{-1}$ was reached. In order to provide the internal electrolyte where needed, all sensors except the pH ISFET were covered with a UV-curable polyacrylamide hydrogel and a polysiloxane gas-permeable membrane to provide an inner electrolyte. A 600-μm-high polysiloxane ridge and glass cover completed the flow cell.

Although the resulting structure had a functional lifetime of more than 5 weeks, the calibration problem was not solved. A calibration concept was demonstrated in a flow cell containing potassium-sensitive chemically sensitive field effect transistors (CHEMFETs).[83] Sample was pumped from a stream into a channel by a piezoelectrically actuated silicon micropump. The CHEMFETs were placed between the channel entrance and the pumps, and care was taken not to pump the sample all the way to the pumps. In this way, the pumps were only in contact with the clean internal solution and pump blocking by particles and dirt was avoided. A second pump was used to pump a calibration solution to the detector. The dead volume of this modular system was still rather high at 5 μl, leading to lag times of several seconds. This concept is developed further in the next subsection.

8.2 Environmental Monitoring

For environmental monitoring, the constraints on the size of analysis systems may be less severe than for continuous blood monitoring. At the same time, the level of system integration tends to be higher, because the analyses often have to be carried out in remote and difficultly accessible places, so that units must be able to run untended for weeks or even months. Because response times tend to be of less importance, remote environmental monitoring is extremely fit for a μTAS based on the modular approach.[173] The MCB concept, adapted from the PCB used in the electronics industry, is a good example of this approach, developed at MESA.[61] An MCB combines electric wiring connecting the electronics needed for data acquisition from the sensors and pumps drivers with fluid circuits connecting the different modules present on the board.

As a demonstration, the system shown in Figure 14(a) and (b) was developed. It consists of three inlets/outlets, two micropumps, two flow sensors and an optical absorption detector module. It can be used to measure chemical reaction products by detection of the (spectral) absorption intensity. Sample and reagent liquids are mixed in the appropriate amounts on-board (the actual mixing takes place during the propagation in channels) and the optical intensity from different colored LEDs is measured by a 64-pixel CCD detector.

The electronic control circuitry is situated in two levels below the fluidic circuitry layer with the modules. It is based on a microcontroller system for the microliquid handling and the chemical analysis data. Implemented in the electrical circuitry are driving circuits for the micropumps, sensing circuits for the flow sensors, optical absorption measurement circuitry, power management and communications using an RS232 interface.

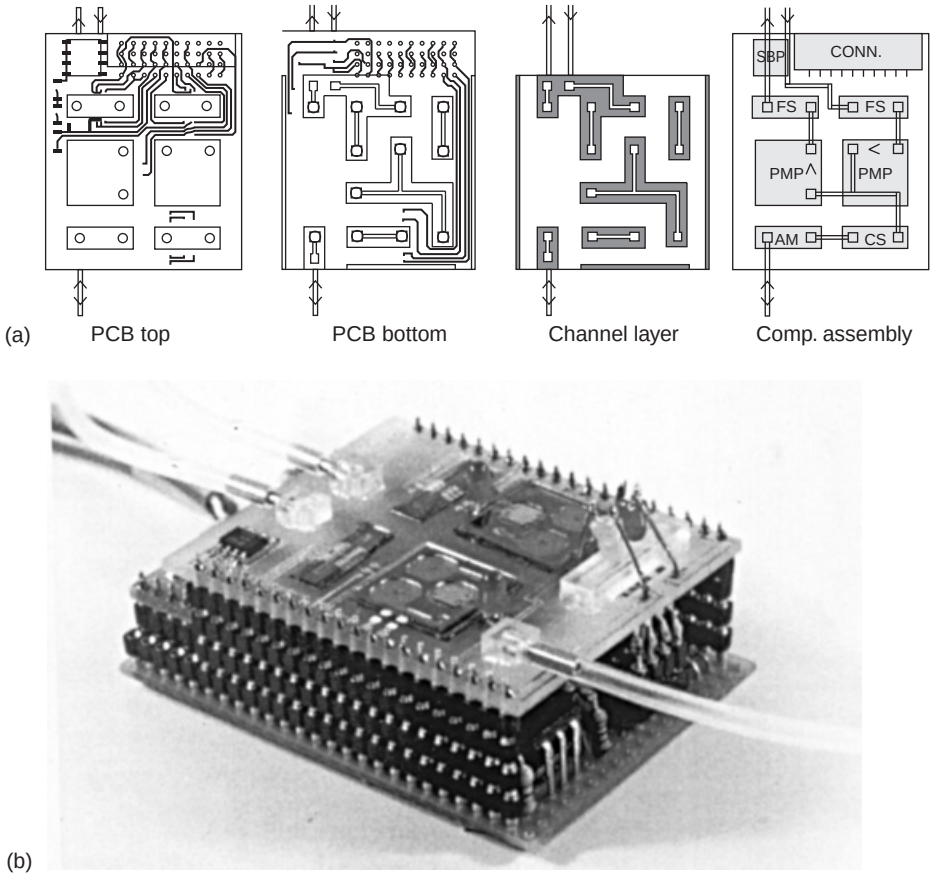

Figure 14 (a) MCB layout and (b) actual device.

A system for automatic groundwater analysis that is to be part of a monitoring network using radio communication has been realized.[174] The system uses ion-selective electrodes to measure nitrate and potassium while a microspectrometer combined with photometric reagents is used to determine iron levels. Microfluidics, including a mixer, were realized by anisotropic etching of p{100} Si in KOH and anodic bonding to Pyrex®. The whole system is mounted on a 75 × 500 mm board, which is rather large. However, it should be noted that everything needed, including tanks filled with calibration fluid, are included on the board.

8.3 Intermittent Bedside Blood Monitoring

Although continuous blood monitoring is still in the research phase, at least one intermittent real-time bedside blood monitoring system is commercially available. Not only is this system an improvement on traditional analyses by a central (hospital) laboratory because the test can be performed before the hospital is reached and results are available within 1 min,[175] but also the immediate analysis also guarantees that a fresh blood sample is used.

A complete blood analysis system was developed by I-STAT (http://www.i-stat.com).[176] The central part of the system is a disposable card consisting of two molded polymer structures sealed by a double-sided adhesive sheet. In this card, a droplet of whole blood is transported by capillary pumping into a channel until a capillary valve described in section 5.2 is met. When a reproducible amount of sample is taken in this way, the channel is sealed off and the card is placed in a hand-held analyzer. A pouch containing calibration fluid is pinched by a barb that is part of the card, and the fluid is led past the sensors. When the sensors are calibrated, the sample is led past the same sensors by operation of an air bladder. Sodium, potassium, chloride, ionized calcium, pH, pCO_2, pO_2, urea, glucose and hematocrit are all detected by a variety of electrochemical detection schemes, as described in section 7.2.

Although the card is at the center of the system, the latter would not function without the development of complementary technology. The analyzer is made to store patients' records and to acquire data. In order to allow electrochemical detection within 1 min after the sensors are wetted for the first time, a method to perform the

analysis before the electrodes have reached equilibrium was developed.[177]

8.4 On-chip DNA Characterization

In on-chip DNA analysis, portability is not an issue and miniaturization efforts are directed towards fast, high-resolution separations and integration. At present integration of detection modules or reactant reservoirs is not believed to lead to significantly better system performance. In contrast, the minimization of dead volumes and maximization of separation speed by miniaturization is given priority. In order to maximize the number of separations even further, chips having a large number of parallel channels are being developed.[178] However, as miniaturization proceeds and the amounts of substances to be detected become smaller, the integration of smaller detection units in DNA characterization chips will be needed to obtain acceptable signal-to-noise ratios.[139]

The usefulness of on-chip electrophoresis as a separation method for DNA characterization first became apparent when the separation of amino acids in open channels on glass chips was demonstrated.[129] Subsequently, channels on thermally bonded glass chips were filled with a sieving matrix and ultrafast (2 min) DNA fragment separations were demonstrated.[179] Although many variations on this theme have been performed since then, the set-up has remained the same in most cases: only the glass chip is miniaturized and detection and high-voltage modules have conventional sizes.[123–131] Later developments include various on-chip mixing schemes, including pre- and postseparation mixing of fluorescent labels,[32,57,102] mixing of the restriction enzyme with complete DNA molecules[180] and mixing of cells and lysing agents.[181] Lately, good results have been obtained with cheaper polymer chips.[38,39]

Moreover, the advantages of fast heat diffusion from miniaturized devices have been demonstrated by on-chip PCR amplification. Silicon reaction chambers with integrated heaters have been reported by several groups.[20] The high surface to volume ratio of the chambers allows the application of fast heating and cooling cycles of the order of 1 min.

In an alternative set-up, on-chip PCR amplification was performed by pumping the sample through a channel meandering across three thermostated temperature zones on a glass chip.[182] The number of meanders traveled by the sample plays the same role as the number of temperature cycles in conventional PCR. Because of the small channel diameter, heat dissipation was very fast and both heating and cooling times were below 100 ms. Because the first part of the meander is never in contact with the amplified, more concentrated DNA, cross-contamination seems less likely when this format is used compared with batchwise PCR reactions.

Although most research is still done on system components, (partly) integrated systems are beginning to appear in the literature[138,180] and are being brought on to the market by Cepheid (http://www.cepheid.com). A system reaching a high level of integration was recently presented by Mastrangelo's group. In their system, integration of components for sample loading, drop metering, PCR amplification, electrophoresis and photodetection into a microfluidic circuit was demonstrated.[95,183] The device was made in silicon and glass and employed an external pressure source for fluid motion, and hydrophobic patches were used to direct the fluids.[80,184]

ABBREVIATIONS AND ACRONYMS

ABS	Acrylonitrile–Butadiene–Styrene
BOE	Buffered Oxide Etch
CCD	Charge-coupled Device
CE	Capillary Electrophoresis
CHEMFET	Chemically Sensitive Field Effect Transistor
CNC	Computer Numerical Control
COMOSS	Collocated Monolith Support Structures
DRIE	Deep Reactive Ion Etching
EOF	Electroosmotic Flow
ESIMS	Electrospray Ionization Mass Spectrometry
FCS	Fluorescence Correlation Spectroscopy
FET	Field Effect Transistor
GC	Gas Chromatography
GOD	Glucose Oxidase
HPLC	High-performance Liquid Chromatography
IEF	Isoelectric Focusing
ISFET	Ion-sensitive Field Effect Transistor
ITP	Isotachophoresis
LC	Liquid Chromatography
LED	Light-emitting Diode
LIF	Laser-induced Fluorescence
LIGA	Lithographie Galvanoformung Abformung
LPCVD	Low-pressure Chemical Vapor Deposition
MCB	Mixed Circuit Board
MEKC	Micellar Electrokinetic Chromatography
MEMS	Microelectromechanical System
METF	Microelectrothermofluidic
MS	Mass Spectrometry
MST	Microsystems Technology

PC	Polycarbonate
PCB	Printed Circuit Board
PCR	Polymerase Chain Reaction
PDMS	Poly(dimethylsiloxane)
PE	Polyethylene
PET	Poly(ethylene terephthalate)
PMMA	Poly(methyl methacrylate)
POC	Point of Care
PU	Polyurethane
RIE	Reactive Ion Etching
μTAS	Micro Total Analytical System
TAS	Total Analytical System
UV	Ultraviolet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application

REFERENCES

1. R.A. Mathies, P.C. Simpson, A.T. Woolley, 'DNA Analysis with Capillary Array Electrophoresis Microplates', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 1–6, 1998.
2. P. Bergveld, 'The Significance of Feedback Control for Chemical Sensors', *Sens. Actuators B*, **8**, 47–52 (1992).
3. B.H. van der Schoot, P. Bergveld, 'Evaluation of the Sensor Properties of the pH-static Enzyme Sensor', *Anal. Chim. Acta*, **233**, 49–57 (1990).
4. H.M. Widmer, 'A Survey of the Trends in Analytical Chemistry Over the Last Twenty Years, Emphasizing the Development of TAS and μTAS', *Anal. Methods Instrum.*, Special Issue μTAS '96, 3–8 (1996).
5. S.C. Terry, J.H. Herman, J.B. Angell, 'A Gas Chromatograph Air Analyzer Fabricated on a Silicon Wafer', *IEEE Trans. Electron. Devices*, **26**, 1880–1886 (1979).
6. A. Manz, N. Graber, H.M. Widmer, 'Miniaturized Total Chemical Analysis Systems: a Novel Concept for Chemical Sensing', *Sens. Actuators B*, **1**, 244–248 (1990).
7. S.J. Haswell, 'Development and Operating Characteristics of Micro Flow Injection Analysis Systems Based on Electroosmotic Flow – a Review', *Analyst*, **122**, 1R–10R (1997).
8. D. Bökenkamp, A. Desai, X. Yang, Y-C. Tai, E.M. Marzluff, S.L. Mayo, 'Microfabricated Silicon Mixers for Submillisecond Quench-flow Analysis', *Anal. Chem.*, **70**(2), 232–236 (1998).
9. R.F. Probstein, *Physicochemical Hydrodynamics: an Introduction*, 2nd edition, Wiley, New York, 1994.
10. S.C. Jacobson, J.P. Kutter, C.T. Culbertson, J.M. Ramsey, 'Rapid Electrophoretic and Chromatographic Analysis on Microchips', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 315–318, 1998.
11. S.D. Senturia, 'CAD Challenges for Microsensors, Microactuators, and Microsystems', *Proc. IEEE*, **86**(8), 1611–1626 (1998).
12. E.B. van Akker, M. Bos, W.E. van der Linden, 'Convection and Diffusion in a Micro-flow Injection System', *Anal. Chim. Acta*, **373**, 227–239 (1998).
13. E.B. van Akker, M. Bos, W.E. van der Linden, 'Continuous, Pulsed and Stopped Flow in a μ-Flow Injection System', *Anal. Chim. Acta*, **378**, 111–117 (1999).
14. O. Français, I. Dufour, 'Dynamic Simulation of an Electrostatic Micropump with Pull-in and Hysteresis Phenomena', *Sens. Actuators A*, **70**, 56–60 (1998).
15. P. Voigt, G. Schrag, G. Wachutka, 'Electrofluidic Full-system Modeling of a Flap Valve Micropump Based on Kirchoffian Network Theory', *Sens. Actuators A*, **66**, 9–14 (1998).
16. S. Carmona, S. Marco, J. Samitier, J.R. Morante, 'Dynamic Simulations of Micropumps', *J. Micromech. Microeng.*, **6**, 128–130 (1996).
17. A. Manz, Y. Miyahara, J. Miura, Y. Watanabe, H. Miyagi, K. Sato, 'Design of an Open-tubular Column Liquid Chromatograph Using Silicon Chip Technology', *Sens. Actuators B*, **1**, 249–255 (1990).
18. Y. Fintschenko, A. van den Berg, 'Silicon Microtechnology and Microstructures in Separation Science', *J. Chromatogr. A*, **819**, 3–12 (1998).
19. J. Drott, K. Lindström, L. Rosengren, T. Laurell, 'Porous Silicon as the Carrier Matrix in Microstructured Enzyme Reactors Yielding High Enzyme Activities', *J. Micromech. Microeng.*, **7**, 14–23 (1997).
20. M.A. Northrup, B. Benett, D. Hadley, P. Landre, S. Lehew, J. Richards, P. Stratton, 'A Miniature Analytical Instrument for Nucleic Acids Based on Micromachined Silicon Reaction Chambers', *Anal. Chem.*, **70**(5), 918–922 (1998).
21. L. Martynova, L.E. Locascio, M. Gaitan, G.W. Kramer, R.G. Christensen, W.A. MacCrehan, 'Fabrication of Plastic Microfluid Channels by Imprinting Methods', *Anal. Chem.*, **69**(23), 4783–4789 (1997).
22. D.E. Raymond, A. Manz, H.M. Widmer, 'Continuous Separation of High Molecular Weight Compounds Using a Microliter Volume Free-flow Electrophoresis Microstructure', *Anal. Chem.*, **68**(15), 2515–2522 (1996).
23. D.J. Harrison, P.G. Glavina, 'Towards Miniaturized Electrophoresis and Chemical Analysis Systems on Silicon: an Alternative to Chemical Sensors', *Sens. Actuators B*, **10**, 107–116 (1993).
24. Y. Fintschenko, P. Fowler, V. Spiering, G.-J. Burger, A. van den Berg, 'Characterization of Silicon-based Insulated Channels for Capillary Electrophoresis', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 327–330, 1998.

25. O.T. Guenat, W.E. Morf, B.H. van der Schoot, N.F. de Rooij, 'Universal Coulometric Nanotitrators with Potentiometric Detection', *Anal. Chim. Acta*, **361**, 261–272 (1998).
26. M. Madou, *Fundamentals of Microfabrication*, CRC Press, Boca Raton, FL, 1997.
27. D.L. Kendal, R.A. Shoultz, 'Wet Chemical Etching of Silicon and SiO_2 and Ten Challenges for Micromachiners', in *Handbook of Microlithography, Micromachining and Microfabrication*, ed. P.R. Choudury, SPIE – The International Society for Optical Engineering, Bellingham, WA, Chapter 2, Vol. 2, 1997.
28. J.W. Berenschot, R.E. Oosterbroek, T.S.J. Lammerink, M.C. Elwenspoek, 'Micromachining of {111} Plates in ⟨001⟩ Oriented Silicon', *J. Micromech. Microeng.*, **8**, 104–106 (1998).
29. R.W. Tjerkstra, J.G.E. Gardeniers, M.C. Elwenspoek, A. van den Berg, 'Etching Technology for Chromatography Microchannels', *Electrochim. Acta*, **42**(20–22), 3399–3406 (1997).
30. Z.H. Fan, D.J. Harrison, 'Micromachining of Capillary Electrophoresis Injectors and Separators on Glass Chips and Evaluation of Flow at Capillary Intersections', *Anal. Chem.*, **66**(1), 177–184 (1994).
31. S.C. Jacobson, A.W. Moore, J.M. Ramsey, 'Fused Quartz Substrates for Microchip Electrophoresis', *Anal. Chem.*, **67**(13), 2059–2063 (1995).
32. S.C. Jacobson, L.B. Koutny, R. Hergenröder, A.W. Moore, J.M. Ramsey, 'Microchip Capillary Electrophoresis with an Integrated Postcolumn Reactor', *Anal. Chem.*, **66**(20), 3472–3476 (1994).
33. M. Grétillat, F. Paoletti, P. Thiébaud, S. Roth, M. Koudelka-Hep, N.F. de Rooij, 'A New Fabrication Method for Borosilicate Glass Capillary Tubes with Lateral Inlets and Outlets', *Sens. Actuators A*, **60**, 219–222 (1997).
34. H. Jansen, H. Gardeniers, M. de Boer, M. Elwenspoek, J. Fluitman, 'A Survey on the Reactive Ion Etching of Silicon in Microtechnology', *J. Micromech. Microeng.*, **6**, 14–28 (1996).
35. J.W. Bartha, J. Greschner, M. Puech, P. Maquin, 'Low Temperature Etching of Si in High Density Plasma Using SF_6/O_2', *J. Microelectron. Eng.*, **27**, 453–456 (1995).
36. H. Jansen, M. de Boer, R. Legtenberg, M. Elwenspoek, 'The Black Silicon Method: a Universal Method for Determining the Parameter Setting of a Fluorine-based Reactive Ion Etcher in Deep Silicon Trench Etching with Profile Control', *J. Micromech. Microeng.*, **5**, 115–120 (1995).
37. S.J. Baldock, N. Bektas, P.R. Fielden, N.J. Goddard, L.W. Pickering, J.E. Prest, R.D. Snook, B.J. Treves Brown, D.I. Vaireanu, 'Isotachophoresis on Planar Polymeric Substrates', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 359–362, 1998.
38. C.S. Effenhauser, G.J.M. Bruin, A. Paulus, M. Ehrat, 'Integrated Capillary Electrophoresis on Flexible Silicone Microdevices: Analysis of DNA Restriction Fragments and Detection of Single DNA Molecules on Microchips', *Anal. Chem.*, **69**(17), 3451–3457 (1997).
39. R.M. McCormick, R.J. Nelson, M.G. Alonso-Amigo, D.J. Benvegnu, H.H. Hooper, 'Microchannel Electrophoretic Separations of DNA in Injection-molded Plastic Substrates', *Anal. Chem.*, **69**(14), 2626–2630 (1997).
40. E. Kim, Y. Xia, X-M. Zhao, G.M. Whitesides, 'Solvent Assisted Microcontact Molding: a Convenient Method for Fabricating Three-dimensional Structures on Surfaces of Polymers', *Adv. Mater.*, **9**(8), 651–654 (1997).
41. Y. Xia, D. Qin, G.M. Whitesides, 'Microcontact Printing with a Cylindrical Rolling Stamp: a Practical Step Toward Automatic Manufacturing of Patterns with Submicrometer-sized Features', *Adv. Mater.*, **8**(12), 1015–1017 (1996).
42. S.Y. Chou, P.R. Krauss, P.J. Renstrom, 'Imprint Lithography with 25-nanometer Resolution', *Science*, **272**, 85–87 (1996).
43. J. Elders, H.V. Jansen, M. Elwenspoek, W. Ehrfeld, 'DEEMO: a New Technology for the Fabrication of Microstructures', in *Proceedings of 1995 MEMS Workshop*, eds. J. Elders, H.V. Jansen, M.C. Elwenspoek, MEMS, Amsterdam, 238–243, 1995.
44. H. Becker, W. Dietz, P. Dannberg, 'Microfluidic Manifolds by Polymer Hot Embossing for μTAS Applications', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 253–256, 1998.
45. D. Qin, Y. Xia, G.M. Whitesides, 'Rapid Prototyping of Complex Structures with Feature Sizes Larger than 20 μm', *Adv. Mater.*, **8**(11), 917–919 (1996).
46. Y. Xia, E. Kim, X.-M. Zhao, J.A. Rogers, M. Prentiss, G.M. Whitesides, 'Complex Optical Structures Formed by Replica Molding Against Elastomeric Masters', *Science*, **273**, 347–349 (1996).
47. Y. Xia, J.J. McClelland, R. Gupta, D. Qin, X.-M. Zhao, L.L. Sohn, R.J. Celotta, G.M. Whitesides, 'Replica Molding Using Polymeric Materials: A Practical Step Toward Nanomanufacturing', *Adv. Mater.*, **9**(2), 147–149 (1997).
48. W.K. Schomburg, J. Vollmer, B. Büstgens, J. Fahrenberg, H. Hein, W. Menz, 'Microfluidic Components in LIGA Technique', *J. Micromech. Microeng.*, **4**(4), 186–191 (1994).
49. L.T. Romankiw, 'A Path from Electroplating Through Lithographic Masks in Electronics to LIGA in MEMS', *Electrochim. Acta*, **42**, 2985–3006 (1997).
50. W. Menz, 'LIGA and Related Technologies for Industrial Application', *Sens. Actuators A*, **54**, 785–789 (1996).
51. K.Y. Lee, N. LaBianca, S.A. Rishton, S. Zolgharnain, J.D. Gelorme, J. Shaw, T.H.-P. Chang, 'Micromachining Application for a High Resolution Ultrathick Photoresist', *J. Vac. Sci. Technol. B*, **13**, 3012–3016 (1995).
52. H. Lorenz, M. Despont, N. Fahrni, N. LaBianca, P. Renaud, P. Vettiger, 'SU-8™: a Low-cost Negative

Resist for MEMS', *J. Micromech. Microeng.*, **7**, 121–124 (1997).
53. L.J. Guérin, M. Bossel, M. Demierre, S. Calmes, P. Renaud, 'Simple and Low Cost Fabrication of Embedded Micro-channels by Using a New Thick-film Photoplastic', *Tech. Dig. Transducers*, **2**, 1419–1422 (1997).
54. M.O. Heuschkel, L. Guérin, B. Buisson, D. Bertrand, P. Renaud, 'Buried Microchannels in Photopolymer for Delivering of Solutions to Neurons in a Network', *Sens. Actuators A*, **48**, 356–361 (1998).
55. W.H. Ko, J.T. Suminto, G.J. Yeh, 'Bonding Techniques for Microsensors', in *Micromachining and Micropackaging for Transducers*, eds. C.D. Fung, P.W. Cheung, W.H. Ko, O.G. Fleming, Elsevier, Amsterdam, 41–61 1985.
56. D.J. Harrison, K. Fluri, K. Seiler, Z. Fan, C.S. Effenhauser, A. Manz, 'Micromachining a Miniaturized Capillary Electrophoresis-based Chemical Analysis System on a Chip', *Science*, **261**, 895–896 (1993).
57. K. Fluri, G. Fitzpatrick, N. Chiem, D.J. Harrison, 'Integrated Capillary Electrophoresis Devices with an Efficient Postcolumn Reaction in Planar Quartz and Glass Chips', *Anal. Chem.*, **68**, 4285–4290 (1996).
58. M. Stjernström, J. Roeraade, 'Methods for Fabrication of Microfluidic Systems in Glass', *J. Micromech. Microeng.*, **8**, 33–38 (1998).
59. M.A. Roberts, J.S. Rossier, P. Bercier, H. Girault, 'UV Laser Machined Polymer Substrates for the Development of Microdiagnostic Systems', *Anal. Chem.*, **69**(11), 2035–2042 (1997).
60. R. Hillman, M.E. Cobb, J.D. Allen, I. Gibbons, V.E. Ostoich, L.J. Winfrey, 'Capillary Flow Device', *US Pat.*, 4 963 498 (1990).
61. A. van den Berg, T.S.J. Lammerink, 'Micro Total Analysis Systems: Microfluidic Aspects, Integration Concept and Applications', *Top. Curr. Chem.*, **194**, 21–49 (1998).
62. M. Elwenspoek, T.S.J. Lammerink, R. Miyake, J.H.J. Fluitman, 'Towards Integrated Microliquid Handling Systems', *J. Micromech. Microeng.*, **4**, 227–245 (1994).
63. P. Gravesen, J. Branjeberg, O.S. Jensen, 'Microfluidics – a Review', *J. Micromech. Microeng.*, **3**, 168–182 (1993).
64. H. Gau, S. Herminghaus, P. Lenz, R. Lipowsky, 'Liquid Morphologies on Structured Surfaces: from Microchannels to Microchips', *Science*, **283**, 46–49 (1999).
65. V.L. Spiering, J.N. van der Moolen, G.-J. Burger, A. van den Berg, 'Novel Microstructures and Technologies Applied in Chemical Analysis Techniques', *Tech. Dig. Transducers*, **1**, 511–514 (1997).
66. N.J. Mourlas, D. Jaeggi, A.F. Flannery, B.L. Gray, B.P. van Drieënhuizen, C.W. Storment, N.I. Malouf, G.T.A. Kovacs, 'Novel Interconnection and Channel Technologies for Microfluidics', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 27–30, 1998.
67. S. Böhm, W. Olthuis, P. Bergveld, 'A μTAS Based on Microdialysis for On-line Monitoring of Clinically Relevant Substances', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 31–34, 1998.
68. R.E. Oosterbroek, J.W. Berenschot, S. Schlautman, T.S.J. Lammerink, A. van den Berg, M.C. Elwenspoek, 'Utilizing the {111} Plane Switch Over Etching Process for Micro Fluid Control Applications', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 137–140, 1998.
69. W.K. Schomburg, J. Vollmer, B. Büstgens, J. Fahrenberg, H. Hein, W. Menz, 'Active Valves and Pumps for Microfluidics', *J. Micromech. Microeng.*, **3**, 216–218 (1994).
70. S. Böhm, W. Olthuis, P. Bergveld, 'A Plastic Micropump Constructed with Conventional Techniques and Materials', *Sens. Actuators A*, **77**(3), 223–228 (1999).
71. W.K. Schomburg, R. Ahrens, W. Bacher, C. Goll, S. Meinzer, A. Quinte, 'AMANDA Low Cost Production of Microfluidic Devices', *Sens. Actuators A*, **70**, 153–158 (1998).
72. S. Shoji, B. van der Schoot, N.F. de Rooij, M. Esashi, 'Smallest Dead Volume Microvalves for Integrated Chemical Analyzing Systems', *Proc. Transducers 91*, 1052–1055 (1991).
73. J. Branjeberg, P. Gravesen, 'A New Electrostatic Actuator Providing Improved Stroke Length and Force', *Proc. MEMS 92*, 6–11 (1992).
74. R.L. Smith, R.W. Bower, S.D. Collins, 'The Design and Fabrication of a Magnetically Actuated Micromachined Flow Valve', *Sens. Actuators A*, **24**, 47–53 (1990).
75. C. Neagu, J.G.E. Gardeniers, M. Elwenspoek, J.J. Kelly, 'An Electrochemical Microactuator: Principle and First Results', *J. Microelectromech. Syst.*, **5**(1), 2–9 (1996).
76. E. Smela, O. Inganäs, I. Lundström, 'Controlled Folding of Micrometer-size Structures', *Science*, **268**, 1735–1738 (1995).
77. E. Smela, O. Inganäs, I. Lundström, 'New Devices Made from Combining Silicon Microfabrication and Conducting Polymers', in *Molecular Manufacturing*, ed. C. Nicolini, Plenum Press, New York, 189–214, 1996.
78. N. Kaartinen, 'Micro Electro Thermo Fluidic (METF) Liquid Microprocessor', *Proc. MEMS '96*, 395 (1996).
79. H. Hartshorne, Y. Ning, W.E. Lee, C. Backhouse, 'Development of Microfabricated Valves for μTAS', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 379–381, 1998.
80. P.F. Man, C.H. Mastrangelo, M.A. Burns, D.T. Burke, 'Microfabricated Capillarity-driven Stop Valve and Sample Injector', *Proc. MEMS 98*, 45–50 (1998).
81. R.C. Anderson, G.J. Bogdan, Z. Barniv, T.D. Dawes, J. Winkler, K. Roy, 'Microfluidic Biochemical Analysis System', *Tech. Dig. Transducers*, **2**, 477–480 (1997).
82. H.T.C. van Lintel, F.C.M. van de Pol, S. Bouwstra, 'A Piezoelectric Micropump Based on Micromachining of Silicon', *Sens. Actuators*, **15**, 153–167 (1988).

83. B.H. van der Schoot, S. Jeanneret, A. van den Berg, N.F. de Rooij, 'A Silicon Integrated Miniature Chemical Analysis System', *Sens. Actuators B*, **6**, 57–60 (1992).
84. V. Gass, B.H. van der Schoot, S. Jeanneret, N.F. de Rooij, 'Integrated Flow-regulated Silicon Micropump', *Sens. Actuators A*, **43**, 335–338 (1994).
85. S. Shoji, M. Esashi, B.H. van der Schoot, N.F. de Rooij, 'A Study of High-pressure Micropump for Integrated Chemical Analysis Systems', *Sens. Actuators A*, **32**, 335–339 (1992).
86. R. Zengerle, J. Uhlrich, S. Kluge, M. Richter, A. Richter, 'A Bidirectional Silicon Micropump', *Sens. Actuators A*, **50**, 81–86 (1995).
87. S. Böhm, W. Olthuis, P. Bergveld, 'A Plastic Micropump Constructed with Conventional Techniques and Materials', *Sens. Actuators A*, **77**(3), 223–228 (1999).
88. F.C.M. van de Pol, H.T.G. van Lintel, M. Elwenspoek, J.H.J. Fluitman, 'A Thermopneumatic Micropump Based on Micro-engineering Techniques', *Sens. Actuators A*, 21–23, 198–202 (1990).
89. I. Ederer, P. Raetsch, W. Schullerus, C. Tille, U. Zech, 'Piezoelectrically Driven Micropump for On-demand Fuel-drop Generation in an Automobile Heater with Continuously Adjustable Power Output', *Sens. Actuators A*, **62**, 752–755 (1997).
90. M. Koch, A.G.R. Evans, A. Brunnschweiler, 'The Dynamic Micropump Driven with a Screen Printed PZT Actuator', *J. Micromech. Microeng*, **8**(2), 119–122 (1998).
91. E. Stemme, G. Stemme, 'A Valveless Diffuser/Nozzle-based Fluid Pump', *Sens. Actuators A*, **39**, 159–167 (1993).
92. S. Matsumoto, A. Klein, R. Maeda, 'Micromachined Bidirectional Liquid Pump with Thermally Controlled Dynamic Valves', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 395–398, 1998.
93. P.H. Paul, D.W. Arnold, D.J. Rakestraw, 'Electrokinetic Generation of High Pressures Using Porous Microstructures', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 48–52, 1998.
94. E. Delamarche, A. Bernard, H. Schmid, B. Michel, H. Biebuyk, 'Patterned Delivery of Immunoglobins to Surfaces Using Microfluidic Networks', *Science*, **276**, 779–781 (1997).
95. M.A. Burns, C.H. Mastrangelo, T.S. Sammarco, F.P. Man, J.R. Webster, B.N. Johnson, B. Foerster, D. Jones, Y. Fields, A.R. Kaiser, D.T. Burke, 'Microfabricated Structures for Integrated DNA Analysis', *Proc. Natl. Acad. Sci. USA*, **93**, 5556–5561 (1996).
96. G. Fuhr, R. Hagedorn, T. Müller, W. Benecke, B. Wagner, 'Pumping of Water Solutions in Microfabricated Electrohydrodynamic Systems', *Proc. MEMS '92*, 25–30 (1992).
97. C.J.M. van Rijn, M.C. Elwenspoek, 'Microfiltration Membrane Sieve with Silicon Micro Machining for Industrial and Biomedical Applications', *Proc. MEMS '95*, 83–87 (1995).
98. G. Stemme, G. Kittilsland, 'New Fluid Structure in Silicon Fabricated Using a Self-aligning Technique', *Appl. Phys. Lett.*, **53**(16), 1566–1568 (1988).
99. J.P. Brody, T.D. Osborn, F.K. Forster, P. Yager, 'A Planar Microfabricated Fluid Filter', *Sens. Actuators A*, **54**, 704–708 (1996).
100. S. Kuiper, C.J.M. van Rijn, W. Nijdam, M.C. Elwenspoek, 'Development and Applications of Very High Flux Microfiltration Membranes', *J. Membr. Sci.*, **150**, 1–8 (1998).
101. J. Branjeberg, P. Gravesen, O. Jensen, 'Micromechanical Filter and Method for the Manufacture Thereof', *World Pat.*, WO 93/11862, 1993.
102. S.C. Jacobson, R. Hergenröder, A.W. Moore, J.M. Ramsey, 'Precolumn Reactions with Electrophoretic Analysis Integrated on a Microchip', *Anal. Chem.*, **66**(23), 4127–4132 (1994).
103. S.B. Cheng, C.D. Skinner, D.J. Harrison, 'Integrated Serial Dilution on a Microchip for Immunoassay Sample Treatment and Flow Injection Analysis', *Proc. µTAS 98*, 157–160 (1998).
104. W. Ehrfeld (ed.), *Microreaction Technology, Proceedings of 1st International Conference On Microreaction Technology*, Springer, Berlin, 1998.
105. T.T. Veenstra, T.S.J. Lammerink, M.C. Elwenspoek, A. van den Berg, 'Characterization Method for a New Diffusion-mixer Applicable in Micro Flow Injection Analysis Systems', *J. Micromech. Microeng.*, **9**(2), 199–202 (1999).
106. R. Miyake, T.S.J. Lammerink, M.C. Elwenspoek, J.H.J. Fluitman, 'Micromixer with Fast Diffusion', *Proc. MEMS '93*, 248–253, 1993.
107. T. Fuji, K. Hosokawa, S. Shoji, A. Yotsumoto, T. Nojima, I. Endo, 'Development of a Microfabricated Biochemical Workbench – Improving the Mixing Efficiency', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 173–176, 1998.
108. A. Yotsumoto, R. Nakamura, S. Shoji, T. Wada, 'Fabrication of an Integrated Mixing/Reaction Micro Flow Cell for µTAS', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 185–188, 1998.
109. L. Licklider, W.G. Kuhr, 'Characterization of Reaction Dynamics in a Trypsin-modified Capillary Microreactor', *Anal. Chem.*, **70**(9), 1902–1908 (1998).
110. T.J. Mason (ed.), *Sonochemistry, the Uses of Ultrasound in Chemistry*, Royal Society of Chemistry, Cambridge, 1989.
111. T.G. Leighton, *The Acoustic Bubble*, Academic Press, New York, 1994.
112. T. Laurell, J. Drott, 'Silicon Wafer Integrated Enzyme Reactors', *Biosens. Bioelectron.*, **10**, 289–299 (1995).
113. P. Önnerfjord, J. Nilsson, L. Wallman, T. Laurell, G. Marko-Varga, 'Picoliter Sample Preparation in

MALDI/TOF/MS Using a Micromachined Silicon Flow-through Dispenser', *Anal. Chem.*, **70**(22), 4755–4760 (1998).
114. T.E. Robinson, J.B. Justice, *Microdialysis in Neurosciences*, Elsevier, Amsterdam, 1991.
115. R. Freany, A. McShane, T.V. Keaveny, M. McKenna, K. Rabenstein, F.W. Scheller, D. Pfeiffer, G. Urban, I. Moser, G. Jobst, A. Manz, E. Verpoorte, M.W. Widmer, D. Diamond, E. Dempsey, F.J. Saez de Viteri, M. Smyth, 'Novel Instrumentation for Real-time Monitoring Using Miniaturized Flow Systems with Integrated Biosensors', *Ann. Clin. Biochem.*, **34**, 291–302 (1997).
116. R. Steinkuhl, C. Sundermeier, H. Hinkers, C. Dumschat, K. Cammann, M. Knoll, 'Microdialysis System For Continuous Glucose Monitoring', *Sens. Actuators B*, **33**, 19–24 (1996).
117. C.L. Liu, Q. Wu, A.C. Harms, R.D. Smith, 'On-line Microdialysis Sample Cleanup for Electrospray Ionization Mass Spectrometry of Nucleic Acid Samples', *Anal. Chem.*, **68**(18), 3295–3299 (1996).
118. N. Xu, Y. Lin, S.A. Hofstadter, D. Matson, C.J. Call, R.D. Smith, 'A Microfabricated Dialysis Device for Sample Cleanup in Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **70**(17), 3553–3556 (1998).
119. J.P. Brody, P. Yager, 'Diffusion-based Extraction in a Microfabricated Device', *Sens. Actuators A*, **58**, 13–18 (1997).
120. W.D. Volkmuth, T. Duke, R.H. Austen, E.C. Cox, 'Trapping of Branched DNA in Microfabricated Structures', *Proc. Natl. Acad. Sci. USA*, **92**, 6887–6891 (1995).
121. A. Marziali, T.D. Willis, R.W. Davis, 'An Arrayable Flow-through Microcentrifuge for High-throughput Instrumentation', *Proc. Natl. Acad. Sci. USA*, **96**, 61–66 (1999).
122. S. Hannoe, I. Sugimoto, T. Katoh, 'Silicon-micromachined Separation Columns Coated with Amino Acid Films for an Integrated On-chip Gas Chromatograph', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 145–148, 1998.
123. S.C. Jacobson, R. Hergenröder, L.B. Koutny, J.M. Ramsey, 'Open Channel Electrochromatography on a Microchip', *Anal. Chem.*, **66**(14), 2369–2373 (1994).
124. B. He, N. Tait, F. Regnier, 'Fabrication of Nanocolumns for Liquid Chromatography', *Anal. Chem.*, **70**(18), 3790–3797 (1998).
125. S.C. Jacobson, J.M. Ramsey, 'Microfabricated Devices for Performing Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 827–840, 1997.
126. C.S. Effenhauser, A. Manz, H.M. Widmer, 'Manipulation of Sample Fractions on a Capillary Electrophoresis Chip', *Anal. Chem.*, **67**(13), 2284–2287 (1995).
127. S.C. Jacobson, R. Hergenröder, L.B. Koutny, J.M. Ramsey, 'High Speed Separations on a Microchip', *Anal. Chem.*, **66**(7), 1114–1118 (1994).
128. D.J. Harrison, A. Manz, Z. Fan, H. Lüdi, H.M. Widmer, 'Capillary Electrophoresis and Sample Injection Systems Integrated on a Planar Glass Chip', *Anal. Chem.*, **64**(17), 1926–1932 (1992).
129. K. Seiler, D.J. Harrison, A. Manz, 'Planar Glass Chips for Capillary Electrophoresis: Repetitive Sample Injection, Quantitation, and Separation Efficiency', *Anal. Chem.*, **65**(10), 1481–1488 (1993).
130. C.S. Effenhauser, A. Manz, H.M. Widmer, 'Glass Chips for High-speed Capillary Electrophoresis Separations with Submicrometer Plate Heights', *Anal. Chem.*, **65**(19), 2637–2642 (1993).
131. S.C. Jacobson, R. Hergenröder, L.B. Koutny, R.J. Warmack, J.M. Ramsey, 'Effects of Injection Schemes and Column Geometry on the Performance of Microchip Electrophoresis Devices', *Anal. Chem.*, **66**(7), 1107–1113 (1994).
132. P.A. Walker, III, M.D. Morris, M.A. Burns, B.N. Johnson, 'Isotachophoretic Separations on a Microchip. Normal Raman Spectroscopy Detection', *Anal. Chem.*, **70**(18), 3766–3769 (1998).
133. P.R. Fielden, S.J. Baldock, N.J. Goddard, L.W. Pickering, J.E. Prest, R.D. Snook, B.J. Treves Brown, D.I. Vaireanu, 'A Miniaturized Planar Isotachophoresis Separation Device for Transition Metals with Integrated Conductivity Detection', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 323–326, 1998.
134. S.J. Baldock, N. Bektas, P.R. Fielden, N.J. Goddard, L.W. Pickering, J.E. Prest, R.D. Snook, B.J. Treves Brown, D.I. Vaireanu, 'Isotachophoresis on Planar Polymeric Substrates', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 359–362, 1998.
135. O. Hofman, D. Che, K.A. Cruickshank, U.R. Müller, 'Adaptation of Capillary Isoelectric Focusing to Microchannels on a Glass Chip', *Anal. Chem.*, **71**(3), 678–686 (1999).
136. C.S. Effenhauser, A. Paulus, A. Manz, H.M. Widmer, 'High-speed Separation of Antisense Oligonucleotides on a Micromachined Capillary Electrophoresis Device', *Anal. Chem.*, **66**(18), 2949–2953 (1994).
137. F. von Heeren, E. Verpoorte, A. Manz, W. Thormann, 'Micellar Electrokinetic Chromatography Separations and Analysis of Biological Samples on a Cyclic Planar Microstructure', *Anal. Chem.*, **68**(13), 2044–2053 (1996).
138. P.F. Gavin, A.G. Ewing, 'Continuous Separations with Microfabricated Electrophoresis–Electrochemical Array Detection', *J. Am. Chem. Soc.*, **118**, 8932–8936 (1996).
139. C.H. Mastrangelo, M.A. Burns, D.T. Burke, 'Microfabricated Devices for Genetic Diagnostics', *Proc. SPIE*, **86**(8), 1769–1786 (1998).
140. P.C.H. Li, D.J. Harrison, 'Transport, Manipulation, and Reaction of Biological Cells On-chip Using Electrokinetic Effects', *Anal. Chem.*, **69**(8), 1564–1568 (1997).

141. J. Yang, Y. Huang, X.-B. Wang, F.F. Becker, P.R.C. Gascoyne, 'Cell Separation on Microfabricated Electrodes Using Dielectrophoretic/Gravitational Field-flow Fractionation', *Anal. Chem.*, **71**(5), 911–918 (1999).
142. S. Fiedler, S.G. Shirley, T. Schnelle, G. Fuhr, 'Dielectrophoretic Sorting of Particles and Cells in a Microsystem', *Anal. Chem.*, **70**(9), 1909–1915 (1998).
143. J. Cheng, E.L. Sheldon, L. Wu, M.J. Heller, J.P. O'Connell, 'Isolation of Cultured Cervical Carcinoma Cells Mixed with Peripheral Blood Cells on a Bioelectronic Chip', *Anal. Chem.*, **70**(11), 2321–2326 (1998).
144. K. Morishima, F. Arai, T. Fukuda, H. Matsuura, K. Yoshikawa, 'Screening of Single *Escherichia coli* in a Microchannel System by Electric Field and Laser Tweezers', *Anal. Chim. Acta*, **365**, 273–278 (1998).
145. J.R. Lakowicz (ed.), *Topics in Fluorescence Spectroscopy*, Plenum Press, New York, 1991.
146. A. Koltermann, U. Kettling, J. Bieschke, T. Winkler, M. Eigen, 'Rapid Assay Processing by Integration of Dual-color Fluorescence Cross-correlation Spectroscopy: High Throughput Screening for Enzyme Activity', *Proc. Natl. Acad. Sci. USA*, **95**, 1421–1426 (1998).
147. C. Eggeling, J.R. Fries, L. Brand, R. Günther, C.A.M. Seidel, 'Monitoring Conformational Dynamics of a Single Molecule by Selective Fluorescence Spectroscopy', *Proc. Natl. Acad. Sci. USA*, **95**, 1556–1561 (1998).
148. K.A. Reshni, M.D. Morris, B.N. Johnson, M.A. Burns, 'On-line Detection of Electrophoretic Separations on a Microchip by Raman Spectroscopy', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 109–112, 1998.
149. G.N. Doku, S.J. Haswell, 'Further Studies into the Development of a Micro-FIA (FIA) System Based on Electroosmotic Flow for the Determination of Phosphate as Orthophosphate', *Anal. Chim. Acta*, **382**, 1–13 (1999).
150. Z. Liang, N. Chiem, G. Ocvirk, T. Tang, K. Fluri, D.J. Harrison, 'Microfabrication of a Planar Absorbance and Fluorescence Cell for Integrated Capillary Electrophoresis Devices', *Anal. Chem.*, **68**(6), 1040–1046 (1996).
151. E. Verpoorte, A. Manz, H. Lüdi, A.E. Bruno, F. Maystre, B. Krattinger, H.M. Widmer, B.H. van der Schoot, N.F. de Rooij, 'A Silicon Flow Cell for Optical Detection in Miniaturized Total Chemical Analysis Systems', *Sens. Actuators B*, **6**, 66–70 (1992).
152. G. Pandraud, T.M. Koster, C. Gui, M. Dijkstra, P.V. Lambeck, 'Evanescent Wave Sensing: New Features for Detection in Small Volumes', Presented at *Eurosensors '99*, The Hague, 1999.
153. B.H. Weigl, P. Yager, 'Microfluidic Diffusion-based Separation and Detection', *Science*, **283**, 346–347 (1999).
154. G.C. Fiaccabrino, M. Koudelka-Hep, 'Thin-film Microfabrication of Electrochemical Transducers', *Electroanalysis*, **10**, 217–222 (1998).
155. A. van den Berg, P.D. van der Wal, B.H. van der Schoot, N.F. de Rooij, 'Silicon-based Chemical Sensors and Chemical Analysis Systems', *Sens. Mater.*, **6**(1), 23–43 (1994).
156. A.G. Ewing, J.M. Mesaros, P.F. Gavin, 'Electrochemical Detection in Microcolumn Separations', *Anal. Chem.*, **66**(9), 527A–536A (1994).
157. A. van den Berg, A. Grisel, E. Verney-Norberg, 'An ISFET-based Calcium Sensor Using a Photopolymerized Polysiloxane Membrane', *Sens. Actuators B*, **4**, 235–238 (1991).
158. D.J. Strike, A. van den Berg, N.F. de Rooij, M. Koudelka-Hep, 'Spatially Controlled On-wafer and On-chip Enzyme Immobilization Using Photochemical and Electrochemical Techniques', *ACS Symp. Ser.*, **556**, 298–307 (1994).
159. A.T. Woolley, K. Lao, A.N. Glazer, R.A. Mathies, 'Capillary Electrophoresis Chips with Integrated Electrochemical Detection', *Anal. Chem.*, **70**(4), 684–688 (1998).
160. P.F. Gavin, A.G. Ewing, 'Continuous Separation with Microfabricated Electrophoresis–Electrochemical Array Detection', *J. Am. Chem. Soc.*, **118**, 8932–8936 (1996).
161. G.C. Fiaccabrino, N.F. de Rooij, M. Koudelka-Hep, J. Hendrikse, A. van den Berg, 'Microtechnology for the Development of In-situ Microanalytical Systems', in *In Situ Monitoring of Analytical Systems: Chemical Analysis and Speciation*, eds. J. Buffle, G. Horvai, IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, John Wiley & Sons, Chichester, 2000.
162. E. Jacobs, E. Vadasi, L. Sarkozi, N. Coman, 'Analytical Evaluation of I-STAT Portable Clinical Analyzer and Use by Nonlaboratory Health-care Professionals', *Clin. Chem.*, **39**(6), 1069–1074 (1993).
163. Ph. Arquint, M. Koudelka-Hep, N.F. de Rooij, H. Bühler, W.E. Morf, 'Organic Membranes for Miniaturized Electrochemical Sensors: Fabrication of a Combined pO_2, pCO_2, and pH Sensor', *J. Electroanal. Chem.*, **378**, 177–183 (1994).
164. R.B. Darling, P. Yager, B. Weigl, J. Kriebel, K. Mayes, 'Integration of Microelectrodes with Etched Microchannels for In-stream Electrochemical Analysis', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 105–108, 1998.
165. J. Hendrikse, W. Olthuis, P. Bergveld, 'Characterization of the ^{E}MOSFET, a novel one electrode chemical transducer for redox measurements', *J. Electroanal. Chem.*, **458**(1–2), 23–29 (1998).
166. J. Hendrikse, W. Olthuis, P. Bergveld, 'A One Spot Electrochemical Transducer for Use in μTAS', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 121–124, 1998.
167. G. Siuzdak, *Mass Spectrometry for Biotechnology*, Academic Press, San Diego, 1996.

168. N.H. Bings, C.D. Skinner, C. Wang, C.L. Colyer, D.J. Harrison, J. Li, P. Thibault, 'Coupling Electrospray Mass Spectrometry to Microfluidic Devices with low Dead Volume Connections', in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 141–144, 1998.
169. D. Figeys, S.P. Gygi, G. McKinnon, R. Aebersold, 'An Integrated Microfluidics–Tandem Mass Spectrometry System for Automated Protein Analysis', *Anal. Chem.*, **70**(18), 3728–3734 (1998).
170. Q. Xue, F. Foret, Y.M. Dunayevskiy, P.M. Zavracky, N.E. McGruer, B.L. Karger, 'Multichannel Microchip Electrospray Mass Spectrometry', *Anal. Chem.*, **69**(3), 426–430 (1997).
171. D. Figeys, R. Aebersold, 'Nanoflow Solvent Gradient Delivery from a Microfabricated Device for Protein Identifications by Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **70**(18), 3721–3727 (1998).
172. P. Bergveld, A. Sibbald, *Comprehensive Analytical Chemistry, Vol. XXIII: Analytical and Biomedical Application of Ion-selective Field-effect Transistors*, Elsevier, Amsterdam, 1988.
173. B.H. van der Schoot, S. Jeanneret, A. van den Berg, N.F. de Rooij, 'Modular Set-up for a Miniaturized Chemical Analysis System', *Sens. Actuators B*, 15–16, 211–213 (1993).
174. A. Ehlert, S. Büttgenbach, 'Automatic Sensor System for Water Analysis,' in *Micro Total Analysis Systems '98*, eds. D.J. Harrison, A. van den Berg, Kluwer, Dordrecht, 427–430, 1998.
175. B.J. Tortella, R.F. Lavery, J.V. Doran, J.H. Siegel, 'Precision, Accuracy, and Managed Care Implications of a Hand-held Whole Blood Analyzer in the Prehospital Setting', *Am. J. Clin. Pathol.*, **106**(1), 124–127 (1996).
176. I.R. Lauks, J.W. Rogers, 'Fluid Sample Collection and Introduction Device and Method', *US Pat.*, 5 779 650 (1998).
177. S.N. Cozzette, G. Davis, L.A. Holleritter, I.R. Lauks, S. Piznik, N.J. Smit, J.A. Tirinato, H.J. Wieck, M.P. Zelin, 'Method for Analytically Utilizing Microfabricated Sensors During Wet-up', *US Pat.*, 5 112 455 (1992).
178. P.C. Simpson, D. Roach, A.T. Woolley, T. Thorsen, R. Johnston, G.F. Sensabaugh, R.A. Mathies, 'High Throughput Genetic Analysis Using Microfabricated 96 Sample Capillary Array Electrophoresis Microplates', *Proc. Natl. Acad. Sci. USA*, **95**, 2256–2261 (1998).
179. A.T. Woolley, R.A. Mathies, 'Ultra-high-speed DNA Fragment Separations Using Microfabricated Capillary Array Electrophoresis Chips', *Proc. Natl. Acad. Sci. USA*, **91**, 11348–11352 (1994).
180. S.C. Jacobson, J.M. Ramsey, 'Integrated Microdevice for DNA Restriction Fragment Analysis', *Anal. Chem.*, **68**(5), 720–723 (1996).
181. L.C. Waters, S.C. Jacobson, N. Kroutchinina, J. Khandurina, R.S. Foote, J.M. Ramsey, 'Microchip Device for Cell Lysis, Multiplex PCR Amplification and Electrophoretic Sizing', *Anal. Chem.*, **70**(1), 158–162 (1998).
182. M.U. Kopp, A.J. de Mello, A. Manz, 'Chemical Amplification: Continuous-flow PCR on a Chip', *Science*, **280**, 1046–1048 (1998).
183. M.A. Burns, B.N. Johnson, S.N. Brahmasandra, K. Handique, J.R. Webster, M. Krishnan, T.S. Sammarco, P.M. Man, D. Jones, D. Heldsinger, C.H. Mastrangelo, D.T. Burke, 'An Integrated Nanoliter DNA Analysis Device', *Science*, **282**, 484–486 (1998).
184. K. Handique, B.P. Gogoi, D.T. Burke, C.H. Mastrangelo, M.A. Burns, 'Microfluidic Flow Control Using Selective Hydrophobic Patterning', *Proc. SPIE*, **3224**, 185–195 (1997).

Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry

Y.M. Dennis Lo, Ching-Wan Lam, and Ivy H.N. Wong
The Chinese University of Hong Kong, Shatin, Hong Kong

1 Introduction 1
2 History 1
3 Sample Preparation 2
3.1 Sources of Materials 2
3.2 DNA Preparation 2
3.3 RNA Preparation 2
3.4 Nucleic Acid Isolation from Archival Specimens 2
4 Techniques for Molecular Analyses 3
4.1 Polymerase Chain Reaction 3
4.2 Other Amplification Methods 3
4.3 Blotting and Hybridization Techniques 4
4.4 Detection of Unknown Mutations 5
4.5 Detection of Known Mutations 6
4.6 DNA Sequencing 7
4.7 Quantitation of Nucleic Acids 7
4.8 Homogeneous Assays for Nucleic Acid Amplification and Detection 8
4.9 In Situ Detection of Nucleic Acids 8
4.10 DNA Chip Technology 8
5 Molecular Pathology in Clinical Chemistry 10
5.1 Cancer 10
5.2 Prenatal Diagnosis 10
5.3 Mitochondrial Diseases 11
5.4 Diabetes Mellitus 12
5.5 Molecular Analysis of Cell-free DNA in Plasma and Serum 13
5.6 The Human Genome Project 13
Acknowledgments 13
Abbreviations and Acronyms 14
Related Articles 14
References 14

The recent introduction of powerful methods for the analysis of nucleic acids has started a revolution in medicine, opening up a new field of molecular diagnosis. The core technologies of molecular diagnosis involve techniques for the amplification and characterization of nucleic acids. These techniques have enabled the diagnosis of many diseases with unprecedented sensitivity and specificity. As research tools, these methods have increased our understanding of many diseases on a molecular level, with the unraveling of the molecular pathology of many common disorders. Foundation for the rapid and continual development of molecular pathology into the next century has received a boost from the Human Genome Project (HGP), which is expected to be completed over the next few years and which will provide the essential genetic blueprint for future investigations.

1 INTRODUCTION

Molecular biological analysis is the use of analytical procedures aimed at the elucidation of the chemical and physical basis of biological systems, especially with regard to the structure and function of genes. Many of the procedures in molecular biology are concerned with the manipulation and characterization of nucleic acid species, especially DNA and RNA. One advantage of molecular biological analysis is that a standard set of techniques can be applied to numerous biological systems because the target materials are mainly nucleic acids. A second advantage is the exquisite sensitivity of many molecular biological techniques.

Molecular pathology is the use of molecular biological techniques towards the understanding of the etiology and pathogenesis of diseases, and the development of diagnostic tests for these disorders. As proteins are the ''downstream'' products of the information flow from DNA to RNA to protein, the understanding of the aberration of ''upstream'' elements (i.e. DNA and RNA) in this pathway can improve our fundamental understanding of many diseases. Such improved understanding may ultimately result in the development of new therapies for many diseases.

2 HISTORY

The foundation of molecular diagnosis is the establishment of DNA as the factor of inheritance, a concept introduced by Avery in 1944.[1] This was followed closely by the elegant elucidation of the structure of DNA by Watson and Crick[2] in 1953. The subsequent application of these discoveries for molecular diagnosis has been

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

made possible by the development of new methods for analysing nucleic acids.

One such technique was introduced in 1975 by Southern[3] and involved the transfer of DNA, following size separation by gel electrophoresis, onto a solid support followed by hybridization to a probe. This technique is now referred to as the Southern blot and has been used for many of the molecular analyses in the 1980s.[4] Many of the subsequently developed methods for RNA and protein analysis, namely the northern and western blots, also bear certain resemblance to the Southern blot. In 1977, two groups have successfully developed methods for sequencing DNA.[5,6] The basic sequencing concept has been modified in numerous ways, including the introduction of fluorescence-based DNA sequencing[7] which is important in pushing the throughput of sequencing efforts.

A key concept that was introduced in the late 1970s was the idea of human DNA polymorphism[8] which has led to the flurry of papers on linkage studies in the 1980s.[9,10] In 1976, the first demonstration of the use of molecular techniques for clinical diagnosis was achieved.[11]

Up to 1985, the performance of molecular analysis generally takes many days or even weeks, and before this type of analysis can have a major clinical impact the turnaround time has to be markedly shortened. An important technique that has been instrumental in bringing molecular analysis to the clinical laboratories is the polymerase chain reaction (PCR).[12,13] This technique has made it possible to provide an analytical result within hours and allows molecular analysis to be carried out with unprecedented sensitivity, even down to a single molecule level.[14] The latter development has been important in the development of new diagnostic modalities, such as the analysis of single biopsied cells from preimplantation embryos.[15]

Up to 1989, the mapping of the human genome and the linkage of specific chromosomal regions to particular diseases had been hampered by the relative paucity of suitable polymorphic markers. In 1989, it was discovered that simple tandem repeat sequences in the human genome could serve as highly informative and frequent polymorphisms.[16,17] This discovery has made possible the rapid advances in linkage studies in the 1990s and the introduction of new concepts of genetic pathologies such as triplet repeat expansion and microsatellite instability in cancers.

The availability of a complementary set of technologies has made possible the launch of the HGP in 1990. The results from this project are expected to have long-term ramifications for science and medicine in the 21st century.

3 SAMPLE PREPARATION

3.1 Sources of Materials

Multiple sources of patients' specimens can be used for molecular analysis in a clinical setting. These include peripheral blood, plasma, serum, bone marrow, tumors and tissue biopsies. To obtain high integrity nucleic acids, especially RNA, from the specimens, one needs to process the samples as soon as possible.[18] In general, DNA samples are less fragile and thus isolated cells and cell-free components may be stored at $-20\,^{\circ}C$ before DNA extraction.

3.2 DNA Preparation

One of the first steps in DNA extraction is cell lysis followed by proteinase digestion.[19] Nucleic acids are then separated from proteins using phenol–chloroform–isoamyl alcohol. To obtain high purity DNA, RNA can be removed by digestion with RNases.[19] After ethanol precipitation and washing in 70% ethanol, DNA is resuspended in water or a Tris-based buffer and subjected to quantity measurement and quality assessment before molecular analyses. Alternatively, the application of genomic columns containing anion-exchange resin that binds DNA may increase the yield and purity of DNA.[19]

3.3 RNA Preparation

A popular method for RNA extraction involves the performance of cell lysis in the presence of guanidinium thiocyanate (which inhibits RNase activities), followed by acid–phenol–chloroform extraction.[18] DNA can be removed by DNase I treatment.[19] Extracted RNA is then subject to quantity measurement and quality assessment before molecular analyses. Alternatively, the application of columns containing anion-exchange resin that binds RNA may increase the yield and purity of RNA.[19] Recently, methods for simultaneous extraction of RNA and DNA from specimens have been developed which can improve the cost-effectiveness and minimize the amount of specimen required for molecular studies.[20]

3.4 Nucleic Acid Isolation from Archival Specimens

A powerful feature of molecular analysis is the ability of this technology to analyze archival materials, thus opening up the possible use of valuable archives in many pathology departments. Bone marrow smears or cell smears on glass slides are well established sources of starting materials.[21] Paraffin-embedded specimens are another very valuable source of test materials. However, the integrity of RNA or DNA may be poor if the archival specimens have been

stored for years or if the fixation steps have resulted in over-crosslinking of the DNA to proteins.

4 TECHNIQUES FOR MOLECULAR ANALYSES

4.1 Polymerase Chain Reaction

The PCR is a primer-directed enzymatic amplification of DNA using a DNA polymerase.[12,13] The PCR can be regarded as a simplified in vitro version of DNA replication which normally operates inside cells. The main components of a PCR consist of: (1) a pair of oligonucleotides, typically between 18 and 30 bases each, which are called the primers; (2) a DNA polymerase which is typically thermostable so as to withstand the high temperatures of the denaturation step; (3) deoxynucleoside triphosphates which provide the raw materials for the synthesis of new DNA strands; and (4) ionic and other cofactors.

The PCR can be regarded as consisting of three steps: (1) thermal denaturation – during this step, the normally double-stranded DNA molecules are denatured into a single-stranded form by heating, typically 94 °C to 95 °C; (2) primer annealing – during this step, the temperature is lowered to typically 50 °C to 60 °C to allow the primers to anneal to target sequences, that is sequences which share an exact or almost exact DNA sequence as the primers; and (3) primer extension using a DNA polymerase – during this step, which typically occurs at 72 °C, the DNA polymerase is allowed to extend the primers by synthesizing new DNA strands. Each combination of these three steps is called a PCR cycle. By the repetition of this process, exponential amplification of the target DNA sequences will result. The temperature alterations are generally carried out using instruments called thermocyclers. Each synthesized DNA molecule is called a PCR product which possesses a length from the 5′-end of one primer to the 5′-end of the second primer. Most PCRs are designed to generate a product length between 100 bp and 400 bp, although very long target lengths have been achieved using special combinations of polymerases.[22–24] The amplification factor (F) is given by Equation (1):

$$F = x(1 + E)^n \quad (1)$$

where x is the starting number of target molecules, E is the amplification efficiency and n is the number of PCR cycles.

With the performance of an additional reverse transcription step, the PCR can also be applied to the amplification of RNA targets, a process known as reverse-transcription polymerase chain reaction (RTPCR).[25] The use of polymerases exhibiting both reverse transcriptase and DNA polymerase activities, such as the *Thermus thermophilus* (*Tth*) DNA polymerase,[26] has enabled the performance of single-tube RTPCR.[27,28]

The PCR is a very sensitive process, capable of detecting a single molecule of target sequence.[14,29,30] This characteristic makes the PCR very valuable for the detection of target sequences which are present at extremely low concentrations, for example, certain pathogens.

The exquisite sensitivity of the PCR is also the inherent weakness of the method, namely, the proneness of the method to generate false-positive results due to exogenous contamination.[31–33] The main culprit for contamination is the carryover of PCR products from previous amplification reactions. Another source of contamination is target nucleic acid species from samples possessing the targets to those in which the targets are not endogenously present. Obsessive care is necessary during the sample preparation and PCR set-up stages to reduce the chance of contamination. Separate areas should be dedicated to the pre- and post-PCR steps. Multiple negative controls should be used to monitor the occurrence of PCR contamination.[34] A number of methods have been devised to reduce the chance of PCR contamination.[35–37] These methods are generally based on the principles of selectively destroying any contaminating PCR products or by rendering them unamplifiable.

4.2 Other Amplification Methods

Since the introduction of the PCR, many researchers and commercial companies have proposed alternative schemes aimed at accomplishing some or all of the functions of the PCR. One such method is the ligase chain reaction (LCR) which utilizes the ligation of oligonucleotides hybridizing in an adjacent fashion onto the target, with oligonucleotide pairs synthesized towards both strands of the DNA.[38] Thus every ligation cycle will in principle double the number of ligation templates. Potential clinical diagnostic systems based on the LCR have been proposed.[39,40]

The PCR and LCR both require the use of thermal cycling. Amplification schemes allowing the performance of isothermal amplification have also been described. Two such methods are nucleic acid sequence-based amplification (NASBA)[41,42] and strand displacement amplification (SDA).[43–45] NASBA is based on the sequential performance of a number of enzyme-mediated reactions: reverse transcription of an RNA target molecule using a reverse transcriptase; and the enzymatic digestion of the RNA template and the synthesis of a DNA strand in its place.[41] The resulting double-stranded DNA molecule

is designed such that an RNA promoter is present at one end. This molecule can then serve as a template for the transcription of multiple copies of RNA molecules, each capable of initiating the same series of reactions. The principle of SDA is based on repeated cycles of primer extension reactions following enzyme-mediated nicks in a double-stranded target DNA molecule.[46] During polymerization, the downstream strand is displaced and the displaced strands can then serve as templates for further cycles of the nicking and strand displacement steps.

Amplification using the Q-beta replicase[47] is an amplification scheme based on the use of an RNA-dependent RNA polymerase derived from the bacteriophage Q-beta.[48] Q-beta replicase is capable of producing a complementary product strand of a template RNA in as little as 12 s.[48] As the product strands are also templates for the replicase, exponential accumulation of the target RNA molecules will result. A single target molecule can yield an easily detectable amount of product in a short 15-min reaction.

In addition to these schemes which involve the amplification of nucleic acids, detection methods have been developed which depend on signal amplification, rather than target amplification. One such approach is the branched DNA (bDNA) assay which achieves signal amplification by secondarily binding large numbers of enzymatic reporters to hybridized probes.[49–52] The bDNA system is a solid-phase sandwich hybridization assay incorporating multiple sets of oligonucleotide probes and several simultaneous hybridization steps. Five to nine target-specific probes serve as capture extenders binding to the target DNA or RNA onto the surface of a microtiter plate.[50] A second set of target-specific probes (18 to 39) is used as labeled extenders binding to the target nucleic acid. In the first-generation assays, the labeled extenders also provide binding sites for the bDNA amplifier molecules with 15 identical arms. Each of these arms can bind three alkaline phosphatase-labeled probes. In other words, up to 3000 enzyme-labeled probes can be hybridized to each target DNA/RNA copy. Bound labeled-probes are detectable by using a chemiluminescent substrate, dioxetane, and quantified by measuring the light emission.[50] The resulting signal is proportional to the concentration of the target nucleic acid. The quantity of the target in the test sample is then determined from a standard curve.

In the second- and third-generation bDNA systems, preamplifier molecules are used to enhance the number of labeled probes bound to the target, and to reduce the background signal.[53] The labeled extender probe sequences are designed to bind to adjacent regions of the target to allow efficient hybridization to the preamplifier molecules. Two kinds of preamplifier molecules are used which contain the same repeat sequence and differ only in the sequences hybridized to the target probes.[53] Each preamplifier has a site specific to the overhanging sequences (15–16 bases) of the target probe set 2. When the overhanging sequences of two target probes are adjacent, the melting temperature increases and hence stabilizes the binding of the preamplifiers. Each preamplifier can hold up to 8 bDNA amplifier molecules and a total of 360 alkaline phosphatase-labeled probes. As a result, each target DNA/RNA may bind up to 10 000 enzyme-conjugated probes.

4.3 Blotting and Hybridization Techniques

Blotting of nucleic acids onto nylon membranes and subsequent ultraviolet (UV) illumination will immobilize the test DNA/RNA, which can then be subjected to hybridization with target probes afterwards. For standard Southern and northern blot analyses, nucleic acids on agarose gels are blotted onto nylon membranes by capillary action or vacuum transfer.[19] Radioactive or nonradioactive hybridization is conducted afterwards using, for example, ^{32}P-labeled or digoxigenin-labeled copy DNA (cDNA) probes or oligonucleotide probes.[19,54] After washing, the blots are exposed to X-ray films. When a nonradioactive technique is applied, an enzyme-conjugated antireporter antibody is added onto the membrane, which is then subject to colorimetric or chemiluminescent detection.[19,54]

4.3.1 Northern Blot Analysis

Northern blot analysis is used for studying messenger RNA (mRNA) expression of specific target genes.[19] Total RNA from samples is separated using gel electrophoresis and then subjected to ethidium bromide staining. The integrity of ribosomal RNA (rRNA) bands reflects the mRNA integrity. Total RNA is blotted onto a nylon membrane and subjected to prehybridization for blocking nonspecific binding and then hybridization with a labeled probe.

4.3.2 Southern Blot Analysis

Southern blot analysis is useful for analysis of deletion, rearrangement and methylation of target genes.[19] Generally, genomic DNA is first digested with specific restriction enzymes. A wide variety of restriction enzymes can be used for different purposes. For example, methylation-sensitive restriction enzymes that only cleave unmethylated sites but will not cut methylated residues could be used for the methylation analysis[55] (see also section 5.1). DNA fragments are then separated by gel electrophoresis, subjected to ethidium bromide staining and blotted onto a nylon membrane. In addition, the

Southern blot can also be used for confirming the identity of PCR products.

4.3.3 In Situ Hybridization

In situ hybridization is a technique for analyzing target nucleic acids within cells fixed on glass slides. Probes can be labeled with either radioactive or increasingly nonradioactive labels (e.g. digoxigenin).[56] After sealing with a coverslip, the targets are denatured by heat and then hybridized with the labeled probe. After appropriate washing steps, the hybridized probes are visualized, for example either by autoradiography for radioactive probes or by using an enzyme-linked colorimetric detection system for nonradioactive probes.

An important group of nonisotopic labels are fluorescent reporters, in which case the technique is called fluorescence in situ hybridization (FISH). FISH is used for many different purposes, namely the analysis of structural and numerical chromosome aberrations, clonality studies and clinical diagnosis. FISH enables the identification of the presence and location of a specific DNA region within morphologically preserved chromosome preparations, fixed cells, and fresh-frozen and paraffin-embedded tissue sections.[57] Chromosomal translocations, deletions, amplifications and other rearrangements are detectable by FISH.[58] The sensitivity of FISH allows the detection down to 1–10 kb and hence the precise localization of genes on chromosomes. FISH can be applied to interphase nuclei using centromere-specific probes or chromosome arm-specific telomeric probes.[59,60] For standard probe preparation, in vitro amplification and labeling with fluorescent dyes are conducted. Telomeric probes are used to improve the sensitivity for detecting microdeletions or other cryptic rearrangements in the telomeric regions.[60] Specific long DNA probes are prepared by vector cloning and oligonucleotide primed in situ labeling, and the signals are detected by an enzyme-linked detection system.[61] For analysis of gene amplification, fluorogenic gene-specific probes together with fluorescent centromeric reference probes are used.[62] Greater than 10 signals per cell or tight clusters of signals in >5% of cells in the confined regions are classified as gene amplification (see also section 5.1).

The recently developed multi-color and ratio-mixing FISH techniques using dual-color and three-color probes are reliable for simultaneous analyses of different chromosomal aneuploidies.[60,61,63,64] The recent development of spectral karyotyping (SKY) has allowed the simultaneous visualization of all human chromosomes[65] and promises to provide much enhanced precision for molecular cytogenetic analysis.[66,67]

4.3.4 Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) is a molecular cytogenetic method based on the co-hybridization of fluorescent-labeled test genomic DNA and normal reference genomic DNA on normal metaphase chromosomes.[68] The relative deficiency or excess of a particular chromosomal region will be detected as a color shift. Standard DNA is labeled by nick translation. Archival DNA (small fragment size of 400–1000 bp) can be labeled using a universal linkage system.[69] CGH enables the identification of all unbalanced chromosome aberrations of the test DNA, such as DNA gains or losses on specific chromosomes. Test DNA samples may be derived from fixed, paraffin-embedded and fresh-frozen cells and tissues. Four-color CGH is based on conventional CGH with an additional cyanine-5 (Cy5)-labeled reference DNA as an internal standard in hybridization.[70] This internal reference standard can identify inconsistently hybridized chromosomal regions and standardize the dynamic range of hybridization.

Chromosome-specific CGH is a newly developed approach for detecting cytogenetic abnormalities.[71] This combines flow sorting of chromosomes, degenerate oligonucleotide primed polymerase chain reaction (DOPPCR) and a modified CGH technique for identifying the site and extent of intrachromosomal gains or losses. Chromosome-specific paint probes for aberrant chromosomes and their normal homologs are first labeled and cohybridized on normal metaphase chromosomes. The ratio of the relative intensities of hybridization signals is determined.

4.4 Detection of Unknown Mutations

Many of the recent developments in molecular genetics have centered around the discovery of disease-causing genes and the elucidation of disease-associated mutations. Tools for the detection of mutations have played a major part in many of these discoveries. DNA sequencing remains the ''gold standard'' in the detection of mutations. However, before high-throughput sequencing at costs acceptable to most laboratories become commonplace, there is still a need for mutation screening.

The ideal method for mutation detection would desirably be accurate, fast, inexpensive, able to provide detailed information regarding the nature and position of the mutation, nonlabor-intensive, automatable and avoid the use of harzardous reagents. None of the currently available methods fulfill all of these requirements. Nonetheless, each of the available methods has its own strength–weakness profile and should be selected according to budgetary allocation, workload and project nature of individual laboratories.

There are two major classes of mutation screening methods, namely those that depend on the conformation of DNA molecules and others which rely on the detection of mismatched bases. The conformation-based methods are based on the principle that DNA molecules of different sequence contexts will exhibit different electrophoretic mobilities under particular conditions. One group of conformation-based mutation screening procedures are based on the use of denaturing agents and include denaturing gradient gel electrophoresis (DGGE),[72,73] temperature gradient gel electrophoresis (TGGE)[74] and constant denaturant gel electrophoresis (CDGE).[75] These techniques utilize the principle that areas of single-stranded DNA within double-stranded segments greatly retard the electrophoretic mobility of DNA through acrylamide gels. Denaturation within a region of DNA is strongly dependent on the sequence of that region and even single base change can be detected using these techniques. For DGGE and CDGE the denaturing agents are chemical in nature (usually formamide and urea), while for TGGE heat is used. DGGE and TGGE use a gradient of increasingly denaturing conditions over the length of the gel. CDGE uses a single concentration of denaturant very close to that which separates the strands of the normal fragment. Methodological variants involving combinations of these methods have been described and reported to exhibit advantages of their constituent methods.[76]

A second group of conformation-based techniques is based on the electrophoresis of DNA molecules under nondenaturing conditions. Examples of such methods include heteroduplex analysis (HA)[77] and single-strand conformation polymorphism (SSCP) analysis.[78] HA is based on the principle that mismatched bases between a mutant and wild-type strand will result in a DNA heteroduplex which has reduced electrophoretic mobility on a polyacrylamide gel.[77] HA has the main advantage that the assay conditions do not have to be determined for each DNA fragment. Recently, high-throughput versions of HA have been developed using denaturing high-performance liquid chromatography (DHPLC), instead of time-consuming gel electrophoresis.[79,80] The major advantages of this development include the use of automated instrumentation, rapidity (approximately 6 min per sample) and the ability to analyze relatively long DNA fragments (up to 1.5 kb).[80]

SSCP is based on the principle that single-stranded DNA molecules of different sequences will take on different secondary conformations in a denaturing polyacrylamide gel and will thus exhibit different electrophoretic mobilities.[78] Even though the electrophoretic conditions for SSCP are carried out under nondenaturing conditions, the DNA fragments to be analyzed are first denatured by formamide and heat prior to loading on the gel. A number of electrophoretic parameters, for example electrophoretic temperature and the presence of glycerol, can be varied which will determine the number of mutations that can be detected. Using a single gel condition, approximately 70% of mutations can be detected. This figure can be increased to approximately 95% by the use of multiple electrophoretic conditions.[81–83] Following electrophoresis, the separated DNA conformers are detected by radioactivity,[78] silver staining[84] or fluorescence.[84–86]

The second class of mutation screening methods are techniques that utilize base mismatch recognition. These techniques generally exploit chemicals or proteins that act at or bind to sites of mismatched bases in DNA.[87] One of the earliest examples of this class of methods is chemical mismatch cleavage (CMC).[88,89] CMC employs hydroxylamine and osmium tetroxide to modify DNA at sites of base mismatch. The modified DNA is then cleaved with piperidine and the reaction products are separated by polyacrylamide gel electrophoresis. The location of the mutation can then be deduced from the sizes of the cleaved fragments. The major disadvantage of CMC is that harzardous chemicals are used and therefore the method is not suitable for use in a routine laboratory.

In addition to the use of chemicals to detect mismatched bases, it is also possible to utilize proteins or enzymes for this purpose. Examples include S1 and mung bean nucleases.[90] More recent methods have utilized bacteriophage enzymes known as resolvases, such as T4 endonuclease VII[91,92] and T7 endonuclease I,[91] for cleaving DNA heteroduplexes. This approach has been called enzyme mismatch cleavage (EMC)[93] or enzymatic mutation detection (EMD).[94]

4.5 Detection of Known Mutations

Following the elucidation of unknown mutations, it is sometimes desirable to develop detection methods to specifically detect a particular mutation, especially if the mutation in question is a common cause of a disorder in a particular population. As DNA amplification is almost invariably used in modern methods for detecting known mutations, these methods can be classified into two main groups, based on the role of the amplification step. In the first group, the function of the amplification step is merely to produce a large quantity of product for subsequent analysis. In the second group, the amplification and diagnostic steps are combined such that the amplification step becomes mutation-specific.

Restriction enzyme analysis of PCR products is one of the first methods for the detection of previously characterized mutations.[12] The main limitation of this method is that a particular mutation may not result in restriction site creation or deletion in any known restriction endonuclease. This limitation can be overcome

to a large extent by the introduction of artificial mismatches in the primers so as to create restriction sites which are affected by the mutation.[95–97] Another well-established approach is the performance of allele-specific oligonucleotide (ASO) hybridization on dot blots of the PCR products.[98] In situations where there are a large number of possible mutations affecting a single gene, it is possible to speed up the procedure by performing a reverse dot blot, where oligonucleotides specific for the different mutations are immobilized on a filter and then a labeled PCR product is hybridized to the probes "in reverse".[99] Newer variations of allele-specific hybridization include the inclusion of fluorescent probes against either mutant or wild-type PCR products during amplification, with different reporters being used for the mutant- and wild-type-specific probes.[100–102]

For the second group of techniques for detecting known mutations, that is, those that rely on allele-specific amplification, the most important method is allele-specific PCR,[103] also known as the Amplification Refractory Mutation System (ARMS).[104,105] This method is based on the fact that the *Thermus aquaticus* (*Taq*) DNA polymerase that is commonly used for PCR does not possess proofreading capability and that a primer with a 3′-end that is mismatched to the template will prime with very low efficiency.[103,104] Thus, allele-specific PCR systems can be designed such that the 3′-end of the primers possesses the same sequence as the mutant or wild-type variety of a particular sequence. It is also possible to achieve allele-specific amplification using the LCR.[40] Apart from using PCR primers that are allele-specific, it is also possible to introduce a blocking oligonucleotide with a 3′-terminal dideoxynucleotide between the two primers such that the amplification of either the mutant or the wild-type allele is blocked by hybridization of this oligonucleotide, a technique known as "blocker PCR".[106] Methods combining the principles of allele-specific PCR and blocker PCR have also been described.[107,108]

4.6 DNA Sequencing

Methods for DNA sequencing were first developed in 1977.[5,6] The method by Maxam and Gilbert[6] is based on the chemical cleavage of DNA at a specific nucleotide using chemical methods. The chemically treated DNA fragments are then electrophoresed and the sizes of the fragments are dependent on the positions of the respective nucleotides. The method by Sanger, on the other hand, relies on the interruption of in vitro DNA synthesis at specific bases of the DNA segment to be sequenced, using dideoxynucleotides.[5] At present, the method by Sanger is the predominant method that is being used by sequencing laboratories around the world.

With the continual requirement for increased sequencing throughput, semiautomated sequencing has been developed.[7] These methods generally depend on the automated detection of fluorescence-labeled DNA fragments. Most of these semiautomated methods have relied on the use of gel electrophoresis. To allow further automation, DNA sequencing based on capillary electrophoresis has been developed[109,110] and instruments with multiple capillaries allow extremely high throughput and are expected to speed up projects with large sequencing requirements, such as the HGP (see section 5.6).

New methods have also been developed for DNA sequencing, including chip-based DNA sequencing (see section 4.10) and methods based on mass spectrometry.[111] These methods are hopeful candidates for further allowing the widespread application of DNA sequencing in clinical diagnosis.

4.7 Quantitation of Nucleic Acids

One of the simplest methods for quantifying nucleic acids is by the use of spectrophotometry. The absorbances at 260 nm and 280 nm are taken and the ratios of these absorbances are taken as an indication of the purity of the nucleic acids.[19] In general, a ratio of approximately 1.8 is taken as a relatively pure preparation of DNA.

Fluorimetry is another method for quantifying nucleic acids. This approach generally requires the addition of nucleic acid-binding fluorescent dyes and measurement using a fluorimeter. Examples of proprietary dyes which can be used include Hoechst 33258 and 33342, and Molecular Probes PicoGreen™.

Specific fragments of DNA or mRNA can also be quantified using quantitative PCR or RTPCR.[112] This method can be performed in a competitive or a noncompetitive manner.[112] Conventional quantitative PCR is generally regarded as a relatively difficult technique to standardize and has a relatively narrow dynamic range. These shortcomings have recently been largely overcome with the introduction of real-time quantitative PCR.[113,114] In this approach continuous optical monitoring of the progress of a PCR is performed by coupling target amplification to the generation of a fluorescent reporter.[113,114] This method allows the online acquisition of quantitative results without any post-PCR processing, thus effectively reducing the risk of PCR product carryover contamination. The availability of this method has allowed the rapid development of quantitative detection methods for gene expression,[115,116] infectious agents[117,118] and gene dosage[119] (see section 4.8).

Variants of the basic real-time quantitative amplification method have been reported based on different amplification strategies[45] or different fluorescent signal generation schemes.[120]

4.8 Homogeneous Assays for Nucleic Acid Amplification and Detection

Nucleic acid amplification technologies such as the PCR are essential for many types of molecular analysis. The main disadvantage of these amplification-based methods is their sensitivity to carryover contamination of amplicons from previous reactions.[31,32] This important shortcoming has hindered the large-scale application of molecular analysis in the clinical laboratories. Recently this obstacle has largely been overcome with the development of a number of homogeneous assays for nucleic acid amplification and detection. The main strength of these methods is that no postamplification processing is necessary, thus avoiding the liberation of amplicons to the laboratory environment.

These homogeneous assays have all relied on the continuous optical monitoring of an amplification reaction in which product amplification is coupled to the liberation of a fluorescent reporter. The kinetics of fluorescence change is then processed by an online computer system and proprietary algorithms are available to rapidly output the data in either a qualitative or a quantitative format. Most of the available systems are based on the PCR[113,114,121,122] although other amplification formats such as SDA have also been converted into a real-time format.[45] A number of signal generation systems have been developed, including those based on the 5′-nuclease assay (or TaqMan™),[123] the BDProbe Tec™ET system[45] and molecular beacons.[120] The performance of real-time nucleic acid amplification and detection requires the use of specialized thermal cyclers with online fluorescence detection capability. A number of such instruments are available commercially including those that are built upon a conventional DNA thermocycler[113] as well as those that utilize rapid thermal cycling using capillaries.[120]

These homogeneous assays have been applied for nucleic acid target quantitation,[113,114,121] infectious agent detection[117,118,120] and allelic discrimination.[100] Their rapidity and robustness are likely to catalyze the introduction of molecular diagnosis by the clinical laboratory.

4.9 In Situ Detection of Nucleic Acids

The extraction of nucleic acids from the target tissues for molecular analysis has the disadvantage that it destroys information regarding the localization of the nucleic acid species with respect to particular cell types and the subcellular location. To preserve such information the performance of nucleic acid analysis in situ is necessary. One method is to carry out in situ hybridization (section 4.3.3). Another method is to perform in situ nucleic acid amplification such as in situ PCR.[124] In situ PCR is the performance of amplification within intact cells or tissue sections of specific gene sequences, or mRNA species, to levels detectable by in situ hybridization and/or immunocytochemistry. In situ PCR has been subdivided into: direct in situ PCR in which reporter labels are incorporated during amplification; and indirect in situ PCR in which in situ amplification is carried out first, followed by in situ hybridization.[125] It has been accepted by many investigators that the indirect approach offers much better specificity than the direct approach.[126] In situ PCR technology requires exhaustive optimization of various empirical variables including tissue fixation, protease digestion, amplification parameters such as the hot start maneuver, hybridization stringency, and the use of DNase digestion (for in situ RNA analysis).[127] Purpose-built thermal cyclers for in situ PCR are available from a number of companies.

Apart from in situ PCR, other molecular methods have also been described for the intracellular analysis of nucleic acids, including the isothermal self-sustained sequence replication (3SR)[128,129] and the oligonucleotide primed in situ DNA synthesis (PRINS).[130]

4.10 DNA Chip Technology

DNA chip assays are referred to as biochips, oligonucleotide array assays, and hybridization array assays. DNA chip assays are composed of a large collection of cDNAs or expressed sequence tags (ESTs) immobilized on glass (termed microarrays) or sets of synthetic oligonucleotides immobilized on silicon chips (termed probe arrays). Regardless of the array chosen, the assay relies on the hybridization of a population of labeled DNAs in solution to DNA sequences immobilized on the solid phase. On oligonucleotide microarrays, different oligonucleotides are arrayed at discrete locations (a few micrometers to several millimeters in size) on a silicon chip. Very large oligonucleotide arrays have been produced with as many as 65 536 sites (all possible 8-mers), and efforts to place even larger numbers of probes on a chip are underway. Affymetrix has created the GeneChip® technology based on high density DNA arrays containing DNA sequences as a method to analyze genetic information in the human genome.[131] Application-specific DNA probe arrays are designed for the rapid screening of gene expression within selected portions of the human genome (described in section 4.10.2). Along with this technology, Affymetrix has developed instrumentation and software for array hybridization, fluorescence detection, data acquisition, and data analysis. In addition to this much higher density of information, DNA chip assays have an advantage over gel-based assays because the nucleic acids are bound to a reusable solid support, allowing for improvements in automation and data analysis. The DNA chip technology

allows biomedical researchers to study entire genes and gather information about expression and gene mutation patterns.

4.10.1 Gene Expression

With DNA chip assays, RNA is extracted from the cells or tissues to be studied, and ''tagged'' cDNA is then made from mRNA in the extract. The product, tagged with fluorescent nucleotide analogs, is hybridized to the array, which is then washed to remove unhybridized material. Hybridized fluorescence-labeled products can be detected directly. To do this, the array is optically scanned or ''read'' by an inverted scanning fluorescence confocal microscope with a triple laser illumination system, which can assess differing red/green/yellow emissions. The resulting data are analyzed by powerful computer packages which integrate and formulate quantitative data generated from the vast number of potential hybridizations on a single chip. For example, one population of cDNAs derived from a cancer cell line can be fluorescence-labeled with a red tag, while a second population of cDNAs derived from a reference cell line can be labeled green. If both populations of cDNAs bind to the same target on a chip, a yellow signal is obtained. In this way, global gene expression can be compared in two populations. The method's great potential is that it allows the simultaneous comparison of the expression of many thousands of genes in a single experiment.[132]

4.10.2 Genotyping

DNA array analysis has also been widely used to detect genetic mutations and polymorphisms. Currently available applications for the oligonucleotide array chips include DNA sequencing and mutation detection, for example, β-*globin* gene mutations in β-thalassemia.[133] The Collins groups at the National Institutes of Health developed DNA arrays of 96 000 oligonucleotides to screen for 24 heterozygous mutations within exon 11 of the hereditary breast and ovarian cancer gene, that is, *BRCA1*.[134] A diagnostic DNA chip able to detect all mutations within the *BRCA1* gene will require 400 000 probes and is currently under investigation.[135] The p53 chip developed by Affymetrix includes over 400 mutations that have been found to be associated with tumors and has been marketed to determine individuals with increased cancer risk.

With the growing understanding of the importance of sequence variations, for example, single-nucleotide polymorphism (SNP), between individuals, large pharmaceutical companies have invested heavily in DNA chip technology. In a related application, the efficacy of therapeutic treatment with protease inhibitors for AIDS patients can be monitored with DNA chips designed to detect sequence variants in the human immunodeficiency virus type 1 (HIV-1)-clad B protease gene that confers protease resistance.[136] It is more cost-effective to use a chip to determine the genetic variant of HIV harbored by a patient before starting therapy, than to prescribe an inappropriate drug to an individual who will then fail to respond.

4.10.3 Sequencing by Hybridization

To detect a gene expressed at levels as low as one copy per cell, the company Hyseq has developed sequencing by hybridization technology.[132] This technology can identify target genes without prior sequence knowledge. A series of short, overlapping probes are hybridized and reconstructed by software to recreate the DNA sequence. An extension of this technology yields a universal sequencing chip that is a potentially powerful tool for both diagnostics and gene discovery.

4.10.4 Current Limitations of DNA Chip Technology

At present, there are publicly available, validated cDNA and EST sequences for only a proportion of expressed genes. Out of a total of 10 000 to 50 000 expressed genes in a given human cell, only 15 000 can be detected by DNA chips because the publicly available Unigene set contains only 15 000 cDNA clusters, each representing a unique human gene. Another limitation is that several cDNA libraries must be analyzed to assess the full complement of expressed genes.

Currently, the most advanced expression array systems are very expensive and, as a result, have only been available to large pharmaceutical companies or the laboratories in which the technologies are being developed. It is possible to reuse chips, but there is a theoretical loss of sensitivity with each successive hybridization. Until the cost per experiment falls, the number of experiments that can be performed and the number of replicates per experiment will be limited. With few replicates, complete confidence in the results will only be obtained in cases with large changes in mRNA levels. Thus, it is possible that some low-abundance mRNA species will not be identified by DNA chip assays.

In addition, software and database systems to design arrays and collect, analyze, and interpret data from gene expression studies are still in their infancy. Very large quantities of data have to be managed both before and after the experiment, because direct access is required to all sequences, annotations, and physical DNA resources for the genes of the organism studied.

5 MOLECULAR PATHOLOGY IN CLINICAL CHEMISTRY

5.1 Cancer

Cancer can essentially be regarded as a genetic disease. Five different types of genetic alterations have been described in cancer cells:

1. Subtle genetic alterations: these include small deletions, insertions and single base pair substitutions. Examples of these subtle genetic alterations include activating mutations for oncogenes[137] and inactivating mutations for tumor suppressor genes.[138]
2. Copy number changes in chromosomes: these include the somatic losses or gains of chromosomes. Some of these changes can be observed by conventional or molecular cytogenetic analysis. Apart from cytogenetic analysis, these changes can also be observed by the study of polymorphic markers on the paternally and maternally inherited chromosomes. The loss of one of these chromosomes results in a phenomenon referred to as loss of heterozygosity (LOH).[139]
3. Chromosomal translocations: these are best characterized in hemopoietic malignancies including leukemias and lymphomas.[140] In most instances, chromosomal translocation fuses sequences of a transcription factor or receptor tyrosine kinase gene to those of a normally unrelated gene, resulting in a chimeric protein with oncogenic properties.[141] Another scenario involves the re-positioning of transcriptional regulatory genes to the neighborhood of highly active promoter/enhancer elements, such as those associated with immunoglobulin genes. Our knowledge of chromosomal translocations in human solid tumors is less detailed than that for hemopoietic malignancies, although a number of nonrandom alterations have been documented.[142] Recently, detailed molecular characterization has been accomplished for several human sarcomas[143,144] and may become the paradigm for other human solid cancers.[145]
4. Gene amplifications: these represent an increase in the copy number of a small region (generally 1 to 2 megabases (Mb) or less) of a chromosome.[146] Gene amplifications are generally only observed in relatively advanced neoplasms. The ''amplicons'' contain one or more genes whose expression can provide the cell with enhanced proliferative potential, with a growth advantage compared with cells not possessing the change.
5. Exogenous sequences: these include tumor viruses, which contribute genes resulting in abnormal growth. Examples include human papillomavirus (HPV) in cervical cancers[147] and hepatitis B virus (HBV) in hepatocellular carcinomas.[148]

Apart from these genetic alterations, it has been increasingly recognized that epigenetic factors are also important in oncogenesis.[149] Epigenetic inheritance is a biochemical modification of the genome which is heritable by progeny cells, without involving a change in DNA sequence. DNA methylation is now known to be an important mechanism of epigenetic inheritance[149,150] and involves the methylation of the C-5 position of cytosine by a DNA methyltransferase. One methylation-dependent oncogenic pathway is the inactivation of tumor suppressor genes by methylation of CpG islands of the promoter region.[151–153] A second epigenetic oncogenic mechanism is the disruption of genomic imprinting, referred to as loss of imprinting (LOI).[154,155] Imprinting is a term used for describing the phenomenon of differential gene expression or nonexpression based on the parent-of-origin of a particular allele of a gene.

The elucidation of the numerous genetic and epigenetic changes that can lead to cancer has provided new tools for the detection and monitoring of cancers. An important advance is the ability of molecular methods to detect the presence of malignancy from clinical materials that can be obtained relatively noninvasively, such as blood. Thus, there is a lot of interest in the detection of circulating tumor cells in the blood of patients suffering from a variety of cancers, not just cancers of hemopoietic origin.[156] Molecular targets that have been detected include tumor-associated genetic alterations, such as oncogene mutations, thus involving the analysis of tumor-derived DNA. A second class of targets include mRNA from the tissue of origin of the malignant tumor, such as tyrosinase mRNA from melanoma[157] and α-fetoprotein transcripts in patients with hepatocellular carcinoma.[54] In addition to blood cells, cell-free blood plasma and serum have also been proposed as valuable materials for tumor detection (section 5.5). Apart from blood, other easily assessable sources of bodily materials have also been used for tumor detection, including sputum,[158,159] stools[160,161] and urine.[162–164] The availability of these new and relatively noninvasive modalities for cancer detection may allow the earlier diagnosis of many neoplasms, leading to a prompter intervention and hopefully an improved survival from these diseases.

5.2 Prenatal Diagnosis

The identification of the causative genes for many inherited diseases has opened up the possibility of prenatal diagnosis for these disorders. The development of molecular cytogenetics has also made possible the performance of rapid chromosome analysis. A source

of fetal materials is necessary for carrying out prenatal diagnosis using these techniques. Fetal materials are usually obtained by invasive sampling methods such as amniocentesis, chorionic villus sampling, fetal blood sampling and fetal biopsy. However, these methods are associated with a finite risk to the fetus and mother.

The development of molecular biology has allowed the development of new methods for prenatal diagnosis which do not carry such risks. Amongst these methods, there is a real hope that the isolation of fetal nucleated cells that have entered the maternal circulation during pregnancy may allow the realization of noninvasive prenatal diagnosis.[165] The first demonstration that fetal nucleated cells do indeed enter into maternal circulation using molecular techniques was achieved in 1989,[166] using the amplification of Y chromosomal sequences from DNA extracted from maternal peripheral blood cells. Since then numerous investigators have repeatedly confirmed these observations[167–170] and have developed many methods aimed at enriching these extremely rare components of the maternal blood. Thus far, the most promising approaches are based on the isolation of fetal nucleated red blood cells that have entered into the maternal blood, using the fluorescence activated cell sorter,[171] magnetic sorting[172] and newer methods such as charge flow separation.[173]

Apart from obvious diagnostic possibilities, the trafficking of nucleated cells between the mother and fetus has raised numerous research opportunities for understanding the fetomaternal relationship on a molecular level.[174] For example, it is now known that the fetomaternal transfer of nucleated cells is actually a two-way process, with maternal nucleated cells being found in the fetal circulation,[174–177] in addition to fetal cells in the maternal circulation mentioned above. Interestingly, recent evidence indicates that certain fetal cell populations can persist for extended periods (e.g. years) following parturition[178] and that this persistence may lead to the development of autoimmune diseases such as systemic sclerosis.[179,180]

5.3 Mitochondrial Diseases

Mitochondria are intracellular organelles specializing in the production of cellular energy in the form of adenosine triphosphate (ATP) by the process of oxidative phosphorylation (OXPHOS). Most cells contain hundreds of mitochondria. Mitochondria possess their own genome which is a circular DNA molecule of approximately 16 500 base pairs.[181] Mitochondrial DNA carries genes for 13 polypeptides, all involved in OXPHOS, 22 transfer RNA (tRNA) genes and genes for the 12S and 16S rRNA. Mitochondrial DNA is transmitted through the cytoplasm of the oocyte at fertilization and thus is maternally inherited.[182] In addition, as there are multiple copies of mitochondria, each containing multiple copies of mitochondrial DNA, a cell can contain a mixture of wild-type and mutant mitochondrial DNA, a condition known as heteroplasmy. The stochastic and quantitative nature of mitochondrial genetics implies that the inheritance and expression of heteroplasmic mutations are highly variable. An additional complexity of the genetics of mitochondrial disorders is that mutations in *nuclear* genes can also affect OXPHOS[183–185] or the stability of mitochondrial DNA.[186,187] In these circumstances, the disorder can be inherited in a Mendelian manner. These features highlight the great care that is needed in counseling families with mitochondrial disorders.

Over the last 10 years, molecular alterations in mitochondrial DNA have been associated with several human pathologies. The first mitochondrial diseases to be elucidated on the molecular level include Leber's hereditary optic neuropathy (LHON), a sudden-onset blindness resulting from a missense mutation of the mitochondrial DNA,[188] and a sporadically occurring group of neurological diseases called chronic progressive external ophthalmoplegia (CPEO) and the Kearns–Sayre Syndrome (KSS), resulting from mitochondrial DNA deletions.[189] Since these discoveries, over 50 pathogenic mitochondrial DNA base substitution mutations and hundreds of mitochondrial DNA rearrangements have been identified in many neurodegenerative diseases (summarized in the website MITOMAP, www.gen.emory.edu/mitomap.html).

Apart from the relatively rare diseases mentioned above, some authors have suggested that mitochondrial dysfunction may underlie common processes such as aging and malignancy. The potential relationship between mitochondrial dysfunction and aging is supported by reports of age-related declines in OXPHOS enzyme activities[190,191] in skeletal muscle, liver, and brain and the associated accumulation of somatic mitochondrial DNA alterations in these same postmitotic tissues.[192–195]

With regard to cancer, multiple somatic mitochondrial DNA mutations, including intragenic deletions, missense and chain-termination point mutations, and alterations of homopolymeric sequences have been reported.[196–198] In addition, alterations in mitochondrial copy numbers have been described in a number of malignancies.[199] The exact relationship between these mitochondrial DNA alterations and malignancy is unclear at present although possible mechanisms include alterations in cellular energy dynamics, mitochondrial oxidative stress and modulation of apoptosis.

The increasing spectrum of disorders related to mitochondrial dysfunction has stimulated interest in the

development of rapid methods for characterizing mitochondrial DNA alterations.[200–202]

5.4 Diabetes Mellitus

Diabetes mellitus affects approximately 5% of the general population, with inter-ethnic and inter-geographical variations. The majority of cases are accounted for by two different types of diabetes mellitus, type 1 and type 2, which account for approximately 10% and 90% of cases, respectively. These two types of diabetes mellitus have distinct etiologies. Type 1 is typically an autoimmune disease in which β-cell destruction is followed by absolute deficiency of insulin. Type 2 diabetes, on the other hand, is characterized by a deficiency of insulin action as a result of a combination of insulin resistance and β-cell dysfunction that is manifested as inadequate insulin secretion in the face of insulin resistance and hyperglycemia. Genetic factors play an important part in both type 1 and type 2 diabetes, with concordance rates among monozygotic twins of 25–50% and 50–95%, respectively.

The recent advances in molecular genetics have allowed us to define the genetic factors associated with diabetes mellitus in more detail. Two approaches have been used to identify genes that predispose to diabetes mellitus: first, by testing candidate genes via the comparison of the frequency of alleles of specific genes in diabetic and control populations (case-control or association studies); and second, by genome scanning to identify chromosomal loci associated with disease susceptibility.[203,204] As type 1 diabetes is an autoimmune disease, one of the first candidate genomic regions to be examined was the human leukocyte antigen (HLA) locus. This locus is now referred to as *IDDM1* and has been estimated to account for 35–40% of the inheritance of the disease.[203,205,206] Detailed molecular studies have identified particular alleles of class II HLA genes that confer susceptibility or protection for type 1 diabetes.[207] Apart from the HLA region, a second susceptibility locus that has been identified using the candidate gene approach is the insulin gene locus on chromosome 11 (referred to as *IDDM2*). Amongst the polymorphisms in this region is the variable number of tandem repeats in the 5′-flanking region of the insulin gene. It has been estimated that the *IDDM2* locus accounts for approximately 10% of the familial inheritance of type 1 diabetes.[203,205,206] Further susceptibility loci for type 1 diabetes have been identified by whole genome scans. Detailed characterization of these candidate loci is still in progress.

Type 2 diabetes is a genetically heterogenous disease, with mutations in a number of genes or combination of genes contributing to the disease phenotype. Further complexity in the study of the genetics of type 2 diabetes is that different combinations of genes may be responsible for the disease in different populations. In addition, environmental factors, many of which are still ill-defined, are important for the pathogenesis of type 2 diabetes. A further complication is that in many instances type 2 diabetes is a late-onset disease. Consequently, for many studied families, the parents, especially those who are affected by the disorder, may have succumbed to complications of diabetes and thus may not be available for study. Similarly, many of the younger members of the studied family, even if they have inherited the diabetogenic genes, may not be old enough to develop the disorder. Because of these difficulties, many investigators have focused on more easily defined but rarer groups of type 2 diabetes. One example is type 2 diabetes which is caused by alterations of the mitochondrial genome.[208] A second example is a form of diabetes called maturity-onset diabetes of the young (MODY). For genetic study, MODY has the advantage that it is monogenic, with an autosomal dominant mode of inheritance and has an early age of onset, typically before 25 years. Detailed molecular genetic studies have recently identified a number of MODY genes including the glucokinase gene (MODY2),[209] hepatocyte nuclear factor-4α (MODY1),[210] hepatocyte nuclear factor-1α (MODY3)[211] and the insulin promoter factor gene (MODY4).[212] It is expected that further MODY genes will be discovered as mutations in these four genes do not account for all cases of MODY. However, the elucidation of at least a proportion of the genetic lesions accounting for MODY does allow for family screening in some affected kindreds. This development may have implications for the prompt institution of preventive measures, such as exercise, diet, weight control, as well as early treatment of symptomatic individuals. Apart from these genetic alterations, rare mutations/polymorphisms of the insulin gene,[213] insulin receptor gene,[214] glucagon receptor gene,[215] amylin gene[216] and insulin receptor substrate 1 gene[217–219] have also been described, with variable effects on glucose homeostasis. The significance of these mutations to "common" late-onset type 2 diabetes, however, remains unclear. Similar to type 1 diabetes, whole genome scans have also been attempted for type 2 diabetes, with a susceptibility locus located on chromosome 2, referred to as *NIDDM1*.[220] The difficulty of this approach is increased by the fact that different genetic factors may be predominant in different populations and thus genomic regions showing positive association with diabetes in a particular population may not exhibit the same relationship in other populations.[221] It is expected that further data will be forthcoming over the next few years to allow a more precise definition of the molecular genetic factors involved in type 2 diabetes.

5.5 Molecular Analysis of Cell-free DNA in Plasma and Serum

It has been known for a number of years that cell-free DNA can be found in the plasma and serum of human subjects.[222,223] Of particular interest, the level of plasma DNA is elevated in individuals with cancer,[224] thromboembolism[225] and systemic lupus erythematosus.[226] Recent molecular characterization indicates that a proportion of the circulating DNA in cancer patients possesses tumor-associated genetic[227–230] or epigenetic[231,232] alterations or neoplasia-associated viral nucleic acids.[233] These data suggest that plasma DNA analysis may be a useful method for the non-invasive detection or monitoring of a number of cancers.

In addition to tumor-derived DNA, other types of cell-free DNA species have also been detected in the plasma/serum. Thus, cell-free fetal DNA has recently been found in the plasma of pregnant women.[234] This discovery has opened up a new approach for noninvasive prenatal diagnosis.[235,236] In addition, abnormal levels of circulating fetal DNA have been found in a number of pregnancy-associated disorders, including preeclampsia[237] and preterm labor.[238] In addition to fetal DNA, cell-free DNA derived from the graft has also been demonstrated in the plasma of liver and kidney transplant recipients.[239] It is likely that other types of circulating nucleic acids will be found in the circulation and that plasma- or serum-based molecular diagnostics will find an increasing number of potential clinical applications.

5.6 The Human Genome Project

The HGP was originally conceived in the mid-1980s as an ambitious effort to characterize the human genome, yielding ultimately the DNA sequence of the human genome, with a projected completion date of 2005. The project represents an enormous undertaking, with the human genome consisting of some 3 billion base pairs, in which coding sequences for 50 000 to 80 000 genes are distributed. At its inception, the HGP was expected to be a project costing some US$ 3 billion over a 15-year period. The project has now evolved into a collaborative effort with input from many countries.

The goals of the project, as initially conceived, were threefold: (1) the creation of genetic maps; (2) the generation of physical maps; and (3) the determination of the entire DNA sequence of the human genome. The first two goals can be regarded as preliminary characterizing efforts, paving the way for the third goal. The genetic and physical maps will provide a series of landmarks for directing sequencing efforts at particular genomic regions and, as of 1999, these first two goals have been completed.[240] Efforts are now focused on the third goal, with a ''working draft'' of the human sequence to be produced by the end of 2001, with a targeted completion date of the entire sequencing project to be 2003, 2 years ahead of the previous projections.[240] Additional goals have now been set, including the development of new sequencing technologies, the promotion of the new science of functional genomics, the studying of human genome sequence variation and the sequencing of other genomes. The potential usefulness of the latter goal has been illustrated by the completion of the genomic sequencing of *Caenorhabditis elegans*, the first completed genome sequence for a multicellular organism.[241]

Implicit in the HGP is the idea that an improved understanding of the human genome will be beneficial to our understanding of the fundamental workings of the human body in health and disease. These developments are likely to improve our ability to diagnose diseases more accurately and precisely and in some cases even before clinical signs of disease manifest themselves. For example, individuals known to possess genes that predispose to cancer or diabetes may be followed up intensively under special programs. Ultimately, it is hoped that these developments will change today's ''treatment-oriented'' health care to one that is more ''prevention-oriented'', by the stratification of individuals with respect to their genetic susceptibility to diseases.[242] The HGP may also catalyze the discovery of new genetic pathways which may be amenable to therapeutic intervention.

The HGP raises numerous ethical and social issues. It also generates concerns regarding the confidentiality of one's own genetic information, the proper institution of genetic counseling prior to DNA testing and potential problems regarding genetic discrimination either at work or by insurance companies. For these reasons, technical advances brought by the HGP should be accompanied by improvement in our understanding regarding the ethical, legal and social implications of these new data. In this regard programs such as the ''Ethical, Legal, and Social Implications'' (ELSI) working group have been set up to address these issues.

ACKNOWLEDGMENTS

We thank Joanne Lee for help during the preparation of this manuscript. Y.M.D.L. is supported by the Hong Kong Research Grants Council and the Industrial Support Fund.

ABBREVIATIONS AND ACRONYMS

ARMS	Amplification Refractory Mutation System
ASO	Allele-specific Oligonucleotide
ATP	Adenosine Triphosphate
bDNA	Branched DNA
CDGE	Constant Denaturant Gel Electrophoresis
cDNA	Copy DNA
CGH	Comparative Genomic Hybridization
CMC	Chemical Mismatch Cleavage
CPEO	Chronic Progressive External Ophthalmoplegia
DGGE	Denaturing Gradient Gel Electrophoresis
DHPLC	Denaturing High-performance Liquid Chromatography
DOPPCR	Degenerate Oligonucleotide Primed Polymerase Chain Reaction
ELSI	Ethical, Legal, and Social Implications
EMC	Enzyme Mismatch Cleavage
EMD	Enzymatic Mutation Detection
EST	Expressed Sequence Tag
FISH	Fluorescence In Situ Hybridization
HA	Heteroduplex Analysis
HBV	Hepatitis B Virus
HGP	Human Genome Project
HIV-1	Human Immunodeficiency Virus Type 1
HLA	Human Leukocyte Antigen
HPV	Human Papillomavirus
KSS	Kearns–Sayre Syndrome
LCR	Ligase Chain Reaction
LHON	Leber's Hereditary Optic Neuropathy
LOH	Loss of Heterozygosity
LOI	Loss of Imprinting
Mb	Megabases
MODY	Maturity-onset Diabetes of the Young
mRNA	Messenger RNA
NASBA	Nucleic Acid Sequence-based Amplification
OXPHOS	Oxidative Phosphorylation
PCR	Polymerase Chain Reaction
PRINS	Primed In Situ DNA Synthesis
rRNA	Ribosomal RNA
RTPCR	Reverse-transcription Polymerase Chain Reaction
SDA	Strand Displacement Amplification
SKY	Spectral Karyotyping
SNP	Single-nucleotide Polymorphism
SSCP	Single-strand Conformation Polymorphism
Taq	*Thermus aquaticus*
TGGE	Temperature Gradient Gel Electrophoresis
tRNA	Transfer RNA
Tth	*Thermus thermophilus*
UV	Ultraviolet
3SR	Self-sustained Sequence Replication

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Nucleic Acid Analysis in Clinical Chemistry

Forensic Science **(Volume 5)**
DNA Extraction Methods in Forensic Analysis • Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Capillary Electrophoresis of Nucleic Acids • Comparative Genomics: Differential Display and Subtractive Hybridization • DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Electron Tomography of Chromosome Structure • Fluorescence In Situ Hybridization • Genome Physical Mapping Using BACs • Mass Spectrometry of Nucleic Acids • Nuclear Magnetic Resonance and Nucleic Acid Structures • Nucleic Acid Structural Energetics • Optical Mapping in Genomic Analysis • PNA and Its Applications • Polymerase Chain Reaction and Other Amplification Systems • Radiation Hybrid Mapping • Restriction Landmark Genomic and cDNA Scanning • RNA Tertiary Structure • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Sequencing Strategies and Tactics in DNA and RNA Analysis • Structural Analysis of Ribozymes • X-ray Structures of Nucleic Acids

REFERENCES

1. M. McCarty, *The Transforming Principle*, W.W. Norton, New York, 1985.
2. J.D. Watson, F.H.C. Crick, 'A Structure for Deoxyribonucleic Acid', *Nature*, **171**, 737 (1953).
3. E.M. Southern, 'Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis', *J. Mol. Biol.*, **98**, 503–517 (1975).
4. J.S. Wainscoat, A.E. Kulozik, M. Ramsay, A.G. Falusi, D.J. Weatherall, 'A TaqI Gamma-globin DNA Polymorphism: an African-specific Marker', *Hum. Genet.*, **74**, 90–92 (1986).
5. F. Sanger, S. Nicklen, A.R. Coulson, 'DNA Sequencing with Chain-terminating Inhibitors', *Proc. Natl. Acad. Sci., USA*, **74**, 5463–5467 (1977).

6. A.M. Maxam, W. Gilbert, 'A New Method for Sequencing DNA', *Proc. Natl. Acad. Sci., USA*, **74**, 560–564 (1977).
7. R.K. Wilson, C. Chen, N. Avdalovic, J. Burns, L. Hood, 'Development of an Automated Procedure for Fluorescent DNA Sequencing', *Genomics*, **6**, 626–634 (1990).
8. Y.W. Kan, A.M. Dozy, 'Polymorphism of DNA Sequence Adjacent to Human Beta-globin Structural Gene: Relationship to Sickle Mutation', *Proc. Natl. Acad. Sci., USA*, **75**, 5631–5635 (1978).
9. J.F. Gusella, N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M.R. Wallace, A.Y. Sakaguchi, 'A Polymorphic DNA Marker Genetically Linked to Huntington's Disease', *Nature*, **306**, 234–238 (1983).
10. S.T. Reeders, M.H. Breuning, K.E. Davies, R.D. Nicholls, A.P. Jarman, D.R. Higgs, P.L. Pearson, D.J. Weatherall, 'A Highly Polymorphic DNA Marker Linked to Adult Polycystic Kidney Disease on Chromosome 16', *Nature*, **317**, 542–544 (1985).
11. Y.W. Kan, M.S. Golbus, A.M. Dozy, 'Prenatal Diagnosis of Alpha-thalassemia. Clinical Application of Molecular Hybridization', *N. Engl. J. Med.*, **295**, 1165–1167 (1976).
12. R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, 'Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia', *Science*, **230**, 1350–1354 (1985).
13. R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich, 'Primer-directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase', *Science*, **239**, 487–491 (1988).
14. H. Li, U.B. Gyllensten, X. Cui, R.K. Saiki, H.A. Erlich, N. Arnheim, 'Amplification and Analysis of DNA Sequences in Single Human Sperm and Diploid Cells', *Nature*, **335**, 414–417 (1988).
15. A.H. Handyside, J.K. Pattinson, R.J.A. Penketh, J.D.A. Delhanty, R.M.L. Winston, E.G.D. Tuddenham, 'Biopsy of Human Preimplantation Embryos and Sexing by DNA Amplification', *Lancet, i*, 347–349 (1989).
16. J.L. Weber, P.E. May, 'Abundant Class of Human DNA Polymorphisms Which can be Typed Using the Polymerase Chain Reaction', *Am. J. Hum. Genet.*, **44**, 388–396 (1989).
17. M. Litt, J.A. Luty, 'A Hypervariable Microsatellite Revealed by In Vitro Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene', *Am. J. Hum. Genet.*, **44**, 397–401 (1989).
18. P. Chomczynski, N. Sacchi, 'Single-step Method of RNA Isolation by Acid Guanidinium Thiocyanate–Phenol–Chloroform Extraction', *Anal. Biochem.*, **162**, 156–159 (1987).
19. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: a Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, New York, 1989.
20. P. Chomczynski, 'A Reagent for the Single-step Simultaneous Isolation of RNA, DNA and Proteins from Cell and Tissue Samples', *Biotechniques*, **15**, 532–534 (1993).
21. R.L. McDonald, B.R. Rose, J. Gibbins, P.J. Baird, 'Use of the same Archival Papanicolanou Smears for Detection of Human Papillomavirus by Cytology and Polymerase Chain Reaction', *Diagn. Mol. Pathol.*, **8**, 20–25 (1999).
22. W.M. Barnes, 'PCR Amplification of up to 35 kb DNA with High Fidelity and High Yield from Lamda Bacteriophage Templates', *Proc. Natl. Acad. Sci., USA*, **91**, 2216–2220 (1994).
23. S. Cheng, S.Y. Chang, P. Gravitt, R. Respess, 'Long PCR', *Nature*, **369**, 684–685 (1994).
24. P. Teynier, M.F. Chretien, F. Savagner, G. Larcher, V. Rohmer, P. Barriere, Y. Malthiery, 'Long PCR Analysis of Human Gamete mtDNA Suggests Defective Mitochondrial Maintenance in Spermatozoa and Supports the Bottleneck Theory for Oocytes', *Biochem. Biophys. Res. Commun.*, **252**, 373–377 (1998).
25. C.D. O'Connell, A. Juhasz, C. Kuo, D.J. Reeder, D.S.B. Hoon, 'Detection of Tyrosinase mRNA in Melanoma by Reverse Transcription-PCR and Electrochemiluminescence', *Clin. Chem.*, **44**, 1161–1169 (1998).
26. T.W. Myers, D.H. Gelfand, 'Reverse Transcription and DNA Amplification by a *Thermus thermophilus* DNA Polymerase', *Biochemistry*, **30**, 7661–7666 (1991).
27. K.K. Young, R.M. Resnick, T.W. Myers, 'Detection of Hepatitis C Virus RNA by a Combined Reverse Transcription–Polymerase Chain Reaction Assay', *J. Clin. Microbiol.*, **31**, 882–886 (1993).
28. F.X. Limbach, B. Jaulhac, Y. Piemont, J.L. Kuntz, H. Monteil, J. Sibilia, 'One-step Reverse Transcriptase PCR Method for Detection of *Borrelia burgdorferi* mRNA in Mouse Lyme Arthritis Tissue Samples', *J. Clin. Microbiol.*, **37**, 2037–2039 (1999).
29. G. Ruano, K.K. Kidd, J.C. Stephens, 'Haplotype of Multiple Polymorphisms Resolved by Enzymatic Amplification of Single DNA Molecules', *Proc. Natl. Acad. Sci., USA*, **87**, 6296–6300 (1990).
30. B. Vogelstein, K.W. Kinzler, 'Digital PCR', *Proc. Natl. Acad. Sci., USA*, **96**, 9236–9241 (1999).
31. Y.M.D. Lo, W.Z. Mehal, K.A. Fleming, 'False-positive Results and the Polymerase Chain Reaction', *Lancet, ii*, 679 (1988).
32. S. Kwok, R. Higuchi, 'Avoiding False Positives with PCR', *Nature*, **339**, 237–238 (1989).
33. J. Loeffler, H. Hebart, R. Bialek, L. Hagmeyer, D. Schmidt, F.P. Serey, M. Hartmann, J. Eucker, J. Einsele, 'Contaminations Occurring in Fungal PCR Assays', *J. Clin. Microbiol.*, **37**, 1200–1202 (1999).

34. R.W. Cone, A.C. Hobson, M.L. Huang, M.R. Fairfax, 'Polymerase Chain Reaction Decontamination: the Wipe Test', *Lancet*, **336**, 686–687 (1990).
35. M.C. Longo, M.S. Berninger, J.L. Hartley, 'Use of Uracil DNA Glycosylase to Control Carry-over Contamination in Polymerase Chain Reactions', *Gene*, **93**, 125–128 (1990).
36. G.D. Cimino, K.C. Metchette, J.W. Treeman, J.F. Hearst, S.T. Isaacs, 'Post-PCR Sterilization: a Method to Control Carryover Contamination for the Polymerase Chain Reaction', *Nucleic Acids Res.*, **19**, 99–107 (1991).
37. G. Sarkar, S.S. Sommer, 'Shedding Light on PCR Contamination', *Nature*, **343**, 27 (1990).
38. F. Barany, 'The Ligase Chain Reaction in a PCR World', *PCR Methods Appl.*, **1**, 5–16 (1991).
39. A. Lindbrathen, P. Gaustad, B. Hovig, T. Tonjum, 'Direct Detection of Mycobacterium Tuberculosis Complex in Clinical Samples from Patients in Norway by Ligase Chain Reaction', *J. Clin. Microbiol.*, **35**, 3248–3253 (1997).
40. C. Bourgeois, N. Sixt, J.B. Bour, P. Pothier, 'Value of a Ligase Chain Reaction Assay for Detection of Ganciclovir Resistance-related Mutation 594 in UL97 Gene of Human Cytomegalovirus', *J. Virol. Meth.*, **67**, 167–175 (1997).
41. E. Fahy, D.Y. Kwoh, T.R. Gingeras, 'Self-sustained Sequence Replication (3SR): an Isothermal Transcription-based Amplification System', *PCR Methods Appl.*, **1**, 25–33 (1991).
42. A.M. Vandamme, S. Van Dooren, W. Kok, P. Goubau, K. Fransen, T. Kievits, J.C. Schmit, E. De Clercq, 'Detection of HIV-1 RNA in Plasma and Serum Samples Using the NASBA Amplification System Compared to RNA-PCR', *J. Virol. Meth.*, **52**, 121–132 (1995).
43. G.T. Walker, M.C. Little, J.G. Nadeau, D.D. Shank, 'Isothermal In Vitro Amplification of DNA by a Restriction Enzyme/DNA Polymerase System', *Proc. Natl. Acad. Sci., USA*, **89**, 392–396 (1992).
44. G.T. Walker, J.G. Nadeau, C.P. Linn, R.F. Devlin, W.B. Dandliker, 'Strand Displacement Amplification (SDA) and Transient-state Fluorescence Polarization Detection of Mycobacterium Tuberculosis DNA', *Clin. Chem.*, **42**, 9–13 (1996).
45. M.C. Little, J. Andrews, R. Moore, S. Bustos, L. Jones, C. Embres, G. Durmowicz, H.J.D. Berger, K. Yanson, C. Rostkowski, D. Yursis, J. Price, T. Fort, A. Walters, M. Collis, O. Llorin, J. Wood, F. Failing, C. O'Keefe, B. Scrivens, B. Pope, T. Hansen, K. Marino, K. Williams, M. Boenisch, 'Strand Displacement Amplification and Homogeneous Real-time Detection Incorporated in a Second-generation DNA Probe System, BDProbe Tec™ET', *Clin. Chem.*, **45**, 777–784 (1999).
46. G.T. Walker, M.S. Fraiser, J.L. Schram, M.C. Little, J.G. Nadeau, D.P. Malinowski, 'Strand Displacement Amplification – an Isothermal, In Vitro DNA Amplification Technique', *Nucleic Acids Res.*, **20**, 1691–1696 (1992).
47. J.S. Shah, J. Liu, J. Smith, S. Popoff, G. Radcliffe, W.J. O'Brien, G. Serpe, D.M. Olive, W. King, 'Novel, Ultrasensitive, Q-beta Replicase-amplified Hybridization Assay for Detection of *Chlamydia trachomatis*', *J. Clin. Microbiol.*, **32**, 2718–2724 (1994).
48. C. Dobkin, D.R. Mills, F.R. Kramer, S. Spiegelman, 'RNA Replication: Required Intermediates and the Dissociation of Template, Product, and Q-beta Replicase', *Biochemistry*, **18**, 2038–2044 (1979).
49. C. Pachl, J.A. Todd, D.G. Kern, P.J. Sheridan, S.J. Fong, M. Stempien, B. Hoo, D. Besemer, T. Yeghiazarian, B. Irvine, J. Kolberg, R. Kokka, P. Neuwald, M.S. Urdea, 'Rapid and Precise Quantification of HIV-1 RNA in Plasma Using a Branched DNA Signal Amplification Assay', *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.*, **8**, 446–454 (1995).
50. F.S. Nolte, 'Branched DNA Signal Amplification for Direct Quantitation of Nucleic Acid Sequences in Clinical Specimens', *Adv. Clin. Chem.*, **33**, 201–235 (1999).
51. R.D. Nargessi, N.F. Khabbaz, X.M. Xu, M. Zamroud, J. Kolberg, M.L. Collins, 'Quantitation of Estrogen Receptor mRNA in Breast Carcinoma by Branched DNA Assay', *Breast Cancer Res. Treat.*, **50**, 47–55 (1998).
52. L.P. Shen, P. Sheridan, W.W. Cao, P.J. Dailey, J.F. Salazar-Gonzalez, E.C. Breen, J.L. Fahey, M.S. Urdea, J.A. Kolberg, 'Quantification of Cytokine mRNA in Peripheral Blood Mononuclear Cells Using Branched DNA (bDNA) Technology', *J. Immunol. Methods*, **215**, 123–134 (1998).
53. D. Kern, M. Collins, T. Fultz, J. Detmer, S. Hamren, J.J. Peterkin, P. Sheridan, M. Urdea, R. White, T. Yeghiazarian, J. Todd, 'An Enhanced-sensitivity Branched-DNA Assay for Quantification of Human Immunodeficiency Virus Type 1 RNA in Plasma', *J. Clin. Microbiol.*, **34**, 3196–3202 (1996).
54. I.H.N. Wong, T. Leung, S. Ho, W.Y. Lau, M. Chan, P.J. Johnson, 'Semiquantitation of Circulating Hepatocellular Carcinoma Cells by Reverse Transcriptase Polymerase Chain Reaction', *Br. J. Cancer*, **76**, 628–633 (1997).
55. I.H.N. Wong, M.H.L. Ng, J.C.K. Lee, K.W. Lo, Y.F. Chung, D.P. Huang, 'Transcriptional Silencing of the *p16* Gene by Hypermethylation in Human Myeloma-derived Cell Lines', *Br. J. Haematol.*, **103**, 168–175 (1998).
56. M. Setoyama, F.A. Kerdel, G. Elgart, T. Kanzaki, J.J. Byrnes, 'Detection of HTLV-1 by Polymerase Chain Reaction In Situ Hybridization in Adult T-cell Leukemia/Lymphoma', *Am. J. Pathol.*, **152**, 683–689 (1998).
57. A. Abati, J.S. Sanford, P. Fetsch, F.M. Marincola, S.R. Wolman, 'Fluorescence In Situ Hybridization (FISH): a User's Guide to Optimal Preparation of Cytologic Specimens', *Diagn. Cytopathol.*, **13**, 486–492 (1995).

58. A. Haegmeijer, A. de Klein, J. Wijsman, E. van Meerten, G.E. de Greef, N. Sacchi, 'Development of an Interphase Fluorescent In Situ Hybridization (FISH) Test to Detect t(8;21) in AML Patients', *Leukemia*, **12**, 96–101 (1998).
59. M.H.L. Ng, A. Kan, Y.F. Chung, I.H.N. Wong, K.W. Lo, N.W. Wickham, K.I. Lei, J.C.K. Lee, 'Combined Morphological and Interphase Fluorescence In Situ Hybridization Study in Multiple Myeloma of Chinese Patients', *Am. J. Pathol.*, **154**, 15–22 (1999).
60. K. Wakui, M. Tanemura, K. Suzumori, E. Hidaka, M. Ishikawa, T. Kubota, Y. Fukushima, 'Clinical Applications of Two-color Telomeric Fluorescence In Situ Hybridization for Prenatal Diagnosis: Identification of Chromosomal Translocation in Five Families with Recurrent Miscarriages or a Child with Multiple Congenital Anomalies', *J. Hum. Genet.*, **44**, 85–90 (1999).
61. L. Wilkens, J. Tchinda, P. Komminoth, M. Werner, 'Single- and Double-color Oligonucleotide Primed In Situ Labeling (PRINS): Applications in Pathology', *Histochem. Cell. Biol.*, **108**, 439–446 (1997).
62. J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J.S. Mihatsch, O.P. Kallioniemi, 'Tissue Microarrays for High-throughput Molecular Profiling of Tumor Specimens', *Nat. Med.*, **4**, 844–847 (1998).
63. S.M. Mohaddes, E. Boyd, A. Morris, N. Morrison, J.M. Connor, 'A Practical Strategy for Detection of Major Chromosome Aneuploidies Using Ratio-mixing Fluorescence In Situ Hybridization', *Mol. Cell. Probes*, **10**, 147–154 (1996).
64. S.K. Banerjee, A.P. Weston, D.L. Persons, D.R. Campbell, 'Quick-FISH: a Rapid Fluorescence In Situ Hybridization Technique for Molecular Cytogenetic Analysis', *Biotechniques*, **24**, 826–830 (1998).
65. E. Schrock, S. du Manoir, T. Veldman, B. Schoell, J. Wienberg, A. Ferguson-Smith, Y. Ning, D.H. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini, T. Ried, 'Multicolor Spectral Karyotyping of Human Chromosomes', *Science*, **273**, 494–497 (1996).
66. M. Macville, E. Schrock, H. Padilla-Nash, C. Keck, B.M. Ghadimi, D. Zimonjic, N. Popescu, T. Ried, 'Comprehensive and Definitive Molecular Cytogenetic Characterization of HeLa Cells by Spectral Karyotyping', *Cancer Res.*, **59**, 141–150 (1999).
67. M. Liyanage, A. Coleman, S. du Manoir, T. Veldman, S. McCormack, R.B. Dickson, C. Barlow, A. Wynshaw Boris, S. Janz, J. Wienberg, M.A. Ferguson-Smith, E. Schrock, T. Ried, 'Multicolor Spectral Karyotyping of Mouse Chromosomes', *Nat. Genet.*, **14**, 312–315 (1996).
68. S. Knuutila, A.M. Bjorkqvist, K. Autio, M. Tarkkanen, M. Wolf, O. Monni, J. Szymanska, M.L. Larramendy, J. Tapper, H. Pere, W. El-Rifai, S. Hemmer, V.M. Wasenius, V. Vidgren, Y. Zhu, 'DNA Copy Number Amplifications in Human Neoplasms: Review of Comparative Genomic Hybridization Studies', *Am. J. Pathol.*, **152**, 1107–1123 (1998).
69. J.C. Alers, J. Rochart, P.J. Krijtenburg, H. van Dekken, A.K. Raap, C. Rosenberg, 'Universal Linkage System: an Improved Method for Labeling Archival DNA for Comparative Genomic Hybridization', *Genes Chromosomes Cancer*, **25**, 301–305 (1999).
70. R. Karhu, J. Rummukainen, T. Lorch, J. Isola, 'Four-color CGH: a New Method for Quality Control of Comparative Genomic Hybridization', *Genes Chromosomes Cancer*, **24**, 112–118 (1999).
71. D.K. Griffin, D. Sanoudou, E. Adamski, C. McGiffert, P. O'Brien, J. Wienberg, M.A. Ferguson-Smith, 'Chromosome Specific Comparative Genome Hybridisation for Determining the Origin of Intrachromosomal Duplications', *J. Med. Genet.*, **35**, 37–41 (1998).
72. V.C. Sheffield, D.R. Cox, L.S. Lerman, R.M. Myers, 'Attachment of a 40-Base-pair G+C-rich Sequence (GC-clamp) to Genomic DNA Fragments by the Polymerase Chain Reaction Results in Improved Detection of Single-base Changes', *Proc. Natl. Acad. Sci., USA*, **86**, 232–236 (1989).
73. Y. Wu, V.M. Hayes, J. Osinga, I.M. Mulder, M.W. Looman, C.H. Buys, R.M. Hofstra, 'Improvement of Fragment and Primer Selection for Mutation Detection by Denaturing Gradient Gel Electrophoresis', *Nucleic Acids Res.*, **26**, 5432–5440 (1998).
74. R.M. Wartell, S. Hosseini, S. Powell, J. Zhu, 'Detecting Single Base Substitutions, Mismatches and Bulges in DNA by Temperature Gradient Gel Electrophoresis and Related Methods', *J. Chromatogr. A*, **806**, 169–185 (1998).
75. A.L. Borresen, E. Hovig, B. Smith-Sorensen, D. Malkin, S. Lystad, T.I. Andersen, J.M. Nesland, K.J. Isselbacher, S.H. Friend, 'Constant Denaturant Gel Electrophoresis as a Rapid Screening Technique for p53 Mutations', *Proc. Natl. Acad. Sci., USA*, **88**, 8405–8409 (1991).
76. Y. Wu, R.P. Stupl, P. Elfferich, J. Osinga, C.H. Buys, R.M. Hofstra, 'Improved Mutation Detection in GC-rich DNA Fragments by Combined DGGE and CDGE', *Nucleic Acids Res.*, **27**, e9 (1999).
77. M.B. White, M. Carvalho, D. Derse, S.J. O'Brien, M. Dean, 'Detecting Single Base Substitution as Heteroduplex Polymorphisms', *Genomics*, **12**, 301–306 (1992).
78. M. Orita, Y. Suzuki, T. Sukiya, K. Hayashi, 'Rapid and Sensitive Detection of Point Mutations and DNA Polymorphisms Using the Polymerase Chain Reaction', *Genomics*, **5**, 874–879 (1989).
79. T.M.U. Wagner, R.A. Moslinger, D. Muhr, G. Langbauer, K. Hirtenlehner, H. Concin, W. Doeller, A. Haid, A.H. Lang, P. Mayer, E. Ropp, E. Kubista, B. Amirimani, T. Helbich, A. Becherer, O. Scheiner, H. Breiteneder, A. Borg, P. Bevilee, P. Oefner, C. Zielinski, 'BRCA1-related Breast Cancer in Austrian Breast and Ovarian Cancer Families: Specific *BRCA1* Mutations and Pathological Characteristics', *Int. J. Cancer*, **77**, 354–360 (1998).

80. W. Liu, D.I. Smith, K.J. Rechtzigel, S.N. Thibodeau, C.D. James, 'Denaturing High Performance Liquid Chromatography (DHPLC) Used in the Detection of Germline and Somatic Mutations', *Nucleic Acids Res.*, **26**, 1396–1400 (1998).
81. D. Glavac, M. Dean, 'Optimization of the Single-strand Conformation Polymorphism (SSCP) Technique for Detection of Point Mutations', *Hum. Mutat.*, **2**, 404–414 (1993).
82. V.C. Sheffield, J.S. Beck, A.E. Kwitek, D.W. Sandstrom, E.M. Stone, 'The Sensitivity of Single-strand Conformation Polymorphism Analysis for the Detection of Single Base Substitutions', *Genomics*, **16**, 325–332 (1993).
83. K. Hayashi, D.W. Yandell, 'How Sensitive is PCR-SSCP?', *Hum. Mutat.*, **2**, 338–346 (1993).
84. Y.M.D. Lo, P. Patel, W.Z. Mehal, K.A. Fleming, J.I. Bell, J.S. Wainscoat, 'Analysis of Complex Genetic Systems by ARMS-SSCP: Application to HLA Genotyping', *Nucleic Acids Res.*, **20**, 1005–1009 (1992).
85. R. Makino, H. Yazyu, Y. Kishimoto, T. Sekiya, K. Hayashi, 'F-SSCP: Fluorescence-based Polymerase Chain Reaction–Single-strand Conformation Polymorphism (PCR-SSCP) Analysis', *PCR Methods Appl.*, **2**, 10–13 (1992).
86. J. Ellison, M. Dean, D. Goldman, 'Efficacy of Fluorescence-based PCR-SSCP for Detection of Point Mutations', *Biotechniques*, **15**, 684–691 (1993).
87. G.R. Taylor, 'Enzymatic and Chemical Cleavage Methods', *Electrophoresis*, **20**, 1125–1130 (1999).
88. R.G. Cotton, N.R. Rodrigues, R.D. Campbell, 'Reactivity of Cytosine and Thymine in Single-base-pair Mismatches with Hydroxylamine and Osmium Tetroxide and its Application to the Study of Mutations', *Proc. Natl. Acad. Sci., USA*, **85**, 4397–4401 (1988).
89. T.P. Ellis, K.E. Humphrey, M.J. Smith, R.G. Cotton, 'Chemical Cleavage of Mismatch: a New Look at an Established Method', *Hum. Mutat.*, **11**, 345–353 (1998).
90. T.E. Shenk, C. Rhodes, P.W. Rigby, P. Berg, 'Biochemical Method for Mapping Mutational Alterations in DNA with S1 Nuclease: the Location of Deletions and Temperature-sensitive Mutations in Simian Virus 40', *Proc. Natl. Acad. Sci., USA*, **72**, 989–993 (1975).
91. looseness=1 R.D. Mashal, J. Koontz, J. Sklar, 'Detection of Mutations by Cleavage of DNA Heteroduplexes with Bacteriophage Resolvases', *Nat. Genet.*, **9**, 177–183 (1995).
92. R. Youil, B.W. Kemper, R.G. Cotton, 'Screening for Mutations by Enzyme Mismatch Cleavage with T4 Endonuclease VII', *Proc. Natl. Acad. Sci., USA*, **92**, 87–91 (1995).
93. J.J. Babon, M. McKenzie, R.G. Cotton, 'Mutation Detection Using Fluorescent Enzyme Mismatch Cleavage with T4 Endonuclease VII', *Electrophoresis*, **20**, 1162–1170 (1999).
94. B.J.J. Del Tito, H.E. Poff, III, M.A. Novotny, D.M. Cartledge, R.I. Walker, II, C.D. Earl, A.L. Bailey, 'Automated Fluorescent Analysis Procedure for Enzymatic Mutation Detection', *Clin. Chem.*, **44**, 731–739 (1998).
95. A. Haliassos, J. Chomel, A. Tesson, M. Baudis, J. Kruh, J.C. Kaplan, A. Kitzis, 'Modification of Enzymatically Amplified DNA for Detection of Point Mutations', *Nucleic Acids Res.*, **17**, 3606 (1989).
96. H.G. Eiken, E. Odland, H.S. Boman, L.F. Engebretsen, J. Apold, 'Application of Natural and Amplification Created Restriction Sites for the Diagnosis of PKU Mutations', *Nucleic Acids Res.*, **19**, 1427–1430 (1991).
97. P. Patel, Y.M.D. Lo, J.I. Bell, J.S. Wainscoat, 'Detection of Susceptibility Alleles to Insulin-dependent Diabetes Mellitus at the DQB1 Locus by Artificial PCR-RFLP', *Immunogenetics*, **36**, 264–265 (1992).
98. R.K. Saiki, T.L. Bugawan, G.T. Horn, K.B. Mullis, H.A. Erlich, 'Analysis of Enzymatically Amplified β-Globin and HLA DQα DNA with Allele-specific Oligonucleotide Probes', *Nature*, **324**, 163–166 (1986).
99. R.K. Saiki, P.S. Walsh, C.H. Levenson, H.A. Erlich, 'Genetic Analysis of Amplified DNA with Immobilized Sequence-specific Oligonucleotide Probes', *Proc. Natl. Acad. Sci., USA*, **86**, 6230–6234 (1989).
100. K.J. Livak, J. Marmaro, J.A. Todd, 'Towards Fully Automated Genome-wide Polymorphism Screening', *Nat. Genet.*, **9**, 341–342 (1995).
101. S. Tyagi, D.P. Bratu, F.R. Kramer, 'Multicolor Molecular Beacons for Allele Discrimination', *Nat. Biotechnol.*, **16**, 49–53 (1998).
102. S.A. Marras, F.R. Kramer, S. Tyagi, 'Multiplex Detection of Single-nucleotide Variations Using Molecular Beacons', *Genet. Anal.*, **14**, 151–156 (1999).
103. D.Y. Wu, L. Ugozzoli, B.K. Pal, R.B. Wallace, 'Allele-specific Enzymatic Amplification of Beta-globin Genomic DNA for Diagnosis of Sickle Cell Anemia', *Proc. Natl. Acad. Sci., USA*, **86**, 2757–2760 (1989).
104. C.R. Newton, A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, A.F. Markham, 'Analysis of Any Point Mutation in DNA. The Amplification Refractory Mutation System (ARMS)', *Nucleic Acids Res.*, **17**, 2503–2516 (1989).
105. Y.M.D. Lo, P. Patel, C.R. Newton, A.F. Markham, K.A. Fleming, J.S. Wainscoat, 'Direct Haplotype Determination by Double ARMS: Specificity, Sensitivity and Genetic Applications', *Nucleic Acids Res.*, **19**, 3561–3567 (1991).
106. T. Seyama, T. Ito, T. Hayashi, T. Mizuno, N. Nakamura, M. Akiyama, 'A Novel Blocker-PCR Method for Detection of Rare Mutant Alleles in the Presence of an Excess Amount of Normal DNA', *Nucleic Acids Res.*, **20**, 2493–2496 (1992).
107. A. Orou, B. Fechner, G. Utermann, H.J. Menzel, 'Allele-specific Competitive Blocker PCR: a One-step Method with Applicability to Pool Screening', *Hum. Mutat.*, **6**, 163–169 (1995).

108. B.L. Parsons, R.H. Heflich, 'Detection of a Mouse H-ras Codon 61 Mutation Using a Modified Allele-specific Competitive Blocker PCR Genotypic Selection Method', *Mutagenesis*, **13**, 581–588 (1998).
109. J.R. Scherer, I. Kheterpal, A. Radhakrishnan, W.W. Ja, R.A. Mathies, 'Ultra-high Throughput Rotary Capillary Array Electrophoresis Scanner for Fluorescent DNA Sequencing and Analysis', *Electrophoresis*, **20**, 1508–1517 (1999).
110. S. Liu, Y. Shi, W.W. Ja, R.A. Mathies, 'Optimization of High-speed DNA Sequencing on Microfabricated Capillary Electrophoresis Channels', *Anal. Chem.*, **71**, 566–573 (1999).
111. D.J. Fu, K. Tang, A. Braun, D. Reuter, B. Darnhofer-Demar, D.P. Little, M.J. O'Donnell, C.R. Cantor, H. Koster, 'Sequencing Exons 5 to 8 of the p53 Gene by MALDI-TOF Mass Spectrometry', *Nat. Biotechnol.*, **16**, 381–384 (1998).
112. F. Ferre (ed.), *Gene Quantification*, Birkhauser, Boston, 1998.
113. C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, 'Real Time Quantitative PCR', *Genome Res.*, **6**, 986–994 (1996).
114. U.E. Gibson, C.A. Heid, P.M. Williams, 'A Novel Method for Real Time Quantitative RT-PCR', *Genome Res.*, **6**, 995–1001 (1996).
115. I. Bieche, I. Laurendeau, S. Tozlu, M. Olivi, D. Vidaud, R. Lidereau, M. Vidaud, 'Quantitation of MYC Gene Expression in Sporadic Breast Tumors with a Real-time Reverse Transcription-PCR Assay', *Cancer Res.*, **59**, 2759–2765 (1999).
116. L. Overbergh, D. Valckx, M. Waer, C. Mathieu, 'Quantification of Murine Cytokine mRNAs Using Real Time Quantitative Reverse Transcriptase PCR', *Cytokine*, **11**, 305–312 (1999).
117. Y.M.D. Lo, L.Y.S. Chan, K.W. Lo, S.F. Leung, J. Zhang, A.T.C. Chan, J.C.K. Lee, N.M. Hjelm, P.J. Johnson, D.P. Huang, 'Quantitative Analysis of Cell-free Epstein–Barr Virus DNA in Plasma of Patients with Nasopharyngeal Carcinoma', *Cancer Res.*, **59**, 1188–1191 (1999).
118. T. Takeuchi, A. Katsume, T. Tanaka, A. Abe, K. Inoue, K. Tsukiyama-Kohara, R. Kawaguchi, S. Tanaka, M. Kohara, 'Real-time Detection System for Quantification of Hepatitis C Virus Genome', *Gastroenterology*, **116**, 636–642 (1999).
119. I. Laurendeau, M. Bahauau, N. Vodovar, C. Larramendy, M. Olivi, I. Bieche, M. Vidaud, D. Vidaud, 'TaqMan™ PCR-based Gene Dosage Assay for Predictive Testing in Individuals from a Cancer Family with INK4 Locus Haploinsufficiency', *Clin. Chem.*, **45**, 982–986 (1999).
120. J.A.M. Vet, A.R. Majithia, S.A.E. Marras, S. Tyagi, S. Dube, B.J. Poiesz, F.R. Kramer, 'Multiplex Detection of Four Pathogenic Retroviruses Using Molecular Beacons', *Proc. Natl. Acad. Sci., USA*, **96**, 6394–6399 (1999).
121. Y.M.D. Lo, M.S.C. Tein, T.K. Lau, C.J. Haines, T.N. Leung, P.M.K. Poon, J.S. Wainscoat, P.J. Johnson, A.M.Z. Chang, N.M. Hjelm, 'Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis', *Am. J. Hum. Genet.*, **62**, 768–775 (1998).
122. Y.M.D. Lo, I.H.N. Wong, J. Zhang, M.S.C. Tein, M.H.L. Ng, N.M. Hjelm, 'Quantitative Analysis of Aberrant p16 Methylation Using Real-time Quantitative Methylation-specific Polymerase Chain Reaction', *Cancer Res.*, **59**, 3899–3903 (1999).
123. P.M. Holland, R.D. Abramson, R. Watson, D.H. Gelfand, 'Detection of Specific Polymerase Chain Reaction Product by Utilising the 5′-3′ Exonuclease Activity of the *Thermus aquaticus* DNA Polymerase', *Proc. Natl. Acad. Sci., USA*, **88**, 7276–7280 (1991).
124. A.A. Long, 'In-situ Polymerase Chain Reaction: Foundation of the Technology and Today's Options', *Eur. J. Histochem.*, **42**, 101–109 (1998).
125. K.H. Wiedorn, H. Kuhl, J. Galle, J. Caselitz, E. Vollmer, 'Comparison of In-situ Hybridization, Direct and Indirect In-situ PCR as Well as Tyramide Signal Amplification for the Detection of HPV', *Histochem. Cell. Biol.*, **111**, 89–95 (1999).
126. A.A. Long, P. Komminoth, 'In situ PCR: an Overview', in *PRINS and In Situ PCR Protocols*, ed. J.R. Gosden, Humana Press, Totowa, 141–161, 1997.
127. G.J. Nuovo, 'In Situ Localization of PCR-amplified DNA and cDNA', *Mol. Biotechnol.*, **10**, 49–62 (1998).
128. I. Zehbe, G.W. Hacker, J.F. Sallstrom, E. Rylander, E. Wilander, 'Self-sustained Sequence Replication-based Amplification (3SR) for the In-situ Detection of mRNA in Cultured Cells', *Cell Vis.*, **1**, 20–24 (1994).
129. H. Hofler, B. Putz, J. Mueller, W. Neubert, G. Sutter, P. Gais, 'In Situ Amplification of Measles Virus RNA by Self-sustained Sequence Replication Reaction', *Lab. Invest.*, **73**, 577–585 (1995).
130. J. Gosden, D. Hanratty, J. Starling, J. Fantes, A. Mitchell, D. Porteous, 'Oligonucleotide Primed In Situ DNA Synthesis (PRINS): a Method for Chromosome Mapping, Banding and Investigation of Sequence Organization', *Cytogenet. Cell. Genet.*, **57**, 100–104 (1991).
131. R.J. Lipshutz, S.P. Fodor, T.R. Gingeras, D.J. Lockhart, 'High Density Synthetic Oligonucleotide Arrays', *Nat. Genet.*, **21**, 20–24 (1999).
132. S. Drmanac, D. Kita, I. Labat, B. Hauser, C. Schmidt, J.D. Burczak, R. Drmanac, 'Accurate Sequencing by Hybridization for DNA Diagnostics and Individual Genomics', *Nat. Biotechnol.*, **16**, 54–58 (1998).
133. A. Drobyshev, N. Mologina, V. Shik, D. Pobedimskaya, G. Yershov, A. Mirzabekov, 'Sequence Analysis by Hybridization with Oligonucleotide Microchip: Identification of Beta-thalassemia Mutations', *Gene*, **188**, 45–52 (1997).

134. J.G. Hacia, L.C. Brody, M.S. Chee, S.P. Fodor, F.S. Collins, 'Detection of Heterozygous Mutations in BRCA1 Using High Density Oligonucleotide Arrays and Two-colour Fluorescence Analysis', *Nat. Genet.*, **14**, 441–447 (1996).
135. G. Ramsay, 'DNA Chips: State-of-the-art', *Nat. Biotechnol.*, **16**, 40–44 (1998).
136. M.J. Kozal, N. Shah, N. Shen, R. Yang, R. Fucini, T.C. Merigan, D.D. Richman, D. Morris, E. Hubbell, M. Chee, T.R. Gingeras, 'Extensive Polymorphisms Observed in HIV-1 Clad B Protease Gene Using High-density Oligonucleotide Arrays', *Nat. Med.*, **2**, 753–759 (1996).
137. W.Y. Tang, J. Elnatan, Y.S. Lee, H.S. Goh, D.R. Smith, 'c-Ki-ras Mutations in Colorectal Adenocarcinomas from a Country with a Rapidly Changing Colorectal Cancer Incidence', *Br. J. Cancer*, **81**, 237–241 (1999).
138. J.A. Newton Bishop, M. Harland, D.C. Bennett, V. Bataille, A.M. Goldstein, M.A. Tucker, B.A. Ponder, J. Cuzick, P. Selby, D.T. Bishop, 'Mutation Testing in Melanoma Families: INK4A, CDK4 and INK4D', *Br. J. Cancer*, **80**, 295–300 (1999).
139. J. Takita, Y. Hayashi, J. Yokota, 'Loss of Heterozygosity in Neuroblastomas – an Overview', *Eur. J. Cancer*, **33**, 1971–1973 (1997).
140. J.E. Rubnitz, A.T. Look, 'Molecular Basis of Leukemogenesis', *Curr. Opin. Hematol.*, **5**, 264–270 (1998).
141. T.H. Rabbitts, 'Perspective: Chromosomal Translocations can Affect Genes Controlling Gene Expression and Differentiation – Why are these Functions Targeted?', *J. Pathol.*, **187**, 39–42 (1999).
142. F. Mitelman, *Catalog of Chromosome Aberrations in Cancer*, 5th edition, Wiley-Liss, New York, 1995.
143. O. Delattre, J. Zucman, T. Melot, X.S. Garau, J.M. Zucker, G.M. Lenoir, P.F. Ambros, D. Sheer, C. Turc-Carel, T.J. Triche, A. Aurias, G. Thomas, 'The Ewing Family of Tumors – a Subgroup of Small-round-cell Tumors Defined by Specific Chimeric Transcripts', *N. Engl. J. Med.*, **331**, 294 (1994).
144. A. Kawai, J. Woodruff, J.H. Healey, M.F. Brennan, C.R. Antonescu, M. Ladanyi, 'SYT-SSX Gene Fusion as a Determinant of Morphology and Prognosis in Synovial Sarcoma', *N. Engl. J. Med.*, **338**, 153–160 (1998).
145. J.L. Bennicelli, F.G. Barr, 'Genetics and the Biologic Basis of Sarcomas', *Curr. Opin. Oncol.*, **11**, 267–274 (1999).
146. M. Schwab, 'Amplification of Oncogenes in Human Cancer Cells', *Bioessays*, **20**, 473–479 (1998).
147. P.A. Lazo, 'The Molecular Genetics of Cervical Carcinoma', *Br. J. Cancer*, **80**, 2008–2018 (1999).
148. P. Pineau, A. Marchio, M.G. Mattei, W.H. Kim, J.K. Youn, P. Tiollais, A. Dejean, 'Extensive Analysis of Duplicated–Inverted Hepatitis B Virus Integrations in Human Hepatocellular Carcinoma', *J. Gen. Virol.*, **79**, 591–600 (1998).
149. P.A. Jones, P.W. Laird, 'Cancer Epigenetics Comes of Age', *Nat. Genet.*, **21**, 163–167 (1999).
150. P.W. Laird, 'Oncogenic Mechanisms Mediated by DNA Methylation', *Mol. Med. Today*, **3**, 223–229 (1997).
151. J.G. Herman, F. Latif, W. Weng, M.I. Lerman, B. Zbar, S. Liu, D. Samid, D.S.R. Duan, J.R. Gnarra, W.M. Linehan, S.B. Baylin, 'Silencing of the *VHL* Tumor-suppressor Gene by DNA Methylation in Renal Carcinoma', *Proc. Natl. Acad. Sci., USA*, **91**, 9700–9704 (1994).
152. J.G. Herman, A. Merlo, L. Mao, R.G. Lapidus, J.P.J. Issa, N.E. Davidson, D. Sidransky, S.B. Baylin, 'Inactivation of the *CDKN2/p16/MTS1* Gene is Frequently Associated with Aberrant DNA Methylation in all Common Human Cancers', *Cancer Res.*, **55**, 4525–4530 (1995).
153. C.T. Liew, H.M. Li, K.W. Lo, C.K. Leow, J.Y.H. Chan, L.Y. Hin, W.Y. Lau, P.B.S. Lai, B.K. Lim, J. Huang, W.T. Leung, S. Wu, J.C.K. Lee, 'High Frequency of p16INK4A Gene Alterations in Hepatocellular Carcinoma', *Oncogene*, **18**, 789–795 (1999).
154. H. Cui, I.L. Horon, R. Ohlsson, S.R. Hamilton, A.P. Feinberg, 'Loss of Imprinting in Normal Tissue of Colorectal Cancer Patients with Microsatellite Instability', *Nat. Med.*, **4**, 1276–1280 (1998).
155. G.S. Randhawa, H. Cui, J.A. Barletta, L.Z. Strichman-Almashanu, M. Talpaz, H. Kantarjian, A.B. Deisseroth, R.C. Champlin, A.P. Feinberg, 'Loss of Imprinting in Disease Progression in Chronic Myelogenous Leukemia', *Blood*, **91**, 3144–3147 (1998).
156. P.W. Johnson, S.A. Burchill, P.J. Selby, 'The Molecular Detection of Circulating Tumour Cells', *Br. J. Cancer*, **72**, 268–276 (1995).
157. B. Smith, P. Selby, J. Southgate, K. Pittman, C. Bradley, G.E. Blair, 'Detection of Melanoma Cells in Peripheral Blood by Means of Reverse Transcriptase and Polymerase Chain Reaction', *Lancet*, **338**, 1227–1229 (1991).
158. L. Mao, R.H. Hruban, J.O. Boyle, M. Tockman, D. Sidransky, 'Detection of Oncogene Mutations in Sputum Precedes Diagnosis of Lung Cancer', *Cancer Res.*, **54**, 1634–1637 (1994).
159. S.A. Belinsky, K.J. Nikula, W.A. Palmisano, R. Michels, G. Saccomanno, E. Gabrielson, S.B. Baylin, J.G. Herman, 'Aberrant Methylation of p16(INK4a) is an Early Event in Lung Cancer and a Potential Biomarker for Early Diagnosis', *Proc. Natl. Acad. Sci., USA*, **95**, 11891–11896 (1998).
160. D. Sidransky, T. Tokino, S.R. Hamilton, K.W. Kinzler, B. Levin, P. Frost, B. Vogelstein, 'Identification of ras Oncogene Mutations in the Stool of Patients with Curable Colorectal Tumors', *Science*, **256**, 102–105 (1992).
161. P. Nollau, C. Moser, G. Weinland, C. Wagener, 'Detection of K-ras Mutations in Stools of Patients with

Colorectal Cancer by Mutant-enriched PCR', *Int. J. Cancer*, **66**, 332–336 (1996).
162. L. Mao, 'Genetic Alterations as Clonal Markers for Bladder Cancer Detection in Urine', *J. Cell. Biochem.*, **25**(Suppl.), 191–196 (1996).
163. S. Mourah, O. Cussenot, V. Vimont, F. Desgrandchamps, P. Teillac, B. Cochant-Priollet, A. Le Duc, J. Fiet, H. Soliman, 'Assessment of Microsatellite Instability in Urine in the Detection of Transitional-cell Carcinoma of the Bladder', *Int. J. Cancer*, **79**, 629–633 (1998).
164. G. Steiner, M.P. Schoenberg, J.F. Linn, L. Mao, D. Sidransky, 'Detection of Bladder Cancer Recurrence by Microsatellite Analysis of Urine', *Nat. Med.*, **3**, 621–624 (1997).
165. D.W. Bianchi, 'Fetal Cells in the Maternal Circulation: Feasibility for Prenatal Diagnosis', *Br. J. Haematol.*, **105**, 574–583 (1999).
166. Y.M.D. Lo, P. Patel, J.S. Wainscoat, M. Sampietro, M.D.G. Gillmer, K.A. Fleming, 'Prenatal Sex Determination by DNA Amplification from Maternal Peripheral Blood', *Lancet, ii*, 1363–1365 (1989).
167. S.M. Kao, G.C. Tang, T.T. Hsieh, K.C. Young, H.C. Wang, C.C. Pao, 'Analysis of Peripheral Blood of Pregnant Women for the Presence of Fetal Y Chromosome-specific ZFY Gene Deoxyribonucleic Acid Sequences', *Am. J. Obstet. Gynecol.*, **166**, 1013–1019 (1992).
168. L.R. Adkison, R.H. Andrews, N.L. Vowell, W.L. Koontz, 'Improved Detection of Fetal Cells from Maternal Blood with Polymerase Chain-reaction', *Am. J. Obstet. Gynecol.*, **170**, 952–955 (1994).
169. M.R. Thomas, B. Tutschek, A. Frost, C.H. Rodeck, N. Yazdani, I. Craft, R. Williamson, 'The Time of Appearance and Disappearance of Fetal DNA from the Maternal Circulation', *Prenat. Diagn.*, **15**, 641–646 (1995).
170. H. Hamada, T. Arinami, T. Kubo, H. Hamaguchi, H. Iwasaki, 'Fetal Nucleated Cells in Maternal Peripheral Blood: Frequency and Relationship to Gestational Age', *Hum. Genet.*, **91**, 427–432 (1993).
171. Y.L. Zheng, D.K. Zhen, M.A. DeMaria, S.M. Berry, R.J. Wapner, M.I. Evans, D. Copeland, J.M. Williams, D.W. Bianchi, 'Search for the Optimal Fetal Cell Antibody: Results of Immunophenotyping Studies Using Flow Cytometry', *Hum. Genet.*, **100**, 35–42 (1997).
172. D. Gänshirt-Ahlert, R. Börjesson-Stoll, M. Burschyk, A. Dohr, H.S.P. Garritsen, E. Helmer, P. Miny, M. Velasco, C. Walde, D. Patterson, N. Teng, N.M. Bhat, M.M. Bieber, W. Holzgreve, 'Detection of Fetal Trisomies 21 and 18 from Maternal Blood Using Triple Gradient and Magnetic Cell Sorting', *Am. J. Reprod. Immunol.*, **30**, 193–200 (1993).
173. S.S. Wachtel, D. Sammons, G. Twitty, J. Utermohlen, E. Tolley, O. Phillips, L.P. Shulman, 'Charge Flow Separation: Quantification of Nucleated Red Blood Cells in Maternal Blood During Pregnancy', *Prenat. Diagn.*, **18**, 455–463 (1998).
174. Y.M.D. Lo, E.S.F. Lo, N. Watson, L. Noakes, I.L. Sargent, B. Thilaganathan, J.S. Wainscoat, 'Two-way Cell Traffic Between Mother and Fetus: Biologic and Clinical Implications', *Blood*, **88**, 4390–4395 (1996).
175. G. Socie, E. Gluckman, E. Carosella, Y. Brossard, C. Lafon, O. Brison, 'Search for Maternal Cells in Human Umbilical Cord Blood by Polymerase Chain Reaction Amplification of Two Minisatellite Sequences', *Blood*, **83**, 340–344 (1994).
176. J.M. Hall, P. Lingenfelter, S.L. Adams, D. Lasser, J.A. Hansen, M.A. Bean, 'Detection of Maternal Cells in Human Umbilical Cord Blood Using Fluorescence In Situ Hybridization', *Blood*, **86**, 2829–2832 (1995).
177. T. Petit, M. Dommergues, G. Socie, Y. Dumez, E. Gluckman, O. Brison, 'Detection of Maternal Cells in Human Fetal Blood During the Third Trimester of Pregnancy Using Allele-specific PCR Amplification', *Br. J. Haematol.*, **98**, 767–771 (1997).
178. D.W. Bianchi, G.K. Zickwolf, G.J. Weil, S. Sylvester, M.A. DeMaria, 'Male Fetal Progenitor Cells Persist in Maternal Blood for as Long as 27 Years Postpartum', *Proc. Natl. Acad. Sci., USA*, **93**, 705–708 (1996).
179. J.L. Nelson, D.E. Furst, S. Maloney, T. Gooley, P.C. Evans, A. Smith, M.A. Bean, C. Ober, D.W. Bianchi, 'Microchimerism and HLA-compatible Relationships of Pregnancy in Scleroderma', *Lancet*, **351**, 559–562 (1998).
180. C.M. Artlett, J.B. Smith, S.A. Jimenez, 'Identification of Fetal DNA and Cells in Skin Lesions from Women with Systemic Sclerosis', *N. Engl. J. Med.*, **338**, 1186–1191 (1998).
181. S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruign, A.R. Coulson, J.E. Drouin, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, I.G. Young, 'Sequence and Organization of the Human Mitochondrial Genome', *Nature*, **290**, 457–465 (1981).
182. R.E. Giles, H. Blanc, H.M. Cann, D.C. Wallace, 'Maternal Inheritance of Human Mitochondrial DNA', *Proc. Natl. Acad. Sci., USA*, **77**, 6715–6719 (1980).
183. J. Loeffen, J. Smeitink, R. Triepels, R. Smeets, M. Schuelke, R. Sengers, F. Trijbels, B. Hamel, R. Mullaart, L. van den Heuvel, 'The First Nuclear-encoded Complex I Mutation in a Patient with Leigh Syndrome', *Am. J. Hum. Genet.*, **63**, 1598–1608 (1998).
184. G. Casari, M. De Fusco, S. Ciarmatori, M. Zeviani, M. Mora, P. Fernandez, G. De Michele, A. Filla, S. Cocozza, R. Marconi, A. Durr, B. Fontaine, A. Ballabio, 'Spastic Paraplegia and OXPHOS Impairment Caused by Mutations in Paraplegin, a Nuclear-encoded Mitochondrial Metalloprotease', *Cell*, **93**, 973–983 (1998).
185. T. Bourgeron, P. Rustin, D. Chretien, M. Birch-Machin, M. Bourgeois, E. Viegas-Pequignot, A. Munnich,

A. Rotig, 'Mutation of a Nuclear Succinate Dehydrogenase Gene Results in Mitochondrial Respiratory Chain Deficiency', *Nat. Genet.*, **11**, 144–149 (1995).

186. J.W. Taanman, A.G. Bodnar, J.M. Cooper, A.A. Morris, P.T. Clayton, J.V. Leonard, A.H. Schapira, 'Molecular Mechanisms in Mitochondrial DNA Depletion Syndrome', *Hum. Mol. Genet.*, **6**, 935–942 (1997).
187. C.T. Moraes, S. Shanske, H.J. Tritschler, J.R. Aprille, F. Andreetta, E. Bonilla, E.A. Schon, S. DiMauro, 'mtDNA Depletion with Variable Tissue Expression: a Novel Genetic Abnormality in Mitochondrial Diseases', *Am. J. Hum. Genet.*, **48**, 492–501 (1991).
188. D.C. Wallace, G. Singh, M.T. Lott, J.A. Hodge, T.G. Schurr, A.M. Lezza, L.J.I. Elsas, E.K. Nikoskelainen, 'Mitochondrial DNA Mutation Associated with Leber's Hereditary Optic Neuropathy', *Science*, **242**, 1427–1430 (1988).
189. I.J. Holt, A.E. Harding, J.A. Morgan-Hughes, 'Deletions of Muscle Mitochondrial DNA in Patients with Mitochondrial Myopathies', *Nature*, **331**, 717–719 (1988).
190. I. Trounce, E. Byrne, S. Marzuki, 'Decline in Skeletal Muscle Mitochondrial Respiratory Chain Function: Possible Factor in Ageing', *Lancet, i*, 637–639 (1989).
191. O.E. Rooyacker, D.B. Adey, P.A. Ades, K.S. Nair, 'Effect of Age on In Vivo Rates of Mitochondrial Protein Synthesis in Human Skeletal Muscle', *Proc. Natl. Acad. Sci., USA*, **93**, 15364–15369 (1996).
192. S. Melov, J.M. Shoffner, A. Kaufman, D.C. Wallace, 'Marked Increase in the Number and Variety of Mitochondrial DNA Rearrangements in Aging Human Skeletal Muscle', *Nucleic Acids Res.*, **23**, 4122–4126 (1995).
193. G. Kopsidas, S.A. Kovalenko, J.M. Kelso, A.W. Linnane, 'An Age-associated Correlation Between Cellular Bioenergy Decline and mtDNA Rearrangements in Human Skeletal Muscle', *Mutat. Res.*, **421**, 27–36 (1998).
194. V.W. Liu, C. Zhang, P. Nagley, 'Mutations in Mitochondrial DNA Accumulate Differentially in Three Different Human Tissues During Ageing', *Nucleic Acids Res.*, **26**, 1268–1275 (1998).
195. M. Corral-Debrinski, T. Horton, M.T. Lott, J.M. Shoffner, M.F. Beal, D.C. Wallace, 'Mitochondrial DNA Deletions in Human Brain: Regional Variability and Increase with Advanced Age', *Nat. Genet.*, **2**, 324–329 (1992).
196. T.M. Horton, J.A. Petros, A. Heddi, J. Shoffner, A.E. Kaufman, S.D.J. Graham, T. Gramlich, D.C. Wallace, 'Novel Mitochondrial DNA Deletion Found in a Renal Cell Carcinoma', *Genes Chromosomes Cancer*, **15**, 95–101 (1996).
197. K. Polyak, Y. Li, H. Zhu, C. Lengauer, J.K. Wilson, S.D. Markowitz, M.A. Trush, K.W. Kinzler, B. Vogelstein, 'Somatic Mutations of the Mitochondrial Genome in Human Colorectal Tumours', *Nat. Genet.*, **20**, 291–293 (1998).
198. W. Habano, S. Nakamura, T. Sugai, 'Microsatellite Instability in the Mitochondrial DNA of Colorectal Carcinomas: Evidence for Mismatch Repair Systems in Mitochondrial Genome', *Oncogene*, **17**, 1931–1937 (1998).
199. J. Boultwood, C. Fidler, K.I. Mills, M. Frodsham, R. Kusec, A. Gaiger, R.E. Gale, D.C. Linch, T.J. Littlewood, P.A.H. Moss, J.S. Wainscoat, 'Amplification of Mitochondrial DNA in Acute Myeloid Leukaemia', *Br. J. Haematol.*, **95**, 426–431 (1996).
200. T.J. Chen, R.G. Boles, L.J.C. Wong, 'Detection of Mitochondrial DNA Mutations by Temporal Temperature Gradient Gel Electrophoresis', *Clin. Chem.*, **45**, 1162–1167 (1999).
201. M. Urata, M. Wakiyama, M. Iwase, M. Yoneda, S. Kinoshita, N. Hamasaki, D. Kang, 'New Sensitive Method for the Detection of the A3243G Mutation of Human Mitochondrial Deoxyribonucleic Acid in Diabetes Mellitus Patients by Ligation-mediated Polymerase Chain Reaction', *Clin. Chem.*, **44**, 2088–2093 (1998).
202. E. Schollen, P. Vandenberk, J.J. Cassiman, G. Matthijs, 'Development of Reverse Dot-blot System for Screening of Mitochondrial DNA Mutations Associated with Leber Hereditary Optic Atrophy', *Clin. Chem.*, **43**, 18–23 (1997).
203. J.L. Davies, Y. Kawaguchi, S.T. Bennett, J.B. Copeman, H.J. Cordell, L.E. Pritchard, P.W. Reed, S.C. Gough, S.C. Jenkins, S.M. Palmer, K.M. Balfour, B.R. Rowe, M. Farrall, A.H. Barnett, S.C. Bain, J.A. Todd, 'A Genome-wide Search for Human Type 1 Diabetes Susceptibility Genes', *Nature*, **371**, 130–136 (1994).
204. J.A. Todd, 'Genetic Analysis of Type 1 Diabetes Using Whole Genome Approaches', *Proc. Natl. Acad. Sci., USA*, **92**, 8560–8565 (1995).
205. C.A. Mein, L. Esposito, M.G. Dunn, G.C. Johnson, A.E. Timms, J.V. Goy, A.N. Smith, L. Sebag-Montefiore, M.E. Merriman, A.J. Wilson, L.E. Pritchard, F. Cucca, A.H. Barnett, S.C. Bain, J.A. Todd, 'A Search for Type 1 Diabetes Susceptibility Genes in Families from the United Kingdom', *Nat. Genet.*, **19**, 297–300 (1998).
206. P. Concannon, K.J. Gogolin-Ewens, D.A. Hinds, B. Wapelhorst, V.A. Morrison, B. Stirling, M. Mitra, J. Farmer, S.R. Williams, N.J. Cox, G.I. Bell, N. Risch, R.S. Spielman, 'A Second-generation Screen of the Human Genome for Susceptibility to Insulin-dependent Diabetes Mellitus', *Nat. Genet.*, **19**, 292–296 (1998).
207. J.A. Todd, J.I. Bell, H.O. McDevitt, 'HLA DQ Beta Gene Contributes to Susceptibility and Resistance to Insulin-dependent Diabetes Mellitus', *Nature*, **329**, 599–604 (1987).
208. M. Odawara, 'Involvement of Mitochondrial Gene Abnormalities in the Pathogenesis of Diabetes Mellitus', *Ann. NY Acad. Sci.*, **786**, 72–81 (1996).

209. S.P. Miller, G.R. Anand, E.J. Karschinia, G.I. Bell, D.C. LaPorte, A.J. Lange, 'Characterization of Glucokinase Mutations Associated with Maturity-onset Diabetes of the Young Type 2 (MODY-2): Different Glucokinase Defects Lead to a Common Phenotype', *Diabetes*, **48**, 1645–1651 (1999).
210. K. Yamagata, H. Furuta, N. Oda, P.J. Kaisaki, S. Menzel, N.J. Cox, S.S. Fajans, S. Signorini, M. Stoffel, G.I. Bell, 'Mutations in the Hepatocyte Nuclear Factor-4 Alpha Gene in Maturity-onset Diabetes of the Young', *Nature*, **384**, 458–460 (1996).
211. K. Yamagata, N. Oda, P.J. Kaisaki, S. Menzel, H. Furuta, M. Vaxillaire, L. Southam, R.D. Cox, G.M. Lathrop, V.V. Boriraj, X. Chen, N.J. Cox, Y. Oda, H. Yano, M.M. Le Beau, S. Yamada, H. Nishigori, J. Takeda, S.S. Fajans, A.T. Hattersley, N. Iwasaki, T. Hansen, O. Pedersen, K.S. Polonsky, R.C. Turner, G. Velho, J.C. Chevre, P. Froguel, G.I. Bell, 'Mutations in the Hepatocyte Nuclear Factor-1 Alpha Gene in Maturity-onset Diabetes of the Young', *Nature*, **384**, 455–458 (1996).
212. D.A. Stoffers, J. Ferrer, W.L. Clarke, J.F. Habener, 'Early-onset Type II Diabetes Mellitus (MODY4) Linked to IPF1', *Nat. Genet.*, **17**, 138–139 (1997).
213. S.T. Bennett, A.M. Lucassen, S.C. Gough, E.E. Powell, D.E. Undlien, L.E. Pritchard, M.E. Merriman, Y. Kawaguchi, M.J. Dronsfield, F. Pociot, J. Nerup, N. Bouzekri, A. Cambon-Thomsen, K.S. Ronningen, A.H. Barnett, S.C. Bain, J.A. Todd, 'Susceptibility to Human Type 1 Diabetes at IDDM2 is Determined by Tandem Repeat Variation at the Insulin Gene Minisatellite Locus', *Nat. Genet.*, **9**, 284–292 (1995).
214. H. Makino, M. Taira, F. Shimada, N. Hashimoto, Y. Suzuki, O. Nozaki, Y. Hatanaka, S. Yoshida, 'Insulin Receptor Gene Mutation: a Molecular Genetical and Functional Analysis', *Cell Signal.*, **4**, 351–363 (1992).
215. J. Hager, L. Hansen, C. Vaisse, N. Vionnet, A. Philippi, W. Poller, G. Velho, C. Carcassi, L. Contu, C. Julier, F. Cambien, P. Passa, M. Lathrop, W. Kindsrogel, F. Demenais, E. Nishimura, P. Froguel, 'A Missense Mutation in the Glucagon Receptor Gene is Associated with Non-insulin-dependent Diabetes Mellitus', *Nat. Genet.*, **9**, 299–304 (1995).
216. S. Sakagashira, T. Sanke, T. Hanabusa, H. Shimomura, S. Ohagi, K.Y. Kumagaye, K. Nakajima, K. Nanjo, 'Missense Mutation of Amylin Gene (S20G) in Japanese NIDDM Patients', *Diabetes*, **45**, 1279–1281 (1996).
217. K. Almind, G. Inoue, O. Pedersen, C.R. Kahn, 'A Common Amino Acid Polymorphism in Insulin Receptor Substrate-1 Causes Impaired Insulin Signaling. Evidence from Transfection Studies', *J. Clin. Invest.*, **97**, 2569–2575 (1996).
218. J.P. Whitehead, P. Humphreys, A. Krook, R. Jackson, A. Hayward, H. Lewis, K. Siddle, S. O'Rahilly, 'Molecular Scanning of the Insulin Receptor Substrate 1 Gene in Subjects with Severe Insulin Resistance: Detection and Functional Analysis of a Naturally Occurring Mutation in a YMXM Motif', *Diabetes*, **47**, 837–839 (1998).
219. R.J. Sigal, A. Doria, J.H. Warram, A.S. Krolewski, 'Codon 972 Polymorphism in the Insulin Receptor Substrate-1 Gene, Obesity, and Risk of Noninsulin-dependent Diabetes Mellitus', *J. Clin. Endocrinol. Metab.*, **81**, 1657–1659 (1996).
220. C.L. Hanis, E. Boerwinkle, R. Chakraborty, D.L. Ellsworth, P. Concannon, B. Stirling, V.A. Morrison, B. Wapelhorst, R.S. Spielman, K.J. Gogolin-Ewens, J.M. Shepard, S.R. Williams, N. Risch, D. Hinds, N. Iwasaki, M. Ogata, Y. Omori, C. Petzold, H. Rietzch, H.E. Schroder, J. Schulze, N.J. Cox, S. Menzel, V.V. Boriraj, X. Chen, 'A Genome-wide Search for Human Non-insulin-dependent (Type 2) Diabetes Genes Reveals a Major Susceptibility Locus on Chromosome 2', *Nat. Genet.*, **13**, 161–166 (1996).
221. S. Ghosh, E.R. Hauser, V.L. Magnuson, T. Valle, D.S. Ally, Z.E. Karanjawala, J.B. Rayman, J.I. Knapp, A. Musick, J. Tannenbaum, C. Te, W. Eldridge, S. Shapiro, T. Musick, C. Martin, A. So, A. Witt, J.B. Harvan, R.M. Watanabe, W. Hagopian, J. Eriksson, S.J. Nylund, K. Kohtamaki, E. Tuomilehto-Wolf, L. Toivanen, G. Vidgren, C. Ehnholm, R.N. Bergman, J. Tuomilehto, F.S. Collins, M. Boehnke, 'A Large Sample of Finnish Diabetic Sib-pairs Reveals No Evidence for a Non-insulin-dependent Diabetes Mellitus Susceptibility Locus at 2qter', *J. Clin. Invest.*, **102**, 704–709 (1998).
222. C.R. Steinman, 'Free DNA in Serum and Plasma from Normal Adults', *J. Clin. Invest.*, **56**, 512–515 (1975).
223. G.L.J. Davis, J.S.I. Davis, 'Detecting of Circulating DNA by Counterimmunoelectrophoresis (CIE)', *Arthritis Rheum.*, **16**, 52–58 (1973).
224. B. Shapiro, M. Chakrabarty, E.M. Cohn, S.A. Leon, 'Determination of Circulating DNA Levels in Patients with Benign or Malignant Gastrointestinal Disease', *Cancer*, **51**, 2116–2120 (1983).
225. J.S. Vargo, D.M. Becker, J.T. Philbrick, F.W. Schoonover, J.S. Davis, 'Plasma DNA: a Simple, Rapid Test for Aiding the Diagnosis of Pulmonary Embolism', *Chest*, **97**, 63–68 (1990).
226. C.R. Steinman, 'Circulating DNA in Systemic Lupus Erythematosus', *J. Clin. Invest.*, **73**, 832–841 (1984).
227. H. Nawroz, W. Koch, P. Anker, M. Stroun, D. Sidransky, 'Microsatellite Alterations in Serum DNA of Head and Neck Cancer Patients', *Nat. Med.*, **2**, 1035–1037 (1996).
228. X.Q. Chen, M. Stroun, J.-L. Magnenat, L.P. Nicod, A.-M. Kurt, J. Lyautey, C. Lederrey, P. Anker, 'Microsatellite Alterations in Plasma DNA of Small Cell Lung Cancer Patients', *Nat. Med.*, **2**, 1033–1035 (1996).
229. K. Hibi, R. Robinson, S. Booker, L. Wu, S.R. Hamilton, D. Sidransky, J. Jen, 'Molecular Detection of Genetic

Alterations in the Serum of Colorectal Cancer Patients', *Cancer Res.*, **58**, 1405–1407 (1998).
230. P. Anker, F. Lefort, V. Vasioukhin, J. Lyautey, C. Lederrey, X.Q. Chen, M. Stroun, H.E. Mulcahy, M.J.G. Farthing, 'K-ras Mutations are Found in DNA Extracted from the Plasma of Patients with Colorectal Cancer', *Gastroenterology*, **112**, 1114–1120 (1997).
231. I.H.N. Wong, Y.M.D. Lo, J. Zhang, C.T. Liew, M.H.L. Ng, N. Wong, P.B.S. Lai, W.Y. Lau, N.M. Hjelm, P.J. Johnson, 'Detection of Aberrant p16 Methylation in the Plasma and Serum of Liver Cancer Patients', *Cancer Res.*, **59**, 71–73 (1999).
232. M. Esteller, M. Sanchez-Cespedes, R. Rosell, D. Sidransky, S.B. Baylin, J.G. Herman, 'Detection of Aberrant Promoter Hypermethylation of Tumor Suppressor Genes in Serum DNA from Non-small Cell Lung Cancer Patients', *Cancer Res.*, **59**, 67–70 (1999).
233. K.W. Lo, Y.M.D. Lo, S.F. Leung, Y.S. Tsang, L.Y.S. Chan, P.J. Johnson, N.M. Hjelm, J.C.K. Lee, D.P. Huang, 'Analysis of Cell-free Epstein–Barr Virus-associated RNA in the Plasma of Patients with Nasopharyngeal Carcinoma', *Clin. Chem.*, **45**, 1292–1294 (1999).
234. Y.M.D. Lo, N. Corbetta, P.F. Chamberlain, V. Rai, I.L. Sargent, C.W.G. Redman, J.S. Wainscoat, 'Presence of Fetal DNA in Maternal Plasma and Serum', *Lancet*, **350**, 485–487 (1997).
235. Y.M.D. Lo, N.M. Hjelm, C. Fidler, I.L. Sargent, M.F. Murphy, P.F. Chamberlain, P.M.K. Poon, C.W.G. Redman, J.S. Wainscoat, 'Prenatal Diagnosis of Fetal RhD Status by Molecular Analysis of Maternal Plasma', *N. Engl. J. Med.*, **339**, 1734–1738 (1998).
236. B.H.W. Faas, E.A. Beuling, G.C.M.L. Christiaens, A.E.G.K. Von dem Borne, C.E. Van der Schoot, 'Detection of Fetal *RHD*-specific Sequences in Maternal Plasma', *Lancet*, **352**, 1196 (1998).
237. Y.M.D. Lo, T.N. Leung, M.S.C. Tein, I.L. Sargent, J. Zhang, T.K. Lau, C.J. Haines, C.W.G. Redman, 'Quantitative Abnormalities of Fetal DNA in Maternal Serum in Pre-eclampsia', *Clin. Chem.*, **45**, 184–188 (1999).
238. T.N. Leung, J. Zhang, T.K. Lau, N.M. Hjelm, Y.M.D. Lo, 'Maternal Plasma Fetal DNA as a Marker for Preterm Labour', *Lancet*, **352**, 1904–1905 (1998).
239. Y.M.D. Lo, M.S.C. Tein, C.C.P. Pang, C.K. Yeung, K.L. Tong, N.M. Hjelm, 'Presence of Donor-specific DNA in Plasma of Kidney and Liver Transplant Recipients', *Lancet*, **351**, 1329–1330 (1998).
240. F.S. Collins, A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland, L. Walters, 'New Goals for the U.S. Human Genome Project: 1998–2003', *Science*, **282**, 682–689 (1998).
241. The *C. elegans* Sequencing Consortium. 'Genome Sequence of the Nematode *C. elegans*: a Platform for Investigating Biology', *Science*, **282**, 2012–2018 (1998).
242. G.J.B. van Ommen, E. Bakker, J.T. den Dunnen, 'The Human Genome Project and the Future of Diagnostics, Treatment, and Prevention', *Lancet*, **354**(Suppl. 1), 5–10 (1999).

Nucleic Acid Analysis in Clinical Chemistry

Michael Thompson and Anil K. Deisingh
University of Toronto, Toronto, Canada

1 Introduction **1**
1.1 Overview 1
1.2 Structure and Function of Nucleic Acids 1
1.3 Hybridization 2
1.4 Probe Labeling 4
1.5 The Polymerase Chain Reaction 5
1.6 DNA Sequencing 5
2 Microbiological Analysis **6**
2.1 Bacteriology 6
2.2 Virology 8
2.3 Parasitology 9
3 Genetic Diseases and Screening **10**
3.1 The Genetic Basis of Disease 10
3.2 The Human Genome Project 10
3.3 A Survey of Some Genetic Diseases 10
3.4 Single Nucleotide Polymorphism 11
4 Other Areas of Human Medicine **11**
4.1 Introduction 11
4.2 Oncology 11
4.3 Alzheimer's Disease 12
5 The Biosensor Approach **12**
5.1 Introduction 12
5.2 Properties of an Ideal Sensor 13
5.3 Types of Sensors 13
5.4 Biosensors for Nucleic Acids 13
5.5 DNA and Oligonucleotide Probe Arrays (Gene Chips) 13
6 Current and Future Developments **14**
Acknowledgments **15**
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

This article examines the role of nucleic acid analysis in clinical chemistry. It is a broad-based article and begins with the background knowledge which will be needed to appreciate the succeeding sections. In the Introduction, structure and function of nucleic acids, hybridization, probe labeling, hybridization protection, polymerase chain reaction (PCR) and sequencing are discussed. Major sections are devoted to microbiological analysis (with particular emphasis on bacteriology), genetic diseases and screening [discussions are provided for cystic fibrosis (CF), heriditary fructose intolerance (HFI) and hemochromatosis along with an introduction to the Human Genome Project and single nucleotide polymorphisms (SNP)], other areas of human medicine (oncology, colorectal cancer and Alzheimer's disease (AD) are considered) and the biosensor approach. A discussion of interesting current and future developments, not only in nucleic acid analysis, but also in gene therapy, concludes the article.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Overview

In the past decade, nucleic acids, in particular DNA, have been widely utilized as diagnostic tools in clinical chemistry. In this article, we shall examine some of the theoretical aspects of the field and then describe various applications in areas such as detection of genetic diseases, diagnosis of infectious disease and aspects of human medicine and, finally, an introduction to progress in biosensor development will be considered.

1.2 Structure and Function of Nucleic Acids

Nucleic acids are high-molecular-weight biomolecular polymers which, upon hydrolysis, yield pyrimidine and purine bases, a sugar component and phosphoric acid.[1] There are two main types of nucleic acid, DNA and RNA. The pyrimidine bases are cytosine (found in RNA and DNA) and thymine (DNA) or uracil (RNA). Both nucleic acids contain two purine bases, adenine and guanine. Structures of these bases are shown in Figure 1.

The sugar in RNA is ribose whereas in DNA it is 2-deoxyribose, both occurring in the furanose form. A purine or pyrimidine linked to a sugar is termed a *nucleoside* while a molecule containing base, sugar and phosphate group is called a *nucleotide*.

1.2.1 Base Pairing[2]

Planar bases can associate in specific ways by means of hydrogen bonding. The adenine : thymine (A.T) base pair has two hydrogen bonds while the guanine : cytosine (G.C) base pair has three. This Watson–Crick base pairing shows that the sugar groups are both attached to the bases on the same side of the base pair (Figure 2a and b). The pairs of bases are flat and may be stacked one above the other so that the molecule is represented as

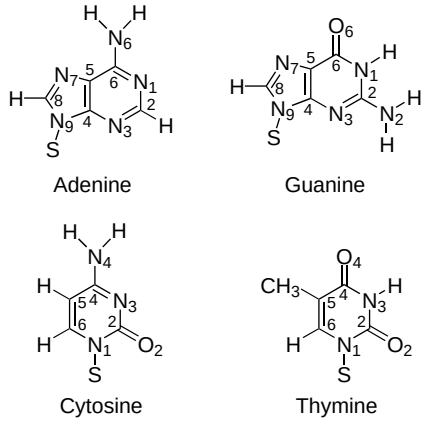

Figure 1 The four bases of nucleic acids.

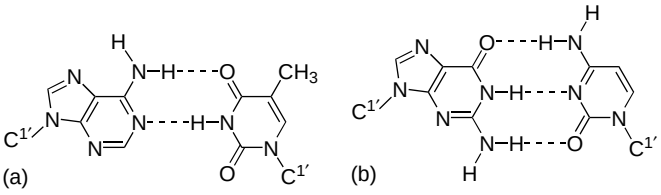

Figure 2 Watson–Crick hydrogen bonding.

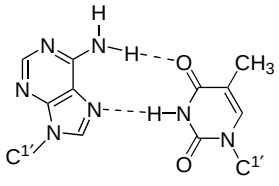

Figure 3 Hoogsteen hydrogen bonding.

a spiral staircase with the base pairs forming the treads. Stacked bases form a hydrophobic core.[1] On the other hand, Hoogsteen base pairing involves atoms N6 and N7 of adenine rather than N1 and N6 of Watson–Crick hydrogen bonding (Figure 3). There are also cases where a triple helix is formed by two pyrimidine strands and one purine strand and it is right-handed with an adenine-strand Watson–Crick hydrogen bonded to a thymine or uracil one and a third pyrimidine-strand Hoogsteen hydrogen bonded to the purine strand and parallel to it (Figure 4).[2]

1.2.2 Denaturation and Renaturation[1]

When double-stranded DNA (ds-DNA) molecules are subjected to extremes of temperature or pH, the hydrogen bonds are broken and the strands become separate. The DNA is denatured and changes from a double helix to

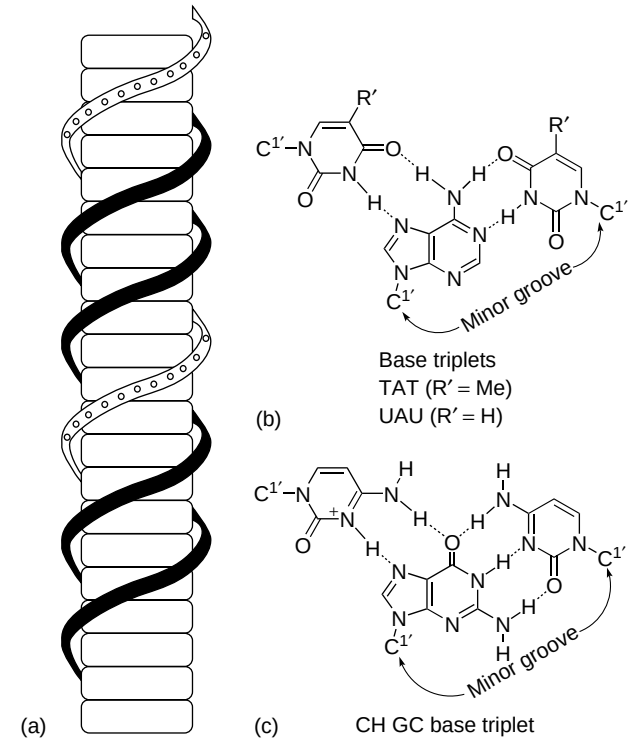

Figure 4 Triple helix formation. (a) Schematic view with third strand dotted. Base triplets [(b) TAT and (c) CH + GC] with extra pyrimidine bound by Hoogsteen hydrogen bonds in the major groove of the duplex.

a random coil (the helix–coil transition). If heat is the denaturant, the DNA is said to melt and the temperature at which the strands separate is termed the melting temperature, T_m. This transition is affected by the G + C content of the DNA, the nature of the solvent and nature of the DNA. There are three hydrogen bonds in G + C but only two in an A + T base pair, thereby making the G + C more stable and leading to a higher T_m. With regard to the solvent, at low concentration of counterion denaturation occurs at low temperature, whereas at high concentration T_m is increased.

Renaturation occurs when two DNA strands are reassociated to form a double helix. This process is dependent on the concentration of DNA molecules and the time allowed for reassociation. Renaturation is usually expressed as C_0t, where C_0 is the initial concentration in moles of nucleotide per liter and t is the time in seconds.

1.3 Hybridization

The employment of nucleic acid analysis primarily depends on the occurrence of hybridization, i.e. the formation of the Watson–Crick double helix from two single-stranded regions of DNA which are

complementary to each other.[3] The process was first reported in 1962[4] and success depends upon the two single strands having some degree of complementarity, the extent of which determines the stability of the ds-DNA which is formed. The stability increases when there is maximum hydrogen bonding and base stacking.[3]

RNA molecules can also take part in hybridization and base pairing can occur between complementary strands of DNA, between DNA and RNA and between complementary strands of RNA. A short (usually 15–30 base pairs but, in some cases, much longer) known segment of nucleic acid which is used to investigate an unknown sample for the presence of complementary sequences via hybridization is called a *nucleic acid probe*.[5]

1.3.1 Types of Probe[6]

Several types of probe are commonly encountered in clinical chemistry and these include cloned DNA probes which contain a known segment of DNA which is inserted into a plasmid vector. Cloned DNA probes are used to detect complementary species on Southern (DNA) or Northern (RNA) blots. These probes are usually >200 base pairs in length. Another common type is the synthetic DNA probe obtained either through oligonucleotide synthesis (by a commercially available DNA synthesizer) or by PCR amplification which allows large amounts of specific DNA probes to be obtained from small quantities of initial material.

RNA probes are usually obtained by the use of cloning vectors. These have found use in the detection of low concentrations of messenger RNA (mRNA).

1.3.2 Factors Affecting Hybridization

Many factors affect hybridization reactions, some of which are as follows:[7,8]

1. Temperature: at low temperature, the reassociation rate is determined by the difference in free energy between the unassociated and transition states. At higher temperature, the stability of the duplex is reduced until it is unstable and the hybrid melts.
2. Monovalent cation concentration: the T_m of a hybrid is reduced at low salt concentration because cations stabilize the DNA duplex.
3. Base composition especially with respect to G + C because rate of hybridization increases to some extent as this percentage increases. A higher G + C percentage also leads to an increase in T_m.
4. Duplex length: the rate of hybridization is proportional to the square root of the length of the smallest fragment.
5. Degree of mismatching of bases in the hybrid.
6. pH: hybridizations are normally performed in the pH range 6.8–7.4. Between pH 5 and 9, it is usually independent of pH.
7. Accelerators: inert polymers such as 10% dextran sulfate and 10% poly(ethene glycol) increase the rate of hybridization, probably by concentrating the probe and by increasing the extent of reaction.

1.3.3 Hybridizations on a Solid Support

This was first described in 1965[9] whereby a single-stranded DNA (ss-DNA) or RNA is immobilized on a solid support, e.g. nylon, nitrocellulose or polystyrene. This is useful for hybridization with the probe in a liquid phase[3] and there are several variations of solid support hybridization assays. We shall briefly consider dot/blot hybridization, Southern, Northern and Western blots and in situ hybridization.

1. Dot/blot hybridization: in this procedure, several samples are immobilized on the solid membrane. Manual addition of samples leads to a random shape (a blot) whereas with commercial devices the shape is regular (a dot).[5] This method allows all samples to be processed simultaneously with both samples and controls being treated under the same conditions.
2. Southern, Northern and Western blots: the Southern and Northern blots use restriction enzymes to cut DNA or mRNA which is obtained from tissues.[10] After digestion, the fragments are separated by electrophoresis, stained with ethidium bromide and the bands are observed under ultraviolet (UV) light. The DNA is then blotted on to a piece of electrostatically charged paper (a nitrocellulose membrane). Finally, hybridization is done using a labeled probe which binds to the fragment of interest with visualization by autoradiography.[10] The Northern blot is more difficult to perform because of the unstable nature of mRNA. The Western blot is used to analyze proteins by using antibodies which are specific for certain regions of the protein.
3. In situ hybridization: this is a special type of assay which uses intact cells, tissues or chromosomes and is used to demonstrate the presence of a particular gene within the material under study. The technique is more difficult than those previously described because the molecules have to reach the target in the tissue.[10] With chromosomal analysis, fluorescent detection is used whereby the technique is called fluorescent in situ hybridization (FISH). Since each

sample must be processed individually, the method is tedious and time-consuming.[3]

1.4 Probe Labeling

The type of labeling depends on the sensitivity of detection which is required for the particular analysis. For high sensitivity, it is common to use radiolabeling techniques even though there are problems with health hazards, storage and length of detection time.[6] Hence nonisotopic methods have become fairly popular.

1.4.1 Isotopic Labeling

The most common isotopes used are ^{32}P and ^{35}S but ^{33}P, ^{125}I, ^{14}C and ^{3}H have found use in several applications. ^{32}P has the highest emission energy and so will lead to the highest detection sensitivity. However, it has a short half-life ($t_{1/2}$) of 14.3 d, which limits its usefulness. If lower emission is required, it is wise to use either ^{35}S ($t_{1/2} = 87.4$ d) or ^{33}P ($t_{1/2} = 25.4$ d).[6] Methods of labeling DNA probes include nick translation,[11] which is used to label ds-DNA, and random priming,[12] where the template probe is denatured and then allowed to anneal with short oligonucleotides of random sequence.[5] It should be noted that, depending on which end of the probe is to be labeled, different enzymes are used. For 5′ labeling (e.g. with ^{32}P) it is common to use T4 polynucleotide kinase, whereas for the 3′ end, terminal deoxynucleotidyl transferase (TdT) is preferred.[5]

1.4.2 Nonisotopic Labeling

Most of the nonisotopic methods involve reactions with the chemical. Detection is usually by absorbance, fluorescent or luminescent systems or by pH electrodes or sensor technology (see section 5). Methods may be classified as either direct or indirect, with the former used for detection of standardized target biomolecules whereas the latter are used for biomolecules which are not specific. Direct methods require the reporter groups (e.g. fluorescein, rhodamine, alkaline phosphatase, horseradish peroxidase) to be directly coupled to the probes. In the indirect approach, however, there is also a noncovalent interaction between the probe and a molecule which binds to the probe. Examples of these include biotin, digoxigenin and immunogold.[13] Some of these probe detection systems are summarized in Figure 5.

1.4.3 Hybridization Protection Assay[14]

The hybridization protection assay (HPA) uses short, sequence-specific probes to detect nucleic acid targets of interest. In the most common format, an acridinium ester (AE) molecule is covalently attached to each probe molecule. The format is homogeneous, requires no

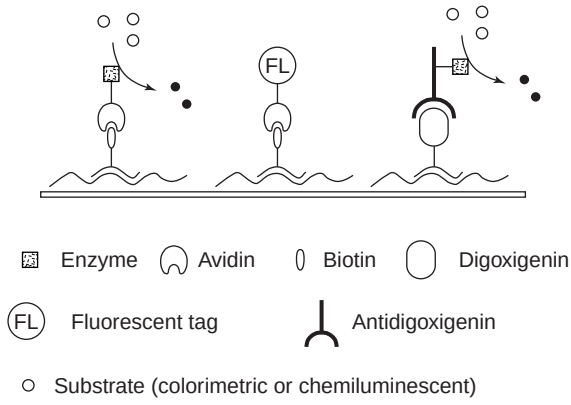

Figure 5 Probe detection systems.

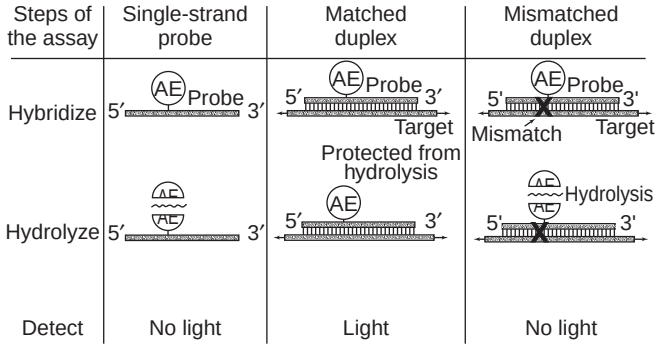

Figure 6 HPA for the detection of mutations.

physical separation steps and can rapidly and sensitively detect all single-base mismatches, multiple mismatches, insertions, deletions and genetic translocations. The HPA method consists of three main steps (Figure 6): (i) the AE probe is added to the sample followed by hybridization; (ii) a mildly alkaline solution is added to hydrolyze AE linked to unhybridized probe; and (iii) detection of remaining AE as a measure of the amount of target present.

HPA is readily used to detect mutations such as a point mutation in the reverse transcriptase (RT) coding region of human immunodeficiency virus (HIV)-1, the F 508 mutation in CF and in cancer mutation (section 4). The technique has also been utilized in Food and Drug Administration (FDA)-approved kits for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* (section 2.1.4)

Advantages of the HPA method include the combination of two separate approaches to discriminate matched and mismatched duplexes, rapidity, simplicity, simultaneous detection of multiple targets and detection requires only a luminometer. However, the method has a few drawbacks, the main one being that samples cannot be re-read since the ester reacts only once with alkaline peroxide to

produce light. Also, multiple analyte detection is actually limited to about five analytes.

1.5 The Polymerase Chain Reaction

The PCR is an in vitro technique which allows the amplification of a specific DNA region which lies between two regions of a known DNA sequence.[15] The method was invented by Kary Mullis,[16] who shared the Nobel Prize for Chemistry in 1993 for this work. The PCR involves three major steps: denaturation, annealing and polymerization (Figure 7).[17]

The template DNA (i.e. the DNA which will be amplified) is usually a ds-DNA fragment. However, for PCR to occur (because of the enzyme used, Taq polymerase derived from *Thermus aquaticus*), ss-DNA is required and so denaturation is carried out at 92–95 °C. The enzyme is stable at these high temperatures. Following denaturation, the mixture is quickly cooled to about 55 °C, where the base sequences anneal to the single strands of DNA and this process is followed by polymerization at about 72 °C (this is the optimum temperature for Taq polymerase). Here, the enzyme incorporates nucleotides into the DNA strand thus producing a complementary copy of the DNA template.[17] The entire cycle of denaturation, annealing and polymerization is repeated 20–30 times.

After each cycle, there is a doubling of the number of DNA molecules present and the final number of copies of target sequence is expressed by the formula[15] $(2^n - 2n)x$, where n = number of cycles, $2n$ = first product obtained after cycle 1 and second products obtained after cycle 2 with undefined length and x = number of copies of the original template. Thus, after about 20 cycles, a 10^6-fold increase in material is obtained.

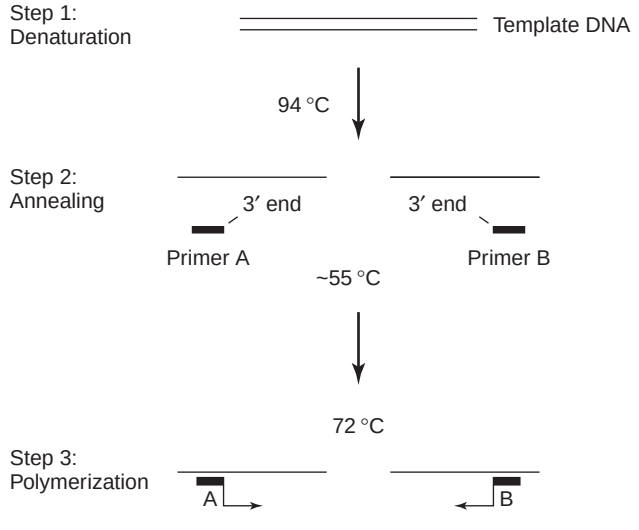

Figure 7 The PCR.

The basic components of a PCR amplification include DNA template, DNA Taq polymerase, deoxynucleoside triphosphates (dNTPs), primers (short ss-DNA molecules which are complementary to the ends of a defined sequence of DNA template), buffer [usually a mixture of tris(hydroxymethyl)aminomethane (TRIS)–HCl, KCl, $MgCl_2$ and gelatin] and $MgCl_2$. These are covered by a thin film of mineral oil in the tube before placing into a thermal cycler. Mg^{2+} ions form a soluble complex with the dNTPs, which is necessary for dNTP incorporation. They also stimulate polymerase activity and increase the T_m of the ds-DNA.[15]

There are several critical parameters which need to be considered when performing PCR to ensure a positive outcome.[18] These include reagent purity, primer selection (especially length and complementarity to template), template purity and concentration, using only the required amounts of dNTPs as increasing the concentration does not lead to greater efficiency and optimization of the time required for each step of the thermal cycling process.

To determine if the reaction is successful, a gel electrophoresis experiment is performed with staining being effected by using ethidium bromide, a fluorescent dye which intercalates with the DNA. Visualization under UV light is done. Molecular weight markers are usually electrophoresed in adjacent cells to allow accurate size determination of the product.[15]

Contamination is a major concern when performing PCR. Sources of possible contaminants include biological samples such as hair and skin cells, sample collection methods, laboratory equipment, pipets and pipet tips, centrifuges and electrophoresis apparatus. To reduce the possibility of contamination occurring, several precautions can be taken.[17,19] These include a physical separation of pre- and post-PCR amplifications, micropipettors specifically devoted to PCR use, the use of positive displacement pipets or plugged pipet tips, aliquoting of reagents and negative controls performed for each experiment.

These considerations are of prime importance in the clinical laboratory where false-positive results can be devastating to both the patient and the credibility of the laboratory.

1.6 DNA Sequencing[20]

In many instances, the DNA sequence must be known so as to allow for further investigation. A rapid method is to do DNA sequencing followed by a computer-assisted search for restriction endonuclease cleavage sites. This is especially useful in obtaining structural information about cloned genes and their products and it also provides a detailed analysis of the 5′ and

3′ noncoding regulatory regions of a gene. Finally, DNA sequence information is essential for site-directed mutagenesis.

There are two main methods for DNA sequencing. (a) The dideoxy (Sanger) sequencing method[21] is an enzymatic system which was developed by Fred Sanger et al. and it uses a DNA polymerase to synthesize a complementary copy of an ss-DNA template. Chain elongation occurs at the 3′ end of a primer DNA which is annealed to template DNA. Overall chain growth is in the 5′ to 3′ direction. (b) The chemical sequencing (Maxam–Gilbert) method was developed by Maxam and Gilbert[22] and four sets of oligodeoxynucleotides are generated by placing a purified 3′ or 5′ end-labeled oligodeoxynucleotide in a base-specific chemical reagent which randomly cleaves DNA at one or two specific nucleotides.[20] In the first stage, bases are modified within the DNA molecule by using one of hydrazine, dimethyl sulfate or formic acid. Strand breakage is then brought about by adding piperidine to the nucleotides. This reaction cannot be allowed to go to completion because all possible cleavage products are needed in the complete mixture.

For both methods, electrophoretic methods such as slab gels and capillary electrophoresis (CE) are used to separate the reaction products and allow reading of the sequences.

2 MICROBIOLOGICAL ANALYSIS

In this section, we shall consider the increasing importance of nucleic acid analyses in bacteriology, virology and parasitology. Several examples will be discussed and references to further reading will be given along the way.

There are several advantages to the use of nucleic acid tests for diagnostic measurements. They can be used to confirm existing immunological tests; nucleic acids are usually stable to degradation; it is possible to amplify nucleic acids (e.g. by PCR), which is not the case with biological molecules such as proteins; and, finally, they are usually easy to develop and modify.[23] With specific reference to infectious pathogens, nucleic acid analyses also possess several advantages such as the potential to differentiate virulent and nonvirulent organisms based on genotype. Several examples will be encountered in succeeding sections. Also, the similarity between nucleic acid probes and related pathogens is very important in determining any genetic links of different pathogens.

It must be borne in mind that the technique is not perfect and several disadvantages may be encountered.[23] High background signals may be present and so it is difficult to detect weak positive samples. There are also cases where the additional sensitivity of nucleic acid tests may actually complicate the analysis of results when compared with culture methods, e.g. by identifying the pathogen or its components at concentrations which are epidemiologically unimportant.

2.1 Bacteriology

Many approaches have been utilized and many organisms have been detected by nucleic acid analysis. In this survey, brief introductions to some of the approaches are given.

2.1.1 Escherichia coli

Escherichia coli, and especially *E. coli* O157 : H7, has been widely studied. This species [which belongs to the group called verocytotoxigenic *Escherichia coli* (VTEC)] has been implicated in numerous outbreaks of food poisoning in the past decade, especially in the USA, Canada, the UK and Japan. The symptoms are particularly severe and include watery, followed by bloody, diarrhea, pus in stools and abdominal cramps. Eventually, conditions such as hemorrhagic colitis (HC) and hemolytic–uremic syndrome (HUS), both of which can lead to kidney failure and death, may result. Cattle are considered to be the major reservoir for O157 : H7, which is spread through fecal contamination of food.[24] Foods implicated in transmission have included ground beef, apple juice, potatoes, lettuce and cider.

Feng[25] has reported on the identification of O157 : H7 by using a DNA probe specific for an allele of the uid A gene. This was done by an oligonucleotide probe, PF-27, which contains a unique base substitution in the allele of the uid A gene. Using colony hybridization analysis of 239 bacteria, it was found that the probe reacted only with isolates of serotype O157 : H7. To verify probe specificity, DNA samples from various *E. coli* subspecies and O157 : H7 were digested with Hin f1 enzyme and examined by Southern blotting. The Southern analysis showed that PF-27 hybridized only with a fragment of about 700 base pairs in the O157 : H7 isolates. In a similar approach, Brunton et al.[26] used the *E. coli* attaching and effacing gene (eae) to detect O157 : H7 in cattle and humans. It was found that the majority of strains isolated from humans with bloody diarrhea and cattle with severe diarrhea were eae positive. In this paper, specific primers used to detect the eae sequences are described and it was suggested that a study of eae gene distribution is a useful method of distinguishing VTEC from other subtypes.

Several examples of the use of the PCR for detecting O157 : H7 have been reported. A group of workers at the FDA[27] have used PCR to detect low levels of the

bacterium in bovine feces. To improve the utility of PCR for this research, several enrichment culture procedures were carried out. It was determined that best results were obtained with boiling for 10 min when compared with enzyme digestion and phenol–chloroform extraction. Upon cooling to room temperature, RNase was added and this extract was used for PCR which involved amplification of primer sequences to detect the 614 base-pair fragment of the shiga-like toxin 1 (slt-1) gene. The authors suggested that boiling may destroy some PCR-inhibiting factor not removed by other procedures. By combining the boiling method with immunomagnetic separation (IMS), O157 : H7 was detected in about 8 h with a sensitivity of about 10^3 cfu g^{-1} feces.

Researchers at the United States Department of Agriculture (USDA)[28] used a multiplex PCR method to detect O157 : H7. In this case, a portion of the 60-MDa plasmid found in O157 : H7 was sequenced and a multiplex PCR was designed by employing primers specific for the eae A gene, conserved sequences of slt-1, slt-2 and the 60-MDa plasmid. This approach allowed the virulence genes (slt-1, slt-2 and eae A) and 60-MDa plasmid sequences to be targeted simultaneously. Thus, specific detection of O157 : H7 was obtained.

2.1.2 *Campylobacter spp.*

Campylobacters are a leading cause of enterocolitis in many developed countries and they have also been associated with bacteremia, reactive arthritis and meningitis.[29] Common sources of infection are unpasteurized milk, contaminated water and undercooked meats. Most infections are thought to be caused by *C. jejuni* and *C. coli*.[30]

Vanniasinkam et al.[31] recently reported a PCR method to detect *Campylobacter* spp. in clinical specimens. The PCR was based on the amplification of the 16S ribosomal RNA (rRNA) gene to detect the organism in human feces. The authors indicated that PCR detected twice as many positives as culture methods and it was proposed that by using rapid DNA extraction methods, PCR could be used as a routine diagnostic test for *Campylobacter*. PCR-based assays using 16S rRNA genes have also been developed for *Helicobacter pylori*,[32] the organism implicated as the causative agent in stomach ulcers.

Recently, it was discovered that surface antigens are important components required for the differentiating ability of serological tests and host recognition. With this in mind, Connerton and Connerton[33] have identified a gene encoding a membrane protein for *C. jejuni*. This was achieved following an immunoscreen of a lambda ZAP II genomic DNA library with antiserum raised against glycine-extractable proteins. Therefore, by using a molecular approach, the antigenic proteins from *C. jejuni* released after acid extraction were identified.

2.1.3 *Salmonella*

Salmonella is a Gram-negative bacterium which invades the mucous membrane and is spread by fecal–oral transmission. There are more than 2200 different *Salmonella* serotypes and most of these are human pathogens.[34] Typhoid fever caused by *S. typhi* leads to 16.6 million cases and 600 000 deaths annually, with the vast majority in Southeast Asia, Africa and South America.[35] *Salmonella* is the most important pathogen reported from foodborne illness in France with *S. enterica* being transmitted to humans by consumption of contaminated eggs and poultry products.[36]

PCR has been widely used as a diagnostic tool in detecting various *Salmonella* serotypes, mainly because it tends to overcome the difficulties encountered with culturing and enumerating these organisms from complex microbial communities. In one case, a primer set of oligonucleotides from the inv A gene has been evaluated.[37] Apart from being specific for *Salmonella* species, it was found that, after PCR amplification, it was possible to identify *S. typhimurium* by restriction enzyme analysis. It was suggested that this polymerase chain reaction/restriction enzyme (PCR/RE) method may be useful as a diagnostic tool for food analysis. In another case, a multiplex PCR with three sets of primers was developed for the detection of all serotypes of *S. enterica* and for the identification of *S. enteritidis* and *S. typhimurium*.[36] The authors found that a combined, modified semi-solid Rappaport Vassiliadis (MSRV) agar (Merck)/PCR method gave similar results to those of bacteriological methods and detection was possible in 2 d.

One interesting application for the differentiation of *S. typhi* and other species involves the use of a random amplified polymorphic DNA (RAPD) fingerprinting method.[35] Five primers were involved and it was concluded that the DNA of various strains of *Salmonella* could be discriminated in a few hours.

2.1.4 *Neisseria gonorrhoeae*

Gonorrhoea is one of the most important sexually transmitted diseases worldwide, with 720 000 cases in the USA in 1988.[38] Gynecological infections may cause infertility, chronic pelvic pain and ectopic pregnancy. There have also been reports of obstetric complications, neonatal pneumonia and conjunctivitis.

Over the past decade, DNA probe arrays have been developed for the detection of *Neisseria gonorrhoeae*, one of which makes use of a 2-h nonisotopic assay for direct detection in urogenital specimens.[37] This is the Gen-Probe test (PACE 2 from Gen-Probe Inc., San

Diego, CA, USA). It was found that with some slight modifications of the manufacturer's specifications, the DNA probe test was essentially equivalent to the culture method with respect to sensitivity and specificity. The test was found to be especially suitable for screening women for gonorrhoeal genital infections. The PACE 2 assay uses a chemiluminescent AE-labeled DNA probe which is specific for the rRNA of *N. gonorrhoeae*.[39] A urogenital swab specimen is placed in the PACE 2 transport medium whereby cells are immediately lysed thereby releasing their rRNA. The target rRNA is then hybridized to a labeled DNA probe and the extent of hybridization is measured in a luminometer.[39] Overall, the PACE 2 system is adequate for detection but it is unable to determine antibiotic resistance of positive specimens.

It has been reported recently that a micromachined chemical amplifier was successfully used to perform the PCR in continuous flow at high speed.[40] Input and output of DNA are continuous and amplification is independent of input concentration. A 20-cycle PCR amplification of a 176 base-pair fragment from the DNA gyrase gene of *N. gonorrhoeae* was performed in less than 20 min.

2.1.5 Mycobacterium tuberculosis

Tuberculosis (TB) is caused by members of the *Mycobacterium tuberculosis* complex and results in three million deaths a year worldwide. *M. tuberculosis* is transmitted from person to person via the airborne route. It is spread on droplet nuclei 1–5 μm in diameter which remain suspended in the air for long periods.[41] The HIV epidemic has led to an increase in the number of cases of TB, which causes bronchopneumonia, fibrosis and calcification. Other symptoms may include fever, night sweats, weight loss, anorexia, cough and blood in sputum.

A recent paper has reported on the identification and rifampin resistance testing of *Mycobacterium* species using high-density DNA probe arrays.[42] The array is based on two sequence databases: one for identification of mycobacteria (using 16S rRNA sequences) and the other for detecting rifampin resistance by targeting rpo B alleles. Species identification and rifampin resistance were determined by hybridizing fluorescently labeled, amplified genetic material generated from bacterial colonies to the array. This approach allows a number of genetic tests to be performed simultaneously but requires labeling of each sample to be tested.

It is also possible to detect and identify *M. tuberculosis* directly from sputum sediments by amplification of rRNA.[43] The results were available in 5 h (culture methods require about 6 weeks). This Gen-Probe Amplified *Mycobacterium tuberculosis* Direct Test™ uses enzymatic amplification of *M. tuberculosis* complex rRNA followed by detection of the amplicon (PCR product) with an AE-labeled DNA probe. The PCR assay involved amplification of insertion sequence IS 6110.

An integrated DNA analysis device has been used to measure nanolitre amounts of solutions containing DNA.[44] It amplifies the DNA and then separates and detects the products. This is done on a microchip 47 mm long, 5 mm wide and 1 mm high. The device requires a blue laser diode, an air pressure source and a programmable control unit. The fluids move around in microcapillaries and extra channels, valves and mixing chambers are needed for adding buffers, oligonucleotides, primers, DNA polymerase and fluorescent dyes. Amplification, purification and detection of a 106 base-pair fragment of *M. tuberculosis* required 1 h.

2.1.6 Bacterial Meningitis

Bacterial meningitis is a serious infection which needs rapid diagnosis and treatment, otherwise brain damage may occur. An evaluation of an extended diagnostic PCR assay for meningitis in cerebrospinal fluid (CSF) was performed using PCR.[45] Organisms evaluated were *Neisseria meningitidis*, *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Streptococcus agalactiae* and *Listeria monocytogenes*. A general bacterial amplicon from the 16S rRNA gene was obtained in a first step followed by species-specific regions. It was reported that the detection level was 4 fg of DNA per reaction, which corresponds to about one bacterial genome per reaction tube. This assay is able to detect bacteria in cultured suspensions as well as in clinical CSF samples.

2.1.7 Commercial Developments

Rapid pathogen detection using a microchip PCR array instrument has been reported.[46] The instrument is called the Advanced Nucleic Acid Analyzer® and uses 10 silicon reaction chambers with thin-film resistive heaters and solid-state optics. Features of the system include efficient heating and real-time monitoring, low power requirements and no moving parts. With this instrument, it is possible to analyze pathogens such as *Bacillus subtilis* and *Erwinia herbicola* vegetative cells.

2.2 Virology

The determination of the virus template sequence is enhanced by the genetic simplicity of viruses.[23] It is also of importance that the viral genes are well defined. In general, a conserved gene from a given virus pathogen is selected as a probe region because it should have the ability to detect various strains of the pathogen. In this brief account, the detection of viruses for hepatitis B and

C and for Epstein–Barr virus (EBV) infection will be discussed.

2.2.1 *Hepatitis C*

Hepatitis C virus (HCV) is an RNA virus which has been identified as the agent responsible for transfusion-associated and community-acquired non-A, non-B hepatitis.[47] It causes persistent infection in more than 90% of infected people and up to 70% of these will develop progressive liver disease over a 20–30-year period. An estimated 3.9 million Americans are currently infected with HCV.[48]

A rapid and reproducible method for assessment of the HCV load in serum has been reported.[47] This combines Taqman® technology (Roche) and the ABI Prism 7700™ (Perkin-Elmer) real-time sequence detection system. A reverse transcription PCR (this involves amplification of RNA by PCR by annealing a primer to the RNA template and then synthesizing a cDNA copy using RT, followed by PCR) which contains a dual-labeled fluorogenic probe to quantify the 5′ noncoding region of HCV was utilized. The probe contains a fluorescent reporter at the 5′ end and a fluorescent quencher at the 3′ end. When combined with the 5′–3′ nuclease activity of Taq polymerase, direct quantification of the PCR product is achieved. The authors concluded that the high sensitivity and reproducibility of real-time HCV RNA quantification allows the screening of a large number of samples, which makes it suitable for monitoring viral load during therapy. In a related method, Stapleton et al.[48] performed a comparison of HCV RNA detection in whole blood and plasma from 50 patients with chronic liver disease. The blood and plasma aliquots were independently tested for HCV RNA by RT PCR and it was found that whole-blood HCV RNA detection is more sensitive than commercially available tests.

2.2.2 *Hepatitis B*

Hepatitis B virus (HBV) is one of the causative agents of viral hepatitis, with an estimated 350 million people worldwide being chronic carriers and 100 million in China alone.[49] About 5–10% of carriers may be chronically infected and, of these, 25–40% may develop cirrhosis or liver cancer. HBV is a partially ds-DNA virus composed of a 42-nm outer shell and a 27-nm inner shell. A nucleic acid photo-cross-linking approach was used to develop a direct assay to quantify HBV DNA levels in serum. Cross-linker-modified DNA probes complementary to the viral genomes of the major HBV subtypes were synthesized and assayed in less than 6 h. It has also been reported that Naxcor (Menlo Park, CA, USA) have developed diagnostic tests for the detection of DNA and RNA targets based on the use of photo-cross-linked DNA probes.[50] It has been proposed[49] that cross-linking of DNA probes has certain advantages over conventional hybridization-type assays. These include the maximum amount of target will be captured and retained throughout the assay, leading to increased signal, and cross-linking allows the use of stringent wash conditions for effective removal of nonspecific background signals.

2.2.3 *Epstein–Barr Virus*

EBV causes infectious mononucleosis and EBV-related malignancies such as Burkitt's lymphoma and nasopharyngeal carcinoma. In immunocompromised individuals, such as acquired immune deficiency syndrome (AIDS) or transplant patients, EBV may cause B-cell lymphoma.[51] A real-time PCR assay was used to quantify the amount of EBV DNA in blood and it was indicated that $2–10^7$ or more copies of EBV DNA could be detected. By using real-time laser scanning coupled with a fluorogenic probe,[52] large numbers of amplified products were rapidly and accurately quantified.

2.3 Parasitology

Nucleic acid-based tests have also been applied to the detection of parasites. In this subsection we shall confine our attention mainly to *Plasmodium*, the agent responsible for malaria. More than 1.6 billion people live in malaria-endemic areas and imported malaria (via traveling) is fairly common. In 1991, there were 2332 cases reported in the UK with 1268 due to *Plasmodium falciparum*.[53] Malaria may cause fever, anemia, jaundice and splenomegaly along with renal failure.

Two PCR-based methods for diagnosing *Plasmodium falciparum* have been reported. One involves a nonradioactive system to detect the four species of parasite causing malaria.[54] Plasmodium-specific primers corresponding to the small-subunit rRNA genes of the parasite were used and a 291 base-pair fragment was amplified. It was possible to detect one parasite in 50 μL of whole blood. The second is an evaluation of a colorimetric assay to detect malaria in travelers.[55] A nonisotopic colorimetric PCR-based assay (Digene SHARP Signal System™) based on hybridization of a single-stranded RNA (ss-RNA) probe to denatured biotinylated PCR products and capture of the RNA–DNA hybrids on a streptavidin-coated microliter plate was developed. By using nested PCR (here, the larger fragment produced by the first round of PCR is used as the template for the second PCR) as the reference standard, it was determined that the colorimetric assay had a sensitivity of 100% and a specificity of 95.4%.

It should be noted that diagnostic tests have been, or are being, developed for visceral leishmaniasis

(caused by *Leishmania donovani*) based on cDNA probes.[56]

3 GENETIC DISEASES AND SCREENING

3.1 The Genetic Basis of Disease

Alterations in the genetic code (genetic defects) may be inherited or acquired. The former may involve a defect in a single gene or in many genes whereas the latter usually involve several genes.[57] These mutations may arise from injury by intrinsic or extrinsic agents. Since the development of nucleic acid probes, it has become much easier to detect alterations in genes. Alterations in the amount of chromosomal DNA may lead to developmental defects such as mental retardation.

One group of autosomal disorders is termed recessive. In these cases, the parents of affected individuals are not themselves affected but are carriers of the disease. This recessive disorder arises by inheriting two copies (one from each parent) of a defective gene. The most common autosomal recessive disorder is CF (see section 3.3.1), which occurs in Caucasian populations with a frequency of about 1 in 2500.[57] Other autosomal disorders are termed ''inborn errors of metabolism'', i.e. they involve a metabolic defect arising from deficiencies of a necessary enzyme.[56] Examples of this category include lipid storage disorders (e.g. Tay–Sachs disease), phenylketonuria (PKU) and abnormalities such as thalassemia, sickle-cell disease and hemophilia.

3.2 The Human Genome Project

It has been estimated that there are over 100 000 human genes (3 billion bases are involved in this genome).[58] As several diseases arise from genetic susceptibility, a map of the genes and their functions will assist in understanding diseases such as cancer, coronary heart disease and the genetic defects mentioned in section 3.1. The purpose of the Human Genome Project is to create a complete map and sequence of the human genome and work is being carried out in several major centers.[57] The task is enormous but it has already led to information on Huntington's disease, CF, breast cancer and AD. In March 1999, it was announced that the Human Genome Project will have its foundations completed by February 2000.[58] At that time, the working draft will be made publicly available. This effort is being undertaken by the Sanger Centre in Cambridge, UK, and laboratories funded by the National Human Genome Research Institute (NHGRI) in the USA, along with several other centers around the world. The final, high-quality sequence is expected to be completed by 2003.

3.3 A Survey of Some Genetic Diseases

3.3.1 Cystic Fibrosis

CF is an inherited disorder characterized by an abnormality of mucus production. The mucus is viscous and tends to clog bronchioles and pancreatic ducts, causing infection, inflammation and tissue damage. Eventually, death due to respiratory failure results.[59]

CF arises when there is a mutation in a gene on chromosome 7. This gene codes for the CF transmembrane conductance regulator which acts as a chloride channel. Thus, in epithelial cells there is impermeability to Cl^- ions which results in reduced production of Na^+ ions and water. Thus, the mucus becomes viscous and sweat is very salty.[59]

The gene for CF has been cloned and the most common mutant among whites is F 508.[60] Antenatal screening is carried out if a risk is known to exist, e.g. if a sibling has the disease. The test is usually performed at 10 weeks. By screening for F 508 and about five of the other most common mutations, about 85% of the cases can be identified.[58] Usually, screening for the F 508 allele makes use of PCR amplification followed by polyacrylamide gel electrophoresis (PAGE) to separate homo- and heteroduplexes (HD).[61] This test is now being combined with in vitro fertilization whereby the individual fertilized cells are screened for CF, affected embryos are discarded and normal embryos are used for implantation.

At present, research is being carried out with the aim of using gene therapy (i.e. inserting a normal working gene into cells containing a defective gene) to assist in providing treatment for CF and other genetic diseases. However, there are many problems, e.g. the gene and associated regulatory DNA must be isolated, sufficient DNA for treatment must be available, the gene has to be delivered to the nucleus of the cell and incorporation of the gene should not produce undesirable effects.[59]

3.3.2 Heriditary Fructose Intolerance

HFI is a genetic disorder resulting from mutation in the aldolase B gene which encodes the isozyme of fructose 1,6-biphosphate aldolase found in the liver, kidney and small intestine.[62] With this disorder, newborns and infants are intolerant of fructose in foods and this condition may lead to hypoglycemia, growth retardation and death. Standard methods of diagnosis are measurement of clinical symptoms upon intravenous fructose introduction and direct assay of aldolase activity in liver biopsy samples.[62] For newborns, the diagnosis is crucial, and rapid, noninvasive tests are required.

Research workers at Boston University[62] have amplified (by PCR) a region of genomic DNA from lymphocytes and prepared a biotin-tagged probe which is

hybridized to complementary poly(dT)-tailed allele-specific oligonucleotides (ASOs). The ASOs are bound to a nylon membrane and detection was by chemiluminescence. It was determined that the assay was able to detect two mutations which cause >70% of HFI worldwide. By using a reverse dot/blot methodology, it is believed that screening newborns is more easily achieved and simultaneous testing for several mutations can be carried out.

3.3.3 Hemochromatosis[63]

Hemochromatosis is an autosomal recessive disorder of iron metabolism which occurs in 0.2–0.5% of Caucasians. The disease leads to excessive accumulation of dietary iron, which may cause cirrhosis, cardiac failure, diabetes, arthritis and hepatocellular carcinoma. To prevent such serious conditions arising, it is vital to have early detection and treatment.

Recently, the HFE gene was identified as being responsible for hemochromatosis and so early genetic diagnosis is now possible, especially as two mutations have been detected in the gene. A multiplex PCR method has been developed which allows both mutations to be simultaneously detected. DNA was extracted from the blood of patients and PCR-mediated site-directed mutagenesis was used to create unique restriction sites in each product of a duplex PCR.

3.3.4 Phenylketonuria

PKU is also an autosomal recessive disorder which, if left untreated, may lead to mental retardation. In the USA, the incidence is from 1 in 10 000 to 1 in 25 000.[61] Phenylketonuriacs do not metabolize phenylalanine, which is an amino acid found in some foods. PKU arises as a result of a large number of mutations and, at present, no DNA analyses are available. This area is expected to become an important research activity.

3.4 Single Nucleotide Polymorphism

In April 1999, it was announced that an innovative collaboration between 10 pharmaceutical companies, the Wellcome Trust and leading academic centers will allow for the identification and analysis of the genetic markers known as SNPs, pronounced ''snips''.[64] The map will be used to identify specific genes involved in diseases, developing novel diagnostic tests and creating new ''personalized'' medicines based on an understanding of small genetic variations.

SNPs are common variations which occur in human DNA and scientists believe that these can assist in determining who will develop diseases such as cancer, AD, diabetes, arthritis and depression.[64] The map will allow answers to be obtained about the genetic factors responsible for causing disease.

4 OTHER AREAS OF HUMAN MEDICINE

4.1 Introduction

DNA probes are finding importance in various areas of human medicine as they are able to identify specific biological molecules associated with pathological conditions.[65] Some aspects of human medicine have already been covered in sections 2 and 3. In this section, special attention will be paid to just two areas: oncology and AD. Readers who require more detailed surveys are referred to Versalovic and Lupski.[65]

4.2 Oncology

Molecular probes can be used to diagnose tumors and to establish cell lineages and also to permit tumor grading.[65] It has also been established that nucleic acid probes may be used with biopsy material for in situ hybridization procedures to obtain more data from a specimen. It is also possible to combine DNA amplification methods with cytology to obtain specific diagnoses.[65]

4.2.1 Breast Cancer

In some countries, it has been estimated that up to 10% of the women may be affected by breast cancer, the strongest risk factor being family history.[66] In the UK, about 25 000 new cases are diagnosed each year and about 15 000 of these women will die. The genetic alterations responsible for breast cancer usually involve one of two genes, BRCA 1 and BRCA 2. The former was isolated in 1994 and since then more than 300 disruptive mutations have been identified in breast and ovarian cancer.[67] In Australia, it has been reported that the proportion of women with breast cancer before age 40 who carry a germline mutation in BRCA 1 was estimated at 3.8%.[68] Many mutations can occur in these genes and it is difficult to identify women who may be most at risk based on a single genetic defect. However, the use of DNA probes has found a valuable role in detection of important parameters.

To determine whether mutations may account for reduced expression of BRCA 1, Catteau et al.[67] screened the promoter region by sequencing. Although no mutations were detected, it was discovered that a new polymorphism consisting of a C to G base change within the β-promoter was present and that it is unlikely to play any role in disease. However, it was found to be suitable for rapid screening for genomic deletions within BRCA 1. The method involved amplifying the promoter region of

BRCA 1 by PCR, using the tumor DNA (obtained from breast cancer patients) as a template. Detection was by the amplification refractory mutation system (ARMS).[69] This allows rapid screening of germline mutations and may also be used to determine loss of heterozygosity in sporadic cancer.

Eccles et al.[70] screened 230 women for BRCA 1 mutations in order to establish the parameters on which to base a regional mutation analysis strategy. Mutation analysis by HD and single-strand conformational polymorphism (SSCP) was performed. Ten different mutations were detected in 17 families with three of these accounting for 59% of the cases. It has been proposed that mutation analysis may become more widely available if it can become cost-effective and if more detailed analysis could be attempted.

Overexpression of the c-myc gene leads to cancer development, especially of the breast.[71] Differential PCR and competing PCR have been developed for gene dosage quantification but both have a tendency to degrade DNA, whereas double-differential PCR, competitive-differential PCR and differential PCR are labor-intensive, expensive and time-consuming. With differential and competitive PCR, both degraded DNA and low competitor concentrations lead to over- or underestimation of gene dosages. In order to overcome these problems, methods using separate comparisons of two single-copy reference genes for any DNA sample (double-differential PCR), dilution series of the competitors (competitive-differential PCR) or sample DNA (differential PCR) have been developed. However, a direct double-differential PCR method was found to be sensitive and reliable.[71] In this approach, DNA fragments from two different single-copy reference genes and the target DNA fragment (c-myc) are amplified simultaneously in the same tube. The method, which involves the use of laser-induced CE is fast and allows the quantification of gene dosages without the use of synthetic competitors even when degraded DNA is present.

4.2.2 Colorectal Cancer[72]

In Europe, there are 210 000 new cases of colorectal cancer annually of which 124 500 will die. It is the second most common malignancy in men and the third most common in women. About a third of the patients have a primary tumor in the rectum with the majority of the remaining tumors occurring in the sigmoid colon. The prognosis is influenced by many factors such as age, gender, duration of symptoms, presence of bowel obstruction, tumor location, need for blood transfusion and the quality of surgical intervention.

A prognostic marker is a variable which provides prospective information on patient outcome which is complementary to the data obtained by the pathologist from the diagnostic sections and on which therapeutic decisions can be guided. The *ras* family is one of the most commonly detected oncogenes in human cancer (examples of *ras* genes are K-*ras*, H-*ras* and N-*ras*). The *ras* genes encode for 21-kDa proteins (p21S) which are found on the inner surface of the plasma membrane. *Ras* mutations may be an early event in leading to colorectal tumors.

PCR methods for the detection of *ras* mutations in DNA isolated from colorectal tumors have allowed prognosis to be determined faster. A *ras* mutation at codon 12 or 13 is associated with a poor prognosis. However, a poorer prognosis in patients with mutations in both *ras* and p53 was discovered.[73] The p53 tumor-suppressor gene is a commonly altered gene in human solid tumors.

4.3 Alzheimer's Disease

AD destroys nerve cells in the brain, which results in memory loss, decline of cognitive ability and incontinence in the later stages. An estimated 10% of people over 65 and about 50% over 85 suffer from some degree of AD.[74]

Abnormal genes for apolipoprotein E (apoE) appear to be a risk factor for the disease. ApoE codes for a family of lipoproteins and is involved in lipid transport in the brain.[74] On its own, apoE genotyping is not sensitive or specific enough to be used as a diagnostic test for AD, but it is used in classifying type III hyperlipidemia and in detecting genetic polymorphisms.[75] It has also been shown that the inheritance of one or more apoE 4 alleles increases the risk of developing artherosclerosis and AD.

Several apoE genotyping techniques have been used and these include minisequencing, SSCP, ASO probes and ARMS.[75] The ARMS technique, which comprises two PCRs each carried out using the same substrate DNA, is advantageous as it is rapid, easy to perform, nonisotopic and requires little sample manipulation.

5 THE BIOSENSOR APPROACH

5.1 Introduction

Figure 8 shows the main regions of a generalized sensor. The most important part of a sensor is the selective membrane/layer which interacts with the sample to produce useful information about the analyte.[76] To allow this, the interaction of the layer with the sample must be selective, a process which is achieved by immobilizing receptor sites for the analyte species in the layer. Mechanisms used for this include ion–cavitand "best-fit" interaction, electrostatic interactions, acid–base reactions, redox reactions, antibody–antigen interactions,

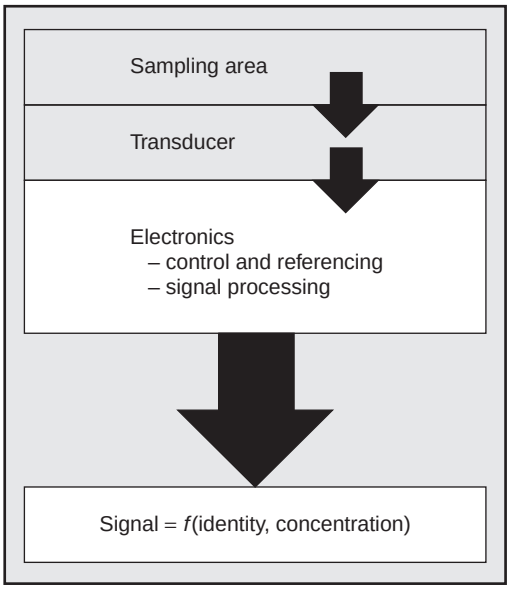

Figure 8 The main parts of a sensor. (Reproduced by permission of Dr D.C. Stone, Department of Chemistry, University of Toronto.)

enzyme–substrate interactions, ion-exchange reactions and nucleic acid hybridization.

5.2 Properties of an Ideal Sensor[76]

An ideal sensor should be specific for the analyte, sensitive to changes in concentration of the analyte, able to produce a rapid response, sturdy, i.e. it should have a lifetime of several months, small and of high precision.

5.3 Types of Sensors

The various types of sensor as classified by transduction categories are electrochemical (parameters monitored include potential, current and conductivity), piezoelectric (mass, viscosity, elasticity, density, conductivity), optical (absorbance, fluorescence, reflectance) and thermal (enthalpy, pyroelectric effect).

5.4 Biosensors for Nucleic Acids

The development of gene probe biosensors has been undergoing rapid technological modification, although it may still be considered to be in a state of infancy. Hybridization reactions have been detected with piezoelectric acoustic wave devices (the first detection was by a piezoelectric resonance sensor[77]) and the evanescent wave methods of total internal reflection fluorescence (TIRF) and surface plasmon resonance (SPR).[78] Readers who are interested in electrochemical devices are referred to references Millan and Mikkelsen[79] and Wilson.[80]

5.4.1 Acoustic Wave Methods

The launching of acoustic waves in solids at ultrasonic frequencies depends on the conversion of electromagnetic into acoustic energy.[81] The acoustic wave generated in piezoelectric materials is determined by crystal cut, thickness of material used and geometry and configuration of the metal electrodes employed to generate the electric field.[81]

Examples of acoustic wave devices include the thickness-shear mode (TSM) sensor (Figure 9), whereby the well-studied avidin–biotin interaction is utilized to measure frequency changes when the mass of substance on the sensor changes. It is also possible to have nonpiezoelectric acoustic wave devices such as the thin-rod acoustic wave (TRAW) sensor. Further details may be obtained in Thompson and Stone.[81] Recent uses of acoustic wave methods have included the detection of hybridization of a biotinylated 25-mer oligonucleotide probe with complementary and single-base mutated oligonucleotides[82] and a study of the various interactions of HIV-1 TAR RNA with tat-derived peptides[83] using a similar approach to that indicated for complementary strands.

TIRF has been used to monitor specific hybridization of labeled oligonucleotides whereby amino-ended 16- or 20-base DNA probes were covalently attached to the surface of optical fibers and hybridization to complementary fluorescein-ended oligonucleotides was detected.[84]

Measurement of nucleic acid hybridization by SPR has also been reported and, in one case, 17- or 50-base oligonucleotide probes were attached to silver-coated waveguides. It was found that a 97-base target sequence could be readily detected.[78]

5.5 DNA and Oligonucleotide Probe Arrays (Gene Chips)[80,85]

Probe arrays involving either DNA or oligonucleotides have been briefly mentioned in sections 2.1.4, 2.1.5 and 2.1.7. In this section, an overview of this potentially exciting area will be provided.

High-density arrays allow rapid sequence analysis of genomic samples and may also provide diagnosis of genetic and infectious diseases. The methodology

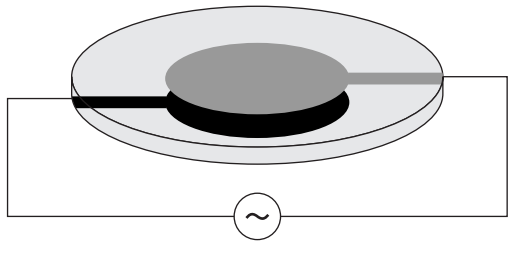

Figure 9 TSM sensor.

employed involves a combination of sequencing by hybridization (SBH), light-directed spatially addressable interrogation, combinatorial chemistry, confocal fluorescence microscopy, robotics and PCR. Advantages of this approach include reduced analysis times through multiplexing, the ability to produce any type of probe sequence and the possibility of having reusable chips which may be incorporated into nanofabricated structures.

Probe arrays have been utilized in several areas, some of which include the development of diagnostic tests for β-thalassemia,[86] studies of the breast cancer gene BRCA 1,[87] simultaneous analysis of the entire human mitochondrial genome[88] and monitoring of metabolic and genetic control of gene expression.[89]

6 CURRENT AND FUTURE DEVELOPMENTS

At present, many approaches to nucleic acid analyses are undergoing research in laboratories around the world. In this section, some of the interesting areas will be briefly discussed so as to give the reader a better perspective of novel developments.

A molecular hula hoop has been developed which is claimed to outperform PCR. The technique, called rolling circle amplification, uses a loop of DNA which sticks to either side of a target sequence and produces millions of copies of it until detection is achieved. This test may prove invaluable in forensic science.[90]

With respect to tumors, it is known that growth of these is dependent on angiogenesis. This has led to the development of new approaches to treatment and new agents directed towards tumor vasculature. One approach involves the use of endogenous inhibitory proteins to counter the angiogenic stimulus produced by tumors, an example of which involves gene transfer of DNA to angiogenesis inhibitors such as angiostatin or platelet factor 4.[91] Another approach to cancer treatment involves the delivery of genetically engineered cures directly to tumors through keyhole surgery. It is believed that this method could reduce unpleasant side effects. Based at the University of Dundee, the project is headed by Professor Sir Alfred Cuschieri, a pioneer of keyhole surgery, and Professor David Lane, who discovered the p53 gene which regulates the production of anti-cancer proteins.[92] A further development in chemotherapy involves the use of cisplatin, widely used to treat testicular and ovarian cancers. It binds to DNA in tumor cells but it also affects DNA in other rapidly dividing cells leading to toxicity. At the University of Edinburgh, Peter Sadler[93] has modified the drug so that it contains Pt(IV), which does not bind to DNA. However, when exposed to blue laser light, it reverts to Pt(II), which can now bind to DNA. By giving the drug to a patient and then illuminating the tumor cells with lasers, it may be possible to kill tumors without damage to other tissues. A 10-year investigation, involving 7500 adults, has just begun in Merseyside, UK, to see if people at high risk of developing lung cancer can be detected by a simple genetic test. Professor John Field believes that genetic alterations which can be detected in sputum may precede the development of lung cancer.[94]

As indicated previously, gene therapies are starting to have a major impact on the treatment of diseases. Research is continuing apace, some of which will be mentioned here. At several universities, researchers have demonstrated that transfer of the normal CF TR gene to nasal tissues of CF mice could restore some normal gene function. Although the mouse is not an ideal model for human CF, it marks the start of what could be an important breakthrough.[95] At Imperial College in London, Eric Alton and colleagues have shown that it is possible to deliver the gene to the lungs of patients with CF. At the Roslin Institute (where Dolly the cloned sheep was created), along with technology from Geron Corporation of California, research is under way to create transplantable, tissue-matched cells which could be used to provide treatments for diabetes, AD, Parkinson's disease and some leukemias.[96] It has also been reported that whole functional genes can also be added to cells and that pressure-treated veins can be successfully used in human heart-bypass patients.[97]

Transgenic therapies are also being investigated in the hope of ''creating'' human babies for infertile couples. At Imperial College, Robert Winston, a pioneer of fertility studies, is researching the effectiveness of inserting genes into sperm rather than into embryos.[98] In a similar approach, biologists in Hawaii have devised a method of producing transgenic animals. Their technique involves freezing and thawing sperm, mixing them with the foreign genes and then injecting them into eggs.[99]

Antiviral gene therapy, which stops the virus from replicating or expressing its genes, is also assuming increased importance. One method uses complementary ''antisense'' molecules: strips of nucleic acid bind to the DNA or RNA of the viruses to form a double helix. When this occurs, the RNA cannot produce protein molecules. Antisense molecules have been found to inhibit strongly HBV replication. Another approach is to use ribozymes, catalytic RNA molecules, which are involved in reactions which break down other RNA molecules. These have been found to be effective against HBV in vitro.[100]

Although much of this section did not deal with nucleic acid analysis per se, discussion of gene therapies was included as these developments all require accurate

methods for identifying and quantifying nucleic acid sequences.

ACKNOWLEDGMENTS

We express our gratitude to Dr David Stone of the Department of Chemistry, University of Toronto, for his assistance with section 5 and to Ms Blanca Granozio for her help with typing of the manuscript.

ABBREVIATIONS AND ACRONYMS

AD	Alzheimer's Disease
AE	Acridinium Ester
AIDS	Acquired Immune Deficiency Syndrome
ARMS	Amplification Refractory Mutation System
ASO	Allele-specific Oligonucleotide
CE	Capillary Electrophoresis
CF	Cystic Fibrosis
CSF	Cerebrospinal Fluid
dNTP	Deoxynucleoside Triphosphate
ds-DNA	Double-stranded DNA
EBV	Epstein–Barr Virus
FDA	Food and Drug Administration
FISH	Fluorescent In Situ Hybridization
HBV	Hepatitis B Virus
HC	Hemorrhagic Colitis
HCV	Hepatitis C Virus
HD	Heteroduplex
HFI	Heriditary Fructose Intolerance
HIV	Human Immunodeficiency Virus
HPA	Hybridization Protection Assay
HUS	Hemolytic–Uremic Syndrome
IMS	Immunomagnetic Separation
mRNA	Messenger RNA
MSRV	Modified Semi-solid Rappaport Vassiliadis
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PCR/RE	Polymerase Chain Reaction/ Restriction Enzyme
PKU	Phenylketonuria
RAPD	Random Amplified Polymorphic DNA
rRNA	Ribosomal RNA
RT	Reverse Transcriptase
SBH	Sequencing by Hybridization
SNP	Single Nucleotide Polymorphism
SPR	Surface Plasmon Resonance
SSCP	Single-strand Conformational Polymorphism
ss-DNA	Single-stranded DNA
ss-RNA	Single-stranded RNA
TB	Tuberculosis
TdT	Terminal Deoxynucleotidyl Transferase
TIRF	Total Internal Reflection Fluorescence
TRAW	Thin-rod Acoustic Wave
TRIS	Tris(hydroxymethyl)aminomethane
TSM	Thickness-shear Mode
USDA	United States Department of Agriculture
UV	Ultraviolet
VTEC	Verocytotoxigenic *Escherichia coli*

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Biosensor Design and Fabrication • DNA Arrays: Preparation and Application • Electroanalytical Chemistry in Clinical Analysis • Micro Total Analytical Systems in Clinical Chemistry

Forensic Science **(Volume 5)**
DNA Extraction Methods in Forensic Analysis • Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Probes • Polymerase Chain Reaction and Other Amplification Systems • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

Electroanalytical Methods **(Volume 11)**
Microbalance, Electrochemical Quartz Crystal

REFERENCES

1. R.L.P. Adams, J.T. Knowles, D.P. Leader, *The Biochemistry of the Nucleic Acids*, 11th edition, Chapman & Hall, London, 5–39, 1992.
2. S. Neidle, *DNA Structure and Recognition*, IRL Press, Oxford, 15–27, 1994.
3. A.K. Bej, 'Nucleic Acid Hybridizations: Principles and Strategies', in *Nucleic Acid Analysis: Principles and Bioapplications*, ed. C.A. Dangler, Wiley-Liss, New York, 1–29, 1996.
4. J. Marmur, P. Doty, 'Determination of the Base Composition of DNA from its Thermal Denaturation', *J. Mol. Biol.*, **5**, 109–118 (1962).
5. E.R. Unger, M.A. Piper, 'Nucleic Acid Biochemistry and Diagnostic Applications', in *Tietz Textbook of Clinical*

Chemistry, 3rd edition, eds. C.A. Burtis, E.R. Ashwood, Saunders, Philadelphia, 421–443, 1999.
6. R.F. Lamb, G.D. Birnie, 'Strategies for Using Gene Probes', in *Gene Probes 1: A Practical Approach*, eds. B.D. Hames, S.J. Higgins, IRL Press, Oxford, 1–29, 1995.
7. G.M. Blackburn, M.J. Gait, *Nucleic Acids in Chemistry and Biology*, 2nd edition, Oxford University Press, Oxford, 1996.
8. M.L.M. Anderson, 'Hybridization Strategy', in *Gene Probes 2: A Practical Approach*, eds. B.D. Hames, S.J. Higgins, IRL Press, Oxford, 1–29, 1995.
9. D. Gillespie, S. Speigelman, 'A Quantitative Assay for DNA/RNA Hybrids with DNA Immobilized on a Membrane', *J. Mol. Biol.*, **12**, 829–842 (1965).
10. D.F. Cowan, J.P. Olano, *A Practical Guide to Clinical Laboratory Testing*, Blackwell, Boston, 362–368, 1997.
11. P.W.J. Rigby, M. Dieckmann, C. Rhodes, P. Berg, 'Labeling Deoxyribonucleic Acid to High Specificity In Vitro by Nick Translation with DNA Polymerase I', *J. Mol. Biol.*, **113**, 237–251 (1977).
12. A.P. Feinberg, B. Vogelstein, 'A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity', *Anal. Biochem.*, **132**, 6–13 (1983).
13. C. Kessler, 'Non-radioactive Labeling and Detection of Nucleic Acid Probes', in *Gene Probes 1: A Practical Approach*, eds. B.D. Hames, S.J. Higgins, IRL Press, Oxford, 93–144, 1995.
14. N.C. Nelson, 'Rapid Detection of Genetic Mutations Using the Chemiluminescent Hybridization Protection Assay (HPA): Overview and Comparison with Other Methods', *Crit. Rev. Clin. Lab. Sci.*, **35**(5), 369–414 (1998).
15. C.R. Newton, A. Graham, *PCR*, 2nd edition, Bios Scientific, Oxford, 1–8, 1997.
16. R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T.Horn, H.A. Erlich, N. Arnheim, 'Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia', *Science*, **230**, 1350–1354 (1985).
17. R.A. Eeles, A.C. Stamp, *Polymerase Chain Reaction (PCR): The Technique and its Application*, ed. R.G. Landes, Austin, TX, 4–11, 1993.
18. D.M. Coen, R.L. Dorit, O. O'hara, C.B. Hwang, 'The Polymerase Chain Reaction', in *Current Protocols in Molecular Biology*, eds. F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, Wiley, New York, Chapter 15, 15.0.1–15.2.11, 1993.
19. K.K. Kidd, G. Ruano, 'Optimizing PCR', in *PCR 2: A Practical Approach*, eds. M.J. McPherson, B.D. Hames, G.R. Taylor, IRL Press, Oxford, 1–22, 1995.
20. F.M. Ausubel, L.M. Albright, 'DNA Sequencing', in *Current Protocols in Molecular Biology*, eds. F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Strul, Wiley, New York, 7.0.1–7.0.13, 1993.
21. F. Sanger, S. Nicklen, A.R. Coulson, 'DNA Sequencing with Chain-terminating Inhibitors', *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467 (1977).
22. A.M. Maxam, W. Gilbert, 'A New Method for Sequencing DNA', *Proc. Natl. Acad. Sci. USA*, **74**, 560–564 (1977).
23. W.C. Wilson, 'Approaches to Infectious Disease Diagnosis', in *Nucleic Acid Analysis: Principles and Bioapplications*, ed. C.A. Dangler, Wiley-Liss, New York, 105–129, 1996.
24. S.A. Renwick, J.B. Wilson, R.C. Clarke, 'Evidence of Direct Transmission of *Eschericia coli* O157:H7 Infection Between Calves and a Human', *J. Infect. Dis.*, **168**, 792–793 (1993).
25. P. Feng, 'Identification of *Escherichia coli* Serotype O157:H7 by DNA Probe Specific for an Allele of uid A Gene', *Mol. Cell. Probes*, **7**, 151–154 (1993).
26. M. Louie, J. de Azavedo, R. Clarke, A. Borczyk, H. Lior, M. Richter, J. Brunton, 'Sequence Heterogeneity of the eae Gene and Detection of Verotoxin-producing *Escherichia coli* Using Serotype-specific Primers', *Epidemiol. Infect.*, **112**, 449–461 (1994).
27. D.S. Stewart, M.L. Tortorello, S.M. Gendel, 'Evaluation of DNA Preparation Techniques for Detection of the slt-1 Gene of *Escherichia coli* O157:H7 in Bovine Faeces Using the Polymerase Chain Reaction', *Lett. Appl. Microbiol.*, **26**, 93–97 (1998).
28. P.M. Fratamico, S.K. Sacitey, M. Wiedmann, M.Y.Deng, 'Detection of *Escherichia coli* O157:H7 by Multiplex PCR, *J. Clin. Microbiol.*, **33**, 2188–2191 (1995).
29. M.B. Skirrow, 'Foodborne Illness: *Campylobacter*', *Lancet*, **336**, 921–923 (1990).
30. Advisory Committee on the Microbiological Safety of Food, 'Interim Report on *Campylobacter*', HMSO, London, 1993.
31. T. Vanniasinkam, J.A. Lanser, M.D. Barton, 'PCR for the Detection of *Campylobacter* spp. in Clinical Specimens', *Lett. Appl. Microbiol.*, **28**, 52–56 (1999).
32. J. Weiss, J. Mecca, E da Silva, D. Gassner, 'Comparison of PCR and Other Diagnostic Techniques for Detection of *Helicobacter pylori* Infection in Dyspeptic Patients', *J. Clin. Microbiol.*, **32**, 1663 (1994).
33. P.L. Connerton, I.F. Connerton, 'Identification of a Gene Encoding an Immuno-reactive Membrane Protein from *Campylobacter jejuni*', *Lett. Appl. Microbiol.*, **28**, 233–237 (1999).
34. J. Kwang, E.T. Littledike, J.E. Keen, 'Use of the Polymerase Chain Reaction for *Salmonella Detection*', *Lett. Appl. Microbiol.*, **85**, 693–702 (1998).
35. Y.H. Shangkuan, H.C. Lan, 'Application of Random Amplified Polymorphic DNA Analysis to Differentiate Strains of *Salmonella typhi* and Other *Salmonella* Species', *Lett. Appl. Microbiol.*, **85**, 693–702 (1998).

36. C. Soumet, G. Ermel, N. Rose, V. Rose, P. Drouer, G. Salvet, P. Colin, 'Evaluation of a Multiplex PCR Assay for Simultaneous Identification of *Salmonella* sp., *Salmonella enteritidis* and *Salmonella typhimurium* from Environmental Swabs of Poultry Houses', *Lett. Appl. Microbiol.*, **28**, 113–117 (1999).
37. L. Cocolin, M. Manzano, C. Cantoni, G. Comi, 'Use of Polymerase Chain Reaction and Restriction Enzyme Analysis to Directly Detect and Identify *Salmonella typhimurium* in Food', *J. Appl. Microbiol.*, **85**, 673–677 (1998).
38. E.S. Panke, L.I. Yang, P.A. Leist, P. Magevney, R.J. Fry, R.F. Lee, 'Comparison of Gen-Probe DNA Probe Test and Culture for the Detection of *Neisseria gonorrhoeae* in Endocervical Specimens', *J. Clin. Microbiol.*, **29**(5), 883–888 (1991).
39. Y.M. Hale, M.E. Melton, J.S. Lewis, D.E. Willis, 'Evaluation of the PACE 2 *Neisseria gonorrhoeae* Assay by Three Public Health Laboratories', *J. Clin. Microbiol.*, **31**(2), 451–453 (1993).
40. M.U. Kopp, A.J. de Mello, A. Manz, 'Chemical Amplification: Continuous Flow PCR on a Chip', *Science*, **280**, 1046–1048 (1998).
41. S.H. Gillespie, '*Mycobacterium tuberculosis*', in *Principles and Practice of Clinical Bacteriology*, eds. A.M. Emmerson, P.M. Hawkey, S.H. Gillespie, Wiley, Chichester, 225, 1997.
42. A. Troesch, H. Nguyen, C.G. Miyada, S. Desvarenne, T.R. Gingeras, P.M. Kaplan, P. Cros, C. Mabilat, 'Mycobacterium Species Identification and Rifampin Resistance Testing with High-density DNA Probe Arrays', *J. Clin. Microbiol.*, **37**(1), 49–55 (1999).
43. V. Jonas, M.J. Alden, J.I. Curry, K. Kamisango, C.A. Knott, R. Lankford, J.M. Wolfe, D.F. Moore, 'Detection and Identification of *Mycobacterium tuberculosis* Directly from Sputum Sediments by Amplification of rRNA', *J. Clin. Microbiol.*, **31**(9), 2410–2416 (1993).
44. M.A. Burns, B.N. Johnson, S.N. Brahmasandra, K. Handique, J.R. Webster, M. Krishnan, T.S. Sammarco, P.M. Man, D. Jones, D. Heldsinger, C.H. Mastrangelo, D.T. Burke, 'An Integrated Nanoliter DNA Analysis Device', *Science*, **282**, 484–487 (1998).
45. A. Backman, P.G. Lontz, P. Radstrom, P. Olcen, 'Evaluation of an Extended Diagnostic PCR Assay for Detection and Verification of the Common Causes of Bacterial Meningitis in CSF and Other Biological Samples', *Mol. Cell. Probes*, **13**, 49–60 (1999).
46. P. Belgrader, W. Bennett, D. Hadley, G. Long, R. Mariella, Jr, F. Milanovich, S. Nasarabadi, W. Nelson, J. Richards, P. Stratton, 'Rapid Pathogen Detection Using a Microchip PCR Array Instrument', *Clin. Chem.*, **44**(10), 2191–2194 (1998).
47. M. Martell, J. Gomez, J.I. Esteban, S. Sauleda, J. Quer, B. Cabot, R. Esteban, J. Guardia, 'High Throughput Real-time Reverse Transcription–PCR Quantitation of Hepatitis C Virus RNA', *J. Clin. Microbiol.*, **37**(2), 327–332 (1999).
48. J.T. Stapleton, D. Klinzman, W.N. Schmidt, M.A. Pfaller, P. Wu, D.R. La Brecque, J.Q. Han, M.J.P. Phillips, R. Woolson, B. Alden, 'Prospective Comparison of Whole-blood and Plasma-based Hepatitis C Virus RNA Detection Systems: Improved Detection Using Whole Blood as the Source of Viral RNA', *J. Clin. Microbiol.*, **37**(3), 484–489 (1999).
49. V.C.H. Lai, R. Guan, M.L. Wood, S.K. Lo, M.F. Yuen, C.L. Lai, 'Nucleic Acid-based Cross-linking Assay for Detection and Quantification of Hepatitis B Virus DNA', *J. Clin. Microbiol.*, **37**(1), 161–164 (1999).
50. M. Wood, D. Albagli, P. Cheng, B. Huan, R. Van Alta, 'Nucleic Acid Crosslinking Probes for DNA/RNA Diagnostics', *Clin. Chem.*, **42**, S196 (1996).
51. H. Kimura, M. Morita, Y. Yabuta, K. Kuzushima, K. Kato, S. Kojima, T. Matsuyama, T. Morishima, 'Quantitative Analysis of Epstein–Barr Virus Load by Using a Real-time PCR Assay', *J. Clin. Microbiol.*, **37**(1), 132–136 (1999).
52. C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, 'Real-time Quantitative PCR', *Genome Res.*, **6**, 986–994 (1996).
53. P.L. Chiodini, S. Clark, 'Malaria', in *Medical Parasitology: A Practical Approach*, eds. S.H. Gillespie, P.M. Hawkey, IRL Press, Oxford, 13–31, 1995.
54. L. Ciceron, G. Jaureguiberry, F. Gay, M. Danis, 'Development of a Plasmodium PCR for Monitoring Efficacy of Antimalarial Treatment', *J. Clin. Microbiol.*, **37**(1), 35–38 (1999).
55. K.J.Y. Zhong, K.C. Kain, 'Evaluation of a Colorimetric PCR-based Assay to Diagnose *Plasmodium falciparum* Malaria in Travellers', *J. Clin. Microbiol.*, **37**(2), 339–341 (1999).
56. J.E. Williams, 'Leishmania and Trypanosoma', in *Medical Parasitology: A Practical Approach*, eds. S.H. Gillespie, P.M. Hawkey, IRL Press, Oxford, 152–161, 1995.
57. S.L. Mera, *The Genetic Basis of Disease in Understanding Disease: Pathology and Prevention*, ed. S.L. Mera, Thornes, Cheltenham, 63–100, 1997.
58. The Wellcome Trust Press Release, 15 March 1999, http://www.wellcome.ac.uk.
59. S.L. Mera, 'Cystic Fibrosis', in *Understanding Disease: Pathology and Prevention*, ed. S.L. Mera, Thornes, Cheltenham, 648–669, 1997.
60. B. Kerem, J.M. Rommens, J.A. Buchanan, D. Markiewicz, T.R. Cox, A. Chakravanti, M. Buchwald, L.C. Tsui, 'Identification of the Cystic Fibrosis Gene: Genetic Analysis', *Science*, **245**, 1073–1080 (1989).
61. L.L. McCabe, R.B. McCabe, 'Approaches to Genetic Analysis: Genetic Disease and Screening', in *Nucleic Acid Analysis: Principles and Bioapplications*, ed. C.A. Dangler, Wiley-Liss, New York, 67–77, 1996.
62. J. Lau, D.R. Tolan, 'Screening for Hereditary Fructose Intolerance Mutations by Reverse Dot–Blot, *Mol. Cell. Probes*, **13**, 35–40 (1999).

63. M.K. Scott, A.P. Fellowes, J.D. Upton, M.J. Burt, P.M. George, 'Simple Multiplex PCR for the Simultaneous Detection of the C282Y and H63D Hemochromatosis (HFE) Gene Mutations', *Clin. Chem.*, **45**(3), 426–428 (1999).
64. The Wellcome Trust Press Release, 15 April 1999, http://www.wellcome.ac.uk.
65. J. Versalovic, J.R. Lupski, 'Diagnostic Nucleic Acid Probes in Human Medicine: Applications, Present and Future', in *Nucleic Acid Analysis: Principles and Bioapplications*, ed. C.A. Dangler, Wiley-Liss, New York, 157–202, 1996.
66. S.L. Mera, 'Breast Cancer', in *Understanding Disease: Pathology and Prevention*, ed. S.L. Mera, Thornes, Cheltenham, 446–465, 1997.
67. A. Catteau, C.F. Yu, M.A. Brown, S. Hodgson, J.Greenman, C.G. Mathew, A.M. Dunning, E. Solomon, 'Identification of a G/C Polymorphism in the Promoter Region of the BRCA 1 Gene and Its Use as a Marker for Rapid Detection of Promoter Deletions', *Br. J. Cancer*, **79**(5/6), 759–763 (1999).
68. M.C. Souther, A. Tesoriero, C.R. Anderson, K.M. Jennings, S.M. Brown, G.S. Dite, M.A. Jenkins, R.H.Osborne, J.A. Maskiell, L. Porter, G.G. Giles, M.R.E.M.C. Credie, J.L. Hopper, D.J. Venter, BRCA 1 Mutations and Other Sequence Variants in a Population-based Sample of Australian Women with Breast Cancer', *Br. J. Cancer*, **79**(1), 34–39 (1999).
69. C.R. Newton, A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, A.F. Markham, 'Analysis of Any Point Mutation in DNA. The Amplification Refractory Mutation System (ARMS)', *Nucleic Acids Res.*, **17**, 1503–1516 (1989).
70. D.M. Eccles, P. Englefield, M.A. Soulby, J.E. Campbell, 'BRCA 1 Mutations in Southern England', *Br. J. Cancer*, **77**(12), 2199–2203 (1998).
71. A. Beckmann, U. Vogt, N. Huda, K.S. Zonker, B.H. Brandt, 'Direct-double-differential PCR for Gene Dosage Quantification of c-myc', *Clin. Chem.*, **45**(1), 141–143 (1999).
72. H.L. McLeod, G.I. Murray, 'Tumor Markers of Prognosis in Colorectal Cancer', *Br. J. Cancer*, **79**(2), 191–203 (1999).
73. S.M. Bell, N. Scott, D. Cross, P. Sagar, F.A. Lewis, G.E. Blair, G.R. Taylor, M.F. Dixon, P. Quirke, 'Prognostic Value of p53 Overexpression and c-Ki-ras Gene Mutations in Colorectal Cancer', *Gastroenterology*, **104**, 57–64 (1993).
74. L. Gopinath, 'Outsmarting Alzheimer's Disease', *Chem. Br.*, **34**(5), 38–40 (1998).
75. G.G. Donohoe, A. Salomaki, T. Lehtimaki, K. Pulkki, V. Kairisto, 'Rapid Identification of Apolipoprotein E Genotypes by Multiplex Amplification Refractory Mutation System PCR and Capillary Gel Electrophoresis', *Clin. Chem.*, **45**(1), 143–146 (1999).
76. D. Diamond, 'Sensors', in *Encyclopedia of Analytical Chemistry*, ed. A. Townshend, Academic Press, London, 4584–4592, Vol. 8, 1995.
77. N.C. Fawcett, J.A. Evans, 'Nucleic Acid Hybridization Detected by Piezoelectric Resonance', *Anal. Lett.*, **21**(7), 1099–1114 (1988).
78. R.W. Titball, D.J. Squirrell, 'Probes for Nucleic Acids and Biosensors', in *Handbook of Biosensors and Electronic Noses*, ed. E. Kress-Rogers, CRC Press, Boca Raton, FL, 91–109, 1997.
79. K.M. Millan, S.R. Mikkelsen, 'Sequence-selective Biosensor for DNA Based on Electroactive Hybridization Indicators', *Anal. Chem.*, **65**, 2317–2324 (1993).
80. E.K. Wilson, 'Instant DNA Detection', *Chem. Eng. News*, 25 May, 47–49 (1998).
81. M. Thompson, D.C. Stone, *Surface Launched Acoustic Wave Sensors: Chemical Sensing and Thin-film Characterization*, Wiley-Interscience, New York, 1997.
82. L.M. Furtado, M. Thompson, 'Hybridization of Complementary Strand and Single-base Mutated Oligonucleotides Detected with an On-line Acoustic Wave Sensor', *Analyst*, **123**, 1937–1945 (1998).
83. L.M. Furtado, H. Su, M. Thompson, D.P. Mack, G.L. Hayward, 'Interactions of HIV-1 TAR RNA with Tat-derived Peptides Discriminated by On-line Acoustic Wave Detector', *Anal. Chem.*, **71**(6), 1167–1175 (1999).
84. C. Graham, D.L. Leslie, D.J. Squirrell, 'Gene Probe Assays on a Fiber-optic Evanescent Wave Biosensor', *Biosens. Bioelectron.*, **7**, 487–493 (1992).
85. M. Thompson, L.M. Furtado, 'High Density Oligonucleotide and DNA Probe Arrays for the Analysis of Target DNA', *Analyst*, **124**, 1133–1136 (1999).
86. G. Yershov, V. Barsky, A. Belgovsky, E. Kirillov, E. Kreindlin, I. Iganov, S. Parinov, D. Guschin, A. Drobishev, S. Dubiley, A. Mirzabekov, 'DNA Analysis and Diagnostics on Oligonucleotide Microchip', *Proc. Natl. Acad. Sci. USA*, **93**, 4913–4918 (1996).
87. J.G. Hacia, K. Edgemon, B. Sun, D. Stern, S.P.A. Fodor, F.S. Collins, 'Two Color Hybridization Analysis Using High Density Oligonucleotide Arrays and Energy Transfer Dyes', *Nucleic Acids Res.*, **26**(16), 3865–3866 (1998).
88. M. Chee, R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, S.P.A. Fodor, 'Accessing Genetic Information with High Density DNA Arrays', *Science*, **274**, 610–614 (1996).
89. J.L. DeRisi, V.R. Iyer, P.O. Brown, 'Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale', *Science*, **278**, 680–686 (1997).
90. 'Going Loopy', *New Sci.*, (2182), 17 April, 10 (1999).
91. A.J. Hayes, L.Y. Li, M.E. Lippman, 'Antivascular Therapy: A New Approach to Cancer Treatment', *Br. Med. J.*, **318**, 853–856 (1999).
92. 'Keyhole Hope in Cancer', *The Times*, 25th May, 12 (1999).

93. 'Seeing Red Will be the Signal to Attack', *New Sci.*, (2187), 22 May, 12 (1999).
94. C. Hall, 'Study Might Lead to a Genetic Test for Lung Cancer', *Electron. Telegraph*, 1 June 1999, http://www.telegraph.co.uk.
95. C. Sanson, 'Of Mice and Men', *Times Higher Educ. Suppl.*, 14 May, 42 (1999).
96. O. Wojtas, 'Dolly's Parent Gains a Californian Son', *Times Higher Educ. Suppl.*, 14 May, 40–41 (1999).
97. 'Pressure Produces Gene Therapy Breakthrough', *BBC News Online*, 25 May 1999, http://www.bbc.co.uk.
98. A. Goddard, 'Fertile with Ideas for the Infertile', *Times Higher Educ. Suppl.*, 14 May, 19 (1999).
99. 'Saved by Sperm', *New Sci.*, (2187), 22 May, 21 (1999).
100. M.A. Feitelson, M.A. Zern, 'Confronting the Epidemic of Hepatitis B and C', *Chem. Ind. (London)*, 15 February, 140–144 (1999).

Pharmacogenetic Testing

Wendell W. Weber
University of Michigan, Ann Arbor, USA

Maureen T. Cronin
ACLARA Biosciences, Mountain View, USA

1 Introduction **2**
1.1 Historical Highlights 2
1.2 The Emergence of Pharmacogenetics 3
1.3 The Scope of Pharmacogenetics 3
1.4 The Pharmacological Profile of Human Drug Response 4
1.5 The Genetic Profile of Human Drug Response 5
1.6 The Rationale and Aims of Pharmacogenetics 5
1.7 Testing Pharmacogenetic Hypotheses 6
1.8 Summary 6

2 The Changing Scene **7**
2.1 Human Protein Diversity 7
2.2 Some Biochemical Generalizations 7
2.3 DNA Polymorphism: Some Practical Matters 7
2.4 Molecular Heterogeneity of Pharmacogenetic Interest 8
2.5 The Effect of Mutations on the Gene Product 10
2.6 Summary 11

3 Genetic Profiling Using DNA Microarrays **11**
3.1 The Origin of DNA Microarrays 11
3.2 DNA Microarray Capabilities and Limitations 13
3.3 Predictive Testing with DNA Microarray Assays 15
3.4 Emerging Applications of Microarray Technologies 16
3.5 The Future of Microarrays and Allied Technology 17
3.6 Summary 18

4 The Outlook for Genomics in Predictive Medicine **18**
4.1 Options for Treating Genetic Disorders 18
4.2 Many Pharmacogenetic Polymorphisms are Exceptions to Experience with Single Gene Disorders 18
4.3 Emerging Applications of Pharmacogenetic Interest 19
4.4 Linking Human Genotyping to Clinical Applications 20
4.5 Summary 20

5 Summary: Progress in Pharmacogenetic Testing **20**

Abbreviations and Acronyms **21**

Related Articles **21**

References **21**

Drugs are administered to patients with the intention of achieving a planned therapeutic response. Yet drug prescription is a medical art because of the wide variation in individual responses to standard drug doses. In the 1960s, the field of pharmacogenetics emerged from the dedicated work of a small group of investigators who began to enrich the discipline of pharmacology by integrating it with human genetics. Pharmacogenetics provides the experimental framework to understand variability in human reactions to drugs and other exogenous substances as a function of intrinsic human genetic variability. The maturation of the human genome initiative has provided a wealth of primary genetic information and functional genomic data to fuel the understanding of genetic polymorphism and its functional consequences. Today, the field of pharmacogenetics is a well-integrated, worldwide network engaging a vast community of academic and industrial scientists.

One hallmark of pharmacogenetics studies is the large amount of genetic data that must be accumulated and integrated for high-resolution drug-response genotyping and subsequent phenotype profiling. This need for highly parallel genetic analysis has, in turn, fueled a demand for technical innovation to provide the tools necessary for its execution. The response has been a flurry of inventions in microfluidics and nanotechnology, some of which are scaled down versions of existing technologies, such as capillary electrophoresis and mass spectrometry, and some are completely new inventions. Outstanding in this technical revolution are DNA microarrays. Array technology has emerged as the most versatile and widely applied tool to support pharmacogenetics, for several reasons: arrays have the largest and most scalable capacity for parallel analysis, their inherent flexibility permits them to be used as independent analysis tools or as integrated components of more complex microfluidic systems and, finally, a wide variety of methods can be used to design and fabricate them. Pharmacogenetics is experiencing a period of rapid growth and definition. Arrays have proven themselves to be a technology capable of responding to the growing and changing needs of the current research

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

environment and are likely to be equally important as pharmacogenetics moves into the clinical arena.

1 INTRODUCTION

Drugs exert characteristic, reproducible effects when administered to individuals, and physicians have come to rely on them for the specific treatment of many human disorders. Most readers need not be convinced of the societal importance of these agents, but patients and physicians alike know that one individual may respond differently from another given the same dose of a drug. Such person-to-person variation in drug response is not a trivial matter. The analysis of the causes of natural variation in human drug response, the central problem of pharmacogenetics, requires a basic understanding of human pharmacology and human genetics.

The initial researches of a small cadre of dedicated investigators saw the emergence of pharmacogenetics in the 1960s as a new field of experimental science primarily concerned with reactions to drugs and other exogenous substances that occur because of a person's distinctive genetic makeup. The field has grown, particularly within the period from the late 1980s, into a well-integrated, worldwide network engaging a vast community of academic and industrial scientists with expertise in a broad spectrum of basic and clinical biomedical disciplines. Until recently, pharmacogenetic studies were usually limited to a few individuals and families in a research setting, but now such an investigation may be directed toward a large fraction of the human genome and involve simultaneous testing of multiple loci in many patients and healthy subjects. Such studies may, at times, be referred to as pharmacogenomics.

The establishment of laboratories dedicated to such large-scale pharmacogenetic testing is rapidly gaining momentum. This action has the attention of physicians who are in the front line of patient care, of those who are responsible for clinical laboratories and of academic and pharmaceutical scientists who are engaged in the discovery and clinical trial of new drugs. Because the genetics and the molecular basis of many pharmacogenetic traits are well established, because advance testing holds out high promise of improved patient care, and because the high potential for acceptance by the biomedical community, pharmacogenetics is in a strong position to link human genotyping and phenotyping rapidly with clinical application.

1.1 Historical Highlights

In an address before the British Medical Association in 1914,[1] Archibald Garrod proposed that enzymes were somehow implicated in the detoxification of exogenous substances. During the intense flurry of research stimulated by the rediscovery of Mendel's laws of heredity around 1900, Cuenot, from studies of mice, and Garrod, from human studies, anticipated the connection of enzymes (''diastases'') with the genetic material. Observations of physiological chemists, intrigued by the fate of chemicals in human subjects, had shown by this time that most drugs were excreted in forms that differed from those that were ingested. These observations, and Garrod's observations pertaining to a case of porphyria brought on by the hypnotic drug sulfonal, led Garrod to conclude that the ability of individuals to transform drugs into nontoxic conjugates, such as hippurates and glycuronates, protected them from the poisonous effects of these agents. Garrod was thus far ahead of his contemporaries in attributing unexpected drug responses of individuals to failure of their enzymes to detoxify these substances.

These ideas regarding the detoxification of chemicals surfaced again and again in Garrod's writings and teachings until the end of his life in the 1930s. He observed that substances in foods, in certain drugs, and in exhalations of animals and plants might produce effects in some persons wholly out of proportion to any they produce in most persons.[2] During the 1920s and 1930s, others began to scrutinize person-to-person differences in the perception of odor and taste.[3] For example, studies of ''taste blindness'' to the bitter-tasting substance *p*-ethoxyphenylurea was demonstrated to be a Mendelian character transmitted from parents to children as an autosomal trait. These deficits in sensory perception were the first indication of the high order of specificity to be expected in human response to chemicals, and the first to establish the heritable nature of such responses.[4]

By the time Garrod presented his ideas before the British Medical Association, chemists had identified virtually every major type of conjugation reaction that we know today. The exact dates of those discoveries is less relevant than the fact that they preceded by about 50 years the discovery of the other major category of detoxifying enzymes (P450s), commonly known as the microsomal enzymes or the drug-metabolizing enzymes. R. Tecwyn Williams, a pioneering pharmacologist at St. Mary's Medical School in London, saw evidence for a new, far-reaching principle of drug metabolism in the studies of the microsomal and conjugating enzymes, namely that the metabolic disposition of drugs in humans and animals occurs in two phases: first, oxidation, reduction, and hydrolysis, then conjugation, designating them as phase I and phase II metabolism respectively.

1.2 The Emergence of Pharmacogenetics

Until around 1950, progress slowed, but then several breakthroughs foreshadowing important events to come were made. In 1948, hemoglobinopathies S and C were identified and their mode of inheritance was determined. Hemoglobin S, the first hereditary protein variant to be identified, was also the first example of a variant protein in which replacement of a single amino acid residue within the protein molecule could be identified unequivocally with a change in the functional effect of the protein. In 1953, the double helix of DNA was described. A few years later, the chromosomes of humans were visualized and enumerated, and the chromosomal basis of one form of cancer (chronic myeloid leukemia) was identified. Following closely on the demonstration of polymorphic forms of hemoglobin S and C by electrophoresis, application of this technique to plasma proteins showed that protein polymorphism was a widespread phenomenon that merited further study.

New technologies that became available in the 1950s, combined with a more genetic approach to investigation, disclosed new relationships between the genetic control of responses to exogenous substances and their metabolic fate. Landmark studies of unexpected responses to primaquine, isoniazid and succinylcholine (suxamethonium) were the first to establish a link between drug response and heredity. The confluence of pharmacology, genetics and biochemistry noted in Arno Motulsky's 1957 seminal paper,[5] concerned with person-to-person differences in drug response that arise because of the unique genetic constitution of individuals, marked the true beginning of pharmacogenetics as an experimental science. In 1962, Werner Kalow published the first systematic account of the field, demonstrating the implications of heredity for responses of microorganisms, insects, and vertebrates, including humans, to environmental chemicals.[6] Since then, the characteristics of many more human traits of pharmacogenetic interest, too many to chart here, have demonstrated beyond any doubt that heredity is a remarkable prognosticator of unexpected response to drugs and other exogenous substances.

1.3 The Scope of Pharmacogenetics

Hereditary differences in toxicity or efficacy that result from exposure to therapeutic agents are usually considered as being within the province of pharmacogenetics, but if the unexpected response results from exposure to other chemicals, or to physical, climatic or atmospheric agents, it may be classified as ecogenetic.[7] Ecogenetics encompasses genetically conditioned responses to exogenous substances from any source and, as such, would include the pharmacogenetics of drugs used in medical practice. It should be noted, however, that, despite this distinction, the literature does not always adhere to this convention. When it is of interest to assess the characteristics of a given response, the criteria to be met in designing and executing a proper study to determine the heritability of the response in twins and families, and the distribution and incidence of genotypes and phenotypes in larger populations, are, in fact, identical for pharmacogenetics and ecogenetics.

Most well-studied polymorphisms derive from investigations of therapeutic agents, because these agents can be administered safely to individuals of all ages in defined amounts, and their disposition patterns can be measured. Historically, the drug-metabolizing enzyme polymorphisms are foremost in pharmacogenetics, and aficionados recognize these traits as the cornerstone of the field.[8] However, since medical drugs and other exogenous chemicals are biotransformed by the same enzymes, unexpected effects, as Garrod noted, may extend to foodstuffs as well as to sporadic disorders associated with exposure to chemicals under occupational and industrial conditions. The enzymes that metabolize exogenous substances may also metabolize hormones and other endogenous messengers that regulate cell signaling; consequently, mutations that affect those pathways may also cause differences in individual response. This evidence serves as an additional hallmark of pharmacogenetic and ecogenetic phenomena. The range of such phenomena is virtually without limit.

The administration of two or more drugs in combination may also set the stage for an unexpected drug response that depends on the genetic makeup of the recipient. The CYP2D6 (debrisoquine/sparteine oxidation) polymorphism is a particularly rich potential source of such interactions. This polymorphism affects the disposition of more than 30 therapeutic agents including β-adrenergic blockers, antidepressants, antiarrhythmics, neuroleptics, and various miscellaneous drugs such as phenformin, dextromethorphan, and codeine. Combined administration of two or more of these agents may result in an aversive interaction.

Smoking represents another environmental hazard that results in a variety of responses and disorders which may depend on genetic predisposition of smokers. In individuals with a genetic defect in α_1-antitrypsin, smoking is the most important environmental factor affecting the rate of deterioration of lung function.[9] Additionally, information is accumulating that a variant form or forms of CYP2A6 may exhibit impaired capacity to metabolize nicotine and may thereby protect against the effects of smoking by reducing the number of cigarettes smoked.[10,11]

Genetic predisposition appears to be involved as a hypersusceptibility factor for workers exposed to complex mixtures of chemicals. The formation of

aniline–hemoglobin adducts and excessive levels of methemoglobin can occur in those who are genetically ''slow acetylators'' after prolonged exposures to aniline derivatives.[12] The sporadic occurrence of urinary bladder cancer in slow acetylators many years after lengthy exposure to the aromatic amine carcinogens[13] employed in the manufacture of dyestuffs, plastics and electrical materials is a case in point.

Important clues to the role of heredity in unexpected responses to drugs and other chemicals are provided by individuals of different ethnogeographic origin. Perhaps the most familiar example is ''primaquine sensitivity'' occurring as a result of glucose-6-phosphate dehydrogenase (G6PD) deficiency among African, Mediterranean and Oriental persons. This is associated with sudden hemolysis, which occurs among susceptible males after exposure to any of some 200 drugs.[14] Other examples of ethnic specificity in drug response include the protective effect against alcohol-induced liver disease of a variant of aldehyde dehydrogenase (ALDH2) that is found in Japanese but not Caucasians[15] and the protective effect against alcohol-related birth defects of a variant form of alcohol dehydrogenase (ADH-3) occurring among African-Americans.[16]

While many of the examples cited above indicate that unexpected responses to exogenous chemicals may be mainly caused by a single (monogenic) factor, experience shows that variation of more than one hereditary element may contribute to the outcome. One example to illustrate the synergism that can occur between interacting factors is provided by a study of liver cancer caused by exposure to aflatoxin and hepatitis B among the Chinese. Aflatoxin exposure alone enhances the relative risk by approximately twofold, while infection with hepatitis B virus alone enhances the risk by approximately fivefold; however, combined exposure yields a 60-fold increase in relative risk of liver cancer. Hepatitis B infection, therefore, enhances the carcinogenic response among Chinese to aflatoxin by 30-fold.[17]

Nowadays, the scope for inherent variation in human response to exogenous chemicals is so great, and the pace of science so rapid, that it is difficult to stay abreast of new developments in the field. Although metabolic polymorphisms of drug-metabolizing enzymes appear most frequently in the literature, polymorphisms of other genes of pharmacogenetic interest, such as those that encode receptors, clotting factors, transporters and immunologic response proteins, appear increasingly frequently. In the 1990s, recombinant DNA technology has contributed immensely to our knowledge of pharmacogenetics by facilitating the development of sensitive and specific diagnostic methods for collecting molecular information about unexpected responses to drugs and other exogenous substances.

1.4 The Pharmacological Profile of Human Drug Response

Responses to a particular drug actually result from the physiological and biochemical attributes of cells that carry the necessary receptors in tissues of the recipient, but these responses are subject to modification by a variety of intrinsic and extrinsic influences.[4] The age, gender, weight and other physiological and pathophysiological attributes of the individual, including heredity, are important intrinsic factors that can modify drug responses in individuals. Foods and other dietary components, exogenous chemicals in household and workplace environments, the use of tobacco and alcohol, and drugs (prescription and nonprescription) themselves are some of the extrinsic factors that may affect these responses.

The response of an individual following administration of a drug (or after exposure to an exogenous chemical) is a complex process that is difficult to analyze. By reducing this process to pharmacokinetic and pharmacodynamic mechanisms, pharmacologists have devised a somewhat idealized picture of the response, which can be analyzed more readily. Pharmacokinetic mechanisms are those that affect the concentration of the drug at receptor sites and its time course of action (absorption, distribution, and elimination by metabolism and excretion) in the individual, while pharmacodynamic mechanisms refer to receptor-mediated and allied events. This maneuver also provides a framework well suited to discussion of the pharmacological nuances of variation in human drug response.

The response of an individual to most drugs can be thought of as a train of events that begins when the drug enters the bloodstream and ceases when the drug and its metabolites are completely eliminated. The administration of a drug, given in ordinary doses, is usually accompanied by a gradual rise in drug concentration in the blood that reaches a peak or steady state; if no more drug is given, the concentration steadily declines until the drug and its metabolites are eliminated. If the drug attains a concentration within the therapeutic range, the characteristic response is expected. If the concentration reaches a level above or below that range, either a toxic response or the absence of the expected response may occur. For unexpected responses, experience indicates that there are two likely explanations. If the drug concentration in plasma is above or below the therapeutic range, one would anticipate a pharmacokinetic mechanism is responsible for the unexpected response, but if the concentration is within the therapeutic range, a pharmacodynamic defect is a more likely explanation.

Consequently, from the pharmacological viewpoint, unexpected drug responses to drugs and other exogenous

substances may be regarded to a first approximation as peculiarities resulting from a defect in either the pharmacokinetics or the pharmacodynamics of the drug.

1.5 The Genetic Profile of Human Drug Response

Geneticists, in contrast to pharmacologists, might consider a drug as an environmental agent on which individuals exert important differential effects; therefore, they might regard an unexpected drug response as a reflection of a genetic difference between normal and abnormal responders.[4] Geneticists, as a rule, are more interested in looking for differences that discriminate individuals rather than focusing on universal similarities, and they would seek to explain an unexpected drug response by drawing on the gamut of information from the normal and affected responders as well as from their biological relatives. From the information obtained, they would attempt to determine the relative contributions of heredity and environment to the unexpected response.

Genetic defects that cause unexpected drug responses occur sufficiently often in many populations to enable members to be divided into two or more relatively common types of responder. Hereditary variation in which such sharply distinct qualities coexist in a given population is referred to as genetic polymorphism. This term is usually used in genetics to refer to genetic loci for which variants occur with a frequency of 1–2% or greater, but in pharmacogenetics the definition is better based on phenotype. This issue has been the source of some discussion[18,19] and is considered further below (section 2.3). To avoid contradictions between pharmacogenetic usage and genetic theory and concepts, this article will refer to the polymorphisms linked to human responses to drugs and other exogenous substances as ''pharmacogenetic polymorphisms''. Pharmacogenetic polymorphism may include variation produced by chromosomal aberrations, which are detectable by cytogenetic techniques, but most result from smaller, genic lesions.

To assess the contribution of heredity to pharmacogenetic polymorphism, unexpected drug responses are explored at all levels of gene action, from the gene molecule itself, to the individual and their biological families, to populations of individuals. Broadly speaking, hereditary differences in drug response are characterized by two types of information: the genetics that characterize the polymorphic phenotypes and their molecular basis. The latter includes a description of the genetic heterogeneity at the level of DNA and of the protein variants that explain the trait. By studying a given trait from several points of view, the relative influence of human ecology and heredity and the mechanisms by which they occur are ascertained. For these purposes, the concepts and techniques of human genetics, biochemistry, population and molecular genetics, and epidemiology are used in concert to analyze pharmacological, toxicological and epidemiological observations of individuals, twins, families and larger populations.

1.6 The Rationale and Aims of Pharmacogenetics

Pharmacogenetic investigation often begins with an anecdotal clinical observation on one or a few individuals who have experienced an unexpected response to a specific drug, or with an epidemiological observation on a group of individuals who have developed a disorder associated with occupational exposure to a chemical or complex mixture of chemicals.[4] The main goal of pharmacogenetics, simply put, is to use knowledge gained in understanding the influence of heredity on human sensitivity (or resistance) to exogenous chemicals to avoid the occurrence of such responses in susceptible persons. Therefore, while the early detection of cancer of environmental origin, for example, may do little to change the outcome in an affected person, identification of the trait that led to its initiation may prevent its occurrence in others.

A number of practical problems can arise in the design and execution of pharmacogenetic studies. A complete assessment of the causes of a given trait may be exceedingly difficult to achieve, may require the collaboration of scientists with differing expertise or may require resources not readily available. For example, epidemiological evidence may suggest a disorder (say bladder cancer) is caused by a particular agent (or agents), but the agent or agents usually cannot be identified unequivocally solely by epidemiological evidence; clinical observation may suggest an adverse reaction is caused by a certain drug, but the physican who contemplates investigation of the reaction by further exposure of humans to the offending drug may face insurmountable ethical and methodological constraints; susceptibility to a toxicant can often be more thoroughly studied under highly controlled conditions in animals than in humans, but studies in animal models are slow and expensive, and animal responses do not necessarily translate closely to those in humans.

Establishment of definable end-points for the investigation of specific differences between normal and abnormal responders poses another problem. Consider the evaluation of drug treatment of hypertension as a case in point. In this instance, the definitive end-point would be measured by the difference in the incidence of strokes and cardiovascular disease between normal and abnormal drug responders, but this would probably occur years after the trait is recognized and modifying treatment undertaken. Another example of pharmacogenetic interest

concerns the evaluation of individual variation in susceptibility to cancer associated with prolonged exposure to an occupational carcinogen. The definitive end-point in this case is the difference in cancer incidence many years later between hypersusceptible and normal responders exposed to the carcinogen. It is apparent that accomplishment of trials to determine definitive end-points might require studies extending over many years involving many individuals and large populations, and the payoff is not guaranteed.

Investigations with more limited goals designed to measure intermediate or short-term changes in biochemical and molecular markers may lessen some of these difficulties. Biological markers that can detect early and subtle differences in individual response would be necessary for this purpose. The most suitable are markers that can be detected in small samples of tissue or body fluids that are accessible to sampling; they should be expressed differentially in accordance with differences observed in normal and abnormal responders and they should have a low probability of spontaneous change.

Several advantages accrue to study designs that focus on the use of intermediate- or short-term biomarkers. First, the marker could be used to document and characterize the heterogeneity in response of individuals to a given drug, and to determine its distribution in individuals, families, and other populations of interest. This would indicate the frequency of a given trait, its mode of inheritance, and its ethnic and geographic specificity. Second, the markers could be used to identify the genes and the mutations responsible for the trait in susceptible people. This information would reveal the molecular basis of the trait and would aid in elucidating disordered physiological mechanism(s) that account for the aberrant response. It would shed light on the mechanism of the response to the agent in normal (wild-type) responders. The marker would also be of further use in the development of noninvasive, inexpensive diagnostic tests to determine the phenotype and genotype of individuals for the trait. Third, for agents used in medical therapy, identification of the molecular basis of specific traits and the elucidation of their underlying mechanisms would rationalize the selection of specific agents for susceptible (or resistant) individuals and make it possible to tailor drug regimens to the individual patient. Fourth, since the exact cellular site that is defective must be known for rational design of therapy, the information obtained in the first three study parameters would help to guide discovery of new drugs and to improve the design and conduct of their clinical trial. Finally, studies with markers could have wider implications for patient welfare by alerting physicians, clinical scientists, and others engaged in patient care to the importance of biomonitoring of individuals who suffer sporadic illnesses while ingesting certain drugs and of individuals who become ill after prolonged exposure to toxic chemicals associated with a specific occupation or a particular environment.

An important goal of studies with intermediate- and short-term biomarkers is validation of the markers by demonstrating their correlation with definitive end-points. It is reasonable to expect that these biomarkers will enable differences in human susceptibilities to exogenous chemicals to be detected, or suspected, at an early stage of exposure, and the liability of susceptible persons to be predicted long before the definitive end-point is reached. It follows that studies using such biomarkers could be accomplished in a much shorter time and at a much lower cost than those aimed at determining definitive end-points.

1.7 Testing Pharmacogenetic Hypotheses

Exploring the pharmacogenetics of unexpected drug responses can be a formidable task. No hard and fast rules exist to guide the investigator's advance planning, but there are some general principles that are widely applicable.[4] Experience indicates that the response of an individual to a drug and its inherent variation are best considered in light of the pharmacokinetic and pharmacodynamic mechanisms that define the disposition and actions of drugs in individuals (see section 1.4). Insights into the relative importance of the pharmacokinetic and pharmacodynamic phases of the response to the overall variability in the response to a given drug in a heterogeneous group of drug responders can often be gained by a preliminary pharmacokinetic analysis.

In general, an experimental approach is almost inevitably based on testing a main hypothesis and one or more auxiliary hypotheses.[20] The novice may, in fact, proceed with only a vague notion of the hypothesis being tested, but skilled investigators know that it is advantageous to have thought about it explicitly. Auxiliary hypotheses are set forth as reasonable alternatives in case the main hypothesis proves false. Investigators should also remember that, for every discovery made with innovative approaches, many important advances are made simply by recognizing that an established method or familiar technique devised for one application can be used for another. In some cases, it is possible for the investigator to design a single experiment that decides the fate of a given hypothesis.

1.8 Summary

Individual differences in human responses to therapeutic agents and other exogenous chemicals often result from some derangement of the genetic material that may be transmitted from one generation to the next. The scientific

study of the effects of heredity on drug response is the province of pharmacogenetics. Its main purpose is to explain the toxicological effects of the derangements by rigorous characterization of their biochemical and molecular basis. In this way, it is hoped to understand the causes of human susceptibility to exogenous substances and avoid the occurrence of unexpected drug responses in susceptible persons.

2 THE CHANGING SCENE

2.1 Human Protein Diversity

Proteins are vital to virtually every process of biological importance because of their capacity to bind other proteins and small molecules with high sensitivity and specificity.[21] The assortment of proteins that resides in cells and tissues requires that the synthesis and targeting of these molecules is carried out with a high level of fidelity according to the well-defined set of instructions encoded in the genes. The properties of distinct proteins ultimately define the link that exists between the genetic constitution of a given individual and his or her response to drugs. It follows that a structural alteration in a given gene, or an error in carrying out the program of instructions encoded in the gene, might reasonably be expected to change the structure or the amount of the protein synthesized, or its final destination. Briefly, mutation could cause an unexpected drug response.

The occurrence of G6PD deficiency, succinylcholine sensitivity and isoniazid acetylation polymorphism, detected in the 1950s, was regarded as a serious threat to drug usage in medical therapy, requiring further investigation. Examination of the pharmacological defect and molecular genetic basis of unexpected drug responses set the agenda for subsequent research, and studies quickly showed that changes in protein structure resulting from genetic change altered human drug responsiveness wherever it was sought, among proteins with widely disparate functions, in health and disease.

2.2 Some Biochemical Generalizations

By the 1970s, Harris and colleagues had sought to examine the genetic causes of human protein diversity from the biochemical, chromatographic and electrophoretic characteristics of proteins associated with certain single gene disorders.[22] Certain hemoglobinopathies and enzymopathies provided the best models for this purpose because information about structural gene mutations was matched by physiological and pathological data. Evidence suggested that mutations might range from a single base change within the coding region of a gene to large deletions that removed a large part, or all, of a gene. A single polypeptide chain that was affected by different mutations in the coding region might thus be represented by several distinct alleles. Most hemoglobinopathies, for example, could be ascribed to missense, nonsense, or frameshift mutations of the coding region of globin genes. Among the enzymopathies, observations suggested that allozymes (i.e. enzymes encoded by allelic loci) generated by point mutations and small deletions would usually display highly homologous amino acid structures and might have similar enzymatic properties, whereas larger deletions resulted in drastically shortened, nonfunctional gene products, or the total absence of the product.

Post-translational modifications were recognized as additional sources of the structural modification of proteins.[22] Should such a modification occur by an enzyme-mediated process, as had been established for oxidation of sulfhydryl groups or the addition of carbohydrate or phosphate groups, or by the cleavage of the polypeptide with loss of a terminal amino group, or a larger part of the chain, it too could be subject to genetic variation.

Unfortunately, the available biochemical data were inadequate to yield a detailed picture of the molecular basis of protein heterogeneity. For instance, a change that alters the charge on a protein can be detected by electrophoresis, but since two-thirds of amino acid substitutions are neutral, they will not change the net charge on the protein and will not be detected in this way. As a consequence of the inability of the standard methods of protein characterization to distinguish variation owing to genetic variation at the level of translation from that which occurs post-translationally, or nongenetically, information regarding the genetic basis of protein diversity was very meager.

2.3 DNA Polymorphism: Some Practical Matters

The analysis of genomic DNA reveals that the human genome is highly polymorphic. Sites that exhibit alternative sequences at a particular chromosomal site can be used as genetic markers for the site, or for the chromosome bearing the site. Three types of polymorphic site are useful for identifying individuals who are predisposed to unexpected drug responses: those within genes that determine drug response, those within short variable DNA repetitive sequences (known as variable number tandem repeat (VNTR)) and those within microsatellites. When such a site is associated with a given trait and tracked within families and larger populations, it can provide information about inheritance patterns and the prevalence of the trait. Polymorphic sites can be studied by restriction fragment (Southern) analysis to survey chromosomal sites for polymorphism, or by DNA sequencing to determine the precise location of base changes that

define a polymorphism. Southern analysis cannot resolve polymorphisms that differ by only one or a few bases, and neither can it identify the polymorphic base, whereas the size and boundaries of the polymorphic site can be precisely determined with the polymerase chain reaction (PCR) used in DNA sequencing. This second approach is essentially a refinement of the first and is preferred in many applications.

Since the development of molecular techniques, remarkable progress has been made toward identification of genes responsible for pharmacogenetic polymorphisms and toward understanding the molecular basis of quite a few of these polymorphisms at the DNA level. Most of the progress has related to traits attributed to a single polymorphic gene, many of which have been cloned and sequenced.[23] For some, the gene of interest has been expressed in heterologous expression systems in quantities sufficient for biochemical and pharmacological characterization.

The VNTRs or minisatellites of the human genome may be repeated 100 times or more in different persons. Many VNTRs, numbering in the thousands, are well characterized. Restriction fragment analysis will produce different size fragments proportional to the number of repeats in the VNTR. Where the identification of traits has been slowed for lack of a sufficient number of suitable genetic markers, the use of VNTRs should alleviate this constraint.[24]

A number of genes have been found to incorporate microsatellites within them.[25] Microsatellites are stretches of repetitive DNA sequences; if mutated, they are capable of disrupting cell function. Such mutations may serve as hereditary markers of genomic instability that may increase individual susceptibility to certain cancers. For example, the loss of microsatellite DNA from the androgen receptor gene has been associated with prostate cancer. This event is of pharmacogenetic interest because a paradoxical response to antiandrogen therapy (i.e. stimulation of tumor growth instead of the inhibition expected) was exhibited by a patient.[26]

The extent of protein diversity in natural populations is immense, but the abundance of DNA polymorphism in the human genome appears even greater. Analysis of genomic DNA reveals the presence of a large number of these polymorphisms. On average about 1 in 500 nucleotides differs between two randomly chosen alleles. Only about 5–10% of these polymorphisms is detected by restriction analysis, and frequently these polymorphisms are unrelated to a clinical phenotype. This phenomenon has been demonstrated for the CYP2D6 polymorphism by Skoda and colleagues,[27] who examined genomic DNA from 53 persons using a total of 20 restriction enzymes to fragment the DNA in different ways. Tests of 13 enzymes yielded 14 allele-specific polymorphic patterns at the *CYP2D6* locus, while the other seven enzymes revealed none. This study and others (see references in Skoda et al.[27]) show that DNA polymorphisms are often functionally silent at the protein level, having no detectable phenotypic effect.

It follows that information about a given genotype is insufficient for unambiguous identification unless combined with information about the associated phenotype. Further evidence pointing to the importance of identification of phenotypes arises from numerous sources. These include studies of the potential dissociation of phenotype from genotype observed in certain races[28,29] and in certain families,[30] the lack of resolution of heterozygous and homozygous subjects by standard metabolic phenotyping procedures,[30] the different phenotypes for different substrates for a given polymorphism,[31,32] structure–function relationships,[33] gene–gene interactions,[34] and the patterns of disease[35] associated with polymorphisms attributed to genes with closely similar sequences.

We should also point out that since the vast majority of pharmacogenetic traits, unlike most genetically determined human disorders, usually cause no recognizable effect on the health of predisposed individuals, the importance of phenotype takes on added significance. Only when susceptible persons suffer an unexpected response from exposure to certain drugs, dietary constituents or other environmental agents is their predisposition revealed. Fortunately, many of the polymorphisms associated with genetically variable human drug-metabolizing enzymes, can be detected in advance with the aid of a number of test probes that are suitable for human phenotyping by in vivo or ex vivo testing procedures. A partial list of these probes is presented in Table 1.

For the reasons stated above, pharmacogenetic polymorphisms must be defined on phenotypic grounds. Any comprehensive description of a given pharmacogenetic polymorphism should include information about the phenotype that characterizes the polymorphism and the genotype(s) that explain the polymorphism.

2.4 Molecular Heterogeneity of Pharmacogenetic Interest

Recombinant DNA techniques rapidly superseded standard methods of protein characterization for genetic analysis during the 1980s, and reports of the molecular characteristics of numerous enzymes, including the P450 (CYP450) enzymes of drug metabolism, soon appeared in the literature. Because of their specificity for endogenous steroids, fatty acids, and prostaglandins, as well as for many drugs, environmental pollutants, and carcinogens, the CYP450 enzymes had already attracted a great deal of

Table 1 Test probe drugs for human phenotyping of human drug-metabolizing enzyme polymorphisms

	Test probe	Refs.
Phase I enzymes		
Aldehyde dehydrogenase	Acetaldehyde	36
Alcohol dehydrogenase	Ethanol	16
CYP1A2	Caffeine	8
CYP2A6	Nicotine, coumarin	11, 37
CYP2C9	Warfarin	8, 37
CYP2C19	Mephenytoin, omeprazole	8
CYP2D6	Dextromethorphan, debrisoquine, sparteine	8
CYP2E1	Chloroxazone, caffeine	8
CYP3A4	Erythromycin	8
CYP3A5	Midazolam	8
Serum cholinesterase	Benzoylcholine, butrylcholine	38
Paraoxonase/arylesterase	Paraoxon	39, 40
Phase II enzymes		
Acetyltransferase (NAT1)	*Para*-aminosalicylic acid	41
Acetyltransferase (NAT2)	Isoniazid, sulfamethazine, caffeine	42
Dihydropyrimidine dehydrogenase	5-Fluorouracil	43
Glutathione transferase (GST-M1)	*Trans*-stilbene oxide	44, 45
Thiomethyltransferase	2-Mercaptoethanol, D-penicillamine, captopril	46
Thiopurine methyltransferase	6-Mercaptopurine, 6-thioguanine, 8-azathioprine	47
UDP-glucuronosyl transferase (UGT1A)	Bilirubin	48, 49
UDP-glucuronosyl transferase (UGT2B7)	Oxazepam, ketoprofen, estradiol, morphine	50, 51

attention among biochemical pharmacologists. Additionally, metabolic studies conducted during the mid-1970s with sparteine or debrisoquine suggested a polymorphic variant of CYP450 might be responsible for the poor metabolism and unexpected responses of susceptible persons. Other studies with mephenytoin suggested a variant of a distinctly different CYP450 appeared to explain the poor metabolism and unexpected responses of susceptible persons to this drug. In 1982, the complete primary structure of a mammalian CYP450 was reported from recombinant DNA studies, and this breakthrough prompted a great deal of further investigation.[52] A recent count shows that more than 150 isoforms of CYP450 have been characterized as products of different genes, and at least 30 different human CYP450 enzymes have been purified, cloned, sequenced and characterized. Of these, less than ten (namely 1A1/1A2, 2A6, 2C19, 2D6, 2E6 and 3A4) are responsible for oxidation of most drugs and other substrates in the human environment.[8]

A chronology of selected human pharmacogenetic polymorphisms that have been identified since the emergence of pharmacogenetics in the 1950s is given in Table 2. It includes a cross-section of polymorphisms: several drug-metabolizing enzyme polymorphisms, several receptor polymorphisms (e.g. malignant hyperthermia, vasopressin resistance, insulin receptor resistance) and polymorphisms of some other protein variants of pharmacogenetic interest. The table also draws attention to the time that elapsed between the initial description of the inheritance of a given polymorphism and the explanation of the polymorphism at the DNA level.

As Table 2 shows, molecular genetic studies identified some of the molecular defects in CYP2D6* associated with poor metabolism of sparteine and debrisoquine some 11 years after the trait was discovered. CYP2D6* (sparteine/debrisoquine oxidation) polymorphism has stimulated an enormous amount of investigation and we now know that this polymorphism results in three separable phenotypes: poor metabolizers, extensive metabolizers, and ultrarapid metabolizers. The CYP2D6* poor metabolizers have an impaired capacity to metabolize more than 30 drugs and are homozygous for an inactive or deficient CYP2D6 enzyme that has been modified by truncation or missense mutations of the gene *CYP2D6**. In contrast, ultrarapid CYP2D6* metabolizers possess an enhanced capacity to metabolize these drugs because they possess more than one copy of an amplified *CYP2D6**.

Pharmacogenetic polymorphism is well established for the genes of several other CYP450 enzymes including 1A1, 2C9, and 2C19, and is rapidly accumulating for 1A2, 2A6, 2E1, and 3A4.[8] Most of the genetically variant CYP450 enzymes occur as high- and low-activity (or null) isoforms that may confer individual susceptibility to the toxic effects of environmental chemicals including carcinogens. Molecular genetic studies have identified the genes that encode these enzymes as well as numerous allelic forms of them.

Table 2 also lists several drug-conjugating enzymes. During the 1990s, considerable progress has also been

Table 2 Chronology of pharmacogenetics

Polymorphism	Inheritance described	Mutation described	Elapsed time (years)
Succinylcholine sensitivity	1957–60	1990–92	33
G6PD deficiency	1958	1988	30
Long QT syndrome	1957–60	1991–97	34
Acetylation	1959–60	1989–93	30
Malignant hyperthermia	1960–62	1991–97	31
Glucuronosyl transferase	1966–69	1992–?	26
Vasopressin resistance	1969	1992	23
Aldehyde dehydrogenase	1969	1988	16
Debrisoquine oxidation	1977	1988–93	11
Retinoic acid resistance	1970	1991–93	21
Thiopurine methyltransferase	1980	1995	15
Mephenytoin oxidation	1984	1993–94	9
Glutathione transferase	1986	1990	4
Fructose intolerance	1986	1988–95	2
Insulin receptor resistance	1988	1988–93	0
Androgen resistance	1990	1990	0
Glucocorticoid remediable aldosteronism	1992	1992	0

made in the molecular genetic analysis of these enzymes. Prominent members of this group are the polymorphic forms of acetyltransferases, glucuronosyltransferases, glutathione-*S*-transferases and thiopurine methyltransferase.

Receptor proteins represent another subset of proteins (see Table 2) with enormous potential to influence human sensitivity to exogenous substances.[4] Comparatively few receptor polymorphisms have been studied so far, since this group of proteins has not yet been explored in the depth achieved with the drug-metabolizing enzymes. Prior to the 1980s, the existence of receptors and receptor subtypes was customarily defined by differences in their interactions with drugs, but this approach did not achieve outstanding success in establishing receptor heterogeneity on a molecular plane. Based on their location and mechanistic features, two main types of pharmacological receptor are recognized: those that are located in the cytoplasm and with a locus of action within the cell nucleus, and those that insert into the cell surface and span the membrane. Cell surface receptors are involved in the actions of most drugs and endogenous first messengers; being on the cell surface, they can bind biogenic amines, protein and polypeptide hormones, autocoids, neurotransmitters and environmental chemicals. The nuclear (cytoplasmic) receptors, by comparison, interact with relatively few, albeit important, first messengers that can enter the cell, such as the steroid hormones. The techniques of molecular biology have enabled sufficient amounts of a given protein to be produced for biochemical and pharmacological characterization and have greatly refined and extended the genetic analysis of receptor heterogeneity.

2.5 The Effect of Mutations on the Gene Product

Mutations have been observed to have two main effects on the gene product: they can cause the production of a structurally variant protein with altered functional properties or the production of a fully functional protein in altered amounts, usually reduced, although traits attributable to increased amounts of protein have been described.[53,54] Mutations that cause the synthesis of structurally variant proteins usually occur within the coding region of the gene, whereas mutations located within the promoter and other regulatory sequences usually alter protein quantity. Most pharmacogenetic polymorphisms result from point mutations, or other small genic lesions that lead to a functional change in a given protein, or loss of all or most of a gene. As these polymorphisms occur at specific genetic loci and within specific regions of the gene, the ability to detect and determine the associated specific structural change(s) allows for diagnosis.

Knowledge of the biochemical, pharmacological or toxicological features associated with an unusual phenotype can sometimes lead to inferences about the identity of the gene and the properties of the allelic variants responsible for the phenotype. Many pharmacogenetic polymorphisms were characterized by this approach long before molecular genetic techniques were developed. For example, observation of low levels of urinary acetylated metabolites of isoniazid suggested that an acetyltransferase, or acetyltransferases, with impaired capacity to acetylate isoniazid accounted for the isoniazid-induced nerve damage seen in susceptible patients. Biochemical studies documented the presence of functionally impaired forms of this drug-conjugating enzyme, which led to the

identification of genetically slow acetylators as a separate class of drug responders in human populations. The properties of acetylating enzymes isolated from tissues of slow acetylators were found to be closely similar to those from normal (rapid) acetylators, which suggested that a small genic lesion, possibly a point mutation, or even a small deletion, might account for slow acetylation of isoniazid.[55]

At the time of the early biochemical studies, the existence of acetyltransferase variants could only be inferred; however, with the aid of molecular genetic techniques, the genes for acetyltransferase have been cloned and sequenced, and expression studies have demonstrated the existence of such variants, affirming predictions made some 30 years earlier. We now know that humans possess two drug-metabolizing acetyltransferase loci, and that variation at the *NAT2** locus is responsible for isoniazid acetylation polymorphism. We also know that the coding region of this gene is the site of multiple mutations that confer slow acetylation ability on individuals. A total of nine single nucleotide changes have been identified within the coding region of *NAT2** that account for more than 95% of slow acetylator alleles extant in human populations. Since these mutations occur singly and in combinations of two and three, they yield more than 20 *NAT2** slow acetylator alleles.[56]

The *NAT2** polymorphism is an example of polymorphisms within the coding region of the gene, but recently polymorphisms within regulatory regions have also been described. The genetically polymorphic form of CYP2E1 with enhanced activity that is associated with ethanol ingestion and with obesity represents an example of a regulatory polymorphism.[57] CYP2E1 is recognized as an important enzyme for the detoxification of ethanol and many other xenobiotics and some drugs, including chloroxazone. Chloroxazone is a test drug that is used to detect intersubject variation in CYP2E1 activity by in vivo and ex vivo procedures. Toxicological studies suggest a possible role for this enzyme in alcohol-related disease among African-American women. Analysis of the upstream region of *CYP2E1* revealed a DNA insertion, 100 base pairs (bp) long. The presence of this insertion mutation significantly increases CYP2E1 metabolic activity as measured by chloroxazone hydroxylation, but only among persons who are obese or have recently consumed ethanol, or both. The frequency of heterozygous carriers of the mutation is 31% in African-Americans and 6.9% among Caucasians. This polymorphism in the regulatory region not only enhances CYP2E1 activity but is sufficiently common to affect susceptibility to CYP2E1-related disorders in at least two ethnic populations.

These examples comprise only a small part of the molecular genetic studies that have been conducted on pharmacogenetic polymorphisms, They afford insights into the genetic basis of phenotypic differences in human drug response, but it is prudent to avoid attempting to classify the effect of the mutation on the gene product from incomplete information. For example, a trait that appears, on limited study, to result from a quantitative defect in protein synthesis, may, on further investigation, be found to be caused by a structurally abnormal protein that disintegrates because it is inherently unstable or undergoes rapid proteolysis in the cellular environment.

2.6 Summary

Proteins are the molecular links between the genetic constitution and the drug response of the individual. The presence of a biologically active protein molecule at its native site in cells and tissues requires the synthesis and targeting of the molecule to occur in accordance with the well-defined program of instructions encoded in the genes. An alteration in the gene or failure to carry out the instructions with complete fidelity could produce a variation in the structure, function or final destination of the protein, any of which might alter the profile of drug response in the individual. Despite the remarkable advances in the understanding of pharmacogenetic polymorphisms at the molecular level, the dissociation of the phenotype from the genotype that is observed under various circumstances requires that both the phenotype and the genotype be characterized for unambiguous identification of a given polymorphism.

3 GENETIC PROFILING USING DNA MICROARRAYS

3.1 The Origin of DNA Microarrays

DNA microarray technology has its origins in Edwin Southern's method described in his landmark article ''Detection of specific sequences among DNA fragments separated by gel electrophoresis'' published in 1975.[58] There, he described how to combine the specificity of DNA restriction endonucleases with gel electrophoresis and sequence-specific hybridization to probe genomic DNA for its sequence composition. Restriction enzymes were used to generate controlled, predictable genomic DNA fragmentation, while gel electrophoresis provided an efficient, economic approach to arraying the fragments by size. The arrayed fragments were transferred by ''blotting'' onto a membrane support with a high affinity for nucleic acids, affording ease of handling and analysis of the same fragmented sample with a variety of probes. Finally, the complementarity of DNA base pairing as it occurs during hybridization permitted access to specific sequence information in the genomic DNA sample when it was hybridized with probes of known sequence.

The simplicity, elegance, and economy of the Southern blot facilitated its adoption into many molecular biology laboratories in a minimum amount of time. It not only became a standard laboratory protocol but also quickly began evolving into a diverse collection of hybridization-based analytical techniques. In 1977, for example, Alwine[59] described the Northern blot version of Southern's electrophoresis: a hybridization probe method to analyze messenger RNA (mRNA) sequences that provided the link between genomic sequence and functional mRNA expression analysis. Alwine's technique was also quickly adopted. Improvements and innovations in Northern and Southern blotting continue to simplify, refine and generalize the analysis of the broad range of sample types studied today.

Continued development in the Southern blotting method has greatly enhanced understanding of genomic DNA structure and organization. For example, routine use of Southern blotting led directly to the realization that restriction enzymes do not yield an identical DNA fragmentation pattern from every genomic DNA sample. Polymorphic variability occurring in the primary sequence occasionally ablates or creates new restriction sites in a genome. Restriction digestion and hybridization probing these samples results in altered banding patterns, called restriction fragment length polymorphism (RFLP).[60] Alex Jeffreys responded to this observation by developing ''DNA fingerprinting'' assays based on the Southern blot method.[61] DNA fingerprinting relies on measuring complex combinations of RFLP fragments to document the degree of variability or similarity among individual genomic DNA samples. For this purpose, Jeffreys developed probes that hybridize with a type of DNA locus he discovered during his study of the myoglobin gene. These sites are characterized by VNTRs called minisatellites.[62,63] These repetitive sequences, found throughout the human genome, are termed hypervariable because they exhibit variable numbers of repeats both within a single locus and between loci. VNTRs are still the most common markers used both for genomic DNA mapping and individual identity genotyping. Routine Southern blotting has resulted in the development of a large collection of informative VNTR probes suitable for many analytical applications.

Hybridization analysis using immobilized DNA includes ''dot'' blots and ''slot'' blots. Dot and slot blots are named for the circular and slotted well shapes in the templates used to present test samples to a membrane surface. They eliminate the need for restriction digestion and electrophoretic resolution steps by depositing samples of DNA to be tested directly onto a hybridizable membrane surface. Dot blot methodology is faster and easier for hybridization screening than Southern blotting, especially when many samples are to be screened simultaneously using specific sets of hybridization probes. They can accommodate DNA samples too small or too damaged to undergo the purification, restriction digestion and electrophoresis demanded for Southern blotting, but tend to be somewhat less sensitive and informative than Southern blots since they do not provide RFLP information and are more prone to nonspecific background hybridization.[64,65]

The invention of PCR permitted the development of the ''reverse dot blot'' in which the probe collections, rather than the DNA test samples, become the immobilized species on the membrane surface. DNA to be analyzed is prepared by PCR amplification of the region of interest, labeled with a detectable marker and hybridized to the array of immobilized probes.[66] The reverse dot blot concept led to the idea of generating generic arrays of probe sets that would allow high-throughput screening applications, as in routine molecular diagnostics. Commercial reagent suppliers began to develop generic products in response to the demand for common genomic analyses. The basic components that enabled individual investigators to customize reverse dot blot probe sets for their own applications have also become commercially available to investigators and are widely used throughout the molecular biology community.[67,68]

As the density of information derived from efforts to sequence, map and identify human genes increased, so did the demand for analytical tools capable of exploiting this information. DNA microarrays were developed in response to this demand. Southern[69] was the first to describe parallel, in situ oligonucleotide synthesis as a means of generating oligonucleotide probe arrays on solid supports for highly parallel hybridization analysis. Southern's method uses standard nucleotide synthetic reactions to synthesize the oligonucleotides. The reactions are carried out in a movable chamber, which provides a physical barrier between the reaction chamber and the intended synthesis area.

This method of generating microarrays has been modified in several different ways since it was introduced. In one, photochemistry and photolithography have been incorporated into the oligonucleotide synthesis.[70] This technique uses novel monomeric reagents with photolabile protecting groups that are activated by light exposure; this permits controlled, localized array synthesis to be achieved. In another approach, piezoelectric nozzles and ink jet heads have been modified to ''print'' DNA synthesis reagents directly onto substrates that support in situ oligonucleotide synthesis localized by surface chemistry.[71]

In a sequence of events reminiscent of the Southern blot evolution into the Northern blot method, arrays of immobilized complementary DNA (cDNA) and expressed sequence tags have emerged on the heels of

DNA oligonucleotide arrays to profile mRNA expression patterns. In these ''spotted'' arrays, purified cDNA clones or PCR products are micropipetted as an array on a substrate surface and immobilized by one of a variety of covalent or noncovalent methods.[72–74]

DNA microarrays are distinguished from reverse dot blots by their relatively high probe density, miniature size and use of solid, nonporous supports. Solid supports offer many functional advantages for arrays, as they become integrated components in automated, high-throughput assay systems. Rigid supports, particularly those that are optically transparent and thermally conductive, are more practical than large, flexible membranes for interfacing with automated fluid delivery or printing equipment, which introduce analytical samples, detection reagents and washing solutions. They also best accommodate the automated scanning systems that are used to image the arrays.

3.2 DNA Microarray Capabilities and Limitations

DNA microarrays already occupy niches in nearly every area of nucleic acid analysis. Both comparative or quantitative measurements and direct sequence analysis can be achieved, and microarrays have been designed and fabricated for applications as diverse as genomic mapping,[75] clone library screening, gene expression profiling,[76] genotyping and reference-based sequence checking (resequencing).[77,78] However, despite their universal appeal and versatile functionality, DNA microarrays are analytical tools and should be evaluated in that context. The probe sets that make up an array and their associated functional assays are designed differently depending on the intended application. Arrays provide only a single component of complex biological assays that depend on well-integrated methods for isolation, amplification and labeling of the target to attain high sensitivity and high-resolution detection.

Sequence information is obtained from microarrays by decoding the complementarity between known array probe sequences and the partners they capture from a solution of labeled, hybridized test targets. Genotyping to identify a given polymorphism in a gene or DNA fragment can be used to illustrate how an array provides sequence information. A successful genotyping assay depends on developing maximum discrimination at the single nucleotide level among sets of closely related probes. This aim is best accomplished by selecting probe sets that display maximal signal uniformity and intensity as perfectly complementary duplexes. A genotyping result is interpreted by comparing positive signals that are anticipated from these duplexes with those obtained from sets of closely related mismatched probes. This method provides a background reference that controls for hybridization specificity and permits confident interpretation of the hybridization result. Because mismatched probes do not hybridize well under optimal assay conditions, positive signals that display high intensity against a low nonspecific background are readily perceived.

During the development and fabrication of an array, candidate probes can be evaluated for optimal performance by using them to genotype target DNA samples that have been genotyped by a separate reference method. Confidence in genotype assignment can be improved further by selecting probes to interrogate the target sequence on both strands of target DNA and then combining the information to make a composite base call. Such an assay is not finished until an algorithmic analysis of its output is shown to be concordant with that of the reference for the same sample set. Discrepancies between the two sets of results must be resolved and the source of any difference must be attributed to the candidate or reference probes.[79]

Defining acceptable performance specifications for a genotyping assay requires that a number of considerations be factored together. Because array hybridization and, therefore, genotyping quality are not likely to be completely uniform across complex hybridization arrays, it is critical to establish acceptable limits for the range of reporter and background signals observed; because the occurrence of different polymorphisms may differ in a given sample, individual polymorphisms should be weighted by the frequency at which each phenotype occurs, to achieve reliable prediction of composite genotypes; and because methods of target amplification and labeling of reporter molecules are variable, they may contribute significantly to the range of quality and confidence that will be observed in the results obtained.

Hybridization efficiency and hybrid stability both increase with probe length, and DNA array probes used for expression profiling tend to be longer than those used for genotyping.[80] This permits maximum detection capability for sequences of low abundance in complex RNA mixtures. Mismatch discrimination decreases as probe length increases, but high-resolution sequence discrimination is not a high priority in most expression profiling studies. Quantification of RNA abundance depends on hybridization signal intensity being directly proportional to hybridization target abundance, and maintaining this relationship depends on having a well-designed target preparation strategy. Linear target amplification methods such as in vitro transcription and ligase chain reaction coupled with controlled label incorporation into the amplified target have been quite successful in meeting this requirement. Appropriate sets of controls for use as internal standards for calibration of the assay can readily be developed.[76,81]

The provision of effective gene expression profiling requires comparative and quantitative measurements of RNA. The target sequences in a given sample of RNA will represent a large concentration range of various transcripts, and the goal is to quantify them as accurately as possible. Strategies for expression profiling to meet this objective must maximize detection sensitivity for unique, low-abundance transcripts in complex hybridization mixtures dominated by highly expressed transcripts.

Effective expression-monitoring probes demonstrate hybridization sensitivity over a wide range of target concentrations. Global optimization of hybridization performance for expression-profiling arrays is as challenging as for genotyping arrays. The large number of probes in an array, coupled with the innate variability of the RNA pools used in the hybridization, make it impractical to test concentration dependence and sequence specificity for each probe independently. Consequently, other broad-based performance optimization strategies have been pursued. Designs that have proven successful for expression profiling include multiple hybridization probes for each mRNA sequence of interest, and related mismatch probes to estimate the contribution of nonspecific interactions to each specific hybridization signal. Redundancy also permits probes to be tailored to detect 3′ and 5′ ends and specific exons in mRNAs. These probes may be useful to gauge the quality of RNA and cDNA target preparations and to dissect splice variants or other closely related members of gene families. It has been reported that from four to twenty probe pairs (one fully matched to a target sequence and one mismatch control in each pair) may define the limits of useful redundancy in oligonucleotide-based gene expression profiling arrays.[76]

A second, more common type of DNA microarray for expression profiling uses cloned cDNA or PCR products amplified from cDNA rather than oligonucleotides as capture probes. These DNA products are collected as libraries and robotically deposited onto chemically prepared solid substrates as addressable hybridization arrays. The original method for making these arrays relied mainly on noncovalent associations between the support surface and the one or two kilobase lengths of deposited DNA to keep the clones immobilized during hybridization.[72,82] However, this method has rapidly evolved to include a wide variety of substrate surfaces, coupling chemistries and methods of delivering the probe DNA accurately and reproducibly to the surface. Both commercial and private efforts to optimize these ''spotted'' arrays are yielding ever more uniform and reproducible arrays. One advantage over oligonucleotide arrays that is gained with immobilized cDNA is that any clone from a cDNA library, whether or not it has been sequenced, can serve as a probe to screen for complementary sequences in other RNA sources. In contrast, some primary sequence must be known for each target to generate synthetic oligonucleotide probes for an array. Hence, these two types of expression-monitoring array actually provide complementary functions. The other obvious contrast is that cDNA, because of its length, is likely to hybridize not only with its intended target but also with other closely related sequences, such as those within gene families. Oligonucleotide probes are more easily customized to discriminate among related sequences and so may offer a more refined expression profile for individual related sequences.

Microarrays based on cDNA or oligonucleotides differ fundamentally in how their experimental outputs are analyzed. Typical cDNA arrays are hybridized with differentially labeled cDNA pools generated from two separate RNA sources. For example, RNA may be harvested from untreated cells grown under a standard set of conditions and the cDNA produced from this RNA pool may be labeled with one fluorescent dye. A second RNA sample is then harvested from the same cell type after it is treated with a chemical or grown under a different set of conditions, and the cDNA is labeled with a second dye. Once the two cDNAs can be distinguished by their labels, equal amounts are mixed and hybridized competitively to the same cDNA array. The ratio of one dye signal to the other at each probe will reflect relative differences in abundance between the two RNA samples for the gene represented.

In contrast, a typical oligonucleotide array is usually hybridized with a single RNA pool labeled using a single reporter molecule. A measure of the hybridization intensity at each probe is obtained using internal calibration standards. The success of this approach depends on having a way to normalize the signal at each probe for any nonspecific background contributing to the signal intensity. This is usually achieved, as mentioned previously, by pairing a closely related, mismatched probe for each specific probe in the array. In addition, use of a redundant pair of probes for each target helps to average out this variability.

As the amount of available information about expressed genes accumulates and the sequence of the entire human genome emerges, successive generations of DNA microarrays have become increasingly dense in probe sites. The result is a dramatic increase in information gathered from each profiling experiment, and greater insight into the complexities of cellular biology. This gain is accompanied by target preparation requirements of increasing complexity to access the information that an array offers. What the useful limit to hybridization complexity will be is not yet clear. Entire cDNA libraries and the entire yeast genome have been successfully hybridized to DNA microarrays to yield large volumes of information

from single experiments, but a method for whole genome typing of a target organism in a single hybridization is still out of reach. Currently, PCR or some alternative strategy is required to achieve informative human genotyping by microarray hybridization. Variability in performance potential among hybridization probes and nonspecific interactions on the array surface (both within targets and between targets and probes) limit full reconstruction of such a complex target from a single hybridization.

The measurements made using DNA microarrays are very difficult to optimize to a tightly specified standard because of the high density of information and its parallel nature. It is equally difficult to validate individual probes globally in microarray hybridization assays at every data point, since it is impractical to check every point by a reference assay method. Statistical approaches to solving this problem include checking a selected sample of output data by an independent assay and testing statistical variability within and among repeated assays. A second strategy to validate the results of microarray hybridization is based on the integration of sets of complementary hybridization controls directly into the array.

The process of establishing and validating DNA microarrays is further constrained by the fact that their functional use and performance evaluation must be in the context of a broader, integrated system. That is, peak array performance occurs when probe design has been tuned to the chosen target, the labeling method is efficient and reproducible, and the imaging system is sensitive, calibrated and well matched to the reporter molecule used. Finally, sensible output analysis depends on the specific assay application and the control elements built into the assay. Arrays, especially very high-density arrays, are most efficiently used with automated interfaces that integrate all assay steps from sample preparation, through hybridization, data acquisition and output analysis. Ultimately, full microarray assay automation should become available from sample input to output of the genotype or expression profile. Progress in microarray development has been impressive, but an equally impressive amount of work remains to be done before these methods will be fully robust clinical tools.

3.3 Predictive Testing with DNA Microarray Assays

Predictive testing based on DNA microarray profiling is not available at the present time, although it is widely promoted. True predictive pharmacogenetic testing relies on having broad individual genotyping capability and integrated rational therapeutic management strategies for disease guided by genetic predisposition and risk profiles. Reaching this goal will require more than just technical development. The pharmacokinetic and pharmacodynamic end-points that are used to describe therapeutic indications and responses today need to be integrated into a coherent picture and translated into molecular end-points. Our fundamental knowledge of the physiological effects of therapeutic agents must also be reduced to a molecular level. Pharmacogenetic traits are remarkable prognosticators of individual responses to drugs, exogenous chemicals and xenobiotics, but our present level of understanding of the pharmacokinetic and pharmacodynamic effects of drug response at the molecular level is rudimentary.

Our current state of knowledge of human drug response stems largely from retrospective genetic analyses guided by phenotypic responses to specific drugs. As these observations are studied in greater depth, it is becoming clear that few, if any, phenotypic outcomes will be reliably predicted from simple genetic analyses at a single locus. It is also clear that even when complex genetic analysis becomes more sophisticated and widely available, predicted phenotypes will still represent a potential rather than a certain outcome. Translating predicted potential into an outcome prediction will still depend on factoring environmental variables such as diet, age, drug use, health status and environmental exposure history into the equation. It is also likely that clinical monitoring of therapeutic responses will not be replaced by, but rather will be guided by, predictive testing. For example, a patient may have a genotype that predicts metabolic tolerance for a cardiotoxic medication with a narrow therapeutic index. The drug may be given to him because it most effectively targets his genetically profiled therapeutic need. However, blood monitoring for toxic drug levels may still be required since the effect of illness on his metabolic performance is not predicted by his genetic profile.

Two radically different approaches have been proposed for developing therapeutics based on known molecular mechanisms for individuals whose phenotypic responses are projected from their individual genetic profiles. One approach is to generate a high-resolution map of the human genome using characterized polymorphic markers, especially single nucleotide polymorphisms. Once the map is available, large-scale population epidemiology and case–control studies are expected to provide high-quality associations between genetic markers, specific disease and therapeutic response phenotypes. These population-derived markers are anticipated to be useful for projecting individual outcomes and risk assessments. The alternative approach has been to find individual genes that give rise to distinctive phenotypes and study them in depth. Each gene is fully characterized for polymorphic variability and each variant is characterized by its impact on the phenotype. As this information is assembled, phenotypes may be recognized that cannot be explained by known genetic

variants, and genetic variants will be recognized that do not lead to separate phenotypes. These observations may, in time, lead to the discovery of new functional relationships with other genes. Both strategies claim some success and both contribute to our basic understanding of human genetic variability. Whether one approach is superior to the other remains to be seen.

3.4 Emerging Applications of Microarray Technologies

Microarray technologies are useful for many applications requiring direct nucleic acid sequence analysis, sequence comparisons or quantitative comparison among nucleic acid samples. Limitations exist where direct hybridization analysis is not the best-suited approach, as for analysis of trinucleotide expansion (fragile X and Huntington's disease) and microsatellites. Microarray applications are not limited to human genetics but can also be used for target organisms such as yeast, model organisms *Arabidopsis thaliana* and *Caenorhabditis elegans*, viral and bacterial human pathogens, mouse, rat and ape genomes, and a wide variety of other agricultural plant, pest and animal targets. It is likely that this entire spectrum of genomic studies will ultimately contribute significantly to clinical applications of pharmacogenetics. Human gene complementation applied to regulatory pathways in model organisms has a strong track record for predicting parallel human biology. In addition, these models provide powerful pharmacokinetic and pharmacodynamic discovery assays and toxicology screening systems. Continued expansion of its technical foundations is expected eventually to provide combined genomic profiling for genetic variability and functional expression profiling capability as single assays.

All early microarray assays relied on hybridizing a labeled target or test nucleic acid directly to probes covalently linked to a substrate surface in an addressable format. Signal location provided sequence information about the hybridization target; signal intensity provided semiquantitative information about target abundance in the hybridization solution. The need for improved quantitative and discriminative performance from microarray assays has led to rapid innovation and improvement in a number of different areas: in the design of initial arrays and assays; in modifications in nucleotide chemistry,[83] linker chemistry,[84] substrate preparation chemistry[71] and hybridization solution compositions,[85] which permit ever-increasing refinement and control over assay performance; in improved array components, which permit performance of increasingly complex and sensitive analyses; and in the coupling of primary hybridization assays to secondary enzymatic processes, such as primer extension and enzymatic mismatch cleavage, to increase the quantity and quality of information yielded by a single assay.[86,87]

Applications of microarrays to genotyping include typing of samples for known sequence variants and discovery of new polymorphisms in specific targets, such as drug-resistant pathogens and functionally variant forms of known genes. Arrays for these types of test perform best with probes that have similar hybridization behavior, and efforts to design such probe sets have proven successful. Hybridization to complete collections of oligonucleotide probes of a given length shows a wide range of variability that correlates directly with individual nucleotide composition. An alternative method of design that avoids such variability is to target array probe sets to a narrow range of thermal stability. Probes designed by this criterion may still vary in length, but they show much more uniform behavior within a single set of hybridization conditions. Discrimination of the hybridization signals of fully complementary hybrids from the uninformative signals that result from associations between partially matched nucleotide fragments is much improved by this approach. Further refinement of hybridization uniformity has been obtained by incorporating nucleoside analogs that smooth thermal stability and disrupt intra-probe hairpin structures and mismatch associations. These assay modifications help to maximize discrimination among specific sets of related nucleotide sequences. These developments have been important in the effort to provide arrays for sequencing applications where samples are hybridized to an array of oligonucleotides that reconstructs a consensus sequence.[83,88] Probes that fail to hybridize to this type of array indicate the presence of a polymorphism in the test target sequence that causes loss of complementarity with the consensus probe.

The density of information from microarrays has increased in parallel with technical advances in robotic microfluidic delivery systems and photolithographic techniques for fabrication of arrays. The result is an increased yield of information from single assays, but this comes at the cost of the increased complexity in preparing a test sample for hybridization in a way that uses the array optimally. For genotyping or sequencing applications, this usually means that a large number of PCR products must be generated in essentially equal quantities and labeled as uniformly as possible with a reporter molecule that permits hybrid detection. For expression-profiling experiments, RNA must be quantitatively isolated from the test source, amplified enough to permit rare mRNA detection but not enough to lose the quantitative relationships among the various mRNA species. Uniform labeling of samples with a reporter molecule is critical to keep the hybridization signal of each mRNA proportional to its abundance. These analytical demands have been met with a variety of innovations in array design and target labeling strategies.

As microarray technology has advanced, an abundance of new and challenging analytical applications have arisen, for example cancer biology. Many cancers are characterized by nucleic acid changes in somatic cells. The challenge here is in finding efficient ways to present the very limited amounts of test sample available from a biopsy to arrays for hybridization analysis. Loss of heterozygosity is another important somatic cell change indicative of cancer progression that can be detected only by quantitative genetic analysis. Detection of drug-resistant pathogens is another quantitative analytical problem, since genotypes of resistant organisms emerge against a background of susceptible genomes. Similarly, assessment of host selectivity for a given pathogen demands the detection of virulent subtypes occurring at low frequency in a background of benign genotypes. Studies to document human genetic variability for association studies to link our genes to our environmental responses on a population basis require true high throughput genotyping for large numbers of genetic markers. These studies present challenges in the technical aspects of genotyping by microarrays and in bioinformatics.

3.5 The Future of Microarrays and Allied Technology

Despite remarkable innovation and rapid technical improvement in microarray technology, several significant limitations in its application remain. Methods of fabrication need improvement to enhance their utility and reduce costs of construction. The availability, utility and efficiency of bioinformatics methods used to access, process and store array information are also limited. Preparation of an adequate quantity of high-quality, labeled target DNA for a given assay still presents a major difficulty in the use of DNA microarray technology. This is particularly problematic for gene expression profiling, where quantitative output is important and the relative abundance of the starting materials must be preserved during amplification and labeling. In addition, as more and more of the genome is accessed for profiling and genotyping, the problem of preparing amplified material from all relevant sites in the genome presents significant technical and economic burdens. At present, methods of multiplexed PCR amplification are not sufficiently robust or reliable to permit quantitative production of target material on a genomic scale. In addition, the cost of enzymatic amplification and labeling reagents for large-scale target preparation is significant.

A better method for highly parallel genetic analysis is needed. One with single molecule sensitivity that eliminates both the requirement for target amplification and the need for a target to be chemically modified or labeled for its detection would be ideal. Innovations in array construction, sample nucleic acid preparation and detection hold this promise for a second generation of microarray technologies. For instance, Heller and colleagues have developed arrays with bioelectronic chips that are electronically addressable.[89] An electric current can be used to address each feature of the chip, enabling direct capture probe immobilization as well as control of hybridization stringency with the target molecules. This type of array has been shown vastly to accelerate target capture rate by using electric fields at each probe site to overcome the passive diffusion that previously had been a rate-limiting factor.[90,91]

A number of completely different approaches to array-based genetic analysis show considerable promise. One consists of a DNA microarray combined with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOFMS) detection. Technical development of this system is moving ahead rapidly. MALDI/TOFMS detects sequences by measuring their mass, which is attractive because it does not require separate labeling of the target. Initially, the principal limitation of MALDI/TOFMS lay in its inability to discriminate large molecules with sufficiently high resolution, but strategies have been devised that mitigate this problem. A second disadvantage of MALDI/TOFMS is the relatively large amount of sample material needed to achieve reliable results. MALDI/TOFMS systems are commercially available, and the development of methods to apply this system to determination of genotypes is accelerating.[92,93]

Another unique approach to microarray technology uses fiber optic bundles to support an array of microbeads that are coated with DNA capture sequences. Each bead is individually associated with a single fiber in the bundle, and hybridization targets labeled with fluorescent dyes are hybridized to the beads. The one-to-one concordance of bead to fiber in the bundle permits every bead in the hybridization array to be addressed and reported individually.[94,95]

DNA microarrays, even at their current state of technical evolution, are proven analytical tools with sufficient capacity to meet future needs of genetic assays. There are, however, technical challenges that must be met before they can reach their full potential, as well as significant challenges in bioinformatics and information management that applications of microarrays have created. Microarray technology generates such vast quantities of information that powerful databases, only now emerging, are necessary to analyze the information obtained and to store it in a format that is accessible. New concepts in database mining will be necessary to relate these large, complex sets of genetic data to the phenotypes recorded in clinical databases. The new knowledge gained from the

discovery of associations between genotypes and phenotypes of individuals in response to exogenous agents should enable clinical applications of pharmacogenetics to develop.

3.6 Summary

Currently, microarray applications span all aspects of molecular biological investigation, enabling the collection of large amounts of genetic data from comparatively few experiments. At present, microarrays are used mainly as research tools to determine the most effective ways of gathering information for individual genetic profiles and management of unexpected responses to medical therapy and environmental substances. With further development, they should contribute substantially to point-of-care analysis and therapeutic management. The full potential of the technique will only be realised when bioinformatics to analyze large datasets are also fully developed.

4 THE OUTLOOK FOR GENOMICS IN PREDICTIVE MEDICINE

4.1 Options for Treating Genetic Disorders

The care of patients with genetic disorders involves many of the approaches to diagnosis and treatment that are used in other medical specialities, but the focus in medical genetics is more toward prevention or avoidance of the disorder. Presymptomatic (including prenatal) diagnosis, genetic screening programs and genetic counseling are concepts central to this approach.

Options for treatment of genetic disorders at the environmental level routinely involve a combination of restriction, replacement and removal of the toxic substance. Restriction of potentially toxic environmental substances could, for example, involve restriction of certain foods or other dietary constituents. Fructose is one of the major constituents of human diets, and intolerance for this sugar is represented by a polymorphism in the aldolase B gene.[96–98] Continued ingestion of fructose or its congeners is required to establish the disease, but the wide distribution of these substances in foods and some drugs places genetically susceptible persons at constant risk from an avoidable nutritional disorder. The fish odor syndrome (trimethylaminuria) is another example of variability in human response to foods.[99] A polymorphism in the gene for flavin-monoxygenase 3 accounts for some cases of this trait.[100] Trimethylamine, a metabolic product of numerous foods that contain choline or carnitine, confers the smell of rotting fish on affected persons, with devastating educational, economic and social consequences. Management of trimethylaminuria remains empirical; the main recourse for affected persons is dietary restriction to reduce their intake of trimethylamine precursors. This may require avoidance of foods rich in choline, such as eggs, liver, soya beans, and marine fish.[4]

Restriction of certain therapeutic agents is an option that is commonly applied by physicians to avoid adverse drug reactions in genetically susceptible patients. Individuals with a deficiency in G6PD (Table 2) must avoid oxidant stresses such as those that accompany intake of antimalarials, sulfonamides and an extensive list of other drugs; failure can result in hemolysis of red blood cells. Individuals with thiopurine methyltransferase deficiency (Table 2) usually must receive much smaller doses of 6-mercaptopurine, 6-thioguanine, and 8-azathiopurine if they are to avoid toxicity (bone marrow suppression) by these agents. These drugs are mainstays of antileukemic and immunosuppressant therapy. Those with this inherited deficiency are intolerant to these agents and may suffer acute and delayed responses that can be life threatening.[46]

Toxicity resulting from exposure to cigarette smoke in persons afflicted with α_1-antitrypsin deficiency has been mentioned above (section 1.3). Avoiding cigarette smoke can prevent its destructive effects on the lungs, which result in emphysema and chronic obstructive lung disease.[9]

Replacement of deficient gene products or even of organs is also utilized in the treatment of genetic disorders: for example, replacement of coagulation factor VIII in hemophilia A, of α_1-antitrypsin in persons deficient in this factor or of pancreatic islet cells in some forms of diabetes mellitus.

The option to remove the toxin from the environment of susceptible persons is taken in hemochromatosis, a common disorder of iron metabolism[101] that affects 1 in 300 persons of northern European descent. Most cases of hereditary hemochromatosis are attributed to a polymorphism of *HFE*, which results in substitution of tyrosine for cysteine at codon 282. The disease, untreated, causes liver cirrhosis, heart failure, diabetes and arthritis and leads to early death; treatment by phlebotomy to remove excess iron allows affected persons to live a normal life span.

4.2 Many Pharmacogenetic Polymorphisms are Exceptions to Experience with Single Gene Disorders

The point was made above (section 1.6) that pharmacogenetic polymorphisms usually have no perceptible effect on the health of predisposed persons, unlike other genes involved in human diseases. The frequent occurrence

and ethnic specificity of human pharmacogenetic polymorphisms are further features that set them apart from many human disease genes.

The gravity of effects that may be in store for susceptible persons is only revealed by exposure to certain drugs, dietary constituents, or other environmental toxins; without this exposure, their genetic predisposition is invisible. Most people perceive heredity as having an important effect on health and disease, but may be unaware of the relevance of genetics to their own responses or of person-to-person differences in response to environmental toxins. The belief is widely held that unexpected responses to these substances are entirely linked to the toxin, but we know that the drug recipient is not a passive participant in these events.

Hemochromatosis (1/300)[101] and cystic fibrosis (1/2000 to 1/2500)[102] are regarded as highly prevalent genetic disorders in human populations, but they are rare compared with many pharmacogenetic polymorphisms. Polymorphisms of *G6PD*, *CYP2D6*, *CYP2C19* and *NAT2* illustrate this point very well.[103] G6PD deficiency is a sex-linked trait that affects 5–10% of dark-skinned races. Males of African, Mediterranean, and Oriental descent are particularly susceptible. Globally, this deficiency is estimated to affect more than 400 million persons.

CYP2D6 polymorphism has been the subject of many population studies. More than 95% of deficient alleles responsible for CYP2D6 poor metabolizers have been identified (see section 3.2). Eleven nucleotide changes belonging to seven *CYP2D6* allelic variants describe the extent of ethnic variation in people of Africa, Asia and Europe.[104,105] The prevalence and ethnic specificity of the ultrarapid CYP2D6 phenotype described recently indicates that Swedish and German populations possess 1–2% of this phenotype, whereas the frequency is higher in Spain (3.5–5%); the carrier (heterozygote) frequency in Ethiopia and Saudi Arabia is also very high (15–20%).[106]

The prevalence and ethnic specificity of CYP2C19, another polymorphic P450 enzyme (Table 2), also varies quite remarkably. CYP2C19 polymorphism, previously named mephenytoin polymorphism, is of major importance because of its role in elimination or activation of drugs such as omeprazole, an antiulcer drug, proguanil, an antimalarial drug, and several barbiturate hypnotics and sedatives.[107] The frequency of poor drug metabolizers owing to the m1 (*CYP2C19*2*) variant is high in Japanese (13–23%) compared with Caucasians (4%). Another variant, m2 (*CYP2C19*3*), occurs in Japanese and Africans but has not been detected in Caucasians. The wild-type and the m1 and m2 variants account for almost 100% of the variation in Asian poor metabolizers, but these and additional variants account for only about 92% of the variation in Caucasians.[108]

Studies of the prevalence and global distribution of *NAT2** polymorphism have involved well over 10 000 subjects in dozens of populations since the polymorphism was identified in the 1950s (Table 2 and section 2.5). The percentage of slow acetylators ranges from 80% or more in Egyptians and certain middle Eastern populations to 20% or less in Japanese and Canadian Eskimos. Populations of European and African origin are, with few exceptions, characterized by intermediate percentages of the slow acetylator phenotype.[55] Molecular studies indicate that *NAT2** allelic frequencies are neither uniformly nor randomly distributed across different populations. For instance, the frequency of the rapid acetylator *NAT2*4* allele is approximately 20–25% among American and European Caucasians, 36% among African-Americans, 42% among Hispanics, and 66–70% among Asiatics from Hong Kong, Korea and Japan. The distribution of *NAT2** alleles for slow acetylation varies with racial origin; in Caucasians, three alleles (4, 5B and 6A) account for about 95% of all *NAT2** alleles, whereas in Oriental populations the 5B allele becomes rarer (5% in Hong Kong Chinese to <1% in Japanese) and 6A (20–30%) and 7A/7B (7–16%) are most frequent.[103]

4.3 Emerging Applications of Pharmacogenetic Interest

Areas of focal importance for the pharmaceutical industry and for those engaged in the development of microarray and allied technologies include adverse drug reactions, diagnostics and targeted therapies. The versatility and broad capability of microarray technology permits it to be applied to any of these areas for the study of variation in genes, pathways and targets associated with disease.

The main challenge for the immediate future is to devise reliable methods and techniques for identifying and scoring all types of genetic variation in the human genome. Several strategies and technologies for identification of such variation at the molecular level are being developed (sections 3.3–3.5). Questions have been raised as to whether microarrays have the capability to identify different types of mutation, and whether they can distinguish heterozygotes. Cronin et al.[79] examined the first question in a study of *CFTR*. They found that the gene, composed of 27 exons, contained more than 500 widely distributed mutations, including transitions, transversions, insertions, deletions and polymorphisms. Two microchip arrays (480 and 1480 probes of 14–16-mers) were used to scan exon 11. Ten unknown genotypes assigned by these assays were identical with PCR product restriction fragment analysis.

Hacia et al.[77] looked at detection of heterozygotes predisposed to breast and ovarian cancer in a study of *BRCA1*. They found that it comprised 22 exons spanning 100 kb and contained more than 110 mutations

(frameshift, nonsense and missense) distributed throughout the gene. A microchip array with 96 500 20-mer nucleotides used to scan 3.45 kb of exon 11 accurately diagnosed 14 of 15 patient samples, and no false positives were found among 20 controls. Even though the full potential of these devices is not yet realized, these studies, among the first to report applications of microarrays to genes associated with disease, demonstrate very clearly the capability of this technology for genetic analysis.

The analysis of pathways affected in disease is another emergent application of this technology. Pathways that regulate cytokine signaling, insulin signaling and apoptosis are among those identified for early investigation. Targets of high priority for study include G-protein transmembrane receptors, ion channels and transcription factors. The nucleic acids targeted for these and some other emerging applications of focal interest are mRNA (cytokine induction, tumor suppression), DNA (cytokine mutants), germline DNA (*BRCA1*), yeast DNA (barcoding mutants), DNA of human immunodeficiency virus (protease-resistant polymorphisms), mitochondrial DNA (natural polymorphisms) and *CYP2D6* and *CYP2C19* DNA (drug-metabolizing enzyme polymorphisms).

4.4 Linking Human Genotyping to Clinical Applications

The ultimate goal of pharmacogenetics is to gather information that explains unexpected responses to exogenous chemicals and that prevents or avoids the occurrence of these responses in genetically susceptible persons (see section 1.6). Among all the possibilities emerging from, or accelerated by, the human genome initiative,[109] pharmacogenetics is an area that may rapidly bring the predictive prospects of human genotyping into the clinical arena. In essence, a given pharmacogenetic polymorphism is characterized by three types of information. These are the genetics (mode of inheritance, allelic frequencies, and ethnic specificity), the molecular basis (genes responsible, and their mutation spectrum) and the medical or biological significance of the polymorphism.

Ethnic specificity not only provides information about unique features of a population that is of evolutionary interest, but may also explain situations of therapeutic interest. Consider the administration of a drug metabolized by the CYP2D6 system, such as codeine, to patients of different ethnicity. Codeine is an analgesic prodrug, the effectiveness of which depends on its conversion to morphine. This conversion is impaired in poor *CYP2D6* metabolizers, and consequently they do not get pain relief from ingestion of codeine.[110] For the same reason, they are also protected against dependence on oral opiates such as codeine.[111] The failure to respond to codeine would, consequently, be more prevalent among Africans and Caucasians (5–10%) than among Asians (<1%). Unexpected neurotoxicity from amphetamine and its analogs (e.g. 3,4 methylenedioxymethamphetamine, also known as ''ecstasy'') as well as interactions with other drugs subject to *CYP2D6* polymorphism would likewise be expected to occur more frequently among Africans and Caucasians than in Asians.[112]

The analysis of the molecular basis for unexpected responses of ultrarapid *CYP2D6* metabolizers to codeine follows a similar line of reasoning. The ultrarapid metabolizers have enhanced capacity to metabolize codeine and hence may exhibit exaggerated responses, such as abdominal cramping, fuzzy vision and disorientation.[113] Because ultrarapid metabolizers occur more frequently among Hispanic, African and Saudi Arabian than Asian populations,[106] the former groups would be more likely to experience *exaggerated* responses to codeine.

4.5 Summary

Genetic disorders with abnormal reactions to environmental factors are no less treatable than many other disorders treated by physicians. These genetic disorders are treated by restricting access to, replacing and/or removing the toxic substance from the environment of the susceptible persons. Pharmacogenetic polymorphisms are invisible until their effects are revealed by exposure of individuals who harbor them to environmental toxins; microarrays and allied technology provide the means to screen for these polymorphisms and to identify susceptible individuals before they are affected.

5 SUMMARY: PROGRESS IN PHARMACOGENETIC TESTING

Individual susceptibility to the effects of drugs and other exogenous chemicals is often the consequence of some derangement of the genetic material that may be transmissible from one generation to the next. The scientific study of these effects is the province of pharmacogenetics. The main purpose of pharmacogenetics is to understand the causes of unexpected responses to exogenous substances and to prevent or avoid their occurrence in susceptible persons. Prior to the development of recombinant DNA technology, the role of changes in the genetic material and the consequent production of aberrant proteins, or failure to produce proteins, in disease could only be inferred. The development of microarrays and allied technologies were direct outgrowths of recombinant DNA technology and can be used to identify and screen for all types of molecular genetic variation affecting genes, pathways and molecules relevant to human genetic disease, including variations of pharmacogenetic importance. Microarray

applications currently span all aspects of molecular biological investigation, mainly in research regarding the analysis of genetic profiles of humans and other targeted organisms. The application of microarrays with bioinformatic technology is on the threshold of contributing substantially to improved diagnosis and therapeutic management of patients in a clinical setting.

ABBREVIATIONS AND ACRONYMS

bp	base pairs
cDNA	Complementary DNA
G6PD	Glucose-6-phosphate Dehydrogenase
MALDI/TOFMS	Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry
mRNA	Messenger RNA
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
VNTR	Variable Number Tandem Repeat

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory • Biosensor Design and Fabrication • Capillary Electrophoresis in Clinical Chemistry • DNA Arrays: Preparation and Application • Drugs of Abuse, Analysis of • Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Product Development for the Clinical Laboratory

Forensic Science **(Volume 5)**
Forensic Science: Introduction • DNA Extraction Methods in Forensic Analysis • Polymerase Chain Reaction in the Forensic Analysis of DNA

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Capillary Electrophoresis of Nucleic Acids • Comparative Genomics: Differential Display and Subtractive Hybridization • DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Fluorescence In Situ Hybridization • Mass Spectrometry of Nucleic Acids • Nuclear Magnetic Resonance and Nucleic Acid Structures • Nucleic Acid Structural Energetics • Optical Mapping in Genomic Analysis • PNA and Its Applications • Polymerase Chain Reaction and Other Amplification Systems • Radiation Hybrid Mapping • Restriction Landmark Genomic and cDNA Scanning • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Sequencing Strategies and Tactics in DNA and RNA Analysis • X-ray Structures of Nucleic Acids

Electronic Absorption and Luminescence **(Volume 12)**
Absorption and Luminescence Probes • Fluorescence Imaging Microscopy • Fluorescence Lifetime Measurements, Applications of • Indirect Detection Methods in Capillary Electrophoresis • Surface Measurements using Absorption/Luminescence

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Multivariate Image Analysis • Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration

REFERENCES

1. A.E. Garrod, 'Medicine from the Chemical Standpoint', *Lancet.*, **ii**, 281–289 (1914).
2. A.E. Garrod, *Inborn Factors in Disease: An Essay*, Oxford University Press, New York, 1931.
3. A.F. Blakeslee, M.R. Salmon, 'Odor and Taste Blindness', *Eugenical News*, **16**, 105–110 (1931).
4. W.W. Weber, *Pharmacogenetics*, Oxford University Press, New York, 1997.
5. A.G. Motulsky, 'Drug Reactions, Enzymes and Biochemical Genetics', *J. Am. Med. Assoc.*, **165**, 835–837 (1957).
6. W. Kalow, *Pharmacogenetics – Heredity and the Responses to Drugs*, W.B. Saunders, Philadelphia, PA, 1962.
7. A.G. Motulsky, 'Pharmacogenetics and Ecogenetics in 1991', *Pharmacogenetics*, **1**, 1–2 (1991).
8. F.P. Guengerich, 'Human Cytochrome P450 Enzymes', in *Cytochrome P450 – Structure, Mechanism, and Biochemistry*, 2nd edition, ed. P.R. Ortiz de Montellano, Plenum Press, New York, 473–517, 1995.
9. D.W. Cox, 'α_1-Antitrypsin Deficiency', in *The Metabolic and Molecular Bases of Inherited Disease*, eds. C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, 7th edition, McGraw-Hill, New York, 2407–2437, 1995.
10. M.L. Piaezza, E.M. Sellers, R.F. Tyndale, 'Nicotine Metabolism Reduces Smoking', *Nature*, **393**, 750 (1998).

11. M. Oscarson, R.A. McLellan, H. Gullsten, Q.-Y. Yue, M.A. Lang, M.L. Bernal, B. Sinues, A. Hirvonen, H. Raunio, O. Pelkonen, M. Ingelman-Sundberg, 'Characterisation and PCR-based Detection of a *CYP2A6* Gene Deletion Found at a High Frequency in a Chinese Population', *FEBS Lett.*, **448**, 105–110 (1999).
12. J. LeWalter, U. Korallus, 'Blood Protein Conjugates and Acetylation of Aromatic Amines: New Findings on Biological Monitoring', *Int. Arch. Occup. Envir. Health*, **56**, 179–196 (1985).
13. R.A. Cartwright, R.W. Glashan, H.J. Rogers, R.A. Ahmad, D.B. Hall, E. Higgins, M.A. Kahn, 'The Role of *N*-Acetyltransferase Phenotypes in Bladder Carcinogenesis: A Pharmacogenetics Epidemiological Approach to Bladder Cancer', *Lancet*, **ii**, 842–846 (1982).
14. E. Beutler, 'Glucose-6-Phosphate Dehydrogenase Deficiency', *N. Engl. J. Med.*, **324**, 169–174 (1991).
15. S. Higuchi, T. Muramatsu, K. Shigemori, M. Saito, H. Kone, M.C. Dufour, T.C. Harford, 'The Relationship between Low K_m Aldehyde Dehydrogenase Phenotype and Drinking Behavior in Japanese', *J. Stud. Alcohol.*, **53**, 170–175 (1992).
16. D.G. McCarver, H.R. Thomasson, S.S. Marteir, R.J. Sokol, T.-K. Li, 'Alcohol dehydrogenase-2,*3 Allele Protects Against Alcohol-related Birth Defects Among African Americans', *J. Pharmacol. Exp. Ther.*, **283**, 1095–1101 (1997).
17. R.K. Ross, J.-M. Yuan, M.C. Yu, G.N. Wogan, G.-S. Qian, J.-T. Tu, J.D. Groopman, Y.-T. Gao, B.E. Henderson, 'Urinary Aflatoxin Biomarkers and Risk of Hepatocellular Carcinoma', *Lancet*, **339**, 943–946 (1992).
18. T.D. Arias, L.F. Jorge, R. Barrantes, 'Uses and Misuses of Definitions of Genetic Polymorphism. A Perspective from Population Pharmacogenetics', *Br. J. Clin. Pharmacol.*, **31**, 117–118 (1991).
19. U.A. Meyer, 'Genotype or Phenotype: The Definition of a Pharmacogenetic Polymorphism', *Pharmacogenetics*, **1**, 66–67 (1991).
20. E.B. Wilson, Jr, *An Introduction to Scientific Research*, McGraw-Hill, New York, Chapter 8, 1954.
21. D. Voet, J.G. Voet, *Biochemistry*, John Wiley, New York, 1990.
22. H.W. Harris, *The Principles of Human Biochemical Genetics*, 3rd edition, Elsevier/North Holland Biomedical Press, New York, 1980.
23. W.W. Weber, 'Influence of Heredity on Human Sensitivity to Environmental Chemicals', *Environ. Mol. Mutagen.*, **25**(Suppl.), 102–114 (1995).
24. J.D. Watson, M. Gilman, J. Witkowski, M. Zoller, *Recombinant DNA*, W.H. Freeman, San Francisco, 1992.
25. K. Muratanai, T. Hada, Y. Yamamoto, T. Kaneko, Y. Shigeto, T. Ohue, J. Furuyama, K. Higashino, 'Inactivation of the Cholinesterase Gene by *Alu* Insertion: Possible Mechanism for Human Gene Transposition', *Proc. Natl. Acad. Sci. USA*, **88**, 11315–11319 (1991).
26. J. Veldscholte, C. Ris-Staplers, G.G.J.M. Kuiper, G. Genster, C. Berrevoets, E. Classen, H.C.J. van Rooij, J. Trapman, A.O. Brinkmann, E. Mulder, 'A Mutation in the Ligand Binding Domain of the Androgen Receptor of Human LNCAP Cells Affects Steroid Binding Characteristics and Response to Antiandrogens', *Biochem. Biophys. Res. Commun.*, **173**, 534–540 (1990).
27. R.C. Skoda, F.J. Gonzalez, A. Demierre, U.A. Meyer, 'Two Mutant Alleles of the Human Cytochrome P-450db1 Gene (*P450C2D1*) Associated with Genetically Deficient Metabolism of Debrisoquine and Other Drugs', *Proc. Natl. Acad. Sci. USA*, **85**, 5240–5243 (1988).
28. N.M. Woolhouse, M. Eichelbaum, N.S. Oates, J.R. Idle, R.L. Smith, 'Dissociation of Co-regulatory Control of Debrisoquine/Phenformin and Sparteine Oxidation in Ghanians', *Clin. Pharmacol. Ther.*, **37**, 512–521 (1985).
29. Q.Y. Yue, L. Bertilsson, M.L. Dahl-Puustinen, J. Sawe, F. Sjoqvist, I. Johansson, M. Ingelman-Sundberg, 'Dissociation between Debrisoquine Hydroxylation Phenotype and Genotype among Chinese', *Lancet*, **ii**, 870 (1989).
30. E. Steiner, L. Iselius, G. Alva, J. Lindsten, F. Sjoqvist, 'A Family Study of Genetic and Environmental Factors Determining Polymorphic Hydroxylation of Debrisoquine', *Clin. Pharmacol. Ther.*, **38**, 394–401 (1985).
31. N. Rothman, G. Talaska, R. B. Hayes, V.K. Bhatnagar, D.A. Bell, V.M. Lakshmi, S.K. Kashyap, M. Dosemeci, R. Kashyap, F.F. Hsu, M. Jaeger, A. Hirvonen, D.J. Parikh, B.B. Davis, T.V. Zenser, 'Acidic Urine pH is Associated with Elevated Levels of Free Urinary Benzidine and *N*-Acetylbenzidine and Urothelial Cell DNA Adducts in Exposed Workers', *Cancer Epidemiol., Biomarkers Prev.*, **6**, 1039–1042 (1997).
32. Q. Zhou, G. Talaska, M. Jaeger, V.K. Bhatnagar, R.B. Hayes, T.V. Zenser, S.K. Kasyap, V.M. Lakshmi, R. Kashyap, M. Dosemeci, F.F. Hsu, D.J. Parikh, B. Davis, N. Rothman, 'Benzidine–DNA Adduct Levels in Human Peripheral White Blood Cells Significantly Correlate with Levels in Exfoliated Urothelial Cells', *Mut. Res.*, **393**, 199–205 (1997).
33. G.H. Goodfellow, K. Durette, D.M. Grant, 'Structure–Function Studies of Human Arylamine Acetyltransferases: The Role of Amino Acids 125 and 127 in Imparting Arylamine Acceptor Substrate Specificity', *First International Workshop on the Arylamine* N-*Acetyltransferases*, Cairns, Australia, October, 22–24, 1998.
34. N.P. Lang, F.F. Kadlubar, S.L. MacLeod, S. Nowell, J. Massengill, K.E. Anderson, M.D. Gross, 'Influence of NAT-1 Genotype on NAT-2 Phenotype', *First International Workshop on the Arylamine* N-*Acetyltransferases*, Cairns, Australia, October, 22–24, 1998.
35. H. Wikman, P. Piirinlä, K. Kääriä, C. Rosenberg, H. Norppa, A. Hirvonen, 'Role of *NAT1* and *NAT2* Polymorphisms in Individual Susceptibility to Isocyanate-Induced Asthma, *First International Workshop on*

the Arylamine N-*Acetyltransferases*', Cairns, Australia, October, 22–24, 1998.
36. H.W. Goedde, D.P. Agarwal, 'Pharmacogenetics of Aldehyde Dehydrogenase (ALDH)', *Pharmacol. Ther.*, **45**, 345–371 (1990).
37. H. Furuya, P. Fernandez-Salguero, W. Gregory, H. Taber, A. Steward, F.J. Gonzalez, J.F. Idle, 'Genetic Polymorphism of CYP2C9 and its Effect on Warfarin Maintenance Dose Requirement in Patients Undergoing Anticoagulation Therapy', *Pharmacogenetics*, **5**, 389–392 (1995).
38. W. Kalow, K. Genest, 'A Method for the Detection of Atypical Forms of Serum Cholinesterase. Determination of Dibucaine Numbers', *Can. J. Biochem.*, **35**, 339–346 (1957).
39. J.R. Playfer, L.C. Eze, M.F. Bullen, D.A.P. Evans, 'Genetic Polymorphism and Interethnic Variability', *J. Med. Genet.*, **13**, 337–342 (1976).
40. B.N. La Du, S. Adkins, 'Analysis of Serum Paraoxonase/Arylesterase Polymorphism in Some Sudanese Families', in *Ethnic Differences in Reactions to Drugs and Xenobiotics*, eds. W. Kalow, H.W. Goedde, D.P. Agarwal, Alan R. Liss, New York, 87–98, 1986.
41. D.M. Grant, P. Vohra, Y. Avis, A. Ima, 'Detection of a New Polymorphism of Human Arylamine *N*-Acetyltransferase NAT1 Using *p*-Aminosalicylic Acid as an In Vivo Probe', *J. Basic Clin. Physiol Pharmacol.*, **3**(Suppl.), 244 (1992).
42. K.P. Vatsis, W.W. Weber, 'Human *N*-Acetyltransferases', in *Conjugation–Deconjugation Reactions in Drug Metabolism and Toxicity*, ed. F.C. Kauffman, Springer-Verlag, New York, 1994.
43. Z. Lu, R. Zhang, R.B. Diasio, 'Dihydropyrimidine Dehydrogenase Activity in Human Peripheral Blood Mononuclear Cells and Liver: Populations Characteristics, Newly Identified Deficient Patients and Clinical Application in 5-Fluorouracil Chemotherapy', *Cancer Res.*, **53**, 5433–5438 (1993).
44. J. Seidegard, R.W. Pero, 'The Hereditary Transmission of High Glutathione Transferase Activity towards *trans*-Stilbene Oxide in Human Mononuclear Leukocytes', *Human Genet.*, **69**, 66–68 (1985).
45. J.K. Wiencke, K.T. Kelsey, R.A. Lamela, W.A. Toscano, Jr, 'Human Glutathione-*S*-transferase Deficiency as a Marker of Susceptibility to Epoxide-induced Damage', *Cancer Res.*, **50**, 1585–1590 (1990).
46. R. Weinshilboum, 'Methyltransferase Pharmacogenetics', *Pharmacol. Ther.*, **43**, 77–90 (1989).
47. C.L. Szumlanski, R. Honcel, M.C. Scott, R.M. Weinshilboum, 'Human Liver Thiopurine Methyltransferase Pharmacogenetics: Biochemical Properties, Liver–Erythrocyte Correlation and Presence of Isozymes', *Pharmacogenetics*, **2**, 148–159 (1992).
48. G. Monaghan, M. Ryan, R. Seddon, R. Hume, B. Burchell, 'Genetic Variation in Bilirubin UDP-gluconosyltransferase Gene Promoter and Gilbert's Syndrome', *Lancet*, **347**, 578–581 (1996).
49. P.J. Bosma, J.R. Chowdhury, C. Bakker, S. Gantla, A. de Boer, B.A. Oostra, D. Lindhout, G.N.J. Guido, P.M. Jansen, L.M. Peter, O. Elferink, P.J. Ronald, N.R. Chowdhury, 'The Genetic Basis of the Reduced Expression of Bilirubin UDP-glucurosyltransferase 1 in Gilbert's Syndrome', *N. Engl. J. Med.*, **333**, 1171–1175 (1995).
50. M. Patel, B.K. Tang, W. Kalow, '(*S*)Oxazepam is Inhibited by Ketoprofen and Other Substrates of UGT2B7', *Pharmacogenetics*, **5**, 43–49 (1995).
51. B.L. Coffman, G.R. Rios, C.D. King, T.R. Tephly, 'Human UGT2B7 Catalyzes Morphine Glucuronidation', *Drug Metab. Dispos.*, **25**, 1–4 (1997).
52. O. Gotoh, Y. Fujii-Kuriyama, 'Evolution, Structure, and Gene Regulation of Cytochrome P-450', in *Basis and Mechanisms of Regulation of Cytochrome P-450*, eds. K. Ruckpaul, H. Rein, Taylor and Francis, London, 195–243, 1989.
53. I. Johansson, E. Lundqvist, L. Bertilsson, M.-L. Dahl, F. Sjoqvist, M. Ingelman-Sundberg, 'Inherited Amplification of an Active Gene in the Cytochrome P450 CYP2D Locus as a Cause of Ultrarapid Metabolism of Debrisoquine', *Proc. Natl. Acad. Sci. USA*, **90**, 11825–11829 (1993).
54. G.K.K. Hershey, M.F. Friedrich, L.A. Esswein, M.L. Thomas, T.A. Chatila, 'The Association of Atopy with a Gain-of-function Mutation in the α-Subunit of the Interleukin-4 Receptor', *N. Engl. J. Med.*, **337**, 1720–1725 (1997).
55. W.W. Weber, *The Acetylator Genes and Drug Response*, Oxford University Press, New York, 1987.
56. G.N. Levy, W.W. Weber, 'Interindividual Variability of *N*-Acetyltransferases in Man', in *Interindividual Variability in Drug Metabolism in Man*, eds. G. Pacifici, O. Pelkonen, Taylor and Francis, London, in press.
57. D.G. McCarver, R. Byun, R.N. Hines, M. Hicheme, W. Wegenek, 'A Genetic Polymorphism in the Regulatory Sequences on Human *CYP2E1*: Association with Increased Chloroxazone Hydroxylation in the Presence of Obesity and Ethanol Intake', *Toxicol. Appl. Pharmacol.*, **152**, 276–281 (1998).
58. E.M. Southern, 'Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis', *J. Mol. Biol.*, **98**(3), 503–517 (1975).
59. J.C. Alwine, D.J. Kemp, G.R. Stark, 'Method for Detection of Specific RNAs in Agarose Gels by Transfer to Diazobenzyloxymethyl-paper and Hybridization with DNA Probes', *Proc. Natl. Acad. Sci. USA*, **74**(12), 5350–5354 (1977).
60. A.R. Wyman, R. White, 'A Highly Polymorphic Locus in Human DNA', *Proc. Natl. Acad. Sci. USA*, **77**(11), 6754–6758 (1980).
61. A.J. Jeffreys, V. Wilson, S.L. Thein, 'Individual-specific Fingerprints of Human DNA', *Nature*, **316**, 76–79 (1985).

62. A.J. Jeffreys, V. Wilson, S.L. Thein, 'Hypervariable Minisatellite Regions in Human DNA', *Biotechnology*, **24**, 467–472 (1992).
63. Y. Nakamura, M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, 'Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping', *Science*, **235**, 1616–1622 (1987).
64. M.S. Ristaldi, M. Pirastu, C. Rosatelli, G. Monni, H. Erlich, R. Saiki, A. Cao, 'Prenatal Diagnosis of Beta-thalassaemia in Mediterranean Populations by Dot Blot Analysis with DNA Amplification and Allele Specific Oligonucleotide Probes', *Prenat. Diagn*, **9**(9), 629–638 (1989).
65. D.Y. Wu, G. Nozari, M. Schold, B.J. Conner, R.B. Wallace, 'Direct Analysis of Single Nucleotide Variation in Human DNA and RNA Using In Situ Dot Hybridization', *DNA*, **8**(2), 135–142 (1989).
66. E. Kawasaki, R. Saiki, H. Erlich, 'Genetic Analysis Using Polymerase Chain Reaction-amplified DNA and Immobilized Oligonucleotide Probes: Reverse Dot–Blot Typing', *Meth. Enzymol.*, **218**, 369–381 (1993).
67. P. Sutcharitchan, R. Saiki, T.H. Huisman, A. Kutlar, V. McKie, H. Erlich, S.H. Embury, 'Reverse Dot–Blot Detection of the African-American Beta-thalassemia Mutations', *Blood*, **86**(4), 1580–1585 (1995).
68. H. Erlich, T. Bugawan, A.B. Begovich, S. Scharf, R. Griffith, R. Saiki, R. Higuchi, P.S. Walsh, 'HLA-DR, DQ and DP Typing Using PCR Amplification and Immobilized Probes', *Eur. J. Immunogenet.*, **18**(1–2), 33–55 (1991).
69. E.M. Southern, U. Maskos, 'Parallel Synthesis and Analysis of Large Numbers of Related Chemical Compounds: Applications to Oligonucleotides', *J. Biotechnol.*, **35**(2–3), 217–227 (1994).
70. A.C. Pease, D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, S.P. Fodor, 'Light-generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis', *Proc. Natl. Acad. Sci. USA*, **91**(11), 5022–5026 (1994).
71. A.P. Blanchard, R.J. Kaiser, L.E. Hood, 'High Density Oligonucleotide Arrays', *Biosens. Bioelectron.*, **11**(6/7), 687–690 (1996).
72. M. Schena, D. Shalon, R.W. Davis, P.O. Brown, 'Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray' [see comments]. *Science*, **270**, 467–470 (1995).
73. V.G. Cheung, M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati, G. Childs, 'Making and Reading Microarrays', *Nat. Genet.*, **21**(Suppl.), 15–19 (1999).
74. D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J. Trent, 'Expression Profiling Using cDNA Microarrays', *Nat. Genet.*, **21**(Suppl.), 10–14 (1999).
75. D.G. Wang, J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, 'Large-scale Identification, Mapping and Genotyping of Single-nucleotide Polymorphisms in the Human Genome', *Science*, **280**, 1077–1082 (1998).
76. D.J. Lockhart, H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, 'Expression Monitoring by Hybridization to High-density Oligonucleotide Arrays', *Nat. Biotechnol*, **14**(13), 1675–1680 (1996).
77. J.G. Hacia, L.C. Brody, M.S. Chee, S.P. Fodor, F.S. Collins, 'Detection of Heterozygous Mutations in *BRCA1* Using High Density Oligonucleotide Arrays and Two-color Fluorescence Analysis', *Nat. Genet.*, **14**(4), 441–447 (1996).
78. J. Hacia, 'Resequencing and Mutational Analysis Using Oligonucleotide Microarrays', *Nat. Genet.*, **21**(Suppl.), 42–47 (1999).
79. M.T. Cronin, R.V. Fucini, S.M. Kim, R.S. Masino, R.M. Wespi, C.G. Miyada, 'Cystic Fibrosis Mutation Detection by Hybridization to Light-generated DNA Probe Arrays', *Hum. Mutat.*, **7**(3), 244–255 (1996).
80. V. Chan, D.J. Graves, S.E. McKenzie, 'The Biophysics of DNA Hybridization with Immobilized Oligonucleotide Probes', *Biophys J.*, **69**(6), 2243–2255 (1995).
81. L. Wodicka, H. Dong, M. Mittmann, M.H. Ho, D.J. Lockhart, 'Genome-wide Expression Monitoring in *Saccharomyces cerevisiae*', *Nat. Biotechnol.*, **15**(13), 1359–1367 (1997).
82. D. Shalon, S.J. Smith, P.O. Brown, 'A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization', *Genome Res.*, **6**(7), 639–645 (1996).
83. J.G. Hacia, S.A. Woski, J. Fidanza, K. Edgemon, N. Hunt, G. McGall, S.P. Fodor, F.S. Collins, 'Enhanced High Density Oligonucleotide Array-based Sequence Analysis Using Modified Nucleoside Triphosphates', *Nucl. Acids Res.*, **26**(21), 4975–4982 (1998).
84. U. Maskos, E.M. Southern, 'Oligonucleotide Hybridizations on Glass Supports: A Novel Linker for Oligonucleotide Synthesis and Hybridization Properties of Oligonucleotides Synthesised In Situ', *Nucl. Acids Res.*, **20**(7), 1679–1684 (1992).
85. W.I. Wood, J. Gitschier, L.A. Lasky, R.M. Lawn, 'Base Composition-independent Hybridization in Tetramethylammonium Chloride: A Method for Oligonucleotide Screening of Highly Complex Gene Libraries', *Proc. Natl. Acad. Sci. USA*, **82**(6), 1585–1588 (1985).
86. T.T. Nikiforov, R.B. Rendle, M.L. Kotewicz, Y.H. Rogers, 'The Use of Phosphorothioate Primers and Exonuclease Hydrolysis for the Preparation of Single-stranded PCR Products and Their Detection by Solid-phase Hybridization', *PCR Meth. Appl.*, **3**(5), 285–291 (1994).
87. T.T. Nikiforov, R.B. Rendle, P. Goelet, Y.H. Rogers, M.L. Kotewicz, S. Anderson, G.L. Trainor, M.R. Knapp, 'Genetic Bit Analysis: A Solid Phase Method for Typing Single Nucleotide Polymorphisms', *Nucl. Acids Res.*, **22**(20), 4167–4175 (1994).

88. M. Chee, R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, S.P. Fodor, 'Accessing Genetic Information with High-density DNA Arrays', *Science*, **274**, 610–614 (1996).
89. C.F. Edman, D.E. Raymond, D.J. Wu, E. Tu, R.G. Sosnowski, W.F. Butler, M. Nerenberg, M.J. Heller, 'Electric Field Directed Nucleic Acid Hybridization on Microchips', *Nucl. Acids Res.*, **25**(24), 4907–4914 (1997).
90. R.G. Sosnowski, E. Tu, W.F. Butler, J.P. O'Connell, M.J. Heller, 'Rapid Determination of Single Base Mismatch Mutations in DNA Hybrids by Direct Electric Field Control', *Proc. Natl. Acad Sci. USA*, **94**(4), 1119–1123 (1997).
91. J. Cheng, E.L. Sheldon, L. Wu, A. Uribe, L.O. Gerrue, J. Carrino, M.J. Heller, J.P. O'Connell, 'Preparation and Hybridization Analysis of DNA/RNA from *E. coli* on Microfabricated Bioelectronic Chips', *Nat. Biotechnol.*, **16**(6), 541–546 (1998).
92. H. Koster, K. Tang, D.J. Fu, A. Braun, D. van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, 'A Strategy for Rapid and Efficient DNA Sequencing by Mass Spectrometry', *Nat. Biotechnol.*, **14**(9), 1123–1128 (1996).
93. D.P. Little, A. Braun, M.J. O'Donnell, H. Koster, 'Mass Spectrometry from Miniaturized Arrays for Full Comparative DNA Analysis', *Nat. Med.*, **3**(12), 1413–1416 (1997).
94. J.A. Ferguson, T.C. Boles, C.P. Adams, D.R. Walt, 'A Fiber-optic DNA Biosensor Microarray for the Analysis of Gene Expression', *Nat. Biotechnol.*, **14**(13), 1681–1684 (1996).
95. K.L. Michael, L.C. Taylor, S.L. Schultz, D.R. Walt, 'Randomly Ordered Addressable High-density Optical Sensor Arrays', *Anal. Chem.*, **70**(7), 1242–1248 (1998).
96. T.M. Cox, 'Aldolase B and Fructose Intolerance', *FASEB J.*, **8**, 62–71 (1994).
97. T.M. Cox, 'Iatrogenic Deaths in Hereditary Fructose Intolerance', *Arch. Dis. Child.*, **69**, 413–415 (1993).
98. M. Ali, T.M. Cox, 'Diverse Mutations in the Aldolase B Gene that Underlie the Prevalence of Hereditary Fructose Intolerance', *Am. J. Hum. Genet.*, **56**, 1002–1005 (1995).
99. R. Ayesh, R.L. Smith, 'Genetic Polymorphism of Trimethylamine *N*-Oxidation', *Pharmacol Ther.*, **45**, 387–401 (1990).
100. C.T. Dolphin, A. Janmohamed, R.L. Smith, E.A. Shephard, I.R. Phillips, 'Missense Mutation in Flavin-containing Mono-oxygenase 3 Gene, *FMO3*, Underlies Fish-odour Syndrome', *Nat. Genet.*, **17**, 491–494 (1997).
101. W. Burke, E. Thomson, M.J. Khoury, S.M. McDonnell, N. Press, P.C. Adams, J.C. Barton, E. Beutler, G. Brittenham, A. Buchanan, E.W. Clayton, M.E. Cogswell, E.M. Meslin, A.G. Motulsky, L.W. Powell, E. Sigal, B.S. Wilford, F.S. Collins, 'Hereditary Hemochromatosis: Gene Discovery and its Implications for Population-based Screening', *J. Am. Med. Assoc.*, **280**, 172–178 (1998).
102. F.S. Collins, 'Cystic Fibrosis: Molecular Biology and Therapeutic Implications', *Science*, **256**, 774–779 (1992).
103. W.W. Weber, 'Populations and Genetic Polymorphisms', *Mol. Diagn.*, **4**(4), 299–307 (1999).
104. D. Marez, M. Legrand, N. Sabbagh, J.-M. Lo Guidice, J.-M. Spire, J.-J. Lafitte, U. Meyer, F. Broly, 'Polymorphism of the Cytochrome P450 CYP2D6 Gene in an European Population: Characterization of 48 Mutations and 53 Alleles, their Frequencies and Evolution', *Pharmacogenetics*, **7**, 193–202 (1996).
105. C. Sachse, J. Brockmöller, S. Bauer, I. Roots, 'Cytochrome P450 2D6 Variants in a Caucasian Population: Allele Frequencies and Phenotypic Consequences', *Am. J. Hum. Genet.*, **60**, 284–295 (1997).
106. M. Ingelman-Sundberg, 'The Gerhard Zbinden Memorial Lecture. Genetic Polymorphism of Drug Metabolizing Enzymes. Implications for Toxicity of Drugs and Other Xenobiotics', *Arch. Toxicol. Suppl.*, **19**, 3–13 (1997).
107. G.R. Wilkinson, F.P. Guengerich, R.A. Branch, 'Genetic Polymorphism of *S*-Mephenytoin Hydroxylation', *Pharmacol. Ther.*, **43**, 53–76 (1989).
108. G. Ibeanu, J. Blaisdell, B.I. Ghanayem, C. Beyeler, S. Benhamou, C. Bouchardy, G.R. Wilkinson, P. Dayer, A.K. Daly, J.A. Goldstein, 'An Additional Defective Allele, *CYP2C19**5*, Contributes to the *S*-Mephenytoin Poor Metabolizer Phenotype in Caucasians', *Pharmacogenetics*, **8**, 129–135 (1998).
109. F.S. Collins, 'Shattuck Lecture – Medical and Societal Consequences of the Human Genome Project', *N. Engl. J. Med.*, **341**, 28–37 (1999).
110. S. Sindrup, K. Brosen, 'The Pharmacogenetics of Codeine Hypoalgesia', *Pharmacogenetics*, **5**, 335–346 (1995).
111. R.F. Tyndale, K.P. Droll, E.M. Sellers, 'Genetically Deficient CYP2D6 Metabolism Provides Protection Against Oral Opiate Dependence', *Pharmacogenetics*, **7**, 375–379 (1997).
112. D. Wu, V. Otton, T. Inaba, W. Kalow, E. Sellers, 'Interactions of Amphetamine Analogs with Human Liver CYP2D6', *Biochem. Pharmacol.*, **53**, 1605–1612 (1997).
113. P. Dalen, C. Fregnell, M.-L. Dahl, F. Sjoqvist, 'Quick Onset of Severe Abdominal Pain after Codeine in an Ultrarapid Metabolizer of Debrisoquine', *Ther. Drug Monit.*, **19**, 543–544 (1997).

Phospholipids of Plasma Lipoproteins, Red Blood Cells and Atheroma, Analysis of

Amir Ravandi and Arnis Kuksis
University of Toronto, Toronto, Canada

1 Introduction **1**
2 General Structure and Nomenclature **1**
2.1 Glycerophospholipids and Sphingomyelins 2
2.2 Phosphatidylinositol Phosphates 3
2.3 Other Phospholipids 4
3 Isolation **4**
3.1 Liquid-phase Extraction 5
3.2 Solid-phase Extraction 6
3.3 Derivatization and Preliminary Segregation 8
4 Resolution of Common Phospholipids **11**
4.1 Principles of Methodology 11
4.2 Chemical Classes 13
4.3 Molecular Species 18
5 Novel Phospholipid Subclasses **26**
5.1 Glycated Aminophospholipids 26
5.2 Oxidized Glycerophospholipids 31
5.3 Phosphoinositide Glycans 32
6 Conclusions **34**
Acknowledgments **34**
Abbreviations and Acronyms **34**
Related Articles **35**
References **35**

The large number of lipid classes and great complexity of molecular species present in blood plasma, red cells, platelets and atheromatous lesions requires a combination of analytical techniques for comprehensive analyses, including chemical and enzymatic derivatization of samples. In many instances only partial analyses are required, which can be accomplished by specific analytical techniques. This chapter describes the full spectrum of the methodology ranging from the most basic thin-layer chromatography (TLC) to the detailed mass spectrometric assays. Traditional sample extraction by liquid–liquid partition is time-consuming and involves large volumes of solvents. Liquid–solid extraction using adsorbent cartridges is more economical. At the present time total lipid extracts can be effectively assayed for lipid class content and molecular species composition by flow injection tandem mass spectrometry (MS/MS), while a more detailed analysis of complex lipid mixtures is provided by a combination of liquid chromatography with on-line electrospray mass spectrometry (LC/ESMS). The new techniques permit completion of the analyses in a few hours, where previously several days or weeks may have been required. The soft ionization mass spectrometric techniques have permitted the recognition and detailed analysis of such minor components of blood as the lipid oxidation and glycation products often observed in disease.

1 INTRODUCTION

Phospholipids occur in blood largely in cell membranes and lipoproteins where they provide structural components and serve as reservoirs of signaling molecules. The plasma lipoproteins, red cells and platelets contain complex mixtures of glycerophospholipids and sphingomyelins (SMs). The choline (Cho) and ethanolamine (Etn) glycerophospholipids and SMs constitute the major and the inositol and serine glycerophospholipids the minor components of the blood system. Each glycerophospholipid class consists of molecular species made up of pairs of fatty acids (FAs) of different chain length and degree of unsaturation. In SMs the nitrogenous base, sphingosine is paired with FAs of differing chain length and degree of unsaturation. In the past the molecular species composition of blood glycerophospholipids and SMs has been extensively studied as an analytical challenge. More recently, the recognition of the special role that these molecules play in metabolic regulation provided a more legitimate basis for such efforts. With detailed analyses permitted by new instrumentation have come discoveries of new phospholipid subclasses and molecular species. In the following both new (LC/ESMS) and old (TLC and gas chromatography) analytical methods are reviewed as an aid to modern studies of the role of glycerophospholipids and SMs in cellular signaling and metabolic regulation.

2 GENERAL STRUCTURE AND NOMENCLATURE

The phospholipids are members of a large class of lipids which are characterized by the fact that on complete hydrolysis they all yield, among other products, inorganic phosphate (P_i). The large variety of phospholipid classes and FAs associated with them has led to the introduction of abbreviated nomenclature and short-hand

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

designation in order to simplify discussion and promote comprehension.[1]

2.1 Glycerophospholipids and Sphingomyelins

The simplest phospholipid is phosphatidic acid (PtdOH, **(1)**), which on hydrolysis yields one equivalent each of glycerol (Gro) and of phosphoric acid (P) and two equivalents of FA, one of which is saturated, the other unsaturated. The P is bound to Gro in ester linkage at the primary *sn*-3-hydroxyl group. In the phosphatides, the P is also bound in ester linkage to one of the following polar head groups: choline (PtdCho, **(2)**), ethanolamine (PtdEtn, **(3)**), L-serine (PtdSer, **(4)**) and inositol (PtdIns, **(5)**). The plasmenyl subclass of phosphatides yields on mild acid hydrolysis one mole each of aliphatic aldehyde, FA, Gro, P and a nitrogen-containing base (Etn or Cho). The aldehyde formed on mild acid hydrolysis of plasmalogens is derived from an aliphatic chain linked to Gro in a vinyl ether linkage. The plasmanyl subclass of the phosphatides on transmethylation yields one mole each of alkylglycerol, FA, Gro and P and a nitrogen-containing base (Etn or Cho). The Etn moiety of PtdEtn may contain a FA in an amide linkage, while the *myo*-inositol residue of the PtdIns glycans may contain a palmitic acid in an ester linkage. The total lipid extracts of platelets may also include phosphatidylglycerol (PtdGro) and cardiolipin **(6)** or diphosphatidylglycerol (Ptd_2Gro) which are characteristic of mitochondria. Each lipid class is made up of numerous molecular species, which contain fatty chains differing in length and degree of unsaturation. The unsaturated and especially the polyunsaturated species may become peroxidized to variable extent depending on the oxidative stress of the tissue and the endogenous protective mechanisms (see below).

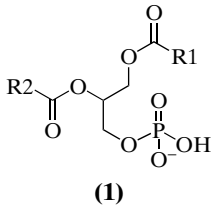

(1)

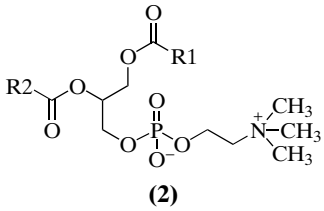

(2)

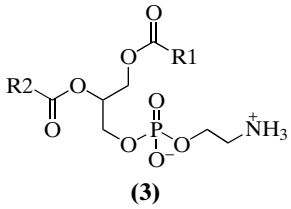

(3)

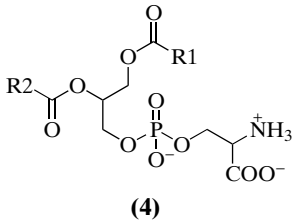

(4)

The glycerophospholipids are subject to hydrolysis by phospholipases (Figure 1). Phospholipase A_2 releases the FA (usually unsaturated) from the *sn*-2-position, while phospholipase A_1 releases the FA (usually saturated) from the *sn*-1-position. Phospholipase C releases the *sn*-1,2-diacylglycerol and phospholipase D the PtdOH moiety of the glycerophospholipids. Many of the phospholipases exhibit specificity for the polar head group of the glycerophospholipid.[2]

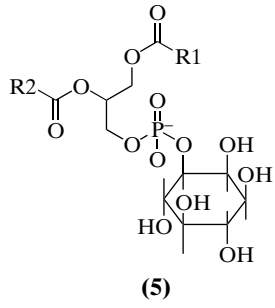

(5)

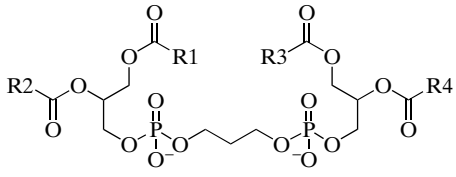

(6)

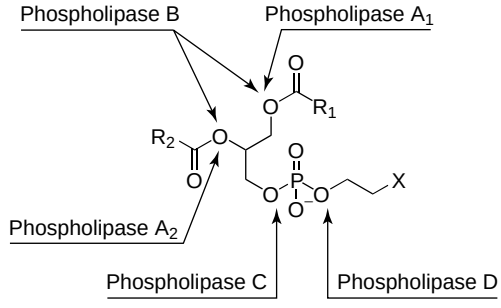

Figure 1 Sites of enzymatic cleavage in a glycerophospholipid. X is Cho, Etn, serine, or inositol headgroup.

The SMs (*N*-acylsphingosine-1-phosphocholines) **(7)** constitute the second largest phospholipid class in plasma lipoproteins and red blood cells. SM contains the amino alcohol sphingosine **(8)** and its homologs. The sphingosine base is esterified to phosphoryl Cho and is involved in an amide linkage to saturated FAs, although some unsaturated species are also present. SMs are relatively resistant to alkaline hydrolysis, which may be taken advantage of during isolation of ceramides **(9)** and SM.

SM is subject to hydrolysis by phospholipase C or sphingomyelinase, which releases the ceramide moiety. Their characterization is usually performed by hydrolysis and separate identification of the FA and long chain base components,[3] but intact ceramides can also be readily determined by gas–liquid chromatography (GLC)[4]

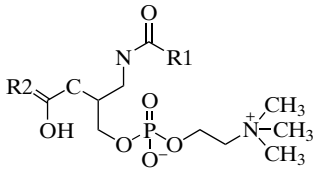

(7)

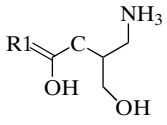

(8)

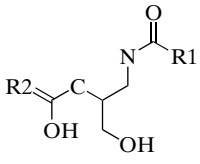

(9)

or high-performance liquid chromatography (HPLC),[5] both techniques permitting on-line combination with mass spectrometry (MS) (see below).

2.2 Phosphatidylinositol Phosphates

The *myo*-inositol residue of PtdIns may contain one to four P groups, which may occur as single or multiple isomers. The numbering of inositol phosphate isomers has been discussed and summarized by Parthasarathy and Eisenberg.[6] Structures **(10–14)** are the structures of the PtdIns phosphates and show the abbreviations recommended by International Union of Biochemistry/International Union of Pure and Applied Chemistry.[1] The structures of the PtdIns phosphates are commonly characterized by chromatography of the intact or deacylated molecules using various chromatographic systems and appropriate reference standards. The distribution of the phosphate groups between the fourth and fifth positions of the inositol ring may be determined by the method of Hawkins et al.,[7] which depends on peroxidate oxidation under conditions which restrict oxidation of the Gro moiety.[6,8] Alternatively, the PtdIns phosphates are characterized on the basis of their susceptibility to specific phosphatases and phospholipases using radioactive substrates. Figure 2 indicates the transformations of the polyphosphoinositides under the influence of the phosphatases and kinases.[9] Thus, PtdIns 3-kinase phosphorylates the D3-position of the inositol ring of PtdIns to produce a PtdIns (3)P that is distinct from PtdIns (4)P, the predominant monophosphate from cellular PtdIns.[6] In addition, the PtdIns 3-kinase also phosphorylates PtdIns (4)P and PtdIns (4,5)P_2 to generate two additional phospholipids, PtdIns (3,4)P_2 and PtdIns (3,4,5)P_3, respectively.[10] Four immunologically distinct forms of PtdIns-specific phospholipase C have been identified.[9] All isozymes require Ca^{2+} and degrade PtdIns, phosphatidylinositol phosphate (PtdIns P), and

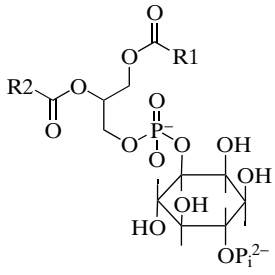

PtdIns (4)P

(10)

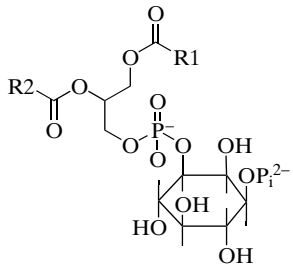

PtdIns (3)P

(11)

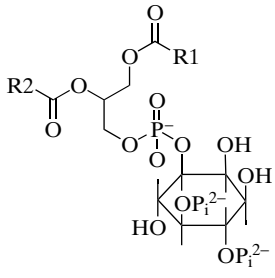

PtdIns (4,5)P_2

(12)

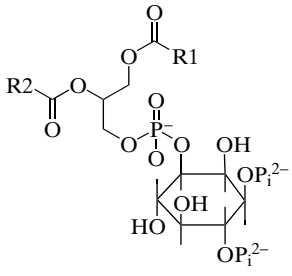

PtdIns (3,4)P_2

(13)

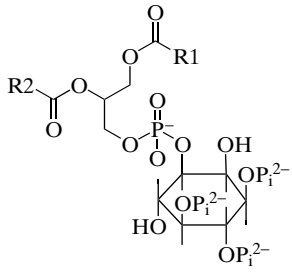

PtdIns (3,4,5)P_3

(14)

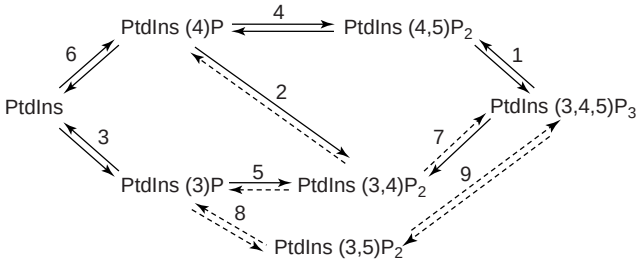

Figure 2 Phosphatase and kinase-mediated pathways of polyphosphoinositide transformation.[9] The kinase pathways catalyzed by known enzymes are as follows: 1–3, PtdIns 3-K; 4, PtdIns (4)P 5-K; 5, PtdIns (3)P 4-K; and 6, PtdIns 4-K. The kinase pathways not yet proven are as follows: 7, PtdIns (3,4)P_2 5-K; 8, PtdIns (3)P 5-K; and 9, PtdIns (3,5)P_2 4-K. The solid lines indicate known enzymatic activities. The dotted lines indicate hypothetical pathways.

PtdIns P_2, but the type of lipid degraded depends on the reaction conditions.

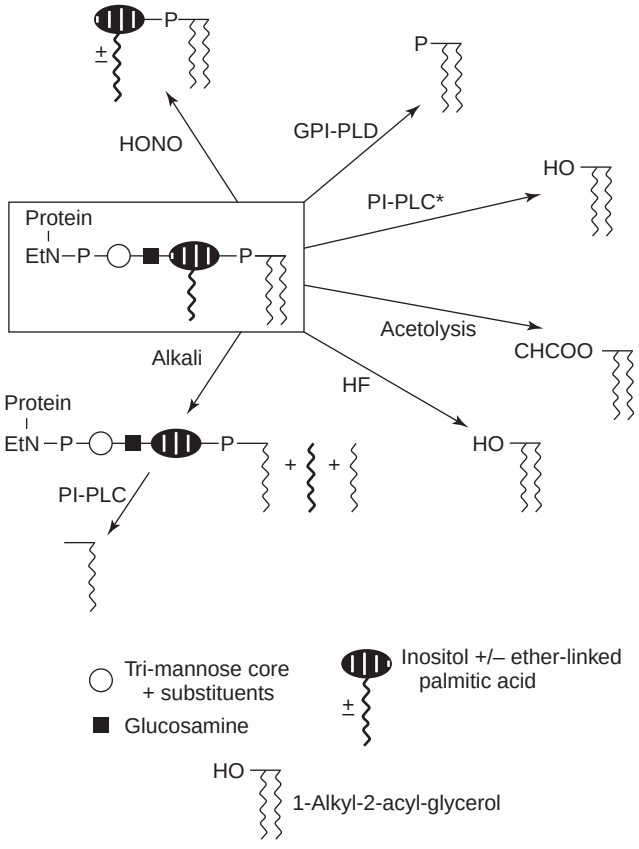

Figure 3 Structure of glycosyl PtdIns and production of PtdIns lipid fragments: cleavages for a PtdIns-anchored protein containing inositol + hydroxyester-linked palmitic acid, and 1-alkyl-2-acyl-glycerol; only lipid products are shown.[13] Phospholipid fragments may be generated by nitrous acid (HONO) treatment, or by cleavage with GroPIns-specific phospholipase D (GPI-PLD). AcylGro may be generated by enzymatic treatment, using bacterial PtdIns-specific phospholipases C (PI-PLC*, the asterisk indicating that PI-PLC will cleave only GPI structures that are not inositol acylated), or by chemical treatment via acetolysis or aqueous HF dephosphorylation. Alkali treatment removes the ester-linked FAs (including the inositol-linked FA if present), and PI-PLC treatment of the resulting lyso-GPI anchor generates 1-alkylglycerol. AcylGro fragments and free FAs are extracted into hexane, toluene, or chloroform–methanol (2:1, v/v). Phospholipid fragments are extracted into water-saturated 1-butanol.

2.3 Other Phospholipids

Glycated PtdIns constitute a diverse family of lipid molecules found in prokaryotes and eukaryotes. The members of this family are characterized by the sequence Etn-PO_4-Manα1-4GlcNα1-6-*myo*-inositol-1-PO_4-diradylglycerol.[11] These structures are involved in anchoring membrane proteins including the acetylcholine esterase of the human red blood cells. The complete structure of the PtdIns anchor of human red blood cell acetylcholine esterase has been determined and shown to possess a C-terminus of the protein linked via Etn phosphate to a glycan with the conserved backbone sequence Manα 1-2Manα 1-6Manα 1-4GlcNH_2, which in turn is linked to the 6-position of the *myo*-inositol ring of PtdIns.[12] The presence of a palmitate residue on one of the inositol hydroxyls renders the anchor resistant to PtdIns-specific phospholipase C hydrolysis. Figure 3 gives the structure of the glycosyl PtdIns anchors along with the chemical and enzymic cleavage points.[13]

3 ISOLATION

The method of lipid extraction and subsequent work-up of the extract are critical to the outcome of the analysis and the reliability of the data. Both must be executed with care. The blood from volunteers is taken into sterile donation bags containing cytidine diphosphate adenine as preservative (115 mM citrate–NaOH, pH 7.3, 160 mM glucose, 21 mM NaH_2PO_4, 2 mM adenine) (Travenol Labs, Thetford, Norfolk, UK). The analysis of blood lipids involves several discrete steps, some of which can be combined depending on the aims of the analysis. The work-up usually involves a separation of plasma and cells, which is accomplished by centrifugation at 500 *g* for 10 min and washing the cells to remove the buffy coat. The erythrocytes are washed four times in ice-cold 154 mM NaCl–1.5 mM Hepes, pH 7.2, the buffy coat and supernatant being also collected. Residual leukocytes are removed by passing the washed cells through a sterile cotton wool column (Cellselect Leukocyte filter).

Platelets are isolated by collecting the buffy coats and recentrifuging them in buffer or saline.[14] However, Low[15] recommends that prostaglandin E_1 (PGE_1) (2.8 μM) is added to the buffer solution since it prevents platelet activation and increases the yield of [^{32}P]polyphosphoinositide. The platelets are prepared from fresh human blood (60 mL) and suspended in 3 mL of 140 mM NaCl, 5 mM KCl, 0.005 mM $CaCl_2$, 0.1 mM $MgCl_2$, 16.5 mM glucose, 0.1 mg serum albumin mL^{-1}, 15 mM Hepes (pH 7.4), containing 2–6 mCi [^{32}P]P_i (carrier free) for 2 h at 37 °C with occasional gentle shaking.

Prior to lipid extraction the plasma membranes of the red cells can be isolated as the normal or inverted ghosts.[16] Furthermore, differentiation can be made between the glycerophospholipids associated with the inner and outer half of the lipid bilayer of the red cell plasma membrane. This is usually accomplished by differential chemical or enzymatic derivatization.[16] In addition to the transbilayer asymmetry of a phospholipid class, some studies have indicated a nonrandom lateral distribution of phospholipid subclasses. The plasma samples are resolved into the major lipoprotein classes by ultracentrifugation using any one of several methods.[17]

3.1 Liquid-phase Extraction

The lipids of plasma and plasma lipoproteins are usually obtained by extraction with chloroform–methanol, while the lipids of the red cells and platelets may be better isolated by chloroform–isopropanol. Complete extraction of the minor acidic phospholipids from all sources requires the use of acidic solvents. Most tissues require special attention to the physicochemical and biochemical properties of the components to be isolated, as described below.

3.1.1 Neutral Solvents

A conventional chloroform–methanol extraction as recommended by Folch et al.[18] or Bligh and Dyer[19] commonly contains the neutral lipids and weakly acidic phospholipids, which make up the bulk of the lipid phase of plasma lipoproteins.[20] These extracts obtained in the absence of acidic solvents, also contain the alkylacyl and the acid labile alkenylacyl glycerophosphocholine (GroPCho) and glycerophosphoethanolamine (GroPEtn). Such extracts have been used for normal phase liquid chromatography/mass spectrometry (LC/MS) with on-line electrospray.[21–23] For complete extraction of the acidic lipids a two-step procedure is necessary (see section 3.1.2).

The extraction of lipids from red cells and platelets can also be performed with the chloroform–methanol routines, especially if the cell membranes are prepared first. However, these methods yield pigmented lipid extracts when applied to whole red cells. Rose and Oaklander[24] have suggested the use of chloroform–isopropanol (7 : 11, v/v) for the extraction of human erythrocytes. This method minimizes the extraction of pigments and has been widely applied. Nakamura et al.[25] have recently provided details for an improved protocol. Briefly, distilled water (1 mL) is added to the red cell pellets and incubated for 15 min, then propan-2-ol (11 mL) is added and vortex mixed thoroughly. The mixture is incubated for 1 h with frequent vortex mixing, which is followed by the addition of chloroform (7 mL). After further vortexing, the mixture is allowed to stand for 1 h, then centrifuged (500 *g*) for 30 min and the supernatant evaporated to dryness under nitrogen. A similar routine may be applied to platelet lipid extraction, but acidic phospholipids and PtdIns phosphates are not adequately recovered without acidification (see section 3.1.2).

Total lipid extracts from atherosclerotic plaques are obtained by Folch extraction of the homogenate of the total tissue.[26,27] However, acidification increases the recovery of the acidic and lysophospholipids (see section 3.1.2). Extraction of largely neutral atheroma lipids has been obtained by Suarna et al.[28] with a mixture of cold methanol (2 mL) and hexane (10 mL) in the presence of lipid-soluble external standards and butylated hydroxytoluene (BHT, 100 μmol L^{-1}) and ethylenediamine tetraacetic acid (EDTA, 1.0 mmol L^{-1}). On the basis of careful investigation, a 5-min homogenization period under these conditions was found to be optimal for the efficient extraction of neutral lipids without substantial destruction of antioxidants.

The neutral phospholipid extracts also recover the glycated PtdEtn and glycated PtdSer,[22] which may be stabilized by reduction with sodium borohydride, and the peroxides. The hydroperoxides may be reduced to the hydroxides by means of cyanoborohydride before further analysis (see section 3.1.2).

3.1.2 Acidic Solvents

A complete extraction of the acidic phospholipids, e.g. lysophospholipids, platelet activating factor (PAF) and other minor acidic phospholipids, PtdSer, PtdIns, and PtdIns phosphates, requires acidification of the extraction solvents. The more strongly acidic phospholipids are isolated by re-extracting the aqueous phase with two volumes of chloroform–methanol–12 M HCl (2 : 1 : 0.012, v/v/v) after an initial removal of the neutral and less acidic phospholipids by chloroform–methanol (2 : 1, v/v) as described above or by diethylamino ethyl (DEAE)-Sephadex columns.[29,30] The organic layer obtained from the acidic chloroform–methanol extract is subsequently neutralized with one drop of 4 M ammonium hydroxide and combined with the first organic extract. For isolation of PtdIns phosphates,[8] samples of an erythrocyte suspension (5 mL, containing 1 mL of packed cells) are lysed in ice-cold 20 mM Tris–HCl–2 mM EDTA, pH 7.2 (70 mL; lysis buffer). This procedure preserves the concentrations of PtdIns-Ps in vivo.[8,31] The membranes are sedimented (15 000 *g* × 20 min) and washed in lysis buffer until white (usually four to six washes). Lipids are extracted from the membranes (final volume, 2 mL) by the addition of chloroform–methanol–HCl (20 : 40 : 1, v/v/v; 7.5 mL). After 20 min at room temperature, this solvent mixture is partitioned by addition of 0.1 M HCl (2.5 mL) and chloroform (2.5 mL). The resulting lower phase is taken, dried in vacuo, and the lipids are dissolved in chloroform.

According to Low[15] platelet lipids (platelets prepared from 60 mL of fresh blood) are extracted with 11.25 mL of chloroform–methanol–HCl (50 : 100 : 1.3, v/v/v) for 30 min at 25 °C. Chloroform (3.75 mL) and 0.1 M HCl (3.75 mL) are then added to separate the phases. The lower phase is washed twice with 5 mL of methanol, 4.5 mL of 2 M NaCl, 0.5 mL of 100 mM EDTA–NaOH (pH 7.4) and then evaporated to dryness in a clean tube

with approximately 1.5 μmol of acid-washed methanol-precipitated bovine brain phosphoinositides (material prepared as above) to act as carrier. The extract is finally redissolved in 0.5 mL of solvent A (chloroform–methanol–water, 20 : 9 : 1, v/v/v) and applied at a flow rate of 1 mL min^{-1} to a 4.5 mm × 250 mm amino normal-phase column equilibrated with solvent A. The column is eluted with 5 mL of solvent A, a 2.25 mL linear gradient from 100% solvent A to 100% solvent B (solvent A containing 0.6 M ammonium acetate). Fractions (1 mL) are collected and radioactivity determined by liquid scintillation counting. Average recovery of applied radioactivity is approximately 70% (see below).

The PtdIns anchors of proteins in neutral detergents, such as Triton X-100 or Nonidet P-40, can be recovered by exhaustive proteolysis.[32] After vortexing and centrifugation the toluene phases are removed, leaving behind the interface. The PtdIns peptides remain in the aqueous phase and can be recovered by rotary evaporation to remove residual solvents and freeze-dried to remove residual ammonium acetate. The PtdIns anchors can be released by enzymatic or chemical degradation. The washed extracts containing the free PtdIns glycan are purified by TLC on Si-60 plates (Merck, Darmstadt, Germany) using either chloroform–methanol–water (4 : 4 : 1, v/v/v; solvent A) or chloroform–methanol–concentrated acetic acid–water (25 : 15 : 4 : 2, v/v/v/v; solvent B). The metabolically labelled glycolipids (HPLC purified or within the glycolipid mixture) are identified as PtdIns by specific enzymatic and chemical treatments including PtdIns-specific phospholipase, PtdIns-specific phospholipase D and HNO_2 deamination, as described elsewhere.[33]

3.2 Solid-phase Extraction

Solid-phase extraction (SPE) may be carried out using normal- and reversed-phase adsorbents, affinity and ion-exchange columns. All of these methods have been employed with various degrees of success for the isolation of individual or mixed lipid classes. However, only a few actual applications have been reported to plasma or blood cell lipids.[34]

3.2.1 Normal-phase Columns

Hamilton and Comai[35] have described a method for the isolation and separation of neutral lipid, free FA and polar lipid classes using small (600 mg) prepacked silica Sep-Pak cartridges. Serum lipids from Bligh and Dyer extraction equivalent to 100 μL of human serum are evaporated to dryness under nitrogen, dissolved in 2.0 mL of hexane–methyl *tert*-butyl ether (MTBE), 200 : 3 and applied to the Sep-Pak column. Combinations of hexane and MTBE are used to progressively elute cholesteryl ester first then triglyceride from the column. After column acidification, FAs are eluted followed by cholesterol. Polar lipids are eluted from the column using combinations of MTBE, methanol and ammonium acetate. PtdEtn and PtdIns eluted together, whereas PtdCho, SM and LysoPtdCho are eluted as a second fraction. Recoveries of each phospholipid are greater than 98%.

Janero and Burghardt[34] have described an optimized procedure for selectively removing blood neutral lipids and noncholine-containing phospholipids from complex lipid mixtures and thereby obtaining a Cho phospholipid fraction markedly enriched in bioactive PAF. From a commercial Sep-Pak cartridge (600 mg silica, Waters, Milford, MA) PAF could be quantitatively eluted along with other Cho phospholipids and free neutral lipid in 10 mL methanol–chloroform–water (2 : 1 : 0.8, v/v/v). With the exception of some PtdSer, negligible noncholine-containing phosphoglycerides remained on the cartridge. The lipids in the PAF-enriched fraction could immediately be re-extracted into chloroform by adding 4 mL water and 3 mL chloroform to the fraction. Figure 4 shows a normal-phase HPLC elution profile obtained for the PAF-enriched fraction using propanol–toluene–acetic acid–water (93 : 110 : 15 : 15, v/v/v/v) as the eluting solvent. A mixture of seven radiolabeled phospholipids (60 000 cpm each) made up of PtdEtn, PtdOH, PtdSer, PtdCho, SM, PAF and lysoPtdCho was subjected to the SPE procedure. The HPLC eluate was collected at 1 min intervals over a 50 min run counted by scintillation spectrometry.

A type of SPE and preliminary segregation of phospholipid classes from a total lipid extract is provided by normal-phase HPLC employed as a method of admission of phospholipids to a mass spectrometer for lipid class and molecular species determination.[22,23,36] A gradient solvent system originally described by Becart et al.[37] has proven to be highly compatible with electrospray ionization (see section 4).

As noted above, none of these methods is effective in separating PtdIns phosphates in pure form from crude tissue extracts. Therefore, the lipid extracts are processed using several chromatographic steps. Singh[38] has described the extraction and analysis of PtdIns, PtdIns (4)P and PtdIns (4,5)P_2 using a Sep-Pak cartridge to provide both mass and radiolabeling. Although this method has not been specifically applied for the isolation of PtdIns phosphates from blood lipids, it is included here because of its apparent potential. Thus, a total lipid extract of rat brain synaptosomes and microvessels is first prepared according to Folch et al.[18] and fraction 1 is collected and further extracted to separate phospholipids and glycolipids as described by Dugan et al.[39] The

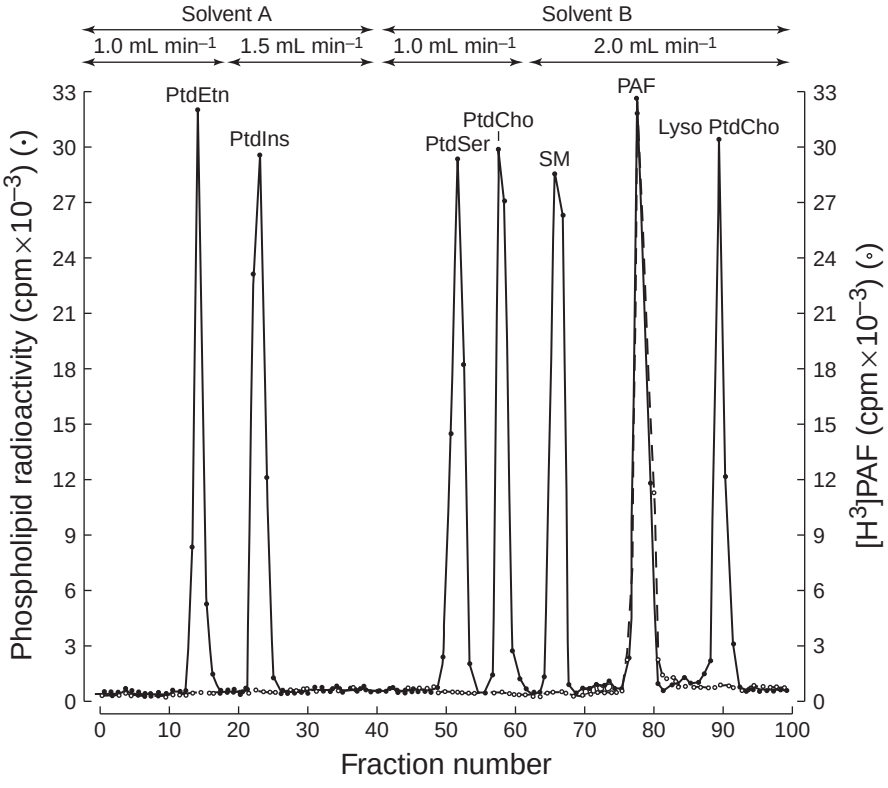

Figure 4 SPE on silica Sep-Pak cartridges as an aid to PAF isolation by HPLC.[34] A mixture of seven radiolabeled phospholipids (60 000 cpm each) was subjected to the SPE procedure. Peak identification is shown. The PAF-enriched fraction was subjected to HPLC with solvent comprising of propanol–toluene–acetic acid–water (93 : 110 : 15 : 15, v/v/v/v) at a flow rate of 2 mL min^{-1} on a Zorbax™ 5 µ silica column (250 × 4.6 mm ID). The HPLC eluate was collected at 1-min intervals over a 50-min run and counted by liquid scintillation spectrometry. The elution profile given is representative of three independent runs. Peak identities were established by the HPLC elution volume and retention time of each radiolabeled lipid standard chromatographed separately. (Reproduced with permission from Janero and Burghardt.[34])

phospholipid fraction was resolved by DEAE-cellulose chromatography (separation sequence 4) as described by Rouser et al.[40] to purify further the inositol lipids. The purified extract is transferred to Sep-Pak cartridge. Various inositol lipids are eluted from the column with a linear gradient of ammonium acetate and water from 0.5% of ammonium acetate (200 mM) and 99.5% of water to 75% of ammonium acetate and 25% of water in 60 min at a flow rate of 0.5 mL min^{-1}. The column eluate for 120 min is collected in 1-mL fractions. Different peaks were identified using PtdIns, PtdIns (4)P and PtdIns (4,5)P_2 as standards. The fractions containing individual inositol lipids were pooled and analyzed for purity by HPLC as described by Van Kessel et al.[41] The phospholipids were detected at 206 nm using a diode array detector programmed to scan from 200 to 350 mm.

Atheroma lipids are isolated from the total chloroform–methanol lipid extracts by TLC and the neutral lipids, PtdCho, SM and LysoPtdCho resolved using chloroform–methanol–water–acetic acid (65 : 25 : 4 : 1, v/v/v/v) as the solvent system.[42]

3.2.2 Reversed-phase Columns

Kaluzny et al.[43] have used aminopropyl bonded phase (Bond-Elut, Analytichem International, Harbor City, CA) for the separation of complex lipids into individual classes in high yield and purity. The method was developed with up to 10 standard lipid classes. It has proven to be rapid and simple for the separation of serum neutral and phospholipids.[44] Figlewicz et al.[45] have shown that the Bond-Elut cartridges can be used to take up lipids from aqueous incubation mixtures. The procedure was simpler than Folch extraction. According to Salari,[46] however, other C_{18}-cartridges (Waters, Milford, MA) did not retain phospholipids significantly when plasma samples diluted with 2 volumes of phosphate-buffered saline acidified to pH 5 with HCl were passed through the adsorbent.

Kim and Salem, Jr[47] have shown that reversed-phase HPLC columns can be effectively employed for a preliminary resolution of phospholipid classes and

molecular species in combination with electrospray mass spectrometry (ESMS) (see below).

3.2.3 Ion-exchange Columns

DEAE-cellulose columns have been extensively utilized for the isolation of particular groups of complex lipids.[40,48] The separation involves partly ionic interactions between the packing material and the polar head groups of complex lipids, and partly adsorption effects with the polar regions of the molecules. According to Christie[48] PtdCho is eluted with chloroform–methanol (9 : 1, v/v), PtdEtn with chloroform–methanol (1 : 1, v/v), PtdSer with glacial acetic acid, and PtdIns with chloroform–methanol (4 : 1, v/v) made 0.05 M with NH_4OAc and 0.56% with respect to NH_4OH. Residual salts are removed from the last fraction by rapid extraction and partition using 0.9% NaCl. Clear-cut fractions are obtained with virtually no cross-contamination using sequential elution from the same DEAE column.

Salari[46] has demonstrated that the Amberlite resins were selective for removal of phospholipids by direct adsorption from plasma. A recovery of greater than 85% was obtained for PtdCho, PtdIns, and PtdSer when XAD-7 or XAD-8 were used as adsorbents. A recovery of about 90% for PtdEtn was obtained when XAD-2 or XAD-4 were used as adsorbents. A solvent mixture of isopropanol–acetonitrile (1 : 1, v/v) was found to be the most effective eluent for removal of phospholipids from the Amberlite polymeric resins. Prior to this time, Salari[49] had demonstrated greater than 95% recovery of the PAF from plasma and incubation media using Amberlite XAD-2 (Serva Feinbiochemica, Heidelberg, Germany) as retaining medium. For this purpose fresh plasma (heparin anticoagulant) is diluted with 2 volumes of Tyrod's buffer and the mixture is acidified to pH 5 with 1 M HCl. The acidified solution is passed through an XAD-2 (1 g dry resin in 10 mL glass syringe) column, which has been previously sequentially washed with 20 mL acetonitrile, 10 mL methanol and 40 mL distilled water. The column is allowed to run at a flow rate of approximately 0.5 mL min^{-1}. Plasma eluates are collected for determination of unretained lipids. The columns are washed with 3×5 mL of distilled water and each individual 5 mL of water eluate is collected. The columns are then eluted with 15 mL methanol at a flow rate of approximately 1.5 mL min^{-1}. Under these conditions 96% of added radioactive 1-*O*-alkyl-2-acetyl-*sn*-glycerol-3-phosphocholine is recovered in the methanol fraction.

Dionex-anion-exchange HPLC has been used for the separation of glucose oligomer standards and PtdIns glycan fragments.[13] The analysis is performed on an HPIC AS6 column (Dionex Corporation, Sunnyvale, CA) using a gradient system. The gradient began with 100% solvent A (0.1 M NaOH), 0% solvent B (0.5 M CH_3COONa, 0.1 NaOH) up to 3 min after sample injection, followed by a linear change to 55% solvent A, 45% solvent B at 33 min, then 0% solvent A, 100% solvent B at 38 min. The flow rate is 1 mL min^{-1}. The glucose oligomer standards and carriers are prepared by partial acid hydrolysis of dextran. Glycosyl PtdIns glycans prepared by nitrous acid treatment of nonprotein-linked glycosyl PtdIns can be analyzed by conventional anion-exchange chromatography to determine their net negative charge.[13] The analysis is conveniently performed with a Mono Q column (HR 5/5, bed size 5×50 mm) and a gradient fast protein liquid chromatography (FPLC) system (Pharmacia, Stockholm).

3.2.4 Affinity Columns

The antibiotic neomycin exhibits a specific affinity for the polyphosphoinositides. The other weakly acidic phospholipids are retained because neomycin also acts as a weak anion exchanger. For the extraction and purification of PtdIns P and PtdIns P_2, Palmer[50] and Letcher et al.[51] recommend the immobilized neomycin columns originally described by Schacht.[52] All anionic lipids present in chloroform–methanol extracts, which have been washed first with acid and then with neutral salts solution are adsorbed. PtdSer and PtdOH are eluted with chloroform–methanol–formic acid mixtures. PtdIns and cardiolipin are eluted sequentially with very low concentrations of ammonium formate in chloroform–methanol–water. All three PtdIns Ps are isolated from washed chloroform–methanol–KCl extracts of rat brain. PtdIns P and PtdIns P_2 are eluted in turn by higher salt concentrations (0.4–1.0 M). Similar methods have been used to isolate the PtdInss from human plasma lipoproteins.[50] Letcher et al.[51] have found that simply using the eluants described by Palmer[50] results in cross-contamination between the three principal inositol lipids, and to prepare really pure PtdIns P, a stepwise elution has to be carried out from 200 mM to 500 mM in 100 mM steps and then taking only the PtdIns P revealed to be pure by TLC analysis.[53]

3.3 Derivatization and Preliminary Segregation

Although combinations of HPLC and MS now allow the detection and quantification of many phospholipid classes and molecular species, there remains a frequent need for a preliminary resolution of the phospholipid mixtures.[54,55] Both molecular subclasses and molecular species of various phospholipids are more readily identified by MS following a preliminary chromatographic resolution and purification with or without derivatization. Preparation of derivatives usually renders the lipid molecules less

polar or endows them with ultraviolet (UV) absorbing or fluorescent properties. The derivatization can be effectively combined with SPE using normal-phase cartridges or TLC (see above).

3.3.1 Enzymatic

The polar headgroups of glycerophospholipids and SMs, which cause peak tailing and interfere with chromatographic resolution of molecular species, may be removed by hydrolysis with phospholipase C and the released diacylglycerols and ceramides subjected to trimethylsilylation or *tert*-butyldimethylsilylation to obtain derivatives with excellent gas chromatographic properties and characteristic mass spectra. *tert*-Butyldimethylsilyl (TBDMS) ethers of diacylglycerols are prepared by reaction with a solution of *tert*-butyldimethylchlorosilane–imidazole (1 : 2.5, mol/mol, diacylglycerol–reagent) in dimethylformamide at 80 °C for 20 min.[56] The mixture is extracted with chloroform–methanol (2 : 1, v/v) and washed with water. The trimethylsilyl (TMS) ethers are prepared by reacting the diradylglycerols with a premixed solution of pyridine–hexamethyldisilazane–trimethylchlorosilane (12 : 5 : 2, v/v/v) for 30 min at room temperature[57] or by reaction with a solution of pyridine–bis(trimethylsilyl)-trifluoroacetamide–trimethylchlorosilane (50 : 49 : 1, v/v/v), which yields volatile byproducts.[58] The reagents are evaporated under nitrogen and the products dissolved in 2 mL of petroleum ether (b.p. 30–60 °C). Derivatization of the diradylglycerols immediately after phospholipase C digestion is needed to prevent isomerization during the subsequent chromatographic analyses. The separation of the TMS ethers by normal-phase HPLC yields the alkylacyl, alkenylacyl and diacyl subclasses of the Cho and Etn glycerophospholipids. Attempts to resolve these derivatives by TLC leads to significant breakdown of the TMS ethers. Figure 5(a) shows the separation of alkenylacyl, alkylacyl and diacyl subclasses of the diradyl GroPEtn moieties of the red cells.[59] A similar separation of the diradylglycerol moieties of red cell PtdCho gave much smaller but readily detectable fractions for the ether-linked species. These separations are performed with the TMS ethers of the diradylglycerols. Based on GLC quantification the red cell diradyl GroPEtn contained 48.8% diacyl, 47.8% alkenylacyl and 3.4% alkylacyl, whereas the diradyl GroPCho contained 93.0% diacyl, 4.6% alkylacyl and 2.5% alkenylacylglycerol subclasses. The diradyl GroPIns is 100% diacyl. The detailed FA composition of the various diradylglycerol fractions derived from them is discussed below under molecular species. Figure 5(b) shows the normal-phase separation of alkenylacyl, alkylacyl and diacylglycerol moieties of human plasma diradyl GroPEtn.[60] Although the ether-linked glycerophospholipid content is not measured in the individual lipoprotein classes, it has been observed that each lipoprotein contained at least some of this lipid class. The plasma diradyl GroPEtn consisted of 71.8% alkenylacyl, 19.9% diacyl and 8.3% alkylacyl, whereas the diradyl GroPCho contained 93% diacyl, 3.4% alkylacyl and 3.6% alkenylacylglycerol. The HPLC procedure was clearly superior to the TLC method used previously,

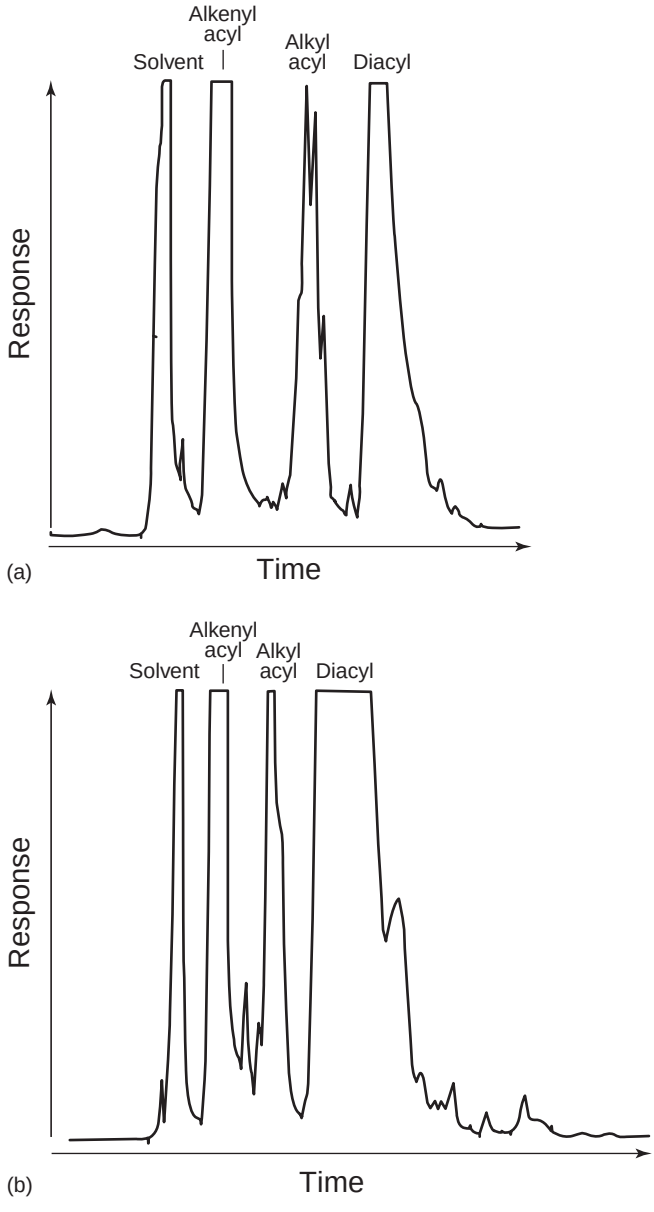

Figure 5 Normal-phase HPLC separation of alkenylacyl, alkylacyl and diacylglycerol moieties of (a) human red blood cell diradyl GroPEtn[59] and (b) human plasma diradyl GroPEtn.[60] Peak identification is given. HPLC conditions: column, Supelcosil LC-Si (5 m) 250 × 4.6 mm ID; solvent, hexane–isopropanol (99.7 : 0.3, v/v), 1 mL min^{-1}; temperature, 30 °C, isothermal. Sample: diradylglycerol TMS ethers. (Reproduced with permission from Myher et al.[59,60])

especially with respect to the resolution of the alkylacyl and diacylglycerol fractions. There is also a partial separation of the molecular species within each diradylglycerol class. The detailed FA composition of the various diradylglycerol fractions derived from them is discussed below under molecular species. Phospholipase C treatment of SM results in release of the ceramide moieties, which also can be converted into the silyl ethers for GLC.[59,60] Phospholipase C dephosphorylation of plasma total lipids has been extensively utilized for a preliminary resolution and quantification of plasma PtdCho, lysoPtdCho and SM,[20] which have yielded estimates comparable to those obtained by combined TLC separation and phosphorus analysis.[61]

For HPLC resolution the diradylglycerols released from the glycerophospholipids by phospholipase C may be converted into benzoates,[62] pentafluorobenzoates,[63] or other UV absorbing or fluorescent derivatives. The diradylglycerols are converted into the naphthylethyl urethanes for resolution as the diastereomers on normal-phase HPLC[64] or into dinitrophenylurethanes for separation of enantiomers on a chiral phase column.[65]

Other useful derivatives for the identification and/or resolution of glycerophospholipids may be prepared by phospholipase D, which releases the PtdOH moiety along with the nitrogenous base or inositol moiety. The PtdOH may be methylated[66] and the lysoglycerophospholipids acetylated[66] for further chromatographic and mass spectrometric characterization.

The positional distribution of FAs in the glycerophospholipids is determined by hydrolysis with phospholipase A_2 (*Crotalus adamanteus*, Sigma Chemical Co.). For this purpose a portion of the purified phospholipid (<0.5 mg) is vortexed for 3 h at 37 °C in a mixture made up of 2 mL of Tris buffer (17.5 mM) trishydroxymethylaminomethane, adjusted to pH 7.3 with dilute HCl, 1.0 mM $CaCl_2$, 2 mL diethyl ether, 50 μg BHT and 50 units of phospholipase A_2.[58] After acidification with one drop of concentrated HCl, the mixture is extracted first with five parts chloroform–methanol (2:1, v/v), then with two parts chloroform–methanol (4:1, v/v). The extract is then resolved by TLC on silica gel H plate (20 × 20 cm). The plates are first developed to a height of 11 cm using chloroform–methanol–acetic acid–water (100:35:10:3, v/v/v/v), dried under nitrogen gas for 10 min and then developed to a height of 15 cm using heptane–isopropyl ether–acetic acid (60:40:4, v/v/v). Alkylglycerols are not determined by this procedure and must be analyzed separately.[67] In order to measure the extent of hydrolysis, one set of the tubes is quantified by GLC after addition of 50 μg of methyl heptadecanoate as internal standard. The residual glycerophospholipids usually correspond to approximately 0.5% of total.

The sidedness of the phospholipid composition of the red cell membrane can also be determined by phospholipase A_2. Using bee venom or *Naja naja naja*, Wilson et al.[68] demonstrated that hyperglycemia induces loss of asymmetry due to increased passive phospholipid flip-flop caused by a secondary effect of hyperglycemia, such as phospholipid peroxidation, or nonenzymatic protein glycosylation.

3.3.2 Chemical

Alternatively, PtdEtn and PtdSer may be subjected to N-derivatization to improve their chromatographic and mass spectrometric properties.[47] It permits reversed-phase chromatography with UV and mass spectrometric detection. Of special interest is the preparation of the trinitrobenzene sulfonate derivatives of the aminophospholipids from intact red cell ghosts, which have been utilized in studies of the molecular species distribution of aminophospholipids between the inner and outer halves of the phospholipid bilayer of the red blood cell.[69,70] Since this approach is not applicable to the Cho-containing glycerophospholipids, chemical methods have been sought for removing the polar head groups. Dephosphorylation of glycerophospholipids comparable to that of phospholipase C hydrolysis may be obtained by acetolysis, which, however, leads to partial isomerization of the released diacylglycerols and destruction of plasmalogens.[58] Acetolysis may provide a desirable choice for chemical degradation of PtdIns glycans, despite isomerization of the diradylglycerol moieties.[12]

Acid vapor decomposition is a crude but effective method for the removal of plasmalogens from glycerophospholipid mixtures.[71,72] Alkaline destruction of the glycerophospholipids is frequently used to purify the SM and its metabolites, which are relatively resistant to alkaline hydrolysis, and can be subjected to mild alkaline hydrolysis in 0.6 N NaOH in methanol, for 1 h at 37 °C to destroy the glycerophospholipids.[73] The phases are separated by adding 1 mL each of chloroform and water, and the organic phase is dried through an anhydrous sodium sulfate column. The alkali-stable lipids are separated on high-performance thin-layer chromatography (HPTLC) plates using chloroform–methanol–acetic acid–water (25:15:4:2, v/v/v/v) as the developing solvent. Lipids are visualized under iodine. For further analysis, ceramide, SM and other alkali-stable compounds are scraped off the plate and the lipid material is eluted from the silica gel with chloroform–methanol (2:1, v/v).

In other instances the hydrophobic moieties of the glycerophospholipids have been removed by deacylation and the water-soluble residues analyzed by chromatography. The deacylation procedure may be performed as follows.[74] Labeled PtdIns phosphates that have been

produced as described above and extracted in chloroform are washed with an equal volume of methanol–0.1 M EDTA (1 : 0.9, v/v) and placed in a glass, screw-capped scintillation vial (20 mL capacity), and then dried under a stream of nitrogen gas. In a fume cupboard, 1.8 mL of methylamine reagent (42.8% of 25% (v/v) methylamine in water, 45.7% (v/v) methanol, 11.5% (v/v) 1-butanol, stored at 4 °C) are added to dissolve and hydrolyze the dried lipids. The tightly capped vial is incubated at 53 °C with constant shaking for 50 min. After this incubation, the contents of the vial are cooled to room temperature, transferred to a 2.0 mL microfuge tube, and dried in vacuo. The dried samples are resuspended in 2.0 mL of distilled, deionized water, transferred to a glass test tube, and extracted two times with an equal volume of 1-butanol–petroleum ether–ethyl formate (20 : 4 : 1, v/v/v) to remove fatty acyl groups. After these extractions, the lower, aqueous phase is dried in vacuo by roto-evaporation and stored at −70 °C until analysis by HPLC. The deacylation may be conveniently performed following TLC purification. The region of the TLC plate that contains a particular phospholipid is carefully excised, placed in a glass screw-capped scintillation vial, and treated with methylamine reagent exactly as described above. The deacylation residues of polyphosphoinositides are resolved by HPLC using a high resolution 5 μm Partisphere SAX column (Whatman) and a shallow discontinuous salt gradient. The HPLC column is equilibrated with water prior to sample loading and is eluted with a discontinuous gradient up to 1 M $(NH_4)_2HPO_4 \cdot H_2PO_4$ (pH 3.8) at a flow rate of 1 mL min^{-1}. The gradient is established from 0 to 1 M of the phosphate buffer over 15 min. Kucera and Rittenhouse[75] have employed a similar HPLC column along with a similar buffer gradient for the separation of GroPIns polyphosphates derived from human platelets.

The methyl esters of phospholipids are prepared by reaction with 6% H_2SO_4 in methanol.[58] The methyl esters are extracted with hexane. The linked FAs of the ceramides are converted into their methyl esters by treating an aliquot of the sample (about 50 μg) with 0.5 N methanolic HCl (80 °C, 20 h) and extracting the FA methyl esters into hexane. Aliquots of the resolved diradylglycerol TMS ethers are converted into the FA methyl esters and dimethylacetals by a 2 h reaction at 80 °C with freshly prepared methanol–sulfuric acid (94 : 6, v/v). Alkylglycerol TMS ethers are converted into the diacetates with acetic anhydride–pyridine (1 : 1, v/v) after transmethylation of the alkylacylglycerol TMS ether with 1 N sodium methoxide in methanol–toluene (3 : 2, v/v) for 15 min at 20 °C. Following methanolysis, the mixture is neutralized with 1% acetic acid in hexane and extracted by adding 200 μL chloroform, 0.5 mL water and 50 μL 3 M aqueous ammonia. The organic phase is washed with 250 μL water and then dried by passing it through a small column of anhydrous sodium sulfate. After drying under nitrogen, the sample is acetylated for 0.5 h at 80 °C with 75 μL acetic anhydride–pyridine (1 : 1, v/v). The reagent is removed by evaporation under nitrogen and the products purified by HPLC.

4 RESOLUTION OF COMMON PHOSPHOLIPIDS

LC/MS and MS/MS with electrospray now offer an analytical method clearly superior in resolution, speed and overall convenience to either HPLC or TLC/GLC methods for the analysis of many phospholipid classes and molecular species. However, the older chromatographic methods are still required for the isolation and purification of standards and occasionally for verification of the mass spectrometric results obtained on the more complex lipid mixtures. A brief description of the principles of the various methods and the merits of each for specific applications is therefore included here.

4.1 Principles of Methodology

4.1.1 Thin-layer, Liquid and Gas–Liquid Chromatography

TLC is the simplest and most effective technique for the separation of phospholipids. The separation is based on the differential migration of the various phospholipid classes on thin layers of adsorbent subjected to continuous flow of appropriate organic solvents. The resolved components are readily located by spraying the plate with fluorescent dyes and viewing it under UV light. The adsorbent (usually silica gel) can be scraped off the plate and extracted with organic solvents to recover the purified compounds. Normal-phase HPLC resolves the phospholipid classes on an adsorbent column usually employing a continuous flow of solvent gradient, although isocratic separations are also practiced. Reversed-phase HPLC resolves both lipid classes and molecular species of phospholipids on a hydrophobic column (usually a silanized adsorbent). The separation is based on both adsorptive and partition properties of the phospholipid molecules and is best performed following a preliminary resolution of the lipid classes by TLC or normal-phase HPLC. HPLC requires a special effluent monitor, such as light scattering or a flame ionization detector, because most lipid classes do not absorb UV light or possess fluorescence unless appropriately derivatized. The resolved components are retrieved by a fraction collector. Modern HPLC systems are fully automated and computer controlled. Gas chromatography is the most effective procedure for

the analysis of the FA, diacylglycerol and ceramide moieties of the phospholipids. Modern gas chromatographs equipped with polar capillary columns up to 100 m in length provide complete resolution and quantification of FA methyl esters ranging from butyric to docosahexaenoic acids. Polar capillary columns up to 15 m in length and high thermal stability provide complete resolution of the molecular species of the diacylglycerol and ceramide moieties of phospholipids as the TMS ether derivatives. In all instances highly sensitive detection and accurate quantification is obtained by the flame ionization detector, which is a standard part of the modern gas chromatographs.

4.1.2 Liquid Chromatography with On-line Electrospray Mass Spectrometry

ESMS is a relatively new method of soft ionization of phospholipid molecules. In ESMS, ions dissolved in a suitable solvent are charged and droplets are formed at atmospheric pressure and room temperature by electrostatic nebulization. The solvent is evaporated by a curtain of gas until only charged solute remains, which is subsequently drawn into the low vacuum of a mass analyzer. HPLC with on-line ESMS (LC/ESMS) is an ideal combination of chromatographic separation with mass spectrometric identification and quantification of phospholipid classes and molecular species. The ions are characterized by their retention time and the mass/charge (m/z) ratio. Phospholipids usually exhibit a charge of 1. Normal-phase HPLC can be used to resolve the lipid classes and to admit them via the column effluent to the ionization chamber. Likewise, reversed-phase HPLC can be used to resolve molecular species of phospholipid classes and to admit them to the ESMS ionization chamber. In the positive ion mode only the Cho-containing phospholipids are effectively determined, while in the negative ion mode only the acidic phospholipids appear, although the chloride adducts of the Cho-containing phospholipids may also be seen. The ions appearing in the mass spectrometer can be subject to collision-induced dissociation (CID) to produce daughter ions as a further aid to identification and quantification of molecular species (see section 4.3). A pseudo MS/MS effect may be achieved by conducting the LC/ESMS analysis at two different capillary exit voltages. At a low voltage only molecular ions are obtained, while at elevated voltage the molecular ions are subject to CID. Provided the chromatographic peaks have been fully resolved LC/ES/CID/MS yields structural information very much like conventional MS/MS. LC/ES/CID/MS has the advantage of up to 10 times greater sensitivity than MS/MS.[21]

4.1.3 Flow Mass Spectrometry/Mass Spectrometry

In MS/MS, precursor ions are selected with the first of three quadrupoles (Q1) for CID with argon in the second quadrupole (Q2). The third quadrupole (Q3) is scanned with a mass step of 0.02 Da and 1 ms per step. Parent ion transmission is maximized by reducing the resolution of Q1 to transmit a 2 to 3 m/z window about the selected parent ion, and Q3 resolution is adjusted to approximately 50% valley between peaks 3 Da apart.[55] Since the MS/MS approach provides nearly all the information necessary for identification of common phospholipid classes and molecular species, several authors have recommended the replacement of the chromatographic step with a simple flow injection.

Furthermore, the flow MS/MS methods can be employed with total lipid extracts thus eliminating the need for time-consuming and labor-intensive preliminary segregation and derivatization of the phospholipid classes in biological samples.[55,72,76] As a result, lipid class and molecular species analyses can be completed within a few hours instead of the many weeks and months necessary for the combined TLC, HPLC and GLC in the past.

The elimination of the laborious work-up of the samples and the continued exposure to air and active surfaces has led to the recovery of greatly increased proportions of the polyunsaturated and more polar lipid classes and molecular species using the LC/MS and MS/MS approach. A comparison of the MS/MS results with the results of reversed-phase HPLC and capillary GLC obtained with the monobenzoate derivatives of the diradylglycerols derived by phospholipase C hydrolysis of the diradylglycerophosphates has led to the discovery of marked quantitative differences although the qualitative results have been the same between the two methods.[76] The quantitative differences in these techniques are due to the fact that reverse-phase HPLC cannot be utilized to accurately quantitate minor constituents (1–3% of total) and that peaks which are thought to represent single molecular species actually contain multiple different molecular species. Large errors in molecular species analysis of Cho glycerophospholipids by reverse-phase HPLC are attributed to the coelution of multiple molecular species (e.g. 18 : 0–18 : 2 and 18 : 1–18 : 1) with 16 : 0–18 : 1 which are subsequently independently documented by gas chromatographic analysis of column effluents. Furthermore, ESMS yielded greatly increased estimates of the plasmenylEtn molecular species containing arachidonate, which had been previously underestimated due to silicic acid catalyzed hydrolysis of the vinyl ether linkage utilizing conventional chromatographic procedures. Similar observations have been made by comparing the LC/MS and TLC/GLC results for the molecular species of glycerophospholipids of plasma and red cells, with the largest discrepancies being found between the estimates of the

more unsaturated long chain species. Both peroxidation and incomplete recoveries of the longer chain molecular species from the polar capillary columns have suggested a reasonable explanation for the quantitative difference. The success of the mass spectrometric analyses can be attributed directly to the development of the mild ionization techniques, such as electrospray and atmospheric pressure ionization, which, in the initial ionization mode, yield largely the molecular ions with a minimum of fragmentation.

4.1.4 Other Methods

Another method potentially applicable to the rapid determination of the lipid classes in plasma, red cells and platelets is nuclear magnetic resonance (NMR) spectroscopy. NMR spectra arise when a powerful external magnetic field is applied to the nucleus and is made to oscillate between different quantized energy levels at specific radiofrequencies. Very small changes in absorbed energy can be detected, amplified and displayed on a chart. The trace obtained of the variation in the intensity of the resonance signal with increasing applied magnetic field is the NMR spectrum. In organic compounds, the isotope of hydrogen (^{1}H) and to a lesser extent the isotope of carbon (^{13}C) exhibit this phenomenon, although at markedly different sensitivities. In the phospholipids phosphorus (^{31}P) exhibits this phenomenon and ^{31}P-NMR has been widely used for phospholipid quantification, although no specific applications to blood plasma or the formed elements have been reported.[77] A solution of phospholipids is treated with Cs-EDTA to destroy the vesicles and improve band width. After this treatment the ^{31}P-NMR spectrum shows very narrow lines with very good dispersion, so that in most cases phospholipids give signals that are separated from each other. The whole liquid is put into the NMR tube, the lower phase containing the phospholipids and the upper phase containing the waste. The molar amount of phospholipid is calculated based on the fact that the ratio of two signal integrals is equal to the ratio of the corresponding molar amounts, that of the internal standard being known. The signal-to-noise ratio depends on the number of accumulated pulses. A minimum of about 250 pulses are sufficient for routine analysis using a 400-MHz instrument. Approximately 15 to 20 mg phospholipid are required in routine analyses, with components down to 0.05% being detectable. However, measurements have been made on samples as little as 50 μg. The phosphorus nucleus needs a certain amount of time (the relaxation time) to return to the ground state after a pulse. This time depends on the chemical environment of the P atom. A total of 18 individual phospholipids have been identified and quantified including many minor and unusual phospholipid classes. Additional information about FAs, oxidation state, and byproducts can be derived from ^{1}H- and ^{13}C-NMR spectroscopy.

4.2 Chemical Classes

4.2.1 Glycerophospholipids and Sphingomyelins

Plasma very low density lipoprotein (VLDL), low density lipoprotein (LDL), high density lipoprotein (HDL) and chylomicrons possess distinct PtdCho, PtdEtn and SM compositions. These compositions differ from those of the red cell and platelet membranes and from the plasma membranes of the cells lining the vascular bed and atheroma. They also differ from those of the liver cell, where they are synthesized and assembled into the lipoprotein particles.

4.2.1.1 Plasma Lipoproteins The common phospholipid classes of plasma lipoproteins are readily determined by normal-phase HPLC with on-line ESMS (LC/ESMS) without prior modification of the lipids by enzymatic or chemical means. Figure 6 shows the total positive and negative ion current profiles recorded for the LDL from normal subjects.[78] Specifically, the phospholipids were resolved with a Supelcosil (25 × 0.46 cm ID) column using a linear gradient of 100% solvent A (chloroform–methanol–30% ammonium hydroxide, 80 : 19.5 : 0.5, v/v/v) to 100% solvent B (chloroform–methanol–water–30% ammonium hydroxide, 60 : 34 : 5.5 : 0.5, v/v/v/v) in 35 min.[22] The Cho-containing phospholipids (PtdCho, SM and lysoPtdCho, and PAF) are analyzed in the positive ion mode while the acidic phospholipids (PtdSer, PtdIns, PtdOH, PtdGro and cardiolipin) are analyzed in the negative ion mode. PtdEtn can be detected in both positive and negative ion modes, but the response is higher in the negative ion mode. This column resolves the Cho and sphingosine phospholipids into two or more peaks on the basis of chain length, with the longer chains being eluted earlier. However, determinations of minor components in the negative ion mode are compromised by the presence of significant amounts of chloride adducts of Cho-containing phospholipids. The phospholipid classes are identified from the specific head group ions and the molecular species from the diradylglycerol fragment ions obtained by MS/MS or pseudo MS/MS (LC/ES/CID/MS), where CID is obtained in a single quadrupole instrument by rerunning the sample at a higher capillary exit voltage. Plasmalogens elute later than the acyl phospholipids of corresponding chain length and degree of unsaturation. The method has been applied to phospholipid analyses in normal and acute phase human plasma HDL before and after hydrolysis with secretory phospholipase A_2.[79] The phospholipid classes are

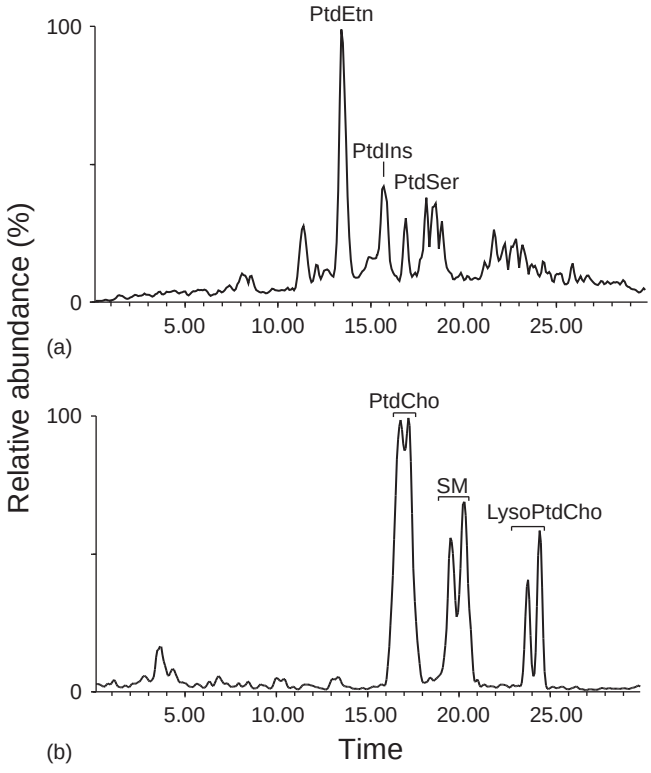

Figure 6 Total negative (a) and positive (b) ion current profiles of total lipid extracts of human LDL by normal-phase LC/ESMS.[78] PtdEtn includes its alkyl and alkenyl ether; PtdCho includes its alkyl ether homologs; LysoPtdCho, lysophosphatidyl-choline including alkyl ether homologs. Column, Spherisorb 3 μ (100 mm × 4.6 mm ID, Alltech Associates, Deerfield, IL); solvent, linear gradient of 100% solvent A to 100% solvent B in 14 min, then solvent B for 10 min; solvent A, chloroform–methanol–30% ammonium hydroxide (80 : 19.5 : 0.5, v/v/v); solvent B, chloroform–methanol–water–30% ammonium hydroxide (60 : 34 : 5.5 : 0.5, v/v/v/v); Model 1060 liquid chromatograph interfaced with Hewlett-Packard Model 5988B quadrupole mass spectrometer equipped with a nebulizer-assisted electrospray interface. Ion spectra were taken in the mass range 300–1100.

readily quantified by including appropriate species of standard phospholipids with the LC/ESMS run (see section 4.3).

The glycerophospholipid and SM classes of plasma lipoproteins are readily resolved by TLC using chloroform–methanol–acetic acid–water (75 : 45 : 12 : 6, v/v/v/v) as solvent.[59,60] This solvent system yields essentially pure fractions of PtdCho, PtdEtn and SM, which are recovered from the silica gel by repeated extraction with chloroform–methanol–acetic acid–water (50 : 39 : 1 : 10, v/v/v/v).[80] To each eluate is added 4 mL of 1 M NH_4OH. The recovered PtdSer and PtdIns, which overlap, are resolved using chloroform–methanol–3 M ammonia (65 : 35 : 6.5, v/v/v) as solvent.[81] The relative proportions of these lipid classes are determined by analysis of phosphorus, FAs or of the dephosphorylated residues of the phospholipids, all of which have given similar results. Thus, while the diacyl GroPCho and SM comprise the bulk, the diradyl GroPEtn and PtdIns constitute minor components of plasma lipoproteins.

The glycerophospholipid fractions may also be resolved by TLC on commercially prepared silica gel G plates (20 × 20 cm, 250 μm thick layer) supplied by Analtech (Fisher Scientific, Boston, MA). PtdCho (retention factor, R_f 0.13) and PtdEtn (R_f 0.59) are isolated using chloroform–methanol–water (25 : 10 : 1, v/v/v, plus BHT) as the solvent.[80] PtdIns (R_f 0.32) is purified similarly using chloroform–methanol–acetic acid–water (50 : 30 : 8 : 4, v/v/v/v, plus BHT).[81] BHT is added as an antioxidant at 50–100 mg L^{-1} of solvent. The phospholipids are located by spraying with 0.1% 2,7-dichlorofluorescein in 95% methanol and by comparing the R_f values to those of reference standards. The phospholipids are recovered from the silica gel scrapings by extraction with chloroform–methanol (1 : 1, v/v) immediately after separation and before the plate has dried.

The phospholipid class composition of whole plasma is of the order reported previously on the basis of TLC separation and phosphorus analyses[82] or GLC analyses of the component FAs,[59] with 67% diradyl GroPCho, 17.7% SM, 2.5% diradyl GroPEtn and 2.1% diradyl GroPIns. The phospholipid composition of individual plasma lipoproteins has been reported by several groups of investigators using a variety of methods. The early data compiled by Skipski[83] remains to date the most complete reference to the phospholipid composition of VLDL, LDL and HDL phospholipids in man (Table 1). Detailed analyses of selected lipid classes of specific lipoproteins have been reported more recently, e.g. chylomicrons,[4] VLDL and LDL[53] and HDL_{2+3}.[84]

Breckenridge and Palmer[54] have reported the PtdIns and PtdIns P content of plasma VLDL, LDL and HDL in comparison to that of platelets and erythrocytes. HDL contained a greater proportion of PtdIns (2.6% of total phospholipid) than did either LDL or VLDL (1.6 and 1.7%, respectively). Breckenridge and Palmer[54] also quantitated the PtdSer of HDL, LDL and VLDL of human plasma.

Kuksis et al.[61] have used GLC to measure the content of the lysoPtdCho, PtdCho and SM of plasma lipoproteins as the monoacylglycerols, diacylglycerols and ceramides, respectively, released from the lipoprotein samples by phospholipase C, following trimethylsilylation and addition of an internal standard. The values derived for total plasma PtdCho (177 mg%) and SM (46 mg%) compared favorably to those obtained by TLC separation of the phospholipid classes and determination of phosphate (PtdCho, 165 mg% and SM, 47 mg%).

Table 1 Phospholipid composition of plasma and plasma lipoproteins as determined by different analytical methods (mol% of total phospholipids)

Components[a]	Plasma[82]	CHYLOS[5]	VLDL[83]	LDL[83]	HDL_{2+3}[83]
PtdCho	64.9 ± 2.3	78.7 ± 1.2	59.7	63.7	74.4
LysoPtdCho	4.3 ± 2.3	1.0 ± 1	5.0	2.7	2.9
SM	18.1 ± 2.2	0.5 ± 0.2	14.8	25.9	13.2
PtdEtn (including plasmalogens)	3.3 ± 1.2	8.1 ± 1.7	4.6	2.2	3.1
PtdSer	3.6 ± 0.7	0.1 ± 0.0	1.5	0.8	0.8
PtdIns	4.3 ± 0.3	2.7 ± 0.5	3.6	1.6	2.4
PtdGro		1.6 ± 0.2			
PtdOH		1.1 ± 0.3	7.6	2.0	2.2
DMPdEtn		1.3 ± 0.5			
Other	1.5		3.2	1.1	0.9

[a] LysoPtdCho, lysophosphatidylcholine; DMPtdEtn, *N*,*N*-dimethyl PtdEtn.
CHYLOS, plasma chylomicrons; VLDL $<1.006\ g\ mL^{-1}$; LDL $1.006-1.063\ g\ mL^{-1}$; HDL_{2-3} $1.063-1.210\ g\ mL^{-1}$.

4.2.1.2 Red Cells and Platelets The major phospholipid classes of red cells and platelets are also readily detected and quantified by normal-phase LC/ESMS or by ESMS/MS as the intact molecules. Han and Gross[72] have reported direct analysis of phospholipid classes and molecular species of phospholipids by flow injection ESMS/MS. The various phospholipid classes are recognized and quantified by the yields of the individual molecular species measured in the presence of appropriate internal standards (e.g. 16 : 0–18 : 1 GroPSer, 14 : 0–14 : 0 GroPIns; 15 : 0–15 : 0 GroPCho; 15 : 0–15 : 0 GroPEtn).

Ravandi et al.[78] have employed LC/ESMS for the analysis of the red cell phospholipids of diabetic subjects. The Cho-containing phospholipids were detected and quantified in the positive ion current, while the Etn-containing phospholipids and the acidic phospholipids were determined in the negative ion mode. Two or three peaks each due to a partial fractionation by chain length on the adsorbent column were found. Summation of the peak areas and correction for differences in ion yields provides valid estimates of the phospholipid class content of the red cells. In the LC/ESMS method the lipid classes are quantified on basis of the parent ion intensities.

Individual phospholipid classes of the red cells can be resolved by unidimensional TLC of the total lipid extract using chloroform–methanol–acetic acid–water (75 : 45 : 12 : 6, v/v/v/v) as solvent,[60] and are recovered from the silica gel by repeated extraction with chloroform–methanol–acetic acid–water (50 : 39 : 1 : 10, v/v/v/v) as described for plasma lipoproteins. The relative proportions of the various lipid classes estimated on the basis of the derived diacylglycerol and ceramide moieties or of FA analyses in the presence of internal standards found by this method are similar to those recorded by earlier workers who estimated the phosphorus content of the resolved lipid classes.[83]

Mallinger et al.[85] have described a two-dimensional TLC method, based on several novel modifications of existing techniques, for concurrently analyzing nanomolar amounts of nine phospholipid classes in a single aliquot of red cell membrane extract. The phospholipid classes are first separated by two-dimensional TLC and then determined using two-dimensional scanning laser densitometry. Tables 2 and 3 give relative phospholipid composition of erythrocyte and platelet membranes from control subjects using acidic extraction of washed membranes. The data obtained by densitometric measurements compare closely to those obtained in an earlier

Table 2 Relative phospholipid composition of erythrocyte membranes from control subjects (mol%)

Phospholipid class	Mallinger et al.[85] (mean ± SD, $n = 11$)	Mitchell et al.[86] (mean ± SD, $n = 32$)
PtdCho	28.7 ± 1.6	28.2 ± 2
LysoPtdCho	2.43 ± 1.65	1.1 ± 1
SM	24.6 ± 2.6	25 ± 3
PtdEtn	30.2 ± 2.3	26 ± 2
PtdSer	12.4 ± 1.2	15 ± 2
PtdIns	0.62 ± 0.21	1.9 ± 0.7
PtdIns (4)P	0.2 ± 0.13	0.9 ± 0.4
PtdIns (4,5)P_2	0.98 ± 0.26	1.5 ± 0.4

Table 3 Relative phospholipid composition of platelet membranes from control subjects (mol%)

Phospholipid class	Mallinger et al.[85] (mean ± SD, $n = 5$)	Mitchell et al.[86] (mean ± SD, $n = 4$)
PtdCho	42.6 ± 1.9	36 ± 5
LysoPtdCho	3.51 ± 1.6	1 ± 1
SM	16.7 ± 1.8	19 ± 4
PtdEtn	24.5 ± 8	27 ± 5
PtdSer	10.2 ± 1.9	12 ± 2
PtdIns	2.36 ± 87	4.5 ± 0.9
PtdIns (4)P	0.11 ± 0.05	0.9 ± 0.4
PtdIns (4,5)P_2	0.08 ± 0.05	0.4 ± 0.2

study by Mitchell et al.,[86] in which radiolabeling with ^{32}P and quantification by phosphate assay were used.

Studies using membrane impermeant reagent TNBS have shown that aminophospholipids are primarily localized on the plasma membrane interior of the red cell, although some of the PtdEtn is found on the cell surface.[70] Han et al.[76] have used ESMS/MS to measure the alterations in individual phospholipid classes and molecular species of platelet phospholipids during thrombin stimulation. By exploiting the analytical power and sensitivity of ESMS they have identified plasmenylethanolamines as the largest source of arachidonic acid mass during thrombin stimulation.

Figure 7 shows a TLC separation of human platelet phospholipids as obtained with chloroform–methanol–acetic acid–water (50 : 37.5 : 3.5 : 2, v/v/v/v) as the developing solvents.[87] This system has been routinely employed to separate PtdIns from PtdSer on precoated commercial silica gel H plates (E. Merck).

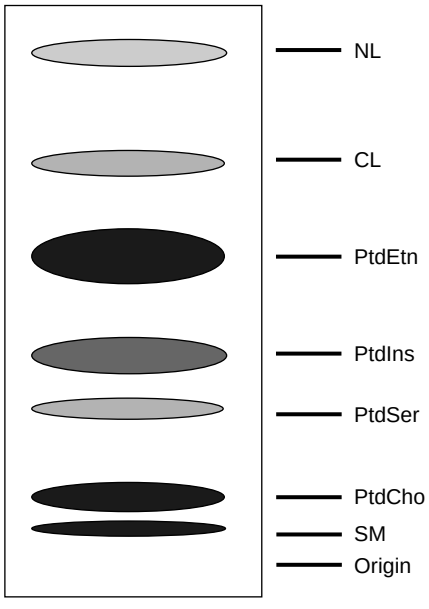

Figure 7 Separation of PtdIns and major cellular phospholipids by one-dimensional TLC.[87] NL, neutral lipid; other lipid classes as described in text. Adsorbent: precoated silica gel H plates (E. Merck). Solvent system: chloroform–methanol–acetic acid–water (50 : 37.5 : 3.5 : 2, v/v/v/v).

4.2.1.3 Atheroma Although the chemical composition of neutral lipids contained in the atherosclerotic plaque has been extensively investigated, relatively few studies have focused on the phospholipid content.[88] The most extensive study has been performed by Katz et al.[42] The techniques utilized for the determinations of the phospholipids were TLC with densitometry which indicated that atherosclerotic plaques contain close to 20% phospholipids. The value is dependent on the age of the plaque and the severity of the lesion. Table 4 shows that the major phospholipid classes present in the atherosclerotic plaque are PtdCho, SM and lysoPtdCho. PtdEtn and PtdIns are also in most atherosclerotic lesions and amount to 2–3% of the total lipid weight.

4.2.2 Phosphatidylinositol Phosphates

The PtdIns phosphates have been selected for special attention here because the usual chromatographic and mass spectrometric methods do not resolve or detect them because of the high polarity of these compounds and their low abundance in blood plasma and cells. This results in a partial repetition of the coverage in previous sections because some of the methods which detect the PtdIns phosphates also detect the other phospholipids.

4.2.2.1 Red Cells Although standard PtdIns phosphates from polymorphonuclear leukocytes (neutrophils from heparinized human blood) have been analyzed by matrix-assisted laser desorption/ionization (MALDI) MS,[89] there appear to have been no such attempts to analyze the phosphoinositides of red cells or platelets. The MALDI mass spectra of polyphosphoinositides with different degrees of phosphorylation were complicated by charge compensation. The content of polyphosphoinositides in stimulated neutrophils was judged too low for recognition of the anticipated ions.

Therefore, the analyses of phosphoinositide phosphates of red cells must rely on TLC and HPLC analyses. TLC on silica gel HL (Analtech, Inc., Newark, DE; not

Table 4 Phospholipid composition for newborn intima and each class of atherosclerotic lesion (wt% of total lipid)[42]

	Lyso PtdCho	SM	PtdCho	Total phospholipids
Newborn intima	0.0	14.5 ± 1.7	35.5 ± 4.7	71.4 ± 2.4
Fatty streak	0.3 ± 0.1	5.6 ± 0.1	4.8 ± 0.4	10.1 ± 1.0
Intermediate lesion	1.0 ± 0.2	11.0 ± 0.7	7.6 ± 0.7	19.6 ± 1.0
Fibrous plaque gruel	0.6 ± 0.1	11.7 ± 0.5	4.5 ± 0.5	16.8 ± 0.8
Gruel plaque	0.9 ± 0.2	10.1 ± 0.5	4.3 ± 0.4	15.3 ± 0.7

activated) with methanol–chloroform–water–concentrated ammonia (48:40:10:5, v/v/v/v) separated the inositides from origin and resolved the three radioactive lipids obtained by incubating the red cells with ^{32}P for 18–30 h.[90] Typical R_f values were PtdIns (4,5)P_2, 0.38; PtdIns (4)P, 0.46; lysoPtdCho, 0.56; PtdOH, 0.60; SM, PtdSer, PtdIns, and lysoPtdEtn, 0.60–0.66; PtdCho, 0.72, PtdEtn, 0.79; and cholesterol, 0.98.

A rapid two-dimensional TLC system which resolves the four phosphoinositide cycle phospholipids as well as all commonly encountered major and minor phospholipids has been described by Mitchell et al.[86] The samples are applied to TLC plates (10 cm × 10 cm), cut from Polygram 20 × 20 cm SIL H-HR plastic backed plates (Camb) and the plates developed first in chloroform–methanol–water–concentrated ammonia (48:40:7:5, v/v/v/v), and then at 90 °C in chloroform–methanol–formic acid (55:25:5, v/v/v). Recoveries are of the order of 85%.

Mallinger et al.[85] have utilized HPTLC plates for two-dimensional separation of red cell and platelet phospholipids including the PtdIns phosphates. Prior to being developed the plates were activated by heating to 100 °C for 15 min. Samples of lipid extract (approximately 2.5 μL) containing 1.8–2.0 μg of phosphorus were applied to HPTLC plates with a 10 μL Hamilton syringe. The plates were developed at room temperature. The first dimension was performed with chloroform–methanol–4.3 M NH_4OH (90:65:20, v/v/v) to resolve the polyphosphoinositides. The plates were then dried and developed a second time in the first dimension with chloroform–methanol–concentrated NH_4OH (130:50:10, v/v/v) in order to separate lysoPtdEtn from PtdCho. Then the plates were rotated 90° and developed in the second dimension with chloroform–methanol–acetic acid–water (100:30:35:3, v/v/v/v) to resolve the remaining phospholipid classes. The plates were then dried thoroughly and visualized and quantified by charring. PtdIns is the major inositol phospholipid in red cells, but appreciable quantities of the PtdIns (4)P and PtdIns (4,5)P_2 are also found (Table 2).

PtdIns P (PtdIns (3)P and PtdIns (4)P), PtdIns P_2 (PtdIns (3,4)P_2 and PtdIns (4,5)P_2) and PtdIns P_3 can clearly be separated by TLC, as noted above; the resolution of the D-3/D-4 isomers of PtdIns P and PtdIns P_2 is critical. A few protocols on this subject have been published. Some of them provide separation of PtdIns (3)P from PtdIns (4)P and others of PtdIns (3,4)P_2 from PtdIns (4,5)P_2. Hegewald[91] has described the first protocol for the separation of D-3 and D-4 PtdIns P lipids. The chromatography is performed on two different types of HPTLC plates under otherwise identical conditions and depends on the ability of the D-4, but not D-3, isomers to form complexes with boric acid. For this purpose precoated HPTLC plates of silica gel 60F254, Art 1.056641, and HPTLC plates NH_2, Art 1.12572, are used (both from Merck). Reference standards, ^{32}P-labeled PtdIns (4)P, PtdIns (4,5)P_2 and PtdOH were produced by in vitro incubation of human erythrocyte membranes with Mg-[γ-^{32}P]ATP and the lipids were extracted with hexane–2-propanol–HCl mixtures, essentially as described by Myher et al.[60] ^{32}P-Labeled PtdIns (3)P, PtdIns (3,4)P_2, and PtdIns (3,4,5)P_3 are synthesized by the incubation of a sonicated mixture of the erythrocyte lipids with Mg-[γ-^{32}P]ATP and the recombinant PtdIns-3-kinase γ. Hegewald et al.[92] impregnated the HPTLC plates by dipping them upside down into 5% boric aid (w/v) solution in methanol and then drying them for 5 min in an air current. Samples containing about 0.01 to 0.1 mg lipid are applied (CAMAG-Linomat IV, Switzerland) as 10 mm lanes, 10 mm above the bottom edge of the plate. The plates, which needed not to be activated, are then developed in 1-propyl acetate–2-propanol–absolute ethanol–6% aqueous ammonia (3:9:3:9, v/v/v/v), in a well equilibrated paper-lined twin-trough chamber (CAMAG). The mobile phase is allowed to reach the top of the plates (about 2 h). Following chromatography, the dried plates are dipped in charring reagent (5% $CuSO_4$ solution in 8% aqueous H_3PO_4) and heated at 180 °C for 15 min. The radioactivity of the ^{32}P-containing phospholipids is visualized and quantified using the GS-250 Molecular Imager System (Bio-Rad Laboratories, Hercules, CA).

4.2.2.2 Platelets The polyphosphoinositides (PtdIns P and PtdIns P_2) of platelets have been resolved by one-dimensional TLC using chloroform–methanol–ammonium hydroxide–water (45:35:6:4, v/v/v/v) as an effective developing solvent in combination with home-made silica gel H plates.[87] In this solvent system, PtdIns P, PtdIns P_2, lysoPtdIns and PtdOH are well resolved, but the separation of PtdIns from other phospholipids is poor. However, one-dimensional TLC separation can be readily used for the inositide separation if the study includes the use of [^{3}H]inositol-labeled lipids from platelets. The basic system containing chloroform–methanol–4M ammonium hydroxide (45:35:10, v/v/v) has been used to separate PtdIns, PtdIns P, and PtdIns P_2 from stimulated human platelets using oxalated silica gel H plates.[87] Using a chloroform–methanol–ammonium hydroxide–water (45:35:8:3, v/v/v/v) system, the following R_f values are obtained: PtdIns, 0.39; lysoPtdIns, 0.32; PtdIns P, 0.19; lysoPtdIns P, 0.07; and PtdIns P_2, 0.02. An increase in the water content from three to four parts in the solvent system will moderately raise the R_f values for PtdIns P and PtdIns P_2.

Mitchell et al.[86] used two-dimensional development on conventional TLC plates to separate the phospholipid

classes of platelets including the polyphosphoinositides as described above for the red cells, while Mallinger et al.[85] employed two-dimensional HPTLC for the resolution of the polyphosphoinositides of human platelets as described above for the red cells. Table 3 compares the relative composition of platelet membrane phospholipids obtained by the methods of Mitchell et al.[86] and Mallinger et al.[85] Excellent separations were obtained in both instances, although the isomeric PtdIns Ps were not resolved. Binder et al.[93] have described an HPLC method for the separation of polyphosphoinositides. A Bondapak NH_2 column (30×0.30 cm) is employed. Separation is carried out utilizing a 20 min isocratic elution with a 60 or 75 mM ammonium acetate–acetic acid buffer, pH 4.0, followed by a 120 min linear gradient to 2 M ammonium acetate–acetic acid, pH 4.0, at a flow rate of 1 mL min^{-1}. The PtdIns is the first one to be eluted from the column (15–20 min) followed by PtdIns P (40–55 min) and PtdIns P_2 (65–80 min). This procedure has been validated by monitoring changes in the phosphoinositides of human platelets stimulated with thrombin.

4.3 Molecular Species

The glycerophospholipids of plasma and cells contain nonidentical and nonrandom associations of fatty chains, attached by ester or ether linkages, with chain lengths of 16–22 carbons and up to six double bonds, whereas SM is made up of a separate pool of FAs and nitrogenous bases joined by amide linkages. This results in several hundreds of chemically distinct species of phospholipids.

In the past, the resolution of molecular species was obtained following an initial separation of the lipid classes, which are then examined separately after a chemical or enzymatic transformation, if necessary. This approach is still the best and can be effectively combined with HPLC and on-line ESMS. However, direct (flow) injection ESMS/MS of total phospholipid mixture saves time and frequently yields comparable results.

4.3.1 Glycerophospholipids

Until recently, when soft ionization MS became available, molecular species of blood glycerophospholipids, with few exceptions, were resolved and quantified by high-temperature GLC[20,58,59,60,94,95] or HPLC[16,62,70,71,87,93–97] following removal of polar head groups and trimethylsilylation or preparation of UV-absorbing derivatives, respectively. Soft ionization MS/MS now permits analysis of molecular species of most phospholipid classes without prior derivatization.

4.3.1.1 Plasma Lipoproteins A normal phase LC/ESMS of a total lipid extract yields separate peaks for different phospholipid classes, which can be examined by means of single ion mass chromatograms extracted from the total ion profile or by single ion monitoring, which is more sensitive. LC/ESMS currently provides the most effective method of determining the molecular species of glycerophospholipids.[21,98,99] Ravandi et al.[99] have used LC/ESMS for the determination of the molecular species of Etn glycerophospholipids of LDL. Table 5 tabulates the molecular species of the diradyl GroPEtn from human LDL. The molecular species were calculated from the total ion current profiles shown in Figure 6. Further accuracy in the LC/ESMS assay of molecular species of glycerophospholipids is obtained by substituting a reversed-phase for the normal-phase HPLC column. Pruzanski et al.[79,84] have used the normal-phase LC/ESMS method to identify and quantify the molecular species of phospholipids of normal and acute phase HDL. The Cho-containing glycerophospholipids were determined in the positive ion mode, while the Etn-containing and acidic phospholipids were determined in

Table 5 Molecular species of diradyl GroPEtn in LDL (mol%)[98]

Species	CN : DB[a]	*m/z*	Mean ± SD ($n = 4$)
Alkenylacyl			
16:0–20:4	36:4	722	4.25 ± 1.39
18:0–18:3	36:3	724	2.09 ± 0.97
16:0–20:3			
16:0–22:6	38:6	746	5.51 ± 1.86
18:2–20:4			
18:0–20:5	38:5	748	10.75 ± 2.14
16:0–22:5			
18:0–20:4	38:4	750	19.30 ± 4.53
18:0–20:3	38:3	752	3.63 ± 0.23
18:0–22:6	40:6	774	3.45 ± 1.87
Total			48.98 ± 3.37
Diacyl			
16:0–18:2	34:2	714	3.84 ± 1.24
16:0–18:1	34:1	716	1.20 ± 0.25
16:0–20:4	36:4	738	3.45 ± 1.32
18:0–18:3	36:3	740	2.29 ± 1.68
18:0–18:2	36:2	742	8.11 ± 3.17
18:0–18:1	36:1	744	1.99 ± 0.79
16:0–22:6	38:6	762	5.51 ± 2.43
18:2–20:4			
18:0–20:5	38:5	764	6.10 ± 1.81
16:0–22:5			
18:0–20:4	38:4	766	12.36 ± 3.92
18:0–20:3	38:3	768	1.89 ± 0.65
18:0–22:6	40:6	790	1.41 ± 1.64
18:0–22:5	40:5	792	2.88 ± 0.85
Total			51.00 ± 4.34

Lipids were extracted and different phospholipid classes were separated by normal phase silica column HPLC and resolved peaks analysed by on-line ESMS.
[a] CN : DB, acyl carbon number : double bond number.

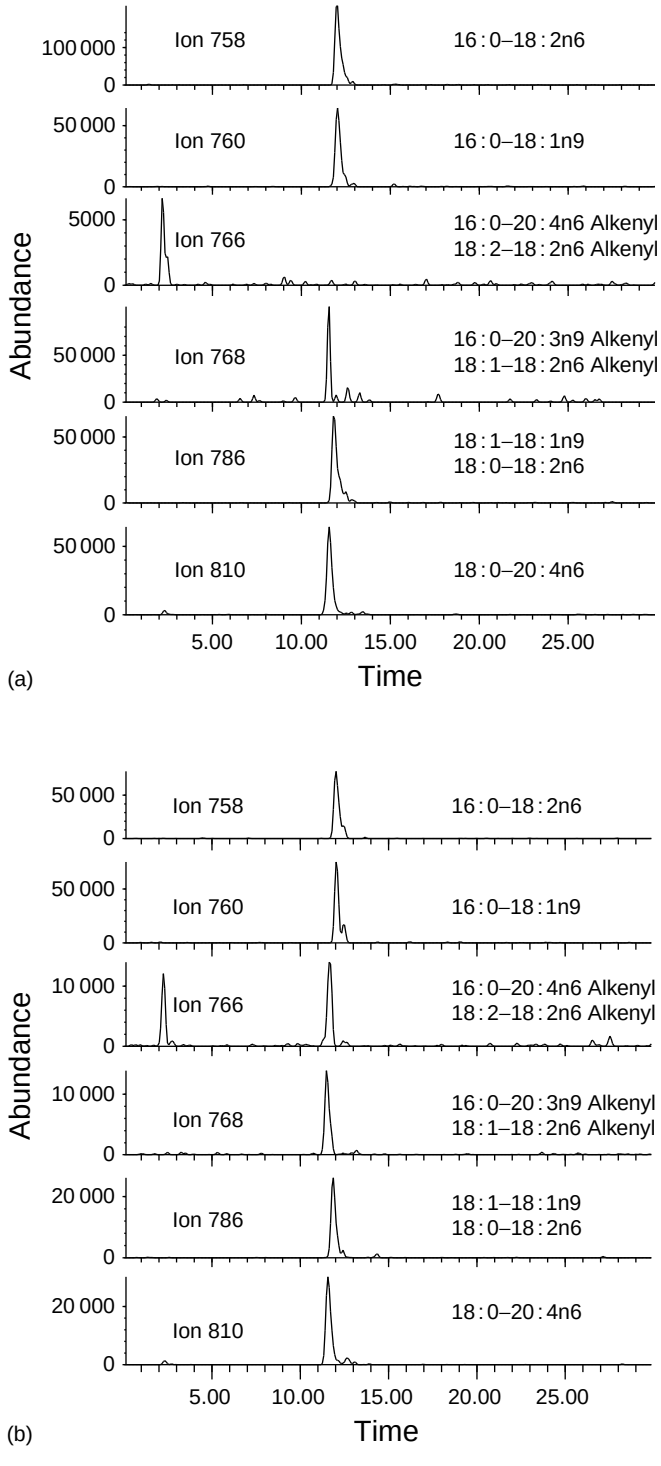

Figure 8 Single ion chromatograms of molecular species of residual PtdChos of acute phase HDL (APHDL) hydrolyzed with secretory group IIA phospholipase A_2 at 0 h (a) and 4 h (b) of hydrolysis.[79] Peak identification is given. LC/ESMS conditions are as given in Figure 6. Single ion chromatograms were retrieved from the total ion spectra by computer. FA abbreviations are as described in the text. Alkenyl designates the vinyl ether species of PtdCho. (Reproduced with permission from Pruzanski et al.[79])

the negative ion mode. Figure 8 shows the single ion mass chromatograms of molecular species of residual PtdCho of acute phase HDL hydrolyzed with secretory group IIA phospholipase A_2 at 0 (Figure 8a) and 4 (Figure 8b) h of hydrolysis. For quantification of the different molecular species an internal standard is included and the response is calibrated for each phospholipid class. Table 6 compares the molecular species of the PtdCho of normal human HDL and acute phase HDL. Only minor differences in the molecular species were noted despite marked differences in the composition of the associated proteins.[84]

Prior to chromatographic and mass spectrometric resolution of intact phospholipids, the molecular species of glycerophospholipids and SMs had been determined by high-temperature GLC of the TMS and TBDMS ethers of the diacylglycerol moieties released by phospholipase C.[59,60] The purified TBDMS ethers are resolved by carbon number using a nonpolar column (8 m × 0.32 mm ID fused silica capillary) coated with a cross-linked 5% phenylmethylsilicone (Hewlett-Packard). The sample is injected on-column and the temperature is programmed in four steps from 40 to 350 °C: 40 to 150 °C at 30 °C min^{-1}; then to 230 °C at 20 °C min^{-1}; then to 280 °C at 10 °C min^{-1}; and to 340 °C at 5 °C min^{-1}. The carrier gas is H_2 at 6 psi (1 psi = 6.9 kPa) head pressure. The purified TBDMS ethers are resolved into molecular species by GLC on a polar column (15 m × 0.32 mm capillary coated) with SP2380 (Supelco, Oakville, Ontario, Canada), which is operated isothermally (260 °C) or is temperature programmed from 240 to 260 °C at 5 °C min^{-1} with hydrogen as carrier gas at 2 psi head pressure. These analyses have provided the most extensive compilation of the analytical data for total

Table 6 Molecular species of control (HDL) and acute phase (APHDL) PtdCho[84] (mean ± SD, $n = 8$)

PtdCho species	HDL[a] (mol%)	APHDL[b] (mol%)
16:0–16:1	1.20 ± 0.21	1.8 ± 0.37
16:0–16:0	2.54 ± 0.35	2.4 ± 0.69
16:0–18:2	28.32 ± 3.53	25.9 ± 2.64
16:0–18:1	13.38 ± 1.47	19.7 ± 0.90
16:0–20:5	0.49 ± 0.32	0.3 ± 0.10
16:0–20:4	10.93 ± 2.77	11.6 ± 1.84
18:1–18:2	10.09 ± 1.86	8.5 ± 1.48
18:1–18:1	13.03 ± 1.31	12 ± 2.65
16:0–22:6	4.38 ± 0.84	3.3 ± 0.76
16:0–22:5	5.88 ± 1.74	3.9 ± 1.32
18:0–20:4	9.74 ± 1.97	9.5 ± 0.60
18:0–22:6	2.43 ± 0.77	1.2 ± 0.75

[a] HDL isolated by sequential ultracentrifugation from normolipidemic subjects in fasting state.
[b] APHDL isolated from eight pools of acute phase plasma from patients in various states of fasting.

plasma lipoproteins.[60] Qualitatively the data for total plasma are identical to the data for individual plasma lipoproteins, but quantitatively significant differences are noted. Figure 9 shows the elution patterns obtained on polar capillary GLC for the TMS ethers of the alkylacylglycerols derived from plasma alkylacyl GroPCho (Figure 9a) and alkylacyl GroPEtn (Figure 9b) respectively. Although both Cho and Etn GPL contain nearly identical proportions of 16:0′–20:4 and 18:0′–20:4 species, the alkylacyl GroPCho contain much more of the saturated, monoenoic and 18:1′–20:4 and 16:0′–22:4, and less of other polyunsaturated species than alkylacyl GroPEtn. Yang et al.[5] have used polar capillary GLC for the resolution of the molecular species of chylomicron PtdCho as the diacylglycerol TMS ethers from corn and fish oil feeding. Breckenridge and Palmer[54] compared the FA composition of the PtdIns among human plasma lipoprotein classes as an indication of the composition of molecular species. It is found that the PtdIns from VLDL, LDL and HDL gave the same FA composition,

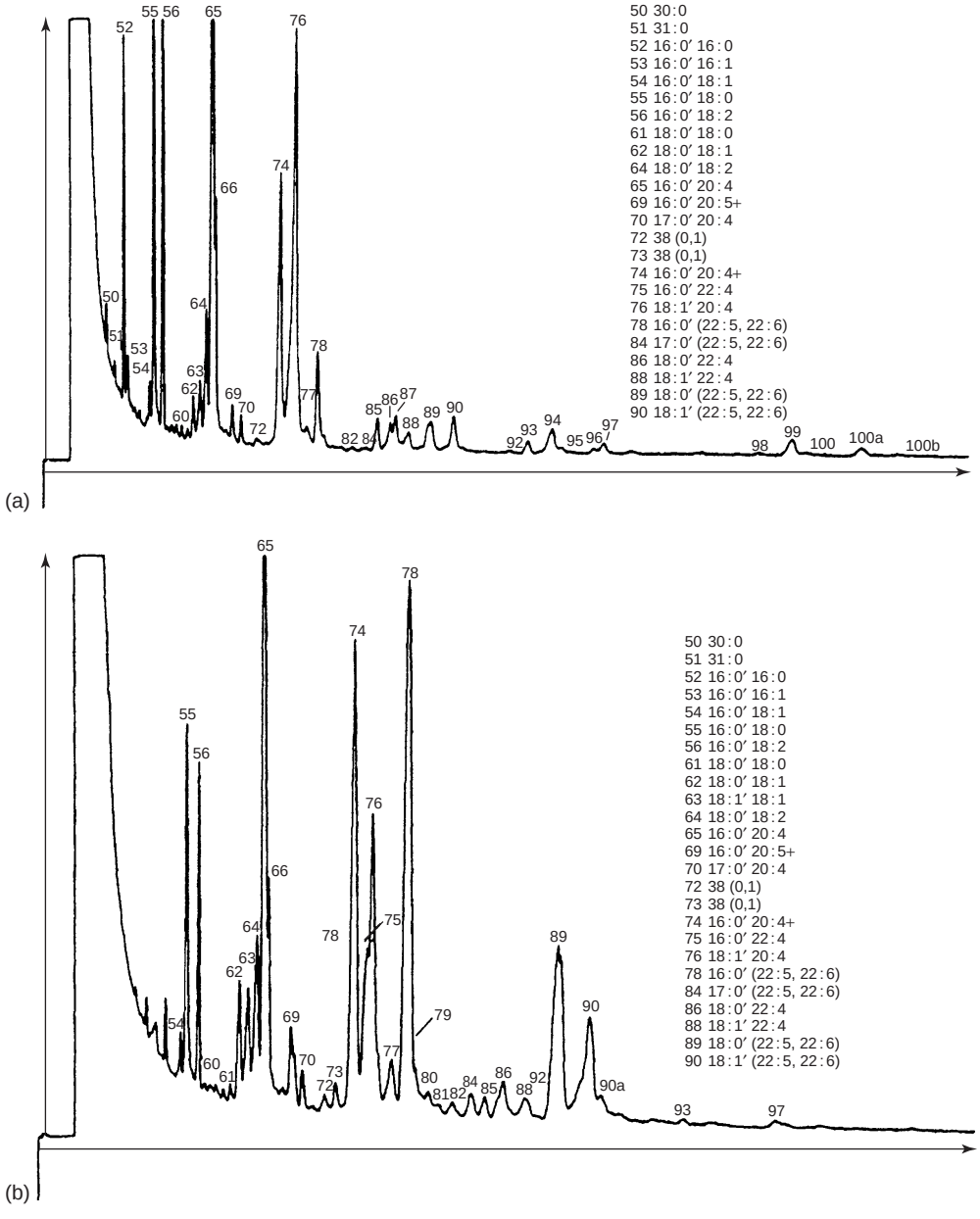

Figure 9 Polar capillary GLC profile of the alkylacylglycerol moieties of human plasma diradyl (a) GroPCho and (b) GroPEtn.[60] Peak identification is given. Sample: 1 mL of 0.1% diradylglycerol TMS ethers in hexane. GLC conditions: column, 15 m × 0.32 mm fused silica capillary coated with cross-bonded RTx 2330; carrier gas, H_2, 3 psi; temperature, 250 °C, isothermal. (Reproduced with permission from Myher et al.[60])

but had less 20:4n6 than platelet PtdIns and less 16:0 than erythrocyte PtdIns.

Myher et al.[4,59] have reported the molecular species of the SMs of plasma lipoproteins. The species were determined by GLC following dephosphorylation by phospholipase C and trimethylsilylation.

4.3.1.2 Red Cells and Platelets Han and Gross[72] have reported direct analysis of the molecular species of erythrocyte phospholipids by flow injection ESMS/MS. Quantitative analysis of individual molecular species from subpicomole amounts of human erythrocyte phospholipids are reported with a sensitivity two to three orders of magnitude greater than that achievable with fast atom bombardment mass spectrometry (FAB/MS). More than 50 phospholipid constituents were identified in human erythrocyte plasma membrane by direct ESMS of chloroform extracts derived from the equivalent of <1 μL of whole blood (Figure 10). Isobaric molecular species of both PtdCho and PtdEtn are identified and quantified by ESMS/MS. The analysis included estimates of the three major Cho phosphatide subclasses: plasmenylcholine, plasmanylcholine and PtdCho. Kerwin et al.,[55] however, observed that ESMS is not capable of differentiating in all instances between alkylacyl and alkenylacyl substituents without prior separation of these lipid subclasses.

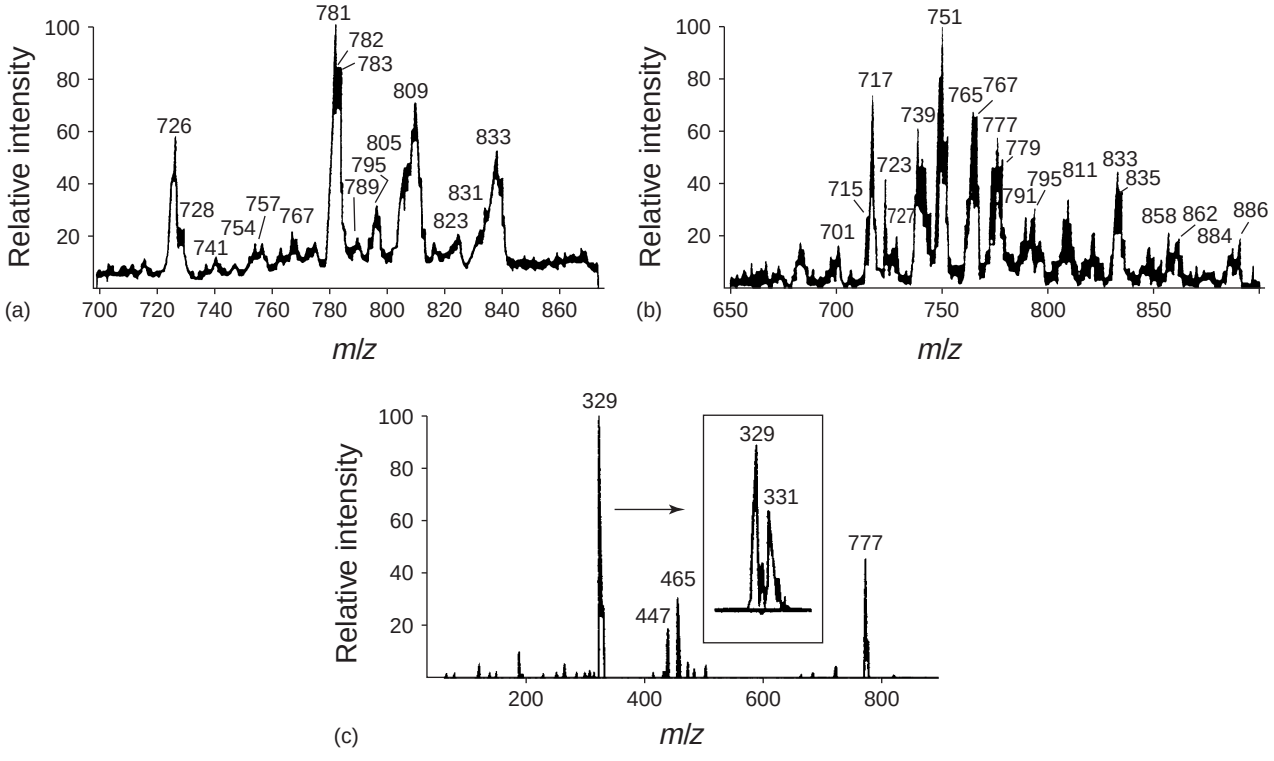

Figure 10 Direct ESMS analysis of human erythrocyte plasma membrane phospholipids.[72] (a) A positive-ion electrospary ionization mass spectrum of erythrocyte plasma membrane phospholipid extract (containing 13.5 pmol of Cho glycerophospholipids from 25 nL of whole blood) shows 14 molecular species of Cho glycerophospholipids, e.g. *m/z* 757, 781, 783, 805, 809, and 833, corresponding to sodiated 16:0–16:0, 16:0–18:2, 16:0–18:1, 16:0–20:4, 18:0–18:2 (18:1–18:1) and 18:0–20:4 GroPCho, respectively, as well as *m/z* 767, 795 and 823, corresponding to sodiated 16:0–18:1, 18:0–18:1 and 18:0–20:1 plasmenyl Cho, respectively, and four molecular species of SM, e.g. *m/z* 726, 754 and 782, corresponding to 16:0, 18:0 and 20:0 amides. (b) A negative-ion electrospary ionization mass spectrum of the same extract of plasma membrane phospholipids shows >25 molecular species of Etn glycerophospholipids, predominantly composed of plasmalogen molecular species, e.g. *m/z* 751, 723, 749, 775, 777, and 779, corresponding to 18:0–20:4, 16:0–20:4, 18:1–20:4, 18:0–22:6 (18:1–22:5), 18:0–22:5 (18:1–22:4) and 18:0–22:4 (20:0–20:4) plasmenyl Etn, respectively. In addition, PtdEtn species were detected at *m/z* 715, 717, 739, 765, 767 and 793, corresponding to 16:0–18:2, 16:0–18:1, 16:0–20:4, 18:1–20:4, 18:0–20:4 (16:0–22:4) and 18:0–22:5 (18:1–22:4), respectively. Molecular species of other acidic phospholipids were seen at *m/z* 811, 833 and 886, corresponding to 18:0–20:4 GroPSer, 18:1–22:6 GroPSer, and 18:0–20:4 GroPIns, respectively. (c) ESMS/MS analysis of isobaric Etn glycerophospholipids (*m/z* 777) extracted from human erythrocyte plasma membranes with selection and CID of *m/z* 777 after negative-ion electrospray ionization demonstrates two carboxylic anions (*m/z* 329 and 331) (see Inset), corresponding to 22:5 and 22:4 FAs, respectively. The sample, in 30 μL of chloroform–methanol (1:2, v/v) in the presence of excess NaOH (i.e. NaOH/lipid molar ratio > 1), was infused directly into the electrospray chamber with a syringe pump at a flow rate of 1.5 μL min^{-1} for mass analyses in both negative- and positive-ion modes. (Reproduced with permission from Han and Gross.[72])

Furthermore, ESMS did not provide information on the position (*sn*-1 or *sn*-2) of FAs, although such differentiation has occasionally been observed.[56]

Furthermore, the major molecular species of the diacyl GroPCho compare favorably to those reported in earlier GLC analyses,[94,100] while those of the diradyl GroPEtn are not unlike those assessed by Hullin et al.[70]

Hullin et al.[16] have obtained distinct HPLC profiles for the trinitrophenyl derivatives of human red blood cell GroPEtn before and after acidolysis, which destroys the plasmalogens. Using this method, Hullin et al.[16] are able to separate and quantify the molecular species of both subclasses (diacyl and alkenylacyl) from human red blood cells and rat brains with a lower limit of detection of about 10 pmol. The method provides a useful addition to the repertoire of analytical procedures for the determination of the molecular species, as well as for the study of molecular species asymmetry in biological membranes. PtdEtn and PtdSer, after N-derivatization to improve their chromatographic properties, have been subjected to mass spectrometric examination with thermospray following reversed-phase HPLC.[16] This compares to over 200 species determined by capillary GLC of the diradylglycerol TMS ethers following dephosphorylation.[59]

The most exhaustive analyses of the molecular species of the glycerophospholipids of red cells have been performed by polar capillary GLC on the alklyacyl, alkenylacyl and diacylglycerol moieties as the corresponding TMS ethers.[59] Figure 11 shows the polar capillary GLC profiles of PtdIns and PtdSer of the red

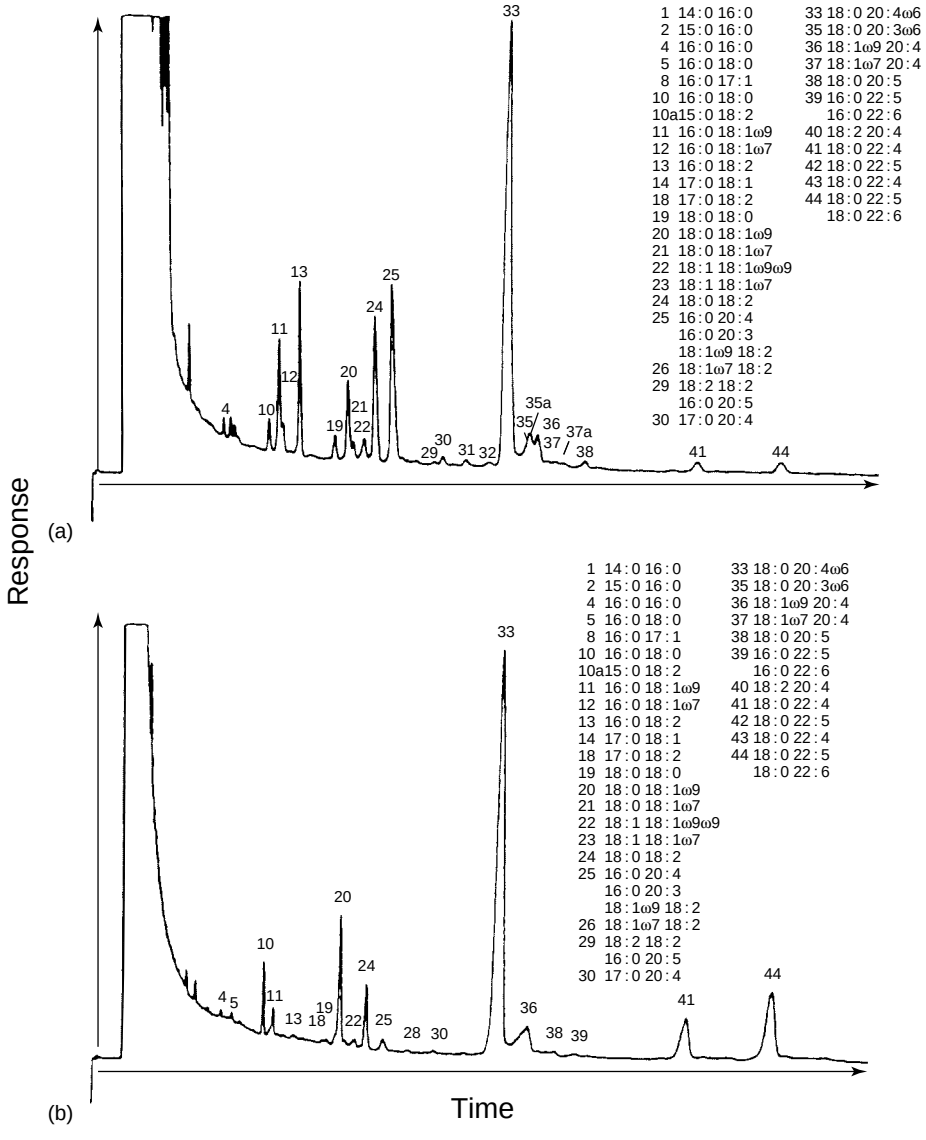

Figure 11 Polar capillary GLC profile of the diacylglycerol moieties of human red blood cell diacyl (a) GroPIns and (b) GroPSer. Peak identification is given. Conditions as in Figure 9. (Reproduced with permission from Myher et al.[59])

blood cells, which contain only the diacylglycerol subclass. The quantitative composition of the molecular species of red cell diacyl glycerophospholipids is given in Table 7 along with that of the diacyl GroPCho and GroPEtn. There are marked differences between the major and minor molecular species among the different glycerophospholipid classes. Thus, while the diacyl GroPCho and diradyl GroPEtn contain closely similar amounts of 16:0–18:1n9 (18.0 vs 15.1%), 18:0–18:1n9 (5.8 vs 4.3%) and 16:0–20:3n6 (5.2 vs 3.5%), they differ greatly in the content of 16:0–18:2n6 (26.9 vs 6.2%), 18:0–18:2n6 (11.2 vs 2.1%), 16:0–20:4n6 (5.1 vs 13.5%), 18:0–20:4n6 (3.9 vs 12.6%) and 18:1n9–20:4n6 (0.6 vs 8.4%). In contrast, the diacyl GroPIns and GroPSer contained closely similar amounts of 18:0–20:4n6 (18:0–20:3n6) (45.3 vs 47.4%), 18:0–18:1n9 (3.6 vs 5.9%) and 18:1–20:3 + 18:1t–20:4 (4.5 vs 4.9%), but differed in the content of 18:0–22:4n6 (2.3 vs 11.3) and 18:0–22:5 plus 18:0–22:6n3 (2.8 vs 21.0).

Table 7 Molecular species of diacylglycerophospholipids of erythrocytes[59] (in mol%)

Molecular species	PtdCho	PtdEtn	PtdIns	PtdSer
14:0–16:0	0.4			
15:0–16:0	0.1			
16:0–16:0	4.6	0.6	0.4	0.1
16:0–16:1	0.9	0.4	0.6	0.1
16:0–17:0	0.2			
16:0–18:0	2.1	0.3	1	2.1
16:0–18:1*w*9	18	15.1	3.8	1.3
16:0–18:1*w*7	2.9	1	0.9	
16:0–18:2	26.9	6.2	5.5	0.2
17:0–18:1	0.4	0.3		
17:0–18:2	0.3	0.3		
18:0–18:0	0.6	0.1	1.1	
18:0–18:1*w*9	5.8	4.8	3.6	5.9
18:0–18:1*w*7	0.3	0.2	0.8	
18:1*w*9–18:1*w*9	1.8	4.6	1.4	0.4
18:19–18:17		0.5		
18:0–18:2		2.1	6.6	2.9
16:0–20:4	11.2	13.5	7.7	0.7
16:0–20:3	5.1	3.5	2.7	
18:1*w*9–18:2	5.2			
18:1*w*7–18:2	0.6			
18:2–18:2	1	0.9	0.3	
16:0–20:5				
17:0–20:4	0.2	0.4	0.6	
18:0–20:4*w*6	3.9	12.6	45.3	47.3
18:0–20:3*w*6				
16:0–22:4*w*6	0.5	5.2		
18:1–20:3 + 18:1t–20:4			4.5	4.9
18:1*w*9–20:4	0.6	8.4	2.7	
18:1*w*7–20:4	0.2	0.9		
18:0–20:5*w*3	0.3	0.6		
16:0–22:5*w*3	2	4.7	0.8	0.3
16:0–22:5*w*3				
18:2–20:4	0.3	1.8	0.8	
18:0–22:4*w*6	0.4	1.9	2.3	11.3
18:0–22:5*w*6		0.7		
18:1*w*9–22:4*w*6		1.3		
18:0–22:5 + 18:0–22:6	1	2	2.8	21
18:2–22:4		0.2		
18:1*w*9–22:5 + 18:1–22:6		1.6		

Detailed GLC analysis has also been performed on alkylacylglycerols derived from the erythrocyte diradyl GroPCho and GroPEtn showing marked differences in both qualitative and quantitative composition, which indicates that the alkylacylglycerols represent different pools of molecular species. While the alkylacyl GroPCho and GroPEtn contain nearly identical proportions of 16:0′–20:4 and 18:1′–20:4 species, the alkylacyl GroPCho contains much more of the saturated and monoenoic and less of the 18:0′–20:4, 16:0′–22:4, 18:0′–22:4 and other polyunsaturated species than alkylacyl GroPEtn.

There are significant differences between the two lipids in the composition of the alkenylacyl species indicating the existence of separate pools of precursors or independent transformation mechanisms. The alkenylacyl GroPCho is much richer in species in which a 16:0″ alkenylglycerol is combined with saturated, monoenoic, dienoic and tetraenoic FAs, whereas alkenylacyl GroPEtn is much richer in 18:0″ alkenylglycerol combined with dienoic, tetraenoic and hexaenoic FAs. The positional distribution of the fatty chains in the individual diacyl GroPCho and GroPEtn is not independently determined: thus, the relative proportions of the reverse isomers have not been estimated. The phospholipase A_2 data are available in the literature[101] and we have taken advantage of them in compiling the tabulated material.

In contrast to red cells, Schick et al.[102] found that the molecular species of the Etn phospholipids of human platelet plasma membrane cell surface did not differ from those of the inside, except in regard to 18:2n6.

Han et al.[76] have applied the flow injection ESMS method to the determination of individual molecular species of human platelet phospholipids. Figure 12 compares the negative-ion electrospray ionization mass spectra of PtdEtn and related lipids in resting and thrombin-stimulated human platelets. The control platelet phospholipids demonstrated that alkenylacyl glycerophosphoethanolamine (PlsEtn) is the predominant Etn glycerophospholipid subclass in human platelets [peaks at *m*/*z* 723 (16:0–20:4 and 18:2–18:2 PlsEtn), 749 (18:1–20:4 and 16:0–22:5 PlsEtn), 751 (18:0–20:4 and 16:0–22:4 PlsEtn), and 779 (18:0–22:4 and 20:0–20:4 PlsEtn)]. The thrombin-stimulation resulted in a 53% decrease in mass during 90 s resulting in the liberation of 20.1 nmol of arachidonic acid/10^9 platelets from the PlsEtn pool (Table 8). Thus, the most abundant PlsEtn species 18:0–20:4 underwent a 57% decrease in mass. The quantitative accuracy of human platelet anionic phospholipid determinations is substantiated by

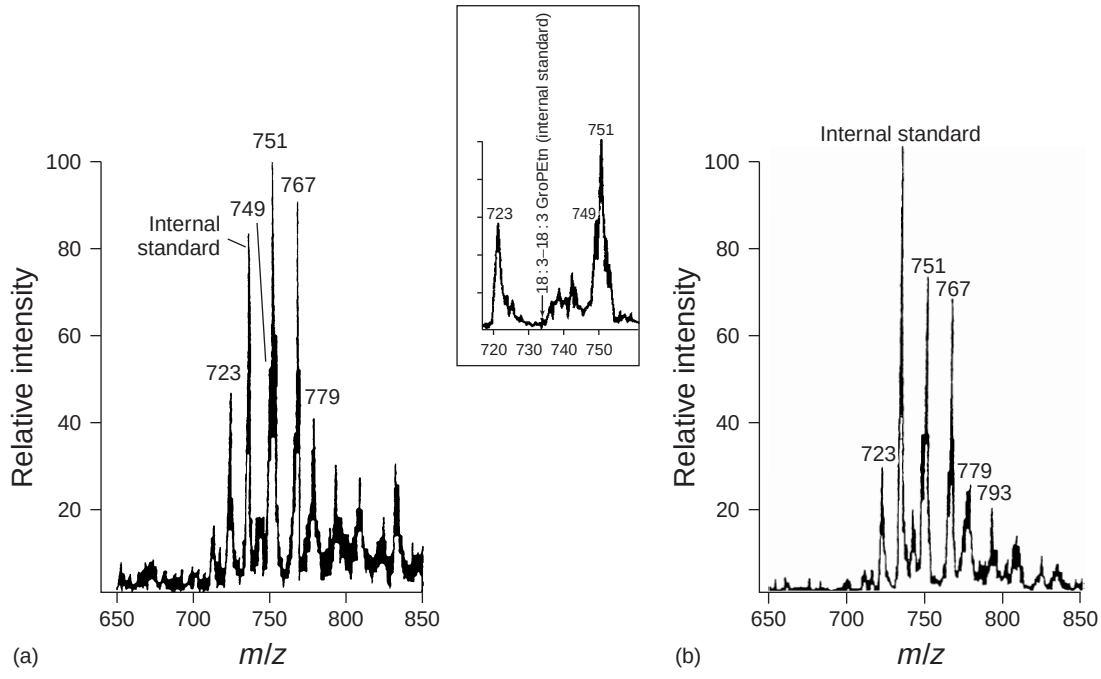

Figure 12 Negative ion electrospray ionization mass spectra of Etn glycerophospholipids in resting and thrombin-stimulated human platelets.[76] (a) A negative-ion electrospray ionization mass spectrum of membrane extract from resting human platelets was obtained in the presence of NaOH in chloroform–methanol (1 : 2, v/v). Individual Etn glycerophospholipid molecular species were identified utilizing ESMS/MS and are listed in Table 8. The internal standard is 18 : 3–18 : 3 PtdEtn. (b) A negative-ion electrospray ionization mass spectrum of membrane extract from thrombin-stimulated human platelets was acquired under identical conditions. The insert is an expanded mass spectrum of the region around *m*/*z* 735 obtained from platelet extracts which did not contain exogenously added 18 : 3–18 : 3 PtdEtn which demonstrates that no endogenous molecular ions are present in this region. (Reproduced with permission from Han et al.[76])

co-addition of known amounts of either 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoserine or bovine liver PtdIns with dimyristoylGroPSer as internal standard in chloroform–methanol (1 : 1, v/v) to the platelet suspension prior to Bligh and Dyer extraction and ESMS analyses.[76] A comparison of the MS/MS estimates to the estimates obtained by HPLC revealed great discrepancies apparently due to incomplete resolution of molecular species under the selected HPLC conditions, which led to inappropriate summing of the molecular species for comparison to MS/MS data. The alkenylacyl molecular species may be distinguished from alkylacyl GroPEtn molecular species by treating human platelet phospholipid extracts with acidic vapors prior to mass spectroscopic analyses as described previously.[71,72]

Kim and Salem, Jr[47] have reviewed the early applications of LC/MS of lipids, including references to plasma lipoproteins and red blood cells. Thermospray spectra of plasmalogens contain prominent monoacylglyceryl fragments resulting from a cleavage of the alkenyl moiety at the *sn*-1 position from the diradylglycerol fragments. The alkylacyl GroPCho, like PtdCho, yields diacylglycerol-like fragments as the base peaks. While a PAF molecule produced diacylglyceryl fragments, head group and molecular ions, 1-acyl analogs also produced monoacylgyceryl fragments including a peak at *m*/*z* 117 which arose from the hydrolysis of the 1-acyl moiety from the diacylglyceryl fragment.

Earlier effective chromatographic resolution of molecular species of intact glycerophospholipids was reported by Patton et al.[96] using reversed-phase HPLC. Patton et al.[97] applied this method to the separation of the molecular species of the glycerophospholipids of rat plasma HDL, and rat lymph HDL and VLDL. Since the various diradylglycerol species are not separated, the diacyl, alkylacyl and alkenylacyl GroPCho and GroPEtn subclasses must be first resolved by some other means (see above). A comparison between the total and the diacylglycerol subclasses is possible, however, as the latter may be recovered along with the alkylacylglycerol species following treatment with acid vapor.[71,72] The molecular species are detected by short-wave UV (205 nm).

4.3.2 Sphingomyelins

Like the molecular species of the glycerophospholipids, those of the SMs in the past were determined by GLC following removal of the polar head group and

Table 8 Molecular species of platelet Etn phospholipids before and after stimulation as estimated by ESMS and HPLC analysis (nmol/10^9 cells)[76]

Species[a]	*m/z*	Resting		Stimulated		% change	
		ESMS (mean ± SD)	HPLC[b]	ESMS (mean ± SD)	HPLC[b]	ESMS	HPLC
P16:0–18:2	699	2.2 ± 0.2		1.9 ± 0.2		−13.7	
P16:0–18:1	701	1.5 ± 0.2		0.9 ± 0.2		−40	
P18:1–16:0	701	0.7 ± 0.2		0.4 ± 0.1		−42.9	
D16:0–18:2	715	2.3 ± 0.4		2 ± 0.3		−13.1	
D16:0–18:1	717	2.3 ± 0.2	4.3 ± 0.7	1.3 ± 0.2	1.9 ± 0.1	−43.5	−55.9
P16:0–20:4	723	7.7 ± 0.7	19.2 ± 2	3.5 ± 0.3	10.2 ± 0.2	−54.6	−46.9
P18:2–18:2	723	3.9 ± 0.7		1.8 ± 0.3		−53.9	
D16:1–20:4	737	2.1 ± 0.6		1.8 ± 0.3		−14.3	
D16:0–20:4	739	5.3 ± 0.4	9.5 ± 0.8	3.1 ± 0.6	4.3 ± 0.5	−41.6	−54.8
D18:1–18:2	741	3.8 ± 0.4		2.6 ± 0.4		−31.6	
D18:1–18:1	743	5.9 ± 0.4		3.3 ± 0.5		−44.1	
D18:0–18:1	745	2.9 ± 0.4		1.7 ± 0.3		−41.4	
P18:1–20:4	749	10.4 ± 0.8	8 ± 1	5.3 ± 0.3	4.8 ± 0.0	−49.1	−40.0
P16:0–22:5	749	2.1 ± 0.8		1.1 ± 0.2		−47.7	
P18:0–20:4	751	17.6 ± 1.2	26 ± 3.1	7.6 ± 0.9	11.8 ± 0.5	−56.9	−54.7
P16:0–22:4	751	5.9 ± 1.2		2.5 ± 0.9		−57.7	
D16:0–22:6	763	3.6 ± 0.4		2.9 ± 0.5		−19.5	
D18:1–20:4	765	5.8 ± 0.4	4.2 ± 0.3	3.3 ± 0.3	2.3 ± 0.6	−43.2	−45.3
D16:0–22:5	765	2.9 ± 0.4		1.6 ± 0.3		−44.9	
D18:0–20:4	767	14.6 ± 1.1	28.7 ± 3.1	7.9 ± 0.8	18.2 ± 3.8	−45.9	−36.6
D16:0–22:4	767	7.3 ± 1.1		3.9 ± 0.8		−46.6	
P18:0–22:6	775	2 ± 0.2		1.1 ± 0.2		−45	
P18:1–22:5	775	2 ± 0.2		1.1 ± 0.2		−45	
P18:1–22:4	777	3.5 ± 0.4		2.1 ± 0.3		−40	
P18:0–22:5	777	3.4 ± 0.4		2.1 ± 0.3		−38.3	
P18:0–22:4	779	6.7 ± 0.5		4.6 ± 0.4		−31.4	
P20:0–20:4	779	2.3 ± 0.5		1.5 ± 0.4		−34.8	
D18:1–22:5	791	2.3 ± 0.2		1.8 ± 0.3		−21.8	
D18:1–22:4	793	2.1 ± 0.3		1.4 ± 0.3		−33.4	
D18:0–22:4	795	2.6 ± 0.5		2.2 ± 0.5		−15.4	
Total		137.7	99.9	78.3	53.5	−43.2	

[a] P represents plasmalogenic species, and D represents diacyl species.
[b] See text for explanation of discrepancy in quantitative estimates between ESMS and HPLC data.

silylation.[4,20,59,60] Soft ionization MS now permits analysis of molecular species of most phospholipid classes without prior derivatization.

4.3.2.1 Plasma Lipoproteins The first detailed analyses of the molecular species of SMs of human plasma lipoproteins were made by Myher et al.,[4] who used high-temperature GLC and trimethylsilylation to resolve the ceramide moieties released by phospholipase C. The molecular species of the di-TBDMS ethers of ceramides were determined by gas chromatography/mass spectrometry (GC/MS)[4] using nonpolar columns. The molecular species are first resolved by carbon number by GLC and the species within each carbon number are determined by GC/MS. The composition by molecular weight was determined from the $[M - 57]^+$ ions. Much improved resolution of the molecular species of the ceramides released from the SMs by phospholipase C is obtained using polarizable capillary columns.[103]

Pruzanski et al.[79,84] have determined the molecular species of HDL SMs using normal-phase LC/ESMS. In one study[79] detailed analyses of the species was used to demonstrate resistance control and acute phase HDL SM to hydrolysis with secretory phospholipase A_2. In the other,[84] detailed comparisons were made between molecular species of control and acute phase HDL SMs. There was a close agreement between the present estimates obtained by LC/ESMS and the earlier GLC analyses of the ceramide moieties of SMs of control HDL.[4]

4.3.2.2 Red Cells and Platelets The first detailed analyses of the molecular species SMs of red cells were also made by nonpolar capillary GC/MS of the ceramide moieties. Much improved resolution and identification of

the ceramide moieties was obtained by polarizable capillary GC/MS of the TBDMS ether derivatives.[103] It must be noted, however, that human erythrocyte membranes have been found to contain free ceramides[104] (5.6 μmol/100 mL of packed cells). The free ceramide is constituted mainly of 16:0, 22:0, 24:0 and 24:1 nonhydroxy FAs and of 18:1 sphingosine.

Han and Gross[72] used flow ESMS and ESMS/MS for the determination of molecular species of human erythrocyte plasma membrane SM. The sensitivity of ESMS is two to three orders of magnitude greater than that achievable with FAB/MS. Multiple molecular species of SM are demonstrated, including species at *m/z* 726, 754, and 782, corresponding to the 16:0, 18:0 and 20:0 amides of sphingosine.

Kerwin et al.[55] have reported the positive and negative ion ESMS and ESMS/MS for commercial preparations of SM. The samples are analyzed by flow ESMS. Molecular ion adducts are the primary products formed by positive ionization, e.g. $[M + H]^+$, $[M + Na]^+$ and others. The negative spectra contained the ions corresponding to the loss of these groups. Positive ion MS/MS of $[M + H]^+$ ions yielded a strong peak at *m/z* 184 corresponding to the phosphocholine head group. In negative ion MS/MS spectra, the major ion found is the result of a loss of acetate from the parent ion described above. Less intense peaks are due to the loss of acetate and an additional CH_3 from the head group. Attempts to increase fragmentation to generate peaks characteristic of the N-acyl and sphingenine/sphinganine constituents using 1% methanolic formate are not successful.

An earlier method for the separation of molecular species of SM by reversed-phase HPLC has been described by Jungalwala et al.[105] SM species from bovine brain and sheep and pig erythrocytes are resolved into 10–12 separate peaks on a μBondapak C_{18} reversed-phase column with methanol–5 mM potassium phosphate buffer, pH 7.4, 9:1 (v/v) as the solvent. Detection is by UV at 203–205 nm. The molecular species are identified by peak collection and determination of the FA and nitrogenous base composition. Jungalwala et al.[105] combined HPLC with MS for improved identification and quantification of the molecular species of the resolved SM.

4.3.2.3 Atheroma The most extensive analysis of atherosclerotic phospholipids has been carried out by Ravandi et al.[98] Utilizing an LC/ESMS system many of the molecular species in each class of phospholipids were identified. In Table 9 over 22 species of Etn phospholipids were identified in which 64.89% consisted of alkenylacyl and 35.11% diacyl species. Although there are differences among species, there is an overall similarity in the composition of PtdCho and SM between LDL and atherosclerotic plaques. Other workers have reported the FA profiles of total lipid and total phospholipid fractions from atheroma.[26,28,42]

Table 9 Molecular species of diradyl GroPEtn from atherosclerotic tissue[a] (mol%)[98]

Species	Alkenylacyl (mean ± SD)	Diacyl (mean ± SD)
16:0–18:2	0.53 ± 0.26	1.24 ± 0.25
16:0–18:1		1.58 ± 0.47
18:0–18:2		2.31 ± 1.25
16:0–20:4	10.54 ± 4.18	
18:0–18:3	2.18 ± 0.83	
16:0–20:3		
16:0–22:6	2.31 ± 1.14	8.22 ± 0.54
18:2–20:4		
18:0–20:5	16.27 ± 3.39	0.64 ± 0.21
16:0–22:5		
18:0–20:4	28.09 ± 4.57	18.34 ± 2.57
18:0–20:3	1.36 ± 0.32	
18:0–22:5		1.15 ± 0.48
18:0–22:6	3.63 ± 0.81	1.63 ± 0.33
Total	64.91 ± 4.79	35.11 ± 4.23

[a] Lipid extracts of atherosclerotic tissue ($n = 6$) analyzed by LC/ESMS.

5 NOVEL PHOSPHOLIPID SUBCLASSES

The recent development of soft ionization mass spectrometric techniques has allowed the isolation and identification of various complex derivatives of the common glycerophospholipids. Analyses of these molecules had been difficult or impossible by conventional chromatographic and mass spectrometric methods because of instability and high polarity of the molecules. The modified compounds include the glycation products of the aminophospholipids, oxidation products of glycerophospholipids, and phosphoinositide glycans. The polyphosphoinositides, which would also belong to this group of compounds because of high polarity and instability, have been considered along with the common phospholipid classes because of their long-standing recognition.

5.1 Glycated Aminophospholipids

5.1.1 Plasma Lipoproteins

Nonenzymatic glycosylation is the post-translational modification of a protein by the covalent attachment of a sugar residue, which results in a spontaneous amino–carbonyl bonding referred to as a Schiff base linkage. Recently it has been shown that aminophospholipids are also susceptible to this reaction and glycated aminophospholipids can be isolated from in vivo sources.[78,99] The complex lipid mixtures require chromatography and MS for complete determination of the

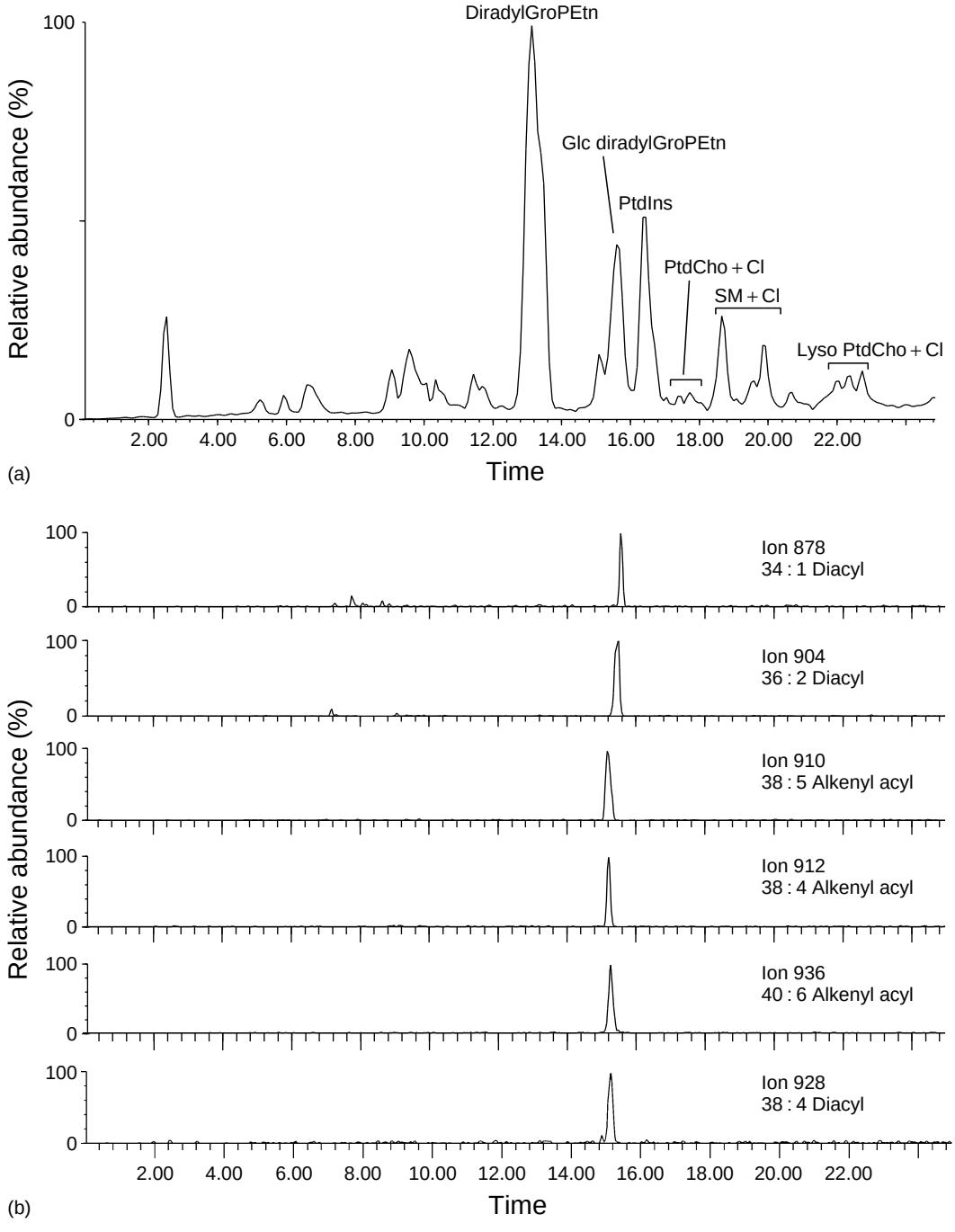

Figure 13 LC/ESMS analysis of total lipid extract of LDL incubated with 50 mM glucose for 7 days.[98] (a) Total negative ion current profile of LDL phospholipids; (b) single ion plots for the major glucosylated PtdEtn and plasmenyl Etn species; (c) averaged spectra from the glycated PtdEtn peak (15.05 to 15.75 min). Peak identification is given. Cl, chloride adducts; diradyl, combined area of diacyl, alkylacyl and alkenylacyl moieties of PtdEtn. LDL total lipid extract was dissolved in chloroform–methanol (2 : 1 , v/v) and 20 µL of the sample containing 10 µg lipid was analyzed. LC/ESMS conditions: normal phase 5 µ Spherisorb column (250 mm × 4.6 mm ID) eluted with a linear gradient of 100% solvent A (chloroform–methanol–30% ammonium hydroxide, 80 : 19.5 : 0.5, v/v/v) to 100% solvent B (chloroform–methanol–water–30% ammonium hydroxide, 60 : 34 : 5.5 : 0.5, v/v/v/v) in 14 min, then at 100% solvent B for 10 min at 1 mL min^{-1}. The effluent was split by 1/50 resulting in 20 µL mL^{-1} being admitted into the mass spectrometer and scanning in the negative ion mode from 400 to 1100 amu. (Reproduced with permission from Ravandi et al.[98])

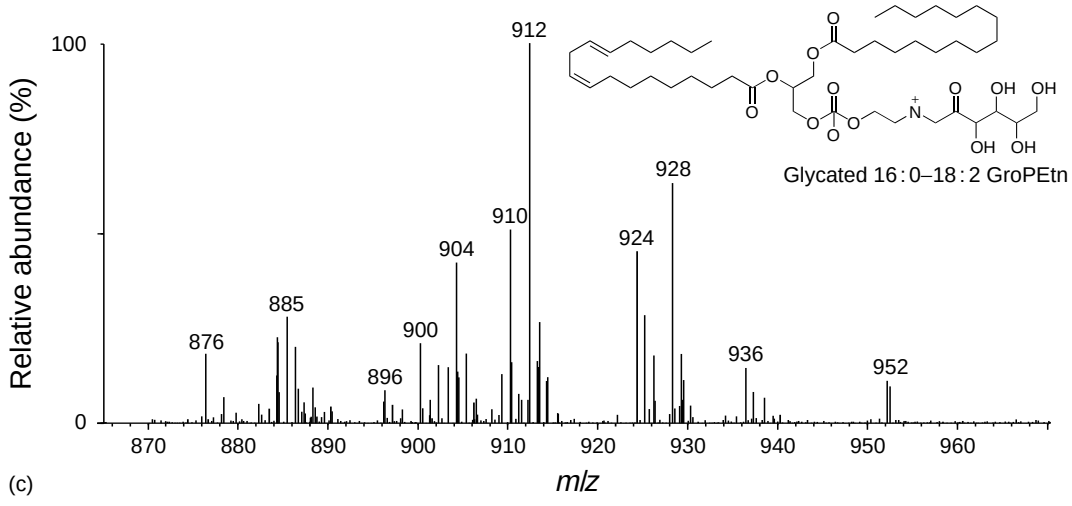

Figure 13 (*Continued*)

structure. As a reference standard Glc PtdEtn is prepared and purified as described previously.[22] Briefly, PtdEtn (2 mg) dissolved in 1 mL of methanol is transferred to a 15 mL test tube and the solvent evaporated under nitrogen. Then, 4 mL of 0.1 M phosphate buffer containing 0–400 mM glucose and 0.1 mM EDTA are added and sonicated at low power for 5 min at room temperature and the mixture incubated under nitrogen at 37 °C for various periods of time. Lipids are extracted into chloroform–methanol (2 : 1, v/v) as described by Folch et al.[18] and the solvents evaporated under nitrogen. Samples are redissolved in chloroform–methanol (2 : 1, v/v) and kept at −20 °C until analysis. Glc PtdEtn (2 mg) is purified by preparative TLC (20 × 20 cm glass plates) coated with silica gel H (250 μm thick layer). The chromatoplates are developed using chloroform–methanol–30% ammonia (65 : 35 : 7, v/v/v) as described.[22] Phospholipids are identified by co-chromatography with appropriate standards, visualizing any lipid bands under UV light after spraying the plate with 0.05% 2,7-dichlorofluorescein in methanol. Both glucosylated and nonglucosylated lipids are recovered by scraping the gel from appropriate areas of the plate and extracting it twice with the developing solvent.

Figure 13 shows the results of the LC/ESMS analysis of phospholipids from LDL preparations containing Glc-PtdEtn (Figure 13a). The Glc-PtdEtn is well resolved from PtdEtn and from other phospholipids seen in the negative ion profile, where the Cho phospholipids yield minimal response. As seen from the single ion mass chromatograms (Figure 13b) all the glycated species are found under the assigned glycated PtdEtn peak. The added glucosylated 16 : 0–18 : 2 GroPEtn (*m/z* 876), $[M-1]^-$ is easily detected in the single ion mass chromatogram (Figure 13c). Using this technique the levels of Glc PtdEtn in plasma have also been estimated[78] in diabetic and normal individuals to be 4.8% and 0.9% GroPEtn/total phospholipids respectively.

Normal-phase HPLC with on-line ESMS (LC/ESMS) is performed by splitting the HPLC flow by 1/50 resulting in 20 μL mL^{-1} being admitted to a Hewlett-Packard Model 5988B quadrupole mass spectrometer equipped with a nebulizer-assisted electrospray interface (HP 59987A).[36] Tuning and calibration of the system is achieved in the mass range of 400–1500 by using the standard phospholipid mix dissolved in the HPLC solvent A and flow-injected at 50 μL min^{-1} into the mass spectrometer. Capillary voltage is set at 4 kV, the endplate voltage at 3.5 kV and the cylinder voltage at 5 kV in the positive mode of ionization. In the negative mode, the voltages are 3.5, 3 and 3.5 kV, respectively. Both positive and negative ion spectra are taken in the mass range 100–1100 amu. The capillary exit voltage is set at 120 V in the positive and 160 V in the negative ion mode. For fragmentation studies the capillary exit voltage is raised to 300 V. Nitrogen gas is used as both nebulizing gas (40 psi) and drying gas (60 psi, 270 °C). Phospholipids, including the glucosylated diradylPEtn, are quantified on basis of standard curves established for each phospholipid class and for the oxidized phospholipids (core aldehydes, hydroperoxy and isoprostanes). The oxidized phospholipids that are used as standards are prepared as previously described.[36] The detection limit for Glc PtdEtn using normal scanning range (400–110 amu) is 20–30 pmol and the response is linear to 100 nmol. The equimolar ion intensities of different species of each phospholipid class varied by less than 5% in each of the ion modes.[99] The LC/ESMS response to different phospholipid classes varied greatly and required the regular use of standards.

5.1.2 Red Cells

Recently, the glycation products of the PtdEtn have also been identified as components of red blood cells.[78] Although PtdSer is present in large amounts, glycated PtdSer was not isolated. The various phospholipid classes and molecular species possess characteristic lipoprotein and cell membrane distribution. As in plasma, the complex lipid mixtures present in the red cells require chromatography and MS for complete determination of the structure. Figure 14 shows the separation of both glycated and nonglycated aminophospholipids from red cells incubated with glucose.[22] The glycated species represent the distribution found in the nonglycated fraction indicating a random glycation process towards both alkenylacyl, diacyl and alkylacyl species. The levels of Glc Etn in diabetic patients (4.8% glycated GroPEtn/total phospholipids) were found to be three times higher than normoglycemic individuals. The levels of glycated Etn have been recently measured by Fountain et al.[106] by acid hydrolysis of the red cell lipid extract and detection of the glycated Etn head group by GC/MS after reduction by $NaBH_4$.

5.1.3 Atheroma

Figure 15 shows the total negative ion current profile (Figure 15a), the single ion chromatograms for major species (Figure 15b), and the mass spectrum averaged over the elution time (15.01–15.89 min) of the glycated Etn phospholipid peak (Figure 15c) as obtained by normal phase LC/ESMS for a total lipid extract of an atherosclerotic plaque from a diabetic male.[98] In order to increase the sensitivity of detection, the scanning is limited to the mass range of 850 to 1000 amu, which eliminated any overlap with the PtdEtn, PtdCho + Cl and SM + Cl ions also present in the total negative LC/ESMS profile. As a result, only the dimers of the molecular species of lysoPtdCho remained visible. There is considerable discrepancy between the glycated and nonglycated sets of species excluding direct in situ interconversion, and suggesting possible deposition of these lipids from the circulation. The composition of the molecular species of the Etn phospholipids of the atheroma also differs significantly from that of human LDL (Table 5), with the relative proportion of the polyunsaturated PtdEtn and specially PlsEtn species being reduced. The most abundant glycated PlsEtn species in the plaque is the 38:4 (*m/z* 928). The possibility of selective glycation and/or oxidation is indicated by the absence of glycated PtdSer species, even though PtdSer species are detected in the atheroma. In samples analysed from nondiabetic individuals we found 2.4 nmol glycated diradyl GroPEtn per milligram of total lipid representing 1.3% of the total atheroma phospholipid. The atherosclerotic tissue from diabetics contained 11.5 nmol glycated diradyl GroPEtn per milligram of total lipid representing 7.3% of total atheroma phospholipid, while an analysis of normal aortic

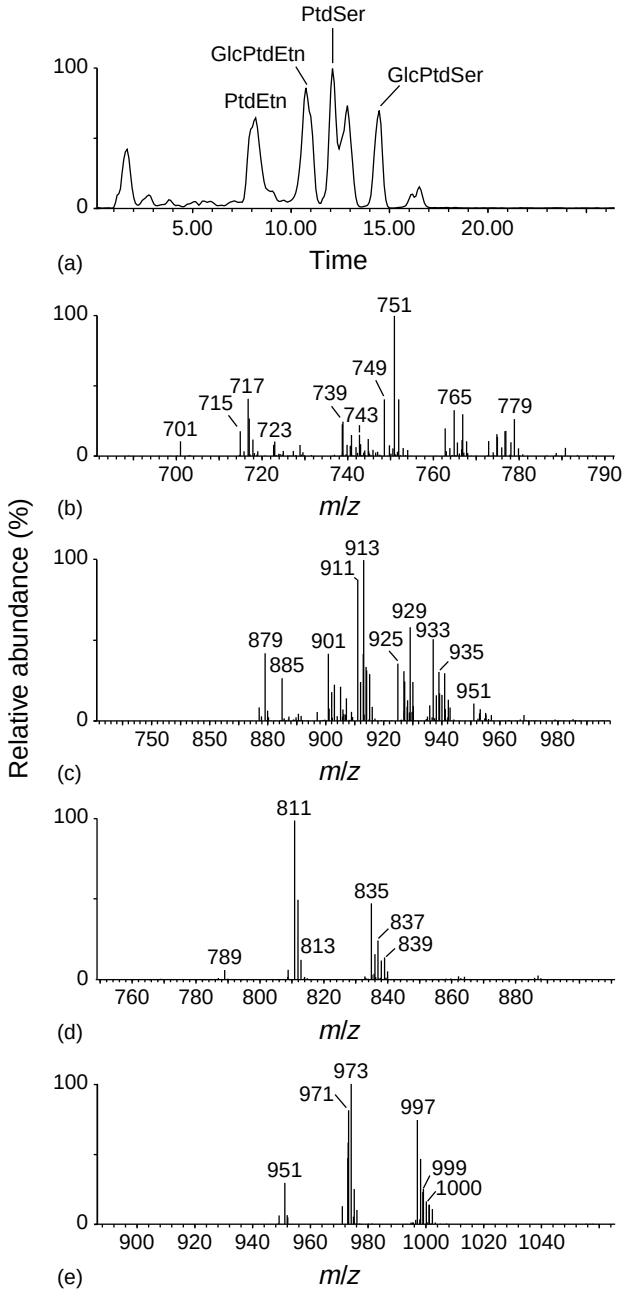

Figure 14 Normal phase LC/MS with electrospray of partially glucosylated aminophospholipids of human red blood cells. (a) Total ion current profile; (b) mass spectrum of nonglucosylated PtdEtn; (c) mass spectrum of glucosylated PtdEtn; (d) mass spectrum of nonglucosylated PtdSer; (e) mass spectrum of glucosylated PtdSer. Values for *m/z* represent major molecular species of PtdEtn and PtdSer as identified elsewhere in the text, or their glucosylated derivatives. (Reproduced with permission from Ravandi et al.[22])

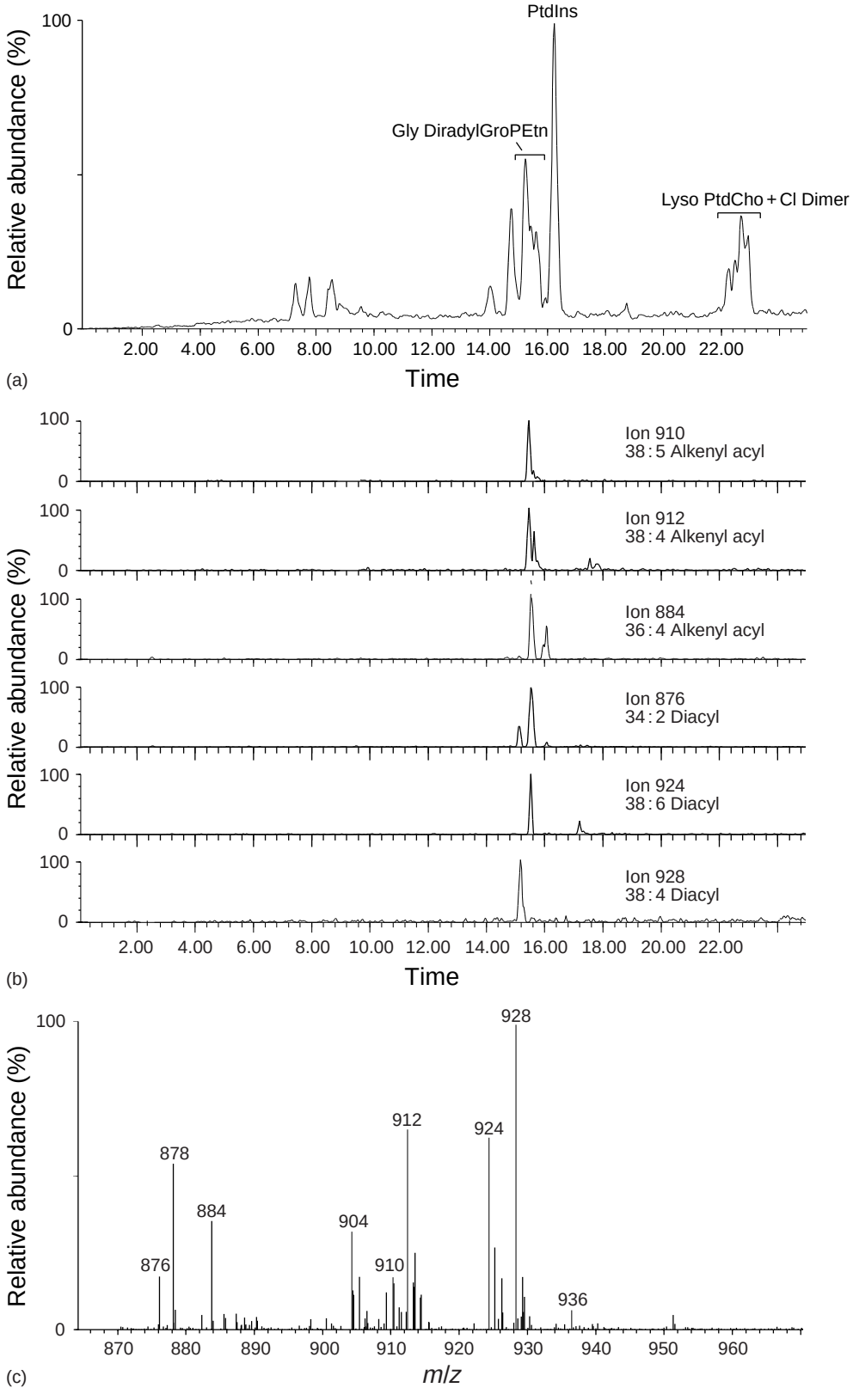

Figure 15 Identification of glycated diradyl GroPEtn from atherosclerotic tissue of a diabetic male.[98] (a) Total negative ion current profile (mass range set to 850–1000 amu); (b) single ion plots of the major glycated diradyl GroPEtn; (c) averaged spectra from the glycated diradyl GroPEtn peak (15.01 to 15.89 min). Peak identification is given. LC/ESMS as described in Figure 13.

tissue did not show any PtdEtn glycation. These results are comparable to the levels of glycated PtdEtn previously measured by our group in plasma of diabetic individuals (4.8% glycated diradyl GroPEtn/total phospholipid).[78]

5.2 Oxidized Glycerophospholipids

5.2.1 Plasma Lipoproteins and Red Cells

It has been claimed that oxidative reactions in vivo are associated with pathological events. To this end much attention has been focused on the products of free radical damage in vivo. Since phospholipids are an abundant source of polyunsaturated FAs, the products of phospholipid oxidation can be markers of peroxidative damage. Ravandi et al.[99] have established an LC/ESMS system for the identification of hydroperoxides of the glycerophospholipids, their aldehyde esters, as well as their isoprostane esters. Figure 16 shows the separation of oxidized phospholipids of human LDL on a normal phase silica column, which allows the identification and quantification of PtdCho oxidation products. It is seen that the various oxidation products are well resolved from the residual Cho-containing phospholipids. Interestingly, the oxidation of PtdCho is accompanied by an extensive release of lysoPtdCho. The isoprostanes are produced only from the arachidonate-containing species. The oxidation products of other long chain polyunsaturated esters were not identified. The relatively saturated SM does not appear to be affected by mild oxidation with copper.[99] Ahmed et al.[107] have successfully extended this method to the identification and quantification of the lipid peroxidation products of human HDL resulting from treatment with peroxynitrite. Lynch et al.[108] have previously demonstrated the formation of the hydroperoxides and isoprostanes in plasma and LDL exposed to oxidative stress in vitro, but these workers did not analyze intact lipid esters.

Most of the efforts in phospholipid oxidation have been focused on PtdCho and PtdEtn. One technique for the quantitation of PtdCho hydroperoxides has been the use of HPLC with a chemiluminescence detector. Akasaka et al.[109] have developed an automatic method for the determination of hydroperoxides by injection of deproteinized plasma sample on to a reverse-phase column and subsequent introduction into two analytical columns and post-column reaction with diphenyl-1-pyrenylphosphine. They reported the presence of 6–7 pmol of PtdCho hydroperoxides in human plasma. The drawback of this technique is the complicated chromatography apparatus and the lack of detailed information about the molecular species of the phospholipid.

5.2.2 Atheroma

Ravandi et al.[110] have demonstrated the presence of hydroperoxides, core aldehydes as well as isoprostanes of glycerophospholipids in samples of atheroma. Figure 17

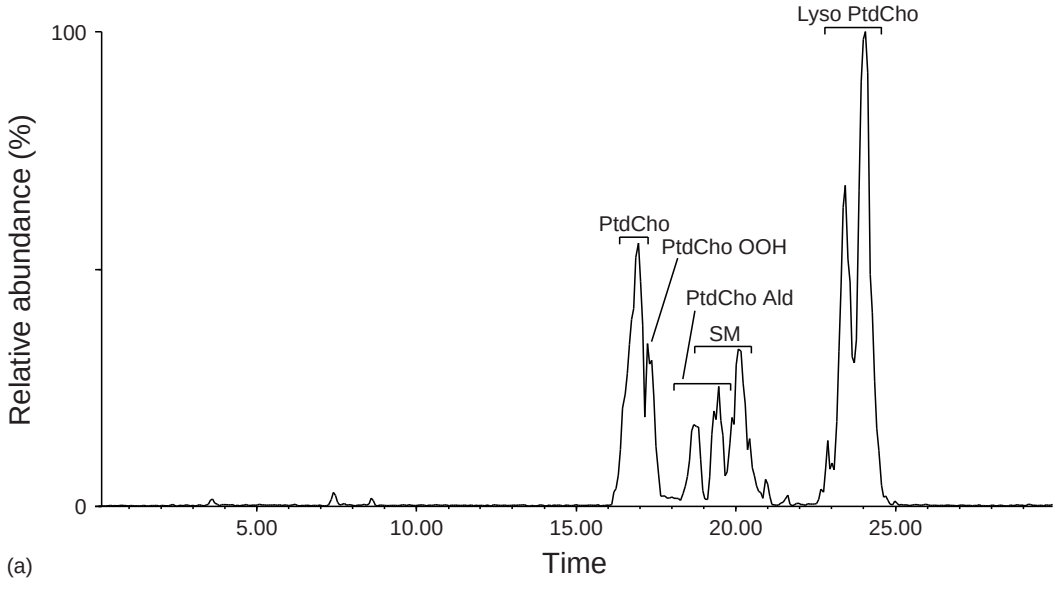

Figure 16 LC/ESMS analysis of oxidized PtdCho in oxidized LDL.[106] (a) Total positive ion current profile of oxidized LDL; (b) single ion plots of representative PtdCho oxidation products. Peak identification is given. LDL was oxidized by incubation with 5 μM $CuSO_4$ in 0.1 M PBS for 12 h at 37 °C. The total lipid extract of the oxidized LDL was dissolved in chloroform–methanol (2 : 1, v/v) and 20 μL of the sample containing 10 μg lipid was analyzed. Structural assignments for aldehydes and hydroperoxides are according to reference standards. Ions 832 and 830 were identified on the basis of retention time and molecular weight. (Reproduced with permission from Ravandi et al.[99])

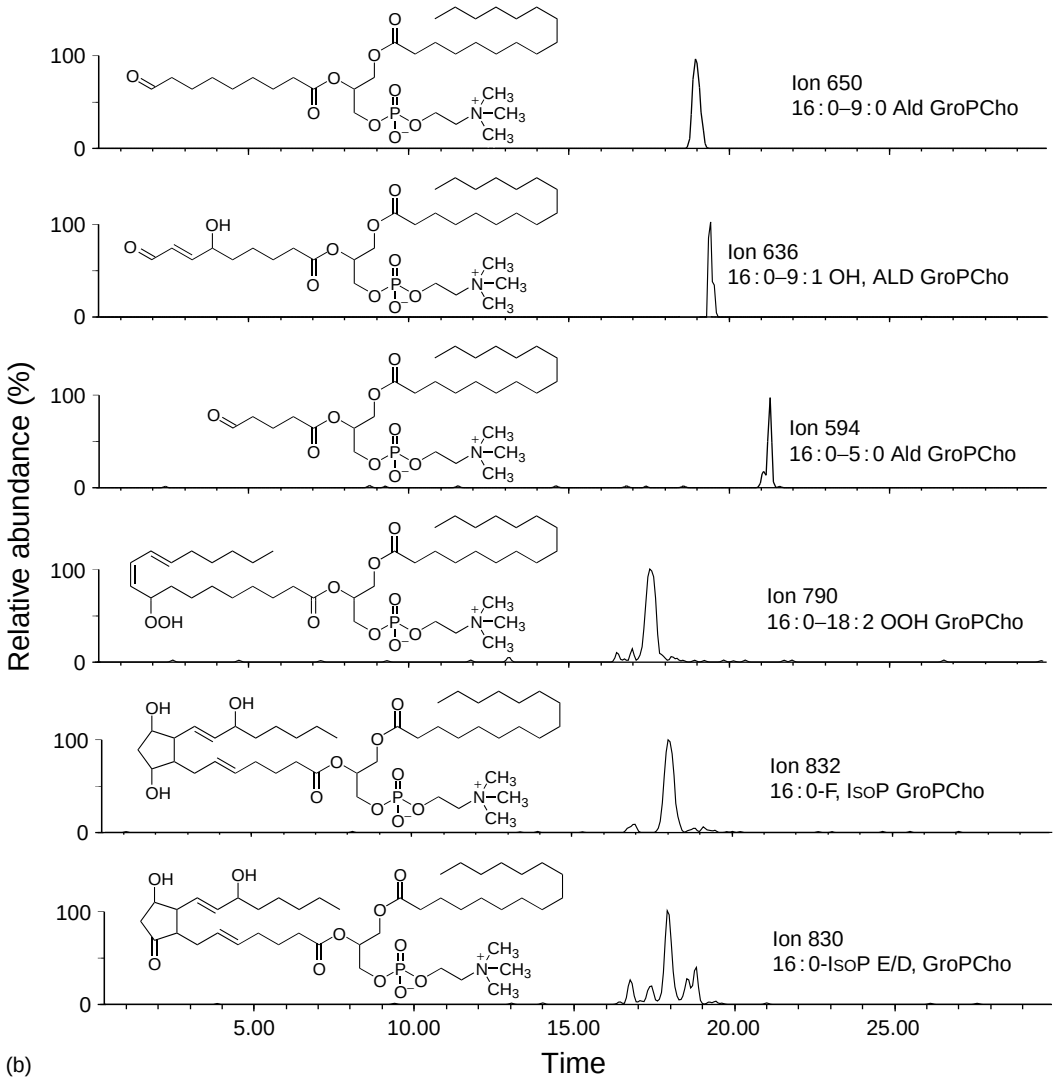

Figure 16 (*Continued*)

shows appropriate single ion chromatograms. The adoption of the single ion monitoring mode allowed the detection limit to be lowered to 10–50 fmol for the different oxidized phospholipid classes.

5.3 Phosphoinositide Glycans

Over 100 glycosyl PtdIns anchored proteins have been described and a number of comprehensive reviews have appeared.[11,13] The anchors have been isolated in the free form and more commonly covalently bound to proteins, from which they may be released by chemical and enzymatic methods.

5.3.1 Plasma Lipoproteins

Although both phospholipases C and D specific for PtdIns glycans have been isolated from plasma, neither free nor bound PtdIns glycans have been detected in plasma or serum,[13,111] and the role of the enzymes has remained obscure. Recently, Nazih-Sanderson et al.[112] have shown that HDL_3 signaling in $HepG_2$ cells involves glycosyl-PtdIns-anchored proteins. In the presence of protease-inhibitors, HDL_3 releases in the culture medium several proteins with a residual inositolphosphoglycan that binds Ab_{IPG} after sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis and immunoblotting. Thus, HDL_3 activates PtdCho hydrolysis through a multistep pathway involving the cleavage of glycosyl PtdIns-anchored proteins.

5.3.2 Red Cells

The nature of the lipid moiety of the glycosyl PtdIns anchor of the red cell acetylcholine esterase

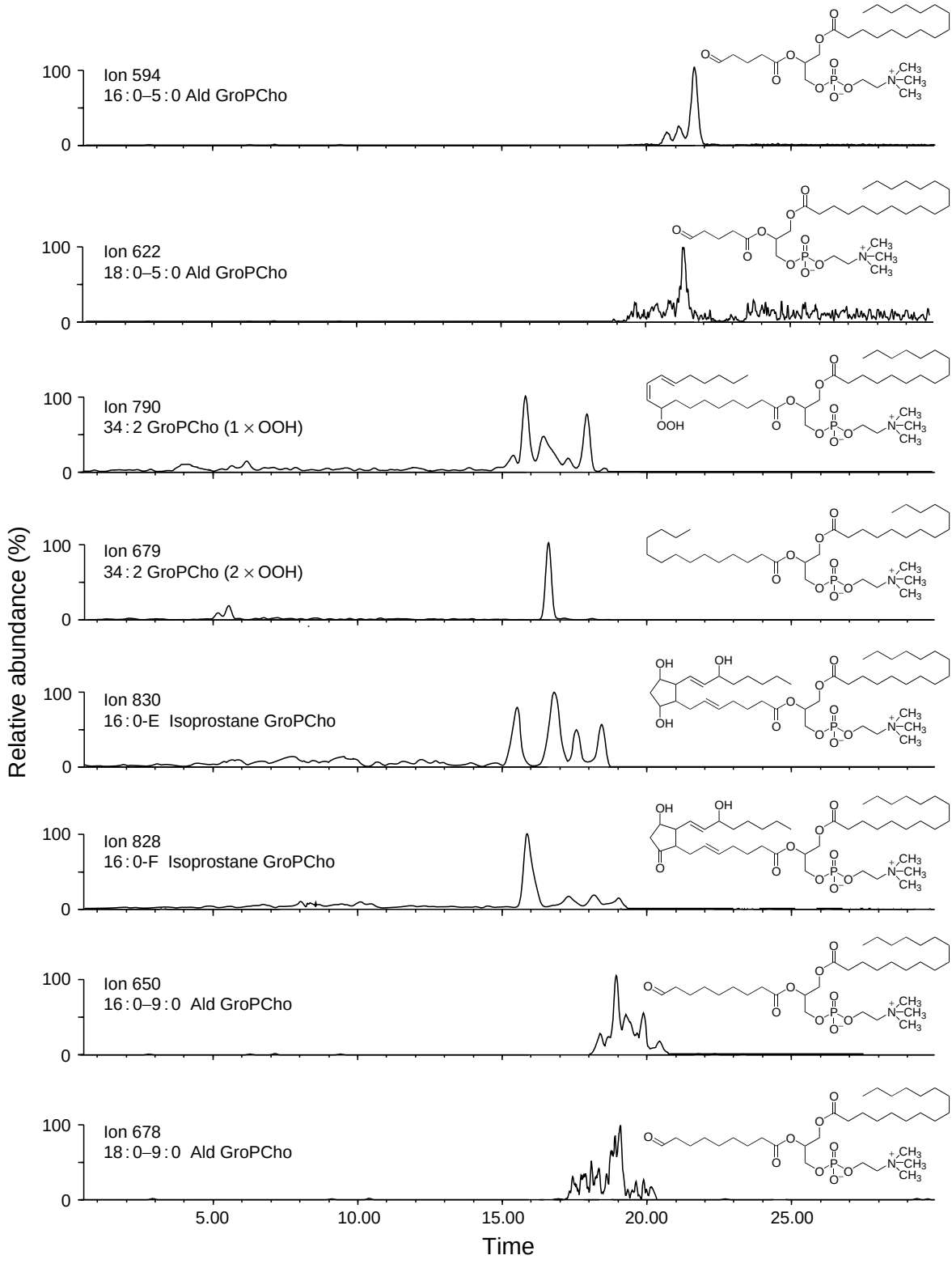

Figure 17 Single ion monitoring LC/ESMS analysis of oxidized phospholipids in human atheroma.[110] Peak identification is given. Ald, core aldehyde. LC/ESMS conditions as in Figure 16.

has been determined[33,58] by GC/MS of lipid derivatives from hydrolyzed material and/or FAB/MS of the PtdIns or diradylglycerol moieties, released by nitrous acid deamination and PtdIns-specific phospholipase C treatment,[12,13,113] respectively. Phospholipid fragments may be generated from a glycosyl PtdIns anchor also by phospholipase D cleavage. The PtdIns anchor specific phospholipase D found in mammalian (including human) serum, cleaves the anchor irrespective of inositol acylation to give PtdOH.[13,114] Nitrous acid treatment of the anchor results in deamination of glucosamine residue and cleavage of glucosamine-inositol glycosidic bond to release an inositol-containing phospholipid fragment.[13]

The lipid moieties of the PtdIns anchor of the human red blood cell acetylcholine esterase are made up largely of alkylacyl species.[33] The human red cell acetylcholine esterase contains an additional FA (palmitate) on the inositol ring.[12] The molecular species of the PtdIns moiety of the acetylcholine esterase anchor have been identified and quantitated (mol%) as follows: 18:0′–18:0 (6.3%); 18:1′–18:0 (2.7%); 18:0–20:4(n-6) (<2.0%); 16:0′–22:4(n-6) (2.0%); 17:0′–22:4(n-6) (1.4%); 18:0′–22:4(n-6) (32.4%); 18:1′–22:4(n-6) (14.2%); 18:0′–(22:5(n-3) and 22:6(n-3)) (25.2%) and 18:1′–(22:5(n-3) and 22:6(n-3)) (11.0%). (The alkyl chains have been identified by primes). The values for the *sn*-1,2- and *sn*-1,3-isomers which were generated by acetolysis were summed. This human red cell acetylcholine esterase anchor is a plasmanylinositol that is palmitoylated on the inositol ring.[12] The early mass spectrometric studies on glycosyl PtdIns glycans have been performed with fast atom bombardment mass spectrometers, which have provided evidence for the presence of a palmitoyl group on one of the inositol hydroxyl groups in the PtdIns moiety of the glycosyl PtdIns anchor of the acetylcholine esterase.[12]

Neither FAB/MS/MS nor ESMS/MS can provide the molecular species composition of the glycosyl PtdIns anchors as long as deacylated materials are being used for the analysis.

6 CONCLUSIONS

The great complexity and the large number of phospholipid classes and molecular species continue to provide a challenge to the best contemporary methods of lipid analysis. Solvent extraction and TLC remain basic techniques for sample preparation, while HPLC on normal- and reversed-phase columns provide effective means for sample purification and preparation of standards.

Traditional sample extraction by liquid–liquid partition is time-consuming and involves large volumes of solvents. More effective is liquid–solid extraction, which is accomplished by passing a dilute plasma or red cell lipid extract through a solid adsorbent bed. However, in the absence of a simple and inexpensive detector for monitoring the column effluents many of these advantages are lost. The experimenter is forced to work blindly (or employ TLC for assessing the success of the SPE) and spend many hours and increasing volumes of solvent in trial and error.

The availability of combined chromatographic and soft ionization mass spectrometric or tandem mass spectrometric methods provide new, rapid and highly specific assays of lipid class and molecular species identification and quantification. In many instances repetition of the older quantitative work by the newer methods would be clearly justified and a revision warranted of much published data of lipid class and molecular species composition of many tissues and cells. Finally, the great potential of NMR as a quantitative tool of phospholipid class analysis remains to be generally exploited.

ACKNOWLEDGMENTS

The research from the authors' laboratory was supported with funds and equipment awarded by the Heart and Stroke Foundation of Ontario, Toronto, Ontario and the Medical Research Council of Canada, Ottawa, Canada.

ABBREVIATIONS AND ACRONYMS

BHT	Butylated Hydroxytoluene
Cho	Choline
CID	Collision-induced Dissociation
DEAE	Diethylamino Ethyl
EDTA	Ethylenediamine Tetraacetic Acid
ESMS	Electrospray Mass Spectrometry
Etn	Ethanolamine
FA	Fatty Acid
FAB/MS	Fast Atom Bombardment Mass Spectrometry
FPLC	Fast Protein Liquid Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GLC	Gas–Liquid Chromatography
Gro	Glycerol
GroPCho	Glycerophosphocholine
GroPEtn	Glycerophosphoethanolamine
HDL	High Density Lipoprotein
HPLC	High-performance Liquid Chromatography

HPTLC	High-performance Thin-layer Chromatography
LC/ESMS	Liquid Chromatography with On-line Electrospray Mass Spectrometry
LC/MS	Liquid Chromatography/Mass Spectrometry
LDL	Low Density Lipoprotein
MALDI	Matrix-assisted Laser Desorption/ Ionization
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MTBE	Methyl *tert*-Butyl Ether
NMR	Nuclear Magnetic Resonance
P	Phosphoric Acid
P_i	Inorganic Phosphate
PAF	Platelet Activating Factor
PGE_1	Prostaglandin E_1
PlsEtn	Alkenylacyl Glycerophospho-ethanolamine
PtdCho	Phosphatidylcholine
PtdEtn	Phosphatidylethanolamine
PtdGro	Phosphatidylglycerol
Ptd_2Gro	Diphosphatidylglycerol
PtdIns	Phosphatidylinositol
PtdIns P	Phosphatidylinositol Phosphate
PtdOH	Phosphatidic Acid
PtdSer	Phosphatidylserine
SM	Sphingomyelin
SPE	Solid-phase Extraction
TBDMS	*tert*-Butyldimethylsilyl
TLC	Thin-layer Chromatography
TMS	Trimethylsilyl
UV	Ultraviolet
VLDL	Very Low Density Lipoprotein

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Glycolipid Analysis

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Lipid Analysis for Important Clinical Conditions

Food **(Volume 5)**
Lipid Analyses in Food

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

REFERENCES

1. IUPAC-IUB Commission on Biochemical Nomenclature, 'Nomenclature of Phosphorus-containing Compounds of Biological Importance', *Hoppe-Seyler's Z. Physiol. Chem.*, **358**, 599–616 (1977).
2. E.A. Dennis, 'Diversity of Group Types, Regulation, and Function of Phospholipase A2', *J. Biol. Chem.*, **269**(18), 13 057–13 060 (1994).
3. A. Poulos, D.W. Johnson, K. Beckman, I.G. White, C. Easton, 'Occurrence of Unusual Molecular Species of Sphingomyelin Containing 28-34-carbon Polyenoic Fatty Acids in Ram Spermatozoa', *Biochem. J.*, **248**(3), 961–964 (1987).
4. J.J. Myher, A. Kuksis, W.C. Breckenridge, J.A. Little, 'Differential Distribution of Sphingomyelins Among Plasma Lipoprotein Classes', *Can. J. Biochem.*, **59**(8), 626–636 (1981).
5. L.Y. Yang, A. Kuksis, J.J. Myher, H. Pang, 'Surface Components of Chylomicrons from Rats Fed Glyceryl or Alkyl Esters of Fatty Acids: Minor Components', *Lipids*, **27**(8), 613–618 (1992).
6. R. Parthasarathy, F. Eisenberg, Jr, 'The Inositol Phospholipids: A Stereochemical View of Biological Activity', *Biochem. J.*, **235**(2), 313–322 (1986).
7. P.T. Hawkins, R.H. Michell, C.J. Kirk, 'Analysis of the Metabolic Turnover of the Individual Phosphate Groups of Phosphatidylinositol 4-Phosphate and Phosphatidylinositol 4,5-Bisphosphate. Validation of Novel Analytical Techniques by Using ^{32}P-Labelled Lipids from Erythrocytes', *Biochem. J.*, **218**(3), 785–793 (1984).
8. C.E. King, L.R. Stephens, P.T. Hawkins, G.R. Guy, R.H. Michell, 'Multiple Metabolic Pools of Phosphoinositides and Phosphatidate in Human Erythrocytes Incubated in a Medium that Permits Rapid Transmembrane Exchange of Phosphate', *Biochem. J.*, **244**(1), 209–217 (1987).
9. S.E. Rittenhouse, 'Phosphoinositide 3-Kinase Activation and Platelet Function', *J. Am. Soc. Hematol.*, **88**, 4401–4414 (1996).
10. A.N. Carter, R. Huang, A. Sorisky, C.P. Downes, S.E. Rittenhouse, 'Phosphatidylinositol 3,4,5-Trisphosphate is Formed from Phosphatidylinositol 4,5-Bisphosphate in Thrombin-stimulated Platelets', *Biochem. J.*, **301**(2), 415–420 (1994).
11. M.J. McConville, M.A. Ferguson, 'The Structure, Biosynthesis and Function of Glycosylated Phosphatidylinositols in the Parasitic Protozoa and Higher Eukaryotes', *Biochem. J.*, **294**(2), 305–324 (1993).
12. W.L. Roberts, S. Santikarn, V.N. Reinhold, T.L. Rosenberry, 'Structural Characterization of the Glycoinositol Phospholipid Membrane Anchor of Human Erythrocyte Acetylcholinesterase by Fast Atom Bombardment Mass Spectrometry', *J. Biol. Chem.*, **263**(35), 18 776–18 784 (1988).

13. A.K. Menon, 'Structural Analysis of Glycosylphosphatidylinositol Anchors', *Methods Enzymol.*, **230**, 418–442 (1994).
14. M. Lagarde, P.A. Bryon, M. Guichardant, M. Dechavanne, 'A Simple and Efficient Method for Platelet Isolation from their Plasma', *Thromb. Res.*, **17**(3–4), 581–588 (1980).
15. M.G. Low, 'Purification of Phosphatidylinositol 4-Phosphate and Phosphatidylinositol 4,5-Bisphosphate by Column Chromatography', in *Methods in Inositide Research*, ed. R.F. Irvine, Raven Press, New York, 145–151, 1990.
16. F. Hullin, H.Y. Kim, N. Salem, Jr, 'Analysis of Aminophospholipid Molecular Species by High Performance Liquid Chromatography', *J. Lipid Res.*, **30**(12), 1963–1975 (1989).
17. G.M. Fless, 'General Preparative Ultracentrifugation and Considerations for Lipoprotein Isolation', in *Analyses of Fats, Oils and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists' Society Press, Champaign, IL, 512–523, 1991.
18. J. Folch, M. Lees, H.S. Stanley, 'A Simple Method for the Isolation and Purification of the Total Lipid from Animal Tissue', *J. Biol. Chem.*, **226**, 497–509 (1957).
19. E.G. Bligh, W.J. Dyer, 'A Rapid Method of Total Lipid Extraction and Purification', *Can. J. Biochem.*, **37**(911), 917 (1959).
20. A. Kuksis, J.J. Myher, 'Gas Chromatographic Analysis of Plasma Lipids', *Adv. Chromatogr.*, **28**, 267–332 (1989).
21. J.J. Myher, A. Kuksis, A. Ravandi, N. Cocks, 'Normal-phase Liquid Chromatography/Mass Spectrometry with Electrospray for Sensitive Detection of Oxygenated Glycerophospholipids', *INFORM*, **5**, 478–479 (1994).
22. A. Ravandi, A. Kuksis, L. Marai, J.J. Myher, 'Preparation and Characterization of Glucosylated Aminoglycerophospholipids', *Lipids*, **30**(10), 885–891 (1995).
23. A. Ravandi, A. Kuksis, N. Shaikh, G. Jackowski, 'Preparation of Schiff Base Adducts of Phosphatidylcholine Core Aldehydes and Aminophospholipids, Amino Acids, and Myoglobin', *Lipids*, **32**(9), 989–1001 (1997).
24. H.G. Rose, M. Oaklander, 'Improved Procedure for the Extraction of Lipids from Human Erythrocytes', *J. Lipid Res.*, **6**, 428–431 (1965).
25. T. Nakamura, D.L. Bratton, R.C. Murphy, 'Analysis of Epoxyeicosatrienoic and Monohydroxyeicosatetraenoic Acids Esterified to Phospholipids in Human Red Blood Cells by Electrospray Tandem Mass Spectrometry', *J. Mass Spectrom.*, **32**(8), 888–896 (1997).
26. S.S. Katz, D.M. Small, 'Isolation and Partial Characterization of the Lipid Phases of Human Atherosclerotic Plaques', *J. Biol. Chem.*, **255**(20), 9753–9759 (1980).
27. J.J. Piotrowski, S. Shah, J.J. Alexander, 'Mature Human Atherosclerotic Plaque Contains Peroxidized Phosphatidylcholine as a Major Lipid Peroxide', *Life Sci.*, **58**(9), 735–740 (1996).
28. C. Suarna, R.T. Dean, J. May, R. Stocker, 'Human Atherosclerotic Plaque Contains both Oxidized Lipids and Relatively Large Amounts of Alpha-tocopherol and Ascorbate', *Arterioscler. Thromb. Vasc. Biol.*, **15**(10), 1616–1624 (1995).
29. M. Saito, Y. Tanaka, S. Ando, 'Thin-layer Chromatography-densitometry of Minor Acidic Phospholipids: Application to Lipids from Erythrocytes, Liver, and Kidney', *Anal. Biochem.*, **132**(2), 376–383 (1983).
30. N.A. Shaikh, 'Assessment of Various Techniques for the Quantitative Extraction of Lysophospholipids from Myocardial Tissues', *Anal. Biochem.*, **216**(2), 313–321 (1994).
31. P. Muller, A. Herrmann, R. Glaser, 'Further Evidence for a Membrane Potential-dependent Shape Transformation of the Human Erythrocyte Membrane', *Biosci. Rep.*, **6**(11), 999–1006 (1986).
32. P. Schneider, M.A. Ferguson, M.J. McConville, A. Mehlert, S.W. Homans, C. Bordier, 'Structure of the Glycosyl-phosphatidylinositol Membrane Anchor of the Leishmania Major Promastigote Surface Protease', *J. Biol. Chem.*, **265**(28), 16 955–16 964 (1990).
33. W.L. Roberts, J.J. Myher, A. Kuksis, M.G. Low, T.L. Rosenberry, 'Lipid Analysis of the Glycoinositol Phospholipid Membrane Anchor of Human Erythrocyte Acetylcholinesterase. Palmitoylation of Inositol Results in Resistance to Phosphatidylinositol-specific Phospholipase C', *J. Biol. Chem.*, **263**(35), 18 766–18 775 (1988).
34. D.R. Janero, C. Burghardt, 'Solid-phase Extraction on Silica Cartridges as an Aid to Platelet-activating Factor Enrichment and Analysis', *J. Chromatogr.*, **526**(1), 11–24 (1990).
35. J.G. Hamilton, K. Comai, 'Separation of Neutral Lipid, Free Fatty Acid and Phospholipid Classes by Normal Phase HPLC', *Lipids*, **23**(12), 1150–1153 (1988).
36. A. Ravandi, A. Kuksis, J.J. Myher, L. Marai, 'Determination of Lipid Ester Ozonides and Core Aldehydes by High-performance Liquid Chromatography with On-line Mass Spectrometry', *J. Biochem. Biophys. Methods*, **30**(4), 271–285 (1995).
37. J.C. Becart, J.P. Chevalier, A. Biese, 'Quantitative Analysis of Phospholipids by HPLC with a Light Scattering Evaporating Detector', *J. High Resol. Chromatogr.*, **10**, 126–129 (1990).
38. A.K. Singh, 'Quantitative Analysis of Inositol Lipids and Inositol Phosphates in Synaptosomes and Microvessels by Column Chromatography: Comparison of the Mass Analysis and the Radiolabelling Methods', *J. Chromatogr.*, **581**(1), 1–10 (1992).
39. L.L. Dugan, P. Demediuk, C.E. Pendley, L.A. Horrocks, 'Separation of Phospholipids by High-performance Liquid Chromatography: All Major Classes, Including Ethanolamine and Choline Plasmalogens, and Most Minor Classes, Including Lysophosphatidylethanolamine', *J. Chromatogr.*, **378**(2), 317–327 (1986).

40. G. Rouser, G. Kritchevsky, A. Yamamoto, G. Simon, Galli, J. Bauman, 'Dimethylaminoethyl and Triethylaminoethyl Cellulose Column Chromatographic Procedures for Phospholipids, Glycolipids and Pigments', *Methods Enzymol.*, **14**, 272–317 (1969).
41. W.S. Van Kessel, W.M. Hax, R.A. Demel, J. de Gier, 'High Performance Liquid Chromatographic Separation and Direct Ultraviolet Detection of Phospholipids', *Biochim. Biophys. Acta*, **486**(3), 524–530 (1977).
42. S.S. Katz, G.G. Shipley, D.M. Small, 'Physical Chemistry of the Lipids of Human Atherosclerotic Lesions. Demonstration of a Lesion Intermediate Between Fatty Streaks and Advanced Plaques', *J. Clin. Invest.*, **58**(1), 200–211 (1976).
43. M.A. Kaluzny, L.A. Duncan, M.V. Merritt, D.E. Epps, 'Rapid Separation of Lipid Classes in High Yield and Purity Using Bonded Phase Columns', *J. Lipid Res.*, **26**(1), 135–140 (1985).
44. J.J. Agren, A. Julkunen, I. Penttila, 'Rapid Separation of Serum Lipids for Fatty Acid Analysis by a Single Aminopropyl Column', *J. Lipid Res.*, **33**(12), 1871–1876 (1992).
45. D.A. Figlewicz, C.E. Nolan, I.N. Singh, F.B. Jungalwala, 'Pre-packed Reverse Phase Columns for Isolation of Complex Lipids Synthesized from Radioactive Precursors', *J. Lipid Res.*, **26**(1), 140–144 (1985).
46. H. Salari, 'Comparative Study of Solid-phase and Liquid-phase Extraction Techniques for Isolation of Phospholipids from Plasma', *J. Chromatogr.*, **419**, 103–111 (1987).
47. H.Y. Kim, N. Salem, Jr, 'Preparation and the Structural Determination of Hydroperoxy Derivatives of Docosahexaenoic Acid and Other Polyunsaturates by Thermospray LC/MS', *Prostaglandins*, **37**(1), 105–119 (1989).
48. W.W. Christie, 'The Analysis of Complex Lipids', in *Lipid Analysis*, ed. W.W. Christie, Pergamon Press, Oxford, 107–134, 1982.
49. H. Salari, 'Solid-phase Extraction and Reversed-phase High-performance Liquid Chromatographic Technique for Isolation and Estimation of Platelet Activating Factor in Plasma', *J. Chromatogr.*, **382**, 89–98 (1986).
50. F.B. Palmer, 'Chromatography of Acidic Phospholipids on Immobilized Neomycin', *J. Lipid Res.*, **22**(8), 1296–1300 (1981).
51. A.J. Letcher, L.R. Stephens, R.F. Irvine, 'Preparation of ^{32}P-labelled Inositol 1,4,5-Triphosphate and ^{14}C-labelled Inositol 1,4-Bisphosphate', in *Methods in Inositide Research*, ed. R.F. Irvine, Raven Press, New York, 31–37, 1990.
52. J. Schacht, 'Purification of Polyphosphoinositides by Chromatography on Immobilized Neomycin', *J. Lipid Res.*, **19**(8), 1063–1067 (1978).
53. I.A. Sobenin, V.V. Tetrov, T. Koschinsky, C.E. Buenting, E.S. Slavina, I.I. Dedov, A.N. Orekhov, 'Modified Low Density Lipoprotein from Diabetic Patients Causes Cholesterol Accumulation in Human Intimal Aortic Cells', *Atherosclerosis*, **100**, 41–54 (1993).
54. W.C. Breckenridge, F.B. Palmer, 'Fatty Acid Composition of Human Plasma Lipoprotein Phosphatidylinositols', *Biochim. Biophys. Acta*, **712**(3), 707–711 (1982).
55. J.L. Kerwin, A.R. Tuininga, L.H. Ericsson, 'Identification of Molecular Species of Glycerophospholipids and Sphingomyelin Using Electrospray Mass Spectrometry', *J. Lipid Res.*, **35**(6), 1102–1114 (1994).
56. J.J. Myher, A. Kuksis, L. Marai, S.K. Yeung, 'Microdetermination of Molecular Species of Oligo- and Polyunsaturated Diacylglycerols by Gas Chromatography–Mass Spectrometry of their *tert*-Butyl Dimethylsilyl Ethers', *Anal. Chem.*, **50**, 557–561 (1978).
57. J.J. Myher, A. Kuksis, 'Improved Resolution of Natural Diacylglycerols by Gas–Liquid Chromatography on Polar Siloxanes', *J. Chromatogr. Sci.*, **13**(3), 138–145 (1975).
58. A. Kuksis, J.J. Myher, 'Mass Analysis of Molecular Species of Diradylglycerols', in *Methods in Inositide Research*, ed. R.F. Irvine, Raven Press, New York, 187–216, 1990.
59. J.J. Myher, A. Kuksis, S. Pind, 'Molecular Species of Glycerophospholipids and Sphingomyelins of Human Erythrocytes: Improved Method of Analysis', *Lipids*, **24**(5), 396–407 (1989).
60. J.J. Myher, A. Kuksis, S. Pind, 'Molecular Species of Glycerophospholipids and Sphingomyelins of Human Plasma: Comparison to Red Blood Cells', *Lipids*, **24**(5), 408–418 (1989).
61. A. Kuksis, J.J. Myher, K. Geher, N.A. Shaikh, W.C. Breckenridge, G.J.L. Jones, J.A. Little, 'Comparative Determination of Plasma Phospholipids by Automated Gas–Liquid Chromatographic and Manual Colorimetric Phosphorus Methods', *J. Chromatogr.*, **182**, 1–26 (1978).
62. M.L. Blank, M. Robinson, V. Fitzgeral, F. Snyder, 'Novel Quantitative Method for Determination of Molecular Species of Phospholipids and Diglycerides', *J. Chromatogr.*, **298**, 473–482 (1984).
63. A. Kuksis, L. Marai, J.J. Myher, Y. Itabashi, S. Pind, 'Applications of GC/MS, LC/MS, and FAB/MS for Determination of Molecular Species of Glycerolipids', in *AOCS Monograph on Lipid and Lipoprotein Analysis*, ed. E.G. Perkins, AOCS Press, Champaign, IL, 464–495, 1991.
64. J. Park, S. Lemieux, G.F. Lewis, A. Kuksis, G. Steiner, 'Chronic Exogenous Insulin and Chronic Carbohydrate Supplementation Increase de novo VLDL Triglyceride Fatty Acid Production in Rats', *J. Lipid Res.*, **38**(12), 2529–2536 (1997).
65. Y. Itabashi, A. Kuksis, J.J. Myher, 'Determination of Molecular Species of Enantiomeric Diacylglycerols by Chiral Phase High Performance Liquid Chromatography

and Polar Capillary Gas–Liquid Chromatography', *J. Lipid Res.*, **31**(11), 2119–2126 (1990).
66. M.G. Luthra, A. Sheltawy, 'The Metabolic Turnover of Molecular Species of Phosphatidylinositol and its Precursor Phosphatidic Acid in Guinea-pig Cerebral Hemispheres', *J. Neurochem.*, **27**(6), 1501–1511 (1976).
67. W.L. Roberts, J.J. Myher, A. Kuksis, T.L. Rosenberry, 'Alkylacylglycerol Molecular Species in the Glycosylinositol Phospholipid Membrane Anchor of Bovine Erythrocyte Acetylcholinesterase', *Biochem. Biophys. Res. Commun.*, **150**(1), 271–277 (1988).
68. M.J. Wilson, K. Richter-Lowney, D.L. Daleke, 'Hyperglycemia Induces a Loss of Phospholipid Asymmetry in Human Erythrocytes', *Biochemistry*, **32**(42), 11 302–11 310 (1993).
69. H.Y. Kim, N. Salem, Jr, 'Liquid Chromatography–Mass Spectrometry of Lipids', *Prog. Lipid Res.*, **32**(3), 221–245 (1993).
70. F. Hullin, M.J. Bossant, N. Salem, Jr, 'Aminophospholipid Molecular Species Asymmetry in the Human Erythrocyte Plasma Membrane', *Biochim. Biophys. Acta*, **1061**(1), 15–25 (1991).
71. L.A. Horrocks, M. Sharma, 'Plasmalogens and O-alkyl Glycerophospholipids', in *Phospholipids*, eds. G.B. Ansell, J.N. Hawthorn, Elsevier Biomedical Press, New York, 51–93, 1982.
72. X. Han, R.W. Gross, 'Electrospray Ionization Mass Spectroscopic Analysis of Human Erythrocyte Plasma Membrane Phospholipids', *Proc. Natl. Acad. Sci. USA*, **91**(22), 10 635–10 639 (1994).
73. H. Oral, G.W. Dorn, D.L. Mann, 'Sphingosine Mediates the Immediate Negative Inotropic Effects of Tumor Necrosis Factor-alpha in the Adult Mammalian Cardiac Myocyte', *J. Biol. Chem.*, **272**(8), 4836–4842 (1997).
74. K.R. Auger, L.A. Serunian, C. Cantley, 'Separation of Novel Polyphosphoinisitides', in *Methods in Inositide Research*, ed. R.F. Irvine, Raven Press, New York, 159–166, 1990.
75. G.L. Kucera, S.E. Rittenhouse, 'Human Platelets Form 3-Phosphorylated Phosphoinositides in Response to Alpha-thrombin, U46619, or GTP Gamma S', *J. Biol. Chem.*, **265**(10), 5345–5348 (1990).
76. X. Han, R.A. Gubitosi-Klug, B.J. Collins, R.W. Gross, 'Alterations in Individual Molecular Species of Human Platelet Phospholipids During Thrombin Stimulation: Electrospray Ionization Mass Spectrometry-facilitated Identification of the Boundary Conditions for the Magnitude and Selectivity of Thrombin-induced Platelet Phospholipid Hydrolysis', *Biochemistry*, **35**(18), 5822–5832 (1996).
77. T. Glonek, T.E. Merchant, '[^{31}P] NMR Profiling of Phospholipids', in *Advances in Lipid Methodology*, ed. W.W. Christie, The Oily Press, West Ferry, Dundee, 45–56, 1995.
78. A. Ravandi, A. Kuksis, L. Marai, J.J. Myher, G. Steiner, G. Lewis, H. Kamido, 'Isolation and Identification of Glycated Aminophospholipids from Red Cells and Plasma of Diabetic Blood', *FEBS Lett.*, **381**(1–2), 77–81 (1996).
79. W. Pruzanski, E. Stefanski, F.C. de Beer, M.C. de Beer, P. Vadas, A. Ravandi, A. Kuksis, 'Lipoproteins are Substrates for Human Secretory Group IIA Phospholipase A2: Preferential Hydrolysis of Acute Phase HDL', *J. Lipid Res.*, **39**(11), 2150–2160 (1998).
80. G.A. Arvidson, 'Structural and Metabolic Heterogeneity of Rat Liver Glycerophosphatides', *Eur. J. Biochem.*, **4**(4), 478–486 (1968).
81. S. Pind, A. Kuksis, 'Isolation of Purified Brush-border Membranes from Rat Jejunum Containing a Ca^{2+}-independent Phospholipase A2 Activity', *Biochim. Biophys. Acta.*, **901**(1), 78–87 (1987).
82. G.B. Phillips, J.T. Dodge, 'Composition of Phospholipids and of Phospholipid Fatty Acids of Human Plasma', *J. Lipid Res.*, **8**(6), 676–681 (1967).
83. V.P. Skipski, 'Lipid Composition of Lipoproteins in Normal and Diseased States', in *Blood Lipids and Lipoproteins: Quantitation, Composition, and Metabolism*, ed. G.J. Nelson, Wiley-Interscience, New York, 471–583, 1972.
84. W. Pruzanski, F.C. de Beer, M.C. de Beer, E. Stefanski, A. Ravandi, A. Kuksis, 'Comparative Analysis of Lipid Composition of Normal and Acute Phase High Density Lipoproteins', *J. Lipid Res.* (in press).
85. A.G. Mallinger, J.K. Yao, A.S. Brown, C.S. Dippold, 'Analysis of Complex Mixtures of Phospholipid Classes from Cell Membranes Using Two-dimensional Thin-layer Chromatography and Scanning Laser Densitometry', *J. Chromatogr.*, **614**(1), 67–75 (1993).
86. K.T. Mitchell, J.E. Ferrell, Jr, W.H. Huestis, 'Separation of Phosphoinositides and Other Phospholipids by Two-dimensional Thin-layer Chromatography', *Anal. Biochem.*, **158**(2), 447–453 (1986).
87. V.G. Mahadeveppa, B.J. Holub, 'Chromatographic Analysis of Phosphoinositides and their Breakdown Products in the Activated Blood Platelets/Neutrophils', in *Chromatography of Lipids in Biomedical Research and Clinical Diagnosis*, ed. A. Kuksis, Elsevier, New York, 225–265, 1987.
88. H. Kamido, K. Nonaka, A. Kuksis, L. Marai, A. Ravandi, 'Identification of lipid ester core aldehydes and hydroperoxides in human atherosclerotic lesions', *Circulation*, **94**(Suppl.), I-707 (1996).
89. J. Schiller, J. Arnold, S. Bernard, Muller, S. Reichl, K. Arnold, 'Lipid Analysis by Matrix-assisted Laser Desorption and Ionization Mass Spectrometry', *Anal. Biochem.*, **267**, 46–56 (1999).
90. J.E. Ferrell, Jr, W.H. Huestis, 'Phosphoinositide Metabolism and the Morphology of Human Erythrocytes', *J. Cell Biol.*, **98**(6), 1992–1998 (1984).
91. H. Hegewald, 'One-dimensional Thin-layer Chromatography of All Known D-3 and D-4 Isomers of Phosphoinositides', *Anal. Biochem.*, **242**(1), 152–155 (1996).

92. H. Hegewald, E. Muller, R. Klinger, R. Wetzker, H. Frunder, 'Influence of Ca^{2+} and Mg^{2+} on the Turnover of the Phosphomonoester Group of Phosphatidylinositol 4-Phosphate in Human Erythrocyte Membranes', *Biochem. J.*, **244**(1), 183–190 (1987).
93. H. Binder, P.C. Weber, W. Siess, 'Separation of Inositol Phosphates and Glycerophosphoinositol Phosphates by High-performance Liquid Chromatography', *Anal. Biochem.*, **148**(1), 220–227 (1985).
94. L. Marai, A. Kuksis, 'Molecular Species of Lecithins from Erythrocytes and Plasma of Man', *J. Lipid Res.*, **10**(2), 141–152 (1969).
95. A. Kuksis, J.J. Myher, K. Geher, 'Quantitation of Plasma Lipids by Gas–Liquid Chromatography on High Temperature Polarizable Capillary Columns', *J. Lipid Res.*, **34**(6), 1029–1038 (1993).
96. G.M. Patton, J.M. Fasulo, S.J. Robins, 'Separation of Phospholipids and Individual Molecular Species of Phospholipids by High-performance Liquid Chromatography', *J. Lipid Res.*, **23**(1), 190–196 (1982).
97. G.M. Patton, S.B. Clark, J.M. Fasulo, S.J. Robins, 'Utilization of Individual Lecithins in Intestinal Lipoprotein Formation in the Rat', *J. Clin. Invest.*, **73**(1), 231–240 (1984).
98. A. Ravandi, A. Kuksis, N.A. Shaikh, 'Glucosylated Glycerophosphoethanolamines are the Major LDL Glycation Products and Increase LDL Susceptibility to Oxidation: Evidence for their Presence in Atherosclerotic Lesions', *Arterioscler. Thromb. Vasc. Biol.* (in press).
99. A. Ravandi, A. Kuksis, N.A. Shaikh, 'Glycated Phosphatidylethanolamine Promotes Macrophage Uptake of Low Density Lipoprotein and Accumulation of Cholesteryl Esters and Triacylglycerols', *J. Biol. Chem.*, **274**(23), 16 494–16 500 (1999).
100. P. Child, J.J. Myher, F.A. Kuypers, J.A. Op den Kamp, A. Kuksis, L.L. van Deenen, 'Acyl Selectivity in the Transfer of Molecular Species of Phosphatidylcholines from Human Erythrocytes', *Biochim. Biophys. Acta.*, **812**(2), 321–332 (1985).
101. G.B. Phillips, J.T. Dodge, 'Composition of Phospholipids and of Phospholipid Fatty Acids of Human Plasma', *J. Lipid Res.*, **8**, 676–681 (1967).
102. P.K. Schick, B.P. Schick, G. Brandeis, D.C.B. Mills, 'Distribution of Phosphatidylethanolamine Arachidonic Acid in Platelet Membranes', *Biochim. Biophys. Acta*, **643**, 659–662 (1981).
103. A. Kuksis, J.J. Myher, K. Geher, 'Quantitation of plasma lipids by gas–liquid chromatography of high temperature polarizable capillary columns', *J. Lipid Res.*, **34**(6), 1029–1038 (1993).
104. K. Samuelsson, 'Identification and Quantitative Determination of Ceramides in Human Plasma', *Scand. J. Clin. Lab. Invest.*, **27**(4), 371–380 (1971).
105. F.B. Jungalwala, J.E. Evans, R.H. McCluer, 'Compositional and Molecular Species Analysis of Phospholipids by High-performance Liquid Chromatography Coupled with Chemical Ionization Mass Spectrometry', *J. Lipid Res.*, **25**(7), 738–749 (1984).
106. W.C. Fountain, J.R. Requena, A.J. Jenkins, T.J. Lyons, E.M. Smyth, J.W. Baynes, S.R. Thorpe, 'Quantification of N-(glucitol)ethanolamine and N-(carboxymethyl)serine: two products of non-enzymatic modification of aminophospholipids formed in vivo', *Anal. Biochem.*, **272**, 48–55 (1999).
107. Z. Ahmed, A. Ravandi, G. Maguire, A. Kuksis, P.W. Connelly, 'Characterization of HDL Phosphatidylcholine Oxidation by Peroxynitrite Donor' (submitted).
108. S.M. Lynch, J.D. Morrow, L.J. Roberts, B. Frei, 'Formation of Non-cyclooxygenase-derived Prostanoids (F2-isoprostanes) in Plasma and Low Density Lipoprotein Exposed to Oxidative Stress In Vitro', *J. Clin. Invest.*, **93**(3), 998–1004 (1994).
109. K. Akasaka, A. Ohata, H. Ohrui, H. Meguro, 'Automatic Determination of Hydroperoxides of Phosphatidylcholine and Phosphatidylethanolamine in Human Plasma', *J. Chromatogr. B Biomed. Appl.*, **665**(1), 37–43 (1995).
110. A. Ravandi, A. Kuksis, N.A. Shaikh, 'Analysis of Core Aldehydes and Isoprostane Esters of Phospholipids in Plasma Lipoproteins and Human Atheroma by Electrospray Mass Spectrometry with Single Ion Monitoring', *Clin. Chem.*, **45**, A16 (1999).
111. D.A. Ford, K.B. Rosenbloom, R.W. Gross, 'The Primary Determinant of Rabbit Myocardial Ethanolamine Phosphotransferase Substrate Selectivity is the Covalent Nature of the sn-1 Aliphatic Group of Diradyl Glycerol Acceptors', *J. Biol. Chem.*, **267**(16), 11 222–11 228 (1992).
112. F. Nazih-Sanderson, G. Pinchon, S. Nion, J.C. Fruchart, C. Delbart, 'HDL3-signalling in HepG2 Cells Involves Glycosyl-phosphatidylinositol-anchored Proteins', *Biochem. Biophys. Acta*, **1346**(1), 45–60 (1997).
113. K. Mensa-Wilmot, J.C. Morris, A. Al Qahtani, P.T. Englund, 'Purification and use of Recombinant Glycosylphosphatidylinositol-phospholipase C', *Methods Enzymol.*, **250**, 641–655 (1995).
114. M.A. Deeg, M.A. Davitz, 'Glycosylphosphatidylinositol-phospholipase D: A Tool for Glycosylphosphatidylinositol Structural Analysis', *Methods Enzymol.*, **250**, 630–640 (1995).

Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Joanna M. Schulman and Stephen G. Schulman
University of Florida, Gainesville, USA

1 Introduction 1
2 Direct Fluorimetry, Phosphorimetry, and Chemiluminometry 2
3 Indirect Methods 3
4 Luminescence Detection in Chromatography and Other Separation Methods 6
5 Fluorescence, Phosphorescence, and Chemiluminescence Detection in Immunoanalysis 7
6 Time- and Phase-resolved Fluorimetry 9
7 Fiber-optic Fluorescence Spectroscopy 10
8 Semiconductor Lasers and Red and Near-infrared Emitting Labels 10
Abbreviations and Acronyms 11
Related Articles 11
Further Reading 11
References 12

Molecular luminescence spectroscopy, including fluorescence, phosphorescence, and chemiluminescence (CL), is among the most sensitive of analytical methods. Its use in clinical chemistry has been extensive since the late 1950s and is constantly expanding. Modern separation and immunoassay methods coupled with luminescence detection have revolutionized the detection and determination of endogenous substances, xenobiotics, and their metabolites. Newer techniques such as time-resolved fluorimetry and phosphorimetry, detection by fiber-optic sensors, and derivatization with infrared (IR)-emitting dyes are expanding the range of clinical analysis into new realms of analytical selectivity, sensitivity, and expediency on a scale unimagined until recently. This article considers some of the aspects of fluorescence, phosphorescence, and CL spectroscopy that have contributed to important recent advances in clinical analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Fluorescence spectroscopy and the closely related areas of phosphorescence spectroscopy and CL spectrometry have become firmly established and widely employed techniques in clinical analytical chemistry. Fluorimetry and CL spectrometry are now routinely used in the detection, quantitation, identification, and characterization of inorganic and organic compounds and of biological structures and processes. Phosphorimetry, although somewhat less popular, has the potential to be as useful as the other two areas if problems associated with sampling can be overcome.

Not long ago, it appeared that luminescence spectroscopy had reached an impasse as an area of active interest in clinical analytical chemistry. Luminescence spectra lacked the information content of, say, mass spectra or nuclear magnetic resonance spectra and quantitative measurements were dramatically influenced by contaminants in solution. Coincident spectra from analytes with nearly identical fluorophores (e.g. catecholamines) made multicomponent analysis nearly impossible.

CL analysis was based on very few luminescent molecules and derivatization of analytes with these luminophores was a major problem. Phosphorescence spectroscopy required the use of very low temperatures. The use of narrow-bore cylindrical quartz tubes (square cells crack at the joints in the freeze–thaw cycle) as cells for phosphorimetry entailed major difficulties with cleaning and filling and also with repositioning for reproducible measurements.

In the past three decades, much of this has changed. Thanks to recent advances in instrumentation and in new techniques, fluorescence and CL spectroscopy can be routinely applied to real analytical problems. In liquid chromatography and immunoassay, fluorescence introduces ultrasensitive detection to extremely selective separation methods. CL spectrometry can also often be coupled with the selectivity of the immune response to allow sensitive quantitation, often without chemical or physical separation of the analyte from an extremely ''dirty'' matrix. CL techniques are also valuable owing to their sensitivity (attomole and subattomole detection limits), their speed (a signal can be generated in a few seconds), their stablity (in some cases signals last for several hours), their relatively uncomplicated procedures, and their use of nonhazardous reagents.

With modern spectroscopic instrumentation, several analytes having nearly identical fluorescence spectra can often be determined simultaneously by taking advantage of differences in their fluorescent decay times or differences between the phase angles in their excitation and emission spectra rather than between their fluorescence spectral band intensities. Phosphorimetry, on the other

hand, seems to be advancing at a slower pace than the other branches of luminescence spectroscopy, although the observation of phosphorescence (albeit weak in intensity) at ambient temperatures on solid surfaces and in micellar solutions shows some promise. Moreover, the microsecond to millisecond duration phosphorescent emissions of some lanthanide chelates, notably those of europium(III) and terbium(III), can be intense and both their intensity and temporal characteristics have been used in clinical analysis.[1,2]

In this article, several modern aspects of luminescence spectroscopy are considered, especially as they pertain to clinical analysis of endogenous substances and of drugs and their metabolites. It is assumed that the reader already has some familiarity with the photophysical, photochemical, and instrumental fundamentals of these spectroscopic methods.

2 DIRECT FLUORIMETRY, PHOSPHORIMETRY, AND CHEMILUMINOMETRY

The systematic study of the production of light by electronically excited molecules is known as fluorescence, phosphorescence, or CL spectroscopy or collectively as luminescence spectroscopy. Luminescence processes can be classified according to the origin of the excitation energy in photoluminescence. In what is usually called fluorescence and phosphorescence, excitation occurs via the absorption of light by a potentially luminescent molecule. CL, on the other hand, entails excitation as the result of an energetic chemical reaction, almost invariably an oxidation reaction.

The simplest way in which to carry out luminescence spectroscopy is by direct photoexcitation or reaction of the analyte to be quantitated. However, if the analyte is not intrinsically fluorescent or phosphorescent, or if it does not form a luminescent product by direct oxidation, more complicated indirect methods entailing chemical derivatization or energy transfer may be necessary. These will be considered in the next section. Owing to its experimental simplicity, most luminescence methods entail measurement of the fluorescence of organic compounds of clinical chemical interest in which the excitation energy is provided by the absorption of visible or ultraviolet (UV) light.

Fluorescent methods comprise the great majority of the ultrasensitive assays carried out by light envision spectroscopy in the clinical sciences. However, in recent years there has been considerable interest in phosphorescence and in CL. Accordingly, these subjects will also be considered here.

Luminescence spectroscopy is used more often in quantitative analysis than in any other application. The quantitative relationship between fluorescence intensity, F, and analyte concentration, C, derives from the Beer–Lambert law (Equation 1):

$$I = I_o 10^{-\varepsilon cl} \tag{1}$$

where I and I_o are the intensities of exciting light transmitted through and incident upon the sample, respectively, l is the pathlength of light through the sample and ε is the molar absorptivity of the molecular species of interest at the nominal wavelength of excitation. The intensity of light absorbed, I_a, is given by Equation (2):

$$I_a = I_o - I_o 10^{-\varepsilon cl} \tag{2}$$

If all molecules that absorb light fluoresced, I_a would be the intensity of fluorescence. However, only the fraction, ϕ_f (the quantum yield of fluorescence), fluoresces. The remaining fraction returns to the ground state by nonradiative means. Hence (Equation 3):

$$F = \phi_f I_o (1 - 10^{-\varepsilon cl}) \tag{3}$$

The exponential term of Equation (3) can be expanded in a Maclaurin series so that (Equation 4):

$$F = \phi_f I_o \quad (2.3\varepsilon cl) - \frac{(2.3\varepsilon cl)^2}{2!} + \frac{(2.3\varepsilon cl)^3}{3!} - \cdots \tag{4}$$

If the absorbance (εcl) of the sample is <0.02, then to within 2.3% error, Equation (4) becomes the linear Equation (5):

$$F = 2.3 f \phi_f I_o \varepsilon cl \tag{5}$$

It then remains only to prepare a standard solution of concentration C_s and having fluorescence intensity F_s to be compared with the unknown solution of concentration C_u and fluorescence intensity F_u according to Equation (6):

$$C_u = \frac{F_u}{F_s} C_s \tag{6}$$

However, it is generally prudent to prepare a calibration curve for several standard solutions to check the linearity of the F versus C curve, before attempting the simple relative fluorimetry expressed in Equation (6) for a large number of samples.

In the case of phosphorescence, the measured intensity, P, is related to the concentration of analytes following similar lines of reasoning applied to fluorescence, according to Equation (7):

$$P = \phi_{st} \phi_p I_o (1 - 10^{-\varepsilon cl}) \tag{7}$$

Table 1 Drugs that exhibit direct CL

Analyte	Method[a]	Limit of detection	Range	Refs.
Tetracycline	Reaction with bromine in alkaline medium, FIA	19 μg mL^{-1} (4×10^{-5} M)	5×10^{-5}–1×10^{-2} M	4
Acetaminophen	Reduction of cerium(IV), FIA	0.070 μg mL^{-1}	1.00–10.0 μg mL^{-1}	5
Isoniazid	Oxidation with *N*-bromosuccinimide	0.024 μg mL^{-1}	0.050–20.0 μg mL^{-1}	6
Morphine	Reaction with permanganate, FIA, HPLC	0.7 pg (2 fmol, 1×10^{-10} M) for FIA 25 ng mL^{-1}, 50 ng mL^{-1} (aqueous solution, biological fluids) for HPLC	50 ng mL^{-1}–500 μg mL^{-1}	7, 8
Loprazolam	Reaction with permanganate, FIA	163 ng (7×10^{-6} mol/50 μL)	1×10^{-5}–1×10^{-3} M	9

[a] FIA, flow injection analysis; HPLC, high-performance liquid chromatography.

where ϕ_{st} and ϕ_p are the quantum yields of intersystem crossing (the fraction of excited molecules radiationlessly converted from the lowest excited singlet state to the lowest triplet state) and phosphorescence (the fraction of molecules arriving in the lowest triplet state that are deactivated by phosphorescence), respectively. As for fluorescence, Equation (7) can be expanded and for weakly absorbing solutions can be made analogous to Equation (6) according to Equation (8):

$$C_u = \frac{P_u}{P_s} C_s \tag{8}$$

where P_u and P_s are the phosphorescence intensities of the unknown and standard samples, respectively. Equation (8) is, therefore, the basis of simple quantitative phosphorimetry.

CL does not depend at all on the intensity of the exciting light as no photoexcitation occurs. Rather, the intensity of light emission observed from a chemi-excited luminescence process (I_{CL}) in which one of the direct products of the reaction fluoresces is governed by the efficiencies of the chemical reaction leading to the luminescent product (ϕ_c), formation of electronically excited molecules in the chemi-excitation process (ϕ_e), and fluorescence of the chemi-excited product (ϕ_f) as well as the rate at which the analyte is consumed in the chemiluminescent reaction (d[L]/dt) (Equation 9):

$$I_{CL} = \phi_c \phi_e \phi_f \left(-\frac{d[L]}{dt} \right. \tag{9}$$

If the CL reaction conditions are controlled so that it is carried out as pseudo-first-order disappearance of L, then (Equation 10):

$$-\frac{d[L]}{dt} = k_L[L] \tag{10}$$

where k_L is the pseudo-first-order rate constant for consumption of L. We then have Equation (11):

$$I_{CL} = \phi_c \phi_e \phi_f k_L[L] \tag{11}$$

which is linear in [L] and if the chemiluminescent reaction is carried out for a solution containing a standard concentration of L ($[L]_s$) an unknown concentration of L ($[L]_u$) can be determined from Equation (12):

$$[L]_u = [L]_s \frac{I_{CLu}}{I_{CLs}} \tag{12}$$

where I_{CLu} and I_{CLs} are the CL intensities of unknown and standard samples, respectively, after allowing both reactions to proceed for the same period of time.

There are two broad categories of luminescence analysis, namely direct and indirect. The direct luminescence method uses the native luminescence of the analyte and is limited to (but not all inclusive of) those molecules containing aromatic rings or highly conjugated aliphatic systems. The fluorescence of proteins, for example, is due almost entirely to emissions from phenylalanine, tyrosine, and tryptophan. Compensating for the lack of universality of the directly excited luminescence is the specificity and great sensitivity of analysis resulting from the luminescence properties inherent to the analyte of interest.

An example representative of direct fluorimetric analysis is the assay of lysergic acid diethylamide (LSD). LSD is highly fluorescent with excitation maximum at 325 nm and emission maximum at 445 nm. Direct fluorimetric estimation of LSD in tissues can determine as little as 3 ng of the compound per gram of tissue.[3] Table 1 gives several examples of drugs that can be analyzed using direct CL.

3 INDIRECT METHODS

Indirect fluorimetric methods are used for analytes that are weakly fluorescent or nonfluorescent. This approach

involves either the conversion of the analyte to a fluorescent derivative, using an appropriate reaction scheme, or the utilization of the capability of the analyte to influence (enhance or quench) the fluorescence of a fluorescent dye.

Conversion of the analyte to a fluorescent derivative can be achieved in several ways. Oxidation, for example, is used to convert phenothiazines into the corresponding sulfoxides, which are highly fluorescent.[10] Morphine can be analyzed fluorimetrically by oxidizing it to fluorescent pseudomorphine.[11] Dehydration of cardiotonic steroids with concentrated H_2SO_4 yields fluorescent products and this reaction has been used to develop several analytical methods for steroids.[12]

Some compounds can be converted to fluorescent derivatives by treating them with strong mineral acids. For example, morphine gives a highly fluorescent derivative after heating it in concentrated sulfuric acid.[13] The reaction product fluoresces when the solution is made alkaline. Codeine, when heated in concentrated sulfuric acid followed by making the solution alkaline, shows analytically useful fluorescence.[14] Chlorprothixene, a thioxanthene derivative used as a major tranquilizer, gives a fluorescent product in concentrated phosphoric acid.[15] Reserpine can be converted to a fluorescent derivative by treating it with nitrous acid.[16]

Weakly fluorescent or nonfluorescent primary arylamines[17] and arylhydrazines[18] can often be condensed with aldehydes or ketones to yield fluorescent derivatives. Aldehydes of analytical interest usually do not fluoresce. They may be condensed with arylhydrazines, such as naphthylhydrazines, or arylamines to yield fluorescent products. A general fluorimetric procedure for assaying primary amines using fluorescamine reagent has been developed.[19,20] Fluorescamine is a high-sensitivity fluorogenic reagent that reacts directly with primary amines to form fluorophores. *o*-Phthalaldehyde and its benzolog 2,3-naphthalenedialdehyde have also been used extensively to form fluorescent azine condensation products with analytes having primary alkylamino groups.[21]

Chelation of the analyte with certain metal ions yields analytically useful fluorescent chelates. For example, the magnesium chelates of tetracyclines are highly fluorescent and have been used for the fluorimetric analysis of tetracyclines.[22]

Analytes that have cationic groups can interact with water-soluble anionic fluorescent dyes, such as eosine, to form neutral complexes which will be extractable into low-dielectric organic solvents. Since the ionic dye will not be extracted alone into the organic solvent, the fluorescence of the extract will be a measure of the concentration of the cationic drug. For example, α-tubocurarine can be analyzed fluorimetrically after co-extraction into chloroform with the anionic dye Rose Bengal.[23]

Some nonfluorescent analytes may interact with fluorescent dyes and quench or enhance the fluorescence of the latter. The extent of quenching or enhancement can be quantitatively related to the concentration of the quencher. For example, halide ions such as Cl^- form precipitates with Ag(I). The adsorption of highly fluorescent fluorescein on the silver chloride particles quenches the fluorescence of the fluorescein and, consequently, can be used to quantitate Cl^-.[24]

The earliest use of the indirect CL method was made using the adenosine triphosphate (ATP)-dependent firefly system, which employs the enzyme firefly luciferase in a reaction that uses ATP and generates visible light. As little as 10^{-15} mol of ATP can be measured using the following firefly luciferase reaction[25] (Equations 13 and 14):

$$\text{ATP} + \text{luciferin} \xrightarrow[Mg^{2+}]{\text{luciferase}} \text{adenylluciferin} \qquad (13)$$

$$\text{adenylluciferin} + O_2 \longrightarrow \text{adenyloxyluciferin} + \text{light} \qquad (14)$$

This CL assay has had a huge impact in the medical community, especially in the fields of immunology and pathology. It has numerous applications because ATP is present in all living cells. This method has been used to measure intracellular ATP in bacterial cultures, allowing an estimate of bacterial growth. It has also been used to determine bacteriuria and the concentration of antibiotics in a serum specimen.[25–28] It is possible to investigate the growth and physiology of mycobacteria and filamental organisms, which are difficult to study using classical techniques.[25,29] The determination of ATP is also useful in the study of the ATP content of erythrocytes in various pathological situations.[30,31] It is used to study the critical effect of ATP and adenosine diphosphate (ADP) on platelet aggregation in vitro[32] and in monitoring of the release of ATP by platelets under the influence of thrombin.[33] It is used in studies regarding shape and lysis of erythrocytes[34] and it is helpful in determining the viability of stocked erythrocytes, leukocytes, and platelets.[35] Assessment of cell death,[36] evaluation of the viability of spermatozoa,[37] and monitoring of the leakage of ATP from erythrocytes during immunolysis are other applications of the CL assay of ATP using firefly luciferase.[38]

The indirect method has an advantage over the direct method in that, provided that a suitable reactant can be found to produce a fluorescent derivative, virtually any drug or metabolite having functional groups can be made amenable to fluorimetric analysis. However, it should be borne in mind that derivatization methods are most often applicable to classes of analytes rather than specific analytes and that the fluorescence of the derivative

usually arises from an electronic transition localized on the derivatizing reagent. As a result, the gain in generality of this method is partially offset by the sacrifice of selectivity. Moreover, because the derivatization reaction is invariably at least bimolecular, the maximum fluorescence intensity obtainable from the derivative will be limited by the concentration of the analyte, which at very low concentrations affects not only spectroscopic detectability but also the completeness of the derivatization reaction. The latter will generally decrease as the analyte concentration decreases. Consequently, indirect luminescence spectroscopy will, in many cases, be less sensitive than direct fluorimetry or phosphorimetry.

Up to the present, discussion has been confined to the determination of the absolute amount of metabolite recovered from a biological sample. However, the most elegant metabolic studies may entail the need for information about the relative fraction of drugs in true solution in the interstitial and intracellular fluids and bound to cellular and extracellular structures such as proteins, nucleic acids, and membranes. Some drugs containing aromatic rings demonstrate alterations in their luminescent properties, such as change in fluorescence efficiency upon passing from aqueous solution to a bound condition on cellular structures. The reason for the alterations of the luminescence properties of these compounds upon binding has been a subject of much research interest. Although it is generally conceded that changes in the intensity and energy characteristics of the fluorescences of these molecules upon binding are due to environmental effects, the specific nature of these effects is the subject of much controversy. Arguments involving viscosity, hydrogen bonding, environmental dielectric strength, and excited-state proton transfer have been invoked, but none of these alone adequately explains all of the observed effects. Nevertheless, the fact that the luminescence properties of drugs are altered upon binding may prove to be of substantial analytical value. For example, the dramatic increase in fluorescence efficiency that occurs in 8-anilino-1-naphthalenesulfonic acid upon binding to serum albumin forms the basis for the fluorimetric analysis of these proteins.[39]

Just as indirect fluorimetry and phosphorimetry entail the derivatization of nonluminescing analytes, CL can have an indirect mode in which derivatization is employed. In fact, owing to the paucity of molecules that demonstrate intense CL, it is often necessary to derivatize nonchemiluminescent molecules with standard intensely emitting chemiluminophores such as luminol (3-aminophthalhydrazide) or isoluminol (4-aminophthalhydrazide). These two CL molecules have been used most often as coupling to other molecules via their primary arylamino group often does not interfere with their CL properties.[40,41]

There is a whole class of chemiluminescent reaction partners which do not demonstrate CL directly. These are the dioxetanes[42] and their precursors such as bis(trichlorophenyl) oxalate (TCPO) and bis(2,4-dinitrophenyl) oxalate (DNPO), which are oxidized to dioxetanes by hydrogen peroxide, superoxide ion, and occasionally even molecular oxygen.[43,44]

The dioxetanes form electronically excited fragments when oxidized. However, in order to see analytically useful luminescence there must usually be present, in solution, a suitable energy acceptor that is also a good fluorophore. The acceptor, usually a fluorescein or rhodamine derivative, collects the excitation energy from the electronically excited dioxetane fragment and after thermal relaxation to the lowest vibrational level of its lowest excited single state, may emit the excess energy as fluorescence. The analyte is usually the oxidizer, an enzyme, or a metal ion catalyzing the oxidation, but may

Table 2 Drugs that can be analyzed using an indirect CL system

Analyte	Method	Limit of detection	Range	Refs.
Penicillin	Enhancement of luminol CL reaction, batch method	$100\,ng\,mol^{-1}$	$100\,ng\,mL^{-1}-100\,\mu g\,mL^{-1}$	45, 46
Cephalothin sodium	Enhancement of luminol CL reaction, batch method	–	–	47
Cortisone	Sensitization of CL of sulfite by cerium(IV), FIA	$0.040\,\mu g\,mL^{-1}$	$0.100-1.00\,\mu g\,mL^{-1}$	48
Hydrocortisone		$0.028\,\mu g\,mL^{-1}$		
Prednisolone		$0.016\,\mu g\,mL^{-1}$		
Methylprednisolone		$0.021\,\mu g\,mL^{-1}$		
Dexamethasone		$0.16\,\mu g\,mL^{-1}$	$0.500-5.00\,\mu g\,mL^{-1}$	
Betamethasone		$0.30\,\mu g\,mL^{-1}$		
Quinine	Sensitization of CL of sulfite by cerium(IV), FIA	$0.64\,\mu g\,mL^{-1}$	$5.00-500\,\mu g\,mL^{-1}$	49
Hydrocortisone	Solid-state peroxyoxalate CL, HPLC	$2\,ng\,mL^{-1}$	–	50
Digoxin		$2\,ng\,mL^{-1}$		
Theophylline		$4\,ng\,mL^{-1}$		

also be the fluorophore. The quantitative relationship between analyte concentration and CL intensity is similar to that in Equation (11) but now the efficiency of energy transfer is also included in the constant of proportionality.

Table 2 gives examples of drugs that can be analyzed using indirect CL.

4 LUMINESCENCE DETECTION IN CHROMATOGRAPHY AND OTHER SEPARATION METHODS

A major limitation of the application of luminescence spectroscopy to clinical analysis is the lack of specificity owing to the similarities in spectral band shapes and spectral positions of the luminescence spectra of many compounds. An obvious way to solve this problem is to separate the analytical sample's interfering constituents from each other prior to quantitation by luminescence. HPLC and related separation methods can be coupled to luminescence spectroscopy in order to take advantage of the sensitivity of the latter and the specificity of the former.

Luminescence affords a very sensitive and selective means of detection in flowing systems such as HPLC, electrophoresis, FIA, and flow cytometry.[51] HPLC fluorescence detectors are similar in operation to normal fluorimeters. Most fluorescence detectors use filters for crude monochromation. At somewhat greater cost, diffraction grating monochromated instruments are available specifically for HPLC. Filters generally pass light in a wider band than do the monochromators. This is often an advantage since the detector does not need to be specifically tuned for each compound as it is eluted. Filters favor spectral sensitivity because more light excites the sample and is collected by the detector. Grating monochromators favor selectivity, which is preferable depending upon the requirements of a specific analysis. For detectors used at concentrations where the absorbance at the excitation wavelength is <0.02, fluorescence intensity is linear with concentration. The fluorimetric detector is susceptible to the usual interferences which hinder fluorescence measurements, mainly background fluorescence and quenching.

In the operation of a fluorimetric HPLC detector the light from a UV or visible source is monochromated to some degree and focused on the cell which is on-line with and subsequent to the separation column. Fluorescence is emitted by the sample in all directions so that the emitted light can be measured with the detector at right-angles to the path of the exciting light and the direction of fluid flow. Stray or scattered excitation light is then blocked from the detector by a filter, and the intensity of the emitted energy is measured by a photocell. It is important to recognize that the solvent has a strong effect on the intensity of fluorescence. For instance, quinoline is nonfluorescent in hexane, but fluorescent in ethanol. The solvent also scatters light owing to the Raman effect. This light is of slightly lower energy than the exciting light. For this reason, the emission wavelength is selected to be longer than that of the scattered Raman light.

Flow cells have been designed which can be used for both fluorescence and absorption measurements. Quantitation can be considerably improved by simultaneously monitoring the absorbance and fluorescence signals, which extends the linear dynamic range for fluorescent samples. At high sample concentrations, where the absorbance at the excitation wavelength is >0.02, fluorescence intensity becomes nonlinear with concentration. At these higher concentrations, however, the light absorbance of the sample is often measurable and linear with concentration. Fluorescence, however, is often detectable at concentrations 10^3–10^9 times lower than those where absorbance is detectable. Fluorescence spectroscopy using flow cells in the chromatographic path and postcolumn with monochromated or filtered arc-lamp excitation sources has been applied to detection in HPLC for several decades and a wide variety of commercial detectors are currently on the market.

The laser, as an excitation source for luminescence detection in chromatography, has become popular because of its high excitation power and its highly focused beam. The laser, although many orders of magnitude more powerful than an arc lamp as an excitation source, has, so far, produced detectability only a few orders of magnitude better than that achieved with arc lamps. This is a result of high background luminescence produced by photo and thermal decomposition of solution components induced by the large energy output of the laser. Thermal lensing (changes in solution refractivity) also produces spectroscopic complications. On the other hand, the high degree of collimation of the laser beam has made it an excellent excitation source with the microbore cells and tubing used in capillary zone electrophoresis (CZE). A typical laser beam can be focused into a spot no more than a few micrometers in diameter. With laser excitation and fluorescence detection CZE, analyte amounts below 1 amol (at nanomolar concentrations) have been detected.

The advantages offered by the power and high degree of collimation of the laser as an excitation source can be exploited to measure the properties of a host of particles, including cells, algae, bacteria, and larvae, which either show native fluorescence or can form fluorescent adducts. A well-defined flow of isolated particles is achieved

by using hydrodynamic focusing. For this purpose, a sample flow is obtained within a larger ''sheath'' flow of different specific density, which allows flow of the sample cells into a well-defined part of the cuvette, where a laser excitation-based analysis is performed. The single cell passage is guaranteed by adjusting the two flows, hence collision between two cells (particles) is avoided. The method allows high-speed analysis of the cells and is called ''flow cytometry'' or ''fluorescent cell sorting''.[51]

CL has been coupled to HPLC with promising results.[52,53] However, the multitude of reagents required to generate CL, especially when energy transfer from the primary chemiluminescing molecule to a better fluorescing acceptor is required, necessitates the use of several pumping systems and a very complex, if not unwieldy, chromatographic system. Moreover, although HPLC (and FIA) provide CL spectrometry with a degree of specificity not inherent to the spectroscopic technique alone, the dynamic characteristics of the flowing separation systems reduce the residence time of the luminescing system in the detector. It is precisely the feature of ability to observe and integrate the CL signal over the entire course of the lumigenic reaction that gives CL its great sensitivity, although even static CL methods are often based on instantaneous measurements for the sake of expedience.

Detection of phosphorimetry in HPLC is an active area of current research interest but is not yet in routine analytical use. To date, much of the work in this area has centered on the use of biacetyl (2,3-butanedione) as a phosphorophore. This molecule is one of very few that phosphoresce in fluid solution at ambient temperatures although rigorous deoxygenation of the solvent is necessary to observe biacetyl phosphorescence. Emissions from phosphorophores such as quinoline entrapped in micelles or cyclodextrins,[54,55] to protect them from dissolved oxygen and, hence, quenching, have been observed but are inadequate for sensitive chromatographic detection. Phosphoresence efficiencies at room temperature tend to be low. A more promising approach is the use of a phosphorescent solute to provide a background phosphorescence which can be quenched by the analyte in the detector cell. Because organic molecules that phosphoresce in fluid solution at ambient temperatures have lifetimes of the order of milliseconds (fluorescent molecules have lifetimes of the order of nanoseconds), dynamic quenching of phosphorescence can be about a 10^6-fold more sensitive than that of fluorescence – so, in principle, although it is impossible to detect quenching of fluorescence at concentrations of lower than, say, 10^{-4} mol L^{-1}, quenching of phosphorescence can be detected at concentrations $> 10^{-10}$ mol L^{-1}.

5 FLUORESCENCE, PHOSPHORESCENCE, AND CHEMILUMINESCENCE DETECTION IN IMMUNOANALYSIS

Immunoassay methods have assumed a role of major importance in clinical analysis. Their ability specifically to quantitate analytes in biological fluids has proven invaluable in the field of biomedical analysis. Their low cost and amenability to automation and also to simple ''dipstick'' methods, coupled with their sensitivity and specificity, have often permitted the circumvention of tedious chromatographic methods in diagnostics.

Immunoassay derives from the ability of mammals to biosynthesize specialized water-soluble globular proteins (antibodies) to complex invasive materials (antigens). The number of molecules capable of eliciting antibody formation (immunogens) is vast, and comprises naturally occurring macromolecules or cells containing such macromolecules on their surfaces.

An antigen is any substance capable of reacting with an antibody but not necessarily capable of introducing antibody formation. Therefore, all immunogens are antigens, although antigens may or may not be immunogens.

Antibodies are a group of heterogeneous, structurally related proteins derived from B-lymphocytes and known as immunoglobulins (Igs), having molecular weights of 150 000–160 000 Da. Each Ig molecule contains two sites of binding which have variable amino acid sequences and appear to be responsible for the great specificity of binding common to antibodies. Binding of these sites by immunogens creates the molecular memory that converts the Igs to antibodies.

Antibodies developed in response to a given immunogen will recognize and bind to only a small section of the immunogen. The ability of a molecule to elicit antibody formation depends on its size. Molecules with molecular weights $< 10\,000$ Da are generally not immunogenic. However, an immunoresponse can be elicited by small molecules or haptens covalently coupled to macromolecules such as albumins or polypeptides. The antibodies produced may react with sites on the hapten even when the hapten is not coupled to the other carrier molecules.

Immunoanalytical methods arise from the competitive binding that occurs between a labeled and an unlabeled ligand for highly specific receptor sites on antibodies. The measurement of some physical or chemical property such as luminescence intensity associated with the label, as a function of unlabeled ligand (analyte) concentration, allows the construction of a calibration curve from which unknown ligand concentrations may be interpolated.

Distinction of the signal corresponding to either the bound or free labeled analyte from that of the total may sometimes be necessary or desirable. This may be

accomplished in two ways. The first involves physical separation of the antibody-bound fraction from the free fraction of labeled analyte. This can be accomplished by using a salt such as ammonium sulfate or a polymer such as poly(ethylene glycol), to precipitate the excess free antibody and analyte-complexed antibody, followed by centrifugation. Alternatively, solid-phase extraction, in which the analyte or antibody is attached to a solid surface (beads, tube wall, dipsticks, etc.) may be employed. Competition between reactants in the liquid phase for the complexant in the solid phase permits the physical separation of the labeled material in the two phases. These techniques are called heterogeneous immunoassays and are required in methods where the labels employed are difficult or impossible to distinguish when present as both free and antibody-bound labeled material. Homogeneous immunoassays, in which physical separation of bound from free labeled ligand is not required, constitute the second category. Signals which differ for the bound and free labeled ligands are obtained from the test solution which contains all the participating analytical species. If the label is a fluorophore, environmental effects imposed on specific label properties can lead to signal modification. For example, a polarization of fluorescence resulting from the increase in molecular volume associated with labeled ligand binding to antibody protein can be monitored. Alternatively, the partitioning of a fluorescent label into an antibody binding site from the bulk aqueous solution may shift, quench, or enhance its fluorescence.

Separation-free, homogeneous immunoassays offer several important advantages over heterogeneous methods. As no separation is involved, the time and cost required per assay is decreased. Additionally, since the physical transfer step is avoided, potential sample loss is reduced.

Fluorescence immunoassay entails the measurement of the fluorescence of a fluorescent label which is bound to an antigen or antibody engaged in competitive immunochemical binding and whose spectral properties vary with the state of complexation of the labeled entity. Fluorescent labels are used in homogeneous and heterogeneous immunoassay systems and may be bound to antigens, antibodies, or solid phases or they may exist free in solution as enzyme substrates.

For a fluorescent label to be used in homogeneous immunoanalysis, it should have certain attributes. As the fluorescence signal must often be measured in a partially opaque or light-scattering biological matrix, the label should have a high quantum yield of fluorescence. To minimize spectroscopic interferences the excitation and emission maxima of the probe should occur at wavelengths longer than those of the other components of the test solution. Excitation spectra of dilute serum, the most commonly tested matrix of biological origin, have excitation maxima at 280 and 340 nm and an emission maximum near 350 nm (also near 520 nm if bilirubin bound to albumin is present). For a fluorescent label to be useful in a serum solution it should emit at wavelengths longer than 400 nm or, even better, longer than 500 nm.

The most popular fluorescent labels for immunoassay have been those derived from the long-wavelength, strongly emitting xanthene dyes fluorescein and rhodamine B. The isothiocyanates or isocyanates of these fluorophores can label primary and secondary aliphatic amines in aqueous solutions. Consequently, they can be used to label a wide variety of analytes. Even those compounds which do not have indigenous alkylamino groups can often be labeled by introducing bridging groups (e.g. aminoethyl) which are amenable to coupling with the isothiocyanato or isocyanato functions.

Heterogeneous fluorescence immunoassay can be carried out with the aid of the separation procedures described earlier and may be desirable when there is not much distinction between the emissive properties of bound and free labels. Occasionally, a second antibody, directed at the antianalyte antibody, will be used in a ''double-antibody'' method to precipitate the bound labeled and unlabeled analyte or to alter the optical properties of the label in such a way as to make the analysis more sensitive. The assay of human chorionic gonadotropin, which forms the basis of most pregnancy tests, is carried out via a double antibody fluorescence immunoassay.[56]

Homogeneous fluorescence immunoassays can often be carried out even when there is no obvious change in the intensity of spectral position of fluorescence of the label upon binding to the antibody. If polymer films or nonlinear crystals are used to polarize the exciting light and then the fluorescence of the sample excited by polarized light is analyzed, it will usually be observed that the intensity of fluorescence reaching the detector will be lower from those samples in which a greater amount of antibody binding occurs. This is a result of the higher degree of polarization of the fluorescence emitted from the labels affixed to the slowly rotating macromolecular antibody. The polarized emission is more efficiently attenuated by the polarizing film whose optical axis is perpendicular to the plane of polarization of the fluorescence than is the unpolarized light emitted by the rapidly rotating, labeled analyte molecules which are not bound to macromolecules. This forms the basis of fluorescence polarization immunoassay (FPIA). The apparent decrease in fluorescence polarization with increasing analyte binding occurs as a result of decreasing labeled analyte binding and can be used to prepare a calibration curve from which the concentration of test samples can be determined when their polarized fluorescences are measured.

Although, in principle, it should be possible to use phosphorescent labels in immunoassay, in practice this is, with the exception of certain lanthanide chelates, not done. The principle reason for this is that in fluid, aerated solutions, very few molecules phosphoresce so that phosphorescence immunoassay as a simple, practical laboratory method would be difficult or impossible.

Ligand labeling with phosphorescent lanthanide chelates has created a form of phosphorescence immunoassay.[1,2] Some chelates of rare earth metals have unique emission characteristics in that upon excitation of aromatic portions of the ligands of the lanthanide complex, the energy of excitation is efficiently transferred to the lanthanide ion. This sensitizes f–f transitions localized on the metal ion which produce very narrow, almost line-like emission bands which arise from spin-forbidden transitions and which permit all of the emitted light to be collected by the detector with narrow emission slits. Eu(III) chelates give a red emission whereas Tb(III) chelates phosphoresce green. In addition, the rare earth chelates have very large Stokes' shifts, excitation usually being effected in the UV region. The excitation of these chelates is broad, which permits excellent sensitivity and selectivity by allowing the use of fairly wide bandwidths for excitation and narrow bandwidths for emission. With the proper combination of a rare earth chelate and a fluorimetric technique, sensitive immunoassays that avoid a large number of interferences can be developed.

CL immunoassay, a technique which is rapidly gaining popularity because of its great sensitivity (limits of detection of $\sim 1 \times 10^{-15}$ mol L^{-1} are sometimes attained), is in a sense a variation of fluorescence immunoassay.[57,58] CL manifests itself in the fluorescence of the oxidation products of some highly strained, highly reduced molecules (e.g. acridinium esters, peroxyoxalate esters, and amino-substituted cyclic hydrazides of phthalic acid). These oxidations are frequently catalyzed by metal ions and occur at appreciable rates only when the precursors of the species (the labels) and the oxidizers are freely diffusable in the solution (i.e. they will not generate light when bound to antibodies or perhaps prior to generation in an enzymic reaction). In this sense, they are analogous to many of the fluorescent labels. However, photofluorescence or photophosphorescence analysis entails the measurement of light emitted by a fraction of an even smaller fraction of molecules absorbing light for an instant. In CL it is possible to gather and integrate the light emitted in the chemiluminescent reaction over the entire course of the reaction. Consequently, very low detection limits can be attained if the chemiluminescent reactions have high luminescence efficiencies so that the choice of labels is much more restricted than in fluorescent immunoassay. Energy transfer to a strongly emitting secondary fluorophore from the originally chemi-excited product is sometimes a remedy for this.

6 TIME- AND PHASE-RESOLVED FLUORIMETRY

The fluorescence spectroscopy described so far is based upon the measurement of the intensity of fluorescence produced under "steady-state" conditions of excitation. "Steady-state" fluorimetry is derived from the excitation of the sample with a continuous temporal output of exciting radiation. The lamps and the power supplies used in conventional fluorimeters are sources of continuous radiation. After a short period of initial excitation of the sample, a steady state is established in which the rate of excitation of the analyte is equal to the sum of the rates of all processes deactivating the lowest excited singlet state including fluorescence. When the steady state is established, the observed fluorescence intensity becomes time invariant and produces the temporally constant signal which is measured by the photodetector. With the development of modern electro-optics however, it has become possible to excite a potentially fluorescent sample with a pulsed flash lamp which emits its radiation in bursts of 2–10 ns duration with about 0.2 ms between pulses. Pulsed lasers generate even shorter duration pulses (typically down to 1 ps). A fluorescent sample excited with such a pulsed source will not fluoresce continuously. Rather, its fluorescence intensity, excited by a single pulse, will decay exponentially until the next pulse again excites the sample. The fluorescence from the sample excited by the pulsed source can be represented, after detection, as a function of time on a fast sampling oscilloscope used in conjunction with a multichannel pulse analyzer. The former approach is called pulse fluorimetry and the latter time-correlated single-photon counting. In either case, fluorescences with decay times much longer than the lamp-pulse characteristics can be analyzed in much the same way as radioactive decay curves. A semilogarithmic plot of fluorescence intensity against time will yield a straight line (or a series of overlapping lines if several fluorophores have comparable but not identical decay times) whose slope is proportional to the decay time and whose vertical axis intercept can be compared with that of a standard solution of the fluorophore for quantitative analysis. If, however, the lamp pulse time and the decay time of the fluorophore are comparable, the lamp characteristics must be subtracted from the observed signal to obtain the fluorophore's decay characteristics. This is usually accomplished by using a computer to deconvolute the composite temporal characteristics of the lamp and the fluorophore output.

The pulsed-source (time-resolved) method effects spectroscopic distinction between the emissions of several fluorescing species by taking advantage of differences in their decay times rather than their fluorescence intensities. This means that several strongly overlapping fluorescences such as those of several proteins or catecholamines can be quantitated simultaneously without chemical or mechanical separation.

Phosphorescence can also be measured in a time-resolved mode. Because of the much longer lifetimes of phosphorescence (10^{-6}–10 s), much longer duration lamp pulses or even mechanical chopping of a continuous source for long-lived phosphorescence (0.1–10 s) can be used. As a result, time-resolved phosphorimetry is much less expensive and electro-optical factors much less demanding than time-resolved fluorimetry. Many of the assays carried out with the lanthanides Eu(III) and Tb(III) are excecuted as time-resolved phosphorimetric analysis.

The pulse-fluorimetric method is best used when one species has a much longer decay time than other species in the solution. It is much less useful when several species in solution have similar decay times. The pulse-fluorimetric approach has seen only limited application to the resolution of more than two fluorescers under any circumstances.

Time-resolved fluorimetry is also useful for the elimination of interferences from stray light due to Rayleigh and Raman scattering. The latter phenomena occur on a timescale of 10^{-14}–10^{-13} s and as they have a much shorter duration than lamp or laser pulses this light associated with them can be eliminated from the signal that ultimately reaches the detector. Time-correlated single-photon counting is superior in its ability to resolve multiple fluorescence from the same solution.

Phase fluorimetry is another useful fluorimetric technique for the determination of substances with overlapping fluorescence spectra. In phase fluorimetry the phase angle between the lamp pulse and the emission of fluorescent light allows discrimination between fluorescences of different origin. Phase-sensitive optics and electronics are rather complicated and, as in the case of time-resolved fluorimetry, will only be mentioned here in passing.

7 FIBER-OPTIC FLUORESCENCE SPECTROSCOPY

A fairly recent but very significant development in fluorescence instrumentation is seen in the fiber-optic fluorimeter. In this instrument a fiber-optic cable replaces the sample compartment and cell and a covalently bound flurophore and perhaps a chemically selective membrane at the distal end of the cable comprise a fluorescence sensor. This device has found extensive use in environmental and biological analysis. In the fiber-optic instrument, light from a xenon lamp, laser, or light-emitting diode (LED) travels along an optically conducting fiber to its distal end, where activation of a covalently bound fluorophore causes fluorescence. Fluorescence sensors based on direct fluorescence or competitive binding are available but most are based upon fluorescence quenching of the sensor fluorophore by the analyte.

The instrumentation consists of a light source, optical filters, the fiber-optic cable, a sensing zone (the fluorophore and auxiliary membranes or reagents) and a detector. Lasers, xenon, hydrogen, deuterium, mercury, and halogen lamps, and LEDs have been used as excitation sources. Fiber-optic fluorescence spectrometers may be operated in a continuous or pulsed excitation mode. The latter mode allows for the application of time-resolved fluorimetry. The material of which the optical fiber is made determines the excitation wavelength range used. Fused-silica, glass, and plastic fibers are the most common fiber materials. Silica can be used from the UV range down to 220 nm but the fibers are expensive. Glass is suitable for use in the visible region and is lower in cost than silica. Plastic fibers are even less expensive but are limited to use above 450 nm.

Fiber-optic sensors are less expensive, more rugged, and smaller than electrodes; in the future we may see the former replacing the latter in various areas of analytical and clinical chemistry. Fields of application include environmental monitoring, process control, remote spectroscopy in high-risk areas with radioactive, explosive, biological, or other hazards, titrimetry, and in vivo bioanalysis.[59,60]

8 SEMICONDUCTOR LASERS AND RED AND NEAR-INFRARED EMITTING LABELS

Semiconductor lasers (laser diodes) have recently appeared as luminescence excitation sources. They have several outstanding advantages over gas and dye lasers and arc lamps as excitation sources. They are more powerful and more coherent light sources than arc lamps. They also have the advantage of producing a polarized beam which can be useful for certain applications. They are smaller, more compact, and much less expensive than other kinds of lasers and the fact that most semiconductor lasers emit in the far-red or IR region means that less energy can be deposited in the sample so that considerably less thermal decomposition can occur than with conventional ion, excimer, and dye lasers. The

fact that emission from semiconductor lasers is largely confined to the far-red and IR region is a ''mixed blessing''. On the one hand, the fact that so few substances demonstrate electronic absorption in that region of the electronic spectrum means that excitation of semiconductor laser light will be very selective. For example, the often interfering luminescences of tryptophan and bilirubin from serum samples will not be problematic with semiconductor laser excitation because these substances are not excited by red light. On the other hand, since so few substances absorb red light, a new generation of fluorescent probes and labels is required for the labeling of analytes to be detected subsequent to chromatographic or immunochemical analysis.

Several of the polymethine dyes absorb in the red, far-red and near-infrared (NIR) region and fluoresce efficiently also in this spectroscopic region. Functional derivatives of these may provide excellent fluorescent labels for semiconductor laser excitation. Several research groups are currently involved in the synthesis and development of large polyunsaturated dye molecules which are excited and show luminescence in the red and NIR region. This promises to be one of the most exciting areas of luminescence spectroscopy for the foreseeable future.

ABBREVIATIONS AND ACRONYMS

ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
CL	Chemiluminescence
CZE	Capillary Zone Electrophoresis
DNPO	Bis(2,4-dinitrophenyl) Oxalate
FIA	Flow Injection Analysis
FPIA	Fluorescence Polarization Immunoassay
HPLC	High-performance Liquid Chromatography
Ig	Immunoglobulin
IR	Infrared
LED	Light-emitting Diode
LSD	Lysergic Acid Diethylamide
NIR	Near-infrared
TCPO	Bis(trichlorophenyl) Oxalate
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Imaging • Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Laboratory Instruments in Clinical Chemistry, Principles of • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science • Immunoassays in Forensic Toxicology

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Probes • Fluorescence In Situ Hybridization

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Fluorescence Spectroscopy in Peptide and Protein Analysis

Surfaces **(Volume 10)**
Photoluminescence in Analysis of Surfaces and Interfaces

Electroanalytical Methods **(Volume 11)**
Ultraviolet/Visible Spectroelectrochemistry

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Absorption and Luminescence Probes • Detectors, Absorption and Luminescence • Fluorescence Imaging Microscopy • Fluorescence in Organized Assemblies • Fluorescence Lifetime Measurements, Applications of • Near-infrared Absorption/Luminescence Measurements • Phosphorescence Measurements, Applications of • Surface Measurements using Absorption/Luminescence • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Ultrafast Laser Technology and Spectroscopy

FURTHER READING

W.R.G. Baeyens, in *Molecular Luminescence Spectroscopy: Methods and Applications, Part I*, ed. S.G. Schulman, J. Wiley & Sons, New York, Chapter 3, 1985.

W.R.G. Baeyens, D. de Keukeleire, K. Korkidis (eds.), *Luminescence Techniques in Chemical and Biochemical Analysis*, Marcel Dekker, New York, 1991.

L. Brand, M.L. Johnson (eds.), *Fluorescence Spectroscopy*, Academic Press, San Diego, 1997.

A.K. Campbell, *Chemiluminescence, Principles and Applications in Biology and Medicine*, VCH, Weinheim, 1988.

M.A. DeLuca, W.D. McElroy (eds.), *Bioluminescence and Chemiluminescence, Part B. Methods in Enzymology*, Academic Press, New York, Vol. 133, 1978.

D. Eastwood, L.J.C. Love (eds.), *Progress in Analytical Luminescence*, ASTM, Philadelphia, 1988.

Fluorescence Spectra, Sadtler Research Laboratories, Philadelphia, Vols. 1–8, 1974–76.

M.C. Goldberg (ed.), *Luminescence Applications in Biological, Chemical, Environmental, and Hydrological Sciences*, American Chemical Society, Washington, DC, 1989.

G.G. Guilbault (ed.), *Practical Fluorescence*, 2nd edition, Marcel Dekker, New York, 1990.

K.D. Gundermann, F. McCapra, *Chemiluminescence in Organic Chemistry*, Springer, Heidelberg, 1985.

L.J. Kricka, 'Chemiluminescent and Bioluminescent Techniques', *Clin. Chem.*, **37**, 1472–1481 (1991).

J.R. Lakowicz (ed.), *Topics in Fluorescence Spectroscopy*, Plenum Press, New York, Vols. 1–5, 1991.

J. Lakowicz, *Principles of Fluorescence Spectroscopy*, 2nd edition, Kluwer Academic/Plenum, New York, 1999.

M.D. Lumb (ed.), *Luminescence Spectroscopy*, Academic Press, London, 1978.

W.T. Mason (ed.), *Fluorescent and Luminescent Probes for Biological Activity: A Practical Guide to Technology for Quantitative Real-time Analysis*, Academic Press, London, 1993.

F. McCapra, K.D. Perring, 'General Organic Chemiluminescence', in *Chemi- and Bioluminescence: Clinical and Biochemical Analysis*, ed. J.G. Burr, Marcel Dekker, New York, Vol. 16, 1985.

K. Nakashima, K. Imai, 'Chemiluminescence', in *Molecular Luminescence Spectroscopy: Methods and Applications: Part 3*, ed. S.G. Schulman, Wiley-Interscience, New York, Chapter 1, 1993.

C.A. Parker, *Photoluminescence of Solutions*, Elsevier, Amsterdam, 1968.

D. Rendell, *Fluorescence and Phosphorescence Spectroscopy*, ed. D. Mowthorpe, J. Wiley & Sons, Chichester, on behalf of ACOL, London, 1987.

S.G. Schulman, *Fluorescence and Phosphorescence Spectroscopy: Physicochemical Principles and Practice*, Pergamon Press, New York, 1977.

S.G. Schulman (ed.), *Molecular Luminescence Spectroscopy: Methods and Applications, Parts I–III*, Wiley, New York, 1985–93.

W.R. Seitz, 'Chemiluminescence and Bioluminescence Analysis Fundamentals and Biomedical Applications', *CRC Crit. Rev. Anal. Chem.*, **13**, 1–58 (1981).

A. Sharma, S.G. Schulman, *Introduction to Fluorescence Spectroscopy*, Wiley-Interscience, New York, 1999.

S.A. Soper, I.M. Warner, L. B. McGown, 'Molecular Fluorescence, Phosphorescence and Chemiluminescence', *Anal. Chem.*, **70**(12), 477R–494R (1998).

T. Vo-Dinh, *Room Temperature Phosphorimetry for Chemical Analysis*, Wiley-Interscience, New York, 1982.

E.L. Wehry (ed.), *Modern Fluorescence Spectroscopy*, Plenum, New York, Vols. 1 and 2, 1976; Vols. 3 and 4, 1981.

J.D. Winefordner, S.G. Schulman, T.C. O'Haver, *Luminescence Spectrometry in Analytical Chemistry*, Wiley-Interscience, New York, 1972.

M. Zander, *Phosphorimetry*, Academic Press, New York, 1968.

REFERENCES

1. K. Peterson, H. Sitari, I. Hemmila, E. Soini, T. Lovgren, V. Hanninen, P. Tanner, U. Stenmon, *Clin. Chem.*, **29**, 60–64 (1983).
2. M.J. Khosravi, E.P. Diamandis, *Clin. Chem.*, **33**, 1994–1999 (1984).
3. G.K. Aghajanian, H.L. Bing, *Clin. Pharmacol. Ther.*, **5**, 611–614 (1964).
4. A.A. Alwarthan, A. Townshend, *Anal. Chim. Acta*, **205**, 261–265 (1988).
5. I.I. Koukli, A.C. Calokerinos, T. Hadjiioannou, *Analyst*, **114**, 711–714 (1989).
6. S.A. Halvatzis, M.M. Timotheou-Potamiu, A.C. Calokerinos, *Analyst*, **113**, 1229–1234 (1988).
7. R.W. Abott, A. Townshend, *Anal. Proc.*, **23**, 25–26 (1988).
8. R.W. Abott, A. Townshend, R. Gill, *Analyst*, **112**, 397–406 (1987).
9. A.R.J. Andrews, A. Townshend, *Anal. Chim. Acta*, **227**, 65–71 (1989).
10. S.L. Tomsett, *Acta Pharmacol. Toxicol.*, **26**, 299–306 (1967).
11. R.A. Chalmers, G.A. Wadds, *Analyst*, **95**, 234–241 (1970).
12. D.B. McDougal, H.S. Farmer, *J. Lab. Clin. Med.*, **50**, 485–488 (1957).
13. G. Nadeau, G. Sobolewski, *Can. J. Biochem. Physiol.*, **36**, 625–631 (1958).
14. H. Kupferberg, A. Burkhalter, E.L. Way, *J. Pharmacol. Exp. Ther.*, **145**, 247–251 (1964).
15. J.B. Ragland, V.J. Kinross-Wright, *Anal. Chem.*, **36**, 1356–1359 (1965).
16. F. Tishler, P.B. Sheth, M.B. Giamimo, *J. Assoc. Off. Agric. Chem.*, **46**, 448–451 (1963).
17. A.C. Mehta, S.G. Schulman, *J. Pharm. Sci.*, **63**, 1150–1151 (1974).
18. E.M. Scott, R.C. Wright, *J. Lab. Clin. Med.*, **70**, 355–360 (1967).
19. S. Udenfriend, S. Stein, P. Bohlen, W. Dairman, W. Leingruber, M. Weigele, *Science*, **178**, 871 (1973).
20. J.A.F. deSilva, N. Strojny, *Anal. Chem.*, **47**, 714–718 (1975).
21. P.A. Shore, J.S. Alpiero, *Life Sci.*, **3**, 551–554 (1964).
22. K.H. Ibsen, R.L. Saunders, M.R. Urist, *Anal. Biochem.*, **5**, 505–514 (1963).
23. E.N. Cohen, *J. Lab. Clin. Med.*, **61**, 338–345 (1963).
24. A.V. Karyakin, G.G. Babicheva, *Zh. Anal. Khim.*, **23**, 788–790 (1968).

25. F. Gorus, E. Schram, *Clin. Chem.*, **25**, 512–519 (1979).
26. R.L. Boeckx, *Hum. Pathol.*, **15**, 104–111 (1984).
27. A. Thore, S. Ansehn, A. Lundin, *Clin. Microbiol.*, **1**, 1 (1975).
28. L. Nilsson, H. Hojer, S. Ansehn, in *Proceedings of the International Symposium of Bioluminescence and Chemiluminescence*, State Printing and Publishing, Westlake Village, CA, 515, 1979.
29. A.M. Dhople, J.H. Hancks, in *Second Biannual ATP-Methodology Symposium: Proceedings*, SAI Technology, San Diego, CA, 316, 1977.
30. P.L. Wolf, P. Walters, P. Singh, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, **35**, 252 (1976).
31. P.L. Wolf, in *ATP-Methodology Seminar: Proceedings*, ed. G. Borun, SAI Technology, San Diego, CA, 340–357, 1975.
32. J.L. David, F. Herion, *Adv. Exp. Med. Biol.*, **34**, 341–354 (1972).
33. T.C. Detwiler, R.C. Feinmann, *Biochemistry*, **12**, 2462–2468 (1973).
34. C.J. Feo, P.F. Leblond, *Blood*, **44**, 639–647 (1974).
35. K. Nakao, T. Wada, T. Kamiyama, *Nature (London)*, **194**, 877–879 (1962).
36. E. Farber, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, **32**, 1534–1539 (1973).
37. D.E. Brooks, *J. Reprod. Fertil.*, **23**, 525–528 (1970).
38. A. Thore, in *Second Biannual ATP-Methodology Symposium: Proceedings*, SAI Technology, San Diego, CA, 171, 1977.
39. V.H. Rees, J.E. Fides, D.J.R. Laurence, *J. Clin. Pathol.*, **7**, 336–340 (1954).
40. H.R. Schroeder, F.M. Yeager, *Anal. Chem.*, **50**, 1114–1120 (1978).
41. S.R. Spurlin, M.M. Cooper, *Anal. Lett.*, **19**, 2277–2283 (1986).
42. K.R. Kopecky, J.H. VandeSande, C. Mumford, *Can. J. Chem.*, **46**, 25 (1968).
43. E.A. Chanhoss, *Tetrahedron Lett.*, 761–765 (1963).
44. M.M. Rauhut, D. Sheehan, R.A. Clarke, A.M. Semsd, *Photochem. Photobiol.*, **4**, 1097–1110 (1965).
45. D.S. Milbrath, in *Bioluminescence and Chemiluminescence*, ed. M. DeLuca, Academic Press, New York, 515–518, 1978.
46. S. Chen, G. Yan, M.A. Schwartz, J.H. Perrin, S.G. Schulman, *Science*, **80**, 1017–1019 (1991).
47. S.G. Schulman, J.H. Perrin, G.F. Yan, S. Chen, *Anal. Chim. Acta*, **255**, 383–385 (1991).
48. I.I. Koukli, A.C. Calokerinos, *Analyst*, **115**, 1553–1557 (1990).
49. I.I. Koukli, A.C. Calokerinos, *Anal. Chim. Acta*, **236**, 463–468 (1991).
50. I. Aichinger, G. Gubitz, J.W. Birks, *J. Chromatogr.*, **523**, 163–172 (1990).
51. A.S. Waggoner, in *Flow Cytometry and Sorting*, 2nd edition, eds. M.R. Melamed, T. Lindmo, M.L. Mendelsohn, Wiley-Liss, New York, 209–225, 1990.
52. K. Imai, A. Nishitani, H. Akimoto, Y. Tsuleamoto, *J. Biochem. Chemilumin.*, **4**, 500–504 (1989).
53. T.A. Nieman, in *Chemiluminescence and Photochemical Reaction Detection in Chromatography*, ed. J.W. Birks, VCH, Weinheim, 99–123, 1989.
54. R.A. Femia, L.J.C. Love, *Anal. Chem.*, **56**, 327–331 (1984).
55. L.J.C. Love, M. Skribc, J.G. Habarta, *Anal. Chem.*, **52**, 754–759 (1980).
56. P. Urios, N. Cittanora, M.F. Jayle, *FEBS Lett.*, **94**, 54–58 (1978).
57. M.L. Grayeski, in *Clinical and Biochemical Analysis*, ed. J.G. Burr, Marcel Dekker, New York, 469–493, Vol. 16, 1985.
58. F. McCapra, K.D. Perring, in *Clinical and Biochemical Analysis*, ed. J.G. Burr, Marcel Dekker, New York, 359–386, Vol. 16, 1985.
59. W.R. Seitz, *CRC Crit. Rev. Anal. Chem.*, **19**, 135–173 (1988).
60. S.M. Angel, *Spectroscopy*, **2**, 37–44 (1987).

Planar Chromatography in Clinical Chemistry

Raka Jain
All India Institute of Medical Sciences, New Delhi, India

Joseph Sherma
Lafayette College, Easton, USA

1 Introduction 1

2 Practical Considerations 2
2.1 Specimen Collection and Storage 2
2.2 Sample Preparation 2
2.3 Sample Application 3
2.4 Equilibration 4
2.5 Development 4
2.6 Qualitative and Quantitative Evaluation 4
2.7 Validation of Results 5

3 Recent Developments in Thin-layer Chromatography Plates 5

4 Comparison of Instrumental Thin-layer Chromatography with Other Analytical Techniques 6

5 Applications of Planar Chromatography in Clinical Chemistry 6
5.1 Amino Acids 7
5.2 Bile Acids 7
5.3 Carbohydrates 8
5.4 Drugs 8
5.5 Lipids 9
5.6 Phospholipids 10
5.7 Porphyrins 11
5.8 Prostaglandins 12
5.9 Steroid Hormones 12
5.10 Other Compounds of Clinical Interest 13

6 Conclusion 13

Abbreviations and Acronyms 13

Related Articles 14

References 14

Planar chromatography (PC) like any other chromatographic technique is a multistage distribution process. It is a form of liquid chromatography in which the stationary phase is supported on a planar surface rather than a column. Separations in PC occur because of differential migration velocities through the sorbent layer in a fixed separation time. The sample zones are fixed in space at the completion of the separation, allowing off-line monitoring. There are two planar techniques: paper chromatography and thin-layer chromatography (TLC). In recent years, TLC has superseded paper chromatography owing to greater speed, versatility, and reproducibility. Improvements in techniques and equipment as well as the ready availability of high-quality plates with uniform layers have extended and increased the use of TLC in clinical chemistry and other disciplines, offering quantitative as well as qualitative analysis. High-performance thin-layer chromatography (HPTLC), in particular, has developed to the extent that separation and quantitation can provide results that are comparable with other analytical methods such as high-pressure liquid chromatography (HPLC). The widespread application of TLC is also due to the fact that it is simple, robust, and flexible, and has a large sample throughput. Furthermore, coupled spectroscopic methods (TLC/mass spectrometry (TLC/MS), TLC/Fourier transform infrared spectrometry (TLC/FTIR) etc.) have considerably enhanced the reproducibility of TLC.

TLC is often critically compared with column chromatography such as HPLC. This is a matter of endless dispute. Each technique has its own advantages and limitations. TLC is only inexpensive when qualitative results are needed. When reliable quantitative data are required, TLC is as expensive as column chromatography. In the authors' opinion, complementary rather than competitive chromatography offers the best solution to analytical requirements.

1 INTRODUCTION

PC was developed in the late 1950s. PC means chromatography in a thin planar stationary field with a length to thickness ratio often exceeding 1000 : 1. It is a physicochemical separation method, in which components to be separated are distributed between two phases. One of these phases consists of a flat stationary bed of large surface area, and the other is a mobile phase that migrates through the stationary phase by capillary action or under the influence of forced flow. Forced flow is achieved by pressure, by an electric field, or by centrifugal force. The movement of the analyte is expressed by its retardation factor (R_f), which is the ratio of the distance moved by the analyte from the origin to the distance moved by the flowing solvent from the origin. Each analyte will have its own R_f value. Thus, the sample zones are fixed in space at the completion of the separation, allowing off-line monitoring.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Initially, paper chromatography was the only planar chromatographic method but TLC has superseded paper chromatography. The use of TLC for separation and quantitative determination of a wide variety of organic and inorganic substances has considerably increased. This increase is due to improved instrumentation for spotting, developing, and scanning layers and for applying detecting reagents. In addition, development of optimization strategies and special techniques, such as scanning of radioactive zones and TLC coupled with spectrometry (TLC/MS, TLC/FTIR, etc.) have enhanced the analytical capabilities of TLC. TLC is considered to be an ideal chromatographic method for clinical laboratories. It is used clinically as an aid in diagnosis.The advantages of TLC over the other methods include simplicity, rapidity, versatility, applicability to a large number of samples in minimal time, and low cost in terms of reagent and equipment. Owing to these advantages, this method of analysis is preferred in clinical chemistry. Several comprehensive texts[1–9] are now available on the PC technique. The attempt here is to not to replace the existing reviews and books available on the subject but to outline recent progress in this area.

2 PRACTICAL CONSIDERATIONS

In this section, the basic steps involved in the analysis of biological samples by PC are discussed.

2.1 Specimen Collection and Storage

Many biological specimens can be used for determination of various analytes. These include whole blood, plasma, serum, urine, saliva, cerebrospinal fluid, gastric fluid, amniotic fluids, body tissues, and so on. Each type has advantages and limitations with respect to its availability, ease of collection, and the amount of information that can be obtained from its analysis. Plasma, serum, and urine samples are most often used in clinical laboratories. Quality control (QC) in the clinical laboratory begins before the sample is collected from the patient. Sample collection is a crucial element in the process of patient diagnosis and treatment. It is important to ensure that a proper and appropriate sample is being collected as the validity of the test results is dependent on the integrity of the specimen.[10,11] All specimens should be collected in a clean sterile unbreakable leakproof bottle. The bottles should be properly stoppered, sealed, and labeled with patients' names, date, and time of collection. To promote clot formation and minimize the potential for hemolysis in blood samples, specimens should be maintained in an upright position and agitated as little as possible during transportation.

Biological specimens for substance-abuse testing programs and forensic purposes must be handled with great care as they can be tampered with by substitution, dilution, or adulteration. In such cases, specimens should be collected under close supervision and transported under custody to avoid pilferage.[10–15]

Proper storage and preservation of samples are also important in clinical chemistry.[10,11] When the samples cannot be analyzed immediately, the collected samples may either be stored at 2–4 °C or for longer periods at −20 °C to −70 °C until analysis. This will, however, depend on the substance to be analyzed. Moreover, some blood and urine samples may need acid, alkali, or solvent stabilization. If samples are left at room temperature for long prior to analysis, their pH changes and they begin to decompose, leading to erroneous results. Long-term storage also enhances the probability of degradation of drugs and metabolites in biological samples. Ideally, prompt delivery of the collected samples to the laboratory and a short period of storage increase the validity of test results. It is also important that samples should be protected from direct light and heat during transportation and storage, and they should, therefore, be kept cold during transport, preferably in an insulated box containing ice or some other cooled packing.

2.2 Sample Preparation

Biological samples are very complex multicomponent mixtures of organic and inorganic compounds in which the specific target analyte is found in minute amounts. Therefore, selective isolations of analytes from the samples and preconcentration procedures are often required prior to analysis. Purification and concentration steps enrich the analyte in the sample so that it falls within the detection limit of the analytical technique employed. Considerable literature on sample preparation exists and has been reviewed.[11,16–22]

The choice of the cleanup procedure may depend on the technique used for detection and determination, the detection limit required, and the speed of analysis and recovery. Sometimes combinations of procedures are used. In this section, frequently used sample preparation procedures are discussed.

Dilution is a very simple and effective means of sample preparation. It is nonspecific and is used where an analyte is present in a sufficiently large concentration. The sample is diluted with water or buffer, centrifuged if needed to remove particulate matter, and then assayed.

Another effective method of sample preparation is precipitation and deproteinization. This is often used on plasma and whole blood samples prior to analysis. The agents commonly used to precipitate proteins are either

acids (viz. trichloroacetic acid, perchloric acid, tungstic acid, sulfosalicyclic acid, and metaphosphoric acids, etc.), organic solvents (methanol, ethanol, acetone, acetonitrile, and dimethyl sulfoxide), inorganic salts (ammonium sulfate) or any metal cations (zinc and copper). The technique is simple and rapid. The main disadvantage of this method is that there may be loss of the analyte by occlusion in the precipitate, and lower recoveries ensue.

Biological samples are also purified by ultrafiltration techniques. The technique is used for rapid removal of low-molecular-weight (<50 000 molecular mass) dissolved solutes from the biological matrix by passing through an anisotropic membrane filter under centrifugal force. This technique is widely used to concentrate proteins and is applicable to small sample volumes.

Dialysis can separate an analyte from the matrix by diffusion through a semipermeable membrane that retains macromolecules, and allows passage of solute molecules of molecular weight below 5000. Diffusion is a slow process and is driven by a concentration gradient until equilibrium is established. It can only be applied to compounds that are relatively weakly protein bound. The use of dialysis for sample preparation is normally not recommended as it is time-consuming and the sample becomes diluted.

Hydrolysis is also used in the preparation of samples for analysis. It is applied where analytes are present in their conjugated form in biological samples. This is accomplished by the use of either enzymes such as ß-glucuronidase and arylsulfatase, or nonspecific strong acids or bases. The aim of the technique is to cleave the conjugate and release the original compound for assay.

Liquid–liquid extraction (LLE) is a versatile technique that is commonly used for sample preparation. In LLE, the pH of the aqueous phase is adjusted so that the analyte to be extracted is virtually undissociated, thus facilitating extraction into the solvent. An advantage of LLE is its selectivity. Depending on the choice of the solvent and pH adjustment of the aqueous phase, the compound of interest can be isolated. The success of an extraction step depends on knowledge of the polarity of the analyte. Additional selectivity can be imparted to an extraction method by multiple extraction procedures or back extraction whenever required. One of the problems frequently encountered in LLEs is the formation of emulsions that causes loss of analyte by occlusion within the emulsion, leading to lower recoveries. Emulsion formation can be minimized by the use of a large volume of the extracting solvent, by less rigorous mixing, centrifugation, and ultrasonication, stirring with a glass rod, or filtration through a loose bed of glass wool or phase-separation filter paper. An increase in ionic strength by addition of salt (sodium chloride, potassium bromide) or a small amount of ethanol or ethyl-1-hexanol may cause a decrease in the forces stabilizing the emulsion. Other drawbacks of the procedure are lack of speed, the need for large solvent volume, and possible low analyte recovery.

The use of the commercial liquid–liquid disposable columns (Chem Elut™, Tox Elut™, Extrelut™, etc.) has gained popularity for isolation of compounds from biological specimens. These columns contain diatomaceous earth as an adsorbent. Some of the limitations of conventional LLE are overcome by the use of these columns.

Solid-phase extraction (SPE) currently is a major sample preparation tool. The isolation of the compound by this method depends upon the relative affinity of the analyte in a biological matrix for the solid adsorbent and the relative ease of eluting the compound for subsequent analysis. The conventional sorbents such as carbon, celite, alumina, nonionic resins, ionic exchangers, and Sephadex™ gel are still widely used for sample preparation, but the recent trend has been to use chemically modified silicas. Many different types of SPE cartridges are commercially available. The main advantage of SPE is its selectivity due to its wide choice of sorbents available. By proper selection of disposable SPE cartridges/minicolumns and the eluting solvents, analytes with a wide range of physicochemical properties can be isolated from biological samples to facilitate chromatographic separation. Other advantages include low solvent consumption, short throughput times, clean extraction, high recovery, elimination of emulsions, and reproducible results. The techniques provide an attractive alternative to traditional sample preparation methods.

At the present time, use of a preadsorbent-zone plate greatly facilitates the sample cleanup procedure. The preadsorbent zone[2,6–7] retains interfering solutes on the preadsorbent zone in the lower part of the plate and causes analyte preconcentration, leading to increased selectivity and sensitivity. The latest development in bioanalysis is the introduction of particle-loaded membranes.[22] Particle-loaded membranes are available in disks with diameter of 4–90 mm. Normally 4-mm disks are used in drug tubes for the analysis. The main advantage of using drug tubes is the small desorption volume and, therefore, increased sensitivity. Moreover, in many cases particle-loaded membranes are more efficient than packed solid-phase cartridges.

2.3 Sample Application

Sample application is a crucial step in all TLC separations. The size and shape of spot are important for

obtaining good chromatographic resolution. For qualitative purposes, the samples are usually spotted by means of a micropipette or microsyringe or disposable glass capillaries, whereas automated application devices are used for quantitative TLC. Generally 5–15 μL of extracts are applied in the form of spots or narrow bands to the plate. The amount of the sample to be spotted will depend on the thickness of the plate and also on the composition of the sample. For accurate spotting, templates are used. Evaporation of the sample solvent is aided and application is made faster if a stream of cool air or inert gas is blown gently across the plate being spotted.

2.4 Equilibration

It is essential to prepare the developing chamber before the TLC separation. This is done by filling the developing tank with the mobile phase to a depth of 0.5–1.0 cm and lining three sides of it with filter paper. The tank is allowed to stand for at least 1 h with a lid over the top to ensure that the atmosphere within the chamber becomes saturated with solvent vapor (equilibration). Saturated chambers give faster separation, a straight solvent front, and reproducible and consistent R_f values.

2.5 Development

After spotting, the TLC plates are developed in the equilibrated developed chamber. The mobile phase rises with the help of capillary force. When the developing distance is reached, the plate is removed from the tank and the solvent front is marked for calculation of R_f values. The plate is then dried at room temperature or in an electric oven. In recent years, various development strategies have been devised that improved the performance and speed of the TLC separation.[23,24] The methods include programmed multiple development (PMD), automated multiple development (AMD), continuous multiple development, circular or radial and anticircular development, overpressured TLC, two-dimensional TLC, two phases two-dimensional TLC, triangular development, and centrifugal development.

2.6 Qualitative and Quantitative Evaluation

In conventional TLC, detection is a static off-line procedure. Colored or fluorescent spots are easily detected under white or ultraviolet (UV) light. Some analytes exhibit native fluorescence characteristics upon UV irradiation (e.g. porphyrins) and are detected at very low concentrations. To improve the sensitivity and selectivity of detection, derivatization techniques are sometimes used in TLC. Some spots can be made visible, absorbing, or fluorescent with various chemical reagents before (prechromatographic) or after (postchromatographic) the separation process. Postchromatographic derivatization reagents can be applied by means of spraying or dipping. Automated spraying devices and the dipping technique facilitate the even distribution of reagent over the layer, compared to manual spraying. This method has the advantages that by-products of the derivatization reagents do not interfere in the chromatography and all samples are derivatized simultaneously.

Qualitative analysis is done by comparing R_f values, or relative R_f values, of the unknowns with those of standards that have been developed under identical experimental conditions on the same layer.

Two approaches are used for quantitation of samples in TLC:

1. Indirect methods: in these methods, quantitative measurements are carried out by visual comparison of colored or chemically visualized zones or excizing of separated zones from the layer followed by elution from the sorbent and spectrophotometric analysis. These methods are insensitive, labor intensive, and tedious.
2. Direct methods: in these methods, the separated spots are evaluated in situ on the plate. The recorded detector signal can be plotted against the amounts of the reference sample providing a calibration curve that is useful for quantitation over the desired range.

Scanning densitometry is the most efficient procedure for quantitative evaluation.[1–9,25] In situ, quantitative measurements of TLC plates can be performed in many ways. The reflectance, transmittance, simultaneous reflectance and transmittance, fluorescence, or fluorescence quenching of a spot can be measured by use of a densitometer. The most sensitive in situ method is fluorimetry. The methods are based on measuring the difference in optical response between blank portions of the medium and regions where a separated substance is present. Significant advances in technology have been made in the densitometric technique. Modern densitometers have a high degree of automation, both in the recording process and in manipulation of the data obtained. Scans can be repeated for all or part of the chromatogram, without the need to rerun samples. This characteristic of TLC permits the use of additional detection techniques between scans that can aid in the identification of the unknown material.

Furthermore, current developments in detection methods for TLC also include in situ methods of detection without mechanical scanning using image analyzers, the use of laser sources to enhance sensitivity, and coupling devices such as TLC/MS and TLC/FTIR to provide structural information as well as quantification.[1–9,25]

2.7 Validation of Results

Validation of results is prerequisite in the analysis of biological samples, as the analytical result aids in making a diagnosis and in treatment of diseases in patients. The validation of PC methods requires the demonstration of calibration linearity, reproducibility, specificity, sensitivity, extraction recovery, precision, and accuracy.

Calibration is one of the important steps in quantitative chemical analysis. Without a good calibration procedure, precision and accuracy cannot be obtained. In general, calibration standards of the test compound in a particular biological fluid are prepared by adding pure reference test standards to blank biological samples. A fixed amount of internal standard in each sample can also be added to evaluate good and consistent recovery. The blank samples are included in the calibration samples to ensure that no interferents that are coeluted with the analyte or the internal standard are present in the sample or introduced during sample preparation. Control spiked samples, blank samples, and the unknown samples are then processed in an identical condition. The processed samples are spotted on the TLC plates, developed, and scanned with the densitometer, with or without spraying of detection reagent. Other details of quantitation and calibration linearity have been discussed elsewhere.[11]

In order to check the reproducibility of the results, chromatography can be repeated three or four times on different plates and the mean value of experimentally obtained data can be plotted for each concentration spotted. Sometimes the same volume of each processed sample can be spotted in triplicate on the plate, the mean being used for calculation. Moreover, each spot can be scanned in triplicate to establish the instrumental error.

Specificity of a test is defined as its ability to distinguish between closely related compounds and their metabolites. Specificity of a chromatography assays largely depends on sample processing as well as on the chromatographic separation.[11] Lack of specificity can lead to false positive results, and, therefore, a high degree of specificity is generally desirable.

Sensitivity is defined as the minimal concentration of an analyte in the original undiluted biological fluid that is detectable with a high probability. The sensitivity of an assay is often dependent on the conditions of analysis, purity of reagent, sample matrices, and instrument system. As mentioned earlier, considerable progress has been achieved in stationary phases, chromogenic spray reagents, instrumentation, and automation of the various steps involved in planar chromatographic procedures. Because of these developments of TLC methods, sensitivity has been considerably enhanced and it is similar or identical to that of other chromatographic methods such as HPLC.

Precision is a measure of variability of the results. The precision of the analysis is estimated as the standard deviation or relative standard deviations (coefficient of variation) of a set of replicate results. The precision may relate to the within-laboratory error of a method (repeatability) or to the between-laboratory error of a method (reproducibility).

Accuracy is a measure of systematic deviation of the results obtained from the true values. The accuracy is estimated as the percentage differences (bias) between the mean values and the true or known concentrations. The accuracy of a method is usually expressed in terms of percentages of recovery. In the authors' view, assay results with $\leq$10% relative standard deviation and $\leq$10% bias may be considered acceptable. Analysts should always attempt to improve the precision and accuracy of their assays.

The recovery of an analyte from the biological matrix is also an important step in all chromatographic methods. Recovery of compounds greatly affects the sensitivity, precision, and accuracy of the test procedure, and recovery should also be reproducible. Recovery may be concentration-dependent. The exact recovery can be estimated by comparing the slope of the calibration curve based on the calibration standards prepared by spiking with accurately measured amounts of test compounds of high purity (reference standard) and processed as per the assay procedure with the slope of the calibration curve of the pure unprocessed reference test compound of the same concentration. Recovery of an analyte in a particular assay procedure can be monitored by the use of an internal standard. As defined, the recovery should be close to 100% to determine the reliability of TLC. Replicate experiments (at least two or more replicates) should be carried out on different representative samples. Use of various analytical techniques (e.g. HPLC, gas–liquid chromatography (GLC), GC (gas chromatography)/MS) is also helpful in validating the results of TLC.

3 RECENT DEVELOPMENTS IN THIN-LAYER CHROMATOGRAPHY PLATES

Traditionally, TLC has been performed using inorganic and organic sorbents such as alumina (neutral and acidic), magnesium silicate, diatomaceous earth (kieselguhr), cellulose, polyamide, ion-exchange resins (Dowex™50-X8), Sephadex™, and so on, coated onto glass, plastic, or aluminum foil. Silica gel continues to be one of the most widely used layer materials for PC. However, during the 1990s considerable progress took place in the preparation of bonded-phase silica layers. A range of

alkyl-bonded silica TLC plates are now commercially available. These include C_{18}, C_{12}, C_8, C_2, aminopropyl, diphenyl, and cyanopropyl bonded plates, as well as plates coated with a chiral complexing agent for the separation of amino acid enantiomers and similar compounds. Furthermore, precoated plates with concentrating zones are also available. The introduction of bonded silica has led to a renewed interest in reversed-phase TLC (RPTLC). RPTLC has proved quite useful for the chromatography of polar compounds. Advantages of RPTLC over normal-phase TLC (NPTLC) are higher recoveries of materials from the plate, ease of optimization of solvent systems, and minimum decomposition of sensitive compounds.

Another major development in TLC has been the introduction of HPTLC plates. Modern HPTLC plates differs from conventional TLC plates in number of ways. These plates are smaller in size (10×10 or 10×20 cm), precoated with smaller size particles (2–7 μm), have narrow particle size distribution, and have smooth surface and thinner layers (0.2 mm). Other characteristics of HPTLC compared with TLC include smaller sample volume, economy, faster separation, reduced diffusion leading to improved separation efficiency, more reproducible results, and lower detection limits (by a factor of 10–15). Thus, separation and quantification can provide results that are comparable with other analytical methods such as HPLC.[11,25]

4 COMPARISON OF INSTRUMENTAL THIN-LAYER CHROMATOGRAPHY WITH OTHER ANALYTICAL TECHNIQUES

Quantitative thin-layer chromatography (QTLC) has gained the analyst's interest. It has been established that HPLC and QTLC have their specific merits and often complement one another. Typical features of QTLC are:[11,26]

1. Simple sample preparation because the stationary phase is disposable and, therefore, contamination of the stationary phase by the sample matrix can be accepted. Conversely, sample preparation for HPLC is critical, since the stationary phase is expensive and intended for multiple reuse.
2. Large sample volumes can be applied using a special spray-on technique. This often eliminates a preconcentration step.
3. The off-line character has many advantages but does not lead to expense.
4. Sample throughput is higher than in any other chromatographic technique. A large number of samples can be chromatographed simultaneously, making TLC even quicker than fully automated HPLC.
5. Any solvent may be used for QTLC irrespective of its UV cut off as the mobile phase (solvent) is evaporated prior to detection (quantitative measurement).
6. Mobile phase use is much less than that required for HPLC, hence QTLC is economical.
7. All substances (fractions) are permanently stored on the plates and are unaffected even after scanning. This makes it possible to repeat detection (scanning) of one, several, or all fractions of the chromatogram with the same or with different parameters. Moreover, the same plate can be used for recording the absorbance of the positively identified fractions and to recover the fractions for further analysis.
8. Two-dimensional chromatography is not possible in HPLC.
9. Any combination of mobile and stationary phases cannot be used in HPLC for the fear of irreversibility damaging the expensive column.
10. Positive identification of fractions is not a problem. Identification can be achieved by cochromatography of standards, in situ spectra recording, identification reactions, and combinations of these.
11. Running costs (solvents) and costs per sample analysis are relatively low in comparison to HPLC.

HPLC has certain distinct advantages over QTLC.

1. The chromatographic resolution power of HPLC is better than that of QTLC owing to the short migration distance of the latter.
2. The quantitative precision of HPLC is somewhat better than that of QTLC.

5 APPLICATIONS OF PLANAR CHROMATOGRAPHY IN CLINICAL CHEMISTRY

PC is widely used in clinical laboratories for detection of a number of substances such as amino acids, bile acids, cerebrosides, drugs and their metabolites, gangliosides, lipids, phospholipids, porphyrins, prostaglandins (PGs), and so on. Owing to the limitations of the article length, it is not possible to cover all of the applications in detail. With these objectives and caveats in view, important selected applications of the technique are discussed below.

5.1 Amino Acids

The use of PC for separation of amino acids is very well documented in the medical sciences. In clinical laboratories, the estimation of free acids in biological fluids and tissues by TLC aids in making a diagnosis of inborn errors of amino acid metabolism,[27–31] classic cystinuria,[32] heterozygous cystinuria,[33] aminoacidopathies,[34] Alports' syndrome,[35] and so on. Prior to TLC analysis of amino acids, purification of the specimen is essential. Cellulose, silica gel, ion exchange, and alkyl-bonded stationary phases are mainly used for separation of amino acids.[2,11] Some researchers[36,37] have developed a method in which amino acid patients with leucinosis, citrullinuria, and phenylketonuria do not require deionized urine or deproteinized serum samples for amino acid analysis.

Amino acid separation[6,7] have been studied in several PC systems: 24 acids on papers impregnated with Sn(IV) and Th(IV) phosphosilicate cation exchangers using water, alcohols, acids, acetone, benzene, ether, and phenol as developing solvents; 15 acids with *n*-butanol–acetic acid and combined with water, chloroform, or ethyl acetate as mobile phases on plain and nickel chloride-impregnated silica gel plates; chitin layers with monocomponent and binary mobile phases; 14 acids on mixed adsorbents containing silica gel and alumina or cellulose with benzene-containing mobile phases; α-amino acids on antimony(V) phosphate silica gel plates in different aqueous, nonaqueous, and mixed-solvent systems; racemic dinitropyridyl, dinitrophenyl, and dinitrobenzoyl amino acids on RPTLC plates developed with aqueous–organic mobile phases containing bovine serum albumin and chiral agent; 4-(*N*-*N*-dibutylamino)azobenzene-4-thiohydantoin derivatives of 15 α-amino acids on antimony(V) phosphate silica gel plates in different aqueous, nonaqueous and mixed-solvent systems; phenyl thiohydantoin(PTH) derivatives of the 20 common protein amino acids by AMD on silica gel plates. Three new solvent systems, pyridine–benzene (2.5 : 20), methanol–carbon tetrachloride (1 : 20), and acetone–dichloromethane (10.3 : 8.0) for the resolution of mixtures of 18 PTH amino acids have been reported and have proved very successful.[36] Direct resolution of enantiomers of amino acids by impregnated TLC has also gained popularity in biomedicine.[37,38]

In general, ninhydrin is used as a staining reagent for qualitative and quantitative detection of amino acids. A modified ninhydrin reagent prepared by addition of D-camphor and various acids for improved identification of amino acids has been reported.[39,40] Sensitivity limits and color differentiation in amino acid detection were further enhanced by spraying layers with 1,3-indanedione or *o*-thiobenzoic acid prior to ninhydrin.[41]

5.2 Bile Acids

Bile acids are 24-carbon steroid derivatives. For clinical purposes, measurement of bile acids in biological specimens (bile, serum, duodenal contents, and crude faecal extracts) aids in making diagnosis of certain liver or intestinal disorders.[42,43] Many methods are used to quantitate either total or individual bile acids in biological samples. These include spectrophotometry, immunoassays, chromatographic techniques, GC/MS, capillary electrophoresis, and so on. Each technique has its own advantages and disadvantages. In clinical practice, TLC is still accepted for the separation of bile acids. Scalia published a review of the chromatographic methods and combined detection system that can be applied to the analysis of bile acids.[44]

Sample pretreatment is essential before the sample is subjected to TLC analysis. Kindel et al.[45] describe an easy, precise, and rapid TLC method for simultaneous determination of five predominant bile acids (cholic, chenodeoxy cholic acid, deoxycholic acid, litocholic acid, and ursodeoxycholic acid) in human stool specimens on a silica gel plate using two different solvent systems. The plates were chromatographed with isooctane–2-propanol–acetic acid (30 : 10 : 1) for 40 min, dried, and chromatographed again with isooctane–ethyl acetate–acetic acid (10 : 10 : 2) for 65 min. Quantification of bile acids was achieved by dipping the plate vertically (for 2 s) into a 0.2% 2,7-dichlorofluorescein ethanolic solution and fluorescence was measured by direct scanning.

Ferreira et al.[46] proposed a binary mixture (methanol–chloroform, 2.5 : 97.5) for separation of pentachlorophenyl esters of bile acids. Ida et al.[47] presented a TLC method for the separation of methyl esters and hydroxylated derivatives of methyl-5-β-cholanate.

The use of alkyl-bonded phases, such as cyano-bonded silica, is increasing for the separation of bile acids.[2] RPTLC plates are also effective for separation of di- and trihydroxy stereoisomers, whereas the less polar monosubstituted isomers are well resolved in a normal phase system. Sometimes combinations of both systems are very efficient.

Fluorimetry is the most sensitive detection technique used for detection and quantification of free and conjugated bile acids.[48] Bile acids are converted into fluorescent derivatives by reaction with 5% perchloric acids in methanol.[49] Spraying with 10% sulfuric acid and subsequent heating yields fluorescent, light blue spots on a dark violet background.[2]

Recently, Tonelli et al.[50] have studied bile acids kinetics in humans by radio TLC and densitometry coupling. Thus, TLC has proved successful for characterizing bile acid metabolism inside the enterohepatic circulation.

5.3 Carbohydrates

Several diseases are accompanied by increased elimination of various sugars in biological samples. Detection and identification of these sugars constitute an important step in establishing diagnosis of certain diseases. Chromatographic separation of sugars is well documented.[2,6,7,11] Paper chromatography and TLC techniques are quite effective for separation of sugars, but now the use of paper chromatography is declining in clinical laboratories mainly owing to long development times.[51]

In biomedicine, the determination of various monosaccharides and disaccharides in biological samples (urine, plasma) by TLC has been established to evaluate the extent of malabsorption and intestinal permeability disruption in several infections and nutritional diseases, including human immunodeficiency virus (HIV) infection.[52–56]

TLC has also proved quite successful for separation of urinary oligosaccharides in diagnosis of inherited lysosomal storage disorders.[57–60] The improved selectivity of urinary oligosaccharide screening using a one-dimensional TLC system has been examined by Abeling et al.[61] Higashi et al.[62] have also reported the quantitation of oligosaccharides by the TLC densitometric technique. Vajda and Pick[63] have developed an improved method for separation of mono-, di-, and trisaccharides on a silica gel HPTLC plate with acetonitrile–aqueous ammonia–aqueous potassium chloride, mobile phase, followed by ethanolic orcinol–sulfuric acid detection. The adsorbents that are used for TLC separation of sugars are magnesium silicate, alumina, kieselguhr, silica gel or a mixture of kieselguhr and silica gel, and aminopropyl-bonded silica.[11] Bhushan and Kaur[64] have reported the separation of some common sugars on silica gel plates impregnated with transition-metal ions in two different solvent systems and using $KMnO_4$ (0.5%) in 0.1 M NaOH as the detection reagent. Separation of sugars[2,11] in biological samples has been recommended on silica gel with binary, ternary, or quaternary mixtures of solvents such as butanol (water-saturated)–triethylamine, ethyl acetate–methanol–acetic acid–water, propanol–methanol–water, propanol–water, ethylacetate–isopropanol–water, isopropanol–*n*-butanol–acetone–aqueous boric acid, acetonitrile–phosphate buffer, acetonitrile–*n*-butanol–ethyl acetate, acetone–aqueous ammonia, and so on. Use of postchromatographic derivatization reagents such as aniline, diphenylamine, phosphoric acid, 4-amino hippuric acid, aniline–phosphoric acid, 2-aminophenol, anthrone, vanillin, and zinc chloride, and so on have also enhanced the detection limit of sugars in nanogram range.[2]

5.4 Drugs

The importance of the analysis of drugs in biological samples in medicine, toxicology, pharmacology, forensic medicine, de-addiction programs, and other biomedical disciplines is well documented.[11] Drug analysis can be carried out in various biological samples, viz. blood, urine, saliva, gastric contents, tissues, and so on. The type and quantity of sample to be tested depend upon the program needs. For instance, urine is used for screening drugs of abuse, because acquisition of sample is noninvasive and most abused drugs can be detected in urine for a reasonable duration after ingestion.[65,66]

Various immunoassays, chromatographic methods, and MS techniques are used for drug analysis in clinical laboratories.[67] The optimal choice to be adopted depends on cost, work load, program needs, and how the data is used for the chromatographic technique. Of the chromatographic techniques, silica gel TLC is widely used for multiple drug screening programs.[68] All samples to be used need prepurification before they are submitted to chromatographic analysis. Further details of sample preparation have been discussed earlier.

Siek et al.[69] have developed a computer-aided identification of thin-layer chromatographic patterns in broad-spectrum drug screening. The computer program assists in matching the data from a particular chromatogram with those obtained for known drugs or their metabolites recovered from serum, urine, or other specimens. The plates were developed in a single mobile phase. Visualization and detection reagents used to characterize unknowns include fluorescamine, ferric chloride/perchloric acid/nitric acid, dragendorff, marquis, mandelic, and iodinated dragendorff solutions, 254-nm UV light, and vapor from chlorine or hydrochloric acid. This method proved to be very sensitive as detection limits of 5–200 ng per sample spot were obtained for drugs in the database.

For toxicological and for drug-abuse screening programs, silica gel TLC remains the most commonly used PC method.[14,15,66,70,71] Recommended solvent systems[70] used for separating these abusable drugs are: ethyl acetate–methanol–ammonium hydroxide (85 : 10 : 5) for opioids and basic drugs, chloroform–acetone (90 : 10) for barbiturates and other acidic drugs, and heptane–diethyl ether–acetic acid (80 : 10 : 4) and chloroform–methanol–ammonium hydroxide (85 : 15 : 2) for cannabinoids. Chromogenic sprays used for detecting abusable drugs include ninhydrin (amphetamine and primary amines), diphenylcarbazone and mercuric sulfate (barbiturate and gluthemide), iodoplatinate (narcotic analgesics and cocaine, phenothiazine, antidepressants, antihistamine), dragendorff's reagent (narcotic analgesics), and fast blue BB salt and fast blue B salt

(cannabinoids). Currently, commercial TLC systems such as Toxi-Lab™[71,72] are used for screening abusable drugs and toxic substances. However, results indicate that in-house horizontal TLC methods are better and more efficient compared to commercial TLC (Toxi-Lab™) for screening urinary opioids.[73] In contact, the Toxi-Lab™ technique was found to be more effective in detection of amphetamine in comparison to TLC.[74]

Ojanpera et al.[75] have reported combination of NPTLC and RPTLC plates for screening 25 amphetamine-type drugs. These authors used fast black K salt for visualization and the mean list length method for evaluation. Identification and quantification of abused drug, viz. cannabinoids,[76] buprenorphine,[77] benzodiazepine,[78,79] and heroin,[80] by densitometric techniques on HPTLC and TLC plates have also been reported.

Currently, many papers have been published using quantitative densitometric HPTLC methods and coupled spectral methods for pharmacokinetic and metabolic studies of drugs in biological fluids[81–85] and for therapeutic monitoring of drug levels to ensure drug dosage.[86–89] These methods proved to be sensitive, economical, and successful.

5.5 Lipids

Lipids play an important role in cell structure and function. They may be polar (phospholipids or sphingolipids), nonpolar, or slightly polar (cholesteryl esters, triglycerides, cholesterol, etc.). Lipid measurement is of greatest interest in blood, serum, faeces, tissues, and amniotic fluids. Lipid profiles are known to change with disease and their analysis holds considerable potential for clinical diagnosis. The investigation of serum lipids, mainly cholesterol ester, has great diagnostic value for hepatic diseases and lipid metabolism disorders. In lipid storage disease, accumulation of cholesterol occurs in blood and tissues. Hyperlipidemia indicates high risk factors of cardiovascular diseases. Serum lipid analysis is important in skin disorders.[90]

The chromatography of lipids and particularly the PC of lipids is well documented in biomedicine. It provides rapid and complete separation of most neutral and phospholipid classes. Numerous reviews and texts dealing with chromatographic analysis of lipids have appeared.[2,11,91–93] Owing to space limitation, these cannot be discussed at length in this article.

Lipids are primarily extracted from biological samples before they are subjected to chromatography. Preliminary extraction is mainly performed by the classical extraction procedure of Folch et al.[94] with chloroform–methanol (2 : 1). There are numerous modifications of this method. Another useful method is the SPE extraction procedure for purification of lipids.[95,96] In sterol storage disease, accumulation of cholesterol occurs in blood and tissue. The use of TLC for cholesterol determination after extraction of serum with hexane has been reported.[93] The extract was subjected to peracetic acid followed by hydrogen sulfide oxidation. Separation of cholesterol was achieved in silica gel 60 HPTLC with heptane–ethyl acetate (60 : 40) and visualization by spraying with phosphomolybdic acid (PMA) reagent. Batta et al.[97] separated cholest-5-enzone and epicholesterol on silica gel plates using mobile phases containing different proportions of acetone in methylene chloride, chloroform, or carbon tetrachloride. Spots were detected by spraying the plates with PMA (3.5% in isopropanol) followed by sulfuric acid (20%) and heating at 110 °C for 2 min. Clinical biochemical and histological analyses of patients with cholesteryl ester storage disease and Wolfman's disease have been reported.[98] The overpressured and conventional TLC methods were utilized by Pucsok et al.[99] for separation of cholesteryl esters. Bhat and Ansari[100] reported an improved method for separation of chloro analogues of cholesterol and its acetate in tissue samples on silica gel TLC plates using mixtures of hexane–ethyl acetate as mobile phases. The plates were sprayed with 50% sulfuric acid followed by heating at 120 °C for 5 min to detect cholesteryl acetate.

The stratum corneum lipids are unique in composition and have been used frequently as a model of the skin's lipid barrier. Serizawa et al.[101] separated cholesterol sulfate from intercellular stratum corneum in studies related to horny layer adhesion. Extracted lipids were spotted on HPTLC plates, and developed in multiple mobile phases. Spots were detected by spraying with 10% cupric sulfate in 8% phosphoric acid followed by heating at 80 °C for 60 min. Edouard et al.[102] reported separation and quantification of free cholesterol and cholesteryl esters in skin biopsies by TLC and suggested this technique can be used as a diagnostic tool to detect pathologies of skin lipid metabolism. Zellmer and Lasch[103] used HPTLC/AMD with a 25-step gradient based on methanol, diethyl ether, and *n*-hexane to separate six major human plantar stratum corneum lipids. Postchromatographic staining of these lipids with a solution of $MnCl_2-H_2SO_4$ at 130 °C or a solution of $CuSO_4-H_3PO_4$ at 140 °C allowed visualization and quantification of the lipids.

Asmis et al.[104] have published a low-cost TLC procedure for the quantification of cholesterol, cholesteryl esters, and triglycerides in small biological samples using TLC and laser densitometry. Alteration of acidic lipids in human sera during the course of pregnancy also suggests a characteristic increase in the concentration of cholesterol sulfate.[105]

TLC is also used for separation of lipids from patients with alcoholic diseases such as alcoholic fatty liver which

is accompanied by a major increase in triglycerides and a smaller increase in cholesteryl esters.[106]

Analyses of the lipids of normal and Gaucher bone marrow using TLC and HPLC methods have also been reported.[107] Yamada et al.[108] have published a quantitative analysis of lipids from Schnyder's corneal dystrophy using the TLC/FID (flame ionization detection) technique. Findings suggest that this disorder involves a disturbance of the metabolism of cholesterol and/or sphingomyelin metabolism in the cornea. Ohta et al.[109] have developed a rapid and quantitative procedure for determination of the fatty acid composition of plasma lipid using silica gel plates with partial development in methanol followed by chloroform–methanol (1:1) and full development in hexane–diethyl ether–acetic acid (80:20:1). This method has proved quite useful for screening a large number of plasma samples. Kaphalia and Ansari[110] reported a simple and reliable single-step TLC method for quantitation of enzymatic formation of fatty acid anilides.

5.6 Phospholipids

Phospholipids are polar lipids and include phosphoglycerides and sphingolipids. These are amphipathic constituents of membrane. They play an essential role in the synthesis of plasma lipoprotein. The determination of phospholipids is important in biomedicine. Surface active phospholipids have an important influence on the mechanical behavior of the lungs. Clinical management of high-risk pregnancies often requires determination of fetal lung maturity status.[111–113] Measuring phospholipids in amniotic fluid as the lecithin/sphingomyelin ratio using the silica gel TLC method has been the established clinical procedure for predicting fetal lung maturity.[111–113] Radial HPTLC has also been employed to analyze phospholipids in more than 2000 specimens of amniotic fluid to assess fetal lung maturity.[114] Iwamori et al.[115] have reported a sensitive TLC-immunostaining procedure for the diagnosis of respiratory distress syndrome. A large number of papers have appeared.[115–118] Prior to chromatography, purification of phospholipids is essential. The extraction of lipids is often carried out by the method of Gluck and Kulovich (described in Siouffi et al.[2] and Jain[11]) or by the modified procedure of Kollins et al.[118] A SPE extraction procedure using C_{18} extraction cartridges has also been reported.[119,120] Both one- and two-dimensional TLC procedures have been reported for the estimation of phospholipids.[2] The common mobile phases used are mainly mixtures of chloroform–methanol–water, chloroform–methanol–ammonium hydroxide, and chloroform–methanol–glacial acetic acid–water.[121] Improved resolution of lecithin and sphingomyelin occurs with addition of either acid or base to the water. The use of HPTLC plates for analysis of phospholipids in amniotic fluid has also been reported.[93] Amniotic fluids were extracted with chloroform–methanol solution. Extracts were spotted on HPTLC plates predeveloped in chloroform–methanol (1:1). The plate was first developed in chloroform–ethanol–triethylamine–water (30:34:30:8) for separation of polar lipids, including sphingomyelin, phosphatidylcholine, lysophatidylcholine, phosphotidylserine, and phosphatidylinositol. After drying the plate was redeveloped in hexane–diethyl ether (50:50) to separate the cholesteryl ester fraction and resolve phosphatidylglycerol and cardiolipid . Detection was achieved by dipping the plates for 5 s in a 10% solution of $CuSO_4$ in 8% H_3PO_4 followed by heating at 120 °C. Quantitation was carried out with a spectrodensitometer at 310 nm.

The determination of phospholipids is also important in patients suffering from cystic fibrosis.[122] Plasma phospholipids were separated on HPTLC silica gel fluorescent plates impregnated with 0.75% boric acid in absolute ethanol with hexane–diethyl ether (6:4) mobile phase and the phospholipid fatty acid pattern was determined by GLC. Li et al.[123] have studied the detection of serum antibodies against phosphocoline containing aminoglycoglycerolipid specific to *Mycoplasma fermentans* in HIV (HIV-1) infected individuals using TLC.

Leray et al.[124] reported a rapid quantitative analysis of vitamin E, cholesterol, and phospholipid fatty acids in a single aliquot of human platelets and cultured endothelial cells. The whole procedure is based on the extraction of total lipids followed by TLC and microcolumn purification of tocopherols and cholesterols. The fatty acid compositions of phospholipid classes were separated on the same silica gel plates and determined by GLC.

Ceramides and 1,2-diacyl glycerol have been demonstrated in intracellular signalling pathways. Okumura et al.[125] developed a rapid method of simultaneous quantitative determination of ceramides and 1,2-diacylglycerol in tissues by the TLC/FID technique.

A sensitive TLC–densitometry method has been reported for the determination of the platelet activating factor (PAF) and phospholipids in human saliva and tears.[126,127] The mobile phase used for separation of lipids and phospholipids was hexane–diethylether–water (65:35:7). Detection was achieved by spraying the silica gel plates with phospholipase C followed by alkaline phosphatase solution at 45 °C to liberate phosphate from each phospholipid. The plate was resprayed with a mixture of ammonium molybdate and malachite green. The liberated phosphates appeared as blue-green spots of molybdophosphate–malachite green aggregation on a yellow-brown background. The absorbance of each spot

on the plate was measured at 620 nm with a densitometer for quantitative analysis.

TLC and laser densitometry have also been used to examine erythrocyte membrane phospholipids in psychotic patients.[128] Alteration of serum phosholipids is also seen in patients with multiple sclerosis.[129] TLC–densitometry was used for separation and quantification of these phospholipids. Analysis of phospholipid and their fatty acid composition from human intestinal mucosa was performed with two-dimensional TLC followed by lipid-phosphorus determination.[130]

There is an increasing interest in the role of antiphospholipid antibodies, and detection of these antibodies by immunostaining has proved very successful.[131] Khan and Glenton[132] have shown that patients with calcified kidney stones excrete more lipids and acidic phospholipids than normal individuals. These authors separated neutral lipids with hexane–ether–glacial acetic acid (70 : 30 : 1), glycolipids with chloroform–methanol–water (100 : 42 : 6) and phospholipids with chloroform–methanol–glacial acetic acid–water (75 : 45 : 3 : 1) on silica gel plates. Individual lipid spots were revealed by exposing the plate to iodine vapor and phospholipids spots were located by staining with molybdenum blue. Existence of a lipid gradient in the upper stratum corneum has also been established by Bonte et al.[133]

Numerous papers have also appeared on blotting of lipids, glycolipids, phospholipids, and gangliosides from an HPTLC chromatogram to a polyvinyldiene difluoride membrane.[134–137] TLC in combination with HPLC has been used for the assay of fatty acids composition of individual phospholipids in platelets from noninsulin-dependent diabetes mellitus patients.[138] Phospholipids of many cancer tumors also contain larger amounts of alkyl ether than healthy tissues. Narayan and Dahiya.[139] studied alterations in sphingomyelin and fatty acids in human prostatic hyperplasia and prostatic cancer by using TLC and GLC techniques. A number of reports have been published for separation and quantification of neutral lipids and phospholipids using TLC coupled with other techniques, and these methods have proved quite successful.[140–145]

5.7 Porphyrins

Porphyrins are tetrapyrrole derivatives with a porphine structure. They are chemical intermediates in the synthetics of hemoglobin, myoglobin, and other respiratory pigments called cytochromes. Detection and identification of porphyrins and their precursors in biological samples are of paramount therapeutic importance.[146–158] They are analyzed in clinical chemistry to aid in the diagnosis of a group of disorders called porphyrias, which result from disturbances in the heme biosynthesis. The accumulation and overproduction of these intermediate compounds in tissues, blood, urine, and feces indicate a metabolic block in heme synthesis. These disorders are associated with acute or cutaneous manifestations (or both).

Analysis of porphyrins in urine or feces involves lipid precipitation from acetone and subsequent extraction with chloroform before they are subjected to chromatography. Increasing numbers of papers advocate the use of HPLC and MS techniques for analysis of urinary and fecal porphyrin excretion patterns in humans.[146–152] Urinary and fecal porphyrin patterns are characteristic of each type of porphyrin. TLC is still widely used for the routine examination of porphyrins in clinical laboratories. Separation of porphyrins by TLC with prior derivatization offers a sensitive, rapid, and reliable method of clinical diagnosis.[151] The developing solvent systems that are commonly used for separation of porphyrins are binary mixture of hexane with methylene chloride, chloroform, or ethyl acetate, and ternary mixtures of benzene–ethyl acetate–methanol, toluene–ethyl acetate–methanol, methylene chloride–carbon tetrachloride–ethyl acetate, chloroform–kerosene–methanol or quaternary mixtures of methylene chloride–carbon tetrachloride–ethyl acetate–ethyl propionate. Free carboxylic porphyrins can be separated with the ion-pairing technique or an alkyl-bonded layer with a mobile phase consisting of acetonitrile-*N*-cetyl-*N*,*N*-trimethyl ammonium bromide, 0.1 mM in aqueous acetate buffer.[153]

Huie and Williams[154] reported an application of HPTLC with laser fluorimetric detection of porphyrin methyl esters in human urine. This method demonstrated good selectivity and detectablity for determination of porphyrin profiles. The detection limits for uro-, heptacarboxy-, hexacarboxy-, pentacarboxy-, copro-, and mesoporphyrin methyl esters were in the 18–35 pg range.

Jacob and Das[155] published the excretion pattern of fecal coproporphyrin isomers (I–IV) in human porphyrias. These authors developed a reliable method of sample preparation using RPTLC for the isolation of naturally occurring coproporphyrin-free carboxylic acids. Accurate separation and quantification of the individual isomers I–IV were achieved with the help of ion-pair, HPLC and on-line detection of their excitation and emission fluorescence spectra. These results clearly demonstrated the potential of fecal coporphyrin I–IV isomer ratio for the diagnosis and differential diagnosis of hereditary porphyrias.

Lai et al.[156] proposed an reversed-phase high-performance thin-layer chromatography (RPHPTLC) method for analysis of free porphyrins for diagnosis of porphyria. These authors observed more than 12 distinct bands of different porphyrins by viewing the plates under long-wavelength UV light. Simultaneous fluorescence detection of fecal urobilins and porphyrins by RPHPTLC

has also been published by Lam et al.[157] Silica gel TLC of free porphyrins for diagnosis of porphyria has also been reported by Handersan.[158] The distinct porphyrin bands were observed under long-wavelength UV light.

Porphyrins exhibit UV spectral absorption bands in the range 400–408 nm. On silica gel, the porphyrin fluorescence may be enhanced by dipping plates in a solution of dodecane–hexadecane–chloroform (1:1:18). The porphyrins can easily be detected in femtomole amounts on reversed-phase (RP) plates with λ_{ex} 404 nm and λ_{em} 560 nm; the calibration graph is linear between 0.15 and 3.0 pmol.[2,11]

5.8 Prostaglandins

PGs are one of the most biologically active family of compounds. They are unsaturated C_{20} fatty acids with a cyclopentane ring. Derivatives that contain this structure (PGs, prostacyclins, and thromboxanes) are known collectively as prostanoids. They occur in the nanogram or even picogram range in human tissues and body fluids. They have diverse biological action. Their full physiological role is not completely understood. A range of methodologies has been used for estimation of PGs in clinical laboratories. These include immunological assays, chromatographic techniques, and GC/MS.[159–165] Often HPLC, GLC, and GC/MS are used for quantitative measurement of PGs. Analyses of PGs by GLC and GC/MS involve sample extraction, thin-layer chromatographic purification, and derivatization. The overall procedure is time-consuming. Sample pretreatment is also essential in analysis of PGs and is mainly performed by using Sep-Pak™ columns[159,166] or styrene–divinylbenzene cross-linked copolymeric macroreticular ion-exchange resin XAD-2 as the solid-phase support.

Silica gel TLC is also used in analysis of PGs either as an analytical method or a cleanup procedure.[11] Raajmakers[167] proposed a TLC method for separation of all PG groups from pure substances using chloroform–methanol–acetic acid–water as a mobile phase. Tsunomoto et al.[168] describe a two-dimensional TLC procedure using ethylacetate isooctane–acetic acid–water (11.5:2:10) and chloroform–methanol–acetic acid (90:8:6) for separation of prostacyclin (PGI2) and thromboxane A2 (TxA2). These compounds are known to be involved in hemostases and oncogenesis. Various PGs (PGE2, PGE1, PGE2, PGA1) and some arachidonic acid metabolites were separated on silica gel modified with phenylmethylvinylchlorosilane RPTLC plates.[2] Alvarez Perez et al.[169] advocated TLC for the presence of PGE2 in normal and inflamed gingiva. Vasodilator PGE2 analogs[11] were separated either on silica gel G with chloroform–methanol–acetic acid (18:2:1) mobile phase or RP-12 plates with acetonitrile–water (35:65). Combinations of TLC and GC/tandem mass spectrometry (GC/MS/MS) or TLC, HPLC and GC/MS have been reported by several researchers for separation of PGs in biological samples.[162–165] Other recommended solvent systems[2] used for separation of major PGs are chloroform–isopropanol–ethanol–formic acid (45:50:0.5:0.3) and ethyl acetate–acetone–acetic acid (90:5:1). Most of the spray reagents used for detection of PGs following TLC separation, for example, concentrated sulfuric acid, PMA, 2,4-dinitrophenyl hydrazide, anisaldehyde/sulfuric acid, acidic ceric sulfate, and vanillin/phosphoric acid/ethanol are destructive. Nondestructive methods for detection include the use of iodine vapor spray and 8-hydroxy-1,3,6-pyrene-trisulfonic acid trisodium salt.[170] To improve detection limits, 4-bromomethyl-7-methoxycoumarin derivatives have been proposed for samples containing PGs.[171]

5.9 Steroid Hormones

Steroid hormones play a vital role in the human organism. They are defined by their physiological functions. The levels of the active steroid hormones in serum and their elimination products in urine are of great clinical significance. Glucocorticoid hormones contribute to adipogenic activity in human serum. Addisons' disease is caused by glucocorticoid deficiency and excessive secretion results in Cushings' syndrome. Estrogenic steroid concentrations in human urine are very important during pregnancy. Progesterone is also involved in the preparation and maintenance of pregnancy. Increased urinary levels of pregnane-3α, 20α-diol and allopregnane-3α, 20α-diol have been observed in early pregnancy. Elevation of 6β-hydroxycortisol is seen in human urine during pregnancy and in cases of hyperadrenocorticism. Urinary aldosterone is assayed to identify patients with hyperaldosteronism.[172] The urinary excretions of andro- and glucocorticoidenesis metabolites have also been studied in patients with gastric cancer.[173] Furthermore, dehydroepiandresterone, and in particular eticholanolone excretions were found to decrease significantly in the urine of women affected by malignant breast disease.[174]

PC, mainly TLC, is widely recognized for steroid analysis in clinical laboratories. Like other compounds, sample preparation is also necessary prior to TLC of steroids. Sample preparation is mainly done by enzymatic and acidic hydrolysis followed by liquid–liquid or liquid–solid extraction. Tang and Crone[175] have developed an efficient, effective, and inexpensive method for hydrolyzing steroid conjugates. Venturelli et al.[176] have suggested an improved multistep procedure for urinary testosterone analysis.

Silical gel is an effective sorbent for separating steroids. Several books and reviews dealing with the PC of steroids have been published.[2,6,7,11,177–179] Separation of urinary steroid hormones on silica gel have been studied in various mobile phases:[2,6,7,11] estrogens in butyl acetate–benzene (85:15), benzene–acetone (90:10), chloroform–ethyl acetate (80:20), chloroform–methanol–water (94:6:0.5); pregnanediol and pregnanetriol in chloroform–acetone (80:10) and chloroform–acetone–methanol (75:15:10); androgens in dichloromethane–methanol–water (225:15:1.5) and light petroleum (37–55 °C)–diethyl ether–acetic acid (48:50:2); corticosteroids in chloroform–methanol (98:2); and glucocorticoids in chloroform–dioxane–methanol (15:4:1). Other effective systems for separation of major estrogens, androgens, and pregnones are chloroform–acetone (9:1) and cyclohexane–ethylacetate–ethanol (77.5:20:2.5). Some researchers have also used two-dimensional and multidimensional development modes for separation of steroids.[180,181]

Tomosova[182] has published a method for analysis of urinary 17-oxosteroids labeled with dansylhydrazine by SPE on a C_{18} minicolumn and silica gel TLC with chloroform–methanol (97:3) mobile phase and scanning of fluorescence. Vingler et al.[183] reported a direct quantitative digital autoradiography thin-layer chromatographic method for separation of eight major androgens derived from testosterone metabolism. The separation was achieved on silica gel plates containing a fluorescent indicator and a concentration zone, and involved two different methods. The first method consisted of a double run using a mobile phase composed of dichloromethane–diethyl ether (90:10) and the second involved a double run with the same mobile phase followed by a second run at 90° with a mobile phase containing of toluene–ethyl acetate (50:50) when a single spot was to be run or toluene–ethyl acetate (70:30) when two tracks were to be run. Fenske and Schonheiter[184] used a silica-coated aluminum sheet separation and elution of glucocorticosteroids and sex steroids prior to radioimmunoassay. Agarwal et al.[185] have reported the TLC separation of regioselective and stereoselective androgen metabolites.

The introduction of alkyl-bonded plates has given rise to new experimental procedures in the PC of steroids. Watkins et al.[186] reported a method for the identification of estrogens on RP-18 and RP-8 layers with the ion-pairing mobile phase methanol–0.5% tetramethyl ammonium bromide solution in water (45:55 or 50:50). Lamparcyzk et al.[187] described methanol–water to be the best solvent system for separation of steroid hormones by RPHPTLC. Diol-bonded and cyano-bonded silica plates also permitted good separations of major estrogens and androgens.

Various spraying reagents have been used for locating steroids in situ. Most of them have been adequately discussed by Lamparczyk[178] The most popular spraying reagent used for detecting steroids is 10% sulfuric acid in methanol, followed by heating for 10 min at 90 °C and fluorescence scanning λ_{ex} 366 nm and λ_{em} 509 nm for quantification. Other recommended reagents are rubeanic acid, a mixture of copper sulfate and *o*-phosphoric acid, fast violet-salt B, manganese(II) chloride with subsequent heating at 120 °C and scanning at 366 nm, or charring of steroids on ammonium sulfate preimpregnated plates after a *p*-toluene sulfuric acid spray. Spraying with 0.1 mM 2,3-dichloro-5,6-dicyano-1,4-benzoquinone in toluene–acetic acid (70:30) and densitometry have been found to be very sensitive for estriol 16β-glucuronide determination. New chromogenic reagents for detection of steroids are: 25% solution of tribromoacetic acid in chloroform, *p*-chloranil, *p*-fluoranil, and 2,3-dichloro 5,6-dicyanobenzoquinone.[7] Moreover, coupling devices[11] have also been advocated for detection and identification of steroids.

5.10 Other Compounds of Clinical Interest

Other compounds of clinical interest are gangliosides,[188] cerebrosides,[93] purines, pyrimidine, derivatives of nucleic acid,[189–192] and urinary organic acids.[193,194] These have been analyzed by TLC in various laboratories.

6 CONCLUSION

In the light of above discussions, it may be seen that considerable advancements have taken place in the area of PC technique. These include chromatographic separations, methods of detection, instrumental TLC, and hyphenated instruments. Thus the use of PC technique in clinical chemistry has widened. Much more advancement is expected in the near future.

ABBREVIATIONS AND ACRONYMS

AMD	Automated Multiple Development
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared Spectrometry
GC	Gas Chromatography
GLC	Gas–Liquid Chromatography
HIV	Human Immunodeficiency Virus
HPLC	High-pressure Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography

LLE	Liquid–Liquid Extraction
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NPTLC	Normal-phase Thin-layer Chromatography
PAF	Platelet Activating Factor
PC	Planar Chromatography
PG	Prostaglandin
PMA	Phosphomolybdic Acid
PMD	Programmed Multiple Development
PTH	Phenyl Thiohydantoin
QC	Quality Control
QTLC	Quantitative Thin-layer Chromatography
RP	Reversed Phase
RPHPTLC	Reversed-phase High-performance Thin-layer Chromatography
RPTLC	Reversed-phase Thin-layer Chromatography
SPE	Solid-phase Extraction
TLC	Thin-layer Chromatography
UV	Ultraviolet

RELATED ARTICLES

Liquid Chromatography **(Volume 13)**
Silica Gel and its Derivatization for Liquid Chromatography • Thin-layer Chromatography

General Articles **(Volume 15)**
Spot Test Analysis

REFERENCES

1. F.A.A. Dallas, H. Read, R.J. Ruane, I.D. Wilson (eds.), *Recent Advances in Thin-layer Chromatography*, Plenum Press, New York, 1988.
2. A.M. Siouffi, E. Mincsovics, E. Tyihak, 'Planar Chromatographic Techniques in Biomedicine: Current Status', *J. Chromatogr.*, **492**, 471–538 (1989).
3. C.F. Poole, S.K. Poole, 'Progress in Densitometry for Quantitation in Planar Chromatography', *J. Chromatogr.*, **492**, 539–584 (1989).
4. K. Bauer, L. Gros, W. Sauer, *Thin-layer Chromatography – An Introduction*, E.M. Science, A Division of E. Merck, Darmstadt, Germany, 1991.
5. J.C. Touchstone, *Practice of Thin-layer Chromatography*, Wiley-Interscience, New York, 1992.
6. J. Sherma, 'Planar Chromatography', *Anal. Chem.*, **64**, 134R–147R (1992).
7. J. Sherma, 'Planar Chromatography', *Anal. Chem.*, **66**, 67R–83R (1994).
8. B. Fried, J. Sherma, *Thin-layer Chromatography Practice and Application*, Dekker, New York, 1994.
9. B. Fried, J. Sherma (eds.), *Practical Thin-layer Chromatography, A Multidisciplinary Approach*, CRC Press, Boca Raton, FL, 1996.
10. J.R. McMurray, 'Collection of Specimens', in *Practical Clinical Biochemistry*, 6th edition, eds. A.H. Gowenlock, J.R. McMurray, M. McLauchlan, CBS Publishers and Distributors, New Delhi, 1996.
11. R. Jain, 'Thin-layer Chromatography in Clinical Chemistry', in *Practical Thin-layer Chromatography, A Multidisciplinary Approach*, eds. B. Fried, J. Sherma, CRC Press, Boca Raton, FL, 131–152, 1996.
12. Report of the Substance Abuse Testing Committee, 'Critical Issues in Urinalysis of Abused Substances', *Clin. Chem.*, **34**, 605–632 (1988).
13. R. Jain 'Interference of Adulterants in Thin-layer Chromatography Methods for Drugs of Abuse', *Indian J. Pharmacol.*, **25**, 240–242 (1993).
14. R. Jain, 'Laboratory Services for Drug Abuse Screening in a Clinical Setting', *J. Forensic. Med. Toxicol.*, 26–31 (1994).
15. R. Jain, R. Ray, 'Detection of Drugs of Abuse and its Relevance to Clinical Practice', *Indian J. Pharmacol.*, **27**, 1–6 (1995).
16. H. Ko, E.N. Petzold, 'Isolation of Samples Prior to Chromatography', in *GLC and HPLC Determination of Therapeutic Agents*, Part 1, eds. K. Tsuji, W. Morozwich, Marcel Dekker, New York, 1978.
17. M.R. Harkey, M.L. Stalowitz, 'Solid Phase Extraction Techniques for Biological Specimens', in *Advances in Analytical Toxicology*, ed. R.C. Baselt, Biomedical Publications, Foster City, 255–270, 1984.
18. R.D. McDowall, J.C. Pearce, G.S. Murkitt, 'Liquid–Solid Sample Preparation in Drug Analysis', *J. Pharm. Biomed. Anal.*, **4**(1), 3–21 (1989).
19. R.D. McDowall, 'Sample Preparation for Biomedical Analysis', *J. Chromatogr.*, **492**, 3–58 (1989).
20. W.A. Baumgartner, V.A. Hill, 'Sample Preparation Techniques', *Forensic Sci. Int.*, **63**, 121–135 (1993).
21. F. Degel, 'Comparison of New Solid-phase Extraction Methods for Chromatographic Identification of Drugs in Clinical Toxicological Analysis', *Clin. Biochem.*, **29**(6), 529–540 (1996).
22. H. Lingeman, S.J. Hoekstra-Oussoren, 'Particle-loaded Membranes for Sample Concentration and/or Clean-up in Bioanalysis', *J. Chromatogr., B*, **689**(1), 221–237 (1997).
23. I.D. Wilson, 'Contemporary Thin-layer Chromatography: An Introduction', in *Recent Advances in Thin-layer Chromatography*, eds. F.A.A. Dallas, H. Read, R.J. Ruane, I.D. Wilson, Plenum Press, New York, 1–20, 1988.
24. J. Sherma, 'Basic Techniques Materials and Apparatus', in *Handbook of Thin-layer Chromatography*, eds. J. Sherma, B. Fried, Marcel Dekker, New York, 1991.

25. T. Cserhati, E. Forgacs, 'Introduction to Techniques and Instrumentation', in *Practical Thin-layer Chromatography, A Multidisciplinary Approach*, eds. B. Fried, J. Sherma, CRC Press, Boca Raton, 1–18, 1996.
26. D.E. Janchen, 'Instrumental Thin-layer Chromatography, State of the Art and Future Trends', *Int. Analyst*, **13**, 36–40 (1987).
27. M.A. Edwards, S. Grant, A. Green, 'A Practical Approach to the Investigation of Amino Acid Disorders', *Ann. Clin. Biochem.*, **25**(2), 129–141 (1988).
28. A. Hestnes, O. Borud, H. Lunde, L. Gjessing, 'Cystathioninuria, in Down's Syndrome', *J. Mental Deficiency Res.*, **33**(Pt3), 261–265 (1989).
29. J.T. Wu, 'Screening for Inborn Errors of Amino Acid Metabolism', *Ann. Clin. Lab. Sci.*, **21**(2), 123–142 (1991).
30. M. Duran, L. Dorland, P.K. Bree, R. Berger, 'Selective Screening for Amino Acid Disorders', *Eur. J. Pediat.*, **157**(7 Suppl), S33–S37 (1994).
31. M. Kaur, G.P. Das, I.C. Verma 'Inborn Errors of Amino Acid Metabolism in North India', *J. Inherit. Metab. Res.*, **17**(2), 230–233 (1994).
32. D.J. Byrd, M. Lind, J. Brodehl, 'Diagnostic and Genetic Studies in 43 Patients with Classic Cystinuria', *Clin. Chem.*, **37**(1), 68–73 (1991).
33. R. Giuglani, I. Ferrari, L.J. Greene, 'An Evaluation of Four Methods for the Detection of Heterozygous Cystinuria', *Clin. Chim. Acta*, **164**, 227–223 (1987).
34. D. Biou, N. Queyrei, M. Visseaux, I. Collignon, M. Pays, 'Separation and Identification of Dansylated Serum and Urinary Amino Acids by Two Dimensional Thin-layer Chromatography', *J. Chromatogr.*, **226**, 477–483 (1981).
35. B. Bartosch, W. Vycudilik, C. Popow, G. Lubec, 'Urinary 3-Hydroxyproline Excretion in Alport's Syndrome: A Non-invasive Screening Test', *Arch. Dis. Child*, **66**(2), 248–251 (1991).
36. R. Bhushan, V.K. Mahesh, A. Varma, 'Improved Thin-layer Chromatographic Resolution of PTH Amino Acids with Some New Solvent Systems', *Biomed. Chromatogr.*, **8**(2), 69–72 (1994).
37. R. Bhushan, J. Mertens, 'Direct Resolution of Enantiomers by Impregnated TLC', *Biomed. Chromatogr.*, **11**(5), 280–285 (1997).
38. R. Bhushan, J. Martens, S. Wallbauem, S. Joshi, V. Prasad, 'TLC Resolution of Enantiomers of Amino Acids and Dansyl Derivatives Using (1*R*, 3*R*, 5*R*)-2-Azabicyclo (3,3,0) Octan-3-Carboxylic Acid as Impregnating Reagent', *Biomed. Chromatogr.*, **11**(5), 286–288 (1997).
39. S. Laskar, U. Bhattacharya, B.K. Barels, 'Modified Ninhydrin Spray Reagent for the Identification of Amino Acids on TLC Plates', *Analyst*, **116**, 625–626 (1991).
40. S. Laskar, B.K. Basek, 'Identification of Amino Acids within Ninhydrin in Acid Media', *J. Planar. Chromatogr.-Mod. TLC*, **3**, 535–536 (1990).
41. S. Laskar, B.K. Basek, 'Identification of Amino Acids on TLC Plates', *J. Planar Chromatogr.-Mod. TLC*, **3**, 275–276 (1990).
42. J.M. Street, D.J.H. Trafford, H.L.J. Makin, 'The Quantitative Estimation of Bile Acids and Their Conjugates in Human Biological Fluids', *J. Lipid. Res.*, **24**(5–8), 491–511 (1983).
43. T.A. Robb, G. Daridson, 'Analysis of Individual Bile Acids and Their Glycine/Taurine Conjugates by High-performance Thin-layer Densitometry', *Ann. Clin. Biochem.*, **21**(Pt2), 137–140 (1984).
44. S. Scalia, 'Bile Acid Separation', *J. Chromatogr., B*, **671**(1–2), 299–317 (1995).
45. M. Kindel, H. Ludwig-Kochin, M. Lembcke, 'New and Versatile Method for the Determination of Faecal Bile Acids by Thin-layer Chromatography with Direct Scanning Fluorimetry', *J. Chromatogr.*, **497**, 139–146 (1989).
46. H.C.S. Ferreira, W.H. Ellial, 'Pre-column Derivatization of Free Bile Acids for High-performance Liquid Chromatographic/Mass Spectrometric Analysis', *J. Chromatogr.*, **562**(1–2), 697–712 (1991).
47. T. Ida, M. Momosa, T. Shinohara, T.N.T. Nambara, F.C. Chang, 'Separation of Allo Bile Acid Stereoisomers by Thin-layer and High-performance Liquid Chromatography', *J. Chromatogr.*, **366**, 396–402 (1986).
48. A. Vanden Ende, C.E. Radecker, W.M. Mairuhu, 'Microanalysis of Free and Conjugated Bile Acids by Thin-layer Chromatography and In-situ Spectrofluorimetry', *Anal. Biochem.*, **134**(1), 153–162 (1983).
49. R. Klaus, 'Procedures for Quantitative High Pressure Liquid Chromatographic Analysis of Non-fluorescing Compounds with Bile Acids as an Example, *J. Chromatogr.*, **333**(1), 276–287 (1985).
50. D. Tonelli, E. Gattarecchyia, G. Mazzell, A. Rodo, 'Bile Acid Kinetics in Man Studied by Radio Thin-layer Chromatography and Densitometry Coupling', *J. Chromatogr., B*, **700**(1–2), 59–66 (1997).
51. K. Robards, M. Whitelaw, 'Chromatography of Monosaccharides and Disaccharides', *J. Chromatgr.*, **373**, 81–110 (1986).
52. I.S. Menzies, J.N. Mount, J.N. Wheeler, 'Quantitative Estimation of Clinically Important Monosaccharides in Plasma by Rapid Thin-layer Chromatography', *Ann. Clin. Biochem.*, **15**, 65–76 (1978).
53. A.J. Flick, R.L. Schnaar, J.A. Perman, 'Thin-layer Chromatographic Determination of Urinary Excretion of Lectulose, Simplified and Applied to Cystic Fibrosis Patients', *Clin. Chem.*, **33**, 1211–1217 (1987).
54. H. Susanne, F. Sorensen, F. Janice Proud, A. Adam, H. Carolein Rutgers, R.M. Batt, 'A Novel HPLC Method for Simultaneous Quantification of Monosaccharides and Disaccharide Used in Tests of Intestinal Function and Permeability', *Clin. Chim. Acta*, **221**, 115–125 (1993).

55. S.G. Lim, I.S. Menzies, W.S. Nukajam, C.A. Lee, M.A. Johnson, R.E. Pounder, 'Intestinal Disaccharides Activity in Human Immuno Deficiency Virus Disease', *Scand. J. Gastroenterol.*, **30**(3), 235–241 (1995).
56. Y. Bao, T.M. Silva, R.L. Gurrant, A.M. Lima, J.W. Fox, 'Direct Analysis of Mannitol, Lactulose and Glucose in Urine Samples by High-performance Anion Exchange Chromatography with Pulse Amperometric Detection. Clinical Evaluation of Intestinal Permeability in Human Immunodeficiency Virus Infection', *J. Chromatogr., B*, **685**(1), 105–112 (1996).
57. R. Humbel, M. Collart, 'Oligosaccharides in Urine of Patients with Glycoprotein Storage Diseases. I. Rapid Detection by Thin-layer Chromatography', *Clin. Chim. Acta*, **60**, 143–145 (1975).
58. A.C. Sewell, 'An Improved Thin-layer Chromatographic Method for Urinary Oligosaccharide Screening', *Clin. Chim. Acta*, **92**, 411–414 (1979).
59. W. Blom, J.C. Luteyn, H. Hkelholt Oijkman, J.B. Huismans, M.S.B. Loonen, 'Thin-layer Chromatography of Oligosaccharides in Urine as a Rapid Indication for the Diagnosis of Lysosomal and Maltose Deficiency (Pome's Disease)', *Clin. Chim. Acta*, **134**, 221–227 (1983).
60. A. Lugowska, A. Tylki-Szymanska, D. Sawnor-Korszynska, 'Thin-layer Chromatography of Urine Oligosaccharides in Diagnosis of Some Lysosomal Storage Disorders', *Pediatr. Pol.*, **70**(10), 847–855 (1995).
61. N.G. Abeling, H. Rusch, A.H. Van Gennip, 'Improved Selectivity of Urinary Oligosaccharide Screening Using Two One Dimensional TLC Systems', *J. Inherit. Metab. Res.*, **19**(2), 260–262 (1996).
62. H. Higashi, H. Hirabayashi, M. Ito, T. Yamagata, M. Matsumoto, S. Veda, S. Kato, 'Immunostaining on Thin-layer Chromatograms of Oligosaccharides Released from Gangliosides by Endoglycoceramidase', *J. Biochem.*, **102**, 291–296 (1987).
63. J. Vajda, J. Pick, *J. Planar. Chromatogr.*, **1**, 347 (1987), quoted from A.M. Siouffi, E. Mincsovics, E. Tyihak, 'Planar Chromatographic Technique in Biomedicine: Current Status', *J. Chromatogr.*, **492**, 471–538 (1989).
64. R. Bhushan, S. Kaur, 'TLC Separation of Some Common Sugars on Silica Gel Plates Impregnated with Transition Metal Ions', *Biomed. Chromatogr.*, **11**(1), 59–60 (1997).
65. K.D. Eskridge, S.K. Ejuthrie, 'Clinical Issues Associated with Urine Testing of Substance of Abuse', *Pharmacotherapy*, **17**(3), 497–510 (1997).
66. R. Jain, 'Analytical Methods', in *Detection of Drugs of Abuse in Body Fluids, A Manual for Laboratory Personnel*, ed. R. Jain, Drug Dependence Treatment Centre, All India Insitute of Medical Sciences. New Delhi, 18–44, 1998.
67. R.A. de Zeeuw, 'Drug Screening in Biological Fluids. The Need for a Systematic Approach', *J. Chromatogr., B*, **689**(1), 71–79 (1997).
68. A.K. Singh, K. Granley, M. Ashraf, U. Mishra, 'Drug Screening and Confirmation by Thin-layer Chromatography', *J. Planar Chromatogr.*, **2**, 410–419 (1989).
69. T.J. Siek, C.W. Stradling, M.W. McCain, T.C. Mehary, 'Computer-aided Identifications of Thin-layer Chromatographic Patterns in Broad-spectrum Drug Screening', *Clin. Chem.*, **43**(4), 619–626 (1997).
70. C.H.W. Gorodetzky, 'Detection of Drugs of Abuse in Biological Fluids', in *Handbook of Experimental Pharmacology, Drug Addiction*, ed. W.R. Martin, Springer-Verlag, New York, 319–407, Vol. 11, 1977.
71. I. Ojanpera, 'Thin-layer Chromatography in Forensic Toxicology', in *Practical Thin-layer Chromatography, A Mutidisciplinary Approach*, eds. B. Fried, J. Sherma, CRC Press, Boca Raton, FL, 193–230, 1996.
72. R.A. Brathwaite, D.R. Jarvie, P.S.B. Minty, D. Simpson, B. Widdop, 'Screening for Drugs of Abuse. I: Opiates, Amphetamines and Cocaine', *Ann. Clin. Biochem.*, **32**, 123–153 (1995).
73. K. Wolff, M.J. Sanderson, A.W.M. Hay, I. Barnes, 'An Evaluation of the Measurement of Drugs of Abuse by Commercial and In House Horizontal Thin-layer Chromatography', *Ann. Clin. Biochem.*, **30**, 163–168 (1993).
74. J.F. Wilson, B.L. Smith, P.A. Toseland, J. Williams, D. Burnett, A.D. Hirst, I.D. Watson, A.N. Horn, 'External Quality Assessment of Techniques for Detection of Drugs of Abuse in Urine', *Ann. Clin. Biochem.*, **31**, 335–342 (1994).
75. I. Ojanpera, P. Lillsunde, J. Vartiavaara, E. Vuori, 'Screening for Amphetamine with a Combination of Normal and Reserved Phase, Thin-layer Chromatography and Visualization with Fast Black K. Salt', *J. Planar Chromatogr.*, **4**, 373–378 (1991).
76. J.A. Vinson, D.J. Lopatofsky, 'A Semi-automated Extraction and Spotting System For Drug Analysis by TLC. I. Procedure for Analysis of the Major Metabolite of 9-Tetrahydrocannabinol in Urine', *J. Anal. Toxicol.*, **9**, 6–9 (1985).
77. R. Jain, 'Determination of Buprenorphine and its *N*-Dealkylated Metabolite in Urine by TLC Densitometry', *Indian J. Pharmacol.*, **26**, 288–291 (1994).
78. R. Jain, 'Simultaneous Determination of Diazepam and its Metabolites in Urine by Thin-layer Chromatography and Direct Densitometry', *J. Chromatogr.*, **615**, 365–368 (1998).
79. T. Inoue, T. Ni Waguchi, 'Determination of Nitrazepam and its Metabolite in Urine by Thin-layer Chromatography and Direct Densitometry', *J. Chromatogr.*, **339**, 163–169 (1985).
80. R. Jain, R. Ray, B.M. Tripathi, C. Singh, 'Opiate Excretion Profile Among Heroin Dependent Subjects', *Indian J. Pharmacol.*, **28**, 220–223 (1996).
81. C.W. Peng, W.L. Chiou, 'Analysis of Drugs and Other Toxic Substances in Biological Samples for Pharmacokinetc Studies', *J. Chromatogr.*, **531**, 3–50 (1990).

82. A. Shigematsu, N. Motoji, A. Hatori, T. Satoh, 'Progressive Application of Autoradiography in Pharmacokinetic and Metabolic Studies for the Development of New Drugs', *Regul. Toxicol. Pharmacol.*, **22**(2), 122–142 (1995).
83. K.K. Pandya, V.P. Mody, M.C. Satia, I.A. Modi, R.I. Modi, B.K. Chakravarthy, T.P. Gandhi, 'High-performance Thin-layer Chromatographic Method for the Detection and Determination of Lansoprazole in Human Plasma and its Use in Pharmacokinetic Studies', *J. Chromatogr., B*, **693**(1), 199–204 (1997).
84. W. Morden, I.D. Wilson, 'The Detection and Characterization of Analgesics and Anti-inflammatory Drugs on High-performance Thin-layer Chromatography Plates Using Tandem Mass Spectrometry: Application to Drugs and Metabolites in Urine', *Rapid. Commun. Mass Spectrom.*, **10**(15), 1951–1955 (1996).
85. K.K. Pandy, M.C. Satia, I.A. Modi, R.I. Modi, B.K. Chakravarthy, T.P. Gandhi, 'High-performance Thin-layer Chromatography for the Determination of Nimesulide in Human Plasma, and its Use in Pharmacokinetic Studies', *J. Pharm. Pharmacol.*, **49**(8), 773–776 (1997).
86. K. Otsuho, H. Seto, K. Futagmi, R. Oishi, 'Rapid and Sensitive Detection of Benzodiazepines and Zopiclone in Serum Using High-performance Thin-layer Chromatography', *J. Chromatogr., B*, **669**(2), 408–412 (1995).
87. M.I. Euegen'ev, S.Iu. Garmonov, V.I. Pogorol'tsev, Kh.I. Kurtbelialova, D.A. Valimukhametova, 'Determination of Acetylation Phenotype in the Therapeutic Monitoring of Drugs', *Klin. Lab. Diag.*, **5**, 24–27 (1996).
88. C.B. Eap, P. Baumann, 'Analytical Methods for the Quantitative Determination of Selective Serotonin Reuptake Inhibitors for the Therapeutic Drug Monitoring Purposes in Patients', *J. Chromatogr., B*, **686**(1), 51–63 (1996).
89. V.U. Dhavse, D.V. Parmar, D.V. Devarajan, 'High-performance Thin-layer Chromatographic Determination of Flurbiproten in Plasma', *J. Chromatogr., B*, **694**(2), 449–453 (1997).
90. P.J. Weldon, 'Thin-layer Chromatography of the Skin Secretions of Vertebrates', in *Practical Thin-layer Chromatography, A Multidisciplinary Approach*, eds. B. Fried, J. Sherma, CRC Press, Boca Raton, FL, 105–130, 1996.
91. E.W. Hammond, *Chromatography for the Analysis of Lipids*, CRC Press, Boca Raton, FL, 1993.
92. K.D. Mukherjee, N. Weber, (eds.), *CRC Handbook of Chromatography, Analysis of Lipids*, CRC Press, Boca Raton, FL, 1993.
93. J.C. Touchstone, 'Thin-layer Chromatographic Procedures for Lipid Separation', *J. Chromatogr., B*, **671**, 169–195 (1995).
94. J.M. Folch, M. Lees, C.H. Sloane-Stanley, 'A Simple Method for the Isolation and Purification of Total Lipids from Animal Tissues', *J. Biol. Chem.*, **226**, 497–570 (1997).
95. J.G. Alvarez, J.C. Touchstone, 'Separation of Acidic and Neutral Lipids by Aminopropyl Bonded Silica Gel Column Chromatography', *J. Chromatogr.*, **577**, 142–145 (1992).
96. T.G. Bernhardt, P.A. Connistraro, D.A. Bird, K.M. Doyle, M. Laposata, 'Purification of Fatty Acid Ethyl Esters by Solid Phase Extraction and High-performance Liquid Chromatography', *J. Chromatogr., B*, **675**(2), 189–196 (1996).
97. A.K. Batta, G. Salen, S. Shefer, 'Chromatographic Separation of Putative Precursors of Cholesterol', *Steroids*, **52**, 109–122 (1988).
98. A. Tylki-Szymanska, J. Rujner, A. Lugowaska, D. Sawnor-Korszynska, B. Wozniewicz, E. Czarnowrka, 'Clinical Biochemical and Histological Analysis of Seven Patients with Cholesteryl Ester Storage Disease', *Acta Paediatr. Jpn.*, **39**(6), 643–646 (1997).
99. W. Pucsok, L. Kovacs, A. Zalka, R. Dobo, 'Separation of Lipids by New Thin-layer Chromatography and Overpressured Thin-layer Chromatography Methods', *Clin. Biochem.*, **21**(2), 81–105 (1988).
100. H.K. Bhat, G.A.S. Ansari, 'Improved Separation of Lipid Esters by Thin-layer Chromatography', *J. Chromatogr.*, **483**, 369–378 (1988).
101. S. Serizawa, K. Osawa, K. Togashi, A. Yamamoto, M. Ito, S. Hamanka, F. Otsuka, 'Relationship between Cholesterol Sulfate and Intercellular Cohesion of the Stratum Corneum: Demonstration using a Push Pull Meter and an Improved High-performance Thin-layer Chromatography Separation System of All Major Stratum Corneum Lipids', *J. Invest. Dermatol.*, **99**, 232–242 (1992).
102. L. Edouard, F. Doucel, J.C. Buxtof, J.L. Beaumont, 'Qualitative Determination of Free Cholesterol and Cholestryl Esters in Skin Biopsies', *Clin. Physiol. Biochem.*, **3**, 323–329 (1985).
103. S. Zellmer, J. Lash, 'Individual Variation of Human Plantar Stratum Corneum Lipids, Determined by Automated Multiple Development of High-performance Thin-layer Chromatography Plates', *J. Chromatogr., B*, **691**(2), 321–329 (1997).
104. R. Asmis, E. Buhler, J. Jek, K.F. Gey, 'Concurrent Quantification of Cellular Cholesterol, Cholesteryl Esters and Triglycerides in Small Biological Samples. Revaluation of Thin-layer Chromatography Using Laser Densitometry', *J. Chromatogr., B*, **691**(1), 59–67 (1997).
105. B. Lin, K. Kubushiro, Y. Akiha, Y. Cui, K. Tsukazaki, S. Nozawa, M. Iwamori, 'Alteration of Acetic Lipids in Human Sera During the Course of Pregnancy: Characteristic Increase in the Concentration of Cholesterol Sulfate', *J. Chromatogr., B*, **704**(1–2), 99–104 (1997).
106. T.J. Peters, S.K. Cairns, 'Analysis and Subcellular Localization of Lipids in Alcoholic Liver Disease', *Alcohol*, **2**, 447–451 (1985).
107. S.P. Miller, G.C. Zirzow, S.H. Doppelt, R.O. Brady, N.W. Barton, 'Analysis of the Lipids of Normal and

Gaucher Bone Marrow', *J. Lab. Clin. Med.*, **127**(4), 353–358 (1996).
108. M. Yamada, H. Mochizuki, Y. Kamata, Y. Nakamura, Y. Mashima, 'Quantitative Analysis of Lipid Deposits from Schnyder's Corneal Dystrophy', *Br. J. Opthalmol.*, **82**(4), 444–447 (1998).
109. A. Ohta, M.C. Mayo, N. Kramer, W.E.M. Lands, 'Rapid Analysis of Fatty Acids in Plasma Lipids', *Lipids*, **25**, 742–427 (1990).
110. B.S. Kaphalia, G.A. Ansari, 'Single Step Thin-layer Chromatographic Method for Quantitation of Enzymatic Formation of Fatty Acid Anilides', *J. Chromatogr., B*, **705**(2), 269–275 (1998).
111. L. Gluck, M.U. Kulovich, R.C. Boerc, Jr, P.H. Brenner, G.G. Anderson, W.N. Spellacy, 'Diagnosis of Respiratory Distress Syndrome by Amniocentesis', *Am. J. Obstet. Gynecol.*, **109**, 440–445 (1971).
112. L. Gluck, M.V. Kulovich, 'Lecithin–Sphingomyelin Ratio in Amniotic Fluid in Normal and Abnormal Pregnancy', *Am. J. Obstet. Gynecol.*, **115**, 539–546 (1973).
113. M.V. Kulovich, M.B. Hallman, L. Gluck, 'The Lung Profile I: Normal Pregnancy', *Am. J. Obstet. Gynecol.*, **135**(1), 57–63 (1979).
114. J.R. Mackenzie, M. Truesdale, 'Radial High-performance Thin-layer Chromatography to Assess Fetal Lung Maturity', *Clin. Chem.*, **36**, 728–731 (1990).
115. M. Iwamori, K. Hirota, T. Utsuki, K. Monoeda, K. Ono, Y. Tisuchida, K. Okumura, K. Hanoka, 'Sensitive Method for the Determination of Pulmonary Surfactant Phospholipid/Sphingomyelin Ratio in Human Amniotic Fluids for the Diagnosis of Respiratory Distress Syndrome by Thin-layer Chromatography–Immunostaining', *Anal. Biochem.*, **238**(1), 29–33 (1996).
116. K.Z. Liu, M.K. Ahmed, T.C. Dembinski, H.H. Mantsch, 'Prediction of Fetal Lung Maturity From Near-infrared Spectra of Amniotic Fluid', *Int. J. Gynaecol. Obstet.*, **57**(2), 161–168 (1997).
117. K.Z. Liu, T.C. Dembinski, H.H. Mantsch, 'Rapid Determination of Fetal Lung Maturity from Infrared Spectra of Amniotic Fluid', *Am. J. Obstet. Gynecol.*, **178**(2), 234–241 (1998).
118. M.D. Kollins, E. Epstein, W.H. Civin, S. Weiner, 'Amniotic Fluid Phospholipids by Continuous Development Thin-layer Chromatography', *Clin. Chem.*, **26**, 403–405 (1980).
119. J.C. Touchstone, J.G. Alvarez, 'Preparation of Acidic and Neutral Glycolipids by Diphasic Thin-layer Chromatography', *J. Chromatogr.*, **436**, 515–516 (1988).
120. J.C. Touchstone, J.G. Alvarez, 'Phospholipids in Ammotic Fluids with Special Reference to the Lecithin/Sphingomyelin Ratio', *J. Chromatogr.*, **429**, 359–371 (1998).
121. D.E. Freer, B.E. Statland, 'Measurement of Amniotic Fluid Surfactant', *Clin. Chem.*, **27**(10–12), 1629–1641 (1981).
122. M. Roulet, P. Frascarolo, I. Rappaz, M. Pilet, 'Essential Fatty Acid Deficiency in Well Nourished Young Cystic Fibrosis Patients', *Eur. J. Pediatr.*, **156**(12), 952–956 (1997).
123. J.L. Li, K. Matsuda, M. Takagi, N. Yamamoto, 'Detection of Serum Antibodies Against Phosphocholine Containing Aminoglycoglycerolipid Specific to Mycoplasma Fermentans in HIV-1 Infected Individuals', *J. Immunol. Methods*, **208**(1), 103–113 (1997).
124. C. Leray, M. Andriamampandry, G. Gutbier, J. Cavadenti, C. Klein-Soyer, C. Gachet, J.P. Cazenave, 'Quantitative Analysis of Vitamin E, Cholesterol and Phospholipid Fatty Acids in a Single Aliquot of Human Platelets and Cultured Endothelial Cells', *J. Chromatogr.*, **696**(1), 33–42 (1997).
125. K. Okumura, K. Hayashi, I. Morishima, K. Murase, H. Matsui, Y. Toki, T. Ito, 'Simultaneous Quantitation of Ceramides and 1,2-Diacylglycerol in Tissues by Iatroscan Thin-layer Chromatography/Flame-ionization Detection', *Lipids*, **33**(5), 529–532 (1998).
126. C. Matsubara, T. Ohyama, K. Takamura, 'Densitometric Quantitation of Platelet Activating Factor and Other Phospholipids in Human Saliva Using Enzyme Reaction on a Silica Plate', *Yakugaku Zasshi*, **114**(9), 681–690 (1994).
127. T. Ohyama, C. Matsubara, K. Takamura, 'Sensitive Densitometry for the Determination of Platelet-activating Factor and Other Phospholipids in Human Tears', *Analyst*, **121**(12), 1943–1947 (1996).
128. M.S. Keshavan, A.G. Mallinger, J.W. Pettegrew, C. Dippold, 'Erythrocyte Membrane Phospholipids in Psychotic Patients', *Psychiatry Res.*, **49**(1), 89–95 (1993).
129. X.W. Tong, Q.M. Xue. 'Alterations of Serum Phospholipids in Patients with Multiple Sclerosis', *Clin. Med. J.*, **106**(9), 650–654 (1993).
130. K. Nakanishi, E. Yasugi, H. Morita, T. Dohi, M. Oshima, 'Plasmenylethanolamine in Human Mucosa Detected by An Improved Method for Analysis of Phospholipid', *Biochem. Mol. Biol. Int.*, **33**(3), 457–462 (1994).
131. M. Sorice, T. Griggi, A. Circella, T. Garafalo, F. d'Agostino, V. Pittoni, G.M. Pontieri, L. Lenti, G. Valesini, 'Detection of Antibodies by Immunostaining on Thin-layer Chromatography Plates', *J. Immunol. Methods*, **173**(1), 49–54 (1994).
132. S.R. Khan, P.A. Glenton, 'Increased Urinary Excretion of Lipids by Patients with Kidney Stones', *Br. J. Urol.*, **77**(4), 506–511 (1996).
133. F. Bonte, A. Saunois, P. Pinguet, A. Meybeck, 'Existence of a Lipid Gradient in the Upper Stratum Corneum and its Possible Biological Significance', *Arch. Dermatol. Res.*, **289**(2), 78–82 (1997).
134. T. Taki, S. Handa, D. Ishikawa, 'Blotting of Glycolipids and Phospholipids from a High-performance Thin-layer Chromatogram to a Polyvinylidene Difluoride Membrane', *Anal. Biochem.*, **221**(2), 312–316 (1994).
135. T. Takao, T. Kasama, S. Handa, D. Ishikawa, 'A Simple Quantitative Purification of Glycosphingolipids and

Phospholipids by Thin-layer Chromatography Blotting', *Anal. Biochem.*, **223**, 232–238 (1994).
136. T. Taki, D. Ishikawa, 'TLC Blotting Application to Microscale Analysis of Lipids and as a New Approach to Lipid Protein Interaction', *Anal. Biochem.*, **251**(2), 135–143 .
137. H. Dreyfus, B. Guerold, L. Freysz, D. Hicks, 'Successive Isolation and Separation of Major Lipid Fractions Including Gangliosides from Single Biological Samples', *Anal Biochem.*, **249**, 67–78 (1997).
138. H. Miwa, M. Yamamoto, T. Fatala, K. Kan, T. Asano, 'Thin-layer Chromatography and High-performance Liquid Chromatography for the Assay of Fatty Acid Compositions of Individual Phospholipids in Platelets from Non-insulin-dependent Diabetes Mellitus Patients: Effect of Eicosapentaenoic Acid Ethyl Ester Administration', *J. Chromatogr., B*, **677**(2), 217–223 (1996).
139. P. Narayan, R. Dahiya, 'Alteration in Sphingomyelin and Fatty Acids in Human Benign Prostatic Hyperplasia and Prostatic Cancer', *Biomed. Biochem. Acta*, **50**, 1099–1108 (1991).
140. R.L. Schanaar, L.K. Needhan, 'Thin-layer Chromatography of Glycosphingolipids', *Methods Enzymol.*, **230**, 371–389 (1994).
141. K. Watarabe, M. Nishiyama, 'A New Type of Two Dimensional Thin-layer Chromatography (TLC Mapping) for Analysis of Acidic Glycosphingolipid Molecular Species', *Anal. Biochem.*, **227**(1), 195–200 (1995).
142. H. Dreyfus, B. Guerolds, L. Ereysz, D. Hicks, 'Successive Isolation and Separation of the Major Lipid Fractions Including Gangliosides from Single Biological Samples', *Anal. Biochem.*, **249**(1), 67–78 (1997).
143. J.I. Ruiz, B. Ocha, 'Quantification in the Subnanomolar Range of Phospholipids and Neutral Lipids by Monodimensional Thin-layer Chromatography and Image Analysis', *J. Lipid. Res.*, **38**(7), 1482–1489 (1997).
144. J. Muthing, M. Radloff, 'Nanogram Detection of Phospholipids on Thin-layer Chromatograms', *Anal. Biochem.*, **257**(1), 67–70 (1998).
145. T. White, S. Bursten, D. Federighi, R.A. Lewis, E. Nudelman, 'High-resolution Separation and Quantification of Neutral Lipid and Phospholipid Species in Mammalian Cells and Sera by Multi-one-dimensional Thin-layer Chromatography', *Anal. Biochem.*, **258**(1), 109–117 (1998).
146. M. Kondo, 'Urinary Porphyrin Screening Tests', *Nippon Rinsho*, **53**(6), 1377–1382 (1995).
147. M. Kondo, 'Methods of Determination of Porphyrins and Their Precursors – Introduction of Analytical Methods for Porphyrins Metabolites', *Nippon Rinsho*, **53**(6), 1357–1363 (1995).
148. C.K. Lim, F.M. Ji, T.J. Peters, 'High-performance Liquid Chromatography of Porphyrins', *J. Chromatogr.*, **429**, 123–153 (1988).
149. S.W. Kennedy, A.L. Maslen, 'Separation of Porphyrin Isomers by High-performance Liquid Chromatography', *J. Chromatogr.*, **493**(1), 53–62 (1989).
150. F.M. Ji, C.K. Lim, T.J. Peters, 'Preparation, High-performance Liquid Chromatographic Separation and Characterization of Carboxylic Porphyrins', *J. Chromatogr.*, **461**, 353–359 (1989).
151. A.M. Muniesa, C. Herrero, M. Lecha, J.M. Mascaro, 'Usefulness of Chromatographic Analysis of Faecal Porphyrins in the Diagnosis of Porphyrias', *J. Dermatol.*, **17**(4), 199–204 (1990).
152. J. Luo, J.M. Lamb, C.K. Lim, 'Analysis of Urinary and Faecal Porphyrin Excretion Pattern in Human Porphyrins by Fast Atom Bombardment Mass Spectrometry', *J. Pharm. Biomed. Anal.*, **15**(9–10), 1289–1294 (1997).
153. A. Junker-Buccheit, H. Jork, 'Urinary Porphyrins: Ion Pair Chromatography and Fluorimetric Determination', *J. Planar Chromatogr.*, **1**, 214 (1988).
154. C.W. Huie, W.R. Williams, 'Laser Fluorometric Detection of Porphyrin Methyl Esters by High-performance Thin-layer Chromatography', *Anal. Chem.*, **61**(20), 2288–2292 (1989).
155. K. Jacob, M.O. Das, 'Excretion Pattern of Faecal Coproporphyrin Isomers I-IV in Human Porphyrias', *Eur. J. Clin. Chem. Clin. Biochem.*, **33**(12), 893–901 (1995).
156. C.K. Lai, C.W. Lam, Y.W. Chan, 'High-performance Thin-layer Chromatography of Free Porphyrins for Diagnosis of Porphyria', *Clin. Chem.*, **40**(11 Pt1), 2026–2029 (1994).
157. C.W. Lam., C.K. Lai, Y.W. Chan, 'Simultaneous Fluorescence Detection of Faecal Urobilins and Porphyrins by Reversed-phase High-performance Thin-layer Chromatography', *Clin. Chem.*, **44**(2), 345–346 (1998).
158. M.J. Handerson, 'Thin-layer Chromatography of Free Porphyrins for Diagnosis of Porphyria', *Clin. Chem.*, **35**(6), 1043–1044 (1989).
159. F. Lellouche, A. Frodin, G. Fitzgerald, J. Meclouf, 'Enzyme Immunoassay Measurement of the Urinary Metabolites of Thromboxane A2 and Prostacyclin', *Prostaglandins*, **40**(3), 297–310 (1990).
160. J.O. Morrow, T.A. Minton, 'Improved Assay for the Quantification of 11-Dehydrothromboxane B2 by Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **612**(2), 179–185 (1993).
161. J.A. Awad, J.D. Morrow, L.J. Roberts, 'Simplification of The Mass Spectrometric Assay for the Major Urinary Metabolite of Prostaglandin D2', *J. Chromatogr.*, **617**(1), 124–128 (1993).
162. A. Ferretti, V.P. Flanagam, 'Assay of Urinary 2,3-Dinor-6-Oxo Prostaglandin F1 Alpha by Gas Chromatography/Tandem Mass Spectrometry', *J. Chromatogr.*, **622**(2), 109–115.
163. H. Schweer, B. Watzer, H.W. Seybeth, 'Determination of Seven Prostanoids in 1 ml of Urine by Gas Chromatography/Negative Ion Chemical Ionization Triple

Stage Quadrupole Mass Spectrometry', *J. Chromatogr.*, **652**(2), 221–227 (1994).
164. H. Schweer, B. Watzer, H.W. Seyberth, R.M. Nusing, 'Improved Quantification of 8-Epi-Prostaglandin F2 Alpha and F2-Isoprostanes by Gas Chromatography/Triple-stage Quadrupole Mass Spectrometry: Partial Cyclooxygenase-dependent Formation of 8-Epi-prostaglandin F2 Alpha in Humans', *J. Mass Spectrom.*, **33**(12), 1362–1370 (1997).
165. A. Ferretti, V.P. Flanagan, 'Isolation and Measurement of Urinary 8-Iso-Prostaglandin F2 Alpha by High Performance Liquid Chromatography and Gas Chromatography/Mass Spectrometry', *J. Chromatogr., B*, **94**(2), 271–276 (1997).
166. J.M. Rosenfeld, M. Mureika-Russell, M. Love, 'Solid Phase Sample Preparation Method for Prostaglandins: Integration of Procedures for Isolation and Derivatization for Gas Chromatographic Separation', *J. Chromatogr.*, **489**(2), 263–272 (1989).
167. J.G.A.M. Raajmakers, 'System for the Identification of Novel Prostaglandins', *J. Chromatogr.*, **138**, 355–372 (1977).
168. K. Tsunomoto, S. Todo, S. Imashuku, 'Separation of Prostaglandins and Thromboxane by Two-dimensional Thin-layer Chromatography', *J. Chromatogr.*, **417**, 414–419 (1987).
169. M.C. Alvarez Perez, C. Miyares Cao, M.C. Averholf Casamayor, R. Delgado Fernandez, 'Prostaglandin E2 (PGE2) in Normal and Inflamed Gingina Clinical Assay', *Rev. Cubana. Estomatol.*, **26**(3), 219–227 (1989).
170. S.K. Goswami, J.E. Kinsella, 'A Non-destructive Spray Reagent for The Detection of Prostaglandins and Other Lipids in Thin-layer Chromatography', *Lipids.*, **16**(10), 759–770 (1981).
171. A. Rintta, E. Seppala, H. Vapaatalo, 'A Rapid Sensitive Method for Detecting Different Arachidonic Acid Metabolites by Thin-layer Chromatography: The Use of Autoradiography', *J. Chromatogr.*, **307**, 185–189 (1984).
172. D. Mattingly, H. Martin, C.M. Tyler, 'Estimation of Urinary Aldosterone Using Thin-layer Chromatography and Fluorometry', *J. Clin. Pathol.*, **46**(12), 1109–1112 (1993).
173. I. Romanishin, I. Sterniuk, 'Hormonal Correlations in Stomach Cancer Patients', *Lik Sprava*, **1**, 72–75 (1992).
174. S. Ivanovic, D. Zivanov-Stakic, I. Konstantinovic, 'The Urinary Dehydroepiandosterone, Andosterone and Eticholanolone Excretion of Healthy Women and Women with Benign and Malignant Heart Disease', *J. Clin. Pharm. Ther.*, **15**(3), 213–219 (1990).
175. P.W. Tang, D.L. Crone, 'A New Method for Hydrolyzing Sulfate and Glucuronyl Conjugates of Steroids', *Anal. Biochem.*, **182**(2), 289–294 (1989).
176. E. Venturelli, A. Cavalleri, G. Secreto, 'Methods for Urinary Testosterone Analysis', *J. Chromatogr., B*, **671**(1–2), 363–380 (1995).
177. J.C. Touchstone, *CRC Handbook of Chromatography: Steroids*, eds. G. Zweig, J. Sherma, CRC Press, Boca Raton, FL, 1986.
178. H. Lamparczyk, *Handbook of Chromatography, Analysis and Characterization of Steroids*, ed. J. Sherma, CRC Press, Boca Raton, FL, 1992.
179. H. Lamparczyk, 'Chromatographic Techniques Used in Steroid Analysis', in *Handbook of Chromatography: Analysis and Characterization of Steroids*, ed. J. Sherma, CRC Press, Boca Raton, FL, 9–18, 1992.
180. L. Vanlook, Ph. Deschuytere, C. Van Peteghem, 'Thin-layer Chromatographic Method for the Detection of Anabolics in Fatty Tissues', *J. Chromatogr.*, **489**, 213–218 (1989).
181. J.W. Brown, A. Carballeria, L.M. Fishman, 'Rapid Two-dimensional Thin-layer Chromatographic System for the Separation of Multiple Steroids of Secretory and Neuroendocrine Interest', *J. Chromatogr.*, **439**, 441–447 (1988).
182. Z. Tomosova, 'Thin-layer Chromatography of Urinary 17-Oxosteroids Using Dansylhydrazine as a Prelabeling Reagent', *J. Chromatogr.*, **570**(2), 396–398 (1991).
183. P. Vingler, M. Kermici, P. Krien, 'Direct Quantitative Digital Autoradiography–Thin-layer Chromatography of 3 Alpha, 3 Beta- and 5 Alpha-Reduced and 17 Beta-Hydrogenated Androgens Derived From Testosterone Metabolism', *J. Chromatogr.*, **571**(1–2), 73–86 (1991).
184. M. Fenske, H. Schonheiter, 'Thin-layer Chromatography on Silica Coated Aluminium Sheet as an Adjunct to Radioimmunoassay of Steroids', *J. Chromatogr.*, **563**(1), 178–183 (1991).
185. A.K. Agarwal, N.A. Pampori, B.M. Shapiro, 'Thin-layer Chromatographic Separation of Regioselective and Stereospecific Androgen Metabolites', *Anal Biochem.*, **224**(1), 458–457 (1995).
186. T.R. Watkins, A. Smith, J.C. Touchstone, 'Reversed-phase Thin-layer Chromatography of Estrogens Glucuronides', *J. Chromatogr.*, **371**, 221–222 (1986).
187. H. Lamparczyk, R.J. Ochacka, P. Zarzyck, J.P. Zielinsk, 'Separation of Steroids by Reversed-phase HPTLC Using Various Binary Mobile Phases', *J. Planar Chromatogr.*, **3**, 34–37 (1990).
188. J. Muthing, 'High-resolution Thin-layer Chromatography of Gangliosides', *J. Chromatogr.*, **720**(1–2), 3–25 (1996).
189. M. Yoshioka, M. Abu-Zeid, T. Kubo, M. el-Merzabani, 'Identification of a Previously Unknown Compound as 2′- Deoxycytidine Found in the Plasma of Breast Cancer Patients Under Combined Chemotherapy', *Biol. Pharm. Bull.*, **17**(2), 169–172 (1994).
190. R. Kumar, K. Hemminki, 'Separation of 7-Methyl and 7-(2-Hydroxyethyl)-Guanine Adducts in Human DNA Samples Using a Combination of TLC and HPLC', *Carcinogenesis*, **17**(3), 485–492 (1996).
191. M. Koskinen, H. Rajaniemi, K. Hemminki, 'Analysis of Tamoxifen-induced DNA Adducts by ^{32}P- Postlabeling

Assay Using Different Chromatographic Techniques', *J. Chromatogr., B*, **691**(1), 155–160 (1997).

192. J.J. Steinberg, G.W. Oliver, Jr, N. Farah, P. Simoni, R. Winiarsky, A. Cajigas, 'In Vivo Determination of 5-Bromo-2-Deoxyuridine Incorporation into DNA Tumor Tissue by a New ^{32}P-Post Labeling Thin-layer Chromatographic Method', *J. Chromatogr., B*, **694**(2), 333–341 (1997).

193. G. Allmany, A. Gamundi, C. Rossello, R. Rial, 'Simultaneous Determination of Vanillyl Mandelic Homovanillic Acids and 5-Hydroxyindoleacetic Acids in Human Urine by Thin-layer Chromatography', *Biomed. Chromatogr.*, **10**(3), 144–145 (1996).

194. E. Marklova, M. Bratova, V. Vorisek, 'Screening for Organic Acid Disorders', *Acta Medica (Hardec Kralove)*, **40**(1), 17–19 (1997).

Point-of-care Testing: Principles, Practice, and Critical-emergency-disaster Medicine

Gerald J. Kost, Nam K. Tran and Richard F. Louie
University of California, Davis, CA, USA

1 Introduction **2**
1.1 Definition of Point-of-care Testing and the Hybrid Laboratory 2
1.2 Goal of Point-of-care Testing and Objectives of the Article 2
2 Principles and Analytical Methods for Point-of-care Testing **2**
2.1 Analytical Methods for Chemistry Analytes 2
2.2 Analytical Methods for Hematology and Hemostasis Analytes 13
2.3 Cardiovascular and Cardiac Injury Markers 13
2.4 Immunoassays 14
2.5 Point-of-care Molecular Diagnostics 19
2.6 Criteria for Instrument Evaluation and Selection 20
3 Critical-emergency-disaster Care **24**
4 Test Clusters (Critical Care Profile) and Medical Indications **24**
5 Accuracy Assessment **27**
5.1 Pattern-recognition Methods 27
5.2 Locally Smoothed Median Absolute Difference Curves 27
5.3 Discrepant Values 28
5.4 Statistical Methods 28
5.5 Future Concerns 29
6 Monitoring Performance **30**
6.1 Total Quality Principle for Point-of-care Testing 30
6.2 Accreditation and Regulation 30
6.3 Performance Enhancement and Continuous Quality Improvement 32
7 Strategies for Optimization of Point-of-care Testing **33**
7.1 Point-of-care Connectivity 34
8 Future Trends **34**
8.1 New Practice Standard for Timeliness 34
8.2 Critical Speed, Test Clusters, and Medical Necessity 34
8.3 Molecular Diagnostics 35
8.4 Recommendations for Point-of-care Glucose Testing 35
8.5 The National Institutes of Health Point of Care Research Network 35
8.6 Connectivity for Medical Response to Disasters 36
8.7 Knowledge Optimization 36
8.8 Enhanced Performance Despite Economic Constraints 36
8.9 Outcomes Optimization and Future Challenges 37
9 Conclusions **37**
Acknowledgments **38**
Abbreviations and Acronyms **38**
Related Articles **39**
References **40**
Further Reading **44**
Appendix **44**

Point-of-care testing (POCT) is defined as testing at or near the site of patient care, wherever that medical care is needed. The goal of POCT is to facilitate immediate evidence-based medical decisions that improve patient outcomes and reduce patient acuity, criticality, morbidity, and mortality, especially during life-threatening crises and emergency resuscitations. POCT is efficient and efficacious because it facilitates rapid diagnosis, faster treatment decisions, high-quality trend monitoring, and patient and physician satisfaction. New point-of-care (POC) technologies are appearing rapidly. This article summarizes principles and analytical methods of whole-blood analysis (WBA) for chemistry, hematology, and hemostasis analytes, cardiac injury markers, and other POC diagnostic tests. Then, the article discusses important clinical aspects, including test clusters and medical indications, methods of monitoring and enhancing performance, and optimization strategies. In the next two decades, diagnostic testing will continue to shift to the POC, necessitating collaborative integration, information consolidation, and Knowledge Optimization® in surgery, intensive care, emergency medicine, public health, and other settings extending from the acute care center to the patient's home.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Definition of Point-of-care Testing and the Hybrid Laboratory

POCT is defined as *diagnostic testing at or near the site of patient care, wherever that medical care is needed*.[1] This broad definition, which includes near-patient testing, derives from the historical roots of the paradigm shift.[2] POCT is performed outside the main laboratory within the hospital, in clinics, and at various other distributed sites (Table 1). POCT facilitates rapid diagnosis, faster treatment decisions, high-quality trend monitoring, and patient and physician satisfaction. The clinical integration of POC analyzers at the site of patient care created the *hybrid laboratory*.[3,4] POCT, customized test clusters, minimized therapeutic turnaround time (TTAT), optimized diagnostic–therapeutic strategies, and, especially, *an emphasis on improved patient outcomes* are the hallmarks of the hybrid laboratory (Table 2).

1.2 Goal of Point-of-care Testing and Objectives of the Article

The goal of POCT is *to facilitate immediate evidence-based medical decisions that improve patient outcomes and reduce patient acuity, criticality, morbidity, and mortality, especially during life-threatening crises and emergency resuscitations*. The objectives of this article are (i) to review the principles of WBA and other POC methods, which enabled the paradigm shift to POCT; (ii) to understand POC test clusters (test menus); (iii) to describe new mathematical–statistical approaches to monitor the performance of POCT; and (iv) to present strategies that optimize POCT in critical care and other settings, including disasters.

Table 1 Sites of POCT

- Hospital
  - Bedside (alternate/alternative, ancillary, decentralized, and waived testing)
  - Critical care workstations in the operating room, emergency department, and intensive care unit
  - Hybrid laboratory instrument sites
  - Modular plug-ins for intensive care systems
  - Near-patient testing locations (permanent, temporary, and mobile)
  - Patient-focused care centers
  - Patient monitoring stations interfaced with *ex vivo* and *in vivo* systems
  - Procedure suites (e.g. cardiac catheterization laboratories)
  - Satellite laboratories
- Outpatient
  - Clinics and specialty care centers
  - Home (e.g. self-monitoring of blood glucose)
  - Patient-focused care facilities (e.g. ambulatory surgery, heart center, and chronic care)
  - Physician offices
  - Urgent care centers
  - Wellness testing areas (e.g. screening)
- Rescue and unique locations
  - Emergency ground vehicles (e.g. ambulances)
  - Helicopter and fixed-wing aircraft for patient transport
  - Military field operations and battlefields
  - Patient travel including remote and inaccessible locations
  - Points of disasters and emergency rescues
  - Ships, submarines, and other nauticals
  - Space shuttles and space station

Table 2 The hybrid laboratory – emergence and definition

Conventional laboratory (1980s)	Hybrid laboratory (1990s)	Personalized laboratory (21st century)
Centralized	Patient focused	Mobile
Separate samples	Critical care profiles (test clusters)	Quantitative diagnostic and therapeutic indices
Confusion	Whole-blood priority	Alternative sample types (e.g. saliva, interstitial fluid)
Slow, high-volume panels	Fewer, more important tests faster	Multiplex sample testing, and multimarker tracking
Data overload	Diagnostic synthesis	Wireless network
Laboratory turnaround time	TTAT	Optimized decision cycles
Cost analysis by function	Cost effectiveness from patient outcomes	Efficient global healthcare
Highly vulnerable during disasters	Reduced risk of failure	Small-world networks prepared for newdemics

2 PRINCIPLES AND ANALYTICAL METHODS FOR POINT-OF-CARE TESTING

2.1 Analytical Methods for Chemistry Analytes

2.1.1 Overview

Table 3 outlines common measurements and principles currently used *in vitro* for WBA. Biosensors found in whole-blood analyzers include the ion-selective electrode (ISE), substrate-specific electrode (SSE), analyte-specific optical sensor (ASOS), and electrical conductance sensor (ECS). The acronym ISE refers to measurement of cations (e.g. K^+) as well as anions (e.g. Cl^-). The term SSE

Table 3 Principles of biosensor-based whole-blood measurements

Analyte	Principle(s)
Ca^{2+}	ISE
Mg^{2+}	ISE
K^+	ISE
Cl^-	ISE
Na^+	ISE
pH	ISE or ASOS
pCO_2	CO_2 sensitive buffer pH electrode or ASOS
Total CO_2	Acidification of sample with pH end point by ISE
pO_2	Amperometric or ASOS
Glucose	SSE
Lactate	SSE
Urea nitrogen	SSE
Creatinine	SSE
Hematocrit	ECS

ASOS, analyte-specific optical sensor; ECS, electrical conductance sensor; ISE, ion-selective electrode; SSE, substrate-specific electrode.

was introduced in 1991[5] and refers to the measurement of glucose, lactate, urea nitrogen, creatinine, and other metabolites. ASOS was introduced recently[6] to describe clinical optical sensors. Most POCT instruments also include hematocrit measurement performed with an ECS or hemoglobin measurement by CO-oximetry or some other method. Some instruments include fiber optics for multiple wavelength measurement of O_2 saturation. *In vitro* instruments use sensors to measure several analytes simultaneously and quickly. *Ex vivo* and *in vivo* blood chemistry monitoring systems generally are limited to relatively few analytes, primarily those that vary rapidly and require fast response or frequent measurements. These analytes include oxygen saturation, glucose, blood gases (partial pressure of oxygen, pO_2; partial pressure of carbon dioxide, pCO_2), pH, potassium, sodium, ionized calcium, hematocrit, and, to a lesser extent, urea, creatinine, and lactate. Most effort has been devoted to the development of *ex vivo* and *in vivo* systems that monitor these analytes in the hospital environment, particularly in the operating room and intensive care unit.[7,8] Glucose also is the focus of ongoing attempts to develop a wearable or implantable monitoring system for outpatient use by diabetics. The next sections briefly discuss the measurement principles of electrochemical, fiber optic, and relatively new fiber-optic chemical sensors. Miniaturized electrochemical and fiber-optic sensors are found in POCT devices and are used in intermittent and continuous monitoring systems as well.

2.1.2 Sensors

Electrochemical sensors fall into two categories: potentiometric and amperometric. Most potentiometric sensors use an ion-selective membrane that incorporates a specific ionophore into the membrane matrix. The membrane functions as the biodetector of the electrode. When in contact with blood or other body fluids, the ionophore selectively interacts with a target ion species and creates an electrical potential across the membrane. This potential is measured between the internal reference electrode and the external reference electrode. The potential is logarithmically proportional to the activity of the ion that is being measured in the sample. In some systems, measurements must be corrected for interference (e.g. correction of Mg^{2+} for Na^+, Ca^{2+}, and pH in the sample). Whole-blood analyzers commonly use ISEs for the measurement of electrolytes and pH *in vitro*. A modification of the basic potentiometric sensor, the Stowe–Severinghaus electrode, is used to measure pCO_2. In this electrode, CO_2 from the sample being measured diffuses across a CO_2-permeable membrane into a bicarbonate buffer that is in contact with a pH sensor. When the CO_2 reaches equilibrium in the bicarbonate buffer, the pCO_2 of the sample can be calculated on the basis of the pH measured in the buffer. A related principle is used to quantitate total carbon dioxide (TCO_2) content in whole blood after automated addition of acid to displace bound CO_2.[9]

The response of potentiometric sensors is logarithmic. Therefore, small errors in the measured potential produce large errors in the results. These errors will affect both accuracy and precision. If there is a change in activity from a_1 to a_2, there will be a change in the sensor potential. If $\Delta E = E_2 - E_1$, it follows from the Nernst equation that $\Delta E = [(RT \ln 10)/(ZF)]\log[\text{ratio of the activities}]$, where R is the gas constant ($8.31431\,J\,K^{-1}\,mol^{-1}$), T is the absolute temperature (K, kelvin), ln 10 is the natural log of 10 (2.303), F is the Faraday constant ($96\,487\,C\,mol^{-1}$), and Z is the valence (charge) of the ion. When ΔE is 1 mV, T is 310.15 K (37 °C), and Z is 1, the ratio of the activities will be equal to $\log^{-1}[(\Delta E)/((RT \ln 10)/(ZF))] \approx \log^{-1}[1\,mV/61.54\,mV] = 1.038$. Hence, a ±1 mV change in the measured potential is equivalent to approximately a ±4% change in the reported value. The reference electrode alone can introduce uncertainty of up to ±0.5 mV from junction potential, sample matrix, and boundary layer phenomena where the electrolyte enters the sample flow path. This is equivalent to a change in the accuracy of approximately ±2% for monovalent ions (i.e. $Z = 1$ and $\log^{-1}[0.5\,mV/61.54\,mV] \approx 1.019$) or approximately ±4% for divalent ions (i.e. $Z = 2$ and $\log^{-1}[0.5\,mV/30.77\,mV] \approx 1.038$). Instabilities during the electronic measurement of the potential of the sensor

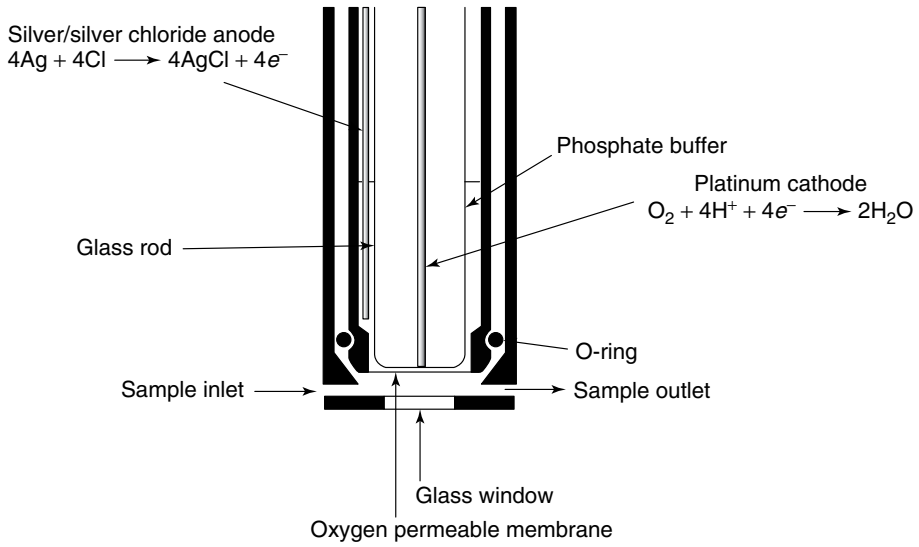

Figure 1 Cross-sectional schematic of a Clarke-type electrode. This figure illustrates a typical Clark electrode for measuring pO_2. The anode generates the electrons to drive the reaction at the cathode where oxygen is converted into water.

electrochemical cell (ISE :: reference electrode) and electrical noise coming from sensor modules or *in vivo* probes detract from the precision of the analyte readings. Inaccuracy and imprecision combine to produce time-variant changes in calibration. As a consequence, *ex vivo* and *in vivo* monitors must either be designed for very high stability or else incorporate periodic recalibrations to correct for this drift.

Amperometric sensors generally are based on the Clark-type oxygen electrode (Figure 1).[10] A barrier membrane placed over the sensing electrode allows oxygen from the sample to diffuse to the electrode, but excludes chemical species that potentially could interfere with electrode function. A voltage applied across the sensing and reference electrodes reduces oxygen and, at steady state, generates an electrical current that is linearly proportional to the pO_2 in the sample. A common enzyme-based amperometric sensor is the glucose SSE. Immobilized glucose oxidase (GO) oxidizes the glucose substrate. Several different approaches have been used to quantitate this oxidation reaction and to determine substrate concentration. These include oxygen consumption, which requires an additional reference oxygen sensor, and hydrogen peroxide generation, which does not. Measurement of the latter is more practical, in view of sensor space constraints. There are many different glucose SSE designs. Some use an electron shuttle and glucose dehydrogenase (GD)[11] to reduce dependence on O_2. Barrier membranes of various configurations are used to exclude interfering species, select for the target substrate, and control the rate of diffusion of the substrate into the sensor. For amperometric sensors, in general, slight changes in membrane permeability or in the electrodes over time can cause drift in the output current. Therefore, for monitoring applications, SSEs also must be designed for high stability or be recalibrated periodically.

Optical sensors differentiate oxygenated blood from deoxygenated blood by spectrophotometry. A narrow waveband light source such as a light-emitting diode (LED) generates light that is transmitted down optic fibers to the tip of a catheter or finger sensor (e.g. pulse oximeter). Light reflects from the blood back through one or more optic fibers to a photodetector. Several wavelengths are chosen such that the differential absorption characteristics of oxyhemoglobin and hemoglobin allow determination of the ratio between the two. Fiber-optic chemical sensors, also referred to as *optodes*, have received considerable attention as a technology that has potential advantages over more conventional electrochemical sensors for automated monitoring. Optodes can be made very small, require no direct electrical connection to the patient, are free from electromagnetic interference, and can incorporate internal referencing to reduce or eliminate drift. On the other hand, commercial experience to date has shown that they can be difficult and costly to manufacture and have low production yields. Four basic approaches to fiber-optic chemical sensors have been described, including absorption, fluorescence intensity, fluorescence lifetime, and surface-enhanced Raman spectroscopy (SERS). Commercial systems are based primarily on either absorption or fluorescence intensity. With both approaches, light is passed through an optic fiber to the sensor that contains an indicator. In absorption sensors, the target analyte is measured by the intensity of the color of the indicator, which changes in

proportion to the analyte concentration. In fluorescence intensity sensors, it is the intensity of a lower wavelength light reemitted by an indicator that changes in proportion to the analyte concentration. Most developmental work with optodes has focused on pO_2, pCO_2, and pH, although fiber-optic sensors for glucose and electrolytes have also been described.

2.1.3 *Transportable, Portable, and Handheld Instrument Formats*

Table 4[12–21] lists whole-blood analyzers and biosensor-based tests that they perform. Biosensors allow simultaneous direct measurement of several indicators of vital functions in one whole-blood sample within 1–2 min. Competition is stimulating companies to expand test menus and to extend linear ranges of the biosensors to better cover critical analyte levels. Transportable instruments are larger whole-blood analyzers that can be placed on carts with uninterruptable (battery) power supplies for short-term operation. These instruments have several advantages, such as the selection of user-defined test clusters (critical care profiles) from large menus of tests, automation with few manual sample processing steps, selectable minimum sample volumes, fast analysis cycle, high throughput, low operating cost per test, long-lasting reagent supply, and self-contained biohazard waste storage. Portable instruments vary in size from dimensions equivalent to personal computers to dimensions nearly comparable to handheld devices. These formats display a wide variety of design features including fiber-optic measurement and multiple- or single-use (unit use) cartridges. Cartridges have self-contained calibrants and reagents for testing. In some cases, for blood gas measurements, solutions in the cartridges are equilibrated with appropriate gas tensions for the calibration of pO_2 and pCO_2. Although a means of temperature control (37 °C) has to be built in, no external gas tanks are required. Hence, bulk and weight are reduced, and mobility is increased significantly. Portables usually have convenient handles for carrying and occupy modest footprints. Other advantages include biohazard containment and battery operation, with the option for benchtop use with conventional power. Test menus generally are limited and operating costs usually are higher compared to transportables. Handheld devices have the advantages of interchangeable test cartridges, convenience, small size, light weight, compact storage, and battery operation. A handheld device can be placed in one hand, while the other hand is free to insert a cartridge or to apply a blood sample with a syringe. Miniaturization of the test cartridge is the result of microfabrication. Chip-based technologies eventually will lead to large test menus in handhelds. Global use of small handhelds will reduce costs.

2.1.4 *Handheld Glucose Monitoring*

The most common handheld bedside device is the glucose meter. Most glucose monitoring systems (GMS) utilize either (i) electrochemistry (amperometry or coulometry), or (ii) photometry. Both methods depend on enzymatic reactions to generate a signal, which correlates to the sample glucose concentration. Table 5 lists representative chemical principles of glucose test strips used in the handheld meters found in Table 4. Most glucose meters today rely on electrochemical methods to measure glucose (Figure 2). Glucose meters typically use a whole-blood sample (arterial, venous, or capillary). Transdermal glucose meters (i.e. GlucoWatch) have been developed to sample interstitial fluid through the skin via reverse iontophoresis.[10] However, these meters have been discontinued by the manufacturer. More recently, continuous glucose monitors have been introduced. Both devices require confirmation of accuracy with a conventional glucose meter. Although promising, both the noninvasive transdermal and continuous glucose monitoring systems are not recommended for routine use in the clinical environment due to limited data from clinical studies.[22,23]

While bedside and home glucose testing represents the standard of care for diabetics, glucose meters must be used with caution in critical care because of potential random and systematic errors. This is a major concern due to the increased use of handheld GMSs for tight glycemic control (TGC).[24,25] TGC protocols require hourly or continuous monitoring to adjust insulin infusion regimens. Figure 3 shows TGC ranges from TGC protocols reported in literature or current use by hospitals.[24–34]

Glucose results must be interpreted in light of regional metabolism, blood supply, and specimen source (if venous or capillary), as well as the patient's clinical status. Additional clinical research and validation will be necessary to identify the relative importance of intrinsic and extrinsic interferences (e.g. confounding factors) and determine if results obtained with glucose meters are accurate enough to successfully guide insulin therapy in critical care patients.

These interferences or "confounding factors" are grave concerns in critical care glucose monitoring. As stated, they are either intrinsic or extrinsic in nature (Figure 2). Intrinsic factors include specific physiological elements, such as hematocrit, blood pO_2, blood pressure, sample source (venous, arterial, or capillary), pH, temperature, and the blood matrix (water, lipid, cellular, and protein contents).[35] Extrinsic factors include drug interferences, user error, and harsh environmental conditions.

Table 4 POCT for mobile, emergency, critical, and primary care

Manufacturer	Instrument (format)	Test(s)
Abaxis www.abaxis.com	Piccolo POC Blood Analyzer (portable)	Albumin, ALP, ALT, amylase, AST, BUN, creat, Ca, Cl^-, K^+, Na^+, Mg, HDL, cholesterol, triglycerides, CK, glucose, TP, LD, direct bilirubin, total bilirubin, phosphorous, TCO_2, GGT, uric acid
Abbott www.abbott.com	Precision PCx (handheld)	Glucose (amperometric)
	FreeStyle Connect (handheld)	Glucose (coulometry)
	Medisense Optium (handheld)	Glucose (amperometric), ketones[17]
	Testpack Plus hCG (handheld, disposable)	hCG (pregnancy test)
Acon Labs www.aconlabs.com	One-Step Myo/CK-MB/TnI Combo (handheld)[a]	Myoglobin, CK-MB, cTnI
	CK-MB One-Step (handheld)[a]	CK-MB
	MYO One-Step Myoglobin Test (handheld)[a]	Myoglobin
	cTnI Troponin I Rapid Test (handheld)[a]	cTnI
	On Call Now Blood Glucose Meter (handheld)	Glucose
	Tumor Marker Test (handheld, disposable)[a]	PSA, α-fetoprotein
Accumetrics www.accumetrics.com	VerifyNow P2Y12 Assay (portable)	Platelet function, aspirin assay
Adeza Biomedical www.adeza.com	TLi Analyzer (transportable)	Fetal fibronectin
American Diagnostica www.americandiagnostica.com	CLEARVIEW Simplify (disposable)	D-dimer
Animas Technologies www.glucowatch.com	Glucowatch G2 (wristwatch)	Glucose (iontophoresis)
APEL www.apel-jp.com (Japan)	BR-501 (transportable)	Neonatal bilirubin
	BR-5000N Digital (transportable)	Neonatal bilirubin
	HG-202 (transportable)	Hb
Axis-Shield www.axis-shield.co.uk	NycoCard Reader II (portable)	HbA_{1c}, microalbumin, CRP, D-dimer
	Afinion Analyzer System (portable)	HbA_{1c}, CRP
Bayer Diagnostics www.bayerdiag.com	RapidLab 348 (transportable)	pH, pCO_2, pO_2, Hct, Na^+, K^+, Ca^{++}, Cl^-
	RapidLab 1200 (transportable)	pH, pCO_2, pO_2, Hct, tHb, HHb, FO_2Hb, sO_2, COHb, MetHb, Na^+, K^+, Ca^{++}, Cl^-
	DCA 2000+ (transportable)[12]	HbA_{1c}, microalbumin
Biosafe International www.ebiosafe.com	Anemia Meter (handheld, disposable)	Hb
	Thyroid Test (handheld, disposable)	TSH
	Prostate Test (handheld, disposable)	PSA
	Diabetes (A_{1c}) Test (handheld, disposable)	HbA^{1c}
	Cholesterol Panel (handheld, disposable)	HDL
Biosite Diagnostics www.biosite.com	Triage Cardiac Panel (portable)	cTnI, CK-MB mass, D-dimer, myoglobin, BNP
	Triage BNP Test (portable)	BNP[b]
	Triage Shortness of Breath Panel (portable)	cTnI, CK-MB mass, D-dimer, myoglobin, BNP
	Triage Tox Screen (portable)	Amphetamines, methamphetamines, cocaine, opiates, phencyclidine, tetrahydrocannabinol, barbiturates, propoxyphene, tricyclic antidepressants
Cholestech Corp. www.cholestech.com	Cholestech LDX (portable)	Cholesterol, HDL, triglycerides, ALT, glucose

Table 4 (*continued*)

Manufacturer	Instrument (format)	Test(s)
Craig Medical www.craigmedical.com	Aimstep PBD (handheld, disposable)	hCG (pregnancy test)
	Aimstep PBC (handheld, disposable)	hCG (pregnancy test)
	Aim MID-OTC (handheld, disposable)	hCG (pregnancy test)
	FertilMARQ (handheld, disposable)	Sperm concentration
Dade Behring www.dadebehring.com	Stratus CS STAT (transportable)	CK-MB, myoglobin, cTnI, D-dimer, β-hCG[13], NT-proBNP[c]
Dainippon Pharmaceuticals www.dainippon-pharm.co.jp (Japan)	RAPIDCHEK (handheld)[b]	H-FABP
Diametrics Medical www.diametrics.com	Paratrend 7 intravascular blood gas monitor (portable, *in vivo*, no longer available)	pH, pO_2, pCO_2, temperature
Draeger Medical www.draeger-medical.com	Konica Minolta/Air-shields JM-103[14] (handheld)	TC bilirubin
Embryotech www.embryotech.com	VasMAQ Diagnostic Screening Test (handheld, disposable)	Sperm concentration
HemoCue AB www.hemocue.com	Hemocue Glucose 201 RT (handheld)	Glucose (approved for screening and diagnosis)
	Hemocue Urine Albumin (handheld)	Urine albumin
	Hemocue HB 201+ (handheld)	Hemoglobin[b]
	Hemocue HB 201 DM (handheld)	Glucose[b] or hemoglobin[b]
	Hemocue Plasma Low/Hb (handheld)	Low Hb concentration
HemoSense www.hemosense.com	INRatio	PT, INR
Horiba www.jp.horiba.com	Palm-LC[a]	WBC
Instrumentation Laboratory www.ilus.com	GEM Premier 3000 (transportable)	pH, pCO_2, pO_2, Na^+, K^+, Ca^{++}, glucose, lactate, Hct
i-STAT www.i-stat.com	i-STAT PCA (handheld)	ACT[d], PT/INR, glucose[e], creat, cTnI, Na^+, K^+, Cl^-, Ca^{++}, pH, pCO_2, pO_2, sO_2, Hct, Hb, BUN, CK-MB[c], BNP[c]
ITC www.itcmed.com	Hemochron Signature Elite (transportable)	PT, aPTT, TT, HNTT, HiTT, fibrinogen, PDAO
	IRMA TruPoint (portable)	pH, pCO_2, pO_2, Na^+, K^+, Cl^-, Ca^{++}, glucose, BUN, hct, creat.
Key Pharmaceuticals www.keypharm.com.au	Pregnosis (handheld, disposable)[a,f]	hCG (pregnancy test)
Khon Kaen University www.kku.ac.th/eng1 (Thai/Eng)	KKU-OF (test kit)[f]	Osmotic fragility[g]
	KKU-DCIP (test kit)[f]	Hemoglobin E[g]
Life Scan www.lifescan.com	OneTouch UltraSmart (handheld)	Glucose (amperometry)
	OneTouch SureStep (handheld)	Glucose (photometry)
	OneTouch Profile (handheld)	Glucose (amperometry)
	InDuo (Handheld)	Glucose monitoring, insulin administration
Masimo www.masimo.com	Rad-57	SpO2, SpCO, SpMet, SpHb (noninvasive hemoglobin)
Matritech www.matritech.com	NMP22 BladderChek (handheld, disposable)[15]	NMP22 protein (bladder cancer marker)
Medtronic www.medtronic.com	ACT Plus Automated Coagulation Timer (portable)	HR-ACT, LR-ACT, RACT
	Minimed CGMS System Gold (handheld)	Glucose (continuous glucose monitoring)
	Minimed Guardian RT (handheld)	Glucose (continuous glucose monitoring)
Metracor Technologies www.viamedical.com	VIA LVM Neonatal Monitor (portable, *ex vivo*)	pH, pCO_2, pO_2, K^+, Na^+, Hct

(*continued overleaf*)

Table 4 (*continued*)

Manufacturer	Instrument (format)	Test(s)
Metrika www.metrika.com	A1cNow (handheld, disposable)[b]	Glycosylated HbA_1C
	LipidNow (handheld, disposable)[c]	Cholesterol, HDL
	MicroalbuminNow (handheld, disposable)[c]	Microalbumin, creatinine
NOVA Biomedical www.novabiomedical.com	Stat Profile Critical Care Express (transportable)	pH, pCO_2, pO_2, Na^+, K^+, Ca^{++}, Cl^-, Mg^{++}, glucose, lac, Hct, Hb, HHb, tHb, BUN, creat, SO_2%, O_2Hb, MetHb, COHb, O_2Cap, O_2Ct
	Stat Profile pHOx (transportable)	pH, pCO_2, pO_2, Na^+, K^+, Ca^{++}, glucose, lac, SO_2%, Hct, HHb, Hb, COHb O_2Cap, O_2Ct, O_2Hb
	StatStrip Glucose Meter (handheld)[c](16)	Glucose, Hct (interference compensation for acetaminophen, ascorbic acid, and uric acid)
Orion Diagnostica www.c-rptest.com	QuikRead CRP (Handheld)	CRP
Osmetech www.osmetech.com	OPTI CCA Touch (transportable)	pH, pO_2, pCO_2, tHb, SO_2, Na^+, K+, Ca^{++}, Cl^-, glucose
Ostex www.ostex.com	Osteomark NTx Point-of-Care (handheld, disposable)	Cross-linked *N*-telopeptides (osteoporosis)
Polymer Technology Systems www.ptspanels.com	BioScanner 2000 (handheld)	Glucose
PTH Testing(17,18)	Roche Elecsys 1010/2010/E170 (bench)	PTH, cardiac biomarkers, other analytes
	QuiCk-Intraoperative Bio-Intact Intact PTH	PTH (transportable)
	QuiCk-Intraoperative Intact PTH	PTH (transportable)
	STAT-10-I-PTH	PTH (transportable)
	Turbo Intact PTH	PTH (transportable)
Radiometer www.radiometer.com	Radiometer ABL 555 (portable)	pH, pCO_2, pO_2, Hct, K^+, Na^+, Ca^{++}, Cl^-, glucose, lac
	Radiometer ABL 700 (transportable)	pH, pCO_2, pO_2, tHb, FO_2Hb, FCOHb, FmetHb, FHHb, FHbF, K^+, Na^+, Ca^{++}, Cl^-, glucose, lac, tBil
	Radiometer ABL 80 Flex (transportable)	pH, pCO_2, pO_2, pH, pCO_2, pO_2, hct
	Radiometer TCM4 (transportable)(19)	TC pO_2 and TC pCO_2
	Radiometer TCM40 (transportable)(19)	SO_2, TC pO_2, and TC pCO_2
	Radiometer TCM400 (portable)(19)	TC pO_2
Respironics bilicheck.dev.respironics.com	BiliCheck (handheld)	TC bilirubin
Response Biomedical www.responsebio.com	RAMP Reader (portable)	Myoglobin, CK-MB, cTnI
Roche Diagnostics www.roche-diagnostics.com	Accu-Chek Advantage (handheld)	Glucose
	Accutrend GCT (handheld)[f,g]	Glucose, cholesterol, triglycerides
	Cardiac Reader (portable)	cTnT, myoglobin, D-dimer, NT-proBNP
	CoaguChek (handheld)	PT
	Cobas h 232	NT-proBNP, cTnT, CK-MB, myoglobin, D-dimer
	Omni S or Omni C (transportable)(20,21)	pH, pCO_2, pO_2, sO_2, Na^+, K^+, Ca^{++}, Cl^-, tHb, CO-oximetry, glucose, lac, Hct, urea, neonatal bilirubin
	Reflotron Plus (portable)	Cholesterol, triglyceride, glucose, HDL, LDL, Hb, bilirubin, alkaline phosphatase, GOT, GPT, GGT, uric acid, creat, urea, K^+, pancreatic amylase

Table 4 (*continued*)

Manufacturer	Instrument (format)	Test(s)
	TropT Sensitive (handheld)	cTnT (qualitative)
Spectral Diagnostics www.spectraldiagnostics.com	Cardiac STATus (handheld, disposable)	cTnI, CK-MB, myoglobin
StanBio www.stanbio.com	Stat-Site M (handheld)[b]	Hb
	Stat-Site Analyzer-Ketone Meter (handheld)[b]	β-Hydroxybutyrate (ketoacidosis)
	Hemopoint H2 (portable)[b]	Hct, Hb
Sysmex www.sysmex.com	pocH-100i Hematology Analyzer (portable)	CBC and differential leukocyte count
Unipath www.unipath.com	Clearplan Fertility Monitor (handheld)[b]	E3G, LH
Wako Diagnostics www.wakousa.com/diagnostics	Apolowako (transportable)[a]	HbA_{1c}, CRP, AST, ALT, GGT, total cholesterol, triglycerides, HDL, creat, BUN, uric acid, total protein, albumin, tBil, glucose, PSA
	Wako Bilirubin Tester (transportable)[f,g]	Neonatal bilirubin

[a] Not FDA approved.
[b] CLIA waived (http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfClia/analyteswaived.cfm).
[c] Under development.
[d] i-STAT proprietary ACT reagent and test method.
[e] Compatible with Precision PCx strips.
[f] Not available in United States.
[g] Available in Thailand.

ALP, alkaline phosphatase; ALT, alanine aminotransferase; aPTT, activated partial thromboplastin time; AST, aspartate aminotransferase; β-hCG, beta-human chorionic gonadotropin; BNP, B-type natriuretic peptide; BUN, blood urea nitrogen; CBC, complete blood count; CGMS, continuous glucose monitoring system; CK, creatine kinase; CK-MB, creatine kinase-myoglobin; CLIA, Clinical Laboratory Improvement Acts (US); COHb, carboxyhemoglobin; Creat, creatinine; CRP, C-reactive protein; cTnI, cardiac troponin I; cTnT, cardiac troponin T; DM, data management; E3G, estrone-3-glucuronide; FCOHb, fraction of carboxyhemoglobin; FDA, Food and Drug Administration (US); FHbF, fraction of fetal hemoglobin; FHHb, fraction of deoxyhemoglobin; FmetHb, fraction of met-hemoglobin; FO_2Hb, fraction of oxyhemoglobin; GGT, gamma glutamyl transferase; GOT, glutamic oxaloacetic transaminase; GPT, Glutamic pyruvic transaminase; Hb, hemoglobin; HbA_{1c}, hemoglobin A_{1c}; hCG, human chorionic gonadotropin; Hct, hematocrit; HDL, high-density lipoprotein; H-FABP, human-type fatty acid–binding protein; HHb, deoxyhemoglobin; HNTT, heparin-neutralized thrombin time; HR-ACT, high-range activated clotting time; INR, international normalized ratio; Lac, lactate; LD, lactate dehydrogenase; LDL, low-density lipoprotein; LH, luteinizing hormone; LR-ACT, low-range activated clotting time; MetHb, met-hemoglobin; O_2Cap, oxygen capacity; O_2Ct, oxygen content; PBC, protein band cassette; PBD, protein band dipstick; PCA, portable clinical analyzer; pCO_2, partial pressure of carbon dioxide; PDAO, protamine dose assay; pO_2, partial pressure of oxygen; PSA, prostate specific antigen; PT, prothrombin time; RACT, recalcified activated clotting time; sO_2, oxygen saturation; SO_2%, percent oxygen saturation; tBil, total bilirubin; TC, transcutaneous; TCM, transcutaneous monitoring; TCO_2, total carbon dioxide; tHb, total hemoglobin; TSH, thyroid stimulating hormone; TT, thrombin time; WBC, white blood cell count.

Hematocrit. The proportion of red blood cells relative to plasma has an inverse effect on measured glucose. Tang et al. found increases in hematocrit decreased glucose meter readings.[36] Conversely; decreases in hematocrit were found to increase glucose values. The mechanisms that cause this effect may be diffusion related or mechanical in nature. High hematocrit can decrease the amount of glucose that diffuses into the biosensor. The large amount of red blood cells in high hematocrit samples may mechanically impede the biosensor and reduce glucose readings. Low hematocrit samples facilitate increased diffusion of glucose toward the biosensor, thereby leading to higher readings on the GMS. Some manufacturers address the hematocrit problem by prelysing red blood cells before testing (HemoCue) or by compensating for hematocrit effects through test strip design coupled to software algorithms (StatStrip) (Figure 4).[35]

pO_2 effects. Critically ill patients may be on oxygen therapy; therefore blood pO_2 levels may be higher than normal. Originally, GO-based systems showed inaccuracies at extreme pO_2 levels. This was common in systems that relied on reactions that generated a hydrogen peroxide (H_2O_2) intermediate. Hydrogen peroxide then disassociates into hydrogen ions, oxygen, and electrons (used to measure glucose).[37] As pO_2 increased (>100 mm Hg), the disassociation reaction would become less favorable, thus causing lower glucose readings. Recent GO-based systems (e.g., StatStrip, NOVA Biomedical, Waltham MA) use a modified GO enzyme method that alleviates pO_2 effects, thereby minimizing this confounding factor.[35] The opposite

Table 5 Principles of amperometric and photometric test strips

Advantage H, Comfort Curve
Glucose + GD/PQQ ⟶ gluconic acid + $GD/PQQH_2$
$GD/PQQH_2$ + ferricyanide ⟶ GD/PQQ + ferrocyanide
Ferrocyanide ⟶ ferricyanide + e^-
Precision PCx, Precision QID
Glucose + GO/FAD ⟶ gluconic acid + $GO/FADH_2$
$GO/FADH_2$ + ferricinium ⟶ GO/FAD + ferrocene
Ferrocene ⟶ ferricinium + e^-
$GO/FADH_2 + O_2$ ⟶ $GO/FAD + H_2O_2$
SureStep
Glucose + GO/FAD ⟶ gluconic acid + $GO/FADH_2$
$GO/FADH_2 + O_2$ ⟶ $GO/FAD + H_2O_2$
H_2O_2 + MBTH + HRP ⟶ $MBTH\text{-}R^+$
$MBTH\text{-}R^+ + ANS + \frac{1}{2}O_2$ ⟶ Blue-green dye
Glucometer Elite
Glucose + GO/FAD ⟶ gluconic acid + $GO/FADH_2$
$GO/FADH_2$ + Ferricyanide ⟶ GO/FAD + ferrocyanide
Ferrocyanide ⟶ ferricyanide + e^-
$GO/FADH_2 + O_2$ ⟶ $GO/FAD + H_2O_2$

GD, glucose dehydrogenase; PQQ, pyrroloquinoline quinone; GO, glucose oxidase; FAD, flavin adenine dinucleotide; H_2O_2, hydrogen peroxide; MBTH, *meta*(3-methyl 2-benzothiazolinone hydrazone)-*N*-sulfonyl benzenesulfonic acid; HRP, horseradish peroxidase; ANS, 8-anilino-1-naphthalenesulfonic acid.

may also be true, where low pO_2 (<40 mmg Hg) can affect glucose results by 15% when using GO-based systems.[38]

Hypotension and hypoperfusion. Sylvain et al. showed GD-based systems to have significantly different results ($P < 0.001$) when testing capillary whole-blood samples from hypotensive patients in comparison to a reference analyzer.[39] GO-based systems showed similar results. In contrast, venous samples tested on GMSs did not show statistically significant differences when compared to venous samples tested on laboratory analyzers.

Sample source. Capillary blood is the most common sample type used by POC GMSs. Several studies have shown variations in accuracy when comparing capillary versus venous blood samples.[40] Recent GMSs showed greater accuracy with arterial samples compared to capillary blood; however, older GMS models showed similar results between the two sample types. Results derived from venous blood samples exhibited improved accuracy compared to capillary blood for GD-based GMSs. Alternate site testing must be considered as well in evaluations of accuracy and may be system dependent.

Specimen matrix. Water, lipid, protein, and cellular content may also be confounding factors. Glucose is more

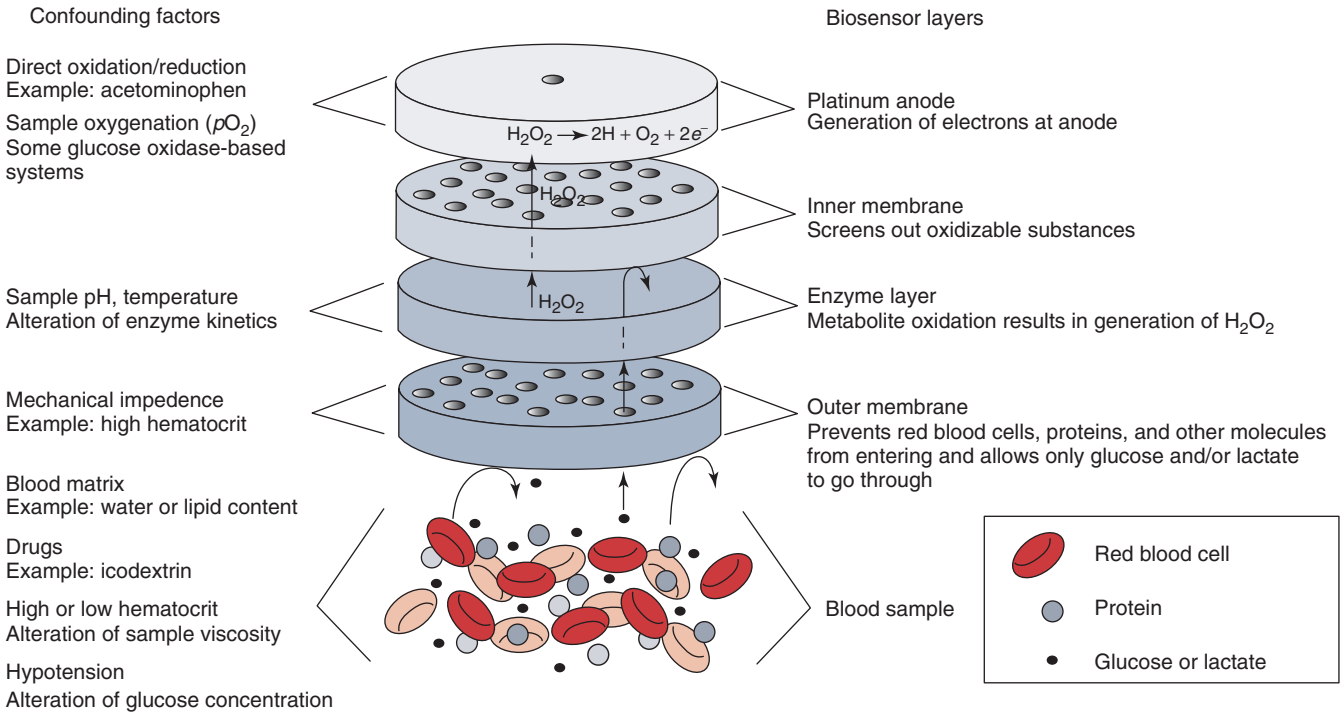

Figure 2 Glucose/lactate electrochemical-based biosensor. Glucose biosensor interferences affect various biosensor layers. Oxidative or reductive reactions typically interfere at the electrode. Blood oxygenation may affect older glucose oxidase–based systems that depend on an oxygen intermediate to generate electrons. Matrix and hematocrit effects interfere with glucose diffusion rates, thereby altering the amount of glucose detected.

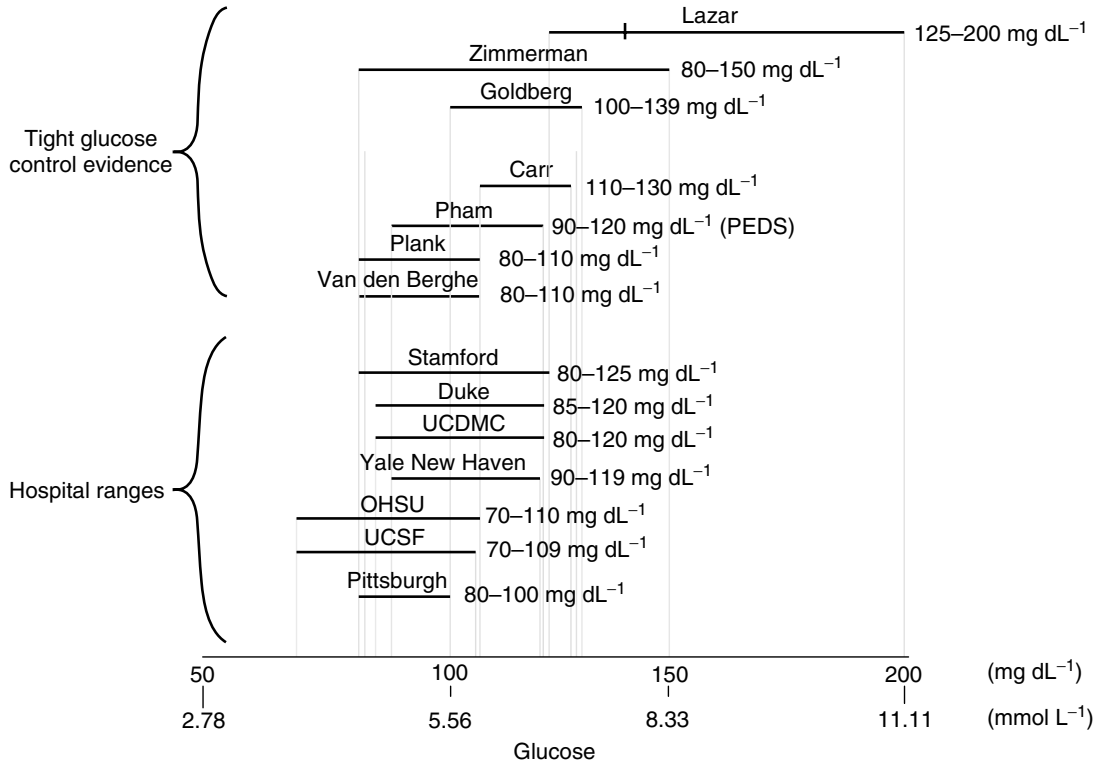

Figure 3 Tight glycemic control ranges as used in hospitals and reported in literature.

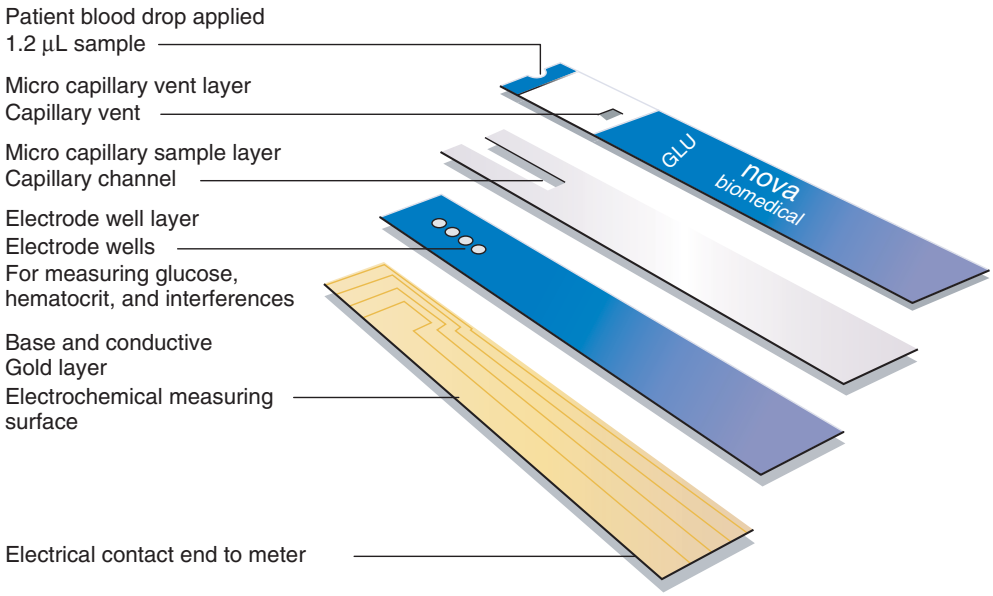

Figure 4 Schematic of a glucose meter test that automatically adjusts for hematocrit effects, and interference from oxidizing agents. Patents 6,287,451/6,837,976/EP 1 212 609/CA 2,375,092.

soluble in water; therefore, high plasma water content can cause higher glucose readings.[41] Lipid, proteins, and cellular components may impede photometric systems by increasing the turbidity of the sample. Such matrix effects may also alter the viscosity and hence reduce the glucose diffusion rate in both amperometric and photometric systems.

Drugs. Drugs that oxidize or reduce glucose biosensor electrodes may interfere with glucose results. Tang et al. conducted a study to determine the effects of 30 different

drugs on 6 different glucose meters.[42] High doses of ascorbic acid were shown to yield lower glucose values on GO-based systems. Acetaminophen caused increased glucose values in all tested GD-based systems and in some GO-based systems. High doses of dopamine increased glucose values in GD-based systems while mannitol increased glucose values in GO-based systems. More recently, icodextrin (osmotic agent for dialysis) has also been shown to generate falsely high results due to the drug being metabolized to maltose, which is indistinguishable from glucose on GD/PQQ-based instruments.[43]

pH and temperature. Since most GMSs convert glucose into a detectable signal via enzyme-catalyzed reactions, pH, and temperature may play a role in altering the enzyme kinetics, thereby causing errors. Kilpatrick et al. reported significant deviation in glucose measurements for samples with pH <6.95 and >7.85.[44] Temperature may also play a role. Oberg et al. suggests cold temperatures producing discrepant results; however, the effects of fever are currently unknown.[38,45]

Other factors. User error and environmental conditions may contribute to inaccuracies. Studies show the potential for operator error to still exist despite advances in more user-friendly instruments and protocols.[46] Over the years, environmental conditions have become more important due to the use of POC instruments under field conditions (e.g. disaster areas, war zones, and rural areas). POC instruments, including GMSs, were used on airborne hospitals to evacuate survivors after the Asian Tsunami of 2004.[47] In a separate study, Kost et al., identified conditions in post-Tsunami Thailand, and post-Hurricane Katrina New Orleans to be unacceptable for routine POC GMS use due to the high humidity and temperatures. This was supported by another study in 2007 that tested the robustness of POC instruments under certain environmental stresses.[48]

If glucose meter results are inaccurate, imprecise, or nonlinear in critical decision ranges, both medical and legal problems may arise, especially if undetected de novo onset of hypoglycemia, hyperglycemia, or diabetic ketoacidosis leads to an adverse outcome or if misleading results trigger inappropriate therapy, especially with TGC protocols.

2.1.5 *Pulse Oximetry*

Pulse oximeters are POC medical devices that noninvasively measures both the amount of oxygen in a patient's blood and changes in blood volume through the skin (Figure 5). These instruments consist of a clip that contains the photosensor, and two LEDs. One LED is red ($\lambda = 660$ nm) and the other emits light in the infrared range ($\lambda = 905$, 910, or 640 nm). The clip is placed on a part of the body, which is translucent to the light emitted by the LEDs (e.g. finger nail and ear). Some new instruments utilize forehead sensors to minimize interference.[49–51] Oxyhemoglobin absorbs more infrared light than deoxyhemoglobin. Conversely, deoxyhemoglobin absorbs more red light than oxyhemoglobin. Therefore, the oxyhemoglobin/deoxyhemoglobin ratio can be calculated on the basis of the amount of red versus infrared light detected by the photosensor. This ratio is then compared to stored reference tables within the instrument to generate oxygen saturation values. Pulsatile blood flow is measured by measuring the rising and decreasing amounts of absorbance that corresponds to blood volume. Absorbance is treated as a wave function where the resultant wave is used to calculate the changes in blood volume through the skin.

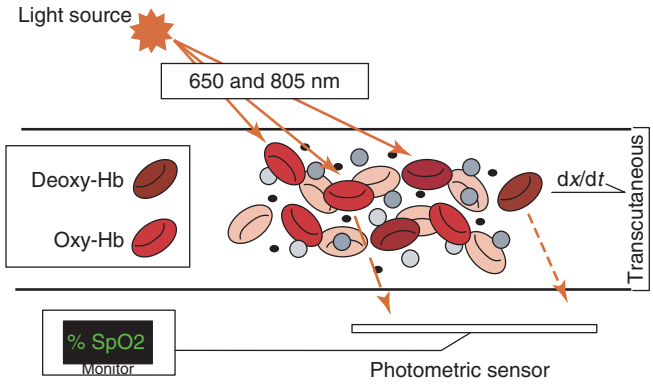

Figure 5 Pulse oximetry methods. Light sources emitting both red (650 nm) and infrared (850 nm) light are transmitted through the skin. Deoxyhemoglobin and oxyhemoglobin have different absorption profiles for red and infrared light. The photosensor calculates the ratio between red and infrared light and compares these values to correlation tables to derive an oxygen saturation level. Since blood flow is pulsatile, the photosensor will detect rises and falls in the red/infrared absorption ratios. The photosensor converts these absorption values into a wave function where the resultant is used to calculate the blood volume going through the skin.

Interferences. Pulse oximeters are susceptible to two types of interference due to (i) photointerference and (ii) hypoperfusion. Interferences have resulted in failure rates of 5–50%, and a potential reduction in sensitivity and specificity from 0.92 to 0.74 and 0.90 to 0.84, respectively.[52,53] Photo interference includes ambient light, sensor motion, surgical diathermy, and presence of either carboxyhemoglobin or methylene blue.[54–56] Hypoperfusion, as the result of cardiovascular disease, hypoxia, sepsis, or respiratory disease also contributes to error rates. In the end, interfering factors alter the detected light (e.g. absorption, refraction, reflection, and

constructive/destructive interference), which results in aberrant oxygen saturation results.

2.2 Analytical Methods for Hematology and Hemostasis Analytes

Table 4[12–21] summarizes representative POC devices, the physical principles of measurements, and the tests that the instruments perform. Generally, the devices use a drop of whole blood from a fingerstick or a venipuncture specimen that may be citrated depending on the method. For general information, please refer to the Clinical Laboratory Standards Institute (CLSI) guideline[57] that describes specimen preparation for coagulation tests. After reagent mixing, clotting of the blood sample in the reaction chamber produces a primary physical change that is detected either mechanically or optically. Measurements may also be performed noninvasively. For example, noninvasive hemoglobin monitoring was cleared by the Food and Drug Administration (FDA) in 2008 (Masimo, Limonest, France). Like the biosensor-based instruments in Section 2.1, instruments in Table 4 vary in size from transportable/portable to portable to nearly handheld, and correspondingly vary in convenience for use at the POC. For these types of instruments, electronic and whole-blood quality control are used to satisfy regulatory requirements. With these new POC technologies, one should evaluate test results for clinical endpoints in specific patient situations in parallel with test results obtained with conventional hematology and hemostasis methods in the parent (main) laboratory. Some new methods produce results that are linearly related or standardized to traditional methods. However, because of the different principles of measurement, some do not. Use of a global standardization parameter, such as the international normalized ratio (INR) for reporting of prothrombin time (PT) results, is an essential strategy. Error can result from admixed tissue fluid when capillary specimens are used. Hence, alternative sources of specimens are more suitable for patients in intensive care.

ECSs involve applying an alternating voltage across two or more electrodes in contact with the fluid sample and measuring impedance (conductance) to the resulting current flow. This method is used to determine the hematocrit in whole blood. The conductivity is dependent on the volume of relatively nonconductive blood cells between the electrodes and on the electrolyte concentration in the blood. The electrolyte concentration can be approximated by the simultaneous measurement of the whole-blood sodium level to produce a more accurate estimate of hematocrit. Conductance-based measurements of hematocrit may not be accurate in patients given autologous blood transfusions where plasma has been largely replaced by crystalloid[58] or in surgery patients with variations in protein levels and osmolality.[59] Therefore, hematocrit determined by ECS must be interpreted with caution in critically ill patients, particularly those patients who receive intravenous therapy and blood products. Hematology analyzers based on electronic particle counting can have an advantage when there is abnormal protein level or osmolality.[13] Platelet function analysis is clinically valuable and is available on some devices. There is a need for both miniaturized hematology endothelial progenitor cells (EPC) instruments and platelet function analyzers for POCT in critical care settings.

2.3 Cardiovascular and Cardiac Injury Markers

The use of cardiac biomarker testing in the emergency department is one of the most difficult challenges faced by emergency medicine physicians today. Admissions of patients with low probability of acute coronary artery disease often times leads to excessive costs,[60] while release of patients with acute myocardial infarction has become the number one leading cause of malpractice lawsuits in the emergency department.

Generally, patient-focused clusters of markers, selected from cardiac troponin I (cTnI), creatine kinase-MB (CK-MB) isoenzyme, B-type natriuretic peptide (BNP), N-terminal pro-B-type natriuretic peptide (NT-proBNP), D-dimer and myoglobin represents a good diagnostic tool for both ruling in and ruling out acute myocardial infarction and related conditions.[61] cTnI, for example, has shown to be a highly sensitive and specific marker compared to myoglobin. It is suggested that the 99th percentile reference serves as a cutoff for the diagnosis of myocardial infarction in the emergency department.[62] Another marker, BNP and NT-proBNP, are also good heart failure biomarkers. Additionally, there is evidence that BNP is released due to ischemic events and thus may also serve as a marker for myocardial infarction.[60]

Several portable and handheld devices are available for POC measurement of cardiac injury markers. For example, one handheld device for the qualitative measurement of cTnI uses a solid phase chromatographic immunoassay, where the sample (200 μL) is transferred into the receptacle region containing monoclonal anti-cTnI antibody–dye conjugates and biotinylated rabbit polyclonal anti-cTnI antibodies. These antibodies bind to cTnI in the sample to form complexes that migrate through the reaction strip. The antigen/antibody–dye complexes are then captured by immobilized streptavidin in the cTnI zone. Additional protein–dye conjugates not bound in the cTnI area are later captured in the control area. Results are available in 15 min. Another handheld model tests for the cardiac isoenzyme of creatine kinase,

BNP, CK-MB, and myoglobin using analogous principles (Table 4). Other handheld or portable models measure cardiac injury markers quantitatively, simultaneously, and quickly (in a few minutes).[60]

The National Academy of Clinical Biochemistry guidelines recommend cardiac biomarker testing instruments to have a high clinical sensitivity and specificity to facilitate early diagnosis of acute myocardial infarction. No signal marker meets all of these criteria; therefore, the multianalyte approach has the most merit.[60] New specific markers are also needed. Future markers may include heart fatty acid–binding protein, myeloperoxidase, ischemia-modified albumin, and soluble CD40 ligand.

POC cardiac biomarker testing is generally recommended at institutions that cannot deliver biomarker TATs of ~1 h. However, faster TAT is more efficient for patient care paths. Laboratory personnel must be involved in the selection of these devices, the training of individuals to perform analysis, the maintenance of POC equipment, verification of the proficiency of operators on a regular basis, and the compliance of documentation with requirements by regulatory agencies.[23,63]

As time progresses, the standard of care for timeliness for biomarker testing will also shift toward faster turnaround times (Figure 6). Contemporary timeliness in critical care testing has set a continuum precedent for chest pain centers that now consider 15 min as a reasonable criterion for accreditation. However, cardiac biomarker results should be available within 30 min, the preference stated by ACC/AHA guidelines.[10,64]

2.4 Immunoassays

Like in cardiac biomarker testing, antibodies can be used to target disease-specific molecules. Devices and test kits are available for the evaluation of patients with metabolic (e.g. β-hydroxybutyrate for diabetes mellitus), infectious (e.g. human immunodeficiency virus, HIV), and other diseases. Samples may include whole blood, sputum, and stool. Other methods include less-invasive or noninvasive samples including urine or saliva to detect sexually transmitted diseases and HIV, respectively.

POC infectious disease tests are predominantly based on immunoassays. Reagent cards, test kits, or swabs are common test formats. Table 6 provides a list of current POC infectious disease tests. Over the years, the convenience and accuracy of POC infectious disease tests has led to increased acceptance by the medical and public health community. In 2006, the Center for Disease Control and Prevention issued a recommendation for POC saliva-based screening for HIV in all Americans from age 16 to 64.[65] With the threat of avian influenza, rapid POC tests for H5N1 and H7N7 strains are being developed to screen patients. Current POC influenza tests may not be specific enough to discern between nonavian versus avian influenza. If not rectified, this lack of sensitivity and specificity may contribute to the spread of a pandemic.[66,67]

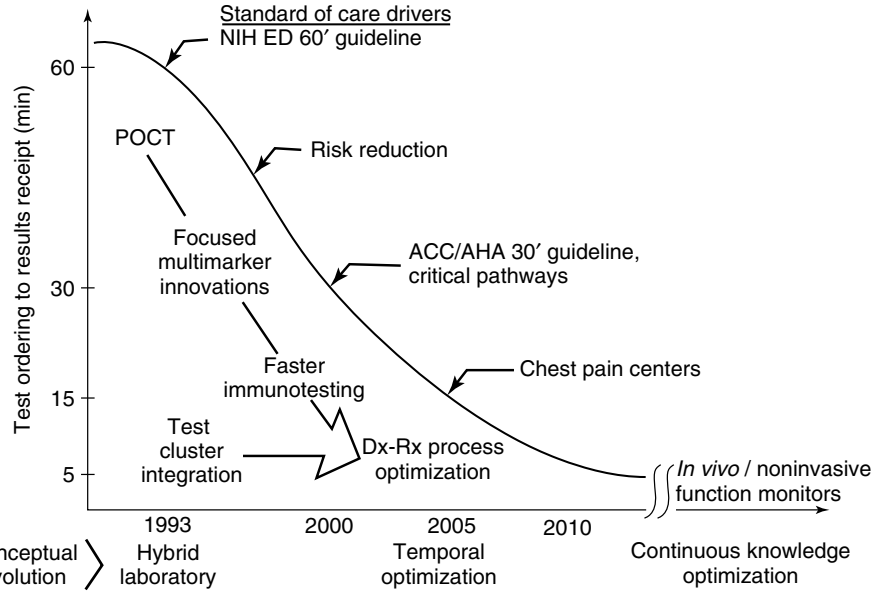

Figure 6 Standards of care for cardiac biomarker testing. The standard of care for timeliness of cardiac biomarker testing. Forceful clinical drivers mold the shape of the rapid response curve while the underlying performance vector enables it, aiming at optimization, first discrete, then in the future, continuous. Contemporary timeliness in critical care testing has set a continuum precedent for chest pain centers, which now consider 15 min as reasonable criterion for accreditation. However, all cardiac biomarker results should be available within 30 min, the preference stated in ACC/AHA guidelines.

Table 6 POCT for infectious diseases

Disease/ pathogen	Device or test	Manufacturer	Format/ method	Sample
Adenovirus	Adeno-Strip	Coris Bioconcept	Disposable reagent kit	Stool
	Immunocard Adenovirus[a]	Meridian Bioscience	Disposable reagent card	Stool
Avian Flu (H5N1, H7N3) flu	Multiplexed POC/flu tests[b]	In development	Chip[b](UCol), others	Depends on method
Bacillus anthracis	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
Bacterial Endotoxin	Sepsis EAA System	Spectral Diagnostics	Disposable reagent kit	Whole blood
Biological weapons	μChemLab BD	Sandia National Labs	Handheld chemical analysis	Various pathogens
	BioBriefcase	Lawrence Livermore	Transportable PCR	Various viruses, Bacteria, toxins
	Bio-Seeq Handheld PCR	Smiths Detection	Handheld PCR	Various pathogens
Borrelia burgdorferi (Lyme Disease)	PreVue B. burgdorferi	Wampole Labs	Disposable reagent card	Whole blood
Candida species	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
Clostridium difficile	Acceava	Biostar	Disposable reagent card	Stool
	Immunocard Toxin A & B[c]	Meridian Bioscience	Disposable reagent card	Stool
	PreVue (toxin)	Wampole Labs	Disposable reagent card	Stool
Chlamydia trachomatis	Abbott Test Pack	Abbott	Disposable reagent card	Cervical swab contents
	Biostar OIA	Biostar	Disposable reagent card	_
	BioSign Chlamydia II[a]	PBM	Disposable reagent card	_
	One-Step Test[a]	Acon Labs	Disposable reagent card	Urine dip stick
	Clearview	Wampole Labs	Disposable reagent card	Cervical swab contents
	QuickView	Quidel Corp.	Disposable reagent card	Cervical swab contents
	Rapid 1-2-3 Hema Test[a]	Hema Diagnostics	Disposable reagent card	Whole blood
Cholera (*Vibrio cholerae*)	BioSign Cholera	PBM	Disposable reagent card	Whole blood
	Megakwik Cholera Ab	Mega Diagnostics	Disposable reagent card	Whole blood
	SMART Cholera 01	New Horizons	Disposable reagent card	Whole blood
Cryptosporidium parvum	Immunocard Crypto/Giardia[a]	Meridian Bioscience	Disposable reagent card	Stool
Dengue Fever Virus	1-Step Test[a] (IgG + IgM)	Teco Diagnostics	Disposable reagent card	Serum/whole blood
	AccuSpot[a] (IgG + IgM)	AccuDx	Disposable reagent card	Serum/whole blood
	Core Dengue[a] (IgG + IgM)	Core Diagnostics	Disposable reagent card	Serum/whole blood
	Dengue Duo[a] (IgG + IgM)	PanBio	Disposable reagent card	Serum

(*continued overleaf*)

Table 6 (*continued*)

Disease/ pathogen	Device or test	Manufacturer	Format/ method	Sample
Enterococcus faecalis/faecium	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
Epstein–Barr Virus	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
	QuickVue	Quidel Corp.	Disposable reagent card	Serum/plasma/whole blood
E. coli O157	Immunocard *E. coli* O157 Plus[a]	Meridian Bioscience	Disposable reagent card	Stool
Giardia lamblia	Immunocard Crypto/Giardia[a]	Meridian Bioscience	Disposable reagent card	Stool
Hepatitis B virus	HBV Hepatitis B Virus Combo[a]	Acon Labs	Disposable reagent card	Serum/plasma
	HbsAb Hepatitis B[a]	Acon Labs	Disposable reagent card	Serum/plasma
Hepatitis C virus	HCV Hepatitis C Virus[a]	Acon Labs	Disposable reagent card	Serum/plasma
	Core HCV[a]	Core Diagnostics	Disposable reagent card	Serum/plasma
Helicobacter pylori	BioSign H. pylori WB[c]	PBM	Disposable reagent card	Whole blood
	FlexSure HP[c]	Beckman Coulter	Disposable reagent card	Serum/whole blood
	One-Step + H. pylori Test	Henry Schein	Disposable reagent card	Serum/whole blood
	Rapid Test Strip[c]	Acon Labs	Disposable reagent card	Serum/whole blood
Helicobacter pylori (continued)	Immunocard H. pylori	Meridian Bioscience	Disposable reagent card	Serum/plasma/whole blood
	Immunocard STAT![c]	Meridian Bioscience	Disposable reagent card	Stool
	QuickVue[c]	Quidel Corp.	Disposable reagent card	Serum/plasma/whole blood
	Signify HP Test[c]	Abbott	Disposable reagent card	Whole blood
	Status H. pylori[c]	LifeSign	Disposable reagent card	Serum/plasma/whole blood
	Uni-Gold H. pylori	Trinity Biotech	Disposable reagent card	Serum/plasma/whole blood
HIV-1 or 2	BioSign HIV[a]	PBM	Disposable reagent kit	Whole blood
	Core HIV-1/2[a]	Core Diagnostics	Disposable reagent kit	Whole blood
	dBest One-Step HIV Test[a]	AmeriTek	Disposable reagent kit	Whole blood
	Determine[a]	Abbott Laboratories	Disposable reagent kit	Whole blood
	Home Access Test System	Home Access	Disposable reagent kit	Whole blood
	Multispot HIV-1/2 Rapid Test[c]	Bio-Rad	Disposable reagent kit	Whole blood
	OraQuick Advanced HIV-1/2[c]	Orasure Technologies	Disposable reagent kit	Oral swab
	CyFlow Counter	Partec	Benchtop/Flow Cytometry	Whole blood
	Rapid 1-2-3 Hema Test[a]	Hema Diagnostics	Disposable reagent kit	Whole blood
	Rapid Test Strip[a]	Acon Labs	Disposable reagent kit	Serum/plasma/whole blood

Table 6 (*continued*)

Disease/ pathogen	Device or test	Manufacturer	Format/ method	Sample
	Serrodia HIV-1/HIV-2[a,d]	Fujirebio	Disposable reagent kit	Serum/plasma/whole blood
	Uni-Gold HIV[c]	Trinity Biotech	Disposable reagent kit	Serum/plasma/whole blood
	Vscan[a]	Medical Service Intl.	Disposable reagent kit	Serum/plasma/whole blood
Herpes Simplex Virus	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
Human papilloma virus	POCT not available, but badly needed	–	–	–
Influenza A/B	Acceava	Biostar	Disposable reagent kit	Nasal swab
	Binax NOW Flu A Test[c]	Binax	Disposable reagent kit	Nasal swab
	Binax NOW Flu B Test[c]	Binax	Disposable reagent kit	Nasal swab
	Now Flu A/B[c]	Meridian Bioscience	Disposable reagent kit	Nasal swab/ aspirate/wash
	QuickVue[c]	Quidel Corp.	Disposable reagent kit	Nasal swab/ aspirate/wash
	ZstatFlu Test[c]	ZymeTx	Disposable reagent kit	Throat swab
Leptospira	Visitect Lepto[a]	Omega Diagnostics	Disposable reagent kit	Serum/plasma/whole blood
Listeria monocytogenes	BioSign Listeria	PBM	Disposable reagent card	Whole blood
Malaria	ASSURE Malaria Rapid Test	GeneLabs Diagnostics	Disposable reagent card	Whole blood
	BioSign Malaria[a]	PBM	Disposable reagent card	Whole blood
	Core Malaria	Core Diagnostics	Disposable reagent card	Whole blood
	dBest One-Step Test[a]	AmeriTek	Disposable reagent card	Whole blood
	Health Test Malaria	Akers Laboratories	Disposable reagent card	Whole blood
	MegaKwik Malaria Card Test	Mega Diagnostics	Disposable reagent card	Whole blood
	NOW Malaria	Binax	Disposable reagent card	Whole blood
	One-Step RapiCard InstaTest	Cortez Diagnostics	Disposable reagent card	Whole blood
	One-Step Test[a]	Acon Labs	Disposable reagent card	Whole blood
	OptiMAL Rapid Malaria Test	DiaMed SA	Disposable reagent card	Whole blood
	Rapid 1-2-3 Hema Test[a]	Hema Diagnostics	Disposable reagent card	Whole blood
Mycobacterium tuberculosis	BioSign TB[a]	PBM	Disposable reagent card	Whole blood
	Core Tuberculosis[a]	Core Diagnostics	Disposable reagent kit	–
	Vscan[a]	Medical Service Intl.	Disposable reagent kit	Whole blood
Multipathogens (Gm+, Gm−, fungi, MRSA)	SeptiFast (Sepsis) (Europe)	Roche Diagnostics	Benchtop Multiplexed PCR	Whole blood
Mycoplasma pneumoniae	Immunocard Mycoplasma	Meridian Bioscience	Disposable reagent card	Serum

(*continued overleaf*)

Table 6 (*continued*)

Disease/ pathogen	Device or test	Manufacturer	Format/ method	Sample
Neisseria gonorrhea	Rapid 1-2-3 Hema Test[a]	Hema Diagnostics	Disposable reagent kit	Whole blood
Parvovirus	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, and capillary)
Pseudomonas aeruginosa	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
Respiratory Syncytial Virus	Immunocard RSV	Meridian Bioscience	Disposable reagent card	Nasal swab/ aspirate/wash
	Clearview NOW RSV	Wampole Labs	Disposable reagent card	Nasal swab/ aspirate/wash
Rotavirus	Immunocard Rotavirus[a]	Meridian Bioscience	Disposable reagent card	Stool
Rubella	BioSign Rubella[a]	PBM	Disposable reagent card	Whole blood
	PreVue	Wampole Labs	Disposable reagent card	Serum/plasma
Salmonella typhi (Typhoid)	KIT dipstick assay	KIT Biomedical	Disposable reagent card	Serum/whole blood
	TYPHIdot[a]	MBDr	Disposable reagent card	Serum
SARS virus	Rapid SARS Test[a]	GeneLabs	Disposable reagent kit	Serum/whole blood
Sepsis Markers	Triage Sepsis Panel (NGAL, CRP, MIP-3)[a]	Biosite	Portable	Whole blood
Shigella flexneri	S. flexneri 2a dipstick[a]	Pasteur Institute, Paris	Dipstick	Stool
Staphylococcus (including MRSA)	LightCycler	Roche Diagnostics	Benchtop PCR	Whole blood (venous, arterial, capillary)
	IDI MRSA[e]	IDI and Cepheid	Benchtop PCR	Nasal swab
Streptococcus agalactiae	Strep B Assay[e]	IDI and Cepheid	Benchtop PCR	Vaginal/rectal swab
Streptococcus pyogenes	Acceava[c]	Biostar	Disposable reagent kit	Throat swab
	AccuStrip Strep A[c]	Jant Pharmacal	Disposable reagent kit	Throat swab
	Binax NOW Strep A[c]	Binax	Disposable reagent kit	Throat swab
	Contrast Strep A[c]	Genzyme	Disposable reagent kit	Throat swab
	ICON DS Strep A[c]	Beckman Coulter	Disposable reagent kit	Throat swab
	Immuno Strep A Detector[c]	Immunostics	Disposable reagent kit	Throat swab
	LLC Strep A[c]	LifeSign	Disposable reagent kit	Throat swab
	One Strep+[c]	Henry Schein	Disposable reagent kit	Throat swab
	PolyStat Strep A[c]	Polymedco	Disposable reagent kit	Throat swab
Streptococcus pyogenes (continued)	QuickVue[c]	Quidel Corp.	Disposable reagent kit	Throat swab
	QuStrip A[c]	Stanbio	Disposable reagent kit	Throat swab
	PreVue[a]	Wampole Labs	Disposable reagent kit	Throat swab

Table 6 (*continued*)

Disease/ pathogen	Device or test	Manufacturer	Format/ method	Sample
	RIM A.R.C. Strep A Test[c]	Remel	Disposable reagent kit	Throat swab
	Signify Strep A Test[c]	Abbott	Disposable reagent kit	Throat swab
	SP Brand Rapid Strep A[c]	Cardinal Health	Disposable reagent kit	Throat swab
	Status First Strep A[c]	PBM	Disposable reagent kit	Throat swab
	Strep A Rapid Twist Test[c]	Acon Labs	Disposable reagent kit	Throat swab
	StrepAim Rapid Dipstick Test[c]	Germaine Labs	Disposable reagent kit	Throat swab
Syphilis (*Treponema pallidum*)	BioSign Syphilis[a]	PBM	Disposable reagent card	Whole blood
	One-Step Test	Acon Labs	Disposable reagent card	Serum/plasma
Trichomonas vaginalis	Genzyme OSOM	Genzyme	Disposable reagent kit	Vaginal swab
West Nile Virus	RAMP Reader WNV Test	Response Biomedical	Disposable reagent kit	Mosquito or bird sample
	Rapid 1-2-3 Hema Test[a]	Hema Diagnostics	Disposable reagent kit	Whole blood
	WNV IgM STATus	Spectral-IDx	Disposable reagent kit	Whole blood

[a] Not FDA-approved.
[b] CDC warns: "... sensitivity of commercially available rapid diagnostic tests for influenza may not always be optimal." (http://www.cdc.gov/flu/avian/professional/han081304.htm) but has evaluated UCol chip for 11 strains, including avian flu, and found the results promising.
[c] CLIA-waived (http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCLIA/search.cfm).
[d] UNFPA website: http://www.unfpa.orghiv/prevention/factsheet2.htm.
[e] Requires Cepheid SmartCycler II instrument.
[f] See http://www.cap.org/apps/docs/proficiency_testing/ waived.html for additional CLIA-waived tests.
CLIA, Clinical Laboratory Improvement Acts (US); CRP, c-reactive protein; FDA, Food and Drug Administration (US); Gm+, Gram's stain positive; Gm-, Gram's stain negative; MIP-3, macrophage inflammatory protein-3; MRSA, methicillin resistant *Staphylococcus aureus*; NGAL, neutrophil gelatinase-associated lipocalin; PCR, Polymerase chain reaction; SARS, severe acute respiratory syndrome.
Websites: Abbott Labs, www.abbottdiagnostics.com; AccuDx, Inc., www.accudx.com; Acon Labs, www.aconlabs.com; Akers Laboratories, www.akerslaboratories.com; AmeriTek, www.ameritek.org; Beckman Coulter, www.beckmancoulter.com; Binax, www.binax.com; Bio-Rad, www.bio-rad.com; Biosite, www.biosite.com; Biostar, www.thermo.com; Cardinal Health, www.cardinal.com; Cepheid, www.cepheid.com; Core Diagnostics, www.corediag.com; Coris BioConcept, www.corisbio.com; Cortez Diagnostics, www.rapidtest.com; Diamed, www.diamed.com;
Websites (continued): Fisher Healthcare, www.fishersci.com; GeneLabs Diagnostic, www.genelabs.com.sg; Genzyme, www.genzymediagnostics.com; Germaine Labs, www.germainelabs.com; Hema Diagnostics, www.rapidbloodtests.com; Henry Schein Inc, www.henryschein.com; Home Access, www.homeaccess.com; IDI, www.infectio.com; Immunostics, www.immunostics.com; Jant Pharmacal, www.accutest.net; LifeSign, www.lifesignmed.com; MBDr, www.mbdr.net; Medical Services International, www.medicalservicesintl.com; Mega Diagnostics, www.mega-dx.com; Meridian Biosciences, www.meridianbioscience.com; New Horizon Diagnostics, www.nhdiag.com/index.htm; Omega Diagnostics, www.omegadiagnostics.co.uk; Orasure Technologies, www.orasure.com; PanBio, www.panbio.com; Partec, www.partec.com; PBM, www.pbmc.com; Polymedco, www.polymedco.com; Quidel Corp., www.quidel.com; Remel, www.remelinc.com; Response Biomedical, www.responsebio.com; Roche Diagnostics, www.rochediagnostics.com; Sandia National Labs, www.sandia.gov; Smiths Detection, www.trace.smithsdetection.com; Spectral Diagnostics, www.spectraldx.com; Stanbio, www.stanbio.com; Teco Diagnostics, www.tecodiag.com; Trinity Biotech, www.trinitybiotech.com; UCol, University of Colorado; Wampole Laboratories, www.wampolelabs.com; and ZymeTx, www.zymetx.com.

2.5 Point-of-care Molecular Diagnostics

Nucleic acid amplification testing (NAAT) methods using polymerase chain reaction (PCR) provide highly sensitive and specific tools at the POC. PCR utilizes DNA polymerases (taq polymerase for DNA or reverse transcriptase for viral RNA) to amplify a target sequence. The amplified genetic materials (i.e., amplicons) are detected using fluorescent probes or absorbance (e.g., turbidity). Coupling of fluorescent-based detection and melting point analysis allows for target differentiation and quantitation. Handheld and transportable PCR-based POC instruments have appeared in the past decade for biowarfare detection (e.g. toxins and pathogens). More recently, in the clinical environment, multiplexed PCR-based bench top systems have been developed to

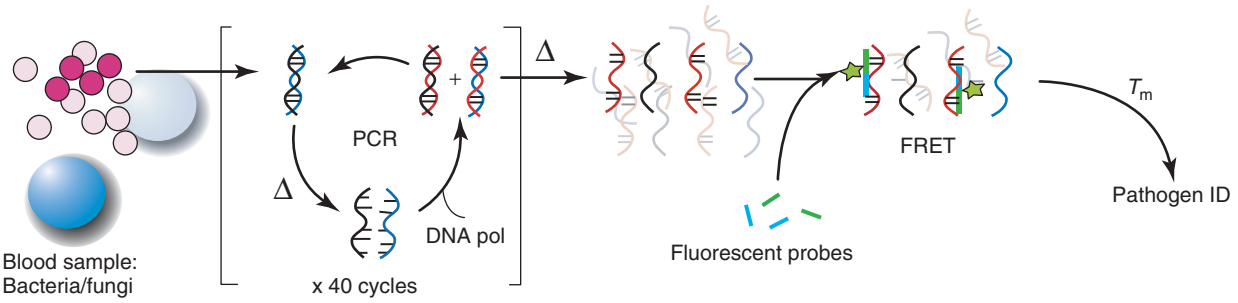

Figure 7 Polymerase chain reaction–based detection of pathogens. Heat (Δ) is used to denature DNA strands. Specific primers bind to target sequences to allow DNA replication by DNA polymerase (DNA pol). Fluorescent probes are used to detect specific target sequences in the amplified DNA. Fluorescent resonance energy transfer (FRET) can be coupled to melting point (T_m) analysis to provide further differentiation between DNA sequences, thereby allowing for multiplexed pathogen detection.

detect pathogens in patients suspected of having sepsis (Figures 8 and 9).[68]

Multiplex PCR NAAT. Existing instruments such as SeptiFast (Roche Diagnostics, Grenzacherstrasse, Switzerland) provide multiplex detection of 25 different gram positive, gram negative, and fungal pathogens. SeptiFast has been shown to detect additional organisms compared to blood culture in blood samples collected in parallel.[68] (Figures 7–9) PCR-positive results may reflect a real infection. Collectively, 60 patients had PCR and/or blood culture positive results, and 140 negative PCR and blood culture results. Polymicrobial bloodstream infections were identified in nine cases. In all four PCR-positive cellulitis cases, PCR detected a pathogen that was not found by blood or other cultures. Thirty-two PCR- and blood culture–positive cases of coagulase negative *Staphylococcus* cases were judged as a contaminant by clinical referees.

Other NAAT. As technology improves, we will see more molecular-based pathogen detection as seen with the PCR-based system approved by the FDA for screening group B *Streptococcus* infections in newborns.[69] Other non-PCR-based tests have emerged such as loop-mediated isothermal amplification (LAMP). LAMP amplifies nucleic acid using multiple primers at a single temperature (63 °C) in less than 1 h. The rapidity and isothermal nature may allow this technology to be applied to disaster and field NAAT.

2.6 Criteria for Instrument Evaluation and Selection

The most important practical criterion to consider when choosing an instrument for POCT is its medical efficacy (Table 7). Test clusters (test menus) should be patient focused and facilitate medical decision making. The standard of care for rapid response testing is evolving. A TTAT of 5 min is achievable with POCT and now is expected during critical emergencies.[1] Speed should be accompanied by cost effectiveness.

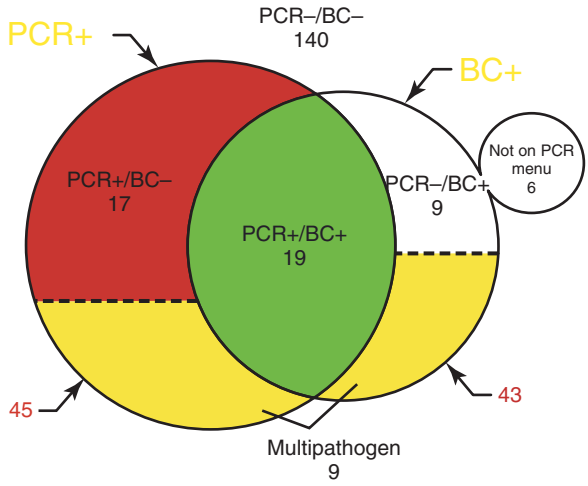

Figure 8 The Venn diagram shows the distribution of PCR and blood culture positive cases. The distribution of cases is based on the comparison of PCR to blood culture results collected at the same time point.

Economic efficacy is measured in terms of various outcome metrics, such as professional productivity and length of stay in critical care. It is important to establish the viewpoint of economic justification with respect to site (e.g. the intensive care unit), diagnosis (e.g. a particular diagnosis-related group), and/or outcome (e.g. the prevention of pseudohyponatremia). System features such as test flexibility, modularity, expandability, and upgrade capability, as well as instrument, consumable, and maintenance costs weigh heavily in the assessment of economic trade-offs. These features must match the workload and its pattern. For example, stationary analyzers with fast throughput may be preferred when there are several operators or bursts in test volume.

Conservation of patient blood volume depends on the specimen volume, type, and matrix (i.e. whole blood, plasma, or serum), and on the number of specimens that are required to perform a given cluster of diagnostic tests.

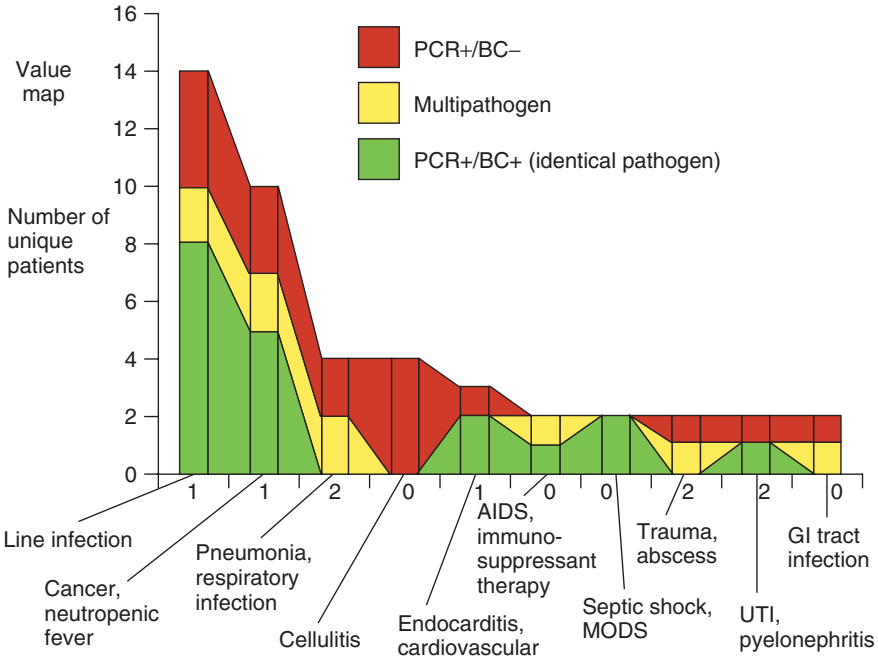

Figure 9 A value map showing the distribution of PCR-positive cases by clinical problem.

If only one or two analyte measurements are desired, it is an advantage if the instrument accepts a smaller sample volume compared to the sample required for the entire test menu. Blood loss from frequent diagnostic testing leads to unnecessary transfusion and ''hidden'' or ''downstream'' risks for the patient and the institution. Most POC devices use whole blood, which conserves volume and eliminates centrifugation. Elimination of processing steps saves time. Fast test results at the POC are most beneficial for the trending of an analyte that changes rapidly or unexpectedly, that is, has a short medical half-life, or for the rapid detection of life-threatening abnormalities. Results are outdated quickly by delays in transporting specimens to the main laboratory or by delays in processing plasma or serum for analysis. Outdated results encourage unnecessary repetition of testing and additional blood loss. Additionally, if an analyte has already been measured and its level is not changing, or if it is not important in medical decision making, there is no need to measure it at the POC. Wasteful duplication should be avoided. The net effect of these strategies is conservation of patient blood volume and institutional resources.

Table 8 also outlines design criteria for safety, ergonomics, security, and risk, as well as technical and quality features. Biohazard containment, space, storage, and disposal are especially important in intensive care. The design of a POC instrument should be ergonomic. Speed, ease, and security of operation are essential. Throughput must match clinical input, especially during resuscitations and emergencies, when there are bursts in testing volume that may overwhelm POC resources. Excellent on-site performance is essential if whole-blood analyzers are to be clinically effective. Handheld and portable instruments should be compact, reliable, durable, lightweight, and efficient, with streamlined requirements for quality control, maintenance, and reagents. Reagents must have adequate shelf life. Instrument operation should be straightforward with few preanalytical processing steps, and, therefore, less potential for error and biohazard exposure. In an emergency, the operator should be able to interrupt an automatic extended (two-point) calibration to save time. The pretest (one point) calibration cycle should be short. Calibration drift should be minimal. Programmability, self-diagnostics, data management, and simplicity are important, since operators with different levels of experience perform measurements. Whole-blood methods should not be sensitive to operator technique or to the type of clinical application (e.g. resuscitation, POC, mobile cart, or satellite laboratory). Whole-blood analyzers may be effective in a main laboratory if there is a fast transport system. Cautiously balance the location of testing, quality of measurements, format of instruments, and total operating expenses. On a per test basis, handheld instruments generally are less expensive to purchase but more expensive to operate, compared to transportables. Handheld instruments and their supplies also run the risk of theft and should be secured.

Table 7 Criteria for the selection and evaluation of POCT systems

- System performance
  - Accuracy, precision, bias, resolution, reproducibility, stability (biosensor drift), and response time
  - Linearity and performance in the high and low extremes of measurements, including critical limits
  - Consistency and relationship to tests performed in the parent (main) laboratory
  - Compatibility with other *in vitro*, and with *ex vivo* and *in vivo* whole-blood systems
  - Artifact elimination, error detection, interferences warning, and specimen flagging
- Technical and quality features
  - Throughput, automated calibration, interrupt capability, and duration of analysis cycle
  - Compact, reliable, durable, lightweight, mobile, and power efficient with battery operation
  - Reagent stability, shelf life, lot size, and variability (biosensor, calibrant, cartridge)
  - Continuous quality improvement (quality control, quality monitors, proficiency testing)
  - Number and efficiency of preanalytical and postanalytical process steps
  - Analytical performance with whole-blood and other sample types (e.g. CSF and body fluids)
  - Compliance with federal, state, and accreditation regulations and training requirements
- Qualification of new devices/tests and education/competency of operators
  - Protocols for evaluation and method comparisons (CLSI) to initiate new devices and tests
  - Instructions for meeting the requirements of federal, state, and voluntary accreditation agencies
  - Resources for education and competency (interactive, video, CD-ROM, virtual, Internet, or other)
- Connectivity, integration, and data storage/retrieval
  - Interfacing with physiological patient monitors in critical care areas
  - Remote review, remote control, robotics, and data management/overview system
  - Networking, interfacing (bidirectional), and wireless communications
  - Information systems, data storage, and archiving capabilities (LIS, HIS, CDR, CIS)
  - Compliance with accepted connectivity standards
- Safety, ergonomics, security, and risk
  - Biohazard control, containment, and disposal
  - Speed, ease, simplicity, and user-friendly operation
  - Identification, validation, notification, and security for operator and patient
  - Risk reduction features and on-site error minimization
- Conservation of patient blood volume
  - Specimen volume, type, and matrix (e.g. whole blood, plasma, or serum) accepted
  - Medically useful half-life of test results
  - Elimination of unnecessary transfusions and their risks
- Economic efficiency
  - Professional productivity and satisfaction
  - Impact on length of stay
  - Cost effectiveness by site, diagnosis, and/or outcome
  - Flexibility, modularity, exchangeability, expandability, and upgrade capability
  - Costs of instruments, consumables, and maintenance
- Medical efficacy in diagnosis, interpretation, treatment, and management
  - Improved decision making and patient outcomes
  - Critical care profile, test clusters, patient focusing, and medical indications
  - TTAT for temporal optimization and trend identification
  - Care path readiness for optimization of diagnostic–therapeutic strategies
  - Suitability for resuscitations, emergencies, peak workflow, and bursts in test volume
  - Alarms, broadcasts, and critical results notifications to physician/clinical team
  - Pattern recognition and interpretation of notable patient findings at point-of-care locations
- Hospital glucose testing – special considerations
  - Performance standards for bedside glucose testing in hospital settings should adopt LS-MAD curves and recommended error tolerance of $\leq 5\,\mathrm{mg\,dL^{-1}}$
  - POCT teams should evaluate performance and cost effectiveness in the context of tight glucose control, therapeutic decisions, and outcomes in critical care
  - Future guidelines should include LS-MAD curve pattern recognition when assessing glucose meter accuracy
  - Licensing agencies, such as the FDA, should require ISO 15197-integrated Bland–Altman and LS-MAD curve plots with highlighted discrepant values and erroneous results explained before validating POC glucose monitoring devices

CDR, clinical data repository; CIS, clinical information system; CLSI, Clinical Laboratory Standards Institute; CSF, cerebrospinal fluid; FDA, Food and Drug Administration; HIS, hospital information system; LIS, laboratory information system; LS-MAD, locally smoothed median absolute difference; POC, point of care; POCT, point-of-care testing; TTAT, therapeutic turnaround time.

Table 8 Test clusters and the critical care profile

Vital function	Diagnostic pivots (physiological indicators)
Energy	Glucose[a]
	Hemoglobin[a]
	pO_2[a]
	O_2 saturation[a]
Conduction	Potassium[a]
	Sodium[a]
	Ionized magnesium[a]
	Ionized calcium[a]
Contraction	Ionized calcium[a]
	Ionized magnesium[a]
Perfusion	Lactate[a]
Acid-base	pH[a]
	pCO_2[a]
	CO_2 content[a]
	End-tidal CO_2 tension
	Bicarbonate
Osmolality	Measured osmolality
	Calculated osmolality
Hemostasis	Hematocrit[a]
	PT[a], INR
	aPTT[a]
	ACT[a]
	D-dimer, platelet count, and function[a]
Homeostasis	Creatinine[a]
	Urea nitrogen[a]
	WBC[a]
	Chloride[a]
	Inorganic phosphate
	Cardiac injury markers[a] (e.g. myoglobin, CK-MB mass, cTnI)
Pathogen detection	Gram positive bacteria
	Gram negative bacteria
	Fungi
	Viruses (e.g. avian influenza)
	Toxins (e.g. endotoxins and exotoxins)

[a] Point-of-care whole-blood method available.
ACT, activated clotting time; aPTT, activated partial thromboplastin time; CK-MB, creatine kinase-MB isoenzyme; cTnI, cardiac troponin I; INR, international normalized ratio; PT, prothrombin time; WBC, white blood cell count.

The accuracy, precision, and linearity of whole-blood analyzers must be (i) excellent in the near-normal ranges; (ii) acceptable in the high and low extremes of measurements; and (iii) comparable for the same analytes measured on different instruments. The performance of systems for POCT is improving as technologies mature. For biosensor-based systems, performance depends on biosensor stability and reproducibility. Cartridge-based and/or single-use (unit use) systems are dependent on the quality of manufacturing for reproducible results. From a practical standpoint, use of the identical biosensor-based or cartridge-based technology in several clinical sites will facilitate consistency and, in effect, provide intramural standardization. Consistency of measurements in different sites should be demonstrated and also related to the parent laboratory as part of the proficiency-testing program. Measurements obtained with different types of systems (*in vitro*, *ex vivo*, and *in vivo*) should be consistent. As devices become more sophisticated, improved software and hardware should be incorporated for the elimination of artifacts, detection of errors, flagging of inappropriate specimens, and warning of interferences (e.g. hemolysis, bilirubin, and lipids). Small bench instruments are used frequently in satellite laboratories in operating rooms, emergency departments, and intensive care units. Inconsistent results due to instrument-specific normal reference intervals (ranges) or measurement inaccuracies become apparent as unexpected shifts that can confound accurate and swift diagnosis as, for example, the patient is transferred from the emergency department, to the operating room and, subsequently, to the intensive care unit. Significant progress has been made with bar coding and menu-driven touch screens. However, the integration of information at the POC and interlinking of satellite sites are not yet developed adequately. Instruments should be interfaced with computerized laboratory information systems (LISs) or networks. Bidirectional interfacing and electronic archiving are important. Interfacing should be flawless with prompt access to the LIS and the hospital information system (HIS). Systems should conform to emerging standards for the connectivity of POC devices.

Instruments should have the potential for integration with physiological monitoring systems used in intensive care. Data management systems to integrate biochemical and physiological data are invaluable and should be anticipated at the time of instrument selection. Cabling from several bedside instruments and critical care monitors should be consolidated, or patient results communicated to a base station by wireless communication and telemetry. Notification of critical patient results is essential. For example, the risk of miscommunication of actionable or critical results during a fast-paced patient resuscitation is great. The data stream of results flows so quickly that assimilation into the patient record, a legal necessity, often lags behind real-time care and can impair the clinical timeliness of whole-blood tests. Conventional local and wide area networks, fiber-optics networks, electronic mail, facsimile machines, two-way radios, beepers, and cellular phones can help to facilitate communication as needed. This situation calls for the increasing use of robotics, remote review, artificial intelligence, and intelligent decision systems.[70–72]

Several publications are available that explain how to qualify new devices or new tests for clinical service, such as publications from the CLSI.[72] This CLSI document covers experimental design, data examination, linear regression, bias, results interpretation, and performance

criteria. The quality division in each institution should design its own protocols for the qualification of new methods of POCT. For example, a designated reference instrument can be moved on-site (e.g. on a mobile cart) for side-by-side anaerobic split-sample comparisons. Transporting and then analyzing whole-blood samples for comparison studies performed at distant sites may introduce preanalytical errors and is inadvisable. Individual instrument histories, tracking records, and comprehensive quality assurance are essential. Performance of operators can be evaluated using control materials, blind samples, and proficiency testing. Instrument performance can be assessed through quality control, proficiency testing, and pattern-recognition techniques. New approaches to quality validation are needed for single-use (unit use) cartridges and test strips and for *ex vivo* and *in vivo* systems. External and internal electronic checks are being accepted by accreditation agencies but must be validated. Some manufacturers and hospitals have developed on-site programs, videos, and teaching aids to educate operators and explain new approaches to quality management and method comparison. Note that education and training programs must satisfy federal, state, and professional accreditation requirements. A quality alliance of laboratory personnel, manufacturers, and regulators is essential to transform laws, rules, and inspections into an integrated quality-based system.

3 CRITICAL-EMERGENCY-DISASTER CARE

The Southeast Asian Tsunami of 2004 and Hurricane Katrina in 2005 revealed the lack of disaster medical preparedness worldwide. POCT was numerous and the disaster response was massive, however, as the results showed, this was quite insufficient. During these two recent disasters, hospitals, roadways, and communications were quickly compromised. POCT and other medical instruments no longer function under these adverse conditions. Ultimately, these recent disasters have shown the current weakness of POCT, but revealed the potential of POCT to play a significant role to augment the damaged, destroyed or overwhelmed healthcare infrastructure.

POC in Thailand and the Asian Tsunami. On Phi Phi Island in Thailand following the tsunami, diagnostic testing was limited to glucose meters, spun hematocrit, and urine dipsticks for protein and glucose. In the costal areas of Thailand, frequent cases of aspiration were assessed for severity by POC pulse oximetry. No quantitative cardiac biomarker testing was available.[47]

POC in New Orleans and Hurricane Katrina. In New Orleans, following Hurricane Katrina, the US disaster response was massive, but inadequate to deal with the loss of 11 Louisiana hospitals.[47] The US military, National Disaster Medical System (NDMS), and Disaster Medical Assistant Teams (DMAT) from across the country responded with mobile hospitals equipped with limited POCT instruments. Military POC instruments included handheld or portable whole-blood analyzers, ABO blood typing, glucose meters, cardiac biomarkers, pulse oximeters, and drugs of abuse tests. POCT for DMATs were limited to handheld whole-blood analyzers, pulse oximeters, and glucose meters. Both the military and DMATs are the quintessential platforms to deploy POCT for any current or future disaster response.

In both disasters, environmental extremes hampered diagnostic medical devices. Flooding, no clean water, 32.22–37.78 °C temperature, 100% humidity, power loss, and destruction of cell and landline phone systems crippled the medical infrastructure.[47] This beckons the question, if POCT can provide results under these conditions. A study by Sumner et al. showed common POC instruments including glucose meters and a handheld whole-blood analyzer to provide erroneous results under extreme conditions. In some cases, instruments failed to operate.[48] GMSs showed sensitivity to high and low temperatures. The test cartridges for handheld blood gas analyzers that are carried by DMATs are affected by high and low temperatures (Figure 10).[48] Some suggest that these instruments should be operated in close proximity to portable air conditioning units.[73]

There is a clear need for POCT instruments for field medicine, and disaster response. Disaster response POCT should be designed with DMAT use in mind and accompanied by educational programs.[73] POC instruments and reagents should be field robust (e.g. vibration, humidity, temperature, salinity, barometric pressure, and exposure). Instruments should be standardized for first responders and DMATs with interchangeability of equipment and common training. Connectivity of instruments is also essential to provide wireless networking capabilities during disaster response management (e.g. patient tracking) and potential integration into telemedicine systems.[74] Health care personnel must be aware of the potential for aberrant results when current POC devices are stressed beyond their normal operating parameters.

4 TEST CLUSTERS (CRITICAL CARE PROFILE) AND MEDICAL INDICATIONS

The innovation of a critical care profile (Table 8) further minimized the response time for whole-blood analyses.[2,70,75–77] Fast TTAT of 2–5 min was necessary

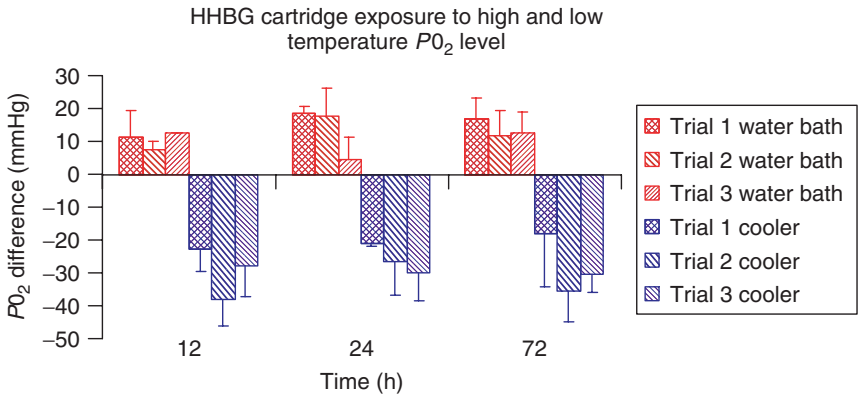

Figure 10 Effects of heating and cooling of test cartridges on pO_2 results obtained with a handheld blood gas analyzer. This figure shows the mean pO_2 paired-differences between heated (40 °C) or cooled (2 °C) test cartridges and control (21 °C). Cartridges were thermally stressed for 12–72 h.

during liver transplantation.[21] Transplant procedures (and massive transfusions) demonstrated the significance of ionized calcium, the time value of diagnostic test results, and the efficacy of the critical care profile. Historically, an important milestone occurred when the critical care profile combined whole-blood ionized calcium (Ca^{2+}) and electrolytes (K^+, Na^+, and Cl^-) with blood gases (pO_2 and pCO_2), pH, and hematocrit on one instrument platform. Now biosensor-based whole-blood measurements in critical care clusters include ionized magnesium, glucose, lactate, urea nitrogen, creatinine, and CO_2 content.[9] Several other medically important critical care tests, combined strategies, and novel approaches are available, or soon will be. User-definable and patient-focused test clusters provide valuable indicators of vital functions, such as those listed in Table 9.

Organ injury markers, such as markers of myocardial injury, and hemostasis–thrombolysis tests (see below), are important recent additions to the POC armamentarium. Each test added provides the physician with extra flexibility and improves the cluster efficiency significantly. In the future, the number and variety of whole-blood tests available at the point of care will expand dramatically as microfabrication/nanofabrication advance, chip-based technologies mature, and the critical care marketplace allows. Availability of a wide range of diagnostic pivots in appropriately selected test clusters expands the spectrum of clinical indications for POCT. These indications now are relatively broad and include critical conditions such as cardiopulmonary arrest, endocrine crises (e.g. diabetic ketoacidosis and hyperosmolar hyperglycemic nonketotic diabetes with coma), pulmonary problems (adult respiratory distress syndrome, ARDS), pulmonary embolism, and respiratory failure), hemorrhage (e.g. acute gastrointestinal or postsurgical bleeding), seizures, and numerous others. Hence, vital functions and their diagnostic pivots form the criteria for the selection of patient-focused test clusters and critical care profiles.

Table 9 summarizes the indications, applications, and limitations of POC tests for hemostasis, thrombolysis, and thrombosis. POC tests such as hemoglobin, hematocrit, and white blood cell count (WBC) are used in several clinical domains. Other tests in this area have relatively specific indications. Patient-focused testing for PT to monitor the status of ambulatory patients on coumadin therapy has been used effectively. In critical care, activated partial thromboplastin time (aPTT) and activated clotting time (ACT), historically an early on-site test used in the operating room, are frequently performed at the POC to monitor heparin anticoagulation, protamine neutralization, and the patient's status before, during, and following cardiac surgery, invasive procedures (such as coronary catheterization), and dialysis. PT and aPTT are valuable indicators in trauma patients who have head injury associated with unconsciousness or focal neurologic deficit, and therefore delayed intracranial hemorrhage is suspected.

This area is one of critical need and rapid expansion. Established large companies have recently invested heavily in research and development. Several new methods that assess platelet function, thrombosis, and thrombolysis are under development, in clinical trials, or moving to the POC. Some tests will be used to predict adverse responses to thrombolytics. Tests for clotting (e.g. TpP, thrombosis precursor protein), if clinically validated, may be useful in clusters with metabolite tests (e.g. lactate) to detect coronary artery thrombosis. These test clusters could complement cardiac injury markers, such as CK-MB, myoglobin, and cTnI, an important "triplet" cluster. Clearly, tying together key clusters of tests is a valuable benefit of POCT and represents the

Table 9 Hemostasis, thrombolysis, and thrombosis tests for the point of care

Diagnostic pivot(s)	Indications, applications, and limitations
ACT (ACT-Plus[a,b,c])	Surgery, invasive procedures, hemodialysis, arterio- and venovenous dialysis
ACT ≥ target threshold (e.g. 400 s)	CPB, PTCA (high heparin levels, strong contact activator in reagents)
ACT ≤ target threshold	ECMO, stents, catheter sheath removal, pharmacological agents[d] (low heparin levels, moderate activator)
aPTT	Heparin anticoagulation monitoring–indications listed above (low heparin levels), intrinsic coagulation screening and monitoring treatment (e.g. hemophilia A)
Clot signature analysis (TEG,[e,c] others[f]	Transplantation, CPB, protracted surgery
D-dimer (XL-FDP)	Deep vein thrombosis, PE, DIC, preeclampsia, reactive fibrinolysis (qualitative point-of-care, quantitative immunoassay devices)
Fibrinogen	Preanalytical processing required; under development for whole blood at point of care[g]; thrombolytic therapy; 1° (congenital) and 2° (acquired consumptive coagulopathy-DIC, liver disease) hypofibrinogenemia; after surgery, ECMO
Heparin dose-response	Improved precision in monitoring anticoagulation and relative insensitivity to hemodilution and hypothermia; various heparin management systems/panels[h]
Platelet count and function	Transfusion therapy, drug effects, screening for acquired and congenital platelet problems
PT (INR)	Coumadin anticoagulation monitoring, extrinsic coagulation screening – bedside, clinics, physician office laboratory, home monitoring, other settings
Thrombin time (HNTT,[i,b,c] HiTT[i,b,c])	Indirect fibrinogen function, heparin therapy, DIC status; rarely used at the bedside
Thrombolysis (fibrinolysis)	Determining or predicting response to thrombolytics, tracking lytic state; tests for thrombolytics (e.g. streptokinase, TPA, urokinase) and lysis onset time (LOT[g,c]) are under development or in clinical trials
Thrombosis (TpP[j,c])	Suggested in acute myocardial infarction, PE, stroke; clinical validation and point-of-care test pending

CPB, cardiopulmonary bypass surgery; DIC, disseminated intravascular coagulation; ECMO, extracorporeal membrane oxygenation; PE, pulmonary embolism; PTCA, percutaneous transluminal coronary angioplasty; TPA, tissue plasminogen activator; XL-FDP, cross-linked fibrin degradation products.

[a] Used in cases of hypothermia, hemodilution, and antifibrinolytic drugs in CPB and cardiac catheterization.

[b] International Technidyne Corporation, Edison, NJ.

[c] Proprietary.

[d] Thrombin inhibitors and platelet IIa–IIIb blockers.

[e] TEG, thromboelastography (Haemoscope, Morton Grove, IL).

[f] Sonoclot (Sienco, Denver, CO) and Xylum (Scarsdale, NY).

[g] Cardiovascular Diagnostics Incorporated, Raleigh, NC.

[h] 1, Heparin Management Test;[g] 2, Hepcon Hemostasis Management System (Medtronic, Parker, CO); 3, RxDx Anticoagulation Management System.[b]

[i] HNTT, fibrinogen function in the presence of heparin; HiTT, CPB, high heparin levels and antifibrinolytics.

[j] TpP, thrombosis precursor protein (American Biogenetic Sciences, Copiague, NY).

precursor of multimarker risk indexing for specific disease and diagnoses.

For coagulation testing, special preanalytical problems may arise due to specimen acquisition and sample handling. To avoid preanalytical problems, blood samples obtained from existing lines or venipunctures must be collected carefully. Clotting and hemolysis, which can also affect electrolyte results (e.g. K^+), must be avoided. In fact, an ideal POC device would not only detect but also flag specimen problems and disallow inaccurate results. Proper sample collection requires training and experience. Preanalytical errors occur, for example, if there are delays, since blood begins clotting immediately after removal from the body, or when a citrated tube is used for anticoagulation but the tube is not properly filled with blood and the sample is not mixed adequately. Additional problems arise in the elderly, who may have poor peripheral circulation, and in premature infants, who have small blood volume and do not tolerate excessive numbers of heelsticks. Nonanticoagulated whole blood obtained by fingerstick can create artifacts in coagulation tests, change over time due to instability, and present biohazards for the operator. Clinicians must be aware of the normal and therapeutic ranges for POC test results used in clinical decisions. Therefore, it is vital to carefully evaluate (i) the preanalytical steps required; (ii) whether the blood sample is anticoagulated or not; (iii) the POC test method used; and (iv) the relationship to results obtained with either whole-blood or plasma-based methods in the parent laboratory (main laboratory) when implementing POC tests for hemostasis, thrombosis, and thrombolysis in individual critical care settings.

5 ACCURACY ASSESSMENT

Numerous methods are available to determine the accuracy of POC instruments. These include national and international standards (e.g. CLSI and ISO), as well as pattern recognition and new statistical methods to determine if POC devices are comparable to reference instruments. Pattern-recognition methods, such as mountain plots, scatter plots, median absolute difference (AD) plots, and Bland–Altman plots, can be used. Illustrative methods such as the Bland–Altman plot may be coupled to standards like the CLSI and International Organization for Standardization (ISO) criteria.[78,79] For statistical methods, examples include the Student's *t*-test, analysis of variance (ANOVA), and least squares linear regression. This section also describes the most recent approach, locally smoothed median AD curves.

5.1 Pattern-recognition Methods

These methods allow for quick determination of biases, trends, and errors. The mountain plot, also called an *empirical cumulative distribution plot*, allows for easy identification of the central 95% of the data and provides clearer differentiation of two distributions (e.g. GMS versus reference).[79] It is generated by computing the percentile for each ranked difference between the device versus the reference. The plot shows the median bias between the two methods, while the tails of the plot show the propensity for the new method to deviate from the comparison method.

POC device versus reference instrument scatter plots are commonly used for linear regression. In glucose monitoring accuracy assessment, an error grid can be applied to the regression plot. Error grids such as the Clarke Error Grid divides a scatter plot (glucose meter versus reference) into five zones (A, B, C, D, and E).[80] Data points falling into zones A and B are considered acceptable, while data falling into zone C indicate that the glucose meter's results may prompt unnecessary corrections. Zone D and E represent dangerous failure to detect a glucose level or results causing erroneous treatment, respectively. More recently, there has been the introduction of the continuous glucose error grid that allows a user to evaluate the accuracy of continuous glucose monitors. The five zones are retained; however, the shape of the error grid allows the investigator to also evaluate the glucose rate (e.g. blood glucose as a function of time).[81]

Median AD curves provide a nonparametric method at illustrating performance data. Medians are calculated from the dataset for each reference value. Data points are plotted on a median AD (y axis) versus reference (x axis) curve. The curve is not biased by extreme values due to its reliance on medians, and does not require a normally distributed data set.

The modified Bland–Altman plot is a bias plot (bias versus reference).[82] Bias (POC instrument–Reference) is on the y axis, and the x axis is the reference. This method is useful in that it shows bias, trends, and errors. When coupled to standards (e.g. CLSI, ISO, American Diabetes Association), these plots serve to identify if an instrument meets the acceptance criteria. For example, a recent draft FDA guideline proposes the use of the ISO 15197 criteria integrated with the modified Bland–Altman plot for glucose meter accuracy assessment.[83] In situations where a reference is not available, the bias can be calculated between two instruments (y value) and compared against the average value between the instruments (x value).

5.2 Locally Smoothed Median Absolute Difference Curves

Locally smoothed median absolute difference (LS-MAD) is a nonparametric method of transforming discrete points on an $x-y$ plane into a continuous curve that helps reveal underlying patterns.[82] This method provides a representation of instrument performance versus a reference through the entire range of testing. The curve contains points (x, y) where y is the median of the values, y^*, for all points (x^*, y^*) in the range $[x-h] \leq y^* \leq [x+h]$. The bandwidth, h, controls the degree of smoothing.

Values are calculated from integer values only. Each value at point x is the median of the AD of all paired instrument and reference measurements where the reference result is within the specified distance, the bandwidth of x. The discrete median points are then connected to generate a single continuous curve. Ideally, LS-MAD curves should be as close as possible to the x axis (low MAD), and be relatively flat.

Figure 11 illustrates the LS-MAD methodology applied to glucose meter versus a reference analysis. ADs are calculated, and a median AD is generated at each reference value based on a bandwidth of 15 mg dL^{-1}. This bandwidth, h, of 15 mg dL^{-1}, where $2h = 30$ mg dL^{-1}, was selected to correspond to a TGC interval of 80–110 mg dL^{-1}.

Improved performance is achievable by correcting for confounding factors such as hematocrit, pO_2, and the effects of drug-oxidizing drugs. The StatStrip GMS (Nova Biomedical, Waltham, MA) is one such example. Recent multicenter evaluations demonstrate the new StatStrip GMS, a sample size of 2767 paired observations having higher accuracy and precision when compared to competing systems. The LS-MAD curve for the StatStrip remained below the 5 mg dL^{-1} tolerance from 35 to 179 mg dL^{-1} (Figure 12a).[84] Figure 12(b) and

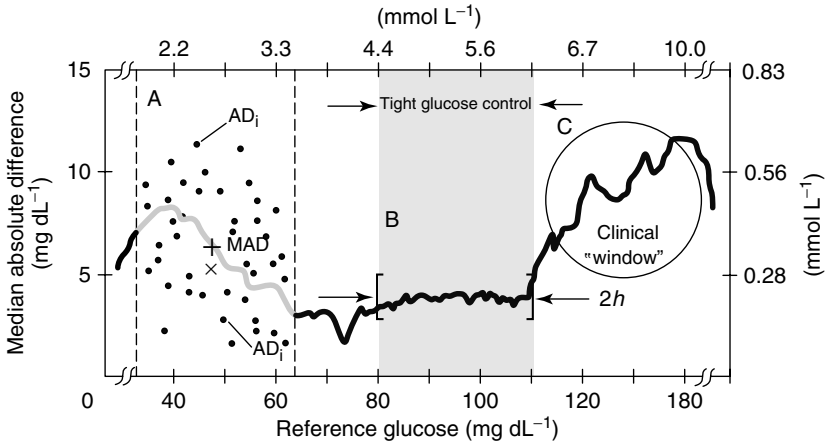

Figure 11 The approach to generating and interpreting an LS-MAD curve is illustrated from left to right in three separate sequential frames. Frame A illustrates the calculation of the LS-MAD y value (y^*) at the point x (see the crosshair). The LS-MAD curve (shadowed) smoothes the variation in the underlying absolute difference (AD) scatterplot; y^* represents the median of the ADs (AD1, AD2, through ADn) for points between $x - h$ and $x + h$, where h is the bandwidth (15 mg dL^{-1}). Frame B shows that twice ($2h$) the bandwidth (h) matches the span of the TGC interval (30 mg dL^{-1}) that we selected on the basis of its efficacy for improving outcomes (15–17). Frame C emphasizes that the LS-MAD curve will instantly reveal regions, such as the high glucose range, where bedside testing performance needs to be improved significantly.

(c) illustrate (e.g. modified Bland–Altman plot and least squares linear regression) methods. We defined erroneous values as values falling outside of ISO 15197 tolerances. Competing meters, which did not correct for confounding factors, performed poorly under LS-MAD analysis and exhibited a high frequency of erroneous values.[85]

5.3 Discrepant Values

Discrepant values represent errors that could significantly impact the effectiveness of TGC protocols.[54] Two classes of discrepant values exist, Class I and Class II. Class I discrepancies are pairs with reference values below the lower TGC limit (e.g. <80 mg dL^{-1}) and GMS values above the high TGC limit (e.g. >110 mg dL^{-1}). Class II discrepancies are pairs with reference values above the high TGC limit and GMS values below the lower TGC limit. Class I discrepancies represent the most serious type of error since they can lead to dangerous clinical decisions and, in the context of TGC, hypoglycemia. Conversely, class II discrepancies are less serious since they could only lead to aggravation of hyperglycemia. The concept of discrepant values can be applied to any POC analyte for which simpler diagnostic-therapeutic decision making is made at the bedside.

5.4 Statistical Methods

Statistical tests can provide p values or correlation coefficients. The most common statistical test is the Student's t-test.[37] Three kinds of t-test's exist: one sample, two sample, and the t-test for paired differences. For the purposes of comparing a GMS versus a reference analyzer, we must use the Student's t-test for paired differences since we are comparing the same blood sample. The Student's t-test for paired differences assumes that the samples are not independent (e.g. using the same blood samples), and the data is normally distributed. ANOVA is another statistical method, which compares the differences of three or more groups of data. This is very useful for comparing variations between different test strip lots. Both the Student's t-test and ANOVA generates a p value where $p \leq 0.05$ is usually considered statistically significant.[86]

Ordinary least squares linear regression, or commonly referred to as *linear regression*, generates a best-fit line onto a scatter plot (POC instrument versus reference analyzer).[37] The correlation coefficient (r), coefficient of determination (r^2), and equation of the line are generated on this plot. The r^2 value is commonly used, with r^2 ranging from 0 to 1, where 0 indicates a nonlinear relationship, and 1 being a perfect fit. Manufactures strive to attain a very high r^2 value since it indicates high correlation between GMS and the reference. However, it must be noted that least squares linear regression is very susceptible to the weighting effects of data points at extreme values. For example, if a single value at a high reference range falls on the regression line, it may provide a high r^2 value. Removal of this data point may then reveal the less than satisfactory nature of data set. Therefore, investigators must be aware of the potential for data to have poor agreement but still produce relatively high correlations.

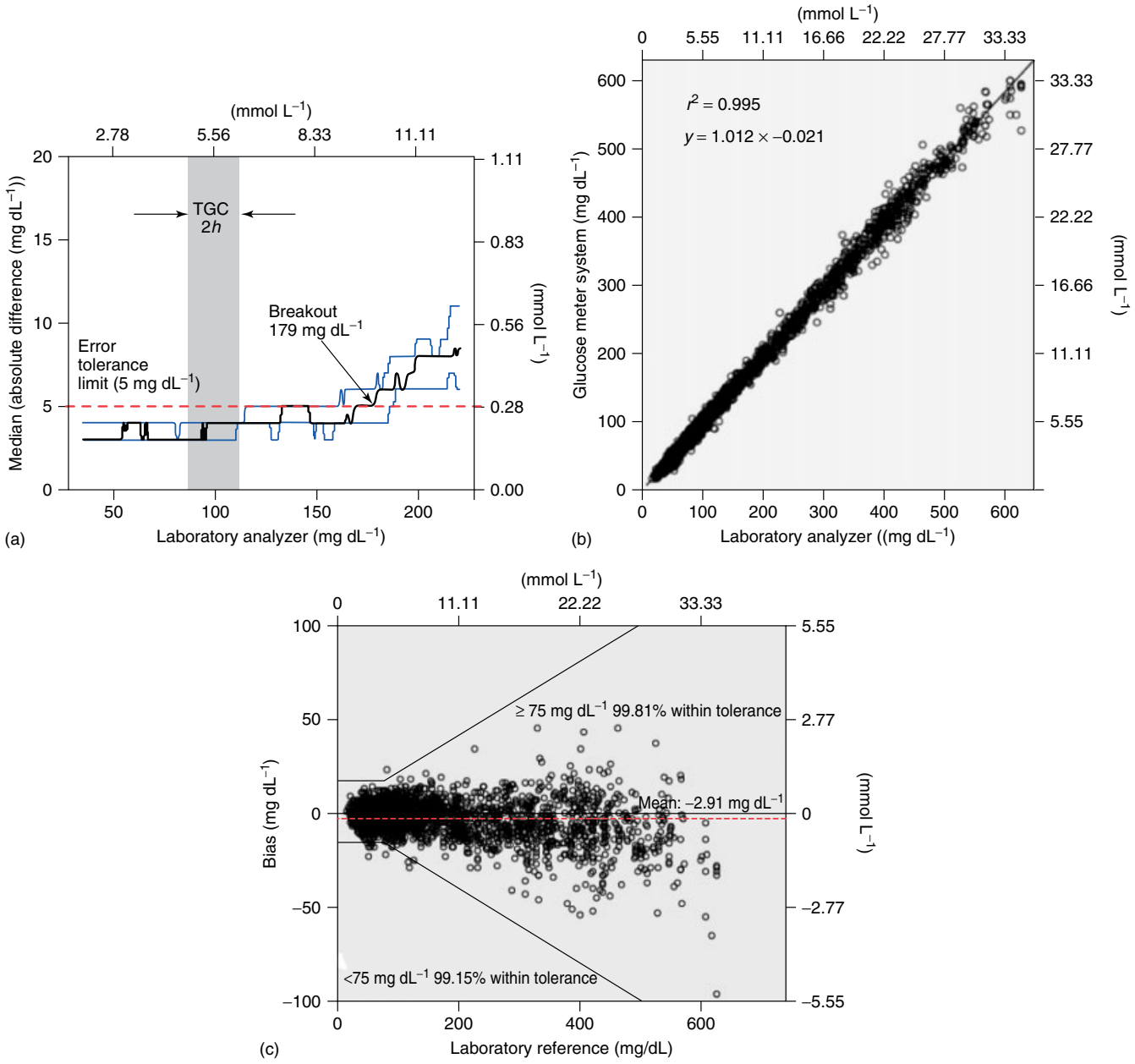

Figure 12 (a) Locally smoothed median absolute difference curve ($n = 2767$ paired observations). This plot shows an LS-MAD curve and the 95% confidence intervals for a glucose meter that automatically adjusts for hematocrit effects when compared to a laboratory analyzer (120 different analyzers). The TGC range (80–110 mg dL^{-1}) is shown in gray, and the arrow identifies where the LS-MAD curve breaks out of the 5 mg dL^{-1} error tolerance at a reference value of 179 mg dL^{-1}. (b) Least squared linear regression plot ($n = 2767$). This plot shows a least squares linear regression plot. $r^2 = 0.995$ and $y = 1.012x - 0.021$. (c) Modified Bland–Altman ISO 15197 integrated plot ($n = 2767$). The ISO criteria requires that 95% of data points less than 75 mg dL^{-1} have a bias of no more than ±15 mg dL^{-1}, while for 95% of data points equal or greater than 75 mg dL^{-1}, the bias must fall within 20% of reference values. When compared to 120 different laboratory analyzers, this glucose meter has 99.15% (581/586) of data points falling within tolerances for reference values <75 mg dL^{-1} and 99.81% (2177/2181) of data points ≥75 mg dL^{-1} falling within tolerances. The mean bias was −2.19 mg dL^{-1} (SD ±9.61, $P < 0.001$).

5.5 Future Concerns

There is a clear need for POC instruments, which account for confounding factors. POC instruments affected by confounding factors may generate erroneous results and discrepant values that lead to adverse treatment decisions, a key concern in settings such as TGC for critically ill patients.[54] Recently, there has been a substantial effort

to account for confounding factors such as hematocrit. Results have shown improved performance over clinically significant ranges. As biosensor technology continues to improve, future instruments should be evaluated on the basis of more stringent performance criteria and unbiased statistical methods to determine accuracy. Additionally, POCT accuracy should be based on absolute global standards, rather than on relative performance and equivalent methods.

6 MONITORING PERFORMANCE

6.1 Total Quality Principle for Point-of-care Testing

Complex critical care settings warrant careful consideration of strategies for enhancing performance. Good strategies transcend statutory requirements. Therefore, the goal of this section is to introduce, as a guiding concept, the *total quality principle* for POCT, and then to discuss the general approach to accreditation, to outline regulatory requirements for POCT programs, to describe how to enhance performance by avoiding common inspection pitfalls and by improving training and education programs, and to identify the role of the clinical laboratory in supporting critical care testing. It may be noted that regulatory requirements are perpetually changing and, hopefully, continuously improving. Accreditation agencies can be contacted via their home pages on the World Wide Web, or directly for the latest information. For general guidelines that reflect the standard of care for POCT, one may refer to Kost.[1]

The Clinical Laboratory Improvement Amendments of 1988[87] (CLIA'88) imposed graded regulatory requirements based on the level of test complexity (the "complexity model") rather than on the location of diagnostic testing, the so-called site neutrality principle. Test categories include waived (e.g. nonautomated spun hematocrit, gastric and fecal occult blood tests, and pH in body fluids such as amniotic fluid and gastric aspirates), moderate complexity (e.g. Ca^{2+}, blood gases, microscopy, and accurate and precise technology), and high complexity. The moderate complexity category includes most diagnostic tests. For a comprehensive discussion of CLIA'88 test categories, personnel descriptions, and regulatory requirements for decentralized testing, please see the summary[88] by authorities, who state, "to know CLIA is to know the other regulations – all of the other regulations are based on CLIA'88".[24] The World Wide Web home page of the FDA updates CLIA-waived tests (see Appendix). The home page of the American Association for Clinical Chemistry offers CLIA information ("CLIANET"). CLIA'88 established the requirement for site neutrality and, in doing so, defined a common denominator of the regulatory process. Revised CLIA rules are anticipated, but the principle of site neutrality is likely to remain a permanent feature in the future.

On the basis of the concept of site neutrality, then, the *total quality principle* for POCT means empowering professionals who proactively integrate quality control, quality monitors, and proficiency testing into one patient-focused package that meets customer needs *irrespective of where the diagnostic testing is performed*. The overall goal is to improve the integrated outcomes of the diagnostic testing process. The practical strategy is total quality management. As part of the total quality management, each institution must design its own performance enhancement program. Performance is enhanced by establishing quality monitors, perpetually assessing trends, and continuously devising new ways to improve quality. Continuous quality improvement is an important objective that all accreditation agencies promote. Another important objective that accreditation agencies share is the demonstration of satisfactory proficiency-testing results (see below). Proficiency testing objectively determines the overall quality and integrity of diagnostic testing services.

6.2 Accreditation and Regulation

6.2.1 Joint Commission on Accreditation of Healthcare Organizations

The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) is a voluntary accreditation agency. The Health Care Financing Administration (HCFA) has determined that JCAHO accreditation standards are equivalent to or exceed those of CLIA'88. Laboratories accredited by JCAHO are deemed to have met CLIA'88 requirements. Therefore, it is important to understand JCAHO requirements for POCT.[89–92] In contrast to the lack of regulatory requirements for tests waived under CLIA'88, JCAHO has several relatively stringent requirements. These requirements include, for example, definition of medical use, identification of operators and supervisors, written procedures (including specimen collection and preservation, equipment performance evaluation, test performance, instrument calibration, and quality control and remedial actions), quality control checks that meet at least the manufacturer's minimum recommendations, and long-term maintenance of patient testing and quality control records. Training and continued competence must be documented by means of split-sample unknowns, observation of technique, or acceptable performance on quality control checks. For both waived and moderately complex tests, it is particularly important to document who is responsible for testing

and what measures are taken when correcting quality control failures or other problems.

Under CLIA'88, most POC tests used in the operating room and intensive care unit fall into the moderate complexity category. Like CLIA'88, for these tests, JCAHO requires compliance with manufacturers' instructions, a procedure manual, periodic calibrations, written control procedures, two-level quality control checks each testing day (or more frequently depending on the type of test), documented remedial actions, and two-year record keeping. JCAHO currently accepts electronic checks, but recommends that liquid controls also be run periodically to help confirm electronic control results. JCAHO also demands evidence that the performance of the method is verified prior to use at the POC. This evidence includes the demonstration and documentation of accuracy (through comparison of split samples, analysis of controls, or proficiency testing), precision (through replicate analysis of patient specimens or of control materials from one lot), reference intervals suitable for the patient population, and validation of high and low values (through linearity studies or comparisons to other instruments). JCAHO requires participation in a proficiency-testing program (except for waived tests) and demonstration of the consistency (correlation) of test results obtained at different POC sites, whether or not the sites participate directly in the proficiency-testing surveys of the parent and/or main laboratory. However, only one methodology or analyzer per analyte (under each CLIA certificate) needs be enrolled in proficiency testing. Corrective actions for proficiency-testing failures must be taken promptly and carefully documented.

For hospitals lacking a consolidated POCT program, vulnerability occurs where quality monitoring activities are not integrated. That is, during inspections, JCAHO may require documentation showing that the loop of total quality management (consisting of instruments, quality monitors, and patient test results) was closed during the same time interval when the audited test results were used in patient management. Typically, POCT sites are widely distributed, and often instrument interfaces to hospital and LISs are limited. Therefore, seamless linkage of patient test results and quality monitor results to the electronic medical record is a major advantage when satisfying JCAHO requests for these audit trails. Institutions that have implemented electronic patient record systems and instrument interfacing are well positioned to track and manage the quality of POCT. JCAHO recently required hospitals to implement one or more performance measurement systems that collect data and report results to JCAHO quarterly, two clinical measures in the form of metrics that relate to more than 20% of the hospital's patient population, and peer group comparisons for both the clinical laboratory and the hospital as a whole.

6.2.2 *College of American Pathologists*

Like JCAHO, the College of American Pathologists (CAP) is a voluntary accreditation agency. The CAP laboratory accreditation program has also received deemed status from HCFA. Additionally, JCAHO will accept CAP accreditation in lieu of a detailed laboratory inspection. The standards of the CAP laboratory accreditation program are the most demanding of any of the accreditation agencies.[93] CAP treats all tests, including those performed at the POC, as highly complex under CLIA'88. A physician or doctoral scientist must be responsible for POCT. When CAP inspects and accredits POCT, bundling it with the clinical laboratory under one CLIA'88 registration number (certificate) and one director is most efficient. Consolidation is also advantageous since fewer directors, licenses, and inspections lower fees and reduce the overheads for POCT programs. The reader is encouraged to consult with CAP directly for details of general requirements and to review the following three checklists: (i) the limited services checklist (no. 25) that evaluates satellite and stat laboratories; (ii) the blood gas laboratory checklist (no. 26) that evaluates sites performing blood gas and whole-blood analyte tests; and (iii) the POCT checklist (no. 30) that evaluates testing performed with handheld or portable instruments that are carried to the bedside and do not require permanent dedicated space. Checklist no. 30, which previously was titled *alternate site testing,* covers only a subset of POCT services. Hence, the other checklists must also be reviewed.

Notable features of the CAP POCT checklist (no. 30) include (i) the strict verification of performance (e.g. accuracy, precision, analytical sensitivity and specificity, reportable range, reference interval, and calibration verification and/or linearity); (ii) a thorough continuous quality improvement program, with quality control required even for tests waived under CLIA'88; (iii) the option to use electronic quality control, that is, ''a scientifically acceptable alternative control system . . . that controls the entire analytic process'' (question 30.0550); and (iv) documented weekly review of quality control by a technical supervisor, plus secondary monthly review by a director or designee; and (v) participation in the CAP proficiency-testing program, processing of survey samples ''in a similar manner as patient specimens'', and analysis by personnel who routinely test patient samples (Q30.0325).

There must be defined criteria for correlating unexpected test results with other clinical findings to validate results (Q30.0370). There is also a question intended

"to minimize unnecessarily large blood draw volume" (Q30.0435) and others (Q30.0455, Q30.0460) to ensure that critical limits are established, critical results are notified, and notifications are documented, unless test results are given directly to the physician. Near-patient testing of blood gases and whole-blood analytes in satellite laboratories is covered by checklist no. 26, if the near-patient sites have separate administration and directorship but are inspected by CAP in parallel with the main laboratory. Near-patient laboratories are common in critical care.[75,76] Therefore, near-patient and POCT should be carefully coordinated and integrated in order to streamline accreditation, improve cost effectiveness, enhance performance, and benefit patient care.

6.3 Performance Enhancement and Continuous Quality Improvement

6.3.1 Avoiding Common Pitfalls of Accreditation Inspections

Common inspection deficiencies under CLIA'88 include (i) weak definition of responsibilities, poor documentation of competency, and ineffective education and training programs; (ii) noncompliance with manufacturers' directions and lack of procedure manuals; (iii) inadequate quality control, quality assurance, proficiency testing, and total quality assessment; and (iv) absence of periodic calibrations and verification of the accuracy of analytes omitted from proficiency testing. During JCAHO inspections, commonly encountered deficiencies include (i) absence of test and control records for methods in the CLIA'88 waived category; (ii) lack of review and remedial actions for proficiency-testing problems with moderately and highly complex tests; (iii) inadequate surveillance of test results and quality monitors; (iv) insufficient compliance with laboratory safety requirements; and (v) various problems with fulfillment in technical areas such as analytical evaluation, preventative maintenance, and adequate space. Note that United States Occupational Safety and Health Administration (USOSHA) requirements (e.g. universal precautions) must be met. Also, always consult the manufacturer to check that the device was cleared by the FDA for POCT and to determine the test complexity category.

CAP checklists are "tools for inspectors and directors to use in evaluating whether or not the laboratory is meeting the standards for laboratory accreditation".[93] On the basis of the CAP POCT checklist (no. 30), common inspection deficiencies encountered are associated with (i) reviewing results (supervisor), protocols, and quality control; (ii) pursuing corrections for quality control results out of tolerance limits; (iii) performing positive and negative controls for qualitative tests; (iv) documenting the quality control program and reference intervals; (v) detecting errors and unusual or unexpected test results; (vi) periodically verifying linearity; (vii) labeling reagents; and (viii) participating in proficiency testing. Deficiencies also arise when proficiency testing is not done by testing personnel or when no corrective actions are taken following proficiency-testing failures.

Analysis of these common CLIA'88, JCAHO, and CAP inspection deficiencies reveals the importance of the total quality principle and continuous quality improvement for POCT. Many of the common pitfalls of accreditation inspections can be avoided through adequate preparation, ongoing supervision, and common sense. For example, testing controls, recording results, remedying problems, and reviewing performance are essential to avoid adverse events associated with the diagnostic testing process, irrespective of where the testing is performed. Proficiency testing is designed to reflect the overall quality of the POCT program and to improve it. Proficiency testing also can be used to assess personnel competency, even, as suggested by JCAHO, for tests waived from this requirement under CLIA'88. Total quality management helps assure that the integrated outcomes of the quality monitoring process reflect the level of performance expected when test results are actually used in clinical decisions.

Hospitals can benefit medically and economically from consolidating total quality management in the hands of one unified group of interdisciplinary "hybrid" personnel.[3,72] The expertise of this multidisciplinary team should span the spectrum from intensive care to laboratory medicine. Laboratorians are professionally trained and experienced in the total quality principle. Their scope is the entire health care system. They can be vested with both authority and responsibility for the design of the quality assurance and proficiency-testing program. Laboratorians can perform important tasks, such as evaluating technology, correlating methods, defining normal ranges, writing protocols, managing instruments, coordinating supplies, providing backup, addressing problems, improving safety, demonstrating technique, and integrating services. They are skilled in computerized informatics for quality control and patient test results tracking. They also can oversee and document the training, education, and certification of clinicians who perform POCT. When part of the performance enhancement team, experience has shown that laboratorians who bridge disciplines and manage quality programs work well with clinicians and accreditation agencies to achieve and, typically, surpass the level of performance required for all types of POCT.

6.3.2 Clinical Laboratory Standards Institute Guidelines

CLSI guidelines emphasize important program objectives, such as performance of instrument maintenance and function checks, patient test performance, recording and reporting of patient results, assessment of problem-solving skills, verification of knowledge by examination, and use of operator proficiency in personnel evaluations.[93,94] It is important that specimen acquisition be performed safely with positive patient identification. Sample transfer should be performed properly to avoid preanalytical errors. Quality control should be rotated among operators who should document quality control results and corrections following quality control failures. Authority to qualify personnel, assure that quality control standards are maintained, and prohibit inappropriate use of POC instruments is pivotal to successful performance. Thus, CLSI guidelines call for a thorough understanding of the principles and practice of POCT. The reader is encouraged to contact the CLSI directly for emerging guidelines that are being developed, undergoing revisions, or pending approval; please send e-mail (exoffice@CLSI.org) or use the CLSI World Wide Web home page (http://www.CLSI.org).

7 STRATEGIES FOR OPTIMIZATION OF POINT-OF-CARE TESTING

Table 10 outlines general strategies to optimize laboratory diagnostic support for critical care and other areas. These include establishing a critical care profile, selecting instruments designed for critical care, minimizing response time, and targeting critical opportunities.

Table 10 Strategies to optimize POCT

Establish critical care/patient-focused care profiles
• Fulfill clinical objectives by selecting key tests
• Combine tests in clusters to improve efficiency
• Monitor vital functions using test clusters
Select instruments designed for POCT
• Use direct whole-blood measurement
• Eliminate centrifugation
• Conserve blood
Minimize response time and maximize connectivity
• Create patient proximity
• Reduce specimen transit time
• Report results immediately
Target critical opportunities
• Identify surrogates of sublethal and lethal events
• Track quantitative metrics
• Focus on medical and economic outcomes

A critical care profile should fulfill clinical objectives, combine tests in clusters to improve efficiency, and monitor vital functions. A technological and clinical hurdle for POC instruments is limited flexibility in the selection of tests or test clusters. POC devices are progressing through a developmental stage limited by test subsets. That is, the subsets are not necessarily patient focused. Test menus are limited relative to actual needs in critical care medicine and other disciplines. Current technologies allow the purchase of a defined test menu on a transportable or portable platform, the use of personal testing devices (PTDs) with a few tests or fixed-menu cartridges, or for remote review systems, the selection of analytes in a ''minimenu'' based on limited methods available at a near-patient workstation (automated testing machine (ATM), or automated testing instrument (ATI)). Storage of excess expiration-dated electrodes, cartridges, or reagents that may go to waste should be avoided. Ideally, a physician should be able to select individual tests or clusters of tests needed at the time of decision making, choose from a menu of at least 30–50 analytes, and pay only for those tests actually used.

Given current technological limits, rapid device diffusion, and significant fiscal constraints, it is important to use direct whole-blood measurement, to eliminate centrifugation and other preanalytical processing steps, and to conserve blood in order to eliminate unnecessary blood transfusions and their hidden and downstream risks. Handheld devices typically require several preanalytical and postanalytical processing steps. Extra labor and supplies costs cannot be ignored. If assessed objectively from a cost-accounting perspective, numerous small processing steps performed by individual caregivers add up to large financial burdens for the integrated health system. In addition, biohazard risks must be taken into consideration. Since blood specimens are handled manually and often without containment, serious biohazard exposures can occur. Staff members are at risk of infection. Critically ill patients may be in isolation, laminar flow, or other special environments that call for biohazard containment during diagnostic testing procedures. All patients must be carefully protected from acquiring hospital-borne pathogens. Instruments with carefully designed biohazards containment occupy a position of high priority in critical care.

To minimize response time, it is necessary to either create patient proximity, that is, perform testing at or near the site of patient care, or to reduce specimen transit time to the main laboratory, and in either case, to report results immediately. Historically, the LIS solved several problems associated with slow reporting of results by reducing or eliminating delays during the postanalytical phase through computerized processing, telecommunications, and networks. Hospitals

are building on this progress by installing infrared, wireless, and other communication modalities, either in conjunction with the LIS (e.g. "virtual" laboratory handheld digital assistants) or as part of other point-of-service systems (e.g. nursing, respiratory, or inventory wireless touchpads). POCT appeared later historically than computerized information transmission.

7.1 Point-of-care Connectivity

Connectivity is the ability to reliably transfer test information between a POCT device and an information system. In February 2000, 49 healthcare institutions, POC diagnostic vendors, diagnostic test system vendors, and system integrators formed the Connectivity Industry Consortium (CIC) to address the POC integration problem. Since then, numerous technologies (e.g. Bluetooth and WiFi) have become available, which may facilitate improved connectivity in the hospital setting. In 2006, CLSI approved a new standard for connectivity (POCT02-A). This document provides the framework for engineers to design devices, workstations, and interfaces to allow multiple POC devices to communicate bidirectionally with access points, data managers, and LIS from a variety of venders. Table 12 lists several connectivity objectives promoted by the CLSI through the standardization process, which will help to optimize the postanalytical phases of POCT.

8 FUTURE TRENDS

8.1 New Practice Standard for Timeliness

Diagnostic efficiency increases as response time decreases and the cluster effectiveness of tests (or other diagnostic procedures) increases.[3,72] For a critical care profile, optimum response time is achieved when the combination of analysis time and transit time (a function of the distance of the measurement from the patient) is minimized. WBA fundamentally decreases response time by eliminating specimen centrifugation and other preanalytical steps. Process simplicity reduces errors. Current guidelines[1] recommend that rapid response testing should be available continuously during critical emergencies (such as cardiac arrest) and should achieve a turnaround time of 5 min or less from test ordering (actional, verbal, or written) to receipt of results by the clinical team. Immediacy of test results avoids empirical treatment, inefficient delays, and physiological obsolescence. Irrespective of where testing is performed, critical data must be both high quality and timely. POCT, that is, testing at or near the site of patient care, enables a new standard for rapid response diagnostic testing. This new standard is now accepted and expected in the practice of critical care medicine and in other fields of patient care.

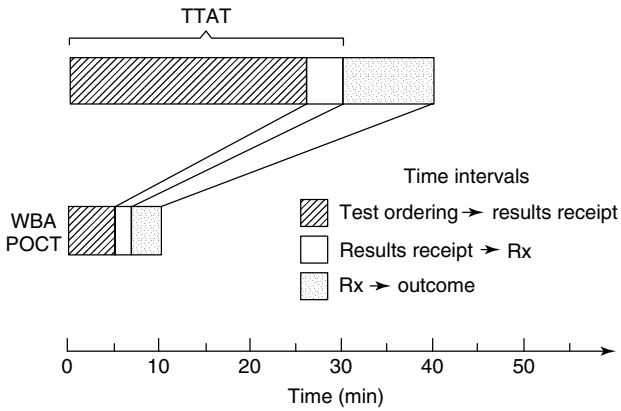

Figure 13 Temporal optimization links rapid response and outcomes. The efficiencies inherent in either WBA or POCT can shorten not only the TTAT but also the interval between treatment and outcome, especially for critically ill patients.

8.2 Critical Speed, Test Clusters, and Medical Necessity

Rapid response shortens TTAT (Figure 13), the time from test ordering to treatment. Fast TTAT accelerates the speed of clinical decision making and diagnostic–therapeutic cycles. Where POCT is available, the recommended median response time for a critical care profile is 5 min.[1,95] In other situations deemed noncritical, it is efficient to fulfill negotiated response times specific to each clinical unit, care path, integrated protocol, or algorithm, with flexibility for individual physician and patient needs. Of course, results should be reported immediately. Technological advances, such as microfabrication, nanofabrication, and wireless technology will bring increasing numbers of networked quantitative tests, possibly as many as 50–100 on only one to three instrument platforms, to or near the bedside in the next 5–10 years, or sooner. These instruments can operate independent of a hospital environment (e.g. field conditions) or become networked into high bandwidth wireless connections compatible with telemedical systems.[94,96] The rapid development of whole-blood and biosensor-based instruments will make the use of conventional instrument platforms in critical care settings obsolete. Numerous qualitative test kits are currently available. POCT is becoming a ubiquitous aid to immediate diagnosis in the United States and abroad. However, POC test clusters must carefully target critical indications, based on medical necessity. The volume of testing should be reduced, while the relevancy should be increased. This impels the discovery of appropriate new diagnostic pivots (such as platelet function tests and multimarker indices), critical decision points, and integrated strategies, in order to avoid excessive costs.

8.3 Molecular Diagnostics

Molecular diagnostics can provide highly specific and sensitive methods for biomarkers and pathogens. As shown in the SeptiFast assay, molecular-based methods such as PCR can detect organisms where traditional methods failed. Future methods will utilize novel, thermal stable methods such as real-time, loop-mediated isothermal amplification (RT-LAMP) to detect nucleic acids without the need for thermocycling. RT-LAMP assays have been published for the detection of severe acute respiratory syndrome corona virus, West Nile virus, Newcastle disease virus, mumps virus, herpes simplex virus, *Porphyromonas gingivitis*, Rubella virus, and Epstein–Barr virus.[97–104] The isothermal nature of the RT-LAMP assay allows it to be used in the field and in disaster response.[105] Other developments may include quantitative nucleic acid testing (NAT). Quantitation will revolutionize infectious disease testing in critical care by allowing us to understand the pathophysiology of bacterial and fungal pathogen invasion and dissemination more clearly.[106] Pathogen kinetics may be revealed with quantitation and allow for monitoring of antimicrobial drug efficacy. Therefore, we may see future development in this area.

8.4 Recommendations for Point-of-care Glucose Testing

The increasing number of diabetics and the continued use of TGC in critically ill patients underscore the need for improved glucose monitoring technology. Performance standards for bedside glucose testing in hospital settings should adopt more strict and unbiased methods such as LS-MAD with a consensus error tolerance (Table 7). POCT teams should evaluate performance and cost effectiveness in the context of tight glucose control, therapeutic decisions, and outcomes in critical care. Future guidelines should include LS-MAD curve pattern recognition when assessing accuracy; LS-MAD theory can be applied to other POC tests, each with their own unique accuracy standard. LS-MAD curves are drawn on the unique human ability to recognize patterns quickly and discriminate performance visually. Both health care personnel (e.g. POC coordinators, critical care physicians, and nurses) and licensing agencies such as the FDA should require ISO 15197-integrated Bland–Altman and LS-MAD curve plots with highlighted discrepant values and erroneous results explained before validating POCT devices. Additionally, different brands of clinical laboratory analyzers and POC glucose devices intended for hospital use, specifically for acute care, should be standardized. These standards should be accuracy based rather than consensus based.

8.5 The National Institutes of Health Point of Care Research Network

In 2006, the National Institutes of Health (NIH), Institute of Biomedical Imaging and Bioengineering put forward a request for applications for the point-of-care technologies research network (POCTRN). The goal of the network is to build expertise in the development of integrated systems that address unmet clinical needs in POCT through the creation of multidisciplinary partnerships.[107] In 2007, four sites were selected including Johns Hopkins University, PATH, the University of Cincinnati, and the University of California Davis – Lawrence Livermore National Laboratory. Each site will have its own theme and is responsible for five cores: (i) development of prototype POC devices that match the center theme, (ii) identification of collaboratory exploratory projects, (iii) clinical needs assessment and dissemination, (iv) training, and (v) administration.[107] Center principal investigators, center websites, and the center-specific themes are shown in Table 11.

Core 1 provides research and development of new POC systems by the Center itself and facilitates clinical evaluation of novel POC designs that emerge from core 2 exploratory projects conducted external to the center. Core 2 will combine the resources of all four

Table 11 The NIBIB point-of-care technologies research network centers

Principal investigators	Center location	Theme
Fred Beyette www.ece.uc.edu/POC-CENT/	University of Cincinnati	Emerging neurotechnologies
Charlotte Gaydos www.hopkinsmedicine.org/medicine/std	Johns Hopkins University	Sexually transmitted diseases
Gerald Kost www.ucdmc.ucdavis.edu/pathology/poctcenter/	UC Davis – Lawrence Livermore National Laboratory (LLNL)	Rapid multipathogen detection for POCT and national disaster readiness
Bernard Weigl www.path.org/dxcenter/index.php	PATH, Seattle and the University of Washington	POC diagnostics for global health
NIBIB POCTRN www.nibib.nih.gov/Research/POCTRN	See for additional details and updates	

NIBIB, National Institute of Biomedical Imaging and Bioengineering; NIH, National Institute of Health; POCTRN, Point-of-Care Technologies Research Network.

sites to solicit, select, and evaluate exploratory projects outside the POC Technologies Research Network. Core 3 assesses diagnostic testing needs to better understand POC technologies that Centers should produce. Results form the basis for soliciting exploratory projects and are disseminated broadly to technology developers and thought leaders. The purpose of core 4 is to train future scientists, engineers, and physicians to become experts in the practical use of POCT. Core 5 oversees the conduct of these four cores and integrates centers in the POCTRN.

8.6 Connectivity for Medical Response to Disasters

Telehealth embedded knowledge packets, in conjunction with POCT strategically placed in small-world networks for regional health care delivery, may provide multilingual education and intellectual connectivity at levels of sophistication appropriate for the practice of modern evidence-based medicine worldwide.[108] Appropriate telehealth integration with respect to demographics and topographies can provide efficient critical, emergency, and disaster care. For example, the NIH-funded wireless internet information system for medical response to disasters (WIISARD).[74] This program used sophisticated wireless technology to coordinate and enhance care of mass casualties in a terrorist attack or natural disaster and constructed a wireless first responder handheld device for rapid triage, patient assessment, and documentation during mass casualty incidents. Additionally, WIISARD integrated 802.11 wireless technologies for a blood pulse oximetry system, and developed an intelligent 802.11 triage tag for medical response to disasters. Ultimately, systems like WIISARD can link together responder wireless devices, patient wireless devices, medical visualization systems, disaster databases, and hospitals into one network. These systems should conform to unitary standards such as those proposed by CLSI (Table 12).[109]

8.7 Knowledge Optimization

Temporal optimization of diagnostic testing focuses on disease evolution and compresses its resolution. Temporal optimization is based on the fact that biochemical variables are coupled or linked, fundamentally, to pathophysiological changes at the cellular level and, ultimately, to patient outcomes. Outcomes integrate preceding effects (Figure 14). Therefore, integration is needed to improve outcomes. Evidence shows that POCT is medically efficacious when integrated within patient-focused protocols, care paths, and treatment algorithms. These approaches streamline essential information and are cost effective in critical care. Prevention of critical problems through early detection yields the greatest medical and economic awards, and is an important area for future development. By targeting surrogates of adverse outcomes, performance maps[70] increase the likelihood of early detection and prevention of sublethal and lethal events. Optimization synthesizes knowledge components. Therefore, you need to synthesize to enhance optimization. The synthesis of temporal optimization and optimization of diagnostic–therapeutic processes[70] produces knowledge accessibility, efficiency, and linkages that are vital to the diagnosis, management, and treatment of critically ill patients. The current shift to POCT points beyond simple efficiency to Knowledge Optimization® in the practice of critical care medicine.

8.8 Enhanced Performance Despite Economic Constraints

Medicine is facing increasingly tough competition in the marketplace. Fiscal constraints challenge our ability to provide the highest quality patient care. At the same time, the clinical team must conserve resources, do more with less, do it faster, and simultaneously improve outcomes. Shortened hospital stays, particularly critical care episodes, are essential. This urgency calls

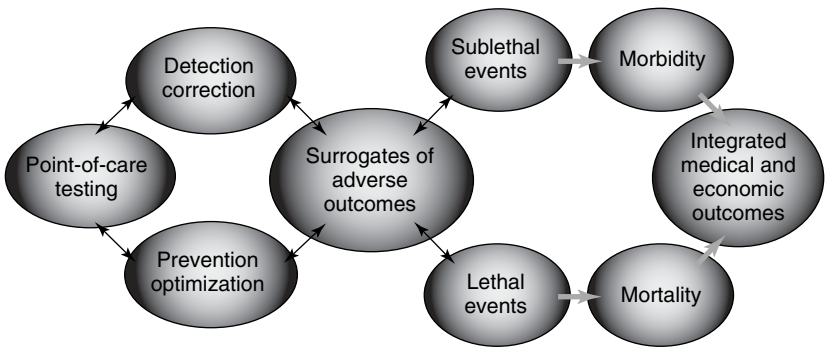

Figure 14 POCT, sublethal/lethal surrogates, and integrated outcomes. The goal is to improve integrated medical and economic outcomes through decreased patient acuity, criticality, morbidity, and mortality. POCT is efficacious when sublethal and lethal surrogates are targeted for early detection and correction, which facilitate prevention and optimization before adverse events can feed forward to untoward outcomes.

Table 12 Connectivity objectives for POCT

Managers, operators, users, and patients

- Ease of use of devices, interfaces, and information (fully integrated with the LIS and HIS)
- Accelerated adoption of point-of-care technologies
- Interoperability with commercial and public domain software (cost savings)
- Facilitated performance enhancement, quality compliance, and training (QC/QA)

Interfaces and networks

- Standardized device docking platform and connections at the point of care (wireless best)
- Use of existing hospital network and computer infrastructures
- Improved regional communications (patient-focused care centers, urgent care, home, and others)

Informatics

- Effortless, fast, and seamless uploading and downloading (forced, if elected)
- Maximum information density and optimized data content (patients, test orders, results, and billing)
- User-configurable informatics use model

Communications

- Bidirectional transmission of instrument, quality, patient, and test result information
- Acceleration (not deceleration) of point-of-care processes (pre- and postanalytical phases)
- Security appropriate for devices, operators, patients, medical records, and legalities

Outcomes

- Faster TTAT and critical results notifications
- Faster, easier, and more complete data retrieving, mining, sharing, analyzing, trending, and managing
- Information consolidation and condensation for outcomes optimization

Source: First meeting of the CIC, Redwood City, CA, October 20, 1999. See www.poccic.org.

for progressive strategies that enhance both personal and team performance, such as using POCT to target surrogates of sublethal and lethal events, potentially reducing morbidity and mortality, and to conserve blood volume, thereby decreasing the need for transfusions, diminishing patient risks and eliminating ''downstream'' and ''hidden'' costs.

The prevention of adverse outcomes benefits the entire health system. Enhanced personal performance, however, depends on accelerated cognitive functions and decision making, and a reduction in the amount of poorly focused, unnecessary, or obsolete information. The new intensive care mosaic combines *in vitro*, *ex vivo*, and *in vivo* multimodal testing and monitoring. Continuous *ex vivo* and *in vivo* monitoring (with telemetry) can be expected to augment critical knowledge. Ultimately, judicious and cost-effective use of these three modalities will lead to continuous improvement in the quality of patient care, despite the financial constraints imposed by managed care and restructured health systems.

8.9 Outcomes Optimization and Future Challenges

Optimization of global POCT requires the development of quantitative technologies, new whole-blood standards, harmonization among standards, and improved quality control.[106,110] This optimization enhances critical knowledge of temporal trends (e.g. lactate), diagnostic–therapeutic patterns (e.g. Ca^{++}), unanticipated changes (e.g. Na^{+}), and organ injury (e.g. cardiac injury markers). Algorithms, performance maps, and new knowledge structures that facilitate synthesis need to be formalized for analytes, clinical problems, and critical care scenarios.[15,16] Formal optimization will be needed when treatment is automated using closed-loop feedback systems and *ex vivo* or *in vivo* patient monitoring. It will be necessary to assess efficacy and to determine, based on medical necessity, whether the changes in practice patterns are economically sound. However, published data indicate that POCT facilitates immediate evidence-based medical decisions and improves patient outcomes in the operating room, the intensive care unit, and other settings.

9 CONCLUSIONS

The hybrid laboratorian focuses on providing healthcare directly as a collaborative member of the clinical team. Therefore, outcome optimization forms a solid basis for setting priorities for the laboratory support of critical care. The new challenge of advancing technology and communication is how to use immediate diagnostic information to improve medical and economic efficiency (Table 13). The focus shifts to trends in information and patterns of change. The currency of the future will be organized knowledge. POCT is a knowledge technology. Networked POCT can organize data packets quickly to provide the right information to the right person

Table 13 Future directions for POCT

- Advancing technology and communication
  - New whole-blood assays, biosensors, and tests (e.g. organ injury markers)
  - Nanotechnology and modularity
  - Noninvasive and minimally invasive monitoring
  - Wireless bidirectional communication
  - Seamless informatics and tracking
  - Uniform regulations and connectivity standards
- Improving medical and economic efficiency
  - Flexible test menus, patient-focused test clusters, and increased pivotal test results
  - Clinical test ranges spanning critical limits (critical values, panic values)
  - Blood volume conservation and biohazard containment
  - Practice standards for fast TTAT (accreditation-required rapid response)
  - Cost effectiveness and peer comparisons (using local, regional, and national databases)
- Integrating knowledge and decisions (Knowledge Optimization®)
  - Diagnostics driven by POCT (increased useful tests, decreased useless tests)
  - Smarter *in vitro*, *ex vivo*, and *in vivo* systems guided by expert systems (artificial intelligence)
  - Simultaneous trending of biochemical and physiological variables
  - Total connectivity of devices, computerized systems, and information
  - Direct physician alerting and real-time decision making (reduced stat testing)
  - Clinical research, evidence-based care, and outcomes optimization
- Disaster preparedness
  - Prepare for regional disasters (e.g. ''newdemics'') by using POCT to facilitate the practice of on-site evidence-based medicine in small-world networks.
  - Develop POC technologies that withstand environmental extremes – high temperature, freezing, long-term humidity, dust exposure, mechanical shock, and different altitude
  - Establish field operating standards that manufacturers must meet to certify POC devices, monitors, and reagents for use in disasters
  - Equip US DMATs with POCT, train members in its use, and validate skills under simulated disaster conditions
  - Determine which tests DMATs need and which instruments they carry by matching treatment capabilities
  - Provide POCT on mobile units (land based, helicopters, and ships) to circumvent obstacles in different geographic and topographic settings

at the right time. Therefore, POCT can quickly create results that are medically valuable in critical care and other medical settings through integrating knowledge and evidence-based decision making.

ACKNOWLEDGMENTS

This article was supported by a Point-of-Care Technologies Research Network Center grant (Dr. Kost, PI, 1U54 EB007959-01) from the National Institute of Biomedical Imaging and Bioengineering, NIH, and by the Point-of-Care Testing Center for Teaching and Research, POCT · CTR℠ (Dr. Kost, Director), School of Medicine, UC Davis. Figures and tables are reproduced by permission of Knowledge Optimization® (Davis, CA).

ABBREVIATIONS AND ACRONYMS

ACT	Activated Clotting Time
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANOVA	Analysis of Variance
aPTT	Activated Partial Thromboplastin Time
ARDS	Adult Respiratory Distress Syndrome
ASOS	Analyte-specific Optical Sensor
AST	Aspartate Aminotransferase
ATI	Automated Testing Instrument
ATM	Automated Testing Machine
β-hCG	Beta-human Chorionic Gonadotropin
BNP	B-type Natriuretic Peptide
BUN	Blood Urea Nitrogen
CAP	College of American Pathologists
CBC	Complete Blood Count
CDR	Clinical Data Repository
CIC	Connectivity Industry Consortium
CK	Creatine Kinase
CK-MB	Creatine Kinase-MB isoenzyme
CLIA'88	Clinical Laboratory Improvement Amendments of 1988
CLSI	Clinical Laboratory Standards Institute
CPB	Cardiopulmonary Bypass Surgery
cTnI	Cardiac Troponin I
DMAT	Disaster Medical Assistance Team
DNA	Deoxyribonucleic Acid
E3G	Estrone-3-glucuronide

ECS	Electrical Conductance Sensor
FDA	Food and Drug Administration
FHHb	Fraction of Deoxyhemoglobin
FmetHb	Fraction of Met-hemoglobin
FRET	Fluorescent Resonance Energy Transfer
GD	Glucose Dehydrogenase
GGT	Gamma Glutamyl Transferase
GMS	Glucose Monitoring System
GO	Glucose Oxidase
GOT	Glutamic Oxaloacetic Transaminase
GPT	Glutamic Pyruvic Transaminase
HBV	Hepatitis B Virus
HCFA	Health Care Financing Administration
hCG	Human Chorionic Gonadotropin
Hct	Hematocrit
HCV	Hepatitis C Virus
HDL	High-density Lipoprotein
H-FABP	Human-type Fatty Acid–Binding Protein
HHb	Deoxyhemoglobin
HIS	Hospital Information System
HIV	Human Immunodeficiency Virus
HNTT	Heparin-neutralized Thrombin Time
HRP	Horseradish Peroxidase
INR	International Normalized Ratio
ISE	Ion-selective Electrode
ISO	International Organization of Standards
JCAHO	Joint Commission on Accreditation of Healthcare Organizations
Lac	Lactose
LAMP	Loop-mediated Isothermal Amplification
LD	Lactate Dehydrogenase
LDL	Low-density Lipoprotein
LED	Light-Emitting Diode
LIS	Laboratory Information System
LR-ACT	Low-range Activated Clotting Time
LS-MAD	Locally Smoothed Median Absolute Difference
MBTH	*meta*(3-Methyl 2-benzothiazolinone hydrazone)–*N*-sulfonyl benzenesulfonic acid
MetHb	Met-hemoglobin
MIP-3	Macrophage Inflammatory Protein-3
MRSA	Methicillin Resistant Staphylococcus Aureus
NDMS	National Disaster Medical System
NGAL	Neutrophil Gelatinase-associated Lipocalin
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NIH	National Institute of Health
NT-proBNP	N-terminal pro-B-type Natriuretic Peptide
PBC	Protein Band Cassette
PBD	Protein Band Dipstick
PCA	Portable Clinical Analyzer
pCO_2	Partial Pressure of Carbon Dioxide
PCR	Polymerase Chain Reaction
PDAO	Protamine Dose Assay
pO_2	Partial Pressure of Oxygen
POC	Point-of-care
POCT	Point-of-care Testing
POCTRN	Point-of-Care Technologies Research Network (NIBIB, NIH)
PQQ	Pyrroloquinoline Quinone
PSA	Prostate Specific Antigen
PT	Prothrombin Time
PTCA	Percutaneous Transluminal Coronary Angioplasty
PTD	Personal Testing Device
RACT	Recalcified Activated Clotting Time
RT-LAMP	Real-time, Loop-mediated Isothermal Amplification
SARS	Severe Acute Respiratory Syndrome
SERS	Surface-enhanced Raman Spectroscopy
sO_2	Oxygen Saturation
SO_2%	Percent Oxygen Saturation
SSE	Substrate-specific Electrode
tBil	Total Bilirubin
TC	Transcutaneous
TCM	Transcutaneous Monitoring
TCO_2	Total Carbon Dioxide
TEG	Thromboelastography
TGC	Tight Glycemic Control
tHb	Total Hemoglobin
TnI	Troponin I
TPA	Tissue Plasminogen Activator
TpP	Thrombosis Precursor Protein
TSH	Thyroid Stimulating Hormone
TT	Thrombin Time
TTAT	Therapeutic Turnaround Time
USOSHA	United States Occupational Safety and Health Administration
WBA	Whole-blood Analysis
WBC	White Blood Cell Count
WIISARD	Wireless Internet Information System for Medical Response to Disasters

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo • Glucose, In Vivo Assay of

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors • Raman Spectroscopy in Analysis of Biomolecules

Clinical Chemistry **(Volume 2)**
Biochemical Markers of Acute Coronary Syndromes • Biosensor Design and Fabrication • Diagnostic Hematology • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Electrolytes, Blood Gases, and Blood pH • Glucose Measurement • Laboratory Instruments in Clinical Chemistry, Principles of • Urinalysis and Other Bodily Fluids

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals

REFERENCES

1. G.J. Kost, 'Guidelines for Point-of-care Testing: Improving Patient Outcomes', *Am. J. Clin. Pathol.*, **104**, S111–S127 (1995).
2. G.J. Kost, 'New Whole Blood Analyzers and their Impact on Cardiac and Critical Care', *Crit. Rev. Clin. Lab. Sci.*, **30**, 153–202 (1993).
3. G.J. Kost, 'The Hybrid Laboratory: Shifting the Focus to the Point of Care', *Med. Lab. Obs.*, **24**(9S), 17–28 (1992).
4. G.J. Kost, 'The Hybrid Laboratory: The Clinical Laboratory of the 1990's is a Synthesis of the Old and the New', *Arch. Pathol. Lab. Med.*, **116**, 1002–1003 (1992).
5. G.J. Kost, D.A. Wiese, T.P. Bowen, 'New Whole Blood Methods and Instruments: Glucose Measurement and Test Menus for Critical Care', *J. Int. Fed. Clin. Chem.*, **3**, 160–172 (1991).
6. G.J. Kost, 'Guidelines for Point-of-care Testing', *Proceedings of the Society of Critical Care Medicine Educational and Scientific Symposium, Audio Recording*, Convention Cassettes Unlimited, Palm Desert, CA, Vol. 24(8), 1995.
7. G.J. Kost, C. Hague, 'The Current and Future Status of Critical Care Testing and Patient Monitoring', *Am. J. Clin. Pathol.*, **104**, S2–S17 (1995).
8. G.J. Kost, C. Hague, 'Vitro, *ex vivo*, and *in vivo* Biosensor Systems', in *Handbook of Clinical Automation, Robotics and Optimization*, ed. G.J. Kost, John Wiley & Sons, New York, 648–753, 1996.
9. G.J. Kost, H.T. Vu, M. Inn, R. Duplantier, M. Fleisher, M.H. Kroll, J.C. Spinosa, 'Multicenter Study of Whole-blood Creatinine, Total Carbon Dioxide Content, and Chemistry Profiling for Laboratory and Point-of-care Testing in Critical Care in the United States', *Crit. Care Med.*, **28**, 2379–2389 (2000).
10. N.K. Tran, C. Promptmas, G.J. Kost, 'Biosensors, Miniaturization, and Noninvasive Techniques', in *Clinical Diagnostic Technology: The Total Testing Process, Preanalytical, Analytical, and Post-analytical Phases*, eds. K.W. Cook, C. Lehmann, L. Schoeff, R. Williams, AACC, Washington, DC, 145–184, Vol. 3 of 3, Chapter 7, 2006.
11. G.J. Kost, H.T. Vu, J.H. Lee, P. Bourgeois, F.L. Kiechle, C. Martin, S.S. Miller, A.O. Okorodudu, J.J. Podczasy, R. Webster K.J. Whitlow, ''US/Canada Multicenter Study of O_2-insensitive Handheld Glucose Point-of-care Testing in Critical Care/Hospital/Ambulatory Patients'', *Crit. Care Med.*, **28**, 666–677 (1998).
12. D.D. Martin, M.D. Sherphard, H. Freeman, M.K. Bulsara, T.W. Jones, E.A. Davis, G.P. Maguire, 'Point-of-care Testing of HbA_{1c} and Blood Glucose in a Remote Aboriginal Australian Community', *Med. J. Aust.*, **182**, 524–527 (2005).
13. B. Beneteau-Burnat, M. Ott, A.L. Fort, J. Guechot, M. Vaubourdolle, 'Analytic Performance of a point-of-care Instrument for Measurement of Human Chorionic Gonadotropin Hormone', *Point Care: J. Near-Patient Test Technol.*, **4**, 98–100 (2005).
14. W.D. Engle, G.L. Jackson, E.K. Stehel, D.M. Sendelbach, M.D. Manning, 'Evaluation of a Transcutaneous Jaundice Meter Following Hospital Discharge in Term and Near-term Neonates', *J. Perinatol.*, **25**, 486–490 (2005).
15. H.B. Grossman, E. Messing, M. Soloway, G. Katz, B. Stein, V. Kassabian, Y. Shen, 'Detection of Bladder Cancer Using a Point-of- care Proteomic Assay', *JAMA*, **293**, 810–816 (2005).
16. L.V. Rao, F. Jakubia, J.S. Sidwell, J.W. Winkelman, M.L. Snyder, 'Accuracy Evaluation of a New Glucometer with Automated Hematocrit Measurement and Correction', *Clin. Chim. Acta*, **356**, 178–183 (2005).
17. L.J. Sokoll, F.H. Wians, A.T. Rema, 'Rapid Intraoperative Immunoassay of Parathyroid Hormone and Other Hormones: A New Paradigm for Point-of-care Testing', *Clin. Chem.*, **50**, 1126–1135 (2004).
18. W.C. Shoemaker, H. Belzberg, C.C. Wo, D.P. Milzman, M.D. Pasquale, L. Baga, M.A. Fuss, G.J. Fulda, K.K. Yarbrough, J.P. Van DeWater, P.J. Ferraro, D. Thangathurai, P. Roffey, G. Velmahos, J.A. Murray, J.A. Asensio, K. ElTawil, W.R. Dougherty, M.J. Sullivan, R.S. Patil, J. Adibi, C.B. James, D. Demetriades, 'Multicenter Study of Noninvasive Monitoring Systems as Alternatives to Invasive Monitoring of Acutely Ill Emergency Patients', *Chest*, **114**, 1643–1652 (1998).
19. W.A. Willems, L.M. van den Berg, H. de Wit, A. Molendijk, 'Transcutaneous Bilirubinometry with the Bilicheck in Very Premature Newboards', *J. Matern. Fetal Neonatal Med.*, **16**, 209–214 (2004).

20. B. Rolinski, A.O. Okorodudu, G.J. Kost, M. Roser, J. Wu, A. Goerlach-Graw, H. Kuester, 'Evaluation of Total Bilirubin Determination in Neonatal Whole-blood Samples by Multiwavelength Photometry on the Roche OMNI S Point-of-care Analyzer', *Point Care: J. Near-Patient Test Technol.*, **4**, 3–8 (2005).
21. B. Bewly, G. Creed, A. Goerlach-Graw, B. Engel, G. Halwachs-Baumann, H. Holzer, N. Katz, G.J. Kost, M. Kraft, H. Kuester, K. Miedema, D. Nagal, H. Normand, A. Okorodudu, B. Olander, B. Rolinski, S. Utzolino, B. Wiley, 'Multicenter Study on Analytic Performance of a New Point-of-care Blood Gas Analyzer and its Use in Critical Care Testing', *Point Care: J. Near-Patient Test Technol.*, **3**, 149–155 (2004).
22. National Academy of Clinical Biochemistry, '*Laboratory Medicine Guidelines, Diabetes Mellitus*', 2002, http://www.nacb.org/lmpg/diabetes_LMPG_PDF.stm, Accessed on September 9, 2007.
23. National Academy of Clinical Biochemistry, '*Laboratory Medicine Guidelines: Evidence Based Practice for POCT*', 2004, http://www.nacb.org/lmpg/poct/POCT_LMPG_41706litsearches.pdf, Accessed on September 9, 2007.
24. G. van den Berghe, P. Wouters, F. Weekers, C. Verwaest, F. Bruyninckx, M. Schetz, D. Vlasselaers, P. Ferdinande, P. Lauwers, R. Bouillon, 'Intensive Insulin Therapy in Critically Ill Patients', *N. Engl. J. Med.*, **345**, 1359–1367 (2001).
25. G.H. Van den Berghe, 'Role of Intravenous Insulin Therapy in Critically Ill Patients', *Endocr. Pract.*, **10**(Suppl 2), 17–20 (2004).
26. H.L. Lazar, S.R. Chipkin, C.A. Fitzgerald, Y. Bao, H. Cabral, C.S. Apstein, 'Tight Glycemic Control in Diabetic Coronary Artery Bypass Graft Improves Perioperative Outcomes and Decreases Recurrent Ischemic Events', *Circulation*, **109**, 1497–1502 (2004).
27. C.R. Zimmerman, M.E. Mlynarek, J.A. Jordan, C.A. Rajda, H.M. Horst, 'An Insulin Infusion Protocol in Critically Ill Cardiothoracic Surgery Patients', *Ann. Pharmacother.*, **38**, 1123–1128 (2004).
28. P.A. Goldberg, M.D. Siegel, R.S. Sherwin, J.I. Halickman, M. Lee, V.A. Bailey, S.L. Lee, J.D. Dziura, S.E. Inzucchi, 'Implementation of a Safe and Effective Insulin Infusion Protocol in a Medical Intensive Care Unit', *Diabetes Care*, **27**, 461–467 (2004).
29. J.S. Krinsley, 'Association Between Hyperglycemia and Increased Hospital Mortality in Heterogeneous Population of Critically Ill Patients', *Mayo Clin. Proc.*, **78**, 1471–1478 (2003).
30. Personal Communication, SICU, Duke University Medical Center, August 9 1995.
31. Dr. Inuzucchi, email correspondence, Yale School of Medicine, June 18 2005.
32. A.P. Furnary, *AACC Teleconference*, OR/ICU OHSU, Portland, OR, May 11 2005.
33. Personal Communication, Department of Endocrinology, UC San Francisco Medical Center, June 20 2005.
34. University of Pittsburgh, Department of Critical Care, http://www.ccm.upmc. edu/pittway/insulin_iv_drip.htm, Accessed on October 26, 2006.
35. K. Dungan, J. Chapman, S.S. Braithwaite, J. Buse, 'Glucose Measurement: Confounding Issues in Setting Targets for Inpatient Management', *Diabetes Care*, **30**, 403 (2007).
36. Z. Tang, J.H. Lee, R.F. Louie, G.J. Kost, 'Effects of Different Hematocrit Levels on Glucose Measurements with Handheld Meters for Point-of-care Testing', *Arch. Pathol. Lab. Med.*, **124**, 1135–1140 (2000).
37. Z.T. Tang, R.F. Louie, G.J. Kost, 'Principles and Performance of Point-of-care Testing Instruments', in *Principles and Practice of Point-of-care Testing*, ed. G.J. Kost, Lippincott Williams and Wilkins, Philadelphia, PA, 67–92, Chapter 6, 2002.
38. D. Oberg, C.G. Ostenson, 'Performance of Glucose Dehydrogenase – and Glucose Oxidase-Based Blood Glucose Meters at High Altitude and Low Temperatures (Letter)', *Diabetes Care*, **28**, 1261 (2005).
39. H.F. Sylvain, M.E. Pokorny, S.M. English, N.H. Benson, T.W. Whitley, C.J. Ferenczy, J.G. Harrison, 'Accuracy of Fingerstick Glucose Values in Shock Patients', *Am. J. Crit. Care*, **4**, 44 (1995).
40. K. Kuwa, T. Makayama, T. Hoshimo, M. Tominaga, 'Relationships of Glucose Concentrations in Capillary Whole Blood, Venous Whole Blood and Venous Plasma', *Clin. Chim. Acta*, **307**, 187 (2001).
41. P. D'Orazio, R.W. Burnett, N. Fogh-Anderson, E. Jacobs, K. Kuwa, W.R. Kulpmann, L. Larsson, A. Lewanstam, A. Maas, G. Mager, J.W. Naskalski, A.O. Okorodudu, 'The International Federation of Clinical Chemistry Science Division Working Group on Selective Electrodes and Point of Care Testing: Approved IFCC Recommendations on Reporting Results for Blood Glucose', *Clin. Chem.*, **51**, 1573–1576 (2005).
42. Z. Tang, X. Du, R.F. Louie, G.J. Kost, 'Effects of Drugs on Glucose Measurements with Handheld Glucose Meters and a Portable Glucose Analyzer', *Am. J. Clin. Pathol.*, **113**, 75–86 (2000).
43. U.S. Food and Drug Administration, *'FDA Reminders for Falsely Elevated Glucose Readings from Use of Inappropriate Test Method'*, 2005. http://www.fda.gov/cdrh/oivd/news/glucosefalse.html. Accessed on July 17, 2007.
44. E.S. Kilpatrick, A.G. Rumley, E.A. Smith, 'Variations in Sample pH and pO_2 Affect ExacTech Meter Glucose Measurements', *Diabet. Med.*, **11**, 506 (1994).

45. A. Haupt, B. Berg, P. Paschen, M. Dreyer, H.U. Haring, J. Smedegaard, S. Matthaei, 'The Effects of Skin Temperature and Testing Site on Blood Glucose Measurements Taken by a Modern Blood Glucose Monitoring Device', *Diabetes Technol. Ther.*, **7**, 597 (2005).

46. R. Bergenstal, J. Pearson, G.S. Cembrowski, D. Bina, J. Davidson, S. List, 'Identifying Variables Associated with Inaccurate Self-monitoring of Blood Glucose: Proposed Guidelines to Improve Accuracy', *Diabetes Educ.*, **26**, 981 (2000).

47. G.J. Kost, N.K. Tran, M. Tuntideelert, S. Kulrattanamaneeporn, N. Peungposop, 'Katrina, the Tsunami, and Point-of-care Testing: Optimizing Rapid Response Diagnosis in Disasters', *Am. J. Clin. Pathol.*, **126**, 513 (2006).

48. R.F. Louie, S.L. Sumner, S. Belcher, R. Mathews, N.K. Tran, G.J. Kost, 'Thermal Stress and Point-of-care Testing Performance: Suitability of Glucose Test Strips and Blood Gas Cartridges for Disaster Response' *Disaster Medicine and Public Health Preparedness*, accepted (2008).

49. M. Fernandez, K. Burns, B. Calhoun, S. George, B. Martin, C. Weaver, 'Evaluation of a New Pulse Oximeters Sensor', *Am. J. Crit. Care*, **16**, 146–152 (2007).

50. G.S. Agashe, J. Coakley, P.D. Mannheimer, 'Forehead Pulse Oximetry: Headband Use Helps Alleviate False Low Readings Likely Related to Venous Pulsation Artifact', *Anesthesiology*, **105**, 1111–1116 (2006).

51. L. Schallom, C. Sona, M. McSweeney, J. Mazuski, 'Comparison of Forehead and Digit Oximetry in Surgical/Trauma Patients at Risk for Decreased Peripheral Perfusion', *Heart Lung*, **36**, 188–194 (2007).

52. W.W. Lee, K. Mayberry, R. Crapo, R.L. Jenson, 'The Accuracy of Pulse Oximetry in the Emergency Department', *Am. J. Emerg. Med.*, **18**, 427–431 (2000).

53. N.S. Trivedi, A.F. Ghouri, N.K. Shah, E. Lai, S.J. Barker, 'Effects of Motion, Ambient Light, and Hypoperfusion on Pulse Oximeters Function', *J. Clin. Anesth.*, **9**, 179–183 (1997).

54. F.A. Workie, K. Rais-Bahrami, B.L. Short, 'Clinical Use of New-generation Pulse Omixeters in the Neonatal Intensive Care Unit', *Am. J. Perinatol.*, **22**, 357–360 (2005).

55. J.W. Berkenbosch, J.D. Tobias, 'Comparison of a New Forehead Reflectance Pulse Oximeters Sensor with a Conventional Digit Sensor in Pediatric Patients', *Respir. Care*, **51**, 715–716 (2006).

56. J. Townshend, B.J. Taylor, B. Galland, S. Willams, 'Comparison of New Generation Motion-resistant Pulse Oximeters', *J. Paediatr. Child Health*, **42**, 359–365 (2006).

57. E. Simson, H.J. Day, C.F. Arkin, *Collection, Transport, and Processing of Blood Specimens for Coagulation Testing and Performance of Coagulation Assays*, National Committee for Clinical Laboratory Standards, Villanova, PA, Document H21-A3, 1997.

58. D.J. McMahon, R.L. Carpenter, ''A Comparison of Conductivity-based Hematocrit Determinations with Conventional Laboratory Methods in Autologous Blood Transfusions'', *Anesth. Analg.*, **71**, 541–544 (1990).

59. R.A.W. Stott, G.L. Hortin, T.R. Wilhite, S.B. Miller, C.H. Smith, M. Landt, ''Analytical Artifacts in Hematocrit Measurements by Whole-blood Chemistry Analyzers'', *Clin. Chem.*, **41**, 306–311 (1995).

60. *'National Academy of Clinical Biochemistry Guidelines, Biomarkers of Acute Coronary Syndromes and Heart Failure'*, 2007, http://www.aacc.org/AACC/members/nacb/LMPG/OnlineGuide/PublishedGuidelines/ACS Heart/ heartpdf.htm. Accessed on September 9, 2007.

61. G.J. Kost, J.D. Kirk, K. Omand, ''A Strategy for the Use of Cardiac Injury Markers (Troponin I and T, Creatine Kinase-MB Mass and Isoforms, and Myoglobin) in the Diagnosis of Acute Myocardial Infarction'', *Arch. Pathol. Lab. Med.*, **122**, 245–251 (1998).

62. F. Di Serio, L.D. Francesca, G. Amodio, L. Varrasco, V. Ruggieri, G. Antonelli, N. Pansini, 'Point-of-care Cardiac Markers: Clinical Impact of the Troponin 99th Percentile Cutoff and Clinical Utility of the Myoglobin Measurement in Early Management of Chest Pain Patients in Low to Intermediate Acute Coronary Syndrome Risk Population Admitted to Emergency Cardiology Department', *Point Care*, **3**, 183–186 (2007).

63. *'National Academy of Clinical Biochemistry Guidelines, Evidence-based Practice for Point-of-care Testing'*, 2007, http://www.aacc.org/AACC/members/nacb/LMPG/OnlineGuide/PublishedGuidelines/poct, Accessed on September 9, 2007.

64. G.J. Kost, N.K. Tran, 'Point-of-care Testing and Cardiac Biomarkers: The Standard of Care and Vision for Chest Pain Centers', *Cardiol. Clin.*, **23**, 467–490 (2005).

65. CDC, Morbidity and Mortality Weekly Report, September 22, 2006, Volume 55. http://www.cdc.gov/mmwr/PDF/rr/rr5514.pdf Accessed on September 9, 2007.

66. World Health Organization Recommendations for Rapid Influenza Tests, 2007, http://www.who.int/csr/disease/avian_influenza/guidelines/RapidTestInfluenza_web.pdf Accessed on September 9, 2007.

67. M.E. Alexander, C.S. Bowman, Z. Feng, M. Gardam, S.M Moghadas, G. Rost, J. Wu, P. Yan, 'Emergence of Drug Resistance: Implications for Antiviral Control of Pandemic Influenza', *Proc. Biol. Sci.*, **22**, 1675–1684 (2007).

68. R.F. Louie, Z. Tang, T.E. Albertson, S. Cohen, N.K. Tran, G.J. Kost, 'Multiplex Polymerase Chain Reaction Detection Enhancement of Bacteremia and Fungemia', *Crit. Care Med.*, **36**, 1487–1492 (2008).

69. *'Cepheid SmartCycler'*, 2007, http://www.cepheid.com/Sites/cepheid/content.cfm?id=215 Accessed on September 9, 2007.

70. G.J. Kost, 'Artificial Intelligence and New Knowledge Structures', in *Handbook of Clinical Automation, Robotics, and Optimization*, ed. G.J. Kost, John Wiley & Sons, New York, 149–196, 1996.

71. G.J. Kost, 'Point-of-care Testing ⇒ The Hybrid Laboratory ⇒ Knowledge Optimization®', in *Handbook of Clinical Automation, Robotics, and Optimization*, ed. G.J. Kost, John Wiley & Sons, New York, 757–838, 1996.

72. J.W. Kennedy, R.N. Caey, C.C. Coolen, *Method Comparison and Bias Estimation Using Patient Samples*, National Committee for Clinical Laboratory Standards, Villanova, PA, Document EP9-A, 1995.

73. *'Pacific Emprints'*, 2008, http://www.emprints.hawaii.edu/ Accessed on July 20, 2008.

74. *'Wireless Internet Information System for Medical Response in Disasters'*, https://www.wiisard.org/, Accessed on September 9, 2007.

75. G.J. Kost, 'Role of New Whole Blood Analytical Techniques in Critical Care', *Clin. Chem.*, **35**, 1232–1233 (1989).

76. G.J. Kost, T.L. Shirey, 'New Whole-blood Testing for Laboratory Support of Critical Care at Cardiac Transplant Centers and US Hospitals', *Arch. Pathol. Lab. Med.*, **114**, 865–868 (1990).

77. G.J. Kost, 'Point-of-care Testing: Patient Focusing for the Future', in *Professional Practice in Clinical Chemistry: A Review*, eds. P.J. Howanitz, J.H. McBride, American Association for Clinical Chemistry Press, Washington, DC, 443–454, 1994; 111–122, 1996.

78. Clinical Laboratory Standards Institute, https://www.nccls.org, Accessed August 15, 2007.

79. International Organization for Standardization, http://www.iso.ch, Accessed August 15, 2007.

80. W.L. Clarke, 'The Original Clarke Error Grid Analysis (EGA)', *Diabetes Technol. Ther.*, **7**, 776 (2005).

81. B.P. Kovatchev, L.A. Godner-Frederic D.J. Cox, W.L. Clarke, 'Evaluating the Accuracy of Continuous Glucose-monitoring Sensors', *Diabetes Care*, **27**, 1922 (2004).

82. J.M. Bland, D.G. Altman, 'Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement', *Lancet*, **327**, 307 (1986).

83. 'Total Produce Life Cycle for Portable Invasive Blood Glucose Monitoring Systems. Draft FDA Guidance Document', October 24, 2006. http://www.fda.gov/cdrh/oivd/guidance 1603.html, Accessed on July 15, 2007.

84. G.J. Kost, N.K. Tran, R.F. Louie, N.L. Gentile, V.J. Abad, 'Assessing the Performance of Handheld Glucose Testing for Critical Care', *Diabetes Technol. Ther.* (2008) [In Press].

85. G.J. Kost, N.K. Tran, V.J. Abad, R.F. Louie, 'Evaluation of Point-of-care Glucose Testing Accuracy Using Locally-smoothed Median Absolute Difference Curves', *Clin. Chim. Acta*, **389**, 31–39 (2008).

86. G.J. Kost, N.K. Tran, V.J. Abad, R.F. Louie, 'Global Standardization and Improved Technologies for Point-of-care Glucose Testing', *Clin. Chim. Acta*, **391**, 127–128 (2008).

87. W.W. Daniel, *Biostatistics: A Foundation for Analysis in the Health Sciences*, 8th edition, John Wiley & Sons, Hoboken, NJ, 2005.

88. US Department of Health and Human Services, *'Medicare, Medicaid and CLIA Programs: Regulations Implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA)'*, Final Rule, Federal Register, 57, 7002–7186, 1992.

89. S.S. Ehrmeyer, R.H. Laessig, 'Regulatory Requirements (CLIA'88, JCAHO, CAP) for Decentralized Testing', *Am. J. Clin. Pathol.*, **104**(Suppl 1), S40–S49 (1995).

90. S.S. Ehrmeyer, R.H. Laessig, 'POCT Regulations and Inspections', *New Approaches to Point-of-care Testing*, American Association for Clinical Chemistry, Washington, DC, 1997.

91. Joint Commission on Accreditation of Healthcare Organizations, *Improving Performance in the Lab: A Case Study Approach*, JCAHO, Oakbrook Terrace, IL, 1997.

92. A. Belanger, 'JCAHO's Perspective on POCT', *Adv. Adm. Lab.*, **6**, 10–11 (1997).

93. Joint Commission on Accreditation of Healthcare Organizations, *Comprehensive Accreditation Manual for Pathology and Clinical Laboratory Services*, JCAHO, Oakbrook Terrace, IL, 2000.

94. J.T. Barr, J. Betschart, A. Bracey, *Ancillary (Bedside) Blood Glucose Testing in Acute and Chronic Care Facilities*, National Committee for Clinical Laboratory Standards, Villanova, PA, Document C30-A, 1994.

95. B.M. Goldsmith, E.M. Travers, R. Bakes-Martin, *Point-of-care in vitro Diagnostic (IVD) Testing*, National Committee for Clinical Laboratory Standards, Villanova, PA, Document AST2-P, 1995.

96. T.C. Hong, Q.L. Mai, D.V. Cuong, M. Parida, H. Minekawa, T. Notomi, F. Hasebe, K. Morita, *Point-of-care Testing Inspection Checklists*, College of American Pathologists Commission on Laboratory Accreditation, Northfield, IL, 2000.

97. G.J. Kost, K. Vansith, M. Tuntideelert, 'Critical Care and Point-of-care Testing in Cambodia and Vietnam', *Point Care: J. Near-Patient Test Technol.*, **5**, 193–198 (2006).

98. H.T. Thai, C.M.Q. Le, et al. 'Development and Evaluation of a Novel Loop-mediated Isothermal

Amplification Method for Rapid Detection of Severe Acute Respiratory Syndrome Coronavirus', *J. Clin. Microbiol.*, **42**, 1956–1961 (2004).

99. M.G. Parida, G. Posadas, S. Inoue, F. Hasebe, K. Morita, 'Real-time Reverse Transcription Loop-mediated Isothermal Amplification for Rapid Detection of West Nile Virus', *J. Clin. Microbiol.*, **42**, 257–263 (2004).
100. H.M. Pham, C. Nakajima, K. Ohashi, M. Onuma, 'Loop-mediated Isothermal Amplification for Rapid Detection of Newcastle Disease Virus', *J. Clin. Microbiol.*, **43**, 1646–1650 (2005).
101. T. Okafuji, N. Yoshida, M. Fujino, Y. Motegi, T. Ihara, Y. Ota, T. Notomi, T. Nakayama, 'Rapid Diagnostic Method for Detection of Mumps Virus Genome by Loop-mediated Isothermal Amplification', *J. Clin. Microbiol.*, **43**, 1625–1631 (2005).
102. H. Kimura, M. Ihira, Y. Enomoto, J. Kawada, Y. Ito, T. Morishima, T. Yoshikawa, Y. Asano, 'Rapid Detection of Herpes Simplex Virus DNA in Cerebrospinal Fluid: Comparison Between Loop-mediated Isothermal Amplification and Real-time PCR', *Med. Microbiol. Immunol. (Berl.)*, **194**, 181–185 (2005).
103. H. Maeda, S. Kokeguchi, C. Fujimoto, I. Tanimoto, W. Yoshizumi, F. Nishmura, S. Takashiba, 'Detection of Periodontal Pathogen Porphyromonas Gingivalis by Loop-mediated Isothermal Amplification Method', *FEMS Immunol. Med. Microbiol.*, **43**, 233–239 (2005).
104. N. Mori, Y. Motegi, Y. Shimamura, T. Ezaki, T. Natsumeda, T. Yonekawa, Y. Ota, T. Notomi, T. Nakayama, 'Development of a New Method for Diagnosis of Rubella Virus Infection by Reverse Transcription-loop-mediated Isothermal Amplification', *J. Clin. Microbiol.*, **44**, 3268–3273 (2006).
105. R.F. Louie, B. Hindson, N.K. Tran, N.L. Gentile, G.J. Kost, 'Three Wishes for Point-of-care Testing in Disaster Care', *Point Care*, (2008) [In Press].
106. N.K. Tran, D. Wisner, T.E. Albertson, S. Cohen, D.G. Greenhalgh, J. Galante, H.H. Phan, G.J. Kost, 'Quantitative Point-of-care Pathogen Detection in Septicemia', *Point Care* (2008) [In Press].
107. G.J. Kost, M. Minear, P.M. Siegel, 'Knowledge → Education → Mind Connectivity: Using Telemedicine to Achieve a Global Vision for Point-of-care Testing', *Point Care*, **7**, 69–71 (2008).
108. G.J. Kost, B. Korte, F.R. Beyette, B. Weigl, C. Gaydos, 'The NIBIB Point of Care Technologies Research Network Center Themes and Opportunities for Exploratory Projects', *Point Care*, **7**, 41 (2008).
109. *'Clinical Laboratory Standards Institute Connectivity Standards'*, http://www.nccls.org/source/orders/free/po-ct1-a2.pdf, Accessed on September 9, 2007.
110. N.L. Gentile, R.F. Louie, J.R. Sifontes, D. Mecozzi, K.R. Hale, G.J. Kost, 'Standardization, Harmonization, and Realization', *Point Care* (2008) [In Press].

FURTHER READING

G.J. Kost, S.S. Ehrmeyer, B. Chernow, J.W. Winkelman, G.P. Zaloga, R.P. Dellinger, T. Shirey, 'The Laboratory–Clinical Interface: Point-of-care Testing', *Chest*, **115**, 1140–1154 (1999).

S. Kulrattanamaneeporn, M. Tuntideelert, G.J. Kost, 'Using Telemedicine with Point-of-care Testing to Optimize Health Care Delivery in Thailand', *Point Care: J. Near-Patient Test Technol.*, **5**, 160–163 (2006).

APPENDIX

Table A CLIA-Waived Tests (109 Total Tests)

Analyte Name
Adenovirus
Aerobic/anaerobic organisms – vaginal
Alanine aminotransferase (ALT) (SGPT)
Albumin
Albumin, urinary
Alcohol, saliva
Alkaline phosphatase (ALP)
Amines
Amphetamines
Amylase
Aspartate aminotransferase (AST) (SGOT)
B-type natriuretic peptide (BNP)
Barbiturates
Benzodiazepines
Bilirubin, total
Bladder tumor associated antigen
Blood lead
Calcium, ionized
Calcium, total
Cannabinoids (THC)
Carbon dioxide, total (CO_2)
Catalase, urine
Chloride
Cholesterol
Cocaine metabolites
Collagen type i crosslink, *n*-telopeptides (NTX)
Creatinine
Erythrocyte sedimentation rate, nonautomated waived
Estrone-3 glucuronide
Ethanol (alcohol)
Fecal occult blood
Fern test, saliva
Follicle stimulating hormone (FSH)
Fructosamine
Gamma glutamyl transferase (GGT)
Gastric occult blood
Gastric PH
Glucose

Table A *(continued)*

Analyte Name
Glucose monitoring devices (FDA cleared/home use)
Glucose monitoring devices (prescription use only)
Glucose, fluid (approved by FDA for prescription home use)
Glycated hemoglobin, total
Glycosylated hemoglobin (HGB A1C)
HCG, urine
HDL cholesterol
Helicobacter pylori
Helicobacter pylori antibodies
Hematocrit
Hemoglobin
HGB, single analyte inst. W/self-cont. . .
HIV antibodies
HIV-1 and HIV-2 antibodies
HIV-1 antibody
Infectious mononucleosis antibodies (mono)
Influenza A
Influenza A/B
Influenza B
Ketone, blood
Ketone, urine
Lactic acid (lactate)
LDL cholesterol
Leukocyte esterase, urinary
Lithium
Luteinizing hormone (LH)
Lyme disease antibodies (*Borrelia burgdorferi* ABS)
Methadone
Methamphetamine/amphetamine
Methamphetamines
Methylenedioxymethamphetamine (MDMA)
Microalbumin
Morphine
Nicotine and/or metabolites
Nitrite, urine
Opiates
Ovulation test (LH) by visual color comparison
Oxycodone
Phencyclidine (PCP)
Platelet aggregation

Table A (*continued*)

Analyte Name
Potassium
Propoxyphene
Protein, total
Prothrombin time (PT)
Respiratory syncytial virus
Semen
Sodium
Spun microhematocrit
Streptococcus, group A
Thyroid stimulating hormone (TSH)
Trichomonas
Tricyclic antidepressants
Triglyceride
Urea (BUN)
Uric acid
Urine dipstick or tablet analytes, nonautomated
Urine HCG by visual color comparison tests
Urine qualitative dipstick ascorbic acid
Urine qualitative dipstick bilirubin
Urine qualitative dipstick blood
Urine qualitative dipstick chemistries
Urine qualitative dipstick creatinine
Urine qualitative dipstick glucose
Urine qualitative dipstick ketone
Urine qualitative dipstick leukocytes
Urine qualitative dipstick nitrite
Urine qualitative dipstick PH
Urine qualitative dipstick protein
Urine qualitative dipstick specific gravity
Urine qualitative dipstick urobilinogen
Vaginal PH

Source: www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCLIA/analyteswaivted.cfm, Accessed on July 20, 2008.

Guidance for Industry and FDA Staff: Recommendations for Clinical Laboratory Improvement Amendments of 1998 (CLIA) Waiver Applications for Manufacturers of *in vitro* Diagnostic Devices. Source: http://www.fda.gov/cdrh/oivd/guidance/1171.html Document Issued: January 30, 2008

Product Development for the Clinical Laboratory

Jan S. Krouwer
Bayer Diagnostics, Medfield, USA

1 Introduction **1**
1.1 Brief History of Clinical Assays 1
1.2 How the AutoAnalyzer Changed the Landscape 2
1.3 The Current Technical, Commercial, Medical, and Regulatory Dimensions of the Product Development Environment 2
2 Product Development as Part of the Commercial Hierarchy **3**
2.1 Company Vision, Mission, and Business Strategy 3
2.2 Technology Strategy 3
2.3 Use of Portfolio Analysis to Select Research and Development Products to Fund 3
3 The Five Stages of Product Development **3**
3.1 Overview of Each Stage 3
3.2 How Companies Manage the Process 4
3.3 Key Metrics to Monitor Progress 4
4 Stage I: Preliminary Assessment: Researching New Opportunities **5**
4.1 The Spectrum of Basic Research to Product Licensing 5
4.2 The Use of Modeling and Empirical Experimentation 5
4.3 Use of Decision Analysis to Value Individual Product Opportunities 5
5 Stage II: Detailed Assessment: Proving Feasibility **6**
5.1 Specifying the Product with Focus Groups and Conjoint Analysis 6
5.2 Quality Function Deployment 8
5.3 Prototyping Techniques for Instruments, Software, and Reagents 8
5.4 What is Feasibility and Experimental Plans to Prove Feasibility 9
6 Stage III: Scheduled Development **9**
6.1 Introduction 9
6.2 The Use of Experimental Design to Build Robust Assays Faster 9
6.3 Reliability Methods for Instrumentation 10
6.4 Design for Assembly 10
7 Stage IV: Validation **10**
7.1 Internal Validation 10
7.2 Customer Trials 11
7.3 Tradeoffs Between Product Launch Date and Features 11
8 Stage V: Commercialization: Product Launch **11**
8.1 The Difference Between Product Claims and (Internal) Specifications 11
8.2 Regulatory Requirements 12
8.3 Product Support Strategies 12
Abbreviations and Acronyms **12**
Related Articles **13**
References **13**

Clinical assays have an important role in medical decision making. Test results help a clinician answer a fundamental question, ''how should I treat the patient?'' Whereas these assays used to be developed in hospital laboratories and were performed by skilled technicians, the availability of automated instruments has created an industry which now provides most of the assays. The industry is highly competitive – development of assay systems requires a product development process that is efficient. Most companies use a development process that breaks product development into stages: (1) researching new opportunities, (2) proving feasibility, (3) scheduled development, (4) validation, and (5) product launch. Each stage has methods and tools that help attain the commercial goal.

1 INTRODUCTION

1.1 Brief History of Clinical Assays

Although clinical assays of some form or another can be traced back to antiquity, it was not until the mid-1800s, as the knowledge of chemistry improved, that urine chemistry tests were developed for glucose and urea.[1] Tests for blood samples were enabled by the invention of the hypodermic syringe in 1845. Assays for glucose in blood were aided by the development of the colorimeter. Jaffe's picric acid reagent for creatinine was published in 1886. In the 1920s and 1930s a spectrophotometric assay of nicotinamide adenine dinucleotide (reduced form) (NADH) was used to determine LDH (lactate dehydrogenase). During the 1950s, a variety of different

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

technologies were invented or perfected, including blood-gas analysis, flame photometry, and electrophoresis of proteins. Immunoassays became possible owing to the invention of radioimmunoassay by Berson and Yallow.

A salient feature of virtually all assays performed during the 1950s was the labor-intensive and technique-sensitive properties of the assays. This restricted the number of assays that could be reasonably performed and also caused quality problems.

1.2 How the AutoAnalyzer Changed the Landscape

In 1957, Skeggs provided the basis for a new industry with the invention of the AutoAnalyzer.[2] The concept was commercialized by Technicon Instruments.

Early automated analyzers were continuous-flow devices. Samples were aspirated through tubing connected to a pumped stream of reagents. The tubing was coiled to allow time for the reagents to react, after which the solution passed through a flow cell which was part of a colorimeter.

The value of the AutoAnalyzer and the instruments that followed, including those used today, is the automation of difficult, labor-intensive steps. This has made laboratory assays cheaper to perform and improved their quality. A typical clinical chemistry device can be considered a robot (or series of robotic devices) controlled by a computer that delivers and moves precise amounts of fluid. With automation, the volume of tests skyrocketed and the founder of the company best known for the initial automation – Technicon – became one of the wealthiest men in the world.

In those days, product development was much simpler than today. Assays could be designed by a single person (the pieces of an AutoAnalyzer cartridge could be assembled like Lego blocks and, although somewhat of an exaggeration, it was boasted that an assay could be developed in the morning, tested in the afternoon, and introduced into the customer training laboratory that evening). Since these early days, the science has advanced, products are more complex, and competition and regulation are part of the landscape. Technicon lost its leadership in the diagnostics field and was sold several times. The purpose of this article is to describe how product development for clinical assays is carried out in this new environment.

1.3 The Current Technical, Commercial, Medical, and Regulatory Dimensions of the Product Development Environment

1.3.1 The Technical Environment

The medical device industry's biggest technical challenge is not mastery of any single technology but rather the *integration* of many complex and diverse technologies such as:

- software including real-time systems, user interfaces, complex algorithms that convert raw signals to concentrations, databases, and communication software;
- hardware including electromechanical, optical, and fluidic systems;
- biochemical reagents, many of which are unstable to heat or light.

Competitive pressures require companies continually to launch new products to remain successful. Technical management is challenged with reducing the length of the product development cycle while meeting product and quality specifications.

1.3.2 The Competitive Environment

The medical device industry was a $20.5 billion industry in 1997 and is expected to grow to $27.5 billion in 2002. As with many competitive environments, successful companies need to anticipate technical trends or technical upsets, and also changing customer demands. Products need to have short development cycles. Since many governments are trying to lower healthcare costs, medical device systems have to be productive. The industry has favored consolidation, so competitors often can leverage expertise in several technical domains and have international sales organizations with efficient distribution channels.

There is a range of product opportunities from commodity types such as routine clinical chemistry analyzers that assay for substances such as cholesterol and glucose to new markers derived from genomics research. Marker concentrations range from millimolar to the ultimate goal of single-molecule detection.

Although there are many patents in the field, few confer a long-term competitive advantage [hepatitis C virus (HCV) is an exception]. Most new products are replicated in some fashion by competitors, hence there is a narrow window of higher profits followed by a return to a near commodity market.

There is a range of ease of sample preparation with the most complex being serum and plasma, followed by whole blood (centrifugation of the sample not needed), then urine, whole blood and urine home use, and finally non-invasive testing.

Market segments include large, national reference laboratories, hospital laboratories, hospital satellite laboratories (''point of care testing'' within the hospital), doctors' offices, and home testing. Each market segment has different primary and secondary customers to be satisfied, which include:

- end users, i.e. the people who conduct tests in hospital laboratories, reference laboratories, doctors' offices, and the home;
- all people involved in the buying decision, ranging from group purchasing agents to home users;
- clinicians and patients;
- regulatory agencies.

2 PRODUCT DEVELOPMENT AS PART OF THE COMMERCIAL HIERARCHY

2.1 Company Vision, Mission, and Business Strategy

Most companies have a hierarchical method of running their business; they include a vision and mission statement, a business strategy, a technology strategy, and a product development process (see Figure 1). The position of the hierarchy depicted in Figure 1 is proportional to the amount of effort required to create each item. Thus vision statements are easy to create, mission statements more difficult, and strategies even more difficult. It has been said that a vision without a strategy is simply a dream.

2.2 Technology Strategy

There is often no clear demarcation between a business and a technical strategy. In some companies a business strategy contains the technology strategy. Either a business or a technology strategy can provide a leadership role. For example, if a business strategy specifies being an efficient "follower", then efficiencies will be valued not innovative technologies. Yet, if there are technical strengths in the organization, a business strategy may revolve around technology. In either case, valuing individual opportunities and the entire portfolio of opportunities *selected* is a key task of a technology strategy.

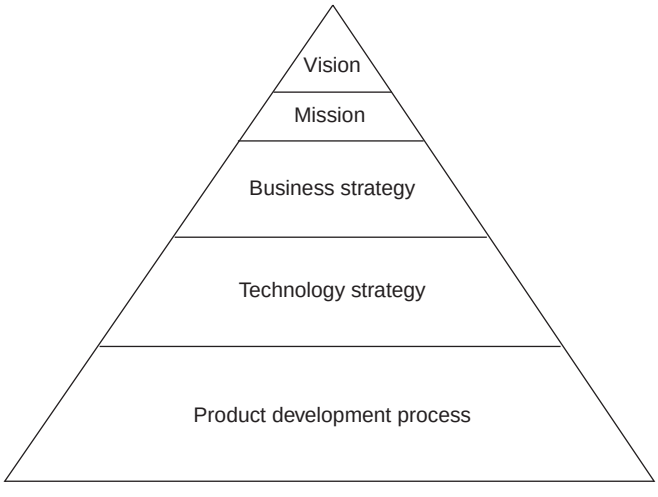

Figure 1 Hierarchical nature of business processes.

2.3 Use of Portfolio Analysis to Select Research and Development Products to Fund

Since virtually all companies have limited funds, there has to be a way to select which projects to fund. Decision analysis is a fundamental building block for this task. This method will be discussed in section 4.3.

3 THE FIVE STAGES OF PRODUCT DEVELOPMENT

3.1 Overview of Each Stage

Many companies use a stage gate system where product development is divided into several stages.[3] At the end of each stage, a decision is made based on the results, whether or not to fund the next stage. (Often in practice, these funding decisions are not associated with the stage gate system.) This system is popular, even if it is partially implemented, because it serves as a useful communication vehicle. As consolidation of the industry continues and projects become larger and more complex, this communication aspect is important. Although the names and numbers of stages may differ in different companies, the distinct stages to this process are briefly defined below.

3.1.1 Stage I: Preliminary Assessment: Researching New Opportunities

In this stage, companies decide which projects to commercialize. Generally, this stage does not include basic research, which is usually decoupled from a product development process. On the other hand, applied research is part of the stage. It may consist of evaluating opportunities from within the basic research division of the same company or from other companies or universities. Many products that are commercialized are licensed from external sources, so that often initial studies are directed to understanding the value of these technologies or assessing the likelihood of technical success.

Studies at this stage are often carried out by modeling. Concomitant with these activities is the use of decision analysis to value individual product opportunities and portfolio analysis to select a set of projects to fund.

3.1.2 Stage II: Detailed Assessment: Proving Feasibility

The decision has been made to commercialize the product. The first step, which is often started and in

some cases completed in Stage I, is to specify the product. There is usually a hierarchy of specifications which flow down from concept to user to technical specifications. The basic task is simple: find out what the customer wants. Techniques used include surveys, focus groups, and conjoint analysis. Quality function deployment (QFD) is a method that translates customer requirements throughout the product development process. At the end of Stage II, studies are conducted to ''prove feasibility''. Companies struggle with defining feasibility, since at this stage release requirements are often not met.

3.1.3 Stage III: Scheduled Development

In Stage III, the lion's share of product development takes place. For diagnostic companies, there are basically two major development paths, assays and instruments. Within assay development, designed experiments often yield robust products faster. For instruments, reliability growth management has proven to lead to faster attainment of reliability goals. Software, while part of the instrument development, is an ever-increasing task. System integration, making the reagent–instrument system work, takes place – whether reagents or instrument work independently is of limited value. In manufacturing, design for assembly (DFA) tools are used. Initial customer trials are performed.

3.1.4 Stage IV: Validation

Customer trials continue and become more formal. Concomitantly, extensive internal studies are carried out to validate assay performance, instrument reliability, and ease of use. Software is formally validated. During this stage, there is often intense pressure to launch the product, hence tradeoffs are made between features (including meeting original specifications) and the product launch date.

3.1.5 Stage V: Commercialization: Product Launch

As the product is launched, marketing campaigns position the product, customer training is carried out and product claims are issued which form the basis of warranty policies. Product-support strategies are deployed to deal with customer issues. Since some product features have been deferred to meet a launch date, they are now developed as are cost-reduction programs.

3.2 How Companies Manage the Process

Many companies use the stage gate system described in section 3.1. Organizationally, companies are set-up from centralized research and development (R&D) to decentralized business units, each containing R&D to in-between organizations. Companies need both good processes and good people. Whereas highly skilled managers previously could overcome poor processes, this is becoming more difficult owing to the complexity and larger scope of projects.

Leadership skills require domain knowledge. One does not have to be a domain expert but should have sufficient knowledge of major techniques and strategies (e.g. more than knowing the buzzwords) to be able to define or review technical strategies not only at a company wide level but also at the project level. Relentless and effective questions – using the scientific method – demanding data rather than relying on anecdotal evidence are often features of successful leaders.

Successful companies often have a process which enables good decisions to be made based in part on quantitative analyses and on a review system that provides checks and balances. The use of quantitative analyses to improve decision making does not replace the judgment and intuition needed. Yet, there can be a danger of foregoing the use of quantitative analyses. A manager who asks, ''what does your gut tell you?'', as his or her primary decision tool is often ignoring valuable information that cannot be gained from intuition or judgment alone. On the other hand, it is also unwise to rely solely on quantitative methods. There is an old adage, ''beware of the phenomenon statistics on, brain off''. Hence, judgement is always used and aided with quantitative analyses.

3.3 Key Metrics to Monitor Progress

Whereas scores of metrics are used by program managers to track project progress, there are only a few key metrics that are needed. Ideally, each of these will have some predictive ability. The goal of the metrics is to alert management of problems (schedule, budget) long before they happen so that contingency plans can be deployed. Management's nightmare is to be told of problems the night before a milestone completion date. Examples of key metrics that should be available for review during the entire project (and updated at appropriate intervals) are:

- a net present value (NPV) analysis of the project's product (note that a well-constructed NPV analysis will include all of the bulleted items below);
- the predicted launch date;
- the total development cost and standard cost;
- the predicted service-call rate or other quality measures such as the predicted complaint rate;
- any other key project specific metrics unique to the product, but limited in number.

One metric that should be avoided is ''percent completed'', because it is misleading. If task completion times were uniformly distributed, then percent completion would be meaningful because a project that was 50% complete would require twice as much time to be 100% complete. For technology projects, there are always unanticipated problems which cause task completion times to have a skewed distribution.[4]

Of course, not every important product feature can be quantified. These should not be ignored but their subjective nature should be realized.

4 STAGE I: PRELIMINARY ASSESSMENT: RESEARCHING NEW OPPORTUNITIES

4.1 The Spectrum of Basic Research to Product Licensing

Companies can be innovative in various ways. It makes little difference if a successful product were developed totally internally or licensed as a (virtually) final product.

Basic research attempts to understand phenomena within an area, without specific product goals. Basic research questions can be more narrowly focused to become applied research. With the appropriate staff, organizations can monitor basic or applied research conducted in other companies or in universities, with the idea of licensing. Licensing can include licensing a basic concept which needs significant applied research for commercialization. Alternatively, a licensed technology may not require any significant development.

4.2 The Use of Modeling and Empirical Experimentation

There is a desirable hierarchy to follow for efficient product development. For example, in the technology evaluation phase, if one has sufficient knowledge, physical modeling can be applied to determine whether a new technology will work. Modeling is attractive because it can give answers relatively quickly and inexpensively. For example, if one has dose-precision goals for an immunoassay and a dose–response curve, one can model, using propagation of errors, whether the dose–response curve is adequate to fulfill the precision goals. If there is insufficient knowledge to permit modeling, designed experiments are an efficient way to assess technology. One-at-a-time experiments are also useful when used to build knowledge.

4.3 Use of Decision Analysis to Value Individual Product Opportunities

The goal of companies is to be profitable. The goal of profitable companies is to be more profitable. In high-tech companies, such as those in medical diagnostics, profit is fueled by new products. Yet all new products are not profitable enough, nor can companies afford to develop all conceivable projects.

Decision analysis is a method to provide decision makers with information to help inform the decision on whether to fund a project.[5] It can be broken down into two phases, an initial set-up phase and an ongoing phase used for each subsequent valuation. The initial phase starts with determination of the frame or the boundaries of the opportunity, followed by an influence diagram, which represents a high-level abstraction of the drivers that contribute to the desired output measure: NPV of cash flows. Cash flow is more important than profit because companies can be profitable with negative cash flow and still go out of business. Influence diagrams set the granularity for the data collection needed for the financial model. The influence diagram is converted into a decision tree, which is then converted into a financial model, often in a spreadsheet form. Key to the decision tree is the use of probabilities to deal with uncertainty. Often a 10–50–90 approach is used, where 10 represents a 1 in 10 chance that the value could be as low as some value, and likewise 90 represents a 1 in 10 chance that the value could be as high as some value. The fundamental exercise in decision analysis, which is conducted in the ongoing phase, is to solicit valid data for all of the inputs. This enables a ''base case'' NPV to be calculated.

Sensitivity analysis is a technique to determine which input variables have the most risk towards NPV. For example, if a product were launched early, its NPV might be $100 million and if launched late, its NPV might be $5 million. In this example, time to market is a risk factor for this program. The results of a sensitivity analysis are often graphed in a bar chart known as a Tornado diagram (see Figure 2).

Scenario analysis takes the top risk factors found from a sensitivity analysis and calculates the NPV for all combinations. Thus, if there were four input factors, each with three levels, there would be 81 calculated NPVs. The probability weighted average of NPVs is known as the expected NPV. One can also evaluate the range of NPVs as a measure of risk.

If one has conducted NPV analysis for each opportunity, one can conduct portfolio analysis to select which projects to fund. This is usually done with linear programming algorithms since there will be several constraints to the problem, such as specifying for each year the minimum cash flows and maximum costs. There may be additional constraints related to strategy or discrete jumps in project funding. Graphs to support these analyses include a graph of technical risk vs commercial return, and a graph of cumulative development cost vs cumulative commercial return (productivity curve – see Figure 3).

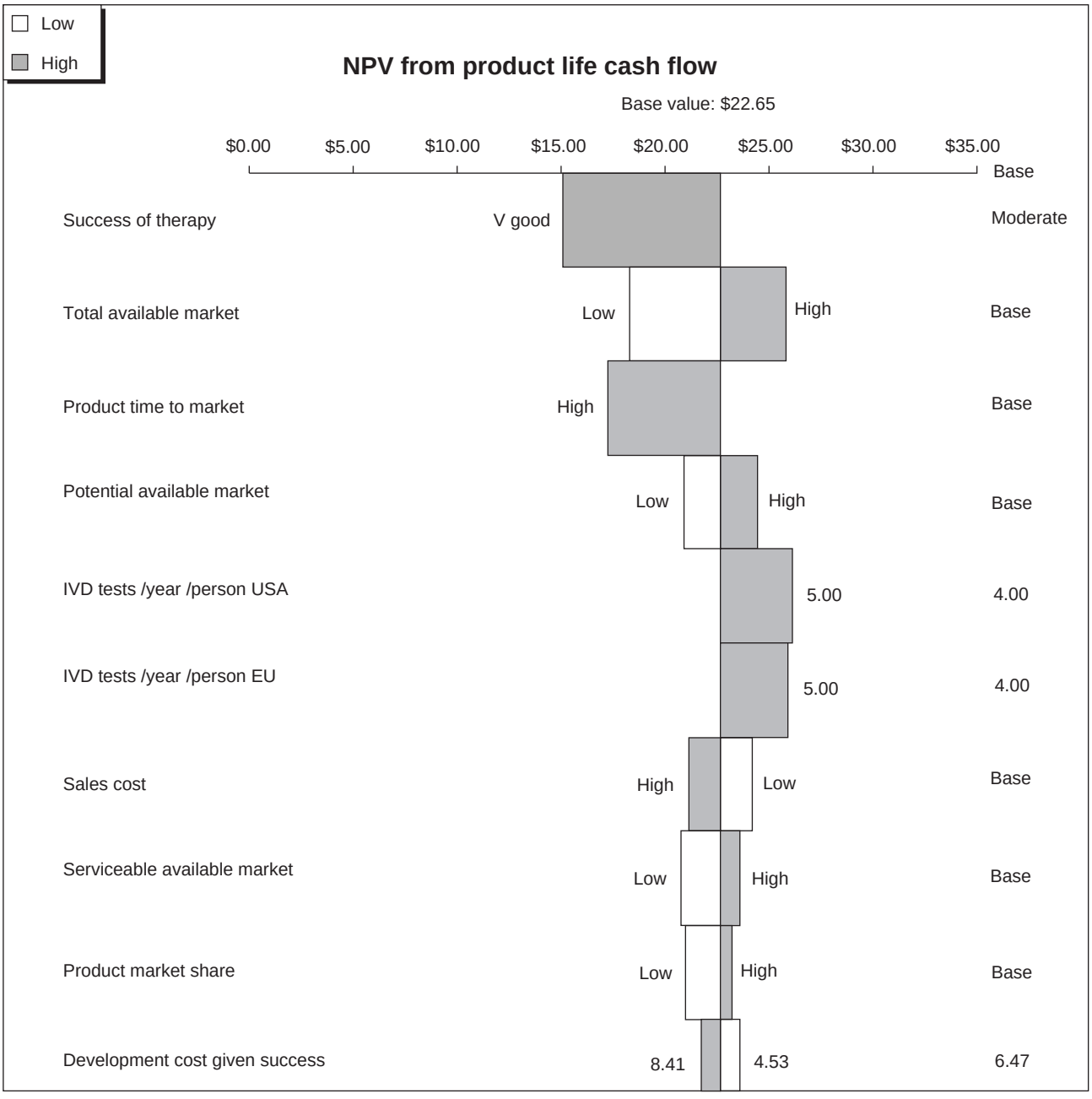

Figure 2 Tornado chart example. IVD, in-vitro diagnostic.

5 STAGE II: DETAILED ASSESSMENT: PROVING FEASIBILITY

5.1 Specifying the Product with Focus Groups and Conjoint Analysis

Engineers and scientists develop a product to meet requirements. These requirements, also called specifications or goals, are important since if they are incorrect the product may fail because it is not what the customer wants. If the specifications are too stringent, product release may be delayed, making the project less profitable.

Whereas it is enticingly simple in concept to find out what customers want, it is surprising difficult in practice. There can be internal barriers to getting this knowledge. For example, if a designer is set on creating a complex design, he or she may be oblivious to finding out if customers value design as an important product attribute. Other organizational issues abound that challenge the

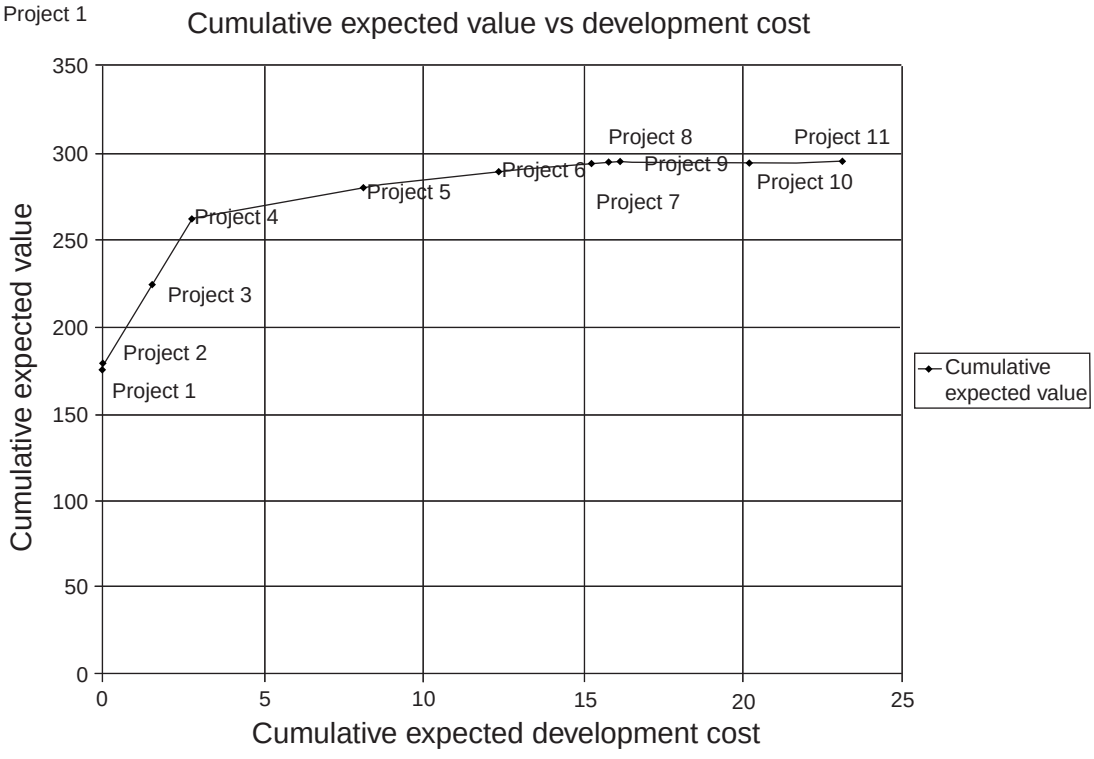

Figure 3 Productivity curve for R&D projects.

process of getting the right requirements. Even when these issues are absent, it is hard for manufacturers to get targets for goals because:

- manufacturers do not know how to ask for customer needs;
- users do not know how to talk about their needs.

Product attributes tend to fall into three categories as shown in Figure 4, often called a Kano diagram.[6] For some attributes, there is a proportional relationship between the quantity of the attribute satisfied and customer satisfaction. As an example, the more menu items are offered for an analyzer, the higher is the degree of customer satisfaction. A second category of attributes is not requested by customers, but when it appears, they are delighted. An example in this category would be the first time remote units were offered with televisions. In the last category, customers expect the attribute to be at a certain level – if it is higher they are not more satisfied but if it is too low, they are very dissatisfied. Product quality often falls into this category.

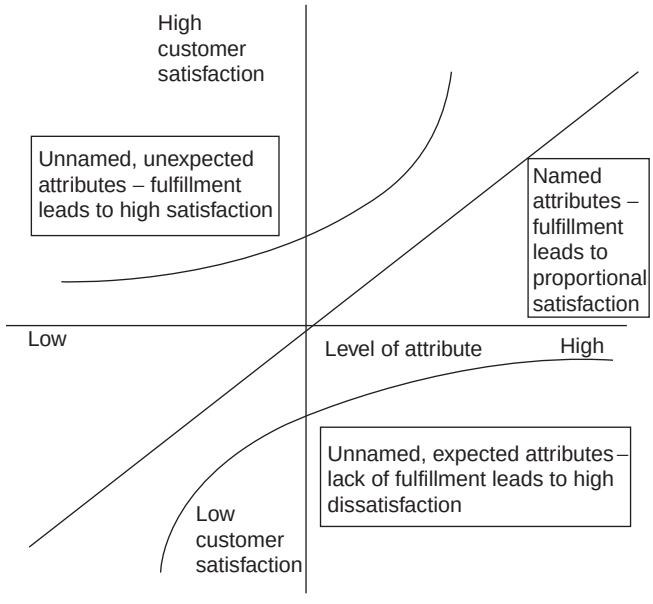

Figure 4 Kano diagram of different types of customer attributes.

5.1.1 Surveys and Focus Groups

Surveys would seem to be an easy way to elicit needs and to set targets. One simply asks customers. However, there are pitfalls. If one asks open-ended questions, such as, ''what are the clinically acceptable limits for a cholesterol assay?'', one could get vague responses such as ''no error'' or ''an error that does not cause diagnostic problems''. One remedy to this is to offer

multiple-choice questions. However, this is not foolproof either. Responses can be checked off without a guarantee that the question was correctly understood by the respondent. Moreover, respondents tend to want ''the best''. Thus, given a choice for the total error for cholesterol to be 1, 3, or 5%, many respondents will simply check off 1%. Skendzel et al. overcame many of these difficulties in a cleverly constructed survey.[7] Focus groups often consist of prototype products shown to customers with a description of features explained. There is then a discussion of the product led by a facilitator who ensures that questions get answered, there is no bias, and there is proper balance among the respondents.

5.1.2 Conjoint Analysis

In real life, one must make tradeoffs. One wants a car that is both luxurious and low priced. For laboratory assays, one wants low total error, low cost, high ease of use, high reliability, etc. For any of these situations, there will be acceptable compromises among the desired values of the goals. Conjoint analysis (also called multiattribute utility analysis) is a form of marketing research that provides a protocol and analysis method for estimating these tradeoffs.[8] Its idea is to present a customer with a series of assays, each with different values for various attributes. The customer ranks his preference for each of the assays. With several customers performing this ranking, the value of each attribute can be found by statistical analysis.

5.1.3 Use of Current Data

Given an assay that is in service, one can ask whether the complaint rate is low enough. A ''yes'' answer may signify that the assay's performance and reliability are adequate. One can then use this performance as a goal for a new version of the same or a similar assay. Of course, this method will not work for new analytes, for which there are no data. Moreover, the problem has not really been solved, it has been transferred. One must now decide what is a low-enough complaint rate.

5.1.4 Aids to Improve Goal Setting

Cause and effect diagrams (also called fishbone or Ishakawa diagrams) help to highlight potential problems and to focus goal setting for quality. The universe of potential quality goals must be narrowed down to those that have a reasonable likelihood of causing problems. Otherwise, the list of goals to be tested could be endless. Reviewing target settings is also important. For example, the goal of zero software bugs is inappropriate. Realistically, there will always be a measurable frequency of software bugs, even for the best systems. A goal of zero is unrealistic and not useful. A nonexistent goal, while a rather obvious category, crops up occasionally. An example is the lack of an outlier goal. Outliers are values that are so far away from the true values that they almost always cause problems. Yet, there is seldom a goal describing how far off a value must be to be called an outlier and how many outliers are acceptable. A goal requires a clear pass–fail criterion. This implies that a metric is in place which will be used to base the decision. Deciding on a metric is a step in agreeing on goals. A metric should be objective, easy to understand, and relevant to the goal.

5.2 Quality Function Deployment

One of the challenges in product development organizations is the variety of teams and personalities involved. There are people who are creative, those who are powerful and persuasive, and so on. To design and build a product that truly satisfies customer needs requires that the customer needs are always the most important priority in design decisions. But consider a powerful functional group (industrial design, for example). If the physical shape of a product is not ranked as important to customers, the powerful design group may nevertheless spend a great deal of time to design the perfect physical shape. This could constrain more important activities by draining scarce resources. QFD attempts to solve this problem by using tools to ensure that what is always heard is ''the voice of the customer'', not the voice of the manager.[9,10]

5.3 Prototyping Techniques for Instruments, Software, and Reagents

Prototypes are used in many of the product development domains. Prototypes are used to communicate ideas to management and potential customers and as a means of learning. Whereas modeling is a form of prototyping, physical prototypes represent an opportunity to expose problems that might be difficult to model.[11]

For hardware, prototypes are built with various levels of functionality focus on elements that require more information. For example, in blood-gas instruments, a new sample-entry method was tested as a prototype because its interaction with the user was different from previous models. In software, user interfaces are prototyped, often with languages such as Visual Basic that can be used by nonprogrammers.

Assays are also prototyped by developing ''laboratory'' lots of reagents on a breadboard or systems designed to model the intended platform.

5.4 What is Feasibility and Experimental Plans to Prove Feasibility

On an abstract level, feasibility is proving that something is technically possible, or conversely, proving that something is *not* technically possible. In these cases modeling based on physical laws will help.

On a more practical level, companies often struggle with the definition and measurement of feasibility. For example, if a product has been shown to meet all of its performance requirements, it is certainly ''feasible'', but if so why would it not at the same time be released for sale. Sometimes feasibility is defined so that a product must meet most of its specifications and be close to meeting the others. Yet, can we be sure that it will ever meet the other specifications? Another variation is to define feasibility as being able to do it once (e.g. with hand-machined parts or hand-formulated reagents). Again, this may not mean much if manufacturing can never make the product.

The bottom line is that each organization must use judgment to measure feasibility whatever definition is employed. For example, if some specifications are allowed to be missed, there must be a believable pathway to meet these specifications. This requires management to have sufficient domain knowledge.

What is still at risk after feasibility is not whether the product can be made but rather when it will be released. Thus, the risk has shifted from no product at all to one that is too late.

6 STAGE III: SCHEDULED DEVELOPMENT

6.1 Introduction

Whereas it would seem simple to turn loose a bunch a scientists and engineers to build products – that is what they are trained to do – there are usually management and other challenges which make this process more complicated. As mentioned, there are huge pressures to reduce the product development cycle time. Often ''quick'' experiments are favored over longer ones, even if the longer one ultimately lead to a shorter cycle time.

6.2 The Use of Experimental Design to Build Robust Assays Faster

Robust products are those that not only meet specifications but do so with each lot of raw materials and process variation.

Most scientists are not trained in the use of experimental design. They often work alone on problems, using creative thinking and hard work. Experiments are their best idea of a solution – this can and often does go on for long periods of time.

This type of experimentation does have a role in industry – sometimes simply trying one's best idea or using intuition will result in the fastest solution. This is particularly true when the state of knowledge is low. If, however, the state of knowledge is higher – but not high enough so that computer models can be built from physical laws – then empirical experimentation can be the fastest way to build robust products. Building robust products through empirical experimentation involves:

- *Flow charting* – A diagram of the process is constructed, where each step is identified. This works for manufacturing processes such as reagent formulation and also for the conceptual process of a sample moving through an instrument system.
- *Cause and effect diagrams* – Often facilitated through brainstorming sessions, cause and effect diagrams are similar to fault trees in reliability. They attempt to enumerate in a diagram all possible failure modes for each identified subsystem.
- *Screening experiments* – Given a large (8–20) possible number of variables that could affect a response, screening experiments are an efficient way to determine which of the factors have no effect and which have the most influence.
- *Factorial experiments* – Traditional experiments, also called one-at-a-time experiments, involve holding all factors constant except the one under study. This is repeated with all other factors. The end result is an ''optimized'' system. However, it has been shown that a more efficient experiment plan is to perform factorial and fractional factorial experiments. One-at-a-time experiments result in $(k+1)/2$ as many runs as factorial designs, where k is the number of variables. Moreover, there are cases where an optimum is much more likely to be reached with factorial experiments.[12]
- *Response surface experiments* – Response surface experiments allow more accurate models to be constructed by modeling second-order effects. This allows empirical equations to be developed which can serve as accurate response predictors within the response surface modeled.
- *Taguchi-type experiments* – These experiments are in fact factorial experiments, but have been developed to ensure that variation is modeled as a response in addition to location. The principle is that with natural manufacturing and raw material variations, the product should behave consistently. Formulation settings can be developed based on these experiments to provide product performance consistency.

6.3 Reliability Methods for Instrumentation

If our state of knowledge of medical diagnostic instrumentation were higher, we could design reliable systems on a computer. When actual systems were built, little testing would be required – in fact, the only test would be a demonstration test that essentially verified the computer model. Whereas this practice has been used in other industries such as for designing and building commercial airplanes, it is not appropriate for medical diagnostic systems because our state of knowledge is not high enough and the product design cycle is too short.

With these constraints, what has proven successful is use of the reliability growth management (MIL-HDBK 781a).[13] This process uses the test, analyze and fix (TAAF) process in which systems are tested in order to expose problems. A corrective action process is established to fix the problems. The process continues until the reliability goal has been met.

Other key elements are:

- *A reliability model* – The reliability metric that is measured after product release must be modeled so that it can be measured in-house. For example, unscheduled service calls may be the reliability metric that the business uses. It would be tempting to model this internally by simply counting how many times a gross failure occurs (e.g. power supply burn out). Yet if some less severe error occurs, from which a customer could recover, and this error occurs too frequently, this problem will also result in a service call and its behavior must be modeled.
- *Failure review analysis and corrective action system (FRACAS)* – This process, often supported by software, reviews the failures that have occurred. Failures are categorized and assigned to appropriate engineers to be fixed. A key element of this step is to prepare Pareto charts of failure modes so that problems to be tackled are selected which will have the biggest impact on improving reliability. When the corrective action is implemented, its effectiveness is monitored.
- *Idealized growth curve* – Given certain assumptions about the problem obtained from initial or historic data, a curve is constructed projecting reliability improvement with instrument cycles (e.g. cycles can be translated to calendar time).
- *Tracking curve* – Actual progress is tracked against the idealized growth curve using Duane or Army Material Systems Analysis Activity (AMSAA) models.

6.4 Design for Assembly

A system that is difficult to assemble will take longer to assemble and be more likely to be incorrectly assembled, both of which lead to higher manufacturing costs. DFA is a process that is used in design to reduce the number of parts and simplify the assembly process.

7 STAGE IV: VALIDATION

7.1 Internal Validation

Internal validations, often conducted by an internal, independent assessment group, are conducted to determine whether the product has met its specifications. There is still tremendous pressure to launch the product, so an efficient set of experiments is required. While specification assessment is the goal, ideally the evaluation will also find out any other problems that might occur with the product (which may not be contained in the set of specifications).

The design of the evaluation protocols has been an area of statistical development[14,15] and standard committees, e.g. National Committee for Clinical Laboratory Standards (NCCLS). As an example of an efficient protocol, the NCCLS protocol EP10A assesses five performance parameters, during each ten-sample run (three replicates of three concentration levels with a primer).[16] The parameters assessed include precision, bias, sample-to-sample carryover, linear drift, and nonlinearity. The key to this protocol is the sample order, which confers a near orthogonal relationship among the variables. Orthogonality is a desirable feature because it allows for ease of interpretation of results.

The set of protocols that are routinely performed to validate a system include:

- *Diagnostic efficacy* – Needed only for new markers, this expensive study establishes the proportion of test answers that correspond to the medical diagnosis (and the proportion of wrong answers, i.e. false positives and false negatives).
- *Method comparison* – Consists of running real or simulated patient samples across the range of the assay over a number of days. The comparison method can be a reference method or popular commercial method. The slope (proportional bias), intercept (constant bias), and scatter about the regression line are estimated.
- *Precision studies* – These are part of the method comparison but also run separately with controls to assess within-run and total precision. Total precision is the precision expected in ''long-term'' use. It ideally would include all possible sources of variation (hence the word ''total''). In reality, sources of variation such as different reagent and calibrator lots are often not included in this estimate.

- *Interference studies* – Whereas we would like to believe that assays are totally specific, virtually all assays suffer from interferences. Protocols to test for interferences usually involve running replicates of a sample spiked with the desired level of interference, compared with the sample spiked with diluent (the control). Establishing the candidate list of interferences requires knowledge of the assay weaknesses in addition to a current list of endogenous and exogenous substances in blood and their concentrations.
- *Sensitivity determination* – Some assays require an extremely low detection limit (for example, in infectious disease assays a single viral particle is sufficient to indicate disease). Sensitivity can be viewed as a precision assessment at zero concentration – the protocol involves running a set of zero concentration samples and or a dilution experiment starting at a low concentration.
- *Dilution* – Some substances have an observed range that spans several orders of magnitude. Human chorionic gonadotropin (hCG) is an example with an expected value of zero, a decision limit of 10 and an abnormal value of 10 000 or more. Often systems cannot directly assay samples spanning this range because the assay response curve does not cover this range. At some level of concentration, the dose–response curve is so flat that precise quantification is impossible. The system has to be able to measure samples after they have been diluted with various ratios of sample to diluent. It is expected that linearity is preserved as a function of dilution.
- *Other special studies* – Random access.

7.2 Customer Trials

Whereas customer trials are conducted throughout the product development cycle, there are issues which make their value somewhat limited. For products that have not been released, customers in the USA cannot report patient results. Thus, the answers provided by the assay cannot be used to help with medical decisions. Moreover, the customer has not purchased the system and it is not under warranty. Therefore, although the product is ''in the hands of the customer'' it is often paradoxically not used by the customer as it would be if the customer had actually purchased the system. Often the data collected at these trials are analyzed solely by the manufacturers, with the customer having a limited role in running samples. Note that a released product that is not approved by the Food and Drug Administration (FDA) may nevertheless be used to report patient results under an ''investigational use'' exemption.

These studies are nevertheless conducted and at times uncover issues that have been missed. The failures to detect these problems in-house may at times be due to unrepresentative sampling, or by using the system differently than a customer. For example, internally based technicians have learned how to work with new systems that have problems. These ''work-arounds'' are often complex and, when the system is field tested, the customer performs the work-arounds incorrectly, causing a system failure. Other issues could be uncovered by assaying samples that are not tested in-house. Testing samples in-house requires different processing to samples tested in a hospital, since shipment and storage are required for in-house samples.

7.3 Tradeoffs Between Product Launch Date and Features

Time-to-market pressure and technical problems force tradeoffs between launching the product ''as is'' or delaying launch to put in more features. The tradeoffs can be modeled with decision analysis tools.

8 STAGE V: COMMERCIALIZATION: PRODUCT LAUNCH

8.1 The Difference Between Product Claims and (Internal) Specifications

During product development, there is a set of product specifications that are used internally. When the product is released, the internal goals are transformed by the manufacturer into customer claims, which may or may not be the same as the internal goals. The claims represent a data source for customers, who can compare different manufacturers' claims. Of course, as most consumers are aware, not all claims are always met! There is an additional problem. A claim can be stated in a way that is not clear, leading to confusion. Basically there are two types of claims:

- The ''typical'' data claim – Half of the customers are expected to observe better performance and the other half worse performance. With statistical tests, one can determine whether observed performance is consistent with the manufacturer or unreasonably far from the claim.
- The ''warranty'' claim – Here, all customers are guaranteed performance better than the limit. If customers don't achieve this level, they are told to contact the service department.

The advantage of using the first method is that claims will appear better, e.g. typical precision 3% coefficient

of variation (CV) instead of less than 5% CV. The disadvantage is that for a customer to verify performance, statistical tests must be conducted. From a manufacturer's standpoint, this can lead to complaint calls that are not real. The second method is easier for both the customer and manufacturer but requires understanding that a claim stated this way does not look as good as stated in the first method (although in reality they are representing the same performance).

8.1.1 Pass–Fail Limits vs Continuous Quality Improvement

Manufacturers, like mostly everyone, are faced with yes–no decisions. Should we release or not release the product (meaning has it met or not met its goals)? However, consider two assays, where assay A is just inside and assay B is just outside specification. From a manufacturer's standpoint, assay A has full value (i.e. identical with an assay that is perfect) whereas assay B has zero value. From a customer standpoint (assuming that the customer could observe performance from the assay that failed the manufacturer's specification), the two assays have similar performance (and value). Yet manufacturers must still treat these two assays, that have similar value to a customer, as either having full or zero value.

There is no easy way to deal with this problem. What happens in practice is that the yes–no region, while conceptually clear (i.e. accept if precision is $\leq$4.0% CV, reject if >4.0% CV) becomes fuzzy: accept if precision $\leq$4.0% CV; conduct further discussions if precision is between 4.0 and 5.0% CV; and reject if precision is >5.0% CV. Thus, targets set at the beginning of a project are revisited throughout the project development cycle and especially near product release. Since many specifications are set close to the technical capability of a system (owing to competitive pressures), the situation occurs frequently.

8.2 Regulatory Requirements

While not having the same scrutiny or oversight as required for new drugs, new medical devices require (with different levels of effort depending on the device) regulatory approval in most countries. Assays that are most regulated are tumor markers and infectious diseases. FDA approval of these assays requires a substantial effort. Instrument systems also require a variety of approvals related to safety or demonstration of lack of interference to other electronic equipment. Medical device manufacturers have become International Organization for Standardization (ISO) certified. In the USA, regulation continues after product launch at both the manufacturer and the customer site. Manufacturers are periodically inspected by the FDA. If problems arise, warnings letters are issued with product recall a possibility. Medical device reporting is a regulation aimed at getting manufacturers to report any possible threat to patient safety due to use of the device (e.g. whether or not they have actually occurred). Hospital laboratories are routinely inspected by accreditation groups.

8.3 Product Support Strategies

Service strategies can play an important role in a company's overall strategy, since customer contact with a company's service department can affect the customer's level of satisfaction and affect future buying decisions. Often this contact occurs when the customer is frustrated. Surveys devoted to assessing customer satisfaction with manufacturers are often a measure of satisfaction with service rather than a measure of the customer's regard for the product itself.

An example of a service strategy follows – but includes required items outside the service group's control. Thus, the service strategy must be built into product development. The key elements are:

- a diagnostic strategy for instruments;
- a means to assess an instrument's status remotely;
- a means to correct faults.

Complex medical instruments require a diagnostic strategy consisting of sensors to record key measures in each subsystem. What needs to be measured is developed from fault trees and a failure mode effect and criticality of action (FMECA). This analysis also dictates how the software will respond when measured values are outside of limits. Measurement results are stored in a database that can be accessed for trend analysis.

Trends need to be monitored not only on the instrument in question but across all instruments. This is facilitated in a variety of ways. If the instrument has a modem, the database can be queried remotely. The Internet can also be used, provided that the instrument is on a network. Remote monitoring can be a two-way communication so that if a fault is detected that requires customer action, the customer can be notified remotely about what action to take.

ABBREVIATIONS AND ACRONYMS

AMSAA	Army Material Systems Analysis Activity
CV	Coefficient of Variation

DFA	Design for Assembly
FDA	Food and Drug Administration
FMECA	Failure Mode Effect and Criticality of Action
FRACAS	Failure Review Analysis and Corrective Action System
hCG	Human Chorionic Gonadotropin
HCV	Hepatitis C Virus
ISO	International Organization for Standardization
IVD	In-vitro Diagnostic
LDH	Lactate Dehydrogenase
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
NCCLS	National Committee for Clinical Laboratory Standards
NPV	Net Present Value
QFD	Quality Function Deployment
R&D	Research and Development
TAAF	Test, Analyze and Fix

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Biosensor Design and Fabrication • Immunochemistry • Laboratory Instruments in Clinical Chemistry, Principles of • Nucleic Acid Analysis in Clinical Chemistry • Point-of-care Testing • Statistical Quality Control in Clinical Laboratories • Urinalysis and Other Bodily Fluids

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

REFERENCES

1. M.K. Schwartz, 'From Antiquity to the 21st Century: the Past and Future of Clinical Chemistry', http://www.aacc.org/divisions/history/ind/resource.html.
2. L.A. Lewis, 'Leonard Tucker Skeggs – A Multifaceted Diamond', *Clin. Chem.*, **27**, 1465–1468 (1981).
3. R.G. Cooper, *Winning at New Products*, Addison-Wesley, Reading, MA, 1986.
4. J.S. Krouwer, 'Beware the Percent Completion Metric', *Res. Technol. Manage.*, **41**, 13–15 (1998).
5. D. Matheson, J. Matheson, *The Smart Organization. Creating Value Through Strategic R&D*, Harvard Business School Press, Boston, MA, 1998.
6. N. Kano, 'Attractive Q. vs. Must be Q.', *Hinshitsu*, **14**, 39–48 (1984).
7. L.P. Skendzel, R.N. Barnett, R. Platt, 'Medically Useful Criteria for Analytic Performance of Laboratory Tests', *Am. J. Clin. Pathol.*, **81**, 200–205 (1985).
8. P.E. Green, V. Rao, 'Conjoint Measurement for Quantifying Judgmental Data', *J. Market. Res.*, **8**, 355–363 (1971).
9. J.R. Hauser, D. Clausing, 'The House of Quality', *Harvard Bus. Rev.*, **66**, 63–73 (1988).
10. S. Marsh, J.W. Moran, S. Nakui, G. Hoffherr, *Facilitating and Training in Quality Function Deployment*, GOAL/QPC, Methuen, MA, 1991.
11. K.T. Ulrich, S.D. Eppinger, *Product Design and Development*, McGraw-Hill, New York, 1995.
12. G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters: an Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, New York, 1978.
13. *Handbook for Reliability Test Methods, Plans, and Environments for Engineering, Development, Qualification, and Production*, MIL-HDBK-781a, SPAWAR, Arlington, VA, 1996.
14. C. Daniel, 'Calibration Designs for Machines with Carryover and Drift', *J. Qual. Technol.*, **7**, 103–108 (1975).
15. C.C. Garber, R.N. Carey, 'Evaluation of Methods', in *Clinical Chemistry, Theory Analysis Correlation*, 3rd edition, eds. L.A. Kaplan, A.J. Pesce, Mosby, St. Louis, MO, 1996.
16. *Preliminary Evaluation of Quantitative Clinical Laboratory Methods, Approved Guideline*, NCCLS Approved Guideline EP10-A, National Committee for Clinical Laboratory Standards, Wayne, PA, 1998.

Serum Proteins

Robert F. Ritchie
Foundation for Blood Research, Scarborough, USA

John T. Whicher
University of Leeds School of Medicine, Leeds, UK

1 Introduction 1
1.1 Definition of ''Protein'' 2
1.2 Definition as it Applies to ''Serum'' Proteins 3
2 History 3
3 Physiology and Pathophysiology 4
3.1 Synthesis 4
3.2 Distribution 5
3.3 Catabolism 6
3.4 Function 7
4 Methods 8
4.1 Immunological Reagents 8
4.2 Immunoassays 10
4.3 Biospecific Affinity Chromatography 15
4.4 Chemical Assay 15
4.5 Physical Properties 16
4.6 Electrophoresis 22
5 Patient and Sample Preparation 26
5.1 Preanalytical Requirements 26
5.2 Intra-analytical Requirements 27
5.3 Postanalytical Requirements 27
6 Limitations 28
6.1 Concentration 28
6.2 Secondary Factors 28
6.3 Proper Reference Ranges 30
6.4 Reference Materials and Calibrators 30
6.5 Other Effects 30
7 Quality Control 31
7.1 Operational Checks 31
7.2 Dynamic Range 31
8 Regulatory Issues 31
9 Clinical Applications 32
9.1 Study of Protein Polymorphism with Clinical Implications 32
Glossary 34
Abbreviations and Acronyms 37
Related Articles 37
References 38

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Plasma contains a large number of different proteins with varying structural characteristics. The ''plasma proteins'' have been defined on the basis of plasma concentration and amenability to measurement in certain types of measurement systems. In the past, they were proteins present with concentrations above 0.05 g L^{-1}, but more recently detection limits have been decreased to about 0.001 g L^{-1}. However, the group of proteins is not defined on a pathophysiological basis and the definition remains idiosyncratic and incomprehensible to the noncognoscenti. The function of an increasing proportion of them is known and the evidence suggests that they are present either with a specific role in the plasma or that they represent cellular proteins shed into the circulation as a result of degradative processes. Many plasma proteins show characteristic changes in concentration or structure either resulting from genetic dictates that may give rise to disease directly or as a result of changes reflecting pathological or physiological processes. All the transcellular fluids, such as cerebrospinal fluid (CSF), urine, synovial fluid and saliva, contain most or all of the plasma proteins, but there are some additional proteins secreted specifically into these fluids.

This chapter reviews the many methods applied to proteins for clinical and research purposes. Also addressed are the pitfalls not often appreciated by the worker presented with a sample to be analyzed in an increasingly impenetrable ''machine''.

1 INTRODUCTION

Proteins are integral to life as we know it; not only are they the building blocks of carbon-based organisms but they act as catalysts, signaling and controlling molecules for metabolism. Thus, it is not surprising that interest in proteins has assumed a logarithmic growth pattern with no signs of abating. The introduction of DNA techniques that allow the identification of the underlying genetic basis of defects, the development of recombinant methods to produce mass quantities of what once were vanishingly small amounts of purified native proteins, and the creation of immunological analyses that precisely measure many proteins from microliter-sized samples have brought us to a level of scientific sophistication that exceeds our ability to interpret the results.

Many of the seminal publications about proteins and the methods to examine them date back to the 1960s and 1970s. These broadly informative texts are increasingly difficult to locate except in the most extensive libraries. Nevertheless, they are listed here as complete sources on this topic. This area of methodological study is essentially matured and well accepted. The application to clinical medicine, however, is in its infancy.

1.1 Definition of "Protein"

The word protein refers to organic substances that are chains of amino acids, in association with a wide variety of other moieties including carbohydrates, lipids, mucins, metals and other inorganic molecules.[1,2] A constant feature is that they are primarily the product of living organisms and are coded for in the genome. In recent years they are also the product of synthetic processes whereby proteins, unlike any seen in nature, are produced. The potential is enormous, but concerns about the safety of such xenobiotic materials is mounting in public awareness.

The basic unit of protein is the amino acid, of which there are about 20. These small molecules are assembled into peptides, then polypeptides and then into macropolypeptides or proteins. This name is apt because even the smallest protein has a molecular mass (M_r) in excess of 15 kDa.

The general formula for an amino acid fits the structure of $CH(NH_2)COOH$ associated with various side chains usually denoted R to produce $R{-}CH(NH_2)COOH$. All the amino acids participating in protein building are α-amino acids with the same structural backbone, with the exception of proline, which contains an imino ($-NH-$) instead of an amino ($-NH_2-$) group. At its most basic, a peptide is the linkage of two or more amino acids between amino and carboxyl groups ($-NH_2$, $-COOH$, respectively) to form repeating chains in excess of 100 units. Living organisms are composed of different forms of amino acids connected together, end to end, in various combinations along with sugar side chains. This level of organization is referred to as the primary structure because it is a scaffolding held together by covalent bonds.

Amino acids connect together in a linear chain whose regularity is interrupted only at the site of proline molecules, which create a deviation in the direction of the chain. Secondary structure results from the association of molecules in the chain, adding complexity to the simple chain and side-chain structure. The arrangement of the chain, often as a helical structure, is held in a relatively firm position by the formation of hydrogen bonds between hydrogen atoms of one amino acid and oxygen atoms of another. The secondary structure does not account for the compactness of proteins in nature. The coiled peptide chains are packed together in the protein molecule in complex ways to form sheets or globular masses held together by intra-chain bonds, often due to the formation of covalent disulfide bridges between cysteine residues. This physical arrangement, known as the tertiary structure, dictates the shape, function, and ultimately how the protein is viewed as the target of the immune system of living multicellular organisms. It also dictates the chemical response of simpler organisms to polypeptides and proteins in their external environment.

Because all life as we know it has evolved in an aqueous environment, the response to water molecules is paramount. Hydrophobic molecules repel water and are often lipid in nature, whereas hydrophilic molecules respond positively toward water molecules. Both are crucial to life and permit cells to contain high-energy lipid within aqueous cell structures. Hydrophobicity is the result of amino acid side chains that are insoluble in water, whereas hydrophilic proteins have side chains that readily form hydrogen bonds with water molecules. This feature, characteristic of each protein, is promoted by the location of the side chains in the protein structure. Hydrophobic peptides are generally oriented such that the water-repelling side chains are in the core of the folded molecule, whereas hydrophilic units are oriented towards the exterior. Many large protein molecules contain both -philic and -phobic aspects, a feature that allows for such basic features as stable cells in an aqueous medium – the plasma – and retaining fat stores in an accessible form with controlled mobilization.

The classification of proteins can present a bewildering scheme often combining every feature of a known protein. Classification can begin with molecular mass, shape, function, simple or complex, and with what prosthetic group(s) they are associated. In fact, every protein can be classified by all techniques, depending on the view of the classifier. Additional schemes of classification reflect functions of the basic structural components, the isoelectric point (pI), sieving properties, affinity for various materials, including ligands, thermal and acid stability, and solubility, to name but a few.

Finally, the primary amino acid sequence of all proteins is determined by a given DNA nucleotide sequence in living cells. The production of a species of protein is the result of transcription of a unique sequence from DNA to complementary sequences of RNA, which then act as a template for assembling the amino acid sequences from available cellular substrates – amino acids. Post-transcriptional modification within the Golgi apparatus proceeds to prune portions from the original sequence or to add carbohydrates or lipids, thus conferring function, unique immunogenicities and physicochemical properties.

It can be appreciated that with such a complex structure, so variable an assembly line and the enormous opportunity for mutational change with each cell division, the family of proteins is almost limitless and constantly under pressure for change, good or bad. It is at this point that the most sensitive mechanism of all comes into play: the ability of the immune system of higher animals to discriminate between minor structural differences, at primary, secondary or tertiary levels. As a result, protein chemistry relies heavily on immunochemical methods for qualitative and quantitative analysis, the description

of constituent units and even the study of phylogenetic origins of today's family members.

1.2 Definition as it Applies to "Serum" Proteins

Limiting this section to serum proteins[3] (or, more correctly, plasm proteins) simplifies the discussion to a degree, although as methods become increasingly sensitive the line between insoluble structural proteins and those in the serum becomes blurred, resulting in the inclusion of a greater number of different proteins (Table 1). The process whereby plasma becomes serum is complex and results in a variety of changes to several proteins in the plasma. Plasma, the liquid phase that circulates in our vessels, becomes serum in vitro if it is allowed to "clot". The change is the conversion of the plasma protein fibrinogen to an insoluble meshwork of fibrin. This does occur in vivo at sites of injury or blood accumulation in tissue spaces, and is the mechanism of homeostasis.

The result of complete clotting is the disappearance of fibrinogen as it polymerizes and comes out of solution. The cleavage of a small end piece of the long yet soluble molecule allows cross-linkage to occur, but remaining in solution are the cleaved pieces, the fibrin split-products. As normal clot resolution proceeds, nonclottable sections of fibrin circulate and can be measured as evidence of clot destruction.

The process of clotting also has significant effects on other plasma proteins. The process activates proteolytic enzymes, which alter other clotting factors as well as proteins not involved in clot formation. In vivo, the meshwork of the protein clot acts as the scaffolding into which the cells of repair migrate. Serum is therefore the end product of the enzymatically induced process that produces a clot; it is a structure that, because of the physical actions of the mesh-like molecules, contracts, trapping cellular elements and expressing the aqueous phase for our examination and discussion.

2 HISTORY

The awareness that there was such a material as protein developed in the mid-1800s. Berzelius coined the term "protein" meaning "of first rank" (in biological importance) in 1838. The methods available were rudimentary and the reduction of the broadly defined material to subspecies required years and new methods.

Table 1 Clinically useful serum proteins

Acid phosphatase	Chymotrypsin	Intrinsic factor
Alanine aminotransferase	C-Reactive protein	Lactic dehydrogenase
Albumin	Creatine kinase	Lysozyme
Alkaline phosphatase	Elastase	Lactoferrin
Amylase	Factor II	Leukocyte alkaline phosphatase
α_1-Lipoprotein	Factor V	Lipoprotein(*a*)
α_1-Fetoprotein	Factor VIII	Melanocyte-stimulating hormone
α_1-Microglobulin	Factor X	Myoglobin
α_1-Antichymotrypsin	Factor XI	Orosomucoid
α_1-Antitrypsin	Ferritin	Parathormone
α_1-Antithrombin III	Fetuin	Procalcitonin
α_2-Macroglobulin	Fibrinogen	Proinsulin
β_2-Glycoprotein 1	Fibronectin	Prolactin
β_2-Lipoprotein	Follicle-stimulating hormone	Prostatic acid phosphatase
β_2-Microglobulin	γ-Glutamyl transferase	Prostate-specific antigen
β_{1h}-Globulin	Glucagon	Prothrombin
C1-Esterase inhibitor	Growth hormone	Renin
C1q	Haptoglobin	Retinol-binding protein
C1r	Chorionic gonadotropin	Serum amyloid A protein
C1s	Hemoglobins	Thyroglobulin
C2	Hemopexin	Thyroid-stimulating hormone
C3	IgA	Tissue transglutaminase
C4	IgD	Transferrin
C5	IgE	Transthyretin
Calmodulin	IgG	Troponin
Cathepsin-G	IgM	Trypsin
Carcinoembrionic antigen	Insulin	Tumor necrosis factor α
Ceruloplasmin	Interleukin-1	Tumor necrosis factor γ
Cholinesterase	Interleukin-6	Vitamin D-binding protein

IgA, immunoglobulin A; IgD, immunoglobulin D; IgE, immunoglobulin E; IgG, immunoglobulin G; IgM, immunoglobulin M.

Thus in 1861 Graham described dialysis as a means of separating proteins (colloids) from crystalloids. In 1862, Schmidt showed that some proteins, the globulins as he called them, were insoluble in water. Analyses were also extremely simple, often relying on visual color changes upon reaction with various reagents. Early in the 20th century it was recognized that if materials containing proteins were injected into animals, those animals produced a substance that had considerable specificity for the injected material, especially microorganisms. It was in 1890 that Emil Von Behring showed that this antibody, as it was called, resided in a chemically definable fraction of the plasma proteins. He received the Nobel Prize in 1901 for this discovery.[4] Thus the concept of specific immune reactions was formed. Soon it was noted that these ''specific'' antisera produced a visible precipitate when they were mixed with the immunizing material, especially a specific species of bacteria. This flocculation of microorganisms became the basis for estimating bacteria of the type that stimulated the antiserum from others that were considered to be different. Visual assessment of the intensity of flocculation became an estimate of potency when antisera were being considered for therapeutic uses.

When antisera were allowed to react in a transparent medium such as the agar used for bacterial culture, it was noted that the precipitate was comparable to that seen in the test tube. This precipitin technique advanced the knowledge of specificity greatly and has been carried to amazing degrees in the form of immunoprecipitin analysis. The nuances of antigenic identity were explored intensely in the 1960s and 1970s. The next step was to replace the agar medium with agarose, which had less reaction with proteins. Two additional steps were required: the development of a synthetic medium, polyacrylamide, whose interstices could be adjusted at will; and the development of immunofixation and immunoblotting, which allowed the immunological identification of proteins separated on electrophoretic media. This remains a prime analytical technique, revealing that many polymorphisms exist and that minor changes in structure often produce different banding patterns.

Of great interest was the fact that proteins could be eluted from gels, after separation, and then used to immunize animals. Not surprisingly, these apparently ''pure'' proteins often produced antisera with multiple specificities. The reasons are the same as for less sophisticated and imprecise separations; the animal's immune system detects differences that the original separation technique cannot. As we will see later, by repeating this process in various media and buffers, apparently pure materials can be reduced further to extreme purity.

An intermediate step in protein purification came with the immunization of an animal with an antigen/homologous antibody complex that produced an even more superior product. Yet again, contaminants remained that at the time caused no difficulty because of the low sensitivity of the methods then in use. The ultimate method is to produce recombinant proteins by the insertion of the specific gene for the protein into a microorganism, insect or plant and more recently into cloned mammals. These proteins can be harvested from their serum or even from their milk.

3 PHYSIOLOGY AND PATHOPHYSIOLOGY

In general, proteins are synthesized at controlled rates during health with measured levels falling within an expected reference range according to age and gender.[1,5,6] This can be viewed as evidence of relative homeostasis. As will be described in more detail below, the assessment of a steady state cannot be made by examining one protein in isolation. However, for the sake of simplicity, a single serum protein level that falls outside the reference range is cause for concern. If the value is above or below normal, the synthetic rate may be depressed or increased, or catabolism altered. The protein may be lost to some other space or consumed by a process that removes it from the circulation. This occurs during immune complex diseases where complement levels fall precipitously, not by destruction but by sequestration onto and into tissue surfaces. Less common is an inability to synthesize normal levels for some genetic reason. The differences between serum and plasma have been described above.

From a practical point of view, however, the blood proteins that are measured most often are largely unaffected and will give the same results in either plasma or serum. However, dilution and mixing of plasma with aqueous fluids, especially those containing antisera, can trigger the clotting reaction in the cuvettes and tubes of the instrument, leading to error in volume transfer, contamination, and all too often complete failure of the system. Thus, plasma is not the ideal fluid for today's automated devices, whereas it could easily be used in manual methods such as in-gel techniques. Therefore, most analyses now use serum and despite the usage of the word plasma, serum is understood. There are exceptions, of course. Virtually all assays of the clotting system require plasma, but again in some instances serum is preferred.

3.1 Synthesis

The liver is a major site of plasma protein synthesis, although some proteins synthesized there are also produced in peripheral tissues[3] (Figure 1). The other major alternative sites of plasma protein synthesis are

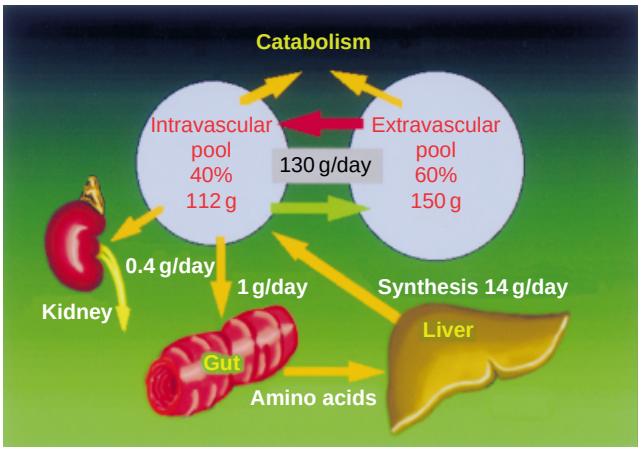

Figure 1 Diagrammatic representation of protein metabolism exemplified by albumin quantities in a 70-kg individual. (Adapted from J. Whicher, C. Spence, *Ann. Clin. Biochem.*, **24**, 572–580 (1987).)

the cells of the immune system. Thus, we know that immunoglobulins are only synthesized by B-cells whereas cell-surface proteins such as β_2-microglobulin are made by many different cells.

The mechanism of protein synthesis is common to all cells. In the cell, proteins are transcribed onto messenger RNA, which is usually translated on cytoplasmic polyribosomes, whereas proteins destined for the cell membrane or for secretion are translated on the membrane-bound ribosomes of the rough endoplasmic reticulum. After synthesis, many proteins possess two additional peptides at the N-terminus. A hydrophobic prepeptide acts as a ''leader'', directing the molecule through the membrane of the endoplasmic reticulum. This is cleaved almost immediately on leaving the protein with a propeptide, which is not removed until maturation occurs in the Golgi vesicles or in storage granules immediately prior to secretion. Glycoproteins have carbohydrate residues added by sugar transferases in the lumen of the endoplasmic reticulum and these may control folding of the polypeptide chain and act as recognition sites for subsequent metabolism. Control of protein synthesis may be exerted on the rate of transcription of messenger RNA from DNA or at the level of translation.

Factors affecting plasma protein synthesis have been studied in the liver. Many factors influence hepatic plasma protein synthesis from specific regulatory factors to amino acid availability. Toxic factors such as ethanol may inhibit synthesis, and hormones have an important, but complex effect. In clinical practice, the effects of androgens and estrogens are the most important. The synthesis of several proteins by the liver is increased in protein-losing states when the plasma albumin level is decreased.

Specific regulatory mechanisms are also being increasingly recognized. It is notable that the synthesis of acute-phase proteins, most of which have specific functions in the inflammatory reaction, are increased within the liver during inflammation in distant tissues due to the action of cytokines released by certain specific cells of the immune system.

3.2 Distribution

Intracellular proteins reach the intravascular space either by secretion or release as the cellular envelope is ruptured or becomes permeable to proteins due to osmotic or metabolic stresses, or as the result of cell turnover and renewal.[3] Regardless of the circumstance, intracellular proteins are often released into the circulation where they can be measured. In this way increased levels of intracellular proteins can signal increased cell turnover or death. This can be seen with tumor markers where measurements afford an exquisitely sensitive means of monitoring tumors that have a propensity to release intracellular proteins. Testicular tumor release of α_1-fetoprotein (AFP), normally absent in normal male serum (it is a normal constituent of the fetal yolk-sac and liver, and during pregnancy), is a good example. Today many people are alive and tumor-free as the result of very early detection of increasing levels of AFP by microscopic metastases and the institution of treatment long before any other sign of tumor recurrence is evident.

All plasma proteins can be found in virtually all spaces within the body, whether they are completely surrounded by intact membranes, such as the intravascular spaces, spinal fluid spaces or contents of the gestational sac, or connecting to the exterior of the body, such as the bronchial tree, urinary bladder or segments of gastrointestinal tract. For smaller plasma proteins the total mass is greater in the extravascular spaces than within the vascular tree, whereas for large proteins the mass within the vascular spaces is usually greater. In practice, the concentration ratio between serum and urine levels or spinal fluid levels can be very informative.

Proteins may be lost into extravascular spaces, e.g. the abdominal or pleural space, when disease results in *transudation* or *exudation* of protein-rich fluids, each with characteristic features (see Table 2).[7] Ascites – the accumulation of a serum-like fluid in the abdominal cavity in cirrhosis or malignancy – can be astonishing; at times, 5 L or more can be present.

Less often, at least in the Western world, are third spaces that are the result of parasitic cysts, which can be enormous, or the accumulation of abdominal fluid as the result of severe nutritional deficiencies. In the same parts of the world, afflicted persons can accumulate enormous quantities of tissue fluid representing a major fraction of

Table 2 Characterization of effusions

	Transudates	Exudates
Appearance	Clear, straw-colored	Clear, turbid, bloody
Specific gravity	<1.015	>1.015
Protein concentration	$<30\,g\,L^{-1}$	$>30\,g\,L^{-1}$
Effusion/serum total protein	<0.5	>0.5
Lactate dehydrogenase concentration	$200\,IU\,L^{-1}$ or 2/3 normal level	$>200\,IU\,L^{-1}$ or >2/3 normal level
Effusion/serum lactate dehydrogenase	<0.6	>0.6
Serum/ascites albumin	>1.1	<1.1
Glucose concentration	Equal to serum	<Serum level
Leukocyte count	$<1000\,mL^{-1}$	$>1000\,mL^{-1}$
Etiology	Noninflammatory Remote systemic disease Plasma ultrafiltrate	Inflammatory Local disease Altered vascular permeability

Reproduced with permission from Ritchie and Navolotskaia.[7]

the total body fluids. These devastating illnesses result from parasitic infestations that block the normal return of protein-rich lymphatic fluids to the general circulation, with resultant massive edema of the lower extremities.

Protein can be lost in tremendous amounts to the exterior through extensive burns, damaged intestinal surfaces or through the genito-urinary tract in the nephrotic syndrome.

The protein content of these sequestered or ''discarded'' fluids varies with the primary process and the type of epithelium that lines the surfaces. Fluid collected in a distended gall bladder will have a different profile from fluid that collects nearby in a pancreatic cyst or outpouchings of the bowel (Meckle's diverticulum). In some cases, study of the characteristics of the fluids can give a precise explanation about the cause: e.g. whether it is a tumor or not and what type of malignancy is responsible.

Plasma proteins pass continuously from the vascular to the extravascular space and as much protein destruction takes place in the capillaries as in the extravascular spaces. There is a concentration gradient between the plasma and the extravascular fluid. The passage across capillary walls is both through interendothelial junctions (especially in the choroid plexus of the central nervous system) and by pinocytic transport across endothelial cells. In both cases, proteins must pass across the basement membrane. It is probable that some of the pinocytic vesicles fuse with lysosomes, resulting in enzymatic degradation of the contents. It is in this way that many plasma proteins are catabolized by the capillary endothelium. The protein content of the transcellular fluids and extracellular fluid is thus dictated by a combination of these two nonspecific mechanisms as well as by specific transport mechanisms. In all such fluids, low-molecular-mass proteins predominate, owing to the molecular sieving effect of the capillary basement membrane. It is important to appreciate that the protein content of transudates will vary from tissue to tissue according to the nature of the capillary interendothelial junctions, and also that the fluid, once formed, is rapidly altered by subsequent metabolic processes. For example, the protein content of bladder urine is very different from that of the glomerular filtrate owing to the reabsorption of proteins in the proximal tubule. Ventricular CSF is very different from lumbar CSF due to the progressive equilibration of the fluid with plasma during its passage down the spinal column.[7] The effect of changes in posture in altering the concentration of proteins in plasma is also of practical importance.

3.3 Catabolism

Plasma protein breakdown probably occurs to a greater or lesser extent in most cells of the body, and degraded plasma proteins provide an important source of amino acids for cellular protein synthesis and tissue repair.[3] In particular, degradation occurs in the capillary endothelial cell during the process of pinocytic transfer from lumen to basement membrane. Proteins that have passed through the endothelium may then be catabolized by the tissue cells they come into contact with, again by pinocytosis and lysosomal degradation.

The liver is also important in protein catabolism. The hepatic sinusoids lack a continuous basement membrane, the endothelial cells have marked intercellular fenestrations and the pericapillary cellular covering is far from complete so that the space of Disse and hepatic sinusoids form a single mixing pool for proteins. Plasma proteins thus enter and leave the hepatocyte without hindrance. This, of course, explains why high-molecular-weight proteins and particles synthesized in liver cells, such as α_2-macroglobulin (725 kDa) and very low-density lipoprotein (VLDL) (512 kDa), enter plasma. Most plasma proteins are glycoproteins and the carbohydrate side chains may have a key role in controlling degradation. Desialation of many glycoproteins results in increased catabolism and it is possible that an intact carbohydrate moiety protects circulating proteins from hepatic catabolism. Removal of these residues as a result of the action of circulating or membrane-bound enzymes may act as the catabolism-initiating process for binding of proteins to hepatocyte membrane receptors, with subsequent pinocytosis and intracellular degradation by

lysosomal enzymes. It is possible that other molecular changes such as polymerization or complex formation may also act as catabolism initiators.

The kidney is the site of breakdown of low-molecular-weight proteins such as immunoglobulin light chains. Between 2 and 4 g of plasma protein are filtered by the glomerular capillaries each day, but only about 100 mg day^{-1} appear in the urine. Proteins pass through the glomerular filter in inverse proportion to their molecular size and are largely reabsorbed in the proximal tubule by pinocytosis and degraded by lysosomal enzymes. Pinocytosis by tubular cells is probably competitive and there may be selective mechanisms for some types of proteins.

3.4 Function

A functional classification of plasma proteins is useful in understanding the changes that occur in disease because proteins of related function often form interacting systems (e.g. immunoglobulins and the complement system). It is probable that most of the plasma proteins present at relatively high concentrations in the blood have a functional role in circulating blood, but it is equally clear that many of those present in trace amounts represent cell-surface or intracellular proteins that have been shed into the bloodstream.[1] Protein function is usually the consequence of secondary and tertiary structure and whether or not they are hydrophobic or hydrophilic.

3.4.1 Transport

Plasma proteins are concerned with the transport of a wide range of substances from the site of production to their site either of action or of catabolism. Another important aspect of their role is the maintenance of a pool of biologically inactive substance in equilibrium with the pool of free active substance. The carrier proteins may have a complex role in the metabolism of the molecules with which they are associated, such as interacting with enzymes or cellular receptors. Transport proteins perform just as one would expect: a moiety is picked up and transported to a target location, where it is released for further use. The divested protein may be free to be reloaded and to transport the same species of moiety again or it may be destroyed after entry into the cell. Serum transferrin is such a protein that interacts with a specific cell-surface receptor to take up and deliver iron. The apoprotein apotransferrin picks up one Fe^{3+} binding to one of the two binding sites. Full saturation of transferrin does not occur, with the average value falling to ca. 1.25 iron atoms per average molecule. Iron is usually bound near the gut mucosal surfaces and is primarily transported to the bone marrow where the iron is transferred to hemoglobin in immature red blood cells: the reticulocytes. It may also transport the iron to the liver cells where ferritin is the intracellular receptacle for iron storage. In hereditary hemochromatosis, ferritin levels rise greatly as a result of leakage from within the cells. By contrast, haptoglobin provides a one-way transport of hemoglobins from wherever red cells are broken down, binding it firmly and then transporting it to the liver where the iron is transferred to ferritin. The heme portion is then released and degraded to biliverdin in the hepatocytes and released into the bile canaliculae, where it is ultimately streamed to the gut and eliminated as bilirubin. Haptoglobin is a suicidal protein, whereas transferrin is a true transport protein. Some protein species have a sharply limited function, such as to bind to a single hormone. Transthyretin is an unusual protein in that it carries two receptors: one site for retinol-binding protein, itself a transport protein for vitamin A; and two sites for thyroid hormones T3 or T4.

Transport of small molecules is also an important function of certain proteins. Albumin, for example, the plasma protein in highest concentration, acts to bind and transport drugs from points of entry to sites of use, preventing rapid loss through the kidney because of their relatively small size. It also provides the constantly available source of amino acids for tissue repair. The amino acid building blocks are derived from the albumin molecules themselves, so it also has a suicidal aspect. In addition to binding drugs, which can have significant effect on the bio- and physicochemical characteristics of albumin, endogenous materials such as bilirubin can act in the same manner. Albumin also acts as a reservoir for bound molecules in equilibrium with the free substance; calcium is such an example.

The binding characteristic of albumin for small molecules such as pigments and dyes is the basis for the more common methods of analysis. Prior binding to endogenous pigments can alter the dye-binding ability of albumin during analysis. As a result, measured values for albumin may be decreased falsely in the serum of patients with high levels of bilirubin or certain drugs such as penicillin. The immunochemical methods, on the other hand, are not affected by dye binding and are gaining favor as the method of choice in those laboratories where immunochemical methods are available.

3.4.2 Proteins of the Immune System

Immunoglobulins provide the adaptive immune response of higher animals for the elimination of foreign antigens or altered autoantigens. Complement proteins provide both effector mechanisms for antibody and an additional nonadaptive defense mechanism. The unique serum protein C-reactive protein may represent a primitive, nonadaptive mechanism-activating complement.

3.4.3 Proteins Associated with the Acute-phase Response

A number of plasma proteins increase in concentration during inflammation in many tissues. The so-called ''acute-phase response'' represents the switching on – by specific messengers, cytokines – of synthetic mechanisms for proteins that are involved in the inflammatory process as mediators, participants or inhibitors.

3.4.4 Signal Proteins

Many plasma proteins are messengers. Most of the classical protein hormones are present at low concentrations, but some, such as human placental lactogen and the placental protein Sp I, are present at high enough concentrations to have been classed as plasma proteins.

3.4.5 Proteins of Blood Clotting

The interacting proteins of the blood clotting system form a quantitatively important group of plasma proteins. The inhibitors and active proteolytic enzymes of this system interact closely with the complement and kallikrein–kinin systems, and thus with inflammation.

3.4.6 Tissue-derived Proteins and Oncofetal Proteins

Some of the proteins present in low concentrations in plasma are cell-membrane proteins shed into the blood during cell-membrane turnover or as a result of cell death. A number of other trace proteins, the oncofetal proteins, are produced by tumors as a result of depression of genes coding for fetal proteins or proteins not normally produced by the tissue of origin or the tumor.

As a first step to viewing blood proteins in an organized fashion, Table 3 categorizes a few of them. Not surprisingly with such a diverse and complex family of molecules, some proteins will have multiple functions and others will have overlapping or redundant functions and be essentially back-up systems.

4 METHODS

Serum proteins have become commodities of research, commercial, diagnostic and therapeutic interest.[8,9] The tremendous leaps in technology and understanding have made individual species of this family big business. A myriad of companies have sprung up around the globe with serum proteins as their principal, if not sole, product. The technological advances, a product of both sponsored and commercial research, have brought this discipline to what will likely be an ever-expanding source of improvement in all fields from medicine to the environment.[5] Because the title of this article focuses our attention on proteins found in serum, our scope is limited. Developments in agriculture, food science and the chemical industry are dealt with by other articles in this Encyclopedia.

The developments in serum protein science fall primarily into two categories: those measured in human or experimental animal serum; and those derived from the serum of animals whose proteins (antibodies) are used in diagnostics or therapeutics. In other words: those proteins to be analyzed and those proteins performing the analyses.

4.1 Immunological Reagents

Measurement of proteins is basically separated into two classes: those that rely on the use of antibodies to target the analyte; and those that separate the protein from, or in, its original matrix for analysis by chemical or physical means. Listed in Table 4 are the characteristics of proteins that can be used either to describe the amounts present or to describe qualitative features. Most methods used today are for medical, veterinary, agricultural or microbiological applications for diagnosis, monitoring or prognostication. In addition, protein analysis forms an integral tool for all biological research. Most methods, for purely economic reasons, analyze proteins in the milieu

Table 3 Classes of serum proteins and a few examples

Transport/storage	Protective	Enzymes	Inhibitors
Albumin	α_1-Fetoprotein	Acid phosphatase	α_1-Antitrypsin
Ceruloplasmin	Chemokines	Alkaline phosphatase	α_1-Antichymotrypsin
Ferritin	Clotting factors	Chymotrypsin	α_2-Macroglobulin
Haptoglobin	Complement series	Creatine kinase	α_2-Antiplasmin
Hemoglobins/myoglobin	C-Reactive protein	Elastase	Antithrombin III
Hemopexin	Fibrinogen	γ-Transglutaminase	C1-Esterase inhibitor
Lipoproteins	Immunoglobulins	Leukocyte alkaline phosphatase	
Orosomucoid	Mucoproteins	Lecethin-cholesterol acyltransferase	
Retinol-binding protein		Plasmin	
Transferrin		Thrombin	
Transthyretin		Tissue transglutaminase	
Vitamin D-binding protein		Trypsin	

Table 4 Methods of protein measurement

Focus of method	Mode of results
Antigenic specificity	Quantitative or qualitative analysis
Antibody activity (for immunoglobulins)	Quantitative or qualitative analysis
Associated moieties	Measurement of metal, lipid or carbohydrate
Electrophoretic mobility	Measure p*I*
Enzyme sensitivity	Cleavage with proteolytic enzymes and measured effect
Function	Enzymatic, ligand, immunospecificity (for immunoglobulins)
Molecular configuration	Stoichiometry
Molecular mass	Ultracentrifugation
Molecular shape	Rotational polarimetry
Molecular size	Sieving electrophoresis
Thermolability	Visible or measurable effect on physical status or function

of the original sample to avoid the inevitable increased cost of separation. Unless the examination is of a general nature, e.g. total protein, or addresses a feature that allows a specific protein to be examined in the presence of a large number of surrounding proteins, e.g. enzyme activity, the methods of analysis require the application of a specific antiserum: immunoanalysis. As this is such a fundamental and complex issue, it will be dealt with first.

4.1.1 Antiserum Production

Antisera are animal immunoglobulins, either raised in intact animals or produced in vitro from clones of antibody-producing cells taken from immunized animals. The fundamental difference between these is that an animal immunized against a macromolecular protein antigen will produce many thousand antibody species reacting with a large number of antigenic determinants on the protein. Monoclonal antibodies, produced from clones of plasma cells grown in vitro, will react with only a single antigenic determinant per protein molecule unless this determinant is repeated in the protein structure. The characteristics of antibodies that are important in immunoanalysis are the avidity, affinity and titer. The avidity is the binding constant of a homogeneous species of antibody to a single antigenic determinant. It is a theoretical concept except in the case of monoclonal antibodies. The affinity is the average binding constant of all the antibody molecules in an antiserum to all the relevant antigenic sites on the target protein and is a characteristic of great importance because it determines the speed of reaction with antigen and the stability of the immune complexes formed. The titer is the concentration of reactive antibodies in the antiserum. The titer can be modified by methods of concentration of immunoglobulin in the antiserum and the affinity can be modified by immunoabsorbtion techniques to concentrate antibody molecules of high avidity and reject those with low affinity (see below). The preparation of specific antisera requires that the protein antigen, the analyte, be isolated from the best available source. Recent regulation requires that the source be shown not to contain infectious agents such as the human immunodeficiency virus and hepatitis B virus. For this reason, source materials from acceptable individual donors have become extremely valuable. In recent years the intermediate step, largely avoiding problems of purification of the starting antigen and also the removal of contaminating specificities from the animal product, has been the production of recombinant antigens; this subject is not addressed in this article. Preparation of the desired immunogen has become more sophisticated, partly because of improved techniques and partly through the realization that the higher the purity of the administered antigen, the more efficient the process of removing contaminating specificities from the antiserum by immunoadsorption.

In the early years of antiserum production it was not unheard of to immunize an animal with a very impure preparation extracted from a serum or tissue believed to have the antigen in high concentration and then to absorb the product with a serum or crude material that was devoid of the desired protein. For example, animals were immunized with crude preparations of serum from small fetuses with high levels of α-fetoprotein. The harvested antiserum was predominantly against contaminants and the specific anti-α-fetoprotein was only a small fraction of the total. To remove this large amount of interfering antibody, whole normal male plasma was added until no further precipitate occurred. Although extremely crude, the method did provide usable antisera. To this day the same problems exist. The purest of immunogens are shown by the immunized animal to be laced with contaminants. Not infrequently, the trace contaminants can be far more immunogenic than the target immunogen.

A further advance in the preparation of high-quality antisera was the development at the Catholic University of Louvain, Belgium by Masson and his associates Mancini and Carbonara of the method of immunizing with low microgram doses of immunogen, injected intracutaneously into multiple sites on the back of the animal, usually a goat.

The antigenicity of the injected material is usually enhanced by emulsification with a lipid-rich material, of which there are several. In the main, whatever material is added, whether mineral oil or tubercle wax D or synthetic adjuvants, the result is an intense inflammatory response with the production of cytokines, such as interleukin-6, which enhance the immune response at the sites. Subsequent booster doses in the same manner are required to elevate the titer of the antibody to the desired

levels. One method of interest was immunization with a washed precipitate produced by an antiserum that was as pure as could be obtained with its respective antigen. The precipitates were developed in relatively dilute solution to avoid the problem of ''occluded'' protein: those that become trapped in the mesh of developing antigen–antibody precipitates. The approach did, in fact, improve the quality in some cases. Another modality is to immunize animals with the fixed and stained proteins separated in polyacrylamide gels. The bands are sliced from the slab, emulsified and injected. Despite fixation and staining, intact antigenic determinants remain.

The inevitable requirement for removal of contaminating antibodies also evolved with time. Initially, soluble antigen was added to the antiserum and the precipitated contaminants were removed by centrifugation. However, as antiserum performance became more demanding, the possibility of inserting unwanted reactants into the finished product and the inevitable dilution by the absorbing solution forced workers to develop efficient adsorptive protocols. Ultimately, large columns of absorbing antigen immobilized on solid-phase materials were used for quick and efficient isolation of products that were relatively monospecific without the addition of other materials or dilution. These methods also had the enormous advantage that, by first eluting and discarding the loosely bound antibodies, the tightly bound, high-avidity antibodies could be collected, rendering the antiserum of higher affinity and thus reactivity. In addition, columns with immobilized antibody could be used to purify antigens. Despite the high cost of preparing a column containing several liters of immunoadsorbant, the columns could be used to strip contaminants from antigens or antibodies (sometimes valuable in themselves) and reused, sometimes over 100 times.

4.1.2 Antiserum Specificity/Relative Purity

As immunoassay methods have increased in sensitivity, the problems of contaminating antibodies have increased. An antiserum usable for in-gel analysis may fail completely when used for an enzyme-linked immunosorbent assay (ELISA) of the same analyte. The problem is simply the presence of low-affinity antibodies, which fail to produce precipitates in solution but are capable of forming complexes when bound to a solid-phase support. For this reason, affinity purification to remove such antibodies has become the norm. As a result, antisera produced today are usually far superior to those produced in the past; however, there is still the practical issue that antisera used for immunonephelometry need not be as monospecific as those for solid-phase fluorescent or enzyme-linked methods. The problem of an antiserum designed for a less-demanding procedure being applied to one requiring high purity, by an unsuspecting worker, is all too common an occurrence.

The subject of monoclonal antisera must be included and put into proper perspective.[10] When first introduced, monoclonal in-vitro-produced antibodies seemed to be a cure-all for immunoassay and the problems of contamination. Within a short time several shortcomings appeared. Producing a monoclonal antibody to a small, often poorly antigenic material was a great improvement over the traditional approach. Not surprisingly, firms that produced kits for serum protein assays felt compelled to explore the use of monoclonal reagents. The very basis of producing a monoclonal cell line dictates that each clone will produce antibody to one antigenic determinant. Unfortunately, this may happen to be of low avidity and nothing can be done to improve this except to search for another clone. This feature, so valuable in the measurement of moieties with few determinants or of restricted heterogeneity, worked against their use for measuring analytes with antigenic diversity – most serum proteins – because the monoclonal antibody only reacted with a small part of the spectrum of molecules present. As a result, the desire to measure large complex proteins failed with monoclonal reagents. More recently, cocktails of multiple monoclonal antibodies are solving these problems. There have been monoclonal antisera produced that clearly solve analytical problems for serum proteins. For example, antisera against cytokines, to the PI*ZZ form of α_1-antitrypsin and to some hormones have been successful. It remains, however, that for serum protein assays with proteins of excellent antigenicity and high concentrations, polyclonal antisera from the intact animal serve well and economically.

4.2 Immunoassays

4.2.1 Light-scattering Immunoassay

The bulk of serum protein assays done at present employ instruments that measure the amount of light scattered or absorbed by the immunoprecipitate developed with high-affinity and high-titer antisera. Two approaches have been used successfully.[11] The first is to direct a light beam of high-intensity light, broad-spectrum light or laser light at a cuvette containing a solution of the antigen-containing sample, to which is added a volume of specific antibody. A precipitate develops rapidly, often augmented by an enhancing polymer such as PEG-8000. At a specified time the amount of light scattered or absorbed by the immunoprecipitate in the sample is measured: the end-point. This is compared to a cuvette containing all reagents except for the sample. By subtracting the blank signal from the sample signal, the resultant can be converted to mass units from a standard curve of known concentrations of antigen.

The second approach is to monitor the rate at which the precipitate develops after the last fluid addition is

made to the cuvette. Rate analysis has several advantages over end-point analysis in that it is less sensitive to initial sample turbidity or pigments such as bilirubin or hemoglobins, it is not as sensitive to fluctuations in the formation of flocculent precipitates that develop after some minutes and it is unaffected by variation in the optical properties of the cuvettes.[12]

The reaction kinetics of all immunoassays can be described by the Heidelberger–Kendall curve (see Figure 2), which rises from near zero on the *x*-axis (the antigen concentration) and *y*-axis (the intensity of scattering) (*the ascending limb*) to a maximum (*the point of equivalence*) and then descends on the *y*-axis (*the descending limb*) at a high value of *x*. Often the signal disappears at very high concentrations. All forms of immunoassay can suffer from the same condition where either of the two reactants, antibody or antigen, is in great excess and the immune complexes become soluble. Of greatest concern in these systems is when the antigen concentration is in great excess. Because most assays of serum proteins have a dynamic range of approximately one order of magnitude, those serum proteins that can vary by more than one order of magnitude present a problem called antigen excess. It is the *bête noir* of immunoglobulin assays. Greatly increased levels, sometimes 20–50 times normal, can produce antigen excess signals low enough to be considered by the instrument as within the normal range. Various approaches, both within the assay system and by confirmation with other methods such as semi-quantitative serum electrophoresis, are used to overcome this problem. Antibody–antigen precipitates can be generated and measured in the fluid phase by immunonephelometry and immunoturbidimetry, in stabilizing gel techniques such as radial immunodiffusion and electroimmunodiffusion, or on solid particulate or planar surfaces by immunosorbent assays.

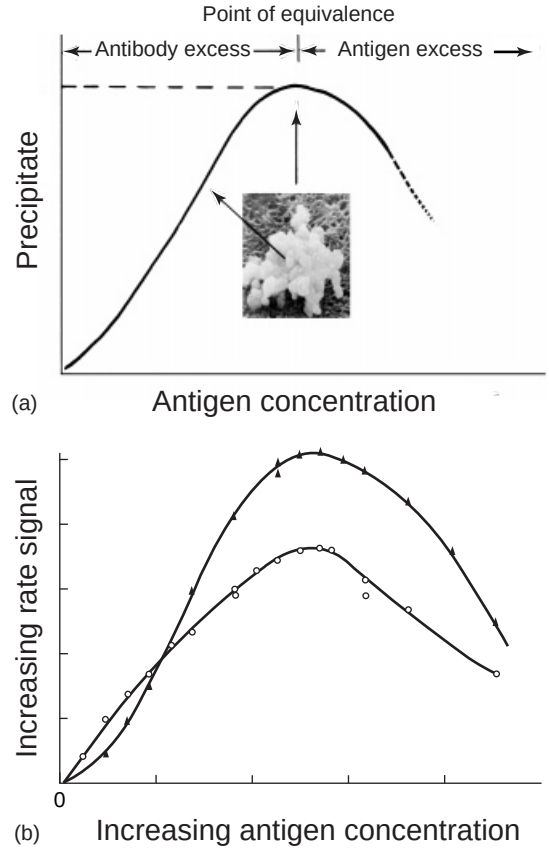

Figure 2 (a) The Heidelberger–Kendall curve, illustrating the relationship between antigen concentration and precipitate. The ascending limb is produced by small units of immunoprecipitate (insert, left arrow), the point of equivalence is the point at which the maximum amount of precipitate has developed (insert, top arrow) and the descending limb is created by increasingly soluble precipitates. (b) Rate nephelometric curves for two antisera produced against the same antigen, demonstrating the differing relationship between antigen concentration and rate signal for seemingly identical products.

4.2.1.1 Immunoturbidimetry Measuring the exiting light at zero degrees directly through the cuvette can perform the procedure described above. With increasing amounts of immune precipitate the light intensity decreases. However, there are a few caveats. At times, the immunoprecipitates are dense and may self-aggregate, thus allowing more than the expected amount of light to pass, or they may sediment out of the light path and give the same error. In most commercial applications, reagent concentrations have been tailored to minimize this problem. There is a conflict that at times can be difficult to resolve: by reducing the amount of reagent and sample to prevent flocculation and sedimentation, the sensitivity is reduced. Conversely, to improve sensitivity by adding larger volumes of sample, as in the case of a low-concentration analyte (e.g. C4, transthyretin), antibody of higher quality is needed. Clearly, antibodies of high affinity and titer are essential.

Intrinsic sample turbidity can be a handicap for low-level assays that require large sample volumes at low dilution. As sample turbidity increases, the baseline or blank value also increases toward a level where precision is compromised (at high optical densities). Clarification of samples by high-speed centrifugation, although a logical way to reduce intrinsic turbidity, can introduce the problem of removing some analytes with the chylomicrons. Serum proteins such as C4 and IgA, and sometimes IgM, will bind to the particles and be reduced to very low levels as a result of the procedure. Ideally, the patient preparation step of fasting must be underscored and a repeat sample taken. In some patients with certain forms of hyperlipemia, even this step will not clarify the

serum and results from measuring such samples should be regarded as suspect.

4.2.1.2 Immunonephelometry The difference between immunonephelometry and immunoturbidimetry is principally that the former measures the light scattered and the latter measures the light transmitted.[11,12] In other words, light that strikes particles and is deflected laterally. Different angles have been employed; a 90° angle has been used frequently with good results. Light reflected from particles strikes many surfaces before leaving the cuvette. Light that exits at 0° has been reflected internally many times. Particle size plays some role in the performance of each system. For small particles whose diameters are less than 1/20th the wavelength of the analyzing beam, the pattern of scatter is called Rayleigh scatter and presents a relatively symmetrical pattern about the axis of the beam. For larger particles, Rayleigh–Debye scatter presents a symmetrical pattern favoring forward angle scatter (toward low angles). Short-wavelength light is scattered much more than that of long wavelength but the requirement for a collimated light beam has made the use of the cheap helium–neon laser of long wavelength popular despite its poor sensitivity. Regardless of the angle at which light is measured, changes in the exiting beam intensity over time constitute rate nephelometry. The analysis time can be very rapid and in some assays time must be reduced to allow mixing transients to have passed before the reaction is measured.

4.2.1.3 Radial Immunodiffusion Not usually considered a light-scattering assay, the application of a set volume of serum to an antibody-containing gel slab is, in fact, just that[13] (see Figure 3). Observation of the immunoprecipitate is by indirect light: the Tyndall effect. The precipitates that have formed as antigen diffuses out concentrically from a central well produce an expanding ring (or rings) of immunoprecipitate that either stabilizes as all the antigen is bound and immobilized in the matrix or is measured at a set time after incubation at a set temperature. Growth and outward migration of the rings is the result of precipitation and resolution until all the antigen has been consumed. As diffusion continues, the precipitate dissolves on the inner aspect in the area of *antigen* excess and forms where *antibody* is in excess. Analyte concentration is derived from measuring the diameter of the ring in more than one axis (or the area) of precipitate as compared to rings of known concentration. Simplicity is the greatest asset of this method, which is still used in underdeveloped countries where budgets do not allow for the expensive instrument packages. It is also extremely efficient for workers developing new assays for research purposes using their own antisera.

There are a myriad of creative ways to employ gel diffusion to study antigen relationships: identity, non-identity or partial identity. There are several texts describing these simple yet elegant methods.[14]

4.2.2 Immunoaffinity Chromatography

Properly designed and prepared, the specificity of antisera allows the worker to perform quantification, purification, recovery and visualization. These techniques require scrupulous attention to the details of antibody specificity. The importance of this issue cannot be overemphasized. Confirmation of specificity remains the responsibility of the worker when applications other than those specified by the manufacturer are employed. Such problems were

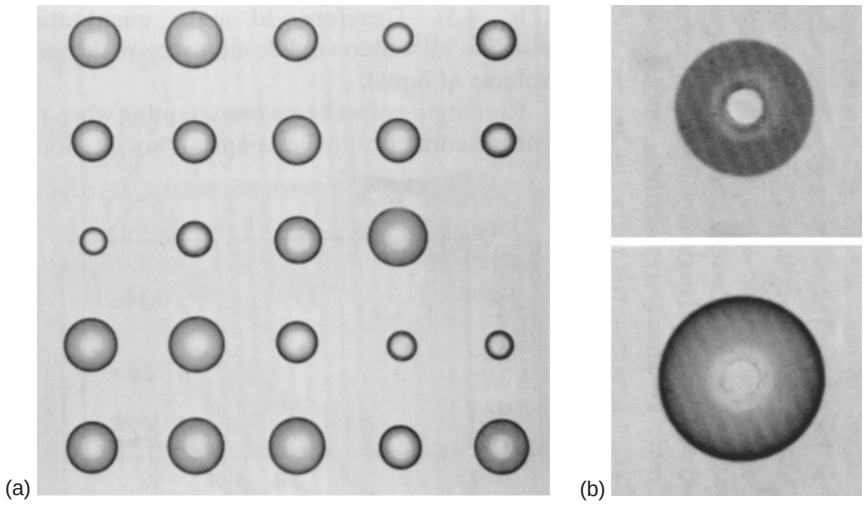

Figure 3 Radial immunodiffusion in agar gel. (a) A demonstration of how many samples, including controls and clinical samples, can be applied to one plate. (b) The appearance of a matured test (bottom) with a sharp outer edge and a still expanding test (top) where the outer edge is diffuse. The images are shown as negatives.

one of the driving forces that propelled the United States Food and Drug Administration (FDA) to require clinical laboratories that used "in-house" tests or reagents to perform extensive validation studies akin to those performed by industry in what has been called the 510 k approval process.

4.2.2.1 Qualitative/Preparative The attachment of antibody or antigen to beads or other insoluble particles has been the mainstay of much of the work in antigen purification, either for detailed study or for the preparation of immunogens for immunization. In principle, the method relies on the attachment of antibody (or antigen) to a large surface area contained in a small volume, such as a porous bead. This large area is then exposed to a solution of protein. This procedure is usually performed in a column, although it can be performed in a flask. The gel is thoroughly washed with a benign solvent and then, when thoroughly flushed of contaminants, a solution of a chaotropic agent that dissociates the antibody–antigen complexes is passed through the column. The released material, now in solution – the effluent – contains the sought-after material. The solution is dialyzed to remove the chaotropic agent, leaving relatively purified material. The process can be repeated, using the recovered material to reduce the contaminants further. Of great importance is that the surface, once cleansed of the bound material, can be reused in the same manner. Industrial columns have been reused hundreds of times, thus greatly reducing the high cost of producing the column.

4.2.2.2 Quantitative/Analytical The same principle can be designed to view the stripped effluent, as it is eluted, as the sample to be analyzed. Proteins can be quantitated by spectrophotometry, which can be readily adapted to measure these effluent concentrations. Because the fluid initially applied to one column can be sent to a second column with different affinity, a serial analytical system can be envisaged. Such a system has been designed and used to assay 25 low-level analytes in series. This elegant work exemplifies all the affinity chromatography work that precedes it.[15]

4.2.3 Labeled Antibodies

The development of fluorescent, radioactive and enzyme markers attached to antibodies (and also to antigens) spawned an entire new industry. The requirement that antibody molecules (usually IgG) be coupled with these amplifying materials, exposed the proteins to rigorous conditions that often resulted in serious damage. Initially, shelf-life was short and performance characteristics were unpredictable within and between batches. Great advances in the quality of these products have been accomplished. The principal labels are listed in Table 5.

Table 5 Materials used to label antibodies

FITC	^{125}Iodine	Alkaline phosphatase
Rhodamine	Tritium	β-Galactosidase
Horseradish peroxidase	Latex particles	Rare-earth chelates
La Jolla Blue®	Gold	Acridinium ester
Isoluminol	Biotin	Luciferase

FITC, fluorescein isothiocyanate.

4.2.3.1 Fluorescence Immunoassay Each type of label presents advantages not offered by the others, but also disadvantages. The most disparate would be the exquisite sensitivity and ease of operation of the radioimmunoassay (RIA) countered with the general public's reluctance to deal with radioactivity, leading to expensive disposal problems. At the other extreme are fluorescence-labeled antibodies (FITC), which provide beautiful visual localization on tissue sections (Figure 4) with a stable and benign reagent that is easily visualized with ultraviolet microscopy, unfortunately countered with the rapid decay of the fluorochrome while being observed.[16,17]

Fluorescence has also been used as a means of quantitative measurement; however, the major problem of autofluorescence at the wavelengths used has required innovative instruments that measure fluorescence emission from fluorophores after a long time delay between stimulation and emission of light. Rare-earth chelates using europium have allowed workers to largely ignore background sample fluorescence at the stimulating wavelength and to delay analysis until the autofluorescent pulse has ended. Time-resolved fluorescence analysis has proven to be highly effective, even for low-level analysis. In a similar manner, sample autofluorescence has been reduced by selecting fluorochromes, which emit at wavelengths significantly removed from the stimulating wavelength. Autofluorescence therefore can be blocked by optical filters and is of little concern. In addition, selecting fluorophores that emit in the infrared region, where the sample contributes little background, resolves the same problem by a different approach.

The use of molecules that emit light directly or indirectly after enzyme activity on them has become a major field of assay development. Called chemiluminescence, the assay design fits that of ELISA assays and has extended the limits of sensitivity considerably. There are homogeneous and heterogeneous assays, as described below.[18] All such assays provide the feature of great amplification of what could be considered as the primary reaction. The steps that lead up to the generation of measurable light can be concatenated to allow precise assays

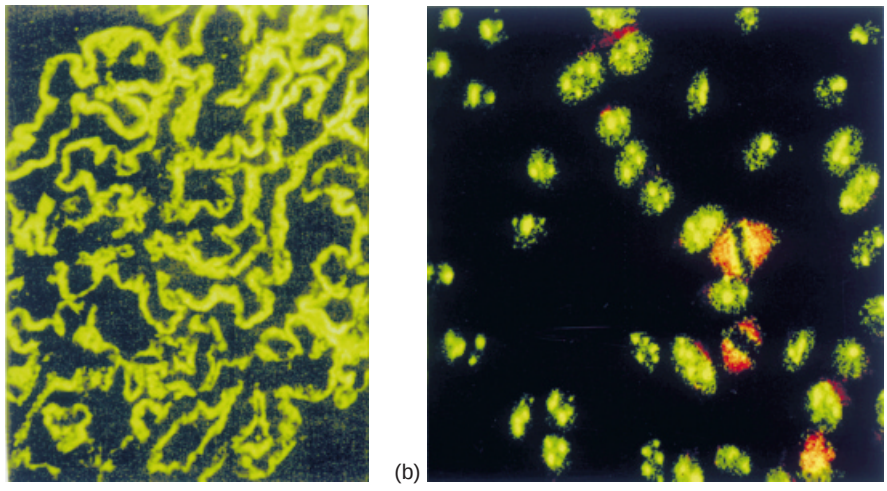

Figure 4 (a) Fluorescent microscopic images of immune complexes visualized in the glomerulus with FITC-labeled anti-IgG. (b) Antinuclear antibodies visualized with FITC-labeled anti-IgG on HEP-2 cell substrate.

into the low picogram range. These assays have become excellent methods for replacing RIA.

Immunoassays can be designed to be complex and multilayered with the goal of increasing sensitivity. These more complex assays fall into two categories: heterogeneous and homogeneous. The former has been supplanted in many cases by the latter for economic reasons; the former requires a separation step whereas the latter does not.

Heterogeneous. In brief, these assays present a specific antibody reagent to the antigen-containing sample where the reaction develops. After a specified time the sample is treated by the addition of precipitating agents such as polyethylene glycol, then by centrifugation or filtration. The precipitated antigen–antibody–fluorochrome complex is then measured optically. The separation step can be confounded by what can loosely be termed coprecipitation, meaning the development or pre-existence of precipitates. Agents that foster precipitation of the immune complexes can also promote precipitation of other proteins in the sample and coprecipitate the fluorophore. Nevertheless, these assays perform very well when properly designed.

Homogeneous. In this format, antigen-containing samples are mixed with unlabeled antibody, a conjugate of the target antigen with a nonfluorescent substrate, and an enzyme (e.g. galactosidase). Unlike the heterogeneous assay, this system will develop fluorescence only if the enzyme cleaves the bound substrate. Reaction of the bound substrate with the added antibody prevents this cleavage, thus inhibiting fluorescence. However, if there is free antigen in the sample it will bind with antibody, thus reducing the amount of antibody available. Some free unreacted nonfluorescent substrate will then be free in the solution and capable of being converted to a fluorescent material by galactosidase. At one extreme the sample devoid of the target antigen will exhibit maximum fluorescence, whereas samples with high levels will exhibit none. The reaction can be measured without the need for a separation step.

4.2.3.2 Enzyme-linked Immunosorbent Assay These extremely sensitive assays have evolved over the last three decades. Antibodies are tagged with an enzyme that, when bound to small amounts of antigen, can convert large amounts of substrate to colored products, thus enormously amplifying the signal.[19] ELISAs usually take advantage of the fact that surfaces such as glass and plastic passively bind materials such as proteins and carbohydrates. The initial step is the coating of the surface of the container (tube or well) with unlabeled antigen (A) or antibody (B) in the sandwich assays (see Figure 5). The diluted sample containing the antigen is applied and incubated under controlled conditions and then washed away. The remaining bound antigen is then exposed to a second antibody that is linked to an enzyme such as alkaline phosphatase, forming the "sandwich". This second antibody (B, C, D) can be identical in specificity to the antibody coating the surface, or be directed to a different aspect of the antigen, thus conferring increased specificity. It may also be an antibody to the species that has provided the first antibody (D). The washed wells are then developed with proper substrates to produce color, which is then read spectrophotometrically. Many automated devices are now available to perform most of the operation with little operator intervention.

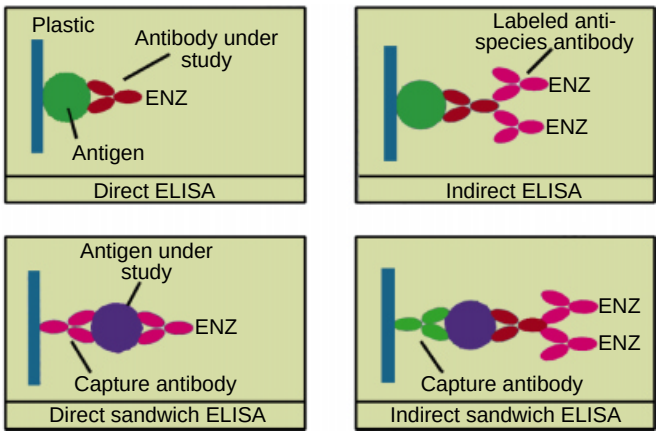

Figure 5 Diagrammatic representation of the direct form of ELISA. (Adapted from Crowther.[19])

As with other immunological methods, creativity has produced a great diversity of protocols with unique features to address special problems. The nature of the multilayered assay allows antigen to be substituted for antibody on the surface of the well, thus providing a means of assaying antibody levels. In such a circumstance, the first antibody is specific for the coating antigen – the antibody being measured (e.g. antibody to the Lyme disease agent, *Borrelia burgdorferi*) – and the second antibody is labeled, which, after separation and washing, provides the signal. It is a simple concept to envisage that, because each step provides amplification, added steps will increase sensitivity. Additional steps can be introduced to inhibit each phase of the assay, thus providing a competitive format. Coupled with a variety of tags and antibodies, virtually any antigen–antibody system can be designed.

4.2.3.3 Radioimmunoassay In exactly the same format as described above, RIAs of extreme sensitivity and specificity can be developed.[20–22] Despite the public aversion to radioactivity, these assays persist and are used safely worldwide. The minuscule amounts of radioisotope employed require Herculean measures for storage and disposal of spent reagents that are far beyond the risks involved. Most RIAs are performed in tubes, in automated devices, with the final step being to measure photon emission in a scintillation fluid (an organic solvent) as the result of radioactive decay of the isotope.

4.3 Biospecific Affinity Chromatography

Although not an immunological method, utilizing the affinity of a protein for a specific material is so similar in all aspects to the method described in section 4.2.2 that it is described in this section.[23] It is an effective means of purifying or quantifying certain proteins. Historically, it was introduced in the early part of the 20th century. It is still used extensively in industrial applications for clarifying, decolorizing and extracting desired or undesired materials. The first affinity adsorbent was charcoal and it is still used in the food industry. There are a myriad of materials that can be used in this way. For serum proteins the activity against target substances such as enzymes, immobilized ions, DNA, dyes, heparin, and the endless list of antigens or antibodies are examples. Very recent use of nanoparticles of lipid for the trapping and removal of specific moieties has particular promise in both purification and diagnostic applications.

Charcoal, microporous glass, ceramics and similar materials possess immobilized ions for trapping or removing proteins, particularly those with hydrophilic groups. Plastics or other hydrocarbon polymers can provide potent binding sites for hydrophobic proteins such as the lipoproteins.

Binding to insoluble matrices also provides a means of further characterizing proteins. The association of a protein with a specific target can vary within a protein species, thus providing a means of further classifying a ligand. Once a material is bound, elution or dissociation can be induced by stepwise changes in eluant pH, ionic strength, ionic content, temperature or specific dissociating agents such as dithiothreitol, glycine, guanidine, detergents and other chaotropic agents.

The ligand can be a plant lectin with specificity for a protein, or even subspecies of animal proteins.[24] These materials are widely employed to prepare and isolate large amounts of material for study. The number of commercially available lectins is steadily increasing as their importance is appreciated: more than 100 have been identified. Many of these materials are extracted from seeds (ex.: concanavalin A) and others from bacteria (staphylococcal protein A). Presumably their original biological purpose is being exploited in these laboratory procedures.

4.4 Chemical Assay

Some of the analytical methods that can be applied to proteins are used infrequently or are so specialized as to be employed only in the research setting. These will not be dealt with other than to mention them and the information that they provide.[5]

4.4.1 Colorimetric

The measurement of protein concentration spectrophotometrically can use many different features of the protein family: the nitrogen content, presence of peptide bonds, the ability to bind dyes, and the ability to form light-scattering precipitates with specific reagents. These methods are all destructive forms of analysis. They are widely employed but, like all forms of analysis, they have

both advantages and disadvantages. In general, these methods cannot be used if unwanted turbidity develops during the analysis. Because many are still in use to determine the protein content of difficult materials, they will be mentioned here.

4.4.1.1 Nitrogen Content Nitrogenous compounds in the serum are principally protein;[5] however, in urine, where protein content is also of great clinical and research importance, the high nitrogen content of urea makes this method unusable unless proteins are separated by dialysis. Nitrogen content assay is now used rarely, except in the investigative setting where fractions or purified materials are examined. This method also has the advantage of being able to measure the nitrogen content in whole tissue or in precipitates that are sufficiently turbid to preclude other common forms of analysis. Several techniques have been devised in the past and still have an important role to play, although in general they require large amounts of sample.

The *Kjeldhal method* relies upon the digestion of nitrogen-containing material by sulfuric acid. The digestion produces ammoniacal compounds that develop a specific yellow color with the highly basic, mercury-containing Nessler's reagent. This method, although excellent, is subject to many interferences and requires careful technical attention to prevent spurious charring and turbidity.

4.4.1.2 Peptide Content The *Biuret reaction* is directed to the presence of at least three adjacent peptide bonds.[5] It therefore has the advantage of not being reactive with other nonpeptide nitrogenous compounds. Free amino acids do not affect the reaction but small peptides can interfere, hence the method again is not usable in measuring urine protein directly. The addition of an alkaline copper sulfate-containing reagent produces color detectable at 557 nm. Because individual species of proteins contain different numbers of peptide linkages, some workers have listed Biuret numbers to correct the reading. It is clear that this method when used with a gravimetrically standardized amount of total serum protein, bovine serum albumin or even human serum albumin gives different results.[25] As a result, clinical samples containing disturbed concentrations of serum proteins can give very aberrant values.

The *Lowry method* is a more sensitive assay. The amino acids tyrosine and tryptophan, characteristic of proteins, are the focus of this method. The phenol Folin–Ciocalteau reagent containing copper ions produces a color detectable at either 720 or 660 nm, but like the other methods the Lowry technique is not suitable for materials that produce turbidity.

A further increase in sensitivity is achieved by the *ninhydrin method*. Like the Biuret reaction, different protein species produce different relative values.

4.4.2 Dye Binding

One of the most commonly used methods for measuring albumin in serum relies on dye binding.[5] Among these are bromocresol green and bromocresol purple, the dyes used in most automated chemistry analyzers. It must be remembered that those proteins that bind dyes also bind other materials as their physiological abilities dictate. Moieties such as bile pigments, anti-inflammatory medication such as acetylsalicylic acid, penicillin and many medications bind to albumin and in the instances where these materials are in high concentration the binding of reagent dyes can be interfered with to a significant degree. The most difficult situation is in severe liver disease, where high levels of bilirubin coupled with low levels of circulating albumin can produce falsely low values. In such situations, the immunochemical determination is believed to be superior.

4.4.3 Precipitation Methods (Nonimmunological)

4.4.3.1 Destructive Primarily applied to CSF total protein analysis, the use of certain acids to precipitate protein is widespread. Trichloroacetic acid and sulfosalicylic acid are the main agents used. A problem with these commonly used techniques is that glycoproteins are very resistant to precipitation and remain in solution.

These techniques are used for measuring urinary total proteins but they are not to be relied upon because other materials also precipitate, such as the Tamm–Horsfall protein and proteins from the epithelium of the genitourinary tract.

4.4.3.2 Nondestructive All proteins are subject to precipitation in the face of an increasing concentration of certain neutral salts, such as ammonium sulfate, sodium sulfate, Rivanol (a long-chain alcohol) and polyethylene glycol.[25] In general, proteins come out of solution in relation to their isoelectric point: fibrinogen first, at lowest salt concentration; and then the immunoglobulins. Ultimately, some proteins will remain in solution in saturated solutions of the salts. Rough quantitative methods have been devised using these properties. Rivanol is the exception, precipitating protein in approximately the reverse manner.

4.5 Physical Properties

Analysis of serum proteins can be accomplished efficiently by measuring individual physical properties of a species in question.[5] With the help of modern instruments many

aspects can be used to describe the characteristics of a sample, both in quantity and quality.

4.5.1 Optical

4.5.1.1 Refractometry As light passes from one medium (air) to another (aqueous solution), the angles to the perpendicular at the boundary can be measured as the refractive index. The total solute content of the liquid affects the index of refraction in a proportional manner so that after dialysis against a ''blank'' solution it can be used to measure protein content. The method has the advantage of being nondestructive. It has been used extensively to measure total protein in the past but it is no longer used where sophisticated instruments are available. It remains of use, however, largely in the research setting when an approximation of the protein content of a small volume of a purified product is needed: the tested volume can be recovered.

The same principles are employed in analytical ultracentrifugation (see section 4.5.11).

4.5.1.2 Spectrophotometry A mainstay of protein quantitation in the research setting has been spectrophotometric measurement of light absorption at a wavelength of 280 nm.[25] This wavelength is usually selected because it represents an absorption peak for proteins rich in tyrosine, tryptophan and phenylalanine. Not surprisingly, when purified proteins are examined for the absorption spectra, each is slightly different from the others as a result of differences in their amino acid composition. Nevertheless, measurement of absorbance at 280 nm is a very useful, nondestructive method of approximating the total protein content at concentrations of $0.05–2\,mg\,mL^{-1}$. This form of analysis is widely employed in monitoring immunoaffinity column effluents. Ultraviolet absorption using shorter wavelengths develops problems because of interference from other substances, including amino acids. At higher wavelengths, protein absorbs very little.

4.5.2 Mass/Gravimetry

Although seemingly crude, there is logic for using this technique as a benchmark for standards. There are several criteria that must be met before the method is usable. The protein to be measured gravimetrically must be obtainable in sufficient quantity to yield an accurate reading in modern balances. New instruments are exceedingly accurate, precise and sensitive when kept on proper maintenance programs. High-sensitivity instruments can easily reach a precision of less than 1 mg. Nevertheless, careful calibration must be performed on a regular basis. In the case of proteins, their propensity to become hydrated from ambient air requires that the purified material be sealed in dry nitrogen after removal of as much water and salts as possible. The method requires no indirect comparisons and when sufficient material is available it is a benchmark means of analysis. Associated nonprotein constituents such as sugars and lipids also contribute to the weight.

4.5.3 Viscosimetry

A measure of molecular size and configuration of solutes versus the solvent alone constitutes a measurement used clinically as an alternative to the trusted erythrocyte sedimentation rate. Measurement of the viscosity of whole human serum or plasma is a very approximate and imprecise measure of the average viscosity of all proteins in the sample and is the result of the concentration and shape of the major fractions.[5,25] Albumin is in the highest concentration range but, being a globular protein, its viscosity-inducing effect is minimal. IgG is more elongate and of intermediate concentration, therefore contributes significantly. The very large proteins IgM and α_2-macroglobulin are both of much lower concentration yet do contribute significantly. All other proteins combined contribute very little effect. In plasma, however, fibrinogen – an elongate, fibrillar protein – contributes to the total viscosity in a major fashion. Although in some countries plasma viscosity has replaced the erythrocyte sedimentation rate, it is clinically less useful. However, it is amenable to the distribution of standardized materials for quality assurance programs.

The greatest clinical use for serum viscosity is for the cluster of diseases known as the monoclonal gammopathies. Some patients have markedly elevated levels of IgM or polymeric forms of IgA or IgG. These intermediate or large-sized immunoglobulins, sometimes elevated from 10- to 100-fold over normal, have a tendency to increase the viscosity of plasma to a hazardous degree. In these instances it is critical that the physician be aware of the presence of the hyperviscosity syndrome, which includes reduced or stagnant blood flow to critical tissues such as the brain, myocardium, pulmonary circulation and the kidney. Symptoms that develop reflect stagnation in these tissues, e.g. seizures, blindness and myocardial, renal and pulmonary failure.

4.5.4 Sieving Chromatography

Column chromatography employing beads of different pore size has been used extensively to purify proteins for special studies and for immunization. A wide range of particulate materials are available commercially for separating protein populations by their molecular size. Even small exclusion volumes allow solvents and salts to enter their pore volume and to be retarded as they

pass through a column. Larger molecules are retarded by entry into gels of appropriate porosity. The simplest use has been to exchange buffers in lieu of dialysis. The protein enters the top of the column in one buffer system, perhaps a denaturing one, and exits in a neutral saline solvent a short time later. Large pore sizes will retard small proteins, allowing larger proteins to exit the column first. Increasing the pore size allows only macromolecules to exit with the exclusion volume. Gels with very large pore size tend to be deformable, resulting in poor flow rates unless the column is pumped in reverse to prevent packing. Often a dye marker such as blue dextran is used to monitor the architecture and performance of a column. The method remains a mainstay of industrial protein purification.

4.5.5 Rotatory Dispersion

All constituent amino acids except glycine are asymmetric. As a result, polarized light is shifted on passage through a solution of polypeptides or serum proteins in a left-handed manner, i.e. levorotatory.[25] This left shift is very sensitive to any denaturation but under carefully controlled conditions and with good instruments it can provide important information about the steric arrangement of amino acids along the polypeptide chain.

4.5.6 X-ray Analysis

Structural features of proteins that can be obtained in pure form and crystallized can be examined at the atomic level with monochromatic X-ray beams.[25] The complex pattern thus produced can be converted to three-dimensional images with modern software programs and modest hardware. The method remains a research tool.

4.5.7 Electron Microscopy

For serum proteins the contributions of electron microscopy[26] have been limited by the resolution of the instruments, which is at best a few angstroms for some instruments, whereas protein size exceeds this. With the addition of heavy-metal shadowing with gold, osmium, and others, features not seen with standard viewing become evident. Nevertheless, combining images of relatively large proteins with X-ray crystallography has allowed researchers to reconstruct what is believed to be the actual physical shape of some proteins, thus enhancing our understanding of function. Blood proteins such as fibrinogen, C-reactive protein, retinol-binding protein, many serum enzymes and immunoglobulins have been imaged, yielding valuable information to aid in understanding their function.

4.5.8 Stoichiometry

All serum proteins whose biological function includes combination with other moieties can be described in terms of molar ratios because they exist in normal and abnormal states. The structure of each molecule includes the description of the number of subunits in the whole and whether they are repeats or are novel. Most serum proteins are composed of an assemblage of subunits linked by bridges or bonds that are subject to dissociation by various enzymes, chaotropic agents or other agents. Such subunits may retain function, as in the case of immunoglobulins, where dissociation can proceed to the point where only the binding site at the end of the Fab′ fragment remains. This ability to bind other moieties by subunits provides useful practical information for clinical medicine: e.g. the binding of one molecule of hemoglobin to a single molecule of haptoglobin, regardless of the haptoglobin subtype; the 1 : 1 ratio indicates that the same mass of hemoglobin will be bound by a given mass of haptoglobin type 1-1 as by many times the mass of the macromolecular isoform type 2-2. The serum level, without the knowledge of the phenotype, can be very misleading as to hemoglobin binding capacity.

A frequently measured ratio is the percent of transferrin saturation with iron or the total iron binding capacity. Normally transferrin, the iron binding protein, remains less than totally saturated with two molecules of Fe^{3+}. In the common hereditary iron storage disease hemochromatosis, high levels of saturation occur, providing a laboratory finding that is often the first sign of the disease. Calculating the transferrin saturation is the most precise means of determining the degree of iron saturation; values above 70% for males and 62% for females are grounds for concern.

In a similar manner, the amount of lipid bound to lipoprotein B, which affects the density, yields information of value in determining risk from atherosclerotic coronary artery disease. The more dense of the low-density lipoproteins (LDLs) appear to have greater atherogenic capacity than the less-dense forms, irrespective of the amount of protein present.

4.5.9 Chain Structure

All proteins possess a linear chain of amino acids as their original backbone. During post-translational modifications, portions of the chain are deleted, cuts occur and new three-dimensional structures are created. Most proteins also possess a secondary branching structure, also composed of amino acids. Most species possess a tertiary structure as the result of portions of these complex chains folding or looping back on themselves and attaching by hydrogen bonds. It is these tertiary structures that

dictate the function for each unique species. Attached to the amino acid backbones or side chains are additional chains composed of carbohydrate, nucleic acids or units of mainly lipid. These also dictate function, either toward other entities or by the manner in which they transport or store the moiety.

The protease inhibitor α_1-antitrypsin has a single chain of 394 amino acids from which carbohydrate side chains are extended (see Figure 6).[3] These side chains can assume either a biantennary configuration or a triantennary form, which combine to create different isoforms whose physicochemical characteristics also differ. The differing isoforms seen in isoelectric focusing result from the carbohydrate side chains terminating with sialic acid residues. As a result, their isoelectric points vary considerably from each other, but under proper conditions of acid pH and gel pore size they separate well in polyacrylamide medium.

The importance of these forms is that certain phenotypic forms of α_1-antitrypsin are synthesized at low levels, placing the host at risk from unrestrained leukocyte proteolytic enzyme activity in the lung spaces, which will progress inexorably to pulmonary failure. The principal reason for searching for these isoforms is to institute lifestyle changes that can mitigate proteolytic digestion and the development of emphysema. Three alleles of the 70 or more described have clinical significance: PI*Z, which produces no more than 10% of the normal amount of inhibitor; PI*S, which produces about 30%; and the exceedingly rare PI*null, which produces none. These figures compare with the normal synthetic product of 50%. In combinations, the Z, S and null alleles can produce subnormal levels of α_1-antitrypsin: PI*ZZ (10–20%); PI*SZ (30–50%); and PI*SS (~60%). Of practical importance is that when most other alleles are present with a deficient allele, levels of the inhibitor are sufficient to prevent proteolysis but the implications are important to the next generation or sibling. A couple who are both heterozygous have a 1 in 4 chance of having an affected child. In Western Europe and North America, the population incidence is 1:1700–1:3000 for the ZZ phenotype.

Haptoglobin, the serum protein that binds hemoglobins in the vascular spaces, exists in three isoforms (see Figure 6).[3] The α-chain is polymorphic (α_1F and α_1S; a third form α_2 is the product of a partial gene duplication with an α-chain twice as long as the α_1 form) whereas the β-chain is not. The products of these combinations are named haptoglobin type 1-1 (one α- and two β-chains), haptoglobin type 2-1 (one α_1-, one α_2- and two β-chains) and haptoglobin type 2-2 (two α_2-chains and two β-chains). These combinations result in molecules of sizes 85, 120 and 160 kDa. The type 2-1 and 2-2 polymers are of very large size, which in high-resolution electrophoretic

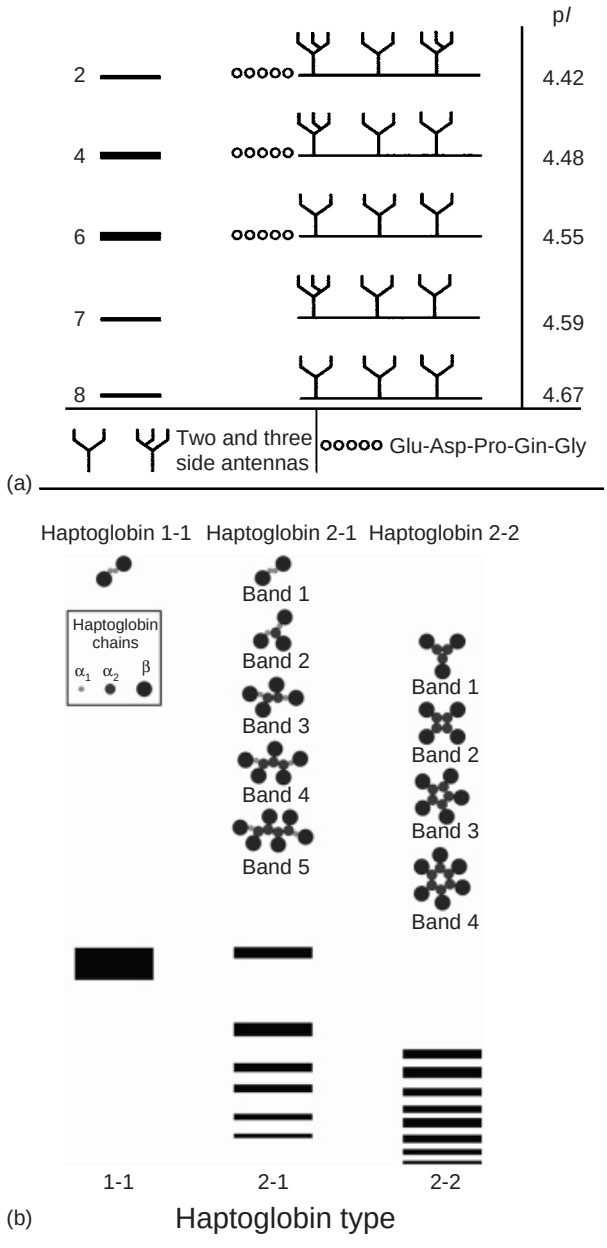

Figure 6 (a) Representation of α_1-antitrypsin isoforms and their branching carbohydrate side chains. (b) Representation of the three isoforms of haptoglobin and their composition of the three chain types α_1, α_2 and β. (Adapted from L.O. Hansson, Karolinska Hospital, Stockholm, Sweden.)

systems produce distinct ladders of haptoglobins, whereas type haptoglobin 1-1 remains a monomer and thus migrates as a single band.

The function of these molecules is stoichiometric binding with free hemoglobin that, as a result of normal red cell turnover, is released into the circulation at low levels. Therefore, any process that increases the release of hemoglobin will consume haptoglobin and depress serum

haptoglobin values if consumption exceeds synthesis. Because of the differing mass of each haptoglobin isoform (e.g. in Caucasian populations, a ratio of 1 : 1.5 : 2), serum levels can only relate to hemoglobin binding capacity as functions of that ratio. Persons with the smallest molecular form will be affected by low-level hemolysis to a greater degree than those with the high-molecular-weight form. Under normal conditions the total haptoglobin pool can bind no more than 3 g of hemoglobin before it becomes depleted. Because of its small molecular weight (14 kDa), hemoglobin will be lost via the kidneys unless it is bound to haptoglobin. There are a variety of ''impact'' sports that result in red cell damage. Red cells are crushed and burst as the result of sudden forceful contact. Originally termed ''bongo hemolysis'' or ''drummer's hemolysis'', it is now recognized that many sports such as hand ball and jogging cause red cell damage. In addition, prosthetic heart valves and roughened surfaces in diseased vessels produce hemolysis. In the first instance cells are physically damaged by valve closure, and in the others because of violent turbulence around the diseased vessel surface: the Venturi effect. All patients placed on cardiopulmonary bypass have accelerated red blood cell damage as a result of cell contact with plastic surfaces and foaming.

4.5.10 Component Parts

Of great importance is the ability of some transport proteins to combine with hydrophobic substances such as lipids, e.g. cholesterol and triglyceride, transporting these hydrophobic substances in plasma to cells as an energy source. Lipids absorbed from the gut are quickly surrounded by the hydrophilic molecules of phospholipids, rendering them soluble in plasma or lymph. This surface coating then accretes the apolipoproteins to form particles whose fate and function are largely dictated by the type of protein associated. Of the major classes, high-density lipoprotein (HDL) follows a different path in that rather than incorporation of lipid particles, HDL incorporates lipid in molecular form. There are at least ten forms of lipoproteins that, in various combinations, create six major classes. Chylomicrons and chylomicron remnants are of intestinal origin, three – VLDL, intermediate-density lipoprotein (IDL) and LDL – are of hepatic origin and one other, HDL, is from hepatic and non-hepatic origin. Yet another classification scheme has been adopted: LpB, LpB : LpE, LpB : CIII, LpB : (a), LpAI, LpAI : AII, LpAI : CIII. Regardless of the classification scheme, each member has a unique role in the metabolism of lipid and the ultimate result of their metabolism, whether normal or predisposing to disease.

4.5.11 Sedimentation Analysis

Between 1923 and 1926, Svedberg developed the technique of ultracentrifugation.[25,27] The technique was so fundamental to the development of protein biochemistry that he was awarded the Nobel Prize in 1926.[4] Under the large gravitational forces of high-speed centrifugation, proteins in solution will sediment differentially at a rate that depends on their mass and, to a minor extent, their shape. By using a solvent with a density gradient, proteins may be separated into layers that reach a stable equilibrium with the density gradient. Under these circumstances, shape no longer influences sedimentation and their molecular mass may be determined. The layers are only stable during centrifugation and so the proteins must be visualized by optical means while the centrifuge is still running and after an appropriate time.

The centrifugal forces employed must be intense and can be achieved only in specially designed instruments. If density gradients are not used, the physical shape of the protein may affect the sedimentation characteristics, especially if the molecule is of a fibrous nature, such as fibrinogen, or globular, such as albumin. When complex solutions are being studied, the proteins may bind to one another, thus altering the rate of sedimentation. This can be used to advantage as evidence of molecular interaction.

Sucrose may be used to create stable gradients through which the test analyte sediments until its buoyant density is reached. Obviously, reaching sedimentation equilibrium can result in flotation when the protein being studied rises rather than sediments. Lipoprotein research has been much advanced by this approach and very large amounts of relatively pure lipoproteins have been recovered for further analysis.

Analysis is performed by continuous observation though the side of the cell using ultraviolet light reflected to an optical visualization system or onto photographic film. The Rayleigh interferometric system is the basis of most systems, where a beam of light is passed through the sides of a cuvette while the ultracentrifugal run is in process. The beams, having been split by the optical system, one passing through the sample and the other directly to the detector, produce interference patterns where boundaries of protein layers occur. More recently, laser illumination has been used to advantage. Measurement of the velocity of the moving boundary between pure solvent and the sedimenting layer of analyte yields the sedimentation coefficient, which is given in Svedberg units (S^0) to honor its discoverer. Molecular masses are recovered from these measurements.

For a period in the late 1960s and early 1970s ultracentrifugal analysis was used to advantage in sophisticated centers caring for patients with diseases of the immune system such as myeloma and Waldenström's disease. With

the advent of inexpensive and precise immunochemical means to quantify these proteins, whose molecular weights have clinical implications for blood flow, sedimentation analysis has become merely a research tool.

Flotation analysis is the application of the same principles to the lipid-rich family of lipoproteins. Ultracentrifugation with the proteins suspended in solutions of varying density results in each protein–lipid micelle floating, to yield flotation constants that provide information about the lipid and protein content.

4.5.12 Physical Features

When whole serum or its fractions can be classified by the effect of external forces on the sample, a new series of descriptors develop, many of which have clinical implications.

4.5.12.1 Thermostability Exposure to cold induces normal physiological events that are all too familiar and often troublesome; however, for individuals who have abnormal serum proteins, especially certain monoclonal immunoglobulins, exposure to cold can be life-threatening. IgM monoclonal components and rarely those comprised of IgG quickly precipitate or gel in the cold, with the result that circulation in chilled areas such as the ears, nose, fingers and especially the toes can halt completely, with resultant gangrene if allowed to persist. In whole blood or plasma, cryofibrinogen can cause gelation, as can very high levels of immunoglobulins with cold-agglutinin activity. There have been rare instances of similar proteins with sensitivity to elevated temperatures (37–45 °C). The consequences are the same as for cryoproteins. In the laboratory, technologists must be constantly alert to thermolability, which is detectable upon thawing of frozen samples or simply in serum samples exposed to refrigerator temperatures overnight or for several days. Gels or precipitates in samples being analyzed can cause serious instrument problems (see Figure 7).

Free monoclonal light chains or Bence Jones protein, usually found in urine, was originally described as a protein that precipitated while urine was heated but dissolved completely on boiling. As the sample is cooled, precipitation again appears but disappears at lower temperatures. This classical behavior is often lacking in this group of interesting patients, however.

4.5.12.2 Binding to Surfaces Some serum proteins, such as fibronectin, have a remarkable affinity for plastics and glass such that preparing a sample for analysis requires effort to prevent this.

4.5.12.3 Color Certain serum proteins can impart unusual color to a serum sample. Normally serum has a yellow tinge as the result of binding of pigments to albumin, but at times the yellow color can be intense or even brownish when the patient is intensely jaundiced, is receiving high levels of penicillin or has been exposed to very large doses of vitamin A as the result of dietary fads (carrot juice). Occasionally, serum takes on a greenish-blue color in pregnancy due to estrogen stimulation of ceruloplasmin synthesis, which contains copper. The serum is green as a product of yellow albumin and blue ceruloplasmin. The most common discoloration of serum is because of the presence of hemoglobin or myoglobin, imparting a pinkish to reddish to brownish color that can interfere with some spectrophotometric analyses if intense. However, artifactual release of hemoglobin and binding to haptoglobin in vitro does not significantly alter the measurement of that protein. In rare instances of crush injuries or severe intravascular hemolysis, free hemoglobin can be visible and will make analysis of haptoglobin impossible: values will be zero.

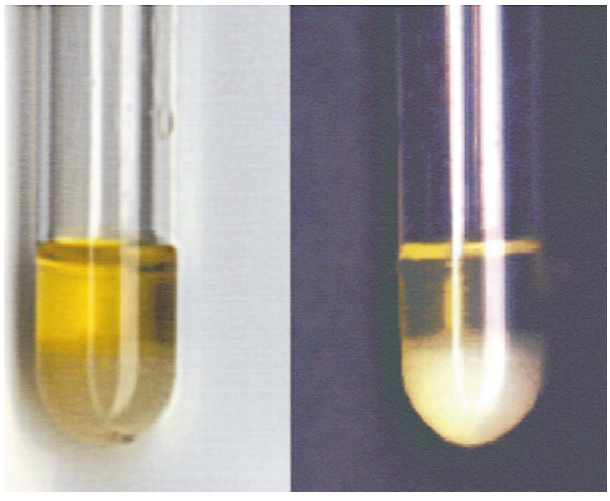

Figure 7 Cryoprecipitate of monoclonal IgM developed after overnight refrigeration in an initially clear serum. Note the large amount of dense material at the base of the tube.

4.5.13 Amino Acid Analysis

A technique now reduced to practical routine in research laboratories is amino acid sequencing. A purified material is digested sequentially and the product of each step is analyzed to determine its nature and contribution to the whole.

4.5.14 Fluorescence Polarimetry

Illuminating molecules at specific wavelengths may result in re-emission of light at different wavelengths or at different axes. The re-emitted light can be quantitated accurately for certain proteins in their native form (see

Table 6 Scale of detection limits

Scale of concentration (log scale)						
Least sensitive						Most sensitive
Albumin ($g\,L^{-1}$) 10^{-1}	C3 ($mg\,L^{-1}$) 10^{-3}	AFP ($\mu g\,mL^{-1}$) 10^{-6}	IgE ($ng\,mL^{-1}$) 10^{-9}	Interleukin-6 ($pg\,mL^{-1}$) 10^{-12}	Research ($fg\,mL^{-1}$) 10^{-15}	Research ($ag\,mL^{-1}$) 10^{-18}

Table 6) or after coupling with compounds that render them fluorescent.

4.6 Electrophoresis

With the introduction of protein separation in an electric field by Tiselius, Svedberg's student, in 1926, investigation of blood proteins took an enormous leap forward.[28,29] At first, methods were labor intensive and took many hours to complete, but since the 1950s they have gradually become rapid, they require microliter quantities of sample and in some cases they are fully automated. Nevertheless, more simple forms of electrophoretic separation remain in widespread use and still retain an economic edge over the automated systems. The interested reader will find that texts on this subject were mostly written during the development phase of the methods and that since these methods, in all their permutations, have become routine little new has been written. Books on the topic are included in the references rather than attempting to select from the enormous number of important papers.

Separation in an aqueous medium results in a variable degree of interaction between the protein and the support matrix used for separating the sample (Table 7).

A great number of supports have been tried over the past 45 years; all share the same basic goal – to prevent ohmic heating from producing convection during migration. Although some convection is inevitable, good separation requires that it be kept to a minimum. In the same way, the effects of diffusion and evaporation, again inevitable, must be minimized. Both goals are met by the introduction of a matrix composed of an insoluble material. Ideally the matrix should not have a charge of its own. In practice, however, many matrices do have electrical charge at the pH of conventional buffers and set up what is called electroendosmotic flow. The effect is that, in its tendency to migrate in the solvent yet being fixed in position, the matrix effectively "pumps" solvent in the opposite direction. Although not totally desirable, it does work to advantage in that proteins with cathodal mobility – a positive charge at the pH of the buffer – are moved backwards and separate in the same manner as those that migrate forward to the anode. With the introduction of hot-water-soluble gels there was the possibility to adjust this endosmosis to suit certain applications. Agarose, a component of agar used in the early phases of this work, provided a variety of features selected by the manufacturer that allow control of endosmosis, in some cases eliminating it completely. Another widely used matrix is a polymerized gel that can be tailored to fit a wide variety of needs. The monomer acrylamide is induced to polymerize by cofactors such as oxidizing agents or ultraviolet light to produce a strong clear gel that can be cast in any form.

Table 7 Separation features for electrophoresis

Solvent	Usually aqueous
Solute	Usually a salt with buffering qualities
pH	Selected to straddle the isoelectric point of the proteins to be studied (~8.6)
Matrix	Selected to prevent convection
Temperature	Usually cool to aid in dissipation of heat generated by the current flow
Voltage	Varying from several hundred to thousands of volts
Current	Direct current and kept to a minimum to prevent ohmic heating
Additives	Selected to modify the species being studied to provide proper separation of bands and to: – dissociate certain species – remove certain or all lipids – prevent coagulation – retain separated species in as near native form as possible for further study

Of great importance was the realization that both specialty agaroses and polyacrylamide could be tailored to produce controlled pore sizes. This provided the ability to produce pore-size gradients so that in an electric field proteins would migrate to a pore size level beyond which their molecular dimensions would prevent further movement. Polyacrylamide, and to a lesser degree agarose, allows pH and size gradients of infinite configurations and in combination. This technique is now a mainstay of protein investigation. After separation, proteins may be visualized by fixation and staining, immunoprecipitation and staining or by direct ultraviolet observation.

4.6.1 Continuous Zone Electrophoresis

As the name implies, the pH and other physical features of the separation matrix and buffer reservoirs are uniform throughout the separation. In an electrical field, ions and

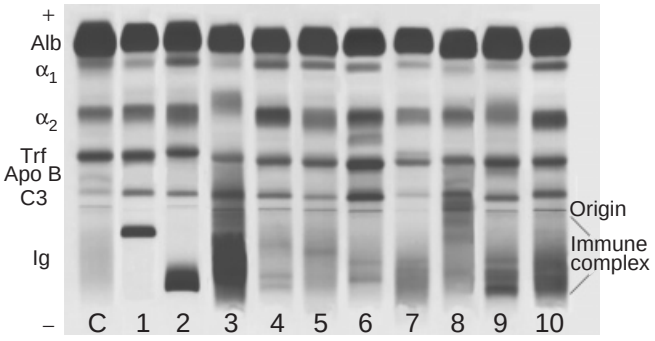

Figure 8 Serum protein electrophoresis in agarose gel at pH 8.6. The anode is at the top and the cathode at the bottom. Proteins separate according to their isoelectric points at the buffer pH. Individual bands (α_1, α_2, C3, Ig, etc.) are actually composites of many superimposed proteins. The apparent exceptions are the heavy bands seen in lanes 1 and 2, which are due to monoclonal immunoglobulins.

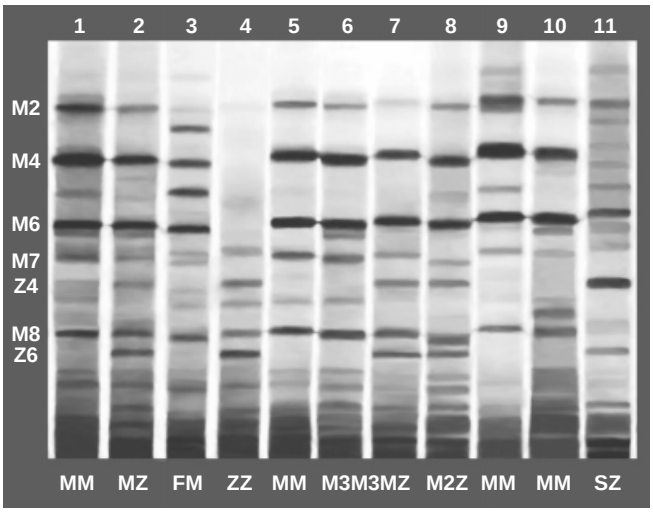

Figure 9 Isoelectric focusing at pH 4.2–4.9 in polyacrylamide gel, showing the distinctive bands of the common phenotypes of α_1-antitrypsin. The anode is at the top.

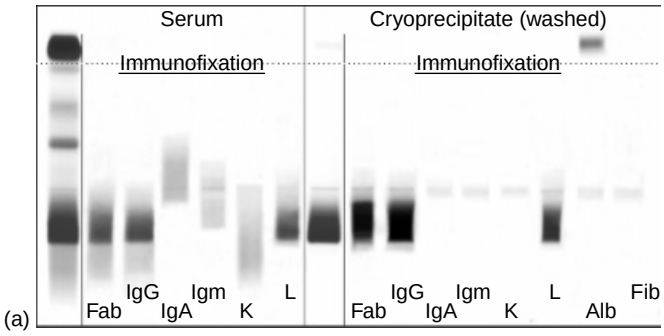

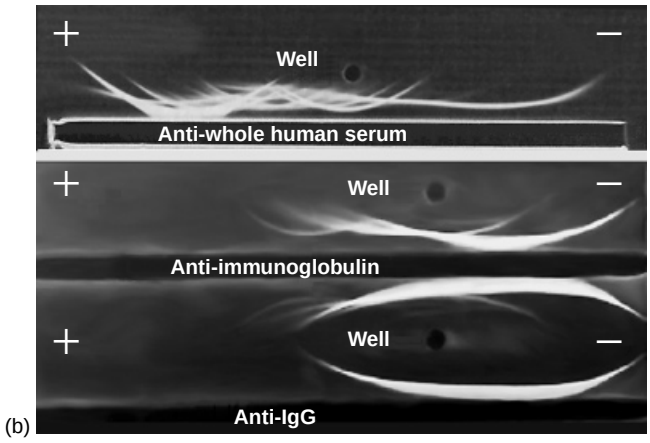

Figure 10 (a) Immunofixation of a cryoprecipitated monoclonal IgG. The sample has been precipitated and washed with saline. The electrophoresis of the original serum and the washed precipitate have been developed with individual antibodies to different proteins, demonstrating that the band seen in the serum electrophoresis on the left is of the IgG class and lambda light chain type. (b) Immunoelectrophoresis developed with three antisera. The samples were placed into the wells and electrophoresed. The troughs in the gel were filled with antiserum and then allowed to diffuse and form specific immunoprecipitate arcs. Fine detail can be observed on magnification.

water move in the direction dictated by their individual p*I* values. Part of this process includes the migration of buffer as a result of electroendosmosis, mentioned above. Many materials have been employed as the separation matrix, whose main purpose is to prevent convection and provide stability within the fluid phase. Early in development came paper, starch, agar gel and even gelatin; later came cellulose acetate, agarose (see Figure 8) and then acrylamide[30] (see Figure 9). Still later, agarose derivatives were introduced as well as combinations of agarose and cellulose acetate. Permutations included antibody cast within the separation gels (see Figure 12) or in wells opposed to the sample well for precipitates to develop in the intervening area (see Figure 10b). The creativity of the worker allowed many novel designs.[14]

A product of advanced technology has been the commercially available, automated capillary zone electrophoresis. Proteins separated on a matrix coated onto the inner surface of a capillary are eluted from the capillary tube to be detected by ultraviolet absorption. Resolution is remarkable, migration time is only a few minutes and sample sizes are minuscule. The format includes automated read-out and graphic result presentation. Coupled with this approach is an application of specific antibody to certain samples, incubation and electrophoresis of the remaining solution. Called ''subtraction capillary zone electrophoresis'', it has application where a monoclonal protein is present and its class identification is desired. By the addition of specific antibody, the protein in question is removed, thus disappearing from the resulting pattern.

4.6.2 Isoelectric Focusing

The next major step came with discontinuous buffers. In this process, strips of matrix are formed in thin glass cylinders or slabs of gel, and recently in very thin hollow silica fibers containing buffer.[28,31] In such a system the acrylamide gel matrix is divided into two phases: the buffer phase at the top and the stacking gel and separation gel at the bottom. The protein sample is placed above a large pore or stacking gel, where the species are able to arrange themselves according to their isoelectric point (p*I*) without being separated by size or configuration. Below this is a third phase, where the pore size is tailored to separate the species according to their velocities (see Figure 9).

The basic structure of each protein species and subspecies dictates differences in p*I*. To capitalize on these differences, which may be very minor, separation occurs over a stable pH gradient created by the incorporation into the matrix of ampholytes (non-protein, low-molecular-weight compounds of set p*I*), which produce pH gradients into which the proteins migrate. Moving toward their isoelectric point, at which their net charge is zero, creates an ever-decreasing zone width for that protein. The result can be a remarkable display of protein bands. This extreme resolution presents a limitation in that at the isoelectric point, where the p*I* is zero, high concentrations of protein tend to form a precipitate unless the procedure is done in very dilute solutions. Many types of ampholines can be acquired commercially to cover the pH zone desired. Gradient pumps are available to assist the worker in producing tailored pH gradients as needed.

4.6.3 Isotachophoresis

In this form of electrophoresis proteins separate into adjacent zones that all migrate at the same rate.[31] Ultimately the bands separate into separate components, with no overlap, and as they are continuously moving down a tube or capillary they can be collected and analyzed. The procedure lends itself to the preparation of pure protein fractions. The separation is by charge alone and can be done in a free liquid but is usually done in a gel slab. A gel of optimal pore size for the molecules of interest is selected, with separation being the result of the pH gradient. As a result, the technique is extremely effective in the study of charge isoforms of a single species. Contaminating proteins are either swept beyond the moving front or are prevented from entering the gel area where the desired proteins separate.

Isotachophoresis can resolve a high concentration of protein into individual bands without promoting precipitation as seen in isoelectric focusing. As a result, the method is able to resolve milligram quantities for preparative purposes.

4.6.4 Immunofixation

This two-step process combines electrophoresis on a support matrix in whatever form, as described above, with a second passive step where antibody is applied to the surface of the gel.[32] This simple step rapidly "fixes" the protein of interest into the matrix. Subsequent washing removes all unreacted antibody and serum proteins so that when the matrix is stained in the usual manner only the antibody–protein complex remains visible. The advantage is that the results are unambiguous and can be compared, side by side, to the original electrophoretic separation. This is often an adjacent comparison strip to minimize slight changes in migration conditions developing during the run (see Figure 10a). Virtually any serum or other body fluid protein can be processed by this technique as long as the antiserum is specific and of sufficiently high titer.

4.6.5 Immunoelectrophoresis

Until replaced by immunofixation, immunoelectrophoresis was the method of choice. It can be credited with raising researcher awareness in the 1950s that serum proteins were an enormously complex family of moieties.[30] A sample is separated in a gel matrix, often on a microscope slide. Adjacent to this, and usually on both sides of each lane, antiserum is introduced into a narrow trench to diffuse laterally into the gel containing the sample. As antiserum and serum proteins diffuse passively towards each other, precipitates develop in the form of arcs. The location and configuration of the arcs (the result of p*I* in the particular system and the molecular weight of the serum protein) dictate where the precipitate forms; the radius of the individual arc is the result of molecular size and concentration (see Figure 10b). The major advantage was that the antigenic identity of the diffusing proteins could be assessed. The principal disadvantage, which eventually led to its replacement, was that the interpretation of the patterns was too often a matter of expert opinion. Artifacts are common and further tax the reader's abilities. Another useful feature of great assistance was that a very creative way could be devised to ascertain partial or complete identity of arcs: by the placement of addition wells, much as was done with simple immunodiffusion. As with all immunodiffusion methods, the potency of the antiserum and its specificity were paramount.

4.6.6 Immunoblotting

Firm binding of protein to membranes provides an efficient means of capturing proteins for subsequent reaction

with antibody or dyes. Several types of membranes are available, each with special characteristics: cellulose acetate, nitrocellulose, nylon, and polyvinylidene fluoride.[33]

Movement of the protein from the matrix to the membrane can be accomplished by simple diffusion aided by blotting papers. This technique is slow and results in some loss of resolution but its simplicity has advantages. A faster technique is to use an electric field: electroelution. The proteins are driven into a membrane with affinity for proteins and then identified by combining with labeled antibody or by direct staining, affording a powerful diagnostic tool analogous to immunofixation (see Figure 11b). Intact or dissociated proteins can be separated, and reagents that interfere with antibody reactions are removed from the membrane prior to reaction with antibody. Blotting can be done on high-resolution acrylamide systems that are not amenable to immunofixation, thus allowing detailed separation and rapid elution to an easily handled membrane.

A variation on this theme is checkerboard immunoblotting (see Figure 11a) where a membrane is exposed to several lanes of fluid, serum or antibody, created by a plastic form but without electrophoretic separation.[34] This step is followed by washing outside the slotted chamber and then its return to the chamber at right angles, where the slots are filled with a reactant: antibody or serum. Final washing and staining provides a qualitative assessment of each combination in the checkerboard

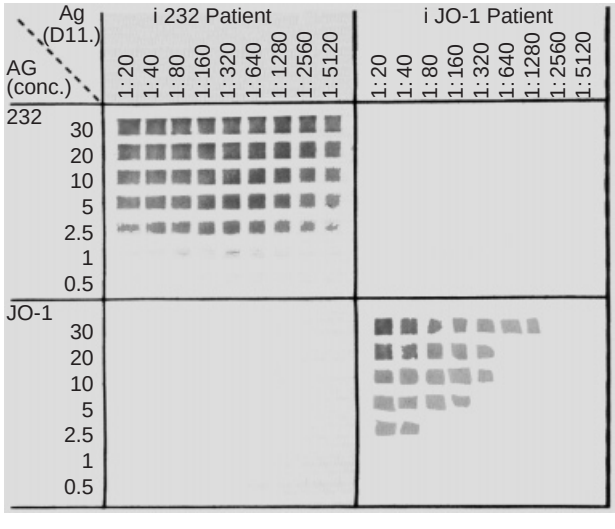

(a)

Figure 11 Two types of immunoblotting: (a) checkerboard immunoblot developed with several dilutions of patient serum and two purified antigens; (b) immunoblot transferred by electroelution after electrophoretic separation of a tissue homogenate and developed with patient sera (lanes 2 and 3), an eluate from a protein A column exposed to the patient shown in lane 2 (lane 4). Lane 1 contains molecular weight standards, electrophoresed and stained.

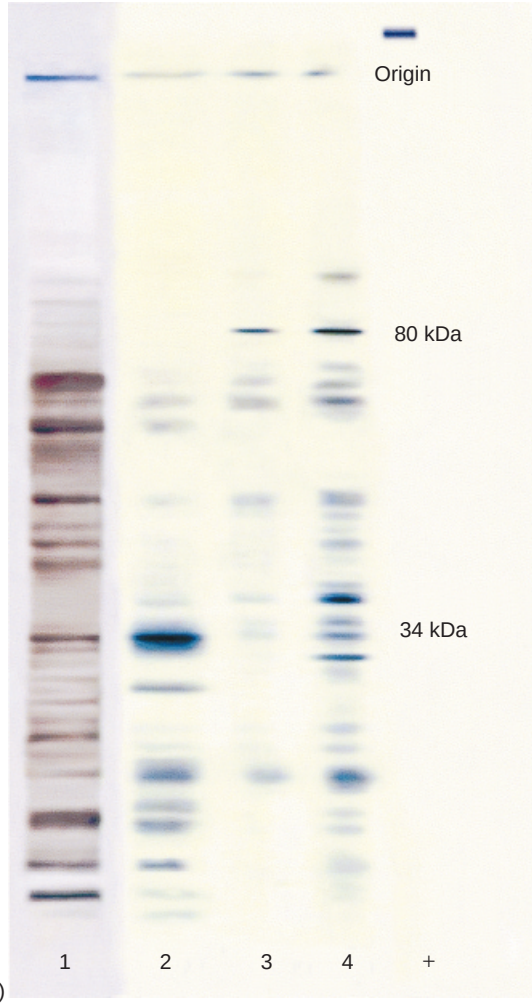

(b)

Figure 11 (*Continued*)

pattern. The method is particularly useful when multiple antibodies are to be used on single samples, or the converse.

4.6.7 Combinations

Great innovation and creativity have characterized investigation in electrophoretic methods. Combinations of methods, one after another, afford a separation power that taxes the workers' ability to understand the results. Only a few merit mention here, but many more can be found in the literature. The main enhancement has been the use of one method followed by a second, applied to all or part of the first separation. Two-dimensional electrophoresis is the most familiar (see Figure 12). In this modification, separation of a single sample along one axis by any of the above methods or in different matrices is completed. The strip is physically removed and applied in the second system, with electrodes rotated at 90°. If the first dimension is simply acrylamide or agarose

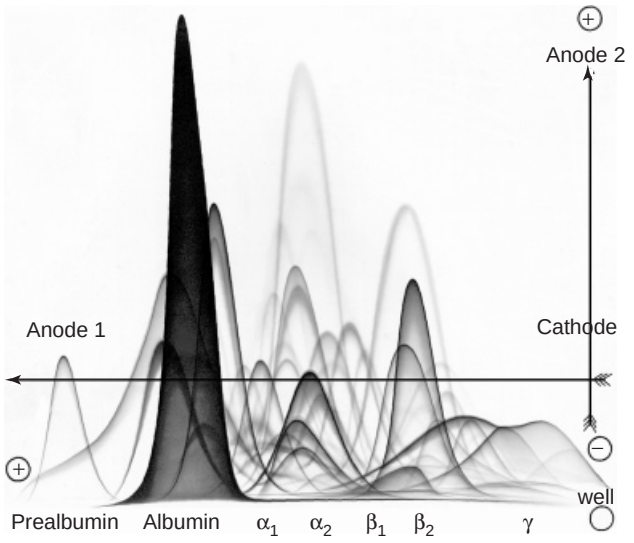

Figure 12 Two-dimensional immunoelectrophoresis in agarose gel. The first dimension was from right to left as a thin strip from the well. The second dimension was from bottom to top, from the strip into a gel containing high-quality anti-whole serum antiserum (Dakopatts). Each "rocket" represents the precipitate developed by one type of antibody and its respective antigen. Over 55 individual antigen–antibody pairs can be identified. Visualization is after washing and staining.

gel separation, the second could be isoelectric focusing, with the result that each apparently single band in the first dimension becomes a family of spots in the second phase. Results can be truly bewildering and interpretation difficult. Devices and protocols have become sufficiently precise to allow the results of different runs to be analyzed by flicker-fusion in the search for significant differences in the hundreds to thousands of spots. Materials to be analyzed can be soluble proteins, tissue extracts reduced with sodium dodecyl sulfate or culture fluids.

Much less sophisticated are combinations employing physical arrangements of sample and antibodies. Two-dimensional immunoelectrophoresis offered creative ways to demonstrate the antigenic identity of related proteins in a manner that was not possible otherwise.[35,36] These methods have largely been relegated to the archives but they may still answer difficult questions not otherwise addressable (see section 4.2.1.3).

5 PATIENT AND SAMPLE PREPARATION

The most elegant and precise assay can be no better than the least precise aspect of a procedure. Uncomfortably, the weakest aspects of every analysis come before and after the actual technological process. Failures at these levels can be so great as to swamp the results of the best analytical performance.[3] Laboratorians tend to forget that the preparation of the subject and the management of the sample itself present risks that can destroy the accuracy and precision of the measured result. To a lesser degree, the reagents are also subject to failure at the laboratory level, at the manufacturer or during shipment. All aspects can be monitored by properly designed quality-control measures but the patient is beyond our control, as will be described below.

All steps in the following sections can affect the final result, such that the most precise measurement can be rendered worthless at best or misleading at worst, by a series of extra-analytical events beyond the control of the measurer. These fall into three broad categories.

5.1 Preanalytical Requirements

The patient is advised to follow a protocol the evening and morning before phlebotomy. The steps are uncomplicated, but should the person or the physician's office staff ignore them, significant alterations in analyte levels can ensue. The request is that the patient follows the steps below:

1. *Have a "light" meal the evening before*. Briefly, this means a meal of modest fat content. A sample taken 1–2 h after a heavy meal may contain a significant amount of chylomicrons, which interfere with the analytical instrument.
2. *Avoid alcoholic beverages from 6 p.m. on the evening prior to the morning of the phlebotomy*. Certain patients may be very sensitive to alcohol-induced hyperlipemia.
3. *Do not smoke cigarettes on the morning of the phlebotomy*. The effects of cigarette smoke are not well understood and do affect certain protein levels.
4. *Drink only water the morning before; no stimulants, no breakfast*. This is for the same reason as in step 1.
5. *Take medications as usual*. If the medication has an effect, it will be long term and should be known to the physician.
6. *Do not engage in vigorous exercise immediately before phlebotomy*. Vigorous exercise can have a marked effect on certain serum proteins in the short term.
7. *Be seated at least 15 min before phlebotomy*. Equilibration of body water into and out of vascular spaces varies with position. Values for bedridden patients are very different from those of ambulatory subjects, but the differences are not usually appreciated.

8. *Use a tourniquet for only a short time and minimally applied*. Prolonged and excessive pressure can alter water content distal to the tourniquet.
9. *Do not shake the vacutainer after filling*. Shaking whole blood or clotted and separated tubes can produce mechanical hemolysis of red cells, which is more marked when red blood cells are already fragile.
10. *Avoid foaming of sample*. Foaming accelerates degradation of sensitive serum proteins.
11. *Cool sample as soon as possible*. Never freeze whole blood. Cooling slows protein degradation markedly. Freezing ruptures red blood cells.
12. *Pay attention to special sample requirements, e.g. anticoagulants, protease inhibitors, chelating agents, etc.* Certain proteins are very sensitive to degradation, and agents to preserve them are needed.
13. *Always mark tubes properly and as ''plasma'' if anticoagulated*. Anticoagulants, if added in liquid form, affect the final concentration as a result of dilution. Demographic information is critical to interpretation.

Each step can introduce analytical or interpretative problems that reduce the accuracy or ultimate meaning of the recovered value.[37] Of critical importance is to supply patient demographics such as age and gender, which affect the reference range against which the result will be compared. Other factors are unquestionably important but at present unknown, such as ethnic group, geography, diet or use of recreational drugs.

Once the sample has been drawn successfully, further care is required. The sample must be allowed to fully coagulate if serum is required. A sample that clots incompletely for any of a variety of reasons may remain on the verge of further clotting. Entry of an incompletely clotted sample into a testing device may result in a partially plugged probe and a reduced aspirated sample volume. In electrophoresis, unclotted or unclottable fibrinogen may mimic a monoclonal immunoglobulin. Once fully clotted, the sample is usually centrifuged to sediment suspended cells and debris. During this procedure the sample should be sealed to avoid desiccation. The device should be temperature controlled to avoid heating and accelerated degradation of sensitive proteins. In some laboratories special separatory tubes are used that do not require centrifugation, but these simple devices can damage cells or allow cells to bypass the plunger if too much pressure is applied. Separation of serum should be performed within 2 h of clotting to avoid release of cellular proteins, including hemoglobin.

Storage temperature is critical. If the samples are to be analyzed during the same working day they should be kept at near 4 °C. If not tested within 48 h the samples should be frozen at −20 °C. Tubes should be tightly capped in either case. During prolonged freezing the sample should fill the tube to more than two-thirds of its volume. If not, use a smaller container to avoid the problem of desiccation within the tube. For long-term storage, −70 °C should be used.

5.2 Intra-analytical Requirements

In laboratories where serum proteins are tested routinely by both quantitative and electrophoretic methods, the procedures should be performed in close proximity to one another. Preferably, the results of serum protein electrophoresis should be known before quantitative analysis is begun, to prevent serious errors as the result of unrecognized extreme elevation in the concentration of some proteins, with resulting antigen excess conditions. Failure can result in the reporting of incorrectly low or normal values when in fact the values are extremely elevated (see section 4.2.1).

Most other intra-analytical failures are approachable by regular use of controls and calibrants. However, there is one very serious situation that often defies monitoring despite the most rigorous training of staff: the mismarking or misinsertion of a sample, with the result that recovered values are attached to the wrong patient. A partial fix for this situation is noted above, where serum protein electrophoresis is available before the quantitative assay; thus, when the final results appear there is a rough check available. Automated bar-coded sample identification is a solution for some situations. However, in a more serious and uncontrollable situation the receipt of a mislabeled sample is irretrievable. It is hoped that the final step by the laboratory physician and the referring physician, upon seeing results that are at odds with clinical judgment, triggers repeat testing for confirmation. Samples that lack the specifics of proper label, name or demographics should be destroyed.

5.3 Postanalytical Requirements

Serum protein studies are usually performed in clusters and not as isolated measurements. Such lone values are virtually meaningless to the physician. In a like manner, the failure to supply patient information seriously compromises the final result and, although numbers are recovered, their clinical meaning is usually lost.[38] The availability of a suspected diagnosis or at least a complaint can increase the diagnostic value greatly. Obvious, but less well appreciated, is the conveying of associated test results such as protein polymorphism, serum description

(color and turbidity), autoantibody test results and the like, for the interpreting physician.

6 LIMITATIONS

The imprecision of reported values for a case will be: $\sqrt{\text{(individual errors)}^2}$. In addition, serum protein values are rarely valuable and useable as single numbers. They are generally analyzed as a cluster, each lending weight and credibility to the others. Possible exceptions are tumor markers, but here also abnormal values can be the result of noncancerous conditions, analytical failures and errors, as discussed above. Clinical use of superior results from carefully prepared patients must still face the rigors of interpretation in the setting of the particular subject.

6.1 Concentration

Alluded to above is the fact that values from patients after prolonged bed-rest can be quite different from the same person when ambulatory, even if no pathology has supervened. The shift of water from extra- to intravascular spaces can result in concentration shifts of as much as 15%. However, in most cases interpretation is carried out without that knowledge. A few workers over the past two decades have tried to address this issue by attempting to normalize values based on the level of oncotically active proteins in the serum sample. Although not ideal, it does mitigate the error induced in this way. Concentration variations as a result of variations in hydration are thus reduced.[38]

6.2 Secondary Factors

Final reported values can be altered by any of several factors either singly or in combination.

6.2.1 Interference

Certain assay techniques are more sensitive to interference than others. As noted above, the clarity of a sample can affect precision, hence the requirement that the patient fast prior to phlebotomy. In most subjects, serum is clearest after an overnight fast and, as a result, interfering chylomicrons are at a minimum. Older methods such as radial immunodiffusion can be performed on even extremely turbid samples and even on tissue extracts or homogenates.

Serum or plasma from some patients with serious autoimmune disease can have very high levels of immune complexes, which at times can produce a visible turbidity that may interfere with the assay systems. The most common cause of opacity, however, is the presence of chylomicrons, as the result of lipid intake before phlebotomy, inherent hyperlipemia or hyperlipoproteinemia. In some patients ingestion of alcohol will produce prolonged hyperlipemia. Administration of corticosteroids for disease control can also produce turbid samples of serum or plasma.

One of the requirements for test kit licensure by the FDA is that the manufacturer has tested the product to ascertain at what level an interfering substance will cause imprecision or inaccuracy. A wide variety of materials, from commonly used drugs to free hemoglobin and bile pigments, are examined. The immunoturbidimetric or immunonephelometric assays are very forgiving of interference by these substances. The sample size used for each assay type also has a significant effect. Assays for albumin or IgG, which use very small sample volumes, can be done on even turbid or hemolyzed samples, whereas assays for complement component C4 and C-reactive protein, which use relatively large sample sizes, are more affected. Heterogeneous ELISAs and RIAs also are largely unaffected by a host of substances that are eliminated from consideration by the washing steps.

6.2.2 Patient Demographics

For a value to be useful to the physician, it must be interpreted by a knowledgeable person or in the future by a capable ''expert system''. Virtually all tests must be interpreted in the light of as many pieces of demographic information as is available. As a result, the information about age and gender is absolutely essential. In time, additional population variables such as diet, medications, geography and ethnicity will also affect the meaning of a test result. Having said that, it must be recognized that once a population variable is thought to be important, reference ranges for that variable must be incorporated into the interpretation. At present, reference ranges for serum proteins covering age and gender are just beginning to appear.[39] In a more complex fashion, information about the individual case will become a crucial factor in extracting clinical meaning from laboratory results. The ideal reference interval in the future will be one based on the single patient's past values, adjusted for age.

6.2.3 Stearic Hindrance

In immunochemical assays, the reaction of the analyte with the antigen can be blunted or totally obliterated by a pre-existing immune reaction that has occurred in vivo. The most common situation is when a test for rheumatoid factor is found to be unreactive when a positive result is expected. In these cases, the IgM rheumatoid factor, which has specificity for altered IgG, appears to be

missing, as though normal. Called ''hidden'' rheumatoid factor, these molecules are partially or totally complexed with circulating altered IgG, thus failing to be detected by the altered IgG-covered reagent particles.

Anti-double-stranded-DNA testing can be converted to a positive reaction if there are anti-histone antibodies in the sample. Because DNA and histones have affinity for each other, the reaction can be mistaken as a true positive anti-double-stranded-DNA when in fact the sample contains free histones that have bound to the substrate DNA and will be recognized by anti-histone antibodies in the same sample. Anti-histone antibodies are induced by a wide variety of medications and do not have the same serious implications as does the presence of anti-double-stranded-DNA.

Many samples contain heterophil antibodies that can produce a positive reaction in a variety of tests. As the name implies, these are human antibodies that react with animal immunoglobulins used in immunological testing. They probably result from immunization in infancy against bovine immunoglobulins present in milk, rabbit proteins present in the diet of some areas and murine proteins seen in impoverished populations. These must be considered as potentially interfering in a significant number of immunological test methods. Examples are the mono-spot test for mononucleosis and other very sensitive immunoassays. Suitable controls are essential.

6.2.4 Dilution

6.2.4.1 Infusions At the time of phlebotomy, several situations can interfere with analyte concentration. Perhaps the most common is the inattention of the phlebotomist to the fact that the patient is being administered intravenous fluids in the limb from which the sample is being drawn.

Less obvious is the fact that the patient may have received infusions of plasma shortly before the sample is drawn. Because most serum proteins have a half-life of many hours or days, an administered unit of plasma or whole blood can significantly alter the true level of an analyte. This is particularly true for those proteins that may be at very low levels prior to samplings, e.g. IgA and complement components.

6.2.4.2 Post-transfusion Banked blood may at times be administered towards the end of the shelf-life of the unit and so red cells are not as robust as normal. As a result, administration of an ''old'' unit of cells may release the hemoglobin of a small fraction of the total, resulting in artifactual hemolysis with subsequent prompt removal of the hemoglobin–haptoglobin complexes thus formed. Five milliliters of lysed red cells will eliminate the total circulating mass of haptoglobin in less than 5 min. Although the administration of aging red cells may be uncommon in sophisticated institutions, it must be considered as a cause of ''hemolysis''. Prudence dictates that serum protein testing should be delayed until 48 h after the last transfusion is completed. It is of interest to appreciate that in vitro release of hemoglobin has little effect on haptoglobin levels because the complexes are still measured in the assay system, whereas in vivo release quickly reduces levels.

Since the blood volume of an average adult is in the neighborhood of 5 L, addition of one unit can affect the concentration by ~7–10%. At times, many units may be required over a short period with a major fraction of the blood volume having been administered. In these circumstances, drawing blood for serum protein levels is not advised.

6.2.5 Accelerated Catabolism In Vivo or Adsorption In Vitro

An analyte of interest may have been removed from the circulation in vivo or from the sample in vitro. In the first instance, intense immune complex formation can remove an analyte from the serum, although this should be viewed as part of the pathological process. A dramatic example is during intense immune complex formation in lupus erythematosus, where complement components are removed as the result of complex formation on glomerular basement membranes. Indeed, the lowered levels are diagnostic. The red cell membrane may also be the target of complement-dependent hemolytic activity.

Certain proteins such as fibronectin have affinity for glass and plastic surfaces, with the result that unless special precautions are taken a blood sample may have markedly reduced levels of such an analyte as the result of vigorous attachment to the inside of the sample tubes. In a very similar manner, antibodies that are temperature sensitive and have a high affinity for red cell membranes can be totally lost to analysis if precautions are not taken to maintain the sample at body temperature throughout the separation process. Cold agglutinins and cryoglobulins present at very high concentrations can be absent from a serum separated at room temperature.

6.2.6 Factitious

Factitious change due to adulteration of samples in an attempt to gain medical attention (Münchausen's syndrome) or to appear ''normal'' for insurance policy acceptance, although rare, is an important consideration. Such a situation was seen by both authors when urine samples sent for protein analysis yielded a very high value and an incredible level of glucose. Further investigation showed that the cause of these very abnormal values in an apparently normal child resulted from the mother, a nurse

in both cases, adding commercial meringue dessert mix to the child's urine sample, ostensibly to prevent the child from having to take part in physical education classes. Although this may be an extreme example, very abnormal results should always raise this issue for consideration. The substitution of another person's serum is, of course, undetectable unless suspicion is raised. Although much more common in drug testing programs, this does also occur in pre-insurance exam testing and should always be considered.

6.3 Proper Reference Ranges

If the laboratory does not use the proper reference ranges, interpretation of the final result will be incorrect[3] (see Figure 13). In addition to variables such as age and gender, several proteins exist as isoforms, each of which is found in the serum at different concentrations. Without knowing which isoform is present in a given patient, the measured value cannot be interpreted with certainty. An example is the measurement of α_1-antitrypsin, which is markedly affected by the genetic phenotype of the subject. At least 75 variants exist, as defined by isoelectric focusing. Only p*I* types MZ, MS, SS and ZZ result in levels that are detectably outside the reference range. These types are also detectable by standard agarose electrophoresis. Therefore, clinical interpretation of α_1-antitrypsin values must take into account the electrophoretic type. Low levels can be functionally normal for certain proteinase inhibitor types as long as total values exceed 0.50–0.60 g L^{-1}. Levels lower than this can lead to progressive damage to lung alveolar membranes from released leukocyte proteases that are uninhibited as a result of inadequate levels of α_1-antitrypsin (see section 4.5.9).

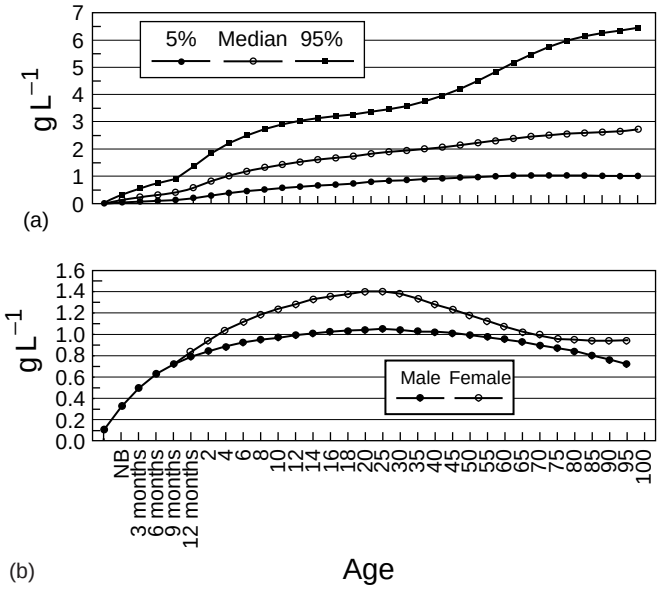

Figure 13 (a) Graphic representation of normal values for one serum protein (IgA) in females from birth to age 100 years. NB, newborn. The 5th, 50th and 95th centiles are shown. (b) Graphic demonstration of the differences in median serum IgM values between males and female during life. (Adapted from Ritchie and Navolotskaia.[3])

Because some serum proteins are influenced by commonplace situations such as pregnancy and use of contraceptive medications, interpreting these proteins should take this knowledge into consideration.

At present, authoritative specific reference ranges for ethnic groups other than Caucasians are unavailable, but such work is in process. Work in the 1960s revealed that Africans domiciled in the United Kingdom had higher immunoglobulin levels than Caucasians. This was also found in African–Americans in the southeastern USA compared to those living in the north.

6.4 Reference Materials and Calibrators

Until recently, serum protein reference materials were only loosely inter-related in spite of the existence of authoritative reference preparations in the possession of the World Health Organization (WHO), The Centers for Disease Control (CDC), the National Institute for Biological Standards and Controls (NIBSC), to name but a few. In 1994, as a result of awareness that there was no reference material consistency among manufacturers of serum protein assay, the International Federation for Clinical Chemistry (IFCC), with financial support from the Bureau de Référence Communitaire (BCR) in Brussels, produced a new certified reference material CRM 470[40] that was manufactured to the highest standards and is now distributed both in Europe by BCR and by the College of American Pathologists (CAP) in the USA. This material, designed for modern immunochemical analysis of serum proteins, is now being used worldwide with the result that interlaboratory harmonization is occurring, as demonstrated by several workers.

6.5 Other Effects

Although not appreciated as yet, a variety of other commonplace circumstances may ultimately be shown to have an effect on what constitutes expected reference ranges. Cigarette smoking has been shown to have a modest effect on IgG, and certain common medications suppress the acute-phase protein C-reactive protein while elevating levels of transthyretin. Other medications and even drugs of abuse may eventually be shown to have an effect. Diet type has a clear long-term effect on levels of lipoproteins, probably not directly but through the intermediary of obesity. Short-term effects of all these factors remain to be investigated.

7 QUALITY CONTROL

The goal of every conscientious laboratorian is to deliver the best and most clinically useful results possible. To achieve this within the confines of the laboratory itself requires that every step of the process be regarded as a pitfall and that ways are designed to uncover errors on a consistent basis[41] (Table 8). Although federal regulation mandates certain checks, each laboratory can add steps that provide a greater degree of certainty that the values that are delivered are of the highest quality.

7.1 Operational Checks

Recognizing that all modern assay methods rely on the stability of increasingly complex instruments and reagents places a requirement on the laboratory to be constantly alert for unforeseen changes. The failure of a system element – reagent or instrument – generally results in easily recognized, aberrant results. Failures of qualitative features are more overt and of quantitative features much less obvious. However, drift can result in gradual change that may be impossible to detect unless specific checks are inserted into every protocol and test run. For example, all illumination devices suffer gradual deterioration in output intensity. In a like manner, detectors also degrade slowly to a given signal strength.

Reagents themselves suffer equivalent degradation as a result of oxidation, denaturation, bacterial contamination or even photooxidation in the case of some fluorescent materials. Yet another source of operational failure is batch-to-batch variation and even vial-to-vial variation, all of which must be anticipated.

Within the reagent family reside calibrants and ultimately reference materials (see below). Often unappreciated by the clinical laboratory will be drift as a result of changes in the tested population, due to either the type of patient (age, gender or race) or the referral of subjects from the health care system itself. Finally, seasonal changes that can expose samples to extremes of temperature can result in accelerated degradation of some, but not all, analytes. Seasonal changes in reference values from a local population have been examined but with little effect noted.

Table 8 Recommendations for optimizing operational performance

- Measure illumination intensity (this may be accomplished automatically in some modern instruments)
- Measure detector performance, e.g. blank readings and baseline levels
- Discard reagents older than the recommended expiration date unless full revalidation is completed
- Do not mix reagent batches
- Aliquot bulk reagents into small, sealed containers
- Insert samples with known concentration values at several concentrations over the dynamic range (at least five)
- Insert samples at the beginning and end of a run whose concentration is known (and in the middle of the run if it is long). Select control samples at various concentrations and consider using occasional samples that are expected to force the system to fail or default to a ''no result'' mode
- Change parts at recommended intervals and never reuse cuvettes that have not been washed properly
- Perform basic quality checks on all purchased reagents regardless of confidence in the supplier

7.2 Dynamic Range

All quantitative assays perform best over a specified span of values: some methods over 10-fold from the minimum level of detection and others over 1000-fold. Called the dynamic range, the span dictates the ideal range for optimal assay performance. Superior performance can be expected toward the center of the range, whereas precision falls off towards the extremes. The degree of imprecision dictates what is the acceptable analytical range within which values can be trusted. Unfortunately, some laboratories attempt to modify assays by decreasing expensive reagent strength, such as antibody, or accepting values that fall near to, but outside, the recommended levels. For immunoassays the result will be that for low values the sample ''noise'' becomes a significant issue and truly undetectable levels will yield an erroneous value or, of more concern, values that fall above a specified level can be in error as a result of antigen excess conditions masking great elevations when recovered raw values exceed the threshold by a small degree. This is a major concern for the immunoglobulin assays, where instrument manufacturers have gone to great lengths to incorporate procedures that prevent a final answer if the value is too high. The operator must then reinsert a diluted sample or more antiserum and re-assay the sample. The instrument itself may do the equivalent without operator intervention. Immunoassays now perform to the same degree of precision/imprecision that can be expected from traditional colorimetric assays.

8 REGULATORY ISSUES

With serum protein testing worldwide having topped $4 billion, it is not surprising that competition is intense and the desire for profit climbs inexorably. With it has come an increasingly complex fabric of regulations that apply equally to the manufacturer and the laboratory. In addition, because the force that drives this commerce is

the physician ordering the tests, regulations are applied to their practices as well.

These regulations, although increasing costs at all levels, are absolutely necessary to protect the entire system, particularly the patient being tested. The result of regulation has been that products today are of greatly increased standardization and efficacy and the effort made by commercial companies in verifying their product claims has contributed significantly to the total medical knowledge base.

Many organizations play roles in stabilizing the laboratory worldwide. Perhaps best known is the WHO, which has shepherded the development and storage of reference materials for decades. Among them have been primary reference preparations for many serum proteins. In a like manner, the NIBSC in the UK and the National Institute for Standards and Technology (NIST) in the USA have contributed to the current excellent state of the art in reference materials. From a somewhat different perspective, the IFCC and the National Committee for Clinical Laboratory Standards (NCCLS) have provided initiatives for the onerous task of providing regulatory guidelines. From yet another aspect, that of quality assurance, CAP and other national quality assurance organizations have developed and executed programs that give all laboratories the means to assess, and correct if need be, their actual performance. The BCR of the European Communities has generously supported protein-related projects, one of which produced CRM 470. Finally, the United States FDA brought all these efforts to focus, at least in the USA, by providing a required framework in which industry, laboratories and professional societies must focus on the goal of providing quality products for the clinical laboratory and appropriate results for clinical use.

9 CLINICAL APPLICATIONS

Changes in the concentration, structure or function of proteins in plasma or other body fluids may be determined genetically and result directly in disease due to an abnormality or deficiency in a specific protein. More commonly, changes are secondary to a disease process affecting protein synthesis, distribution or catabolism. The laboratory methods for the investigation of plasma proteins are those of measurement, separation and characterization.[42–45]

Genetic variants, although not common individually, form an important group for the clinical laboratory to investigate because they are often essential for the diagnosis of a disease process. Secondary changes vary enormously in clinical usefulness; they may be pivotal to a diagnosis, such as the finding of a monoclonal component in myeloma, or simply be a part of the jigsaw of diagnostic information that allows us to assess the probability of a particular disease. In an increasing number of cases, sequential protein measurements are crucial in the monitoring of disease activity and response to treatment. More recently, the concept of triaging or sorting according to urgency, to allow rational use of expensive treatments, has become important in medicine. The plasma proteins, as sensitive markers of key processes such as inflammation, have an important and developing role in this field.

It is thus important that both qualitative and quantitative techniques are available in the laboratory. Plasma proteins fall into a series of functionally related groups, which makes it useful to define profiles of protein measurements to address pathophysiological issues. In many cases, alterations in molecular characteristics occur alongside changes in concentration, thus demanding the complementary use of separation and characterization techniques. Simple electrophoresis on cellulose acetate or agarose is a most useful tool providing a considerable amount of information that complements quantitative specific protein measurements. It also allows the identification of a number of genetic variants that result in changes in electrophoretic mobility. The identification of proteins separated by electrophoresis and by the use of immunofixation or immunoelectrophoresis, further increases the discriminatory power of qualitative examination.

For proteins present in plasma at concentrations above $1\,mg\,L^{-1}$, quantitation may be done by gel diffusion, electroimmunodiffusion, nephelometry and turbidimetry. Below this concentration, RIA and ELISA are the most widely used techniques. Functional assays are essential for proteins that are present in normal concentrations but whose function may be defective.

9.1 Study of Protein Polymorphism with Clinical Implications

Any large biological molecule can present variations definable by some technique, either chemical or functional, or by any method that will separate one isoform from another. For serum proteins, electrophoresis has been used for decades. For example, a serum protein with more than one phenotype is measured by at least one technique. For example, there are three major genetically controlled phenotypes of the iron-binding protein transferrin – a ''normal'' TfC, a fast mobility TfB, and a slow-form TfD – all of which have slightly different amino acid content. However, there are an additional 38 rare variants known. To demonstrate the various forms, high-resolution electrophoresis is required. If phenotype description includes alterations in molecular structure as

a result of nongenetic effects, additional forms can be identified for transferrin. The normal form of transferrin can be separated by isoelectric focusing into four isoforms appearing at p*I* 5.2, 5.4 (the major fraction), 5.6 and 5.7. Carbohydrate-deficient transferrin produced in high percentage (>1%) as the result of alcohol abuse can be separated into two bands at p*I* 5.7 and 5.9, the latter not being present in normal serum. In addition, transferrin in whatever form can bind one or two molecules of Fe^{3+} in the circulation. The iron-devoid form, apotransferrin, exists near the sites of iron incorporation in the gut and represents yet another form of the molecule.

This phenotypic division and subdivision of a single protein can be seen often among serum proteins. In certain circumstances the presence of a phenotypic variant leads to overt clinical disease directly, whereas in other circumstances disease results as an indirect consequence of inadequate or dysfunctional synthesis of the variant protein. An example of disease resulting directly from overproduction of an apparently normal protein, IgM, is Waldenström's macroglobulinemia. Normal levels are in the 0.4–3 g L^{-1} range in adults but, when a malignancy of IgM B-lymphocytes exists, synthesis can reach extreme levels. This protein with an M_r of 97 000 Da can reach 60–100 g L^{-1}, producing hyperviscosity, sludgeing in small vessels and resultant cyanosis, seizures, chest pain, poor lung perfusion, neuropathy, etc. A very large number of diseases result from the failure of specific protein synthesis. It is beyond the scope of this chapter to discuss even a limited number, but a few examples are listed in Table 9.

9.1.1 Example: Waldenström's Macroglobulinemia

Case history An 83-year-old, apparently healthy, elderly physics professor came to his local physician with a history that dated to the previous winter, of bluish discoloration of the tip of his nose, fingers and toes and the pinnae of both ears, on being out of doors. The discoloration had not been noticed during the warm months. This gentleman offered the opinion that "something is making my blood thick in the cold".

The physical examination revealed little except some palor and some scaring of the tip of his nose. He was otherwise well nourished, alert and without complaint. His entire physical examination was otherwise unremarkable.

His laboratory examination, however, provided many abnormal findings. The man's chilled blood, before the serum was separated, was noted to contain gelatinous strands that disappeared on warming. After centrifugation at room temperature the serum was clear yellow and unremarkable. The sample was then refrigerated routinely to await analysis. Two hours later, the sample was removed in preparation for transfer to a chemistry autoanalyzer. The technologist noted that the sample had separated into two phases, as shown in Figure 7. Placing the sample in a water bath at 37 °C cleared it and the tests were performed without difficulty. One of the tests run later was serum protein electrophoresis, which showed a very large, dense protein band in the region of the application point and an apparent precipitate that could be removed easily from the surface. Subsequent measurement of immunoglobulins revealed that the IgG level was at the 20th centile, IgA at the 26th centile and IgM was 35 g L^{-1} (about 35 times normal). Other chemistries were unremarkable. His red cell smear showed marked aggregation of red cells into chains, called rouleaux, and the red cells were hypochromic. A urine sample was found to contain a very large amount of protein, which later proved to be free monoclonal light chains (Bence Jones protein) of the kappa type. Immunofixation electrophoresis confirmed the band and the gel seen in the chilled tube to be monoclonal IgM of the kappa type as well.

His electrocardiogram showed some evidence of ischemia and the chest X-ray showed prominent pulmonary marking throughout. A careful re-examination of his optic fundi revealed mild evidence of segmenting of the blood column in the retinal vessels. His diagnosis was Waldenström's macroglobulinemia. His bone marrow aspirate was not remarkable, with the exception of a

Table 9 List of some phenotype-related diseases

Protein	Disease	Pathology
Failure of synthesis of lipoprotein AI	Tangier disease	Xanthomatosis
Overproduction of lipoprotein B100	Arteriosclerotic disease	Myocardial infarction
Synthetic failure of immunoglobulins	Immune deficiencies	Infection
Synthetic failure of complement factors	C3 deficiency	Infection
Dysfunctional ceruloplasmin	Wilson's disease	Liver and neurological disease
Failure of haptoglobin synthesis	Cholelithiasis	Excess bilirubin excretion
Diminished synthesis of transthyretin		Diminished ability to bind T3 and T4
Hypomorphic forms of α_1-antitrypsin	Emphysema, cirrhosis	Progressive lung disease in certain homozygous states
Failure of synthesis of blood clotting factors	Von Willebrand's disease, etc.	Bleeding diatheses

slight increase in mature lymphocytes. The patient refused chemotherapy of any sort, but agreed to return visits.

Four months later he returned, having been brought to the office at the insistence of a family member. The history now included bouts of disorientation, shortness of breath and weakness. Physical examination revealed that blood flow in the retinal vessels was interrupted by segments of clear plasma, his respirations were rapid and he was clearly disoriented as to time, person and place. An electrocardiogram showed evidence of myocardial ischemia and a chest X-ray showed more marked dilatation of the small pulmonary vessels.

Repeat laboratory examination revealed that his IgM level had risen to $60\,g\,L^{-1}$. He was now showing signs of renal failure, reduced pulmonary perfusion resulting in shortness of breath, early congestive heart failure and myocardial ischemia, poor cerebral perfusion resulting in disorientation and confusion and worsening of his cold sensitivity. The unusual fundoscopic finding of segmented blood flow was also the result of sludgeing. It was clear that all these physical features were being caused by increased blood viscosity as a result of the very high levels of the large immunoglobulin IgM (molecular weight 971 kDa). He was plasmaphoresed of 4 units over 24 h (removal of plasma) to rapidly reduce his blood viscosity and to prevent clinical collapse. After removal of the 4 units, all his acute symptoms cleared. He was then started on a program of chemotherapy to reduce the neoplastic cell mass responsible for the massive production of the macroglobulin.

GLOSSARY

The definitions for this glossary come from a glossary produced by the NCCLS,[46] the OVID Technology Medline Database and miscellaneous textbooks. The form in which it is presented is taken directly from Ritchie and Navolotskaia,[3] with permission.

Accuracy	The nearness of a measurement to its accepted or true value.
Agglutination	The clumping of cells, microorganisms, and so forth, caused by antibody or other natural or synthetic chemical factors.
Amino acids	Simple nitrogenous organic compounds that constitute all peptides, polypeptides and proteins.
Analyte	The substance, a set of substances, or ''factor'' to be assayed.
Antibody	A glycoprotein that is produced by B cells in response to exposure to an antigen and reacts specifically with that antigen; a part of humoral immunity.
Antibody diversity	The phenomenon of immense variability characteristic of an antibody, which enables the immune system to react specifically against unlimited antigens and is generated by the rearrangement of variable-region gene segments during the differentiation of antibody-producing cells.
Antibody specificity	The property of antibodies that enables them to react with some antigen determinants and not with others.
Antigen	A substance that is capable, under appropriate conditions, of inducing a specific immune response and of reacting with the product of that response (antibody or specifically sensitized T lymphocyte).
Antiserum	A serum containing antibodies.
Autoantibody	An antibody directed against a self-antigen, i.e. against a normal tissue constituent.
Biantennary	A chemical structure including two major arms.
B Lymphocytes	The bone-marrow-derived lymphocytes that are defined by the presence of endogenously produced immunoglobulins; a part of humoral immunity
Calibrator	A material of known characteristics (concentration, activity, reactivity) used to calibrate or adjust a measurement procedure or to compare the response obtained with the test specimen. The material must be of the same performance characteristics and can be used to standardize field methods.
Cathode	Negatively charged pole in electrophoresis.

Chaotrope	A substance that promotes dissociation of immunological complexes with minimal damage to the constituents.
Cytokine	Nonantibody protein, secreted by inflammatory leukocytes and some non-leukocytic cells, that acts as intercellular mediator.
Denaturation	Loss of the native configuration of the macromolecule, usually with resulting loss of biological or immunological reactivity or solubility.
DNA	Deoxyribonucleic acid, the principal chemical structure of heredity. Composed of ribose phosphate units in various combinations.
Electroendosmosis	Migration of buffers as the result of the presence of charged groups on the fixed matrix. The result is the flow of fluid rather than the charged particles that are anchored, e.g. agarose or cellulose acetate.
Electroimmunoassay	An immunoassay where antigen and antibody are forced into association, e.g. ''rocket immunoassay'' or ''electroimmunodiffusion''.
Electrophoresis, agarose	A form of electrophoresis in which protein molecules move through buffer suspended in an agarose gel matrix.
Electrophoresis, cellulose acetate	A form of electrophoresis in which protein molecules move through buffer suspended in a cellulose acetate matrix.
Electrophoretic mobility	A number that describes the movement of a protein in an electric field, usually in pH 8.6 barbital buffer of given ionic strength. The term can also be qualitative, e.g. α-, β-, γ-, etc.
ELISA	ELISA – an immunoassay utilizing an antibody labeled with an enzyme marker. The change in enzyme activity as a result of the enzyme–antibody–antigen reaction is proportional to the concentration of the antigen and can be measured.
Epitopes	The sites either on or within the antigen with which antibodies and T-cell receptors react.
Fab region	Fragment antigen-binding region of immunoglobulin molecule that contains an entire L-chain and amino-terminal half of the H-chain and is responsible for binding to the antigen.
False negativity	Negativity in a diagnostic or screening test performed on a specimen from an individual having the disease that the test is designed to detect.
False positivity	Positivity in a diagnostic or screening test performed on a specimen from an individual not having the disease, or other finding, that the test is designed to detect.
Genetic variant	A variation in a single molecular species that is defined by heredity.
Genotype	The genetic constitution of the individual.
Hemolysis	Disruption of the integrity of the red cell membrane, causing release of hemoglobin.
Heterozygote	A condition where two or more alleles exist at one genetic locus.
Immune complexes	The complexes formed by the binding of antigen and antibody molecules.
Immunoblotting	Immunological method for isolating and semi-quantitatively measuring immunoreactive substances.
Immunodiffusion	Technique involving the diffusion of antigen or antibody through a semisolid medium, usually agar or agarose gel.
Immunofixation	Method of precipitating an electrophoresed protein in an electrophoretic matrix.
Immunofluorescence microscopy	Microscopy of specimens stained with fluorescent dye bound to an antibody or antigen that emit light when

	exposed to ultraviolet or blue light.
Isoelectric focusing	Electrophoresis in which a pH gradient is established in a gel medium allowing proteins to migrate until they reach the site at which the pH is equal to their p*I*.
Isoelectric point (p*I*)	The pH at which the protein has a net charge of zero.
Isoform	One of several usually genetically determined forms of a single protein having the same antigenic structure yet differing in amino acid or carbohydrate content or configuration.
Isotachophoresis	Electrophoresis in which a pH gradient is established in a gel allowing proteins to migrate until they reach the zone at which their p*I* prevents further migration.
L (light)-Chains	The shorter polypeptide chain of the immunoglobulin molecule.
Microheterogeneity	Defines different forms of the same molecular species that are evident after special treatment.
Monoclonal antibodies	Antibodies produced in vitro by a cell line arising from a single cell. All molecules will be of a single class and subclass and with a single antigenic specificity.
Multiple myeloma	A malignant tumor of plasma cells usually arising in the bone marrow.
Nephelometry	A means of analyzing the amount of material in suspension in an optically clear fluid, by measuring the amount of light reflected from the suspended particles.
Phenotype	The end result of both the genetic and environmental factors giving the identifiable structural and functional characteristics of an organism or a protein.
pH Gradient	A system of buffers that create a controlled change in pH in an electrophoretic separation medium, e.g. agarose, polyacrylamide gel, hydrolyzed starch. The slope of change can be gradual, linear, discontinuous, etc.
Plasma cell	The type of mature B cell that actively secretes large amounts of a specific antibody.
Polymerization	The formation of a large molecule from identical smaller units.
Polymorphism	The condition in which a protein species exists in different forms.
Polypeptide	A substance composed of two or more peptides.
Precision	The extent to which replicate analyses of a sample agree with each other; usually expressed as imprecision (the coefficient of variation of a population of values, equal to the standard deviation divided by the mean).
Proteinase	A class of enzymes that split peptide bonds to form simpler compounds.
Proteinase inhibitor	A protein or other chemical that is capable of blocking the digestive enzymatic activity of a proteinase.
Radioimmunoassay	A quantitative assay for the detection of antigen–antibody reactions using a radioactively labeled substance to measure the binding of the unlabeled substance to a specific antibody or other receptor system.
Reference material	A material or substance in which one or more properties are sufficiently well established for use in calibrating or verifying a measurement, method or apparatus.
Reference values	The range or frequency distribution of a measurement in a population.
Rheumatoid factor	Antibody directed against antigenic determinant in the Fc region of IgG.
Sensitivity	A measure of the probability of correctly diagnosing a condition (for

	assessing the results of diagnostic and screening tests).
Specificity	A measure of the probability of correctly identifying a nondiseased person (for assessing the results of diagnostic and screening tests).
Steric hindrance	Interference based upon shape or proximity of adjacent structures.
Triantennary	A chemical structure including three major arms.
Turbidimetry	A means of analyzing the amount of material in suspension in an optically clear solution, by means of measuring the amount or percentage of light that passes through a given suspension of material.
Waldenström's macroglobulinemia	A malignant neoplasm of cells with lymphocytic, plasmacytic, or intermediate morphology that secrete an IgM component.

ABBREVIATIONS AND ACRONYMS

AFP	α_1-Fetoprotein
BCR	Bureau de Référence Communitaire
CAP	College of American Pathologists
CDC	Centers for Disease Control
CSF	Cerebrospinal Fluid
ELISA	Enzyme-linked Immunosorbent Assay
FDA	Food and Drug Administration
FITC	Fluorescein Isothiocyanate
HDL	High-density Lipoprotein
IDL	Intermediate-density Lipoprotein
IFCC	International Federation for Clinical Chemistry
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgM	Immunoglobulin M
LDL	Low-density Lipoprotein
NCCLS	National Committee for Clinical Laboratory Standards
NIBSC	National Institute for Biological Standards and Controls
NIST	National Institute for Standards and Technology
RIA	Radioimmunoassay
VLDL	Very Low-density Lipoprotein
WHO	World Health Organization

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Fluorescence-based Biosensors

Carbohydrate Analysis **(Volume 1)**
Glycoprotein Analysis: General Methods

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Capillary Electrophoresis in Clinical Chemistry • Immunochemistry • Phospholipids of Plasma Lipoproteins, Red Blood Cells and Atheroma, Analysis of • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Point-of-care Testing • Statistical Quality Control in Clinical Laboratories • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry • Urinalysis and Other Bodily Fluids

Food **(Volume 5)**
Electrophoresis and Isoelectric Focusing in Food Analysis • Proteins, Peptides, and Amino Acids Analysis in Food • Sample Preparation for Food Analysis, General

Forensic Science **(Volume 5)**
Immunoassays in Forensic Toxicology

Particle Size Analysis **(Volume 6)**
Centrifugation in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Chromatography of Membrane Proteins and Lipoproteins • Fluorescence Spectroscopy in Peptide and Protein Analysis • Gel Electrophoresis in Protein and Peptide Analysis • Protein Purification: Theoretical and Methodological Considerations • Proteolytic Mapping

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Refractive Index Technology as a Real Time Viscosity Technique

Pharmaceuticals and Drugs **(Volume 8)**
Proteins and Peptides Purification in Pharmaceuticals Analysis

Polymers and Rubbers **(Volume 9)**
X-ray Scattering in Analysis of Polymers

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. F. Putnam, *The Plasma Proteins*, Academic Press, New York, Vols. I–V, 1975–1987.
2. M. Sela, *The Antigens*, Academic Press, London, Vols. I–VII, 1973–1987.
3. R.F. Ritchie, O. Navolotskaia, *Serum Proteins in Clinical Medicine*, FBR, Scarborough, Vol. I, 1996.
4. http://www/nobel.se/prize/index.html.
5. A.J. Pesce, L.A. Kaplan, *Methods in Clinical Chemistry*, CV Mosby, St. Louis, 1987.
6. M.A. Rothschild, T. Waldmann, *Plasma Protein Metabolism; Regulation of Synthesis, Distribution and Degradation*, Academic Press, New York, 1970.
7. R.F. Ritchie, O. Navolotskaia, *Serum Proteins in Clinical Medicine*, FBR, Scarborough, Vol. II, 1999.
8. N.R. Rose, H. Friedman, J.L. Fahey, *Manual of Clinical Laboratory Immunology*, 3rd edition, American Society of Microbiology, Washington, 1986.
9. N.R. Rose, E.C. deMacario, J.D. Folds, H.C. Lane, R.M. Nakamura, *Manual of Clinical Laboratory Immunology*, 5th edition, American Society of Microbiology, Washington, 1997.
10. J.W. Goding, *Monoclonal Antibodies: Principles and Practices*, Academic Press, London, 1986.
11. J. Kusnetz, H.P. Mansberg, 'Optical Considerations: Nephelometry', in *Automated Immunoanalysis*, ed. R.F. Ritchie, Marcel Dekker, New York, 1–43, Vol. 7, 1978.
12. J.C. Sternberg, 'A Rate Nephelometer for Measuring Specific Proteins by Immunoprecipitin Reactions', *Clin. Chem.*, **23**, 1456–1464 (1977).
13. N.R. Rose, P.E. Bigazzi, *Methods in Immunodiagnosis*, 2nd edition, John Wiley & Sons, New York, 1980.
14. A. Crowle, *Immunodiffusion*, 2nd edition, Academic Press, New York, 1973.
15. T.M. Phillips, J.M. Krum, 'Recycling Immunoaffinity Chromatography for Multiple Analyte Analysis in Biological Samples', *J. Chromogr., B*, **715**, 55–63 (1998).
16. J.J. Cavallaro, D.F. Palmer, P.E. Bigazzi, *Immunofluorescence Detection of Autoimmune Diseases*, US Department of Health and Education and Welfare, CDC, Atlanta, GA, 1976.
17. R.C. Nairn, *Fluorescent Protein Tracing*, 4th edition, Churchhill Livingstone, Edinburgh, 1976.
18. N.R. Rose, E.C. deMacario, J.D. Folds, H.C. Lane, R.M. Nakamura, *Manual of Clinical Laboratory Immunology*, 5th edition, American Society of Microbiology, Washington, 1997.
19. J.R. Crowther, *ELISA: Theory and Practice. Methods in Molecular Biology*, Humana Press, Totowa, NJ, Vol. 42, 1995.
20. P.L. Howard, T.D. Trainer, *Radionuclides in Clinical Chemistry*, Little Brown and Company, Boston, 1980.
21. T. Chard, *Laboratory Techniques in Biochemistry and Molecular Biology*, North-Holland Publishing, Amsterdam, 1981.
22. W.H. Hannon, et al., *Assessing the Quality of Radioimmunoassay Systems*, 2nd edition, NCCLS Document LA1-A2, NCCLS, Villanova,Vol. 14, No. 17, 1994.
23. J. Porath, T. Kristiansen, 'Biospecific Affinity Chromatography and Related Methods', in *The Proteins*, ed. N. Neurath, Academic Press, London, 95–178, 1975.
24. N. Sharon, H. Lis, 'Lectins as Cell Recognition Molecules', *Science*, **246**, 227–246 (1989).
25. H.E. Schultze, J.F. Heremans, *Molecular Biology of Human Proteins*, Elsevier, Amsterdam, 1966.
26. J.T. Finch, 'Electron Microscopy of Proteins', in *The Proteins*, ed. N. Neurath, Academic Press, London, 413–497, 1975.
27. K.E. Van Holde, 'Sedimentation Analysis of Proteins', in *The Proteins*, ed. N. Neurath, Academic Press, London, 225–291, 1975.
28. B.D. Hames, D. Rickwood, *Gel Electrophoresis of Proteins*, IRL Press, Oxford, 1987.
29. L.A. Lewis, J.L. Opplt, *Handbook of Electrophoresis*, CRC Press, Boca Raton, FL, Vol. I, 1980.
30. H.E. Whipple, M.I. Spitzer, R. Morea, *Gel Electrophoresis*, New York Academy of Sciences, New York, Vol. 121, 1964.
31. N. Catsimpoolas, *Isoelectric Focusing and Isotachophoresis*, New York Academy of Sciences, New York, Vol. 209, 1973.
32. R.F. Ritchie, R. Smith, 'Immunofixation I. General Principles and Application to Agarose Gel Electrophoresis', *Clin. Chem.*, **22**, 497–499 (1976).
33. B.A. Bardo, E.R. Tovey, *Protein Blotting*, Karger, Basel, 1989.
34. R.F. Ritchie, M. Collins, T.B. Ledue, 'A Flexible, Efficient, Checkerboard Immunoblot System for the

Detection and Semiquantitation of Specific Antinuclear Antibodies', *J. Clin. Lab. Anal.*, **6**, 391–398 (1992).
35. N.H. Axelsen, J. Krøll, B. Weeke, *Quantitative Immunoelectrophoresis: Methods and Applications*, Blackwell Scientific, Oxford, 1978.
36. N.H. Axelsen, *Handbook of Immunoprecipitation Techniques*, Blackwell Scientific, Oxford, 1983.
37. D.S. Young, *Effects of Preanalytic Variables on Clinical Laboratory Tests*, 2nd edition, AACC Press, Washington, DC, 1997.
38. R.B. Friedman, D.S. Young, *Effects of Disease on Clinical Laboratory Tests*, 3rd edition, AACC Press, Washington, DC, 1997.
39. R.F. Ritchie, G.E. Palomaki, L.M. Neveux, O. Navolotskaia, T.B. Ledue, W.Y. Craig, 'Reference Distributions for Immunoglobulins A, G and M: A Practical, Simple and Clinically Relevant Approach', *J. Clin. Lab. Anal.*, **12**, 363–370 (1998).
40. J.T. Whicher, R.F. Ritchie, A.M. Johnson, S. Baudner, J. Bienvenu, S. Blirup-Jensen, et al., 'New International Reference Preparation for Proteins in Human Serum (RPPHS)', *Clin. Chem.*, **40**, 934–938 (1994).
41. J.H. Rippey, R.M. Nakamura, *Diagnostic Immunology: Technology Assessment and Quality Assurance*, CAP, Skokie, IL, 1983.
42. N.B. Tietz, R.B. Conn, E.L. Pruden, *Applied Laboratory Medicine*, WB Saunders, Philadelphia, 1992.
43. N.W. Tietz, *Clinical Guide to Laboratory Tests*, 3rd edition, WB Saunders, Philadelphia, 1995.
44. S.E. Ritzmann, *Protein Abnormalities. Physiology of Immunoglobulins: Diagnostic and Clinical Aspects*, Alan Liss, New York, Vol. 1, 1982.
45. J.B. Henry, *Clinical Diagnosis and Management by Laboratory Methods*, 16th edition, WB Saunders, Philadelphia, 1979.
46. NCCLS, *Glossary and Guidelines for Immunodiagnostic Procedures, Reagents, and Reference Materials*, NCCLS Document DI1-A, NCCLS, Villanova, Vol. 12, 1992.

Statistical Quality Control in Clinical Laboratories

Jerry Dechert
Dechert and Moses, Inc., Riverton, USA

1 Introduction 1
2 Historical Development 1
3 Quality Assurance Versus Quality Control 2
4 Traditional Statistical Quality Control Approaches 2
4.1 Westgard's Multirule Procedure 3
4.2 Various Runs Rule Combinations 3
5 Recently Proposed Clinical Quality Control Approaches 3
5.1 Exponentially Weighted Moving Average Charts 3
5.2 Multivariate Approaches 4
6 Evaluation of Clinical Quality Control Performance 5
7 Clinical Significance 6
8 Economic Aspects 7
9 Conclusion 8
Abbreviations and Acronyms 8
Related Article 9
References 9

The importance of reporting quality patient results by clinical laboratories cannot be overstated. Maintaining and improving the quality of laboratory results is a never-ending process, and the thoughtful application of statistical quality control techniques is the key to this continuous pursuit. Using statistical methods, laboratories can ensure that their patient results are reliable and can continue to improve their performance.

It is clear that the method for monitoring quality control in the clinical laboratory will have a direct impact on the results reported by the laboratory. This article reviews the historical development of clinical quality control and details methods traditionally applied in the field. Recently proposed approaches are identified and their statistical performance is evaluated. Quality control methods involving clinical significance are also described, and economic modeling applicable to clinical quality control is presented.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

One of the most important aspects of diagnostic testing in a clinical laboratory is maintaining the quality of the results being reported by the laboratory. Since the reported results are used by physicians and pathologists to diagnose patient conditions and monitor patient progress, the quality of the results is of utmost importance. Therefore, quality control and assurance efforts regarding these results must be well planned and effective.

For many years, control materials have been tested and monitored to ensure the continued high quality of the patient results reported. These controls are stable materials of known (or assumed known) concentrations tested in the same fashion as patient samples, and they serve the purpose of evaluating testing system performance. The assumption is that changes detected by control materials will also impact on patient test results accordingly.

It is clear that the method for monitoring quality control in the clinical laboratory will have a direct impact on the results reported by the laboratory. This article reviews the historical development of clinical quality control and details methods traditionally applied in the field. Recently proposed approaches are identified and their statistical performance is evaluated. Quality control methods involving clinical significance are also described, and economic modeling applicable to clinical quality control is presented.

2 HISTORICAL DEVELOPMENT

The original application of statistical methods for monitoring quality control in the clinical laboratory dates back to the early 1950s.[1] In this original application, Levey and Jennings applied a Shewhart chart from industrial quality control and used average and range charts for subgroups of size two. Laboratory personnel today often refer to plots of quality control values as Levey–Jennings charts, but they are typically referring to charts of individual values plotted on a chart with limits set at ± 2 standard deviations (SD). This evolution from the original Levey–Jennings application of average charts to charts of individual values with ± 2 SD limits occurred gradually. Henry and Segalove's paper[2] discusses the plotting of single replicates against ± 2 SD limits, and the approach gained popularity in laboratories owing to its ease of use.

This evolution of quality control systems began in the 1950s and continues in many instances today. Typical approaches for monitoring controls in the clinical laboratory use three different levels of control materials (low, medium, and high) with targeted concentrations

across the measurement range of the diagnostic test. Each of these levels is tracked separately with control limits often placed at ±2 SD limits and applied to individual observations. Linnet's work[3,4] advocates averaging replicate observations of control values for monitoring quality control materials and shows the increased power of such an approach, but laboratory personnel have not embraced the practice.

In the late 1970s, Westgard et al.[5] published the first article assessing the statistical performance characteristics of clinical quality control methods. Using computer simulation, their work develops power curves for various combinations of runs rules. The graphs display the probability of detecting specific errors as a function of the number of control observations evaluated. The graphs are based on the probability of detecting a true error (p_{ed}) and the probability of a false rejection (p_{fr}). Additionally, the article discusses the concepts of random error (RE) and systematic error (SE). RE is the inherent imprecision, or noise, that a testing system will experience in a state of statistical control (SOSC). An SE is a change in the centering of the measurement system. Therefore, a change in the RE is a change in the measurement system variability, and an SE represents a change in the centering of the measurement system.

This work by Westgard et al. is extremely important as it set the direction for future evaluation of quality control approaches. The use of simulation became standard, and future models assumed a common shift to all levels used for monitoring. These assumptions, therefore, set the stage for work continuing into the 1990s.

Westgard et al.[6] used the developments of power curves to evaluate and propose a monitoring method combining a Shewhart control chart with a cumulative sum chart. While the approach performs well statistically, the sophistication of the method makes it unattractive to users and it has not been widely implemented.

Continuing the use of computer simulation to evaluate control methods comprised of combinations of control rules, Westgard and Groth[7] published what they termed "power function graphs". The graphs are refinements of earlier power curves, displaying probabilities of error detection and false rejection versus shifts in centering (ΔSE) and spread (ΔRE). These power function graphs provide a means for assessing a control monitoring system's performance and determining the statistical acceptability of a given approach.

3 QUALITY ASSURANCE VERSUS QUALITY CONTROL

A general distinction between quality assurance and quality control in the clinical laboratory must be made. As a rule, quality control refers to the internal assessment of control material testing. Alternatively, quality assurance often refers to larger scale activities and external assessments. Such external assessments are initiated by organizations such as the College of American Pathologists (CAP) or local regulatory bodies. In these quality assurance surveys, samples of unknown concentrations are tested at individual laboratories and their results reported to the governing body. Results across laboratories are compared to identify laboratories which are not producing results in agreement with the other laboratories participating in the survey. In contrast, quality control focuses on the internal assessment of control data and works towards continuous improvement of the testing system. It is clear that the results achieved through quality control will have a dramatic effect on the level of quality assurance evidenced by the laboratory through the independent survey assessments.

This article focuses on the issues and various approaches for applying quality control in the clinical laboratory. It is evident that a variety of factors must be considered when selecting the appropriate quality control approach for a given laboratory. A great deal of freedom is provided to laboratories in their selection since the Clinical Laboratory Improvement Amendments of 1988 (CLIA 88)[8] require that two levels of control material must be run each 24-h period. Within these general requirements, there are a variety of options available to clinicians for achieving quality control of their testing system. Determining the best approach for a given laboratory essentially consists of assessing the statistical performance, ease of use and implementation, and economic implications of the method selected. Given the proliferation of computer support available for analyzing quality control data, the options for clinicians continue to increase, and educated decisions regarding the selection of a quality control approach are definitely required.

4 TRADITIONAL STATISTICAL QUALITY CONTROL APPROACHES

A detailed review of the literature reveals that a large number of quality control approaches are available for application to clinical quality control. A few methods, however, have proven to stand the test of time and become the most common approaches in practice. These common methods include the strict application of ±2 SD limits (described in a previous section), the use of the Westgard Multirule Procedure (WR), or some subset of the WR. The following sections describe the WR and other sets of runs rules commonly encountered in practice.

4.1 Westgard's Multirule Procedure

Clinical Chemistry published the culmination of Westgard et al.'s work in the late 1970s and early 1980s as a ''selected method''.[9] Termed the ''Multirule Procedure'', Westgard et al. describe a combination of control rules to apply to monitoring control materials. The WR works as follows:

Reject if:

one point outside ±3 SD limits (1_{3s})

If one point outside ±2 SD limits, then consider:

(a) two consecutive points outside ±2 SD on the same side of the centerline (2_{2s});
(b) range of two points greater than 4 SD (R_{4s});
(c) four consecutive points outside ±1 SD limits on the same side of the centerline (4_{1s});
(d) 10 consecutive points above or below the mean (10_x).

As described in the original publication, all of the rules are applied both within a level of control material and across levels of control material. Suppose a laboratory is using two levels of control material (medium and high concentrations) for quality monitoring purposes. The 2_{2s} rule could be violated in two different ways. The first possible violation would be for both the medium and high control to plot outside ±2 SD limits on the same side of the mean on the same analytical run. The 2_{2s} rule would also be violated if the medium control plotted outside ±2 SD limits on the same side of the mean in two consecutive analytical runs. The result is that application of the WR as originally published results in a large number of comparisons which must be made in practice.

Although many clinicians endorse the procedure, there is still considerable confusion about its application in laboratories. The method can be difficult to implement manually as there are many comparisons required; many laboratories which claim to be using the WR are probably using some variation of the approach rather than the procedure as originally published.

Westgard et al. continued their work in the area of clinical quality control with the development of selection grids for planning quality control procedures.[10] The grids are tools which allow a user to select a set of control rules for a given application. The parameters for selecting the set of rules are the true frequency of actual errors in the laboratory and the critical systematic shift (ΔSE_c) that the user wishes to detect. Based on these parameters, the user can determine the number of replicates of controls to run and the control rules to employ.

4.2 Various Runs Rule Combinations

Following the appearance of the WR, there continued to be publication of other combinations of runs rules. Blum[11] published a method that incorporates the use of 10 different runs rules, justifying the selected rules using computer simulation. The method has not been widely implemented in the laboratory setting, probably a result of the complexity of applying the proposed runs rules.

A common practice of clinicians is to adjust the WR to fit their application. Certain runs rules in the procedure (primarily the 4_{1s} and 10_x rules) are sometimes ''turned off'' or dropped from the procedure because they tend to take longer to detect shifts in the process and often detect shifts that are relatively small in magnitude.

Another popular approach to monitoring quality control in the clinical laboratory is the use of ±2 SD limits verified by a retest. Since a number of laboratory personnel are aware that ±2 SD limits strictly applied will result in a high rate of false rejection, they choose to give the measurement system a ''second chance''. If the clinician tests controls and a replicate is outside ±2 SD limits, then the clinician reruns the controls and classifies the measurement system as unstable only if a replicate exceeds ±2 SD limits on the second run. If all control values are within ±2 SD limits on the second run, then the measurement system is considered stable.

5 RECENTLY PROPOSED CLINICAL QUALITY CONTROL APPROACHES

Recent developments in the field of clinical quality control have proposed new approaches to monitoring clinical quality control data. These methods are more sophisticated than some of the traditional methods which have been employed, but the availability of computer support for these new methods allows for their implementation in clinical laboratories. The option of implementing these more sophisticated methods on a real-time basis has not been available until the last few years.

5.1 Exponentially Weighted Moving Average Charts

In the spirit of improving the sensitivity of statistical quality control methods, Neubauer[12] recently proposed the use of the exponentially weighted moving average (EWMA) chart for application in clinical laboratories. The method has received attention for application in other industries[13,14] for some time and shown promising performance.

Application of the EWMA chart is relatively straightforward. A statistic is continually updated as new data are collected, and this statistic is plotted against control

limits. The statistic of interest, $EWMA_t$, is determined as follows in Equation (1):

$$EWMA_t = (1 - \lambda)EWMA_{t-1} + \lambda y_t \quad (1)$$

where t = time point, λ = weighting parameter ($0 < \lambda \leq 1$) and y_t = actual testing observation value. The parameter λ is used as a weighting factor to control how much emphasis is placed on recent data. The larger the value of λ, the more weight is placed on recent data. For example, setting λ to 1.0 would be the same as plotting the actual data themselves.

The selection of λ is a necessary decision when using the EWMA chart, and it is clear that its selection will impact the variability seen in the EWMA statistic. Larger values of λ will result in more variability, but the resulting control limits will also be more responsive to changes in the mean of the measurement process. Therefore, selection of λ is a tradeoff between reducing variability in the plotted statistic while at the same time adequately representing recent data in the statistic. Typical values of λ range between 0.05 and 0.25.

5.2 Multivariate Approaches

Another approach to statistical quality control in the clinical laboratory capitalizes on the inherent multivariate nature of clinical laboratory quality monitoring.[15] Multiple levels of control materials are typically tested at the same time when monitoring quality control. This results in a natural correlation among the control materials being monitored. Testing variation in clinical laboratories is generally largest from run to run with the sources of variability including environmental variability, technician variability, variation within the instrument, and materials variation. When the quality control levels are tested side by side, they are all affected by essentially the same run to run variation. This creates correlation among the levels. The correlation, however, is not perfect as there are some sources of variation that can affect individual levels differently. However, a positive correlation among control material testing data is fairly common.

While there are a number of multivariate approaches available for quality monitoring, the χ^2 chart is an attractive candidate owing to its relative ease of use. The χ^2 chart allows for the simultaneous control of p quality characteristics. For the clinical application, the number of characteristics, p, will correspond to the number of control levels being monitored. For the quality control application to multiple quality characteristics, one can test the hypothesis that $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0$ is a specified vector (the historical mean vector of the control level data). The critical region to test this hypothesis is $j(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) > \chi^2_p(\alpha)$, where j is the sample size (in the clinical application, $j = 1$), p is the number of characteristics being monitored ($p = 2$ or 3 control levels for the clinical case), α is the type I error associated with the chart, and $\bar{\mathbf{x}}$ is the observed sample mean vector.

The assumptions underlying the application of the χ^2 chart (besides that of multivariate normality) are that the specified mean vector ($\boldsymbol{\mu}_0$) and covariance matrix ($\boldsymbol{\Sigma}$) are known. While theoretically impractical, these assumptions may be applied in cases where a great deal of data has been collected surrounding the diagnostic measurement system.

In some instances, the assumption of a known mean vector and covariance matrix are not supportable, resulting in the requirement of another approach instead of the χ^2 chart. Such an approach is the T^2 chart, which assumes an unknown covariance structure and an unknown mean vector. Developed by Hotelling,[16,17] the T^2 statistic allows one to apply a multivariate control monitoring approach to p characteristics when the covariance matrix is unknown.

In all respects with the exception of the unknown covariance matrix and the mean vector, the T^2 chart is completely analogous to the χ^2 chart. As the number of subgroups used in establishing the T^2 chart limits increases to infinity, the T^2 chart will approach the χ^2 chart. Therefore, the question arises as to how many subgroups need to be collected to provide reliable T^2 chart limits. Additionally, how many subgroups are necessary before the T^2 chart can be effectively replaced with the χ^2 chart?

Similarly to the χ^2 chart, Anderson[18] showed that the T^2 statistic is defined according to Equation (2):

$$T^2 = m(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \quad (2)$$

where $\mathbf{S}$ is the sample covariance matrix, m is the number of subgroups of size j collected, $\bar{\mathbf{x}}$ is the mean vector of the sample with dimension p, and $\boldsymbol{\mu}_0$ is the historical mean vector. The upper limit for the T^2 chart is calculated as shown in Equation (3):

$$\text{upper limit} = \frac{(m-1)p}{m-p}F_{p,m-p,\alpha} \quad (3)$$

The effect of the number of subgroups on the control limits of the T^2 chart can easily be evaluated. Setting the α level for the T^2 chart at 0.01, Figure 1 shows the impact of the number of subgroups collected on the performance of the T^2 chart when a 2.0 SD shift occurs across all three levels of control material. Three different correlation coefficients (0.80, 0.50, and 0.10) among the three levels of controls are evaluated. This graphic indicates that there is diminishing return in increasing the number of subgroups collected for establishing the T^2 chart limits past 20 and that the correlation structure has little effect on the selection of the number of subgroups collected. The

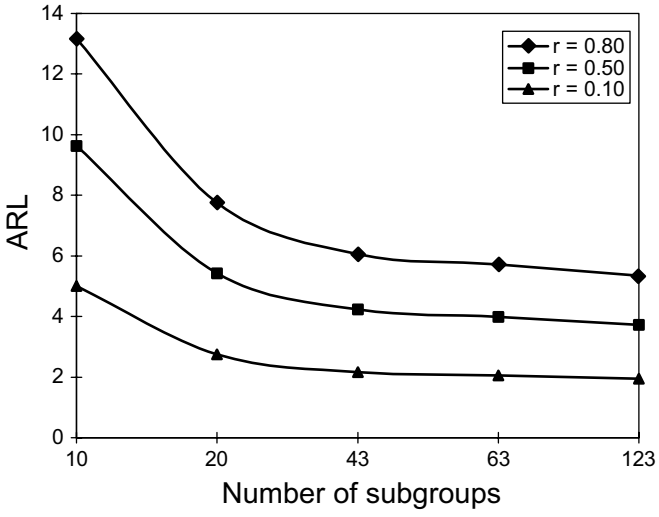

Figure 1 Number of subgroups comparison for the T^2 chart.

graph also illustrates that the χ^2 chart can be effectively substituted for the T^2 chart with little loss of performance following a relatively small number of subgroups.

6 EVALUATION OF CLINICAL QUALITY CONTROL PERFORMANCE

Given the proliferation of approaches available for clinical quality control, it is necessary to select among the options based on objective criteria. Of primary importance is the statistical performance of the methods. Based on their abilities to avoid false rejections and detect true changes of importance in the testing system, the appropriate method can be selected for application. One of the difficult aspects of selecting the quality control method is that the two keys to statistical performance (i.e. avoiding false rejections and detecting true system changes) are difficult to satisfy simultaneously. Typically, methods which are very sensitive to system changes also have high false rejection rates. Conversely, methods with low false rejection rates may often be insensitive to changes in the measurement system. Evaluation of the statistical performance of proposed methods is of the utmost importance.

In order to evaluate the statistical performance of quality control approaches, researchers often use computer simulation models. There has been a great deal published in the clinical literature concerning the computer simulations used for this work, and the editorial by Westgard[19] provides an excellent summary of the assumptions employed in the simulators and lists many of these simulations. Hatjimihail[20] and Parvin[21,22] provide examples of recent simulation efforts. The literature also contains other approaches including neural networks[23] and genetic algorithms.[24]

Analytical approaches to analyzing control methods utilizing runs rules have been limited in the clinical laboratory literature until recently. Parvin[25] performed some analytical analysis by looking at data within a run, but his results are not widely applicable. Bishop and Nix[26] were the first authors in the clinical chemistry literature to obtain analytical results by applying Markov chains to procedures incorporating runs rules. Using Markov modeling to analyze the Westgard Rules as previously published, their work compares their results with previous computer simulation analysis. Bishop and Nix also proposed the use of a cumulative sum chart for monitoring control levels and provide results supporting its utility.

Lee[27] has also done some analytical work regarding the use of the WR. The sample sizes investigated by Lee were indicative of those actually implemented in practice, and his results provide good insight into the workings of modified approaches to the WR (i.e. not including the R_{4s} rule or using ± 2 SD limits as a screening criterion).

The methods discussed in this article can be evaluated and compared on a statistical basis. These methods include the strict application of ± 2 SD limits (2 SD), the use of ± 2 standard deviation limits with a retest (SD/RT), the original description of the WR, the χ^2 chart, and the EWMA chart.

Comparison of the quality control approaches is achieved through the average run length (ARL) of the approaches. The ARL is the expected number of test points before a quality control approach will signal. In other words, the number of times control materials must be evaluated (on average) before the quality control method indicates that there has been a change in measurement system performance. When evaluating methods that incorporate runs rules, it is important to evaluate their performance based on ARLs. For instance, adding the 4_{1s} rule does not affect the probability of detecting a shift in the first quality control sample taken after the shift, but it does reduce the number of sampling points required to detect the shift in the long run. For methods with a readily defined probability of detection, the ARL is simply the inverse of the probability of detection.

Using the ARL concept, the previously described quality control approaches can be readily evaluated from a statistical perspective. For the 2 SD, 2 SD/RT, and χ^2 chart approaches, probabilities of detection can readily be calculated analytically. Simulation results are used for assessing both the WR and the EWMA chart. For the χ^2 and EWMA charts, certain assumptions or parameter settings are necessary. This comparison sets the in-control ARL for both methods at 100 (i.e. a 1% chance of a false

rejection). Additionally, the covariance structure for the χ^2 chart must be specified, and it is fixed at a correlation of 0.50 (i.e. moderate correlation among levels). For the EWMA chart, a weighting factor, λ, of 0.20 is assumed.

For comparison purposes, it is assumed that two levels of control material are being monitored and that the same relative magnitude of shift is applied to both levels. ARLs are plotted as a function of shifts in the mean of the measurement process where these shifts are represented in multiples of the process SD (i.e. a 0.8 SD shift, etc.). The ARLs for the methods are shown in Figure 2.

From an examination of Figure 2, a number of conclusions are available. The first is the unacceptably low in-control ARL for the 2 SD approach. Laboratories employing this approach will incur a large number of false rejections, potentially leading to distrust of the quality control approach. This false rejection issue is alleviated through the use of a retest, and the long in-control ARL for 2 SD/RT is evidenced in the figure. The WR succeeds in reducing the false rejection rate incurred by the strict use of 2 SD limits and is more sensitive to shifts than the use of 2 SD with a retest.

Performance of the recently proposed methods of clinical quality control can also be evaluated using Figure 2. The EWMA chart has an attractive performance owing to the relatively long in-control ARL, but is also sensitive to true changes in the measurement system. Additionally, the χ^2 chart shows a good in-control ARL, but relative insensitivity to true shifts in the measurement process. As indicated by Dechert and Case,[15] the sensitivity of the multivariate approach is affected by the nature of the shift incurred, thus resulting in the relative insensitivity for the constant shift assumed here across both levels of control materials.

Finally, Figure 2 indicates that large shifts (i.e. shifts greater than 1.5 SD) are essentially detected equivalently by all of the methods considered. Hence if a laboratory is only concerned with detecting large shifts in the measurement process, any of the methods would be acceptable from an error detection perspective. Consideration of the associated false rejection rate would also be required.

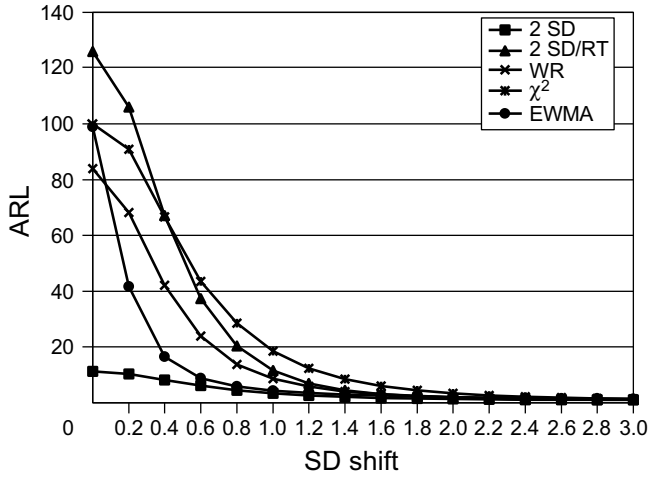

Figure 2 Statistical comparison of quality control approaches.

7 CLINICAL SIGNIFICANCE

The quality control approaches described in the previous sections are all directed towards detecting any amount of true shift in the measurement process. However, as the imprecision of analytical devices continues to decrease through technological innovation, the use of such methods can become problematic. The evaluation of the quality control approaches previously presented evaluated the ARLs as functions of shifts in the mean in terms of process SD. This is an accepted method of evaluating the statistical performance of quality control approaches. However, as assay variability is continually decreased, the multiples of SD shifts can take on new relative meanings. For example, consider an assay with a 5% coefficient of variation (CV). A 1.0 SD shift would constitute a 5% shift in the mean of the process. If the CV of that assay is reduced to 2.5%, then a 1.0 SD shift would represent a 2.5% shift in the mean of the process. Eventually, the clinical utility of the assay must be considered in the selection and establishment of quality control procedures.

Towards this end, Westgard et al.'s selection grids[10] involve the use of ΔSE_c (i.e. the critical SE to detect) as a parameter for selecting a method. This signifies the appearance of methods designed to detect some specific change identified to be clinically significant or relevant. In other words, changes less than this amount (ΔSE_c) are considered to be of no consequence in the clinical setting while changes this large or larger can have clinical implications. These implications may involve changing a patient's dosage or initiating a change in treatment.

Attempts to define the requirements of clinical laboratory testing systems go back to Skendzel et al.[28] Skendzel et al. mailed a questionnaire to physicians to determine the total precision required for a variety of diagnostic tests. Results at that time indicated that almost all the tests considered provided adequate precision for physicians' requirements. The work was also useful for establishing medically useful guidelines for analytical precision.

Linnet[29] then used Skendzel et al.'s work to calculate what he called "maximum clinically allowable analytical error". Represented by ΔSE_c, this error represents the largest error that can be tolerated according to the requirements outlined through Skendzel et al.'s survey results. Linnet's approach for calculating ΔSE_c is shown

in Equation (4):

$$\Delta SE_c = \Delta_{med} - 1.65 s_t \quad (4)$$

where Δ_{med} = the median difference of medical importance as reported by the physicians in Skendzel et al.'s work and s_t = total precision including analytical measuring system, sample handling, and patient biological variability. Linnet's paper includes a similar approach for defining a critical change in the spread of the measurement system, ΔRE_c. It is clear that the parameters required to calculate the maximum clinically allowable error are not easy to estimate and can be sources of debate.

Koch et al.[30] described an application that incorporates the use of the maximum clinically allowable error in the design of the quality control system for a specific analytical system. Their application uses a consensus process at the testing site to determine the total allowable error (TE_a). Then, they calculate ΔSE_c and ΔRE_c using Equations (5) and (6):

$$\Delta SE_c = \frac{TE_a - |bias|}{s} - 1.65 \quad (5)$$

$$\Delta RE_c = \frac{TE_a - |bias|}{1.96 s} \quad (6)$$

where TE_a = total allowable error for the testing method, bias = known difference between the laboratory's mean and the true mean of the control material, and s = analytical testing system precision. Control methods (number of control replicates and control rules) can then be selected for the individual tests with the requirement of a 90% probability of detecting ΔSE_c.

Many other methods are available for determining the maximum clinical allowable error, as noted by Fraser et al.[31] and Westgard and Burnett.[32] Also, Petersen and Fraser[33] provide an excellent editorial discussing the issues involved with determining this allowable error. Westgard et al.[34] further developed their approach for defining maximum clinically allowable errors using criteria from CLIA 88. This federal regulation outlines TE_a for many analytical tests. Westgard et al. recommend using the total error indicated from CLIA 88 according to Equation (7):

$$TE_a = bias_{meas} + \Delta SE_c s_{meas} + z \Delta RE_c s_{meas} \quad (7)$$

where $bias_{meas}$ = bias in the measurement system, s_{meas} = total precision of the measurement system, and z = standard normal value that sets the maximum percentage beyond TE_a. Using the TE_a associated with the test being analyzed, the user can calculate ΔSE_c and ΔRE_c and select an appropriate quality control scheme.

Although there does not yet appear to be any general consensus in the literature regarding the best approach for defining changes which are clinically significant, it is clear that quality control approaches in the clinical setting are moving towards some amount of allowable error. As the variability inherent in these diagnostic tests continues to become smaller and smaller, there will be stronger movement in the clinical field away from methods which define acceptability limits solely on assay variation. Traditionally accepted methods (e.g. ±2 SD limits) will be replaced with methods incorporating some allowable error or having a total error specification.

8 ECONOMIC ASPECTS

The traditional emphasis for statistical quality control approaches applied to clinical laboratories has been with respect to the statistical performance of the proposed methods. Although the statistical performance of the methods is of utmost importance, the economic aspects of the selected methods must not be ignored. Often, multiple approaches can result in the same statistical power for detecting changes in the measurement process, but the cost associated with the approaches can be drastically different.

Although the application of economic models in the clinical setting is limited, Westgard et al.[35,36] have done some work in the area. In both studies, the authors used a predictive value model to assign costs to various approaches for monitoring quality control. Basically, their model develops the following four situations:

1. the quality control method detects a true change;
2. the quality control method detects a change when there is no change;
3. the quality control method does not detect a true change;
4. the quality control method does not detect a change when there is no change.

Situations 1 and 4 correspond to a correctly working method whereas situation 2 corresponds to a type I error and situation 3 is a type II error. The authors develop probabilities for being in each of the four situations and the costs associated with these situations, comparing methods based on these costs. They can then use their model to predict "quality" (essentially the defect rate) and "productivity" (the test yield).

Recent work with respect to economic modeling of clinical statistical quality control has focused on developing cost equations related to the natural cycle encountered in clinical quality control.[37] The cycle here refers to the testing process beginning in a SOSC (i.e. on target and stable). After operating in control for a period of time, a shift or special cause of variation affects

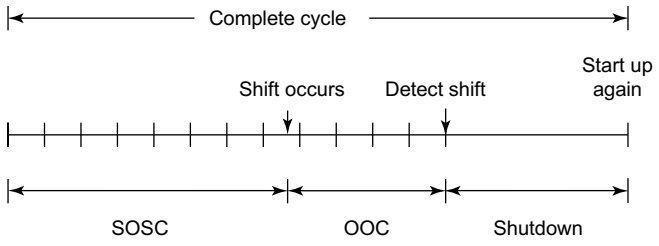

Figure 3 Illustration of a complete cycle for the economic model.

the testing process driving the state of the process out of control. Eventually, the quality control monitoring approach signals that a shift has occurred and the testing process is shutdown until the cause of the shift is identified. A complete cycle is shown in Figure 3.

Using this model, the testing process is in one of the following states: operating under control (SOSC), operating out of control (OOC), or shutdown. Costs with being in each of these associated states can be specified, resulting in an overall cost equation. The cost of SOSC is essentially a function of testing volume. It includes costs for labor, testing consumables, equipment depreciation, overheads, etc. It can consider the cost of doing business when things are performing as expected. The cost of OOC includes the basic costs related to SOSC, but there is an additional cost component attached. This additional cost relates to reporting patient results which are actually shifted from their true mean. Failing to detect large shifts can result in governing bodies shutting the laboratory down, so this additional cost is very large for major shifts in the measurement process. Finally, there is a cost associated with not generating results when they are needed. This cost reflects responsiveness of the laboratory and may be extremely high for situations when a physician needs rapid turnaround of patient results, but the laboratory is down or not operating. This cost will include basic overhead costs, additional costs for having idle resources on hand, and may include overtime operations for making up lost throughput.

The economic models allows for the incorporation of clinically significant errors. Various magnitudes of process shifts are used as inputs to the model, and the cost associated with these specified shifts can be used to reflect the clinical relevance of the shifts.

9 CONCLUSION

It is clear that a variety of challenges remain with respect to statistical quality control in the clinical laboratory. While traditional methods have been instrumental in ensuring high-quality patient results historically, new needs and requirements for clinical quality control must be addressed. The ever-increasing capabilities of computer technology with respect to summarizing quality control data offer new opportunities. Methods that had been previously infeasible owing to difficulties in implementation can now capitalize on the computer technology available. Training in the application of the outputs of such computer models will definitely be necessary, but complex methods can be applied in a fashion that is transparent to the end user.

Additionally, the consideration of what truly constitutes a clinically relevant change in the measurement process must be integrated into future statistical quality control approaches for laboratories. As instrumentation and assay technology improve, the inherent imprecision of diagnostic assays will continue to decrease. As a result, changes in system performance which are statistically identifiable, yet not clinically important, will become problematic. While there are recommendations as to the level of variability which is clinically tolerable available (CLIA 88), these recommendations are not exhaustive and a general consensus as to the magnitude of allowable errors has not been achieved.

The economic aspect of statistical quality control monitoring for clinical laboratories must also be addressed in the future. While the clinical performance of the laboratory is the first priority, economic aspects of selected methods must not be ignored.

All of these issues constitute future challenges facing statistical quality control in clinical laboratories. Through continued research and transfer of results into the clinical laboratories, the quality of reported patient results will continue to improve in the future.

ABBREVIATIONS AND ACRONYMS

ARL	Average Run Length
CAP	College of American Pathologists
CLIA 88	Clinical Laboratory Improvement Amendments of 1988
CV	Coefficient of Variation
EWMA	Exponentially Weighted Moving Average
OOC	Operating Out of Control
RE	Random Error
SD	Standard Deviation
SD/RT	Standard Deviation Limits with a Retest
SE	Systematic Error
SOSC	State of Statistical Control
TE_a	Total Allowable Error
WR	Westgard Multirule Procedure

RELATED ARTICLE

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. S. Levey, E.R. Jennings, 'The Use of Control Charts in the Clinical Laboratory', *Am. J. Clin. Pathol.*, **20**, 1059–1066 (1950).
2. R.J. Henry, M. Segalove, 'The Running of Standards in Clinical Chemistry and the Use of the Control Chart', *J. Clin. Pathol.*, **5**, 305–311 (1952).
3. K. Linnet, 'The Between-run Component of Variation in Internal Quality Control', *Clin. Chem.*, **35**, 1416–1422 (1989).
4. K. Linnet, 'Mean and Variance Rules are More Powerful or Selective Than Quality Control Rules Based on Individual Values', *Eur. J. Clin. Chem. Clin. Biochem.*, **29**, 417–424 (1991).
5. J.O. Westgard, T. Groth, T. Aronsson, H. Falk, C.H. de Verdier, 'Performance Characteristics of Rules for Internal Quality Control: Probabilities for False Rejection and Error Detection', *Clin. Chem.*, **23**, 1857–1867 (1977).
6. J.O. Westgard, T. Groth, T. Aronsson, C.H. de Verdier, 'Combined Shewhart–Cusum Control Chart for Improved Quality Control in Clinical Chemistry', *Clin. Chem.*, **23**, 1881–1887 (1977).
7. J.O. Westgard, T. Groth, 'Power Functions for Statistical Control Rules', *Clin. Chem.*, **25**, 863–869 (1979).
8. United States Department of Health and Human Services, 'Medicare, Medicaid and CLIA Programs: Regulations Implementing the Clinical Laboratories Improvement Amendments of 1988 (CLIA). Final Rule', *Fed. Regist.*, **57**, 7002–7186 (1992).
9. J.O. Westgard, P.L. Barry, M.R. Hunt, 'A Multi-rule Shewhart Chart for Quality Control in Clinical Chemistry', *Clin. Chem.*, **27**, 493–501 (1981).
10. J.O. Westgard, E.F. Quam, P.L. Barry, 'Selection Grids for Planning Quality Control Procedures', *Clin. Labo. Sci.*, **3**, 271–278 (1990).
11. A.S. Blum, 'Computer Evaluation of Statistical Procedures and a New Quality-control Statistical Procedure', *Clin. Chem.*, **31**, 206–212 (1985).
12. A.S. Neubauer, 'The EWMA Control Chart: Properties and Comparison with Other Quality-control Procedures by Computer Simulation', *Clin. Chem.*, **43**, 594–601 (1997).
13. S.V. Crowder, 'Design of Exponentially Weighted Moving Average Schemes', *J. Qual. Technol.*, **21**(3), 155–162 (1989).
14. J.S. Hunter, 'The Exponentially Weighted Moving Average', *J. Qual. Technol.*, **18**, 203–210 (1986).
15. J. Dechert, K.E. Case, 'Multivariate Approach to Quality Control in Clinical Chemistry', *Clin. Chem.*, **44**, 1959–1963 (1998).
16. H. Hotelling, 'The Generalization of Student's Ratio', *Ann. Math. Stat.*, **2**, 360–378 (1931).
17. H. Hotelling, 'Multivariate Quality Control', in *Techniques of Statistical Analysis*, eds. C. Eisenhart, M. Hastay, W.A. Wallis, McGraw-Hill, New York, 111–184, 1947.
18. T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 1958.
19. J.O. Westgard, 'Simulation and Modeling for Optimizing Quality Control and Improving Analytical Quality Assurance', *Clin. Chem.*, **38**, 175–178 (1992).
20. A.T. Hatjimihail, 'A Tool for the Design and Evaluation of Alternative Quality-control Procedures', *Clin. Chem.*, **38**, 204–210 (1992).
21. C.A. Parvin, 'Estimating the Performance Characteristics of Quality-control Procedures when Error Persists Until Detection', *Clin. Chem.*, **37**, 1720–1724 (1991).
22. C.A. Parvin, 'Comparing the Power of Quality-control Rules to Detect Persistent Systematic Error', *Clin. Chem.*, **38**, 358–363 (1992).
23. C.R. Schweiger, G. Soeregi, S. Spitzauer, G. Maenner, A.L. Pohl, 'Evaluation of Laboratory Data by Conventional Statistics and by Three Types of Neural Networks', *Clin. Chem.*, **39**, 1966–1971 (1993).
24. A.T. Hatjimihail, 'Genetic Algorithms-based Design and Optimization of Statistical Quality-control Procedures', *Clin. Chem.*, **39**, 1972–1978 (1993).
25. C.A. Parvin, 'New Insight into the Comparative Power of Quality-control Rules That Use Control Observations Within a Single Analytical Run', *Clin. Chem.*, **39**, 440–447 (1993).
26. J. Bishop, A.B.J. Nix, 'Comparison of Quality-control Rules Used in Clinical Chemistry Laboratories', *Clin. Chem.*, **39**, 1638–1649 (1993).
27. W.K. Lee, 'Performance Characteristics of Multi-rule Shewhart Control Charts Used in Clinical Chemistry Laboratories', Master's Report, Oklahoma State University, Stillwater, OK, 1996.
28. L.P. Skendzel, R.N. Barnett, R. Platt, 'Medically Useful Criteria for Analytical Performance of Laboratory Tests', *Am. J. Clin. Pathol.*, **83**, 200–205 (1985).
29. K. Linnet, 'Choosing Quality-control Systems to Detect Maximum Clinically Allowable Analytical Errors', *Clin. Chem.*, **35**, 284–288 (1989).
30. D.D. Koch, J.J. Oryall, E.F. Quam, D.H. Feldbruegge, D.E. Dowd, P.L. Barry, J.O. Westgard, 'Selection of Medically Useful Quality-control Procedures for Individual Tests Done in a Multi-test Analytical System', *Clin. Chem.*, **36**, 230–233 (1990).
31. C.G. Fraser, H. Petersen, M.L. Larsen, 'Setting Analytical Goals for Random Analytical Error in Specific Clinical Monitoring Situations', *Clin. Chem.*, **36**, 1625–1628 (1990).

32. J.O. Westgard, R.W. Burnett, 'Precision Requirements for Cost-effective Operation of Analytical Processes', *Clin. Chem.*, **36**, 1629–1632 (1990).
33. P.H. Petersen, C.G. Fraser, 'Setting Quality Standards in Clinical Chemistry: Can Competing Based on Analytical, Biological, and Clinical Outcomes be Harmonized?', *Clin. Chem.*, **40**, 1865–1868 (1994).
34. J.O. Westgard, J.J. Seehafer, P.L. Barry, 'Allowable Imprecision for Laboratory Tests Based on Clinical and Analytical Test Outcome Criteria', *Clin. Chem.*, **40**, 1909–1914 (1994).
35. J.O. Westgard, T. Groth, 'A Predictive Value Model for Quality Control: Effects of the Prevalence of Errors on Performance of Control Procedures', *Am. J. Clin. Pathol.*, **80**, 49–56 (1983).
36. J.O. Westgard, J.J. Oryall, D.D. Koch, 'Predicting Effects of QC Practices on the Cost-effective Operation of a Multi-test Analytical System', *Clin. Chem.*, **36**, 1760–1764 (1990).
37. J.D. Dechert, K.E. Case, 'An Economic Model for Clinical Quality Control', *J. Appl. Stat.*, **26**, 553–562 (1999).

Supercritical Fluid Chromatography in Clinical Chemistry

Steven W. Graves
Brigham Young University, Provo, USA

1 Introduction 1
2 Fundamental Characteristics of Supercritical Fluid Chromatography 1
3 Analytical Considerations 2
3.1 Instrumentation 2
3.2 Sample Selection 2
3.3 Sample Introduction 2
3.4 Columns 3
3.5 Mobile Phase 4
3.6 Elution 4
3.7 Detection 4
4 Applications in Clinical Chemistry 5
4.1 Fundamental Advantages 5
4.2 Limitations 6
4.3 Use for Extraction 6
4.4 Previous Applications 6
4.5 Comparisons with Other Methods 10
5 Future Developments 10
Abbreviations and Acronyms 11
Related Articles 11
References 11

Supercritical fluid chromatography (SFC) employs a gas taken above its critical temperature and critical pressure to where it forms a supercritical fluid (SF). SFs have properties of both gases and liquids. They have lower densities and viscosities and higher diffusion coefficients than liquids, but they also have the solvating power of a liquid and can dissolve nonpolar and many polar compounds. As such, these favorable properties combine to make SFs an excellent mobile phase for column chromatography. The SFC technique is highly versatile in that it can be used with either packed columns (similar to high-performance liquid chromatography (HPLC) columns with the same variety of stationary phases available) or open-tubular columns (similar to gas chromatography (GC) columns and employing the same types of coatings) which can be interfaced with a very wide array of detectors. In addition, SFC avoids high temperatures during the chromatographic separation and in most cases the need to derivatize analyzed compounds. These features make SFC particularly well suited to the separation and assay of clinically important biomolecules. As SFC instruments have become commercially available, numerous applications have been explored in the areas of drug monitoring, toxicology, environmental contaminants, and the measurement of physiological biomolecules. Despite its potential, SFC enjoys limited use in clinical chemistry laboratories currently. Greater efforts are needed to improve sensitivity, standardize and validate assays, and work toward high-volume throughput capabilities.

1 INTRODUCTION

The principles underlying SFs have been understood for decades. In 1822 de la Tour[1] reported the disappearance of the meniscus between gas and liquid in a closed system at its critical temperature and in 1879 Hannay and Hogarth[2] demonstrated that salts could be dissolved in condensed gases. However, the use of SFs as the mobile phase of chromatographic separations is much more recent.[3] Moreover, SFC instruments and SFC columns have only been commercially available since about the mid-1980s. This highly versatile technique frequently provides separation efficiencies approaching those of GC, and eliminates the need to work at high temperatures and to derivatize analytes. Hence, it is especially well suited for chemically reactive, heat-sensitive, nonvolatile or higher-molecular-weight compounds. Despite these advantages SFC is still rarely used in the clinical laboratory. Partly this may be related to the existence of well-established and adequate methods for many routinely assayed compounds, to a lack of SFC method development that is specific to clinical applications, to the reality that most SFC instruments are designed for the research laboratory and do not accommodate high-volume throughput, and finally to the limited experience of most clinical chemists with this technology.

2 FUNDAMENTAL CHARACTERISTICS OF SUPERCRITICAL FLUID CHROMATOGRAPHY

A substance becomes a SF when its temperature and pressure exceed their critical points. In this condition, the substance has physical properties that are intermediate between its gaseous and liquid states. For example, SFs have lower densities, lower viscosities, and higher diffusion coefficients than liquids, while still maintaining many of the favorable solvating properties of liquids

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

not present in gases. SFs persist without undergoing the gas-to-liquid transition, despite increases in the external pressure applied to them.

Because of these unique properties, SFs have found many useful applications as mobile phases for column chromatographic separations. The theoretical considerations and background of SFC have been considered in a number of publications.[4–7] As mentioned above, SFC offers the combination of advantages normally associated with GC or LC (liquid chromatography) alone. Because of higher diffusion coefficients, SFs allow for greater radial mass transfer than in the liquid state and, consequently, more efficient separations are achieved than in LC. This affords a shorter analysis time and/or the ability to resolve complex mixtures of analytes. Because of the solvation characteristics of SFs, compounds for analysis need not be volatile as required in GC, but need only be soluble in the SF to allow the technique to be applied.

3 ANALYTICAL CONSIDERATIONS

3.1 Instrumentation

Currently, there are still only a few commercial suppliers of SFC systems.[8] As of this writing, the major vendors of SFC systems are:

- Berger Instruments, Inc., Delaware, Maryland, USA, Tel. 1-(302)266-8201. Their systems are modular and support packed column applications. Components include an autosampler, pumps with pressure regulators, an oven, and an ultraviolet/visible (UV/VIS) absorption detector. Their systems can be interfaced with other instrumentation, including flame ionization detection (FID), nitrogen–phosphorus thermionic ionization detection (NPD), electron capture detection (ECD) and mass spectrometry (MS). Complete systems range in price from about US$45 000 to US$75 000.
- Gilson, Inc., Middleton, Wisconsin, USA, Tel. 1-(608)836-1551, www.gilson.com. Gilson provides complete systems that employ packed column technologies and are based on LC instrumentation. The Gilson system has an external computer unit that operates the pumps and allows for programmable gradient elution capabilities. The detector included with the standard system is a UV absorption detector fitted with a special flow cell. Although other detectors are not provided, different detectors including FID and MS can be used. The cost of a complete system is about US$50 000 currently. They also market packed SFC columns.
- JASCO, Easton, Maryland, USA, Tel. 1-(800)33-5272, www.jascoinc.com; Dunmow, Essex, UK, Tel. (1371)876988, www.jasco.co.jp; Hachioji City, Tokyo, Japan, Tel. (426)66-0682. They provide customized systems and market all of the relevant components. Their systems support packed column SFC applications. The components include extraction cells, pumps, back pressure regulators, ovens, cooling jackets, and a variety of detectors, including fluorescence, diode array, UV absorption, circular dichroism and refractive index. Their systems can also perform HPLC. The list price for a complete system begins at about US$40 000.
- Selerity Technologies, Salt Lake City, Utah, USA, Tel. 1-(801)936-0564. This company produces complete SFC systems that can employ either open tubular or packed microbore SFC columns. Injection is accomplished by means of a computer-controlled valving system; the system has programmable pressure gradient capability. The standard system comes equipped with FID, but can be interfaced with ECD, UV absorption, NPD, or MS. The current retail price is about US$36 000. Both packed and open tubular columns are available.

3.2 Sample Selection

SFs that are compatible with commercial SFC systems, i.e. that are not corrosive or requiring excessive pressures or temperatures, are most easily used for the analysis of nonpolar or weakly polar compounds. Most polar compounds can also be separated on SFC by modifying the mobile phase, which is described hereafter, or after analyte derivatization. Analytes can range in molecular weight from very small molecules up to polymers having molecular weights in excess of 10 000.[9] Many compounds that cannot be separated by GC because of their low volatility, reactivity, or chemical instability are handled with ease by SFC. Hence, many biological analytes are amenable to SFC separation. The technique has recently been applied to the resolution of chiral compounds with excellent success,[10] and this has led to its having extensive application in the manufacture and quality control of drug production. Like most chromatographic separations used in clinical laboratory analysis, appropriate internal standards are required. As with other chromatographic separations, the analyte of interest must typically be extracted from its biological matrix before it can be introduced into an SFC system.

3.3 Sample Introduction

Sample introduction on commercial systems is usually accomplished by means of a high-pressure valving

system. Consequently, the materials to be analyzed are most often dissolved in a small volume of an organic solvent. The sample port typically comprises a loop or an extraction cell in which the sample is introduced. For open tubular columns, loop volumes are of the order of 0.2–1.0 μL of which 10–100 nL are moved onto the analytical column by a time-plotted technique and carried through the analytical column to the detector. Other commercial injection systems support packed column SFC and typically accommodate larger injection volumes of up to a few microliters. Several approaches have been developed to increase the sample volume or, more specifically, to increase the amount of analyte introduced to improve sensitivity. These techniques include solvent venting,[11] which has allowed in some cases sample volumes of 100 μL or more to be introduced, pressure focusing with solvent venting,[11,12] solid-phase extraction,[13–15] or use of a retention gap.[16] This last method, unlike some of the former ones mentioned, is a modification that can be readily applied to commercial systems. Other simple approaches have been employed that involved the use of a reversed-phase HPLC guard column (or precolumn) for off-line solid-phase extraction (our unpublished results). In this approach the analyte in aqueous solution was passed through an octadecylsilanol column and the analyte was retained. Thereafter all of the aqueous solvent was flushed out of the guard column prior to it being placed in series with the SFC column. The analyte (which in this case was a steroid) was eluted from the guard column efficiently by SF carbon dioxide and delivered to an open tubular SFC column for chromatography and detection. Any of these modifications not yet commercially available require additional time for development and validation and a measure of expertise in working with instrumentation. In many cases commercial instruments are not easily adapted.

3.4 Columns

One of the favorable features of SFC systems is that they can employ a wide variety of columns and stationary phases. These include both ''packed'' and ''open tubular'' column types. Packed columns (typically 1–5 mm in diameter) contain spherical silica particles (3–5 μm) to which various organic phases have been covalently bound through the silanol groups of the solid support. The chemical nature of the organic phase provides the separation characteristics. These columns are analogous to HPLC columns and most commercial SFC systems accommodate the use of standard, unmodified HPLC columns. Open tubular columns are typically microcapillary columns made of fused silica which have been coated with nonpolar materials that possess some polarizable capability. The stationary phases are commonly the same materials used to coat GC columns. Typically the stationary phase is covalently bonded to the capillary surface and cross-linked to prevent the solid phase from bleeding from the column. These columns are very similar in nature to GC columns except they have a smaller inner diameter. This is required, both on a theoretical basis and in practice, to achieve the desired column efficiency.[17,18] The theoretical considerations for column selection have been reviewed in greater detail elsewhere.[19,20] Briefly, commercially available, packed columns are short (5–30 cm in length) and have a wider inner diameter (ID) of 1.0–5.0 mm compared to open-tubular columns which are longer with smaller IDs (1.0–30.0 m in length, 25–100 μm ID). Packed columns typically have fewer theoretical plates per meter, although more than 200 000 theoretical plates have been achieved by placing packed columns in series.[21] Even with the use of a single conventional 25 cm packed column, efficiencies of over 20 000 plates are common and would be comparable to a 10 m open tubular column when it used at 20 times its optimal velocity, which is a frequent practice. Recent work on columns has primarily focused on the stationary phase of packed columns. Many new packings have emerged, including those that allow for chiral separations.[9,10] Despite these new packings, there are still concerns about pressure drops over the length of the packed column, which can greatly compromise a column's efficiency, a reminder that longer may not always be better.[9] Despite these concerns, careful determination of the solid phase, temperature, modifier, modifier gradient and other additives can achieve excellent selectivity irrespective of the number of theoretical plates. Packed columns also have higher sample capacities and are well suited for relatively simple mixtures, preparative work, or very dilute analytes. Use of packed column SFC may limit the detection method or at least require specialized approaches to remove or reduce organic modifiers for certain detection modalities. Because of the high mobile-phase flow rates (1–20 mL min^{-1}) used with packed column SFC, it is not possible to use flame-based detectors such as FID. However, most other detection methods are compatible with packed column SFC including MS which is routinely done with packed columns at flow rates of 1 mL min^{-1} or greater without splitting the eluate.

Open tubular columns typically have more theoretical plates per meter and are often better suited for the resolution of complex mixtures. In general these separations are slower, and smaller amounts of specimen can be loaded. As noted, open tubular columns are typically coated with the same types of materials used in GC columns. Most frequently, these are polysiloxanes which

are cross-linked to provide column stability. Some of the side-chain constituents of these polymers include alkyl, aromatic, cyanopropyl, and ethylene oxide groups, again allowing for selective separations. Open tubular column SFC applications typically omit the organic modifier and have much lower flow rates ($1-10\,\mu L\,min^{-1}$) which allow them to be interfaced with all types of detectors, including flame-based detectors.

3.5 Mobile Phase

There are only a few gases that have physicochemical properties that allow them to form SFs under conditions that are compatible with the physical limitations (e.g. heat tolerance, pressure tolerance) of commercial instruments. These gases all have relatively low critical temperatures and relatively low critical pressures, and are also noncorrosive as SFs. However, in a clinical laboratory there are added constraints, including toxicity, flammability, and cost, and hence only supercritical carbon dioxide is used. As a SF, carbon dioxide has a somewhat intermediate solubility parameter. Consequently, there are a number of compounds of analytical interest that are not directly soluble in SF carbon dioxide because of their high polarities. In most cases these analytes can be made soluble by the addition of small percentage of a polar solvent, termed a modifier, to the SF carbon dioxide. Among these methanol is the most common, but others include ethanol, acetonitrile, formic acid, and water.[6,7] The addition of methanol or another modifier changes the solvating properties of the SF carbon dioxide as well as increasing its critical temperature and pressure. These latter changes tend to limit the amount of modifiers that can be added to the SF carbon dioxide. Typically, concentrations of less than 20% organic solvent are used. It should be noted that a specific modifier may have differential effects on solubility, depending on whether the application employs an open tubular or a packed column.[7] Such phenomena as ion pairing[22] or even reverse micelle formation can occur in the presence of certain modifiers.[23,24] Employing organic modifiers typically prevents the use of flame-based detectors and can complicate the use of other detectors such as MS. More recently small amounts of a third mobile phase component, termed an additive, may also be added to optimize separations.[9]

3.6 Elution

As mentioned earlier, SFs have properties not attendant when they exist as gases. For example, the ability to solvate molecules is greater in SFs than in the gaseous state. Moreover, the solvating capacity of an SF increases with its density (and its density is increased with increases in pressure and/or decreases in temperature). Consequently, the ability of the mobile phase to solvate the analyte of interest and displace it from the stationary phase is increased as its density increases. The effect is conceptually like the elution of a compound from a normal-phase HPLC column as the aqueous composition (i.e. polarity) of the mobile phase is increased. Elution in open tubular SFC is most frequently accomplished by increasing the density in a gradient fashion in a programmed manner, typically by increasing pressure. The relationship between pressure and density can be nearly linear for certain regions of the phase diagram. A more detailed treatment of these considerations is provided by Lee and Markides.[6] Even so, the density is not simply a function of pressure, even when temperature is fixed. Temperature programming has been used with some experimental systems to increase both separation efficiency and solvating power.[25] Nevertheless, pressure is much more easily measured and regulated, and it is the means by which all current commercial open tubular SFC systems provide gradient capability.

For packed column SFC the density is usually maintained constant and the portion of organic modifier is increased to accomplish elution of the analyte(s) in a manner that is completely analogous to gradient elution from a reversed-phase HPLC column.

3.7 Detection

Because of the high flow rate or the presence of organic modifiers associated with packed column SFC, flame-based detection does not work well. Of the flame-based methods, the most common is still FID, but other methods include flame photometric and thermionic ionization detection, but all of these have very limited application for clinical targets. Many other types of detectors exist, and in some cases the presence of organic modifiers with packed or open tubular column SFC may limit use of some of these detectors or may require splitting the eluate or require the substantial elimination of solvent prior to detection. Significant progress has occurred in interfacing detectors of many types to SFC, including many commercially available detection systems: spectrofluorimetric; UV absorption detectors, including diode array; Fourier transform infrared (FTIR) spectroscopy, evaporative light-scattering detection (ELSD); refractive index detectors; circular dichroism; and MS, especially MS instruments that have atmospheric pressure ionization capabilities. Open tubular SFC requires that there be a flow restrictor at the end of the column to maintain the supercritical conditions through the column prior to detection. Packed columns need no frit to maintain the pressure but have downstream pressure regulators. To interface packed column SFC with MS, the detector is

positioned after the pressure regulator so as to preserve the pressure control.

In addition, there are several other detector types that have been more recently interfaced to SFC. Although these are still research approaches, they demonstrate the rapid advances being made in SFC, expanding its capabilities. Undoubtedly interfaces for many of these detectors will soon be available commercially. These detectors include the following: proton NMR (nuclear magnetic resonance),[26,27] which was limited to unmodified SF carbon dioxide; chemiluminescent nitrogen detection, which tolerates modifier well and was made sensitive to 25 pg of nitrogen in a drug;[28] sulfur chemiluminescence, which was sensitive to 3 pg of S;[29] electrochemical detection, both amperometric which was sensitive to 20 pg[30] and voltametric which was sensitive to low-nanogram levels of compounds, did not require the introduction of salts, and could be applied to SF carbon dioxide with as little as 1% methanol or acetonitrile;[31] and inductively coupled plasma MS,[32] which allowed for detection of toxic elements in ultratrace quantities.

In terms of the measurement of clinically important or interesting molecules there are some specialized detection needs. For an assay to be useful, the detector must be both specific and quantitatively accurate at physiological concentrations. In the setting of clinical chemistry, one is constantly dealing with very complicated matrices such as serum, plasma, and urine which contain many, chemically diverse analytes. With these considerations in mind some SFC approaches are more attractive than others. The issue of specificity would be anticipated to be very challenging for any detector, but currently the most specific and most universal detection method is MS. In terms of sensitivity, MS, fluorescence, nitrogen and sulfur chemiluminescence and electrochemical detection are quite sensitive and compatible with the physiological concentrations of many clinical chemical targets. Although no one detection method is completely comprehensive, SFC/MS is very promising and has generated substantial interest and attention.[33]

4 APPLICATIONS IN CLINICAL CHEMISTRY

4.1 Fundamental Advantages

4.1.1 Analyte Integrity

Many biological molecules are air or heat sensitive. Certain physiological compounds are susceptible to chemical modification by means of reactions promoted by an aqueous environment, especially once out of the body's reductive environment. Consequently, attempts to use instrumental techniques to separate these compounds with subsequent quantitation has been problematic. In many instances these challenges can be overcome when SFC is employed. Specimens can be introduced onto SFC systems in nonprotic solvents (e.g. methylene chloride, ethyl acetate, dimethyl sulfoxide (DMSO), or acetone), maintained in an oxygen-free environment and handled at temperatures close to the body's physiological temperature. Moreover, SFC is a very clean method, in particular when SF carbon dioxide is used alone. SF carbon dioxide can be obtained in very high purity and any trace impurities present are predominantly other gases, which are quickly lost with depressurization. Even the trace impurities found in organic HPLC solvents can on occasion complicate the analysis of biological analytes present in very minute amounts. Additionally, both packed column and open tubular SFC rarely require the derivatization of analytes, which is commonly required for GC, leaving the analyte intact.

4.1.2 Speed of Use

Because SFC possesses high resolving power and high efficiency, analytes can be resolved rapidly, frequently in under 5 min, when packed columns are used. Times required for capillary open tubular columns often compare favorably with HPLC separation times.

4.1.3 Chromatographic Efficiency

Chromatographic efficiencies that approach those of GC can be achieved by SFC without having to modify the analyte chemically or carry out the separation at high temperature. This of course depends on the analyte and the column used. Open tubular SFC columns can produce over 3000 theoretical plates per meter. For a 30 m column, efficiencies can exceed 100 000 theoretical plates when optimal flow rates are used. In comparison, the use of a 25 cm HPLC column would provide about 20 000 theoretical plates. Efficiencies for packed SFC columns are comparable to the same columns used in HPLC, but packed columns also offer several ways to optimize separations (packing, modifier, additives, gradients, temperatures) and excellent separations can be achieved with the appropriate conditions.

4.1.4 Flexibility

One of the theoretical advantages of instrumental approaches to analyte measurement compared with individualized immunoassays or colorimetric assays is that several molecular species can be measured simultaneously. In principle, even species that are quite different

chemically may be part of the same assay. For this to be successful requires highly efficient separations and highly selective (and quite sensitive) detection, such as MS.

4.2 Limitations

4.2.1 Analyte Solubility

Many clinically important molecules are highly polar or ionic. Such compounds have very limited solubilities in SF carbon dioxide. As mentioned above, modifiers, such as methanol, afford an increased range of solubility for SFC; however, there are likely to be several analytes that are not compatible with this technique.

4.2.2 Clinical Chemical Assays

Clinical chemistry methods are somewhat unique in that there are additional demands placed on these assays that go beyond the traditional requirements of accuracy and reproducibility. Some of these revolve around the issue of sensitivity. In many cases biomolecules of clinical importance exist in picogram per milliliter concentrations or even below. In such cases the limitations of injection volume prohibit the application of SFC currently. In reviewing the current literature, it is also clear that very few SFC applications directly measure endogenous quantities of analytes in plasma and urine. In other words there are few standardized methods for clinical targets in actual biological specimens. Most assays measure standards, occasionally standards in a biological matrix, but very few have measured enough clinical material to validate an assay for potential clinical use. Finally, the clinical chemistry laboratory is expected to provide timely results. Although most SFC applications are in fact rapid, SFC systems are not typically set up for high-volume throughput. All of these limitations are not inherent in the technology per se, but reflect the limited amount of research and development that has targeted the unique needs of a modern clinical chemistry laboratory.

4.2.3 Cost

The cost of an SFC system is comparable to or slightly greater than that of a gradient-capable HPLC system. In addition to the initial cost of an instrument there are the incidental costs of mobile phases and columns. These compare favorably with HPLC or GC. Detectors are no more expensive for SFC than for other column technologies, but they can be expensive. Many clinical laboratories make very limited use of chromatographic techniques such as GC or HPLC and may have a reluctance to use them or SFC because of a lack of experience and expertise. Consequently, with few applications reported in the literature and no commercial method kits available, a laboratory would need to invest substantial time in the development of useful SFC applications. Any instrumental technique requiring significant method development and evaluation may exceed a hospital laboratory's limited resources. Commercial clinical reference laboratories, pharmaceutical industry laboratories, or SFC instrument vendors are much better positioned to take on the time and expense for such assay development.

4.3 Use for Extraction

Supercritical fluid extraction (SFE) has become widely used in the chemical industry to perform analyte extractions from solid matrices. The value of this technology is the ability to perform such extractions without creating organic waste. However, this technique has had very limited use in biomedical applications. The SFE technique has been combined with HPLC to extract and separate compounds of interest from plant materials,[34] coupled with GC to extract and separate flavors or fragrances from plant materials,[35] or coupled with SFC in the analysis of caffeine in coffee beans.[36] A more recent study by Ramsey et al.[37] used SFE to extract drugs from porcine kidney followed by SFC separation and tandem mass spectrometry (MS/MS). This approach took advantage of the drugs being relatively polar compared to other components in the kidneys. To accomplish the extraction, the kidneys were homogenized and lyophilized and a defined amount of the powdered tissue placed in an empty HPLC guard column or precolumn. Neat SF carbon dioxide was then passed through the material and onto an SFC column. The polar drugs were nonetheless solubilized in the mobile phase but were retained by the SFC column which was an amino-bonded Spherisorb® packed column. The drugs were then eluted from the column by a gradient increase of methanol in the mobile phase.[37]

4.4 Previous Applications

The SFC methods that have been developed as clinical assays are still limited in number and are not routinely employed in clinical chemistry laboratories. The analytes for which assays have been developed have included drugs or their metabolites in human specimens such as urine or plasma, environmental toxins in human specimens or in water, and finally physiological molecules. As mentioned above, these assays have not been used in the chemistry laboratory sufficiently to validate their ability to provide reliable therapeutic drug monitoring or to differentiate patients with a particular disease from a normal, reference population.

4.4.1 Drug Analysis

Many drugs have been separated by SFC. Early applications have been reviewed[38–40] as well as more recent

Table 1 Selected references for the use of SFC in drug analysis

Drug category	Refs.
Alkaloids	41, 42[a], 43[a]
Anesthetics	44
Anti-anxiety	45[a]
Antibiotics	44, 46[a], 47, 48
Antidepressants	49–51[a]
Anti-epileptics	44
Anti-inflammatory (non-steroidal)	41[a], 52–54[a]
Antipsychotics	50[a]
Asthma	41
Beta blockers (antihypertensive)	53[a], 55[a], 56[a]
Calcium channel blockers (antihypertensive)	57[a]
Cancer	58[a]
Estrogens	52[a], 59[a]
Illicit drugs	41, 44, 60[a], 61[a]
Malaria	62, 63
Muscle relaxants	41, 64
Sedatives	53[a]
Steroids	44, 66[a]
Vasodilators (antihypertensive)	67[a]

[a] Indicates analysis of pure standards only.

developments.[40] Drug analysis represents the largest single SFC application in the biomedical area. Frequently, these methods represent or have grown out of the monitoring of manufacturing processes. Table 1 lists some general drug classifications to which SFC has been applied together with selected references. The analyses limited only to standards are indicated, and are not discussed extensively. Those that involve the measurement of a drug or metabolite in serum, plasma or urine are given in more detail to provide specifics for SFC approaches that have had a measure of success in working with biological matrices.

Among the first reports of SFC applied to actual biological specimens is the work of Crowther and Henion.[41] They studied a number of drugs using packed columns, coupled to an HPLC system refurbished to accommodate SFC which used MS for detection. In their report they studied caffeine, codeine, and cocaine as pure standards. They also measured phenylbutazone, an anti-inflammatory drug, and its metabolites in an extract of equine urine. For this analysis they used a packed, amino column (20 cm by 2.1 mm ID, 5 μm amino particles, at 70 °C) and SF carbon dioxide modified with 10% methanol as the mobile phase. Elution of phenylbutazone and its metabolites was accomplished in less than 5 min.

Another early set of experiments was reported by Later et al.[44] using SFC to study several groups of drugs. These included glucocorticoids, including the measurement of prednisolone in equine urine. The urine had been previously extracted and details of specimen processing were not given. This group used an open tubular column (18 m by 50 μm ID) that had been coated with methylpolysiloxane bonded to the silica surface and neat SF carbon dioxide was employed as the mobile phase. The column was kept at 130 °C, and a stepped and ramped density gradient was used. Peaks were monitored by FID. Although relatively slow (40 min), SFC provided excellent resolution of these steroids, even for the epimers betamethasone and dexamethasone which differ only by the orientation of the hydroxyl group at the 16-carbon of the steroid ring system. This same paper reported on the separation of tetrahydrocannabinol and six of its metabolites both as pure standards and in extracts of human urine. This was achieved in under 20 min using a 15 m column (with the same inner diameter and coating) at 120 °C but with a somewhat different gradient. Using standards, they also applied the same column and mobile phase to separate tetracycline and oxytetracycline (both antibiotics) at 100 °C and later erythromycin (another antibiotic) at 40 °C. Atenolol, a beta blocker used in the treatment of hypertension, was also readily isolated using the same SFC column and mobile phase but at a temperature of 120 °C and SF carbon dioxide at a constant 0.4 g mL^{-1} density. Finally, they looked at a series of standard drugs of abuse. These again were done on the same column with the same mobile phase and included cocaine, a central nervous system (CNS) stimulant, phencyclidine (hallucinogen), methaqualone (CNS depressant), methadone (narcotic), phenobarbital (CNS depressant), and propoxyphene (analgesic). Clearly, this paper demonstrated the remarkable versatility of the approach and the wide spectrum of analytes that can frequently be separated with a single column type.

Mefloquine, an antimalarial drug, has been analyzed in an extract of human serum by Mount et al.[62] The analyte is basic and its pH was exploited to extract it first from blood (made basic by a sodium phosphate solution) into methyl *tert*-butyl ether. The mefloquine was next back extracted into an aqueous, dilute acid solution. The pH was then modified to be basic and the mefloquine was re-extracted into the ether, which was removed under a stream of nitrogen, and a small volume of a mixture of organic solvents was used to re-solubilize the analyte. The method employed SF pentane modified with 1% methanol and 0.15% *n*-butylamine, a packed column (20 cm by 0.75 mm ID, 7 μm silica particles) at 210 °C, and ECD. Constant-density elution resulted in baseline separation in about 15 min.

Perkins et al.[48] used packed column (100 mm by 4.6 mm ID, 5 μm amino-bonded Spherisorb®) SFC/MS to separate a series of sulphonamide antibiotics which had been added to a porcine kidney extract. The SF carbon dioxide modified with 15% methanol was maintained at constant density through the column at 75 °C and required less than 5 min for the sulphonamides to be separated. Yang and Griffiths have also developed

methods for the analysis of another related antibiotic, sulfanilimide.[68] This method utilized open-tubular SFC (with biphenyldimethyl silicone as the coating), higher than usual pressures (200–500 atm), and FTIR spectroscopy detection.

However, the issue of sensitivity was not considered extensively in earlier SFC drug studies. Wong[69] compared SFC with capillary electrophoresis and HPLC in the analysis of some drugs. He studied FK-206, an experimental immunosuppressive agent, in whole-blood extracts but was not successful in obtaining a measurement by SFC because of the low concentration of drug in a real sample. Pinkston et al.[64] set out to evaluate the feasibility of using SFC/MS on real biological samples. They used solid-phase (octadecylsilanol) extraction of dog plasma fortified with mebeverine, a smooth muscle relaxant. Samples (1.0 mL) were extracted and the extracted material reconstituted in a small volume of mixed organic solvents (0.025 mL). They used an open tubular column (10 m × 50 μm ID, SB-methyl 100 coating) and achieved excellent extraction efficiencies and reasonable separations using this approach. In this early approach of theirs, for a typical run of the 0.6 ng loaded on the system, only 2.6 pg were actually introduced into the analytical SFC column. However, dramatic advances in specimen introduction have been achieved, including later work by Pinkston himself.[33]

Several newer SFC approaches have achieved acceptable sensitivities. A recent study of the anticancer drug paclitaxel used packed-column SFC (with a diol stationary phase) and UV detection to achieve detection limits in the low-microgram range.[70] Others have worked to increase sensitivity to allow for the measurement of other diagnostic targets. One of these involved solid-phase extraction of serum or plasma followed by packed column SFC with either UV or ECD. This approach was used to study mitomycin C, an antibiotic, and employed a 150 mm by 4.6 mm ID column, packed with 5 μm octadecylsilanol bonded particles.[47] Using SF carbon dioxide as the mobile phase and UV absorption for detection, the method was sensitive to 20 ng of drug in 1.0 mL of plasma. A second paper reported on the use of a 200 mm by 1 mm ID column with 5 μm cyano packing, employing SF carbon dioxide as the mobile phase and ECD for measurement of artemisinin, an antimalarial drug.[65] This group also reported a limit of detection of 20 ng mL^{-1}. Although such serum concentrations are achieved for several drugs, measurement of many clinically important endogenous compounds, such as hormones, which occur at even lower concentrations, presents a greater challenge. However, a recent study by Combs et al. suggests that SFC can, with selected detectors, achieve such sensitivities.[28] They used packed column SFC with SF carbon dioxide, methanol as a modifier, and nitrogen chemiluminescence for detection. With the use of a membrane drier and a microbore column, they were able to reduce the minimum detectable quantity to 125 pg of sulfamethazine, an antimicrobial drug. Sulfur chemiluminescence,[29] electrochemical detection[30,31] and MS detection have provided comparable sensitivities in other applications.[71] Nevertheless, improvements in volume capacity in specimen introduction are probably necessary for several potential analytical targets.

Chiral separation of drugs is a rapidly developing industry. Propanol, a beta blocker and an antihypertensive drug, has been separated from several analogs using a packed SFC column that contains chyrosine-A.[10] This approach provided separations superior to those achievable using LC. Alasandro used a Chiracel OD® packed column to separate enantiomers of an oral oxazolidinone anti-bacterial agent, DuP 105.[72] Resolution of the manufactured racemic mixture was achieved by making use of temperature, pressure, modifier concentration, and modifier type. Chiral coatings have also been developed that allow for open tubular SFC. Shurig et al. used Chirasil-DEX® as a stationary phase in a standard open tubular SFC column.[73] Using this approach they were able to separate hexobarbital, a member of the barbituate family.

The use of SFC in the pharmaceutical industry continues to expand. It is no longer a novelty method but is a routine part of many protocols in the monitoring of drug production. Its applicability to so many chemically divergent analytes and its increased ease of operation have made it a valuable analytical tool.

4.4.2 Environmental Toxin Analysis

SFC has also been used in the detection of environmental toxins in human materials. There have been limited studies of this type, but this appears to be a very important application of SFC in the future. Niessen et al.[74] measured the herbicide diuron in plasma. They carried out solid-phase extraction on a SFC precolumn, which retained about 80% of the material loaded. They then used SF carbon dioxide modified with 2% methanol to desorb the analyte from the precolumn. A packed column (150 mm by 4.6 mm ID, 5 μm octadecylsilanol particles) was used for the chromatographic separation, using SF carbon dioxide modified with 2% methanol as the mobile phase, and coupled to UV absorption detection. A detection limit of 30 ng mL^{-1} was reported.

More recently, the urinary metabolites of styrene, used commonly in the production of plastics, were measured in urine by SFC/UV.[75] Organic-phase extraction of urine was used after the urine had been saturated with NaCl. A packed column (150 mm by 3.2 mm ID, 5 μm diol packing) eluted with 1.75 mL min^{-1} SF carbon dioxide modified with methanol–water–methane sulfonic acid, 95 : 5 : 0.01 v/v, was employed and the

analysis was completed in about 12 min. A detection limit was not determined, but samples spiked to 1.5 μg mL^{-1} were readily detected by UV absorption. Another study evaluated 2-thiothiazolidine-4-carboxylic acid (TTCA), which is held to be the best indicator of occupational exposure to carbon disulfide, in urine by SFC.[76] They found that diet accounted for a substantial amount of TTCA after eating uncooked cabbage.

Other work has evaluated environmental contaminants in water by SFC, work that may have implications for clinical chemistry. Phenol and nitrophenols were extracted from both tap and river water and separated by packed column SFC coupled with UV detection.[77] Another approach by the same group used disks with either polystyrene–divinylbenzene or octadecylsilanol as a sorbent to remove polycyclic aromatic hydrocarbons from water. In those experiments the extraction disks could be coupled on-line to the packed column SFC system.[15] They later applied this approach to the detection of low-level organic contaminants in water.[14]

4.4.3 Endogenous Physiologically Important Analytes

The applications of SFC to the analysis of physiological molecules has been limited. The approaches used for these analytes are the same as those used for drugs or environmental toxins. For biomolecules of clinical interest, there are several reports of SFC separations using pure standards, but fewer for the analytes in human specimens such as serum, plasma, or urine. As before, the methods that actually measure biomolecules in biological samples are more extensively reviewed, so as to provide a sense of the range and diversity of approaches that have been used. Other SFC separations of biomolecules that involved only the analysis of pure standards are listed selectively in Table 2 but are not discussed in detail.

David and Novotny[91] analyzed steroids both as standards and as components added to serum or urine. Nitrous oxide was used as the mobile phase (nitrous oxide is extremely dangerous and should not be used). The steroids were quantitatively derivatized to form thiophosphinic esters of their hydroxyl groups prior to SFC. This was done to allow for phosphorus thermionic detection. Sample preparation was consequently involved. Serum or urine was passed over an octadecyl-bonded packing material to extract the analytes and then they were eluted with methanol, dried, and reconstituted in water. Samples were then put through a cation-exchange step (SP-Sephadex), the steroid conjugates being fractionated on a TEAP-LH-20 column (strong cation exchange), and the separated conjugates enzymatically treated to remove the glucuronide or the sulfate groups. Then the various fractions were submitted to anion-exchange chromatography. Thereafter, the steroids were derivatized, which involved a 12 h incubation, and further cleaned up. Finally, the sample was injected onto a 10 m by 50 μm ID column coated with polymethylhydrosiloxane and overlaid with SE-33 (a wax) which was cross-linked. A pressure gradient of 8 atm min^{-1} over 40–50 min was used to accomplish the separation. The detection limit was not determined. Although this approach was successful in providing excellent separations, it is obviously not practical for routine use in the clinical chemistry laboratory where timely results are critical. Edlund and Henion[90] also measured steroids in equine urine. The specimen was collected 2 h after a large-bolus intramuscular injection of prednisolone. In this case the steroid was extracted from sodium borate-treated urine with ethyl acetate. Thereafter, the organic layer was streaked onto a thin-layer chromatography (TLC) plate and chromatographed. The appropriate spot was removed, cleaned of particulates, and introduced onto a packed SFC column (240 mm by 2.1 mm ID, 3 μm cyano-Spherisorb® particles) and delivered to MS by means of an SF carbon dioxide mobile phase. No attempt was made to determine the detection limit. Although effective in extracting and cleaning up the specimen, these methods are simply too time intensive for routine use in the clinical chemistry laboratory.

Scalia and Games[78] studied conjugated bile acids in human bile by means of packed column SFC. Bile was homogenized and centrifuged and the supernatant collected, diluted, and passed over a solid phase (octadecylsilanol) cartridge. The conjugated bile acids were eluted in 0.5 mL of methanol. Of this, 10 μL were injected onto the SFC system which employed a packed column (250 mm by 4.6 mm ID, 5 μm octadecylsilanol coated particles), SF carbon dioxide modified with 20% methanol as the mobile phase, and UV absorption detection. Eight different bile acids were identified in a 13 min run. Bile from patients with liver cirrhosis and hepatitis were compared to normal subjects. This is perhaps the only study to date in which there was an attempt made to evaluate the diagnostic power of an SFC assay approach (even though it was qualitative and not quantitative), even though the

Table 2 Selected references for the use of SFC in the analysis of physiological compounds

Compound classification	Refs.
Bile acids	78, 79
Cholesterols	80, 81, 82[a]
Fatty acids	83[a], 84[a]
Glycolipids	85[a], 86[a]
Prostanoids	87[a]
Retanoids	88, 89[a]
Steroids	52[a], 67[a], 90[a], 91[a]
Sugars	92[a], 93[a], 94, 95[a]
Triacylglycerols	96[a], 97[a]

[a] Indicates analysis of pure standards only.

number of specimens assayed was quite limited. The use of SFC to separate and assay bile salts has been recently reviewed and compared to other analytical instrumental approaches.[98]

Serum retanoids have been analyzed by Xie et al.[88] This involved the direct injection of rat serum. Serum (5 μL) was injected as five 1 μL aliquots onto a deactivated, open tubular column using solvent venting. The stream of nitrogen used caused the specimen to be deposited on the surface of the precolumn (2 m by 100 μm ID). Then SF carbon dioxide was used to desorb the materials from the precolumn and carry them into the first analytical column (3 m by 50 μm ID, deactivated with polycyanopropylhydrosiloxane and coated with oligoethylene oxide-substituted polysiloxane (glyme)). This provided a powerful clean-up step. The SFC system was a two-dimensional system that allowed for moving the relevant regions of the first analytical SFC column onto a second analytical SFC column (8 m by 50 μm ID, coated with a liquid crystalline polysiloxane). Both columns employed SF carbon dioxide as the mobile phase and FID was used for detection. Using a spiked specimen, 13-*cis*-retinoic acid was separated from the all-*trans* retinoic acid.

An unusual oligosaccharide was isolated from the urine of patients with gangliosidosis and characterized using SFC/MS or SFC followed by NMR.[94] The neutral oligosaccharides were isolated by a complex series of steps including a celite column, followed by a paper chromatography step that lasted from 4 to 12 days. An open tubular column (10 m by 50 μm ID, coated with diphenyl/dimethylpolysiloxane, 1:1 at 120 °C) was used for SFC with SF carbon dioxide gradient as the mobile phase and chemical ionization MS as the detector.

Others have analyzed cholesterol and cholesterol esters in human serum by SFC. Kim et al.[81] extracted serum with methanol–chloroform (2:1, v/v), the organic phase was collected and dried and the specimen was reconstituted in hexane. Capillary SFC was employed using a column (10 m by 50 μm ID) coated with *n*-octyl- and methylpolysiloxane (1:1). The mobile phase used was SF carbon dioxide with a linear pressure gradient applied over 2 h at 65 °C coupled to FID as the detector. The esters were converted to cholesterol. The method successfully accomplished baseline resolution of 10 separate cholesterol compounds. Nomura et al.[80] also measured these same analytes using a packed column. Serum was processed in a very similar manner, employing methanol/chloroform to extract the cholesterols. Specimens were reconstituted in 50 μL of hexane of which 1 μL was injected. The column was 250 mm by 4.6 mm ID, packed with 5 μm octadecylsilanol bonded particles, and the mobile phase was SF carbon dioxide maintained at 45 °C and 200 atm. Detection was by FID with the run requiring about 20 min. Sjörberg and Markides have recently interfaced SFC with atmospheric-pressure chemical-ionization MS and applied it to cholesterol and fatty acids.[99] Laasko and Manninen have also recently interfaced SFC with atmospheric-pressure chemical-ionization MS to measure lipids.[100] Using a nonpolar packed column they achieved excellent separation of both saturated and unsaturated triacylglycerols. This approach may be readily adapted to more physiological specimens.

4.5 Comparisons with Other Methods

As described throughout, SFC has several favorable chromatographic properties that provide it with a versatility and chromatographic efficiency that compare well with and in many cases exceed other instrumental approaches. The cost for SFC systems and their operation are somewhat, though not dramatically, higher than GC or HPLC systems. The variety of detectors that can be interfaced with SFC continues to grow, affording more sensitive and selective assays. However, the limited use of any of the these instruments in the clinical laboratory currently may perpetuate the status quo given the increasing pressures for laboratories to adopt standardized methodologies used widely at many institutions, and a regulatory environment that discourages method development by individual laboratories themselves. Although automated instruments are in high demand in the hospital setting, most of these are systems developed around the concept that each analyte is measured by a single assay involving unique reagents, whether immunological, enzymatic, or colorimetric. That is to say instruments generally used in laboratories do not rely on separation efficiencies and characteristic retention times to provide isolation of an analyte from other matrix components which can be uniquely delivered to a universal detector. Although SFC provides this added selectivity, such approaches are foreign to most clinical laboratorians and will require education to broaden their perspectives. Additionally, the attractiveness of SFC would be greatly enhanced if sample introduction methods improve sufficiently to allow the capture of most of the analyte in a specimen and thereby increase assay sensitivity. Finally, the general lack of SFC method development specifically for clinical chemistry has been a barrier to its broader use.

5 FUTURE DEVELOPMENTS

The future for SFC is bright. Even in clinical chemistry, rapid and sensitive detection instruments now offer increased specificity and sensitivity while managing large numbers of rapidly eluting analytes from complex matrices such as serum or plasma. Some state-of-the-art systems are expensive and may be beyond the

current resources of hospital laboratories. Consequently the financial means to take advantage of the remarkable potential of SFC seems more likely to reside with commercial reference laboratories, pharmaceutical companies, or analytical instrument companies, which may also create new markets for their services or products in this area. The vast number of broadly used, competitively priced, off-the-shelf assays already marketed specifically to clinical laboratories means that SFC will initially need to develop niche markets, such as in toxicology, assessment of environmental contaminant exposure, and/or therapeutic drug monitoring. It is also incumbent on researchers who develop SFC techniques to apply them more rigorously to clinical chemistry targets, using actual specimens and providing sufficient numbers of assayed patient and control materials to validate these promising methods.

ABBREVIATIONS AND ACRONYMS

CNS	Central Nervous System
DMSO	Dimethyl Sulfoxide
ECD	Electron Capture Detection
ELSD	Evaporative Light-scattering Detection
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
ID	Inner Diameter
LC	Liquid Chromatography
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NPD	Nitrogen–Phosphorus Thermionic Ionization Detection
SF	Supercritical Fluid
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
TLC	Thin-layer Chromatography
TTCA	2-Thiothiazolidine-4-carboxylic Acid
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Circular Dichroism in Analysis of Biomolecules • Infrared Spectroscopy of Biological Applications • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Drugs of Abuse, Analysis of • Electroanalytical Chemistry in Clinical Analysis • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Infrared Spectroscopy in Clinical Chemistry • Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Supercritical Fluid Extraction of Inorganics in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Polymers and Rubbers **(Volume 9)**
Supercritical Fluid Chromatography of Polymers

REFERENCES

1. 'Cagniard de la Tour', in *World Who's Who in Science*, ed. A. Debus, Marquis Press, Chicago, 287–288, 1968.
2. J.B. Hannay, J. Hogarth, 'On the Solubility of Solids in Gases', *Proc. R. Soc. London*, **29**, 324–326 (1879).
3. E. Klesper, A.H. Corwin, D.A. Turner, 'High Pressure Gas Chromatography above Critical Temperature', *J. Org. Chem.*, **27**, 700–701 (1962).
4. M.N. Myers, J.C. Giddings, 'Ultra-high-pressure Gas Chromatography in Microcolumns to 2000 Atmospheres', *Sep. Sci.*, **1**, 761–776 (1966).
5. T.H. Gouw, R.E. Jentoff, 'Supercritical Fluid Chromatography', *J. Chromatogr.*, **68**, 303–323 (1972).
6. M.L. Lee, K.E. Markides, 'Chromatography with Supercritical Fluids', *Science*, **235**, 1342–1347 (1987).
7. M.V. Novotny, 'Recent Developments in Analytical Chromatography', *Science*, **246**, 51–63 (1989).
8. B. Erickson, 'SFC in Flux', *Anal. Chem.*, **69**, 683A–686A (1997).
9. T.L. Chester, J.D. Pinkston, D.E. Raynie, 'Supercritical Fluid Chromatography and Extraction', *Anal. Chem.*, **70**, 301R–319R (1998).
10. N. Bargmann-Leyder, C. Sella, D. Bauer, 'Supercritical Fluid Chromatographic Separation of β-Blockers on Chyrosine A: Investigation of the Chiral Recognition Mechanism Using Molecular Modeling', *Anal. Chem.*, **67**, 952–958 (1995).

11. R.M. Campbell, H.J. Cortes, L.S. Green, 'Large-volume Injection in Capillary Supercritical Fluid Chromatography', *Anal. Chem.*, **64**, 2852–2857 (1992).
12. I.J. Koski, K.E. Markides, B.E. Richter, M.L. Lee, 'Microliter Sample Introduction for Open Tubular Column Supercritical Fluid Chromatography Using a Packed Capillary for Solute Focusing', *Anal. Chem.*, **64**, 1669–1675 (1992).
13. I.J. Koski, E.D. Lee, I. Ostrovsky, M.L. Lee, 'Solid-phase Injector for Open Tubular Column Supercritical Fluid Chromatography', *Anal. Chem.*, **65**, 1125–1129 (1993).
14. L. Toribio, M.J. del Nozal, J.L. Bernal, J.J. Jimenez, M.L. Serna, 'Packed-column Supercritical Fluid Chromatography Coupled with Solid-phase Extraction for the Determination of Organic Microcontaminants in Water', *J. Chromatogr. A*, **823**, 163–170 (1998).
15. J.L. Bernal, M.J. Nozal, L. Toribio, M.L. Serna, F. Borrull, R.M. Marce, E. Pocurull, 'Determination of Polycyclic Aromatic Hydrocarbons in Waters by Use of Supercritical Fluid Chromatography Coupled On-line to Solid-phase Extraction with Disks', *J. Chromatogr. A*, **778**, 321–328 (1997).
16. T.L. Chester, D.P. Innis, 'Quantitative Open-tubular Supercritical Fluid Chromatography Using Direct Injection onto a Retention Gap', *Anal. Chem.*, **67**, 3057–3063 (1995).
17. P.A. Peaden, M.L. Lee, 'Theoretical Treatment of Resolving Power in Open Tubular Column Supercritical Fluid Chromatography', *J. Chromatogr.*, **259**, 1–16 (1983).
18. S.M. Fields, R.C. Kong, J.C. Fjeldsted, M.L. Lee, 'Effect of Column Diameter on Efficiency in Capillary Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr.*, **7**, 312–318 (1984).
19. T.L. Chester, 'The 1988 Workshop on Supercritical Fluid Chromatography', *GC/LC*, **6**, 344–348 (1988).
20. V. Berry, L.T. Taylor, 'Supercritical Fluid Chromatography '87' ', *Am. Lab.*, **20**, 23–29 (1988).
21. T.A. Berger, W.H. Wilson, 'Packed Column Supercritical Fluid Chromatography with 220 000 Plates', *Anal. Chem.*, **65**, 1451–1455 (1993).
22. W. Steuer, M. Schindler, G. Schill, F. Erni, 'Supercritical Fluid Chromatography with Ion-pairing Modifiers. Separation of Enantiomeric 1,2 Aminoalcohols as Diastereomeric Ion Pairs', *J. Chromatogr.*, **447**, 287–296 (1988).
23. G.J. Lye, J.A. Asenjo, L. Pyle, 'Mass Transfer Studies on the Extraction and Separation of Proteins Using Reverse Micelles', *Biochem. Soc. Trans.*, **19**, 275S (1991).
24. R.W. Gale, J.L. Fulton, R.D. Smith, 'Reverse Micelle Supercritical Fluid Chromatography', *Anal. Chem.*, **59**, 1977–1979 (1987).
25. S.M. Fields, M.L. Lee, 'Effects of Density and Temperature on Efficiency in Capillary Supercritical Fluid Chromatography', *J. Chromatogr.*, **349**, 305–316 (1985).
26. K. Albert, U. Braumann, L.-H. Tseng, 'On-line Coupling of Supercritical Fluid Chromatography and Proton High-field Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **66**, 3042–3046 (1994).
27. U. Braumann, H. Handel, S. Strohschein, M. Spraul, G. Krack, R. Ecker, K. Albert, 'Separation and Identification of Vitamin A Acetate Isomers by Supercritical Fluid Chromatography – ^{1}H NMR Coupling', *J. Chromatogr.*, **761**, 336–340 (1997).
28. M.T. Combs, M. Ashraf-Khorassani, L.T. Taylor, 'Optimization of Chemiluminescent Nitrogen Detection for Packed-column Supercritical Fluid Chromatography with Methanol-modified CO_2', *Anal. Chem.*, **69**, 3044–3048 (1997).
29. H. Shi, L.T. Taylor, E.M. Fujinari, 'Open-tubular Supercritical Fluid Chromatography with Simultaneous Flame Ionization and Chemiluminescent Nitrogen Detection', *J. High Resolut. Chromatogr.*, **19**, 213–216 (1996).
30. S.R. Wallenborg, K.E. Markides, L. Nyholm, 'Development of an Amperometric Detector for Packed Capillary Column Supercritical Fluid Chromatography', *Anal. Chem.*, **69**, 439–445 (1997).
31. S.F. Dressman, A.M. Simeone, A.C. Michael, 'Supercritical Fluid Chromatography with Electrochemical Detection of Phenols and Polyaromatic Hydrocarbons', *Anal. Chem.*, **68**, 3121–3127 (1996).
32. M.J. Tomlinson, L. Lin, J.A. Caruso, 'Plasma Mass Spectrometry as a Detector for Chemical Speciation Studies', *Analyst*, **120**, 583–589 (1995).
33. J.D. Pinkston, T.L. Chester, 'Putting Opposites Together', *Anal. Chem.*, **67**, 650A–656A (1995).
34. K.K. Unger, P. Roumeliotis, 'On-line High-pressure Extraction–High Performance Liquid Chromatography. I. Equipment Design and Operation Variables', *J. Chromatogr.*, **282**, 519–526 (1983).
35. S.B. Hawthorne, M.S. Krieger, D.J. Miller, 'Analysis of Flavor and Fragrance Compounds Using Supercritical Fluid Extraction Coupled with Gas Chromatography', *Anal. Chem.*, **60**, 472–477 (1988).
36. K. Sugiyama, M. Saito, T. Hondo, M. Senda, 'New Double-stage Separation Analysis Method. Directly Coupled Laboratory-scale Supercritical Fluid Extraction–Supercritical Fluid Chromatography, Monitored with a Multiwavelength Ultraviolet Detector', *J. Chromatogr.*, **332**, 107–116 (1985).
37. E.D. Ramsey, J.R. Perkins, D.E. Games, J.R. Startin, 'Analysis of Drug Residues in Tissue by Combined Supercritical-fluid Extraction–Supercritical-fluid Chromatography–Mass Spectrometry–Mass Spectrometry', *J. Chromatogr.*, **464**, 353–364 (1989).
38. W.M.A. Niessen, U.R. Tjaden, J. Van Der Greef, 'Bioanalytical Applications of Supercritical Fluid Chromatography', *J. Chromatogr.*, **492**, 167–188 (1989).

39. S.H.Y. Wong, 'Supercritical Fluid Chromatography and Microbore Liquid Chromatography for Drug Analysis', *Clin. Chem.*, **35**(7), 1293–1296 (1989).
40. M.L. Lee, K.E. Markides (eds.), *Analytical Supercritical Fluid Chromatography and Extraction*, Chromatographic Conferences, Provo, UT, 462–468, 1990.
41. J.B. Crowther, J.D. Henion, 'Supercritical Fluid Chromatography of Polar Drugs Using Small-particle Packed Columns with Mass Spectrometric Detection', *Anal. Chem.*, **57**(13), 2711–2716 (1985).
42. G. Holzer, L.H. Zaldow, C.F. Asibal, 'Capillary Supercritical Fluid Chromatography of Pyrrolizidine Alkaloids', *J. Chromatogr.*, **400**, 317–322 (1987).
43. J.L. Janicott, M. Caude, R. Rosset, 'Separation of Opium Alkaloids by Carbon Dioxide Sub- and Supercritical Fluid Chromatography with Pack Columns', *J. Chromatogr.*, **437**(12), 351–364 (1988).
44. D.W. Later, B.E. Richter, D.E. Knowles, M.R. Andersen, 'Analysis of Various Classes of Drugs by Capillary Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **24**(6), 249–253 (1986).
45. M.Z. Wang, M.S. Klee, S.K. Yang, 'Achiral and Chiral Analysis of Camazepam and Metabolites by Packed-column Supercritical Fluid Chromatography', *J. Chromatogr., B*, **665**(1), 139–146 (1995).
46. S. Schmidt, L.G. Blomberg, E.R. Campbell, 'Modifier Effects on Packed and Capillary Columns in Supercritical Fluid Chromatography', *Chromatographia*, **25**, 775–780 (1988).
47. W.M.A. Niessen, P.J.M. Bergers, U.R. Tjaden, J. Van der Greef, 'Phase-system Switching as an On-line Sample Pretreatment in the Bioanalysis of Mitomycin C Using Supercritical Fluid Chromatography', *J. Chromatogr.*, **454**, 243–251 (1988).
48. J.R. Perkins, D.E. Games, J.R. Startin, J. Gilbert, 'Analysis of Sulphonamides Using Supercritical Fluid Chromatography and Supercritical Fluid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **540**, 239–256 (1991).
49. R.N. Gupta, 'Drug Level Monitoring: Antidepressants', *J. Chromatogr.*, **576**(2), 183–211 (1992).
50. T.A. Berger, W.H. Wilson, 'Separation of Drugs by Packed Column Supercritical Fluid Chromatography. 2. Antidepressants', *J. Pharm. Sci.*, **83**(3), 287–290 (1994).
51. T.A. Berger, W.H. Wilson, 'Separation of Basic Drugs by Packed-column Supercritical Fluid Chromatography. 3. Stimulants', *J. Pharm. Sci.*, **84**(4), 489–492 (1995).
52. N.K. Jagota, J.T. Stewart, 'Supercritical Fluid Chromatography of Selected Oestrogens', *J. Pharm. Biomed. Anal.*, **10**(9), 667–673 (1992).
53. A. Kot, P. Sandra, A. Venema, 'Sub- and Supercritical Fluid Chromatography on Packed Columns: A Versatile Tool for the Enantioselective Separation of Basic and Acidic Drugs', *J. Chromatogr. Sci.*, **32**(10), 439–448 (1994).
54. B.R. Simmons, N.K. Jagota, J.T. Stewart, 'A Supercritical Fluid Chromatographic Method Using Packed Columns for Phenylbutazone and Oxyphenbutazone in Serum, and for Phenylbutazone in a Dosage Form', *J. Pharm. Biomed. Anal.*, **13**(1), 59–64 (1995).
55. L. Siret, N. Bargmann, A. Tambute, M. Caude, 'Direct Enantiomeric Separation of Beta-blockers on ChyRoSine-A by Supercritical Fluid Chromatography: Supercritical Carbon Dioxide as Transient In Situ Derivatization Agent', *Chirality*, **4**(4), 252–262 (1992).
56. C.J. Bailey, R.J. Ruane, I.D. Wilson, 'Packed-column Supercritical Fluid Chromatography of Beta-blockers', *J. Chromatogr. Sci.*, **32**(10), 426–429 (1994).
57. J.T. Strode, 3rd, L.T. Taylor, A.L. Howard, D. Ip, M.A. Brooks, 'Analysis of Felodipine by Packed Column Supercritical Fluid Chromatography with Electron Capture and Ultraviolet Absorbance Detection', *J. Pharm. Biomed. Anal.*, **12**(8), 1003–1014 (1994).
58. S.M. Musser, P.S. Callery, 'Supercritical Fluid Chromatography/Chemical Ionization/Mass Spectrometry of Some Anticancer Drugs in a Thermospray Ion Source', *Biomed. Environ. Mass. Spectrom.*, **19**(6), 348–352 (1990).
59. E.D. Lee, S.H. Hsu, J.D. Hennion, 'Electron-ionization-like Mass Spectra by Capillary Supercritical Fluid Chromatography/Charge Exchange Mass Spectrometry', *Anal. Chem.*, **60**, 1990–1994 (1988).
60. S.H. Wong, 'Advances in Chromatography for Clinical Drug Analysis: Supercritical Fluid Chromatography, Capillary Electrophoresis, and Selected High-performance Liquid Chromatography Techniques', *Ther. Drug Monit.*, **15**(6), 576–580 (1993).
61. B. Backstrom, M.D. Cole, M.J. Carrott, D.C. Jones, G. Davidson, K. Coleman, 'A Preliminary Study of the Analysis of Cannabis by Supercritical Fluid Chromatography with Atmospheric Pressure Chemical Ionization Mass Spectroscopic Detection', *Sci. Justice*, **37**(2), 91–97 (1997).
62. D.L. Mount, L.C. Patchen, F.C. Churchill, 'Determination of Mefloquine in Blood by Supercritical Fluid Chromatography with Electron-capture Detection', *J. Chromatogr.*, **527**, 51–58 (1990).
63. D.L. Mount, G.D. Todd, V. Navaratnam, 'Packed-column Supercritical Fluid Chromatography of Artemisinin (Qinghaosu) with Electron-capture Detection', *J. Chromatogr., B*, **666**(1), 183–187 (1995).
64. J.D. Pinkston, C.J. Venkatramani, L.J. Tulich, D.J. Bowling, K.R. Wehmeyer, 'Evaluation of Capillary Supercritical Fluid Chromatography with Mass Spectrometric Detection for the Analysis of a Drug (Mebeverine) in a Dog Plasma Matrix', *J. Chromatogr.*, **622**, 209–214 (1993).
65. D.L. Mount, G.D. Todd, V. Navaratnam, 'Packed-column Supercritical Fluid Chromatography of Artemisinin (Qinghaosu) with Electron-capture Detection', *J. Chromatogr., B*, **666**, 183–187 (1995).

66. J.S. Loran, K.D. Cromie, 'An Evaluation of the Use of Supercritical Fluid Chromatography with Light Scattering Detection for the Analysis of Steroids', *J. Pharm. Biomed. Anal.*, **8**(7), 607–611 (1990).
67. I.C. Bhoir, B. Raman, M. Sundaresan, A.M. Bhagat, 'Separation and Estimation of Seven Vasodilators Using Packed Column Supercritical Fluid Chromatography', *J. Pharm. Biomed. Anal.*, **17**(3), 539–546 (1998).
68. J. Yang, P.R. Griffiths, 'Separation and Identification of Sulfanilamide by Capillary Supercritical Fluid Chromatography–Fourier Transform Infrared Spectroscopy', *J. Chromatogr.*, **785**, 111–119 (1997).
69. S.H.Y. Wong, 'Advances in Chromatography for Clinical Drug Analysis: Supercritical Fluid Chromatography, Capillary Electrophoresis, and Selected High-performance Liquid Chromatography Techniques', *Ther. Drug Monit.*, **15**(6), 576–580 (1993).
70. N.K. Jagota, J.B. Nair, R. Frazer, M. Klee, M.Z. Wang, 'Supercritical Fluid Chromatography of Paclitaxel', *J. Chromatogr.*, **721**, 315–322 (1996).
71. I.M. Lazar, E.D. Lee, A.L. Rockwood, M.L. Lee, 'Evaluation of an Electrospray Interface for Capillary Electrophoresis Time-of-flight Mass Spectrometry', *J. Chromatogr. A*, **791**, 269–278 (1997).
72. M. Alasandro, 'Separation of Diastereoisomers of DuP 105, a Novel Oxazolidinone Antibacterial Agent, by Supercritical Fluid Chromatography on a Chiracel OD® Column', *J. Pharm. Biomed. Anal.*, **14**, 807–814 (1996).
73. V. Shurig, M. Jung, S. Mayer, M. Fluck, S. Negura, H. Jakubetz, 'Unified Enantioselective Capillary Chromatography on a Chirasil-DEX® Stationary Phase. Advantages of Column Miniaturization', *J. Chromatogr., A*, **694**, 119–128 (1995).
74. W.M.A. Niessen, M.A.G. de Kraa, E.R. Verheij, P.J.M. Bergers, G.F. La Vos, U.R. Tjaden, J. van der Greef, 'Combined Supercritical-fluid Chromatography in the Analysis of Diuron in Plasma Using On-line Phase-system Switching', *Rapid Comm. Mass Spectrom.*, **3**(1), 1–4 (1989).
75. P. Simon, T. Nicot, 'Capillary Electrophoresis and Supercritical Chromatography, Complementary and Alternative Techniques for the Determination of Urinary Metabolites of Styrene', *J. Chromatogr., B*, **679**, 103–112 (1996).
76. P. Simon, T. Nicot, M. Dieudonne, 'Dietary Habits, a Non-negligible Source of 2-Thiothiazolidine-4-carboxylic Acid and Possible Overestimation of Carbon Disulfide Exposure', *Int. Arch. Occup. Environ. Health*, **66**, 85–90 (1994).
77. E. Pocurull, R.M. Marce, F. Borrull, J.L. Bernal, L. Toribio, M.L. Serna, 'On-line Solid Phase Extraction Coupled to Supercritical Fluid Chromatography to Determine Phenol and Nitrophenols in Water', *J. Chromatogr., A*, **755**, 67–74 (1996).
78. S. Scalia, D.E. Games, 'Analysis of Conjugated Bile Acids by Packed-column Supercritical Fluid Chromatography', *J. Chromatogr.*, **574**(2), 197–203 (1992).
79. S. Scalia, 'Bile Acid Separation', *J. Chromatogr., B*, **671**(1–2), 299–317 (1995).
80. A. Nomura, J. Yamada, A. Takatsu, Y. Horimoto, T. Yarita, 'Supercritical Fluid Chromatographic Determination of Cholesterol and Cholesteryl Esters in Serum on ODS-silica Gel Column', *Anal. Chem.*, **65**(15), 1994–1997 (1993).
81. D.H. Kim, K.J. Lee, G.S. Heo, 'Analysis of Cholesterol and Cholesteryl Esters in Human Serum Using Capillary Supercritical Fluid Chromatography', *J. Chromatogr., B*, **655**(1), 1–8 (1994).
82. E.B. Hoving, 'Chromatographic Methods in the Analysis of Cholesterol and Related Lipids', *J. Chromatogr., B*, **671**(1–2), 341–362 (1995).
83. K.E. Markides, S.M. Fields, M.L. Lee, 'Capillary Supercritical Fluid Chromatography of Labile Carboxylic Acids', *J. Chromatogr. Sci.*, **24**, 254–257 (1986).
84. G. Gutnikov, 'Fatty Acid Profiles of Lipid Samples', *J. Chromatogr., B*, **671**, 71–89 (1995).
85. J. Kuei, G.R. Her, V.N. Reinhold, 'Supercritical Fluid Chromatography of Glycosphingolipids', *Anal. Biochem.*, **172**(1), 228–234 (1988).
86. M.V. Merritt, D.M. Sheeley, V.N. Reinhold, 'Characterization of Glycosphingolipids by Supercritical Fluid Chromatography–Mass Spectrometry', *Anal. Biochem.*, **193**(1), 24–34 (1991).
87. I.J. Koski, B.A. Jansson, K.E. Markides, M.L. Lee, 'Analysis of Prostaglandins in Aqueous Solutions by Supercritical Fluid Chromatography', *J. Pharm. Biomed. Anal.*, **9**(4), 281–290 (1991).
88. L.Q. Xie, K.E. Markides, M.L. Lee, N.K. Hollenberg, G.H. Williams, S.W. Graves, 'Bioanalytical Application of Multidimensional Open Tubular Column Supercritical Fluid Chromatography', *Chromatographia*, **35**(7/8), 363–371 (1993).
89. R. Wyss, 'Chromatographic and Electrophoretic Analysis of Biomedically Important Retinoids', *J. Chromatogr., B*, **671**, 381–425 (1995).
90. P.O. Edlund, J.D. Henion, 'Packed-column Supercritical Fluid Chromatography/Mass Spectrometry via a Two-stage Momentum Separator', *J. Chromatogr.*, **27**(6), 274–282 (1989).
91. P.A. David, M. Novotny, 'Analysis of Steroids by Capillary Supercritical Fluid Chromatography with Phosphorus-selective Detection', *J. Chromatogr.*, **461**, 111–120 (1989).
92. R. Kornfeld, 'Reliable Analytical Determinations Using Supercritical Fluid Chromatography', *Am. Lab., Sept.*, 25–33 (1992).
93. J.P. Ceasar, Jr, D.M. Sheeley, V.N. Reinhold, 'Femtomole Oligosaccharide Detection Using a Reducing-end Derivative and Chemical Ionization Mass Spectrometry', *Anal. Biochem.*, **191**(2), 247–252 (1990).

94. J.C. Michalski, J. Lemoine, J.M. Wieruszeski, B. Fournet, J. Montreuil, G. Strecker, 'Characterization of a Novel Type of Chain-terminator Gal Beta 1-6 Gal Beta 1-4 GlcNAc in an Oligosaccharide Related to N-Glycosylated Protein Glycans Isolated from GM1 in the Urine of Patients with Gangliosidosis', *Eur. J. Biochem.*, **198**(2), 521–526 (1991).
95. D.M. Sheeley, V.N. Reinhold, 'Characterization of N-linked Glycans by Supercritical Fluid Chromatography', *Anal. Biochem.*, **193**(2), 240–247 (1991).
96. H. Kallio, P. Laasko, 'Effect of Carbon Dioxide Flow-rate on the Separation of Triacylglycerols by Capillary Supercritical Fluid Chromatography', *J. Chromatogr.*, **518**(1), 69–75 (1990).
97. P. Laasko, P. Manninen, J. Makinen, H. Kallio, 'Post-parturition Changes in the Triacylglycerols of Cow Colostrum', *Lipids*, **31**(9), 937–943 (1996).
98. A. Roda, F. Piazza, M. Baraldini, 'Separation Techniques for Bile Salt Analysis', *J. Chromatogr., B*, **717**, 263–278 (1998).
99. P.J.R. Sjöberg, K.E. Markides, 'New Supercritical Fluid Chromatography Interface Probe for Electrospray and Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr.*, **785**, 101–110 (1997).
100. P. Laasko, P. Manninen, 'Identification of Triacylglycerols by Capillary Supercritical Fluid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Lipids*, **32**, 1285–1295 (1997).

Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Stephen L. Upstone
PerkinElmer Ltd., Beaconsfield, UK

1 Introduction 1

2 Principles of Analysis and Instrument Parameters 2
- 2.1 Fundamentals of Ultraviolet/Visible Spectroscopy and the Beer–Lambert Law 2
- 2.2 Linearity and Deviations from the Beer–Lambert Law 2
- 2.3 Cuvettes and Solvents 3
- 2.4 Resolution (Band-pass) and Slit Width 4
- 2.5 First- and Second-derivative Spectroscopy 5
- 2.6 Verification of Spectrophotometer Performance 5

3 Overview of Instrument Designs and Their Advantages and Disadvantages 6
- 3.1 Dispersive and Diode-array Systems 6
- 3.2 Dispersive Ultraviolet/Visible Spectrophotometers 6
- 3.3 Single-beam Spectrophotometers 6
- 3.4 Double-beam Spectrophotometers 6
- 3.5 Dual-beam Spectrophotometers 8
- 3.6 Photodiode-array Spectrophotometers 8
- 3.7 Microplate Reader Spectrophotometers 8
- 3.8 Reflectance-based Analyzers 9

4 Common Clinical Applications Using Ultraviolet/Visible Absorption Spectroscopy 9
- 4.1 Enzyme Rate Assays 9
- 4.2 Colorimetric and End-point Assays 10
- 4.3 Immunoassays, Enzyme-linked Immunosorbent Assays and Microplate Assays 11
- 4.4 Porphyrin Analysis 11
- 4.5 Hemoglobin Analysis 12
- 4.6 Protein Assays 12
- 4.7 Molecular Biology 13

Acknowledgments 13

Abbreviations and Acronyms 13

Related Articles 14

References 14

Ultraviolet/visible (UV/VIS) absorption spectroscopy has been used in the clinical laboratory for many years. The technique has appeal, as it is almost universal in its application. Although much of the routine work is performed using high-throughput dedicated clinical analysis systems, absorption spectroscopy still has a place in most clinical laboratories.

This article discusses the range of application types for which absorption spectroscopy can be used and some examples of common analyses are given. The article also discusses the merits of various instrument types and discusses some more advanced spectroscopic techniques, such as derivative spectroscopy, to enhance the data measured by the spectrophotometer. Brief reference is also made to the use of reflectance in clinical analysis.

1 INTRODUCTION

Ultraviolet/visible (UV/VIS) absorption spectroscopy has been used in the clinical laboratory for many years. The technique has appeal, as it is almost universal in its application. The reliance on UV/VIS spectrophotometers (and other spectroscopic techniques) has diminished over the years, however, as dedicated multichannel clinical analyzers and readers have appeared on the market. These analyzers offer high sample throughput and clinically validated proprietary chemistries with ready-mixed reagents for greater speed and convenience. These analyzers normally utilize a range of analytical techniques such as ion-selective electrodes, turbidometry and UV/VIS absorption in one integrated instrument. Reagent manufacturers are also trying to simplify the analysis even further so that much simpler instrumentation is required. There is some work on solid-phase chemistry systems in which the completed reaction on the solid surface can be measured on a simple reader which uses a simple light-emitting diode (LED) source and measures light reflectance from the active face.

Spectroscopic analysis, in general, is in a slow decline in the clinical laboratory as the increasing adoption of molecular biological techniques such as the polymerase chain reaction (PCR) and DNA sequencing are now being used to measure (for example) genetic abnormalities rather than the more indirect biochemistry-based approach.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

I have spent some length explaining the fundamentals of the technique and about instrument design. This is deliberate – too often people treat UV/VIS spectrophotometers as a simple ''meter'' without understanding their limitations or suitability for a particular analysis. This article should also be useful for understanding spectrophotometric assays on dedicated analyzers.

2 PRINCIPLES OF ANALYSIS AND INSTRUMENT PARAMETERS

2.1 Fundamentals of Ultraviolet/Visible Spectroscopy and the Beer–Lambert Law

UV/VIS spectroscopy is the study of how a sample responds to light. When a beam of light passes through a substance or a solution, some of the light may be absorbed and the remainder transmitted through the sample. The ratio of the intensity of the light entering the sample (I_0) to that exiting the sample (I_t) at a particular wavelength is defined as the transmittance (T). This is often expressed as the percentage transmittance ($\%T$), which is the transmittance multiplied by 100 (Equation 1):

$$\%T = \left(\frac{I_0}{I_t}\right. \times 100 \qquad (1)$$

The absorbance (A) of a sample is the negative logarithm of the transmittance (Equation 2):

$$A = -\log T \qquad (2)$$

The UV/VIS range of the electromagnetic spectrum covers the range 190–700 nm (most instruments are capable of measuring at longer wavelengths than this, depending on their detector type). For clinical analysis this is useful, as water (most assays are in aqueous solution) is almost completely transparent in this region.

Most clinical assays are concerned with quantitation rather than identification (there are more powerful techniques available to perform the latter) as absorption spectra tend to be featureless – they lack the fine structure which is found, for example, in an infrared (IR) spectrum. Some techniques to enhance resolution and qualitative information will be discussed later.

The most important principle in absorption analysis is the Beer–Lambert law. This law states that, for a given ideal solution, there is a linear relationship between concentration and absorbance provided that the path length is kept constant; the absorptivity (ε) is a constant for each molecule for each wavelength (Equation 3):

$$A = \varepsilon cl \qquad (3)$$

where ε = the absorptivity of the substance, c = concentration and l = path length. Provided that ε and l are kept constant for a given set of experiments, a plot of the sample absorbance against the concentration of the absorbing substance should give a straight line. In practice, a calibration curve is prepared by plotting the absorbance of a series of standard samples as a function of their concentration. If the absorbance of an unknown sample is then measured, the concentration of the absorbing component can be assessed from this graph. Another consequence of the Beer–Lambert law is that it is possible to change the path length to affect the absorbance. This can be useful where lower detection limits are required as the path length can be increased (longer path length cuvettes are available) or, where the absorbance is too high to be measured on the instrument, the path length can be reduced. Alternatively, it is possible to reduce the absorbance by diluting the sample, but one has to take care when dealing with biologically active samples, particularly enzyme-based solutions, as this may have a profound effect on the activity.

2.2 Linearity and Deviations from the Beer–Lambert Law

Above, the Beer–Lambert Law has been shown to be a linear relationship between sample concentration and absorbance. If this relationship is tested experimentally by measuring samples of increasing concentration and the results are plotted, the relationship will be seen to break down with increasing absorbance. With general laboratory UV/VIS spectrophotometers of recent design, this will probably be around 3 absorbance units (AU).

What is observed here is not a breakdown in the Beer–Lambert law (although some assumptions about the sample being infinitely dilute are made in the law) but a limitation in the instrument's performance.

The greatest contributor to this nonlinearity is the instrument's stray light. Stray light can be loosely defined as the amount of light present in the instrument at nonanalytical wavelengths or, put another way, the amount of unwanted light present in the system. This unwanted light can come from several sources. First, there may be light from outside entering the optical system. Manufacturers usually go to some lengths to exclude this light and so it ought to be negligible in a well-designed instrument. The main source of stray light comes from the monochromator itself.

Monochromators (in dispersive instruments) are usually diffraction gratings. These gratings, although highly efficient producers of monochromatic radiation, are not entirely efficient. They will also allow small amounts of light to pass through at other discrete wavelengths, in addition to small amounts of white light. We have seen

Table 1 Observed absorbance readings for three different instrument stray light specifications

True absorbance	1% T stray light[a]	0.01% T stray light[b]	0.0001% T stray light[c]
1.0	0.9788	0.9996	1.0000
2.0	1.8239	1.9957	2.0000
3.0	2.2218	2.9586	3.0000
4.0	2.2924	3.6999	3.9957
5.0	2.3009	3.9586	4.9586

[a] Low-cost or old instrument.
[b] High-performance double-beam instrument.
[c] "Top of the range" double-monochromator, double-beam instrument.

earlier that absorbance is a logarithmic scale and so 1 AU corresponds to 10% T, 2 AU corresponds to 1% T, 3 AU to 0.1% T and so on.

If the light throughput is expressed in percentage terms (%T) and a stray light value (from the instrument manufacturer's data) of 0.05% transmittance is quoted, this will represent a third of the total light seen by the detector at 3 AU (i.e. 0.1% T). This will cause an underreporting of the true absorbance. By 4 AU (0.01% T), there will be more stray light than sample signal, rendering the instrument unusable at this value. Table 1 shows the anticipated effect on the observed absorbance for a range of stray light values which may be encountered on various instruments found within the clinical laboratory.

2.2.1 Other Causes of Non-linearity

Occasionally, there are other reasons why there may be unexpected deviations from linearity.

One major cause is sample fluorescence. If a fluorophore is excited at the same wavelength at which one is trying to measure absorption (the chromaphore), the detector will see a combination of absorbed light and the fluorescence emission. This emission will occur at a longer wavelength than the absorption. It should be borne in mind that there is no monochromator (on a dispersive instrument) between the sample and detector, so the emission will be detected. The observed result will be a lower absorbance reading than expected (as the detector is receiving light from two phenomena instead of one).

If sample fluorescence is severe, it can be removed by inserting a cutoff filter between the sample and the detector. This filter should allow light of the absorbance wavelength to pass through but block higher wavelengths.

Sometimes the linearity can be affected by changes in pH. Some chromophores (the functional group in the molecule responsible for the light absorption) will shift wavelength with a change in pH. If the absorption peak is particularly sharp, it may mean that measurement is no longer at the actual peak but at its side. More sophisticated instruments can detect the actual peak for every measurement or measure areas under the absorption curve between two user-defined points.

Another possible reason for nonlinearity is sample scatter. The instrument expects light to travel through the sample in a straight line (unless there is special provision in the instrument design). If the sample is turbid, this light will be scattered and the light will be splayed out in a cone-like fashion. As a result, less light will fall on the detector and a falsely high absorbance reading will be observed. Turbidity is relatively easy to observe as it is not wavelength specific. A scattering sample, therefore, will have a raised baseline. In bad cases, the baseline may be elevated over 1 AU.

2.3 Cuvettes and Solvents

Apart from ensuring that the instrument design is sufficient for the required analysis, the user must also check that the other components of the analytical system – the cuvettes and the solvents – are applicable for the required task. Cuvettes are offered in either optical glass or quartz. Optical glass will transmit light above 320 nm whereas quartz will transmit light well below 190 nm (the usual low-wavelength limit for UV/VIS spectrophotometers).

As an alternative to glass and quartz cuvettes, several manufacturers offer disposable cuvettes. These are made from various plastics including polystyrene and acrylic. These cuvettes offer the main advantage that they do not need cleaning between samples, which is very useful when handling biohazardous samples. The disadvantages are that they have a restricted wavelength range (one should carefully check the manufacturer's stated wavelength specification before using disposable cuvettes) and they have a lower optical performance. In some cases striations across the face of the cuvette might be observed. These are due to the molding process in the cell manufacture. There will also be a molding mark at the bottom of the cuvette (from the extrusion process) and this can make stirring with a small magnetic flea difficult (with instruments which provide a stirring facility). Another disadvantage with disposable cuvettes is that the cell path length (an important consideration for the Beer–Lambert law) will not be as accurate as with glass or quartz cells and this may have quantitative accuracy implications. Figure 1 shows the light transmission charateristics of some common cuvette materials.

As regards solvents, water has the widest range and the lowest cutoff wavelength. Cutoff wavelengths are somewhat arbitrary and there will be variations in the stated values from different literature sources. I have used the definition as being the wavelength

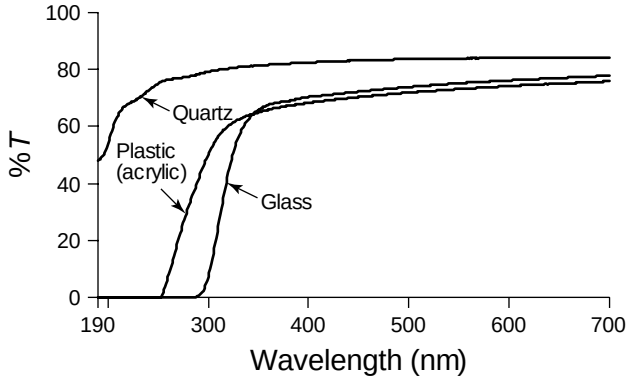

Figure 1 Light transmission characteristics of various cuvette materials.

Table 2 Cutoffs for common solvents

Solvent	Cutoff (transmittance of 10% in a 1-cm cell) (nm)
Water	190
2-Propanol (isopropyl alcohol)	210
Ethanol	210
Methanol	210
Acetonitrile	210
Dichloromethane	235
Toluene	270
Acetone	330
Chloroform	250

at which a 1-cm cuvette filled with pure solvent has a transmittance of 10% (equivalent to 1 AU). The cutoffs for common solvents are given in Table 2. Some buffers [e.g. tris(hydroxymethyl)aminomethane (TRIS)] have organic components and so it should not be assumed that the cutoff is identical with that for water.

2.4 Resolution (Band-pass) and Slit Width

UV/VIS instruments have a specified band-pass. This term relates to the instrument's ability to resolve peaks which are very close together or "shoulders" (small features on the side of a larger peak). The band-pass will either be fixed, usually between 1 and 2 nm, or, in the case of more expensive instruments, variable (either continuously or in finite steps). Photodiode-array instruments have a fixed optical band-pass as there is no physical slit aperture on these instruments (although the software may be able to simulate other slit conditions); the resolution is defined by the light dispersion, the polychromator and the number of elements in the array.

The band-pass will affect the instrument's ability to discriminate between sharp features. Most molecules possess a broad spectrum, so a small band-pass will not change the appearance of the spectrum (apart from increased noise). Some molecules commonly encountered in clinical analysis, such as porphyrins and the various types of hemoglobin (oxy-, met- and carboxy-) give sharp structures and the position of the peak has to be accurately measured or a misdiagnosis may otherwise result.

Figure 2 shows the effect on a sample with fine structure, i.e. benzene vapor in a sealed cuvette. This is a highly artificial situation, as typical samples in the clinical laboratory do not exhibit such sharp features.

Figure 3 shows the effect of variable resolution on a more typical sample with a single broad peak. In this instance there is little or no difference between a high- and low-resolution (band-pass) scan.

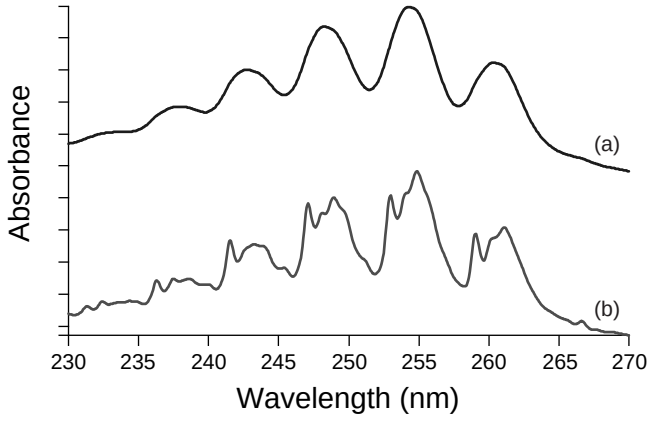

Figure 2 Effect of changing the instrument resolution on a sample with sharp peaks (benzene vapor) showing the effect on band shapes and illustrating sharper discrimination using a 0.5-nm slit (b) compared with a 2-nm slit (a).

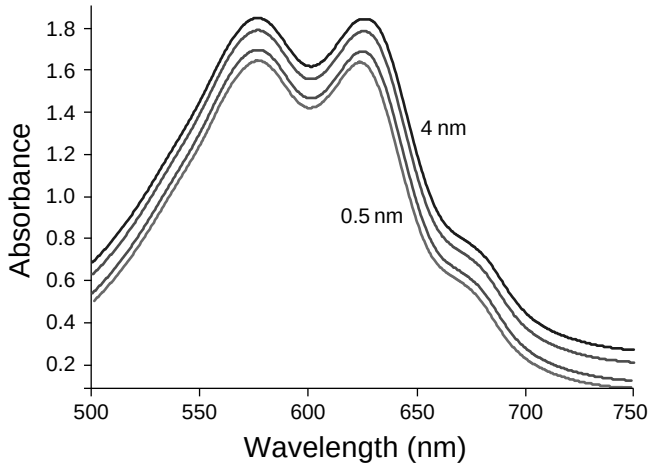

Figure 3 Effect of using various slit settings on a typical spectrum with broad bands (spectra offset for clarity) showing that the slit setting is largely irrelevant for broad peaks.

2.5 First- and Second-derivative Spectroscopy

UV/VIS spectra are rather featureless at room temperature, which is why the technique is primarily used for quantitative analysis rather than for sample identification purposes. A broad ultraviolet (UV) spectrum is composed of many separate electronic transitions which, at room temperature, become ''blurred'' to give the appearance of a single entity. Spectra only become more defined at very low temperatures and in the gas phase. Neither of these two conditions is applicable in routine clinical analysis.

One major advantage that UV/VIS analysis possesses for biological measurements is that water (which is the main component of all living systems) is virtually transparent across the entire range (there is some very weak activity around 860 nm) and so any technique to improve the qualitative data is welcome.

Derivative spectroscopy is a useful tool to give some improvement. As the name suggests, the technique consists of plotting the rate of change of the absorbance spectrum versus wavelength. This will give a plot (shown in Figure 4 with the original spectrum). At peak maxima and minima (and also points of inflection), the graph is seen to pass through zero on the ordinate. As a result, the technique can be used to identify peak maxima and minima. This is termed a first-derivative spectrum.

First-derivative spectra have their uses but they are difficult to interpret (as the brain often sees the 'peaks' rather than the true information). If we take the second-derivative spectrum (by derivatizing the absorbance spectrum twice), we obtain a plot such as that in Figure 5.

The second-derivative spectrum can be thought of as an inverted spectrum. Sharp peaks will be made even sharper. Broad peaks (and also broad background

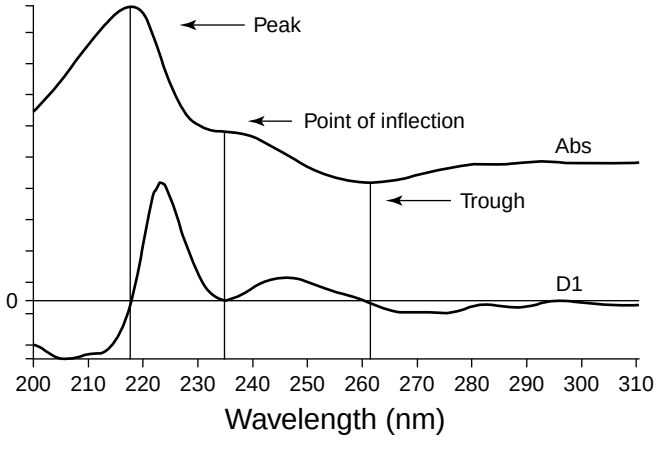

Figure 4 Explanation of the first-derivative (D1) spectrum, and its relationship to the absorption spectrum (Abs).

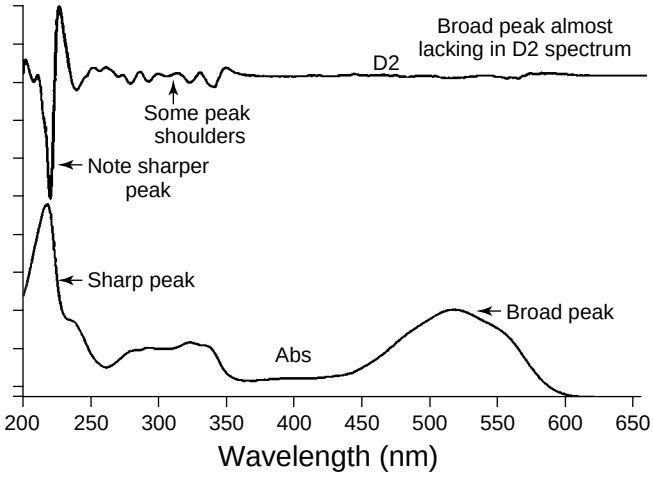

Figure 5 Second-derivative (D2) spectrum showing sharp band discrimination.

features) will become flattened and so the technique can be used both as a peak-enhancement and background-rejection tool. The second-derivative spectrum, like its absorbance counterpart, still contains quantitative information. If two points are consistently used on the spectrum and the difference in their values on the ordinate is taken, this can be plotted against concentration and a linear relationship established.

Derivative spectra represent an alternative presentation of the original data. Information cannot be created and so there is a cost involved. This cost is at the expense of a much poorer signal-to-noise ratio (as the noise is also being derivatized) and so if a requirement for derivative spectroscopy is anticipated, an instrument with a good noise specification is highly desirable. Single-beam instruments should not be considered for serious derivative spectroscopy.

Higher order derivatives also are possible, and usually up to the fourth derivative is offered on current instrumentation, but the increased noise makes their general analytical use questionable.

2.6 Verification of Spectrophotometer Performance

Clinical analysis is concerned with producing an accurate result to enable the clinician to make an accurate diagnosis and to provide the correct treatment. It is important, therefore, that all equipment used to make the diagnosis is well maintained and that the operators have a sufficient skill level to understand the limitations of the instrument and to use the correct operating and sampling procedures.

Instruments should be checked regularly, using a certified reference material (CRM). As a minimum, the user should possess a sample for checking wavelength

accuracy and a set of absorbance filters. Wavelength standards should never be used to check absorbance accuracy and vice versa. Wavelength standards (such as holmium oxide glass filters) have very sharp peaks whereas absorbance standards such as the National Institute of Standards and Technology (NIST) 930D filters are of neutral density.

Additional checks such as for stray light and noise may also be advantageous. Most instrument companies will be able to give advice on instrument performance checks and routine maintenance.

Microplate reader users are more limited in the choice of calibration aids. For filter-based instruments (the majority at the time of writing), there is no way to check wavelength accuracy (other than to scan the filter on a spectrophotometer) and there are few CRMs available for all formats of microplates. If one is using a microplate reader, the manufacturer should be contacted for advice.

3 OVERVIEW OF INSTRUMENT DESIGNS AND THEIR ADVANTAGES AND DISADVANTAGES

3.1 Dispersive and Diode-array Systems

Before going into details of analysis, it is important to discuss the design of a UV/VIS specrophotometer and the implications of the various design types for a particular analysis, and to understand the limiting factors in making that measurement.

UV/VIS instruments are available in two main types, dispersive and photodiode array (usually shortened to diode array). The dispersive design uses a monochromator before the sample to convert the white light produced by the light source into a single pure wavelength of light. This single wavelength is then passed through the sample and detected. The monochromator can be either a fixed-wavelength filter or a variable-wavelength design such as a prism (rare nowadays) or a diffraction grating. On higher specification instruments, there may be two monochromators linked in series to permit higher performance – particularly at high absorbance values (e.g. over 3.0 AU).

3.2 Dispersive Ultraviolet/Visible Spectrophotometers

Dispersive UV/VIS spectrophotometers have four main optical configurations:

- single beam
- double beam (single detector)
- dual beam (dual detector)
- photodiode array.

3.3 Single-beam Spectrophotometers

The single-beam spectrophotometer is the simplest optical configuration. It consists of a light source, monochromator (either a grating or a filter), the sample area and the detector, as shown in Figure 6. In the single-beam design, it is necessary first to zero the instrument (to establish the I_0 value) and then to measure the sample.

Single-beam instruments are used mainly on grounds of cost. They have poorer noise specifications compared with their dual- and double-beam counterparts and are prone to drifting with time. This makes accurate kinetics and applications involving repetitive sampling over time more difficult. The cheaper models often lack the spectral resolution demanded by some clinical applications (e.g. porphyrin analysis).

3.4 Double-beam Spectrophotometers

In the double-beam design as shown in Figure 7, there is still a single detector – usually a photomultiplier (a phototube). The beam is then sent alternately through sample and reference positions using a chopper wheel or vibrating mirror. The electronics in the instrument are also able to synchronize the beam switching with the detector so that it can distinguish whether the detector is measuring the sample or reference beam at any one time.

Even in complete darkness, a photomultiplier will output a signal. This is known as the dark current. As a result, the beam shuttling device will also have some means of temporarily blocking the light from both beams so that this residual signal (the dark current) can be measured and subtracted from each sample and reference measurement.

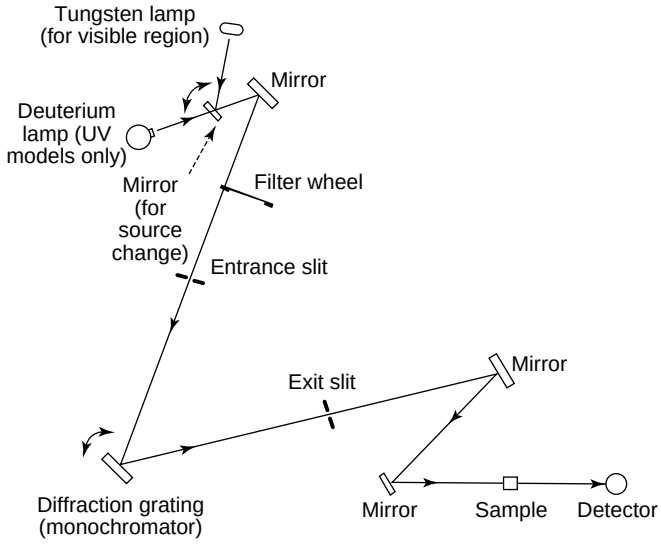

Figure 6 Single-beam optical configuration.

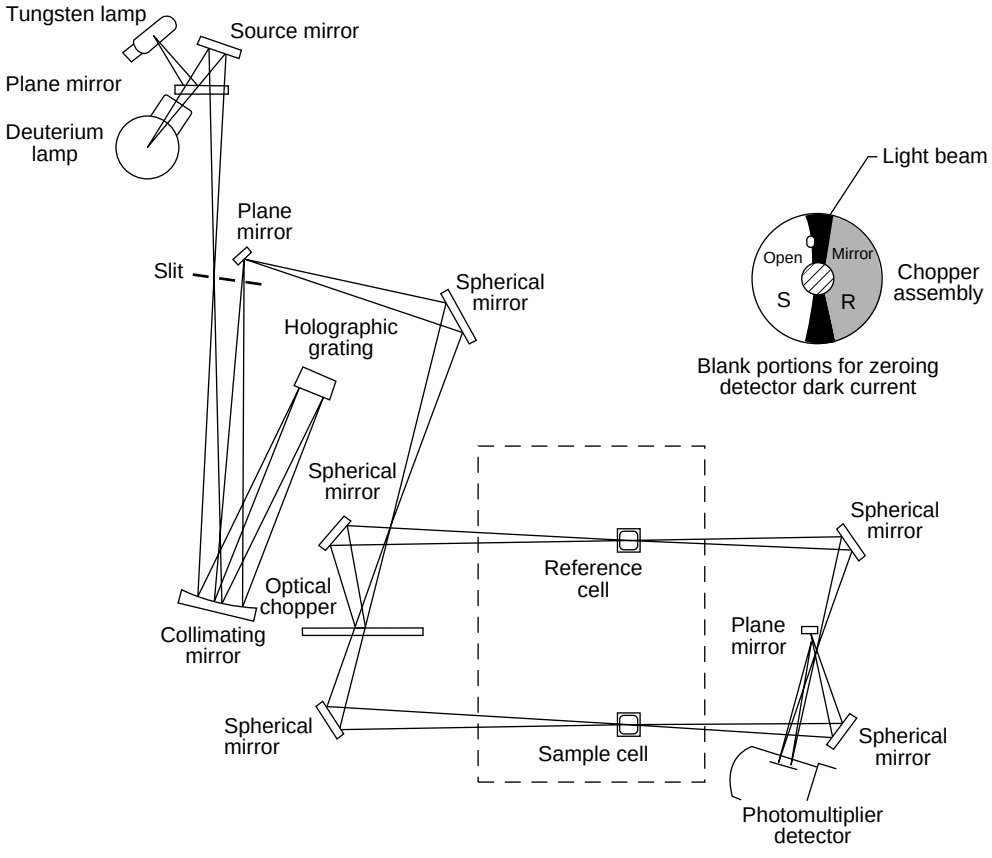

Figure 7 Double-beam (chopper wheel) optical layout.

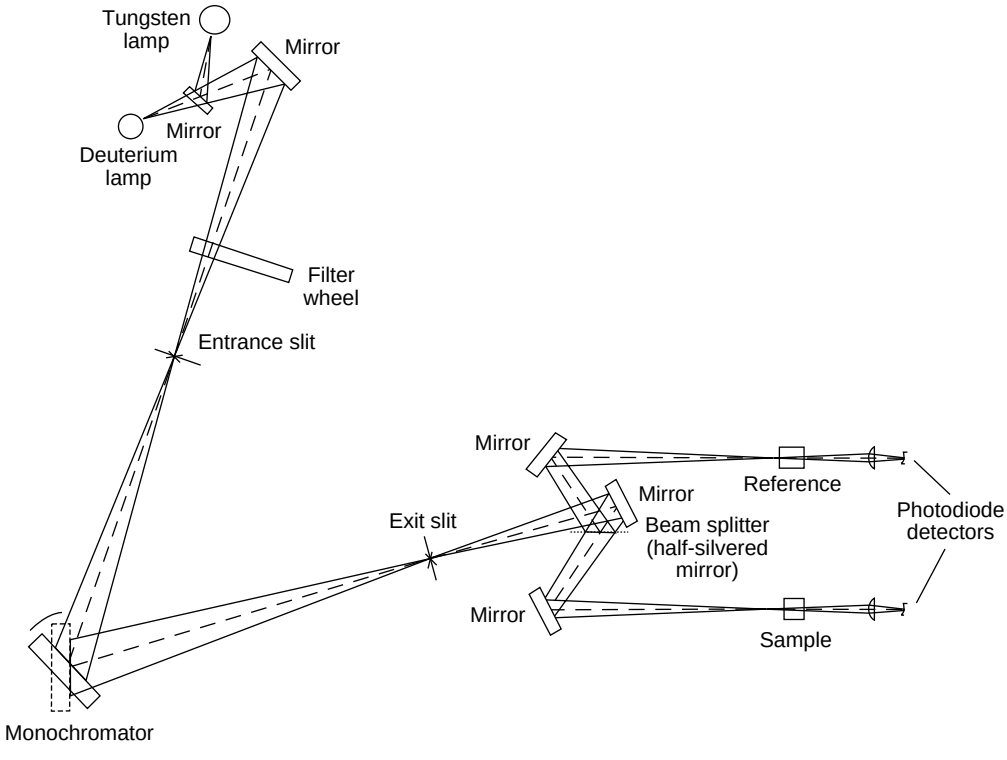

Figure 8 Dual-beam spectrophotometer optical layout.

3.5 Dual-beam Spectrophotometers

An alternative to the double-beam design is the dual-beam layout (Figure 8). In this design, the beam is split using a half-silvered mirror (beam splitter) into its sample and reference components. Each beam has its own detector.

This design is only practical with solid-state detectors (which have fairly constant dark current) and gives equivalent results compared with a double-beam system for most analyses. The design removes the need for a mechanical beam shuttling device (and hence improves reliability) and, as solid-state detectors are used, there is no need for a high-tension power supply to provide the high voltages (around 1000 V) which are required to set the photomultiplier gain. The design also produces very good baseline noise characteristics as, again, there is no contribution from the mechanical beam shuttle. If the design has a weakness, it is at very high absorbance (over 3 AU) as the photomultiplier is able to detect lower light levels than solid-state detectors. Routine clinical analyses rarely exceed these absorbance values and, in any case, a single monochromator instrument is limited by its stray light (see later). If samples exceed the upper absorbance limit of a spectrophotometer, they can either be diluted or a shorter path length cell can be used.

3.6 Photodiode-array Spectrophotometers

Photodiode-array spectrophotometers have been available since the early 1980s with the advent of the early designs from Hewlett-Packard (now Agilent) (Figure 9). The design is basically a single-beam instrument but the sample is irradiated with white light (as opposed to monochromatic light in a dispersive instrument). This light, after it has passed through the sample, is then dispersed by a dispersion monochromator (often referred to as a polychromator) on to a special solid-state detector with individual segments, one for each wavelength. The main appeal of these instruments is that the measurement of a spectrum takes only a few seconds. The system has some disadvantages, however. A diode-array spectrophotometer is a single-beam instrument (although a nonanalytical wavelength may be used as a pseudo-reference to overcome nonwavelength-specific drift). It is also less suitable for some single-wavelength measurements as a whole spectrum has to be collected irrespective of whether the data points are required or not. The design is also more prone to errors from sample fluorescence (as the sample irradiation is at all wavelengths and so any fluorophore present will be also excited) and any nonparallel surfaces in the sample (as this will affect the beam dispersion on to the individual elements in the array). Nevertheless, the diode-array system offers a high throughput of scanned data and the ability also to visualize whole spectrum for single-wavelength analyses so that any unexpected results can be investigated further.

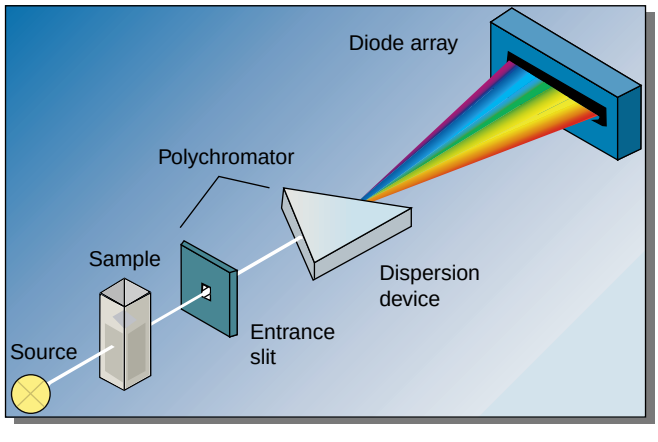

Figure 9 Diode-array spectrophotomer optical layout.

3.7 Microplate Reader Spectrophotometers

Over the past 20 years, many clinical analyses have been transferred from the traditional, cuvette-based spectrophotometer to a microwell (normally 96-well) format. The microwell started life in the early 1960s for microbiological culture. Later, it was realized that the format could be applied to bulk analysis using clear microplates and a dedicated reader.

Microplates are manufactured from plastic (usually polystyrene or acrylic, depending on the required wavelength range). They are cheap, disposable and provide a universal format. The plates also have the advantage that they are compatible with liquid handling devices such as plate fillers and washers.

Although 96 wells is the most common format, the requirement for greater speed and throughput has seen the introduction of even higher density formats such as 384 wells per plate.

The dedicated reader is really a spectrophotometer in an applied form. Most plate readers work in the visible region only (as the polystyrene microplates absorb in the UV region) and use optical filters rather than monochromators. More sophisticated readers may also offer multiple reading modes such as fluorescence and bioluminescence in addition to absorbance. Many modern readers can also be used in conjunction with a robotic system.

In terms of performance, there is a compromise compared with using a spectrophotometer, but for many assays this is far outweighed by the reduced costs and greater convenience which this format offers.

3.8 Reflectance-based Analyzers

In recent years, reagent manufacturers have investigated ways of making tests even more simple to perform and less dependent on expensive analytical equipment. An example of this is the portable glucose monitoring systems which diabetics use to check their blood sugar levels.

The color reaction is performed on a solid support which is coated with the required reagents. This support is usually a small stick or strip. After exposure to a blood or urine sample, this support is inserted in the analytical device and the reading displayed. These devices are based on reflectance rather than absorbance (as the supports are opaque it is not possible to pass light through them).

When light falls on a surface it can be reflected in two main ways, as either specular or diffuse reflectance. Specular reflectance (from the Latin word *speculum* for mirror) is the study of mirror-like reflectance. The path of the reflected ray of light is entirely predictable as it should be at the same angle as the angle of light incidence. In diffuse reflectance, the surface has a matt surface which scatters the reflected light in all directions and this scattered reflectance can be collected by a detector. In practice, a sample may exhibit both types of reflectance (e.g. a coating with a glossy surface) and so the geometry of the collection sphere can be adjusted either to include or exclude the specular component. Reflectance is normally expressed as a percentage compared with a standard – normally a white surface such as Teflon®. The relationship between reflectance and concentration is much less clearly defined when compared with absorbance as there are physical factors to consider (e.g. particle size, layer thickness). A fairly accurate quantitative result can be obtained by taking the logarithm of reflectance or by using the Kubelka–Munk equation. Manufacturers of reflectance-based devices need to spend considerable development work in providing an accurate calibration which can be stored inside the instrument to produce the correct results. These results will need to be verified by regulatory bodies such as the US Food and Drug Administration (FDA) before such a device can be marketed.

As many of these devices are designed for use in the home or general practitioner's surgery, they must be manufactured as cheaply as possible. As a result, an LED can be used in place of a light source and monochromator. Detection is by means of a silicon photodiode or other solid-state detector. A diagram of a reflectance-based analyser for the measurement of blood sugar for diabetes monitoring is shown in Figure 10.

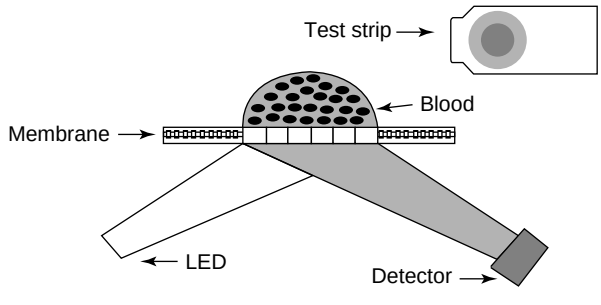

Figure 10 Principle of reflectance-based hand-held analyzer for diabetes monitoring.

4 COMMON CLINICAL APPLICATIONS USING ULTRAVIOLET/VISIBLE ABSORPTION SPECTROSCOPY

There are many clinical tests which employ UV and visible spectrophotometry. For readers who want a detailed and exhaustive list, there are many textbooks on the subject.[1] In this section I have highlighted a range of these tests and detailed how they work.

4.1 Enzyme Rate Assays

Enyzmes are biocatalysts, which are extremely efficient in converting their chosen substrates into product. Apart from their efficiency, they are also highly specific and often will not work with a slightly different substrate.

Often it is useful to study the rate of this catalysis by measuring either the rate of depletion of substrate or the formation of products. This may be a matter of simply measuring the absorbance of one of the reaction components directly or by forming an absorbing conjugate with another molecule.

Enzyme kinetics are usually zero order. This means that, after an initial lag phase, there should be a linear relationship between substrate (or product) concentration with time until one of these components becomes limiting. It is therefore useful if the instrument software is able to allow the user to choose where this linear portion is (either prior to or, better, after data collection) and to use this portion to calculate the rate. The enzyme activity is normally expressed in International Units (IU) by applying a simple factor to the measured slope (absorbance/time).

Enzyme rate assays are nearly always performed at a single wavelength (340 and 405 nm are commonly used) and require a temperature-controlled environment. This may take the form of a simple water-jacketed cell holder where water (or other liquid) is passed through the water jacket at a constant temperature (supplied by a thermostated water circulator). Some cell holders use thermoelectric (or Peltier effect) cell holders. These cell

holders have the advantage of much more precise temperature control and the ability to work below ambient temperatures (Peltier cells can cool by reversing the electric current flow). Some Peltier designs, particularly those requiring high or low temperatures, will still require a flow of cold water in order to operate correctly whereas those covering a more restricted range do not.

Most enzyme reactions are fairly slow, taking place for 5 min or more. In order to increase productivity, most UV/VIS instruments offer a cell changer as an accessory. This is a shuttle device, which can hold six or more cuvettes at once. The instrument then cycles through each of the cell positions taking a reading on each cell every 30 s during the course of the reaction.

The collected data can be analyzed either using the instrument's own kinetic software or externally, using either a computer or manual calculation. Some statistical data on the quality of the curve fit are also useful.

An example of a clinical rate assay is the determination of butyrylcholinesterase (BchE). Certain individuals express a mutant form of the BchE gene. This then encodes for a defective form of the enzyme, which lacks the ability to hydrolyze succinylcholine. In some rare cases, the complete BchE gene is missing. A defective or missing gene will not, normally, be of any consequence. If, however, succinylcholine is used during tracheal intubation in the administration of inhalation anesthetics, this will then cause the patient to undergo complete paralysis. The test for this enzyme[2] is commonly performed using a UV/VIS spectrophotometer with a temperature-controlled cell holder (most tests will be performed at 37 °C).

One important class of enzyme-catalyzed reactions involves the oxidation and reduction of pyridine nucleotides [nicotinamide adenine dinucleotide (NAD^+) and nicotinamide adenine dinucleotide (reduced form) (NADH), respectively]. If the reaction is followed at 340 nm, NAD^+ does not absorb whereas NADH shows a strong absorbance. Examples of NAD^+/NADH kinetic reactions include glucose dehydrogenase,[3] aspartate aminotransferase[4] and urea.[5]

Other enzymatic tests often include a colorless compound, which is added to the reaction. This is then enzymatically converted to a colored product (e.g. *p*-nitrophenol). Examples of this type of test include alkaline phosphatase,[6–8] acid phosphatase[4] and amylase.[1] These tests are included in the summary of common clinical tests in Table 3.

4.2 Colorimetric and End-point Assays

It is also possible to perform a reaction to completion and then to calculate the quantity of initial substrate by extrapolation. This type of assay is termed an end-point assay. There are a vast number of such tests and, if an exhaustive list is sought, the reader is directed to a compendium of methods such as *Tietz*.[1] A summary of some common clinical end-point analyses is given in the summary of common clinical tests in Table 3. The methods will prescribe a time after which the reading can be taken. In the case of enzyme-catalyzed reactions, the incubation time will need to be at constant temperature.

Most diagnostics companies provide premixed reagents to perform these tests together with a detailed protocol describing their use. Different protocols may use slightly

Table 3 Some common clinical tests

Analyte	Method	UV wavelength (nm)
Acid phosphatase[4]	Kinetic: PNP	405
Alanine aminotransferase[9]	Kinetic: NAD^+/NADH	340
Alkaline phosphatase[6–8]	Kinetic: PNP	405
α-Amylase[1]	Kinetic: 2-chloro-4-nitrophenylmaltoheptaoside 2-chloro-PNP	405
Aspartate aminotransferase[10]	Kinetic: NAD^+/NADH	340
Bilirubin[11]	Evelyn–Molloy	555
Cholesterol[12]	Kinetic: cholesterol oxidase	500
Creatinine[13]	Jaffe	510
GGT[14]	Kinetic: carboxy substrate	405
Glucose[15]	Kinetic: hexokinase (NAD^+/NADH)	340
LDH[16]	Kinetic: lactate/pyruvate	340
Porphyrins (total)[17]	Acidification using HCl	Absorbance ~ 405 (Soret peak)
Pseudocholinesterase[28]	NAD^+/NADH	340
Triglycerides[18,19]	Kinetic: GPO colorimetric	520
Urea[20]	Kinetic: NAD^+/NADH	340

PNP, *p*-nitrophenyl phosphate; GGT, γ-glutamyl transferase; LDH, lactate dehydrogenase; GPO, glycerol-3-phosphate oxidase.

altered wavelengths for each analyte. As with rate assays, the test may be based on reading the analyte directly or an additional color reaction may be required to develop the test.

4.3 Immunoassays, Enzyme-linked Immunosorbent Assays and Microplate Assays

Immunoassays rely on the very strong affinity between an antibody and its target molecule (antigen). This strong affinity has been used to design a wide range of assays for a variety of targets. These include pathogens (where a chemical associated with the organism or virus will be targeted), tumor markers and drug monitoring (either therapeutic or drugs of abuse). In fact, anything which can elicit an antibody response can be developed into an immunoassay. These assays are ideal where high specificity is required. They are not good at dealing with a range of possible compounds. For example, it is difficult to design an immunoassay to detect all abused drugs, only specific types. If a ''designer'' drug is synthesized with an additional functional group, this may adversely affect the assay.

Immunoassays are competitive assays. The reaction mixture will contain antigens labeled with some kind of tag (this is either a radiolabel, a fluorophore or a site which will bind to a chromophore). When the (unlabeled) sample antigens are introduced, there will be competition between the labeled and unlabeled antigens and this can be calibrated against a binding curve (which will be sigmoidal in nature). Prior to measurement, it will be necessary to separate the free and bound antigens and so a separation step will be required (heterogeneous assay). Much effort has been spent in previous years trying to simplify this process (e.g. using magnetized, latex-covered beads) or using a technique which inherently separates bound from unbound antigens (e.g. fluorescence polarization) to produce a homogeneous assay. For the highest sensitivity immunoassays, either radioactivity or fluorescence (or time-resolved fluorescence) has to be used. In situations where lower sensitivity is acceptable, an absorption-based assay can be employed.

A popular assay type is the enzyme-linked immunosorbent assay (ELISA). In this assay, an antibody (specific to the analyte) is coated on the bottom of the microplate. In some cases, any exposed area on the microplate (i.e. any area where the antibody has not been coated) is blocked using bovine serum albumin (BSA) and the excess antibody is washed away. The blocking agent is to help prevent unbound antigen from adhering directly to the plate, thus affecting the final result.

The analyte and enzyme conjugate [e.g. horseradish peroxidase (HRP)] is added. The analyte will bind to the antibody, which has previously been coated on the microplate. The enzyme conjugate also binds to the analyte. This will be used later in conjunction with a specific dye to produce the color, which will then be measured on the reader.

4.4 Porphyrin Analysis

Porphyria is the name given to a number of related conditions, some genetic, which result from an over production of porphyrins, which are precursors in the production of hemoglobin. Sufferers are prone to bouts of severe abdominal pain, vomiting, severe personality changes and sensitivity to light. Some types of porphyria produce characteristic dark-colored (''port wine'') urine and this is a good first indication that porphyria is present.

One of the main types of porphyria (a generic term for porphyrin-related disease) is variegate porphyria. In this case a mitochondrial enzyme called protoporphyrinogen oxidase is defective (owing to incorrect genetic coding) and as a result excess protoporphyrin (one of the porphyrin types) is produced. The protoporphyrinogen reacts with oxygen to produce protoporphyrin in an uncontrolled reaction. Other porphyria types include acute intermittent porphyria. Accurate diagnosis of the exact type of porphyria is vital as incorrect treatment could have very serious consequences.

Porphyrins[17] have a characteristic UV/VIS absorption peak (Soret peak) in the region 400–410 nm depending on the type of porphyrin present (coproporphyrin has a peak between 402 and 403 nm whereas the uroporphyrin peak lies between 406 and 407 nm).

The urine sample is filtered, diluted with distilled water and acidified with hydrochloric acid. The sample is scanned between 300 and 500 nm and the spectrum peak positions are noted. The measurement is taken by first constructing a baseline at points at either side of the main peak (usually around 380 and 430 nm) and then measuring the height of the peak down to this baseline (Figure 11).

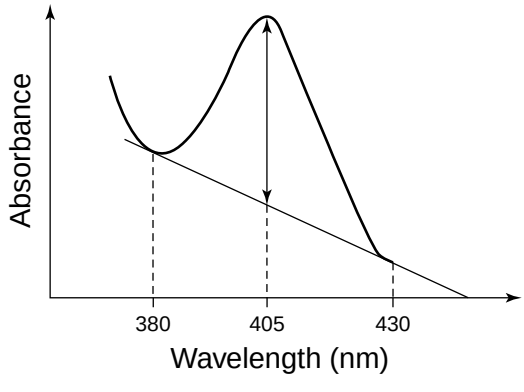

Figure 11 Porphyrin analysis using three wavelengths.

The total porphyrin concentration is given by Equation (4):

$$\text{total porphyrin}(\mu g\, L^{-1}) = 2A(\lambda_{max}) - (A_{380} + A_{430})\varepsilon \quad (4)$$

where A is absorbance, λ_{max} is the wavelength at maximum absorption and ε is the molar absorptivity (a constant) for the analyte; for porphyrins (in a 1-cm cell) $\varepsilon = 4740\,\mu g\, L^{-1}$.

Alternatively, second-derivative spectroscopy has been used[1] (as this reduces background effects and produces sharper peaks). It should be borne in mind that, for a full and correct diagnosis, the type of porphyrin must be accurately identified. This can only be done using a good UV/VIS spectrophotometer offering narrow slits and a skilled user, as interpretation of the corrected spectrum or the second-derivative spectrum may be involved. UV/VIS absorption spectroscopy is generally used for screening and other techniques such as high-performance liquid chromatography (HPLC) or fluorescence spectroscopy (which gives much better selectivity as each porphyrin has a different excitation and emission wavelength) are often used to make the final, confirmatory, diagnosis. These techniques are usually offered by porphyria reference centers.

4.5 Hemoglobin Analysis

Hemoglobin is a protein with a nonprotein core consisting of an iron atom surrounded by heme groups. It has remarkable oxygen transportation properties where it can change its conformation to accept oxygen (oxyhemoglobin). This process can be inhibited by carbon monoxide, which has a 200 times stronger affinity for hemoglobin (carboxyhemoglobin) than oxygen, resulting in severe respiratory problems and death in cases of carbon monoxide poisoning. Hemoglobin possesses an iron atom core in its ferrous (Fe^{2+}) state. If the iron is oxidized to its ferric (Fe^{3+}) state, its oxygen transport capabilities are diminished and the molecule is called methemoglobin (metHb).

Total hemoglobin can be measured by performing a reaction of the total hemoglobin present with potassium cyanide to form a hemoglobin–cyanide complex (Figure 12).

This is the basis of the method of Zijlstra et al.[20] In this method, hemoglobin is reacted with potassium cyanide (taking great care with pipetting and certainly never by mouth!) and sodium (or potassium) hexacyanoferrate(III) to produce a hemoglobin–cyanide complex (Equation 5):

$$\text{hemoglobin} + CN^- + Fe(CN)_6{}^{3-} \longrightarrow \text{hemoglobin–cyanide} \quad (5)$$

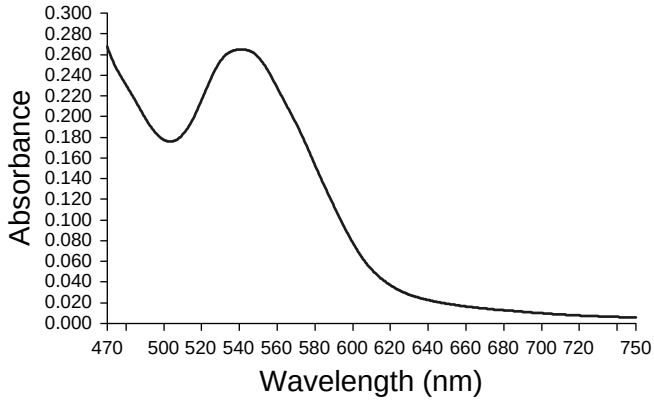

Figure 12 Absorption spectrum of hemoglobin–cyanide complex.

This complex has a peak around 546 nm, which can be measured and quantitated.

Franzini et al.[21] have described a method based on second-derivative spectroscopy to measure hemoglobin and its homologs. This method helps to overcome the interference from bilirubin, which is also often present in these analyses.

Shih et al.[22] also used multicomponent analysis in order to quantitate carboxy-, met- and oxyhemoglobin independently from a single scan. This is a statistical approach which compares each data point in a spectrum with a calibration set of known references – either mixtures or single components – and attempts to calculate the relative proportions of each in the unknown spectrum. This approach has its attractions as it is potentially able to measure each component without the need for a separation step. The main drawback is when there is significant spectral overlap or in situations where the concentrations of the various components differ widely. Problems may also occur if there is a strong background matrix which is not constant. More powerful techniques exist, such as principle component regression or partial least-squares fitting. These techniques are a science in themselves (chemometrics) and go beyond this article.

4.6 Protein Assays

Proteins are composed of amino acid building blocks. Protein has some intrinsic absorbance at approximately 280 nm (from the aromatic amino acids – tyrosine and tryptophan) but it is more common to perform a reaction to produce a colored complex which can be assayed in the visible region. The three most common procedures for protein analysis are biuret, Bradford and bicinchoninic acid (BCA) assays.

Table 4 lists the various protein methods. These are all simple colorimetric determinations.

Table 4 Summary of protein assays

Assay method and reference	Principle	Range ($\mu g\,mL^{-1}$)	Interferences
Lowry[23]	As biuret plus determination of aromatic amino acids	5–200	Phenols, aromatic amino acids
Bradford[24]	Dye binding with coomassie brilliant blue	10–200	Detergents
Biuret[25]	Determination of number of peptide bonds	200–5000	Amines, ammonium salts
BCA[26]	Reduction of copper by protein and formation of Cu(I)–BCA complex	200–1200	High concentrations of metals, strong reducing agents, chelating agents
Absorbance at 280 nm[23]	Determination of tyrosine and tryptophan in protein	>50	Nucleic acids, phenols, aromatics
Warburg–Christian[27]	Determination of aromatic amino acids with compensation for nucleic acids	50–3000 purity check	Phenols, aromatics

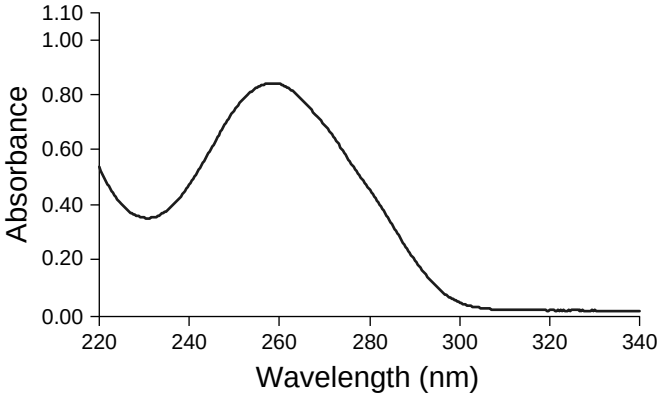

Figure 13 Absorption spectrum of DNA.

4.7 Molecular Biology

The use of UV/VIS spectroscopy for molecular biology strictly falls outside the scope of clinical analysis. On the other hand, more routine clinical laboratories are using molecular biology techniques (such as the PCR and automated dideoxy sequencing) as these give more direct diagnosis for genetically based disorders.

As a result of these techniques, UV/VIS spectroscopy is useful for assessing the purity of the starting template in either a sequencing or PCR reaction and, as a result, saves time in optimizing the reaction and helps to reduce reagent costs. Pure DNA and RNA absorb at 260 nm (Figure 13). Protein (which is the main source of contamination) absorbs at around 280 nm (this is dependent on the exact amino acid composition as individual amino acids have slightly different absorption maxima). There is a rule of thumb which is now widely adopted in molecular biology laboratories as a quick method for assessing purity. This method measures the absorbance at 260 and 280 nm and the ratio (A_{260}/A_{280}) is calculated. Optionally, a third reference point can be taken at 320 nm (to assess the amount of turbidity and scatter) and this absorbance value can be subtracted from either of the two absorbance values prior to calculating the ratio. If the ratio of the two absorbances is between 1.7 and 2.0, then the DNA preparation is considered to be pure. If it is >2, then there is probably a high RNA content. A lower ratio would indicate a high protein or phenol content (a reagent commonly used in DNA extraction).

One major issue in molecular biology is sample volume. It is not uncommon to have volumes of 10 μL and so special low-volume cells are available. These special cuvettes are made from quartz (as plastic and glass cuvettes generally absorb below 300 nm). Dedicated low-cost instruments are available for the assessment of nucleic acid purity.

ACKNOWLEDGMENTS

I thank Hanswilly Müller at PerkinElmer, Überlingen, Germany, for providing some of the spectra shown here, Jackie Woolf at the Porphyria Reference Centre, Heath Park Hospital, Cardiff, and Ipswich Hospital, UK, for some information and spectra of porphyrins and Chris Royle at the Brompton Hospital, UK, for keeping me up to date with the latest developments in clinical analysis. I also thank Agilent Technologies (formerly Hewlett-Packard Instruments) and Hypoguard for permission to use their diagrams.

ABBREVIATIONS AND ACRONYMS

BchE	Butyrylcholinesterase
BSA	Bovine Serum Albumin

CRM	Certified Reference Material
ELISA	Enzyme-linked Immunosorbent Assay
FDA	Food and Drug Administration
GGT	γ-Glutamyl Transferase
GPO	Glycerol-3-phosphate Oxidase
HPLC	High-performance Liquid Chromatography
HRP	Horseradish Peroxidase
IR	Infrared
LDH	Lactate Dehydrogenase
LED	Light-emitting Diode
metHb	Methemoglobin
NAD^+	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
NIST	National Institute of Standards and Technology
PCR	Polymerase Chain Reaction
PNP	*p*-Nitrophenyl Phosphate
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Glucose, In Vivo Assay of • Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy in Microbiology • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis • Glycolipid Analysis • Monosaccharides and Sugar Alcohol Analysis

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Diagnostic Hematology • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Immunochemistry • Infrared Spectroscopy in Clinical Chemistry • Laboratory Instruments in Clinical Chemistry, Principles of • Nucleic Acid Analysis in Clinical Chemistry • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Serum Proteins

REFERENCES

1. C. Burtis, E.R. Ashwood (eds.), *Tietz Textbook of Clinical Chemistry*, 2nd edition, W.B. Saunders, Philadelphia, 1994.
2. R.T. Evans, J. Wroe, 'Is Serum Cholinesterase Activity a Predictor of Succinylcholine Sensitivity? An Assessment of Four Methods', *Clin. Chem.*, **24**, 1762–1766 (1978).
3. D. Barham, P. Trinder, 'An Improved Colour Reagent for the Determination of Blood Glucose by the Oxidase System', *Analyst*, **97**, 142–145 (1972).
4. Deutsche Gesellschaft für Klinische Chemie, *Z. Clin. Chem. Klin. Biochem.*, **8**, 658 (1970); **9**, 464 (1971); **10**, 182 (1972).
5. H. Talke, G.E. Schubert, *Klin. Wochenschr.*, **42**, 172 (1965).
6. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, VCH, Weinheim, 1986.
7. 'Biochemica Information', Boehringer Mannheim, Mannheim, 1975.
8. W.F. Gohara, 'Rate of Decrease of Glutamyltransferase and Acid Phosphatase Activities in the Human Vagina after Coitus', *Clin. Chem.*, **26**, 254–257 (1980).
9. Expert Panel on Enzymes, International Federation of Clinical Chemistry, 'IFCC Methods for the Measurement of Catalytic Concentrations of Enzymes', *Clin. Chim. Acta*, **105**, 147F–154F (1980).
10. R.J. Henry, N. Chiamori, O.J. Golub, S. Berkman, 'Revised Spectrophotometric Methods for the Determination of Glutamic–Oxaloacetic Transaminase, Glutamic–Pyruvate Transaminase and Lactic Acid Dehydrogenase', *Am. J. Clin. Pathol.*, **34**, 381–398 (1960).
11. J.B. Henry (ed.), *Clinical Diagnosis and Management by Laboratory Methods*, 18th edition, W.B. Saunders, Philadelphia, 1991.
12. G.R. Cooper, P.H. Duncan, J.S. Hazelhurst et al., 'Cholesterol, Enzymatic Method', in *Selected Methods of Clinical Chemistry: Selected Methods for the Small Clinical Chemistry Laboratory*, eds. W.R. Faulkner, S. Meites, American Association for Clinical Chemistry, Washington, DC, 1982.
13. R.J. Henry, D.C. Cannon, J.W. Winkelman (eds.), *Clinical Chemistry: Principles and Techniques*, 2nd edition, Harper and Row, Hagerstown, MD, 1974.
14. S.B. Rosalki, 'Gamma Glutamyl Transpeptidase', *Adv. Clin. Chem.*, **17**, 53–107 (1975).
15. 'Chemstrip 10 Package Insert', Boeringer-Mannheim, Indianapolis, IN, 1992.
16. D. McKenzie, A.R. Henderson, 'Electrophoresis of Lactate Dehydrogenase Isoenzymes', in *Standard Methods of Clinical Chemistry*, ed. G.R. Cooper, American Association for Clinical Chemistry, Washington, DC, Vol. 10, 1983.
17. G.H. Elder, S.G. Smith, S.J. Smyth, 'Laboratory Investigation of the Porphyrias', *Ann. Clin. Biochem.*, **27**, 395–412 (1990).

18. G.R. Warnick, 'Enzymatic Methods for Quantification of Lipoprotein Lipids', *Methods Enzymol.*, **129**, 101–123 (1986).
19. M.W. McGowan, J.D. Artiss, D.R. Strandbergh, B. Zak, 'A Peroxidase-coupled Method for the Colorimetric Determination of Serum Triglycerides', *Clin. Chem.*, **29**, 538–542 (1983).
20. W.G. Zijlstra, A. Buursma, W.P. Meeuwsen-van der Roest, 'Absorption Spectra of Human Fetal and Adult Oxyhemoglobin, De-oxyhemoglobin, Carboxyhemoglobin and Methemoglobin', *Clin. Chem.*, **37**, 1633 (1991).
21. C. Franzini, G. Cattozzo, A. Pagani, 'Serum Hemoglobin Measurement by Second Derivative Spectroscopy', *Int. Lab.*, October, 33–39 (1988).
22. M.L. Shih, W.D. Korte, C.R. Clark, 'Multicomponent Spectroscopic Assay for Hemoglobin and Ferrihemoglobin Species in Methemoglobin Treatment of Cyanide Poisoning', *J. Anal. Toxicol.*, **21**(7), 543–547 (1997).
23. E. Layne, *Methods Enzymol.*, **3**, 447–454 (1957).
24. M.M. Bradford, 'A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein–Dye Binding', *Anal. Biochem.*, **72**, 248–254 (1976).
25. A.G. Gornall, C.S. Bardawill, M.M. David, *J. Biol. Chem.*, **177**, 751 (1949).
26. C.V. Sapan, R.L. Lundblad, N.C. Price, 'Colorimetric Protein Assay Techniques', *Biotechnol. Appl. Biochem.*, **29**(2), 99–108 (1999).
27. O. Warburg, W. Christian, *Biochem. Z.*, **310**, 384 (1941).
28. J.R. Huizenga, C.H. Gips, 'Evaluation of the UV-340 Spectrophotometric Determination for Pseudocholinesterase Activity (EC 31.1.8) in Human Serum', *J. Clin. Chem. Clin. Biochem.*, **25**(3), 161–165 (1987).

Urinalysis and Other Bodily Fluids

Helen M. Free and Alfred H. Free (deceased)
Bayer Corporation, Elkhart, USA

1 Introduction 1
2 History of Urine Testing 1
2.1 Ancient Times 2
2.2 Microscopic Examination of Urine Sediment 2
2.3 Chemical Tests 2
2.4 Convenient Urine Tests 3
2.5 Dip-and-read Reagent Strip Tests 3
2.6 Automation 3
3 Evolution of Urine and Blood Testing Using Diabetes Tests as the Example 4
4 Most Common Urine Analytes 5
4.1 Glucose 5
4.2 Ketones 5
4.3 Protein 5
4.4 Occult Blood 5
4.5 Bilirubin 6
4.6 Urobilinogen 6
4.7 Nitrite 6
4.8 Leukocyte Esterase 6
4.9 pH 6
4.10 Specific Gravity 7
4.11 Clinitek Microalbumin 7
4.12 Human Chorionic Gonadotropin 7
5 Less Common Urine Analytes 7
5.1 Inborn Errors of Metabolism 7
5.2 Drugs of Abuse 7
6 Other Bodily Fluids 8
6.1 Serum or Blood 8
6.2 Sweat 8
6.3 Cerebrospinal Fluid 8
6.4 Amniotic Fluid 8
6.5 Tears or Saliva 8
6.6 Stools 8
6.7 Miscellaneous Bodily Fluids 8
7 Where We are in the Year 2000 8
8 The Future – Into the New Millennium 9
Trademarks 9
Abbreviations and Acronyms 9
Related Articles 9
References 9

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The chemical analysis of urine and other body fluids is an integral part of laboratory medicine today. Clinical chemistry measurements are made in laboratories ranging from large medical research/teaching centers to small rural physicians' offices, and with varying degrees of sophistication in all parts of the world. Indeed, the use of chemical tests at home is currently an important part of monitoring chronic conditions such as diabetes, and this aspect of chemical testing is expected to expand.

This article traces the historical methods of urine testing to provide information on health and disease, from crude visual examination in ancient times to modern techniques using sophisticated chemical instrumentation. The information obtained is useful in the diagnosis of disease, in monitoring the course of the treatment of disease, and in helping to maintain wellness in normal populations.

The chemical tests most widely used in urinalysis are described in some detail along with a discussion of the development of commercial tests leading to the evolution of the diagnostics industry. Chemical tests on other bodily fluids are briefly summarized.

1 INTRODUCTION

Urinalysis has helped throughout the ages to diagnose disease. We briefly review the history of its development, from primitive urinalysis consisting mainly of observation of specimens to modern uses of chemical reagents and automated instrumentation. An older but more detailed presentation is found in our book, *Urinalysis in Clinical Laboratory Practice*.[1] With the creation of the diagnostics industry, the application of analytical chemistry to the diagnosis of disease has allowed the clinical laboratory to provide more precise, more accurate, more useful and more rapid information to the physician. Moreover, chemistry allows the patient to monitor chronic conditions such as diabetes directly and thus assume more responsibility for his or her own health. Because the analysis of urine does not require invasive specimen collection procedures, the immediate future appears bright for its continuing utility. It also seems that the kidney's role in excreting a variety of metabolites assures urinalysis a continuing place in analytical chemistry.

2 HISTORY OF URINE TESTING

The practice of using information from urine testing to indicate disease or abnormalities begins in ancient times,

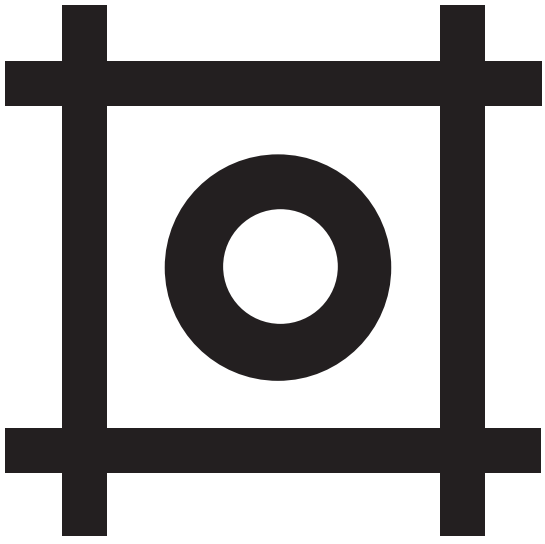

Figure 1 An ancient symbol for urine, as one of the elements of nature.

continues through the present time and is likely to be expanded into the foreseeable future.

2.1 Ancient Times

Before there was written language, symbols were used for important natural elements and shown in Figure 1 is the symbol used for urine.[2] As early as 400 BC Hippocrates noted the different odors and color of urine. About 1000 AD, Ismail of Jurjani recorded seven observations which could be made on urine for various interpretations. These were: quantity, consistency, color, odor, transparency, sediment, and froth.

Visual observations continued, and in the Middle Ages physicians were known as uroscopists and were nearly always painted peering into a container of urine – the urine glass that is seen in many engravings of the 13th to 15th centuries. This was a round-bottomed flask called a matula, and is depicted in Figure 2. Uroscopy later developed a poor reputation because charlatans, or uromancers, used the examination of urine not only to provide information about health and disease, but also as sort of a fortune-telling exercise to foretell the future. Such quacks were known as ''Pisse Prophets''. A review of the role of urinalysis in the history of diagnostics medicine gives many more details of urine study before chemical analysis of urine was used.[3]

2.2 Microscopic Examination of Urine Sediment

Of course, the visual examination of urine continues today with the microscopic examination of urine sediment. Chemical stains are often used to enhance the differentiation of the various elements formed. The addition of acetic acid to the sediment causes red cells to hemolyze and aids in their identification. Knowledge of the urine pH is also useful in the identification of sediment. Chemical tests are often used to identify various crystals in urine.

Figure 2 The painting *The Physician* by Franz Christophe Janneck showing the matula, the urine collection vessel used by physicians of the time. (Reproduced by permission of Fisher Scientific.)

2.3 Chemical Tests

In the 1600s, Scribonius described the black urine in alcaptonuria, Frederick Dekker of Leyden detected protein in urine using heat and acetic acid, Herman Boerhaave measured the specific gravity of urine, and Thomas Willis tasted urine to detect the sweetness of urine caused by diabetes (and it was only a century later in 1776 that Matthew Dobson proved that the sweet taste was due to sugar).

In the 18th century, several chemical tests were developed specifically for use in urine studies. They included yeast fermentation for urine sugar, nitric acid precipitation of protein in urine, and the nitric acid colorimetric bilirubin test made popular by Leopold Gmelin. In 1841, Trommer introduced his test for sugar, which involved the reduction of blue cupric ions in hot alkaline solution; this test was later improved by Herman

von Fehling. Perhaps the most widely known liquid test for urine sugar was described by Stanley Benedict in 1911, when he was a student at the University of Cincinnati.[4] He not only devised Benedict's qualitative solution for the estimation of urine sugar, but also developed a quantitative volumetric method in which the extent of copper reduction in a given volume of Benedict's quantitative solution was carefully determined by titration while the solution was kept boiling over a Bunsen burner. Benedict made many contributions to analytical chemical methods during his career as Professor of Chemistry at Cornell University. Otto Folin of Harvard University Medical School was another famous biochemist who contributed to the early advances in clinical analytical chemistry, and Meites[5] has written his biography. It is our belief that Victor C. Myers, who received his doctorate from Yale University in 1909, created the first real clinical laboratory with the extensive application of chemical measurements to blood and urine at the New York Postgraduate Hospital. This can be considered the beginning of 20th century urinalysis. In 1906, Dr Charles P. Emerson of Johns Hopkins University wrote the textbook *Clinical Diagnosis*,[6] a significant part of which is devoted to urine study. For the first quarter of the 20th century, this text profoundly affected the importance of urine study as viewed by the medical profession.

2.4 Convenient Urine Tests

The idea of providing convenience to the clinical study of urine was first noted by Dr Walter Compton, who with Maurice Treneer (the chief chemist at Miles Laboratories in Elkhart, Indiana), devised a simple tablet containing cupric sulfate, sodium hydroxide, citric acid and a little carbonate.[7] (After all, Miles was the maker of Alka-Seltzer®, which also fizzed!) This test, Clinitest®, is still widely used and when a tablet is dropped into a test tube containing urine, the heat of solution of the sodium hydroxide and the exothermic reaction of sodium hydroxide with citric acid supply enough heat to allow the sugar of the urine to reduce blue cupric ions to orange or red cuprous ions. Color blocks from blue through green to orange represent the amount of color change, and can be used to estimate the sugar concentration of the urine specimen tested. Essentially, this allows the copper reduction test to be performed without an external source of heat.

Our group, the Biochemistry Section of the Miles Ames Research Laboratory, continued to advance the convenience of urine testing by producing other single-use reagent tablets: Acetest® for urine ketones and Ictotest®, the ultrasensitive test for urine bilirubin. Both of these tests are still widely used today even after the introduction of the very convenient ''dipstix'' in 1956.

2.5 Dip-and-read Reagent Strip Tests

At the 1956 International Congress of Clinical Chemistry in New York, the first dip-and-read test was described by our group. This was Clinistix®, a simple specific test for urine glucose. In the same year, Eli Lilly & Co., leaders in supplying insulin to those with diabetes, introduced Tes-Tape® which is based on the same chemical principle. Glucose was, and still is, the most common urine analyte in the clinical laboratory. Urine protein is the second most common test, and the colorimetric test for protein was introduced in 1957. Since these two tests were performed on almost every urine specimen reaching the laboratory, it seemed logical to provide both tests on the same dip-and-read strip, and thus the first multiple reagent dip-and-read test was created, for glucose and protein. The next additions to the multiple reagent strip were tests for pH, occult blood and ketones. In 1969 the liver function tests for bilirubin and urobilinogen were added. In 1972 the nitrite test for urinary tract infections (UTIs) or bacteriuria was added. And in 1981 the colorimetric test for specific gravity finally came into being. In 1984 another indirect test for UTIs, a test for leukocyte esterase, was added as the final test of the ten-way multiple reagent strip. Each of these tests is a colorimetric test, and the reagent is matched to a series of color blocks of increasing intensity, proportional to the amount of analyte being determined. A color chart is shown in Figure 3.

2.6 Automation

In 1972, automation was introduced to the urinalysis laboratory with the large-volume laboratory instrument called Clinilab®. Reagent areas for each test were spaced on reels and fed through the sampler apparatus; the color intensity of the reacted reagent was measured by reflectance spectroscopy. Specific gravity was measured by the falling drop method, in which the time a drop of urine takes to travel between two marks, through a column of viscous fluid, is converted into specific gravity units. The Clinilab® was a ''walk-away'' instrument with automatic printout of results on a data tag attached to the urine specimen tube, so that no recording errors could occur. A semi-automated instrument, a ''strip reader'' called Clinitek®, was introduced for small-volume laboratories and physicians' offices. This required the analyst to dip the reagent strip into the specimen and place the reacted strip on a tray table. Results were displayed on a lighted panel and copied by the analyst. Many improvements have been made over the years, and the latest version of the larger instrument is called the Clinitek® 500, while the smaller instrument is the Clinitek® 50. This latter instrument can also be used with the newest products, Clinitek® hCG, a test

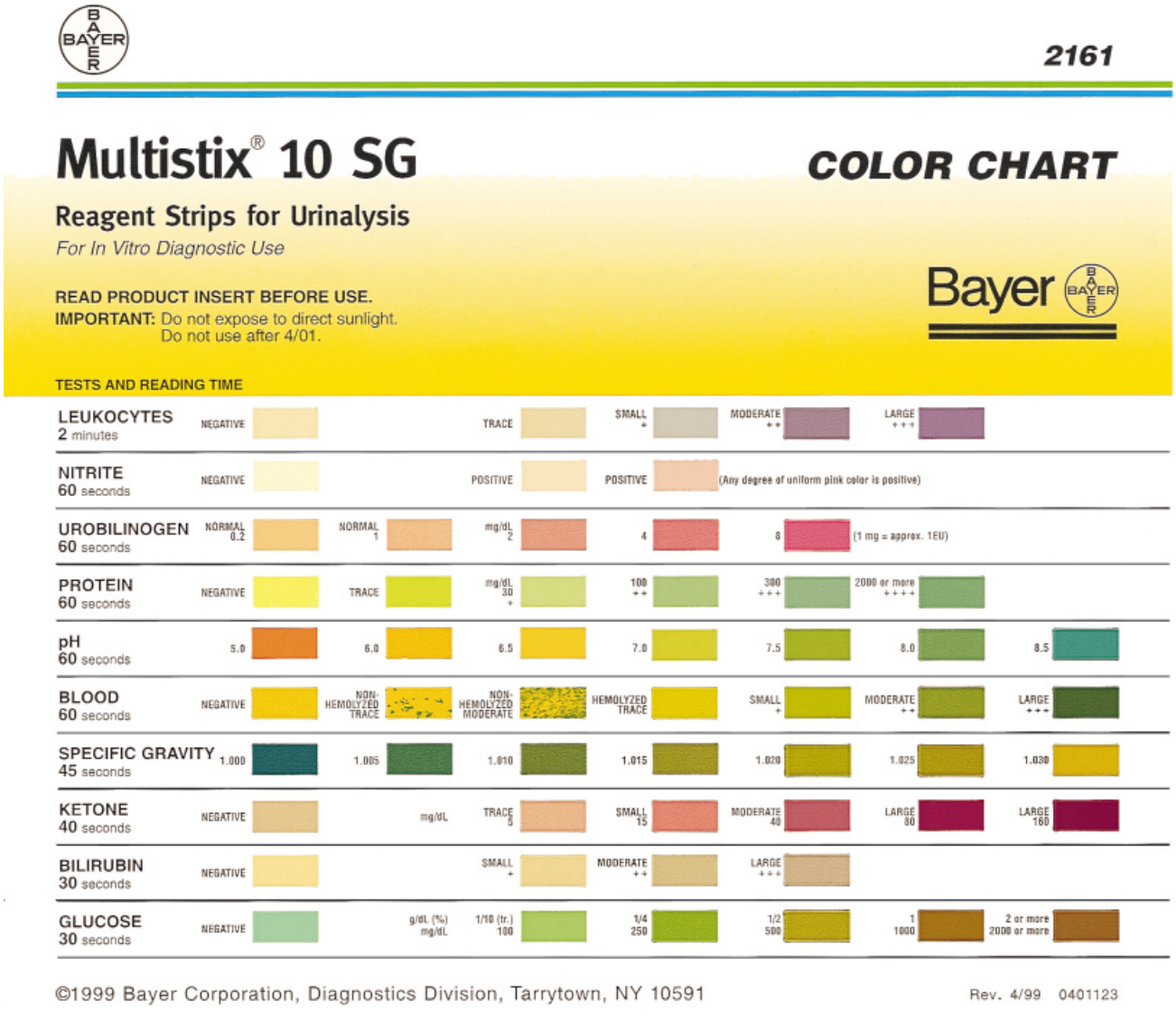

Figure 3 Color chart for the ten-test multiple dip-and-read urine reagent strip. (Reproduced by permission of Bayer Corporation.)

for human chorionic gonadotropin (HCG), and Clinitek® Microalbumin, a test for microalbuminuria. A timeline adapted from a previous publication[8] and noting many of the important milestones in urine testing is shown in Table 1.

3 EVOLUTION OF URINE AND BLOOD TESTING USING DIABETES TESTS AS THE EXAMPLE

In ancient times, urine was poured on the ground to see if it attracted insects. The attraction of insects indicated an abnormality then known as 'honey urine'' which was correlated with what later became known as diabetes. As mentioned in section 2.3, the method of ''taste testing'' was used by Dr Thomas Willis in the 1600s. Metal reduction tests that relied on the fact that glucose is a strong reducing agent to detect and quantitate its level in urine or blood were used in the early 1900s. Copper was the usual reagent but bismuth reduction was also utilized. The development of the convenient copper reduction test in tablet form was the culmination of the reduction tests. The biggest disadvantage of copper reduction tests was the fact that any reducing substance present in the urine specimen would react in the same way as glucose. This led to the use of the enzyme glucose oxidase to produce a test formulation specific for glucose. The same disadvantage of nonspecificity was also

Table 1 Timeline with various milestones in the development of urine testing

Ancient times	Insects – honey urine
400 BC – Hippocrates	Color/odor/fever
1000 AD – Ismail of Jurjani	Seven observations
Middle Ages – Uroscopy	''Pisse Prophets''
1673 – Dekkers	Boiling acetic acid – protein
1674 – Willis	Sweet taste – diabetes
1674 – Boerhaave	Specific gravity – dehydration
1827 – Bright	Heat/acid/protein – renal
1911 – Benedict	Alkaline Cu reduction – sugar
1942 – Compton	Self-heating tablet – Clinitest®
1956 – Free et al.	''Dipstix'' – Clinistix®
1958 – Ames	First multiple – two tests
1972 – Ames/Miles	Automation
1984 – Miles	Ten-test multiple
1990 – Miles	Clinitek® 100
1997 – Bayer	Clinitek® Microalbumin Clinitek® hCG

true of blood glucose measurements, and the enzymatic reaction gradually replaced the reduction methods in the laboratory. The adaption of the convenient dip-and-read methodology for use on whole blood from finger punctures provided a convenient blood glucose analysis usable in any place, and at any time of the day or night. It thus allowed stricter control of blood glucose levels and provided a convenient alternative to trips to the doctor's office or the hospital laboratory. To help standardize the reading of results of the colorimetric test, an easy-to-use reflectance spectrophotometer was devised to read the intensity of color formed in the chemical reaction. The most recent tests utilized in the control of diabetes are the tests for glycated protein and microalbuminuria. Glycated hemoglobin, or Hb A_{1c}, is increased in persons with diabetes whose blood glucose is out of control for several days. It is a good monitor for checking the success of the therapeutic regime the diabetic patient is supposed to follow. The test for microalbuminuria is a predictive test. It has been shown that very small amounts of albumin excreted consistently by the kidney indicate that further deterioration of the kidney is likely to occur. Early detection of microalbuminuria allows steps to be taken to delay the onset of renal failure.

4 MOST COMMON URINE ANALYTES

Throughout this section, the chemical principles described are those commercially available under the series of Multistix® trademarks from Bayer Diagnostics. Similar products may be available under the Chemstrip® trademark from Roche Diagnostics, under the Rapignost® trademark from Behring Hoechst, and from KDK in Japan and others in the Far East.

4.1 Glucose

Glucose appears in the urine of a person whose pancreas does not produce active insulin to allow the body's cells to metabolize glucose. Consequently the glucose level rises in the blood and when it becomes so high that the kidney tubules cannot reabsorb it all, it spills over into the urine. The specific test for glucose is a double sequential enzymatic reaction. The reagent contains glucose oxidase, peroxidase and a potassium iodide chromogen. Glucose oxidase reacts with glucose to form gluconic acid and hydrogen peroxide. In turn the peroxidase of the reagent reacts with the hydrogen peroxide to oxidize the chromogen to a colored compound. The intensity of color is proportional to the glucose concentration.

4.2 Ketones

Ketones are breakdown products of fat metabolism. When individuals cannot utilize carbohydrates to fuel their body cells, the cells use fat for fuel. The most common instance of fat metabolism is in diabetes where insulin is not available. However, in starvation or vomiting with no or little food metabolized, or in high fat diets, ketones may also appear in the urine. The ketones, acetone and acetoacetic acid react in strongly alkaline media with nitroprusside to form a purple color. The intensity of color is proportional to the concentration of ketones.

4.3 Protein

The principal protein appearing in the urine as a result of kidney dysfunction is albumin. This protein leaks from the serum through the damaged glomerular filter into the urine to signify renal disorder. Albumin is determined by using the principle of the protein error of indicators. Some acid–base indicators are subject to protein error; that is, they turn the color that indicates alkalinity in the presence of protein even if the pH is kept constant. In the protein reagent, the indicator tetrabromophenol blue is buffered strongly at an acid pH and is a yellowish color. Thus any blue color formed when the strip is dipped into a urine specimen is due to protein and not to a pH change. A highly buffered alkaline urine may occasionally overcome the strong acid to give a pH color change.

4.4 Occult Blood

Blood appears in the urine as a result of kidney disorder. The glomerular filter damage is great if cells can get into the urine. Red cells will hemolyze in a dilute urine

specimen or when they are damaged during the course of urine flow through the kidney. Hemolysis releases hemoglobin from the cell and it is the peroxidase-like activity of hemoglobin which is detected with this reagent. The reagent contains a solid peroxide and the chromogen 3,3′,5,5′-tetramethylbenzidine. The reaction is similar to the second enzyme reaction of the glucose test. The peroxidase-like hemoglobin catalyzes the reaction of the peroxide to form the oxidized (blue) form of the chromogen.

4.5 Bilirubin

Bilirubin is the normal breakdown product of hemoglobin and it circulates in the bloodstream bound to the plasma albumin. Thus there is normally no bilirubin in the urine since albumin is too large a molecule to filter through the kidney. However, the liver conjugates bilirubin into the glucuronide, a soluble product excreted through the bile into the intestine where the normal intestinal bacteria reduce bilirubin to urobilinogen, which is further reduced to urobilin and excreted in the stool. A small amount of urobilinogen circulates in the bloodstream and appears in the urine. If the liver and biliary system do not function properly, then bilirubin glucuronide backs up into the bloodstream and appears in the urine. The principal causes of bilirubinuria are biliary obstruction and hepatitis. In fact, detection of bilirubin in urine may be an early sign of hepatitis. The test is based on the coupling of bilirubin with diazotized dichloroaniline in the presence of strong acid to give a tan color. Bilirubin is light sensitive and it is essential that the urine be tested as soon as possible after excretion.

4.6 Urobilinogen

As noted in section 4.5, a small amount of urobilinogen is not excreted into the bile but circulates in the bloodstream. Therefore a small amount of urobilinogen is normally present in the urine. Urobilinogen is *absent* from the urine in cases of obstruction of the bile duct and is *increased* in other types of liver damage such as hepatitis, cirrhosis and in blood disorders such as hemolytic anemia where excess hemoglobin passes through the liver. Table 2 shows how testing for bilirubin and urobilinogen in urine can help provide a differential diagnosis of liver disease. The test is based on the Wallace and Diamond reaction; in strong acid, urobilinogen reacts with *p*-diethylaminobenzaldehyde to give a reddish-orange color, the intensity of which is proportional to the amount of urobilinogen present.

Table 2 An indication of how testing the urine for bilirubin and urobilinogen aids in the differential diagnosis of liver dysfunction

Diagnosis	Urobilinogen	Bilirubin
Health	Normal	Negative
Hemolytic disease	Increased	Negative
Hepatic disease	Increased	Positive or negative
Biliary obstruction	Low or absent	Positive

4.7 Nitrite

The test for nitrite in urine is an indirect chemical test for bacteriuria or UTIs. Most bacteria which cause kidney or bladder infections secrete the enzyme reductase which reduces nitrate to nitrite. There is nitrate present in urine because it is an end product of protein metabolism, but nitrite is never normally present in urine. Therefore if the test for nitrite is positive it always indicates the presence of significant numbers of bacteria – generally 10^5 organisms per milliliter of urine. However, a few cases of UTIs are caused by bacteria which do not secrete reductase and therefore will not reduce nitrate to nitrite. Thus the nitrite test does not give a false positive result but in some cases does not react in the presence of bacterial infection. In this test, nitrite reacts with *p*-arsanilic acid to form a diazonium salt, which couples with 1,2,3,4-tetrahydrobenzo(h)-quinolin-3-ol to produce a pink color.

4.8 Leukocyte Esterase

Leukocytes or white blood cells are usually present in urine when there is an infection in the urinary tract because they are the cells which fight to destroy invading bacteria. Leukocytes release the enzyme leukocyte esterase into the urine and thus the test for leukocyte esterase is another indirect test for bacteriuria or UTIs. The chemical test for leukocyte esterase is based on the hydrolysis of the derivatized pyrrole amino acid ester to liberate 3-hydroxy-5-phenyl pyrrole. This pyrrole reacts in turn with a diazonium salt to produce a purple color.

4.9 pH

The normal urine pH ranges from 5 to almost 8. Urine pH varies in relation to the diet and the time of urine collection after eating, with acid urine excreted by individuals on high protein diets and alkaline urine excreted by those on diets high in vegetables and dairy products. Certain medications may cause typically acid or alkaline urines. The urine is likely to be acid in uncontrolled diabetes, and alkaline in UTIs. The reagent for the colorimetric determination of urine pH is a combination of methyl red and bromothymol blue. This gives a broad range of colors from orange at pH 5 through yellow and green to dark blue at pH 8.

4.10 Specific Gravity

The specific gravity (density) of urine varies with the amount of liquid ingested and excreted and is usually inversely related to the volume of urine. It is a measure of the amount of solids dissolved in the urine. Specific gravity values may be close to those of water (1.000) after intake of large amounts of liquid or in diseases such as diabetes or kidney disease. In contrast, in states close to dehydration from high salt ingestion, or excess perspiration, vomiting and diarrhea, the specific gravity may be as high as 1.030 or more. The unique colorimetric specific gravity test is based on the apparent change in pK_a of the polyelectrolyte poly(methyl vinyl ether/maleic anhydride) in relation to ionic concentration. The indicator, bromothymol blue, is a deep blue-green color at low ionic concentration and changes to a yellow-green color at high ionic concentration. Ionic concentration of urine is directly related to specific gravity as determined by physical measurements.

4.11 Clinitek® Microalbumin

This is a reagent strip with two tests, one for creatinine and one for microalbumin. It is a useful test to identify the early stages of kidney damage, which is particularly important for patients with diabetes or hypertension. In order to allow the use of random specimens, the albumin-to-creatinine ratio is provided as milligrams albumin per gram or millimole creatinine. Albumin-to-creatinine ratios of $30–300\,mg\,g^{-1}$ creatinine indicate microalbuminuria. The test is read on the Clinitek® 50 (described in section 2.6). The creatinine reagent is based on the peroxidase-like activity of a copper creatinine complex, which catalyzes the reaction of diisopropyl-benzene dihydroperoxide with 3,3′,5,5′-tetramethylbenzidine. The microalbumin reagent is based on dye binding using the buffered high affinity dye 3′,3″-diido-4′,4″dihydroxy-5′,5″-dinitrophenyl-3,4,5,6-tetrabromo-sulfonephthalein, with the resulting color ranging from orange to blue. Positive results from microalbuminuria testing are an indication of the likelihood of impending renal damage and proper intervention can delay serious kidney problems.

4.12 Human Chorionic Gonadotropin

HCG is the time-honored analyte which indicates pregnancy. The placenta produces this hormone soon after the fertilized egg is implanted in the uterus. The Clinitek® hCG reagent is based on an immunochemical reaction. The reagent strip is dipped into the urine specimen and the urine migrates through an area containing a monoclonal antibody specific for the beta subunit of HCG. Any HCG in the urine is bound to the antibody; the complex continues to migrate and is bound to a polyclonal antibody for the alpha subunit of HCG, which is immobilized on the strip. The instrument then displays or prints ''negative'' or ''positive''; a positive result is obtained with HCG concentrations of $25\,mIU\,mL^{-1}$ or greater.

5 LESS COMMON URINE ANALYTES

There are literally hundreds of analytes in urine which in special circumstances can provide useful information. The fact that urine is an easy-to-obtain noninvasive specimen makes it the specimen of choice in many instances. Two of these groups of tests will be briefly discussed.

5.1 Inborn Errors of Metabolism

Several important analytes for which chemical analysis has been developed are in the excretory products caused by inborn errors of metabolism. Analytes appear in urine as a result of defects of various body systems. Most of them are best tested for on newborns since they may lead to mental retardation or other disabilities, and such consequences can be prevented in some instances if early detection is made. One inborn error was originally detected by testing urine for phenylpyruvic acid as a sign of phenylketonuria (PKU). This compound reacts with ferric ions to give a blue-gray or green-gray color. It is the product of faulty metabolism of the essential amino acid phenylalanine. If PKU is detected during the first several weeks of life, the infant can be placed on special protein hydrolysates with phenylalanine removed thus preventing the mental retardation which occurs in PKU. All infants in the USA are by law tested for PKU at birth, but blood phenylalanine assays are now used. Amino acid chromatography of urine is also used on infants suspected of having other amino acid abnormalities such as maple syrup urine disease, cystinuria or homocystinuria. Galactosemia is an inborn error in which galactose cannot be metabolized by the liver and is excreted in the urine. It is rare but testing is important since removal of lactose from the diet will prevent the mental retardation and early death seen in infants with this metabolic error. The screening test combination for galactose is a positive reducing test and a negative glucose oxidase test, with confirmation of galactose by more specific measurement.

5.2 Drugs of Abuse

Urine is the test substance of choice for detection of substances of abuse. Thin-layer chromatography (TLC) is often used. Several commercial products are available for testing for the metabolites of such drugs as morphine, cocaine, marijuana, amphetamines, and barbiturates.

The most widely used immunochemical test systems are EMIT® from Syva and Abuscreen® from Roche Diagnostics. Roche has recently introduced a rapid test called OnTrack TesTcup® er.

6 OTHER BODILY FLUIDS

Most other bodily fluids used for chemical analysis are obtained by invasive techniques. And the most important, blood or serum, will be covered in other chapters of this encyclopedia. The following brief discussion summarizes bodily fluids other than urine.

6.1 Serum or Blood

There are many blood analytes of importance in health and disease. Presently most analyses are performed on serum rather than whole blood since analytes such as glucose are more stable if the red cells are removed and reagents can be reacted directly with serum while whole blood requires protein removal by precipitation and filtration. All tests are ordered from the clinical chemistry laboratory as single tests by physicians, and several may be grouped into panels pertaining to specific types of disease or particular possible sites of dysfunction, e.g. liver panel, cardiac panel. Tests range from easy-to-perform bedside analyses to those involving sophisticated complicated instrumentation that must be carried out in the main or specialized laboratory.

6.2 Sweat

The usual analyte of importance in sweat is sodium or chloride, as a presumptive test for cystic fibrosis. The most difficult part of the analysis is the collection of the specimen, since testing is usually performed on infants or small children.

6.3 Cerebrospinal Fluid

The important chemical analytes in cerebrospinal fluid (CSF) are total protein and glucose – especially in relation to meningitis, in which total protein is elevated beyond the normal 14–45 mg dL^{-1} while glucose is decreased from the normal level of 50–60% of the serum concentration. The enzymes lactate dehydrogenase and creatine kinase are also occasionally measured in CSF.

6.4 Amniotic Fluid

Specific chemical tests are performed on amniotic fluid in order to detect serious conditions in the fetus. Spectrophotometric readings for bilirubin/bilirubinoids are used to detect hemolytic disease of the newborn. Fetal lung maturity is determined by measurement of the lecithin–sphingomyelin ratio or by TLC determination of phosphatidylglycerol. Alpha fetal protein levels are measured to predict neural tube defects such as spina bifida.

6.5 Tears or Saliva

Tears and saliva have often been suggested as substitutes for serum, mainly to avoid the trauma associated with obtaining blood specimens. Diabetics who often collect fingertip blood specimens several times a day would particularly appreciate substitutes.

6.6 Stools

Although stools are technically not ''other bodily fluids'', these specimens are grouped in this category in most clinical discussions and texts. An important analyte in stools is blood. Blood appears in the stool in the presence of colorectal cancer and is an easy, inexpensive cancer test. Blood can also appear in the stool as a result of bleeding stomach or duodenal ulcers or, in some cases, as a result of more benign causes such as bleeding gums. However, repeat tests after discussion with the patient can often clarify the source. Colorectal cancer detection must always be confirmed with physical tests such as colonoscopy or sigmoidoscopy.

6.7 Miscellaneous Bodily Fluids

A recent textbook[9] describes the analysis of a variety of body fluids other than urine. It include several less often measured analytes, in addition to those mentioned here, as well as physical and microscopic analysis of seminal fluid; glucose and protein determinations in synovial fluid; and protein, various enzymes and pH in transudates and exudates of pleural, pericardial and peritoneal cavities.

7 WHERE WE ARE IN THE YEAR 2000

The chemical analysis of urine has attained an important role in the clinical laboratory, in the physician's office and in home testing by the patient. We have emphasized the more common analyses performed on urine by professional chemists, medical technologists and technicians in laboratory settings. But there are millions of tests performed on urine specimens by individuals. Among the most common are pregnancy tests. These are based on the same chemical principles as those used for laboratory testing. Fertility testing, and urine glucose and ketone monitoring by diabetics are other home tests often used. In addition, tests for UTIs are performed by patients who have recurring infections, and

tests for glucose and protein are used by women during the course of their pregnancy as suggested by their obstetricians to detect problems early enough to avoid severe consequences.

Routine urinalysis provides a cost-effective, easy-to-do procedure on an easily collected clinical specimen, instantly yielding important information on the status of health or disease of an individual. Much discussion has been held on the pros and cons of widespread testing of urine performed on select population groups or on the general healthy population as a whole, such as in health fairs, yearly physical examinations, etc. It is our opinion that such chemical testing serves a unique and important function. Those who argue that the expense of testing is higher than the perceived benefit may justify, do not consider the positive value of a negative or normal result. Not only is the finding of positive reactions important, but the fact that the majority of those tested yield normal results is of just as much value to those individuals as is the detection of some abnormality.

8 THE FUTURE – INTO THE NEW MILLENNIUM

With the rapid explosion of chemical reagents and automated instrumentation which occurred with the evolution of the diagnostics industry, and with the significant attention to genetic diseases and gene manipulation of life forms, it is difficult to predict what will happen in the future. However, we feel confident that even when blood analyses can be performed without invasive specimen collection, chemical analyses on urine will provide important information that is not available by testing other clinical specimens.

We have long been proponents of the concept of testing for wellness. If there were easy-to-do tests to detect a variety of health problems, we could help those who were identified to delay or perhaps avoid major problems. For example, if there were a test for general cancer in the early stages, those showing a positive result could then be tested further with specific tests for specific kinds of cancer. Or a test for immunocompetence could be used to identify those with a low level of immunocompetence who could then be better protected against viral diseases such as human immunovirus or hepatitis.

It would seem that health care funds could better be spent to prevent disease rather than cure it once it has been established. And so our wish for the new millennium is that a way could be found to persuade individuals to become interested enough in maintaining responsibility for their own wellness that they change their lifestyle. If everyone would embrace a healthy lifestyle and quit smoking, get optimal exercise, maintain a healthy weight, minimize fat intake, limit alcohol, and avoid or learn to cope with stress, there would be far fewer illnesses and enough funds to take care of those with unavoidable diseases.

TRADEMARKS

Alka-Seltzer®, Clinitest®, Acetest®, Ictotest®, Clinistix®, Clinilab®, Clinitek®, and Multistix® are registered trademarks of Bayer Corporation.
Tes-Tape® is a registered trademark of Eli Lilly & Co.
Abuscreen®, OnTrack TesTcup®, and Chemstrip® are registered trademarks of Roche Diagnostics.
Rapignost® is a registered trademark of Behring Hoechst.
EMIT® is a registered trademark of Syva Company.

ABBREVIATIONS AND ACRONYMS

CSF	Cerebrospinal Fluid
HCG	Human Chorionic Gonadotropin
PKU	Phenylketonuria
TLC	Thin-layer Chromatography
UTIs	Urinary Tract Infections

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Automation in the Clinical Laboratory • Drugs of Abuse, Analysis of • Glucose Measurement • Laboratory Instruments in Clinical Chemistry, Principles of • Point-of-care Testing • Product Development for the Clinical Laboratory

REFERENCES

1. A.H. Free, H.M. Free, *Urinalysis in Clinical Laboratory Practice*, CRC Press, Boca Raton, 1975.
2. R. Koch, *The Book of Signs*, Dover, New York, 71, 1955.
3. W. White, 'A New Look at the Role of Urinalysis in the History of Diagnostics Medicine', *Clin. Chem.*, **37**(1), 119–125 (1991).
4. S. Benedict, 'The Detection and Estimation of Glucose in the Urine', *JAMA*, **57**, 1193–1194 (1911).
5. S. Meites, *Otto Folin, America's First Clinical Biochemist*, AACC Press, Washington, 1989.

6. C.P. Emerson, *Clinical Diagnosis*, 1st edition, J.B. Lippincott, Philadelphia, 1906.
7. W.A. Compton, J.M. Treneer, 'Tablet and Method of Dissolving Same', US Patent 2.387, 244, October, 1944.
8. H. Free, 'Modern Urine Chemistry', Bayer Corporation, New York, Revised 1996.
9. L. McBride, *Textbook of Urinalysis and Body Fluids*, J.B. Lippincott, Philadelphia, 1997.

Coatings Analysis: Introduction

Dennis G. Anderson
CK Witco Corporation, Dublin, USA

An organic coating is applied to a substrate for either or both of the following reasons: 1) to hide the surface over which the coating is applied (beautification) and/or 2) to protect the substrate from corrosion, degradation from the elements, etc. The types of organic coatings to be examined in the following chapters can be classified according to their end-use or method of application. Typical coating applications include the protection of architecture, automobiles, appliances, aerospace vehicles, aircraft, and food and beverage containers. Application mechanisms include brush, roll coating, coil coating and electrodeposition. Many excellent references are available for the characterization of coatings systems. Among the best is the ''Fourteenth Edition of the Gardner-Sward Paint and Coating Testing Manual'' published by the American Society for Testing and Materials (ASTM).[1] Numerous chapters deal with different characterization techniques and their application to specific coating types. Heavy emphasis is given to the use of tests published by the ASTM under the direction of Committee D-1. Due to their wide applicability, references cited in this introduction will also refer to ASTM methodology.

A typical organic coating contains three major ingredients. An organic and/or inorganic pigment provides decoration and hiding of the underlying substrate. The presence of an organic binder or resin provides a mechanism for adhering the pigment to the substrate and a solvent allows for viscosity adjustment to permit easy application. A fundamental characteristic of a coatings system, therefore, is the percentage of the total coating formulation which remains on the substrate following application. In general, this is a gravimetric determination whereby the solvent is allowed to evaporate under controlled conditions. A general nonvolatile procedure has been developed (ASTM D-2832), as well as methods to examine latex emulsions (ASTM D-4758), aerosols (ASTM D-5200), resins (ASTM D-1259), varnishes (ASTM D-4209) and specific procedures for the measurement of volatile organic content (VOC) (ASTM D-2369). Most of these methods involve heating in an oven at a given temperature for a fixed time; however, increased emphasis on the use of microwave heating has been noted recently. Inorganic pigment content is generally measured by heating a dried sample of the coating in a muffle furnace to remove organic binder species (ASTM D-3723). This technique is not generally useful with organic pigments due to their limited temperature stability. In this case, centrifugation techniques have been employed (ASTM D-2698 and ASTM D-2371). The determination of nonvolatile by volume poses a number of difficulties, which have been addressed in ASTM D-2697 and a newly developed procedure based on helium displacement (ASTM D-6093). Increased concern over the release of VOCs into the environment prompted the development of ASTM D-3960, which analytically defines this quantity as 100 – % nonvolatiles – % water or exempt solvent. Flash point is an important safety feature related to the use and transportation of coating materials. Among the methods currently in use are TAG open cup (ASTM D-1310), TAG closed cup (ASTM D-56), Setaflash closed cup (ASTM D-3278), Setaflash open cup (ASTM D-4206) and Cleveland open cup (ASTM D-92). Density and specific gravity measurements are common for coating finished products and raw materials. For liquids, hydrometers (ASTM D-2935 and D-1555), pyncnometers (ASTM D-1475) and digital density meters (ASTM D-4052) are useful. For solids, it is necessary to differentiate between ''apparent'' and ''true'' density. The ''apparent'' density of a powder is measured following compaction by vibration, leaving air in the spaces between the individual powder particles. In ''true'' density, no air contribution is present during the density measurement (ASTM D-153).

In many cases it will be necessary to prepare cast films of coatings prior to physical or chemical property measurements. Measurement of appearance, for example, should be made using films prepared using the same application method expected to be employed during normal use of the product. For the measurement of physical properties, emphasis on the preparation of ''defect free'' films of appropriate film thickness is stressed. Typically, a coating is applied to a nonstick surface, such as: poly(ethylene), Teflon®, silicone coated paper or glass. Major concerns, in this regard, deal with proper wetting of the substrate by the coating and an ability to remove the film from the substrate without excessive force (ASTM D-823). Amalgamation off a tin foil surface has also been used (ASTM D-4708), but is no longer in general use due to hazards associated with the use of mercury. The simplest device for preparing a uniform cast film is the use of an aluminum or steel bar with a gap related to the film thickness desired. Excess paint is placed on the substrate and the bar is drawn across the substrate in a slow, uniform, motion. The use of automated devices is reported to produce more uniform films due to less variation in draw speed and downward force upon the bar (ASTM D-823). An alternate method for producing uniform cast films involves the use of a wirewound rod. Typically, the film applied using a wirewound rod is approximately one-tenth

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the diameter of the wire on the rod. Once cast, wet film thickness can be measured using an Inmont or Pfund wet film gauge (ASTM D-1212). Notch gauges, having notches of varying depth, have been used where thickness is estimated between highest notch which has been coated with paint and next highest uncoated notch. The estimation of dry film thickness on a nonferrous substrate makes use of eddy currents (ASTM D-1400), whereas the same measurement on a ferrous substrate uses the magnetic properties of film and substrate (ASTM D-1186). Measurement of free film thickness is required if mechanical film properties are to be measured. In this case, microscopic techniques and a number of micrometers and dial gauges (ASTM D-1005) are generally adequate. For mechanical film property measurements, specimens of known cross-sectional area and length are prepared. As these specimens are stretched, force and elongation are continuously recorded. From this data, tensile strength, elongation at break and tensile modulus at a given elongation can be calculated. A critical physical property for floor coatings is coefficient of friction. In one method, the force required to pull a weighted sled across the surface is measured (ASTM D-1894 and ASTM C-1028), while in a second approach a platform containing a sled is slowly raised until the angle is sufficiently steep to cause the coated panel to begin sliding down the incline (ASTM-4518). In a third case, the traction of footware on painted surfaces is measured using a variable incidence tester (ASTM D-2370). The flexibility of coatings on metallic substrates is measured by bending around a series of cylindrical mandrels 0.125 to 1.00 inches in diameter. Bending around successively smaller diameter mandrels takes place until the coating cracks. An alternate procedure makes use of a conical mandrel (ASTM D-522), where the coated sample is bent around a metal cone (1.5″ dia. to 0.125″ dia.). The diameter at which cracking occurs is taken as the point of failure. With some types of coatings, only solvent evaporation occurs during the drying process and the dried film continues to be sensitive to solvent. In other cases, chemical reactions occur following solvent removal and the solvent resistance of the finish improves as curing is complete. The resistance of dried coating films to solvent (ASTM D-5402), therefore, can be a convenient way to measure the extent of cure. Another useful measure of cure includes the use of a series of pencils of varying hardness (ASTM D-3363). The softest pencil required to scratch through the film is the hardness of the coating. Knoop hardness utilizes the indentation of a pyramidal diamond indenter (ASTM D-1474) and Sward rocker hardness (ASTM D-2134) measures the swings of a pendulum after rocking a given degree from an equilibrium position. Resistance of a coating to forward and reverse impact is also a measure of cure (ASTM D-2794). A coated panel is placed under a spherical weight dropped from a variety of heights. The number of ''inch pounds'' of force required to cause coating failure is used as a measure of cure. Tape adhesion (ASTM D-3359) continues as a measure of adhesion. The paint film is crosshatched down to the substrate using a sharp blade. Pressure sensitive adhesive tape is then applied to the crosshatched area. When the tape is removed, the percentage of the coating removed by the tape is used as a measure of coating adhesion. Balanced beam scrape adhesion (ASTM D-2197) uses a balanced beam to which a platform with load weights and a rod holding a scraping loop at a 45° angle are attached. The load on the scraping loop is increased to a point where the coating is barely removed from the substrate. Differential scanning calorimetry, thermomechanical analysis and dynamic mechanical analysis are also useful in measuring cure, since the glass transition temperature of a material increases as cure increases. Any exothermic or endothermic processes occurring during the curing process are also readily examined in this manner. Evaporative rate analysis measures the evaporation of a ^{14}C labeled solvent from the surface of a coating. The greater the degree of cure, the smaller the degree of penetration of solvent into the film and the more rapid the evaporation rate. Abrasion resistance is a major performance parameter in certain applications. The use of rotating, weighted, abrasive wheels is very popular (ASTM D-4060), as is the use of falling sand (ASTM D-968) and gravel (ASTM D-3170).

One of the initial stages in coating production is dispersion of the pigments. This process involves breaking down any pigment agglomerates into the ultimate particle size formed during manufacture. The extent to which dispersion has been achieved is generally monitored using a Hegman fineness of grind gauge (ASTM D-1210). A metal bar with a calibrated, tapered, slot 0–4 mm in depth is generally used. An excess of sample is placed at the deep end of the channel and the excess dispersion is drawn to the shallow end of the channel using a scraper. The presence of large particles is observed at some point along the slot when the film in the slot becomes broken. Similar information is obtained by casting a film of the pigment dispersion onto a sealed paper chart using a doctor blade. In this case, the sensitivity of the test is dependent upon the gap of the doctor blade and visual ability to observe oversize particles. Instrumental methods for the examination of average particle size and particle size distribution include the use of a disc centrifuge, low angle laser light scattering (LALLS), a Coulter® counter, sieves (ASTM E-11) and optical microscopy (ASTM E-20). The viscosity and rheological properties of coatings have a very marked effect on application characteristics and ease of use. In measuring viscosity, it is necessary to differentiate between several shear related phenomena.

If the viscosity of a material is independent of shear rate, the material exhibits Newtonian behavior. If viscosity increases as shear rate increases, the material is dilatent, while a material that decreases in viscosity with shear rate is deemed thixotropic. It is desirable, for example, for an architectural coating to be somewhat thixotropic. The coating should be viscous when in the can, such that insertion of a brush picks up a useable amount of paint. Once shear is applied, when brushing begins, we want the coating to lose viscosity, so that a thin uniform layer can be applied. The rheological properties desired for a given coating type will generally be dictated by the application mechanism to be applied during use. Many instruments are used to measure viscosity. Among the most common are the Stormer rotating paddle (ASTM D-562), a circulating spindle (ASTM D-2196) and a cone and plate (ASTM D-4287). A desire to measure the shear dependence of viscosity has also led to the development of numerous variable shear rate viscometers. Timed flow through an orifice is a convenient measurement technique for lower viscosity materials, Typical configurations involve the use of a Ford cup (ASTM D-1200), or a Zahn ''dip type'' cup (ASTM D-4212). The time necessary for an air bubble to rise in a glass tube of known dimensions is the basis for the Gardner–Holdt technique (ASTM D-1545). This procedure is relatively simple, but lacks fine definition between materials of similar viscosity. Since viscosity is quite temperature dependent, extremely good temperature control is required for accurate measurements. Several other practical methods have also been developed to meet specific application needs. These include a thixotropic index test (ASTM D-2196), a leveling test (ASTM D-4062), a procedure to measure brush drag (ASTM D-4958) and a sag resistance test (ASTM D-4400).

When light enters a paint film, some is scattered, some is absorbed and some bounces off the substrate and comes back out of the film. The smaller the fraction of light that emerges from the coating film, the greater the ability of the coating to hide any imperfections, etc. in the substrate. The minimum film thickness required to provide for a given degree of light bouncing off the substrate and back to the surface is called the ''hiding power'' of a coating system. The thinner the film required, the more efficient the coating, since more area can be effectively covered by a given volume of paint. The simplest way to measure hiding power is to cast a film of an experimental coating and a control material on a sealed paper chart having white and black sections and visually comparing how effectively the substrate features are hidden. For a more quantitative measure, the ratio of reflectance over a dark and light substrate is measured photometrically. This is deemed the ''contrast ratio'' for a given film thickness and permits the calculation of $ft^2\,gal^{-1}$ the coating can cover at a given degree of hiding. Another useful property related to substrate masking is the change in hiding which occurs upon solvent evaporation (ASTM D-5007). The gloss of a coating is its ability to reflect incident light from the air-coating interface. Specular gloss is the ratio of incident energy flux to reflected energy flux as measured at angles such as 20 and 60° from the perpendicular (ASTM D-523). Sheen, on the other hand, is measured at 85° from the perpendicular (ASTM E-284), a much more grazing angle than specular reflectance.

A number of atmospheric factors influence the weathering of coatings. These include sunlight intensity and wavelength distribution (spectrum), temperature, humidity and rainfall amount, type and amount of pollutants present and the presence of fungus, mildew, mold and insect attack. By far the most reliable testing of weathering involves the exposure of coated panels to a variety of atmospheric conditions for varying lengths of time. This process, while providing extremely useful data, is very expensive and time-consuming. To speed up this process, a number of artificial weathering mechanisms have been employed. ASTM D-5894 utilizes cyclic corrosion/UV exposure by cycling a salt fog/dry cabinet and a fluorescent UV/condensation cabinet. Supplements to this procedure include fluorescent/UV condensation and water exposure (ASTM D-4587) and recommended guidelines for the operation of this apparatus (ASTM G-53). The use of a carbon arc increases the intensity of ultraviolet light considerably (ASTM D-822, ASTM D-5031 and ASTM G-23) as does the use of a xenon arc (ASTM G-26). Application of water (ASTM D-1735, ASTM D-4585 and ASTM D-2247) or salt fog (ASTM B-117 and ASTM G-85) also increase severity of corrosion and coating degradation. Other measures of chemical resistance include the evaluation of rusting phenomena (ASTM D-610 and ASTM D-1654), blistering (ASTM D-710) and resistance to household chemicals, such as water, ethanol, acid, base, soaps and detergents, condiments, grease, oil, fats and beverages (ASTM D-1308). Electrochemical impedance measurements are also growing in popularity relative to evaluating the corrosion resistance properties of a wide variety of coating systems.

The previous sections of this introduction have concentrated on the characterization of formulated coatings systems. To obtain more complete compositional information, the remaining contributions to the encyclopedia must be consulted. Classical analysis provides functional group information as well as the measurement of volatile, pigment and binder content. Infrared and Raman spectroscopy completes qualitative analysis for organic and inorganic functional groups. Organic species can be further characterized via nuclear magnetic resonance spectroscopy. Volatile components are most conveniently

examined using gas chromatography. Nonvolatile species can also be examined in some cases following derivitization, thermal or chemical degradation. Thermal analysis provides useful data concerning decomposition mechanisms and thermal stability. Elemental analysis for trace and major components is possible using atomic spectroscopy. Mechanical property measurements yield valuable information on cured coatings systems, where rheology and viscometry are uniquely suited to the examination of uncured coating application properties. Finally, microscopy and related techniques give the coatings analyst powerful tools to study coating defects, problems with exterior durability and substrate interactions.

REFERENCES

1. J.V. Koleske, *Paint and Coating Testing Manual; Fourteenth Edition of the Gardner-Sward Handbook*, American Society for Testing and Materials, West Conshohocken, PA, 1995.

Atomic Spectroscopy in Coatings Analysis

Dwight G. Weldon
MATCO, Inc., Pittsburgh, USA

1 Introduction 1
2 Electromagnetic Radiation 1
2.1 Theory of Electromagnetic Radiation 1
2.2 Theory of Absorption and Emission Processes 2
3 Atomic Absorption and Atomic Emission Spectroscopy – Equipment 2
3.1 Burners and Atomizers 2
3.2 Spectrophotometer 3
3.3 Sources 3
4 Sources of Error 3
5 Inductively Coupled Plasma Spectroscopy 4
5.1 Theory 4
5.2 Inductively Coupled Plasma Instrumentation 4
6 Applications 5
Abbreviations and Acronyms 8
Related Articles 8
Further Reading 8
References 8

In atomic spectroscopy, vaporized neutral atoms are detected by measuring either the absorption or emission of light at characteristic frequencies. In atomic absorption (AA) spectroscopy, radiation is absorbed by the atom in the process of exciting an electron from a lower energy state into a higher energy state. In atomic emission (AE) and inductively coupled plasma (ICP) spectroscopy, radiation is released when an electron from an excited energy state relaxes into a lower energy state. The energy required for these transitions occurs in either the visible or ultraviolet region of the electromagnetic spectrum.

Atomic spectroscopy is a very useful technique in the field of coating analysis. It is primarily used for the analysis of metallic constituents, such as certain pigments and coating additives. It is extremely sensitive, and can detect several dozen different elements. It is especially useful in determining whether or not a particular coating contains environmentally hazardous levels of heavy metals, such as lead, cadmium and chromium. Disadvantages of the technique include its inability to determine the chemical form in which the element is present, somewhat laborious sample preparation procedures, and the inability to determine other important constituents of coatings, such as resins and solvents.

1 INTRODUCTION

Coatings are complex materials, which contain pigments, plasticizers, additives, solvents, and one or more polymeric materials. Therefore, a thorough analysis of a coating requires the use of several analytical techniques. One technique which is useful for the analysis of metallic constituents such as certain pigments and additives is atomic spectroscopy.

There are several variations of atomic spectroscopy utilized in the coatings industry, including AA, AE, and ICP spectroscopy. This chapter will briefly address the theory of these techniques, followed by examples of their use in the field of coatings analysis.

2 ELECTROMAGNETIC RADIATION

A detailed discussion of the theory of electromagnetic radiation and of atomic spectroscopy is contained elsewhere in this work. Therefore, only a brief introduction is contained here.

2.1 Theory of Electromagnetic Radiation

The light by which we see the world around us is a form of electromagnetic radiation which, in its simplest terms, is energy traveling through space. In a vacuum, the velocity of electromagnetic radiation is equal to the speed of light ($C = 2.99 \times 10^{10}\,\mathrm{cm\,s^{-1}}$). The classical interpretation of this radiation describes it as a series of waves propagating through space with a certain wavelength (λ) and frequency (ν). This simple wave-like interpretation was able to explain certain macroscopically observed phenomena, such as diffraction, but was found to be insufficient to explain other phenomena, such as the photoelectric effect.

The photoelectric effect, whereby light impinging on a photoelectric cell results in the generation of current, was more readily explained by the quantum mechanical interpretation of electromagnetic radiation. Quantum mechanics describes electromagnetic radiation as small discrete particles of energy referred to as photons.

In either case, electromagnetic radiation can be described by the simple relationship that $\lambda = V/\nu$. In other words, the wavelength of the radiation (the distance

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

between two successive maxima or minima of the wave front) is equal to the velocity of the wave front divided by its frequency in cycles per second, or hertz. Because, in a vacuum, the velocity is equal to the speed of light, the equation can also be stated as $\lambda = C/\nu$.

The quantum mechanical, or particle-like interpretation of electromagnetic radiation relates the energy of a photon with its frequency, according to the equation $E = H\nu$, where H is Planck's constant (6.624×10^{-27} erg s), and ν is again the frequency in cycles per second, or hertz.

As this simple equation illustrates, the energy of electromagnetic radiation is directly proportional to its frequency. Therefore, it is reasonable to expect that light of varying frequencies can interact differently with matter. This is indeed true and accounts for the wide variety of spectroscopic techniques in use today. Frequencies ranging from about 20 000 to 10^6 cm^{-1} encompass the visible and ultraviolet region of the spectrum, and involve relatively high-energy processes such as outer electron transitions. These transitions form the basis of atomic spectroscopy.

2.2 Theory of Absorption and Emission Processes

Of AA, AE, and ICP spectroscopy, the one most widely used in the coatings industry is AA spectroscopy.

The electrons in an atom occupy certain energy levels about the nucleus. The electronic configuration is such that the greatest population of atoms resides in the lowest energy state, or ground state. If the atom is exposed to light of an energy exactly matching the difference in energy between a ground state orbital and a higher energy orbital, a portion of this light can be absorbed and the electron will be raised, or excited, from the ground state to a higher energy excited state. Thus, if a suitable light source is coupled with a detector capable of sensing the attenuation of this light source as a portion of it is absorbed by the atoms in its path, the concentration of a particular species can be determined. This is the basis of AA spectroscopy.

It seems reasonable that if an electron in an atom is excited to a higher energy level by the absorption of electromagnetic radiation, it will eventually decay back to the initial ground state. When it does, a photon of light equal in energy (frequency) to that initially absorbed will be generated. The detection and quantification of this emitted radiation is the basis of AE spectroscopy.

3 ATOMIC ABSORPTION AND ATOMIC EMISSION SPECTROSCOPY – EQUIPMENT

Detailed descriptions of the processes and equipment involved in absorption and emission spectroscopy can be found elsewhere in this text. However, this section contains a brief description of the three major components of an AA/AE spectrometer: burners and atomizers, spectrophotometers, and sources.

3.1 Burners and Atomizers

The atomic processes which are responsible for the basis of AA/AE spectroscopy occur in the flame. Therefore, it is appropriate to begin a discussion by considering burners and atomizers.

The purpose of an atomizer is to produce a fine uniform spray of the solution containing the sample. This fine suspension is carried into the reaction zone of the flame by the gas stream used to support combustion, usually air or oxygen. In order to reduce noise and increase precision, the atomizer must be able to deliver this solution at a steady rate, and the temperature of the flame should be relatively constant. Therefore, precisely controlled flow rates of both the fuel and the oxidant are essential.

Most spectrometers employ what are referred to as premix burners. In a premix burner, the aspirated sample, fuel, and oxidant are mixed prior to reaching the burner opening. This opening consists of a long narrow strip, and the flame produced is smooth, nonturbulent, and quiet. Most premix burners are designed with a series of baffles which are located before the burner head and which promote mixing and allow larger droplets to be drained off. Indeed, the majority of the sample aspirated into a premix burner actually goes down the drain, contributing to a loss of sensitivity. However, the advantages of such an aspirator usually outweigh the disadvantage of sample loss.

Several complicated events occur once the sample actually reaches the flame. These include the evaporation of the water or other solvent carrying the sample, the decomposition of the vaporized analyte to produce neutral atoms or radicals, and the excitation of a certain percentage of these individual atoms.

Since it requires considerable energy to vaporize and dissociate inorganic salts, the temperature of the flame is one of its key parameters. The most common type of flame is the air–acetylene flame, which can maintain a temperature of approximately 2200 °C. The other common flame is the nitrous oxide–acetylene flame, which reaches a temperature of approximately 2900 °C.

Not only do hotter flames more effectively vaporize and decompose samples, but they can also reduce the number of chemical interferences. The primary source of chemical interference is compound formation within the flame, which can result in broad band absorption and emission. As the flame temperature is increased, many of these compounds are decomposed into atoms, thus reducing or eliminating the problem. Hot flames

are also desirable in AE spectroscopy, where their higher energy results in the excitation of an increased number of elements. However, if too high a temperature is used, it is possible that a significant portion of the neutral atoms may become ionized, and therefore useless analytically.

It should be pointed out that nonflame atomizers exist for AA equipment. By far, the most common type of nonflame atomizer is the graphite furnace. A typical graphite furnace consists of a hollow graphite cylinder a few centimeters in length, and a few millimeters in cross-section. The tube is heated electrically to produce very high temperatures. The sample is usually introduced via a micropipet through a hole in the top of the tube.

The heating of the graphite tube is usually accomplished in stages. The first stage is a low-temperature step used to volatilize any water or solvent in the sample. The second higher temperature step is used to ash the sample, followed by a rapid increase to temperatures of 2000–3000 °C resulting in atomization of the elements.

The main advantage of the graphite furnace is improved detection limits. In a graphite furnace, there is no loss of sample down a drain tube and the actual residence time of the element in the analytical zone is greatly increased.

3.2 Spectrophotometer

The best burner/flame assembly would be of little use without an efficient monochrometer to sample the analytical line of interest accurately. Both single beam and double beam monochrometers are available. Most instruments are of the single beam type, which use a rotating chopper. This allows the source radiation to pass alternately through the flame or to be sampled unattenuated.

In a double beam instrument, the source radiation is split into two beams, one of which passes through the flame, while the second, or reference beam, is detoured around the flame by a series of mirrors. The two beams are then recombined after the flame, and the ratio between the sample beam and the reference beam are electronically measured.

3.3 Sources

By far, the most common source used in AA spectroscopy is the hollow cathode lamp. The hollow cathode lamp consists of a cathode manufactured from the element of interest, situated adjacent to an anode in an enclosed glass tube which has a glass or quartz window. An inert gas, such as neon or argon, at a few millimeters of pressure, is used as a filler gas inside the tube. A potential difference across the anode and cathode results in the sputtering of a small number of metal atoms from the cathode, which in turn collide with the inert gas, resulting in the production of radiation. Although most lamps use cathodes manufactured from a single element, it is possible to purchase multielement lamps which contain from two to perhaps six different elements.

As of yet, no broad band continuous sources have been found suitable for AA spectroscopy. This is a severe limitation, because it means that the technique is quantitative, not qualitative. In order to analyze for a particular element, a hollow cathode lamp manufactured from that element must be mounted in the instrument. ICP spectrometers, which will be discussed later in this chapter, are a means of solving this problem.

4 SOURCES OF ERROR

Although the absorption and emission lines used in flame photometry are much narrower than those encountered in many other spectroscopic techniques, there are still several sources of error to be considered. Common sources of error include background effects, ionization effects, and spectral and chemical interferences.

With the possible exception of flourescence techniques, almost any analytical measurement has a background signal associated with it. In AA and AE spectroscopy, the two main sources of background signal are the flame itself and matrix effects, with matrix effects usually predominating. In emission spectroscopy, these effects can usually be accounted for by slowly scanning over the region of interest and drawing in a base line, either manually or electronically. This is not possible, however, in AA spectroscopy, using monochromatic line sources such as hollow cathode lamps.

Instrument manufacturers have devised several techniques for greatly reducing errors due to background radiation. These techniques are discussed in greater detail in other sections of this encyclopedia.

While it may be debatable whether or not ionization is actually a source of interference, it can certainly affect detection limits. Both absorption and emission spectroscopy rely on the population of neutral metal atoms in the flame. In absorption spectroscopy, the concentration is determined by measuring the attenuation of the source beam, which is being absorbed by the ground state atoms. In emission spectroscopy, the concentration is determined by measuring the intensity of radiation emitted when atoms which have been thermally elevated to an excited state subsequently decay and emit radiation. In both processes, a reduction in the population of neutral atoms results in a decreased signal. One mechanism whereby this population is decreased is ionization.

The ionization process occurs when a neutral metal atom absorbs sufficient energy to eject an electron, thus

forming an ion. This process will be more prevalent at higher flame temperatures. However, selecting a relatively low temperature flame may not be the best alternative, because it may result in greater spectral and chemical interferences.

For many metal atoms, the ionization potential is much higher than the energy available from the flame, and ionization is not a problem. However, for easily ionizable elements such as potassium, the percentage ionization can be well over 50% in some of the hotter flames. One way to reduce this effect is the use of a radiation buffer or ionization suppressant. An ionization suppressant is simply another easily ionized element, an excess of which is added to the sample. In essence, the use of an ionization suppressant causes a large increase in the density, or concentration, of electrons in the flame, and thus suppresses the ionization of the element of interest as a result of the Le Chatelier effect.

Spectral interferences are caused by the overlapping of lines or bands from an impurity at or near the wavelength of the analyte. A common example of spectral interference is the effect of calcium on the determination of sodium. The characteristic yellow line of sodium, at 590 nm, overlaps a band produced by calcium hydroxide, which is a thermally stable compound that can exist in the flame. This type of interference can be minimized by using a high-resolution monochromator.

Spectral interference is usually more of a problem in AE spectroscopy than in AA spectroscopy. In absorption spectroscopy, the source line produced by the hollow cathode lamp is extremely narrow and would therefore not be appreciably absorbed by other species whose absorption bands are close to the element of interest. However, there are still examples of spectral interference even in AA spectroscopy. For example, iron and platinum both have absorption lines at 2719 Å. If this type of interference is suspected, it may be possible to perform the analysis using a line other than the main absorption line.

Chemical interferences are caused by the element of interest forming a thermally stable compound in the flame, thus decreasing the population of free metal atoms. For example, some elements can form stable oxides or hydroxides within a flame, such as calcium. One of the best ways of reducing chemical interference is to use a flame of higher temperature, although ionization effects must be considered.

Another source of interference is instability in the flame itself. This is usually not a problem with modern instrumentation, which can produce very consistent flow rates. However, particulate matter in the sample can result in flame fluctuations which can affect the accuracy and sensitivity of the analytical measurement.

5 INDUCTIVELY COUPLED PLASMA SPECTROSCOPY

Both the theory and instrumentation of ICP spectroscopy are covered in detail elsewhere in this publication. Therefore, the current chapter provides only a brief outline of this field.

5.1 Theory

ICP spectroscopy is closely related in theory to the more conventional AE spectroscopy discussed previously. In essence, the technique relies on the measurement of the intensity of light emitted when excited state atoms relax to their ground state. Instead of using a flame to accomplish this, ICP uses a plasma. The plasma is formed as a result of argon gas being excited by an inductively coupled radiofrequency generator. The plasma is formed in what is referred to as the plasma torch, into which the sample is carried by a separate stream of argon gas. Temperatures in the plasma are much hotter than any conventional flame, and can reach 7000 °C. As a consequence, several elements can be analyzed by ICP spectroscopy which could not be analyzed by AE spectroscopy.

An advantage of ICP spectroscopy, which it shares with AE spectroscopy, is the fact that individual hollow cathode lamps are not required. Therefore, when coupled with an appropriate monochrometer, ICP is a very efficient technique for conducting simultaneous multielement analysis. Indeed, it would rarely be used to analyze for a single element, as the cost of both the instrumentation and the analysis would be prohibitive.

Earlier in this chapter, it was stated that one of the disadvantages of very hot flames was the increased ionization of the sample, resulting in poorer detection limits. This is not the case, however, in ICP spectroscopy because of the suppression effect caused by the very large number of argon ions present in the plasma. In general, detection limits are an order of magnitude better than AA spectroscopy, and calibration curves also span a wider linear range.

5.2 Inductively Coupled Plasma Instrumentation

Briefly, there are two fundamental types of ICP instrument on the market, distinguished by the type of monochrometer employed, the sequential or simultaneous.

In most sequential spectrometers, the light emitted when excited state atoms relax to the ground state is dispersed by a movable diffraction grating and subsequently focused on a fixed detector. Another design is to keep the grating in a fixed position and move the detector. Although modern monochrometers can perform this function in a relatively short time, the technique is still

sequential, since the radiation emitted from a specific element must first be collected before the grating or detector is moved to determine the next element in sequence. Current instruments can do approximately 20–50 elemental determinations per minute. This is dramatically faster than a conventional AA instrument.

In simultaneous spectrometers, a bank of detectors, typically photomultiplier tubes, diode arrays, or some type of charge-coupled device, are positioned at fixed points about the grating so that light at several preselected frequencies can be measured simultaneously. Therefore, the analysis is truly simultaneous: multiple elements can be detected and quantified virtually instantly.

While simultaneous spectrometers would appear to have a significant advantage over sequential instruments, not only are they typically more expensive, but they lack the flexibility of a sequential instrument. Because only a limited number of detectors can be used, only a limited number of preselected analytical lines can be measured. Most instruments can be set up to detect approximately 30–40 elements. However, this deficiency has been mitigated by the introduction of a segmented array charge coupled device as a detector. One such device is claimed by its manufacturer to be able to measure 60 elements in less than a minute at multiple wavelengths.

6 APPLICATIONS

Approximately 70 elements can be detected by atomic spectroscopy. Detection limits vary, but are typically on the order of approximately 1 part per million or less. Typically, paint samples are digested in hot acid, sometimes after a preliminary ashing step in a muffle furnace to remove resins and polymers. Although several hours may be required for sample preparation, the actual analysis time is extremely short. Samples are generally introduced in a dissolved state. Many of the methods have been devised to address environmental health and safety issues relating to heavy metals such as lead, cadmium, and chromium.

The following are brief outlines of many of the published test methods for analyzing coatings or coating-related materials by atomic spectroscopy, and Table 1 summarizes the methods in terms of analytes and conditions. Many of them are published annually by the American Society for Testing and Materials (ASTM).[2] Another excellent source for published test methods is the journal *Analytical Chemistry*. Since 1987, this journal has published its *Application Reviews* in the June 15th edition of every odd-numbered year. Prior to 1987, they were published in the April editions. For example, the 1993 review[1] lists 13 AA/ICP methods for the analysis of coatings.

ASTM D 3335[2] is a method for determining low concentrations of lead, cadmium, and cobalt in paint by AA spectroscopy. For samples containing approximately 500 parts per million lead, the specimen size is approximately 2–3 g of liquid coating, or 1–2 g of dried paint film. Smaller quantities can be used for samples containing higher amounts of lead.

Liquid specimens are first dried on a hot plate or in a low-temperature oven, after which time they are treated the same as dried specimens, and subjected to an ashing procedure in a 500 °C muffle furnace. After ashing, the residue is digested in 10 mL of 1 : 1 nitric acid on a hot plate and filtered through medium porosity filter paper into a 50-mL volumetric flask. The container and filter paper are washed with small quantities of hot ammonium acetate solution, and the volume finally adjusted to 50 mL with deionized water. The diluted solution is aspirated

Table 1 Summary of methods

Reference[a]	Analytes	Sample preparation	Flame	Instrument
2 D3335	Pb, Cd, Co	Ash/hot plate	Air/acetylene	AA
2 D4834	Pb	Dilution	Air/acetylene	AA
3	Pb	Hot plate	Air/acetylene	AA
2 D4358	Pb, Cr	Hot plate	Air/acetylene	AA
2 D3718	Cr	Ash/pressure vessel	Nitrous oxide/acetylene	AA
2 D3717	Sb	Ash/reflux	Air/acetylene	AA
2 D3624	Hg	Pressure vessel	Cold vapor	AA
2 D4563	Ti	Ash/hot plate	Nitrous oxide/acetylene	AA
4	Surfactants	Solvent extraction	Graphite furnace	GFAA
5	Ba, Cd, Zn, P	Various	Air/acetylene	AA/ICP
6	Sn	H_2O leaching	Furnace	GFAA
7	Pb, Ba, Zn, Ca, Ti, S, Mg, Al, Cu, Sb, Mn	Microwave	N/A	ICP
8	Fe, Cu, Cd, V	Direct aspiration	N/A	ICP

[a] Includes ASTM numerical number where appropriate. GFAA, graphite furnace atomic absorption.

into an AA spectrometer, which has first been calibrated with lead, cadmium, or cobalt standards.

Another method for determining lead in liquid paint is ASTM D 4834.[2] The first step in this procedure is to prepare three standard paints by accurately weighing a compatible lead compound to a paint which has already been shown to contain less than 10 parts per million of lead. The standards are prepared such that they contain approximately 300, 600, and 900 parts per million lead based on their nonvolatile content.

Samples and standards are analyzed by accurately weighing approximately 0.5 g of liquid specimen into a 10-mL volumetric flask. The flask is then diluted to the mark with methyl isobutyl ketone (MIBK) for solvent-reducible coatings, or water for water-reducible coatings. The standards and specimens are then aspirated immediately after mixing, and their absorption recorded. The amount of lead can then be determined and related to the coating's nonvolatile content once this has been determined in accordance with other appropriate ASTM methods. It is important that MIBK or water is aspirated immediately after each standard or specimen, to prevent the fouling of the aspirator/nebulizer system.

Another method for determining the lead content of dried paint samples is the Association of Official Analytical Chemists (now known as AOAC international) (AOAC) Method 974.02.[3] In this procedure, a few tenths of a gram of either liquid or solid paint is weighed into 16×150-mm test tubes, and digested in approximately 5 mL of concentrated nitric acid using a heating block or sand bath. The temperature is brought to 90–100 °C until the initial fuming has subsided, and then increased such that the vapors are condensing in the top 1–2 cm of the test tube. The digestion is continued for 3 h, after which time the specimens are cooled, transferred to 25-mL volumetric flasks, and rinsed with several small portions of deionized water. The specimens are diluted to volume, and then allowed to settle for at least one-half hour. Any floating residue may be removed by aspiration through a disposable pipet prior to introduction into the spectrophotometer.

A method for determining both lead and chromium in air particulate filter samples by AA spectroscopy is described in ASTM D 4358.[2] The procedure covers the analysis of pigment dust on filters, expressing the results as micrograms of lead or chromium. The method does not address sampling procedures, only the analytical procedure. Briefly, the filter media is digested in 6 mL of one part concentrated nitric acid diluted with three parts water for approximately 10 min on a hot plate. Addition of 2 mL of 50% ammonium acetate solution follows, along with water as needed, with continued heating for 1 h. The samples are then cooled, transferred to a 10-mL volumetric flask, diluted, and analyzed by AA.

ASTM D 3718[2] describes a test method for determining low concentrations of chromium in paint by AA spectroscopy. Calibration of the instrument is performed by using chromium working solutions of various concentrations, which bracket the expected chromium concentration of the sample to be tested. The standards are made from a 100 part per million working solution prepared by dissolving potassium chromate in water. This working solution is first adjusted by adding an oxidizing solution (potassium permanganate in sulfuric acid), followed by dropwise addition of a reducing solution (hydroxylamine hydrochloride in water) until the permanganate color has just been discharged.

Once the calibration curve has been constructed, samples are analyzed by weighing approximately 5 g of liquid coating or 3 g of dried coating into porcelain crucibles. For liquid coatings, the crucibles are heated on a hot plate or in a low-temperature drying oven to remove solvents. The specimens are then heated further on the hot plate until the material chars, at which time the crucibles are ashed in a muffle furnace at 500 °C for 1–2 h. The percentage ash is determined. After grinding to a uniform powder, 0.02–0.2 g of ash is then weighed directly into polytetrafluoroethylene (PTFE) decomposition vessel inserts along with 10 mL of oxidizing solution. The digestion vessels are then placed in a 105 °C oven for 1.5 h, removed, and allowed to cool. The contents are then quantitatively filtered directly into 50- or 100-mL volumetric flasks. The filter paper is then washed several times with small volumes of water, and hydroxylamine hydrochloride solution added dropwise to the filtrate until the permanganate color has been discharged. Finally, the volumetric flasks are diluted to the mark with water and analyzed by AA spectroscopy.

ASTM D 3717[2] describes a method for determining low concentrations of antimony in paint by AA spectroscopy. Standards are made using suitable aliquots of a 0.1 mg mL^{-1} standard antimony solution, to which 40 mL of concentrated HCl and 8 drops of a stannous chloride solution have been added. Specimens consist of 5–10 g of liquid coating or 2–6 g of dried coating. The liquid specimens are first dried on a hot plate, and are then treated the same as dried specimens. The temperature of the hot plate is increased until the material chars, after which time the specimens are ashed in a 500 °C muffle furnace for 1–2 h. They are allowed to cool, and then 40 mL of concentrated hydrochloric acid and 8 drops of stannous chloride solution are added. Condensers are added and the flasks refluxed for 1 h.

The digested specimen is filtered into a 100-mL volumetric flask. After rinsing and washing, the flask is diluted to the mark, and the specimens aspirated into an AA spectrometer.

ASTM D 3624[2] describes a method for determining low concentrations of mercury in paint by a cold vapor AA technique. The apparatus is calibrated using various size aliquots of a working mercury standard containing 1000-ppm mercury. These aliquots are added to a series of 250-mL gas washing bottles. Enough water is added to each bottle to make a total volume of 100 mL and then 5 mL of sulfuric acid and 3 mL of nitric acid are added. Dropwise additions are made of 0.1 N potassium permanganate solution to each bottle until a pink color persists, and then 1 mL of hydroxylamine hydrochloride solution is added until the pink color is discharged. The bottles are then sealed using fritted glass inlet tubes and stoppers, and nitrogen is allowed to flow through the bottles.

The apparatus is constructed such that the nitrogen flows from the specimen bottle into a water vapor trap and a drying tube, and finally into a suitable absorption cell positioned in the spectrometer compartment. This is a cell 100 mm in length having quartz windows, aligned within the spectrometer compartment to provide maximum transmittance. A calibration curve is constructed using the mercury standards and then the various specimens are analyzed.

The sample preparation for ASTM D 3624 involves weighing liquid paint or dried specimens into PTFE inserts of acid decomposition vessels. The inserts are then placed in an oven at 60 °C for 1 h to remove all volatiles. Sulfuric acid and nitric acid are then added to the decomposition vessel inserts, and the vessels are sealed and placed in a 140 °C oven for 1.5 h. After cooling, 5 mL of cold water are added, and the contents are diluted and filtered into a 100-mL volumetric flask. The residue and filter paper are washed several times with small volumes of water, and the flask filled to the mark with deionized water. The samples are then treated in the same fashion as the calibration standards.

ASTM method D 4563[2] describes a method for determining the titanium dioxide content of pigments by AA spectroscopy. The initial step is to isolate the pigment from the liquid paint by drying at 110 °C followed by ashing at 450 °C. The percentage pigment must also be determined by first drying and subsequently ashing a known weight of liquid sample. A portion of the pigment thus obtained is dried for 2 h at 105 °C. The dried specimens are then transferred to 500-mL flasks, to which 8 g of ammonium sulfate and 50 mL of concentrated sulfuric acid are added, along with five drops of concentrated nitric acid. This mixture is heated on a hot plate until solution is complete (usually not more than 5 min of boiling), cooled, and diluted with 100 mL of water.

The contents of the flasks are then transferred to 1000-mL plastic volumetric flasks, and 50 mL of hydrochloric acid and 30 mL of hydrofluoric acid are added. After cooling, the flasks are diluted to the mark with water, and the solutions are then filtered prior to being analyzed by AA spectroscopy using a nitrous oxide–acetylene flame. The standard for the test procedure is a different sample of pigment which has previously been analyzed in accordance with ASTM D 1394.

Surfactants are widely used as additives in various types of paints. Chattaraj and Das[4] have described an indirect AA spectroscopy method for the determination of cationic surfactants. The method involves diluting the sample with water, and extracting 1–4 mL of it in a 100-mL separatory funnel with 2 mL of a sodium hexanitrocobaltate(III) solution, 10 mL of a sodium phosphate/citric acid pH 7 buffer, and 5 mL of 1,2-dichloroethane. The conditions favor the formation of an ion pair between the cationic surfactant and the hexanitrocobaltate(III) anion, which is soluble in the 1,2-dichloroethane. Successive extracts are obtained, combined, and diluted to 10 mL. The cobalt concentration in the organic phase is then determined via graphite furnace atomic absorption spectroscopy (GFAAS), using the 240.7 nm line of a cobalt hollow cathode lamp.

A discussion of the methods used for the AA determination of barium, cadmium, zinc, and phosphorous in vinyl stabilizers was presented by Nagourney and Madan.[5] Vinyl stabilizers are combinations of metal salts and various organic additives which are used to inhibit color formation and increase the effective service life of polyvinyl chloride materials. The authors discussed several extraction methods, including aqueous extraction using a variety of acids, and organic extractions using appropriate solvents. Analysis of the final extracts was performed by AA spectroscopy, GFAAS, and ICP/AE spectroscopy.

Alkyl tin compounds are widely used as additives in paints designed for ship hulls to prevent biological growth. Katsura et al.[6] have described the sensitivity enhancement by organopalladium complexes of the GFAAS determination of such compounds leached from ship paints.

GFAAS is usually a very sensitive technique. However, the technique is not as sensitive as might be expected for tin in organotin compounds because of their volatile nature. This volatility results in significant evaporation of the analyte from the graphite boat prior to the actual analytical measurement. Katsur et al. evaluated several chemical modifiers in an attempt to improve the sensitivity of the technique, and found that $PdCl_2(CH_3CN)_2$ dissolved in ethyl acetate/hexane enhanced the detection limit of organotin materials by a factor of approximately 19. The authors then conducted a study to evaluate the efficacy of this modifier in determining the amount of total tin leached from ship paints by artificial seawater.

A half circular specimen painted with a coating containing tributyl tin chloride was rotated in 1.5 L of

artificial seawater at 60 rpm. Seawater in 5-mL aliquots was taken after various time intervals, and placed in 100-mL separatory funnels. Hydrochloric acid and saturated sodium chloride solution were added, and the mixture was diluted with 50 mL of water. Ethylacetate–hexane (10 mL) was then added, and the funnels were shaken vigorously and allowed to sit in such a way that the aqueous layer separated. This extraction procedure was repeated three times and all the organic layers were collected and concentrated to less than 4 mL. A 1-mL volume of $PdCl_2(CH_3CN)_2$ and ethyl acetate–hexane was added, and the total volume of the sample was diluted to 5.0 mL with additional ethyl acetate–hexane. A 10-mL quantity was then injected into the graphite furnace and analyzed for tin using the 246 nm line.

Paudyn and Smith[7] described the determination of various elements in paints by ICP/AE spectroscopy using microwave-assisted digestion. Approximately 15–20 mg of powdered paint samples were transferred into Teflon® microwave digestion vessels. Liquid paints were also evaluated, after first evaporating any solvents present by leaving the vessels uncovered in the microwave oven for 1 min. Nitric acid and hydrofluoric acid were then added to the vessels, which were then capped, weighed, and heated in the microwave oven for 35 min. After washing into PTFE beakers the specimens were reduced to approximately 1 mL on a hot plate, and finally diluted to 50 mL with deionized water.

The diluted specimens were then analyzed on an ICP spectrometer against multielement working standards. Elements reported on included lead, barium, zinc, calcium, titanium, sulfur, magnesium, aluminum, copper, antimony, and manganese.

Rademeyer and Fischer[8] reported a method to determine metals in wax. The method uses a V groove nebulizer and a heated spray chamber to nebulize molten waxes directly into an ICP spectrometer. Calibration of the technique was done via the method of standard additions by mixing a multielement standard oil with melted wax. Detection limits of 0.3–0.9 mg g^{-1} were reported for iron, copper, cadmium, and vanadium. However, problems were encountered with precision as the result of nebulizer instability, and the authors offered suggested courses of action.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
AE	Atomic Emission
AOAC	Association of Official Analytical Chemists (now known as AOAC international)
ASTM	American Society for Testing and Materials
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
ICP	Inductively Coupled Plasma
MIBK	Methyl Isobutyl Ketone
PTFE	Polytetrafluoroethylene

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

FURTHER READING

J.T. Vandeberg, H.D. Swafford, R.W. Scott, 'Determination of Low Concentrations of Lead in Paint', *J. Paint Technol.*, **47, 604**, 84–91 (1975).

D.R. Brezinski, 'Review of Factors Determining Solubility and Absorption of Lead and Other Trace Metals in the Gastrointestinal Tract', *J. Coatings Technol.*, **48, 615**, 48–57 (1976).

REFERENCES

1. D.G. Anderson, 'Coatings', *Anal. Chem.*, **65**(12), 1R–11R (1993).
2. *Annual Book of ASTM Standards*, Vol. 06.01, American Society for Testing and Materials, Philadelphia, published annually.
3. ASTM, *Official Methods of Analysis of the Association of Official Analytical Chemists*, Association of

Official Analytical Chemists, 1st edition, Arlington, VA, 1990.
4. S. Chattaraj, A.K. Das, 'Indirect AAS Determination of Cationic Surfactants in Waste Effluents and Shampoos', *Anal. Lett.*, **25**(12), 2355–2366 (1992).
5. S.J. Nagourney, R.K. Madan, 'Metals Analysis in the Plastics Industry: Quality Assurance Considerations', *J. Testing Evaluation*, **19**(1), 77–82 (1991).
6. T. Katsura, F. Kato, K. Matsumoto, 'Sensitivity Enhancement by Organopalladium Complexes for Graphite Furnace Atomic Absorption Spectrometry of Alkyltin Compounds in Organic Solvents and its Application to the Determination of Total Tin Leached from Ship Paints', *Anal. Chim. Acta*, **252**, 77–81 (1991).
7. A.M. Paudyn, R.G. Smith, 'Determination of Elements in Paints and Paint Scrapings by Inductively Coupled Plasma Atomic Emission Spectrometry Using Microwave Assisted Digestion', *Fresenius' J. Anal. Chem.*, **345**, 695–700 (1993).
8. C.J. Rademeyer, J.L. Fischer, 'Determination of Metals in Waxes by Inductively Coupled Plasma Atomic Emission Spectrometry Using High Temperature Nebulization of the Molten Wax', *J. Anal. At. Spectrom.*, **8**, 487–492 (1993).

Gas Chromatography in Coatings Analysis

J.K. Haken
The University of New South Wales, Sydney, Australia

1 Introduction 1
2 Solvent Analysis 3
3 Monomers 4
4 Plasticizers 4
5 Vegetable Oils 4
6 Fish Oils 5
7 Tall Oil 5
8 Aerosol Coatings 5
9 Coating Additives 5
10 General Methods of Resin Analysis 6
10.1 Pyrolysis 6
10.2 Chemical Degradation 7
10.3 Pyrolytic Methylation 8
11 Condensation Systems 8
11.1 Alkyd and Polyester Resins 8
11.2 Silicone Polyesters 10
11.3 Polyurethanes 10
11.4 Epoxy Resins 10
11.5 Polysiloxanes 11
11.6 Polyamides 11
11.7 Vinyl Esters 12
11.8 Phenol Formaldehyde Resins 12
11.9 Rosin Derivatives 12
11.10 Cellulose and Derivatives 12
12 Addition Systems 13
12.1 Polyolefins 13
12.2 Polyethers 13
12.3 Polyvinyl Chloride 14
12.4 Methacrylate Polymers 14
12.5 Acrylates 14
12.6 Polystyrene 14
12.7 Polyacrylonitrile 14
12.8 Polyvinyl Acetate 15
12.9 Polyvinyl Alcohol 15
13 General Analytical Scheme for Resins 15
14 Quantitative Nature of Resin Degradation 15
15 Forensic Applications 16
16 Archaeological Uses 16
17 Art Frauds and Conservation 16
18 Headspace Analysis 17
19 Inverse Chromatography 17
Abbreviations and Acronyms 17
Related Articles 17
References 17

This work describes the analysis of coating materials using gas chromatography (GC). An element of volatility is necessary for samples to be amenable to the technique and early applications concerned the analysis of solvents and their mixtures. The examination of low molecular weight and volatile components of coatings followed; these included trace or residual products in resins. The resinous or binder components are essentially nonvolatile, and to be amenable to GC a considerable reduction in molecular weight is necessary. This has been achieved by chemical cleavage of condensation systems, with examination of the reconstituted reactants or their derivatives, or by thermal means in the absence of oxygen where pyrolysis provides fragments which may be related to the initial composition. Chemical cleavage was applied to alkyd resins, acrylics, polysiloxanes, polyurethanes and polyethers during the 1960s and while the results were most satisfactory, the time required was considered excessive. Several decades later, some of the problems have been eliminated but the procedures have found little application. Pyrolysis was also introduced at the same time and while the early reports were of low reproducibility, refinements in technique allow reproducible results to be achieved rapidly. The major area of the application of GC to coating materials is pyrolysis, where, in combination with mass spectroscopy (MS), valuable information is achieved in minutes.

Other coating applications of GC include inverse chromatography which has found negligible use and headspace chromatography which is widely used in determining the source of odors or taints in coated food packaging. Special applications of GC include use in forensic science where accelerants used by arsonists are readily detected and also examination and identification of paint flakes by pyrolysis GC/MS. Similarly pyrolysis techniques are used in conservation studies, with the analysis of old paints on art works.

1 INTRODUCTION

GC was employed early in coating analysis and the applications have continued to increase. GC is now probably the most widely used analytical tool of the coatings manufacturer. The technique was reported

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

in 1952, followed by reports on solvent analysis, in 1956, based on a presentation at an early international conference, in 1957, and in 1958 together with reports on monomeric plasticizers, drying oils, and other general works.[1]

While the analysis of coatings, including the use of GC, has been regularly reviewed since 1961,[2] few comprehensive works have appeared, although a monograph was published in 1974.[3] This work detailed packed column separation and tentative identification methods which were applicable at the time but have now been largely replaced by capillary column separation and spectroscopic detection. The development of GC in the analysis of coatings is best considered by study of the regular reviews.[2]

Spectroscopic, particularly infrared, spectroscopy and MS, and chromatographic techniques including GC, liquid chromatography and size exclusion chromatography, are the principal techniques used in coating analysis. The approximate molecular weight range applicable to these three major chromatographic techniques is shown in Figure 1.

For GC, the sample must possess some element of volatility. The efficiency of separation may be improved by modification of the compound studied. Many compounds contain highly polar groups and frequently exhibit poor chromatographic behavior, although by alteration of the chemical nature of the compound, the chromatography is usually improved. Derivative formation generally employs acetylation or esterification. The formation of a simple or trimethylsilyl derivative of an alcohol or acid increases the molecular weight of the compound but greatly reduces the polarity and increases the volatility. When polymers are used as coating binders, a considerable reduction in molecular weight, typically by several orders of magnitude, is necessary to provide fragments amenable to GC.

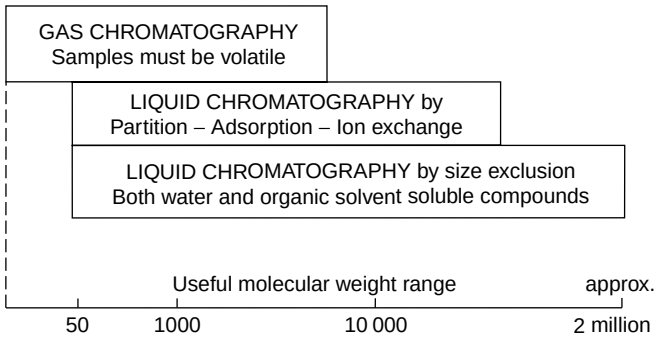

Figure 1 Approximate useful molecular weight range (Daltons) of the common chromatographic techniques. (Reproduced from J.K. Haken, *TRAC Trends in Analytical Chemistry*, **9**(1), 14–20 (1990) with permission from Elsevier.)

This reduction is achieved with condensation systems by chemical degradation with examination of the reconstituted reactants or their derivatives, or by thermal means, where, in the absence of oxygen, pyrolysis occurs, providing fragments which may be related to the initial composition. Chromatography is a separation method in which the components of a sample partition between two phases, one of the phases, the stationary phase, is a bed with a large surface area, the second phase is the mobile phase which in GC is a gas. The sample is vaporized and carried by the mobile phase or carrier gas through the column. Based on solubility, each component partitions between the two phases. The distribution constant is the major factor in the partitioning equilibrium and is the ratio of the concentration of a compound in the stationary phase to the concentration of the compound in the gas phase. The greater the solubility of the compound in the stationary phase the greater the distribution constant and the greater the retention or slower the elution.

This factor, formerly described as the solvent efficiency, and the column efficiency determine the resolution of chromatographic peaks. The column efficiency is concerned with the peak broadening as a sample moves along the column. The usual measure of column efficiency is the plate number, which is readily determined from a chromatogram but which varies for different types of compounds. Packed columns possess 500–2000 plates per meter, while typically a capillary column may exhibit 2000–4000 plates per meter. While a packed column may be 2 m in length, a capillary column may be 60 m in length, the total number of plates present being almost two orders of magnitude higher and the resolution accordingly very much greater.

Early reports of coating analysis tended to focus on stationary phase selectivity but the greater use of capillary columns has focused on thermally stable phases of modest polar character. These phases have found considerable use and have superseded many of the specialist selective phases which were of low thermal stability and of very limited application.

The most commonly used detector is the flame ionization detector, with the limits of detection being approximately 10–11 g L^{-1}, this being about two orders of magnitude greater than the thermal conductivity detectors. These have increased greatly in sensitivity and the unstable thermistors once used have disappeared. The other general detector used is the mass spectrometer, with detection limits of approximately 10–12 g L^{-1}; widespread acceptance has been achieved. The amount of sample required for analysis is minuscule; readily identifiable mass spectra are obtained with 40–60 pg of material. A variety of specific detectors, i.e. electron capture detectors and nitrogen–phosphorus detectors,

are in common use but have found little application in coating analysis.

2 SOLVENT ANALYSIS

The analysis of solvent mixtures was traditionally carried out by examination using chemical and physical testing of fractions obtained by fractionation. The initial report of Whitham[4] supplemented these methods with GC. Even the low resolution of packed columns made fractionation unnecessary, and very small amounts of sample were required for analysis. Elution depended on the stationary phase used, so that with a nonpolar phase such as dimethyl polysiloxane, elution followed a boiling point sequence, while with polar phases specific interactions occurred and elution of low boiling species could be retarded. Highly selective stationary phases allowed difficult separations to be readily made. Separation of the isomeric xylenes was reported in 1964 using a chemically treated bentonite clay. The boiling points of the isomers of xylene are *o*-xylene 144 °C, *m*-xylene 139 °C and *p*-xylene 138 °C, the isomers boiling at 138 °C and 139 °C being superimposed. With this similarity the isomers are eluted as two peaks on normal nonpolar and polar columns. When using a capillary column, unsatisfactory resolution occurs on a nonpolar column, but if the stationary phase has modest polarity three component peaks are resolved to the baseline.

When using packed columns, aliphatic and aromatic hydrocarbons, which are both groups of compounds of low polarity, are eluted in approximate boiling point sequence on low polarity stationary phases and are greatly intermingled. During the 1960s it was found that highly polar stationary phases including *N*,*N*′-bis(2-cyanoethyl)formamide (CEF) retarded the elution of aliphatic hydrocarbons, and the application to solvent mixtures[5] showed that the C_{13} aliphatic hydrocarbon, *n*-tridecane (bp 234 °C), was eluted well before benzene (bp 61 °C). While these stationary phases were of value in the petroleum industry, they were of limited use with coating solvents because CEF has a low maximum-use temperature and the determinations were made at 115 °C which is inadequate for the resolution of many hydrocarbon fractions. Separations of solvent mixtures have been carried out since the late 1970s using capillary columns with modest to polar stationary phases. The thickness of the coating layer of stationary phase is of importance; the thicker the film the greater the retention of the solvents. A film thickness of 0.25 μm is frequently used. If analysis is restricted to packed column operation, the peak overlap that frequently occurs can be minimized by conducting the analysis on several columns of different polar character or by the use of differences made by extraction with water, sodium chloride solution and sulfuric acid of varying strength.

Where the solvent mixture is free of resins and pigments, direct injection is obvious, but very often the solvents are present in liquid coatings and the treatment must be modified. Traditionally, steam distillation or dry distillation, often in the presence of added inert material, was carried out. With the widespread use of water-soluble or partly water-soluble solvents, steam distillation is of little value, similarly neither is dry distillation because of the minor degradation of resins and organic pigments before complete solvent removal is achieved and these materials produce disproportionately large spurious peaks. With packed and capillary columns, direct injection of coating has been practised, with nonvolatile material retained in inert injection liner packing. Direct injection of paint is not recommended, as the hot injector bakes paint in the syringe needle.

The solvents are conveniently removed from the coating prior to analysis. A technique and device which has been found effective is shown in Figure 2. The small device is based on the work of Haslam, Jeffs and Willis.[6] A sample of coating (0.1–0.5 g) is injected into the bulb using a coarse syringe, ensuring that the opening is not contaminated. The sample is frozen in liquid nitrogen and the opening sealed. After evacuation, the opening above the v-shaped receiver is sealed. The device is removed from the coolant and the receiver is cooled. The sample is placed in boiling water and the solvents are distilled off, condensing in the receiver. If the coating contains 30% solvent, 0.1 g of coating will produce hundreds of microliters of solvent for examination. A small amount of coating provides a generous amount of solvent for examination. When a coating is air-dried simply by

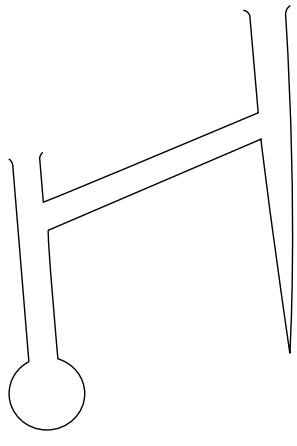

Figure 2 Apparatus for semimicrodistillation of solvent from a liquid coating.

evaporation or is cross-linked by a chemical reaction, the film initially contains solvent which is gradually lost and over time reduced by diffusion. The residual solvent in a film is readily determined by GC in association with low temperature pyrolysis. A filament pyrolyzer at temperatures 300–500 °C, or a Curie point pyrolyzer with a low-temperature element, may be employed, because little degradation of the constituent resin occurs at these temperatures. Alternatively the solvent may be determined by headspace GC.

3 MONOMERS

Like solvents, monomers form a significant part of coatings: the polymers formed providing the basis of many resins. Monomers like solvents are generally very volatile and are thus readily amenable to GC, which is a readily applied quality control tool. Both packed and capillary chromatography are used with a wide variety of stationary phases and column temperatures. Many monomers are extremely volatile and have low boiling points, so that the column temperatures required are low, often less than 100 °C. Analyses of these monomers are frequently directed towards the determination of small amounts of related materials, often polymerizable but with different reactivities, or nonpolymerizable or odorous which occur in different monomer deliveries, so that the resolution and consequently the column used are dependent on these factors. Other analyses concern residual monomer determination. Most resins have an indication of the specified maximum levels of residual monomer that they contain, i.e. frequently 0.1%, and determinations of headspace monomer content can be readily carried out. A sample of the resin or formulated coating is either heated in a sealed jar and a large sample (i.e. gas) injected into the chromatograph, or the sample is examined in a commercial headspace analyzer and gas chromatograph. In producing a counter or copy of a commercial material, headspace analysis will normally show what monomers have been used in the manufacture of the resin and valuable preliminary information can be obtained before pyrolysis. With both these latter applications, considerable resolution is required and a capillary column is indicated. When injecting a large gas sample, an appropriate split ratio is needed if a mass spectroscopic detector is used.

4 PLASTICIZERS

While monomeric plasticizers must be of low volatility to be effective, they are readily determined by GC. Packed column chromatography has allowed homologous aliphatic (i.e. succinates, adipates, suberates) and aromatic (i.e. phthalates) and phosphate plasticizers to be resolved. The use of capillary columns allows greater resolution, although these are frequently not required, as short packed columns rapidly allow determination of the plasticizer in a coating. High temperatures for separation are frequently unnecessary; many separations being carried out at 200–250 °C.

Many plasticizers, particularly ester types, are manufactured by esterification procedures and as it is not practical to carry the reaction to completion, it is necessary to remove residual alcohols, which give the product an unacceptable odor before sale. The alcohols are frequently mixtures of isomers which are of different reactivities and possess different odors. Separation of the components and any half esters may be effected on a long packed column (4–5 m, 1/8″) or more effectively on a capillary column. The analysis may be hastened using a wide bore (0.53 mm) capillary column. The separation of the low and high molecular weight species may be optimized using a multistep temperature regime. The alcohols may be resolved by an isothermal or temperature programmed operation, after which the temperature is rapidly increased in order to separate the esters isothermally or by temperature programming.

5 VEGETABLE OILS

The first report on GC concerned fatty acids and their esters, and since that time thousands of papers have appeared concerning these materials. Vegetable oils are readily amenable to GC and the complete oil may be examined as glycerides of molecular weight approximately 800, or much more commonly and successfully as esters, usually methyl esters. The oils or glycerides are subjected to hydrolysis or less frequently transesterification. A multitude of esterification procedures have been reported[7,8] and include reaction with methanol or higher alcohols in the presence of various catalysts, reaction with trimethylsilyl reagents or reaction with diazomethane. Use of the latter material has the advantage that the only product, apart from the methyl group, is nitrogen. Care must be exercised when using diazomethane because a number of accidents have been reported.

The unsaturated nature of vegetable oils is not normally lost during either esterification or GC, which may be carried out using either a packed or capillary column with either nonpolar stationary phases or more suitably with polar columns. The temperature of examination is usually within the range 200–250 °C. Early studies considered the examination of polymerized or partly polymerized

vegetable oils. These results were unsatisfactory and have continued to be unsatisfactory. Once the unsaturation has been lost by autoxidative polymerization, it cannot be restored. Early studies where the viscosity of vegetable oils was increased by heating have shown that the constituent oil may be readily determined by analysis of the unpolymerized portion. While such oils are now of no importance, partial and minor polymerization occurs during alkyd manufacture but the constituent oils are readily identified. Where a high degree of autoxidative polymerization has occurred, the constituent oils may be determined by gas chromatographic examination of the degradative fragments obtained after oxidative cleavage, as was frequently carried out more than half a century ago.

6 FISH OILS

These triglyceride oils are significant in countries that have a significant fishing industry and are used in coatings in the same manner as vegetable oils. Fish oils, as produced, contain large amounts of saturated oils, but these are largely removed before sale. Most vegetable oils are largely C_{18} fatty acids with up to three double bonds. Fish oils contain significant amounts of C_{20}, C_{22}, C_{24} and small amounts of C_{26} fatty acids and up to six double bonds.

Analysis follows the procedures for vegetable oils; the differences in fatty-acid chain length and unsaturation readily allow identification.

7 TALL OIL

Tall oil is a by-product of the Kraft Process for producing wood pulp from pine wood. Unlike other oils, the material exists as fatty acids and consists of approximately equal parts of long chain fatty acids and rosin acids. Where the Kraft process is carried out, use of tall oil in coatings occurs. Analysis as esters follows the procedures for vegetable oils; only a slightly higher temperature is required to elute the rosin esters.

8 AEROSOL COATINGS

Aerosol coatings form a lucrative niche market. The coatings are generally of a fairly basic type packed by contract aerosol fillers for the coating manufacturers and largely sold to householders. Much attention has been applied to developing containers which do not block easily and prevent efficient application. With aerosols, two types of GC analyses are possible, both for the propellant and for the coating. In the simplest case the aerosol container may be refrigerated and the contents discharged into a chilled container, providing in effect a headspace atmosphere. More suitably, the container is frozen, pierced and placed in a sealed pressure vessel fitted with an injection port. The propellant, which is either a hydrocarbon, a halogenated hydrocarbon or an inert gas, thaws first and is analyzed using a gas sampling syringe. The analysis generally detects some of the solvent but this does not present any problem. By weighing the intact container and the container on thawing, the percentage propellant is determined within a few percent. The thawed coating is transferred to a sealed container and analyses conducted as for a liquid coating.

9 COATING ADDITIVES

Many additive materials are used in coatings, usually in small concentrations, to introduce or contribute to a special property. Effective control of these additives is necessary and GC where applicable is ideal, although many of the materials are polymeric or single compounds of high molecular weight. The materials should usually be separated from the coating prior to analysis, and techiques include liquid extraction or supercritical fluid extraction.

Materials which are considered to be of low thermal stability may often be determined and include inhibitors such as hydroquinone, cathecol and resorcinol. Similarly promoters used in reactive polyester resins, such as *N*,*N*′-dialkylanilines which may be estimated directly and cobalt salts which require hydrolysis and estimation as methyl esters, may be determined. Driers are metal soaps and after hydrolysis, the methyl esters or, in some cases, the free acids, are examined.

Many coatings contain ultraviolet absorbers and several dozen of these may be estimated directly. Tin compounds are also used and these require conversion into derivatives. Tin compounds are also used as toxins in coatings and similar analytical techniques are applied. Antioxidants have been traditionally used with natural rubber and later with synthetic rubbers, and have found some acceptance with vinyl and polyolefins. Many antioxidants have been identified by GC.

Preservatives may often be subjected to direct chromatography, but in a film or on a substrate can be detected by low-temperature pyrolysis, the temperature being such that minimal degradation of the resin occurs while the more volatile components are expelled. Limitations to the use of such techniques are that before all the additive is liberated, significant pyrolysis of the resin has occurred.

10 GENERAL METHODS OF RESIN ANALYSIS

All of the materials mentioned so far have been of relatively low molecular weight and suitable for direct chromatographic examination or after simple derivative formation. The major component of a coating film is the resin, which is of high molecular weight and low volatility. To be suitable for examination considerable reduction in the molecular weight is necessary. This is achieved by degradation which may be of two types, namely thermal degradation or pyrolysis and chemical degradation. Pyrolysis is applicable to resins in general, because if sufficient thermal energy is applied then the various constituent bonds are broken. Among the strongest bonds are the C–C bond and the C–H bond, and the latter is stronger; the former is often broken. Unfortunately, rupture is random and a wide range of products is frequently produced. In condensation polymers the main polymer chain is not made up of C–C bonds but is composed of functional groups which may be cleaved by chemical means. With a few materials direct chromatography is possible; these are low molecular weight materials that are monofunctional and react into the final polymer. Epoxy diluents used to reduce the viscosity of epoxy products are identified by GC. Other related analyses are for the detection of impurities in resin-making ingredients such as trace materials in bisphenol A and for phenols, amines and halogenated materials.

Pyrolysis finds much greater application than chemical degradation and has the advantage that analysis is rapid, requiring effectively only the time required for the components to pass through the column, and the technique is applicable to all resin samples. The major disadvantage is that a multitude of components are produced and the results are usually not quantitative. Chemical degradation suffers from the disadvantages that it is only applicable to condensation polymers and the time required is much greater than for pyrolysis. The advantages are that a complete analysis, rather than an analysis of reaction products amenable to GC, is obtained and the results are normally quantitative. Pyrolysis initially required a much smaller sample than chemical degradation, but currently little difference in sample size for the two techniques is required. A recent modification of pyrolysis is the technique widely known as pyrolytic methylation developed by Challinor[9] and renamed thermally assisted hydrolysis and methylation. A small sample of resin or coating film, typically 5 μg, is mixed with 5 μL of tetramethyl or tetrabutylammoniun hydroxide, and subjected to Curie point pyrolysis at 770 °C. The results are qualitative – partial esters and ethers of the degradation products being formed.

10.1 Pyrolysis

Early studies utilized pyrolysis external to the chromatograph, but, currently, in situ pyrolysis with pyrolyzers attached or built into the chromatograph are universally used. Two basic types of pyrolyzer are available namely: (1) the continuous mode- or furnace-type where the resin sample is introduced into a heated microfurnace attached to the injection port of the gas chromatograph with the volatile pyrolysis products being rapidly swept into the column by the carrier gas and (2) the pulse mode-type where the resin is attached to the pyrolysis element which is rapidly heated to a predetermined temperature with the volatile pyrolysis products being swept into the chromatograph column. The pyrolysis element may either be a filament or ribbon of noble-metal alloy heated resistively or a ferromagnetic wire heated to its Curie point inductively. A range of ferromagnetic wires are available, with Curie points of 358 °C for a nickel wire to 980 °C for an alloy wire consisting of 50 : 50 iron and cobalt. Details of some commercially available ferromagnetic wires are shown in Table 1.

The requirements for pyrolysis GC to produce reliable and reproducible results were not known when the technique first attracted attention in the 1960s. Relatively large samples were used, not intimately in contact with the pyrolysis element and often with slowly increasing temperatures. It is now recognized that small samples, typically 2–5 mg, are desirable, although forensic casework has been conducted with 1 μg samples. The temperature of the sample should rapidly be increased and heat transfer should be maximized.

The temperature of pyrolysis is important and partly depends on the nature of the polymer. Polymers of high thermal stability or those heavily cross-linked require a higher temperature than does a simple thermoplastic. The bond strengths of the constituent atoms and the association of these bonds influences both the ease and type of degradation. In pyrolysis, stable free-radicals

Table 1 Curie points and composition of ferromagnetic wires

Metal	Composition (%)	Curie point (°C)
Ni	–	358
Fe–Ni–Mo	17 : 94 : 4	420
Fe–Ni–Cr	48 : 51 : 1	420
Fe–Ni	55 : 45	440
Fe–Ni	49 : 51	510
Fe–Ni	40 : 60	590
Fe–Ni	30 : 70	610
Ni–Co	67 : 33	660
Fe	–	770
Ni–Co	40 : 60	900
Fe–Co	50 : 50	980

are usually formed and the fragments are stabilized by rearrangement. Temperatures of 500–600 °C are regarded as minima, although lower temperatures, like those that use the lowest Curie point wire (i.e. 358 °C) are widely used to liberate loosely held additives or residues. An excessively high temperature is not recommended as increased amounts of low molecular weight compounds and gases are formed. It is not helpful to identify these products as they are often typical of organic compounds in general, being often formed by secondary reactions, rather than products of a particular polymer. Pyrolysis temperatures of 700–800 °C are suitable.

Pyrolysis GC of polymers has been the subject of several thousand reports and has been detailed in a number of monographs,[10–13] and in many reviews. Two recent reviews respectively detail almost 600 representative references to polymer pyrolysis, the majority being of polymers applicable to coatings,[14] and almost 500 references concerning coatings materials.[15]

10.2 Chemical Degradation

Chemical cleavage has been applied for decades and was used with various polymers before the development of GC. The best known example is alkyd resins, materials whose analysis was revolutionized by the use of chromatographic detection. Other early uses of chromatographic detection after chemical cleavage, normally hydrolysis, were the hydrolysis of polyurethanes, the determination of methoxy groups in cellulose, the degradation of polyethers and the acidic hydrolysis of polysiloxanes and linear polyamides. The microanalytical Zeisel reaction, where all alkoxyl groups were determined as methoxy groups, was transformed by the use of GC such that each individual alkoxy group is determined as the appropiate alkyl halide. The technique was first applied to acrylic copolymers and later to copolymers with other esters.

Prechromatographic reaction GC was extensively developed by the American chemist, Siggia, in 1972.[16] A microreactor was built from an old pyrolyzer and connected to the injection point of a gas chromatograph. Many students were employed in testing many types of compound including some classes of polymers.

A 1–10 mg sample was mixed with a prefused fusion reagent consisting of potassium hydroxide containing approximately 15% water, approximating the hemihydrate, with sodium acetate (1–10%) as a flux. The alkali was present as a substantial excess, i.e. 30–50 : 1, the fusion temperature was 250–350 °C and the reaction period 0.5–1.0 h. Reaction products amenable to GC were rapidly determined but high molecular weight reaction products and the half of the compound that was converted into potassium soaps remained in the reactor or were gradually carried into the column. All the alkali metal hydroxides have been used to achieve cleavage. The melting points of the common hydroxides are shown in Table 2. In addition to possessing a desirable melting point, potassium hydroxide melts have a greater solubility for organic compounds and the viscosity is lower.

Table 2 Melting points of alkali metal hydroxides

Hydroxide	Melting point	
	Anhydrous	Hydrate
Potassium hydroxide	360	125[a]
Sodium hydroxide	318	64.3[b]
Lithium hydroxide	417	–[b,c]

[a] Commercial potassium hydroxide contains approximately 15% water and is present as the hemihydrate.
[b] Present as the monohydrate.
[c] Decomposes to form lithium hydroxide and water.

External fusion using the same fusion reagent and comparable reaction conditions has been carried out with many polymer materials.[17,18] The sample used has varied in size, initially being 100 mg, but with developments in reactors a sample size of 2 mg is adequate. The advantages of external fusion are that the reaction products of the hydrolysis may be worked up and materials present as metal salts are converted into volatile compounds that are readily chromatographed. In this way complete analysis of the resin is possible. In addition the reaction products have been subjected to other chromatographic techniques, i.e. gel permeation chromatography and high-pressure liquid chromatography.

A recently described microfusion reactor[19] is shown in Figure 3 and consists of two parts, namely the reactor shell and the flange. The reactor shell is a square stainless-steel block containing a cylindrical cavity 23 mm deep and 20 mm in diameter. The flange has an orifice (1-mm diameter) in the center which allows entry to the cavity and combines with the main reactor cavity of the shell. The orifice is sealed with a cap fitted with a heat-resistant silicone rubber septum. A thin stainless-steel gasket seals the cover and the reactor shell. The cover is fitted to the reactor shell by four screws.

Degradation of condensation polymers has also been reported using acidic reagents, where cleavage of both ester links occurs in addition to cleavage of ether groups, the reactions having been reported individually and simultaneously. Most organic and mineral acids as shown in Table 3 have been used, but often with the formation of undesirable by-products. A reagent that has been used successfully is a mixed anhydride of *p*-toluenesulfonic acid and acetic anhydride. Other monoprotic acids have been used with highly cross-linked or thermally stable polymer systems and include trifluoroacetic acid and trifluoroacetic anhydride.

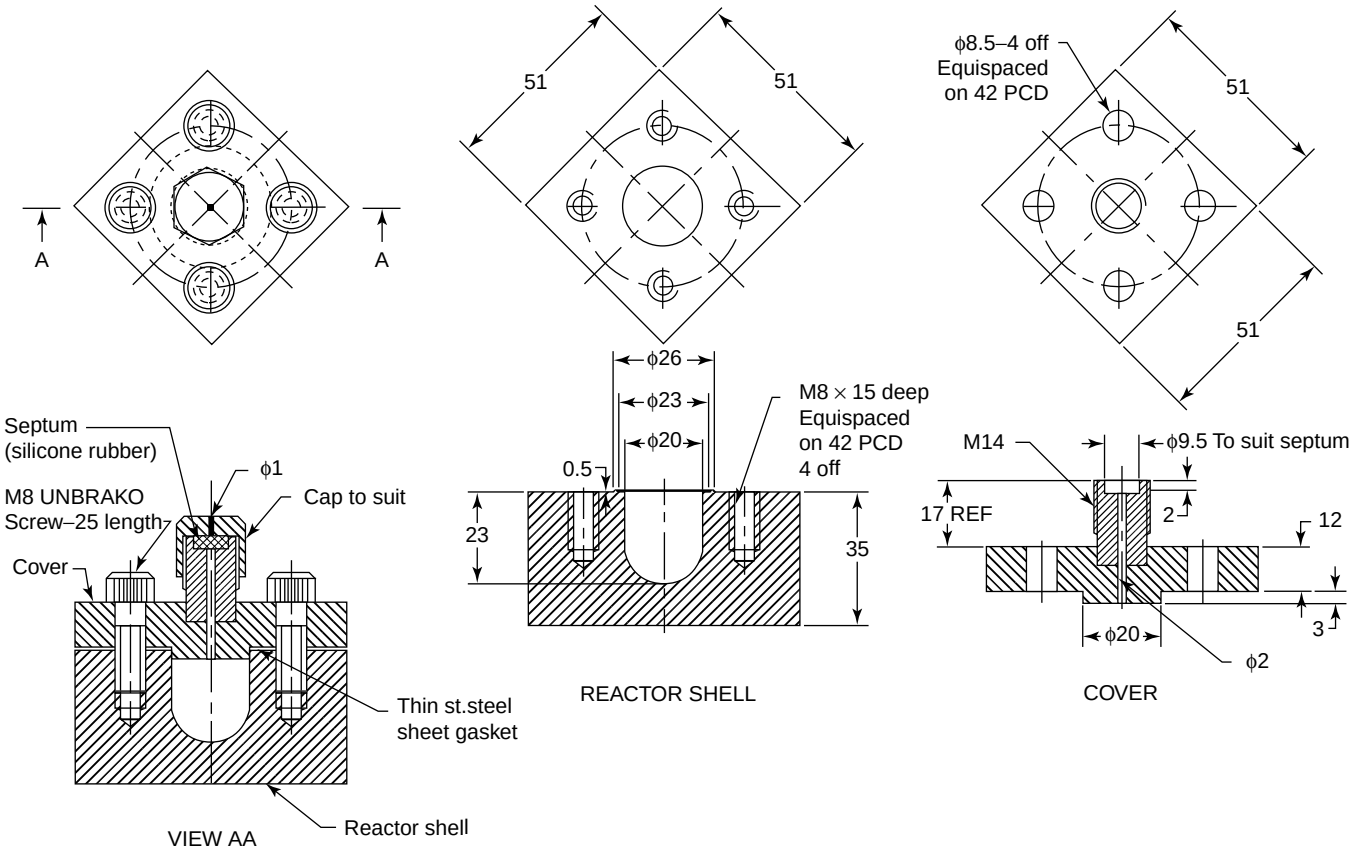

Figure 3 Design of stainless-steel semimicroreactor for prechromatographic fusion. PCD, pitch guide diameter. (Reproduced from J.K. Haken, P.I. Iddamalgoda, *J. Chromatography*, **600**, 352–357 (1992) with permission from Elsevier.)

Table 3 Acids used for chemical degradation

Hydrochloric acid
Phosphoric acid
Hydrobromic acid
Hydriodic acid
Sulfuric acid
p-Toluenesulfonic acid
Mixed anhydrides of *p*-toluenesulfonic acid and acetic anhydride
Mixed anhydride of acetic acid and acetic anhydride
Mixed anhydride of trifluoroacetic acid and trifluoroacetic anhydride

The chemical degradation of polymers has been described in several reviews.[17,18,20,21]

10.3 Pyrolytic Methylation

The coinjection of a free carboxylic acid and tetramethylammonium hydroxide was reported in 1979 and the technique was extended by Challinor a decade later.[9] The technique as described previously has been directed towards coating materials for forensic purposes. It has been postulated that the mechanism of the procedure concerning reaction of the polymer when intimately mixed with tetramethylammonium hydroxide and heated above 400 °C is hydrolysis, forming salts which undergo thermal fragmentation to the methyl derivatives.

The extent of application of pyrolytic methylation has been included in a review.[21]

11 CONDENSATION SYSTEMS

Condensation systems were developed earlier than addition systems – such systems in many applications having now been largely displaced. The majority of coating resins and many other specialty polymers remain as condensation systems.

11.1 Alkyd and Polyester Resins

These resins have remained the backbone of the coating industry for more than 50 years, and remain as the major tonnage film former despite the general acceptance of water-based systems for architectural finishing. The chemical analysis of alkyd resins dates from the

1930s and appeared as a standard specification[22] in 1940. The first step, hydrolysis, was retained, and gravimetric detection of the reactants was replaced by GC, with estimation of polyols and acids as acetates and methyl esters, respectively. Packed columns were first used, but separations were restricted. The three dimethyl phthalate isomers were resolved as two peaks on a non-polar column, with the *ortho*-ester being well resolved while the *iso*-esters and *tere*-isomers were superimposed, while on a polar polyester column two peaks result, but the *tere*-ester appears as a single identity. The three isomers have been separated by temperature programming on a cyanopropylpolysiloxane[23] packed column, but are readily separated by isothermal operation on a dimethylpolysiloxane capillary column. A wide variety of other acids are used in alkyd and polyester resins. The linear or vegetable oil free resins contain homologous aliphatic dicarboxylic acids, and while homologs are not well resolved by infrared spectroscopy, good separation is achieved using GC. Many resins contain small amounts of maleic or fumaric acids, and while it is not possible to decide which acid was initially present because thermal isomerization occurs, analysis is readily achieved. Fiberglass reinforced plastics (FRP) applications, where the ethylenically unsaturated (this term refers to the ethylene double bond in the maleic acid) acids copolymerize with styrene monomer, contain larger amounts of the acids. All polyester resins may contain small amounts of monobasic acids such as benzoic or *p-tert*-butylbenzoic acids which are used as chain regulators or terminators, and these are readily identified by GC as esters in the same determination as the major acids.

Polyester resins based on *o*-phthalic acid in the liquid form are readily hydrolyzed by solution hydrolysis, but with isophthalic acid or with chemical modification, the ease of hydrolysis is dramatically reduced, while as a cured film the reaction is minimal. Using vigorous hydrolysis with molten alkali, simple alkyd and polyester resins, both as liquids and as films, are readily hydrolyzed, as are FRP laminates where the acids, polyols and the fiberglass may be determined concurrently. Similarly polyesters cross-linked with aminoplasts are cleaved. The aminoplast is liberated and appears as a sludge except where butylated ureaformaldehyde resin is used and this is degraded.

This later hydrolysis is carried out with a trifluoroacetic acid and trifluoroacetic anhydride mixture, with reaction for 2 h. The urea forms carbon dioxide and ammonium trifluoroacetate while the polyester forms polyol trifluoroacetates, with *n*-butyltrifluoroacetate being formed from the butoxy groups. Other aminoplasts examined were melamine and benzoguanamine–formaldehyde resins which were cleaved from the polyester but not themselves cleaved. Figure 4 shows a series of chromatograms resulting from analysis of a soya oil alkyd cross-linked with 30% butylated urea formaldehyde resin.

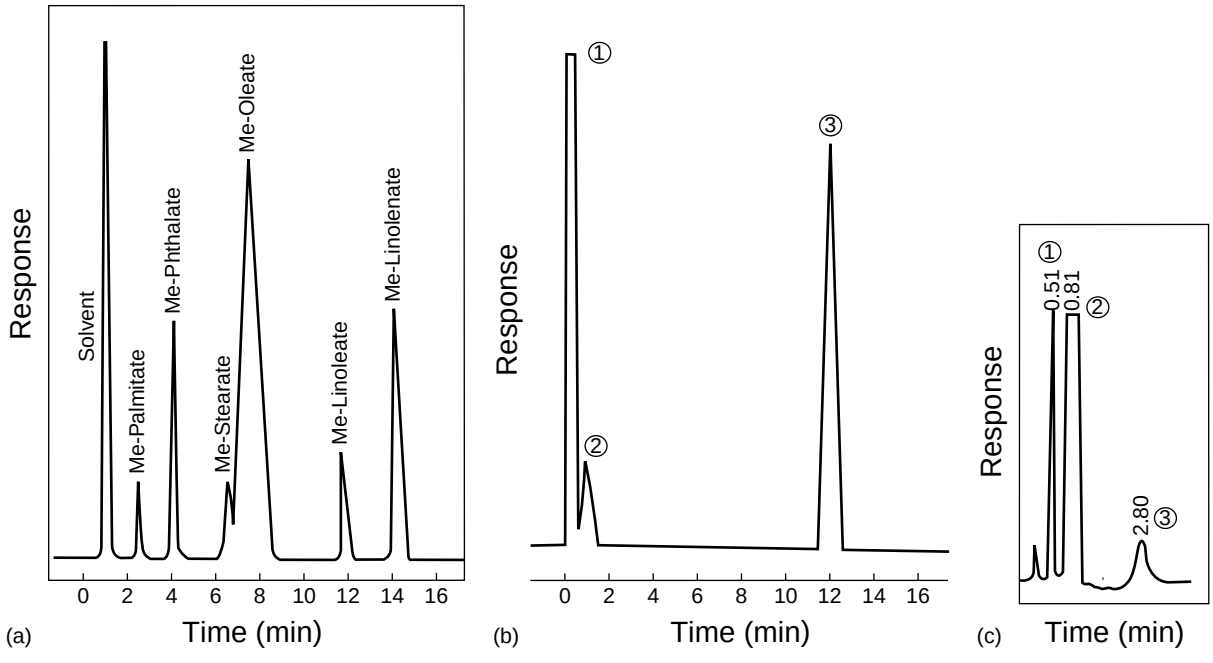

Figure 4 Analysis of a soybean alkyd cross-linked with 30% butylated urea formaldehyde resin. Chromatogram (a) methyl esters of fatty acids, (b) trifluoroacetates of polyols (1) solvent, (2) *n*-butyl trifluoroacetate, (3) glyceryl trifluoroacetate, (c) carbon dioxide and ammonium trifluoroacetate. (Reproduced from J.K. Haken, M.R. Green, *J. Chromatography*, **403**, 145–150 (1987) with permission from Elsevier.)

Pyrolysis of alkyd and polyester resins has been reported, but the results are relatively poor and vastly inferior to those obtained with other techniques, and, as such, the procedure is not recommended. If speed of pyrolysis is desired, pyrolytic methylation is superior to direct pyrolysis.

11.2 Silicone Polyesters

Silicone polyester resins are extensively used in coil coating enamels and are produced by simple condensation of hydroxy and methoxy functional, low molecular weight silicone compounds coreacted with polyesters.

The resins are readily cleaved using 2 mg of resin and 100 mg of the prefused potassium hydroxide reagent. The fusion was conducted for 1 h at 250 °C. The fusion was incomplete although qualitative results were obtained if the silicone polyester was examined as a film that had been cross-linked with an aminoplast. Total cleavage was achieved using trifluoroacetic acid and trifluoroacetic anhydride as fusion reagent. Figure 5 shows the separation of the products of degradation of a silicone polyester with simultaneous determination of the polyols, dicarboxylic acids and some species derived from the silicone portion of the polymer, all as trimethylsilyl derivatives.

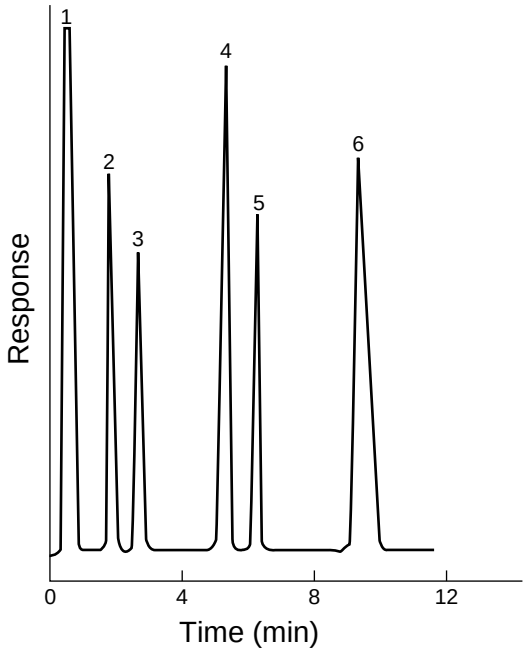

Figure 5 Chromatogram showing simultaneous separation of polyols, dicarboxylic acids and silica as trimethylsilyl derivatives. Peak 1, solvent; peak 2, neopentyl glycol; peak 3, a type of silica species; peak 4, trimethylolpropane; peak 5, adipic acid; peak 6, isophthalic acid. (Reproduced from J.K. Haken, N. Harahap, R.P. Burford, *J. Chromatography*, **440**, 329–332 (1988) with permission from Elsevier.)

11.3 Polyurethanes

The polyurethane molecule contains both carbonyl and amino functional groups and, while of considerable stability, is amenable to chemical degradation. In carrying out analyses, the level of residual isocyanate, which in most countries is subject to strict statutory limits, or the level of polyurethane composition, should be considered. Residual isocyanate may be determined by direct injection or by headspace analysis.

The polyurethane molecule is generally resistant to solution hydrolysis although systems containing long chain reactants have been successfully hydrolyzed. Vigorous hydrolysis using molten alkali gives rapid reaction, and cleavage has been obtained using reaction external to and within the gas chromatograph. Ether-type polyurethane cleavage is carried out with external fusion using alkali hydrolysis to liberate hydroxyl-terminated polyethers which are subsequently cleaved at the ether links by acid reaction to form low molecular weight fragments that are readily examined by GC. The molecular weight of the polyether may be determined on the isolated compound using size exclusion chromatography. The polyurethane may be chain-extended by amines or diols and both may be determined. Often the same compound is obtained from the polyol present and from the polyether, and care in interpretation is necessary. In addition to polyethers from ether polyurethane, the reaction products obtained are a diamine corresponding to the diisocyanate, acetates of the polyols present and methyl esters of dicarboxylic acids.

Pyrolysis GC is the subject of several dozen reports and is readily conducted at 700–800 °C. Early reports had some difficulty in differentiating between ester and ether-type polyurethanes but capillary column operation readily allows identification to be made. Pyrolysis GC of polyurethanes produces many fragments, is temperature dependent and never provides quantitative results.

The products are numerous and complex and include oxides of nitrogen, hydrogen cyanide, acetonitrile, acrylonitrile, pyridine, benzonitrile from the diisocyanate, carbon dioxide and specific fragments of the ester and ether components.

11.4 Epoxy Resins

Epoxy resins have been extensively examined by pyrolysis GC. The most widely used epoxy resins are based on bisphenol A or more recently bisphenol F. The major pyrolysis products are phenol, isopropenyl phenol and bisphenol A and if cross-linked with amino materials there are nitrogenous products. With bisphenol F the complex group joining the two benzene rings is absent and isopropenylphenol is not observed.

The epichlorhydrin does not produce any characteristic fragments.

Epoxy resins in the cross-linked state, while theoretically amenable to cleavage using vigorous conditions, have, on examination, produced unsatisfactory results. Pyrolytic methylation of epoxy resins in the uncross-linked state produces phenol, isopropenylphenol and the monomethyl ether of these compounds and of bisphenol A together with the diether of bisphenol A.

11.5 Polysiloxanes

Polysiloxanes are produced by the hydrolysis and polymerization of silanes. The materials have long been used as additives in coatings but are finding increasing use in the major film formers. In addition to use in silicone alkyds, combinations with epoxy resins are often used as heavy-duty coatings. Pyrolysis GC is readily carried out and a variety of methyl-substituted cyclic polysiloxanes are formed with dimethylpolysiloxanes. With low molecular weight products, cyclic products may be present before pyrolysis. If present, these low molecular weight oligomers may be determined seperately by direct GC.

The polysiloxane bond is amenable to chemical degradation – hydrolysis with molten alkali being essentially quantitative. A number of atoms and functional groups are attached to silicon atoms and such derivatives find considerable use as reactive intermediates used in coatings. The organic constituent may be cleaved from the polymer to produce low molecular weight compounds that are readily estimated. The presence of small amounts of other alcohol groups, vinyl or aryl groups may be detected in dimethylpolysiloxanes. With either external reaction or reaction in situ, in a reactor attached to the gas chromatograph, alkalis produce silicates together with ethylene from vinyl groups, ethane from ethyl groups and benzene from aromatic groups. Many other procedures have been developed to determine the pendant vinyl content of polysiloxanes using the mineral acids. Alkoxy groups present are converted to the corresponding alcohol. Table 4 shows pendant groups commonly present, cleavage reagents and reaction products which are readily determined by GC.

Table 4 Pendant groups and cleavage reagents for polysiloxanes

Pendant group	Cleavage reagent	Cleavage product
Alkyl	Potassium hydroxide	Corresponding alcohol
	Phosphoric acid	Corresponding alcohol
	Boron trifluoride	Corresponding alcohol
Phenyl	Potassium hydroxide	Benzene
	Phosphoric acid	Benzene
	Potassium hydrogen sulfate	Benzene
Vinyl	Potassium hydroxide	Ethylene
	Phosphoric acid	Ethylene
	Sulfuric acid	Ethylene
Alkoxy	Boron tribromide	Corresponding alcohol
Ethylene oxide	Potassium hydrogen sulfate	Acetaldehyde
Propylene oxide	Potassium hydrogen sulfate	Propionaldehyde
Glycerol	Potassium hydrogen sulfate	Acrolein

With alkali fusion, the silicates may be converted to derivatives suitable for GC. Trimethylsilation of the liberated silicates is achieved by reaction of the silycic acids with bis-silyltrifluoroacetic acid (BSTFA) and trimethylchlorosilane (TMCS).

11.6 Polyamides

The best known polyamides are the nylons, which are not used in coatings. The nylons and polyamides in general are of two types, α, w-aminocarboxylic acid condensates or diamine–dicarboxylic acid condensates. The dimer acid polyamides used as cross-linking agents in epoxy resin systems are of the latter type. The nitrogenous part of the polyamide pyrolyzes similarly to other aliphatic polyamides, producing a mixture of products including hydrocarbons, mononitriles, hydrocarbons and mononitriles containing an amide group, amines, dinitriles, cyclic ketones and lactams. The lactam fragment is a minor product as opposed to being the major or a major product of α, w-aminocarboxylic acid condensates. The pyrolysis products show many degradation fragments from the long-chain fatty acids, for example C_{36} with some C_{18} and some C_{54} mixed acids originally being present. The dimer polyamides are capable of pyrolysis at lower temperatures than nylons. When cross-linked with epoxy resins, the thermal stability is increased.

The polyamides are amenable to chemical cleavage by aggressive hydrolysis, and nylons which are of definite composition provide quantitative results. A small sample, i.e. 1–2 mg of polyamide, is reacted with 50–100 mg of potassium reagent at 250 °C for 0.5 h. The diamines are identified as trifluoroacetates and the fatty acids as dimethyl esters. Several dozen amines have been determined and are usually present in the polyamides as mixtures.

Aromatic polyamides and poly(amide–imides) are used in high-temperature-resistant coatings and have been subjected to both pyrolysis and chemical degradation, forming a wide range of products comparable to those of simpler homologs. Despite the thermal stability, the imide linkage is readily cleaved, with molten alkali generally forming esters of 1,2,4,5-tetracarboxylic acid

and 1,2,4- tricarboxylic acid, with *m*-phenylenediamine and 4,4-methylenedianiline being the usual amines.

11.7 Vinyl Esters

The vinyl esters incorporate some of the properties and reactions of polyesters, epoxies and methacrylates, namely glycerol, bisphenol A and methacrylic acid, and products representative of each material are obtained on pyrolysis. The vinyl ester may be more effectively cleaved chemically, since the formation of methyl methacrylate is essentially quantitative. The terminal methacrylate moiety is hydrolyzed with molten alkali finally producing methylmethacrylate and the ether polyol. The ether is cleaved with acetic anhydride to form glycerol triacetate and bisphenol A diacetate. When this procedure is used with vinyl ester laminates, like polyester–fiberglass laminates or coatings containing glass flakes, it allows estimation of silica as an organic derivative.

11.8 Phenol Formaldehyde Resins

The chemistry of phenolic resins is very complex and the resins developed more than a century ago have found use in a myriad of applications, all of which are relatively minor. As such, these resins have not been as extensively examined as many other polymers using modern analytical techniques.

Phenolic resins are produced by the reaction of phenol or *para*-substituted phenols with formaldehyde to form methylol groups on the phenol rings. Condensation polymers are produced by reaction of the methylol groups to form methylene bridges with the elimination of water. Either alkali or acid catalysis may be used, the former being a slower reaction allowing isolation of the methylol phenols, the latter being much more rapid, with a cross-linked polymer being formed more rapidly.

The resins are not amenable to chemical cleavage, but have been extensively studied by pyrolysis. Pyrolysis products include both gaseous and liquid fragments and gaseous products include methane, carbon dioxide and carbon monoxide. Rupture of the weaker aliphatic carbon–carbon bonds occurs predominantly and as there are three possible methyl phenols and 16 possible polymethyl phenols, the products are numerous. Studies show that the constituent phenol/s appear as the major or a major fragment.

Blocked or *para*-substituted phenols have been introduced, including *para-tert*-butylphenol, *para-tert*-amylphenol and *para*-nonylphenol. These reactants limit the possible reaction products and the phenols are evident in pyrolysis. Pyrolytic methylation has been used with the reaction products of these blocked phenols with formaldehyde. Pyrolysis of these resins produces the phenol and methyl and dimethyl phenols, while pyrolytic methylation produces the methyl ether of the phenol and the methyl ethers of the methyl and dimethylphenols, all of which elute before the three phenols.

11.9 Rosin Derivatives

Rosin is produced commercially as wood and gum rosin, the former from the stumps of trees, the latter from the gum of living trees. Solid rosin is approximately 90% rosin acids and 10% neutral material. Abietic acid, a monobasic acid containing a conjugated diene, is the major constituent, accompanied by a number of closely related acids.

Rosin is extremely reactive, forming esters with glycerol and pentaerythritol. The diene allows the Diels–Alder addition with maleic, fumaric and trimellitic anhydride, while oil solubility is achieved by reaction with phenol formaldehyde resins. Metal resinate soaps are formed by reaction with zinc oxide or calcium oxide. Cobalt, lead and manganese oxides produce resinate driers.

The pyrolysis of rosin and rosin products produces many fragments, the most characteristic of these being from rosin. The pyrolytic methylation of these products is considered superior, as multiple peaks of the rosin are obtained together with those of the reaction product. The major products of rosin are the methyl esters of abietic, dehydroabietic, pimaric and isopimaric acids. Esterified rosin glycerol produces the trimethyl ether, while pentaerythritol gives both pentaerythritol tetramethyl ether and pentaerythritol trimethyl ether. Polybasic acids form dimethyl maleate and dimethyl fumarate. Resins formed with *tert*-butylphenol and *p*-nonylphenol, formaldehyde and rosin, form characteristic rosin products together with methyl ethers of the phenol, of the methylphenol and of the dimethylphenols, all of which elute before the free phenols.

11.10 Cellulose and Derivatives

Cellulose is readily pyrolyzed and the major saccharidic degradation product is levoglucosan formed by scission of the 1,4-ether bridges between the structural units. This is followed by intramolecular rearrangement to produce a complex mixture of volatile low-molecular-weight compounds which include carbon monoxide, carbon dioxide, methane, ethane, ethylene, acetaldehyde, propionaldehyde, acrolein, methanol and water.

On pyrolysis, cellulose esters produce acetic acid from cellulose acetate, propionic acid from cellulose propionate and butyric acid from cellulose butyrate, in addition to products characteristic of cellulose.

When pyrolyzed, cellulose nitrate produces a complex mixture which includes nitrogen, oxides of nitrogen and carbon, hydrocarbons, aldehydes, esters and water.

Cellulose and its derivatives are readily susceptable to chemical reaction – the esters may be transesterified, hydrolyzed by solution or molten alkali hydrolysis, with the reaction products determined by GC.

Cellulose is partly alkylated to introduce solubility and the ethers are readily hydrolyzed, usually with dilute sulfuric acid, with the hydrolyzate being converted to methyl glucosides, using methanolic hydrochloric acid. Nine possible glucoside ethers are formed and all may be resolved using a short-packed polyester column. Hydroxyalkyl and carboxyalkyl celluloses are similarly susceptible to chemical degradation. These products, like methyl and ethyl cellulose, are usually used in small concentrations, and the analysis of resins and paints containing cellulose ethers, by either pyrolysis or chemical degradation, is very difficult.

12 ADDITION SYSTEMS

Addition systems are finding greater use in coatings. Water-based architectural coatings are addition systems, as are acrylic automobile lacquers and many ultraviolet-cured finishes. Some major groups of resins used in coatings are shown below.

12.1 Polyolefins

The polyolefins alone are not generally used in coatings, although polyethylene is commonly used as a coating on food containers, while polyolefins generally are widely experienced as comonomers in many copolymer systems.

Addition polymers commonly incorporate a carbon–carbon backbone and with the simplest polyolefin, polyethylene, the bonds are equivalent and rupture is random. The bond strength of the C–C bond is approximately 83 kcal mol^{-1} (347 kJ mol^{-1}) and of the C–H bond 94 kcal mol^{-1} (393 kJ mol^{-1}) and rupture of the weaker bond occurs. A large number of fragments result and the mechanism is described as random scission. The fragments produced have free radical ends and require stabilization. The fragment with a free radical may extract a hydrogen atom from an adjacent fragment and become a saturated end. Stabilization frequently occurs by β-scission where the induced free radical becomes an unsaturated molecule end. The process continues and produces a sequence of three hydrocarbons, the first saturated, the next with a double bond at one end, the third with a double bond at both ends. A series of *n*-alkanes, α-olefins and, α, w-diolefins up to carbon-number approximately 40 are formed, with the chromatogram showing an approximately normal distribution. In situ hydrogenation of the pyrolysis products of polyolefins is used, with hydrogen as the carrier gas and a precolumn containing a hydrogenation catalyst being inserted between the pyrolyzer and the column. In use the triplets become single peaks, with isocompounds of comparable structure forming single isoalkane peaks.

With polypropylene and other olefins such as isoprene, none of the C–C bonds are equivalent and a regular effect is not achieved. With polypropylene a series of diastereomers are formed and the polymer is synthesized with highly stereoregular structures – the various structures being readily detected by pyrolysis GC. Behavior is similar to polyethylene and a series of hydrocarbons with methyl branches is observed. The microstructure of polyolefins has been extensively studied and reviewed by several workers.[24,25]

Polybutadiene is found in styrene–butadiene latices and the diene polymers may be linked in either the 1,4-position or the 1,2-position. The former has *cis*- and *trans*-isomers and the latter has different tacticities. The prevailing structures may be resolved by pyrolysis GC.

Polyisoprene or 2-methyl-1,3-butadiene finds some use in metal primers. The thermal degradation of natural rubber has been extensively studied. At low temperatures, isoprene dimerizes to dipentene, but at higher temperatures, little more than identifiable fragments are produced, together with a wide variety of products. Isomers are found as with butadiene, but more products are formed due to the complexity of the methyl group.

12.2 Polyethers

Alkylene oxide polymers are frequently hydroxyl terminated and are widely used as reactants in polyurethanes, while similar polymers are used in the myriad of alkyloxylated long-chain surfactants that are available. Pyrolysis of these copolymers produces fragments characteristic of polyethylene, polypropylene and hybrid products of ethylene and propylene units. The ratio of ethylene and propylene is readily determined together with sequence distributions of the respective units. Pyrolysis with in situ hydrogenation leads to a simplification of the pyrogram – the saturated and unsaturated triplets of peaks forming single peaks.

Polyethers are susceptible to chemical cleavage; most mineral acids and some organic acids have been used for this. A few milligrams of sample are degraded by heating at 150 °C with a few milliliters of HBr in acetic acid to produce ethylene dihalide from polyoxyethylene products and propylene dihalide from polyoxypropylene products. A baseline separation of the ethylene and

propylene dibromides is achieved on a short packed column.

Polytetramethylene ether as glycol is also used in polyurethanes, and chemical cleavage forms 1,4-butanediol diacetate. *p*-Toluenesulfonic acid and the mixed anhydride with acetic anhydride are effective cleavage reagents, with fewer by-products being formed than often occurs with mineral acids.

12.3 Polyvinyl Chloride

Polyvinyl chloride homopolymer does not find use in coatings, except plastisols, but is widely used in copolymers. Polyvinyl chloride degrades by a side-chain scission mechanism. The bond strength of the C−Cl bond is 73 kcal mol^{-1} (305 kJ mol^{-1}), such that hydrogen chloride is readily split out leaving unsaturation in the chain. The unsaturated backbone fragments and produces aromatic compounds, of which benzene, toluene and naphthalene predominate. These fragments particularly are readily apparent in the pyrograms of copolymers because of their abundance.

12.4 Methacrylate Polymers

Methacrylate and acrylate polymers normally occur together and are generally termed acrylics. The two classes, while being close homologs, are treated separately, as the mechanisms of degradation are completely different. Methacrylates and other polymers possessing methyl substitution on thermal degradation revert almost completely to monomer by a process termed depolymerization, unzipping or monomer reversion. β-Scission with fragmentation occurs to form the monomer. The yield of monomer depends on the alkyl chain length, with methyl methacrylate and other short alkyl group polymers, the monomer yield is near 100%, but with long-chain polymethacrylates the yield is decreased due to some degradation of the alkyl chain. Pyrolysis GC of polymethacrylates is readily conducted and was first reported in the 1960s. Other polymers with α-methyl substitution such as α-methylstyrene also produce high yields of monomer.

The acrylic polymers are a class of vinyl polymers that are amenable to chemical degradation due to the presence of the pendant ester group. The backbone remains and analyses are of the pendant alkyl groups.

The Zeisel reaction has been used to determine the composition of acrylic copolymers. The pendant alkyl groups are estimated quantitatively as the alkyl halides. The resin is reacted with hydriodic acid in phenol and the halides formed are separated by GC. A packed polyester column resolves the methyl, ethyl and butyl esters using a low temperature, i.e. 70–75 °C. 2-Ethylhexyl acrylate is frequently used in coating terepolymers and this halide together with the lower homologs is readily separated using temperature programming. A limitation of the procedure is where the same alkyl acrylate and alkyl methacrylate are used – the two compounds are resolved together as the same alkyl halide. The procedure is applicable where acrylic and other esters are present.

The polymethacrylates are resistant to solution hydrolysis, but are readily cleaved by fusion with molten alkali to form the corresponding alcohol and the alkali metal salt.

12.5 Acrylates

The acrylate esters although producing less than 5% monomer on pyrolysis, do produce a series of saturated and unsaturated oligomers. Both series produce homologous saturated and unsaturated compounds, and these compounds have been detected as dimers, trimers, tetramers and pentamers and thus a particular acrylate ester produces many characteristic compounds.

The acrylates are also amenable to Ziesel-type cleavage and to cleavage with molten alkali as detailed above for the methacrylate esters. Longer chain polyacrylates have been degraded by solution hydrolysis but the digestion times required are extremely long and the analysis is considered unsatisfactory.

12.6 Polystyrene

Polystyrene is used in coatings both in condensation systems, i.e. as styrenated alkyds, and more extensively as a comonomer in addition polymerization systems. Polystyrene degrades by a complex mechanism, producing approximately 40% monomeric styrene depending on the temperature and the dimer and trimer. The dimer fraction is a complex mixture of structural isomers: about 20 peaks corresponding to the molecular formula $C_{16}H_{16}$ have been partly identified by MS.

12.7 Polyacrylonitrile

Polyacrylonitrile because of its insolubility is not used as a coating, but is widely used as a comonomer in combination with styrene, butadiene and vinyl systems. The polymer degrades on pyrolysis, the reaction products and their concentrations being dependent on the temperature. The products include ammonia, hydrogen cyanide, methacrylonitrile, acetonitrile and monomer. At lower temperatures cyanogen and vinylacetonitrile are also produced together with traces of benzene, toluene and pyridine derivatives. With polyacrylonitrile pyrolysis, some complex nitrogenous materials always remains. Pyrolysis GC with MS detection readily determines fragments characteristic of polyacrylonitrile.

Polyacrylonitrile is amenable to chemical cleavage. With hydrolysis with molten alkali, the nitrile pendant

group is cleaved essentially quantitatively and is determined as ammonia.

12.8 Polyvinyl Acetate

Polyvinyl acetate homopolymer is now rarely used in coatings, although copolymers containing 5–20% of acrylate esters are widely used, while copolymers with other esters find limited use. Of more recent origin are ethylene–vinyl acetate copolymers used extensively in adhesives. Polyvinyl acetate degrades by side-chain scission, with acetic acid being liberated and readily determined, together with benzene and toluene.

Higher homologs of vinyl acetate are available, and when degraded produce the corresponding acids which may be resolved concurrently.

12.9 Polyvinyl Alcohol

Polyvinyl alcohol, usually resulting from the hydrolysis of polyvinyl acetate and usually containing some unhydrolyzed ester, is widely used in emulsion systems. When pyrolyzed yields mainly water and acetaldehyde and usually acetic acid from unhydrolyzed polyvinyl acetate present. This type of degradation is typical of side-chain scission.

13 GENERAL ANALYTICAL SCHEME FOR RESINS

A complete analysis of many coating materials is a daunting and often impractical task, but with experience, the desired information can usually be obtained. A myriad of analytical techniques are available and for specific determinations a number of techniques are used. It is pointless to attempt an analysis using a single major technique. For coating analysis in general, three techniques are first applied, namely infrared spectroscopy, GC and liquid chromatography. All are techniques which are applied quickly and hence many hours of expensive time are not required, as with gravimetric analysis. The cost of the instruments needed for application of the three techniques is relatively low. The first step in a coating analysis is the determination of the infrared spectrum. If the resin is soluble in solvents, this uses transmission techniques, if insoluble or as a film, conventional reflectance or Fourier transform infrared techniques may frequently be used. From the spectrum the functional groups present are readily determined, and a strategy of further analyses may be determined.

Pyrolysis GC rapidly provides a pyrogram which may contain a few, although frequently many, component peaks. The peaks that are present or the absence of a particular peak or peaks provide much valuable information about composition.

Mass spectrometers are widely used and available in most laboratories. Attachment to the gas chromatograph allows detailed or sufficient information to be obtained about each pyrolysis peak as it is resolved. Bar charts, compilations and books of polymer pyrograms have been available since the 1960s, but are of little value as modern resins are extremely complex and the pyrogram often provides a huge amount of data. Programs are available or may be developed employing a database of pyrolysis GC data or the mass spectra produced to allow the identification of an unknown resin.[26] The database need only be restricted, as it soon becomes apparent that a sample type is not included. The database may be accessed with regard to a specific type of resin or may be quite general.

It is generally possible to determine the basic composition from the infrared spectrum of the resin and from identification of the pyrogram fragments. From the information available, the value of utilizing chemical degradation or other analytical techniques is apparent.

14 QUANTITATIVE NATURE OF RESIN DEGRADATION

With few exceptions, notably the α-methyl substituted polymers such as the polymethacrylates, pyrolysis is not quantitative, but often produces an array of products. The products are extremely valuable in identifying the resin and also in showing the resin structure. A simple mixture of equal weights of polyacrylate homopolymers does not produce the same pyrogram as a copolymer containing equal weights of the comonomers. The particular value of pyrolysis GC is the small sample required, the rapid time of analysis and the capacity for rapid identification using coupled MS.

Chemical degradation or cleavage has been found to be essentially quantitative; early studies showed that errors were introduced in the extraction and separation steps used in the derivative formation steps, but with elimination or reduction in the number of these steps, the errors are reduced or minimized. The sample size required is little greater than for pyrolysis, but the reaction time both for the degradation and the workup is much greater. A limitation of chemical degradation is that some of the initial reactants are mixtures and the reconstituted products or their derivatives are similarly mixtures. In resins, however, some of the functional classes are essentially pure and the overall composition can be established. Table 5 shows resin types that are chemically degraded,

Table 5 Resins Subjected to Chemical Degradation

Resin	Products identified
Dimer polyamides	Amines and trifluoroacetyl derivatives[a]
Alkyds and polyesters	Polyols and dicarboxylic acids as derivatives[b]
Silicone polyesters	Polyols, dicarboxylic acids and silica as derivatives
FRP polyester laminate	Polyols, dicarboxylic acids and silica as derivatives[c]
FRP vinyl ester laminate	Polyols, dicarboxylic acids and silica as derivatives[d]
Cross-linked alkyds	Polyols, dicarboxylic acids as derivatives, *n*-butanol[e]
Alkyl polysiloxanes	Hydrocarbon and silica derivatives
Aryl polysiloxane	Benzene and silica derivatives
Cellulose esters	Corresponding alcohol
Polyamides and imides	Acids and amines as derivatives
Cellulose esters	Corresponding acids
Cellulose ethers	Methyl glucosides
Polyethers	Alkyl dihalides
Polymethacrylates	Corresponding alcohol
Polyacrylates	Corresponding alcohol
Polyacrylonitrile	Ammonia[f]
Polyvinyl esters	Corresponding acid[f]
Butylated urea formaldehyde	Carbon dioxide, ammonium and *n*-butyl trifluoroacetate

[a] Dimer acids resolved as esters but quantitative analysis doubtful.
[b] Vegetable oils, not subject to autoxidation determined.
[c] Copolymerized maleic anhydride and styrene, not determined.
[d] Styrene, not determined.
[e] Aminoplasts cleaved off but not determined, except urea formaldehyde which is degraded.
[f] Pendant group cleaved.

together with the products isolated and limitations of analysis.

Pyrolytic methylation is related more closely to pyrolysis than to chemical degradation. The sample size, apparatus and time required are the same, but species not resolved or poorly resolved by GC are converted into derivatives more suitable for GC. The technique is temperature dependent and is not quantitative because a series of partial ethers and esters are formed in some analyses. The technique, however, greatly extends the value of pyrolysis.

15 FORENSIC APPLICATIONS

GC is an ideal tool for use in forensic science because only a small sample is required and the sensitivity is high. Numerous uses of GC exist in forensic science but two major applications concern the examination of coating fragments from accident or crime scenes and the examination of solvent residues from material used as accelerants in arson investigations. Coating examination has been the topic of numerous reports, the vast majority involving pyrolysis GC. Case work has utilized samples of 1 μg, but usually slightly larger samples are used. Both pulse mode and furnace type pyrolyzers are used. Pyrolytic methylation may also be used by the coatings analyst and has been extensively employed in the forensic examination of coating samples.

While automotive fuel or refined petroleum is widely used as an accelerant, the residue after ignition and/or after weathering due to the accentuated loss of the lower boiling components is such that the chromatogram of the sample bears little resemblance to the chromatogram of the initial fuel. Other accelerants include alcohols, where the denaturant is frequently concentrated and higher boiling petroleum fractions with comparable homologs including kerosene, white spirits and mineral turpentine. Lacquers, lacquer thinners, solvent-based wood stains and plastics have all been detected in accelerant investigations by GC.

Other related uses of GC concern the examination of nail lacquers, waxes, automotive body putties and fatty acids. The topic has been the subject of several reviews.[27,28]

16 ARCHAEOLOGICAL USES

GC, usually in association with pyrolysis and MS, has become a standard tool in the examination of coatings on ancient objects, including buildings, art works and funeral works. Organic media, including gums, waxes, naturally occurring resins and synthetic polymers, are frequently examined. Pyrolytic methylation is used as a replacement for simple pyrolysis – low volatility products being converted to esters or ethers and facilitating or allowing gas chromatographic analysis.

Fatty acids and triglycerides have been extracted from coating samples and from human remains that have dated from biblical times. Several reviews[29,30] have appeared, prepared by leading groups of researchers.

17 ART FRAUDS AND CONSERVATION

The investigation of the authenticity of art works and antiques is widely employed and some notable forgeries have been uncovered. Studies frequently employ pyrolysis and pyrolytic methylation and the techniques employed are analogous to those used in forensic and archeological studies.

In conservation and restoration studies the same techniques can be employed. It is often essential to have knowledge of the extent of prior restoration and the degree and type of the overcoat varnish before further restoration is commenced.

18 HEADSPACE ANALYSIS

Headspace analysis is the examination of an equilibrium mixture of volatile materials in a gaseous mixture, the procedure being widely used in the investigation of coating problems associated with packaging. Many food products, including nuts, chocolate and biscuits, are packaged in plastic containers, frequently clear but also in multilayer or laminated films. The packages are printed with inks or coatings which contain solvents, and similarly adhesives used to laminate films usually contain solvents, as do many clear plastic films.

Residual solvents often not detected by the nose, equilibrate on storage, are released into the usually sealed container, and are often transferred to the food, causing the food to be tainted or to assume an abnormal taste. The residual solvents may be readily detected by headspace GC. By the use of suitable controls it is often possible to determine if the solvent has come from the coating or from which coating in a multicolored system: the adhesive or the plastic film. Solvents not conventionally used in coatings, such as tetrahydrofuran or dimethyl formamide, have been found in food products and wrappers and are used in the manufacture of plastic films.

The technique may employ simple laboratory constructed apparatus, which may simply be a perforated metal lid on a screw-capped glass jar fitted with a silicone washer, or a lid with an injection port, or alternatively a dedicated commercially available headspace gas analyzer. Plastic lids and lids with polymer seals are to be avoided as these frequently liberate volatiles on heating and lead to spurious results. The glass vessel should be relatively small, typically 100–150 mL to avoid considerable dilution or the use of a large sample. The sample in the jar is heated in an oven to 100–200 °C for 0.25–1.0 h and a gas sample injected into the chromatograph. Conditions depend on the particular sample and it is essential that controls be used and that the sample and controls be examined under the same conditions. Capillary GC with mass spectroscopic identification of the products is desirable as peaks characteristic of the food concerned are often detected.

The technique has been the subject of several monographs.[31,32]

19 INVERSE CHROMATOGRAPHY

Many hundreds of reports have appeared concerning inverse GC, a variant of simple GC where the column is packed with polymeric or resinous materials, many of which are used in coatings. Simple molecules are used as solutes and while thermodynamic parameters are determined, the technique is considered to have little general utility and to be of negligible value to the coatings industry.

ABBREVIATIONS AND ACRONYMS

BSTFA	bis-Silyltrifluoroacetic Acid
CEF	*N*,*N*′-bis(2-cyanoethyl)formamide
FRP	Fiberglass Reinforced Plastics
GC	Gas Chromatography
MS	Mass Spectroscopy
PCD	Pitch Guide Diameter
TMCS	Trimethylchlorosilane

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction • Infrared and Raman Spectroscopy and Imaging in Coatings Analysis • Mechanical Properties of Solid Coatings • Microscopy of Coatings

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Pyrolysis Gas Chromatography in Forensic Science

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers • Pyrolysis Techniques in the Analysis of Polymers and Rubbers

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. S.T. Preston, G. Hyder, *A Comprehensive Bibliography and Index to the Literature on Gas Chromatography*, 1952–1962, Preston Technical Abstract Company, Evanston, Illinois, 1965.

2. M.H. Swann, 'Reviews of Applications of Analyses of Coatings', *Anal. Chem.*, **33**(5), 33R–37R (1961) and subsequent reviews.
3. J.K. Haken, *The Gas Chromatography of Coating Materials*, Marcel Dekker, New York, 334, 1974.
4. B.T. Whitham, *Application of Gas–Liquid Partition Chromatography to Solvent Analysis, Vapour Phase Chromatography*, ed. D.H. Desty, Butterworth, London, 395–412, 1956.
5. G.G. Esposito, M.H. Swann, 'Determination of Aromatic Content of Hydrocarbon Paint Solvents by Gas Chromatography', *J. Paint Technol.*, **38**, 377–380 (1966).
6. J. Haslam, A.R. Jeffs, H.A. Willis, 'The Examination of Mixed Solvents Obtained from Plastic Adhesives, Lacquers and Surface Coating Preparations', *J. Oil Color Chem. Assoc.*, **45**, 325–334 (1962).
7. V.Y. Taguchi, *Derivatization Reactions in Gas Chromatography: Biomedical and Chemical Applications*, ed. R.E. Clement, Wiley, New York, 129–177, 1990.
8. A.E. Pierce, *Silylation of Organic Compounds*, Pierce Chemical Company, Rockford, Illinois, 1968.
9. J.M. Challinor, 'The Scope of Pyrolytic Methylation Reactions', *J. Anal. Appl. Pyrolysis*, **20**, 15–24 (1991).
10. R.W. May, E.F. Pearson, D. Scothern, *Pyrolysis–Gas Chromatography*, The Chemical Society, London, 1977.
11. V.G. Berezkin, V.R. Alishoyev, I.B. Nemirovskaya, *Gas Chromatography of Polymers*, Elsevier, Amsterdam, 1977.
12. S.A. Liebman, E.J. Levy (eds.), *Pyrolysis and GC in Polymer Analysis*, Marcel Dekker, New York, 1984.
13. T.P. Wampler (ed.), *Applied Pyrolysis Handbook*, Marcel Dekker, New York, 1995.
14. J.K. Haken, 'Pyrolysis Gas Chromatography of Synthetic Polymers–A Bibliography', *J. Chromatogr. A*, **825**, 171–187 (1998).
15. J.K. Haken, 'Pyrolysis Gas Chromatography of Coatings Materials – A Bibliography', *Prog. Org. Coatings*, **36**, 1–10 (1999).
16. S.P. Frankoski, S. Siggia, 'Analysis of Carboxylic Esters using Alkali Fusion Reaction Gas Chromatography', *Anal. Chem.*, **44**, 507–511 (1972).
17. J.K. Haken, *Fusion Gas Chromatography; A Powerful Analytical Technique for the Analysis of Condensation Polymers in Advances in Chromatography*, eds. P.R. Brown, E. Gruskha, Marcel Dekker, New York, 177–231, Vol. 33, 1993.
18. J.K. Haken, 'Degradative Analysis of Reactive Coating Resins', *Prog. Org. Coatings*, **21**, 111–133 (1992).
19. J.K. Haken, P.I. Iddamalgoda, 'Reactor for Prechromatographic Fusion Reactions', *J. Chromatogr.*, **600**, 352–357 (1991).
20. L.R. Whitlock, S. Siggia, 'Fusion Reaction Gas Chromatography, Separation and Purification Methods', **3**, 299–337 (1974).
21. J.K. Haken, P.I. Iddamalgoda, 'Degradative Polymer Analysis by Chromatography', *J. Chromatogr. A*, **756**, 1–20 (1996).
22. American Society for Testing and Materials Gravimetric Analysis of Phthalic Anhydride in Alkyd Resins, ASTM D-631 1940T American Society for Testing Materials, Philadelphia, 1940.
23. J.K. Haken, M.A. Rohanna, 'Chromatographic Analysis after Chemical Degradation of Polyester Resins in Liquid and Cured Laminate Forms', *J. Chromatogr.*, **298**, 263–271 (1984).
24. S. Tsuge, H. Ohtani, 'Microstructure of Polyolefins', in *Applied Pyrolysis Handbook*, Marcel Dekker, New York, 1995.
25. D.H. Ahlstrom, 'Microstructure of Synthetic Polymers', in *Pyrolysis and GC in Polymer Analysis*, eds. S.A. Liebman, E.J. Levy, Marcel Dekker, New York, 1984.
26. M.J. Matheson, T.P. Wampler, L. Johnson, L. Atherly, L. Smucker, 'A Pyrolysis/GC–MS Library for Polymers', *Am. Labs* (Shelton, Conn), **29**(11), 24C–F (1997).
27. R. Saferstein, 'Forensic Aspects of Analytical Pyrolysis', in *Pyrolysis and GC in Polymer Analysis*, eds. S.A. Liebman, E.J. Levy, Marcel Dekker, New York, 339–372, 1984.
28. J.M. Challinor, 'Examination of Forensic Evidence', in *Applied Pyrolysis Handbook*, ed. T.P. Wampler, Marcel Dekker, New York, 207–241, 1995.
29. A.M. Shedrinsky, N.S. Baer, 'The Application of Analytical Pyrolysis to the Study of Cultural Materials', in *Applied Pyrolysis Handbook*, ed. T.P. Wampler, Marcel Dekker, New York, 1995.
30. A.M. Shredinsky, T.P. Wampler, N. Indictor, N.S. Baer, 'Application of Analytical Pyrolysis to Problems in Art and Archeology: A Review', *J. Anal. Appl. Pyrolysis*, **15**, 393–412 (1989).
31. B. Kolb (ed.), *Applied Head-space Gas Chromatography*, Heyden, London, 186, 1980.
32. B.V. Ioffe, A.G. Vitenberg, *Head-space Analysis and Related Methods in Gas Chromatography*, Wiley-Interscience, New York, 276, 1984.

Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Marek W. Urban
The University of Southern Mississippi, Hattiesburg, USA

1 Introduction 1
2 Attenuated Total Reflectance Spectroscopy 2
3 Reflection–Absorption Infrared Spectroscopy 5
4 Photo-acoustic Fourier Transform Infrared Spectroscopy 7
4.1 Photo-acoustic Experimental Methods 9
4.2 Continuous-scan Photo-acoustic Methods 10
4.3 Step-scan Photo-acoustic Spectroscopy Methods 11
4.4 Step-scan Photo-acoustic Spectroscopy Phase Analysis 11
5 Infrared and Raman Imaging 13
6 Summary 15
Acknowledgments 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

This monograph provides fundamental concepts utilized in infrared (IR) and Raman analysis of coatings as well as applications of attenuated total reflectance (ATR), reflection–absorption (RA), photo-acoustic (PA), and IR/Raman imaging techniques. This choice of techniques is dictated by the fact that these techniques alone or their combination allow one to obtain a three-dimensional molecular level representation of polymeric coatings, quantitative or semi-quantitative analysis, surface/interfacial depth profiling, and determination of structure–property relationships in coatings. As a result, the effect of molecular level interactions/processes on macroscopic properties can be elucidated.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The primary difference between IR spectroscopy of polymers and IR spectroscopy of polymers deposited on substrates is that while analysis of the former may be a relatively straightforward task, the latter is usually more challenging. This is because when a polymer is deposited on a substrate, usually referred to as an organic coating, more sensitive and selective spectroscopic analytical techniques are required. These analyses may often be accomplished by either removal of a polymeric coating from a substrate, or utilization of analytical techniques that overcome this requirement. Ideally, one would like to have a nondestructive means of analysis that would provide qualitative and quantitative information. Because such approaches require high-energy throughput, selected surface-sensitive IR spectroscopic approaches have been developed in conjunction with Fourier transform infrared (FTIR) and Fourier transform (FT) Raman spectroscopy.

The advantages of FTIR spectroscopy, in comparison to dispersive techniques, include the Jacquinot,[1] Fellgett,[2] and Connes[3] advantages. While the Jacquinot advantage arises from a greater energy throughput from the source, thereby improving signal-to-noise ratio,[4] the Fellgett advantage derives from the ability to collect all frequencies from a single scan. In contrast, dispersive instruments acquire the spectrum by measuring the intensity of one frequency at a time. The Connes advantage arises from the monochromatic laser used as a standard reference to effectively control the optical retardation, allowing co-addition of scans in an experiment. In essence, this feature reduces the error of mirror position resulting from a series of scans, therefore minimal spectral artifacts are produced by the co-addition of scans.

A Michelson two-beam interferometer, shown in Figure 1, is the basic design used in most modern FTIR spectrometers.[5,6] In this interferometer, IR light is split into two orthogonal beams by a beamsplitter, where one beam strikes a movable mirror, and the other impinges on a stationary mirror. Reflection of the radiation from these mirrors returns to the beamsplitter, where light interference due to recombination produces a collimated beam of given intensity directed toward a sample. Modulation of the IR light intensity, which generates an interferogram of intensity as a function of time, is accomplished by changing the optical retardation (i.e. by displacing the movable mirror position) of the light.

The interferogram is represented as the sum of all cosine or sine waves multiplied by a factor reflecting the intensity for all wavelengths from a polychromatic source during one movement of a mirror, that is one scan.[4] The relationship between light intensity for a

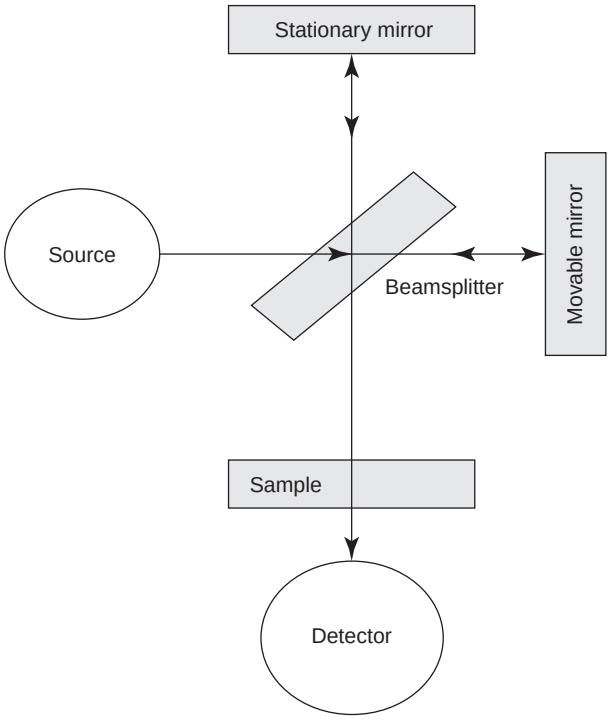

Figure 1 A schematic diagram of a Michelson interferometer.

given wavelength of light is expressed as Equation (1):

$$I(\delta) \propto \sum_{\lambda=-\infty}^{\lambda=\infty} I_\lambda \cos\left(\frac{2\pi\delta}{\lambda}\right) \quad (1)$$

where I is the light intensity, δ is the optical retardation, and λ is the wavelength of light. Conversion of the interferogram to the frequency domain is accomplished by moving the movable mirror at a constant rate, followed by multiplying the mirror velocity (in centimeters per second) by the optical retardation (in centimeters). Separation of the interferogram into its resolved cosine/sine waves is achieved by a FT, which expresses the intensity of light as a function of energy of IR radiation, which is an FTIR spectrum. Due to overwhelming advantages of this interferometric approach over dispersive methods, FTIR is used in surface and interfacial analysis of organic coatings.

In analysis of polymeric films and coatings there are three special directions of analysis. Figure 2 schematically illustrates x, y, and z directions and correlates the most desirable surface techniques with the direction of analysis. While the z direction may be effectively analyzed using ATR, RA and PA techniques, x–y directions can be imaged using FTIR and FT Raman microscopy/imaging in conjunction with surface selected methods. This choice is dictated by their ability to carry out a nondestructive analysis, and qualitative and possibly quantitative surface and interfacial molecular level analysis. The following sections describe principles and selected applications utilized in analysis of polymeric films and coatings.

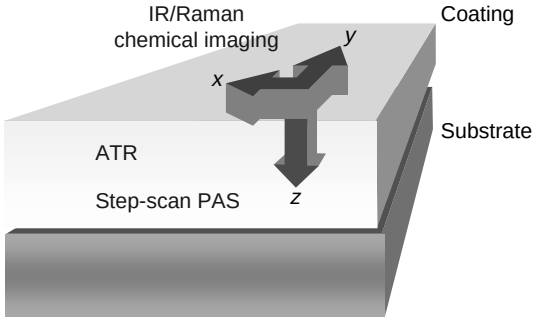

Figure 2 Surface sensitive techniques utilized in polymeric/organic coatings analysis.

2 ATTENUATED TOTAL REFLECTANCE SPECTROSCOPY

ATR belongs to a category of internal reflection spectroscopy which utilizes an IR transparent crystal of higher refractive index in contact with a sample to measure optical spectra. The intensity of IR light that passes through the crystal is dictated by the critical angle (θ_c), defined as $\sin^{-1} n_2/n_1$. Refractive index differences between the crystal (n_1) and sample (n_2) cause the light to reflect inside the crystal when $\theta > \theta_c$, and refracts out of the crystal for $\theta < \theta_c$. This is schematically illustrated in Figure 3. Expounding upon the internal reflection element, numerous configurations examining a wide variety of samples can be analyzed.[7]

When the sample absorbs a portion of radiation, the propagating wave is attenuated. ATR is a contact method (Figure 4a), where the IR beam is directed at an

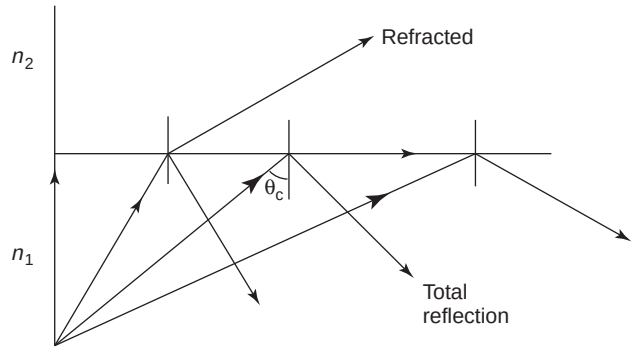

Figure 3 Light propagation and directions of reflection through two media with different refractive indices ($n_1 > n_2$). (Reproduced by permission from Urban.[7])

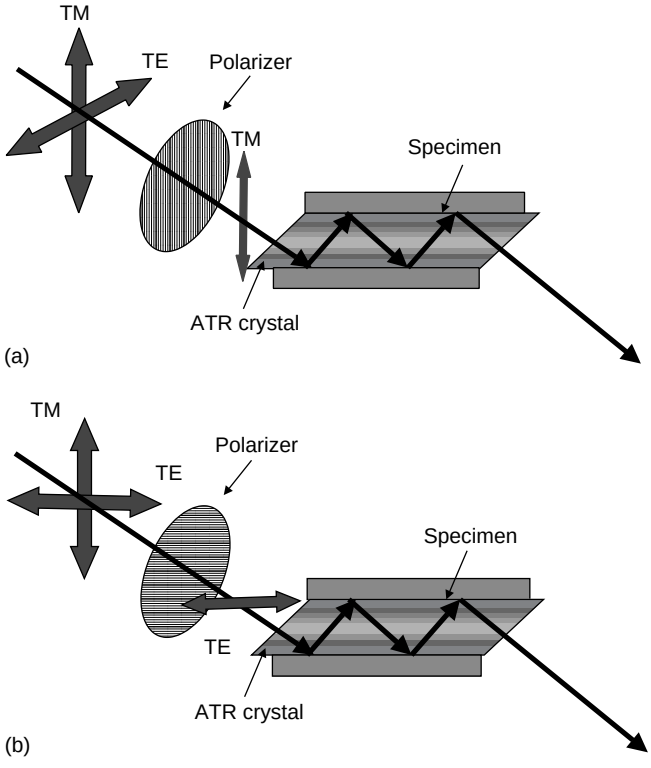

Figure 4 Propagation of light through an ATR cell: (a) transverse electric (TE) polarization; (b) transverse magnetic (TM) polarization.

incident angle (θ) through a crystal of refractive index n_1 onto a sample of refractive index n_2. Using polarized IR light, as shown in Figure 4(b), the orientation of surface functional groups can be determined.[8] The TE wave is defined as having its electric vector parallel to the plane, whereas the TM polarization has an electric vector perpendicular to the plane according to convention.[9,10] ATR is a surface-sensitive technique commonly utilized to determine surface information from polymeric samples.[7,11]

One of the main advantages of ATR is the ability to perform surface depth-profiling experiments from 0.2 to 5 μm. This is accomplished by variation in the angle of incidence and/or utilizing ATR crystals with different refractive indices. The penetration depth, d_p, is given by Equation (2):[7]

$$d_p = \frac{\lambda}{2\pi n_1(\sin^2\theta - n_{21}^2)^{1/2}} \qquad (2)$$

where n_{21} is the refractive index ratio of the sample and crystal (n_2/n_1), and λ is the wavelength of electromagnetic radiation. However, it should be kept in mind that although this relationship is quite useful, due to optical effects resulting from optical effects between the sample and the crystal, it cannot be used directly. Furthermore, because Equation (2) was derived for homogeneous samples, any variation of surface concentration or conformational changes will invalidate this relationship. The following sections describe ways to get around this problem.

ATR spectra are a complex function of absorption and refractive index spectra, which incur optical effects between the ATR crystal and sample as a result of contact. To account for these effects, ATR spectra are converted to optical constants, followed by conversion to Beer–Lambert absorbance spectra. While detailed description of the entire process is documented in the literature, a schematic diagram of the algorithm is shown in Figure 5, and it is based on a double Kramers–Kronig transform (KKT).[12,13]

Although the above correction methodology provides meaningful data, as indicated earlier, quantitative analysis in terms of surface depth profiling is limited because Equation (2) was derived for homogeneous surfaces. In other words, if there are surface concentration changes, this useful relationship, which allows the angle of incidence and the refractive index of a crystal changes to perform a surface depth profiling experiment, cannot be used. In an effort to account for this phenomenon, the following approach was developed. The surface is numerically sliced to form a stack of parallel thin homogeneous films. The films are homogeneous within each layer, but the layers among themselves are not. Thus, Equation (2) can be applied to each layer, and applying

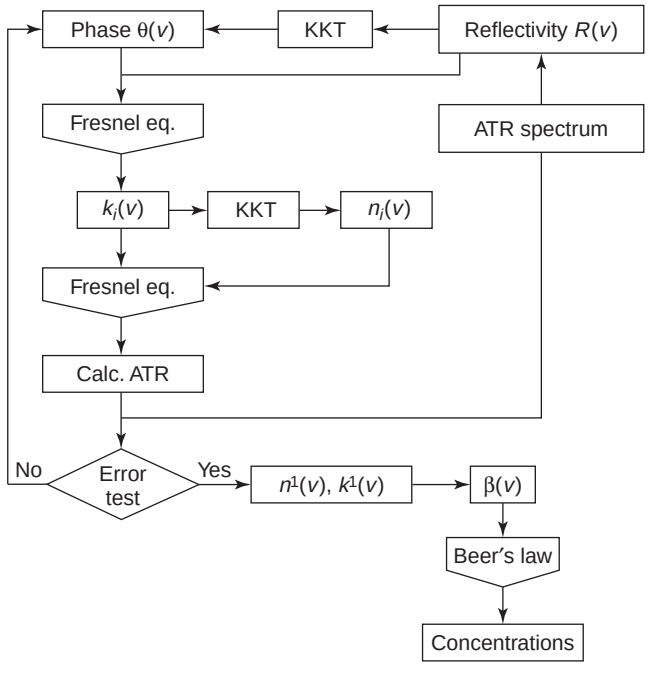

Figure 5 ATR algorithm utilized to correct ATR spectra for optical distortions. (Adopted from Urban.[11])

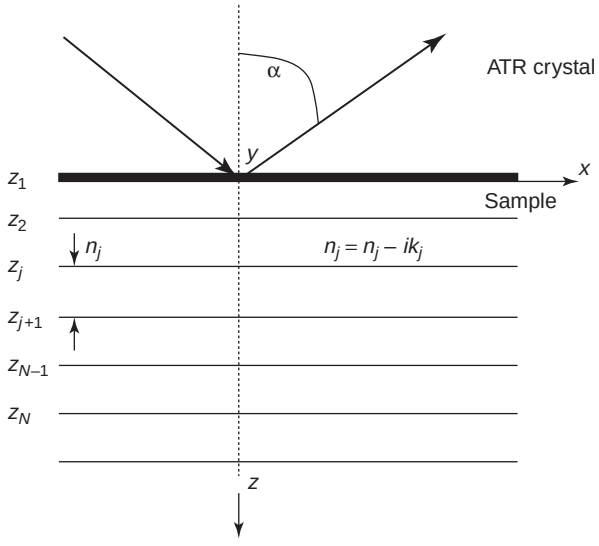

Figure 6 Schematic diagram of a step-wise analysis of ATR spectra. (Adopted from Urban.[11])

well-established reflectivity optical theory for each layer and step-wise treatments, one can obtain distribution of the species for nonhomogeneous surfaces. Figure 6 illustrates the reflection of light propagating in the $x-y$ plane at the boundary between ATR crystal and an N-phase sample stratified into z layers. Each layer has a thickness of h_j and refractive $n_j = n_j - ik_j$.

ATR surface depth profiling was used to elucidate stratification of organic coatings in numerous systems, including melamine-polyesters,[14] alkyds,[15] urethanes,[16–18] and latexes.[19–29] Using ATR FTIR quantitative aspects of polyurethane (PUR) formation, including stratification of individual components as well as kinetics of crosslinking reaction near surfaces and interfaces can be studied. For example, Figure 7 illustrates NCO concentration changes plotted as a function of reaction time for urethane reactions conducted at different humidity conditions. As seen, the relative concentration of the NCO groups changes not only as a function of time, but also as a function of depth from the film–air and film–substrate interfaces.

Quantitative aspects of ATR were fully explored in the studies of biological systems where thin, organic multi-layers were attached to poly(vinyl chloride) (PVC)[19] in order to chemically bond antithrombocytic layers of heparin. While the ultimate goal of these studies was to prevent blood cluding of artificial organs made out of PVC, one of the interesting aspects of this analysis was discovery that depending upon an application method,

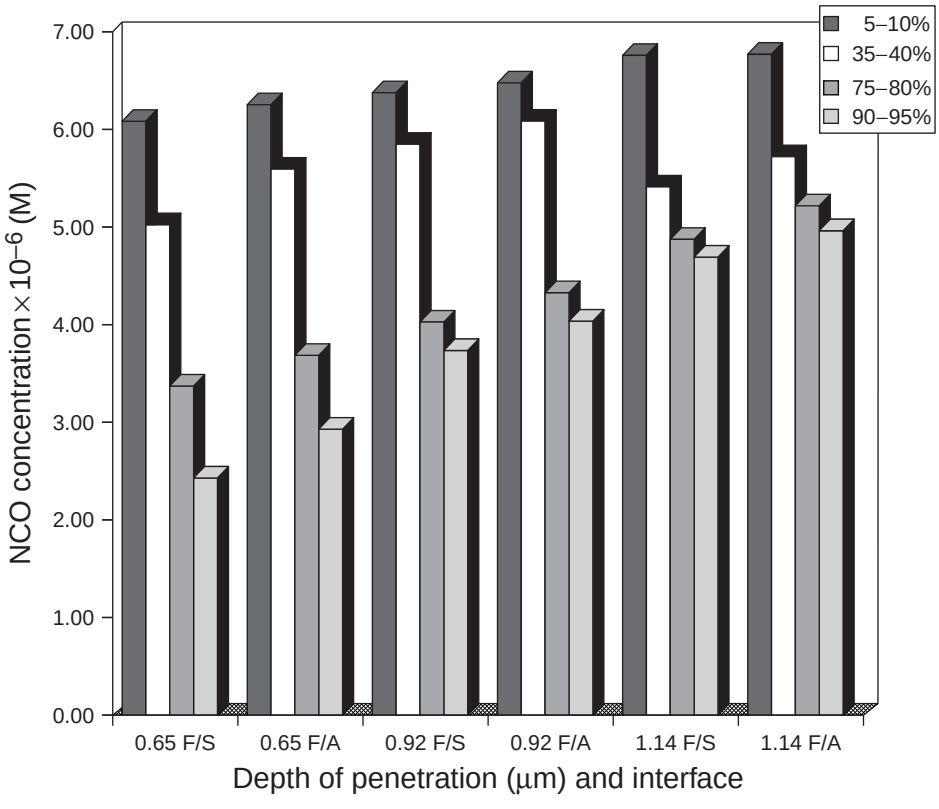

Figure 7 NCO concentration changes plotted as a function of reaction time for urethane reactions conducted at different humidity conditions.

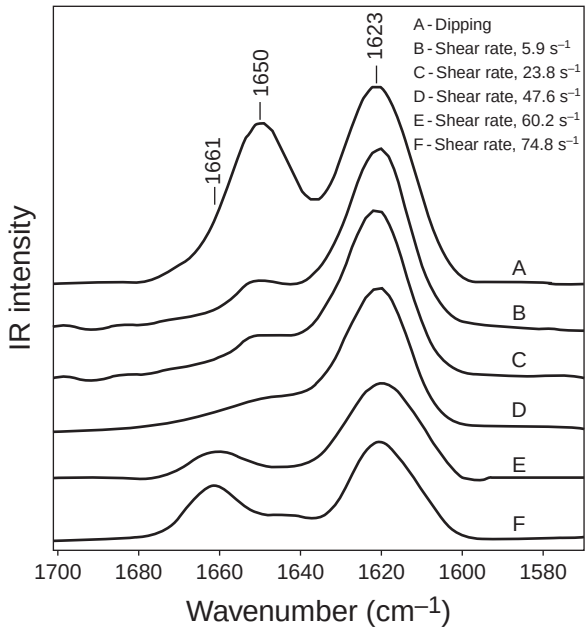

Figure 8 ATR FTIR spectra in the O–H bending region for heparin reacted with a PVC surface at various shear rates. (Reproduced by permission from Kim and Urban.[19])

heparin molecules may align parallel or perpendicular to the surface. As shown in Figure 8, application of higher shear rates results preferentially in parallel orientation of heparin to the surface. In these studies the ATR configuration illustrated in Figure 4(b) was used.[20–30]

3 REFLECTION–ABSORPTION INFRARED SPECTROSCOPY

While details concerning applications and theory of reflection–absorption infrared spectroscopy (RAIRS) can be found in the literature,[31] Figure 9 illustrates a schematic diagram of a typical RAIRS measurement configuration,

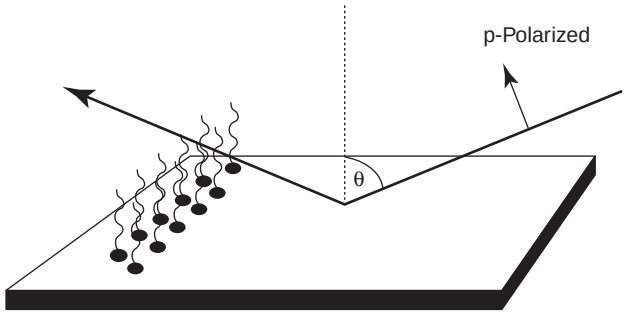

Figure 9 Schematic diagram of a typical RAIRS measurement configuration, where IR polarized light hits the surface of a metal substrate and reflects off. (Reproduced by permission from Claybourn.[31])

where IR polarized light hits the surface of a metal substrate and reflects off. In RAIRS experiments, only those dipoles that are aligned with the electric vector of the incident light, or with the direction of polarization, will interact with the light, resulting in an absorption band and providing information about the orientation of molecules at a surface. In contrast, for a thin film with randomly oriented molecules, all active dipoles will contribute to the spectrum. The earliest use of IR reflection spectroscopy to characterize adsorbed monolayers on metal substrates were measurements performed by Pickering and Eckstrom[32] and Francis and Ellison.[33] These as well as other measurements were made using a multiple reflectance approach as a means to enhance interactions with the thin film and give more intense absorption bands.[34–37] However, the optimum experimental conditions were established by Greenler,[38–40] who determined that only a single reflection was necessary for angles of incidence close to 88°. While Greenler[41] formulated the early theory for RAIRS and demonstrated the underlying physics that allowed us to understand the enhancement in the response from a thin surface layer for p-polarized measurements, the most recent monograph by Claybourn[42] provides an excellent overview of the current status of RA spectroscopy.

Figure 10 shows the expected reflectivity as a function of the angle of incidence for a three-phase system at 1000 cm^{-1}. The s-polarized component, R_s is very close to 1 because there are negligible interactions with the film. However, in the R_p curve, there is a minimum at a grazing angle close to 88°, that corresponds to a strong coupling with the coating. If one uses Greenler's description of an absorption factor A, defined for s- and p-polarizations as shown in Equations (3) and (4):

$$A_s = \left\{ \frac{R_0 R_s}{R_0} \right\} \tag{3}$$

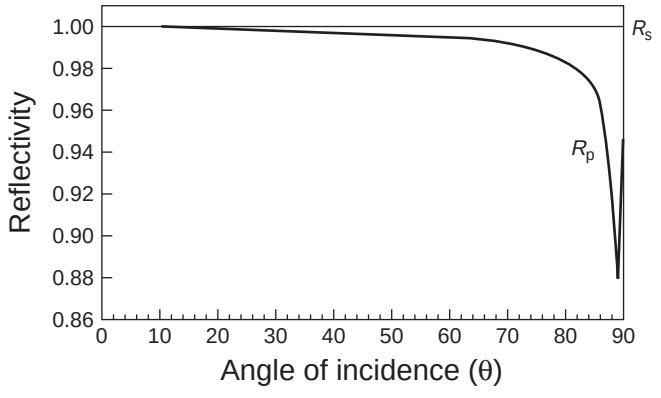

Figure 10 Reflectivity plotted as a function of the angle of incidence for s- and p-polarized light. (Reproduced by permission from Claybourn.[31])

$$A_p = \left\{ \frac{R_0 R_p}{R_0} \right\} \quad (4)$$

the absorption by the film can be estimated as a function of the angle of incidence. The meaning of R_0 and $R_{s,p}$ is shown in Figure 11(a). Figure 11(b) illustrates the absorption factor for s- and p-polarized light obtained using the same parameters for the reflectivity calculations depicted in Figure 10. From this simple calculation for a three-phase system, there is a strong enhancement only for p-polarization in the absorption factor for the surface film at grazing angles. If this is compared with the results shown for a 100 nm poly(methyl methacrylate) (PMMA) film in Figure 12 taken at normal incidence, no spectral signal is present, as one would expect from Figure 11(b). This is because $A_{s,p} \approx 0$ at this angle of incidence.

RAIRS can be used for studying the behavior of thin films on metal substrates, how they bond to the metal, structural ordering, and modifications with sample processing. Garton[43] examined the chemical binding of sulfonated polystyrene to a polished copper substrate. Figure 13 shows the RAIRS spectrum for this sample consisting of a 100 nm film of the polymer. The apparent

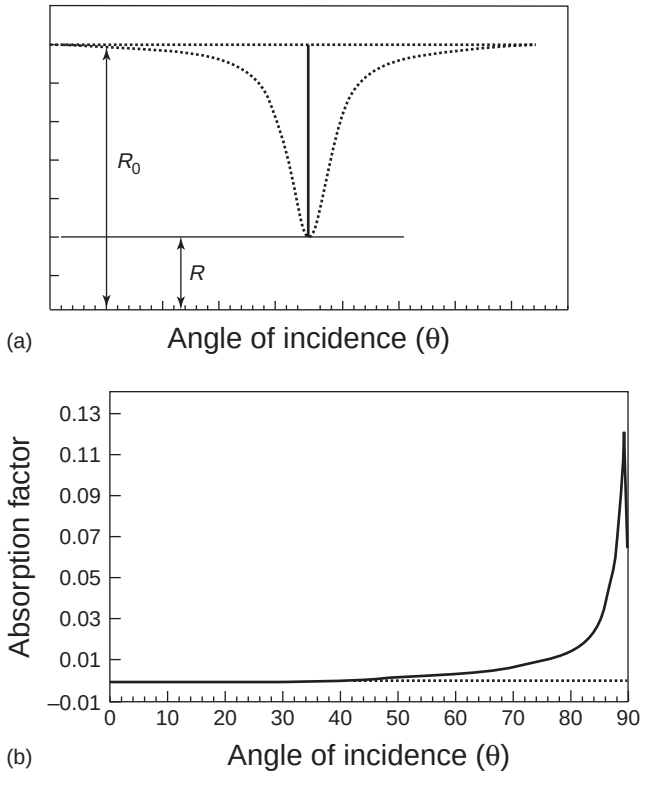

Figure 11 (a) R_0 and $R_{s,p}$ plotted as a function of the angle of incidence. (b) Absorption factor plotted as a function of the angle of incidence for s- and p-polarized light components. (Reproduced by permission from Claybourn.[31])

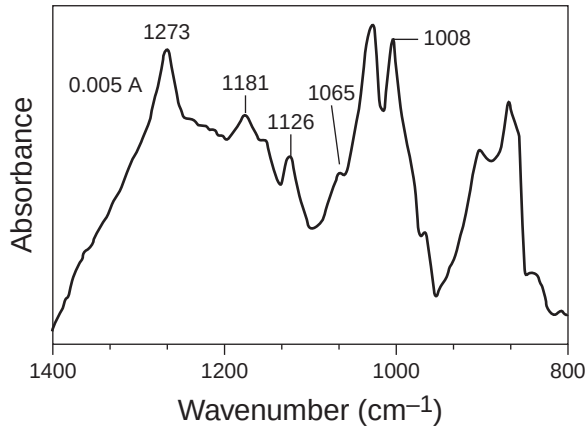

Figure 12 RA FTIR spectra of a 100 nm PMMA film recorded using p-polarized light. No spectral signal is present for s-polarized light because $A_{s,p} \approx 0$ at this angle of incidence. (Reproduced by permission from Claybourn.[31])

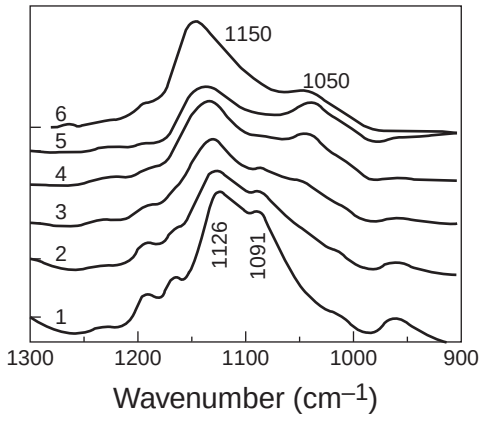

Figure 13 RAIRS spectrum of a 100 nm film of sulfonated polystyrene on a polished copper substrate. (Reproduced by permission from Claybourn.[31])

absence of the band at 1101 cm^{-1} due to the sulfonic acid groups is obvious, and the presence of the bands at 1008 cm^{-1} and 1126 cm^{-1} due to sulfonate results from the acid conversion to salt at the copper surface.

This technique can also be used for monitoring chemical and physical processes. Kurth and Bein[44] have used RAIRS for studying the hydrolysis and condensation reactions of siloxanes on gold. Figure 14 shows the RAIRS data for a film undergoing thermal processing, which initially forms low molecular weight oligomers followed by a low degree of crosslinking. This is expected because a monolayer would not give the same level of crosslinking as the bulk material, where the molecular interactions occur in three dimensions. Duevel and Corn[45] have used phase modulation (PM) RAIRS to study the amide and ester derivitization of alkanethiols on gold surfaces.

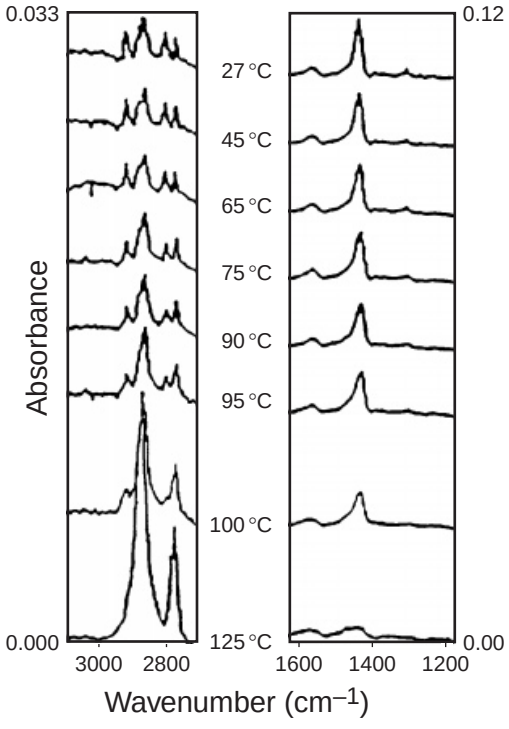

Figure 14 RAIRS spectra of hydrolysis and condensation reactions of siloxanes undergoing thermal processing on gold. (Reproduced by permission from Claybourn.[31])

Naselli et al.[46,47] have studied the thermal processing of cadmium arachidate films on silver substrates, and demonstrated that at room temperature the molecules are oriented, with progressive disordering of the molecules even below the melting temperature of 110 °C. This type of application can give insight into the structuring of polymeric materials at polymer–metal interfaces. RAIRS is potentially an important technique for characterization of binding of coatings to metal substrates.

The above examples illustrate the feasibility of RAIRS for characterization of thin films on metal substrates. Because not all samples are conveniently deposited on highly reflecting, optically flat metal substrates, it is also important to be able to utilize RAIRS in other sampling situations. Monolayers, in particular Langmuir–Blodgett films or self-assembling layers, have applications as boundary lubricants, such as component layers in electrical and optical devices. In these cases, the substrate is not a metal, and RAIRS has been used for studying thin films on semiconductors[48–51] and water.[52–56] In the case of water being a substrate, many polymers, surfactants, proteins, and other species are deposited on water, and such experiments require in-situ characterization. Dluhy and Cornell[57] first demonstrated the facility for characterizing monolayers at an air–water boundary using RAIRS. In this case, expected reflectivities for s- and p-polarized

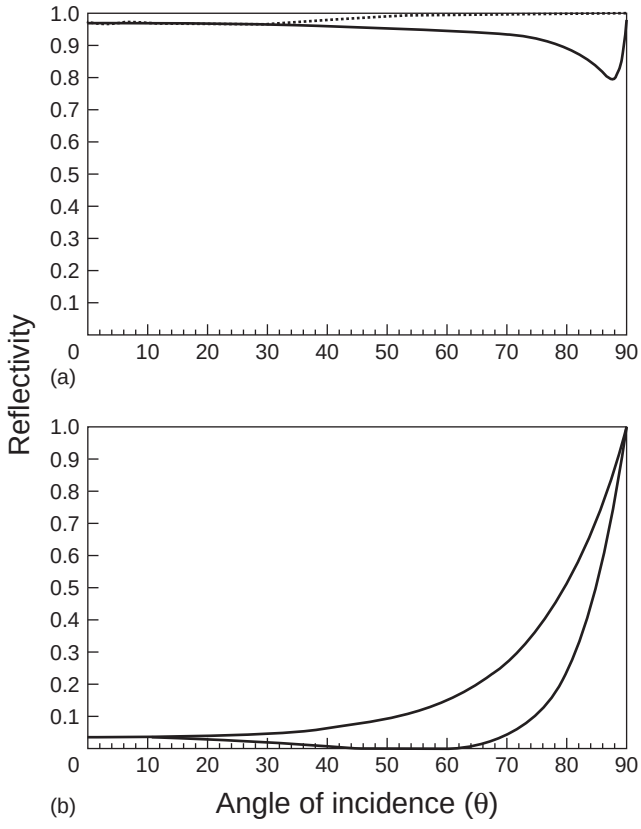

Figure 15 Calculated reflectivity on (a) metallic and (b) water substrates plotted as a function of the angle of incidence. (Reproduced by permission from Claybourn.[31])

light can be calculated from the Fresnel equations for a three-phase system.[31] If one considers a 100 Å film with $\hat{n}_2 = 1.5 + 0.1\,\mathrm{i}$ on water with $\hat{n}_3 = 1.44 + 0.03\,\mathrm{i}$, the expected reflectivities as a function of angle of incidence in the C=O spectral region are shown in Figure 15. For comparison, behavior of a similar film on a metal surface ($\hat{n} = 3 + 30\,\mathrm{i}$) is also illustrated. For metal substrates, both s- and p-polarized reflectivities remain high, with R_s close to unity for all angles of incidence, and R_p close to unity except at grazing angles where there is a sharp dip. Reflectivities for an air–film–water system are much lower, but rise sharply at higher angles.

4 PHOTO-ACOUSTIC FOURIER TRANSFORM INFRARED SPECTROSCOPY

The first PA experiment was discovered over a century ago by Alexander Graham Bell while investigating various forms of optical communications.[58,59] In Bell's experiments, the modulation of sunlight directed onto an absorbing sample produced an audible sound capable of

being detected by an observer positioned away from the apparatus. At the time, Bell recognized that his discovery may be used to investigate IR absorption spectra of various materials. Bell's initial experiments were later followed by Tyndall[60] and Röntgen,[61] who discovered that a gas enclosed in a cell can produce an acoustic signal when modulated by a light source.

The technique remained relatively unknown until its renaissance in the 1970s.[62,63] The development of FTIR, low-noise electronics, and improved detector sensitivity during the 1980s caused a renewal of interest in photoacoustic spectroscopy (PAS), which proved to be a unique and valuable analytical technique.

In PA FTIR, modulated IR light is absorbed by a sample, followed by nonradiative de-excitation processes converting light to heat. Thermal waves produced within a depth x travel to the sample surface, periodically heating an adjacent, inert gas atmosphere directly above the surface. A microphone within the gas atmosphere detects pressure fluctuations produced by this periodic heating of the sample, and subsequent Fourier transformation of the interferogram generates PA FTIR spectra. An illustration of a PA cell, along with frequently used parameters in a PA system discussed later, is shown in Figure 16.[64,65] For optically and thermally homogeneous samples, the light source attenuates by a factor A, at a distance x into the material as shown in Equation (5)

$$A = \beta(\bar{\nu})\mathrm{e}^{-\beta(\bar{\nu})x} \tag{5}$$

where $\beta(\bar{\nu})$ is the wavenumber dependent optical absorption coefficient. As a result, dampening of source radiation by a factor of e^{-1}, e^{-2}, and $\mathrm{e}^{-2\pi}$ occurs at distances of $x = l_\beta$, $2l_\beta$, and $2\pi l_\beta$ into the sample, where the optical absorption length, l_β, is defined by Equation (6):

$$l_\beta = \frac{1}{\beta(\bar{\nu})} \tag{6}$$

After the internal conversion of light to heat occurs within a sample, thermal waves propagate from a depth determined by the thermal diffusion length, μ_s, defined by Equation (7):[67]

$$\mu_s = \left(\frac{\alpha}{\pi f}\right)^{1/2} \tag{7}$$

where f is the Fourier frequency, and α denotes the thermal diffusivity of the sample, which is represented by Equation (8):

$$\alpha = \frac{\kappa}{\rho C_p} \tag{8}$$

where κ is the thermal conductivity, ρ is the density, and C_p denotes the heat capacity at constant pressure.

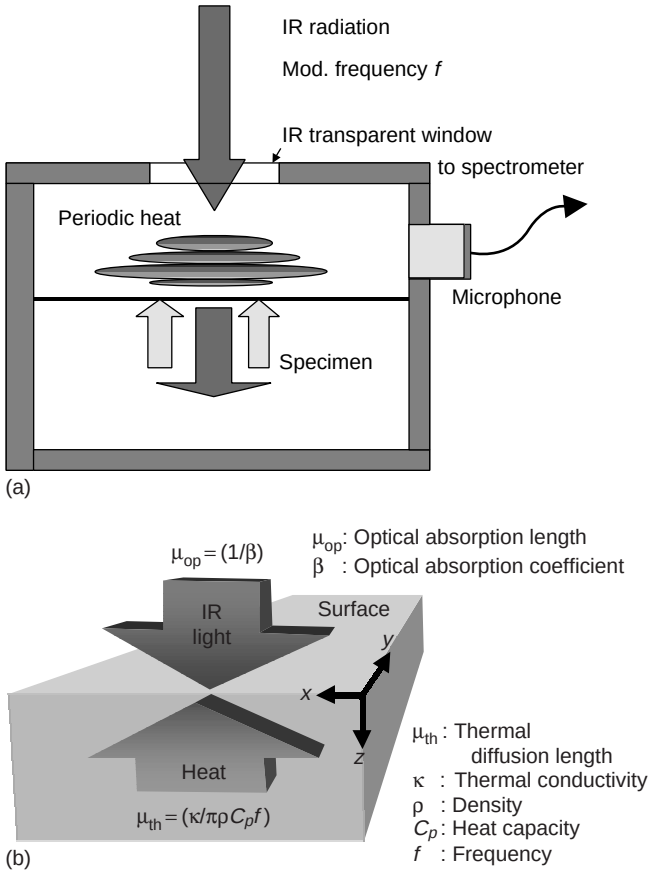

Figure 16 Schematic diagram of (a) a PA cell and (b) parameters involved in the signal detection.

Thermal waves evolving from a depth x are proportional to e^{-ax}, where the thermal diffusion coefficient, $a = \mu_s^{-1}$; therefore, thermal wave amplitude is reduced by a factor of e^{-1}, e^{-2}, and $\mathrm{e}^{-2\pi}$ at distances of $x = \mu_s$, $2\mu_s$, and $2\pi\mu_s$ from the surface.

Previously, numerous theories were proposed to account for the detection of the PA signal in condensed phase materials.[12,68] Rosencwaig and Gersho proposed that propagation of thermal waves through the gas layer from a sample produces an acoustic signal detected by a microphone, often referred to as the thermal piston model (TPM).[69] In this theory, the photothermal wave emanating from a sample surface can be considered a complex wave with real and imaginary components, $M\mathrm{e}^{-\mathrm{i}\theta}$, where M is the magnitude, and θ is the phase. This is depicted in Figure 17. McDonald and Wetsel[70] considered the composite piston model, where expansions and contractions of the sample produce a mechanical force on the gas atmosphere within the cell, in addition to thermal waves contributing to the overall PA signal. In the case for most polymer systems, mechanical deformations of the sample contribute somewhat to the PA signal due to

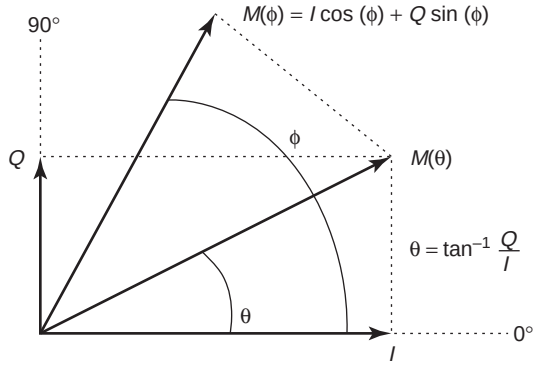

Figure 17 Phase rotational analysis of a PA signal.

low thermal expansion coefficients. Therefore, for quantitative purposes, the TPM is often used as a reference to account for various PA effects in polymers.[71–75]

Other theoretical approaches of PAS for nonhomogeneous samples have been derived from the TPM theory using the PA signal magnitude.[76–78] Grosse and Wynands.[79] used a general expression for the surface temperature to simulate a PA spectrum sample with any number of homogenous lamellae with various optical and thermal properties. This theory accounts for multiple reflections occurring within the sample chamber as well as thermal wave interference using matrix calculations.

4.1 Photo-acoustic Experimental Methods

Two modes of data collection for PA FTIR experiments are continuous-scan (CS) and step-scan (S^2), where the primary distinction between methods depends on whether the modulation frequency is varied, or constant for a given wavelength.[80] The expression for the Fourier frequency at a specific wavenumber is given by Equation (9):[4]

$$f = 2v\bar{\nu} \tag{9}$$

where v is the velocity of the interferometer mirror and $\bar{\nu}$ is the wavenumber. In CS PAS mode, the interferometer mirror moves at a constant velocity for all retardation points, resulting in a different modulation frequency for each wavenumber. Conversely, the mirror moves incrementally to each retardation point in S^2 PAS experiments, where application of a constant light frequency eliminates spectral multiplexing of the instrument often found when using CS PAS.[81]

The S^2 PAS experiments can be further classified, according to the method used to produce a modulated light waveform, as amplitude modulation (AM) and PM.[82] In AM experiments, a mechanical beam chopper produces modulated radiation, whereas PM utilizes oscillation of the interferometer mirror back and forth across a set point (light interference) for signal generation. Both methods effectively modulate the IR light source, but PM allows continuous irradiation, resulting in a higher signal-to-noise ratio. Moreover, AM experiments produce a large direct current offset, resulting in substantial noise contributions outside the zero-path-difference point of the interferogram, i.e. the ''wings'' on the sides of the centerburst region, creating significant spectral artifacts.

Equation (8) illustrates a straightforward method of sample depth profiling using PAS, which is obtained by manipulating the Fourier frequency dependence of μ_s. Figure 17 shows the spectra obtained at 0° phase shifted (I), and 90° phase shifted (Q), with respect to the incident radiation waveform, which can be mathematically processed to increase spatial resolution. The I spectra represent molecular information closer in proximity to the surface, whereas the Q spectra evolve primarily from the specimen bulk. Both spatially distinct spectra are easily extracted from a PA experiment by employing a digital-signal-processor or lock-in-amplifier.[83,84] With these mutually orthogonal component spectra, spatially dependent PA phase information, such as PA signal phase and phase rotation spectra, are made possible.

A physical interpretation of the PA signal phase is the time delay for thermal diffusion to occur, and the subsequent generation and detection of an acoustic wave. Layers deeper within a sample are resolved from shallower layers by the presence of larger phase shifts from the source radiation. The degree of phase shift can be measured using phase spectra, which are used to determine PA signal depth, especially when a nonabsorbing top layer is present in conjunction with an absorbing bottom layer at a given wavenumber.[85,86] Phase spectra are mathematically derived by applying an arctangent function on the I and Q spectra for all wavenumbers, as shown in Equation (10):

$$\theta = \tan^{-1}\frac{Q}{I} \tag{10}$$

Similar to the PA signal phase, phase rotated spectra employ both I and Q spectra as well. As shown in Figure 17, two signals are projected onto a specific angle, ϕ, to acquire angle-dependent spectra. Furthermore, phase spectra allow calculations of the PA signal phase delay with respect to the source, whereas phase rotation spectra determine the intensity of a band as a function of ϕ using Equation (11):[87]

$$M(\phi) = I\cos(\phi) + Q\sin(\phi) \tag{11}$$

Previously, few methods have explored the possibilities of using phase rotated spectra due to a limited knowledge of this angle specific information. Although several recent studies investigated the capabilities of depth-profiling

with phase spectra, there are discrepancies in this theory.[71,85,86] The major problems include collection angle determination, correlating data obtained from different frequencies, and determining spatial resolution of phase rotated spectra.

4.2 Continuous-scan Photo-acoustic Methods

Because early spectrometers lacked the sophistication of current instrumentation, CS PAS methods were more commonly used for data collection. Due to its ability to obtain spectra for highly opaque samples,[88–91] the first PA experiments established the usefulness of this technique for the characterization of polymers.[92–95] In spite of the fact that CS PAS methods were the earliest types of PA experiments performed, a significant amount of current research employs this technique.

During the 1980s, CS PAS was established as a viable method of polymer characterization with unique attributes, such as minimal sample preparation, and a wide scope of possible analytes, as opposed to other IR techniques.[96–98] Some of the earliest studies of PAS showed the depth-profiling capabilities of this technique for both ultraviolet/visible and IR regions.[99,100] One particular study displayed the ability to depth-profile layered polymer films by changing the mirror velocity of the interferometer, thereby changing the modulation frequency for each data set.[101] Ultimately, changing interferometer mirror velocity allows control of recorded surface or bulk information by manipulating the experimental thermal diffusion length, μ_s. In these studies, the spectra from a layered system of poly(vinylidine difluoride) (PVDF) on poly(ethylene terephthalate) (PET), recorded at mirror velocities of 0.15, 0.30, 0.60, 0.90, and 1.20 cm s^{-1}, showed the presence of the PET carbonyl contribution (1738 cm^{-1}) at low mirror velocities (greater thermal diffusion lengths).

A similar study, using the mirror velocity dependence of the experimental thermal diffusion length, investigated treatments of PET fiber with a poly(ethylene oxide) (PEO) block copolymer.[102,103] From spectra of the treated and untreated samples collected at four mirror velocities, the intensity of the CH stretch (2880 cm^{-1}) of the PEO copolymer increased with mirror velocity, demonstrating a higher concentration of PEO copolymer at the surface.

Applications of CS PAS have proven to be an effective method for determining component stratification from a study of thermoplastic olefins (TPO), which are extensively used in automotive applications.[104] These injection-molded polymers consist primarily of polypropylene (PP), with ethylene–propylene rubber (EPR), and talc filler. Characterization of various components were performed utilizing the IR bands at 2953 and 2849 cm^{-1} (C–H stretching region) for PP and EPR, respectively. By acquiring spectra at various mirror velocities from samples consisting of a series of PP and EPR ratios, a PA calibration plot was devised to determine the amount of each component at various depths.

PAS has also been used to investigate the diffusion of monomers in concentrated polymers, along with thermal relaxation processes near the glass transition temperature for amorphous polymers.[105] In this study, a mixture of ethanol (EtOH) and hydroxyl-terminated poly(dimethyl siloxane) were situated in a PA cell such that the IR probe light was mechanically blocked in order to determine the gas phase concentration of EtOH. The partition coefficient, J, of the EtOH in the gas phase and sample, and the diffusion coefficient, D, of the system were calculated by inspection of the 1065 cm^{-1} band due to C–O stretching normal vibrations. Thermal relaxation processes were analyzed by using poly(butyl methacrylate) and poly(vinyl acetate) at various temperatures near the glass transition temperature. These studies indicated that molecular dynamics cause changes in specific heat near the glass transition temperature, which, in turn, have an affect on PA intensity.

Novel applications of CS PAS were illustrated in rheo-photo-acoustic (RPA) spectroscopy studies, where IR measurements were performed on a stretched polymer sample to monitor the static stress–strain response.[106–108] For example, this method was used to determine diffusion rates of ethyl acetate (EtAc) in PVDF, which was shown to be a function of several variables including structural changes, polymer morphology, amount of applied strain, and EtAc concentration.[109] Further examination of the same system using RPA spectroscopy allowed quantitative calculation of diffusion coefficients of EtAc in PVDF, which were shown to depend upon the fraction of amorphous phase.[110]

Mongeau et al.[111] presented one of the first methods to separate the PA bulk and surface signals using vectorial relationships (Figure 18). In this study, it was determined that signals originating from the surface, S, and bulk, B, can be calculated by setting the phase channel equal to one of these signals. The technique employed CS PAS methods, therefore an inherent weakness of this approach is a consequence of the thermal diffusion length wavenumber dependence. This ultimately results in a discrete collection angle for each wavenumber in order to collect the appropriate vector component.

One of the recent studies showed that it is possible to determine the work of adhesion using stress–strain experiments inside a PA cell.[112,113] In the experimental set-up shown in Figure 19, substrate is elongated, and at the same time PA FTIR spectra are recorded. Interfacial forces responsible for the adhesion of coatings are

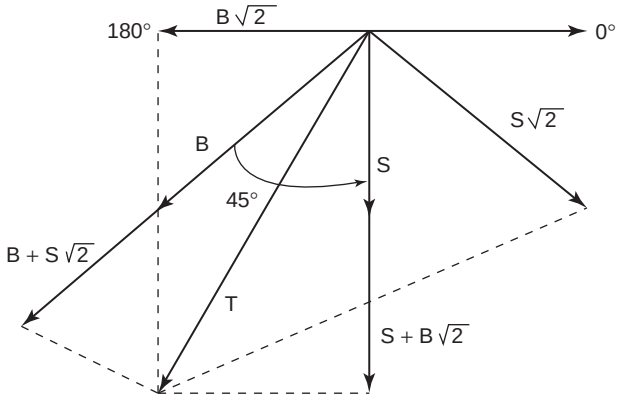

Figure 18 Vectorial analysis of surface (S) and bulk (B) PA signals. T, total.

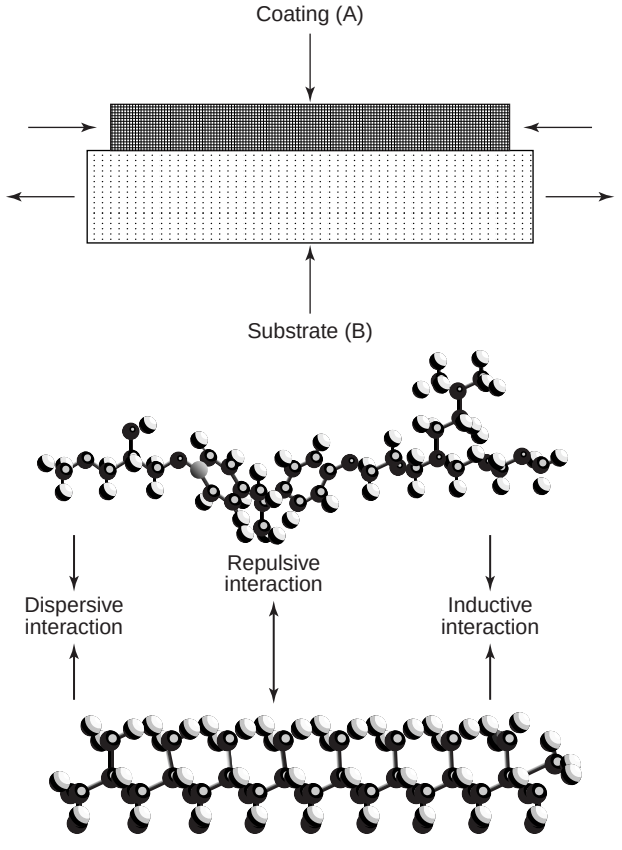

Figure 19 Schematic diagram of RPA set-up and interfacial interactions responsible for adhesion. (Reproduced by permission from Pennington.[113])

influenced by the nature and magnitude of the interactions between the coating and the substrate. When stresses are applied, these interactions are altered, resulting in potential energy changes within the neighboring layers of macromolecules forming an interfacial region, such as dispersive, repulsive, and inductive interactions illustrated in Figure 19. Therefore, vibrational energy and band intensities of the most outer species at the interface will change, which allows one to determine the work of adhesion.

4.3 Step-scan Photo-acoustic Spectroscopy Methods

The most recent developments in PAS resulted in S^2 PAS, where the ability to acquire spectra at the same sampling depth for all wavenumbers makes this method most appealing.[114–118] Similar to CS PAS, depth-profiling capabilities are possible, however in this case the modulation frequency is varied to obtain information from a specific depth. Studies performed by Palmer and Dittmar[66] showed the utility of this technique by applying a series of modulation frequencies to various samples. Most notably, these experiments illustrated an S^2 PAS method similar to Mongeau et al.,[111] where selective control of recording surface or bulk information is obtained by changing the experimental collection angle. For a PP sample surface treated with cobalt tetraphenylporphyrin (CoTPP) and hexamethyldisilazane (HMDSiN) plasmas, it was determined that a greater amount of surface information was acquired at lower collection angles, whereas higher collection angles showed more PP contributions.

A more recent study used a similar depth-profiling approach by changing the collection angle to characterize fibers.[119] In this study, a spectrometer frequency of 400 Hz was employed to determine absorbing layer depth of various polyamide bands due to the outer cuticle and inner cortex of a keratin fiber. The bands selected for characterization of the cuticle and cortex were the amide I band at 1650 cm^{-1} (CO stretching), and the 1540 cm^{-1} amide II band (CN stretch and NH deformation), respectively. It was determined that in order to calculate absorbing layer depth, a phase shift correction, which accounts for experimental conditions, such as sample and spectrometer characteristics, must first be subtracted from the experimental data.

4.4 Step-scan Photo-acoustic Spectroscopy Phase Analysis

An attribute unique to PAS in comparison to other FTIR techniques is the ability to extract additional sample information by performing PA phase analysis. Although this method was introduced in the 1970s, there is significant interest involving this technique,[120–122] because it allows extraction of spectral information from layers underneath the surface with no interference from the top surface layers. One useful example illustrating this capability is shown in Figure 20, which illustrates a model of PUR/PP sandwiched polymer films on which phase

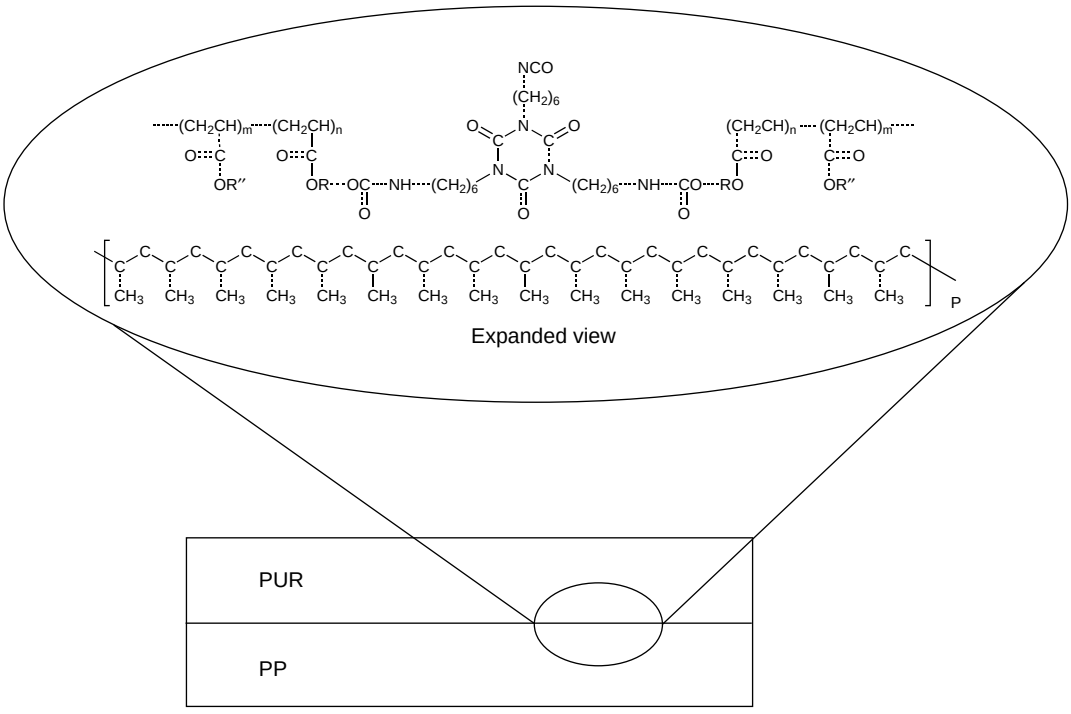

Figure 20 Model double-layer PUR/PP polymer sandwich.

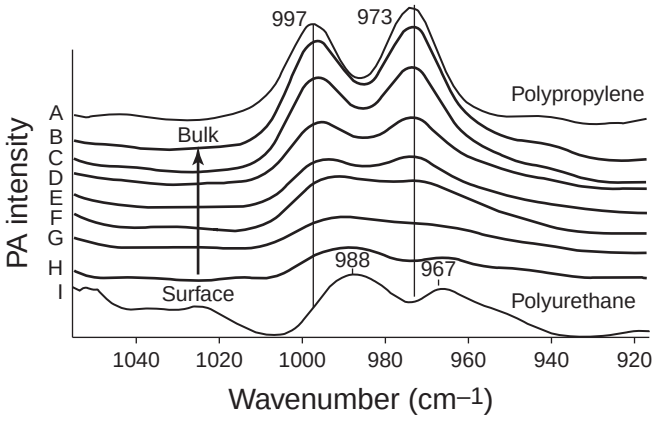

Figure 21 Magnitude S^2 PA FTIR spectra of PUR/PP in the 900–1000 cm^{-1} region recorded at various modulation frequencies: A, 23 Hz; B, 45 Hz; C, 91 Hz; D, 195 Hz; E, 488 Hz; F, 585 Hz; G, 632 Hz; H, 688 Hz; I, 732 Hz.

rotational analysis at different modulation frequencies was performed.[123] Figure 21 illustrates a series of PA FTIR spectra recorded at specific modulation frequencies and it is quite clear that as the modulation frequency increases, more PUR is detected, as demonstrated by the appearance of the enhanced intensity of the 988 and 967 cm^{-1} bands. With this in mind one can choose desired modulation frequencies and collect information at two orthogonal detector channels referred to as in-phase (0°) and in-quadrature (90°). This represents conceptually a surface and a bulk signal, respectively, and Figure 22(a) and (b) illustrates this process conducted on PUR/PP layers. In this case the spectra were recorded at 732 and 23 Hz modulation frequencies. Since the bands at 998 and 997 cm^{-1} are attributed to the PUR layer, while the bands at 997 and 973 cm^{-1} are due to PP, their intensity changes will represent the contribution of each layer to the spectra. As seen, the I and Q spectra recorded at 732 Hz exhibit the presence of PUR and PP bands, indicating that the phase rotational analysis provides spectral information from PP (Q spectrum). However, when the modulation frequency is lower, indicating that the penetration depth will be higher, the same analysis results in the presence of the PP bands only. Thus, the entire spectral information comes from PP, and there is no interference from the top PUR layer.

One of the approaches to PA phase analysis was developed by Jiang et al.[85,124] who calculated the surface temperature contributions for sublayer species with various thermal and optical properties. Experimental confirmation of this approach was obtained by determining a surface phase reference angle with a 60% carbon black-filled rubber elastomer, followed by determination of the experimental phase of a HMDSiN plasma-treated PP substrate, ethyl vinyl acetate on PP, and a CoTPP plasma on HMDSiN and PP at various frequencies. The differences between the phase measurements for bands

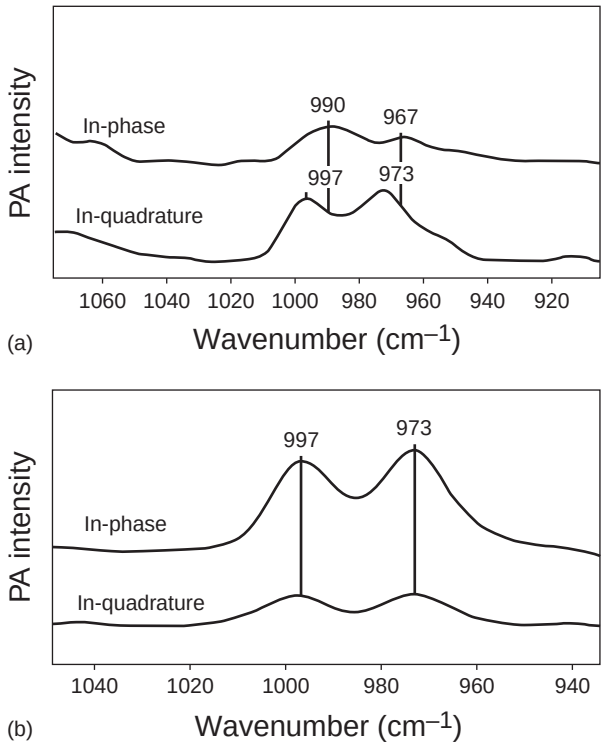

Figure 22 PUR/PP in-phase and in-quadrature PA FTIR spectra recorded (a) at 732 Hz and (b) 23 Hz modulation frequencies.

distinct to an absorbing layer were then calculated and compared to the theoretical values. It was shown that layers deeper within the sample have greater phase lags than shallow layers, in addition to the fact that phase depends on modulation frequency.

Another study used a similar phase difference approach with various plasma treated samples.[86] However, in this case depth-profiling was performed by using the experimental phase angles acquired at 400 Hz to calculate the absorbing layer depth of a particular species, by subtracting phase angles of spatially different species. The authors concluded that obtaining phase angles from other modulation frequencies is unnecessary to acquire all spatial information, which is inconsistent with the prior theory.

A method to determine absorbing layer depth for samples with overlapping IR bands was performed by Jones and McClelland[125] using vector analysis. In this study, PET films of known thickness were placed on a polycarbonate (PC) substrate, followed by the subsequent monitoring of the phase angles of the carbonyl C=O stretching bands at 1730 (PET) and 1778 cm^{-1} (PC). Similar to previous studies, the spectra were collected at a reference phase angle determined by a strongly absorbing sample, in addition to collecting data at one modulation frequency (400 Hz) to correlate all absorbing layer depths. Phase analysis was performed using analytical trigonometric equations to extract PA signal phase and subsequent depth of a species with overlapping contributions. More importantly, this study experimentally showed PA magnitude as a function of phase angle.

Recent studies focused on investigating PET films of known thickness and determined discrepancies in using carbon black samples to determine phase reference angles.[71] This conclusion was based solely on the fact that large phase errors occurred between carbon black reference angles from different manufacturers. Marginal errors resulted from an examination of PET with various thickness, therefore transparent spectral regions of PET were used to determine a collection angle. Similar to other studies,[85,86] this study showed that for all specimens the reference angle displayed frequency dependence, which was attributed to the phase contribution from the experimental apparatus.

A promising development using the PA signal phase is displayed in two-dimensional IR correlation analysis.[126,127] This method extends the uses of the I and Q spectra of S^2 PAS even further to create a two-dimensional array showing depth correlation of various species, similar to the combined use of both PA phase and magnitude. An implementation of all phase analyses to various systems may enhance PA spatial resolution beyond the limits of most current surface sensitive techniques.

5 INFRARED AND RAMAN IMAGING

In analysis of coatings it is often essential to correlate surface morphology with molecular level information responsible for this morphology. For that reason IR and Raman microscopy and imaging are quite useful because they are capable of obtaining $x-y$ direction information illustrated in Figure 2 and are capable of surface mapping. Principles of IR microscopy are the same as for ordinary optical microscopy, where a specimen is illuminated using a visible light in order to obtain a magnified image of the area of interest. Once the area for analysis is chosen and illuminated, the mode of operation is switched to IR light and the same area is illuminated by IR light, followed by collecting the IR spectrum. Figure 23 illustrates an example of IR mapping of a polymer latex surface. The spectra were collected from areas labeled A–E and indicate the presence of surfactant islands on the surface resulting from stratification across film thickness. Chemical imaging using the recently developed Stingray system by Bio-Rad

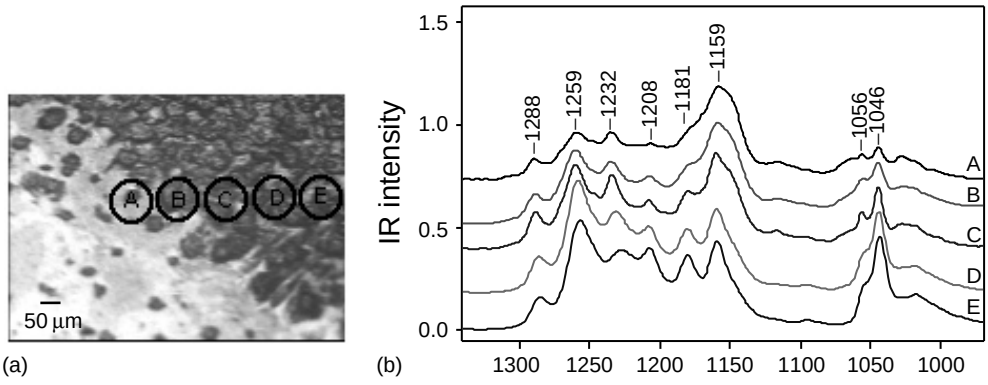

Figure 23 An example of IR mapping of a polymer latex surface.

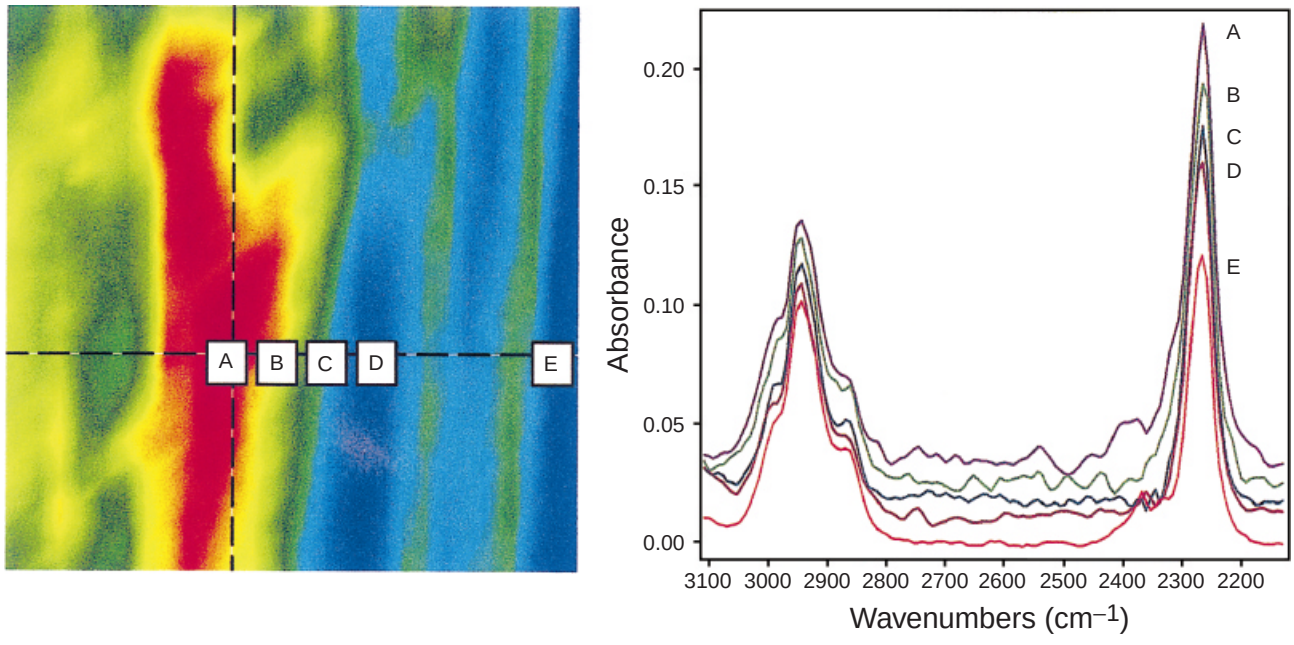

Figure 24 An example of Stingray imaging data of PUR films obtained using the Bio-Rad system. Image based on absolute intensity at 2920 cm^{-1}. Spectra A, B, C, D and E correspond to points A–E of an IR image, and each pixel contains spectral information.

represents state-of-the-art imaging technology. Figure 24 illustrates Stingray images of 128 PUR films recorded in the transmission mode. As seen, nonhomogeneous cross-linking reactions occur, as demonstrated by changing NCO intensities.

Similar principles apply to Raman microscopy where a laser beam illuminates a designated area and inelastically scattered light is collected to obtain Raman or FT Raman spectra. The most recent approach to Raman microscopy involves Raman chemical imaging in which a laser is focused in such a way that wide illumination is obtained at the sample surface. The laser light is directed to a holographic notch filter optimized for oblique illumination placed in the optical path of an infinity-corrected microscope. Figure 25 shows a chemical imaging set-up in which reflected light propagates through an objective and is transmitted through a holographic notch filter that rejects light at the laser wavelength. A second notch filter is positioned after the first filter and provides additional rejection of the elastically scattered (Rayleigh) light. The Raman emission is filtered with a 9 cm^{-1} bandpass liquid crystal tunable filter (LCTF) constructed in the Evans split-element geometry. The LCTF is an electronically tunable filter capable of high fidelity Raman imaging and has been described in the literature.[129,130] In a typical experiment the images are collected using a slow scan charge-coupled device from an approximately 20-µm square of 512 × 512 pixels. While Figure 25 is a schematic diagram of Raman chemical imaging, Figure 26 shows an image of an epoxy surface

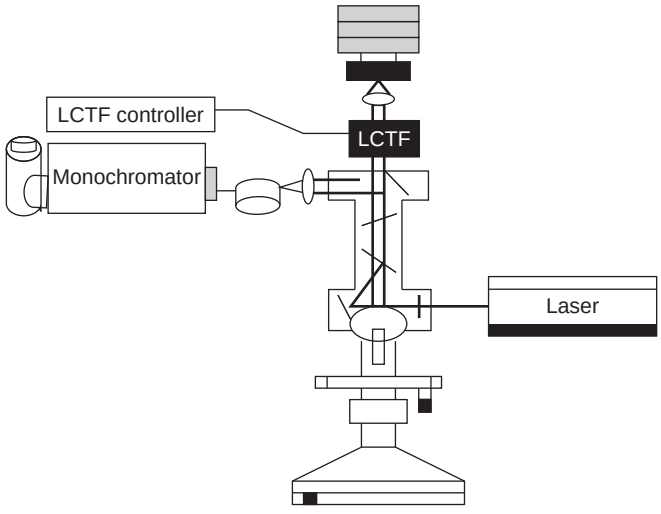

Figure 25 Schematic diagram of Raman chemical imaging.

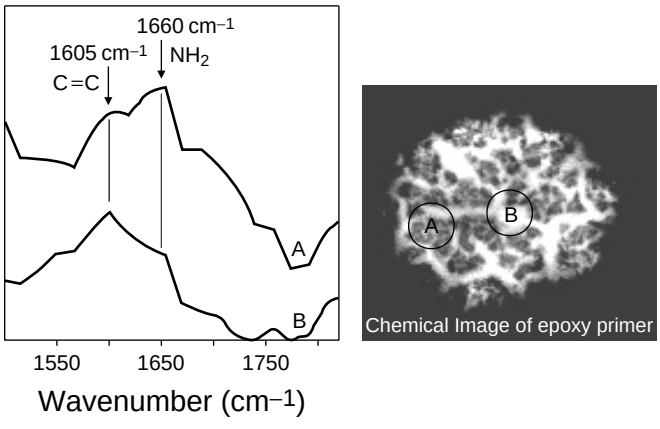

Figure 26 Raman chemical image of an epoxy surface exposed for 13 weeks to ultraviolet light.

exposed for 13 weeks to ultraviolet light. As seen, areas marked A and B show distinctly different images and their spectra provide clear evidence that area A is as epoxy-rich region, whereas area B is polyamine-NH_2-rich. Studies of adhesion of TPO also utilized Raman chemical imaging and showed that it is possible to identify surface morphology of rubber domains in chlorinated polyolefins (CPO) as well as low surface energy TPO.[131] Raman chemical imaging is capable of measuring the intrinsic vibrational signatures of CPO, PP, and EPR.

6 SUMMARY

With the recent technological advances there are numerous methods available for the analysis of

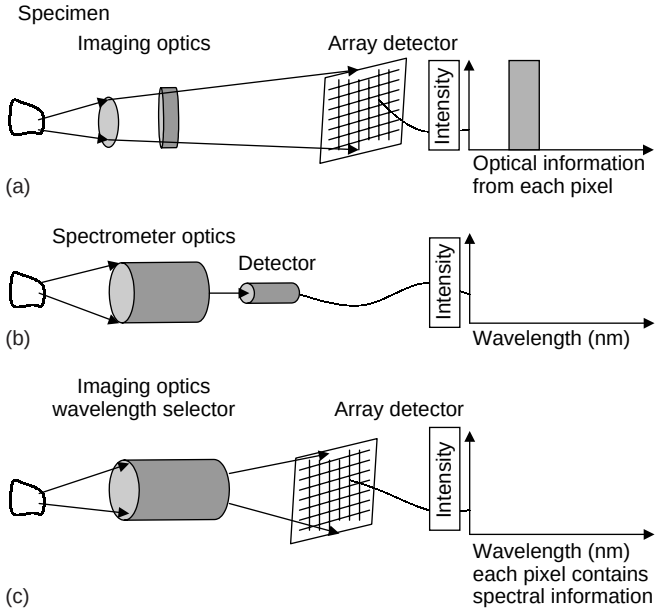

Figure 27 Schematic diagram of imaging techniques used in vibrational spectroscopy: (a) optical imaging; (b) spectroscopic analysis; (c) chemical imaging.

coatings, especially when nondestructive measurements are needed, and IR and Raman spectroscopy play a key role in surface depth profiling of coatings. Since both approaches or their variations do not require high vacuum environments, specimens are not altered. While ATR, RA, and PA FTIR techniques will continue to be utilized as powerful surface tools, Figure 27 illustrates three processes for obtaining surface information: optical imaging, spectroscopic analysis, and chemical imaging. The primary difference between optical imaging combined with spectroscopic analysis is that the combination of both allows us to obtain visual representation of the analyzed area. In contrast, in chemical imaging the analyzed surface is represented by individual pixels that contain spectral information. Because of recent and anticipated future technological advances it is possible to combine imaging and spectroscopy into a new area called chemical imaging, and it will not be long before numerous techniques will be combined together to provide multi-measurement capabilities on a molecular level.

ACKNOWLEDGMENTS

The author is grateful to the National Science Foundations Industry/University Cooperative Research Center in Coatings for partial support of these studies.

ABBREVIATIONS AND ACRONYMS

AM	Amplitude Modulation
ATR	Attenuated Total Reflectance
CoTPP	Cobalt Tetraphenylporphyrin
CPO	Chlorinated Polyolefins
CS	Continuous-scan
EPR	Ethylene–Propylene Rubber
EtAc	Ethyl Acetate
EtOH	Ethanol
FT	Fourier Transform
FTIR	Fourier Transform Infrared
HMDSiN	Hexamethyldisilazane
IR	Infrared
KKT	Kramers–Kronig Transform
LCTF	Liquid Crystal Tunable Filter
PA	Photo-acoustic
PAS	Photo-acoustic Spectroscopy
PC	Polycarbonate
PEO	Poly(ethylene oxide)
PET	Poly(ethylene terephthalate)
PM	Phase Modulation
PMMA	Poly(methyl methacrylate)
PP	Polypropylene
PUR	Polyurethane
PVC	Poly(vinyl chloride)
PVDF	Poly(vinylidine difluoride)
RA	Reflection–Absorption
RAIRS	Reflection–Absorption Infrared Spectroscopy
RPA	Rheo-photo-acoustic
S^2	Step-scan
TE	Transverse Electric
TM	Transverse Magnetic
TPM	Thermal Piston Model
TPO	Thermoplastic Olefins

RELATED ARTICLES

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers

Raman Spectroscopy **(Volume 15)**
Raman Microscopy and Imaging

REFERENCES

1. P. Jacquinot, *J. Opt. Soc. Am.*, **44**, 761 (1954).
2. P. Fellgett, *J. Phys. (Radium)*, **19**, 197 (1958).
3. J. Connes, *Optical Instruments and Techniques*, Oriel, Paris, 1970.
4. P.R. Griffiths, J.A. DeHaseth, *Fourier Transform Infrared Spectrometry*, Chemical Analysis Series 83, Wiley & Sons, New York, 1980.
5. A.A. Michelson, *Philos. Mag.*, **31**, 256 (1891).
6. D.C. Harris, *Quantitative Chemical Analysis*, W.H. Freeman, New York, 1995.
7. M.W. Urban, *Vibrational Spectroscopy of Molecules and Macromolecules on Surfaces*, Wiley-Interscience, New York, 1993.
8. C.S.P. Sung, *Macromolecules*, **14**, 591 (1981).
9. J.P. Hobbs, C.S.P. Sung, K. Krishan, S. Will, *Macromolecules*, **16**, 93 (1983).
10. P.A. Fluornoy, W.J. Schaffer, *Spectrochim. Acta.*, **22**, 5 (1966).
11. M.W. Urban, *Attenuated Total Reflectance Spectroscopy of Polymers*, American Chemical Society, Washington, DC, 1996.
12. J.B. Huang, M.W. Urban, *Appl. Spectrosc.*, **46**, 1666 (1992).
13. J.B. Huang, M.W. Urban, *Appl. Spectrosc.*, **47**, 973 (1993).
14. T. Hirayama, M.W. Urban, *Prog. Org. Coat.*, **20**, 81 (1992).
15. E.M. Salazar-Rojas, M.W. Urban, *Prog. Org. Coat.*, **16**, 371 (1989).
16. B.W. Ludwig, M.W. Urban, *J. Coat. Technol.*, **66**, 59 (1994).
17. B.W. Ludwig, M.W. Urban, *J. Coat. Technol.*, **68**, 93 (1996).
18. A.M. Kaminski, M.W. Urban, *J. Coat. Technol.*, **69**, 55 (1997).
19. H. Kim, M.W. Urban, *Langmuir*, **14**, 7235 (1999).
20. M.W. Urban, K.W. Evanson, *Polym. Commun.*, **31**, 279 (1990).
21. K.W. Evanson, M.W. Urban, *J. Appl. Polym. Sci.*, **42**, 2287 (1991).
22. K.W. Evanson, M.W. Urban, *J. Appl. Polym. Sci.*, **42**, 2297 (1991).
23. B.J. Niu, M.W. Urban, *J. Appl. Polym. Sci.*, **56**, 377 (1995).
24. C.L. Zhao, Y. Holl, T. Pith, M. Lambla, *Coll. Polym. Sci.*, **265**, 823 (1987).
25. T.A. Thorstenson, K.W. Evanson, M.W. Urban, *Polym. Mater. Sci. Eng.*, **4**, 195 (1991).
26. K.W. Evanson, M.W. Urban, *J. Appl. Polym. Sci.*, **42**, 2309 (1991).
27. T.A. Thorstenson, M.W. Urban, *J. Appl. Polym. Sci.*, **47**, 1381 (1993).
28. T.A. Thorstenson, M.W. Urban, *J. Appl. Polym. Sci.*, **47**, 1387 (1993).
29. B.J. Niu, M.W. Urban, *J. Appl. Polym. Sci.*, **60**, 389 (1996).
30. B.J. Niu, M.W. Urban, *J. Appl. Polym. Sci.*, **62**, 1903 (1996).

31. M. Claybourn, *Infrared Reflectance Spectroscopy of Polymers; Analysis of Films, Surfaces, and Interfaces*, Global Press, Chicago, 1998.
32. H.L. Pickering, H.C. Eckstrom, *J. Phys. Chem.*, **63**, 512 (1959).
33. S.A. Francis, A.H. Ellison, *J. Opt. Soc. Am.*, **49**, 131 (1959).
34. R.A. Gardner, R.H. Petrucci, *J. Am. Chem. Soc.*, **82**, 5051 (1960).
35. G.A. Bird, *J. Opt. Soc. Am.*, **51**, 579 (1961).
36. H.C. Ekstrom, W.H. Smith, *J. Opt. Soc. Am.*, **57**, 1132 (1967).
37. H. Pobimer, *Anal. Chem.*, **39**, 90 (1967).
38. R.G. Greenler, *J. Chem. Phys.*, **50**, 1963 (1969).
39. R.G. Greenler, *J. Appl. Phys. Suppl.*, **2**(2), 265 (1974).
40. R.G. Greenler, *J. Vac. Sci. Technol.*, **12**, 1410 (1975).
41. R.G. Greenler, *J. Chem. Phys.*, **44**, 310 (1966).
42. M. Claybourn, *Infrared Spectroscopy of Polymers; Analysis of Films, Surfaces, and Interfaces*, Global Press, Chicago, 1998.
43. A. Garton, *Infrared Spectroscopy of Polymer Blends, Composites and Surfaces*, Hanser Publishers, Munich, 1992.
44. D.G. Kurth, T. Bein, *Langmuir*, **11**, 578 (1995).
45. R.V. Duevel, R.M. Corn, *Anal. Chem.*, **84**, 337 (1992).
46. C. Naselli, J.F. Rabolt, J.D. Swalen, *J. Chem. Phys.*, **82**, 2136 (1985).
47. C. Naselli, J.P. Rabe, J.F. Rabolt, J.D. Swalen, *Thin Solid Films*, **134**, 173 (1985).
48. R.G. Snyder, M. Maroncelli, H.L. Strauss, V.M. Mark, *J. Phys. Chem.*, **90**, 5623 (1986).
49. A. Udagawa, T. Matsui, S. Tanaka, *Appl. Spectrosc.*, **40**, 794 (1986).
50. J.S. Wong, Y.S. Yen, *Appl. Spectrosc.*, **42**, 598 (1988).
51. Y.S. Yen, J.S. Wong, *J. Phys. Chem.*, **93**, 7208 (1989).
52. J.A. Mielkzarski, R.H. Yoon, *J. Phys. Chem.*, **93**, 2034 (1989).
53. J.A. Mielkzarski, R.H. Yoon, *Langmuir*, **7**, 101 (1991).
54. A.N. Parikh, D.L. Allara, *J. Chem. Phys.*, **96**, 15 (1992).
55. R.A. Dluhy, *J. Phys. Chem.*, **90**, 1371 (1986).
56. M.L. Mitchell, R.A. Dluhy, *J. Am. Chem. Soc.*, **110**, 712 (1988).
57. R.A. Dluhy, D.G. Cornell, *J. Chem. Phys.*, **89**, 3195 (1985).
58. A.G. Bell, *J. Sci.*, **20**, 305 (1880).
59. A.G. Bell, *Philos. Mag.*, **11**, 510 (1881).
60. J. Tyndall, *Proc. R. Soc. London*, **31**, 307 (1881).
61. W.C. Röntgen, *Philos. Mag.*, **11**, 308 (1881).
62. A. Rosencwaig, *Opt. Commun.*, **7**, 305 (1973).
63. J.G. Parker, *Appl. Opt.*, **12**, 2974 (1973).
64. A. Rosencwaig, *Photoacoustics and Photoacoustic Spectroscopy*, Wiley & Sons, New York, 1980.
65. J.F. McClelland, *Anal. Chem.*, **55**, 89A (1983).
66. R.A. Palmer, R.M. Dittmar, *Thin Solid Films*, **223**, 31 (1993).
67. R.M. Dittmar, J.L. Chao, R.A. Palmer, *Appl. Spectrosc.*, **45**, 1104 (1991).
68. L.C. Aamodt, J.C. Murphy, *J. Appl. Phys.*, **49**, 3036 (1978).
69. A. Rosencwaig, *J. Appl. Phys.*, **49**, 2905 (1978).
70. F.A. McDonald, G.C. Wetsel, *J. Appl. Phys.*, **49**, 2313 (1978).
71. M.W. Wahls, J.P. Toutenhoofd, L.H. Leyte-Zuiderweg, J. de Bleijser, J.C. Leyte, *Appl. Spectrosc.*, **51**, 552 (1997).
72. J.F. McClelland, R.N. Kinsley, *Appl. Opt.*, **15**, 2658 (1976).
73. L. Bertrand, *Appl. Spectrosc.*, **42**, 134 (1987).
74. G.C. Wetsel, F.A. McDonald, *Appl. Phys. Lett.*, **30**, 252 (1977).
75. P. Poulet, J. Chambron, *J. Appl. Phys.*, **51**, 1738 (1980).
76. Y. Fujii, A. Moritani, J. Nakai, *Jpn. J. Appl. Phys.*, **20**, 361 (1981).
77. J. Brunn, P. Grosse, R. Wynands, *Appl. Phys. B*, **47**, 343 (1988).
78. Akira Harata, T. Sawada, *J. Appl. Phys.*, **65**, 959 (1989).
79. P. Grosse, R. Wynands, *Appl. Phys. B*, **48**, 59 (1989).
80. M.J. Smith, C.J. Manning, R.A. Palmer, J.L. Chao, *Appl. Spectrosc.*, **42**, 542 (1988).
81. C.J. Manning, P.R. Griffiths, *Appl. Spectrosc.*, **51**, 1092 (1997).
82. R.M. Dittmar, R.A. Palmer, R.O. Carter, *App. Spec. Rev.*, **29**, 171 (1994).
83. C.J. Manning, P.R. Griffiths, *Appl. Spectrosc.*, **47**, 1345 (1993).
84. D.L. Drapcho, R. Curbelo, E. Jiang, R.A. Crocombe, W.J. McCarthy, *Appl. Spectrosc.*, **51**, 453 (1997).
85. E.Y. Jiang, R.A. Palmer, *J. Appl. Phys.*, **78**, 460 (1995).
86. E.Y. Jiang, R.A. Palmer, N.E. Barr, N. Morosoff, *Appl. Spectrosc.*, **51**, 1238 (1997).
87. R.W. Jones, J.F. McClelland, *Appl. Spectrosc.*, **50**, 1258 (1996).
88. J. Chamberlain, H.A. Gebbie, *Infrared Phys.*, **11**, 57 (1971).
89. P.L. Richards, *J. Opt. Soc. Amer.*, **54**, 1474 (1964).
90. J. Strong, G.A. Vanasse, *J. Opt. Soc. Amer.*, **49**, 844 (1959).
91. G.W. Chantry, H.M. Evans, J. Chamberlain, H.A. Gebbie, *Infrared Phys.*, **9**, 85 (1969).
92. S.I. Yaniger, S.M. Riseman, T. Frigo, E.M. Eyring, *J. Chem. Phys.*, **76**, 4298 (1982).
93. F.G. Will, R.S. McDonald, R.D. Gleim, M.R. Winkle, *J. Chem. Phys.*, **78**, 5847 (1983).
94. S.M. Riseman, S.I. Yaniger, E.M. Eyring, D. Macinnes, A.G. Macdiarmid, A.J. Heeger, *Appl. Spectrosc.*, **35**, 557 (1981).
95. J. Langer, *Synth. Met.*, **35**, 301 (1990).
96. D.W. Vidrine, *Appl. Spectrosc.*, **34**, 314 (1980).
97. M.G. Rockley, *Appl. Spectrosc.*, **34**, 405 (1980).
98. M.G. Rockley, *Chem. Phys. Lett.*, **68**, 455 (1979).
99. G.F. Kirkbright, *Opt. Pura. Appl.*, **11**, 125 (1977).

100. K. Krishnan, S. Hill, J.P. Hobbs, C.S.P. Sung, *Appl. Spectrosc.*, **36**, 257 (1982).
101. M.W. Urban, J.L. Koenig, *Appl. Spectrosc.*, **40**, 994 (1986).
102. C.Q. Yang, W.G. Fateley, *Anal. Chim. Acta.*, **194**, 303 (1987).
103. C.Q. Yang, R.R. Bresee, W.G. Fateley, *Appl. Spectrosc.*, **41**, 889 (1987).
104. B.D. Pennington, M.W. Urban, R.A. Ryntz, *Polymer*, (1998).
105. J. Huang, M.W. Urban, *J. Chem. Phys.*, **100**, 4509 (1994).
106. W.F. McDonald, H. Goettler, M.W. Urban, *Appl. Spectrosc.*, **43**, 1387 (1989).
107. M.W. Urban, *Polym. Mater. Sci. Eng.*, **64**, 31 (1991).
108. W.F. McDonald, M.W. Urban, *Polym. Mater. Sci. Eng.*, **60**, 739 (1989).
109. B.W. Ludwig, M.W. Urban, *Polymer*, **33**, 3343 (1992).
110. B.W. Ludwig, M.W. Urban, *Polymer*, **34**, 3376 (1993).
111. B. Mongeau, G. Rousset, L. Bertrand, *Can. J. Phys.*, **64**, 1056 (1986).
112. B. MacDonald, M.W. Urban, *Appl. Spectrosc.*, **43**, 1387 (1989).
113. B. Pennington, M.W. Urban, *J. Adhes. Sci. Technol.*, **13**, 19 (1999).
114. M.J. Smith, C.J. Manning, R.A. Palmer, J.L. Chao, *Appl. Spectrosc.*, **42**, 546 (1988).
115. D. Débarre, A.C. Boccara, D. Fournier, *Appl. Opt.*, **20**, 4281 (1981).
116. L.B. Lloyd, S.M. Riseman, R.K. Burnham, E.M. Eyring, M.M. Farrow, *Rev. Sci. Instrum.*, **51**, 1488 (1980).
117. K.H. Michaelian, *Appl. Spectrosc.*, **43**, 145 (1989).
118. K.H. Michaelian, *Infrared Phys.*, **29**, 87 (1989).
119. L.E. Jurdana, K.P. Ghiggino, I.H. Leaver, P. Clarke-Cole, *Appl. Spectrosc.*, **49**, 361 (1995).
120. A. Mandelis, Y.C. Teng, B.S. Royce, *J. Appl. Phys.*, **50**, 7138 (1979).
121. M.J. Adams, G.F. Kirkbright, *Analyst*, **102**, 281 (1977).
122. J.C. Roark, R.A. Palmer, *Chem. Phys. Lett.*, **60**, 112 (1978).
123. M.W. Urban, J. Stegge, *Polymer* (in press).
124. R.A. Palmer, E.Y. Jiang, *J. Physique*, **4**, C7–337 (1994).
125. R.W. Jones, J.F. McClelland, *Appl. Spectrosc.*, **50**, 1258 (1996).
126. C. Marcott, G.M. Story, I. Noda, *Abstracts of the Pittsburgh Conference on Analytical and Applied Spectroscopy*, New Orleans, LA, Paper 1043, 1995.
127. G.M. Story, C. Marcott, I. Noda, *SPIE Proc.*, 2089 (1993).
128. Q. Han, M.W. Urban (studies in progress).
129. J.F. Turner, P.J. Treado, *SPIE Int. Soc. Opt. Eng.*, **3061**, 280 (1997).
130. P.J. Treado, *Laser Focus World* (1995).
131. H.R. Morris, J.F. Turner, B. Munro, R.M. Ryntz, P.J. Treado, *Polymer*, **68**, 2344 (1997).

Mechanical Properties of Solid Coatings

Joseph V. Koleske
Charleston, USA

1 Introduction 1
2 History 2
3 Architectural Coatings 2
4 Industrial Coatings 3
4.1 Film Formation 3
4.2 Mechanical Properties 4
4.3 Flexibility and Toughness 6
4.4 Adhesion 7
4.5 Hardness 8
4.6 Abrasion 8
4.7 Slip 9
4.8 Stress in Coatings 10
4.9 Chemical Resistance 11
5 End-uses 11
5.1 Tests Required for Specific End-uses 11
Abbreviations and Acronyms 12
Related Articles 12
References 12

Coatings can be liquid or solid materials; they have been known and used from the time of early man and are of major importance for protecting and decorating the myriad of items in use today. These thin protective films are used to coat commercial and residential buildings (architectural coatings) and for many products in use today (industrial coatings). This article is concerned with the formation of coating films and the properties of these films. Static and dynamic mechanical properties, flexibility, toughness, adhesion, hardness, abrasion resistance, slip, chemical resistance, and stress within coatings are discussed and ways to investigate these characteristics are given. Some of the end-uses and the relationship of coating properties to the uses are detailed.

1 INTRODUCTION

The term ''coating or coatings'' is used to designate liquid or solid materials. A product that is to be applied to a substrate in a continuous or discontinuous film by one of many application methods is a liquid coating. After the liquid coating is dried by evaporative means or is cured (cross-linked) by oxidative, thermal, ultraviolet light or other method it is a solid coating film. This article will deal for the most part with solid coating films and the properties of liquid coating films can be found elsewhere in the encyclopedia. Coatings can be colorless or colored; they are thin, solid films that are transparent, translucent, or opaque in nature. The terms ''coatings'' and ''paint'' should be synonymous, but often coatings refer to industrial coatings such as those used on appliances, office furniture, paper, automobiles, beverage cans, etc. and paint refers to architectural coatings such as house paints, wall and ceiling paints, trim paints, etc. The materials used by artists[1] are referred to as artists' paints. While this is not a clear distinction, it is useful to be aware of a general differentiation in the terms and of the way they are used on a daily basis. Consider also that coatings are often referred to as finishes.

Paints can be defined as dispersions of pigments, optionally including fillers, in a fluid vehicle. The fluid vehicle comprises a liquid binder that will solidify during cure and, if necessary, a liquid carrier that serves as a viscosity reducing aid and will provide desirable application characteristics. The liquid carrier is lost to the local environment or recovered during the drying or curing process. The binder portion of the fluid vehicle is an oil, dissolved polymer, and, when appropriate, a cross-linking agent, or a dispersed polymer in latex or other dispersion form. The binder holds the pigment, filler particles, and miscellaneous formulating ingredients when it solidifies into a film on exposure to air or some other curing media. Varnish, a term that is often replaced by the term clear coating, is a clear or transparent solution that solidifies into a functional film. Lacquers are opaque and/or colored varnishes. Collectively or individually, paints, varnishes, and lacquers are termed coatings.

Coatings can be functional and/or decorative in nature. One need only to look at almost any surroundings to see how widely coatings are used. Many things that can be seen are coated with a decorative and/or a functional material. For example, within an office there might be walls and ceiling coated with a colored decorative paint; a desk coated with a clear, functional finish, which may be applied over a colored, decorative stain coating; alternatively the desk could be coated with a colored, opaque, functional and decorative coating; a floor coated in a manner similar to that of the desk with a functional and decorative coating; there may also be an aluminum beverage can coated on the outside with a decorative coating that serves as an advertising and identification medium and coated on the inside with a functional coating that protects the metal from the chemical nature of the can's contents. In addition to the above, there may be

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

many books, perhaps a newspaper, and other printed material. Some of these are printed on coated paper, and the printing in itself is a coating that is discontinuous in nature. The book jackets and some of the illustrations in the books are continuous-film coatings. Sheets of labels found in an office are coated with a pressure-sensitive adhesive. The list could go on, but it is readily apparent that we are surrounded by coatings.

Decorative coatings can be brilliant and bright to attract visual senses quickly; pastel and soft to provide a restful atmosphere; single color or multicolor to provide a variety of dramatic effects; glossy, semi-glossy, or matt to provide different mood effects or to affect cleanability; rough or smooth, and so on. Functional coatings often protect substrates – wood, metal, plastic, or other – from the ravages of nature, which often provide a hostile environment and can cause wear through rusting, erosion, light attack, etc. The protection provided by functional coatings saves natural resources and is friendly to the environment because it minimizes corrosion and other means of degradation, thus allowing a substrate to last for a much longer period than it would without protection. Such coatings could also protect an expensive part of an assembly that cannot be reached or is difficult to reach for repair, such as the role conformal coating plays in the electronics industry. Here the coatings protect printed circuit assemblies from the hostile environment found in outer space and sometimes within plants, laboratories, or living quarters. Pressure-sensitive adhesives applied to the back of heavily coated paper or polyester film form widely used functional label products of various designs.

Coatings can be divided into two broad groups: architectural coatings and industrial coatings. Architectural coatings are those used in decorating and protecting houses and other buildings. If the coatings are used on the portion of the building exposed to atmospheric conditions, they are termed exterior coatings. Such coatings are usually made from materials that are not light (radiation) sensitive or are stabilized against attack by radiation of different wavelengths. If they are used on the inner portions of the buildings, they are referred to as interior coatings; this group is further broken down into ceiling paints, wall paints, varnishes, masonry paints, and stains.

2 HISTORY

Early humans used plant extracts, tree saps, animal fats, berry juices, and metal oxides to create paints that were used to decorate and communicate by means of pictures, often on cave walls.[2,3] Such pictures have remained well defined and vibrant for over 15 000 years. Early Egyptians coated dead bodies with bitumen and other materials in the mummification process (the word mummy is derived from the Persian word *mumia*, which means bitumen or pitch)[3,4] and it is said that Noah's ark was coated with pitch. About 2500 years ago, Egyptians developed clear varnishes by heating amber and vegetable oils; colored coatings or lacquers were made by adding ground minerals such as malachite, azurite, and iron compounds to the clear varnishes. Early Romans developed the method of fresco painting in which paints composed of pigments, fillers, and carrier vehicles were applied to wet plaster for interior surface coatings and paintings. Ancient Hebrews used milk-based paints for decorating walls and ceilings. Over 3000 years ago, the Japanese developed lacquers based on sap from the varnish tree (the Japanese sumac, *Rhus verniciflua*). In colonial days, water-slacked lime – whitewash – was extensively used for a variety of coating purposes.[5] Later whitewash was modified with milk and protein-based materials to improve durability and adhesion. This was followed by the addition of pigments, fillers such as clay, and whiting to provide a variety of products with improved aesthetic appeal and economics. Finally, the lime was replaced with milk phosphoproteins, and whitewash became casein paint, a forerunner of today's architectural paints.

Although paints were used and modified in such ways for centuries, it was not until the Industrial Revolution (1700–1950) that the paint and coating industry took on far-reaching importance. This historical period resulted in the production of a multitude of bridges, factories, manufacturing machinery, and allied equipment, as well as other items; all of these needed to be coated to provide protection from hostile natural and derived environments. Protection had to be provided from moisture, salt water, barnacles, mildew, mold, wind, rain, hail, heat and cold, sunlight, sulfurous fumes from coal fired furnaces, etc. This protection was provided by coatings that extended the lifetime of the world's infrastructure, the manufacturing facilities, and the ever increasing number of manufactured items. Today, the same items plus a myriad of other items – packaging, cars, trucks, trains, boats, furniture, beverage can liners, wallpaper, etc. – that grew out of the ever expanding manufacturing base are coated with a protective and very often aesthetically pleasing film that is usually thinner than a sheet of writing paper. If it were not for paint and coatings, our world would certainly be a dull, corroding place in which to live.

3 ARCHITECTURAL COATINGS

Architectural coatings are those coatings used on interior and/or exterior surfaces such as those found in or

on commercial, institutional, industrial, and residential buildings as well as on various structures such as bridges.[6] The surfaces coated are wood, metal, composition, plaster and wallboard, plastic, or masonry. The coatings may be latex, alkyd, oil, solvent borne, and so on, and they are applied by brushing, spraying, rolling, as well as other methods. Gloss is an important optical characteristic of these coatings and varies from low to high depending on any particular end-use. These coatings are often termed trade sales paints, because they are usually purchased by consumers who will apply them on-site under ambient conditions.

Standard methods are available to test the characteristics of architectural coatings.[7–10] Abrasion, dry and wet adhesion, flexibility, chemical resistance, block resistance, print resistance, and cleanability are important properties.

Abrasion resistance, adhesion, and flexibility are discussed elsewhere in this article. Mechanical properties of architectural coatings are not usually directly measured, but rather such properties manifest themselves in the results of an end-use-related test. Testing of architectural coatings and industrial coatings is often a measure of a complex interaction of various coating physical characteristics. For example, washability and related characteristics[11–14] of a wall coating requires the coating to be chemically resistant to water and detergent and to be sufficiently strong to withstand a scrubbing action that applies tensile and shear stresses to the coating as its hardness, adhesion, cohesion, and abrasion resistance are brought into play.

Ease of application and aesthetic characteristics are more important than mechanical characteristics for interior coatings. However, mechanical and other physical properties are important to exterior architectural coatings that are subjected to outdoor exposure. Exterior coatings are exposed to heat and associated temperature changes, moisture, oxygen, and sunlight. These factors individually or in concert contribute to coating failure. Temperature changes alter properties and can result in significant alternating strains, along with concomitant stresses,[15] being placed on the coating–substrate matrix. This, coupled with the effects of exposure to relatively high temperatures, results in cracking, checking, embrittlement, and peeling. Moisture can cause blistering, erosion, loss of adhesion, and mildew. Oxygen can cause surface degradation and eventually internal degradation, resulting in embrittlement, cracking, and crazing. Sunlight and, in particular, the ultraviolet light component of sunlight can cause surface chalking and loss of gloss, degradation, and embrittlement, with accompanying cracking and discoloration.

Effects of outdoor exposure are often measured with test devices that attempt to simulate and accelerate changes that would be encountered in the environment.[16] Accelerated exposure test results are often difficult to reproduce and may not correlate with actual or natural exposure testing. However, such testing is widely used because natural weathering can take years to effect changes, and the accelerated tests do give a good indication of coatings that will fail early. They are particularly important for comparison purposes, for new product development, and to improve the durability of existing products.

Natural weathering[17] is a true measure of the ravages of nature; however, it is reproducible only if properly planned, conducted, inspected, and reported. Natural weathering tests are not carried out by merely placing a test specimen outdoors and letting sunlight, rain, and so on fall on it. Rather, the testing is carried out at selected sites with the exposed specimens set at a particular angle. Conditions such as these affect the four major factors listed above and allow other factors such as humidity or lack of it, biodegradation, and pollution to be brought into the testing scheme. The synergistic interaction of these factors, which will vary with exposure site, determines how a coating's failure through outdoor exposure takes place. The variation of solar energy radiation dosage with season and with the angle of the test specimen to the sun is an important variable to be considered when reproducible test results are expected. The closer the test conditions simulate the actual use conditions, the more accurate will be the prediction of long-term results.

4 INDUSTRIAL COATINGS

Industrial coatings are coatings applied to factory-manufactured products. These include, but are not limited to, transportation coatings (those coatings applied to aircraft, appliances, automobiles, buses, recreational vehicles, trucks, and trains) beverage-can and spray-can coatings, packaging items, business machine and office furniture coatings, wood cabinet and furniture coatings, pipeline coatings, printed circuit board and assembly coatings, sign coatings, marine coatings, and masonry coatings.

4.1 Film Formation

Coating films are formed from either thermoplastic (soluble) or thermoset (insoluble) polymers combined with other ingredients, including, if desired, but not limited to, pigments, fillers, colorants, plasticizers, surfactants, solvents, catalysts, or initiators.[18–20] Thermoset coatings are formed from initially soluble ingredients that react and undergo a change termed cross-linking as a consequence of an energy input.

When thermoplastic polymers are dissolved in a solvent, films are formed by evaporation of the solvent under ambient conditions or in the presence of controlled heating. Since physical characteristics of the final coating are dependent on the polymer's properties, polymers of high or relatively high molecular weight are used. The high molecular weight limits the amount of polymer that can be dissolved because the viscosity of the final system must be sufficiently low to allow the coating to be applied by brush, spray, roll-coater, etc. Polymers that are used to form coating films include nitrocellulose, cellulose acetate butyrate, vinyl chloride/vinyl acetate copolymers and terpolymers, poly(vinyl acetate), and poly(methyl methacrylate). Coatings such as these are easily removable with a solvent or marred by a plasticizing compound; in addition particular liquids can attack the coating, for example water or alcohol can cause a ring on nitrocellulose-coated furniture.

Thermoset coatings are produced when multifunctional low-molecular-weight polymers or oligomers are reacted with a multifunctional cross-linking compound that contains appropriate functionality. If a solvent is present, it is removed by evaporation to form a film, which is then heated to effect reaction between the different functionalities in the presence of (or without) a catalyst. If the cross-linking reaction is between hydroxyl-containing and isocyanato-containing or epoxide-containing compounds, for example, the reaction proceeds by an addition process without emissions to form either urethane linkages or ether linkages (Scheme 1).

If the reaction is between a hydroxyl-containing compound and, for example, a methoxymelamine, material is lost by emission of a volatile by-product, in this case methanol (Scheme 2).

In addition to forming films by evaporation of solvent from solutions of polymers and oligomers, films can be formed from aqueous and non-aqueous dispersions, organosols, plastisols, electrodeposition, powders, and radiation-activated systems. The American Society for Testing and Materials (ASTM) and other organizations have a variety of test methods for determining various aspects of film formation and the films formed. These include detailed procedures for preparing[21] and testing[22] organic films, for drying, curing, and formation of films,[23] for determination of minimum film formation and coalescence temperatures of aqueous dispersions,[24] for permeability,[25] for block resistance,[26] and for numerous other particular physical factors. A number of the physical property tests are briefly described below.

The degree of cure or solvent resistance of films, particularly thermoset coatings, is often determined by means of a solvent rubbing procedure that is formalized for zinc-rich, ethyl silicate coatings.[27] A gauze cloth is made into a pad and saturated with a solvent such as methyl ethyl ketone or acetone. Then, using thumb pressure, the solvent-wet pad is rubbed back and forth over the coating. Although the test is said to be imprecise because of variability in thumb pressure from operator to operator, it does provide quick, useful results, particularly on a comparative basis. Other tests can be found to determine solvent resistance.[28–31]

4.2 Mechanical Properties

Mechanical properties such as tensile strength, elongation, and toughness and related parameters are important characteristics of coatings.[32] Outdoor coatings must withstand the rigors of the tensile and compressive forces that occur and cause expansion and compression during each day and with the changing seasons as the temperature changes. Hail and sleet challenge the toughness of coatings. The effects of temperature on coatings for wood substrates are further complicated by the differences in hardness, chemical composition, and expansion coefficients of spring wood and summer wood. Obviously, these changes occur many, many times over the course of a number of years. Many coatings are applied to a substrate before forming operations are carried out. Consider the metal white caps that are applied to a variety of packaged foodstuffs. The caps are formed after the white coating is applied to sheet steel. The coated metal is subjected to severe tensile and shear forces during the strong bending and twisting operations required to form the cap. The coating–steel composite must have excellent adhesion, strength, and toughness to withstand the torturous operation.[33] Three-piece beverage cans undergo severe bending and twisting when the can ends are combined with the can body in a flanging operation. The same is true

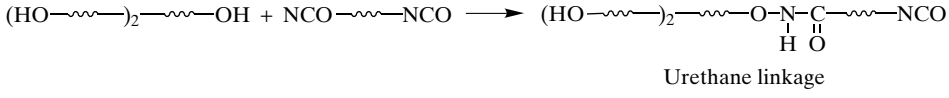

Scheme 1

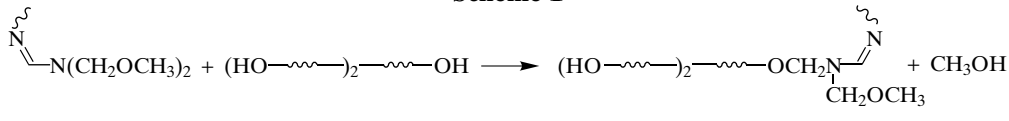

Scheme 2

when the lid is attached to a two-piece can. In addition, when the filled cans are transported by truck, the coated, filled can is again subjected to twisting and flexing during handling and with each bump and turn in the road during transportation. The severity of these operations is such that pin holes and failure can develop in the container and cause product loss as well as potential liability.

Tensile properties are those characteristics that a material exhibits when a uniaxial force F is applied to a specimen of length L_0 and cross-sectional area A, as depicted in Figure 1. Under no-load conditions, the specimen is at rest. When the force is applied, the specimen experiences a tensile stress, σ, that is equal to the applied force per unit area and an elongation to a final length, L, involving a tensile strain, ε, which is given by change in length per unit length or $(L - L_0)/L_0$. In the region where stress is directly proportional to strain, a tensile modulus, E, can be defined as the slope of the stress–strain relationship as shown in Equations (1) and (2). (Similar expressions can be defined for shear and bulk deformation conditions.)

$$E = \frac{\text{Tensile stress}}{\text{Tensile strain}} = \frac{\text{Force per unit area}}{\text{Change in length per unit length}} \quad (1)$$

$$E = \frac{F}{A} \div \frac{\Delta L}{L_0} = \frac{\sigma}{\varepsilon} \quad (2)$$

However, polymers and coatings actually have no measurable region where stress is proportional to strain because polymer molecules can flow under an applied force. That is, polymers are not elastic in nature; rather they are viscoelastic and slowly flow when placed under conditions of loading. The modulus of viscoelastic materials has a storage or completely recoverable elastic component (E') and a viscous or loss component (E'') that is not recoverable and the energy of which is lost through viscous heating during the stressing process. To circumvent this problem, when measured under static conditions, the value of E is calculated by measuring the stress at some stated, fixed strain, usually 1% or less, and dividing it by the strain. Such a measured value of E is denoted as the secant modulus[34] measured at the selected strain.

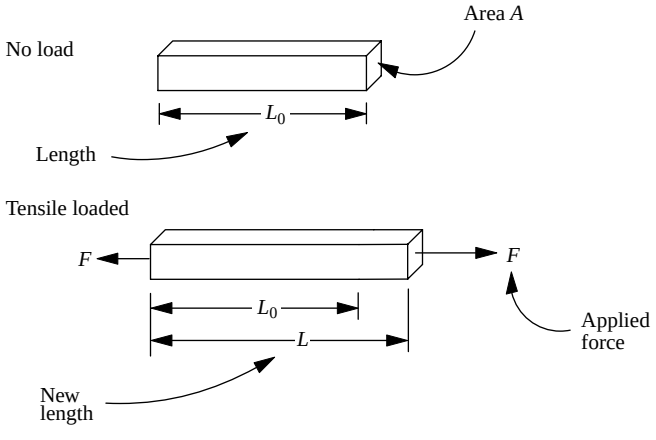

Figure 1 A specimen at rest and in a tensile-loaded condition.

Dynamic mechanical analysis (DMA) is a technique that enables the two components E' and E'', which together are known as the complex tensile modulus, E^*, to be distinguished and measured.[35,36] In DMA, the stress or strain described in Figure 1 is applied to the specimen in an oscillatory manner that is usually described as being sinusoidal in nature, though the exact nature of the deformation depends on the particular instrument used for the measurements. A sample is held under sufficient tension that it is not limp at the lowest applied oscillatory strain. Although the impressed wave motion is the same, say sinusoidal, for the stress and the strain, because of the material's viscoelastic nature these properties are out of phase by an amount or angle, δ, the phase lag. For an ideal elastic material, δ is zero and the stress and strain are in phase. For a Newtonian liquid, if it could be so tested, δ would be 90°. For viscoelastic materials, δ is between 0 and 90°. These components of the modulus can be represented as two vectors that are 90° out of phase with each other, as described in Figure 2.

Equations (3–6) describe the stress wave as two waves, one in phase with the strain (the elastic response) and the other 90° out of phase with the strain (the viscous response).

$$\text{Storage modulus} \quad E' = \frac{\sigma_t \cos\delta}{\varepsilon} \quad (3)$$

$$\text{Loss modulus} \quad E'' = \frac{\sigma_t \sin\delta}{\varepsilon} \quad (4)$$

$$\text{Complex modulus} \quad E^* = [E'^2 + E''^2]^{1/2} \quad (5)$$

$$\text{Loss tangent} \quad \frac{E''}{E'} = \frac{\sigma_t \sin\delta}{\sigma_t \cos\delta} = \tan\delta \quad (6)$$

The complex tensile modulus, E^*, or the one that is actually sensed, can be calculated from the components, E' and E'', by the Pythagorean theorem. The ratio of the viscous response to the storage response is equal to the

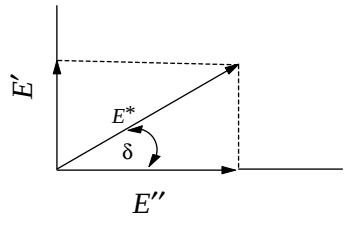

Figure 2 Vector representation of the components, E' and E'', of the complex shear modulus, E^*, and the phase angle, δ.

tangent of the out-of-phase angle and is known as the loss tangent. When these parameters are measured as a function of temperature, the loss modulus and the loss tangent exhibit maximums at the glass transition temperature, T_g, and at other temperatures if there are other secondary loss mechanisms. Below T_g, materials are glass-like in nature – hard and brittle – and above T_g they are soft and flexible in nature. If the system is cross-linked, above T_g the material will exhibit an extensive rubbery nature.

To visualize the storage and loss components of a material, consider a rubber ball that is dropped from some height onto a hard, elastic surface. The ball will bounce up and down with a decreasing height as each succeeding bounce takes place until it is finally at rest. The elastic nature of the ball causes it to try and return to the release point by means of stored energy, but the viscous nature of the ball results in it losing energy as heat to the surroundings; consequently, the ball reaches a height somewhat less than the original height. The motion is gradually damped until finally the ball ceases to bounce. One readily knows that a markedly different result is obtained if this experiment is performed on a summer day than on a winter day. Actually one can measure important physical properties of materials by this technique, but a number of more practical methods than this one exist.[36]

4.3 Flexibility and Toughness

Flexibility is the ability of a coating to be bent or flexed in forming operations without cracking, losing adhesion, or failing in some other manner. Toughness is the ability of a coating to withstand large stress forces imposed over a short time without cracking, rupturing, shattering, or tearing.[37] Coatings must properly perform during manufacturing operations, during use, and often during misuse. To do this, they must have sufficient flexibility and toughness to withstand failure when subjected to bending and twisting, as is encountered in forming operations, to expansion and shrinking during temperature changes, and to mechanical abuse.

Flexibility is usually measured by a mandrel bend test[33] or a T-bend test.[35] The mandrel bend test involves bending a coated substrate, usually sheet metal or rubber-type materials, over either a conical mandrel or over cylindrical mandrels of various diameters. The standard, smooth-steel, conical mandrel has a length of 203 mm (8 in) and a diameter of 3 mm (0.125 in) at one end and 38 mm (1.5 in) at the other end. The coated substrate, coating side up, is bent around the mandrel with a lever device and the extent of cracking, if it exists, is determined. The distance from the small end of the mandrel to the crack is determined visually and can be used graphically to determine the percent elongation. (However, there is no indication in the test method that elongation determined from tensile studies will yield a value related to the cracking-failure point.) The mandrel diameter at the point where cracking ceases is reported as the resistance to cracking resistance or flexibility.

The cylindrical mandrel test is a pass/fail test that involves placing the coated substrate over a mandrel, coating side up, and bending the specimen about 180° around the mandrel by hand at a uniform velocity in a specified time. Usually six mandrels having diameters ranging from 25 mm (1.0 in) to 3.2 mm (0.125 in) are used. The panel is bent over the largest diameter mandrel and then immediately examined for cracking. If none occurs, the next smaller mandrel is used and so on until failure occurs or the smallest diameter mandrel has been passed. The smallest diameter at which cracking does not occur is reported. The test can be used to calculate coating elongation.

The T-bend test involves placing a coated metal panel with a 50 mm (2 in) minimum width in a smooth jaw bench vise and holding it firmly.[38] The panel must be sufficiently long that the needed number of bends can be made, i.e. about 150 mm (6 in). Then the panel is bent 90° with the coating on the outside of the bend, removed, and further bent by hand until the bent end can be inserted in the vise; the vise is tightened to complete the 180° bend. The apex end of the bend should be as flat as possible. This is termed a 0T (zero-T) bend. The bend is then examined with a 5 to 10 power magnifier for cracks and pressure-sensitive tape is applied and removed to determine if coating can be picked off. The process is then repeated by placing the bent end in the vise and bending through 180° around the 0T bend. This forms the 1T bend. This is continued for 2T, 3T, etc. bends. The lowest T bend at which no cracks are visible and there is no pick off of coating is the value reported. Note that the radius of curvature of the bend increases with each succeeding bend and coating elongation required to make the bend decreases with each succeeding bend.

Flexibility of pipeline coatings that are to be subjected to short-radius bends is determined by bending the coated pipe around a designed, variable-radius mandrel to produce a range of short-radius bends.[39] Coating failure is apparent by visual and/or electrical inspection of cracking or loss of adhesion.

Toughness can be defined as the ability of a coating to withstand an impact without cracking or breaking. It is dependent on the nature of the polymer or polymers used in the coating and on adhesion. Impact resistance, which is related to formability, can be measured by dropping a weight from various heights through a guide tube onto an indenter that rests on the surface of the coated substrate.[40] The test can be made on the coated side (face impact) and/or the uncoated side (reverse impact)

of the coated substrate. Cracking or other failure is noted on or around the dimple caused by the indenter. The cited ASTM test gives three procedures for ascertaining failure: visual inspection, application of an acidified copper sulfate solution, and use of a pin hole detector. Several impacts are made at different impact values and at the same impact value. The value where the force required changes the result from mainly passing to failing is the test end-point. The result at this point is reported as kilogram-meters (inch-pounds) impact resistance.

A wedge bend device is used to determine impact resistance and formability of metal strips that have been factory coated by a roll coating or other application technique.[41] Coated strips are bent 170–180° over a 3.2-mm (0.125 in) cylindrical mandrel that is attached to the impact platform. A 1.82-kg (4 lb) guided rod with a flat end is then dropped onto the test specimen. Variation in the height of drop allows the force needed to crack the coating to be measured. A test that involves high-pressure pressing of an indenter ball into a zinc-rich primer-coated metal substrate tests the formability of the coated metal.[42] Formability tests that ascertain the flexibility and impact resistance of coatings by stamping a die into coated metal exist.[37]

The impact resistance of pipeline coatings is determined by a limestone drop test,[43] a falling weight test,[44] and a penetration resistance test.[45] The limestone drop test involves dropping weighed amounts of a particular type of limestone through a chute onto a coated pipe. The number of drops required to penetrate the coating by either visual or electrical inspection is reported as the impact resistance. The falling weight test is similar to that described above[40] except special pipe-holding devices and impacter surface characteristics are involved. Coating breaks or penetrations are detected by measurement of electrical resistance changes; the impact resistance is the amount of energy required to cause a break. The penetration resistance test involves applying a blunt rod loaded with a dead weight to a coated steel pipe. The depth or rate of penetration of the rod into the coating is measured as a function of time. This and any failure (cracking or other penetration) are reported.

4.4 Adhesion

The importance of adhesion, the ability of a coating to resist removal from the surface to which it is applied, is self evident.[46] Such adhesion can be between substrate and coating, between a primer coating and a top coating, between coatings applied to an existing coating, etc. In addition, the coating must adhere under various weathering and cleaning, usually aqueous, conditions. The adhesion can be between the same – in a chemical sense – materials or between a broad variety of materials including plastics, wood and other cellulosics, metals, ceramics, etc. There are two aspects involved in adhesion: ''basic'' adhesion, which is the combination of all intermolecular and interfacial forces, and ''practical'' adhesion, which is the work needed to disrupt the adhering combination. Practical adhesion is almost always the quantity measured in the coating industry.

The most common method of testing coating adhesion involves applying an adhesive tape to the coating, which is either uncut or cut in some manner, and then removing the tape under specified conditions. The cut surface is observed and the degree to which the coating is removed is compared against standards. The test is considered simple to perform and low in cost. A widely used test method for coatings on metallic substrates[47] involves making an X-cut in the film (method A) or making a lattice cut with a device that makes six or eleven cuts in each direction (method B) with each of the cuts made through the film to the substrate. A transparent, pressure-sensitive tape is applied to the cut area and removed in a prescribed manner. The coating is then visually examined and rated on a zero to five comparison scale. On this scale five indicates no removal and zero indicates greater than 65% removal from the scored area, with various descriptions for the values between the extremes. Method A is meant to be used on the job and method B is meant for use in the laboratory. The test and results are qualitative in nature, and the results are considered reproducible within one unit when the substrate is metal. On plastic substrates, reproducibility is poor since the test is not designed for relatively soft substrates that are usually coated with brittle coatings.

Adhesion of coatings to flat substrates can be determined by pushing the panel beneath a rounded stylus on a balance-beam device that is increasingly loaded until the coating is removed.[48] This scrape adhesion test is used to differentiate the degree of adhesion to substrates. It provides relative rating values for coatings with considerably different degrees of adhesion.

The pull-off strength[49] or adhesion of a coating is measured by applying an increasing tensile force perpendicular to the coating surface until a plug of material is detached. Alternatively, this test can be a pass or fail test if a prescribed stress is applied and it is determined if the surface remains intact under this stress. The tests are carried out with a portable device with a loading fixture that is secured to the coating surface with an adhesive. The adhesive is either a two-part epoxide or acrylic system. The fixture is aligned normal to the surface, and the tensile stress is applied in a slow (less than 1 MPa s^{-1}, 150 psi s^{-1}), continuous manner until a plug of material is removed. The force attained at failure or at maximum force applied is reported. In addition, the plug is examined to determine the percentage adhesive and

cohesive failures, and the interfaces and layers involved in failure are reported.

4.5 Hardness

Coating hardness is the ability to resist permanent indentation, scratching, cutting, and penetration by a hard object.[50] Different methods of evaluating hardness yield different results because they measure different qualities of the material. There is no absolute scale and each method has its own scale of defined hardness.

Determining hardness by gouging or scratching the coating with drawing leads or wood pencils of different hardness (from 6B to 6H) is simple and inexpensive; it is widely used in laboratory development work and production control testing.[51] To conduct the test, the pencil is sharpened with a draftsman-type sharpener. The sharpened lead point is then held at a 90° angle to horizontal on No. 400 grit abrasive paper and rubbed until a smooth, flat, circular cross-section is obtained. To carry out the test, the coated panel is firmly held on a level surface and the hardest, sharpened pencil is held on the coating at a 45° angle. The pencil is then pushed away from the operator while using sufficient downward pressure to either cut through (gouge) or scratch the film or to crumple the edge of the lead. This procedure is repeated with softer and softer leads until a pencil is found that will not cut through or scratch the coating. The gouge hardness is reported as the hardest pencil that will leave the coating uncut for a push stroke of at least 3 mm (0.125 in). The scratch hardness is reported as the hardest pencil that will not scratch the coating. Because of the nature of this test, it is operator dependent and results may vary between different operators and laboratories.

Indentation hardness of coatings is determined with sophisticated devices that determine the resistance to penetration by an indenter.[52] Knoop indentation hardness (method A) is determined by bringing a pyramidal diamond indenter into contact with the coating and then applying a selected load to the indenter and maintaining the load for 18 ± 0.5 s. After this time, the indenter is withdrawn. The Knoop device is equipped with a microscope that has a movable micrometer stage; immediately after the indenter is withdrawn, the microscope is adjusted and focused so that the indentation is in the field of the microscope. With the indentation sharply focused, the length of the long indentation diagonal is determined. The indentation length is converted into Knoop hardness numbers (KHN) with tables supplied by the instrument manufacturer. If the tables are not available, KHN hardness numbers can be calculated.

Pfund indentation hardness (method B) is determined with a device equipped with a microscope that will apply a 1.0-kg (2.2 lb) load to a hemispherical (3.18 mm, 0.125 in radius) transparent quartz or sapphire indenter that is in contact with the coating surface. The load is applied for 60 s; after that time, with the loaded indenter still in place, the diameter of the circular impression is rapidly measured. An instrument constant (1.27) is then divided by the square of the indentation diameter in millimeters to obtain the Pfund hardness numbers (PHN).

Numerous other methods exist that determine hardness by scraping and indenting as well as by marring and abrasion.[50] In addition, hardness can be measured with pendulum damping devices[53] and rocker devices.[54] The three methods described above are in common usage.

4.6 Abrasion

Abrasion resistance is the ability of a coating to resist having its original appearance and structure altered when it is subjected to the influence of erosion, rubbing, scraping, or other ablative action.[55] Both temperature and environment can have an effect on abrasion resistance, but the relationship between these factors and interrelated mechanical properties is not simple. For example, hardness and modulus increase with decreasing temperature, and this may be detrimental to abrasion resistance if the coating loses flexibility or toughness. Increases in humidity around an object or subjecting an object to a moist environment as in washing a wall can soften a coating and alter its resistance to abrasion. Other factors that can have an effect on abrasion resistance include toughness, strength, and other mechanical properties. Because of the complex interrelationship between characteristics that affect abrasion resistance, it is important that the test method subjects test specimens to conditions that are similar to those encountered in actual use.

Many, it not almost all, coated items are subjected to some sort of ablative action, and such action can cause marring and/or wear. These items include appliances, automobiles and other transportation equipment, beverage cans, business machines, farm equipment, floors, furniture, highways (traffic paints), interior and exterior structural walls, and so on. The actions include: wind, rain, hail, and other natural periodic actions; wear that can be relatively continuous, such as automobile traffic or walking; polishing or other rubbing of furniture or an automobile with a harsh fabric; or accidental contact of a coated surface with a button, a toy, or a similar object. An example of rubbing effects caused by similar objects rubbing against each other are those that are encountered when beverage cans rub against each other in a multipack during shipping and handling.[56]

Abrasion resistance of coatings applied to flat, rigid surfaces can be measured by rotating the coating against an abrasive-filled, weighted wheel.[57] The results are reported as the number of cycles to remove a unit amount

of coating (wear cycles per 25.4 μm), as the loss in weight per cycle multiplied by 1000 (wear index), or as the weight loss determined at a specific number of cycles (weight loss). Although this test method is fairly simple to carry out, reproducibility is poor. It is recommended that results be limited to testing in only one laboratory when numerical values are to be used. Agreement of results between laboratories is markedly improved if different coatings are merely ranked rather than trying to compare numerical values.

Coatings on non-planar surfaces such as those found on pipelines are tested for abrasion resistance by placing the externally coated pipe through a specially designed steel drum apparatus and eroding it with an aqueous, abrasive slurry contained in the horizontally revolving drum.[58] The specimens are electrically insulated from contact with the drum. The test is applicable to all types of electrical insulating coating including thermoplastic and thermoset coatings and bituminous materials. Measurement of electrical resistance changes between the pipe and the drum as the coating erodes indicates the coating abrasion resistance. Because of electrical requirements, metallic protective coatings such as zinc are not tested by this method. However, such coatings and others designed to function as electrical barriers are tested for cathodic disbonding by other tests.[59,60]

Painted interior house walls are soiled near doorways, windows, play areas, cooking areas, etc. Such soiled areas and often the entire walls are cleaned by repeated scrubbings, and during the scrubbing the paint is subjected to corrosion. The relative erosion resistance of interior, flat wall paints to wet scrubbing can be determined by applying the paint to particular black plastic panels and scrubbing with a nylon bristle brush that is wet with an aqueous, detergent-based scrub medium.[61] The wet brush is driven in one direction across the coated surface; after each set of 400 cycles the brush is removed, the scrub medium is replenished, and the brush is replaced. This procedure is repeated until the paint film has been removed. The number of cycles to failure is reported. The test is designed for freshly painted surfaces rather than aged surfaces. The degree of erosion of exterior paints, which occurs mainly by chalking, can be determined by comparison of the surface with pictorial standards.[62] Pictorial comparisons are also used to evaluate the wear resistance of traffic paints.[63]

Abrasion resistance is also determined by air-blasting silicon carbide grains at the coated panel at a 45 g min^{-1} flow rate.[64] The abating is continued until the coating is worn through. At this point, the blasting is terminated, and the amount of ablative used is determined. The abrasion resistance is determined as the grams of ablative used per 25.4-μm film thickness. A similar test involves dropping a silica or silicon carbide abrasive through a guide tube from a specified height onto a coated planar surface.[65] Silica (sand) is a milder abrasive than silicon carbide and the slower rate of abrasion it causes can be useful in discriminating between different coatings. Basically this test differs from the previous test in that the ablative contacts the coating under a gravity flow rate rather than an air-forced flow rate. The abrasion resistance is determined as the kilograms of ablative used per 25.4-μm film thickness.

4.7 Slip

Usually slip,[66] meaning the opposite of traction or clinging, is not an inherent property of coatings unless one is discussing the tetrafluoroethylene polymers and copolymers. Coatings are said to have good slip when they have a low coefficient of friction and poor slip when they have a high coefficient of friction. Slip indicates the ease with which two contacting surfaces can move by each other. Coatings are said to have slip when they have a tack-free surface and behave as if they were lubricated.

Slip is an important characteristic of coated objects for it is the property that allows coated materials to slide by one another in forming operations, during filling, handling and shipping, as well as in other manufacturing and use operations. However, it is worth pointing out that the surface can be too slippery, that is have too low a coefficient of friction. Imagine a beverage can with a surface so slippery that it could not easily be held in a person's hand. Also, a low coefficient of friction can be an undesirable characteristic in floor coatings, since people walking on the surface could slip and fall or vehicles could slide and cause damage or harm. Gymnasium floors, porch and deck floors, concrete work-area floors, and kitchen floors are areas where this is of particular concern.

Slip can be imparted to films by incorporating a compound into a coating formulation that is incompatible with the dried or cured coating; it will then exude to the surface of the coating. A way of imparting slip to flat, coated metal sheets is to spray lightly a very low-volatility lubricant onto the coating just after the coating is cured and prior to stacking for the next manufacturing operation. Compounds such as wax esters, fatty esters, alkanolamides, metallic stearates, waxes, and silicones are used to decrease frictional resistance or to control slip.

Slip is determined by measuring the frictional properties of coatings. Friction is the force between surfaces that opposes imposed sliding motion. It is the characteristic that determines the resistance to slip or the magnitude of slip.

In one method,[67] the static friction of coatings is determined by an inclined plane sliding test or a horizontal pull test. The inclined plane test employs one or more weighted sleds that are individually placed on the coated

surface, which is fixed to a flat, movable surface. The movable surface is then inclined from the horizontal at a rate of $1.5 \pm 0.5^\circ\ s^{-1}$ until the sled begins to slide down the inclined coating surface. The tangent of the angle of inclination at this point is reported as the static friction. The horizontal pull test has a weighted sled placed on a specimen that is fixed to a flat, horizontal base. The sled is then pulled across the specimen with a mechanical power unit, and the force required to start the sled moving is determined. This force divided by the mass of the sled is reported as the static friction. Static friction determined by this method is useful for ascertaining the slipperiness of floor polishes, the slip resistance of footwear on floor tiles and floor coatings, the appropriateness of coatings for the exterior of cans, etc. The measurements are also useful to determine the effect of coating additives or spray lubricants on the slipperiness of coatings. A number of methods for determining friction can be found in the literature.[66]

4.8 Stress in Coatings

Stresses can develop within coatings during film formation, through temperature changes, and through relative humidity (RH) changes.[15] These internal stresses have an effect on coating degradation. They affect adhesion and/or cohesion and have an effect on delamination and cracking. Thermoset coatings have higher internal stresses than coatings that do not involve cross-linking compounds, such as lacquers and alkyds.

Although internal stresses can have a detrimental effect on adhesion, they originate through the process of adhesion. This seeming paradox can be readily understood if the following is considered. To protect a substrate adequately, good adhesion between the substrate and coating is required. However, adhesion causes immobility of the coating at and near this interfacial area, which, in turn, does not allow the coating to move in a normal manner, for example when the temperature changes.

When a solid coating film forms, a liquid is changed into a solid. While the film is liquid, the coating is mobile and volume contraction can take place with no stress development. As a solid coating film forms, in almost every instance contraction continues to take place but is restricted by adhesion. As a result of this restriction, tensile stresses develop within the coating. However, as soon as stress develops, the molecules seek to relieve the stress and a relaxation process begins. Therefore, as film development continues, stresses within the film can increase, decrease, or remain constant depending on the rate of stress development and of stress relaxation. It should be noted that stress development begins when the T_g of the changing system is reached. In the case of a coating that is formed from a solution of polymer, this is at the point where the T_g of the solvent/polymer solution is equal to the experimental temperature.

If it is assumed that the internal stress is in a plane parallel to the substrate and is isotropic in nature, the internal strain ε_i can be described by Equation (7):

$$\varepsilon_i = \frac{V_s - V_t}{3V_s} \tag{7}$$

where V_s is the coating volume at the solidification point and V_t is the coating volume at time t after solidification. It is readily apparent that, as the volume decreases as a function of time during final film formation, the internal strain and, therefore, the internal stress increase.

Changes in temperature will cause the dimensions of a coating/substrate combination to change. Since the expansion coefficients of the coating, $\alpha_{c,T}$, and the substrate, $\alpha_{s,T}$, are almost always different, an internal strain, ε_T, is set up. This is described by Equation (8).

$$\varepsilon_T = (\alpha_{c,T} - \alpha_{s,T})\,\Delta T \tag{8}$$

Since absorption and desorption of water can cause similar changes in dimensions of the coating, $\alpha_{c,RH}$, and substrate, $\alpha_{s,RH}$, Equation (9) similarly expresses the internal strain that is caused by RH changes.

$$\varepsilon_{RH} = (\alpha_{c,RH} - \alpha_{s,RH})\,\Delta RH \tag{9}$$

These stresses act together and may augment each other and be very important or they may negate each other and be small and relatively unimportant (Equation 10).

$$\sigma_{total} = \sigma_i \pm \sigma_T \pm \sigma_{RH} \tag{10}$$

The component σ_i is always positive, but the contributions from temperature and RH effects can be positive or negative. Positive effects occur in coatings that tend to contract and set up internal tensile stresses. Negative effects occur in coatings that tend to expand and set up compressive stresses. A dry, cold winter day will involve low temperatures and low RH, with high resultant internal tensile stresses. Conversely, a humid, summer day will involve high temperatures and RH, with resultant high internal compressive stresses.

There are a number of ways[15] to measure internal stresses, including brittle lacquer materials, cantilever beams, optical, strain gauges, and X-ray diffraction. The cantilever beam method is most widely used and gives suitable measurements. This method depends on the fact that a coating under stress on a substrate will deflect in the direction that will relieve the stress. There are two types of cantilever beam used. A one-side coated substrate is either fixed at one end or is freely supported on two knife edges. The deflection in either case can be measured;

if the elastic properties of the substrate are known, the internal stress can be calculated.

4.9 Chemical Resistance

Coatings are the first-line defense for a product that contacts hostile environments. They protect many products from a variety of chemicals.[28] In addition to protecting the product, it is preferable that the coating does not stain, does not lose adhesion, does not lose gloss, and is not permanently altered in any way by its contact with the hostile conditions.

Household chemicals include alkaline and acid solutions, beverages, condiments, cosmetics, edible and inedible oils and greases, ethyl alcohol, fruit juices, hot and cold water, soap and detergent solutions, vinegar, as well as many other common compounds. These compounds can be placed on coatings and either left open to the air or covered by a watch-glass to determine the effect of the compounds on adhesion, blistering, gloss, softening, and other properties of the coating.[29] Furniture finishes are tested for resistance to alcohol, boiling water, cosmetics, hot coffee, and other chemicals.[68] Each compound is examined in a particular manner. For example, hot coffee is poured onto a coated panel held in a horizontal position and allowed to dry. The coating is then examined for spotting, softening, graying, staining, or any other deterioration. Cosmetics are applied to the coating and placed overnight in a 50 °C oven. The coating is then examined for film failure and discoloration.

Various coatings used in the transportation industry are tested in a somewhat similar manner.[69] However, in some instances, the test includes exposure to either sunlight or ultraviolet light for a specified time; increased temperature is an important additional variable if hot, sunny climates are being simulated. Some of the chemicals important to transportation coatings are alcoholic windshield-washing solutions, antifreeze compounds, gasoline, hydraulic fluids, lubricating greases and oils, polishes, and road oils and tars. An immersion technique[70] is used to determine the solvent and fuel resistance of traffic paint. Important factors are adhesion loss, blistering, softening, and wrinkling. Other immersion techniques exist for examining chemical resistance.[28]

Resistance to moisture and water is determined in accelerated cabinets that are intended to determine long-term durability. However immersion in warm water (37.7 °C) can be used for comparison purposes.[71] Resistance to aqueous saline fogs is particularly important for aircraft, automobile, and marine coatings or for coatings that are used near oceans or in salty road conditions.[72] In this test, X-scored coated metal panels are placed in a cabinet and exposed to a condensing fog of an aqueous sodium chloride solution. The panels are periodically examined to determine the degree of rusting. Details for nonmandatory construction of a suitable cabinet are given in the test method.

5 END-USES

Earlier discussions of coating uses indicated the wide variety of end-uses. Testing of coatings can be fundamental and sophisticated in nature or it can be quick and simple to accomplish. Testing may be done to meet a set of specifications set by the supplier and by the seller or user or it may be generalized to meet what a company or group feels is important to a number of end-uses. Simple tests include the pencil hardness test,[51] rubbing tests for solvent resistance (rubbing a coating with a solvent such as acetone, methyl ethyl ketone, or xylene until it fails or passes a given number), or merely scratching a nickel coin across the coating while applying pressure. These simple tests tell one skilled in the art if the coating will be hard or soft, if it will be solvent resistant, and if it is tough and adherent while having mar resistance. Combinations of such test are useful to the in-house experts who wish to test coatings as they are produced. Each expert readily knows if the tests or others they have devised will mean that the coating being manufactured will meet requirements. However, particular end-uses require the use of specific tests, some of which were mentioned above. Others are developed to meet the requirements of specific industries.

5.1 Tests Required for Specific End-uses

Although the specific tests described below are referenced, often a buyer and a seller will define certain requirements that must be met by tests such as these or those that will define mutually acceptable testing criteria.

Aerospace and aircraft coatings[73] are tested for adhesion under ambient conditions and under specific environments, such as after water immersion with the Scotch tape peel test.[47] Scrape adhesion[48] is also important. Flexibility is determined by the mandrel bend test[33] at temperatures as low as −51 °C. Toughness is determined by measuring impact resistance with the falling weight test (see above)[40] and/or with a G. W. Impact-flexibility Tester.[74] The latter test involves dropping a steel cylinder that has spherical knobs on its surface onto a coated panel. The knobs will subject the coating to elongations of 0.5–60% under the conditions of the test. Other important mechanical properties include hardness and mar resistance.[48,51] Tensile properties are determined with free films.

The aluminum and steel beverage container industry[56] is a large consumer of coatings and requires coatings to have high-quality mechanical characteristics in addition to meeting governmental regulations for safety and health. The inside and the outside of the cans are coated but with different coatings. The inside coating provides protection of the metal can from its contents as well as protection of the contents from taste alteration or contamination by contact with the metal can. The outside coating provides attractive, product identification. Important mechanical properties include abrasion resistance,[57] adhesion,[47] hardness,[51] and flexibility.[38,40]

Pipelines carry oil, natural gas, water, and chemicals to plants and to consumers. Often the product is transported for long distances and often these pipes are buried underground and, thus, have a constant pressure applied. During the burial process, dirt, stones, and rocks are thrown into the trench and the coatings must have sufficient toughness to withstand this rough handling. Chemical plants and refineries use extensive pipeline systems to carry raw materials to reactors, intermediates to separation or other reaction systems, and final products to shipping facilities. The coatings must have sufficient integrity to withstand the hostile environments associated with such use. Above- and below-ground pipes are subjected to expansions and contractions as the temperature changes; again toughness and flexibility are important factors. Although pipelines are efficient means to carry out such operations, they must be protected by coatings to ensure dependable service and long life.[75]

Automotive coatings[76] are subjected to mechanical abuse from ordinary usage and from nature. Although individual automobile manufacturers have specific testing protocols involving test methods devised in-house, they often use many of the tests described above. A variety of substrates are involved and range from those that are flexible to those that are rigid; the coatings range from primers, to guide coats, to topcoats, etc. Hardness is determined by the Tukon indentation method[52] and wear resistance by the Taber Abraser method.[57] Adhesion is determined by either X-scribing or cross-hatching and applying pressure-sensitive tape under ambient conditions[47] and after exposure to 100% RH.[77] Coatings must have at least 99% adhesion when tested by these methods. To test resistance to impact from stones and road debris, a gravelometer test is used.[78] Toughness is ascertained by scraping coatings on rigid substrates with a dime and by scraping with a knife when the coating is on flexible substrates. When these scraping tests are used, the coating should not flake, peel, or lose adhesion. Resistance to water[71] and to saline solutions are important aspects of transportation coatings.

Testing of coatings is usually carried out before the coating is chosen for an end-use in order to assess when or how the coating will fail, if it does, when in use. Many coatings are used to protect substrates and thus preserve materials and conserve natural resources; as such, coatings are environmentally sound in nature. Although they can be well designed, coatings can and do fail. Therefore, it is also important to investigate coatings when they have failed to determine the reasons for failure. Mills[79] has given an excellent description of coating failure analysis.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
DMA	Dynamic Mechanical Analysis
KHN	Knoop Hardness Numbers
PHN	Pfund Hardness Numbers
RH	Relative Humidity

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction • Atomic Spectroscopy in Coatings Analysis • Microscopy of Coatings • Rheology in Coatings, Principles and Methods

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Dynamic Mechanical Analysis of Polymers and Rubbers • Mechanical Properties of Polymers and Rubbers

REFERENCES

1. B. Gavett, 'Artists Paints', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 60, 1995.
2. R.B. Seymour, H.F. Mark, *Organic Coatings: Their Origin and Development*, Elsevier Science, New York, 1990.
3. Federation of Societies for Coating Technology, *Sciquest* (CD-ROM), Federation of Societies for Coating Technology, Blue Bell, PA, Vol. 1, 1995.
4. P.G. Bahn, *Nature*, **356**, 109 (1992).
5. C.R. Martens, *Waterborne Coatings: Emulsion and Water-soluble Paints*, Van Nostrand Reinhold, New York, 1981.
6. H.E. Ashton, 'Architectural Coatings', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 59, 1995.

7. ASTM D 2833, 'Standard Index of Methods for Testing Architectural Paints and Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
8. ASTM D 3730, 'Testing High-performance Interior Architectural Wall Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
9. ASTM D 5146, 'Testing Solvent-borne Architectural Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
10. ASTM D 5324, 'Testing Water-borne Architectural Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
11. ASTM D 2486, 'Scrub Resistance of Interior Latex Flat Wall Paints', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
12. ASTM D 3450, 'Washability Properties of Interior Architectural Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
13. ASTM D 4213, 'Wet Abrasion Resistance of Interior Paints', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
14. ASTM D 4828, 'Practical Washability of Organic Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
15. D.Y. Perera, 'Stress Phenomena in Organic Coatings', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 49, 1995.
16. V.D. Sherbondy, 'Accelerated Weathering', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 53, 1995.
17. L.S. Hicks, M.J. Crewdson, 'Natural Weathering', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 52, 1995.
18. Z.W. Wicks, Jr, *Federation Series on Coatings Technology: Film Formation*, Federation of Societies for Coatings Technology, Blue Bell, PA, 1986.
19. R.D. Athey, Jr, 'Film Preparation for Coating Tests', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 37, 1995.
20. T.J. Miranda, 'Cure: The Process and Its Measurement', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 36, 1995.
21. ASTM D 4708, 'Preparation of Free Films of Organic Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
22. ASTM D 4712, 'Testing Industrial Water-reducible Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
23. ASTM D 1640, 'Drying, Curing or Film Formation of Organic Coatings at Room Temperature', in *Annual Book of ASTM Standards*, V 06.02, ASTM, West Conshohocken, PA.
24. ASTM D 2345, 'Minimum Film Formation Temperature of Emulsion Vehicles', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
25. ASTM D 1653, 'Water Vapor Permeability of Organic Coating Films', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
26. ASTM D 2793, 'Block Resistance of Organic Coatings on Wood Substrates', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
27. ASTM D 4752, 'Measuring MEK resistance of Ethyl Silicate (Inorganic) Zinc-rich Primers by Solvent Rub', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
28. A.H. Brandau, 'Chemical Resistance', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 55, 1995.
29. ASTM D 1308, 'Effect of Household Chemicals on Clear and Pigmented Organic Finishes', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
30. L.W. Hill, *Federation Series on Coatings Technology, Mechanical Properties of Coatings*, Federation of Societies for Coatings Technology, Philadelphia, PA, 1987.
31. L.W. Hill, K. Kozlowski, 'The Relationship between Dynamic Mechanical Measurements and Coatings Properties', in *Proceedings of the 12th International Conference in Organic Coatings Science and Technology, Advances in Organic Coatings Science and Technology*, ed. A.V. Patsis, Technomic Inc., Lancaster, PA, Vol. 10, 1986.
32. ASTM D 2370, 'Tensile Properties of Organic Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
33. ASTM D 522, 'Mandrel Bend Test of Attached Organic Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
34. ASTM D 4885, 'Initial/Offset/Secant Modulus – Geomembranes by Wide Strip Tensile Method', in *Annual Book of ASTM Standards*, V 04.08, ASTM, West Conshohocken, PA.
35. L.E. Nielsen, *Mechanical Properties of Polymers*, Reinhold' New York, 1962.
36. L.W. Hill, 'Dynamic Mechanical and Tensile Proprieties', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 46, 1995.
37. M.P. Morse, 'Flexibility and Toughness', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 47, 1995.
38. ASTM D 4145, 'Coating Flexibility of Precoated Sheet', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
39. ASTM G 10, 'Specific Bendability of Pipeline Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
40. ASTM D 2794, 'Resistance of Organic Coatings to the Effects of Rapid Deformation (Impact)', in *Annual*

Book of ASTM Standards, V 06.01, ASTM, West Conshohocken, PA.
41. ASTM D 3281, 'Formability of Attached Organic Coating with Impact-wedge Bend Apparatus', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
42. ASTM D 4146, 'Formability of Zinc-rich Primer/Chromate Complex Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
43. ASTM G 13, 'Impact Resistance of Pipeline Coatings (Limestone Drop Test)', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
44. ASTM G 14, 'Impact Resistance of Pipeline Coatings (Falling Weight Test)', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
45. ASTM G 17, 'Penetration Resistance of Pipeline Coatings (Blunt Rod)', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
46. G.L. Nelson, 'Adhesion', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 44, 1995.
47. ASTM D 3359, 'Measuring Adhesion by Tape Test', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
48. ASTM 2197, 'Adhesion of Organic Coatings by Scrape Adhesion', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
49. ASTM D 4541, 'Pull-off Strength of Coatings Using Portable Adhesion Testers', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
50. P.R. Guevin, Jr, 'Hardness', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 48, 1995.
51. ASTM D 3363, 'Film Hardness by Pencil Test', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
52. ASTM D 1474, 'Indentation Hardness of Organic Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
53. ASTM D 4366, 'Hardness of Organic Coating by Pendulum Damping Tests', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
54. ASTM D 2134, 'Softening of Organic Coatings by Plastic Compositions', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
55. M.P. Morse, 'Abrasion Resistance', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 45, 1995.
56. M.B. Price, 'Can Coatings', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 62, 1995.
57. ASTM D 4060, 'Abrasion Resistance of Organic Coatings by the Taber Abraser', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
58. ASTM G 6, 'Abrasion Resistance of Pipeline Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
59. ASTM G 8, 'Cathodic Disbonding of Pipeline Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
60. ASTM G 80, 'Specific Cathodic Disbonding of Pipeline Coatings', in *Annual Book of ASTM Standards*, V 14.02, ASTM, West Conshohocken, PA.
61. ASTM D 2486, 'Scrub Resistance of Interior Latex Flat Wall Paints', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
62. ASTM D 662, 'Evaluating Degree of Erosion of Exterior Paints', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
63. ASTM D 913, 'Evaluating Degree of Resistance to Wear of Traffic Paint', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
64. ASTM D 658, 'Abrasion Resistance of Organic Coatings by Air Blast Abrasive', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
65. ASTM D 968, 'Abrasion Resistance of Organic Coatings by Falling Abrasive', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
66. P.R. Guevin, 'Slip Resistance', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 50, 1995.
67. ASTM D 4518, 'Measuring Static Friction of Coating Surfaces', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
68. ASTM D 2571, 'Guide for Testing Wood Furniture Lacquers', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
69. ASTM D 1540, 'Effect of Staining Agents on Organic Finishes Used in the Transportation Industry', in *Annual Book of ASTM Standards*, V 06. 01, ASTM, West Conshohocken, PA.
70. ASTM D 2792, 'Solvent and Fuel Resistance of Traffic Paint', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
71. ASTM D 870, 'Water Resistance of Coatings Using Water Immersion', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
72. ASTM B 117, 'Salt Spray (Fog) Testing', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
73. C.R. Hegedus, S.J. Spadafora, D.F. Pulley, A.T. Eng, D.J. Hirst, 'Aerospace and Aircraft Coatings', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 58, 1995.
74. *Federal Test Method Standard 141B, Method 6226, Impact Flexibility*, US Government Printing Office, Washington, DC, 1979.
75. L.B. Odell, A.L. Siegmund, 'Pipeline Coatings', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 64, 1995.

76. R.A. Ryntz, 'Automotive Product Tests', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 61, 1995.
77. ASTM D 2247, 'Water Resistance of Coatings in 100% Relative Humidity', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
78. ASTM D 3170, 'Chip Resistance of Coatings', in *Annual Book of ASTM Standards*, V 06.01, ASTM, West Conshohocken, PA.
79. G.D. Mills, 'The Analysis of Coatings Failures', in *Paint and Coating Testing Manual*, 14th edition, ed. J.V. Koleske, ASTM, Philadelphia, PA, Chapter 69, 1995.

Microscopy of Coatings

Edda Rädlein
Institut für Nichtmetallische Werkstoffe, Clausthal-Zellerfeld, Germany

1 The Role of Imaging Techniques During Development and Investigation of Coatings 1
2 Methods 2
2.1 History of Modern Microscopy Techniques 2
2.2 Principles of Scanning Probe Microscopy 2
2.3 Principles of Atomic Force Microscopy 3
2.4 Operating Modes of the Atomic Force Microscope 4
2.5 Various Types of Scanning Probe Microscopy Instrument 7
2.6 Comparison of Microscopy Methods 10
2.7 Determination of Roughness 11
3 Sample Preparation 13
3.1 Cleaning 13
3.2 Cross-sections of Coatings 14
3.3 Fixation, Coated Fibers and Small Particles 14
3.4 Preparation for Scanning Tunneling Microscopy 15
4 Artifacts and Calibration 15
4.1 Data Processing and Measurement System 15
4.2 Tip Geometry 15
4.3 Sample Response 18
4.4 Calibration 18
5 Applications 19
5.1 Nucleation, Oxidation 19
5.2 Surface of Coatings: Defects, Roughness 20
5.3 Thickness 22
5.4 Microstructure and Nanostructure 23
5.5 Mechanical Properties of Coatings 26
5.6 Corrosion and Aging 28
Acknowledgments 28
Abbreviations and Acronyms 28
Related Articles 29
References 29

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Fundamental research on films and the development and optimization of coatings are not possible without high-resolution imaging. This article describes the advantages and disadvantages of the newly introduced scanning probe microscopy (SPM) in comparison with well-established microscopy techniques with light and particles [e.g. scanning electron microscopy (SEM) and transmission electron microscopy (TEM)], which are limited by the diffraction limit, i.e. 200 nm for conventional light microscopy and 4 nm for SEM. This limit does not hold for SPM, which uses near-field interactions for imaging. Thus, even simple atoms or molecules can be resolved on well-ordered structures by scanning tunneling microscopy (STM) or atomic force microscopy (AFM). Both possess an extremely good height sensitivity of 0.02 nm.[1] In the constant-height operation mode, AFM can detect vertical forces as low as 10^{-15}–10^{-4} N. Atomic resolution is not always possible or not always necessary for many practical applications; a lateral resolution of a few nanometers is sufficient. Besides the capability of achieving this value with ease, simple sample preparation, versatile environmental conditions and observation of reactions in situ are advantages of STM and AFM.

Especially AFM has proved to be a useful technique for investigating coatings, because it is independent of conductivity and provides topographic data together with information on structure, local variations of mechanical properties and surface alterations in different media. However, some artifacts and restrictions have to be considered which can differ considerably from customary electron micrographs with comparable resolutions, such as onset overswing and tip convolution.

Microscopy helps in solving coating problems according to substrate quality, nucleation, wettability, thickness, roughness, microstructure, crystal phases (texture, orientation), lateral and vertical homogeneity, density, porosity, interfaces, adhesion, hardness, stress, scratch and abrasion resistance, tribology, wear and corrosion. Practical examples of microscopy results on these questions are presented together with suitable preparation methods and interpretation of artifacts.

1 THE ROLE OF IMAGING TECHNIQUES DURING DEVELOPMENT AND INVESTIGATION OF COATINGS

The performance of coatings is dependent on their morphology to a great extent. Microscopy is an indispensable tool for the development of coatings as it provides topographic information from nucleation to corrosion and wear during application. Thin coatings with a thickness below 1 μm are becoming more and more important and

many different coating and substrate materials are being combined, hence instruments with high resolution are demanded. Instruments are preferred which can cope, e.g. with both conducting and nonconducting materials and combine topographic information with information on other properties such as crystal structure or local variations in chemistry. Because of their versatility, simple handling and unpretentious sample preparation, the use of scanning probe microscopes has expanded rapidly and is still expanding.

2 METHODS

2.1 History of Modern Microscopy Techniques

Microscopy means magnified imaging of objects. Conventional microscopy techniques can be classified with respect to the type of interaction that they use and the resulting magnification:

1. electromagnetic radiation, waves:
   (a) optical microscopy
   (b) X-ray microscopy
   (c) scanning acoustic microscopy
2. particles:
   (a) SEM
   (b) TEM
   (c) field ion microscopy (FIM).

Two-dimensional mapping and magnification of objects are achieved by either geometrical optics or scanning in small steps. The methods mentioned above detect reflected or secondary electromagnetic or particle beams in the far field.

The newly developed scanning probe microscopes measure near-field interactions. Therefore, they are not restricted by the length of the probing wave.

The first ideas on exploiting near-field interactions for microscopy were published by Synge.[2] In 1928 he proposed scanning an illuminated sample with a fine glass tip and measuring the near-field interaction of the evanescent electromagnetic field. With the help of an extremely fine aperture, the Abbe limit of diffraction can be avoided and the lateral resolution is no longer limited to half of the wavelength of the applied beam. Synge even proposed piezoelectric material for scanners.[3] However, it was not before the 1980s that piezoelectric actuators with sufficient precision were available and his ideas were realized.

In 1986, Gerhard Binnig and Heinrich Rohrer received the Nobel Prize for Physics for the realization of the first SPM technique, STM, which 4 years before had visualized atoms for the first time.[4–6] In 1985, Binnig et al. developed also the second type of SPM technique, AFM, together with Gerber and Quate.[7]

2.2 Principles of Scanning Probe Microscopy

The common principle of all scanning probe microscopes is depicted in Figure 1. A fine tip is brought close to the sample surface. Within distances of few nanometers or even less than 1 nm, near-field interactions are measured (see Table 1[9]). Scanning either the tip or the sample in the x and y directions with a step precision of 1 pm[10] results in images of the height or other sample properties.

The measured types of interaction are as follows:

- the current of electrons tunneling between a conductive sample and a conductive tip in STM;
- Coulomb, Van der Waals and other forces between the sample and a fine tip on a spring cantilever in AFM;
- the evanescent electromagnetic field which is detected with a glass fiber above an illuminated sample in scanning near-field optical microscopy (SNOM).

Further types of SPM instruments are described in section 2.5. In this section, only a brief explanation of the method principles is given, as far as is necessary to comprehend preparation methods and possible artifacts during the measurement of coatings. Fundamentals and further information on SPM can be found in the article

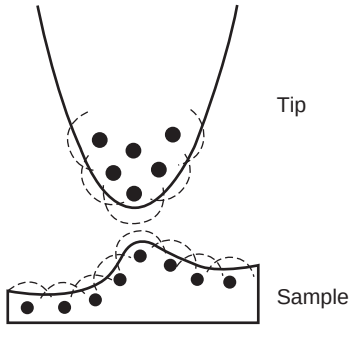

Figure 1 Common principle of all scanning probe microscopes: near-field interactions between a fine tip and the sample surface serve for imaging.[8]

Table 1 Typical interaction distances for SPM[9]

Method	Tip–sample distance
STM	1 nm
Contact AFM	<few 0.1 nm
Noncontact AFM	1 nm–some 10s nm

Scanning Probe Microscopy, Industrial Applications of and in textbooks, monographs and reviews.[11–22] For quick help on practical aspects, the book by Howland and Benatar is recommended.[9]

2.3 Principles of Atomic Force Microscopy

2.3.1 Components of an Atomic Force Microscope

Of all types of SPM, AFM is most widespread and frequently used for the investigation of coatings. In Figure 2 the five main parts of an AFM instrument are depicted:

- a fine tip at a soft cantilever spring;
- a detector for the spring deflection;
- a feedback loop which controls the movements of the microscope;
- actuators which move either the tip or the sample;
- an image processor which produces three-dimensional pictures of the $x-y$ position on the sample and the height or tip deflection as a z value.

As a first step, the tip is brought close to the sample surface manually or automatically. Attractive forces can be detected at a distance of some tens of nanometers. In the most commonly used contact mode, the tip is made to approach the surface until attractive and repulsive forces are roughly balanced, i.e. within less than 1 nm.[9] At a certain deflection of the cantilever, the actuator starts scanning the sample.

Common cantilevers are produced from Si or Si_3N_4 with integrated tips of the same material (see also section 2.3.2 on AFM tips). Typical spring constants for commercial cantilevers are 0.06–0.6 N m^{-1}. An easy way to detect the cantilever deflection is by reflection of a laser beam towards a partitioned photodiode (see Figure 2).[24–27] This optical geometry is the only way to detect extremely low forces of about 10^{-11} N. With a typical cantilever length of 100–200 μm, the optical system enhances deflections by a factor of about 1000 and allows the detection of deflections below 0.1 nm.[22]

Other ways to detect spring deflections as low as 10^{-3} nm are as follows:[28]

- tunneling contact on a conductive cantilever;[7]
- optical interferometry;[29–32]
- measurement of capacity;[28,33,34]
- piezoelectric cantilevers.[34–37]

Force detection can also be achieved with a minute quartz resonator instead of the cantilever spring.[38]

The signal of the photodiode is fed into the feedback circuit to keep the deflection constant by adjusting the height of the sample during scanning. The image processor then produces a three-dimensional image of the point $x-y$ on the sample and the z voltage of the actuator.

A reliable type of actuator is a single-tube piezoelectric ceramic,[27,39] but three orthogonally arranged piezoceramic bars[4,27] and bimorphs[40,41] have also been used. A common piezoceramic material is doped lead zirconium titanate with a typical elongation of 5 nm V^{-1}. Single-tube scanners with partitioned electrodes on their surfaces bend in the x and y directions and shrink and elongate in the z direction as shown in Figure 3.[39,42]

The maximum scan range depends on the length of the scanner. Typically, $x-y$ ranges of 1 μm^2 are achieved with a length of 1 cm. Large scanners with a length of some centimeters can scan over 100 μm^2. In Figure 2, the scanner moves the sample and the tip is fixed. In STM or AFM instruments with large sample stages or other than contact mode imaging, the sample is fixed and the tip is moved.

To protect the system from spurious vibrations, a rigid construction of the measuring head and good isolation from external vibrations are demanded. Means of vibration isolation such as silicone-rubber pads, bungee cords, eddy current damping and damping stacks with Viton® fluoroelastomer are employed for protection from structural-borne noise in the 1-Hz frequency range. The first STM system was isolated from disturbances such as door slamming by sophisticated eddy current damping

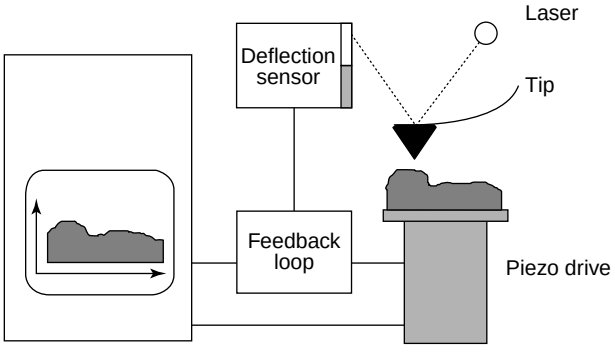

Figure 2 Schematic diagram of an atomic force microscope.[23]

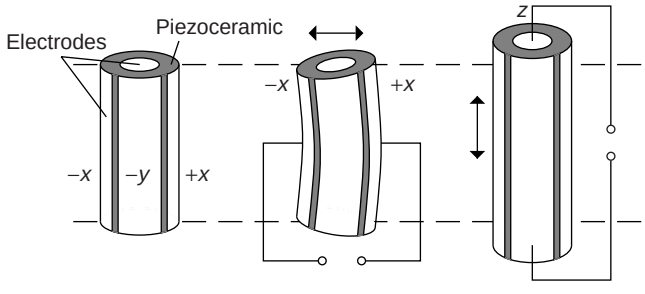

Figure 3 Function of a single-tube scanner.

with superconductive liquid He,[43] which proved to be oversized. In addition to structural-borne noise, constant sources of acoustic noise should also be avoided.

2.3.2 Tips in Atomic Force Microscopy

Commercial tips are produced via photolithography and chemical vapor deposition (CVD) processes.[44] For easy handling, the cantilever and tip are integrated into a glass or Si chip. Typical materials for cantilevers and tips are Si and Si_3N_4,[45,46] but diamond,[7,26] diamond-like carbon (DLC),[47] thermal oxide grown on Si,[25,48,49] GaAs,[50] metals[51–55] and glasses[56–58] have also been used.

The most important parameters are the spring constant k of the cantilever and the radius R of the tip apex. For a rectangular cantilever, k can be calculated from Young's modulus E, length l, width b, thickness d and density ρ of the spring according to Equation (1):[22,27,44,59]

$$k = Ed^3b(4l^3)^{-1} \tag{1}$$

However, d varies from tip to tip and cannot be measured precisely. Therefore, k is better calculated from the measured resonance frequency ν_0 as given in Equation (2):[59]

$$k = 2\pi^3 l^3 b \nu_0^3 \left(\frac{\rho^3}{E}\right)^{1/2} \tag{2}$$

For Si, density and Young's modulus are $\rho_{Si} = 2.33\,g\,cm^{-3}$, and $E_{Si} = 1.79 \times 10^{11}\,N\,m^{-2}$.[60] Many cantilevers are V-shaped, and their spring constant k_V is approximated by Equation (3):[61]

$$k_V \approx 2k \tag{3}$$

However, ρ and E also vary within wide ranges (by 50% and 600%, respectively), as the properties of thin CVD films may differ considerably from bulk material.[59,62] For optical deflection detection, gold coatings are applied to enhance the reflectivity. They also change the mechanical properties. Therefore, each sensor has to be calibrated individually for precise force measurements. A precise but time-consuming approach is to measure the resonance frequency with and without an attached additional mass;[59,63] k can than be deduced from the two resonance frequencies and the known additional mass.

Magonov and Whangbo propose to vibrate the tip by enhancing the gains, which is a fast way to determine ν_0.[22] Sensors can also be bought with individual calibration.

For many practical applications, it is sufficient to know the relative force differences in arbitrary units instead of absolute values. Then, one can rely on average k values. A sketch of the most commonly used Si_3N_4 and Si tips is given in Figure 4.

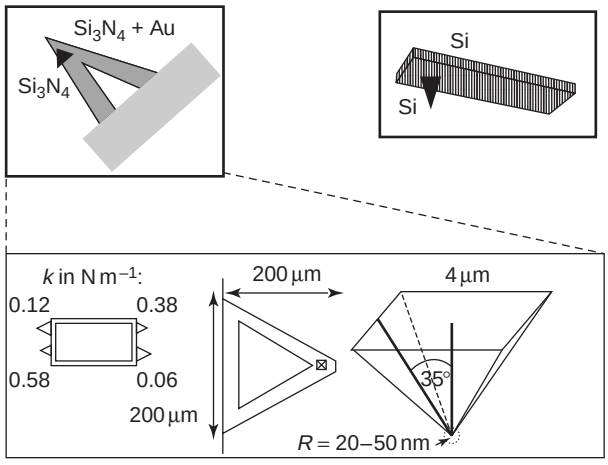

Figure 4 Sketch of the most commonly used Si_3N_4 and Si cantilevers and tips.

2.4 Operating Modes of the Atomic Force Microscope

2.4.1 Contact Atomic Force Microscopy Height and Force Mode

The simplest mode of operation of AFM, the contact mode, has been briefly explained before. Height mode images are produced from the x, y and z voltage of the scanner with constant spring deflection. For force mode images, the z voltage is kept constant and the deflection detector signal is transformed into force values. Although the physical interpretation of the force variations might be unclear (because of unknown k or unknown absolute force), these images can be valuable for the interpretation of height mode images. Owing to a different way of amplification, they are more prone to overswinging and thus grain boundaries, coating interfaces, etc, can be readily recognized.

A typical AFM image consist of 256^2, 400^2, 512^2 or 1000^2 pixels. Generally, x is called the fast scanning direction and y the slow scanning direction. The time to sample an image depends on the scan size. For 400^2 pixel images larger than $1\,\mu m^2$, a frequency of 1–8 Hz (lines per second) is useful; for images smaller than $1\,\mu m^2$, 8–60 Hz might be sufficient. Scans of $1\,\mu m^2$ can be finished after less than 1 min. This is fast enough for in situ observations of changes due to chemical reactions or physical processes which can cause changes of the coating topography.

2.4.2 Local Spectroscopy

In the STM spectroscopy mode, the tip is fixed at one point x–y and the bias voltage is varied. A current versus voltage characteristic is registered, which gives information on local electronic properties. Corresponding to this local current spectroscopy, local force spectroscopy

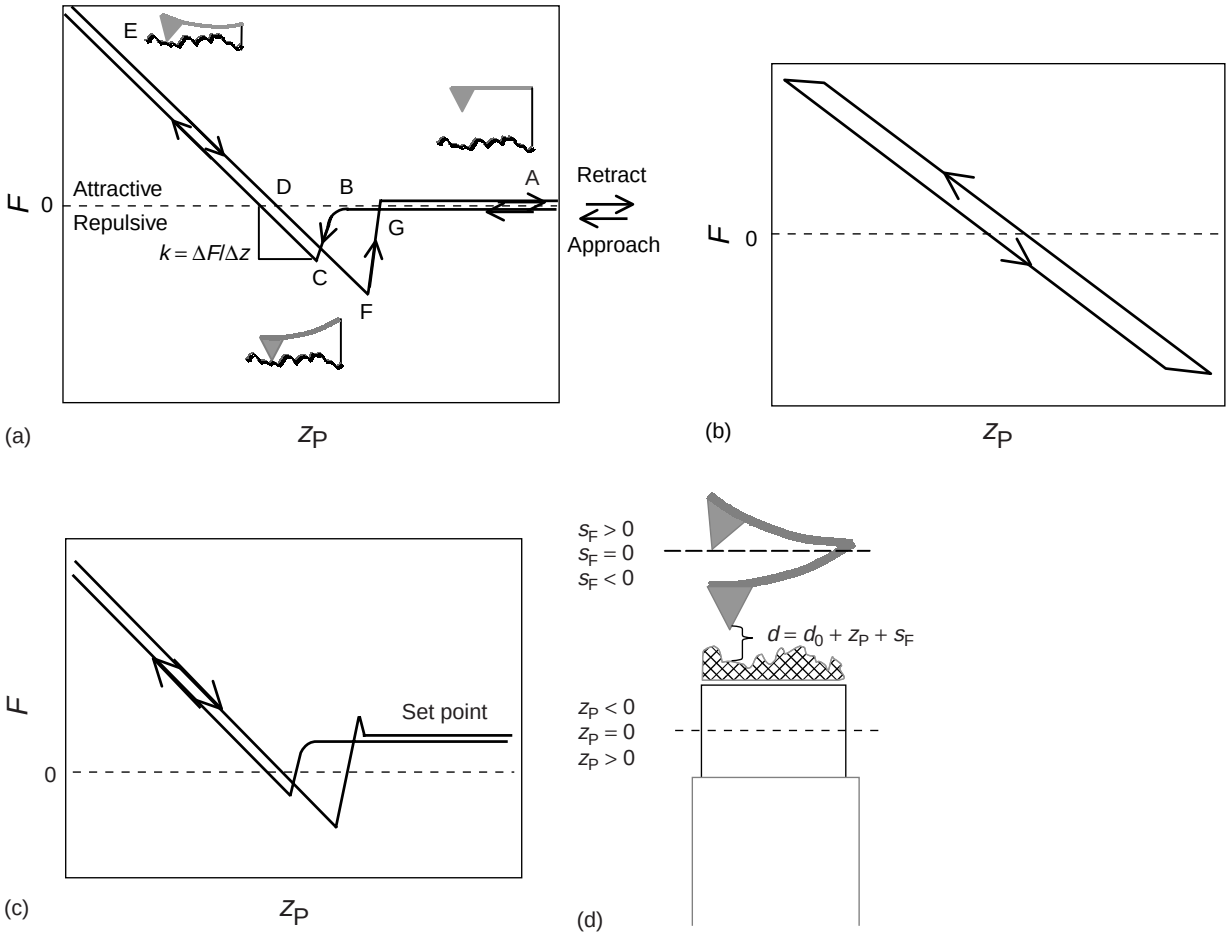

Figure 5 Idealized force–distance curves. (a) Force F at the cantilever versus piezo elongation z_P. (b) Force–distance curve (hysteresis) if the scanner elongation is not sufficient for retracting the tip from the surface. (c) Force–distance curve with optimized scanning force. (d) Tip deflection s_F and scanner elongation z_P.

(LFS) with an AFM instrument is achieved by keeping the tip at one point x–y and varying the z height.

Force–distance curves give information on local forces between the tip and sample[64,65] and local mechanical properties of coatings.[1,66] LFS is recommended for adjusting a proper tip–sample distance for delicate samples. If the Young's modulus of the tip is much larger than the sample modulus, local variation in sample elasticity can be imaged.[50] By recording force–distance curves in electrolytes, the isoelectric point of the tip and/or the sample can be determined.[67,68]

Figure 5(a) displays an idealized force–distance curve, i.e. cantilever deflection versus z-elongation of the piezo. For long distances between the tip and sample, there is no deflection (A). Approaching the range of attractive forces, the cantilever bends downwards to balance the attraction. As soon as the gradient of the attractive forces is higher than the spring constant, the tip abruptly deflects downwards and jumps to contact[69] (B–C). If the distance is reduced still further, repulsive forces occur. Between D and E, the modulus of the tip and sample surface govern the force–distance curve.[66] Because of adhesive forces and elastic, viscoelastic or plastic deformation and in the presence of liquid films, the tip has to be retracted further than the former point of jump to contact before it is set free (F–G).[70]

The gradient $k = \Delta F(\Delta z)^{-1}$ of the repulsive part of the LFS curve (C–E) depends upon the spring constant of the cantilever and the Young's modulus of the sample.[66] This part serves for investigations of the elastic properties.[1] Hardness is calculated from the retraction behavior by subtracting the elastic component. Hardness is defined by the maximum load divided by the contact area after elastic recovery of the sample. The force which is necessary to break the contact depends on the sample and can serve for material characterization.[57,60,71,72]

The great advantage over conventional indenters is that AFM can provide information also on thin coatings, as the indentation depth and load can be much smaller. Besides the thin liquid water film present in laboratory

air, other contaminant layers and lubricants can disturb the measurement and lead to additional kinks in the force–distance curve.[9]

Figure 5(b) shows the force–distance curve for too small z variations. In this case, the tip cannot spring free. Figure 5(c) gives an example of how to minimize interaction forces on delicate samples. With a contact area of 100–400 nm^2 and a load of some nanonewtons, the mechanical pressure under the tip is 10^7 Pa.[50,63] Loubet et al. gave an estimation of 3 nm for the contact radius r_C and 2×10^9 Pa for the pressure.[73] To minimize sample or tip destruction, measurements should be made at zero force. Therefore, the tip is retracted to a point where it still keeps in contact[22,69,74] and the piezo elongation for scanning is set to the setpoint depicted in Figure 5(c).

The tip–sample distance d depends on the distance d_0 before the spectroscopy measurements, the elongation z_P of the piezo and the cantilever deflection $s_F = Fk^{-1}$ (see Figure 5d) (Equation 4):

$$d \approx d_0 + z_P + s_F \tag{4}$$

Taking into account only cantilever deflection (no other deformations), one can write:

$s_F = 0$	from A to B and at D
$s_F < 0$	from B to D
$d = r_C$	from C to E (r_C = contact radius)
$s_F > 0$	from D to E

With unknown sample deformation, an exact value of d cannot be given.

Depending on the strength of the interaction and on the sample reaction, there are different types of LFS curves.[1] With strong tip–sample interaction, the tip jumps into contact and out of contact (B–C and F–G are steep). With relatively high spring constant k, the tip slides in and out (B–C and F–G are flat). On some surfaces, a mix of these two types is observed: jump in/slide out has been observed, e.g. on graphite.[1] When the tip is retracted, the graphite sticks on the tip and deforms, because of the strong tip–sample interaction and its low modulus.

With known tip radius R, the Hamaker constant A can be calculated from the interaction energy E_{SP} and the force F_{SP} in the attractive part of the LFS curve (Equations 5 and 6):

$$E_{SP} = -AR(6d)^{-1} \tag{5}$$

$$F_{SP} = -\partial E_{SP}(\partial d)^{-1} = -AR(6d^2)^{-1} = -kz \tag{6}$$

A depends on the density of the tip and sample material, their dipole interaction and the surrounding medium. From the LFS of graphite in air, Burnham and Colton calculated $A = 3.2 \times 10^{-19}$ J.[1] However, the true effective interaction radius can be different from the microscopic tip radius. The interaction is dominated by the smaller radius of curvature of rough asperities on the tip.[60]

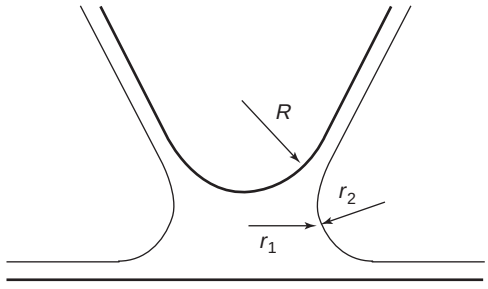

Figure 6 The Kelvin radius r_{Kelvin} is calculated from the radius r_1 of the liquid capillary and the radius of curvature r_2 of its surface.

Capillary forces affect the force measurements in a humid environment. A meniscus is formed between the tip and liquid layers if the radius of curvature of the tip–sample contact is smaller than the Kelvin radius r_{Kelvin} (see Figure 6); r_{Kelvin} is determined from the radius of the capillary, r_1, and the radius of curvature, r_2, of its surface (Equation 7):[75]

$$r_{Kelvin} = r_1 r_2 (r_1 + r_2)^{-1} \tag{7}$$

where r_1 and r_2 depend on the relative vapor pressure.[75] With a relative humidity of 90% at 20 °C, r_{Kelvin} for water is of the order of 10 nm, and at 30–40% it becomes $r_{Kelvin} \approx 1$ nm. For low vapor pressures (about 10% relative humidity), Equation (7) would render r_{Kelvin} in the range of molecular radii and it can no longer be used.[28,75] Below 10% relative humidity, the water capillary evaporates.

2.4.3 Lateral Force Microscopy

Deflection detectors for the lateral force microscopy (LFM) mode are split into four sectors and can separate normal and lateral forces.[76] Lateral force differences, i.e. torsion of the cantilever, are due to variations in friction resulting from different gradients, different roughnesses or different materials on the sample surface. By changing the fast scanning direction from parallel (ideally no lateral forces) to perpendicular (ideally maximum lateral forces) to the cantilever beam and by comparing scans in the x and $-x$ directions, roughness and height differences can be distinguished.[22] On flat surfaces, material contrasts can be visualized, which is extremely helpful for the investigation of multiphase coatings and the first stages of layer growth.

LFM opens up a new research field for tribology in the nanometer world.[73,77–81] First theoretical explanations for the nanotribological behavior of the tip–sample system are under development.[82,83]

2.4.4 Non-contact Mode (Attractive Mode)

In the non-contact mode, which was introduced by Martin et al.,[29] a relatively stiff cantilever with $k \approx 100\,\mathrm{N\,m^{-1}}$[84] is scanned within a distance from 1 nm to some tens of nanometers above the sample.[9] The tip is vibrated and the frequency shift or the variation of the free amplitude is detected. At this distance, attractive Van der Waals and capillary forces dominate,[34,85] which are about 10^3 times weaker than the repulsive forces in the contact mode. At smaller distances, covalent forces between the dangling bonds of adatoms on the sample and the top also contribute to the contrast.[86]

Noncontact mode scanning is recommended for soft surfaces which are not stable when scanned in the contact mode. Van der Waals forces can be measured with a sensitivity of 10^{-13} N and force gradients with $10^{-6}\,\mathrm{N\,m^{-1}}$.[87] However, 10–30 monolayers of $H_2O + N_2$ on samples in humid air can hinder stable noncontact mode measurements, as the thickness of these layers might be in the same range as the scanning distance.

2.4.5 Tapping™ or Intermittent Contact Mode

With tip vibration amplitudes of 10–100 nm, the liquid layer on samples under ambient conditions no longer causes unstable oscillations and the tip springs free from the layer with each oscillation.[88] Vertical forces might be high in this intermittent contact mode. Its was found, however, that vertical forces are not as detrimental for delicate surfaces as the lateral forces which occur during contact mode scanning.[22,89] Digital Instruments chose the name Tapping™ for this mode, as the tip taps each image pixel from above, instead of scratching along a line. With tapping frequencies of about 200–400 Hz, each pixel is touched a few times.[90] Height images are gained from the variation in the free amplitude.

Phase contrast can also be recorded in this mode; it can contain information on topography and on local chemical variations.[90–92]

2.4.6 Force Modulation

For force modulation measurements, the tip is also vibrated, but with a smaller amplitude than for noncontact or intermittent contact mode. The tip stays in contact with the surface and images material contrasts by recording local variations in elasticity.[93,94]

Modes with modulations of the tip–sample distance such as noncontact, intermittent contact or force modulation are called AC methods or dynamic modes. Quasi-static methods without distance modulation are called DC methods.

2.4.7 Environmental Conditions: From Ultrahigh Vacuum to Liquid Media

SPM is very versatile regarding the environment for measurement: scanning is possible in ultrahigh vacuum (UHV), in air or in liquids,[6,95,96] and temperatures between 400 and 4 K are possible.[34,35,41,42,97–100] The atmosphere can be controlled by performing SPM in a glove-box[101] or by sweeping the sample with well-defined gases.[102]

In liquids, capillary forces are avoided and Van der Waals forces are one to two orders of magnitude smaller than in humid air,[74] and electric double layers can be observed. Scanning in a liquid cell opens up new fields for AFM in biology and electrochemistry and for the investigation of crystal growth and corrosion of coating materials. Depending on the scanning frequencies, dissolution or growth rates from 10^{-10} to $10^{-16}\,\mathrm{mol\,m^{-2}\,s^{-1}}$ are observable.[103] Another way of avoiding the liquid water film is the investigation of frozen samples.[104]

2.5 Various Types of Scanning Probe Microscopy Instrument

Based on the first STM technique, more than 20 various types of near-field microscopes have been developed[105] and overviews have been published.[85,87,106–111] The main difference between these types of SPM instruments is the type of sensor, which renders locally resolved near-field information. Some common SPM instruments, which are useful for coating investigations, are presented in Table 2. They are briefly discussed below, ordered according to the type of interaction measured.

2.5.1 Current, Electric Interaction

An STM instrument measures the current of tunneling electrons between a conductive tip and a conductive surface. Typically, the sensor is a fine metal wire. On the tip apex, minitips with an effective radius of curvature < 5 nm can provide atomic resolution on many ordered surfaces. With typical distances of 0.3–1 nm, a bias of 1 mV to some volts between the tip and surface renders a detectable current of 20 pA–10 nA. The distance tolerance is 0.001 nm.[16]

The distance tolerance in scanning noise microscopy (SNM) is smaller. This microscope works like an STM instrument without external bias. Therefore, it is not as well suited for rough coatings. Distance control by tunneling noise can be helpful for other applications. The low signal is acceptable if for electrochemistry investigations or for sensitive samples a small or disappearing current is desired.[106] Scanning ion conductance microscopy

Table 2 Scanning probe microscopes[107,112]

SPM principle	Acronym of technique	Name of microscope	Sensor	Measured near-field interaction	Refs.
				Current	
	STM	Scanning tunneling microscope	Conductive (metal) tip	Tunneling current (electrons)	11–13, 16, 113
	SNM	Scanning noise microscope		= STM without bias	106
	SICM	Scanning ion conductance microscope	Micropipet (in electrolyte)	Ion current	85
				Forces	
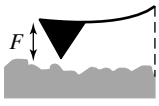	AFM	Atomic force microscope	Cantilever with tip	Van der Waals, Coulomb and other forces	
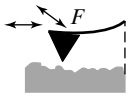	LFM	Lateral force microscope	Cantilever with tip	Forces parallel to sample surface	73, 77–80
	–	Nanoindenter	Cantilever with tip	Forces perpendicular to sample surface, depending on indentation depth	114, 115
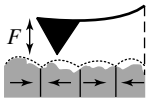	MFM	Magnetic force microscope	Cantilever with tip with a magnetic coating or Fe, Ni wire	Magnetic forces	22, 45, 54
	HFSEM	High-frequency scanning electrical force microscope	Cantilever with tip, with an integrated coplanar wave conductor	Noncontact AFM for the investigation of microwave components in the 10-GHz range	50
				Light	
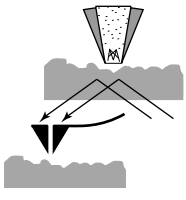	SNOM	Scanning near-field optical microscope	Glass fiber or AFM tip with a fine aperture	Optical near-field	55, 116–122
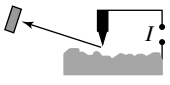	STM-LE	Inverse photoemission spectrometry	STM plus photodetector	Inverse photoemission	87, 116, 123
				Temperature	
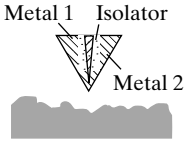	SThM (NFTM)	Scanning thermal microscope	Fine thermoelectric couple (or thermic contact between metal tip and sample)	Thermoelectric voltage	85, 124, 125
				Acoustic waves	
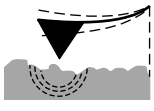	TAM	Tunneling acoustic microscope	Cantilever with tip	Acoustic waves	16, 105

(SICM) measures ion currents. The sensor consists of a glass micropipet with a tip radius of about 100 nm. The micropipet and the sample are contained in an electrolyte solution.[85,126]

The scanning capacity microscope measures the capacity of the tunneling barrier with a metal wire as a sensor; this capacity is of the order of attofarad (10^{-16} F).[34] The combination of metal wire, tunneling gap and Si acts like a MIS-diode (metal–insulator–semiconductor) and can be used for recording dopant profiles. Variation of the external bias serves for scanning capacitance spectrometry and is helpful for the investigation of, for example, charge storage devices.[33,107]

2.5.2 *Forces and Mechanical Sample Properties*

AFM and LFM, which measure vertical and lateral forces with a tip and cantilever, were described in section 2.4.

Force modulation measurements of organic coatings on mica in air led to the surprising result that mica appeared softer than monolayers of a self-organizing hydrosilicon.[92] A combined instrument with an optical microscope, an LFM instrument and a Vickers nanohardness tester was developed by the Centre Suisse d'Electronique et de Microtechnique (CSEM) in Neuchâtel, Switzerland.[114] The sample is transferred between the three stations with extremely high precision to hit the same area in all three tests. Scholl et al. use the AFM instrument itself as an indenter; they indented the sample with a special W-tip and imaged the indentation with a commercial AFM tip.[115] Bhushan et al. used three-sided diamonds as tips and cantilevers of steel as nanoindenters with normal loads of 10–150 μN, as scratch testers at 1–100 μN and for AFM imaging at 0.3 nN.[127]

For the measurement of magnetic forces only a small modification of the AFM system is necessary. AFM cantilevers with a magnetic coating or L-shaped wire of Fe, Ni or W can render information on magnetic domains and are especially interesting for the investigation of thin magnetic films.[54,107,128] Because of the relatively long range of magnetic dipole forces, the resolution of magnetic force microscopy (MFM) is restricted to 10–50 nm.[22] Jarvis and Pethica detected environmental water films with the help of hydrophobized MFM.[129]

2.5.3 *Microwaves and Light*

Coaxial metal hollow wave conductors with an aperture of 2 mm and a central wire with a radius of 6 μm can measure the tunneling current and microwave conductance of thin films or semiconductors instantaneously.[130] High-frequency scanning electrical force microscopy (HFSEM) is a noncontact AFM technique with a coplanar waveguide integrated into the cantilever for the investigation of microwave components in the 10-GHz range.[50]

In SNOM, nonradiative electromagnetic near-field interaction serves for imaging.[116,117] These interactions have a decay length shorter than the wavelength. Therefore, the diffraction limit of optical microscopy (see Equation 8, section 2.6) does not hold. Any inhomogeneity in a radiation field becomes a source of radiation, an antenna, whose dielectric and magnetic near-field properties can be characterized with the help of its far field.[116] The antenna can be used as a transmitter and receiver or only as a receiver. Investigations can be conducted in transmission or in reflection. This gives rise to a variety of different construction principles of SNOM with different acronyms. SNOM is used here for all these different types of optical near-field microscopy techniques.

Micropipets or glass fibers with tip radii < 100 nm serve as sensors. Another way of producing a small aperture is a metallic coating with a tiny hole on a glass fiber or a glass slide.[116,117,121]

In contrast to AFM, SNOM is not sensitive to electrical charges.[131] As in all near-field interactions, the optical near-field signals are strongly dependent on the distance. Typical measurement distances for SNOM in the infrared region are 0.2–5 μm,[16,119] for SNOM with visual wavelengths below 10 nm.[132] There are different concepts for distance control: glass fibers with conductive coatings can use photon tunneling; in other SNOM techniques control is effected by shearing forces on the tip[133] or by capacity.[119,134]

The resolution is limited by the aperture and measuring distance. By reducing the measuring distance to 2–5 nm the resolution could be improved from about 700 to 25 nm.[135]

Inverse photoemission spectrometry (STM-LE) derives advantage from the inverse photo effect by detecting the light emission caused by tunneling electrons.[15,87,116,123] STM-LE is comparable to cathodoluminescence imaging in SEM. The quantum yield is much lower, about 10^{-4} photons per tunneling electron, and most of the tunneling electrons only heat the sample.[106] In spite of this fact, the aim is to develop a chemical surface analysis via STM-LE.

2.5.4 *Temperature, Heat and Acoustic Waves*

The probe in scanning thermal microscopy (SThM) is a fine thermoelectric couple made of a W wire coated with an isolator and Ni. Tip radii of about 15 nm have been achieved.[85] For scanning thermal microscopes with Pt/PtRh thermoelectric couples a lateral resolution of from micrometers[124] to 35 nm has been reported.[85] In addition to local temperature measurements with a precision of 10^{-4} K, SThM can serve for chemical

analysis if it is combined with a photothermal absorption spectrometer. The sample is irradiated with a tunable laser. Absorption and thus heating of the sample are wavelength dependent and specific for elements.[85]

A calorimeter with nanometer precision must itself have a small heat capacity. This demand is fulfilled by Si AFM tips with an Al coating which bend like a bimetallic strip owing to temperature changes. They can indicate 1 nJ.[125]

In tunneling acoustic microscopy (TAM) the probing tip is excited with ultrasonic vibrations close to the resonance frequency of sample, sample holder and piezoelectric transducer. Depending on the distance, another transducer indicates the transmittance to the sample.[16] Theoretically, a local resolution of 20 nm should be possible; in practice commercial systems reach 0.5 μm.[136] An instrument with two springs, one as a transmitter and the other as a receiver, detects pores, flaws or other inhomogeneities in the sample volumes.[105]

2.6 Comparison of Microscopy Methods

Because of the diffraction limit, the spatial resolution of conventional microscopes is dependent essentially on the wavelength of the type of radiation used (see Table 3). According to a relation which was initially established by Ernst Abbe for light microscopy, the spatial resolution limit d_c of conventional microscopes is dependent on the aperture α, the refractive index n of the medium and a constant k describing the brightness of the object[137–140] (Equation 8):

$$d_c = k\lambda(n \sin \alpha)^{-1} \quad (8)$$

The essential factor is the wavelength λ of the type of radiation used. Equation (8) can be roughly estimated by $d_c = 0.5\lambda$.[137,138] Lens aberrations can prevent this theoretical limit from being reached.

Optical polarization microscopes can image height distances of around 0.1 nm and measure roughness[143] (see section 2.7), with a rather poor lateral resolution of 1 μm. An optical microscope with an outstanding capability of imaging the topography of coatings is the Nomarsky or differential interference contrast microscope.[144,145] Its depth of focus is smaller than that of electron microscopes, which makes it more sensitive to surface features. Nomarsky images can also be transferred to digital data.

Table 3 Best achievable spatial resolution of different types of microscopes[136–139,141,142]

Method	Lateral resolution (nm)
Conventional optical microscope	200
X-ray microscope	30
Ion microscope	40
Scanning electron microscope	3–6
Transmission electron microscope	0.2
Scanning acoustic microscope	600
Optical near-field microscope	20
Scanning probe microscope	0.2

The lateral resolution of confocal laser scanning microscopy is better than that of conventional light microscopy by a factor of 1.4. Further advantages of this method are the reduction of stray light from neighboring structures and instantaneous three-dimensional data sets.[146]

In an electron microscope, a beam of accelerated electrons is used instead of electromagnetic radiation, and electromagnetic lenses take the role of the glass lenses in an optical microscope. The sample is either scanned (SEM) or directly imaged (TEM) by electrons on a fluorescent screen.

In SEM, the image can be reconstructed from secondary electrons, backscattered electrons, cathodoluminescence and some other interactions between the incident beam and sample.[139] Its resolution is mainly limited by the diameter of the incident electron beam; optimum focusing results in $d_c = 3–6$ nm[139] and $d_c \geq 10$ nm is typically achieved.[147]

For direct imaging as in TEM, Equation (8) holds; in this case λ is the De Broglie wavelength λ_c of the electrons (Equation 9):

$$\lambda_c = h(m_e v)^{-1} \quad (9)$$

where $h = 6.626 \times 10^{-34}$ J s is Planck's constant, m_e = electron mass and v = electron velocity;[139] v and thus λ depend on the acceleration bias, which is in the range of some tens of kilovolts to megavolts; relativistic effects are to be considered above 50 kV. Acceleration with 50 kV results in $\lambda_e = 0.0053$ nm.[148]

Spherical and chromatic aberration lead to a worse practical resolution. In TEM, the chromatic aberration also depends on the sample thickness t, $d_c \approx 0.1t$.[137,138] An aperiodic resolution limit of about $d_c = 0.2$ nm can be achieved in TEM with a 10 nm thick sample.[147] TEM resolution can be sufficient for imaging large atomic distances in some ordered materials. With an acceleration voltage of 200 kV, even a lattice-fringe image resolution of 0.1 nm has been reached.[148]

One large advantage of electron microscopes is that material properties contribute to the contrast as electron scattering and transmission depend on the element mass. Additionally, many electron microscopes are combined with elemental microanalysis via energy-dispersive X-ray spectrometry (EDX), wavelength-dispersive X-ray spectrometry (WDX) or crystal structure analysis via diffraction.[137,138]

Some modern instruments no longer require high vacuum in the sample chamber. With a diaphragm between the microscope column and the specimen, wet materials and materials with volatile components can be imaged.[139,149] The gas pressure between 10 and 1000 Pa in the sample chamber of such an environmental scanning electron microscopy (ESEM) instrument can also provide charge compensation during imaging of nonconducting samples.[150]

In conventional electron microscopy, however, a conductive coating, generally C or Au, must be applied. This decoration masks sample features smaller than the grain size. A further restriction is given by the electron radiation damage in the irradiated zone of the sample. Therefore, TEM investigations on, for example, the crystal phases in nonconducting samples must be conducted fast and with low bias.

Scanning X-ray microscopes employ soft X-rays with a wavelength between 2 and 5 nm and achieve a spatial resolution of about 30 nm.[141] Ion microprobe mass analysis (IMMA) images element distributions with a lateral resolution of micrometers.[151,152]

Field ion microscopes achieve a magnification of the order of 10^6 and can image individual atoms. However, they are restricted to atoms on fine tips with typical apex radii of 10 nm[153] and cannot display flat coated samples.

In addition to the spatial resolution and information depth, other criteria for the choice of a certain type of microscope are the following:

- the contrast resolution (i.e. height resolution of topography images);
- the restrictions on sample size and sample properties;
- the expense of sample preparation (cleaning, decoration);
- the environment needed (high or medium vacuum, air, liquids);
- the time to sample one image;
- the availability and price of the instrument.

In any case, a comparison of images from different microscope types, starting with a low magnification and resolution, is recommended for unknown surfaces. Combined instruments utilizing AFM in an optical microscope or SEM and SPM in the same vacuum chamber are commercially available and eliminate the problems of sample transfer.

2.7 Determination of Roughness

A quantitative determination of roughness is part of a complete surface characterization of substrates and coatings. Roughness is an important parameter for topics such as surface energy, adhesion, optical quality, tribology and wear. According to Bennett,[154] surface formations can be classified as follows depending on their lateral dimensions:

- microroughness or roughness with wavelengths up to about 1 mm;
- waviness or mid-frequency roughness with wavelengths from millimeters to centimeters;
- surface shape or optical shape with wavelengths from centimeters up to the dimensions of the object, especially deviations from the ideal shape.

A practical example of waviness is the ''orange peel'' of chemically polished Si wafers. Microroughness is the parameter which describes the quality of substrates and coatings. Deviating from these definitions, microscopists often use the term waviness for formations with lateral dimensions of the order of the image size, i.e. from $<1\,\mu m$ to some hundreds of micrometers.

Roughness is given, for example, as average roughness R_a (arithmetic average of all z values) (Equation 10):

$$R_a = \frac{1}{N^2}\sum_{i=1}^{N}|z_i| \qquad \left(\sum_{i=1}^{N} z_i = 0 \right. \tag{10}$$

where N = number of pixels per line and lines per image and z_i = measured height at point i, or as root mean square (RMS) roughness (geometric average of all z values, also called R_q or σ) (Equation 11):[144]

$$\mathrm{RMS} = \sqrt{\frac{1}{N^2}\sum_{i=1}^{N} z_i^2} \tag{11}$$

The RMS roughness can be calculated from total integrated scattering (TIS) data according to Equation (12):

$$\mathrm{RMS} = \frac{\lambda}{4\pi}\sqrt{\frac{S}{R_0}} \tag{12}$$

where λ = wavelength of light, S = ratio of total stray light intensity to incident intensity and R_0 = reflection.[144] For ideally isotropic surfaces with a Gaussian distribution of heights, the RMS values of line scans and area scans are identical.[155]

Methods for determining roughness are described in DIN 4760 (RMS), ISO 4278/1 and ANSI/ASME B46.1[156] and in a report of the ASTM STM/AFM Subcommission E-42-141 (''standard practices for analyzing and reporting surface roughness as measured by STM/AFM'', http://www.atm.org). The ASTM Subcommission F-1.19 is working on a draft standard for calculating power spectral density (PSD) from surface profiles.[157]

There are mainly two types of methods for the measurement of roughness: diffraction of electromagnetic

radiation or particle beams and optical or mechanical profilers. An overview of the methods and some recommendations according suitable methods for different surfaces were given by Bennett and Mattson[144,154] and Rothe et al.[156] Guenther et al.[158] and Bhushan et al.[155,159,160] compared optical and mechanical profilometry with different types of scanning force microscopes.

Roughness values can only be compared if the bandwidth range and the height sensitivity of the instrument used are known.[144,155,161] The lower bandwidth limit is defined by the best achievable resolution or the number of measurement points per scan and the upper limit by the measured length or area on the sample surface.

Optical methods such as scattering of light [TIS, angle resolved scattering (ARS)], small-angle diffraction of X-rays (SAXS), or small-angle scattering of neutrons (SANS) and grazing incidence X-ray reflection (GIXR) give average roughness values for spots in the order of square millimeters.

The upper bandwidth limit in SPM can range from about 100 μm down to nanometers. For analysis with GIXR or optical interferometry, an assumption regarding the height distribution is required. In many cases, a Gaussian distribution works well.[162] In contrast, the real height distribution is measured with mechanical profilers or by SPM.

Anisotropies of diffraction data give indirect hints on the type of the defects (e.g. scratches, holes, dust) that contribute to the roughness of optical surfaces. True defect profiles cannot be deduced from the distribution of diffraction intensities.[163] Two-dimensional profilers and scanning probe microscopes simultaneously provide data on roughness and direct images of the profiles in real space.

An evaluation of the information from different roughness measurement methods was given by Rothe et al.[156] and the capabilities of AFM and X-ray scattering have been compared.[164] As is shown in Figure 7, SPM has the largest bandwidth. Owing to its capability to resolve even single atoms (at least on some favorable flat surfaces), it can determine differences between extremely smooth samples. It can detect wavelengths from the resolution limit (atomic distances to at least some ten of nanometers) up to the maximum x–y range of the scanners (about 100 μm) and height differences from 0.01 nm to some micrometers.

The bandwidth of the measurement must be chosen with respect to the practical application and the physical parameters implicated. The important length scale can be the area of contact between the recording head and magnetic storage media, the wavelength of visible light for optical coatings or atomic distances for tribology of extremely smooth surfaces.[155] Frequently, the measured

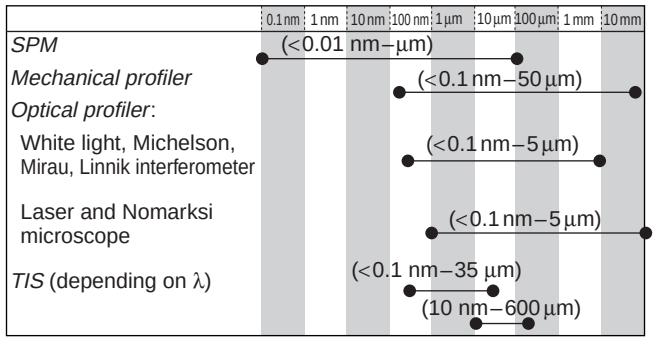

Figure 7 Bandwidths of different roughness measurement methods. The detectable roughness (RMS) is given in parentheses [according to Bennett and Mattson[144]].

roughness values increase with increasing upper bandwidth limit up to a certain saturation value. In such cases, the upper bandwidth limit may be set to the point, where no longer wavelengths are observed.[155]

For the comparison of bandwidth ranges, a surface profile $z(x)$ can be described by sine waves as in Equation (13):

$$z(x) = Z \sin\left(\frac{2\pi x}{\lambda}\right. \tag{13}$$

where Z = amplitude and λ = wavelength. The working range of an instrument for roughness analysis is defined by straight lines in a $\log Z = f(\log \lambda)$ plot (see Figure 8):[165]

- minimum and maximum detectable wavelength;
- minimum and maximum detectable height;
- maximum detectable gradient; for scanning probe microscopes, the detectable gradient is dependent on the tip angle 2α, limited by the straight line according to Equation (14):

$$\log Z = \log\left(\frac{90^\circ - \alpha}{2\pi}\right. + \log \lambda \tag{14}$$

- minimum detectable radius of curvature $R_{\min} = R + r_w$, depending on the tip radius R and the interaction distance r_w, limited by the straight lines according to Equation (15):

$$\log Z = -\log(4\pi^2 R_{\min}) + 2 \log \lambda \tag{15}$$

The lower limit for the detectable wavelengths is defined by the resolution limit or the distance between to measurement points. AFM has a lower wavelength limit in the nanometer range; for mechanical profilers, this limit is between 0.1 and 1 μm. The upper wavelength limit, which is defined by the scan size, is about 100 μm for SPM and some millimeters for profilers.

Because of different bandwidth limitations, roughness values measured on the same samples with different

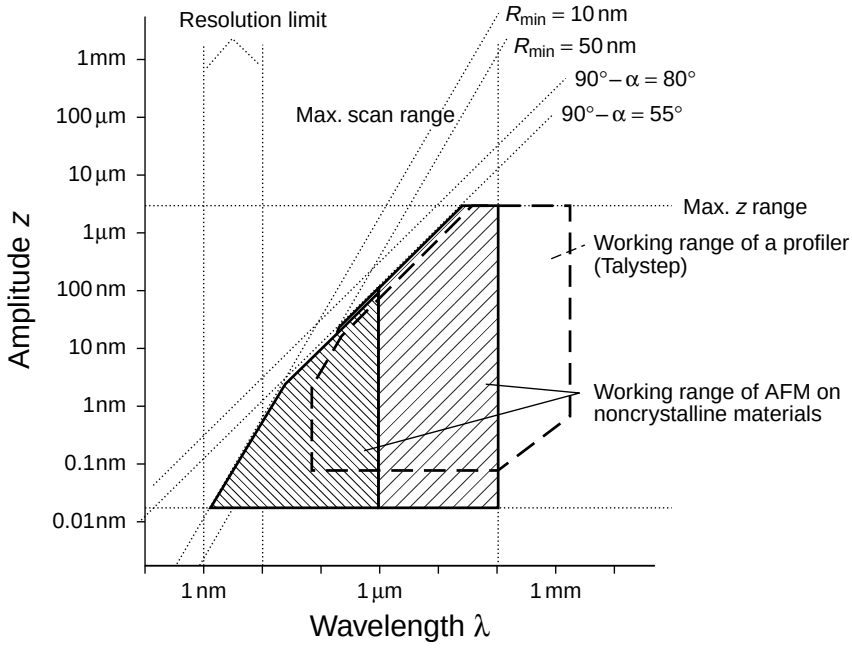

Figure 8 Working ranges of profiler and AFM for roughness determination.[165]

instruments can deviate. Apparent contradictions can be explained by a two-dimensional Fourier analysis of the profiles. By plotting the PSD against frequency, the fractal dimension can be calculated. With PSD curves from measurements with different bandwidths, the surface roughness can be analyzed over seven decades of wavenumbers.[162]

3 SAMPLE PREPARATION

Similarly to all other high-resolution surface analyses, the sample preparation for SPM must be carried out with care to avoid unwanted alterations of the surface of interest. SPM samples must be sufficiently stable mechanically and chemically. Mechanically unstable specimens, e.g. areogels, are vibrated by the tip in contact mode AFM. There are also less interactive measuring modes such as noncontact AFM which can image such delicate materials[166] and soft coatings, e.g. not fully consolidated gels.

3.1 Cleaning

The best way to avoid unwanted alterations of unknown surfaces is to apply no cleaning at all. If there is too much visible dust or a test scan results in unacceptable disturbances, blowing the dirt off with a bellows can be sufficient in many cases. Pressurized air can contain unctuous contaminants from the compressor. Large solid dust particles can interfere with the cantilever beam and soft dirt affects the image quality by sticking to the tip or smearing across the sample.

Many chemical, mechanical and physical standard procedures for the cleaning of optical surfaces or of substrates prior to coating can also be used for SPM samples. An overview of such methods was given by Dunken.[167] Most oxidic coatings withstand a procedure with the following steps:

1. brushing with a nylon brush in hot water with a mild surfactant;
2. rinsing with deionized water;
3. ultrasonic bath in
   (a) acetone
   (b) ethanol
   (c) methanol;
4. drying in air, with a bellows or in dry nitrogen.

Steps 3 and 4 are repeated until the surface dries without visible stains. Some contaminants require wiping with soft tissue between the ultrasonic treatments. One way of removing hydrocarbons without mechanical contact is ultraviolet irradiation in a flow of oxygen. Some carbon-containing remnants, however, cannot be removed by this treatment.[162]

OPTICLEAN™ (producer, Caliope; sold in Europe by Caliope Ltd, 3 Marfleet Close, Great Shelford, Cambridge CB25LA, UK, or by LOT, Im Tiefen See 58, 64293

Darmstadt, Germany) is a solution of polyurethane (PU) in a mixture of ethyl acetate, alcohol and water, which can be applied to plane surfaces like a lacquer. This polymer film can also protect from dust during sample storage, but not from humidity. Before measurements, the film is torn off and takes contamination with it.

New surfaces which had no contact to air can be obtained by cleaving under inert liquids,[91,168] cleaving in vacuum or ion sputtering.[169] Argon ions (2–3 keV)[170] or krypton ions (5 keV)[171,172] can produce surfaces which are less contaminated than after cleaning with ultrasound or with polymer film.

3.2 Cross-sections of Coatings

To gain information on the nanostructure of the bulk of coatings by SPM, fracturing, ultramicrotomy, grinding and polishing, etching, replica techniques and ion beam sputtering can be employed.

Many coatings can be investigated after simple fracturing of notched bars.[23,173–177] On notched glass bars, a fracture mirror of some square millimeters can be obtained, which is characteristic for the glass volume and oxidic vitreous or crystalline coatings. Thus, coatings with a thickness of more than some tens of nanometers can be investigated. A deep sawn notch is to be preferred to notching with a diamond, which produces more fine dust particles on the fracture surface (see Figure 10d).

Cross-sections can be produced by sandwiching two coated surfaces face to face or a coated sample to a dummy. For TEM investigations, Becker and Bange marked the surface of interest with an additional Cr coating.[178] Wilder et al. imaged cross-sections of metal oxide semiconductor devices with AFM after sandwiching, sawing, lapping, polishing and selective etching.[179]

Duparré et al. applied preparation techniques which have been developed for TEM for cross-sections of optical coatings and compared AFM and TEM images;[180,181] 150 nm thick single and multilayers of MgF_2 and LaF_3 on Si were embedded in Araldite and fractured or ground, ion polished and ion beam etched. Fracturing resulted in too high steps between the Araldite and the sample, but in higher AFM resolution on single layers. The fractured triple layer was too rough for AFM imaging; it could be scanned only after ion polishing and ion etching. In both cases, the columnar structure of the fluorides was better made visible with TEM. Ion beam preparation of Mo–Ti and TiN hard coatings allowed similar results for imaging with TEM and AFM.[182]

Ultramicrotomy[178,183–185] and replica[186] techniques have also been developed for imaging in electron microscopes and can also be used for SPM. Ultramicrotomes cut extremely thin slices with a diamond blade. It has been shown that this routine method for medical or biological materials can also be applied for oxidic coatings.[178]

For the replica technique, a Pt–C and/or C film is evaporated on the sample surface (e.g. a fracture surface) and subsequently removed in water or acidic or alkaline solutions. Then the replica film is imaged instead of the sample itself. In contrast to replica coatings for TEM, which are evaporated with oblique incidence to render high contrast, replica coatings for SPM are evaporated vertically to avoid artifacts.[187] Capillary forces between the SPM tip and the replica film, however, have been reported to alter the film topography during scanning.[187,188] Soft samples can be shock frozen and fractured before replica preparation.[187,188]

3.3 Fixation, Coated Fibers and Small Particles

With ''stand-alone'' scanning probe microscopes, many bulk samples can be investigated without cutting them into pieces; their sample holder is large enough for, for example, semiconductor wafers. In such microscopes, the tip is scanned instead of the sample, because of its smaller mass. In most commercial SPM instruments, however, the piezo drive is arranged as in Figure 2 and the sample holder is designed for samples with an area of around 1 cm^2 and a thickness < 0.5 cm. Larger objects have to be cut.

With sufficient sample mass and a piezo design as in Figure 2, the specimen can just be laid on the sample stage. Light samples, which could be moved by the tip, need fixation. Flat samples can be fixed with double-sided adhesive tape to magnetic plates. For UHV or liquid cell applications, a glue must be chosen which does not contaminate the environment. In such cases, tailored clamps have proved to be reliable.[189–191] Nonflat samples can be embedded in plasticine with the surface of interest parallel to the scanning plane.

For coated ''small'' particles such as fibers, grains or powder, adequate fixation must be tested. Double-sided tape does not always immobilize sufficiently and carries the risk of irreparable contamination of the tip. Indium is a very ductile metal. Pressing fibers into In foils with a stamp press has been shown to fix the fibers well and furthermore provides antistatic effects.[8,192] Curtin Carter et al. recommend a carbon suspension in 2-propanol as a glue for fibers. This suspension is also used for electron microscopy and reduces charging.[193] Other workers prepared fibers by fixing them to the sample platelet at two points at the ends.[194,195] Embedding in resin as for electron microscopy and cutting and polishing or ultramicrotomy can provide information on fiber cross-sections.[93,196–198] For the investigation of fully consolidated sol–gel coatings on glass fibers, however, a further preparation step was necessary: the coating thickness could be determined after mild etching with HF.[194]

Solid grains with diameters between 20 and 70 μm have been fixed with gelatine;[199] for small flat particles, drying a water suspension on microscope slides works well.[8] Garcia and Edwards developed special sample holders for scanning individual particles with diameters between 70 and 500 μm.[200] They drew fine glass pipets, which could be fixed on SPM sample holders and connected to a water jet pump.

3.4 Preparation for Scanning Tunneling Microscopy

STM is suitable only for conductive surfaces. For the investigation of nonconductive coatings or of cross-sections of conductive coatings on nonconductive substrates, AFM is recommended. If mainly surface features larger than 10 nm are to be expected, nonconductive samples can be coated with gold as for electron microscopy.[201] Smaller details can be imaged only after special evaporation methods with cooled samples[202,203] or on replicas.

4 ARTIFACTS AND CALIBRATION

All deviations of the measured profile from the true sample topography are to be considered as artifacts. Artifacts in SPM can be caused by

- the data processing;
- the measuring system;
- the tip geometry;
- the behavior of the sample during scanning. (See references 9, 42, 62, 204, 205.)

In many cases, combinations of these effects are observed. One impressive example is the SPM images of DNA strands on mica. The true diameter of DNA strands is 2.5 nm. On SPM images, an apparent height of 1 nm is measured. The apparent width is 15–20 nm in contact mode AFM under water, 5 nm in Tapping™ mode AFM under water (see section 2.4.5) and 3.5 nm in STM in humid air.[22] The error in the height is ascribed to tip-induced deformation of the molecule and the error in the lateral dimensions is ascribed to the tip geometry and the interaction distance.

4.1 Data Processing and Measurement System

The most important issue in data processing is the applied filters.[205,206] Not all types of representations (e.g. quasi-three-dimensional, top view, with color encoding height or in illumination mode) give similar evidence.[207] Therefore, it is recommended to judge unknown surfaces from many different representation modes.

Images can be affected by external vibrations, feedback in the measuring system or wrong gains. For each scan, a new compromise of optimum scanning speed and gain has to be found. With low speeds, low gains can be realized. Because of thermal drift of the actuator, the speed should not be too low. Reasonable scanning frequencies are in the range ca. 20–1 lines s^{-1}. For large scans (100 μm^2), low scanning frequencies are necessary; small scans can be executed faster.

The piezo actuators can affect images in the form of hysteresis, nonlinearity, thermal drift and creep.[9,42] Single tube scanners do not scan an ideally flat surface, but move along a spherical surface. This nonlinearity can be removed with mathematical algorithms. For extremely flat samples, however, the angle between the sample surface and scanning plane is too small to be recalculated properly.[208] Hysteresis artifacts are avoided by imaging only traces in one direction.[42] Thermal drift decreases with running time and in well-conditioned air, but cannot be avoided totally in ambient conditions.

Creep means that the piezo material does not react instantaneously to voltage pulses and does not stop instantaneously after the voltage pulse. Therefore, dimensions can be wrong in the slow scanning direction (generally y) and overswinging can occur at steep steps. On three-dimensional representations, overswings give the illusion of additional peaks or troughs. On top-view images, overswings can enhance the contrast.

Piezoelectric materials age, especially when they are not used. At temperatures above 150 °C, they depolarize. They should be recalibrated from time to time and must be recalibrated after the baking of vacuum instruments[209] (see section 4.4).

4.2 Tip Geometry

Tip geometry is the decisive detail of SPM. Because of the small range of near-field interactions, it is easier to gain high resolution on flat samples than to image rough samples truly. Artifacts result if the tip radius R is larger than the radius of curvature r_S of surface features,[166] but even at $r_S < 2R$, tip convolution is essential.[210,211] Diameters of single elevated objects are imaged too large and diameters of holes are imaged too small. In most cases, the height of single elevated objects is imaged correctly and the depth of narrow trenches and pores is underestimated.[212]

SPM sensors can have fins or blunted tips if the holes in the etching masks used for tip production are not perfect squares.[62] Fins result in an apparently preferred orientation (see Figure 9a–d). Measuring a sample with different tips or turning the sample can help to decide whether a preferred orientation is true or apparent.[9]

At peaks which are narrower than the tip, the roles of object and probe are exchanged, and such peaks can

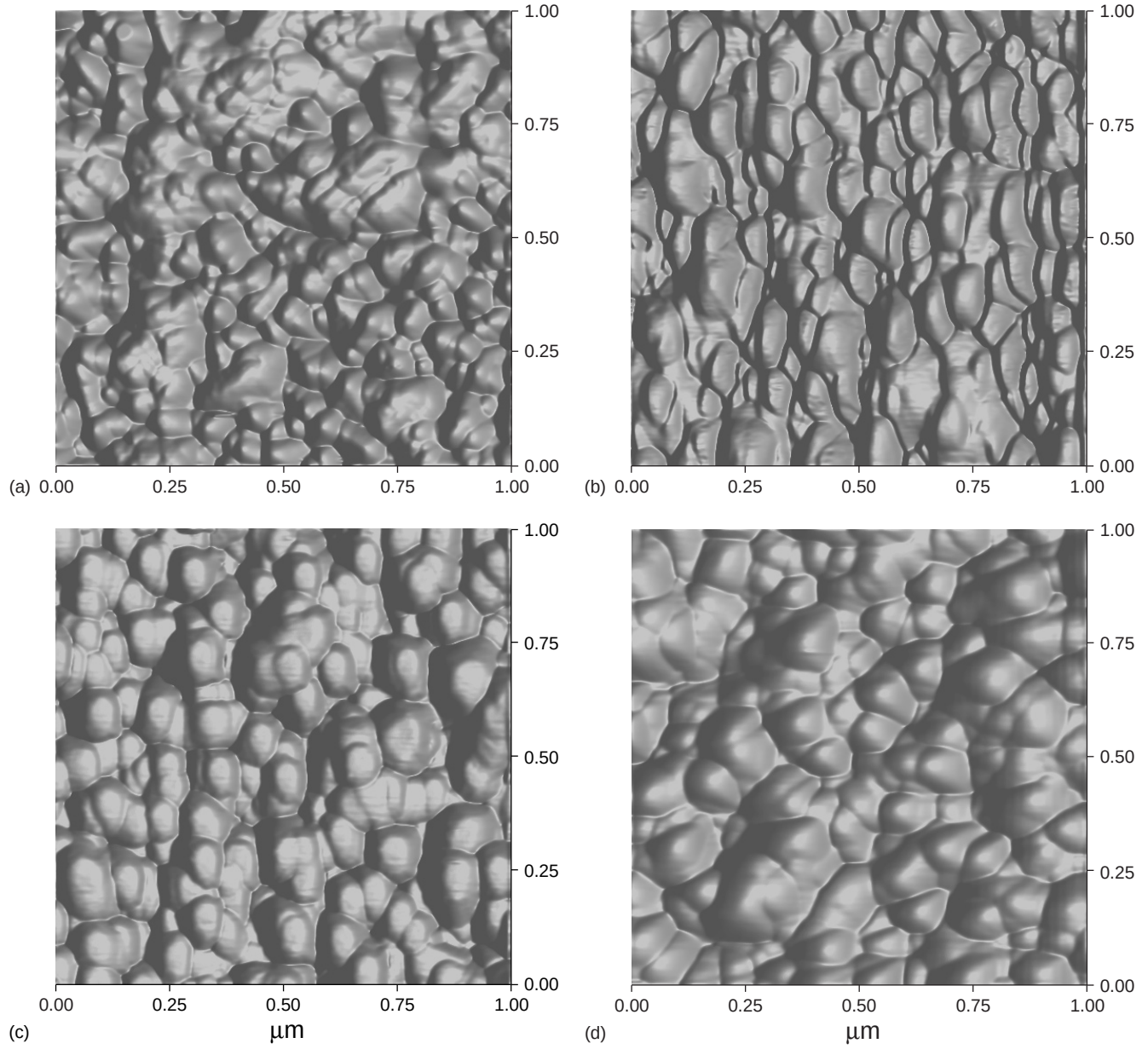

Figure 9 Images of a crystalline SnO_2 coating with Si_3N_4 tips of different quality: (a) good image with a new tip; (b) the *y* dimensions are overestimated because of a fin in the *y* direction; (c) overestimation of *y* dimensions because of a used, truncated tip; (d) imaging with a used, blunt tip: no preferred orientation, but few details. $\Delta z = 100$ for all four images.

serve for imaging the tip. Images of steps which are steeper than the tip are convolutions of true surface profile and geometry of the contact area of the tip (see Figure 10a and b). Steep steps are imaged with the half apex angle α of the tip (Si_3N_4 pyramids: $2\alpha = 70.5°$). Cantilevers are installed in the scanning head with a flat angle of 10–20° with respect to the sample plane. Thus, e.g. cubes or spheres seem to be asymmetric.[204]

Examples of multiple tips are given in Figure 10(c). The contact to the sample changes between the individual tips. Trenches between the crystallites of the SnO_2 coating are imaged twice; all of the individual glass dust particles on the fracture surface of Figure 10(d) seem to have the same shape (small particles are oblong, large particles seem to be double). In Figure 10(e), a gold coating which is used for SEM is scanned by AFM with a multiple tip: the metal grains resemble berries. So-called 'ghosts' appear if the contact changes from one tip flank to the other in narrow holes.[213]

Schwarz et al. compared tips from different suppliers and stated that Si_3N_4 pyramids with a relatively low aspect ratio are to be preferred for unknown samples.[62] They

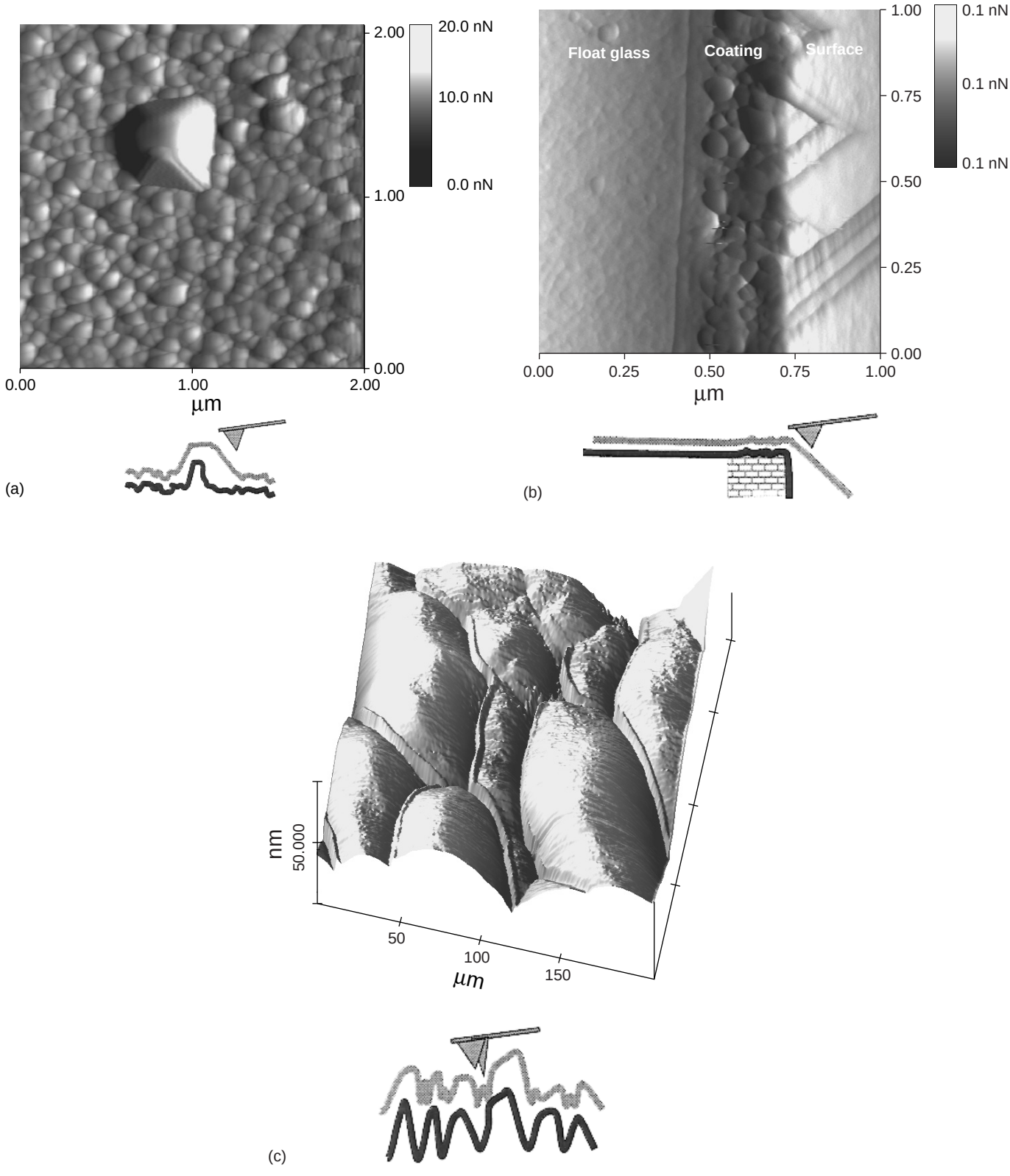

Figure 10 Imaging the tip: (a) tin oxide coating on glass with a defect that is narrower than the tip; (b) fracture surface of SnO_2-coated glass: the coating surface is imaged by the flat side of the tip; (c) a multiple tip images the trenches between the SnO_2 crystallites twice [same sample as (a)]; (d) fracture dust particles on the glass surface are also imaged twice by a multiple tip; (e) 'berry'-like appearance of the grains on a gold coating for SEM, due to a multiple tip.

have the advantage of a well-defined geometry with small deviations. They are harder than Si tips and perform well on rough samples, with high scanning frequencies and large scanning areas. Etched Si tips are sharper, but exhibit larger variations, they are more sensitive to damage and tend to develop multiple tips. This makes it harder to recognize tip defects.

4.3 Sample Response

It can be difficult to scan extremely rough samples. If there are elevations higher than the tip (Si_3N_4: about 3 μm), the cantilever bar touches the sample and the image shows a flat plane. Contamination particles or parts of soft samples can stick to the tip and let it appear blunt; abrasive surfaces also affect the tip quality during scanning. Strong local variations of interaction energy can convolute with real topography. On polymers and other organic substances holes can appear, because the tip sticks to certain spots on the surface or because of local deformations under too high normal forces.[204] Investigation of organic substances showed that tip geometry plays a decisive role. With asymmetric tips, the probability of destructive shearing loads is higher.[214]

Depending on the affinity to water, many materials give rise to a marked onset overswing (see Figure 11) during scanning in a humid environment. This overswing is due to the acceleration of the water layer on the sample at the start of each scan line. Products of reactions with the environment and contamination can also affect the images.

The following nine rules are recommended for avoiding artifacts as far as possible:[204,215]

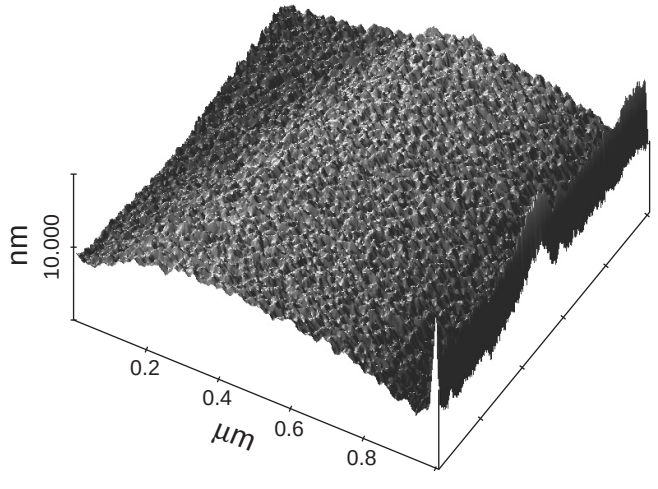

Figure 11 Onset overswing due to the acceleration of the water film on a sample scanned in humid air. At the start of each scan line (scanning direction: $-x$), the tip dips into the adsorbed water. This leads to the overswing at the right-hand side of the image [sample: glass fiber[23]].

- know the artifacts to which your system is prone;
- use different tips;
- take images with different sample rotation;
- rotate the scanning direction;
- change the scanning velocity;
- change the scanning area;
- compare with other imaging techniques;
- represent images with as few filters as possible;
- keep in mind that there are as yet no commonly accepted standards for product claims.

4.4 Calibration

For the use of SPM as a reliable metrological technique for coatings, piezos and tips must be calibrated taking into account artifacts and systematic errors. Unknown variables for the deconvolution of SPM images are:

- sample geometry;
- tip geometry;
- contact area;
- local variations of interaction;
- piezo behavior.

Theoretical deconvolution calculations exist for STM and AFM images with atomic resolution, even for multiple tips,[28,216,217] but there is no standard routine which is valid for all surfaces and tips.[213]

A rather laborious way of eliminating one of the unknown variables is to scan a used tip with a new one.[218] Michely et al. combined STM with FIM to provide in situ analysis of the tip.[219]

Usually, standard samples with known surface features in the nanometers to micrometers range are imaged. One can exploit the effect of reversed imaging: objects which are smaller than the tip give information on the tip geometry.[220,221] Calibrations must be performed for each scan area and scanning velocity.

For small scans, the x and y axes can be calibrated with the help of atomically resolved images of known crystal lattices such as highly oriented pyrolytic graphite (HOPG), Si or mica. The parameters for the z height of atom images, however, are not yet fully understood. For $z > 10$ nm, commercial profiler standards are available. Standards for electron microscopy, e.g. latex spheres, have dimensions of some tens of nanometers. Some authors proposed polystyrene and latex spheres (diameters > 100 nm) or gold colloids (diameters 10–40 nm) on mica,[213,222,223] others developed techniques to grow Cu peaks[221] or etch InP columns[224] or used well-known coatings such as SnO_2 or Cr.[173,189,225,226] Structured wafers from chip fabrication or diffraction gratings serve for scan areas $xy > 1\ \mu m^2$.[22] For topographic features of the order of nanometers, test methods are under development. For full characterization

of the tip geometry, not only the dimensions of standard topographies are important, but also their geometrical form, e.g. angles of inclination and radii of curvature.

Standard samples with steep parallel ripples can be obtained from $SrTiO_3$ crystals. On the (305) surface, steps of about 10 nm height are observed. Commercial standards are Nanoedge by IBM with parallel ripples[227] or Si grid with pyramid holes, which are produced with similar processes to the tips (by Digital Instruments, see Figure 12a and b). The tip radius itself cannot be determined exactly, as the apex radii of the etch holes are of the same order.

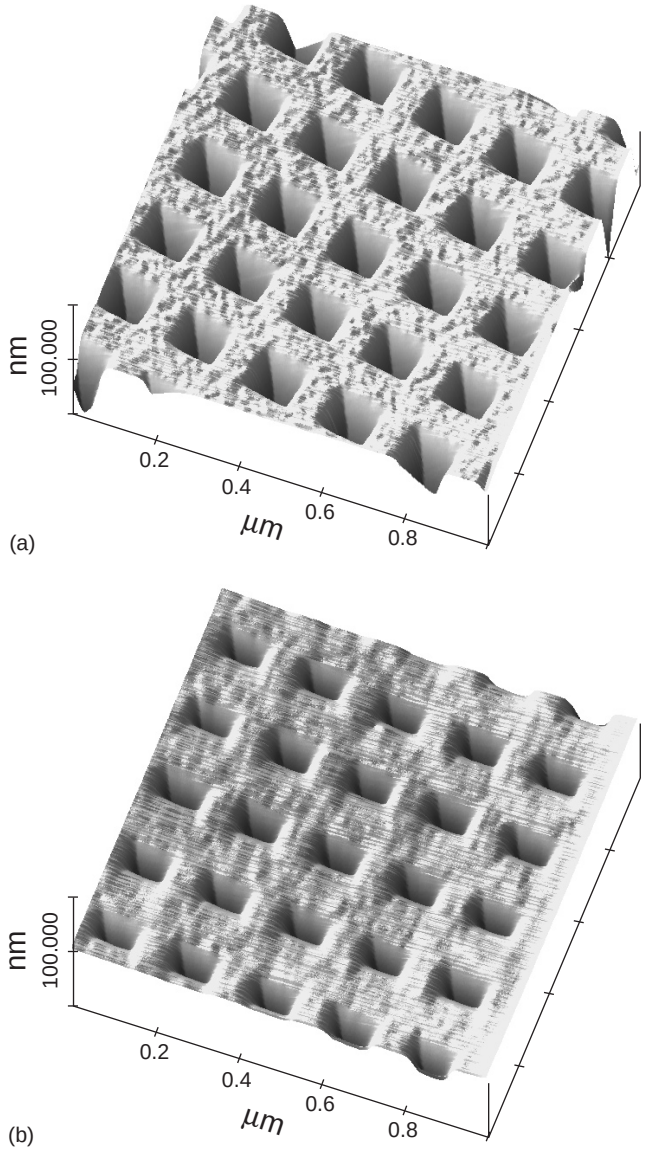

Figure 12 Si calibration grid by Digital Instruments: AFM images with (a) a sharp and (b) a blunt Si_3N_3 tip. The grid consists of pyramid-shaped etch holes with a width and a distance of 100 nm and a depth of 70 nm.

In some cases, SPM is combined with SEM, so a direct SEM imaging of the SPM tip can be achieved.[228] Transfer of the tip to an external microscope bears the risk of contamination or destruction. Periodical tip tests on well-known surfaces are recommended; tests before and after the imaging on unknown samples can give hints on wear-induced tip blunting.[9,227]

5 APPLICATIONS

In the following sections some practical examples are given of coating questions such as nucleation, defects, roughness, thickness, microstructure, crystal phases, homogeneity, interfaces, porosity, adhesion, hardness, scratch resistance, tribology, wear and corrosion, which can be solved with the help of SPM.

5.1 Nucleation, Oxidation

For the investigation of film nucleation, flat substrates with well-defined surfaces must be chosen. As there is no possibility of chemical identification, substrate irregularities, contaminants and nuclei must be identified by careful investigations of the substrate before coating. Early stages of the growth of metal films on glass were investigated by SPM by Duyne et al.,[229] Roark and Rowlen,[230] Smith et al.[231] and Creuzet.[232]

5.1.1 Silver on Glass: Silver Island Films

The first steps of sputtering Ag on glass display Vollmer–Weber growth (three-dimensional growth of islands) and the islands are much larger than during evaporation.[232] A thin physical vapor deposition (PVD) silver island film (AgIF) consisting of flat, spheroidal islands with dimensions below 100 nm and an equivalent layer thickness of some nanometers is used as a substrate for surface-enhanced Raman spectroscopy (SERS). SERS is an extremely sensitive method for identifying adsorbates and provides information on their binding, orientation and dynamics[229] with an information depth of about 10 nm.[233] SERS is restricted to surfaces with a well-defined roughness. The analytical relationship between the roughness and the surface enhancement factor is not yet fully understood. With AFM, the roughness of AgIFs could be characterized over the whole range of wavelengths. Histograms of different AgIFs showed a Gaussian distribution of the island radius, the ratio of radius to height decreasing with increasing radius. The direct determination of this ratio facilitated the interpretation of the measured SERS enhancement factors.[229] A tip convolution was not considered, although with particle radii of 10–100 nm and tip radii of 20–50 nm tip artifacts

cannot be ruled out. Different optical properties of AgIFs on microscope slides with and without a Formvar™ coating (polyvinyl Formar solution, trade mark by Monsanto) were not caused by different particle radii, but only by smaller heights of the islands on the uncoated glass.[229]

5.1.2 Gold

The dimensions and the shape of the first nucleated islands are responsible for the film roughness in the subsequent growth steps, when the coverage becomes complete. This was shown by Vancea et al. with the help of STM images of PVD Au films on Corning 7039 glass, Si(100) and freshly cleaved NaCl.[234] These films completely cover the substrates if the thickness exceeds 8–12 nm. On the cleaved crystals, the films were much smoother than on glass and Si, with height differences below 0.1 nm. On HOPG, even 80-nm thick gold layers were still extremely smooth, displaying regular hillocks with the periodicity of the C atoms of the substrate (0.29 nm). On Corning 7039, the hillocks had diameters of 20–40 nm and height differences of 6–17 nm. On Si(100), they were even larger. The distinctive island growth on Si was explained by gold silicide nuclei on which pure gold can grow. The role of the number of such nuclei where tight bonds to Au adatoms can be established was revealed by a Corning 7039 substrate, on to which <1.5-nm Cr was evaporated before the Au coating. Cr offered a higher number of stable nucleation centers and thus led to a smooth Au film with height differences of 2–3 nm.

5.1.3 Silicon: Amorphous and Crystalline Islands

The first phases of layer growth of microcrystalline Si (μc-Si) from a plasma-enhanced CVD process are strongly dependent on the substrate type. Smith et al. investigated the growth of μc-Si for thin-film transistors on crystalline Si and on polished glass with AFM.[231] After a deposition time of 15 s, the first Si islands on glass with an equivalent film thickness of 1.5 nm appeared less uniform and flatter than on Si. After 1 min, at an equivalent film thickness of 5 nm, the film on Si showed complete coverage. After 5 min, at 30 nm, the differences between the two substrates disappeared. Reflection high-energy electron diffraction (RHEED) measurements confirmed that at an equivalent film thickness of 4–5 nm, amorphous a-Si transforms to randomly oriented μc-Si. The island shape suggested three-dimensional growth on Si and a more two-dimensional character on glass. The inhomogeneous appearance of the Si islands on glass is considered to be due to selective erosion of nuclei with highly stressed bonds during plasma CVD.[231]

5.1.4 Surface Oxidation

One of the reasons for the development of AFM was the wish to pursue surface oxidation on an atomic scale. STM had already rendered images of the first stages of oxidation of Ni, Si and Al.[6,235–237] On Mg, oxide layers with a thickness of up to 2.5 nm could be investigated via STM;[238] MgO grew in domains with diameters of 2–6 nm.

In contrast to STM, AFM allows imaging not only of oxidized islands, but also of totally covering oxide layers. On AFM images with atomic resolution, a 2.3-nm thick natural oxide layer on Si(111) displayed 0.05-nm high regular waves resembling the crystalline structure of Si.[236] Images with lower resolution showed hillocks with diameters of about 40 nm and height differences < 1 nm.[26]

Surface reactions such as oxidation give rise not only to roughness changes, but also to changes in the measured forces (see section 5.5.4).[239]

5.1.5 Island Growth of Self-assembling Organosilanes

n-Octadecyltrichlorosilane (OTS) tends to form islands[57,240] before a molecular layer is closed. AFM served for determining the number, width and height of these islands.

According to Fourier transform infrared (FTIR) spectroscopy, OTS develops a 2.6-nm thick, closed monomolecular film on float glass after an immersion time of 90 min.[240] This finding was supported by AFM measurements. After immersion times of <90 min, individual islands with a height < 2.6 nm emerged. The height, which was less than the thickness of one monolayer, was explained by the lower order of the molecular chains. After longer immersion times, the roughness increased from the minimum of $R_a = 0.07$ nm at 90 min because of individual islands with low orientation on the first oriented covering monolayer. FTIR measurements give average values over some surface layers and could not resolve the local arrangement of ordered monolayers and islands with low order.[240]

In contrast to OTS, octadecylsiloxane (ODS) submonolayers show two types of growth: an island-type growth with roughly vertically aligned hydrocarbon chains and formation of disordered, liquid-like submonolayers.[241] Which type of growth prevails depends on the relative coverage and other preparation parameters.

5.2 Surface of Coatings: Defects, Roughness

5.2.1 Defects: Holes and Cracks

Figure 13(a) and (b) gives an example of the capability of defect detection with AFM. An $Fe_2O_3 \cdot SiO_2$ sol–gel dip coating was applied on alkali silicate glass without

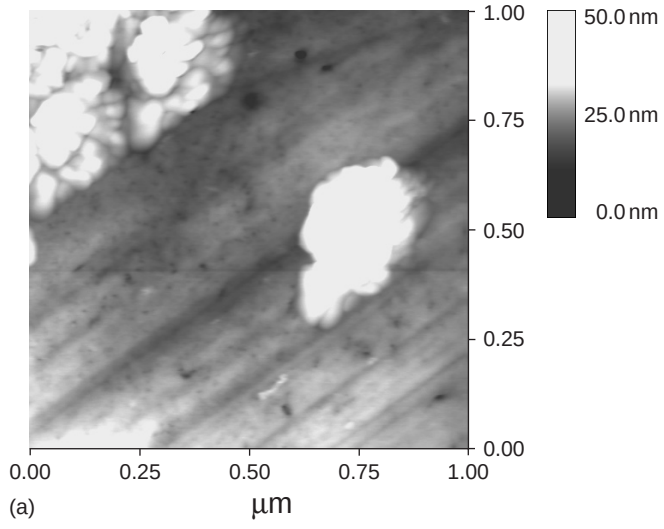

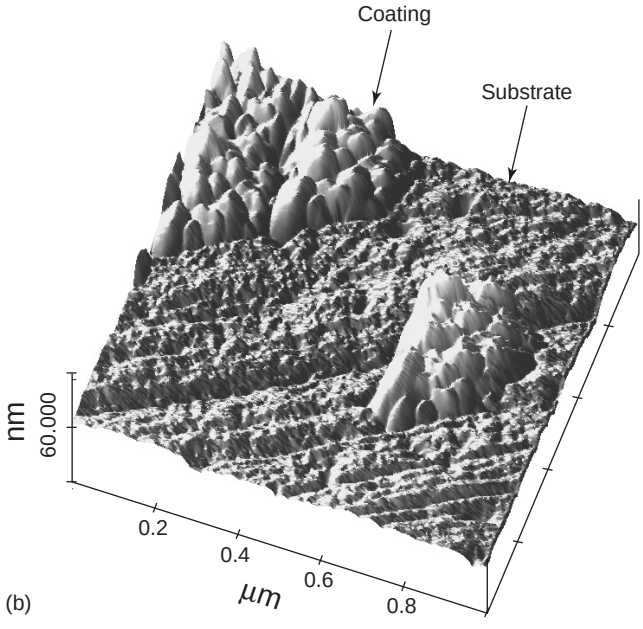

Figure 13 Hole in an $Fe_2O_3 \cdot SiO_2$ coating on alkali silicate glass. The glass was polished with $BaSO_4$ prior to sol–gel dip coating; uncoated parts of the substrate are recognized from the polishing grooves.

a coupling agent.[242,243] The sample appeared homogeneous under an optical microscope, but AFM revealed partially uncovered substrate. The glass could clearly be discerned from the coating because of the polishing grooves. With an adequate coupling agent, no such holes were detected.

Submicrometer defects in a sol–gel coating are displayed in Figure 14. A sol–gel dip coating of STZ was heated to 200 °C within 45 min and then investigated by secondary neutrals mass spectrometry (SNMS).[245] By sputtering with Kr ions for 330 s, 60 nm of the 90-nm thick coating were removed. The sample was then transferred

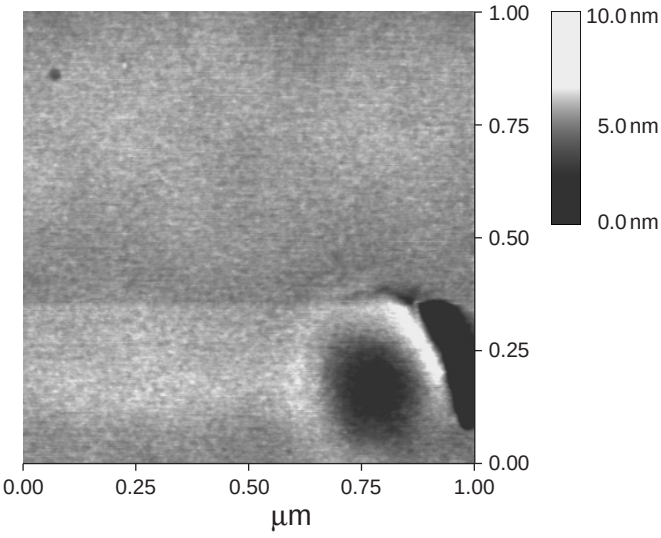

Figure 14 Pores in a partially consolidated and ion sputtered $SiO_2 \cdot TiO_2 \cdot ZrO_2$ (STZ) sol–gel coating. In the top left corner, a small hole with a diameter of 40 nm and a depth of 2 nm is detected. The large pore with a diameter of 300 nm is most probably due to an interfacial bubble and was partially opened by the AFM tip.

into a contact mode AFM instrument to investigate the sputtering crater. The topography of the sputtered crater is homogeneous with a few exceptions: in the top right corner a small pore with a diameter of 40 nm and a depth of 2 nm is detected. The large pore with a diameter of 300 nm is most probably due to a bubble at the interface between coating and glass. The shape of the bubble suggests that it was partially opened by the AFM tip interaction.[8]

Figure 15 displays a coating system with a mismatch in thermal expansion.[177] The conductivity of ITO, (In_2O_3–SnO_2) coatings is affected by Na^+ diffusing from the substrate glass into the ITO. The charge carrier concentration is reduced by this contamination. Fused silica would be a good substrate material because of its purity. The thermal expansion coefficient of fused silica ($\alpha = 0.5 \times 10^{-6}\,K^{-1}$), however, is much smaller than that of ITO ($\alpha = 7.2 \times 10^{-6}\,K^{-1}$). ITO is applied at elevated temperature and during the cooling the coating cracks, so no homogeneous products can be obtained.[246]

Commercial glass containers are coated with an oxide (SnO_2 or TiO_2) and an organic protection (waxes or polyolefins) against friction and wear. Figure 16 displays the surface of a green bottle with defects in the organic coating.[8] These rectangular and triangular flat holes are presumably due to NaCl crystals originating from the oxide coating. The oxide coating is applied by pyrolysis of, e.g. $SnCl_4$ before the annealing lehr. Sodium from the glass can react with the chloride and the resulting salt crystals fall off and leave 12 ± 0.5 nm deep holes in the second coating.

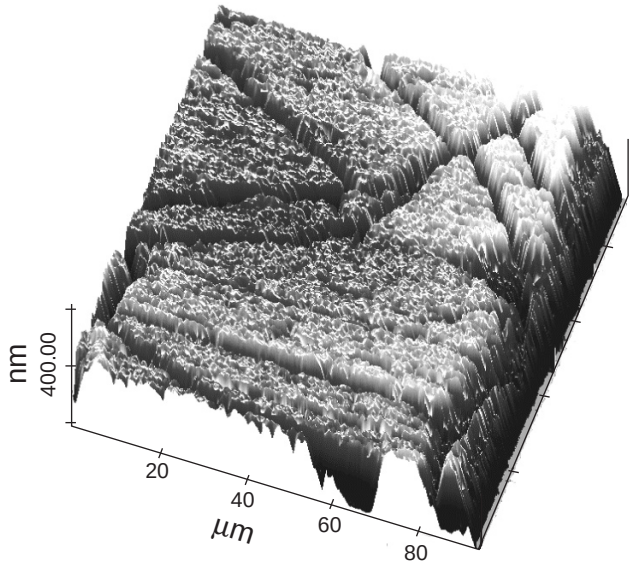

Figure 15 AFM image of a cracked indium tin oxide (ITO) sol–gel coating on fused silica.[246]

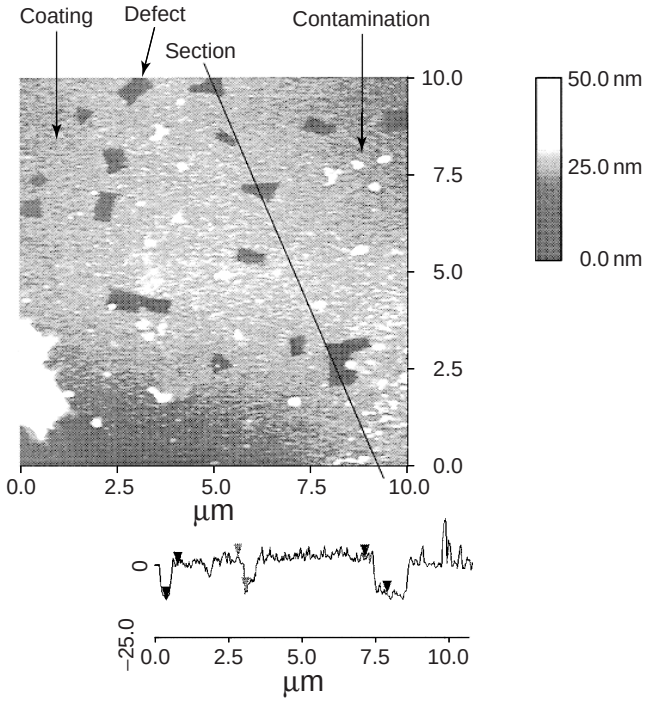

Figure 16 Surface of a used bottle. The regular flat holes in the organic coating are ascribed to NaCl crystals which originated from the first coating and fell off during the later life of the bottle.[8]

5.2.2 Roughness

Roughness values measured by SPM can be too high if the instrumental noise is too high and too low if the tip is blunt. For credible measurements, optimum resolution and low noise are necessary. If both are guaranteed, AFM can measure roughness even below RMS = 0.1 nm on areas in the square micrometers range on polished surfaces and films.[145] (All RMS values in the following text are valid for areas of square micrometers unless stated otherwise).

For the miniaturization of semiconductor devices, removing the uncontrolled oxidation layer is decisive. Especially the roughness after wet chemical or plasma etching is important. AFM roughness measurements help to improve the etching processes.[239,247]

During production of metal oxide semiconductor components, the natural oxide is removed by HF etching before the controlled growth of the gate oxide. On the natural oxide on Si(100) and Si(111), Crossley et al. measured RMS = 0.8 and 1.4 nm, respectively.[247] HF etching increased the roughness to RMS = 1.2 and 1.8 nm, respectively.

A much lower roughness of RMS = 0.1 nm was found for the natural oxide layer on Si wafers by Duparré and Jakobs.[161] They detected a similar roughness on superpolished fused silica. In addition to SiO_2, also 500–1200 nm thick borophosphosilicate glass (BPSG) CVD coatings are used as isolators in integrated circuits. The roughness of these coatings first increases with increasing thickness and then reaches saturation at RMS = 18 nm.[248] Subsequent densification of the coatings at 850 °C reduces the roughness back to RMS = 2 nm. Such smooth surfaces are desired for the detection of unwanted defects smaller than 0.2 μm.

5.3 Thickness

In some cases, coating defects such as holes or scratches can help to measure the thickness without any further preparation (see e.g. Figure 19 section 5.4.1). If the nanostructure of the coating and substrate allow the detection of the interface, thickness can be measured on AFM images of fracture surfaces.[23,174,246] One advantage over conventional thickness determination methods is the high lateral resolution which permits the detection of local thickness variations and measurements with difficult sample geometry.

Yoshino coated float glass partially with fluoralkyltriethoxysilane.[249] At the boundary with the uncoated glass, the exact thickness of the coating could be measured to be 6 and 9 nm, which correspond to two and three molecular layers, respectively.

The easiest way to measure the thickness of sol–gel coatings is to make a scratch with a plastic blade, immediately after coating in the soft layer. After annealing, the densified gel can be investigated in a mechanical profiler. An AFM image of such a scratch is displayed in Figure 17. A float glass dip coated with

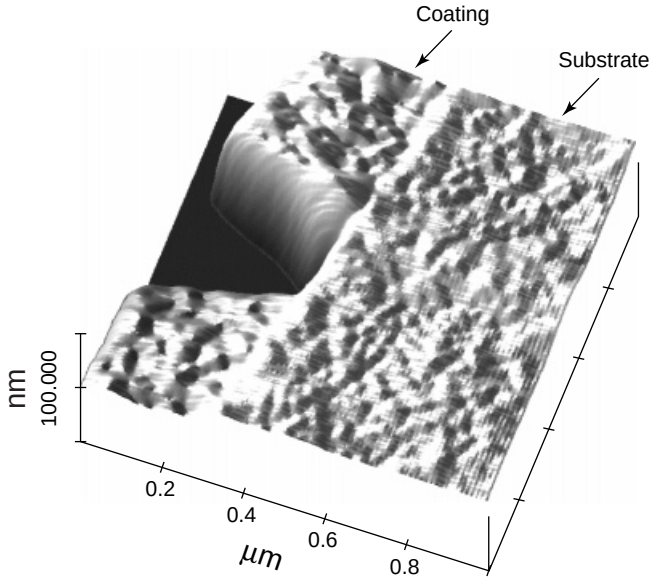

Figure 17 Fracture surface of float glass with a 300 nm thick sol–gel coating.[8] The STZ dip coating was scratched with a plastic knife, heated to 100 °C, notched and fractured for imaging the fracture mirror in the AFM instrument. The interface between the not fully densified film and the glass is clearly recognizable and allows thickness determination.

STZ was scratched with a plastic blade, heated to 100 °C, notched and fractured. Such coatings are not yet fully densified and do not withstand mechanical profilometry. AFM images help to monitor the thickness changes during consolidation even at temperatures as low as 100 °C. The STZ coating in Figure 17 is 300 nm thick.

Samples with a difficult geometry for the determination of coating thickness include fibers. Röder developed a preparation method for cross-sections of sol–gel–coated glass fibers.[194] She embedded STZ-coated Cemfil glass fibers in Araldite and sawed and polished the samples perpendicular to the fiber axis. Polishing contaminants and Araldite on the fiber cross-section were removed by 5–10 s of etching in 0.5% HF. The HF etching rate of Cemfil is higher than that of STZ, so the coating can be visualized as an elevated wall around the fiber (see Figure 18). The AFM measurement supported the assumption that similar sol–gel coatings on glass fibers are thinner than adequate coatings on flat substrates because of the lower roughness and surface tension of the fibers. In the case of the STZ coatings, the difference amounts to 15%.[194,250]

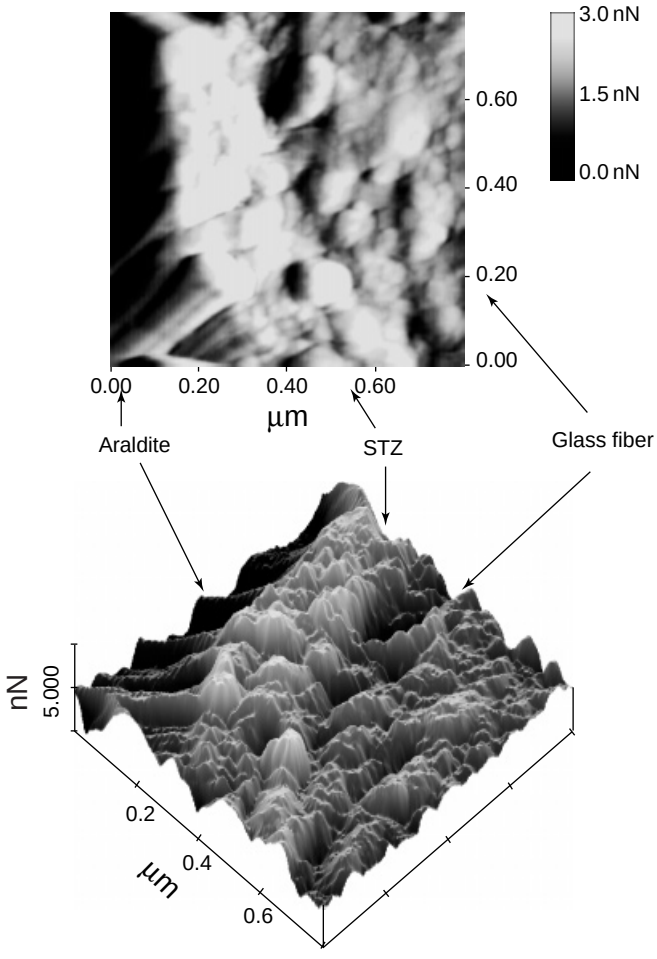

Figure 18 AFM images for thickness measurements on a sol–gel coating on glass fibers.[194,250] The coated fibers are embedded in Araldite, the cross-sections have been etched in HF and the STZ coating is visible as an elevated wall with a thickness of 120–160 nm.

5.4 Microstructure and Nanostructure

5.4.1 Crystal Phases, Homogeneity

ITO coatings have great commercial significance as transparent conducting electrodes, e.g. for flat panel displays. The crystal phases of ITO from different coating methods have been observed with SPM by several workers.[177,189,246,251–255] STM and AFM helped to understand that below a nominal thickness of 20 nm epitaxial ITO does not yet cover the substrate glass, but forms connected islands.[253]

ITO can be sputtered with metal (ITO_{me}) or oxide targets (ITO_{ox}). On ITO_{ox} SPM reveals a homogeneous surface with columnar growth of one type of crystal. The lower conductivity of ITO_{me} was shown to be due to an inhomogeneous microstructure of columns perpendicular to the surface, oblique columns and large, flat grains.[189,255]

ITO from the sol–gel process resembles the ITO_{ox} sputtered type, with a homogenous surface. The columnar growth can be seen on the scratch in Figure 19. The representation exaggerates the columnar character of the crystallites. Their average diameter is 25 nm and the

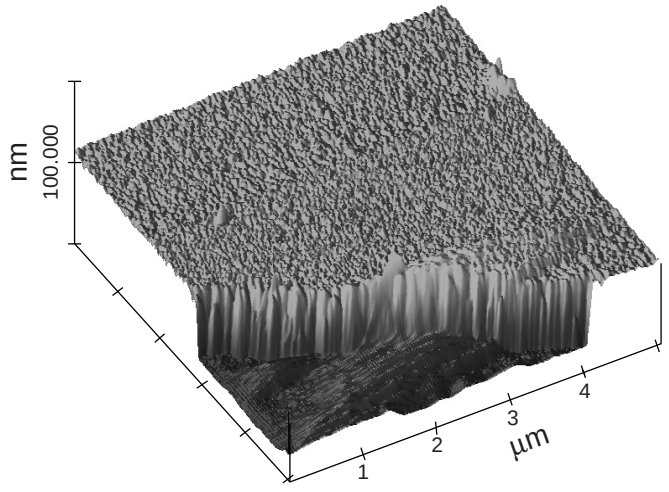

Figure 19 Scratch in a 50 nm thick ITO sol–gel coating with columnar crystal growth[177,246] (the height : diameter ratio of 2 : 1 is exaggerated by the representation).

height of most of the columns is equal to the coating thickness of 50 nm.[177,246]

The crystallization of ITO coatings was also pursued with AFM.[177,246] The crystal phase was identified by X-ray diffraction (XRD) and AFM served as a fast method to measure the growth of the crystallites. Figure 20(a–d) shows ITO sol–gel coatings annealed at different temperatures. Between 250 and 400 °C, crystallization starts and the average diameter of the crystals increases from 20 nm at 400 °C to 35 nm at 500 °C. Homogeneity and crystal diameters are two of the important parameters for the conductivity of the semiconductive coating materials.

The ideal SnO_2 concentration for ITO coatings is about 10%.[246,253,256,257] At 14%, the solubility of SnO_2 in the In_2O_3 lattice is exceeded. On AFM images a larger amount of small crystals is detected, which can be correlated with metallic Sn in the XRD spectra.[246]

The performance of scratch-resistant Al_2O_3 sol–gel coatings is dependent on the crystallization of α-Al_2O_3 (corundum).[258] The best results were achieved with 70 nm thick cryptocrystalline coatings from alkoxide solutions. These coatings were too thin for XRD analysis. The coatings were optimized by comparing AFM images

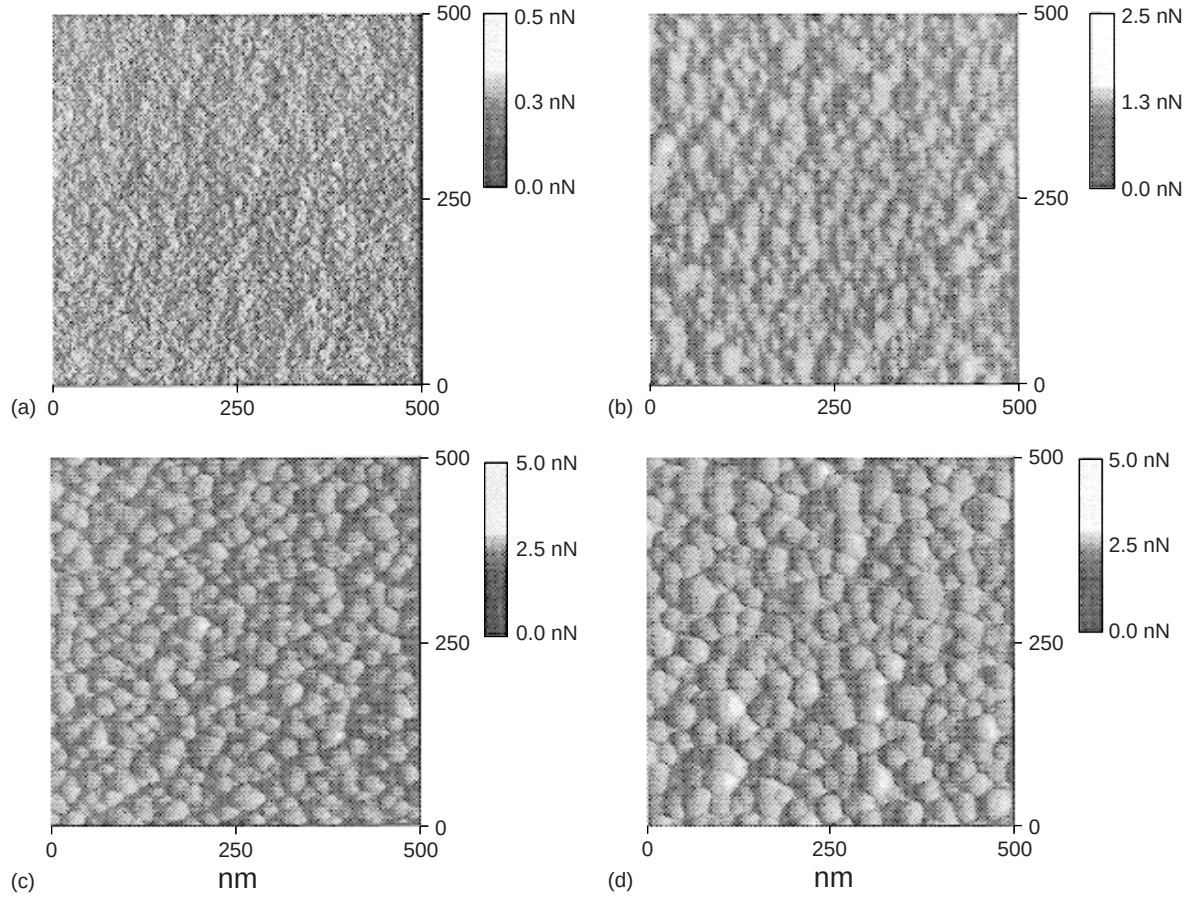

Figure 20 AFM images of the surface of ITO sol–gel coatings after annealing at (a) 250, (b) 400, (c) 450 and (d) 500 °C.[177,246] (a) Appears vitreous; on (b), (c) and (d) the average diameter of the crystals is 20, 25 and 35 nm, respectively.

with those of α-Al_2O_3 coatings which were thick enough for identification of crystal phases by XRD. AFM images of the optimized coatings displayed the same morphology.

Another example of the complementary information obtained by XRD and AFM is the investigation of lead titanate coatings by Lai et al.[259] All coatings annealed above 300 °C displayed crystal peaks in the XRD spectra. In contrast, the AFM images displayed the typical pattern of vitreous coating even after annealing at 400 °C. On samples annealed at 450 °C, the AFM revealed crystallites which grew through the thin superficial vitreous film on the coating surface. Hence both the phases and their local distribution could be made clear.

5.4.2 *Correlation of the Topography of Coatings and Substrates*

Meyer et al. investigated the dependence of the topography of TiO_2 and Ta_2O_5 coatings on the substrate and the coating process.[260] Electron microscopy of cross-sections had given hints on differences before, but AFM resolved details much better. It was shown that the Ta_2O_5 coatings grow mainly in a two-dimensional manner and hence replicate the original substrate topography. TiO_2 coatings cover the polishing traces and develop a microstructure which is independent of the substrate. In spite of the thickness of 100 nm, the roughness of the Ta_2O_5 coatings, RMS = 1 nm, was similar to the polished optical glass substrate. The TiO_2 coatings displayed RMS = 4.6 nm (ion plating) and 7.7 nm (reactive evaporation). The last, high value is presumably caused by the columnar growth of the crystalline layer even in the first stages.

MgF_2 is also an optical coating with pronounced columnar growth. This had been shown by TEM images of replicas before.[261] The exact determination of crystallite diameters and roughness was possible only by AFM.[161,180,181]

5.4.3 *Interfaces*

Cross-sections of coatings with good adhesion to the substrate can easily be imaged on fracture surfaces. The coating itself can only be recognized on AFM images if the fracture morphology is different from that of the substrate.

One example of such a coating is SnO_2 on glass, which serves for insulating windows (see Figure 10a–c and Figure 21a). The interface line is dominated by the shape of the crystals. The shape of the crystal grains is represented truly, in contrast to TEM images of ultramicrotome cuts. On those images,[185] the crystals appear with a slim columnar shape because of a preparation artifact and the grooves between the grains on the surface are not resolved because of the integral image of a slice. Ultramicrotomy tends to disguise grain boundaries, especially of coatings with low relative density.[178]

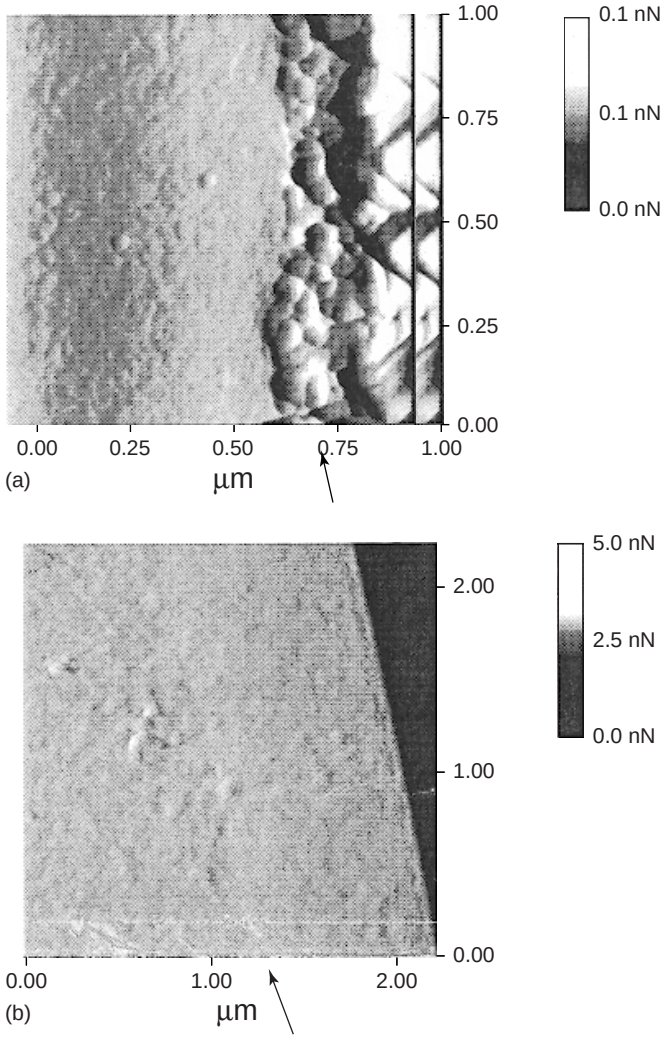

Figure 21 Fracture surfaces of coated glasses. The arrow marks the interface. (a) Edge of SnO_2-coated glass. In the glass fracture mirror, the 300 nm thick pyrolytic crystalline coating displays a rougher surface and larger hillocks than the smooth glass fracture surface. (b) Edge of SiO_2-coated glass. In the glass fracture mirror, this 1 µm thick vitreous coating from the sol–gel process displays a microstructure similar to that of the float glass.

In contrast to the AFM images, TEM images of ultramicrotome cuts reveal an interface layer between tin oxide and glass. This hydrosilicon type of interlayer is applied to prevent iridescence and as an Na^+ diffusion barrier. On TEM images, it appears similar to the vitreous substrate, but brighter.[185]

The sample in Figure 21(b) is an SiO_2 sol–gel coating. After annealing at 500 °C for 1 h, the microstructure of the coating is similar to that of fused silica; the fracture surface resembles that of the float glass and the interface is barely visible.

Stathatos et al.[262] and Liu[177] showed with the help of AFM that thicker sol–gel coatings of TiO_2 and ITO which are manufactured by multiple dipping are homogeneous. In the coating volume, no interfaces or inhomogeneities were found at the boundaries of the single layers.

5.4.4 Porosity

AFM images of porous materials do not give the same three-dimensional impression as TEM images with extremely high resolution, because the AFM tip cannot see ''inside'' the pores. The literature contains examples of AFM images of porous material,[263–268] and of TEM images,[269,270] while a comparison of the two is also available.[271] Generally, the actual radius of the AFM tip is unknown, and therefore the lower detection limit for small pores is also unknown. Typically, this radius is >10 nm. On conductive coatings such as porous Si, fine STM tips can detect pores with radii of 1–2 nm.[272]

The rougher fracture surface of the STZ sol–gel coatings in Figures 17 and 22 is due to the porous microstructure of not fully consolidated gels. AFM images can help to pursue microstructural changes after different steps of consolidation. In STZ, even after annealing at 500 °C for 1 h, the pores remain. Figure 22 shows clusters, rings and linear alignments of particles with diameters of 70–400 nm in such a consolidated 1 μm thick coating.[23] Pore diameters can only be estimated to be in the same range. For exact values, the tip radius must be determined independently.

In contrast to STZ, fractured dense SiO_2 sol–gel coatings from tetraethylorthosilicate (TEOS) show a hillock pattern similar to the pattern of the substrate glass, and the interface is barely visible. Vong and Sermon investigated porous SiO_2 coatings from silica sols with different diameters of the colloid particles.[271] Sols with colloidal particles with average diameters of 17 nm resulted in porous coatings with particle diameters of 25–30 nm. In porous coatings from sols with average diameters of 100 nm, the particle diameter did not change during annealing.

AFM investigations supported other measurements of the development of crystal phases and optical properties of TiO_2 coatings during annealing.[273] Martin et al.[273] coated fused silica and Si(100) via magnetron sputtering and annealed the TiO_2 coating for 1 h at different temperatures. Up to 700 °C, the hillocks with diameters of 40–50 nm did not change much. Between 300 and 700 °C, the roughness decreased from RMS = 4 to 3 nm, which correlated with a densification of the porous coating and index increase. Rutile crystallization causes dramatic topography changes from 700 °C on. Crystal facets developed, diameters increased to 100–200 nm at 900 °C and >500 nm at 1100 °C and roughness increased to RMS = 20 nm at 1100 °C.

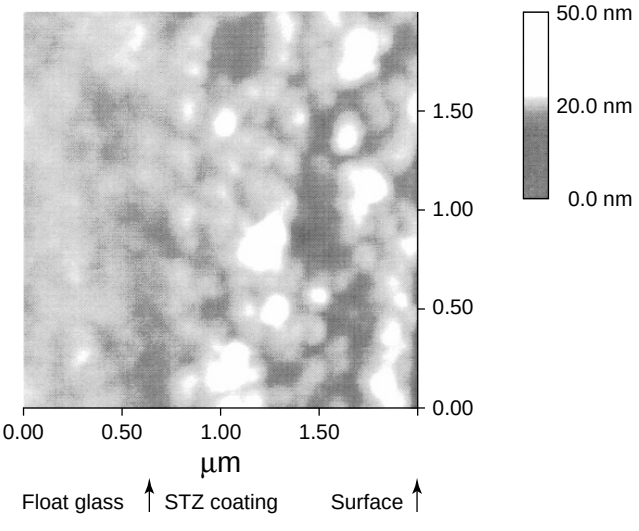

Figure 22 Pores, channels, clusters and rings of particles with diameters of 70–400 nm on the fracture surface of a 1000 nm thick STZ sol–gel coating consolidated at 500 °C for 1 h.

5.5 Mechanical Properties of Coatings

5.5.1 Friction and Wear of Magnetic Storage Media

Customary magnetic storage media consist of an Al–Mg alloy substrate disk with a hard coating of Ni–P, the magnetic film (e.g. $Co_{79}Pt_{14}Ni_7$) and a top coating. Special glass or glass ceramic disks are suggested instead of the metal alloy because of their smoother surface and higher temperature and shape stability.[159] They can be manufactured with higher data density and lower thickness. Especially for small, hard disks and portable computers, glass disks are already used. Bhushan et al. compared the friction and wear of commercial hard disks of Al–Mg and glass and of laboratory glass ceramics and established parameters for a convenient determination of roughness.[159,160]

As the distance between the hard disk and recording head is to be reduced from >50 nm to <25 nm, each single high protrusion is detrimental. Therefore, the highest z elongation is important in addition to RMS and correlation length. The adhesion forces due to the meniscus of lubricant between the recording head and disk and the friction were higher for rougher samples. Wear was dependent on the top coating and the environment. The durability of common carbon top coatings is reduced by moist or O_2-containing atmospheres. The glass disks were protected with sputtered SiO_2 with higher durability, which is even improved by a humid atmosphere.

5.5.2 *Wear Tests of Hard Coatings*

Oberländer et al. used AFM to investigate the wear processes of hard coatings of TiN and partially oxidized TiN_xO_y during friction against steel.[274] The most important damage to the harder TiN coatings were scratches and delamination, caused by oxidized spots.

The scratch resistance of α-Al_2O_3 hard coatings on float glass was investigated with the help of AFM by Hauk et al.[258,275] A scratch tester, which simulated the load on windscreens by dirty wipers, moved a quartz grain with a constant velocity of $2\,cm\,s^{-1}$ and constant loads up to 100 g across the samples. Figure 23(a) and (b) displays AFM force mode images of scratches with a load of 100 g on normal float glass (Flachglas, Weiherhammer, Germany) and on a 70-nm thick alkoxide gel coating prepared from aluminum butoxide. The scratch on bare float glass is about 60 µm wide. In the trace, stripes of undamaged glass and grooves are found; most of the damage is due to prominent stripes of material dragged out of the surface. These up to 1.2-µm high elevations cause stray light on degraded windscreens. On the coated part of the sample, the scratch can only be detected by AFM. The width of the trace is about the same, but the elevations of <20 nm do not interfere with visible light.[174,258]

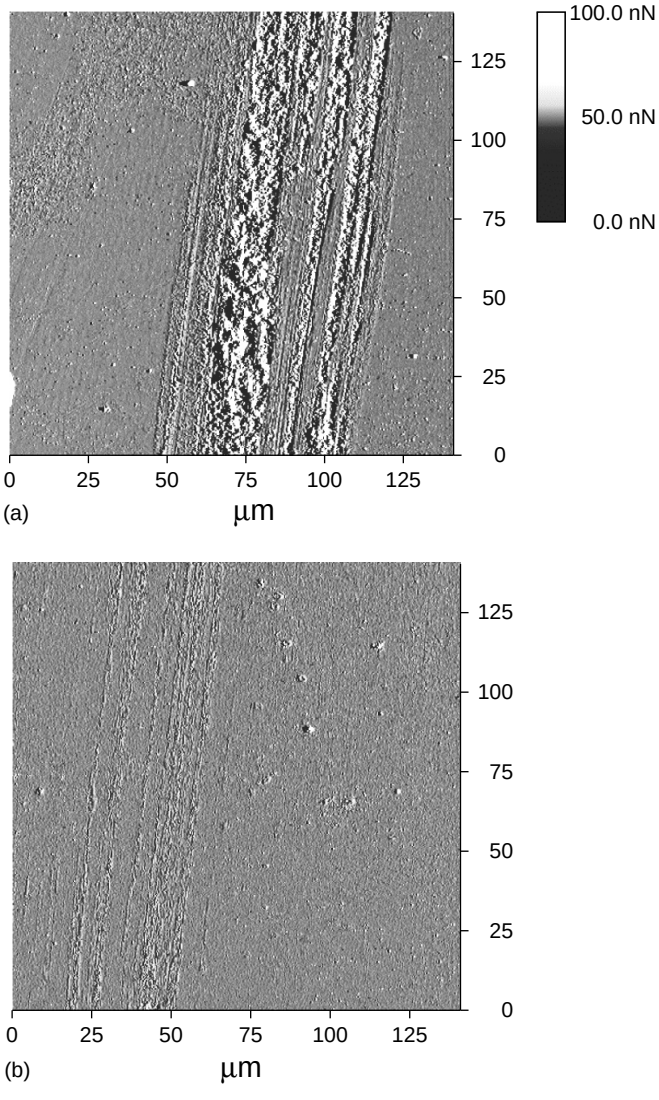

Figure 23 AFM force mode images of scratches on (a) bare float glass and (b) float glass with a 70 nm thick hard coating of α-Al_2O_3.[258,275] The 1.2 µm high elevations give rise to stray light on degraded windscreens; the scratch on the hard coated part of the sample can only be detected by a high-resolution microscope technique such as AFM.

5.5.3 *Tribology of Carbon Coatings*

Snitka et al. produced 180–360 nm thick C coatings on Si, GaAs and fused silica via electron beam evaporation.[276] With ion beam support, the coatings became DLC; without ion beam support, porous non-crystalline a-C developed. AFM served for the estimation of surface energy from the tip adhesion and for the judgement of the scratches from two tribological tests (diamond tip on plate and steel ball on plate). The two tribological tests caused different damage to the a-C coating on Si(100). The diamond tip produced traces typical of plastic behavior; under the steel ball, the coating delaminated as one would expect for a vitreous material. The DLC coatings proved to be much more scratch resistant than a-C. On a-C–Si an adhesion of 1.8 nN was measured, on DLC–Si values of about 20 nN were found and on DLC–fused silica the adhesion amounted 76 nN. The surface energy was estimated to be $6\,mJ\,m^{-2}$ (a-C–Si) and $68\,mJ\,m^{-2}$ (DLC–Si, DLC–GaAs), with an estimated tip radius of roughly 50 nm. On DLC–fused silica, a jump to contact occurred at a distance of 20 nm, which was a sign of a liquid capillary bridge of adsorbed water; the surface energy was estimated to be about $120\,mJ\,m^{-2}$.

5.5.4 *Atomic Force Microscopy Tip Adhesion as a Hint to Chemical Modification*

Friedbacher et al. measured the roughness of differently pretreated Si substrates and the pull-off force of the AFM tip.[239] The removal of the oxide layer resulted in a lower pull-off force: on the natural oxide layer, this force was 2 nN, and after etching in an H_2 plasma or in HF, it decreased below 0.5 nN. The force–distance curves allowed conclusions on the species on the surface. The bias treatment in hydrogen created an H-terminated surface; in a mixture of hydrogen and methane, SiC was found.

Kiridena et al. removed the hydrosilicone film partially with the AFM tip and measured the Young's modulus

of film and substrate.[92] The result, which contrasts with experience in the macroscopic world, is explained by the fact that modulus measurements via phase shift are sensitive only to the first atom layers. In air, mica is covered with a water film, which is detected by force modulation instead of the mica. With butanol as environment, the measured Young's modulus of mica was higher than that of the hydrosilicone. With increasing modulation amplitude, the modulus increased to values close to the bulk values.[92]

5.5.5 Peeling Tests

AFM improves conventional peeling tests. The decision as to whether a fracture is adhesive or cohesive depends on the detection of small remnants on the surface, which is better performed with a high-resolution method such as AFM. Creuzet et al. aged a compound of float glass and PU and analyzed the glass after the peeling test.[277,278] The aged elastomer delaminated cohesively. However, also without any aging, 20-nm high polymer particles with diameters of 80 nm were found on the glass surface after delamination. These hillocks were identified to be different from the glass composition by force–distance curves. From these results, it can be concluded that the delamination takes place in an interface layer in the PU which had been damaged by ion migration from the glass surface. The hillocks are interpreted as PU fibrils which developed during the peeling, fractured and collapsed on the glass surface.[279]

5.6 Corrosion and Aging

Crystallization and corrosion of sol–gel coatings of TiO_2 and in the system $Al_2O_3 \cdot TiO_2 \cdot SiO_2$ were investigated with AFM by Du et al.[280,281] In both cases, the corrosion behavior depended on the crystal phases. The roughness of $Al_2O_3 \cdot TiO_2 \cdot SiO_2$ increased gradually with time after treatment in 1 M NaOH at 60 °C, and the sensitivity to scratches by the scanning tip also increased. According to their shape, the crystallites which developed during corrosion could be divided into three groups. At high TiO_2 contents, quadrangles with the habit of anatase developed, presumably by precipitation from the solution. At an $Al_2O_3 : SiO_2$ molar ratio of 1 : 2, hexagonal crystallites were found, which could be due to nepheline. On coatings with other compositions, irregular hillocks without any crystal facets could not be correlated with a distinct crystal shape.

Helsch investigated the aging of porous SiO_2 antireflective coatings in different media.[268] The chemical processes were analyzed by SNM and the habit and distribution of reaction products on the surface were imaged by AFM. Beside other reactions, the long-term transmission losses of the 100-nm thick coatings in normal air were found to be due to 500-nm thick individual crystals on the surface, consisting of $CaCO_3$. The large inner surface of the coatings allowed Ca^{2+} diffusion from the substrate glass to the surface, where Ca^{2+} reacted with CO_2.

ACKNOWLEDGMENTS

The financial support of the Deutsche Forschungsgemeinschaft (Ra 706/2-1) for parts of this work is gratefully acknowledged. The author thanks Professor G.H. Frischat for many helpful discussions.

ABBREVIATIONS AND ACRONYMS

AFM	Atomic Force Microscopy
AgIF	Silver Island Film
ARS	Angle Resolved Scattering
BPSG	Borophosphosilicate Glass
CSEM	Centre Suisse d'Electronique et de Microtechnique
CVD	Chemical Vapor Deposition
DLC	Diamond-like Carbon
EDX	Energy-dispersive X-ray Spectrometry
ESEM	Environmental Scanning Electron Microscopy
FIM	Field Ion Microscopy
FTIR	Fourier Transform Infrared
GIXR	Grazing Incidence X-ray Reflection
HFSEM	High-frequency Scanning Electrical Force Microscopy
HOPG	Highly Oriented Pyrolytic Graphite
IMMA	Ion Microprobe Mass Analysis
ITO	Indium Tin Oxide
LFM	Lateral Force Microscopy
LFS	Local Force Spectroscopy
MFM	Magnetic Force Microscopy
ODS	Octadecylsiloxane
OTS	*n*-Octadecyltrichlorosilane
PSD	Power Spectral Density
PU	Polyurethane
PVD	Physical Vapor Deposition
RHEED	Reflection High-energy Electron Diffraction
RMS	Root Mean Square
SANS	Small-angle Scattering of Neutrons
SAXS	Small-angle Diffraction of X-rays
SEM	Scanning Electron Microscopy
SERS	Surface-enhanced Raman Spectroscopy
SICM	Scanning Ion Conductance Microscopy

SNM	Scanning Noise Microscopy
SNMS	Secondary Neutrals Mass Spectrometry
SNOM	Scanning Near-field Optical Microscopy
SPM	Scanning Probe Microscopy
SThM	Scanning Thermal Microscopy
STM	Scanning Tunneling Microscopy
STM-LE	Inverse Photoemission Spectrometry
STZ	$SiO_2 \cdot TiO_2 \cdot ZrO_2$
TAM	Tunneling Acoustic Microscopy
TEM	Transmission Electron Microscopy
TEOS	Tetraethylorthosilicate
TIS	Total Integrated Scattering
UHV	Ultrahigh Vacuum
WDX	Wavelength-dispersive X-ray Spectrometry
XRD	X-ray Diffraction

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction

Polymers and Rubbers **(Volume 8)**
Atomic Force Microscopy in Analysis of Polymers

Polymers and Rubbers cont'd **(Volume 9)**
Surface Energetics of Polymers and Rubbers, Characterization of

Surfaces **(Volume 10)**
Electron Microscopy and Scanning Microanalysis • Ellipsometry in Analysis of Surfaces and Thin Films • Proximal Probe Techniques • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Probe Microscopy, Industrial Applications of • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

REFERENCES

1. N.A. Burnham, R.J. Colton, 'Measuring the Nanomechanical Properties and Surface Forces of Materials Using an Atomic Force Microscope', *J. Vac. Sci. Technol. A*, **7**, 2906–2913 (1989).
2. E.H. Synge, 'A Suggested Method for Extending Microscopic Resolution into the Ultramicroscopy Region', *Philos. Mag.*, **6**, 356–362 (1928).
3. E.H. Synge, 'An Application of Piezo-electricity to Microscopy', *Philos. Mag.*, **13**, 297–300 (1932).
4. G. Binnig, H. Rohrer, C. Gerber, E. Weibel, 'Surface Studies by Scanning Tunneling Microscopy', *Phys. Rev. Lett.*, **49**, 57–61 (1982).
5. G. Binnig, H. Rohrer, 'Scanning Tunneling Microscopy', *IBM J. Res. Dev.*, **30**, 355–369 (1986).
6. G. Binnig, H. Rohrer, 'Scanning Tunneling Microscopy – From Birth to Adolescence', *Rev. Mod. Phys.*, **59**, 615–625 (1987).
7. G. Binnig, C.F. Quate, C. Gerber, 'Atomic Force Microscopy', *Phys. Rev. Lett.*, **56**, 930–933 (1986).
8. E. Rädlein, *Werkstoffkundliche Beurteilung von Gläsern und Schichten Mittels Rastersondenmikroskopie*, Habilitation, TU Clausthal, Clausthal-Zellerfeld, 1999.
9. R. Howland, L. Benatar, *A Practical Guide to Scanning Probe Microscopy*, Park Scientific Instruments, Sunnyvale, CA, 1997.
10. V. Stary, 'Atomar ebene Kristalloberflächen', Diploma Thesis, Fachhochschule Jena, Jena, 1994.
11. H.-J. Güntherodt, R. Wiesendanger (eds.), *Scanning Tunneling Microscopy I*, Springer Series in Surface Science, 20, Springer, Berlin, 1992.
12. R. Wiesendanger, H.-J. Güntherodt (eds.), *Scanning Tunneling Microscopy II*, Springer Series in Surface Science, 28, Springer, Berlin, 1992.
13. R. Wiesendanger, H.-J. Güntherodt (eds.), *Scanning Tunneling Microscopy III*, Springer Series in Surface Science, 29, Springer, Berlin, 1993.
14. R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, 1994.
15. R. Wiesendanger, 'Scanning Probe Microscopy', in *Nanoscience and Technology*, eds. K. von Klitzing, R. Wiesendanger, Springer, Berlin, 1998.
16. C. Hamann, M. Hietschold, *Raster-Tunnel-Mikroskopie*, Akademie Verlag, Berlin, 1991.
17. R.J. Behm, N. Garcia, H. Rohrer (eds.), *Scanning Tunneling Microscopy and Related Methods*, NATO ASI Series E: Applied Sciences, 184, Kluwer Academic, Dordrecht, 1989.
18. J.A. Stroscio, W.J. Kaiser, 'Scanning Tunneling Microscopy', in *Methods of Experimental Physics*, eds. R. Celotta, T. Lucatorto, Academic Press, San Diego, 27, 1993.
19. S.H. Cohen, M.T. Bray, M.L. Lightbody, *Atomic Force Microscopy/Scanning Tunneling Microscopy*, Plenum Press, New York, 1994.
20. N.J. Di Nardo, *Nanoscale Characterization of Surfaces and Interfaces*, VCH, Weinheim, 1994.
21. H.-J. Güntherodt, D. Anselmetti, E. Meyer (eds.), *Forces in Scanning Probe Methods*, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 1995.
22. S.N. Magonov, M.-H. Whangbo, *Surface Analysis with STM and AFM*, VCH, Weinheim, 1996.
23. E. Rädlein, G.H. Frischat, 'Atomic Force Microscopy as a Tool to Correlate Nanostructure to Properties of Glasses', *J. Non-Cryst. Solids*, **222**, 69–82 (1997).

24. G. Meyer, N.M. Amer, 'Novel Approach to Atomic Force Microscopy', *Appl. Phys. Lett.*, **53**, 1045–1047 (1988).
25. E. Meyer, H. Heinzelmann, H. Rudin, H.-J. Güntherodt, 'Atomic Resolution on LiF (001) by Atomic Force Microscopy', *Z. Phys. B*, **79**, 3–4 (1990).
26. S. Alexander, L. Hellemanns, O. Marti, J. Scheir, V. Eilings, P.K. Hansma, M. Longmire, J. Gurley, 'An Atomic Resolution Atomic-force Microscope Implemented Using an Optical Lever', *J. Appl. Phys.*, **65**, 164–167 (1989).
27. O. Marti, J. Colchero, 'Scanning Probe Microscopy Instrumentation', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 15–34, 1995.
28. E. Meyer, H. Heinzelmann, 'Scanning Force Microscopy (SFM)', in *Scanning Tunneling Microscopy II*, eds. R. Wiesendanger, H.-J. Güntherodt, Springer Series in Surface Science, 28, Springer, Berlin, 99–150, 1992.
29. Y. Martin, C.C. Williams, H.K. Wickramasinghe, 'Atomic Force Microscope–Force Mapping and Profiling on a Sub-100-Å Scale', *J. Appl. Phys.*, **61**, 4723–4729 (1987).
30. G.M. McClelland, R. Erlandsson, S. Chiang, *Rev. Prog. Qual. Non-destruct. Eval.*, 6, 1987.
31. C.M. Mate, M. Lorenz, V.J. Novotny, 'Atomic Force Microscopy of Polymeric Liquid Films', *J. Chem. Phys.*, **90**, 7550–7555 (1989).
32. G. Meyer, N.M. Amer, 'Optical-beam-deflection Atomic Force Microscopy: The NaCl (001) Surface', *Appl. Phys. Lett.*, **56**, 2100–2101 (1990).
33. N. Blanc, J. Brugger, N.F. DeRooij, 'Electrostatically Actuated Silicon Micromachined Sensors for Scanning Force Microscopy', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 79–84, 1995.
34. C.F. Quate, 'The AFM as a Tool for Surface Imaging', *Surf. Sci.*, **299/300**, 980–995 (1994).
35. H.J. Hug, A. Moser, O. Fritz, B. Stiefel, I. Parashikov, 'Low Temperature Scanning Force Microscopy', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 35–62, 1995.
36. F.J. Giessibl, 'Atomic Resolution of the Silicon (111)–(7 × 7) Surface by Atomic Force Microscopy', *Science*, **267**, 68–71 (1995).
37. R. Linnemann, T. Gotszalk, L. Hadjiiski, I.W. Rangelow, 'Characterization of a Cantilever with an Integrated Deflection Sensor', *Thin Solid Films*, **264**, 159–164 (1995).
38. K.H. Besocke, 'An Easily Operable Scanning Tunneling Microscope', *Surf. Sci.*, **181**, 145–153 (1987).
39. G. Binnig, D.P.E. Smith, 'Single-tube Three-dimensional Scanner for Scanning Tunneling Microscopy', *Rev. Sci. Instrum.*, **57**, 1688–1689 (1986).
40. P. Muralt, D.W. Pohl, W. Denk, 'Wide-range, Low-operating-voltage, Bimorph STM: Application as Potentiometer', *IBM J. Res. Dev.*, **30**, 443–450 (1986).
41. J. Burger, 'Aufbau eines Rastertunnelmikroskops für tiefe Temperatuen und äußere Magnetfelder – Oberflächenuntersuchungen und Wachstumsstudien an Epitaktischen Hochtemperatursupraleiter-Dünnfilmen und -Einkristallen mit dem Rastertunnel- und Kraftmikroskop', *PhD Thesis*, FAU Erlangen-Nürnberg, Erlangen, 1992.
42. R.S. Howland, 'The Scanning Probe Microscope as a Metrology Tool', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York, 347–358, 1994.
43. G. Binnig, H. Rohrer, C. Gerber, E. Weibel, 'Tunneling Through a Controllable Vacuum Gap', *Appl. Phys. Lett.*, **40**, 178–180 (1982).
44. O. Wolter, T. Bayer, J. Greschner, 'Micromachined Silicon Sensors for Scanning Force Microscopy', *J. Vac. Sci. Technol. B*, **9**, 1353–1357 (1991).
45. D. Rugar, P. Hansma, 'Atomic Force Microscopy', *Phys. Today*, October, 23–30 (1990).
46. R.-H. Peters, 'Nur Davon Träumen', *Wirtschaftswoche*, **4/18.1**, 98–99 (1996).
47. W. Kulisch, A. Malave, G. Lippold, W. Scholz, C. Mihalcea, E. Oesterschulze, 'Fabrication of Integrated Diamond Cantilevers with Tips for SPM Applications', *Diamond Relat. Mater.*, **6**, 906–911 (1997).
48. G. Binnig, C. Gerber, E. Stoll, T.R. Albrecht, C.F. Quate, 'Atomic Resolution with Atomic Force Microscope', *Europhys. Lett.*, **3**, 1281–1286 (1987).
49. T.R. Albrecht, C.F. Quate, 'Atomic Resolution Imaging of a Nonconductor by Atomic Force Microscopy', *J. Appl. Phys.*, **62**, 2599–2602 (1987).
50. E. Oesterschulze, 'The Potential of Scanning Probe Microscopy for Thin Film Characterization', *Surf. Coat. Technol.*, **97**, 694–706 (1997).
51. H. Heinzelmann, P. Grütter, E. Meyer, H. Hidber, L. Rosenthaler, M. Ringger, H.-J. Güntherodt, 'Design of an Atomic Force Microscope and First Results', *Surf. Sci.*, **189/190**, 29–35 (1987).
52. H. Heinzelmann, P. Grütter, E. Meyer, H. Hidber, L. Rosenthaler, M. Ringger, H.-J. Güntherodt, 'Atomic Force Microscopy: General Aspects and Application to Insulators', *J. Vac. Sci. Technol. A*, **6**, 275–278 (1988).
53. C.M. Mate, G.M. McClelland, R. Erlandsson, S. Chiang, 'Atomic-scale Friction of a Tungsten Tip on a Graphite Surface', *Phys. Rev. Lett.*, **59**, 1942–1945 (1987).
54. S. Porthun, M. Rührig, J.C. Lodder, 'Magnetic Force Microscopy on Thin Film Magnetic Recording Media', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI

Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 471–476, 1995.
55. S. Münster, S. Werner, C. Mihalcea, W. Scholz, E. Oesterschulze, 'Novel Micromachined Cantilever Sensors for Scanning Near-field Optical Microscopy', *J. Microsc.*, **186**, 17–22 (1997).
56. A. Ruf, J. Diebel, M. Abraham, T.R. Dietrich, M. Lacher, 'Ultra-long Glass Tips for Atomic Force Microscopy', *J. Micromech. Microeng.*, **6**, 254–260 (1996).
57. Y.I. Rabinovich, R.-H. Yoon, 'Use of Atomic Force Microscope for the Measurement of Hydrophobic Forces between Silanated Silica Plate and Glass Sphere', *Langmuir*, **10**, 1903–1909 (1994).
58. M. Mantel, Y.I. Rabinovich, J.P. Wightman, R.-H. Yoon, 'A Study of Hydrophobic Interactions between Stainless Steel and Silanated Glass Surface Using Atomic Force Microscopy', *J. Colloid Interface Sci.*, **170**, 203–214 (1995).
59. J.P. Cleveland, S. Manne, D. Bocek, P.K. Hansma, 'A Nondestructive Method for Determining the Spring Constant of Cantilevers for Scanning Force Microscopy', *Rev. Sci. Instrum.*, **64**, 403–405 (1993).
60. D.M. Schaefer, M. Carpenter, B. Gady, R. Reifenberger, L.P. DeMejo, D.S. Rimai, 'Surface Roughness and its Influence on Particle Adhesion Using Atomic Force Techniques', in *Fundamentals of Adhesion and Interfaces*, eds. D.S. Rimai, L.P. DeMejo, K.L. Mittal, VSP, Utrecht, 35–48, 1995.
61. J.L. Hazel, V.V. Tsukruk, 'Quantitative Force Measurements in SPM: Vertical and Torsional Spring Constants for V-shaped Cantilevers', *Polym. Prepr.*, **37**, 567–568 (1996).
62. U.D. Schwarz, H. Haefke, P. Reimann, H.-J. Güntherodt, 'Tip Artefacts in Scanning Force Microscopy', *J. Microsco.*, **173**, 183–197 (1994).
63. G.J.C. Braithwaite, A. Howe, P.F. Luckham, 'Interactions between Poly(ethylene oxide) Layers Adsorbed to Glass Surface Probes by Using a Modified Atomic Force Microscope', *Langmuir*, **12**, 4224–4237 (1996).
64. R.M. Overney, R. Lüthi, H. Haefke, J. Frommer, E. Meyer, H.-J. Güntherodt, S. Hild, J. Fuhrmann, 'An Atomic Force Microscopy Study of Corona-treated Polypropylene Films', *Appl. Surf. Sci.*, **64**, 197–203 (1993).
65. T. Han, J.M. Williams, T.P. Beebe, Jr, 'Chemical Bonds Studied with Functionalized Atomic Force Microscopy Tips', *Anal. Chim. Acta*, **307**, 365–376 (1995).
66. R.J. Colton, E.I. Altman, S.M. Hues, 'Surface Science at the Nanoscale: Molecular Imaging and Surface Forces', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum, New York & London, 311–319, 1994.
67. T. Arai, M. Fujihira, 'Effect of Tip Shape on Force–Distance Curves for AFM in Aqueous Electrolytes', *J. Electroanal. Chem.*, **374**, 269–273 (1994).
68. X.Y. Lin, F. Creuzet, H. Arribart, 'Atomic Force Microscopy for Local Characterization of Surface Acid–Base Properties', *J. Phys. Chem.*, **97**, 7272–7276 (1993).
69. F. Ohnesorge, G. Binnig, 'True Atomic Resolution by Atomic Force Microscopy through Repulsive and Attractive Forces', *Science*, **260**, 1451–1456 (1993).
70. J.B. Pethica, A.P. Sutton, 'Nanomechanics: Atomic Resolution and Frictional Energy Dissipation in Atomic Force Microscopy', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 353–366, 1995.
71. H.A. Mizes, K.-G. Loh, R.J.D. Miller, S.K. Ahuja, E.F. Grabowski, 'Submicron Probe of Polymer Adhesion with Atomic Force Microscopy: Dependence on Topography and Material Inhomogeneities', *Appl. Phys. Lett.*, **59**, 2901–2903 (1991).
72. G.I. Rudd, S.H. Garofalini, D.A. Hensley, C.M. Mate, 'Atomic Force Microscopy and X-ray Photoelectron Spectroscopy Investigation of the Onset of Reactions on Alkali Silicate Glass Surfaces', *J. Am. Ceram. Soc.*, **76**, 2555–2560 (1993).
73. J.L. Loubet, M. Belin, R. Durand, H. Pascal, 'Triboscopic Description of Local Wear under an AFM Tip', *Thin Solid Films*, **253**, 194–198 (1994).
74. A.L. Weisenhorn, P.K. Hansma, T.R. Albrecht, C.F. Quate, 'Forces in Atomic Force Microscopy in Air and Water', *Appl. Phys. Lett.*, **54**, 2651–2653 (1989).
75. U. Hartmann, 'Theory of Non-contact Force Microscopy', in *Scanning Tunneling Microscopy III*, eds. R. Wiesendanger, H.-J. Güntherodt, Springer Series in Surface Sciences, 29, Springer, Berlin, 293–360, 1993.
76. M.G. Heaton, C.B. Prater, K.J. Kjoller, 'Lateral and Chemical Force Microscopy Map Surface Friction', *Adv. Mater. Processes*, **149**, 27–28 (1996).
77. R. Kaneko, 'Microtribology Today and Tomorrow', *Wear*, **168**, 1–5 (1993).
78. R. Kaneko, M. Igarashi, T. Miyamoto, E. Hamada, 'Mechanical Property Evaluations of Solid Surfaces as a Technological Application of SPM', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, Kluwer Academic, Dordrecht, 405–424, 1995.
79. M.C. Mate, 'Nanotribology Studies of Carbon Surfaces by Force Microscopy', *Wear*, **168**, 17–20 (1993).
80. M. Binggeli, 'Neue Wege in der Tribologieforschung', *Oberflächen Werkst.*, **5**, 8–10 (1993).
81. B. Bhushan, S. Sundararajan, 'Micro/nanoscale Friction and Wear Mechanisms of Thin Films Using Atomic Force Microscopy and Friction Force Microscopy', *Acta Mater.*, **46**, 3793–3804 (1998).
82. H. Hölscher, U.D. Schwarz, R. Wiesendanger, 'Modelling of the Scan Process in Lateral Force Microscopy', *Surf. Sci.*, **375**, 395–402 (1997).
83. D. Tománek, 'Theory of Atomic-scale Friction', in *Scanning Tunneling Microscopy III*, eds. R. Wiesendanger,

H.-J. Güntherodt, Springer Series in Surface Sciences, 29, Springer, Berlin, 269–292, 1993.
84. Omicron, 'Combined UHV STM/AFM', Data sheet V05/06.94, Omicron, Taunusstein, 1994.
85. H.K. Wickramasinghe, 'Raster-Sonden-Mikroskopie', *Spektrum Wisse.*, December, 62–70 (1989).
86. R. Pérez, M.C. Payne, I. Stich, K. Terakura, 'Contrast Mechanism in Non-contact AFM on Reactive Surfaces', *Appl. Surf. Sci.*, **123/124**, 249–254 (1998).
87. K. Wickramasinghe, 'Extensions of STM', in *Scanning Tunneling Microscopy*, eds. J.A. Stroscio, W.J. Kaiser, Methods of Experimental Physics, 27, Academic Press, San Diego, 77–94, 1993.
88. C.B. Prater, 'Vibrating Probe (AC) Methods in Atomic Force Microscopy', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York, 447–450, 1994.
89. M.-H. Whangbo, G. Bar, R. Brandsch, 'Description of Phase Imaging in Tapping™ Mode Atomic Force Microscopy by Harmonic Approximation', *Surf. Sci.*, **411**, L794–801 (1998).
90. I. Schmitz, M. Schreiner, G. Friedbacher, M. Grasserbauer, 'Phase Imaging as an Extension to Tapping™ Mode AFM for the Identification of Material Properties on Humidity Sensitive Surfaces', *Appl. Surf. Sci.*, **115**, 190–198 (1997).
91. I. Schmitz, M. Schreiner, G. Friedbacher, M. Grasserbauer, 'Tapping™-mode AFM in Comparison to Contact-mode AFM as a Tool for In Situ Investigations of Surface Reactions with Reference to Glass Corrosion', *Anal. Chem.*, **69**, 1012–1018 (1997).
92. W. Kiridena, V. Jain, P.K. Kuo, G. Liu, 'Nanometer-scale Elasticity Measurements on Organic Monolayers Using Scanning Force Microscopy', *Surf. Interface Anal.*, **25**, 383–389 (1997).
93. P. Maivald, H.J. Butt, S.A.C. Gould, C.B. Prater, B. Drake, J.A. Gurley, V.B. Elings, P.K. Hansma, 'Using Force Modulation to Image Surface Elasticities with the Atomic Force Microscope', *Nanotechnology*, **2**, 103–106 (1991).
94. L.E. Walther, N.E. Israeloff, 'Mesoscopic Dielectric and Structural Fluctuations at the Glass Transition', presented at the 14th University Glass Conference, Lehigh University, Bethlehem, PA, 1997.
95. M. Tosa, 'Measurement of Atomic Forces between Surface Atoms Using an Ultra-clean AFM Incorporated in an XHV Integrated System', *Adv. Colloid Interface Sci.*, **71/72**, 233–241 (1997).
96. T. Fries, 'Rastersondenmikroskopie', *Metalloberfläche – Beschicht. Metall Kunststo.*, **50**, 322–325 (1996).
97. M.B. Shattuck, M.G.L. Gustafsson, K.A. Fisher, K.C. Yanagimoto, A. Veis, R.S. Bhatnagar, J. Clarke, 'Monomeric Collagen Imaged by Cryogenic Force Microscopy', *J. Micros.*, **174**, RP1–RP2 (1994).
98. F.J. Giessibl, G. Binnig, 'Investigation of the (001) Cleavage Plane of Potassium Bromide with an Atomic Force Microscope at 4.2 K in Ultra-high Vacuum', *Ultramicroscopy*, **42–44**, 281–289 (1992).
99. S.A. Elrod, A. Bryant, A.L.d. Lozanne, S. Park, D. Smith, C.F. Quate, 'Tunneling Microscopy from 300 to 4.2 K', *IBM J. Res. Dev.*, **30**, 387–395 (1986).
100. P.K. Hansma, J. Tersoff, 'Scanning Tunneling Microscopy', *J. Appl. Phys.*, **61**, R1–R23 (1987).
101. C. Wünsche, 'Bruchvorgänge in Gläsern Untersucht mit dem Rasterkraftmikroskop', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1996.
102. I. Schmitz, 'In-situ Investigations of Corrosion Processes on Glass and Metal Surfaces by Scanning Probe Microscopy (SPM)', PhD Thesis, Technische Universität, Vienna, 1996.
103. P.M. Dove, F.M. Platt, 'Compatible Real-time Rates of Mineral Dissolution by Atomic Force Micrsocopy (AFM)', *Chem. Geol.*, **127**, 331–338 (1996).
104. Z. Shao, Y. Zhang, 'Biological Cryo Atomic Force Microscopy: A Brief Review', *Ultramicroscopy*, **66**, 141–152 (1996).
105. F. Grotelüschen, 'Legospiel mit Atomen und Molekülen', *VDI Nachri.*, **8.8**, 9 (1997).
106. H.K. Wickramasinghe, 'Related Scanning Techniques', in *Scanning Tunneling Microscopy II*, eds. R. Wiesendanger, H.-J. Güntherodt, Springer Series in Surface Sciences, 28, Springer, Berlin, 209–231, 1992.
107. R. Wiesendanger, 'Physik in Unserer Zeit auf der Nanometerskala', *PhiuZ*, **26**, 206–216 (1995).
108. S.R. Snyder, H.S. White, 'Scanning Tunneling Microscopy, Atomic Force Microscopy, and Related Techniques', *Anal. Chem.*, **64**, 116R–134R, 1992.
109. J. Fricke, 'Folgeentwicklungen des Rastertunnelmikroskops', *PhiuZ*, **21**, 219–221 (1990).
110. H. Rohrer, 'Scanning Tunneling Microscopy: A Surface Science Tool and Beyond', *Surf. Sci.*, **299/300**, 956–964 (1994).
111. B. Müller, 'Das Schweizer Messer', *Bild Wiss.*, **4**, 62–65 (1998).
112. J. Zarzycki, 'Near-field Microscopies in the Study of Glass', *J. Non-Cryst. Solids*, **196**, 7–15 (1996).
113. L.L. Soethout, H.v. Kempen, G.F.A.v.d. Walle, 'Scanning Tunneling Microscopy: A Mature Surface-science Technique', *Adv. Electron. Electron Phys.*, **79**, 155–269 (1990).
114. N.X. Randall, R. Christoph, S. Droz, C. Julia-Schmutz, 'Localised Micro-hardness Measurements with a Combined Scanning Force Microscope/Nanoindentation System', *Thin Solid Films*, **290/291**, 348–354 (1996).
115. D. Scholl, M.P. Everson, R.C. Jaklevic, 'Measurement of Surface Topography and Area-specific Nanohardness in the Scanning Force Microscope', *J. Mater. Res.*, **10**, 2503–2506 (1995).
116. D.W. Pohl, 'Nano-optics and Scanning Near-field Optical Microscopy', in *Scanning Tunneling Microscopy II*,

eds. R. Wiesendanger, H.-J. Güntherodt, Springer Series in Surface Sciences, 28, Springer, Berlin, 233–271, 1992.
117. U. Dürig, D. Pohl, F. Rohner, 'Near-field Optical Scanning Microscopy with Tunnel-distance Regulation', *IBM J. Res. Dev.*, **30**, 478–483 (1986).
118. J. Samuel, A. Strinkovski, S. Shalom, K. Lieberman, M. Ottolenghi, D. Avnir, A. Lewis, 'Miniaturization of Organically Doped Sol–Gel Materials: A Microns-size Fluorescent pH Sensor', *Mater. Lett.*, **21**, 431–434 (1994).
119. A. Piednoir, F. Creuzet, C. Licoppe, J.M. Ortéga, 'Locally Resolved Infrared Spectroscopy', *Ultramicroscopy*, **57**, 282–286 (1995).
120. A.K. Kirsch, C.K. Meyer, T.M. Jovin, 'Shear Force Imaging of DNA in a Near-field Scanning Optical Microscope (NSOM)', *J. Microsco.*, **185**, 396–401 (1997).
121. U. Fischer, 'Resolution and Contrast Generation in Scanning Near Field Optical Microscopy', in *Scanning Tunneling Microscopy and Related Methods*, eds. R.J. Behm, N. Garcia, H. Rohrer, NATO ASI Series E: Applied Sciences, 184, Kluwer Academic, Dordrecht, 475–496, 1990.
122. K. Nakajima, Y. Mitsuoka, N. Chiba, H. Muramatsu, T. Ataka, K. Sato, M. Fujihira, 'Polarization Effect in Scanning Near-field Optic/Atomic-force Microscopy (SNOM/AFM)', *Ultramicroscopy*, **71**, 257–262 (1998).
123. S. Ushioda, 'Characterization of Surface Nanostructures by STM Light Emission Spectroscopy', *Appl. Surf. Sci.*, **113/114**, 335–342 (1997).
124. A. Hammiche, H. Pollock, 'Scanning Thermal Microscopy and Calorimetry', *Appl. Newsl.*, **96-1**, 1–3 (1996).
125. J.K. Gimzewski, C. Gerber, E. Meyer, R.R. Schlittler, 'Micromechanical Heat Sensor: Observation of a Chemical Reaction, Photon and Electrical Heat Pulses', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, Kluwer Academic, Dordrecht, 123–131, 1995.
126. W. Häberle, D.P.E. Smith, J.K.H. Hörber, J.P. Czerny, 'Scanning Force Microscopy on Living Virus-infected Cells', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York & London, 7–17, 1994.
127. B. Bhushan, 'Nanotribology and its Application to Magnetic Storage Devices and MEMS', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 367–395, 1995.
128. J.P. Rabe, 'Imaging Surfaces, Electric Charges and Magnetic Domains with the Atomic Force Microscope', *Angew. Chem. Adv. Mater.*, **101**, 1610–1613 (1989).
129. S.P. Jarvis, J.B. Pethica, 'Hydrophobic Surface Interactions Studied Using a Novel Force Microscope', in *Forces in Scanning Probe Methods*, eds. H.-J. Güntherodt, D. Anselmetti, E. Meyer, NATO ASI Series E: Applied Sciences, 286, Kluwer Academic, Dordrecht, 105–112, 1995.
130. A. Kramer, F. Keilmann, B. Knoll, R. Guckenberger, 'The Coaxial Tip as a Nano-antenna for Scanning Near-field Microwave Transmission Microscopy', *Micron*, **27**, 413–417 (1996).
131. V. Drach, 'Nahfeldoptik mit Nanometerauflösung', *PhiuZ*, **27**, 34 (1996).
132. P. Moyer, D. Higgins, P. Barbara, K. Walzer, 'Near-field Scanning Optical Microscopy Offers Solutions to Scientific Problems', *Appl. Newsl.*, **95-1**, 1–2 (1995).
133. M. Mertesdorf, M. Schönhoff, F. Lohr, S. Kirstein, 'Scanning Near-field Optical Microscope Designed for Operation in Liquids', *Surf. Interface Anal.*, **25**, 755–759 (1997).
134. A. Piednoir, C. Licoppe, F. Creuzet, 'Imaging and Local Infrared Spectroscopy with a Near Field Optical Microscope', *Opt. Commun.*, **129**, 414–422 (1996).
135. S. Ragona, P. Haydon, 'Near-field Optical Microscopy: A Breakthrough in Resolution', *Appl. Newsl.*, **96-1**, 5 (1996).
136. J. Baumann, G. Fritsch, 'Das Ultraschall-Raster-Mikroskop', *PhiuZ*, **19**, 16–28 (1988).
137. S.L. Flegler, J.W.J. Heckman, K.L. Klomparens, *Scanning and Transmission Electron Microscopy*, Oxford University Press, Oxford, 1993.
138. S.L. Flegler, J.W.J. Heckman, K.L. Klomparens, *Elektronenmikroskopie: Grundlagen – Methoden – Anwendungen*, Spektrum Akademischer Verlag, Heidelberg, 1995.
139. L. Reimer (ed.), *Scanning Electron Microscopy*, Springer Series in Optical Sciences, 45, Springer, Berlin, 1998.
140. S. Amelinckx, D. van Dyk, J. van Landuyt, G.G. Tendeloo (eds.), *Handbook of Microscopy*, VCH, Weinheim, 1992.
141. A. Bauer, 'Röntgenmikroskopie', *PhiuZ*, **23**, 151–158 (1992).
142. M. Hoppe, A. Thaer, 'Scanning Acoustic Microscopy', in *Microscopic Methods in Metals*, ed. U. Gonser, Springer, Berlin, 7–28, 1986.
143. J.P. Roger, P. Gleyzes, H. ElRhaleb, D. Fournier, A.C. Boccara, 'Optical and Thermal Characterization of Coatings', *Thin Solid Films*, **261**, 132–138 (1995).
144. J.M. Bennett, L. Mattsson, *Introduction to Surface Roughness and Scattering*, Optical Society of America, Washington, DC, 1989.
145. J.M. Bennett, J. Jahamir, J.C. Podlesny, T.L. Balter, D.T. Hobbs, 'Scanning Force Microscope as a Tool for Studying Optical Surfaces', *Appl. Opt.*, **34**, 213–230 (1995).
146. J. Engelhardt, W. Knebel, 'Konfokale Laserscanning-mikroskopie', *PhiuZ*, **24**, 70–78 (1993).
147. N.M. Brown, H.X. You, 'Tyranno (Si–Ti–C–O) Fibres', *J. Mater. Sci.*, **28**, 121–128 (1993).
148. U. Gonser (ed.), *Microscopic Methods in Metals*, Springer, Berlin, 1986.

149. K. Wetzig, D. Schulze (eds.), *In Situ Scanning Electron Microscopy in Materials Research*, Akademie Verlag, Berlin, 1995.
150. J. Hopfe, M. Füting, 'Fundamentals and Applications of Environmental Scanning Electron Microscopy', in *In Situ Scanning Electron Microscopy in Materials Research*, eds. K. Wetzig, D. Schulze, Akademie Verlag, Berlin, 219–240, 1995.
151. M. Spitzer-Aronson, 'A New Parallel Research on Glass with Ion Imaging and Electron Microprobe', *Proceedings of XVII International Congress on Glass, Peking, China*, ICG, 544–547, 1995.
152. A. Hattori, 'Measurement of Glass Surface Contamination', *J. Non-Cryst. Solids*, **218**, 196–204 (1997).
153. N. Ernst, G. Ehrlich, 'Field Ion Microscopy', in *Microscopic Methods in Metals*, ed. U. Gonser, Springer, Berlin, 75–116, 1986.
154. J.M. Bennett, 'Precision Metrology for Studying Thin Films and Surfaces', *Thin Solid Films*, **220**, 227–233 (1992).
155. C.Y. Poon, B. Bhushan, 'Comparison of Surface Roughness Measurements by Stylus Profilometer, AFM and Non-contact Optical Profiler', *Wear*, **190**, 76–88 (1995).
156. H. Rothe, A. Duparré, S. Jacobs, 'Generic Detrending of Surface Profiles', *Opt. Eng.*, **33**, 3023–3030 (1994).
157. Digital Instruments, 'NanoScope III User's Manual Version 2.5 and Additions to Command Reference Manual, especially Version 2.53', Digital Instruments, Mannheim, 1992.
158. K.H. Guenther, J.M. Bennett, F.K. Urban, III, M.F. Tabet, M. van Gunten, R. Bevis, M.M. Tehrani, 'Comparative Study of the Surface of Oxide Films', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York, 281–299, 1994.
159. B. Bhushan, L. Yang, C. Gao, S. Suri, R.A. Miller, 'Friction and Wear Studies of Magnetic Thin-film Rigid Disks with Glass–Ceramic, Glass and Aluminium–Magnesium Substrates', *Wear*, **190**, 44–59 (1995).
160. C.Y. Poon, B. Bhushan, 'Surface Roughness Analysis of Glass–Ceramic Substrates and Finished Magnetic Disks, and Ni–P Coated Al–Mg and Glass Substrates', *Wear*, **190**, 89–109 (1995).
161. A. Duparré, S. Jakobs, 'Combination of Surface Characterization Techniques for Investigating Optical Thin-film Components', *Appl. Opt.*, **35**, 5052–5058 (1996).
162. D. Abriou, Y. Levillain, H. Arribart, F. Creuzet, 'Analysis of the Roughness of Glass Surfaces: From the Nanoscopic to the Macroscopic Scale', *International Symposium on Glass Problems, Istanbul, Turkey*, 427–433, 1996.
163. H. Truckenbrodt, A. Duparré, U. Schuhmann, 'Roughness and Defect Characterisation of Optical Surfaces by Light Scattering Measurements', *Proc. SPIE*, **1781**, 139–151 (1992).
164. V.E. Asadchikov, A. Duparré, S. Jakobs, A.Y. Karabekov, I.V. Kozhevnikov, Y.S. Krivonosov, 'Comparative Study of the Roughness of Optical Surfaces and Thin Films by Use of X-ray Scattering and Atomic Force Microscopy', *Appl. Opt.*, **38**, 684–691 (1999).
165. M. Stedman, 'Limits of Topographic Measurement by the Scanning Tunnelling and Atomic Force Microscopes', *J. Microsc.*, **152**, 611–618 (1988).
166. A. Borne, B. Chevalier, J.L. Chevalier, D. Quenard, E. Elaloui, J. Lambard, 'Characterization of Silica Aerogel with the Atomic Force Microscope and SAXS', *J. Non-Cryst. Solids*, **188**, 235–242 (1995).
167. H. Dunken, *Physikalische Chemie der Glasoberfläche*, VEB Deutscher Verlag für Grundstoffindustrie, Leipzig, 1981.
168. I. Schmitz, T. Prohaska, G. Friedbacher, M. Schreiner, M. Grasserbauer, 'Investigation of Corrosion Processes on Cleavage Edges of Potash–Lime–Silica Glasses by Atomic Force Microscopy', *Fresenius' J. Anal. Chem.*, **353**, 666–669 (1995).
169. H. Lüth, *Surfaces and Interfaces of Solid Materials*, Springer, Berlin, 1995.
170. T.M. Schaub, D.E. Bürgler, C.M. Schmidt, H.-J. Güntherodt, 'Investigation of Non-crystalline Surfaces by Scanning Tunneling Microscopy', *J. Non-Cryst. Solids*, **205–207**, 748–754 (1996).
171. R. Ambos, E. Rädlein, G.H. Frischat, 'SNMS and AFM: Combination of Two Surface Sensitive Methods for the Characterisation of Oxide Layers and Coatings', *Glastech. Ber. Glass Sci. Technol.*, **68**(C1), 351–359 (1995).
172. E. Rädlein, R. Ambos, G.H. Frischat, 'Atomic Force Microscopy of Coated Glasses', *Fresenius' J. Anal. Chem.*, **353**, 413–418 (1995).
173. E. Rädlein, C. Wünsche, G.H. Frischat, 'Atomic Force Microscopy Investigations on Vitreous Coatings and Glass Fracture Surfaces', *Proceedings of XVII International Congress on Glass, Peking, China*, ICG, 3–8, 1995.
174. E. Rädlein, J. Liu, R. Hauk, G.H. Frischat, 'The Nanostructure of Sol–Gel Coatings on Glass Investigated by Atomic Force Microscopy (AFM)', *Proceeding of XVIII International Congress on Glass, San Francisco, CA, USA*, ICG, ICG656-E609-006, 1998.
175. E. Rädlein, G.H. Frischat, 'Nanotopographie von Gläsern und Beschichtungen, Untersucht Mittels Atomkraftmikroskopie', *Kurzreferate der 72. Glastechnische Tagung der DGG (Vorträge), Münster*, Verlag der Deutschen Glastechnischen Gesellschaft, 281–284, 1998.
176. G. Helsch, E. Rädlein, G.H. Frischat, 'Poröses SiO_2 als Antireflexionsschicht', *Kurzreferate der 72. Glastechnischen Tagung der DGG (Vorträge), Münster*, Verlag der Deutschen Glastechnischen Gesellschaft, 215–218, 1998.
177. J. Liu, 'Sol–Gel-Herstellung und Charakterisierung von ITO-Schichten', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1998.

178. O. Becker, K. Bange, 'Ultramicrotomy: An Alternative Cross-section Preparation for Oxidic Thin Films on Glass', *Ultramicroscopy*, **52**, 73–84 (1993).
179. K. Wilder, C. Quate, B. Singh, R. Alvis, W.H. Arnold, 'Atomic Force Microscopy for Cross-section Inspection and Metrology', *J. Vac. Sci. Technol. B*, **14**, 4004–4008 (1996).
180. A. Duparré, N. Kaiser, H. Truckenbrodt, M. Berger, A. Köhler, 'Microtopography Investigations of Optical Surfaces and Thin Films by Light Scattering, Optical Profilometry, and Atomic Force Microscopy', *Proc. SPIE*, **1995**, 181–192 (1993).
181. A. Duparré, C. Ruppe, K.A. Pischow, M. Adamik, P.B. Barna, 'Atomic Force Microscopy on Cross-sections of Optical Coatings: A New Method', *Thin Solid Films*, **261**, 70–75 (1995).
182. K.A. Pischow, M. Adamik, P.B. Barna, A.S. Korhonen, 'Cross-section X-scanning Force Microscopy Analysis of Different Hard Coatings and Thin Films', *J. Vac. Sci. Technol. B*, **12**, 1716–1721 (1994).
183. F. Oberlies, 'Über Ultramikrotom-schnitte von Silikatgläsern', *Glastechn. Ber.*, **37**, 122–125 (1964).
184. C. Quintana, 'Ultramicrotomy for Cross-sections of Nanostructure', *Micron*, **28**, 217–219 (1997).
185. H. Bach, 'Advanced Surface Analysis of Silicate Glasses, Oxides and Other Insulation Materials: A Review', *J. Non-Cryst. Solids*, **209**, 1–18 (1997).
186. W. Vogel, *Glaschemie*, Springer, Berlin, 1992.
187. P.K. Hansma, V.B. Elings, O. Marti, C.E. Bracker, 'Scanning Tunneling Microscopy and Atomic Force Microscopy: Application to Biology and Technology', *Science*, **242**, 209–216 (1988).
188. J.T. Woodward, J.A. Zasadzinski, 'Scanning Tunneling Microscopy of Freeze Fracture Replicas of Biomaterials', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York & London, 339–345, 1994.
189. R. Vollmers, 'Charakterisierung leitfähiger Schichten auf Gläsern und leitfähiger Gläser mit Hilfe der Rastersondenmikroskopie', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1996.
190. T. Prohaska, G. Friedbacher, M. Grasserbauer, 'In-situ Investigation of Surface Reactions on Alkali Halides by Atomic Force Microscopy', *Fresenius' J. Anal. Chem.*, **349**, 190–194 (1994).
191. K. Techmer, E. Rädlein, 'Experimentelle In-situ Beobachtungen von Alterationsvorgängen an natürlichen Glasoberflächen mit Hilfe der AFM-fluidzelle', *Kurzreferate der 72. Glastechnischen Tagung der DGG (Vorträge), Münster*, Verlag der Deutschen Glastechnischen Gesellschaft, 289–292, 1998.
192. A. Czymai, 'Eine Methode zur Bestimmung von Tiefenprofilen an korrodierten Glasfasern', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1995.
193. M.M. Curtin Carter, N.S. McIntyre, R. Davidson, H.W. King, 'Imaging Fibers by Atomic Force Microscopy', *J. Vac. Sci. Technol. B*, **14**, 1867–1869 (1996).
194. S. Röder, 'Probenpräparation zur Schichtdickenbestimmung und zur Rauhigkeitsmessung mit dem atomaren Kraftmikroskop', Studienarbeit, TU Clausthal, Clausthal-Zellerfeld, 1997.
195. P.A. Zhdan, D. Grey, J.E. Castle, 'Surface Structure of the PAN-based Carbon Fibres Studied by Scanning Probe Microscopy', *Surf. Interface Anal.*, **22**, 290–295 (1994).
196. P.W. de Bont, P.M.L.O. Scholte, M.H.J. Hottenhuis, G. Kempen, van, J.W. Kerssemakers, F. Tuinstra, 'Atomic Structure of Longitudinal Sections of a Pitch-based Carbon Fiber Studied by STM', *Appl. Surf. Sci.*, **74**, 73–80 (1994).
197. L.A. Titcombe, M.G. Huson, P.S. Turner, 'Imaging Internal Cellular Structure of Merino Wool Fibres Using Atomic Force Microscopy', *Micron*, **28**, 69–71 (1997).
198. G.C. Rutledge, D. Snévety, G.J. Vansco, 'Studies of High Performance Fibers by Atomic Force Microscopy and Molecular Simulation', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York & London, 251–263, 1994.
199. E. Olbrich, 'Charakterisierung von Hüttensandglas hinsichtlich Struktur und Reaktivität', Diploma Thesis, TU Clausthal, Clausthal-Zellerfeld, 1997.
200. A.A. Garcia, M. Edwards, 'Sample Holders for Imaging Intact Particles with the Scanning Force Microscope', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York & London, 327–338, 1994.
201. R.S. Robinson, H.H. Yuce, 'Scanning Tunneling Microscopy of Optical Fiber Corrosion: Surface Roughness Contribution to Zero-stress Aging', *J. Am. Ceram. Soc.*, **74**, 314–318 (1991).
202. R.C. Jaklevic, L. Elie, W. Shen, J.T. Chen, 'Ultrathin Au Films Studied with the Scanning Tunneling Microscope', *Appl. Phys. Lett.*, **52**, 1656–1657 (1988).
203. R.C. Jaklevic, L. Elie, W. Shen, J.T. Chen, 'Application of the Scanning Tunneling Microscope to Insulating Surfaces', *J. Vac. Sci. Technol. A*, **6**, 448–453 (1988).
204. T.G. Lenihan, A.P. Malshe, W.D. Brown, L.P. Schaper, 'Artifacts in SPM Measurements of Thin Films and Coatings', *Thin Solid Films*, **270**, 356–361 (1995).
205. C.S. Lindquist, F.K. Urban, 'Surface Analysis Algorithms for Scanning Probe Microscopy', *Thin Solid Films*, **270**, 399–405 (1995).
206. E. Stoll, O. Marti, 'Restoration of Scanning-tunneling-microscope Data Blurred by Limited Resolution, and Hampered by $1/f$-like Noise', *Surf. Sci.*, **181**, 222–229 (1987).

207. S.M. Shevchenko, G.W. Bailey, Y.S. Yu, 'Nanoscapes: Coupled SPM–VR Analysis of Environmental Surfaces', *Am. Lab.*, **29**, 18 (1997).
208. P.E. West, R. Jobe, T. Van Slambrouck, 'Linearity and Calibration of Scanning Probe Microscope Images', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York & London, 321–325, 1994.
209. S. Viera, 'The Behavior and Calibration of Some Piezoelectric Ceramics Used in the STM', *IBM J. Res. Dev.*, **30**, 553–556 (1986).
210. K.L. Westra, A.W. Mitchell, D.J. Thomson, 'Tip Artifacts in Atomic Force Microscope Imaging of Thin Film Surfaces', *J. Appl. Phys.*, **74**, 3608–3610 (1993).
211. K.L. Westra, D.J. Thomson, 'Atomic Force Microscope Tip Radius Needed for Accurate Imaging of Thin Film Surfaces', *J. Vac. Sci. Technol. B*, **12**, 3176–3181 (1994).
212. J.K. Gimzewski, A. Humbert, 'Scanning Tunneling Microscopy of Surface Microstructure on Rough Surfaces', *IBM J. Res. Dev.*, **30**, 472–477 (1986).
213. M. Van Cleef, S.A. Holt, G.S. Watson, S. Myhra, 'Polystyrene Spheres on Mica Substrates: AFM Calibration, Tip Parameters and Scan Artefacts', *J. Microsc.*, **181**, 2–9 (1996).
214. G. Kaupp, J. Schmeyers, U. Pogoda, M. Haak, T. Marquardt, M. Plagmann, 'AFM Imaging of Large and Steep Submicroscopic Features, Artifacts and Scraping with Asymmetric Cantilever Tips', *Thin Solid Films*, **264**, 205–211 (1995).
215. R. Cassidy, '7 Ways to Weed Out Artifacts in SPM Images', *Res. Dev.*, **36**, 93–94 (1994).
216. H.A. Mizes, S. Park, W.A. Harrison, 'Multiple-tip Interpretation of Anomalous Scanning-tunneling-microscopy Images of Layered Materials', *Phys. Rev. B*, **36**, 4491–4494 (1987).
217. H. Tang, C. Joachim, J. Devillers, 'Interpretation of AFM Images: The Graphite Surface with a Diamond Tip', *Surf. Sci.*, **291**, 439–450 (1993).
218. A. Kurshudov, K. Kato, 'Wear of the Atomic Force Microscope Tip Under Light Load, Studied by Atomic Force Microscopy', *Ultramicroscopy*, **60**, 11–16 (1995).
219. T. Michely, K.H. Besocke, M. Teske, 'A Combined Scanning Tunnelling and Field Ion Microscope', *J. Microsc.*, **152**, 77–83 (1988).
220. F. Jensen, '*z* Calibration of the Atomic Force Microscope by means of a Pyramidal Tip', *Rev. Sci. Instrum.*, **64**, 2595–2597 (1993).
221. F. Atamny, A. Baiker, 'Direct Imaging of the Tip Shape by AFM', *Surf. Sci.*, **323**, L314–L318 (1995).
222. C. Odin, J.P. Aimé, Z. ElKaakour, T. Bouhacina, 'Tip's Finite Size Effects on Atomic Force Microscopy in the Contact Mode: Simple Geometrical Considerations for Rapid Estimation of Apex Radius and Tip Angle Based on the Study of Polystyrene Latex Balls', *Surf. Sci.*, **317**, 321–340 (1994).
223. S. Xu, M.F. Arnsdorf, 'Calibration of the Scanning (Atomic) Force Microscope with Gold Particles', *J. Microsc.*, **173**, 199–210 (1993).
224. L. Montelius, J.O. Tegenfeld, P.v. Heeren, 'Direct Observation of the Atomic Force Microscopy Tip Using Inverse Atomic Force Microscopy Imaging', *J. Vac. Sci. Technol. B*, **12**, 2222–2226 (1994).
225. S. Hamelmann, 'Aufbau eines Spitzenätzplatzes für die Raster-Tunnel-Mikroskopie', Studienarbeit, TU Clausthal, Clausthal-Zellerfeld, 1993.
226. L. Anselm, 'Charakterisierung von SnO_2-beschichteten Glasoberflächen mit dem Raster-Tunnel-Mikroskop', Diploma Thesis, TU Clausthal, Clausthal-Zellerfeld, 1993.
227. J. Schneir, J.S. Villarrubia, T.H. McWaid, V.W. Tsai, R. Dixson, 'Increasing the Value of Atomic Force Microscopy Process Metrology Using a High-accuracy Scanner, Tip Characterization, and Morphological Image Analysis', *J. Vac. Sci. Technol. B*, **14**, 1540–1546 (1996).
228. J.M. Gómez-Rodriguez, L. Vázquez, A.M. Baró, 'Application of an STM/SEM Instrument for the Study of Materials: Tip Effects and Data Correlation', *Surf. Interface Anal.*, **16**, 97–104 (1990).
229. R.P. Van Duyne, J.C. Hulteen, D.A. Treichel, 'Atomic Force Microscopy and Surface-enhanced Raman Spectroscopy. I. Ag Island Films and Ag Film Over Polymer Nanosphere Surfaces Supported on Glass', *J. Chem. Phys.*, **99**, 2101–2115 (1993).
230. S.E. Roark, K.L. Rowlen, 'Atomic Force Microscopy of Thin Ag Films', *Chem. Phys. Lett.*, **212**, 50–57 (1993).
231. L.L. Smith, E. Srinivasan, G.N. Parsons, 'Investigations of Substrate-dependent Nucleation of Plasma-deposited Microcrystalline Silicon on Glass and Silicon Substrates Using Atomic Force Microscopy', *J. Appl. Phys.*, **82**, 6041–6046 (1997).
232. F. Creuzet, 'Atomic Force Microscopy Study of Glass Surface and Interface', *Proceedings of 4th International Conference on Joining Ceramics, Glass and Metal, Königswinter, Germany*, Verlag der Deutschen Glastechnischen Gesellschaft, 28–35, 1993.
233. H.W. Werner, R.P.H. Garten, 'A Comparative Study of Methods for Thin-film and Surface Analysis', *Rep. Prog. Phys.*, **47**, 221–344 (1984).
234. J. Vancea, G. Reiss, F. Schneider, K. Bauer, H. Hoffmann, 'Substrate Effects on the Surface Topography of Evaporated Gold Films – A Scanning Tunnelling Microscopy Investigation', *Surf. Sci.*, **218**, 108–126 (1989).
235. G. Binnig, H. Rohrer, 'Das Raster-Tunnelmikroskop', *Spektrum Wiss.*, **12**, 62–68 (1985).
236. O. Marti, B. Drake, S. Gould, P.K. Hansma, 'Atomic Resolution Atomic Force Microscopy of Graphite and the ''Native Oxide'' on Silicon', *J. Vac. Sci. Technol. A*, **6**, 287–290 (1988).

237. T. Bertrams, A. Brodde, H. Hannemann, C.A. Ventrice, G. Wilhelmi, H. Neddermeyer, 'STM of Manipulated Structures: Characterization of Metal Oxide Films', *Appl. Surf. Sci.*, **75**, 125–132 (1994).
238. M.C. Gallagher, M.S. Fyfield, J.P. Cowin, S.A. Joyce, 'Imaging Insulating Oxides: Scanning Tunneling Microscopy of Ultrathin MgO Films on Mo(001)', *Surf. Sci.*, **339**, L909–L913 (1995).
239. G. Friedbacher, E. Bouveresse, G. Fuchs, M. Grasserbauer, D. Schwarzenbach, R. Haubner, B. Lux, 'Pretreatment of Silicon Substrates for CVD Diamond Deposition Studied by Atomic Force Microscopy', *Appl. Surf. Sci.*, **84**, 133–143 (1995).
240. R. Banga, J. Yarwood, A.M. Morgan, B. Evans, J. Kells, 'FTIR and AFM Studies of the Kinetics and Self-assembly of Alkyltrichlorosilanes and (Perfluoroalkyl) trichlorosilanes onto Glass and Silicon', *Langmuir*, **11**, 4393–4399 (1995).
241. T. Vallant, H. Brunner, U. Mayer, H. Hoffmann, T. Leitner, R. Resch, G. Friedbacher, 'Formation of Self-assembled Octadecylsiloxane Monolayers on Mica and Silicon Surfaces Studied by Atomic Force Microscopy and Infrared Spectroscopy', *J. Phys. Chem.*, **B102**, 7190–7197 (1998).
242. G. Mezinskis, G.H. Frischat, E. Rädlein, 'Micro- and Nanostructures of $Fe_2O_3 \times SiO_2$ Sol–Gel-derived Thick Coatings', *J. Sol–Gel Sci. Technol.*, **8**, 489–492 (1997).
243. G. Mezinskis, 'Thick Sol–Gel Coating Preparation by Generation of In Situ Metallic Particles', *Proceedings of XVI International Congress on Glass, Madrid, Spain*, ICG, 321–326, 1992.
244. G. Mezinskis, G.H. Frischat, 'Preparation of SiO_2–TiO_2–ZrO_2 Gel Compositions by Means of a Liquid Crystal Approach', *Chim. Chron., New Ser.*, **23**, 157–162 (1994).
245. R. Ambos, 'Analyse von silicatischen Gläsern und oxidischen Beschichtungen auf Gläsern mit der Sekundärneutralteilchen-Massenspektrometrie', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1996.
246. J. Liu, E. Rädlein, G.H. Frischat, 'Preparation, Nanostructure and Properties of Indium-tin-oxide (ITO) Films on Glass Substrates. Part I: Preparation and Nanostructure. Part II: Optimization of Properties', *Phys. Chem. Glasses*, **40**, 277–286 (1999).
247. A. Crossley, C.J. Sofield, J.P. Goff, A.C.I. Lake, M.T. Hutchings, A. Menelle, 'A Study Comparing Measurements of Roughness of Silicon and SiO_2 Surfaces and Interfaces Using Scanning Probe Microscopy and Neutron Reflectivity', *J. Non-Cryst. Solids*, **187**, 221–226 (1995).
248. H. Rojhantalab, M. Moinpur, N. Peter, M.L.A. Dass, W. Hough, R. Natter, F. Moghadam, 'AFM Studies of Surface Roughness of Borophosphosilicate Glass (BPSG) Films and their Impact on Defect Detection Capability for Sub-micron VLSI Technology', *Materials Research Society Symposium*, eds. H.A. Atwater, E. Chason, M.H. Grabow, M.G. Lagally, Materials Research Society, Pittsburgh, PA, 147–152, 1993.
249. N. Yoshino, 'Atomic Force Microscopy of Glass Surface Modified with Silane Coupling Agent Containing Fluorocarbon Chain', *Chem. Lett.*, **4**, 735–736 (1994).
250. S. Priller, 'Frühstadien der Korrosion von Technischen Glasfasern', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1998.
251. I.A. Rauf, 'A Comparison of Scanning Tunnelling Microscopy with Conventional and Scanning Transmission Electron Microscopy Using Tin-doped Indium Oxide Thin Films', *Surf. Sci.*, **325**, L413–L419 (1995).
252. I.A. Rauf, F. Czerwinski, J.D. Boyd, J.A. Szpunar, 'Detection of Impurity Segregation in Zone-confined, Polycrystalline Tin-doped Indium Oxide Thin Films by STM and AFM', *Micron*, **26**, 565–569 (1995).
253. T. Asikainen, M. Ritala, M. Leskelä, T. Prohaska, G. Friedbacher, M. Grasserbauer, 'AFM and STM Studies on In_2O_3 and ITO Thin Film Deposited by Atomic Layer Epitaxy', *Appl. Surf. Sci.*, **99**, 91–98 (1996).
254. R. Vollmers, G.H. Frischat, 'Tunnelmikroskopie an leitfähig beschichteten Gläsern', *Kurzreferate der 69. Glastechnischen Tagung der DGG, Würzburg*, Verlag der Deutschen Glastechnischen Gesellschaft, 108–111, 1995.
255. R. Vollmers, E. Rädlein, G.H. Frischat, 'Nanotopography of Thin Films on Glasses Measured with AFM and STM', *Fundamentals of Glass Science and Technology, Växjö, Sweden*, ESG, 23–32, 1997.
256. K.L. Chopra, S. Major, D.K. Pandya, 'Transparent Conductors – A Status Review', *Thin Solid Films*, **102**, 1–46 (1983).
257. I. Hamberg, C.G. Granqvist, 'Evaporated Sn-doped In_2O_3 Films: Basic Optical Properties and Applications to Energy-efficient Windows', *J. Appl. Phys.*, **60**, R123–R159 (1986).
258. R. Hauk, 'Herstellung kratzfester Al_2O_3-Schichten nach dem Sol–Gel-Verfahren', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1997.
259. Y.-C. Lai, J.-C. Lin, C. Lee, 'Nucleation and Growth of Highly Oriented Lead Titanate Thin Films Prepared by a Sol–Gel Method', *Appl. Surf. Sci.*, **125**, 51–57 (1998).
260. E. Meyer, H. Haefke, H.-J. Güntherodt, O. Anderson, K. Bange, 'Surface Topography of Polished and Oxide-coated BK-7 Glasses Investigated by Atomic Force Microscopy', *Glastechn. Ber.*, **66**, 30–37 (1993).
261. K.H. Guenther, 'Nonoptical Characterization of Optical Coatings', *Appl. Opt.*, **15**, 3487–3502 (1981).
262. E. Stathatos, P. Lianos, F.d. Monte, D. Levy, D. Tsiourvas, 'Nanocrystalline Titanium Dioxide Films made by the Sol–Gel Method Using Reverse Micelles', *J. Sol–Gel Sci. Technol.*, **10**, 83–89 (1997).
263. W.R. Bowen, N. Hilal, R.W. Lovitt, P.M. Williams, 'Visualization of an Ultrafiltration Membrane by Non-contact Atomic Force Microscopy at Single Pore Resolution', *J. Membr. Sci.*, **110**, 229–232 (1996).

264. W.R. Bowen, N. Hilal, R.W. Lovitt, C.J. Wright, 'A New Technique for Membrane Characterization: Direct Measurement of the Force of Adhesion of a Single Particle Using an Atomic Force Microscope', *J. Membr. Sci.*, **139**, 269–274 (1998).
265. D. Beyer, H. Richter, G. Tomandl, 'AFM and STM Characterization of TiO_2 Ultrafiltration Membranes', *Surf. Interface Anal.*, **25**, 593–595 (1997).
266. G.B. Amisola, R. Behrensmeier, J.M. Galligan, F.A. Otter, F. Namavar, N.M. Kalkhoran, 'Scanning Tunneling Microscopy of Porous Silicon-based Surfaces', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, eds. S.H. Cohen, M.T. Bray, M.L. Lightbody, Plenum Press, New York, 181–188, 1994.
267. G. Helsch, E. Rädlein, G.H. Frischat, 'On the Origin of the Aging Process of Porous SiO_2 Antireflection Coatings', *J. Non-Cryst. Solids*, **265**, 193–197 (2000).
268. G. Helsch, 'Ursache der Alterung von porösen Sol–Gel-Antireflexionsschichten', PhD Thesis, TU Clausthal, Clausthal-Zellerfeld, 1999.
269. H.M. Kerch, F. Cosandey, R.A. Gerhardt, 'Imaging of Fine Porosity on a Colloidal Silica : Potassium Silica Gel by Defocus Contrast Microscopy', *J. Non-Cryst. Solids*, **152**, 18–31 (1993).
270. H. Kaji, K. Nakanishi, N. Soga, 'Formation of Porous Gel Morphology by Phase Separation in Gelling Alkoxy-derived Silica. Phenomenological Study', *J. Non-Cryst. Solids*, **185**, 18–30 (1995).
271. M.S.W. Vong, P.A. Sermon, 'Observing the Breathing of Silica Sol–Gel-derived Anti-reflection Optical Coatings', *J. Non-Cryst. Solids*, **293**, 185–195 (1997).
272. T. Yu, R. Laiho, L. Heikkila, 'Investigation of Porous Silicon by Scanning Tunneling Microscopy and Atomic Force Microscopy', *J. Vac. Sci. Technol. B*, **12**, 2437–2439 (1994).
273. N. Martin, C. Rousselot, D. Rondot, F. Palmino, R. Mercier, 'Microstructure Modification of Amorphous Titanium Oxide Thin Films during Annealing Treatment', *Thin Solid Films*, **300**, 113–121 (1997).
274. B.C. Oberländer, M. Frank, P. Campbell, J.P. Celis, J. Roos, L. Hellemans, J. Snauwaert, 'Comparative Investigation by Laser Profilometry, Scanning Electron Microscopy and Atomic Force Microscopy of Wear on Solar-beam-irradiated Partially Oxidized TiN Coatings', *Thin Solid Films*, **241**, 222–229 (1993).
275. R. Hauk, G.H. Frischat, K. Ruppert, 'Herstellung Kratzfester Al_2O_3-Schichten auf Glas nach dem Sol–Gel-Verfahren', *Glastech. Ber. Glass Sci. Technol.*, **71**, N114 (1998).
276. V. Snitka, A. Ulcinas, M. Rackaitis, D. Zukauskas, M. Fukui, 'Microtribological Properties of Diamond-like and Hydrogenated Carbon Coatings Grown by Different Methods', *Diamond Relat. Mater.*, **6**, 1–5 (1997).
277. F. Creuzet, G. Ryschenkow, H. Arribart, 'A New Tool for Adhesion Science: The Atomic Force Microscope', *J. Adhesion*, **40**, 15–25 (1992).
278. F. Creuzet, 'Near Field Microscopy of Glass Surfaces and Interfaces', *Microsc. Microanal. Microstruct.*, **5**, 11–18 (1994).
279. G. Ryschenkow, H. Arribart, 'Adhesion Failure in the Stick–Slip Regime: Optical and AFM Characterizations and Mechanical Analysis', *J. Adhesion*, **58**, 143–161 (1996).
280. Y. Du, E. Rädlein, G.H. Frischat, 'AFM Study in Crystallization and Corrosion of TiO_2 Coating on Glass by NaOH Solution', *International Symposium on Glass Problems, Istanbul, Turkey*, 462–468, 1996.
281. Y. Du, E. Rädlein, G.H. Frischat, 'Atomic force Microscopy Study of NaOH Corroded Al_2O_3–TiO_2–SiO_2 Coatings on Glass', *J. Sol–Gel Sci. Technol.*, **13**, 763–767 (1998).

Nuclear Magnetic Resonance of Coating and Adhesive Systems

Anita J. Brandolini
Mobil Chemical Company, Edison, NJ, USA

1 Introduction 1
2 Analysis with Nuclear Magnetic Resonance 1
2.1 Brief Overview 4
2.2 Identification of Materials 5
2.3 Curing/Drying Reactions and Kinetics 8
2.4 Adsorption/Surface Phenomena 8
3 Spectral Features of Common Coating and Adhesive Materials 9
3.1 Vinyl Polymers 9
3.2 Epoxies and Phenolics 9
3.3 Cellulosics 10
3.4 Acrylates, Polyesters, and Alkyds 10
3.5 Oil-based Coatings 11
3.6 Polyamides, Amino Resins, and Polyurethanes 11
3.7 Silicones and Silicates 12
3.8 Metal Oxides 13
3.9 Additives and Pigments 13
4 Conclusion 13
Abbreviations and Acronyms 13
Related Articles 13
References 14

The interaction of the magnetic moment of an atomic nucleus with a static magnetic field gives rise to nuclear magnetic resonance (NMR). Examination of NMR can reveal both the chemical structure and physical characteristics of a chemical system and this makes the technique particularly useful for the study of materials such as those encountered in coating and adhesive applications. No other spectroscopic technique can offer such a detailed description, both qualitative and quantitative, of molecular structure, including features such as stereochemistry, reaction kinetics and surface phenomena. Each nucleus has a distinct precession frequency, allowing different (e.g. ^{13}C, ^{1}H, ^{29}Si*) to be observed within a sample. The technique does, however, suffer from a few drawbacks. First, it is not a particularly sensitive technique, and it is most readily applied to features that constitute at least a few percent of the sample. Second, sample preparation can be difficult; this can be especially true for coatings and adhesives materials, as will be discussed. The most detailed chemical information is obtained from the spectra of liquids or solutions, which precludes cross-linked or cured materials. The spectra of solids, including cross-linked and cured materials, must be studied using complex experimental approaches, and these often suffer from poor resolution.*

1 INTRODUCTION

Because of the various functions they serve – mainly decorative or protective – coatings can be quite complicated formulations that must form a strong bond with the surface to be covered (the *substrate*). In general, coatings are composed of a *binder*, which originates as, or becomes, a polymer. A volatile solvent mixed with the binder lowers the viscosity enough to allow the coating to be spread evenly across the substrate. Finally, additives such as pigments, opacifiers, drying agents, and dispersants may be included in the formulations to achieve a desired color, intensity, or other effect. Some coatings form a film over the substrate when the solvent evaporates. A latex is water based, although organic solvents are used in other applications. The polymer component is usually a vinyl or acrylic polymer, or a cellulose derivative; in other formulations, the coating forms when monomeric binder molecules polymerize **(1–34)**. The monomeric binders are usually based on alkyds, epoxies, unsaturated polyesters, or vegetable oils. Oxidation of metal surfaces creates yet another class of coating. Adhesives **(3–9, 15, 25–28)** are related to coatings in that they are also thin films in which effective bonding, in these cases between two substrates, is required. In pressure-sensitive applications, the adhesive is inherently sticky and bonds after moderate stress is applied; such adhesives are usually removable. Reactive adhesives, such as epoxies, rely on chemical reactions, which are often similar to those involved in coating formation. Materials commonly used as adhesives include the vinyl polymers, acrylates, and cyanoacrylates [**(1–34)** in Table 1] in addition to the epoxy resins.

2 ANALYSIS WITH NUCLEAR MAGNETIC RESONANCE

The phenomenon of NMR arises from the interaction of an atomic nucleus' magnetic moment with a static magnetic field. The magnetic moment precesses (rotates)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Polymers commonly used as coatings

(1) Poly(ethylene-co-acrylic acid) [*9010-77-9*]

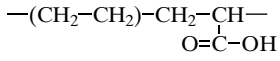

(2) Poly(ethylene-co-ethyl acrylate) [*9003-21-1*]

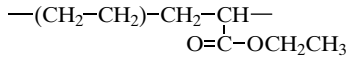

(3) Poly(ethylene-co-vinyl acetate) [*9010-77-9*]

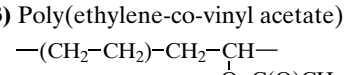

(4) Polypropylene [*9003-07-0*]

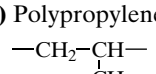

(5) Polyisobutylene [*9003-27-4*]

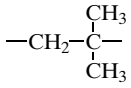

(6) Poly(isobutylene-co-isoprene) [*9010-85-9*]

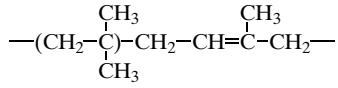

(7) Poly(styrene-co-butadiene) [*9003-55-8*]

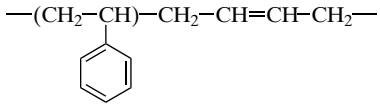

(8) Poly(vinyl acetate) [*9003-20-7*]

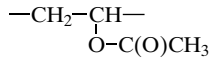

(9) Poly(vinyl alcohol) [*9002-89-5*]

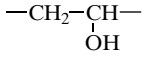

(10) Poly(vinyl fluoride) [*24981-14-4*]

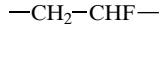

(11) Poly(vinyl chloride) [*900-86-2*]

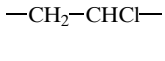

(12) Poly(vinyl chloride-co-vinyl acetate) [*9003-22-9*]

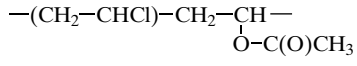

(13) Poly(vinylidene chloride) [*9010-76-8*]

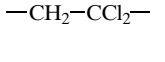

(14) Poly(vinyl methyl ether) [*9003-09-2*]

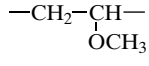

(15) Poly(bisphenol A-co-epichlorohydrin) [*26402-79-9*]

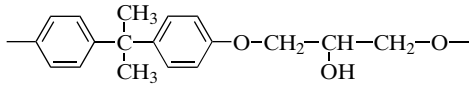

(16) Phenol-formaldehyde resole (example)

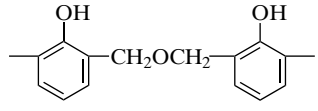

(17) Phenol-formaldehyde novolac (example)

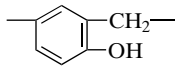

(18) Cellulose derivative

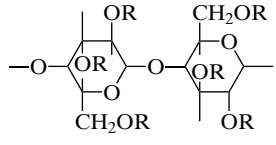

(19) Poly(methyl acrylate) [*9003-86-2*]

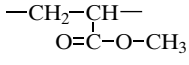

(20) Poly(ethyl acrylate) [*9003-32-1*]

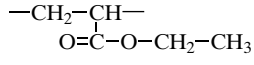

(21) Poly(*n*-butyl acrylate) [*9003-49-0*]

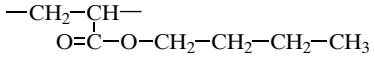

(22) Poly(isobutyl acrylate) [*9004-49-0*]

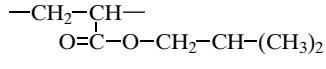

(23) Poly(methyl methacrylate) [*9011-14-7*]

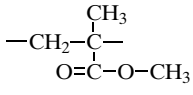

(24) Poly(ethyl methacrylate) [*9003-42-3*]

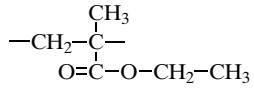

(*continued overleaf*)

Table 1 (*continued*)

(25) Poly(*n*-butyl methacrylate) [*9003-63-8*]

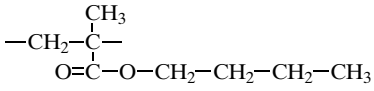

(26) Poly(isobutyl methacrylate) [*9011-15-8*]

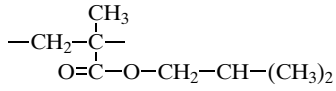

(27) Poly(ethyl cyanoacrylate) [*25067-30-5*]

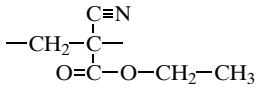

(28) *cis*-Polyisoprene [*104389-31-3*]

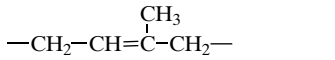

(29) Poly(ethylene terephthalate) [*25038-59-9*]

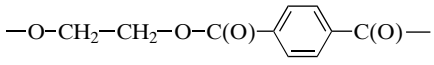

(30) Poly(hexamethylene adipamide) [*32131-17-2*]

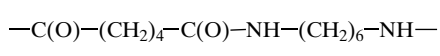

(31) Urea-formaldehyde resin (example)

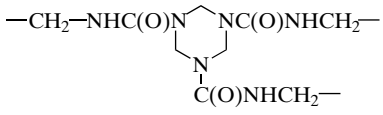

(32) Poly(ester urethane) (example)

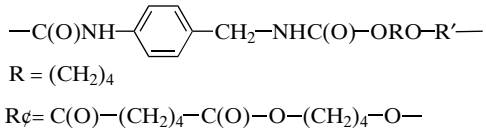

(33) Polydimethylsiloxane [*9016-00-6*]

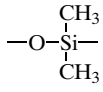

(34) Poly(ethyl silicate) [*68412-37-3*]

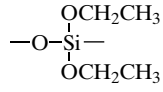

about the static field with a frequency, the Larmor frequency (w_0), that is a property of a given nuclide. Smaller frequency variations called chemical shifts (δ) depend on the exact chemical structure around a particular nucleus. Other features of this precessional motion, such as the coherence of the moments, provide information about molecular reorientations; these relaxation times can be used to calculate motional frequencies and activation energies. No other spectroscopic technique can offer such a detailed description of molecular structure, including stereochemistry,[1–4] reaction kinetics, and surface phenomena. Because distinct nuclei exhibit separate precession frequencies, in many cases several different nuclei within a sample can be independently observed (e.g. ^{13}C, ^{1}H, and ^{29}Si) and may provide complementary information.[5,6] NMR experiments have even been designed so that two or more nuclear properties can be correlated and compared.[7,8]

While NMR is not the most widely used technique for the study of coatings, it can provide useful information in many cases.[9,10] It is, in particular, well suited for elucidating the chemical structure of a coating or adhesive formulation before application and can often provide a fairly complete description of the material's composition. For example, **(35)**, poly(*n*-butyl methacrylate-co-isobutyl methacrylate) [9011-53-4] is often used as a pressure-sensitive adhesive; part of its ^{13}C spectrum is shown in Figure 1. The resonances owing to each comonomer can be clearly identified; if the experimental conditions are appropriately set, the relative amounts of each can be easily calculated based on a comparison of peak areas. Such quantitative compositional analyses can be readily done for many polymer types.

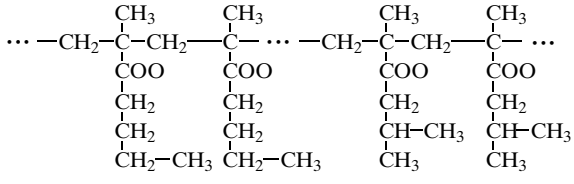

(35)

Adhesives and coatings that solidify through a network-formation reaction, such as epoxies, usually cannot be dissolved in their final state (although the original constituents can be), and such materials must be studied as solids. The mechanisms and kinetics of the reaction can be characterized, and intermediate species identified in some systems. Because of the limited sensitivity of NMR, such studies are usually performed on samples of the coating or adhesive alone, although the substrate can be included in some cases.

The efficacy of an adhesive or coating depends largely on the strength of the bond formed between the binder and the substrate. This bond may be chemical in nature, but more commonly it is a physical attachment. The

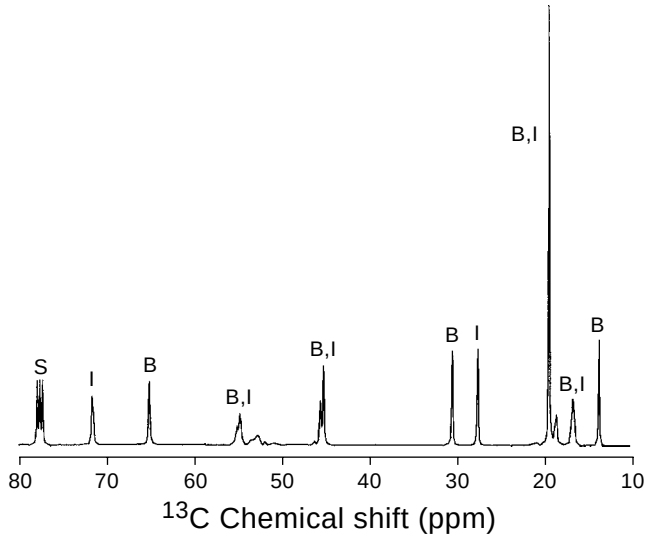

Figure 1 ^{13}C-NMR spectrum of poly(*n*-butyl methacrylate-co-isobutyl methacrylate) **(35)** dissolved in $CDCl_3$ (''S''). Resonances owing to the *n*-butyl methacrylate moiety are marked by ''B''; those arising from isobutyl methacrylate are indicated by ''I''.

conformations taken by the large, polymeric binder molecules, both at the surface and in the bulk of the coating, often govern the effectiveness of the bond. Since NMR can be profitably used to study molecular dynamics, such studies can reveal important information about the physical nature of the adhesion between coating and substrate.

2.1 Brief Overview

There are many publications describing the basic features of NMR,[3,4,11–14] in addition to some that specifically cover its practical applications to polymer systems.[9,10]

As with many other analytical techniques, proper sample preparation is a crucial first step in the successful acquisition of an NMR spectrum. For solution experiments, obtaining a homogeneous sample is critical; this process can take up to several hours for polymers, which often dissolve poorly and/or slowly. Gentle heating of the sample (below the solvent's boiling point, obviously) often hastens this step. Deuterated solvents are normally used to prepare NMR samples; the ^{2}H signal is then monitored to stabilize the instrument's magnetic field and to optimize its homogeneity. Labeled analogs of many common solvents are available from a variety of sources. Samples for solid-state NMR analysis are, of course, not dissolved but must often be processed into a fine powder so that they can be stably spun at high speeds (several kilohertz). In all cases (solutions and solids), the inclusion of metal as an additive or substrate may adversely affect the results. Brown-colored materials, which usually contain iron oxide as a pigment, are particularly notorious in this regard.

As mentioned briefly above, NMRs arise from the altered behavior of certain atomic nuclei in the presence of a magnetic field. Any nuclide with nonzero nuclear spin, I, exhibits this interaction. While many common and chemically significant nuclides, such as ^{12}C or ^{16}O, have no nuclear spin, most elements possess at least one isotope that is NMR-active; the ease with which a signal can be observed and the amount of chemical detail that is contained in the spectrum vary widely, however. Examples of commonly observed nuclides are listed in Table 2. Many of the NMR-observable isotopes are naturally present at only low abundances, although in favorable cases, artificial enrichment can be employed to make the signals more easily detected. Furthermore, nuclides with $I > \frac{1}{2}$ also have quadrupole moments, which tend to broaden the resonances.

The NMR experiment is accomplished by a series of radiofrequency pulses that serve as probes of precessional frequencies of the constituent nuclei. The parameters (e.g. number, length, phase, and spacing) of this pulse sequence determine exactly what data are recorded and whether the end results are quantitative. Different chemical species of a particular nuclide (such as ^{1}H, ^{2}H, ^{13}C, or ^{29}Si) resonate in a narrow range around a specific frequency called the Larmor frequency. In a 9.4 tesla (T; 9.4×10^4 Gauss) magnet, for example, ^{1}H nuclides resonate at 400 MHz, ^{13}C nuclides at 100 MHz, and ^{29}Si nuclides at 79 MHz. The chemically induced differences, chemical shifts, δ, are only a few hertz removed from the precessional frequency; consequently, the chemical shift is usually expressed as the ratio of the frequency shift to the Larmor frequency (Hz/MHz) and cited in parts per million (ppm). Even though different nuclides may exhibit widely separated resonant frequencies, there are interactions, or couplings, among them; these interactions can provide additional information, although they generally also complicate the

Table 2 Nuclei useful for NMR studies of coatings

Isotope	Spin	Natural abundance (%)	Frequency[a] (MHz)	Chemical shift range (ppm)
^{13}C	1/2	1.1	101.8	250
^{1}H	1/2	99.99	400.0	12
^{2}H	1	0.015	61.4	12
^{19}F	1/2	100	376.4	1200
^{29}Si	1/2	4.7	79.5	300
^{31}P	1/2	100	161.9	1200
^{15}N	1/2	0.37	40.6	1000
^{27}Al	5/2	100	104.2	450

[a] In a 9.4 T magnetic field.

spectrum. For example, under certain experimental conditions, a ^{13}C signal will split into several ($n + 1$, where n is the number of directly bonded ^{1}H's) lines of differing intensities, separated by a coupling constant, J. At times, even longer-range interactions can be observed. While such $^{13}C–^{1}H$ coupling can provide useful information (identifying CH_3, CH_2, etc.), it also complicates the spectrum by producing many, often overlapping, lines. It can be easily removed or reduced, however, by a technique called heteronuclear scalar, or J, decoupling. Homonuclear couplings (such as $^{13}C–^{13}C$ or $^{1}H–^{1}H$) also exist, but their importance depends on the abundance of the NMR-active isotope. Juxtapositions of $^{13}C–^{13}C$, for instance, are rare because of the relative paucity of ^{13}C compared with ^{12}C in unenriched materials. By comparison, since ^{1}H's are essentially 100% abundant, $^{1}H–^{1}H$ couplings are intense and can dominate the spectrum; if necessary, they can be removed or reduced by homonuclear decoupling techniques.

During the course of an NMR experiment, the pulse sequence is repeated many times; the pause between cycles is called the relaxation delay. If the relative peak areas in the final spectrum are required to be quantitative, the relaxation delay must be set long enough to allow the system to return to equilibrium between cycles; this time depends on the chemical identity of the nucleus and on the system's dynamic behavior (typical lengths are a few seconds to several minutes). If this time is so long that the duration of the experiment will be too great, then a paramagnetic relaxation agent, such as chromium(III) acetylacetonate [21679-31-2], can be added to a sample solution; a shorter relaxation delay can then be used. This is commonly done for observation of slow-relaxing nuclides such as ^{29}Si and ^{15}N.

A consequence of the decoupling process, called the nuclear Overhauser enhancement (NOE), can also result in further peak-intensity distortions. While this side effect generally increases the signal-to-noise ratio (normally a benefit), it can affect peaks disproportionately. For quantitative analyses, the NOE is usually removed by gated decoupling, in which the decoupling field is turned off for part of the experiment. Some nuclei, such as ^{29}Si and ^{15}N, exhibit a negative NOE, which can actually decrease the signal intensity; in studies involving these nuclei, it is critical to employ gated decoupling.

Over the course of an NMR experiment, the pulse sequence is repeated many times in order to build up a sufficiently large signal to distinguish the details of interest. The signal-to-noise ratio increases as the square root of the number of scans; consequently the duration of an acquisition must be quadrupled in order to double the ratio. For weak signals, this can make for very long experiments (up to several days). Other parameters, such as the length and phase of the radiofrequency pulses are also important in certain NMR experiments; information on these experimental protocols can be found in many references.[1–8,11–15]

While NMR spectra of solids are acquired in a superficially similar manner, the intrinsic differences between solids and liquids necessitate some changes in sample handling, instrumentation, and experimental approach. Molecular motions in solids tend to be anisotropic and slow (on the timescale of a typical experiment) and internuclear interactions are not motionally averaged, as they are in liquids. As a consequence, special techniques must be applied if spectra with relatively narrow, symmetric resonances are desired. For example, in order to obtain a high-resolution ^{13}C spectrum of a solid, samples are commonly packed into small cylindrical sample holders called *rotors*, which are then spun rapidly (several kilohertz) about an axis oriented 54.7° to the static magnetic field. This magic-angle spinning reduces anisotropies of chemical shift that result from the lack of rapid molecular tumbling. Decoupling of $^{13}C–^{1}H$ is applied, although at a much higher power level than for solution-state NMR, in order to overcome through-space dipolar interactions, as well as the much weaker J couplings. For ^{13}C, the basic data-acquisition pulse sequence is designed to enhance the signal and reduce the relaxation time (which can be quite long because of the restricted molecular motion) via cross-polarization. The details of such experiments are discussed in greater detail elsewhere.[15,16]

The discussion above clearly demonstrates that there are many different types of information obtained from the NMR phenomenon. Spectra in which all the inter- and intranuclear interactions are fully expressed are so full of overlapping information that they may become difficult to interpret with great specificity. In an attempt to reach a compromise between these inherent complexities and the simplicity needed for readily understandable results, clever experiments have been devised that allow for the spectral information to be selectively recovered. For example, rather than showing the multiplicity of a J-coupled ^{13}C resonance by its splitting pattern, this feature can be indicated by the phase of its signal (i.e. whether it is upright or inverted in the final spectrum), by specially crafted pulse sequences such as distortionless enhancement through polarization transfer (DEPT) and insensitive nuclei enhanced by polarization transfer (INEPT).[17] Even more complex experimental protocols produce spectra in which the interactions are shown along more than the normal two (frequency–intensity) axes. Examples will be discussed in the following sections.

2.2 Identification of Materials

The ^{13}C spectrum shown in Figure 1 is a typical representation of a one-dimensional spectrum of a polymer

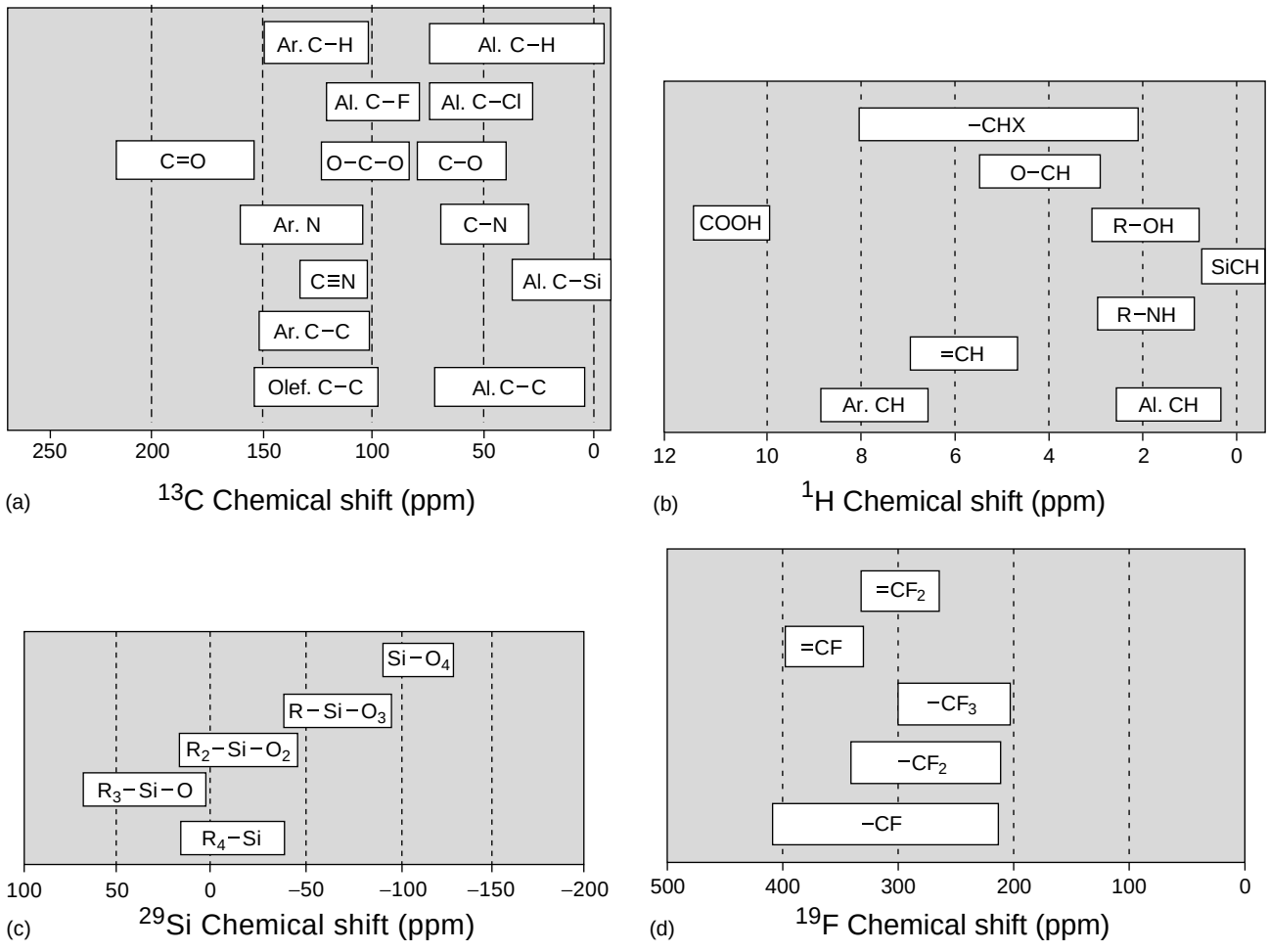

Figure 2 Chemical-shift correlation charts for (a) ^{13}C; (b) ^{1}H; (c) ^{29}Si; and (d) ^{19}F. Ar., aromatic; Al., aliphatic; Olef., olefinic (alkenic).

solution. While it is the particular set of resonance positions and relative intensities that ultimately reveals the material's identity, generalizations about chemical functionality can be made quickly because the correlations between structures and shifts are well established for most NMR-active nuclei. For example, carbonyl carbons resonate over a fairly narrow range, as do aromatic or alkenic carbons, carbons bonded to nitrogen or oxygen, and so on. For aliphatic structures, it is known that methyl groups tend to appear at lower chemical shift (δ) values than do methylene groups, which are at lower values than methines, which resonate, in turn, at lower shifts than quaternaries. Figure 2 shows these correlations for ^{13}C, ^{1}H, ^{29}Si, and ^{19}F, the nuclei most commonly of interest for coatings and adhesives. Material identification can be done through deduction from chemical-shift correlations by comparison with published spectra or model compounds[11,18–20] or (most likely) through a combination of approaches.

While NMR is undoubtedly useful for basic materials identification, it also serves as one of the most powerful structural characterization techniques, which is able to distinguish subtle molecular details. Stereochemistry, for example, is an important feature of many polymers. Materials made from monomers such as vinyl halides, acrylate esters, or 1-alkenes higher than C_2 exhibit tacticity, which arises from pendant-group stereo-inequivalence.[21] In isotactic materials, all of the side groups (X) are on one "side" of the backbone chain [as shown in **(36)**] while in atactic polymers **(37)**, they are randomly distributed. Syndiotactic chains have alternating placement of the X group. This feature, which can significantly affect bulk properties, is clearly reflected in the NMR spectrum as a pattern of resonances that can be assigned to pairs of stereosequences. In the spectrum of poly(vinyl chloride) for example, [**(11)**; Figure 3], the resonances can be designated as *mm* [**(38)**: for two sequential *meso* placements], as *rr* [**(39)**: for two racemic placements] or as *mr* [**(40)**: for *meso*-racemic placement]. A simple comparison of

quantitative peak areas in a spectrum reveals the relative amounts of *mm*, *rr*, and *mr*. Furthermore, some polymers, such as poly(vinyl fluoride) **(10)** exhibit regiosequences that can be 'head to tail' **(41)** or 'head to head' **(42)**. These are also analyzable from the NMR spectrum, as shown in Figure 4.

The primary use of NMR in chemical analysis is as an identification tool based on the great sensitivity of chemical shift to the details of chemical structure, as seen in the examples above. The technique's utility can be further enhanced, in many cases, by the observation of more than

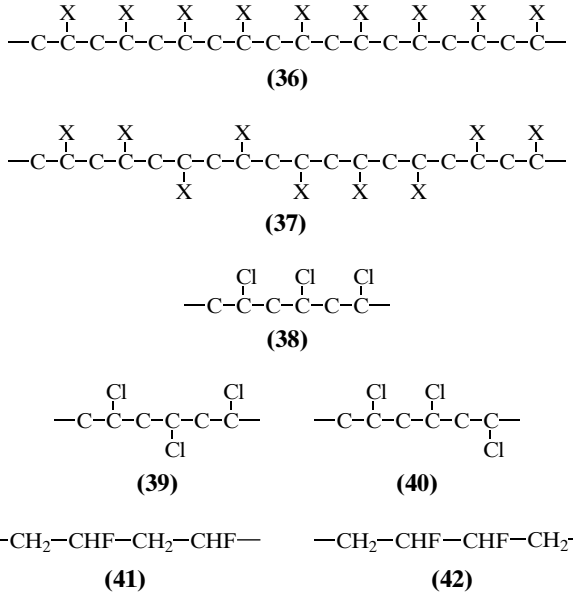

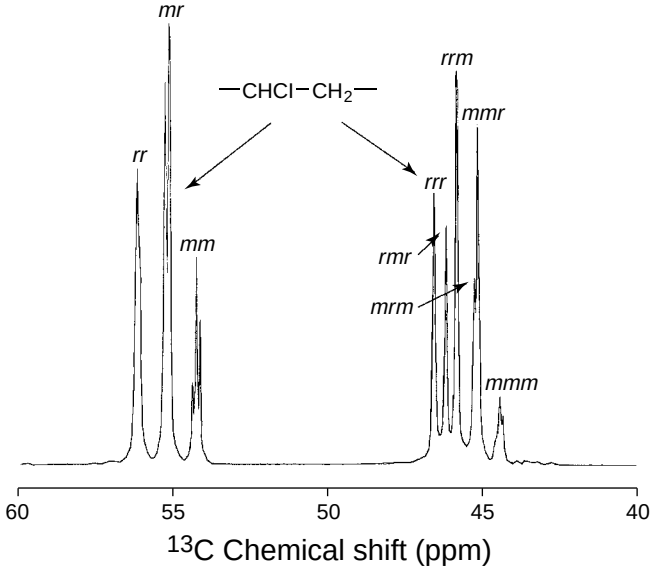

Figure 3 ^{13}C-NMR spectrum of poly(vinyl chloride) dissolved in *o*-dichlorobenzene-d_4, showing fine structure resulting from tacticity. *m*, *meso*; *r*, racemic; [see **(38–40)**].

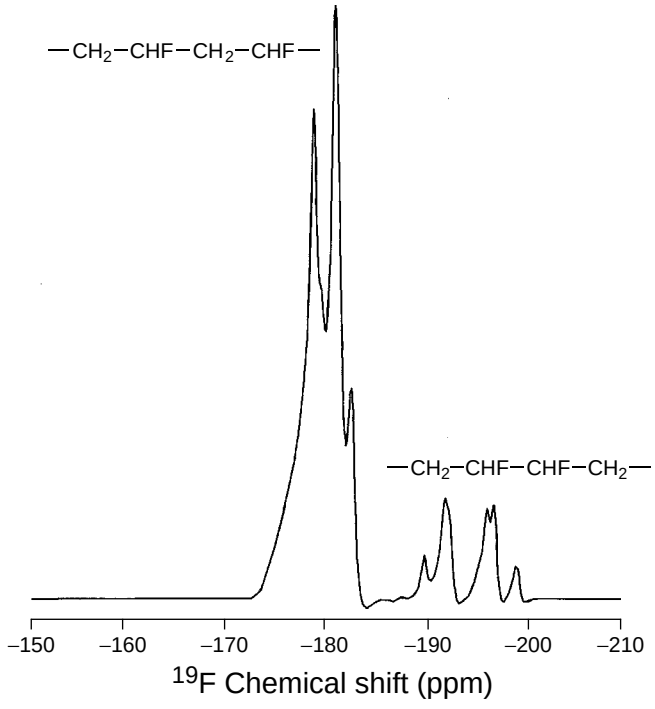

Figure 4 ^{19}F-NMR spectrum of poly(vinyl fluoride) dissolved in acetone-d_6, showing fine structure resulting from regiosequences [see **(41)** and **(42)**].

one nuclide. For example, the ^{29}Si spectra of polysiloxanes indicate the material's molecular weight more unambiguously than the spectra of either ^{13}C or ^{1}H. Furthermore, since the ^{29}Si experiment only reflects species that contain that nuclide, components that do not contain silicon do not interfere. Other nuclides, listed in Table 2, are also useful in studying coatings, although the ease with which they can be observed varies; in general, higher resonant frequencies and natural abundances make it simpler to acquire signals of sufficient intensity. Unfortunately, some nuclides that would be expected to be informative, such as ^{17}O, are generally not studied because of the difficulties of signal detection in an unenriched sample. Other nuclides, such as the metals, tend to give broad peaks, which can obscure chemical details.

Coatings and adhesives are generally complex formulations, and the presence of ingredients other than the binder can complicate the spectrum. For example, water does not interfere in a ^{13}C spectrum, but an organic solvent, such as a ketone, does; by comparison, water would contribute a large signal to the ^{1}H spectrum. Solid additives, such as inorganic pigments, may degrade the spectral quality because they disturb the magnetic-field homogeneity. Brown or orange pigments are particularly troublesome because they often contain ferromagnetic iron species, which can make the spectrum nearly uninterpretable as a consequence of severe line broadening.

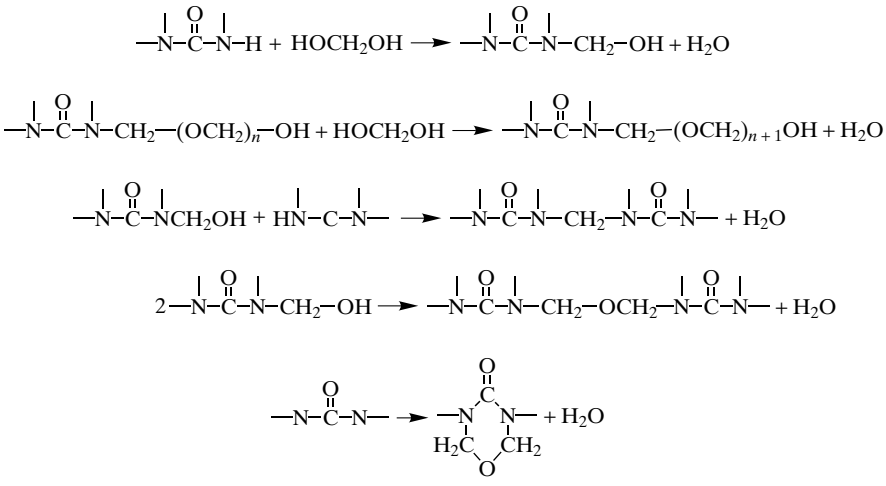

Scheme 1

2.3 Curing/Drying Reactions and Kinetics

Characterization of the mechanism and kinetics of curing and/or drying reactions first requires an understanding of correlation of chemical shifts with structure, as described above. If spectra are recorded as a function of time, changes in the positions and/or intensities of resonances reveal the details of the reactions taking place. For example, the evaporation of water from a latex material can be readily followed by the loss of the ^{1}H signal due to water. Since this process is fairly slow (hours to days), the length of the actual NMR measurement (minutes to hours) is not a significant detriment, as it might be for faster reactions.

The curing of commercial coating materials typically involves processes more complex than simple evaporation. In the urea-formaldehyde resin system **(31)**, for example, there are many possible reaction sites, leading to a complex chemistry of reaction products (Scheme 1).[22] Complementary studies (^{13}C and natural abundance ^{15}N) reveal many details of the curing reactions, as each of these structures exhibits unique resonances for one or both nuclei. Solid-state NMR techniques can be used to distinguish them and characterize the final material.[22]

2.4 Adsorption/Surface Phenomena

The efficacy of a coating or adhesive relies heavily on its ability to bond with the substrate, and on the sensitivity of this bond to environmental conditions. Any experimental tool that can elucidate the nature of this chemical or (more commonly) physical phenomenon is valuable. NMR, particularly the solid-state experimental variants, fills this need, as illustrated by a recent study of the adsorption of poly(methyl acrylate)-d_3 **(19)** onto silica.[23] Incorporation of a deuterium label into the polymer and subsequent observation of the ^{2}H resonance clearly revealed the presence of at least two dynamically distinct components. The polymer chains at the silica interface were motionally constrained, while those on the surface of the coating exhibited considerably more freedom, as illustrated in Figure 5. The temperature-dependent behavior of this system was also studied, and since the correlation between molecular dynamics and ^{2}H lineshapes is well established,[10] a detailed model of polymer chain motion occurring in this model coating could be developed.

The ability of NMR to distinguish chemical entities within a polymer can provide insight into the adsorption of more complex systems.[24] In particular, in the blocky poly(styrene-co-4-vinylpyridine) **(43)** [26222-40-2] the vinylpyridine moiety was found to be preferentially

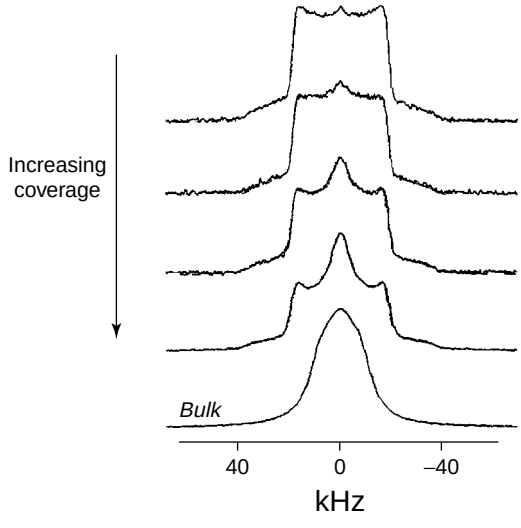

Figure 5 ^{2}H-NMR spectrum of poly(methyl acrylate)-d_3 adsorbed onto silica. (Reprinted with permission from Lin and Blum,[23] Copyright (1997) American Chemical Society.)

absorbed onto silica [*112945-52-5*], titanium dioxide [*1317-80-2*], and calcium carbonate [*471-34-1*]. In this case, solid-state NMR was not used; instead, the disappearance of the copolymer from the supernatant solution was tracked.

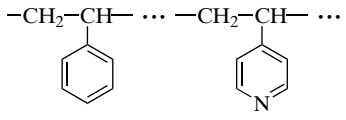

(43)

3 SPECTRAL FEATURES OF COMMON COATING AND ADHESIVE MATERIALS

It should be obvious from the foregoing discussions that a basic understanding of the chemical attribution of each observed resonance in an NMR spectrum is critical to the technique's effective use. The following sections, therefore, briefly review these correlations for the types of material commonly found in coating and adhesive formulations and show some typical spectra. In all cases, the ^{13}C and ^{1}H chemical shifts are included, but other relevant nuclides are discussed where appropriate.

3.1 Vinyl Polymers

Vinyl polymers commonly used as coatings or adhesives include ethylene copolymers such as poly(ethylene-co-ethyl acrylate) **(2)**, polypropylene **(4)**, isobutylene homopolymer **(5)** and copolymers, poly(styrene-co-butadiene) **(7)**, vinyl acetate homopolymer **(8)** and copolymers, poly(vinyl alcohol) **(9)**, poly(vinyl fluoride) **(10)**, vinyl chloride homopolymers **(11)** and copolymers, vinylidene chloride homopolymer **(13)** and copolymers and poly(methyl vinyl ether) **(14)**. All have essentially linear, hydrocarbon backbones; the pendant groups vary (see Table 1 for structures). Because the chemical functionalities of the side groups are different (e.g. alkyl chain, halide, ether, or ester), it is difficult to predict the spectrum a priori; identification is best made by comparison with published data for known materials or model compounds.[18–20] Backbone carbons not attached to anything other than hydrogen or another carbon, and any carbon-only pendant groups, resonate between 10 and 60 ppm, depending on the exact type of carbon (methyl, methylene, etc.) and on the exact structure surrounding it. Approximate peak positions for the carbons in other groups (such as esters) can be deduced from Figure 2(a).

Many of these spectra exhibit the subtle features associated with tacticity (section 2.2); this effect is particularly significant for the backbone and for any carbons immediately attached to it, as illustrated in the ^{13}C-NMR spectrum of atactic polypropylene (Figure 6). In solid-state spectra, the tacticity features are usually obscured by residual broadening.

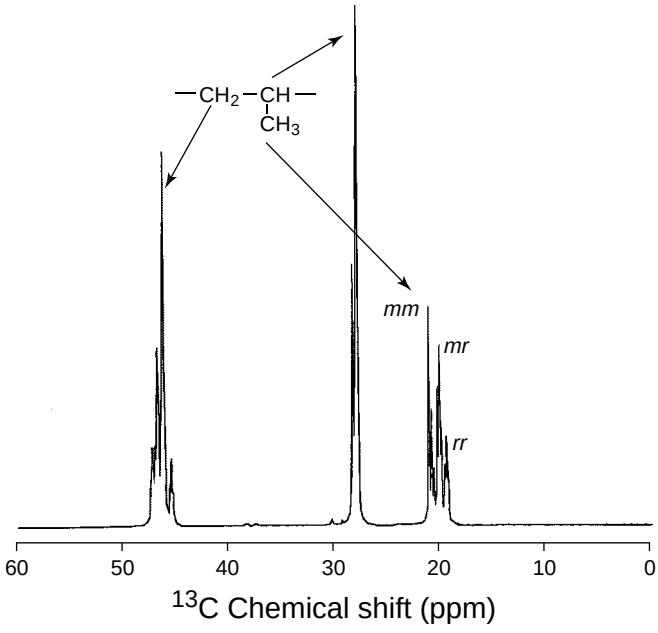

Figure 6 ^{13}C-NMR spectrum of atactic polypropylene dissolved in *o*-dichlorobenzene-d_4/1,2,4-trichlorobenzene, showing fine structure resulting from tacticity. *m*, *meso*; *r*, racemic.

In some cases, ^{1}H-NMR can provide useful information about coating and adhesive systems, but it is not usually the nucleus of choice for study of these materials because polymers tend to give broad resonances that can be difficult to resolve. ^{19}F spectra of fluoropolymers can be used to characterize both tacticity and regioselectivity; two-dimensional NMR has been particularly useful for studying these structural features.[25,26] In general, neither ^{35}Cl nor ^{37}Cl provides enough chemical detail to be useful for the investigation of vinyl chloride or vinylidene chloride polymers.

3.2 Epoxies and Phenolics

Epoxies generally have a phenolic structure, but not all phenolic resins **(15–17)** encountered in coatings applications are epoxies. As illustrated in Figure 7, which shows the ^{13}C spectrum of poly(bisphenol A-co-epichlorohydrin) **(15)**, peaks indicating aliphatic ether linkages appear at 60–100 ppm for ^{13}C (4–6 ppm for ^{1}H); the phenolic moiety gives rise to ^{13}C resonances in the aromatic region, ranging from 110 to 150 ppm. In particular, the aromatic carbon *ortho* to the oxygen resonates at approximately 115 ppm, in an uncommon chemical shift range; the appearance of a peak at this point

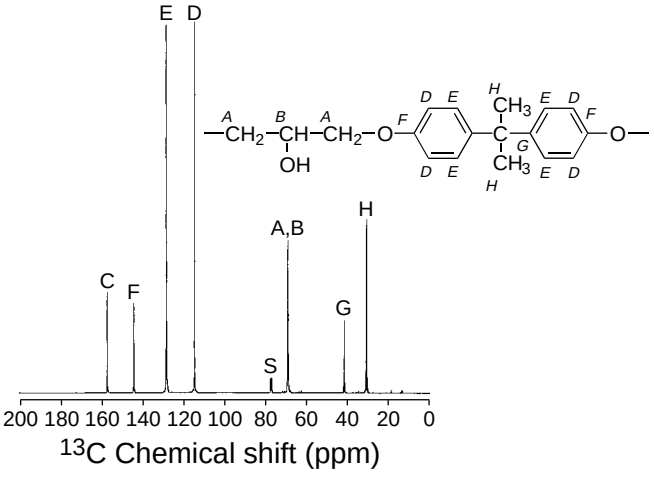

Figure 7 ^{13}C-NMR spectrum of poly(bisphenol A-co-epichlorohydrin) dissolved in $CDCl_3$ (solvent peak indicated by "S").

strongly suggests the presence of a phenolic functionality. Epoxy materials are generally formed in situ from two components that react to create the final product. Solution NMR can easily describe the structure of the starting materials, and the technique can also be used to follow the polymerization reaction if spectra are acquired as a function of curing time.

The other types of phenolic resin used in coatings are the resoles **(16)** and the novolacs **(17)**, which are phenol-formaldehyde products. The use of a base as the catalyst for a resole, and an acid for a novolac, ultimately leads to slightly different structures. The prepolymers can be studied by solution NMR (^{13}C, ^{1}H, and ^{15}N), but the final cured coatings are insoluble owing to their highly cross-linked nature. Solid-state techniques can elucidate these structures.[27,28]

3.3 Cellulosics

Cellulosics are basically polyethers, comprising linked six-membered saccharide rings **(18)**. The hydroxyl groups are usually derivatized, most commonly as ethers and esters. The resonances of these materials are often rather broad because of their limited solubility. The ^{13}C ring resonances appear between 70 and 105 ppm, and the carbons in the pendant groups appear where appropriate for the specific type. An example of this, cellulose triacetate [**(18)**, where R is acetate [9012-09-3]], is shown in Figure 8; this compound exhibits the carbonyl and methyl signals typical of acetate groups, in addition to the saccharide peaks. For the substituted cellulosics, incomplete derivatization may introduce additional resonances; some of the small peaks in Figure 8 result from the presence of mono- and diacetate forms. Starch-based adhesives exhibit spectra very similar to those of the cellulosics.

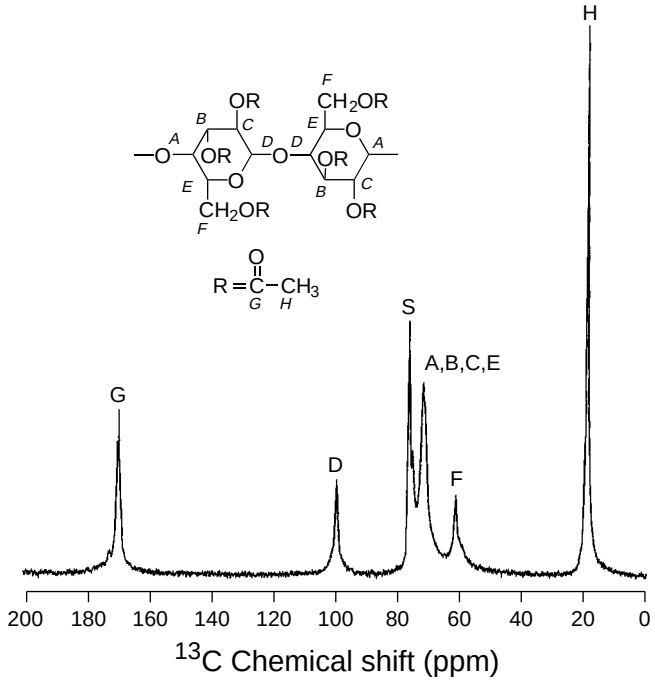

Figure 8 ^{13}C-NMR spectrum of cellulose triacetate dissolved in $CDCl_3$ (solvent peak indicated by "S").

3.4 Acrylates, Polyesters, and Alkyds

While acrylic polymers and aliphatic polyesters have different backbone structures in Table 1, their spectra often appear similar because ester groups are present in both (see Figure 1 for example). Such carbonyl carbons resonate between 165 and 180 ppm, while C–O is observed in the 50–80 ppm range (for acrylates and aliphatic polyesters). Acrylates (along with the related methacrylates and cyanoacrylates) are used as bases for many latex paints and adhesives. Such water-based systems also contain an elastomer, most commonly *cis*-polyisoprene **(28)** or natural rubber (Figure 9). The cyanoacrylates, which are used in "super glues", are important components of many adhesives; the α-C≡N peak **(27)** appears around 120 ppm. ^{1}H spectra of these materials can be quite informative, as the position of the pendant OCH may change significantly with the identity of this group.[16]

Polyesters used as coatings are typically based on terephthalic acid and give ^{13}C spectra similar to that of poly(ethylene terephthalate) [**(29)**; Figure 10]. It should be noted that many polyurethanes (section 3.6) contain ester segments; these should not be confused with polyester resins. Alkyd resins are based on polyesters made with unsaturated diacids such as maleic [*110-16-7*] and fumaric [*110-17-8*] acids, or with unsaturated diols like allyl alcohol [*107-18-6*]. During curing, a network forms through the double bonds; this can be followed

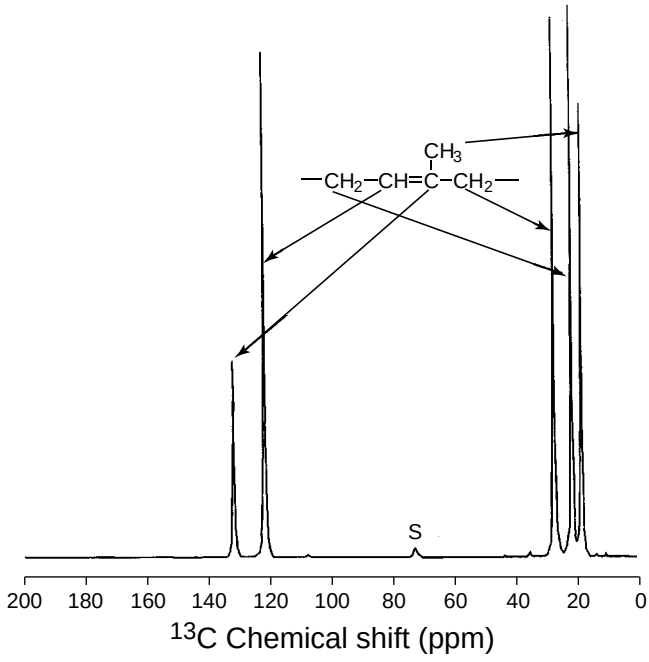

Figure 9 ^{13}C-NMR spectrum of *cis*-polyisoprene dissolved in $CDCl_3$ (solvent peak indicated by "S").

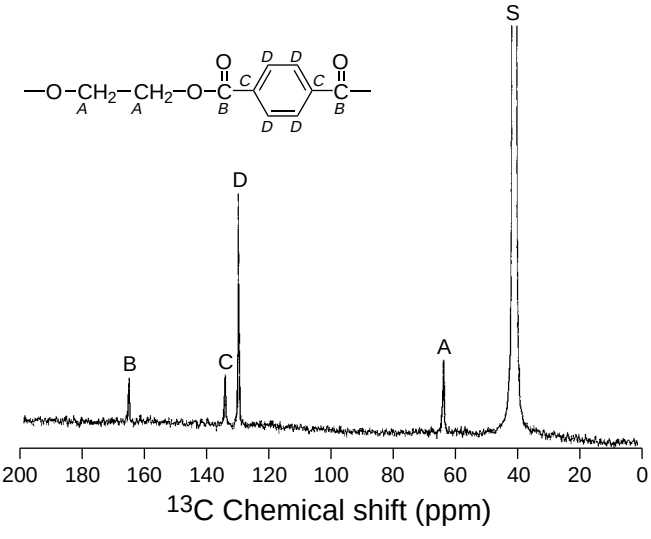

Figure 10 ^{13}C-NMR spectrum of poly(ethylene terephthalate) dissolved in DMSO-d_6 (solvent peak indicated by "S").

from the decrease in signal intensity in the ^{13}C chemical shift region 110–140 ppm.

3.5 Oil-based Coatings

Oil paints are based on unsaturated plant oils, such as linseed oil [*8001-26-1*]; the ^{13}C spectrum of an epoxidized form of this material is shown in Figure 11. These oils are usually glycerol esters of various unsaturated fatty

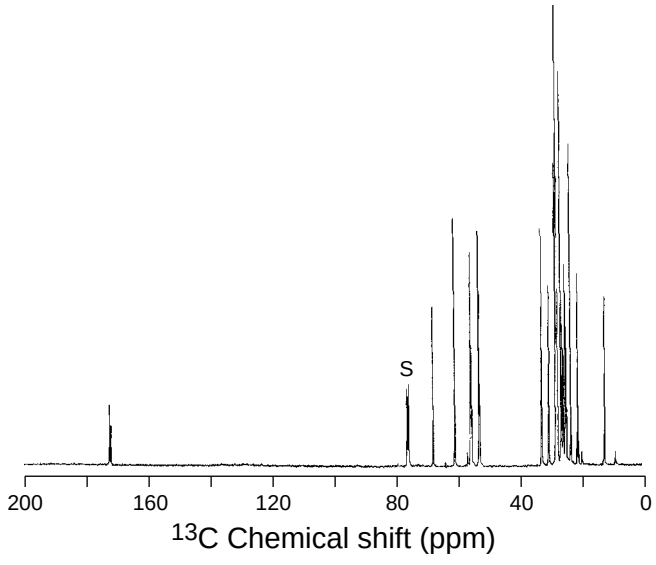

Figure 11 ^{13}C-NMR spectrum of epoxidized linseed oil dissolved in $CDCl_3$ (solvent peak indicated by "S").

acids, such as oleic, linoleic, and linolenic acids. The ^{13}C spectrum, therefore, exhibits a complex pattern of resonances, with carbonyls at 170–180 ppm, alkenic carbons at 110–140 ppm, C–O at 50–80 ppm, and aliphatic groups at 10–60 ppm. As the paint cures, the double bonds open, leading to network formation and a concurrent decrease in intensity in the 110–140 ppm region.

3.6 Polyamides, Amino Resins, and Polyurethanes

The nitrogen-containing materials fall into a number of different chemical classes. Amide linkages are found in polyamides (nylons), some amino resins, and the polyurethanes. The ^{13}C spectra of these materials may appear, at first inspection, similar to those of the polyesters, but there is no C–O resonance between 50 and 80 ppm. The nylons are linear polymers that are formed by ring-opening polymerization of a lactam or by co-reaction of a dicarboxylic acid and a diamine. The resulting material is designated as "Nylon X" or "Nylon XY", where X indicates the number of carbons in the lactam or in the diacid, and Y the number in the diamine. It is possible to determine X and Y based on relative peak areas, but the distinction is more readily achieved using melting points. Amino resins are products of the reaction of formaldehyde [*50-00-0*] with either urea [*57-13-6*] or melamine [*108-78-1*]. The system discussed in section 2.3 is an example of a urea-formaldehyde resin **(31)**.[22] Since the resulting structure of this type of polymer is, essentially, a highly cross-linked polyamide, its spectrum is similar to that of a nylon. Melamine-formaldehyde resins [*68002-25-5*] contain triazine rings, which give rise

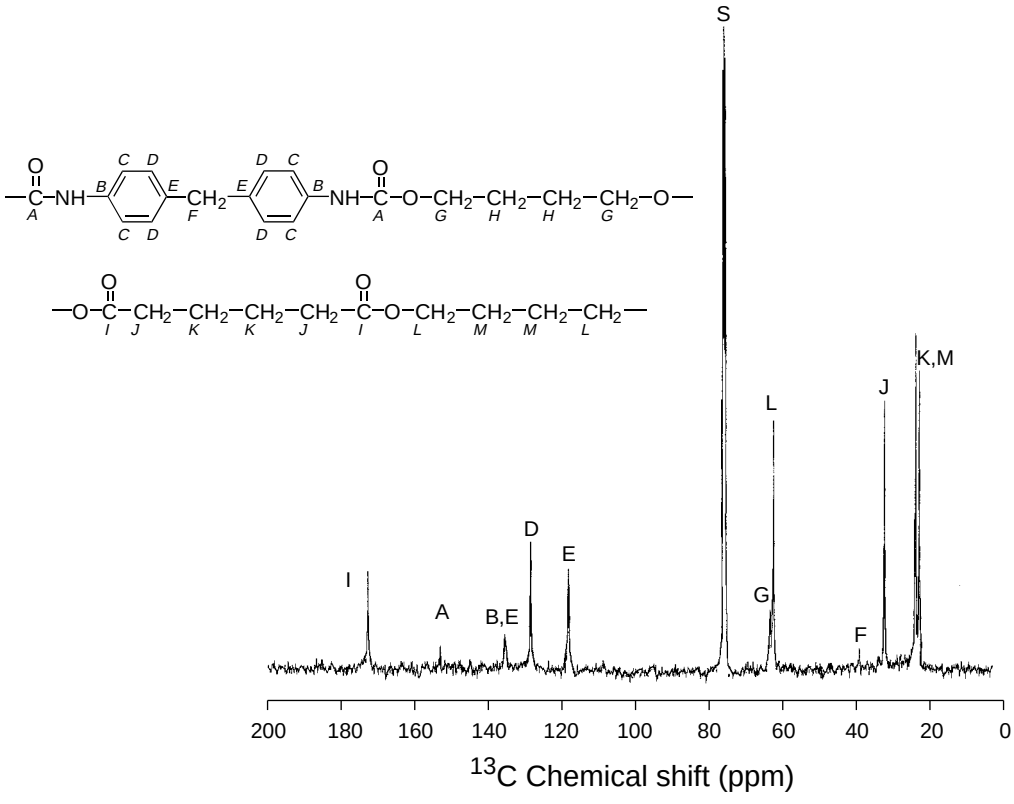

Figure 12 ^{13}C-NMR spectrum of poly(ester urethane) dissolved in $CDCl_3$ (solvent peak indicated by 'S').

to ^{13}C resonances in the aromatic region between 110 and 150 ppm. Polyurethanes also contain amide structures in the backbone, but these segmented materials also include long polyether and/or polyester blocks. In many cases, these will be the dominant resonances in the spectrum, as shown in Figure 12. In addition to ^{13}C NMR, ^{15}N experiments (either with an enriched sample or in natural abundance) can be used specifically to study reaction sites, without possibly complicating signals from the carbon-containing parts of the molecule.

3.7 Silicones and Silicates

Silicon-containing polymers such as the silicones and silicates [for example, **(33)** and **(34)**] are relatively easy to identify by either ^{13}C or ^{1}H NMR. The presence of the silicon atom dramatically affects the shift of any carbon (or of any hydrogen bonded to a carbon) directly attached to it, causing it to resonate near 0 ppm. In fact, the common chemical shift reference, tetramethylsilane [*75-76-3*] is defined as 0 ppm for both ^{13}C and ^{1}H. Furthermore, ^{29}Si NMR is a particularly useful tool for the study of these materials, as it provides much structural information. There is, for example, a significant difference in chemical shift, depending on whether the silicon is bonded to one, two, three, or four oxygens (types M, D, T and Q, respectively; see Figure 2c). The most common silicon-containing coating materials are type D, such as polydimethylsiloxane **(33)**. In some cases, even more details are available; the spectrum shown in Figure 13 illustrates this point for poly(ethyl silicate) **(34)**. The four groups of resonances clearly indicate silicon atoms that are attached (through an oxygen bridge, Q type) to one, two, three, or four other silicon atoms [**(44)**, **(45)**, **(46)**, and **(47)**, respectively]. The relative intensities of each resonance can be used to determine the amount of each structure. The broad peak that underlies the narrow resonances arises from silicon in the glass tube that contains the sample.

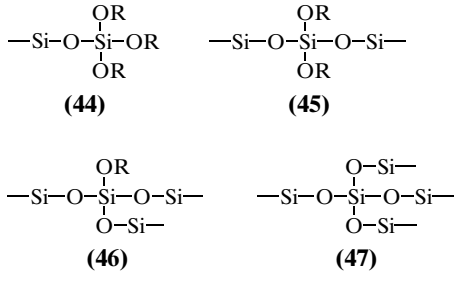

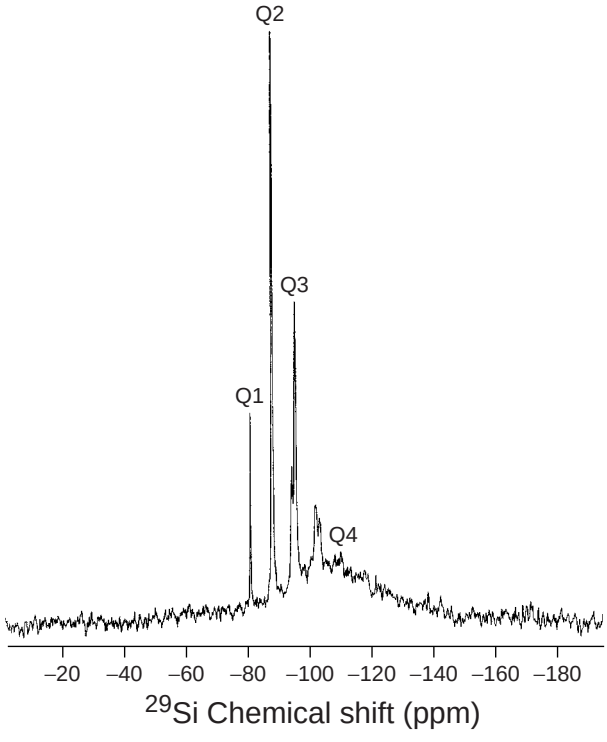

Figure 13 ^{20}Si NMR spectrum of poly(ethyl silicate) dissolved in $CDCl_3$.

3.8 Metal Oxides

Many metals are provided with protective coatings by allowing the surface to oxidize slightly. This thin coating can be difficult to observe directly by NMR for several reasons. First, the amount is quite small. Second, the metal substrate may interfere with signal acquisition by creating radiofrequency inhomogeneities;[29] lastly, the particular nucleus may not be appropriate for NMR analysis. However, in some cases, a great deal of information can be obtained from model metal oxide systems. One such example is the investigation of aluminum oxide [*1344-28-1*] by ^{27}Al-NMR.[30] The invisibility of some of the surface aluminum atoms suggests that they are different in nature from the bulk. Low-temperature studies indicated that this difference is a consequence of greater molecular motion at the surface.

3.9 Additives and Pigments

Coatings and adhesives are complex formulations, often containing many colorants and other additives. NMR is not a particularly useful technique for the study of pigment species per se, since many are simple inorganic oxides or salts. For example, the most common white colorant is titanium dioxide, and neither titanium nor oxygen is a very facile nucleus for NMR analysis. As discussed previously, brown or orange iron-based colors can severely degrade the quality of the observed signal since their presence induces magnetic-field inhomogeneities. Organic colorants can, in principle, be observed by NMR, but a mutually suitable solvent for pigment and binder must be found, which may be difficult. However, NMR can be used to study the behavior of other formulation additives, such as the antiplasticizers that modify the physical properties of the base material. Sophisticated multidimensional solid-state techniques (^{2}H, ^{15}N, and ^{13}C rotational-echo double resonance) have been employed to pinpoint the location of such an additive in a cross-linked epoxy resin.[31] Since the experiment involves the measurement of geometry-dependent $^{2}H-^{15}N$ and $^{15}N-^{13}C$ couplings, isotopically enriched samples must be prepared. These measurements revealed the average distance of the additive's carbonyl from the amine nitrogen and the quaternary carbon of the epoxy resin and indicated that the additive has no preference for either the free or the cross-linked amine sites.

4 CONCLUSION

Because of the great wealth of molecular detail that can be extracted from NMR spectroscopy, the technique finds applications even to systems as complex as coatings and adhesives. It is particularly useful for studying model systems examining the mechanisms and kinetics of the drying or curing reaction and to achieve insight into the physical basis of adhesion.

ABBREVIATIONS AND ACRONYMS

DEPT	Distortionless Enhancement Through Polarization Transfer
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Enhancement
T	Tesla

RELATED ARTICLES

Coatings **(Volume 2)**
Thermal Analysis of Coatings

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. E. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy*, VCH, New York, 1987.
2. W.S. Brey, *Pulse Methods in 1D and 2D Liquid-Phase NMR*, Academic Press, San Diego, CA, 1988.
3. M.D. Bruch (ed.), *NMR Spectroscopy Techniques*, Marcel Dekker, New York, 1996.
4. C. Dybowski, R.L. Lichter (eds.), *NMR Spectroscopy Techniques*, Marcel Dekker, New York, 1987.
5. R.K. Harris, B.E. Mann, *NMR and the Periodic Table*, Academic Press, London, 1978.
6. J. Mason (ed.), *Multinuclear NMR*, Plenum Press, New York, 1987.
7. J. Schraml, J.M. Bellama, *Two-dimensional NMR Spectroscopy*, John Wiley & Sons, New York, 1988.
8. F.J.M. van de Ven, *Multidimensional NMR in Liquids*, VCH, New York, 1995.
9. R.A. Komoroski (ed.), *High-resolution NMR Spectroscopy of Synthetic Polymers in Bulk*, VCH, Deerfield Beach, FL, 1986.
10. L.J. Mathias (ed.), *Solid State NMR of Polymers*, Plenum Press, New York, 1991.
11. W. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy*, VCH, New York, 1987.
12. W.S. Brey, *Pulse Methods in 1D and 2D Liquid-Phase NMR*, Academic Press, San Diego, CA, 1988.
13. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, UK, 1987.
14. C. Farrar, E.D. Becker, *Pulse and Fourier Transform NMR*, Academic Press, New York, 1971.
15. C.A. Fyfe, *Solid-State NMR for Chemists*, CFC Press, Guelph, ON, 1983.
16. B.C. Gerstein, C.R. Dybowski, *Transient Techniques in NMR of Solids*, Academic Press, New York, 1985.
17. S.W. Homans, *A Dictionary of Concepts in NMR*, Clarendon Press, Oxford, 1995.
18. A.J. Brandolini, D.D. Hills, *NMR Spectra of Polymers and Polymer Additives*, Marcel Dekker, New York, 2000.
19. Q.T. Pham, R. Petiaud, H. Waton, *Proton and Carbon NMR Spectra of Polymers*, J. Wiley & Sons, New York, 1983.
20. L.J. Mathias, R.F. Colletti, R.J. Halley, W.L. Jarrett, C.G. Johnson, D.G. Powell, S.C. Warren, *Solid-state NMR Polymer Spectra Collected Vol. 1*, MRG Press, Hattiesburg, MS, 1990.
21. J.C. Randall, *Polymer Sequence Determination: Carbon-13 Method*, Academic Press, New York, 1977.
22. I.-S. Chang, G.E. Maciel, 'Natural Abundance ^{15}N Cross-polarization/Magic-angle Spinning Nuclear Magnetic Resonance Study of Urea-formaldehyde Resins', *Polymer*, **35**(8), 1621–1628 (1994).
23. W.-Y. Lin, F.D. Blum, 'Segmental Dynamics of Bulk and Adsorbed Poly(methyl acrylate)-d_3 by Deuterium NMR: Effect of Adsorbed Amount', *Macromolecules*, **30**, 5331–5338 (1997).
24. F. Bossé, H.P. Schreiber, A. Eisenberg, 'Specific Adsorption of some Styrene/Vinylpyridine Diblocks from Selective Solvents onto Solid Substrates – an NMR Study', *Macromolecules*, **26**, 6447–6454 (1993).
25. M.D. Bruch, 'Measurement of Homonuclear Coupling Constants Associated with Poly(vinyl fluoride) by ^{19}F Two-dimensional J-Resolved Spectroscopy', *Macromolecules*, **22**, 151–155 (1989).
26. M.D. Bruch, 'Microstructural Analysis of Poly(vinyl fluoride) by Fluorine-19 Two-dimensional J-Correlated NMR Spectroscopy', *Macromolecules*, **17**, 2547–2551 (1984).
27. I.-S. Chang, B.L. Hawkins, G.E. Maciel, G.E. Myers, 'Nitrogen-15 Nuclear Magnetic Resonance of Cured Urea-formaldehyde Resins Using Cross-polarization and Magic-angle Spinning', *Macromolecules*, **18**, 1482–1485 (1985).
28. L.E. Bogan, Jr, S.K. Wolk, 'Synthesis and Assignment of Carbon-13 NMR Resonances to *m*-Cresol Novolak Resins', *Macromolecules*, **25**, 161–165 (1992).
29. E. Fukushima, S.B.W. Roeder, *Experimental Pulse NMR: A Nuts and Bolts Approach*, Addison-Wesley, Reading, MA, 1981.
30. B.A. Higgins, P.D. Ellis, '^{27}Al Nuclear Magnetic Resonance of Aluminas and their Surfaces', *J. Am. Chem. Soc.*, **114**, 2098–2108 (1992).
31. M.E. Merritt, J.M. Goetz, D. Whitney, C.-P.P. Chang, L. Heux, J.L. Halary, J. Schaefer, 'Location of the Antiplasticizer in Cross-linked Epoxy Resins by ^{2}H, ^{15}N, and ^{13}C REDOR NMR', *Macromolecules*, **31**, 1214–1220 (1998).

Rheology in Coatings, Principles and Methods*

Richard R. Eley
ICI Paints, Strongsville, USA

1 Introduction **2**
2 Basic Definitions **2**
2.1 Rheology 2
2.2 Deformation (Strain) 2
2.3 Strain Rate 2
2.4 Stress 3
2.5 Viscosity 3
2.6 Modulus 3
2.7 Units 3
3 General Classes of Fluid Behavior **4**
3.1 Newtonian Fluids 4
3.2 Non-Newtonian Fluids 4
4 Classes of Non-Newtonian Behavior **4**
4.1 Shear-dependent Viscosity 4
4.2 Shear-thinning Fluids 6
4.3 Shear-thickening Fluids 7
4.4 Time-dependent Fluids 8
4.5 Plastic (Yield) Behavior 11
4.6 Elastic Liquids (Linear Viscoelasticity) 12
5 Rheology of Dispersions **17**
6 Rheology of Film Formation **19**
6.1 Sagging 20
6.2 Leveling 20
7 Rheology Instrumentation **21**
7.1 Industrial Viscometers 21
7.2 Absolute Viscometers 21
8 Rheometry: Experimental Methods **22**
8.1 Rotational Rheometry 22
8.2 Oscillatory Shear 26
List of Symbols **28**
Abbreviations and Acronyms **28**
References **28**

Rheology by definition is the study of the deformation and flow behavior of materials. An operational definition would be "the study of the response of certain materials to the stresses imposed on them". Rheology seeks to understand the relationship between applied force, or stress, and the resulting deformation, particularly for materials showing nonsimple responses. Such a relationship expressed mathematically is termed a constitutive equation. Perhaps the most familiar such equation is Hooke's law for elastic solids, which states the linear relationship between the stress σ and the deformation or strain γ [Equation 1]:

$$\sigma = G\gamma \qquad (1)$$

where the modulus G *is the constant of proportionality. However, rheology is primarily concerned not with the solid state but with the study of fluid behavior, and the simplest constitutive relation for a liquid is Newton's law [Equation 2]:*

$$\sigma = \eta\dot{\gamma} \qquad (2)$$

where the coefficient of viscosity or, more commonly, the viscosity η *expresses the proportionality between stress and, now, not strain but the strain rate,* $\dot{\gamma}$.

For the ideal, simple materials described above, G *and* η *are called material constants because they are dependent only on the thermodynamic variables of temperature, pressure, and concentration, similar to a material property such as the density. However, rheology is not concerned with either Hookean solids or Newtonian liquids. It is rather concerned with materials whose constitutive relationships involve complex behavior. For such materials,* G *and* η *are no longer material constants but material functions. Their values will now depend on variables such as the magnitude of applied stress, strain, strain rate, and even the deformation history. In fact, for rheologically interesting materials, the behavior is somewhere between that of a Newtonian liquid and an elastic solid. Such materials are termed non-Newtonian, and may display time-dependent, plastic and viscoelastic behaviors.*

Paints and industrial coatings, creams and lotions, inks, adhesives, ceramic slips, solder pastes, foods, medicines, etc. are representative of the range of materials whose commercial viability depends on having the correct rheology for the application. In turn, the required rheological properties must be defined with reference to the specific process at hand. This article introduces the science of rheology, with description of basic terms, models, and methods used. The mechanisms responsible for complex rheological behavior and the use of rheological measurements to understand and control the flow of materials are described. The emphasis is on coatings and coatings processes as points of reference, but the principles discussed apply to a wide range of materials and industrial products.

* Portions of this article are excerpted, by permission, from the chapter 'Rheology and Viscometry' by R.R. Eley, in the *Paint and Coatings Testing Manual: 14th Edition of the Gardner-Sward Handbook* (1995), ASTM, 100 Bar Harbor Drive, West Coshocken, PA 19429-2959.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

This article presents a brief review of principles and methods of rheological analysis, directed toward the coatings industry. The title of the article mentions coatings, but this could also include adhesives, inks, and sealants. These products share a common task: they must be applied to a substrate and function as a thin film. This process of application and film formation obviously requires not only a large total deformation, but also a high degree of control of flow, to achieve success. Flow cannot be controlled unless it can be properly measured. A persistent fault of industrial viscosity tests is that the methods do not closely relate to actual coating processes. This deprives the test of relevance to product performance. Moreover, the majority of commercial fluid products are non-Newtonian, and the rugged but simple viscometers commonly used in the industrial laboratory are generally not well suited for the characterization of non-Newtonian fluids. The objective for the applied rheologist, therefore, is to develop methods of rheological characterization that:

1. yield accurate data for complex fluids;
2. are related specifically to the critical processes that paints must undergo.

To meet the latter goal requires characterization methods that cover a wide range of stresses and timescales.

Even though all other properties be acceptable, a coating will usually not meet with success if the rheology is not. Experienced formulators say that more than half the cost of new product development is consumed in getting the rheology right. Furthermore, apparently minor changes in a raw material or process can cause significant and unexpected variability in product rheology, which naturally requires urgent solution. For all these reasons, rheological analysis is a vital and cost-effective tool for the coatings industry.

2 BASIC DEFINITIONS

2.1 Rheology

The term rheology derives from the Greek ρειν, ''to flow''. Rheology is defined classically as ''the study of the deformation and flow of matter'', and operationally as ''the study of the response of certain materials to stress''. In order to quantify the deformation and flow behavior of materials, three basic terms must be defined. The first two of these relate to the measurement of deformation (strain and strain rate) and the third to the measurement of the force necessary to deform the material (stress).

2.2 Deformation (Strain)

A deformation (or strain) is a change in shape and/or volume of a material in response to an applied stress. There are two main types of deformation: reversible (elastic, rubber-band-like), and irreversible (inelastic or viscous). Irreversible deformation is the definition of flow. Rheologically interesting materials display a third type, in which either or both a viscous and an elastic response to stress can be seen, depending on the rate or the duration of applied stress.

During the processing and end-use of a coating the deformation history can be complex. The most important types of deformation it will experience are simple shear and extension. (Extensional rheometry will be excluded from this presentation, because (a) a very limited number of studies on paint extensional rheology exist, and (b) in the author's opinion, there are no commercial instruments well suited to extensional characterization of paints.) For purposes of definition, the discussion is limited to simple shear. Simple shear deformation is exactly analogous to the spreading of a deck of playing cards, each card representing a thin volume element displaced relative to its nearest neighbor (Figure 1). If a force F is applied to the uppermost volume element (thickness dy), the material will deform by the displacement dx of adjacent volume elements. The total thickness is Δy, and the total displacement is Δx. The shear strain γ is the ratio of the net displacement Δx to the distance between the confining surfaces Δy [Equation 3]:

$$\gamma = \frac{\Delta x}{\Delta y} \tag{3}$$

2.3 Strain Rate

Unlike solids, such as a rubber band, liquids cannot support a constant deformation or strain. In order to measure the viscosity or resistance to flow of a fluid we therefore do not measure the deformation, but instead

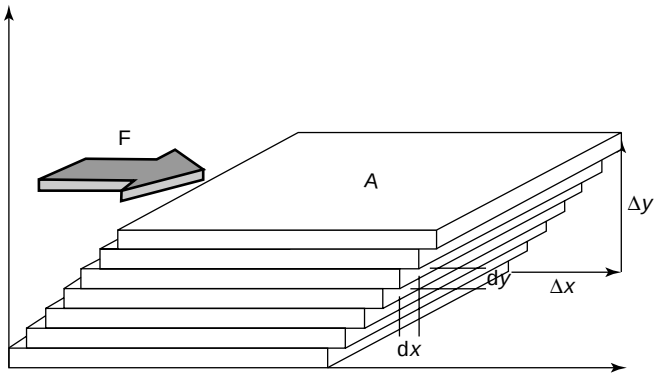

Figure 1 Term definitions for simple shear.

the rate of deformation (strain rate). The strain rate is the change in strain per unit time or the time derivative of the strain, symbolized $\dot{\gamma}$ (where the dot signifies "time derivative of"). Therefore, Equation (4) holds:

$$\dot{\gamma} = \frac{d\gamma}{dt} = \frac{d(dx/dy)}{dt} = \frac{d(dx/dt)}{dy} = \frac{dv}{dy} \tag{4}$$

If the deformation is simple shear, the strain rate is called the shear rate. The dimensionality of $\dot{\gamma}$ is $LT^{-1}L^{-1}$ (e.g. $cm\,s^{-1}\,cm^{-1}$). Unit cancellation leaves reciprocal time (s^{-1}) as the unit of shear rate. It is conceptually helpful to remember that, as shown by Equation (4), the shear rate is actually a velocity gradient, dv/dy (change in velocity per unit gap between shearing surfaces).

2.4 Stress

A force applied to a material creates a state of stress within the material. Stress can be expressed in units of force per unit area (e.g. dyne cm^{-2}) or, equivalently, energy per unit volume (e.g. erg cm^{-3}). In terms of Figure 1, the shear stress σ is the force F necessary to maintain steady shearing motion against the resistance of the fluid, divided by the area of the shear face, A [Equation 5]:

$$\sigma = \frac{F}{A} \tag{5}$$

The stress is determined not only by the magnitude of the applied force, but also by the area over which it is applied. Stress can therefore be thought of as a kind of force density, or the "concentratedness" of the force. Consider the analogy of a weight applied to a mattress. The depth to which the weight sinks into the mattress depends on the area over which the force is applied. The smaller the area of contact of the weight with the mattress, the higher the stress, and the greater the resulting strain, or the deeper the weight sinks.

2.5 Viscosity

The measure of the resistance to flow of a fluid is the viscosity. The resistance to flow is caused by friction between the flow units of the fluid (e.g. molecules), or may also be due in part to forces of attraction between the flow units. Thus, a highly viscous fluid (one reluctant to flow) may be so because of high molecular weight (as in motor oil) or may be of relatively low molecular weight, but have strong intermolecular interactions (e.g. hydrogen bonds, as between sugar molecules in honey).

The interaction of molecules in flow dissipates energy, chiefly as frictional heat. Flow, therefore, is a process that costs energy, of which the viscosity is a measure. For the case of shear deformation, the viscosity η is calculated as the ratio of shear stress to shear rate. As stated above, the shear stress represents the energy input per unit volume to achieve flow. The viscosity, therefore, is the energy dissipated per unit volume per unit velocity gradient [Equation 6]:

$$\eta = \frac{\sigma}{\dot{\gamma}} \tag{6}$$

Note that the rate of viscous energy dissipation (the power per unit volume) is the viscosity times the square of the shear rate, $\eta\dot{\gamma}^2$. The latter is the cause of viscous heating in fluids under shear.

2.6 Modulus

Materials comply with an applied stress by deforming, or undergoing strain. For ideal Hookean materials, the strain will be proportional to the applied stress. The modulus is the proportionality constant between the stress and strain. For example,

$$G = \frac{\sigma}{\gamma} \tag{7}$$

where in Equation (7) the shear modulus G is equal to the ratio of the shear stress and shear strain. Most polymeric materials and coatings systems in the solid state are non-Hookean, i.e. the modulus is not a material constant, but will depend on both rate and extent of deformation.

2.7 Units

Until recently, the most common system of units for rheological terms was the cgs (centimeter–gram–second) or "small metric" system. However, international convention now specifies Système International (SI) units

Table 1 Units in various systems

Variable	cgs	mks	SI
Strain	dimensionless	dimensionless	dimensionless
Strain rate	s^{-1}	s^{-1}	s^{-1}
Stress	dyne cm^{-2}	newton m^{-2}	pascal (Pa) (=1 newton m^{-2})
Viscosity	poise (P) (=1 dyne s cm^{-2})	–	Pa s (=10 P)
	centipoise (cP) (=0.01 P = 1 m Pa s)		m Pa s ($=10^{-3}$ Pa s = 1 cP)
Modulus	dyne cm^{-2}	newton m^{-2}	Pa

as the standard. The SI system is based on the large metric, or mks (meter–kilogram–second) units, with some additional named units. The units associated with the above variables according to the various systems are given in Table 1.

3 GENERAL CLASSES OF FLUID BEHAVIOR

3.1 Newtonian Fluids

It was postulated by Newton that the force resisting flow of a liquid is proportional to the rate at which it is caused to flow. Restated in terms of the above-defined quantities, the stress is proportional to the strain rate and the constant of proportionality is η, the coefficient of viscosity or, simply, the viscosity:

$$\sigma = \eta\dot{\gamma} \tag{8}$$

Equation (8) is the simplest example of a constitutive relationship, a model relating an applied stress and the response of a material. Fluids that obey Equation (8) over a range of shear rate are said to be Newtonian over that range. The viscosity of a Newtonian fluid is a material constant and depends only on the thermodynamic variables of temperature, pressure, and concentration. If the viscosity of such a material is measured at any shear rate or shear stress, the viscosity under all conditions of deformation is known (excluding turbulent flow).

3.2 Non-Newtonian Fluids

Commercial fluid products comprise a variety of materials, with a wide range of consistencies. In general, polymer solutions, emulsions, colloidal dispersions, and other suspensions of particulate solids will be non-Newtonian. For non-Newtonian materials, the viscosity is no longer a material constant, but is now a material function – in this case a function of the shear rate (or shear stress). For non-Newtonian fluids, a viscosity measured at a single shear rate is not an adequate representation of the rheology of the system.

4 CLASSES OF NON-NEWTONIAN BEHAVIOR

4.1 Shear-dependent Viscosity

Materials for which the viscosity falls with increasing shear rate are said to be shear thinning. (Any descriptions of behavior stated in terms of shear rate-dependence may equally well be stated in terms of shear stress dependence.) Simple shear thinning without either time dependence or a yield stress is termed pseudoplasticity. Viscosity rising with increasing rate of shear is called shear thickening. The term ''dilatancy'' is often applied to shear-thickening behavior, although this refers strictly to shear thickening accompanied by a volume increase, as the term implies. Figure 2 shows curves illustrating viscosity–shear rate relationships for Newtonian and non-Newtonian fluids.

Equation (8) describes simple fluid or Newtonian behavior, for which the viscosity η is a material constant, independent of shear rate (contour N, Figure 2). As flow behavior becomes more complicated, a more complex expression is needed to model it. First, to the Newtonian model we will add a yield stress (section 4.5), giving the Bingham equation:[1]

$$\sigma - \sigma_0 = \eta_p\dot{\gamma} \tag{9}$$

Equation (9) says that, when the applied stress σ is greater than the yield stress σ_0, the difference $\sigma - \sigma_0$ is directly proportional to shear rate. Note that the Bingham model and others containing a yield stress term describe flow behavior only when $\sigma > \sigma_0$. That is, if $\sigma \leq \sigma_0$, $\dot{\gamma} = 0$ (i.e. $\eta = \infty$), and there is no flow. Contrary to what is often stated, this does *not* correspond to Newtonian behavior above the yield stress. As illustrated in Figure 2 the Bingham model (curve B) displays shear-thinning behavior (when $\sigma > \sigma_0$), but the viscosity approaches the plastic viscosity η_p as a limit at higher shear rate. Casson[2] derived an equation similar to Bingham's for the flow of a dispersion, except all terms are present as the square

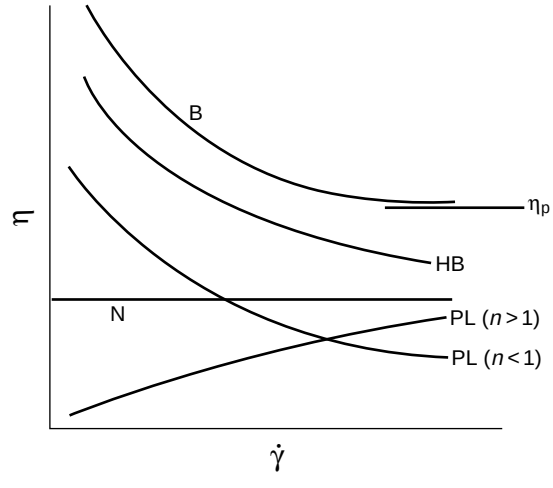

Figure 2 Viscosity–shear rate curves illustrating flow behavior according to various flow models: N = Newtonian; B = Bingham (η_p = plastic viscosity); HB = Herschel–Bulkley; PL = power law (n = flow index).

root [Equation 10]:

$$\sigma^{1/2} - \sigma_0^{1/2} = \eta_\infty^{1/2}\dot{\gamma}^{1/2} \tag{10}$$

Casson's equation is reputed to fit a variety of paint systems, particularly as modified by Asbeck[3]

$$\eta^{1/2} - \eta_\infty^{1/2} = \sigma_0^{1/2}\dot{\gamma}^{-1/2} \tag{11}$$

In fact, the author's experience is that the Casson model does not represent paint flow as well as certain other models (see below). The term η_∞ in Equations (10) and (11), sometimes called the Casson viscosity, is not literally an ''infinite-shear viscosity'' but is a limiting value that is approached at a relatively high shear rate. The analysis of flow data using these models is simply an exercise in curve fitting, and the value obtained for η_∞ can vary depending on the completeness of the dataset used in the analysis. The user should be aware that the parameters resulting from a Casson analysis, as with all such models, are not necessarily true material constants. For example, a finite yield stress will generally be obtained whether or not the material really possesses yield behavior.

Next, we will let the exponent of the shear rate in the Newtonian law to be other than unity. The shear stress is now proportional to the shear rate to the power n:

$$\sigma = K\dot{\gamma}^n \tag{12}$$

Equation (12) is known as the power law or Ostwald–deWaele model. The constant K, sometimes called the consistency, has replaced the coefficient of viscosity η. (In numerical magnitude, though not dimensionally, K is the viscosity at unit shear rate, $1\ s^{-1}$.) This is necessary because the exponent n can be other than unity, in which case K will not have proper viscosity units associated with it. The power law exponent n is sometimes called the flow index because a value of $n < 1$ corresponds to shear thinning behavior and $n > 1$ to shear thickening (Figure 2 curves PL). Dividing Equation (12) through by $\dot{\gamma}$ yields the viscosity power law [Equation 13]:

$$\eta = K'\dot{\gamma}^{n-1} \tag{13}$$

Substituting $n = 1$ in Equations (12) and (13), yields back the Newtonian law.

Adding a yield stress term to the power law expression gives the Herschel–Bulkley relationship [Equation 14]:

$$\sigma - \sigma_0 = K\dot{\gamma}^n \tag{14}$$

which describes power law behavior above the yield point (curve HB, Figure 2). Lapasin and Torriano[4] showed the Herschel–Bulkley model to be superior to a number of others in fitting the flow behavior of a wide variety of coatings materials over a useful range of deformation conditions.

4.1.1 Generalized Equilibrium Flow Curve

The generalized equilibrium flow curve (Figure 3) represents the general features of the shear rate dependence of viscosity for non-Newtonian fluids (after Hoffman,[5] Krieger,[6] and Choi[7]). The condition ''equilibrium flow'' means that any time-dependent or relaxation effects have been experimentally removed. Figure 3 consists of:

1. a low shear rate Newtonian regime, region I;
2. a shear-thinning regime, the power law regime, region II;
3. a high-shear Newtonian regime, region III;
4. a possible shear-thickening regime, region IV.

The figure is explained in more detail in section 4.2. The chief shortcoming of the power law models is that their range of validity is limited to region II. They cannot account for the upper or lower Newtonian regions, and predict infinite viscosity at zero shear rate and zero viscosity at infinite shear rate, both unrealistic limiting behaviors. However, the power law models (Ostwald–DeWaele, Casson, Herschel–Bulkley) are useful within their limitations, particularly for engineering-type calculations. More elaborate models are required to fit the generalized non-Newtonian curve beyond the power law region. A simple extension of the power law model is to add an upper Newtonian limiting viscosity, η_∞:

$$\eta = K'\dot{\gamma}^{n-1} + \eta_\infty \tag{15}$$

Equation (15) is known as the Sisko model,[8] and includes regions II and III of Figure 3.

Modern rheometers have made the low-shear Newtonian plateau viscosity η_0 more experimentally accessible, and models incorporating region I are quite useful. Perhaps the two most successful of several proposed models encompassing regions I, II, and III inclusively are the

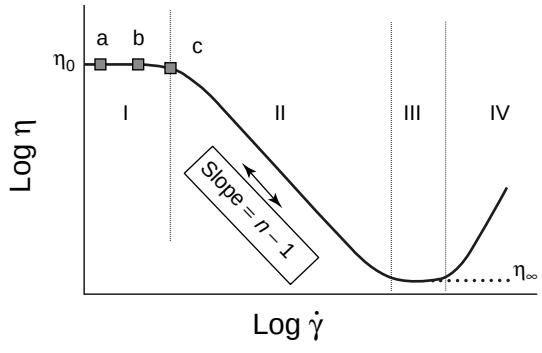

Figure 3 Generalized non-Newtonian equilibrium flow curve. η_0 = zero-shear viscosity; η_∞ = high-shear limiting viscosity; region I = first Newtonian plateau; region II = power law regime; region III = second Newtonian plateau; region IV = shear-thickening behavior. (Adapted from Choi.[7])

Cross[9] and Carreau[10] models. Hieber and Chiang[11] have written a general form of which the Cross and Carreau models are special cases (modified here to include region III; Equation 16):

$$\eta = \eta_\infty + \frac{\eta_0 - \eta_\infty}{(1 + [\beta\dot{\gamma}]^a)^{(1-n)/a}} \tag{16}$$

where η_0 is the first (lower) Newtonian plateau viscosity and η_∞ the second (upper) Newtonian viscosity. The constant a determines the curvature of the transition region between the first Newtonian regime and the power law regime. The value of a can be a measure of the breadth of the molecular weight distribution of a polymer[11] or the particle-size distribution of a colloidal dispersion. Setting $a = 1 - n$ in the above expression yields the Cross equation, whereas $a = 2$ gives the Carreau-B model. The parameter n has the identical meaning as in the simple power law model [Equations 12 and 13], i.e. it is the slope of the power law region in a log–log plot of shear stress versus shear rate. The constant β has the dimension of time and is a time constant characteristic of the available mechanisms permitting the system to accommodate to a shear field. It may thus be related to the diffusional or rotational relaxation time of the flow units (e.g. colloidal particles or aggregates) or perhaps to the time for rupture of particle flocs or aggregates under shear. The characteristic shear rate of the transition from the initial Newtonian plateau (region I) to the shear-thinning regime (region II) in Figure 3 is determined by the value of β [Equation 17]:[11,12]

$$\dot{\gamma}_{tr} = \frac{1}{\beta} \tag{17}$$

For colloidal systems, β is the time constant for Brownian diffusion. The Stokes–Einstein equation leads to the value of β for a spherical-particle dispersion [Equation 18]:

$$\beta = \frac{6\pi\eta R^3}{k_B T} \tag{18}$$

where η is the viscosity of the continuous phase, R is the particle radius, k_B is the Boltzmann constant, and T is the absolute temperature. It should be said that Equation (18) is quantitatively correct only for very dilute dispersions. However, this expression is qualitatively useful for rationalizing the behavior of more concentrated dispersions that display Brownian motion. Here, β is the time to diffuse a distance equal to a particle radius.[13]

Shear thinning involves a transition from Brownian to hydrodynamic dominance of the microstructure. (This event corresponds to the point c in Figure 3.) This transition from Newtonian to shear thinning can be expressed in terms of the Peclet number, *Pe*, a dimensionless shear rate [Equation 19]:[14]

$$Pe = \frac{6\pi\eta\dot{\gamma}R^3}{k_B T} \tag{19}$$

The numerator represents the viscous resistance to particle diffusion and the denominator the thermal energy driving Brownian motion, so that the transition from the Brownian to the hydrodynamic regime occurs as *Pe* exceeds unity. It is useful, however, to express this transition in dimensional terms, and the dimensional shear rate of transition is $1/\beta$ [Equation 17].

Equations (17) and (18) suggest ways of controlling the onset of shear thinning for a dispersion. Any change that increases the value of β (such as increasing the effective particle size, the continuous-phase viscosity, or lowering the temperature) will reduce the onset of shear thinning to lower shear rates. Decreasing β extends Newtonian behavior to higher shear rates. Although Equation (18) is strictly valid only for very dilute dispersions, it still provides qualitative guidelines for manipulating the rheology of dispersions. For concentrated dispersions, η can be taken as the viscosity of the dispersion.[16,17]

A broad particle-size distribution will possess a range of β values, a corresponding range of $\dot{\gamma}_{tr}$ values, and consequently the transition to shear thinning will be spread out over a wide range of shear rates. Thus, the effect of polydispersity in the flow unit size distribution will be to make more gradual the transition from region I to region II. The mean shear rate of transition corresponds to a mean diffusion relaxation time, whereas the value of a in Equation (16) will vary according to the *distribution* of relaxation times. Thus, a colloidal dispersion having a wide particle size distribution will exhibit a broad, gradual transition from region I to region II. A similar effect will be seen in a polymer melt of broad molecular-weight distribution.

4.2 Shear-thinning Fluids

Fluids that show decreasing viscosity with increasing shear rate (or shear stress) are often called pseudoplastic, an obsolescent term that implies shear-thinning behavior without yield stress. Currently, the more descriptive (but longer) phrase ''time-independent shear thinning'' is used to refer to this behavior. Particulate dispersions, polymer colloids, and polymer solutions can display this behavior above a concentration threshold. Pseudoplastic fluids display shear thinning without time dependence, meaning that at a constant shear rate and temperature the viscosity also remains constant. In a ramp experiment (discussed below), the up-ramp and down-ramp curves for such a material will be coincident. This behavior implies that the fluid microstructure is always at equilibrium despite

conditions of changing shear rate. In other words, the microstructure can adjust to a change in shear rate faster than the shear rate is changing. Put another way, the microstructural relaxation time is faster than the time-rate-of-change of the shear rate $\ddot{\gamma}$. Thus, the microstructure effectively equilibrates instantaneously when the shear rate is changed. Therefore, one always measures the equilibrium viscosity for such fluids. Of course, it may be possible to increase the shear rate rapidly enough (as by a shear rate step-jump transient experiment) to observe the microstructural relaxation time for a given system.

4.2.1 Mechanisms of Shear Thinning

Two principal mechanisms for shear thinning in dispersions are:

1. For unstable (flocculating) systems, shear thinning results from the break-up of floc structures, releasing interstitially trapped liquid phase, reducing the effective volume fraction (section 5), hence lowering the viscosity. Once flocs have broken down, shear thinning can proceed further by mechanism (2).
2. For stable (nonflocculating) colloidal systems, shear thinning results from the progressive domination of microstructure-randomizing Brownian motion by hydrodynamic shear forces. This induces ordering of particles in strings (ordered layers of particles in concentrated dispersions), which is a more efficient-flowing microstructure. Hydrodynamic ordering also leads to shear thinning for noncolloidal (non-Brownian) systems.

The following discussion applies to nonflocculating systems and excludes, for the present, the case of flocculated systems, which are addressed in section 4.4.2.

The viscosity is a measure of the energy cost to flow. Therefore, shear thinning implies a change in the microstructure that allows more efficient flow, consequently with less energy dissipation. The mechanism of this involves a shear-induced increase in order, or anisotropy, within the system. Thermal (Brownian) motion tends to keep systems disordered (of random order). Shear forces work against this, tending to impose order. For particulate dispersions, this means that individual particles tend to align with the shear field, in a string-of-pearls fashion (Figure 4). This alignment tends to reduce the degree of frictional interaction between individual particles and the surrounding fluid, making the flow more energy efficient, hence reducing the viscosity. For the case of polymers in solution, polymer random coils tend to stretch in the direction of shear, and become ellipsoidal in shape, with the major axes aligned at a 45° angle to the flow direction. The result is a steadily decreasing

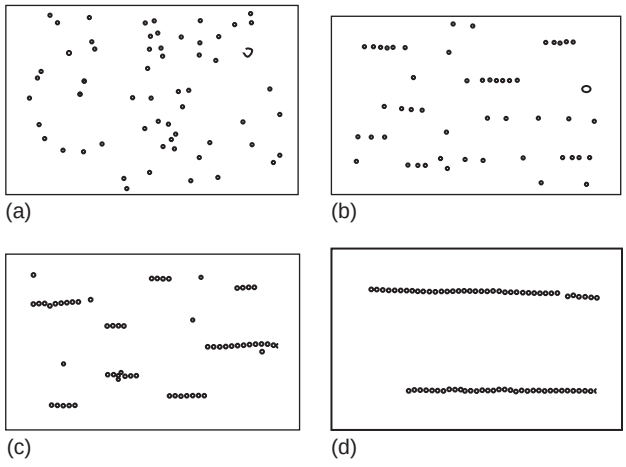

Figure 4 Suspension of glass microspheres in a viscoelastic polymer solution, confined in narrow gap between glass plates: (a) before shear, particles are randomly distributed; (b–d) after being sheared in a side-to-side manner with increasing vigor (higher shear rate). (By permission from J. Michel, R. Pätzold, R. Donis, *Rheol. Acta*, **16**, 317 (1977).)

viscosity with increasing shear rate as the degree of order increases. This describes region II in Figure 3. Ultimately, at high shear rates, the maximum amount of shear ordering possible is attained and the viscosity again becomes independent of shear rate (region III in Figure 3, the upper Newtonian plateau η_∞). Particle dynamics simulations confirm this description in a general way, although some of the details are still in question.[18] The general flow curve involves an initial constant-disorder regime (first Newtonian), followed by a disorder to order transition (to the shear-thinning regime), succeeded by a metastable constant-order regime (second Newtonian), and finally an order–disorder transition (shear thickening, regime IV). The onset of shear thickening occurs at progressively lower shear rates as the dispersion concentration increases,[19] with the consequence that one or more of the other regimes are obliterated.

If the dispersion is unstable (flocculated) the viscosity may or may not plateau at low shear. Something similar to the dashed curve in Figure 5 may be seen, the viscosity generally increasing as the shear rate is reduced.[7] Such systems may possess an apparent yield stress (section 4.5). (section 5 provides an additional discussion of particulate system rheology.)

4.3 Shear-thickening Fluids

It is described above how shear thinning results from a shear-induced increase in the order of a system, which allows the fluid elements to flow with minimum expenditure of energy. Conversely, shear thickening can result from a decrease in the order of a system.

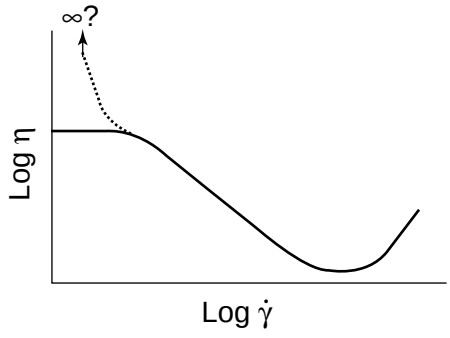

Figure 5 Representation (dotted line) of behavior of flocculated systems at low shear. Instead of a Newtonian plateau, the curve ascends toward possible yield stress (infinite viscosity).

A disordered system flows less efficiently, hence is more viscous. An example of this is the catastrophic increase in viscosity observed by Hoffman in concentrated polyvinylchloride dispersions,[5,20] postulated to result from the buckling or collapse of ordered, layered arrays of particles.

A common type of shear thickening behavior is dilatancy. The term has sometimes been applied loosely or generically to shear thickening, but dilatant behavior is, strictly, time-independent shear thickening accompanied by an increase in volume of the fluid (implying a decrease in packing efficiency or order). This specific behavior has been referred to as volumetric dilatancy and usually occurs in relatively concentrated disperse systems. In such a dilatant system, the disperse-phase particles are minimally wetted by the liquid continuous phase. At rest, the particles of the disperse phase are in a random close-packed structure, for which the interstitial volume is relatively minimal (Figure 6). If the dispersion in this state is deformed slowly, adequate time exists for the meager liquid phase to flow sufficiently to maintain the dispersed phase in a wetted state, and the system is able to nearly maintain its close-packed structure. Faster deformation causes a liquid-starved condition because the interstitial volume increases when the system is deformed or made to flow (Figure 6). There is no longer enough liquid to lubricate the system. The particles are, therefore, incompletely wetted, may come into direct contact, and further flow would ultimately create microscopic voids leading to fracture of the material. The surface of a dilatant material may appear dry when stressed, due to the withdrawal of surface liquid into the increased interstitial volume. This is seen when walking on wet sand on the beach. The resistance to deformation of the material can increase sharply due to these effects.

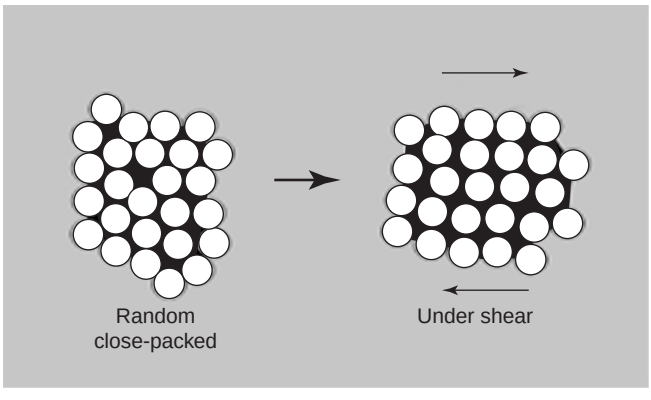

Figure 6 Dilatant behavior (shear thickening with volume expansion). Random close-packed structure gives way to less-efficient packing, with increase in interstitial volume.

The mechanism for shear thickening described above clearly applies to true dilatant behavior, such as can be observed in wet sand and in cornstarch dispersions. For such systems, dilatancy appears when the system at rest is disturbed. Shear thickening occurring in flowing systems with increasing shear rate has been termed rheological dilatancy, and the mechanism here is probably somewhat different. Hoffman's postulated collapse of particle-layered structures has been challenged by more recent neutron-scattering experiments as well as by computer particle dynamics simulations.[18] These studies indicate that shear thickening is the result of particle cluster formation, occurring when hydrodynamic forces overcome stabilizing forces, allowing the formation of shear-induced aggregates.[21]

Dilatancy is very sensitive to dilution or warming. Also, introducing polydispersity, adding flocculants (preventing random close packing) or electrolyte (reducing the effective volume fraction of the dispersed phase, section 5) can alleviate undesirable dilatancy. Flow curve measurement provides an excellent means of quantifying the degree of reduction of dilatancy achieved by these measures.

4.4 Time-dependent Fluids

For time-dependent fluids, the viscosity depends on the shear history. That is, the viscosity of time-dependent fluids will depend on both shear rate (or shear stress) and time. Time-dependent shear thinning is termed thixotropy, and is common in paints and other flocculated or otherwise structured fluids.[22] Its rarer opposite, time-dependent shear thickening, has been called rheopexy, though the preferred term is antithixotropy.

4.4.1 Characteristics of Thixotropy

If a thixotropic fluid is subjected to a constant shear rate and temperature the viscosity will fall, eventually reaching a constant value. If the shear rate is then changed, the viscosity will approach a new equilibrium

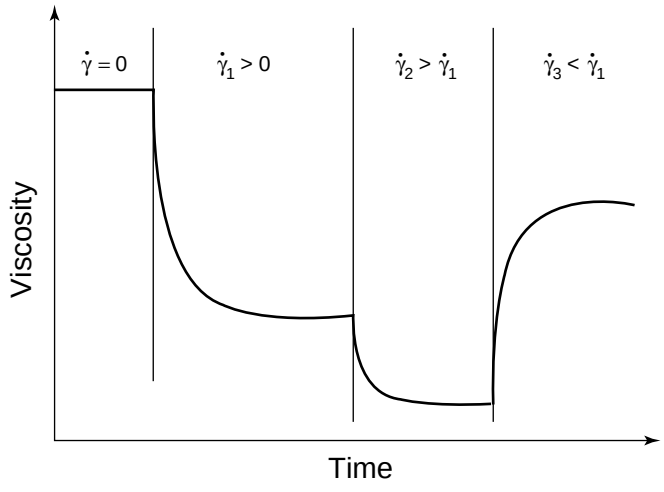

Figure 7 Time-dependent shear-thinning (thixotropic) behavior, with $\eta_{\dot{\gamma}=0}$ = high or infinite. Subsequent step changes in shear rate cause re-equilibration of structure accompanied by exponential approach to equilibrium viscosity.

value at a characteristic rate of approach. This behavior is illustrated in Figure 7, showing the way viscosity changes for a time-dependent system when the imposed shear rate (or shear stress) is changed in a series of steps. In Figure 7 the shear rate is stepped from zero to a finite value ($\dot{\gamma}_1$), following which the viscosity drops exponentially, reaching an equilibrium value. If the shear rate is increased further to $\dot{\gamma}_2$ the viscosity drops again to a new, lower equilibrium value. Now the shear rate is decreased to $\dot{\gamma}_3$ resulting in an exponential rise of viscosity to a higher equilibrium value. If, instead of the multiple steady-shear experiment just described, a shear rate (or shear stress) ramp experiment were to be carried out (Figure 8a), a loop would be observed (Figure 8b). The explanation of the loop is given below, but has to do with the fact that in a ramp experiment the equilibrium structure is never attained. The circuit in Figure 8(b) is called a thixotropic loop. Roughly bisecting the loop is the equilibrium–structure curve generated from the equilibrium viscosity data of Figure 7.

4.4.2 Mechanism of Thixotropy

Thixotropy is defined as "an isothermal, reversible sol–gel (liquid–solid) transformation".[23] It is experienced as a viscosity that is both time- and shear-rate (or stress) dependent. Its origin is the breakdown of internal fluid structure to smaller flow units under shear or the reassembly of structure from smaller units, when stress is relaxed. The cause of time dependence is that both breakdown and recovery processes are slow, relative to the rate of change of stress, or to the time of observation at constant stress.

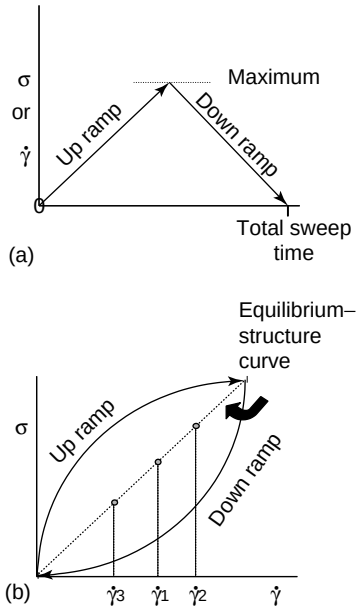

Figure 8 (a) Shear rate or shear stress ramp experiment (nonequilibrium flow curve). The shear rate or stress increases to a selected maximum value then decreases, linearly (or logarithmically). Variables are the ramp rate ($\ddot{\gamma}$ or $\dot{\sigma}$) and the ramp duration (sweep time). (b) Thixotropic loop. Equilibrium–structure curve from data in experiment illustrated in Figure 7.

In simple shear thinning or pseudoplastic systems, rapid structural equilibration quickly accommodates to changes in shear rate. The accommodation process essentially is particle diffusion, which is quite rapid for submicron particles (section 5.4.1). The viscosity (a measure of microstructure) thus always keeps up with changes in shear rate for such systems. In contrast, for thixotropic systems, the rate of structural re-organization is slower than the experimental rate of change of shear rate ($\ddot{\gamma}$, Figure 8a). In thixotropic systems, a viscosifying structure exists at rest, and the viscosity falls because of the collapse of that structure under shear. This viscosifying structure may consist of, for example, flocculated particulates (flocculated systems will be higher in viscosity than a deflocculated system of the same composition, section 5), an associated network, or may be due to polymers interacting through hydrogen bonds. The necessary attribute of the structure is a significantly long characteristic time (seconds) for breakdown and reassembly. Consequently, on the up side of a shear rate ramp experiment (Figures 8a and 8b), the structural breakdown lags behind the ever-increasing shear rate, so that the viscosity on the up ramp will be higher than the equilibrium viscosity at any given shear rate. On the down side of the ramp, the rebuilding structure lags the reducing shear rate, so that the measured viscosity is always lower than the equilibrium value. Thus, a thixotropic loop is seen, the size of which

should be proportional to the structural time constant, for a given ramp rate and duration. Holding the shear rate constant allows the structural reorganization to catch up until equilibrium between structural breakdown and recovery rates is achieved (as in Figure 7).

4.4.3 Thixotropy Test Methods

4.4.3.1 Thixotropic Index Test The thixotropic index test is very widespread in practice, but a misnomer. The test determines the ratio of viscosities measured at arbitrary high- and low-shear conditions. ASTM test method D2196 specifies taking the ratio of viscosities measured on the Brookfield Synchro-Lectric viscometer at two speeds, representing a 10-fold speed ratio. The test measures only equilibrium viscosities, and therefore provides no information about time dependence and should be called the ''shear-thinning index'' (STI) instead (it is so named in ASTM D2196). It should be emphasized that thixotropic behavior has two important aspects: shear thinning and time dependence. However, it is the time dependence that makes thixotropy valuable to the formulator. Hence, it is important to correctly characterize the time-dependent aspect of the behavior. For this reason, the so-called thixotropic index test is of limited benefit for characterization of such systems.

4.4.3.2 Thixotropic Loop Method In the thixotropic loop test, a programmed increase and decrease of shear rate (or shear stress) is applied to a thixotropic material (as in Figure 8a), resulting in the generation of a loop (Figure 8b). The size of the loop can be obtained by numerical integration and is taken as an indicator of the rate of structural breakdown and recovery for a given ramp time (the larger the loop, the slower the breakdown-recovery rates). This method is a difficult one to do correctly, however, and is more complex than it at first appears. The loop size depends on the shear history of the sample (time lapse since previous deformation, manner of sample handling and mounting), the shear rate (or shear stress) maximum in the ramp experiment, and the ramp rate ($\ddot{\gamma}$ or $\dot{\sigma}$).

4.4.3.3 Step-shear Method In the step-shear test, a high shear rate (10^3–$10^4\,s^{-1}$ preferred) is applied to the material, allowing time for the viscosity (and hence the structure) to reach equilibrium, then the shear rate is suddenly reduced to a very low value (order of $1\,s^{-1}$; Figure 9). As reassembly of the fluid structure occurs at the low shear rate, the viscosity recovers in approximately exponential fashion. Useful physical parameters can be extracted from the data by fitting the following equation to the recovery curve:

$$\eta(t) = \eta(0) + [\eta(\infty) - \eta(0)][1 - e^{-t/\tau}] \qquad (20)$$

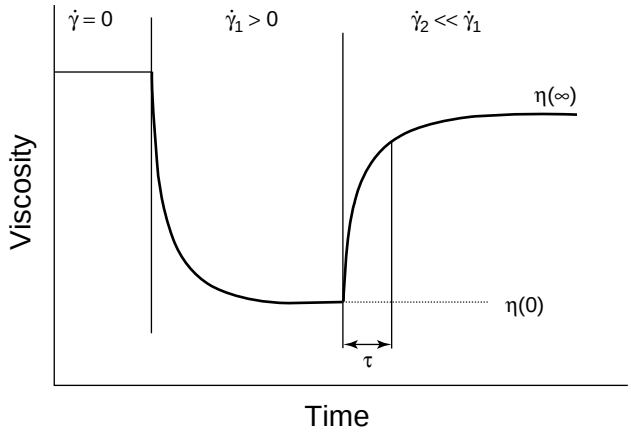

Figure 9 Step-shear method for characterization of thixotropy. The shear rate is stepped down from a high value $\dot{\gamma}_1$ (simulates application process) to a low value $\dot{\gamma}_2$ (emulating applied coating layer). The viscosity recovers from zero-time value $\eta(0)$, characterized by time constant τ, ultimately reaching equilibrium viscosity $\eta(\infty)$. The recovery parameter is the ratio $\eta(\infty)/\tau$.

In Equation (20) $\eta(t)$ is the viscosity as a function of time t, $\eta(0)$ is the sheared-out viscosity (at time zero), $\eta(\infty)$ is the infinite-time recovered viscosity level, and τ is the time constant characterizing the recovery rate. The ratio $\eta(\infty)/\tau$, is as the recovery parameter (a true thixotropic index), and has been found to correlate well to thixotropy-related properties such as sag resistance and air entrainment.[24] Table 2 gives data for two gel-coat systems, one that sagged excessively and one that did not. The conventional ''Thix Index'' results do not predict the sag behavior, and in fact are contrary to the observed behavior. The step-shear experimental parameters τ and $\eta(\infty)$ are given, together with their ratio. The nonsagging formula (with additive) has both a shorter recovery time and a higher final recovered viscosity, and the recovery parameter takes both of these into account to predict significantly better sag resistance for the with-additive material.

The step-shear test can be performed on an inexpensive viscometer (such as the Brookfield Synchro-Lectric or Wells-Brookfield Cone/Plate Viscometer) as well as on more sophisticated instrumentation. The latter offer

Table 2 Gel coat thixotropy test data comparison with step-shear data (τ = recovery time constant; $\eta(\infty)$ = equilibrium recovered viscosity; $\eta(\infty)/\tau$ = recovery parameter)

Composition	τ (s)	$\eta(\infty)$ (P)	$\eta(\infty)/\tau$ ($P\,s^{-1}$)	Thix index	Sag?
With additive	8.9	226	13	4.04	no
Without additive	18.2	97.3	5.4	5.24	yes

advantages, of course, particularly viscoelastic characterization capability. The preferred way to characterize thixotropic recovery would be to break down the structure under high, steady shear (as in the step-shear method), then step down to a small-amplitude (small strain) oscillatory-shear test, as described by Dodge.[25] The rebuilding of structure can then be followed by means of the viscoelastic parameters (Figure 10), which are sensitive probes of the fluid structure responsible for thixotropy. This method more realistically simulates processes occurring at low strain in the relatively quiescent coating layer after application. Figure 11 shows the recovery curves of the viscoelastic moduli G' and G'' (section 4.6.1) from such a step-shear experiment.

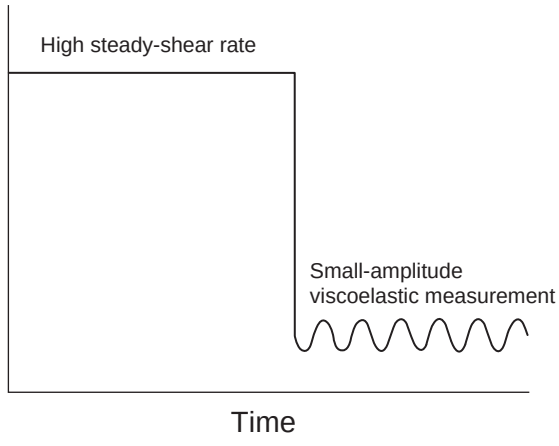

Figure 10 Measurement of thixotropic recovery by the step-shear method using small-strain oscillatory characterization of viscoelastic properties in recovery phase.

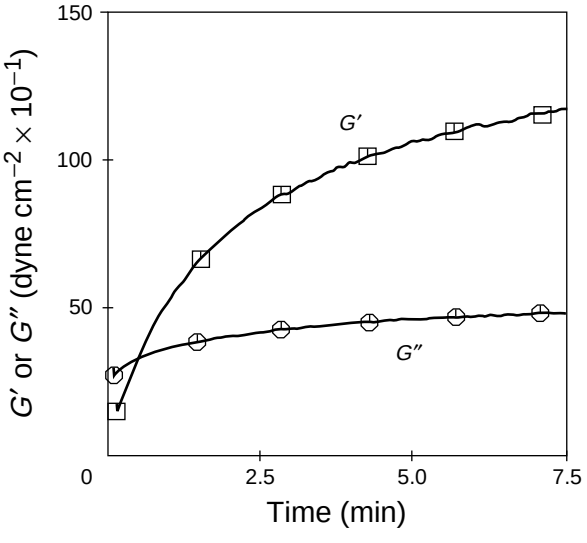

Figure 11 Experimental curves of G' (elastic modulus,) and G'' (viscous modulus, ○) versus time, illustrating viscoelastic characterization of thixotropic recovery.

4.5 Plastic (Yield) Behavior

A plastic material has the properties of an elastic solid until a critical stress is applied, whereupon it suddenly yields and becomes fluid. This critical stress is the yield stress, the minimum stress necessary to initiate flow. Ideal plastic behavior includes the following characteristics:

1. Hookean behavior below the yield stress, i.e. the steady-shear viscosity is infinite (or undefined) and the deformation is linear with stress;
2. the yield value is a material constant.

Metals typically exhibit ideal plastic behavior. Ironically, what we commonly call ''plastics'' (i.e. polymeric materials) do not. For most solid polymeric materials, deformation below the yield point is a combination of elastic strain and viscous flow; hence the measured yield stress is not a material constant but will depend on the deformation history of the material. In a similar way, structured fluids such as paints can exhibit yield-like behavior, apparently changing from solid-like to readily-flowing fluid when a critical stress is exceeded. However, the measured apparent yield stress is not single-valued, but will vary depending on experimental conditions. This is because the interparticle forces (secondary bonds) presumed responsible are of a range of types with a corresponding range of relaxation times.[26] Therefore, as stress is applied the material undergoes creep accompanied by viscous relaxation, such that some of the secondary bonds rupture before overall yield occurs. As a consequence, the measured yield value will depend on the rate at which the stress is increased up to the point where flow occurs: the faster the rate of stress increase, the higher the measured yield value, and vice versa. For this reason, yield stress values cited in the literature are meaningless unless the exact experimental methodology is provided.

Barnes and Walters[27] challenged the yield concept with the assertion that most materials with an apparent yield stress will be found, in reality, to have a high but finite viscosity if measured at sufficiently low stresses. Barnes recently published a comprehensive examination of the question.[28] The idea is that few, if any, structured fluids possess true yield behavior. Although in principle this view is probably correct, it is also true that, in practical terms, yield behavior can have important consequences for the processing, stability, and end-use of materials. There seems to be, from practical experience if not from clear experimental evidence, some sort of rather sudden viscosification (possibly some kind of flow discontinuity) that can occur at low stresses, the magnitude of which is apparently related to the number and strength of interparticle attractive forces.[26,29] The yield stress is therefore an engineering reality[30] that must be taken into account when formulating paints or dispersions. A

yield stress may be desirable or undesirable depending on the process in question. Materials with an apparent yield stress will often exhibit thixotropy and viscoelasticity as well.

4.5.1 Static versus Dynamic Yield Stress

The static yield stress is measured by the start-up of flow from rest conditions, by gradually increasing the stress until a finite viscosity is measured. The dynamic yield stress is obtained by reducing the stress on a fluid until it ceases to flow at some nonzero stress value.[31] The method to be used should be chosen with reference to a specific process. For example, for long-term suspension of solids or for start-up flow in pumps, the static yield stress is relevant. Post-application leveling and sag behavior would be governed by dynamic yield stress characteristics, including the kinetics of structural recovery (section 4.4).

4.5.2 Yield Stress Measurement Methods

4.5.2.1 Curve-fitting and Extrapolation Methods

Curve-fitting techniques may be used to obtain a value for the yield stress, by fitting the Bingham, Herschel–Bulkley, or Casson models, for example, to experimental flow data (section 4.1). However, such methods are very questionable because one is attempting to infer a property of the solid state from behavior of the fluid. Limitations of the curve-fitting method are evident in the fact that one usually gets a finite value for σ_0, even when the specimen is clearly a mobile fluid! This begs the question as to whether such extrapolated yield data are very meaningful. Even less satisfactory is the practice of extrapolation of a shear stress–shear rate curve to zero shear rate, taking an intercept on the stress axis as the yield value. Such a method is highly arbitrary and subjective, as well as again being based on the uncertain assumption that the yield stress is related to the fluid viscosity.

4.5.2.2 Start-up of Flow Methods

Penetrometer Method. Start-up-of-flow methods include the use of penetrometer-type instruments such as a thermomechanical analyzer, where increasing force is applied to a probe. The yield stress is determined as the break point of a force–penetration curve.[32]

Controlled-stress Rheometer Method. The controlled-stress (CS) rheometer is well suited to the characterization of the properties of structured fluids, including the yield stress. For this instrument, the applied torque is the independent (or controlled) variable, whereas the dependent (or measured) variable is the angular displacement. With this instrument, the yield stress can be directly measured by a stress ramp experiment. If the stress is gradually increased from zero or a very low value until flow is detected, the instrument then records this value as the measured yield stress. However, it should be remembered that the measured yield stress will be dependent on the experimental protocol, including shear history, ramp parameters, and the test geometry used.

Vaned Rotor Method. A still rather novel technique employs a special vaned rotor to remedy problems of slip in dense suspensions.[33,34] The vaned rotor consists of rectangular blades or vanes fixed to a rotating shaft. Advantages of this geometry are: (a) there is little disturbance of the sample when this type of probe is inserted; and (b) when the rotor turns, the material moves as a solid cylinder. Thus, the yield surface is within the material itself, avoiding problems of wall slip. Barnes and Carnali[35] showed that the vaned geometry can be used in place of the bob-and-cup geometry for flow curve measurement of shear-thinning fluids.

With the vaned rotor on a controlled-strain-rate (CR) type of rheometer (section 7.2.1.1), the yield stress is measured by monitoring the torque as the geometry is rotated at a constant low rate. The torque will go through a maximum, corresponding to concerted yielding along the virtual cylindrical surface defined by the vane radii. The yield stress is related to the maximum torque T_m according to Equation (21):[34]

$$\sigma_0 = \frac{T_m}{K} \tag{21}$$

where K is the vane geometric constant given by

$$K = \frac{\pi D^3}{2}\left(\frac{H}{D} + \frac{1}{3}\right. \tag{22}$$

In Equation (22) D is the rotor diameter and H is the vane height.

Using a CS rheometer, measurement of the yield stress using the vaned rotor is straightforward. The torque is ramped slowly and the stress at which a finite displacement is detected is reported as the yield stress.

4.6 Elastic Liquids (Linear Viscoelasticity)

When a purely viscous fluid is in steady flow, the energy of deformation dissipates as frictional heat. No potential energy is stored. Therefore, viscous flow is an irreversible deformation. Newtonian liquids show essentially ideal viscous (or ''inelastic'') behavior over a wide range of deformation rate. For ideal elastic (''Hookean'') substances, all the energy of deformation is stored, similar to a stretched rubber band. Consequently, pure elastic deformation is not permanent, but is in fact completely reversible. Real materials can display a mixture of elastic and viscous character, in varying

degrees. Hence, the term viscoelastic is applied to such materials. In a viscoelastic fluid, elementary flow units interact together in some fashion such that the structural relaxation time (see below) is measurably long. This may be due to a flocculated or percolating particulate structure or to polymeric chain entanglement, for example. As stated, viscoelasticity is a material response that is a combination of viscous and elastic behavior. Viscous flow superimposed on elastic strain results in the relaxation, or gradual disappearance, of stress within the strained object. This is manifested as an imperfect or fading stress memory. Figure 12 illustrates the material response to an applied stress for a material with permanent stress memory, such as a rubber band, and one with fading stress memory, such as ''silly putty''. Viscoelasticity can have significant consequences for coatings processes,[36–38] as discussed in section 4.6.3.

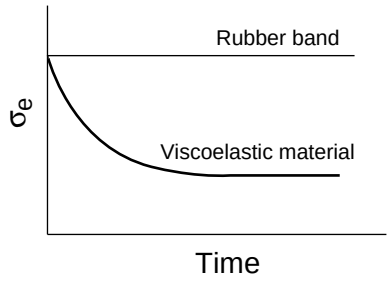

Figure 12 Stretch–force experiment illustrating viscoelastic stress-memory loss. The rubber band is a cross-linked polymer that does not relax tensile stress (σ_e) when stretched. The viscoelastic material adjusts to the strain by viscous flow, allowing elastic stress to decay.

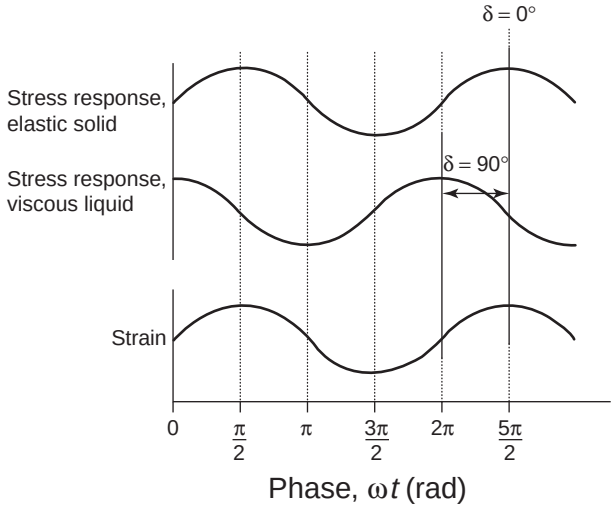

Figure 13 Stress response to oscillatory strain for ideal viscous and elastic bodies. The stress is in phase with the strain for an elastic solid; the stress leads strain by phase shift $\delta = 90°$ for a viscous liquid.

4.6.1 Viscoelastic Parameters and their Measurement

Viscoelastic materials are often characterized by sinusoidally alternating the direction of strain (or stress; Figure 13). For example, a sinusoidally varying strain γ may be applied at frequency w (rad s^{-1}). The strain amplitude varies with time t according to Equation (23),

$$\gamma = \gamma_0 \cos wt \tag{23}$$

which describes a cosine wave of maximum amplitude γ_0. The strain amplitude must be small such that the material response remains in the linear viscoelastic region (LVER) where stress and strain are linearly related. In purely elastic behavior (Hookean spring), the stress follows the strain instantaneously; the two are said to be in phase with each other. For materials with a viscous (energy-loss) component in the material response, the stress and strain will not be coincident, but will be out of phase. This happens because viscous flow continually relieves the stress within the material, causing the stress to reach a maximum and die away before the strain maximum. That is, as the rate of strain decreases near the maximum (or minimum) of the strain sine wave, stress relaxation catches up and overtakes the stress-building effect of the strain. The separation of the stress and strain maxima is called the phase shift (or phase angle – the angular part of a complete cycle represented by the phase shift). The phase shift is often given the symbol δ. The faster the viscous stress–relaxation process, the earlier in the cycle the stress maximum will occur, i.e. the larger the phase shift will be. The limiting value of the phase angle for purely viscous liquids is 90° (Figure 13). This is because the maximum strain rate (maximum slope of the strain sine wave) occurs 90° ahead of the maximum strain, and for Newtonian liquids the stress is proportional to strain rate. Therefore, the maximum stress must occur at the maximum strain rate, which corresponds to a phase shift of 90°.

The complex shear modulus G^* (Figure 14) is the ratio of the maximum stress σ_0 to the maximum strain γ_0, i.e. $G^* = \sigma_0/\gamma_0$. The complex modulus G^* may be

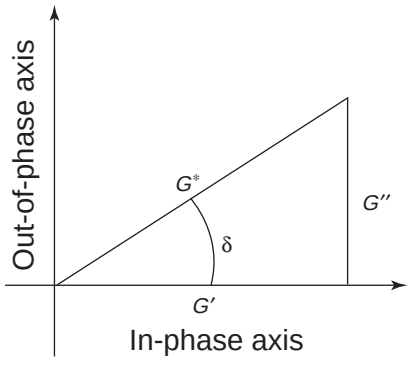

Figure 14 Viscoelastic relationships in a complex plane.

separated into G', the modulus of elasticity (storage modulus), and G'', the viscous modulus (loss modulus). Figure 14 demonstrates the geometric relationship of G^* to its in-phase and out-of-phase components, G' and G'', governed by the phase angle δ. It is obvious from geometry that when δ is zero, $G^* = G'$, and when $\delta = 90°$, $G^* = G''$. The physical meaning is that, when $\delta = 0$, the mechanical response is purely elastic, whereas when $\delta = 90°$, the response is purely viscous. Trigonometry yields Equations (24) and (25):

$$G' = G^* \cos\delta \tag{24}$$

$$G'' = G^* \sin\delta \tag{25}$$

Put another way, G' is the component of the complex shear modulus that is in phase with the strain. G' therefore represents the elastic part of viscoelastic behavior. G'' is the viscous component of the modulus, the part that is 90° out of phase with the strain, but in phase with the strain rate (i.e. in phase with the velocity).

We can define a dynamic viscosity η' as the quotient of the stress in-phase with the strain rate divided by the strain rate:

$$\eta' = \frac{\sigma_0 \sin\delta}{\dot{\gamma}_0} = \frac{\sigma_0 \sin\delta}{\gamma_0 w} = \frac{G^* \sin\delta}{w} = \frac{G''}{w} \tag{26}$$

Equation (26) is obtained using the relationships $\dot{\gamma}_0$(strain-rate amplitude) $= \gamma_0 w$ and $G^* = \sigma_0/\gamma_0$. (The angular frequency $w = 2\pi f$ (f in hertz).)

The magnitudes of G'' and G' reveal the relative importance of viscous and elastic behavior in the mechanical response of a material. In Figure 14, again by trigonometry, the tangent of the phase angle equals the ratio G''/G'. Thus, $\tan\delta$ quantifies the balance of energy loss and energy storage, measured under certain conditions of temperature, pressure, and frequency or rate of deformation. For solids, $\tan\delta$ can be useful in predicting the likelihood of brittle or ductile failure of a polymer or the sound absorption or vibration-damping properties. For liquids, $\tan\delta$ can monitor the progress of fluid restructuring in thixotropic recovery. The latter may be useful in computing sag resistance, for when $\tan\delta$ becomes less than unity the system has essentially become immobilized (i.e. it has reached a dynamic yield point). When $\tan\delta > 1$, viscous flow (i.e. a steady-state viscosity) is possible. However, when $\tan\delta < 1$ the material is more elastic than viscous and can be regarded as a gel.

4.6.2 Models for Linear Viscoelasticity

It is useful to employ simple mechanical models as aids to conceptualizing linear viscoelasticity. As stated, viscoelasticity is a combination of two idealized behaviors: Hookean elasticity and Newtonian viscosity. The mechanical analog of Hookean behavior is a spring of force constant G, and that of Newtonian flow is a dashpot – a piston-in-cylinder filled with a viscous fluid of viscosity η. These elements are combined in various ways to model viscoelastic response.

The first model consists of a spring and dashpot in series (Figure 15), an arrangement known as the Maxwell model. To imagine what response the model has, apply a constant stress σ_0 to one end, the other being fixed. The dashpot is considered to be infinitely long, so that the piston never runs out of travel. This being the case, an equilibrium strain would never be reached, but the dashpot would continually move as long as the stress is applied. As the model can flow without limit, this is obviously a model for viscoelastic liquids (also called elasticoviscous liquids). The foregoing describes a creep experiment, in which a sudden stress σ_0 is applied, and the evolution of strain, or deformation, is followed with time. The creep of a Maxwell liquid is not very interesting, however. As Equation (27) shows, the Maxwell strain–time curve consists merely of a straight line with intercept equal to σ_0/G and slope of σ_0/η:

$$\gamma = \frac{\sigma_0}{G} + \frac{\sigma_0 t}{\eta} \tag{27}$$

An alternative to the creep method, that is useful for viscoelastic solids, is to apply a sudden deformation and follow the decay of stress with time. This is similar to the experiment depicted in Figure 12 and is known as stress relaxation. A sudden strain γ_0 results in instantaneous lengthening of the spring. The dashpot experiences an initial stress σ_0 from the extended spring, causing a gradual movement of the dashpot's piston, relaxing both the spring and the stress. Figure 16 illustrates the decay of stress in a stress relaxation experiment for the Maxwell model, described by Equation (28):

$$\sigma = \sigma_0 e^{-Gt/\eta} \tag{28}$$

The quotient η/G has the dimension of time; this quotient is given the symbol λ and is referred to as

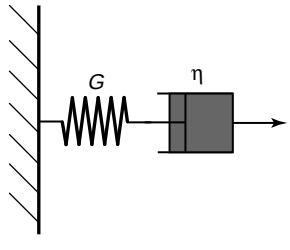

Figure 15 Maxwell model for viscoelastic liquid behavior. G is the force constant (modulus) of spring, η is the viscosity of dashpot fluid, and σ_0 is the applied stress.

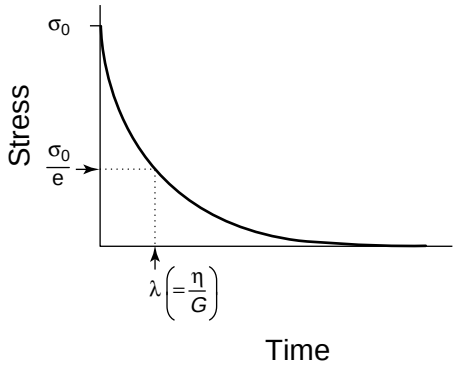

Figure 16 Stress relaxation experiment. A fixed strain is imposed, creating an initial stress σ_0, which relaxes with time constant $\lambda(=\eta/G)$. λ is the time at which the stress has decayed to σ_0/e.

the stress-relaxation time constant. It can be seen from Equation (28) that when $t = \eta/G$, then $\sigma = \sigma_0/e$.

Therefore, λ represents the time for the stress to fall to 1/e of its initial value (Figure 16). The time constant λ could also be called the stress memory time constant, as it is the time taken for a viscoelastic material to forget its initial elastic stress level when subjected to a constant strain.

In modern rheometers, sinusoidal oscillation is a common method of viscoelastic characterization. Equations (29–31), derived for the Maxwell model,[39,40] show how the viscoelastic time constant λ may be obtained from the oscillatory data:

$$\eta' w = G'' = \frac{G_\infty w\lambda}{1 + w^2\lambda^2} \tag{29}$$

and

$$G' = \frac{G_\infty w^2\lambda^2}{1 + w^2\lambda^2} \tag{30}$$

from which,

$$\tan\delta = \frac{G''}{G'} = \frac{1}{w\lambda} \tag{31}$$

where G_∞ is the limiting value of G' reached at high frequency, equal to the Maxwell spring constant. Equation (31) says that where G' and G'' cross over ($\tan\delta = 1$), then $\lambda = 1/w_c$ (w_c is the cross-over frequency). The high-frequency modulus can be obtained from Equation (30), because at the cross-over point $G_\infty = 2G'$. Similarly, at cross-over, the zero-frequency plateau viscosity $\eta = 2\eta'$ [from Equation 29].

If the mechanical elements are connected in parallel rather than series, they each experience identical strain, but the stresses are now additive. This arrangement is called the Kelvin–Voigt model (Figure 17). Because the elements are now in parallel, the Kelvin–Voigt model

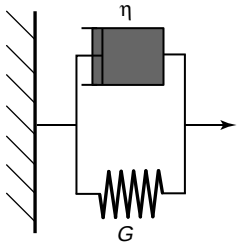

Figure 17 Kelvin–Voigt model for viscoelastic solid. (Notation as in Figure 15.)

can only undergo finite strain, limited by the extensibility of the spring. Therefore, this is a model for a viscoelastic solid below its yield point. Equations (32–34), for creep, stress relaxation, and $\tan\delta$, for the Kelvin–Voigt model are

$$\gamma = \frac{\sigma_0}{G}(1 - e^{-t/\lambda}) \tag{32}$$

$$\sigma = \gamma_0 G \tag{33}$$

$$\tan\delta = w\lambda \tag{34}$$

These equations describe an exponentially increasing strain at constant stress (see Figure 18 and Equation 32) and a nonrelaxing stress at constant strain [Equation 33], respectively.

The Maxwell and Voigt models each have a single relaxation time and cannot accurately describe viscoelastic materials that have multiple relaxation mechanisms and associated relaxation times. A model must possess multiple elements in order to describe them. A Maxwell element connected in series with a Kelvin–Voigt element is a first step in that direction, and does indeed model the linear viscoelasticity of real systems better than the single element models. Figure 19 shows such an arrangement, known as the Burgers model. Equations (35) and (36) describe creep and stress relaxation behavior for the

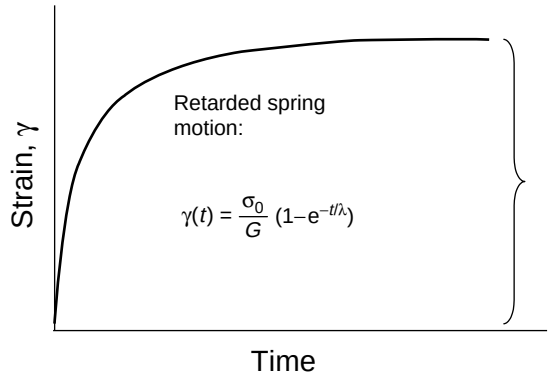

Figure 18 Kelvin–Voigt creep experiment. At constant applied stress σ_0, the strain γ increases exponentially with time constant $\lambda = \eta/G$.

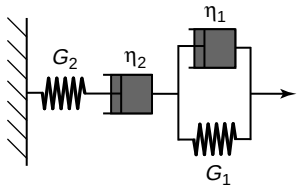

Figure 19 Burgers model: subscript 1 refers to the Voigt element; subscript 2 refers to the Maxwell element.

Burgers model:

$$\gamma = \frac{\sigma_0}{G_2} + \frac{\sigma_0 t}{\eta_2} + \frac{\sigma_0}{G_1}(1 - e^{-t/\lambda_1}) \quad (35)$$

$$\sigma = \gamma_0 G_2 e^{-t/\lambda_2} + \gamma_0 G_1 \quad (36)$$

where $\lambda_1 = \eta_1/G_1$ and $\lambda_2 = \eta_2/G_2$. The creep and stress relaxation behaviors for the Burgers model are shown in Figures 20 and 21. Modern rheometers of the controlled stress type are capable of performing creep measurements. Creep data are usually analyzed according to an extended Burgers model, to which has been added a series of additional Voigt elements.

4.6.3 Viscoelasticity and Industrial Processes

The role of elasticity in an industrial process depends on the rate (or frequency) of deformation. Put another way, the response of a viscoelastic material depends on how long a stress is applied ($1/w$), relative to the time required for any elastic extra stress to decay. This suggests taking a ratio of the stress relaxation time constant λ to the time (duration) of the process stress, t. This ratio is a defined rheological term known as the Deborah Number,

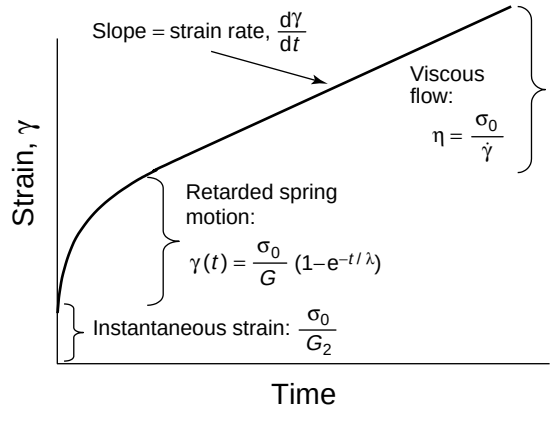

Figure 20 Burgers creep experiment. Under constant applied stress σ_0, the Burgers model undergoes instantaneous deformation of magnitude σ_0/G. The next segment of the curve is retarded spring motion where strain increases exponentially with time. Finally, the spring is fully extended and viscous flow occurs with a constant strain rate $d\gamma/dt$.

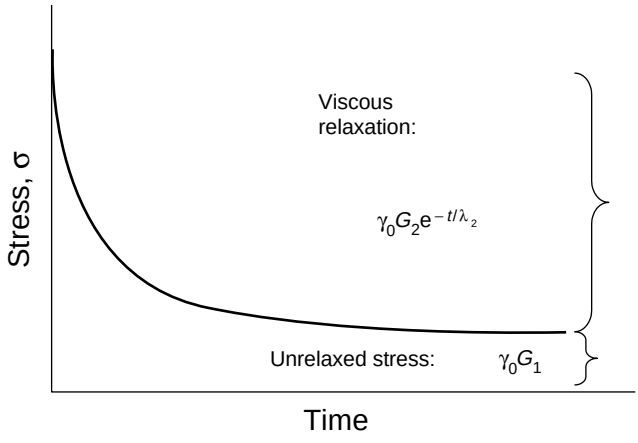

Figure 21 Burgers model stress relaxation. Imposition of constant strain γ_0 causes stress, which decays exponentially, governed by the relaxation time λ_2 ($=\eta_2/G_2$). The Voigt element contributes unrelaxed stress $\sigma_\infty = \lambda_0 G_1$.

De [Equation 37]:

$$De = \frac{\lambda}{t} \quad (37)$$

It is named for the Biblical prophetess Deborah, who prophesied that the "mountains flow before the Lord".[41] This is an accurate statement, long before being verified by the science of geology, of the fact that, on God's timescale, rock formations can be observed to undergo permanent deformation, or flow. In other words, if the duration of the applied stress t greatly exceeds the relaxation time λ ($De \ll 1$) the material will respond as a viscous fluid (because elastic stress has time to decay). Conversely, if $De \gg 1$ a material behaves as if an elastic solid. When $De \approx 1$ the material will behave viscoelastically (i.e. stress relaxation will take place on the timescale of the process). Thus, the Deborah number quantifies the proportion of elastic to viscous control of a process. This is one reason why a determination of the viscoelastic properties of paints and coatings is important. An equivalent way of writing the Deborah number is Equation (38),

$$De = \lambda\dot{\gamma} \quad (38)$$

where $\dot{\gamma}$ is the shear rate of a coating process. As an example of how the role of elastic extra stress can be determined for a given process, assume a stress relaxation time constant of about 10 s for a paint. Thus, for a pigment particle settling at $\dot{\gamma} \approx 10^{-5}\,s^{-1}$, $De \ll 1$ and the paint would offer no elastic resistance to the settling pigment. However, for leveling ($\dot{\gamma} \approx 0.1\,s^{-1}$), $De \approx 1$ and elastic stress is likely to retard the leveling process.

Viscoelastic dispersions[42] and associative polymer solutions[43] can show Maxwellian behavior with a single relaxation time. In general, however, real materials may not exhibit simple exponential stress decay (i.e. a single relaxation time), but rather may possess a spectrum of relaxation times. However, the mechanical response can be dominated by the longest relaxation time,[36] obtained from experiments such as described in foregoing sections.

The process consequences of viscoelasticity are also partly due to the stabilization of otherwise unstable liquid structures by the elastic extra stress. Thus, liquid fibers and webs which would ordinarily collapse due to surface tension are stabilized, leading to, for example, rollcoat spatter[44] (or misting) and ribbing, and also inhibiting atomization of sprayed materials. Once again, the Deborah number gages the importance of viscoelasticity for a particular process. Of course, both the magnitude and lifetime of the elastic stress will be important, for together they will govern the degree of stabilization.

Many processes, particularly of coatings application, involve large strains outside the range of linear viscoelastic behavior. There is no general theoretical basis for treatment of nonlinear viscoelasticity at present. However, there are promising experimental methods, based on an enhanced creep analysis, for characterizing nonlinear viscoelastic behavior, even at high frequencies.[45] It may be possible to build experimental correlations between nonlinear viscoelasticity and coatings performance.

5 RHEOLOGY OF DISPERSIONS

Rheology is particularly useful for dispersion characterization because of its sensitivity to the microstructure. Both the stability and quality of dispersions can be evaluated. Dispersion rheology has two broad aspects: (a) the dependence of viscosity on the concentration of the dispersed phase, and (b) the dependence of viscosity on shear stress and shear rate. (A third important aspect is the viscoelastic properties of dispersions, which will not be discussed here.) Additional factors are particle shape, rigidity, particle size and size distribution, and interparticle forces, both attractive and repulsive. All these elements combine to determine the microstructure of the dispersion, its response to stresses, and hence its rheology.

A dispersion consists of a discontinuous phase dispersed within a continuous phase (often mobile). For example, this might be a system in which a fine-particle-size solid is wetted by, and thoroughly mixed with, a liquid. The dispersed phase may be a liquid (emulsions), solid (suspensions), or gaseous (foam). The terminology is somewhat inconsistent in the literature, but the dividing line between colloidal dispersions and less-stable dispersions that it seems logical to call "suspensions" is, in practical terms, one of particle size. Colloidal particles range in size from about 10 nm to 1 μm. Due to their small size and mass, colloidal particles often are very slow settling, or are nonsettling, because Brownian motion effectively keeps them suspended. However, even colloidal-size dense particles will eventually settle. Suspensions above 1 μm in size generally exhibit rapid settling because Brownian forces are ineffective with such massive particles. Van der Waals attractive forces are proportional to particle size, and so larger particles tend to flocculate, leading to more rapid settling.

Addition of a particulate phase to a liquid increases its viscosity in proportion to the volume fraction ϕ of the dispersed phase. ϕ is the fraction of the total volume of the dispersion occupied by the dispersed material (a dimensionless number; Equation 39):

$$\phi = \frac{\text{volume of particles}}{\text{volume of dispersion}} \tag{39}$$

A convenient way of expressing the effect of a dispersed phase on the viscosity of a liquid is by normalizing the dispersion viscosity to the pure-liquid viscosity. This ratio is termed the relative viscosity η_r and is also dimensionless [Equation 40]:

$$\eta_r = \frac{\eta_d}{\eta_l} \tag{40}$$

The contribution of a single particle to the viscosity of a dispersion is characterized by its intrinsic viscosity (or limiting viscosity number) $[\eta]$, a function of particle shape and deformability. $[\eta]$ is obtained as the slope of a plot of $\eta_r - 1$ (formerly known as the specific viscosity) versus ϕ [Equation 41]:

$$[\eta] = \lim_{\phi \to 0} \frac{\eta_r - 1}{\phi} \tag{41}$$

It is reasonable that the viscosity of a liquid will be augmented by a factor equal to the product of the particle intrinsic viscosity and the concentration of particles [Equation 42]:

$$\eta_r = 1 + [\eta]\phi \tag{42}$$

Einstein[46] calculated the intrinsic viscosity for noninteracting rigid spheres in a Newtonian liquid, obtaining the number 2.5 [Equation 43]:

$$\eta_r = 1 + 2.5\phi \tag{43}$$

The intrinsic viscosities of other particle shapes, such as prolate and oblate spheroids (discs and rods), have been calculated[47] and are always greater than 2.5. This

means that any deviation from spherical particle shape will increase dispersion viscosity. Equation (43) is valid only in the very dilute regime ($\phi < 0.05$), where the flow around a single particle is not affected by the presence of other particles. At higher concentrations the motion of particles is perturbed by interactions with other particles. These interactions are accounted for by higher-order terms in ϕ, as for example by Batchelor [Equation 44]:[48]

$$\eta_r = 1 + 2.5\phi + 6.2\phi^2 \tag{44}$$

Figure 22 illustrates the typical dependence of η_r on ϕ for actual dispersions (using the data of Eilers[49]), and shows curves corresponding to the predictions of Equations (43) and (44). It can be seen that Equation (44) predicts a finite viscosity at $\phi = 1$, which is not realistic. It is also seen in Figure 22 that η_r goes toward infinity at a volume fraction considerably less than unity. The volume fraction ϕ_m corresponding to $\eta_r \to \infty$ is called the maximum volume fraction, or maximum packing fraction. At $\phi = \phi_m$, the density of particle packing is such that the dispersion can no longer flow. The value of ϕ_m will be system dependent (also shear-rate dependent) and will be determined by particle shape, particle size distribution, the ionic strength of the medium, the degree of particle flocculation, and the manner in which the particles arrange themselves (pack) in three-dimensional space.

Numerous models have been proposed containing ϕ_m.[16,49–51] Probably the most successful in fitting a variety of data is the Krieger–Dougherty model [Equation 45]:[52]

$$\eta_r = \left(1 - \frac{\phi}{\phi_m}\right)^{-[\eta]\phi_m} \tag{45}$$

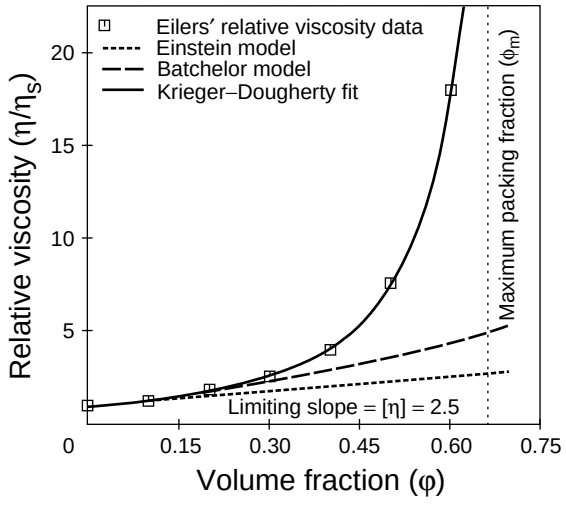

Figure 22 Dispersion relative viscosity data of Eilers together with best-fit curves for the Einstein model, the higher-order model of Batchelor, and the Krieger–Dougherty model. Also indicated are the hard-sphere limiting behaviors: a straight line with slope equal to the Einstein intrinsic viscosity $[\eta] = 2.5$, and the maximum packing fraction $\phi_m = 0.63$.

Figure 22 shows the fit of the Krieger–Dougherty model to data of Eilers, using a value of $[\eta]$ close to the Einstein value. The maximum packing fraction will have different, unique values at low and high shear rates because the strength of the shear field determines how efficiently particles pack together. Reviews of the rheology of polymer colloids are recommended for further reading.[14,47,53]

Having said that dispersion viscosity depends on the disperse-phase volume fraction ϕ, it is often the case that the effect of the particle phase on the rheology may be greater than expected on the basis of the volume of material used to make up the dispersion. This is because it is the *hydrodynamically effective* particle volume fraction that determines the rheology, rather than the formulated volume fraction. Thus, the rheological behavior will scale with the effective volume fraction, ϕ_{eff}, rather than with the formulated volume fraction, ϕ.

There are two mechanisms that can cause ϕ_{eff} to be greater than ϕ. In the first, the effective particle radius increases as a consequence of various methods of dispersion stabilization. For example, adsorption of a polymeric stabilizer onto the surface of a particle adds the thickness of the stabilizer layer to the particle radius. By steric–osmotic interactions between stabilizer layers, the close approach and flocculation of particles is prevented, resulting in steric stabilization. If the thickness of the steric stabilizing layer on the particle is δ, the effective volume fraction is given by Equation (46),[53]

$$\phi_{eff} = \phi \left(1 + \left(\frac{\delta}{R}\right)^3\right) \tag{46}$$

where R is the original particle radius. Quantitative viscometric methods have been developed for inferring the adsorbed-layer thickness.[53,54]

The effective particle radius and hence the effective volume fraction ϕ_{eff} can also be increased by the effects of particle electrical charge. Aqueous dispersions are often stabilized by association of repulsive electrical charge with the particle, which is called electrostatic stabilization. The charge may be due to adsorbed ions, anionic or cationic surfactants or polyelectrolytes, or, as in the case of polymer colloids, to the presence of ionizable groups which are part of the polymer molecules. (In the case of functionalized polymer latexes, such groups tend to migrate to the particle surface.) The surface ionic charge surrounds the particle with an electrical field, which propagates into the aqueous phase at a

distance inversely proportional to the ionic strength of the medium. In the presence of dissolved counterions, an ionic atmosphere envelops the charged particle. These phenomena are the origin of the so-called electroviscous effects, which have to do with the way the electrical field surrounding the particle affects the collision cross-section and the effective volume fraction, and also the nature of the hydrodynamic interaction between particle and surrounding liquid.[47,55] The rheology of a dispersion can be greatly altered as a consequence of these effects, as the solution ionic strength or the pH are varied. In a study by Krieger and Eguiluz, η_r for a dispersion of uniform polystyrene latex spheres ($\phi = 0.4$) decreased from 10^6 to 10, as the result of increasing the concentration of electrolyte by two orders of magnitude.[56] In that study, the low-electrolyte latexes exhibited apparent yield stress behavior, a consequence of the increase of the effective volume fraction of the latex particles due to the expansion of the repulsive electrostatic field. The effect is so great that the particles become locked into crystalline arrays that can diffract light, producing striking iridescent colors. As the electrolyte concentration is increased, the counterion cloud both shrinks the electrostatic field and shields particle fields from each other. Consequently, the viscosity drops dramatically (remember that the volume change depends on the cube of the radius change).

A second mechanism by which ϕ_{eff} can exceed ϕ is flocculation. When particles form flocs, the liquid phase contained within the interstices of the floc is not free-draining, but effectively becomes part of the floc. Thus the volume of a flocculated phase is augmented by the volume of trapped interstitial liquid. Consequently, the disperse-phase volume fraction is higher than for the deflocculated system.

Dispersions can exhibit the full range of rheological behavior mentioned previously, including Newtonian, shear-dependent, time-dependent, plastic, and elastic behavior. A dispersion of noninteracting spherical particles will be Newtonian up to about $\phi = 0.2$.[16] Above this point, onset of non-Newtonian character begins due to particle interactions and hydrodynamic factors. Mechanisms for these effects are discussed in section 4.2.

6 RHEOLOGY OF FILM FORMATION

In the following sections two coating flows, sagging and leveling, are described. These flow processes occur in an applied coating layer during what is known as the film-formation phase of drying. In describing the way in which the coating rheology controls the flow, there is an important point to be made. Both sagging and leveling are driven by specific shear stresses, which can be calculated from the forces acting (gravity and surface tension, respectively) and the geometry of the film. However, these processes can occur over a wide range of shear rates, depending on the coating's viscosity at the acting shear stress. For this reason it is preferable to graph flow behavior as viscosity versus shear stress plots, as opposed to viscosity versus shear rate, which is the more common practice. The shear rate is the dependent variable, for real processes. The appropriate independent variable to use to differentiate the performance of paints according to their rheology is the shear stress. Not to do so is wrong in principle and will result in incorrect relative ranking of paints by their rheology with respect to their rates of sagging, leveling, pigment settling, or to their ease of application. The discussion below neglects much in the way of detail, which can be found in an earlier review.[57] The focus is on the former issue, as it is key to using rheological data correctly in the interpretation of flow curves.

Plotting with shear stress as the independent variable (Figure 23) allows straightforward and correct comparison of paints at the specific shear stress for a particular process. The process stresses illustrated here are for (a) gravity-driven sagging of a 3-mil wet film, (b) surface tension-driven leveling of surface roughness, and (c) application by brush or roller. The point is that, no matter what the rheology of a paint, the shear stress acting on a coating layer for a given process is the same (for a given geometry, density, surface tension). The shear rates for these processes, however, are not the same, as shown in Figure 24. Figure 24 shows the same two paints as in Figure 23, this time with the viscosity plotted as a function of shear rate. Note that process shear rates are shifted to the left for the higher viscosity paint.

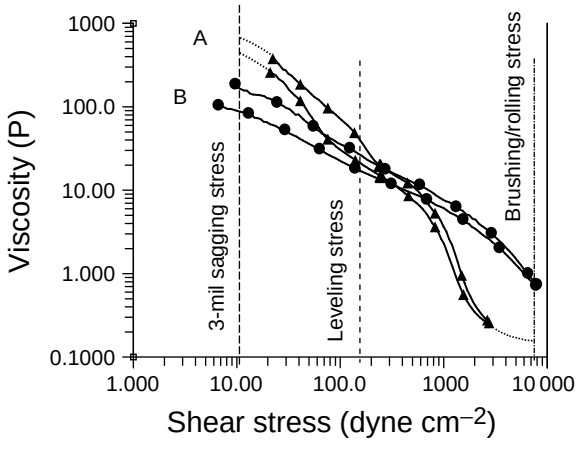

Figure 23 Viscosity versus shear stress plot for two paints A ( ) and B (●). Vertical dashed lines indicate gravitational shear stress driving sagging for a 3-mil wet paint layer; surface-tension-driven leveling stress; brushing/rolling application stress.

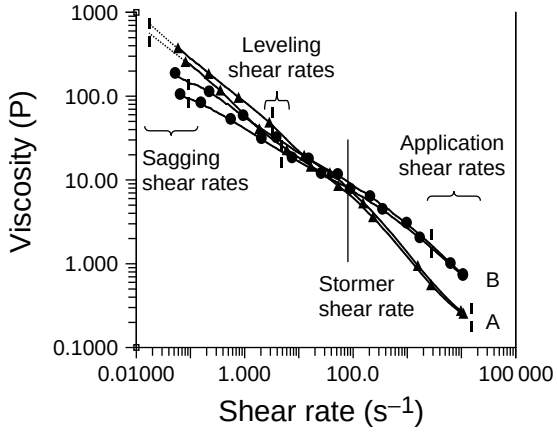

Figure 24 Viscosity versus shear rate plot for paints A ( ) and B (●). Shear rates for sagging, leveling, application are different for A and B, shifted to the left for higher-viscosity paint. Vertical dashed lines indicate shear rates for each paint and process. Vertical solid line is the approximate shear rate for the Stormer™ Viscometer.

6.1 Sagging

On a nonhorizontal substrate, a liquid coating layer will flow or drain downward under the influence of gravity. For a coating on a vertical wall, this is usually called sagging. Sagging is driven by gravitational shear stress σ_g whose magnitude depends solely on the wet film thickness and density [Equation 47]:

$$\sigma_g = \rho g h \cos\theta \tag{47}$$

(where $\cos\theta = 1$ for a vertical substrate). Predictions of sagging from viscosity measured at an arbitrary shear rate will be misleading because paints sag at different shear rates, depending on the rheology. The proper way to predict relative sagging tendency is to first select the governing viscosity from a plot of viscosity versus shear stress, rather than shear rate. The sagging shear stress calculated from Equation (47) determines the viscosity controlling sagging (η_{sag}) from the flow curve. Figure 23 compares two paints in this manner, with gravitational shear stress levels corresponding to the 3 mil wet-film thickness indicated. Figure 24 shows the flow curves for the same two paints plotted as a function of shear rate, with the calculated sagging shear rates at 3 mils indicated ($\dot{\gamma}_{sag} = \sigma_{sag}/\eta_{sag}$).

The solid vertical line in Figure 24 is approximately the shear rate of the Stormer™ viscometer. The viscosities of the two paints at the Stormer™ shear rate are about the same (paint B actually slightly higher). At an arbitrary shear rate of $1\,s^{-1}$, paint A is about 50% higher, whereas at the true sagging shear rates paint A is 600% higher in viscosity. Clearly, comparison of viscosities measured at arbitrary shear rates can lead to incorrect predictions of relative sagging behavior, particularly if the flow curves cross over. In contrast, when plotted as a function of shear stress (Figure 23), the sagging shear stresses are single vertical lines, unambiguously cutting the viscosity curves at the value appropriate to the particular situation.

6.2 Leveling

A freshly applied coating layer will be of variable thickness, having a pattern of surface roughness generated by hydrodynamic factors in the application process. It is desirable to have uniform coating thickness for both esthetic and functional reasons. The process of smoothing of the initial rough surface is driven by surface tension and is called leveling. Numerous studies of the subject have been published,[36,58–64] but the role of rheology and other factors in the leveling of paint films is not fully understood. One might expect leveling to correlate reasonably well to low-shear viscosity. However, factors such as viscoelasticity, time-dependent viscosity, evaporation, wicking, and surface tension gradient flows make leveling a complex process. These complications are not dealt with here, but a fuller discussion of leveling and sagging is available in an earlier review.[57]

The basic model for the leveling of a thin liquid layer is due to Smith et al.[65] Orchard later[66] developed a complicated expression for the leveling of a more realistic, arbitrarily rough surface profile in terms of a Fourier series. However, the expression simplifies considerably when the coating thickness is assumed to be small relative to the roughness wavelength. Assuming a single sinusoidal wavelength (which is a reasonable approximation as the leveling behavior will be dominated by the longest wavelength), the Orchard equation is

$$\ln\frac{a_t}{a_0} = -\frac{16\pi^4 h^3 \gamma t}{3\lambda^4 \eta} \tag{48}$$

for constant surface tension γ and viscosity η. Equation (48) says that the roughness amplitude a_t decays exponentially with time as a function of the coating layer thickness h to the third power and of the inverse roughness wavelength λ to the fourth power.

The Orchard equation is a deficient model for the leveling of real paints because it assumes Newtonian fluid behavior and neglects time-dependent effects, such as thixotropy, viscosity increase due to drying and wicking, and surface tension gradient (Marangoni) effects. Overdiep[62] and others[67,68] have discussed the importance of Marangoni flows for coatings. Overdiep[62] showed that a strong Marangoni rebound occurs during the leveling of solvent-based paints which develop surface tension gradients during drying. Nevertheless,

the Orchard model conveys the effect of film geometry and physical properties on leveling rates and allows estimation of the effects of changes in these variables. For example, it should generally hold, despite complicating factors, that doubling the wet film thickness will increase the leveling rate eight-fold. However, leveling is a very nonlinear process and more realistic modeling will require the use of computational fluid dynamics methods, such as numerical simulation.[64,69] Such methods require realistic constitutive models obtained from rheological characterization as necessary input.

6.2.1 Rheological Measures of Leveling

Kornum and Raaschou–Nielsen[70] and Dodge[25] advocate rheological test conditions which simulate real paint processes, i.e. the best rheological test for leveling should most closely simulate the application and postapplication experience of the paint. This would mean that: (a) the paint should experience shear rates/stresses/strains similar in magnitude and duration to those of the application process; (b) the shear rate/stress/strain should then be step-changed to values representative of the leveling process; (c) the change of viscosity or viscoelasticity with time should then be followed under the latter conditions. This protocol is essentially identical to the step-shear method of characterizing thixotropic recovery described in section 4.4.3. This kind of procedure is most easily accomplished by employing a rheometer capable of job-streaming or linked methods. However, the foregoing method still suffers from the limitation that the effect of drying on viscosity during the leveling period is neglected, and that the leveling stress decays with time as leveling proceeds.

Equation (49) calculates the stress driving leveling from the roughness amplitude a, roughness wavelength λ, wet film thickness h, and surface tension γ:

$$\sigma_{\text{lev}} = \frac{8\pi^3 \gamma a h}{\lambda^3} \tag{49}$$

This leveling stress decays exponentially as leveling progresses.[71] The variation of viscosity with decaying shear stress can be characterized using a CS rheometer with programmable logarithmic stress ramp. Leveling efficiency might then be judged by numerically integrating the fluidity over the leveling period [Equation 50]:

$$\Phi = \int_0^t \frac{1}{\eta}\, \mathrm{d}t \tag{50}$$

Leveling and gloss have been found to be logarithmic functions of the fluidity integral for powder coatings.[72]

7 RHEOLOGY INSTRUMENTATION

7.1 Industrial Viscometers

Many different instruments have been devised to measure viscosity or consistency in the industrial laboratory. However, a number have serious design limitations that make them generally poorly suited for the measurement of non-Newtonian fluids. This becomes an important issue when one realizes that paints and other commercial products such as pastes, slurries, dispersions, emulsions, gels, etc. are non-Newtonian. The problem with many of the familiar laboratory viscometers is that the flow field is complex and not subject to mathematical analysis. The result is that an accurate stress or strain rate cannot be calculated. Therefore, the true viscosity cannot be known. Furthermore, the nonuniformity of the shear field causes the viscosity of the material to vary within the test specimen for non-Newtonian materials. For complex flow fields, such an error cannot be corrected. The measurement error will be in proportion to the sample's deviation from Newtonian character. A second important deficiency of most simple viscometers is a lack of temperature control, which can be a source of significant error. These quality-control viscometers have been discussed elsewhere[39,57,73] and are not further considered here.

7.2 Absolute Viscometers

An absolute viscometer produces viscometric flow (steady simple shear) and yields data in absolute units for non-Newtonian fluids. Broadly, these are of two types: rotational viscometers and efflux viscometers. Rotational viscometers employ steady (uniform) shear, oscillatory shear, and transient (time-dependent) shear experiments of various types. Efflux viscometers extrude the fluid through a small-gap orifice, which can be a capillary tube, a slit, or an annulus (concentric cylindrically walled tube). Absolute devices based on acoustic wave propagation exist, but are less common. The focus in this abbreviated treatment will be on rotational instruments. Several other instruments capable of absolute measurement are being excluded from discussion because of a limited range. These include the Brookfield viscometers, the falling needle viscometer, the ICI cone and plate viscometer, and others. A number of summaries of other instrumentation and methods are available.[57,73–75] Although dated, Van Wazer et al.'s book is still a useful resource.[39]

7.2.1 Rotational Instruments

7.2.1.1 Controlled Stress and Controlled Rate The choice of working principles in modern rotational rheometers is between controlled strain (or controlled

strain rate) and controlled stress. The difference is whether the torque or the angular displacement is the controlled variable. In a controlled-strain instrument the angular displacement (or perhaps angular velocity) is the independent (controlled) variable and the opposing viscous torque the dependent (measured) variable. Controlled-stress instruments control the torque, and measure the resulting rate of angular displacement. Instruments of the controlled-strain type include the Weissenberg Rheogoniometer, Haake CV, Rheometrics ARES, and Bohlin CVO. Instruments of the CS type include the TA Instruments AR-1000, Rheometrics SR5, Bohlin CSR, Haake RS, and Physica MC rheometers.

Each instrument type has characteristic advantages and limitations. The choice between the two types of instrument depends on the material under test and the intended experiments. In terms of performance, CS instruments can measure much lower angular velocities than can CR instruments, but CS tend to be more limited at the high angular velocity and oscillation frequency end. Both CS and CR instruments have inertia issues affecting the accuracy for high frequency or rapid accelerations. In the case of CS the issue is measurement system inertia and is correctable; for CR the problem is torque system resonance and is not correctable. The CS type is better suited to measure long relaxation times.[75]

CR instruments impose a shear-rate sweep while measuring the drag-torque response of materials. Structured fluids tend to be shear-sensitive (more precisely, strain-sensitive). Consequently, as strain increases exponentially under a linear strain-rate sweep protocol, the fluid structure tends to collapse rapidly, with the result that relatively few data points are obtained to provide information on structure. In contrast, CS devices use linear (and logarithmic) rates of stress increase, an intrinsically gentler mode of deforming materials, which allows materials to obey their own rules of stress–strain response. CS instruments are especially useful for characterizing structured fluids, such as paint, printing ink, adhesives, ceramic slips, coal slurries, cement, pigment and colorant dispersions, drilling muds, medicines, foodstuffs, cleaners, personal care products and cosmetics, solder pastes, etc. This is particularly true where materials exhibit yield behavior. In principle, CS instruments can directly measure the stress at the onset of yield, avoiding errors associated with extrapolation or curve-fitting methods (section 4.5.2.2).

Considering coating flows as a case in point, it is important to realize that the proper variable for correlating coating rheology to real-world coating processes is the shear stress, not the shear rate. First of all, coating flows are the outcome of the sum of forces acting on a fluid coating layer. That is, the rheological response to those forces determines the resulting coating flow. Therefore, coating flows are not driven at a given shear rate, but rather the observed shear rate is the result of the stress driving the process and the corresponding viscosity at that stress. Using stress as the applied variable is, in this sense, the more natural way to characterize coatings.

Both CS and CR instruments each do their own version of a transient experiment: creep for CS and stress relaxation for CR. The stable torque and high strain resolution capabilities of CS mean creep experiments can be used to probe very slow relaxation processes. A modified creep method – a programmed series of small torque increments – can be used for determining the first Newtonian plateau down to very low shear rates (section 5.4.2). A limiting factor in the accuracy of such methods, however, is the ability to minimize or correct for spurious air-bearing torque.

7.2.1.2 Inertia Correction One of the issues in CS rheometry is errors due to inertia. If the test fluid is low in viscosity, and/or the test rotor has a large moment of inertia, and/or the acceleration is too high (short ramp time), the applied torque is partially expended to accelerate the mass of the measuring system. This is the cause of ''bell-mouth'' loops in flow curve ramp experiments. Krieger has published a correction method for this error.[76] Commercial CS instruments now feature on-line inertia correction.

8 RHEOMETRY: EXPERIMENTAL METHODS

Rheometry is the experimental aspect of rheology. A rheometer is an instrument that can suitably measure some rheological property, or material function (section 3.2), such as the viscosity as a function of shear rate or shear stress. Normally, a curve will be produced representing the functional dependence of the measured property. In contrast, a viscometer will generally measure viscosity at a single (sometimes even unknown) value of the shear rate (or shear stress), or perhaps over a very limited range of values. These types of devices are typically found in the industrial quality control or formulation laboratory, and have been discussed previously.[57] This section focuses on techniques for the use of rotational rheometers to characterize fluids with complex rheological properties.

8.1 Rotational Rheometry

Simple shear deformation was defined in section 2.2, in which a fluid is confined between surfaces that are in parallel, rectilinear motion (i.e. a sliding plate rheometer). However, very few rheometric instruments utilize this test

geometry. This is because it is not a practical geometry, as either the experiment must be rather short in duration, or one of the shear faces must be very long. This predicament is solved by using rotational motion to shear a material between, say, a fixed plate and a rotating disk. The majority of commercial rheometers operate thus, using a variety of rotating fixtures or geometries, mainly (a) cone and plate, (b) parallel plate, and (c) concentric cylinder (bob and cup). Detailed descriptions of these standard geometries are available.[39,77–79]

Another important test geometry used in steady-shear flow is tube flow or capillary rheometry. Its use is somewhat limited outside the characterization of polymer melts, and consequently is not discussed further here.

One novel geometry that offers some advantages in measuring difficult materials, particularly those exhibiting wall slip is the vaned rotor.[33,34] It consists of rectangular vanes fixed to a rotating shaft (section 4.5.2). Barnes and Carnali[35] have shown that, for sufficiently shear-thinning fluids, this geometry is equivalent to a conventional bob and cup, except that wall slip is avoided because the effective shear surface is within the fluid itself, not at a solid surface.

Three principal rheometric techniques are typically used for rotational rheometers:

1. steady shear (large strain) – for equilibrium viscosity measurement, nonequilibrium viscosity measurement, and normal stress measurement;
2. creep (medium strain);
3. oscillation (small strain).

8.1.1 Steady Simple Shear (Equilibrium Viscosity)

Steady simple shear involves a uniform flow field in which the shear rates and stresses are independent of position. Not all rheometer test geometries meet this criterion exactly, but they do generally satisfy the requirements of viscometric flow, defined by Dealy and Wissbrun[80] as a flow field that is ''everywhere indistinguishable from steady simple shear'' (but not necessarily uniform shear). Viscometric flow is a necessary requirement in order to properly measure material functions.

Some rotational rheometers utilize steady-shear equilibrium viscosity measurements to characterize flow behavior, by means of a series of shear rate steps. If, for example, a constant torque is applied to a specimen until a constant shear rate is observed (after inertial and structural relaxation effects have equilibrated with time), the system is undergoing steady simple shear and the steady-state viscosity or equilibrium viscosity can be measured. (Even with thixotropic materials, it should be possible to observe a steady-state viscosity value eventually.) An equilibrium flow curve (as in Figure 3) can be generated in this way.

8.1.2 Time-dependent Materials: Nonequilbrium Flow Curve

As mentioned, one can construct a curve of equilibrium viscosity versus shear rate or shear stress by a series of step-equilibrium experiments (a variant of the creep test, see below). However, the nonequilibrium flow curve is less time-consuming experimentally and perhaps more informative with regard to the processability of time-dependent materials. The nonequilibrium flow curve is obtained by recording the viscosity while the angular velocity or torque is ramped, or increased with time (usually followed by a decreasing torque or velocity ramp; Figure 8a). Time-dependent effects such as hysteresis can be seen in this way (Figures 8b and 25). The collapse of the structure and also relaxation effects can be seen, which are lost in an equilibrium measurement. Figure 25 shows a flow curve for a structured fluid, measured by a stress-ramp procedure with the curves corresponding to the increasing-stress and decreasing-stress ramps indicated.

8.1.2.1 Shear-history Effects: Sample Mounting and Conditioning A great many industrial and commercial materials possess time-dependent rheology, usually thixotropy, but occasionally antithixotropy (section 4.4). Time-dependent materials pose a special set of challenges to the experimentalist, and it is easy to make mistakes in characterizing them. For time-independent systems the pre-experiment shear history is unimportant. In contrast, the shear history of time-dependent materials prior to the experiment will influence the results.

Both the degree of deformation in mounting a sample and the time elapsed since mounting will affect the rheological state of a time-dependent material and hence the experimental results. Failure to take account of this issue often leads to poor repeatability of data on such systems. Some approaches to control this shear-history problem are:

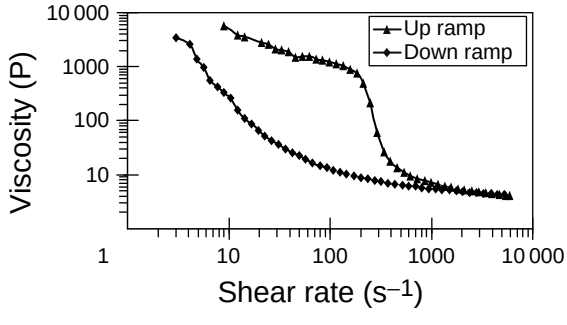

Figure 25 Viscosity versus shear rate curve for a thixotropic paint done by a stress-ramp experiment. Up ( ) and down ( ) ramp data are indicated.

1. Controlling the rate of closure of the platens in a parallel-plate or cone-plate experiment (e.g. exponentially decreasing the closure rate).
2. Allowing a rest time for the sample after mounting (limited by evaporation and productivity issues).
3. Giving a controlled shear history to the sample, such as preshearing the specimen for a set period, then initiating the experiment immediately on cessation of shear. (An up–down ramp cannot be applied in this case, because the sample would begin to recover in an uncontrolled way concurrent with the initial low-shear ramp onset. Instead, a series of step reductions from high shear can be applied, with the time delay between steps being varied, in order to simulate the effects of varying degrees of structural recovery.)[81]

Method (1) reduces the strain rate (but not the strain) the sample experiences as the instrument gap closes, but will be limited in its effectiveness because practical rates of closure (plus the coupled deformation) are likely to be significant in terms of their effect on the specimen's structural integrity. Furthermore, depending on the geometry (e.g. cone and plate) the strain may not be uniform. Method (2) is limited by the time one can afford to wait for structure recovery. It is not uncommon for a material to require hours in order to regain its rest-state condition. The use of an arbitrary rest time is a compromise that provides a practical but not ideal method of characterizing thixotropic materials. Method (3) is somewhat arbitrary, too, in that it is not known a priori at what shear stress or shear rate or for how long a material must be preconditioned in order to achieve adequate or consistent structural breakdown. Whatever the method chosen, it is important to adhere strictly to a constant protocol if results are wanted on a comparable basis.

8.1.2.2 Ramp Methods Many rotational rheometers have the ability to apply a time-based ramp of torque or angular velocity for purposes of flow-curve measurement. Some allow a choice of either a linear time-based ramp or one that is logarithmic with time. The latter method spaces the data points very close together in the beginning of the ramp, which increases the amount of information obtained at low shear rates and stresses, but also imparts a different shear history. For time-dependent systems, it is as important to keep the experimental conditions constant as to keep the pre-experiment shear history constant, in order to obtain repeatable results and to assure valid comparison of materials. This means keeping the following experimental settings constant for a given series of experiments:

1. ramp duration;
2. choice of linear or log time-based ramp;
3. maximum shear stress (or shear rate);
4. duration of any hold-time at ramp maximum.

8.1.3 Transient Methods

Transient methods are nonsteady state, i.e. the stresses, strains, and material properties are not at equilibrium, and it is their behavior as a function of time that is being studied, as they move toward equilibrium. Transient methods include:

1. step shear stress (creep and recovery);
2. step shear strain (stress relaxation);
3. step shear rate (thixotropic recovery);
4. start-up of flow

The first two methods are briefly discussed in section 4.6.2. The third method is described in section 4.4.3. The last method will not be discussed here. We will now discuss creep methods in more detail.

8.1.3.1 Creep Methods The viscoelastic character of a material can derive from a range of molecular and/or particle interactions, each with their individual force constants, length scales, and dissipative characteristics. Viscoelastic deformation can therefore occur over a range of strains and stresses. Another consequence is that viscoelastic materials will display a range of relaxation times, sometimes referred to as a relaxation spectrum. An oscillatory shear or dynamic method for determining linear viscoelastic behavior (sections 4.6.1 and 8.2) has a practical lower limit of frequency w, due to the time required to complete an oscillation at very low frequencies. Since the Deborah number $De = \lambda/t = w\lambda \approx 1$ for viscoelastic fluids (section 4.6.3), the measurement of very long relaxation times λ (characteristic of polymers) would necessitate the use of very low frequencies – probably below the practical limit (we do not want to wait hours to complete enough oscillations to analyze). An answer to this is the creep method, in which a constant stress σ_0 is applied to a material, which then undergoes strain. Creep conveniently probes long relaxation timescales, but also short- and intermediate-time behavior, as well.

As described in section 4.6.2, in a creep-recoil shear experiment a stress σ_0 is suddenly imposed and the evolution of strain is followed with time. This initial phase of creep is one of retarded compliance and is characterized by retardation time constants. Retardation data can be represented as strain γ versus time or as creep compliance J versus time. The complex shear compliance J^* is the quotient of the strain and the stress (the inverse of the complex shear modulus G^*). Being the inverse of the modulus, a material of higher compliance is less rigid,

or softer [Equation 51]:

$$J^*(t) = \frac{\gamma(t)}{\sigma(t)} \tag{51}$$

Equation (35) gives the relationship between strain γ and time for the Burgers model (Kelvin–Voigt element and Maxwell element in series), and Figure 20 shows the corresponding creep curve. Dividing Equation (35) through by the stress, gives Equation (52) for the time-dependent creep compliance as a function of time t:

$$J(t) = \frac{1}{G_2} + \frac{t}{\eta_2} + \frac{1}{G_1}(1 - e^{-t/\lambda_1}) \tag{52}$$

or, expressed only in terms of the compliance [Equation 53],

$$J(t) = J_2 + \frac{t}{\eta_2} + J_1(1 - e^{-t/\eta_1 J_1}) \tag{53}$$

where the subscripts refer to Figure 26.

Real materials will generally have a spectrum of relaxation times; hence a general form of Equation (53), an extended Burgers model, is given by Equation (54) for n Voigt elements:

$$J(t) = J_0 + \sum_{i=1}^{n} J_i(1 - e^{-t/\eta_i J_i}) + \frac{t}{\eta_0} \tag{54}$$

where J_0 is the instantaneous compliance and η_0 is the equilibrium Newtonian viscosity reached at long times. Figure 26 shows a representation of a general creep mechanical model corresponding to Equation (54), for $n = 2$. A test for linear viscoelasticity is that $J(t)$ is independent of σ_0. A rule of thumb is that one viscoelastic unit is needed per decade of experimental frequency.

A property of Voigt units is that they cannot move for ever, i.e. individually, they are not capable of continuous flow. The spring will lengthen under stress, retarded by the viscous dissipation represented by the dashpot, but will reach an equilibrium extension and stop. The rate of lengthening of a given Voigt element is given by η/G, the ratio of the dashpot viscosity to the spring constant, called the retardation time. In a real material, represented by a series of Voigt elements, the Voigt element with the shortest retardation time (with a low viscosity and/or high rigidity) reaches equilibrium extension first and then ceases to move. Further movement of the material is due to the remaining Voigt elements as they approach their own equilibrium deformation, in order of increasing retardation time. If the applied stress is greater than the "yield stress" (in quotes because of the uncertain nature of the concept), then springs will start to break and the system will achieve steady flow, described by the single dashpot η_0 (the Newtonian viscosity). When steady-state flow (steady-state compliance) has been achieved in the retardation step, the stress σ_0 is removed and, if any elastic stress remains, the system will recover some strain, or show recoil (Figure 27). The recovery curve can be fitted by a recoil function or by an extended Burgers model analysis as above.

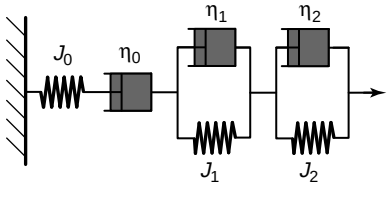

Figure 26 Generalized Burgers model for creep retardation analysis corresponding to Equation (54) for $n = 2$.

8.1.3.2 Analysis of Nonlinear Viscoelasticity Using Creep

Viscoelastic materials are not only sensitive to the rate of strain but also to the total strain. Thus, there is a critical strain beyond which the material response is nonlinear (non-Hookean), in which the springs are breaking (i.e. the structures or interactions responsible for the elastic character are breaking down). Such a system displays nonlinear viscoelasticity, for which a general theoretical model is lacking. The equations describing linear viscoelastic behavior are invalid in the nonlinear regime, thus oscillatory methods are normally restricted to the linear regime. However, few real-world processes for either coatings or other commercial fluids actually take place within the linear regime. Many processes are high strain or even infinite strain, and therefore the nonlinear viscoelastic properties will govern the material behavior. Creep offers a way of characterizing materials in the nonlinear, as well as the linear, regime.

In a creep experiment a torque is suddenly applied to the specimen (or suddenly removed, in the recovery step).

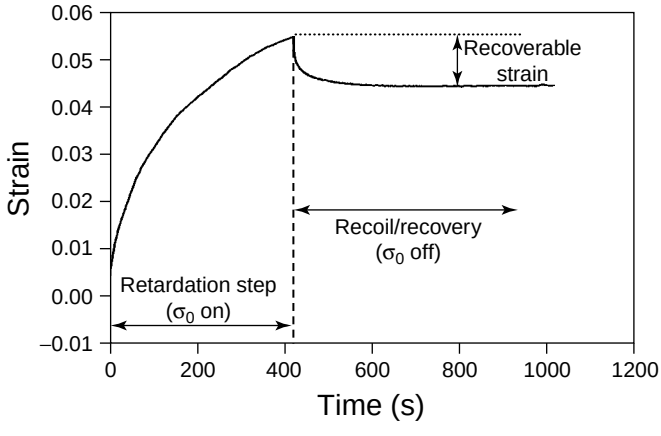

Figure 27 Showing elastic recoil or recovery after removal of the stress σ_0.

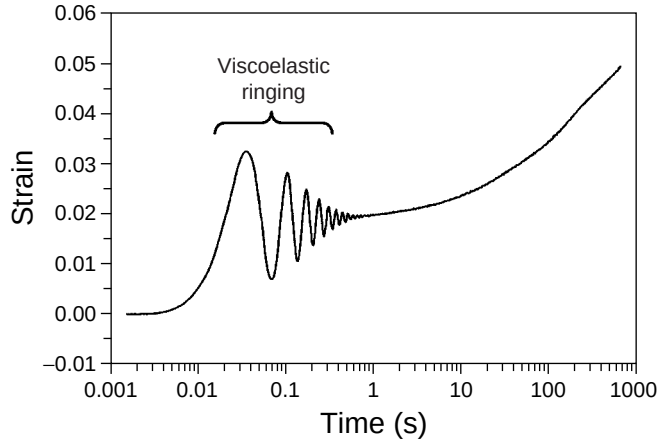

Figure 28 Illustrating viscoelastic ringing, the damped oscillation of strain when a stress σ_0 is rapidly imposed, at the onset of a creep experiment. A similar oscillation occurs when σ_0 is suddenly removed for the recoil step.

The sudden acceleration, together with the measurement system's inertia, causes a strain overshoot, and for viscoelastic materials this can result in viscoelastic ringing, where the material undergoes a damped oscillation just like a bowl of jello that is bumped (Figure 28). The oscillations can be rather short-lived, and may not be apparent unless the time axis is logarithmic. Normand and Ravey[82] showed that this coupling between the instrumental inertia and material elasticity can be analyzed according to standard viscoelastic models. They also found that this modified creep analysis can both replace standard forced oscillation methods and extend the accessible frequency beyond the typical range of such methods. Baravian and Quemada[45] have shown in addition that the method can be used to characterize nonlinear viscoelasticity and to separate the effects of thixotropy from viscoelastic relaxation.

8.2 Oscillatory Shear

If, instead of applying a strain (or, interchangeably, stress) in a constant direction, the direction of the strain undergoes cyclic reversal, then the specimen is subjected to oscillatory deformation, sometimes also referred to as a dynamic test (section 4.6.1). (Note that it is immaterial whether the independent variable is the stress or the strain, for oscillatory linear viscoelastic measurements. Wherever the discussion mentions stress, one may equally well read strain.) Most commercial rotational rheometers are capable of oscillatory shear using a variety of test geometries, although there is now available an oscillatory compression rheometer for liquids, utilizing the parallel plate geometry (TA Instruments CP-20). Oscillatory tube flow has also been used for dynamic testing.

Oscillatory testing is normally done in the linear viscoelastic regime, where the stress and strain are related by a linear differential equation. The usual practice is to employ sinusoidal variation of strain or stress. Testing modes in oscillation include:

1. torque sweep or strain sweep at constant frequency;
2. frequency sweep at constant strain;
3. time or temperature sweep at constant frequency.

Oscillation is a convenient way of characterizing the medium to short timescale response of viscoelastic materials, in the LVER. The employment of small strains (below the critical strain for the LVER) at low frequencies is a reasonable simulation of the physical state of a newly applied coating layer in the process of film formation. Hence, the low-frequency linear viscoelastic properties of coatings are of interest in understanding this phase of coating performance, especially if measured subsequent to a high-shear high-strain experience.

8.2.1 Torque Sweep/Strain Sweep

Torque sweep on a CS rheometer or strain sweep on a CR rheometer are used to determine the strain sensitivity of the viscoelastic parameters, and thus determine the critical strain γ_{cr}, which is the strain limit for linear viscoelastic response. The elastic modulus G' is a sensitive indicator (more so than the dynamic viscosity) of the breakdown of structure, and is often used to determine the critical strain. The experiment is done by applying a constant frequency (e.g. 1 Hz) while increasing the torque gradually, so that the strain amplitude steadily increases. Modulus G' will be constant within the linear viscoelastic regime. The critical strain is the point where G' begins to decrease (Figure 29).

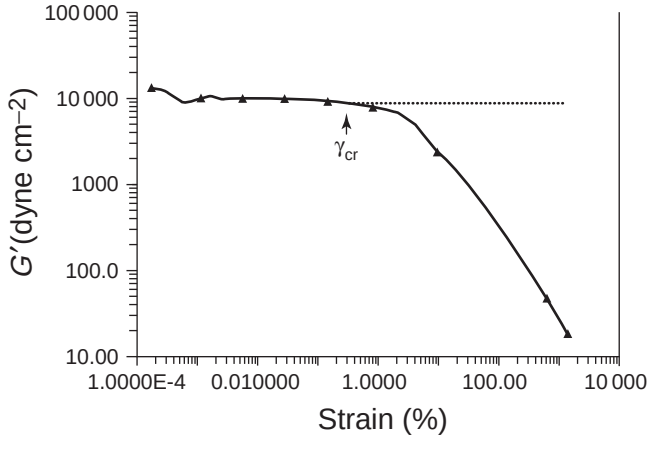

Figure 29 Plot of G' versus strain showing the critical strain, where G' ceases to be constant.

The strain is then held constant at the critical value for subsequent frequency sweep experiments, ensuring that the deformation remains within the linear regime.

The critical strain is not a single value for a given material, however, but is a decreasing function of frequency, a consequence of the superposition of viscous and elastic response to deformation in a viscoelastic material. At lower frequencies, the material can undergo more strain due to viscous flow, without fracturing. However, a plot of G' versus oscillatory stress shows that the critical stress increases with frequency.[83] This behavior is analogous to the observation that the measured yield stress value is an increasing function of the rate of loading, and in fact the mechanisms underlying the two effects are identical.

Polymeric systems can have quite large critical strains (above 10%), whereas γ_{cr} for colloidal systems can be quite small (below 0.1%). The critical strain value is a measure of the brittleness or ductility of the microstructure, and an indicator of the relative length scale of the forces responsible for the elastic behavior. Difficulty is often encountered in finding the linear regime for thixotropic materials composed of flocculated or associated particles, because the structure is rather brittle – the attractive forces are short range, and the critical strain is consequently small.

8.2.2 *Frequency Sweep*

In a frequency sweep experiment, the oscillation frequency is steadily increased, while the strain is held constant below the critical value. Typically, the frequency is swept logarithmically with time, so that the discrete frequency points are linear on a log axis. A typical plot of viscoelastic functions for a frequency sweep experiment is shown in Figure 30. G' and G'' normally increase with frequency, whereas η' decreases with frequency, trends which can be understood by examination of the mechanical models' response. As pointed out in section 4.6.2, values for the viscoelastic functions can be obtained from frequency-sweep data, e.g. the frequency, w_c, where, G' and G'' cross, equals the inverse of the relaxation time (in the Maxwell model). A gel is distinguished by having a nearly constant value of G' with frequency, since a gel will not have a finite relaxation time.

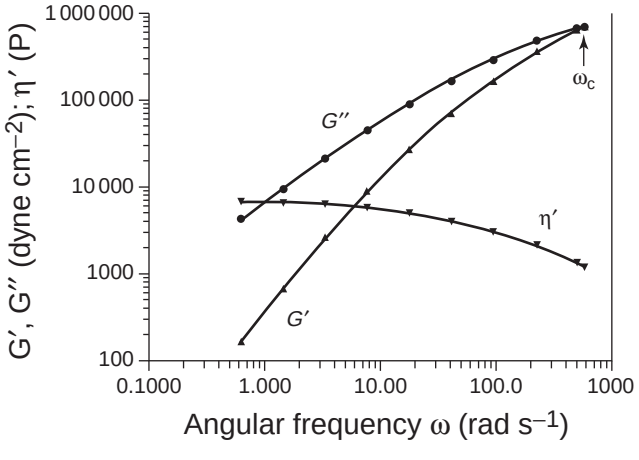

Figure 30 Typical plot of the dependence of the viscoelastic rheometrical functions on oscillatory frequency.

Frequency sweep experiments are in some ways analogous to the flow curve measured in steady shear. In fact, the empirical Cox–Merz rule asserts the equivalence of steady shear rate and angular frequency in oscillatory deformation. The Cox–Merz rule means that steady shear and oscillatory data should be superimposable, and typically works well for polymeric materials. However, the rule generally fails for colloidal or particulate systems, because the average microstructure during the period of oscillation is not representative of that achieved in steady shear, due to the rapid structural equilibration of such systems. The value of the Cox–Merz rule is probably greatest for characterization of polymers, as an answer to the problem of elastic artifacts (such as the Weissenberg effect), which usually limit the steady shear rate for entangled polymer melts to around $1\,s^{-1}$. By use of the rule, one can extend polymer steady shear data to higher shear rates using oscillatory data. A modified Cox–Merz rule, called the Rutgers–Delaware rule,[84] apparently works for dispersions, provided the structural recovery time constant is slow relative to the period of oscillation. The microstructure must equilibrate to that characteristic of steady shear at an equivalent effective maximum dynamic shear rate. This effective maximum dynamic shear rate is determined from the maximum strain rate amplitude [Equation 55]:

$$\dot{\gamma}_{max} = w\gamma_0 = \dot{\gamma}_{eff} \tag{55}$$

8.2.3 *Time or Temperature Sweep*

A time sweep at constant frequency is used to characterize recovery of structure after shear, such as for a thixotropic material (section 4.4.3.3). Emulsion stability can be indicated by temperature sweep experiments at constant frequency.[83] An oscillatory temperature sweep is an excellent way to characterize the chemical advancement of thermosetting polymers, a technique used in a subdiscipline known as chemorheology (the study of the rheology of chemically reacting systems).[85,86]

8.2.4 *Experimental Protocol*

Similar to the discussion of shear-history effects (section 8.1.2.1), the rheological state of a specimen is of concern, with an eye to obtaining repeatable results.

Consequently, the following series of experimental steps is recommended:

1. The specimen should be prior-sheared in steady shear at an arbitrary high shear rate or stress value, to put the specimen in a low-structure state.
2. A constant-frequency small-strain oscillatory time-sweep step should be commenced immediately, following the structural recovery to a constant-property state, if possible.
3. Once the specimen has recovered to equilibrium properties, either of two experiments can be done:
   (a) oscillatory frequency-sweep experiment;
   (b) steady-shear flow curve ramp experiment.

The above protocol is a logical sequence which enables tests to be done on a specimen which is in a repeatable rheological state.

8.2.5 Linearity

As discussed above, the test of linearity in the viscoelastic response is that the rheometrical functions G', G'', η' are independent of the strain amplitude or stress amplitude. When they cease to be constant, the critical value of stress or strain has been exceeded. Another manifestation of nonlinearity is that the response function contains not only the fundamental exciting frequency w, but also harmonics of the fundamental.

LIST OF SYMBOLS

Φ	fluidity integral
β	time constant characteristic of a transition shear rate
γ	dimensionless shear strain; surface tension
$\dot{\gamma}$	shear strain rate (s^{-1})
δ	phase shift; thickness of adsorbed polymer layer
η	coefficient of viscosity
η_0	zero-shear viscosity
η_∞	high-shear limiting Newtonian viscosity
η_d	dispersion viscosity
η_l	liquid-phase viscosity
η_r	relative viscosity (η_d/η_l)
η'	dynamic viscosity
$[\eta]$	intrinsic viscosity (limiting viscosity number)
θ	angle of inclination with respect to the vertical
λ	wavelength of coating surface striations; elastic stress relaxation time constant
π	ratio of circumference of a circle to its diameter
ρ	density
σ	shear stress; generalized stress
σ_0	yield stress; stress at zero-time
σ_e	tensile stress
ϕ	volume fraction of internal (dispersed) phase of a dispersion
ϕ_{eff}	effective volume fraction for a dispersion
ϕ_m	maximum packing fraction of a dispersion, where $\eta \to \infty$
w	angular frequency ($=2\pi f$ rad s^{-1})
A	area of shear face
De	Deborah number (λ/t)
F	force
G	shear modulus
G_∞	plateau modulus
G^*	complex shear modulus
G'	storage modulus
G''	loss modulus
K	consistency in power law model
R	universal gas constant; spherical particle radius
T	absolute temperature
T_g	glass transition temperature
a	curvature parameter in Equation (16)
a_t	amplitude of coating surface striations at time t
a_0	time-zero amplitude of coating surface striations
e	base of the natural logarithmic scale = 2.71828
f	frequency
g	gravitational acceleration (cm s^{-2})
h	film thickness
k_B	Boltzmann's constant
l	length
n	power law exponent
t	time
v	velocity
x	coordinate parallel to surface
y	coordinate normal to surface

ABBREVIATIONS AND ACRONYMS

CR	Controlled-strain-rate
CS	Controlled-stress
LVER	Linear Viscoelastic Region

REFERENCES

1. E.C. Bingham, U.S. Bureau of Standards Scientific Paper No. 278, 1916.
2. N. Casson, in *Rheology of Disperse Systems*, ed. C.C. Mill, Pergamon Press, New York, 1959.
3. W.K. Asbeck, 'Fundamentals of the Rheology of Pigment Dispersions', *Official Digest, Federation of Societies for Paint Technology*, **33**(432), 65–83 (1961).

4. R. Lapasin, G. Torriano, 'Rheological Models Suitable for Application in the Paint Field', *Proc. 18th FATIPEC Congress, Venice*, **4**, 1–41 (1986).
5. R.L. Hoffman, 'Discontinuous and Dilatant Viscosity Behavior in Concentrated Suspensions. I: Observation of a Flow Instability', *Trans. Soc. Rheol.*, **16**, 155–172 (1972).
6. I.M. Krieger, 'Rheology of Polymer Colloids', in *Polymer Colloids*, eds. R. Buscall, T. Corner, J.F. Stageman, Elsevier Applied Science, New York, 1986.
7. G.N. Choi, 'Rheology of Sterically-stabilized Model Colloidal Dispersions', Ph.D. Thesis, Case Western Reserve University, Cleveland, OH, 1982.
8. A.W. Sisko, *Ind. Eng. Chem.*, **50**, 1789 (1958).
9. M.M. Cross, 'Rheology of Non-Newtonian Fluids: A New Flow Equation for Pseudoplastic Systems', *J. Coll. Sci.*, **20**, 417–437 (1965).
10. P.J. Carreau, 'Rheological Equations from Molecular Network Theories', *Trans. Soc. Rheol.*, **16**, 99–127 (1972).
11. C.A. Hieber, H.H. Chiang, 'Shear-rate-dependence Modeling of Polymer Melt Viscosity', *Polym. Eng. Sci.*, **32**(14), 931–938 (1992).
12. S.A. Jenekhe, 'The Rheology and Spin Coating of Polyimide Solutions', *Polym. Eng. Sci.*, **23**(15), 830–834 (1983).
13. J.W. Goodwin, R.H. Ottewill, 'Properties of Concentrated Colloidal Dispersions', *J. Chem. Soc. Farad. Trans.*, **87**(3), 357–369 (1991).
14. J. Mewis, P. D'Haene, 'Prediction of Rheological Properties in Polymer Colloids', *Makromol. Chem., Macromol. Symp.*, **68**, 213–225 (1993).
15. J. Laven, H.N. Stein, 'The Peclet Number for Shear Thinning in Hard Sphere Dispersions', *J. Rheol.*, **36**(3), 497–501 (1992).
16. I.M. Krieger, T.J. Dougherty, 'A Mechanism for Non-Newtonian Flow in Suspensions of Rigid Spheres', *Trans. Soc. Rheol.*, **3**, 137 (1959).
17. A. Toussaint, 'Choice of Rheological Model for Steady Flow', *Prog. Org. Coat.*, **21**, 255–267 (1992).
18. L.E. Silbert, J.R. Melrose, R.C. Ball, 'The Rheology and Microstructure of Concentrated, Aggregated Colloids', *J. Rheol.*, **43**(3), 673–700 (1999).
19. D.C.-H. Cheng, 'Viscosity–Concentration Equations and Flow Curves for Suspensions', *Chem. Ind.*, 17 May, 403 (1980).
20. R.L. Hoffman, 'Explanations for the Cause of Shear Thickening in Concentrated Colloidal Suspensions', *J. Rheol.*, **42**(1), 111–123 (1998).
21. W.H. Boersma, P.J.M. Baets, J. Laven, H.N. Stein, 'Time-dependent Behavior and Wall Slip in Concentrated Shear-thickening Dispersions', *J. Rheol.*, **35**(6), 1093–1120 (1991).
22. D.C.H. Cheng, 'Thixotropy', *Int. J. Cosmetic Sci.*, **9**, 151 (1987).
23. H. Freundlich, *Kapillärchemie*, 3rd edition, Leipzig, 1932.
24. R.R. Eley, Unpublished Data.
25. J.S. Dodge, 'Quantitative Measures of Leveling', *J. Paint Technol.*, **44**(564), 72–78 (1972).
26. E.R. Lang, C. Rha, 'Rheology', in *Fluids*, eds. G. Astarita, G. Marucci, L. Nicolais, Plenum Press, New York, 659, Vol. 2, 1980.
27. H.A. Barnes, K. Walters, 'The Yield Stress Myth?', *Rheol. Acta*, **24**, 323–326 (1985).
28. H.A. Barnes, 'The Yield Stress – a Review or ($\pi\alpha\nu\tau\alpha\rho\epsilon\iota$) – Everything Flows?', *J. Non-Newtonian Fluid Mech.*, **81**, 133–178 (1999).
29. Z. Zhou, M.J. Solomon, P.J. Scales, D.V. Boger, 'The Yield Stress of Concentrated Flocculated Suspensions of Size-distributed Particles', *J. Rheol.*, **43**(3), 651–671 (1999).
30. G. Astarita, 'The Engineering Reality of the Yield Stress', *J. Rheol.*, **34**(2), 275–277 (1990).
31. D.C.H. Cheng, 'Yield Stress: A Time-dependent Property and How to Measure It', *Rheol. Acta*, **25**, 542–554 (1986).
32. K.L. Johnson, *J. Mech. Physics Solids*, **18**, 115–126 (1970).
33. Q.D. Nguyen, D.V. Boger, 'Thixotropic Behaviour of Concentrated Bauxite Residue Suspensions', *Rheol. Acta*, **24**, 427–437 (1985).
34. Q.D. Nguyen, D.V. Boger, 'Yield Stress Measurement for Concentrated Suspensions', *J. Rheol.*, **27**(4), 321–349 (1983).
35. H.A. Barnes, J.O. Carnali, 'The Vane-in-cup as a Novel Rheoloeter Geometry for Shear Thinning and Thixotropic Materials', *J. Rheol.*, **34**(6), 841–866 (1990).
36. R. Keunings, D.W. Bousfield, 'Analysis of Surface Tension-driven Leveling in Viscoelastic Films', *J. Non-Newtonian Fluid Mech.*, **22**, 219–233 (1987).
37. D.F. Massouda, 'Analysis and Prediction of Roll-spatter From Latex Paints', *J. Coat. Technol.*, **57**(722), 27–36 (1985).
38. J.E. Glass, 'Dynamics of Roll-spatter and Tracking. Part IV. Importance of G^* Recovery and N_1 in Tracking', *J. Coat. Technol.*, **50**(641), 72–78 (1978).
39. J.R. Van Wazer, J.W. Lyons, K.Y. Kim, R.E. Colwell, 'Viscosity and Flow Measurement', Interscience, New York, 1963.
40. R.B. Bird, R.C. Armstrong, O. Hassager, 'Dynamics of Polymeric Liquids, Vol. 1: Fluid Mechanics', John Wiley & Sons, New York, 1987.
41. Judges 5 : 5, Translation of M. Reiner.
42. T.A. Strivens, 'The Viscoelastic Properties of Concentrated Suspensions', *Coll. Polym. Sci.*, **261**, 74–81 (1983).
43. T. Annable, R. Buscall, R. Ettelaie, D. Whittlestone, 'The Rheology of Solutions of Associating Polymers: Comparison of Experimental Behavior with Transient Network Theory', *J. Rheol.*, **37**(4), 695–726 (1993).
44. J.E. Glass, 'Dynamics of Roll-spatter and Tracking. Part III: Importance of Extensional Viscosities', *J. Coat. Technol.*, **50**(641), 56–71 (1978).
45. C. Baravian, D. Quemada, 'Using Instrumental Inertia in Controlled Stress Rheometry', *Rheol. Acta*, **37**, 223–233 (1998).

46. A. Einstein, *Annal. der Physik*, **17**, 459 (1905); **19**, 271 (1906); **34**, 591 (1911).
47. I.M. Krieger, 'Rheology of Polymer Colloids', in *Polymer Colloids*, eds. R. Buscall, T. Corner, J.F. Stageman, Elsevier Applied Science, New York, 1985.
48. G.K. Batchelor, *J. Fluid Mech.*, **83**, 97 (1977).
49. H. Eilers, *Koll. Zeitschrift*, **97**, 913 (1941).
50. I.R. Rutgers, *Rheol. Acta*, **2**, 202 and 305 (1962).
51. M. Mooney, *J. Coll. Sci.*, **6**(2), 162 (1951).
52. I.M. Krieger, T.J. Dougherty, *Trans. Soc. Rheol.*, **3**, 137 (1959).
53. J.W. Goodwin, R.H. Ottewill, 'Properties of Concentrated Colloidal Dispersions', *J. Chem. Soc., Farad. Trans.*, **87**(3), 357–369 (1991).
54. A. Doroszkowski, R. Lambourne, 'A Viscometric Technique for Determining the Layer Thickness of Polymer Adsorbed on Titanium Dioxide', *J. Coll. Interf. Sci.*, **26**, 214–221 (1968).
55. J.W. Goodwin, 'Some Uses of Rheology in Colloid Science', in *Colloidal Dispersions*, ed. J. Goodwin, Henry Ling, Ltd., Dorset Press, Dorchester, 1982.
56. I.M. Krieger, M. Eguiluz, 'The Second Electroviscous Effect in Polymer Latices', *Trans. Soc. Rheol.*, **20**, 29–45 (1976).
57. R.R. Eley, 'Rheology and Viscometry', in *ASTM Paint and Coatings Testing Manual*, 14th edition, ASTM, Philadelphia, 333–368, 1995.
58. D.W. Bousfield, 'Long-wave Analysis of Viscoelastic Film Leveling', *J. Non-Newtonian Fluid Mech.*, **40**, 47–54 (1991).
59. D.W. Bousfield, 'A Model to Predict the Leveling of Coating Defects', *J. Tech. Assoc. Pulp Paper Ind.*, May, 163–170 (1991).
60. M. Camina, D.M. Howell, 'The Leveling of Paint Films', *J. Oil Col. Chem. Assoc.*, **55**, 929–939 (1972).
61. R.E. Smith, 'Brookfield Viscometers for Determining Low-shear Viscosity and Leveling Behavior', *J. Coat. Technol.*, **54**(694), 21–29 (1982).
62. W.S. Overdiep, 'The Leveling of Paints', *Prog. Org. Coat.*, **14**, 159–175 (1986).
63. M. Klarskov, J. Jakobsen, A. Saarnak, 'Verification of the Orchard Leveling Analysis', *Rheology 92*, March, 30 (1992).
64. L.W. Schwartz, R.R. Eley, *Prog. Org. Coat.* (in press).
65. N.D.P. Smith, S.E. Orchard, A.J. Rhind-Tutt, 'The Physics of Brushmarks', *J. Oil Col. Chem. Assoc.*, **44**(1), 618–633 (1961).
66. S.E. Orchard, *Appl. Sci. Res., A*, **11**, 451 (1962).
67. D.E. Weidner, L.W. Schwartz, R.R. Eley, 'Role of Surface Tension Gradients in Correcting Coating Defects in Corners', *J. Coll. & Interf. Sci.*, **179**, 66–75 (1996).
68. G.P. Bierwagen, 'Surface Defects and Surface Flows in Coatings', *Prog. Org. Coat.*, **19**(1), 59 (1991).
69. S.K. Wilson, 'The Derivation and Analysis of a Model of the Drying Process of a Paint Film', *Surf. Coat. Int.*, **4**, 162 (1997).
70. L.O. Kornum, H.K. Raaschou-Nielsen, 'Surface Defects in Drying Paint Films', *Prog. Org. Coat.*, **8**, 275–324 (1980).
71. M.L. Colclough, N.D.P. Smith, T.A. Wright, 'A Low-shear Viscometer: An Instrument for Measuring Flow and Sag Resistance in Coatings', *J. Oil Color Chem. Assoc.*, **63**, 183 (1980).
72. R.R. Eley, Unpublished Results.
73. C.K. Schoff, 'Rheology', Federation Series on Coatings Technology, Federation of Societies for Coatings Technology, Blue Bell, PA, 1991.
74. K. Walters (ed.), 'Rheometry: Industrial Applications', Research Studies Press, J. Wiley & Sons, New York, 1980.
75. C.W. Macosko, 'Rheology: Principles, Measurements and Applications', VCH Publishers, Inc., New York, 352, 1994.
76. I.M. Krieger, 'The Role of Instrument Inertia in Controlled-stress Rheometers', *J. Rheol.*, **34**(4), 471–483 (1990).
77. R.W. Whorlow, 'Rheological Techniques', 2nd edition, John Wiley & Sons, New York, 1992.
78. C.W. Macosko, 'Rheology: Principles, Measurements and Applications', VCH Publishers, Inc., New York, Chapter 5, 1994.
79. A.A. Collyer, D.W. Clegg (eds.), 'Rheological Measurement', Chapman & Hall, New York, 1998.
80. J.M. Dealy, K.F. Wissbrun, 'Melt Rheology and its Role in Plastics Processing', Chapman & Hall, New York, 1995.
81. T. Annable, R.A. Brown, J.C. Padget, A. van den Elshout, 'Improvements in the Application Properties of Water-based Low VOC Coatings', *Surf. Coat. Int.*, **7**, 321 (1998).
82. V. Normand, J.-C. Ravey, 'Dynamic Study of Gelatin Gels by Creep Measurements', *Rheol. Acta*, **36**, 610–617 (1997).
83. P. Whittingstall, Unpublished Lecture Notes, TA Instruments Rheology Symposium, Lake Tahoe, 1999.
84. I.M. Krieger, 'Comments on a Manuscript by Doraiswamy et al.', *J. Rheol.*, **36**(1), 215–216 (1992).
85. R.R. Eley, 'Thermosetting Coatings: Analytical and Predictive Capability by Chemorheology', *J. Coat. Technol.*, **56**(718), 49–56 (1984).
86. S.G. Yeates, T. Annable, B.J. Denton, G. Ellis, R.M.D. Nasir, D. Perito, I. Parker, 'Rheology of Acrylic Powder Coatings', *J. Coat. Technol.*, **68**(861), 107–114 (1996).

Thermal Analysis of Coatings

Edwin F. Meyer III
ICI Paints Research Center, Strongsville, USA

1 Introduction **1**
2 Differential Scanning Calorimetry **2**
2.1 History 2
2.2 Differential Scanning Calorimetry Fundamentals 2
2.3 Sample Preparation 3
2.4 Experimental Conditions 3
2.5 Applications 4
2.6 Method Development 8
3 Dynamic Mechanical Analysis **8**
3.1 History 8
3.2 Dynamic Mechanical Analysis Fundamentals 8
3.3 Sample Preparation 9
3.4 Experimental Conditions 9
3.5 Applications 9
3.6 Dynamic Mechanical Analysis Method Development 10
4 Thermomechanical Analysis **10**
4.1 History 10
4.2 Thermomechanical Analysis Fundamentals 11
4.3 Sample Preparation 11
4.4 Experimental Conditions 11
4.5 Applications 11
4.6 Thermomechanical Analysis Method Development 11
5 Thermogravimetric Analysis **12**
5.1 History 12
5.2 Thermogravimetric Analysis Fundamentals 12
5.3 Sample Preparation 12
5.4 Experimental Conditions 12
5.5 Applications 12
5.6 Method Development 13
6 Summary **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **14**

The thermal analysis of coatings involves measuring the response of a material when subjected to a thermal stimulus, sometimes in combination with a mechanical stress. The thermal stimuli used in the thermal analysis of coatings include linear temperature ramps, a sinusoidal modulation around a fixed temperature, the superposition of a linear ramp and a sinusoidal modulation, and a simple isotherm in which some property of the material is monitored as a function of time. The responses measured by the thermal analyst include: melting, crystallization, evaporation, degradation, stress relaxation, softening point, and glass transition, as well as physical properties such as the elastic and viscous moduli. These responses aid in the development, characterization, and determination of the proper application and bake conditions for a wide range of coatings.

In general, thermal analysis techniques are very sensitive and, for many applications, a sample mass of the order of 10 mg is sufficient. However, the assignment of a thermal event to a particular chemistry or morphology requires knowledge of the system being analyzed or another analytical technique. For example, thermal analysis can accurately determine the temperature and enthalpy of a crystalline polymer melt but it does not provide any other information as to the chemistry of the polymer. Assigning a cause to a thermal event becomes increasingly difficult as the nature of the material becomes more complicated. Such is the case with a fully formulated coating which contains a wide range of polymeric and oligomeric materials, as well as various additives. To determine which chemistry is responsible for a particular thermal transition, other analytical techniques are used, frequently in direct combination with the thermal analysis technique.

1 INTRODUCTION

The end users of coatings often demand that a particular product meet specifications that involve end-use properties, such as: water resistance, scrub resistance, detergent resistance, flexibility, hardness, impact resistance, and degree of tackiness. Besides these requirements of the final coating, there are also application-related criteria that have to be met. These include low-temperature film formation, the degree of cure at a given time and temperature, and the time taken to dry. All these specifications can be related back to fundamental chemical and physical properties of the coating, using thermal analysis. In general, thermal analysis is simply the characterization of the properties of a material as a function of temperature. This definition is frequently expanded to include isothermal experiments performed on conventional thermal analysis instruments. Standard techniques include differential scanning calorimetry (DSC), thermogravimetric analysis (TGA), dynamic mechanical analysis

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

(DMA), dielectric analysis (DEA), and thermomechanical analysis (TMA). Variations on these techniques include modulated temperature differential scanning calorimetry (MTDSC) and modulated thermogravimetric analysis (MTGA). In these two techniques the temperature program of the sample is modulated, which provides superior results in virtually all cases.

Recent advances in thermal analysis include microthermal calorimetry. This technique involves the marriage of modulated temperature spectroscopy and atomic force microscopy (AFM) by effectively placing a modulated temperature probe on the tip of an AFM probe.

2 DIFFERENTIAL SCANNING CALORIMETRY

2.1 History

A great step toward modern DSC was achieved in 1899 by William Chandler Roberts-Austen, then chemist to the Royal Mint, when he conceived the idea of measuring the temperature difference between the sample and a thermally inert material. Previously, all attempts at determining the heating curve for a material – at that time mostly clays and minerals – involved only the measurement of the temperature of the material as a function of time. The advantage of the differential method is twofold. First, a differential method is more precise and, second, the differential method eliminates the need to use time as a dependent variable. That is, the differential method produces a ΔT vs T curve, where T is the temperature of the material and ΔT is the temperature difference between the material and the thermally inert reference.

From this time until very recently the advances toward modern DSC were mainly instrumental. For example, better temperature measurement and control, control of the atmosphere within the cell, and a greater symmetry between the sample and reference positions. Another great leap forward in DSC came in 1992 when Dr Mike Reading, then of ICI Paints, conceived the idea of superimposing a modulation on top of the linear sample temperature rate. Prior to this development the linear temperature rate performed two functions: to supply the temperature difference between the sample side and the reference side and the temperature regime of interest. A modulation of the temperature of the sample side can supply a useful temperature difference between the two cell positions, allowing the experimentalist to independently adjust the rate at which the temperature regime of interest is scanned. Indeed, the temperature scan rate or linear underlying temperature rate can be set to zero. The technique of modulating around a fixed temperature and monitoring the output signal as a function of time has been called quasi-isothermal MTDSC. Of course, modulating the temperature of the sample side will produce a modulated temperature difference between the sample and reference. This modulated temperature difference acts as a continuous probe of the heat capacity of the sample – irrespective of any kinetic events occurring within the sample. If the heat capacity of the sample increases, the amplitude of the thermal energy transfer between the furnace and the sample, needed to produce the programmed temperature modulation, will increase proportionally. This change in the amplitude of the thermal energy transfer will be manifest in the temperature difference between the sample and the reference.

2.2 Differential Scanning Calorimetry Fundamentals

The technique of DSC is used to detect and quantify thermal events in a material. These thermal events include glass transitions, melts, crystallizations, stress relaxations, chemical reactions, and volatilizations. The working part of the instrument is the cell (Figure 1). The cell contains two sites – sample and reference – and is enclosed by a furnace block that, ostensibly, is thermally symmetric with respect to the sample side and reference side. Thermocouple junctions just underneath each position monitor both the temperature of the sample side, $T_{\text{samp side}}$, and the difference between the temperature of the sample side and the temperature of the reference side, $T_{\text{ref side}} - T_{\text{samp side}}$. If the cell is empty and thermally symmetric, the temperature of the reference side will always be the same as the temperature of the sample side, irrespective of the manner in which the furnace block is heated or cooled.

To perform a typical DSC experiment a sample is placed in a small aluminum pan and the pan is then placed on the sample side. To thermally balance the pan

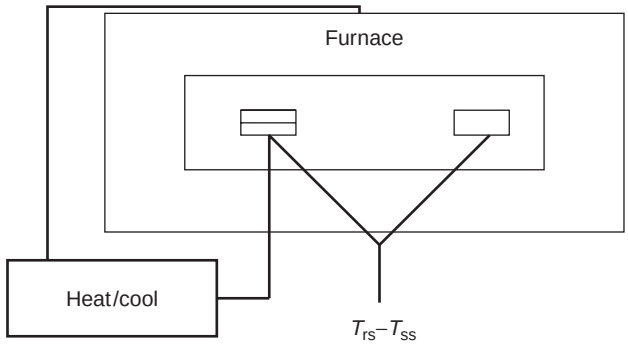

Figure 1 Simple diagram of the DSC sample cell. The output signal is the temperature difference between the sample side and the reference side.

containing the sample, an empty, identical reference pan is placed on the reference side.

With this configuration any temperature difference between the reference side and the sample side can be attributed solely to the presence of the sample. Further, it can be shown – assuming Newton's law of cooling describes the thermal energy transfer – that this temperature difference is directly proportional to the rate of thermal energy transfer between the furnace and the sample.

2.3 Sample Preparation

A thermal analyst for a coatings company can divide all sample submissions into two groups, commercial and research. Research samples are usually less problematical because their history is well known – the samples have been prepared and baked on-site. Commercial samples offer a different set of challenges because the coating has been applied at the customer's site under line conditions that are usually not as well controlled or as well characterized as those in a research center.

For this reason, when trouble-shooting a problem with a commercial coating it is important to obtain a coating sample directly from the substrate upon which the problem occurred. For example, if the thermal analyst is to trouble-shoot a film-whitening problem on the exterior of tuna fish cans at a plant in Ecuador, it is important that he/she analyzes coating taken directly from a can that exhibited the film-whitening. The reason for this is twofold. First, the coating may have changed from the time it left the facility where it was manufactured and when it was applied and, second, virtually all the physical properties of the coating are a strong function of the bake conditions.

One method of obtaining a coating that has been baked on a substrate is simply to carefully scrape off a small amount with a bladed instrument. Ideally, the sample mass for a DSC experiment is about 5 mg, although lighter samples can be used. The thinner the coating, the larger the area that must be scraped to obtain a massive enough sample. In the case of interior beer and beverage coatings, which can be as thin as 2 μm, about 25 cm^2 of coating will need to be removed to obtain a 5 mg sample. The scrapings from the coating are placed in a cylindrical aluminum pan which, typically, is about 1 mm high and 5 mm in diameter. To optimize the thermal energy transfer, the sample should be spread out evenly in the pan. To ensure good thermal contact between the sample and the pan, an aluminum lid is placed on top of the sample and then crimped with the pan, thus forming a aluminum sandwich with the sample. If the sample may contain volatiles, it is good practice to make a hole in the pan lid before crimping. This will prevent a pressure increase inside the sample pan during heating.

One problem that may arise when scraping a coating from a soft metal is that some metal is scraped off with the coating. This will make the results nonquantitative because the total mass of the sample does not represent the mass of the coating. However, the results will be qualitative and, in many cases, this is good enough.

2.4 Experimental Conditions

For both linear temperature differential scanning calorimetry (LTDSC) and MTDSC the technician must choose a purge gas and a flow rate. The purge gas, which should be inert, will displace moisture and oxygen in the DSC cell and assist in the transfer of thermal energy between the furnace, the reference, and the sample. Although helium, due to its high speed at a given temperature, provides a more efficient thermal energy transfer, nitrogen is conventionally used for virtually all standard applications. For specialized applications where fast heating rates or high modulation amplitudes are used, helium may be used. A typical nitrogen flow rate is 50 $cm^3\ min^{-1}$, but this can be adjusted based on the type of sample being run and the programmed temperature rate of the sample.

Another experimental parameter that is common to both LTDSC and MTDSC is the temperature range. For optimum results it is common practice to choose a starting temperature that is at least 50 °C before any transition and, of course, the final temperature should be sufficiently above any transitions of interest.

The lone experimental parameter that is the reason for the difference between LTDSC and MTDSC is the temperature rate. In LTDSC the operator selects only the linear temperature rate which is constant throughout the run. This temperature rate creates the steady-state temperature difference between the temperature of the sample side and the temperature of the reference side – in other words the output signal. So, a higher temperature rate provides a higher signal-to-noise ratio. Unfortunately, because the linear temperature rate is also the mechanism by which the temperature regime of interest is scanned, a high temperature rate tends to broaden the transitions and thus meld transitions that are close to each other. Often, when running an LTDSC experiment, the operator must compromise between sensitivity and precision.

In MTDSC the temperature difference between the sample side and the reference side – the output signal – can be generated solely by the modulation, leaving the operator free to choose a linear temperature rate that will scan the temperature regime of interest at a slower rate. Choosing the temperature protocol in an MTDSC experiment requires the assignment of three parameters: the underlying linear temperature rate, the temperature modulation amplitude, and the period of the temperature modulation. The first consideration when assigning these parameters is whether the furnace is able to carry out the

modulation. If the period is too small or the amplitude too high for a given linear temperature rate, the actual temperature rate of the sample side will not follow the set protocol. That is, the sine wave will be distorted. A typical set of MTDSC temperature parameters may be an underlying heating rate of $2\,^{\circ}\mathrm{C\,min^{-1}}$, a modulation amplitude of $1\,^{\circ}\mathrm{C}$, and a modulation period of 1 min. If a faster underlying temperature rate is desired, the modulation amplitude may have to be lowered or the period of the modulation raised to ensure that the furnace can keep the sample-side temperature on protocol. A simple technique to check whether the sample side temperature is a sine wave is to plot the modulated temperature of the sample side versus time.

2.5 Applications

In most coatings applications the desired result from a DSC experiment is the T_g (glass transition temperature; or T_gs) of the polymer system. The T_g of a polymer is an important parameter for coatings formulators because it gives an indication of the processing conditions needed in manufacturing, the application conditions needed to produce a film from the wet coating, as well as the performance properties of the final film. Due to the complexity of today's coatings, the characterization of the T_g of a particular coating system can be confounded by various thermal events. These include surfactant melts and crystallizations, stress relaxations, volatilizations, and all kinds of rearrangements of the polymer micromorphology with temperature – especially near to the T_g of one of the components. For this reason, MTDSC is the method of choice for coatings analysts. Also, MTDSC can separate the heat capacity of the coating from kinetic thermal events. A simple example that demonstrates this involves a pilot plant batch and a laboratory batch of a resin.

One of the problems frequently encountered by plant engineers is the scale-up of a reaction. Frequently, a resin made in a 1 gal laboratory reactor will exhibit different physical properties to one made in a 100 gal pilot plant reactor or a 10 000 gal plant reactor, even though the recipe was the same. This is because larger reactors often have different heat flow and mixing characteristics. When the product from two different reactors are significantly different, the thermal analyst is often called upon to characterize the two resins. Figure 2 shows the total heat flow versus temperature for two samples – one manufactured in a 1 gal laboratory reactor and the other manufactured in a 100 gal pilot plant reactor. Here the assignment of T_g is problematical due to kinetic events near the T_g transition. The heat capacity and its derivative are shown in Figure 3. Here the assignment of T_g is a trivial matter, using the peak in the derivative curve. Clearly, the samples have different T_g values.

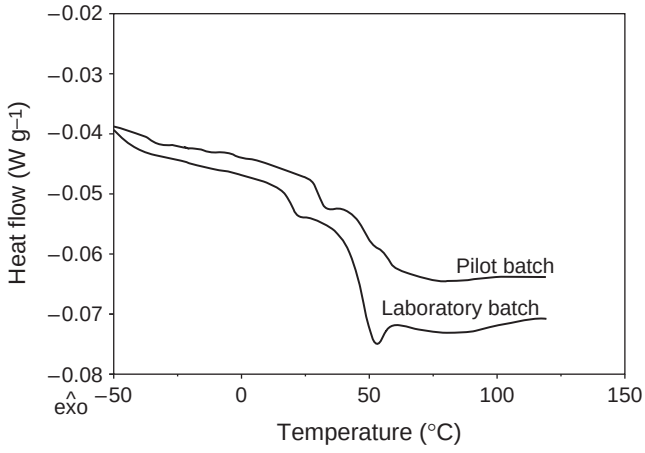

Figure 2 The total heat flow versus temperature curves for a pilot plant batch and a laboratory batch of a particular coating.

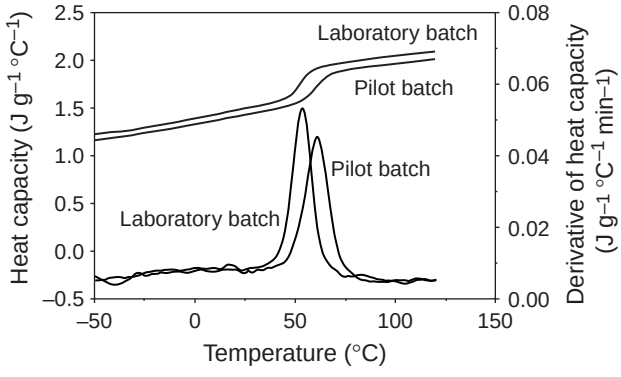

Figure 3 The heat capacity and its derivative versus temperature from the same experiment shown in Figure 2.

Another advantage MTDSC offers over LTDSC is sensitivity. In LTDSC the temperature scan rate through a region of interest must be kept high in order to produce a useful temperature difference between the sample and the reference sides. This precludes the accurate characterization of many minor thermal events. The first example that demonstrates the enhanced sensitivity of the MTDSC protocol involves the emulsion polymerization of vinyl acetate (VA) and of butyl acrylate (BA) monomers with a ladder of concentrations. The ladder consists of five evenly spaced steps from 50/50 VA/BA to 90/10 VA/BA. The samples were prepared by drawing down the emulsion and allowing it to dry for three days.

Figure 4 shows a coplot of the heat capacity and its derivative with respect to temperature plotted versus temperature for the 90/10 concentration ratio. Here there is a sharp transition and an easily defined T_g. The 80/20 results, shown in Figure 5, exhibit a major T_g from the copolymer and a minor T_g from VA homopolymer (ca. $40\,^{\circ}\mathrm{C}$). The VA homopolymer was from the pure

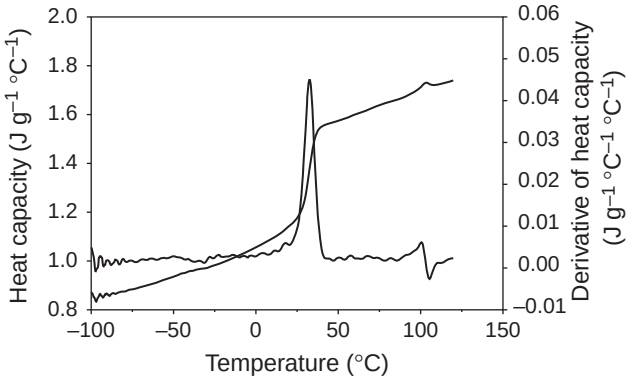

Figure 4 The heat capacity and its derivative for a 90/10 blend of VA/BA.

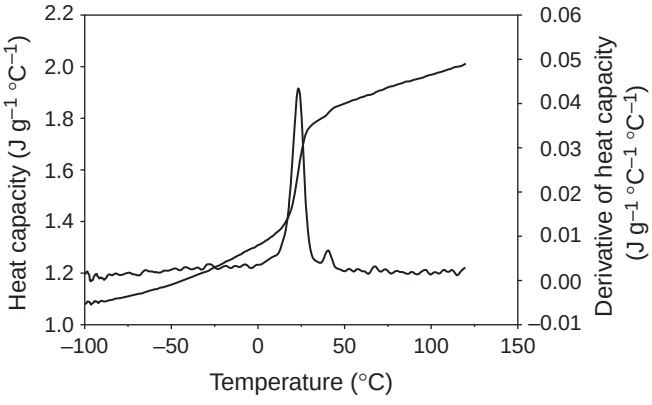

Figure 5 The heat capacity and its derivative for a 80/20 blend of VA/BA.

VA seed used in manufacturing the emulsion. The results from the 50/50 sample are shown in Figure 6. Here four transitions are evident, identified as follows: the VA/BA copolymer (ca. −10 °C), the VA seed (ca. 40 °C), a small amount of BA homopolymer at about −50 °C and a small transition at about 25 °C, due to the region of interpenetration between the VA seed and the VA/BA copolymer.

A second example that demonstrates the sensitivity of the MTDSC technique involves the homogenization of a latex with time. It is well known that a latex paint continues to cure for a couple of weeks after it is applied, even though it is dry to the touch after a few hours. Coatings manufacturers can characterize this slow ageing with MTDSC. Figure 7 shows the total heat flow traces for a latex sample at 4, 11 and 23 days after application. Due to the kinetic activity, the T_g is obscured. However, the derivative of the heat capacity traces, shown in Figure 8, is free from the kinetic activity that confounds the total heat flow. From the gradual melding of the peaks with ageing time it can be imagined that the coating is becoming more homogeneous with time.

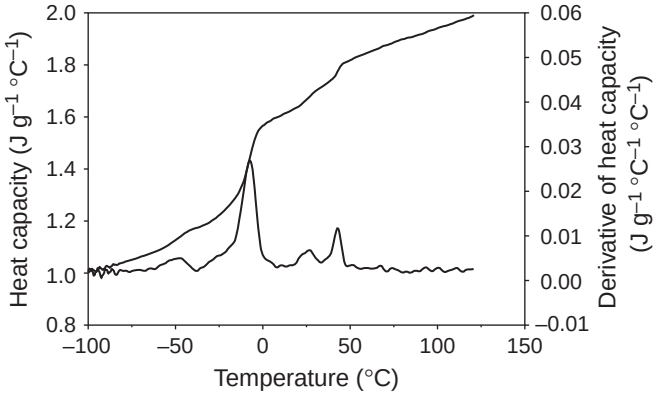

Figure 6 The heat capacity and its derivative for a 50/50 blend of VA/BA.

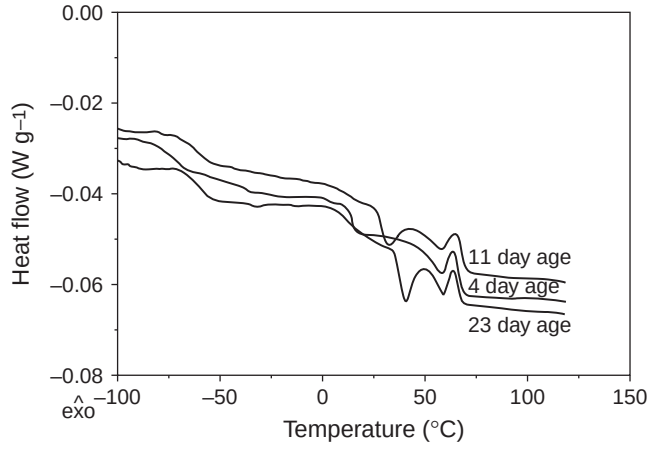

Figure 7 The total heat flow curves for a latex taken at three different ageing times.

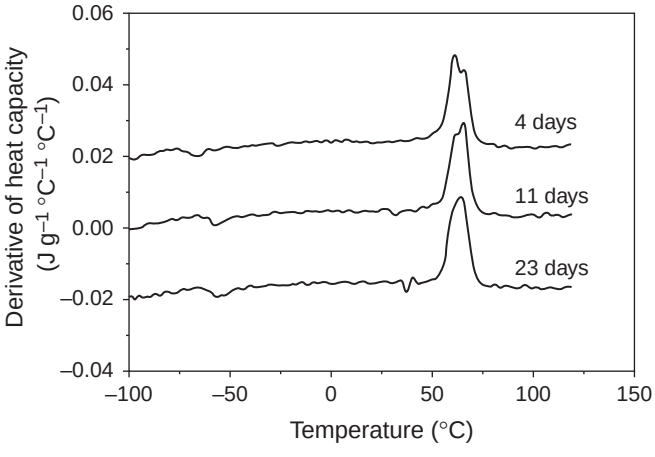

Figure 8 The derivative of the heat capacity curves for the runs shown in Figure 7.

Another example that demonstrates the precision of the T_g determination using MTDSC involves the

determination of the relative concentration of each monomer in a copolymer. An unknown sample was suspected to be a copolymer of only two monomers, monomer A and monomer B. The first experiment performed was to determine the T_g of the unknown. Next tested were three concentration ratios that were polymerized from monomer blend ratios of 100/0, 50/50, and 0/100. From a plot of the T_g versus concentration ratio using only these three points, the estimated ratio of the unknown was about 84/16. To check this, three samples were prepared with concentration ratios of 80/20, 85/15, and 90/10. These results confirmed the original estimate of 84/16 as the concentration ratio of the unknown. A coplot of all the MTDSC results is shown in Figure 9 and a plot of T_g versus concentration ratio is shown in Figure 10. The curve in Figure 10 is that calculated from the Fox equation and the horizontal line represents the T_g of the polymer with the unknown monomer blend ratio.

Whereas most coatings sold to consumers are air-dried, most commercially applied coatings are baked. The optimal bake for a given coating is often bounded on both sides. An underbake can result in a weak, soft, and permeable coating, whereas an overbake can result in a brittle coating that has a tendency to crack. Avoiding an overbake is especially important when the coating used in a precoat, postforming application. In such a process the flat metal stock is coated and baked. Following this, the pieces are stamped and formed from the coated metal. Many food cans, such as those used for soup, tuna, vegetable, and fruit are manufactured using the precoat postform process. The T_g of a coating can be a strong function of bake temperature and thus an indicator of the bake a coating received. The following example involves a corrosion problem with an interior food coating

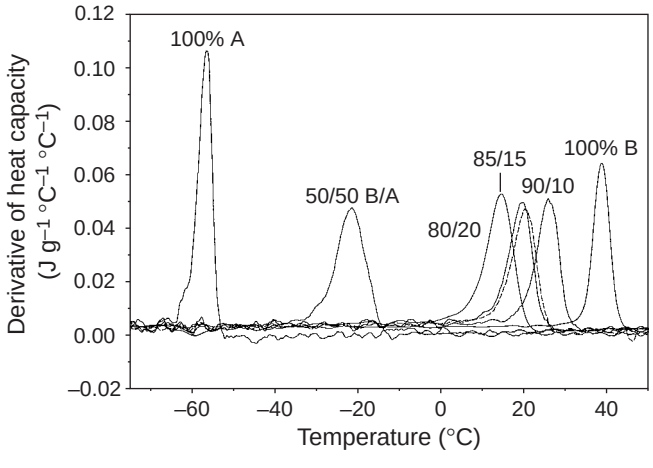

Figure 9 The derivative of the heat capacity curves for a series of experiments designed to determine the blend ratio of the unknown, which is shown as a dotted line in the plot.

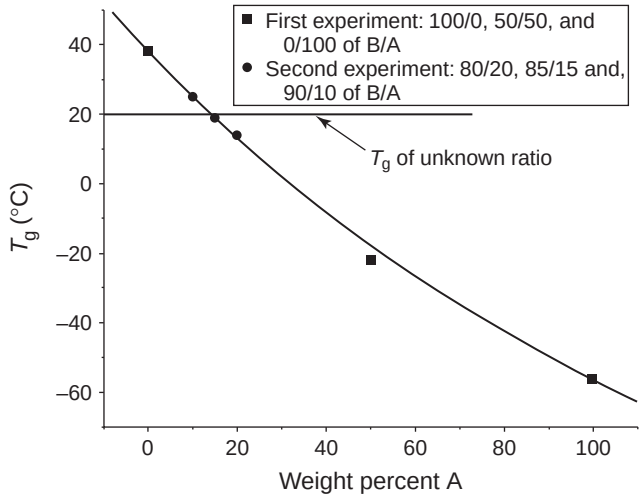

Figure 10 Determination of the polymer blend ratio with MTDSC. A plot of the T_g versus blend ratio of the controls; the horizontal line represents the T_g of the unknown.

at a customer's plant. To investigate this complaint, bake extremes were prepared in the laboratory and compared with the coating taken directly from a problem can. The temperature derivative of the heat capacity results is shown in Figure 11. Clearly, the T_g is a function of bake and thus can be used as the T_g of a coating baked at a customer's plant as an indicator of the bake it received at the customer's plant.

The degree to which the T_g of a coating is a function of the bake depends upon the nature of the coating. The T_g of a thermoplastic coating increases with bake due to the loss of plasticizing agents, such as water, whereas the increase in the T_g of a thermoset coating during a bake is due mainly to an exothermic cross-linking reaction which transforms a low-molecular-weight liquid into an infinite network. The rate of the cure reaction will be

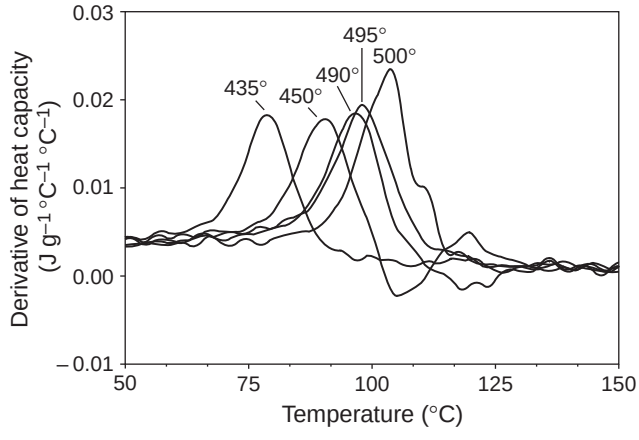

Figure 11 A coplot of the derivative of the heat capacity for five control bakes of a particular coating.

limited by the reactivity of the components when its temperature is below its T_g, and limited by diffusion when it has vitrified. However, as the T_g is changing due to cross-linking, a thermoset system may – if the conditions are right – vitrify, devitrify, and then revitrify. That is, a chemical reaction causes the T_g to exceed the temperature of the coating and the coating vitrifies, thus drastically altering the rate of reaction. Then the oven brings the coating above its current T_g, where the reaction rate will again be governed by the reactivity ratio of the components. Finally, the cure proceeds to a point where the T_g again exceeds its temperature and the coating revitrifies. To optimize the efficiency of a thermoset cure it is helpful to know the T_g of the coating as a function of the degree of cure and the reaction rate of the components as a function of temperature. With such information the plant engineers can set the temperatures at various zones in the oven in a manner that minimizes the time the coating is in the oven and thus maximizes the efficiency of the oven.

To obtain these parameters requires access to the T_g as a function of time during the cure reaction. This makes it an ideal candidate for interrogation with MTDSC which can separate the heat capacity signal from the exotherm due to the cure reaction.

Indeed, MTDSC has been used to fully characterize the bake of a thermoset coating.[1,2] Note that this is virtually impossible with LTDSC because the relatively small T_g transition is difficult to deconvolute from the larger exotherm.

Another application of DSC to coatings involves the detection of additives – specifically waxes. These lubricants are added to industrial coatings for a variety of reasons. They provide slip during the metal-forming process in precoat postform applications, they provide smooth transport through the trackwork that transports the work throughout the plant and they act as an interior release agent in food cans – so the food slides out of the can easily. There are a tremendous variety of commercially available lubricants which range from natural carnauba wax to Teflon®. The identification of a wax is accomplished not through its T_g but through its melt temperature. Because of this it is not necessary to run MTDSC – a linear temperature rate can be used. One application where DSC can be useful is the identification of the wax a competitor is using in their coatings. Another application involves the identification of a substance that has built up on the metal tooling at a customer's plant. Is it the wax? Is it the polymer? Or is it not the coating at all? Removing a few milligrams from the tool and running it in a DSC will usually provide the answer.

The final example involves a baseline shift which is difficult, if not impossible, to uncover using LTDSC. The total heat flow for an MTDSC run is shown in Figure 12.

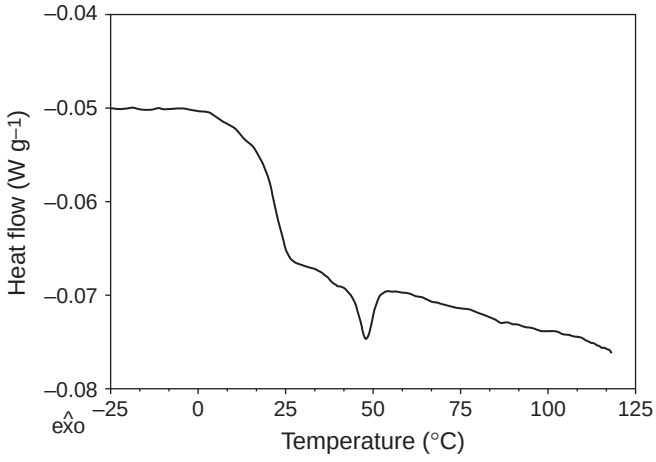

Figure 12 The total heat flow trace for a test coating.

It appears that there is a T_g at about 20 °C. However, the event at 20 °C had nothing to do with the polymer. For some reason the baseline shifted at about room temperature. That is, the symmetry of the cell changed at about room temperature. Regardless of the reason for this shift, a downward shift in baseline appears T_g-like when looking at the total heat flow. However, with MTDSC you will not be fooled because the amplitude of the modulation will not change during a baseline shift. Figure 13 shows the output signal from the cell versus the temperature of the sample side (averaged over a modulation) and its amplitude, which is simply the heat capacity. Here it is clear that the event at about room temperature is not a T_g, but rather a baseline shift because the amplitude of the output signal does not undergo a transition. The true T_g is at about 45 °C where the amplitude of the output signal increases significantly, indicating that the sample's heat capacity has increased. That is, the amplitude of the thermal energy modulation

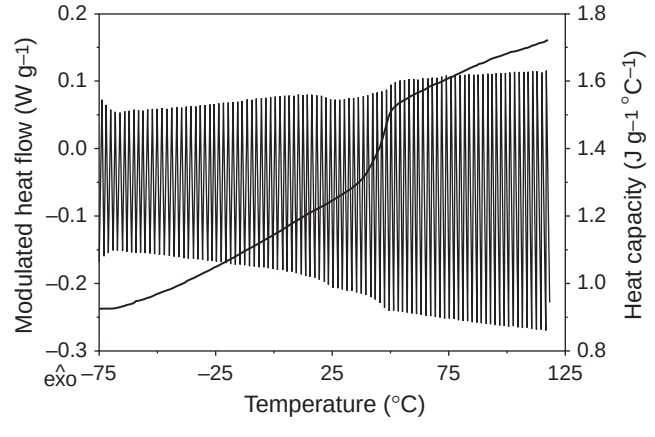

Figure 13 The modulated heat flow and heat capacity for a test coating from the same run shown in Figure 12.

of the furnace needed to produce a 1 °C modulation in the sample side temperature has increased.

2.6 Method Development

In MTDSC the temperature scan rate or linear underlying temperature rate can be set to zero. The technique of modulating around a fixed temperature and monitoring the output signal as a function of time has been called quasi-isothermal MTDSC. This technique can be used to elucidate the nature of crystallizations and glass transitions.[3,4] Typical experimental conditions for an isothermal step are an amplitude of a 0.1 K and periods ranging from 30 to 100 s. The entire quasi-isothermal experiment will consist of a series of isothermal steps across the transition of interest that are perhaps 0.1–0.2 K apart. Typically, each isothermal experiment will run for 20 min with data accumulation over the last 10 min. In contrast with a continuous temperature scan which results in a continuous line representing heat capacity versus temperature, the results of a quasi-isothermal MTDSC experiment consist of a series of points each representing the heat capacity at (or around) a particular temperature. The advantage of a series of quasi-isothermal steps is that a steady state is achieved by simply waiting and there is no need to deconvolute the effect of the underlying linear temperature rate from the effect of the modulation because the underlying linear heating rate is zero.

Another novel development in MTDSC is the light-heating technique, developed by Saruyama.[5] This technique allows higher modulation frequencies to be accessed by employing a light beam directed on the sample, which is attenuated by a polarizer rotating at a constant frequency. The intensity of the beam determines the modulation amplitude, and the angular frequency of the rotating polarizer determines the period of the modulation. The underlying heating rate is supplied by a conventional heating source, allowing access to high frequencies of modulation.

Still another, and perhaps the most technically impressive, development in this fertile field is the marriage of DSC and AFM into a technique called microthermal calorimetry. In this new technique the AFM tip is fitted with an ultraminiature temperature probe that acts as both the heat source and the detector. Obvious applications include the characterization of micromorphology, and detecting and identifying contaminants. The accessible frequencies with this technique are much higher than with normal MTDSC simply because the mass whose temperature is being modulated is much smaller. Indeed, varying the frequency of the modulation is a means of depth profiling from the surface.

3 DYNAMIC MECHANICAL ANALYSIS

3.1 History

The development of modern DMA began with the torsion pendulum. This instrument consists of a clamping mechanism that allows the sample to be deformed in the torsion mode (twisted). When the sample is released from the twisted state it recovers with a damped oscillation around its equilibrium position. The frequency of the oscillations can be related to the shear modulus of the sample and the damping of the oscillations (rate of decrease in amplitude) can be related to the loss modulus of the sample.[6]

A torsion pendulum was used to study the mechanical properties of free films as early as 1966.[7,8] The development of DMA has progressed significantly from the custom-made torsion pendula of the 1960s. Commercial instruments of the late 1990s have numerous deformation modes, precise temperature control, and a wide range of deformation frequencies. Along with the sophisticated hardware comes software with the ability to analyze, fit, and smooth the data.

3.2 Dynamic Mechanical Analysis Fundamentals

In general, DMA is a technique that is used to characterize the response of a material to a periodic perturbation as a function of frequency, temperature, or time. This periodic perturbation can consist of a constant oscillation amplitude at a fixed frequency or a free oscillation at the resonant frequency. The response includes both the elastic component, which is in phase with the perturbation, and a viscous component, which is out of phase with the perturbation.

To obtain quantitative results for the absolute value of the moduli from a DMA test, the sample must be uniform and its geometry well known. For this reason the DMA testing of coatings is usually qualitative because the two most widely used sample configurations – coated braid and free film – preclude precise knowledge of the sample geometry, the sample uniformity, or both. However, in this case qualitative does not imply inadequate. In many cases DMA testing of coatings involves a rank ordering of coatings rather than a determination of absolute moduli.

A couple of advantages that the DMA technique has over other thermal analysis techniques are the ability to detect lower-order transitions and the ability to simultaneously determine the viscous and elastic moduli of the material. The lower-order transitions, which have been attributed to side chain motions of the polymer, are related to the ability of the polymer to respond to fast deformations without exhibiting catastrophic failure. A polymer coating may be subject to such deformations in a precoat postform process, when

exposed to vibrations, or when subject to an impact from, say, a rock. The DMA is a useful instrument to employ when developing, for example, sound-absorbing coatings, impact resistant coatings, and coatings for a flexible substrate. Any exterior coating falls into the latter category due to the thermal expansion and contraction of the substrate with temperature and – in the case of wood substrates – relative humidity.

3.3 Sample Preparation

Basically there are two sample configurations used to test coatings in a DMA. One is a free film and the other is a supported coating. There are advantages and disadvantages to each method. A free film usually has a measurable thickness and the preparation of the film can usually be controlled to a greater degree. The drawbacks of using a free film are numerous. First, in most cases it is quite difficult, to obtain a free film that is representative of the film as prepared on the substrate. Removal of the film from the substrate often irreversibly stretches or damages it, whereas chemical removal of the substrate from the coating (for example, with acid) raises questions of the effect of the chemical on the coating. If a representative free film can be obtained, mounting it in the DMA presents a significant experimental challenge. If, somehow, a free film is removed from its substrate without changing it and is successfully mounted in the DMA, there still remains the question of film thickness. The percentage variation in thickness of a thin free film (1 cm by 4 cm by 0.0025 cm) will usually be greater than the percentage variation in a rectangular slab of a polymer (1 cm by 4 cm by 0.25 cm) prepared in a mold. This is due to incomplete leveling of the thin film. If a film is not uniform, the modulus calculated by the software from the force data will be lower than the true modulus of the polymer because the film was not deformed uniformly.

The second option is to provide a support for the film. The two basic methods of supporting a coating are with a flat, solid substrate and with a mesh or absorbent fiber. An example of the former type is a coating applied to a metal panel and an example of the second type is a coating-impregnated fiberglass mesh. A flat substrate still provides some information about the geometry of the film, and semiquantitative results can be obtained if the response from the substrate can be deconvoluted from the response from the coating.

A mesh can be used when the film or coating is too weak to support itself. This is the case when studying the cure or drying of a coating. Just prior to testing, a known amount of liquid coating (typically about 25 μL) is spread evenly across the mesh. The surface tension of the coating is sufficient to keep the liquid from dripping off the braid. During the cure or drying of the film the elastic modulus of the coating will increase, and this increase will be evident in the results.

3.4 Experimental Conditions

The difficulty in choosing the optimal experimental conditions has increased with the complexity of commercially available instruments. First, the operator has to choose the mode of deformation. The most common mode is the two-point bend whereby one end of the sample is held fixed and the other end is oscillated sinusoidally. Other modes include the three-point bend, shear, torsion, compression, and tension. After the mode of deformation is chosen, the operator must select the experimental conditions. These include the sample length, the amplitude of the deformation, the frequency or frequencies of the deformation, and the temperature protocol. If the sample has a relatively low modulus a shorter sample should be used, and if the sample has a high modulus a longer sample should be used. The amplitude of the deformation should be large enough to create a useful restoring force but not so large that it exceeds the linear response region. Consult the instrument manual for guidelines on setting experimental parameters for a given sample stiffness.

Ideally, a range of frequencies should be chosen. This will provide information regarding the relationship between both the elastic and loss moduli and the frequency of deformation. The temperature program may consist of a constant scan rate, a constant temperature or a combination of both. For example, the temperature program of a cure experiment may consist of a jump from room temperature to 400 °F, a 10 min isotherm at 400 °F, a jump to −100 °F, and finally a scan at 2 °C min^{-1} from −100 °F to 400 °F.

3.5 Applications

The application of DMA to coatings can be divided into two main groups – consumer and industrial. Consumer coatings consist of paints, lacquers, etc. that are applied by the consumer and dried at ambient temperature. Most industrial coatings are subjected to a bake which accelerates drying and, in the case of a thermoset, promotes a chemical reaction. DMA can be used either to characterize the final cured or dried film or to characterize the transition from a wet coating to a final film.

The first example involves the ranking of exterior paints for low-temperature film formation. That is, whether the films will dry completely at low temperatures. The temperature used in this study is a stretch target for exterior house paint application, namely 30 °F. Three paints were studied. The experimental protocol consisted of applying 25 μL of coating to a fiberglass braid and holding its temperature at 30 °F for 700 min. Following the 700 min isotherm, the temperature of the sample was

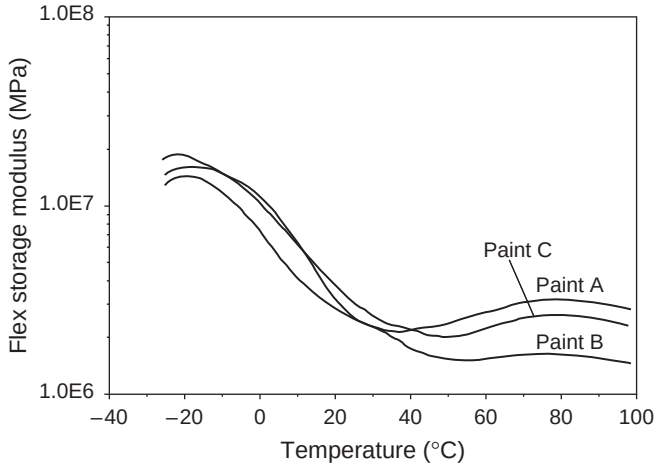

Figure 14 A coplot of the flexural storage modulus for test paints A, B, and C after drying at 30 °F.

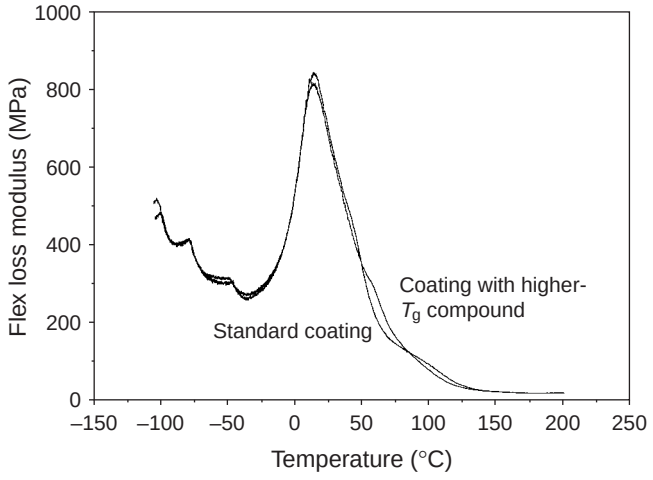

Figure 15 Coplot of the flexural loss modulus of a standard coating and an experimental coating with a higher-T_g component.

scanned from −25 °C to 100 °C. The storage modulus during these temperature scans is coplotted in Figure 14. If the paint is completely dry, the modulus should not increase with temperature. Any increase in elastic modulus with temperature is evidence of an incomplete dry. From the results of the scan we can conclude that the paint B has the best low-temperature film formation, paint C is second best, and paint A has the most incomplete film formation at low temperature as its elastic modulus increased significantly during the temperature scan.

The second example involves the determination of whether a higher-T_g additive would adversely affect the flexibility of an exterior house paint. An exterior house paint must be soft enough to expand and contract with the outside of the house as the exterior temperature changes. However, it also must be hard enough to resist dirt pick-up when the wind blows dust and dirt against the house. Paint formulators often combine hard and soft polymers in the formula in an attempt to produce a paint with both sufficient flexibility and low dirt pick-up. In this example, the formulators were considering replacing one of the components with a higher-T_g polymer, but were concerned that this would compromise the excellent flexibility of the product. To investigate this, the standard coating and the coating with the higher-T_g component were dried on a fiberglass braid under identical conditions. A coplot of the flexural loss modulus, as determined during a temperature scan at a rate of 2 °C min^{-1}, is shown in Figure 15. It can be seen that the low temperature loss modulus is virtually unaffected by the addition of the higher-T_g component. Only at about 50 °C does the difference between the coatings become evident. Clearly, the higher-T_g component can be used in the formula without sacrificing the ability of the coating to respond to the contraction and expansion of the house with temperature.

3.6 Dynamic Mechanical Analysis Method Development

Recent developments of DMA have involved expanding the range of applied forces, increasing the range of allowable frequencies, increasing the temperature range over which experiments can be performed, and increasing the number of modes of deformation. That is, much of the recent development has involved engineering rather than science. Current instruments have the ability to oscillate samples at 200 Hz using forces of up to 18 N with a temperature range from −150 °C to 600 °C. With modern instrumentation, all of this can be accomplished with a strain resolution of about 1 nm.

With these engineering developments comes an increase in the capability and speed of the software used both to control the instrument and to analyze the data. Temperature programs can consist of rapid coolings or heatings and they can contain numerous temperature excursions as well as isotherms. Software permits extensive coplotting as well as advanced data handling and curve-fitting procedures.

4 THERMOMECHANICAL ANALYSIS

4.1 History

The precursor to the modern thermomechanical analyzer is simply a dilatometer that measures a sample's dimensions as a function of temperature. In this way the sample's

linear or volume expansion coefficient can be determined. With the widespread use of polymers in engineering applications, the need to monitor the sample's change in dimension as a function of time under a load at constant temperature (creep) became apparent. Further, the need to determine a material's response to either a tensile or compressive force as a function of temperature and time became necessary. These requirements were the driving force behind the development and commercialization of today's modern thermomechanical analyzers.

4.2 Thermomechanical Analysis Fundamentals

Conventionally, TMA involves the measurement of a linear dimension of a material as a function of temperature or time when the material is subject to a stress (either compressive or tensile). TMA is similar to DMA in that the instrument is monitoring the material's response to an applied force. However, the instrument-naming convention is such that instruments that perturb the sample with a force of constant magnitude are called TMAs, and instruments that perturb the sample with a force that varies sinusoidally are called DMAs. A typical TMA sample cell set-up to measure a sample under load is shown in Figure 16. The sample under investigation rests on a stage and a probe is brought down on top of the sample with a predetermined force. The measured parameter is the distance between the tip of the probe and the stage upon which the sample rests. If the sample expands, as it might when heated, the measured distance increases; if the sample contracts or softens, the measured distance will decrease. A modern instrument comes with a variety of probe tips. Sharp or hemispherical probes are used to measure softening points and large flat probes are used to measure coefficients of linear expansion. In the coatings laboratory, TMA is used primarily to determine softening points of coatings.

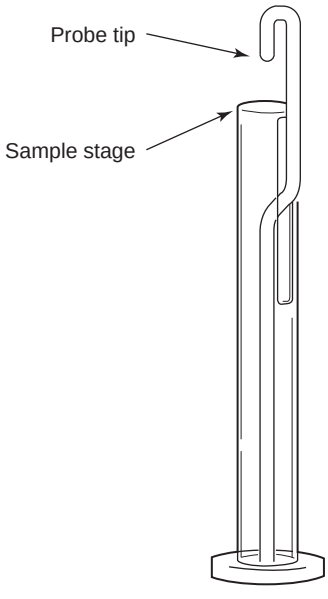

Figure 16 Diagram of a TMA cell.

4.3 Sample Preparation

When attempting to measure the softening point of a coating there is no compelling reason to remove the coating from the substrate. Therefore, coatings are usually tested in situ. Of course, it is helpful if the substrate is flat and will lie flat on the instrument's stage. Frequently, the only sample preparation required to analyze a coating is to cut a piece of the coated substrate that is small enough to fit on the stage (typically about 1 cm across).

4.4 Experimental Conditions

The conditions chosen by the user are probe type, the force exerted on the sample by the probe, and the temperature program. For a penetration experiment, the probe tip should be relatively sharp, such as a hemisphere of small radius. The applied force should be sufficient to eliminate the noise in the distance measurement, but not so great as to cause premature damage to the coating. A typical value for the applied force when a hemispherical probe tip is used is 0.25 N, which is the force of gravity on a mass of about 25 g. Frequently, the temperature program consists of a simple linear scan through the temperature regime of interest. A typical scan rate is $5\,^{\circ}\mathrm{C\,min^{-1}}$.

4.5 Applications

A wet decorative coating may be dry to the touch after only a few hours but it continues to develop physical properties for a couple of weeks after it is applied. To quantify this, a series of experiments were performed on a test coating after a series of ageing times.

The results, coplotted in Figure 17, show that the probe starts to penetrate significantly into the coating at roughly 50 °C. At 7 days drying, the penetration is deep, being about 15 μm. After 14 days of drying, the maximum penetration depth is less than at 7 days, and after 28 days of drying the penetration depth into the coating is less than 5 μm. From these results we can conclude that the coating is still building integrity after a month of drying. By adding specifically designed ingredients the cure of the coating can be accelerated and the effect quantified with TMA.

4.6 Thermomechanical Analysis Method Development

A recent improvement in the software that controls the TMA instrument provides an option for the linear temperature scan through the regime of interest. This

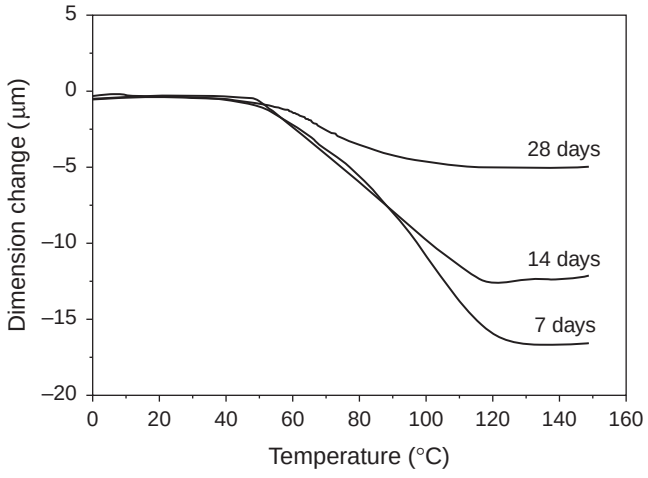

Figure 17 The dimension change versus temperature for a test coating at three drying times.

MaxRes[9] technique will lower the temperature rate in direct proportion with the rate at which the distance between the probe tip and the sample stage is changing. That is, the temperature rate will be decreased during softening and increased when the sample dimension is constant. Thus, the temperature rate through the transitions is relatively small and the temperature rate through steady-state regions is relatively large. This decreases the total time to run a series of experiments, while increasing the sensitivity in the temperature regimes of importance.

A significant advance in TMA is the capability to apply a dynamic load to the sample. That is, a force that changes, linearly or sinusoidally, with time. Dynamic load TMA will provide the user with a tool to rapidly map the softening rate profile of a coating for any temperature and applied pressure.

Another recent advancement in TMA involves the monitoring of the temperature of the sample in a manner that allows the characterization of thermal events that occur within the sample. That is, when the temperature of the sample deviates from the steady state, a thermal event is occurring within the sample. With such an instrument, simultaneous measurements of the displacement trace and the temperature difference trace can be plotted against temperature. Thus, changes in the modulus of the sample can be related to thermal events within the sample.

5 THERMOGRAVIMETRIC ANALYSIS

5.1 History

The concept that the mass of an object will change upon heating goes back at least 2000 years to when the Egyptians heated materials on a balance over a fire. The term thermobalance was coined by Honda[10] in 1915 in his paper on the degradation of various oxides. The appearance of commercial thermobalances occurred in 1963.

5.2 Thermogravimetric Analysis Fundamentals

The TGA technique accurately measures the change in mass of a material as a function of temperature. The change in mass can be due to a number of phenomena, such as degradation, evaporation, oxidation, absorption, or desorption. Modern electrobalances used in quality TGAs are sensitive to a change in mass of 100 ng.

5.3 Sample Preparation

For most TGA experiments the sample preparation is very simple. All that is required is to isolate a representative piece of the material to be tested. Sample masses are typically tens of milligrams. Typically, the sample is placed in the center of a pan which is hanging from a microbalance. The pans are usually aluminum or platinum, the latter being required for high-temperature applications.

5.4 Experimental Conditions

There are a number of factors to be chosen to run a simple TGA experiment. First, there is the temperature rate and the range over which the temperature is scanned. The linear scan rate should be small enough to accurately assign a temperature or temperature range to the phenomenon of interest. If the sample has a low surface-to-volume ratio (as for a large single piece), the temperature rate should be decreased when attempting to characterize a bulk phenomenon. However, if the sample has a high surface-to-volume ratio (as with a fine powder), higher temperature rates can be used. Often, an unknown sample is run first with a high temperature rate for the purpose of discovering the temperature (or temperatures) at which the mass changes. Once these are known the temperature can be scanned slowly through the regions of interest. Of course, if the instrument is more than a simple TGA, there are other experimental parameters that must be chosen.

5.5 Applications

In coatings development, especially in coatings for food or beverage containers, it is important to know whether the coating will absorb certain gases that may affect the flavor of the contents. In this example, a polymer is tested for selected absorbance of carbon dioxide. Over the first

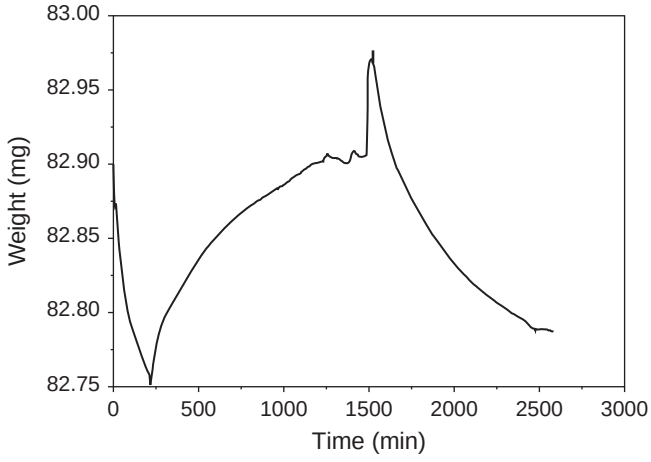

Figure 18 A plot of weight against time for a polymer in an effort to characterize its absorption of carbon dioxide.

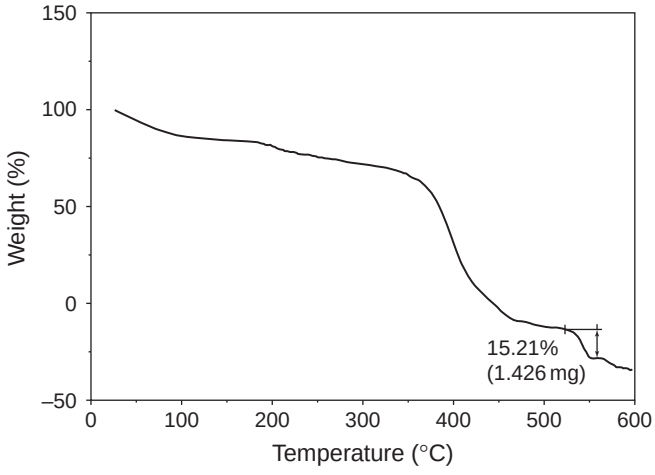

Figure 19 A plot of weight percent versus temperature, which shows a degradation between 530 °C and 550 °C.

part of the experiment the polymer is exposed to high-purity nitrogen gas. After the mass of the sample has settled somewhat, the purge gas is switched to carbon dioxide. Following the apparent decrease in mass due to the buoyancy of the heavier carbon dioxide, there is a slow absorption of carbon dioxide into the polymer. Later in the experiment the purge gas is switched back to nitrogen. Here the desorption of carbon dioxide from the polymer can be monitored. The force on the balance is plotted against time in Figure 18. The jump in weight after the switch from carbon dioxide back to nitrogen is due to the greater density of the carbon dioxide. Various coatings and components thereof can be tested to qualitatively rank their relative permeability to carbon dioxide and the solubility of carbon dioxide within them.

A second example involves the detection of a contaminant in a coating. The contaminant is known to have a degradation temperature of about 530 °C, which is fortunately higher than the rest of the coating. The contaminant was present at a concentration of only a few parts per million. At this concentration its presence would be virtually impossible to detect by TGA. However, the liquid coating was placed in a centrifuge which increased the level of contaminant in the solids (if it was present) by a factor of roughly 1000. The concentrated sample was placed in the TGA and the temperature was scanned from room temperature to 600 °C at a rate of 10 °C per minute. The results are shown in Figure 19. The presence of the contaminant is revealed by the mass loss between 500 °C and 550 °C. The amount present in the original sample can be approximated by dividing the concentration of contaminant in the sample tested by the factor by which the sample was concentrated with the help of the centrifuge.

5.6 Method Development

There are many possibilities for the combination of a TGA with other analytical instruments. The addition of some thermocouples to a TGA provides simultaneous TGA and differential thermal analysis (DTA). That is, a differential thermal analyzer as well as a TGA. With this set-up the user is better able to characterize various weight losses. For example, even though both evaporation and degradation result in a loss of sample mass, an evaporation is endothermic whereas a degradation is exothermic. The difference in the character of the mass loss will be evident with the DTA.

By adding a Fourier transform infrared or a mass spectrometer to the purge gas output from the sample cell, the gas coming off the sample can be analyzed for chemical composition. This is called EVA (evolved gas analysis).

Another advance in basic TGA is controlled-rate thermogravimetric analysis (CRTGA). With this method the temperature rate is controlled in such a way as to make the mass loss constant. In regions where the sample is virtually steady state, the temperature rate will be high and in regions of degradation or evaporation the temperature rate will be slowed considerably. This has the effect of maximizing the sensitivity of the weight loss in regions of interest.

The final development in basic TGA is the addition of a modulated temperature signal to the linear temperature rate – MTGA. By subjecting the sample to a temperature modulation, the rate of weight loss is measured at a wide range of temperature rates rather than the single temperature rate used in ordinary TGA. This allows a more precise calculation of activation energies and pre-exponential factors.

6 SUMMARY

A well-equipped thermal analysis laboratory has the potential to provide significant guidance to a coatings research and development team. However, to fully capitalize on the capability of the equipment, the users must possess the ability to skillfully prepare the samples, efficiently design an experiment, and properly interpret the results.

ABBREVIATIONS AND ACRONYMS

AFM	Atomic Force Microscopy
BA	Butyl Acrylate
CRTGA	Controlled Rate Thermogravimetric Analysis
DEA	Dielectric Analysis
DMA	Dynamic Mechanical Analysis
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
EVA	Evolved Gas Analysis
LTDSC	Linear Temperature Differential Scanning Calorimetry
MTDSC	Modulated Temperature Differential Scanning Calorimetry
MTGA	Modulated Thermogravimetric Analysis
T_g	Glass Transition Temperature
TGA	Thermogravimetric Analysis
TMA	Thermomechanical Analysis
VA	Vinyl Acetate

RELATED ARTICLES

Coatings **(Volume 2)**
Mechanical Properties of Solid Coatings • Rheology in Coatings, Principles and Methods

Polymers and Rubbers **(Volume 9)**
Dynamic Mechanical Analysis of Polymers and Rubbers • Mechanical Properties of Polymers and Rubbers

Thermal Analysis **(Volume 15)**
Differential Scanning Calorimetry and Differential Thermal Analysis • Inorganic Systems, Thermal Analysis Applications to • Simultaneous Techniques in Thermal Analysis • Thermogravimetry

REFERENCES

1. G. Van Assche, A. Van Hemelrijck, H. Rahier, B. Van Mele, 'Modulated Differential Scanning Calorimetry: Isothermal Cure and Vitrification of Thermosetting Systems', *Thermochim. Acta*, **268**, 121–142 (1995).
2. G. Van Assche, A. Van Hemelrijck, H. Rahier, B. Van Mele, 'Modulated Differential Scanning Calorimetry: Non-isothermal Cure, and Vitrification and De-vitrification of Thermosetting Systems', *Thermochim. Acta*, **286**(2), 209–224 (1996).
3. K. Ishikiriyama, M. Pyda, G. Zheng, T. Forschner, J. Grebowicz, B. Wunderlich, 'Heat Capacity of Poly-*p*-dioxanone', *J. Macromol. Sci., Phys.*, **B37**(1), 27–44 (1998).
4. B. Wunderlich, 'The Heat Capacity of Polymers', *Thermochim. Acta*, **300**, 43–65 (1997).
5. Y. Saruyama, 'Development of Light Heating Dynamic DSCII: Determination of Complex Calibration Constants', *Thermochim. Acta*, **283**, 157–163 (1996).
6. L.E. Nielsen, 'Mechanical Properties of Polymers', Reinhold Publishing Corp., New York, 1962.
7. P.E. Pierce, R.M. Holsworth, 'The Mechanical Properties of Coatings – The Measurement of Torsion Modulus and Logarithmic Decrement of Free Films', *J. Paint Technology*, **38**(496), 263–268 (1966).
8. P.E. Pierce, R.M. Holsworth, 'The Mechanical Properties and Performance of Wood Primers', *J. Paint Technology*, **38**(501), 263–268 (1966).
9. B. Riesen, B. Schenker, 'MaxRes in TMA and TGA: A New Tool for Fast Maximum Resolution By Rate Adjustment', Proceedings of the 26th Conference of the North American Thermal Analysis Society, 116–121, 1998.
10. K. Honda, 'A Thermal Balance', *Sci. Rep. Tohoku Univ. Ser. 2*, **4**, 97–105, 1915.

Environmental Trace Species Monitoring: Introduction

Markus W. Sigrist
ETH Zurich, Zurich, Switzerland

Today, air pollution is an issue of great concern worldwide. This is manifested by tremendous scientific and technological effort in various fields as well as by the activities of many international organizations such as the Intergovernmental Panel on Climatic Change (IPCC) and the United Nations Conference on Environment and Development (UNCED). The first conference held in Rio de Janeiro, Brazil, was succeeded by various conventions such as the Framework Convention on Climatic Change (UNFCCC) that resulted in the Kyoto protocol in 1997. Finally, the Nobel Prize in Chemistry in 1995 was awarded for the first time to scientists researching in atmospheric chemistry (Paul Crutzen, Mario Molina and Sherwood Roland).

The main components of the air we breathe are nitrogen, oxygen, argon, carbon dioxide and water vapor which account for 99.9998% of its density. It is recognized, however, that the remaining trace concentrations of atmospheric species have a great impact in diverse areas. One cubic meter of air contains 1.1 mg of methane, 60–250 μg of carbon monoxide, 50 μg of hydrogen, 30–200 μg of ozone, 0.6 mg of nitrous oxide, 8–170 μg of nitric oxides, 0.2–100 μg of sulfur dioxide and 1–100 μg of particulate matter.[1] The given values obviously vary with location and altitude. Both trace gases and aerosols contribute to well-known phenomena such as global climatic change, photochemical smog formation, acid rain, stratospheric ozone depletion, and forest decline (Figure 1). Ever increasing concentrations of numerous trace species during the 20th century, arising from increasing population, industrialization, road and air traffic, have contributed to the negative consequences that have become a matter of public concern. Another aspect is the danger of catastrophic emissions of toxic substances into the environment, most dramatically manifested by recent past disasters, such as in Seveso (Italy) in 1976, in Bhopal (India) in 1984, in Tschernobyl (Ukraine) in 1986, or in Schweizerhalle (Switzerland) in 1986.

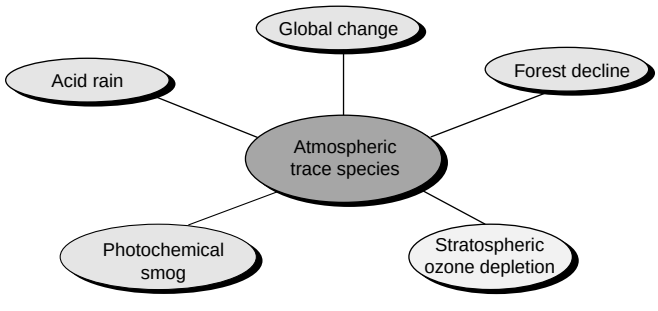

Figure 1 Illustration of the significance of atmospheric trace species for environmental issues.

The role of atmospheric aerosols which contribute only about 10^{-9}–10^{-7} to the mass of the air is manifold.[1] An aerosol represents a two-to-three phase system consisting of liquid and/or solid particles and a gaseous medium in which the particles are suspended. Aerosols provide reaction sites for pollutant gases, influence cloud formation and modify precipitation by functioning as condensation nuclei, and act as carriers for pollutant transport. The aerosol particles have various origins including both natural (e.g. volcanoes) and man-made processes. Globally, the natural aerosols dominate but in urban areas the anthropogenic aerosols are more important. As the tropospheric lifetime is about 1 week, atmospheric aerosols are widely distributed and they can have a large spatial and temporal variability.

Today, there are regulations in many countries to govern the maximum concentrations of selected gaseous species at the workplace, emission limits for sources of pollution, or concentration standards for the open atmosphere. Furthermore, there exist lists of hazardous air pollutants (HAPs) like the one issued by the US Environmental Protection Agency (EPA). Ambient air quality standards for fine airborne particles, such as those emitted by Diesel engines, have been introduced according to their size and concentration. As example, the PM_{10} and $PM_{2.5}$ values for particulate matter include 24-hour and annual average standards on the mass concentration for particles with aerodynamic diameters below 10 μm and below 2.5 μm, respectively. These standards are considered as an important measure for health hazards. They are often not met in urban, suburban or even rural areas. A detailed knowledge of source contributions to PM_{10} and $PM_{2.5}$ values is still lacking, despite their significance for the development of reduction strategies.

Both the global impact of trace species and the examples of local hazardous emissions demonstrate the need for versatile analytical techniques and for remediation schemes. Measurements of ambient air compounds include several diverse fields: urban (traffic emissions); industrial (petrochemical industry, combustion sites, waste incinerators, semiconductor manufacturing process control); rural (background concentrations, greenhouses, fruit storage); medical diagnostics (exhaled breath gas monitoring); and spacecraft habitat air quality monitoring.

The large number of trace species, their distribution and diversity pose a great challenge with regard to the development, implementation and application of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

such methods. Different kinds of devices are used for aerosol detection including simple instruments for light transmission measurements, or porous filters to collect material for determining the mass concentration, as well as sophisticated sensors to characterize the particle size distribution and chemical composition.[1,2]

Ideal analytical tools for trace gas sensing have to meet several requirements, such as multicomponent capability, high sensitivity and selectivity, high accuracy and precision, large dynamic range, good temporal resolution, ease of use, versatility, reliability and robustness. However, most techniques cannot simultaneously fulfill all requirements so that one has to compromise between the most important features for a given analytical application or task. In the following the desired properties are briefly discussed individually.

- *Multicomponent capability* This is often important because in most situations the trace species, such as a gas to be detected, is not isolated but contained in a mixture. Hence, it is desirable to know additional compounds present in the mixture that may cause harm to the environment. Of course, this could be achieved by measurements with a variety of different instrumentation but at the expense of time, easiness, cost, etc.
- *Sensitivity* Usually, concentrations are given in units of ppm (10^{-6}, corresponding to micromole per mole), or ppb (10^{-9}, nanomole per mole), or else in milligrams per cubic meter or in micrograms per cubic meter volume of air mass. It is evident that instruments with extremely high sensitivity are required for measurements in rural areas, for instance in the monitoring of background concentrations of trace compounds.
- *Selectivity or specificity* This represents a crucial aspect in cases where an individual species is to be monitored in a mixture of various compounds. This scenario is very often encountered and is thus related to the multicomponent capability. It permits quantitative analysis or at least selective monitoring of specific compounds in mixtures. However, there are also situations where it is sufficient to quantify the concentration of just a group of compounds. An example is the total hydrocarbon concentration monitored in regular exhaust checks in a car garage.
- *Accuracy and precision* These two scientific terms are often confused. Accuracy is a measure on how close the measurement value is to the true value, whereas precision is a measure of the range of values obtained from repeated measurements on samples with identical concentrations. As a result accuracy is an important criterion when testing the observance of regulatory standards or when comparing different measurement schemes, whereas precision is the important feature when recording the trend of a species concentration at the same location. This property is connected to the sensitivity – the closer the measured concentration approaches the detection limit, the smaller the precision of the measurement.
- *Dynamic range* A large dynamic range allows to sense pollutants in rather different environments with one and the same instrumentation. At the source of emission, such as at a gas leak, concentrations can be as high as in the percent range, whereas far away from a source in a rural area the same pollutant can be present in concentrations at ppb or even ppt levels.
- *Temporal resolution* This is important when concentration fluctuations on short timescales of, say, minutes, are to be expected, when the interest lies on flux measurements or on the temporal evolution of species, as after a gas leakage. In these situations on-line and real-time sensing is often required. However, time resolution is of minor concern for cases like the surveillance of the ambient atmospheric CO_2 concentration.
- *Ease of use* This is desirable in several respects, such as easy implementation and preferentially unattended automated operation of the measurement system, preferentially room-temperature operation, easy maintenance and tolerable costs.
- *Versatility, reliability and robustness* These are all important for applications at all various locations outside the laboratory such as in cold, hot, humid, noisy environments in the field, in airplanes, or even in spacecraft or satellites.

Other features not mentioned here can be important for selecting one or another technique for a certain monitoring task (e.g. requirements for sample preparation, size, weight, etc.).

In the past, numerous techniques ranging from chemiluminescence to gas chromatography have been developed and successfully applied to trace gas monitoring.[3] Some of these schemes have become standard methods and are widely used on a routine basis. However, several requirements cannot be fulfilled by conventional techniques. This is particularly true for the number of species that can be detected with a single-measurement device, i.e. the multicomponent capability, and problems related to sample preparation. In recent times, optical techniques have attracted considerable interest.[4] They offer some unique features that are relevant for the monitoring of atmospheric constituents. A key issue is the contact-free measurement of the sample, the lack of any preconcentration, or any kind of pretreatment of the sample. Finally,

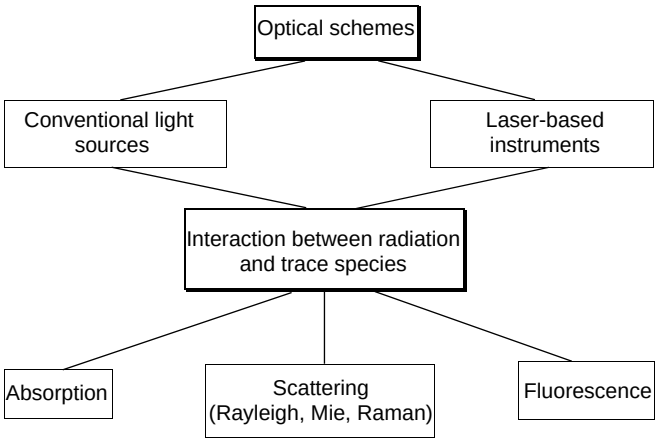

Figure 2 Diagram of light sources and interaction processes used in the optical monitoring of trace species.

as only schemes they offer the possibility of remote three-dimensional sensing in the open atmosphere from the ground or from space.[1,5,6]

Today, there is a great variety of optical schemes. They can be separated into those that use conventional light sources such as Xe lamps and the laser-based devices (Figure 2). The latter essentially rely on lasers that are tunable in appropriate wavelength ranges. The advantage of lasers is their extremely high spectral brightness, which is not achievable with conventional sources combined with filters or gratings. This characteristic laser property forms the basis for a high detection sensitivity and selectivity. However, conventional light sources often offer a larger tuning range which can compensate for the lack in spectral brightness and which is also important when a multicomponent capability is required. Furthermore, broadly tunable lasers are still often complex and expensive devices, which can hinder their wide distribution. However, this situation may change in the future with the development of compact diode laser-based systems.

As Figure 2 implies, the basis of all optical schemes is some kind of interaction process between the radiation and the air mass (gases and aerosols) to be monitored. The essential processes involved are absorption, Rayleigh and Raman scattering, fluorescence, and combinations thereof. Hence, these methods are based on spectroscopic properties of the species under study. Apart from these techniques there are further optical schemes, such as fiberoptic sensing, laser mass spectrometry and laser-induced breakdown spectroscopy (LIBS), that use other (nonspectroscopic) effects for the production of the relevant monitoring signal.

Finally, it should be emphasized that there is no single technique that meets all the requirements mentioned above. However, there is a large variety of instruments available (sometimes with complementary performance) from which one can select a device that can be tailored to the needs of the application.

This section of the *Encyclopedia* describes the various techniques available and their application to the sensing of trace species in different environments. Emphasis is placed on optical schemes and trace gases. The selection of articles has been made on the basis of the current distribution, significance and impact of the method. A broad coverage of application areas has also been taken into account. The techniques selected are either already well established or else represent very promising methods with a great potential in various application fields. Although lasers have had a great impact in diverse areas including environmental sensing, non-laser-based spectroscopic systems also carry great potential and the presently most promising schemes are treated here as well. The applied wavelengths span from the ultraviolet to the infrared (IR) region. New developments can be expected from such techniques mainly in combination with future laser developments, such as quantum cascade lasers, and other technological achievements. The individual articles are reasonably self-contained. They are written by authors who were selected for their expertise in the field and their own original contributions.

A comprehensive presentation is given of the fundamentals and of the state-of-the-art, illustrated by application examples. Although the list of articles is by no means complete, it should still give an impression on optical methods, their performance, and their limitations in trace species monitoring.

The first article, by L. Stefanutti and S. Borrmann, National Research Council, Florence (Italy), is on **Airborne Instrumentation for Aerosol Measurements**. This includes remote sensing instruments, direct reading in situ methods, and sampling devices followed by sophisticated laboratory analyses. The detection limits and ranges, the reliability and areas of applications as well as the different platforms used are addressed.

U. Platt, University of Heidelberg (Germany), presents the article **Differential Optical Absorption Spectroscopy, Air Monitoring by**. After a discussion of the fundamentals of molecular scattering and absorption, the author outlines the different modes of differential optical absorption spectroscopy (DOAS) operation followed by a description of practical designs and characterization of DOAS systems, evaluation techniques for DOAS spectra and concludes with a discussion of some applications.

M. Tacke et al., Fraunhofer Institut fuer Physikalische Messtechnik, Freiburg (Germany), gives an overview on **Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode**. The lasers employed are cryogenically cooled lead salt diode lasers. The basics of this technique are discussed and various

examples of applications in air monitoring are presented, including ground and airborne measurements, rocket experiments, and exhaust analyses. Finally, a comparison with alternative schemes is given.

F.K. Tittel, Rice University, Houston (USA), and K.P. Petrov, Gemfire Corporation, Palo Alto (USA), discuss a different approach to **Diode Laser Spectroscopic Monitoring of Trace Gases** by employing room-temperature difference frequency mixing in nonlinear optical materials to access the important mid-IR spectral region. After discussing some fundamentals of gas phase spectroscopy in general, they compare the merits of different tunable laser sources followed by a discussion of the available techniques for measuring small gas concentrations. They conclude with a presentation of several examples of trace gas sensors based on this novel diode laser technology.

D. Griffith, University of Wollongong (Australia), describes **Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis**. He includes both open path and closed long path measurements in specially designed gas cells. Remote sensing is presented as ground-based solar absorption Fourier transform infrared (FTIR) spectroscopy. IR emission studies on stack plumes, volcanoes and hot sources are examples of applications of this technique.

The potential of **Photoacoustic Spectroscopy in Trace Gas Monitoring** is treated by F.J.M. Harren, University of Nijmegen (The Netherlands). The author gives a history of this effect which was discovered in 1880 by A.G. Bell, followed by a detailed discussion of the devices and equipment used nowadays. Possible limitations, detection limits and selectivity issues are also presented. The performance is illustrated with many examples including stack gas and car exhausts emissions, as well as ambient air monitoring. The wide scope is manifested by biological and medical applications. These include studies on postharvest research, plant physiology, microbiology, entomology, molecular biology, soil science and noninvasive human breath analysis.

Matrix Isolation Spectroscopy in Atmospheric Chemistry as a further quantitative analytical tool is outlined by K.S. Viswanathan, K. Sankaran and K. Sundararajan, Indira Gandhi Centre for Atomic Research, Kalpakkam (India). The authors highlight the advantages of this technique particularly for applications in atmospheric chemistry, such as quantitative estimates of atmospheric trace gases and reactive species. The evaluation of the appropriate matrix, the detection system and the sampling methods are presented as well as the use of the technique for studying reactions of atmospheric species such as NO_x, halomethanes, ozone, and OH radicals.

The LIDAR (light detection and ranging) scheme is presented in two contributions differentiating between ultraviolet-to-visible (UV/VIS) LIDAR and IR LIDAR. Both chapters demonstrate how lasers can be employed to detect trace species at a distance. The **Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring** article, written by J.P. Wolf, University of Lyon (France) describes the DIAL (differential absorption LIDAR) principle and gives examples of DIAL systems. On the one hand, the mapping of gaseous pollutants is discussed for traffic-induced NO_x distribution in cities, chemical dynamics of ozone production, industrial SO_2 plume studies, and diffuse emissions of VOCs (volatile organic compounds). On the other hand, LIDAR systems are employed for the measurement of urban aerosols, the detection of soot, or the multispectral monitoring of stratospheric aerosols active in the ozone layer depletion. The article also includes the validation of numerical models, comparisons with other techniques, and a discussion on new standards.

The **Infrared LIDAR Applications in Atmospheric Monitoring** is addressed by B.J. Orr, Macquarie University, Sydney (Australia) and starts with a survey on optical and atmospheric processes like the forms of light scattering, atmospheric transmission and spectroscopic signature of common atmospheric molecular species. These play a crucial role for the choice of laser wavelength, bandwidth, sensitivity, etc. An overview of the different types of LIDAR schemes including IR/DIAL and Raman systems follows, with a detailed presentation of coherent IR light sources appropriate for LIDAR sensors. Finally, illustrative examples of air monitoring are given and perspectives and future developments outlined.

Although fiberoptic sensors are commonly employed for sensing the parameters temperature, pressure, or pH values, there is an increasing interest in their application to gas analysis. The article on **Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments** by G. Gauglitz, University of Tuebingen (Germany), discusses the basics of these sensors with emphasis on planar waveguides. Applications are illustrated by some representative examples.

U. Boesl et al., Technical University in Munich (Germany), present the technique of **Laser Mass Spectrometry in Trace Analysis**. The main characteristics of the method are outlined including its development. The applications discussed comprise exhaust emissions from combustion engines (traces of aromatic compounds behind the catalytic converter at fast changes of engine conditions and control of oil traces in the exhaust) and from incinerators (on-line measurements of polycyclic aromatic compounds, isomer-selective analysis of polychlorinated aromatics). An example of great current interest concerns the possible realization of a fast dioxin analysis. Furthermore, the on-line monitoring of volatiles in the headspace of food products and during

food processing is discussed. Future developments concern calibration standards and the highly selective trace analysis of organic compounds in solid samples.

Three articles focus on specific applications rather than on the technique as such, namely on explosives detection, combustion diagnostics, and elemental analysis.

J.P. Singh and D.L. Monts, Mississippi State University (USA), present Laser- and Optical-based Techniques for the Detection of Explosives. These include IR, Raman, UV/VIS, laser-induced fluorescence, as well as mass and ion mobility spectrometries and chromatography. The authors address solid, liquid and gas-phase samples, and focus on water and soil in their discussion of environmental applications.

J. Wolfrum et al., University of Heidelberg (Germany), discuss **Laser-based Combustion Diagnostics** which include the linear techniques of absorption spectroscopy, laser-induced fluorescence and incandescence, Rayleigh and Raman scattering, as well as nonlinear techniques such as coherent anti-Stokes Raman scattering, degenerate four-wave mixing, polarization spectroscopy, sum frequency mixing and resonance-enhanced multiphoton ionization. The applications comprise diagnostics of laminar and turbulent flames, engine combustion, and coal and waste combustion.

In their second contribution, J.P. Singh and D.L. Monts, Mississippi State University (USA), address **Laser-induced Breakdown Spectroscopy, Elemental Analysis**. After a presentation of the technique and some fundamental studies, they focus on some examples of environmental applications, including molten metal, glass, soil, paint, air sampling filters, and radioactive species. The performance of LIBS with regard to accuracy and precision, sample preparation, sampling rate and detection limits is compared with competing techniques.

R. Hirschberger, Institut für Spektrochemie und Angewandte Spektroskopie (ISAS), Dortmund (Germany), focuses on **Automotive Emissions Analysis with Spectroscopic Techniques**. After a discussion of traditional techniques such as nondispersive IR and UV or chemiluminescence systems the author discusses more recent IR techniques based on FTIR and diode-laser systems with applications for on-line measurement and analysis of hydrocarbons. Linear and nonlinear Raman techniques are also presented, with applications for on-line and in situ measurements of hydrocarbons, hydrogen, oxygen, and nitrogen. The different methods are compared on the bases of handling, sensitivity, robustness, and costs. It is mentioned that most of these modern optical techniques are still laboratory research systems and have not yet been validated for field use.

Additional analytical methods for the trace analysis of air, not treated in this section, are presented in related sections of this *Encyclopedia*: Field Portable and Transportable Air and Vapor Measurement, Industrial Hygiene, Chemical Warfare Analyses, Environment (Water and Waste), Atomic Spectroscopy, Gas Chromatography, and X-Ray Spectrometry.

It should also be emphasized that the development and implementation of novel and improved instrumentation is an active field, so that devices with ever higher performance are regularly becoming available. It is thus hoped that this section may stimulate further research interest in the field of trace species characterization, analysis, and monitoring Although detection schemes per se cannot solve the challenging problems associated with air pollution, they can nevertheless provide a solid basis toward a sustainable improvement of the environment.

ACKNOWLEDGMENTS

As editor of this section I wish to thank all the authors for the overviews of their specialties, which point out the advantages and drawbacks of the numerous techniques available and which demonstrate the state-of-the-art performance by illustrative applications. I acknowledge their cooperation in following the guidelines provided, thus ensuring a useful and comprehensive section with little overlap between chapters. I am also indebted to the editor-in-chief, R.A. Meyers, and the managing editor, A. Holyoake of J. Wiley & Sons, Inc., for their initiative, helpful suggestions, critical comments, and (last but not least) for their understanding and patience with late manuscripts.

ABBREVIATIONS AND ACRONYMS

DIAL	Differential Absorption LIDAR
DOAS	Differential Optical Absorption Spectroscopy
EPA	Environmental Protection Agency
FTIR	Fourier Transform Infrared
HAPs	Hazardous Air Pollutants
IPCC	Intergovernmental Panel on Climatic Change
IR	Infrared
ISAS	Institut für Spektrochemie und Angewandte Spektroskopie
LIBS	Laser-induced Breakdown Spectroscopy
LIDAR	Light Detection and Ranging
UNCED	United Nations Conference on Environment and Development
UNFCCC	United Nations Framework Convention on Climatic Change

UV/VIS	Ultraviolet-to-visible
VOCs	Volatile Organic Compounds

REFERENCES

1. H. Horvath (guest ed.), 'Atmospheric Aerosols', *J. Aerosol Sci.*, **27**(3), 355–475 (1996).
2. G.M. Hidy, *Aerosols: An Industrial and Environmental Science*, Academic Press, Orlando, 1984.
3. B.J. Finlayson-Pitts, J.N. Pitts, *Atmospheric Chemistry, Fundamentals and Experimental Techniques*, J. Wiley & Sons, New York, 1986.
4. M.W. Sigrist (ed.), *Air Monitoring by Spectroscopic Techniques*, Chemical Analysis, J. Wiley & Sons, New York, Vol. 127, 1994.
5. E.D. Hinkley (ed.), *Laser Monitoring of the Atmosphere*, Topics in Applied Physics Series, Springer, Berlin, Heidelberg, Vol. 14, 1976.
6. R.M. Measures (ed.), *Laser Remote Chemical Analysis*, Chemical Analysis Series, J. Wiley & Sons, New York,Vol. 94, 1988.

Airborne Instrumentation for Aerosol Measurements

Leopoldo Stefanutti
Instituto di Ricerca Sulle Onde Elettromagnetiche ''Nello Carrare'' del Consiglio Nazionale delle Richerche, Florence, Italy

Stephan Borrmann
Institute for Physics and Dynamics of the Geosphere ICG1, Juelich, Germany

1 Introduction 1
2 The Atmospheric Aerosol 1
2.1 Atmospheric Aerosols and Clouds 1
2.2 Tropospheric Aerosols 4
2.3 Stratospheric Aerosols 5
2.4 Measurement Techniques 6
3 Aerosol Instrumentation 6
3.1 Remote Sensing Instruments for Satellites, Balloons, and Aircraft 6
3.2 Airborne in situ Instrumentation for Direct Reading Measurements 13
Acknowledgments 19
Abbreviations and Acronyms 19
Related Articles 20
References 21

This paper gives a description of airborne and balloon-borne instrumentation for the measurement of microphysical properties and the chemical composition of aerosol and cloud particles. Different techniques are described. In particular the instrumentation is subdivided into two broad categories: remote sensing instrumentation and in situ instrumentation. In the first category are instrumentation like cloud RADARs (radio detection and ranging), LIDARs (light detection and ranging), backscattersondes, limb and occultation aerosol detection devices. They can determine the microphysical properties of the aerosols and, only with complex a priori assumptions, their chemical composition. Among the in situ instrumentation, for example, optical particle counters (OPCs), condensation particle counters (CPCs) determine the microphysical properties of aerosols, while their chemical composition is determined by means of mass spectrometers, and the counterflow virtual impactor (CVI) technique associated with tunable diode laser and Lyman-α spectrometers.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Solid-, liquid- and mixed-phase aerosol particles are present in our atmosphere. Their characteristics and distribution are highly variable and change spatially and temporally with altitude. Aerosols play a fundamental role as nuclei for haze, cloud and hydrometeor formation, as well as in meteorology and atmospheric chemistry, and have great influence on the atmosphere's radiative budget. Therefore it is important to study the fundamental mechanisms involving particles and monitor changes related to the atmospheric aerosol. Various aerosol properties can be measured by means of satellites, airborne instrumentation, ground-based techniques and analytical methods in the laboratory. Ground-based in situ instruments measure physical and chemical properties of the aerosol on specific locations and constitute point measurements conducted usually for extended periods.

Remote sensing methods measure some optical properties of the aerosol and are capable of scanning a specific (often large) volume of air in the vicinity of the station. Similarly, satellite-based remote sensing instruments observe significant portions of the global atmosphere detecting large scale and coarsely resolved distributions of aerosols and clouds.

In order to obtain specific physical and chemical aerosol data, sometimes down to the scale of a single particle, airborne techniques have been developed to be deployed on aircraft or balloons utilizing passive and active remote sensing methods, as well as in situ techniques. Unlike aircraft, balloons are susceptible to uncontrolled wind drifts, but capable of reaching much higher altitudes.

The atmospheric region covered by airborne instrumentation extends from the planetary boundary layer (PBL) (from the ground to approximately 2 km) through the free troposphere (up to 17 km depending on geographical latitude and season) well into the stratosphere (from the tropopause up to 50 km).

In this article the different types of airborne instrument are discussed considering their detection limits and ranges, their reliability and range of applications. Also the different platforms available for airborne aerosol instrumentation and payload configurations are briefly described.

2 THE ATMOSPHERIC AEROSOL

2.1 Atmospheric Aerosols and Clouds

Aerosols are a natural constituent of the Earth's atmosphere and their origin is mostly natural, although lately the anthropogenic contribution is increasing. The total

mass of natural aerosols was evaluated to be four times larger than the one due to human activity in 1968, and by the year 2000 the anthropogenic component is estimated to double.[1,2] An aerosol is defined as a stable suspension of particles, liquid or solid, in a carrier gas like atmospheric air. Therefore dust, haze, smoke, smog, fog, mist, and clouds can be considered to be specific aerosol types. Atmospheric clouds are volumes of supersaturated air containing hydrometeors which are microscopic objects (e.g. aerosol particles, cloud droplets, ice and snow crystals) as well as macroscopic objects (snowflakes, graupel grains, raindrops, and hailstones). The individual particles have sizes with radii ranging from 0.001 μm to several hundreds of micrometers, or even centimeters as in the case of hailstones. According to Junge[3] a first classification can be made based on particle size. The particles with diameters below 0.1 μm are named ''Aitken particles'', those with sizes between 0.1–1 μm are ''large particles'', and those larger than 1 μm are called ''giant'' particles.[1] The aerosols may also be classified according to their origin into marine, continental, rural, remote, background, and urban aerosol, where each category has different characteristics like chemical composition, size distribution, shapes and so on. Aitken nuclei, i.e. the smallest atmospheric aerosol particles, are naturally produced from supersaturated vapors by a process called gas-to-particle conversion (GPC), often in connection with photochemical reactions between gas phase consituents. Anthropogenic activities like combustion and industrial processes increasingly lead to the release of aerosol particles, or precursor gases from which aerosol particles are formed, into the atmosphere. Large aerosol particles can be produced by combustion processes and other anthropogenic or natural processes (e.g. the coagulation of smaller Aitken nuclei). Natural sources for giant particles are bulk-to-particle conversion (BPC) processes like sea-salt aerosol production, or desert dust mobilization, and human activity contributes, for example through emission from industries, mining, biomass- and woodburning, and agricultural dust generation.

Figures graphically displaying the volume distribution of an atmospheric aerosol as function of particle size can exhibit up to three peaks (modes).[1] These peaks are termed as ''nuclei mode'', ''accumulation mode'', and ''coarse mode''. The nuclei mode consists of the primary particles formed by condensation of gases and vapors. Coagulation of such small particles, GPC and condensation of vapors onto existing small particles create the accumulation mode, and the particles contained in the coarse mode are predominantly generated by BPC. Typical modal particle sizes are diameters of 0.02 μm for the nuclei mode, and 0.2 μm and 5 μm for the accumulation and coarse modes, respectively. These sizes depend on sampling time and location of the atmospheric aerosol under consideration, and not necessarily all modes are present in a given sample. Although this ''three-mode concept'' is useful for theoretical and experimental studies, it has to be considered as a conceptual model because real atmospheric measurements may exhibit very different features.

Methods for aerosol measurement apply either physical or chemical principles. The number density of particles in the air for given sizes can be determined by using physical techniques. The thermodynamic phase of the particles, total mass concentration, total particulate surface area, total volume, and various optical properties can also be measured. On the other hand different chemical analyses are adopted in order to determine particle composition, density, and reactivity. Since the shapes of individual aerosol particles may be very complex and highly variable, the analysis of size-dependent parameters requires the introduction of an equivalent particle size diameter. This is defined as the diameter of a spherical particle which has the same physical properties (e.g. volume, cross-section for optical scattering) as the aspherical particle under investigation.

Because the particle size range encountered in the atmosphere spans as much as five orders of magnitude (or even seven if unstable aerosol populations like rain drops, snow flakes, and graupel grains are included), different techniques or even different physical principles have to be adopted in order to measure the aerosol property of interest. This cannot be achieved with a single instrument. For example, the sizes of small particles can be measured by detecting their mechanical or electrical mobility, while optical methods are employed for sizes above 0.3 μm up to roughly 100 μm.[4] For the sampling of particles above approximately 0.1 μm, with the option of subsequent chemical analyses, techniques based on impacting particles on surfaces are suitable. Physical and chemical analyses on particle samples are performed by means of electron microscopy, X-ray fluorescence (XRF), atomic absorption spectrometry (AA), emission spectrometry (ES), infrared (IR) spectroscopy, anodic stripping voltammetry (ASV), electron spectroscopy for chemical analysis (ESCA), ion chromatography, laser microprobe mass spectrometry (LMMS), proton (or particle)-induced X-ray emission (PIXE), neutron activation (NA), among a variety of other methods.[4,5]

The analysis of the vertical distribution of the physical and chemical aerosol properties is fundamental in the study of aerosol and cloud effects on the natural and anthropogenically disturbed environment.

For aerosol and cloud-related research, airborne platforms (like airplanes and balloons) are commonly utilized where platform type and instrumentation have to be

carefully selected depending on the specific scientific goals. The part of the atmosphere which is accessible to aircraft or balloons is structured in different layers with the PBL (from the ground up to approximately 1–2 km), the free troposphere (from the top of the boundary layer up to the local tropopause), and the stratosphere.

The characteristics and distribution of atmospheric aerosols are highly variable, changing spatially, temporally, and with altitude. The aerosol in the lower troposphere is particularly inhomogeneously distributed where large aerosol loads are found near the source regions (e.g. deserts) and during times of ''outbreak events''. The measured particle size distributions can exhibit nuclei, accumulation, and coarse modes, where most of the time not all modes are simultaneously present in a particular size distribution measurement. Mathematically these distributions are described by a log–normal function, and such multimodal distributions are used as models characterizing most of the atmospheric aerosol types.[1] The particle number concentration decreases with increasing particle size and Aitken particles dominate the total number of particles in an air sample. On the other hand their contribution to the total aerosol volume is small, while the largest contribution comes from the large and giant particles.

For particles with sizes ranging from a few nanometers up to tens of micrometers typical number concentrations can be coarsely specified, for example at the (continental) ground level total concentrations are near $1.4 \times 10^4\,\mathrm{cm}^{-3}$ particles with values decreasing towards $6 \times 10^2\,\mathrm{cm}^{-3}$ in remote locations. The total concentration in large cities is as high as $10^5\,\mathrm{cm}^{-3}$ particles, in small towns values around $3 \times 10^4\,\mathrm{cm}^{-3}$ are detected, while over the oceans, far from coastlines, $5 \times 10^2\,\mathrm{cm}^{-3}$ can be found. In clean Arctic environs measurements show total particle concentrations near $50\,\mathrm{cm}^{-3}$, in the remote Antarctic continent concentrations can be even lower. As with the atmospheric air density profile, the particle number concentration (in number of particles per cubic centimeter of air) shows an exponential decay for the lowest 6 km of the atmosphere. Above this level a background tropospheric aerosol is observed with values around 200–300 particles per cm^3 in the Aitken mode. Above the tropopause the total number concentration decreases again to a minimum below $1–10\,\mathrm{cm}^{-3}$ particles which remains constant to altitudes around 30 km.

In the case of the large particles (i.e. those with sizes between 0.1–1.0 μm) the total concentration decreases with altitude up to a minimum around 0.01–0.02 particles per cm^3 in the upper troposphere. In the stratosphere the values increase and a maximum around 0.1–1 particles per cm^3 is reached at altitudes between 17 km (in the polar regions) and 25 km (equatorial region) depending on geographical latitude, corresponding[6,7] to up to $1.5 \times 10^{-7}\,\mu\mathrm{g\,cm}^{-3}$. Above this maximum the concentrations decrease again. This stratospheric aerosol layer of binary sulfuric acid solution droplets is named the Junge layer and the altitude of its maximum also varies with latitude.[8]

Aerosols play a fundamental role as nuclei for haze, fog, cloud, and hydrometeor formation. If the particles are solid they offer a surface on which trace gases can undergo chemical reactions. If they are liquid they absorb gases which can react inside the solution droplets, thus forming and subsequently releasing new compounds into the atmosphere. This chemical role is very important especially in the case of polluted air, both in the lower troposphere and in the stratosphere as it is, for instance, for stratospheric ozone chemistry[9,10] and the acid rain phenomenon in the troposphere.[11] Furthermore cloud and aerosols have great influence on the atmosphere's radiative budget. Figure 1 from Intergovernmental Panel for Climate Change (IPCC)[2] demonstrates this by displaying the global mean radiative forcing (also referred to as ''greenhouse forcing'') for various radiatively important atmospheric trace substances, as well as the aerosol. These estimations include the changes (increases) in atmospheric concentrations of the displayed radiatively active compounds between the preindustrial era and 1992. Estimations of the change in forcing originating from variations in solar energy flux are indicated at the right side of the figure. The first bar indicates the contributions of four gases to the direct greenhouse effect. The indirect effects of the increase in tropospheric ozone and the decrease in stratospheric ozone on the greenhouse forcing is displayed by the next two bars. The following set of three bars pertains to the direct effect (see below) of the tropospheric aerosol. The single standing error bar following to the right represents the indirect effect on the aerosol of changing the cloud properties (see also below). This quantity is so uncertain that not even a mean value can be specified in the form of a bar as is possible for the other quantities in the figure. Of particular note are the error bars and indications of the confidence placed into these estimations which are given. These show large uncertainties for the role of the direct and indirect effects of the tropospheric aerosol. Here direct effects refer to the reflectance and extinction of radiation by the aerosols through scattering and absorption, whereas indirect aerosol-related effects are the modification of the formation and lifetime of clouds, the global and regional cloud coverage, as well as changes in the cloud optical properties. Therefore it is important to study the fundamental mechanisms involving particles and monitor changes related to the atmospheric aerosol and its effects.

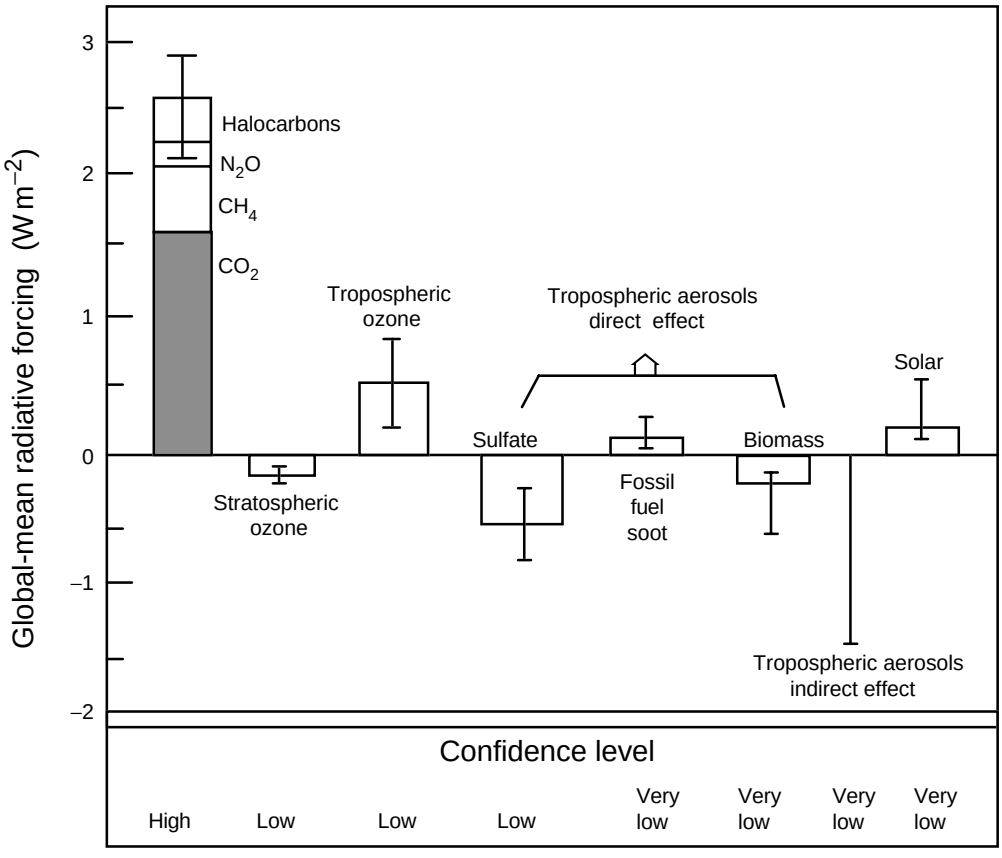

Figure 1 Global-mean radiative forcing (1850–1990) of various trace gases and aerosols. Both the direct and indirect effect of the aerosol on the atmospheric radiative budget are considered. The uncertainties are indicated as error bars, and confidence levels are also provided. The estimation of the indirect effects inherent in the modification of the cloud properties is associated with the highest uncertainties. (Reproduced by permission of IPCC (1996).[2])

Atmospheric aerosols and clouds are closely related, because each cloud droplet, ice crystal, and precipitation element in tropospheric clouds originates on an a priori existing aerosol particle. The levels of saturation with water vapor are never high enough in the atmosphere to allow for homogeneous homomolecular nucleation, i.e. the creation of liquid-phase water droplets directly from water vapor. Cloud droplets always form via heterogeneous nucleation on already existing aerosol particles.[12] If the Earth's atmosphere were free of aerosol particles, clouds and precipitation could not form. (As an exception, homogeneous heteromolecular condensation is possible in the atmosphere, as for example for the co-condensation of sulfuric acid and water vapor in the stratosphere.) Radiatively clouds influence the albedo of the planet, and chemically the cloud droplets act as little reactors influencing the chemical composition of the surrounding air. The distribution of the clouds, their optical, properties, as well as the chemistry associated with them are issues of primary importance in the understanding of climatic changes.

2.2 Tropospheric Aerosols

2.2.1 Planetary Boundary Layer

The largest part of the tropospheric aerosol is contained inside the PBL, where the majority of energy exchanges between the Earth's surface and the atmosphere occur. Also most of the aerosol generating BPC and anthropogenic processes take place there. The global rate of aerosol production is estimated to be of the order of 2000 Tg per year, 15% of which are attributed to anthropogenic sources. One-fifth of these 15% is believed to originate from direct anthropogenic emissions, which implies that 80% of the anthropogenic contribution derives from GPC. The natural contribution to the global aerosol generation rate is 60% on account of direct emissions, while 40% are due to GPC.[13] Urban aerosols are essentially anthropogenic and derive from combustion of fossil fuels, industrial and chemical production, as well as GPC and BPC. For example photochemical and sulfurous smog are typical phenomena connected with human activity in cities.[11] In the vicinity of the ground the aerosol

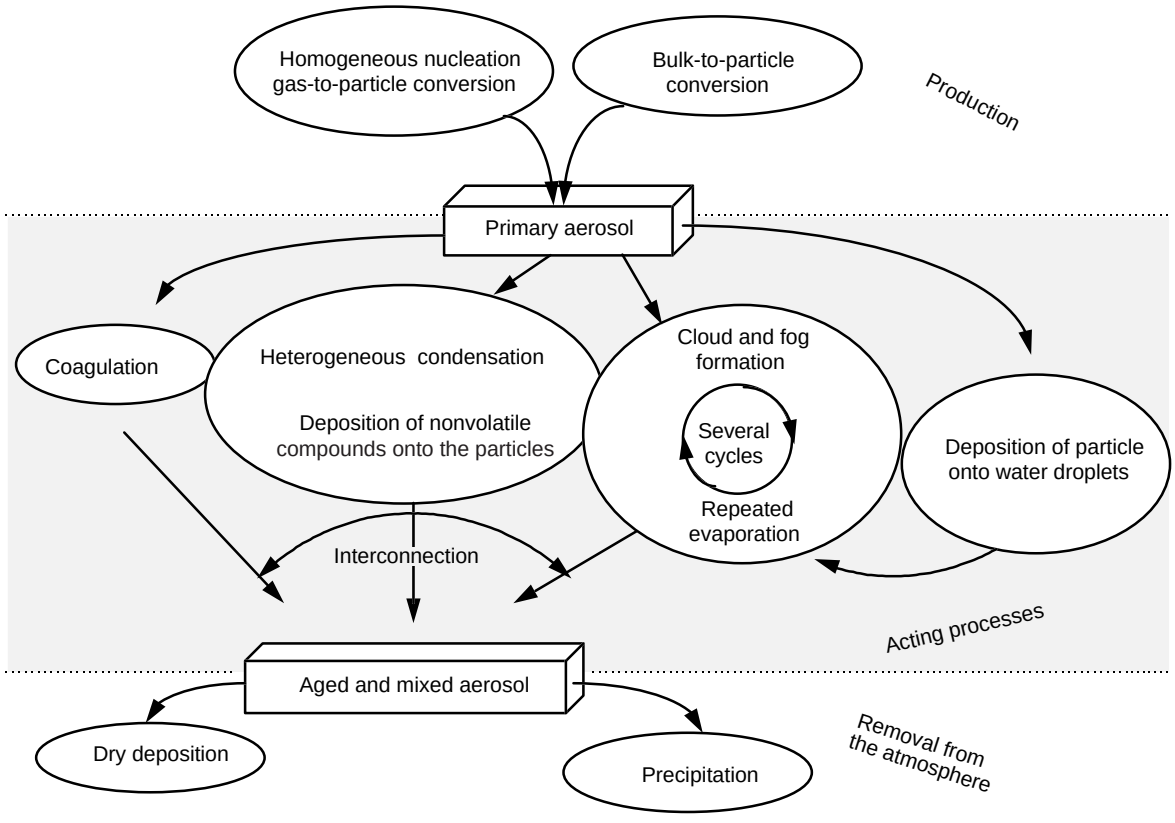

Figure 2 Aerosol-related processes in the atmosphere involving particle formation, processes altering particle characteristics, and removal from the atmosphere.

content of the air is rarely less than 1 μg of particulate matter per cm^3 of air. In cities values above 100 μg per m^3 are measured, and in sandstorms aerosol loads of as much as 30 000 μg per m^3 can occur. In general, if the measured mass concentrations of the aerosol are compared with those of the other atmospheric trace gases, then one finds that the contribution of the aerosol is equivalent to that of a trace gas in the ppbv (parts per billion by volume) range.

2.2.2 Free Troposphere

The free troposphere has less aerosols than the PBL. The largest part of the total atmospheric water vapor resides in the troposphere and the free troposphere is the location where clouds are found and the majority of meteorological events occur. Sand storms in desert areas produce clouds of dust and mobilize submicometer-sized particles (silicates),[14] which then can be transported in the free troposphere over distances of several thousand kilometers.[13] Similarly forest fires cause the transport of smoke and soot over long distances. Aircraft exhaust emissions are sources of tropospheric and lower stratospheric aerosols, the concentrations, characteristics and effects of which are largely unknown and are the subject of intense research.[15] The generation of water vapor from the aircraft exhaust also influences condensation, the formation of contrails and additional cirrus clouds in the upper troposphere.[16,17]

The residence time of freshly generated aerosol in the lower troposphere is of the order of a few days due to removal processes like coagulation, sedimentation, dry deposition, incorporation into cloud droplets, and scavenging by precipitation (wet deposition). Residence times in the upper troposphere can be 1–2 weeks. Figure 2 shows the most relevant processes to (1) aerosol production creating the primary aerosol, (2) the changes the aerosol is subjected to in the troposphere, and (3) the removal processes.

2.3 Stratospheric Aerosols

While tropospheric aerosols and clouds include a wide range of different particle types, the situation is less complicated in the stratosphere. The types of stratospheric aerosol can be categorized as stratospheric sulfate aerosols (SSA) and polar stratospheric clouds (PSC), where the latter are divided into PSC of Type I and Type II.[18] In addition, soot particle injections into the stratosphere from aircraft may play a role.

SSA are formed by a liquid binary solution of H_2SO_4 and water, and they are present at all altitudes from the tropopause up to about 30 km. Above that altitude the stratospheric temperatures are too high and the H_2SO_4 partial pressure is too low, such that the droplets would evaporate. In periods during which the stratosphere is not perturbed by heavy volcanic aerosol injections the distribution is relatively similar at all latitudes and the density is of the order of 10 particles cm^{-3}.[13,19,20]

PSCs of Type I are believed to be essentially composed of nitric acid in liquid solution, or of solid, or hydrate phase. These particles have sizes with radii of the order of 0.5–2 μm. PSCs of the sub category "Type Ia" are commonly believed to consist of nitric acid trihydrate particles (NAT; $HNO_3 \cdot 3H_2O$). For clouds of the second subcategory "Type Ib", it is generally accepted that they are formed by supercooled ternary solutions (STS) of H_2SO_4, HNO_3, and H_2O, with traces of HCl, HOCl, and possibly some bromine-containing compounds. Type II category PSCs consist of ice crystals, possibly with some HNO_3 and H_2SO_4. These particles are significantly larger having size diameters between 4 and 40 μm. Mass mixing ratios for Type I and Type II PSCs are in the order of 10 and 1000 ppbm, respectively.[21] Other hydrates like sulfuric acid tetrahydrate (SAT), nitric acid monohydrate or nitric acid dihydrate (NAM, NAD) could also occur because these are thermodynamically stable under stratospheric conditions.

Soot is found in the lower stratosphere and its presence can be particularly important due to its potential heterogeneous chemical reactivity and radiative effects. The most important source of soot in the stratosphere is due to commercial air traffic, as confirmed by recent measurements.[22,23] The maximum concentration of soot in the stratosphere is of the order of 1 $ng\,m^{-3}$ in the northern hemispheric aircraft corridors around the cruise altitude of the aircraft.

2.4 Measurement Techniques

Various aerosol and cloud properties can be measured by means of satellites, airborne instrumentation, ground-based techniques, and particle sampling accompanied by analytical methods in the laboratory. Ground based in situ instruments measure physical and chemical properties of the aerosol on specific locations and constitute point measurements conducted usually for extended periods.

Remote sensing methods measure an optical property of the aerosol and are capable of scanning a specific (often large) volume of air in the vicinity of the station. Similarly, satellite-based remote sensing instruments observe significant portions of the global atmosphere detecting large scale and coarsely resolved distributions of aerosols and clouds.

Satellite observations allow global coverage of aerosols and clouds, which serves as a fundamental input as in the case of weather satellites for meteorological predictions. Global aerosol distributions can be obtained, as for instance with satellite instruments like SAGE (stratospheric aerosol and gas experiment). But remote sensing techniques cannot give a detailed insight into the microphysics and chemistry of the particles. A priori assumptions on particle shape and index of refraction are necessary to retrieve aerosol parameters from the measured optical properties (mostly extinction, or backscatter).

In order to obtain specific physical and chemical aerosol data, sometimes down to the scale of a single particle, airborne techniques have been developed to be deployed on aircraft or balloons utilizing passive and active remote sensing methods, as well as in situ techniques.

Aircraft are versatile enough to allow for experiments inside the PBL, in the free troposphere, and in the lower stratosphere up to altitudes around 21 km. Balloons are susceptible to uncontrolled wind drifts, but capable of reaching much higher altitudes (approximately 40 km) than aircraft.

3 AEROSOL INSTRUMENTATION

A coarse categorization of aerosol instruments includes:

1. remote sensing instruments,
2. direct reading in situ methods, and
3. sampling devices followed by, in most cases, sophisticated laboratory analyses.

A second classification divides the instruments into those which detect physical, or optical properties, and those which determine particle chemical composition and chemical properties.

3.1 Remote Sensing Instruments for Satellites, Balloons, and Aircraft

Remote sensing for satellite or airborne deployment may be based on passive sounders or on active probing of the atmosphere.

These instruments are only capable, in most cases, of measuring the properties of large populations of particles, while in situ instrumentation permits individual particle analyses. The components of active remote sensing instruments consist of a transmitter in the optical or microwave bands, a reception antenna, and the data acquisition system. In principle, an electromagnetic pulse is emitted into the atmosphere and the receiving

system detects those parts of the radiation which are scattered back by the aerosol particle population or the hydometeors present in the air. Typical instruments of this class are RADARs and LIDARs. LIDARs are particularly well suited for aerosol and (thin) cloud detection and characterization, while RADARs are used for (thick) cloud and precipitation measurements. Passive sounders use the sun (or the moon) as a light source performing some kind of optical extinction measurement in most cases.

Generally, remote sensing instruments for aerosol measurements deliver optical characterization of the particle population, movement velocities of hydometeors, rain hail fall intensities, as well as low resolution particle size distributions under certain circumstances. They are widely used to observe individual, or groups of clouds, and monitor their temporal and spatial evolution. For most of these instruments complex mathematical methods with inherent assumptions, uncertainties, and ambiguities need to be applied in order to convert the measured return signal into the desired physical quantity.

3.1.1 Radio Detection and Ranging Measurements

The RADAR technique was originally developed during World War II for military purposes. It consists of a microwave transmitter, a receiving antenna, and data acquisition electronics. A microwave pulse, generated initially by a magnetron, was transmitted through the atmosphere. The beam was collimated using the same transmitting and receiving antenna to reduce the area of search. A small part of the pulse energy is scattered back from the target and detected by the receiving antenna. The distance of the target is equal to half the time-of-flight of the pulse multiplied by the speed of light. Over the years the RADARs have become more sophisticated and suitable for meteorological applications. Conceptually this is a qualitative improvement, because what was initially only a background noise for the RADAR, became the object of investigation. Since the atmosphere is formed by molecules, aerosols and clouds, clouds and raindrops can become the targets for a RADAR. Meteorological RADARs are used today on aircraft for monitoring meteorological conditions during the flight. A new class of RADARs have been developed in the field of science for the investigation of clouds and rain. Such systems have been studied essentially to obtain information on the atmospheric water budget. The main aim of such systems is to be deployed on satellites thus providing global coverage. Still for many scientific aircraft operations a cloud RADAR can be a useful tool, to operate with other active sensors, like LIDARs. Depending on the electromagnetic band in which such RADARs operate, they may be used to detect clouds (35, 70, or 94 GHz), in the high frequency band, or rain in the low frequency band (10 GHz). RADARs operating at high frequencies permit the detection of the cloud's droplet population. Therefore they yield estimates for the cloud liquid water content. Since such RADARs operate on much longer wavelengths than the LIDARs, they penetrate much deeper into optically thick clouds with the ability to give more complete information on cloud types like cumulus, cumulonimbus, thunderstorm cells and so on. Comparisons were carried out between the GKSS (Hamburg, Germany) cloud RADAR MIRACLE (Microwave Radar Cloud Layer Exploration) operating at 95 GHz and the GKSS backscattering LIDAR system (BELINDA, Broadband Emission LIDAR with Narrowband Determination of Absorption) in Germany in September 1997.[24] Both systems are ground-based instruments, but the result is of general interest: it appears that in particular conditions of relatively optically thick clouds, the cloud RADAR penetrates much better inside the cloud as expected, while the LIDAR can detect optically thin layers, formed by small particles, which are transparent to the RADAR. A cloud RADAR can give information which is complementary to that obtained by a LIDAR, and the benefits of simultaneous measurements with both are the subject of research. Cloud RADARs are also used to derive information on rain, snow and hail, and precipitation fall rates. Doppler RADARs additionally yield information concerning in-cloud wind velocities. In the presence of precipitation, the strong signal originating from the larger precipitating elements masks the signal from the smaller droplets. To measure rainfall, RADARs in lower wavelength bands have been developed operating in the C or X band, 5 GHz or 10 GHz, and are designed to operate on board satellites. During 1993 the NCAR (National Center for Atmospheric Research) ELDORA (Electra Doppler Radar) was installed on the NCAR's aircraft during the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean Atmospheric Research Experiment (COARE) campaign.[25] It was used in the X-band and was a dual doppler RADAR. It was able to collect dual Doppler information as the aircraft flew past or through storms (storm reflectivity and kinematic structure).

3.1.2 The Light Detection and Ranging Technique

Closely related to RADARs and of similar principle are LIDARs.[26,27] The LIDAR transmitter is a pulsed laser beam in the optical, UV (ultraviolet), or IR spectral range, while the receiving antenna is an optical telescope in connection with one or more detection channels, and a suitable electronic data acquisition and storage system. The optical detectors can be photomultiplier tubes or avalanche photodiodes. The emitted pulse (or sequence of pulses) may be at one single or several different wavelengths, generated by one or several lasers. Since

aerosols have dimensions ranging from 0.001 to a few hundreds of micrometers, the most suitable wavelength depends on the types of particle under investigation. The bulk of the aerosols range in diameters from 0.1 to several micrometers and in this region the most suitable wavelengths are in the visible part of the spectrum.

The LIDAR Equation (1) is:

$$P(R) = E\frac{K}{R^2}A\tau(\beta_{\text{ray}}(R) + \beta_{\text{a}}(R))\text{e}^{-2\eta \int_o^R (\sigma_{\text{ray}}(r)+\sigma_{\text{a}}(r))\,\text{d}r} \quad (1)$$

where $P(R)$ indicates the received power as a function of the distance R, E is the pulse emitted energy, A is the receiver's area, K is an instrument-dependent constant which takes into account the various losses of the system and τ is the length of the laser pulse. Since they are exposed to visible electromagnetic waves the air molecules scatter the light according to Rayleigh scattering theory. For aerosol particles having dimensions of the order of the wavelength or larger, the scattering is described by the Mie theory. The variables β_{ray} and β_{a} are the backscatter cross-sections for molecules (Rayleigh scattering) and for aerosols (Mie scattering). Similarly in the exponent the term η indicates a possible contribution due to multiple scattering, and σ_{ray} and σ_{a} denote the extinction coefficient due to the gas molecules and the aerosols. While the relation between β_{ray} and σ_{ray} is well known and can be computed, the relation between the backscatter coefficient and the extinction coefficient for aerosols depends on the specific aerosol, and particularly on the shape of the particles.

The light scattered by molecules S_{ray}, and hence the received signal, is proportional to the fourth power of the inverse of the wavelength λ (Equation 2):

$$S_{\text{ray}} \propto \lambda^{-4} \quad (2)$$

When the light interacts with particles of the size of the wavelength a signal proportional to the second power of the inverse of the wavelength is produced (Equation 3):

$$S_{\text{M}} \propto \lambda^{-2} \quad (3)$$

For tropospheric cloud particles in the range 20–100 μm, the optical wavelengths produce a flat response, since such wavelengths are much smaller than the particle size.

A laser pulse propagating in the atmosphere is scattered therefore both by aerosol particles and molecules. The ratio between the signal due to aerosols and the signal due to molecules results (Equation 4):

$$\frac{S_{\text{M}}}{S_{\text{ray}}} \propto \lambda^2 \quad (4)$$

This means that in order to enhance the difference between the signal due to aerosols from the molecular contribution it is necessary to shift toward larger wavelengths. For particles, ranging from 0.1 to a few micrometers (i.e. aerosols and PSC), wavelengths in the optical region are particularly suitable. For such cases, if the particles are liquid and hence spherical, it is possible, by means of multispectral investigation and Mie theory calculations, to derive the particle size distribution.[28,29] For solid aspherical particles general solutions of the Maxwell equations can be obtained for simple cases (i.e. spheroids, Chebyshev particles, etc.) and T-matrix[30] calculations can give some indication of particle size.

If the layer is optically thin, then contribution of multiple scattering can be neglected and considered as equal to one, which is certainly the case for stratospheric aerosols, PSC, and generally true for optically thin cirrus clouds.

The variables β_{ray} and β_{a} are related to the number of molecules and particles present in the atmosphere, according to the relation (Equation 5):

$$\beta_{\text{a}}(\lambda, R) = \int_0^\infty N_{\text{a}}(r)\sigma_{\text{abac}}\,\text{d}r \quad (5)$$

where r is the scatter dimension, $N(r)$ is the number of scattering elements present in the atmosphere having a dimension comprised between r and $r + \text{d}r$, and σ_{abac} is the backscattering cross-section. The contribution to backscattering due to the molecular atmosphere (Rayleigh contribution) is given in Equation (6):

$$\beta_{\text{ray}}(\text{m}^{-1}) = 3.44 \times 10^{-8}\frac{p(\text{mbar})}{T(\text{K})\lambda(\mu\text{m})^{4.08}} \quad (6)$$

where p the atmospheric pressure and T is its temperature. It is possible to derive direct information from LIDAR measurements on the particle size distribution, on the total volume, and total surface of the particles if the contribution due to molecular atmosphere can be separated and subtracted from the total signal. Then, in the case of spherical particles, using multispectral LIDARs, it is possible to invert the LIDAR signatures to obtain the particle size distribution and index of refraction by means of Mie theory computations. A minimum of five wavelengths are necessary. The minimum requirement at one wavelength is that one receiving channel has to be in the plane of polarization of the transmitted pulse, which must be linearly polarized, and one must be in the perpendicular plane. The presence of a depolarized signal from an optically thin medium is indicative of the presence of aspherical particles, and in such a case T-matrix computations or other techniques must be applied. If there is no depolarization, then the different LIDAR signatures, obtained simultaneously, can be processed to derive the optical parameters of the medium, and also information on the particle size distribution. This has already been accomplished by ground-based LIDARs

and a wide network of multispectral LIDAR systems is presently operating at very different latitudes and longitudes. For cloud particle size evaluation, with particles in the range between tens and hundreds of micrometers, an equivalent set of wavelengths in this range would be necessary. IR LIDARs covering this wide range do not yet exist. In addition to this, for wavelengths in or near the IR, the aerosol signal is much stronger than the molecular signal, but is difficult to detect using currently available detectors unless coherent detection is used, which is still quite a complex technique. This is generally applied only for wind measurements and is not recommended for aerosol detection. Therefore the most suitable and the most used spectral region for LIDAR application is in the visible spectrum between approximately 350–1000 nm.

Calibration of a LIDAR operating in the visible spectrum is relatively simple as the molecular signal can be detected. If the LIDAR is firing upward in the stratosphere, the signal above 25 km is generally caused only by air molecules and is used to calibrate the system. If the system is used only in the troposphere, there are generally areas in the free troposphere where aerosols loading contributing to the LIDAR signal is minimum. The signal detected from this region can be used for calibration. This simple technique avoids the necessity to measure pulse per pulse laser energy and the instrument constant, two quantities which may change either from pulse to pulse or in time. For IR LIDAR systems topographic targets are used as no molecular signal can be detected.

Ground-based LIDARs produce very good climatologies above the ground station, but are limited because they can give information only above the site of observation; this is why a wide ground-based LIDAR network has been established, but horizontal resolution is still nonexistent. Therefore space or aircraft systems have been developed.

A space system can give global coverage of the atmosphere of our planet.[31] The first LIDAR in space, mounted on board the shuttle LITE (Laser in Space Technology Experiment) was equipped with a multispectral LIDAR to monitor tropospheric clouds and aerosols and stratospheric aerosols. The laser transmitter was a Nd-YAG laser operating at the following wavelengths: 1064 nm, 532 nm, 355 nm. LITE flew successfully from September 9 to September 19, 1994,[32] demonstrating the capability of a space LIDAR to measure tropospheric clouds and aerosols from the tropopause to the Earth's surface and even measure stratospheric background aerosol distribution.

On aircraft several LIDARs have been flown since the start of the 1980s. Early in their history LIDARs were used from aircraft to monitor phenomena connected with PBLs, like the tracking of plumes or the monitoring of urban aerosols.

LIDARs have been operated both on tropospheric aircraft and presently on stratospheric aircraft to measure from the aircraft downward (nadir) or upward (zenith). The two operations are quite different. In fact the LIDAR signal, like the case of the ground-based LIDARs, reduces with the square of the distance and because the backscattering coefficients decrease almost by an exponential term with altitude, since they depend on the air density when the LIDAR is looking upward (see the LIDAR Equation 5), but remain almost constant if the pulse is fired downward, as the reduction in the signal by the square of the distance is compensated by the increase in the air density. This is the main reason why for the case of a spaceborne LIDAR, if the pulse can be kept collimated, the great distance of the spacecraft from the atmosphere is not a problem; even from space, eye safety remains one of the main problems when firing a laser, at optical wavelengths, downward.

For tropospheric studies, depending of the flight altitude of the aircraft, both upward and downward looking solutions are possible. If stratospheric measurements have to be carried out, then only the up-looking solution is used.

Finally airborne LIDAR systems have been flown since the mid-1980s. Some examples are: the large National Air and Space Administration (NASA) DC-8 airborne LIDAR systems, used by Browell and co-workers,[33,34] McCormik and co-workers,[35,36] and others,[37] which measured stratospheric aerosols, like for instance the very early Pinatubo cloud, PSCs, tropospheric clouds and aerosols. On the ER-2 a small compact LIDAR system was developed by Spinhirne and Hart.[38] Finally coherent LIDARs have also been flown, operating in the 10 μm range. Such systems were mainly devoted to wind measurements, but could also deliver very good aerosol data as a by-product. Vaughan et al.[39] have developed, for example, these types of system. As examples, three European recent airborne LIDAR systems are described in more detail in sections 3.1.2.1, 3.1.2.2 and 3.1.2.3. These three systems represent different concepts of airborne LIDARs: OLEX (Ozone Lidar Experiment), is similar to a ground-based system, but operates inside a pressurized aircraft and needs an operator during the flight; ABLE (Airborne Lidar Experiment) is an automated high power LIDAR, installed in an unpressurized bay of a high altitude aircraft; and finally MAL (Miniaturized Airborne Lidar) is an automated microjoule LIDAR, using a diode-pumped Nd-YAG laser and seems to be one of the most promising technologies for future space operation.

3.1.2.1 The Ozone Lidar Experiment System Among the operational LIDARs are ALEX (Aerosol Lidar Experiment) and OLEX of the German DLR (Deutsches

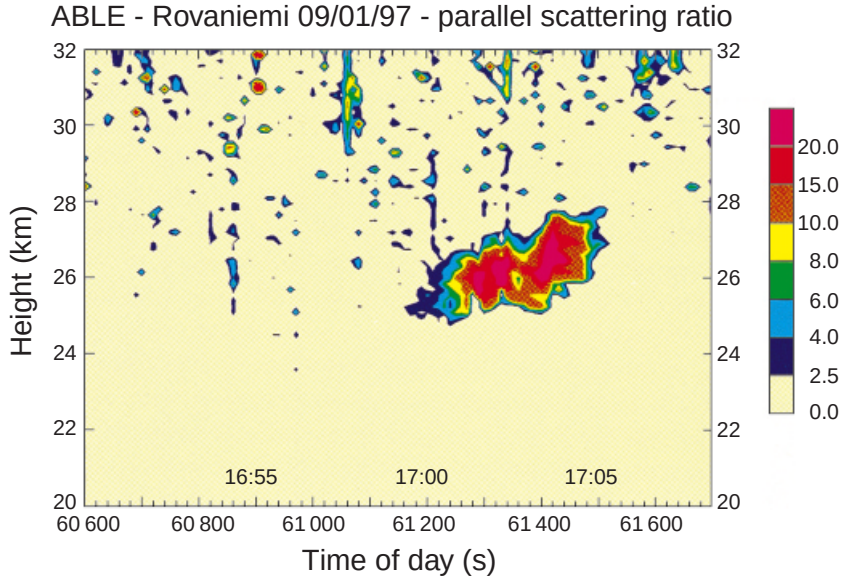

Figure 3 High altitude PSCs measured from ABLE during the airborne polar experiment–polar stratospheric clouds, leewaves, chemistry, aerosols and transport (APE-POLECAT) campaign held in Rovaniemi (Arctic circle) during the winter 1996/1997. (Courtesy of Giorgio Fiocco, University of Rome.)

zentrum für Luft-und Raumfahrt), which have flown on different scientific missions to monitor tropospheric clouds, stratospheric aerosols, PSC, and stratospheric ozone.[40,41] Such LIDARs are representative of large airborne LIDARs, flying on tropospheric aircraft, which need an operator on board. The OLEX LIDAR is used as an aerosol and ozone LIDAR. It has been mounted the German TRANSALL and the DLR Falcon. The system consists of two laser transmitters: (1) a Nd-YAG laser operating at a repetition rate of 10 Hz in the fundamental, second harmonic, and third harmonic with output energies of 200 mJ at 1064 nm, 120 mJ at 532 nm, and 180 mJ at 355 nm, and (2) a Xe–Cl laser operating at 10 Hz, emitting 200 mJ at 308 nm.

The receiver system is based on a 35 cm cassegrain telescope with a field of view (FOV) of 1 mrad and a focal length of 500 cm. The system uses five detection channels, which are one channel for 1064 nm, two channels for 532 nm in two polarization planes, one channel for 355 nm, and one for 308 nm. The 308 and 355 nm channels are specifically dedicated to ozone profiling in the stratosphere, while the other wavelengths are used for aerosol and cloud monitoring. The acquisition electronics can acquire each single signature, with the lasers firing at 50 Hz repetition rate, and the aerosol distributions measured by the system being available in real time. Thus decisions on measurement strategies can be made on board by means of the LIDAR. The total weight of the system is above 270 kg and the total power consumption above 1.6 kW.

Similar LIDARs have been developed in the USA by NASA and mounted the NASA DC-8 for tropospheric and stratospheric studies, as indicated in the previous section. Other airborne LIDARs have been developed by the French Centre National de la Recherche Scientifique (CNRS), as the Lidar Embarqué Pour L'étude des Aérosols et des Nuages, de Interaction Dynamique-Rayonnement et du cycle de L'Eau (LEANDRE), used for stratospheric aerosol and PSC measurements.

3.1.2.2 The Airborne Lidar Experiment System for a High Altitude Research Aircraft An example of LIDAR mounted on a stratospheric aircraft is ABLE,[42] developed for the M-55 Geophysica stratospheric aircraft. This system emits pulses both in the upward and in the downward direction. ABLE was developed by the University of Rome and consists of a flash lamp pumped Nd-YAG laser transmitter. The laser operates with a repetition rate of 10 Hz and produces output energies of 250 mJ at 1064 nm (near IR), 150 mJ at 532 nm (visible), and 70 mJ at 355 nm (near UV) which can be selected depending on the scientific goals. The receiving optics is formed by a 40 cm telescope receiver with the availability of two or more acquisition channels. The transmitted pulse is linearly polarized and an option is implemented to detect the direct and the perpendicularly polarized backscattered light signals. The system operates in a coaxial mode, i.e. the output of the laser beam is coaxial with the receiving telescope. Total overlapping between the FOV of the receiver and the emitted beam is reached after slightly over 1 km. With dimensions of

1700 × 900 × 800 mm, and weighing over 250 kg, ABLE is a large instrument, operating fully automated. A computer controls all operations and maintains the acquisition of the data. The LIDAR starts operation when the aircraft has reached a given altitude and the agreement of the pilot is needed to start emitting the light beams. The instrument is mounted in the main bay of the M-55 Geophysica, and by means of a plane rotating mirror it can fire both in the nadir and the zenith direction. In the latter case the laser beam emerges vertically only a little over 1 m from the pilot's cockpit. Due to the low pressure at high altitude and the low optical density of the high cirrus and PSC, the side scattering of the laser beam poses no hazard for the pilot. Due to the high energy of the system, LIDAR signatures can be obtained up to above 30 km altitudes, when looking upward. From Figure 3 it can be seen that very thin PSC at altitudes up to 29 km can be measured. In the nadir-looking mode, the system can penetrate deeply into relatively thick clouds and in general it is possible to obtain signals from near the Earth's surface. In its nadir configuration ABLE is designed to be eyesafe in the UV wavelength region. It is not eyesafe in the green at full energy, which can be used for measurements in the zenithal mode. The instrument was flown on the M-55 Geophysica inaugural mission (the Airborne Polar Experiment) in the Arctic in December 1996–January 1997.

3.1.2.3 The Miniaturized Airborne LIDAR A different type of airborne LIDAR is the MAL,[43] also mounted on board the M-55 Geophysica. A conceptually similar system was flown before on the ER-2.[38] The aim of the MAL system is to monitor the first 1–2 km above or below the aircraft, a role complementary to a high energy LIDAR like ABLE. Therefore it is necessary that the FOV of the MAL transmitter and receiver overlaps at distances of only a few tens of meters from the instrument. The spatial resolution of the retrieved signal has to be of the order of 50–100 m. In order not to saturate the detector in the near-field range, MAL utilizes a laser diode in the near-IR as a laser source. The laser operates at a high frequency repetition rate (10 kHz). MAL, developed under a joint Swiss–Russian cooperation between the Observatoire Neuchatel and the Space Research Institute of Moscow, consisted of a laser diode operating at 836 nm, with a bandwidth of 7 nm, and 0.8 μJ per pulse. Two detection channels were available to measure the direct signal and the crossed-polarized signal. MAL is a compact lightweight instrument, mounted behind the cockpit of the M-55 Geophysica. This type of LIDAR is similar to backscattersondes, with the main difference that it presents a range-resolved signature. MAL flew in the Airborne Polar Experiment in the Arctic on board the M-55 Geophysica aircraft. It operated only at night. The maximum range of the return signal was a kilometer. An upgraded system was designed using a diode pumped Nd-YAG laser. This laser operates in the second harmonic (532 nm) of the YAG output energies in the range of 2 μJ per pulse and a pulse repetition rate of 10 kHz. This permits the use of small photomultiplier tubes with incorporated photon-counting electronics and operation in a spectral range of high sensitivity thus enhancing greatly the operational capabilities of MAL and consequently permitting daytime operation.

3.1.2.4 Backscattersondes Backscattersondes are similar to the microjoule LIDAR. In principle a backscattersonde operates like a MAL, having a low energy optical source and measuring the backscattered radiation from an air volume close to the instrument. The backscattered signal is not range-resolved. Since the scattering volume is close to the instrument, the extinction term in Equation (4) can be neglected and a direct measurement of the backscattering coefficient can be achieved. However, the backscattersonde signals require calibration.

Backscattersondes have been flown on stratospheric balloons for several years. Generally a backscattersonde has a flashlamp as transmitting source and at least two receiving channels at two different wavelengths (e.g. 490 nm and 940 nm, like the type of backscattersonde developed by Hofmann and Rosen[44] at the University of Wyoming, USA). Backscattersondes are produced as small lightweight packages which can be flown on standard ozone sounding stratospheric balloons. The detection of two different wavelengths permits information to be gathered on the particle size.

More sophisticated instruments have been developed which have operated on both balloons and aircraft using as emitting source diode lasers operating at different wavelengths as the light source. An example of such a system is MAS (Multiwavelength Aerosol laser Scatterometer).[45]

A laser scatterometer can simultaneously emit a few polarized laser beams at different wavelengths and can detect the light scattered back towards the instrument by the particles and molecules in the air. The detected backscattered light and its change of polarization (the so-called depolarization) can be interpreted, if combined with OPC measurements, in terms of particle shape and refractive index. The OPCs (see section 3.2.2) measure in situ particle size distributions by means of a light scattering technique. For this, Mie theory has to be applied, which necessarily requires information about the particle's refractive index and shape. Usually for liquid stratospheric sulfuric acid droplets, justified a priori assumptions concerning these particle properties can be made. However, if the in situ measurements

take place at ambient temperatures low enough for freezing of sulfuric acid droplets to occur, or close enough to nitric acid hydrate saturation, or water vapor saturation temperatures, similar a priori assumptions cannot easily be made. Then a scatterometer operated on the same platform as the OPC can actually measure particle refractive index and shape. Conversely, the refined interpretation of the data delivered by the scatterometer critically depends on the simultaneously measured particle size distribution.

The laser scatterometer MAS is equipped with three laser diodes contemporarily emitting three 100 : 1 polarized beams in the atmosphere at 680, 780, and 830 nm. The emitted power is about 20 mW for each single laser. The telescope used to collect the backscattered light has a diameter of 120 mm. The atmospheric backscattering is collected by six different channels: three are tuned on the three laser wavelengths at parallel polarization, two are tuned on two laser wavelengths but for depolarized backscattering, and one is to measure the atmospheric background in intermediate wavelengths. The maximum spatial resolution is 200 m. A first prototype of a balloon-borne laser scatterometer has already flown in Antartica and other tests where performed in the Arctic during the Second European Stratospheric Arctic and Middle latitude Experiment (SESAME) campaign with the aim of studying the stratospheric clouds which form in the cold polar stratosphere.

3.1.3 Balloon-borne Optical Instruments for Aerosol Analysis Using Limb Scanning Techniques

The Laboratoire d' Optique Atmospherique (LOA) of Lille in collaboration with the Belgian Institute for Space Aeronomy (BIRA-IASB) has developed two instruments to be flown on stratospheric balloons to obtain information on aerosols and trace gases.[46] These instruments are BALLAD (Balloon Limb Aerosol Detection) and BOCCAD (Balloon Occultation for Aerosol Detection).

The BALLAD (Figure 4) instrument was built at LOA during 1993 to observe the Earth's limb for stratospheric study. The objective of the instrument forms the Earth's limb image on a linear charge-coupled device (CCD) detector of 1728 pixels, the vertical FOV of the system is about 6°. A filter wheel, located between the objective and the detector, allows multispectral radiance and polarization measurements to be performed. Two interference filters at 850 and 450 nm are dedicated to aerosol study, a third channel centered at 600 nm is deployed for ozone detection. Three other filters centered at 850 nm are equipped with polaroids rotated by 60° with respect to one other to evaluate the linear degree of polarization of scattered light.

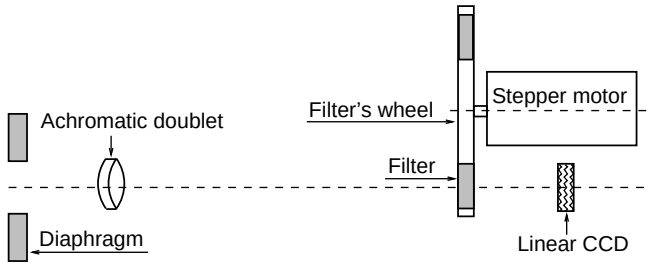

Figure 4 Schematic diagram of BALLAD. The filter wheel located between the objective and the detector (a CCD) allows the multispectral radiance and polarization measurements to be made. For measurements of aerosol the 850 and 450 nm interference filters are used. See text for details. (Courtesy of Colette Brogniez, University of Lille.)

BALLAD is installed on board of the gondola of a stratospheric balloon and operates at the ceiling level of 30 km. Its optical axis, pointing approximately 2° below the balloon horizon combined with its large vertical FOV allows observation of all the stratospheric layers, from the tropopause up to the balloon level, simultaneously with one exposure. The vertical resolution is about 350 m for a balloon ceiling at 30 km and a tangent height equal to 15 km.

Measurements are carried out for solar elevation ranging from 2–10° and the rotation of the gondola around the vertical axis (1/3 rpm) permits BALLAD to observe Earth's limb for various azimuth angles (angular sampling 16°).

The inversion of the nonpolarized radiance measurements at 850 and 450 nm is achieved at LOA; the inversion of the polarized measurements is being performed at BIRA-IASB. They both work in a single scattering approximation. Assuming that the layers on the line of sight are reasonably homogeneous over a sufficient horizontal extent, the circular scans provide measurements for different scattering angles. The normalized radiance, called reflectance r ($=L/E$, where L is the measured radiance and E the spectral solar irradiance at the top of the atmosphere), depends on the aerosol extinction coefficient and on the aerosol scattering phase function, as well as the same parameters for molecules. Consequently the directional behavior of r at a fixed tangent altitude contains information about the phase function multiplied by the aerosol concentration. The retrieval procedure[47] is based on a nonlinear fit to a Henyey–Greenstein function for the azimuthal dependence. The Henyey–Greenstein function is a good approximation of the phase function in the forward direction so that the inversion is restricted to the 0–80° scattering angle range, which also allows inhomogeneity effects to be limited.

The data inversion scheme based on the Mill method then provides an aerosol extinction profile versus altitude

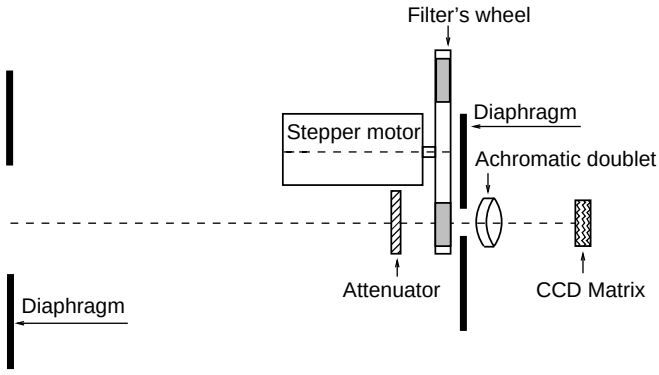

Figure 5 Schematic diagram of BOCCAD. In this case the instrument observes the solar occultation through the atmosphere. For the aerosols measurement interference filters centered at 850, 780 and 443 nm are adopted. A new channel, corresponding to that used by POAM II (Polar Ozone and Aerosol Measurements II), is used compared to BALLAD. See text for details. (Courtesy of Colette Brognez, University of Lille.)

and an asymmetry factor profile for each channel, from ca. 12–15 km (depending on the channel) up to ca. 25 km.

The BOCCAD (Figure 5) instrument was built by LOA in 1994[48] to observe solar occultation through the atmosphere. It is planned to provide useful information for the other LOA experiment RADIBAL (RADIometer BALloon), and to provide a direct comparison with satellite measurements SAGE II and a POAM II. The objective forms the Sun's image on a CCD matrix of 244×550 pixels. The vertical and horizontal FOV of the instrument are 10° and 7.5°, respectively. A filter wheel, located between the objective and the detector, allows multispectral measurements to be performed. Three interference filters centered at 850, 780 and 443 nm (i.e. identical to BALLAD with one additional channel corresponding to the POAM II channel) are used for aerosol studies, and a fourth channel at 600 nm is dedicated to ozone detection.

BOCCAD is installed on the same gondola as BALLAD and also operates at the balloon ceiling level. The gondola stops rotating after the BALLAD measurements and the instrument points 2° below the balloon horizon so that the large FOV allows observation of the whole occultation event.

The measurements begin when the solar elevation reaches ca. 2° and the Sun's image can be obtained almost outside the atmosphere to provide a reference image in each channel. Typically about 300 images are formed successively in each channel during a sunset at high latitudes.

Radiance measurements averaged on a few pixels located in the middle of each Sun's image are converted into transmission data by using a reference image. These transmissions are a function of the local (solar) time. Knowledge of the Sun's position with ephemeris allows the time–transmission data to be converted into tangent altitude–transmission data taking into account the atmospheric refraction. The transmission is obtained with 350 m vertical resolution. The inversion is conducted in the same way as was done by LOA with SAGE II[49] and POAM II data:[50] after correction for molecular transmission the 850, 780 and 600 nm channels give the aerosol plus ozone transmission and the 443 nm channel represents the aerosol plus nitrogen dioxide transmission. A Chahine's inversion procedure gives the global extinction coefficients and the separation of the species can be effected.

1. The 850 nm channel is dominated by the aerosol at the lowest levels (typically 20 km), and then leads directly, for these levels, to the aerosol extinction coefficient. The 443 nm channel, corrected for a weak nitrogen dioxide contribution by assuming a standard profile, gives the aerosol extinction coefficient at all altitudes.
2. The 600 nm channel is dominated by ozone at altitudes higher than ca. 20 km and gives a first estimation of the ozone extinction coefficient. For the lower levels an aerosol correction is needed to derive extinction.
3. A complete ozone profile is then available and this allows retreival of the ozone contribution from the 850 and 780 nm channels to retrieve the aerosol extinction coefficient at these wavelengths.
4. By refining the spectral variation of the aerosol extinction coefficient at all altitudes, an iterative procedure can be adopted in order to get a better estimation of the ozone extinction.

The data inversion provides aerosol extinction profiles versus altitude for the 850, 780 and 443 nm channels, as well as the ozone number density profile, from the tropopause up to the balloon altitude.

3.2 Airborne in situ Instrumentation for Direct Reading Measurements

A variety of instruments was designed for use on aircraft or balloons in order to measure physical and chemical properties of individual (or small populations of) aerosol particles and cloud elements. The major advantage of airborne instrumentation is that a number of different instruments can be simultaneously carried to a particular atmospheric region of scientific interest. The combined data of the different variables measured by all instruments allow detailed studies of atmospheric physics and chemistry to be conducted.

Optical methods determine the sizes, shapes, thermodynamical phase, and velocity vectors of individual aerosol particles. Additionally particle size distributions and integral properties like number concentrations, particulate surface area, or volume densities, cloud liquid water content, and distributions of particle populations in space are measured optically. By these means it is also possible to obtain the refractive index, the optical extinction, and the backscatter properties of the aerosol. Most in situ direct reading techniques employed on research aircraft, or balloons for aerosol-related measurements, require (1) the representative sampling of the aerosol. The instruments used subsequently for analyses fall into one of the categories: (2) OPCs (3) thermodynamic instruments, (4) other physical techniques, (5) methods requiring particle sampling on substrates and (6) mass spectrometric devices for chemical analyses.

For the deployment of in situ measuring aerosol instrumentation on a fast moving platform particular care has to be exercised concerning the sampling of the particles as it needs to be achieved isokinetically and nondestructively. Certain aspects concerning the sampling are described in section 3.2.1, and the instrument types (2) through (6) are subject of the subsequent sections.

3.2.1 Inlets and Nondestructive Aerosol Sampling

Sampling can be performed either directly in the unperturbed free air flow, or sample air containing the aerosol particles can be brought inside the aircraft for analysis through a specially designed inlet. When a parcel of sample air is brought from the free atmosphere to a measuring device through inlets and tubing, it is important to ensure that the sampled air arriving at the measuring instrument has the same characteristics as the undisturbed atmosphere, i.e. in the case of sampling aerosols, the number density of aerosol particles should remain unchanged, like the total surface area, the total volume, chemical particle composition and so on. Since a slowdown of the flow occurs in most inlet systems for aircraft the air is adiabatically heated and this may modify the thermodynamic state and the chemical composition of the particles. For the slower moving balloons this difficulty is not as severe. Additional problems arise as particle size dependent losses to the walls of the sampling inlet and its tubing occur and because impaction losses in curves and bends may alter the representativeness of the extracted air sample. Aerosols and gases may stick to, or are absorbed by, the walls of the tubes. If the tubes are not straight, then deposits of aerosol particles accumulate in bends, and particles from these deposits may be released back into the flow at later times. The vibrational environment of an airplane is especially prone to such ''memory effects''. Similar problems are associated with the sampling of reactive gases.

Additional requirements are imposed on the transport system in order to minimize the transit time, thus reducing the alteration introduced to the sample.

An inlet efficiency η is defined for a given particle size as Equation (7):

$$\eta = \eta_{\text{asp}}\eta_{\text{trans}} \tag{7}$$

where η_{asp} is the aspiration efficiency representing for each size the ratio of the concentration of particles entering the inlet to those present in the sampled volume. η_{trans} represents the fraction of particles of a given size which are transmitted by the inlet.

The inlet can collect air in two different configurations, the isoaxial one, in which the free air flow and the sample flow inside the inlet are aligned, and the anisoaxial one where the two flows occur in nonparallel directions. Particularly on fast moving airplanes, anisoaxial sampling is often connected to the generation of turbulent boundary layers inside the inlet system and boundary layer detachments which should be avoided because of increased turbulent wall deposition losses of small aerosol particles.

The two sampling configurations are further subdivided into three subcategories, which are:

- Isokinetic sampling, which occurs when the sample flow gas velocity (V) and the free ambient air velocity (V_{f}) are identical, or $V = V_{\text{f}}$, at the point where the inlet encounters the free air. (Further downstream inside the inlet the flow velocity can be decreased by diffusors etc.) Under conditions of perfect isokinetic sampling the relative ratios of the concentrations of small to large particles are the same in the sample as in the free atmosphere. For isokinetic sampling conditions the η_{asp} factor is equal to 1 for the axial sampling and below 1 in the case of the anisoaxial sampling. In both cases the transmission losses arise from the gravitational settling of larger particles and from the freestream turbulence inducing wall deposition losses.
- Superisokinetic sampling, which is the case $V > V_{\text{f}}$. Under these conditions the flow velocity V inside the inlet tip is larger than V_{f} outside which is usually due to some pump driving the sample flow through the inlet. The small particles may be able to follow the resulting curved tramlines around the inlet and enter, while the larger, more inert particles may fly by the inlet without entering. Therefore enhancement of the small particles could be introduced as an artefact to the sample. The aspiration efficiency is one or less, limited by its dependency on the ratio V_{f}/V for larger particles.
- Subisokinetic sampling, which is the case of $V_{\text{f}} > V$, leads to an enhancement of the larger particles over

the smaller ones in the sample. For the subisokinetic sampling the aspiration factor will be one or higher, up to a limit depending on the ratio V_f/V for the larger particles. Transmission losses are due again to the gravitational settling, turbulent wall deposition, and also to the inertial impaction of particles moving toward the inlet internal walls.

During airplane operation un-isoaxial sampling often occurs if the angle of attack changes, i.e. in situations when the plane flies in a curve, or when the air flow has a nonzero angle with respect to the aircraft's longitudinal axis. So-called ''shrouded inlets'' have been utilized[5] where the inlet performing the representative sampling itself is surrounded by a shroud in order to prevent turbulent boundary layer generation or detachments.

An extreme case of subisokinetic sampling is applied by the so-called CVI[51] for chemical and volatility analyses on the aerosol particles. This inlet permits the collection only of those particles larger than a certain equivalent size radius, while particles smaller than that cut-size, as well as the ambient air, are not permitted to enter the inlet. This is achieved by having the sonde produce a counterflow inside of opposite direction to the free atmospheric flow. Particles larger than the cut-size have sufficient inertia to ''overcome'' the barrier imposed by this counterflow and enter the inlet, while those particles which are too small are carried together with the air around the inlet tip.

In the Figure 6 the layout of a CVI set-up is shown.

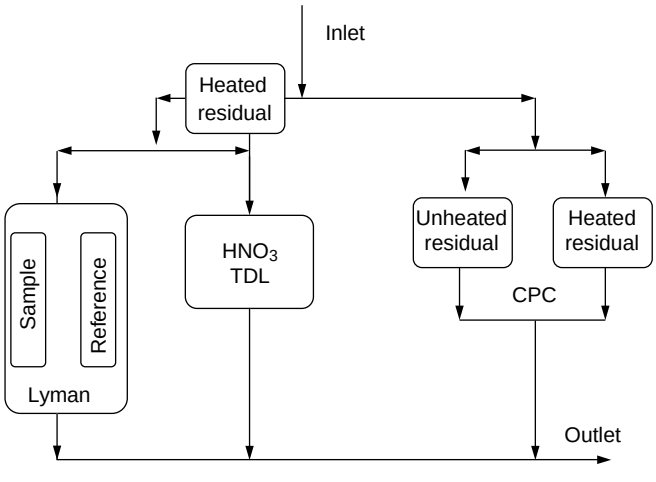

Figure 6 Schematic diagram of the CVI. The aersol sample air collected by the inlet is distributed to two different instruments. In the instrument shown in the left panel the air is heated and the gases evaporating from the particles are chemically analyzed by means of a Lyman-α and a TDL HNO_3 spectrometer. In the other instrument (right panel) volatile particles and nonvolatile residue particles are counted by the CPC.

If the instrument flow rate (the so-called return flow rate) is kept constant, then the flow rates define the distance L where the velocity relative to the sonde's axis is equal to zero in Equation (8):

$$L = \frac{F_1 - F_2}{F_1} X = \frac{F_3}{F_1} X \tag{8}$$

where F_1 is the return flow rate in $\mathrm{L\,min^{-1}}$, F_2 is the sample flow rate in $\mathrm{L\,min^{-1}}$, F_3 is the excess flow rate in $\mathrm{L\,min^{-1}}$, R is the sonde's radius in cm, D is the diameter of the inlet tube in cm, X is the length of the porous tube in cm and L is the distance of the tip of the sonde to the stagnation plane inside the sonde in cm.

The cut-size depends on the flow rate, on R and L.

Particles with inertia sufficient to permit deviation from the stream lines enter the CVI inlet and it is these particles crossing over the stagnation plane which are sampled. In other words, only those particles are collected which travel a distance larger than the so-called ''stop-distance'' (Equation 9):

$$d_s > L + g(R) \tag{9}$$

where $g(R)$ is a function of the radius of the sonde. d_s is used to evaluate the ''cut-size'' of a CVI inlet design, for the different input parameters. In the case of $g(R) \approx R$ it is possible to compute the dimensional range of the droplets which can be collected with an efficiency close to 50%. For particles larger than 1 μm this results in Equation (10):

$$d_s = V_\infty \times m_d \times B \times f \tag{10}$$

where V_∞ is the air flow velocity in $\mathrm{cm\,s^{-1}}$, m_d is the droplet mass $= (4/3)\pi r_d^3 \rho d$ in g, B is the droplet mechanical mobility $= 1/(6\pi\eta r_d)$ in St $\mathrm{g^{-1}}$ and f is a nondimensional correction factor for flows outside of the Stokes regime.

Thus, by a proper dimensioning of the CVI together with proper control of the flow rates, it is possible to sample particles larger than a predefined dimension.

If the air constituting the counter flow is heated and warmer than the ambient air, the CVI does not maintain the characteristics of the sampled particles unaltered. The particles can be forced to evaporate, and subsequent chemical analysis of the resulting gas yields information on the previous chemical composition of the sampled aerosol. The return flow (system flow) is formed by a known air mixture, which is carried on the aircraft in a pressurized tank. The additional trace species detected downstream in the flow then results from the aerosol components which evaporated after passage of the stagnation plane. In the instrumental package developed for the Russian M-55 stratospheric research aircraft, Geophysica, the subsequent chemical analyses

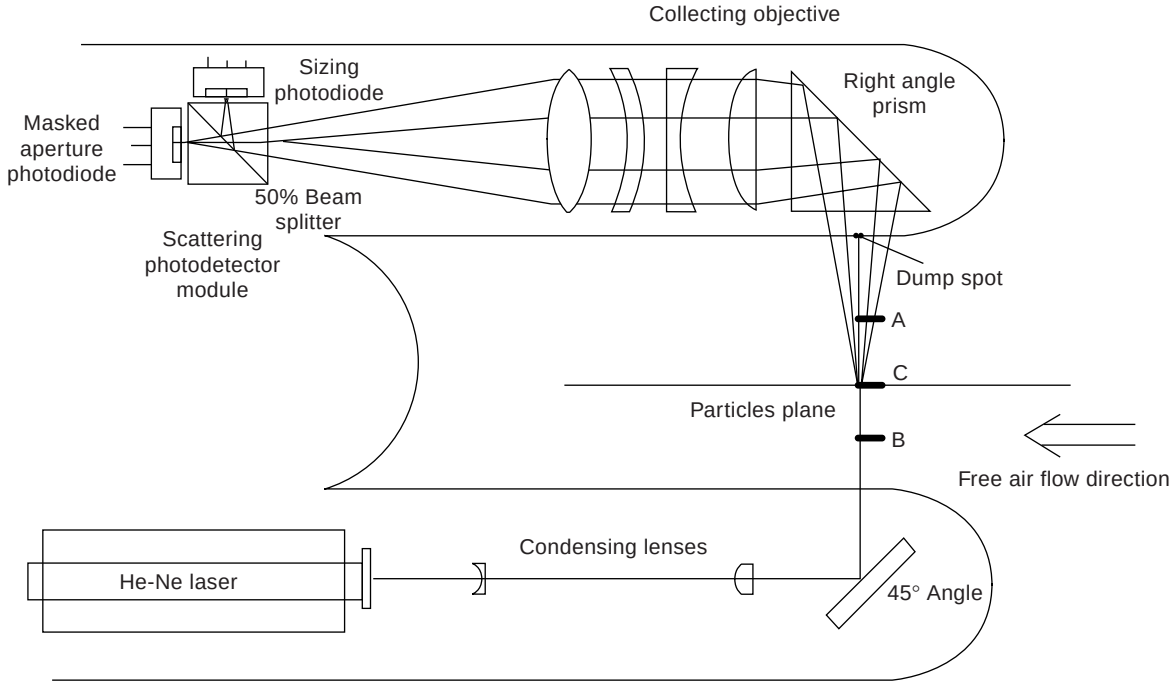

Figure 7 Schematic diagram of an airborne OPC (FSSP 300) where the particles are nondestructively detected directly in the free atmosphere. See text for details.

of the CVI-generated sample air are performed by a Lyman-α spectrometer,[52] which detects the amount of water vapor present in the sample, by a tunable diode laser absorption spectrometer (TDLAS) for the measurement of nitric acid vapor[53] as well as a variety of other trace gases, and by a two-channel CPC (see section 3.2.3) to count volatile and nonvolatile residual particles in the flow. This instrument package is designed specifically for the purpose of analyzing the chemical composition of PSC cloud particles.

In summary, the design of inlets for representative sampling of aerosols and gases poses a major task for the instrument developer, and sophisticated flow mechanical modeling including particle trajectory calculations, as well as cost intensive wind tunnel studies have to be performed in many cases.

3.2.2 Optical Particle Counters

In OPCs the air sample is drawn through an optically sensitive volume such that each individual particle is illuminated by a laser beam. The resulting light reflex scattered by the particle is detected by some optics and the particle size is inferred from the measured light intensity by means of Mie theory, or by the numerical T-matrix method in the case of aspherical particles. Analyzing a large number of particles in this way gives the particle number size distribution (the number of particles of a certain size per volume of sampled air), as well as particulate surface area and particulate volume distribution. From these primary deliverables of the instrument higher order variables can be derived such as total particle number density, surface area and volume densities, as well as the optical extinction, backscattering coefficients and so on. The OPCs for airborne use either adopt a sampling inlet and perform the particle size measurement inside the balloon or aircraft,[54,55] or they conduct the representative detection outside in the free flow.[56] Particle size diameters measured by such instruments typically range from 0.05–100 µm, but the whole range is not covered by one single instrument with high resolution. For example the FSSP-300 (forward scattering spectrometer probe model 300, manufactured by Particle Measuring Systems in Boulder, Colorado, USA) detects particles in the free flow outside the aircraft, which have sizes between 0.4–23 µm, and classifies them into 31 size bins, while the model FSSP-100 probe equipped with an optical system nearly identical to the FSSP-300 covers particle sizes up to 100 µm albeit less resolved with respect to size.

As seen in the scheme in Figure 7 the individual particles cross the collimated beam of a He-Ne laser in the particle plane scattering a light reflex into the detection optics. This optical system consists of the right angle prism, collecting lens system, a 50% beam splitter,

and two photodiodes as shown in the diagram. The laser beam itself ends on the blind dump spot on the prism such that only light scattered by the particles into an angular cone between 4–12 degrees in the forward direction is intercepted by the detection optics. The two photodiodes, one of which is partially covered by a mask, are necessary in order to ensure that the processing electronics accept only particles which cross the laser beam close enough to the particle plane (Position C in Figure 7). Particles crossing too far away (Positions A and B in Figure 7) are not detected under the scattering geometry given by the 4–12 degree angular cone. The electronics processing the signals from the photodiodes essentially consists of timing, gating, and pulse height analysis circuitry which counts each accepted particle into the 31 ''size'' bins (more precisely ''pulse height bins'') according to the intensity of the particle's scattered light reflex. As a result the data acquisition system records the number of the size bin for each particle. If proper calibration is available the upper and lower pulse height thresholds of each size bin can be associated with specific particle size diameters. This last step involves the application of solutions for the Maxwell equations (usually from Mie theory) under the boundary conditions given from the scattering geometry and particle and carrier gas properties. These theoretical calculations are subjected to many caveats, as the particle refractive index (real and imaginary parts) must be a priori known (or measured with a different instrument), the particle shape is assumed to be spherical, and the solutions imply various ambiguities over certain size ranges. In the case of aspherical particles Mie theory cannot be used to convert the measured scattered light cross-sections of the particles into the corresponding size, and other numerical methods (as the T-matrix method) have to be adopted.[57,58]

The most severe disadvantage is the unavailability of the particle refractive index in many experimental circumstances because the upper and lower pulse thresholds for each individual size bin are given from Mie theory and depend critically on the refractive index. Either reasonable assumptions regarding the refractive index need to be made, which is possible for the stratospheric aerosol types, or measurements from systems like MAL or MAS provide the information needed. Modified OPCs became available for use on aircraft which detect the light scattered by individual particles simultaneously under two or several different scattering directions. With these additional signals the size and the refractive index of the particle can be determined as for example by the MASP (Multi Angle Spectrometer Probe).[59]

The FSSP-300 was deployed on the NASA ER-2 during (Airborne Arctic Stratospheric Experiment) AASE I (1988/89), AASE II (1991/92), and SPADE (1993, Stratospheric Photochemistry Aerosol and Dynamics Experiment), as well as on the Russian M-55 Geophysica high altitude research aircraft during the Airborne Polar Experiment (1996/97) with important results concerning the freezing of stratospheric sulfuric acid droplets and on the evolution of the stratospheric aerosol in the aftermath of the 1991 Mount Pinatubo volcanic eruption. At the speed of 200 m s^{-1} the OPC acquires particle size distributions in 10 s averaging periods in order to keep errors due to count statistics small. Thus the system provides a spatial resolution of 2 km in the horizontal direction. Since 1994 the MASP has been utilized on the ER-2.

3.2.3 Thermodynamic Instruments

The smallest particles (condensation nuclei or Aitken particles) encountered in the atmosphere in the nanometer size range cannot be detected and sized easily by means of OPCs. Thermodynamic instruments like CPCs are widely used to measure the small particles. In these instruments an air sample flow containing the aerosol particles is at first ducted through a saturator unit, where the air is supersaturated with butanol or other suitable vapors. Afterwards the supersaturated air passes through a condenser unit where the air is cooled such that the supersaturated vapor condenses onto the aerosol particles. The particles then grow, attaining sizes large enough to make them amenable to optical detection. In the detection unit following downstream the grown particles are individually detected by means of a simple OPC similar to those described above. Thus the instrument delivers a time series of total particle concentrations. The most critical parameter of such instruments is the lower particle size detection limit which is influenced by many different instrument parameters and is determined experimentally for the most past.

As the particles are grown from subdetectable sizes to sizes above the optical detection threshold a size differentiation is not possible because the original particle size is masked during the condensation process. However, a sequence of simultaneously operating CPCs with different lower size detection limits can provide some coarse size resolution. Of course the CPCs will also count all particles which are larger than 0.05 μm and which could be detected directly by an OPC, but under atmospheric conditions the number of these ''coarse particles'' is small compared with the number of Aitken or fine particles and the error introduced is small. On the American high altitude research aircraft ER-2 condensation nuclei counters have been operated,[60,61] addressing the scientific issues of new particle formation by GPC in the tropopause region, the new particle formation and coagulation in the stratosphere after

the Mount Pinatubo eruption in June 1991, and the question how the global Junge aerosol layer is created and maintained.

3.2.4 Other Physical Techniques for Airborne Aerosol Measurements

For those atmospheric aerosol particles larger than 50 μm other optical techniques are adopted for aircraft implementation. These include imaging techniques as realized by one- or two-dimensional optical array probes (Particle Measuring Systems, Boulder, Colorado, USA). These linear photo diode arrays are exposed outside the aircraft's boundary layer perpendicular to the flow. Shadow cast images from individual cloud droplets and hydrometeors are projected on the photodiode detector by means of He-Ne or diode laser illumination. The number of diodes shadowed by a hydrometeor depends on the particle size and the instrument delivers size and concentration estimates. If the data from the one-dimensional array are read at a fast rate as some large hydrometeors passes through, the time series essentially provides a two-dimensional image of the hydrometeor. For particles larger than 50 μm, size and crystal habit can be resolved by such instruments.

Holographic methods have been used on aircraft in order to record in situ holograms of small volumes of cloudy air (approximately 1 L) containing aerosol particles cloud droplets and hydrometeors utilizing pulsed ruby lasers. Once recorded on the airplane a three-dimensional stationary reconstruction of the original scene can be obtained in the laboratory by means of these holograms. This image can be analyzed aided by a suitably designed reconstruction device with an automated or a semi-automated algorithm. With this methodology the sizes, shapes, distances between droplets, and velocity vectors of each individual hydrometeor can be retrieved[62,63] for all objects with size diameters larger than 6 μm.

Other methods usually applied in the laboratory have been modified for use on aircraft. Among these are differential mobility analyzers (DMA), diffusion batteries (DB), aerodynamic particle sizers (APS), Aitken nuclei counters (ANC), cloud nuclei counters (CNC) and ice nuclei counters (INC), nephelometers, optical extinction cells and a variety of others. Detailed descriptions of these instruments are beyond the scope of this section. Their physical principles and applicability for airborne atmospheric measurements are the subject of various chapters in Hinds[4] and Willeke and Baron.[5]

Most of the instruments described in section 3.2.2 through to section 3.2.4 need careful calibration or determination of instrument characteristics before they can be used for atmospheric in situ measurements. This involves, in many cases, sophisticated laboratory technology, for example for the generation of test aerosols with known size, optical properties and sometimes chemical composition. For the size calibration of OPCs, monodisperse aerosols (i.e. a population of particles with identical sizes) of a variety of sizes are needed covering the whole detection range of the OPC. This constitutes a difficult task especially for particles larger than 10 μm. For some instruments it is desirable to generate a known number concentration of test particles, which is also difficult. For CPCs it is necessary to determine experimentally the so-called ''efficiency curve'', which characterizes for each particle size the instrument's ability to detect accurately the number of particles per sample air volume. To perform these tasks laboratory devices are utilized like aerosol suspension atomizers, spraying devices, vibrating orifice generators, spinning disc atomizers, fluidized bed generators, tube furnace nanometer particle generators, DMAs, aerosol electrometers and others. The basic principles and characteristics of most of these techniques are described by Hinds[4] and Willeke and Baron[5] and details can be obtained from the references therein.

3.2.5 Methods Requiring Particle Sampling on Substrates

Most of the instruments described so far exploit physical principles to obtain information about a particular aerosol property. By these methods usually little or no information concerning the particle chemical composition is obtained, although this knowledge is key to many scientific issues involving atmospheric aerosol. There are several techniques which require the sampling of aerosol particles on matrices like filters, impactor slides, coated or pretreated surfaces, grids for electron microscopy and so on prior to chemical analysis. Impaction devices, filter samplers and other particle collection techniques are used on aircraft in order to sample and preserve aerosol particles, transport them into a laboratory, and subject them to detailed analyses by e.g. ion content, solubility and acidity measurements, by electron microscopy, LMMS, PIXE, radioactive radiation emission measurements, NA, micro Raman spectroscopy, and various mass spectrometric methods (see section 2.1 and Willeke and Baron).[5] Since the airborne part for these methods is mostly confined to the sampling of particles they are beyond the scope of this article and only the wire impactor is included here as example.

Wire impactors with chemically reactive coatings were deployed on aircraft in order to study the chemical composition of the particles. They are designed to be mounted perpendicularly to the free air flow outside the aircraft's boundary layer. Wire impactors consist of wires with dimensions from a few tens to a few hundreds of micrometers onto which atmospheric aerosol particles

impact and stick during exposure. After exposure the impactor is recovered and housed in a sealed container. In the laboratory the collected particles are analyzed with an electron microscope or by other techniques. Coatings which are chemically reactive have been deposited on the wires prior to the in situ exposure (see e.g. Pueschel et al.)[64] such that particles of a certain composition undergo reactions with the substances of the coating.

The impaction efficiency as a function of particle size and the sticking efficiency of the particles on the wire cause uncertainties in these measurements and therefore other instruments (e.g. OPCs) are commonly deployed simultaneously on the research aircraft.

3.2.6 Mass Spectrometric Devices for Chemical Analyses

In order to obtain a direct reading, in situ analyses of the chemical composition of single particles or small populations of them, airborne mass spectrometry has become available in conjunction with laser ablation,[65,66] a method which is similar to LMMS. The particle ambient air containing the particles is sampled through an inlet into the aircraft, where the air is ''stripped off'' the sample and the particles enter a vacuum (approximately 1.3×10^{-4} Pa) through what is called an ''aerodynamical lens'' or an differentially pumped inlet (DPI). In Figure 8 the following treatment of the aerosol particles is shown schematically. After exiting from the DPI into the vacuum, the particles are sized via laser scattering and then hit by the pulse of an excimer laser (4–5 mJ energy, 3 ns pulse duration) in the source region of a mass spectrometer. The energy deposited on the particle by the excimer laser suffices to ablated ions and molecule fractions from the aerosol particle surface or body. These ablated ions and charged molecule fractions enter into a time-of-flight mass spectrometer (TOFMS), where the mass spectrum of the ion cloud is measured. Subsequent analyses of the mass spectra of single particles or averages over groups of particles yield information concerning the chemical composition of the aerosol. The introduction of the particles to the vacuum requires only times in the order of milliseconds such that the particle chemical composition is not significantly altered during the sampling. The interpretation of the mass spectra obtained is difficult and may imply ambiguities. Since as many as 10 particles can be treated per second in the mass spectrometer and the spectrum of each individual particle is recorded, the amount of spectra and data acquired can become very large during one research flight. For these reasons automated methods for analyses and interpretation are required. The advantage of these individual particle analyses is that it is possible to resolve details of short events during a research flight (e.g. the crossing of aircraft exhaust emission plumes or contrails).

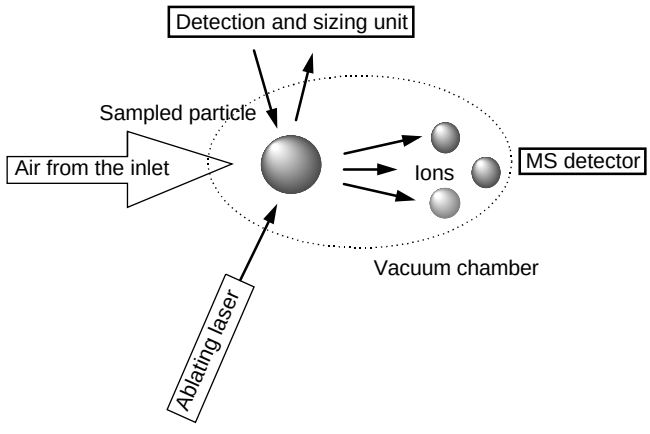

Figure 8 Basic principle of an aerosol composition mass spectrometer. The air is sampled through an inlet, the particles are conducted into a vacuum chamber by means of an aerodynamical lens and subsequently they are sized by a laser scattering technique. An excimer laser flash ablates ions and molecules from the particle surface. The ions and charged molecules enter the drift tube of a TOFMS where the mass spectrum of the ion cloud is measured.

ACKNOWLEDGMENTS

The authors want to express their gratitude to Dr. Vincenzo Santacesaria of IROE-CNR, of Florence, Italy, who has revised the whole manuscript. He has also given invaluable help with very stimulating and constructive discussions. The authors also want to thank Dr. Colette Brogniez, of the Institute of Atmospheric Optics of Lille, France, for providing figures and revising the part relative to the instruments BALLAD and BOCCAD and Prof. Giorgio Fiocco, of the University of Rome, Italy, for producing the ABLE LIDAR measurements.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
AASE	Airborne Arctic Stratospheric Experiment
ABLE	Airborne Lidar Experiment
ALEX	Aerosol Lidar Experiment
ANC	Aitken Nuclei Counters
APE-POLECAT	Airborne Polar Experiment–Polar Stratospheric Clouds, Leewaves, Chemistry, Aerosols and Transport
APS	Aerodynamic Particle Sizers
ASV	Anodic Stripping Voltammetry
BALLAD	Balloon Limb Aerosol Detection

BELINDA	Broadband Emission LIDAR with Narrowband Determination of Absorption
BIRA-IASB	Belgian Institute for Space Aeronomy
BOCCAD	Balloon Occultation for Aerosol Detection
BPC	Bulk-to-Particle Conversion
CCD	Charge-coupled Device
CNC	Cloud Nuclei Counters
CNRS	Centre National de la Recherche Scientifique
COARE	Coupled Ocean Atmospheric Research Experiment
CPC	Condensation Particle Counter
CVI	Counterflow Virtual Impactor
DB	Diffusion Batteries
DLR	Deutsches zentrum für Luft-und Raumfahrt
DMA	Differential Mobility Analyzers
DPI	Differentially Pumped Inlet
ELDORA	Electra Doppler Radar
ES	Emission Spectrometry
ESCA	Electron Spectroscopy for Chemical Analysis
FOV	Field of View
FSSP	Forward Scattering Spectrometer Probe
GPC	Gas-to-particle Conversion
INC	Ice Nuclei Counters
IPCC	Intergovernmental Panel for Climate Change
IR	Infrared
LEANDRE	Lidar Embarqué Pour L'étude des Aérosols et des Nuages, de Interaction Dynamique-Rayonnement et du cycle de L'Eau
LIDAR	Light Detection and Ranging
LITE	Laser in Space Technology Experiment
LMMS	Laser Microprobe Mass Spectrometry
LOA	Laboratoire d' Optique Atmospherique
MAL	Miniaturized Airborne Lidar
MAS	Multiwavelength Aerosol Laser Scatterometer
MASP	Multi Angle Spectrometer Probe
MIRACLE	Microwave Radar Cloud Layer Exploration
NA	Neutron Activation
NAD	Nitric Acid Dihydrate
NAM	Nitric Acid Monohydrate
NASA	National Air and Space Administration
NAT	Nitric Acid Trihydrate
NCAR	National Center for Atmospheric Research
OLEX	Ozone Lidar Experiment
OPC	Optical Particle Counter
PBL	Planetary Boundary Layer
PIXE	Proton (or Particle)-induced X-ray Emission
POAM II	Polar Ozone and Aerosol Measurements II
PSC	Polar Stratospheric Clouds
RADAR	Radio Detection and Ranging
RADIBAL	Radiometer Balloon
SAGE	Stratospheric Aerosol and Gas Experiment
SAT	Sulfuric Acid Tetrahydrate
SESAME	Second European Stratospheric Arctic and Middle Latitude Experiment
SPADE	Stratospheric Photochemistry Aerosol and Dynamics Experiment
SSA	Stratospheric Sulfate Aerosols
STS	Supercooled Ternary Solutions
TDLAS	Tunable Diode Laser Absorption Spectrometer
TOFMS	Time-of-flight Mass Spectrometer
TOGA	Tropical Ocean Global Atmosphere
UV	Ultraviolet
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Infrared LIDAR Applications in Atmospheric Monitoring • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Electrochemical Sensors for Field Measurements of Gases and Vapors

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Diffraction in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting

REFERENCES

1. J. Wallace, P. Hobbs, 'Atmospheric Sciences, an Introductory Survey', *Atmospheric Aerosols and Cloud Microphysical Processes*, Academic Press, 143–209, 1987.
2. Intergovernmental Panel on Climate Change (IPCC) Climate Change 1995, 'The Science of Climate Change, Contribution of Working Group I to the second assessment report of the Intergovernmental panel on Climate Change', WMO-UNEP, Technical Summary, Cambridge University Press, Cambridge, 1996.
3. C.E. Junge, J.E Manson, 'Stratospheric Aerosols Studies', *J. Geophys. Res.*, **66**, 2163–2182 (1961).
4. W.C. Hinds, *Aerosol Technology, Properties, Behavior, and Measurement of Airborne Particles*, Wiley Interscience, New York, 1982.
5. K. Willeke, P.A. Baron, *Aerosol Measurement, Principles, Techniques, and Applications*, Van Nostrand Reinhold, New York, 1993.
6. H. Jäger, D. Hoffmann, 'Midatitude Lidar Backscatter to Mass, Area, and Extinction Conversion Model Based on in situ Aerosol Measurements from 1980 to 1987', *Appl. Optics*, **30**, 127–138 (1991).
7. H. Jäger, T. Deshler, F. Homburg, V. Freudenthaler, 'Five Years of Lidar Observation of the Pinatubo Eruption Cloud', Proceedings 18th ILRC, Berlin, 22–26 July 1996, A. Ansmann, R. Neuber, P. Rairoux, U. Wandinger, Springer-Verlag, Berlin, 485–488, 1996.
8. T.E. Graedel, P.J. Crutzen, *Atmospheric Change: An Earth System Perspective*, W.H. Freeman, New York, 1993.
9. 'European Research in the Stratosphere–The Contribution of EASOE and SESAME to our Current Understanding of the Ozone Layer', European Commission, Directorate General XII, EUR 16986 en, 1997.
10. World Meteorological Organization (WMO), 'Scientific Assessment of Ozone Depletion: 1994', Global Ozone Research and Monitoring Project, Report No. 37, Geneva, 1995.
11. P. Fabian, *Atmosphäre und Umwelt*, Springer Verlag, New York, 91–97, 1987.
12. H.R. Pruppacher, J.D. Klett, *Microphysics of Clouds and Precipitation*, 2nd edition, Kluwer Academic, Dordrecht, 1997.
13. P. Warneck, *The Chemistry of Natural Atmospheres*, Academic Press, San Diego, 1988.
14. S. Borrmann, R. Jaenicke, 'Wind Tunnel Experiments on the Resuspension of Sub-micrometer Sandparticles from a Sand Surface', *Atmos. Environ.*, **21**(9), 1891–1898 (1987).
15. 'Pollution from Aircraft Emissions in the North Atlantic Flight Corridor (POLINAT)', Air Pollution Research Report 58, EUR 16978 EN, ed. U. Schumann, 1997.
16. P. Fabian, B. Karcher, 'The Impact of Aviation upon the Atmosphere', *Phys. Chem. Earth*, **22**, 503–598 (1997).
17. G.P. Brasseur, R.A. Cox, D. Hauglustaine, I. Isaksen, J. Lelieveld, D.H. Lister, R. Sausen, U. Schumann, A. Wahner, P. Wiesen, 'European Scientific Assessment of the Atmospheric Effects of Aircraft Emission', *Atmos. Environ.*, **32**, 2329–2418 (1998).
18. L. Stefanutti, M. Morandi, M. Del Guasta, S. Godin, G. Megie, J. Brechet, J. Piquard, 'Polar Stratospheric Cloud Observations Over the Antarctic Continent at Dumont d' Urville', *J. Geophys. Res.*, **96**, 12 975–12 987 (1991).
19. R.P. Turco, R.C. Whitten, O.B. Toon, 'Stratospheric Aerosols: Observations and Theory', *Rev. Geophys. Space Phys.*, **20**, 233–279 (1982).
20. D.J. Hofmann, 'Increase in Stratospheric Background Sulfuric Acid Aerosol Mass in the Past 10 years', *Science*, **248**, 996–1000 (1990).
21. K. Drdla, 'Applications of a Model of Polar Stratospheric Clouds and Heterogeneous Chemistry', PhD Thesis, University of California at Los Angeles, 1996.
22. D.F. Blake, K. Kato, 'Latitudinal Distribution of Black Carbon Soot in the Upper Troposphere and Lower Stratosphere', *J. Geophys. Res.*, **100**, 7195–7202 (1995).
23. R.F. Pueschel, 'Stratospheric Aerosols-formation; Properties, Effects', *J. Aerosol Sci.*, **27**, 383–402 (1996).
24. C. Weitkamp, O. Danne, H. Flint, W. Lahmann, M. Quante, E. Raschke, F.E. Theopold, 'Lidars for Cloud Measurements at GKSS and First Radar–Lidar Comparisons', Proceedings of the Workshop on Synergy of Active Instruments in the Earth Radiation Mission, eds. M. Quante, J.P.V. Poiares Baptista, E. Raschke, *GKSS/98/E/10*, GKSS, Geesthacht, Germany, 1998.
25. S.E. Yuter, R.A. Houze, B.F. Smull, F.D. Marks, J.R. Daugherty, S.R. Brodzik, 'TOGA COARE Airgraft Mission Summary Images: An Electronic Atlas', *Bull. Am. Met. Soc.*, **76**, 319–333 (1995).
26. L. Stefanutti, F. Castagnoli, M. Del Guasta, M. Morandi, V.M. Sacco, V. Venturi, L. Zuccagnoli, J. Kolenda, H. Kneipp, P. Rairoux, B. Stein, D. Weidauer, J.P. Wolf, 'A Four-wavelength Depolarization Lidar for Polar Stratospheric Cloud Monitoring', *Appl. Phys. B*, **55**, 13–17 (1992).
27. L. Stefanutti, F. Castagnoli, M. Del Guasta, M. Morandi, V.M. Sacco, L. Zuccagnoli, S. Godin, G. Megie, J. Porteneuve, 'The Antarctic Ozone Lidar System', *Appl. Phys. B*, **55**, 3–12 (1992).
28. B. Stein, M. Del Guasta, J. Kolenda, M. Morandi, P. Rairoux, L. Stefanutti, J.P. Wolf, 'Stratospheric Aerosol Size Distribution from Multispectral Lidar Measurements at Sodankylä During EASOE', *Geophys. Res. Lett.*, **21**, 1311–1314 (1994).
29. M. Del Guasta, M. Morandi, L. Stefanutti, B. Stein, J.P. Wolf, 'Derivation of Mt. Pinatubo Stratospheric Aerosol Mean Size Distribution by Means of a Multiwavelength Lidar', *Appl. Optics.*, **33**, 5690–5697 (1994).
30. C.F. Bohren, S.B. Singham, 'Backscattering by NonSpherical Particles: A Review of Methods and Suggested

New Approaches', *J. Geophys. Res.*, **96**, 5269–5277 (1991).
31. C. Flood, H. Grassl, G. Megie, Backscatter LIDAR, Report of the 'ATLID' Consultancy Group, Vol. ESA SP-1121, May 1990.
32. Proceedings of the ELITE Workshop 9–10 November 1995, eds. P. Fletcher, F. Lodge, E. Attema, Vol. ESA WWP-107, March, 1996.
33. E.V. Browell, L. Poole, M.P. McCormick, S. Ismail, C.F. Butler, S.A. Kooi, M.M Szedlmayer, R. Jones, A. Krueger, A. Tuck, 'Large Scale Variations in Ozone and Polar Stratospheric Clouds Measured with Airborne Lidar during Formation of the 1987 Ozone Hole over Antartica', Polar Ozone Workshop, NASA C.P. 10014, 61–64, 1988.
34. E.V. Browell, 'Differential Absorption Lidar Detection of Ozone in the Troposphere and Lower Stratosphere', *Optoelectronics for Environmental Science*, eds. S. Martellucci, A.N. Chester, Plenum Press, New York, 77–99, 1990.
35. M.P. McCormick, G.S. Kent, W.H. Hunt, M.T. Osborne, L.R. Poole, M.C. Pitts, 'Arctic Polar Stratospheric Clouds Observed by Airborne Lidar', *Geophys. Res. Lett.*, **17**, 381–383 (1990).
36. M.P. McCormick, 'Stratospheric Impact of the Eruption of Pinatubo', V Topical Meeting – Optical Remote Sensing of the Atmosphere, Williamsburg, Virginia, USA, Nov. 1991.
37. S. Spinhirne, J.F. Cavanough, J.L. Bufton, 'Aerosol and Cloud Backscatter at 1.06 1.54, and 0.53 mm by Airborne Hard-target-calibráted Nd : YACY Methane Raman Lidar', *Appl. Optics.*, **36**, 3475–3490 (1997).
38. J.D. Spinhirne, W.D. Hart, 'Lidar Study of Storm Tops', XIII ILRC, NASA CP-2431, 230–232, 1986.
39. J.M. Vaughan, D.W. Brown, C. Nash, S.B. Alejandro, G.G. Koening, 'Atlantic Atmospheric Aerosol Studies: 2. Compendium of Airborne Backscatter Measurements at 10.6 mm', *J. Geophys. Res.*, **100**, 1043–1065 (1995).
40. M. Wirth, G. Ehret, P. Moerl, W. Renger, 'Two Dimensional Stratospheric Aerosol Distributions During EASOE', *Geophys. Res. Lett.*, **21**, 1287–1290 (1994).
41. M. Wirth, W. Renger, 'Evidence of Large Scale Ozone Depletion within the Arctic Polar Vortex 94/95 Based on Airborne Lidar Measurements', *Geophys. Res. Lett.*, **23**, 813–816 (1996).
42. G. Fiocco, P. Calisse, M. Cacciani, S. Casadio, G. Pace, D. Fua, 'ABLE: Development of an Airborne Lidar', *JOAT*, **16**, 1336–1343 (1999).
43. R. Matthey, V. Mitev, 'Miniature Airborne Lidar for M-55 'Geophysica' High Altitude Research Aeroplane', personal communication.
44. D.J. Hofmann, J.M. Rosen, 'Balloon-borne Observations of Stratospheric Aerosols and Condensation Nuclei During the year following the Mt. St. Helens Eruption', *J. Geophys. Res.*, **87**, 11 039–11 061 (1982).
45. A. Adriani, F. Cairo, M. Viterbini, S. Mandolini, G. Di Donfrancesco, 'Multiwavelength Aerosol Scatterometer for Airborne Experiments to Study the Stratospheric Particle Optical Properties', Special issue of *JAOT* on APE, **16**, 1328–1335 (1999).
46. M. Herman, C. Brogniez, D. Ramon, J. Lenoble, P.C. Simon, 'Characterization of the Stratospheric Aerosols from Balloon-borne Measurements During the SESAME Campaign', personal communication.
47. D. Ramon, 'Detection des Aerosols Stratospheriques Par mesure ballon du rayonnement diffus au limbe: mise au point d'un nouvel instrument et premiers resultats'. PhD Thesis, Universite de Lille 1, 1995.
48. C. Brogniez, Ramananhérisoa, J. Lenoble, J.-P. Pommereau, C. Pillips, L. Denis, 'Balloon Measurements of Aerosol Extinction, Ozone and NO_2 Profiles for the Validation of GOME Algorithm and Products', Proceedings Workshop on the GOME Geophysical Validation Campaign, ESA WPP-108, 179–183, 1996.
49. W.P. Chu, M.P. McCormick, J. Lenoble, C. Brogniez, P. Pruvost, 'SAGE II Inversion Algorithm', *J. Geophys. Res.*, **94**, 8339–8351 (1989).
50. J.D. Lumpe, R.M. Bevilacqua, K.W. Hoppel, S.S. Krigman, D.L. Kriebel, C.E. Randall, D.W. Rusch, C. Brogniez, R. Ramananaherisoa, E.P. Shettle, J.J. Olivero, J. Lenoble, P. Pruvost, 'POAM II Retrieval Algorithm and Error Analysis', *J. Geophys. Res.*, **102**, 23 593–23 614 (1997).
51. K.J. Noone, J.A. Ogren, J. Heinzenberger, R.J. Charson, D.S. Covert, 'Design and Calibration of a Counterflow Virtual Impactor for Sampling Atmospheric Fog and Cloud Droplets', *Aerosol Sci. Technol.*, **8**, 235–244 (1988).
52. J. Stroem, J. Heinzenberger, K.J. Noone, K.B. Noone, J.A. Ogren, F. Albers, M. Quante, 'Small Crystals in Cirriform Clouds: A Case Study of the Residue Size Distribution, Cloud Water Content and Related Cloud Properties', *J. Atmos. Res.*, **32**(1–4), 125–141 (1994).
53. G. Toci, P. Mazzinghi, M. Vannini, 'A Diode Laser Spectrometer for the in situ Measurements of the HNO_3 Content of PSC', *JAOT*, special issue on APE, **16**, 1294–1301 (1999).
54. D.J. Hofmann, 'Twenty Years of Balloon-borne Tropospheric Aerosol Measurements at Laramie, Wyoming', *J. Geophys. Res.*, **98**, 12 753–12 766 (1993).
55. H.H. Jonsson, J.C. Wilson, C.A. Brock, R.K. Knollenberg, R. Newton, J.E. Dye, D. Baumgardner, S. Borrmann, G.V. Ferry, R. Pueschel, D.C. Woods, M.C. Pitts, 'Performance of the Focused Cavity Aerool Spectrometer for Measurements in the Stratosphere of Particle Size in the 0.06–2 μm Diameter Range', *J. Atmos. Oceanic Technol.*, **12**(1), 115–129 (1995).
56. D. Baumgardner, J.E. Dye, B.W. Gandrud, R.G. Knollenberg, 'Interpretation of Measurements Made by the

Forward Scattering Spectrometer Probe (FSSP-300) During the Airborne Arctic Stratospheric Expedition', *J. Geophys. Res*, **97**, 8035–8046 (1992).
57. C.F. Bohren, S.B. Singham, 'Backscattering by Nonspherical Particles: A Review of Methods and Suggested New Approaches', *J. Geophys. Res.*, **96**, 5269–5277 (1991).
58. S. Borrmann, S. Solomon, J.E. Dye, B. Luo, 'The Potential of Cirrus Clouds for Heterogeneous Chlorine Activation', *Geophys. Res. Lett.*, **16**, 2133–2136 (1996).
59. D. Baumgardner, J.E. Dye, B. Gandrud, K. Barr, K. Kelly, K.R. Chan, 'Refractive Indices of Aerosols in the Upper Troposphere and Lower Stratosphere', *Geophys. Res. Lett.*, **23**, 749–752 (1996).
60. J.C. Wilson, J.H. Huyn, E.D. Blackshear, 'The Function and Response of an Improved Stratospheric Condensation Nucleus Counter', *J. Geophys. Res.*, **88**, 6781–6785 (1983).
61. C.A. Brock, P. Hamill, J.C. Wilson, H.H. Jonsson, K.R. Chan, 'Particle Formation in the Upper Tropical Troposphere: A Source of Nuclei for the Stratospheric Aerosol', *Science*, **270**, 1650–1652 (1995).
62. S. Borrmann, R. Jaenicke, 'Application of Microholography for Ground Based in situ Measurements in Stratus Cloud Layers: A Case Study', *J. Atmos. Oceanic Technol.*, **10**, 277–293 (1993).
63. P.R.A. Brown, 'Use of Holography for Airborne Cloud Physics Experiments', *J. Atmos Oceanic Technol.*, **6**, 293–306 (1989).
64. R.F. Pueschel, K.G. Snetsinger, J. Goodman, O.B. Toon, G.V. Ferry, V.R. Overbeck, J.M. Livingtson, S. Verna, W. Fong, W.L. Starr, R.K. Chan, 'Condensed Nitrate, Sulfate, and Chloride in Antarctic Stratospheric Aerosols', *J. Geophys. Res.*, **94**, 11 271–11 284 (1989).
65. D.M. Murphy, D.S. Thomson, A.M. Middlebrook, 'Bromine, Iodine, and Chlorine in Single Aerosol Particles at Cape Grim', *Geophys. Res. Lett.*, **24**, 3917–3200 (1997).
66. Th. Peter, 'Airborne Particle Analysis for Climate Studies', *Science*, **273**, 1352–1353 (1996).

Automotive Emissions Analysis with Spectroscopic Techniques

Rainer Hirschberger
ISAS Institut für Spektrochemie und angewandte Spektroskopie, Dortmund, Germany

1 Introduction 1
2 Physical Fundamentals of Ultraviolet, Infrared and the Raman Effect 2
2.1 Movement of Electrons – Ultraviolet Spectroscopy 2
2.2 Movement of Atomic Nuclei – Infrared Spectroscopy 2
2.3 Combined Electron and Atomic Nucleus Movement – Raman Spectroscopy 2
2.4 Discussion 2
3 Techniques 4
3.1 Traditional Optical Techniques 4
3.2 Infrared Absorption-based Modern Techniques 5
3.3 Exhaust Gas Measurement using Raman-based Analytical Systems 12
4 Potential and Comparison of Different Techniques 21
Acknowledgments 21
Abbreviations and Acronyms 21
Related Articles 22
References 22

Quantitative analysis of automotive emissions gives detailed information about a motor's combustion behavior and gives information about the contribution from traffic to air pollution. Following the introduction, which discusses the problems of exhaust gas analysis, section 2 covers the basics of the physical effects, that the different spectroscopic systems are based on. These physical effects are discussed especially with respect to gas analysis.

Section 3 describes the functioning of different spectroscopic systems, the measuring procedures and different examples of exhaust gas analysis. Simple examples are used to illustrate the functioning of Fourier transform infrared (FTIR) and diode laser (DL) systems. The construction of a completely home-made Raman polychromator is described in detail. Specific problem areas are pointed out. Selected examples of exhaust gas analysis show typical applications of the particular methods.

Section 4 discusses the present and future potential of the systems described.

1 INTRODUCTION

Although emissions from gasoline-powered vehicles have been reduced significantly since the 1960s, motor vehicles are one of the major sources of air pollutants such as volatile organic compounds (VOC), nitrogen oxides (NO_x), CO, benzene and polycyclic aromatic hydrocarbons (PAH). While VOC and NO_x are well known to react with sunlight, forming harmful ozone, CO, benzene and PAH are toxic or carcinogenic. The contribution to air pollution by motor vehicles in Germany in 1992 was 70% for NO_x, more than 50% for VOC and 90% for benzene.[1] Further reductions in all these components is the goal of gasoline motor development, catalytic conversion systems and improved fuels. These developments all require to be supported by effective exhaust gas-analyzing systems.

Well-established and legally certified analyzing methods have been in operation for many years. The main problem is that different systems are required for the quantitative analysis of different gas components, for example:

- total hydrocarbons (THC) by FID (flame ionization detector);
- CO and CO_2 by NDIR (nondispersive infrared spectroscopy);
- total NO_x by chemiluminescence;
- C_1-C_8 aldehydes and ketones by liquid chromatography (LC);
- C_1-C_{12} hydrocarbons, ethers and alcohols by gas chromatography (GC).

Another consideration is that chromatographic systems are not suitable for on-line measurements. In the early 1980s strategy papers for new analyzing systems suggested the following requirements for new systems:[2]

- only one analytical method;
- the possibility of in situ or at least on-line measurement;
- simultaneous detection of, ideally, all compounds;
- easy calibration and easy handling.

Looking at these requirements it is clear that only a few spectroscopic methods can approximate these

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

conditions. In the following years systems for automotive emissions analysis were developed, based in the main on infrared (IR) and Raman spectroscopies. At that time mass spectroscopic systems were of no interest because mass spectrometers could not be coupled directly to the exhaust gas stream. This situation was changed by the development of atmospheric pressure ionization (API) sources, so that nowadays such systems are good alternatives or are used as complementary systems.

2 PHYSICAL FUNDAMENTALS OF ULTRAVIOLET, INFRARED AND THE RAMAN EFFECT

The basis of these optical spectroscopies is the interaction of objects of investigation with electromagnetic radiation. This interaction is possible only when the objects (molecules) are able to work as an antenna for electromagnetic radiation. That is, following irradiation, charges move within these objects. This causes a dipole to change its magnitude, its direction or both.

2.1 Movement of Electrons – Ultraviolet Spectroscopy

The electrons in a molecule can be moved from their atomic positions over the complete molecular framework. For a diatomic molecule this movement will be along the molecular axis. The eigenfrequencies of this movement are very high, due to the small mass of the electrons and the small dimensions of the molecule, and correspond to the frequencies of the ultraviolet (UV) and visible (VIS) spectral regions. Quantum physics describes this process as the excitation of electrons from their electronic ground state to higher electronic levels.

2.2 Movement of Atomic Nuclei – Infrared Spectroscopy

In molecular vibrations atoms, not electrons, move. Consider the case of a diatomic molecule involving two different atoms. The electronic charge is asymmetric and the molecule forms an electric dipole along its axis, which can change its magnitude with the rhythm of incident electromagnetic radiation. The eigenfrequencies of these vibrations, where the relatively large atomic nuclei are moved, are about two orders of magnitude smaller than for electronic movement and correspond to the frequencies of IR radiation. This process is called an excitation of a molecule from a vibrational ground state to higher vibrational levels.

2.3 Combined Electron and Atomic Nucleus Movement – Raman Spectroscopy

When a diatomic molecule consists of two identical atoms, the electric charge is symmetric over the molecule, even while vibrating. Due to the absence of a dipole the molecule cannot absorb IR radiation, even if the frequency of irradiation corresponds to its vibrational frequency. In this case the vibrational energy has to be provided indirectly. A solution to this problem was first suggested by Raman and Krishnan[3] as follows.

Electromagnetic radiation from the UV and VIS spectral regions can enforce electronic vibrations, so that the electrons will partially move over the molecule. This movement will loosen and stretch the bond, and the atoms will start vibrating. In this way the radiation indirectly loses energy to nuclear movement, namely molecular vibrations, while exciting electronic movement. After this interaction the radiation will have less energy – its frequency is lowered by the molecular vibration frequency. The Raman process is not an absorption, but is based on an inelastic scattering process. Quantum physics describes this process as the excitation of electrons from the ground state to a virtual electronic state, resulting in excitation of higher vibrational states.

2.4 Discussion

As described above, a change in dipole moment during a molecular vibration can be observed by absorbing IR radiation. If there is no change in dipole moment, the vibration can be excited indirectly by UV/VIS radiation, resulting in Raman scattering.

Any molecule has $3N$ degrees of freedom for movement, where N is number of atoms. Subtracting three degrees for translation and three degrees for rotation (or two for a linear molecule), gives the number of vibrational modes. Therefore the linear CO_2 molecule has $3N - 5 = 4$ vibration modes (Figure 1).

In the symmetric vibration mode, both bonds are stretched simultaneously and the vibration is a Raman-active one. Both dipole moments, left and right in the molecule, are equal in magnitude and in opposite directions. The resulting dipole moment for the whole molecule is therefore zero and this vibration is not IR active.

In the asymmetric vibration mode, the bonds are not stretched simultaneously and the vibration is not Raman active. However, one dipole will increase as the other gets smaller, so that there is a resulting dipole moment (changing its direction in the rhythm of vibration) and this vibration is IR active.

The deformation vibrations (Figure 1) are both equivalent, resulting in identical frequencies and are called

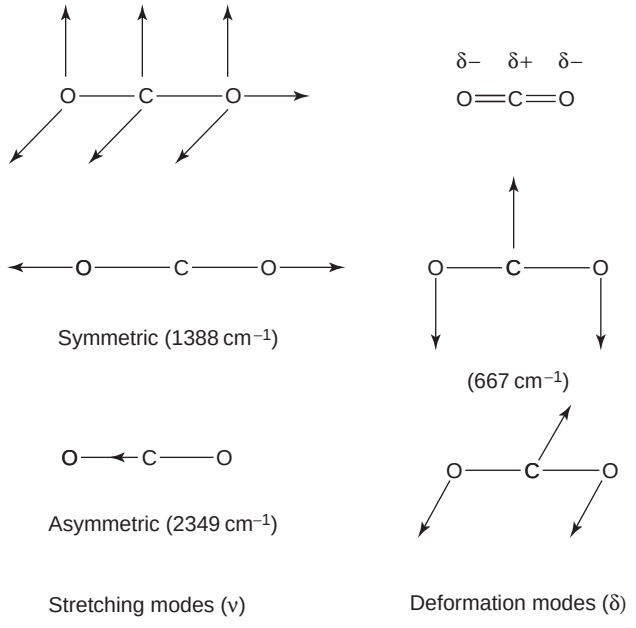

Figure 1 Vibrational modes of the CO_2 molecule.

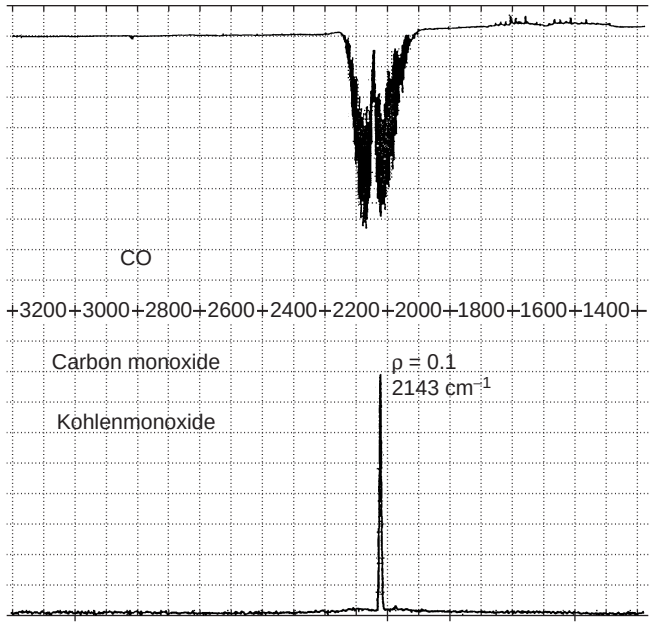

Figure 2 IR (top) and Raman (bottom) spectra for carbon monoxide.

doubly degenerate. These vibrations are Raman inactive but IR active.

Using these selection rules now for the CO molecule shows that the only possible vibration is symmetric and that the molecular dipole is changed. So this vibration is both Raman active and IR active. However, the IR and Raman spectra (Figure 2) are quite different. In the Raman spectrum the vibrational excitation is represented by a small signal at 2143 cm^{-1}, whereas the IR spectrum shows two broad signals, below and above that frequency. This is because there are additional selection rules, due to the fact that molecular vibration is always accompanied by rotation. Again, as in the case of vibration, these rules are different for IR and Raman excitation.

The lower part of Figure 3 shows the IR absorption spectrum of CO, but now with much higher resolution than in Figure 2. It can be seen that the broad bands of CO (Figure 2) consist of individual small lines. The upper part of Figure 3 gives the explanation.

If a molecule is excited from the vibrational ground level $v'' = 0$ to the upper level $v' = 1$, there is also a change in molecular rotation. In this case the rotational quantum number J will change exactly by $\Delta J = \pm 1$. Therefore, the different rotational states of the vibrational ground state will result in different rotational states of the excited vibrational state, each changing to the next higher (R-branch) or next lower (P-branch) state. Against this, when exciting vibrational states by Raman scattering, the rotational number J will not change, thus $\Delta J = 0$. The Raman spectrum of CO therefore only consists of

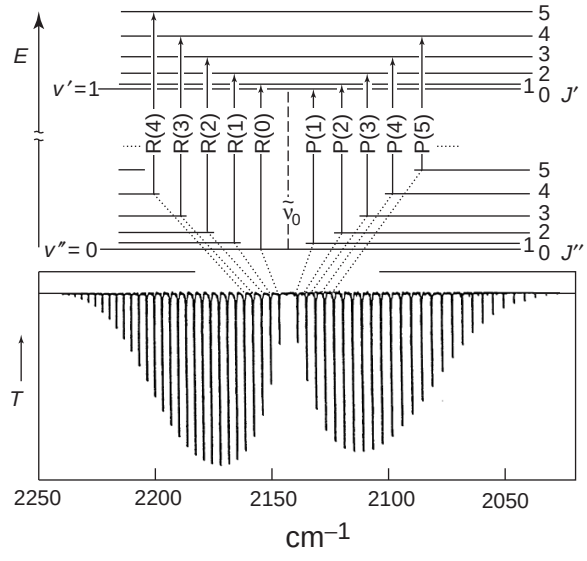

Figure 3 Rotational transitions of CO. (Reproduced by permission of Wiley-VCH Weinheim Germany, taken from B. Schrader, *IR and Raman Spectroscopy* (1995).)

the small band (the Q-branch) located exactly between the two IR absorption bands, as shown in Figure 2. This principal difference between IR and Raman excitation is very important for multicomponent gas analysis, as this leads to a lot of broad bands (R- and P-branches) in the IR spectrum and it is often difficult to find spectral regions where bands from different components will not interfere. The Raman spectrum consists of small bands

(Q-branch) and normally there is no interference. This can easily be seen by comparing the IR and Raman spectra for the two components CO_2 and H_2O, which are always present in an exhaust gas in high concentrations. With these components the IR spectrum in wide regions is dominated by the broad signals from these components and, especially in the case of low resolution, these signals can obscure those of other components. In most cases this problem can only be solved using high resolution and special evaluation. The Raman spectrum for these components shows only two small signals for CO_2 (the reason that there are two signals, even though there is only one Raman-active vibration, is the so-called Fermi resonance[4]) as can be seen in the Raman spectra from exhaust gases in section 3.3.2 and a sharp signal for H_2O at $3652\,cm^{-1}$, which is outside the spectral region of interest. For completeness and correctness it should be mentioned here that, for some molecules such as N_2 and O_2, during vibrational excitation by Raman scattering different rotational levels can be excited resulting in R- and P-branches (see Figures 24 and 25), but the signals of these transitions are orders of magnitude smaller than for the Q-branch.

Considering the results from IR absorption, it is evident that UV spectra are much more complex, as the transition from the electronic ground state to higher levels is accompanied by vibrational and rotational transitions.

The above results can be summarized with respect to a multicomponent gas analysis:

- IR absorption techniques can be used for multicomponent analysis by consideration of interference of absorption bands from different molecular species, requiring high-resolution instruments. Analysis of homonuclear components such as H_2, O_2, and N_2 is not possible, as these molecules do not have a dipole moment.
- UV absorption, except in a few cases (e.g. NO, NO_2, SO_2), leads to spectra having insufficient structure to be useful for selective concentration analysis. Small-band laser systems (frequency-doubled tunable dye laser, optoparametric oscillators) or high-resolution spectrometers with echelle gratings have been developed for the analysis of air pollution. The development of UV absorption-based exhaust gas analyzers may be encouraged when spectral databases, as in the case of IR absorption, become available.
- Raman spectroscopy has the advantage that each molecule has at least one Raman-active vibration mode, so that in principle all exhaust gas species can be analyzed. The problem of the poor quantum efficiency of the Raman effect had to be solved, as described below.

3 TECHNIQUES

As previously mentioned most optical techniques are based on IR or Raman effects. The different types of IR and Raman spectrometers are discussed in sections 3.2 and 3.3. But first we will have a look at traditional systems, which are in widespread use when checking the emission of motor cars with respect to CO and NO_x.

3.1 Traditional Optical Techniques

For applications where only one or few compounds have to be analyzed, nondispersive (ND) systems can be used. This means that the radiation passing through the sample is not decomposed by dispersing optical components such as prisms or gratings. In all cases a relative broad spectral region, which is selected by filters or selective detectors, is used for the measurement. The systems are in widespread use for gas analysis because of their simple construction, which leads to a high optical throughput and so to high sensitivity. They are also compact in design, easy to handle and are not sensitive to working conditions, such as in an automotive workshop. ND systems are based on IR and UV/VIS absorption.

NDIR systems can be classified as single beam and double beam arrangements, each using filters or selective detectors, giving four basic types of NDIR spectrometer.

Figure 4(a) shows the design of a double-beam CO analyzer. IR radiation emitted by an IR radiation source (in general a coiled filament) alternately passes the sample cell (where the automotive exhaust gas is pumped through) and the reference cell (filled with a CO-free gas, such as air) and focused to a detector (usually a semiconductor material such as PbSe). The optical filter limits the spectral region of the radiation to the CO absorption band ($2000–2250\,cm^{-1}$). The sector shutter blocks up the two courses of rays alternately, so that there are alternately two signals from the detector, associated with the two paths. If there is no CO in the sample cell, the two signals have the same magnitude. If there is CO in the sample cell, the radiation in this path will be weakened, while radiation through the reference cell will stay constant. The resulting difference of the two signals is directly converted into the CO concentration.

Figure 4(b) shows the arrangement of a single-beam CO analyzer. In this arrangement there is only one path of rays with one cell, the sample cell. By rotating a wheel, two filters (F1 and F2) alternately cross the rays path, of which F1 will only transmit radiation between $2000\,cm^{-1}$ and $2250\,cm^{-1}$, while F2 will only transmit radiation from a small spectral region outside the CO absorption band. In this arrangement F2 takes the role of the reference cell in the system described above, and the transmitting window of F2 has to be chosen so that no component

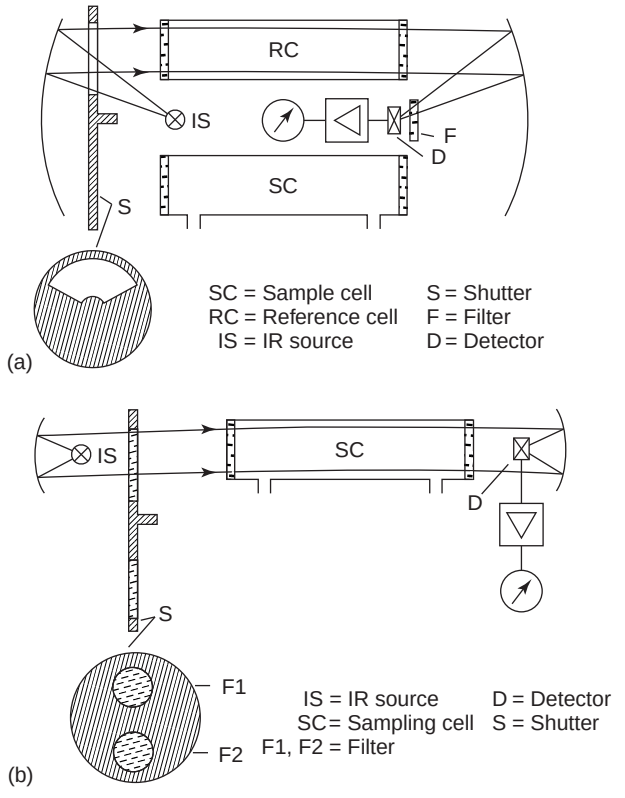

Figure 4 Working principles of NDIR analyzers: (a) double-beam arrangement, (b) single-beam arrangement.

of the exhaust gas will absorb this transmitted radiation. If there is no CO in the sample cell, the two signals, alternately produced with the frequency of the rotating filters, will be of the same magnitude. If there is CO in the sample cell, the signal is weakened when F1 is in the path whereas the signal is always the same when F2 is in the path. Again, the difference between the two signals is directly converted into the CO concentration.

Removing the filters in the arrangement of Figure 4(a) and (b) and replacing the semiconductor detectors by frequency-selective detectors,[5] leads to the remaining two systems mentioned above.

Changing the filter band to the corresponding absorbance of other compounds gives a system for measuring each exhaust compound having an absorption in the IR region.

Changing the IR radiation source to a UV/VIS radiation source (that is, discharge lamps such as mercury vapor high-pressure lamp, deuterium lamp, continuous-wave xenon lamp, or flashlight xenon lamp) and using the corresponding filters and a photodetector for that spectral region gives, respectively, a nondispersive-ultraviolet (NDUV) and a nondispersive-visible (NDVIS) analyzer. In most cases NDUV/VIS systems work with a filter wheel with a lot of different optical filters, so that by turning this wheel the spectral range can be covered quickly. The application of these analyzers is essential for the quantification of molecules consisting of identical atoms, such as molecules with no absorption in the IR region (see section 2).

ND analyzers can be coupled directly to the exhaust pipe of a motor car, but in each case and especially for IR-based measurements, the exhaust gas has to be filtered to remove soot and water. Such filter systems are directly integrated in commercially available analyzers.

With ND analyzers a large number of gases can be quantified in an on-line measurement; those of interest in exhaust gas analysis are

$CO \longrightarrow IR$	$NH_3 \longrightarrow IR$
$CO_2 \longrightarrow IR$	$N_2O \longrightarrow IR$
$SO_2 \longrightarrow IR/UV$	$NO_2 \longrightarrow UV$
$NO \longrightarrow IR$	$H_2S \longrightarrow UV$
$CH_4 \longrightarrow IR$	$C_2H_6 \longrightarrow IR$
$H_2O \longrightarrow IR$	$C_3H_6 \longrightarrow IR$
$C_2H_2 \longrightarrow IR$	$C_6H_6 \longrightarrow IR$
$C_2H_4 \longrightarrow IR$	$HCHO \longrightarrow UV$

3.2 Infrared Absorption-based Modern Techniques

The two IR absorption-based systems for analyzing multicomponent gas mixtures are FTIR spectrometers and DL spectrometers. Wavelength dispersive IR spectrometers, which record a spectrum step by step, are not suitable for automotive emission analysis, as they take too much time scanning a complete spectrum, especially when high spectral resolution is necessary.

3.2.1 Fourier Transform Infrared Principle

The design of a spectrometer using the FTIR technique is shown in Figure 5. The principal item of such a system is a Michelson interferometer. Consider the IR light source to produce monochromatic light of a wavelength λ_0. This light beam is divided at a beam splitter, reflecting 50% to mirror m1 and transmitting 50% to mirror m2. Reflected by the mirrors, the beams are combined again and focused onto a detector, such as mercury cadmium telluride (MCT). If the optical pathlengths are equal, the combining of both will result in constructive interference and the detector will show a maximum signal. By moving mirror m1 there will be a path difference which leads, when a difference of $\lambda_0/4$ is reached, to total extinction (as the new distance to mirror m1 is passed twice, the real path difference is $\lambda_0/2$ and adding two waves with pathlengths difference $\lambda_0/2$ leads to total extinction). Moving mirror m1 continuously with a fixed velocity v, gives a cosine signal with a frequency of $v/\lambda_0/4$ Hz.

For polychromatic radiation the conditions for constructive and destructive interference are different with

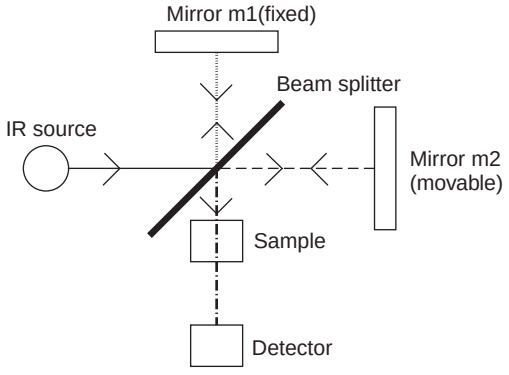

Figure 5 Design of an FTIR spectrometer.

respect to pathlength differences. The detector signal now takes the form of alternating maxima and minima, both decreasing in amplitude, and this interferogram contains all light frequencies. Calculating from the time domain to the frequency domain gives the spectrum of the light source. Figure 6 shows this Fourier transformation for the simple case of a two-line radiator. The interferogram (Figure 6a) can be decomposed into two periodic sine functions of $t_1 = 1$ s and $t_2 = 3$ s (Figure 6b) and from this we directly get the frequency distribution with two peaks at frequencies 1 Hz and 4 Hz (Figure 6c). In the case that there is an absorption of one frequency by a gaseous compound, its concentration can be calculated using the Beer–Lambert law (Equation 1):

$$I(\lambda) = I_0(\lambda) \exp(-\alpha l c) \tag{1}$$

where I is the intensity of a spectral absorption line, I_0 is the intensity from the spectrum baseline (where no specific absorption occurs), α is the specific absorption coefficient, l is the pathlength and c is the concentration. The specific absorption line intensities are obtained from spectra databases,[6] which are additionally used for identifying gaseous compounds and for the selection of noninterfering absorption lines.

As mentioned above, the main problem when analyzing complex gas mixtures, such as an exhaust gas, is the interference between the signals from different compounds. Müller et al.[7] have developed and tested algorithms to evaluate such complex IR spectra. The example in Figure 7 shows the complete procedure of evaluation for a multicomponent gas mixture. The measurements were done at a refinery by an open path FTIR spectrometer. Fourier transformation of the interferogram (Figure 7a) gives the complete absorption spectrum (Figure 7b) from which the part of interest can be selected (Figure 7c). The concentrations were determined from reference data[6] and the Beer–Lambert law. The spectra of the single components were calculated with respect to the concentrations to show the excellent agreement to the measured spectrum (see the residue in Figure 7c).

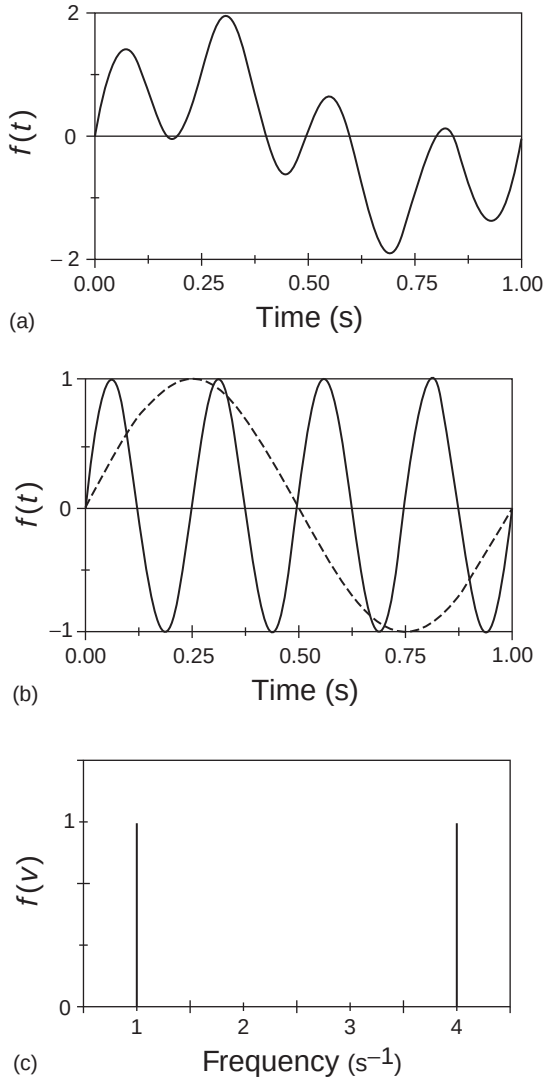

Figure 6 Simple example of a Fourier transformation: (a) interferogram, (b) transformed to time domain, (c) transformed to frequency domain.

3.2.2 Exhaust Gas Measurements using Fourier Transform Infrared Spectrometry

In recent years some FTIR spectrometer systems have been designed specifically for simultaneous on-line analysis of automotive exhaust gases. Hohenberg et al.[8] selected three types for detailed tests at a motor test stand:

- SESAM II (NIC 800 interferometer) – Siemens AG, München, Germany

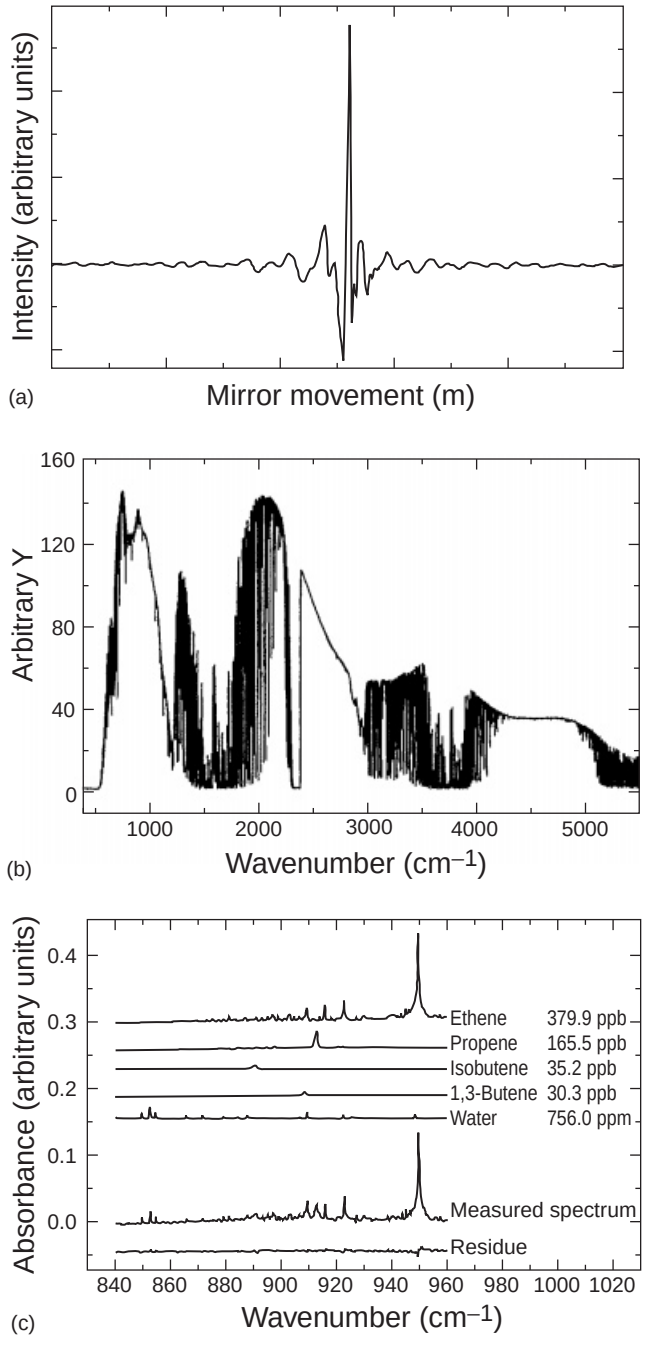

Figure 7 Simultaneous quantification of air pollution by FTIR: (a) interferogram, (b) complete IR spectrum, (c) spectral parts for quantification. (Reproduced by permission of Wiley & Sons, NY.)

- AMA-FTIR (Mattson interferometer) – Pierburg GmbH, Neuss, Germany
- MEXA-4000FT (Horiba rotary-type interferometer) – Horiba Ltd, Japan.

With these systems nearly all small molecules in an exhaust gas can be quantified, and up to 30 compounds can be detected simultaneously. Owing to the limited resolution of the FTIR spectrometer in use, the quantification of larger molecules, e.g. most hydrocarbons in an exhaust gas, is restricted.

For on-line sampling there is, in principle, only the need for a heated transfer line and a heated sample cell. A special gas-conditioning device is described in detail by Herget et al.[9]

The SESAM device is constructed only for measurements of undiluted exhaust gas (gas temperature 185 °C), the AMA-FTIR for diluted exhaust gas (gas temperature 50 °C), and MEXA-4000FT for both (gas temperature 113 °C).

All systems have to be calibrated. As the concentration of the different exhaust gas compounds can vary over some orders of magnitude, the calibration[10] has to be done with different test gases and different concentrations, to check linearity. A new test gas source has been developed by Quaß and Schilling,[11] where the concentration of the compound of interest is continuously produced and can be varied easily and continuously, so that extensive production of test gases for different compounds and different concentrations is not necessary. When analyzing gas mixtures by FTIR systems for compounds with high concentration one can expect an inaccuracy of about 2%, and for compounds with concentration below 10 ppm of about 10%, with respect to linearity and reproducibility.[8] The detection limits for different compounds determined for the three FTIR systems by test gas measurements can be taken from Figures 8 and 9.

Measurements on an Otto motor showed that, for the determination of CO_2, CO, NO, NO_2, N_2O, H_2O, and NH_3, the FTIR systems are a good alternative to conventional techniques with the advantage of simultaneous measurement. The determination of THC in an exhaust gas from an Otto motor is possible only when the evaluation algorithms are well fitted to the kind of fuel in use. In most cases this determination must be done by FID. However, the development of new algorithms is progressing rapidly.[8]

For diesel motors, in addition to the compounds mentioned above, SO_2 can be quantified. The determination of THC is not possible, because the concentration is almost below 100 ppm. Likewise, the determination of aromatic and alkenic compounds was not possible with the systems tested.

Figures 10 and 11 show examples for quantitative time-resolved on-line measurements of CO, NO, HCHO and NH_3 while running an Otto motor.

A special application of FTIR is testing the efficiency of catalytic activity, as in this case the important compounds are detected in a time-resolved manner, whereas they can only be determined integrally in traditional chemical

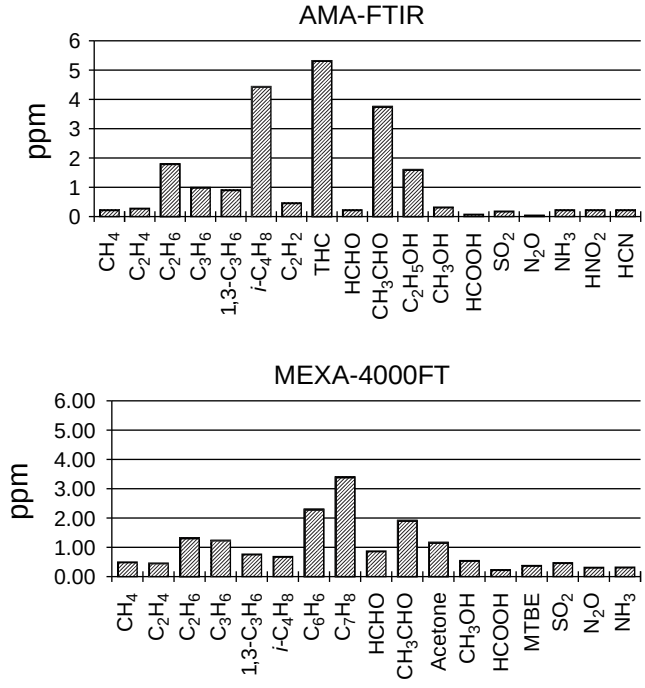

Figure 8 Detection limits for AMA and MEXA systems for diluted exhaust gas. MTBE = methyl tertiary butyl ether. (Reproduced by permission of FVV, Frankfurt, Germany.)

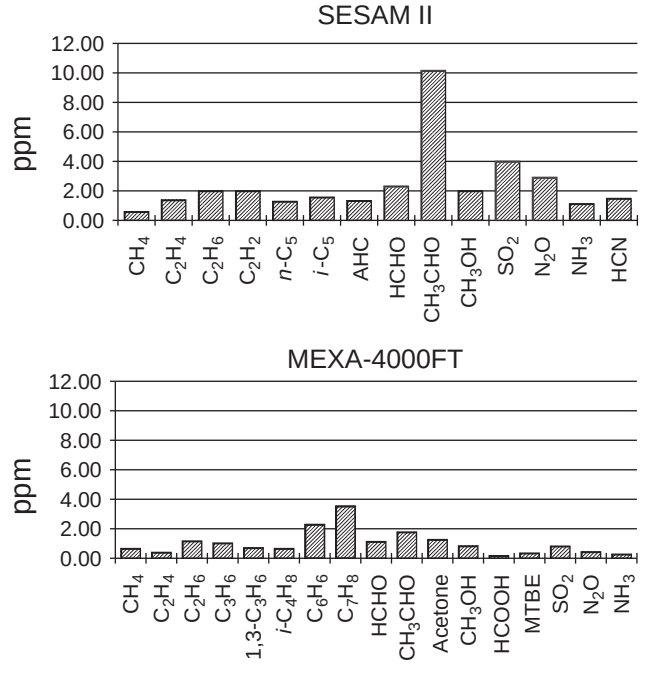

Figure 9 Detection limits for SESAM and MEXA systems for undiluted gas. AHC = aliphatic hydrocarbons. (Reproduced by permission of FVV, Frankfurt, Germany.)

analysis. As an example Figure 12 shows the creation of ammonia in a catalytic system.

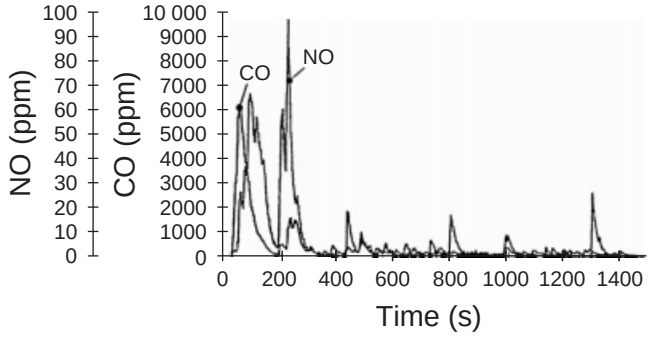

Figure 10 FTIR simultaneous on-line quantification of CO and NO. (Reproduced by permission of FVV, Frankfurt, Germany.)

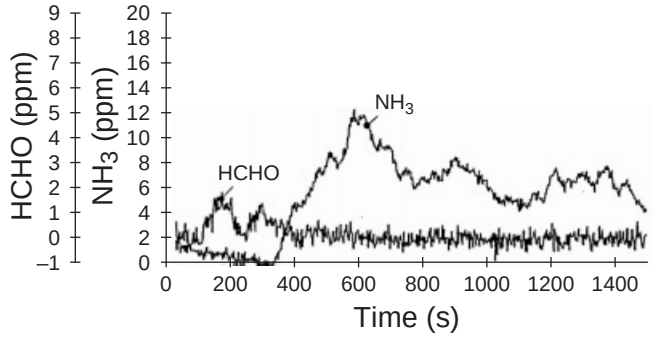

Figure 11 FTIR simultaneous on-line quantification of HCHO and NH_3. (Reproduced by permission of FVV, Frankfurt, Germany.)

At the Idaho National Engineering Laboratory Hybrid & Electric Vehicle Laboratory, automotive emissions measurements[12] were made with an FTIR system based on REGA 7000 (Nicolet) in accordance with the U.S. Federal Test Procedure. To demonstrate the real-time monitoring capability, a Dodge Dakota pickup, driven with liquified natural gas (LNG) and with gasoline, was used. For raw gas measurements the spectrometer was configured with a 4 m path length gas cell. The entire sampling apparatus was heated to prevent condensation of water and of higher-molecular-weight hydrocarbons. The complete IR spectrum was obtained using about 1 s intervals and 23 species of the exhaust gas were quantified simultaneously.

Figure 13 shows the THC emissions from the cold-start phase of the test for gasoline and LNG fuel. These THC levels were determined by summing the concentrations of the single hydrocarbons. As one can see, the most hydrocarbon emission occurs in the first minutes, i.e. before the catalytic converter has reached its operating temperature.

Comparing the THC concentration for gasoline fuel and for LNG fuel shows that, for latter, the THC concentration is much higher during the cold phase. This could lead

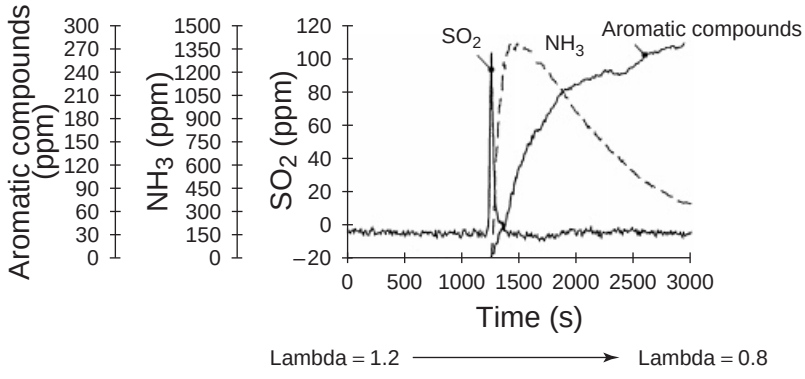

Figure 12 Quantification of NH_3, SO_2, and aromatic compounds during catalyst warm-up. (Reproduced by permission of FVV, Frankfurt, Germany.)

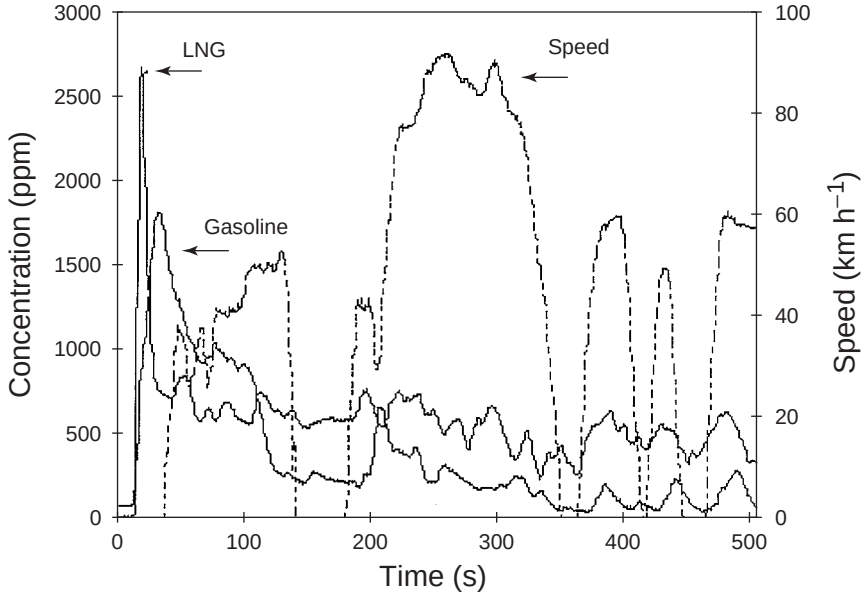

Figure 13 On-line quantification of THC for gasoline and LNG fuel. (Courtesy of the Idaho National Engineering and Environmental Laboratory.)

to the interpretation that gasoline-fueled cars will affect the environment less. However, detailed analysis shows that the initial spike of unburned hydrocarbons contains 98% methane in the case of LNG fuel, whereas for gasoline fuel the initial spike contains only 43% methane, with the rest being higher-molecular-weight hydrocarbons. It is debatable whether methane, as a global contributor to the greenhouse effect, or the higher-molecular-weight hydrocarbons, as precursors for the formation of local ozone, are more harmful to the atmosphere. However, this example clearly shows that there is an essential need for analyzers with the capability of measuring on-line and providing simultaneous exact determinations of exhaust gas compounds. The standard FID method, which determines THC only, is clearly not sufficient.

Figure 14 shows the simultaneous detection of CO and CO_2 for some minutes, beginning with a cold start. The principal CO emissions occur at engine ignition and during the warm-up period of the catalyst. When the catalyst has reached its operating temperature, known as the light-off temperature, the CO concentration suddenly decreases, while the CO_2 concentration increases because of efficient conversion of CO into CO_2. The small spikes in CO emissions from 200 s to about 250 s are the result of the rapid acceleration, where the engine is fuel rich. In this case there is a stoichiometric excess of fuel over the oxygen present, so that complete oxydation up to CO_2 cannot occur. This example shows the ability to monitor such transient events during a driving cycle, so that it is possible to study simultaneously the changes in

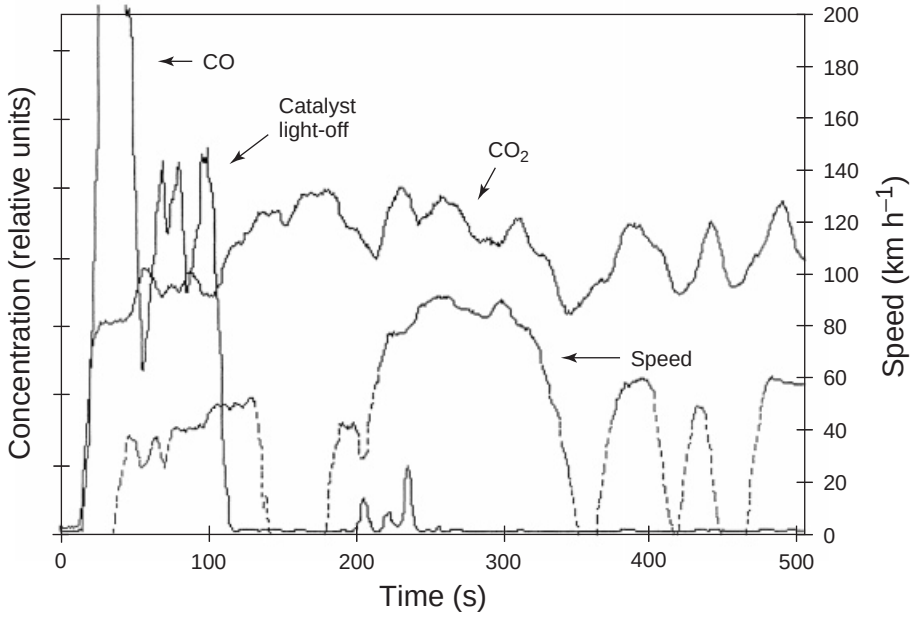

Figure 14 On-line quantification of CO and CO_2 during catalyst warm-up. (Courtesy of the Idaho National Engineering and Environmental Laboratory.)

emission due to the different conditions encountered and the sometimes complex relationships occurring between different exhaust compounds.

3.2.3 Laser Photometers with Infrared Diode Lasers

Due to their excellent beam quality as well as high intensity, lasers are ideal light sources for spectroscopic gas analysis. In addition, DL have advantages over other laser systems:

- small dimensions, so that compact spectrometers can be built up;
- available for the complete region of near and middle IR;
- low initial cost and cost of maintenance;
- easily tuneable;
- long life, especially in comparison with gas lasers.

There are some commercial applications of DL photometers[13] such as continuous determination of ammonia in flue gas or continuous determination of methane on mineral oil platforms. Much recent work has concentrated on systems for automotive emissions measurements.

3.2.4 Measuring Procedure Principles

The design and principles of operation of DL photometers is easy and similar to that of conventional spectroscopics. The DL emission passes a multiple reflection cell and

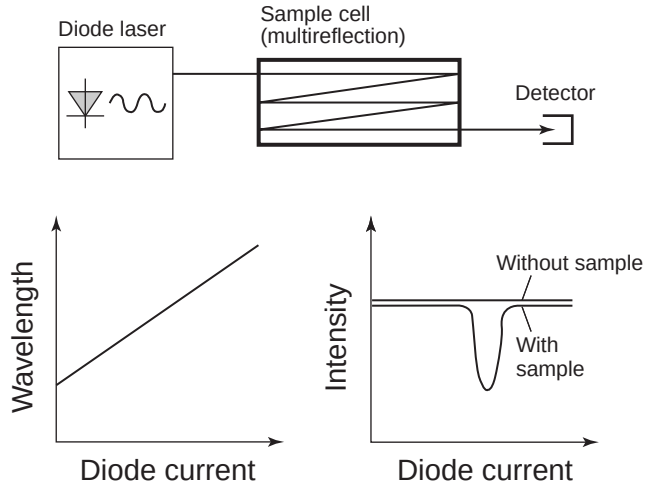

Figure 15 Principle of DL measurement.

is focused on to a detector (Figure 15). The emission wavelength can be tuned by the laser temperature or by the laser current, the latter being the most simple way. The laser parameters have be chosen so that the emission wavelength is near to an absorption line of the gas compound to be analyzed. If the laser emission is tuned to the absorption line, this line will appear in detected laser intensity (Figure 15). Absorption lines as well as laser lines are extremely small, so that it is difficult to match the selected absorption line. Gas analyzers therefore need a reference sample, such as shown in Figure 16. The reference sample is filled with the

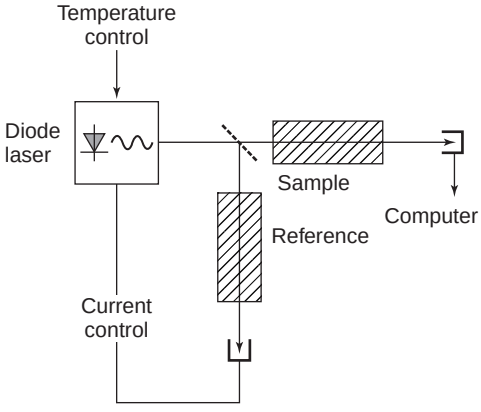

Figure 16 DL system tuning with a reference gas cell.

compound to be analyzed in high concentration and with repeated tuning a suitable absorption line is easily found. This reference signal is used to tune the laser emission reproducibly over this absorption line (Figure 16). The detected signal in the region of the absorption line is used to calculate the concentration from the Beer–Lambert law (see section 3.2.1). Quantitative calibration is done as described for FTIR.

A special IR DL system DEGAS III (dynamic exhaust gas analyzer system) for analyzing exhaust gas was developed at Fraunhofer Institut Physikalische Meßtechnik, Freiburg Germany.[14,15] The principle of DEGAS III is shown in Figure 17, in which two IR DLs are operated at controlled temperature and current. After combining them to a single beam, the laser beams are split into one reference and two measurement channels. The reference beam serves as an exact control of the laser emission wavelength, as described above. The DLs are operated by current pulses of typically 60 µs duration with a repetition rate of 5 kHz. The tuning range of each laser includes one selected absorption line of the respective molecular species.

As can be seen from Figure 18, the absorption lines become well separated without interference only for low pressures, so that a special sampling device has been developed (Figure 19). With the special cell geometry and the low-pressure conditions, a particle filter is not needed. The exhaust gas is fed into a 2 mm inner diameter tube of a coaxial heat exchanger about 1 m in length. The cell pressure is kept at 50 mbar by the large buffer volume. The windows of the sample cell are made of saphire and can easily be cleaned, which is necessary once a day. When

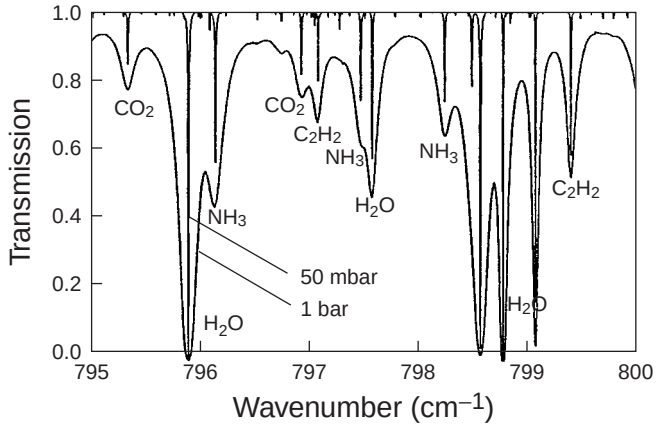

Figure 18 IR absorption spectra for different gas pressures. (Reproduced by permission of R. Oldenbourg Verlag, München, Germany.)

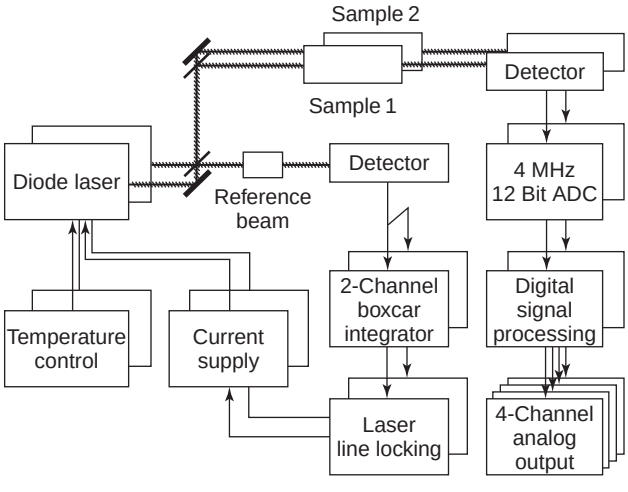

Figure 17 Design of the DL system DEGAS III. (Reproduced by permission of R. Oldenbourg Verlag, München, Germany.)

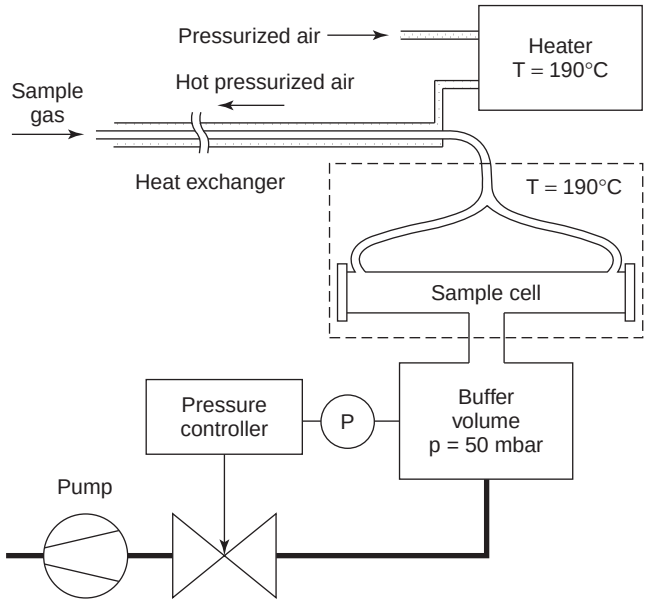

Figure 19 Sampling system for DEGAS III. (Reproduced by permission of R. Oldenbourg Verlag, München, Germany.)

sampling diesel engine exhaust gas with a high particulate content, cleaning is required after about half an hour. DEGAS III was conceived for two gas components, which are measured at two sampling locations, yielding four real-time values of concentrations. For an optical pathlength of 30 cm the detection limits are 5 ppm for CH_4, 10 ppm for CO, 100 ppm for NO, and 500 ppm for CO_2 with respect to the selected nonoverlapping absorption lines. Shorter pathlengths are possible with the result of faster response, but then the detection limit increases. Longer pathlengths can be achieved with a multiple reflection cell with a pathlength of 150 cm, resulting in a fivefold lower detection limit.

Calibration and zeroing is performed by flooding the whole analyzer system with a high flow of test and zero gas, which has to be done once a day.

3.2.5 Results

The system was tested during a number of trials in European and US automotive companies by the designers mentioned above, from which some selected results are shown below.

Figure 20 shows concentration profiles for NO and CO emissions taken from behind two different outlet valves from a single-cylinder research engine. Under these conditions the signal-to-noise ratio is better than 50 : 1, so that the observed variations represent real concentration profiles. Although the emissions of NO are largely synchronous, the CO values show distinct differences. This indicates inhomogeneous combustion due to improper air–fuel mixing. Such results are extremely helpful in optimizing combustion and have not been observable with conventional techniques.

Another result from these measurements is shown in Figure 21, which represents a comparison between NO analysis by a conventional instrument and by DEGAS III. Again, the variations represent real NO level fluctuations and are not due to noise, which is determined during the ignition-off period to be about 150 ppm. In addition to these fluctuations a large peak directly after ignition turn-on is registered, which may be the result of an unusual high temperature during the first combustion cycle. From these tests it is concluded that there is good agreement between conventional and DEGAS III measurements, but DL system provides more detailed information, which is essential for new motor design.

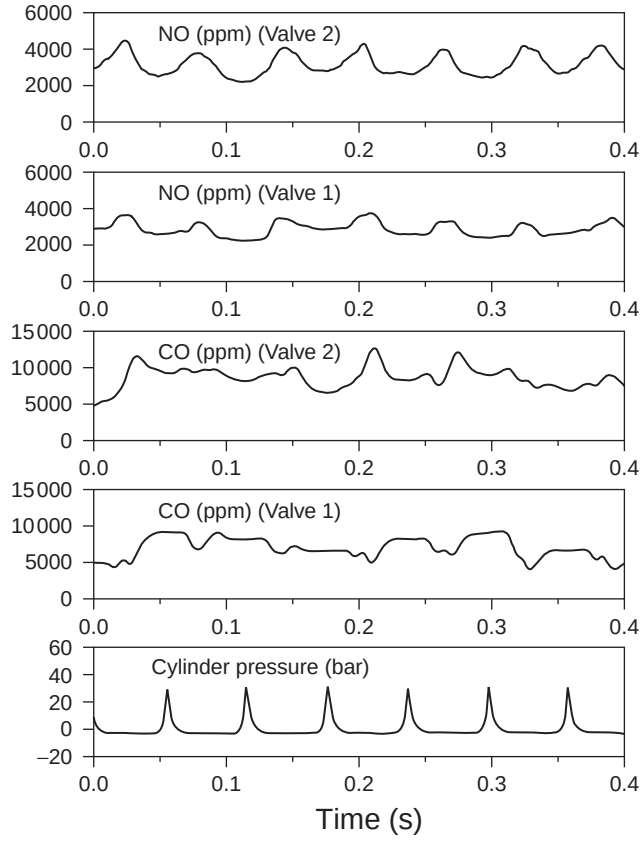

Figure 20 On-line quantification of NO and CO at two different outlet valves. (Reproduced by permission of R. Oldenbourg Verlag, München, Germany.)

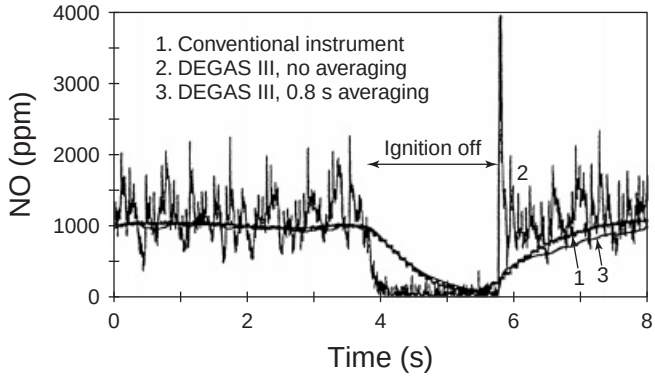

Figure 21 Comparison of NO measurement with DEGAS III and a conventional commercial system. (Reproduced by permission of R. Oldenbourg Verlag, München, Germany.)

3.3 Exhaust Gas Measurement using Raman-based Analytical Systems

As mentioned in section 2, the Raman process is based on an inelastic scattering of photons by molecules. Thus when electromagnetic radiation passes through a sample the main part will not interact with sample molecules, a minor part will interact by elastic scattering (Rayleigh scattering), and only a very small part will interact inelastically (Raman scattering). The quantum efficiency of the Raman process is therefore very poor. For that reason early Raman spectroscopic analyses of gases was

difficult – for gases, the Raman signals are about three orders of magnitude lower than for condensed media, due to their low density and lack of an internal field. However, this drawback can be compensated by the use of a powerful laser as the light source. In addition, when it is possible to remove dust particles and aerosols from the sample, the proportion of false light will be very small, which allows the employment of technically simple but very sensitive Raman spectrometers. Special types of holographically formed concave gratings have recently become available, which makes it possible to design systems with very large optical throughput and which, in addition, are very suitable for adaption to multichannel detectors.

3.3.1 Design of a Raman Polychromator

The request for large-throughput systems suitable for use with multichannel detectors has been met by various manufacturers of Raman instruments by the provision of mirror mechanisms, allowing the employment of only one monochromator for double or triple systems.[16,17]

The Raman polychromator shown in Figure 22 was designed for simultaneous on-line measurement at exhaust gases at the Institut for Spectrochemistry and Applied Spectroscopy, Dortmund, Germany. The optical elements were mounted on a cast-iron ground plate. The holographically manufactured concave grating was a Type IV model HG 20 U, 110 mm by 110 mm, 2000 grooves mm^{-1}, by Jobin Yvon. This grating allows the omission of any additional collimator optics in the polychromator (there are many alternative types of gratings available). The detector was a 1205 D SIT with 500 channels from Princeton Applied Research. The width of one channel was 25 µm. For the measurement of gases, the detector was cooled to about −15 °C. The laser was a Spectra Physics Ar^+ laser, and all measurements were performed with the laser line at 488 nm operating at 5 W.

The objectives, both for entrance and exit slit, were combined by three achromatic lenses with a focal length of 80 mm and a diameter of 50 mm (Figure 23a) resulting in powerful optics with an f number of 0.7. The corresponding data for the field lens are a focal length of 110 mm and a diameter of 50 mm. This lens may be omitted when using a grating which produces a flat spectrum in its focal plane. The laser focus was imaged onto the slit plane with an image ratio of 1 : 7. Figure 23(b) shows a Raman signal at the detector at various slit widths. For the measurements described, the slit width was chosen to be 500 µm, so that the slit acts only as a sort of optical stop for the suppression of false light from outside the laser focus.

The imaging ratio between the exit plane of the polychromator and the plane of the detector window was

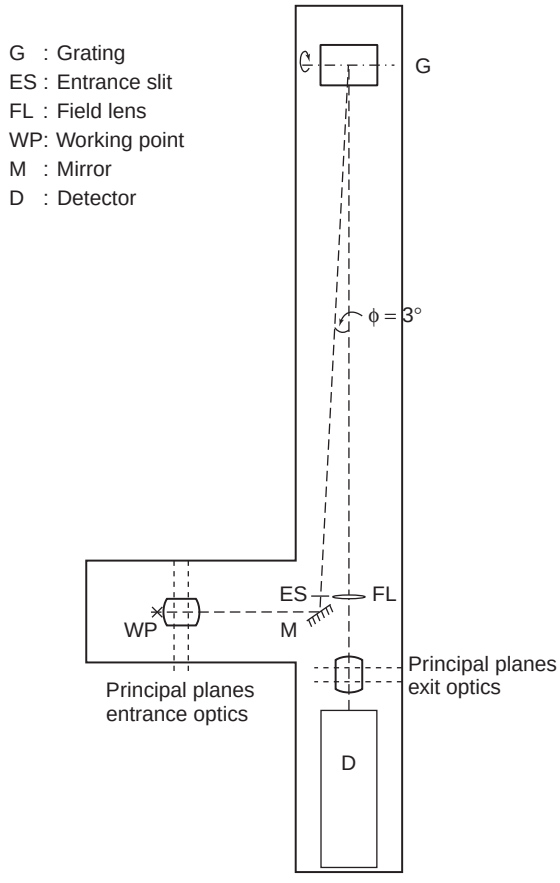

Figure 22 Design of a Raman polychromator; the laser beam is coupled into a sample cell placed at WP, as shown in Figure 27.

3.5 : 1. This means that the 12.5 mm width of the cathode array of the detector corresponds to about 45 mm in the exit plane. As the linear dispersion of the grating is 16 $cm^{-1}\,mm^{-1}$ this means that about 700 cm^{-1} of a spectrum can be detected simultaneously.

3.3.2 Measuring Procedure and Results for Two Different Raman Systems

For stationary measurements using the Raman polychromator described above, a flexible bag containing an evacuated sample cell was filled immediately at the exhaust pipe. After some minutes, when the soot had settled, the cell was opened and filled. This settling of the soot simulates a fairly efficient filtering procedure, which is necessary as mentioned above.

For dynamic measurements a mineral wool covered metallic plait probe was placed into the exhaust with the end of the probe about 50 cm inside the exhaust pipe. The probe was connected by a flexible pipe with a filter, membrane pump (24 l min^{-1}) and an open ended Raman cell. All measurements were performed on a 1.1-l Otto motor of 40 kW rating at 5700 rpm without catalyst.

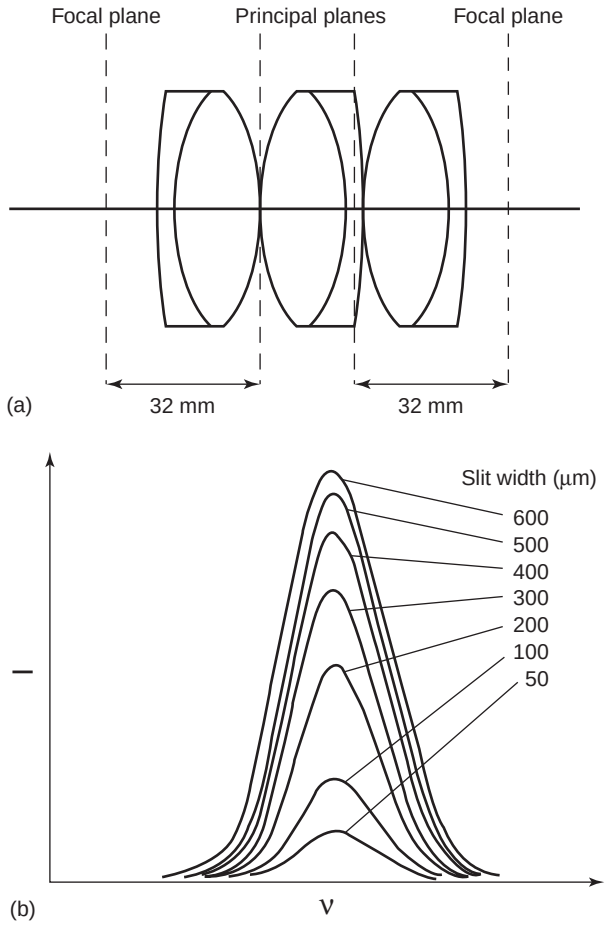

Figure 23 (a) The design of the powerful optics and (b) the shape of Raman signal as function of slit width.

All signals were related to the signal of N_2 of the air as standard, accounting for the spectral sensitivity k_λ which was determined according to D'Orazio and Schrader.[18] The Raman cross-sections of the compounds of interest were taken from the literature.[19] As the width of the entrance slit was set as mentioned above, the effective halfwidth of the signals was much larger than the physical halfwidth of the bands of the gaseous compounds, i.e. it was approximately the same for all signals. Therefore, the cross-sections could be correlated directly to the peak height of the signals, which facilitates the quantitative analysis considerably. This procedure yields the concentration c_X of compound X directly from Equation (2)

$$c_X = \frac{I_X}{I_{N_2}} \frac{\sigma_{N_2}}{\sigma_X} k_\lambda c_{N_2} \tag{2}$$

where I_X is the signals peak height for compound X, I_{N_2} is the signals peak height for nitrogen, c_{N_2} is the concentration of nitrogen in air, σ_{N_2} is the relative cross-section for nitrogen (=1), and σ_X is the relative cross-section for compound X (relative to $\sigma_{N_2} = 1$). As nitrogen is universal and its concentration in air is constant, it is a nearly ideal standard and additional calibration gases are not needed.

The cross-sections for the various compounds are very different. For example, the cross-section for benzene is more than 10 times larger than that for N_2, CO, CO_2, NO, etc., so that the detection limit for benzene or substituted benzene derivatives is more than 10 times lower.

The spectrum of Figure 24 shows the complete Raman spectrum from an exhaust gas taken with the flexible bag. The measuring time was 1 s and it is appropriate to correlate the signals to atmospheric N_2 as the external standard. As the exhaust gas was sampled directly after starting the cold motor with a choke, which decreased the supply of air, the sample shows comparatively high concentrations for low-weight hydrocarbons (50 ppm for ethene, 130 ppm for acetylene and nearly 180 ppm for methane) as well as for CO, O_2 and H_2. Against this the concentration of CO_2 is fairly low. The Raman band at about 1000 cm^{-1} labelled "c" can be assigned to a mono- or metadisubstituted benzene molecule.

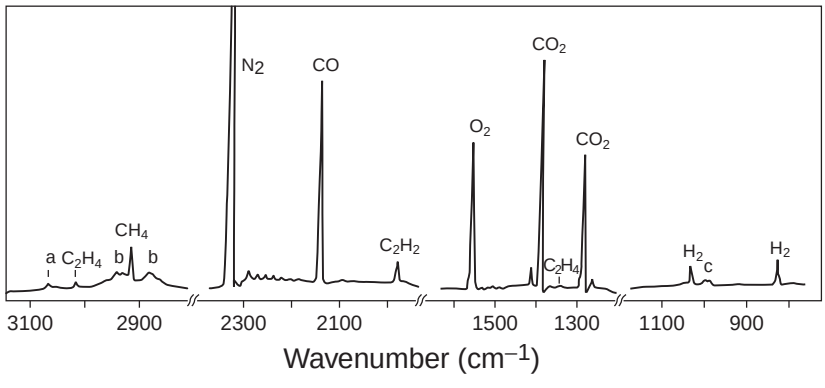

Figure 24 Raman spectrum of an exhaust gas, taken with a measuring time of 1 s. a, aromatic hydrocarbons; b, AHC; c, mono- or metadisubstituted benezene ring. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

Increasing the measuring time to 20 s enables additional signals to be registered, labelled "*" in Figure 25. Figure 26 shows the spectral region where all signals are exclusively produced by C–H vibrations from aliphatic and aromatic exhaust gas compounds. Although this part of spectrum cannot be used for a precise determination of THC concentration, it is useful in estimating the total number of C–H bonds from all hydrocarbons in an exhaust gas. Although this can suggest the order of magnitude of total hydrocarbon concentration, the true THC concentration has to be calculated by summing the concentrations of all hydrocarbons detected in the exhaust gas.

As mentioned above, dynamic measurements were performed by connecting the Raman sample cell to the exhaust pipe of the car by a transfer line (Figure 27). Raman spectra from the exhaust gas were registered with a time resolution of 1 s, so that the calculated concentrations represent mean values from an exhaust gas volume of about 400 ml. The sample cell was flushed with clean air both before and after exhaust gas measurements for detecting the signal of atmospheric nitrogen (see Equation 2).

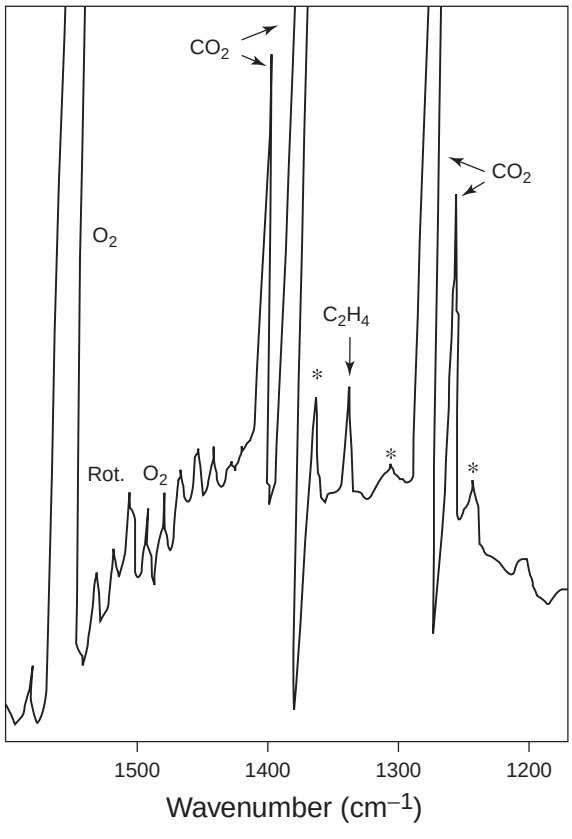

Figure 25 Section of a Raman spectrum of exhaust gas, taken with a measuring time of 20 s. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

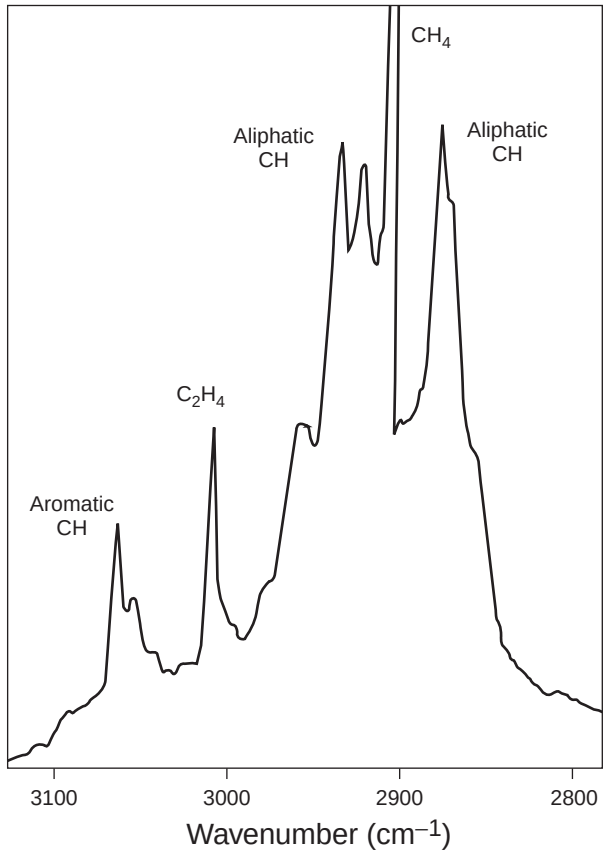

Figure 26 The C–H section of a Raman spectrum of exhaust gas, measuring time 20 s. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

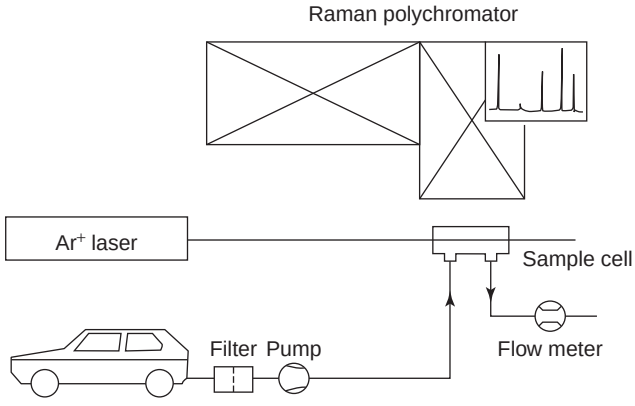

Figure 27 On-line measurement with a Raman polychromator. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

Figure 28 shows the concentrations of N_2, CO, CO_2, H_2, and O_2 while varying the speed of the motor from 750 rpm to 4000 rpm. These measurement was done after warming up the motor in idling cycle. Figure 29 shows the same components for a cold motor with a choke reducing

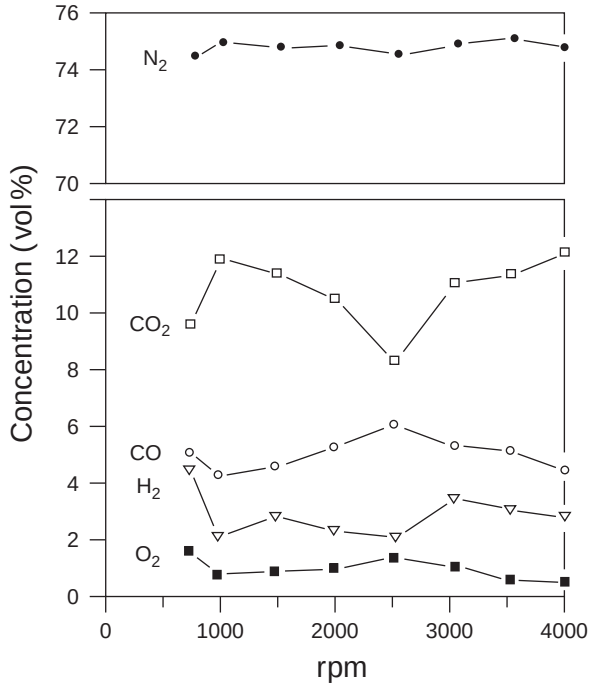

Figure 28 Simultaneous quantification of exhaust gas compounds as a function of the motor speed for a hot motor during idling cycle.

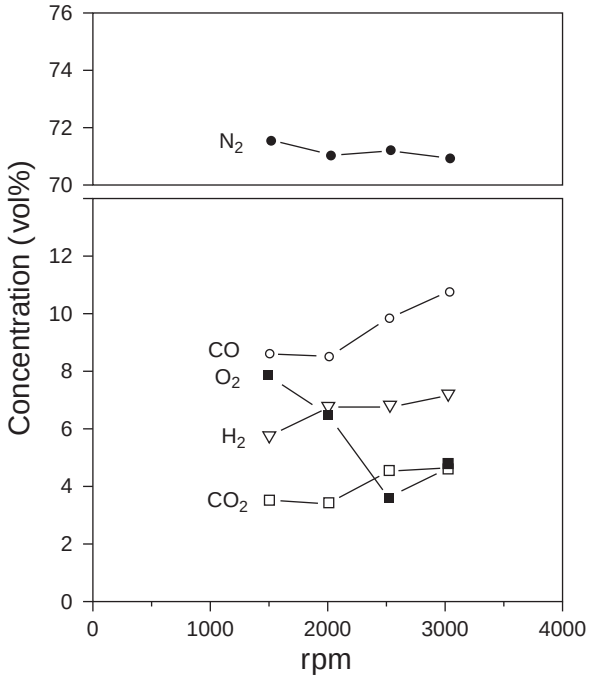

Figure 29 Simultaneous quantification of exhaust gas components as a function of the motor speed for a cold motor with reduced air supply.

the air supply. The reduced air supply can be seen directly from the resulting N_2 concentrations, which are about 4% lower than in case of the warmed-up motor. Furthermore, the nearly reversed concentrations for CO and CO_2, as well as the higher concentration of H_2, demonstrate the extreme motor conditions. It should be mentioned here that similar results would be obtained from the cold motor measurements of a car with catalyst – at least during the first few minutes until the light-off temperature is reached (see section 3.2.2).

A lot of exhaust gas compounds are missed in the measurements described here, even though they should have been in detectable concentrations in the exhaust gas. The reason is that in both cases (static measurement with the bag sample probing and dynamic measurements with a transfer line) the exhaust gas cooled down to room temperature, so that the following compounds could not reach the sample cell:

- water vapor, which condenses in the elastic bag and in the transfer line;
- SO_2, NH_3, and NO_x which react with condensed water;
- benzene and similar compounds which are adsorbed at the soot particles;
- compounds with low vapor pressures, such as PAH which condense for the most part.

Figures 30–33 show Raman spectra of some of these compounds, mixed in air to demonstrate the potential of Raman measurement, on one hand, and to demonstrate the limit of detection, on the other hand. Figure 30 shows the Raman signal of water vapor in air, taken with measurement time of 1 s. The relative humidity during this measurement was 43%, which means that the signal represents a water vapor concentration of 1%. Figure 31 shows the Raman spectrum of NO_2/air mixture with an NO_2 concentration of 40 ppm and a measuring time of 1 s. The NO_x species present in an exhaust gas depend critically on the temperature of the gas. In an internal combustion engine the NO_x should be represented nearly exclusively by NO with concentrations of about 4000 ppm (see Figure 20). Below temperatures of 150 °C, in contact with oxygen, NO will be increasingly converted into NO_2, resulting at normal temperatures in a well-defined equilibrium between NO_2 and N_2O_4. Therefore, the ratio of the concentration of the various NO_x will depend entirely on the distance between the motor outlet and the measuring point. Although the minimum detectable concentration of the system described here is about 500 ppm at a measuring time of 1 s and consequently much higher than for NO_2 (Figure 31), the Raman measurement should be done directly at the motor to ensure that the NO has not been converted. As can be seen from Figure 31, the resonance-enhanced Raman signal from NO_2 is

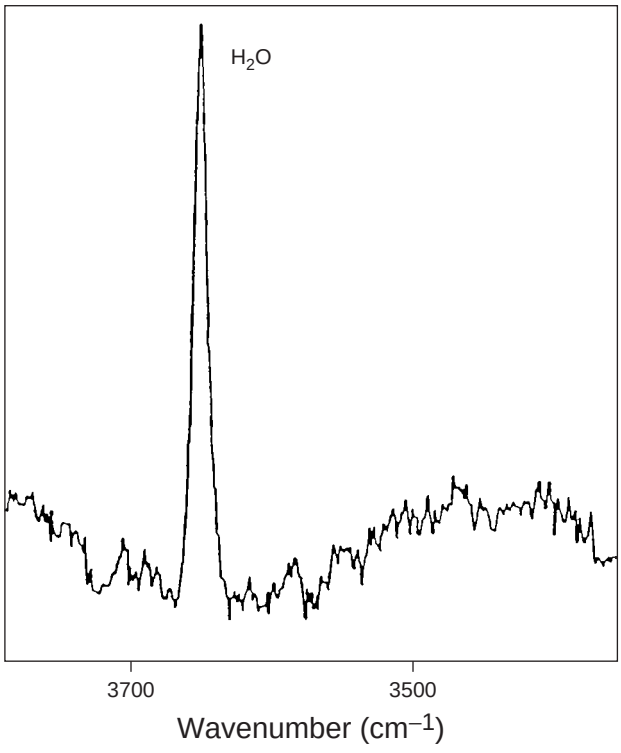

Figure 30 Raman spectrum of H_2O in ambient air. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

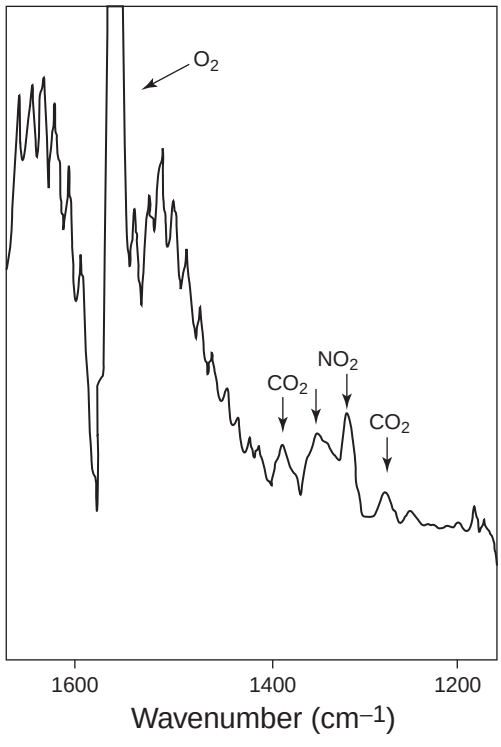

Figure 31 Raman spectrum of NO_2/air mixture. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

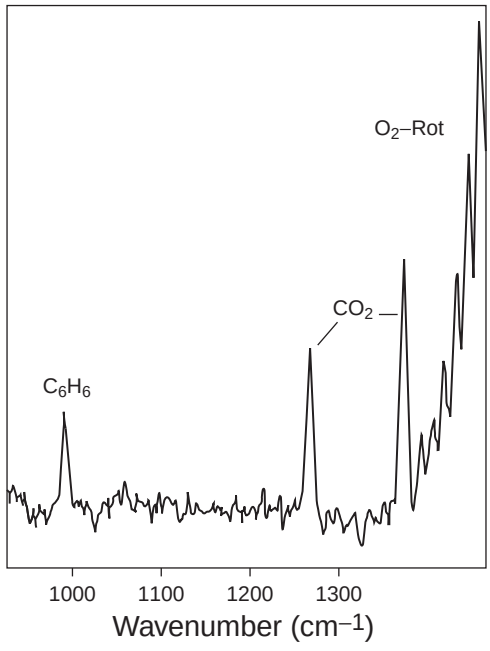

Figure 32 Raman spectrum of C_6H_6/air mixture. (Reproduced by permission of Vieweg Verlag, Wiesbaden, Germany.)

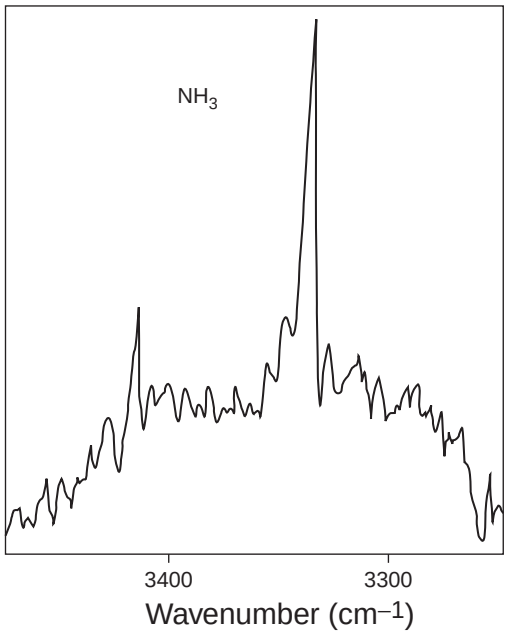

Figure 33 Raman spectrum of NH_3/air mixture.

accompanied by a broad fluorescence signal. This signal, increasing with increasing concentration of NO_2, may complicate a quantification.

Figure 32 shows the Raman spectrum of a benzene/air mixture with a benzene concentration of 10 ppm. As the Raman scattering cross-sections will increase with increasing the number of rings we can conclude from

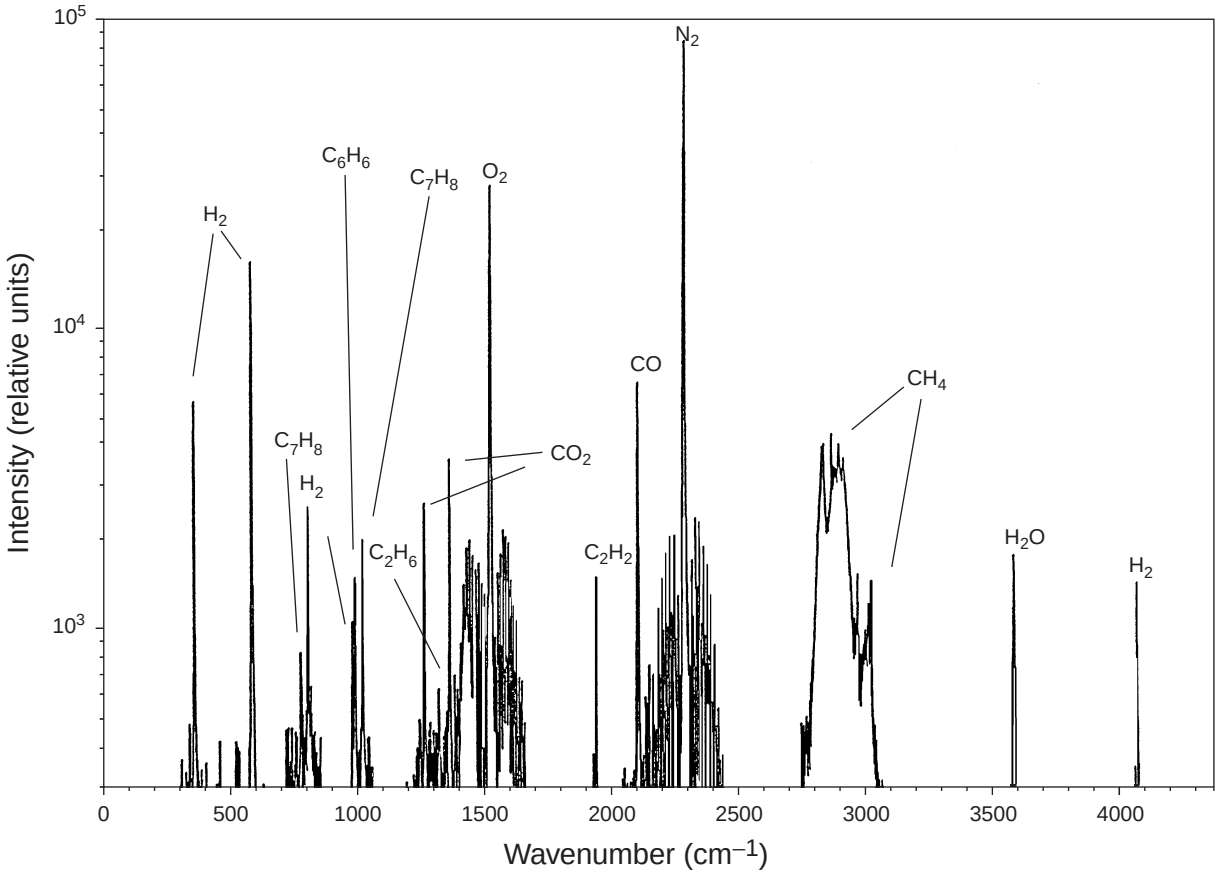

Figure 34 Raman spectrum of exhaust gas from a small two-stroke motor. (Reproduced by permission of FVV, Frankfurt, Germany.)

Figure 32, that the detection limits for PAHs will be in the range of some hundred ppb to some ppm.

Figure 33 shows part of the Raman spectrum of ammonia, for an ammonia concentration of 50 ppm in air.

Another Raman experiment for analyzing exhaust gas was done by Brüggemann and Kasal.[20] They used a 0.85 m Czerny–Turner double monochromator (SPEX, Edison NJ, USA), which was modified to a polychromator. The Raman spectra are registered with a charge coupled device (CCD) detector (SI, Gilching, Germany) with 1024 pixels, of which 700 pixels were intensified. After passing through the sample cell, the laser light is reflected into the sample cell again by a concave mirror, so that the laser power is extended to 30 W (using the 514.5 nm line from an Ar^+ laser). A second option is a high-energy pulsed laser with 20 J per pulse and a pulse width of 4 μs.

On-line measurements were done directly behind the exhaust pipe, after removing water, at a small two-stroke motor with measuring time of 6 s. Figure 34 shows the complete Raman spectrum, which had to be composed from some partial spectra due to the limited detector size.

Figure 35 shows the Raman spectrum of an off-line exhaust gas sample taken from a four-stroke motor. As the measurements were taken about 4 h after sampling, the composition should have changed. A good hint to this changing is the relatively high concentration of NO_2, which is not normally present in an exhaust gas in such high concentrations (see above). It obviously derives from NO, which is the main compound of NO_x in exhaust gas.

The results of a special Raman application are shown in Figure 36.[21] The quantification of H_2 before and after catalytic reactions can be used for checking the warm-up time, the light-off temperature and the technical constitution of a catalyst. Figure 36 shows the H_2 concentration to be fairly constant when a gas temperature of 90 °C is reached. The catalyst's light-off seems to be at about 100 °C, where the H_2 concentration in the exhaust gas behind the catalyst rapidly decreases from 0.5% to less than 100 ppm at 135 °C. The warm-up time was determined to be 140 s.

Figure 37 shows the simultaneous quantification of four exhaust-gas components (measured after the catalyst),

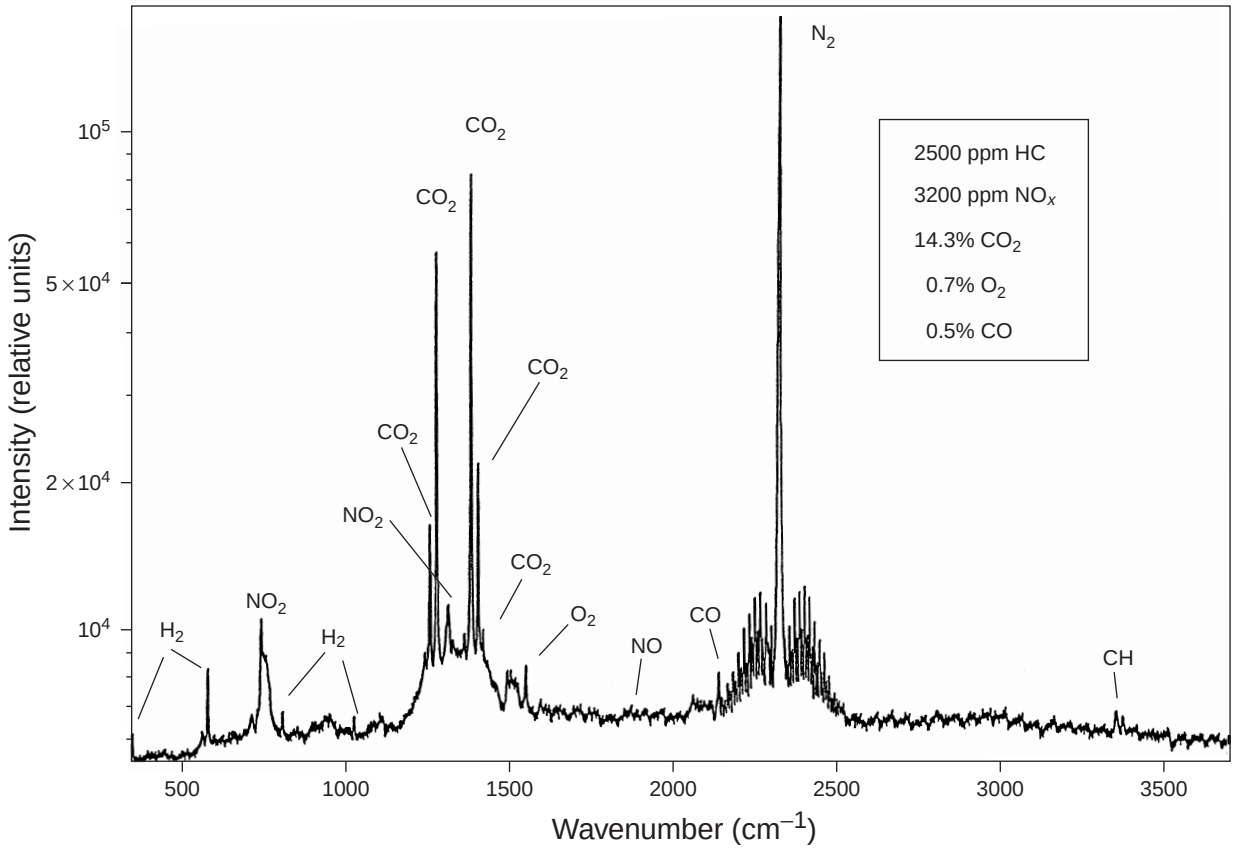

Figure 35 Raman spectrum of exhaust gas from an Otto motor, gas sampled in a bag. (Reproduced by permission of FVV, Frankfurt, Germany.)

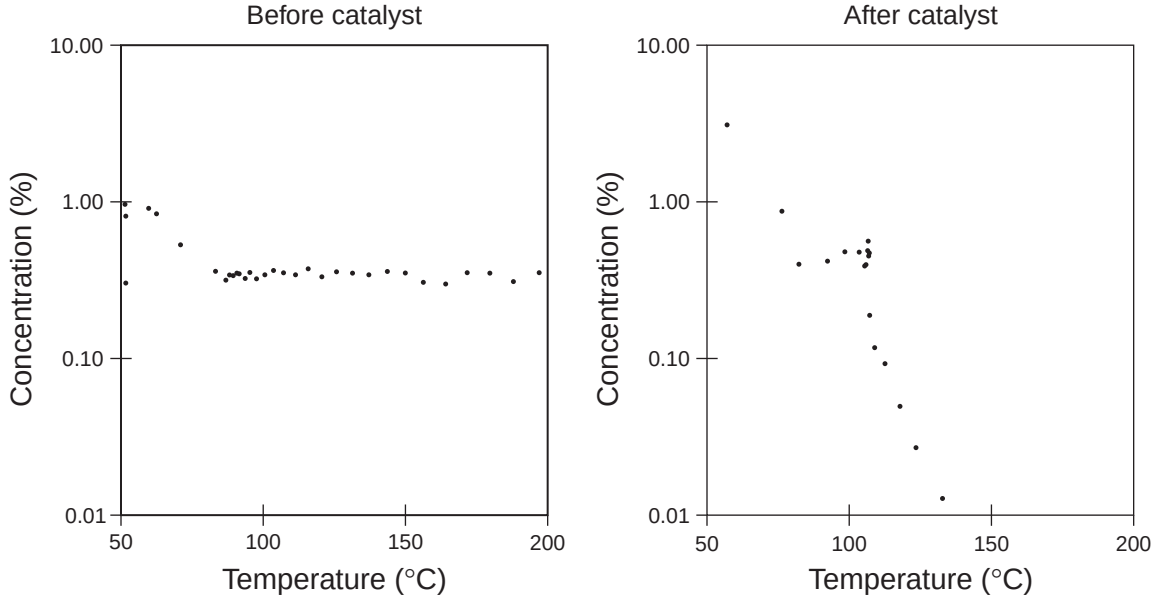

Figure 36 Quantification of H_2 dependence on exhaust gas temperature, measured before and after the catalytic reaction.

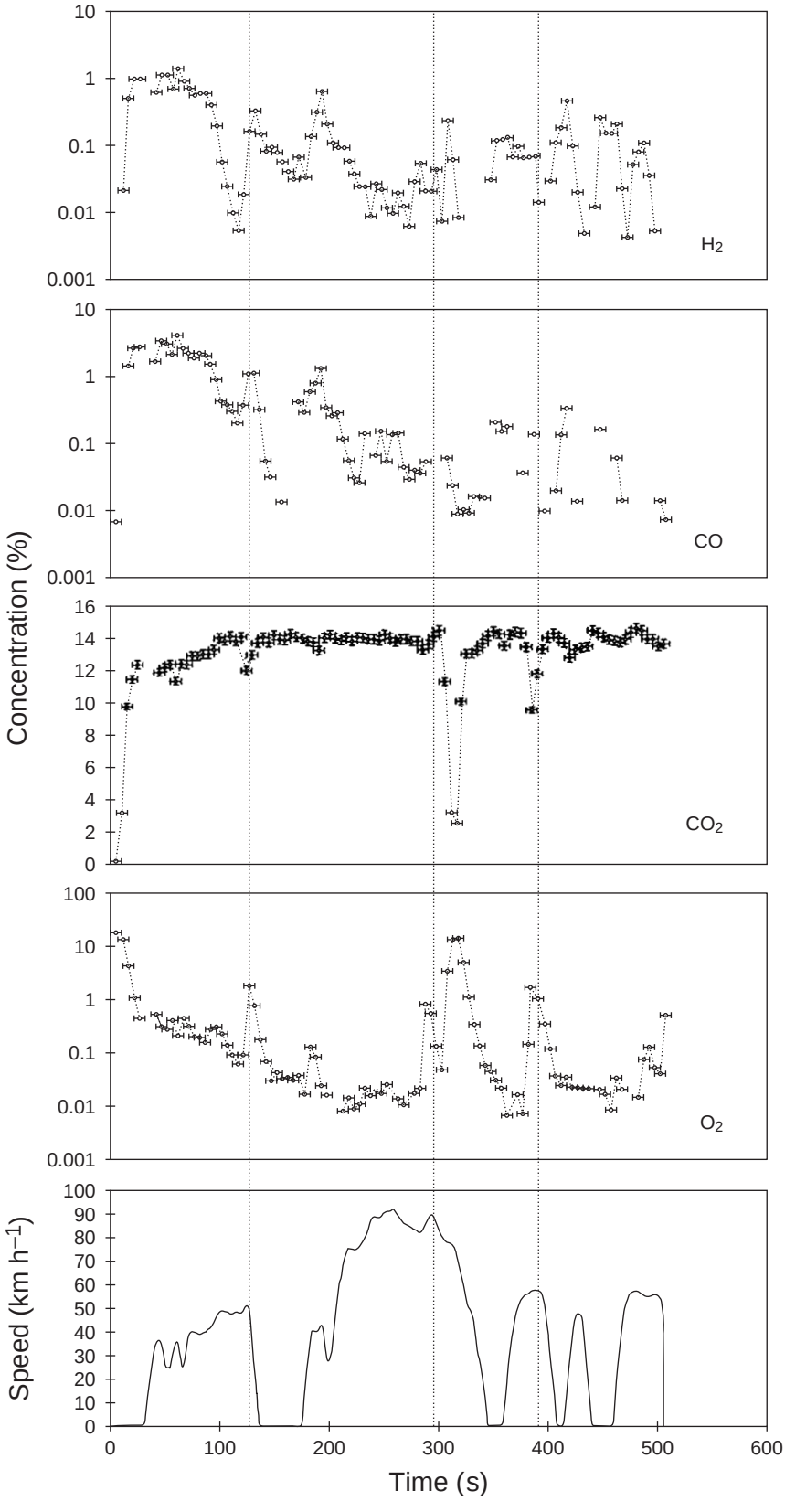

Figure 37 Simultaneous quantification of four exhaust-gas components during simulated town and country driving.

showing their dependence on vehicle speed. The measurements were done at a car test stand simulating town and country driving. The results show the capability of the Raman system to quantify exhaust gas components over four orders of magnitude with a time resolution of 5 s.[20]

4 POTENTIAL AND COMPARISON OF DIFFERENT TECHNIQUES

ND techniques, as used at NDIR and NDUV/VIS analyzers, are well established and have been legally certified for some years. Their application is restricted to the measurement of one component with relatively low time resolution, so these systems cannot be applied to simultaneous detection of the various exhaust gas compounds and are not able to resolve fast changes in exhaust gas composition.

FTIR spectrometers allow on-line and simultaneous measurement. They are a good alternative to conventional systems such as CO, CO_2, NO, and NO_2 analyzers, and are additionally capable of quantifying low-weight hydrocarbons. They are particularly suited to checking catalytic effects on exhaust gas, as the compounds created by catalytic processes (NH_3, HCHO, HCOOH, NO_2, N_2O) can be detected very sensitively. With time resolutions of about 1 s the detection limits are in the range of lower ppm. The detection limits can be improved by magnification of the optical path in the sample cell. This improvement has to be purchased by much lower time resolution – halving the detection limit results in a 10-fold lower time resolution. The interference problem may be solved by the development of sophisticated evaluation algorithms, which are in rapid development.

Due to such molecular properties, homonuclear molecules in principle can not be detected.

DL analyzers are now equipped with lasers emitting radiation from the near IR. The detection limits are comparable to that of FTIR systems, but can be improved by at least two orders of magnitude by using DL with emission from the middle infrared (MIR). However, these DL are of limited applicability and problems can occur when special wavelengths are required. In addition they have to be cooled to the temperature of liquid nitrogen and in some cases below. Each component requires a separate DL, and this leads to highly extended systems. New semiconductor materials may enable a middle infrared diode laser (MIRDL) to be developed that can be cooled thermoelectrically or even work at room temperature. Telecommunication systems use DL systems with external resonators. This technique is transferable to the MIRDL and will extend the tuning range, so that several constituents of a gas can be quantified by a single DL. These developments are progressing rapidly. For reasons of high selectivity and high sensitivity, DL measurements should be done on diluted gas samples. This may be a disadvantage for systems intended to analyze raw exhaust gases. Homonuclear compounds cannot be detected.

Raman-based analyzers can quantify exhaust gas components (including the homonuclear compounds H_2, O_2, N_2) on-line and simultaneously. In principle the measurements can be done in the raw gas stream directly behind the exhaust pipe, without the use of a sample cell, provided the exhaust pipe has an efficient soot filter. The detection limits for benzene and other compounds containing benzene rings are comparable to IR systems. They are higher (10–100 ppm) for other organic and inorganic compounds. The sensitivity can be increased using lasers with emissions in the UV region. As the Raman quantum efficiency depends on the fourth power of the laser frequency, this results in increased sensitivity of about one order of magnitude. Over that, UV radiation may be at or near to absorption bands of exhaust gas molecules, and the Raman signals will be in resonance or preresonance, enhanced to up to five orders of magnitude. If fluorescence occurs, these fluorescence signals may be used for compound quantification as, in general, the fluorescence of gases reveals distinct vibration structures and a consequent gain in specifity. However, as could be seen for NO_2 (Figure 31), fluorescence may also complicate a quantification. At the moment laser sources for that spectral region can be nitrogen, excimer, frequency-doubled dye lasers, or frequency quadrupled Nd : YAG laser. The development of a powerful DL, emitting in the violet or UV spectral region, will lead to an analyzer with high sensitivity, no need of calibration gas, easy handling and low price.

In summary, IR DL and Raman systems have great potential for further development, whereas FTIR systems have nearly reached their maximum possible sensitivity.

ACKNOWLEDGMENTS

The author wishes to thank Dr R. Grisar from Frauenhofer Institut Freiburg (Germany), B. Mewes from ITLR Universität Stuttgart (Germany), Ch. Wagner from Technische Universität Darmstadt (Germany) and George H. Cole from INEL, Idaho Falls (USA) for support with latest research results.

ABBREVIATIONS AND ACRONYMS

AHC	Aliphatic Hydrocarbons
API	Atmospheric Pressure Ionization
CCD	Charge Coupled Device

DL	Diode Laser
FID	Flame Ionization Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
IR	Infrared
LC	Liquid Chromatography
LNG	Liquified Natural Gas
MCT	Mercury Cadmium Telluride
MIR	Middle Infrared
MIRDL	Middle Infrared Diode Laser
MTBE	Methyl Tertiary Butyl Ether
ND	Nondispersive
NDIR	Nondispersive Infrared Spectroscopy
NDUV	Nondispersive-Ultraviolet
NDVIS	Nondispersive-Visible
NO_x	Nitrogen Oxides
PAH	Polycyclic Aromatic Hydrocarbons
THC	Total Hydrocarbons
UV	Ultraviolet
VIS	Visible
VOC	Volatile Organic Compounds

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Laser-based Combustion Diagnostics

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Fuel Performance Specifications, Mid-infrared Analysis of • Fuels Analysis, Regulatory Specifications for

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Raman Scattering, Fundamentals

REFERENCES

1. German parliament, 5. Emissionsschutzbericht of German Government, Drucksache 12/4006, Bonn, 1992.
2. J. Staab, H. Klingenberg, D. Schürmann, 'Strategy for the Development of a New Multicomponent Exhaust Emissions Measurement Technique', SAE Paper 830437, 1983.
3. C.V. Raman, K.S. Krishnan, 'A New Type of Secondary Radiation', *Nature*, **121**, 501 (1928).
4. N.B. Colthup, L.H. Daly, St.E. Wiberly, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, 31–33, 1990.
5. H.M. Heise, *Infrarotspektrometrische Gasanalytik – Verfahren und Anwendungen in Infrarot-Spektroskopie*, eds. H. Günzler, A.M. Bahadir, R. Borsdorf, K. Danzer, W. Fresenius, R. Galensa, W. Huber, I. Lüderwald, G. Schwedt, G. Tölg, H. Wisser, Springer, 1–67, 1996.
6. P.L. Hanst, S.T. Hanst, G.M. Williams, *Infrared Spectra, Database and Quantitative Analysis Programs for Measurement of Gases*, Infrared Analysis Inc. Anaheim, CA, 1995.
7. U. Müller, H.M. Heise, A. Gärtner, T. Häusler, 'Improved Strategies for a Quantitative Evaluation of Atmospheric FTIR-Spectra Obtained in Open Path Monitoring', *Field Anal. Chem. Fact*, **3**, 141–159 (1999).
8. G. Hohenberg, G. Fehl, Ch. Wagner, *Erprobung der FTIR-Abgasmeßtechnik zur Analyse der limitierten und nicht limitierten Abgaskomponenten von Diesel- und Ottomotoren bei stationärem und dynamischem Betrieb*, FVV, Frankfurt, 1995.
9. W.F. Herget, J. Staab, H. Klingenberg, W.J. Riedel, 'Progress in the Prototype Development of a New Multicomponent Exhaust Gas Sampling and Analyzing System', SAE Paper 840470, 1984.
10. H.M. Heise, H.-H. Kirchner, W. Richter, 'Calibration Method for the Infrared-spectrometric Trace Gas Analysis', *Fresenius' J. Anal. Chem.*, **322**, 397–400 (1985).
11. U. Quaß, M. Schilling, 'A New Test Gas Generator for Atmospheric Trace Compounds', *Fresenius' J. Anal. Chem.*, **350**, 461–466 (1994).
12. G.H. Cole, 'Automotive Emissions Analysis at the INEL', INEL Publications, IDAHO National Technical Information Centre (http://ev.inel.gov).
13. R. Grisar, H. Böttner, M. Tacke, G. Restelli (eds.), *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, Kluwer Academie Publishers, Dordrecht, 1992.
14. W. Thiel, W. Hübner, R. Grisar, W.J. Riedel, H. Wolf, 'Dynamic Laser Analysis of Exhaust Gas', SAE Technical Paper Series No. 940825, 15–24, 1994.
15. E. Wagner, R. Grisar, W.J. Riedel, M. Tacke, H. Wolf, 'Dynamic Exhaust Gas Analysis: A Tool to Study and to Improve Engine Management', CD Auto 98, 98A5017, Genua, 1998.
16. M. D'Orazio, R. Hirschberger, 'Multichannel Raman Spectrometer for the Study of Dynamical Processes in Analytical Chemistry', *Opt. Eng.*, **22**(3), 308–313 (1983).
17. R. Hirschberger, M. D'Orazio, 'Analysis of Automobile Exhaust Gases by a Raman Polychromator System', *Intern. J. Environ. Anal. Chem.*, **48**, 115–126 (1992).

18. M. D'Orazio, B. Schrader, *J. Raman Spectrosc.*, **2**, 585–592 (1974).
19. H.W. Schrötter, 'Raman Scattering Cross Sections in Gases and Liquids', in *Raman Spectroscopy of Gases and Liquids*, ed. A. Weber, Springer, 137–141, 1979.
20. D. Brüggemann, P. Kasal, 'Potential der Roman – Meßtechnik zur motorischen Abgasanalyse', FVV Workshop, Frankfurt, R483, 1995.
21. B. Mewes, P. Kasal, D. Brüggemann, 'Application of Laser Raman Spectroscopy for Quantitative, Time Resolved Gas Analysis', Poster Session at Laser '97, Munich, 1997.

Differential Optical Absorption Spectroscopy, Air Monitoring by

Ulrich Platt
University of Heidelberg, Germany

1 Introduction **2**
2 Light Absorption in the Atmosphere, Long-path Absorption Spectroscopy **2**
3 Differential Optical Absorption Spectroscopy Principle and Modes of Operation **3**
3.1 Differential Optical Absorption Spectroscopy Principle 3
3.2 Long-path Differential Optical Absorption Spectroscopy Using Artificial Light Sources 4
3.3 Differential Optical Absorption Spectroscopy with Direct Sunlight or Moonlight 5
3.4 Differential Optical Absorption Spectroscopy with Zenith Scattered Sunlight 6
3.5 Satellite-based Earthshine Differential Optical Absorption Spectroscopy 7
3.6 Determination of the Geometrical Photon Pathlength of Cloud-penetrating Sunlight 7
4 Practical Design of Differential Optical Absorption Spectroscopy Systems **8**
4.1 System Components 8
4.2 Design Considerations 10
4.3 Example Systems 11
4.4 Software Issues 12
5 Operation and Characterization of Differential Optical Absorption Spectrometry Systems **12**
5.1 Typical Measurement Sequences 12
5.2 Scattered Light with Artificial Light Sources 12
5.3 Instrument Function 13
5.4 Residual Spectral Structures 13
5.5 Characterizing 13
6 Evaluation Techniques for Differential Optical Absorption Spectra **14**
6.1 Preparation of Reference Spectra 14
6.2 High-pass Filtering of Spectra 14
6.3 Dealing with Spectral Shift and Changes in Dispersion 14
6.4 Spectral Fitting Procedures 14
6.5 Ring Effect and Related Problems with Solar Scattered Light 15
6.6 Realistic Estimation of Errors 15
6.7 Interferences 15
7 Sensitivity and Detection Limits of Differential Optical Absorption Spectroscopy **16**
7.1 Species Measurable by Differential Optical Absorption Spectroscopy 16
7.2 Detection Limits 16
7.3 Airmass Factors for Scattered Light Operation 18
8 Applications of Differential Optical Absorption Spectroscopy **18**
Acknowledgments **19**
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

Differential optical absorption spectroscopy (DOAS) allows the quantitative determination of atmospheric trace gas concentrations by recording and evaluating the characteristic absorption structures (lines or bands) of the trace gas molecules along an absorption path of known length in the open atmosphere. The DOAS technique is characterized by the following: (1) measuring the transmitted light intensity over a relatively (compared to the width of an absorption band) broad spectral interval; (2) high-pass filtering of the spectra to obtain a differential absorption signal and eliminating broad-band extinction processes such as Rayleigh and Mie scattering (RS and MS); and (3) quantitative determination of trace column densities by matching the observed spectral signatures to prerecorded (reference) spectra by, for instance, least-squares methods.

DOAS shares the advantages of most other spectroscopic techniques, including inherent calibration, sub-parts per trillion (ppt) to ppt sensitivity and precision (1–10%), good specificity, wall-less operation, and the capability for remote measurements. In particular, the concentration of very reactive species such as the free radicals OH, NO_3, ClO, BrO, and IO are determined with DOAS. Other species of interest to atmospheric chemistry are also measurable such as SO_2, CS_2, O_3, NO, NO_2, HONO, NH_3, CH_2O, and most monocyclic aromatic hydrocarbons.

A description of the technique is given, and the various modes of operation and light-path configurations are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

explained. The emphasis of this article is on the practical aspects of instrument design, components of DOAS systems, operation of DOAS instruments, and state-of-the-art techniques for the evaluation of DOAS spectra and realistic determination of detection limits. Finally some examples of DOAS applications are given.

1 INTRODUCTION

The quantitative determination of atmospheric constituents by DOAS makes use of the characteristic absorption structures (lines or bands) of trace gas molecules along a path of known length in the open atmosphere. As a spectroscopic technique DOAS has the advantages of inherent calibration, high sensitivity (sub-ppt to ppt, depending on the trace gas) and precision (1–10%), good specificity, wall-less operation, and the capability for remote measurements. The concentration of very reactive species such as the free radicals OH, NO_3, ClO, BrO, and IO can be determined with DOAS. In addition a wide variety of other trace gases is measurable with DOAS at very high sensitivities (mixing ratios in the sub-ppt to parts per billion (ppb) range; Table 1) including sulfur dioxide, ozone, oxides of nitrogen, nitrous acid, ammonia, aldehydes, and a large number of aromatic hydrocarbons.

Table 1 Substances detectable by UV/visible absorption spectroscopy

Species	Wavelength interval (nm)	Differential absorption cross-section per molecule (10^{-19} cm^2)	Detection limit with 5 km light-path (ppt)
SO_2	200–230	65	100[a]
	290–310	5.7	35
CS_2	320–340	0.4	500
NO	200–230	24	350[a]
NO_2	330–500	2.5	80
NO_3	600–670	200	1
NH_3	200–230	180	40[a]
HNO_2	330–380	5.1	40
O_3	300–330	0.045	4000
CH_2O	300–360	0.48	400
ClO	260–300	35	6
BrO	300–360	104	2
IO	400–470	170	1
Benzene	240–270?	21.9	400[a]
Toluene	250–280?	12.8	660[a]
Xylene (*o*/*m*/*p*)	250–280?	2.1/6.6/20.3	4000/1300/420[a]
Phenol	260–290?	198	40[a]
Cresol (*o*/*m*/*p*)		20.1/31.8/87.2	420/270/100[a]
Benzaldehyde	280–290?	44	200[a]

[a] 500 m Light path, 0.001 minimum detectable optical density.
UV, ultraviolet

In contrast to laboratory spectroscopy, the true intensity $I_0(\lambda)$, as would be received from the light source in the absence of any atmospheric absorption, is usually difficult to determine, as it would involve removing the air from the open light path. This problem is solved by measuring the so-called differential absorption. This quantity can be defined as the part of the total absorption of any molecule rapidly varying with wavelength, in contrast to the total light extinction it is readily observable. Thus DOAS is characterized by:

- Measuring the transmitted light intensity over a relatively (compared to the width of a molecular absorption band) broad spectral interval.
- High-pass filtering of the spectra to obtain the differential absorption signal
- Mathematically matching the observed spectral signatures to prerecorded information.

2 LIGHT ABSORPTION IN THE ATMOSPHERE, LONG-PATH ABSORPTION SPECTROSCOPY

Quantitatively the absorption of radiation is expressed by the Lambert–Beer law (Equation 1):

$$I(\lambda) = I_0(\lambda)\exp(-Lc\sigma(\lambda)) \tag{1}$$

where $I_0(\lambda)$ denotes the initial intensity emitted by some suitable source of radiation, $I(\lambda)$ is the radiation intensity after passing through a layer of thickness L, where the species to be measured is present at the concentration c (Figure 1). The quantity $\sigma(\lambda)$ denotes the absorption cross-section at the wavelength λ. The absorption cross-section is a characteristic property of any species, $\sigma(\lambda)$ can be measured in the laboratory, the determination of the light pathlength L is usually trivial. Once these quantities are known the (average) trace gas concentration c can be calculated from the measured ratio $I_0(\lambda)/I(\lambda)$ as shown in Equation (2):

$$c = \frac{\log\left(\dfrac{I_0(\lambda)}{I(\lambda)}\right)}{\sigma(\lambda)L} = \frac{D}{\sigma(\lambda)L} \tag{2}$$

where $D = \log(I_0/I)$ denotes the optical density of a layer of a given species (note that the decadic as well as the natural logarithm are used in the definition of D; here we always use the natural logarithm). Typical optical densities recorded in DOAS applications are 10^{-4}–10^{-2}.

Although the attenuation of a light beam by absorption due to atmospheric constituents is in principle described

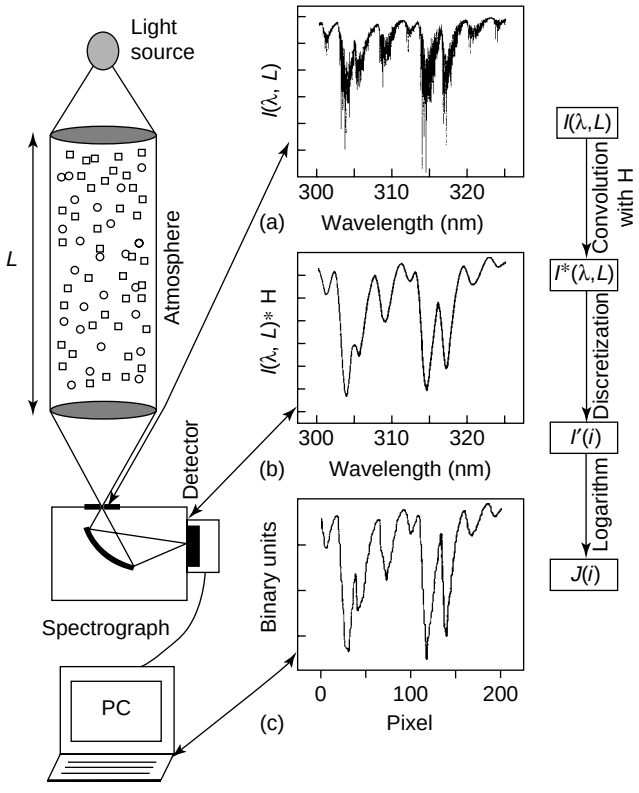

Figure 1 Trace gas detection by absorption spectroscopy makes use of the wavelength-dependent absorption of radiation originating from a suitable light source. Also shown is the relationship between (a) the true spectrum, (b) its transformation by the instrument function, and (c) the digitized version that is analyzed by the software.

by Equation (1) for practical measurements in the atmosphere, this equation is oversimplified in that it neglects other light-extinction mechanisms, in particular:

- Extinction due to RS, i.e. scattering by air molecules. This is not an absorption process; however, light scattered out of the probing light beam will normally not reach the detector, thus for DOAS applications it is justifiable to treat RS as an absorption process with the cross-section $\sigma_R(\lambda) \approx \sigma_{R0}\lambda^{-4}$ ($\sigma_{R0} \approx 4.4 \times 10^{-16}\,\mathrm{cm^2\,nm^4}$ for air).
- Extinction due to MS, i.e. scattering by atmospheric aerosol particles, described by $N_A\sigma_{M0}\lambda^{-n}$. with the Junge exponent n in the range 1–4[(1)] and an average absorption cross-section σ_M. By similar arguments to the case of RS, MS can be treated as an absorption process, even though it is only partly an absorption process.

In addition, the real atmosphere will usually have more than one absorbing species present, thus a more comprehensive description of atmospheric absorption is expressed in Equation (3):

$$I(\lambda) = I_0(\lambda)\exp\ -L\ \sum(\sigma_i(\lambda)c_i) + \sigma_{R0}(\lambda)\lambda^{-4}c_{AIR} + \sigma_{M0}\lambda^{-n}N_A \quad (3)$$

where $\sigma_i(\lambda)$ and c_i denote the absorption cross-section and the concentration of the ith species, c_{AIR} that of air molecules ($2.4 \times 10^{19}\,\mathrm{cm^{-3}}$ at 20 °C, 1 atm), and N_A the average aerosol number density. Typical extinctions due to RS and MS at 300 nm are $1.3 \times 10^{-6}\,\mathrm{cm^{-1}}$ and $1\text{–}10 \times 10^{-6}\,\mathrm{cm^{-1}}$, respectively.

3 DIFFERENTIAL OPTICAL ABSORPTION SPECTROSCOPY PRINCIPLE AND MODES OF OPERATION

3.1 Differential Optical Absorption Spectroscopy Principle

As noted in the introduction, DOAS makes use of the characteristic absorption features of trace gas molecules along a path of known length in the open atmosphere. Thereby the problem of determining the true intensity $I_0(\lambda)$, as would be received from the light source in the absence of any extinction, is solved by measuring the differential absorption. It is defined as the part of the total absorption of any molecule rapidly varying with wavelength and is readily observable as will be shown below. Accordingly, the absorption cross-section of a given molecule is split in two portions (Equation 4):

$$\sigma_i(\lambda) = \sigma_{i0}(\lambda) + \sigma_i'(\lambda) \quad (4)$$

where σ_{i0} varies only slowly (i.e. essentially monotonously) with the wavelength λ, for instance describing a general slope, (as for MS and RS), whereas $\sigma_i'(\lambda)$ shows rapid variations with λ, for instance due to an absorption line (Figure 2). The meaning of 'rapid' and 'slow' variation of the absorption cross-section as a function of wavelength is a matter of the observed wavelength interval and the width of the absorption bands to be detected. Introduction of Equation (4) into Equation (3) gives Equation (5):

$$I(\lambda) = I_0(\lambda)\exp\ -L\sum(\sigma_i'(\lambda)c_i)\ \exp\ -L \times\ \sum(\sigma_{i0}(\lambda)c_i) + \varepsilon_R(\lambda) + \varepsilon_M(\lambda)\quad A(\lambda) \quad (5)$$

where the first exponential function describes the effect of the structured differential absorption of trace species, and the second exponential constitutes the slowly varying absorption of atmospheric trace gases as well as the influence of RS and MS (described by $\varepsilon_R(\lambda) = \sigma_{R0}(\lambda)\lambda^{-4}c_{AIR}$ and $\varepsilon_M(\lambda) = \sigma_{M0}\lambda^{-n}N_A$, respectively). The attenuation

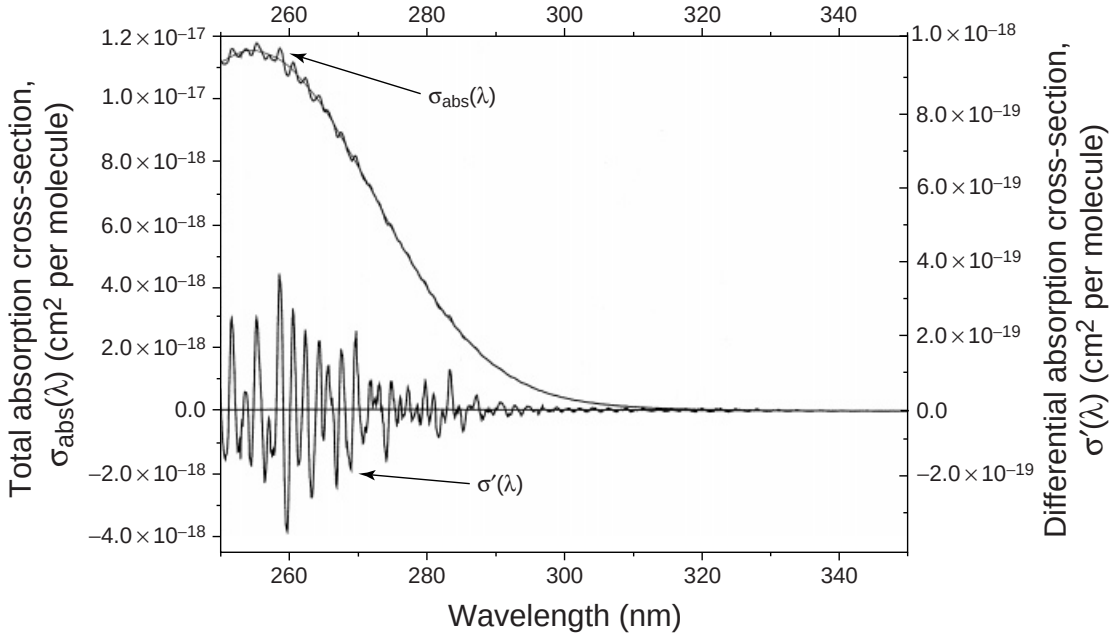

Figure 2 The total absorption cross-section $\sigma_{abs}(\lambda)$ and the differential absorption cross-section $\sigma'(\lambda)$ of ozone at around 300 nm. Although the UV bands of ozone present a dramatic example with which to illustrate the difference between σ and σ', for most other relevant trace gases in the atmosphere the differences between absolute and differential cross-sections are quite small.

factor $A(\lambda)$ describes the (slowly) wavelength-dependent transmission of the optical system used. Thus Equation (6) defines the quantity I'_0 as the intensity in the absence of differential absorption:

$$I'_0(\lambda) = I_0(\lambda) \exp\ -L\ \sum(\sigma_{i0}(\lambda)c_i) + \varepsilon_R(\lambda) + \varepsilon_M(\lambda)\quad A(\lambda) \qquad (6)$$

The corresponding differential absorption cross-section $\sigma'(\lambda)$ is then substituted for $\sigma(\lambda)$ in Equations (1) and (2), $\sigma'(\lambda)$ is determined in the laboratory (i.e. taken from literature data), just like $\sigma(\lambda)$. Likewise Equation (7) defines a differential optical density D' in analogy to Equation (2) as the logarithm of the quotient of the intensities I'_0 and I_0 (as defined in Equations (5) and (6), respectively):

$$D' = \log \frac{I'_0(\lambda)}{I(\lambda)} = L \sum_i \sigma'_i(\lambda) c_i \qquad (7)$$

Atmospheric trace gas concentrations can then be calculated according to Equation (2) with the differential quantities D' and $\sigma'(\lambda)$ substituted for D and $\sigma(\lambda)$, respectively. Figure 2 illustrates the relationship of $\sigma(\lambda)$ and $\sigma'(\lambda)$ and the determination of D' from a measured spectrum.

In many cases (such as for the free radicals OH and NO_3, see below) total and differential cross-sections are nearly identical. Obviously DOAS can only measure species with reasonably narrow absorption features. Continuous absorptions of trace gases will be neglected by DOAS. However, DOAS is insensitive to extinction processes, which vary only monotonously with wavelength, such as MS by aerosol, dust, or haze particles. Likewise slow variations in the spectral intensity of the light source or in the transmission of the optical system (telescope, spectrometer etc.) are also essentially eliminated.

The DOAS principle can be applied in a wide variety of light-path arrangements and observation modes as sketched in Figure 3.

3.2 Long-path Differential Optical Absorption Spectroscopy Using Artificial Light Sources

The most basic arrangement of a light path in the open atmosphere (Figure 4) consists of a detector (coupled to a dispersing device, for instance a spectrometer, and a telescope) looking into the light beam emitted from the light source such as a searchlight. The instrument will average the concentration over this extended volume of air. This is of advantage in applications where average concentration levels are desired. In particular, local influences such as small emission sources (for instance near the instrument site) have little effect on the result of the measurement. However investigations of, for example, fast chemical processes in the atmosphere require small

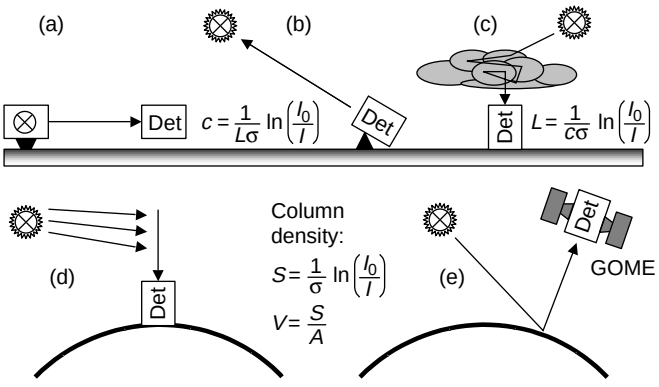

Figure 3 The DOAS principle can be applied in a wide variety of light-path arrangements and observation modes using artificial (a) as well as natural (b–e) light sources. Either the (light-path averaged) trace gas concentration (a, b), the trace gas column density (d, e), or the length of the light path (as in clouds, c) can be determined (GOME, Global Ozone Monitoring Experiment).

measurement volumes. A logistical disadvantage of this arrangement is the requirement to align the optical elements at either end of the light path; also, power is needed at two places.

Although early DOAS instruments were based of this arrangement,[2,3] which is still in use in some commercial instruments, a series of variants have been developed.

On variation of this design uses a reflector (usually a corner-cube retroreflector as indicated in Figure 5) in the field to return the light of a source located next to the spectrometer, thus the total light pathlength is twice the distance between light source/spectrometer and reflector. A coaxial merge of transmitting and receiving telescope is usually used.[4] The advantages and limitations are similar to the basic arrangement (Figure 4). However, the power source and delicate alignment is only necessary at the spectrometer/light source site.[5–8] Another advantage is in reducing the effect of atmospheric turbulence on the light beam. This is due to the ability of corner-cube retroreflectors to return the incident light exactly (although with some lateral offset) into the direction of incidence. Because of this last advantage, this arrangement has become the standard set-up for long-path DOAS instruments.

Another alternative is for the path to be folded in a multiple-reflection cell.[9–12] This approach has two main advantages. First, the measurement volume is greatly reduced, thus making the assumption of homogeneous trace gas distribution much more likely to be fulfilled. Second, the length of the light path is easily changed (in known increments of small multiples of the base path, depending on the particular design of the cell). It is therefore possible to keep the length of the light path close to its optimum (section 7) for the atmospheric conditions.

Disadvantages are the more complicated optical design, reflection losses (high reflectivity $R > 0.99$ of the cell-mirrors over spectral intervals of several 10 nm are difficult to obtain, in particular for UV), and the high radiation intensity inside the folded light path can cause photochemical reactions with consequent changes in the concentrations of the species to be measured.

3.3 Differential Optical Absorption Spectroscopy with Direct Sunlight or Moonlight

The total atmospheric trace gas contents can be observed by using extra terrestrial light sources, namely sunlight or moonlight. From the optical density D as derived from the recorded spectra, only total column densities

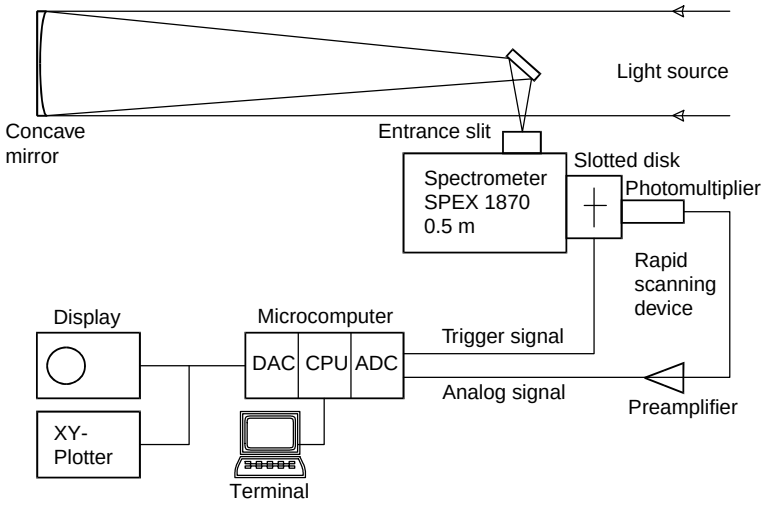

Figure 4 Simple set-up of a low-resolution DOAS (the searchlight-type light source is not shown).

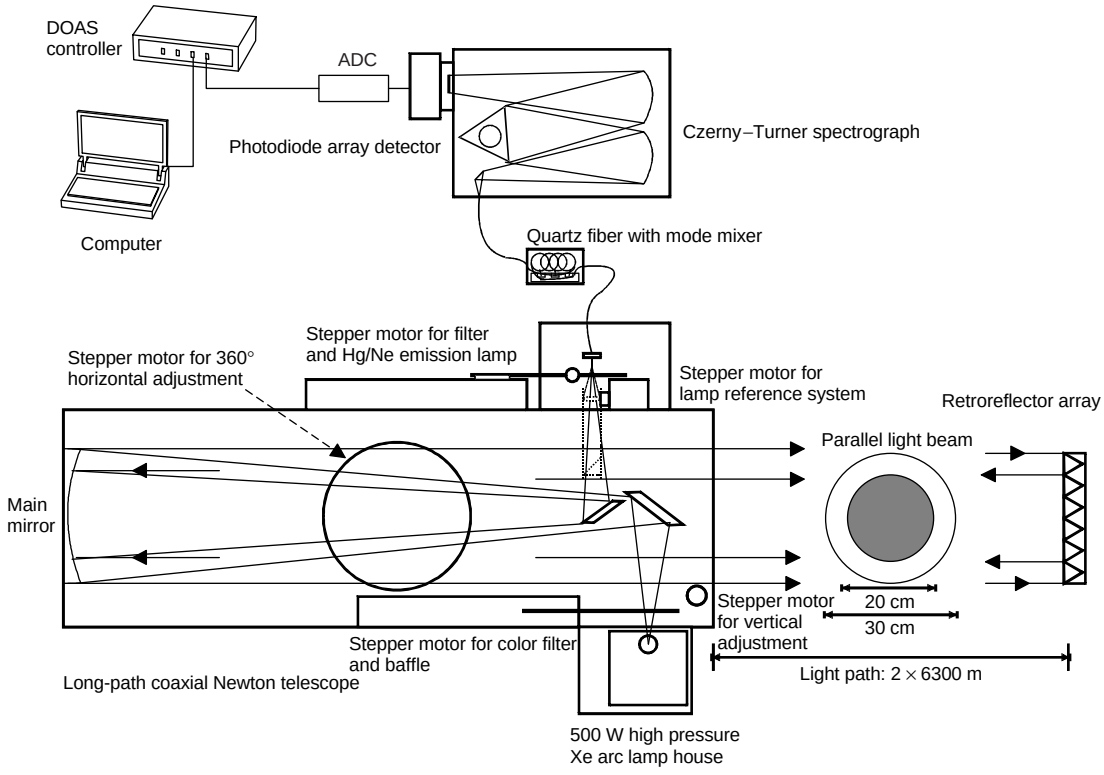

Figure 5 Typical set-up of a modern low-resolution DOAS. It uses a coaxial merge of transmitting and receiving telescope. An outer ring of the receiving mirror is used as light source mirror, thus emitting a hollow beam. Due to beam divergence the center hole is filled after a short distance. The high-precision retroreflector array returns the light exactly (within seconds of the arc) towards the source; however, due to the lateral offset of the reflected rays and to atmospheric turbulence the reflected light also reaches the center of the mirror acting as the primary of the receiving telescope.

$S = \int c(z)\,\mathrm{d}z$ (with $\mathrm{d}z$ denoting an element of the light path) can be derived. An important step in the evaluation of the spectra is the elimination of the strong solar Fraunhofer structure. This is readily done by recording spectra at different solar zenith angles (SZA, ϑ) – while the contribution of the earth's atmosphere then varies proportionally to the airmass factor (AMF, $A \approx 1/\cos\vartheta$, neglecting the curvature of the earth's surface), the solar lines stay constant. Thus ratioing two spectra taken at SZAs ϑ_1 and ϑ_2 will yield a spectrum with the effective AMF A_{12} given by $A_{12} \approx (1/\cos\vartheta_2 - 1/\cos\vartheta_1)$. The deviations from the $1/\cos\vartheta$ relationship become noticeable at $\vartheta > 75°$, (where $A(75°)$ is about 4% smaller than $(\cos 75°)^{-1}$), in particular at $A(90°) \approx 20$ instead of infinity.

An interesting application of this technique is the determination of vertical trace gas concentration profiles by taking a succession of spectra during an ascent of a balloon-borne spectrometer. From these spectra the column densities S_z above the balloon altitude z can be derived. After conversion into vertical column densities and differentiating, the concentration profile $c(z)$ can be derived.[13,14]

3.4 Differential Optical Absorption Spectroscopy with Zenith Scattered Sunlight

A spectrograph directed to the zenith (Figure 6) observes the optical density of trace gas absorption bands indirectly via the scattering process of the sunlight at different layers of the atmosphere (assuming that the instrument located is north or south of the respective tropic so that it never receives direct sunlight). At large SZAs this involves averaging over the contribution of nearly horizontal rays, that reach the zenith at different altitudes with different intensities and information content about the trace gas layer.[15–17] Simplified, the detected light from the zenith can be represented by a most probable light path through the atmosphere defined by the most likely scattering height Z_o in the zenith. Typically Z_o reaches about 26 km at 90° (327 nm) down to 11 km (505 nm). Thus, the wavelength-dependent beam height allows, in principle, a certain degree of height resolution. Those slant light paths through the atmosphere are quite long under twilight conditions, thus at 90° zenith angle (and 327 nm) a horizontal light path running 600 km through the stratosphere integrates 90% of the trace gas absorption signature. As in the case of

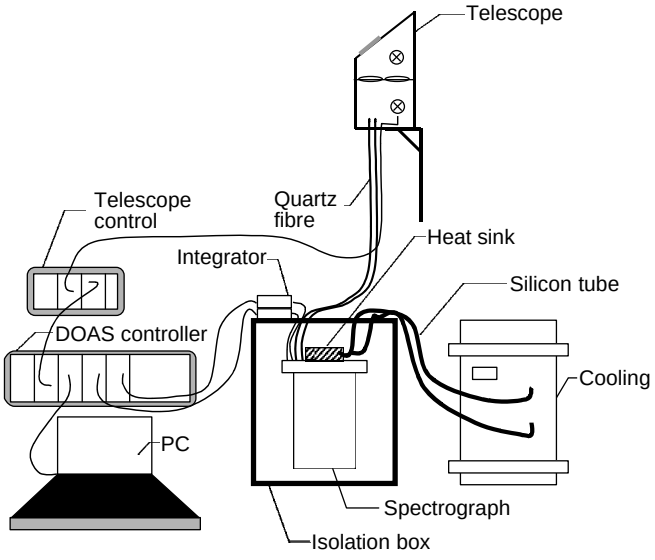

Figure 6 Optical set-up of a zenith-scattered light-observing spectrometer. Two independent spectrographs (for the UV and the visible spectral ranges) are used, therefore two telescopes (of about 4 cm aperture) are pointed to the zenith. The detectors are cooled to reduce the dark current; insulation and thermostating of the spectrometers reduces the thermal drift of the instruments.[99]

direct sunlight from the measured optical density D, only effective (or slant) column densities S can be derived. The effective term indicates the fact that the light reaching the spectrometer is an average over a multitude of rays actually taking somewhat different routes through the atmosphere. The solar Fraunhofer structure is eliminated in a way analogous to the direct sunlight observation.

Modeling the physical process of radiation transport, including RS, refraction and MS, allows the AMF to be determined. As in the case of direct sunlight DOAS, this is defined as ratio of the effective column S and the vertical column density (VCD, V) of a given trace gas ($A = S/V$). Typically the AMF increases from values near unity at $\vartheta = 0$ to ≈ 20 at $\vartheta = 90^\circ$. It is a function of SZA and wavelength, and depends to some extent on assumptions of the relative vertical trace-gas, stratospheric-aerosol, and air-density profiles.[17–19]

DOAS of scattered, rather than direct, sunlight has been applied to the determination of VCDs for O_3, NO_2,[20–26] NO_3,[27,28] OClO,[29–33] BrO,[34,35] and $(O_2)_2$.[36,37]

This technique has two major advantages: (1) no adjustment of the telescope direction is required (which is relatively difficult in the case of direct sunlight or moonlight); (2) stratospheric species can be detected even in the presence of (tropospheric) cloud cover. These advantages come at the expense of relatively difficult calculations of the effective light pathlength (see below); in addition, several effects of light scattering in the atmosphere have to be carefully compensated for.[38]

In a recently developed variant of the zenith scattered sunlight (ZSL) DOAS technique, the off-axis spectroscopy instrument is pointed away from the zenith by an off-axis angle Θ, usually towards the sun (but not looking directly to the sun). The advantages of this arrangement are better sensitivity at very large SZAs ($>95^\circ$) and better sensitivity to tropospheric species; however, the radiation transport calculations are even more complex than in the case of ZSL/DOAS.

3.5 Satellite-based Earthshine Differential Optical Absorption Spectroscopy

Spectroscopy of sunlight back-scattered from the earth's surface and atmosphere allows determination of trace gas column densities in a way similar to ZSL/DOAS. VCDs are derived by applying appropriately calculated AMFs. As the light traverses the atmosphere twice the AMFs are larger than in the case of ZSL/DOAS for the same SZA. In contrast to ZSL/DOAS the instrument cannot determine the part of the trace gas column below tropospheric clouds. Therefore the presence of clouds has to be detected by the techniques described in section 3.6. An example of this approach is the GOME on board the ERS-2 satellite.[39]

3.6 Determination of the Geometrical Photon Pathlength of Cloud-penetrating Sunlight

The usual goal of DOAS measurements is the determination of the unknown concentration of trace gases where the (differential) absorption cross-section and the length of the light path are known. However, Equation (2) can also be solved for an unknown light path L when the concentration of the absorber is known. Using oxygen (O_2), tropospheric ozone, or oxygen dimer $(O_2)_2$ absorptions the average lengths of photon paths in clouds or haze can be determined.[37,40,41] Moreover, by analyzing the absorption of individual rotational lines (e.g. of the O_2 a-band around 765 nm) it is possible to infer the distribution of photon pathlengths inside clouds.[42] These data are of importance in studying the internal structure of clouds or for assessing the solar heating of the atmosphere.[43]

Once the optical pathlengths L inside clouds are known this information can be used to determine the concentration of interstitial gases, such as ozone or NO_2 in thunderstorm clouds.[41,44]

4 PRACTICAL DESIGN OF DIFFERENTIAL OPTICAL ABSORPTION SPECTROSCOPY SYSTEMS

Instruments based on spectroscopic techniques have been designed in a variety of ways.[3–7,21,45–62] Depending on the particular type of measurement desired they consist of the following key components.

4.1 System Components

4.1.1 Light Source

An important requirement for DOAS light sources is a minimum of spectral variation of $I_0(\lambda)$, in particular at the scale of molecular band widths (i.e. below about 1 nm). In other words the light sources should emit white light at high brightness. In use are the following types.

4.1.1.1 Thermal Light Sources such as Incandescent Lamps or Arc Lamps Criteria are spectral brightness (watts per unit of radiating area and wavelength interval). This quantity is essentially given by the Planck function and thus the temperature T and emissivity ε of the radiating area ($T \approx 3000$ K, $\varepsilon \approx 1$ for incandescent lamps and $T \approx 6000$–$10\,000$ K, $\varepsilon \ll 1$ for Xenon-arc lamps). The spectrum of incandescent lamps is very similar to a blackbody radiator. Xenon-arc lamps have a much larger brightness and also emit an essentially smooth spectrum in the spectral range of interest, with the exception of a group of emission lines between 400 and 450 nm.[63] For special purposes other types of arc lamps are employed; for instance, good UV intensity is obtained from deuterium lamps.

4.1.1.2 Lasers Lasers give a much lower beam divergence and higher spectral intensity than thermal light sources, but are usually much more complex devices (although this may change with the availability of room-temperature diode lasers for the visible and UV spectral regions). Problems are due to the narrow spectral emission bandwidths or tuning range, which makes it usually difficult to observe molecular bands. Therefore, lasers have rarely been used as DOAS light sources.

4.1.1.3 Light Emitting Diodes Light emitting diodes (LED) are potentially useful light sources as they emit a generally smooth spectrum with an emission bandwidth of several 10 nm, also there is a steady improvement of LED brightness (recently marked by the introduction of GaN devices). LEDs are available throughout the visible spectral range (starting with blue at about 400 nm) into the near-infrared (IR). Although the emitted power is orders of magnitude smaller than from lasers and thermal light sources, the emission only occurs in a relatively narrow band and – more importantly – from a very small area. In fact the energy emitted from a LED per unit area can be comparable with at least an incandescent lamp.

4.1.1.4 Light Sources Outside the Atmosphere The largest problem with such light sources (sunlight, moonlight, or starlight) is their structured spectra which contain many Fraunhofer lines (section 3.3).

4.1.2 Optical Elements Adapting the Absorption Path to Light Source and Spectrometer

4.1.2.1 Long-path Systems In long-path or ZSL/DOAS systems a receiving telescope of sufficiently large aperture (usually 150–300 mm) is required to collect the light. The telescope has to be precisely pointed at the light source or retroreflector array (section 3.2), this can usually only be accomplished by active control of the telescope. Note that even a very rigid mounting of the optical building blocks of the system will usually not solve the problem because, due to an 'El Mirage' effect, the light path will bend due to changing thermal gradients in the atmosphere. For ZSL/DOAS instruments much smaller telescopes (aperture several 10 mm) are sufficient. In either case the aperture ratio of the telescope must match that of the spectrometer (or of the connecting quartz fiber).

4.1.2.2 Quartz Fibers Quartz fibers are conveniently employed to connect the receiving telescope to the spectrometer. This arrangement gives freedom in the placing of the components, so that the telescope can be placed outside while the spectrometer is sheltered. However, optical fibers can be an important component, improving the overall sensitivity of the system considerably. Nonuniform illumination of the field of view of the spectrometer can cause residual structures in the spectra, which can limit the minimum detectable optical density to several 10^{-3}.[64] Introducing mode mixers into the quartz fiber can greatly reduce this effect.

4.1.2.3 Transfer Optics Transfer optics are frequently needed to match the aperture ratio of the optical building blocks e.g. telescope to spectrometer, quartz fiber to spectrometer or a multireflection cell to the spectrometer.

4.1.3 Spectrometers

Depending on the wavelength and sensitivity requirements, a wide variety of detector designs are used, ranging from nondispersive semiconductor or photomultiplier detectors to spectrograph–photodiode array combinations. A number of spectrometer designs are in use.

4.1.3.1 Czerny–Turner This design (and its variants) are traditionally selected for DOAS instruments.[3,53] As well as the grating at least two additional reflecting surfaces, the collimating mirror and the camera mirror are required. The additional reflection losses are usually offset by high-efficiency (60–80%) plane blazed gratings. The wavelength range is readily selected by rotating the grating.

4.1.3.2 Flat-field Spectrometers Modern holographic grating technology[65] allows the design of spectrometers consisting of only entrance slit, grating, and detector (i.e. optomechanical scanning device or diode array; section 4.1.4). The present low efficiency of most holographic gratings (20–40%) is approximately compensated by the reduced losses due to the low number of optical surfaces. Although the stray light produced by holographic gratings itself is lower compared to ruled gratings, the higher intensity of unwanted orders may increase the stray-light level of the whole spectrometer. However, the unused orders are also focused and thus are relatively easily removed by light traps. A disadvantage of most flat field gratings is their limitation to a relatively small spectral range that cannot normally be changed.

4.1.3.3 Fourier-transform Spectrometers Although this type of instrument has been used for IR absorption spectroscopy for a long time, its application for DOAS in the UV–visible spectral range is quite recent.[66]

Regardless of the type of spectrometer selected, stray light requires careful consideration; its significance for DOAS systems is discussed in section 4.2.

4.1.4 Radiation Detectors to Measure $I(\lambda)$ and $I_0(\lambda)$ Spectra

These devices must rapidly measure a spectrum (quasi) simultaneously to minimize the additional noise arising from atmospheric turbulence; good quantum efficiency is also desirable. However, the most significant requirement is the ability to reliably detect relative intensity differences within the spectrum of 10^{-4}. Presently two different detector designs are in use.

4.1.4.1 Optomechanical Scanning Devices These scan a section of the dispersed spectrum (typically from several nanometers to several tens of nanometers) in the focal plane of the spectrograph by a moving exit slit. In the most common design a series of slits are etched radially on a thin metal disk (slotted disk) rotating in the focal plane. At a given time, one particular slit serves as an exit slit where the start of a scan is signaled by an IR light barrier. The light passing through this slit is received by a photomultiplier tube (PMT) the output signal of which is digitized by a high-speed analog-to-digital converter (ADC) and recorded by a computer. The signal is adapted to the ADC range by variable preamplifier gain and/or changing the PMT high voltage. During one scan (i.e. one sweep of an exit slit over the spectral interval of interest), several hundred digitized signal samples are taken. Consecutive scans are performed at a rate of typically 100 scans (±0.1%) per second and are digitized into several hundred channels, which are signal averaged by the software (thus individual samples of the signal are recorded at several tens of kilohertz), thereby meeting the requirement to detect very small signal differences. Alternatively, vibrating exit slits can be used. The disadvantages are less uniform scanning speeds and possible influences of the vibration on other parts of the optics.

As a single scan takes ≈10 ms, the effect of atmospheric scintillations is negligible, because the frequency spectrum of atmospheric turbulence close to the ground peaks at around 0.1–1 Hz and contains very little energy at frequencies above 10 Hz. In addition, typical spectra obtained during several minutes of integration time represent an average over 10 000–40 000 individual scans. Thus, effects of noise and temporal signal variations are very effectively suppressed; even momentary blocking the light beam (e.g. when a vehicle is driven through the beam) has no noticeable effect on the spectrum.

The main disadvantage of optomechanical scanning of the absorption spectra is the large multiplex loss, as only the section of the spectrum passing through the exit slit reaches the detector. Thus only of the order of 1% of the light is utilized, which makes relatively long integration times necessary. In addition, the available PMT photocathodes (multialkali, S20) limit the spectral range to the longer wavelengths of the visible range.

4.1.4.2 Solid-state Multichannel Detectors Solid-state detectors consisting of several hundred individual photodetectors arranged in a linear row are, in principle, far superior for recording absorption spectra. Additional advantages are the absence of moving parts, (nearly) simultaneous sampling of all wavelength intervals, no high voltage is required, and the quantum efficiency (at least in the red and near-IR spectral regions) is better than that of most PMT cathodes.[67] Figure 7 shows the cross-section of a typical silicon photodiode array detector system (using a Reticon RL 1024 R detector chip). It consists of 1024 diodes, each measuring about 0.025 mm × 2.5 mm integrated on a single silicon chip. Depending on the read-out mode, solid-state detectors are distinguished as diode arrays (read-out by sequentially connecting each diode individually to the read-out amplifier via a common rail) and charge-coupled devices (CCD; here read-out is performed by shifting the charge through all diodes to a collecting electrode connected to

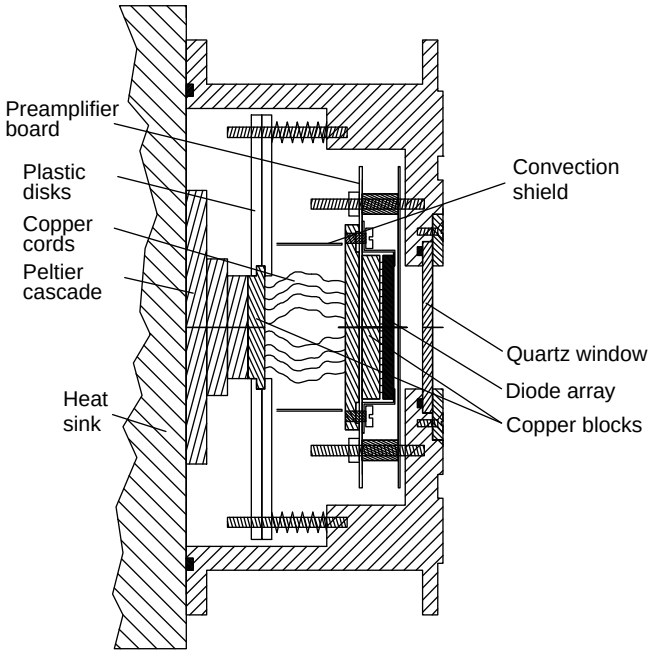

Figure 7 Cross-section of a solid-state photodetector (photodiode array) housing.[68] An air-cooled three-stage Peltier cascade allows cooling to about −40 °C when the housing is filled with argon.

the read-out amplifier). CCD detectors usually contain two-dimensional arrays of pixels, and are thus able to simultaneously record several spectra. The detectors are easily adapted to varying light levels by changing the integration time (from a fraction of a second to a few seconds in artificial light source systems, to minutes when recording zenith scattered light).

Although solid-state array detectors have clear advantages, they pose some problems when used to record weak absorption features in DOAS instruments. First, the signal dependence of the photodiode dark current has to be considered. This apparent paradox is due to the different average voltages across the diode junctions during exposures, when the array is darkened or irradiated, respectively.[68] As the dark current decreases exponentially with temperature it can always be reduced to negligible levels by cooling the array to low temperatures (−40 to −80 °C). Another problem is the presence of Fabry–Perot etalon structures in the recorded spectra, which are caused by the protective overcoat of many commercially available detector arrays and also by possible vapor deposits on the array.[68–70] Particular problems arise from changes in the thickness of the layer of vapor deposit, causing continuous changes in the etalon modulation of the array sensitivity. Diode sensitivity (and dark current) varies by more than ±1% between individual diodes, and this (usually constant) fixed pattern has always to be compensated for (see sections 5 and 6).

Diode array detectors have clear advantages in applications where strong absorptions ($D > 10^{-3}$) occur, or where only low light levels are available (as in zenith scattered light observation of stratospheric species). For the observation of very low optical densities ($D \approx 10^{-2}$–10^{-4}) optomechanical scanning devices give comparable results; however, compared to diode arrays, inferior light utilization necessitates longer exposure times.

4.2 Design Considerations

DOAS has become a versatile technique adapted to many different measurements problems. Once the basic arrangement has been selected, various decisions and tradeoffs have to be made in the design of a DOAS system:

- Choice of spectrometer type and the wavelength range (fixed, variable). This choice is largely dictated by the range of species to be measured (see Figure 8 and Table 1).
- Choice of the light path arrangement (long path, folded, see Figure 3). As outlined in section 3.2 this depends on the desired measurement volume, but also on the required wavelength range.
- Choice of light source.
- Choice of detector type (optomechanical or solid-state).
- Determination of the proper tradeoff between light throughput and size, weight, and cost of the optics.
- Proper tradeoff between spectral resolution and size, cost of the spectrometer.

Some guidelines can be given for the last two points. It can be shown that the amount of light received in a long-path DOAS system with sufficiently large optics is proportional to:

- the luminous intensity of the light source;
- the dispersion of the spectrometer;
- the aperture of the spectrometer;
- the area covered by retroreflectors (if applicable).

However, no gain in light throughput is achieved by increasing the power of the light source (at constant luminous intensity) or by increasing the size of the receiving telescope beyond a certain limit.[71]

Care must be taken that the instrument transmission and the spectral response curve of the detector (as summarized by $A(\lambda)$ in Equation 6) are only slowly varying functions of the wavelength. The remaining undesired influences on the shape of the spectrum are then eliminated by the mathematical treatment of the spectrum, as discussed in section 6.

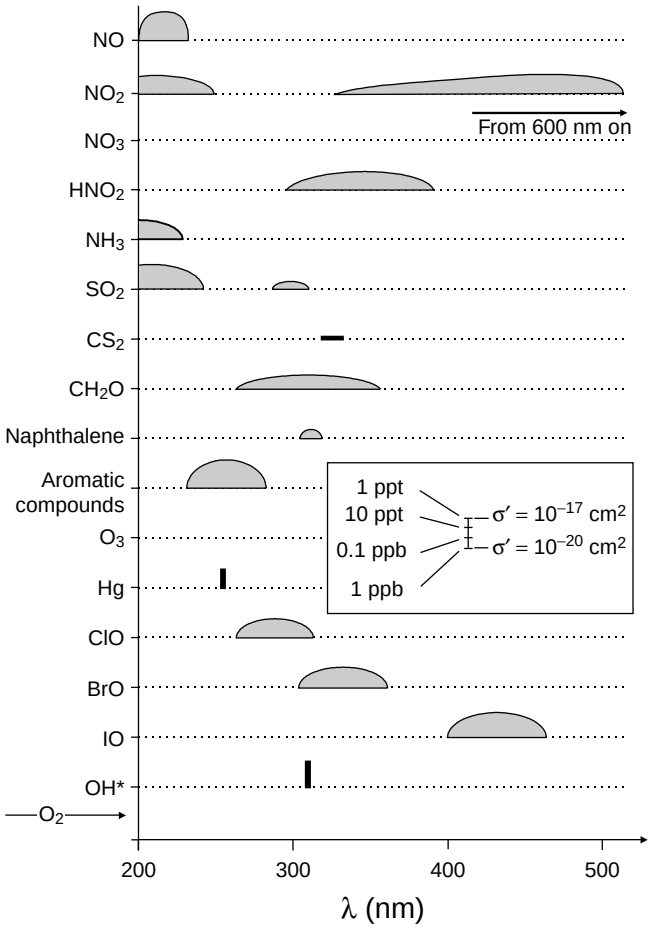

Figure 8 Overview of UV/visible spectral ranges useable for the detection of atmospheric trace gases. The insert is the log of absorption cross-section of the molecule (10^{-20}–10^{-17} cm^2 per molecule), which also indicates the approximate detection limit with a 5 km light pathlength (1 ppt to 1 ppb). Molecules only measurable by high-resolution DOAS are marked by an asterisk (*).

Another point that requires careful consideration when designing a DOAS system is spectrometer stray light. Sources of stray light include: light scattered from optical elements (grating and mirrors), reflection of unused diffraction orders off the spectrometer walls, reflection of unused portions of the spectrum from walls near the focal plane, and reflections from the detector surface.[72]

Stray light tends to be comparatively high in spectrometers filtering a relatively broad wavelength interval from a continuous spectrum, as in DOAS applications. To illustrate this point, consider a typical stray light level of 10^{-5} for a Czerny–Turner spectrometer.[72,73] This figure gives the fractional light intensity found anywhere in the spectrum, when essentially monochromatic light (e.g. of a He-Ne laser) enters the spectrometer. The actual width of the line seen in the focal plane will then just be equal to the spectral resolution of the instrument. Continuous light entering the spectrometer can be thought of as being composed of a series of lines spaced at center-to-center distances equal to their width (i.e. the spectral resolution of the instrument). As the total spectral range of light entering the spectrometer (for instance 300–600 nm) is of the order of 1000 times larger than the spectral resolution (typically 0.3 nm for a low-resolution DOAS instrument) the level of stray light (per wavelength interval) is roughly three orders of magnitude larger than would be expected from the single line definition usually considered. Thus, in DOAS applications stray light levels I_{SL}/I are can be expected to be closer to 10^{-2} than to 10^{-5}. For example, a stray light level of 10^{-2} would reduce a true optical density of $D = 1.00 \times 10^{-3}$ to an apparent $D_{SL} \approx 0.99 \times 10^{-3}$, thus introducing a 1% error into the result.[71]

Although this might be considered a tolerable error, stray light can become a major problem at short wavelengths because most light sources have their maximum intensity in the visible part of the spectrum with a sharp drop throughout the UV range. Therefore, below about 250 nm the stray light will normally exceed the desired signal, leading to large errors unless additional wavelength-selective elements (e.g. filters) are used.

4.3 Example Systems

A state-of-the-art design using a coaxial merge transmitting and receiving telescope in conjunction with an array of corner-cube retroreflectors is shown in Figure 5. Unattended operation is possible by stepper motor-driven change of spectrometer wavelength setting, introduction of shutters and filters into the light path, and activation of a shortcut optics to record lamp intensity spectra. Automatic alignment of the telescope also allows compensation for small movements of the instrument support and diurnal changes in the temperature gradient in the atmosphere. In addition, different light paths can be selected by pointing the telescope to retroreflector arrays mounted at different distances.

The primary mirror diameter and focal length are 300 mm and 1800 mm, respectively. Smaller instruments with accordingly reduced light throughput and sensitivity are commercially available.

Scattered light DOAS systems are in comparison much simpler (Figure 6), only requiring a spectrometer and a small telescope pointed to the zenith. (Note that the telescope cannot improve the brightness of the sky, but rather will reduce the aperture of the system and thus make radiation transport calculations simpler.)

For the measurement of OH radicals a special, high-resolution DOAS system has been developed,[74] this instrument uses a wide-band dye laser as light source. In combination with a high-resolution spectrometer and innovative measures to reduce the instrument-induced

spectral structures,[74] individual rotational lines of the OH molecule with optical densities of a few times 10^{-5} can be recorded at a spectral resolution of about 10 pm. Coupled to a multireflection cell with a base path of 40 m for a total light path of several kilometers the resulting detection limit is of the order of 10^6 OH radicals per cubic centimeter.

4.4 Software Issues

The operating and evaluation software of a state-of-the-art DOAS system is an essential part of the system. The tasks include:

- Recording and storage of the spectra and associated information such as meteorological data to give at least the air pressure and temperature required to convert measured concentrations into mixing ratios.
- Cycle through measurement sequences (e.g. take atmospheric and background spectra, section 5.1) and calculate the true atmospheric spectra (Equation 8), make measurements in different wavelength regions, switch to different light paths etc.
- Align optics, for instance in long-path systems the telescope has to be precisely pointed at the light source or retroreflector array as described in section 4.1.2.
- Evaluate spectra (section 6) and display trace gas concentrations. This is the most complex task and is frequently performed off-line.

Several software packages have been developed to perform some or all of the above tasks.[62,75,76]

5 OPERATION AND CHARACTERIZATION OF DIFFERENTIAL OPTICAL ABSORPTION SPECTROMETRY SYSTEMS

5.1 Typical Measurement Sequences

A complete DOAS measurement involves taking a series of several different spectra. The example given here is for a system using a diode array detector and an artificial light source, but other arrangements follow similar procedures:

1. Record the atmospheric raw spectrum (ARS, denoted $I_M(i)$); several (N_M) individual spectra with integration time T_M are usually co-added.
2. Record the background spectrum ($I_B(i)$) with integration time T_B. As explained in section 5.2, scattered sunlight might disturb the measurements. A scattered sunlight or background spectrum is taken with the telescope pointed beside the light source.
3. Record the lamp intensity spectrum (LIS, denoted $I_L(i)$). This spectrum is recorded by admitting direct light from the lamp to the spectrometer (using some optics to bypass the light path).
4. Record the dark current (and electronic offset) spectrum ($I_D(i)$). Normalizing the sum of the N_D co-added dark current spectra gives $I_{D1}(i) = I_D(i)/N_D$.

The true atmospheric spectrum $I(i)$ can then be calculated from the above series of spectra from Equation (8):

$$I(i) = [(I_M - N_M I_{D1}) - T_M/T_B(I_B - N_B I_{D1})] \times (I_L - N_L I_{D1})^{-1} \quad (8)$$

where N_B and N_L denote the numbers of co-added background and lamp intensity spectra, respectively.

Note that background spectra need not always be taken and compensated for, for instance at night or at wavelengths below 300 nm where scattered sunlight levels are low. Also, although lamp intensity spectra are always required to compensate for diode-to-diode sensitivity variations (section 4.1.4), it might not be necessary to take one with each ARS.

An altogether different approach to taking lamp intensity spectra is the multichannel scanning technique (MCST).[77,78,70] In MCST a series of (typically about 10) ARSs is taken and stored. They are then successively shifted in wavelength (for instance by slightly rotating the grating) in such a way that the spectral information is essentially smoothed out when they are co-added (after subtracting I_D) to give a LIS. Now each ARS is divided by the thus obtained LIS and digitally shifted back to the same wavelength position (for instance of the first ARS in the series) to obtain a true atmospheric spectrum. The advantage of MCST is that the diode-to-diode sensitivity variations are even more precisely eliminated, because spectra taken to determine these variations are recorded under precisely the same conditions as the ARS. Measurement time is also saved, as no separate LIS has to be taken.

5.2 Scattered Light with Artificial Light Sources

During daytime sunlight scattered within the measurement volume (not to be confused with stray light inside the spectrometer) may also reach the spectrometer. In contrast to a more or less continuous offset caused by spectrometer stray light, scattered sunlight exhibits the strong Fraunhofer structure (containing signatures of, e.g. stratospheric O_3 and NO_2; section 3.3), which can disturb measurements even at stray light levels as low as a fraction of a percent of the received light source intensity. Fortunately, due to the very narrow field of view of the instrument (of the order of 10^{-9} st), levels of

scattered sunlight are usually found to be quite low, with the exception of hazy conditions.

In the latter cases, however, the scattered light is easily subtracted from the total spectrum by frequently interrupting the measurement and taking a scattered light (i.e. background) spectrum with the telescope pointed beside the light source. As only an offset of a fraction of a degree is required, this leaves the amount of scattered light entering the system practically unchanged.

In folded light path DOAS systems scattered light from the artificial light source can cause also problems. In arrangements using retroreflector arrays some of the light might be scattered directly into the spectrometer and thus bypass the light path. Fortunately, this type of scattered light is eliminated by the procedure described above.

Somewhat less satisfactory is the situation in multireflection cells. Although the amount of scattered sunlight can be readily determined by blocking the light source, deadjustment procedures to determine the fraction of light bypassing the full light path are difficult to apply.

5.3 Instrument Function

A spectrometer is characterized by its response $W(\lambda)$ to monochromatic signal. The half-width Γ_o of $W(\lambda)$ is frequently called the spectral resolution of the spectrometer. At the typical resolution of a DOAS spectrometer (a few 10ths of a nanometer) atomic emission lines (e.g. of a Hg lamp with typical linewidths of the order of 1 pm) are a sufficiently good approximation to a monochromatic light source. Thus $W(\lambda)$ can be readily determined for each instrument during the characterization procedure; note that W might be a function of the wavelength. The instrument function is then used to calculate the proper $\sigma'(\lambda)$ and thus allows reliable absolute calibration of any DOAS instrument (see Equation 9 in section 6.1).

5.4 Residual Spectral Structures

The spectral structures that remain after all known spectral features have been removed from the measured spectrum (see section 6 and Figure 9, bottom trace) can be due to several physical possibilities:

- Absorbing species present in the atmosphere, that are unknown or unaccounted for.
- Unknown instrumental features.
- Random noise (due to photoelectron statistics)

Although the last possibility can be tested for (and can be reduced by co-adding more spectra), the first two may ultimately determine the detection limit of the DOAS system. Techniques for treating residual spectral structures as randomly occurring features are described elsewhere.[70,79]

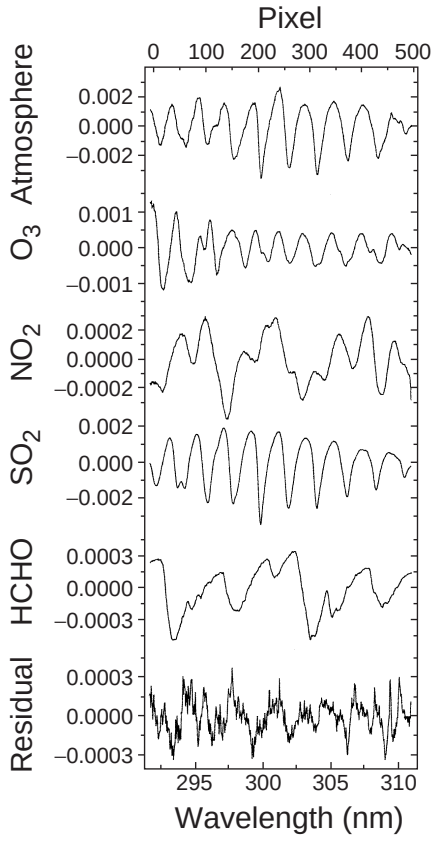

Figure 9 Deconvolution process of an atmospheric spectrum (uppermost trace) recorded in Heidelberg on 27 August 1994, with overlapping absorptions due to O_3, NO_2, SO_2, and HCHO (traces 2–5 from the top, note different scales). The least-squares fitting procedure determines the four weighting factors for the above reference spectra in order to model the measured spectrum. From these factor mixing ratios of 16.6 ± 0.8, 4.6 ± 0.6, 1.25 ± 0.02, and 2 ± 0.3 ppb were derived for the above species, respectively. Further fitted parameters compensate for slight errors in the wavelength–pixel mapping.[70] In addition a fifth-order polynomial was fitted. The remaining residual spectrum after removal of the absorption structures is shown as the bottom trace.

5.5 Characterizing

The most important steps in characterizing a DOAS system include determination of the following quantities:

- The instrument function W (and thus the spectral resolution) of the spectrometer.
- The stray light level of the spectrometer.
- The light throughput of the system (e.g. in photons per pixel and per second) under defined conditions (e.g. atmospheric extinction and scattering properties).
- The residual spectral structure of the system.

6 EVALUATION TECHNIQUES FOR DIFFERENTIAL OPTICAL ABSORPTION SPECTRA

As the absorption cross-section σ and the light path L in Equation (2) are usually known. In practice two additional problems arise:

- the appropriate differential absorption cross-section must be found;
- the I'_0 spectrum has to be determined.

In addition, in certain spectral regions overlapping absorptions by several trace species may occur (Figure 9), which must be separated. This problem is discussed in section 6.7.

6.1 Preparation of Reference Spectra

In principle there are two ways to determine reference spectra for DOAS spectrometers. The first uses a reference cell filled with a known amount of the species under consideration in the light path of the spectrometer. This procedure has the advantage that it yields spectra exactly as seen in the atmosphere; however, it is usually practical only for stable species. A cell filled with an unknown amount of a given species can still be used to obtain relative values of $\sigma(\lambda)$ (i.e. values known to a constant factor). Relative reference spectra thus derived can be used in a least-squares fitting procedure (section 6.3) to completely account for the presence of a species in the spectrum. However, in order to determine the concentration of that species, an absolute value for σ is required.

In the second method reference spectra taken at a high (in principle infinite) resolution compared to the resolution of the DOAS instrument $\sigma_{\mathrm{hr}}(\lambda)$ are folded with the instrument function $W(\lambda)$ (see section 5.3) to obtain the effective absolute absorption cross-section as seen by the DOAS spectrometer (Equation 9):

$$\sigma(\lambda) = \int \sigma_{\mathrm{hr}}(\lambda) W(\lambda - \lambda')\, \mathrm{d}\lambda' \qquad (9)$$

The differential absorption cross-section $\sigma'(\lambda)$ is then determined from the thus-calculated spectrum $\sigma(\lambda)$.

6.2 High-pass Filtering of Spectra

The spectrum I'_0 is found by digitally high-pass filtering the spectra during the evaluation procedure. The cut-off frequency of the filter in wavelength space is chosen so as to minimize attenuation of the molecular absorption bands to be detected. As the width of these bands is roughly 1 nm, this suggests cut-off frequencies of the order of a fraction of a nm^{-1}. However, the slowly varying slope of the spectra should be removed as completely as possible. There are many possible choices of suitable filtering techniques:

- Experience shows that the slowly varying light extinction factors $\exp\left(-\sum \sigma_{i0}(\lambda) c_i\right)$, $\exp(-\varepsilon_{\mathrm{R}}(\lambda))$, $\exp(-\varepsilon_{\mathrm{M}}(\lambda))$, and $A(\lambda)$ of Equations (5) and (6) are reasonably approximated by a least-squares fitted polynomial $P(\lambda) = \sum p_n \lambda^n$ of suitable order (usually fifth order[3]). Thus the expression for $I'_0(\lambda)$ given in Equation (6) becomes (Equation 10):

$$I'_0(\lambda) = I_0(\lambda) \exp\left[-L \left(\sum (\varepsilon_{i0}(\lambda) \mathrm{c_i}) + \varepsilon_{\mathrm{R}}(\lambda) + \varepsilon_{\mathrm{M}}(\lambda)\right)\right] A(\lambda) \approx P(\lambda) \qquad (10)$$

  In the case of only a single absorbing species this corresponds exactly to Equation (1) with $P(\lambda)$ as an approximation for $I_0(\lambda)$. Then the optical density D' can then be determined from the light intensities at either side of one of the bands as well as in the center of the band (see Figure 2).
- Division by a digitally smoothed copy $S(\lambda)$ of the spectrum can be used instead of dividing by $P(\lambda)$.
- Fourier transform techniques have been applied to obtain suitable filtering of the spectra.[66]

Note that this filtering procedure is not usually performed as a separate step, but rather as a part of the overall fitting process with $P(\lambda)$ sometimes being called a closure term.

6.3 Dealing with Spectral Shift and Changes in Dispersion

Different spectra, even those recorded with the same instrument, may show slightly different wavelength calibration. If $I(i)$ is the recorded light intensity in the ith channel (pixel; i.e. photodiode of a detector array, see Figure 1), then in different spectra $I(i)$ might not cover exactly the same wavelength interval $(\lambda(i) - \Delta\lambda, \lambda(i) + \Delta\lambda)$. For instance, the center wavelength $\lambda(i)$ might show a constant offset $\delta\lambda$ for all values of i (which would be called a shift of the spectrum) or, more complicated, the offset $\delta\lambda = \delta\lambda(i)$ could be a function of i itself (usually described by a polynomial $\delta\lambda(i) = \sum a_j i^j$).

Although this effects of spectral shift and squeeze can also be readily compensated by the software, these features introduce additional parameters in the fit and thus increase the uncertainty of the desired trace gas column densities (section 6.6).

6.4 Spectral Fitting Procedures

The central part of spectral evaluation consists of determining the parameters c_i in Equation (3), which are

proportional to the trace gas concentration. In addition, parameters describing the filtering of spectra and spectral distortion (section 6.5) are determined. In essence the spectral fitting procedure models the measured spectra as the sum of reference spectra, a general spectral slope (which slowly varies with wavelength), and possible spectral shift and changes in dispersion. Although only the parameters c_i are of direct importance, the other parameters can also give information, such as the polynomial parameters on the aerosol contents or the parameters describing spectral shift (section 6.3) on drift effects in the instrument.

The fitting procedure can be transformed into a linear problem if $J(i) = \log(I(i))$ is modeled instead of I, thus the parameters c_k describing the trace gas concentrations (or column densities) and p_i characterizing the polynomial can be obtained by a straightforward matrix inversion procedure.[15,70] In contrast, the parameters a_i that describe spectral shift and distortion (section 6.3) can only be determined by iterative methods, such as Levenberg–Marquardt techniques,[70] so that the determination of these nonlinear parameters can easily dominate the processing time in DOAS spectral fitting procedures. Here advanced computational techniques can improve the processing speed by orders of magnitude.[76]

6.5 Ring Effect and Related Problems with Solar Scattered Light

As discovered by Grainger and Ring[80] Fraunhofer lines observed at large SZAs appear somewhat weaker (i.e. are filled in) than the same lines at small SZAs. Meaningful solar scattered light DOAS measurements can only be made if this effect is compensated, otherwise complete removal of Fraunhofer lines by division of spectra taken at small and large SZAs, respectively, is impossible. Fish and Jones[81] and Burrows et al.[82] present good evidence that rotational Raman scattering is the main cause of the ring effect. As pointed out by Fish and Jones[81] the ring effect will also reduce the strength of absorption lines due to atmospheric trace species. This will in particular affect relatively narrow lines (width of the order of the wavelength shift due to rotational Raman scattering, i.e. roughly 1 nm). For the NO_2 bands around 450 nm the authors calculate an underestimation of the true line strength of up to 7%; the effects on BrO and OClO are probably of similar magnitude. In addition, multiple scattering due to aerosol scattering might lead to further underestimation of the absorption line strength and thus the derived trace gas column densities.

Vibrational Raman scattering is also nonnegligible. Its effect is that Stokes lines of O_2 and N_2 of strong Fraunhofer lines will appear as red-shifted vibrational Raman ghost (VRG) lines in the spectrum. It could be argued that those lines are an always occurring part of the Fraunhofer spectrum, but this is not true. First, the angular dependence of RS and Raman scattering are different, so the relative contribution of Raman light is twice as large at $SZA = 90^\circ$ compared to $SZA = 0^\circ$. Second, there may be multiple scattering enhancing the VRG. According to calculations by Haug[83] VRG may show up as lines with optical densities of up to 0.003.

6.6 Realistic Estimation of Errors

The main sources of error in the derived trace gas column densities or concentrations are:

- statistical errors of the fitting procedure;
- uncertainty of the differential absorption cross-section;
- the uncertainty of the AMF (in the case of scattered-light DOAS).

The error of the fitting procedure normally calculated assumes that the noise in the spectra follows a Gaussian distribution and that the individual pixels are statistically independent. Both criteria are only approximately fulfilled; for instance, low-path filtering of a spectrum (digital smoothing) prior to the fitting procedure will increase the correlation between the individual channels and thus lead to a gross underestimation of the statistical fit error.

In addition the effects of spectral shift and squeeze, although readily compensated by software, introduce additional parameters in the fit and thus increase the uncertainty of the desired trace gas column densities.

A detailed analysis of the problem, taking the above point into account, shows that the true uncertainty of the optical density due to a given trace gas might be several times larger than the statistical error of the associated fitting parameter alone.[70]

The problems of aliasing are well known from many areas of electronic signal evaluation. It will arise in cases where the spectral resolution is too high for a given number of detector pixels in the focal plane of the spectrometer. Although it is tempting to make ''best possible use'' of a given number of pixels, there must be a sufficiently large number of signal samples $I(i)$ (or detector pixels) across one resolution interval. Sample calculations show that about 5–10 samples per resolution interval lead to acceptable errors.[70,84] In principle advanced interpolation techniques could be used to correct for this problem (to some extent), but to the author's knowledge this has not yet been attempted in DOAS spectra evaluation.

6.7 Interferences

If the absorption features of several species overlap in the same spectral region, a simultaneous least-squares

fit of the respective absorption spectra is usually used to separate the contributions of the individual species (section 6.4). The fit coefficients will then be a measure for the $\sigma'_k c_k$ products of the absorbing species. As Equation (2) is nonlinear in intensity I, the logarithm of $I(\lambda)/I'_0$ is usually formed, thus allowing a (mathematically simpler) linear fit.

Examples of the deconvolution process are shown in Figures 9 and 10. Figure 9 shows a deconvolution process of an atmospheric spectrum (uppermost trace) recorded in Heidelberg on 27 August 1994 with a 7.5 km light path, with overlapping absorptions due to O_3, NO_2, SO_2, and HCHO (formaldehyde). The least-squares fitting procedure determines the four weighting factors for the above reference spectra in order to model the measured spectrum. From these, factor mixing ratios of 16.6 ± 0.8, 4.6 ± 0.6, 1.25 ± 0.02, and 2 ± 0.3 ppb, respectively, for the above species were derived. Further fitted parameters compensate for slight errors in the wavelength–pixel mapping.[70] In addition a fifth-order polynomial was fitted. The remaining residual spectrum after removal of the absorption structures is shown as the bottom trace.

A section of the OH absorption spectrum, is presented in Figure 10 together with the spectra of several interfering species.[85,86] It can be seen that the interference by those species can be readily eliminated, because the OH spectrum consists of several isolated groups of lines, whereas the interfering absorption features are more or less evenly distributed over the observed spectral interval. Therefore the strengths of the interfering spectra can be quite accurately measured in the gaps between the OH lines. Once the strength is known the interfering spectra are easily subtracted from the observed spectra.

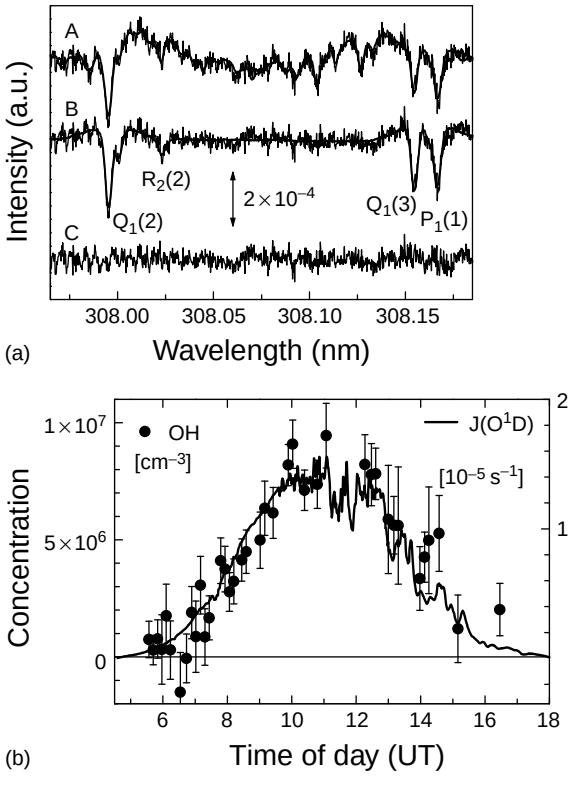

Figure 10 (a) Atmospheric OH absorption spectrum (uppermost trace, A) measured on 9th August 1994 with a 3 km light path. The center (trace B) shows the same spectrum after subtraction of the spectra of several interfering species (SO_2, CH_2O, naphthalene, etc.). An OH reference spectrum (corresponding to 9.6×10^6 OH cm^{-3}) is superimposed (thin line). Trace C indicates the residual spectrum after removal of all known absorption features. The vertical arrow indicates a differential optical density of 2×10^4. (b) Diurnal profiles of OH (filled dots with error bars) and the photolysis frequency of ozone (drawn line) on 17th August 1994. (Reproduced by permission of American Geophysical Union from Dorn et al.[74])

7 SENSITIVITY AND DETECTION LIMITS OF DIFFERENTIAL OPTICAL ABSORPTION SPECTROSCOPY

7.1 Species Measurable by Differential Optical Absorption Spectroscopy

A summary of the species detectable by DOAS is given in Table 1 and Figure 8, which also give detection limits for typical measurement conditions. The list is constantly growing; for instance Etzkorn et al.[87] report structured absorption spectra of 24 monocyclic aromatic species, which allow their sensitive detection by DOAS.

7.2 Detection Limits

The detection limit for a particular substance (with known differential absorption cross-section) at a given length L of the light path can be calculated according to Equation (2), by substituting the minimum detectable optical density D_o for D.

The accuracy of the derived concentration values thus depends on the residual noise in the spectrum being of the order of the minimum detectable optical density D_o as well as on the error in $\sigma'(\lambda)$. Typical uncertainties of $\sigma'(\lambda)$ are in the 1–10% range. At concentrations close to the detection limit the former source of error will dominate, whereas at high concentrations the accuracy is determined by $\sigma'(\lambda)$. Additional work still needs to be done on the issue of optimum filtering of DOAS spectra.

In general D_o is determined from photoelectron statistics (shot noise), which is a function of light intensity, and residual absorption structures, denoted here as B and being of the order of 10^{-4} for a typical DOAS system. Photoelectron shot noise is proportional to $N^{-1/2}$, where N is the total number of photons recorded around the

center of the absorption line during the time interval t of the measurement (Equation 11).

$$D_o = (1/N + B^2)^{1/2} \tag{11}$$

If sufficient light is available, D_o is limited by B. For DOAS systems using artificial light sources in typical situations an integration time of the order of seconds (for solid-state multichannel detectors) to minutes (for optomechanical detection) should be sufficient. ZSL/DOAS systems require from below 1 min to 10 min integration time, depending on the SZA.

Although for a given minimum detectable optical density the detection limit improves proportionally to the length of the light path, the actual detection limit will not always improve with a longer light path. This is because with a longer light path the received light intensity $I(\lambda)$ is reduced, leading to increased photoelectron shot noise associated with the measurements of $I(\lambda)$ and thus increasing the minimum detectable optical density.

For a typical DOAS system there are no geometrical light losses up to pathlengths of about 2 km (for longer light paths, the size of the light source mirror could be increased). However, in addition to geometrical light losses, light attenuation due to atmospheric absorption as well as to scattering affect the magnitude of the received light signal (Equation 12):

$$I_{\text{received}} = I_{\text{source}} \exp(-L/L_o) \tag{12}$$

where the absorption length L_o reflects the combined effects of broad band atmospheric absorption, MS and RS. The noise level of the detector is essentially dependent on the number of photons received and thus is proportional to the square root of the intensity. The signal-to-noise ratio of a DOAS system as a function of the light pathlength L can be expressed (Equation 13) as

$$\frac{D'}{N} \approx LG(\lambda) \exp\left(-\frac{L}{2L_o}\right. \tag{13}$$

where $G(\lambda)$ varies from a constant at very short light paths L to $G(\lambda) = L^{-1}$ (intermediate L), to $G(\lambda) = L^{-2}$ (at very large L). Equation (13) yields optimum D'/N ratios for light pathlengths of $L = 2L_o$ (small L), $L = L_o$ (intermediate L), or a boundary maximum at the shortest light path (very large L). Although there are no lower limits for L_o in the atmosphere (fog), the upper limits are given by RS, and the scattering by atmospheric background aerosol which indicate optimum light pathlengths ($2L_o$) in excess of 10 km for wavelengths above 300 nm.

Another point to consider is the width Γ of the absorption line (or band). At a given spectral intensity of the light source $N(\lambda)$ the light intensity around the line center will be $N(\lambda)\mathrm{d}\lambda \approx N(\lambda)\Gamma$ (in photons per nanometer and second), thus the detection of a spectrally narrow absorption line requires a proportionally higher spectral intensity of the light source (section 4.1.1).

For a given spectrometer the light throughput varies in proportion the square of the spectral resolution (here expressed as the half-width of the instrument function Γ_o), thus the light intensity at the output becomes $I \approx I_i/\Gamma_o^2$. The signal in the spectrum is the differential optical density D', which is (for a given species, wavelength, and light path) proportional to the concentration and is also a function of the spectral resolution (Figure 11). In the case of shot noise limitation (see above) D_o in the spectrum is inversely proportional to the square root of I, thus in general the signal as a function of the spectral resolution will vary proportionally to $\sigma'(R)$ as a function of resolution. Frequently $\sigma'(\Gamma_o)$ can be approximated as a linear function $\sigma'(\Gamma_o) \approx f(\Gamma_o) \approx \sigma'_o(1 - a\Gamma_o)$. Where a is a constant ($a \approx 1\,\text{nm}^{-1}$). In this case the optimum D'/N is thus obtained at the resolution $\Gamma_{\text{opt}} = 1/2a$.

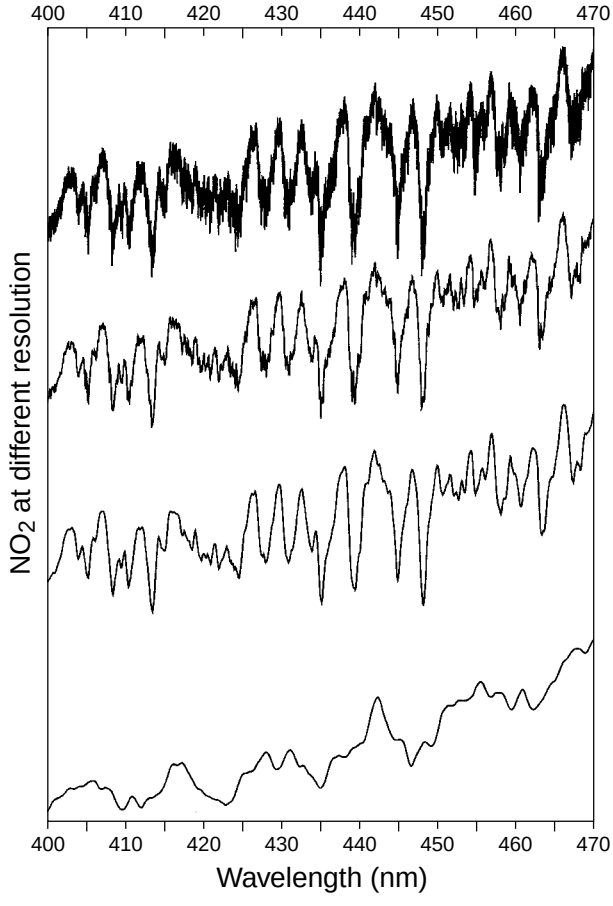

Figure 11 NO_2 spectra at different spectral resolution for the same NO_2 column density. Traces from top to bottom: 0.01 nm, 0.3 nm, 1.0 nm, and 3.0 nm resolution. The apparent absorption cross-section decreases by almost one order of magnitude, when the spectral resolution is reduced from 0.3 to 3 nm.

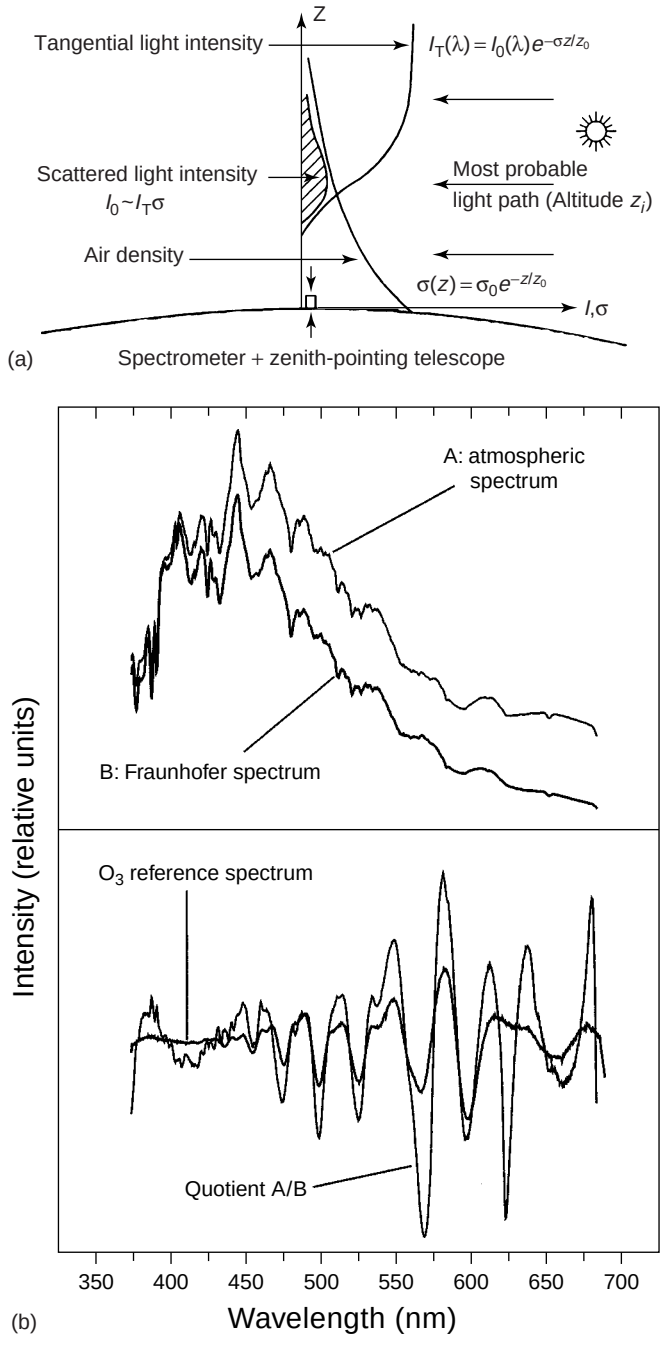

Figure 12 (a) ZSL/DOAS geometry: at high solar zenith angles (around 90° as shown here) ZSL observed from the ground mostly originates from a certain altitude range (most probable light path at altitude z_i). This is due to the tangential light intensity IT increasing with altitude, while the air density (and thus density (s) of air molecules) scattering tangential light down into the instrument, decreases with altitude. (b) ZSL/DOAS spectra: upper panel shows spectra taken at large and small zenith angles, respectively; note the dominating Fraunhofer structure. The lower panel shows the ratio of spectra taken at large/small zenith angles: atmospheric absorption features due to the O_3 Chappuis bands become visible.

7.3 Airmass Factors for Scattered Light Operation

The slant column density (SCD) is an average of the column densities seen by many nearly horizontal rays (at SZA = 90°) traversing the atmosphere at long paths and then being scattered in the zenith, see Figure 12(a) and (b) (In view of the origin of the absorption the term apparent column density (ACD) would describe the situation better.) Interpretation of the SCD (slant column density) ACD, therefore, requires radiation transport modeling.[15,19,88,89] The results of those calculations are usually expressed as the AMF, which describes the ratio of the SCD to the VCD of the trace species. For medium and small SZAs, i.e. $\vartheta \leq 70°$, the AMF can be approximated by $1/\cos(\vartheta)$. At $\vartheta = 90°$ (sunset or sunrise) the AMF typically reaches a value of about 20 for a stratospheric species near 25 km, but the actual value can vary widely as a function of wavelength, vertical trace gas profile, and stratospheric aerosol loading.[15,88,90]

8 APPLICATIONS OF DIFFERENTIAL OPTICAL ABSORPTION SPECTROSCOPY

DOAS applications (Figures 13 and 14) encompass studies in urban air, measurements in rural areas, observations in the background troposphere, as well as investigations of the distribution of stratospheric ozone and species leading to its destruction. In addition, geometric light pathlengths in clouds or haze can be determined. The following discussion gives a few examples.

An important early result obtained with DOAS was the first unambiguous detection of nitrous acid (HONO)

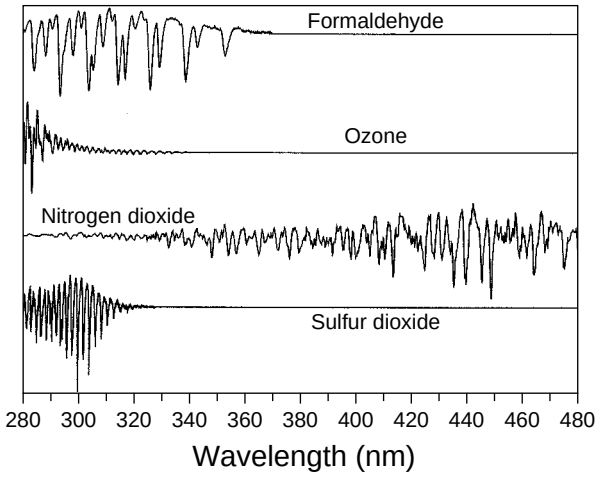

Figure 13 Spectra of CH_2O, O_3, NO_2, and SO_2 in the spectral range 280–485 nm. Absorption strengths are not to scale. Note the strong, virtually interference-free features of NO_2 at wavelengths above 360 nm.

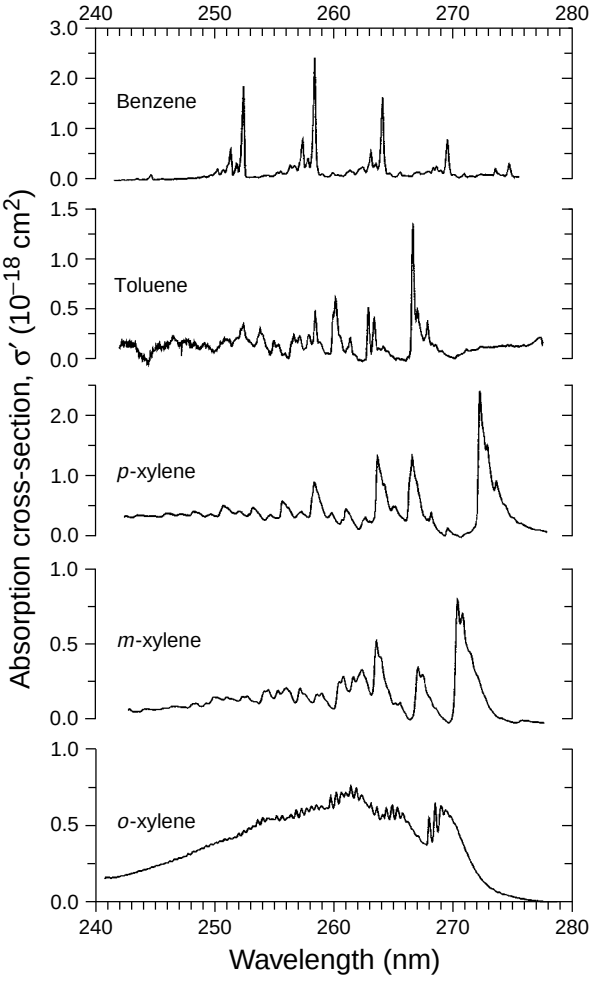

Figure 14 Sample absorption cross-section spectra $\sigma(\lambda)$ of aromatic compounds. In addition to the species shown most monocyclic aromatics exhibit strongly structured absorptions in the spectral region from 230 to 290 nm, allowing their sensitive (low-ppt range) detection in the atmosphere.[87] A point to watch is the structured O_2 absorption due to the forbidden Herzberg I system[100] (not shown).

in urban air.[2] Nitrous acid is produced from NO_2 and water at various types of surfaces. Although many subsequent DOAS investigations confirmed that HONO levels rarely exceed 5% of the NO_2 it is, nevertheless, significant for atmospheric chemistry because its photolysis (Equation 14)

$$HONO + h\nu \longrightarrow OH + NO \qquad (14)$$

leads to the production of OH radicals which, in turn, initiate most chemical degradation processes of air pollutants.

The detection of OH radicals presented a major challenge for DOAS for a long time, because (daytime) atmospheric OH levels are of the order of 10^6 molecules per cubic centimeter (roughly 0.04 ppt). After early attempts[45] steady progress was made[91,92] until excellent atmospheric spectra of OH were reported.[74]

Another radical species, the nitrate radical (NO_3) was also discovered in the stratosphere[50] and troposphere[51,93] by DOAS techniques. Nitrate radicals are formed via oxidation of NO_2 by ozone (Equation 15):

$$NO_2 + O_3 \longrightarrow NO_3 + O_2 \qquad (15)$$

The nitrate radical is a strong oxidant, initiating the degradation of many (unsaturated) hydrocarbons. Some the oxidation products lead to the formation of organic peroxy and HO_2 radicals, which in turn can yield OH radicals. As the formation of NO_3 does not require sunlight, this OH source is also active at night.

Among the first indications of the involvement of chlorine species in the formation of the Antarctic ozone hole was the detection of OClO by ZSL/DOAS.[29] Bromine monoxide (BrO) could also be detected in stratospheric air.[25,33,34,94]

Ground-based DOAS systems with artificial light sources have also detected reactive halogen species in the troposphere, such as bromine monoxide BrO and possibly ClO,[7,95] and IO.[96] The distribution of stratospheric BrO[94] and tropospheric BrO[97,98] was mapped by satellite-borne DOAS. As in the stratosphere, halogen monoxide radicals lead to very efficient, catalytic ozone destruction. A spectacular phenomenon caused by BrO (and possibly ClO) is the complete, episodic destruction of boundary layer ozone during the polar spring (the polar tropospheric ozone hole).[7]

By reversing the usual DOAS approach, so as to determine unknown trace gas concentrations for a known light pathlength, the average lengths of photon paths in clouds could be determined by making use of the known concentrations of oxygen (O_2), tropospheric ozone, or oxygen dimers $(O_2)_2$ (section 3.5).[37,41] By analyzing the absorption of individual rotational lines (e.g. of the O_2 a band around 765 nm) it was possible to infer not only the average light pathlength in clouds but also elements of its distribution.[42] These data give new insights into the internal structure and properties of the radiation field inside clouds.

ACKNOWLEDGMENTS

The help of Kai Hebestreit and Birgit Greiner in preparing the final version of the manuscript is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

ACD	Apparent Column Density
ADC	Analog-to-digital Converter
AMF	Airmass Factor
ARS	Atmospheric Raw Spectrum
CCD	Charge-coupled Devices
DOAS	Differential Optical Absorption Spectroscopy
GOME	Global Ozone Monitoring Experiment
IR	Infrared
LED	Light Emitting Diode
LIS	Lamp Intensity Spectrum
MCST	Multichannel Scanning Technique
MS	Mie Scattering
PMT	Photomultiplier Tube
ppb	parts per billion
ppt	parts per trillion
RS	Rayleigh Scattering
SCD	Slant Column Density
SZA	Solar Zenith Angle
UV	Ultraviolet
VCD	Vertical Column Density
VRG	Vibrational Raman Ghost
ZSL	Zenith Scattered Sunlight

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Automotive Emissions Analysis with Spectroscopic Techniques • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Laser-based Combustion Diagnostics • Photoacoustic Spectroscopy in Trace Gas Monitoring • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Formaldehyde, Environmental Analysis of

Environment: Water and Waste cont'd **(Volume 4)**
Phenols Analysis in Environmental Samples

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Process Instrumental Methods **(Volume 9)**
Ultraviolet/Visible Spectroscopy in Process Analyses

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Satellite and Sensor Systems for Environmental Monitoring • Stellar Spectroscopy

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. C.E. Junge, *Air Chemistry and Radioactivity*, Academic Press, New York and London, Vol. 4, 1963.
2. D. Perner, U. Platt, 'Detection of Nitrous Acid in the Atmosphere by Differential Optical Absorption', *Geophys. Res. Lett.*, **6**, 917–920 (1979).
3. U. Platt, D. Perner, H. Pätz, 'Simultaneous Measurement of Atmospheric CH_2O, O_3, and NO_2 by Differential Optical Absorption', *J. Geophys. Res.*, **84**, 6329–6335 (1979).
4. H. Axelsson, B. Galle, K. Gustavsson, P. Regnarsson, M. Rudin, 'A Transmitting/Receiving Telescope for DOAS Measurements Using Retroreflector Technique, Digest of Topical Meeting on Optical Remote Sensing of the Atmosphere, OSA 4, 641–644, 1990.
5. J.M.C. Plane, C.F. Nien, 'Differential Optical Absorption Spectrometer for Measuring Atmospheric Trace Gases', *Rev. Sci. Instrum.*, **63**, 1867–1876 (1992).
6. H. Edner, P. Ragnarson, S. Spännare, S. Svanberg, 'Differential Optical Absorption Spectroscopy (DOAS) System for Urban Atmospheric Pollution Monitoring', *Appl. Opt.*, **32**, 327–333 (1993).
7. M. Hausmann, U. Platt, 'Spectroscopic Measurement of Bromine Oxide and Ozone in the High Arctic During Polar Sunrise Experiment 1992', *J. Geophys. Res.*, **99**, 25 399–25 413 (1994).
8. F. Heintz, H. Flentje, R. Dubois, U. Platt, 'Long-term Observation of Nitrate Radicals at the TOR-Station Kap Arkona (Rügen)', *J. Geophys. Res.*, **101**, 22 891–22 910 (1996).
9. J.U. White, 'Long Optical Paths of Large Aperture', *J. Opt. Soc. Am.*, **32**, 285–288 (1942).
10. J.U. White, 'Very Long Optical Paths in Air', *J. Opt. Soc. Am.*, **66**, 411–416 (1976).
11. D. Horn, G.C. Pimentel, '2.5 km Low-temperature Multiple-reflection Cell', *Appl. Opt.*, **10**, 1892–1898 (1971).
12. D. Ritz, M. Hausmann, U. Platt, 'An Improved Open-path Multireflection Cell for the Measurement of NO_2 and NO_3', in *Optical Methods in Atmospheric Chemistry*, eds. H.I. Schiff, U. Platt, The International Society for Optical Engineering, Bellingham, USA, 1715, 200–211, 1992.
13. I. Pundt, J.P. Pommereau, F. Lefevre, 'Investigation of Stratospheric Bromine and Iodine Oxides Using the

SAOZ Balloon Sonde', in *Proc. XVIII Quadrennial Ozone Symp. L'Aquila*, Italy, 12–21 Sept., eds. R.D. Bojkov, D. Visconti, 575–578, 1996.

14. F. Ferlemann, C. Camy-Peyret, R. Fitzenberger, H. Harder, T. Hawat, H. Osterkamp, M. Schneider, D. Perner, U. Platt, P. Vradelis, K. Pfeilsticker, 'Stratospheric BrO Profile Measured at Different Latitudes and Seasons: Instrument Description, Spectral Analysis and Profile Retrieval', *Geophys. Res. Lett.*, **25**, 3847–3850 (1998).
15. S. Solomon, A.L. Schmeltekopf, R.W. Sanders, 'On the Interpretation of Zenith Sky Absorption Measurements', *J. Geophys. Res.*, **92**, 8311–8319 (1987).
16. H. Frank, U. Platt, 'Advanced Calculation Procedures for the Interpretation of Skylight Measurements', in *Proc. First European Ozone Workshop*, Schliersee, 65–68, 1990.
17. L.C. Marquard, T. Wagner, U. Platt, 'Improved Approaches for the Calculation of Air Mass Factors Required for Scattered Light Differential Optical Absorption Spectroscopy', *J. Geophys. Res.*, **105**, 1315–1327.
18. A. Sarkissian, D. Fish, M. Van Roozendael, M. Gil, H.B. Chen, P. Wang, J.P. Pommereau, J. Lenoble, 'Ozone and NO_2 Air-mass Factors for Zenith Sky Spectrometers: Intercomparison of Calculations with Different Radiative Transfer Models', *Geophys. Res. Lett.*, **22**, 1113–1116 (1995).
19. J. Slusser, K. Hammond, A. Kylling, K. Stamnes, L. Perliski, A. Dahlback, D. Anderson, R. DeMajistre, 'Comparison of Air Mass Computations', *J. Geophys. Res.*, **101**, 9315–9321 (1996).
20. A.W. Brewer, C.T. McElroy, J.B. Kerr, 'Nitrogen Dioxide Concentrations in the Atmosphere', *Nature*, **246**, 129–133 (1973).
21. J.F. Noxon, 'Nitrogen Dioxide in the Stratosphere and Troposphere Measured by Ground-based Absorption Spectroscopy', *Science*, **189**, 547–549 (1975).
22. J.P. Pommereau, 'Observation of NO_2 Diurnal Variation in the Stratosphere', *Geophys. Res. Lett.*, **9**, 850–853 (1982).
23. G. Mount, R. Sanders, A. Schmeltekopf, S. Solomon, 'Visible Spectroscopy at McMurdo Station, Antarctica: 1. Overview and Daily Variation of NO_2 and O_3, Austral Spring, 1986', *Geophys. Res.*, **92**, 8320–8328 (1987).
24. J.P. Pommereau, F. Goutail, 'An Advanced Visible–UV Spectrometer for Atmospheric Composition Measurements', ESA SP-270, 197–200, 1987.
25. A. Wahner, J. Callies, H.-P. Dorn, U. Platt, C. Schiller, 'Near UV Atmospheric Absorption Measurements of Column Abundances During Airborne Arctic Stratospheric Expedition, Jan.–Feb. 1989: 3. BrO Observations', *Geophys. Res. Lett.*, **17**, 517–520 (1990).
26. P.V. Johnston, R.L. McKenzie, J.G. Keys, W.A. Matthews, 'Observations of Depleted Stratospheric NO_2 Following the Pinatubo Volcanic Eruption', *Geophys. Res. Lett.*, **19**, 211–213 (1992).
27. R. Sanders, S. Solomon, G. Mount, M.W. Bates, A. Schmeltekopf, 'Visible Spectroscopy at McMurdo Station, Antarctica: 3. Observation of NO_3', *J. Geophys. Res.*, **92**, 8339–8342 (1987).
28. S. Solomon, H.L. Miller, J.P. Smith, R.W. Sanders, G.H. Mount, A.L. Schmeltekopf, J.F. Noxon, 'Atmospheric NO_3, 1, Measurement Technique and the Annual Cycle at 40°N', *J. Geophys. Res.*, **94**, 11 041–11 048 (1989).
29. S. Solomon, G. Mount, R.W. Sanders, A. Schmeltekopf, 'Visible Spectroscopy at McMurdo Station, Antarctica: 2. Observation of OClO', *J. Geophys. Res.*, **92**, 8329–8338 (1987).
30. C.Schiller, A. Wahner, U. Platt, H.-P. Dorn, J. Callies, D.H. Ehhalt, 'Near UV Atmospheric Absorption Measurements of Column Abundances During Airborne Arctic Stratospheric Expedition, January–February 1989: 2. OClO observations', *Geophys. Res. Lett.*, **17**, 501–504 (1990).
31. D. Perner, T. Klüpfel, U. Parchatka, A. Roth, T. Jörgensen, 'Ground-based UV–Vis Spectroscopy: Diurnal OClO – Profiles During January 1990 above Söndre Strømfjord, Greenland', *Geophys. Res. Lett.*, **18**, 787–790 (1991).
32. K. Kreher, J.G. Keys, P.V. Johnston, U. Platt, X. Liu, 'Ground-based Measurements of OClO and HCl in Austral Spring 1993 at Arrival Heights, Antarctica', *Geophys. Res. Lett.*, **23**, 1545–1548 (1996).
33. C. Otten, F. Ferlemann, U. Platt, T. Wagner, K. Pfeilsticker, 'Groundbased DOAS UV/visible Measurements at Kiruna (Sweden) During the SESAME Winters 1993/94 and 1994/95', *J. Atmos. Chem.*, **30**, 141–162 (1998).
34. S. Solomon, R.W. Sanders, M.A. Carroll, A.L. Schmeltekopf, 'Visible and Near-ultraviolet Spectroscopy at McMurdo Station, Antarctica, 5, Observations of the Diurnal Variations of BrO and OClO', *J. Geophys. Res.*, **94**, 11 393–11 403 (1989).
35. K. Kreher, P.V. Johnston, S.W. Wood, U. Platt, 'Ground-based Measurements of Tropospheric and Stratospheric BrO at Arrival Heights (78°S), Antarctica', *Geophys. Res. Lett.*, **24**, 3021–3024 (1997).
36. A. Sarkissian, J.P. Pommereau, F. Goutail, 'Identification of Polar Stratospheric Clouds from the Ground by Visible Spectrometry', *Geophys. Res. Lett.*, **18**, 779–782 (1991).
37. F. Erle, K. Pfeilsticker, U. Platt, 'On the Influence of Tropospheric Clouds on Zenith-scattered-light Measurements of Stratospheric Species', *Geophys. Res. Lett.*, **22**, 2725–2728 (1995).
38. U. Platt, L. Marquard, T. Wagner, D. Perner, 'Corrections for Zenith Scattered Light DOAS', *Geophys. Res. Lett.*, **24**, 1759–1762 (1997).
39. F. Bednarz (ed.), 'GOME Users Manual SP-1182, European Space Agency', Noordwijk, The Netherlands, 1995.
40. M.E. Feigelson, *Radiation in a Cloudy Atmosphere*, D. Reidel, Dordrecht, 1981.

41. T. Wagner, F. Erle, L. Marquard, C. Otten, K. Pfeilsticker, Th. Senne, J. Stutz, U. Platt, 'Cloudy Sky Photon Pathlengths as Derived from DOAS Observations, *J. Geophys. Res.*, **103**, 25 307–25 321 (1998).
42. K. Pfeilsticker, F. Erle, O. Funk, H. Veitel, U. Platt, 'First Geometrical Pathlengths Probability Density Function Derivation of the Skylight from Spectroscopically Highly Resolved Oxygen A-band Observations. 1. Measurement Technique, Atmospheric Observations and Model Calculations', *J. Geophys. Res.*, **103**, 11 483–11 504 (1998).
43. K. Pfeilsticker, F. Erle, U. Platt, 'Absorption of Solar Radiation by Atmospheric O_4', *J. Atmos. Sci.*, **54**, 933–939 (1997).
44. T. Winterrath, T. Kurosu, A. Richter, J.P. Burrows, 'Enhanced O_3 and NO_2 in Thunderstorm Clouds: Convection or Production', *Geophys. Res. Lett.*, **26**, 1291–1294 (1999).
45. D. Perner, D.H. Ehhalt, H.W. Pätz, U. Platt, E.P. Röth, A. Volz, 'OH-radicals in the Lower Troposphere', *Geophys. Res. Lett.*, **3**, 466–468 (1976).
46. U. Bonafe, G. Cesari, G. Giovanelli, T. Tirabassi, O. Vittori, 'Mask Correlation Spectrophotometry Advanced Methodology for Atmospheric Measurements', *Atmos. Environ.*, **10**, 469–474 (1976).
47. B.I. Kuznetzov, K.S. Nigmatullina, 'Optical Determination of the Nitrogen Dioxide Content in the Atmosphere', *Izv. Atmos. Oceanic Phys.*, **13**, 614–617 (1977).
48. M. Millan, 'Recent Advances in Correlation Spectroscopy for the Remote Sensing of SO_2', Proceedings of 4th Joint Conference on Sensing of Environmental Pollutants, 40–43, 1978.
49. M.M. Millan, 'Remote Sensing of Air Pollutants. A Study of Some Atmospheric Scattering Effects', *Atmos. Environ.*, **14**, 1241–1253 (1980).
50. J.F. Noxon, R.B. Norton, W.R. Henderson, 'Observation of Atmospheric NO_3', *Geophys. Res. Lett.*, **5**, 675–678 (1978).
51. J.F. Noxon, R.B. Norton, E. Marovich, 'NO_3 in the Troposphere', *Geophys. Res. Lett.*, **7**, 125–128 (1980).
52. U. Platt, 'Mikrometeorologische Bestimmung der SO_2-Abscheidung am Boden', PhD Thesis, University of Heidelberg, 1977.
53. U. Platt, D. Perner, 'Ein Instrument zur Spektroskopischen Spurenstoffmessung in der Atmosphäre', *Fresenius' Z. Anal. Chem.*, **317**, 309–313 (1984).
54. P.V. Johnston, R.L. McKenzie, 'Long-path Absorption Measurements of NO_2 in Rural New Zealand', *Geophys. Res. Lett.*, **11**, 69–72 (1984).
55. H. Edner, A. Sunesson, S. Svanberg, L. Uneus, S. Wallin, 'Differential Optical Absorption Spectroscopy System used for Atmospheric Mercury Monitoring', *Appl. Opt.*, **25**, 403–409 (1986).
56. H. Edner, R. Amer, P. Ragnarson, M. Rudin, S. Svanberg, 'Atmospheric NH_3 Monitoring by Long-path UV Absorption Spectroscopy', Proceedings of International Congress on Optical Science and Engineering: Environment and Pollution Measurement Sensors and Systems, Den Haag, 1990.
57. H. Edner, P. Ragnarson, S. Svanberg, 'A Multi-path DOAS System for Large Area Pollution Monitoring', *Proceedings of EUROTRAC Symposium '92*, eds. P.M. Borrell et al., SPB Academic Publishing BV, Den Haag, 220–223, 1992.
58. H.-P. Dorn, U. Platt, 'Eine Empfindliche Optische Nachweismethode für Spurenstoffe in der Atmosphäre', *Elektrizitätswirtschaft*, **24**, 967–970 (1986).
59. H.E. Wagner, 'Bestimmung der Spurengaszusammensetzung in der Troposphäre mit Langpfad-DOAS (Differenzielle Optische Absorptionsspektroskopie)', Diplomarbeit, Inst. f. Physikalische und Theoretische Chemie, TU München, 1990.
60. H.W. Biermann, E.C. Tuazon, A.M. Winer, T.J. Wallington, J.N. Pitts, 'Simultaneous Absolute Measurements of Gaseous Nitrogen Species in Urban Ambient Air by Long Pathlength Infrared and Ultraviolet–Visible Spectroscopy', *Atmos. Environ.*, **22**, 1545–1554 (1988).
61. R.K. Stevens, T.L. Vossler, 'DOAS (Differential Optical Absorption Spectroscopy) Urban Pollution. Measurements of Atmospheric Gases', *Proc. SPIE-Int. Soc. Opt. Eng.*, **1433**, 25–35 (1991).
62. C. Fayt, P. Dufour, C. Hermans, M. van Roozendael, P.C. Simon, 'Instrument and Software Development for DOAS Measurements of Atmospheric Constituents', Proceedings of EUROTRAC Symposium '92, ed. P.M. Borrell et al., SPB Academic Publishing BV, Den Haag, 231–233, 1992.
63. Canrad-Hanovia, Compact Arc Lamps, Data Sheet, Canrad-Hanovia Inc., Newark NJ, USA, 1986.
64. J. Stutz, U. Platt, 'Improving Long-path Differential Optical Absorption Spectroscopy (DOAS) with a Quartz-fiber Mode-mixer', *Appl. Optics*, **36**, 1105–1115 (1997).
65. J.M. Lerner, A. Thevenon, 'The Optics of Spectroscopy', Jobin-Yvon Optical Systems/Instruments SA, 1988.
66. J. Bösenberg, D. Brassington, P.C. Simon (eds.), *EUROTRAC Final Report, Springer Verlag*, New York, Vol. 8, Chapter 11, 1997.
67. D.A. Yates, T. Kuwana, 'Evaluation of a Self-contained Linear Diode Array Detector for Rapid Scanning Spectrophotometry', *Anal. Chem.*, **48**, 510–514 (1976).
68. J. Stutz, U. Platt, 'Problems in Using Diode Arrays for Open Path DOAS Measurements of Atmospheric Species', Proc. EUROPTO Series. *Optical Methods in Atmospheric Chemistry*, eds. H.I. Schiff, U. Platt, **1715**, 329–340 (1992).
69. G.H. Mount, R.W. Sanders, J.W. Brault, 'Interference Effects in Reticon Photodiode Array Detectors', *Appl. Opt.*, **31**, 851 (1992).
70. J. Stutz, U. Platt, 'Numerical Analysis and Error Estimation of Differential Optical Absorption Spectroscopy Measurements with Least-squares Methods', *Appl. Optics*, **35**, 6041–6053 (1996).

71. U. Platt, 'Differential Optical Absorption Spectroscopy (DOAS)', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, John Wiley & Sons, 1994.
72. A. Pierson, J. Goldstein, 'Stray Light in Spectrometers: Causes and Cures', *Lasers Optron.*, September, 67–74 (1989).
73. J.K. Pribram, C.M. Penchina, 'Stray Light in Czerny–Turner and Ebert Spectrometers', *Appl. Optics*, **7**, 2005–2014 (1968).
74. H.P. Dorn, U. Brandenburger, T. Brauers, M. Hausmann, D.H. Ehhalt, 'In-situ Detection of Tropospheric OH Radicals by Folded Long-path Laser Absorption. Results from the POPCORN Field Campaign in August 1994', *Geophys. Res. Lett.*, **23**, 2537–2540 (1996).
75. T. Gomer, J. Stutz, F. Heintz, U. Platt, 'MFC Handbook', 1995.
76. C. Leue, M. Wenig, U. Platt, *Retrieval of Atmospheric Trace Gas Concentrations*, Handbook of Image Processing, Vol. III, Academic Press, San Diego, 783–805, 1999.
77. P. Knoll, R. Singer, W. Kiefer, 'Improving Spectroscopic Techniques by a Scanning Multichannel Method', *Appl. Spectr.*, **44**, 776–782 (1990).
78. T. Brauers, M. Hausmann, U. Brandenburger, H.-P. Dorn, 'Improvement of Differential Optical Absorption Spectroscopy Using Multichannel Scanning Technique', *Appl. Opt.*, **34**(21), 4472–4479 (1995).
79. M. Hausmann, U. Brandenburger, T. Brauers, H.-P. Dorn, 'Simple Monte Carlo Methods to Estimate the Spectra Evaluation Error in Differential-optical-absorption Spectroscopy', *Appl. Optics*, **38**(3), 462–475 (1998).
80. J.F. Grainger, J. Ring, 'Anomalous Fraunhofer Line Profiles', *Nature*, **193**, 762 (1962).
81. D.J. Fish, R.L. Jones, 'Rotational Raman Scattering and the Ring Effect in Zenith-sky Spectra', *Geophys. Res. Lett.*, **22**, 811–814 (1995).
82. J. Burrows, M. Vountas, H. Haug, K. Chance, L. Marquard, K. Muirhead, U. Platt, A. Richter, V. Rozanov, 'Study of the Ring Effect', Final Report for ESA Contract 109996/94/NL/CN, 1995.
83. H. Haug, Diploma Thesis, University of Heidelberg, 1996.
84. H.K. Roscoe, D.J. Fish, R.L. Jones, 'Interpolation Errors in UV–Visible Spectroscopy for Stratospheric Sensing: Implications for Sensitivity, Spectral Resolution, and Spectral Range', *Appl. Opt.*, **35**, 427–432 (1996).
85. G. Hübler, D. Perner, U. Platt, A. Tönnissen, D.H. Ehhalt, 'Groundlevel OH Radical Concentration: New Measurements by Optical Absorption', *J. Geophys. Res.*, **89**, 1309–1319 (1984).
86. R. Neuroth, H.P. Dorn, U. Platt, 'High Resolution Spectral Features of a Series of Aromatic Hydrocarbons and BrO: Potential Interferences with OH Measurements', *J. Atmos. Chem.*, **12**, 12 287–12 298 (1991).
87. T. Etzkorn, B. Klotz, S. Sörensen, I.V. Patroescu, I. Barnes, K.H. Becker, U. Platt, 'Gas-phase Absorption Cross-sections of 24 Monocyclic Aromatic Hydrocarbons in the UV and IR Spectral Ranges', *Atmos. Environ.*, **33**, 525–540 (1999).
88. M. Fiedler, H. Frank, T. Gomer, M. Hausmann, K. Pfeilsticker, U. Platt, 'Groundbased Spectroscopic Measurements of Stratospheric NO_2 and OClO in Arctic Winter 1989/90', *Geophys. Res. Lett.*, **20**, 963–966 (1993).
89. L. Marquard, U. Platt, 'AMFTRAN: A New Monte Carlo Radiative Transfer Model for Calculating Air Mass Factors', *Proceedings of the NATO Advanced Research Workshop*, Athens, Nov. 1995, Springer, Heidelberg, London, New York, 1996.
90. K. Pfeilsticker, U. Platt, 'Airborne Measurements During the Arctic Stratospheric Experiment: Observation of O_3 and NO_2', *Geophys. Res. Lett.*, **21**, 1375–1378 (1994).
91. U. Platt, M. Rateike, W. Junkermann, J. Rudolph, D.H. Ehhalt, 'New Tropospheric OH Measurements', *J. Geophys. Res.*, **93**, 5159–5166 (1988).
92. H.-P. Dorn, J. Callies, U. Platt, D.H. Ehhalt, 'Measurement of Tropospheric OH Concentrations by Laser Long-path Absorption Spectroscopy', *Tellus*, **40B**, 437–445 (1988).
93. U. Platt, D. Perner, G.W. Harris, A.M. Winer, J.N. Pitts, 'Detection of NO_3 in the Polluted Troposphere by Differential Optical Absorption', *Geophys. Res. Lett.*, **7**, 89–92 (1980).
94. A. Richter, M. Eisinger, A. Ladstätter Weißenmayer, F. Wittrock, J.P. Burrows, 'DOAS Zenith-sky Observations: Seasonal Variations of BrO Over Bremen (53°N) 1994–1995', *J. Atmos. Chem.*, **32**, 83–99.
95. M. Tuckermann, R. Ackermann, C. Gölz, H. Lorenzen-Schmidt, T. Senne, J. Stutz, B. Trost, W. Unold, U. Platt, 'DOAS-Observation of Halogen Radical-catalyzed Arctic Boundary Layer Ozone Destruction During the ARCTOC-campaigns 1995 and 1996 in Ny-Alesund, Spitsbergen', *Tellus*, **49B**, 533–555 (1997).
96. B. Alicke, K. Hebestreit, J. Stutz, U. Platt, 'Detection of Iodine by DOAS in the Marine Boundary Layer', *Nature*, **397**, 572–573 (1998).
97. A. Richter, F. Wittrock, M. Eisinger, J.P. Burrows, 'GOME Observation of Tropospheric BrO in Northern Hemispheric Spring and Summer 1997', *Geophys. Res. Lett.*, **25**, 2683–2686 (1998).
98. T. Wagner, U. Platt, 'Observation of Tropospheric BrO from the GOME Satellite', *Nature*, **395**, 486–490 (1998).
99. K. Kreher, M. Fiedler, T. Gomer, J. Stutz, U. Platt, 'The Latitudinal Distribution (50 °N – 50 °S) of NO_2 and O_3 in October/November 1990', *Geophys. Res. Lett.*, **22**, 1217–1220 (1995).
100. R. Volkamer, T. Etzkorn, A. Geyer, U. Platt, 'Correction of the Oxygen Interference with UV Spectroscopic (DOAS) Measurements of Monocyclic Aromatic Hydrocarbons in the Atmosphere', *Atmos. Environ.*, **32**, 3731–3747 (1998).

Diode Laser Spectroscopic Monitoring of Trace Gases

Frank K. Tittel
Rice University, Houston, USA

Konstantin P. Petrov
Gemfire Corporation, Palo Alto, USA

1 Introduction **1**
2 Fundamentals of Gas-phase Spectroscopy **2**
2.1 Atmospheric Trace Gases 2
2.2 Absorption Line Shape 3
2.3 Spectra of Multicomponent Gas Mixtures 4
3 Laser Spectroscopic Sources **5**
3.1 Advantages of Diode Lasers 5
3.2 Overtone Band Detection with Near-infrared Diode Lasers 5
3.3 Fundamental Band Detection with Lead-salt Diode Lasers 7
3.4 Diode Laser Frequency Conversion 8
4 Techniques for the Measurement of Small Gas Concentrations **9**
4.1 Long Path Length and Cavity-enhanced Spectroscopy 9
4.2 Balanced and Balanced-ratiometric Detection 11
4.3 Frequency and Wavelength Modulation Spectroscopy 12
4.4 Opto-acoustic Detection 13
5 Examples of Trace-gas Sensors **14**
5.1 Gas Sensors Based on Near-infrared Diode Laser Overtone Spectroscopy 14
5.2 Tuneable Mid-infrared Diode Laser Absorption Spectrometer 15
5.3 Gas Sensors Based on Difference Frequency Generation Absorption Spectroscopy 16
6 Conclusion **18**
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **18**
References **18**

Modern diode laser spectroscopy is becoming an important and a more widely used tool for detection and measurement of trace gases. The change is driven by the recent advances in diode laser technology and diode laser frequency conversion techniques, which now push the limits of emission wavelength, output power, operating temperature, miniaturization, and cost. Basic features of the diode laser now find new, interesting, and unique use in well-established laboratory and field applications related to trace-gas detection, and in some instances even transform the traditional spectroscopic methods. It is the purpose of this chapter to familiarize the reader with the key elements of diode laser technology and analytical methods used in diode laser spectroscopy.

1 INTRODUCTION

Sensors based on tuneable diode lasers have found widespread applications to the sensitive and selective detection of environmentally important atmospheric trace gases in real time. The motivation for such precise species concentration measurements of gaseous compounds in ambient air includes such diverse fields as urban (e.g. automobile traffic, air quality in large enclosed structures), industrial (e.g. fence line perimeter monitoring in the petrochemical industry, combustion sites and waste incinerators), rural (e.g. horticultural greenhouses, rice agro-ecosystems) emission studies, chemical analysis and control for manufacturing processes (e.g. HF, NH_3 and HCl in semiconductor fabrication facilities and HF in aluminum smelters), biomedical sensing of physiologically important molecules (e.g. NO and CO), atmospheric chemistry (e.g. CO, CO_2, CH_4 and H_2CO in global studies of the environment), and spacecraft habitat air-quality monitoring.[1,2]

Numerous analytical instruments based on optical and nonoptical techniques have been developed in the past. In general, these instruments have some, but not all, of the desired characteristics that include: high sensitivity, selectivity, multicomponent detection capability, room temperature operation, large dynamic range, automatic operation, small size, high reliability, ease of use, and cost-effectiveness in terms of initial and maintenance costs. An ideal gas sensor technology that meets all these diverse requirements is, however, a challenging research and development task. Today's state-of-the-art laser spectroscopy, for example, is far from being able to meet these requirements, but recent significant technological advances may change the way it is used in future applications. Areas of technology that have seen a particularly strong improvement are tuneable diode lasers, both near-infrared (NIR) and mid-infrared (MIR); novel nonlinear materials for optical frequency conversion; optical fiber and semiconductor amplifiers;

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

low-noise, room-temperature detectors; and advanced data acquisition and signal processing techniques.

Such new diode laser based spectroscopic sensors are beginning to provide excellent sensitivity and selectivity of an increasing number of organic and inorganic gaseous compounds in the infrared spectral region. This region consists of both the (NIR, 0.8 μm to 2 μm) that can be accessed by an increasing number of distributed feedback (DFB) telecommunications diode lasers and the (MIR, 2 μm to 20 μm) by lead-salt and antimony-based diode lasers, quantum cascade lasers, and compact sources based on difference-frequency mixing of commercially available diode and fiber lasers. Performance characteristics of several recent optical architectures of MIR gas sensors that have been developed and applied to real-world spectroscopic applications in the field and laboratory will be discussed. An effective method to increase sensitivity is to increase the effective optical path. This is possible for in situ, open path atmospheric measurements, but numerous applications require a compact extractive technique based on multipass cells (White, Herriott designs).

The NIR region is characterized by the presence of the first and second overtone and combination bands for many important gases. However, the transition strengths of the fundamental vibrational bands in the MIR are at least one to two orders of magnitude stronger, making this spectral regime the region of choice for high-sensitivity laser absorption spectroscopy. The detection sensitivity of current diode laser based gas sensors using direct absorption spectroscopy is limited by etalon fringes and laser feedback noise to 10^{-5} to 10^{-6} relative absorption. In addition to direct absorption spectroscopy, various techniques such as frequency modulation (FM), two-tone frequency modulation (TTFM) spectroscopy, balanced homodyne or balanced ratiometric detection (BRD) and cavity ring down spectroscopy (CRDS) can be utilized. Of these, only CRDS offers a means for avoiding the baseline variations caused by etalon fringes which often limit sensitivity.

2 FUNDAMENTALS OF GAS-PHASE SPECTROSCOPY

2.1 Atmospheric Trace Gases

Spectroscopic trace-gas detection is a method allowing one to compute concentration of a known gas, or gases, from a measured optical absorption spectrum of the gas mixture (in practice, a small fragment of the spectrum may be measured). The procedure requires a good quantitative knowledge of the gas absorption characteristics. This knowledge is the realm of molecular spectroscopy, a complex and highly developed subject that is outside the scope of this chapter. A few fundamental spectroscopic concepts and formulae that are directly applicable to gas detection are, however, summarized in this section.

Each atom or molecule, small or large, is uniquely characterized by a set of energy levels. Transitions between levels by absorption or emission of electromagnetic radiation result in highly specific spectroscopic features. These features allow both the identification and quantification of the molecular species, such as atmospheric trace gases. Molecules may undergo transition between electronic, vibrational, and rotational states when exposed to electromagnetic radiation, resulting in absorption spectra. These spectra consist of a number of discrete absorption lines. Each line will have a certain linewidth and shape that depend on temperature and what surrounds the molecule, so the lines may form resolved or unresolved bands (see Figure 1); some will be intense while others will be weak. Transitions between molecular rotational–vibrational (ro-vibrational) states occur in the infrared ''fingerprint'' region of the electromagnetic spectrum, approximately between the wavelengths of 2.5 μm and 25 μm. Often also overtone and combination-overtone ro-vibrational bands are possible with significantly lower line intensity as compared to those for fundamental vibrational bands and the corresponding wavelengths are in the 0.8 μm to 2.5 μm spectral region. Transitions between electronic states of atoms and molecules occur in the ultraviolet and visible spectral region, 200 nm to 500 nm.

All polyatomic molecules, with the exception of homonuclear diatomic molecules (e.g. N_2), absorb infrared radiation. The absorption changes the state of molecular rotation and vibration. An absorption spectrum therefore depends on the physical properties of the molecule such as size and shape and hence each molecule

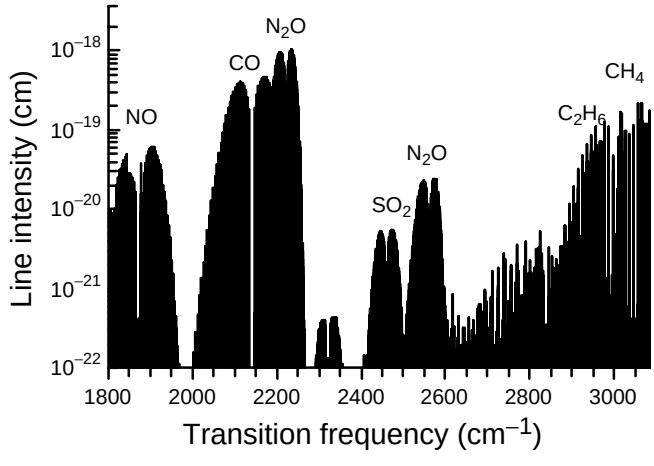

Figure 1 Survey absorption spectra of several atmospheric trace gases in the 3.1–5.6 μm wavelength region (3200–1800 cm^{-1}[3,4]).

Table 1 Analysis of atmospheric gas traces

Molecule	Absorption line (μm)	Overtone-MDC (ppm m)	Fundamental-MDC (ppm m)
CO_2	4.23	8[6]	0.0072
OCS	4.85	NA	0.023
N_2O	4.54	220[6]	0.036
	2.87		1.3
CO	4.6	100[7]	0.044
HCl	3.5	1.3[8]	0.050
HCN	3.0	0.29[9]	0.081
C_2H_6	3.34	NA	0.081
HF	2.4	0.2[10]	0.090
C_2H_2	3.08	0.08[9]	0.1
CH_4	3.3	15[8]	0.23
HBr	3.81	0.6[9]	0.32
H_2CO	3.6	50[9]	0.54
NO_2	3.46	0.5[11]	1.8
NH_3	2.94	2[12]	2.7
NO	5.3	60[9]	4
O_3	4.75	NA	9
SO_2	4.0	NA	40
H_2S	3.72	0.2[10]	54
H_2O	3.3	NA	NA

Values computed for the following parameters:
– Fundamental: $p = 100$ torr; $T = 298$ K; Min. detectable absorbance $= 2 \times 10^{-4}$ using direct absorption.
– Overtone: Min. detectable absorbance $\cong 1 \times 10^{-5}$ using either balanced detection/FM-techniques.
MDC, minimal detectable concentration.

is characterized by a unique spectral "signature". Spectra of linear and some nonlinear polyatomic molecules consist of an array of individual or small groups of lines. In the case of large polyatomic molecules (e.g. benzene, C_6H_6) at atmospheric pressure, there are many lines overlapping each other, resulting in broad spectral features with some occasional peaks. These features occur in the majority of spectra and are the principal reason for the sensitivity and selectivity of trace-gas detection in the infrared.

There are numerous atmospheric trace gases (see Table 1). Their concentrations would normally be in the parts per billion (ppb, 10^{-9}) to parts per million (ppm, 10^{-6}) range. Minimum and maximum concentrations range from parts per trillion (ppt, 10^{-12}) to percent (%, 10^{-2}). There is much spectroscopic data available in the literature[1,2] and in electronic form which are important tools in the identification and development of specific detection strategies, especially in the presence of interfering species.[3–5]

2.2 Absorption Line Shape

In the absence of optical saturation and particulate-related scattering, the intensity of light $I(x)$ propagating in a homogeneous gas follows the Beer–Lambert law (Equation 1):

$$I(x) = I_0 \exp(-\sigma_m(\nu) N_m x) \tag{1}$$

Here m is the gas species index that labels the molecular concentration N_m and absorption cross-section $\sigma_m(\nu)$. The molecular absorption cross-section depends on frequency and has the units of cm^2 per molecule. It is the sum of cross-sections of all individual ro-vibrational transitions (Equation 2):

$$\sigma_m(\nu) = \sum_n S_n \Gamma(\nu - \nu_n) \tag{2}$$

We label the transitions using the index n; thus ν_n is the frequency of the n-th transition and S_n is its intensity. Technically both these quantities should also bear the index m to remind that they refer to a certain gas species, but for simplicity it is not done here.

The function $\Gamma(\nu)$ describes the line shape and has the same analytical form for all transitions. In MIR spectroscopy, the broadening of an individual transition due to finite upper-level lifetime is insignificant compared to broadening by the other two important mechanisms – thermal motion and molecular collisions. Their individual and combined effects on a molecular transition at a frequency ν_n are expressed as follows (Equations 3–7):

Thermal motion (Gaussian):

$$\Gamma(\nu) = \frac{1}{\Delta\nu_T \sqrt{\pi}} \exp\left[-\left(\frac{\nu}{\Delta\nu_T}\right)^2\right] \tag{3}$$

$$\Delta\nu_T = \frac{\nu_n}{c}\sqrt{\frac{2kT}{M}} \tag{4}$$

Molecular collisions (Lorentzian):

$$\Gamma(\nu) = \frac{1}{\pi\gamma_n P}\left[1 + \left(\frac{\nu}{\gamma_n P}\right)^2\right]^{-1} \tag{5}$$

Combined broadening (Voigt):

$$\Gamma(\nu) = \frac{1}{\gamma_n P} V\left(\frac{\nu}{\gamma_n P}; \frac{\Delta\nu_T}{\gamma_n P}\right) \tag{6}$$

$$V(x, y) = \frac{1}{\pi\sqrt{\pi}} \int_{-\infty}^{+\infty} \frac{\exp(-t^2)}{1 + (x + yt)^2}\, dt \tag{7}$$

Here c is the speed of light in vacuum, k the Boltzmann's constant, T the gas temperature, M the mass of the molecule, P the gas pressure, and γ_n the coefficient of pressure broadening. The quantities $\Delta\nu_T$ and $2\gamma_n P$ are referred to as the Doppler- and pressure-broadened linewidths. The line shape that results from the combined effect of Doppler- and pressure-broadening is a convolution of the two respective line shapes. It is known

as the Voigt profile, and its mathematical expression $V(x, y)$ can not be further simplified. The physical significance of the convolution is that the Voigt profile has different asymptotic shapes for very low and very high gas pressure. At low pressure, molecular collisions are less frequent, leaving thermal motion the dominant broadening mechanism – the corresponding line shape is near-Gaussian. As the gas pressure increases the collisions take over, and the resulting line shape is near-Lorentzian (see Figure 2).

The previous expressions do not include the effect of pressure shift, which is typically in the range of several megahertz per atmosphere. The shift is very small compared to the width of an atmospheric-pressure-broadened line, typically several gigahertz.

It is common practice in infrared spectroscopy to express transition frequencies in inverse centimeters (cm^{-1}), or wavenumbers, defined simply as the inverse of the transition wavelength in vacuum, $\nu = \lambda^{-1}$. Multiplying this quantity by c gives the frequency in Hertz; thus 1 cm^{-1} is roughly 30 GHz. We shall use both units throughout this chapter, where appropriate.

One can verify, by integrating the absorption cross-section of an individual transition over frequency, that it is independent of the broadening mechanism and is equal to the line intensity S_n, in the units of cm per molecule. Line intensity is proportional to the lower-state population density of a transition and thus depends on temperature (see Figure 3).

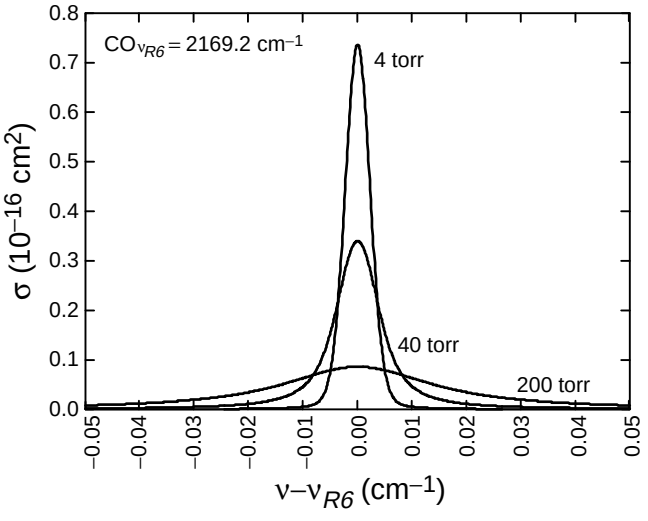

Figure 2 Frequency-dependent absorption cross-section of the R(6) fundamental transition of carbon monoxide at room temperature, plotted for different values of background air pressure. The horizontal axis is in cm^{-1} relative to the line center, $\nu_{R6} = 2169.2\,cm^{-1}$. Area under each curve, or line intensity, is the same: $S = 4.44 \times 10^{-19}$ cm. At 40 torr background air pressure, the Doppler- and collision-broadened linewidths are approximately equal: 0.003 cm^{-1}.

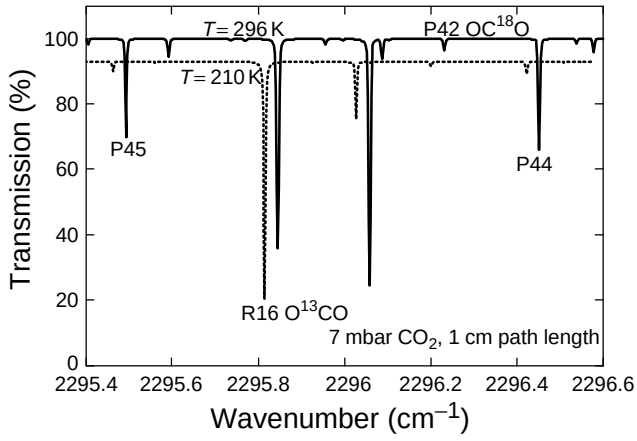

Figure 3 Absorption spectra of pure CO_2 at a pressure of 7 mbar over a 1 cm path calculated at two different temperatures using the HITRAN database.[4] The dotted trace has been offset for clarity and corresponds to $T = 210$ K, simulating the surface atmosphere of Mars. The strong dependence of line intensity on temperature for the transitions P44 and P45 should be noted.

To a good approximation, especially at low temperatures, the sum of line intensities of all transitions in a band is independent of temperature and is known as the band strength B_m (Equation 8):

$$B_m = \int_{-\infty}^{+\infty} \sigma_m(\nu)\,\mathrm{d}\nu = \sum_n S_n \tag{8}$$

It contains information on the vibrational electric dipole moment.[3] Thus in order to compute the absorption cross-section $\sigma_m(\nu)$ at any frequency, one must know the values ν_n, S_n, and γ_n for all transitions. These parameters have been measured and calculated for many lightweight gas molecules across the microwave and infrared spectrum, and compiled into extensive databases such as HITRAN[4] and GEISA.[5] Numerically accurate absorption spectra can be computed based on these data, not only for single gas species but for gas mixtures as well.

2.3 Spectra of Multicomponent Gas Mixtures

Analytical formulae given in the preceding section apply also to gas mixtures. The total absorption cross-section $\sigma(\nu)$ is then a weighted average of absorption cross-sections of individual species, with the mole fraction C_m of each species used as the weight coefficient (Equations 9 and 10):

$$\sigma(\nu) = \sum_m C_m \sigma_m(\nu) \tag{9}$$

$$\sum_m C_m = 1 \tag{10}$$

For each of the m species, the pressure broadening coefficients γ_n generally depend on the transition. They also

depend on the type of molecule with which the collisions occur. In general, partial pressures in conjunction with the appropriate pressure-broadening coefficients should be used to compute the overall pressure broadening from all gases present in the background (this includes self-broadening). Air-broadening coefficients are particularly useful in calculations, and are listed in spectroscopic databases.[4,5]

In trace-gas sensing applications, however, the species of interest are often present in very low concentrations, so that self-broadening and broadening against other trace gases can be neglected in calculations, and air-broadening alone will suffice. For the conditions of atmospheric pressure broadening, $\gamma P \gg \Delta\nu_T$, the Doppler contribution to the overall linewidth can often be neglected, and the line shape be treated as pure Lorentzian. Likewise, at pressures low enough to ensure $\gamma P \ll \Delta\nu_T$, the line shape can be treated as pure Gaussian. In either case, calculation of the line profile is simplified considerably.

At intermediate total pressures, $\gamma P \sim \Delta\nu_T$, which for most lightweight gases range from 5 to 100 torr, calculation of the Voigt profile is necessary to obtain numerically accurate absorption spectra. Methods for approximate calculation of the Voigt profile, and the related plasma dispersion function, are now a well-developed subject. We have found the approximations published by Humlicek[13] to be particularly useful.

3 LASER SPECTROSCOPIC SOURCES

3.1 Advantages of Diode Lasers

Soon after the first semiconductor (diode) laser was developed in 1962, diode laser spectroscopy became a firmly entrenched technique for detection, identification, and measurement of molecular and atomic species in the gas phase. The reason for its wide acceptance in the scientific community was twofold. First, narrow-linewidth light sources such as lasers were well suited for probing the inherently narrow molecular and atomic transitions, so that both the line shape and intensity could be measured accurately and rapidly. Lasers simply offered better wavelength resolution (linewidth) and spectral brightness (power emitted per unit linewidth) than conventional grating or prism monochromators. Second, among all lasers, diode lasers offered a unique combination of tuneability, output power, small size, and modulation capabilities (see Table 2). In other words they are significantly more convenient to use than other sources. Perhaps the most important advantages of diode lasers are the simple excitation in the form of electric current and the fact that the laser wavelength and output power depend on the current. For small changes in the injection current, that dependence is nearly *linear* and *instantaneous*, allowing predictable and fast wavelength control. This advantage is discussed in section 4.3.

Table 2 Performance characteristics of room-temperature single-frequency diode lasers

Parameter	Unit	Typical values
Center wavelength	nm	670, 780, 810, 860, 980, 1060, 1310, 1550
Gain bandwidth	nm	20–50
Output power	mW	3–200
Mode size	μm	1×3
Linewidth	MHz	10–300
Tuning response	nm/K	0.1–0.3
Tuning response	GHz/mA	0.3–2
Threshold current	mA	20–90
Modulation bandwidth	GHz	2–40

Diode laser spectroscopy is divided into two categories according to the spectral region used. MIR spectroscopy is used in the "fingerprint" region from 2.5 μm to 25 μm where most molecular species exhibit fundamental absorption. It offers the highest detection sensitivity among spectroscopic methods, primarily because the fundamental absorption bands are very strong. Diode lasers operating at MIR wavelengths typically deliver 100-μW output power and require cryogenic cooling (more on this in section 3.3). NIR, or "overtone", spectroscopy employs room-temperature diode lasers at wavelengths below 2 μm. Detected here are the short-wavelength molecular overtone transitions that are typically a factor of 30 to 300 weaker than the fundamental transitions. The two types of molecular spectroscopy just described represent the two different choices made in the tradeoff between absorption strength and the optical power available to detect it. MIR spectroscopy is used whenever higher sensitivity is needed, but at the expense of cryogenic cooling. Overtone spectroscopy is the method of choice in applications where lower sensitivity can be tolerated, but where low cost, reliability, and room-temperature operation are paramount. Alternative techniques for the generation of MIR light for spectroscopic applications will be discussed in section 3.4.

3.2 Overtone Band Detection with Near-infrared Diode Lasers

NIR spectrometers usually employ commercial diode lasers with emission wavelengths from 780 nm to 1900 nm. Gas detection at these wavelengths is based on the molecular vibration overtone and combination-overtone bands that are significantly less intense than the fundamental

bands. For example, the lines of the first overtone stretch vibration band of methane centered at 1.6 μm is roughly 160 times weaker than the fundamental band. Thus ambient methane would cause a 0.005% absorption over a 1-m path length at this wavelength at room temperature.[4] Reliable measurement of such low absorption is difficult even under laboratory conditions, so longer path lengths or special measurement techniques are necessary to render a satisfactory signal-to-noise ratio for applications requiring trace-level monitoring.

Overtone spectrometers usually have plenty of optical probe power to deal with this problem. NIR diode lasers emit anywhere from 1 to 100 mW of single-frequency radiation with low excess noise: 15 to 35 dB over shot noise limit is typical, and it can be detected with a silicon photo-diode.[14] They are fast, with modulation bandwidths of over a gigahertz allowing rapid scanning, fast FM, and leading to gas detection in real time.

High output power levels have two important benefits. First, detector noise can be neglected and the measurement of absorption be performed near the optical shot-noise limit. For example, Allen et al.[15] report a detection sensitivity of $2 \times 10^{-7}\,\mathrm{Hz}^{-1/2}$ absorption units with the use of a 1.3-μm diode laser and a balanced ratiometric detector. Second, with high initial power available in a beam, one can employ a multi-pass cell to propagate the beam back and forth through a gas sample, achieving long effective path lengths and thus increasing the observed absorption signal. Although the throughput of such a cell decreases exponentially as the number of passes increases (also see section 4.1), there is still plenty of light left at the end to permit measurements that are not limited by detector noise.

NIR overtone spectrometers, such as one developed by Uehara and Tai,[16] are usually built to detect one or a few specific gases, for two reasons. First, the NIR wavelength region from 780 to 1900 nm is not covered completely; diode lasers are available at only few discrete wavelengths, and each has a limited tuning range when operated without an external cavity. Second, these devices are relatively inexpensive – sometimes under $50 in small quantities – making it economical to have several dedicated lasers, each detecting one gas species, in a single instrument. This configuration is also attractive because different gas species can be measured in parallel. For example, several groups have developed a gas sensor that can do this using one common beam path through the sample and one detector.[17]

Alternatively, a single external-cavity diode laser (ECDL) with a large tuning range can be used. In an ECDL, one or both faces of the laser chip are anti-reflection-coated to eliminate optical feedback. The feedback is provided by a larger, external cavity. The cavity acts as a wavelength selector, picking a specific wavelength out of the usually broad gain spectrum of the semiconductor laser material. Several cavity configurations have been developed that differ in the method of tuning, component count, output beam characteristics, and output coupling efficiency. Figure 4 shows a frequently used external cavity type, known as the Littman configuration. It employs a diffraction grating as a wavelength discriminator. The first-order diffracted beam is folded back into the cavity by a mirror that acts as a tuning element – its angle and position determine the output wavelength. The grating also functions as an output coupler, producing the zero-order (reflected) beam whose angle and point of origin are independent of wavelength. Mode-hop-free single frequency tuning ranges of over 1000 GHz have been demonstrated for an ECDL.[18] A spectrometer based on such a widely tuneable laser is a very useful tool in that it can acquire spectra of an entire molecular band in a single electronically controlled scan in a matter of seconds. Oh and Hovde,[19] for example, used a widely tuneable 1.5-μm ECDL to record a spectrum of the $\nu_1 + \nu_3$ stretch-vibration combination band of acetylene.

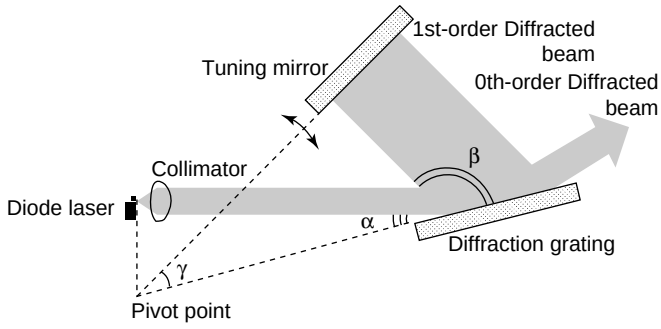

Figure 4 Optical diagram of an ECDL. Light at a wavelength λ emitted by the laser chip is collimated and is incident upon a diffraction grating with a period b. The first-order diffracted beam emerges at an angle to the grating, such that $\cos\alpha - \cos\beta = \lambda/b$, and is folded back into the cavity by an adjustable mirror. The round-trip distance traveled by light in the cavity must be an integer number of wavelengths. Continuous wavelength tuning results when both these conditions are met; this is accomplished by rotation of the tuning mirror around a pivot point located outside the laser cavity as shown.

NIR diode lasers often come in compact sealed packages that include a convenient fiber-coupled output. Thus the probe light can be delivered from a single source to several sampling locations. Likewise, the radiation passing through the sample can be returned to a detector via a fiber, sometimes even the *same* fiber (see Figure 5).

A general problem that is faced is interference by water vapor. Long path lengths do not help here as they increase not only the signal of interest but the interfering absorption as well. Multicomponent spectral

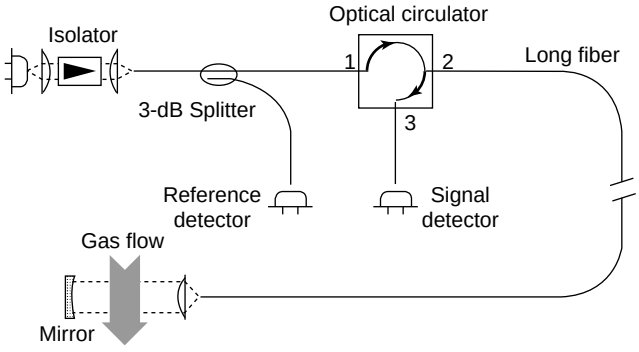

Figure 5 Schematic of a NIR diode laser spectrometer using a single fiber to deliver the light to and from the sampling location. The reference detector monitors the laser output power during frequency scans. The optical circulator transmits light from port 1 to 2 and from 2 to 3, and is used to route the returned beam to the signal detector.

fitting algorithms have been developed that can resolve weak absorption lines of interest in the presence of heavy interference by a known gas. The problem becomes far more severe, however, when the identity of the interfering species is unknown.

3.3 Fundamental Band Detection with Lead-salt Diode Lasers

Lead-salt diode lasers have been developed for operation at wavelengths from 3 to 30 μm. They typically deliver 10–500 μW of output power in a near-diffraction-limited beam and can be tuned by temperature or current control. These lasers are based on PbS, PbSe, or PbTe semiconductors and generally require cooling near the temperature of liquid nitrogen, although recently a continuous wave (CW) operation has been achieved at temperatures as high as 195 K.[20]

MIR spectrometers employing lead-salt diode lasers have shown perhaps the most impressive performance to date in terms of minimum detectable gas concentration. There are four reasons for this. First is the very high absorption strength of the fundamental molecular vibrational bands. For example, carbon dioxide at a 3.3 ppm concentration in air (which is roughly 100 times lower than its typical ambient level) will cause absorption of 15% at 4.23 μm over a 1 m path. Such macroscopic absorption signals can be easily measured even without sophisticated signal processing techniques. Second, lead-salt diode laser spectrometers use cryogenically cooled InSb or HgCdTe detectors with noise-equivalent powers in the range of 0.5 to 50 pW $Hz^{-1/2}$, and although the lead-salt lasers produce less output power than their NIR counterparts, there is still enough light to render the detector noise virtually nonexistent and allow shot-noise-limited detection. Third, near-diffraction-limited beam characteristics of lead-salt diode lasers allow beam shaping and propagation over long distances; this makes it possible to use multipass cells or remote sampling in open air. Fourth, lead-salt diode lasers have FM capabilities similar to those of NIR diode lasers, allowing the use of harmonic detection and two-tone modulation techniques – efficient methods of noise bandwidth reduction, to be discussed in section 4.3.

These advantages add up to a real-world instrument performance that is impressive.[21] Schiff et al.[22] report the detection limit 75 ppt for airborne measurements of HNO_3 in the continental boundary layer with a response time of 3 minutes. Werle and company[23] achieved a detection limit of 10 ppt for NO_2 under laboratory conditions using a signal averaging time of 25 s.

Although well entrenched in a spectroscopic laboratory, lead-salt diode lasers are seldom used in industrial applications. Preventing their wide acceptance are several important practical drawbacks. Perhaps the most serious one is the large manufacturing spread of operating wavelengths. The situation is aggravated by the fact that each individual laser has a rather limited tuning range, typically 100 cm^{-1} with temperature control, and that the range itself is not free of mode hops. This makes it difficult to find a laser chip that actually tunes to a specific wavelength – a situation potentially catastrophic in applications where there is a limited choice of absorption lines free from interference by other lines, e.g. detection of formaldehyde.[24]

A more subtle problem with lead-salt diode lasers has to do with thermal cycling, a process in which a laser is simply warmed up to room temperature and then cooled back to its normal operating temperature. The simple process tends to cause irreversible changes in prior laser tuning characteristics and emission wavelength.

Despite these and other technical difficulties, several specialized trace-gas sensors based on lead-salt diode lasers have been developed for field use. These instruments show excellent performance, proving just how effective the use of a MIR diode laser can be. For example, Webster et al.[25] reported the design and operation of a fully automatic, compact sensor for the measurement of five trace gases in the lower stratosphere on board an airplane. Their instrument houses four lasers and four detectors mounted on the same liquid-nitrogen-cooled platform, beam shaping optics, a compact multipass absorption cell with 80 m path length, analog electronics, and a computer-controlled data storage system. It is capable of detecting optical absorption as small as 10^{-5} which corresponds to detection limits in the range of several tens of ppt for species such as HCl, NO_2, HNO_3, CH_4, and N_2O. Podolske and Loewenstein[26] report construction and performance of a similar instrument that uses an additional, ''zero'', beam for removing the background

absorption signal caused by gases within the instrument case. Both instruments employ wavelength modulation (WM) and second harmonic detection to achieve high sensitivity.

3.4 Diode Laser Frequency Conversion

A more recent technique for producing spectroscopic MIR light is laser difference-frequency generation (DFG). Two laser beams at different frequencies combined in a nonlinear material with suitable dispersion characteristics can generate a beam at the difference-frequency (''idler''). Narrow emission spectra of the two lasers, usually referred to as ''pump'' and ''signal'', translate into a similarly narrow spectrum of the idler wave. Idler wavelength tuning is accomplished by tuning of the pump laser, or signal laser, or both (see Figure 6).

In the early demonstration of this method by Pine,[27] an argon-ion laser output and a dye laser output were combined in bulk lithium niobate to produce a narrowband (15 MHz), 2.2–4.2 μm tuneable radiation. However, NIR diode lasers can now be used instead,[28] making it feasible to construct compact MIR spectrometers that operate at room temperature (see Figure 7). Thus the convenience and practicality of NIR diode laser technology are combined with analytical power of MIR spectroscopy in a single instrument. Such an instrument inherits the single-frequency operation and high speed modulation capabilities of diode lasers, and takes advantage of their relatively wide tuning range. For example, a typical 780-nm diode laser (see Table 2) can be tuned over 20 nm, or 2.6% in wavelength, by temperature control without appreciable change in output power. When the output of such a laser is down-converted by mixing with a 980 nm diode laser, the tuning range in frequency units remains the same, in this case a very significant tuning range: 3.6–4.1 μm, or 13% in wavelength.

The practicality of diode-pumped MIR frequency conversion received a significant boost from the development of novel nonlinear materials, such as periodically poled lithium niobate ($LiNbO_3$), lithium tantalate ($LiTaO_3$), and potassium titanyl phosphate ($KTiOPO_4$, or KTP) at wavelengths in the 2.5–5.2 μm spectral region.[29] Quasi-phase matching properties of each of these crystals can be engineered for interaction of any pump and signal wavelengths within transparency range of the crystal, allowing significant flexibility in the choice of laser sources.[30]

Although routinely used for spectroscopy and gas detection, the DFG in bulk nonlinear crystals is characterized by low conversion efficiency, typically in the range of 0.002–0.05% $W^{-1}\,cm^{-1}$.[31] A detailed quantitative theory of this nonlinear optical process, developed by Boyd and Kleinman,[32,33] is beyond the scope of this review chapter, so we will state simply that the maximum idler power generated in a given crystal is proportional to the product of crystal length, pump power, signal power, and the squared second-order nonlinear coefficient of the crystal. Maximum DFG output power is achieved by means of optimal focusing,[34] for which any further increase in beam intensity through tighter focusing is offset by a decrease in interaction length due to diffraction, resulting in no gain in output power. For DFG radiation longer than 5 μm it is possible to use birefringent bulk

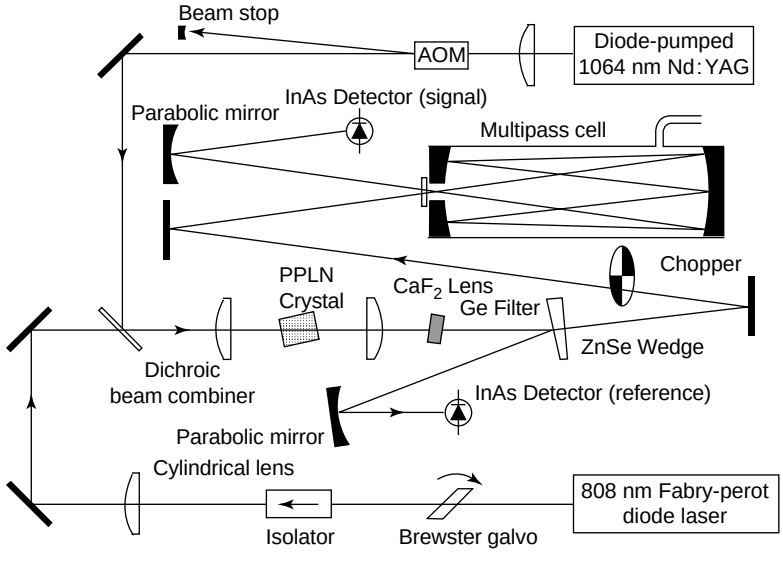

Figure 6 Schematic of difference-frequency spectrometer. A free-running Fabry–Perot diode laser at 808 nm, the ''pump'', is mixed with a diode-pumped Nd:YAG laser, the ''signal'', in a periodically poled $LiNbO_3$ (PPLN) crystal to produce a 3.4-μm tuneable radiation. An antireflection-coated germanium filter blocks the residual pump and signal light.[44]

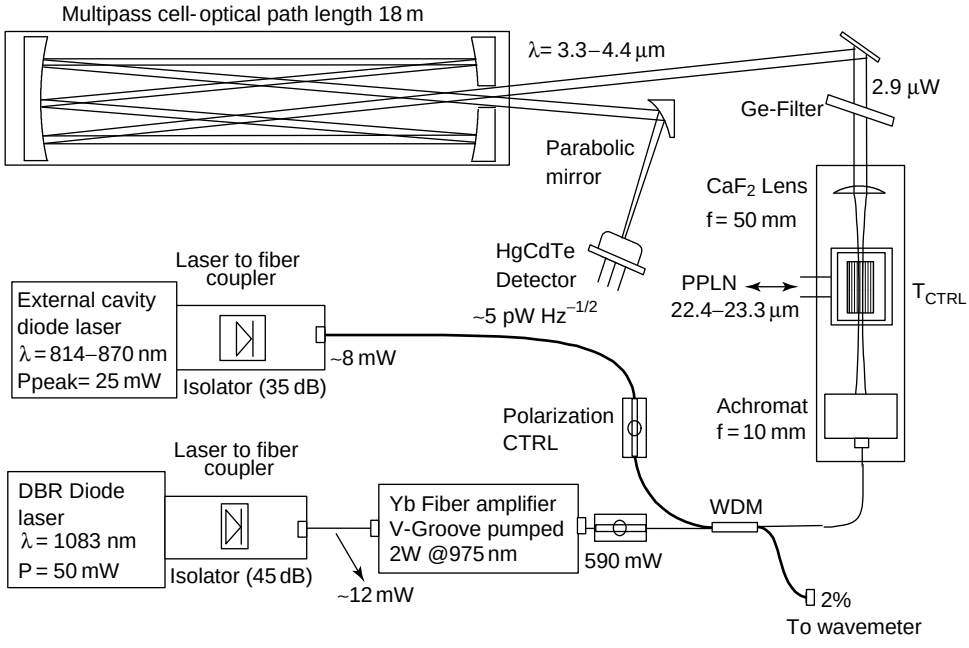

Figure 7 Portable diode-pumped difference-frequency spectrometer for detection of trace gases.[69] DBR: distributed Bragg reflector.

nonlinear optical materials, such as $AgGaS_2$, $AgGaSe_2$, GaSe and in the future, quasi-phase matched GaAs.

The tradeoff between beam size and interaction length can be eliminated altogether in guided-wave DFG. Tight optical confinement of pump and signal radiation near the waveguide core creates a region of high intensity and good modal overlap which are both maintained throughout, and independent of, the length of the waveguide.[35] Interaction length is now limited by the length of the waveguide, not diffraction. Guided-wave parametric processes, such as OPO, SHG, and DFG, have been demonstrated in periodically poled $LiNbO_3$, $LiTaO_3$, and KTP. In $LiNbO_3$, for example, a waveguide can be formed by titanium in-diffusion,[36] or a $^+Li \longleftrightarrow {}^+H$ ion exchange.[37] The latter process is normally followed by several hours of annealing at elevated temperature to create a graded index distribution.

A DFG waveguide designed to carry a single spatial mode at the idler wavelength is necessarily multimode at the shorter, pump and signal wavelengths. The presence of multiple spatial modes complicates the waveguide phase matching characteristics considerably. For example, a TM_{00} (fundamental) mode at the signal wavelength will interact with TM_{02} and TM_{10} modes at the pump wavelength, but not with TM_{01} or TM_{11} modes. Efficient and reproducible fundamental-mode excitation of a DFG waveguide was first achieved by Chou et al.,[38] who used a combination of a mode filter and an adiabatic taper (see Figure 8). An improved device featuring separate inputs for the pump and signal beams followed by a directional coupler has also been demonstrated.[39] DFG waveguides have been used to build viable sources of MIR radiation for spectroscopic purposes.[46]

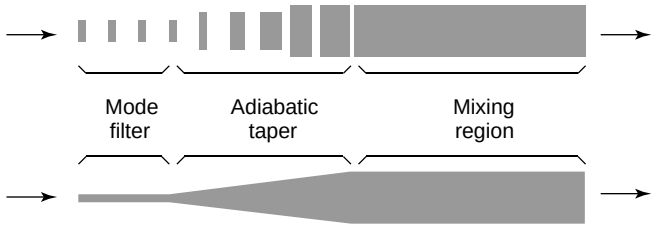

Figure 8 Alternative geometries of a tapered waveguide for difference-frequency mixing. Shaded regions indicate material with higher index of refraction. Laser light is launched into the mode filter that carries a single (fundamental) spatial mode. The mode size is then adiabatically converted to fit the mixing waveguide which is multimode. Periodic segmentation[38,39] of the waveguide acts to reduce the effective index. It also provides independent control of the effective waveguide width and depth.

4 TECHNIQUES FOR THE MEASUREMENT OF SMALL GAS CONCENTRATIONS

4.1 Long Path Length and Cavity-enhanced Spectroscopy

Gas sensors often deal with gas concentration and sample size so small that the detected full-scale absorption signal is indistinguishable from system noise. This is

no surprise, since most photons in the probe beam pass through the sample without being absorbed. If one were able to somehow ''recycle'' these unabsorbed photons, by forcing them to pass through the sample many times, one would obtain better contrast between signals on and off resonance of the molecular transition. Two different forms of this idea have been developed to date.

Long-path length spectroscopy, not to be confused with remote sensing or fence-line monitoring,[41] refers to the use of an optical set-up that provides multiple passes of the probe beam through a relatively small sample volume. An example is the Herriott cell, a device with two identical spherical mirrors facing each other and separated by nearly their radius of curvature (see Figure 9). A probe beam launched through a hole in one of the mirrors at an angle to the optical axis, completes a certain number of passes between the mirrors, and exits through the same hole. The mirror curvature, applied to the beam at each reflection, keeps it from diverging, as if in a cavity. The beam bounce pattern and path length can be controlled by adjusting the mirror separation. The picture is deceptively simple, but certain design rules must be followed to ensure that the beam exit the cell after a controlled number of passes, especially in the case of astigmatic mirrors.[42] The number of passes can routinely exceed 100, thus providing a commensurate improvement in signal strength (see Figure 10). It is important to recognize that mirrors are not perfect, and a small portion of the probe light is lost to absorption and scattering each time the beam bounces. Optical throughput of the cell thus decreases exponentially with the number of passes, whereas the detected peak-to-peak absorption only increases linearly (in the small-signal limit) with the number of passes.

Several configurations of the multipass cell are available that offer different mirror counts, input and output ports, and beam patterns. They can be sealed, for the measurement in static gas samples or controlled gas flows, or open to ambient air. Performance of long path length multipass cells, especially those with dense beam patterns, suffers from optical interference (''fringes'') due to light scattering by the cell mirrors. The fringe magnitude is

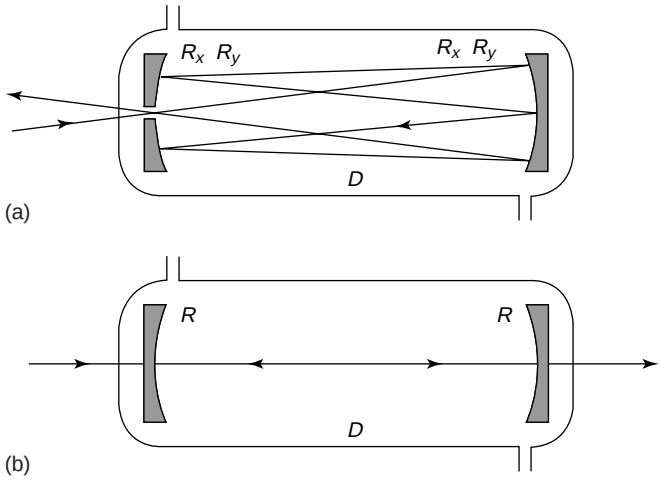

Figure 9 Schematic of a multipass cell (a). The standard Herriott cell has two spherical mirrors separated by nearly their radius of curvature $D \approx R_x = R_y$. Unlike with the confocal cavity (b), a resonance does not occur in the multipass cell because the beam never hits the same spot twice on any mirror before it exits through the coupling hole.

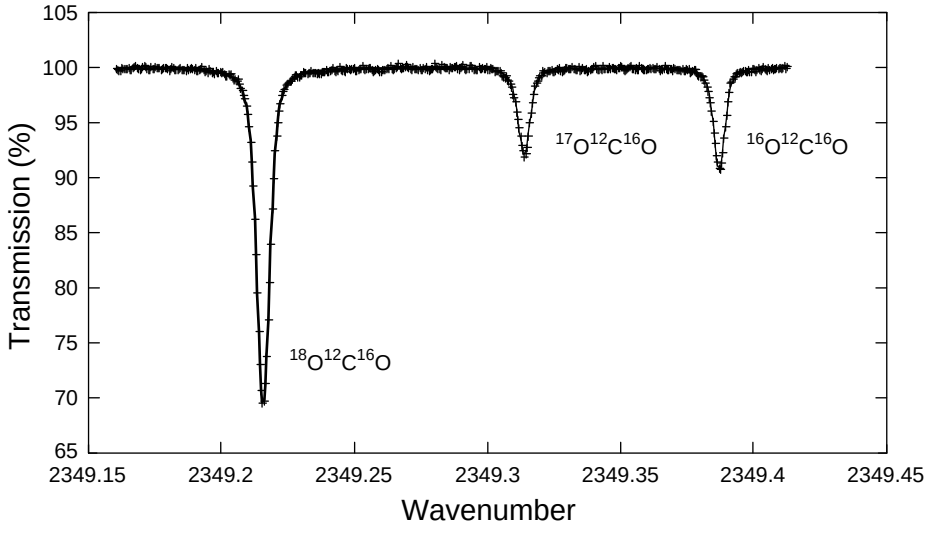

Figure 10 Direct-absorption spectrum of three different isotopes of natural CO_2 in air at a pressure of 3.6 kPa in an 18-m multipass cell. The probe beam makes 90 passes between the cell mirrors that are 20 cm apart. Given the cell transmission of 33%, the overall gain in signal strength is 30.[65] Interference fringes, due to light scattered by the mirrors, can be seen in the baseline.

sensitive to optical alignment and is typically on the order of 0.01–0.1% full-scale transmission (see Figure 10). Mirror drift and vibration can also become a problem, as they modulate the cell transmission. Mirror vibration, on the other hand, reduces the effective magnitude of interference fringes, as it scrambles the phase of optical feedback within the cell.

Cavity-enhanced spectroscopy is another method to increase the magnitude of absorption signal when only a small gas sample is available. Although its implementation and treatment are different from those described above for the multipass cell, the similarity is evident if one considers the idea of ''recycling'' photons in the probe beam. In this method, a weakly absorbing gas sample is placed inside a cavity, and the cavity is tuned to resonance with the probe beam, e.g. by a piezo-driven cavity mirror. At the peak of resonance, the cavity photon lifetime significantly exceeds the cavity round-trip time, thus increasing the probability of interaction with gas molecules in the sample. The effective number of passes between cavity mirrors, N, constitutes the improvement in the magnitude of absorption signal and is proportional to cavity finesse, F, which can be made rather high Equation (11):[43]

$$N = 2\frac{F}{\pi} \tag{11}$$

This approximation is good for weak absorption signals. A point to consider here is that increased gas absorption acts to reduce (''spoil'') the cavity finesse. This leads to nonlinear distortion of the observed molecular line shapes, making a quantitative measurement difficult. Another point to consider is that the cavity must be kept in resonance with the probe beam when the wavelength is tuned. A feedback loop is often used, which adjusts cavity length to track the changes in probe wavelength.[44] The rate of this electromechanical adjustment is much too slow to keep up with the laser tuning rate normally used for sensitive spectroscopy (more on this in section 4.3). Another form of cavity-enhanced spectroscopy is CRDS.[45–48]

An interesting, and remarkably simple, method to increase effective path length through a gas sample has recently been demonstrated by Tranchart et al.[49] It involves the use of a broadband integrating sphere whose inside walls diffuse light rather than reflect it. Effective path length for the probe beam inside the sphere is then given by Equation (12):

$$L_{\text{eff}} = \frac{4R}{3(1-\rho)} \tag{12}$$

Here R is the sphere radius, and ρ the diffuse reflection coefficient. Highly reflective walls ($\rho \to 1$) make it possible to obtain several meters of effective path length from a sphere as small as 10 cm in diameter. This method does not require sophisticated alignment of the probe beam into the sphere, but has the disadvantage of low detected power that is proportional to the ratio of the detector active area to the inside surface area of the sphere.

4.2 Balanced and Balanced-ratiometric Detection

Laser intensity noise and drift may limit the sensitivity with which absorption signals are measured. This is particularly true for gas lasers, or externally pumped solid-state ring lasers, where output power fluctuations induced by acoustic vibration and variations in pump power can exceed 1% full scale. Balanced detection may be used to recover small absorption signals in this situation. Noise detected in an equal-intensity replica of the probe beam, such as that created by a variable-ratio beamsplitter, is subtracted from noise detected in the probe beam, thus leaving only the uncompensated weak absorption signals of interest. A variable-ratio beamsplitter can be made by placing a polarization rotator (a half-wave plate) in series with a polarizing beamsplitter cube, as shown in Figure 11.

With the input polarization rotated about 45°, the beams emerging from the beamsplitter cube carry equal amounts of power P, and power noise ΔP. In the absence of absorption, the photo-currents generated by the (identical) signal and reference detectors can be subtracted to cancel each other exactly. If one of the beams is attenuated due to small absorption a, by a gas, for example, the balance of photo-currents is disturbed, and a signal is seen at the output of the amplifier (Equation 13):

$$V_{\text{out}} = R(I_{\text{ref}} - I_{\text{sig}}) = aRSP + aRS\Delta P \tag{13}$$

Here R is feedback resistance, and S is the detector element response, in amperes per watt. The second term constitutes noise, and the corresponding signal-to-noise ratio is $P/\Delta P$, independent of absorption a and limited basically by the quality of the probe beam. With the use of a single detector (no balancing), the signal-to-noise ratio would be a mere $aP/\Delta P$, i.e. absorption on the order of $a_{\min} = \Delta P/P$ would be indistinguishable from noise. In practice, adjustment of the beam splitting ratio (''zeroing'') is necessary to compensate for unequal response of the signal and reference detectors and difference in optical transmission of the signal and reference arms. Moreover, detectors must be sufficiently quiet to resolve fluctuations in probe power, and have equal delay times.

An advanced implementation of this method, proposed by Hobbs[50] and known as balanced ratiometric detection (BRD), uses circuitry to compute a log ratio of photo-currents, rather than their difference, and to cancel noise currents at the same time. The analog divider, whose simplified schematic is shown in Figure 12, uses

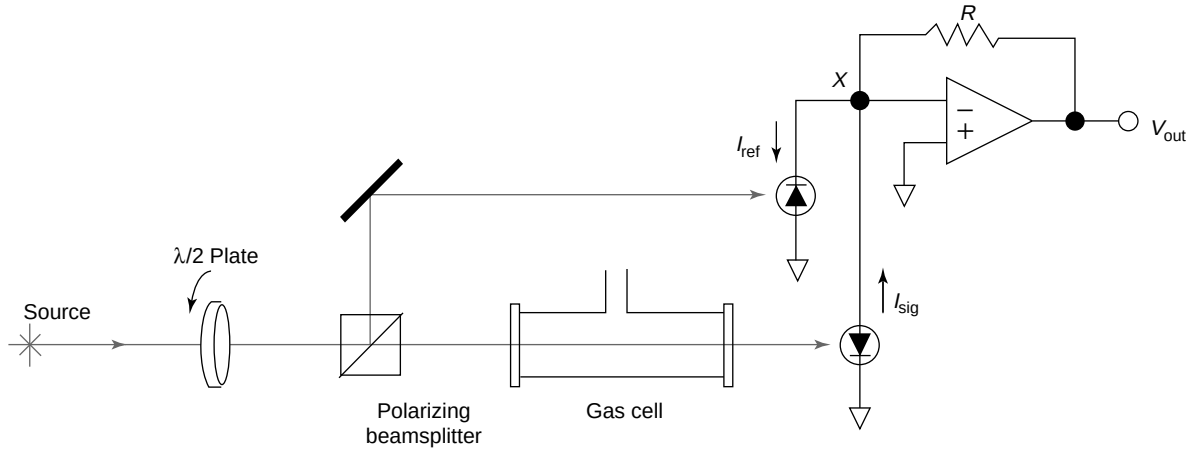

Figure 11 In the simple balanced detection method, a variable-ratio beamsplitter (a combination of a half-wave plate and a polarizing beamsplitter) is used to send nearly equal amounts of optical power to the signal and reference detectors, bringing V_{out} to zero. The balance at the current-summing junction X is maintained despite the laser amplitude noise, since the generated noise photo-currents are of opposite sign and nearly equal magnitude. An imbalance of photo-currents due to absorption in the signal beam results in a nonzero output.

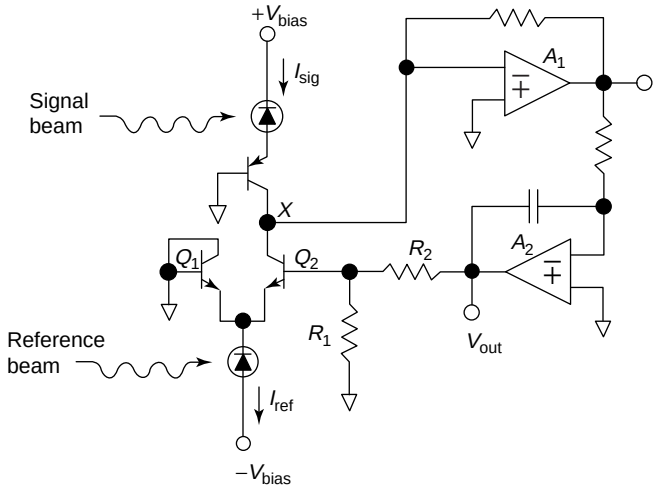

Figure 12 Circuit diagram of a balanced ratiometric detector. When currents at the summing junction X are balanced, the output of the error amplifier A_1 is zero, and the output of the integrator A_2 is proportional to $\ln(I_{ref}/I_{sig} - 1)$. A nearly perfect cancellation of noise photo-currents is achieved when $I_{ref} \approx 2I_{sig}$.

logarithmic conformance and tight symmetry of base-emitter curves of a matched transistor pair. The summing junction X is held at ground by an error amplifier A_1 whose output is integrated and applied to the base of transistor Q_2. When currents at the summing junction are balanced, the output of A_1 is zero, and the output of A_2 is given by Equation (14):

$$V_{out} = \left(\frac{R_1 + R_2}{R_1}\right)\left(\frac{k_B T}{e}\right) \ln\left(\frac{I_{ref}}{I_{sig}} - 1\right) \qquad (14)$$

Here k_B is the Boltzmann's constant, T the absolute temperature of the matched transistor pair Q_1Q_2, and e the electron charge. This scheme provides nearly perfect cancellation of noise currents when the reference beam carries roughly twice the power of the signal beam. The signal-to-noise ratio is thus increased well beyond the $P/\Delta P$ limit of the simple balancing scheme described above.

Since the signal versus reference current balancing is performed by means of electronic feedback, no physical adjustment of the beam splitting ratio is necessary. The BRD differential response to absorption signals is nonlinear, however, in that it depends on the ratio of the signal and reference currents which changes when the signal beam is partially absorbed. It also depends on temperature in the same way the transistor base-emitter voltage does, and additional compensation circuitry is needed to produce a useful output voltage V_{out}. Nevertheless, detection sensitivities presently achieved with BRD are quite stunning. Noise-equivalent absorbances as low as 2×10^{-7} Hz$^{-1/2}$ have been demonstrated by Allen et al.,[15] close to the limit imposed by the laser shot noise.

4.3 Frequency and Wavelength Modulation Spectroscopy

The ability of a diode laser to change its emission wavelength with injection current, and to do so very rapidly, is one of the reasons why the diode laser has been so effective in spectroscopic applications. The wavelength change is driven by two effects. Current-induced heating of the semiconductor junction leads to change in optical length of the laser cavity and red-shifts the laser's broad gain curve; this constitutes ''coarse'' tuning. The current

also controls the carrier density which in turn affects the cavity refractive index; this ''fine'' tuning mechanism is normally at least ten times less pronounced at low frequencies, but becomes dominant at high frequencies (>1 MHz). The tuning response thus depends on the modulation rate and ranges from 2–3 GHz mA^{-1} at low frequencies to under 300 MHz mA^{-1} at high frequencies. FM is always accompanied by amplitude modulation, as the injection current also controls the laser output power (Equation 15):

$$E(t) = A\{1 + m\cos(\Omega t)\}\sin\{wt + \beta\cos(\Omega t + \Phi)\} \quad (15)$$

Here $E(t)$ is the laser electric field, $w = 2\pi c/\lambda$ the laser frequency, and Ω the modulation frequency. By analogy with radio communication, w is the ''carrier'' frequency. The quantities m and β are the amplitude and FM indices, respectively, and Φ is the generally nonzero phase shift. Sine-wave modulation of the diode laser has the effect of creating multiple side bands in its otherwise nearly monochromatic emission spectrum. Each side band is separated from the carrier by an integer multiple of the modulation frequency Ω, and its relative intensity depends on β.[51]

In *FM spectroscopy*, for example, Ω significantly exceeds the laser linewidth that is typically several tens of megahertz (see Table 2), and m, β are both small, so that only the two first-order side bands, $w + \Omega$ and $w - \Omega$, have appreciable magnitude. After uniform attenuation, such as that encountered in nonresonant optical systems or media (e.g. imaging optics or vacuum), the side bands add up coherently with the carrier and balance each other to produce a beam of nearly constant intensity, A^2. If the attenuation strongly depends on frequency, however, as is the case with most gases, one of the side bands may become unbalanced and lead to the appearance of multiple harmonics of Ω in the detected laser intensity. The strength of absorption determines the magnitude of these harmonics, which may be measured separately and with high noise immunity, by using a lock-in amplifier for example. This is usually done while the laser carrier frequency w is scanned in the vicinity of the absorption line of interest.

This detection technique was first applied by Bjorklund to a CW dye laser.[52] It proved extremely powerful and is widely used in diode laser spectroscopy today, sometimes in modified form such as TTFM,[53] or amplitude modulated phase modulation (AMPM) spectroscopy.[54]

Wavelength modulation (WM) is really another form of FM spectroscopy, in which case the modulation frequency Ω is smaller than the laser linewidth, and the modulation indices m and β are both large.[55] The aforementioned side bands are then present to a very high order and, by virtue of their small separation from each other, merge into a continuous spectrum. The detection is again performed at the first, second, or higher harmonics of Ω as the laser carrier frequency w is scanned in the vicinity of a gas absorption line. WM spectroscopy dominates applications that rely on relatively low-speed detectors, and its sensitivity is limited by the laser amplitude $1/f$ noise.

4.4 Opto-acoustic Detection

Optical power absorbed by a gas from the probe beam leads to local heating of the gas and inevitably generates a local pressure differential. The pressure differential can, in principle, be detected as an acoustic pulse, some distance away from the source in any direction. A train of acoustic pulses at a known frequency (a ''tone''), such as that produced in a gas by a chopped beam, for example, can be detected using a microphone and a lock-in amplifier (see Figure 13). Since power absorbed from the probe beam depends on wavelength, a change in the intensity of the acoustic wave will result as the laser wavelength is tuned. Spectroscopy based on this principle is called *opto-acoustic spectroscopy* (*OAS*), or *photo-acoustic spectroscopy* (*PAS*). It has one important

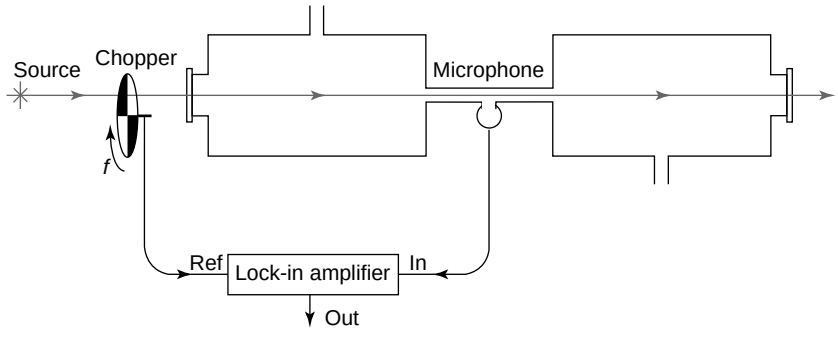

Figure 13 Simplified schematic of an opto-acoustic trace-gas detection system using a resonant acoustic cell. The beam is chopped at a frequency f, and a synchronous acoustic signal is detected by a microphone and a lock-in amplifier. Laser WM on and off the gas absorption line can produce an acoustic signal in much the same way as a chopped beam that is tuned to the line center.

advantage: in the absence of an absorbing gas there is no acoustic signal at all. External acoustic noise is unimportant, as it has random phase relative to the chopper and is therefore filtered away effectively by the lock-in amplifier. At some level of scrutiny, however, one might find interfering synchronous acoustic signals, such as those generated by the slightly absorbing windows of the sample cell, etc.

Since the opto-acoustic signal is proportional to beam intensity, the method has been used primarily with high-power infrared lasers, such as CO_2 and CO.[56] Low-power diode lasers are now being used as well, although gas cells are often built as acoustic resonators in order to increase the observed signal.[57] Surprisingly, the WM capabilities of diode lasers have found an elegant use in OAS – periodic tuning of the laser beam on and off the line resonance is also a good way to generate an acoustic wave. The theory for WM opto-acoustic signal generation involves coupled heat transfer and the Navier–Stokes equations[58] and is beyond the scope of this review.

5 EXAMPLES OF TRACE-GAS SENSORS

Here we will describe practical embodiments of the three types of laser spectroscopic sensors described in section 3. These are sensors based on NIR semiconductor lasers, lead-salt diode lasers, and diode laser frequency conversion. We have selected one instrument in each category as an example. These instruments have either stimulated a development of a broad range of similar sensors and related measurement techniques, or are an example of how multiple elements of the developed sensor technology function together in a complete spectroscopic instrument.

5.1 Gas Sensors Based on Near-infrared Diode Laser Overtone Spectroscopy

Intense development of gas sensors based on visible and NIR semiconductor lasers has been reported by groups in the USA, Europe, and Japan.[42,50] Applications in which these sensors find increasing use are combustion diagnostics, landfill emissions monitoring, natural-gas leak detection, optical sensing of flows, and industrial process control. One of the first applications of a NIR diode laser to gas detection was reported in 1992 by Uehara and Tai.[16] Figure 14 shows a schematic diagram of spectrometer built for the purpose of detecting methane in ambient air; it uses the same transmitter design in two remote detection schemes. The transmitter employs a 1.665-μm InGaAsP DFB diode laser developed

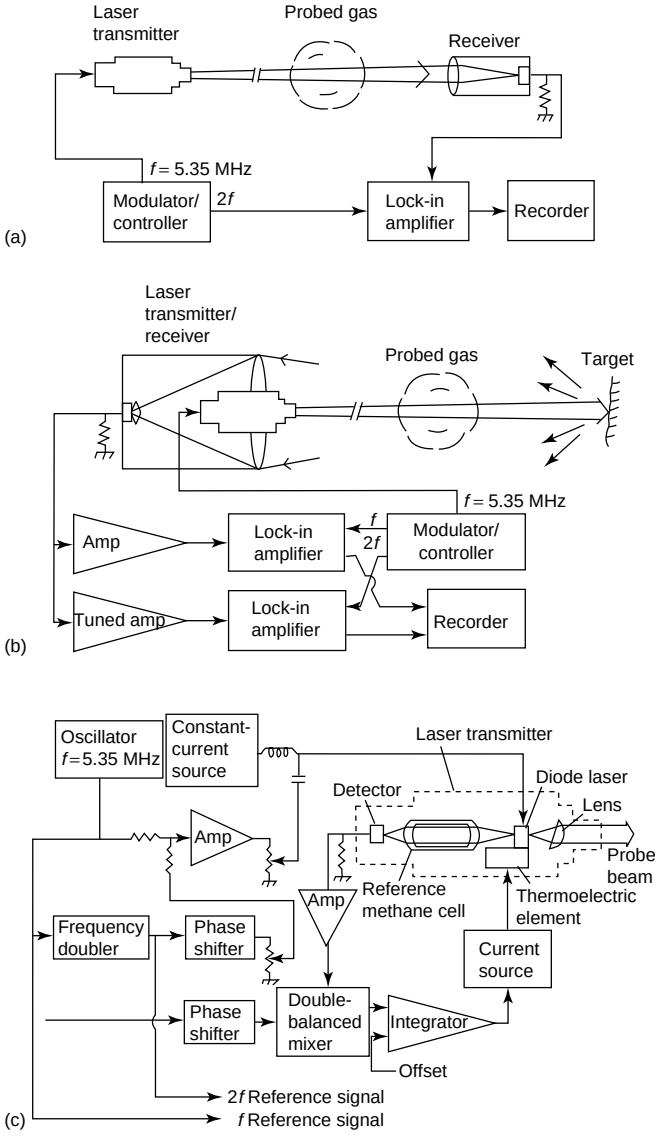

Figure 14 Diode laser methane sensor developed by Uehara and Tai,[16] in two remote detection schemes: transmission (a) and reflection (b). Schematics of the laser transmitter and its modulation controller are shown in (c). The transmitter housing is 4 cm in diameter.

specifically for the detection of methane in the $2\nu_3$ stretch vibration overtone band.

Laser emission in this instrument was available from both front and rear facets of the laser chip, and was detected using InGaAs pin photodiodes. The backward beam was used to dither-lock the laser wavelength to the center of the Q(6) line of methane in a reference cell. The 3-cm-long pyrex cell contained 200 torr of methane mixed with 560 torr of air. The cell windows functioned as lenses: one collimated the laser output while the other focused it on to a reference detector. The laser injection current was modulated at a rate of $f = 5.35$ MHz and the output of the

reference detector was processed by a double-balanced mixer to recover the amplitude of the first harmonic ($1f$) of the modulation frequency. Due to AM–FM cross-talk in the diode laser, the amplitude is nonzero even in the absence of absorption, therefore offset cancellation was used to produce a suitable error signal. The error signal crosses zero when the laser frequency is scanned across the center of the absorption line, and therefore can be used, upon integration, as feedback to control the laser temperature. Such a simple locking scheme was reported to stabilize the center wavelength to within 10 MHz of the absorption line. After frequency stabilization was obtained, a small-amplitude second harmonic ($2f$) signal was added to the laser injection current to compensate the residual $2f$ signal amplitude in the detected *forward* laser beam in the absence of absorption. Absorption caused by the presence of methane in the signal beam path was then registered as a positive $2f$ amplitude.

Figure 15 shows an example of instrument output in the transmission scheme, in which the second harmonic amplitude alone was recorded. The signal was caused by a 100-ppm mixture of methane in air in a 50-cm-long cell. The mixture was introduced into the cell and purged subsequently with nitrogen, resulting in zero output after two minutes. The signal-to-noise ratio in the trace corresponds to a minimum detectable CH_4 column density of 0.3 ppm m, achieved using a signal averaging time of 1.3 s.

In the reflection detection scheme, the registered amplitude of the second harmonic depends not only on absorption, but on the range to the scattering target and on the target reflectance (wooden boards and concrete blocks were used). The amplitude of the first harmonic is directly proportional to the received optical power; quantitative detection was thus performed by computing a ratio of the $2f$ and $1f$ amplitudes. The corresponding detection sensitivity was 50 ppm m. Although higher modulation frequencies are advantageous for noise reduction, caution must be exercised when they are used in the reflection scheme. When the beam path length from transmitter to receiver changes, the detected phases of the first and second harmonics also change. For a modulation frequency of 5.35 MHz, for example, a 3.5-m change in distance to target results in a 90° phase shift in the second-harmonic signal. The phase shift must be tracked and compensated in order to maintain maximum signal intensity.[51,59–63]

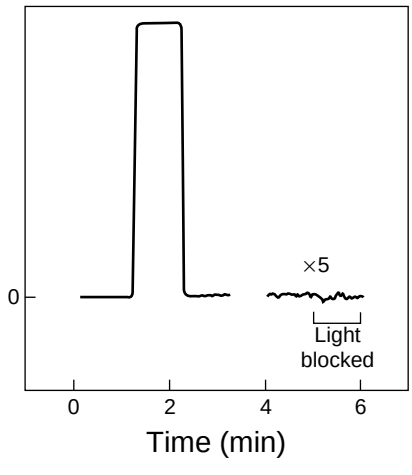

Figure 15 A $2f$ signal derived from a methane column density of 50 ppm m, lock-in-detected in the transmission scheme.[16] An inset on the right shows a ×5 magnified noise. No background shift is observed when the laser beam is blocked.

5.2 Tuneable Mid-infrared Diode Laser Absorption Spectrometer

In the area of MIR spectroscopy, the airborne tuneable laser absorption spectrometer (ATLAS) developed by Podolske and Loewenstein[19] is an example of a self-contained field instrument. It was developed for in situ measurement of nitrous oxide and carbon monoxide in the lower stratosphere from an ER-2 high-altitude aircraft. Optical layout of the instrument is shown in Figure 16. A pressure-regulated liquid nitrogen Dewar housed two different lead-salt lasers: one operating at 2169.2 cm^{-1} for the detection of carbon monoxide, and another at 1270.2 cm^{-1} for the detection of nitrous oxide. After collimation by an off-axis parabolic mirror,

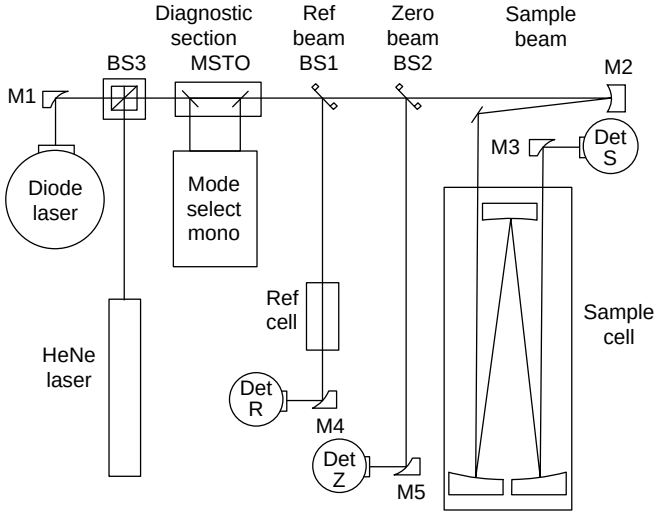

Figure 16 Optical layout of ATLAS, by Podolske and Loewenstein.[26] The sample, reference, and zero beams are detected by detector S, R, and Z, respectively. The zero beam has the same optical path length as the sample beam external to the White cell. A bidirectional alignment beam is formed by a beamsplitter BS3, illuminated by a HeNe laser. Mode selecting monochromator and mode-selecting transfer optics are optional.

the beam passed through an adjustable iris and the diagnostics section, and was then partitioned into three beams (sample, reference, and zero) by organic pellicle beamsplitters. These have the advantage of negligible thickness, hence beam displacement, and small, 5–10%, amount of deflected light, so that most of optical power remained in the sample beam. All three beams were measured by identical detectors. InAs detectors were used for the measurement of carbon monoxide at 4.7 μm, and HgCdTe detectors were used for the measurement of NO at 8 μm.

The instrument employed a noticeably complex laser current modulation and control scheme. Up to eight signals were superimposed to determine the instantaneous laser wavelength. Two of them were generated by an on-board computer for coarse and fine adjustment of the center wavelength. An adjustable amplitude, 2-kHz sine wave was added to perform WM spectroscopy. The 1*f* feedback signal derived from the reference channel was also added for locking of the laser wavelength to the center of a chosen absorption line (R(6) for carbon monoxide, or P(17) for nitrous oxide). A 12.8-kHz triangle wave dither was also added to suppress the effective magnitude of optical interference fringes. Another signal, a 125-Hz triangle wave was used to perform frequency scans for display purposes during instrument set-up. For the measurement of dark detector output, each period of the waveform was preceded by a 125 μs pulse that gated off the laser injection current. Each gate pulse was immediately followed by another, adjustable compensation pulse that provided heating of the laser junction that would have occurred in the absence of the gate pulse. The use of this compensation diminished the frequency drift associated with thermal recovery of the laser junction after each gate pulse.

In addition to optics, the ATLAS included a calibration and gas flow system (see Figure 17). The gas flow system routed and conditioned the flow of both the reference gas (100 ppm CO in nitrogen or 1000 ppm N_2O in nitrogen) and the sample. Air was sampled from outside the aircraft boundary layer by means of a narrow inlet tube, into which the air was propelled by ram pressure. The air then passed through a multipass cell with a residence time of approximately 1 s. Gas pressure and temperature inside the cell were maintained approximately equal that of the reference cell. Pressure equalization was accomplished by a feedback-controlled flow of the reference gas through a frit, rather than by making it a static specimen. Gas temperatures were made equal by heating of the incoming air to 25 °C; this also prevented temperature-induced alignment drift of the cell mirrors. For airborne measurements of nitrous oxide, whose typical abundance in air is approximately 300 ppb, the uncertainty of ±10% was estimated for ATLAS when using a signal averaging time of 1 s.

Apart from the systems described, the ATLAS had several electronic subsystems: laser control, analog signal processing, data acquisition, computer, and the pressure and temperature control. To fully appreciate the complexity of an actual instrument designed for field applications it is helpful to consult Fried et al.[24,64] and Nelson et al.[63]

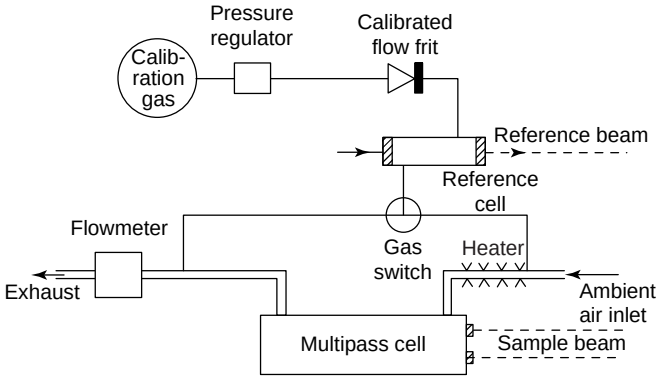

Figure 17 Diagram of the calibration and gas flow system of the ATLAS instrument.[26]

5.3 Gas Sensors Based on Difference Frequency Generation Absorption Spectroscopy

As described in section 3.4, difference frequency mixing of near infrared CW or pulsed laser sources is another convenient technique to access the molecular "fingerprint" region. Until recently, DFG sensors used discrete optical components to generate microwatt-level CW radiation.[31,65–68] More recently, the technologies developed for optical telecommunications – optical fiber, and rare-earth-doped fiber amplifiers – have been used in DFG sensors. This approach has improved the robustness and reliability, since it ensures permanent alignment of DFG pump channels without an increase in device complexity and cost. Lancaster et al.[67] built a portable sensor that is tuneable from 3.25 to 4.4 μm (see Figure 7).

A 20-mW ECDL with a tuning range of 814–870 nm, and an Yb-doped fiber amplifier seeded by a 1083-nm DBR were mixed in a multi-grating, temperature-controlled PPLN crystal. A 7.2-m long fiber amplifier, pumped by a 2-W diode laser at 975 nm, produced 590 mW output power when seeded with 12 mW. All amplifier components were packaged into a $9 \times 11 \times 2\,cm^3$ housing.

The pump wavelengths were combined by a low-loss fiber directional coupler. An $f = 1$ cm Achromat lens was used for imaging the fiber output into the PPLN crystal. A 19-mm long, 0.5-mm thick crystal was used, with broadband antireflection coatings applied to both end faces. The crystal had 8 domain grating periods,

from 22.4 to 23.1 μm in 0.1 μm increments, and was temperature-controlled in the 10–85 °C range. The DFG radiation is collimated by an $f = 5$ cm CaF_2 lens and the residual pump light was blocked by a germanium filter. The radiation is either focused directly on a Peltier-cooled HgCdTe detector or aligned through a compact multipass cell with an effective path length from 18 to 100 m. The data acquisition system used was similar to the one described in Topfer[31] and Lancaster,[70] and consists of a compact 16-bit analog-to-digital converter card and a notebook computer. Detector voltage was sampled at a rate of 100 kHz and processed using a 9 kHz digital low-pass filter.

Coarse frequency tuning of the Littman-cavity ECDL (Figure 4) was performed by rotation of its tuning mirror with respect to the diffraction grating. The advantage of this configuration is a beam direction and a point of origin that are independent of wavelength; this allows stable coupling into an optical fiber. Fine tuning and repetitive frequency scans of up to 25 GHz were accomplished by linear current modulation of the seed DBR diode laser at a rate of 200 Hz. The instrument linewidth of 42 ± 5 MHz was measured indirectly by spectroscopy of methane at low pressure. Figure 18 shows the individual direct absorption spectra of six species of interest for trace-gas detection that are within the 3.25–4.4 μm tuning range of the sensor, including CO_2, N_2O, H_2CO, HCl, NO_2 and CH_4. These spectra were taken at reduced pressure (88 torr) in a multipass cell using either calibrated gas mixtures or room air. Fitted Lorentzian line shapes were used to determine gas concentrations. Stability of instrument performance was assessed by monitoring of a methane line at 3028.751 cm^{-1} for an extended time period. In this experiment a calibrated 1772.7 ± 1 ppb mixture of CH_4 in air was continuously flown through the multipass cell at a pressure of 88 torr. Observed standard deviation of the measurement was 0.8% (15 ppb).

The instrument occupies a 45×45 cm^2 optical breadboard mounted in a reinforced plastic suitcase for portability and weighing 25 kg. Total power consumption is approximately 60 W. To provide a continuous gas flow through the multipass cell at a controlled pressure, a two-stage diaphragm pump was used in series with a pressure flow controller. A capacitive manometer was used to measure the gas pressure just prior to the multipass cell. For the measurement of reactive gases such as H_2CO, HCl and NO_2, the gas handling system was maintained at 40 °C.

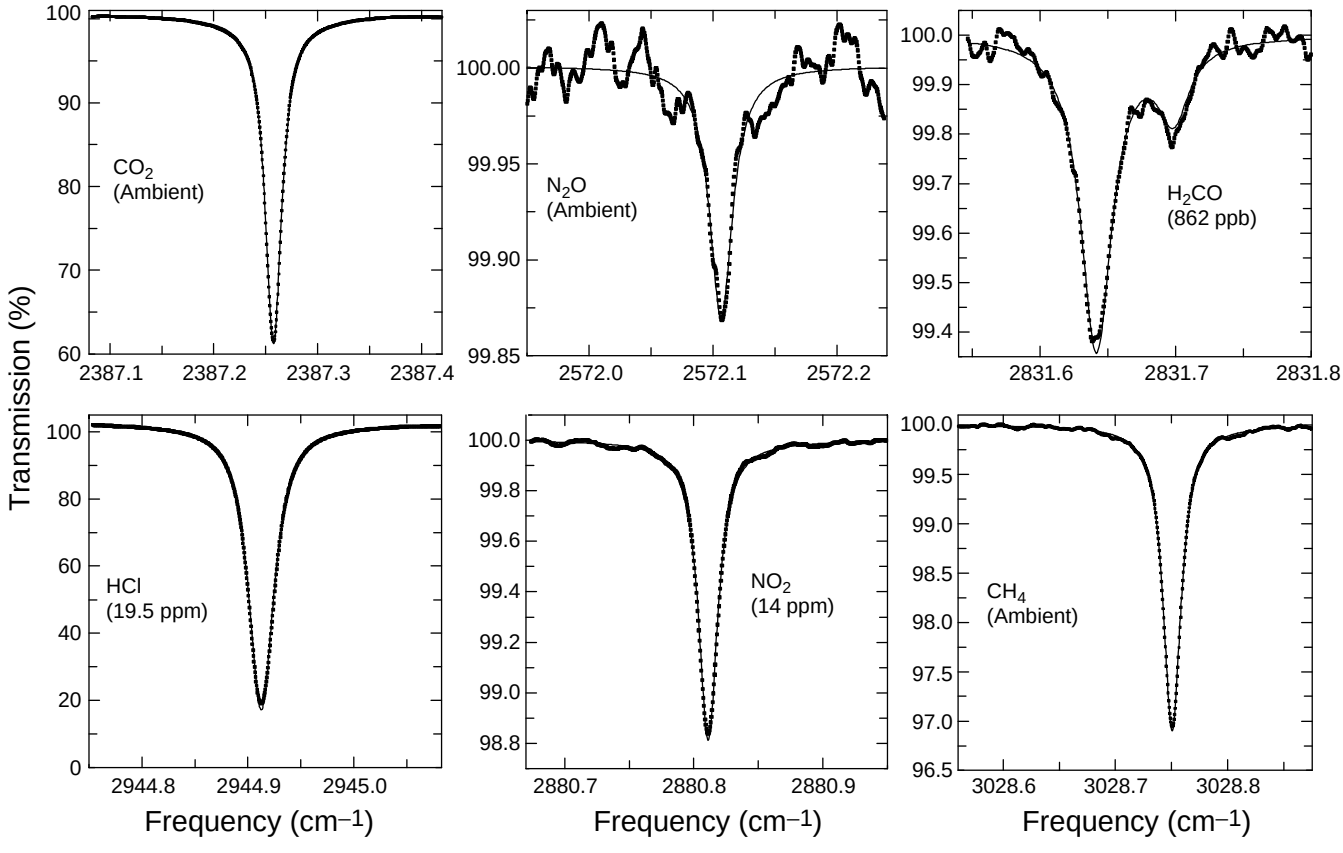

Figure 18 Detected absorption lines of several atmospheric trace gases.[69]

6 CONCLUSION

Spectroscopic sources for trace-gas monitoring, based on diode lasers, find increasing use in both established and new fields, including air-quality control; atmospheric chemistry; industrial, traffic, and agricultural emissions; chemical analysis and process control; and medical diagnostics. The moderate cost, compact gas sensor using all-solid-state technology is capable of highly sensitive, selective detection and real-time analysis of a large number of gas species by means of absorption spectroscopy in the overtone and fundamental spectral regions. Reductions in cost and complexity, coupled with improved reliability and ease of operation, are now made as a result of the availability of several novel technologies developed by the telecommunications and computer industries that are equally applicable and useful in state-of-the-art gas sensors.

ACKNOWLEDGMENTS

The authors gratefully acknowledge stimulating input by a number of present and previous coworkers, graduate students and collaborators. This chapter also benefited from helpful suggestions and critical comments by Mark Allen of Physical Sciences, Inc. (Boston) and Adam Holyoake (John Wiley & Sons, Inc.).

ABBREVIATIONS AND ACRONYMS

AMPM	Amplitude Modulated Phase Modulation
ATLAS	Airborne Tuneable Laser Absorption Spectrometer
BRD	Balanced Ratiometric Detection
CRDS	Cavity Ring Down Spectroscopy
CW	Continuous Wave
DBR	Distributed Bragg Reflector
DFB	Distributed Feedback
DFG	Difference-frequency Generation
ECDL	External-cavity Diode Laser
FM	Frequency Modulation
KTP	Potassium Titanyl Phosphate
MDC	Minimal Detectable Concentration
MIR	Mid-infrared
NIR	Near-infrared
OAS	Opto-acoustic Spectroscopy
PAS	Photo-acoustic Spectroscopy
TTFM	Two-tone Frequency Modulation
WM	Wavelength Modulation

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Differential Optical Absorption Spectroscopy, Air Monitoring by • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Laser-based Combustion Diagnostics • Photoacoustic Spectroscopy in Trace Gas Monitoring

REFERENCES

1. M.W. Sigrist (ed.), *Air Monitoring by Spectroscopic Techniques*, Chemical Analysis, John Wiley & Sons, Inc., New York, Vol. 27, 1994.
2. M.W. Sigrist, 'Air Monitoring, Optical Spectroscopic Methods', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, Inc., New York, 84–117, 1998.
3. A.G. Maki, J.S. Wells, *Wavenumber Calibration Tables from Heterodyne Frequency Measurements*, Spec. Publ. 821 (Natl. Inst. Stand. Technol., Gaithersburg, MD), 1991.
4. L.S. Rothman, R.R. Gamache, A. Goldman, L.R. Brown, R.A. Toth, H.M. Pickett, R.L. Poynter, J.-M. Flaud, C. Camy-Peyret, A. Barbe, N. Husson, C.P. Rinsland, M.A.H. Smith, 'The HITRAN database: 1986 edition', *Appl. Opt.*, **26**, 4058 (1987).
5. GEISA database, Laboratoire de Meteorologie Dynamique du CNRS, Ecole Polytechnique, F-91128 Palaiseau Cedex, France.
6. R.M. Mihalcea et al., *Appl. Phys. B* – Lasers and Optics, Special Issue, August (1998).
7. D.M. Sonnenfroh, M.G. Allen, *Appl. Opt.*, **36**, 18 (1997).
8. Norsk Elektro Optikk, 'Laser Gas Monitors, Technical Description', Norway, 1996.
9. Southwest Sciences Inc., 'Detection Limits', http://www.iac.net/~dchovde/.
10. Bovar – Western Research, 'Spectrascan: Technical Specifications', Ref. No. 908-0001 Rev. 0.
11. M.G. Allen et al., *Appl. Opt.*, **34**, 18 (1995).
12. M. Feher et al., *Appl. Opt.*, **32**, 12 (1993).
13. J. Humlicek, 'An Efficient Method for Evaluation of the Complex Probability Function: The Voigt Function and its Derivatives', *J. Quant. Spectrosc. Radiat. Transfer*, **21**, 309 (1979).
14. C.E. Wieman, L. Hollberg, 'Using Diode Lasers for Atomic Physics', *Rev. Sci. Instrum.*, **62**, 1 (1991).
15. M.G. Allen, K.L. Carleton, S.J. Davis, W.J. Kessler, C.E. Otis, D.A. Palombo, D.M. Sonnenfroh, 'Ultra Sensitive Dual Beam Absorption and Gain Spectroscopy:

Applications for Near-infrared and Visible Diode Laser Sensors', *Appl. Opt.*, **34**, 3240 (1995).
16. K. Uehara, H. Tai, 'Remote Detection of Methane with a 1.66 μm Diode Laser', *Appl. Opt.*, **31**, 809 (1992).
17. H. Tai, K. Yamamoto, M. Uchida, S. Osawa, K. Uehara, 'Long Distance Simultaneous Detection of Methane and Acetylene by Using Diode Lasers Coupled with Optical Fibers', *IEEE Photon. Technol. Lett.*, **4**, 804 (1992).
18. A.T. Schremer, C.L. Tang, *IEEE Photon. Technol. Lett.*, **2**, 3 (1990).
19. D.B. Oh, D.C. Hovde, 'Wavelength-modulation Detection of Acetylene with Near-infrared External Cavity Diode Laser', *Appl. Opt.*, **34**, 7002 (1995).
20. Z. Feit, D. Kostyk, R.J. Woods, P. Mak, 'Molecular Beam Epitaxy Grown PbEuSeTe Buried-heterostructure Lasers with Continuous Wave Operation @ 195 K', *Appl. Phys. Lett.*, **57**, 2891 (1990).
21. J. Reid, M. El-Sherbiny, B.K. Garside, E.A. Ballik, 'Sensitivity Limits of a Tunable Diode Laser Spectrometer with Application to the Detection of NO_2 @ the 100 ppt. Level', *Appl. Opt.*, **19**, 3349 (1980).
22. H.I. Schiff, D.R. Karecki, G.W. Harris, D.R. Hastie, G.I. Mackay, 'A Tunable Diode Laser System for Aircraft Measurements of Trace Gas', *J. Geophys. Res.*, **95**, 10 147 (1990).
23. M. Kohl, D. Wohrle, et al., 'Delayed Observation of Laser Induced Fluorescence for Imaging of Tumors', *Appl. Phys. B*, **57**, 131 (1993).
24. A. Fried, S. McKeen, S. Sewell, J. Harder, B. Henry, P. Golden, 'Photochemistry of Formaldehyde During the 1993 Tropospheric OH Photochemistry Experiment', *J. Geophys. Res.*, **102**, 6283 (1997).
25. C.R. Webster, R.D. May, 'Simultaneous In Situ Measurements and Diurnal Variations of NO, NO_2, O_3, jNO_2, CH_4, H_2O, and CO_2 in the 40–26-km Region Using an Open Path Tunable Diode Laser Spectrometer', *J. Geophys. Res.*, **92**, 11 931 (1987).
26. J. Podolske, M. Loewenstein, 'Airborne Tunable Diode Laser Spectrometer for Trace Gas Measurement in the Lower Stratosphere', *Appl. Opt.*, **32**, 5324 (1993).
27. A.S. Pine, *J. Opt. Soc. Am.*, **64**, 1683 (1974).
28. U. Simon, C.E. Miller, C.C. Bradley, R.G. Hulet, R.F. Curl, F.K. Tittel, 'Difference Frequency Generation in $AgGaS_2$ by Use of Single Mode Diode Laser Pump Sources', *Opt. Lett.*, **18**, 1062 (1993).
29. G. Rosenman, Kh. Garb, A. Skliar, M. Oron, D. Eger, M. Katz, 'Domain Broadening in Quasi-phase Matched Nonlinear Optical Devices', *Appl. Phys. Lett.*, **73**, 865 (1998).
30. L.E. Myers, W.R. Bosenberg, 'Periodically Poled Lithium Niobate and Quasi-phase-matched Optical Parametric Oscillators', *IEEE J. Quant. Electron.*, **33**, 1663 (1997).
31. T. Topfer, K.P. Petrov, Y. Mine, D. Jundt, R.F. Curl, F.K. Tittel, 'Room Temperature Mid-infrared Laser Sensor for Trace Gas Detection', *Appl. Opt.*, **36**, 8042 (1997).
32. G.D. Boyd, D.A. Kleinman, *J. Appl. Phys.*, **39**, 3597 (1968).
33. T.B. Chu, Q. Broyer, *J. Phys.*, **46**, 523 (1985).
34. L. Goldberg, W.K. Burns, R.W. McElhanon, 'Difference Frequency Generation of Tunable Mid-infrared Radiation in Bulk Periodically Poled $LiNbO_3$', *Opt. Lett.*, **20**, 1280 (1995).
35. M.L. Bortz, M.A. Arbore, M.M. Fejer, 'Quasi-phase Matched Optical Parametric Amplification and Oscillation in Periodically Poled $LiNbO_3$ Waveguides', *Opt. Lett.*, **20**, 49 (1994).
36. J. Webjörn, F. Laurell, G. Arvidsson, *J. Lightwave Technol.*, **7**, 1597 (1989).
37. C.E. Rice, *J. Solid State Chem.*, **64**, 188 (1986).
38. M.H. Chou, M.A. Arbore, M.M. Fejer, 'Adiabatically Tapered Periodic Segmentation of Channel Waveguides for Node-size Transformation and Fundamental Mode Excitation', *Opt. Lett.*, **21**, 794 (1996).
39. M.H. Chou, J. Hauden, M.A. Arbore, M.M. Fejer, '1.5-μm Band Wavelength Conversion Based on Difference Frequency Generation in $LiNbO_3$ Waveguides with Integrated Coupling Structures', *Opt. Lett.*, **23**, 1004 (1998).
40. K.P. Petrov, A.T. Ryan, T.L. Patterson, L. Huang, S.J. Field, D.J. Bamford, 'Mid-infrared Spectroscopic Detection of Trace Gases Using Guided-wave-difference-frequency Generation', *Appl. Phys. B*, **67**, 357 (1998).
41. W.B. Grant, R.H. Kagann, W.A. McClenny, *J. Air Waste Manage. Assoc.*, **42**, 18 (1992).
42. J.B. McManus, P.L. Kebabian, M.S. Zahniser, 'Astigmatic Mirror Multipass Absorption Cells for Long-path Length Spectroscopy', *Appl. Opt.*, **34**, 3336 (1995).
43. K. Nakagawa, T. Katsuda, A.S. Shelkovnikov, M. de Labachelerie, M. Ohtsu, 'Highly Sensitive Detection of Molecular Absorption Using a High Finesse Optical Cavity', *Opt. Commun.*, **107**, 369 (1994).
44. K.P. Petrov, S. Waltman, U. Simon, R.F. Curl, F.K. Tittel, E.J. Dlugokencky, L. Hollberg, 'Detection of Methane in Air Using Diode-laser Pumped Difference Frequency Generation Near 3.2 μm', *Appl. Phys. B*, **61**, 553 (1995).
45. G. Meijer, M.G.H. Boogaarts, R.T. Jongma, D.H. Parker, A.M. Wodtke, 'Coherent Cavity Ring Down Spectroscopy', *Chem. Phys. Lett.*, **217**, 112 (1994).
46. G. Meijer, M.G.H. Boogaarts, R.T. Jongma, D.H. Parker, A.M. Wodtke, 'Cavity Enhanced Absorption and Cavity Enhanced Magnetic Rotation Spectroscopy', *Rev. Sci. Instrum.*, **69**, 3763 (1998).
47. J.J. Scherer, J.B. Paul, C.P. Collier, A. O'Keefe, D.J. Rakestra, R.J. Saykally, *Spectroscopy*, **11**, 46 (1996).
48. J.J. Scherer, J.B. Paul, C.P. Collier, A. O'Keefe, D.J. Rakestra, R.J. Saykally, *Spectroscopy*, **14**, 24 (1999).
49. S. Tranchart, I.H. Bachir, J.-L. Destombes, 'Sensitive Trace Gas Detection with Near Infrared Laser Diodes and an Integrating Sphere', *Appl. Opt.*, **35**, 7070 (1996).
50. P.C.D. Hobbs, IBM Corp, Armonk, N.Y. 'Noise Canceling Circuitry for Optical Systems with Signal Dividing and Combining Means', US Patent No. 5,134,276 (July 28, 1992).

51. P. Werle, 'A Review of Recent Advances in Semiconductor Laser Based Gas Monitors', *Spectrochimica Acta*, Part A (1998).
52. G.C. Bjorklund, 'Frequency-modulation Spectroscopy: A New Method for Measuring Weak Absorption and Dispersions', *Opt. Lett.*, **5**, 15 (1980).
53. D.E. Cooper, T.F. Gallagher, 'Double Frequency Modulation Spectroscopy: High Modulation Frequency with Low-bandwidth Detectors', *Appl. Opt.*, **24**, 1327 (1985).
54. R. Großkloß, P. Kersten, W. Demtröder, 'Sensitive Amplitude- and Phase-modulated Absorption-spectroscopy with a Continuously Tunable Diode Laser', *Appl. Phys. B*, **58**, 137 (1994).
55. F.S. Pavone, M. Inguscio, 'Frequency and Wavelength Modulation Spectroscopies: Comparison of Experimental Methods Using an AlGaAs Diode Laser', *Appl. Phys. B*, **56**, 118 (1993).
56. M.W. Sigrist, S. Bernegger, P.L. Meyer, 'Atmospheric Pollution Monitoring Using CO_2 Laser Photoacoustic Spectroscopy and Other Techniques', *Rev. Sci. Instrum.*, **61**, 1779 (1990).
57. M. Fehér, Y. Jiang, J.P. Maier, A. Miklós, 'Optoaccoustic Trace-gas Monitoring with Near Infrared Diode Lasers', *Appl. Opt.*, **33**, 1655 (1994).
58. A. Miklós, Z. Bozóki, Y. Jiang, M. Fehér, 'Experimental and Theoretical Investigation of Photoacoustic Signal Generation by Wavelength Modulated Diode Lasers', *Appl. Phys. B*, **58**, 483 (1994).
59. M.G. Allen, 'Diode-laser Absorption Sensors for Gas-dynamic and Combustion Flows', *Meas. Sci. Technol.*, **9**, 545 (1998).
60. D.M. Sonnenfroh, W.J. Kessler, J.C. Magill, B.L. Upschulte, M.G. Allen, J.D.W. Barrick, 'In-situ Sensing of Tropospheric Water Vapor Using an Airborne Near Infrared Diode Laser Hygrometer', *Appl. Phys. B*, **67**, 275 (1998).
61. R.M. Mihalcea, M.E. Webber, D.S. Baer, R.K. Hanson, G.S. Feller, W.B. Chapman, 'Diode-laser Absorption Measurements of CO_2, H_2O, N_2O, and NH_3 near 2.0 μm', *Appl. Phys. B*, **67**, 283 (1998).
62. G. Modugno, C. Corsi, M. Gabrysch, F. Marin, M. Inguscio, 'Fundamental Noise Source in a High Sensitivity Two Tone Frequency Modulation Spectrometer and Detection of CO_2 at 1.6 μm and 2 μm', *Appl. Phys. B*, **67**, 289 (1998).
63. D.D. Nelson, M.S. Zahniser, J.B. McManus, C.E. Kolb, J.L. Jimenez, 'A Tunable Diode Laser System for the Remote Sensitive of On-road Vehicle Emissions', *Appl. Phys. B*, **67**, 433 (1998).
64. A. Fried, B. Henry, B. Wert, S. Sewell, J.R. Drummond, 'Laboratory, Ground-based, and Airborne Tunable Diode Laser Systems: Performance Characteristics and Applications in Atmospheric Studies', *Appl. Phys. B*, **67**, 317 (1998).
65. K.P. Petrov, R.F. Curl, F.K. Tittel, 'Compact Laser Difference Frequency Spectrometer for Multi-component Trace Gas Detection', *Appl. Phys. B*, **66**, 531 (1998).
66. M. Seiter, D. Keller, M.W. Sigrist, 'Broadly Tunable Difference-frequency Spectrometer for Trace Gas Detection with Nonlinear Critical Phase-matching in $LiNbO_3$', *Appl. Phys. B*, **67**, 351 (1998).
67. B. Sumpf, D. Rehle, T. Kelz, H.-D. Kronfeldt, 'A Tunable Diode-laser Spectrometer for the MIR Region Near 7.2 μm Applying Difference Frequency Generation in $AgGaSe_2$', *Appl. Phys. B*, **67**, 369 (1998).
68. W. Chen, G. Mouret, D. Boucher, 'Difference-frequency Laser Spectroscopy Detection of Acetylene Trace Constituent', *Appl. Phys. B*, **67**, 375 (1998).
69. D.G. Lancaster, D. Richter, F.K. Tittel, Accepted by *Appl. Phys. B*, 459 (1999).
70. D.G. Lancaster, D. Richter, R.F. Curl, F.K. Tittel, 'Real-time Measurements of Trace Gas Using a Compact Difference Frequency Based Sensor Operation at 3.5 μm', *Appl. Phys. B*, **67**, 339 (1998).

Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

David W.T. Griffith and Ian M. Jamie
University of Wollongong, Wollongong, Australia

1 Introduction **1**

2 Fourier Transform Infrared Spectrometry Essentials **2**
2.1 Basic Principles 2
2.2 Spectral Resolution 3
2.3 Optical Throughput 3
2.4 Advantages of Fourier Transform Infrared Spectrometry 4
2.5 Signal-to-noise Ratio 4
2.6 Sources and Detectors 4

3 Extractive and Open-path Measurements **5**
3.1 Multiple-pass Absorption Cells for Gas Samples 5
3.2 Open-path Systems 6
3.3 Open-path versus Extractive Sampling 7
3.4 Sampling Methods for Extractive Analysis 7
3.5 Detection Limits 9
3.6 Choice of Spectral Resolution 11

4 Atmospheric Infrared Spectra **11**

5 Calibration and Quantitative Analysis **12**
5.1 Measurement of Background Spectra 13
5.2 Peak Height or Area Analysis 13
5.3 Interactive Subtraction 13
5.4 Classical Least Squares 14
5.5 Partial Least Squares 14
5.6 Calculation of Synthetic Calibration Spectra 15
5.7 Limitations to Beer's Law 16
5.8 Nonlinear Least Squares Analysis 16

6 Selected Applications **17**
6.1 Early History – Measurements of Urban Pollution Chemistry by Fourier Transform Infrared Spectrometry Using Long-path Cells 17
6.2 Open-path Monitoring of Air Pollution and Fugitive Gases 17
6.3 High-precision Measurements of CO_2, CO, CH_4, N_2O and $^{13}CO_2/^{12}CO_2$ in Clean Air 18
6.4 Measurements of Trace Gas Fluxes by Fourier Transform Infrared Spectrometry and Micrometeorological Techniques 18
6.5 Open-path and Extractive Measurements of Biomass Burning and Combustion Gases 19
6.6 Volcanic Emissions 19
6.7 Remote Sensing by Solar Fourier Transform Infrared Spectrometry 20
6.8 Infrared Emission Spectrometry of Plumes and Stack Emissions 21
6.9 Cryogenic Sampling: Matrix Isolation 21

Abbreviations and Acronyms **21**

Related Articles **22**

Appendix: Spectral Libraries and Databases **22**
Electronic Databases 22
Molecules in the HITRAN Database 24

References **24**

Fourier transform infrared (FTIR) spectrometry is the dominant technique used to measure the infrared (IR) absorption and emission spectra of most materials, with substantial advantages in signal-to-noise ratio (S/N), resolution, speed and detection limits over conventional dispersive spectroscopy. In this application, FTIR spectrometry is used to measure the trace gas composition of the atmosphere. Applications to both clean and polluted air are described. Techniques include measurements over open paths in situ, sampling and measurement in closed cells in the field or laboratory and, briefly, remote sensing using the sun, sky or natural hot objects as an IR radiation source. The techniques are applicable to a very wide range of compounds, including labile or reactive species, and all species are measured simultaneously. Detection limits vary with the application but are typically of the order of parts per billion (nmol mol^{-1}) in air.

1 INTRODUCTION

Atmospheric analysis by IR spectrometry began over a century ago with measurements of the solar absorption spectrum from the earth's surface, providing the first identification of minor atmospheric gases such as nitrous oxide[1] and carbon monoxide.[2] The technique developed rapidly from the late 1940s and early 1950s

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

using grating spectrometers. Around the same time, Stephens, Hanst and co-workers used long-path cells to study the chemistry of urban pollution, identifying for example the existence of peroxyacetyl nitrate (PAN), an irritant and intermediate in ozone smog chemistry, in urban smog.[3,4] However, it was the advent of commercially available FTIR spectrometers in the 1970s that really allowed the techniques to flourish. FTIR spectrometry provides significant advantages in S/N and spectral resolution over conventional dispersive spectrometry, advantages that can be turned to better detection limits and discrimination between absorbing species. Since the early 1970s, the number and scope of FTIR atmospheric measurements has increased steadily to include

- measurements of ambient air at the earth's surface, in both clean and polluted air, by either extractive sampling or over open paths in ambient air;
- remote sensing of the whole atmosphere by measurement of the solar absorption spectrum through the atmosphere, or the atmospheric emission spectrum, from the earth's surface, aircraft and space;
- IR emission measurements for analysis of hot trace gas sources such as stack emissions and volcanic gases.

FTIR atmospheric analysis has a number of advantages over other techniques, including simultaneous analysis of a wide range of trace gas species, in situ, nonintrusive and nondestructive measurements requiring no chemical transformation of the sample, and speed, typically being able to analyze samples in a few seconds to minutes. The principal limitation is the typical detection limits, of the order of a few nanomoles per mole (parts per billion by volume, ppbv) in air.

The objective of this article is to introduce the reader to the principles and applications of FTIR spectrometry in the measurement of atmospheric composition. Within this scope, we focus on the first of the applications above – measurements of ambient air at the earth's surface. The principles are readily extended to the other methods, and some examples are given in section 6. The article is not intended to be a comprehensive review, nor is it prescriptive in the sense of being a ''how-to'' manual. It is rather an exposition of the general principles and a selection of applications to illustrate the range of the technique. With this knowledge, extended by relevant references to the literature, it is our intention that the analyst can design and construct his or her own experiment according to the particular requirements of the measurement. Section 2 describes some essential underlying principles of FTIR spectrometry that are required throughout the discussion. Section 3 describes specific details of the FTIR technique as applied to atmospheric measurements. Section 4 provides a brief introduction to the nature of atmospheric IR spectra, and section 5 reviews methods for quantitative analysis of spectra to retrieve atmospheric concentrations. Finally, section 6 describes a selection of applications that illustrate the range of applicability of the technique. Additional information can be found in the books and review articles referred to in each section.

2 FOURIER TRANSFORM INFRARED SPECTROMETRY ESSENTIALS

2.1 Basic Principles

FTIR spectrometry has almost completely replaced dispersive spectrometry in the IR region during the 1980s and 1990s. The principles and practice of FTIR spectrometry are well documented in many sources,[5–7] including articles in this encyclopedia, and it is not the purpose of this article to repeat them. However, a reasonable understanding is crucial in optimizing FTIR operating conditions for atmospheric analysis, and in this section we provide a basic introduction to FTIR spectrometry and in particular the principles governing S/N, spectral resolution and optical throughput.

Figure 1 is a schematic diagram of a typical FTIR spectrometer based on the Michelson interferometer.

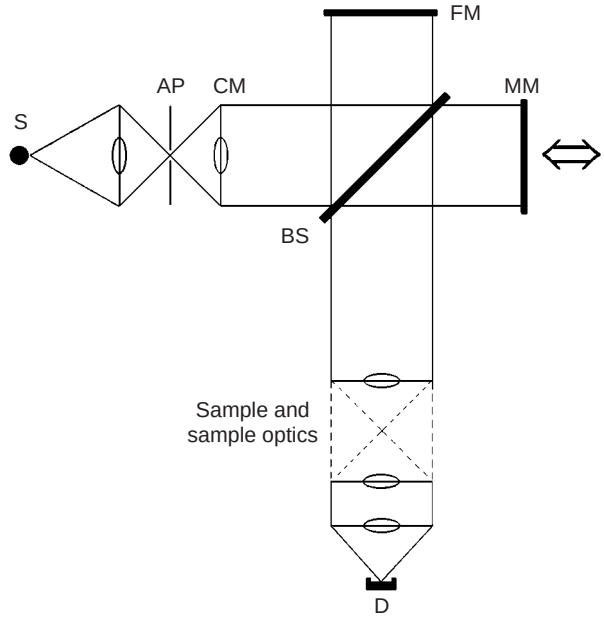

Figure 1 Schematic diagram of an FTIR spectrometer. S = source; AP = aperture; CM = collimating mirror; BS = beamsplitter; FM = fixed mirror; MM = moving mirror; D = detector.

Radiation from a source S is focused on to an entrance aperture AP and collimated by a mirror or lens CM. The collimated beam is split into two by the beamsplitter BS. The two beams travel to a fixed mirror FM and a moving mirror MM and return to and recombine at the beamsplitter, half the intensity being directed through the sample to the detector and half returning to the source. (When FM and MM are corner-cube mirrors, the beam returning to the source arm of the interferometer may be laterally displaced and may also be used for sample measurement.) In operation, the moving mirror MM is scanned at a nearly constant speed through the point at which the mirrors FM and MM are equidistant from the beamsplitter and beyond. When the mirrors are equidistant from the beamsplitter, the two beams have an equal optical pathlength or zero optical path difference (OPD), and the recombining beams are perfectly in-phase and interfere constructively in the output beam. As the mirror MM is scanned away from the equidistant point and the OPD increases, the two beams move progressively in and out of phase, and the output intensity measured by the detector for each monochromatic frequency from the source varies sinusoidally with OPD. The variation of output intensity with OPD is called the interferogram, $I(x)$.

In an ideal spectrometer, if the source were monochromatic, that is, a single frequency of radiation of wavenumber $\tilde{\nu}$ and intensity $B(\tilde{\nu})$, the interferogram would be a pure cosine wave (Equation 1):

$$I(x) = B(\tilde{\nu}) \cos 2\pi\tilde{\nu}x \tag{1}$$

If the source is broad band and contains many frequencies, the interferogram is a sum or integral over all frequencies present (Equation 2):

$$I(x) = \int B(\tilde{\nu}) \cos 2\pi\tilde{\nu}x \, \mathrm{d}\tilde{\nu} \tag{2}$$

Equation (2) is one half of a Fourier transform pair, and the spectrum $B(\tilde{\nu})$ can be recovered by computing the Fourier transform of the interferogram (Equation 3):

$$B(\tilde{\nu}) = \int I(x) \cos 2\pi\tilde{\nu}x \, \mathrm{d}x \tag{3}$$

Equations (2) and (3) provide the simplest description of the operation of an ideal FTIR spectrometer. We provide further details only as they become necessary for our discussion of atmospheric composition measurement.

2.2 Spectral Resolution

If the source spectrum is spread over a range of frequencies, ideally they will all interfere constructively at zero path difference (ZPD) and progressively go out of phase as the OPD increases. Thus the interferogram of a broad-band source will be sharply peaked at ZPD and decay as OPD increases. Two frequencies will go out of phase with the OPD quickly if they are widely spaced, and slowly if they are close together. The more closely spaced the two frequencies, the greater must be the maximum OPD to separate them. The apodized spectral resolution $\Delta\tilde{\nu}$ of an FTIR spectrometer is usually defined as the inverse of the maximum OPD (Equation 4):

$$\Delta\tilde{\nu} = \frac{1}{\mathrm{OPD}_{\max}} \tag{4}$$

2.3 Optical Throughput

In Figure 1, the source is focused on to an entrance aperture AP and collimated by the mirror CM. The throughput, Ω, of the optical system defines the amount of energy transmitted through the optical system. Throughput is analogous to electrical current through a circuit or flow through a pipe – in the absence of losses such as leaks, it is conserved. It is defined by the product of the area of the beam and its solid angle at any focus, for example, the aperture area and the solid angle subtended at the aperture by the collimator, or equivalently the product of the area of the collimated beam and the solid angle of its divergence (Equation 5):

$$\Omega = \frac{\pi}{4}\phi_{\mathrm{AP}}^2 \, \Theta_{\mathrm{AP}} = \frac{\pi}{4}\phi_{\parallel}^2 \, \Theta_{\parallel} \tag{5}$$

where ϕ_{AP} and Θ_{AP} are, respectively, the aperture diameter and the solid angle of the beam at the aperture focus (i.e. the solid angle subtended by CM at AP in Figure 1) and $\phi_{\parallel}$ and $\Theta_{\parallel}$ are the diameter and divergence solid angle of the collimated beam. (For a cone of light with half-angle α, the solid angle $\Theta = 2\pi(1 - \cos\alpha)$. If α is small, from the series expansion of $\cos\alpha$, $\Theta = \pi\alpha^2$. A divergent beam is often described by its f number, the ratio of the focal distance to the diameter of the (assumed circular) collecting mirror or lens. Thus an $f/4$ beam has a focal distance-to-diameter ratio of 4. The half-angle of the divergent beam, α, is related to the f/number of the beam by $f/\mathrm{no.} = 1/2\alpha$.) In some low-resolution spectrometers there is no collimator aperture, and the source or detector acts as the effective aperture and limits the throughput.

To obtain a perfectly collimated (i.e. perfectly parallel) beam, the aperture must be infinitesimally small, and the throughput would be zero. The larger the aperture, the greater is the throughput, but the more divergent is the collimated beam. However, the beam divergence in the interferometer, and hence the throughput, are limited by the spectral resolution required because rays traversing the interferometer at different angles to the true optical axis have different path differences

for a given mirror displacement, and their individual contributions to the total interferogram signal "smear out" the OPD at each mirror displacement. Thus the higher the resolution, the less divergence can be tolerated. The optimum throughput is the largest consistent with the spectral resolution at the highest frequency of interest. Quantitatively, the maximum throughput is proportional to the spectral resolution and in practical terms most medium- and high-resolution commercial spectrometers allow a choice of the optimum throughput via the entrance aperture diameter ϕ_{AP} according to Equation (6):

$$\phi_{AP} = 2FL_c \sqrt{\frac{\Delta\tilde{\nu}}{\tilde{\nu}_{max}}} \tag{6}$$

where FL_c is the collimator focal length (AP–FM in Figure 1), $\Delta\tilde{\nu}$ is the spectral resolution (defined by Equation 4) and $\tilde{\nu}_{max}$ is the maximum frequency to be measured. In open-path optical systems for atmospheric analysis, the throughput may be limited by the external optical system rather than the spectrometer resolution and becomes an important design parameter. This is discussed in section 3.2.

2.4 Advantages of Fourier Transform Infrared Spectrometry

One of the two major advantages of FTIR spectrometry over dispersive spectrometry is the multiplex or Fellgett advantage, which arises because the detector detects and measures all frequencies in the signal simultaneously, unscrambling them later with the Fourier transform. If the spectrum contains N resolution elements (e.g. a spectrum from 400 to 4000 cm^{-1} at 1 cm^{-1} resolution has 3600 resolution elements), each individual element is measured for N times longer than in an otherwise equivalent dispersive spectrometer which measures the resolution elements consecutively. Provided that the noise is detector limited, this leads immediately to a $\sqrt{N}$ improvement ($\times 60$ in the above example) in S/N; this is the multiplex advantage. The second major advantage is the throughput or Jacquinot advantage. The throughput of both FTIR and dispersive spectrometers is limited by spectral resolution – in the case of FTIR it is the entrance aperture which is limiting and in the case of dispersive instruments it is the monochromator entrance slit. At a given resolution the FTIR throughput is considerably greater (typically 10–100 times) than that of a typical grating spectrometer.

2.5 Signal-to-noise Ratio

The precision and accuracy of a quantitative measurement are determined primarily by the single-beam spectrum S/N in the measured spectrum. Following Griffiths and de Haseth,[5] the S/N is given by Equation (7):

$$\text{S/N} = \frac{S(\tilde{\nu})\xi\Delta\tilde{\nu}\Omega\sqrt{t}}{NEP} \tag{7}$$

where $S(\tilde{\nu})$ is the source radiance in W cm^{-2} sr^{-1} $(cm^{-1})^{-1}$, $\xi(\tilde{\nu})$ is the efficiency of the optical system (fraction of source energy at $\tilde{\nu}$ transmitted), $\Delta\tilde{\nu}$ is the spectral resolution (cm^{-1}), Ω is the optical throughput (cm^2 sr), t is the measurement time (s) and NEP is the noise-equivalent-power of the detector (W $Hz^{-1/2}$). Of these quantities, S and ξ are dependent on the source type and temperature and the spectrometer design and cannot normally be influenced by the experimenter. The detector NEP may be improved by detector choice, as described below. However for a given hardware configuration, S/N is determined by a trade-off between resolution, throughput and measurement averaging time. If the resolution is doubled (i.e. $\Delta\tilde{\nu}$ is halved), S/N is halved and the measurement time required to obtain equivalent S/N increases fourfold. If in addition the optimum throughput is maintained (halved) by reducing the aperture size, the measurement time increases a further fourfold, a total of 16-fold. Clearly, resolution is an expensive commodity on S/N. We shall analyze this trade-off further as we discuss different optical configurations and applications for gas measurements.

2.6 Sources and Detectors

Good throughput requires a source with high brightness (i.e. high power emitted per unit area), which for a blackbody emitter depends only on temperature. However, high-temperature filament-type sources such as quartz halogen lamps or xenon arc lamps do not make effective IR sources because the glass or quartz envelope is not transparent in the IR region, and not hot enough to be an effective source itself. Hence the most common IR sources must operate in air, which limits their temperature owing to oxidation. The most commonly used source is the resistively heated SiC globar, typically operating at 1200–1600 K. The size of the source need only be sufficient to fill the field of view of the detector through the optical system (at the largest collimator aperture employed). Once the detector field of view is filled, the only means to improve throughput is to increase the temperature of the source.

Although there are numerous types of IR detector, three types dominate general use in the mid-IR region. Pyroelectric detectors such as deuterated triglycine sulfate (DTGS) are thermal detectors; they are the cheapest and simplest, operating at room temperature, but are also both the slowest and the least sensitive, with response times of the order of milliseconds and detectivities of

Table 1 Characteristics of four commonly used detectors for FTIR spectrometry

Detector	Wavenumber range (cm^{-1})	D^* at peak $\tilde{\nu}$ ($cm\,Hz^{1/2}\,W^{-1}$)
DTGS	200–10 000	3×10^8
MCT (wide band)	500–6000	1×10^{10}
MCT (narrow band)	800–6000	4×10^{10}
InSb	1800–10 000	2×10^{11}

the order of $10^9\,cm\,Hz^{1/2}\,W^{-1}$. (Detectivity is defined as $D^* = \sqrt{A_D}/NEP$, where A_D is the detector area. D^* is a measure of the detector's sensitivity on S/N performance independent of its area.) Mercury cadmium telluride (MCT) and indium antimonide (InSb) detectors are quantum detectors and have both faster response times (microseconds) and higher detectivities (10^{10}–$10^{11}\,cm\,Hz^{1/2}\,W^{-1}$) than pyroelectric detectors. Table 1 shows a comparison of typical spectral coverage and detectivity for commonly used, commercially available 1-mm detectors.

The best choice of detector depends on the application. Pyroelectric detectors do not require cryogenic cooling and may be the only choice if liquid nitrogen supply is difficult, as in some in situ and field measurements. For low-resolution studies with high optical throughput, they may be perfectly adequate. In low light level situations (e.g. large multipass cells or long open-path optics as described below), MCT and InSb detectors provide better S/N in spectra (lower NEP, see Equation 7). However, the response of both detectors saturates at high light levels such as in the unattenuated beam of a typical commercial spectrometer, requiring either that overall throughput be reduced (which reduces S/N) or, preferably, that an optical interference filter be used in the optical train to pass only energy in the spectral range of interest. InSb detectors provide the best performance, but only at frequencies above $1800\,cm^{-1}$.

3 EXTRACTIVE AND OPEN-PATH MEASUREMENTS

The preceding section outlined essentials of FTIR spectrometry applicable to any FTIR analysis. This section is concerned with the particular case of measurements in ambient air. There is a fundamental choice of two strategies: to collect the sample and contain it in a closed cell for measurement (extractive sampling), or to collect the spectra over an open path in air. In this section we describe simple and multiple pass (multipass) absorption cells, open-path optical systems and air sampling methods, and summarize the advantages and disadvantages of each.

Concentration detection limits decrease linearly with increasing sample pathlength and S/N in the spectrum. For typical molecular absorptivities, simple short-path cells of 10–20 cm pathlength suffice only to measure relatively concentrated air samples such as combustion gases; pathlengths of tens of meters and more are required to obtain detection limits in the parts per million (by volume; $1\,ppmv = 1\,\mu mol\,mol^{-1}$) range and below in ambient air. Long optical paths are obtained by either multiple passing of the beam through the sample in an open or closed cell, or by collimating and projecting the IR beam over a long path in open air. These options are discussed in the following sections. Detection limits are discussed in detail in section 3.4.

3.1 Multiple-pass Absorption Cells for Gas Samples

To achieve long pathlengths in an absorption cell of practical size, the beam must be passed through the air sample many times. The most common multipass cell is the White cell, first described by White.[8] In the White cell (Figure 2), typically located in the focused beam in the spectrometer sample compartment, the convergent FTIR beam is focused into one end of the cell in the plane of the field mirror (FM) and reflects many times between the FM and objective mirrors (OM) before emerging, divergent, on the opposite side of the FM. The beam is refocused on to the FM on each pass so that, after the first pass, the only losses are due to reflective losses on the mirrors. Typically, up to 40 passes of the cell are possible, allowing pathlengths up to 40 m from a cell of 1-m base path. Several variants of the original White cell that double or triple the maximum number of reflections by adding

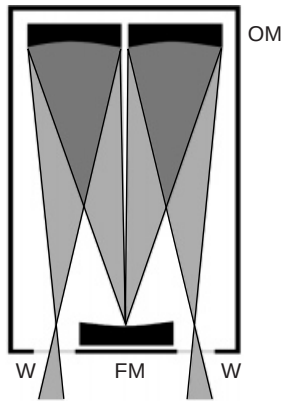

Figure 2 Schematic diagram of a White cell, shown set for four passes. The beam enters the cell through one of the windows, focused in the plane of the FM. The OMs collect and refocus the beam on to the FM on every pass. The FM images each OM on to the other. The number of passes is selected by lateral adjustment of one of the OMs such that the beam is refocused at a different spot on the FM at each pass until it exits through the second window after the requisite number of passes.

retroreflectors have been described.[9–12] Multipass cells with pathlengths from <2 m (volume ca. 100 mL) to >100 m (volume 20 L) are commercially available.

For N reflections and a mirror reflectivity ρ, the transmission due solely to reflective losses is ρ^N. For example, for 40 passes this corresponds to 67% and 45% transmission for reflectivities of 0.99 (e.g. silver) and 0.98 (e.g. gold), respectively. As the pathlength in a White cell is increased, the absorption increases but the radiation intensity and hence S/N decrease; the optimum ratio of absorption to noise is obtained when the transmission due to reflection losses is 1/e, that is, $\rho^N = 1/e$ or 0.37.[4] For reflectivities of 0.99 and 0.98 this corresponds to 100 and 50 reflections, respectively. Thus dirty or degraded mirrors with poor reflectivity lead quickly to a loss of signal and lower S/N in spectra. The other major loss is due to throughput of the cell, which is typically less than the optimum throughput of the spectrometer except at high resolution. For example, for a large 100-m path White cell the beam is typically $f/15$, and a medium-sized cell with pathlength 10–20 m it is $f/6$. A typical FTIR beam in the sample compartment is in the range $f/4$–$f/6$, so that only for smaller White cells will the entire FTIR beam be captured by the White cell mirrors. After all losses, multipass cells typically have a net transmission of 10–20%.

A different type of multipass cell, based on the Heriott cell,[13] is commonly used for tunable IR diode applications. This type of cell provides a much better pathlength to volume ratio than the White cell, but has very low throughput and is not generally suitable for FTIR applications. Recently an integrating sphere has been described for multipassing gas-phase samples in a small volume with a tunable diode laser.[14] Based on earlier measurements of cryogenically sampled gases in an integrating sphere,[15] this technique should also be applicable to gas-phase FTIR analysis, but to our knowledge has not been described in the open literature.

3.2 Open-path Systems

For many applications, such as fugitive emissions monitoring and measurements of reactive gases that are difficult to sample reliably, open path in situ measurements are more appropriate than extractive measurements. To transmit and receive an IR beam over a long path, the beam must be well collimated to minimize inverse-square-law losses due to divergence; according to Equation (5), for a given throughput, the larger the diameter of the collimated beam, the less is its divergence. Two configurations are commonly used and usually referred to as monostatic and bistatic.

In the bistatic configuration, the IR source is collimated by a transmitting telescope and directed to a receiving telescope across the atmospheric path to be analyzed. The receiving telescope collects the beam and focuses it on to an intermediate aperture from which it is then recollimated into the interferometer by the usual collimating mirror. The bistatic configuration is shown schematically in Figure 3(a). It is useful and instructive to examine the theoretical throughput and absorption-to-noise ratio of such a long open-path system. Following Equation (5) and assuming that the solid angle Θ can be approximated by $\pi\alpha^2$, the throughput of the interferometer and telescope systems are given by Equations (8) and (9), respectively:

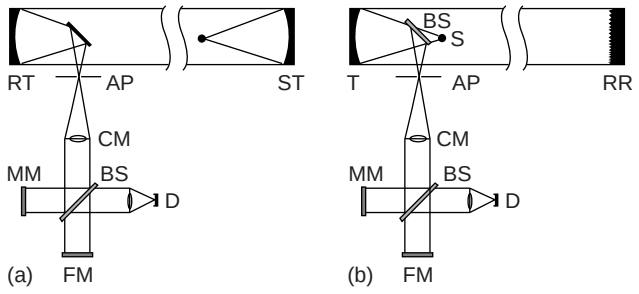

Figure 3 (a) Bistatic and (b) monostatic configurations for long open-path FTIR spectroscopy. T = telescope; ST = source telescope; RT = receiving telescope; RR = retroreflector; other labels as in Figure 1.

$$\Omega_{\mathrm{FTIR}} = \frac{\pi^2}{16}\frac{\phi_{\mathrm{AP}}^2\phi_{\mathrm{c}}^2}{FL_{\mathrm{c}}^2} \tag{8}$$

$$\Omega_{\mathrm{tele}} = \frac{\pi^2}{16}\frac{\phi_{\mathrm{tele}}^4}{d^2} \tag{9}$$

where ϕ_{AP}, ϕ_{c}, and ϕ_{tele} are the diameters of the intermediate aperture, collimator mirror and telescopes, respectively, (assuming both telescopes have the same diameter), FL_{c} is the focal length of the collimator and d is the distance between telescopes. According to Equation (6), ϕ_{AP} and hence the throughput of the FTIR system are limited by spectral resolution, while the telescope system throughput decreases as $1/d^2$ (as required by the inverse square law). As the path d is increased, up to some critical distance $d_{\max}$ the throughput of the whole system will be constant and limited by the interferometer, but beyond $d_{\max}$ it will be limited by the telescope subsystem and decrease as $1/d^2$. (Equivalently, this can be understood as the image of the transmitting telescope formed by the receiving telescope on the collimator aperture AP being larger than ϕ_{AP} for $d < d_{\max}$ so that the aperture is overfilled, and smaller than ϕ_{AP} for $d > d_{\max}$, underfilling the aperture). The ratio of absorption to noise increases linearly with d up to $d_{\max}$, as the pathlength increases but throughput remains constant. For $d > d_{\max}$, the ratio decreases as $1/d$ as the throughput decreases as $1/d^2$ while the absorption increases linearly with d.

Table 2 Critical telescope separations for maximum absorbance to noise ratio in open-path FTIR spectra (see text for details)

d_{max} (m)	Resolution (cm^{-1})		
	4.0	1.0	0.2
f/6, 150 mm/25 mm	57	114	255
f/4, 200 mm/50 mm	28	57	127

Under the assumption that the interferometer collimator and telescopes are f-matched (i.e. that both have the same f/number and the beam from the receiving telescope exactly fills the collimating mirror after the aperture), the critical distance d_{max} can be calculated from Equations (6), (8) and (9) as Equation (10):

$$d_{max} = \frac{\phi_{tele}^2}{\phi_{AP}} \frac{FL_c}{\phi_c} = \frac{\phi_{tele}^2}{2\phi_c} \sqrt{\frac{\tilde{\nu}_{max}}{\Delta\tilde{\nu}}} \qquad (10)$$

The critical distance thus depends on the resolution, wavenumber and beam diameters and most importantly increases as the square of the telescope diameter ϕ_{tele}. As an example, Table 2 shows the critical distance d_{max} for two typical interferometer configurations: 25-mm diameter beam, 150-mm focal length collimator (f/6), such as the Bomem MB100 commonly used for open-path measurements, and 50-mm beam diameter, 200-mm collimator (f/4). The table assumes a 300-mm diameter telescope system f-matched to the interferometer collimator.

In the monostatic configuration, illustrated in Figure 3(b), the source telescope is replaced by a retroreflector, normally either a corner-cube array or a cat's eye, which returns an incoming beam exactly back along the input direction. The single telescope acts as both transmitter and receiver by using a beamsplitter to fold the IR source beam into the path and transmit the returned beam to the interferometer. The monostatic configuration has the principal advantage that the remote retroreflector is passive and requires no power or fine alignment. Several retroreflectors may be positioned around a study site and thus several paths measured from a single spectrometer that is able to point at the several reflectors. In the monostatic configuration half of the signal is lost at the source beamsplitter both at transmission and reception, resulting in a throughput loss of a factor of four. There are additional losses if the whole area of the corner cube array is not retroreflective. These losses are offset by a fourfold increase in throughput compared with a bistatic experiment of equivalent pathlength because the sample is double passed, the telescope–reflector distance is halved and the signal loss through beam divergence is correspondingly decreased.

Spectral artefacts may arise in bistatic systems because the source radiation is unmodulated by the interferometer before passing through the sample path. All radiation in the field of view of the detector is modulated and contributes to the measured interferogram, not just the source radiation. In monostatic systems, stray radiation from the transmitter–receiver beamsplitter may also contribute to the measured signal and contribute photometric errors. Corrections for these effects have been discussed by several authors.[16–18]

The throughput of the telescope system ultimately limits the usefulness of such open-path measurements over very long pathlengths, and for most applications a few hundred meters is the normal limit (although measurements over paths of several kilometers have been reported[19,20]). Galle and Mellqvist[21] described an alternative configuration for pathlengths up to 1 km based on 40 passes of a triple-passed open cell with a base length of 25 m. The design, due to Ritz et al.,[10] is self-correcting for mirror instability and optical distortions that occur over such an open path. The throughput of the f/100 cell as described is around a factor of 10 better than that of a telescope-based system of equivalent beam diameter using the same interferometer for a pathlength of 1 km. The distance at which it becomes more efficient than a telescopic system is around 350 m. Compared with telescope-based systems, the open-path multipass cell thus achieves higher absorption over the longer path and is more compact, but probes only a 25-m path of air. The particular application will determine which system is more suitable.

3.3 Open-path versus Extractive Sampling

Advantages and disadvantages of open-path versus extractive sampling are summarized in Table 3.

3.4 Sampling Methods for Extractive Analysis

For extractive analysis in closed cells, the sample must be collected from the atmosphere and transferred to the cell. Sampling lines should be constructed of inert materials, such as stainless steel or Teflon®. If the line is of significant length or the gas being sampled is at higher than ambient temperature, the intake line may need to be heated to prevent condensation. The simplest method for in situ measurements is to draw the ambient air into the cell for direct measurement if target species concentrations are high enough for detection. The cell may be continually flushed or repeatedly filled and measured in a batch mode.

Water vapor can be removed from the sample during collection if desired. The main advantage is that regions of the spectrum that are otherwise obscured by the extensive water vapor spectrum become accessible. The major disadvantage is that drying the air may also remove other species, in particular polar compounds. Water vapor

Table 3 Comparison of advantages and disadvantages of open-path and extractive FTIR analysis of air

	Extractive/multipass cell	Long open path
Advantages	Sample may be dried or otherwise pretreated, reduces interference Preconcentration possible, improves detection limits Pressure and temperature control improve accuracy and precision Calibration spectra possible in cell	Noninvasive sampling, suited to reactive, labile or polar compounds Pseudo-continuous monitoring Line integrating Long paths to 1 km Multiple sample paths from single spectrometer
Disadvantages	Reactive, labile, polar compounds difficult to sample Time delays for sample treatment and transfer Detection limits sensitive to mirror degradation	Low S/N for long paths No temperature or pressure control Cannot run calibration spectra over same sample path Weather-proofing of optics and spectrometer required

can be removed from the sample by use of semipermeable membranes (see, for example, United States Environmental Protection Agency (USEPA) Method TO-14A[22]), cryogenic traps or chemical drying agents such as Drierite, magnesium perchlorate or phosphorus pentoxide.

If the spectrometer is remote from the sample site, the air must be collected and returned to the laboratory for analysis. Established methods exist for the collection of air samples in canisters in USEPA Methods TO-14A (nonpolar volatile organic compounds (VOCs))[22] and TO-15 (polar VOCs).[23] While these methods address analysis by gas chromatography (GC) and mass-selective detection, the sampling protocol is also applicable to FTIR analysis. Treated stainless-steel canisters or absorption traps are the most commonly used methods to collect and store the samples. ''SUMMA''-type canisters have a single valve and a passivated and electropolished or coated inner surface. (SUMMA is a brand name of a particular manufacturer, but it is also commonly used as a generic name for this general type of canister. USEPA Methods TO-14A and TO-15 list a number of suppliers of suitable canisters.) This reduces the potential for loss of sample through adsorption on the canister walls. Preparation of canisters for sampling involves multiple evacuations and filling with humidified zero air, which is used as the water competes with polar compounds for adsorption sites on the canister and thus reduces loses of trace compounds through this mechanism. The canister should be evacuated for the final time only immediately before sample collection. USEPA Methods TO-14A and TO-15 give procedures for ensuring high-quality canister preparation.

The canister can be passively filled to ambient pressure through natural pressure differential or actively filled (and pressurized above ambient) using a clean, leak-free pump. Under pressurization there is potential for condensation of liquid water, especially with samples taken under conditions of high humidity, which may lead to the loss of water-soluble components from the gas phase. It is also possible to increase the sample pressure with zero air or pure nitrogen at a later stage. This lowers the mixing ratios of the compounds but increases the sample volume, thus allowing multiple measurements.

There have been a number of studies on the long-term storage capabilities of SUMMA-type canisters. Brymer et al.[24] determined that 168 of 194 (86%) of compounds tested could be stored in canisters for up to 30 days without detectable losses. Evans et al.[25] showed that for 31 compounds it was possible to recover 80% or more after 32 weeks of storage. Kelly and Holdren[26] suggested that polar VOCs present more problems, but that canisters are adequate for at least 4 days of storage when the sample is humidified. Both high mixing ratios and high absolute pressure enhanced stability.

Alternative sampling containers include rigid glass canisters. These may have two valves, which allow the canister to be flushed thoroughly with the sample gas before sealing, and the pump may be placed downstream of the container, reducing the risk of contamination. Another method is the use of inflatable bags of inert materials such as Teflon®, Tedlar® or Mylar; the material must be chosen to be both inert to the target gases and not to contribute impurities to the spectra. Sampling bags may be packed into a much smaller volume than rigid canisters, and are always at ambient pressure. They must be inflated with a pump, a potential source of contamination.

An alternative to canister-based collection is the use of adsorption traps. Canister and sorbent-trap sample strategies were compared by Castellnou et al.,[27] Evans et al.[25] and Camel and Caude[28] Such traps are commonly used for collecting samples for GC analysis, and their use is covered extensively in the GC literature (see Camel and Caude[28] for example). For closed-path FTIR analysis the disadvantage of adsorption traps is that the recovered gas volume is generally low, but may be suitable for particular gases. Sorbent-trap methods may be useful in combination

with a preconcentration strategy. By using adsorbents that are selective, it may be possible to increase the mixing ratio of some components. For instance, by using hydrophobic adsorbents, water vapor may be rejected.[29]

Consideration must be given to the volume of the sample container relative to the volume of the cell and sampling manifold. SUMMA-type canisters are available in sizes up to at least 15 L, while cell and manifold volumes are typically 10 L or less. If the sample and cell/manifold volumes are approximately equal, then the maximum cell pressure possible would be half of that originally in the sample container, if the sample is transferred by pressure equalization. This may be too low to achieve the desired detection limits. It is necessary then to have a significantly larger sample volume over cell/manifold volume, or incorporate a method to transfer more of the sample from the container to the cell. This may include the use of a mechanical pump or a cryogenic condenser. Hanst and Hanst[30] described a system where the manifold is comprised of relatively large-volume but narrow-bore tubing. After pressure equilibration between the sample container and the cell, the sample container is opened to the atmosphere. This then pushes the sample gas through the manifold and into the cell. Because of the narrow bore there is no mixing between the sample and the ambient atmosphere. The result is that the cell is filled to ambient pressure with sample gas only, while the ambient atmosphere progresses no further than the manifold. It is also possible to transfer nearly the complete volume through the intermediate step of cryogenic condensation of the sample from the sample container, then vaporization into the cell.

3.5 Detection Limits

Absorption in IR spectroscopy is ideally described by Beer's law. Radiation intensity decreases exponentially with the concentration and pathlength through an absorber (Equation 11):

$$I(\tilde{\nu}) = I_0(\tilde{\nu})\mathrm{e}^{-\alpha'(\tilde{\nu})cl} \tag{11}$$

or equivalently the absorbance $A(\tilde{\nu})$ increases linearly with concentration and pathlength (Equation 12):

$$A(\tilde{\nu}) = \log_{10}\frac{I_0(\tilde{\nu})}{I(\tilde{\nu})} = \alpha(\tilde{\nu})cl \tag{12}$$

where $I_0(\tilde{\nu})$ and $I(\tilde{\nu})$ are respectively the radiation intensity at frequency $\tilde{\nu}$ before and after traversing the sample, $\alpha(\tilde{\nu})$ is the absorptivity of the absorber (also called the absorption coefficient), c is the absorber concentration and l the optical pathlength through the sample. (All logarithms are assumed here to be decadic. Thus $\alpha'(\nu)$ in Equation 11 is 2.303 times $\alpha(\nu)$ in Equation 12.)

Quantitative analysis should be performed on absorbance spectra because in the Beer's law approximation absorbance is linear in absorber amount. The minimum detectable amount (i.e. concentration–pathlength product) of an analyte is given by Equation (13):

$$[cl]_{\min} = \frac{A_{\min}(\tilde{\nu})}{\alpha(\tilde{\nu})} \tag{13}$$

where $A_{\min}(\tilde{\nu})$ is the minimum detectable absorbance, assumed here to be equal to the peak-to-peak (p–p) noise level in the spectrum (For normally distributed noise, the p–p noise level is typically 4–5 times the root mean square (RMS) noise level; the single-beam S/N is equal to 1/(2.303 × noise in absorbance)). At this level of absorption the feature would normally be identifiable by eye, and when using full spectrum fitting techniques such as classical least squares (CLS) or partial least squares (PLS) (see section 5) for quantitative analysis a quantitative determination should be possible. Other authors variously use two or three times the spectral noise level; the actual choice should be clearly stated. The p–p noise level in absorbance spectra will typically be in the range 0.01 (long open path, DTGS detector, poor throughput) to 0.0001 (good throughput, closed cell, MCT or InSb detector).

For the best *concentration* detection limit, the pathlength, absorptivity and S/N should therefore be as large as possible. The pathlength will be determined by the experiment as discussed earlier. The true absorptivity is a molecular property and clearly the strongest absorption bands of the analyte should be used if possible. However, for sharp absorption lines the apparent absorptivity will decrease at low spectral resolution, as discussed below. The most important determining factor for detection limits then becomes the spectral S/N. Interference from other absorbing species, in particular water vapor and carbon dioxide (section 4), may reduce detection limits below the S/N-limited values.

We can distinguish two limiting cases, depending on whether the spectral bands are much wider or much narrower than the instrument spectral resolution.

(a) Spectral bands broader than instrument resolution. This is the simpler case and applicable for larger molecules where the fine structure of the absorption bands (due to molecular rotation) is not spectrally resolved. Such bands may be 10–20 cm^{-1} wide and will be well resolved with 4–8 cm^{-1} resolution or better. In this case the observed peak heights will be independent of resolution and determined only by the absorptivity and absorber amount. The lowest resolution that still resolves the broad bands should be selected since this will maximize S/N (Equation 7). Absorptivities are normally quoted in units of $\mathrm{atm}^{-1}\,\mathrm{cm}^{-1}$ or $\mathrm{cm}^2\,\mathrm{molecule}^{-1}$. For 1-atm pressure at 296 K, 1 atm cm

Table 4 Dependence of detection limit on absorptivity and spectral noise level

Absorptivity		Detection limit (ppmv m)	
$atm^{-1}\,cm^{-1}$	$ppmv^{-1}\,m^{-1}$	Abs. noise = 0.01	Abs. noise = 0.0001
1	10^{-4}	100	1
100	10^{-2}	1	0.01

is equivalent to 2.5×10^{19} molecules cm^{-2} or 10^4 ppmv m. Typical peak absorptivities for medium to large molecules are in the range $1–100\,atm^{-1}\,cm^{-1}$, corresponding to $10^{-4}–10^{-2}\,ppmv^{-1}\,m^{-1}$. For these extreme values of α and typical noise levels of 0.01–0.0001 in absorbance, Table 4 shows the corresponding detection limits in ppmv m; for example, a detection limit of 1 ppmv m corresponds to 1 ppmv over a 1-m path, 0.01 ppmv over a 100-m path, and so on. Over a 100-m path, detection limits typically lie in the range 1 ppmv–50 pptv.

(b) Spectral bands narrower than instrument resolution. This will be the case for many smaller molecules with resolved rotational structure, for which pressure-broadened linewidths at 1 atm pressure are typically $0.05–0.15\,cm^{-1}$. Unless high resolution ($\Delta\tilde{\nu} < 0.1\,cm^{-1}$) is used, the measured peak absorptivities will be approximately inversely proportional to the resolution $\Delta\tilde{\nu}$. Peak heights will increase with resolution (inversely with $\Delta\tilde{\nu}$), but S/N will decrease in proportion to either $\Delta\tilde{\nu}$ or $\Delta\tilde{\nu}^2$ depending on whether or not throughput is reduced by reducing the spectrometer aperture (section 2.5). In the former case there is no penalty

Table 5 Examples of absorptivities and detection limits for a range of gaseous and volatile compounds in air

Compound	Peak position (cm^{-1})	α_{peak} $(ppmv\,m)^{-1} \times 10^3$	Peak fwhm (cm^{-1})	Minimum detection limit[a]
1,3-Butadiene	908	1.1	1.1	0.88
2-Butanone (methyl ethyl ketone)	1175	0.31	36	3.3
4-Methyl-2-pentanone (methyl isobutyl ketone)	1172	0.32	27	3.1
Acetaldehyde (ethanal)	2729	0.17	158	5.9
Acetonitrile (ethanonitrile)	1042	0.05	0.4	19
Acrolein (2-propen-1-one)	959	0.27	2.6	3.6
Acrylonitrile (propenonitrile)	954	0.39	1.3	2.6
Ammonia	967	1.3	3.4	0.76
Benzene	674	8.9	0.5	0.11
Carbon dioxide	2363	3.7	0.6	0.27
Carbon monoxide	2169	0.43	0.8	2.4
Carbon tetrachloride	795	5.4	7.9	0.18
Chlorobenzene	741	1.3	1.0	0.79
Chlorodifluoromethane (Freon 22)	809	1.2	22	0.85
Chloroform (trichloromethane)	773	3.2	14	0.31
Cyclohexane	2933	2.9	5.8	0.34
Dichloromethane	749	0.91	22	1.1
Diethyl ether (ether)	1143	1.4	27	0.70
Ethane	2980	0.58	0.7	1.7
Ethylene oxide	3066	1.4	0.8	0.70
Methanol (methyl alcohol)	1033	1.0	2.1	1.0
m-Xylene	769	0.62	3.2	1.6
n-Hexane	2965	1.0	64	0.96
Nitric oxide	1894	0.2	0.7	5.6
Nitrogen dioxide	1603	1.6	26	0.64
Nitrous oxide	2210	1.4	30	0.72
o-Xylene	741	2.2	1.0	0.45
Ozone	1056	0.43	15	2.3
Phosgene	857	3.4	29	0.29
p-Xylene	796	0.45	2.5	2.2
Sulfur dioxide	1361	1.0	41	1.0
Sulfur hexafluoride	948	42.0	1.0	0.02
Tetrachlorethylene	916	1.6	18	0.64
Tetrafluoromethane	1283	82	0.6	0.01
Toluene	729	1.1	2.4	0.95

[a] Minimum detection limit in ppbv, 100-m path, 10^{-4} absorbance noise (base 10).
fwhm, full width at half-maximum.

in detection limits for higher resolution since the peak heights and noise level increase together, and the higher resolution may help in identifying and separating spectral lines. In the latter case, however, higher resolution will lead to lower S/N and consequently lower detection limits. The lowest resolution acceptable for analysis of the resulting spectra with acceptable discrimination between absorbers should then be chosen.

Strong individual lines of small molecules, for example, CO_2 in the ν_3 band near 2350 cm^{-1}, may have true peak absorptivities up to 500 $atm^{-1}\,cm^{-1}$ at 1 atm pressure and high resolution. The apparent absorptivities will be reduced if the resolution is lower than the intrinsic width of the pressure-broadened lines, 0.05–0.15 cm^{-1}. Since both individual coefficients and linewidths vary widely, it is more difficult to generalize on detection limits in these cases.

Collections of quantitative reference spectra are available both commercially and in the public domain; details are given in the Appendix. Once absorptivities and experimental parameters such as resolution and noise level are known, the principles of this section may be used to estimate detection limits. Interference from atmospheric water vapor and carbon dioxide absorption may affect detection limits for some bands.[31] Table 5 provides examples of gaseous and volatile compounds which may be observed in a polluted atmosphere or plume, together with apparent peak absorptivities at 0.5 cm^{-1} resolution and corresponding detection limits assuming a 100-m path and noise level of 0.0001 in the absorbance spectrum. The data in Table 5 are taken from the Hanst library (IR Analysis, Anaheim CA, USA).[30] For features of smaller molecules where the listed peak width appears as 0.5 cm^{-1}, the peak absorptivity can be expected to increase at higher resolution since the apparent peak width is limited by 0.5 cm^{-1} resolution.

3.6 Choice of Spectral Resolution

From the above, it is clear that the selection of spectral resolution for a measurement is a critical choice. While the spectroscopist would argue for high resolution to resolve all bands, the analytical chemist would argue for the lowest acceptable resolution to maximize S/N and quantitative precision. In practice, a balance must be struck to suit the demands of the individual measurement, and this requires an informed judgement by the operator. To some extent resolution will, of course, be limited by available hardware. Low-resolution spectrometers have the advantage of being cheaper, smaller and in principle more robust since the scanning mechanism need not maintain alignment over long scan lengths. Throughput is higher, leading to better detection limits. The main limitation will be for conditions where higher resolution is required to discriminate spectral features that cannot be separated by the quantitative analysis methods employed. For example, peak height- or peak area-based methods require that the band being analyzed be reasonably well resolved from interfering absorptions with a well-defined baseline, whereas full spectrum methods such as CLS and PLS (section 5) can often discriminate between absorbers even when their bands overlap.

Several authors have discussed the choice of resolution in long open-path monitoring when using CLS analysis (section 5.4).[32–36] While there is no definite 'best' resolution, there is a general consensus towards 1–2 cm^{-1} as a reasonable compromise between spectral discrimination of overlapping bands and quantitative accuracy. Griffiths and co-workers[37,38] make the case that lower resolution (8 cm^{-1}) may be practical when using PLS (section 5.5) for quantitative analysis.

4 ATMOSPHERIC INFRARED SPECTRA

All molecules except homonuclear diatomics exhibit IR spectra due to the vibrations and rotations of the molecules. Thus we are fortunate that the major fixed components of air, N_2 (78% of dry air), O_2 (21%) and Ar (1%), are transparent to IR radiation, and atmospheric absorption in clean air over pathlengths up to a few hundred meters is dominated by absorption of minor and trace gases, principally water vapor (0–4% by volume depending on humidity) and carbon dioxide (360 ppmv). Figure 4 shows a typical atmospheric transmission spectrum over a 100-m path in clean air with the principal absorption bands shown. The most useful windows for detection of additional absorbers are between the water vapor and carbon

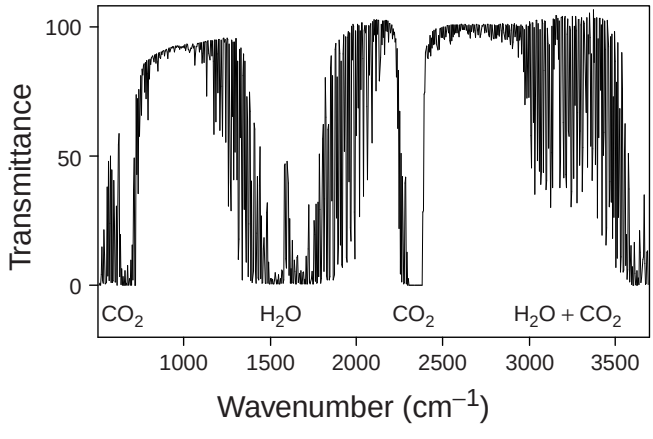

Figure 4 Typical mid-IR absorption spectrum of 100-m path of clean air at 1 atm, 25 °C. Absorption bands due to H_2O and CO_2 are marked.

dioxide absorptions in the approximate ranges 800–1200, 2000–2300 and 2400–3000 cm^{-1}, of which the first corresponds largely to the so-called "fingerprint region" in which most molecules have characteristic absorption bands.

Each absorption band in the IR spectrum corresponds to a particular molecular vibration, with shape or fine structure determined by concomitant rotational transitions. Individual vibration–rotation transitions are broadened by two mechanisms:

- Pressure broadening, due to the perturbation of molecular energy levels by collisions between molecules. This broadening varies from molecule to molecule and line by line and is proportional to pressure, has a Lorentzian lineshape, and typical linewidths (fwhm) of 0.05–0.15 $cm^{-1}\,atm^{-1}$.
- Doppler broadening, due to the random distribution of molecular velocities around the mean velocity. Doppler broadening increases with the square root of temperature and decreases as the square root of molecular mass. For small to medium-sized molecules at ambient temperatures, Doppler linewidths (fwhm) are typically 0.002–0.004 cm^{-1}.

The general, combined lineshape is called the Voigt lineshape. Under normal ambient atmospheric conditions, collision or pressure broadening dominates and individual absorption lines generally have a near-Lorentzian lineshape and linewidths of the order of 0.1 cm^{-1}. In smaller molecules the individual vibration–rotation transitions are usually well separated compared with their linewidths. In larger molecules the rotational lines become more closely spaced, as to a first approximation the spacing scales inversely with the mass and the square of the molecular size. Typically, in molecules with molecular masses much over 100, the rotational lines are more closely spaced than their linewidths and merge together into a broad band profile for each vibrational transition. Figure 5(a–c) illustrates typical vibration–rotation band spectra of molecules which show the gradation from resolved rotational structure in carbon monoxide (CO) through partially merged rotational lines in nitrous oxide (N_2O) to a smooth profile in toluene ($C_6H_5CH_3$).

The actual appearance of measured spectra depends in addition on the spectral resolution. If the resolution is three or more times higher (i.e. $\Delta\tilde{\nu}$ lower) than the intrinsic linewidths, the true spectra will be faithfully reproduced, but this may come at a significant cost in S/N (section 2.5) and complexity in the experimental configuration. At lower resolution ($\Delta\tilde{\nu}$ greater than the intrinsic linewidths), the detailed band structure may be wholly or partially lost so that only a broad band profile is observed. This may lead to a loss of selectivity between compounds but with a gain in S/N at the lower resolution. The ideal resolution for a particular analysis requires a considered judgement by the experimenter.

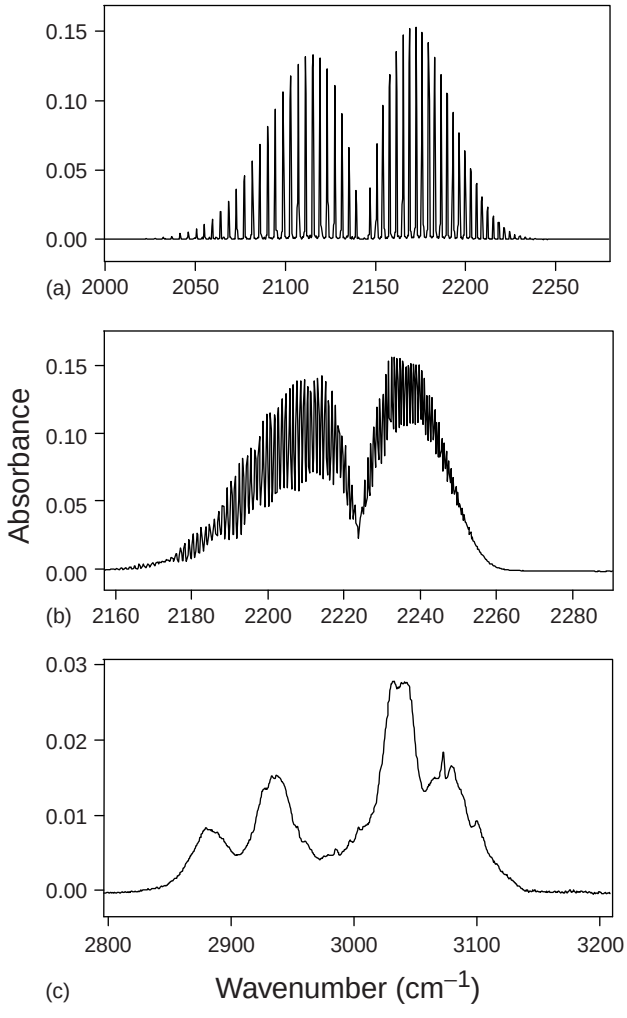

Figure 5 Examples of vibration–rotation bands in gas-phase spectra: (a) CO, resolved rotational structure; (b) N_2O, partially resolved; (c) toluene, unresolved.

5 CALIBRATION AND QUANTITATIVE ANALYSIS

Several methods present themselves for calibration and quantitative analysis of measured spectra, and the choice of the best method is an integral part of the analysis strategy. Most methods are based on Beer's law (Equation 12), in which the absorbance is proportional to the amount of absorber present in the beam. Deviations from Beer's law are fairly common, however, especially at low resolution, and are discussed

in section 5.7. In increasing order of complexity, we briefly review here peak height/area measurements, spectral subtraction, CLS, factor analysis methods such as PLS and principal components regression (PCR) and nonlinear least squares. Haaland[39] provides a detailed overview of calibration methods with references to further details of each individual method.

5.1 Measurement of Background Spectra

In all but nonlinear least squares, Beer's law requires that absorbance spectra be used, which in turn requires that a background (I_0) spectrum be recorded in addition to the sample spectrum. Closed-cell measurements present no great difficulty in recording a background spectrum of either the evacuated cell or the sample cell flushed with clean background air or a nonabsorber such as N_2. Open-path measurements are more difficult since the sampled volume cannot be so easily manipulated. Possible methods for background spectrum measurement include the following:[18,38]

1. Synthetic background spectrum. By selecting points in a sample spectrum estimated to have negligible absorption, a smooth curve through these points can be constructed. Most spectrometer software will allow this procedure but it requires user skill and judgement and has many potential pitfalls. It should be used only when the collection of real background spectra is impractical, and preferably only over narrow frequency intervals.
2. Upwind background spectra. Collect spectra upwind of the path to be investigated so that none of the target gases appear in the background spectrum. This requires an experimental set-up in which the optical path can easily be redirected, e.g. by having two retroreflectors in a monostatic open-path configuration.
3. Short-path spectra. Record background spectra over a very short path relative to the path for sample measurements by bringing the source or retroreflector close to the spectrometer. Care must be taken not to saturate detectors with this method since there will be much greater throughput for the short path.

None of these methods is ideal in that none provides a background spectrum identical to the sample spectrum except for the lack of sample. The best choice will depend on the experimental configuration, measurement site and conditions and available hardware.

5.2 Peak Height or Area Analysis

The simplest and most traditional quantitative analysis method is to measure either peak height at a particular frequency or integrated peak area over a restricted frequency range in an absorbance spectrum. The amount of absorber cl is calculated directly from the measured absorbance (or integrated absorbance) and the known absorptivity assuming Beer's law, $cl = A/\alpha$. The peak or integrated absorptivity is determined by an equivalent analysis of a set of calibration spectra of known amounts; the slope of a Beer's law plot of absorbance versus amount yields the absorptivity. If the Beer's law plot is nonlinear it may still be used as a calibration curve. Peak area is in principle better than peak height because the absorbance *integrated* over a peak is to a good approximation independent of the resolution and intrinsic lineshape provided the absorbance is not too large. These peak-based methods, although quick and relatively simple, have potential drawbacks and should be used with caution or only for initial estimates of concentrations. Both methods require that the peak be reasonably isolated from other spectral features so that a baseline from which the height or area is measured can be estimated; this requirement is particularly stringent for peak areas. Software provided with commercial FTIR instruments normally has the ability to perform quantitative analysis based on peak area and height.

5.3 Interactive Subtraction

If Beer's law is obeyed, the absorbance spectrum of a mixture of gases is simply the sum of the spectra of the individual gases weighted according to their respective amounts. At any wavenumber the absorbance is given by Equation (14):

$$A_{\tilde{\nu}} = \alpha_{1,\tilde{\nu}}(cl)_1 + \alpha_{2,\tilde{\nu}}(cl)_2 + \alpha_{3,\tilde{\nu}}(cl)_3 + \cdots \qquad (14)$$

where the subscripts $1, 2, 3, \ldots$ refer to the individual compounds contributing to the spectrum at frequency $\tilde{\nu}$. If reference absorbance spectra of the individual compounds are available, they can be sequentially subtracted from the total spectrum until in principle only noise remains in the residual spectrum. Most spectrometer software provides the facility for interactive subtraction of a scaled reference spectrum from a measured spectrum. The scaling factor is interactively adjusted until the residual spectrum appears by eye not to contain any contribution from the reference spectrum. The scaling factor then provides the amount of reference compound in the measured spectrum. Clearly a region of the spectrum should be chosen for this process where the absorption features of the reference compound dominate all others in the spectrum. This method is intuitive and instructive for gaining a good "feel" for the nature of measured spectra. It also uses more of the measured spectral data over a wider range of frequencies than peak-based methods, which adds to the information content of the method.

However, it requires a degree of practice and skill for consistent quantitative results, which are somewhat subjective since the ultimate criterion used to decide the scaling factor is a judgement by eye. Care should be taken to match the resolution, apodization, lineshape and spectral shifts of sample and reference spectra or a good fit will never be achieved.

CLS methods are based on the same premise (Equation 14) but use a more rigorous criterion for goodness of fit and are described next.

5.4 Classical Least Squares

CLS is the simplest of the multivariate least squares techniques and perhaps the most commonly used method for quantitative analysis of atmospheric spectra. CLS is also based on Beer's law in the form of Equation (14), which can be written at each frequency $\tilde{\nu}$ over a suitable spectral range for analysis of the target gases to provide a set of equations linear in the unknown amounts $(cl)_i$. Provided there are more frequencies in the selected spectral region than unknown components (i.e. more equations than unknowns), the set of equations is overdetermined and the amounts can be determined by standard methods of least squares regression. In essence, a linear combination of single-component reference spectra of the individual compounds is fitted to the measured spectrum such that the sum of the squares of the residuals at each frequency is minimized. The amounts of each absorber required in the best fit provide the required quantitative result. The method is similar to spectral subtraction except that the rigorous least squares criterion is used to determine the best fit. The equations for the regression can be readily written and solved in matrix form so that the method is computationally fast and efficient. The matrix of absorptivities is often called the K-matrix and CLS regression equivalently called the K-matrix method.

Calibration in CLS is the determination of the basis set of single component spectra to be used in fitting the unknown spectra. Ideally the reference spectra should be measured with the same spectrometer under the same conditions as the unknown measurement so that the environmental (e.g. temperature, pressure) and instrumental (e.g. instrument lineshape, background, wavenumber shifts) effects on spectra are the same in the reference as in the sample spectra. The spectra can be measured directly as single pure component spectra or as mixtures, in which case a CLS calibration step is used to determine (again by least squares regression) the best single-component spectra which fit the ''training set'' of calibration spectra. Such calibration may be very laborious and time-consuming, especially if many compounds are involved; in practice, such calibration is seldom undertaken in the field. Reference spectra are more commonly generated in the laboratory and stored, or taken from publicly or commercially available libraries (see Appendix). For open-path measurements, the reference calibration spectra cannot be recorded in an equivalent experimental configuration, and cell or library spectra must be used. As for all subtraction methods, great care must be taken to match the resolution, apodization, lineshape and spectral shifts of sample and reference spectra or poor fits with large residuals will result.[5,40] An alternative approach to generation of reference spectra is to calculate the spectra, as described in section 5.6.

CLS has the advantage that it is conceptually and computationally simple, and is often well suited to gas-phase analysis. It is in principle more accurate than peak height or area measurements because it uses the information contained in an extended range of frequencies in the quantitative analysis, rather than just a few over a single peak. However, there are two major drawbacks. First, CLS requires that *all* compounds absorbing in the analyzed spectral region be included in the CLS reference set. If not, the residual spectrum will still contain the spectrum of the missing components and the method may determine incorrect amounts of other gases in a false attempt to minimize the total spectral residual. This sort of error is readily identified by observing the full spectrum residual after the CLS fit – if all absorbers have been accounted for, and the residual should consist of only noise. If reference and sample spectra have different instrumental lineshapes and shifts, these will also be apparent in the residuals. The second major drawback is that CLS relies implicitly on Beer's law, since the fitted spectrum is constructed as a simple linear combination of reference absorbance spectra. Breakdown of Beer's law is discussed in section 5.7; for gas-phase spectra the problems are most severe for medium-to-strong absorption in low-resolution spectra of molecules with resolved rotational structure. In this case care must be taken to minimize the impact of Beer's law nonlinearity.

The USEPA currently recommends CLS as the preferred method for analysis of open-path spectra, as set out in detail in the Compendium Method TO-16, ''Long-path open-path FTIR monitoring of atmospheric gases''.[41] This document sets out explicit methods and recommendations for most aspects of CLS analysis including window determination, avoidance of interferences and minimization of the effects of Beer's law breakdown. Software for general-purpose CLS is available in many commercial spectrometer packages.

5.5 Partial Least Squares

PLS and its relation PCR are, like CLS, full spectrum methods which construct a fit to a measured

spectrum from a linear combination of basis spectra. In PLS and PCR, however, the basis spectra are not pure single-component spectra. The first step of a PLS or PCR calibration is to create a training set of mixture spectra which should span the range of natural variation to be expected in the samples to be analyzed. A factor analysis is performed on the training set to find a set of factors or ''pseudo spectra'' which form the basis set and are sufficient to explain all significant variance in the training set, that is, each training spectrum (and, during prediction, each sample spectrum) can be reconstructed as a linear combination of the factor spectra weighted by a score for each factor. PLS and PCR differ only in the method by which the factor analysis is performed. In the first instance, it is not necessary to know the complete composition of the training spectra. For those components for which concentrations in the training spectra are known, the concentrations are regressed against the factor scores to link the factor analysis to the known concentrations in the calibration spectra. For analysis of an unknown sample, first the unknown spectrum is fitted as a linear combination of the factors, much the same as CLS where the fitting spectra are the pure single- component spectra. Second, the amount of each component known to the calibration is calculated from the fitted factors and the scores in a second linear regression.

PLS also assumes Beer's law to hold, but to some extent nonlinearities due to Beer's law breakdown will be modeled during the factor analysis and accounted for as extra factors in the basis set. These non-Beer's law components include baseline variations, nonlinearities and the existence of complexes or even reaction products between the individual compounds. For analysis of multicomponent mixtures, PLS thus has some significant advantages over CLS in the general case, but if the limitations of CLS are recognized and can be taken into account as described above and in section 5.7, CLS is simpler and provides equivalent results which are easier to interpret.

Johansen et al.[42] provided a good example of PLS use for open-path monitoring of air quality in a spacecraft capsule, building a calibration model for 23 separate gases, including corrections for baseline drift and mirror degradation. The calibration training spectra are built from single-component spectra, digitally combined into mixture spectra and baseline variations are built into the model. High absorbance regions are avoided to minimize Beer's law limitations (section 5.7). More recently, Griffiths and co-workers[37,38,43] proposed PLS methods for general open-path monitoring. A large database of real atmospheric spectra collected over a range of pathlengths, temperatures, pressures, spectrometer variations and background concentrations of water vapor and carbon dioxide is used as a set of background spectra to which known amounts of reference spectra of target compounds from the USEPA library (see Appendix) are added digitally to create the calibration training set. Good results have been demonstrated using wide spectral windows at relatively low resolution ($8\,cm^{-1}$), which allows the use of smaller, cheaper and more compact FTIR spectrometers employing only pyroelectric detectors. Genetic algorithms have been used in optimizing window selection[44] and neural network techniques in compound identification.[43] These PLS methods provide an alternative approach to CLS methods as embodied for example in the USEPA TO-16 document.[41]

5.6 Calculation of Synthetic Calibration Spectra

An alternative method of acquiring calibration spectra is to calculate them. Griffith[45] described a method in which synthetic calibration spectra are calculated from a database of absorption line parameters such as HITRAN.[46,47] The program MALT (Multiple Atmospheric Layer Transmission) first takes the standard absorption line parameters, corrects and scales them to the concentrations, pathlengths, temperature and pressure assumed for the calculated spectrum, and convolves them with the pressure and Doppler lineshape contributions to provide the true monochromatic atmospheric transmission spectrum. The monochromatic spectrum is then convolved with an instrument lineshape function including the effects of finite resolution, field of view divergence, apodization and spectral shift. In later improvements (to be published), effects of phase error, spectrometer misalignment and other spectrometer imperfections can be included. The calculated spectra then closely model true measured spectra on a real spectrometer. The program can be used to calculate whole sets of calibration spectra for any environmental and instrument operating conditions quickly and accurately, and the calibration spectra used, for example, as a training set for CLS or PLS analysis.

The synthetic spectrum approach is fast, reproducible and convenient and when combined with CLS analysis as described below (section 6.3) provides precision (reproducibility) of much better than 1% and absolute accuracy of 1–5% without any need for calibration gases. It is a particularly useful option for open-path spectra where true calibration or reference spectra cannot easily be measured, but is also a practical basis for cell-based gas analysis. Calculating reference spectra is the only practical option for nonlinear least squares (NLLS) analysis and in particular the analysis of solar absorption spectra (section 6.7). The major drawback is the range of species for which the method is applicable–line parameters must be available. The HITRAN database is publicly available and covers 37 molecules

listed in the Appendix.[47] Other databases have also been developed.[48,49] It is also possible to combine calculated and library spectra in a single CLS or PLS calibration training set.

5.7 Limitations to Beer's Law

Beer's law holds strictly only at infinite resolution, that is, when the spectrometer does not distort the true spectral lineshapes. The most common way in which Beer's law breaks down is when the true linewidths of absorption features are comparable to or narrower than the spectral resolution. In this case the absorbance is linear with absorber amount at low absorbance but increasingly falls off at higher absorbance. Anderson and Griffiths[50] and Zhu and Griffiths[51,52] give a detailed account of this behavior. This can be a serious problem for small molecules with resolved rotational structure in their spectra. Typical linewidths of individual lines at 1-atm pressure are 0.05–0.15 cm^{-1}, which would require spectra at high resolution to avoid such Beer's law nonlinearities. Since many atmospheric measurements are made with compact spectrometers with resolution typically 0.5–4 cm^{-1}, this form of Beer's law breakdown can often not be neglected. Haaland[53] discusses several means to minimize the effect of Beer's law nonlinearity on CLS analysis; these are illustrated in Figure 6 and may be summarized as listed below.

1. Calibrate only over a narrow concentration range about the concentration range of interest to avoid excessive curvature in the Beer's law (absorbance vs absorber amount) calibration curve.
2. Allow for an intercept in Beer's law, i.e. assume that $A = \alpha_0 + \alpha_1 cl$ in the calibration model (line (c) in Figure 6). This allows a curved section of the Beer's law calibration curve to be better approximated by a straight line since the straight line is not constrained to pass through the origin.
3. Allow for quadratic terms in the Beer's law model, i.e. $A = \alpha_0 + \alpha_1 cl + \alpha_2 (cl)^2$. This model allows an approximation to the curvature in the true A vs cl curve but is still linear in the coefficients α and the formalism of CLS can still be applied.

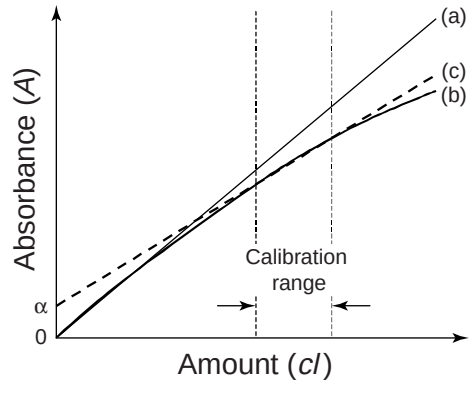

Figure 6 Illustration of deviations from Beer's law. (a) Ideal; (b) typical deviation at low-resolution; (c) linear approximation to (b) with intercept α. The dotted lines illustrate the desired calibration range.

Beer's law also breaks down if components in the sample interact and perturb the individual pure-component spectra and/or add spectra of some interaction complex, so that the spectrum of the mixture is not a linear sum of the pure-component spectra. This is not a serious problem in gas-phase spectra of the ambient atmosphere, but may be in concentrated samples such as stack effluent or combustion products. PLS would be better used in these cases.

5.8 Nonlinear Least Squares Analysis

NLLS analysis provides a different approach to quantitative analysis which does not rely on the implied linearity of Beer's law. NLLS constructs a fit to a measured spectrum by calculating spectra iteratively until the spectral residual between calculated and measured spectrum converges to an acceptable minimum.[54,55] An initial guess is made about the parameters describing the unknown spectrum and the spectrum is calculated with a program such as MALT described above. The two spectra are compared, the calculation parameters refined, the spectra recalculated and recompared, and the process repeated until satisfactory agreement between measured and calculated spectra is obtained. The search method for the best fit is usually based on the commonly used Levenberg–Marquardt algorithm.[56]

This method is used almost exclusively in the analysis of solar absorption spectra through the atmosphere, since the inhomogeneity and variability of the atmosphere with altitude and time make the use of any sort of measured reference spectra impossible for quantitative analysis. Some examples are described in section 6.7. NLLS could be readily adapted to long open-path or closed-cell measurements, and would have particular advantages in cases where Beer's law is not obeyed since it makes no assumptions about linearity between concentrations and measured absorbance. One drawback with this method is that it is computationally intensive since the spectrum must be recalculated on each iteration of the fit. However, with the power of modern desktop computers, calculation times have

been reduced to a few seconds per spectrum and this is no longer a serious limitation. To our knowledge there is currently no ready-written software commercially available for NLLS analysis of gas phase spectra.

6 SELECTED APPLICATIONS

In this section we describe a selection of applications of atmospheric analysis by FTIR spectrometry. The coverage here is not intended to be a thorough review, rather it is a selection which illustrates various applications of the technique. More detailed reviews are quoted in the text, and the biannual reviews of IR spectrometry in *Analytical Chemistry* provide a regular, thorough review of work published. The two most recent reviews are by McKelvy and co-workers.[57,58]

6.1 Early History – Measurements of Urban Pollution Chemistry by Fourier Transform Infrared Spectrometry Using Long-path Cells

Long-path IR studies of atmospheric chemistry were pioneered by Stephens and co-workers in the 1950s.[3,4,59,60] Using long-path cells of up to 400-m pathlength and dispersive spectrometers, these studies achieved sensitivities in the ppmv range, and laid much of the important groundwork for the technique. This work concentrated on the chemistry of urban photochemical pollution and formation of photo-oxidants such as ozone, and provided the first identification of PAN,[3] one of the most important secondary pollutants in urban smog.

The sensitivity of these measurements was limited, however, by the use of dispersive spectrometry, and was insufficient for routine measurements except in the most polluted ambient air. The advent of FTIR spectrometers in the 1970s increased the utility and potential of the method enormously. Throughout the 1970s and 1980s, Hanst, Pitts and co-workers, working at the USEPA and the Statewide Air Pollution Research Center in Riverside, California, developed the technique further with commercial FTIR spectrometers and multipass cells with pathlengths of >1 km.[11,61–66] This work also focused on urban pollution and photochemical smog in the Los Angeles basin, and with FTIR spectrometry was able to detect and quantify pollutant species at ppbv levels in ambient air. One of the major attractions of the technique was the ability to measure a wide range of volatile organic, nitrogen- and sulfur-containing compounds simultaneously and nonintrusively. The large cells could be used for both ambient air monitoring and as smog chambers for photochemical kinetic studies. They played a major role in elucidating the chemistry of urban smog.

Hanst and Hanst[30] provide a more recent, detailed description of the techniques of gas-phase FTIR measurements using long-path cells.

6.2 Open-path Monitoring of Air Pollution and Fugitive Gases

The USEPA also fostered the early development of open-path monitoring by long-path FTIR spectrometry with the ROSE (Remote Optical Sensing of Emissions) system,[67–69] which was trialed in a variety of applications, including emissions from phosphate fertilizer waste treatment ponds (HF, SiF_4), engine exhausts (CO, hydrocarbons, CH_2O), oil refinery emissions (hydrocarbons, methanol, SO_2) and power plant and waste incinerator stack emissions (CO, CO_2, hydrocarbons, NO, NH_3, HCl). However, the real growth in open long-path FTIR monitoring techniques and applications took place in the 1990s, driven largely by tighter government clean air regulations such as the US Clean Air Act Amendments of 1990.[70] The applicability of open long-path FTIR spectrometry to monitor many species simultaneously and nonintrusively at ppbv levels in many cases fits well with increased requirements on industry to monitor and control fugitive emissions from chemical plants and refineries, and monitoring the clean-up of chemically contaminated sites.

For example, Fateley and co-workers[19,71,72] describe fenceline monitoring at chemical plants, contaminated site remediation studies and remote measurements of gaseous emissions from volcanoes using open-path systems based on Bomem and Midac spectrometers at medium (typically $0.5\,cm^{-1}$) resolution. Haus and co-workers[73,74] describe a higher resolution ($0.25\,cm^{-1}$) spectrometer for both long-path absorption and IR emission studies in clean air, pollution studies and smoke stack and flare emissions. Galle and co-workers[75] used 200-m and 1-km open-path white cells combined with a $1\,cm^{-1}$ Bomem spectrometer described in section 3.2 as an open-path monitor in a number of industrial fugitive gas and greenhouse gas monitoring situations. They described the use of tracer gases such as SF_6 and N_2O deliberately released with the target gases to obtain quantitative estimates of fluxes from various sources.[75]

Although earlier work was mostly done at relatively high resolution to resolve the individual spectral lines, long open-path FTIR monitors have been increasingly based on lower resolution ($0.5–4\,cm^{-1}$), more compact, mobile and affordable spectrometers and combined with CLS and PLS quantitative analysis to resolve overlapping absorptions. Recent reviews, and published proceedings of a regular series of conferences by the

Air and Waste Management Association and the SPIE (International Society for Optical Engineering) on optical remote sensing for environmental monitoring, provide a more detailed overview and many further examples of applications.[76–83] The USEPA has published two detailed reports on open-path FTIR monitoring, the first a guidance document for use of the technique,[18] and the second the more formal and prescriptive Compendium Method TO-16 ''Long-path open-path FTIR monitoring of atmospheric gases''.[41]

6.3 High-precision Measurements of CO_2, CO, CH_4, N_2O and $^{13}CO_2/^{12}CO_2$ in Clean Air

CO_2 and CH_4 are the two most important greenhouse gases, N_2O is both a minor greenhouse gas and the principal source of stratospheric nitrogen oxides for ozone depletion and CO plays a crucial role in the background photochemistry of the atmosphere.[84] High-precision measurements of these gases are regularly made by GC and nondispersive infrared (NDIR) techniques in clean air at a network of global monitoring stations in order to understand their global source and sink budgets and impacts on global change. These four gases are also the most abundant IR-absorbing atmospheric trace gases in clean air (excluding H_2O), with typical mixing ratios of around 360 ppmv, 1700 ppbv, 310 ppbv and 50–100 ppbv, respectively. We have developed long-path FTIR spectrometry to monitor these gases simultaneously, cost effectively and with high precision.[85,86]

Spectra of dried air in 10- or 22-m White cells are recorded with a Bomem MB100 spectrometer at 1 cm^{-1} resolution using an InSb detector for low noise levels. The spectrometer and cell are carefully temperature controlled to maintain good stability. We use MALT-calculated calibration spectra and CLS quantitative analysis as described in section 5.6,[45] and achieve relative precision (i.e. repeatability) of better than 0.1% for CO_2, CH_4 and N_2O and 0.5% for CO. Absolute accuracy is of the order of 1–5% without any need for measurement of calibration gases, that is, with calibration based only on the MALT–HITRAN calculation. Higher absolute accuracy can be obtained by comparison with measurements of standard gas mixtures. This level of precision is similar to or better than that obtainable by the usual high-precision analyses currently used at background clean air monitoring stations, but is available from a single FTIR instrument.

Natural variations in the $^{13}C/^{12}C$ ratio in atmospheric CO_2 are a valuable diagnostic for deciphering the atmospheric budget of CO_2, because individual source and sink processes fractionate the isotopes differently and leave a characteristic isotopic ''signature'' in the resultant CO_2.[87] The spectrum of $^{13}CO_2$ is partially resolved from that of $^{12}CO_2$ at 1 cm^{-1} resolution and provides the basis for FTIR measurements of the $^{13}C/^{12}C$ isotopic ratio. With the same experimental configuration as described above, we routinely obtain precision for the $^{13}CO_2/^{12}CO_2$ ratio of better than 0.02%,[86,88] which is sufficient to resolve important natural variability. We are also developing methods for other isotopic ratios in trace gases using high-resolution FTIR spectrometry, such as $^{18}O/^{17}O$ in CO_2 and CO and $^{15}N/^{14}N$ in N_2O; we currently obtain a precision of the order of 0.1–0.2%.[88]

6.4 Measurements of Trace Gas Fluxes by Fourier Transform Infrared Spectrometry and Micrometeorological Techniques

Knowledge of the rates of exchange of trace gases such as CO_2, CO, CH_4 and N_2O between the earth's surface and the atmosphere in different environments is also crucial in assessing global trace gas budgets. The flux-gradient technique for measuring trace gas exchange fluxes requires high-precision measurements of the vertical profile of mixing ratios in the lowest few meters of the atmosphere, combined with measurements of profiles of temperature, water vapor, wind speed and direction. To measure typical fluxes in agricultural environments, analytical precision of the order of 0.2% is required in an in situ measurement in a few minutes. We have applied high-precision in situ FTIR analysis for this purpose in a number of field campaigns.

We use equipment similar to that described for high-precision concentration measurements (section 6.3), except that a dual-beam configuration with two matched multipass cells based on a Bomem MB100 spectrometer is used.[88–93] This design effectively doubles the rate at which samples can be measured, since one sample can be analyzed by the spectrometer while the other is being filled with the next sample to be measured, and the duty cycle of the spectrometer is maximized. In a typical experiment, air samples drawn through continuous tubing from several heights above ground on a mast are analyzed sequentially in a 2-min measurement cycle so that a complete vertical profile is determined every 20–30 min. The trace gas flux is then determined for each 30-min period from the diffusion equation and the vertical gradient of trace gas concentration (Equation 15):

$$\text{Flux} = K\frac{\partial c}{\partial z} \tag{15}$$

where K is the eddy diffusion coefficient (determined concurrently by micrometeorological techniques) and $\partial c/\partial z$ is the vertical gradient of trace gas concentration. Using high-precision FTIR trace gas analysis we can resolve small gradients of <1 ppmv CO_2, 5 ppbv CH_4,

0.2 ppbv N_2O, 0.5 ppbv CO and 4 ppbv NH_3 per meter in the required measurement times of a few minutes.

6.5 Open-path and Extractive Measurements of Biomass Burning and Combustion Gases

Biomass burning is a major contributor of many trace gases to the atmosphere. Open-path FTIR measurements are well suited to the analysis of smoke and plumes from biomass burning because they are applicable to a wide range of species and sufficiently sensitive for such relatively concentrated samples. Since many compounds in smoke are reactive, labile or otherwise difficult to sample, the noninvasive advantage of FTIR monitoring also applies. In 1989 we made a series of open-path FTIR analyses through smoke from prescribed burning of open rangeland and forest logging waste using a trailer-mounted bistatic system at 0.12 cm^{-1} resolution over paths from 30 to 100 m.[94] We were able to quantify a number of compounds simultaneously in the smoke, including CO_2, CO, CH_4, CH_2O, N_2O, NO, NO_2 and NH_3. The nitrogen-containing species in particular are difficult to measure together by other techniques, and these measurements provided a good overview of the fuel nitrogen budget during combustion.

We also used the spectrometer over a simple folded 3-m open path across the stack of a large-scale combustion research laboratory to investigate the combustion products from a range of biomass fuels.[94,95] Concentrations in the stack are much higher than in field measurements and a wide range of species has been identified despite the shorter path. The technique has since been developed further to include a dedicated FTIR spectrometer and an open-path White cell with a 1.5-m base path.[96] Of particular interest from this work has been the identification and quantification of many oxygenated organic molecules which play an important role in photochemistry and ozone formation in the smoke plume downwind of the combustion source. The spectrometer and White cell have also been installed in an aircraft and used to grab and analyze samples from flights through smoke plumes from wild and prescribed forest fires.[97]

Canister samples of smoke collected from biomass burning sources can also be conveniently analyzed for CO_2, CO, CH_4, N_2O and other minor species in a single measurement by FTIR spectroscopy in the laboratory. We have analyzed a large number of samples from Australian biomass burning studies with this technique, with samples collected in pre-evacuated 600-mL glass flasks. After drying with $MgClO_4$, the samples were analyzed on a Bomem DA3 spectrometer at 0.2 cm^{-1} resolution in a 5.6-m White cell.[98–100]

FTIR spectrometry is also used extensively in the analysis of stack gases from industrial incinerators and power plants, where again the nonintrusive and multicomponent advantages can be exploited. In general, measurements can be made either directly in an open path across the stack, or extractively, where stack gases are sampled into a measurement cell in the FTIR spectrometer. In the latter case, heated lines and cells must be used or the sample diluted to avoid condensation of the hot gases (especially water vapor) on cooler surfaces. A detailed treatment is beyond the scope of this article. Mao et al.[101] provide a detailed example.

6.6 Volcanic Emissions

Volcanoes are a significant source of many gases (such as CO_2, CO, SO_2, OCS, HF and HCl) to the atmosphere, and analysis of the plume gases from active volcanoes is also an important tool in assessing volcanic activity and eruptive state. Clearly, in situ sampling is logistically difficult and often dangerous, and remote measurements are desirable if possible. Several authors have described FTIR open-path remote measurements in the plumes of active volcanoes by a number of methods. Francis et al.[102,103] made bistatic measurements over paths of 500–1900 m across active fumaroles and craters using a conventional globar source. They measured emissions of HCl, SO_2, SiF_4 and other gases and found good correlations between different species and plume temperature. They also made measurements using naturally occurring areas of hot ground surface as a natural IR source, as did Mori and Notsu.[104] Fateley et al.[19] made similar measurements and also emission measurements of hot plumes ratioed to the colder sky background. Francis et al.[105] later made further measurements by solar occultation, where the FTIR spectrometer was located far from the target volcanic plume so that the direct solar beam passed through the plume. The sun was manually tracked into the spectrometer to provide an absorption spectrum, allowing detection of HF in addition to other molecules as above. In a different approach, McGee and Gerlach[106] used an aircraft-mounted FTIR spectrometer with White cell to collect and analyze in situ samples from traversals of the plume from Kilauea volcano, Hawaii. Finally, Love et al.[20] used two passive FTIR techniques at long distance to monitor the plume of an actively erupting volcano, Popocatépetl, Mexico. The spectrometer was located 17 km from the erupting volcano. For spectra below 1200 cm^{-1}, the emission from the hot plume itself was measured against the background of the colder clear sky. For spectra above 2000 cm^{-1}, scattered sunlight from high-altitude cloud was used as the radiation source. FTIR measurements of volcanic emissions have been reviewed by Oppenheimer et al.[107]

6.7 Remote Sensing by Solar Fourier Transform Infrared Spectrometry

The sun is a very good radiation source, with an effective blackbody temperature of 5800 K and an angular size large enough to overfill the field of view of a typical FTIR optical system. When spectra of the sun are recorded through the atmosphere, rich spectra with high S/Ns and high spectral resolution can be obtained in a few seconds to minutes. Remote sensing of atmospheric composition by solar IR spectroscopy is routinely carried out from the ground, aircraft, high-altitude balloons and from space. Beer[108] and Rao and Weber[109] provide a detailed introduction to FTIR remote sensing and atmospheric spectrometry.

Atmospheric pressure drops off approximately exponentially with altitude; the total vertical atmospheric column at sea level is equivalent to that of a single layer 8 km thick at 1 atm pressure and 273 K, and increases approximately as the secant of the solar zenith angle. In the lower atmosphere individual spectral lines are predominantly pressure broadened with near-Lorentzian lineshapes and pressure-dependent widths, whereas in the upper atmosphere pressure broadening is dominated by Doppler broadening and the lineshapes are predominantly Gaussian and dependent on temperature but not pressure. Thus atmospheric solar spectra display lineshapes that depend on the altitude range of the light path and vertical distribution of the absorber molecules in the atmosphere. Figure 7 illustrates this effect, with a portion of the solar spectrum measured at 0.004 cm^{-1} resolution near 1150 cm^{-1}. The broader lines are due to N_2O, which resides predominantly in the troposphere (<15-km altitude), and the narrow lines are due to ozone, most of which is in the stratosphere (>20 km). Quantitative analysis of solar FTIR spectra is normally carried out by NLLS fitting of calculated spectra (see, for example, sections 5.6 and 5.8) to measured spectra. The equivalent width of each line (i.e. the area under the line) is related to the total column amount of the absorbing species, and detailed analysis of the lineshape provides a method for retrieving limited information on the vertical distribution.[45,110,111]

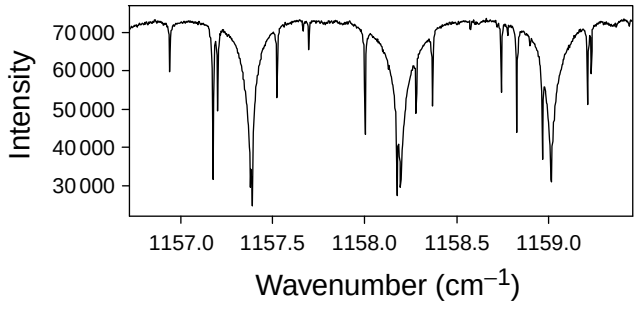

Figure 7 Small section of a solar FTIR absorption spectrum near 1156 cm^{-1} illustrating the lineshape dependence on the vertical distribution of the absorber. The three broader lines are due to N_2O, which resides predominantly in the troposphere, and the narrow lines are due to ozone, most of which is in the stratosphere.

For ground-based solar FTIR spectrometry, a sun tracker is used to image the solar disk on to the entrance aperture of the FTIR spectrometer, which normally has high resolution (0.004 cm^{-1}) to resolve the lineshapes of stratospheric gases and provide some altitude resolution. MCT detectors are normally used below 1800 cm^{-1} and InSb detectors above 1800 cm^{-1}, with optical bandpass filters usually required to reduce the total flux on the detector and avoid detector saturation. Single scans typically provide S/N of 200–1000 at high resolution. Despite its much lower intensity as a radiation source, the moon has also been used as a source for atmospheric measurements in the high Arctic.[112] A global network of ground-based solar FTIR stations operates as part of the network for detection of stratospheric change (NDSC).[113] Atmospheric constituents that can be detected include ozone (four different isotopomers), major chlorine and fluorine species (HCl, $ClONO_2$, HF and COF_2), species related to ozone destruction including ClO, HNO_3, NO, NO_2, halocarbons such as CFC-11, CFC-12 and CFC-22, major greenhouse gases such as CO_2, H_2O, HDO, CH_4 and N_2O and a number of other mainly tropospheric species including C_2H_2, C_2H_4, C_2H_6, CO, H_2CO, HCN, HO_2NO_2, NH_3, OCS and SO_2. For further details we refer readers to a sample of publications from some of the NDSC groups.[110,111,114–121]

Ground-based solar spectroscopy can be carried out from any altitude but is best from higher altitudes to reduce absorption by water vapor (which is heavily concentrated near the ground) and undesirable effects of pollution in the atmospheric boundary layer. Thus two of the best sites are at Mauna Loa in Hawaii (3400 m above sea level)[115] and the Jungfraujoch in Switzerland (3580 m above sea level).[118] Solar FTIR spectrometry from aircraft[122–124] and balloons[125–127] provides more detailed data on the upper atmosphere. For spectra collected with the sun above the observation altitude (zenith angles < 90°), the principles are the same as for ground-based measurements. For spectra collected near sunrise or sunset with the solar zenith angle below the observation altitude (>90°), absorption is dominated by the atmospheric layers nearest the tangent altitude of the solar beam. An ''onion-peeling'' technique can then be used to determine vertical profiles, in which spectra are measured as the sun sets (or rises), and the spectra of upper layers are analyzed first and the analyses successively subtracted from those of the lower layers.

Measurements can be made from up to 40-km altitude with large stratospheric balloons.

Satellite-borne FTIR solar absorption spectrometers effectively observe from above the entire atmosphere and can provide pseudo-global coverage of atmospheric composition. For example, the ATMOS spectrometer, which made several flights on the Space Shuttle in the 1980s and 1990s, has provided a large body of atmospheric composition data with broad geographic coverage. ATMOS typically measures 100 spectra at $0.017\,cm^{-1}$ resolution (60 cm OPD) during each 4-min solar occultation from the 300-km altitude orbit of the Space Shuttle; in the November 1994 mission, it measured 190 such occultations. For an overview and further details, see Gunson et al.[128] and following papers in a special issue of *Geophysical Research Letters*. A number of other satellite-borne FTIR spectrometers have already been flown, with many more planned.

6.8 Infrared Emission Spectrometry of Plumes and Stack Emissions

Sources of hot gases such as stack plumes and flares emit radiation above the ambient background level, allowing a passive, single-ended method of remote FTIR trace gas detection where no separate IR source is required. Early measurements of stack emission were made with the ROSE system described in section 6.2.[69] Haus et al.[74] have more recently described this type of measurement, shown schematically in Figure 8.

The hot plume is viewed by the FTIR spectrometer, typically through a telescope to restrict the field of view. The total intensity may be thought of as (1) the background intensity I_b transmitted through the plume and foreground atmosphere (transmittance t_p and t_f, respectively) plus (2) the plume emission I_p, transmitted through the foreground atmosphere, plus (3) any emission I_f from the foreground atmosphere. The plume emission term I_p is the product of the blackbody or Planck function $B(\tilde{\nu}, T_p)$ at the temperature T_p of the plume and the emissivity $(1 - t_p)$ of the plume. Thus (Equation 16):

$$I = I_b t_p t_f + I_p t_f + I_f \tag{16}$$

All terms in Equation (16) are implicitly dependent on frequency, $\tilde{\nu}$. Analysis of the spectrum $I(\tilde{\nu})$ requires calculation of the atmospheric transmittance and plume emission spectra and least squares fitting to the measured spectra. The contribution from the background term I_b can be measured by recording the spectrum with the telescope pointed beside the plume, and the contribution removed from the calculation. Haus et al.[74] described examples of the analysis of stack emissions from power plant plumes for a range of gases, for example, H_2O, CO_2, CO, CH_4, N_2O, NO, NO_2, SO_2, NH_3, HCl and CH_2O. While this technique is very attractive for its true remote sensing characteristic, quantitative analysis of the spectra is more complex, and both accuracy and detection limits are inferior to those achievable by in situ or cross-stack measurements if these can be made.

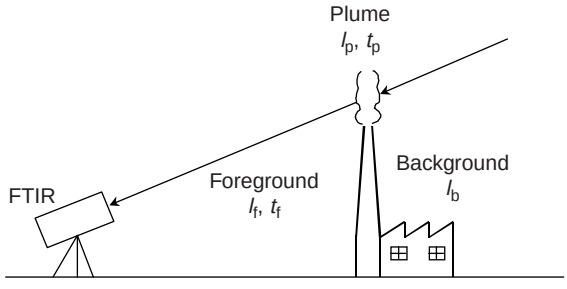

Figure 8 Schematic illustration of FTIR plume emission monitoring configuration. I_b, I_p, I_f, intensities of background, plume emission and foreground radiation, respectively; t_p, t_f, transmittance of plume and foreground, respectively.

6.9 Cryogenic Sampling: Matrix Isolation

Finally, we mention one further but somewhat different approach to atmospheric analysis by FTIR spectrometry, in which the air sample is collected and analyzed cryogenically at liquid nitrogen temperature, 77 K. The technique is described in detail elsewhere,[129,130] and we give here only a brief overview. After water vapor has been removed from an air sample, the main condensable component at 77 K is CO_2, which forms a matrix in which all other trace gases are embedded. This CO_2 is deposited as a matrix on a flat mirror or on the inner surface of an integrating sphere[15] at 77 K, and its absorption spectrum recorded for analysis of the trace gas composition of the sample. The technique is applicable to a wide range of species since virtually all trace gases are condensable, including labile or reactive species, which are effectively trapped and stable in the cold matrix. There is an effective preconcentration in the CO_2 that leads to low detection limits, in the pptv (pmol mol^{-1}) range in some cases. The technique has been used to detect and quantify phosgene ($COCl_2$) in samples collected from an aircraft in the upper troposphere and lower stratosphere[131] and to analyze for a wide range of volatile compounds in smoke samples from natural biomass burning.[98]

ABBREVIATIONS AND ACRONYMS

CLS	Classical Least Squares
DTGS	Deuterated Triglycine Sulfate

FM	Field Mirror
FTIR	Fourier Transform Infrared
fwhm	full width at half-maximum
GC	Gas Chromatography
InSb	Indium Antimonide
IR	Infrared
MCT	Mercury Cadmium Telluride
NDIR	Nondispersive Infrared
NDSC	Network for Detection of Stratospheric Change
NLLS	Nonlinear Least Squares
OM	Objective Mirror
OPD	Optical Path Difference
PAN	Peroxyacetyl Nitrate
PCR	Principal Components Regression
PLS	Partial Least Squares
RMS	Root Mean Square
S/N	Signal-to-noise Ratio
USEPA	United States Environmental Protection Agency
VOCs	Volatile Organic Compounds
ZPD	Zero Path Difference

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Automotive Emissions Analysis with Spectroscopic Techniques • Matrix Isolation Spectroscopy in Atmospheric Chemistry

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Field-based Analysis of Organic Vapors in Air

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air • Spectroscopic Techniques in Industrial Hygiene

Pulp and Paper **(Volume 9)**
Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Chemometrics **(Volume 11)**
Chemometrics • Multivariate Calibration of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

APPENDIX: SPECTRAL LIBRARIES AND DATABASES

Electronic Databases

Commercial

Company	Database	Electronic address	Ref.
Nicolet Instrument Corporation, 5225 Verona Road, Madison, WI 53711, USA	1. Vapor Phase 2. EPA Vapor 3. TGA Vapor Phase	www.nicolet.com/Spectral_libr.html	

APPENDIX (*continued)*

Company	Database	Electronic address	Ref.
Aldrich Chemical, 1001 West Saint Paul Avenue, Milwaukee, WI 53233; Mail: P.O. Box 355, Milwaukee, WI 53201, USA	1. The Aldrich Library of FTIR Spectra: Vapor Phase, Vol. 3 2. Nicolet/Aldrich FTIR Vapor Phase Library, on $3\frac{1}{2}$-in disks	www.sigmaaldrich.com/sacatalog.nsf/ GenericSearch?OpenForm (full text search on ''Vapor'')	132
Bio-Rad Laboratories, Sadtler Division, 3316 Spring Garden Street, Philadelphia, PA 19104-2596, USA	4610 Sadtler EPA Vapor Phase	www.biorad.com/621137.html	
Infrared Analysis Inc., 1558-B S. Anaheim Boulevard, Anaheim, CA 92805, USA	QASoft (Hanst Atlas)	www.infrared-analysis.com	30
National Institute of Standards and Technology, Bldg. 820, Rm. 113, Gaithersburg, MD 20899-0001, USA	NIST/EPA Gas-phase Infrared Database	www.nist.gov/srd/nist35.htm	133 134

Public Domain

Supplier/producer/organizer	Description	Electronic address	Ref.
Galactic Industries Corp.	Mixed phases	http://www.galactic.com/spconline/	
D. Sullivan, University of Texas, Chemical Engineering Department	Mixed phases	http://www.che.utexas.edu/~dls/ir/ir_dir.html	
USEPA–AECD Public Domain Gas Phase HAP Database	Low vapor pressure hazardous air pollutants (HAPs) and other species of interest	http://www.epa.gov/ttn/emc/ftir/welcome.html	
USEPA Public Domain Vapor Phase Database	EPA Vapor Phase Reference Absorbance Spectra	http://www.epa.gov/ttn/emc/ftir/welcome.html	
National Institute of Standards and Technology Chemistry WebBook	Mixed phases	http://webbook.nist.gov/chemistry/	133 134

(*continued overleaf*)

APPENDIX (*continued)*

Molecules in the HITRAN Database

HITRAN No.	Molecule	HITRAN No.	Molecule	HITRAN No.	Molecule
1	H_2O	7	O_2	13	OH
2	CO_2	8	NO	14	HF
3	O_3	9	SO_2	15	HCl
4	N_2O	10	NO_2	16	HBr
5	CO	11	NH_3	17	HI
6	CH_4	12	HNO_3	18	ClO
19	OCS	26	C_2H_2	33	HO_2
20	H_2CO	27	C_2H_6	34	O
21	HOCl	28	PH_3	35	$ClONO_2$
22	N_2	29	COF_2	36	NO^+
23	HCN	30	SF_6	37	HOBr
24	CH_3Cl	31	H_2S		
25	H_2O_2	32	HCOOH		

REFERENCES

1. S.P. Langley, 'The Absorption Lines in the Infrared Spectrum of the Sun', *Ann. Astrophys. Obs. Smithsonian Inst.*, **1**, 7 (1900).
2. M. Migeotte, 'The Fundamental Band of Carbon Monoxide at 4.7 μm in the Solar Spectrum', *Phys. Rev.*, **75**, 1108–1109 (1949).
3. E.R. Stephens, W.E. Scott, P.L. Hanst, R.C. Doerr, 'Recent Developments in the Study of the Organic Chemistry of the Atmosphere', *J. Air Pollut. Control Assoc.*, **6**(3), 159–165 (1956).
4. E.R. Stephens, 'Long Path Infrared Spectroscopy for Air Pollution Research', *Appl. Spectrosc.*, **12**, 80–84 (1958).
5. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley, New York, 1986.
6. R.J. Bell, *Introduction to Fourier Transform Spectroscopy*, Academic Press, New York, 1972.
7. J.W. Brault, 'Fourier Transform Spectroscopy', in *High Resolution in Astronomy*, eds. A. Benz, M. Huber, M. Mayor, Geneva Observatory, Geneva, 1–62, 1985.
8. J.U. White, 'Long Optical Paths of Large Aperture', *J. Opt. Soc. Am.*, **32**, 285–288 (1942).
9. D. Horn, G.C. Pimentel, '2.5 km Low Temperature Multiple Reflection Cell', *Appl. Opt.*, **10**, 1892–1898 (1971).
10. D. Ritz, M. Hausmann, U. Platt, 'An Improved Open Path Multireflection Cell for the Measurement of NO_2 and NO_3', *Proc. SPIE*, **1715**, 200–211 (1993).
11. E.C. Tuazon, R.A. Graham, A.M. Winer, R.R. Easton, J.N. Pitts, Jr, P.L. Hanst, 'A Kilometre Pathlength Fourier Transform Infrared System for the Study of Trace Pollutants in Ambient and Synthetic Atmospheres', *Atmos. Environ.*, **12**(4), 865–875 (1978).
12. J.U. White, 'Very Long Optical Paths in Air', *J. Opt. Soc. Am.*, **66**, 411–416 (1976).
13. J.B. McManus, P.L. Kebabian, M.S. Zahniser, 'Astigmaitc Mirror Multipass Cells for Long-path-length Spectroscopy', *Appl. Opt.*, **34**(18), 3336–3348 (1995).
14. S. Tranchart, I.H. Bachir, J.L. Destombes, 'Sensitive Trace Gas Detection with Near-infrared Laser Diodes and an Integrating Sphere', *Appl. Opt.*, **35**(36), 7070–7074 (1996).
15. E. Berger, D.W.T. Griffith, S.R. Wilson, G. Schuster, 'Spectroscopy of Matrices and Thin Films with an Integrating Sphere', *Appl. Spectrosc.*, **43**, 320–324 (1989).
16. U. Mueller, R. Kurte, H.M. Heise, 'Investigation of Photometric Errors in FTIR Spectra', *J. Mol. Struct.*, **482/483**, 539–544 (1999).
17. R.L. Richardson, P.R. Griffiths, 'Reduction of Stray Light in Monostatic Open Path FTIR Spectrometers with a Plane Correction Mirror', *Appl. Spectrosc.*, **51**(8), 1254–1256 (1997).
18. G.M. Russwurm, J.W. Childers, 'FTIR Open-path Monitoring Guidance Document', EPA/600/R-96/040, US Environmental Protection Agency, Research Triangle Park, NC, 1996.
19. W.G. Fateley, R.M. Hammaker, C.T. Chaffin, T.L. Marshall, 'Open Path and Extractive FTIR Environmental Monitoring Above and Below Ground Level', *Proc. SPIE*, **2504**, 2–14 (1995).
20. S.P. Love, F. Goff, D. Counce, C. Siebe, H. Delgado, 'Passive Infrared Spectroscopy of the Eruption Plume at Popocatepetl Volcano, Mexico', *Nature (London)*, **396**(6711), 563–567 (1998).
21. B. Galle, J. Mellqvist, 'A 1 Kilometer Open Path Multiple Reflection System Applied in Atmospheric FTIR Trace Gas Analysis', *Appl. Spectrosc.*, submitted (1999).

22. J.A. Manning, J.O. Burckle, S. Hedges, F.F. McElroy, 'Compendium Method TO-14A: Determination of Volatile Organic Compounds (VOCs) in Ambient Air Using Specially Prepared Canisters with Subsequent Analysis by Gas Chromatography', in *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, US Environmental Protection Agency, Cincinnati, OH, 86, 1999.
23. W.A. McClenny, M.W. Holdren, 'Compendium Method TO-15: Determination of Volatile Organic Compounds (VOCs) in Air Collected in Specially-prepared Canisters and Analyzed by Gas Chromatography/Mass Spectrometry (GC/MS)', in *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, US Environmental Protection Agency, Cincinnati, OH, 63, 1999.
24. D.A. Brymer, L.D. Ogle, C.J. Jones, D.L. Lewis, 'Viability of Using Summa Polished Canisters for the Collection and Storage of Parts Per Billion By Volume Level Volatile Organics', *Environ. Sci. Technol.*, **30**(1), 188–195 (1996).
25. J.C. Evans, J.L. Huckaby, A.V. Mitroshkov, J.L. Julya, J.C. Hayes, J.A. Edwards, L.M. Sasaki, '32-week Holding-time Study of Summa Polished Canisters and Triple Sorbent Traps Used to Sample Organic Constituents in Radioactive Waste Tank Vapor Headspace', *Environ. Sci. Technol.*, **32**(21), 3410–3417 (1998).
26. T.J. Kelly, M.W. Holdren, 'Applicability of Canisters for Sample Storage in the Determination of Hazardous Air Pollutants', *Atmos. Environ.*, **29**(19), 2595–2608 (1995).
27. A. Castellnou, N. Gonzalezflesca, J.O. Grimalt, 'Onsite Comparison of Canister and Solid-sorbent Trap Collection of Highly Volatile Hydrocarbons in Ambient Atmospheres', *Environ. Monit. Assess.*, **52**(1–2), 97–106 (1998).
28. V. Camel, M. Caude, 'Trace Enrichment Methods for the Determination of Organic Pollutants in Ambient Air', *J. Chromatogr.*, **710**(1), 3–19 (1995).
29. K.D. Oliver, J.R. Adams, E.H. Daughtrey, W.A. McClenny, M.J. Yoong, M.A. Pardee, E.B. Almasi, N.A. Kirschen, 'Technique for Monitoring Toxic VOCs in Air: Sorbent Preconcentration, Closed-cycle Cooler Cryofocusing, and GC–MS Analysis', *Environ. Sci. Technol.*, **30**(6), 1939–1945 (1996).
30. S.T. Hanst, P.L. Hanst, 'Gas Measurement in the Fundamental Infrared Region', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, 335–470, 1994.
31. U. Mueller, H.M. Heise, H. Mosebach, A.G. Gaertner, T. Hausler, 'Improved Strategies for Quantitative Evaluation of Atmospheric FTIR Spectra Obtained in Open Path Monitoring', *Field Anal. Chem. Tech.*, **3**(3), 141–159 (1999).
32. J.W. Childers, G.M. Russwurm, E.L. Thompson, B. Phillips, 'Resolution Revisited – Practical Considerations in Open Path FTIR Monitoring', in *Proceedings of Optical Sensing for Environmental and Process Monitoring – VIP-55*, Air and Waste Management Association, Pittsburgh, PA, 167–178, 1996.
33. W.F. Herget, 'Effects of Spectral Resolution on Classical Least Squares Analysis of Open Path FTIR Spectra', in *Proceedings of Optical Sensing for Environmental and Process Monitoring – VIP-55*, Air and Waste Management Association, Pittsburgh, PA, 191–204, 1996.
34. R.H. Kagann, W.T. Walter, C.D. Wang, 'Field Test to Determine Optimum Resolution for Open Path FTIR Measurements', in *Proceedings of Optical Sensing for Environmental and Process Monitoring – VIP-55*, Air and Waste Management Association, Pittsburgh, PA, 179–190, 1996.
35. T.L. Marshall, C.T. Chaffin, V.D. Makepeace, R.M. Hoffman, R.M. Hammaker, W.G. Fateley, P. Saarinen, J. Kaupinnen, 'Investigation of the Effects of Resolution on the Performance of Classical Least Squares (CLS) Spectral Interpretation Programs when Applied to Volatile Organic Compounds (VOCs) of Interest in Remote Sensing Using Open-air Long-path Fourier Transform Infrared (FTIR) Spectrometry', *J. Mol. Struct.*, **324**(1–2), 19–28 (1994).
36. P. Jaakkola, J.D. Tate, M. Paakkunainen, J. Kauppinen, P. Saarinen, 'Instrumental Resolution Considerations for Fourier Transform Infrared Gas-phase Spectroscopy', *Appl. Spectrosc.*, **51**(8), 1159–1169 (1997).
37. R.J. Berry, B.K. Hart, R.L. Richardson, P.R. Griffiths, 'A Low Resolution Spectrometer for Open Path Fourier Transform Infrared Spectrometry', *Field Anal. Chem. Tech.*, **3**(2), 131–138 (1999).
38. B.K. Hart, R.J. Berry, P.R. Griffiths, 'Effects of Resolution, Spectral Window and Background on Multivariate Calibrations used for Open Path Fourier Transform Infrared Spectrometry', *Field Anal. Chem. Tech.*, **3**(2), 117–130 (1999).
39. D.M. Haaland, 'Multivariate Calibration Methods Applied to Quantitative FTIR Analyses', in *Practical Fourier Transform Infrared Spectroscopy*, eds. J.R. Ferraro, K. Krishnan, Academic Press, San Diego, 395–468, 1990.
40. T. Hirschfeld, 'Instrumental Requirements for Absorbance Subtraction', *Appl. Spectrosc.*, **30**, 550–551 (1976).
41. G.M. Russwurm, 'Compendium Method TO-16: Long-path Open-path Fourier Transform Infrared Monitoring of Atmospheric Gases', in *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, US Environmental Protection Agency, Cincinnati, OH, 16.1–16.41, 1999.
42. I.R. Johansen, G.T. Lines, A. Honne, T. Midtgaard, 'Calibration of an FTIR Spectrometer for Ambient Air Monitoring Using PLS', *Appl. Spectrosc.*, **51**(10), 1540–1546 (1997).
43. H. Yang, P.R. Griffiths, 'Application of Multilayer Feedforward Neural Networks to Automated Compound

Identification in Low-resolution Open Path FTIR Spectrometry', *Anal. Chem.*, **71**(3), 751–761 (1999).
44. R.J. Berry, P.R. Griffiths, 'Genetic Algorithms used for Spectral Window Selection in Open Path Fourier Transform Infrared Spectrometry', in *Proceedings of 11th International Conference on Fourier Transform Spectroscopy*, Athens, GA, 1997, ed. J.A. de Haseth, American Institute of Physics, College Park, MD, 237–240, 1998.
45. D.W.T. Griffith, 'Synthetic Calibration and Quantitative Analysis of Gas Phase Infrared Spectra', *Appl. Spectrosc.*, **50**(1), 59–70 (1996).
46. L.S. Rothman, R.R. Gamache, R.H. Tipping, C.P. Rinsland, M.A.H. Smith, D.C. Benner, V.M. Devi, J.-M. Flaud, C. Camy-Peyret, A. Perrin, A. Goldman, S.T. Massie, L.R. Brown, R.A. Toth, 'The HITRAN Molecular Database: Editions of 1991 and 1992', *J. Quant. Spectrosc. Radiat. Transfer*, **48**(5/6), 469–507 (1992).
47. L.S. Rothman, C.P. Rinsland, A. Goldman, S.T. Massie, D.P. Edwards, J.M. Flaud, A. Perrin, C. Camypeyret, V. Dana, J.Y. Mandin, J. Schroeder, A. Mccann, R.R. Gamache, R.B. Wattson, K. Yoshino, K.V. Chance, K.W. Jucks, L.R. Brown, V. Nemtchinov, P. Varanasi, 'The HITRAN Molecular Spectroscopic Database and HAWKS (HITRAN Atmospheric Workstation): 1996 Edition', *J. Quant. Spectrosc. Radiat. Transfer*, **60**, 665–710 (1998).
48. N. Jacquinet-Husson, E. Arie, J. Ballard, A. Barbe, G. Bjoraker, B. Bonnet, L.R. Brown, C. Camy-Peyret, J.P. Champion, A. Chedin, A. Chursin, C. Clerbaux, G. Duxbury, J.M. Flaud, N. Fourrie, A. Fayt, G. Graner, R. Gamache, A. Goldman, V. Golovko, G. Guelachvili, J.M. Hartmann, J.C. Hilco, J. Hillman, G. Lefevre, 'The 1997 Spectroscopic GEISA Databank', *J. Quant. Spectrosc. Radiat. Transfer*, **62**(2), 205–254 (1999).
49. L.R. Brown, M.R. Gunson, R.A. Toth, F.W. Irion, C.P. Rinsland, A. Goldman, 'The 1995 Atmospheric Trace Molecule Spectroscopy (ATMOS) Linelist', *Appl. Opt.*, **35**(16), 2828–2848 (1996).
50. R.J. Anderson, P.R. Griffiths, 'Errors in Absorbance Measurements in Infrared Fourier Transform Spectrometry because of Limited Instrument Resolution', *Anal. Chem.*, **47**(14), 2339–2347 (1975).
51. C. Zhu, P.R. Griffiths, 'Extending the Range of Beer's Law in FTIR Spectrometry. Part I: Theoretical Study of Norton–Beer Apodization Functions', *Appl. Spectrosc.*, **52**(11), 1403–1408 (1998).
52. C. Zhu, P.R. Griffiths, 'Extending the Range of Beer's Law in FTIR Spectrometry. Part II: Theoretical Study of Continuous Apodization Functions', *Appl. Spectrosc.*, **52**(11), 1409–1413 (1998).
53. D.M. Haaland, 'Methods to Include Beer's Law Nonlinearities in Quantitative Spectral Analysis', in *Computerised Quantitative Infrared Analysis*, ASTM-STP-934, ed. G.L. McClure, American Society for Testing and Materials, Philadelphia, PA, 78–94, 1987.
54. C.L. Lin, E. Niple, J.H. Shaw, J.G. Calvert, 'The Quantitative Analysis of Absorption Spectra', *Appl. Spectrosc.*, **33**(5), 481–491 (1979).
55. E. Niple, 'Nonlinear Least Squares Analysis of Atmospheric Absorption Spectra', *Appl. Opt.*, **19**(20), 3481–3490 (1980).
56. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd edition, Cambridge University Press, Cambridge, 1992.
57. M.L. McKelvy, T.R. Britt, B.L. Davis, J.K. Gillie, L.A. Lenz, A. Leugers, R.A. Nyquist, C.L. Putzig, 'Infrared Spectroscopy', *Anal. Chem.*, **68**(12), R93–R160 (1996).
58. M.L. McKelvy, T.R. Britt, B.L. Davis, J.K. Gillie, F.B. Graves, L.A. Lenz, 'Infrared Spectroscopy', *Anal. Chem.*, **70**(12), R119–R177 (1998).
59. E.R. Stephens, P.L. Hanst, R.C. Doerr, W.E. Scott, 'Auto Exhaust: Composition and Photolysis Products', *J. Air Pollut. Control Assoc.*, **8**, 333–335 (1959).
60. E.R. Stephens, 'Long Path Infrared Spectroscopy for Air Pollution Research', *Infrared Phys.*, **1**, 187–196 (1961).
61. P.L. Hanst, A.S. Lefohn, B.W. Gay, 'Detection of Atmospheric Pollutants at Parts-per-billion level by Infrared Spectroscopy', *Appl. Spectrosc.*, **27**, 188–198 (1973).
62. P.L. Hanst, W.E. Wilson, R.K. Patterson, B.W. Gay, L.W. Chaney, C.S. Burton, 'A Spectroscopic Study of Californian Smog', EPA 640/4-75-006, US Environmental Protection Agency, Research Triangle Park, NC, 1975.
63. P.L. Hanst, L.L. Spiller, D.M. Watts, J.W. Spence, M.F. Miller, 'Infrared Measurement of Fluorocarbons, Carbon Tetrachloride, Carbonyl Sulfide and other Atmospheric Gases', *J. Air Pollut. Control Assoc.*, **25**, 1220–1226 (1975).
64. P.L. Hanst, N.W. Wong, J. Bragin, 'A Long-path Infrared Study of Los Angeles Smog', *Atmos. Environ.*, **16**(5), 969–981 (1982).
65. E.C. Tuazon, A.M. Winer, R.A. Graham, J.N. Pitts, Jr, 'Atmospheric Measurements of Trace Pollutants by Kilometer Pathlength FTIR Spectroscopy', in *Advances in Environmental Science and Technology, Vol. 10: Advances in Environmental Science and Technology*, eds. E.C. Tuazon, A.M. Winer, R.A. Graham, J.N. Pitts, Jr, Wiley, New York, 259–299, 1980.
66. H.W. Biermann, E.C. Tuazon, A.M. Winer, T.J. Wallington, J.N. Pitts, Jr, 'Simultaneous Absolute Measurements of Gaseous Nitrogen Species in Urban Ambient Air by Long Path Infrared and Ultraviolet–Visible Spectroscopy', *Atmos. Environ.*, **22**, 1545–1554 (1988).
67. W.F. Herget, 'Air Pollution: Ground-based Sensing of Source Emissions', in *Fourier Transform Infrared Spectroscopy*, eds. L.J. Basile, J.R. Ferraro, Academic Press, New York, Vol. 2, 111–127, 1979.
68. W.F. Herget, J.D. Brasher, 'Remote Measurement of Gaseous Pollutant Concentrations Using a Mobile

Fourier Transform Interferometer System', *Appl. Opt.*, **18**, 3404–3420 (1979).
69. W.F. Herget, 'Remote and Cross-stack Measurement of Stack Gas Concentrations Using a Mobile FTIR System', *Appl. Opt.*, **21**, 635–641 (1982).
70. B. Lee, 'Highlights of the Clean Air Act Amendments of 1990', *J. Air Waste Manag. Assoc.*, **41**, 16 (1991).
71. W.G. Fateley, R.M. Hammaker, M.D. Tucker, M.R. Witkowski, C.T. Chaffin, T.L. Marshall, M. Davis, M.J. Thomas, J. Arello, J.L. Hudson, B.J. Fairless, 'Observing Industrial Atmospheric Environments by FTIR', *J. Mol. Struct.*, **347**, 153–168 (1995).
72. M.L. Spartz, M.R. Witkowski, J.H. Fateley, J.M. Jarvis, J.S. White, J.V. Paukstelis, R.M. Hammaker, W.G. Fateley, R.E. Carter, M. Thomas, D.D. Lane, G.A. Marotz, B.J. Fairless, T. Holloway, J.L. Hudson, D.F. Gurka, 'Evaluation of a Mobile FTIR System for Rapid VOC Determination', *Am. Environ. Lab.*, **11//89**, 15–30 (1989).
73. K. Schaefer, R. Haus, J. Heland, A. Haak, 'Measurements of Atmospheric Trace Gases by Emission and Absorption Spectroscopy with FTIR', *Ber. Bunsen-Ges. Phys. Chem.*, **99**(3), 405–411 (1995).
74. R. Haus, K. Schaefer, W. Bautzer, J. Heland, H. Mosebach, H. Bittner, T. Eisenmann, 'Mobile Fourier Transform Infrared Spectroscopy Monitoring of Air Pollution', *Appl. Opt.*, **33**(24), 5682–5689 (1994).
75. B. Galle, J. Samuelsson, B. Svensson, G. Boerjesson, 'Measurements of Methane Emissions from Landfills Using a Time Correlation Tracer Method Based on FTIR Absorption Spectroscopy', *Environ. Sci. Technol.*, submitted (1999).
76. T.L. Marshall, C.T. Chaffin, R.M. Hammaker, W.G. Fateley, 'An Introduction to Open Path FTIR Atmospheric Monitoring', *Environ. Sci. Technol.*, **28**(5), 224A–232A (1994).
77. W.B. Grant, R.H. Kagann, W.A. McClenny, 'Optical Remote Measurement of Toxic Gases', *J. Air Waste Manage. Assoc.*, **42**(1), 18–30 (1992).
78. *Proceedings of Optical Remote Sensing for Environmental and Process Monitoring – VIP-55*, Air and Waste Management Association, Pittsburgh, PA, 1996.
79. *Proceedings of Optical Sensing for Environmental and Process Monitoring – VIP-37*, Air and Waste Management Association, Pittsburgh, PA, 1995.
80. P. Fabian, V. Klein, M. Tacke, K. Weber, C. Werner (eds.), *Proceedings of the International Society for Optical Engineering, Munich, 1995*, International Society for Optical Engineering, Bellingham, WA, Vol. 2506 (Air Pollution and Visibility Measurements).
81. K. Schaefer (ed.), *Proceedings of SPIE – The International Society for Optical Engineering, Munich, 1997*, SPIE – The International Society for Optical Engineering, Bellingham, WA, Vol. 3106 (Spectroscopic Atmospheric Monitoring Techniques).
82. G. Cecchi, T. Lamp, R. Reuter, K. Weber (eds.), *Proceedings of SPIE – The International Society for Optical Engineering, Munich, 1997*, SPIE – The International Society for Optical Engineering, Bellingham, WA, Vol. 3107 (Remote Sensing of Vegetation and Water, and Standardization of Remote Sensing Methods).
83. R.R. Romano (ed.), *Proceedings of SPIE – The International Society for Optical Engineering, Munich, 1996*, SPIE – The International Society for Optical Engineering, Bellingham, WA, Vol. 2883 (Optical Remote Sensing for Environmental and Process Monitoring).
84. T.E. Graedel, P.J. Crutzen, *Atmospheric Change: An Earth System Perspective*, W.H. Freeman, New York, 1993.
85. M.B. Esler, D.W.T. Griffith, S.R. Wilson, L.P. Steele, 'Precision Trace Gas Analysis by FTIR Spectroscopy 1. Simultaneous Analysis of CO_2, CO, CH_4 and N_2O in Air', *Anal. Chem.*, **72**(1), 206–215 (2000).
86. M.B. Esler, D.W.T. Griffith, S.R. Wilson, L.P. Steele, 'Precision Trace Gas Analysis by FTIR Spectroscopy 2. The $^{13}C/^{12}C$ Isotope Ratio of CO_2', *Anal. Chem.*, **72**(1), 216–221 (2000).
87. P. Ciais, P.P. Tans, J.W.C. White, M. Trolier, R.J. Francey, J.A. Berry, D.R. Randall, P.J. Sellers, J.G. Collatz, D.S. Schimel, 'Partitioning of Ocean and Land Uptake of CO_2 as Inferred by $\delta^{13}C$ Measurements from the NOAA Climate Modelling and Diagnostics Laboratory Global Air Sampling Network', *J. Geophys. Res.-Atmos.*, **100**(D3), 5051–5070 (1995).
88. D.W.T. Griffith, M.B. Esler, S.R. Wilson, 'Isotopic Analysis of Atmospheric Trace Gases by FTIR Spectroscopy', in *Proceedings of 11th International Conference on Fourier Transform Spectroscopy*, Athens, GA, 1997, ed. J.A. de Haseth, American Institute of Physics, College Park, MD, 207–210, 1998.
89. S. Christensen, P. Ambus, J.R.M. Arah, H. Clayton, B. Galle, D.W.T. Griffith, K.J. Hargreaves, L. Klemedtsson, A.-M. Lind, M. Maag, A. Scott, U. Skriba, K.A. Smith, M. Welling, F.G. Wienhold, 'Nitrous Oxide Emission from an Agricultural Field: Comparison Between Measurement by Flux Chamber and Micrometerological Techniques', *Atmos. Environ.*, **30**(24), 4183–4190 (1996).
90. B. Galle, L. Klemedtsson, D.W.T. Griffith, 'Application of a Fourier Transform IR System for Measurements of N_2O Fluxes Using Micrometeorological Methods, an Ultralarge Chamber System, and Conventional Field Chambers', *J. Geophys. Res.*, **99**(D8), 16 575–16 583 (1994).
91. D.W.T. Griffith, I.M. Jamie, P.A. Beasley, O.T. Denmead, R. Leuning, I.E. Galbally, C.P. Meyer, 'FTIR in the Paddock: Trace Gas Soil Flux Measurements Using FTIR Spectroscopy', in *Proceedings of 11th International Conference on Fourier Transform Spectroscopy*, Athens, GA, 1997, ed. J.A. de Haseth, American Institute of Physics, College Park, MD, 211–214, 1998.

92. D.W.T. Griffith, B. Galle, 'Flux Measurements of NH_3, N_2O and CO_2 Using Dual Beam FTIR Spectroscopy and the Flux-gradient Technique', *Atmos. Environ.*, **34**(7), 1087–1098 (2000).
93. K.J. Hargreaves, F.G. Wienhold, L. Klemedtsson, J.R.M. Arah, I.J. Beverland, D. Fowler, B. Galle, D.W.T. Griffith, U. Skriba, K.A. Smith, M. Welling, G.W. Harris, 'Measurement of Nitrous Oxide Emission from Agricultural Land Using Micrometeorological Methods', *Atmos. Environ.*, **30**(10–11), 1563–1571 (1996).
94. D.W.T. Griffith, W.G. Mankin, M.T. Coffey, D.E. Ward, A. Riebau, 'FTIR Remote Sensing of Biomass Burning Emissions of CO_2, CO, CH_4, CH_2O, NO, NO_2, NH_3 and N_2O', in *Global Biomass Burning*, ed. J.S. Levine, MIT Press, Cambridge, MA, 230–239, 1991.
95. R.J. Yokelson, D.W.T. Griffith, D.E. Ward, 'Open Path FTIR Studies of Large-scale Laboratory Biomass Fires', *J. Geophys. Res.*, **101**, 21 067–21 080 (1996).
96. R.J. Yokelson, R. Susott, D.E. Ward, J. Reardon, D.W.T. Griffith, 'Emissions from Smoldering Combustion of Biomass Measured by Open-path FTIR Spectroscopy', *J. Geophys. Res.*, **102**(D15), 18 865–18 877 (1997).
97. R.J. Yokelson, J.G. Goode, D.E. Ward, R.A. Susott, R.E. Babbit, D.D. Wade, I. Bertschi, D.W.T. Griffith, W.M. Hao, 'Emissions of Formaldehyde, Acetic Acid, Methanol and Other Trace Gases from Biomass Fires in North Carolina Measured by Airborne Fourier Transform Infrared Spectroscopy, *J. Geophys. Res.*, **104**(D23), 30 109–30 125, 1999.
98. D.F. Hurst, D.W.T. Griffith, J.N. Carras, D.J. Williams, P.J. Fraser, 'Measurements of Trace Gases Emitted by Australian Savanna Fires During the 1990 Dry Season', *J. Atmos. Chem.*, **18**, 33–56 (1994).
99. D.F. Hurst, D.W.T. Griffith, G.D. Cook, 'Trace Gas Emissions from Biomass Burning in Tropical Australian Savannas', *J. Geophys. Res.*, **99**(D8), 16 441–16 456 (1994).
100. D.F. Hurst, D.W.T. Griffith, G.D. Cook, 'Trace Gas Emissions from Biomass Burning in Australia', in *Biomass Burning and Global Change, 2, Biomass Burning in South America, Southeast Asia, and Temperate and Boreal Ecosystems, and the Kuwait Oil Fires*, ed. J.S. Levine, MIT Press, Cambridge, MA, 787–792, 1996.
101. Z. Mao, J. Demergian, A. Mathew, R. Hyre, 'Use of Fourier Transform Infrared Spectrometry as a Continuous Emission monitor', *Waste Manag.*, **15**(8), 567–577 (1995).
102. P. Francis, A. Maciejewski, C. Oppenheimer, C. Chaffin, T. Caltabiano, 'SO_2–HCl Ratios in the Plumes from Mt. Etna and Volcano Determined by Fourier Transform Spectroscopy', *Geophys. Res. Lett.*, **22**(13), 1717–1720 (1995).
103. P. Francis, C. Chaffin, A. Maciejewski, C. Oppenheimer, 'Remote Determination of SiF_4 in Volcanic Plumes: A New Tool for Volcano Monitoring', *Geophys. Res. Lett.*, **23**(3), 249–252 (1996).
104. T. Mori, K. Notsu, 'Remote CO, COS, CO_2, SO_2 and HCl Detection and Temperature Estimation of Volcanic Gas', *Geophys. Res. Lett.*, **24**(16), 2047–2050 (1997).
105. P. Francis, M.R. Burton, C. Oppenheimer, 'Remote Measurements of Volcanic Gas Compositions by Solar Occultation Spectroscopy', *Nature (London)*, **396**(6711), 567–570 (1998).
106. K.A. McGee, T.M. Gerlach, 'Airborne Volcanic Plume Measurements Using an FTIR Spectrometer, Kilauea Volcano, Hawaii', *Geophys. Res. Lett.*, **25**(5), 615–618 (1998).
107. C. Oppenheimer, P. Francis, M. Burton, A.J.H. Maciejewski, L. Boardman, 'Remote Measurements of Volcanic Gases by Fourier Transform Infrared Spectroscopy', *Appl. Phys. B*, **67**(4), 505–515 (1998).
108. R. Beer, *Remote Sensing by Fourier Transform Infrared Spectrometry*, Wiley, New York, 1992.
109. K.N. Rao, A. Weber (eds.), *Spectroscopy of the Earth's Atmosphere and Interstellar Medium*, Academic Press, San Diego, 1992.
110. X. Liu, F.J. Murcray, D.G. Murcray, J.M. Russell, 'Comparison of HF and HCl Vertical Profiles from Ground-based High-resolution Infrared Solar Spectra with Halogen Occultation Experiment Observations', *J. Geophys. Res.-Atmos.*, **101**(D6), 10 175–10 181 (1996).
111. C.P. Rinsland, N.B. Jones, B.J. Connor, J.A. Logan, N.S. Pougatchev, A. Goldman, F.J. Murcray, T.M. Stephen, A.S. Pine, R. Zander, E. Mahieu, P. Demoulin, 'Northern and Southern Hemisphere Ground Based Infrared Spectroscopic Measurements of Tropospheric Carbon Monoxide and Ethane', *J. Geophys. Res.*, **103**(D21), 28 197–28 218 (1998).
112. J. Notholt, 'The Moon as a Light Source for FTIR Measurements of Stratospheric Trace Gases during the Polar Night – Application for HNO_3 in the Arctic', *J. Geophys. Res.-Atmos.*, **99**(D2), 3607–3614 (1994).
113. M.J. Kurylo, 'Network for the Detection of Stratospheric Change', *Proc. SPIE*, **1491**, 169–174 (1991).
114. G.P. Adrian, T. Blumenstock, H. Fischer, L. Gerhardt, T. Gulde, H. Oelhaf, P. Thomas, O. Trieschmann, 'Column Amounts of Trace Gases Derived from Ground-based Measurements with MIPAS During CHEOPS III', *Geophys. Res. Lett.*, **18**(4), 783–786 (1991).
115. A. Goldman, F.J. Murcray, F.H. Murcray, R.D. Blatherwick, D.G. Murcray, 'Infrared Measurements of Atmospheric Gases Above Mauna Loa, Hawaii, in February 1987', *J. Geophys. Res.*, **93**, 12 607–12 626 (1988).
116. D.W.T. Griffith, N.B. Jones, W.A. Matthews, 'Interhemispheric Ratio and Annual Cycle of Carbonyl Sulfide (OCS) Total Column from Ground Based Solar FTIR Spectra', *J. Geophys. Res.*, **103**(D7), 8447–8454 (1998).
117. B. Galle, J. Mellqvist, D.W. Arlander, I. Floisand, M.P. Chipperfield, A.M. Lee, 'Ground Based FTIR Measurements of Stratospheric Species from Harestua, Norway

During SESAME and Comparison with Models', *J. Atmos. Chem.*, **32**(1), 147–164 (1999).

118. E. Mahieu, R. Zander, L. Delbouille, P. Demoulin, G. Roland, C. Servais, 'Observed Trends in Total Vertical Column Abundances of Atmospheric Gases from IR Solar Spectra Recorded at the Jungfraujoch', *J. Atmos. Chem.*, **28**(1–3), 227–243 (1997).
119. J. Notholt, A. Meier, S. Peil, 'Total Column Densities of Tropospheric and Stratospheric Trace Gases in the Undisturbed Arctic Summer Atmosphere', *J. Atmos. Chem.*, **20**(3), 311–332 (1995).
120. N.B. Jones, M. Koike, W.A. Matthews, B.M. McNamara, 'Southern Hemisphere Mid-latitude Seasonal Cycle in Total Column Nitric Acid', *Geophys. Res. Lett.*, **21**(7), 593–596 (1994).
121. R. Sussmann, K. Schaefer, 'Infrared Spectroscopy of Tropospheric Trace Gases: Combined Analysis of Horizontal and Vertical Column Abundances', *Appl. Opt.*, **36**(3), 735–741 (1997).
122. W.G. Mankin, 'Airborne Fourier Transform Spectroscopy of the Upper Atmosphere', *Opt. Eng.*, **17**(1), 39–43 (1978).
123. G.C. Toon, C.B. Farmer, L.L. Lowes, P.W. Schapper, J.-F. Blavier, R.H. Norton, 'Infrared Aircraft Measurements of Stratospheric Composition over Antarctica During September 1987', *J. Geophys. Res.*, **94**(D14), 16 571–16 596 (1989).
124. M.T. Coffey, W.G. Mankin, A. Goldman, 'Simultaneous Spectroscopic Determination of the Latitudinal, Seasonal, and Diurnal Variability of Stratospheric N_2O, NO, NO_2 and HNO_3', *J. Geophys. Res.*, **86**(C8), 7331–7341 (1981).
125. F.J. Murcray, J.J. Kosters, R.D. Blatherwick, J. Olson, D.G. Murcray, 'High Resolution Solar Spectrometer System for Measuring Atmospheric Constituents', *Appl. Opt.*, **29**(10), 1520–1525 (1990).
126. B. Sen, G.C. Toon, J.F. Blavier, E.L. Fleming, C.H. Jackman, 'Balloon-borne Observations of Midlatitude Fluorine Abundance', *J. Geophys. Res.-Atmos.*, **101**(D4), 9045–9054 (1996).
127. A. Goldman, R.D. Blatherwick, F.J. Murcray, D.G. Murcray, 'University of Denver Infrared Spectral Atlases', *Appl. Opt.*, **35**(16), 2821–2827 (1996).
128. M.R. Gunson, M.M. Abbas, M.C. Abrams, M. Allen, L.R. Brown, T.L. Brown, A.Y. Chang, A. Goldman, F.W. Irion, L.L. Lowes, E. Mahieu, G.L. Manney, H.A. Michelsen, M.J. Newchurch, C.P. Rinsland, R.J. Salawitch, G.P. Stiller, G.C. Toon, Y.L. Yung, R. Zander, 'The Atmospheric Trace Molecule Spectroscopy (Atmos) Experiment – Deployment on the Atlas Space Shuttle Missions', *Geophys. Res. Lett.*, **23**(17), 2333–2336 (1996).
129. D.W.T. Griffith, G. Schuster, 'Atmospheric Trace Gas Analysis Using Matrix-isolation-Fourier Transform Infrared Spectroscopy', *J. Atmos. Chem.*, **5**, 59–81 (1987).
130. D.W.T. Griffith, 'Matrix Isolation Spectroscopy in Atmospheric Chemistry', in *Air Monitoring by Spectroscopic Techniques*, Chemical Analysis, ed. M.W. Sigrist, Wiley, New York, 471–514, Vol. 127, 1994.
131. S.R. Wilson, P.J. Crutzen, G. Schuster, D.W.T. Griffith, G. Helas, 'Phosgene Measurements in the Upper Troposphere and Lower Stratosphere', *Nature (London)*, **334**, 689–691 (1988).
132. C.J. Pouchert, *The Aldrich Library of Infrared Spectra*, 3rd edition, Aldrich Chemical, Milwaukee, WI, 1981.
133. P.M. Chu, F.R. Guenther, G.C. Rhoderick, W.J. Lafferty, 'The NIST Quantitative Infrared Database', *J. Res. Natl. Inst. Stand. Technol.*, **104**(1), 59–81 (1999).
134. P.M. Chu, G.C. Rhoderick, D. Vanvlack, S.J. Wetzel, W.J. Lafferty, F.R. Guenther, 'A Quantitative Infrared Spectral Database of Hazardous Air Pollutants', *Fresenius. J. Anal. Chem.*, **360**(3–4), 426–429 (1998).

Infrared LIDAR Applications in Atmospheric Monitoring

Brian J. Orr
Macquarie University, Sydney, Australia

1 Introduction **1**
1.1 Historical Perspectives 2
1.2 Sources of Information 3
1.3 The Earth's Atmosphere 3
1.4 Nonresonant Optical Scattering Processes 3
1.5 Resonant Optical Processes 4
2 Theory and Operating Principles **6**
2.1 Nonresonant Infrared LIDAR 6
2.2 Differential Absorption LIDAR 7
3 Infrared LIDAR Instrumentation **8**
4 Coherent Light Sources for Infrared LIDAR Applications **10**
4.1 Fixed-wavelength Lasers 10
4.2 Line-tunable Lasers 12
4.3 Continuously Tunable Lasers 13
4.4 Nonlinear-optical Coherent Infrared Sources 14
4.5 Multiwavelength Coherent Sources for Differential Absorption LIDAR and Mutiplex Sensing 16
5 Perspectives and Future Developments **18**
Acknowledgments **19**
Abbreviations and Acronyms **19**
Related Articles **20**
References **20**

The acronym LIDAR stands for light detection and ranging, an optical analog of RADAR (radio detection and ranging). The conventional version of LIDAR requires a laser transmitter to launch short pulses of coherent light, which are scattered from atmospheric targets of interest back to an optical receiver, with a time delay that is determined by the range of the target. Optical phenomena in the Earth's atmosphere (e.g. Rayleigh scattering, Raman scattering, Mie scattering, refraction, and resonant absorption) contribute to the amplitude of optical signals returning to the receiver and their characteristic wavelength dependence allows us to measure the concentration and velocity distributions of different atmospheric molecules and aerosol particles. LIDAR backscattering in the infrared (IR) region, on which this article concentrates, is well suited to detecting aerosols (as in clouds or industrial particulate emissions). IR DIAL (differential absorption LIDAR), a variant in which two wavelengths are used simultaneously to separate resonant molecular signals from background, enables most molecular species to be monitored by means of their IR absorption spectra. A closely related approach comprises long-path IR laser absorption, with retro-reflection from a topographic target (e.g. a strategically located mirror); this sacrifices optical range information but gains sensitivity because it integrates over all molecules in the optical column between the transmitter/receiver and the reflector. Such techniques are vitally dependent on pulsed IR laser technology.

1 INTRODUCTION

This article is concerned with laser-based atmospheric sensing in the IR region (arbitrarily defined here by the wavelength range $\sim$0.7–20 μm, or $\sim$14 000–500 cm^{-1} in wavenumber units). Atmospheric sensing by IR LIDAR relies on two key factors: the strength of IR scattering from aerosols and particulates (relative to that from molecules) and the amenability of most molecular species to be monitored by means of their IR absorption spectra. These factors are vital to the general range-resolved form of IR DIAL, which depends on elastic scattering from atmospheric aerosols (to act as a ''distributed mirror'') and on characteristic vibration–rotation absorption spectra as signatures of specific molecules in the atmosphere (e.g. water, methane, ozone, and various pollutant species). The latter attribute also enables retro-reflected long-path IR laser absorption which entails a trade-off between range resolution and sensitivity. Such IR laser-based techniques contrast with atmospheric sensing in the visible and ultraviolet (UV) regions, where the optical processes involved arise predominantly from electronic properties of molecules: Rayleigh and Raman scattering, electronic absorption, and fluorescence.

Several atmospheric sensing strategies involving IR LIDAR and related techniques are shown in Figure 1(a–c). These will be discussed in more detail at appropriate stages of this article. In early literature, terms such as ''optical RADAR'' and ''laser RADAR'' were synonymous with ''LIDAR'', in view of its range-resolution capabilities. However, LIDAR tends now to embrace techniques that monitor an optically defined column of the atmosphere, as in Figure 1(b) and (c), in addition to true range-resolved LIDAR. That broader view of IR LIDAR will be adopted in this article.

Reviews of LIDAR-based atmospheric sensing usually concentrate either on particular detection techniques (e.g.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

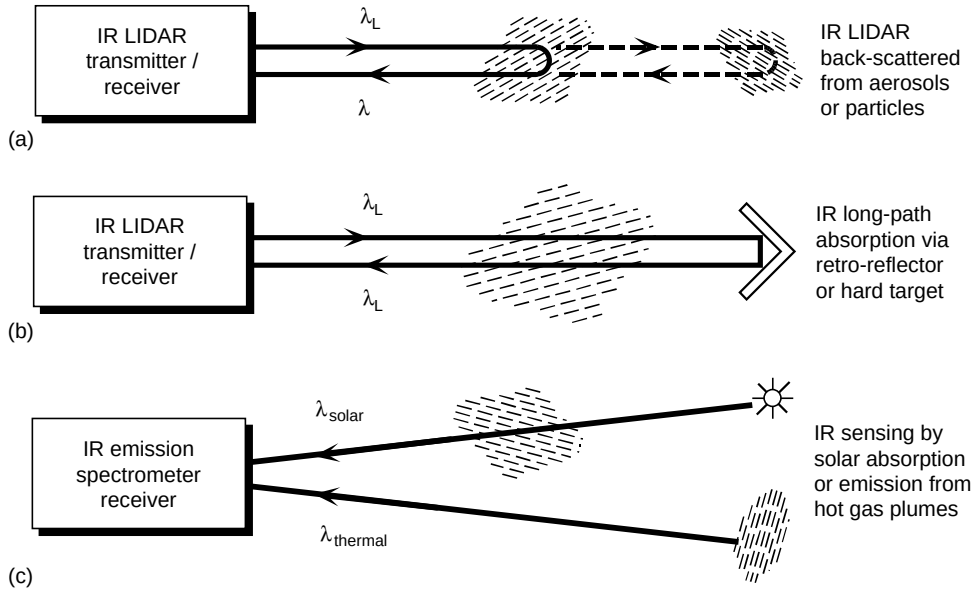

Figure 1 IR LIDAR, DIAL, and related atmospheric sensing strategies.

DIAL, Raman LIDAR or laser heterodyne radiometry) or on a particular class of atmospheric phenomena (such as aerosols, particulates, or pollutant emissions), rather than on a single spectral region. This article is concerned with LIDAR-based atmospheric monitoring in the IR region (arbitrarily defined here by the wavelength range ~0.7–20 μm, or ~14 000–500 cm^{-1} in wavenumber units). Nevertheless, IR LIDAR cannot be treated in isolation from complementary LIDAR techniques using visible or UV laser radiation, so comparisons will be made where appropriate. The article aims to integrate principles of IR LIDAR instrumentation and its applications to atmospheric sensing.

Some of those applications are treated in this Introduction. Others are classified in section 3, according to the forms of coherent light source employed, notably fixed-wavelength solid-state lasers (e.g. Nd : YAG (yttrium aluminum garnet) lasers at 1.06 μm), line-tunable gas lasers (e.g. CO_2 lasers at 9–11 μm), tunable solid-state lasers (e.g. $Ti^{3+} : Al_2O_3$ lasers at 0.7–1 μm, alexandrite at 0.7–0.8 μm, and various semiconductor diode lasers), and nonlinear-optical wavelength converters, such as Raman shifters, difference-frequency generators, and optical parametric oscillators (OPOs).

1.1 Historical Perspectives

Interactions between light and the Earth's atmosphere produce many natural spectacles, such as rainbows, sunsets, blue skies, and clouds.[1] In a sense, therefore, there is not much new about using optical phenomena to monitor and predict atmospheric conditions, as this probably comprised one of the earliest forms of human observation and inference. During the last few decades, however, natural phenomena arising from sunlight have been supplemented by using artificial light sources (notably lasers) to probe the atmosphere and yield meteorological and pollution-related information in unprecedented detail. In the same period, our knowledge of atmospheric chemistry and physics[2] has been much enhanced by the analytical capability of optical and spectroscopic sensing.[3–6]

For at least half a century, it has been well understood that a short pulse of light could be used to locate and map atmospheric objects such as cloud ceilings. This approach, for which the acronym "LIDAR" was coined as early as 1953,[7] entails timing the interval between the light pulse from an optical transmitter (e.g. a powerful flashlamp) and a pulsed return signal, detected by an optical receiver. With the advent of pulsed ruby lasers in 1960,[8] LIDAR advanced rapidly[6,9] – no doubt exceeding expectations of its early pioneers from the searchlight era. Developments in laser technology[10–14] have since then continued to drive many of the key advances in LIDAR methodology.

DIAL[15,16] is an important differential spectroscopic adjunct to LIDAR; two wavelengths of pulsed light are used simultaneously to detect signals from an atmospheric species of interest in the presence of unwanted background signals. Long-path laser absorption, where range resolution is exchanged for sensitivity, is closely related to DIAL. Other related techniques include DOAS [differential optical absorption spectroscopy, usually recorded by dispersing broadband ultraviolet/visible (UV/VIS) light][17] and DISC (differential scattering).

IR LIDAR is well served by pulsed IR lasers, of which Nd : YAG and carbon dioxide (CO_2) lasers have been most prominent, superseding earlier ruby lasers[8,10,14] operating at ~694 nm. Nd : YAG lasers have a solid-state gain medium,[14] in which neodymium ions (Nd^{3+}) in a YAG ($Y_3Al_5O_{12}$) host, are flashlamp- or diode-pumped to yield high-quality laser output at 1.064 μm (or its harmonics at 532, 355 or 266 nm for UV/VIS LIDAR applications). The CO_2 laser is a discharge-excited gas laser that is line tunable over rotation–vibration bands of the IR emission spectrum of CO_2, in the range 9–11 μm.[18–21] Apart from these two ''workhorse'' lasers for IR LIDAR, there are other useful lasers and nonlinear-optical devices (such as OPOs)[22] to be discussed later in this article. Some are continuously tunable and many are adaptable for field-based sensing. A key theme of this article concerns advances in IR laser technology that have steered the progress of IR LIDAR over the last 30 years and that are set to continue to do so in the future.

1.2 Sources of Information

IR LIDAR is a mature technological field, well served by many excellent literature resources. The following brief bibliographic survey is not intended to be comprehensive, but rather to provide a broad foundation and an indicative sample of the vast body of LIDAR-related literature. For instance, Measures has produced a major textbook[3] and an edited work.[4] Grant et al. have generated several useful reviews[6,23,24] and a compilation of research papers,[9] as well as an ongoing electronic LIDAR bibliography.[25] Notable review collections include those edited by Hinkley,[26] Killinger and Mooradian,[27] and Sigrist.[5] Various other early review articles[28–33] give useful perspectives on the subject. The reader is also referred to authoritative reference books on lasers,[10–14] laser spectroscopy,[34,35] and applications of quantum electronics and nonlinear optics.[36–38]

1.3 The Earth's Atmosphere

The objective of IR LIDAR methodology reviewed in this article is the Earth's atmosphere: its composition, structure, dynamics, and chemistry.[2,39] These factors influence problems such as global warming, stratospheric ozone depletion, and photochemical smog formation.

Meteorology drives much of the interest in LIDAR, which is able to provide information on aerosols, atmospheric chemistry, clouds, winds, and pollution. In the case of aerosols,[40] LIDAR can measure and map their characteristic spatial, size and shape distributions, as well as their chemical and physical composition. The temperature-inversion mixing layer that traps urban pollution can be monitored by ground-based LIDAR, as can localized plumes of particulates and chemicals emitted from industrial sites. Airborne LIDAR platforms can cover large areas quickly, e.g. to monitor migration of aerosols across international boundaries or to study the effects of major volcanic eruptions.

Meteorological parameters of the Earth's troposphere and stratosphere that can be measured by various forms of LIDAR include wind velocities, temperature, pressure, density, and humidity.[6] LIDAR measurements in the IR region do not generally suffice to obtain all of these measured parameters. As will become apparent in sections 1.4 and 1.5, two or more IR, visible, or UV laser wavelengths are often used to determine different atmospheric properties and to separate the distinct signals from molecules and aerosols. IR LIDAR (particularly DIAL)[15,16] is a useful way to monitor the molecular composition of the atmosphere, by means of spectroscopic signatures contained in vibration–rotation bands of particular molecules. Atmospheric concentration profiles of water and ozone are of particular interest.

1.4 Nonresonant Optical Scattering Processes

Various forms of interaction between light and matter are at the heart of optical remote sensing methods such as LIDAR. Relevant atmospheric sensing schemes have already been depicted schematically in Figure 1(a–c). In this section, we consider ''nonresonant'' processes, which do not require the laser light to be tuned to a specific resonant wavelength. Resonant optical processes will be considered separately in section 1.5.

As already stated, this article emphasises phenomena that are observable with IR light sources and are therefore not visible by the human eye (400–700 nm). Even in the natural world, however, the refraction of sunlight from water droplets gives rise to IR rainbows which can be photographed with IR-sensitive film at 800–900 nm.[1] Likewise, there are advantages for LIDAR in accessing the IR region, as well as the visible and UV regions.

Two forms of quasi-elastic light scattering are important here: Rayleigh scattering (from individual molecules) and Mie scattering (from particles that are generally of size comparable to the wavelength of the light itself).[41,42] Neither process involves a spectroscopic resonance, so that the choice of laser wavelength λ is not critical or selective for particular scattering species (note, however, that scattering cross-sections at wavelengths close to resonance may be enhanced markedly relative to those off resonance; resonance fluorescence occurs when the laser wavelength is exactly resonant).

The λ^{-4} wavelength dependence of Rayleigh scattering causes it to be relatively unimportant for laser wavelengths longer than ~1 μm and therefore only barely relevant to the IR LIDAR topic of this article. Clear skies

appear blue and sunsets red because blue light is Rayleigh scattered more strongly than red light.[1]

Mie scattering[41,42] from particles (e.g. aerosols, clouds, ice particles, dust) remains of critical interest in the context of IR LIDAR.[40] The original theory, which dates back to work published in 1908 by Gustav Mie (and independently, in 1909, by Peter Debye),[43] refers only to spherical particles and their characteristic angle and polarization dependences. Mie's theory can be adapted to the case of ''real-life'' nonspherical particles.[44,45] The wavelength dependence of Mie scattering from aerosols and other particles is more complicated than the simple λ^{-4} law for Rayleigh scattering from small molecules. It usually varies as $\lambda^{-\delta}$, where $0 < \delta < 3$, with δ smaller for bigger particles; if λ greatly exceeds the particle diameter then the scattering will be weak, but it will increase markedly as λ becomes equal to, or smaller than, the particle diameter. There is therefore a trade-off between particle diameter in the atmospheric target and the wavelength, efficiency, and availability of laser. Clouds often appear white because cloud aerosols are relatively large and Mie scatter visible sunlight with a much less pronounced wavelength dependence than for Rayleigh scattering.[1]

Many aerosol scattering measurements are made in the spectral region 0.5–1.5 μm, so that Nd : YAG lasers (at 1.06 μm) are useful here. CO_2 lasers (at 10.6 or 9.11 μm) are also widely used, with their relatively weak scattering compensated by the high pulse energy and electrical efficiency that is attainable with a CO_2 laser, combined with the atmosphere's relatively high transparency at 9–11 μm. The reliable application of such a CO_2 laser system to aerosol scattering requires careful calibration of scattering efficiency.[40]

Multiwavelength LIDAR systems are sometimes used to determine variations between size distributions of different aerosol types.[46] Most of these use UV and visible laser wavelengths but one[47] in the IR LIDAR context has employed a $Ti^{3+} : Al_2O_3$ laser to generate radiation at 750 and 850 nm and also Nd : YAG laser harmonics; this system has allowed polarization-sensitive measurements of aerosol size distribution in polar stratospheric clouds and the Mount Pinatubo volcanic eruption.[40]

We refer to Rayleigh and Mie scattering as ''quasi-elastic'' since each is sensitive to motion of the molecules or particles along the direction of the laser beam, yielding a Doppler shift $\Delta\nu_D$ of the 180°-backscattered radiation relative to the frequency ν_L of the incident laser light (Equation 1):

$$\Delta\nu_D = 2\nu_L \frac{v}{c} \tag{1}$$

where v is the particle velocity along the laser beam and c is the speed of light; such differences between laser wavelength λ_L and received wavelength λ are implicit in Figure 1(a).

Mie scattering from atmospheric aerosol particles, moving with the air mass, displays small Doppler shifts that yield wind speed profiles. CO_2-laser Doppler velocimetry LIDAR systems have achieved this objective in various applications.[6,48,49] Sensing of tropospheric horizontal wind velocities from airborne and space-based Doppler LIDAR platforms receive much ongoing attention.[6,49] Doppler shifts are recorded by combining the backscattered return signals (at frequency $\nu_L + \Delta\nu_D$) coherently with part of the outgoing laser radiation (at ν_L), using a heterodyne detection system[50] in which beat frequencies yield the Doppler shifts $\Delta\nu_D$. Wind velocities are thereby determined to an accuracy of 1 m s^{-1}.

Accompanying quasi-elastic Rayleigh scattering is its inelastic counterpart, Raman scattering. This too is a nonresonant molecular process, in the sense that it is a slowly varying function of laser wavelength λ_L. However, Raman scattering carries information about the molecular species from which the scattering emanates, with characteristic Raman shifts $\Delta\nu_R$ separating the frequency ν_R of scattered radiation from the exciting laser frequency ν_L[51,52] (Equation 2):

$$\Delta\nu_R = \nu_L - \nu_R \tag{2}$$

The strongest Raman scattering is for positive values of $\Delta\nu_R$ (i.e. $\lambda_R > \lambda_L$) and is known as Stokes scattering. Raman scattering, like Rayleigh scattering, has a λ^{-4} wavelength dependence and so is more efficiently excited by UV and visible lasers than by the IR lasers on which this article concentrates. However, Raman LIDAR[53–55] is an important adjunct to IR LIDAR techniques used to measure molecular species and concentrations in the atmosphere. Its mode of detection is similar to that depicted in Figure 1(a), except that Raman scattering is usually excited in the UV or visible region, with received wavelength $\lambda = \lambda_R$, and arises from molecules rather than aerosol particles. Raman LIDAR is particularly useful for detecting water vapor (for which $\Delta\nu_R \approx 3652\,\text{cm}^{-1}$);[55] see section 5 for aspects of recent progress. It can also monitor homonuclear diatomic species, such as N_2 ($\Delta\nu_R \approx 2331\,\text{cm}^{-1}$), which has no IR-active vibrational mode and is therefore not amenable to IR LIDAR; this approach can measure temperature and density profiles of N_2 in the upper troposphere and stratosphere.[55]

1.5 Resonant Optical Processes

The preceding section covered a variety of nonresonant optical scattering processes (notably Rayleigh, Raman, and Mie scattering) in which the scattering cross-section is generally a slowly varying function of laser wavelength and depends on the real part of the optical susceptibility

of the scattering medium.[56] In this sense, Rayleigh scattering is closely related to refraction and normal dispersion in a dilute optical medium. In regions of anomalous dispersion, however, the imaginary part of the (complex) optical susceptibility becomes significant and rapid variations of scattering cross-section occur as the laser wavelength is varied. More significantly, the phenomenon of optical absorption then arises, yielding a rich source of spectroscopic information and a principal area of interest in IR LIDAR.

Molecular absorption spectra provide a characteristic signature that is useful for identification and analytical determination. UV and visible absorption spectra (which are beyond the scope of this article) involve transitions to excited molecular electronic states. For instance, stratospheric ozone (O_3) has strong UV absorption bands that provide a protective filter against harmful solar UV-B radiation (280–320 nm); its depletion has created the infamous ''ozone hole''.[39] The same UV absorption bands can enable O_3 to be detected and mapped using, for instance, XeCl excimer lasers operating at 307.9 nm.[16]

In the contrasting context of IR LIDAR, the key transitions are between quantized vibrational states associated with normal modes of molecular vibration. The theory and practice of IR absorption spectroscopy are established,[57–59] as are correlations between such spectra and molecular structure.[52] Heliostatically tracked IR spectra of solar radiation transmitted through the atmosphere, as shown in the upper part of Figure 1(c), are used traditionally to characterize atmospheric constituents[28,59] and can be monitored with high sensitivity by laser heterodyne radiometry,[60] either airborne or ground-based. IR LIDAR absorption spectroscopy in general, and DIAL in particular, will be discussed later in this article. In the meantime, we note that CO_2 lasers have a fortuitous coincidence at $\sim$9.6 μm with a vibrational absorption band of ozone; this enables IR DIAL methods to measure trace concentrations of tropospheric O_3 at a range of 1–2 km.[16]

Fluorescence, an optical phenomenon that is effectively the reverse of resonant absorption, occurs when an atom, ion, or molecule spontaneously emits radiation from a high-energy quantum state that has been excited (e.g. by electric discharge, optical absorption, or heating); it has a finite fluorescence quantum yield that is subject to collisional quenching.

Laser-induced fluorescence (LIF) can occur either elastically (resonance fluorescence) or inelastically (with emitted wavelength $\lambda \neq \lambda_L$) and its LIDAR counterpart employs a detection scheme similar to that depicted in Figure 1(a). Fluorescence LIDAR[61] can be used, for example, to study concentrations of the hydroxyl radical (OH) in the troposphere,[62,63] but it is generally confined to UV and visible excitation. Trace alkali metal atoms in the atmosphere are particularly amenable to LIDAR monitoring by laser-induced visible or UV resonance fluorescence.[15,54,64,65]

The intrinsically long radiative lifetime of laser-induced infrared fluorescence (LIIRF) yields a low intensity, which is further diminished by collisional quenching.[66] Early proposals to use LIIRF for range-resolved atmospheric monitoring[29,31,32,67] were optimistic, with traces of diatomic species such as CO and NO predicted to be detectable at a range of 100 m by high-power pulsed IR lasers. However, it was soon recognized[68,69] that far higher sensitivity could be attained by range-resolved IR DIAL methods, with atmospheric aerosols and particulates in the atmosphere serving as distributed reflectors, as in Figure 1(a). In this IR DIAL approach, elastically backscattered laser radiation of this type is collected by the LIDAR receiver and the variation of its intensity with time is used to extract range information.[3,15,24,70,71] Details are discussed in section 2.2.

While remote sensing by LIIRF may be impractical, a viable alternative is thermally induced infrared fluorescence (TIIRF) from hot molecules (e.g. in industrial smokestacks), as shown in the lower portion of Figure 1(c). TIIRF is measurable by passive IR laser heterodyne radiometry methods, using a laser as local oscillator tuned to a spectroscopic region that characterizes the molecules of interest.[31,32,50] However, Fourier-transform infrared (FTIR) spectroscopy[59] appears generally to be a more sensitive means of mid-IR radiometry.

Resonant absorption plays two key roles in IR LIDAR applications: in addition to being the source of the signal of interest, it also provides a background interference effect (along with scattering by aerosols) that attenuates the outgoing and return optical signals. Studies of atmospheric transmission therefore play an important part in optimizing and calibrating IR LIDAR measurements. There are atmospheric ''windows'' (relatively free of absorption by H_2O, CO_2 and other minor species) in various parts of the IR region, notably 1.0–1.4, 1.6–1.8, 4.3–5.3, and 7.8–13 μm.[15,23,24,59,72] An extensive compilation of atmospheric IR absorption data is available commercially in the form of the HITRAN database.[73] HITRAN simulations of atmospheric transmission match real conditions well (unless, of course, unanticipated pollutant molecules or particles are present).

Absorption coefficients for atmospheric IR LIDAR spectroscopy depend on several critical factors:[24] pressure (which influences collisional line broadening); temperature (which controls Doppler broadening of molecular lines, as well as thermal population of molecular energy levels); and the local density of energy levels (since adjacent transitions can cause ''spectral wing'' overlap effects). The pressure and temperature dependences of absorption coefficients need to be measured if

Table 1 Forms of optical interaction involved in laser-based atmospheric sensing. [Adapted from data of Killinger and Menyuk[33]]

Optical interaction	Spectral region	Scattering medium	Attenuation factor, $\kappa(m^{-1})$[a]	Effective reflectivity[b]
Resonant optical processes				
Absorption	UV	Atoms or	10^4	10^{-2}
	1 μm	molecules	10^3	10^{-3}–10^{-1}
	10 μm		10^2	10^{-7}–10^{-4}
Nonresonant optical processes				
Rayleigh scattering	UV	High-abundance	10^{-4}	10^{-3}
	1 μm	species (O_2, N_2)	10^{-6}	10^{-5}
	10 μm		10^{-10}	10^{-9}
Mie scattering	UV	Aerosols, clouds,	10^{-4}	10^{-3}
	1 μm	fine particles, etc.	10^{-4}	10^{-4}
	10 μm		10^{-5}	10^{-5}
Raman scattering	UV	Vibrating and	10^{-8}	10^{-7}
	1 μm	rotating molecules	10^{-10}	10^{-9}
	10 μm		10^{-14}	10^{-13}
Backscattering from topographic targets				
Retro-reflection	UV–IR	Distributed mirrors	–	~1
Other hard targets	UV–IR	Buildings, hillsides, trees, dense clouds, etc.	–	10^{-2}–1

[a] The atmospheric attenuation coefficient $\kappa(\lambda)$ is defined in section 2.1. The values listed here are for atmospheric pressure concentration.

[b] The effective reflectivity is calculated[33] for a range resolution $\Delta R = 100$ m, an atmospheric pressure concentration, and optical scattering assumed to be isotropic. The entries for absorption refer to backward-directed fluorescence. Those for nonresonant scattering processes are related to the volume backscattering coefficient (defined in section 2.1), and those for topographic targets refer to surface reflectivity of the target.

observations are made at other than ambient conditions. For example, altitude profiles of tropospheric O_3 can be resolved to ~5 km merely from the pressure dependence of its absorption coefficients at CO_2 laser wavelengths.[74]

Table 1 summarizes various optical interactions involved in laser-based sensing of the atmosphere, together with typical attenuation and efficiency factors.[33] These processes contribute to the atmospheric sensing strategies depicted in Figure 1(a–c).

2 THEORY AND OPERATING PRINCIPLES

The preliminary background in section 1 covers many of the constraints on IR LIDAR system design and its optimal mode of operation. We now present principles of both nonresonant, elastic-backscatter IR LIDAR (section 2.1) and IR DIAL (section 2.2). Details have been presented by Measures.[3,70] We merely give a simplified outline as in other reviews.[15,24,71] A serious IR LIDAR practitioner needs to dig deeper for refinements.

2.1 Nonresonant Infrared LIDAR

Adapting previously published treatments,[3,24,70] the LIDAR equation for atmospheric elastic backscatter can be presented in the simplified form shown in Equation (3):

$$\frac{P_r(R)}{P_L} = \frac{A}{R^2}\ \eta(\lambda_L) f(R)\ \frac{c\tau_L}{2} \times \beta(\lambda_L, R) \exp[-2\kappa(\lambda_L)R] \qquad (3)$$

where $P_r(R)$ and P_L are the optical powers (W) received by the detector and transmitted by the laser source, respectively. These two devices are assumed to be coaxially located. The factor $[A/R^2]$ is the solid angle (sr) of the collecting aperture, in terms of the receiving mirror area A (m^2) and the distance R (m) between the transmitter–receiver and the region of the atmosphere doing the backscattering. The factor $\eta(\lambda)$ allows for the optical efficiency of the receiver and detector as a function of wavelength λ, while $f(R)$ defines the overlap between the transmitted beam and the field of view of the receiver. The factor $[c\tau_L/2]$ represents the pathlength covered by the laser light, assumed to be a rectangular pulse of duration τ_L (s), at any instant; note that the time taken for the laser pulse to propagate at the speed of light c ($m\,s^{-1}$) to range R and back to the receiver is $[2R/c]$, which is usually much greater than τ_L.

The volume backscattering coefficient $\beta(\lambda, R)$ in Equation (3) is the fractional amount of incident energy backscattered per steradian per unit atmospheric pathlength; its units are therefore $m^{-1}\,sr^{-1}$. The coefficient

$\beta(\lambda, R)$ is usually the sum of many scattering contributions from atmospheric species j, each with R-dependent number density $N_j(R)$ (Equation 4):

$$\beta(\lambda_L, R) = \sum_j N_j(R) \left.\frac{d\sigma(\lambda_L)}{d\Omega}\right|_j \quad (4)$$

where $[d\sigma(\lambda_L)/d\Omega]_j$ is the electric DISC cross-section for species j at laser wavelength λ_L. Additional factors are needed when the scattering is inelastic (as in the Doppler LIDAR or Raman LIDAR cases briefly discussed in section 1.4), since the scattered wavelength λ differs from the laser wavelength λ_L. It is also assumed in deriving Equation (3) that lateral distribution effects across the area of the laser beam at range R are negligible, that is, the laser beam is assumed to have a ''top-hat'' spatial profile and the scattering medium is taken to be homogeneous within that profile.

The final exponential factor in Equation (3) allows for attenuation of the outgoing and returning light by a combination of absorption and scattering with a column-average atmospheric attenuation coefficient $\kappa(\lambda)$ (m^{-1}). The factor of 2 in the exponent [...] in Equation (3) arises from the two-way transmission to and from the target through the atmosphere. In the case of inelastic scattering over an inhomogeneous atmospheric column, the exponent [...] in Equation (3) needs to be replaced by a more elaborate form (Equation 5):

$$-2\kappa(\lambda_L)R = -\int_0^R [\kappa(\lambda_L, R') + \kappa(\lambda, R')]\, dR' \quad (5)$$

where the coefficients κ are now taken to be R-dependent (as would be the case in vertical atmospheric profile studies, for instance).

Equation (3) evaluates the instantaneous optical power $P_r(R)$ received by the detector, ratioed to the transmitted laser power P_L. A more pertinent quantity is the radiative energy $E_r(R)$ received by the detector within the response time τ_d of the detector (Equation 6):

$$\frac{E_r(R)}{E_L} = \frac{A}{R^2}\,\eta(\lambda_L) f(R)\,\frac{c\tau_d}{2} \times \beta(\lambda_L, R)\exp[-2\kappa(\lambda_L)R] \quad (6)$$

where the laser pulse energy E_L equals $(P_L\tau_L)$ in this alternative form of LIDAR equation. The effective range resolution for such a LIDAR system is limited to (Equation 7):

$$\Delta R = \frac{c(\tau_d + \tau_L)}{2} \quad (7)$$

Measures[3,70] presented various refinements of LIDAR Equations (3) and (6). It should be noted that some of the factors in Equations (3) and (6), namely P_L, E_L, A, $\eta(\lambda_L)$, $f(R)$, τ_L and τ_d, are properties of the LIDAR apparatus itself whereas others, namely $\beta(\lambda_L, R)$ and $\kappa(\lambda_L)$, are solely dependent on the atmosphere. The former comprise a LIDAR response function that needs to be calibrated carefully, as has already been mentioned in section 1.3, in the context of the CO_2 laser LIDAR.[40]

Grant[24] surveyed some typical LIDAR system parameters. The case of Mie scattering from aerosols in the lower atmosphere (with number densities in the range 10^8–$10^{11}\,m^{-3}$ and Mie scattering cross-sections up to $10^{-12}\,m^2\,sr^{-1}$) yields an upper limit of $0.1\,m^{-1}\,sr^{-1}$ for $\beta(\lambda_L, R)$. Using an Nd:YAG laser ($\lambda_L = 1.064\,\mu m$, pulse energy = 1 J, pulse duration = 10 ns, outgoing photons per pulse = 5×10^{18}) and typical LIDAR apparatus parameters with $\kappa = 0.11\,km^{-1}$, the calculated return signal from aerosols at a range of 1 km yields a readily detectable $P_r/P_t = 4 \times 10^{-10}$. Other orders of magnitude are summarized in Table 1.[33]

2.2 Differential Absorption LIDAR

DIAL measurements compare the atmospheric attenuation at a characteristic absorption wavelength, λ_{on}, of a certain molecule with that at a neighboring wavelength, λ_{off}, where the specific absorption is much lower. Absorption by the target molecule can then be separated from attentuation associated with other atmospheric species (molecules, aerosols, etc.).

Two arrangements for IR DIAL-type measurements are depicted in Figure 1(a) and (b). These are distinguished by the mode of returning the laser radiation to the optical receiver. One method – actually differential long-path laser absorption,[57,58] rather than DIAL[15,16,71] – uses a topographic target (e.g. the ground in the case of a downward-directed airborne laser) or a strategically located retro-reflector, as in Figure 1(b). This approach gains sensitivity by sacrificing range information and measuring only column content. The other (true) DIAL method, on which we now focus, uses elastic scattering from atmospheric aerosols and particulates and thereby preserves the range-dependent advantage of LIDAR methods.[15,68,69,71] This range-resolved form of IR DIAL is shown schematically in Figure 1(a), with two aerosol plumes shown but only one of two laser wavelengths marked explicitly.

Our treatment of IR DIAL is guided by that of others[3,15,24,70,71] and is based on the elastic scattering LIDAR Equation (3). Assuming that P_L and τ_L are identical at the two wavelengths λ_{on} and λ_{off}, we can evaluate the ratio of the received power signals by allowing several instrumental factors to cancel out and using the R-dependent form of attenuation coefficients κ that was introduced in the context of Equation (5). This

yields Equation (8):

$$\frac{P_r(\lambda_{on}, R)}{P_r(\lambda_{off}, R)} = \frac{\eta(\lambda_{on})}{\eta(\lambda_{off})}\frac{\beta(\lambda_{on}, R)}{\beta(\lambda_{off}, R)} \times \exp\left\{-2\int_0^R [\kappa(\lambda_{on}, R') - \kappa\lambda_{off}, R')]\,dR'\right\} \quad (8)$$

It is often acceptable to assume that the difference between $\kappa(\lambda_{on}, R)$ and $\kappa(\lambda_{off}, R)$ arises entirely from absorption of specific target molecules and that there is no significant change in other forms of optical attenuation (from aerosol scattering or other molecular absorption). Moreover, it is also often reasonable to assume that the ratios of instrumental factors $\eta(\lambda)$ and volume backscattering coefficients $\beta(\lambda, R)$ are equal to unity. With these approximations (which we shall adopt tentatively), the ratio of received powers is solely dependent on the difference in optical absorption of the target species at wavelengths λ_{on} and λ_{off}.

From the Beer–Lambert law, we can then simplify Equation (8) to the form Equation (9):

$$\frac{P_r(\lambda_{on}, R)}{P_r(\lambda_{off}, R)} \approx \exp\left\{-2\,[\sigma_{abs}(\lambda_{on}) - \sigma_{abs}(\lambda_{off})] \times \int_0^R N(R')\,dR'\right\} \quad (9)$$

where $\sigma_{abs}(\lambda)$ is the absorption cross-section (m^2 molecule^{-1}) of the molecular species of interest at wavelength λ and $N(R)$ is its number density (molecule m^{-3}) at range R. The molecular number density $\overline{N}_{\Delta R}$ averaged over increment $\Delta R = R_2 - R_1$ of the range can then be expressed in terms of the differential absorption coefficient $\Delta\sigma_{abs} = [\sigma_{abs}(\lambda_{on}) - \sigma_{abs}(\lambda_{off})]$ according to Equation (10):

$$\overline{N}_{\Delta R} \approx [2\Delta R\Delta\sigma_{abs}]^{-1} \times \left\{\ln\frac{P(\lambda_{on}, R_1)}{P(\lambda_{on}, R_2)} - \ln\frac{P(\lambda_{off}, R_1)}{P(\lambda_{off}, R_2)}\right\} \quad (10)$$

It is evident that four independent backscattered optical power measurements are needed for each determination of number density of the particular molecular species of interest. With appropriate signal averaging, the accuracy of determination of logarithmic factor [...] in Equation (10) can be typically 10% or less, but the accuracy of $\overline{N}_{\Delta R}$ itself within a particular range interval ΔR will depend on how well $\Delta\sigma_{abs}$ is known. Careful laboratory experiments are needed to determine $\Delta\sigma_{abs}$ for each molecule of interest, preferably with the same laser wavelengths, optical bandwidths, etc., as are used in actual field-based LIDAR measurements.

Differential long-path laser absorption, with topographic target or retro-reflector as in Figure 1(b), loses range resolution ΔR but gains in sensitivity and the range R for which measurements are feasible. A relationship comparable to Equation (9), with similar approximations, yields the integrated column number density N_{col} along the path of the laser beam [Equation 11]:

$$N_{col} \approx [2R_T\Delta\sigma_{abs}]^{-1} \ln\left\{\frac{[E(\lambda_{off}, R_T)]}{[E(\lambda_{on}, R_T)]}\right\} \quad (11)$$

where R_T is the range position of the topographic target or retro-reflector and $E(\lambda, R_T)$ are relative (not absolute) radiative energies received from the reflector at R_T by a detector whose integration period, τ_d, exceeds the laser pulse duration, τ_L. Equation (11) applies to a spectrally flat reflector (e.g. an ideal retro-reflector) but this is not always the case in the IR region (for instance, topographic targets often have different scattering efficiencies at different CO_2 laser wavelengths; it is then necessary to include additional corrections).

As in section 2.1, the reader seeking a more advanced theoretical treatment of DIAL and differential long-path laser absorption is referred to other sources.[3,70,71]

3 INFRARED LIDAR INSTRUMENTATION

A detailed description of operating LIDAR systems is beyond the scope of this article. This section merely outlines the essential elements of standard types of transmitter/receiver system for IR LIDAR measurements of the atmosphere. Many factors influencing instrument design have already been covered in sections 1 and 2, and details of laser sources suitable for IR LIDAR are surveyed in section 4. This article concentrates on IR LIDAR, but descriptions of visible and UV LIDAR systems in the literature remain useful in terms of general layout and data processing; various reviews[3,5,6,24,70,71] are helpful in this regard.

The basic elements of any LIDAR system are as follows: a laser transmitter operating at one or more wavelengths; an optical receiver with return wavelength discrimination and photodector(s); and electronics and data processor. Distinctive features of IR LIDAR systems (relative to those operating in visible and UV regions) are the following: the type of laser (see section 4); the need for IR-transparent optical materials for windows, partial reflectors, etc.; the virtually exclusive use of reflective optics rather than lenses; and the special IR photodetectors used (usually less sensitive than photomultipliers). There is no essential difference in broad terms between the electronics and data processing of an IR LIDAR system, and those for visible and UV LIDAR. A typical LIDAR system directs intense pulses of laser radiation towards atmospheric targets of interest, with a portion sampled to provide a zero-time electronic

trigger. Output optics can provide some form of beam collimation, spatial and/or spectral filtering and dichroic beam splitters to combine laser beams if more than one laser is used (as in DIAL).

The radiation that returns from the atmospheric target (as discussed in sections 1 and 2) is collected by a reflective telescope and usually passes through some form of wavelength discriminator (monochromator, polychromator, interferometer, and/or set of optical filters) on its way to the photodetector(s). Narrowband spectral filters are of little or no advantage in IR LIDAR for wavelengths longer than $\sim 3\,\mu m$, owing to thermal background noise (which can be limited by reducing the field of view of the detector using liquid-nitrogen-cooled baffles).

The transmitter and receiver can be located separately, but they are usually grouped together as in Figure 1(a) and (b), so that outgoing and return beams are virtually coaxial. Many LIDAR systems need to be mobile, either in a van or trailer or on an aircraft (or satellite), with compactness an essential design element. A stable platform is needed for the laser radiation and viewing direction of the receiving telescope to be steered towards the atmospheric region of interest, aided by a viewfinder and often with an associated video camera. Aligning the transmitter, receiver, and viewfinder of an IR LIDAR instrument is not a trivial task.

Photodetectors for IR LIDAR need to provide high sensitivity, low noise, and fast time response.[34,35,38,58] A limitation for IR LIDAR is that photomultipliers cannot access wavelengths longer than $\sim 1\,\mu m$. Beyond this limit, various high-performance IR detectors[75,76] are available but they generally fail to match the high gain and low noise that photomultipliers provide in the UV/VIS region, nor can they be used in a photon-counting configuration.

IR detectors can be classified as thermal detectors (including pyroelectric detectors), photovoltaic devices (photodiodes), and photoconductors.[75,76] The last two classes of detector are based on a wide range of semiconductor materials, which usually need liquid nitrogen (77 K) cooling and specially designed preamplifiers for optimal performance. There is usually a three-way performance trade-off between bandwidth (rise time), detector area, and detector noise. Popular cooled photoconductive detectors include those based on HgCdTe for the 5–13 μm region (which covers the wavelength range of CO_2 lasers) and on InSb for the near-infrared (NIR) region (1–5 μm). InGaAs photodiodes are well suited to the eye-safe IR region around 1.5 μm. Closer to the visible region, silicon avalanche photodiodes (APDs) (a solid-state counterpart of photomultipliers) are favored in some recent LIDAR applications.[77,78] Vidicons, diode arrays, and CCD (charge-coupled device) cameras have revolutionized spatially resolved imaging in the visible and UV regions. However, the availability of IR-sensitive imaging devices and APDs, based on materials such as Hg(Cd, Mn)Te,[75,76] is limited but nevertheless an enticing prospect for future IR remote-sensing applications.

Most LIDAR receivers rely on a transient digitizer to record the variation of signal strength as a function of time delay (which corresponds to range R in the LIDAR context). Such electronic devices and associated computer facilities have progressively improved in both performance and price. Likewise, another area of rapid growth is the computing power available to process raw LIDAR signals and for image formation and presentation of results.

The twin issues of accuracy and signal-to-noise ratio are crucial to reliable IR LIDAR measurements.[3,6,15,24,58,70,71,78,79] Aspects are briefly mentioned here, against the background of sections 1 and 2. Calibration of the LIDAR response function is especially important for single-wavelength LIDAR backscatter measurements, but still vital in DIAL systems. It is difficult to predict transmitter/receiver overlap functions, so that it is advisable to calibrate these empirically by means of a target of known reflectivity, positioned at several locations. Measurements at short range can be complicated by non-ideal overlap factors. In DIAL measurements, the two laser beams must be identically aligned and their spatial distributions and temporal profiles must coincide to avoid a serious source of systematic error. Changes in LIDAR alignment after calibration are also a persistent problem.

The installation of large high-power Nd : YAG, CO_2, and other lasers (with associated provision of electrical power, cooling, and gas or dye handling) in mobile IR LIDAR systems is nontrivial but feasible. For further insight into state-of-the-art LIDAR instrumentation, the reader is referred to several descriptions of airborne IR LIDAR facilities deployed by the US National Aeronautics and Space Administration (NASA).[48,77,80–83]

A notable recent accomplishment by NASA is the project LIDAR Atmospheric Sensing Experiment (LASE),[83] which was able to make IR DIAL measurements of water vapor, aerosol, and cloud profiles from a high-altitude (16–21 km), extended-range aircraft in September 1995.[81,82] LASE employed a $Ti^{3+} : Al_2O_3$ laser pumped by a double-pulsed Nd : YAG laser and injection-seeded by a diode laser to operate sequentially on and off resonance with absorption lines in the 813–819-nm band of H_2O vapor. The distribution of atmospheric water vapor is of fundamental importance to weather and climate, atmospheric radiation studies, the global hydrological cycle and atmospheric chemistry. It has been a key target for IR DIAL, in the pioneering work of Schotland[84] and in subsequent early studies[85,86] and ongoing developments.[80,82,87]

Such airborne LIDAR programs are seen as forerunners of space-based sensing. One instance is NASA's program LIDAR In-space Technology Experiment (LITE),[82,83–88] comprising a cloud and aerosol LIDAR system that was deployed in the space shuttle Discovery in September 1994. LITE employed an Nd : YAG laser to generate three wavelengths (1.064 μm, 532 nm, and 355 nm) and a 1-m diameter lightweight receiver telescope mirror; the system operated for 53 h, resulting in 40 GB of data covering 1.4×10^3 km of ground track. These data have provided a highly detailed global view of the vertical structure of clouds and aerosols, extending from the Earth's surface up to the middle stratosphere (~20 km), coupled to a series of concurrent airborne and ground-based LIDAR validation tests. Future plans include the USA–France collaborative project Pathfinder Instruments for Cloud and Aerosol Spaceborne Observations/Climatologie Etendue des Nuages et des Aerosols (PICASSO/CENA),[83] a satellite-borne Nd : YAG LIDAR emitting at 1.064 μm and 532 nm for polarization-sensitive aerosol and cloud measurements; it is scheduled to be launched in early 2003 into a sun-synchronous 705-km circular orbit. This will perform active LIDAR measurements, flying in formation with a second satellite that carries a suite of passive optical and spectroscopic instruments. The technological feasibility of putting an 815-nm water-vapor IR DIAL instrument into space has been discussed.[82,87]

4 COHERENT LIGHT SOURCES FOR INFRARED LIDAR APPLICATIONS

This section surveys lasers and nonlinear-optical instruments suited to IR LIDAR applications. Pulsed coherent IR sources (with attributes needed for range-resolved sensing) are featured, but mention is also made of shorter wavelength coherent sources such as tunable dye lasers (since dye laser radiation can be nonlinear-optically downconverted into the mid-IR region and some dyes operate in the NIR region). Likewise, continuous-wave (CW) IR lasers are not excluded, as they are useful for optical heterodyne Doppler velocimetry and long-path laser absorption. Moreover, narrowband CW devices, such as a semiconductor tunable diode laser (TDL), can act either as local oscillators in laser heterodyne systems or as wavelength-control devices in higher-power coherent sources entailing pulse amplification or injection seeding.

Eye safety[89] is an important issue in designing lasers for atmospheric sensing,[6] notably with regard to high-power pulsed solid-state lasers operating in the UV, visible or NIR regions.[14] Wavelengths longer than 1.4 μm are considered to be eye-safe,[90] which partly accounts for the acceptability of CO_2 lasers (9–11 μm) as sources for IR LIDAR. There is much interest in high-power pulsed coherent sources in the vicinity of 1.5–2.5 μm, either by nonlinear-optical means[80,90] or by using new solid-state laser materials.[14,49,91–93]

4.1 Fixed-wavelength Lasers

Fixed-wavelength pulsed solid-state IR lasers have long been central to LIDAR instruments, such as those recording Mie scattering from aerosols. Ruby lasers,[8,10,14,94] flashlamp-pumped devices operating at ~694 nm, were used almost exclusively in pathfinding applications of the mid-1960s.[6] Schotland's original DIAL studies of atmospheric water vapor[84] used a temperature-tuned ruby laser. Laser-grade ruby ($Cr^{3+} : Al_2O_3$) comprises sapphire (Al_2O_3) in which a small fraction of Al^{3+} ions are replaced ("doped") by Cr^{3+} ions. Its tuning curve in the range 200–300 K is given by Equation (12):[14]

$$\lambda(T)(\text{nm}) = 694.325 + 0.068(T - 293) \qquad (12)$$

This enables a ruby laser's wavelength to be varied from 694.3 nm (at ~300 K, with line width ~11 cm^{-1}) to 693.4 nm (cooled by liquid nitrogen at 77 K, with line width ~0.15 cm^{-1}).

Ruby lasers have been largely superseded in most applications by other forms of solid-state IR laser, of which the most commonly used is the Nd : YAG laser,[14] first demonstrated in 1964.[95] The Nd : YAG gain medium comprises a Nd^{3+}-doped YAG ($Y_3Al_5O_{12}$) crystal lattice with ~1% of the Y^{3+} sites occupied by neodymium ions (Nd^{3+}). The host material has good optical, thermal, and mechanical properties and its structure favors a narrow fluorescent line width, yielding high laser gain and low operational threshold. The fundamental output wavelength from a Nd : YAG laser is 1.06414 μm, which is directly usable for some IR LIDAR applications. Its harmonics at 532, 355 and 266 nm are commonly used in UV/VIS LIDAR and in pumping devices such as dye lasers and OPOs.

A typical pulsed Nd : YAG laser oscillator[14] uses cylindrical Nd : YAG rods with polished, antireflection (AR) coated end faces, enclosed in a water-cooled reflective flashlamp housing (e.g. with an elliptical reflector coupling the Nd : YAG rod to a linear xenon flashlamp at its respective foci). These flashlamp-pumped Nd : YAG rod modules are located in a laser cavity, often with mirrors chosen for unstable optical operation (to extract maximum power from the laser rod, and to tailor the laser beam to a diffraction-limited, quasi-Gaussian profile).[96] The Nd : YAG laser cavity also usually incorporates a Q-switch (for regular pulsed operation) and optical elements such as waveplates and polarizers

(to compensate for thermal birefringence, refine beam quality and couple out a high-quality pulsed laser beam). An amplifier stage, using further flashlamp-pumped Nd : YAG rod modules, is often added to boost the output power. Such Nd : YAG laser oscillator–amplifier systems are commercially available, with routinely available operating characteristics as follows: pulse repetition rate = 10 Hz, pulse duration = 5–10 ns, pulse energy = 1 J, and peak power ≥100 MW.

Nd : YAG lasers typically run on several longitudinal modes, yielding an optical bandwidth of $\sim 1\,cm^{-1}$ and an uneven temporal profile owing to beating between those modes. Single longitudinal mode (SLM) operation of a pulsed Nd : YAG laser can be achieved by injection seeding[97,98] with a narrowband 1.064-μm laser source (such as a single-mode diode laser or a miniature monolithic Nd : YAG laser). This yields a smooth temporal profile (free of mode-beating effects) and an optical bandwidth of ~50 MHz, close to the limit imposed by the Fourier transform of the pulse duration), in a high-quality, diffraction-limited laser beam. Injection-seeded Nd : YAG lasers have several advantages, all arising from SLM operation and compensating for its complexity: transform-limited optical bandwidth, smooth temporal profile and higher nonlinear-optical conversion efficiency.

The last decade has seen a dramatic growth in the availability, affordability and utility of laser-diode arrays as an alternative to flashlamp excitation of Nd : YAG and other pulsed solid-state lasers.[14] All-solid-state laser systems pumped by laser-diode arrays offer higher overall efficiency relative to flashlamp-pumped systems, since the laser-diode emission can give a closer spectral match to the absorption band of a solid-state gain medium such as Nd : YAG. This in turn reduces the thermal load on the gain medium, yielding improved laser beam quality and much-increased pulse repetition frequency (e.g. diode-pumped Nd : YAG lasers can readily be run at ≥1 kHz, compared with ~10 Hz for typical flashlamp-pumped systems). The lower voltage requirements and compactness of diode laser pump power supplies provides a further significant advantage, especially for field-based applications such as LIDAR. The cost of laser-diode arrays is still relatively high (but decreasing steadily), although partially offset by their longer component lifetimes. Design parameters for all-solid-state Nd : YAG and other lasers are well defined[14] and numerous systems are commercially available. They will certainly find widespread use in future LIDAR instruments,[99] although their implementation may be slowed by the inertia of replacing flashlamp-pumped Nd : YAG lasers installed in existing mobile and airborne LIDAR facilities.

Nd : YAG lasers may be regarded as the industry standard against which other pulsed solid-state lasers are gauged.[14] On the one hand, other materials can serve as alternative hosts for Nd^{3+}, notably GSGG (gadolinium scandium aluminum garnet; $Gd_3Sc_2Al_3O_{12}$) doped with Nd^{3+} and Cr^{3+} gives Nd : Cr : GSGG lasers a better spectral match than Nd : YAG to flashlamp emission; glasses are used in very high-power Nd : glass laser amplifiers, although these are subject to poorer lasing thresholds, thermal properties, and optical distortion effects than Nd : YAG; YLF (yttrium lithium fluoride; $YLiF_4$) is transparent out to 1.5 μm, is less prone than YAG to thermal lensing and birefringence, and has higher energy storage capacity (owing to its lower gain). Mode-locked and Q-switched Nd : YLF laser oscillators (emitting with orthogonal polarizations at 1.053 and 1.047 μm) closely match the 1.054-μm wavelength of large phosphate glass Nd^{3+} laser amplifiers regularly used in fusion research. Diode-pumped Nd : YLF lasers have been used in IR LIDAR measurements of the atmospheric boundary layer, cirrus clouds, and volcanic aerosols.[40]

Variation of transition metal ion as well as host material provides flexibility and diversity in pulsed solid-state lasers.[14] This is particularly advantageous in extending IR LIDAR applications into the IR region beyond 1.06 μm. Pulsed IR lasers based on erbium ions (Er^{3+}) are of considerable interest: Er : glass lasers emit at ~1.54 μm, in the eye-safe region used for optical telecommunications, and are crucial as fiber-optic amplifiers; Er : YAG lasers emit at ~2.94 μm, which coincides with the fundamental IR absorption band of water and is amenable to atmospheric sensing of H_2O by IR LIDAR. Pulsed IR lasers based on holmium (Ho^{3+}) and thulium (Tm^{3+}) ions emit in the 2-μm region: for instance, Tm : YAG and Tm : YLF emit at ~2.01 μm. Diode-pumped Tm : Ho : YAG and flashlamp-pumped Cr : Tm : Ho : YAG lasers yield output at ~2.08 μm and are useful for wind sensing.[6,49,91,92] A CW 2.066-mm Ho : YLF laser has been developed recently for remote sensing of CO_2 and H_2O.[93]

Many forms of "fixed-wavelength" laser can run at a variety of wavelengths other than that with maximum gain, by using wavelength-selective reflectors or intracavity elements. For instance, 1.06414 μm is not the only possible emission wavelength of a Nd : YAG laser; others (in descending order of intensity at ~300 K) include 1.0615, 1.0738, 1.0646, 1.1121, 1.05205, 1.1159, 1.1227, 1.0780, and 1.3188 μm.[14,100]

The same is true of some gas lasers,[101] such as the helium–neon (He–Ne) laser.[10,102] The ubiquitous red He–Ne laser line at 632.8 nm is not in fact that with the highest gain. Many NIR He–Ne laser lines are accessible, notably those at 1.1526 μm (the first gas laser emission line ever observed), 2.3951 μm, and 3.3913 μm.[101]

Sensing with fixed-wavelength lasers makes the most of chance coincidences between discrete laser wavelengths and spectroscopic transitions in molecules of

interest. Fixed-wavelength lasers have the advantage of ''auto-calibration'' of detection wavelength and tend to be more robust, compact and/or easier to operate than line- or continuously tunable sources (such as dye lasers or OPOs). In one instance, an étalon-scanned Nd:YAG laser line at ~1.34 μm was used to monitor CH_4 via the P(4) feature in its $(\nu_2 + 2\nu_3)$ combination band.[103]

Laser-spectroscopic monitoring of hydrogen fluoride (HF) gas in aluminum smelters has been performed by long-path absorption of IR He–Ne laser radiation at ~2.4 μm.[104,105] This HF sensor employs two He–Ne lasers: one operates at $\lambda_{on} = 2.3958\,\mu m$, coinciding with the R(5) fundamental absorption line of HF; the other is at $\lambda_{off} = 2.3985\,\mu m$, providing an interference-free reference. In proving trials,[105] the two CW laser beams were transmitted along the ceiling of the potroom of an aluminum smelter to retro-reflectors at a range of up to 400 m. The He–Ne laser sensor is able to detect changes of airborne HF concentration as low as $0.01\,mg\,m^{-3}$ (0.05 ppm) at a range of 270 m; its response time is routinely set at 15 s – much more rapid than conventional wet-chemical HF sensors.[106] It is thereby feasible to monitor abrupt changes of HF emission inside the aluminum smelter potroom, arising from regular operations such as anode changing, crust breaking, and metal tapping.

4.2 Line-tunable Lasers

Several forms of molecular gas laser are line-tunable, able to operate on successive vibration–rotation emission transitions by simple cavity adjustments. The most notable are CO_2 lasers,[18–21] which operate in the wavelength range 9–11.5 μm with high wallplug efficiency (~20%) and have long been vital to atmospheric sensing by IR LIDAR.

A typical CW CO_2 laser comprises a water-cooled tube with Brewster-angle IR-transparent windows, through which a low-pressure gas mixture of CO_2, N_2, and He flows, supporting a longitudinal glow discharge excited electrically. The tube is placed in an IR-resonant optical cavity with one or more wavelength-selective elements such as a Littrow-mounted diffraction grating, a prism, or an étalon; adjustment of this element allows successive laser lines to be selected. The emission spectrum of a regular CO_2 laser comprises the P and R branches of two rotation–vibration bands of $^{12}C^{16}O_2$, centered at 10.4 and 9.4 μm, respectively; without wavelength selection, lasing is on the highest gain transition, 10.59-μm P(20).

The number of available CO_2-laser wavelengths can be increased markedly[21] by operating on sequence bands (with an intracavity absorption cell to suppress regular laser lines and enhance hot-band emission). Isotopic modifications of CO_2 such as $^{13}CO_2$ or $C^{18}O_2$ can also be used in sealed laser discharge tubes (with Xe added to minimize dissociation of CO_2); the 9.1145-μm R(20) line of a $C^{18}O_2$ laser is an agreed standard for atmospheric IR LIDAR[6,40,49] since it has high gain and is far from absorption lines of natural-abundance CO_2 in the air.

Not all CO_2 lasers are CW. Most forms of IR LIDAR need CO_2-laser radiation that is emitted in pulses, rather than continuously. This can be attained by incorporating a Q-switch in the laser cavity[18,20] or by atmospheric-pressure operation with a pulsed electrical power supply in the transversely excited (TE) mode. CO_2 TE lasers can also be operated at higher pressures up to ~10 bar, where they become continuously tunable as a result of pressure broadening.[19] Limited ranges of continuous tunability can also be obtained from waveguide CO_2 lasers, which are typically excited by radiofrequency at atmospheric pressure and are relatively compact.[11,74]

Many molecules have functional groups that cause their IR absorption spectra to fall within the wavelength range (9–11 μm) of line-tunable CO_2 lasers.[52,107] This makes CO_2 lasers useful for many DIAL and long-path absorption applications, of which tropospheric ozone detection at ~9.6 μm has already been mentioned in section 1.5.[16] In an early application of a continuously tunable CO_2 waveguide laser operating in the vicinity of 9.5–9.6 μm, Menzies studied the variation of atmospheric ozone line widths and absorption coefficients as a function of pressure and altitude.[74] There are numerous examples of line-tunable CO_2-laser DIAL systems used to map atmospheric pollutants.[15,16,71]

A recently reported IR DIAL technique[108] employed CO_2 lasers for quantitative multiwavelength detection of four gas-phase molecules that are of particular environmental interest: CH_3CCl_3, C_2HCl_3, $CFCl_2CF_2Cl$ (Freon 113), and C_2H_4. Their respective absorption characteristics at each of 40 CO_2-laser wavelengths were established in laboratory measurements with a 16-W CW CO_2 laser; trailer-borne, range-resolved field trials were then conducted with a pair of pulsed TE CO_2 lasers to identify and quantify mixtures of the four gases in controlled releases at ranges of ≥3 km, aided by chemometric data-analysis methods.

Another promising recent CO_2-laser application, known as BAGI (backscatter absorption gas imaging),[109] uses a CW CO_2 laser (typically 20 W, able to image gases at ranges up to 125 m) to produce rastered gas-plume images of IR absorption by backscattering from hard targets. The BAGI database covers ~80 detectable gases, most of them significant industrial pollutants. In a BAGI variant, a 3.3966-μm He–Ne laser is able to monitor methane.[109]

Other recent papers[49,110–113] demonstrate the ongoing utility of CO_2 lasers for a variety of IR LIDAR applications. Such papers confirm that careful, systematic

calibration of various CO_2-laser LIDAR parameters, particularly backscatter coefficients from hard targets[110] and aerosols,[111,113] is no less critical than in earlier work.[16] For example, DIAL-type sensing by backscattering of two CO_2-laser wavelengths (λ_{on} and λ_{off}) from atmospheric aerosols must incorporate variability with wavelength λ_L of volume backscattering coefficients, $\beta(\lambda_L, R)$ in Equation (4).[16,113] Corrections also need to be made for potential interferences from the many atmospheric species that absorb in the CO_2-laser region (9–11 μm)[16,97,98] and for strong (altitude-related) temperature and pressure dependences of absorption cross-sections, $\sigma_{abs}(\lambda_L)$, and other optical parameters.[57,74,111,112]

Other useful forms of high-power line-tunable molecular IR laser[101] include discharge-excited lasers based on carbon monoxide (CO) (emitting in the fundamental range 4.75–8.25 μm and in the overtone range 2.6–4.0 μm),[114] nitrous oxide (N_2O) (10.3–11.0 μm), carbon disulfide (CS_2) (11.5–12.3 μm),[19] and carbon oxysulfide (COS) (8.2–8.4 μm).[19] Also significant are various forms of hydrogen halide chemical laser,[115] based on HF, HCl, HBr and their deuterated analogs, covering most of the IR range between 2.4 and 6.3 μm.[101]

4.3 Continuously Tunable Lasers

Continuously tunable lasers[11–14] offer flexibility in LIDAR applications, such as DIAL, which require coherent radiation that is adjustable either on or off resonance with molecules of interest in the atmosphere. Fixed-wavelength or line-tunable lasers rely on chance spectroscopic coincidences. Continuously tunable lasers have no such limitation, but they generally require some form of active wavelength control and calibration. This section reviews several forms of pulsed, continuously tunable laser that are applicable to IR LIDAR applications, namely tunable solid-state lasers,[14,94,116] notably titanium–sapphire ($Ti^{3+}:Al_2O_3$) and alexandrite ($Cr^{3+}:BeAl_2O_4$), TDLs,[34,117–121] and semiconductor TDL sources.[122]

Some optically pumped solid-state lasers are confined to one or more discrete wavelengths (as treated in sections 4.1 and 4.2), but others have spectrally broad gain profiles and can be operated as continuously tunable coherent sources via wavelength-selective intracavity elements or injection seeding. One such device is the titanium–sapphire laser ($Ti^{3+}:Al_2O_3$), which operates over the wavelength range 0.7–1.05 μm, either CW (pumped by an argon-ion laser) or pulsed (pumped by an Nd:YAG or Nd:YLF laser).[14,83,116,123] Such a laser has been used in a multiwavelength LIDAR system to measure aerosol size distributions.[40,47] A double-pulsed TDL-seeded $Ti^{3+}:Al_2O_3$ laser operating around 820 nm has been incorporated in airborne water-vapor IR DIAL systems used in the context of project LASE[81–83] and its predecessors.[86,87]

Alexandrite lasers[14,94,116,124] are also important solid-state sources for airborne water-vapor NIR DIAL measurements.[16,77,78,125] Alexandrite, a flashlamp-pumped laser material in which Cr^{3+} ions are doped in a chrysoberyl ($BeAl_2O_4$) lattice, avoids some of the "three-level-laser" difficulties of ruby. Its room-temperature tuning range is 700–820 nm. Commercially available alexandrite lasers typically deliver Q-switched output pulses of ≥100 mJ, continuously tunable (by means of intracavity étalons and birefringent or Lyot filters) in sweeps of ~150 cm^{-1}, with optical bandwidths of ~1 cm^{-1}.

Injection seeding by a narrowband laser (e.g. a CW TDL or another narrowband, low-power laser) provides a useful wavelength-control strategy for either pulsed $Ti^{3+}:Al_2O_3$[81,87,126–128] or alexandrite[78,129–132] lasers. Relevant applications of such lasers are aimed at IR DIAL sensing of humidity via the third-overtone ($4\nu_{OH}$ at ~730 and ~820 nm) and second-overtone ($3\nu_{OH}$ at ~940 nm) absorption bands of H_2O vapor[78,79,81,82,87,129–132] and of gas density and temperature via the 760-nm magnetic-dipole vibronic band of O_2.[79,129,132] For reliable IR DIAL measurements such as these, spectral purity (i.e. the narrowband proportion of laser output energy) needs to be as high as possible.[77,79,87,125] and has been optimized to at least 99.9% in the various injection-seeded $Ti^{3+}:Al_2O_3$[81,82,126,127] and alexandrite[78,129–132] lasers used. Such lasers, when used in the "λ_{on}" mode of IR DIAL, also generally need some form of active wavelength stabilization and locking to the exact peak of the spectroscopic feature of interest.

Ground-based IR LIDAR has been used recently[133] to detect LIF from potassium atoms in meteor trails blown by winds into the path of vertically directed 769.9-nm radiation (resonant with the D_1 fine-structure line of K) from a tunable TDL-seeded alexandrite laser.[134] Other recent developments include a tunable, narrowband, CW 2.066-μm Ho:YLF laser, injection-seeded by a $Ti^{3+}:Al_2O_3$ laser;[93] this has been evaluated for remote sensing of atmospheric CO_2 and H_2O.

Tunable dye lasers[117–121] use optically pumped fluorescent organic dyes in solution and are vital to many forms of laser spectroscopy.[34–37] Continuous narrowband tunability is achieved by resonating the fluorescent emission in an optical cavity incorporating a diffraction grating and/or étalon). Pulsed dye lasers, pumped by flashlamps or another laser, are useful for UV and visible LIDAR applications, particularly laboratory trials when the handling of dyes and flammable solvents can be tolerated. However, their direct utility as pulsed, coherent sources for IR LIDAR applications is less evident. NIR dyes extend to ~0.95 μm, but solid-state lasers such as

$Ti^{3+}:Al_2O_3$ or alexandrite are preferred in mobile IR LIDAR facilities. NIR dye-laser DIAL systems have been used for ground-based[85] and airborne[135] remote sensing of tropospheric water vapor at ~725 nm. Tunable dye lasers occasionally have a secondary role to play in the area of IR LIDAR, in that visible dye laser radiation can be downconverted into the IR region by nonlinear-optical means, such as coherent difference-frequency generation (DFG)[100] or by injection-seeding a pulsed OPO[22] (see section 4.4).

Semiconductor TDLs[58,121,122,136] are not only vital to communications technology, but are also key devices in laser spectroscopy[34–37,137,138] and atmospheric sensing.[30,31,57,58,139] The spectral operating range of a TDL depends on its band gap, according to the composition of the semiconductor; a few representative materials and their wavelength ranges are $GaAs_{1-x}P_x$, 0.65–0.9 μm; $In_{1-x}Ga_xAs$, 0.9–2 μm; $PbS_{1-x}Se_x$, 4–8.5 μm; $Pb_{1-x}Sn_x$ (Se, Te), 8.5–<15 μm. The precise TDL output wavelength is varied by adjusting the diode's operating temperature T or its applied current i.

Much spectroscopy and atmospheric sensing[30,31,58,137,139–141] has been performed with cryogenically cooled, (i, T)-tuned TDLs, usually with an auxiliary monochromator for mode isolation. In the NIR region, continuous single-mode tunability over wide ranges is more readily achievable by means of a room-temperature external-cavity diode laser (ECDL),[121,122] in which the TDL is coupled to a grating in a wavelength-selective external optical cavity and tuned by simultaneously varying diffraction angle and cavity length; the advantages of such ECDLs for spectroscopic sensing are clearly established,[138,142–144] with extremely narrow optical bandwidth ($<10^{-4}\,cm^{-1}$), high stability, durability, and compactness. Other TDL wavelength-control methods include distributed feedback (DFB), distributed Bragg reflector (DBR), and vertical-cavity surface-emitting laser (VCSEL) geometries.[122]

CW operation and low output intensity generally prohibit the direct use of TDLs as coherent sources for range-resolved IR LIDAR measurements. However, they are useful for long-path IR laser absorption studies of atmospheric species, with either a multipass absorption cell or a remote-sensing retro-reflector arrangement as in Figure 1(b). TDLs have other indirect applications to IR LIDAR, as local oscillators for heterodyne detection[50] or injection seeders for higher-power tunable lasers[81,88,126–128,133,134] or OPOs[22] (see section 4.4 for details).

Other applications of TDLs to atmospheric monitoring entail nonlinear-optical difference-frequency mixing[100] of two room-temperature NIR TDLs to generate mid-IR frequencies more readily than can be done with a single conventional cryogenically cooled lead salt TDL. Examples of such TDL DFG methods include long-path absorption studies of atmospheric pollutants such as SO_2[145] and formaldehyde (H_2CO)[146] (see section 4.4 for details).

4.4 Nonlinear-optical Coherent Infrared Sources

Nonlinear optics[36–38,56,100] involves the interaction of a number (σ, >2) of optical waves in a medium with a nonlinear-optical susceptibility tensor $\chi^{(\sigma-1)}$. These waves (each with a frequency $\nu_1, \nu_2, \ldots, \nu_\sigma$) obey two conservation conditions, one for energy [Equation 13]:

$$\nu_1 + \nu_2 + \cdots + \nu_\sigma = 0 \tag{13}$$

and the other (expressed in terms of wave vectors $\boldsymbol{k}_j$, with $j = 1, 2, \ldots, \sigma$ with magnitudes $k_j = 2\pi n_j \nu_j / c = 2\pi n_j / \lambda_j$, where n_j is the refractive index at vacuum wavelength λ_j and c is the speed of light) for momentum [Equation 14]:

$$\boldsymbol{k}_1 + \boldsymbol{k}_2 + \cdots + \boldsymbol{k}_\sigma + \Delta\boldsymbol{k} = 0 \tag{14}$$

where $\Delta\boldsymbol{k}$ is a phase-mismatch increment that must be minimized to optimise the efficiency of the nonlinear-optical process of interest. Each frequency component has a positive or negative sign associated with phase relationships. We now consider three nonlinear-optical processes that can generate coherent IR radiation suitable for atmospheric sensing.

One such nonlinear-optical process is stimulated Raman scattering (SRS), the first-Stokes version of which is useful as a passive (i.e. adjustment-free) means of converting high-power pulsed laser radiation at frequency ν_L to a lower IR frequency ν_R determined by the Raman shift $\Delta\nu_R$, as in Equation (2), of a suitable Raman-active medium. First-Stokes SRS from isotropic fluids is a coherent four-wave process ($\sigma = 4$) with $|\nu_1| = |\nu_2| = \nu_L$, and $|\nu_3| = |\nu_4| = \nu_R$ which is nonparametric and automatically phase-matched ($\Delta k = 0$) in the forward-scattered direction. Commonly used Raman-shifting media are high-pressure gases such as H_2 ($\Delta\nu_R/c = 4155\,cm^{-1}$), D_2 ($2987\,cm^{-1}$), and CH_4 ($2917\,cm^{-1}$).[34–37] For instance, pulsed Nd:YAG laser radiation (1.064 μm) can be Raman-shifted in high-pressure CH_4 to an eye-safe wavelength (1.543 μm); this approach has proved convenient in a number of IR LIDAR instruments.[80,90] Likewise, SRS is a way to generate longer wavelength tunable radiation for spectroscopic sensing of the atmosphere, by downconverting continuously tunable sources such as dye lasers.[147] The tuning range (0.70–0.82 μm) of an alexandrite laser can be extended by first-Stokes SRS in H_2, D_2, and CH_4 to cover most of the range 0.9–1.2 μm[101,111] (and beyond, using higher order Stokes and anti-Stokes processes[34–37]).

Another form of nonlinear-optical wavelength conversion involves coherent three-wave processes ($\sigma = 3$) in noncentrosymmetric solid-state media,[100] including the second-harmonic ($\nu_3 = 2|\nu_1|$) and sum-frequency ($\nu_3 = |\nu_1 + \nu_2|$) generation processes widely used to upconvert pulsed laser radiation. In the IR LIDAR context, two such processes are important. One case is coherent DFG, with two input laser frequencies $|\nu_1| = \nu_L$ and $|\nu_2| = \nu'_L$ ($<\nu_L$), an IR output frequency $\nu_{IR} = |(\nu_L - \nu'_L)|$, and a phase mismatch $\Delta k = |\boldsymbol{k}_{IR} - \boldsymbol{k}_L + \boldsymbol{k}'_L|$. The other is that associated with an OPO or optical parametric amplifier (OPA).[22,148–154] In this case, a single laser input wave (''pump'', frequency ν_P) and two coherent output waves (''signal'', frequency ν_S; ''idler'', frequency ν_I; $\nu_S \geq \nu_I$) obey both energy conservation ($\nu_P = \nu_S + \nu_I$) and phase-matching ($\boldsymbol{k}_P - \boldsymbol{k}_S - \boldsymbol{k}_I - \Delta\boldsymbol{k} = 0$) conditions, as in Equations (13) and (14).

The close complementarity between DFG and an OPO should be noted.[151] In fact, the cases of intracavity DFG and injection-seeded OPOs (with $\nu_{seed} = \nu_S$ or ν_I) are distinguished only by the relative amplitudes A of the waves at ν_L and ν'_L (A_L typically comparable to A'_L) in the former case and those at ν_P and ν_{seed} ($A_P \gg A_{seed}$) in the latter.

In both DFG and OPOs, phase matching is needed to minimise $\Delta k = |\Delta\boldsymbol{k}|$. Traditionally, this is done by adjusting the angle and/or temperature of a birefringent nonlinear-optical crystal via its ordinary- and extraordinary-ray refractive indices.[100] A more recent alternative is to use quasi-phase-matched (QPM) media tailored for specific wavelength sets by periodic structuring; periodically poled lithium niobate (PPLN) is a prominent example.[155]

DFG in bulk media such as lithium niobate ($LiNbO_3$) has long been a dependable source of narrowband, high-quality tunable IR radiation, as illustrated by CW IR spectroscopic studies of CH_4 and other molecules by Pine.[156] Recent progress has been made in developing compact tunable CW mid-IR sources based on DFG in birefringently phase-matched media such as $AgGaSe_2$ (generating 8.7-μm radiation as the difference of 1.554 and 1.319 μm, for sensing of SO_2)[145] and in QPM media such as PPLN (by DFG mixing of TDLs in the 0.8–1.6 μm range, yielding 3.5-μm radiation to monitor trace gases such as H_2O, H_2CO, and CH_4).[145,146,157,158] Such advances combine the wavelength-extension advantage of nonlinear optics with photonics technology (e.g. TDLs, optical fibers, and fiber amplifiers); they have contributed to NASA's Lunar–Mars Life Support Test program.[146]

Relative to DFG, OPOs offer the prospect of high output power and coherence, both of which are advantageous for remote sensing of the atmosphere. Of the many possible varieties of OPO, we concentrate discussion on those with 2–10-ns pulse duration, which is compatible with the range-resolution requirements of many IR LIDAR and DIAL applications.

A simple pulsed OPO uses a suitable nonlinear-optical crystal (either birefringently phase-matched or QPM) in a two-mirror optical cavity that is resonant at either λ_S or λ_I. It is pumped at λ_P by a pulsed monochromatic coherent source (e.g. harmonics of an Nd : YAG laser). Such a free-running OPO (with no wavelength-selective elements) has a broad optical bandwidth ($\sim 10\,cm^{-1}$ or more), depending on several factors: the refractivity, dispersion, and absorption of the OPO medium; the wavelengths λ_S, λ_I, and λ_P; the type of phase matching; the crystal dimensions; the cavity reflectivity and effective number of passes of the resonated wave; and the bandwidth, divergence, pulse duration and pulse energy of the pump radiation.

Narrowband operation of a pulsed OPO can be achieved by means of wavelength-selective elements, such as intracavity gratings and étalons. This was realised before 1980 by Byer and colleagues,[22,149–151,159] whose pulsed $LiNbO_3$ OPOs provided a practical narrowband pulsed spectroscopic source, continuously tunable under computer control over wide IR ranges with an optical bandwidth of $\sim 0.1\,cm^{-1}$. This form of tunable OPO system was successfully applied in a number of atmospheric remote-sensing demonstrations, applied to CO (at 2.3 μm and a range of >100 m),[160] SO_2 (at 4.0 μm and 120-m range),[161,162] CH_4 (at 1.66 μm and over a 2.7-km column),[162] and H_2O [around 1.75 μm, both for a 2-km atmospheric column[163] and range-resolved up to ~ 1 km[164]].

After that early (pre-1982) progress on pulsed OPOs, relatively few OPO-based advances in IR LIDAR appear to have been made during the next 15 years. This is probably because such tunable pulsed OPOs proved difficult to operate and damage prone (largely because intracavity losses from gratings and étalons caused the operating threshold to approach the damage threshold of materials such as $LiNbO_3$). A revival in OPO technology in the last decade[22,152–154] is attributable to new nonlinear-optical materials[100] such as BBO (β-barium borate) and KTP (potassium titanyl phosphate), combined with high-performance pump lasers[14,97,98] and advanced IR OPO system designs.[165] OPOs now appear to have regained acceptance as high-power pulsed tunable sources suitable for IR LIDAR applications.

One particularly successful approach to controlling the wavelengths and optical bandwidth of a pulsed OPO tuning is injection seeding[22,152,154,166–173] with narrowband, tunable radiation from a low-power, narrowband external source. This simplifies OPO construction, by separating its wavelength-control function from that of power amplification.

A TDL-seeded $LiNbO_3$ OPO has been used by Milton et al.[169] to demonstrate range-resolved 3.4-μm IR DIAL measurements of atmospheric CH_4 at ranges up to 0.5 km. Single-mode TDLs have been exploited in the author's laboratory for continuously tunable injection seeding of pulsed IR OPOs of $LiNbO_3$ in both its bulk[170] and QPM PPLN[171,172] forms; optical bandwidths are as low as $\sim 0.005\,cm^{-1}$, approaching the Fourier-transform limit of the pulsed radiation.[10] Their performance has been verified by recording OPO-scanned spectra of CH_4, both as a gas (Doppler-broadened) and in a supersonic free jet (sub-Doppler).

Another significant TDL-seeded OPO system designed for airborne water-vapor IR DIAL measurements is that of Ehret et al.;[173,174] it comprises a BBO or KTP ring-cavity OPO pumped at 532 nm and TDL-seeded at its 942-nm signal wavelength. This system is able to monitor the second-overtone absorption band ($3\nu_{OH}$) of H_2O, which offers higher sensitivity than the third-overtone band ($4\nu_{OH}$) region at ~725 nm accessed in earlier airborne dye-laser-based IR DIAL studies by the same group.[135]

Finally, we note an intrinsic advantage for IR-spectroscopic sensing that is shared by both DFG and OPOs, namely coherent downconversion of available pump-laser radiation to longer wavelengths.[151] It is now relatively routine to be able to tune OPOs throughout the NIR region to beyond 4 μm. Even longer OPO output wavelengths [as far out as 12 μm[22]] can be accessed by using lasers with wavelengths longer than 1.064 μm [e.g. ~2 μm[14]] to pump OPO gain media such as $ZnGeP_2$, $AgGaS_2$, or $AgGaSe_2$.[100]

4.5 Multiwavelength Coherent Sources for Differential Absorption LIDAR and Mutiplex Sensing

An essential requirement for atmospheric sensing based on differential long-path laser absorption[57,58] or (range-resolved) DIAL[15,16,70,71] is a coherent light source capable of generating two specific wavelengths, λ_{on} and λ_{off}. These are effectively the "signal" (resonant with the molecule of interest) and "reference" (nonresonant background) wavelengths for differential detection, as explained in section 2.2. There are several ways to implement such dual-wavelength laser-based sensing techniques:

1. *Separate lasers* – two discrete lasers with their output beams combined and aligned to copropagate to the remote target regions;
2. *Sequential multiwavelength operation* – a single tunable laser that can be sequentially scanned or switched between λ_{off} and λ_{on}, preferably in a time interval [typically <1 ms[16]] that is short compared with appreciable atmospheric turbulence;
3. *Simultaneous multiwavelength operation* – a single laser with a broad spectral gain profile from which the two wavelengths λ_{on} and λ_{off} can simultaneously be extracted (e.g. by a doubly resonant optical cavity or by injection seeding).

A simple example of strategy 1 is the 2.4-μm He–Ne laser long-path absorption sensor for HF emissions in aluminum smelters,[104–106] as recorded in section 4.1. Likewise, an early implementation of water-vapor IR DIAL used 724-nm dye-laser radiation for λ_{on} and a portion of the 649-nm ruby-laser pump radiation for λ_{off}. Concerning strategy 2, various schemes[16] enable a CO_2 laser to be scanned rapidly from one wavelength to another; these usually entail a rotatable intracavity mirror or diffraction grating and can be run either quasi-CW or pulsed (at rates >100 Hz). Various schemes of type 3 enable a dye laser to operate simultaneously on two output wavelengths; these entail splitting the cavity into two paths that are separately tunable (e.g. by insertion of an adjustable wedge plate or by using two separate diffraction gratings)[71,175] or by using adjacent modes of an intracavity étalon.[135] Injection-seeded Nd : YAG lasers are simultaneously operable on two output wavelengths,[176–178] separated by a multiple of the free spectral range (FSR) of the laser cavity.

The above-mentioned FSR-interval restriction has been relaxed in a flashlamp-pumped 800-nm $Ti^{3+} : Al_2O_3$ laser[128] by sustaining high-order transverse modes in the laser cavity and injection-seeding with a single TDL whose wavelength was current-tuned from one laser shot to the next; broadband background limited the device's utility for H_2O-vapor IR DIAL but it was usefully frequency doubled for DIAL detection of NO_2 at ~400 nm.[128]

Strategy 1 has been adopted by NASA for early airborne IR DIAL sensing of H_2O vapor, with λ_{on} and λ_{off} in the 730-nm region generated respectively by a tunable alexandrite laser and a Nd : YAG-pumped dye laser;[16,77,125] these were fired sequentially with a temporal separation of ~0.3 ms, allowing a common detector and transient digitizer to be time shared. Similar time sharing was adopted in NASA's subsequent LASE project,[16,81–83,87,127] where an airborne H_2O-vapor IR DIAL system used method 2 with a double-pulsed 820-nm $Ti^{3+} : Al_2O_3$ laser injection-seeded by two separate TDL wavelengths (controlled by rapidly varying the current i applied sequentially to a single diode). Strategy 2 is also employed in a narrowband IR DIAL system designed for atmospheric water-vapor and temperature profiling,[78,79,130–132] using an intracavity electro-optic switch in the CW single-mode $Ti^{3+} : Al_2O_3$ "master" laser that seeds the Q-switched alexandrite "slave" laser

to generate a double-pulsed sequence of two output wavelengths (λ_{on} and λ_{off}); these are typically separated by a frequency of 50 GHz (1.7 cm^{-1}) and a delay interval of 10 ms.[132]

A novel three-wavelength technique has been adopted in the most recent phase of NASA's project LASE,[81] with alternate shots of the TDL-seeded, double-pulsed $Ti^{3+}:Al_2O_3$ laser operated on pulse pairs (λ_{on}, λ_{off}) and (λ_{side}, λ_{off}), where λ_{side} is accurately positioned on the slope of a strong absorption feature of H_2O vapor. This approach permits almost simultaneous airborne measurements of atmospheric water vapor at high altitude (up to 14 km) and at lower altitudes (down to sea level), along a single ground track.

Injection-seeded OPOs provide another convenient form of dual-wavelength coherent source suitable for spectroscopic remote sensing. Sequential strategy 2 has been used in recent OPO-based IR DIAL demonstrations, [169,174] either switching the TDL-seeded OPO output between λ_{on} and λ_{off} (both narrowband) on alternate shots of the pump laser[169] or switching between narrowband, TDL-seeded (λ_{on}) or broadband, unseeded (λ_{off}) operation.[174]

Injection seeding also enables OPOs to be used as a simultaneous, dual-wavelength coherent source for spectroscopic remote sensing. This has been verified in the author's laboratory,[179,180] via a variant of strategy 3: bichromatic 607-nm signal output from a pulsed BBO OPO, pumped at 355 nm and injection-seeded by two separate 855-nm single-mode TDLs has been combined with monochromatic 532-nm radiation to record single-shot coherent-Raman thermometric spectra of N_2 in furnace air.[179] This is an instance of "spectroscopic tailoring" of OPO output radiation to match spectral features of interest. There are other reports of dual-wavelength OPO operation by simultaneous bichromatic pumping or seeding.[177,178] Our approach[22,166,170,179,180] avoids the limitations of cavity FSR effects by passive control of an OPO cavity with one of its reflectors slightly misaligned (by ~2 mrad). Continuous tuning of the injection-seeded OPO signal and idler outputs is then facilitated by decreasing the effective finesse of the OPO cavity, thereby avoiding mode hops as the TDL seed wavelength is scanned (without needing to lock the OPO cavity length to the TDL wavelength). Two separate TDLs can then be used to injection seed such a misaligned OPO cavity and simultaneously generate spectroscopically tailored dual-wavelength signal and idler outputs that are unrestricted by OPO-cavity FSR.[179] However, backconversion sideband effects arise when the OPO is run much above its operating threshold.[180]

This dual-wavelength spectroscopic tailoring concept is capable of extension to multiwavelength remote sensing applications. This would require a source of coherent, pulsed radiation to generate simultaneously a structured set of discrete wavelengths, each of which is set to be on- or off-resonance with characteristic features in spectra of molecular target species of interest. Such remote sensing could be by DIAL or long-path absorption. Figure 2 portrays a multiplex system of this type that is under development in the author's laboratory. It uses a set of single-mode DBR TDLs and a fiber-optic switch (all computer-controllable) to injection seed a multiwavelength passive-cavity OPO. Appropriate modulation and

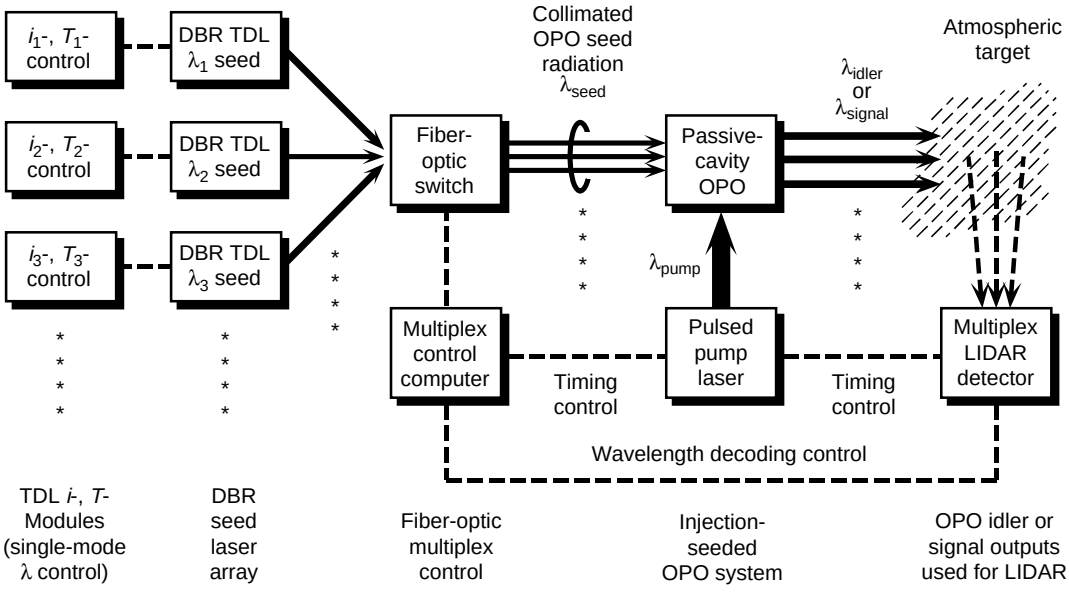

Figure 2 Conceptual schematic of a spectroscopically tailored multiwavelength IR DIAL system.

demodulation sequences can then decode the resulting spectroscopic signals, with a multiplex advantage for sensitivity and specificity. Beyond this preliminary proof of principle based on discrete DBRs as injection seeders, the ultimate multiplex spectroscopic tailoring capability could be by computer-addressable monolithic devices such as two-dimensional rastered multiwavelength VCSEL arrays.[181,182] The potential of multiple-wavelength diode arrays has also been noted by NASA[183] as a means of automated, sequential atmospheric sensing.

5 PERSPECTIVES AND FUTURE DEVELOPMENTS

LIDAR measurements in the IR region (including long-path IR laser absorption, according to this classification) cannot meet all of our atmospheric monitoring requirements and should not be set apart from what can be determined by other optical techniques: both UV/VIS LIDAR[3–6] and "laser-free" techniques such as DOAS, FTIR, and gas-filter correlation methods.[184] IR LIDAR, on which we have necessarily focused, is not always the preferred or sole optical technique needed in a given atmospheric monitoring application. For instance, in atmospheric sensing that characterises molecular species by their IR spectra, FTIR spectroscopy (e.g. long-path absorption with broadband light)[185] yields minimum detection limits in the parts per billion range[59,184] – generally more sensitive than IR DIAL or long-path IR laser absorption. IR LIDAR techniques come into their own, relative to FTIR, when range-resolved or highly specific molecular information is sought over very long optical paths (>0.5 km).

Another relevant comparison can be made between range-resolved IR DIAL and Raman LIDAR, particularly in the context of monitoring of atmospheric water vapor. Both techniques are based on molecular vibrational spectra,[52] but Raman LIDAR[53–55] requires only a fixed-wavelength laser in contrast to the more elaborate tunable or multiwavelength sources needed for IR DIAL (as surveyed in section 4). Some of the traditional drawbacks of Raman LIDAR[24,33,34,78] are its relatively low scattering cross-section (which necessitates high UV or visible laser power and large telescope area), its limited applicability in daylight conditions (owing to signal contamination by background radiation), possible interferences from fluorescing media, and its lack of inbuilt absolute calibration. Many of these limitations appear to have been minimized in recent water-vapor Raman LIDAR developments,[186] using 355-nm pulsed laser excitation and high-rejection optical bandpass filters for simultaneous short- and long-range LIDAR sensing of aerosols (by elastic Mie scattering), H_2O vapor ($\Delta\nu_R \approx 3652\,cm^{-1}$), and N_2 ($\Delta\nu_R \approx 2331\,cm^{-1}$, for real-time ratiometric calibration) in a transportable unit capable of unattended day- and night-time operation. A critical comparison[78] of IR DIAL and Raman LIDAR, in the context of atmospheric water-vapor profiling, concludes that the two techniques are complementary for ground-based sensing (but that Raman LIDAR from air- or spaceborne platforms is unrealistic) and that there is little to separate their performance in tropospheric monitoring applications.

Subtle choices are also sometimes needed between different modes of IR LIDAR detection, for instance, whether to employ heterodyne detection or not,[187] or trade-offs between using shorter-wavelength IR lasers (which are easier to use, more readily available, or of higher performance than those at longer wavelengths) and correspondingly weaker absorption cross-sections of IR overtone- or combination-band spectra relative to longer-wavelength IR fundamental bands.

The distinctive utility of IR LIDAR techniques can be attributed to three key features:

1. The higher efficiency of scattering of IR radiation from aerosols (relative to that from molecules, which are favored at shorter wavelengths) enables vital meteorological and environmental measurements on clouds, winds and particulates[6,40,49] and yields the "distributed mirror" advantage[68] that is used in range-resolved IR DIAL.[15,16,70,71]
2. The molecular information content of IR absorption spectra,[52,57–59] provides characteristic signatures for many significant atmospheric species including H_2O, CO_2, CO, NO, NO_2, O_3, CH_4, H_2CO and myriad other organic molecules, SO_2, and even O_2 (via its NIR magnetic-dipole bands[143] – but not its diatomic partner in the air, N_2).
3. The versatility of coherent tunable IR sources (lasers, OPOs, and other nonlinear-optical devices) is well suited to DIAL and long-path laser absorption measurements.

Aspect 1 has been adequately discussed in sections 1 and 2. With regard to aspect 2, it should be noted that the complex mixtures of molecules in the air can often lead to severe interference effects when IR spectra are employed for analysis purposes. FTIR, with its broadband advantage, can then be helpful in providing survey spectra to identify a set of characteristic wavelengths suitable for multiwavelength IR LIDAR measurements. Recent IR LIDAR work has been making increasing use of chemometric techniques, such as multiple linear regression[108] or principal component analysis[188] to address this type of problem.

Aspect 3 provides the dominant theme of this article. As in the past, we can expect further advances in IR LIDAR technology to rely heavily on new laser, nonlinear-optical, and photonics technology. One instance of this is the spectroscopically tailored multiwavelength OPO source proposed in the context of Figure 2, with its blend of such technologies (DBR TDLs, fiber-optic switches, VCSEL arrays, multiplex OPO injection seeding, etc.) involved.

Further development (and improved availability) of solid-state IR photodetector technology[75,76] is needed for LIDAR to take full advantage of the current state of the art in tunable IR lasers, OPOs, OPAs, and DFG devices. Silicon APDs have already been accepted in NIR LIDAR applications[77,88] and their Hg(Cd, Mn)Te counterpart[76] is an enticing prospect for mid-IR remote sensing. The same sense of anticipation applies to array detectors, such as HgCdTe CCDs,[75,76] with their potential for mid-IR imaging and multiplex spectroscopy, as has recently been demonstrated in the context of broadband UV/VIS DIAL.[189]

Whatever the future prospects of remote sensing of the atmosphere,[190] it seems certain that spaceborne LIDAR systems will play an increasingly significant role, with satellite-based mapping of water vapor, ozone, clouds, aerosols, and meteorological conditions in the Earth's atmosphere among their top priorities. LIDAR-in-space systems[40,146,191] in turn pose formidable technological challenges in terms of compactness, ruggedness, sensitivity, automation, and long-term reliability. This is an obvious new frontier for IR LIDAR. However, it should be recognized that spaceborne LIDAR projects,[82,83,88] either realized (e.g. LITE) or immediately projected (e.g. PICASSO/CENA), are so far confined to fixed-wavelength "nonresonant" studies of aerosol backscattering. Payload restrictions and technological complexities appear to cause spaceborne IR DIAL systems (using a tunable or multiwavelength laser for spectroscopic applications such as global sensing of atmospheric water vapor) to be held at a feasibility study stage.[82,87,191]

ACKNOWLEDGMENTS

Research funding from the Australian Research Council is gratefully acknowledged. Thanks are due to Charles Halloway, Dr Yabai He, Maria Hyland, and Richard Shorten for valuable assistance. It is also appropriate to pay tribute to the many authors who have built up the formidable body of literature on laser-based atmospheric sensing, particularly that cited in section 1.2; they have helped to make this author's task both a pleasure and a revelation.

ABBREVIATIONS AND ACRONYMS

APD	Avalanche Photodiode
AR	Antireflection
BAGI	Backscatter Absorption Gas Imaging
BBO	β-Barium Borate
CCD	Charge-coupled Device
CW	Continuous-wave
DBR	Distributed Bragg Reflector
DFB	Distributed Feedback
DFG	Difference-frequency Generation
DIAL	Differential Absorption LIDAR
DISC	Differential Scattering
DOAS	Differential Optical Absorption Spectroscopy
ECDL	External-cavity Diode Laser
FSR	Free Spectral Range
FTIR	Fourier-transform Infrared
GSGG	Gadolinium Scandium Aluminum Garnet
HF	Hydrogen Fluoride
IR	Infrared
KTP	Potassium Titanyl Phosphate
LASE	LIDAR Atmospheric Sensing Experiment
LIDAR	Light Detection and Ranging
LIF	Laser-induced Fluorescence
LIIRF	Laser-induced Infrared Fluorescence
LITE	LIDAR In-space Technology Experiment
NASA	National Aeronautics and Space Administration
NIR	Near-infrared
OPA	Optical Parametric Amplifier
OPO	Optical Parametric Oscillator
PICASSO/CENA	Pathfinder Instruments for Cloud and Aerosol Spaceborne Observations/Climatologie Etendue des Nuages et des Aerosols
PPLN	Periodically Poled Lithium Niobate
QPM	Quasi-phase-matched
RADAR	Radio Detection and Ranging
SLM	Single Longitudinal Mode
SRS	Stimulated Raman Scattering
TDL	Tunable Diode Laser
TE	Transversely Excited

TIIRF	Thermally Induced Infrared Fluorescence
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VCSEL	Vertical-cavity Surface-emitting Laser
YAG	Yttrium Aluminum Garnet
YLF	Yttrium Lithium Fluoride

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Airborne Instrumentation for Aerosol Measurements • Differential Optical Absorption Spectroscopy, Air Monitoring by • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Chemical-sensing Networks: Satellite-based

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Satellite and Sensor Systems for Environmental Monitoring

REFERENCES

1. R. Greenler, *Rainbows, Halos, and Glories*, Cambridge University Press, Cambridge, 1980.
2. R.P. Wayne, *Chemistry of Atmospheres*, 2nd edition, Clarendon Press, Oxford, 1991.
3. R.M. Measures, *Laser Remote Sensing, Fundamentals and Applications*, Wiley, New York, 1984.
4. R.M. Measures (ed.), *Laser Remote Chemical Analysis*, Wiley, New York, 1988.
5. M.W. Sigrist (ed.), *Air Monitoring by Spectroscopic Techniques*, Wiley, New York, 1994.
6. W.B. Grant, 'LIDAR for Atmospheric and Hydrospheric Studies', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, Chapter 7, 213–305, 1995.
7. W.E.K. Middleton, A.F. Spilhaus, *Meteorological Instruments*, 3rd edition, University of Toronto Press, Toronto, 208, 1953.
8. T.H. Maiman, 'Stimulated Optical Radiation in Ruby', *Nature (London)*, **187**, 493–494 (1960).
9. W.B. Grant, E.V. Browell, R.T. Menzies, K. Sassen, C.-Y. She (eds.), *Selected Papers on Laser Applications in Remote Sensing*, SPIE Optical Engineering Press, Bellingham, WA, MS141, 1997.
10. A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986.
11. L.F. Mollenauer, J.C. White (eds.), *Tunable Lasers*, Springer, New York, 1987.
12. F.J. Duarte (ed.), *Tunable Laser Applications*, Marcel Dekker, New York, 1995.
13. F.J. Duarte (ed.), *Tunable Lasers Handbook*, Academic Press, San Diego, CA, 1995.
14. W. Koechner, *Solid-state Laser Engineering*, 4th edition, Springer, New York, 1996.
15. S. Svanberg, 'Differential Absorption LIDAR (DIAL)', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, Chapter 3, 85–161, 1994.
16. W.B. Grant, 'Differential Absorption LIDAR', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, 231–245, 1995.
17. U. Platt, 'Differential Optical Absorption Spectroscopy (DOAS)', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, Chapter 2, 27–84, 1994.
18. P.K. Cheo, 'CO_2 Lasers', in *Lasers*, eds. A.K. Levine, A.J. DeMaria, Marcel Dekker, New York, Chapter 2, 111–267, Vol. III, 1971.
19. T. Jaeger, G. Wang, 'Tunable High-pressure Infrared Lasers', in *Tunable Lasers*, eds. L.F. Mollenauer, J.C. White, Springer, New York, Chapter 8, 303–330, 1987.
20. W.J. Witteman, *The CO_2 Laser*, Springer, New York, 1987.
21. C. Freed, 'CO_2 Isotope Lasers and Their Applications in Tunable Laser Spectroscopy', in *Tunable Lasers Handbook*, ed. F.J. Duarte, Academic Press, San Diego, CA, Chapter 4, 63–165, 1995.
22. B.J. Orr, M.J. Johnson, J.G. Haub, 'Spectroscopic Applications of Pulsed Tunable Optical Parametric Oscillators', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, Chapter 2, 11–82, 1995.
23. W.B. Grant, R.T. Menzies, 'A Survey of Laser and Selected Optical Systems for Remote Measurement of Pollutant Gas Concentrations', *J. Air Pollut. Control Assoc.*, **33**, 187–194 (1983).
24. W.B. Grant, 'Laser Remote Sensing Techniques', in *Laser Spectroscopy and Its Applications*, eds. L.J. Radziemski, R.W. Solarz, J.A. Paisner, Marcel Dekker, New York, Chapter 8, 565–621, 1987.

25. W.B. Grant, *LIDAR Bibliography*, OpticsNet, Optical Society of America, Washington, DC, 1998, at URL <http://www.osa.org/homes1/general/biblio/lidar97.htm>
26. E.D. Hinkley (ed.), *Laser Monitoring of the Atmosphere*, Springer, New York, 1976.
27. D.K. Killinger, A. Mooradian (eds.), *Optical and Laser Remote Sensing*, Springer, New York, 1983.
28. P.L. Hanst, 'Infrared Spectroscopy and Infrared Lasers in Air Pollution Research and Monitoring', *Appl. Spectrosc.*, **24**, 161–174 (1970).
29. H. Kildal, R.L. Byer, 'Comparison of Laser Methods for the Remote Detection of Pollutants', *Proc. IEEE*, **59**, 1644–1663 (1971).
30. E.D. Hinkley, P.L. Kelley, 'Detection of Air Pollutants with Tunable Diode Lasers', *Science*, **171**, 635–639 (1971).
31. E.D. Hinkley, 'Tunable Infrared Lasers and Their Application to Air Pollution Measurements', *Opto-electronics*, **4**, 69–86 (1972).
32. R.T. Menzies, 'Remote Sensing with Infrared Heterodyne Radiometers', *Opto-electronics*, **4**, 179–186 (1972).
33. D.K. Killinger, N. Menyuk, 'Laser Remote Sensing of the Atmosphere', *Science*, **235**, 37–45 (1987).
34. S. Svanberg, *Atomic and Molecular Spectroscopy*, 2nd edition, Springer, New York, 1992.
35. W. Demtröder, *Laser Spectroscopy*, 2nd edition, Springer, New York, 1996.
36. M.D. Levenson, *Introduction to Nonlinear Laser Spectroscopy*, Academic Press, New York, 1982.
37. Y.R. Shen, *The Principles of Nonlinear Optics*, Wiley, New York, 1984.
38. A. Yariv, *Quantum Electronics*, 3rd edition, Wiley, New York, 1989.
39. M.W. Sigrist, 'Introduction to Environmental Sensing', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, Chapter 1, 1–26, 1994.
40. W.B. Grant, 'Aerosols', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, 218–230, 1995.
41. H.C. van de Hulst, *Light Scattering by Small Particles*, Wiley, New York, 1957.
42. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, Academic Press, New York, 1969.
43. M. Born, E. Wolf, *Principles of Optics*, 6th edition, Pergamon Press, New York, 633–664, 1980.
44. S.C. Hill, A.C. Hill, P.W. Barber, 'Light Scattering by Size/Shape Distributions of Soil Particles and Spheroids', *Appl. Opt.*, **23**, 1025–1031 (1984).
45. C.F. Bohren, S.B. Singham, 'Backscattering by Nonspherical Particles, a Review of Methods and Suggested New Approaches', *J. Geophys. Res.*, **96**, 5269–5277 (1991).
46. Y. Sasano, E.V. Browell, 'Light Scattering Characteristics of Various Aerosol Types Derived from Multiple Wavelength LIDAR Observations', *Appl. Opt.*, **28**, 1670–1679 (1989).
47. L. Stefanutti, F. Castagnoli, M. Del Guasta, M. Morandi, V.M. Sacco, V. Venturi, L. Zuccagnoli, J. Kolenda, H. Kneipp, P. Rairoux, B. Stein, D. Weidauer, J.P. Wolf, 'A Four-wavelength Depolarization Backscattering LIDAR for Polar Stratospheric Cloud Monitoring', *Appl. Phys. B*, **55**, 13–17 (1992).
48. J.W. Bilbro, C.A. DiMarzio, D. Fitzjarrald, S. Johnson, W. Jones, 'Airborne Doppler LIDAR Measurements', *Appl. Opt.*, **25**, 3952–3960 (1986).
49. R.M. Huffaker, R.M. Hardesty, 'Remote Sensing of Atmospheric Wind Velocities Using Solid-state and CO_2 Laser Systems', *Proc. IEEE*, **84**, 181–204 (1996).
50. R.T. Menzies, 'Laser Heterodyne Detection Techniques', in *Laser Monitoring of the Atmosphere*, ed. E.D. Hinkley, Springer, New York, Chapter 7, 297–353, 1976.
51. A. Weber (ed.), *Raman Spectroscopy of Gases and Liquids*, Springer, New York, 1979.
52. N.B. Colthup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, San Diego, CA, 1990.
53. H. Inaba, T. Kobayasi, 'Laser-Raman RADAR–Laser-Raman Scattering Methods for Remote Detection and Analysis of Atmospheric Pollution', *Opto-electronics*, **4**, 101–123 (1972).
54. H. Inaba, 'Detection of Atoms and Molecules by Raman Scattering and Resonance Fluorescence', in *Laser Monitoring in the Atmosphere*, ed. E.D. Hinkley, Springer, New York, Chapter 5, 153–236, 1976.
55. W.B. Grant, 'Raman LIDAR', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, 245–248, 1995.
56. M.P. Bogaard, B.J. Orr, 'Electric Dipole Polarizabilities of Atoms and Molecules', in *MTP Review of Molecular Structure and Properties, Series II*, ed. A.D. Buckingham, Butterworth, London, Chapter 5, 149–194, Vol. 2, 1975.
57. C.R. Webster, R.T. Menzies, E.D. Hinkley, 'Infrared Laser Absorption: Theory and Applications', in *Laser Remote Chemical Analysis*, Wiley, New York, Chapter 3, 85–162, 1988.
58. H.L. Schiff, G.I. Mackay, J. Bechara, 'The Use of Tunable Diode Laser Absorption Spectroscopy for Atmospheric Measurements', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, Chapter 5, 239–333, 1994.
59. P.L. Hanst, S.T. Hanst, 'Gas Measurement in the Fundamental Infrared Region', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, Chapter 6, 335–479, 1994.
60. F. Allario, S.J. Katzberg, J.M. Hoell, 'Tunable Laser Heterodyne Spectrometer Measurements of Atmospheric Species', in *Optical and Laser Remote Sensing*, eds. D.K. Killinger, A. Mooradian, Springer, New York, 50–73, 1983.

61. T.J. McIlrath, 'Fluorescence LIDAR', *Opt. Eng.*, **19**, 494–502 (1980).
62. I.S. McDermid, J.B. Laudenslager, T.J. Pacala, 'New Technological Developments for the Remote Detection of Atmospheric Hydroxyl Radicals', *Appl. Opt.*, **22**, 2586–2591 (1983).
63. D.E. Heard, 'Measuring the Elusive Tropospheric Hydroxyl Radical', *Atmos. Environ.*, **32**, 801–802 (1998).
64. G. Mégie, 'Laser Measurements of Atmospheric Trace Constituents', in *Laser Remote Chemical Analysis*, Wiley, New York, Chapter 5, 333–408, 1988.
65. W.B. Grant, 'Metal Atoms and Ions', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, 248–252, 1995.
66. J.T. Yardley, *Introduction to Molecular Energy Transfer*, Academic Press, New York, 1980.
67. J.W. Robinson, J.D. Dake, 'Remote Sensing of Air Pollutants by Laser-induced Infrared Fluorescence – a Review', *Anal. Chim. Acta*, **71**, 277–288 (1974).
68. R.L. Byer, M. Garbuny, 'Pollutant Detection by Absorption Using Mie Scattering and Topographic Targets as Retroreflectors', *Appl. Opt.*, **12**, 1496–1505 (1973).
69. T. Kobayasi, H. Inaba, 'Infrared Heterodyne Laser RADAR for Remote Sensing of Air Pollutants by Range-resolved Differential Absorption', *Opt. Quantum Electron.*, **7**, 319–327 (1975).
70. R.M. Measures, 'Fundamentals of Laser Remote Sensing', in *Laser Remote Chemical Analysis*, Wiley, New York, Chapter 1, 1–84, 1988.
71. K.A. Fredriksson, 'Differential Absorption LIDAR for Pollution Mapping', in *Laser Remote Chemical Analysis*, Wiley, New York, Chapter 4, 273–332, 1988.
72. R. Beer, 'Transmission Through the Atmosphere', in *Laser Remote Chemical Analysis*, Wiley, New York, Chapter 2, 153–236, 1988.
73. L.S. Rothman, C.P. Rinsland, A. Goldman, S.T. Massie, D.P. Edwards, J.-M. Flaud, A. Perrin, C. Camy-Peyret, V. Dana, J.-Y. Mandin, J. Schroeder, A. McCann, R.R. Gamache, R.B. Watson, K. Yoshino, K.V. Chance, K.W. Jucks, L.R. Brown, V. Nemtchinov, P. Varanasi, 'The HITRAN Molecular Spectroscopic Database and HAWKS (HITRAN Atmospheric Workstation): 1996 Edition', *J. Quantum Spectrosc. Radiat. Transfer*, **60**, 665–710 (1998).
74. R.T. Menzies, 'Ozone Spectroscopy with a CO_2 Waveguide Laser', *Appl. Opt.*, **15**, 2597–2599 (1976).
75. M. Schlessinger, *Infrared Technology Fundamentals*, Marcel Dekker, New York, 1994.
76. A. Rogalski (ed.), *Infrared Photon Detectors*, SPIE Optical Engineering Press, Bellingham, WA, 1995.
77. N.S. Higdon, E.V. Browell, P. Ponsardin, B.E. Grossmann, C.F. Butler, T.H. Chyba, M.N. Mayo, R.J. Allen, A.W. Heuser, W.B. Grant, S. Ismail, S.D. Mayor, A.F. Carter, 'Airborne Differential Absorption LIDAR System for Measurements of Atmospheric Water Vapor and Aerosols', *Appl. Opt.*, **33**, 6422–6438 (1994).
78. V. Wulfmeyer, J. Bösenberg, 'Ground-based Differential Absorption LIDAR for Water-vapor and Temperature Profiling: Assessment of Accuracy, Resolution, and Meteorological Applications', *Appl. Opt.*, **37**, 3825–3844 (1998).
79. J. Bösenberg, 'Ground-based Differential Absorption LIDAR for Water-vapor and Temperature Profiling: Methodology', *Appl. Opt.*, **37**, 3845–3860 (1998).
80. J.D. Spinhirne, S. Chudamani, J.F. Cavanaugh, J.L. Bufton, 'Aerosol and Cloud Backscatter at 1.06, 1.54, and 0.53 μm by Airborne Hard-target-calibrated Nd : YAG/Methane Raman LIDAR', *Appl. Opt.*, **36**, 3475–3490 (1997).
81. A.S. Moore, K.E. Brown, W.M. Hall, J.C. Barnes, W.C. Edwards, L.B. Petway, A.D. Little, W.S. Luck, I.W. Jones, C.W. Antill Jr, E.V. Browell, S. Ismail, 'Development of the LIDAR Atmospheric Sensing Experiment (LASE) – An Advanced Airborne DIAL Instrument', in *Advances in Atmospheric Remote Sensing with LIDAR*, eds. A. Ansmann, R. Neuber, P. Rairoux, U. Wandinger Springer, New York, 281–288, 1996.
82. E.V. Browell, S. Ismail, W.B. Grant, 'Differential Absorption LIDAR (DIAL) Measurements from Air and Space', *Appl. Phys. B*, **67**, 399–410 (1998).
83. NASA's Atmospheric Sciences Division Home Page, at <http://asd-www.larc.nasa.gov/ASDhomepage.html>, provides links to the NASA Langley LIDAR Applications Group and specific projects such as LASE and LITE also the joint NASA–CNES (Centre National d'Etudes Spatiales) project PICASSO/CENA.
84. R.M. Schotland, 'Some Observations of the Vertical Profile of Water Vapor by a Laser Optical RADAR', in *Proceedings of the Fourth Symposium on Remote Sensing of the Environment (University of Michigan, Ann Arbor, Michigan, 12–14 April 1966)*, 273–283, 1966; reprinted in Grant et al., 423–430.
85. E.V. Browell, T.D. Wilkerson, T.J. McIlrath, 'Water Vapor Differential Development and Evaluation', *Appl. Opt.*, **18**, 3474–3483 (1979).
86. W.B. Grant, 'Differential Absorption and Raman LIDAR for Water Vapor Profile Measurements: a Review', *Opt. Eng.*, **30**, 40–48 (1991).
87. S. Ismail, E.V. Browell, 'Recent LIDAR Technology Developments and Their Influence on Measurements of Tropospheric Water Vapor', *J. Atmos. Ocean. Technol.*, **11**, 76–84 (1994).
88. D.M. Winker, R.H. Couch, M.P. McCormick, 'An Overview of LITE: NASA's LIDAR In-space Technology Experiment', *Proc. IEEE*, **84**, 164–180 (1996).
89. Occupational Safety and Health Administration (OSHA), *Guidelines for Laser Safety and Hazard Assessment*, OSHA Directives Publication 8-1.7, US Department of Labor, Washington, DC, 1991; accessible at <http://www.osha-slc.gov/OshDoc/Directive_data/PUB_ 8-1_7.html>

90. E.M. Patterson, D.W. Roberts, G.G. Gimmestad, 'Initial Measurements Using a 1.54-μm Eyesafe Raman Shifted LIDAR', *Appl. Opt.*, **28**, 4978–4981 (1989).
91. S.W. Henderson, C.P. Hale, J.R. Magee, M.J. Kavaya, A.V. Huffaker, M.J. Kavaya 'Eye-safe Coherent Laser RADAR System at 2.1 μm Using Tm,Ho : YAG Lasers', *Opt. Lett.*, **16**, 773–775 (1991).
92. S.W. Henderson, P.J.M. Suni, C.P. Hale, S.M. Hannon, J.R. Magee, D.L. Bruns, E.H. Yuen, 'Coherent Laser RADAR at 2 μm Using Solid-state Lasers', *IEEE Trans. Geosci. Remote Sens.*, **GRS-31**, 4–14 (1993).
93. T.M. Taczak, D.K. Killinger, 'Development of a Tunable, Narrow-linewidth, CW 2.066-μm Ho : YLF Laser for Remote Sensing of Atmospheric CO_2 and H_2O', *Appl. Opt.*, **37**, 8460–8476 (1998).
94. N.P. Barnes, 'Transition Metal Solid-state Lasers', in *Tunable Lasers Handbook*, ed. F.J. Duarte, Academic Press, San Diego, CA, Chapter 6, 219–291, 1995.
95. J.E. Geusic, H.M. Marcos, L.G. Van Uitert, 'Laser Oscillations in Nd-doped Yttrium Aluminum, Yttrium Gallium and Gadolinium Garnets', *Appl. Phys. Lett.*, **4**, 182–184 (1964).
96. R.L. Herbst, R.L. Byer, 'The Unstable Resonator YAG', *Laser Focus*, **14**(7), 48–57 (1978).
97. Y.K. Park, G. Giuliani, R.L. Byer, 'Single Axial Mode Operation of a Q-switched Nd : YAG Oscillator by Injection Seeding', *IEEE J. Quantum Electron.*, **QE-20**, 117–125 (1984).
98. R.L. Schmitt, L.A. Rahn, 'Diode-laser-pumped Nd : YAG Laser Injection Seeding System', *Appl. Opt.*, **4**, 629–633 (1986).
99. W. Krichbaumer, H. Herrmann, E. Nagel, R. Häring, J. Streicher, Ch. Werner, A. Mehnert, Th. Halldorsson, S. Heinemann, P. Peuser, N.P. Schmitt, 'A Diode-pumped Nd : YAG LIDAR for Airborne Cloud Measurements', *Opt. Laser Technol.*, **25**, 283–287 (1993).
100. V.G. Dmitriev, G.G. Gurzayan, D.N. Nikogosyan, *Handbook of Nonlinear Optical Crystals*, 2nd edition, Springer, New York, 1997.
101. R. Beck, W. Englisch, K. Gürs, *Table of Laser Lines in Gases and Vapors*, 2nd edition, Springer, New York, 1978.
102. L. Allen, D.G.C. Jones, 'The Helium–Neon Laser', *Adv. Phys.*, **14**, 479–519 (1965).
103. J.C. Scott, R.A.M. Maddever, A.T. Paton, 'Spectroscopy of Methane using a Nd : YAG Laser at 1.34 μm', *Appl. Opt.*, **31**, 815–821 (1992).
104. J.E. Eberhardt, A.W. Pryor, 'Application of a CW Helium Neon Laser for Measurement of Gaseous Hydrogen Fluoride Concentration', *Appl. Phys. B*, **27**, 43–47 (1982).
105. J.E. Eberhardt, D.L. Death, R.P. Read, C.A. Rogers, 'Potroom Trial of a 400 m Path Laser Hydrogen Fluoride Measuring Instrument', in *Light Metals 1995*, ed. J. Evans, Minerals, Metals and Materials Society, Warrendale, PA, 335–340, 1995.
106. D.L. Death, J.E. Eberhardt, R.P. Read, C.A. Rogers, M. Atkinson, D. Collins, F. Fleer, K. Whiteley, 'Potroom Comparison Trial of a Laser HF Measuring Instrument and an Intensive Air Sampling Array over a 400 m Path', in *Light Metals 1997*, ed. R. Huglen, Minerals, Metals and Materials Society, Warrendale, PA, 173–177, 1997.
107. M.W. Sigrist, 'Air Monitoring by Laser Photoacoustic Spectroscopy', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Wiley, New York, Chapter 4, 163–238, 1994.
108. J.R. Quagliano, P.O. Stoutland, R.R. Petrin, R.K. Sander, R.J. Romero, M.C. Whitehead, C.R. Quick, J.J. Tiee, L.J. Jolin, 'Quantitative Chemical Identification of Four Gases in Remote Infrared (9–11 μm) Differential Absorption LIDAR Experiments', *Appl. Opt.*, **36**, 1915–1927 (1997).
109. T.G. McRae, T.J. Kulp, 'Backscatter Absorption Gas Imaging: a New Technique for Gas Visualization', *Appl. Opt.*, **32**, 4037–4050 (1993).
110. M.A. Jarzembski, V. Srivastava, 'Comparison of Continuous-wave CO_2 LIDAR Calibration by Use of Earth-surface Targets in Laboratory and Airborne Measurements', *Appl. Opt.*, **37**, 7120–7127 (1998).
111. M.A. Jarzembski, V. Srivastava, J. Rothermel, 'Vertical Variability of Aerosol Backscatter from an Airborne-focused Continuous-wave CO_2 LIDAR at 9.1-μm Wavelength', *Appl. Opt.*, **38**, 908–915 (1999).
112. P. Drobinski, A.M. Dabas, P. Delville, P.H. Flamant, J. Pelon, R.M. Hardesty, 'Refractive-index Structure Parameter in the Planetary Boundary Layer: Comparison of Measurements Taken with a 10.6-μm Coherent LIDAR, a 0.9-μm Scintillometer, and in situ Sensors', *Appl. Opt.*, **38**, 1648–1656 (1999).
113. A. Ben-David, 'Backscattering Measurements of Atmospheric Aerosols at CO_2 Laser Wavelengths: Implications of Aerosol Structure on Differential-absorption LIDAR Retrievals of Molecular Species', *Appl. Opt.*, **38**, 2616–2624 (1999).
114. W. Urban, 'The Carbon Monoxide Laser as Spectroscopic Source', *Laser Optoelektron.*, **23**(1), 56–61 (1991).
115. A. Ben-Shaul, Y. Haas, K.L. Kompa, R.D. Levine, *Lasers and Chemical Change*, Springer, New York, 1981, especially, Chapter 4, 289–359.
116. J.C. Walling, 'Tunable Paramagnetic-ion Solid-state Lasers', in *Tunable Lasers*, eds. L.F. Mollenauer, J.C. White, Springer, New York, Chapter 9, 331–398, 1987.
117. F.P. Schäfer (ed.), *Dye Lasers*, 3rd edition, Springer, New York, 1991.
118. F.J. Duarte (ed.), *High-power Dye Lasers*, Springer, New York, 1991.
119. M. Stuke (ed.), *Dye Lasers: 25 Years*, Springer, New York, 1992.
120. F.J. Duarte, 'Dye lasers', in *Tunable Lasers Handbook*, ed. F.J. Duarte, Academic Press, San Diego, CA, Chapter 5, 167–218, 1995.

121. F.J. Duarte, 'Dispersive External-cavity Semiconductor Lasers', in *Tunable Laser Applications*, ed. F.J. Duarte, Marcel Dekker, New York, Chapter 3, 11–82, 1995.
122. P. Zorabedian, 'Tunable External-cavity Semiconductor Lasers', in *Tunable Lasers Handbook*, ed. F.J. Duarte, Academic Press, San Diego, CA, Chapter 8, 349–442, 1995.
123. P.F. Moulton, 'Spectroscopic and Laser Characteristics of $Ti:Al_2O_3$', *J. Opt. Soc. Am. B*, **3**, 125–133 (1986).
124. J.C. Walling, D.F. Heller, H. Samuelson, D.J. Harter, J.A. Pete, R.C. Morris, 'Tunable Alexandrite Lasers: Development and Performance', *IEEE J. Quantum Electron.*, **QE-21**, 1568–1581 (1985).
125. P. Ponsardin, N.S. Higdon, B.E. Grossmann, E.V. Browell, 'Spectral Control of an Alexandrite Laser for an Airborne Water-vapor Differential Absorption LIDAR System', *Appl. Opt.*, **33**, 6439–6450 (1994).
126. T.D. Raymond, A.V. Smith, 'Injection-seeded Titanium-doped-sapphire Laser', *Opt. Lett.*, **16**, 33–35 (1991).
127. J.C. Barnes, N.P. Barnes, L.G. Wang, W. Edwards, 'Injection Seeding II: $Ti:Al_2O_3$ Experiments', *IEEE J. Quantum Electron.*, **29**, 2684–2692 (1993).
128. J. Yu, P. Rambaldi, J.-P. Wolf, 'Dual-wavelength Diode-seeded Ti:sapphire Laser for Differential Absorption LIDAR Applications', *Appl. Opt.*, **36**, 6864–6868 (1997).
129. D. Bruneau, T. Arnaud des Lions, P. Quaglia, J. Pelon, 'Injection-seeded Pulsed Alexandrite Laser for Differential Absorption LIDAR Application', *Appl. Opt.*, **33**, 3941–3950 (1994).
130. V. Wulfmeyer, J. Bösenberg, S. Lehmann, C. Senff, St. Schmitz, 'Injection-seeded Alexandrite Ring Laser: Performance and Application in a Water-vapor Differential Absorption LIDAR', *Opt. Lett.*, **20**, 638–640 (1995).
131. V. Wulfmeyer, J. Bösenberg, 'Single-mode Operation of an Injection-seeded Alexandrite Ring Laser for Application in Water-vapor and Temperature Differential Absorption LIDAR', *Opt. Lett.*, **21**, 1150–1152 (1996).
132. V. Wulfmeyer, 'Ground-based Differential Absorption LIDAR for Water-vapor and Temperature Profiling: Development and Specifications of a High-performance Laser Transmitter', *Appl. Opt.*, **37**, 3804–3824 (1998).
133. J. Hoeffner, U. von Zahn, W.J. McNeil, E. Murad, 'The 1996 Leonid Shower as Studied with a Potassium LIDAR: Observations and Inferred Meteoroid Sizes', *J. Geophys. Res.*, **104**, 2633–2643 (1999).
134. U. von Zahn, J. Hoeffner, 'Mesopause Temperature Profiling by Potassium LIDAR', *Geophys. Res. Lett.*, **23**, 141–144 (1996).
135. G. Ehret, C. Kiemle, W. Renger, G. Simmet, 'Airborne Remote Sensing of Tropospheric Water Vapor with a Near-infrared Differential Absorption LIDAR System', *Appl. Opt.*, **32**, 4534–4551 (1993).
136. J.J. Coleman, 'Semiconductor Lasers', in *Electro-Optics Handbook*, eds. R. Waynant, M. Ediger, McGraw-Hill, New York, Chapter 6, 1994.
137. E.D. Hinkley, K.W. Nill, F.A. Blum, 'Infrared Spectroscopy with Tunable Lasers', in *Laser Spectroscopy of Atoms and Molecules*, ed. H. Walther, Springer, New York, Chapter 2, 125–196, 1976.
138. C.E. Wieman, L. Hollberg, 'Using Diode Lasers for Atomic Physics', *Rev. Sci. Instrum.*, **62**, 1–20 (1991).
139. E.D. Hinkley, T. Ku, P.L. Kelley, 'Techniques for Detection of Molecular Pollutants by Absorption of Laser Radiation', in *Laser Monitoring in the Atmosphere*, ed. E.D. Hinkley, Springer, New York, 237–296, 1976.
140. L.W. Chaney, D.G. Rickel, G.M. Russwum, W.A. McClenny, 'Long-path Laser Monitor of Carbon Monoxide: System Improvements', *Appl. Opt.*, **18**, 3004–3009 (1979).
141. D.T. Cassidy, J. Reid, 'Atmospheric Pressure Monitoring of Trace Gases Using Tunable Diode Lasers', *Appl. Opt.*, **21**, 1185–1190 (1982).
142. F.S. Pavone, F. Marin, M. Inguscio, K. Ernst, G. Di Lonardo, 'Sensitive Detection of Acetylene Absorption in the Visible Using a Stabilized AlGaAs Diode Laser', *Appl. Opt.*, **32**, 259–262 (1993).
143. Q.-V. Nguyen, R.W. Dibble, T. Day, 'High-resolution Oxygen Absorption Spectrum Obtained with an External-cavity Continuously Tunable Diode Laser System", *Opt. Lett.*, **19**, 2134–2136 (1994).
144. M. Fehér, P.A. Martin, 'Tunable Diode Laser-based Monitoring of Atmospheric Trace Gas Constituents', *Spectrochim. Acta, Part A*, **51**, 1579–1599 (1995).
145. K.P. Petrov, R.F. Curl, F.K. Tittel, L. Goldberg, 'CW Tunable 8.7-μm Spectroscopic Source Pumped by Fiber-coupled Communications Lasers', *Opt. Lett.*, **21**, 1451–1453 (1996).
146. D.G. Lancaster, D. Richter, R.F. Curl, F.K. Tittel, 'Real-time Measurements of Trace Gases Using a Compact Difference-frequency-based Sensor Operating at 3.5 μm', *Appl. Phys. B*, **67**, 339–345 (1998).
147. N. Sugimoto, A. Minato, 'Long-path Absorption Measurement of CO_2 with a Raman-shifted Tunable Dye Laser', *Appl. Opt.*, **32**, 6827–6833 (1993).
148. S.E. Harris, 'Tunable Optical Parametric Oscillators', *Proc. IEEE*, **57**, 2096–2113 (1969).
149. R.L. Byer, 'Parametric Oscillators', in *Laser Spectroscopy*, ed. R.G. Brewer, A. Mooradian, Plenum Press, New York, 77–101, 1973.
150. R.L. Byer, 'Optical Parametric Oscillators', in *Quantum Electronics: a Treatise, Volume I, Nonlinear Optics, Part B*, eds. H. Rabin, C.L. Tang, Academic Press, New York, Chapter 9, 587–702, 1975.
151. R.L. Byer, R.L. Herbst, 'Parametric Oscillation and Mixing', in *Nonlinear Infrared Generation*, ed. by Y.-R. Shen, Springer, New York, Chapter 3, 81–137, 1977.
152. A. Fix, T. Schröder, R. Wallenstein, 'The Optical Parametric Oscillators of Beta-barium Borate and Lithium Borate: New Sources of Powerful Tunable Laser Radiation in the Ultraviolet, Visible and Near Infrared', *Laser Optoelektron.*, **23**(3), 106–110 (1991).

153. C.L. Tang, W.R. Bosenberg, T. Ukachi, R.J. Lane, L.K. Cheng, 'Optical Parametric Oscillators', *Proc. IEEE*, **80**, 365–374 (1992).
154. N.P. Barnes, 'Optical Parametric Oscillators', in *Tunable Lasers Handbook*, ed. F.J. Duarte, Academic Press, San Diego, CA, Chapter 7, 293–348, 1995.
155. L.E. Myers, R.C. Eckardt, M.M. Fejer, R.L. Byer, W.R. Bosenberg, J.W. Pierce, 'Quasi-phase-matched Optical Parametric Oscillators in Bulk Periodically Poled Lithium Niobate', *J. Opt. Soc. Am. B*, **12**, 2102–2116 (1995).
156. A.S. Pine, 'High-resolution Methane ν_3-Band Spectra Using a Stabilized Tunable Difference-frequency Laser System', *J. Opt. Soc. Am.*, **66**, 97–108 (1976).
157. D. Richter, D.G. Lancaster, R.F. Curl, W. Neu, F.K. Tittel, 'Compact Mid-infrared Trace Gas Sensor Based on Difference-frequency Generation of Two Diode Lasers in Periodically Poled $LiNbO_3$', *Appl. Phys. B*, **67**, 347–350 (1998).
158. L. Goldberg, J. Koplow, D.G. Lancaster, R.F. Curl, F.K. Tittel, 'Mid-infrared Difference-frequency Generation Source Pumped by a 1.1–1.5-μm Dual-wavelength Fiber Amplifier for Trace-gas Detection', *Opt. Lett.*, **23**, 1517–1519 (1998).
159. S.J. Brosnan, R.L. Byer, 'Optical Parametric Oscillator Threshold Linewidth Studies', *IEEE J. Quantum Electron.*, **QE-15**, 415–431 (1979).
160. T. Henningsen, M. Garbuny, R.L. Byer, 'Remote Detection of CO by Parametric Tunable Laser', *Appl. Phys. Lett.*, **24**, 242–244 (1974).
161. R.A. Baumgartner, R.L. Byer, 'Remote SO_2 Measurements at 4 μm with a Continuously Tunable Source', *Opt. Lett.*, **2**, 163–165 (1978).
162. R.A. Baumgartner, R.L. Byer, 'Continuously Tunable IR LIDAR with Applications to Remote Measurements of SO_2 and CH_4', *Appl. Opt.*, **17**, 3555–3561 (1978).
163. M. Endemann, R.L. Byer, 'Remote Single-ended Measurements of Atmospheric Temperature and Humidity at 1.77 μm Using a Continuously Tunable Source', *Opt. Lett.*, **5**, 452–454 (1980).
164. D.J. Brassington, 'Differential Absorption LIDAR Measurements of Atmospheric Water Vapor Using an Optical Parametric Oscillator Source', *Appl. Opt.*, **21**, 4411–4416 (1982).
165. W.R. Bosenberg, D.R. Guyer, 'Broadly Tunable, Single-frequency Optical Parametric Frequency-conversion System', *J. Opt. Soc. Am. B*, **10**, 1716–1722 (1993).
166. J.G. Haub, R.M. Hentschel, M.J. Johnson, B.J. Orr, 'Controlling the Performance of a Pulsed Optical Parametric Oscillator: a Survey of Techniques and Spectroscopic Applications', *J. Opt. Soc. Am. B*, **12**, 2128 (1995).
167. A.V. Smith, W.J. Alford, T.D. Raymond, M.S. Bowers, 'Comparison of a Numerical Model with Measured Performance of a Seeded, Nanosecond KTP Optical Parametric Oscillator', *J. Opt. Soc. Am. B*, **12**, 2253–2267 (1995).
168. A. Borsutzky, 'Frequency Control of Pulsed Optical Parametric Oscillators', *Quantum Semiclass. Opt.*, **9**, 191–207 (1997).
169. M.J.T. Milton, T.D. Gardiner, F. Molero, J. Galech, 'Injection-seeded Optical Parametric Oscillator for Range-resolved DIAL Measurements of Atmospheric Methane', *Opt. Commun.*, **142**, 153–160 (1997).
170. G.W. Baxter, H.-D. Barth, B.J. Orr, 'Laser Spectroscopy with a Pulsed, Narrow-band Infrared Optical Parametric Oscillator System: a Practical, Modular Approach', *Appl. Phys. B*, **66**, 653–657 (1998).
171. G.W. Baxter, Y. He, B.J. Orr, 'A Pulsed Optical Parametric Oscillator System, Based on Periodically Poled Lithium Niobate (PPLN), for High-resolution Spectroscopy', *Appl. Phys. B*, **67**, 753–756 (1998).
172. Y. He, G.W. Baxter, B.J. Orr, 'Locking the Cavity of a Pulsed Periodically Poled Lithium Niobate Optical Parametric Oscillator to the Wavelength of a Continuous-wave Injection Seeder by an 'intensity-dip' Method', *Rev. Sci. Instrum.*, **70**, 3203–3213 (1998).
173. G. Ehret, A. Fix, V. Weiss, G. Poberaj, T. Baumert, 'Diode-laser-seeded Optical Parametric Oscillator for Airborne Water Vapor DIAL Application in the Upper Troposphere Lower Stratosphere', *Appl. Phys. B*, **97**, 427–431 (1998).
174. A. Fix, V. Weiss, G. Ehret, 'Injection-seeded Optical Parametric Oscillator for Airborne Water Vapor DIAL', *Pure Appl. Opt.*, **7**, 837–852 (1998).
175. K.A. Fredriksson, 'DIAL Techniques for Pollution Monitoring: Improvements and Complementary Systems', *Appl. Opt.*, **24**, 3297–3394 (1985).
176. T.D. Raymond, A.V. Smith, 'Two-frequency Injection-seeded Nd : YAG Laser', *IEEE J. Quantum Electron.*, **31**, 1734–1737 (1995).
177. W.A. Neuman, S.P. Velsko, 'Spectral Properties of Optical Parametric Oscillators', in *Proceedings of International Conference on Lasers '95*, eds. V.J. Corcoran, T.A. Goldman STS Press, Society for Optical and Quantum Electronics, McLean, VA, 718–725, 1996.
178. M.S. Webb, K.B. Stanion, D.J. Deane, W.A. Cook, W.A. Neuman, S.P. Velsko, 'Multi-wavelength Injection Seeded Mid-infrared Optical Parametric Oscillator', *Proc. SPIE*, **2700**, 269–300 (1996).
179. G.W. Baxter, M.J. Johnson, J.G. Haub, B.J. Orr, 'OPO CARS: Coherent Anti-Stokes Raman Spectroscopy Using Tunable Optical Parametric Oscillators Injection-seeded by External-cavity Diode Lasers', *Chem. Phys. Lett.*, **251**, 211–218 (1996).
180. G.W. Baxter, J.G. Haub, B.J. Orr, 'Backconversion in a Pulsed Optical Parametric Oscillator: Evidence from Injection-seeded Sidebands', *J. Opt. Soc. Am. B*, **14**, 2723–2730 (1997).
181. C.J. Chang-Hasnain, J.P. Harbison, C.-E. Zah, M.W. Maeda, L.T. Florez, N.G. Stoffel, T.-P. Lee, 'Multiple

wavelength Tunable Surface-emitting Laser Arrays', *IEEE J. Quantum Electron.*, **27**, 1368–1376 (1991).
182. E. Zeeb, B. Möller, G. Reiner, M. Ries, T. Hackbarth, K.J. Ebeling, 'Planar Proton Implanted VCSEL's and Fiber-coupled VCSEL Arrays', *IEEE J. Sel. Top. Quantum Electron.*, **1**, 616–623 (1995).
183. D.L. Wall, J.C. Sproul, Z. Feit, G.W. Sachse, 'Development of IR Tunable Diode Lasers and Source Assemblies for Atmospheric Monitoring and Related Applications', *Proc. SPIE*, **2112**, 2–11 (1994).
184. W.B. Grant, R.H. Kagann, W.A. McClenny, 'Optical Remote Measurement of Toxic Gases', *J. Air Waste Manage. Assoc.*, **42**, 18–30 (1992).
185. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectroscopy*, Wiley, New York, 1986.
186. J.E.M. Goldsmith, F.H. Blair, S.E. Bisson, D.D. Turner, 'Turn-key Raman LIDAR for Profiling Atmospheric Water Vapor, Clouds, and Aerosols', *Appl. Opt.*, **37**, 4979–4990 (1998).
187. D.K. Killinger, N. Menyuk, W.B. DeFeo, 'Experimental Comparison of Heterodyne and Direct Detection for Pulsed Differential Absorption CO_2 LIDAR', *Appl. Opt.*, **22**, 682–689 (1985).
188. D.P. Donovan, A.I. Carswell, 'Principal Component Analysis Applied to Multiwavelength LIDAR Aerosol Backscatter and Extinction Measurements', *Appl. Opt.*, **36**, 9406–9424 (1997).
189. I.M. Povey, A.M. South, A. t'Kint de Roodenbeke, C. Hill, R.A. Freshwater, R.L. Jones, 'A Broadband LIDAR for the Measurement of Tropospheric Constituent Profiles from the Ground', *J. Geophys. Res.*, **103**, 3369–3380 (1998).
190. R.M. Measures, 'Laser Remote Sensing: Present Status and Future Prospects', *Proc. SPIE*, **3059**, 2–8 (1997).
191. J.J. Ewing, J.C. Barnes, 'Conceptual Designs for Advanced Space Based LIDAR Source', *Proc. SPIE*, **3104**, 111–122 (1997).

Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode

M. Tacke
Forschungsinstitut für Optronik und Mustererkennung, Ettlingen, Germany

F. Wienhold and R. Grisar
Fraunhofer Institut für Messtechnik, Freiburg, Germany

H. Fischer
Max-Planck-Institut für Chemie, Mainz, Germany

F.-J. Lübken
Institut für Atmosphärenphysik, Ostseebad Kühlungsborn, Germany

1 Introduction 1
2 Tunable Diode Laser Absorption Spectroscopy Basics 2
3 Applications 7
3.1 Review of Monitoring Applications 7
3.2 Airborne Monitoring 9
3.3 Rocket-based Experiments 15
3.4 Exhaust Gas Analysis 20
4 Relevant System Parameters 25
4.1 Gas Handling 25
4.2 Diode Lasers 25
4.3 Conclusions 29
5 Comparison to Other Gas-monitoring Methods 29
Abbreviations and Acronyms 31
Related Articles 31
References 32

Infrared (IR) spectroscopy is a proven tool for gas analysis, and tunable mid-IR diode laser absorption spectroscopy is a special technique giving optimum sensitivity, selectivity and measurement speed for most molecules of interest. Molecules such as the NO_x family (N_2O, NO, NO_2, HONO, and HNO_3) can be detected selectively down to well below part per billion levels. The technique works well with all molecules having line spectra and thus is quite universal. It is conceptually simple and detects gases by repetitively scanning absorption lines, which is possible due to the narrow laser line width and easy laser tuning. The scanning procedures of diode laser spectroscopy optimally reduce background fluctuations. Basic operation principles are explained in the article.

Alternative techniques are well established and often allow detection of single components at lower system cost, but in general need specific sample pretreatment that makes the results often difficult to compare. Mass spectroscopy is an alternative multicomponent technique that reaches comparable selectivity, but only with preselection processes which induce calibration problems of their own, such as matrix-dependent sensitivity. In order to give an insight into the capabilities of tunable mid-IR diode laser absorption spectroscopy, three exemplary systems are treated in detail for air monitoring from airplanes, from rockets, and gas analysis of combustion exhaust. System parameters are explained, and system performance is compared with alternative techniques.

1 INTRODUCTION

A large number of molecules of interest in air monitoring are accessible to IR spectroscopy, such as O_3, N_2O, NO, NO_2, HONO, HNO_3, CO, CO_2, H_2O_2, HCHO, NH_3, CH_4, C_2H_2, C_2H_4, C_2H_6, HCl, HF, SiF_6, CF_2Cl_2, H_2S, and OCS. All of these trace constituents of the atmosphere have been detected and their concentrations measured by laser spectroscopy. Detection levels to below 100 ppt are reported with temporal resolution of the order of minutes. Thus, tunable mid-IR diode laser spectroscopy is clearly a universal monitoring technique.

The concentrations of single components are of interest when monitoring air quality by representative species. However, the detection of a number of compounds is important for an understanding of atmospheric chemical reactions. At present many details of the processes that govern the natural destruction of man-made pollutants and the resultant formation of new compounds are unknown. In the longer term, it will be necessary to monitor a list of compounds in order to keep track of the development of the atmosphere's composition, to ensure that it does not change too rapidly.

This article describes tunable mid-IR diode laser spectroscopy in terms of three specific systems. The first has been used for air monitoring on board aircraft. This system can be regarded as typical of laboratory systems but shows that such systems are rugged and can be used for mobile campaigns.

The second example was developed for the special purpose of atmospheric measurement during a rocket mission. This system was designed to demonstrate the ability of the mid-IR diode laser technique to withstand the demanding conditions of a rocket flight.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The last example comes from emissions monitoring in industrial environments. This differs from the other two in that the air samples may be at temperatures up to several hundred degrees centigrade and can be contaminated with material such as soot, that make analysis difficult. Tunable mid-IR diode laser spectroscopy can cope with such samples, in many cases even without filtering. Additionally, this system demonstrates the capabilities of a diode laser gas analyzer designed to be operated by non-specialist personnel. Most air monitoring systems with really superior performance (such as the first two examples) are routinely operated by well trained scientific personnel.

These examples cover all possible questions as to the practical applicability of the technique, showing it also to be sensitive, selective, and fast. At present, however, tunable diode laser gas analyzers for high performance air monitoring are custom made, custom adopted, or even home-built, which makes them expensive and needing specialist operators. Tunable mid-IR diode laser spectroscopy is therefore used mostly for tasks that other techniques fail to resolve.

The following section introduces the basic technical and scientific features of these instruments, and compares them with other, competing techniques.

2 TUNABLE DIODE LASER ABSORPTION SPECTROSCOPY BASICS

In essence, tunable diode laser absorption spectroscopy (TDLAS) is just another spectroscopic technique for gas analysis. However, whereas conventional IR spectroscopy uses a thermal radiation source combined with a spectroscopic filter (such as an interference filter, a monochromator, or a Fourier transform spectrometer), the diode laser radiation source is itself tunable. This achieves gains in spectral resolution, spectral power, and speed and ease of tuning. Also, very good local resolution is obtained which is diffraction limited, thereby offering the potential of a superior long optical path. One loses wide access to spectral data due to the comparably small total spectral width available from a single laser. Use of laser radiation leads to unwanted interferences due to the high optical quality of the beam, which may get annoying in some applications, and one pays for the gained laser advantages by the fact that present prices (1999) of systems are high.

The main factor affecting cost is that of mass production. Single diode lasers of good stability cost between €2000 and €4000, if made in quantities of less than a few hundred per year. Typical spectroscopic applications of TDLAS require spectral resolution of the radiation. The laser drive current also has to be extremely stable and the laser temperature stable to some millikelvin in order to achieve laser stability to 10^{-5} of the wavelength. Although such requirements can be met by dedicated laboratory electronics, and might well be attained in mass production, the price of the corresponding electronic equipment is of the order of some tens of €1000.

Systems for atmospheric analysis with TDLAS are today made in small quantities or single units, and the production costs (often including special requirements) range from some €50 thousand to €1 million. At present mass production cannot be expected, but the market is slowly growing.

Returning to technical and scientific topics, good spectral resolution is one of the keys to laser spectroscopy success. IR gas spectra of the smaller molecules are composed of many lines. These lines are broadened by molecular movement relative to the radiation propagation. The corresponding Doppler line width depends on temperature (molecular motion increases with thermal energy) and on the molecular mass (smaller molecules with the same energy travel faster). At room temperature, typical Doppler line widths are of the order of several 10^{-6} of the peak wavelength. A wavelength of $\lambda = 10\,\mu m$ corresponds to a wavenumber $\tilde{\nu} = 1/\lambda = 1000\,cm^{-1}$. The lines are further widened by collisions between molecules. At atmospheric pressure this width is several 10^{-4} of the peak wavenumber. This width is proportional to pressure, and at atmospheric pressure frequently leads to the overlap of neighboring lines, so that spectral bands are formed rather than isolated lines. Also, lines from different gas species present may overlap. Therefore, selectivity of spectroscopic analysis is hindered by line width.

An example of molecular line structure is shown in Figure 1. The lines shown are calculated numerically, and correspond precisely to experimental observations. The line width of this light molecule (HCl) is much narrower than the line separation, hence even a medium resolution spectrometer would be able to resolve these lines. The main lines are found close to $\lambda = 3\,\mu m$ or $\tilde{\nu} = 3000\,cm^{-1}$ ($3333\frac{1}{3}\,cm^{-1}$, to be exact). Higher harmonics of these are observed at a half and a quarter of the wavelength and can also be used for atmospheric analysis, albeit at a sensitivity reduced by roughly two orders of magnitude per harmonic step. This observation can be generalized – most molecules of interest absorb strongly and specifically somewhere in the wavelength range 3–15 μm (corresponding to 700–3000 cm^{-1}), and they have corresponding absorption replica at shorter wavelength, equivalent to higher quantum energy or wavenumber. As mid-IR equipment for the wavenumber range of the highest absorption sensitivity is more costly and somewhat more demanding technologically, harmonic lines are used whenever possible. However,

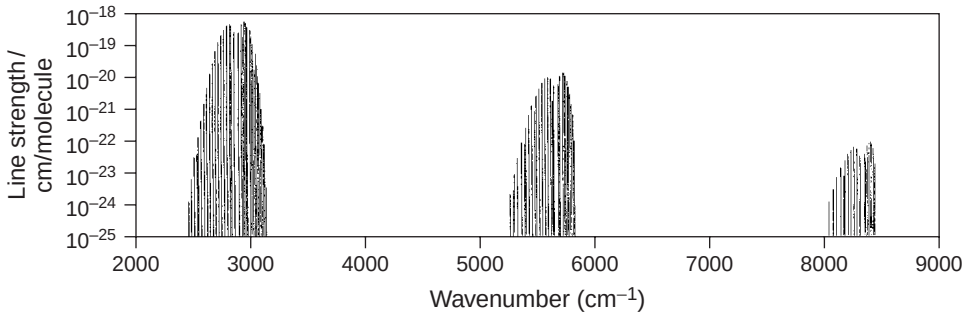

Figure 1 Molecular absorption lines of HCl at room temperature. The line strength is proportional to the absorption coefficient (and proportional to the absorbed amount of radiation at small absorbances). The strongest absorption and hence the highest detection sensitivity is observed close to 3 μm wavelength around the wavenumber of 3000 cm^{-1}.

atmospheric analysis typically is very demanding in respect of sensitivity and selectivity, and hence the 700–3000 cm^{-1} region is used nearly exclusively, because it allows detection limits well below parts per billion (ppb). The qualitative results and conclusions usually hold for the near-IR as well, or are easily adapted.

The heavier molecules have more numerous vibration and rotation freedom and hence more lines than HCl, the example of Figure 1. For typical atmospheric analysis, the lines of the various gas constituents are densely packed (Figure 2) and overlap due to their line width. For this reason, absorption data are usually taken at reduced pressure, so that the pressure broadening is negligible. The residual Doppler width is of the order of some 10^{-3} cm^{-1}, and usually enables a line to be found for detection of a molecule of interest, isolated from the other lines of the atmospheric sample. Doppler broadening is more benevolent to the spectroscopist in that the corresponding line shape drops exponentially from the maximum, whereas pressure broadening leads to an inverse square reduction of the line wings with separation from the line center, and hence leads to more overlap. An example of a problematic mixture is shown in Figure 3. It is obvious, that detection of NO is hindered strongly by overlapping absorptions of the other gas components, that have orders of magnitude higher concentrations. At reduced pressure, the absorption lines are well separated and become accessible to high selectivity and high sensitivity analysis by measurement of spectral absorption.

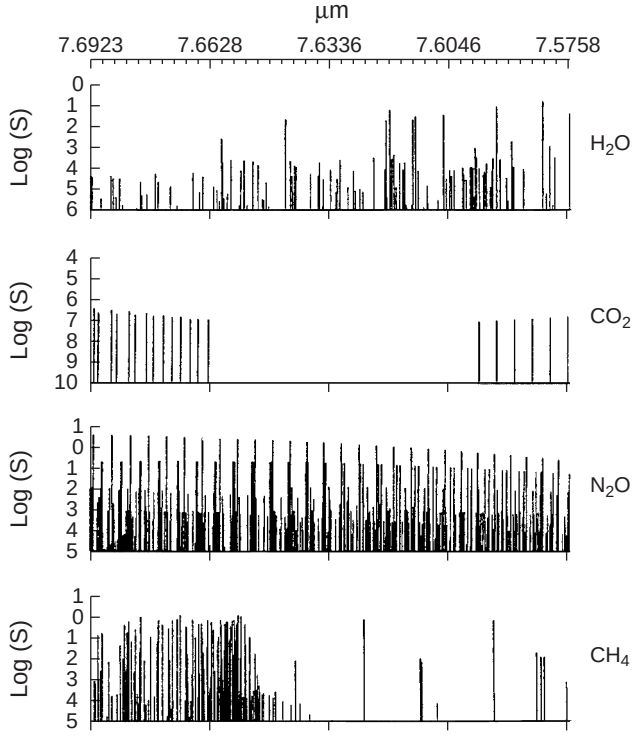

Figure 2 Molecular transitions of some relevant compounds in a limited wavelength range. (Taken from ref. 1.)

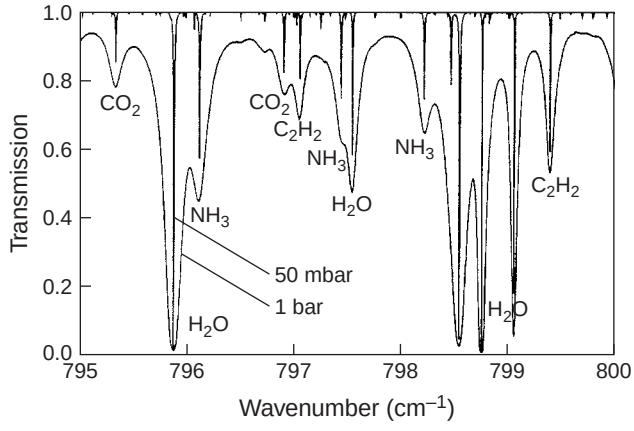

Figure 3 Transmission spectrum of a gas mixture typical for emissions of technical combustion containing H_2O, CO_2, NH_3, and C_2H_2 in N_2 at 470 K and 1 m optical path. (Taken from ref. 32). Such a mixture can be found in combustion exhaust to the atmosphere. While at atmospheric pressure the absorption lines merge, at low pressure (50 mbar) lines are only Doppler broadened and well isolated.

The key to such an analysis technique is equipment that allows a factor of 10^{-6} spectral resolution. Lower resolution would make the low line width absorption advantages inaccessible, as they result in large system line width. Although 10^{-6} resolution can be reached with the most sophisticated Fourier spectrometers, such equipment does not allow routine atmospheric analysis because the measurement speed is low and takes hours rather than minutes for most tasks of interest. Diode lasers lend themselves well for atmospheric analysis because they can be made and operated to have corresponding emission band widths and because they are easily tunable by changing the drive current. The principle of operation of a corresponding laser system is simple, as shown in Figure 4. The laser radiation is fed through a sample optical path, and is tuned across absorption lines. This is done in order to discriminate background effects from the absorption line. The laser is tuned repetitively and at high speed, so that any background effects that may vary with temperature are effectively frozen in for a single sweep; single-sweep evaluation and integration for noise reduction then leads to minimum signal drifts.

The true details and operation parameters of such systems are more complicated than the scheme of Figure 4 shows. For instance, the laser power depends on the drive current. From a given threshold current, the laser emits radiation, the intensity of which increases from close to zero to maximum powers of the order of a few 100 μW in the wavenumber range of interest. Laser powers are often specified well above this value and

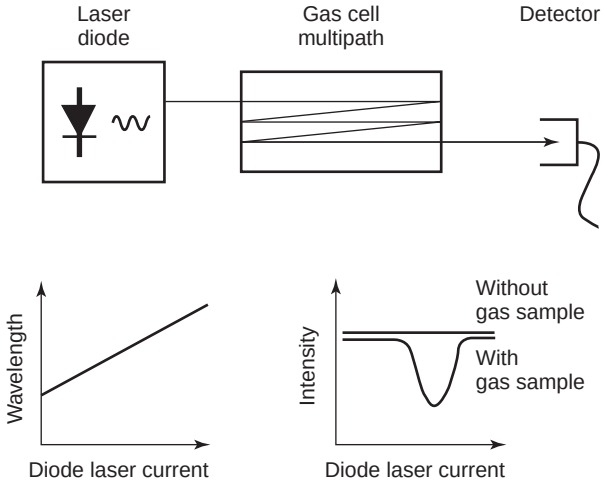

Figure 4 Operation principle of a diode laser spectrometer for atmospheric analysis. The laser is tuned by change of its drive current, the laser radiation fed into an optical path through the atmospheric sample of interest, and the laser intensity is recorded as a function of current, thus indicating the concentration of the molecular species of interest by the specific absorption line spectral position and strength.

reported to many milliwatts; however, such power levels are gained with laser designs not suited for atmospheric measurement purposes. The increase of laser intensity with drive current gives the baseline without radiation an ascending slope, so that it is not horizontal, as the figure. Various techniques are used to separate the line absorption depression from the baseline variation. One consists of sampling the transmission at the line peak and at two points located symmetrically either side of the line of interest, where negligible absorption is found. The latter are then used to extrapolate the baseline and to determine the base point at the peak.

Another technique consists of modulating the current sinusoidally (or similarly), and then analyzing the transmission signal with a lock-in amplifier tuned to twice the modulation frequency. This method is called $2f$ method, as many commercial lock-in amplifiers had a special '$2f$' button for changing from the reference frequency to an internally generated second harmonic. This picks out the curvature of a graph of the spectral transmission curve; that is, the resulting signal is zero for any signal which varies linearly with the wavenumber. As the power ramp is very close to being linear with current (and hence with laser wavelength), this variation is effectively suppressed, and only the extra curvature due to the absorption line shows up in the processed transmission signal.

Another alternative is to record a complete neighborhood of the line under consideration. Then in a second numerical data processing step, a ramp is fitted and subtracted from the data. Within the same procedure, a line fit from stored data can be obtained in order to extract the information of interest – the line position is used for gas recognition and the line strength for concentration measurement. With improving computing power and electronics, this technique is found in an increasing number of laser gas analysis systems.

The lasers have a tuning rate of the order of $0.1-1\ \mathrm{cm}^{-1}$ per milliamp of drive current. This is what makes tuning so easy compared with conventional means, such as rotation of a diffraction grating. However, in order to keep the laser radiation at a stable 10^{-6} match to the absorption wavenumber, the current must be stable to within a few microamps.

The laser radiation wavenumber also depends on the laser temperature. This is due to material properties, that inherently depend on the temperature. The laser emission temperature drift is of the order of inverse centimeters per Kelvin temperature drift. As a consequence, laser temperature has to be stabilized to millikelvin levels. As the passive stability due to independent stabilization of the current and temperature is, in most cases, found insufficient for long-term system stability, most laser spectroscopy systems dedicated to atmospheric analysis have active stabilization. This is usually achieved by

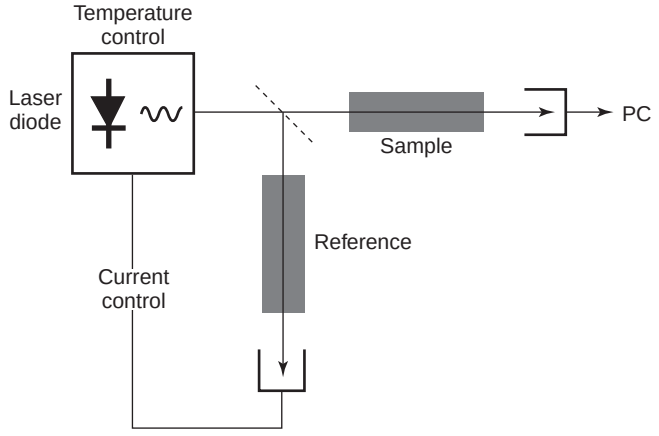

Figure 5 Atmospheric analysis diode laser systems usually are stabilized by temperature stabilization of the laser, stabilized current supplies, and active stabilization to reference absorption obtained with a transmission cell filled with the gas of interest.

splitting some of the radiation away from the main beam and guiding it through a reference cell containing the gas of interest in concentrations leading to easily detectable absorption (Figure 5). Electronic or numeric stabilization components evaluate the reference signal and add stabilization currents to the drive current such that the laser radiation always sweeps a well-defined wavenumber range.

Stability is not only limited by laser wavelength drifts. The background spectrum may also change with time. A typical problem of high-sensitivity systems is optical interference due to scattered radiation. Assume for instance that 10^{-8} of the power received at the detector is scattered radiation that traveled a distance of a few meters difference to the main beam. The corresponding fields of the direct and scattered beams overlap and give rise to interference structures. The fields vary as the square root of the power, and hence the field modulation is of the order of 10^{-3}, and comparable 10^{-3} modulation is found on the power and the detector signal. Such modulation has periodicities of $10^{-3}\,\mathrm{cm}^{-1}$ for the above mentioned path difference of a few meters, and hence may become similar to the absorption lines. High-sensitivity systems cope with such effects by sequentially filling the sample cell first with nitrogen or synthetic air for background measurement. Synthetic air is preferable, as it resembles atmospheric air more closely and models, for instance, H_2O contributions to the transmission interferences. An alternative is scrubbed ambient air, where the species under investigation is removed by a selective scrubber. The second step is filling the sample cell with calibration gas containing the target gas species in low concentration for calibration. Finally, the sample atmosphere is introduced to the cell. This is done automatically, and the final signal is gained from a numerical treatment of all three signal components, which reduces background effects and optimizes calibration stability.

Clearly, the relative presence of spurious signals will limit the detectivity. Generally, the minimum detectable concentration is a direct consequence of the minimum detectable change of the transmission T, which is connected to the gas absorption constant α by Lambert–Beer's law, $T = \exp(-N\alpha L)$. This states that transmission varies with the exponential function of the product of the gas concentration N, α, and the optical pathlength L. One way to optimize sensitivity is to maximize the optical length L, a point that is returned to later on. For a given length L, the minimum detectable concentration ΔN is calculated from Lambert–Beer's law to depend on the minimum detectable transmission change ΔT by $\Delta N = \ln(\Delta T/\alpha L)$. In typical situations, $\Delta T/\alpha L$ is much smaller than unity, so that as an approximation $\Delta N = \Delta T/\alpha L$ holds well. If one wants to compare system performance for different systems with different optical cell length and different ΔT, one uses a system parameter ΔNL, usually given in the dimension ppm m. This parameter obviously only depends on the gas specific absorption constant and the system transmission resolution.

The minimum detectable transmission change relates directly to the minimum detectable intensity step which is resolvable in the detector signal. In general, this step is equal to the signal noise under operational conditions. The signal noise has various sources, the most prominent of which is the noise of the detector itself. This can be observed with the optical path blocked. This noise is limiting because the power of thermal radiation sources is comparatively low, so that the signal due to spectrally filtered radiation is only some orders of magnitude larger than the detector noise. The optical beam power is proportional to the spectral power of the radiation source and the radiation bandwidth; if one wants to use thermal sources and the 10^{-6} spectral resolution that is sought with high sensitivity and high selectivity in mind, then typical integration times in order to get acceptable transmission measurements would be of the order of hours to days, which is totally unacceptable. This is where the high spectral power of diode lasers comes in. In a well-designed laser gas analysis system, the optical power is so high that the signal-to-noise level is no longer dominated by the constant detector noise.

The next noise source to be taken into consideration is the intensity fluctuation due to the laser radiation source itself. This noise is such that 10^{-7}–10^{-8} signal-to-noise levels can be reached regarding the source only, and this level makes the theoretical sensitivity level for transmission measurements, ΔT, equal to 10^{-7}–10^{-8}. Such signal-to-noise levels have often been

reported by laboratory models of diode laser gas analysis systems. However, realistic systems, as used for atmospheric investigations, never attain such precision due to inevitable adverse environmental parameters such as changing temperatures or acoustic vibrations. At present (and in the foreseeable future) a ΔT range of 10^{-4}–10^{-6} is realistic. The dominating system noise is usually identified to be the unwanted optical interference signal components discussed above.

In order to give an impression of the sensitivities resulting from such noise levels, typical values are noted in Table 1. These values are theoretical, and derived from the known absorption constants. A ΔT value of 10^{-4} is used; with a very high performance system $\Delta T = 10^{-6}$ can be obtained, which improves sensitivity by a factor of 100. The table is based on absorption data both at the longer wavelengths beyond 3 μm with optimum sensitivity, and at shorter wavelengths in the near-IR between 1 μm and 2 μm which allow the use of more readily available components. The direct comparison shows the sensitivity advantage of the longer-wavelength IR to be 2–5 orders of magnitude. In comparing, note that the numbers are given in ppm for the near-IR and ppb for the mid-IR, a difference of three orders of magnitude.

For the high sensitivities demanded by atmospheric analysis, long optical pathlengths are necessary. There are optical limits to the practical dimensions of the pathlength. Basically, all optical instruments with a long optical path have a radiation source of a given diameter a (or an equivalent internal aperture of diameter a), that is matched to the parallel open beam by an optical element with focal length f. Even neglecting diffraction, such an element cannot produce a parallel beam; this would only be possible for a light source of zero extension in the focal spot of the collimating optics. The best one can do in real systems is to image the source aperture to a plane at the end of the sample optical path L. This image will be magnified by the ratio L/f. Let us assume a desirable pathlength of 100 m (typical for high sensitivity atmospheric gas analysis systems), and an aperture a of 1 mm with thermal radiation source. Such an aperture is the lower limit acceptable for optical power reasons. In this case a typical $f = 10$ cm lens will generate an image with $A = 1$ m diameter. The receiving optics thus must be this large in order to catch all the signal power. A laser can be made to have an emitting aperture of some micrometers, which results in an image spot of some millimeters. This is another major advantage of laser transmission spectroscopy.

In practice, beam widths are larger than discussed above for diffraction reasons, but qualitatively the above arguments do hold. As mentioned above, the gas samples should be contained in sample cells that allow a suitable pressure reduction. Such cells must be limited to volumes of some liters, to allow acceptable time constants due to the gas pumping speed. Long pathlengths are obtained by folding the optical beam with mirror systems, so that the gas volume is traversed many times by the beam. Pathlengths of 100 m are readily available, but the above beam-spreading argument also holds for the long-path optical cells that must be very large in diameter for conventional radiation sources. Thus practical considerations regarding closed sample cells also favor strongly the application of diode lasers for spectroscopy.

Long open optical pathlengths in the field are attractive for atmospheric measurements, as the data are representative for a whole area, as compared to sampling the gas locally at the system gas inlet. Also, closed gas sampling systems may induce artefacts; for example, some gases adhere strongly to technical tubing material or the gas cell walls and thus give rise to residual gas concentrations if the gas component is no longer in the ambient air, but freed from the inside walls. As discussed before, cross-sensitivity due to overlapping lines will hinder such measurements at atmospheric pressure. Corresponding data are contained in Table 2. It is worthwhile noting in comparison with Table 1, that ΔT of the order of some 10^{-5} was achieved with the low-pressure systems.

Table 1 Minimum detectable concentration of some gases for a system transmission resolution of $\Delta T = 10^{-4}$, comparing the best suited absorption lines in the near-IR and the mid-IR. With one meter of optical path, the numbers give the detection limit in ppm for near-IR and ppb for mid-IR; at 100 m pathlength, the detection limits are lowered by a factor of 100. (Taken from ref. 2)

Molecule	Minimum detectable concentration	
	Near-IR (ppm · m)	Mid-IR (ppb · m)
CO_2	470	2
CO	360	10
CH_4	7	100
HCl	1	10
NH_3	20	20
H_2O	0.6	14

Table 2 Experimental values of system sensitivities in the mid-IR with open pass and reduced pressure gas cells. Original data are scaled to 100 m pathlength. Roughly two orders of magnitude are gained in sensitivity by the use of closed cells and pressure reduction. (Taken from ref. 2)

Molecule	Minimum detectable concentration (ppb · m)	
	Open path	Reduced pressure
CO	40	0.2
SO_2	80	1
HNO_3	18	0.13
NH_3	20	0.1

In sum, a typical gas analysis system consists of a radiation source component including temperature control of the lasers. Separate lasers are usually provided for each gas component to be detected. Mid-IR lasers must be operated at low temperatures, typically close to that of liquid nitrogen; appropriate cooling is provided by liquid nitrogen cryostats or closed cycle coolers. The optical beam is formed by mirror optics for minimum unwanted spurious reflections that might originate from lens surfaces. A reference beam with reference cell serves for laser stabilization. The gas cell has a folded optical beam of around 100 m optical pathlength for a cell below a meter in length; the cell volume of one to a few liters allows about one minute gas change times. The detectors are of low noise and usually cooled to below 100 K by standard equipment. Signal analysis and system control by computers is standard. Such systems achieve $\Delta T = 10^{-4}$ sensitivities in unattended operation, and ΔT below 10^{-5} for the best routine atmospheric analysis equipment.

3 APPLICATIONS

3.1 Review of Monitoring Applications

In recent years TDLAS has found widespread application in the field of atmospheric science. The TDLAS technique meets the major requirements for atmospheric trace gas monitoring: sub-ppb sensitivity, high detection speed, and the potential for simultaneous in situ measurements of several compounds. Researcher laboratories, such as at the Max Planck Institute for Chemistry in Mainz, Germany, have developed own multilaser TDLAS systems and used them extensively for trace gas detection throughout the troposphere and the lower stratosphere on ground-based, ship-based and airborne platforms.[3] Highly specific instruments have been adapted for the direct measurement of trace gas fluxes by the application of micrometeorological techniques[4] and for the measurement of the ratios of the stable isotopomers in methane from various sources.[5]

Operation of these instruments in the field, especially on airborne or ship-based platforms, implies a special design to meet the stringent demands of vibration isolation, shock resistance and weight minimization. Typically, modern systems are computer controlled, multilaser systems which produce pseudosimultaneous measurements on up to four channels by interrogating each laser in turn using different time multiplexing devices.

Table 3 lists trace gases relevant to the chemistry of the atmosphere. Typical mixing ratios for these species in the atmosphere are of the order of ppb (10^{-9}) to ppt (10^{-12}). The table lists detection limits for a TDLAS

Table 3 Theoretical estimates of the detection limit of a TDLAS field instrument

Gas	Wavelength, λ (μm)	Wavenumber, ν (cm^{-1})	Line strength, S (10^{-19} cm per molecule)	Line width, γ_{Voigt} ($10^{-3}\,cm^{-1}$)	Detection limit (pptv)
OH	2.87	3484	0.37	5.00	1800
HO_2	7.08	1411	0.12	3.63	970
H_2O_2	7.79	1284	0.45	3.52	256
NO	5.33	1876	0.34	3.26	278
NO_2	6.25	1600	2.18	2.90	40
HNO_3	5.90	1694	0.29	3.75	430
N_2O	7.69	1301	1.56	2.77	56
NH_3	10.74	931	5.20	2.92	18
O_3	9.50	1053	0.42	2.64	206
CH_4	7.56	1322	0.53	3.25	180
CO	4.65	2151	1.89	3.85	58
HCHO	3.56	2781	1.19	5.17	128
SO_2	7.29	1372	0.49	3.86	264
OCS	4.87	2053	10.30	3.55	10
HF	1.24	7856	0.76	12.09	400
HCl	3.40	2945	5.03	4.30	24
HBr	3.78	2649	0.45	2.90	192
HI	4.39	2278	0.02	2.21	3200
ClO	11.60	860	0.08	2.73	1200
CH_3Cl	7.52	1330	0.30	4.12	600
CO_2	4.40	2370	0.33	3.51	312

The estimate is based on the system parameters: OD, optical density = 10^{-5}; SNR, signal-to-noise ratio = 1; $\Delta f = 0.01$ Hz (30 s integration time); $L = 126$ m, $p = 30$ mbar.

measurement based on a detectable minimum OD of 10^{-5}. This is a typical in-field sensitivity obtained for high-performance systems. To compare these theoretical estimates based on the line strength of the individual rotational-vibrational transitions, Table 4 lists measured sensitivities for various TDLAS instruments during field employment to detect species with sub-ppb mixing ratios in the atmosphere. In general, there is good agreement between theoretical and deduced detection limits. The table demonstrates that the diode laser atmospheric analysis is highly specific and sufficiently sensitive to measure the concentration of a large number of trace gas species in the atmosphere at sub-ppbv levels.

The inherent capability of the technique to perform high-precision measurements at time resolutions of less than 1 s permits the use of TDLAS systems for airborne tracer measurements to study dynamic processes in the stratosphere (section 3.2) and direct determination of trace gas fluxes by the application of micrometeorological techniques. Trace gas fluxes are commonly deduced from the gas-chromatographic measurement of the concentration increase in closed chambers placed at ground level. The shortcomings of this technique are the limited spatial coverage of the boxes and the uncontrolled disturbance of the ecosystem's microclimate caused by the closed box. To overcome these drawbacks micrometeorological techniques can be applied, such as the eddy covariance technique, which derives the trace gas flux from the covariance between the vertical wind speed and the fluctuations of the trace gas concentration at a given height above the ground. Alternatively the flux gradient technique can be employed, where the trace gas flux is determined from the measured concentration gradient of the emitted trace gas at several different altitudes above the ground and the mean atmospheric transfer coefficient, which can be deduced from momentum and heat flux measurements. The advantage of these techniques is that they average the fluxes over larger areas while not disturbing the microclimate of the ecosystem. In order to make use of such techniques, extremely fast and precise trace gas detectors are needed for eddy covariance measurements, criteria that are both met by TDLAS systems. Tunable diode laser spectroscopy has the potential for eddy covariance measurements of trace gases such as CH_4, N_2O, CO, NH_3, or COS.

Table 4 Measured sensitivities of TDLAS instruments during field employment

Species	Reference	Sensitivity (ppt)	Time constant
NO_2	6	25	70 s
	7	50	1 min
	8	25	10 min
	9	20	5 min
	10	150	3 min
	11	100	
	12	25	150 s
	13	27	120 s
HCHO	14	25	30 min
	10	300	3 min
	6	250	3 min
	15	100	30 min
	6	70	70 s
	13	22	120 s
	16	31	25 s
H_2O_2	14	75	30 min
	17	600	4 min
	15	150	60 min
	6	400	2 min
	18	50	60 min
	13	177	120 s
HNO_3	10	300	3 min
	11	100	
	7	300	1 min
	15	150	30 min
	8	75	10 min
NH_3	19	100	4 min
	20	100	5 min
HCl	9	50	1 min
	19	100	1 s
	21	40	30 s
SO_2	7	300	1 min
NO	7	300	1 min
OCS	22	5	2 min

Table 5 gives an estimate of quantities relevant to the measurement of fluxes by eddy covariance using a TDLAS, which are the typical ambient air concentration c, expected fluxes F, absorption line wavelength n and strength S, OD at concentration c, required as well as measurable precision R and the ratio r of both. The numbers are derived from the formulae given in the table footnote. The measurable precision R is estimated based on the center absorption of Doppler-broadened lines. Typical pressure broadening is represented with a Voigt parameter of 0.5. The Voigt parameter is a measure of the similarity of the line shape to Doppler-broadened line shape or pressure-broadened lineshapes. Intermediate line shapes are said to have a Voigt profile and can be described by a special line-shape formula that makes use of the Voigt parameter. The required R value is estimated from Gaussian error propagation of the covariance of wind and concentration data with neglected wind measurement error. For a half-hour measurement with 10 Hz, the standard deviation of the vertical wind fluctuations is assumed to be $0.2\,m\,s^{-1}$. Errors introduced by micrometeorological variables are not considered, meaning that the ratio r should be well above unity for satisfactory measurements. Whereas NH_3 seems to be detectable from a spectroscopic point of view, COS might be beyond the scope of today's instruments.

The high selectivity of TDLAS measurements can be used to study the isotope ratios for CH_4, N_2O,

Table 5 Detectivity of typical trace gas fluxes by eddy covariance using a TDLAS. For explanation see text

Gas	c (ppbv)	F ($\mu g\,m^{-2}\,h^{-1}$)	n (cm^{-1})	S ($10^{-20}\,cm\,m^{-1}$)	OD (10^{-3})	R (ppbv)	Required R (ppbv)	r
CH_4	1700	5×10^3	1322	5.3	25	7.5	1.3×10^2	17
	1700	5×10^3	3067	21.3	43	4.3	1.3×10^2	30
N_2O	300	5×10^2	1300	15.6	21	1.5	4.7	3.1
	300	5×10^2	2236	100.4	80	0.40	4.7	12
NH_3	5	50	931	52.0	1.0	0.52	1.2	2.3
COS	0.5	5	2053	103.0	0.18	0.31	3.5×10^{-2}	0.1

Equations:
Required precision: $R = 670\,s\,m^{-1} \times \Delta F$ with $\Delta F = 0.1F$.
Achievable precision: $R(\text{ppbv}) = 0.12 \dfrac{\nu(cm^{-1})}{S(10^{-20}\,cm\,molecule^{-1})\sqrt{M(g\,mol^{-1})}}$.
Achievable precision requirements: OD $= 1 \times 10^{-4}$, pathlength $= 18$ m, cell pressure $= 50$ hPa.

or O_3. Measurements of the stable isotopomer ratios $^{13}CH_4/^{12}CH_4$ and $^{12}CH_3D/^{12}CH_4$ in atmospheric methane sources are an important tool with which to establish the global atmospheric methane budget and furthermore provide important information about methane production and consumption processes occurring within the various sources. At the above-mentioned Max Planck Institute, for example, a methane isotope spectrometer (MISOS) based on the TDLAS technique has been developed and directly applied to isotope ratio measurements on environmental samples collected from natural wetlands, rice paddies, landfills, and natural gas. Evaluation of the isotopic composition is based on the simultaneous measurement of a pair of suited absorption lines for $^{13}CH_4$ and $^{12}CH_4$ or $^{12}CH_3D$ and $^{12}CH_4$, respectively, which are selected to give similar optical densities on one single laser scan.

The main advantage of the TDLAS instrument, compared to mass spectrometry (MS), is its ability to perform direct measurements on methane samples without the need for any sample preparation (apart from drying) or for chemical conversion of CH_4 to CO_2 and H_2. A comparison between TDLAS and MS-based determinations of the isotopic signature of various samples from natural wetlands, rice paddies and landfills for $\delta^{13}C$ and for δD ($\delta^{13}C$ and δD = isotope ratio in the sample relative to the V-PDB standard in ‰) showed an excellent agreement. The mean deviation in $\delta^{13}C$ between TDLAS and MS was 0.5‰ which is comparable to the reproducibility of the TDLAS measurements (±0.5‰). The mean deviation between TDLAS and MS for δD is 5.1‰, whereas reproducibility for the TDLAS measurements is ±2‰ and ±3‰ for the MS measurements. The $\delta^{13}C$ and δD precision and accuracy of MISOS is therefore well suited to the investigation of atmospheric CH_4 sources, for which the isotopic variation is typically markedly larger.

3.2 Airborne Monitoring

3.2.1 Introduction

This section describes as a typical example the set-up and deployment of an aircraft-based tunable diode laser spectrometer which was developed during 1995 and 1996 and which, since early 1997, has been used extensively in a variety of airborne atmospheric measurement campaigns. The TRISTAR (tracer in situ tunable diode laser absorption spectroscope for atmospheric research) instrument has the ability to measure three atmospheric trace gases simultaneously at high time resolutions (below 1 s) using an individual laser diode for each component. The main target species to be measured with TRISTAR are atmospheric components of low chemical reactivity, such as carbon monoxide or nitrous oxide, whose chemical lifetime is comparable or long as compared to that of atmospheric transport or mixing processes. The abundance of these so-called atmospheric tracers and their correlation with each other thus reveals information about the origin of the air mass encountered during the measurement. This information can serve to support other observations of in situ processes, such as atmospheric particle formation, or it can be of principle interest, for example if the nature of atmospheric exchange processes is studied.

In contrast to true atmospheric trace components – chemically active species (such as nitric oxide, formaldehyde, or hydrogen peroxide) with volume mixing ratios usually near or below 1 ppbv – the atmospheric tracers are present in the atmosphere in the range from a few tens to a few hundred ppbv, or even up to 1700 ppbv mixing ratio as in the case of methane. For TRISTAR this implies that ultimate measurement sensitivity is not required in terms of low detection limits, rather the instrument and the data postprocessing aim to achieve high absolute measurement precision (% range) on a high mixing ratio background of the target species.

In the following, the TRISTAR set-up is described with attention given to the optical, electronic and data acquisition layout of the instrument, the gas flow system which allows in situ calibration, and the data postprocessing that uses the in situ calibration and related diagnostics data. More detailed information can be taken from an instrument-related publication.[23] TRISTAR has been used on two different aircraft types: the German DLR research aircraft Falcon, and a Cessna Citation research aircraft run by the Dutch NLR. Some exemplary results of the measurement campaigns carried out are presented at the end of this section.

3.2.2 Optical Set-up

The optical bench of TRISTAR is a mechanically independent module which can be mounted according to the space requirement of the aircraft used. In the Falcon aircraft, the optics are located on top of a 19 in rack containing the electronics and the gas flow system. Due to the limited cabin height of the Cessna Citation, the optics module is carried by a support structure holding the gas flow system only, with the instrument electronics contained in a separate rack (Figure 6).

Operation of a TDLAS system, especially on an airborne platform, requires careful design to meet the stringent demands of vibration isolation, shock resistance and weight minimization. This affects in particular the optical and mechanical set-up of the system. For TRISTAR a very rigid optical layout was developed consisting of a $60 \times 45 \times 3\,cm^3$ aluminum board recessed for weight reduction, a single liquid-nitrogen cryostat, housing for three lasers and two detectors, a multi-pass cell based on White's design,[24] and ellipsoidal, parabolic and plane mirrors for imaging and steering the optical beam. Figure 7 presents a view of the optical bench facing the laser station side of the liquid nitrogen dewar laser cooler.

The principle of operation of the liquid-nitrogen cryostat is based on the evaporation of the liquid nitrogen

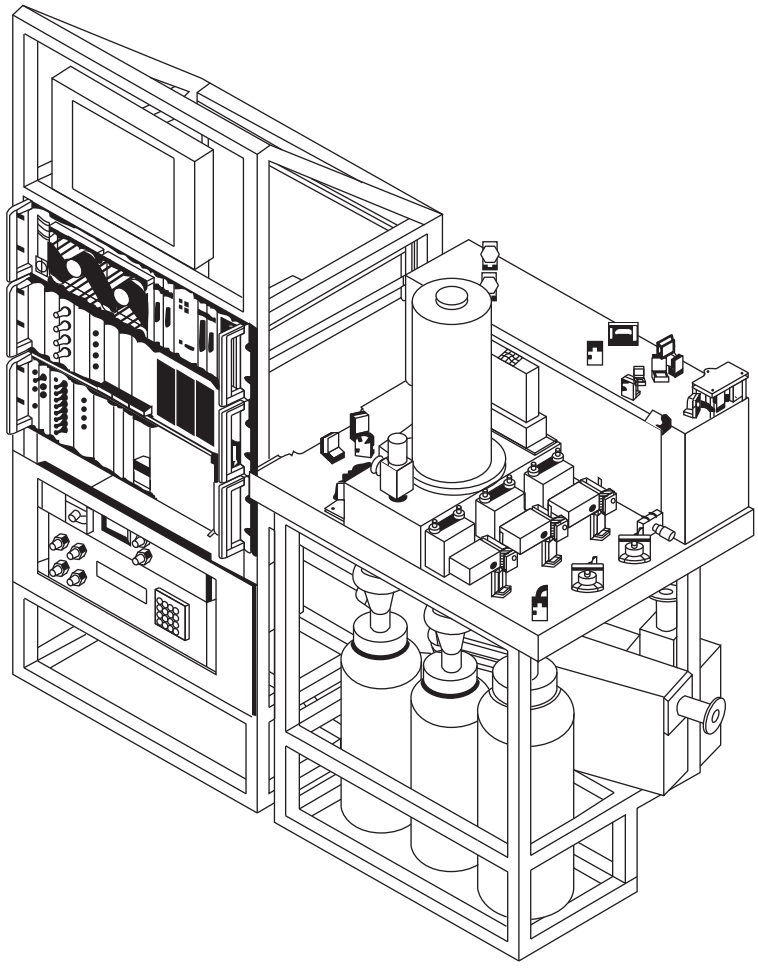

Figure 6 Installation of the TRISTAR instrument in the Cessna Citation aircraft. The left rack houses the instrument electronics, the optical bench is mounted on top of the right rack, which also contains the pump and the calibration gas containers of the gas handling system.

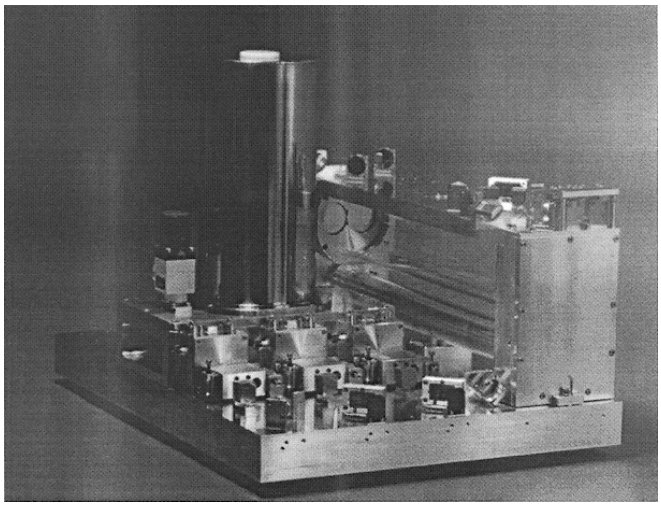

Figure 7 This image of the optical bench is taken from the laser station side of the liquid nitrogen dewar. The three laser cold stations are attached to the base section of the dewar with their lower part partly covered by the mirror objectives that image the emerging beams to an intermediate focus. At the front edge two beam combiners, placed at 45 degrees with their blackened beam dumps can be seen. The right part of the board is occupied by the measurement White cell, whose circular objective mirrors are clearly visible. The detector cold station is mounted on the rear part of the dewar base section, hidden in this image as well as the reference cells.

at the laser/detector cold station. By controlling the flow rate of the evaporated nitrogen the resulting cooling power of each cold station is approximately 1.5 W, independent of the actual operating temperature which can be varied between from near 80 K to 150 K by a heating device. The constant cooling power allows optimization of the electronic temperature controller over the whole operating temperature range of the diode lasers, exhibiting a temperature stability of better than 3 mK. Although the temperature operating point of the three lasers is set in three individual cold stations, the two mercury cadmium telluride detectors are operated in a single station at liquid nitrogen temperature.

Each laser is imaged through a mirror objective onto an intermediate focus producing a real image of the laser front facet. Precise coupling of the laser radiation into the optical path is achieved by aligning the mirror objective in three degrees of freedom, two in the image plane and one in the focal direction, while visually inspecting the front facet image of the laser diode through a crosshair ocular located at the intermediate focus. The rest of the optical path is aligned using visible radiation of a fiber-coupled laser diode, placing the fiber tip at the nominal intermediate focus location. The light emerging from the three intermediate foci is collimated by off-axis paraboloids and combined into a single beam, which is imaged through another intermediate focus into the measurement White cell. This cell contains the sample gas at reduced pressure and provides an optical absorption pathlength of up to 64 m through multiple reflection between the cell front and rear mirror groups spaced 0.5 m from each other. The gold coated mirrors provide reflectivity in excess of 99%. The light emerging from the cell is directed to an exit intermediate focus and then collimated again. The beam is then split with a 50% intensity fraction traversing a chain of reference cells. The cells contain the target gas at high concentration and produce a strong optical absorption signal which is primarily used to stabilize the laser emission wavelengths.

The direct signal and the reference beams are individually focused onto the signal and reference detectors, respectively, by off-axis parabolic mirrors.

3.2.3 Electronics and Data Acquisition

To evaluate the tracer concentration of the gas pumped through the measurement cell, the electronic output from the signal detector is recorded as a function of time while a well defined pulse on the laser current causes the emission wavelength to sweep across the absorption feature. A relative measurement is performed in the sense that this recorded signal is quantitatively related to a corresponding signal resulting from a prior calibration measurement, in which the ambient gas in the measurement cell was replaced by a standard gas of known target gas concentration during a calibration operation.

The key tasks of the instrument electronics and data acquisition are: first, to electronically process the detector signals in order to increase signal-to-noise and signal to background ratios; second, to establish a user interface to set and control the instrumental parameters (such as laser current and temperature), to monitor the absorption signals, and to switch either manually or through a predefined time schedule between measurement and calibration mode; third, to stabilize the laser emission wavelength to the absorption line by the determination of shifts in the spectral signal of the reference detector and corresponding minute variations to the laser current pulse and/or the laser operating temperature; and fourth, to analyze the spectral signals in real time deducing the trace gas mixing ratio and an error estimate from the measured absorption spectra by comparison to the spectrum recorded in the preceding calibration. The last task is governed by the conditions of the measurement situation – at typical aircraft speeds of 50–200 m s^{-1} a measurement time resolution of approximately 1 s is required to obtain adequate spatial resolution. For a typical measurement flight lasting 3–4 h it would be impractical to store all the spectral data for a later off-line analysis, and furthermore the opportunity to monitor the

measured trace gas concentration on-line during flight is of great advantage for the mission, allowing other activities to be triggered, such as to perform grab sample measurements or to even modify the planned mission schedule in case of unexpected events.

The TRISTAR electronic layout is shown in Figure 8. All electronic subunits are controlled by a digital signal processor (DSP) housed within an Intel 80486-based personal computer (PC), except for the diode laser controller. The laser bias current and temperature are set via a standard laboratory IEEE 488 link. To increase the signal-to-noise and signal to background ratios a phase-sensitive detection scheme is implemented. The optical frequency emitted by a tunable diode laser can be modulated efficiently up to the radiofrequency (RF) range by adding an electrical RF signal into the laser current. The optical sidebands produced by the frequency modulation of the laser are balanced in the case of absence of differential optical absorption. The absorption feature destroys this balance, resulting in an RF signal on the detector which is phase coupled to the RF laser feeding signal and thereby can easily be demodulated (lock-in technique). By this method, the signal generation is shifted in a frequency regime where laser, detector and electronic noise are significantly reduced due to the component's $1/f$ frequency characteristic. In the two-tone frequency modulation technique[25] used here, an electronic modulator on the laser side creates two RF signals, typically 105 MHz and 125 MHz, that are added to the laser current. On the detector side the signal is retrieved by demodulators (individually for the optical signal and reference paths) at the laser signals' frequency difference of 20 MHz. This technique offers some advantage in the reduction of the background signal caused by spurious laser amplitude modulation as compared to the method where a single RF signal is used for modulation and demodulation.

The high demodulation frequency also allows fast sampling of the absorption signals and thereby fast scanning of the laser emission across the absorption lines. To allow simultaneous measurement of the three target species, the three lasers are sequentially pulsed for 2 ms each with an individual pulse shape for each laser generated by the DSP. During the scan the absorption signal is recorded at a sampling rate of 100 kHz, while at the same time the preceding scan is processed. This includes an algorithm that determines the line peak position from the reference detector signal and imposes the adequate correction in the signal averaging and for the next pulse of that laser, following 6 ms later (line locking). Time-multiplexed absorption signals are thus averaged for all three lasers in basically the same timeframe, which is determined by the PC and synchronized to the aircraft board time.

Typically each full second the PC requests a data analysis from the DSP, which automatically starts a new signal integration. For the data analysis, the DSP optionally smoothes the spectrum.[26] If necessary, a background spectrum previously generated with neutral gas (dry nitrogen) in the measurement cell is subtracted and the resulting spectrum is used as a target for a linear least-squares fit of the calibration spectrum and a second-order polynomial. The fit parameters are converted to concentration values using the known mixing ratio of the

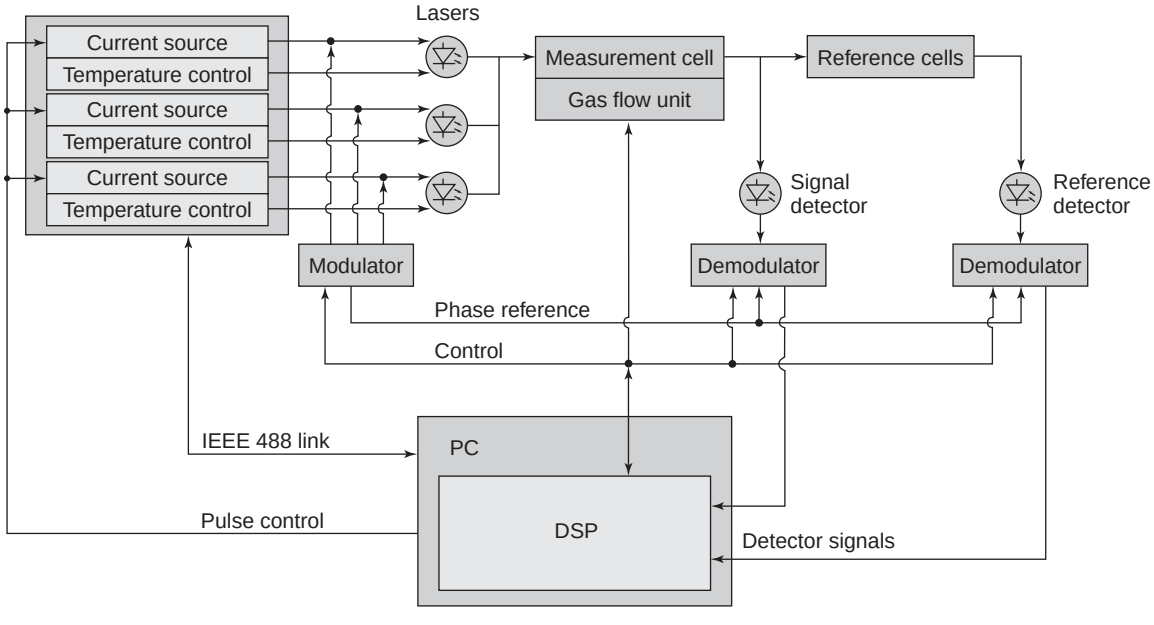

Figure 8 The electronic set-up of TRISTAR.

calibration gas, and transferred to the PC together with items of diagnostic data. The most important of these is a measure of the laser power determined from the detector signal difference sampled between the scan end and a subsequent 0.1 ms gap time, in which all lasers are off, i.e. not pulsed. While the averaged spectra and the resulting trace gas concentration values and their time series are displayed on the PC screen, the DSP generates the sampled signal waveforms on analog outputs to be displayed on an oscilloscope for diagnostic purpose. The calibration spectra, trace gas concentrations and diagnostic data are stored on the PC hard disk

3.2.4 Gas Flow and Calibration

The ambient air is conducted to the measurement cell by a gas sampling system which is fully included in the calibration, as indicated in Figure 9. The air is sampled with an inlet directed against the flight direction from outside the aircraft boundary layer through a 0.5 in PFA tube and expanded into the measurement cell through an adjustable critical orifice, which keeps the cell pressure at approximately 40 hPa independent from the outside pressure. The sample air is driven through the cell by a 16 $m^3 h^{-1}$ rotary pump at the cell exit. With the cell volume of 2.7 L a gas exchange time of 1 s is achieved. The calibrations are performed at intervals of 10–20 min depending on the flight schedule and typically last for less than 1 min. For this purpose, an excess flow of the calibration gas is introduced very close to the outside end of the intake line, with the surplus gas ejected outwards. A gas tank of compressed air provides nitrous oxide, carbon monoxide and methane at ground-level ambient mixing ratios which are calibrated before and after the measurement campaigns to a primary gas standard traced to a standard from the National Institute of Standards and Technology (NIST, Gaithersburg, USA).

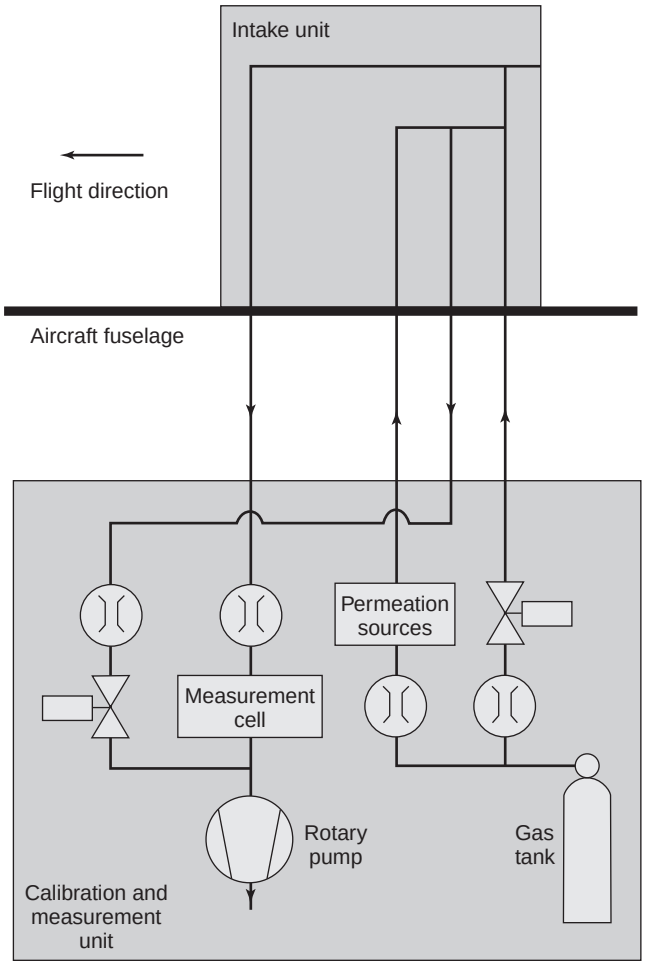

Figure 9 Gas flow and calibration system of TRISTAR.

The calibration system is also prepared for sticky or reactive gases such as H_2O_2, HCl, or HNO_3, using the constant mass flow of permeation sources. For stable operation of the permeation devices constant purging is required, and the gas lines need conditioning. Therefore, a continuous low gas flow is driven from the tank through the devices, circulated close to the calibration gas injection point of the sampling line and removed by the rotary pump when ambient air is measured. For calibration, a valve in the pump line is closed and the gas is added to the high flow of calibration gas from the tank. With this approach, no valves need to be placed remote from the instrument at the intake point or between the permeation sources and the calibration gas inlet, and possible loss of the calibration gas caused by wall reactions are minimized.

3.2.5 Data Postprocessing

To assure the quality of the measured data, the calibration spectra are analyzed on the ground. During a flight typically 10–20 calibration spectra are generated for each species. First, the line shape and shift of the spectra are inspected to ensure that the line locking procedure worked correctly and that the laser operating point was stable. Then, the signal height change is analyzed by a linear fit of each calibration spectrum to its successor. The deviation of the fit result from unity identifies the instrumental responsivity drift during the time interval between the two calibrations. Correlating this deviation with the corresponding deviation in the measured laser power reveals whether power changes caused the drift. These changes may arise from minute alterations in the optical alignment, due for instance to temperature changes in the cabin or in certain cases from changes in the composition of cabin air, which changes broad-band absorption in the optical path outside the measurement cell. A clear correlation between sensitivity and measured laser power will only arise if the laser operates on a single mode which, although desirable,

is not always achievable with the lasers available. In the case of single-mode laser operation, the data quality can be improved by appropriate normalization of the raw data to the laser power. For multimode operation, a drift in sensitivity between two successive calibrations is not necessarily correlated with the power signal. For multimode operation and mode competition noise, the drift appears to be of stochastic origin, and a linear time interpolation for the instrument response between two successive calibrations is carried out. Typically these corrections are smaller than 2%, which is the precision that can be quoted for the measurement data.

Since 1998 a further diagnostic parameter has been introduced by applying the same fit algorithm that is used to determine the ambient gas concentration from the signal spectrum to the reference channel. In multimode operation of the laser, the mode spacing is large compared to the absorption linewidth, therefore the fit result is a direct measure of the laser power present in the measurement mode – provided that the absorption in the reference cell is stable and large compared to the measurement cell, which in general is the case. The beam splitter separating signal and reference channel is placed behind the White cell, so changes in alignment or in mode competition affect both channels equally and the data can be normalized to the relative power of the measurement mode deduced from the fit of the reference cell signal.

The most prominent aspect of the data postprocessing is that drift-induced level shifts of the measured concentration time series before and after a calibration are significantly reduced. The calibrations are usually performed during stable phases of the mission schedule when no sudden change in the atmospheric concentration is expected.

3.2.6 Field Deployment

The TRISTAR system has been installed on the German research aircraft Falcon operated by the DLR, Oberpfaffenhofen, and the Dutch research aircraft Cessna Citation II operated by the Technical University of Delft and used during a total of six field missions in the years 1997 to 1999 to study dynamical and chemical processes in the atmosphere up to 13.5 km altitude. A first scientific mission with the Falcon was undertaken from Kiruna (Northern Sweden) to study the trace gas composition of the lower stratosphere, the development of hygropause clouds and particle formation in lee-waves (the POLSTAR mission). The TRISTAR system was used successfully to perform high-precision tracer measurements (N_2O, CO) in the tropopause region. Additional CO and N_2O measurements with TRISTAR, were obtained in March 1997 during the European Community-funded STREAM (stratosphere troposphere experiments by aircraft measurements) campaign in Kiruna. Successful operation of all three channels (CO, CH_4, N_2O) was achieved during the second Arctic POLSTAR expedition in January 1998. Further campaigns with the Cessna followed in March 1998 (LBA-CLAIRE: cooperative LBA airborne regional experiment) to study trace gas emissions over the Amazon, July 1998 (STREAM-Canada) to study biomass burning of the boreal forest in Canada, and in February–March 1999 during the Indian Ocean Experiment (INDOEX) to study trace gas distribution and transport associated with the Indian winter monsoon. Within these missions TRISTAR provided a large amount of trace gas measurements in the tropopause region in tropical, mid and high latitudes. These data give a wealth of information on chemical and dynamical processes that influence the chemical composition of the upper troposphere and lower stratosphere.

As an example Figure 10 shows measurements of N_2O and CO in the Arctic winter lower stratosphere obtained during the POLSTAR and STREAM expeditions in January and March 1997, respectively. Both species have sources in the troposphere and major sinks in the stratosphere, leading to nearly constant mixing ratios in the upper troposphere and decreasing concentrations in the lower stratosphere. The mixing ratio of N_2O in the troposphere measured during both campaigns is constant at a level of approximately 312 ppbv. The average free tropospheric CO concentration measured during the STREAM winter campaign in March 1997 was 147 ± 8 ppbv. This is significantly lower than CO concentrations measured 2 months earlier in the same location during the POLSTAR campaign. During POLSTAR average free tropospheric concentrations were 168 ± 18 ppbv. Above the tropopause N_2O and CO exhibit decreasing concentrations with increasing altitude. Although the altitude profile for N_2O is similar for both campaigns, CO

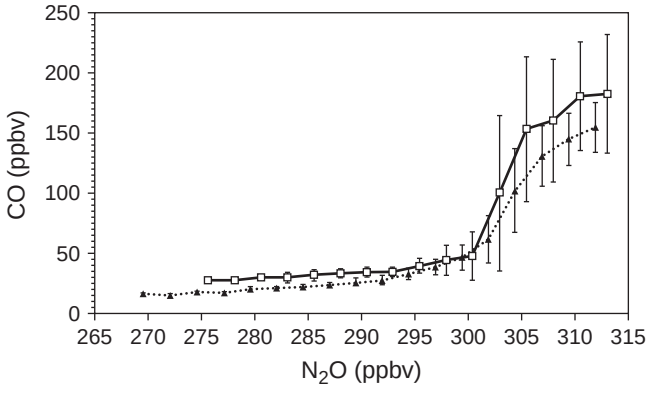

Figure 10 Average CO concentration as a function of N_2O for STREAM (filled triangles) and POLSTAR (open squares) data. The CO data have been binned according to N_2O. Notice that the two correlations have been shifted by 1 ppbv against each other on the *x*-axis to avoid overlap of data points in the figure.

shows a temporal trend towards lower concentrations in the lower stratosphere, similar to the trend observed in the troposphere. So the CO concentrations on the highest flight levels ($N_2O = 275–280$ ppbv) during POLSTAR were of the order of 30–35 ppbv, comparable to earlier observations, whereas CO concentrations at similar altitudes during STREAM decreased to 20–25 ppbv.

The general trend of decreasing CO concentrations in the lower stratosphere between January 97 (POLSTAR) and March 97 (STREAM) could be due to a seasonal increase of the OH radical concentration, leading to a higher CO loss rate later in the year. Observations demonstrate that tropospheric CO exhibits a strong latitude-dependent seasonal cycle, with CO enhancements of up to 100 ppbv at high northern latitudes in the winter compared to the summertime minimum. This seasonal variation, whose amplitude decreases with decreasing latitude, is mainly due to the seasonal variation of the OH concentration. It must be assumed, that the increase of the OH concentration from January to March will also influence the CO concentrations in the lower stratosphere, and can thus at least partly explain the observed temporal trend between POLSTAR and STREAM. In addition, mixing processes have to be taken into account, in particular transport of CO-poor air from the tropical tropopause region.

3.3 Rocket-based Experiments

3.3.1 Introduction

In the 1990s a tunable diode laser spectrometer was designed, built and tested to be flown on a sounding rocket in order to measure the abundance of water vapor and carbon dioxide in the upper atmosphere. This instrument was named MASERATI (middle-atmosphere spectrometric experiment on rockets for analysis of trace-gas influences). Details on the technical aspects of this instrument and first scientific results from the two rocket flights performed in 1998/99 have been published.[27] In the following discussion the most important scientific and technical aspects are summarized.

High-resolution in situ measurements of H_2O and CO_2 in the mesosphere and lower thermosphere (the atmosphere region from approximately 50 to 120 km above sea level) are of importance for our understanding of the atmosphere. The concentration of water vapor in the upper atmosphere is critically determined by photochemical processes and by transport through convection, advection, and turbulence. Therefore, a measured water vapor profile allows for a sensitive test of theoretical models that combine both atmospheric dynamics and chemistry. During summer, the existence of various layers such as noctilucent clouds and polar mesosphere summer echoes is controlled by the concentration of water vapor and temperature.[28]

Despite their scientific importance, in situ measurements of water vapor are rather sparse and contradictory.[29] Remote sensing by ground- or satellite-based techniques cannot achieve the spatial and temporal resolution required to study small-scale processes. Mass spectrometers have been flown on sounding rockets, but cannot be used for water vapor, as this trace gas (as do others), reacts with the surfaces inside the mass spectrometer which causes an unknown modification of the mixing ratio. In summary, there is a lack of data on the concentration of water vapor, and simultaneous studies of water vapor profiles and turbulence have not been performed at all before.

Carbon dioxide may serve as a reference for water vapor measurements in order to deduce mixing ratios from number densities. Carbon dioxide is an appropriate reference as its concentration is constant from the ground up to 90–95 km, as has been shown by in situ and satellite measurements.[30,31] Above 95 km the CO_2 mixing ratio decreases rapidly due to photochemical destruction and demixing by molecular diffusion. Apart from being a reference for the H_2O measurements, the carbon dioxide profile is of interest on its own account.

From the decrease of the CO_2 mixing ratio with altitude, one can deduce the turbopause altitude and the mean turbulent mixing coefficient, also called the 'eddy diffusion coefficient', K. Model studies have shown that the mixing ratio of an inert tracer around the turbopause reflects the long-term effect of turbulent mixing and vertical convection. This may be very different from the actual turbulence field encountered during the rocket flight.[32] In addition to K, one can also deduce actual turbulent parameters, such as turbulent energy dissipation rates, from the small-scale fluctuations of the carbon dioxide mixing ratios. The underlying idea of this analysis and the results from recent neutral gas density measurements are described in the literature.[33] The combined measurement of CO_2 and H_2O allows an interesting comparison between the long-term mixing effect and the actual strength of turbulence measured during flight.

Carbon dioxide is the major coolant of the upper atmosphere due to excitations of the CO_2 molecules by collisions and subsequent emission of IR radiation. The efficiency of this cooling strongly depends on the excitation/deexcitation mechanism of the CO_2 molecules, thus on the coupling between the radiation field and the gas. As the lifetime of the excited states is of comparable magnitude to the mean time between collisions the CO_2 molecules may not be in local thermodynamic equilibrium (LTE). It turns out that collisions with atomic oxygen are most important for the radiative

coupling between CO_2 and the background atmosphere. Non-LTE of CO_2 has been investigated with models and measurements for several years,[34] but there are still large uncertainties regarding the quantitative effect of non-LTE on the CO_2 cooling rate. The problem is the measurement of the deactivation coefficient k_O of atomic oxygen on carbon dioxide at realistic temperatures, and the uncertainty of the actual atomic oxygen number density O. Diode laser spectroscopy can be a tool to measure non-LTE effects directly by studying two appropriate CO_2 absorption lines simultaneously. From this, one can determine the vibrational temperature of the excited level T_{vib} which deviates significantly from the kinetic temperature T_{kin} if CO_2 is not in LTE with the environment. In Figure 11 we show model calculations demonstrating the difference between the vibrational and the kinetic temperature under non-LTE conditions. The kinetic temperature profile used in these calculations exhibits wavy fluctuations, as expected during a flight performed in the winter season. As explained in more detail elsewhere,[35] the non-LTE effect is enhanced if the temperature profile is disturbed, for example by gravity waves. The intercomparison of T_{vib} and T_{kin} allows for a determination of the product $k_O \cdot O$, which is exactly the quantity needed to estimate the radiative cooling efficiency of CO_2.

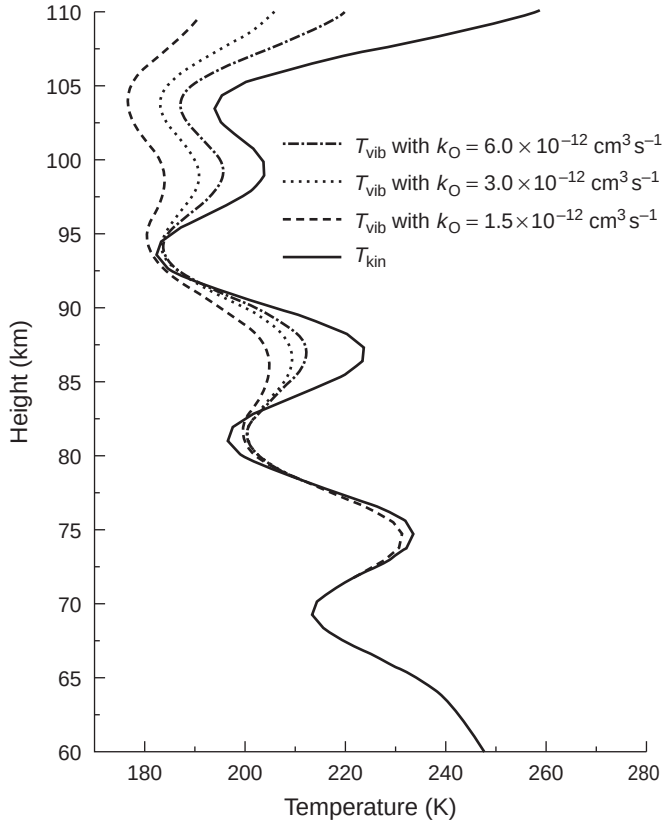

Figure 11 Model calculations of vibrational temperatures T_{vib} (broken lines) under non-LTE conditions in comparison with the kinetic temperature profile T_{kin} (solid line). Three different deactivation coefficients k_O have been used to calculate the T_{vib} profiles. (This figure is taken from ref. 35.)

3.3.2 Middle-atmosphere Spectrometric Experiment on Rockets for Analysis of Trace-gas Influences: A Rocket-based Tunable Diode Laser Spectrometer

In order to perform measurements described in the previous section, a dedicated diode laser system was designed, built, and tested. Two IR diode lasers are mounted on a cold head which is cooled by liquid nitrogen. The operating temperature of each laser is around 80–120 K and is stabilized to within a few millikelvin. For reasons explained in more detail elsewhere,[27] laser emission lines at 1539.06 cm^{-1} (i.e. in the 6.5 μm band) for water vapor, and at 2347.57 cm^{-1} (i.e. in the 4.3 μm band) for carbon dioxide were chosen.

Part of the laser beam is directed through a reference cell containing both carbon dioxide and water vapor. The absorption signals from this reference cell are used for laser stabilization and monitoring. The main part of the laser beam is directed into a multipath absorption cell (White cell) which is open to the atmospheric environment. The optical cell is mounted in front of the rocket payload, so that the atmosphere is sampled without the risk of contamination due to gases leaking from the rocket body. Inside the White cell the laser beam is reflected 104 times which results in a total pathlength of 32 m.

Both lasers are operated sequentially, i.e. one laser is turned off while the other is emitting. The switch over from one laser to the other occurs every 7.37 ms. While one laser is emitting, its frequency is tuned by varying the injection current in order to scan the absorption line. A small modulation with high frequency (500 kHz) is superimposed on the laser current which leads to a corresponding modulation of the laser frequency. The absorption signal is detected at the second harmonic of the modulation frequency by a lock-in amplifier in order to improve the SNR. The laser spectrometer was designed to be able to detect gases with a relative absorbance as low as 10^{-4}–10^{-5} for an integration period of approximately 1 s. The height resolution at the expected payload speed of 1000 m s^{-1} varies from a few meters for CO_2 (no integration required) to about 1 km at the upper altitude limit (integration period 1 s).

An instrument to be flown on a sounding rocket has to be specially designed, qualified and tested in order to survive the large mechanical stress in the acceleration phase of the flight. It must be ruggedly constructed and minimized in terms of weight and dimensions. Table 6 gives more details on the mechanical specifications of MASERATI. The MASERATI instrument consists of

Table 6 Mechanical specifications of MASERATI

Number of lasers	2
Total mass	112 kg
Mass of electronics section	37 kg
Mass of White cell	7 kg
Mass of laser section	68 kg
Total length	1497 mm
Length of White cell	350 mm
Length of laser section	621 mm
Length of electronics section	526 mm
Distance between White cell mirrors	305 mm
Diameter of flange with White cell	438 mm
Max. content of liquid nitrogen in the dewar	~2 L
Max. operation time	6 h

three main mechanical subsystems: the White cell, the laser and detector section, and the electronics section (Figure 12). A photo of the MASERATI section is shown in Figure 13.

The basic details of the three sections are as follows: The White cell (No. 8 on Figure 12) ensures a long optical path to maximize absorption with minimum physical space. The total optical path in the White cell is 32 m. The White cell is built in an open configuration, that is, with direct contact to the ambient air (no tubing, valves etc.). It is placed at a considerable distance (336 mm) from the base deck in order to minimize shock-front effects. Aerodynamical calculations were performed to demonstrate that under flight conditions the major part of the shock front is located behind, i.e. downstream of, the White cell. One therefore expects very little influence from shock front effects on the measurements and practically no contamination from outgasing of the payload.

The laser section (No. 9) contains the two tunable diode lasers, the optics for laser monitoring and stabilization (power and frequency), and the detectors for the reference and the White cell signals. It also contains the liquid nitrogen dewar for cooling of the lasers and the detectors. The dewar has a maximum capacity of 2 L which corresponds to a maximum operation time of approximately 6 h. Figure 14 shows a photo of the rear end of the laser section.

The electronics section (No. 10 on Figure 12) contains various power supplies, digital and analog electronics for laser current and temperature control, lock-in amplifiers, and analog and digital interfaces to the control computers. The $2f$ signals are amplified with an automatic gain adjusted to the actual signal strength. The largest available gain factor is $2^{14} = 16\,384$.

The laser and the electronics sections are sealed to ensure that no seawater can enter the payload when it is floating in the ocean after flight waiting for recovery. The White cell is located underneath the rocket nosecone and is covered by a separate hood (No. 7) to be protected during lift-off and during the first phase of the rocket flight. The hood is vacuum tight, and on the ground the gas composition and pressure can be controlled from outside which allows to test and monitor the laser performance during tests and countdown. The hood is mechanically fixed to the nosecone; both are ejected at approximately 50 km on the upleg part of the rocket flight.

In the MASERATI electronic section the raw data received from the detectors are amplified and preprocessed. Only the important information is stored and used for later data evaluation. The absorption spectra and various other data that monitor the performance of the instrument (e.g. the laser temperatures) are digitized in the MASERATI electronic section and are then transferred to the so-called housekeeping section of the payload (No. 3) where the data are transmitted to the ground by means of standard telemetry techniques. The most important parameters of the electronics are listed in Table 7.

The MASERATI instrument is mounted on top of a sounding rocket called RONALD, a German–Norwegian collaborative project for middle-atmosphere studies. The payload is equipped with a sea recovery unit (No. 4) which basically consists of a parachute and a floatation bag. The parachute ensures deceleration before impact, and the floatation bag prevents the payload from sinking. A beacon and a strobe light are mounted on top of the floatation bag to help locate the payload in the ocean. The recovery unit also contains a despin mechanism which reduces the spin of the payload during flight from 3 Hz to practically zero (the residual spin during the first flight was 0.1 Hz). The total weight of the rocket is 1.5 metric tons, the payload weight is approximately 300 kg. The total length of the rocket is 8.6 m, nearly half of which (4.3 m) consists of the rocket motor (No. 6) and the conical adapter (No. 5; Figure 12). The diameters of the payload and the motor are 44 cm and 56 cm, respectively.

A sounding rocket launch is a strong mechanical burden for a complicated instrument. In the initial phase of the rocket flight (for about 26 s after lift-off), there is a nearly constant acceleration of 6–10 g superimposed by strong vibrations. After motor burn-out, accelerations are very small. The rocket motor transports the payload to an apogee of about 120 km within 170 s. The payload falls into the sea approximately 7.5 min after lift-off.

3.3.3 Laboratory Calibrations, Environmental Tests, and First Results

A special vacuum chamber had to be set-up in the laboratory to test the sensor in the pressure range from 10^{-6} to 10 mbar. The instrument was calibrated with a wide range of CO_2 and H_2O mixing ratios, total pressures, and gas temperatures. Several calibration

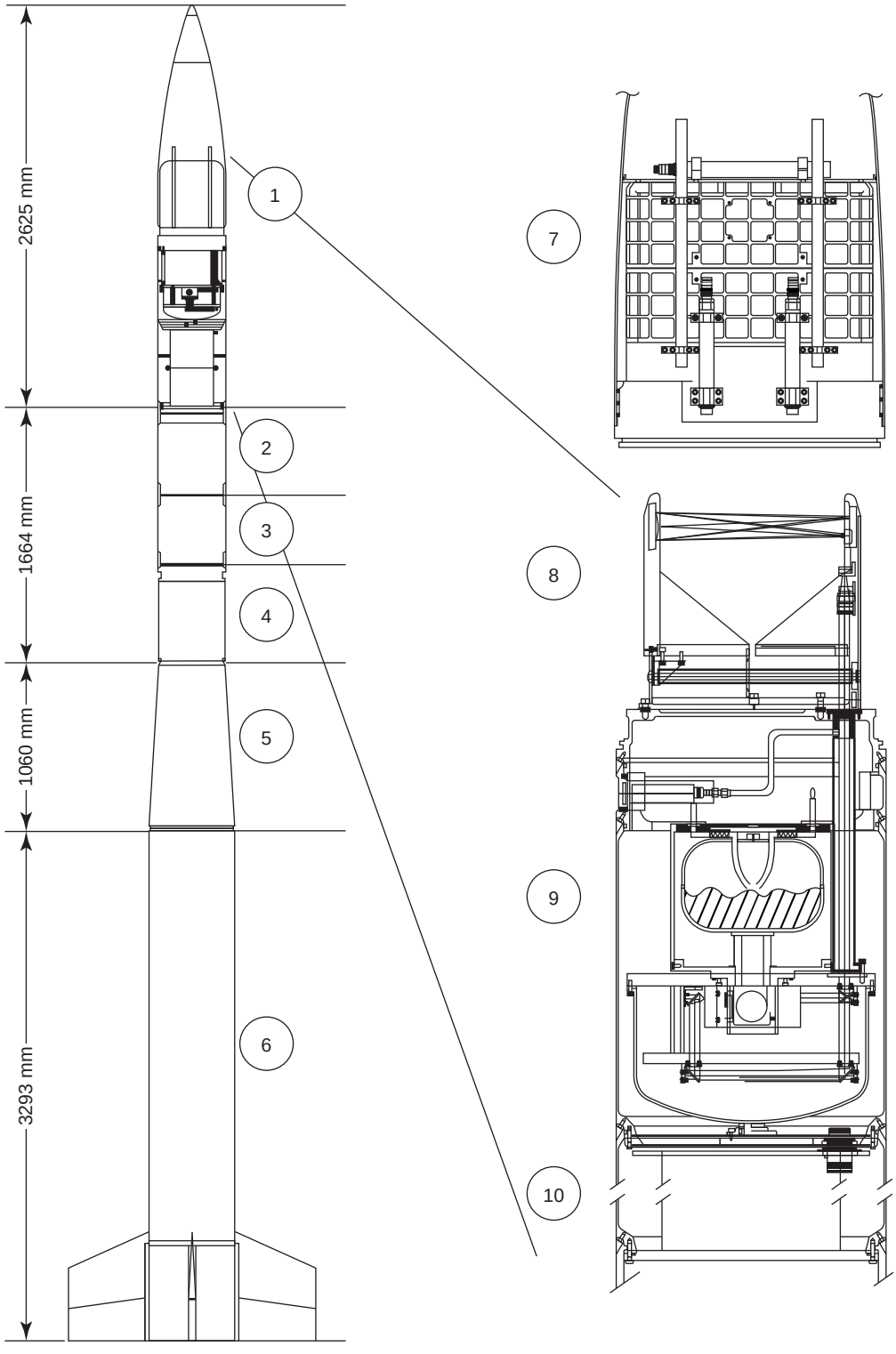

Figure 12 Sketch of the sounding rocket rocket-borne optical neutral gas analyzers with laser diodes (RONALD) with the MASERATI instrument. Apart from MASERATI (1) the other main instrument on board the payload is TROLL (transmitter and receiver of laser light)[36] (2). Furthermore, the payload consists of a telemetry section (3) and a recovery unit (4). A conical adapter (5) is used as an interface between the payload and the rocket motor. (6) More details on the MASERATI instrument are given in the text.

Figure 13 Photograph of the MASERATI section of the rocket payload. The two cylindrical sections contain the lasers and the electronics of the MASERATI instrument. At the top of the figure one of the mirrors of the open White cell can be seen.

series were performed to study and maximize the SNR and to determine the absolute sensitivity of the instrument.

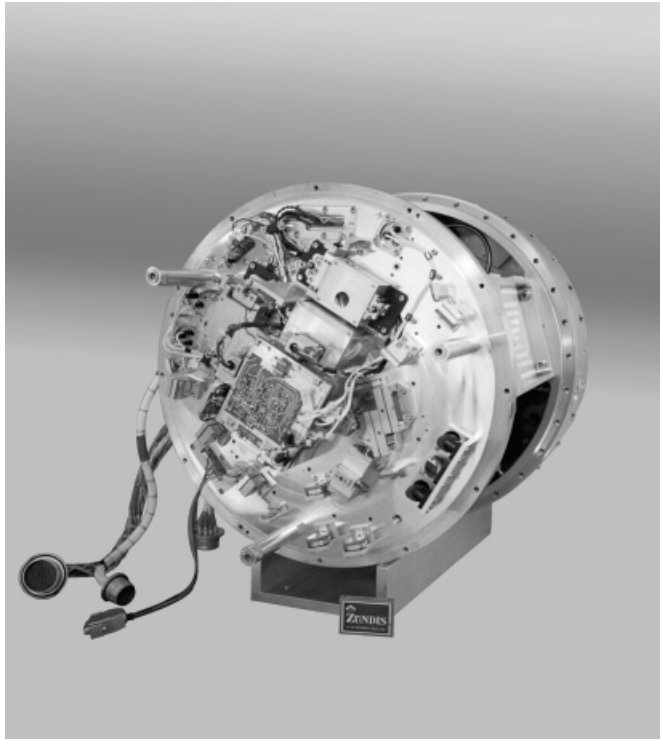

Figure 14 Photograph of the rear end of the laser section of the MASERATI instrument. Two diode lasers are mounted in the mechanical structure in the center of the picture. Various optical components direct the laser beam to the White cell in front of the payload (see Figure 13).

Table 7 Electrical specifications of MASERATI

Supply voltage	28 V
Supply current	2.2 A (3 A max)
Scans per second per laser	68
Time for one laser scan	7.37 ms
Samplings per scan	64
A/D bit resolution	12
Total bit rate	434 kbit s^{-1}

One can only dare fly such a singular and costly experiment on a rocket after extended simulations have been performed, for instance regarding the pressure changes expected during flight. The instrumental parameters must be chosen carefully because they cannot be changed during the rocket flight. The laboratory tests demonstrated the sensitivity of MASERATI to be in the range 10^{-4}–10^{-5} of relative absorbance for an integration time of 1 s.[27] This corresponds to an upper altitude limit of 120 km and 95 km for CO_2 and H_2O, respectively, taking into account the expected concentrations, the performance of the particular lasers, and the selected spectral lines. Concerning the small-scale fluctuations in the CO_2 channel, the noise at high sampling rates was also analyzed at signal strengths which correspond

to mesospheric altitudes. Random noise fluctuations of 0.5% (root mean square) for sampling rates of 14 ms were measured which is sufficient to detect turbulence-induced density fluctuations. Several vibration tests with accelerations expected during the rocket flight were performed for various subsystems as well as for the entire instrument. After minor problems were identified and solved the instrument was found to be well suited for a sounding rocket flight.

MASERATI was launched twice from the Andoya Rocket Range (69 °N) which is located on an island in northern Norway. The first flight took place on October 12, 1997, at 01 : 19 : 00 UT under perfect operational conditions, i.e. no troposphere clouds, very little ground wind, no aurora. The rocket reached an apogee of 120.9 km. Other ground-based and rocket-borne instruments were in operation in order to independently determine atmospheric parameters such as densities and temperatures in the mesosphere.

In Figure 15 a sequence of $2f$ signals from the reference channel and from the White cell channel is shown, which had been measured during an early stage of the first flight when the acceleration of the rocket is still about $6\,g$ and the spin is 3 Hz (the de-spin of the payload occurs later during flight, some 48 s after lift-off). At this time the White cell is still tightly surrounded by the vacuum cover. In both channels (reference and White cell) there are two CO_2 absorption signals which originate from lines at 2347.58 cm^{-1} and 2347.63 cm^{-1}, respectively. The corresponding line strengths S at room temperature are 6.8×10^{-19} and 1.3×10^{-19} cm per molecule, respectively. The ratios of the $2f$ signals in Figure 15 differ from the ratio of S due to two effects: the second CO_2 $2f$ signal in the scan is given preference relative to the first by optimizing the amplitude of the frequency modulation for the second line, as it is weaker compared to the first; in addition, the absorption in the reference cell is close to saturation (the amount of CO_2 in the reference cell is rather large in order to get a strong signal, used for laser stabilization).

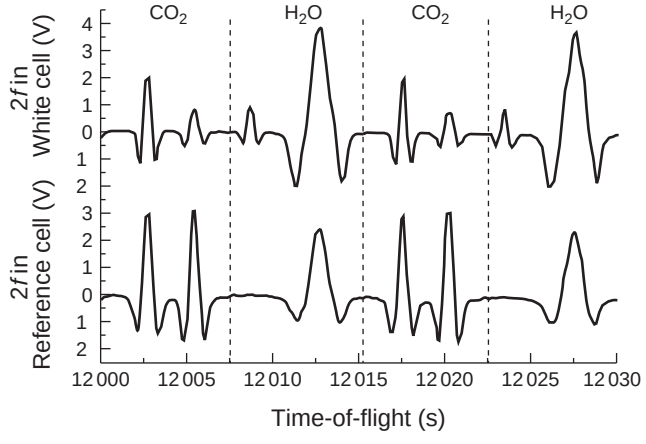

Figure 15 Sequence of $2f$-signals measured during the first RONALD flight launched on October 12, 1997, at 01 : 19 : 00 UT. Note the following amplifier gain factors for the CO_2 and the H_2O signals: 16 and 4 for the White cell signal, and 4 and 1 for the reference cell signal. More details are given in the text.

The two absorption features in the H_2O White cell (Figure 15) stem from lines at 1539.01 cm^{-1} and 1539.06 cm^{-1}. The corresponding line strengths at room temperature are 2.4×10^{-21} and 2.0×10^{-19} cm per molecule. The main absorption line in the White cell is close to saturation, which is caused by a large amount of H_2O under the vacuum cover (the vacuum cover cannot be actively pumped during the last 2 h before launch). The second line cannot be seen in the reference cell because the absorption is too weak.

The signals in Figure 15 demonstrate that the MASERATI instrument is functioning perfectly and that it has successfully survived the rough mechanical load encountered during lift-off. However, it must be noted that some subsystems of MASERATI failed later during flight.

The second MASERATI flight took place on January 31, 1998, at 23 : 43 UT. In Figure 16 a sequence of scans of the CO_2 channel is shown as measured during flight in the lower mesosphere at approximately 66–67 km. Very persistent and stable spectra are observed which confirms the good technical performance of the instrument. More details on this flight are available.[27]

Unfortunately, there was a failure in the rocket hardware which prevented the payloads from being recovered from the sea. Still, the MASERATI instruments have demonstrated the suitability of the TDLAS technique for in situ measurements in the mesosphere and lower thermosphere on board sounding rockets.

3.4 Exhaust Gas Analysis

3.4.1 Introduction

In the field of exhaust gas analysis, tunable diode laser spectroscopy has been applied both to the analysis of stack gas of power plants[37] and to automotive exhaust within industrial test beds. In the latter case, two different directions were followed. On one hand, a couple of exhaust gas components were measured simultaneously for both Otto engines[38] and Diesel engines[39] with response times of the order of 1 s. A special version, equipped with a long-path cell, for detection limits down to ppb levels was also reported.[40] On the other hand, fast response gas analysis with response time distinctly lower than 1 ms, but with lower sensitivities, was demonstrated in the early 1980s.[41,42] Later developments[43,44] led to more sophisticated systems with a response time of the order of 5 ms.[44]

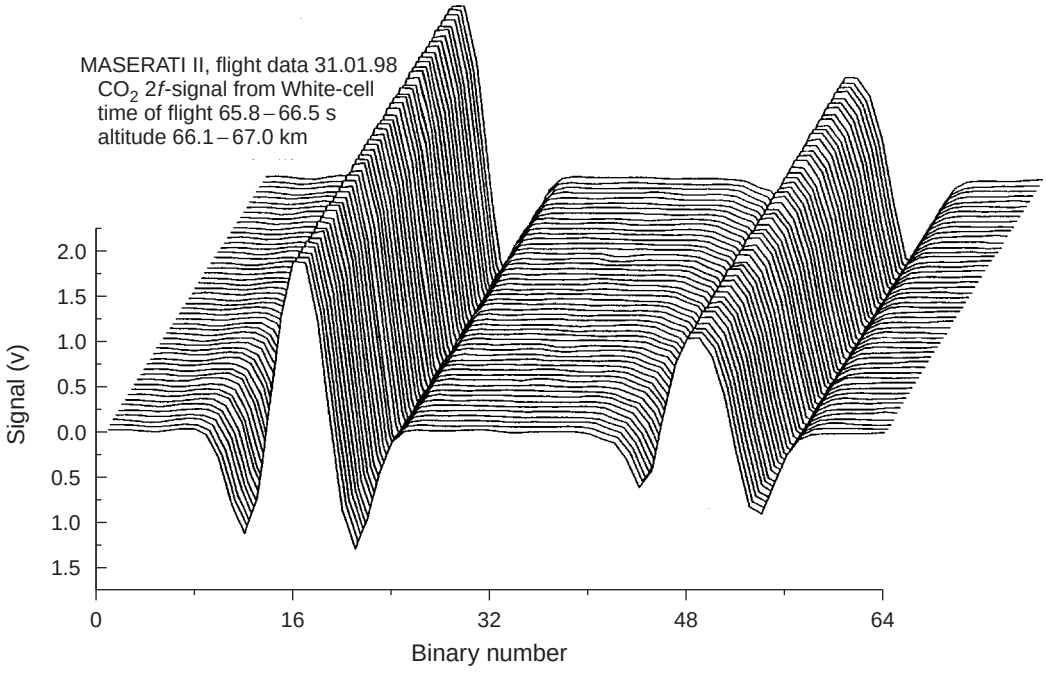

Figure 16 A sequence of 51 successive 2*f* signals in the CO_2 channel measured in the atmosphere during the second MASERATI flight on 31 January 1998, 23 : 43 UT. The scans were recorded 65.8–66.5 s after lift-off, which corresponds to an altitude range of 66.1–67.0 km. Very persistent and stable spectra are measured (maximum deviation from the mean: less than 1%) which demonstrates the good technical performance of the instrument.

Modern cars are characterized by remarkably low exhaust gas emission levels, essentially realized by improved engine design and by sophisticated exhaust gas after-treatment. As a result, the emission contributions of state-of-the-art vehicles mainly occur at cold-start and during nonstationary operational phases. Better control of these two conditions represents the major potential for further emission reductions.

3.4.2 Dynamic Car Exhaust Gas Analysis: The Diode Laser System

Figure 17 is the block diagram of the optics and the electronics of a dynamic car exhaust gas analyser system.[44] Two IR diode lasers are operated at a precisely controlled individual temperature ($\Delta T = 3\,\text{mK}$) and current ($\Delta I = 10\,\mu\text{A}$). Both laser beams are combined at first to form one beam and then split in order to generate one reference and two measurement channels with respective detector elements and further signal processing units. All set-points and control-circuits are microcomputer-controlled and can be operated from a PC located up to 20 m away from the instrument.

The diode lasers are operated by applying current pulses of typical 60 µs duration in time multiplex with a repetition rate of 5 kHz. The concentration is evaluated after each pulse.

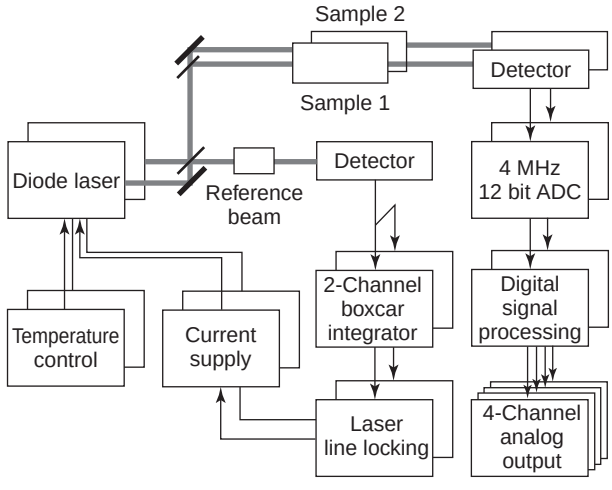

Figure 17 Block diagram of the dynamic exhaust gas analyzer.

The principle of fast gas sampling is given in Figure 18. Two sampling channels are incorporated in parallel into the system in order to extract sample gas at two locations out of the car exhaust gas line. Due to the special geometry at the cell and the low-pressure conditions, no particle filter is required, thus preventing unwanted sample gas delays. The exhaust gas, taken from a selected position of the car exhaust gas system, is fed into the 2 mm inner diameter tube of a coaxial heat exchanger approximately

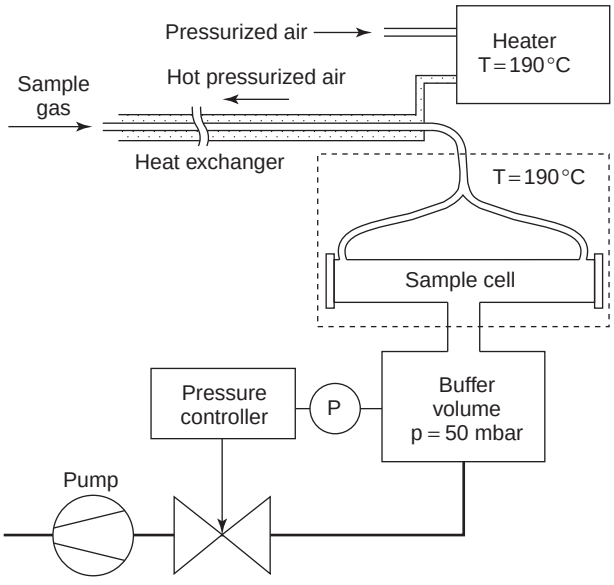

Figure 18 Diagram of the fast gas-sampling system.

100 cm in length, with a counterstream of pressurized air at 190 °C in the outer jacket. The sample gas is then fed into the temperature-controlled sample cell, which has a length of 30 cm and a diameter of 0.6 cm. The pressure in the cell is kept at 50 mbar by a large buffer volume, which is common to both sampling channels. Pressure control is achieved by means of a fast butterfly valve between the buffer vessel and the pump, the latter having a relatively large pumping capacity of 60 $m^3 h^{-1}$. The fast sample gas velocity within the heat exchanger results in a delay along the sampling system of about 20 ms. It has been found that the turbulent flow of the sample gas does not intermix the gas significantly; therefore, variations in composition are transferred to the sample cell without noticeable modification.

The system response time is essentially determined by the gas exchange time in the sample cell. The windows of the sample cells are made of sapphire, which can be easily interchanged or cleaned if necessary. Under practical operating conditions, cleaning is required once a day. Even Diesel engine exhaust gas with a high particle content can be analyzed for more than half an hour before cleaning of the windows is required.

Figure 19 gives a view of the analyzer system taken during a measurement campaign. In the rear, a large Diesel engine can be seen. The laser spectrometer optics are located on the left side of the black upper portion, with a liquid nitrogen cooler dewar protruding on the top. Below on the left are laser operating and data processing electronics. To its right, the sample gas conditioning system with the pressure control and pump (at the bottom) can be seen. The two sampling lines with integrated heat exchangers are on top of the right

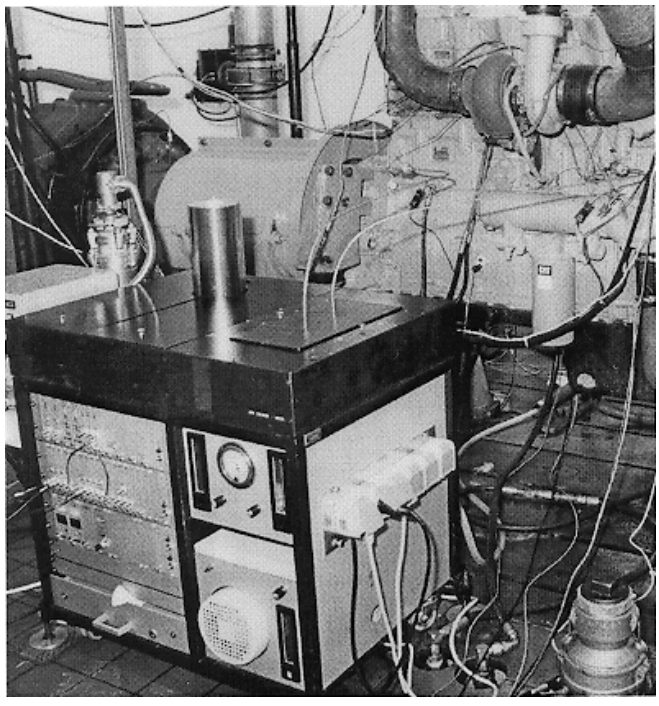

Figure 19 The dynamic exhaust gas analyzer system in operation with a large Diesel engine. (Courtesy of Caterpillar Inc., Peoria.)

side; they lead directly to the sample cells, which are not visible. The sampling lines can be rotated to come as close as possible to the sampling location. In order to extract the exhaust gas sample, welded pipe sockets at the respective positions of the exhaust gas line of the engine are required. Flexible connections between the engine and analyzer are obtained with short 2 mm stainless steel tubes and standard Swagelok components.

The major specifications of the analyzer are listed in Table 8. The instrument has been conceived for two exhaust gas components which are measured simultaneously at two sampling locations, yielding a total of

Table 8 Basic specifications of the dynamic exhaust gas analyzer system

Gas components		2
Sampling locations		2
Sampling line length		~120 cm
Gas delay time		~20 ms
Time resolution		~5 ms
Detection limit	CO	10 ppm
	NO	100 ppm
	CO_2	500 pmm
	CH_4	5 ppm
Span	CO	2000 ppm
	NO	10 000 ppm
	CO_2	20%
	CH_4	2000 ppm

four real-time values of concentrations. Four molecular species (Table 8) have been determined so far. The detection limit values are obtained with selected strong (with the exception of CO_2) absorption lines of the respective molecular species which do not overlap with those of other exhaust gas components. This is a prerequisite for gas analysis free of cross-interference.

System calibration and zeroing is performed by flooding the total car exhaust gas pipe by a sufficiently high flow of test gas and zero gas, respectively. This way not only the spectroscopic part of the instrument, but also the complete sampling system is included in the calibration procedure. In our experience, one calibration cycle per day is sufficient.

3.4.3 Results of Dynamic Car Exhaust Gas Analysis

The results discussed below[44] were obtained during a number of measurement campaigns by European and US American automotive companies. Although most of the results were classified as proprietary to the respective companies, a few findings can be given in order to illustrate the potential of tunable diode lasers in fast exhaust gas analysis.

The first two examples were obtained with a single-cylinder research engine. In Figure 20 the NO and CO emissions are plotted as measured simultaneously behind two different outlet valves at an engine revolution rate of 2000 rpm. The SNR of the analyzer is better than 50 : 1 under these conditions, which means that the observed variations in Figure 20 represent real concentration profiles. The cylinder pressure is given for reference.

As can be seen, the time constant is sufficient to resolve each cycle; even faster variations can be distinguished. Note that the emissions of NO at the two outlet valves are largely synchronous, whereas the CO values exhibit distinct differences. This behavior indicates inhomogeneous combustion due to improper air–fuel mixing. Such facts are not observable using conventional exhaust gas techniques. This kind of result is extremely helpful when testing detailed combustion models and to optimize the combustion in experimental engines.

Another result obtained with this single-cylinder engine is given in Figure 21. It represents a comparison between the NO emission measured directly with 0.8 s averaging time as well as with a commercial chemoluminescence (CLD) system with approximately 0.8 s response time. For the plot, the time delay between both measurements was corrected. The ignition was turned off for approximately 1 s in order to intentionally generate a strong transient. The variations observed with the ignition turned on are again essentially not due to noise but represent real fluctuations of NO level. The noise level of some 150 ppm can be deduced from variations during the ignition-off period.

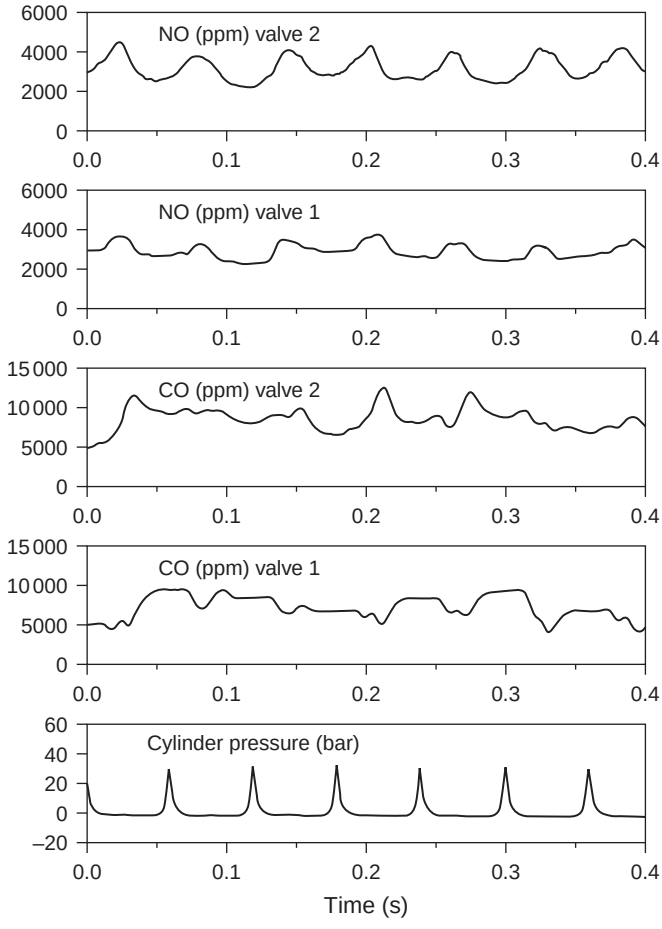

Figure 20 NO and CO behind two different outlet valves of a one-cylinder engine. (Courtesy of Renault, Direction de la Recherche, Rueil Malmaison.)

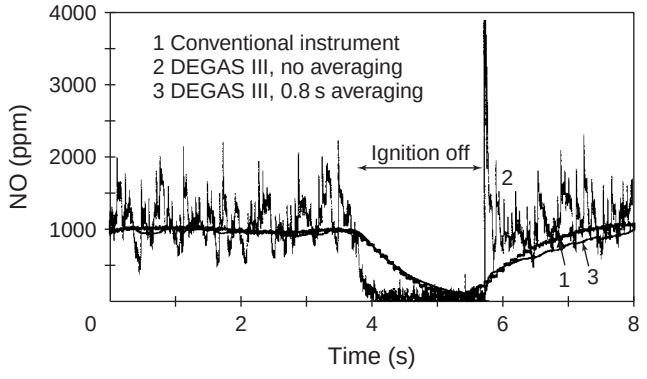

Figure 21 Comparison of NO concentrations as measured by the system of Figure 19 (DEGAS III) and by a CLD instrument with approximately 0.8 s rise time. (Reproduced by courtesy of Renault, Direction de la Recherche, Rueil Malmaison.)

The overall agreement between the two curves taken with the 0.8 s time constant is excellent, although the CLD under these conditions yields some more

information, including a step pattern connected to the single combustion cycles. Obviously, the dynamic analyzer with full time resolution provides considerable additional information as compared to the CLD. In addition to the fluctuations mentioned there a large peak upon turn-on, which may point to an unusual high temperature during the first combustion cycle. The corresponding results for CO, not shown here, exhibit a similar fine structure with a somewhat poorer agreement for the 0.8 s average with respect to a conventional measurement with nondispersive IR techniques.

A further investigation concentrated on the influence of fuel composition on the emissions as a function of crank angle. Figure 22 is an example for CO and CH_4 emissions obtained for a four-valve single-cylinder engine under stationary operation. The temporal profiles of these two exhaust components were required in order to test different thermodynamic models of the combustion process. The cylinder pressure was again recorded for reference. The data were synchronously averaged over a number of strokes in order to produce a sufficiently high SNR for the relatively low CH_4 emission values. By this procedure, repetitive features in every stroke can still be distinguished with full system time resolution. To obtain the graph in Figure 22, the time delay of the gas measurement was again compensated.

Figure 23 is an example for a Diesel engine measurement. Plotted are the NO and CO concentrations emitted from one cylinder of a six-cylinder 14.6 L engine in stationary operation at about 1400 rpm. The upper two traces

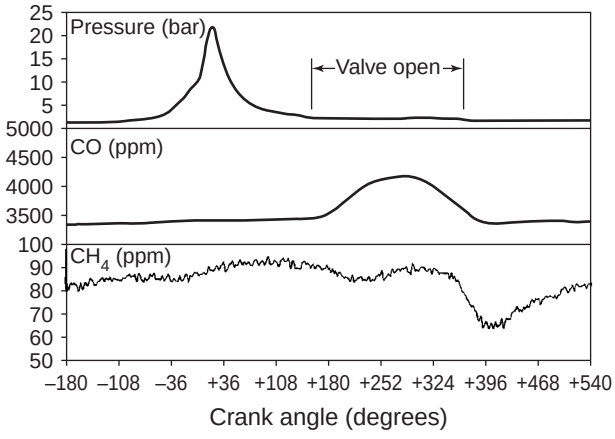

Figure 22 CO and CH_4 emissions of a single-cylinder engine as a function of crank angle. (Reproduced by courtesy of Shell Research, Thornton, UK.)

show the valve control signals and the cylinder pressure, respectively. Whereas the NO concentration proceeds quite regularly from cycle to cycle, the CO emission exhibits distinct irregularities. During zero CO emission periods, the residual noise during this measurement of about 50 ppm may even lead to a negative reading, as seen at 0.45 s.

The result shown in Figure 23 for an engine running at a relatively low revolution rate illustrates that high time resolution yields valuable information even in the case of stationary engine operation. In this particular case, the resolution is high enough to observe non-reproducible

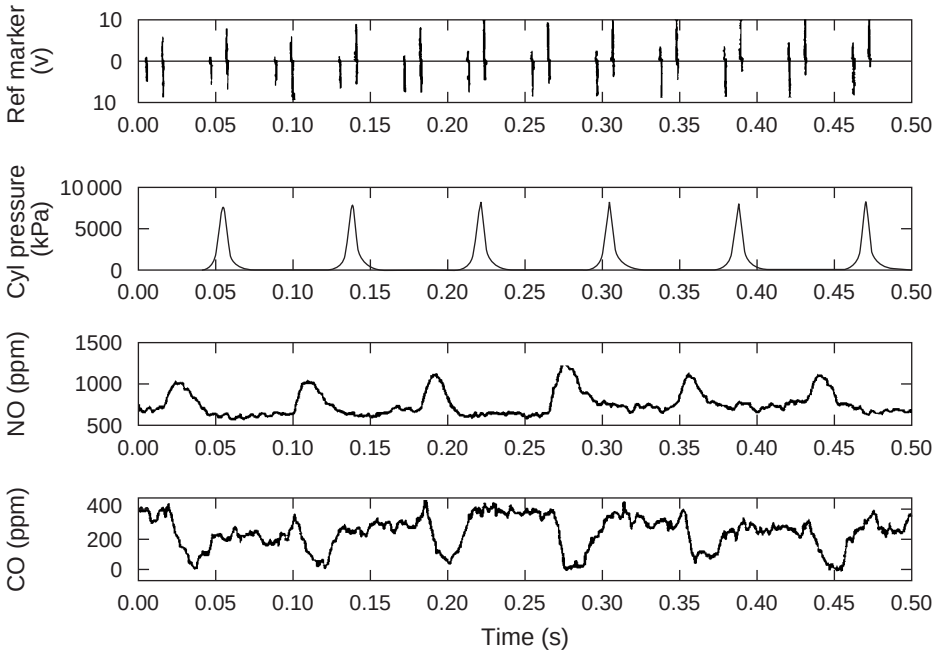

Figure 23 NO and CO emissions of a large Diesel engine. (Reproduced by courtesy of Caterpillar Inc.)

structures within single combustion cycles. During this investigation, the effect of modifications to the engine management was often observed on-line in the emission concentrations.

The above examples show that dynamic car exhaust gas analysis by tunable diode lasers is a valuable tool in engine research as well as in engine development and fine tuning. A net advantage at present seems to be in the investigation of instationary engine operation and the corresponding emissions; these may exhibit millisecond emission peaks that can considerably contribute to the total emission, such as over a driving cycle. A detailed interpretation of the results requires additional information on the engine operation parameters and the gas flow dynamics in the cylinder and the exhaust system. The benefit is not only a better knowledge of the behavior of individual engines, with the potential of efficient optimization of the engine operating parameters, but also a better understanding of the combustion thermodynamics within the cylinder. The first can be immediately transformed to higher fuel economy and lower emission of existing engines, the latter will have an appreciable impact on future engine development.

4 RELEVANT SYSTEM PARAMETERS

4.1 Gas Handling

Sampling systems usually consist of the intake and a tube connecting it to the sampling cell. These parts require special attention if the target molecules adsorb to the walls, or if they react with or on the walls. This can only be accounted for by proper choice of the wall material and by sufficient air flux through the intake system.

The same holds for the sample cell itself. Choice of materials is limited here, as the optical windows and optical components needed inside long-path cells must be of good optical quality. However, very rarely is a satisfying compromise not found. These material problems are not specific to laser systems, but are found for nearly all analysis systems except open optical path systems. Most other analysis techniques face stronger sampling restrictions.

Long optical pathlengths are necessary in order to obtain high sensitivity. For low concentrations and consequently low absorptions, the measurement signal (which is the reduction of the transmission) is proportional to the pathlength. For a length of more than some meters, long-pathlength cells with folded optical beams are inevitable. Such cells scale in length proportional to the optical path. The partial beams are stacked side by side, so that the cross-section of the cell is filled proportionally to the number of partial beams. As the number of partial beams again is proportional to the whole length, the cell cross-section area scales with the number of beams, and hence the cell volume scales with the square of the optical beam length and thus with the square of system sensitivity. This discussion neglects the fact that the beams expand within the cells, which asks for increased cell diameters proportional to the cell length. This results in an overall increase of cell volume with the third power of the pathlength (and third power of sensitivity, too). In practical cases, very small cells are often limited in volume by other technical demands; cell volumes of commercial cells range from a few tens of milliliters for a 3 m cell to some liters for a 100 m optical path.

Trace gas analysis typically requires an optical pathlength of 100 m, which results in gas exchange times of the order of many seconds, depending on the pump efficiency and the cell design. The gas throughput is about two orders of magnitude smaller, due to the reduced pressure in the sampling system, so that the total gas throughput is quite limited and usually of no concern.

For calibration purposes, one usually flushes the whole intake system, so that all wall effects enter into the calibration and need not be corrected later on. Systems with very high sensitivity are usually operated by probing consecutively a calibration gas, then zero gas, and finally the atmospheric sample, as discussed in section 2. Thus system drifts can be accounted for. The total time for taking a data point this way is of the order of 1 min.

The temporal resolution of the reduced sensitivity emission analysis is also limited by the gas exchange, which was reduced down to the order of milliseconds by proper sample cell design. Open-path measurements can be much faster, the sampling time being limited by the acceptable noise which in general is the product of the system noise level and the signal averaging bandwidth. This may be of importance for model studies of atmospheric reactions at higher reaction rates; the system time constants are quite easily brought down to the order of microseconds and below.

4.2 Diode Lasers

The characteristics of these components are dominant on system design and performance, an exception being optical interference features that are common to all tunable laser spectroscopy systems.

Diode lasers are usually made from semiconductor wafers a few hundred micrometers thick, by cleaving or breaking them into small cubes, the laser chips. Two faces of these cubes serve as mirrors for the laser resonator; sometimes they are given optical coatings to optimize the resonator properties. Two laser chip faces perpendicular to the resonator mirror facets are contacted and (usually) soldered to electrical connections. One of

these also acts as a heat sink. This is necessary, since the delicate semiconductor–metal contacts have to sustain current densities up to 10 kA cm^{-2}; most of this current is converted to Joulian heat, which must be led away to prevent the laser from overheating.

The active layer producing the radiation usually has a thickness of the order of a fraction of the radiation wavelength and is located parallel to one of the contact faces, and a few micrometers below it. This face usually is located close to the heat sink, as the active layer produces most of the energy losses and must be well cooled. Typical dimensions of mid-IR lasers are some 300 μm × 300 μm in length and width (of the contact plane) and 300 μm in thickness (Figure 24).

Most lasers have active layers that are limited to a narrow stripe of some micrometers to some tens of micrometers so as to confine the laser radiation to only one propagation direction.

Figure 25 is of a mid-IR laser diode, showing the diode in its mount. The chip can be seen in the reflection of the crystal facet behind the laser. The crystal is PbSe, and the laser chips are made of wafers cut from such crystals parallel to the crystal facet. The PbSe material is a member of the IV–VI semiconductor family. At present, IV–VI diode lasers are available commercially for the 3–15 μm wavelength range, and are used nearly exclusively for atmospheric gas analysis.

As a voltage is applied to the laser, a current can be observed that is typical of the current–voltage characteristics of a rectifying diode. This is because the active area giving optical gain has an internal p–n junction. To generate laser radiation, the diode is driven in forward direction so that the electrons of the n-type material and holes, the charge carriers in the p-type material, recombine near the p–n junction. One of the recombination channels effects spontaneous emission of photons that

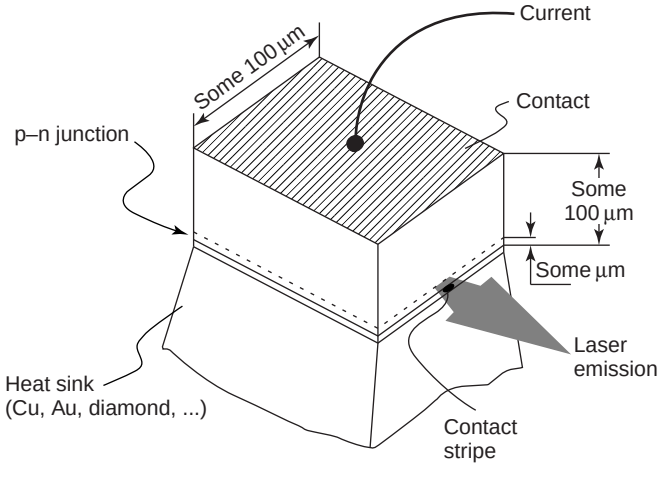

Figure 24 Schematic of a diode laser.

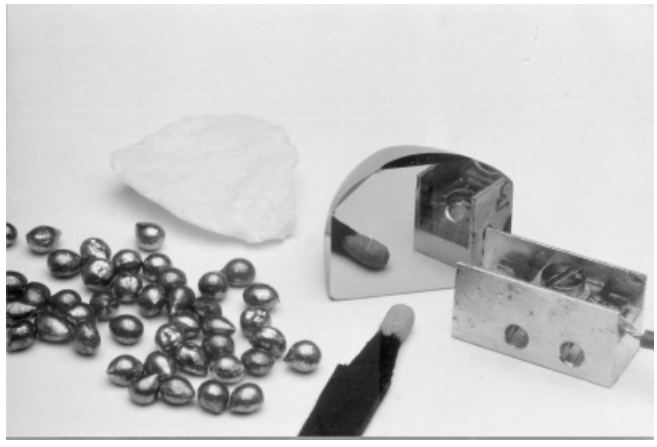

Figure 25 A mid-IR diode laser. The tiny laser chip is mounted in a metal housing, which is positioned here in front of a semiconductor crystal that the laser wafer is made with, and the elements lead and sulfur of such crystals.

carry the difference in energy between the conduction band and the valence band, E_g. These photons can be observed by a detector that detects radiation coming from the facet and near the p–n junction. The intensity of this radiation increases with current, as expected from the increasing number of carriers that recombine. A faster increase of the emission with current results, due to stimulated emission, and finally an even stronger and nearly linear increase occurs which is combined with a high coherence of the radiation, the laser emission. The output power P of a diode laser is usually described as an approximately linear function of current, $P \propto (I - I_{th})$, with I_{th} being the threshold current. This power usually is comparatively low, and in usable lasers can be up to some 100 μW. This is moderate compared to other wavelength ranges and other laser types. However, the spectral power is orders of magnitude larger than what is available from conventional blackbody sources. This is one major advantage of diode laser systems as compared to conventional IR spectroscopic systems.

The second advantage stems from the fact that the refractive index of the laser material depends on temperature. As the geometrical dimensions vary negligibly with temperature, the change of refractive index causes the optical length of the laser resonator to vary. This length determines the emission wavelength, because a multiple of half the emission wavelength must fit into the resonator length. As mentioned before, most of the laser current is lost and transformed to heat generated in the active section. This heat changes the temperature which in turn results in a variation of the emission wavelength. Hence the lasers are tuned by the current, which gives a very fast and efficient tuning mechanism that is optimally suited for electronic control.

For technical reasons, diode lasers mostly are made of double hetero structures, with a thin active semiconductor layer embedded between different (passive) semiconductors. This active layer is where the carriers are made to recombine and add their energy to the laser radiation. The active layer is also designed to guide the radiation. Thus the output area has a thickness of the order of the thickness of this layer. In practice, the output area has about the height of a wavelength and a width of one to several wavelengths. The radiation coming from such an aperture is governed by diffraction effects such as radiation originating from a narrow slit. The corresponding diffraction lobe of the lasers is elliptic, usually with a strong difference between the two diameters of the ellipse. It is hard to cope with such strongly divergent beams; the optical situation is analogous to that of a microscope objective. Usually a high aperture is sought for the first optical system, the laser collimator. Most optics use off-axis paraboloids or ellipsoids for the first nonplanar mirrors. The following beam diameter is usually chosen to be of the order of 2 cm, or larger for laboratory systems with reduced stability and mobility demands. This diameter is one influence on the total system dimensions. As diffraction-limited beam sizes scale proportionally to the wavelength, one must always expect long-wavelength IR optics to have larger dimensions than optical systems for visible light.

On its path from the laser through the system optics and sample cell to the detector, a number of optical elements have to be traversed (windows and lenses) or touched (mirrors). As a number of different wavelengths are used for different target components, all lens systems have to be corrected to a high chromatic tolerance. This is difficult to accomplish for the IR range. Moreover, all surfaces give rise to reflections and scattering. Such spurious radiation gets onto the detector and leads to interference, as discussed in section 1. Typically, interferences fluctuate with time and give rise to optical system noise on the detector signal, which limits system sensitivity.

The lasers are also sources of noise, as the laser emission has some residual fluctuation. When close to mode hops (described below), this noise may be large and dominant. Usually, however, a free-running laser that is well protected from beam reflections has a noise floor that does not limit the system sensitivity. Reflections into the laser are another type of interference noise, which must be minimized in order to retain system sensitivity. For this reason, all optical elements are usually made from mirrors rather than lenses.

The optical gain within the laser has a limited bandwidth of some 10 cm^{-1}. As stated before, a multiple of half the laser radiation wavelength must fit into the laser resonator. This can be fulfilled with a number of different wavelengths. The emission can hence consist of many different electric field configurations called modes, with different wavelengths. The mode spacing usually is given in wavenumbers, $\Delta\tilde{\nu} = 1/(2nL)$, where L is the length of the resonator (some 100 μm) and n is the refractive index. Typically, such modes have separations of the order of inverse centimeters, and many modes may be emitted.

As the laser material changes its temperature, the gain band also changes its wavelength location. It does so at roughly fourfold the tuning rate of the modes. Hence, single modes have varying gain, as their wavelength position changes relative to the gain band. The usable tuning length of a single mode is from one to a few inverse centimeters. In general, therefore, laser diodes are selected for wavenumber bands that correspond to the specific substances to be analysed. Whenever a mode dies out as a result of tuning, due to insufficient gain, and a new mode is observed, one talks of mode hops. Close to such mode hops, the laser power fluctuations increase and may dominate the system noise. Figure 26 shows mode spectra of a mid-IR diode laser. The spectra are taken at the same laser operating temperature, but in two different operating current ranges as evidenced by the different wavenumber ranges emitted. These spectra have remarkably few modes and relatively high powers, which shows that they were well selected.

Single-mode lasers can be made by special external optics or by special laser structures. The latter are usually periodic structures along the laser resonator, and corresponding lasers are called distributed feedback (DFB) or distributed Bragg reflector (DBR) lasers, dependant on the specific design of corrugation. Such lasers have been demonstrated for the wavelengths of interest by laser developers, but have reached production stage. This is mainly because diode laser gas analysis is a limited market that cannot support the development of a specific laser technology.

In summary, the diode laser technology and characteristics strongly influence the system design. The small laser emission bandwidths that are possible allow for good selectivity. Their fast tuning response is the basis of high system stability, as the absorption lines can be scanned rapidly, absorption features determined for each scan, and then averaged, a procedure that eliminates drifts. Multimode emission is often found and is detrimental to system performance. Specific combinations of laser current and temperature are usually sought, that reduce these detrimental mode effects such that they do not govern system performance.

Diode laser technology is more advanced at wavelengths of interest for telecommunications. These are in the near-IR and are made from the III–V semiconductors, such as GaAs. Both DFB and DBR lasers are available with 1.54 μm emission for use in gas analysis systems. Such equipment can be bought for emission control of NH_3

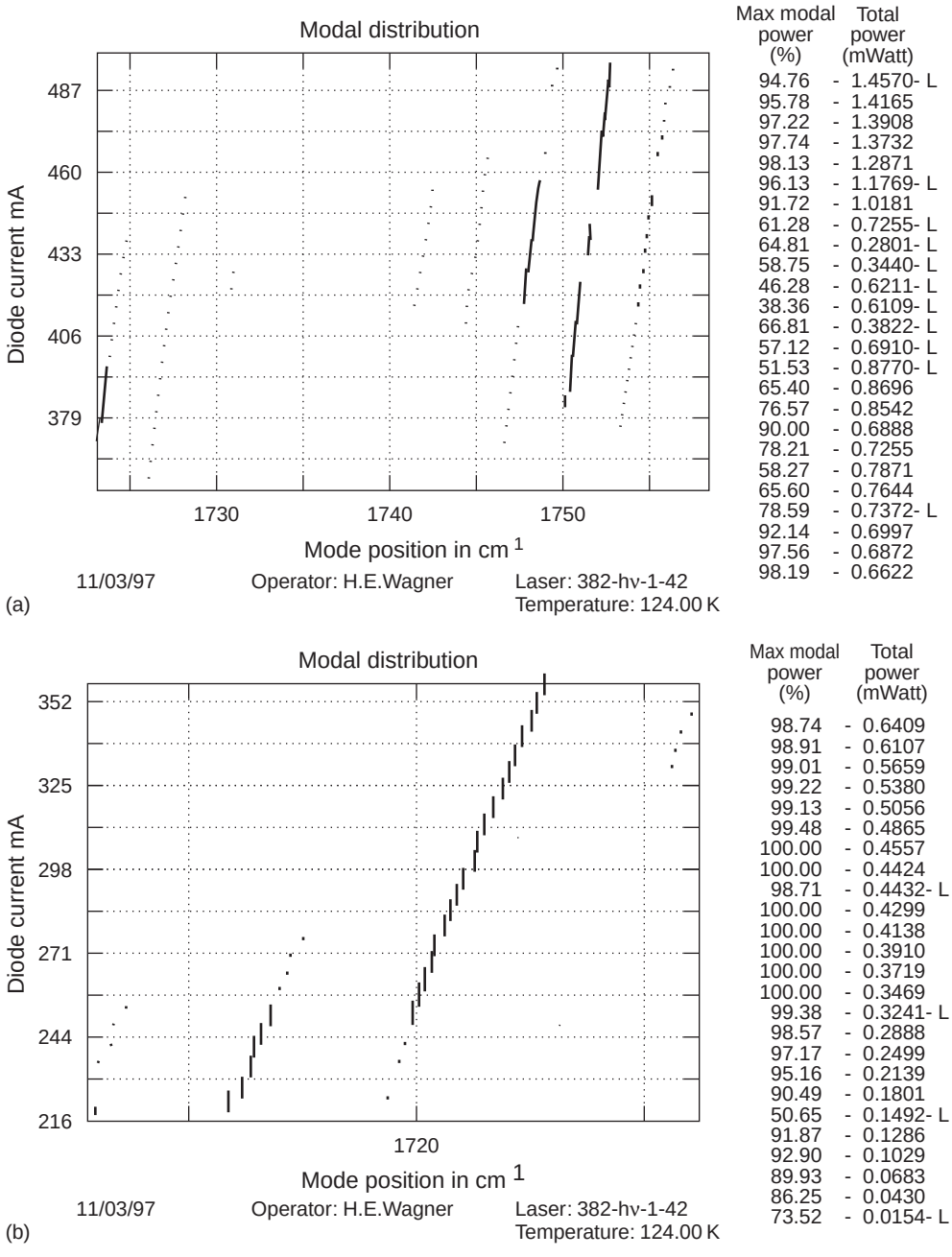

Figure 26 Mode spectra of a mid-IR diode laser. The spectra were gained by recording emission spectra at a temperature of 124 K and two laser operating currents around 450 mA (a), and 300 mA (b). Spectra were taken for variable currents indicated on the vertical axis, and spectral positions with laser mode emission are indicated in the figure by vertical bars that are proportional in length to the mode powers. The relative amount of power in the leading mode and the total mode power are listed. (Courtesy of laser components, a producer of mid-IR lasers).

in power stations with catalytic NO conversion. As the achievable sensitivity is strongly limited by the reduced absorption coefficients at such small wavelengths, the advantages of the short-wavelength technology are not available to most atmospheric analysis tasks.

At the time of writing (1999), a new laser type, the so-called quantum cascade lasers, has become available in some laser development laboratories. They make use of some technology features developed for short-wavelength III–V lasers. The major difference to the IV–VI diode lasers are their higher powers. Their operation temperatures are higher in pulsed mode, and lower in continuous mode, which is the mode used in most air monitoring systems. The tuning characteristics are

quite different, in that tuning by current is much slower. If used in an air-monitoring system, this imposes much less burden on temperature and current stability, but leads to larger power increase with tuning, which in turn burdens the techniques needed in order to discriminate absorption effects. Lasers of this type are at present not used in air-monitoring systems.

4.3 Conclusions

Diode laser gas analysis systems for air monitoring have sample cells 10–100 m in optical length and one to some liters in volume. The air is kept inside at reduced pressures close to 30 mbar, so that in the common flow-through mode only a negligible volume of air is sampled. System time constants are in minutes for very high sensitivity measurements with complete sample–calibration gas–background gas sequences. If fast measurement is sought, the air is directly sucked through the long path cell, and the time constants are of the order of seconds. System response times for very fast measurements at reduced sensitivity are still limited by the gas exchange time, although orders of milliseconds are achieved. If the air sample does not have to be changed, the system time constant is limited by the available laser power and detector time constants, and microseconds are reported.

Nearly all smaller molecules of interest can be monitored. Notable exceptions are all gases with vanishing or very small permanent dipole moments, such as N_2 and O_2. In general, no pretreatment of gas samples is needed. Only if heavily contaminated samples have to be treated, for instance due to soot or corrosive components, are the optical surfaces of mirrors and windows cause for concern. The measurement process itself is usually not affected by admixtures in air monitoring samples.

Optically, the bandwidth is of the order of 10^{-5}–10^{-6} of the wavelength. This high spectral resolution is needed for low-pressure spectroscopy; this is less critical at ambient pressure, but at the price of reduced sensitivity and selectivity. Single lasers are operated to scan intervals of 10^{-2} to about one cm^{-1}; scan times vary from some microseconds to milliseconds.

5 COMPARISON TO OTHER GAS-MONITORING METHODS

As a rule of thumb, whenever small to medium size molecules are to be analyzed with high temporal resolution, high sensitivity, high selectivity, and with quantitative comparison – all at the same time – the diode laser technique is very well suited. The major disadvantage at present lies in the price and sometimes the size of such equipment. Additionally, for highest performance, as is often needed in atmospheric analysis, skilled operation is required.

The most commonly used gas-monitoring methods are chemical in nature. Air samples are fed through volumes filled with sensitive materials that respond to specific molecules, for instance with a color change. These techniques are relatively inexpensive and the necessary equipment can be stored at locations of interest; however, they do not give exact quantitative results and can only be used once. Hence they are not suited to monitoring applications.

Similar arguments hold for sampling with enrichment and subsequent transfer to a laboratory. Again, only singular detections are possible; however, collection can be performed over prolonged time intervals, so that mean values can be determined. Although the collection step limits this technique, be it temporal resolution or possible chemical reactions between sampling and analysis, this type of analysis is very sensitive, very selective, and can be performed on a large number of compounds. Additionally, such analyses are capable of detecting large molecules such as polychlorinated biphenyls, which are not accessible to laser gas analysis in their typical atmospheric concentrations. As a rule for atmospheric analysis, laboratory techniques of this type and diode laser spectroscopy do not compete directly; if both are accessible there will be a clear decision, dependent on the technical and scientific goal.

Typical air-monitoring techniques are CLD for nitrogen oxides and flame ionization detection (FID) for hydrocarbons. Both are proven and their use is widespread. The disadvantages of both as compared to the laser technique is their limited selectivity; improved selectivity always stems from additional sample treatment. Although these are standard analysis techniques, their different operation principles and sample treatments can lead to different systematic errors, and hence they only give limited bases for detailed theoretical conclusions on atmospheric chemistry. The laser technique is advantageous in this regard, because it detects the different gas components with the same principle and from only one sample, which is not affected by the analysis technique itself.

Conventional spectroscopic techniques are also used for air monitoring. Nondispersive and filter techniques use large-bandwidth spectra compared to laser spectroscopy. Such systems are quite rugged and are standard for CO_2, for instance. Although useful for many applications, disadvantages lie in their limited sensitivity, selectivity, and multicomponent ability. IR spectroscopy is used frequently. If used with comparable optical pathlength and hence comparable potential sensitivities, these instruments suffer from the low intensity of their conventional

radiation sources, which makes large detector time constants necessary. In principle, IR systems exist that have the same 10^{-6} spectral resolution as diode laser spectroscopy. However, such systems are quite costly and bulky, can only be operated in laboratory environments, and have spectrum recording times of the order of an hour. This makes them a very good tool for gaining high-resolution reference spectra, but does not allow routine atmospheric analysis.

The final point about the analysis of clean air is the use of spectroscopic equipment with open-path measurement. In this case, the radiation is made to travel a considerable distance through the atmosphere, so that high sensitivity is gained. Measurements yield a column concentration, which is the average of the concentration along the optical path which may be as long as 1 km. Depending on the application, this may be an advantage or a disadvantage over local sampling. In general, open-path techniques can be performed both with diode laser spectroscopy and with conventional spectroscopy. The advantages of laser systems are higher sensitivity and higher measurement speed, whereas conventional systems offer better multicomponent capability. Both lack superior spectral resolution due to pressure broadening.

The discussion so far concentrates on a ''clean'' atmosphere. Close to pollution emitters, the situation is different, in that higher concentrations of many compounds make cross-sensitivity a problem; problems must also be expected from other contaminants, such as particulates. As an example, consider exhaust gas analysis. Standard techniques for car exhaust gas measurement essentially are influenced by the Californian legislation for the reduction of detrimental exhaust components. Established techniques for single components in exhaust gases are FID for hydrocarbons, CLD for nitrogen oxides and nondispersive IR analysis for gases such as carbon monoxide. Their underlying physical principles make these techniques unsuitable for multicomponent analysis and their sensitivities do not satisfy the increasing analytical requirements. In addition, FID cannot resolve individual hydrocarbons and the contribution to the signal to some extent depends on the compound itself and on the composition of the gas matrix. This matrix dependence is far more important than in clean air.

The analytical results obtained from single-component methods are not strictly comparable because the sample-conditioning procedures differ considerably from component to component. For example, CLD detection of NO requires an admixture of dried ambient air after separation of particles in order to generate excited NO_2 molecules from NO and oxygen and to subsequently determine the ultraviolet light emission. If, in addition, NO_2 has to be determined, an additional catalytic converter is required to reduce NO_2 to NO. The NO_2 concentration is then determined only indirectly, i.e. as the difference. Such gas-conditioning processes essentially determine the temporal response. In general, the temporal behavior is not strictly comparable to that of other gas components which are measured by other techniques, such as by FID with corresponding sample conditioning, including particle filtering and the admixture of other gases. Different sample-conditioning procedures potentially also result in different chemical reactions within the sample gas, which may lead to systematic errors that are difficult to identify.

Instruments for Fourier transform infrared (FTIR) analysis of exhaust gas have only been commercially available since the late 1990s.[45] They enable selective analysis of various exhaust gas components and allow for temporal resolution in seconds, all with a common sampling technique. The FTIR technique enables the measurements of exhaust gas components in those spectral ranges where absorption by water vapor is weak. Whereas inorganic exhaust components can be detected in a reliable manner, detection of individual hydrocarbons is possible only in some cases. The absorption bands around a wavelength of 3 μm allow detection of the sum of hydrocarbons; however, this also depends on the type of fuel. In practice, different analytical wavenumbers and band widths are used for petrol and diesel fuel. The FTIR technique therefore replaces the conventional techniques, with comparable results. Its noise-equivalent concentration limit is commonly accepted to be of the order of 1 ppm, which is mainly due to its limited selectivity. This degree of sensitivity, together with its relatively poor time resolution, make FTIR inadequate to meet the needs of car exhaust gas analysis in motor development.

One prime advantage of FTIR is its high calibration stability, which is a consequence of the IR detection principle. This property is also expected to hold for the IR diode laser technique. Moreover, FTIR has proved, that superimposed IR spectra can be evaluated very reliably by numerical methods. In principle, these properties will also hold for IR diode laser spectroscopy, and the more complex laser absorption signature is expected to achieve selective resolution of a large number of organic molecules.

An increasingly popular method is chemical ionization mass spectroscopy (CIMS),[46,47] based on selective chemical reactions of added gases, which are selected according to the exhaust gas component to be detected. The method is characterized by extreme sensitivity, multicomponent capability and high speed of response if applied to a single component. Stability of the signal is a problem, because it depends on the ion yield and hence also on the exhaust gas composition. As a result of the insufficient mass resolution, the necessary

selectivity can be achieved only by special ionization. Exhaust gas components with similar ionization energy and similar mass cannot be separated. These problems can be handled, if only classes of compounds as different hydrocarbons are measured.

Mass spectroscopy with preceding laser excitation[48] can in principle yield similar sensitivities to CIMS, but is limited by the number of available lasers with the required high power. Multicomponent analysis requires specific wavelengths for single groups of exhaust gas components, which are only partly achieved by fixed-frequency lasers, and are realized for example by using tunable dye lasers. The high expense for such lasers, if they are feasible, does not permit their economical operation under engine test bed conditions.

For scientific investigations, other methods have been successfully applied. Among these are nonlinear optical methods, such as coherent Raman spectroscopy.[49] This technique enables high spatial resolution, high selectivity, and high speed of measurement. For routine operation with test beds, this method is supposed to be too complex at present. Further problems originate in the high laser power levels under field conditions and in the skilled manpower requirement for operation. The same can be stated for other laser techniques that rely on nonlinear optics. They may represent an optimum choice for specific scientific investigations, but are not suited as universal methods for the large number of compounds that need to be analyzed with typical test beds for engine development.

In summary, many measurement systems are used for air monitoring. At the present state of development, system price and the demand for trained operators often are the main arguments against buying equipment using mid-IR diode laser spectroscopy for air monitoring. Technically and scientifically this technique is definitely superior to the others, whenever multicomponent measurement (with a limited number of separate components) is needed, with at the same time high selectivity, high absolute correctness, and high measurement speed.

ABBREVIATIONS AND ACRONYMS

CIMS	Chemical Ionization Mass Spectroscopy
CLD	Chemoluminescence
DBR	Distributed Bragg Reflector
DFB	Distributed Feedback
DSP	Digital Signal Processor
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
INDOEX	Indian Ocean Experiment
IR	Infrared
LBA-CLAIRE	Cooperative LBA Airborne Regional Experiment
LTE	Thermodynamic Equilibrium
MASERATI	Middle-atmosphere Spectrometric Experiment on Rockets for Analysis of Trace-gas Influences
MISOS	Methane Isotope Spectrometer
MS	Mass Spectrometry
OD	Optical Density
PC	Personal Computer
RF	Radiofrequency
RONALD	Rocket-borne Optical Neutral Gas Analyzers with Laser Diodes
SNR	Signal-to-noise Ratio
STREAM	Stratosphere Troposphere Experiments by Aircraft Measurements
TDLAS	Tunable Diode Laser Absorption Spectroscopy
TRISTAR	Tracer in situ Tunable Diode Laser Absorption Spectroscope for Atmospheric Research

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Clinical and Diagnostic Analysis

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Automotive Emissions Analysis with Spectroscopic Techniques • Differential Optical Absorption Spectroscopy, Air Monitoring by • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Mass Spectrometry in Trace Analysis • Laser-based Combustion Diagnostics • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Field-based Analysis of Organic Vapors in Air

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. J.H. Park, L.S. Rothman, C.P. Rinsland, H.M. Pickett, D.J. Richardson, J.S. Namkung, *Atlas of Absorption Lines from 0 to 17 900 cm^{-1}*, NASA Reference Publication 1188, NASA Langley Research Center, Hampton, 1987.
2. M. Tacke, R. Grisar, 'Laserphotometer für die Gasanalyse: Absehbare Eigenschaften und Entweicklungsaussichten (Laser Photometers for Gas Analysis: Prospective Properties and Development Potential)', *Laser und Optoelektronik*, **27**, 51 (1995).
3. H. Fischer, P. Bergamaschi, F.G. Wienhold, T. Zenker, G.W. Harris, 'Development and Application of Multilaser TDLAS Instruments for Ground-based, Shipboard and Airborne Measurements of Atmospheric Trace Gases', in *Application of Tunable Diode and Other Infared Sources for Atmospheric Studies and Industrial Process Monitoring*, ed. A. Fried, Proc. SPIE 2834, 130–141, 1996.
4. F.G. Wienhold, H. Fischer, G.W. Harris, 'Fast Response Tunable Diode Laser Spectroscopy for Trace Gas Flux Measurements', *Infrared Physics and Technology*, **37**, 67–74 (1996).
5. P. Bergamaschi, M. Schupp, G.W. Harris, 'High-precision Direct Measurements of $^{13}CH_4/^{12}CH_4$ and $^{12}CH_3D/^{12}CH_4$ Ratios in Atmospheric Methane Sources by means of a Long-path Tunable Diode Laser Absorption Spectrometer', *Appl. Opt.*, **33**, 7704–7716 (1994).
6. G.W. Harris, G.I. Mackay, I. Iguchi, L.K. Mayne, H.I. Schiff, 'Measurements of Formaldehyde in the Troposphere by Tunable Diode Laser Absorption', *J. Atm. Chem.*, **8**, 119–137 (1989).
7. G. Schmidtke, W. Kohn, U. Klocke, M. Knothe, W.J. Riedel, H. Wolf, 'Diode Laser Spectrometer for Monitoring up to Five Atmospheric Trace Gases in Unattended Operation, *Appl. Opt.*, **28**, 3665–3670 (1989).
8. H.I. Schiff, D.R. Karecki, G.W. Harris, D.R. Hastie, G.I. Mackay, 'A Tunable Diode Laser System for Airborne Measurements of Trace Gases', *J. Geophys. Res.*, **95**, 147–153 (1990).
9. G.W. Harris, D. Klemp, T. Zenker, J.P. Burrows, B. Matthieu, 'Tunable Diode Laser Measurements of Trace Gases During the 1988 Polarstern Cruise and Intercomparison with Other Methods', *J. Atmos. Chem.*, **15**, 315–326 (1992).
10. G.I. Mackay, K. Anlauf, H.I. Schiff, A. Wiebe, 'Measurements of NO_2, H_2CO and HNO_3 by Tunable Diode-laser Absorption-spectroscopy During the 1985 Claremont Intercomparison Study', *Atmos. Environ.*, **22**, 1555–1564 (1988).
11. J.E. Sickles et al., *Atmos. Environ.*, **24A**, 155–165 (1990).
12. G.L. Gregory et al., 'An Intercomparison of Airborne Nitric Acid Measurements', *J. Geophys. Res.*, **95**, 10 089–10 102 (1990).
13. J. Roths, T. Zenker, U. Parchatka, F.G. Wienhold, G.W. Harris, 'Four-laser Airborne Infrared Spectrometer for Atmospheric Trace-gas Measurements', *Appl. Opt.*, **35**, 7075–7084 (1996).
14. G.I. Mackay, D.R. Karecki, H.I. Schiff, 'Tunable Diode Laser Measurements of H_2O_2 and HCHO During the Mauna Loa Observatory Photochemistry Experiment', *Proc. SPIE*, **1715**, 126–137 (1992).
15. J.W. Drummond, *Proceedings of APAC '89*, 1989.
16. A. Fried, B. Henry, B. Wert, S. Sewell, J.R. Drummond, 'Laboratory, Ground-based, and Airborne Tunable Diode Laser Systems: Performance Characteristics and Applications in Atmospheric Studies', *Appl. Phys. B*, **67**, 317–330 (1998).
17. F. Slemr, G.W. Harris, D.R. Hastie, G.I. Mackay, H.I. Schiff, 'Measurements of Gas Phase Hydrogen Peroxide in Air by Tunable Diode Laser Spectroscopy', *J. Geophys. Res.*, **91**, 5371–5376 (1986).
18. H.I. Schiff, G.I. Mackay, D.R. Karecki, J.T. Pisano, S.D. Nadler, 'Measurements of H_2O_2 in the Grand Canyon by Tunable Diode Laser Spectrometry During the Navajo Generating Station Visibility Study', in *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, eds. R. Grisar, H. Boettner, M. Tacke, G. Restelli, Kluwer, Dordrecht, The Netherlands, 21–30, 1992.
19. D.J. Brassington, 'Measurements of Atmospheric HCl and NH_3 with a Mobile Tunable Diode Laser System', in *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, eds. R. Grisar, G. Schmidtke, M. Tacke, G. Restelli, Kluwer, Dordrecht, The Netherlands, 16–24, 1989.
20. H.I. Schiff, G.W. Harris, G.I. Mackay, 'Measurement of Atmospheric Gases by Laser Absorption Spectroscopy', *ACS Symp. Ser.*, **349**, 274–288 (1987).
21. C.R. Webster, R. May, C.A. Trimble, R.G. Chave, J. Kendall, 'Aircraft (ER-2) Laser Infrared Absorption Spectrometer (ALIAS) for In-situ Stratospheric Measurements of HCl, N_2O, CH_4, NO_2, and HNO_3', *Appl. Opt.*, **33**, 454–472 (1994).
22. A. Fried, J.R. Drummond, B. Henry, J. Fox, 'Versatile Integrated Tunable Diode Laser System for High Precision: Application for Ambient Measurements of OCS', *Appl. Opt.*, **30**, 1916–1932 (1991).
23. F.G. Wienhold, H. Fischer, P. Hoor, V. Wagner, R. Königstedt, G.W. Harris, J. Anders, R. Grisar, M. Knothe, W.J. Riedel, F.-J. Lübken, T. Schilling, 'TRISTAR – a Tracer in situ TDLAS for Atmospheric Research', *Appl. Phys. B*, **67**, 411–417 (1998).

24. J.U. White, *J. Opt. Soc. Am.*, **66**, 411 (1976).
25. D.E. Cooper, R.E. Warren, *J. Opt. Soc. Am. B*, **4**, 470 (1987).
26. A. Savitzky, M.J.E. Golay, *Anal. Chem.*, **36**, 1627 (1964).
27. F.-J. Lübken et al., *Appl. Opt.*, **38**, 5338–5349 (1999).
28. F.-J. Lübken et al., *J. Geophys. Res.*, **101**, 9489–9508 (1996).
29. K. Grossmann et al., *J. Atmos. Terr. Phys.*, **47**, 291–300 (1985).
30. H. Trinks, K.-H. Fricke, *J. Geophys. Res.*, **83**, 3883–3886 (1978).
31. C.P. Rinsland et al., *J. Geophys. Res.*, **97**, 20479–20495 (1992).
32. U. von Zahn et al., *J. Geophys. Res.*, **95**, 7443–7456 (1990).
33. F.-J. Lübken et al., *J. Geophys. Res.*, **102**, 13441–13456 (1997).
34. V.I. Fomichev et al., *J. Geophys. Res.*, **101**, 4041–4055 (1996).
35. F. Dingler, H. von Lucke, F.-J. Lübken, A. Kutepov, *Application of the MASERATI Instrument for the Detection of Non-LTE in the Mesosphere*, ESA special publication, SP-397, European Space Agency, Neuilly, France, 311–316, 1997.
36. T. Eriksen et al., *Appl. Opt.*, **38**, 2605–2613 (1999).
37. K. Gregorius, H. Schörner, 'Stack Gas Control by Diode Laser Spectrometers in Power Plants', in *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, eds. R. Grisar, H. Preier, G. Schmidtke, G. Restelli, D. Reidel, Dordrecht, 127–135, 1987.
38. W.F. Herget, J. Staab, H. Klingenberg, W.J. Riedel, *Progress in the Prototype Development of a New Multicomponent Exhaust Gas Sampling and Analyzing System*, SAE Paper No. 840470, 1984.
39. G.W. Harris, G.I. Mackay, T. Iguchi, H. Schiff, 'Measurement of NO_2 and HNO_3 in Diesel Exhaust Gas by Tunable Diode Laser Absorption Spectrometry', *Environ. Sci. Technol.*, **21**, 299–304 (1987).
40. W.J. Riedel, R. Grisar, U. Klocke, M. Knothe, H. Wolf, E. Bessey, P. Schottka, N. Pelz, 'Analysis of Trace Components in Automotive Exhaust Gas by TDLS', in *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, eds. R. Grisar, H. Böttner, M. Tacke, G. Restelli, Kluwer, Dordrecht, 319–324, Vol. 3, 1992.
41. J.A. Sell, R.K. Herz, D.R. Monroe, *Dynamic Measurement of Carbon Monoxide Concentrations in Automotive Exhaust Using Infrared Diode Laser Spectroscopy*, SAE Paper No. 800463, 1980.
42. J.A. Sell, R.K. Herz, E.C. Perry, *Time Resolved Measurements of Carbon Monoxide in the Exhaust of a Computer Command Controlled Engine*, SAE Paper No. 810276, 1981.
43. E. Nitzschke, H. Wolf, 'Dynamische Abgasmessung Mit Diodenlasern (Dynamical Exhaust Gas Measurement with Diode Lasers)', *MTZ Motortechnische Zeitschrift*, **52**, 362–368 (1991).
44. E. Wagner, R. Grisar, W.J. Riedel, M. Tacke, H. Wolf, 'Dynamic Exhaust Gas Analysis: A Tool to Study and to Improve Engine Management', in *Proceedings of 2nd International Conference on Control and Diagnostics in Automotive Applications*, Genoa, Italy, Paper No. 98A5017, 1998.
45. G. Hohenberg, G. Fehl, Ch. Wagner, 'Erprobung der FTIR-Abgasmeßtechnik zur Analyse der limitierten und nichtlimitierten Abgaskomponenten von Diesel- und Ottomotoren bei stationärem und dynamischem Betrieb', FVV, Frankfurt, Heft R 483, 1995.
46. H.P. Lenz und, G. Schopp, 'Messung nichtlimitierter Abgaskomponenten mit dem Massenspektrometer CIMS 500 am Institut für Verbrennungskraftmaschinen und Kraftfahrzeugbau der TU Wien', FVV, Frankfurt, Heft R 483, 1995.
47. H. Schubert, U. Guntow, K. Hofmann, R. Schlögl, 'Performance and Application Potential of Ion–Molecule Reaction Mass Spectrometry (IMR-MS) in the Analysis of Complex Gas Mixtures', FVV, Frankfurt, Heft R 483, 1995.
48. U. Boesl, 'Laser-Massenspektrometrie: eine empfindliche und multikomponentenfähige Abgasmeßtechnik', FVV, Frankfurt, Heft R 483, 1995.
49. D. Brüggemann, 'Potential der Raman-Meßtechnik zur motorischen Abgasanalyse', FVV, Frankfurt, Heft R 483, 1995.

Laser- and Optical-based Techniques for the Detection of Explosives

David L. Monts, Jagdish P. Singh, and Gary M. Boudreaux
Mississippi State University, Starkville, USA

1 Introduction 1
2 Physical and Chemical Properties of Explosives 2
3 Fundamental Characteristics 3
3.1 Effects of Physical Variables 3
3.2 Analytical Considerations 3
3.3 Interferences 5
4 Experimental Techniques 5
4.1 Infrared Spectroscopy 5
4.2 Raman Spectroscopy 7
4.3 Photoacoustic Spectroscopy 10
4.4 Ultraviolet/Visible Absorption Spectroscopy 11
4.5 Laser-induced Fluorescence 12
4.6 Ionization-based Detection Techniques 14
5 Environmental Applications 16
6 Future Development 18
Acknowledgments 18
Abbreviations and Acronyms 19
Related Articles 19
References 19

A wide variety of optical- and laser-based techniques have been and are currently being investigated as means of identifying and characterizing explosive materials. These efforts are hampered by the very low volatility of explosive compounds at room temperature and their tendency to ignite at higher temperatures where their vapor pressures are beginning to reach a regime where the species can be readily detected. Traditional, condensed-phase absorption spectroscopy (both infrared (IR) and ultraviolet/visible (UV/VIS)) requires larger samples than are frequently available and, although useful as confirmatory techniques, do not provide signatures that enable unique identification of species. The choice of nontraditional techniques for analysis of explosive materials is dependent upon the amount of sample available, the sample matrix, and sample-imposed constraints upon the measurement. Raman spectroscopy can identify and quantify condensed-phase explosive materials, but the detection limit depends upon substrate, excitation wavelength, and illumination area: limits of detection (LODs) vary from 0.05 ng for 2,4,6-trinitrotoluene (TNT) on glass to 10 μg for nitroglycerin (NG) on silica gel. The other detection techniques considered require volatilization of either the explosive compound itself or more often of characteristic decomposition products, such as NO, NO_2, or N_2O. Using IR irradiation, researchers have demonstrated that photoacoustic spectroscopy (PAS) has detection limits ranging from 0.55 ppb (parts per billion) for vapor-phase NG with 9 μm excitation to 220 ppm (parts per million) for vapor-phase 2,4-dinitrotoluene (DNT) with 6-μm excitation. IR laser differential absorption detection of dissociation products NO, NO_2, and/or N_2O can achieve detection limits of a few picograms. Laser-induced fluorescence (LIF) detection of the photofragments (typically NO) of explosive compounds yields detection limits in the tens to hundreds of ppb (by weight) range. Low LODs can be achieved using ion detection techniques: resonance-enhanced multiphoton ionization (REMPI) spectrometry is capable of LODs for detecting vapor-phase explosives compounds in the hundredths to tens of ppb range; ion mobility spectrometry (IMS) has vapor-phase detection limits for common explosive compounds of typically 200 pg, but for some species can be significantly lower: for example, 1 pg for TNT.

1 INTRODUCTION

Chemical detection and analysis of explosives is playing an increasingly important role in the modern world.[1–3] There are basically four common situations requiring analysis of explosive materials:

1. Criminal investigations of a bombing require forensic identification of the explosive materials involved in order to provide information on who the perpetrators may be and ultimately to link suspects with the type of explosive materials used. The thermal instability of most explosive materials leaves very little of the original explosive material following an explosion. Thus detection and identification of explosive residues on debris materials is very difficult. Consequently high detection sensitivity is required for successful analysis.
2. Detection of explosive materials concealed in mail or in airline baggage is also important in order to prevent crimes. Highly sensitive instruments capable of detecting explosives in air sampled from mail or baggage is required. Since identification can

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

be hindered by the use of less common explosive materials,[4] the repertoire of species detected needs to be expanded.

3. Since composition variation affects stability, detonation sensitivity, ignitability, and performance of explosives, accurate knowledge of the exact composition of explosive materials is important for their safe handling and storage.
4. Analyses may be necessary in order to address environmental and/or health-related concerns. Because many explosive materials exhibit some toxicity,[5] ingestion or inhalation of explosive materials presents a significant health hazard. Sites where analysis may be required include industrial facilities where explosive materials are manufactured in order to insure the safety of the workers and of nearby residents. In addition, sites may have become contaminated by improper storage or disposal of explosive materials, and analysis may be required as part of an environmental clean-up of the area.

2 PHYSICAL AND CHEMICAL PROPERTIES OF EXPLOSIVES

The chemistry of decomposition of explosive materials has been extensively investigated[6–10] with many efforts still on-going. We will concentrate on those physical and chemical properties of explosives that are pertinent to their detection and characterization.

Laser-based explosive detection involves recording information on the intensity and wavelengths of characteristic absorptions and/or emissions of samples that may contain explosive materials. It is the physical and chemical properties of the explosives that govern or constrain the methods and techniques of laser-based detection. Most laser-based techniques focus on vapor and/or particle detection and were developed from knowledge of molecular properties and spectroscopic signatures of the target.

Explosives are categorized or distinguished by their chemical structure. They can be divided into the following groups:[11] nitro compounds; nitric esters; nitramines; salts of nitric, chloric, and perchloric acids; azides and other compounds capable of exploding; and mixtures. Nitro compounds are distinguished by the carbon–nitrogen bond ($C-NO_2$). Most compounds which have three or more nitro groups for one benzene ring and some which have two nitro groups (dinitrobenzenes and dinitrotoluenes) are used as explosives. Nitric esters are characterized by the presence of a nitroxy group $-C-O-NO_2$. This means the nitro group in nitric esters is attached to a carbon atom by means of an oxygen atom. Nitramines are characterized by nitro groups attached to carbon atoms via a bridging nitrogen atom $-C-N-NO_2$. They are basically derivatives of the simple inorganic nitramine NH_2NO_2. Salts of nitric, chloric, and perchloric acids can be used as explosives or in explosive mixtures as oxygen carriers. Azides include compounds such as hydrazoic acid (HN_3), lead azide ($Pb(N_3)_2$), and silver azide (AgN_3). Other compounds such as fulminates and diazo compounds also fall under this category. Mixtures are simply common explosives combined to obtain the required characteristics.

Hence most organic compound explosives have a chemical structure with a bound NO_2 functional group and are subdivided by which atom attaches to this group. Table 1 lists properties for five common explosives with this property. (For more information about other explosive materials, see Urbanski[11] and Köhler and Meyer.[12]) The several different methods of measuring these properties have led to several different values, especially for vapor pressures. The vapor detection methods are limited because the room temperature vapor pressures of these explosives are small. There is a large increase in vapor pressure at elevated temperatures, but other properties such as ignition temperature limit the maximum detection temperature and introduce various dangers. Also, explosives undergo various physical and/or chemical changes at these elevated temperatures. Table 1 lists equations[13] used to calculate the vapor pressure

Table 1 Physical and chemical properties of common explosives

Acronym	Name	Melting point (°C)	Ignition temperature (°C)	Vapor pressure equation[a]
TNT	2,4,6-Trinitrotoluene	82	300	$\text{Log}\,P(\text{ppb}) = 19.37 - 5481/T$
RDX	1,3,5-Trinitro-1,3,5-triazacyclohexane	204	229	$\text{Log}\,P(\text{ppt}) = 22.50 - 6473/T$
HMX	1,3,5,7-Tetranitro-1,3,5,7-tetrazacyclooctane	278	335	
PETN	Pentaerythritol tetranitrate	141.3	205	$\text{Log}\,P(\text{ppt}) = 25.56 - 7243/T$
NG	Nitroglycerin	13.2		$\text{Log}\,P(\text{ppb}) = 18.21 - 4602/T$
AN	Ammonium nitrate	169.6		$\text{Log}\,P(\text{ppb}) = 12.97 - 541.77/T$

[a] T is in Kelvin.
ppt, parts per trillion.

for any given temperature, which were developed using measurements from a vapor generator apparatus and a least squares analysis.

3 FUNDAMENTAL CHARACTERISTICS

3.1 Effects of Physical Variables

For each of the situations requiring explosive analysis, the explosives may react with and be affected by their environment. Detection of explosives in open air and in underwater environments provides significant challenges when compared with detection in buildings or vehicles. Many detection techniques employ in situ or noncontact measurements and applications often require sensor mobility, minimal sensor size/weight, and unique processing methods. Detection systems must be sensitive enough to find small amounts hidden in a complex environment, and must be able to discriminate between threatening and harmless materials.

Explosive detection devices must be designed differently for different threats. Aviation security, law enforcement, counterterrorism, process quality, health safety, and environmental remediation groups require widely different instruments and techniques. For explosives detection, the largest physical barrier is the method by which the explosive sample is collected.

There are primarily two methods for introducing the sample to the detector: vapor or particle sampling.[3] There are also two distinct techniques, portal screening (a fixed installed sampler) and portable/hand-held samplers. Portals are designed to sample explosives mainly from people. They use techniques such as vacuums, high-velocity airflow, wiping, or a combination of these. Portable or hand-held samplers are used in the field or a laboratory to collect samples from the ground, groundwater, or atmosphere.

There are several orders of magnitude difference between the density of a solid explosive sample and a typical saturated explosive vapor. Hence detection methods for solid-phase explosives may require only a microscopic sample, while vapor detection requires a large sampling volume and methods to increase vapor pressure and overcome the strong surface adsorption of these vapors by the transport mechanism. Vapor sampling is also limited because many explosives are deliberately concealed or sealed and the vapors may be attenuated or adsorbed by the containment materials.

3.2 Analytical Considerations

Factors affecting the detection limits include sample purity and sample thermal history. To produce readily useable commercial and military devices, explosives are commonly mixed with binders, such as waxes, thermoplastics, or rubbers. This has the effect of reducing the relative amount of explosive present. If macroscopic quantities of sample are available, then chemical and/or physical separation may be necessary prior to analysis. Subjecting explosive materials to heating may induce physical and/or chemical changes in the substances; thus the characteristic spectroscopic signatures may be different depending upon the amount and extent of heating.

3.2.1 Solid Sample Mixtures

Explosive material molecules tend to cling tightly to surfaces. Handling of explosives is sufficient to transfer detectable amounts of explosive materials to hands and subsequently to other surfaces touched by hands.[14,15] In the case of RDX, eight consecutive hand washings with soap and water were required before the RDX could no longer be detected on the hands.[16] Although the vapor pressure is very low at room temperature, TNT has been shown to contaminate surfaces (such as wood, rubber, aluminum) placed near, but not touching the TNT.[17] In the case of porous materials such as wood, significant TNT contamination was found not only on the surface, but throughout the wood.[17] This tendency of explosive materials to spread can be an advantage for forensic investigations and for efforts to prevent terrorist activities. However, it also implies that great care must be taken to insure that any explosive material detected is from the sample of interest and not from previous contaminated samples (memory effects), from other samples stored in the laboratory, or from handling of the sample.

With regard to forensic investigations or efforts to prevent terrorism, one of the primary difficulties is locating the small samples (often milligrams or less) that might be located on a large item. As noted above, this search is hindered by the very low vapor pressures of most explosive materials at room temperature. Once located, the investigator must decide whether to attempt analysis in situ or to separate the explosive from the matrix by heating or chemical means. The advantage of analyzing the sample in situ is that there is no loss to adhesion to sampling apparatus walls or to thermal degradation. The disadvantage of analyzing in situ is that the presence of other compounds may preclude definitive identification and quantification. Preconcentration may be required for successful detection.[3] With regard to analyses for process control or to address environmental or health-related concerns, analyses are subject to the usual concerns about how to representatively sample (heterogeneous) solid mixtures.

Solid mixture samples can be characterized using IR spectroscopy, Raman spectroscopy, UV/VIS spectroscopy, and fluorescence spectroscopy. Vaporization or solvation of the sample also allows these and other techniques to be employed. As mentioned above, the use of chemical or physical separation facilitates definitive identification and quantification.

3.2.2 Liquid Mixture Samples

In principle, liquid mixture samples are the easiest to analyze since: (1) chemical separation prior to analysis is facilitated in the liquid phase and the results of analysis of purified samples are more definitive than those of mixtures; (2) preconcentration techniques are simple for liquid mixtures; (3) liquids avoid the difficulties arising from low vapor pressures; and (4) several of the techniques require the sample to be introduced as a liquid. In practice, the presence of liquids interferes with detection by some of the techniques, for example, LIF where water quenches the fluorescence of explosive materials.

3.2.3 Gas-phase Mixture Samples

A major difficulty in measurement of gas-phase explosives arises from their pronounced tendency to stick to surfaces, especially when metal surfaces are present.[18] This can preclude accurate determination of the vapor pressure because, even for the case of an in situ sample, the vapor pressure will equilibrate between the rate of vaporization and the rate of plating out.

Calibration is one of the most important parts in verifying or developing any analytical technique. For calibration, one requires a standard source to produce with good precision a known amount of analyte. Several research groups have reported vapor generator calibration sources for explosive materials.[19–24] Pella[19] constructed a vapor generator for TNT, DNT, and ethylene glycol dinitrate (EGDN), obtaining concentrations as low as 0.05 ppb with a systematic error within 15 to 20% of the values expected for TNT. He was able to generate an equilibrium vapor concentration of the explosive at a known temperature by passing N_2 through a column containing explosives. Gas chromatography was used to evaluate the operation of this system.

Davies et al.[20] have designed and calibrated a pulsed vapor generator for TNT, RDX, and PETN (Figure 1). The pulsed generator could deliver between 1 and 1000 pg of explosive in a pulse of less than 5-s duration with outputs within ±20% of the expected value. Their vapor generator had four main components: a temperature-controlled explosives vapor reservoir; a flow control manifold; thermostated air supply; and data acquisition and control system. The explosives reservoir contained ~0.1 g of solid explosive suspended on quartz beads. The temperature of the explosives reservoir was monitored by two thermocouples and controlled to within ±0.1 °C by controlling the flow rate of the cooling and heating air. A nichrome wire heater was used to keep the outlet temperature 4 ± 1 °C higher then rest of the generator to avoid the deposition of explosive vapors at the outlet. The carrier gas was controlled by valve V-1. A three-way solenoid valve SV-1 was used to switch the flow of carrier gas between set-up manifold and the generator coil. During initial set-up and between explosive vapor pulses, the carrier gas flowed through the set-up manifold. In order to calibrate the vapor generator,

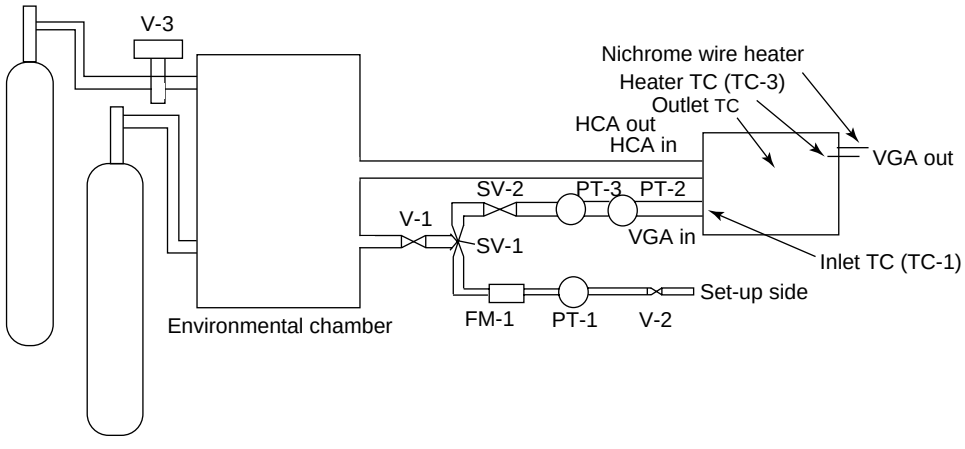

Figure 1 Schematic of pulsed vapor generator system for TNT, RDX, and PETN.[20] FM-1, mass flow meter, 0–250 cm^3; PT-1, pressure transducer, 0–5 psig; PT-2, pressure transducer, 0–3.6 psig; PT-3, pressure transducer, 0–25 psig; V-1, flow control, micrometer needle valve; V-2, set-up line resistance, micrometer needle valve; V-3, flow control valve, 0–200 SLM; SV-1, three-way solenoid valve; SV-2, two-way solenoid valve; TC, thermocouple, type T. (Reproduced by permission of American Chemical Society.)

they collected explosive vapors in vapor collection tubes; the samples were subsequently transferred to an ion mobility spectrometer (see section 4.6.1) for quantification.

Eiceman et al.[24] have tested and characterized a vapor generator for TNT, RDX, and PETN from 79 to 150 °C. The mass output rate was stable for several hundreds of hours. This instrument was calibrated using the National Institute of Standards and Technology (NIST) gravimetric method. The mass flow rates of TNT and RDX were consistent whereas PETN was higher. This is due to molecular decomposition of PETN at higher temperature. The vapor output of TNT and RDX was in pure molecular form in the operating range (79–150 °C) of the instrument. This system was good for calibrations involving TNT or RDX, but had limited use for PETN.

3.3 Interferences

All techniques are potentially subject to interferences that can result in an incorrect determination of a species' concentration. There are several types of possible interferences, including chemical, ionization, physical, and spectral.[25] Chemical interferences occur when other species present chemically react with the species of interest to transform it into a species that is not detected or whose spectral (or mass) signature is different from that of the parent species. Ionization interference occurs when the presence of another species enhances/reduces ionization of the species of interest. Physical interferences are related to changes in the physical characteristics of the sample, such as changes in viscosity, vapor pressure, surface tension/adhesion, etc. Spectral interference occurs when another species present absorbs light at the same wavelengths as the species of interest. Techniques that monitor signal intensity at a specific wavelength rather than by recording the spectrum over a selected region are particularly prone to spectral interference; however, it is often possible to achieve lower detection limits with single-wavelength techniques than with spectrum-recording techniques so the experimenter must decide whether sensitivity or selectivity is more important. Chemical and/or physical separation prior to analysis can substantially reduce interference effects. However, for in situ or screening measurements, purification of the sample may not be feasible.

4 EXPERIMENTAL TECHNIQUES

4.1 Infrared Spectroscopy

Vibrational transitions occur in the IR spectral region. Because the location of the peaks depends on the molecular structure, the IR is often referred to as a "fingerprint" region because the pattern of the spectrum can be used to identify the associated species. In the IR region, the NO_2 group produces two high-intensity characteristic bands that are associated with the symmetric (ν_s) and asymmetric (ν_{as}) stretching vibrations of the N–O bonds within the NO_2 moiety.

In 1960, Pristera et al.[26] reported the IR spectra of 68 explosive ingredients. Based upon their work, characteristic frequencies have been identified; these are summarized in Table 2. Chasan and Norwitz[27] extended this work by recording the IR spectra of 43 of the most common ingredients of primers, tracers, igniters, incendiaries, boosters, and delay compositions. Subsequent investigations have reported the IR spectra of selected explosive materials.[28–36]

Analysis of an unknown complex explosive solely by IR spectroscopy is generally not recommended because of the uncertainty in interpretation of the spectrum. The certainty with which analysis of the IR spectrum can provide definitive identification of a material greatly increases if prior chemical separation is employed. Beveridge et al.[37] reported that in a large

Table 2 IR bands of explosives. (Adapted from Yinon[2] and Pristera et al[26])

Type of explosive	Band peak	
	cm^{-1}	μm
Sym-trinitro compounds that have the following additional groups: CH_3, C_2H_5, OCH_3, OC_2H_5, COOH, OH, NH_2	1081	9.25
m-Dinitro compounds that have the following additional groups: CH_3, C_2H_5, OCH_3, CHO, COOH, $N{=}NH_2$, CH_3NH, C_2H_5NH	930–922	10.85–10.95
In trinitro compounds (addition of NO_2)	939–909	10.75–11.0
o-Mononitro compounds that have one of the following groups ortho to the nitro: CH_3, C_2H_5, CHO, COOH, NH_2	787–781	12.7–12.8
p-Mononitro compounds that have one of the following groups para to the nitro: CH_3, C_2H_5, CHO, COOH, NH_2	1111	9.0
Nitramines	1282	7.8
Organic nitrates	1667, 1282, 833	6.0, 7.8, 12.0
Inorganic nitrates	833	12.0

number of analyses of explosive residues, IR spectroscopy detected explosive components in 50% of the tests in contrast to 80% for thin-layer chromatography. Thus IR spectroscopy is a good confirmatory test if sufficient quantity of sample is available.

Wormhoudt et al.[38] have reported their efforts to develop a tunable infrared laser differential absorption spectroscopy (TILDAS) system for in situ identification of the presence of explosives in subsurface soil.[39] In laboratory tests, a 900 °C pyrolyzer ring heater was placed a short distance (about 0.25 cm) from the soil surface and the soil heated for periods on the order of minutes. The heat from the pyrolyzer first volatilized and then thermally decomposed the explosive TNT in the laboratory tests. The volatile pyrolysis products (which include NO, NO_2, and CO_2) were collected by a ceramic tube located at the center of the ring heater, as shown in Figure 2. The gases then passed through flexible tubing (to permit motion of the pyrolyzer over the soil) and into the multipass detection cell, which is sometimes referred to as a White cell.[40] Spherical mirrors within the multipass cell reflect the laser beam back and forth many times, producing an effective path length for detection of many times the physical length of the cell. Since according to the Beer–Lambert law the absorption magnitude is related to the product of the absorption coefficient, the concentration, and the path length, increasing the path length lowers the minimum detectable concentration. Two tunable lead-salt laser diodes were used for excitation: one excited a NO vibrational transition, and the other excited a NO_2 vibrational transition. The absorptions of the two species as a function of time were monitored. NO and NO_2 had roughly parallel heating time behavior, but the NO signal was much larger than the NO_2 signal. Low NO_2 signal levels were ascribed to rapid secondary reactions of NO_2 produced in the initial bond-breaking steps rather than to a lack of NO_2 formation from the earliest stages of TNT thermal decomposition. The NO_2 disappeared as soon as the heat was removed while some NO was detected for a longer time. Currently the fieldable TILDAS instrument has minimum detectable fractional absorptions in the range 10^{-4} to 10^{-5}. Assuming a minimum absorbance of 10^{-5}, the resulting detection limits for NO, NO_2, and N_2O are presented in Table 3. N_2O is not a major product of TNT decomposition, but it is for decomposition of nitramines (such as HMX and RDX). Therefore, monitoring of N_2O would allow nitramines to be distinguished from nitroaromatics.

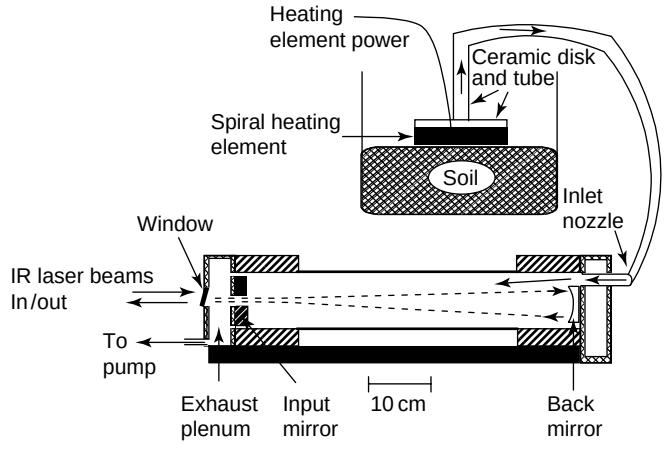

Figure 2 Schematic drawing of key elements of a laboratory investigation by TILDAS[38] of heating explosives-contaminated soil, including soil sample, planar spiral heating element with central sampling tube, and laser absorption multipass cell. (Reproduced by permission of Optical Society of America.)

Riris et al.[41] have demonstrated detection of explosives using catalysts to facilitate decomposition of explosives and subsequent detection using a frequency modulation technique. The frequency modulation technique consists of rapidly changing the laser wavelength between two chosen wavelengths: one of which corresponds to a transition of a species of interest and the other is a nearby wavelength that does not excite that species. By comparing the on-resonance and off-resonance signals, the concentration of that species can be deduced. A schematic of their apparatus is presented in Figure 3. Three different lead-salt laser diodes were used: one was tuned to 4.57 μm where N_2O absorbs, another to 5.25 μm where NO absorbs, and the third to 6.24 μm where NO_2 absorbs. The three laser beams were combined and sent through the sample cell and then through a calibration cell that was filled with known quantities of N_2O, NO, and NO_2. Dichroic beam splitters were used to separate the laser beams exiting the cells and to direct the light onto IR detectors that were dedicated to each species. In this manner, the concentrations of N_2O, NO, and NO_2 were simultaneously determined. For their prototype, their LOD for RDX was 5–10 pg, although the theoretical LOD was 0.005 pg RDX. They anticipate that in a fieldable system, it will be possible to reduce the achievable LOD by an order of magnitude.

Table 3 Minimum detectable mixing fractions by TILDAS[38] for a 5.2-m total path and an assumed minimum detectable peak absorbance of 10^{-5}. (Reproduced by permission of Optical Society of America)

Molecular species	Frequency (cm^{-1})	Minimum detectable mixing fraction (parts in 10^9 by volume)
NO	1900	4
NO_2	1630	0.6
N_2O	2210	0.3

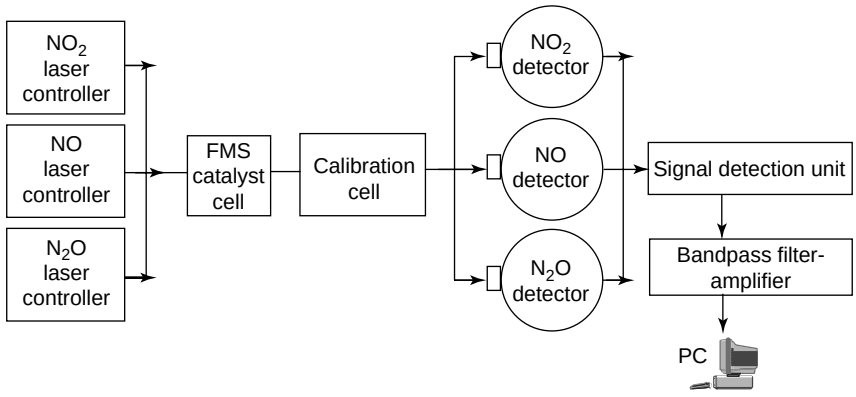

Figure 3 Schematic of IR frequency modulation spectrometer detection system.[41] (Reproduced by permission of Optical Society of America.)

4.2 Raman Spectroscopy

In simplistic terms, the Raman effect signal results from an inelastic collision between a photon and a molecule.[42,43] A Raman transition is separated from the excitation laser wavelength by an energy difference (Δv or Δw, usually measured in units of cm^{-1}) that corresponds to a vibrational energy difference in the molecule's ground electronic state. As shown in Figure 4, those Raman transitions that are at lower energy than the excitation laser are said to be Stokes transitions and those that are at higher energy are termed anti-Stokes transitions. Both types of transitions can simultaneously occur in a single Raman spectrum. Since the intensities of anti-Stokes transitions is much, much weaker than those of Stokes transitions, the Stokes transitions are generally used for analytical characterization.

Most of the Raman spectroscopy research directed toward detection of explosives has concentrated on the use of two approaches: Fourier transform Raman spectroscopy (FTRS) and Raman laser microprobe spectroscopy. Figure 5 shows a schematic of a FTRS apparatus. A rejection filter is placed between the laser and the beam separator to insure that only one wavelength of light is used to excite the sample; this is done to filter out stray flashlamp light from neodymium : yttrium aluminum garnet (Nd : YAG) laser beams or other lines from argon ion (Ar^+) laser beams. The laser beam irradiates the sample and the reflected light (which includes the Raman signal) is directed into a scanning interferometer. The resulting interferogram is detected and recorded on a computer. Fourier transforming the interferogram yields the entire Raman spectrum. Since the entire spectrum is recorded simultaneously, this technique requires a much shorter data acquisition time than those that require scanning of a spectrometer. Typically 100–500 scans are co-added to improve the signal-to-noise ratio (S/N).

A schematic diagram of a Raman laser microprobe apparatus is presented in Figure 6. Many researchers have devised different Raman laser microprobe configurations.[45] In the variation developed for explosives by Hayward et al.,[46] a fiber optic cable was utilized to transport the light from the laser to the microprobe head. Inside the microprobe head, the laser beam was filtered to remove any unwanted laser wavelengths and then directed onto a microscope objective lens that focused the light to a small diameter, typically on the

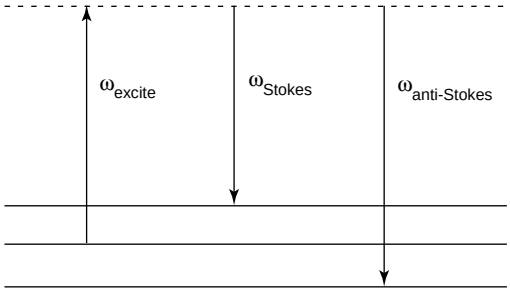

Figure 4 Schematic of the Raman effect process.

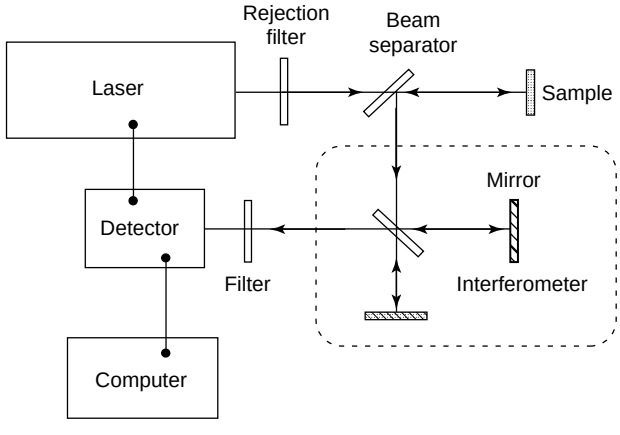

Figure 5 Schematic of laser FTRS apparatus.[44]

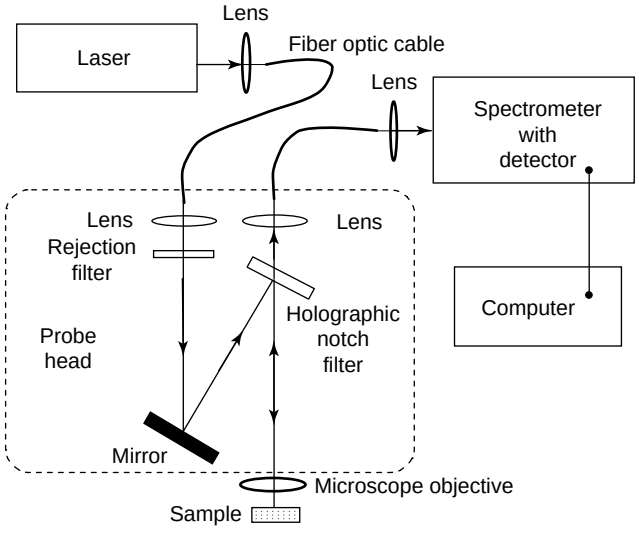

Figure 6 Schematic of laser Raman microprobe apparatus.[46] (Reproduced by permission of ASTM.)

order of tens of micrometers. The Raman signal passed back through the microscope lens and then through a holographic notch filter that transmitted the Raman-shifted signal, but not the laser wavelength. The signal was focused into a fiber optic cable for delivery to a scanning spectrometer for detection by a multichannel charge-coupled device (CCD) detector. The advantage of the Raman laser microprobe technique is that by increasing the laser flux by focusing the laser beam to a small diameter, it is possible to detect lower concentrations than is possible with the FTRS technique. One of the limitations of the Raman laser microprobe technique is the necessity of precisely positioning the laser beam. Most reports have not dealt with the problem of identifying the exact location of the sample; sample location is one of the primary challenges facing those who perform forensic analyses of bomb explosion sites since minuscule explosive residuals may be present on objects hurled significant distances by the force of the explosion. Hayward et al.[46] utilized an XYZ stage to translate the sample while monitoring the Raman signal in the 400–1300 cm^{-1} range and more thoroughly investigating those regions exhibiting a strong Raman peak near 880 cm^{-1} from RDX or PETN.

Table 4 compares the LODs achieved using macrosampling and microsampling (Raman laser microprobe) as determined by Carver et al.[47] By moving a mirror, they could change from microprobe to macrosampling. Using the microprobe, slightly lower LODs could be obtained than with macrosampling (FTRS): this is due to the opposing contributions from more efficient signal collection by the microprobe and the smaller sample area for microprobes (typically a few tenths of a square millimeter for microprobes vs a few square millimeters for macrosampling).

Factors affecting the detection limits include: sample purity; sample thermal history; backscattered radiation from the substrate; the Raman spectrum of the substrate; fluorescence of substrate; packing; excitation wavelength; fluorescence of analyte; laser power; detector sensitivity; and laser beam spot size on the sample. Backscattered radiation is much more pronounced from silica gel than from activated charcoal or glass.[47] However, activated charcoal produces its own Raman signals, which may interfere with those of explosive analytes.[48] Carver et al.[47] believe that glass is the ideal support material for Raman detection of explosives because it has a weak Raman spectrum and gives only minimal backscatter. Lewis et al.[49] have reported both solid and liquid explosives could be successfully analyzed in common glass and plastic containers when near-infrared (NIR) excitation is used.

Whether the sample is loosely packed or tightly compressed can have a strong effect on the spectral S/N.[49] Compressing the solid sample increased the rate of heat transfer from the sample to the container and hence reduced the effects of laser heating; for the loosely packed particles, the effects of laser heating of the sample were manifested by thermal emission above 2500 cm^{-1} and a significant baseline offset.

The effect of laser excitation wavelength on the Raman signal is also important.[49] The Raman intensity increases with increasing laser excitation frequency (and hence

Table 4 Raman spectroscopy detection limits for macrosampling (FTRS) and for microsampling (Raman laser microprobe) for a variety of common substrates.[47] The excitation wavelength was 514.5 nm, and the laser power was 20 mW

Explosive	Silica gel macrosampling (μg)	Silica gel microsampling (μg)	Charcoal macrosampling (ng)	Glass microsampling (ng)
NG	10	–	–	–
PETN	5	1	40	0.5
RDX	5	1	40	1
TNT	1	0.5	20	0.05

decreasing laser excitation wavelength) since the Raman intensity is related to the scattered frequency ν as shown in Equation (1):[43]

$$I_{v,J} = \left(\frac{2\pi}{c}\right)^4 \nu^4 I_L \frac{N_{v,J}}{2J+1} \langle\alpha\rangle^2, \Omega\ell \qquad (1)$$

where $I_{v,J}$ is the Raman intensity, c is the speed of light, ν is the frequency of the Raman scattered radiation, I_L is the intensity of the incident radiation, $N_{v,J}$ is the population in vibrational state v and rotational state J, α is the polarizability matrix element, Ω is the solid angle, and ℓ is the scattering length. Thus the shorter the excitation wavelength, the larger the Raman intensity. Since fluorescence is typically orders of magnitude more intense than the Raman signal,[42,43] it is essential to minimize fluorescence in order to obtain quantitatively reliable Raman spectra. Because fluorescence typically increases with decreasing excitation wavelength, the gain in Raman intensity is counterbalanced and often overpowered by an increase in fluorescence with decreasing wavelength. Thus, using longer excitation wavelengths decreases fluorescent interference, but increases data acquisition time: in order to obtain Raman spectra of 3,3′-diamino-2,2′,4,4′,6,6′-hexanitrophenyl (DIPAM) of approximately the same quality, Lewis et al.[49] needed 1 s integration time at 632.8 nm with 10 mW laser power, 10 s integration time at 785 nm with 5 mW laser power, and for a FTRS system at 1064 nm, 10 min integration time with 100 mW laser power. They[49] tested 32 pure or composite explosives and found that 23 of the samples could be adequately analyzed using all three excitation wavelengths that they investigated (632.8, 785, and 1064 nm); however, they concluded that the optimum laser excitation wavelength for analysis of explosives is probably in the range between 785 nm and 1064 nm.

Thermal and/or laser heating of the sample can thermally decompose explosive samples. McNesby et al.[50] noted that TNT can be decomposed due to exposure to laser radiation (and hence laser heating); for TNT they had to use lower laser powers to prevent the sample from combusting. They[51] also noted that thermal decomposition begins below the melting point of RDX and RDX-based propellants. In addition to thermal decomposition, laser heating may result in sample melting or sample vaporization. Based on similar chemistry, similar thermal effects can be expected for other explosives. Hodges and Akhavan[52] encountered problems when black or darkly colored samples were investigated: these samples efficiently absorbed the IR radiation, leading to sample heating and eventual burning. They[52] investigated a variety of techniques in an effort to overcome this difficulty. These efforts included defocusing the laser beam, reducing the incident laser power, accumulating single scans, and collecting data at lower resolution (and hence with shorter data acquisition times). None of these efforts resulted in success. McNesby et al.[51] report that laser sample heating is more important for 1064-nm excitation than for 514-nm excitation. Moreover, laser sample heating becomes increasing significant as the bulk sample temperature is raised.

McNesby et al.[50,51] studied the effects of bulk and laser heating on the Raman spectrum of solid explosive samples. In general as the temperature of a sample increased, the intensity of Raman transitions decreased and their peak frequencies moved to lower energy. There was also a noticeable variation in relative Raman peak intensity with increasing temperature, which has been interpreted as an indication of an increase in the lattice spacing and hence a loss of order within the crystal. They[51] believe that in order of decreasing importance, for their experimental set-up, the factors contributing to the decrease in Raman signal intensity with increasing sample temperature are laser sample heating, density changes, population changes, and changes in refractive index and reflectivity. They emphasize the importance of minimizing the intensity of the excitation radiation.

Recently Lewis et al.[53,54] have made a major advance in developing Raman spectroscopy as an automated analytical technique for identification of explosives. They recorded the FTRS spectrum of 32 nitro-containing explosive materials and have obtained a classification scheme based upon the symmetric and asymmetric NO_2 stretching vibration transitions, which is graphically presented in Figure 7. By the use of these two frequencies, it is possible to at least identify a cluster of explosive materials that will include the analyte. Using this classification scheme as a foundation, they have studied various means of computer-based, unsupervised pattern recognition involving neural networks, fuzzy logic, and statistical models. The use of computer-based data analysis will significantly reduce the time required for characterization of a sample and will ultimately permit the development of a commercial Raman system for explosives analysis that does not require a highly trained Raman spectroscopist for rapid and successful analysis.

Kneipp et al.[55] have reported NIR surface-enhanced Raman scattering (SERS) of TNT on colloidal gold and silver. Addition of TNT to colloidal silver produced a strong coagulation with the resulting particles too big to produce a large signal enhancement. In contrast to silver, the SERS spectrum of TNT absorbed on gold exhibited significant signal enhancement, enabling concentrations down to 10^{-7} M TNT (about 1 pg of TNT) to be detected. Moreover, the SERS spectrum of TNT on gold displayed many Raman bands strongly related to the spectrum of "free" TNT molecules, while TNT on silver displayed only a single Raman transition. This work demonstrates

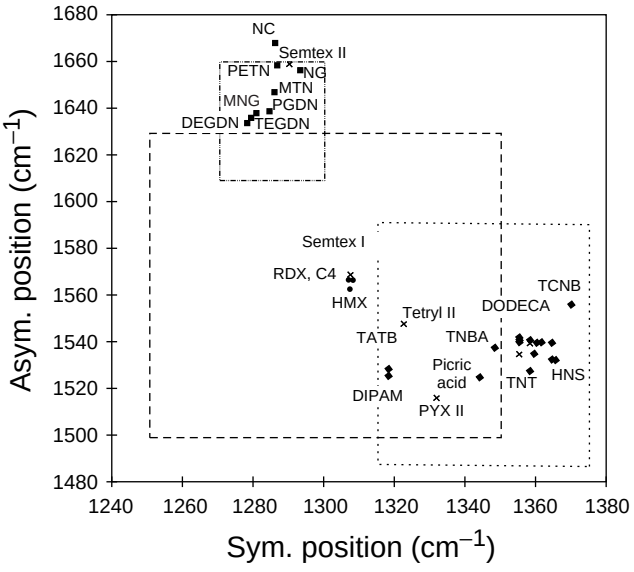

Figure 7 Plot of frequency of asymmetric NO_2 stretching transition vs frequency of symmetric NO_2 stretching transition for nitro-containing explosives. Dashed lines indicate the range of each corresponding explosive class: nitro-aromatic ( ), nitrate ester (■), and nitramine (●). The classes are nitro-aromatic (· · · ·), nitramine (- - - -), nitrate ester (-·-·-·-·-), and exceptional (x) explosives. This classification scheme was devised by Lewis et al.[53] (Reproduced by permission of Society for Applied Spectroscopy.)

the feasibility of using SERS for ultratrace analysis of explosive materials.

McNesby and Coffey[56] have recently demonstrated the use of Raman spectroscopy for determination of the impact sensitivities of explosives (such as HMX, PETN, RDX, TNT, and 1,3,5-triamino-2,4,6-trinitrobenzene (TATB)).

4.3 Photoacoustic Spectroscopy

PAS, which is also referred to as optoacoustic spectroscopy, is performed by irradiating a sample cell with a tunable pulsed-light source, typically a laser, and monitoring the resulting pressure changes using a microphone.[57,58] If the wavelength of the tunable light source is scanned, a PAS spectrum is obtained that is analogous to an absorption spectrum. A schematic of a typical PAS system is shown in Figure 8.

Using CO and CO_2 lasers in the 6-, 9-, and 11-μm regions, Claspy et al.[59] have reported PAS detection of vapors of NG, EGDN, and DNT. Their detection limits are presented in Table 5. In addition to investigating the optoacoustic spectroscopy of these explosives, they also investigated the PAS of possible spectral interferents (NO, NO_2, CH_4, C_4H_{10}, and H_2O). In the 6-μm region, water vapor, NO_2, and NO give rise to spectral interferences of varying degrees of severity, depending on

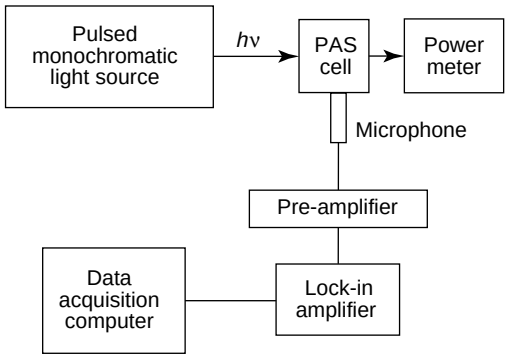

Figure 8 Schematic of PAS apparatus.

Table 5 Laser optoacoustic spectroscopy minimum detectable concentrations (in ppm) for explosive vapors as reported by Claspy et al.[59] (Reproduced by permission of Optical Society of America)

Wavelength region (μm)	Explosive	Minimum detected concentration (ppm vapor)
6	NG	0.024
	EGDN	2.5
	DNT	220
9	NG	0.00055
	EGDN	2.5
	DNT	16
11	NG	0.00028
	EGDN	1.5

the explosive involved; of these, water vapor presented the most difficulty. The water was present in the air and was also entrained in the NG and EGDN explosives. In the 9- and 11-μm regions, none of the possible interferents absorbed laser radiation. Consequently, Claspy et al. recommended that PAS systems designed to detect explosives operate in the 9- to 10-μm or 11- to 12-μm regions.

Crane[18] extended the work of Claspy et al.[59] by measuring the PAS absorption coefficients of EGDN, NG, and DNT for about 77 CO_2 laser transitions in the 9.6-μm and 10.6-μm regions. The minimum detectable concentrations reported are significantly lower than those reported by Claspy et al.[59] (Table 5): 8.26, 0.23, and 0.50 ppb by weight for EGDN, NG, and DNT respectively.[18] Crane noted that the tendency of explosives to adhere to metal surfaces (such as the microphone) made determination of detection limits difficult. In contrast to the report by Claspy et al., Crane observed no detectable saturation with laser power at power levels 40 times those used by Claspy et al. Since Claspy et al. did not provide a sufficiently detailed description of their system, Crane was unable to offer an explanation for this discrepancy.

Williamson et al.[60] have used ultraviolet (UV) laser radiation near 226 nm and obtained PAS LODs of 1.2, 2.8, and 4.9 ppm for detection of NO in Ar, N_2, and air respectively. Since NO is generated by the decomposition of many explosive materials, it may be possible to use detection of NO as a screening method for detection of explosive materials.

Recently Hasue et al.[61] have reported the PAS spectrum of 18 energetic material powders in the 400- to 1600-nm region. Tunable radiation in this spectral region was obtained by using a monochromator as a tunable bandpass filter for a 500-W xenon lamp in the 400 to 800-nm region and a 300-W halogen lamp in the 800 to 1600-nm region. As with the IR laser work described in section 4.1, pulsed radiation was achieved by employing an optical chopper to alternatively transmit and block the light. Figure 9 presents the PAS spectrum of TNT. Hasue et al. found that, in general, the energetic materials studied exhibit PAS peaks in the 600 to 800-nm and 1400 to 1600-nm regions. Furthermore, they found that the energy required to initiate explosives using a ruby laser (694 nm) was correlated with the PAS signal intensities. Hasue et al. did not report detection limits, but they did report PAS signal intensities from which relative detection efficiencies can be inferred.

The current status of PAS detection of explosives is that the feasibility of detecting explosives at low concentrations has been established. Using vapor-phase samples, it is possible to achieve sub-ppb detection. However to date, it has not been demonstrated that different explosives can be distinguished using PAS because at the resolution reported, the spectra are not sufficiently distinct from other explosive materials or from other organic compounds. Moreover, vapor-phase PAS detection of explosives is subject to the experimental difficulties that elevated temperatures are required to obtain appreciable vapor and the pronounced tendency of explosives to adhere to surfaces, especially metal surfaces; the latter places restrictions on the materials used in the apparatus and also means that care must be taken to avoid memory effects.

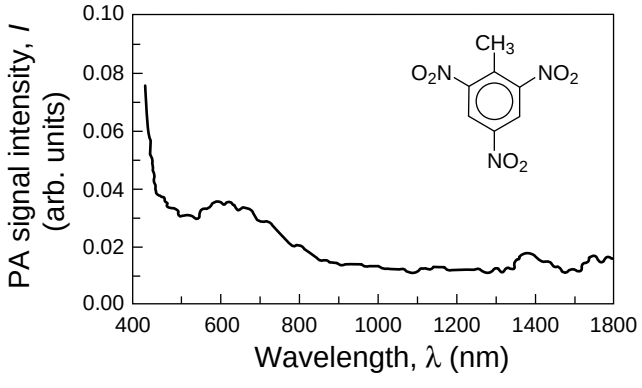

Figure 9 Photoacoustic spectrum of TNT powder recorded by Hasue et al.[61] (Reproduced by permission of Wiley-VCH, Verlag, GmbH.)

4.4 Ultraviolet/Visible Absorption Spectroscopy

The UV spectrum of nitro compounds is dominated by an $n \rightarrow \pi^*$ electronic transition primarily associated with the nitro group. Because of their low vapor pressures and pronounced tendency to decompose at elevated temperatures, the available UV spectra of nitro compounds have primarily been recorded either in solution or in the solid phase; gas-phase spectra are rare.[62,63] The UV electronic transitions of nitro compounds appear as broad, structureless transitions; this is due in part to the broadening induced by the medium and in part by a tendency toward photofragmentation (PF) (and hence broadening due to the Heisenberg Uncertainty Principle). Aliphatic nitro compounds exhibit very low intensity absorption bands in the 260- to 270-nm region.[11,64] Aromatic nitro compounds reveal sharp, intense absorption bands in the 210- to 280-nm region.[65–67] Owing to the influence of a large number of nitro groups on the aromatic ring, the absorption maximum can be shifted and the intensity altered. For the case of nitric esters, an extremely weak absorption band at about 265 nm arises because of the nitroxy group.[66,68] Nitramines in solution exhibit a broad UV absorption between 225 and 240 nm.[67–71]

Changing the solvent changes the position of the maximum and may affect the intensity. Changing from polar to nonpolar solvents causes a shift to shorter wavelengths for aromatic nitro compounds[72] and a shift to longer wavelengths for aliphatic nitro compounds.[73]

Schroeder et al.[66] have published the UV/VIS absorption spectra of 135 explosive or explosive-related compounds in ethanol solution; their paper includes the molecular absorption coefficients of peak maxima. Figure 10 presents the UV absorption spectra of selected nitramines and nitroalkylbenzenes in ethanol solution. Recently solid-state UV/VIS spectra of such explosive compounds as PETN, RDX, HMX, and 2,2′,4,4′,6,6′-hexanitrostilbene (HNS) have been published.[71,74] In comparison with spectra recorded in acetonitrile solution, the solid-state spectra generally have absorption maxima shifted to longer wavelengths.

Because the UV absorptions of nitro compounds in solution and in the solid phase are broad, weak, and generally structureless, UV/VIS absorption spectroscopy cannot provide definitive identification of specific nitro compounds without chemical separation. However, if a separation technique (such as high-performance liquid chromatography (HPLC)) is used, minimum detection limits of 0.03 to 0.21 $\mu g\,L^{-1}$ can be achieved.[67]

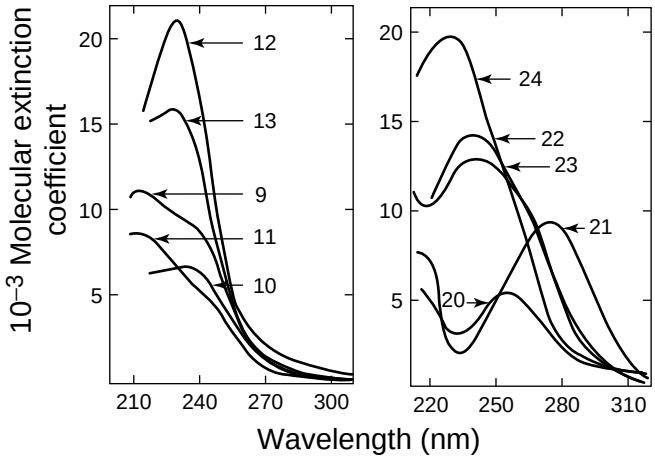

Figure 10 UV absorption spectra of selected explosive compounds in ethanol solution:[66] 9, RDX; 10, TAX; 11, cyclonite oxide; 12, HMX; 13, 20, 2-nitrotoluene; 21, 4-nitrotoluene; 22, 2,4-DNT; 23, 2,4-dinitro-1-ethylbenzene; and 24, TNT. (Reproduced by permission of the American Chemical Society.)

Because of the very low vapor pressures of the common explosive materials, application of UV/VIS absorption spectroscopy to vapor-phase samples is not practical using current commercial UV/VIS spectrometer systems; however, it may be feasible using cavity ringdown spectroscopy.[75–77] (See also **Cavity Ringdown Laser Absorption Spectroscopy** in this encyclopedia.)

4.5 Laser-induced Fluorescence

LIF[78–81] is simply the optical emission from molecules that have been excited to higher energy levels by absorption of laser radiation. The main advantage of using fluorescence detection compared to absorption measurements is the greater sensitivity achieved because the fluorescence signal has a very low background. LIF has been used to study the electronic structure of molecules and to make quantitative measurements of concentrations. Analytical applications of LIF include monitoring gas-phase concentrations in the atmosphere, flames, and plasmas; and remote sensing using light detection.

LIF for the detection of explosives is usually combined with PF in order to photodissociate a target molecule and subsequently detect the fluorescence of the generated photofragments.[82–85] This is done because the fluorescence of most explosives is weak while the fluorescence intensity of the nitric oxide (NO) produced by PF is strong enabling low concentrations of explosives to be detected. This method has the advantage of being very sensitive and is easily applied for in situ measurements.

One example of this technique is the detection of TNT.[83–85] Figure 11 shows a partial energy level diagram of NO and NO_2 and illustrates the PF/LIF

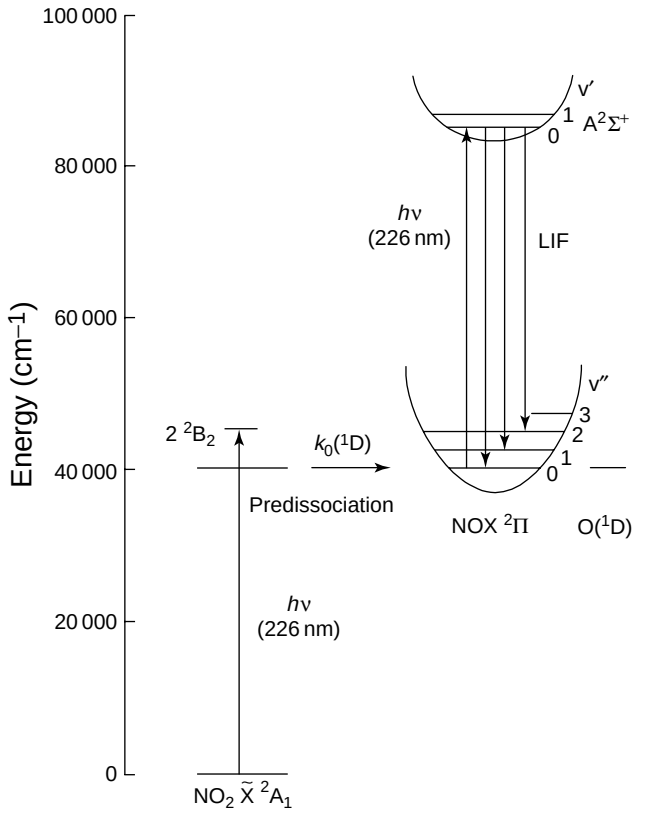

Figure 11 Schematic of energy level diagram illustrating PF/LIF process.

process. The process begins with photofragmenting TNT to release a weakly bound NO_2 functional group (40–50 kcal mol^{-1}).[6–9,62] This NO_2 functional group is responsible for the weak and structureless UV absorption of TNT near 226 nm. The 226-nm laser radiation photodissociates TNT and produces NO_2 in the ground electronic state ($X\ ^2A_1$). The NO_2 generated then absorbs a 226-nm photon and undergoes a transition to the $2\ ^2B_2$ electronic state. The $2\ ^2B_2$ state predissociates rapidly to produce NO in the ground vibrational state of the ground electronic state ($X\ ^2\Pi$, $v'' = 0$) and atomic oxygen. The ground state NO then absorbs another 226-nm photon to undergo a resonant transition ($A\ ^2\Sigma^+ v' = 0 \leftarrow X\ ^2\Pi\ v'' = 0$) which produces NO ($A\ ^2E^+ v' = 0 \rightarrow X\ ^2\Pi\ v'' = n$, where $n = 0, 1, 2, 3, \ldots$) fluorescence. PF and predissociation are achieved with the same UV laser frequency. This allows for a simple experimental set-up for detecting nitro compounds. The intensity of the NO fluorescence is related to the concentration of TNT in the sample.

Figure 12 shows a schematic of a PF/LIF experimental system used to detect TNT.[83–85] A UV laser power of 2 μJ at 226 nm is sufficient for the PF and to induce fluorescence. A photomultiplier tube (PMT) detects the

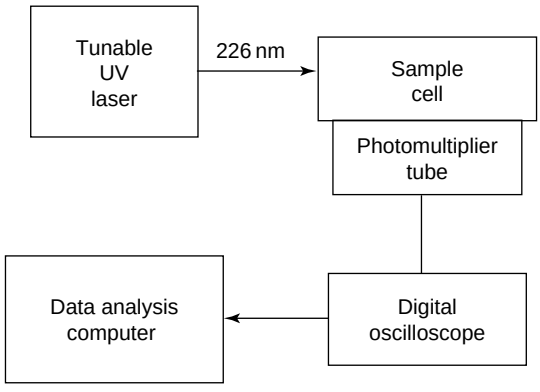

Figure 12 Schematic of PF/LIF apparatus.[84,85]

PF/LIF signal that can be averaged and recorded on a digital oscilloscope. The dye laser wavelength can also be scanned to collect PF/LIF spectra. These spectra can be used to identify the wavelength of the highest intensity LIF signal.

This application of using PF/LIF to detect TNT has also included studies of effects on the PF/LIF signal. Increasing the sample temperature increases the PF/LIF signal intensity, but also produces physical and chemical changes to the sample.[83] This is also true for reheating the samples and increasing the heating time. Increasing sample concentration and laser power also increases the intensity of the PF/LIF signals.[84,85] The PF/LIF signals show a large intensity increase and then a long decay. Analysis of the decay identifies characteristic decay lifetimes and further characterizes the PF/LIF signals. Studies have also been performed to identify the effects of water, fertilizer, and other common soil additives.[85,86] Water quenches the fluorescence and must be boiled or removed from the sample. Fertilizer shows similar signals to small concentrations of TNT, but has little effect when added to TNT samples, at least at the concentrations studied to date.

A common error in measurements with LIF is the variation in collisional quenching. This process competes with fluorescence as a method for excited-state molecules to become de-excited. The rate of collisional de-excitation determines the fluorescence yield (fraction of fluorescing molecules/atoms to the total number excited) and is dependent on the mixture of molecular/atomic species, and their densities and temperature. Fluorescence from other species can also compete and lead to errors in measurements. Modern lasers provide sufficient power to saturate transitions; this results in the fluorescence no longer being linearly related to the laser energy. Under these saturated conditions, the fluorescence becomes insensitive to quenching variations (since stimulated emission dominates quenching as a loss mechanism), and thus some techniques employ this method.[86]

The technique of LIF has been applied to several nitro compounds, including TNT,[83–85] nitrobenzene,[87] and dinitrobenzene.[88] The sensitivity of PF/LIF is very dependent upon the spectroscopic characteristics of the target molecule, the detected fragment and the laser system used for the measurement. The reported LODs for nitro compounds are presented in Table 6. Since the technique has a strong dependence on the laser radiation characteristics, the applications and sensitivity of PF/LIF will increase with improvement of laser systems.

In addition to PF/LIF, other fluorescence-based techniques have been used to detect explosive materials. Jian and Seitz[89] have reported optical detection of nitro compounds via fluorescence quenching. When aqueous nitro compounds were passed through a fluorophore-containing membrane, fluorophore fluorescent intensity was reduced. The response is generic to nitro compounds and cannot be used to identify which compound is involved. Nitrogen purging prior to measurement eliminated interference from oxygen and enhanced sensitivity. LODs are $\sim 2\,\mathrm{mg\,L^{-1}}$ for DNT and TNT and $10\,\mathrm{mg\,L^{-1}}$ for RDX; it is desirable to have even lower LODs. When combined with fiber optic excitation and detection, the membrane can be used for remote in situ screening of groundwater for contamination by explosive materials. Crowson et al.[90] have investigated detection of explosives by gas-phase chemiluminescence. They reported that when vapors from a variety of explosive compounds were thermally decomposed, a blue-green chemiluminescent emission was observed. Simple alkyl nitrates are more thermally stable and hence can be detected at lower concentrations than RDX or PETN. No signal was observed for TNT. For pure vapors, the LODs approached the limit of accurate pressure measurements. The use of carrier gases significantly decreased the achievable LODs: significant improvement occurred with nitrogen and even greater improvement with helium. The authors believed that in the absence of carrier gases, the explosives tended to either be retained within or decomposed within the heated injector. The LOD under static conditions for isopropyl nitrate was 0.013 Pa in 1.3×10^4 Pa of helium (1 ppm). Like the membrane fluorescence quenching technique described above, the chemiluminescence technique does not identify which explosive material is present, but can be used for screening.

Table 6 PF/LIF spectroscopy LODs for nitro compounds

Compound	LOD (ppb by weight)	Ref.
Nitrobenzene	500	88
Dinitrobenzene	11	89
TNT	40	90

4.6 Ionization-based Detection Techniques

4.6.1 Ion Mobility Spectrometry

Mass spectrometry (MS) is a very powerful and sensitive technique for detection and characterization of a wide array of species. IMS[91–93] is a modified form of time-of-flight (TOF) MS which operates at atmospheric pressure. IMS is neither a laser- nor an optical-based detection technique, but it is included for completeness. The principle of operation of IMS can easily be understood from Figure 13. An IMS analyzer consists of a sample ionizer, reaction region, shutter grid, drift region, aperture grid, and a signal collector. The vapor sample is introduced into the IMS cell with the help of a carrier gas, such as air, CO_2, argon, helium, or nitrogen. The sample constituents are ionized along with the carrier gas. Ionization is usually accomplished by bombarding the gas with electrons produced by the radioactive decay of ^{63}Ni, but ionization can also be accomplished by a laser or by an UV/vacuum UV light source. The ionized species travel across the reaction region under the influence of an electric field and reach the shutter grid. The shutter grid consists of sets of thin wires with a voltage bias between them, establishing a localized electric field perpendicular to the direction of flow. This electric field controls entrance of ions into the drift region. When the bias voltage is applied to the shutter grid, the ions are attracted to the grid and lose their charge. When the voltage across the shutter grid is momentarily turned off, ions are transmitted into the drift region.

In the drift region, ions with different mass, charge, and shape travel with different velocities for a given applied constant electric field. The motion of ions through the drift region is hindered by a countercurrent flow of drift gas. The drift velocity of ions defines the ion mobility. Thus the mobility of different ions are different. Ions with smaller mass travel with higher velocity than those with larger mass and reach the collector in shorter time. After the drift region, the ions pass through an aperture grid which prevents an electrical signal being induced on the collector before the ion clouds reach the collector. The ions accumulated on the collector produce an electrical signal that is measured by an electrometer. The different mass ion clouds arrive at the collector at different times. The resulting plot of collector current versus time is referred to as the mass spectrum of the sample. This spectrum is recorded by a data acquisition and analysis system.

The arrival time at the collector depends on the electric field strength in the drift region. Typically ions are introduced into the drift region by the shutter grid every 20–40 ms. The shutter grid is open for typically about 0–0.4 ms. The shutter grid is used to synchronize the signal collection of IMS spectra. To improve S/N, several hundred scans may be averaged. The IMS spectra of both reactants and products are recorded. The recorded spectrum is compared with the library of reference data available on different ions. These data are used to identify the species. The signal intensity at a particular time corresponds to the concentration of the associated species.

As an analytical technique, IMS has several advantages.[91] First for many compounds, the detection limits are in the picogram or sub-ppb range. Second, IMS has selectivity based on ion mass, electric field strength, and electric field polarity; these allow specificity of

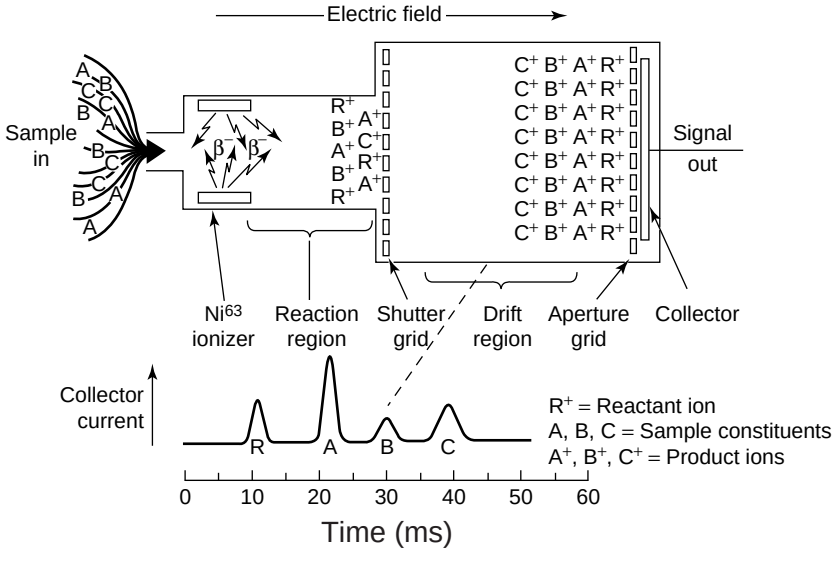

Figure 13 Schematic of IMS apparatus.[95] (Reproduced by permission from *Sensors* (www.sensorsmag.com), April, 1994, Helmers Publishing, Inc., Copyright 1994.)

detection even in the presence of matrix interferents. Third, the system can be easily automated to perform suitable sampling, calibration, and data analysis. Fourth, IMS analysis is rapid, typically under a minute for data acquisition and processing. All techniques have limits/weaknesses; those of IMS include (1) the response varies with background gas composition (consequently peak intensity alone is not sufficient for quantitative analysis), and (2) for some compounds, IMS has a limited linear range (in some cases as small as an order of magnitude[92] although some compounds have linear ranges of up to five orders of magnitude).

IMS detection of explosive materials has been reported since the early 1970s.[92–95] In 1978, Spangler and Lawless[97] investigated the processes of negative ion formation in nitrotoluene compounds. They reported a practical detection limit of 10 ppt (mol mol^{-1}) for TNT in air. In general, negative ionization provides a high degree of selectivity because only highly electronegative compounds, such as explosives, are ionized. Addition of Cl^-, Br^- or I^- ions leads to an enhanced ability to detect NO_3^- from fragmentation of explosive materials.[98] Qualitative identification of explosives by IMS has been reported by various researchers with a variety of very low detection limits. According to Cohen et al.,[99] rapid explosive vapor detection by IMS is feasible down to 10^{-14} parts by volume in air; other researchers report that laboratory detection limit is as low as 200 pg for common explosives (such as NG, PETN, RDX, and TNT)[16] and as low as 0.01 ppb or 1.0 pg for TNT vapor.[100] IMS quantitative detection of TNT has a linear response between 300 and 5000 pg of TNT with a precision of 0.2–1.0%.[101] Thus the concentrations at which explosive materials can be identified by IMS are significantly lower than concentrations for which quantitative determinations can be made. Recently Eiceman et al.[24] have compared the results of an IMS with those of the NIST gravimetric method for TNT vapor from a vapor generator at 79 °C and obtained exact agreement within experimental uncertainty.

The ability of IMS to detect explosives in practical situations has also been investigated. NG, PETN, and RDX have been detected by IMS in post-blast analysis of the remains of pipe bombs, steel plates, suitcases, and cassette/radios.[16] IMS analysis of post-blast residues requires that the sample be dissolved in a solvent for injection into the IMS spectrometer. Solvation of the sample permits purification of the compounds by chemical and/or physical methods, enabling IMS detection of sub-nanogram quantities of explosives in post-blast residues.[102] Handling of commercial[14] and military[15] explosives leaves residuals on hands that can subsequently be transferred to other surfaces. IMS has detected RDX on hands and various car parts following handling of the explosive;[16] eight consecutive hand washings with soap and water were required before IMS could no longer detect the RDX.

TNT vapors can easily be detected by IMS, but the vapors of RDX and of PETN cannot be readily detected without the use of a preconcentrator.[103] The vapor pressure of RDX is about four orders of magnitude less than that of TNT, making detection of RDX more difficult than TNT. Field samples are typically collected by vacuuming a pre-selected surface area. IMS detection of explosives using laser desorption was demonstrated in 1987.[104]

In order to be a viable field instrument, IMS must have an automated data analysis system capable of determining whether or not a peak has been detected. Currently research efforts are underway to combine IMS with real-time data processing using neural network techniques.[106]

4.6.2 Resonance-enhanced Multiphoton Ionization Spectroscopy

MS combined with various preparation techniques has been very extensively utilized for identification and characterization of explosive materials.[1,2,106] A review of the extensive MS literature on detection of explosives is beyond the scope of this article. We will instead restrict ourselves to laser- and optical-based ionization techniques that have been combined with mass spectrometric detection, specifically REMPI spectrometry,[82,107–110] which is sometimes also referred to as resonant ionization mass spectrometry (RIMS). Generally RIMS refers to detection of atoms and REMPI to detection of molecules. A related technique, photofragmentation/photoionization (PF/PI) spectroscopy, uses charged plates to collect the ion signal rather than a TOF mass spectrometer as REMPI does. For the purposes of this review article, we will group REMPI, RIMS, and PF/PI together.

A schematic of a typical REMPI apparatus is presented in Figure 14. A gas-phase sample is introduced into the apparatus, typically in the form of a molecular beam. A high-power, pulsed laser is used to induce multiphoton absorption, resulting in fragmentation and ionization. The resulting ions are detected in a manner similar to that for IMS. With regard to detection, the primary difference between IMS and REMPI is that an IMS experiment uses a carrier gas and a REMPI experiment is performed under vacuum. As with IMS, entry of the ions into the drift region is controlled by grids at high potential, and the mass of the ions is related to the time the ions require to move from the ion formation region to the detector. Again the plot of detector signal vs time is referred to as a mass spectrum.

Marshall et al.[111] have demonstrated that PF of nitroaromatic compounds occurs by predissociation,

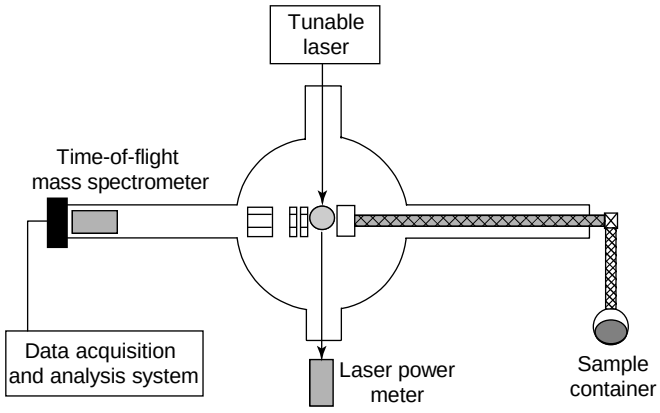

Figure 14 Schematic of REMPI apparatus.[110] (Reproduced by permission of Elsevier Science.)

releasing an NO_2 fragment, which subsequently predissociates to form NO. The NO generated can subsequently undergo multiphoton ionization to produce NO^+, which appears prominently in the mass spectrum. Different nitroaromatics can be distinguished from each other by the wavelength dependence of the NO^+ ion formation,[112] provided that enough sample is available. Moreover, NO^+ ion formed from NO or NO_2 can be distinguished from NO^+ formed from nitroaromatic compounds in vacuum by the wavelength dependence of the NO^+ ion formation.[112]

Multiphoton fragmentation and ionization can occur by two distinct mechanisms.[107–110] First, molecules absorb a number of photons, exciting them to a dissociative state below the ionization limit. If the laser pulse duration is much shorter than the lifetime of the dissociative state, then the up pumping rate may be so high that ionization occurs before dissociation. The ions may absorb additional photons before fragmenting. This process is ionization followed by dissociation (ID) or ladder climbing. However, if the laser pulse duration is longer than the dissociative state lifetime, then the molecule dissociates forming neutral fragments. If the laser intensity is sufficiently high, these fragments may absorb additional photons to further fragment or to ionize. This process is dissociation followed by ionization (DI) or ladder switching. For analytical purposes, DI fragmentation is not very useful because neither parent nor structurally distinctive ions are formed. Thermally labile molecules, which includes many nitro compounds, fragment primarily by the DI route while many molecules, such as aromatics, fragment by the ID route alone. Clark et al.[113,114] have reported REMPI detection of C, H, and O atoms following photodissociation of nitrobenzene. Recently Ledingham et al.[115] have shown that use of femtosecond lasers enhances the ability to identify the parent ion by promoting the ID route. This is important since identification of parent species can be difficult if the original molecule is extensively fragmented by the excitation process or if only a few fragment moieties are ionized. For example, only the NO^+ ion is present in the REMPI mass spectrum of nitrate esters using nanosecond excitation.[116] Among the nitro compounds studied to date by femtosecond REMPI are NO_2,[117] nitromethane,[118] nitrobenzene,[119] and nitrotoluene isomers.[119]

Table 7 Sensitivity levels for nitro compounds detected by REMPI spectroscopy using nanosecond excitation

Compound	Equivalent concentration	Ref.
NO	80 ppt	126
NO_2	15–75 ppb	123
NB	1.3 ppb	116
DNT	5.3 ppb	116
TNT	0.5 ppb	116
PETN	65 ppt	116
RDX	65 ppt	116

Quantitative REMPI determinations of nitro compounds have been reported in a variety of papers.[82,112,116,120–125] The lowest detection limits are presented in Table 7. Several researchers have noted that the detection limits are wavelength dependent.[112,120,123,126] In general, the shorter the excitation wavelength, the lower the detection limit. Recently Elwood and Simeonsson[126] have demonstrated that substantially lower LOD for NO can be achieved using 215 nm + 355 nm excitation rather than with single-color excitation.

As a detection technique for explosives, REMPI has the advantage of having very low detection limits, typically 10–100 pg with nanosecond excitation. To date, REMPI analysis of explosive materials has only occurred in academic rather than forensic laboratories.

5 ENVIRONMENTAL APPLICATIONS

Explosives are dangerous not only owing to their explosive nature, but because they also present a potential health hazard since they are toxic.[5] There are contaminated sites near military test ranges and ammunition plants all over the world. Unexploded or partially exploded ordinance buried in the ground slowly releases the explosive into the soil. These explosive compounds slowly migrate deeper and deeper and, in time, reach groundwater level. These toxic compounds can subsequently be ingested by people from drinking water. Contamination is difficult to clean up once the

toxic substance has reached groundwater level. Thus, detection of explosive materials in soil before it reaches groundwater is important.

Various techniques for monitoring explosives in soil and groundwater have been evaluated. A mobile cone penetrometer system (Site Characterization and Analysis Penetrometer System (SCAPS))[39,127] that can position sensors at selected distances below the surface in order to measure the concentration of explosives or other pollutants (such as petroleum, oils and lubricants (POLs), volatile organic compounds, metals, and radioactive contaminants) as a function of depth has been developed and field tested. In general, the sensors provide rapid screening information – for example, a sensor might quickly indicate that explosives are present and what their concentration is, but might not indicate which explosive is involved. A precise chemical analysis can then be performed by sending a sample to an analytical laboratory. By combining concentration vs depth information for a variety of different surface locations with high-power computers, three-dimensional maps of subsurface pollution distributions can be obtained. These three-dimensional maps are critical to insure that all contaminated areas are cleaned up while effort and funds are not expended on uncontaminated areas.

The SCAPS was initially developed during the late 1980s. Figure 15 shows a typical cone penetrometer system for environmental site characterization and monitoring.[39,127] The SCAPS system includes a custom-engineered, 20-ton truck capable of hydraulically pushing an instrumented probe down to depths of up to 50 m. The truck is separated into two parts: one contains the penetrometer push/pull system, and the other contains most of the measurement equipment. Optical, electrochemical, and chemical cone penetrometer sensors have been developed and evaluated. Laser-based sensors are being developed in the laboratory which can be incorporated into a SCAPS probe.

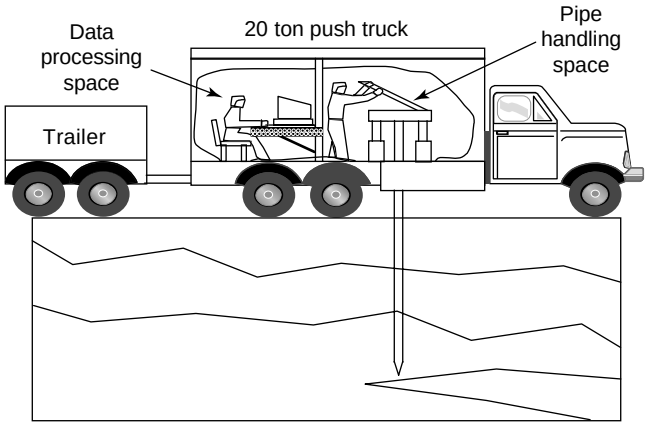

Figure 15 Schematic of SCAPS.[39]

Previously in this article, we have described the detection of explosive pyrolysis products from soil detected by TILDAS[38] (Figure 2 and Table 3) and the detection of TNT in soil by PF/LIF spectroscopy[84–86] (Figure 12 and Table 6). In the following paragraphs, some of the applications of SCAPS for explosive detection for environmental application along with analytical methods will be discussed. Traditional methods are also presented for the sake of completeness.

Jenkins et al.[128] have summarized the analytical methods currently used for characterization of explosive contaminated sites at US Army installations. These include on-site field screening and off-site laboratory analysis. The laboratory techniques are those in general use in forensic laboratories.[1,2] In the laboratory, explosives are extracted from soil and water samples prior to analysis. In general, Soxhlet extraction and ultrasonic methods recovered similar amounts of explosives from samples. They recovered more than mechanical shaking does, particularly when analyte concentration is low. Analysis of water samples with low explosives concentration is often performed after pre-concentration. This is achieved by solvent extraction or solid-phase extraction. The high polarity of the aromatics limits the types of solid phases that can be used. Reversed-phase HPLC is used most of the time to analyze the explosives after extraction and pre-concentration.

In an accompanying paper, Jenkins et al.[129] have reviewed colorimetric methods used for field screening. The field screening results were compared with standard laboratory techniques and were found to be in reasonably good agreement. In colorimetry, explosive materials react with appropriate chemicals and the mixture undergoes a color change. The intensity of the color (as measured by absorption spectroscopy) gives the concentration of the explosive. The LODs of different colorimetric reagents for various explosives are typically in the range 0.1 to 0.3 absorptivity units $mg^{-1}L^{-1}$ for polynitroaromatics and 0.004 to 0.16 absorptivity units $mg^{-1}L^{-1}$ for nitramines and nitrate esters.

Levsen et al.[130] have described the detection of explosives and related compounds in aqueous samples from former ammunition production sites in Germany. They have presented the analysis of the sample using hydrogen nuclear magnetic resonance spectroscopy both for the identification and quantitation of unknowns in wastewater.

Buttner et al.[131] have developed an electrochemical sensor for in situ detection of TNT and other nitrated explosives. TNT, RDX, HMX, and other explosives are rapidly heated with a nichrome wire to 900 °C for a short period (~30 s). The explosive compounds

pyrolyze, producing NO. The NO fragment is detected by amperometry. The NO concentration is correlated with explosives concentration. They installed this sensor in a SCAPS probe and demonstrated its capabilities for field measurements.

Charles et al.[132] have reported their progress on development of a continuous flow, fluorescence-based immunosensor for detection of explosives (TNT and RDX). Antibodies specific to the pollutant of interest are covalently immobilized in a micro column, which is then saturated with a fluorescent analog of the pollutant. When the pollutant of interest is injected into the micro column, it displays the fluorescent analog which is subsequently detected by an on-line fluorometer. LODs for TNT and RDX are 20 ppb. Their sensor was demonstrated on-site using samples collected at two military bases.

6 FUTURE DEVELOPMENT

Detection of explosives is mainly divided into three major areas, all of which have their own difficulties. The first is forensic investigation of crimes; the second is detection to prevent terrorist activity; and the third is detection to address environmental and/or public health concerns. Although several techniques have recently been evaluated for explosives detection as described in the previous sections, further work is still needed to improve them before they will be routinely used with confidence for nonforensic investigations.

With regard to forensic identification of explosive materials, several optical- and laser-based techniques (such as IR and Raman spectroscopy) are now sufficiently developed that they can be routinely used in the better equipped forensic laboratories for screening and/or confirmatory measurements. The main impediment for their further adoption is the need for automation of detection and analysis.

The present airport security check is performed by monitoring the shape and size of explosive packages inside suitcases or by using dogs to detect vapors emitted by explosive materials. These provide the initial information for human evaluation of the suspected bag or site. Based on the success with dogs, vapor-phase detection techniques should be improved to monitor on-line and in real time luggage and packages in airports and other security locations. As mentioned before, low vapor pressure and concealing explosive materials in packages reduces explosive vapors to very low concentrations for detection. The sensitivity of vapor detection needs to be further improved. This can be achieved by developing new technologies or improving current ones. Further work is also needed on fast, near real-time pre-concentration techniques that can improve detection sensitivity.

The selectivity of the explosives detection techniques needs further attention. A few techniques being evaluated (such as PF/LIF, PF/PI, REMPI, and IMS) are based on the detection of fragmentation products. Most of the fragmentation detection techniques lack selectivity. A farmer or gardener who works with fertilizer might provide nitro compound vapors, giving a false alarm when traveling though a security area. Additional work is needed to characterize the fragmentation processes from different explosives and other nitro compounds. Detection techniques can be improved by monitoring more than one fragment species and correlating the fragments with specific explosive materials. Thus work is also needed to enhance data acquisition and analysis in order to provide information in near real time by correlating the fragment concentrations. This will help in increasing traffic flow at airports and other security places. Improved detection will provide rapid information about the presence of explosive materials so that safety measures can be taken in time to avoid accidents or terrorist activities.

The third major area of explosives detection is environmental monitoring in the condensed phase. Measuring explosives in soil and groundwater has to be accomplished in the presence of major interferences, because soil is a very complicated medium in which the matrix can vary rapidly and unexpectedly from place to place at monitoring sites. Most of the evaluated explosives detection techniques recently developed for screening suspected sites have major problems with selectivity – while they indicate the presence of a nitro compound (and its concentration), they are usually not specific as to which nitro compound is present. Techniques for monitoring explosive vapors from condensed-phase samples can have even further complications due to interferences from more volatile compounds. The area of vapor detection needs further attention. Further work is needed towards the development of biochemical and laser-based techniques for explosives detection. Both sensitivity and selectivity of the techniques need further studies. Tagging of explosives with higher vapor pressure additives will help with the detection of new explosive packages.[133]

Automation of detection and analysis is essential for developing the instruments described here into systems that can be routinely used in practical situations. Extensive efforts are currently under way to combine signal recognition through the use of neural network techniques with instruments capable of continuous monitoring.

ACKNOWLEDGMENTS

We are grateful for assistance from Ms Fang-Yu Yueh and Ms Paula Jordan-Neely for preparation of this

manuscript. This work was supported by US Department of Defense grant ARO DAA H04-96-1-0137.

ABBREVIATIONS AND ACRONYMS

AN	Ammonium Nitrate
Ar^+	Argon Ion
CCD	Charge-coupled Device
DI	Dissociation Followed by Ionization
DIPAM	3,3′-Diamino-2,2′,4,4′,6,6′-hexanitrophenyl
DNT	2,4-Dinitrotoluene
EGDN	Ethylene Glycol Dinitrate
FTRS	Fourier Transform Raman Spectroscopy
HMX	1,3,5,7-Tetranitro-1,3,5,7-tetrazacyclooctane
HNS	2,2′,4,4′,6,6′-Hexanitrostilbene
HPLC	High-performance Liquid Chromatography
ID	Ionization Followed by Dissociation
IMS	Ion Mobility Spectrometry
IR	Infrared
LIF	Laser-induced Fluorescence
LOD	Limit of Detection
MS	Mass Spectrometry
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
NG	Nitroglycerin
NIR	Near-infrared
NIST	National Institute of Standards and Technology
PAS	Photoacoustic Spectroscopy
PETN	Pentaerythritol Tetranitrate
PF	Photofragmentation
PF/PI	Photofragmentation/Photoionization
PMT	Photomultiplier Tube
POL	Petroleum, Oils and Lubricant
ppb	parts per billion
ppm	parts per million
ppt	parts per trillion
RDX	1,3,5-Trinitro-1,3,5-triazacyclohexane
REMPI	Resonance-enhanced Multiphoton Ionization
RIMS	Resonant Ionization Mass Spectrometry
SCAPS	Site Characterization and Analysis Penetrometer System
SERS	Surface-enhanced Raman Scattering
S/N	Signal-to-noise Ratio
TATB	1,3,5-Triamino-2,4,6-trinitrobenzene
TILDAS	Tunable Infrared Laser Differential Absorption Spectroscopy
TNT	2,4,6-Trinitrotoluene
TOF	Time-of-flight
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Photoacoustic Spectroscopy in Trace Gas Monitoring

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science • Ion Mobility Spectrometry in Forensic Science

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy

Infrared Spectroscopy **(Volume 12)**
Cavity Ringdown Laser Absorption Spectroscopy • Microspectroscopy • Theory of Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry

Raman Spectroscopy **(Volume 15)**
Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. J. Yinon, S. Zitrin, *Modern Methods and Applications in Analysis of Explosives*, John Wiley & Sons, Chichester, 1993.
2. J. Yinon, 'Analysis of Explosives', *CRC Crit. Rev. Anal. Chem.*, **7**(1), 1–35 (1977).
3. J.I. Steinfeld, J. Wormhoudt, 'Explosives Detection: A Challenge for Physical Chemistry', *Annu. Rev. Phys. Chem.*, **49**, 203–232 (1998).
4. J.C. Oxley, 'Explosives Detection: Potential Problems', in *Law Enforcement Technologies: Identification*

Technologies and Traffic Safety, eds. B. Debuisson, G. Harding, SPIE, Bellingham, WA, 217–226, Vol. 2511, 1995.
5. J. Yinon, *Toxicity and Metabolism of Explosives*, CRC Press, Boca Raton, FL, 1990.
6. G.F. Adams, R.W. Shaw, Jr, 'Chemical Reactions in Energetic Materials', *Annu. Rev. Phys. Chem.*, **43**, 311–340 (1992).
7. T.B. Brill, K.J. James, 'Kinetics and Mechanisms of Thermal Decomposition of Nitroaromatic Explosives', *Chem. Rev.*, **93**(8), 2667–2692 (1993).
8. M.H. Alexander, P.J. Dagdigian, M.E. Jacox, C.E. Kolb, C.F. Melius, H. Rabitz, M.D. Smooke, W. Tsang, 'Nitramine Propellant Ignition and Combustion Research', *Prog. Energy Combust. Sci.*, **17**, 263–296 (1991).
9. T.L. Boggs, 'Thermal Behavior of RDX and HMX', in *Fundamentals of Solid-propellant Combustion*, eds. K.K. Kuo, M. Summerfield, American Institute of Astronautics and Aeronautics, New York, 121–175, Vol. 90, 1984.
10. D.S. Anex, J.C. Allman, Y.T. Lee, 'Studies of Initial Dissociation Processes in 1,3,3-Trinitroazetidine by Photofragmentation Translational Spectroscopy', in *Chemistry of Energetic Materials*, eds. G.A. Olah, D.R. Squire, Academic Press, San Diego, CA, 27–54, 1991.
11. T. Urbanski, *Chemistry and Technology of Explosives*, Pergamon Press, New York, Vol. 1, 1964.
12. J. Köhler, R. Meyer, *Explosives*, 4th revised and extended edition, VCH Publishers, Weinheim, 1993.
13. B.C. Dionne, D.P. Rounbehler, E.K. Achter, J.R. Hobbs, D.H. Fine, 'Vapor Pressure of Explosives', *J. Energ. Mater.*, **4**(1–4), 447–472 (1986).
14. J.D. Twibell, J.M. Home, K.W. Smalldon, D.G. Higgs, 'Transfer of Nitroglycerine to Hands During Contact with Commercial Explosives', *J. Forensic Sci.*, **27**(4), 783–791 (1982).
15. J.F.B. Lloyd, R.M. King, 'Detection and Persistence of Military Explosives on the Hands', *J. Forensic Sci.*, **29**(1), 284–290 (1984).
16. D.D. Fetterolf, T.D. Clark, 'Detection of Trace Explosive Evidence by Ion Mobility Spectrometry', *J. Forensic Sci.*, **38**(1), 28–39 (1993).
17. E. Bender, A. Hogan, D. Leggett, G. Miskolczy, S. MacDonald, 'Surface Contamination by TNT', *J. Forensic Sci.*, **37**(6), 1673–1678 (1992).
18. R.A. Crane, 'Laser Optoacoustic Absorption Spectra for Various Explosive Vapors', *Appl. Opt.*, **17**(13), 2097–2102 (1978).
19. P.A. Pella, 'Generator for Producing Trace Vapor Concentrations of 2,4,6-Trinitrotoluene, 2,4-Dinitrotoluene, and Ethylene Glycol Dinitrate for Calibrating Explosives Vapor Detectors', *Anal. Chem.*, **48**(11), 1632–1637 (1976).
20. J.P. Davies, L.G. Blackwood, S.G. Davis, L.D. Goodrich, R.A. Larson, 'Design and Calibration of Pulsed Vapor Generators for 2,4,6-Trinitrotoluene, Cyclo-1,3,5-trimethylene-2,4,6-trinitramine, and Pentaerythritol Tetranitrate', *Anal. Chem.*, **65**(21), 3004–3009 (1993).
21. E.E.A. Bromberg, D. Dussalult, S. MacDonald, W.A. Curby, 'Vapor Generation for Use in Explosive Portal Detection Devices', in *Proceedings of the 4th International Symposium on Analysis and Detection of Explosives*, ed. J. Yinon, 473–484, 1992.
22. D.W. Hannum, 'Characteristics of Selected Glass Tube Preconcentrators Used with an Ion Mobility Spectrometer', Sandia National Laboratory Report SAND89-0242-UC-15, 1989.
23. C.S. Giam, L.W. Chang, G.E. Reed, T.L. Holliday, J.M. Budzik, 'Results from Measurements of a Portable Calibrator for Trace Levels of RDX, PETN, and TNT Using IMS', in *Proceedings of the 4th International Workshop on Ion Mobility Spectrometry*, 1995.
24. G.A. Eiceman, D. Preston, G. Tiano, J. Rodriguez, J.E. Parmeter, 'Quantitative Calibration of Vapor Levels of TNT, RDX, and PETN Using a Diffusion Generator with Gravimetry and Ion Mobility Spectrometer', *Talanta*, **45**(1), 57–74 (1997).
25. L.H.J. Lajunen, *Spectrochemical Analysis by Atomic Absorption and Emission*, Royal Society of Chemistry, London, 1992.
26. F. Pristera, M. Halik, A. Castelli, W. Fredericks, 'Analysis of Explosives Using Infrared Spectroscopy', *Anal. Chem.*, **32**(4), 495–508 (1960).
27. D.E. Chasan, G. Norwitz, 'Qualitative Analysis of Primers, Tracers, Igniters, Incendiaries, Boosters, and Delay Compositions on a Microscale by Use of Infrared Spectroscopy', *Microchem. J.*, **17**(1), 31–60 (1972).
28. R.J. Karpowicz, T.B. Brill, 'Comparison of the Molecular Structure of Hexahydro-1,3,5-trinitro-*s*-triazine in the Vapor, Solution, and Solid Phases', *J. Phys. Chem.*, **88**(3), 348–352 (1984).
29. W.R. Carper, J.J.P. Stewart, 'Effects of Isotopic Substitution on the Vibrational Spectra of 2,4,6-Trinitrotoluene', *Spectrochim. Acta A*, **43**(10), 1249–1255 (1987).
30. J. Janni, B.D. Gilbert, R.W. Field, J.I. Steinfeld, 'Infrared Absorption of Explosive Molecule Vapors', *Spectrochim. Acta A*, **53**(9), 1375–1381 (1997).
31. M. Rey-Lafon, C. Trinquecoste, R. Cavagnat, M.T. Forel, 'Étude des Spectres de Vibration de la Trinitro-1,3,5 hexahydro-*s*-triazine. I. Analyse des Spectres Infrarouges et Raman entre 3200 et 200 cm^{-1}', *J. Chim. Phys. Phys.-Chim. Biol.*, **68**(209), 1533–1542 (1971).
32. M. Rey-Lafon, R. Cavagnat, C. Trinquecoste, M.T. Forel, 'Étude des Spectres de Vibration de la Trinitro-1,3,5 hexahydro-*s*-triazine. II. Analyse des Spectres du Monocristal entre 200 et 10 cm^{-1}', *J. Chim. Phys. Phys.-Chim. Biol.*, **68**(210), 1573–1577 (1971).
33. B.L. Deopura, V.D. Gupta, 'Vibration Spectra of 1,3,5-Triamino-2,4,6-trinitrobenzene', *J. Chem. Phys.*, **54**(9), 4013–4019 (1971).

34. Z. Iqbal, S. Bulusu, J.R. Autera, 'Vibrational Spectra of β-Cyclomethylene Tetranitramine and Some of its Isotopic Isomers', *J. Chem. Phys.*, **60**(1), 221–230 (1974).
35. T.G. Townes, 'Vibrational Spectrum of 1,3,5-Triamino-2,4,6-trinitrobenzene', *Spectrochim. Acta A*, **39**(9), 801–804 (1983).
36. J.J.P. Stewart, S.R. Bosco, W.R. Carper, 'Vibrational Spectra of 2,4,6-Trinitrotoluene and its Isotopically Substituted Analogs', *Spectrochim. Acta A*, **42**(1), 13–21 (1986).
37. A.D. Beveridge, S.F. Payton, R.J. Audette, A.J. Lambertus, R.C. Shaddick, 'Systematic Analysis of Explosive Residues', *J. Forensic Sci.*, **20**(3), 431–454 (1975).
38. J. Wormhoudt, J.H. Shorter, J.B. McManus, P.L. Kebabian, M.S. Zahniser, W.M. Davis, E.R. Cespedes, C.E. Kolb, 'Tunable Infrared Laser Detection of Pyrolysis Products of Explosives in Soils', *Appl. Opt.*, **35**(21), 3992–3997 (1996).
39. P.V. Doskey, E.R. Cespedes, 'Cone-penetrometer Deployed Chemical Sensors', in *Encyclopedia of Environmental Analysis and Remediation*, J. Wiley & Sons, New York, 1284–1312, 1998.
40. J.U. White, 'Long Optical Paths of Large Aperture', *J. Opt. Soc. Am.*, **32**(5), 285–288 (1942).
41. H. Riris, C.B. Carlisle, D.F. McMillen, D.E. Cooper, 'Explosives Detection with a Frequency Modulation Spectrometer', *Appl. Opt.*, **35**(24), 4694–4704 (1996).
42. T.J. Vickers, C.K. Mann, 'Quantitative Analysis by Raman Spectroscopy', in *Analytical Raman Spectroscopy*, eds. J.G. Grasselli, B.J. Bulkin, J. Wiley & Sons, New York, 107–135, 1991.
43. D.A. Long, *Raman Spectroscopy*, McGraw-Hill, New York, 1977.
44. N.F. Fell, J.M. Widder, S.V. Medlin, J.B. Morris, R.A. Pesce-Rodriguez, K.L. McNesby, 'Fourier Transform Raman Spectroscopy of Some Energetic Materials and Propellant Formulations. II', *J. Raman Spectrosc.*, **27**(2), 97–104 (1996).
45. I.R. Lewis, P.R. Griffiths, 'Raman Spectrometry with Fiber-optic Sampling', *Appl. Spectrosc.*, **50**(10), 12A–30A (1996).
46. I.P. Hayward, T.E. Kirkbride, D.N. Batchelder, R.J. Lacey, 'Use of a Fiber Optic Probe for the Determination and Identification of Explosive Materials by Raman Spectroscopy', *J. Forensic Sci.*, **40**(5), 883–884 (1995).
47. F.W.S. Carver, D.P. Wyndham, T.J. Sinclair, 'Spectroscopic Studies of Explosives II-Detection of Nitro Compounds on Silica and Glass with a Raman Microprobe', *J. Raman Spectrosc.*, **16**(5), 332–336 (1985).
48. F.W.S. Carver, T.J. Sinclair, 'Spectroscopic Studies of Explosives. 1-Detection of Nitro Compounds on Silica Gel and Carbon by Non-resonant Raman Spectroscopy', *J. Raman Spectrosc.*, **14**(6), 410–414 (1983).
49. I.R. Lewis, N.W. Daniel, Jr, N.C. Chaffin, P.R. Griffiths, M.W. Tungol, 'Raman Spectroscopic Studies of Explosive Materials: Towards a Fieldable Explosives Detector', *Spectrochim. Acta A*, **51**(12), 1985–2000 (1995).
50. K.L. McNesby, J.E. Wolfe, J.B. Morris, R.A. Pesce-Rodriguez, 'Fourier Transform Raman Spectroscopy of Some Energetic Materials and Propellant Formulations', *J. Raman Spectrosc.*, **25**(1), 75–87 (1994).
51. K.L. McNesby, N.F. Fell, Jr, J.A. Vanderhoff, 'Detection and Characterization of Explosives Using Raman Spectroscopy: Identification, Laser Heating, and Impact Sensitivity', in *Electro-optical Technology for Remote Chemical Detection and Identification II*, eds. M. Fallahi, E. Howden, SPIE, Bellingham, WA, 121–135, Vol. 3082, 1997.
52. C.M. Hodges, J. Akhavan, 'The Use of Fourier Transform Raman Spectroscopy in the Forensic Identification of Illicit Drugs and Explosives', *Spectrochim. Acta A*, **46**(2), 303–307 (1990).
53. I.R. Lewis, N.W. Daniel, Jr, P.R. Griffiths, 'Interpretation of Raman Spectra of Nitro-containing Explosive Materials. Part I: Group Frequency and Structural Class Membership', *Appl. Spectrosc.*, **51**(12), 1854–1867 (1997).
54. N.W. Daniel, Jr, I.R. Lewis, P.R. Griffiths, 'Interpretation of Raman Spectra of Nitro-containing Explosive Materials. Part II: The Implementation of Neural, Fuzzy, and Statistical Models for Unsupervised Pattern Recognition', *Appl. Spectrosc.*, **51**(12), 1868–1879 (1997).
55. K. Kneipp, Y. Wang, R.R. Dasari, M.S. Feld, B.D. Gilbert, J. Janni, J.I. Steinfeld, 'Near-infrared Surface-enhanced Raman Scattering of Trinitrotoluene on Colloidal Gold and Silver', *Spectrochim. Acta A*, **51**(12), 2171–2175 (1995).
56. K.L. McNesby, C.S. Coffey, 'Spectroscopic Determination of Impact Sensitivities of Explosives', *J. Phys. Chem. B*, **101**(16), 3097–3104 (1997).
57. V.P. Zharov, V.S. Letokhov, *Laser Optoacoustic Spectroscopy*, Springer-Verlag, Berlin, 1986.
58. M.W. Sigrist, 'Air Monitoring by Laser Photoacoustic Spectroscopy', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, J. Wiley & Sons, New York, 163–238, 1994.
59. P.C. Claspy, Y.H. Pao, S. Kwong, E. Nodov, 'Laser Optoacoustic Detection of Explosive Vapors', *Appl. Opt.*, **15**(6), 1506–1509 (1976).
60. C.K. Williamson, R.L. Pastel, R.C. Sausa, 'Detection of Ambient NO by Laser-induced Photoacoustic Spectrometry Using $A\,^2\Sigma^+ - X^2\Pi(0, 0)$ Transitions Near 226 nm', *Appl. Spectrosc.*, **50**(2), 205–210 (1996).
61. K. Hasue, S. Nakahara, J. Morimoto, T. Yamagami, Y. Okamoto, T. Miyakawa, 'Photoacoustic Spectroscopy of Energetic Materials', *Propellants, Explosives, Pyrotechnics*, **20**, 187–191 (1995).
62. M.J. McQuaid, R.C. Sausa, 'Absorption Cross-sections of Gaseous Dimethylnitramine at Selected Wavelengths between 185 and 325 nm', *Appl. Spectrosc.*, **45**(5), 916–917 (1991).
63. S. Nagakura, 'Ultra-violet Absorption Spectra and π-electron Structures of Nitromethane and the Nitromethyl Anion', *Mol. Phys.*, **3**(2), 152–162 (1960).

64. P.A. Mullen, M.K. Orloff, 'Ultraviolet Absorption Spectrum of Pentaerythritol Tetranitrate', *J. Phys. Chem.*, **77**(7), 910–911 (1973).
65. C.P. Conduit, 'Ultraviolet and Infrared Spectra of Some Aromatic Nitrocompounds', *J. Chem. Soc. (London)*, (665), 3273–3277 (1959).
66. W.A. Schroeder, P.E. Wilcox, K.N. Trueblood, A.O. Dekker, 'Ultraviolet and Visible Absorption Spectra in Ethyl Alcohol: Data for Certain Nitric Esters, Nitramines, Nitroalkylbenzenes, and Derivatives of Phenol, Aniline, Urea, Carbamic Acid, Diphenylamine, Carbazole, and Triphenylamine', *Anal. Chem.*, **23**(12), 1740–1747 (1951).
67. E.S.P. Bouvier, S.A. Oehrle, 'Analysis and Identification of Nitroaromatic and Nitramine Explosives in Water Using HPLC and Photodiode-array Detection', *LC-GC*, **13**(2), 120–130 (1995).
68. R.N. Jones, G.D. Thorn, 'The Ultraviolet Absorption Spectra of Aliphatic Nitramines, Nitrosamines, and Nitrates', *Can. J. Res. B*, **27**(10), 828–860 (1949).
69. J. Stals, C.G. Barraclough, A.S. Buchanan, 'Molecular Orbital Interpretation of Ultra-violet Absorption Spectra of Unconjugated Aliphatic Nitramines', *Trans. Faraday Soc. (London)*, **65**(4), 904–914 (1969).
70. J. Stals, 'Chemistry of Aliphatic Unconjugated Nitramines. Part 4. Photophysical Processes of Secondary Nitramines', *Trans. Faraday Soc. (London)*, **67**(6), 1739–1748 (1971).
71. P.L. Marinkas, J.E. Mapes, D.S. Downs, P.J. Kemmey, A.C. Forsyth, 'Absorption of Cyclic Polynitramines in the Solid and Solvated States', *Mol. Cryst. Liq. Cryst.*, **35**(3/4), 15–25 (1976).
72. G.S. Hammond, F.J. Modic, 'Aromatic Nitration. I. The Ultraviolet Spectra of Aromatic Nitro Compounds in Sulfuric Acid', *J. Am. Chem. Soc.*, **75**(6), 1385–1388 (1953).
73. N.S. Bayliss, C.J. Brackenridge, 'The Perturbation of the Ultraviolet Spectrum of Nitromethane in Mixed Solvents, with a Note on Complex Formation', *J. Am. Chem. Soc.*, **77**(15), 3959–3963 (1955).
74. K.J. Smit, 'Ultraviolet and Visible Absorption Spectroscopy of Some Energetic Molecules in the Solid State', *J. Energ. Mater.*, **9**(1–2), 81–103 (1991).
75. R.T. Jongma, M.G.H. Boogaarts, I. Holleman, G. Meijer, 'Trace Gas Detection with Cavity Ring Down Spectroscopy', *Rev. Sci. Instrum.*, **60**(4), 2821–2828 (1995).
76. P. Zalicki, R.N. Zare, 'Cavity Ring-down Spectroscopy for Quantitative Absorption Measurements', *J. Chem. Phys.*, **102**(7), 2708–2717 (1995).
77. R. Vasudev, A. Usachev, W.R. Dunsford, 'Detection of Toxic Compounds by Cavity Ring-down Spectroscopy', *Environ. Sci. Technol.*, **33**(11), 1936–1939 (1999).
78. R.N. Zare, P.J. Dagdigian, 'Tunable Laser Fluorescence Method for Product State Analysis', *Science*, **185**(4153), 739–747 (1974).
79. J.L. Kinsey, 'Laser-induced Fluorescence', *Annu. Rev. Phys. Chem.*, **28**, 349–372 (1977).
80. A.C. Eckbreth, *Laser Diagnostics for Combustion Temperature and Species*, 2nd edition, Gordon and Breach, Amsterdam, 1996.
81. W. Demtröder, *Laser Spectroscopy: Basic Concepts and Instrumentation*, 2nd enlarged edition, Springer Verlag, Berlin, 1996.
82. J.B. Simeonsson, R.C. Sausa, 'A Critical Review of Laser Photofragmentation/Fragment Detection Techniques for Gas-phase Chemical Analysis', *Appl. Spectrosc. Rev.*, **31**(1/2), 1–72 (1996).
83. D. Wu, J.P. Singh, F.Y. Yueh, D.L. Monts, '2,4,6-Trinitrotoluene Detection by Laser-photofragmentation–Laser-induced Fluorescence', *Appl. Opt.*, **35**(21), 3998–4003 (1996).
84. G.M. Boudreaux, 'Development of a Photofragmentation/Laser Induced Fluorescence (PF/LIF) Laser Sensor for Detection of 2,4,6-Trinitrotoluene (TNT) in Soil and Groundwater', MS thesis, Mississippi State University, 1997; *Masters Abstr. Int.*, **36**(1), 201–202 (1998).
85. G.M. Boudreaux, T.S. Miller, A.J. Kunefke, J.P. Singh, F.Y. Yueh, D.L. Monts, 'Development of a Photofragmentation/Laser-induced-fluorescence for Detection of 2,4,6-Trinitrotoluene in Soil and Groundwater', *Appl. Opt.*, **38**(9), 1411–1417 (1999).
86. J.W. Daily, 'Saturation Effects in Laser Induced Fluorescence Spectroscopy', *Appl. Opt.*, **16**(3), 568–571 (1977).
87. N. Daugey, J. Shu, I. Bar, S. Rosenwaks, 'Nitrobenzene Detection by One-color Laser-photolysis/Laser-induced Fluorescence of NO ($v'' = 0 - 3$)', *Appl. Spectrosc.*, **53**(1), 57–64 (1999).
88. J. Shu, I. Bar, S. Rosenwaks, 'Dinitrobenzene Detection by One-color Laser Photolysis/Laser Induced Fluorescence of Vibrationally Excited NO', *Appl. Opt.*, **38**(21), 4705–4710 (1999).
89. C. Jian, W.R. Seitz, 'Membrane for In Situ Optical Detection of Organic Nitro Compounds Based on Fluorescence Quenching', *Anal. Chim. Acta*, **237**(2), 265–271 (1990).
90. A. Crowson, R.W. Hiley, T. Ingham, T. McCreedy, A.J. Pilgrim, A. Townshend, 'Investigation Into the Detection of Nitrated Organic Compounds and Explosives by Direct Chemiluminescent Emission During Thermally Induced Gas Phase Decomposition Reactions', *Anal. Commun.*, **34**(8), 213–216 (1997).
91. R.H. St. Louis, H.H. Hill, Jr, 'Ion Mobility Spectrometry in Analytical Chemistry', *CRC Crit. Rev. Anal. Chem.*, **21**(5), 321–355 (1990).
92. G.A. Eiceman, Z. Karpas, *Ion Mobility Spectrometry*, CRC Press, Boca Raton, FL, 153–164, 1994.
93. G.A. Eiceman, 'Advances in Ion Mobility Spectrometry: 1980–1990', *CRC Crit. Rev. Anal. Chem.*, **22**(1,2), 17–36 (1991).
94. J.A. Reategui, 'Ion Mobility Spectrometry: Theory and Applications', *Sensors*, **11**(4), 14–59 (1994).

95. F.W. Karasek, D.W. Denney, 'Detection of 2,4,6-Trinitrotoluene Vapors in Air by Plasma Chromatography', *J. Chromatogr.*, **93**(1), 141–147 (1974).
96. Z. Karpas, 'Forensic Science Applications of Ion Mobility Spectrometry', *Forensic Sci. Rev.*, **1**(2), 104–119 (1989).
97. G.E. Spangler, P.A. Lawless, 'Ionization of Nitrotoluene Compounds in Negative Ion Plasma Chromatography', *Anal. Chem.*, **50**(7), 884–892 (1978).
98. C.J. Proctor, J.F.J. Todd, 'Alternate Reagent Ions for Plasma Chromatography', *Anal. Chem.*, **56**(11), 1794–1797 (1984).
99. M.J. Cohen, R.F. Wernlund, R.M. Stimac, 'The Ion Mobility Spectrometer for High Explosive Vapor Detection', *Nucl. Mater. Manage.*, **13**, 220–225 (1984).
100. G.E. Spangler, J.P. Carrico, D.N. Campbell, 'Recent Advances in Ion Mobility Spectrometry for Explosives Vapor Detection', *J. Test. Eval.*, **13**(3), 234–240 (1985).
101. F. Garofolo, F. Marziali, V. Migliozzi, A. Stama, 'Rapid Quantitative Determination of 2,4,6-Trinitrotoluene by Ion Mobility Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**(11), 1321–1326 (1996).
102. F. Garofolo, V. Migliozzi, B. Roio, 'Application of Ion Mobility Spectrometry to the Identification of Trace Levels of Explosives in the Presence of Complex Matrices', *Rapid Commun. Mass Spectrom.*, **8**(7), 527–532 (1994).
103. F.J. Conrad, 'Explosive Detection–The Problems and Prospects', *Nucl. Mater. Manage.*, **13**, 212–215 (1984).
104. S.D. Huang, L. Kolaitis, D.M. Lubman, 'Detection of Explosives Using Laser Desorption in Ion Mobility Spectrometry/Mass Spectrometry', *Appl. Spectrosc.*, **41**(8), 1371–1376 (1987).
105. A.H. Lawrence, R.A. Goubran, H.M. Hafez, 'Signal Improvement in Ion Mobility Spectrometry', in *Applications of Signal and Image Processing in Explosives Detection Systems*, eds. J.M. Connelly, S.M. Cheung, SPIE, Bellingham, WA, 97–106, Vol. 1824, 1992.
106. J. Yinon, 'Forensic Applications of Mass Spectrometry', *Mass Spectrom. Rev.*, **10**(3), 179–224 (1991).
107. D.M. Lubman, 'Analytical Multiphoton Ionization Mass Spectrometry. Part I. Theory and Instrumentation', *Mass Spectrom. Rev.*, **7**(5), 535–554 (1988).
108. D.M. Lubman, 'Analytical Multiphoton Ionization Mass Spectrometry. Part II. Applications', *Mass Spectrom. Rev.*, **7**(6), 559–592 (1988).
109. K.W.D. Ledingham, 'The Use of Lasers to Detect Strategic and Environmentally Sensitive Materials', *Phys. Scr., T*, **58**, 100–103 (1995).
110. K.W.D. Ledingham, R.P. Singhal, 'High Intensity Laser Mass Spectrometry–A Review', *Int. J. Mass Spectrom. Ion Process.*, **163**(3), 149–168 (1997).
111. A. Marshall, A. Clark, R. Jennings, K.W.D. Ledingham, J. Sander, R.P. Singhal, 'Laser-induced Dissociation, Ionization and Fragmentation Processes in Nitroaromatic Molecules', *Int. J. Mass Spectrom. Ion Process.*, **116**(2), 143–156 (1992).
112. A. Clark, K.W.D. Ledingham, A. Marshall, J. Sander, R.P. Singhal, 'Attomole Detection of Nitroaromatic Vapors Using Resonance Enhanced Multiphoton Ionization Mass Spectrometry', *Analyst*, **118**(6), 601–607 (1993).
113. A. Clark, K.W.D. Ledingham, A. Marshall, R.P. Singhal, 'Resonant Ionization Spectroscopy of Carbon Atoms Following Laser-induced Fragmentation of Nitro-aromatic Molecules', *Spectrochim. Acta B*, **47**(6), 799–808 (1992).
114. A. Clark, C. Kosmidis, K.W.D. Ledingham, A. Marshall, J. Sander, R.P. Singhal, M. Campbell, 'Resonant Ionization of Oxygen and Hydrogen Atoms Following Laser-induced Photodissociation of Nitrobenzene Vapor', *J. Phys. B*, **26**(19), L665–L670 (1993).
115. K.W.D. Ledingham, H.S. Kilic, C. Kosmidis, R.M. Deas, A. Marshall, T. McCanny, R.P. Singhal, A.J. Langley, W. Shaikh, 'A Comparison of Femtosecond and Nanosecond Multiphoton Ionization and Dissociation for Some Nitro-molecules', *Rapid Commun. Mass Spectrom.*, **9**(15), 1522–1527 (1995).
116. A. Marshall, A. Clark, K.W.D. Ledingham, J. Sander, R.P. Singhal, C. Kosmidis, R.M. Deas, 'Detection and Identification of Explosives Compounds Using Laser Ionization Time-of-flight Techniques', *Rapid Commun. Mass Spectrom.*, **8**(7), 521–526 (1994).
117. K.W.D. Ledingham, C. Kosmidis, S. Georgiou, S. Couris, R.P. Singhal, 'A Comparison of the Femto-, Pico- and Nano-second Multiphoton Ionization and Dissociation Processes of NO_2 at 248 and 496 nm', *Chem. Phys. Lett.*, **247**(4/6), 555–563 (1995).
118. H.S. Kilic, K.W.D. Ledingham, C. Kosmidis, T. McCanny, R.P. Singhal, S.L. Wang, D.J. Smith, A.J. Langley, W. Shaikh, 'Multiphoton Ionization and Dissociation of Nitromethane Using Femtosecond Laser Pulses at 375 and 750 nm', *J. Phys. Chem. A*, **101**(5), 817–823 (1997).
119. C. Kosmidis, K.W.D. Ledingham, H.S. Kilic, T. McCanny, R.P. Singhal, A.J. Langley, W. Shaikh, 'On the Fragmentation of Nitrobenzene and Nitrotoluenes Induced by a Femtosecond Laser at 375 nm', *J. Phys. Chem. A*, **101**(12), 2264–2270 (1997).
120. J.B. Simeonsson, G.W. Lemire, R.C. Sausa, 'Trace Detection of Nitrocompounds by ArF Laser Photofragmentation/Ionization Spectrometry', *Appl. Spectrosc.*, **47**(11), 1907–1912 (1993).
121. G.W. Lemire, J.B. Simeonsson, R.C. Sausa, 'Monitoring of Vapor-phase Nitro Compounds Using 226-nm Radiation: Fragmentation with Subsequent NO Resonance-enhanced Multiphoton Ionization Detection', *Anal. Chem.*, **65**(5), 529–533 (1993).
122. J.B. Simeonsson, G.W. Lemire, R.C. Sausa, 'Laser-induced Photofragmentation/Photoionization Spectrometry: A Method for Detecting Ambient Oxides of Nitrogen', *Anal. Chem.*, **66**(14), 2272–2278 (1994).

123. J.B. Simeonsson, R.C. Sausa, 'Trace Analysis of NO_2 in the Presence of NO by Laser Photofragmentation/Fragment Photoionization Spectrometry at Visible Wavelengths', *Appl. Spectrosc.*, **50**(10), 1277–1282 (1996).
124. V. Swayambunthan, R.C. Sausa, unpublished results reported in J.B. Simeonsson, R.C. Sausa, 'Laser Photofragmentation/Fragment Detection Techniques for Chemical Analysis of the Gas Phase', *Trends Anal. Chem.*, **17**(8/9), 542–550 (1998).
125. S.H. Lee, J. Hirokawa, Y. Kajii, H. Akimoto, 'New Method for Measuring Low NO Concentrations Using Laser Induced Two Photon Ionization', *Rev. Sci. Instrum.*, **68**(7), 2891–2897 (1997).
126. S.A. Elwood, J.B. Simeonsson, 'Studies of the Spectral Dependence of Photoionization Spectrometry Measurements of NO in Air', *Appl. Spectrosc.*, **54**(12), 190–196 (2000).
127. W.M. Davis, E.R. Cespedes, L.T. Lee, J.F. Powell, R.A. Goodson, 'Rapid Delineation of Subsurface Petroleum Contamination Using the Site Characterization and Analysis Penetrometer System', *Environ. Geol.*, **29**(3/4), 228–237 (1997).
128. T.F. Jenkins, M.E. Walsh, P.G. Thorne, 'Analytical Methods for Characterization of Explosives-contaminated Sites on US. Army Installations', in *Environmental Monitoring and Hazardous Waste Site Remediation*, eds. T. Vo-Dinh, R. Nießner, SPIE, Bellingham, WA, 342–349, Vol. 2504, 1995.
129. T.F. Jenkins, M.E. Walsh, W. Schumacher, P.G. Thorne, 'Development of Colorimetric Field Screening Methods for Munitions Compounds in Soil', in *Environmental Monitoring and Hazardous Waste Site Remediation*, eds. T. Vo-Dinh, R. Nießner, SPIE, Bellingham, WA, 324–333, Vol. 2504, 1995.
130. K. Levsen, A. Prieß, E. Berger-Prieß, 'Analysis of Organic Compounds in Aqueous Samples of Former Ammunition Plants', in *Environmental Monitoring and Hazardous Waste Site Remediation*, eds. T. Vo-Dinh, R. Nießner, SPIE, Bellingham, WA, 350–361, Vol. 2504, 1995.
131. W.J. Buttner, M. Findlay, W. Vickers, W.M. Davis, E.R. Cespedes, S. Cooper, J.W. Adams, 'In Situ Detection of Trinitrotoluene and Other Nitrated Explosives in Soils', *Anal. Chim. Acta*, **341**(1), 63–71 (1997).
132. P.T. Charles, J.C. Bart, L.L. Judd, P.R. Gauger, F.S. Ligler, A.W. Kusterbeck, 'Continuous Flow Fluorescence Based Immunosensor for Detection of Explosives and Environmental Pollutants', in *Chemical, Biochemical, and Environmental Fiber Sensors IX*, ed. R.A. Lieberman, SPIE, Bellingham, WA, 80–87, Vol. 3105, 1997.
133. A.A. Peterson, 'A Report on the Detection and Identification of Explosives by Tagging', *J. Forensic Sci.*, **26**(2), 313–318 (1981).

Laser Mass Spectrometry in Trace Analysis

Ulrich Boesl and Hans-Jörg Heger
Technische Universität München, Garching, Germany

Ralf Zimmermann
GSF Forschungszentrum für Umwelt und Gesundheit GmbH, Oberschleissheim, Germany

Holger Nagel
Bruker Franzen Analytic GmbH, Bremen, Germany

Peter Püffel
BMW AG, München, Germany

1 Introduction 1

2 Resonant Laser Mass Spectrometry: Scientific and Technical Principle 3
- 2.1 Resonance-enhanced Multiphoton Ionization 3
- 2.2 Time-of-flight Mass Spectrometry 5
- 2.3 The Combining of Resonant Laser Ionization and Reflectron 6
- 2.4 High-resolution Laser Ultraviolet Spectroscopy 6
- 2.5 Gas Inlet System 8

3 Historical Retrospective and Modern Developments with Relevance to Trace Analysis 8

4 Application to Trace Analysis of Complex Chemical Mixtures 13
- 4.1 Application to Trace Analysis of Exhaust Emissions from Motor Cars: Dynamic Processes Behind the Combustion Chamber and in the Catalytic Converter 13
- 4.2 Application to Trace Analysis of Exhaust Emissions from Motor Cars: Traces of Motor Oil in the Exhaust 16
- 4.3 Application to Trace Analysis of Exhaust Emissions from Industrial Incinerators 17
- 4.4 Application to Further Analytical Problems 20

5 Further Specifications of Resonant Laser Mass Spectrometry in Trace Analysis 22
- 5.1 Calibration Methods and Detection Limits 22
- 5.2 Specific Features of Resonant Laser Mass Spectrometry 23

Acknowledgments 24

Abbreviations and Acronyms 24

Related Articles 25

References 25

Laser mass spectrometry (laser-MS) in molecular trace analysis involves two analytical tools: Ultraviolet (UV)-spectroscopy in the gas phase and time-of-flight mass selection. Both tools are combined by resonance-enhanced multiphoton ionization (REMPI). This type of mass spectrometry is therefore called resonant laser-MS in the following text. The special features of resonant laser-MS are high selectivity, rapidity, multicomponent capacity, and adaptability for application to many different problems. Beneath the fundamental principle of this analytical technique (i.e. REMPI and time-of-flight mass selection) several options are described in this article, such as medium or high resolution of mass selection and UV-spectroscopy. These options allow more or less sophisticated instrumentation and optimized adaptation to specific analytical problems. The reader will be introduced to the principles of resonant laser-MS and then to its history and the state-of-the-art of its application. This organization should help the reader to recognize the importance of the early experiments and of modern tendencies for application of laser-MS in trace analysis. The large experimental and instrumental variety and the tables of numerous researchers and applications presented in section 3 may demonstrate the flexibility and growing acceptance of resonant laser-MS. The examples presented, which illustrate the usefulness of resonant laser-MS in practice, include on-line trace analysis of emissions from combustion processes, e.g. from combustion engines or municipal incinerators. The article ends with a discussion of possible calibration processes, detection limits, and a summary of specific features.

1 INTRODUCTION

The combination of lasers and mass spectrometry has been widely accepted in research as well as in industrial laboratories in the form of matrix-assisted laser desorption/ionization (MALDI) for the detection of large biological molecules[1,2] and also of laser ablation for elemental analysis of surfaces.[3,4] However, for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the application of laser-MS in trace analysis (particularly in molecular trace analysis, which is the main subject of this article) only insufficient selectivity can be achieved by mass spectrometric means alone, and an additional species-selective discrimination is necessary. For this reason, in conventional mass spectrometry, the highly sophisticated technique of coupled gas chromatography/mass spectrometry (GC/MS) has been developed. This is a two-dimensional spectrometric method applying the selection parameters retention time and mass. Its twofold selectivity has permitted the sensitivity of conventional ultratrace analysis of complex mixtures to be extended to extremely low detection limits. Extensive, time-consuming and laboratory-bound procedures for sample preparation have to be accepted, though. New methods of trace analysis, which may not be capable of reaching these low detection limits, but exhibit short detection times and possibly significant levels of mobility, are thus needed. In particular, trace analysis, which is planned for incorporation into industrial processes for the purpose of feedback control, or which is needed for the analysis of strongly fluctuating emissions of gases, or which should permit the attainment of fast results in the case of chemical accidents, ought to be able to be performed on a timescale of minutes or even seconds. However, even a performance speed in the region of 1 h would represent progress in many cases. Laser-MS using pulsed lasers and time-of-flight mass analyzers (one mass spectrum for every laser pulse) fulfills the requirement of rapidity exceptionally well. With regard to the requirement of two-dimensional selectivity, a new option is possible in laser-MS, which is not available in conventional mass spectrometry.

This option is resonant laser-MS, which combines the two analytical tools, UV spectroscopy (in the gas phase) and mass spectrometry (time-of-flight). The combining element is REMPI. Several modes of laser-MS are possible in trace analysis, e.g. involving high or reduced resolution of UV spectroscopy and mass selection. In the case of actual industrial tasks of chemical analysis (e.g. arising in test facilities of the automobile industry or in municipal incinerators) we found that in addition

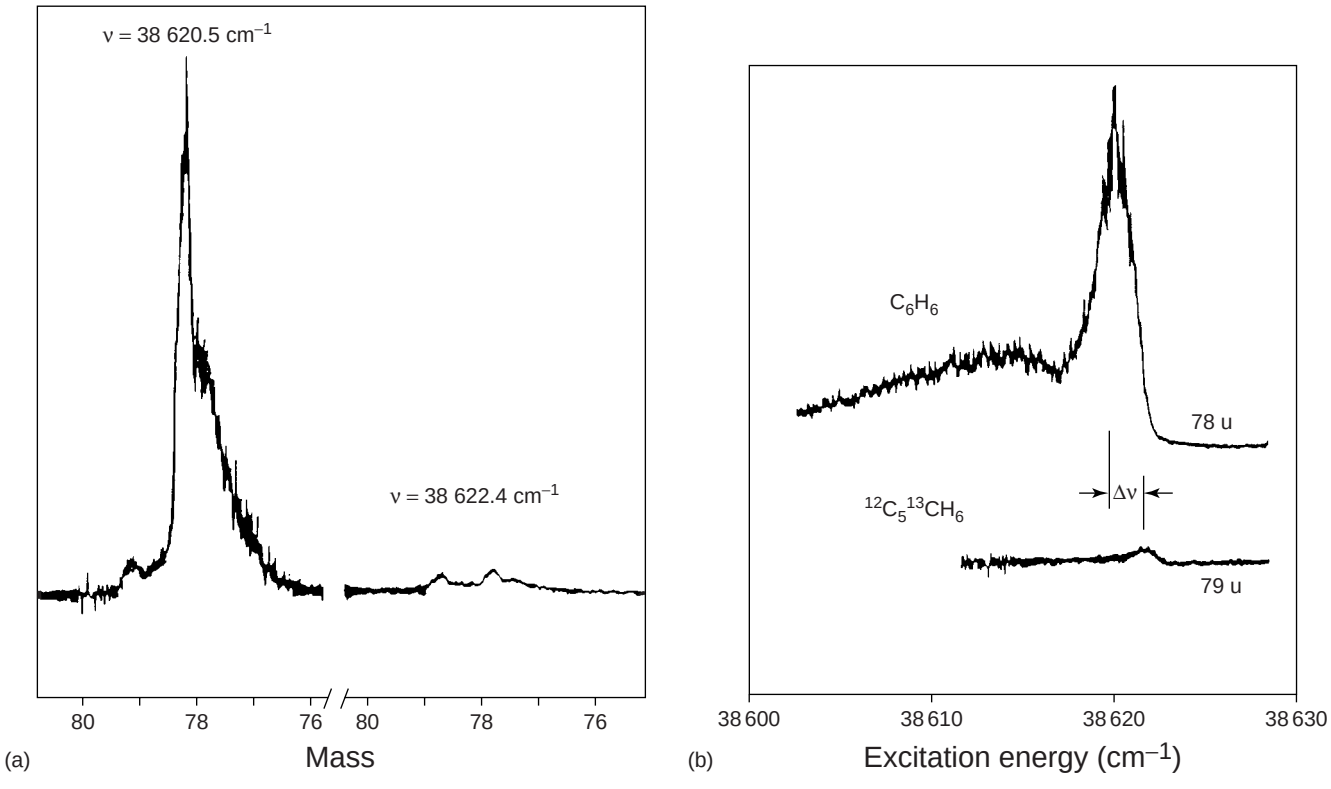

Figure 1 The first mass spectrum (a) and UV spectrum (b) of a polyatomic molecule (benzene) measured using a resonant laser-MS apparatus employing laser light with tunable wavelength. (a) Mass spectra of benzene in its natural isotopic mixture taken at two excitation energies. In the right mass spectrum the abundant mass 78 is suppressed, thus relatively enhancing the $^{13}C^{12}C_5H_6$ isotopomer. (b) Mass-selective UV spectra taken at the two masses 78 and 79 for abundant and $^{13}C_1$-substituted benzene. The lower spectrum demonstrates that even in a mixture of components pure UV spectra can be measured for one selected trace component. It should be noted that these experiments were performed at a very early stage in the development of laser technology and of the adaptation of laser ionization in mass analyzers.

to high-resolution laser-MS the version with reduced UV and mass resolution is of particular practical use. Most of the applications presented in section 4 have been achieved using this option. The high-resolution version of laser-MS is necessary for specific problems, such as discriminating between structural isomers (e.g. traces of chlorinated aromatics within highly complex gas mixtures) or even isobaric isotopomers.

To illustrate the two-dimensional selectivity of laser-MS, an early mass-selected resonance ionization spectrum is presented in Figure 1 together with a REMPI mass spectrum.[5,6] Both of these, the mass-selected laser spectrum and the spectroscopically enhanced mass spectrum, may be considered as being complementary, with either the mass or the laser wavelength fixed, and the other parameter being freely variable. The resonance ionization spectra in Figure 1(b) were recorded mass selectively (at 78 and 79 u), and show the $S_1(\nu_6)-S_0(0)$ transition of the abundant $^{12}C_6H_6$ and the rare $^{13}C^{12}C_5H_6$ natural isotopomer. They reveal an isotopic shift of $\Delta\nu = +2\,cm^{-1}$ for the ν_6 vibronic band. This shift is sufficiently large to selectively excite these two isotopomers. In Figure 1(a), two UV resonance-selected mass spectra are shown which were produced by REMPI at 38 620.5 and 38 622.4 cm^{-1}. The latter represents the band maximum position of the 79 u selective laser spectrum, and clearly demonstrates a significant relative enhancement of the heavier mass in the mass spectrum.

Furthermore, the mass peaks in Figure 1(a) exhibit a wing to smaller masses, producing a considerable broadening of the peak and a reduction in mass resolution. This is not due to an intrinsically small resolution in the quadrupole mass analyzer used for this experiment, but rather to metastable decay induced by photon absorption during laser ionization. This effect, which was observed very early without its origin being recognized, shines light on a different feature of resonance-enhanced ionization, and concerns fragmentation and metastable decay. A renaissance in ion kinetic decay studies and the development of new methods and new applications have been initiated by this feature. The possibility of producing both surprisingly soft and exceptionally hard ionization was recognized already in the earliest stage of development of resonance ionization and mass selection. Its high sensitivity (e.g. one atom detection) and high selectivity (isomer-selective ionization) represent further remarkable characteristics. In the meantime, the importance of mass-selective laser spectroscopy and optically selected mass spectrometry for trace analysis has been recognized.

The method of resonant laser-MS has been so successfully introduced into research laboratories that an extensive body of survey literature already exists, e.g. an introduction to lasers in chemical analysis and mass spectrometry,[7–13] a perspective of molecular spectroscopy by resonance ionization and characterization of molecules with regard to their behavior during this ionization method,[14–24] resonant laser-MS,[12,22–30] and resonance ionization mass spectrometry of atoms.[31–33] The objective of this article is to provide an introduction to the principles and possibilities of resonant laser-MS, applied in molecular trace analysis. Therefore, in section 2, the scientific and technical features of the method are introduced. In section 3, the initial stages of resonant laser-MS and its modern developments are presented. The large range of subjects, accessible molecular species, and research groups demonstrate the wide distribution and acceptance of this method. In section 4, examples from our own work have been selected to illustrate the versatility of laser-MS. Of course, these examples do not cover all the possibilities and facets of resonant laser-MS in trace analysis exploited and developed by the many research active groups in this field. They rather are intended to provide a first insight rather than an overall review, the latter being disproportionate in terms of an introduction to the method. Finally, in section 5, calibration methods, detection limits, and special features of applied resonant laser-MS are discussed.

2 RESONANT LASER MASS SPECTROMETRY: SCIENTIFIC AND TECHNICAL PRINCIPLE

2.1 Resonance-enhanced Multiphoton Ionization

The combining of laser spectroscopy with mass spectrometry to provide a real two-dimensional analytical tool is made possible by REMPI. On the one hand, this laser-induced process is used to produce the ions which are needed for mass spectrometry; on the other hand, resonant intermediate states, which enhance this ionization process, intrinsically involve molecular UV spectroscopy. The principle behind REMPI is presented in Figure 2: a UV photon is absorbed exciting a higher energy level of the neutral molecule. A second photon is absorbed, which reaches above the molecular ionization threshold. This $(1+1)$ ionization represents the most effective multiphoton-ionization (MPI) process. However, this process is not generally applicable, e.g. it cannot be used for some small molecules or to alkanes. This is due to unfavorable energies of the accessible intermediate states and of the ionization threshold. In this case, MPI processes of a higher order have to be used, such as $(2+1)$ multiphoton absorption (see Figure 2a). All these processes involve the molecular spectroscopy of neutral intermediate states (see Figure 2b). Each molecule has its own characteristic UV absorption spectrum. Using an appropriate

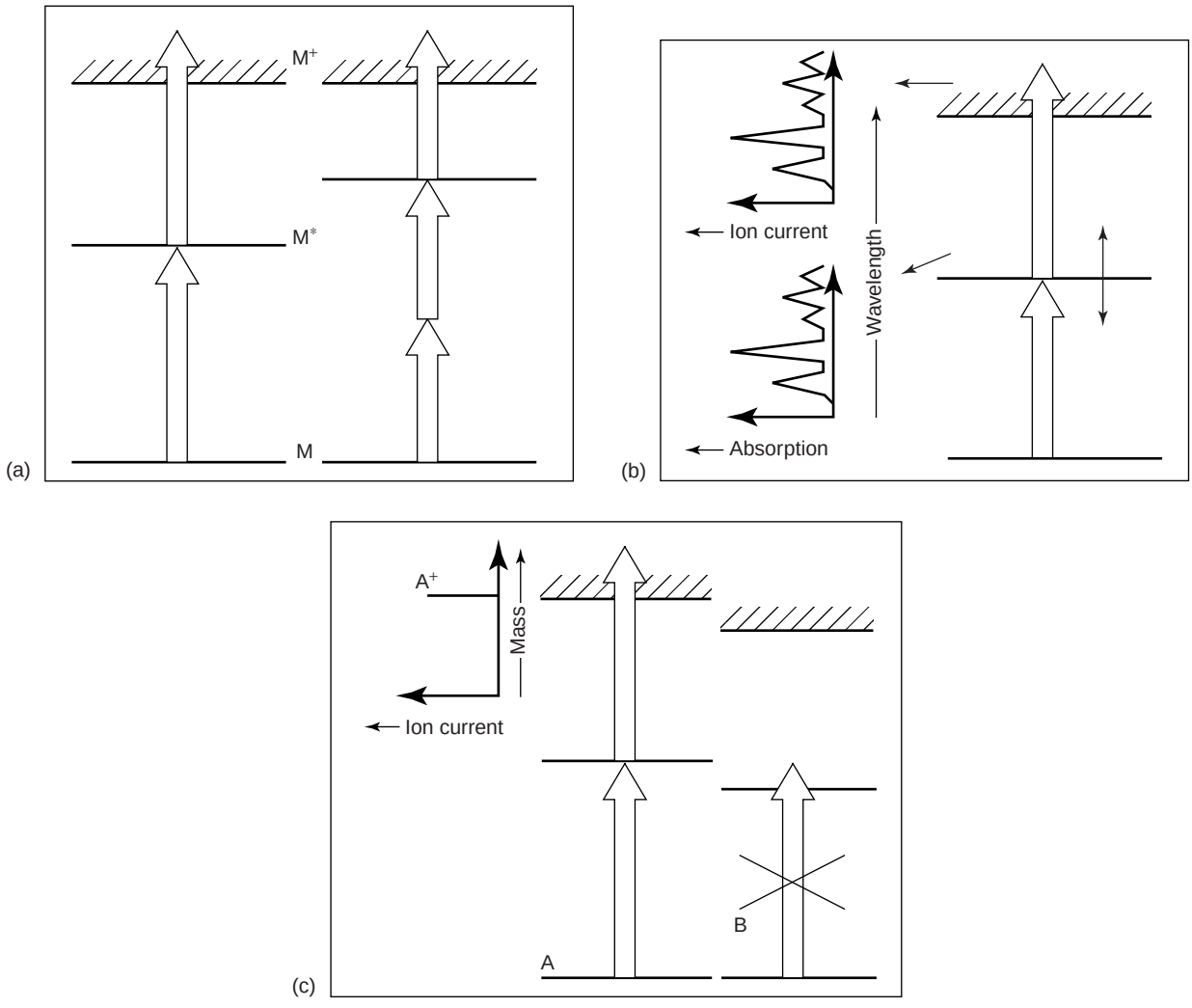

Figure 2 The process of REMPI. (a) The principle of $(1+1)$ and of $(2+1)$ MPI. (b) The tuning of the laser wavelength and recording of the ion current permits one to measure molecular spectra which are similar to UV absorption spectra. (c) Keeping the laser wavelength in resonance with an absorption band of one selected molecule allows selective ionization of this molecular species. The prerequisite for this is the knowledge of UV spectroscopy for this particular molecular species.

laser wavelength, selective absorption and (by resonant MPI) selective ionization of a particular molecule can be achieved (Figure 2c). Thus, trace species can be analyzed selectively by REMPI within a gas mixture consisting of many other components in considerably larger concentrations.

In contrast to atoms, with their low number of energy states and narrow absorption bands, the density of states in molecules is considerably larger. This often results in heavily congested spectra, with regions of broad, unstructured absorption bands, particularly at higher temperatures (e.g. 500 K). As a result, selectivity is reduced. The advantage of this in itself unfavorable factor is that it enables synchronous analysis of several preselected trace components, due to an accidental overlapping of the molecular absorption bands. The molecular absorption cross-sections are smaller than typical atomic ones, to a level of several orders of magnitude, but nevertheless sufficient to provide high sensitivity. Typical ionization yields [ratio of the number of ionized molecules to the total number of molecules within the laser focus[34]] are 10^{-2} for $(1+1)$ MPI and 10^{-5} for $(2+1)$ MPI. These can be enhanced by increasing laser intensities. However, at higher laser intensities, which may be necessary for higher-order MPI processes, multiphoton-induced fragmentation takes place[35,36] in addition to ionization, reducing selectivity considerably. Therefore, laser wavelength and intensity have to be chosen carefully, in order to prevent fragments from other molecules from overlapping with the molecules of interest, within a mass spectrum. Other unfavorable effects, which may reduce ionization efficiency significantly, are

fast relaxation processes affecting the resonant intermediate electronic state or unfavorable Franck–Condon factors of the first or second absorption step. In the case of these occurring, alternative intermediate states or MPI schemes have to be chosen, or techniques such as species-specific fragment ion detection,[37] preselection by fast gas chromatography (GC) and, in particular, application of femtosecond lasers[22] have to be involved.

2.2 Time-of-flight Mass Spectrometry

For REMPI of molecules, mostly pulsed lasers are used. This permits ion production within a short time and in small volumes, these being ideal preconditions for time-of-flight mass spectrometry (TOFMS). In pulsed laser ion sources, the ions are usually produced in a static electric field. These are accelerated at the moment of ionization, and reach a field-free drift tube after some microseconds. As all ions are produced within the confines of a very small volume, and therefore at nearly the same electrical potential, they have very similar kinetic energies in the field-free drift region. As a result, the velocities, and therefore the flight times of the ions, are proportional to the square root of the mass. Finally, the ion current measured as a function of time supplies mass information about the ionized species. A schematic view of a time-of-flight (reflection) mass analyzer is shown in Figure 3. Several factors, such as a finite volume of the laser focus, initial velocities of the neutral molecules and space charge effects, result in a wider distribution of kinetic energy and therefore in an increased distribution of flight times. This reduces the mass resolution, and is a well-known problem in linear time-of-flight mass spectrometers.[26,38]

The so-called reflectron time-of-flight mass analyzer[26,39] compensates for most of these factors by reflection of the ions within a special reflecting field. The principles of its operation are as follows. The originally diverging ion clouds are compressed at the space focus of the pulsed ion source. The ion reflector projects this space focus onto the surface of the ion detector. This results in short ion pulses at long ion flight times, thus providing high mass resolution. This principle is indicated in Figure 3 by the different widths of the ion clouds in front of, inside, and behind the ion reflector. An optimal mass resolution as high as $R(50\%) = 20\,000$[26,40,41] has been achieved in this way. The routine mass resolution of reflectrons lies in the range $R(50\%) = 3000–5000$. A very good mass resolution helps to achieve not only efficient mass selection, but also a good signal-to-noise ratio: the ions investigated appear within a very narrow time window (e.g. 10–20 ns) while background signals are randomly

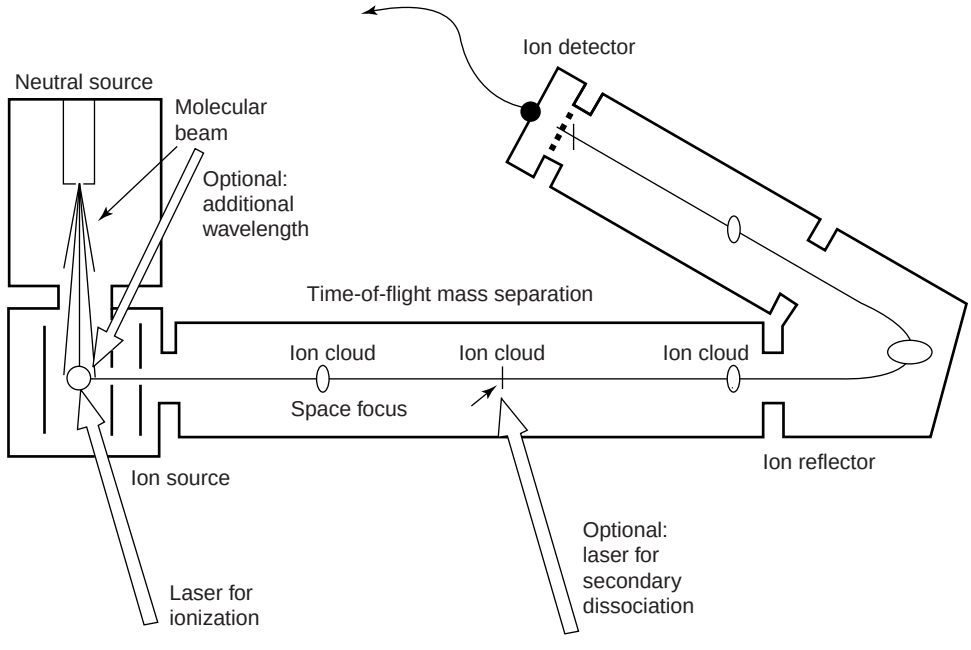

Figure 3 Schematic diagram of the laser mass spectrometer consisting of the neutral source (gas inlet), the ion source (ion optics, position of laser ionization), time-of-flight regions for mass separation, the ion reflector for compensation of time-of-flight distribution (high mass resolution), and the ion detector. For two-color ionization a second laser wavelength is needed which can generally be provided by the pump laser of the tunable laser source for resonant excitation of the molecular intermediate states (first laser wavelength fully tunable, second laser wavelength harmonics of neodymium–yttrium-aluminum garnet (Nd : YAG) laser: 355, 266 or 213 nm).

distributed over a large time region (e.g. 100 ms). As indicated in Figure 3, there is also a special location within the field-free drift region, the so-called space focus. This is a typical feature of pulsed ion sources and represents a point of maximum ion pulse compression, and therefore also of optimized mass resolution. By secondary excitation, with laser-induced fragmentation, a type of tandem mass spectrometry is achieved[36] without any major change necessary in the reflectron time-of-flight spectrometer itself.

Further to high sensitivity and good mass resolution, speed is another specific beneficial feature of reflectron time-of-flight mass spectrometers. Each ionizing laser pulse produces a complete mass spectrum. In principle, the repetition rate is limited by the overlap of two consecutive mass spectra. A repetition rate of 10 kHz could be achieved, limited by the typical flight times of the slowest ions of some 100 μs. Until now, suitable lasers with such a high repetition rate have not been available; besides this, the temporal behavior of typical gas inlet systems does not allow such a high time resolution. We have achieved a repetition rate of 100 Hz with available pulsed lasers and mass spectrometers fitted with a special gas inlet system. An additional advantage of time-of-flight mass analyzers is their simple and sturdy construction.

2.3 The Combining of Resonant Laser Ionization and Reflectron

The combination of these two methods results in a two-dimensional analysis technique applying the parameters of laser wavelength and mass. The choice of the suitable wavelength allows species-selective ionization, and enhances the discrimination of chemical compounds which cannot be separated by means of their mass alone. An illustrative example of this is the carbon monoxide (CO) in the exhaust emissions of motor cars. CO has the same nominal mass as nitrogen but is much less abundant there. Whereas CO is easily ionized at 230 nm, N_2 does not show any noticeable degree of ionization efficiency at all at this wavelength. Some further typical exhaust components with identical nominal masses are, for example, formaldehyde, ethane, and nitrogen monoxide with a mass of 30; acetaldehyde, carbon dioxide, and dinitrogen oxide with a mass of 44; propionaldehyde and acetone with a mass of 58; polycyclic and substituted aromatics with all their isomers and fragments, e.g. the harmless perylene and the carcinogenic benzo[*a*]pyrene, both with a mass of 252. Without selective ionization, these components can only be separated by means of extremely high mass resolution, which brings with it the disadvantages of reduced sensitivity and speed, and increased costs.

An example of a scenario for the combination of resonant ionization and laser-MS for trace analysis is the following case of motor exhaust gas monitoring. Typical points for the collection of gas samples are behind the valve (single-cylinder emission), and in front of and behind the catalytic converter. The exhaust gas, with its rapidly fluctuating concentrations of trace compounds, is guided into the mass spectrometer by means of a 6 mm tube of about 500 cm in length, and introduced into the ion source by a pulsed valve, forming supersonic molecular beams, or by a thin metal needle emitting effusive beams. Thus, parts of the sampled gas are channeled to the spot in the ion source, where the ionizing laser is focused. This laser is fired releasing pulses of typically 10 ns in pulse length. The laser pulse defines flight time $t = 0$, the time at which the ions start to be accelerated. They reach the ion detector after a flight time which is determined by their mass. When the slowest ion has arrived, a new gas pulse can enter the ion source and be ionized by the next laser pulse. By synchronizing the motor ignition cycle with the ionizing laser pulse, and delaying the laser pulse in relation to the motor cycle in steps (e.g. of a few milliseconds or less), even the temporal profile of concentrations in the original exhaust gas can be analyzed species selectively with a sub-millisecond time resolution, provided the conditions within the motor remain stable during this sampling process.

2.4 High-resolution Laser Ultraviolet Spectroscopy

UV spectroscopy in chemical analysis is mostly applied to compounds in solutions and therefore exhibits only minor species-selective information in comparison with methods such as infrared (IR) spectroscopy. For larger molecules [e.g. polycyclic aromatic hydrocarbons (PAHs)], even UV spectroscopy in the gas phase often does not supply good molecular fingerprints. This is due to low-frequency vibrations and rotational motion which are present in a highly excited state at room temperature. A typical example of this is the internal rotation of the two phenyl rings of biphenyl around their connecting single bond. Thus, the gas-phase UV spectrum of biphenyl is at room temperature totally congested and does not show any clear structure. These broadening effects impede the achievement of optimum spectral resolving power, and therefore of the selectivity which would be possible by means of UV spectroscopy (and consequently also REMPI).

Modern laser UV spectroscopy, however, often provides a well separated fine structure which is comparable to or has even better resolution than in IR spectroscopy. This is achieved by cooling the molecules in an isenthalpic gas expansion produced by an external pressure of several bar which expands into a vacuum through a narrow hole (e.g. width 0.2 mm, length 1 mm); in this way a so-called

supersonic beam is formed. During the expansion process, low-frequency vibrational and rotational motions freeze out (rotational quasi-temperatures below 10 K are achievable) without condensation of the gas molecules occurring. This reduces the number of possible vibronic transitions, and therefore also the number of lines in the UV spectrum, considerably. In addition, these lines exhibit a significant narrowing of their line width, due to the rotational cooling. The effect of cooling in a supersonic molecular beam is illustrated in Figure 4, which shows the vibrationless electronic origin of the S_1-S_0 transition of one of the three different monochloro anthracene isomers. These spectra were measured using resonance ionization spectroscopy in an effusive beam at room temperature (Figure 4, top trace) and in a supersonic beam (Figure 4, bottom trace). The significant narrowing of the vibronic band is obvious. This is due to the lower level of rotational excitation in the S_0 ground state, and therefore to the occurrence of fewer transitions contributing to the rotational envelope of this band.

To generate such a supersonic molecular beam one uses mostly Ar or He as the carrier gas, with the molecule of interest seeded into this. To avoid putting the vacuum system under a too heavy load, instead of a continuous gas inlet, pulsed expansion is used, delivering gas pulses of typically 0.2 ms duration at a rate of repetition of 20–50 Hz (compatible with typical pulsed laser repetition rates). Even molecules with low vapor pressure can be investigated, by means of high-temperature valves, which, nevertheless, allow sufficient cooling of internal molecular degrees of freedom.

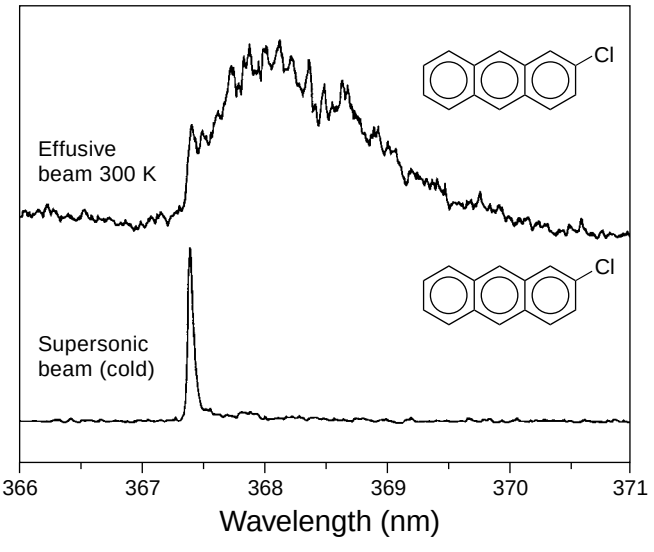

Figure 4 Illustration of the effects of supersonic beam cooling on the structure of molecular spectra (here 2-chloroanthracene). Top trace, effusive beam at room temperature; bottom trace, supersonic beam at less than 20 K rotational quasi-temperature.

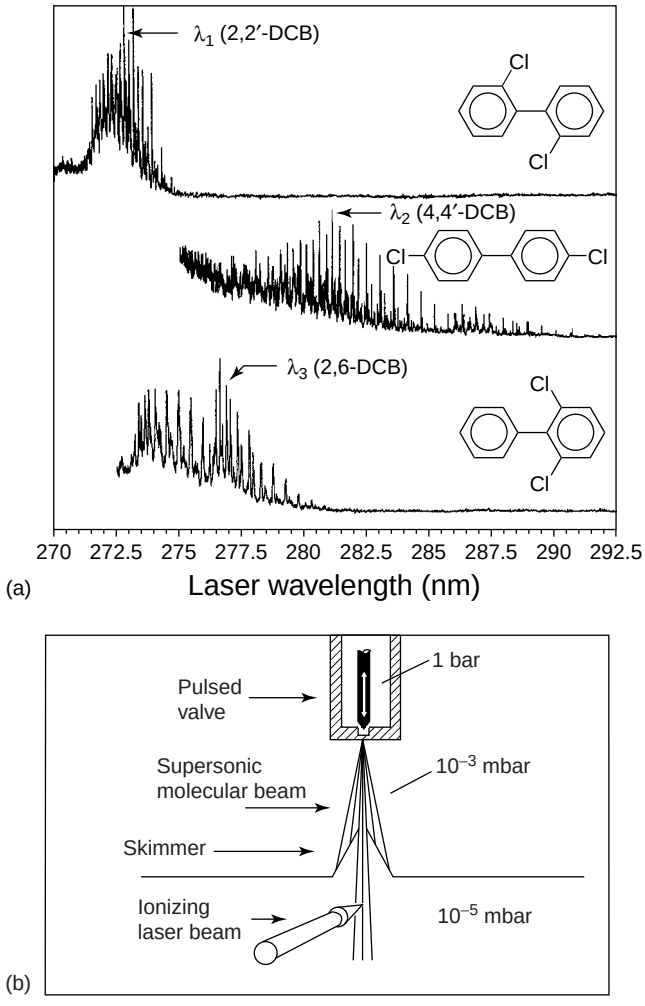

Figure 5 (a) Supersonic beam UV absorption spectra for three isomers of dichlorobiphenyl. Isomer-selective laser excitation is possible for instance at the marked wavelengths λ_1, λ_2, and λ_3. These spectra were taken by resonance ionization spectroscopy [see Figure 2b] using two-color excitation; this is possible with one single laser system (dye laser pumped by an Nd/YAG laser): the first, tunable wavelength is achieved using the dye laser, while the second fixed wavelength can be obtained from the pumping Nd/YAG laser. This allows for higher ionization efficiency and the achieving of low-lying, but sharp and therefore highly selective intermediate states in the resonant ionization process. (b) Schematic diagram of the supersonic beam arrangement with pulsed valve, skimmer and pressure regimes.

Today, supersonic beam spectroscopy represents a highly effective technique and has been successfully introduced in laboratories for molecular physics.[42–44] Its virtues for analytical applications were recognized at an early stage. Very well structured and simplified molecular spectra now even allow isomer selection, e.g. of large aromatic molecules. As an example, Figure 5 shows the laser ionization spectra (comparable to UV absorption spectra) of three dichlorobiphenyl isomers. The spectra are

well structured, making it possible to find excitation wavelengths (e.g. λ_1, λ_2, and λ_3) at which every single isomer can be excited and ionized selectively. Isomer selectivity has been achieved for many different PAHs (e.g. perylene and benzo[*a*]pyrene) and for chlorinated aromatics. Depending on the special analytical problem involved, one may choose high-resolution REMPI with supersonic beams or medium-resolution REMPI with effusive molecular beams. The latter is sufficient if polycyclic and substituted aromatics are to be monitored as a class of chemical compounds contained within a complex mixture of pollutants (e.g. in emissions from industrial combustion processes) but without isomer selectivity. Medium-resolution REMPI is also sufficient for trace analysis of emissions from combustion engines.

2.5 Gas Inlet System

Suitable sampling is crucial for reliable trace analysis. Since mass spectrometry is primarily a gas analytical tool, it is well suited for trace analysis of gaseous emissions. One major problem arises from the large pressure differences between atmospheric pressure samples and the vacuum system ($<10^{-9}$ bar). Special means of pressure reduction are required, although at the point of ionization (defined by the laser beam) the gas density should be as high as possible. The only way to fulfill these conditions is to use molecular beams which can be produced in different ways. Examples of these arrangements are described in the following. Effusive continuous molecular beams can be formed using a thin metallic needle (e.g. 3 cm long, 0.5 mm inner diameter). This needle reaches into the ion source; its end is positioned near the axis of the electrode arrangement making laser ionization possible about 1 mm below the end of the tubing (minimum spatial broadening of the molecular beam) and, simultaneously, on the electrode axis (optimum ion trajectory) (see Figure 6a).

However, thin needles do not reduce the total gas flow sufficiently if a gas pressure of 1 bar or more is applied on the high-pressure side. Therefore, these have to be opened periodically for short intervals or the use of an additional pressure reduction system is necessary. In Figure 6(a), diagrams of both versions are shown with a U-shaped by-pass for the sample gas. In Figure 6(b), a supersonic beam gas inlet is presented; a small nozzle (0.2-mm inner diameter, 1-mm length) is periodically opened for very short times ($<200\,\mu s$). On the high-pressure side several bar may be applied. The large pressure drop to the vacuum side of the nozzle enables the formation of a supersonic beam. The special features of this gas inlet have been described in section 2.4. Additional pressure reduction, which is mostly necessary due to the large gas load of supersonic beams, is achieved by using a skimmer and a separate vacuum chamber. Unfortunately, this results in a large distance between beam nozzle and point of ionization, and therefore a reduced concentration at the laser focus. An improvement of this situation can be achieved using an in-line arrangement of the molecular beam as shown in Figure 6(c).

In Figure 6(d), a further important version of the set-up is presented, the combination of a supersonic molecular beam with laser desorption of neutral molecules. The latter can be selectively ionized by means of laser light and mass-selectively detected. Many applications are possible or have been put into use already, such as the study of biomolecules, analysis of solids and surfaces, testing of semiconductor surfaces for organic pollutants, or analysis of soil samples for dangerous polycyclic or substituted aromatics (see section 3). In addition, the combination of a supersonic beam valve with a GC capillary has been realized as shown in Figure 6(e). A special construction for this coupling is necessary to avoid first the reduction of GC resolution due to gas diffusion in slits and dead volumes or due to reduced gas flow, and second the inefficient cooling of the supersonic beam due to unfavorable pressure conditions.

Of course, it is of fundamental importance to use heated gas inlet systems for larger molecules, e.g. chlorinated or polycyclic aromatics. This can be achieved in resonant laser-MS with commercial heated transfer lines, heating within the vacuum system, optimum heat conductance to the very end of the gas beam-emitting device (needle or valve), and avoidance of any kind of cold spots (e.g. at the transit of gas lines through vacuum flanges).

3 HISTORICAL RETROSPECTIVE AND MODERN DEVELOPMENTS WITH RELEVANCE TO TRACE ANALYSIS

Already in the first 4–5 years since its introduction, most features of resonant laser-MS of molecules have been defined and studied and the foundations for most of the later steps of its development have already been laid at this early stage. By presenting these features, developments, and current topics of interest relating to trace analysis in three tables, together with the research groups involved, with particular work cited in the bibliography, and a summarizing discussion, we hope to provide an informative survey without overwhelming the reader with too many details. In this respect we would like to apologize for the fact that not all groups working in the field of resonant laser-MS with relevance to trace analysis could be mentioned in this survey.

Successful laser-induced (although not resonance-enhanced) ionization of molecules in the gas phase

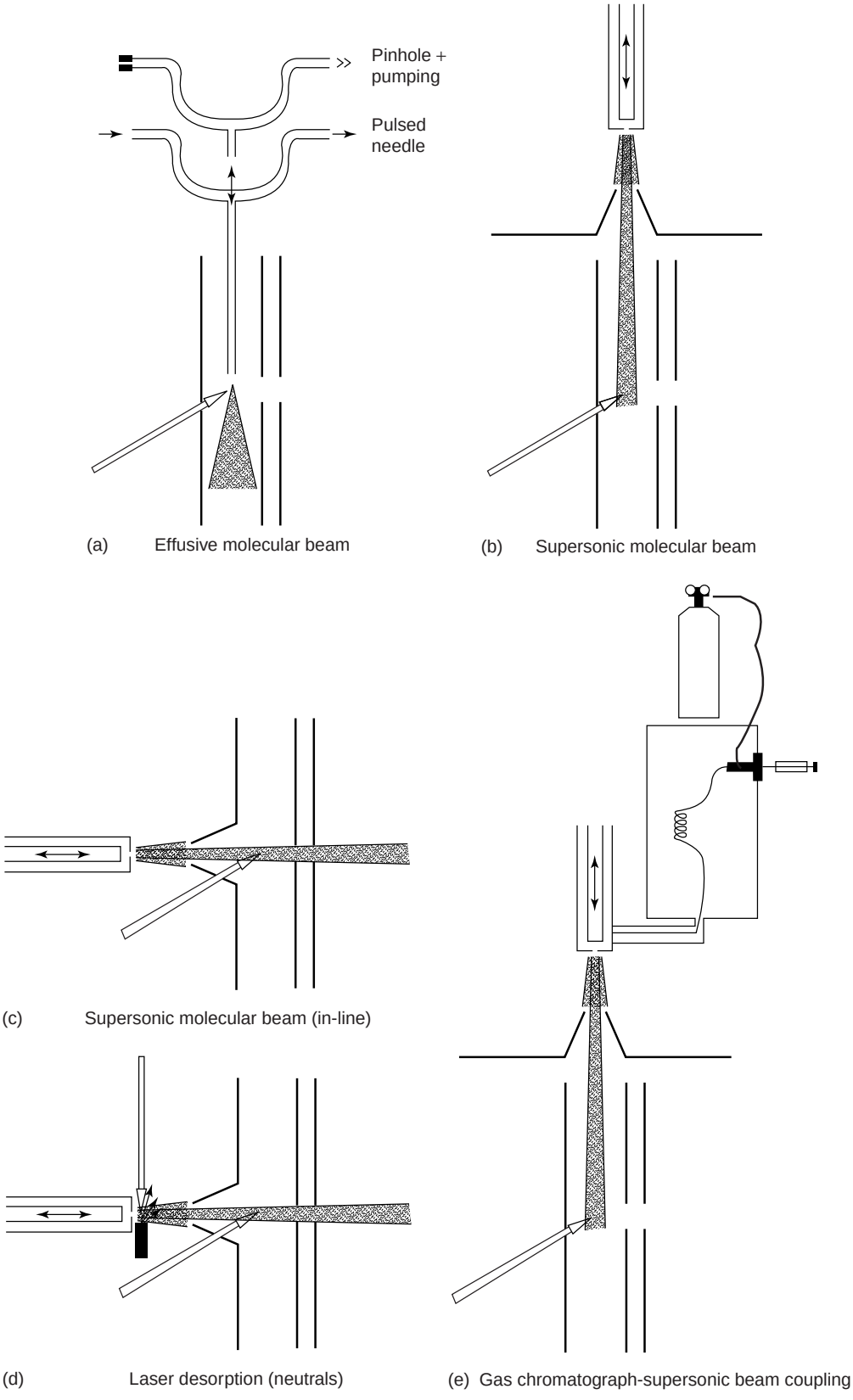

Figure 6 Different gas inlet systems for the supply of neutral sample molecules in the isolated gas phase, necessary for resonant laser ionization and time-of-flight mass analysis.

followed by mass selection was reported on as early as 1970 (Table 1).[45,46] Resonant laser ionization was applied to molecules 5 years later.[14] The method was introduced as a new means for UV and visible molecular spectroscopy rather than mass spectrometry. Even its combination with mass selection[48–52] was originally rather a further step towards a new form of molecular spectroscopy. In this context, it is interesting to observe that another new technique, namely the cooling of molecular rotational and vibrational internal motions in a supersonic gas beam,[42,47] revolutionized molecular spectroscopy. It did not take long for resonant laser ionization and supersonic molecular beams to be combined, thus starting a fruitful period of molecular spectroscopy on larger molecules.

While the first application of laser ionization [both resonant[48–52] and non-resonant[45,46]] was on small, mostly two-atom molecules, first mass-selective laser ionization experiments on polyatomic molecules were reported in 1977 and 1978. The first polyatomics were benzaldehyde, benzophenone, and nitrotoluene, which were still ionized using a fixed-frequency laser,[53] and benzene, for the ionization of which the wavelength of a dye laser was tuned to its $S_1(\nu_6)-S_0$ transition.[5,6] These experiments were followed by many publications on all kinds of molecules, their mass-selected laser spectroscopy or spectroscopically selected mass spectra.[54–65] Already in the early years of REMPI (1982–85) several groups therefore tried to characterize molecules according to their behavior under resonant laser ionization conditions.[16–18]

Two outstanding features of resonant laser ionization – very soft ionization and variability from metastable decay to extreme fragmentation – were already detected in the first experiments. Therefore, several research groups went on to investigate the mechanisms of fragment ion formation during resonant laser ionization.[57,58,66–71]

Table 1 The beginning of resonant laser-MS

State of development	Year	Authors
Mass-selective ionization of molecules by lasers		
(fixed frequency lasers)	1970/71	Berezhetskaya et al.[45]
H_2, I_2, D_2O, CCl_4		Chin[46]
Molecular resonance ionization spectroscopy		
(in the gas cell)	1976	Johnson[14]
Supersonic beams and molecular spectroscopy		
(no ionization)	1975	Smalley et al.[42,47]
Mass-selective resonance ionization spectroscopy		
(diatomic molecules)	1977/78	Wöste et al.[48,49]
Na_2, K_2, I_2, Li_2, BaCl, NO		Zare et al.[50]
		Bernstein et al.[51]
		Rothe et al.[52]
Mass-selective resonance ionization of polyatomic molecules		
(fixed frequency lasers)	1977/78	Letokhov et al.[53]
Benzaldehyde		
Benzophenone, nitrotoluene		
Mass-selective resonance ionization spectroscopy of polyatomic molecules		
(tuneable lasers, spectroscopy and mass spectrometry)		
Benzene	1978	Boesl et al.[5,6]
Aniline	1979/80	Brophy, Rettner[54,55]
Triethylenediamine, benzene, butadiene		Bernstein et al.[56–58]
Metal carbonyls, aniline, bromobenzene, iodobenzene		Smalley et al.[59–61]
Azulene, naphthalene	1980	Lubman et al.[62]
Acetaldehyde		Robin et al.[63]
Pyrrole, *N*-methylpyrrole, furan		Compton et al.[64]
Thiophene, toluene, furan (cross-sections)		Boesl et al.[65]
Resonance ionization and fragmentation of polyatomic molecules		
Extreme fragmentation	1979	Bernstein et al.[57,58]
Extreme fragmentation	1980	Boesl et al.[66]
Photoelectron analysis		Reilly et al.[67]
The role of the neutrals	1981	El-Sayed et al.[68]
2,4-Hexadiyne		Baer et al.[69]
Tertiary amines		Bernstein et al.[70]
Aniline		Zare et al.[71]

Different methods have been used to this purpose, such as secondary fragment ion excitation using a time-delayed second laser (representing the first steps towards tandem laser-MS), a new kind of photoelectron spectroscopy revealing fast intramolecular processes in the neutral intermediate state, and picosecond laser excitation. In fact, in addition to molecular spectroscopy, the new tool of resonant laser ionization has been applied mainly to the study of internal molecular dynamic processes.

The combination of resonant laser-MS with other techniques proved to be an important factor in its development since the very beginning. In this respect, some early experiments are listed in Table 2. The combination with supersonic molecular beams has already been mentioned above; its special features and benefits are presented in section 2.4. In addition to providing the possibility of performing isomer- and even isotopomer-selective molecular ionization, it was also able to deliver important spectroscopic information (a prerequisite for highly selective resonant laser ionization) about larger molecules.[59–61,72–75] Another intriguing early combination was that with GC.[76,77] On the one hand, this allows the GC of very complex gas mixtures without the need for laborious chemical separation of interfering species in advance (due to the high level of selective detection achieved by resonant laser-MS). On the other hand, this combination permits to quantify or calibrate resonant laser mass spectra, or to test its sensitivity.

In Table 2, some citations are given under the heading Resonance ionization of molecules and trace analysis. Again, very early, the advantages and new possibilities of resonant laser ionization for ultratrace analysis were recognized and detection limits for aniline,[54] and NO,[78] and the detection of single naphthalene molecules in the laser focus volume[79] were reported between 1979 and 1981. Another important technique performed in combination with resonant laser-MS is reflectron TOFMS (see also sections 2.2 and 2.3). In addition to high mass resolution, new means of performing tandem mass spectrometry and of studying and putting into practical application of metastable decay of molecular ions[2,26] were provided by this combination. Since resonant laser ionization also permits the enhancement of metastable decay (as mentioned above; see also Figure 1), the connection of both methods proved to be very promising.

A third combined technique is neutral laser desorption in a resonant laser mass spectrometer. Its different facets such as picosecond laser desorption,[82] supersonic

Table 2 The beginning of resonant laser-MS: combination with other techniques

Method	Year	Authors
Supersonic beams	Since 1979	
Metal carbonyls, aniline, bromobenzene, iodobenzene		Smalley et al.[59–61]
Benzene, naphthalene: two-color		Smalley et al.[72]
Biphenyl		Ito et al.[73]
Na_3		Wöste et al.[74]
Aniline: isotope-selective		Leutwyler, Even[75]
GC	Since 1980	
		Klimcak, Wessel[76]
		Reilly et al.[77]
Trace analysis	Since 1979	
Atmospheric pollutants (aniline)		Brophy, Rettner[54]
Sensitivity of NO detection		Welge et al.[78]
Single naphthalene molecules		Letokhov et al.[79]
Reflectron TOFMS	Since 1982	
Mass resolution $R = 4000$, metastable ion decay		Boesl et al.[80]
Neutral laser desorption	Since 1982/87	
Desorption of neutrals, non-laser postionization		Vastola et al. (1968)[81]
Picosecond laser		Letokhov et al.[82]
Supersonic beam		Von Weyssenhoff et al.[83,84]
		Trembeuill, Lubman[85]
		Boesl et al.[86]
Biomolecules		Letokhov et al.[82]
		Boesl et al.[86,88]
		Zare et al.[87]
Picosecond laser excitation	Since 1984	
Study of dissociation processes		El-Sayed et al.[89,91]
		Zewail et al.[90]

beam cooling of desorbed neutral molecules[83–86] and application to large biomolecules[82,86–88] were exploited at an early stage. By contrast with the other laser-MS method for solids, MALDI (see Introduction), resonant laser-MS of laser-desorbed neutrals is more suitable for trace analysis of complex mixtures of solid samples. This is due to its two-dimensional selectivity. It is not unexpected that laser desorption resonant laser ionization mass spectrometry has developed into an important field of research within its own right, with many highly interesting applications.

Another early combination,[89–91] namely that of picosecond lasers and resonant laser ionization, should also be mentioned. This has been applied to study fast intramolecular processes in the neutral intermediate states. These processes may reduce ionization efficiency.

Table 3 Tendencies of resonance multiphoton ionization mass spectrometry with relevance to trace analysis

Method	Application	Authors
Supersonic beam spectroscopy of molecules with environmental relevance (polycyclic aromatics, substituted, e.g. halogenated aromatics)	Non-ionization	Amirav et al.,[92] Hayes,[44,93] Levy,[43] Imasaka et al.[94]
	Resonant laser-MS	Ito et al.,[73] Smalley et al.,[59,60] Wallace et al.,[95] Levy et al.,[96] Rohlfing et al.,[97,98] Lubman et al.,[19,99,100] Imasaka et al.,[101] Boesl et al.[102,103]
Combustion processes (flames, incinerators, combustion engines) (polycyclic and halogenated aromatics, NO_x, acetaldehyde, etc.)	Flame sampling	Cool et al.,[104–106] Homann et al.,[107] Rohlfing et al.,[108] Siegmann et al.,[109,110] Senkan, Castaldi[111]
	Incinerators	Cool et al.,[112,113] Grotheer et al.,[114,115] Zimmermann et al.[116,117]
	Combustion engine	Boesl et al.[118]
Femtosecond laser excitation	Nitromolecules, NO_2	Ledingham et al.[119–122]
	Halogenated benzenes	Imasaka et al.[123]
	Biomolecules	Weinkauf et al.,[124] Imasaka et al.,[125,126] Lockyer, Vickerman,[127] Winograd et al.[128]
	Metal-organic compounds	Grotemeyer et al.[129,130]
	Comparison with vacuum-UV excitation	Winograd et al.[131]
GC Coupling	PAHs, alkylbenzenes,	Wessel, Klimcak[76] Fassel et al.,[132] Reilly et al.[77,133,134]
	Nitro, nitroso compounds	Reilly et al.[135]
	Aniline derivatives	Imasaka et al.[100]
	Supersonic beams	Peppich et al.[136] Imasaka[101], Grotemeyer et al.,[137] Zimmermann et al.[138]
Laser desorption and postionization	Biomolecules	Zare et al.,[139] Grotemeyer et al.,[140] Boesl et al.,[86] Weinkauf et al.,[141,142] Langridge-Smith et al.[143]
	PAHs, amino acids	Lubman et al.,[85] Boesl et al.,[144] Hankin et al.[145]
	Aromatics, PAHs, polymers	Hunziker et al.[146–148]
	Organics, natural samples	Zenobi et al.,[149–152] Zare[153–155] Langridge-Smith et al.,[156] Hanley[157]
	Mechanisms of neutral laser desorption	Zare et al.,[158] Zenobi et al.,[159–161] Li et al.[162]
	Frozen matrices	Alimpiev et al.[163,164]
	Aerosols, particles	Prather[165], Zenobi[152]
	Thermal desorption and REMPI	Johnston et al.,[166] Li et al.[167]
Resonance ionization of atoms	Single-atom detection	Hurst[122]
	Strontium, plutonium	Wendt,[168] Kluge[169]
	Laser ablation/resonant ionization	Hurst et al.,[170] Miller et al.[171]

Their study, and means of surpassing these effects, are also of importance for laser-MS in trace analysis. Recently, very short high-intensity laser pulses (femtosecond lasers) have been used for this purpose, thus closing the cycle which began with the early short pulse experiments.[89–91]

Whereas Tables 1 and 2 summarize experiments and developments of the beginning of resonant laser-MS, Table 3 presents recent and future tendencies in the development of resonant laser-MS which have a significant relevance to trace analysis. Numerous research groups (see Table 3) are using the technique in all kinds of trace analysis applications.

Supersonic beams play an increasingly important role in achieving highly selective trace analysis. This is true for pure spectroscopy and also for its combination with resonance MPI. Not a scientific technical step of development, but rather a very concrete field of application, is the analysis of combustion processes. In the process of combustion, hundreds of different components in the gas phase are produced, in significantly varying concentrations. Highly selective, sensitive, on-line detection methods are necessary to monitor these emissions. Resonant laser-MS fulfills these conditions, and has been applied to study flames, rough and flue gas in waste incineration plants, and exhaust emissions from motor cars.

Since femtosecond lasers have become commercially available, their use in resonant laser-MS has increased considerably. The hope is to achieve high ionization efficiencies also for problematic molecules such as nitroaromatics and metallo-organic compounds. Fast internal molecular processes mostly represent the reason for the problematic nature of these molecules. Another means of overcoming these problems involved is single-step photoionization by laser-induced vacuum UV light, putting up with reduced or even lost selectivity. A comparison of vacuum UV and femtosecond laser ionization, which likewise does not allow very high selectivity either, has been published.[131]

A third way to overcome these problems is to introduce a further parameter of selectivity, such as the retention time of a GC capillary, and detect fragment ions if dissociation processes in the neutral molecules are involved in the ionization process. Thus, for instance, nitroaromatics can be separated by the GC capillary while selective NO monitoring by resonant laser-MS may discriminate between all other trace compounds in the untreated sample. However, GC resonant laser-MS is of considerable interest as a method of extreme selectivity, as well as an extension of GC itself.

Another combination technique which represents a subject of increasing popularity within the community of laser-MS researchers and users is laser desorption resonant laser-MS. Although this has not achieved the same level of importance as MALDI (no separate ionization step there), it is much more appropriate for trace analysis and has been applied to many interesting problems. Examples are the comparison of aerosol particles near busy roads and from industrial areas (with regard to their PAH content)[152] and the detection of organic molecules in extraterrestrial (meteoric) particles.[153,154] Different variations of this technique have been studied, such as combination with supersonic beams, desorption from frozen samples and the comparison of thermal and laser desorption.

Although this article is not directly concerned with atomic trace analysis, we want to mention at least a few important topics in this respect. (This is not representative for atom detection by resonant laser ionization.) Since atoms usually show much higher ionization efficiencies than molecules, single-atom detection was discussed, studied and acclaimed at a very early stage. Extreme sensitivity and selectivity have been demonstrated for strontium and plutonium detection.[168,169] Not extreme trace analysis, but rather elemental analysis of solids and surfaces, is the goal of combined laser ablation and resonant laser ionization.[170,171]

4 APPLICATION TO TRACE ANALYSIS OF COMPLEX CHEMICAL MIXTURES

4.1 Application to Trace Analysis of Exhaust Emissions from Motor Cars: Dynamic Processes Behind the Combustion Chamber and in the Catalytic Converter

For all measurements presented in this section,[118,172,173] a medium-resolution linear time-of-flight mass analyzer was used. The ionizing laser wavelength was either 363 nm (for simultaneous NO and acetaldehyde monitoring) or 266 nm at which most aromatics but not other hydrocarbons absorb. In Figure 7, fast on-line measurements of the concentrations of acetaldehyde and nitrogen monoxide in the exhaust gas under rapidly changing motor conditions (speed and load) are shown. The laser repetition rate was 50 Hz, giving a maximum time resolution of 20 ms. Whereas NO is a product of total combustion of hydrocarbons, acetaldehyde production is due to incomplete combustion and is formed from ethanol contributions (important for alternative fuels). We found that products of total and partial combustion usually show the same temporal behavior when speed and load are synchronously changed (corresponding to a car moving with traffic). However, under special conditions (in the case of Figure 7a rapid decrease of speed after exceeding 3000 min^{-1}) a strong emission peak of acetaldehyde is observed with a delay of 300 ms (this delay time was similar for different measurements on the same motor). Together with other observations, this indicates that the acetaldehyde peak in Figure 7 is not due to direct internal

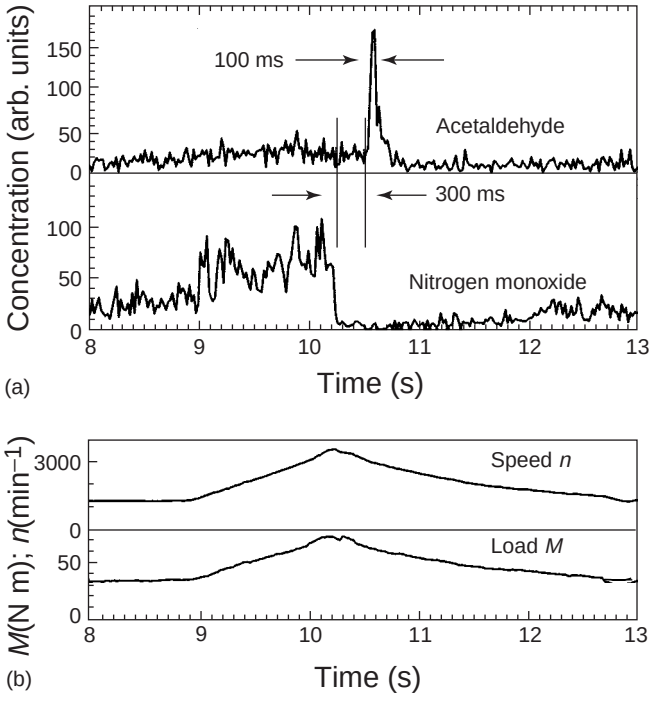

Figure 7 (a) Concentrations (arbitrary units) of acetaldehyde and NO measured synchronously to take into account rapidly changing engine speed and load (b). Note the sharp acetylene peak with a width of 100 ms and a delay compared with NO of 300 ms. This indicates the existence of two acetaldehyde formation processes, one occurring during the combustion process (low intensity) and one after the combustion process (peak). Both signals (for both acetaldehyde and NO) illustrate the high time resolution of resonant laser-MS.

combustion, but rather to chemical reactions at very hot surfaces of the exhaust pipe directly after the emission valve. In fact, combustion is interrupted in the motor type investigated if the speed has exceeded 3000 min^{-1} and then suddenly decreased below this value. This is a means of regulation initiated by the motor electronics for gasoline-saving reasons.

In Figure 8, two measurements of aromatic compounds are shown, one taken in front of (Figure 8a) and one behind (Figure 8b) the catalytic converter of a motor car, at an industrial test facility. In particular, the concentrations of benzene, toluene, xylene, trimethylbenzene, and naphthalene were observed synchronously during a so-called FTP (federal test program) cycle. This is a program of well defined variations in velocity (see Figure 8c) for testing pollutant emissions from motorized vehicles and has its origins in governmental regulations in the USA. In the measurement made in front of the catalytic converter, numerous short peaks of emission (of up to 300 ppm and more) for all observed aromatics appear, which show a more or less constant ratio for the different compounds.

The contribution of these short emission peaks to the overall sum of emitted aromatics is considerable. A reduction of this ''dynamic'' part of the emission should reduce the concentrations of aromatics produced in the automobile exhaust significantly, at least in the case of city traffic. From other highly time-resolved measurements, we deduce that the origin of these peaks is to be found in motor-controlling effects. For instance, strong fluctuating pollutant concentrations have been observed when the oxygen content in the fuel mixture of regulated motors is subject to rapid changes.

The measurements behind the catalytic converter reveal a significant reduction in pollutant concentrations as expected. The emitted concentrations lie far below 10 ppm after a warming-up time of the converter of 200 s. It is interesting to note that in the first 10 s near-zero emissions are observed even in the cold converter, while the measurements in front of the catalytic converter (Figure 8a) show exceptionally high emissions during the first few seconds (corresponding to the starting phase of the engine). This effect may be explained by adsorption of the aromatics on the cold catalytic surfaces. After some delay, these surfaces are warm enough for the adsorbed aromatics to evaporate (see the emission peak after a delay time of >10 s in Figure 8b). The molecule-dependent delay times (the larger the molecule the larger the delay) observed in spectra with better time resolution support this explanation. Far more astonishing, however, is the appearance of emission peaks of benzene even for a hot catalytic converter, i.e. after 700 and 1200 s. In the latter case, the concentration even reaches a value of more than 100 ppm. In laboratory experiments, we found that under special conditions (i.e. high hydrogen or water, low oxygen concentrations) dealkylation takes place along with efficient production of benzene from alkylbenzenes. Even the formation of naphthalene has been observed. The assumption that these processes are the reason for the high benzene peaks is supported by the absence of corresponding significant peaks for the alkylated benzenes in Figure 8(b).

These few examples demonstrate that analytical techniques with high time resolution in the 10 ms range are necessary to detect and study fast dynamic effects in vehicle exhaust emissions and to improve the electronic control of motorized vehicles. Understanding the origins of pollutants both during and after combustion phases may help to reduce the pollutant concentrations in the exhaust; understanding the dynamic effects of electronic control may even help to reduce the total consumption of fuel. REMPI mass spectrometry is able to supply the necessary time resolution for such on-line studies and works for different trace compounds, e.g. aromatics, aldehydes, or typical catalytic converter products (e.g. HCS, CS_2, OCS).

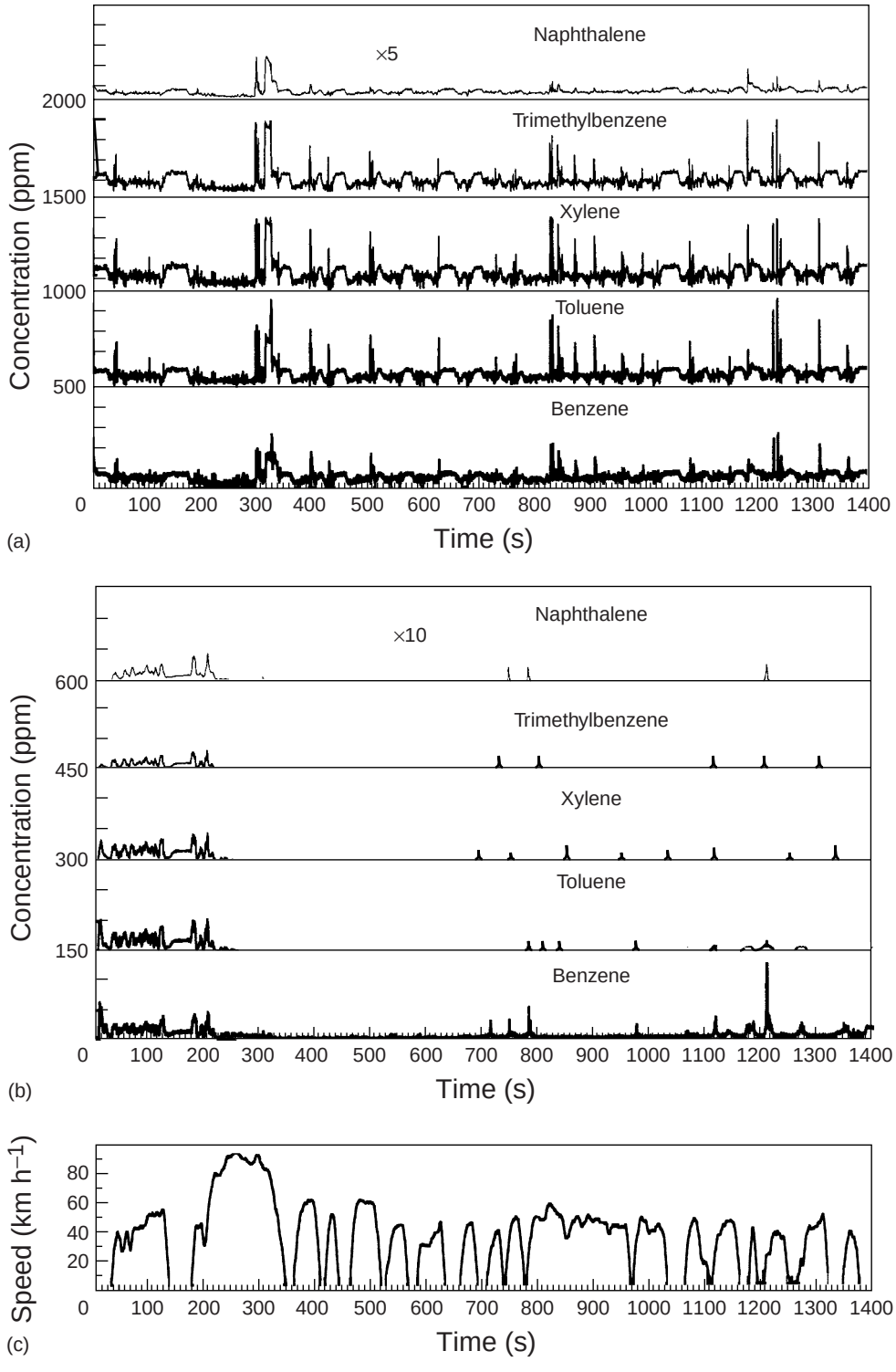

Figure 8 Concentration (ppm) of aromatics in the exhaust emission of a motor car (industrial test facility) measured (a) in front of and (b) behind the catalytic converter during a fixed-velocity program [FTP cycle (c)]. The rapid changes in motor operating conditions (speed, load, changing of gear) induce significant peaks of pollutant emission. Even within the catalytic converter, unfavorable chemical gas conditions may result in the formation and emission of peaks of aromatics (benzene, naphthalene) instead of preventing them.

4.2 Application to Trace Analysis of Exhaust Emissions from Motor Cars: Traces of Motor Oil in the Exhaust

In section 3, aromatic compounds originating in the fuel and appearing in the exhaust of combustion engines were investigated. However, motor oil too is a source of aromatic species in the exhaust, these being on average of larger molecular size. The first experiments to detect these motor oil-based emissions[174] were performed on a hydrogen motor (a conventional combustion engine with minor changes to permit use of hydrogen as a fuel). All aromatics detected in the exhaust could therefore only be due to the motor oil. In Figure 9, two mass spectra are presented which were measured in the exhaust of a hydrogen motor, under two different sets of operating conditions. At the higher speed but lower load, the smaller masses up to 170 (mostly due to alkylated naphthalenes) are enhanced in contrast to the situation with lower speed but higher load. The overall concentration of all aromatics determined by conventional unselective methods was about 1 ppm. A sensitivity of less than 10 ppb for most of these aromatics can be deduced from this value along with the intensity distribution presented in Figure 9. Again, a small, medium-resolution linear time-of-flight spectrometer and a semiselective laser wavelength of 266 nm are sufficient for this analysis.

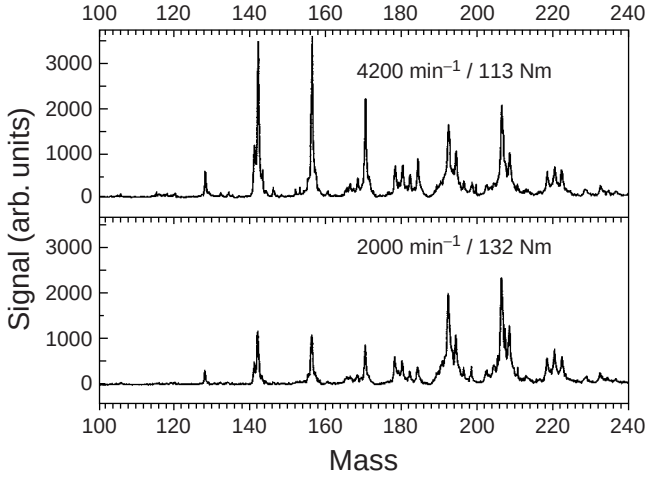

Figure 9 Aromatic compounds in the exhaust emissions of a combustion engine (hydrogen motor) resulting mainly from emissions from the motor oil film coating the combustion chamber. The mass spectra were measured using two different sets of parameters (speed, load) revealing significant variations in the relative concentrations of masses between 120 and 170 and between 170 and 240.

In Figure 10, the attempt to perform the same measurement on a combustion engine working with conventional gasoline as fuel is presented. From the bottom trace, one realizes that above a mass of 150 there is practically no signal, whereas between mass 90 and 130 exceptionally large peaks occur, which obviously saturate the ion detector, and represent unburned compounds from the fuel. By using a pulsed mass filter (pulsed deflection plates located after the ion source in the field-free drift region), this

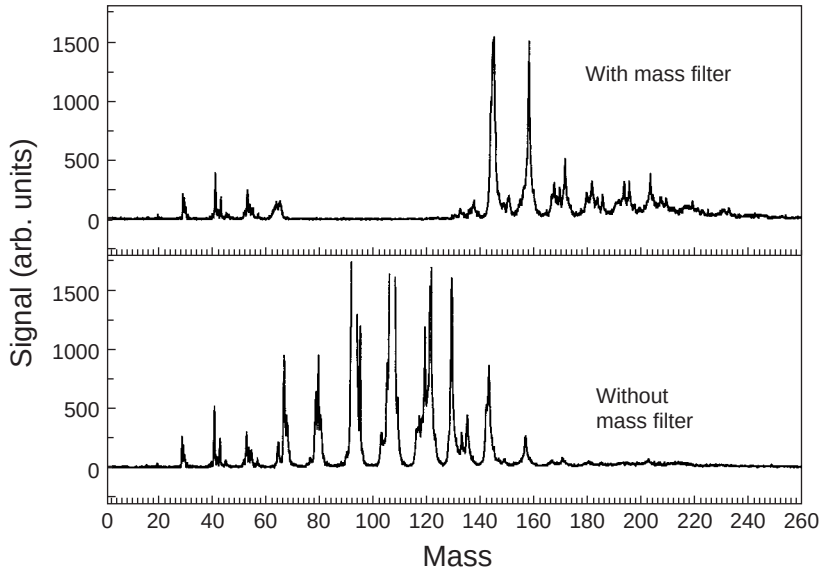

Figure 10 Aromatic compounds in the exhaust of a combustion engine (conventional gasoline). Lower trace: unburned aromatic compounds from the fuel give rise to such high peaks that the ion detector saturates and higher masses disappear within the mass spectrum. Upper trace: by means of a pulsed mass gate, masses in the range from 60 to 130 are deflected, therefore do not reach the ion detector, and thus are eliminated from the mass spectrum; now, higher mass peaks appear. Good correlation with the peaks in Figure 9 and with experiments involving the introduction of tracer molecules into the motor oil, which are not present in the fuel, allow for an assignment of these higher mass peaks to emissions from the motor oil film coating the combustion chamber.

range of masses has been suppressed in the upper mass spectrum. Now, masses above 130 occur which correlate well with the mass peaks of Figure 9 and are classified as ''motor-oil'' peaks. This has been verified by measurements carried out under different motor conditions, by comparison with conventional methods under static motor conditions, and by the addition of tracer molecules to the motor oil which are not present in the fuel.

Besides information about the mechanism of oil consumption, these results can be used to determine the dynamic behavior of oil consumption, or to obtain quantitative values for it under many different static motor conditions, e.g. characterized by load and speed. Conventionally, oil consumption is measured by weighing, which is extremely time consuming and very expensive if performed for a number of pairs of parameters (i.e. load and speed). In Figure 11, a measurement of motor oil consumption is presented, which was performed by means of resonant laser-MS within the space of one day. Similar conventional measurements would take many weeks to be performed. The interpretation of Figure 11 is still in progress. Some effects such as oscillations for constant load but increasing speed indicate rotation and twisting of the piston rings which control the oil film on the inner wall of the combustion chamber. Other effects, e.g. the decrease in motor oil consumption at constant (high) speed and at high increasing load, indicate cracking of the oil compounds at high temperatures. It has been found (in measurements shown here) that under most motor conditions more than 90% of emitted oil compounds leave the combustion chamber unburned.

The examples given in this section illustrate that fast analytical methods represent a purposeful or even necessary approach not only from the point of view fast fluctuating concentrations but also from that of a large number of ''data points''.

4.3 Application to Trace Analysis of Exhaust Emissions from Industrial Incinerators

Combustion processes causing air pollution are not present in combustion engines alone, they also occur on an industrial level, e.g. in the case of municipal incinerators. Gas cleaning after combustion is the most expensive element in such plants, and any degree of reduction

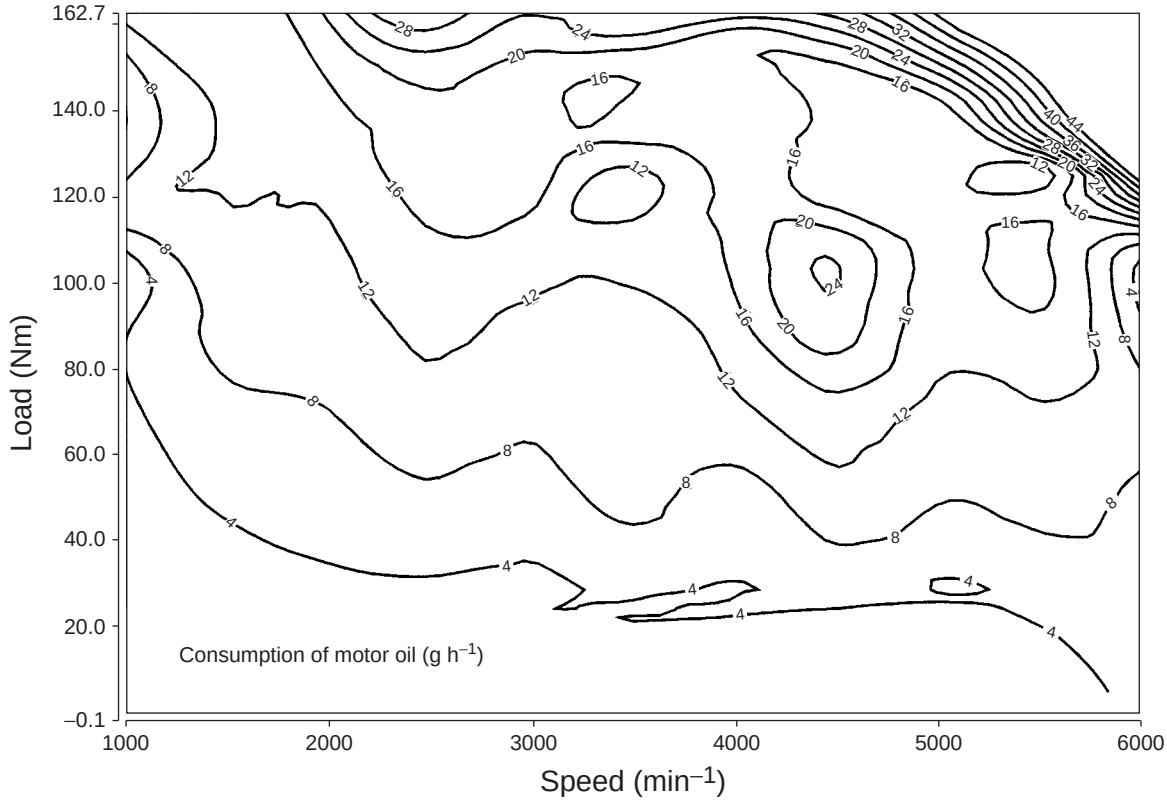

Figure 11 Measurement of the consumption of motor oil for the different pairs of parameters load (N m) and speed (min^{-1}). Lines are shown for constant oil consumption ($g\,h^{-1}$). Notice the significant variations which occur at constant load or speed in certain parts of this figure; this allows conclusions to be drawn regarding the actual mechanical and chemical processes at work during and after combustion.

of these costs would be economically very productive. One way of achieving such economies is to monitor the combustion products and minimize pollutants produced (e.g. due to malfunction) by feedback control.

We performed measurements of the flue gas directly from the second combustion chamber of a pilot plant for waste incineration with 1 MW heating power.[116,117] One goal of the pilot project was to optimize overall combustion by transition from a precombustion on the grate of the first combustion chamber to a gasification process. This step was to allow more complete combustion in the second chamber. Monitoring of the combustion products was to help one to perform this transition in an optimum way, to confirm the improved combustion conditions, and permit the carrying out of a control of the whole plant. The first such measurements are presented below. Figure 12, shows a mass spectrum taken at the second combustion chamber, with gas temperatures of 900 K. Details of the gas sampling can be found elsewhere.[116,117] The combustion conditions were fairly unfavorable, giving rise to many different compounds in significant concentrations. In fact, this mass spectrum is a snapshot of trace concentrations which develop differently over time. For a more complete view of the situation one has to study concentration–time profiles.

Some such profiles are presented in Figure 13. They were obtained by centering time windows at the flight time of selected masses. In Figure 13, the masses 78 (benzene), 128 (naphthalene), 170 (alkylated naphthalenes), and 178 (phenanthrene) have been selected for presentation. These mass signals provide illustrative examples of how single species behave differently over time and under changing combustion conditions, e.g. the transition from combustion to gasification. This transition is indicated by three short, small concentration peaks of masses of 78 and 128 while the higher masses (170, 178) do not show corresponding fluctuations. In fact, phenanthrene even exhibits a significant but constant signal over the whole time, from the beginning of the measurement up to 500 s. At 500 s a malfunction in the post-combustion occurred, inducing a very strong increase in masses 78 and 128, strong oscillations in mass 170, and a remarkable decrease in phenanthrene production which lasted for more than 100 s. It took more than 300 s for the combustion to become stabilized again, this showing up as a decrease in masses 78 and 128 and a delayed decrease in mass 170. The behavior of mass 178 should be noted. At the supposed end of the malfunction near 800 s, the phenanthrene signal increased significantly and became the dominant signal in the mass spectrum (see Figure 12) for a further 800 s.

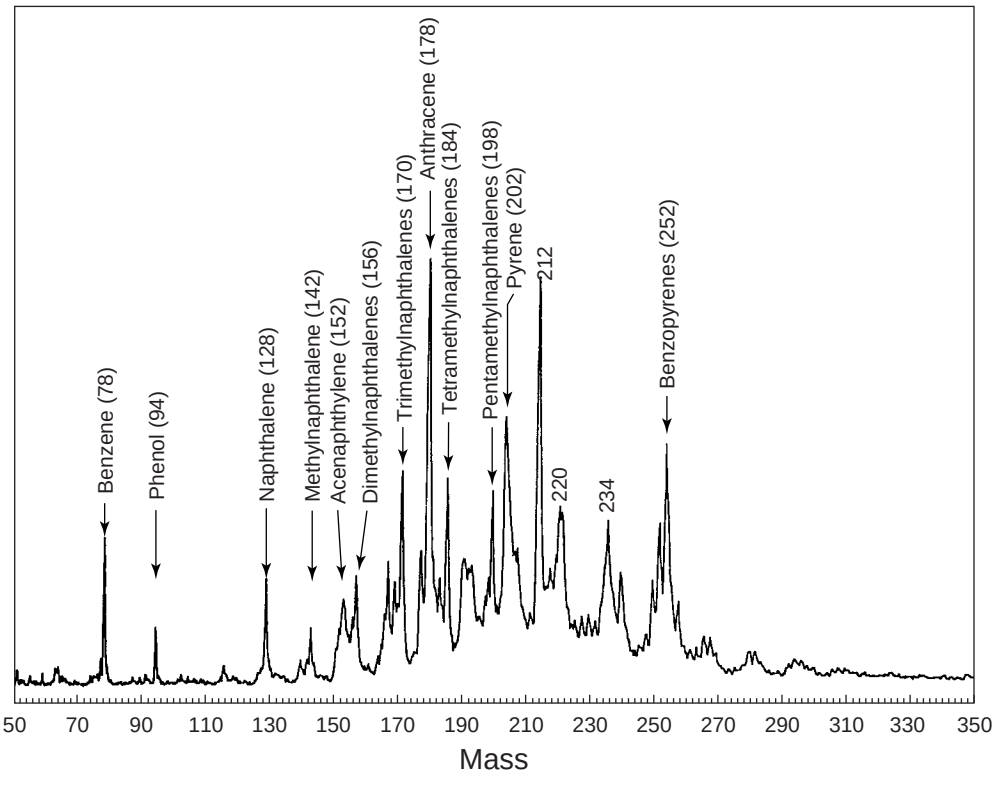

Figure 12 On-line PAH-selective mass spectrum for the flue gas of a waste incinerator obtained using a resonant laser-MS instrument with a ionization wavelength of 248 nm (excimer laser). The sampling point was located directly at the postcombustion chamber, the gas temperature being 900 K. The recording time (and time resolution) was 0.5 s.

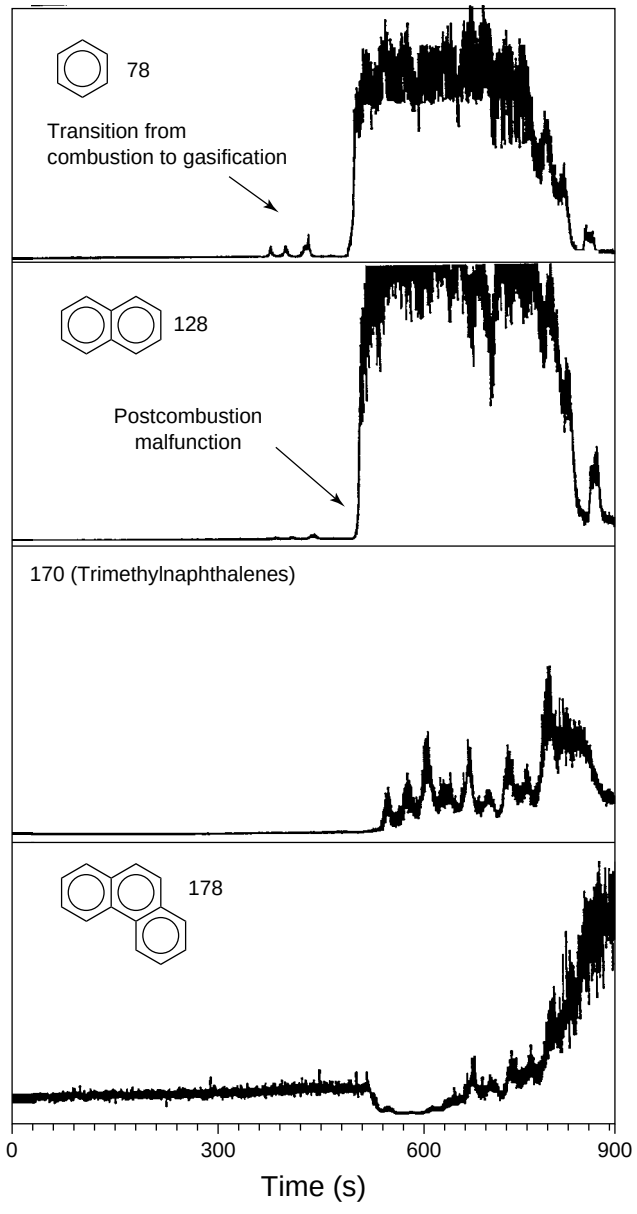

Figure 13 Concentration profiles over time for selected masses (78, 128, 170, 178). Selection was performed using time windows within the time-of-flight mass spectrum. Instability in combustion between 350 and 450 s and the occurrence of a malfunction between 500 and 800 s are indicated by the variations in masses 78 and 128. Both of the larger masses show significantly different behavior.

In Figure 14, the same situation is shown for different species (besides naphthalene) and using the optimum sensitivity setting of the data acquisition system. It should be noted that the concentration scale is now calibrated and shows quantitative values for the concentrations. The spectrum for naphthalene is magnified (ppt) and in this form reveals a slight increase in concentration by about 100 ppt at 200 s. This means that the transition from combustion to gasification started considerably earlier. The three peaks obviously indicate a significant degree of instability in the combustion condition before a serious malfunction occurred. While benzene and naphthalene behave similarly to each other, obviously the fluctuations in fluorene, dibenzofuran, and alkylnaphthalenes also show common features, although their relative concentrations at between 500 and 700 s and between 700 and 900 s are different. However, the remarkable fluctuation frequency, with a time constant of 30 s, indicates correlated processes of formation. A totally different process seems to be responsible for the formation of phenanthrene and larger polycyclic aromatics.

This kind of time-resolved trace analysis may therefore (in addition to its application as a controlling tool for

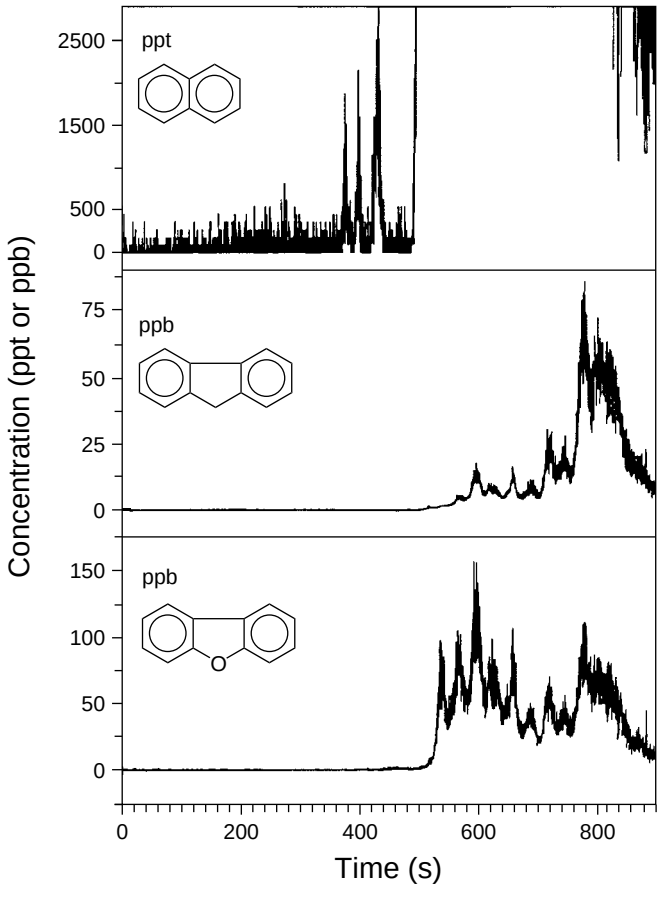

Figure 14 Concentration profiles over time for selected masses (128, 166, 168) measured synchronously with those shown in Figure 13. The concentrations are calibrated (in ppt for naphthalene and ppb for fluorene and dibenzofuran). Masses 166 and 168 show similar, but not equal, variations compared with mass 170 given in Figure 13. Naphthalene (now on a significantly expanded concentration scale) exhibits variations in concentrations (in addition to those already mentioned in the caption of Figure 13) even before the elapse of 300 s within the 100 ppt range.

environmental protection) be of help in understanding complicated combustion processes. The development of motors emitting less pollutants or of plants based on combustion processes is at least as important as emission control carried out on such motors and plants. In both cases, adequate (i.e. fast, selective, and sufficiently sensitive) analytical tools are required.

4.4 Application to Further Analytical Problems

Resonant laser-MS is usable in many other fields of application than just those of combustion processes. Three of these are presented in the following subsections.

4.4.1 Food Processing

The possibility of carrying out rapid controls would represent a very interesting, but often not yet feasible, option for the food processing industry. For example, the degree of roasting determines the quality and character of coffee. Many compounds which are responsible for the aroma of coffee are formed during the roasting process, while others already exist in the coffee bean. These compounds are subject to very different temporal changes in their concentrations. The ability to control their relative concentrations would supply an excellent means of regulating the roasting process on-line. This has been experimentally tested using resonant laser-MS, in a model roasting arrangement, with very promising success.[175]

4.4.2 Gas Chromatography Coupling

The use of resonant laser-MS also opens up new possibilities for GC (as reported in section 3), giving rise to a three-dimensional technique. The combination GC and TOFMS has great advantages, since both techniques rely on selection by means of time. Gas pulses of species separated in time by the GC capillary can be mass analyzed with a high time resolution in the time-of-flight instrument. Species of different mass which have not yet been resolved by GC can then be discriminated. Species-selective ionization even allows one to distinguish between species of the same mass which are not separated by GC.

An example is shown in Figure 15. A gas chromatogram of the gas phase above a gasoline sample is shown. The coupling of the GC capillary and a supersonic beam valve provided the possibility of isomer selection. A fairly short, medium-resolution GC capillary was

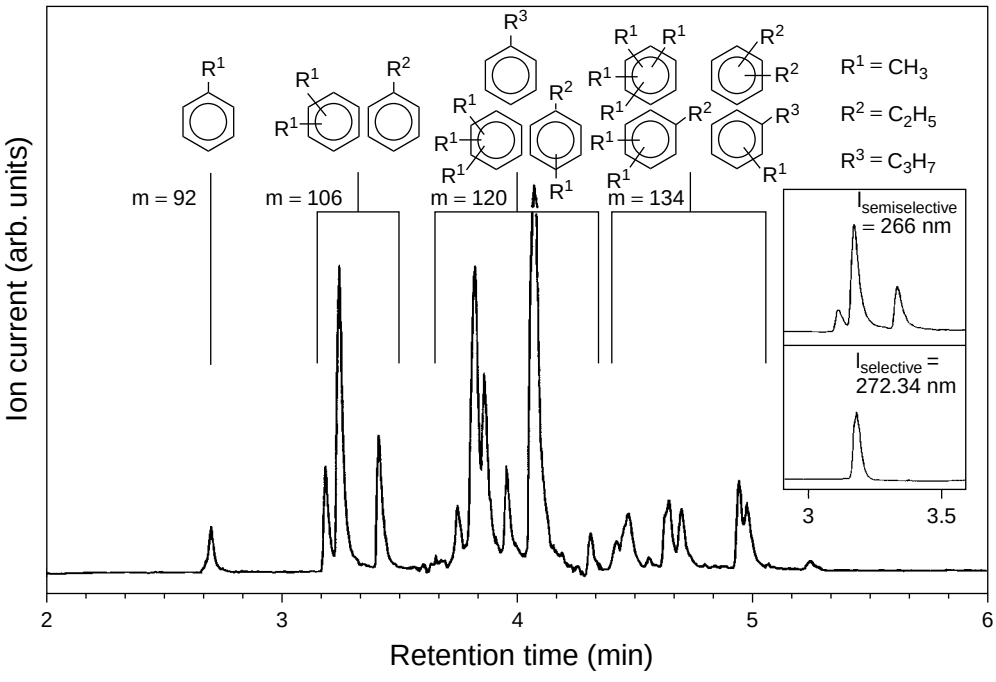

Figure 15 Detection of alkylated benzenes emitted from a gasoline sample by means of a combination of GC and resonant laser-MS. The spectrum shows the GC signal as a function of GC retention time, detected by means of semiselective ionization at 266 nm and a linear time-of-flight mass analyzer. During the GC run, a time window was transferred from mass 92 to 106, 120 and 134. The possibly present isomers with these masses (1 × at 92, 4 × at 106, 7 × at 120, and 15 × at 134) are displayed at the top of the figure. The GC capillary used cannot, however, separate all of these isomers, e.g. *m*- and *p*-xylene (the highest of the $m = 106$ peaks). Inset: selective ionization allows for further discrimination of isomers. At the laser wavelength 272.32 nm, only *p*-xylene was ionized and this appears in the $m = 106$ mass window; the other isomers ethylbenzene and, *m*- and *o*-xylene have been suppressed.

used, thus providing short retention times. During the GC run, time-of-flight mass spectra were continuously measured and the signal within a preset time window was recorded. This time window was shifted during the scan from mass 92 to mass 106, 120 and 134. An ionizing laser wavelength of 266 nm was used, resulting in semiselective ionization of aromatic molecules. As a result, a well-structured gas chromatogram is obtained from a very complex multi-component gas sample. The combination of all three methodical components (GC, REMPI, and TOFMS) obviously results in a selective, but still reasonably fast and sensitive device, requiring only low levels of effort for chemical sample preparation (these being significantly reduced for a fast, low-resolution GC capillary). To demonstrate the high level of selectivity achieved (e.g. for discrimination of the different isomers with mass 106: ethylbenzene and *o*-, *m*-, and *p*-xylene) with this instrumental set-up, ionization was performed with a semiselective (266 nm) and an isomer-selective wavelength (272.34 nm). The results are presented in the inset in Figure 15. *p*-Xylene is obviously selectively ionized at 272.34 nm, and at a mass of 106 it is the only sample component recorded out of the whole, unprepared gas sample.

4.4.3 Laser Desorption Coupling

Up to now, only gaseous samples have been considered. In many situations, however, samples in the condensed phase have to be analyzed for pollutants. If the sample is solid, laser vaporization (desorption of neutrals) may be the connecting element between solid sample and the gas phase necessary for mass spectrometry. Many groups are working in this field (see section 3; one of the most active is the research group of Zenobi et al). Nevertheless, laser desorption resonant laser ionization mass spectrometry as a method of trace analysis of solids and surfaces is still in the developmental stage. To address just one important example of this, sparse experience only exists for such complicated types of sample as soil, and considerable research efforts are still necessary. The need for a fast, but nevertheless selective and sensitive tool for the analysis of soil is obvious, if one considers such cases where larger areas of polluted soil have to be monitored within a dense grid of testing points.

In Figure 16(a), a schematic diagram of a set-up for the combination of laser desorption of neutral molecules and resonant laser-MS is shown along with an exemplary mass spectrum. The probe consists of a rod which can be moved continuously by a motor and gear assembly. The top of this rod is placed very near ($\Delta x < 1$ mm) to the nozzle of the pulsed supersonic beam valve. A desorbing laser pulse is focused in this region. Neutral desorbed molecules are emitted into the supersonic beam,

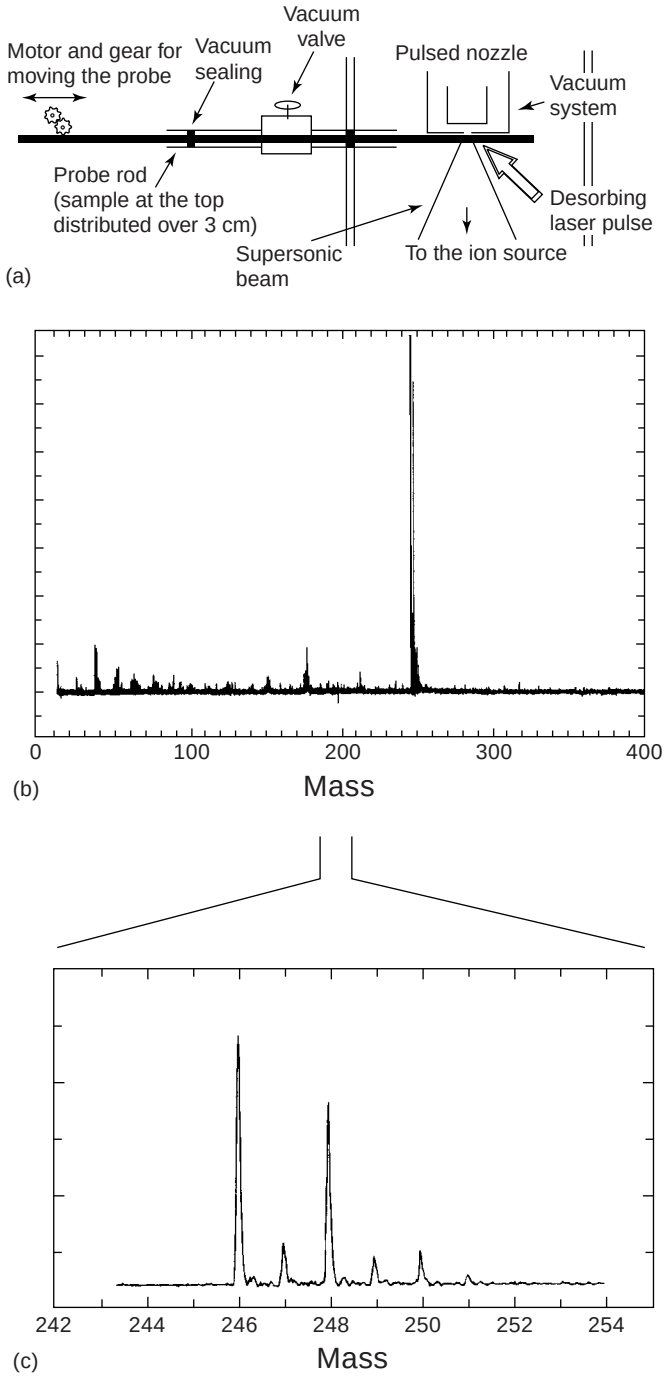

Figure 16 (a) Experimental set-up for the coupling of neutral laser desorption of trace molecules from solid samples with resonant laser-MS. The sample is positioned within the top three centimeters of a continuously moved rod. (b) Mass spectrum (reflectron TOFMS) of laser-desorbed and laser-ionized (266 nm) dichloroanthracene. Note the low degree of fragmentation. (c) Mass region of the molecular ion displaying the $^{13}C_0$, $^{13}C_1$, as well as the $^{37}Cl_0$, $^{37}Cl_1$ and $^{37}Cl_2$ isotopomers.

cooled by a large number of gas collisions and finally drift together with the gas pulse into the ion source

where the resonant ionization takes place. Figure 16(b) shows the mass spectrum of laser-desorbed and laser-ionized (266 nm) dichloroanthracene. Figure 16(c) shows the molecular ion region displaying the different ^{13}C and ^{37}Cl isotopomers of this molecules in its natural isotopic mix.

5 FURTHER SPECIFICATIONS OF RESONANT LASER MASS SPECTROMETRY IN TRACE ANALYSIS

5.1 Calibration Methods and Detection Limits

Multiphoton processes are nonlinear optical processes. The ion yields are therefore correlated with laser intensities in a highly nonlinear way. This dependence changes when one of the single absorption steps becomes saturated at high laser intensities. In addition, fast intramolecular processes occurring in the excited intermediate molecular state may influence the ion yield and its dependence on laser intensity and pulse length. Even though this complicated behavior can be derived theoretically by rate equations,[34] absolute ion yields cannot be predicted in a satisfactory way, since mostly single absorption cross-sections and intramolecular rate constants are not known. Quantification can be performed by measuring absolute cross-sections.[65] The problem then is to transfer these results to the individual measurements (variations in gas density, laser intensity, focus geometry, charge density, etc.). Such measurements may help to estimate ion yields and thus also concentrations, but are usually not appropriate for reliable quantification.

An adequate calibration procedure is therefore important for analytical applications. For quantification, one can rely on the linear dependence of the REMPI–TOFMS signal on the concentration of a standard. This has been proven for many orders of magnitude down to the pptv concentration range.[113,115,138] In general one may discriminate in quantification of analytical results by means of off-line comparison to an external calibration standard and quantification by on-line addition of an internal calibration standard. In the first case (off-line) an analyte standard can be used with a well-defined standard sampling. The standard signal is not disturbed by interfering mass peaks and the integrated peak intensity is well related to the concentration. However, external calibration standards can be used occasionally only. Long-term variations may be checked by means of external calibration before and after a measurement. In Figure 14, the external calibration method has been applied using ppb standards prepared dynamically with permeation and diffusion devices.[176,177]

Using on-line calibration with an internal standard, the short-term variations produced by the instrument (e.g. laser intensity and spatial laser beam profile) can also be compensated for. This is performed either by adding a non-analyte calibration gas[178] or by using a standard addition method with alternating measurement of the sample and of the sample with an additional standard.[178] The first method permits the carrying out of rapid measurements resolving single shot-to-shot fluctuations, but the possible differences in the ionization yield of analyte and non-analyte molecules have to be considered. The second method uses an analyte standard; here, every second shot has to be used for calibration purposes. Shot-to-shot variations of the analyte signal or interfering substances may disturb the differential calibration.

All components which are available as standards may be quantified by these calibration and quantification techniques. For all other components the relative REMPI cross-sections between trace component and standard molecule have to be determined. These relative ionization yields are much more reliable than the quantitative ones (as mentioned above), if trace species and standard species are of similar molecular type. Measurements of relative cross-sections have been performed for a number of PAHs (naphthalene, biphenyl, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, and pyrene) using GC/laser-MS at two different laser wavelengths.[33,179] GC/laser-MS is well adapted to the problem of relative quantification, since it allows for introduction of a well-defined quantity of the sample molecule into the mass spectrometer at well-defined time intervals (widths of the GC peaks). A different method proposed for relative quantification[180] involves laser desorption. Specially prepared thin polymer samples are used with homogeneous and well-defined trace concentrations. In response to a defined number of laser shots, holes (volume defined by laser focus and thickness of film) are ablated out of this thin film, which then releases well-defined quantities of trace molecules. Seventeen PAHs have been investigated by means of this method in the wavelength range 218–310 nm.

The detection limits of resonant laser-MS have been studied by several research groups, mostly employing supersonic molecular beams and tunable laser wavelengths. In the following, all detection limits refer to a signal-to-noise ratio of 1. Cool et al. have investigated the detection limits for numerous toxic combustion byproducts.[113,181] Detection limits in the low-ppb (5 ppb for naphthalene at 302 nm; however, see comment below) and 10-ppt ranges (50 ppt for toluene at 267 nm) have been achieved for highly selective ionization of aromatic compounds using the first excited singlet state of the jet-cooled target molecules. With improved free-jet setups reducing the distance between nozzle and ionization

region and using different, less selective REMPI transitions, detection limits in the low-ppt range have been reported.[111,114]

As for resonant laser-MS with an effusive molecular beam and fixed-frequency laser ionization (see Figures 7–14), sensitivity has been determined on the basis of a mass spectrum for a naphthalene permeation standard. A detection limit of 90 ppt has been achieved for naphthalene at 266 nm. This is more favorable than the results of Cool and et al. for naphthalene. The reason for this is not an intrinsically lower level of sensitivity in the supersonic beam arrangements (the example of toluene proves high sensitivity), but rather that 302 nm is an unfavorable ionization wavelength for (1 + 1) MPI, by contrast with 266 nm. The 302 nm wavelength has been used to reach the UV spectral range of high selectivity (e.g. for isomers), while 266 nm only allows medium selectivity (e.g. for substance class-selective ionization) where, owing to overlapping broad bands in the UV spectrum, PAHs may equally be excited and ionized. The problem of the low ionization efficiency of naphthalene at 302 nm can be solved by using a two-color ionization, with the first absorption step at 302 nm and the second (ionization) step at 266 nm. In a Nd/YAG (laser active crystalline medium) laser pumped dye or optical parametric oscillator (OPO) (optical principle for tunable solid-state laser) laser, ample intensity is available from the pump laser at 266 nm.

5.2 Specific Features of Resonant Laser Mass Spectrometry

In this last section, specific features and options of resonant laser-MS and also drawbacks will be discussed.

5.2.1 Compactness and Mobility

The laser mass spectrometers used for practical and realistic applications (e.g. analysis of combustion engine exhaust and municipal incinerator emissions) already had very reasonable dimensions of typically 1.20 m (length) × 1.00 m (width) × 1.20 m (height). The laser power supply and personal computer were located elsewhere (e.g. in the controlling room of the motor test facility). Both instruments were fitted with wheels; on one of the instruments hooks for lifting by crane and placing in special positions were even mounted. No time-consuming adjustment (in several cases even none at all) was necessary after transportation in a van for an entire day. For automobile exhaust analysis a commercially available instrument already exists; for incinerator measurements a prototype instrument has been developed at the Technical University, Munich.

5.2.2 No or Minimum Sample Preparation

For gas-phase analysis, no sample preparation or sample accumulation is necessary and thus real on-line measurement is possible. Of course, the gas inlet system has to be designed carefully to avoid memory effects. This is achieved by allowing for an overall heating temperature of the inlet system of 500–600 K. The efforts involved in the preparation of solid samples for laser desorption resonant laser-MS are mainly concerned with the physical consistency of the sample. A chemical separation of abundant or troublesome compounds is not necessary, since the level of selectivity of resonant laser-MS is already sufficiently good. For instance, high levels of water content, which can represent a serious problem in conventional analytical methods, or large concentrations of oxygen, which may lead to the burning out of electron-emitting wires, do not affect resonant laser ionization at all.

5.2.3 High Speed of Recording of Mass Spectra

The high speed achieved in connection with selectivity is one of the central features of resonant laser-MS. It is limited by the time characteristics of the gas inlet system and the repetition rate of the ionizing laser. A real time resolution of 20 ms is easily achievable (50 Hz lasers). However, in many cases a time resolution of 1 s is perfectly sufficient. This allows an averaging over 50 mass spectra per measurement (using 50 Hz lasers) and thus better reproducibility and signal-to-noise ratio.

5.2.4 High Selectivity

Owing to its two-dimensionality (wavelength and mass), the selectivity of resonant laser-MS is intrinsically good. Depending on the analytical problem to be dealt with, one can choose between an instrument with a good resolution (one or few fixed wavelengths and linear time-of-flight mass spectrometer) with reduced costs and dimensions and a very high-resolution instrument involving supersonic molecular beams, tunable lasers and reflectron mass analysis. In the first case, groups of molecular species (e.g. aromatics) can be discriminated from other species (mostly hydrocarbons) during the ionization process and detected mass selectively, by means of time-of-flight separation. In the second case, structural isomers (e.g. isobaric conformers of chlorinated aromatics) or even isotopomers involving isotope substitution in chemically different positions (e.g. ^{13}C in the aromatic ring or the methyl group of xylene) can be selectively detected.

5.2.5 High Sensitivity

The detection limit depends significantly on the type of multiphoton process and the cross-section absorption

processes involved. For most components detection can be achieved at least down to the ppm region, for many aromatics ppb sensitivity could be achieved even in very complex mixtures and in some experiments the ppt limit was even reached.

5.2.6 Applicable to Many Different Compounds

The range of different chemical compounds detectable by resonant laser-MS has been illustrated in studies of more than 25 representative gaseous components of the exhaust emissions from combustion engines,[118] in which different wavelength ranges and multiphoton absorption and ionization schemes were investigated. These representative components include substances from different classes of chemicals, such as alkenes, carbonyls (in particular aldehydes), aromatic hydrocarbons, nitrogen and carbon oxides or catalytic converter products (e.g. HCN, NH_3, H_2S, OCS). Additionally, the resonance-enhanced (and thus selective) MPI of innumerable individual chemical compounds has been successfully achieved and described in the literature (see section 3).

Of course, resonant laser-MS also has drawbacks. One of these is that particular species can only be ionized by REMPI with very low or diminishing efficiency. Another problem concerns the commercial availability of user-friendly large laser systems.

5.2.7 Ionization Yields?

One restriction on resonant laser ionization is the inavailability of tunable, easy to handle vacuum-UV laser sources. For this reason, molecules whose lowest excited electronic states lie far into the UV range (e.g. the alkanes) have to be ionized using multiphoton processes of higher order, which are less efficient. Furthermore, in some molecules resonant intermediate states are extremely short lived due to the occurrence of fast electronic relaxation or even dissociation. In these cases no molecular ion becomes visible in the mass spectrum. A typical example of this kind of behavior is to be found in nitroaromatics. Sometimes, the fragments (here NO) are nevertheless sufficient to permit discrimination and selective or partially selective detection. If not, other intermediate states or multiphoton processes have to be selected or ultrafast laser pulses (femtosecond laser) can be used.[119–122] These solutions tend to be expensive or complicate the whole analytical method. In addition, selectivity may be reduced considerably. An alternative, less expensive approach could be provided by the combination using a short GC capillary for the separation of the problematic species without paying any attention to the separation from all other compounds. This is achieved by resonant laser-MS.[135]

5.2.8 Customer-friendly Laser Systems?

Large pulsed laser systems in general, and lasers with tunable wavelength in particular, are still not well accepted by industrial users. In particular, pulsed, tunable lasers have not yet reached a state of development which permits them to be handled in an uncomplicated and convenient way. This is, however, more a question of supply and demand than one of technical state-of-the-art. The consequence of this is that high-resolution instruments with isomer selectivity may not get beyond a prototype stage in the near future. Resonant laser-MS instruments with reduced performance (i.e. few or even only one fixed laser wavelength) can be and have indeed already been commercialized. Even additional options such as high mass resolution achieved with reflectron mass spectrometry or additional selectivity achieved by coupling with a GC capillary could be incorporated in such commercial instruments to permit fast trace analysis.

ACKNOWLEDGMENTS

The examples presented here received the support of the German Ministry for Education, Science, Research and Technology (BMBF), the German automobile industry (FVV), and the Fonds der Chemischen Industrie (FCI). The authors would like to thank Prof. E.W. Schlag (TU Munich) and Prof. A. Kettrup (GSF Oberschleissheim) for their continuous interest and support in relation to many projects reported on in this article. We are indebted to Prof. M. Beckmann and Dr H. Griebel (both from the CUTEC Institut Clausthal Zellerfeld), to Dr Thiel and Dr Fröchtenicht (both BMW AG), to Prof. H.E. Rikeit (Mercedes Benz AG), and to Dr Franzen and Dr R. Frey (both from Bruker Franzen Analytik GmbH) for many fruitful discussions and opportunities to perform resonant laser-MS analysis at industrial facilities.

ABBREVIATIONS AND ACRONYMS

GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
IR	Infrared
laser-MS	Laser Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/ Ionization
MPI	Multiphoton-ionization
Nd/YAG	Neodymium–Yttrium-Aluminum Garnet
OPO	Optical Parametric Oscillator
PAH	Polycyclic Aromatic Hydrocarbon

REMPI	Resonance-enhanced Multiphoton Ionization
TOFMS	Time-of-Flight Mass Spectrometry
UV	Ultraviolet

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Automotive Emissions Analysis with Spectroscopic Techniques • Laser- and Optical-based Techniques for the Detection of Explosives

Environment: Water and Waste **(Volume 3)**
Dioxin-like Compounds, Screening Assays

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction

Mass Spectrometry **(Volume 13)**
Time-of-flight Mass Spectrometry

REFERENCES

1. U. Bahr, M. Karas, F. Hillenkamp, 'Analysis of Biopolymers by Matrix-assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **348**, 783–791 (1994).
2. R. Kaufmann, D. Kirsch, B. Spengler, 'Sequencing of Peptides in a Time-of-Flight Mass Spectrometer: Evaluation of Postsource Decay Following Matrix-assisted Laser Desorption Ionization (MALDI)', *Int. J. Mass Spectrom. Ion Proc.*, **131**, 355–385 (1994).
3. J.S. Becker, R.S. Soman, T. Becker, V.K. Panday, H.J. Dietze, 'Trace and Ultratrace Analysis of Gallium Arsenide by Different Mass Spectrometric Techniques', *J. Anal. At. Spectrom.*, **13**, 983–987 (1998).
4. C.H. Becker, 'Surface Analysis by Laser Ionization', *Curr. Opin. Solid State Mater. Sci.*, **1**, 493–497 (1996).
5. U. Boesl, H.J. Neusser, E.W. Schlag, 'Two-photon Ionization of Polyatomic Molecules in a Mass Spectrometer', *Z. Naturforsch., Teil A*, **33**, 1546–1548 (1978).
6. U. Boesl, H.J. Neusser, E.W. Schlag, 'Mass Selective Two-photon Ionization of a Polyatomic Molecule', in *Laser Induced Processes in Molecules*, Springer Series of Chemical Physics, Springer, Berlin, 219–221, Vol. 6, 1979.
7. V.S. Letokhov (ed.), *Laser Analytical Spectrochemistry*, Adam Hilger, Bristol, 1986.
8. D.L. Andrews, *Lasers in Chemistry*, Springer, Berlin, 1990.
9. K. Kompa, V. Sick, V.J. Wolfrum (eds.), 'Laser Diagnostics for Industrial Processes', Special Issue of *Ber. Bunsenges. Phys. Chem.*, **97**, 1503–1745 (1993).
10. D.L. Andrews (ed.), *Applied Laser Spectroscopy: Techniques, Instrumentation and Application*, VCH, Weinheim, 1994.
11. R.J. Cotter, 'Lasers and Mass Spectrometry', *Anal. Chem.*, **56**, 485A (1984); D.M. Lubman, comments on 'Lasers and Mass Spectrometry', *Anal. Chem.*, **56**, 1256A–1257A (1984).
12. D.M. Lubman, *Lasers in Mass Spectrometry*, Oxford University Press, New York, 1990.
13. A. Vertes, R. Gijbels, F. Adams (eds.), *Laser Ionization Mass Analysis*, John Wiley & Sons, New York, 1993.
14. P.M. Johnson, 'Molecular Multiphoton Ionization Spectroscopy', *Appl. Opt.*, **19**, 3920–3925 (1980).
15. D.M. Lubman, M.N. Kronick, 'Mass Spectrometry of Aromatic Molecules with Resonance-enhanced Multiphoton Ionization', *Anal. Chem.*, **54**, 660–665 (1982).
16. R.B. Bernstein, 'Systematics of Multiphoton Ionization – Fragmentation of Polyatomic Molecules', *J. Phys. Chem.*, **86**, 1178–1184 (1982).
17. A. Gedanken, M.B. Robin, N.A. Kuebler, 'Nonlinear Photochemistry in Organic, Inorganic, and Organometallic Systems', *J. Phys. Chem.*, **86**, 4096–4107 (1982).
18. D.A. Gobeli, J.J. Yang, M.A. El-Sayed, 'Laser Multiphoton Ionization – Dissociation Mass Spectrometry', *Chem. Rev.*, **85**, 529–554 (1985).
19. D.M. Lubman, 'Optically Selective Molecular Spectrometry', *Anal. Chem.*, **59**, 31A–40A (1987).
20. H.J. Neusser, 'Lifetimes of Energy and Angular Momentum Selected Ions', *J. Phys. Chem.*, **93**, 3897–3907 (1989).
21. U. Boesl, 'Multiphoton Excitation and Mass Selective Ion Detection for Neutral and Ion Spectroscopy', *J. Phys. Chem.*, **95**, 2949–2962 (1991).
22. K.W.D. Ledingham, R.P. Singhal 'High Intensity Laser Mass Spectrometry – A Review', *Int. J. Mass Spectrom. Ion. Proc.*, **163**, 149–168 (1997).
23. O.P. Haefliger, R. Zenobi 'Laser Mass Spectrometric Analysis of Polycyclic Aromatic Hydrocarbons with Wide Wavelength Range Laser Multiphoton Ionization Spectroscopy', *Anal. Chem.*, **70**, 2660–2665 (1998).
24. V.S. Letokhov, *Laser Photoionization Spectroscopy*, Academic Press, Orlando, FL, 1987.
25. J. Grotemeyer, E.W. Schlag, 'Multiphoton-ionization Mass Spectrometry', *Angew. Chem., Int. Ed. Engl.*, **27**, 447–459 (1988).

26. U. Boesl, R. Weinkauf, E.W. Schlag, 'Reflectron Time-of-Flight Mass Spectrometry and Laser Excitation for the Analysis of Neutrals, Ionized Molecules and Secondary Fragments', *Int. J. Mass Spectrom. Ion Proc.*, **112**, 121–166 (1992).
27. D.M. Lubman, 'Methods Utilizing Low and Medium Laser Irradiance', in *Laser Ionization Mass Analysis*, eds A. Vertes, R. Gijbels, F. Adams, John Wiley & Sons, New York, 321, 1993.
28. U. Boesl, R. Weinkauf, C. Weickhardt, E.W. Schlag, 'Laser Ion Sources for Time-of-Flight Mass Spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **131**, 87–124 (1994).
29. R. Zenobi, 'In Situ Analysis of Surfaces and Mixtures by Laser Desorption Mass Spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **145**, 51–77 (1995).
30. R. Zenobi, Q. Zhan, P. Voumard, 'Multiphoton Ionization Spectroscopy in Surface Analysis and Laser Desorption Mass Spectrometry', *Mikrochim. Acta*, **124**, 273–281 (1996).
31. G.S. Hurst, M.G. Payne, S.D. Kramer, J.R. Young, 'Resonance Ionization Spectrometry and One-atom Detection', *Rev. Mod. Phys.*, **51**, 767–819 (1979).
32. G.S. Hurst, M.G. Payne, S.D. Kramer, C.H. Chen, 'Counting the Atoms', *Phys. Today*, Sept., 24–29, (1980).
33. J.D. Fassett, L.J. Moore, J.C. Travis, J.R. DeVoe, 'Laser Resonance Ionization Mass Spectrometry', *Science*, **230**, 262–267 (1985).
34. D.S. Zakheim, P.M. Johnson, 'Rate Equation Modelling of Molecular Multiphoton Ionization Dynamics', *Chem. Phys.*, **46**, 263–272 (1980).
35. W. Dietz, H.J. Neusser, U. Boesl, E.W. Schlag, S.H. Lin, 'A Model for Multiphoton Ionization Mass Spectrometry with Application to Benzene', *Chem. Phys.*, **66**, 105–127 (1982).
36. U. Boesl, R. Weinkauf, K. Walter, C. Weickhardt, E.W. Schlag, 'Tandem Time-of-Flight Techniques and Multiphoton Mass Spectrometry: The Ladder Switching in Benzene', *J. Phys. Chem.*, **94**, 8567–8573 (1990).
37. S.T. Fountain, D.M. Lubman, 'Wavelength-specific Resonance-enhanced Multiphoton Ionization for Isomer Discrimination via Fragmentation and Metastable Analysis', *Anal. Chem.*, **65**, 1257–1266 (1993).
38. W.C. Wiley, I.H. McLaren, 'Time-of-Flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, **26**, 1150 (1955).
39. B.A. Mamyrin, V.I. Karataev, D.V. Shmikk, V.A. Zagulin, 'The Mass-reflectron, a New Nonmagnetic Time-of-Flight Mass Spectrometer with High Resolution', *Sov. Phys. JETP*, **37**, 45 (1973).
40. R. Grix, R. Kutscher, G. Li, U. Grüner, H. Wollnik, 'A Time-of-Flight Mass Analyzer with High Resolving Power', *Rapid Commun. Mass Spectrom.*, **2**, 83 (1988).
41. T. Bergmann, T.P. Martin, H. Schaber, 'High-resolution Time-of-Flight Mass Spectrometers: Parts I to III', *Rev. Sci. Instrum.*, **60**, 347–349 (1989) and **61**, 2585–2600 (1990).
42. R.E. Smalley, L. Wharton, D.H. Levy, 'Molecular Optical Spectroscopy with Supersonic Beams and Jets', *Acc. Chem. Res.*, **10**, 139–145 (1977).
43. D.H. Levy, 'The Spectroscopy of Very Cold Gases', *Science*, **214**, 263–269 (1981).
44. J.M. Hayes, 'Analytical Spectroscopy in Supersonic Expansions', *Chem. Rev.*, **87**, 745–760 (1987).
45. N.K. Berezhetskaya, G.S. Voronov, G.A. Delone, N.B. Delone, G.K. Piskova, 'Effect of a Strong Optical-frequency Electromagnetic Field on the Hydrogen Molecule', *JETP*, **31**, 403–406 (1970).
46. S.L. Chin, 'Multiphoton Ionization of Molecules', *Phys. Rev. A*, **4**, 992–996 (1971).
47. R.E. Smalley, L. Wharton, D.H. Levy, 'The Fluorescence Excitation Spectrum of Rotationally Cooled NO_2', *J. Chem. Phys.*, **63**, 4977 (1975).
48. A. Herrmann, S. Leutwyler, E. Schumacher, L. Wöste, 'Multiphoton Ionization: Mass Selective Laser Spectroscopy of Na_2 and K_2 in Molecular Beams', *Chem. Phys. Lett.*, **52**, 418–425 (1977).
49. A. Herrmann, S. Leutwyler, E. Schumacher, L. Wöste, 'On Metal-atom Clusters. Photoionization Thresholds and Multiphoton Ionization Spectra of Alkali Metal Molecules', *Helv. Chim. Acta*, **61**, 453–487 (1978).
50. D.L. Feldman, R.K. Lengel, R.N. Zare, 'Multiphoton Ionization: a Method for Characterizing Molecular Beams and Beam Reaction Products', *Chem. Phys. Lett.*, **52**, 413–417 (1977).
51. L. Zandee, R.B. Bernstein, D.A. Lichtin, 'Vibronic/Mass Spectroscopy via Multiphoton Ionization of a Molecular Beam: the I_2 Molecule', *J. Chem. Phys.*, **69**, 3427–3429 (1978).
52. E.W. Rothe, B.P. Mathur, G.P. Reck, 'Isotope Fractionation in Two-step Photoionization of Li_2', *Chem. Phys. Lett.*, **53**, 74–78 (1978).
53. V.S. Antonov, I.N. Knyazev, V.S. Letokhov, V.M. Matiuk, V.G. Movshev, V.K. Potapov, 'Stepwise Laser Photoionization of Molecules in a Mass Spectrometer: a New Method for Probing and Detection of Polyatomic Molecules', *Opt. Lett.*, **3**, 37–39 (1978).
54. J.H. Brophy, C.T. Rettner, 'Laser Two-photon Ionization Spectroscopy: a New Method for Real-time Monitoring of Atmospheric Pollutants', *Opt. Lett.*, **4**, 337–339 (1979).
55. J.H. Brophy, C.T. Rettner, 'Laser Two-photon Ionization of Aniline in a Molecular Beam and the Bulk Gas Phase', *Chem. Phys. Lett.*, **67**, 351–355 (1979).
56. L. Zandee, R.B. Bernstein, 'Laser Ionization Mass Spectrometry: Extensive Fragmentation via Resonance-enhanced Multiphoton Ionization of a Molecular Benzene Beam', *J. Chem. Phys.*, **70**, 2574–2575 (1979).
57. L. Zandee, R.B. Bernstein, 'Resonance-enhanced Multiphoton Ionization and Fragmentation of Molecular Beams: NO, I_2, Benzene, and Butadiene', *J. Chem. Phys.*, **71**, 1359–1371 (1979).

58. K.R. Newton, D.A. Lichtin, R.B. Bernstein, 'Two-color Laser Study of the Multiphoton Ionization and Fragmentation of Triethylenediamine', *J. Phys. Chem.*, **85**, 15–17 (1981).
59. M.A. Duncan, T.G. Dietz, R.E. Smalley, 'Efficient Multiphoton Ionization of Metal Carbonyls Cooled in a Pulsed Supersonic Beam', *Chem. Phys.*, **44**, 415–419 (1979).
60. T.G. Dietz, M.A. Duncan, M.G. Liverman, R.E. Smalley, 'Efficient Multiphoton Ionization of Jet-cooled Aniline', *Chem. Phys. Lett.*, **70**, 246–249 (1980).
61. T.G. Dietz, M.A. Duncan, M.G. Liverman, R.E. Smalley, 'Resonance-enhanced Two-photon Ionization Studies in a Supersonic Molecular Beam: Bromobenzene and Iodobenzene', *J. Chem. Phys.*, **73**, 4816–4821 (1980).
62. D.M. Lubman, R. Naaman, R.N. Zare, 'Multiphoton Ionization of Azulene and Naphthalene', *J. Chem. Phys.*, **72**, 3034–3040 (1980).
63. G.J. Fisanick, T.S. Eichelberger, B.A. Heath, M.B. Robin, 'Multiphoton Ionization Mass Spectrometry of Acetaldehyde', *J. Chem. Phys.*, **72**, 5571–5580 (1980).
64. C.D. Cooper, A.D. Williamson, J.C. Miller, R.N. Compton, 'Resonantly-enhanced Multiphoton Ionization of Pyrrole, *N*-methylpyrrole, and Furan', *J. Chem. Phys.*, **73**, 1527–1537 (1980).
65. U. Boesl, H.J. Neusser, E.W. Schlag, 'Multiphoton Ionization in the Mass Spectrometry of Polyatomic Molecules: Cross-sections', *Chem. Phys.*, **55**, 193–204 (1981).
66. U. Boesl, H.J. Neusser, E.W. Schlag, 'Visible and UV Multiphoton Ionization and Fragmentation of Polyatomic Molecules', *J. Chem. Phys.*, **72**, 4327–4333 (1980).
67. J.T. Meek, R.K. Jones, J.P. Reilly, 'Photoelectron Energy Distribution Following UV Laser Ionization of Gas Phase Benzene', *J. Chem. Phys.*, **73**, 3503–3505 (1980).
68. R.S. Pandolfi, D.A. Gobeli, M.A. El-Sayed, 'Mechanism of Laser Multiphoton Ionization–Dissociation. A New Technique for Evaluating the Role of Neutrals', *J. Phys. Chem.*, **85**, 1779–1782 (1981).
69. T. Carney, T. Baer, 'Laser Wavelength Dependence of the REMPI Mass Spectrum of 2,4-Hexadiyne: Direct Evidence for Dissociation Through Ionic States', *J. Chem. Phys.* , **75**, 477–478 (1981).
70. D.H. Parker, R.B. Bernstein, D.A. Lichtin, 'Multiphoton Ionization – Fragmentation Patterns of Tertiary Amines', *J. Chem. Phys.*, **75**, 2577–2582 (1981).
71. D. Proch, D.M. Rider, R.N. Zare, 'Unimolecular Fragmentation Kinetics by Multiphoton Ionization', *Chem. Phys. Lett.*, **81**, 430–434 (1981).
72. M.A. Duncan, T.G. Dietz, R.E. Smalley, 'Two-color Photoionization of Naphthalene and Benzene at Threshold', *J. Chem. Phys.*, **75**, 2118–2125 (1981).
73. J.I. Murakaini, M. Ito, K. Kayal, 'Electronic Absorption Spectra of Biphenyl as Studied by Multiphoton Ionization–Supersonic Beam Technique', *J Chem. Phys.*, **74**, 6505–6506 (1981).
74. A. Hermann, M. Hoffmann, S. Leutwyler, E. Schuhmacher, L. Wöste, 'Optical Spectroscopy of Na_3 by Two-photon Ionization in a Supersonic Molecular Beam', *Chem. Phys. Lett.*, **62**, 216–222 (1979).
75. S. Leutwyler, U. Even, 'Isotopically Selective Two-photon Ionization of Aniline in Supersonic Beams', *Chem. Phys. Lett.*, **81**, 578–581 (1981).
76. C.M. Klimcak, J.E. Wessel, 'Gas Chromatography with Detection by Laser Excited Resonance Enhanced 2-Photon Photoionization', *Anal. Chem.*, **52**, 1233–1239 (1980).
77. G. Rhodes, R.B. Opsal, J.T. Meek, J.P. Reilly, 'Analysis of Polyaromatic Hydrocarbon Mixtures with Laser Ionization Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **55**, 280–286 (1983).
78. H. Zacharias, H. Rottke, K.H. Welge, 'Sensitivity and Mass Selectivity of NO Detection by Stepwise Photoionization', *Appl. Phys.*, **24**, 23–28 (1981).
79. V.S. Antonov, V.S. Letokhov, A.N. Shibanov, 'Laser Photoionization and Mass-spectral Detection of Single Naphthalene Molecules', *Opt. Commun.*, **38**, 182–184 (1981).
80. U. Boesl, H.J. Neusser, R. Weinkauf, E.W. Schlag, 'Multiphoton Mass Spectrometry of Metastables: Direct Observation of Decay in a High Resolution Time-of-Flight Instrument', *J. Phys. Chem.*, **86**, 4857–4863 (1982).
81. F.J. Vastola, A.J. Pierre, 'Ionization of Organic Solids By Laser Irradiation', *Adv. Mass Spectrom.*, **4**, 107–111 (1968).
82. V.S. Antonov, V.S. Letokhov, Y.A. Matveyets, A.N. Shibanov, 'Sputtering of Neutral Molecules and Molecular Ions from the Adenine Crystal Surface Induced by the UV Picosecond Laser Pulse', *Laser Chem.*, **1**, 37–43 (1982).
83. H. von Weyssenhoff, H.L. Selzle, E.W. Schlag, 'Verfahren und Einrichtung zum Erzeugen von gepulsten Molekularstrahlen, die grosse, thermisch instabile Moleküle enthalten', Ger. Patent, P32 24 801.6-09, 1986.
84. H.v. Weyssenhoff, H.L. Selzle, E.W. Schlag, 'Laser-desorbed Large Molecules in a Supersonic Jet', *Z. Naturforsch., Teil A*, **40**, 674–676 (1985).
85. R. Trembeuill, D.M. Lubman, 'Pulsed Laser Desorption with Resonant Two-photon Ionization Detection in Supersonic Beam Mass Spectrometry', *Anal. Chem.*, **58**, 1299–1303 (1986).
86. U. Boesl, J. Grotemeyer, K. Walter, E.W. Schlag, 'A High Resolution Time-of-Flight Mass Spectrometer with Laser Desorption and Laser Ionization Source', *Anal. Instrum.*, **16**, 151–171 (1987).
87. F. Engelke, J.H. Hahn, W. Henke, R.N. Zare, 'Determination of Phenylthiohydantoin-amino Acids by Two-step Laser Desorption/Multiphoton Ionization', *Anal. Chem.*, **59**, 909–912 (1987).
88. J. Grotemeyer, U. Boesl, K. Walter, E.W. Schlag, 'A General Soft Ionization Method for Mass Spectrometry:

Resonance-enhanced Multiphoton Ionization of Biomolecules', *Org. Mass Spectrom.*, **21**, 645–653 (1986).
89. D.A. Gobeli, J.R. Morgan, R.J. Pierre, M.A. El-Sayed, 'Studies on Rapid Dynamics in Laser-Multiphoton Ionization Dissociation Mass Spectrometry by Using Pump–Pump Two-color Picosecond Lasers: 2,4-Hexadiyne', *J.Phys. Chem.*, **88**, 178–181 (1984).
90. J.L. Knee, L.R. Khundkar, A.H. Zewail, 'Picosecond Mass Spectrometry of a Collisionless Photodissociation Reaction', *J. Chem. Phys.*, **82**, 4715–4716 (1985).
91. D.A. Gobels, M.A. El-Sayed, 'The Pump–Pump Picosecond Mass Spectrometric Technique and the Rate of Energy Distribution Prior to Ionic Fragmentation of 2,4-Hexadiyne by Laser Multiphoton Processes', *J. Phys. Chem.*, **89**, 1722–1729 (1985).
92. A. Amirav, U. Even, J. Jortner, 'Analytical Applications of Supersonic Jet Spectroscopy', *Anal. Chem.*, **54**, 1666–1673 (1982).
93. J.M. Hayes, G.J. Small, 'Rotationally Cooled Laser-induced Fluorescence/Gas Chromatography', *Anal. Chem.*, **54**, 1202–1204 (1982).
94. T. Imasaka, H. Fukuoka, T. Hayashi, N. Ishibashi 'Supersonic Jet Fluorescence Spectrometry of Perylene and Benzo[*a*]pyrene', *Anal. Chim. Acta*, **156**, 111–120 (1984).
95. J.W. Hager, S. Wallace, 'Two-laser Photoionization Supersonic Jet Mass Spectrometry of Aromatic Molecules', *Anal. Chem.*, **60**, 5–10 (1988).
96. J.R. Cable, M.J. Tubergen, D.H. Levy, 'Electronic Spectroscopy of Small Tryptophan Peptides in Supersonic Molecular Beams', *J. Am. Chem. Soc.*, **110**, 7349–7355 (1988).
97. E.A. Rohlfing, C.M. Rohlfing, 'Resonant Two-photon Ionization Spectroscopy of Jet-cooled *p*-Dichlorobenzene', *J. Phys. Chem.*, **93**, 94–101 (1989).
98. C.M. Gittins, E.A. Rohlfing, C.M. Rohlfing, 'Experimental and Theoretical Characterization of the S-1–S-0 Transition of Benzo[*a*]pyrene', *J. Chem. Phys.*, **105**, 7323–7335 (1996).
99. R. Tembreull, C.H. Sin, H.M. Pang, D.M.Lubman, 'Applicability of Resonant Two-photon Ionization in Supersonic Beam Mass Spectrometry to Halogenated Aromatic Hydrocarbons', *Anal. Chem.*, **57**, 1186–1192 (1985).
100. R. Trembeull, C.H. Sin, H.M. Pang, D.M. Lubman, 'Resonance Enhanced Multiphoton Ionization Spectroscopy for Detection of Azabenzenes in Supersonic Beam Mass Spectrometry', *Anal. Chem.*, **57**, 2911–2917 (1985).
101. T. Imasaki, K. Tashiro, N. Ishibashi, 'Gas Chromatographic Determination of Aniline Derivatives by Supersonic Jet Resonance Multiphoton Ionization Mass Spectrometry', *Anal. Chem.*, **58**, 3242–3244 (1986).
102. C. Weickhardt, R. Zimmermann, U. Boesl, E.W. Schlag, 'Laser Mass Spectrometry of Dibenzodioxin, Dibenzofuran and Two Isomers of Dichlorodibenzodioxins: Selective Ionization', *Rapid Commun. Mass Spectrom.*, **7**, 183–185 (1993).
103. C. Weickhardt, R. Zimmermann, K.W. Schramm, U. Boesl, E.W. Schlag, 'Laser Mass Spectrometry of the Di-, Tri- and Tetrachlorobenzenes: Isomer Selective Ionization and Detection', *Rapid Commun. Mass Spectrom.*, **8**, 381–384 (1994).
104. S. Satyapal, J.H. Werner, T.A. Cool, 'An Extended Flame Zone in the Combustion of CH_3Cl', *Combust. Sci. Technol.*, **106**, 229–238 (1995).
105. J.H. Werner, T.A. Cool, 'Flame Sampling Photoionization Mass Spectrometry of CH_3PO_2 and CH_3OPO_2', *Chem. Phys. Lett.*, **275**, 278–282 (1997).
106. J.H. Werner, T.A. Cool, 'Flame Sampling Photoionization Mass Spectrometry of Dichloroethanol', *Chem. Phys. Lett.*, **290**, 81–87 (1998).
107. J. Ahrens, R. Kovacs, E.A. Shafranowskii, K.H. Homam, 'On-line Multi-photon Ionization Mass Spectrometry Applied to PAH and Fullerenes in Flames', *Ber. Bunsenges. Phys. Chem.*, **98**, 265 (1994).
108. C.M. Gittins, M.J. Castaldi, S.M. Senkan, E.A. Rohlfing, 'Real-Time Quantitative Analysis of Combustion-generated Polycyclic Aromatic Hydrocarbons by Resonance-enhanced Multiphoton Ionization Time-of-Flight Mass Spectrometry', *Anal. Chem.*, **69**, 286–293 (1997).
109. H. Hepp, K. Siegmann, K. Sattler, 'New Aspects of Growth Mechanisms for Polycyclic Aromatic Hydrocarbons in Diffusion Flames', *Chem. Phys. Lett.*, **233**, 16–22 (1995).
110. K. Siegmann, H. Hepp, K. Sattler, 'Reactive Dimerization: a New PAH Growth Mechanism in Flames', *Combust. Sci. Technol.*, **109**, 165–181 (1995).
111. M.J. Castaldi, S.M. Senakn, 'Real-time, Ultrasensitive Monitoring of Air Toxics by Laser Photoionization Time-of-Flight Mass Spectrometry', *J. Air Waste Manage. Assoc.*, **48**, 77–81 (1998).
112. T.N. Tanada, Valezquez, N. Hemmi, T.A. Cool, 'Detection of Toxic Emissions from Incinerators', *Ber. Bunsenges. Phys. Chem.*, **97**, 1516–1527 (1993).
113. T.N. Tanada, J. Velazquez, N. Hemmi, T.A. Cool, 'Surrogate Detection for Continuous Emission Monitoring by Resonance Ionization', *Combust. Sci. Technol.*, **101**, 333–348 (1994).
114. H. Oser, R. Thanner, H. Grotheer, 'Jet-REMPI for the Detection of Trace Gas Compounds in Complex Gas Mixtures, a Tool for Kinetic Research and Incinerator Process Control', *Combust. Sci. Technol.*, **116/117**, 567–582 (1996).
115. H. Oser, R. Thanner, H. Grotheer, R. Walter, R. Richers, A. Merz, 'Jet-REMPI for Process Control in Incineration', *Proc. SPIE*, **3108**, 21–29 (1997).
116. R. Zimmermann, H.J. Heger, A. Kettrup, U. Boesl, 'A Mobile Resonance-enhanced Multiphoton Ionization Time-of-Flight Mass Spectrometry Device for On-line Analysis of Aromatic Pollutants in Waste Incinerator

Flue Gases: First Results', *Rapid Commun. Mass Spectrom.*, **11**, 1095–1102 (1997).
117. H.J. Heger, R. Zimmermann, R. Dorfner, M. Beckmann, H. Griebel, A. Kettrup, U. Boesl, 'On-line Emission Analysis of Polycyclic Aromatic Hydrocarbons at pptv Concentration Levels in the Flue Gas of an Incineration Pilot Plant with a Mobile Resonance-enhanced Multiphoton Ionization Time-of-Flight Mass Spectrometer', *Anal. Chem.*, **71**, 46–57 (1998).
118. U. Boesl, H. Nagel, C. Weickhardt, R. Frey, E.W. Schlag, 'Vehicle Exhaust Emission Time-resolved Multicomponent Analysis by Laser Mass Spectrometry', in *The Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, 5000–5022, 1998.
119. C. Kosmidis, K.W.D. Ledingham, A. Clark, A. Marshall, R. Jennings, J. Sander, R.P. Singhal, 'On the Dissociation Pathways of Nitrobenzene', *Int. J. Mass Spectrom. Ion Proc.*, **135**, 229–242 (1994).
120. K.W.D. Ledingham, H.S. Kilic, C. Kosmidis, R.M. Deas, A. Marshall, T. McCanny, R.P. Singhal, A.J. Langley, W. Shaikh, 'A Comparison of Femtosecond and Nanosecond Multiphoton Ionization and Dissociation for Some Nitro-molecules', *Rapid Commun. Mass Spectrom.*, **9**, 1522–1527 (1995).
121. K.W.D. Ledingham, C. Kosmidis, S. Georgiou, S. Couris, R.P. Singhal, 'A Comparison of the Femto-, Pico- and Nano-Second Multiphoton Ionization and Dissociation Processes of NO_2 at 248 and 496 nm', *Chem. Phys. Lett.*, **247**, 555–563 (1995).
122. D.J. Smith, K.W.D. Ledingham, R.P. Singhal, H.S. Kilic, T. McCanny, A.J. Langley, P.F. Taday, C. Kosmidis, 'Time-of-Flight Mass Spectrometry of Aromatic Molecules Subjected to High Intensity Laser Beams', *Rapid Commun. Mass Spectrom.*, **12**, 813–820 (1998).
123. J. Matsumoto, C.H. Lin, T. Imasaka, 'Enhancement of the Molecular Ion Peak from Halogenated Benzenes and Phenols Using Femtosecond Laser Pulses in Conjunction with Supersonic Beam Multiphoton Ionization Mass Spectrometry', *Anal. Chem.*, **69**, 4524–4529 (1997).
124. R. Weinkauf, P. Aicher, G. Wesley, J. Grotemeyer, E.W. Schlag, 'Femtosecond Versus Nanosecond Multiphoton Ionization and Dissociation of Large Molecules', *J. Phys. Chem.*, **98**, 8381–8391 (1994).
125. T. Imasaka, 'Supersonic Jet Multiphoton-ionization Mass Spectrometry Using a Femtosecond Laser', *Trends Anal. Chem.*, **17**, 551–556 (1998).
126. C.H. Lin, J. Matsumoto, S. Ohtake, T. Imasaka, 'Supersonic Jet Time-of-Flight Mass Spectrometry of Adenine using Nanosecond and Femtosecond Lasers', *Talanta*, **43**, 1925–1929 (1996).
127. N.P. Lockyer, J.C. Vickerman, 'Multiphoton Ionization Mass Spectrometry of Small Biomolecules with Nanosecond and Femtosecond Laser Pulses', *Int. J. Mass Spectrom.*, **176**, 77–86 (1998).
128. K.F. Willey, V. Vorsa, R.M. Braun, N. Winograd, 'Postionization of Molecules Desorbed from Surfaces By keV Ion Bombardment with Femtosecond Laser Pulses', *Rapid Commun. Mass Spectrom.*, **12**, 1253–1260 (1998).
129. C. Weickhardt, C. Grun, R. Heinicke, A. Meffert, J. Grotemeyer, 'The Application of Resonant Multiphoton Ionization By Sub-picosecond Laser Pulses for Analytical Laser Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 745–748 (1997).
130. C. Grun, C. Weickhardt, J. Grotemeyer, 'Multiphoton Ionization Mass Spectrometry of Metal Organic Compounds: Avoiding the Ultrafast Neutral Dissociation Channels By Femtosecond Laser Activation', *Eur. Mass Spectrom.*, **2**, 197–202 (1996).
131. A. Wucher, R. Heinrich, R.M. Braun, K.F. Willey, N. Winograd, 'Vacuum Ultraviolet Single Photon Versus Femtosecond Multiphoton Ionization of Sputtered Germanium Clusters', *Rapid Commun. Mass Spectrom.*, **12**, 1241–1245 (1998).
132. R.L.M. Dobson, A.P. D'Silva, S.J. Weeks, V.A. Fassel, 'Multidimensional, Laser-based Instrument for the Characterization of Environmental Samples for Polycyclic Aromatic Compounds', *Anal. Chem.*, **58**, 2129–2137 (1986).
133. C.W. Wilkerson, S.M. Colby, Jr, J.P. Reilly, 'Determination of Polycyclic Aromatic Hydrocarbons Using Gas Chromatography/Laser Ionization Mass Spectrometry with Picosecond and Nanosecond Light Pulses', *Anal. Chem.*, **61**, 2669–2673 (1989).
134. R.B. Opsal, J.P. Reilly, 'Ionization of Alkylbenzenes Studied By Gas Chromatography/Laser Ionization Mass Spectrometry', *Anal. Chem.*, **60**, 1060–1065 (1988).
135. R.B. Opsal, J.P. Reilly, 'Selective Analysis of Nitro- and Nitroso-containing Compounds by Laser Ionization Gas Chromatography /Mass Spectrometry', *Anal. Chem.*, **58**, 2919–2923 (1986).
136. B.V. Peppich, J.B. Callis, D.H. Burns, M. Gounterman, D.A. Kalman, 'Capillary Gas Chromatography/Pulsed Supersonic Jet/Fluorescence Excitation Spectroscopy for the Identification of Methylanthracenes in a Complex Environmental Sample', *Anal. Chem.*, **58**, 2825 (1986).
137. C. Köster, J. Grotemeyer, E.W. Schlag, 'A High Pressure Pulsed Valve for Gases, Liquids, and Supercritical Fluids', *Z. Naturforsch., Teil A*, **45**, 1285–1292 (1990).
138. R. Zimmermann, C. Lermer, K.W. Schramm, A. Kettrup, U. Boesl, 'Three-dimensional Trace Analysis: Combination of Gas Chromatography, Supersonic Beam UV Spectroscopy and Time-of-Flight Mass Spectrometry', *Eur. Mass Spectrom.*, **1**, 341–351 (1995).
139. J.H. Hahn, R. Zenobi, R.N. Zare, 'Subfemtomole Quantification of Molecular Adsorbates by Two-step Laser Mass Spectrometry', *J. Am. Chem. Soc.*, **109**, 2842–2843 (1987).
140. J. Grotemeyer, E.W. Schlag, 'Peptides Investigated by Laser Desorption Multiphoton Ionization Mass Spectrometry', *Org. Mass. Spectrom.*, **23**, 388–396 (1988).

141. R. Weinkauf, P. Schanen, D. Yang, S. Soukara, E.W. Schlag, 'Elementary Processes in Peptides: Electron Mobility and Dissociation in Peptide Cations in the Gas Phase', *J. Phys. Chem.*, **99**, 11255–11265 (1995).
142. R. Weinkauf, P. Schanen, A. Metsala, E.W. Schlag, M. Bürgle, H. Kessler, 'Highly Efficient Charge Transfer in Peptide Cations in the Gas Phase: Threshold Effects and Mechanism', *J. Phys. Chem.*, **100**, 18567–18585 (1996).
143. M.J. Dale, K.F. Costello, A.C. Jones, P.R.R. Langridge-Smith, 'Investigation of Porphyrins and Metalloporphyrins Using Two-step Laser Mass Spectrometry', *J. Mass Spectrom.*, **31**, 590–601 (1996).
144. H. Böll, P. Püffel, C. Weickhardt, U. Boesl, 'Laser Desorption/Resonance Post-ionization of Large PAH's: Ionization and Surface Specific Fragmentation of Decacylene', in *Resonance Ionization Spectroscopy 1994*, eds. H.-J. Kluge, J.E. Parks, K. Weudt, AIP Press, New York, 68–71, Vol. 329, 1995.
145. S.M. Hankin, P. John, A.W. Simpson, G.P. Smith, 'Spatially Resolved Time-of-Flight Mass Spectrometry of Polycyclic Aromatic Hydrocarbons: Quantification Studies', *Anal. Chem.*, **68**, 3238–3443 (1996).
146. G. Meijer, M.S. de Vries, H.E. Hunziker, Wendt, 'Laser Desorption Jet-cooling of Organic Molecules', *Appl. Phys.*, **B51**, 395–403 (1990).
147. G. Meijer, M.S. de Vries, H.E. Hunziker, H.R. Wendt, 'Laser Desorption Jet-cooling Spectroscopy of *para*-Aminobenzoic Acid Monomer, Dimer and Clusters', *J. Chem. Phys.*, **92**, 7625–7635 (1990).
148. P. Arrowsmith, M.S. de Vries, H.E. Hunziker, H.R. Wendt, 'Pulsed Laser Desorption Near a Jet Orifice, Concentration Profiles of Entrained Perylene Vapor,' *Appl Phys.*, **B46**, 165–173 (1988).
149. P. Voumard, Q. Zhan, R. Zenobi, 'A New Instrument for Spatially Resolved Laser Desorption/Laser Multiphoton Ionization Mass Spectrometry', *Rev. Sci. Instrum.*, **64**, 2215–2220 (1993).
150. O.P. Haefliger, R. Zenobi, 'New Sample Preparation for Quantitative Laser Desorption Mass Spectrometry and Optical Spectroscopy', *Rev. Sci. Instrum.*, **69**, 1828–1832 (1998).
151. S.J. Wright, M.J. Dale, P.R.R. Langridge-Smith, Q. Zhan, R. Zenobi, 'Selective in Situ Detection of Polymer Additives Using Laser Mass Spectrometry', *Anal. Chem.*, **68**, 3585–3594 (1996).
152. Q. Zhan, P. Voumard, R. Zenobi, 'Application of Two-Step Laser Mass Spectrometry to the Chemical Analysis of Aerosol Particle Surfaces', *Rapid Commun. Mass Spectrom.*, **9**, 119–127 (1995).
153. R. Zenobi, J-M. Philippoz, P. Buseck, R.N. Zare, 'Spatially Resolved Organic Analysis of the Allende Meteorite', *Science*, **246**, 1026 (1989).
154. L.J. Kovalenko, C.R. Maechling, S.J. Clemett, J.M. Philippoz, R.N. Zare, C.M. Alexander, 'Microscopic Organic Analysis Using Two-step Laser Mass Spectrometry: Application to Meteoric Acid Residues', *Anal. Chem.*, **64**, 682–690 (1992).
155. S.J. Clemett, X.D.F. Chillier, S. Gillette, R.N. Zare, M. Maurette, C. Engrand, G. Kurat, 'Observation of Indigenous Polycyclic Aromatic Hydrocarbons in ''giant'' Carbonaceous Antarctic Micrometeorites', *Origins Life Evol. Biosphere*, **28**, 425–448 (1998).
156. M.J. Dale, A.C. Jones, S.J.T. Pollard, P.R.R. Langridge-Smith 'Direct Determination of Polycyclic Aromatic Hydrocarbons in Environmental Matrices Using Laser Desorption Laser Photoionization Time-of-Flight Mass Spectrometry', *Analyst*, **119**, 571–578 (1994).
157. O. Kornienko, E.T. Ada, L. Hanley, 'Organic Surface Analysis By Two-laser Ion Trap Mass Spectrometry', *Anal. Chem.*, **69**, 1536–1542 (1997).
158. C.R. Maechling, S.J. Clemett, F. Engelke, R.N. Zare, 'Evidence for Thermalization of Surface-desorbed Molecules at Heating Rates of 10^8 K/s', *J. Chem. Phys.*, **104**, 8768–8776 (1996).
159. M. Handschuh, S. Nettesheim, R. Zenobi, 'Laser Induced Desorption of Thin Molecular Films Investigated with High Time Resolution', *J. Chem. Phys.*, **108**, 6548–6549 (1998).
160. M. Handschuh, S. Nettesheim, R. Zenobi, 'Is Laser Heating Advantageous for Thermal Desorption of Large Polar Molecules', *J. Chem. Phys.*, **107**, 2603–2610 (1997).
161. Q. Zhan, S.J. Wright, R. Zenobi, 'Laser Desorption Substrate Effects', *J. Am. Soc. Mass Spectrom.*, **8**, 525–531 (1997).
162. J.Y. Zhang, D.S. Nagra, L. Li, 'Molecular Cooling and Supersonic Jet Formation in Laser Desorption', *Anal. Chem.*, **65**, 2812–2818 (1993).
163. S.S. Alimpiev, V.V. Mlynsky, N.E. Belov, S.M. Nikiforov, 'Selective Detection of Phenol Impurities in Water', *Anal. Chem.*, **67**, 181–186 (1995).
164. N.E. Belov, S.S. Alimpiev, V.V. Mlynsky, S.M. Nikiforov, P.J. Derrick, 'Ablation of Organic Molecules from Frozen Matrices', *Rapid Commun. Mass Spectrom.*, **9**, 1431–1436 (1995).
165. B.D. Morrical, D.P. Fergenson, K.A. Prather, 'Coupling Two-step Laser Desorption/Ionization with Aerosol Time-of-Flight Mass Spectrometry for the Analysis of Individual Organic Particles', *J. Am. Soc. Mass Spectrom.*, **9**, 1068–1073 (1998).
166. J.B. Morris, M.V. Johnston, 'Multiphoton Ionization of Transition-metal Tetraphenylporphyrins. Metal Complexes Which Display Molecular Ionization', *Int. J. Mass Spectrom. Ion Proc.*, **73**, 175–180 (1986).
167. D.S. Nagra, J.Y. Zhang, L. Li, 'Pulsed Rapid Heating Method for Volatilization of Biological Molecules in Multiphoton Ionization Mass Spectrometry', *Anal. Chem.*, **63**, 2188–2193 (1991).

168. K. Wendt, G.K. Bhowmick, B.A. Bushaw, G. Herrmann, J.V. Kratz, J. Lantzsch, P. Muller, W. Nortershauser, E.W. Otten, R. Schwalbach, U.A. Seibert, N. Trautmann, A. Waldek, 'Rapid Ultratrace Analysis of Sr-89, Sr-90 in Environmental Samples by Collinear Laser Resonance Ionization Mass Spectrometry', *Radiochim. Acta*, **79**, 183–190 (1997).
169. P. Peuser, G. Herrmann, H. Rimke, P. Sattelberger, N. Trautmann, W. Ruster, F. Ames, J. Bonn, H.J. Kluge, U. Krönert, E.W. Otten, 'Trace Detection of Plutonium by Three-step Photoionization with a Laser System Pumped by a Copper Vapor Laser', *Appl. Phys.*, **B38**, 249–253 (1985).
170. D.W. Beekmann, T.A. Calcott, S.D. Kramer, E.T. Arakawa, G.S. Hurst, E. Nussbaum, 'Resonance Ionization Source for Mass Spectrometry', *Int. J. Mass Spectrom. Ion Phys.*, **34**, 89–97 (1980).
171. N.S. Nogar, R.C. Estler, C.M. Miller, 'Pulsed Laser Desorption for Resonance Ionization Mass Spectrometry', *Anal. Chem.*, **57**, 2441–2444 (1985).
172. H. Nagel, R. Frey, U. Boesl, 'On-line Analysis of Formaldehyde and Acetaldehyde in Non-stationary Engine Operation Using Laser Mass Spectrometry', *SAE Technical Papers Series*, 9 610 84, 131–136, 1996.
173. H. Nagel, R. Frey, C. Hertgerink, H.E. Rikeit, R.D. Greiner, C. Klein, U. Boesl, 'Online Analysis of Individual Aromatic Hydrocarbons in Automotive Exhaust: Dealkylation of the Aromatic Hydrocarbons in the Catalytic Converter', *SAE Technical Papers Series*, 9 71 606, 47–53, 1997.
174. P.K. Püffel, W. Thiel, R. Frey, U. Boesl, 'A New Method for the Investigation of Unburned Oil Emissions in the Raw Exhaust of SI Engines', *SAE Technical Papers Series*, 9 82 438, 1–7, 1998.
175. R. Zimmermann, H.J. Heger, C. Yeretzian, H. Nagel, U. Boesl, 'Application of Laser Ionization Mass Spectrometry for On-line Monitoring of Volatiles in the Headspace of Food Products: Roasting and Brewing of Coffee', *Rapid Commun. Mass Spectrom.*, **10**, 1975–1979 (1996).
176. J. Namiesnik, 'Generation of Standard Gaseous Mixtures', *J. Chromatogr.*, **300**, 79–108 (1984).
177. A.E. O'Keefe, G.C. Ortman, 'Primary Standards for Trace Gas Analysis', *Anal. Chem.*, **38**, 760–763 (1966).
178. U. Boesl, C. Weickhardt, S. Schmidt, H. Nagel, E.W. Schlag, 'A Calibration Method for the Quantitative Analysis of Gas Mixtures By Means of Laser Mass Spectrometry', *Rev. Sci. Instrum.*, **64**, 3482–3486 (1993).
179. R.B. Opsal, J.P. Reilly, 'Chromatographic Effluent Detection with Laser Ionization Mass Spectrometry', *Opt. News*, June, 18–21 (1986).
180. O.P. Haeflinger, R. Zenobi, 'Laser Mass Spectrometric Analysis of Polycyclic Aromatic Hydrocarbons with Wide Wavelength Range Laser Multiphoton Ionization Spectroscopy', *Anal. Chem.*, **70**, 2660–2665 (1998).
181. B.A. Williams, T.N. Tanada, T.A. Cool, 'Resonance Ionization Detection Limits for Hazardous Emissions', in *24th Symposium on Combustion*, Combustion Institute, Pittsburgh, 1587–1596, 1992.

Laser-based Combustion Diagnostics

Jürgen Wolfrum, Thomas Dreier, Volker Ebert, and Christof Schulz
Institute of Physical Chemistry, University of Heidelberg, Germany

1 Introduction **2**
2 Basic Principles **2**
2.1 Linear Techniques 2
2.2 Nonlinear Techniques 8
3 Application of Laser-based Techniques **14**
3.1 Detection Sensitivities 14
3.2 Laminar Flames 14
3.3 Turbulent Combustion 18
3.4 Engine Combustion 19
3.5 Coal and Waste Combustion 20
4 Future Developments **22**
Acknowledgments **23**
Abbreviations and Acronyms **24**
Related Articles **24**
References **25**

In laser spectroscopy the interaction of light emitted from various types of laser sources – tunable or nontunable in their output frequency – with the atomic or molecular species of interest is used to probe the sample through a variety of spectral responses. In order to perform laser spectroscopy suitable laser sources must be selected which meet the requirements of the chosen spectroscopic method. This means that the laser has to provide radiation which is in the wavelength range of interest, has the appropriate emission characteristics (lineshape) and a suitable energy to perform the measurements. Further requirements are pulse length (milliseconds to femtoseconds or continuous wave), repetition rate and beam profile. Nowadays, laser radiation can be generated with most of the required parameters necessary for the respective spectroscopic application, either directly or by generating new radiation frequencies by frequency mixing of one or several laser beams in a nonlinear medium (gas, liquid, solid).

As an example, the most direct probe is absorption of laser radiation (LAS, laser absorption spectroscopy) by suitable spectroscopically allowed transitions in atoms or molecules which are known from conventional spectroscopic methods. The increase or decrease in the laser radiation transmitted through the sample is then a measure of the amount of substance probed in the sample which characteristically absorbs at the required wavelength.

Laser light scattering methods, elastic (Rayleigh scattering (RS)) or inelastic (spontaneous Raman scattering (SRS)), are other techniques to probe the medium. In the first method, which is nonspecies specific, the density of the medium can be interrogated, whereas the second is able to probe all species with Raman-active vibration–rotation transitions.

There are several advantages in using laser spectroscopy instead of conventional spectroscopic techniques using conventional thermal light sources. The spectral brightness of laser beams is many orders of magnitude higher than that of thermal radiation sources, which correspondingly increase the detection sensitivity *of laser spectroscopic techniques. In addition, the small linewidth of the emitted radiation increases dramatically the spectral resolution such that minor details of the spectroscopic branch investigated can be resolved. This enables more quantitative interpretations of all parameters influencing the lineshape and intensity of the probed transition, and as such the physical and chemical environment of the probed species: temperature, pressure, velocity, chemical species and so on. It makes laser spectroscopic techniques much more* selective *than conventional methods, which often are not able to separate closely spaced spectral features from different species. A third advantage of laser spectroscopic techniques is connected with the variable pulse duration and repetition frequency of lasers: the very short pulse lengths can be used successfully to probe the sample within time periods which are short compared to any other physical or chemical time development – flow, chemical reaction, pressure changes and so on. Finally, the small spatial regions which can be probed by focusing diffraction limited laser beams makes laser spectroscopic techniques ideally suited for applications where high spatial resolution is required. All these advantages of laser spectroscopy are beneficial when the various techniques are applied as a diagnostic tool in combustion processes: flames constitute a complex interaction of fast chemistry with flow fields and surfaces and, therefore, a detailed understanding of combustion events often needs in situ, species-specific optical diagnostics with high spatial and temporal resolution.*

In many applications laser spectroscopy is a developed technique and can even be applied by untrained people. However, numerous laser spectroscopic techniques require detailed theoretical knowledge of the spectroscopy underlying the respective technique, and use sophisticated equipment in order to obtain meaningful results. Future development is aimed towards simplifying experimental set-ups, data evaluation and maintenance. This development runs parallel to the breathtaking development in laser technology.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Combustion is the oldest and one of the most successful technologies of humanity. An early important spectroscopic observation in a combustion process was made in 1859 by Robert Wilhelm Bunsen and Gustav Robert Kirchhoff,[1] who recognized the origin of some of the famous dark lines in the solar spectrum first seen by Wollaston and indexed in 1814 by Josef Fraunhofer. Bunsen and Kirchhoff realized that it could not happen by chance that 60 of the Fraunhofer lines coincided with sixty iron emission lines measured in the hot nonluminous flame gases of the Bunsen burner. Their experimental success, however, was based on the ideas of Isaac Newton, who introduced the word spectrum into physics. After Bunsen and Kirchhoff, spectroscopy in flames became a rapidly growing field. However, the meaning of all the observed lines was not known until 1913 when Niels Bohr formulated his model of the atom. A few years later, Albert Einstein[2] published his famous paper *Zur Quantentheorie der Strahlung* in which he derived Planck's law and Bohr's rule by discussing the possible elementary process of energy and momentum exchange through radiation. Despite Einstein's clear description of the stimulated emission, the practical application of this principle was achieved only many years later in 1955, first in the microwave region (microwave amplification by stimulated emission of radiation (MASER))[3] and in 1960, in the visible spectral region (light amplification by stimulated emission of radiation (LASER)).[4] The principle of stimulated emission has now been realized in all states of matter: solids, liquids, gases and free electrons. This allows generation of coherent radiation from the far-infrared to the X-ray region. Compared to flame spectroscopy with conventional light sources[5] the introduction of tunable lasers and the development of nonlinear optical techniques[6] greatly expanded the potential of combustion spectroscopy. Now the spectroscopic states of atoms and molecules in combustion process can be observed nonintrusively with high temporal, spectral and spatial resolution. An excellent overview can be found in the book by Eckbreth.[7]

Technical combustion processes consist of a complex multidimensional and time-dependent interaction of a large number of elementary chemical reactions with different transport processes for mass, momentum and energy. To increase the rate of chemical conversion, mostly turbulent flow conditions are applied in technical combustion. Similar to the laminar case, turbulent reactive flows can be described by solving the conservation equations for total species mass, momentum and enthalpy (Navier–Stokes equations).[8] However, solving the Navier–Stokes equations by direct numerical simulation (DNS) is, even in the days of modern parallel computing a very demanding task. For a realistic system of liquid hydrocarbon oxidation in an internal combustion (IC) engine one would need more than 10^{21} computing steps. Therefore, at present and for the near future a DNS of three-dimensional turbulent reactive flows in technical combustion systems will not be possible. Instead, turbulence has to be modeled by the use of averaged Navier–Stokes equations, while mean reaction rates can be evaluated with the help of probability density functions (PDFs). These models can then be validated by multidimensional nonintrusive laser spectroscopic measurements. In addition, such laser spectroscopic techniques can be applied directly in technical combustion systems to study details of the complex interaction of chemical kinetics and transport processes and act as sensors for species concentration, temperature, flow and other parameters in active control loops to improve fuel efficiency and reduce the environmental effects of combustion.

2 BASIC PRINCIPLES

Laser-based combustion diagnostics as well as other optical diagnostic methods exploits extensively the direct interaction of light and matter, providing an important tool to observe the spectroscopic states of molecules and atoms with high spectral and spatial resolution in time regimes from hours to the femtosecond (10^{-15} s) scale, using optical power densities from femtowatts (10^{-12}) to multiterawatts (10^{+12}) per square centimeter. A main feature of optical techniques is the nonintrusive in situ investigation of the sample without any interference between the sample and the measuring device itself. This is especially important for reactive and aggressive environments like combustion processes where probe sampling techniques are of questionable value. In all cases light with a certain intensity and temporal and spectral distribution is sent onto or through a sample. The energy, in most cases in form of light, re-emitted from the interaction region of the irradiated sample is observed and used to characterize the composition and/or structure of the sample. This is done by a detailed analysis of the intensity, spectral composition or angular and temporal distribution of the scattered light. Only absorption techniques determine the losses of the incoming light along the line of sight.

2.1 Linear Techniques

A spectroscopic technique is called linear if the photometric effect shows a linear relationship to the amount of photons to which the sample is exposed. This is the case as long as the population of the initial state, mostly

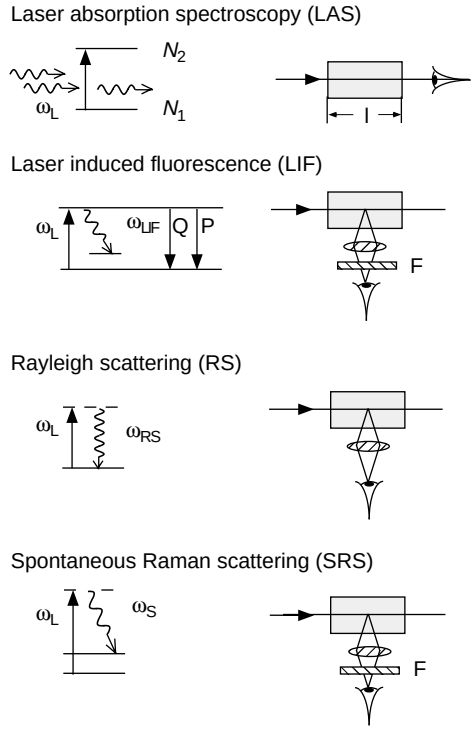

Figure 1 Energy level diagrams showing photon–species interaction (left) and schematic experimental set-up (right) for linear laser–spectroscopic techniques discussed in the text. $w_i = 2\pi\nu_i$, circular frequency of respective incident and/or emitted radiation; N_i, number density (cm^{-3}) of ith energy level; Q, P, quenching, predissociation rate (s^{-1}), respectively. Solid arrows indicate radiation- or collision-induced processes, and wavy arrows incident/emitted radiation, respectively.

a ground state, is not changed significantly during the interaction process. Especially with the development of modern laser sources it is fairly easy to achieve photon densities high enough to work beyond the linear power regime. Therefore, great care has to be taken to avoid nonlinear interactions. Owing to the complexities (and new possibilities) introduced by entering the nonlinear regime the most important laser-based diagnostic techniques are explained for the linear and the nonlinear case separately. Figure 1 depicts the basic excitation schemes for the most important linear techniques used for combustion diagnostics: LAS, laser-induced fluorescence (LIF), RS and SRS.

2.1.1 Laser Absorption Spectroscopy

Absorption spectroscopy is one of the oldest techniques for the nonintrusive investigation of gaseous media. The subject of interest in LAS is mostly the absolute concentration or absorber density of one or more atomic or molecular species which are present in a mixture.[9,10] Apart from this, LAS is able to determine simultaneously the physical boundary conditions (such as temperature, velocity, pressure or mass flux etc.) to which the absorbers are exposed.[11]

The basic set-up of an absorption spectrometer can be described as a radiation source, which emits collimated light with an initial intensity $I_0(\nu)$ through the sample under investigation towards a radiation detector. Part of the radiation, $dI(\nu)$, is absorbed by atomic or molecular constituents along each path increment dl in a simple one step process. The light intensity $I(\nu)$ behind the sample is monitored to determine the total loss and used as a quantitative measure of the number of absorbers along the beam path. The transition frequency ν at which the absorption occurs is characteristic of the absorbing substance, so that a simultaneous species identification is possible (spectral fingerprint). Integration of the light losses over the total absorption path l leads, under the assumption of a homogeneous absorber distribution and a monochromatic light source, to the Lambert–Beer Equation (1)

$$I(\nu, l) = I_0(\nu) \exp[-\kappa(\nu) conc.\, l] \tag{1}$$

where *conc.* is the absorber concentration, and $\kappa(\nu)$ the linear absorption coefficient (in the respective concentration units).[6,9]

At thermal equilibrium the fractional population $f_B(j, T)$ of level j with energy E_j is described by the Boltzmann Equation (2)

$$f_B(j, T) = \frac{N_j}{N} = \frac{g_j}{Q(T)} \exp\left(\frac{-E_j}{k_B T}\right. \tag{2}$$

where N is the total number density (cm^{-3}) of species, g_j is the statistical weight of level j, which is closely related to the total angular momentum quantum number j, T is the temperature, k_B is Boltzmann's constant, and $Q(T)$ is the partition function. Under equilibrium of particle ensemble and radiation field the upward and downward radiative transition rates have to be equal and the Einstein coefficients are related by Equation (3):

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h \nu_{12}^3}{c^3} \tag{3}$$

At low laser intensity the total amount ΔI absorbed at the transition frequency ν_{12} by any upward transitions from E_1 to E_2 is then described as, Equation (4):

$$-\Delta I(\nu_{12}) = I_0(\nu_{12}) N_1 \Delta l \frac{A_{21} c^2}{8\pi \nu_{12}^2} \tag{4}$$

2.1.1.1 Absorption Line Parameters: Line Strength and Shape In order to remove its lineshape dependence $\kappa(\nu)$ is often expressed as a product of a frequency independent line strength factor S, called integrated absorption coefficient, and a normalized lineshape function

$\phi(\nu - \nu_{12})$ with half width γ centered at ν_{12}, Equation (5):

$$\kappa(\nu) = S\phi(\nu - \nu_{12}), \qquad \int_{-\infty}^{\infty} \phi(\nu - \nu_{12})\,\mathrm{d}\nu = 1$$
$$S = \int_{-\infty}^{\infty} \kappa(\nu)\,\mathrm{d}\nu \tag{5}$$

With these relations we can rewrite the Lambert–Beer relation as Equation (6)

$$I(\nu, l) = I_0(\nu)\exp[-S'\phi(\nu)N_1 l] \quad \text{with} \quad S' = \frac{S}{N_1} \tag{6}$$

and remove the lineshape dependence by integrating over the entire absorption line, so that the total spectral "area" S, occupied by the absorption line, reflects only the absorption strength and the number density of absorbers in the initial level E_1. In this case the relation between the Einstein coefficient for spontaneous emission, A_{21}, and S becomes, $S = A_{21}N_1 g_2/g_1\nu_{12}^2 c^2/8\pi$,[12] which shows that the integrated absorption coefficient only is proportional to absorber specific constants and the number of atoms in the initial state.

2.1.1.2 Temperature Dependencies In any combustion-related sensor for molecular species it is very important to recognize that S is a strong function of temperature. If the pressure remains constant during the temperature change the absorber density will be modified by T^{-1}. In addition, the fractional population f_B (Equation 2) at the initial energy level changes as a function of temperature and energy of the probed state. A change in temperature of several hundred degrees can cause variations in the integrated absorption coefficient over two orders of magnitude. On the other hand, this strong temperature dependence can be used to measure the temperature itself by ratioing the integrated absorption of two suitable transitions.[11,13]

2.1.1.3 Lineshapes and Broadening The finite lifetimes of all excited states lead to natural or radiative line broadening, which is described by a Lorentzian profile with a width γ_N. In addition, the thermal motion of the absorbers with mass M in a gas in thermal equilibrium at temperature T causes a velocity dependent Doppler shift in transition frequency, resulting in a distribution of transition frequencies, which is modeled by a Gaussian lineshape function of width γ_D. Furthermore, collisional broadening is generated by frequent perturbations of the absorber through collisions with other gas molecules, which results in a shortening of the lifetime of the excited state. This lineshape often is sufficiently described by a Lorentzian function. To take into account the effect of different collision partners, separate foreign-gas (γ_f) and self-broadening (γ_s) coefficients are introduced and empirical expressions for the pressure and temperature dependence of the total halfwidth are combined to give Equation (7)

$$\gamma_{\text{Lorentz}} = \left(\gamma_s \frac{p_s}{p_{\text{tot}}} + \gamma_f \frac{p_f}{p_{\text{tot}}}\right)\left(\frac{T_0}{T}\right)^n \tag{7}$$

with T_0 a reference and T the actual temperature, and p_s, p_f the partial pressures of the absorber and the perturber species and p_{tot} the total pressure respectively. n reflects the temperature dependence and from kinetic gas theory is a power law with $n = 0.5$. Finally, at intermediate pressures, when none of the above mentioned mechanisms dominates, the experimental lineshape is described by a convolution of Gauss and Lorentz profiles, the so-called Voigt function, which generally cannot be expressed in an analytical form. Approximations and further literature on lineshapes are accessible through.[9,12,14]

LAS can be applied to any absorber which has either a permanent dipole moment or an allowed electronic transition. In the first case almost any heteronuclear molecule relevant to combustion like H_2O, CO, CO_2, CH_4, NO, NO_2, NH_3 and so on[10,13,15,16] can be detected via strong ground state rotational–vibrational transitions in the mid-infrared or via weaker overtone and combination bands in the near-infrared (0.75–3 μm) spectral region. Even stronger electronic transitions of all atoms, ions, molecules and free radicals, e.g. H, O, N, O_2, NO, CH_4, CH, OH, NH_2, CH_3, HO_2,[9,17,18] important in combustion are accessible in the visible (350–750 nm) and most often in the ultraviolet (UV) part (<350 nm) of the spectrum.

2.1.1.4 Long-path Absorption Techniques To increase detection sensitivities in absorption measurements the pathlength of the sample traversing beam has to be maximized. White- or Herriot-type mirror arrangements often are employed at the cost of spatial resolution. Another line-of-sight absorption technique which has emerged recently as a species-specific combustion diagnostic tool to monitor reactants, intermediates and products with very high sensitivity ($<10^{-6}$ fractional absorption) is cavity ring-down laser absorption spectroscopy (CRLAS).[19] In CRLAS a laser beam is coupled into a linear cavity formed by two concave mirrors with high reflectivity (>99.98%). The intensity decay in the cavity is given by $\mathrm{d}I/\mathrm{d}t = -IMTc/L$, where L is the length of the cavity with mirrors of transmissivity MT. This leads to the exponential solution $I = I_0\exp(-MTtc/L)$ with the decay constant $\tau = L/MTc$. One can define a total round trip loss of $\Gamma = 1 - \exp[-2L/c\tau]$, since the round trip time in the cavity is $\tau_{RT} = 2L/c$.[19] The intensity of the trapped light pulse is monitored in time via transmission through the output mirror on successive passes inside the cavity. Any absorption generated by the probed species increases the overall photon decay rate of the light pulse

in its multiple travels between the mirrors (Equation 8)

$$\frac{1}{\tau} = \frac{(MT + \Lambda + \sigma(\lambda)h)c}{L} \tag{8}$$

from that of the empty resonator with τ_{RT}. In Equation (8) $MT + \Lambda$ are cavity losses caused by mirror transmission (MT), and nonspecific absorption and scattering (Λ); $\sigma(\lambda)$ is the absorption cross-section at wavelength λ of the sample (concentration n) placed inside the cavity. Therefore, the decay time is a quantitative measure of the absorption strength and hence the concentration of the species. Since only the decay time is relevant the method also has the advantage of being insensitive to pulse-to-pulse energy fluctuations of the laser source. However, any other intensity losses should be avoided (scattering, surface reflection, bulk material absorption etc.), which reduce the beam intensity inside the cavity, thereby reducing the decay time.

Intermediates in hydrocarbon combustion, such as hydroxy, methylene and methyl radicals, have been detected in laminar flames[20,21] and low pressure plasmas,[22] respectively, using CRLAS or intracavity laser absorption spectroscopy (ICLAS) with UV light to stimulate electronic transitions. The exceptionally high signal-to-noise level in such measurements leads to very accurate, path-integrated temperature determinations of less than 2% in atmospheric pressure flames. Using CRLAS, HCO,[23] hydrocarbons (CH_4, CH_3), water and CO_2 were also recently detected quantitatively in their characteristic infrared spectrum around 3 μm with very high spectral resolution ($0.007\,cm^{-1}$) in low pressure (35 torr) laminar methane/air flames using an optical parametric oscillator (OPO),[24] and ICLAS has made possible the detection of CH_2 in flames.[21] Apart from combustion diagnostics CRLAS is increasingly being applied in other fields of physical chemistry including high resolution spectroscopy[25] and chemical kinetics.[26]

2.1.2 Laser-induced Fluorescence

LIF is the process of spontaneous emission from an excited electronic state populated upon absorption of a laser photon. Large cross-sections mainly in the visible and UV spectral range enable species detection down to the sub-parts per million range. Many combustion-relevant species like OH, O_2, NO, CH, CN, NH, C_2 can be accessed selectively (Eckbreth[7] and references therein). In many cases this technique is easily extended to two-dimensional imaging yielding concentration measurements with high spatial resolution simultaneously with high temporal resolution upon nanosecond laser pulse excitation.

2.1.2.1 Linear Laser-induced Fluorescence Equation (9) describes the LIF intensity I_{LIF} in the linear range for one species as

$$I_{LIF} = N(p, T) f_B V B_{ik} I_L g_\lambda(p, T) \phi \eta \varepsilon \frac{\Omega}{4\pi} \tag{9}$$

The LIF intensity is proportional to the number density $N(p, T)$ of the excitable molecules in the probed volume V which is determined by $N(p, T)$ times the temperature-dependent Boltzmann fraction f_B giving the population of the initial level i. The Einstein B_{ik} coefficient describes the absorption probability for transition $i \rightarrow k$. The fluorescence intensity depends linearly on the laser intensity I_L and the spectral overlap $g_\lambda(p, T)$ of the laser profile and the absorption line. All these factors determine the number density of excited molecules and, therefore, the absorption part of the LIF process. The fluorescence quantum yield ϕ in Equation (10) gives the ratio of the spontaneous emission rate (from level k) versus the total rate of (radiative and nonradiative) relaxation processes:

$$\phi = \frac{A_{ki}}{\sum_j A_{jk} + Q_k + P_k} \tag{10}$$

ϕ therefore depends on the Einstein A_{ki} coefficient for spontaneous emission divided by the sum of the rates of all depopulation processes of the excited state (spontaneous emission from state k to all possible ground state levels: A_{kl}, quenching $Q_k(p)$ and predissociation P_k). The detection efficiency of fluorescence photons depends on the observed solid angle $\Omega/4\pi$, the transmission of the optical system ε and the response of the detector η. Effects of polarization are discussed by Rothe et al.[27]

When quantifying signal intensities obtained from linear LIF one usually has to account for the temperature-dependent population of the ground state f_B and the temperature- and pressure-dependent spectral overlap g_λ. Furthermore, at atmospheric pressure and above (for nonpredissociative states, see below) the fluorescence quantum yield is strongly influenced by the quenching rate Q, which depends on temperature, pressure and local gas composition. The temperature dependence of the ground-state population can be minimized by choosing a transition with minimized temperature sensitivity in the expected temperature range. The spectral overlap factor can be calculated from spectra simulations which are available for the most important combustion relevant species.[28,29] This is mainly important in high pressure applications where collision broadening causes the individual rotational lines to blend giving absorption features which are spectrally broader than the laser line. At atmospheric pressure and above, the denominator in the fluorescence quantum yield ϕ (see Equation 9) is dominated by fluorescence quenching ($Q \gg A$) and may introduce large uncertainties when not corrected for.

2.1.2.2 Saturated Laser-induced Fluorescence In the fully saturated regime the population of the excited level reaches steady state and depends only on the ratio of the probabilities of stimulated absorption and emission (B_{12}/B_{21}). The influence of other depopulation processes of the excited state like collisional quenching disappears and the fluorescence intensity I_{LIF} is described according to Equation (11):

$$I_{\mathrm{LIF}} = N(p, T) f_{\mathrm{B}} V \frac{B_{ki} A_{ik}}{B_{ki} + B_{ik}} \eta \varepsilon \frac{\Omega}{4\pi} \tag{11}$$

Therefore, the fluorescence intensity only depends on the number of excitable molecules $N f_{\mathrm{B}} V$, on molecular constants and on the detection probability defined by the experimental set-up. From the two-level assumption saturation is reached at the saturation spectral irradiance I_{sat} (Equation 12), which at atmospheric pressure is in the range of $1.3 \times 10^6\,\mathrm{W\,cm^{-1}cm^{-2}}$ for OH and $1.2 \times 10^7\,\mathrm{W\,cm^{-1}cm^{-2}}$ for NO (see references in Eckbreth[7]):

$$I_{\mathrm{sat}} = \frac{(A_{21} + Q_{21})c}{B_{12} + B_{21}} \tag{12}$$

Experimentally however, full saturation is not easy to obtain since both temporal and spatial contributions from nonsaturated LIF have to be excluded from signal detection. With very short detection times excluding the beginning and the end of the laser pulse, and referencing of LIF acquired along two different optical paths simultaneously (two-optical path laser-induced fluorescence[30] (TOPLIF)), saturated LIF measurements are feasible.

2.1.2.3 Laser-induced Predissociative Fluorescence Curve crossings of bound and dissociative states in the excited level cause a very short lifetime of the excited molecule. The predissociation rate is a molecular constant and, therefore, independent of pressure and colliding molecules. If choosing appropriate excited levels (i.e. OH A, $v' = 3$ or O_2 B-state) the denominator in the fluorescence quantum yield (Equation 10) is dominated by P with $P \gg Q$. In this case, the quenching influence can be omitted leaving the fluorescence quantum yield independent of pressure and temperature. However, since Q is rising with the collision frequency, the pressure range where the assumption of LIPF (laser-induced predissociative fluorescence) applies is limited.

2.1.2.4 Computation of Quenching Rates For a large number of colliders quenching rates for, e.g. OH and NO have been determined as a function of temperature and pressure enabling the development of simulation models.[31,32] Therefore, quenching can be quantified as long as local gas composition and temperature are known. However, close to the flame front where composition changes quite dramatically, because of transient species like atoms and radicals, concentration data are not available from experiment. Chemical kinetics simulations of laminar flames using full chemistry allow estimation of the influence of the transient species in the flame front. In the burned gas region, however, where concentrations are close to equilibrium and where transient species are absent, the database provided by the quenching models is good for correcting LIF intensity data.

2.1.2.5 Time-resolved Laser-induced Fluorescence The fluorescence quantum yield ϕ can also be expressed as a ratio of fluorescence lifetimes according to Equation (13)

$$\phi = \frac{\tau_{\mathrm{eff}}}{\tau_{\mathrm{rad}}} \tag{13}$$

where τ_{rad} is the fluorescence lifetime related to the spontaneous emission rate only. τ_{eff} is the effective fluorescence lifetime in the system under study, influenced by all depopulation processes. Therefore, with a measurement of τ_{eff} the fluorescence quantum yield can be calculated and from Equation (9) concentration information can be gained without detailed knowledge about the depopulation processes of the excited state. In low-pressure systems this method is frequently applied since timescales are sufficiently long to use nanosecond lasers for excitation. At atmospheric pressure picosecond laser systems and ultrafast detectors have to be used.[33,34] When calculating quenching rates from time-dependent LIF measurements the simple two-level model must be expanded including neighboring rotational states which are populated via collision induced rotational energy transfer (RET) for both the ground state and the excited state. The two-level model is sufficient when either no RET is present or full thermalization over the complete level manifold can be assumed. For intermediate cases however, four-level models[7] have been suggested allowing for time-dependent calculation of LIF signals.

2.1.2.6 Energy Transfer Processes Besides fluorescence quenching which describes nonradiating transitions to the ground state, other collisionally induced energy transfer processes have to be considered. Electronic energy transfer (EET), vibrational energy transfer (VET) and RET populate excited levels which subsequently fluoresce. This especially gains importance if levels with significantly different fluorescence lifetimes are involved. Furthermore, the fluorescence signal may be shifted to different spectral regions compared to the direct transitions from the laser-populated level. The implications of these effects are discussed in further detail by Eckbreth and by Rothe et al.[7,27]

2.1.2.7 Quantification of Laser-induced Fluorescence Signal Intensity With all the considerations mentioned above, from Equations (9) or (11) semi-quantitative concentration data can be derived from LIF signal intensities. Quantification, however, requires knowledge of the detection efficiency $\eta\varepsilon\Omega/4\pi$ determined by the detection system. These factors are usually obtained via calibration procedures. Different attempts at calibration have been applied using Raman or RS from a known number density in the probe volume with excitation and detection channels being exactly equal to the LIF measurements[35] to correct for the visualized solid angles and detection efficiency of optics and detectors. However, to include transmission effects of filters used for detection it is advantageous to calibrate with the species that is to be measured. Therefore, calibration cells filled with known gas mixtures are applied for stable species.[36] Transient species on the other hand can be calibrated by using standard flames where species concentrations and temperatures obtained, respectively, from absorption and CARS (coherent anti-Stokes Raman scattering) measurements are well documented for different operating conditions.[37] In the case of NO, calibration is possible by seeding known quantities of NO to the flame under study.[38] In stationary combustion, objects like laminar flames absorption measurements were carried out to assess absolute concentrations.[39]

2.1.2.8 Laser-induced Fluorescence Thermometry The temperature-dependent population of different ground state levels allows temperatures to be assessed by LIF. In stationary systems local temperatures can be obtained by detecting and simulating fluorescence excitation spectra, for example of NO.[40] When choosing two transitions where the ground state levels have an appropriate energy difference $(F_2 - F_1)$, two-color excitation[41] provides a possibility for time-resolved temperature measurements which can be extended to two-dimensional temperature mapping. Temperature T is calculated according to Equation (14) from the fluorescence intensities $I_{\mathrm{LIF},1}$ and $I_{\mathrm{LIF},2}$ obtained upon excitation with the respective laser wavelengths:

$$T = \frac{F_2 - F_1}{k[\ln(I_{\mathrm{LIF},1}/I_{\mathrm{LIF},2}) + c_{\mathrm{calib}}]} \tag{14}$$

k_{B} is the Boltzmann constant and c_{calib} is a constant obtained from independent calibration measurements. Ideally the excited levels populated upon excitation of either laser photon should be identical. In this case all depopulation processes in the excited state like RET and quenching are the same for both transitions and cancel out when calculating temperatures. Local concentration cancels out as well. However, saturation and ground state depopulation have to be considered very carefully[27] to prevent large systematic errors. Two-line temperature measurements have been shown using OH,[42] O_2[43] and NO.[44] Another possibility for temperature measurement is provided by molecular tracers like ketones, which exhibit a shift in their absorption spectrum with temperature which can be used for two-line temperature measurements in inhomogeneously-mixed combustion systems.[45,46] Single-line techniques give information about the local gas density which can be transferred to temperature when pressure and species concentration are known (and spatially invariant).

2.1.3 Rayleigh Scattering

RS takes advantage of the statistical density fluctuations in a polarizable medium. For an electromagnetic wave travelling in y-direction with its $\boldsymbol{E}$ vector linearly polarized in the z-direction the induced dipole moments p are proportional to the electric field and to the respective tensor component of the polarizability α (of an assumed space fixed scattering molecule) which causes scattered light to be polarized in the y- or z-direction when observed in the x-direction in Equation (15)

$$\begin{aligned} p_y &= \varepsilon_0 \alpha_{yz} \boldsymbol{E}_z \\ p_z &= \varepsilon_0 \alpha_{zz} \boldsymbol{E}_z \end{aligned} \tag{15}$$

The radiant intensity of scattered light into a unit solid angle perpendicular to the propagation direction of the incident light beam of an ensemble of arbitrarily oriented molecules is proportional to the square of the induced dipole moment, and can be written as (Equation 16)[7]

$$\begin{aligned} I_{yz}^{\Omega} &= \frac{\mathrm{d}P(\pi/2)}{\mathrm{d}\Omega} \propto \overline{(\alpha_{yz})^2}\boldsymbol{E}_z^2 = \frac{1}{15}\gamma^2 \boldsymbol{E}_z^2 \\ I_{zz}^{\Omega} &= \frac{\mathrm{d}P(\pi/2)}{\mathrm{d}\Omega} \propto \overline{(\alpha_{zz})^2}\boldsymbol{E}_z^2 = \frac{1}{45}(45a^2 + 4\gamma^2)\boldsymbol{E}_z^2 \end{aligned} \tag{16}$$

where a and γ are the mean value and the anisotropy of the polarizability tensor. In a quantum mechanical derivation the harmonic oscillator approach for the oscillating charge leads to the ν^4-dependence of the scattered radiant intensity (Equation 17):

$$I_{nm}^{\Omega} = \frac{\pi^2}{2\varepsilon_0 c^3}(\nu_0 + \nu_{nm})^4 p_{nm}^2 N_1 \tag{17}$$

with a proportionality to the number of species N_1, to the squared matrix element of the induced transition p_{nm} from state n to m and to the fourth power of the frequency of the exciting radiation ν_0. For RS $m = 1$ and one arrives at Equation (18) for the scattered radiant power of the Rayleigh scattered light[7]

$$P_{\mathrm{RS}} = P_L n \left(\frac{\partial \sigma}{\partial \Omega} \right)_{\mathrm{mix}} \Omega l \varepsilon \tag{18}$$

where P_L is the incident laser power and n the number density of scatterers in the probe volume sampled along a length l. The RS cross-section of the gas mixture can be calculated from the mole fractions x_i of the individual species in the mixture (Equation 19):

$$\left(\frac{\partial\sigma}{\partial\Omega}\right)_{\text{mix}} = \sum_i x_i \left(\frac{\partial\sigma}{\partial\Omega}\right)_i \tag{19}$$

Therefore, from the Rayleigh intensity data the density in mixtures with known scattering cross-section and the temperature (and pressure for a known temperature) can be determined on a single shot basis in a one-(collimated laser beam) or two-(laser light sheet) dimensional area.

2.1.4 Spontaneous Raman Scattering

While many combustion species can be detected by single photon absorption techniques, there are important exceptions for species at major concentration levels (homonuclear molecules with singlet-ground states like H_2 and N_2), which have to be monitored by, for instance Raman spectroscopy or two-photon techniques. Since almost all molecular species exhibit Raman active vibrational modes, SRS has developed to the method of choice for the determination of major species concentrations in turbulent combustion processes. Although inherently a weak process (the ratio of the captured Raman signal intensity to the incoming laser intensity, $I_{\text{SRS}}/I_{\text{L}}$, amounts to approx. 10^{-14}), and being susceptible to stray light, with high power KrF-, Nd:YAG- or dye lasers, pointwise[47] or spatially resolved measurements along the laser beam axis are possible within a single exciting light pulse. The Raman radiant intensity in Equation (20) emitted from a single species for a Q-branch transition ($\Delta v = +1$, $\Delta J = 0$) into a small solid angle Ω,

$$P_{\text{SRS}} = P_L n \left(\frac{\partial\sigma}{\partial\Omega}\right)_{zz} \Omega l \varepsilon \quad \text{with} \quad \left(\frac{\partial\sigma}{\partial\Omega}\right)_{zz} = \frac{I^{\Omega}_{zz}}{N I_{\text{L}}} \tag{20}$$

$$I^{zz}_{\text{SRS}} = \frac{hN(\nu_0 - \nu_{12})^4}{8mc^4[1 - \exp(hcw_{\text{e}}/kT)]} \left\{(a')^2 + \frac{4}{45}(\gamma')^2\right\} I_{\text{L}} \tag{21}$$

is determined by the differential Raman scattering cross-section, $(\partial\sigma/\partial\Omega)_{zz}$, the laser radiant power P_L and the illuminated and imaged probe length l. In Equation (21) a' and γ' are the respective derivatives of a and γ with respect to the spatial coordinate of the Raman excited vibration. This enables mixture fractions to be readily determined from single pulse measurements.

2.1.5 Laser-induced Incandescence

The exploration of the potentials of laser spectroscopic techniques in sooting flames is much less advanced than in clean particulate-free environments. The possibility for absorption or scattering of laser radiation in penetrating combustion media makes diagnostic methods a challenge in fuel rich and sooting environments. One technique with great prospects for obtaining spatially resolved one- and two-dimensional imaging of soot distributions is laser-induced incandescence (LII).[48–51] In this technique the soot particles are exposed to a high flux of radiation, usually from a fundamental or frequency doubled Nd:YAG laser beam, whose absorption leads to their strong heating with subsequent total evaporation and cooling. The time-resolved detection of the emitted blackbody radiation constitutes a quantitative measure of the soot volume fraction if an independent calibration by extinction was performed.[52,53] In an elegant experimental set-up this calibration was done simultaneously with the LII signal generation by splitting the two-dimensional laser sheet into a low intensity part (for extinction) and a high intensity part (for LII). Both were sent through the same spatial region in the turbulent flame with a temporal separation of several nanoseconds and then the required images (2 for extinction, 1 for LII) were placed onto separate partitions of the CCD (charge coupled device) camera chip.[54]

The detected LII signal is given by Equation (22)[52]

$$I_{\text{LII}} = \int_{t_1}^{t_2} J(t, q\lambda_{\text{em}})\,\mathrm{d}t \tag{22}$$

where $(t_2 - t_1)$ is the opening time of the intensifier gate of the CCD camera which is activated at time t_1 after the laser pulse. The temporal evolution of the soot particle size distribution function, $J(t, q, \lambda_{\text{em}})$, is dependent in a complicated manner on the chosen emission wavelength λ_{em} in the Planck equation for the blackbody radiation flux q[52] through the conductive heat loss and radiation balance of the superheated particle. Through analysis of the temporal profile of the LII signal intensity the particle size can be deduced as an important parameter in the history of particle growth in the flame.[55]

2.2 Nonlinear Techniques

Nonlinear wave mixing phenomena usually are not observed at the low light intensities of incoherent thermal radiators. However, the development of high-power tunable laser systems has triggered the study of numerous new nonlinear optical phenomena in liquid, solid and gaseous samples. The attained signal intensities in molecule specific resonantly enhanced higher order wave mixing processes can be large due to their strong dependence on the intensities of all incoming beams, the concentration of the sample molecule and the interaction length (e.g. the effective spatial extent in the sample where the signal is produced inside the

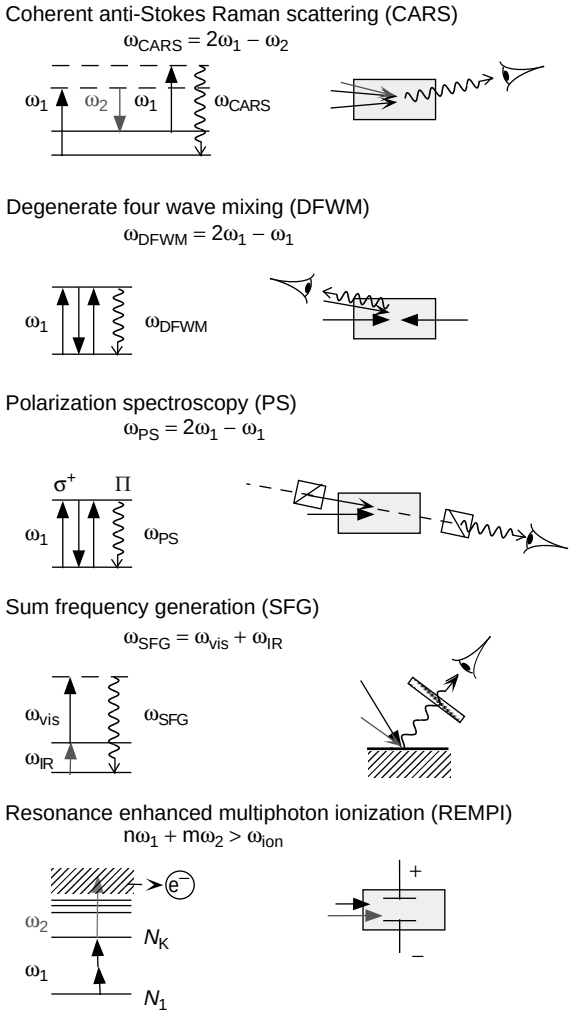

Figure 2 Energy level diagrams showing photon–molecule interaction (left) and schematic experimental set-up (right) for nonlinear laser spectroscopic techniques discussed in the text. See Figure 1 caption for more detail.

interaction region of all beams). The energy level schemes of multibeam techniques most often established in combustion diagnostics today are summarized in Figure 2 together with the sketched experimental set-ups. The nonlinear susceptibilities are enhanced whenever one or more of the frequencies of the interacting laser beams coincide with allowed one- or multiphoton transitions. In the macroscopic picture the induced higher order polarization terms arise from the nonlinear mixing of the input waves and give rise to the radiation of the same or various new sum and/or difference frequency combinations of the frequencies of the input waves. This produces a coherent signal beam in a direction determined by the phase matching condition, Equation (23)

$$\vec{k}_{\text{sig}} = \sum \vec{k}_j \tag{23}$$

where $|\vec{k}_j| = n_j w_j / c$, is the magnitude of the respective beam wave vector. Through focusing of the beams good spatial resolution transverse to the beam propagation direction can be obtained with sample volumes of several cubic micrometers without sacrifice of signal intensity loss due to larger crossing angles.

2.2.1 Coherent Anti-Stokes Raman Scattering

Among the coherent Raman processes CARS is almost exclusively being used in practical combustion diagnostics. In CARS experiments (compare Figure 2) one fixed frequency (pump) and one broadband (Stokes) laser beam in the visible region with their frequency difference equal to a Raman allowed transition are aligned and focused into the sample to create, in a four-wave mixing process, a coherent beam at the anti-Stokes frequency $(2w_1 - w_2)$, well-separated spatially and spectrally from all incoming beams. The CARS radiation then is dispersed in a spectrometer and detected with a CCD camera (Figure 3). In the nonsaturated regime the CARS intensity at frequency $w_{\text{CARS}} = w_3 = 2w_1 - w_2$ for an isolated transition[7] in Equation (24)

$$I_{\text{CARS}} = I_3 = \frac{w_3^2}{n_1^2 n_2 n_3 c^4 \varepsilon_0^2} (lN\Delta_j)^2 I_1^2 I_2 |\chi^{(3)}_{\text{CARS}}|^2 \tag{24}$$

with

$$\chi^{(3)}_{\text{CARS}} = K_j \frac{\Gamma_j}{2\Delta w_j - i\Gamma_j} + \chi_{\text{nr}}$$

$$K_j = \frac{w_3^2 (4\pi)^2 n_1 \varepsilon_0 c^4}{n_2 \hbar w_2^4 \Gamma_j} \left(\frac{\partial \sigma}{\partial \Omega} \right)_j \tag{25}$$

is dependent on higher powers of the pump (I_1) and Stokes (I_2) laser intensities as well as on the square of the number density difference $N\Delta_j$ between lower and upper Raman level of the jth transition. Equation (24) is written for perfect phase matching of the beams, i.e. when Equation (23) holds and the corresponding signal intensity is maximized. Γ_j is the total broadening rate of the transition and $\Delta w_j = w_1 - w_2 - w_j$ the detuning from the Raman transition frequency. The spectral structure of the CARS signature over thermally populated levels is contained in the expression of the third-order susceptibility $\chi^{(3)}$ (Equation 25), which, therefore, is an accurate measure of the gas temperature. This is exploited for thermometry applications through least squares fitting of computer generated to experimental CARS spectral shapes, if proper account is taken of the various physical effects contributing to the CARS spectral signature, such as coherences in the CARS pump beams,[56,57] saturation and collisional effects[58,59] or the noise characteristics when broadband Stokes lasers are employed.[60]

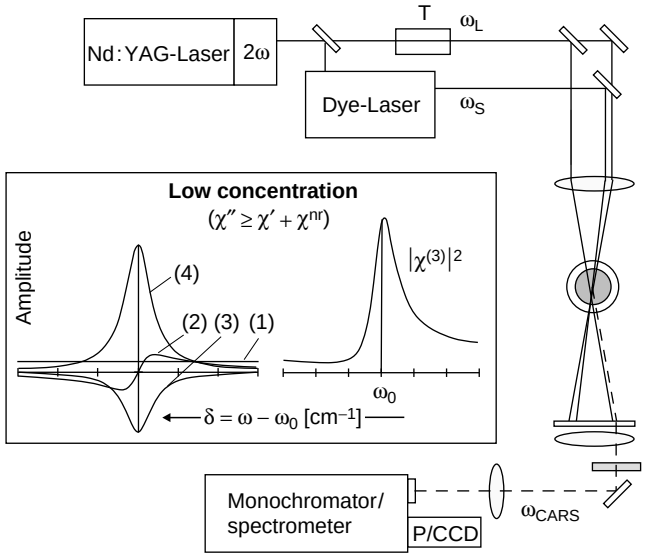

Figure 3 Experimental arrangement for a CARS spectrometer using either a narrowband scanning or broadband dye laser and a photomultiplier (P) or a CCD camera © as a detector, respectively. The inset shows a typical CARS lineshape function for an isolated transition at low concentration of the resonant species. It is composed of the constant nonresonant background (1), the interference contributions with χ_{nr} (2,3) and the Lorentzian-shaped resonant part (4) of the third order susceptibility $\chi^{(3)}$.

In their pioneering work Regnier and Taran[61] recorded the first CARS spectra of molecular hydrogen in an atmospheric pressure Bunsen flame using a narrowband scanning dye laser as a Stokes beam. Eckbreth[62] introduced single pulse temperature and species measurements by using a broadband Stokes dye laser which covers Raman shifts of the whole investigated spectral branch. Species concentrations can be determined from either the integrated CARS line intensity or its spectral shape[63,64] through calibration measurements and spectral modeling, respectively. For low concentrations the (real) nonresonant contribution χ_{nr} in the susceptibility expression, Equation (25), is significant and limits the detection sensitivity due to interference with the resonant contribution. To illustrate this behavior, the real and imaginary susceptibility components determining the CARS lineshape are depicted in the inset of Figure 3 for an assumed low concentration of the resonant species in a nonresonant buffer gas.

Owing to the $\Delta v = 1$, $\Delta J = 0$ selection rules in vibrational Q-branch Raman spectra, individual lines are closely spaced and different laser dyes are needed to cover the whole Stokes wavelengths to detect different species. Alternatively, Aldén et al. excited the pure rotational transitions to enable temperature and species concentration measurements to be made in the single shot broadband rotational-CARS technique.[65,66] The dicyanomethylene-2-methyl-6-(*p*-dimethylaminostyryl)-4H-pyran (DCM) laser dye covers most of the rotational Stokes shift even for spectra at 1000 K or higher, and multiple species can be detected within the same spectral range as well. Since both, first pump- and Stokes beam in the respective level scheme (cf. Figure 2) originate from the same Stokes laser, additional spectral noise averaging takes place which makes this technique less susceptible to mode noise in the laser sources. Problems in this technique are stray light from the pump beam as well as a lower-temperature measurement accuracy at high temperatures.

2.2.2 Degenerate Four Wave Mixing

Degenerate four wave mixing (DFWM) has emerged as a spatially resolving diagnostic technique with high sensitivity, which can be applied to all species exhibiting an absorption spectrum. Since the method does not need the species to fluoresce, it constitutes an alternative to LIF. As a special case of more general multicolor resonant four wave mixing techniques, DFWM can be visualized as a laser-induced grating (LIG) process,[67,68] in which two beams of equal intensity at wavelength λ are crossed at a half angle θ in the probe volume. If the wavelength of these ''writing'' beams is tuned into resonance with a molecular transition, then for parallel linear electric field polarizations, a spatially sinusoidally modulated population grating (PG) with wavelength Λ equal to $\Lambda = \lambda/(2\sin 2\theta)$ will be formed by absorption in the interference region of both beams. This PG can be 'read out' by a third, weak probe beam at the same wavelength under the phase matching angle. From a perturbative treatment of the nonlinear interaction process in a two-level atom with a saturating pump- and weak probe and signal beams by Abrams and Lind,[69] the spectrally integrated DFWM signal intensity (Equation 26) can be written as[7,70]

$$I_{\mathrm{DFWM}} = \frac{3\pi}{2T_2}(k_{12}l)^2 \frac{4(I/I_{\mathrm{sat}}^0)^2}{[1+4(I/I_{\mathrm{sat}}^0)^2]^{5/2}} I_{\mathrm{pr}}$$
$$k_{12} = \left(\frac{wT_2}{2\hbar c\varepsilon_0} \quad |\mu_{12}|^2 N_1 \right. \tag{26}$$

In this expression N_1 is the population of the lower electronic state, and μ_{12} the dipole transition moment between the excited (index 2) and ground (index 1) energy levels coupled by the laser radiation. I_{sat}^0 is the saturation intensity at line center (Equation 27):

$$I_{\mathrm{sat}}^0 = \frac{\hbar c\varepsilon_0}{2T_1T_2|\mu_{12}|^2} \tag{27}$$

T_1 and T_2, respectively, denote the population and coherence decay times of the coupled transitions. From

Equation (26) it is obvious that larger signal intensities can be obtained by increasing the intensities of the incoming beams, although saturation effects will make quantitative interpretation of experimental results difficult.[71,72] In the limit of laser intensities much larger than the saturation intensity of the probed transition, the expression for the DFWM signal intensity (Equation 26) reduces to[70]

$$I_{\mathrm{DFWM}} = \frac{k_{12}^2}{T_2}\left(\frac{I}{I_{\mathrm{sat}}}\right)^{1/2} I_p \propto \mu_{12}^3 N_1^2 \left(\frac{T_2}{T_1}\right)^{1/2} I^{1/2}\, I \gg I_{\mathrm{sat}} \tag{28}$$

which shows a reduced sensitivity on the dipole moment μ_{12} and relaxation times T_i of the transition. Since in practical measurements laser intensities often are high enough at least partially to saturate the transitions, the DFWM signal intensity varies with laser intensity in between those given by Equations (26) and (28), and the correct power dependence of the dipole moment has to be determined experimentally.[73,74] In addition, some theoretical modeling of saturated DFWM has been performed to quantify temperature and concentration measurements for NO and OH in flames at various pressures.[75,76]

Also, the strong pressure dependence of the DFWM signal intensity in the linear regime can be reduced by working at saturating laser intensities. As has been confirmed by experiment[70,74] and numerical calculations,[71,76] a most useful range is $I_{\mathrm{p}} \approx 2I_{\mathrm{sat}}$. Modeling of the pressure dependence of the DFWM signal at saturating laser intensities, using direct numerical integration (DNI) instead of the stationary absorber model of Abrams et al. gave good agreement with corresponding measurements with NO.[71] The capability of DFWM to probe nonfluorescing species sensitively has been exploited for monitoring highly predissociative states such as in CH_3[78] in flames. Since the increasing availability of tunable infrared radiation from OPOs the opportunity to monitor hydrocarbons (CH_4, C_2H_2) via DFWM by excitation of their characteristic rovibrational transitions has been exploited.[79] If narrow bandwidth tunable lasers are employed in the phase conjugated beam geometry, where both pump beams are counterpropagating, sub-Doppler DFWM lineshapes can be obtained. This is demonstrated in Figure 4, where individual LIF (solid line without dots) and DFWM rotational lines from the O-branch of the $A^2\Sigma$–$X^2\Pi(0, 0)$ transition of NO at 226 nm are compared at low pressure.[77] In contrast to the DFWM spectrum, which shows almost no background interference, the LIF profile is dominated by the room temperature Doppler line profile with superimposed saturation dips at line center, where the frequencies from all beams observed in the molecular frame of reference coincide.

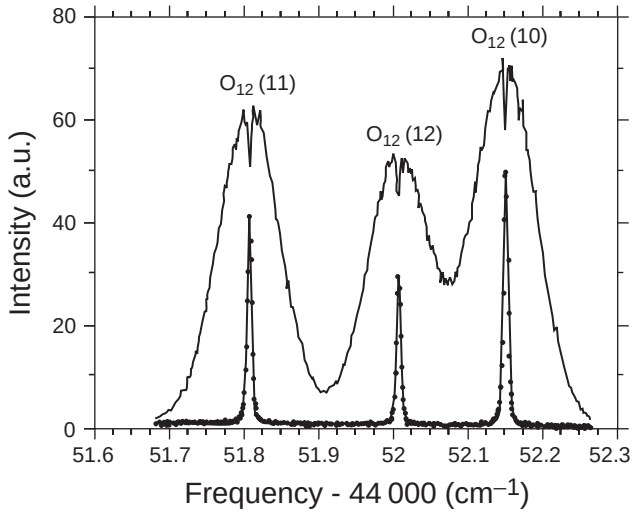

Figure 4 Comparison of LIF (pure solid line) and DFWM (line with dots) lineshapes in the O-branch spectrum of the $A^2\Sigma$–$X^2\Pi(0, 0)$ electronic band of NO. Using a narrowband laser system the Doppler broadened LIF and the almost Doppler-free DFWM lineshapes acquired in the phase conjugated beam geometry[77] are clearly distinguished. Total NO pressure in the cell was 2.7 Pa.

At pressures high enough to quench efficiently the excited state population created by the two grating forming beams within the laser pulse duration, the excitation energy is translated into heat, which gives rise to a thermal grating (TG) signal in the medium.[68] The purely TG signal contribution can be probed in the direction of the Bragg angle ϕ_{B} for the grating wavelength Λ, $\sin\phi_{\mathrm{B}} = \lambda_{\mathrm{pr}}/2\Lambda$, if the wavelength λ_{pr} of the reading beam is not resonant with any transition in the molecule. These data enable single shot measurements to be made of the sound velocity and thermal diffusivities in a localized volume of the sample at high pressures. Efficient electronic quenching of excitation energy, therefore, is beneficial in LIG spectroscopy compared to LIF.[80] The method also has been applied for the detection of OH in a high-pressure premixed methane/air flame. Through fitting of the experimental LIG temporal signals trace good agreement of the gas temperature with nitrogen CARS thermometry at the same location in the flame has been obtained.[81]

2.2.3 Polarization Spectroscopy

Polarization spectroscopy (PS) is a two-beam technique, in which one pump and one probe beam are crossed in the sample and the depolarization of the probe beam is monitored through crossed polarizers.[6] The first implementation of PS by Teets et al.[82] showed it to be a versatile alternative to high-resolution gas phase saturation spectroscopy. For combustion diagnostics

Nyholm et al.[83] demonstrated its potential as a spatially resolved and background free absorption technique with good sensitivity for the detection of minor species.

The principles of PS are outlined in early publications[82,84] and textbooks.[6] The pump beam creates an anisotropy in the sample by selective absorption within the degenerate magnetic sublevels of the probed transition. This anisotropy causes rotation of the plane of polarization of the probe beam and a depolarization through differential absorption and refraction for the left and right circularly polarized components of the linearly polarized probe light, respectively, which is detected through a crossed polarizer in the probe beam path. For this case the transmitted beam intensity can be described according to Equation (29) as[82]

$$I_{\mathrm{PS}} = I_0 \left[\xi + \theta^2 + b^2 + \frac{1}{2}\frac{\theta \Delta k_{12} l}{1+x^2} + \frac{1}{2} b \Delta k_{12} l \frac{x}{1+x^2} + \left(\frac{1}{4} \Delta k_{12} l \right)^2 \frac{1}{1+x^2} \right] \tag{29}$$

which is a combination of Lorentzian and dispersive contributions to the total line profile: the first three terms are frequency independent and are caused by residual transmission of the polarizers (ξ), their accidential uncrossing (θ) and a possible residual dichroism (b) of the optics between both polarizers in the probe beam path. The differential absorption coefficient for left and right circularly polarized light, Δk_{12}, at resonance $x = w_{12} - w_l = 0$ is given by Equation (30)

$$\Delta k_{12}(x=0) = \Delta k_{12}^0 = k^+ - k^- = k_{12}^0 S_0 C_{jj_1}^* \tag{30}$$

where S_0 is a saturation parameter and $C_{jj_1}^*$ are Clebsch–Gordan coefficients for the respective transition and coupling cases. For sensitive species detection ξ, b and θ should be minimized, which makes this diagnostic difficult when probing, for example, through optical windows. Depending on the relative importance of these terms in I_{PS}, purely dispersive or Lorentzian lineshapes can be obtained. For the special arrangement with perfectly orthogonal beam polarizations ($\theta = 0$), and stress free window mounting ($b = 0$) the signal intensity at line center ($\Delta x = 0$) and for a linearly polarized pump beam is given by Equation (31):

$$I_{\mathrm{PS}} = I_0 \mathrm{e}^{-k_{12} l} \left[\xi - \frac{1}{2}\theta^* k_{12} l + \left(\frac{\Delta k_{12} l}{4} \right)^2 \right] \tag{31}$$

2.2.4 Sum Frequency Generation

Many combustion processes are influenced and can be guided by heterogeneous chemical reactions. Prominent examples are the exhaust gas treatment in automotive engines for reduced emission of hydrocarbons and NO_x for better fuel economy and environmental acceptance. To obtain most direct information on the chemistry in heterogeneous catalytic combustion processes, apart from linear absorption/reflection UV/visible and infrared spectroscopy, surface sensitive laser spectroscopic probes are of great value. Sum frequency generation (SFG) provides such possibilities since it is surface and species selective within the whole technically interesting pressure range.[85]

SFG was first observed in anisotropic crystals, where the proper superposition of two laser beams at frequencies w_1 and w_2 can generate a second order polarization $P^{(2)}(w_3 = w_1 + w_2)$, which gives rise to the coherent radiation of a third beam at the sum of both frequencies.

Because of symmetry considerations, second order phenomena only arise to a significant extent in non-centrosymmetric media. In reflection from a surface this condition is not satisfied for the bulk material but only for a small surface layer of thickness $\sim\lambda/2\pi$, where the symmetry is broken. As in any nonlinear spectroscopic technique, the SFG signal intensity can be enhanced several orders of magnitude if the waves interact resonantly with a vibrational mode at w_v of species which form the molecular surface layer. This property has triggered numerous activities using SFG as a sensitive surface-specific monitor for probing surface coverage, molecular structure and chemical reactions at solid or liquid interfaces. SFG, therefore, is of fundamental interest for the study of heterogeneous combustion and catalysis, sensor development, etc.

The SFG intensity at frequency $w_{\mathrm{SFG}} = w_{\mathrm{vis}} + w_v$ is proportional to the squared modulus of the second order susceptibility, $\chi_{\mathrm{SFG}}^{(2)} = \chi_{\mathrm{r}}^{(2)} + \chi_{\mathrm{nr}}^{(2)}$, and can be written for a surface adsorbed species with n active vibrations according to Equation (32) as[86,87]

$$I_{\mathrm{SFG}} = K_{\mathrm{sys}} |F(\chi_{\mathrm{nr}}^{(2)} + \chi_{\mathrm{r}}^{(2)})|^2 \quad \text{with} \quad \chi_{\mathrm{r}}^{(2)} = \sum_{v=1}^{\mathrm{n}} \chi_{v,\mathrm{r}}^{(2)} \tag{32}$$

In this expression K_{sys} is a geometry-dependent factor which also contains a correction for the local field effects in the interaction layer. The nonlinear optical susceptibility, χ_{SFG} is the macroscopic form of the molecular hyperpolarizability β. From a quantum mechanical derivation of this expression, Equation (33)

$$\beta_{v,lmn}(w_{\mathrm{SFG}}) = \frac{A_{v,lm} M_{v,n}}{\hbar(w_v - w_{\mathrm{IR}} + i\Gamma_v)} \tag{33}$$

it can be seen that since both the infrared ($M_{v,n}$) and Raman ($A_{v,lm}$) transition moments to appear, the vibrational mode to be excited in the surface bound molecule must be infrared and Raman active.

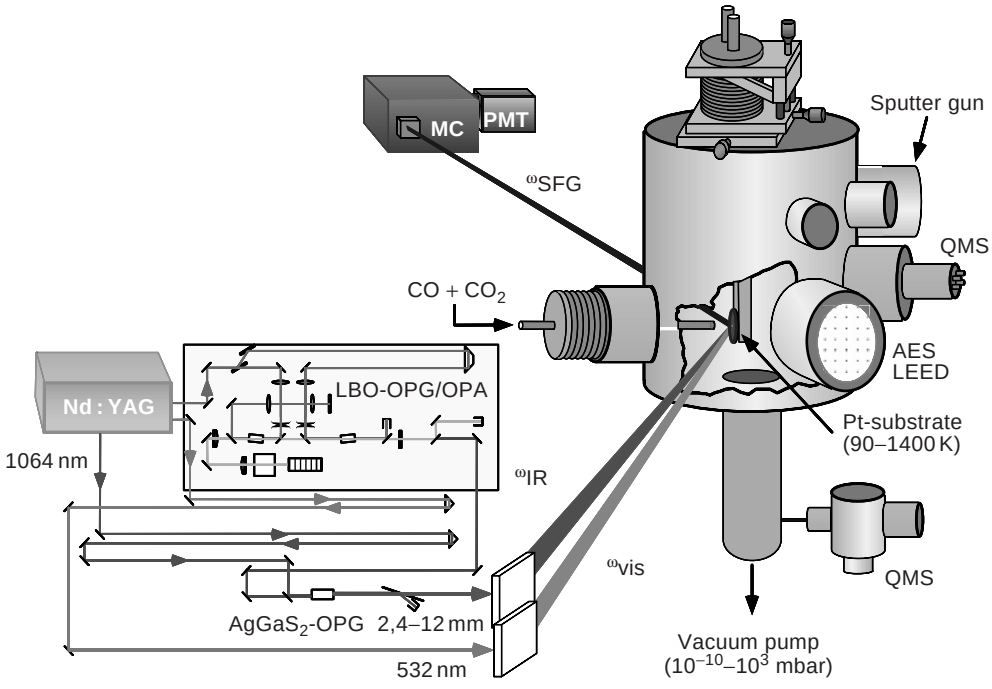

Figure 5 Experimental arrangement for SFG for the in situ detection of chemisorbed molecules on platinum surfaces. The visible fixed frequency (532 nm) and tunable infrared beam are derived from the second harmonic of a ps-Nd : YAG-laser and a lithium borate (LBO) optical–parametric generator/amplifier (OPG/OPA), respectively. AgGaS2-OPG, silver-gallium-sulfide optical parametric generator; QMS, quadrupole mass spectrometer; AES, Auger-electron spectrometer; LEED, low energy electron diffraction; MC, monochromator; PMT, photomultiplier tube.

Figure 5 depicts an experimental example of the use of a SFG spectrometer to study chemical reactions on surfaces. It consists of a laser system which provides fixed visible (532 nm from a frequency doubled Nd : YAG-laser) and tunable infrared (2–10 μm) wavelengths in short (30 ps) high-intensity pulsed beams that generate the SFG signal in the visible for detection with a filtered photomultiplier. The two beams are sent through CaF_2 windows to the specimen surface mounted in a high-vacuum chamber for preparation and analysis. The system is capable of investigating desorption phenomena on surfaces, bridging the pressure gap from ultrahigh vacuum up to 1 atmosphere (1 atmosphere (technical) = 10^2 kPa).

2.2.5 *Multiphoton Absorption Spectroscopy*

Several constraints limit the direct excitation of single photon transitions in atoms or light molecules in flame environments: deep UV radiation can be absorbed nonspecifically by other species in the flame because of the dense electronic bands in this wavelength range. This can cause interfering fluorescence or bias absorption measurements. An additional complication arises from laser-induced chemistry at short wavelengths, that is, the production of reactive atoms through multiphoton absorption (MPA) and subsequent photodissociation or ionization processes during the laser pulse duration.[88] The first constraint can be circumvented by MPA spectroscopy, where visible or near UV laser radiation can be used. Owing to the high intensity of focused laser beams this method still shows sufficient sensitivity as a viable combustion diagnostic technique. As an example, the MPA technique enabled detection of CO at 230 nm[89] via the strong $B^1\Sigma^+ \leftarrow\leftarrow X^1\Sigma^+$ two-photon absorption, and of NO[90] in atmospheric pressure flames. Quenching and possible predissociation of the upper excited state limit detection sensitivity and make calibration procedures necessary.[91]

The transition probability $W^{(n)}$ for n-photon transitions is determined by the MPA cross-section $\sigma^{(n)}$, the relevant laser intensities $I_i(w_i)$ and their respective photon energies $\hbar w_i$ (Equation 34):[86]

$$W^{(n)} = \frac{\sigma^{(n)} I_1(w_1) \ldots I_n(w_n)}{\hbar^n w_1 \ldots w_n} \tag{34}$$

Crosley and Smith[92] determined experimentally an absolute two-photon cross-section of approximately 10^{-50} cm^4 s for excitation in the $A^2\Sigma^+ - X^2\Pi_i(0,0)$ electronic system of OH in the burnt gases of a methane air flame. This value already is 3–4 orders of magnitude lower than a corresponding two-photon transition

of atoms. Therefore, high pulse energies are beneficial in multiphoton spectroscopy as long as disturbing effects originating from power broadening, gas breakdown and ac-Stark effect can be neglected.

2.2.6 Resonance Enhanced Multiphoton Ionization

Absorption of multiple photons can ionize the probed species and the ions/electrons formed in the last step can be detected with very high sensitivity. This can be done either in situ through measurement of the current between two electrodes or through withdrawal of small gas samples from the probe volume with subsequent ionization and detection in a mass spectrometer. If a $(2+1)$-REMPI (resonance enhanced multiphoton ionization) process is considered (two-photon allowed transition to a resonantly excited intermediate state k at frequency $2w_1$, followed by a one-photon ionization step, see Figure 2), a simplified expression for the ionization current as a measure of species density results in (Equation 35)[6]

$$S_{\mathrm{REMPI}} = N_1 \frac{\sigma_{1k} n_{\mathrm{L1}}^2}{1 + (A_k/\sigma_{kl} n_{\mathrm{L2}})} \tag{35}$$

In this expression, N_1 is the population of the ground electronic state, σ_{1k} the absorption cross-section (per molecule) for the two-photon transition involved, n_{L1} the number of incident laser photons at frequency w_1, σ_{kl} the ionization cross-section from intermediate level k, and n_{L2} the number of photons in the ionizing laser pulse.

3 APPLICATION OF LASER-BASED TECHNIQUES

3.1 Detection Sensitivities

Detection sensitivities attainable in combustion environments are essential parameters for the application of the various laser-based techniques described in the previous paragraphs. These limits cannot be stated generally since they depend on the line strength of the species to be detected, system parameters like absorption length, response time or optical output power,[93] while especially for in situ measurements temperature and pressure in the probe region and disturbing effects of the in situ beam path (dust, background emission) are most important factors for system specific detection limits. A figure of merit for absorption spectrometers is the minimum detectable absorption change (MDA $= \Delta I/I_0$). In a precisely controlled absorption cell an MDA of 10^{-3} is achievable with noncoherent light sources on a routine basis, while the best laser-based spectrometers provide MDAs in the 10^{-8} range, resulting in a dynamic range of 10^2 to 10^7. Table 1 is a far from complete collection of experimental work in the literature with stated detection limits of combustion of relevant minor species for the various techniques introduced here under different pressures and environments to give a general impression of their practical applicability.

Species selectivity is another feature which has improved greatly through the introduction of lasers, since high radiance and small source linewidth are available simultaneously. For the case of continuous wave-lasers (diode- and dye-lasers) the laser linewidth can be of the order of 10 MHz, compared to molecular linewidths in the gas phase of several 100 MHz under low pressure conditions and several gigahertz at atmospheric pressure. Ultimate selectivity can be ensured, if a spectral region, a so-called ''spectral window'', can be found where the species under investigation is the only (or at least the dominant) absorber.

3.2 Laminar Flames

3.2.1 Low-pressure Flames

Laminar premixed flames at low pressure on a flat flame burner constitute an ideal experimental arrangement for studying the interaction of elementary chemical combustion mechanisms. Experimental data on temperature as well as on concentration profiles for stable and unstable species are used to validate and develop further mathematical models which predict these profiles by numerical solution of the underlying conservation equations, including convection and molecular transport processes.

Figure 6(a) shows such an arrangement used for a low pressure $CH_4/O_2/NO$ flame. Absolute concentration profiles of methyl and hydroxy radicals as well as nitric oxide are measured by LAS. CH and CN radicals are followed by LIF. As shown in Figure 6(b) the shape and absolute values of the concentration profiles for OH and CH_3 radicals as well as the initial reduction of nitric oxide are predicted well by the models. However, absolute CN radical concentrations[111] are significantly overestimated by all models (lines). More work has to be done on the details of the formation and destruction of CN during the reduction of nitric oxide in the reburn process. Further improvement in sensitive absorption spectroscopy in flames can be obtained by using longpath absorption techniques such as CRLAS[19] and intracavity absorption (cf. section 2.1.1).

REMPI, combined with molecular beam sampling mass spectrometry, is a highly selective and sensitive technique for combustion diagnostics and environmental chemical analysis.[95] Detection sensitivities in the lower parts per billion to parts per trillion (per volume) range of chloroethylenes, toluene and *p*-chlorophenol have been obtained in helium carrier gas, whereas

Table 1 Detection limits for minor combustion-relevant species of various linear and nonlinear laser spectroscopic techniques. The spectroscopic transitions probed for each species, their transition frequency (in cm^{-1}), total pressure p, an estimate of the temperature T, and the sample environment are given. The detection limits are approximate values, either determined directly or extrapolated from the respective experimental data

Method	Species	Transition	Energy of transition (cm^{-1})	p (total) (bar)	T (K)[c]	Environment	Detection limit (absolute) (cm^{-3})	Detection limit (ppm)	Ref.
LAS	HO_2	$2\nu_1$ band	6625.8	6.7 (−3)	295	photolysis cell	3 (13)[b]	16.8	17
	NO	(3,0) band	5524	1	1040	H_2-air flame	2.7 (15)	100	11
	CH_4	$2\nu_3$ band	6048	1	300	absorption cell	1.8 (14)	7	16
	NH	$A^3\Pi - X^3\Sigma(0,0)$, $R_2(8)$	29 762	1	2100	$NH_3/N_2/O_2$ flat flame, $\Phi = 1.28$	3.0 (13)	7.9	98
CRLAS	OH	$A^2\Sigma - X^2\Pi(0,0)$	32 500	0.04	1800	CH_4/air-flame	2.0 (10)	0.12	20
	CH_3	$^rR(6,6)$	3224.42	0.05	1400	CH_4/air-flame	1.5 (13)	57.6	24
ICLAS	HCO	$A - X\ (09^00) - (00^10)$	16 260	0.047	1800	$CH_4/O_2/N_2$-flame	1.4 (11)	0.57	23
LIF	OH	$A^2\Sigma - X^2\Pi(0,0)$	32 500	1	2000		25.6 (11)	0.07	90
		$A^2\Sigma - X^2\Pi(0,0)$	32 500	9.2	1700	$C_2H_6/O_2/N_2$-flame	1 (14)	2.5	100
	NO	$A^2\Sigma - X^2\Pi(0,0)$	44 247.8	1	2000		8 (11)	0.22	7
DFWM	OH	$A^2\Sigma - X^2\Pi(0,0)$	32 500	1	1700	premixed CH_4/air-flame	2.0 (13)	4.66	101
	OH	$A^2\Sigma - X^2\Pi(0,0)$	32 500	1	2200	premixed CH_4/air-flame	7.0 (13)	21.1	102
	NH	$A^3\Pi - X^3\Sigma(0,0)$	29 762	1	2100	$NH_3/O_2/N_2 = 2.1/1.5/1.0$			103
	CH_4	ν_3, Q(5)(1,0)	3017.5	1 (−6)	300	CH_4/N_2	1.5 (11)	6174	79
	HF	ν_1, R(5)(1,0)	4000	1 (−3)	300	HF/He	1.0 (10)	0.41	104
	CH_3	$3s\ ^2A'1 - 2p^2A_2''$	46 185	1	1600	$CH_4/N_2/O_2$ flame, $\Phi = 1.25$	3.0 (14)	65	78
	C_2	$d^3\Pi_g - a^3\Pi_u$	19 354	1	3000	C_2H_2/O_2 flame (welding torch)	5 (11)	0.19	105
CARS	C_2	Q(10)(1,0)	1611.7	1	2500	C_2H_2/O_2 flame (welding torch)	1.0 (10)	0.003	94
	OH	$O_1(7.5)(1,0)$	3065.3	1	≈1800	H_2/air-flame (premixed)	1.0 (13)	2.5	97
	CO	Q(10)(1,0)	2143	1	2000	flame	4.8 (16)	13 000	106
	CO	Q(10)(1,0)	2143	1	2000	flame	7.3 (16)	20 000	64
	OH	$Q_1(7.5)(1,0)$	3560	0.0106	300	HNO_3/He	1.4 (15)	5400	107
	NH_2	$\nu_1, 2\nu_4$, (1,0)	3210	0.0005	300	NH_3	1.2 (12)	99	108
REMPI	1,1-DCE[d]	$p - ^3d\ (1+1)$	33 863.9	–	5	molecular beam	6 ppbv[a]		95
	C_6H_6	$S_0 - S_1\ (1+1)$	38 610	–	5	molecular beam	0.09 ppbv[a]		95
	C_2H_4	3 + 1	68 181	–	5	molecular beam	10 ppbv[a]		96
PS	C_2	$d\ ^3\Pi_g - a^3\Pi_u(0,0)$	19 357	1	1800	C_2H_2/O_2-flame	1.0 (12)	0.19	109
	OH	$A^2\Sigma - X^2\Pi(0,0)$	32 500	1	2100	propane/air flame	1.0 (13)	2.9	110

[a] Mass spectrometric inlet system with molecular beam sampling; the detection limits are referred to the concentration levels in the original sample mixture.
[b] $1(11) = 1 \times 10^{11}$.
[c] Mostly estimated temperatures.
[d] 1,1-Dichloroethene.

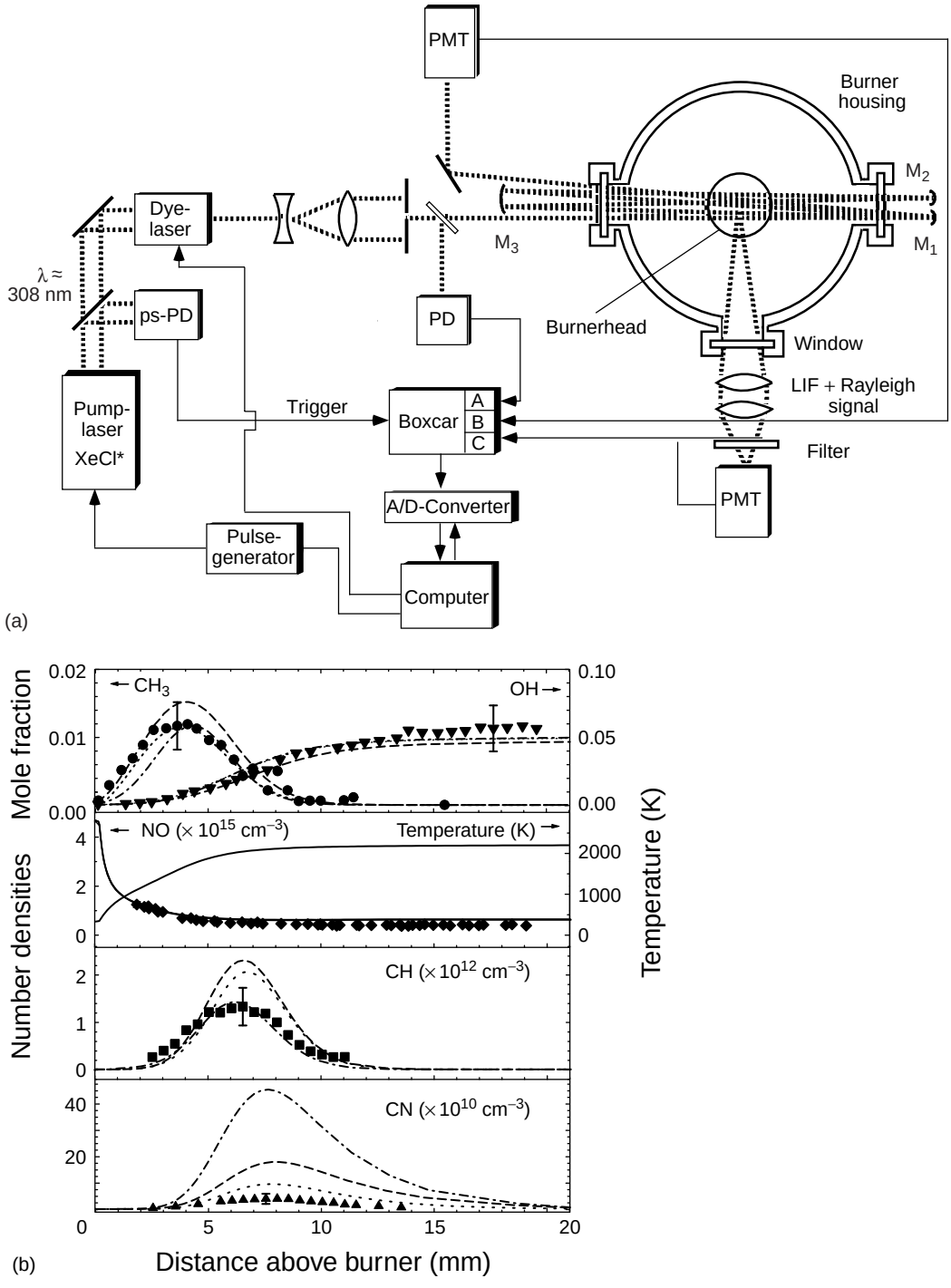

Figure 6 (a) Experimental set-up for laser absorption and LIF spectroscopy in low-pressure flames using tunable dye lasers and a long-path White-mirror arrangement. PMT = photomultiplier tube, PD = photo diode. (b) Comparison of measured and calculated concentration profiles in a laminar $CH_4/O_2/NO$ premixed flame at 1.33×10^3 Pa. Models: GRI (dashed), Warnatz (dash-dotted) and Lindstedt (dotted).[111]

because of cross-sensitivities and nonresonant multiple species ionization, this limit increases into the percent range in sample mixtures directly withdrawn from real world industrial devices. To increase selectivity, with an appropriate choice of excitation wavelengths various electronic transitions in large molecules can be accessed in one- or multiphoton transitions from which ionization is accomplished in a second step at the same or a

separate laser frequency. Portable instruments have been designed which offer sufficient flexibility in switching between different laser wavelengths for the excitation of numerous unsaturated hydrocarbons, radicals and polycyclic aromatic hydrocarbons (PAHs) for monitoring in real time. An example is the extraction and sensitive on-line REMPI detection of automotive exhaust gas pollutants at different engine loads and speeds.[96]

Laser ionization mass spectrometric detection also has been employed successfully for spatially resolved species analysis in low-pressure sooting flames to enlighten the chemical kinetic mechanisms responsible for soot precursors and particle growth. Most soot precursors possess absorption maxima between 200–300 nm with ionization thresholds between 6 and 9.5 eV. Special scavenging techniques allow the sampling and subsequent REMPI/mass spectrometric analysis of possible radical precursors as important building blocks for the growth history of soot particles and carbon cages such as C_{60}, C_{70} in flame gases. Homann et al.[112] determined correlations between the number of hydrogen and carbon atoms contained in highly condensed isomers of PAHs as precursors of C_{60}, and observed their formation and destruction in low-pressure sooting flames. These authors claim the C_{60} cage is the final product in the chain of hydrogen abstraction, folding, condensation and isomerization of smaller fragments of five- and six-membered aromatic ring condensates.[112,113] The detailed study of the fragmentation behavior, ionization efficiencies and absolute calibration of a variety of PAHs is essential for quantitative comparisons with numerical modeling of soot in flames.[114]

3.2.2 Higher Pressure Flames

In laminar flames spatially resolved temperature measurements using CARS are possible with high precision. Figure 7 shows a comparison of the temperature readings deduced from the pure rotational spontaneous Raman spectrum (squares) and the Q-branch vibrational CARS spectrum of molecular nitrogen (triangles) acquired at the same location and under the same operating conditions in a stable high-pressure premixed methane/air flame. Both data sets agree within the 3% limit of precision estimated from the spontaneous Raman data.[115] The pure rotational spontaneous Raman spectra exhibit well-resolved lines even at the highest pressures investigated. The spectrally dense CARS Q-branch of N_2 suffers partial collisional collapse of its rotational structure at the band origin which usually degrades the temperature measurement accuracy at low temperatures (300–700 K) and higher pressures.[116] Temperature measurements using pure rotational CARS have also been done in laminar flames and often are an alternative

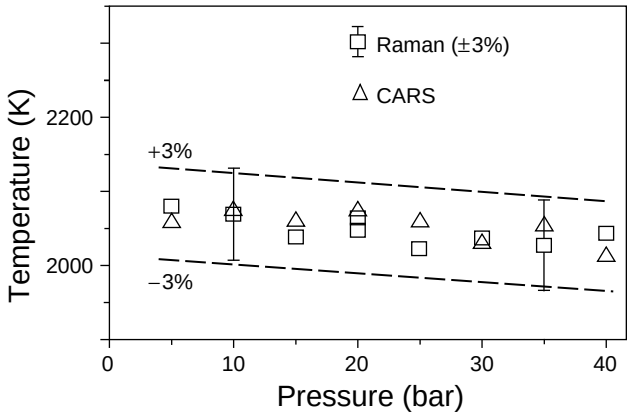

Figure 7 Comparison of measured temperatures (deduced from fitting spectral shapes) in a methane/air flame at various pressures. Spectra were obtained either from spontaneous pure rotational Raman scattering (squares) or from Q-branch vibrational CARS (triangles) measurements of molecular nitrogen at the same location in the post flame gases.[115] Dashed lines are the ±3% temperature measurement accuracies inherent in fitting results from the spontaneous Raman data.

if higher precision is required in the low temperature (300–870 K) higher pressure (1–15 bar) regime.[117]

To detect minor species, fully resonant nonlinear techniques, such as DFWM, PS and resonance enhanced CARS are neccessary to achieve sufficient detection sensitivity. DFWM spectra of electronic transitions have been recorded and gas temperatures have been determined in flames from OH, NH,[103] NO,[75,118] CN,[119] CH,[72] and in cell experiments from NO_2[120] and HCO.[121] Using PS important radicals in flames such as OH and NO were analyzed by Nyholm et al.[83] and Löfstedt et al.,[122] respectively. The potentials of PS for two-dimensional imaging of these species and for single shot temperature measurements were also shown.[123,124] Because the total angular momentum of the photon-molecule system is conserved during the interaction process, a proper choice of beam polarizations can be exploited to excite selectively P-, Q- or R-branch transitions. Careful modeling of the PS spectral structure for OH and NH is essential for in situ temperature measurements in atmospheric pressure flames.[125] Recently, the intensity and collision frequency dependence of PS signal intensities has been modeled using the rigorous approach of DNI of the Liouville equation for the density operator,[126] which promises to address more closely these effects in practical measurement situations. To increase the Raman-resonant susceptibility in a CARS interaction process, Stokes-, pump- and the generated anti-Stokes-signal frequency should be tuned close to one-photon allowed electronic transitions in the molecule studied. This makes the experimental approach rather complicated, since up to three narrowband tunable laser systems

have to be employed. Nevertheless, spatially resolved measurements of C_2 and OH distributions in various pressure flames have been performed by Attal et al., proving high detection sensitivities.[94,97]

3.3 Turbulent Combustion

Fundamental studies of flames are based on different experimental set-ups tailored to select certain aspects of combustion. Technical processes usually rely on non-premixed, turbulent combustion for reasons of efficiency and safety. This information about chemistry–turbulence interaction and complex recirculating flows is necessary to invent models suitable for simulating technically relevant flames. The understanding of the underlying processes is best verified through a direct comparison of numerical model results with experimental measurements of properties such as temperature and species concentrations. Furthermore, detailed knowledge of the underlying flow field patterns is essential.

Laser-based techniques are ideal for studying turbulent processes since they do not disturb the flow and, using pulsed lasers, they are fast enough to resolve the smallest relevant timescales. Statistical investigations of correlations of temperature and concentration of different reactive species were investigated by using point measurements based on Rayleigh and Raman scattering and LIF diagnostics. Quantitative comparisons of the thermochemical states in turbulent flames with idealized representations, such as steady strained laminar flames, perfectly stirred reactors, or adiabatic equilibrium are thus possible. Comparisons with DNS calculations on the other hand give insight into the validity of some of the basic assumptions that turbulent combustion models are built upon and allow for the evaluation of the predictive accuracy, strengths and limitations of a wide variety of combustion models.

There has been increasing interest in revealing the structure of turbulent flames and this was addressed using laser imaging for temperature, mixture fraction and reactive species such as OH and CH.[127] Simultaneous measurements of these quantities yielding scalar dissipation rates are critical for the validation of turbulent combustion models describing molecular mixing and extinction processes. Simultaneous measurements of RS and fuel Raman imaging provide information about temperature distribution and mixing. Both can be used to quantify further OH-LIF data by correcting for temperature and collisional quenching effects. These measurements (cf. Figure 8) showed the correlation of reaction zone width and jet Reynolds numbers and addressed the influence of local dissipation rate on flame extinction.[128] A review article that covers this topic in detail is given by Masri et al.[129]

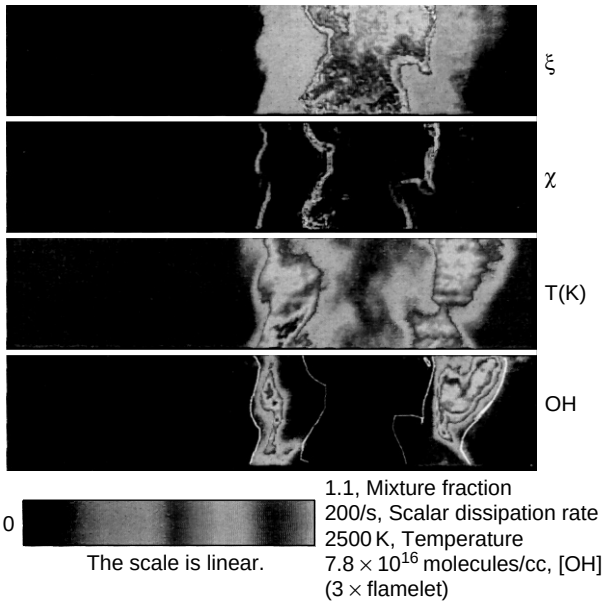

Figure 8 Images of mixture fractions (ξ), scalar dissipation rates (χ), temperatures (T) and OH concentrations in turbulent CH_4/air diffusion flames obtained by two-dimensional-Raman, Rayleigh- and LIF-spectroscopy.[128]

With careful spectral filtering Raman measurements are possible even in dense media involving strong Mie scattering from small droplets. By forming a thin light sheet from the output beam of a XeF-excimer laser at 351 nm, Decker et al.[130] were able to image the Raman scattered light from a dense cryogenic jet of liquid nitrogen onto a CCD camera to study the evaporation and breakup of the fluid nitrogen exiting through a small (1.9-mm diameter) nozzle as employed in rocket engines, and performed temperature and density measurements (from the Stokes/anti-Stokes intensity ratio and the Rayleigh intensity, respectively).

There are numerous applications of CARS as a temperature measurement technique in turbulent combustion processes. From the Q-branch spectrum of nitrogen temperature measurement accuracies of 2% or better within a single laser pulse have been obtained,[131] and applications have spread into diverse fields of technical combustion systems such as liquid fuel combusters and swirl burners,[62] engine combustion,[132] gas turbines and jet propulsion devices.[133–135] High-resolution lineshape measurements of combustion relevant species, such as N_2,[136] O_2,[137] H_2,[138] CO,[139] CO_2,[140] and NO[141] have improved modeling of CARS spectra, an important prerequisite for precise temperature measurements through spectral simulation. Figure 9 shows temperature profiles from single shot broadband nitrogen CARS thermometry across the shear layer region of a supersonic air-fed combuster fueled with hydrogen.[134] The symbols connected

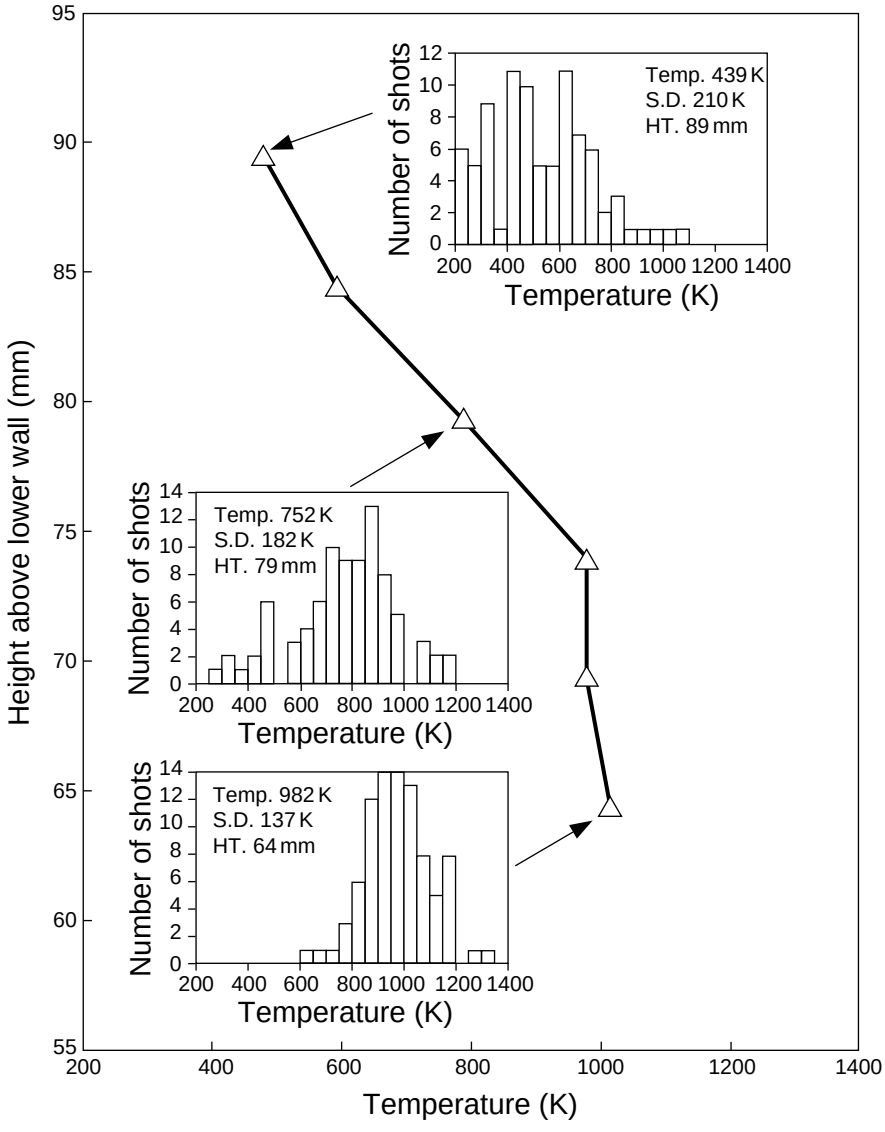

Figure 9 Temperature distribution deduced from H_2 Q-branch CARS spectra across the shear layer of a supersonic hydrogen fueled combuster. The inserts at three measurement locations also depict single shot histograms with relative standard deviations (S.D.) of temperature at different heights (HT) from the wall.[134]

by straight lines depict the mean temperature across the shear layer which is formed after injection of the fuel hydrogen into the supersonic air flow; it slowly decreases as the measurement point moves upstream. The neccessity to do single pulse measurements is clearly demonstrated, however, in the temperature histograms (inserts, at three measurement locations), illustrating the fluctuation in this parameter as a function of position in the flow. This gives a more detailed picture on the mixing properties and burning behavior of the combuster.

3.4 Engine Combustion

Reducing the release of pollutants is of particular interest in improving the environmental acceptance of combustion processes. The following discussion focuses on NO which is with soot and unburned hydrocarbons one of the major combustion-related pollutants. NO concentration measurements in spark ignition (SI) and diesel engines exciting NO with tunable ArF excimer lasers at 193 nm (ε-bands)[142] and 225 nm (γ-bands)[36] are reported. However, to circumvent the strong attenuation[143] of these short wavelengths in many technical combustion systems, excitation of NO with tunable KrF excimer lasers at 248 nm was suggested.[144] Since then, this technique is frequently applied for quantitative measurements of NO in IC engines.[145] The excitation laser is tuned to the O_{12} bandhead of the A–X(0,2) transition where the fluorescence excitation

spectrum of molecular oxygen has a local minimum. NO fluorescence emitted in the A–X(0,1) and A–X(0,0) bands at shorter wavelengths is detected which further minimizes the influence of interfering species. Quantification of LIF signal intensities requires detailed knowledge of several temperature and pressure-dependent factors. Variation of absorption spectra due to line broadening and shifting[146] change the spectral overlap of laser profile and absorption line.[29] Collisional quenching leads to a decrease of the fluorescence lifetime by radiationless deactivation of excited molecules and therefore, has to be included in the data reduction.[147] The simultaneous application[145] of RS and LIF for measurements of temperatures and NO concentrations in an IC engine is shown in Figure 10. From the image pairs with corresponding NO concentrations and temperature fields it can be seen that NO is formed in areas of high temperature. While the overall spatial distribution of NO and temperature is strongly correlated, the profiles in Figure 10 indicate that the temperature distribution is more uniform than the NO concentration distribution. This is an important result for comparison with mathematical models which are developed for engine design. Owing to the strong nonlinear temperature dependence of the NO production rate, careful control and homogenization of the combustion conditions should allow significant reduction of primary NO formation.

Valuable information for most major species in engines can again be obtained from SRS. Grünefeld et al.[148] performed spontaneous Raman imaging along a line with noise levels of a few per cent in a single shot using the beam of a KrF excimer laser at 248 nm with spectral as well as spatial dispersion of the scattered light accomplished by mounting an intensified CCD camera in the exit plane of a spectrograph. The larger scattering cross-sections in the UV allows imaging of the scattered light with smaller collection optics and therefore measurements in technically relevant devices such as engines and boilers. By taking advantage of the highly polarized Raman and Rayleigh scattered light in comparison with unwanted laser excited fluorescence and stray light from enclosure walls, they were able to monitor major species in an isooctane spray flame burner and a four cylinder in-line engine by subtracting spectra taken with mutually perpendicular polarization of the exciting laser beam.

The most accurate spatially resolved temperature measurements in engines can be performed with CARS as long as beam steering effects which severely degrade the CARS signal generation process can be tolerated. Therefore, precise cycle-resolved thermometry is possible in the compression stroke prior to ignition[149] and in the expansion stroke when the flame front has traversed through the focal volume of the beams.[150]

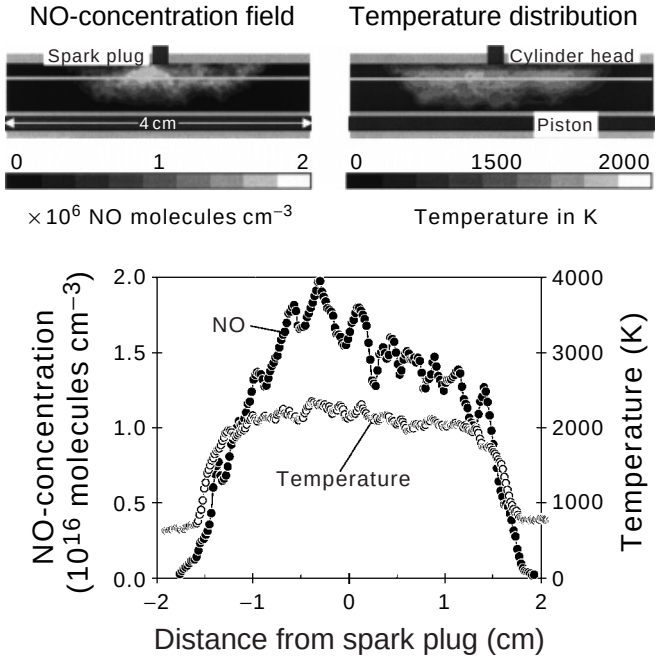

Figure 10 Simultaneous single-shot absolute NO concentration (left image) and temperature (right image) fields in a transparent SI-engine fueled with propane/air ($\lambda = 1$) using two-dimensional-LIF and RS.[145] In the lower graph are depicted plots of the respective data points (NO: filled circles, Temperature: open circles) along the straight lines given in the respective images.

In any case the USED (unstable resonator enhanced detection) CARS phase matching beam geometry[133] is preferred in such measurements to minimize beam steering effects.

3.5 Coal and Waste Combustion

For large scale combustion systems such as power plants, incinerators and heavy industries, tight monitoring and control of combustion is of great importance for a clean environment. However, these applications need ruggedized, compact and reliable laser spectroscopic instruments. Species concentrations,[16,151] temperature[152] and other physical parameters like mass flux and pressure[11] can be directly inferred from absorption lines of the majority species such as H_2O, O_2, CO_2, or CH_4 within the combustion chamber or exhaust channel by inexpensive and compact room temperature near-infrared diode lasers (NIRDL) via overtone and combination bands in the near-infrared (0.75–2 μm) spectral region. Water present in high concentrations in all hydrocarbon combustion processes has strong rovibrational absorption bands around 1.4 μm and is an ideal molecule for temperature measurements. This was successfully used to develop a closed-loop combustion temperature control system with a time resolution of 100 μs by extracting

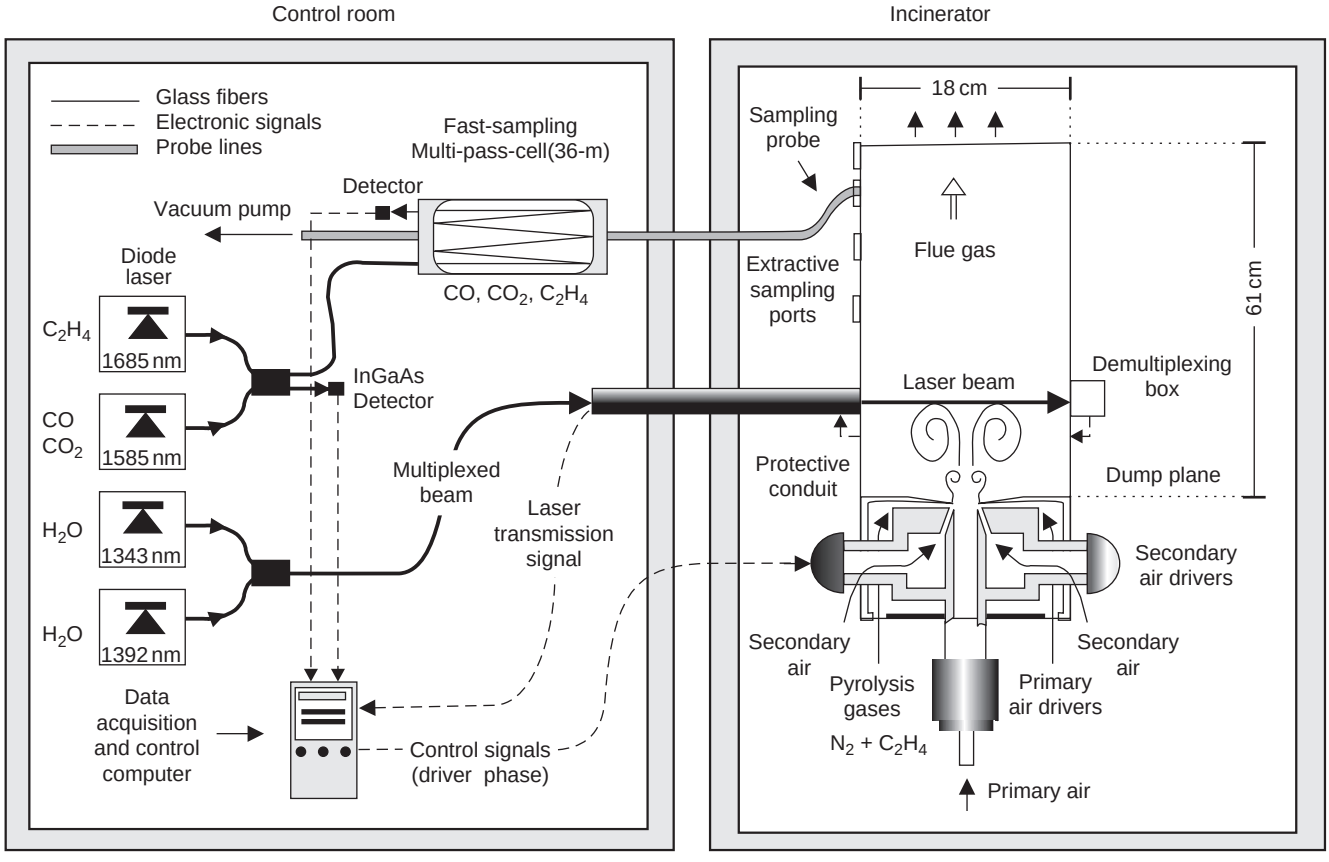

Figure 11 Closed loop control of a pulsed 50-kW model incinerator using a multiplexed fiber-coupled multispecies diode laser system for the in situ detection of temperature and H_2O and fast-sampling detection of CO, CO_2 and C_2H_4.[153]

the temperature from the ratio of the peak absorption coefficients. An improved version of this device extracted characteristic fast temperature variations within the flue gas duct (diameter 18 cm) of a pulsed 50-kW incinerator and used the root mean square average (T_{RMS}) of the 183-Hz frequency component to establish an active closed-loop control of the CO emission by optimizing the relative phase angle ϕ between the primary and secondary air injection valves which are pulsed at identical frequency ϕ_0 (Figure 11).[153] Another successful NIRDL-based sensor development for active control purposes is the simultaneous in situ detection of O_2 and H_2O in the combustion chamber (diameter 4 m) of a 20-MW-waste incinerator, which aims to optimize the secondary air injection.[154] This sensor provides an important contribution to the complex combustion control system (CCS) depicted in Figure 12. It was used to ensure a highly efficient and environmentally safe waste incineration by actively optimizing the selective non-catalytic reduction (SNCR) with a CO_2-LISA.[155] In addition, the primary air injection, grate movement and fuel input are under closed-loop control through an infrared scanner camera (TACCOS)[156] measuring the waste bed temperature. Numerous disturbances caused by the in situ path such as transmission fluctuations and thermal background emission, which interfere with the molecular absorption signal had to be suppressed completely.[157] The laser signals were verified by extractive reference sensors, which probe the cold flue gas duct and provided a linear response with a sensitivity of about 0.3 volume% and a temporal resolution of 1 s. Since gas transport to the sensor could be avoided, the laser signals are up to 60 s faster than the reference signals. The recent version of this sensor was improved further by fiber coupling the laser and by an inexpensive, compact and robust analog data evaluation[15] so that the NIRDL-based in situ species sensors are attractive tools for industrial control purposes.

Excimer laser-induced fragmentation fluorescence (ELIF),[158] is another laser spectroscopic sensing technique, which proved its suitability for industrial combustion control applications, by in situ detection of alkali compounds. These substances which are present in the flue gas of various combustion processes can cause severe corrosion damage to the high-temperature gas turbines needed for high efficiency pressurized

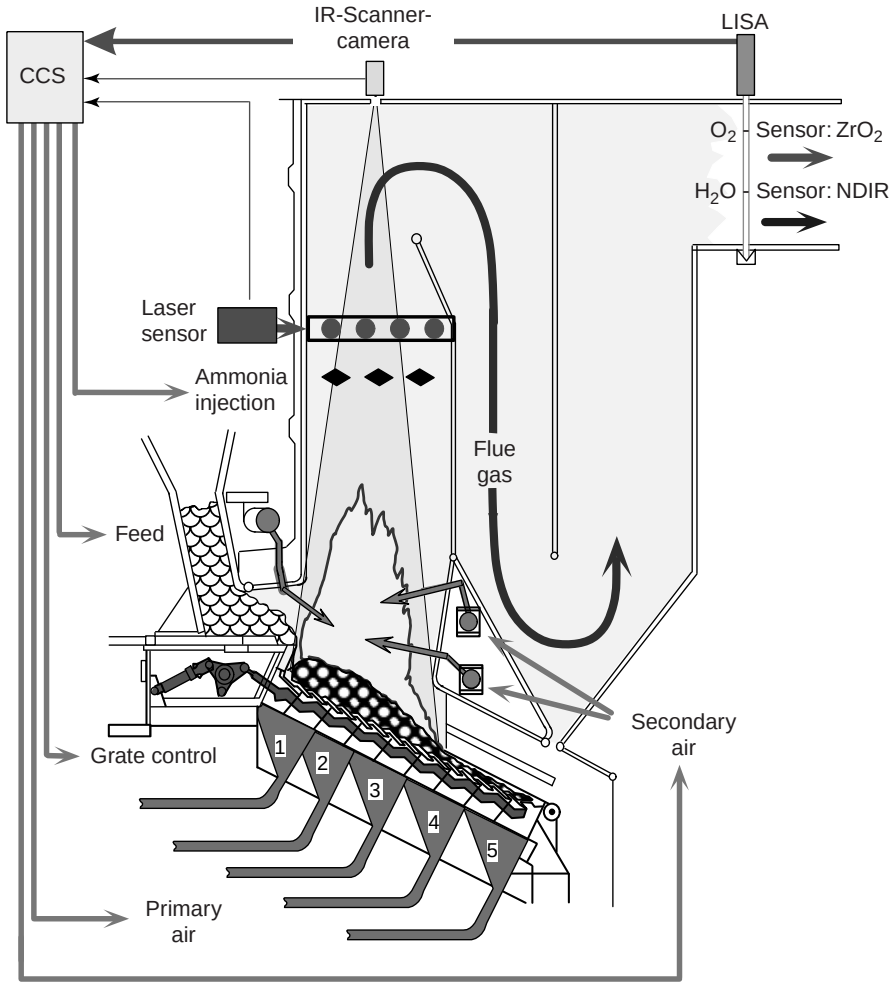

Figure 12 Intersection through the waste incinerator at which the in situ detection of O_2 and H_2O has been performed. A selection of important sensors (diode laser in situ sensor, laser in situ ammonia monitor, laser based in situ ammonia monitor (LISA), and infrared scanner camera, thermography-assisted combustion control system (TACCOS)) as well as the possible actors which are connected to the CCS, are indicated.[156]

coal combustion.[159,160] In ELIF an excimer laser photodissociates gaseous alkali compounds within the flue gas duct and generates electronically excited alkali atoms ($Na(3^2P)$, $K(4^2P)$), which are then readily detected by their fluorescence in the visible region via a fiber-coupled two species detection system as depicted in Figure 13. The measured signals are converted to absolute concentrations by calibration of the complete detection apparatus under known thermodynamic conditions. A detection limit of 0.2 ppb could be established for both alkalis at 10 bar total pressure and 800 °C gas temperature. An expansion of the principle to the detection of heavy metals is in preparation, so that this sensor promises to become an important tool to control the efficiency of alkali flue gas filters and to optimize the combustion process in order to minimize the release of alkali and heavy metal compounds.

4 FUTURE DEVELOPMENTS

After about two decades of research and development in the area of laser-based concepts for combustion diagnostics many of the spectroscopic methods have matured from qualitative to quantitative techniques. A closer and more direct relation to engineering and development purposes is the application of laser-imaging diagnostics to IC engines. Once optical access has been achieved for a given device, temperature, gas and particle velocities as well as a variety of species can be studied. By identifying problems with fuel intake and mixing in engines via gaseous fuel imaging and detection of fuel droplets by Mie scattering together with the corresponding temperature information, one can optimize the cold start behavior of Otto engines and control soot formation in direct-injection diesel engines.

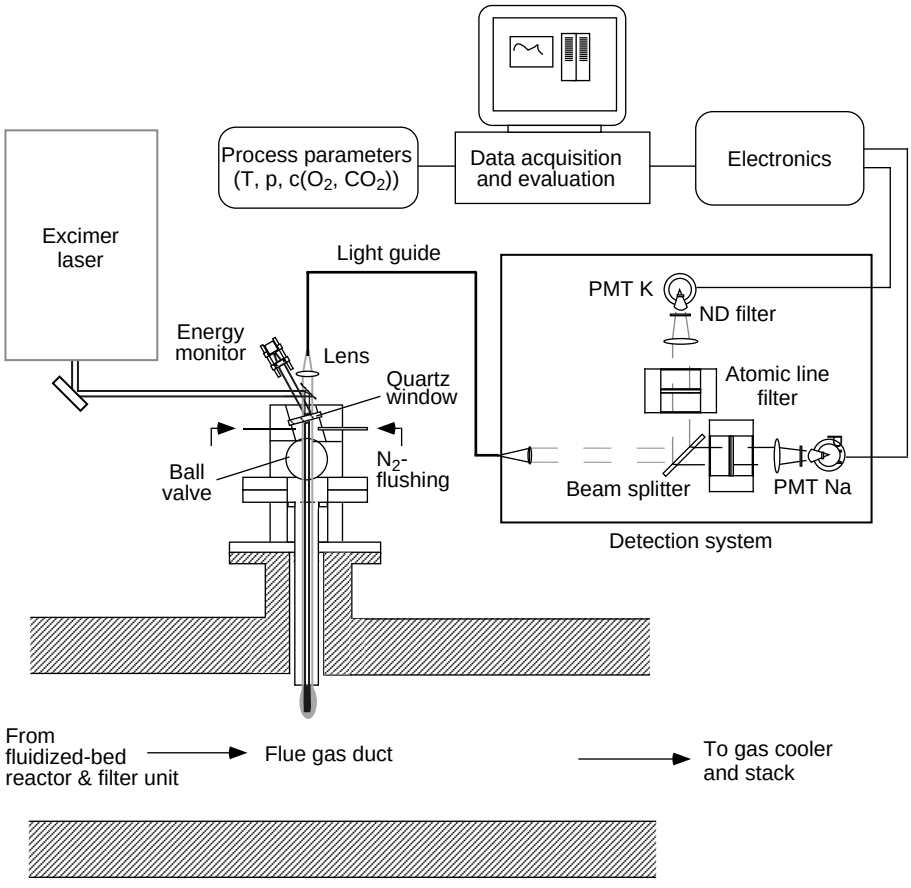

Figure 13 Detection of alkali species in the flue gas from a pressurized coal combustion fluidized bed (PCFB) plant with a single-port ELIF system.

The cold start accounts for a major part in unburnt hydrocarbon emission from automobiles. A detailed knowledge of the fuel evaporation process is also of basic importance for gas turbines in power plants and jet engines. In all these areas, laser-based techniques are now on the way to move from the research laboratories into the development departments.

Another area of high technological relevance is the interaction of turbulent reactive flows with the walls of the technical installations. Heterogeneous processes at the surface as well as mass and energy transfer to the turbulent boundary layer are important parameters of the optimization, not only of combustion processes, but also for industrial chemical reactors. Using imaging experiments together with detailed numerical simulations one can elucidate microscopic details in terms of underlying elementary processes and optimize industrial installations.

The application of modern high power UV-laser systems as well as robust and low cost semiconductor-based detectors and laser diodes offers many new possibilities for intelligent CCSs in new power plants with combined cycle operation as well as in the efficient use of refuse for the production of thermal and electrical energy. The large amount of information provided by these new laser-based detection schemes can be used in modern fuzzy logic and neural network-based control systems.

Laser multidimensional imaging of the interaction of turbulent flow fields with combustion can be done by diode pumped high repetition rate solid state lasers which finally will provide a full time development visualization. Short pulse lasers with picosecond and femtosecond pulse length can be used to investigate directly collision and transfer rates which are necessary for quantification of LIF experiments in technical combustion systems.

ACKNOWLEDGMENTS

TD is thankful to Thomas Fernholz for critically reading the manuscript. The authors are grateful for the generous support by the Deutsche Forschungsgemeinschaft (SFB 359), the BMBF and the State of

Baden-Württemberg (TECFLAM) and many industrial companies.

ABBREVIATIONS AND ACRONYMS

AES	Auger-electron Spectrometer
CARS	Coherent Anti-Stokes Raman Scattering
CCD	Charge Coupled Device
CCS	Combustion Control System
CRLAS	Cavity Ring-down Laser Absorption Spectroscopy
DFWM	Degenerate Four Wave Mixing
DNI	Direct Numerical Integration
DNS	Direct Numerical Simulation
EET	Electronic Energy Transfer
ELIF	Excimer Laser-induced Fragmentation Fluorescence
IC	Internal Combustion
ICLAS	Intracavity Laser Absorption Spectroscopy
LAS	Laser Absorption Spectroscopy
LASER	Light Amplification by Stimulated Emission of Radiation
LEED	Low Energy Electron Diffraction
LIF	Laser-induced Fluorescence
LIG	Laser-induced Grating
LII	Laser-induced Incandescence
LIPF	Laser-induced Predissociative Fluorescence
LISA	Laser Based In Situ Ammonia Monitor
MASER	Microwave Amplification by Stimulated Emission of Radiation
MC	Monochromator
MDA	Minimum Detectable Absorption
MPA	Multiphoton Absorption
NIRDL	Near-infrared Diode Lasers
OPO	Optical Parametric Oscillator
PAHs	Polycyclic Aromatic Hydrocarbons
PCFB	Pressurized Coal Combustion Fluidized Bed
PDF	Probability Density Function
PG	Population Grating
PMT	Photomultiplier Tube
PS	Polarization Spectroscopy
QMS	Quadrupole Mass Spectrometer
REMPI	Resonance Enhanced Multiphoton Ionization
RET	Rotational Energy Transfer
RS	Rayleigh Scattering
SFG	Sum Frequency Generation
SI	Spark Ignition
SNCR	Selective Non-catalytic Reduction
SRS	Spontaneous Raman Scattering
TACCOS	Thermography-assisted Combustion Control System
TG	Thermal Grating
TOPLIF	Two-optical Path Laser-induced Fluorescence
USED	Unstable Resonator Enhanced Detection
UV	Ultraviolet
VET	Vibrational Energy Transfer

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Automotive Emissions Analysis with Spectroscopic Techniques • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Laser-induced Breakdown Spectroscopy, Elemental Analysis • Laser Mass Spectrometry in Trace Analysis • Photoacoustic Spectroscopy in Trace Gas Monitoring

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Luminescence in Environmental Analysis • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Atomic Spectroscopy **(Volume 11)**
Graphite Furnace Atomic Absorption Spectrometry • Laser Spectrometric Techniques in Analytical Atomic Spectrometry • Laser-induced Breakdown Spectroscopy

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Lifetime Measurements, Applications of

Infrared Spectroscopy **(Volume 12)**
Cavity Ringdown Laser Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach

• Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Time-of-flight Mass Spectrometry

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Raman Scattering, Fundamentals

X-ray Spectrometry **(Volume 15)**
Ultrafast Diffraction Techniques

General Articles **(Volume 15)**
Ultrafast Laser Technology and Spectroscopy

REFERENCES

1. G.R. Kirchhoff, R.W. Bunsen, 'Chemische Analyse Durch Spectralbeobachtungen', *Poggend. Annal.*, **110**, 161–176 (1860).
2. A. Einstein, 'Zur Quantentheorie der Strahlung', *Phys. Z.*, **18**, 21–128 (1917).
3. J.P. Gordon, H.H. Zeiger, C.H. Townes, 'The Maser – New Type of Microwave Amplifier, Frequency Standard and Spectrometer', *Phys. Rev.*, **99**, 1264–1274 (1955).
4. T.H. Maiman, 'Stimulated Optical Radiation in Ruby', *Nature*, (London), **187**, 493–494 (1960).
5. A.G. Gaydon, *The Spectroscopy of Flames*, Chapman and Hall, London, 1957.
6. W. Demtröder, *Laser Spectroscopy – Basic Concepts and Instrumentation*, Springer, New York, 1997.
7. A.C. Eckbreth, *Laser Diagnostics for Combustion, Temperature and Species*, Gordon and Breach, Amsterdam, 1996.
8. J. Warnatz, U. Maas, R.W. Dibble, *Combustion – Physical and Chemical Fundamentals, Modeling and Simulation, Experiments, Pollutant Formation*, Springer-Verlag, Berlin, 1996.
9. V. Ebert, J. Wolfrum, 'Absorption Spectroscopy', in *Optical Measurements – Techniques and Applications*, ed. F. Mayinger, Springer-Verlag, Berlin, 273–312, 1994.
10. D.S. Baer, R.K. Hanson, M.E. Newfield, N.K.J.M. Gopaul, 'Multiplexed Diode-laser Sensor System for Simultaneous H_2O, O_2 and Temperature Measurements', *Opt. Lett.*, **19**, 1900–1902 (1994).
11. M.G. Allen, 'Diode Laser Absorption Sensors for Gasdynamic and Combustion Flows', *Meas. Sci. Technol.*, **9**, 545–562 (1998).
12. A.P. Thorne, *Spectrophysics*, Chapman and Hall, London, 1988.
13. E.R. Furlong, D.S. Baer, R.K. Hanson, 'Combustion Control and Monitoring using a Multiplexed Diode-laser Sensor System', Twenty-sixth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 2851–2858, 1996.
14. I.I. Sobel'man, L.A. Vainshtein, E.A. Yukov, 'Excitation of Atoms and Broadening of Spectral Lines', *Springer Series on Atoms and Plasmas*, Springer-Verlag, Berlin, 1995.
15. V. Ebert, J. Fitzer, I. Gerstenberg, K.-U. Pleban, H. Pitz, J. Wolfrum, M. Jochem, J. Martin, 'Online Monitoring of Water Vapor with a Fiber Coupled NIR-diode Laser Spectrometer', 5th International Symposium on Gas Analysis by Tunable Diode Lasers (VDI-Berichte 1366), VDI-Verlag, Düsseldorf, 145–154, 1998.
16. S. Schäfer, M. Mashni, J. Sneider, A. Miklos, P. Hess, V. Ebert, K.-U. Pleban, H. Pitz, 'Sensitive Detection of Methane with a 1.65 µm Diode Laser by Photoacoustic and Absorption Spectroscopy', *Appl. Phys. B*, **66**, 511–516 (1998).
17. C.A. Taatjes, D.B. Oh, 'Time-resolved Wavelength Modulation Spectroscopy Measurements of HO_2 Kinetics', *Appl. Opt.*, **36**, 5817–5821 (1997).
18. K. Kohse-Höinghaus, D.F. Davidson, A.Y. Chang, R.K. Hanson, 'Quantitative NH_2 Concentration Determination in Shock-tube Laser-absorption Experiments', *J. Quant. Spectrosc. Radiat. Transfer*, **42**, 1–17 (1989).
19. J.J. Scherer, J.B. Paul, A. O'Keefe, R.J. Saykally, 'Cavity Ringdown Laser Absorption Spectroscopy: History, Development, and Application to Pulsed Molecular Beams', *Chem. Rev.*, **97**, 25–51 (1997).
20. S. Cheskis, I. Derzy, V.A. Lozovsky, A. Kachanov, D. Romanini, 'Cavity Ring-down Spectroscopy of OH Radicals in Low Pressure Flame', *Appl. Phys. B*, **66**, 377–381 (1998).
21. S. Cheskis, I. Derzy, V.A. Lozovsky, A. Kachanov, F. Stoeckel, 'Intracavity Laser Absorption Spectroscopy Detection of Singlet CH_2 Radicals in Hydrocarbon Flames', *Chem. Phys. Lett.*, **277**, 423–429 (1997).
22. P. Zalicki, Y. Ma, R.N. Zare, E.H. Wahl, J.R. Dadamio, T.G. Owano, C.H. Kruger, 'Methyl Radical Measurement by Cavity Ring-down Spectroscopy', *Chem. Phys. Lett.*, **234**, 269–274 (1995).
23. J.J. Scherer, D.J. Rakestraw, 'Cavity Ringdown Laser Absorption Spectroscopy Detection of Formyl (HCO) Radical in a Low Pressure Flame', *Chem. Phys. Lett.*, **265**, 169–176 (1997).
24. J.J. Scherer, K.W. Aniolek, N.P. Cernansky, D.J. Rakestraw, 'Determination of Methyl Radical Concentrations in a Methane/Air Flame by Infrared Cavity Ringdown Laser Absorption Spectroscopy', *J. Chem. Phys.*, **107**, 6196–6203 (1997).
25. Y. He, M. Hippler, M. Quack, 'High-resolution Cavity Ring-down Absorption Spectroscopy of Nitrous Oxide

and Chloroform Using Near-infrared cw Diode Laser', *Chem. Phys. Lett.*, **289**, 527–534 (1998).
26. T. Yu, M.C. Lin, 'Kinetics of the C_6H_5 + NO Association Reaction', *J. Phys. Chem.*, **98**, 2105–2109 (1994).
27. E.W. Rothe, Y. Gu, A. Chryssostomou, P. Andresen, F. Bormann, 'Effect of Laser Intensity and of Lower State Rotational Energy Transfer upon Temperature Measurements Made with Laser-induced Predissociative Fluorescence', *Appl. Phys. B*, **66**, 251–258 (1998).
28. J. Luque, D.R. Crosley, 'Lifbase: Database and Spectral Simulation Program', SRI International Report, MP 96-001, 1996.
29. C. Schulz, V. Sick, J. Heinze, W. Stricker, 'Laser-induced Fluorescence Detection of Nitric Oxide in High-pressure Flames with $A-X(0,2)$ Excitation', *Appl. Opt.*, **36**, 3227–3232 (1997).
30. P. Desgroux, M.J. Cottereau, 'Local OH Concentration Measurements in Atmospheric Pressure Flames by a Laser-saturated Fluorescence Method: Two-optical Path Laser-induced Fluorescence', *Appl. Opt.*, **30**, 90–97 (1991).
31. P.H. Paul, J.A. Gray, J.L. Durant, Jr, J.W. Thoman, Jr, 'A Model for Temperature-dependent Collisional Quenching of NO $A^2\Sigma^+$', *Appl. Phys. B*, **57**, 249–259 (1993).
32. R. Kienle, M.P. Lee, K. Kohse-Höinghaus, 'A Scaling Formalism for the Representation of Rotational Energy Transfer in OH($A^2\Sigma^+$) in Combustion Experiments', *Appl. Phys. B*, **63**, 403–418 (1996).
33. R. Schwarzwald, P. Monkhouse, J. Wolfrum, 'Picosecond Fluorescence Lifetime Measurement of the OH Radical in an Atmospheric Pressure Flame', *Chem. Phys. Lett.*, **158**, 60–64 (1987).
34. A. Dreizler, R. Tadday, P. Monkhouse, J. Wolfrum, 'Time- and Spatially Resolved LIF of OH $A^2\Sigma^+(v'=1)$ in Atmospheric Pressure Flames Using Picosecond Excitation', *Appl. Phys. B*, **57**, 85–87 (1993).
35. J.T. Salmon, N.M. Laurendeau, 'Calibration of Laser-saturated Fluorescence Measurements Using Rayleigh Scattering', *Appl. Opt.*, **24**, 65–73 (1985).
36. A. Bräumer, V. Sick, J. Wolfrum, V. Drewes, R.R. Maly, M. Zahn, 'Quantitative Two-dimensional Measurements of Nitric Oxide and Temperature in a Transparent SI Engine', SAE paper, No. 952462, 1995.
37. E.W. Kaiser, K. Marko, D. Klick, L. Rimai, C.C. Wang, B. Shirinzadeh, D. Zhou, 'Measurements of OH Density Profiles in Atmospheric Pressure Propane/Air Flames', *Combust. Sci. Technol.*, **50**, 163–183 (1986).
38. C. Schulz, V. Sick, U. Meier, J. Heinze, W. Stricker, 'Quantification of NO $A-X(0,2)$ LIF: Investigation of Calibration and Collisional Influences in High-pressure Flames', *Appl. Opt.*, **38**, 1434–1443 (1999).
39. T. Etzkorn, J. Fitzer, S. Muris, J. Wolfrum, 'Determination of Absolute Methyl- and Hydroxyl-radical Concentrations in a Low Pressure Methane–Oxygen Flame', *Chem. Phys. Lett.*, **208**, 307–310 (1993).
40. A.O. Vydorov, J. Heinze, M. Dillmann, U.E. Meier, W. Stricker, 'Laser-induced Fluorescence Thermometry and Concentration Measurements on NO A–X(0,0) Transitions in the Exhaust Gas of High Pressure CH_4/Air Flames', *Appl. Phys. B*, **61**, 409–414 (1995).
41. R. Cattolica, 'OH Rotational Temperature from Two-line Laser Excited Fluorescence', *Appl. Opt.*, **20**, 1156–1166 (1981).
42. A. Arnold, B. Lange, T. Bouché, T. Heitzmann, G. Schiff, W. Ketterle, P. Monkhouse, J. Wolfrum, 'Absolute Temperature Fields on Flames by 2D-LIF of OH Using Excimer Lasers and CARS Spectroscopy', *Ber. Bunsenges. Phys. Chem.*, **96**, 1388–1393 (1992).
43. J.H. Grinstead, G. Laufer, J.C. McDaniel, Jr, 'Single-pulse, Two-line Temperature Measurement Technique Using KrF Laser-induced O_2 Fluorescence', *Appl. Opt.*, **34**, 5501–5512 (1995).
44. B.K. McMillin, J.L. Palmer, R.K. Hanson, 'Temporally Resolved, Two-line Fluorescence Imaging of NO Temperature in a Transverse Jet in a Supersonic Cross Flow', *Appl. Opt.*, **32**, 7533–7545 (1993).
45. F. Großmann, P.B. Monkhouse, M. Ridder, V. Sick, J. Wolfrum, 'Temperature and Pressure Dependencies of the Laser-induced Fluorescence of Gas-phase Acetone and 3-pentanone', *Appl. Phys. B*, **62**, 249–253 (1996).
46. M.C. Thurber, F. Grisch, B.J. Kirby, M. Votsmeier, R.K. Hanson, 'Measurement and Modeling of Acetone Laser-induced Fluorescence with Implications for Temperature-imaging Diagnostics', *Appl. Opt.*, **37**, 4963–4978 (1998).
47. T.S. Cheng, J.A. Wehrmeyer, R.W. Pitz, 'Simultaneous Temperature and Multispecies Measurement in a Lifted Hydrogen Diffusion Flame', *Combust. Flame*, **91**, 323–345 (1992).
48. N.P. Tait, D.A. Greenhalgh, 'PLIF Imaging of Fuel Fraction in Practical Devices and LII Imaging of Soot', *Ber. Bunsenges. Phys. Chem.*, **97**, 1619–1625 (1993).
49. B. Quay, T.-W. Lee, T. Ni, R.J. Santoro, 'Spatially Resolved Measurement of Soot Volume Fraction Using Laser-induced Incandescence', *Combust. Flame*, **97**, 384–392 (1994).
50. R.C. Shaddix, J.E. Harrington, K.C. Smith, 'Quantitative Measurement of Enhanced Soot Production in Flickering Methane/Air Diffusion Flame', *Combust. Flame*, **99**, 723–732 (1994).
51. P.-E. Bengtsson, M. Aldén, 'Soot-visualisation Strategies Using Laser Techniques. Laser-induced Fluorescence in C_2 from Laser-vaporized Soot and Laser-induced Soot Incandescence', *Appl. Phys. B*, **60**, 51–59 (1995).
52. J. Appel, B. Jungfleisch, M. Marquardt, R. Suntz, H. Bockhorn, 'Assessment of Soot Volume Fractions From Laser-induced Incandenscence by Comparison with Extinction Measurements in Laminar, Premixed Flat Flames', 26th Symposium (International) on

Combustion, The Combustion Institute, Pittsburgh, 2387–2395, 1996.
53. L.A. Melton, 'Soot Diagnostics Based on Laser Heating', *Appl. Opt.*, **23**, 2201–2208 (1984).
54. H. Geitlinger, Th. Streibel, R. Suntz, H. Bockhorn, 'Two-dimensional Imaging of Soot Volume Fractions, Particle Number Densities and Particle Sizes in Laminar and Turbulent Diffusion Flames', 27th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 1613–1621, 1998.
55. S. Will, S. Schraml, K. Bader, A. Leipertz, 'Performance Characteristics of Soot Primary Particle Size Measurements by Time-resolved Laser-induced Incandescence', *Appl. Opt.*, **37**, 5647–5652 (1998).
56. R.J. Hall, 'Intensity Convolutions of CARS Spectra', *Opt. Commun.*, **52**, 360–366 (1985).
57. R.E. Teets, 'Accurate Convolutions of Coherent anti-Stokes Raman Spectra', *Opt. Lett.*, **9**, 226–228 (1984).
58. M. Péalat, M. Lefebvre, J.-P.E. Taran, P.L. Kelley, 'Sensitivity of Quantitative Vibrational Coherent Anti-Stokes Raman Spectroscopy to Saturation and Stark Shifts', *Phys. Rev.*, **38**, 1948–1965 (1988).
59. M.L. Koszykowski, R.L. Farrow, R.E. Palmer, 'Calculation of Collisionally Narrowed Coherent anti-Stokes Raman Spectroscopy Spectra', *Opt. Lett.*, **10**, 478–480 (1985).
60. S. Kröll, M. Aldén, T. Berglind, R.J. Hall, 'Noise Characteristics of Single Shot Broadband Raman-resonant CARS with Single- and Multimode Lasers', *Appl. Opt.*, **26**, 1068–1073 (1987).
61. P.R. Regnier, J.P. Taran, 'On the Possibility of Measuring Gas Concentrations by Stimulated anti-Stokes Scattering', *Appl. Phys. Lett.*, **23**, 240–242 (1973).
62. A.C. Eckbreth, 'CARS Thermometry in Practical Combustors', *Combust. Flame*, **39**, 133–147 (1980).
63. W.A. England, J.M. Milne, S.N. Jenny, D.A. Greenhalgh, 'Application of CARS to an Operating Chemical Reactor', *Appl. Spectrosc.*, **38**, 867–876 (1984).
64. A.C. Eckbreth, R.J. Hall, 'CARS Concentration Sensitivity with and without Nonresonant Background Suppression', *Combust. Sci. Technol.*, **25**, 175–192 (1981).
65. P.-E. Bengtsson, L. Martinsson, M. Aldén, 'Dual-broadband Rotational CARS Measurements in an IC Engine', Twenty-fifth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 1735–1742, 1994.
66. P.-E. Bengtsson, L. Martinsson, M. Aldén, 'Combined Vibrational and Rotational CARS for Simultaneous Measurements of Temperature and Concentrations of Fuel, Oxygen, and Nitrogen', *Appl. Spectrosc.*, **49**, 188–192 (1995).
67. J.F. Reintjes, *Nonlinear Optical Parametric Processes in Liquids and Gases*, Academic Press, New York, 1984.
68. H.J. Eichler, P. Günter, D. Pohl, *Laser-induced Dynamic Gratings*, Springer Series in Optical Sciences, Springer Verlag, Berlin, 1986.
69. R.L. Abrams, R.C. Lind, 'Degenerate Four-wave Mixing in Absorbing Media', *Opt. Lett.*, **4**, 94–96 (1978).
70. R.L. Farrow, D.J. Rakestraw, T. Dreier, 'Investigation of the Dependence of Degenerate Four-wave Mixing Line Intensities on Transition Dipole Moment', *J. Opt. Soc. Am. B*, **9**, 1772–1777 (1992).
71. R.P. Lucht, R.L. Farrow, D.J. Rakestraw, 'Saturation Effects in Gas-phase Degenerate Four-wave Mixing Spectroscopy: Nonperturbative Calculations', *J. Opt. Soc. Am. B*, **10**, 1508–1520 (1993).
72. S. Williams, R.N. Zare, L.A. Rahn, 'Reduction of Degenerate Four-wave Mixing Spectra to Relative Populations II. Strong-field Limit', *J. Chem. Phys.*, **101**, 1093–1107 (1994).
73. L. Lehr, M. Motzkus, G. Pichler, K.L. Kompa, P. Hering, 'Temperature Determination of Transient Species by Degenerate Four-wave Mixing: Application of the Independently Determined Power Law of the Transition Dipole Moment and Geometric Factors', *J. Chem. Phys.*, **104**, 9698–9703 (1996).
74. R.L. Farrow, D.J. Rakestraw, 'Analysis of Degenerate Four-wave Mixing Spectra of NO in a $CH_4/N_2/O_2$ Flame', *Appl. Phys. B*, **68**, 741–747 (1999).
75. G.N. Robertson, K. Kohse-Höinghaus, S. LeBoiteux, F. Aguerre, B. Attal-Trétout, 'Observation of Field Effects and Rotational Line Coupling in DFWM Processes Resonant with $^2\Sigma-^2\Pi$ Electronic System', *J. Quant. Spectrosc. Radiat. Transfer*, **55**, 71–101 (1996).
76. B. Attal-Trétout, H. Bervas, J.P. Taran, S. Le Beuteux, P. Kelley, T.K. Gustafson, 'Saturated FDFWM Lineshapes and Intensities: Theory and Application to Quantitative Measurements in Flames', *J. Phys. B: At. Mol. Opt. Phys.*, **30**, 497–522 (1997).
77. R.L. Farrow, D.J. Rakestraw, 'Determination of Trace Molecular Species Using Degenerate Four-wave Mixing', *Science*, **257**, 1894–1900 (1992).
78. R.L. Farrow, M.N. Bui-Pham, V. Sick, 'Degenerate Four-wave Mixing Measurements of Methyl Radical Distributions in Hydrocarbon Flames: Comparison with Model Predictions', Twenty-sixth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 975–983, 1996.
79. G.J. Germann, R.L. Farrow, D.J. Rakestraw, 'Infrared Degenerate Four-wave Mixing Spectroscopy of Polyatomic Molecules: CH_4 and C_2H_2', *J. Opt. Soc. Am. B*, **12**, 25–32 (1995).
80. P.M. Danehy, E.J. Friedman-Hill, R.P. Lucht, R.L. Farrow, 'The Effect of Collisional Quenching on Degenerate Four-wave Mixing', *Appl. Phys. B*, **57**, 243–248 (1993).
81. H. Latzel, A. Dreizler, T. Dreier, J. Heinze, M. Dillmann, W. Stricker, G.M. Lloyd, P. Ewart, 'Thermal Grating and Broadband Degenerate Four-wave Mixing Spectroscopy of OH in High-pressure Flames', *Appl. Phys. B*, **67**, 667–673 (1998).
82. R.E. Teets, F.W. Kowalski, W.T. Hill, N. Charlson, T.W. Hänsch, 'Laser Polarization Spectroscopy', in *Advances*

in Laser Spectroscopy, Proceedings Society Phot. Optical Instrument Engineers, San Diego SPIE, Bellingham, WA, 80–87, Vol. 113, 1977.
83. K. Nyholm, R. Maier, C.G. Aminoff, M. Kaivola, 'Detection of OH in Flames by Using Polarization Spectroscopy', *Appl. Opt.*, **32**, 919–924 (1993).
84. M. Sargent, III, 'Polarized Field Saturation Spectroscopy', *Phys. Rev. A*, **14**, 524–527 (1976).
85. Y.R. Shen, *The Principles of Nonlinear Optics*, J. Wiley & Sons, New York, 1984.
86. Y.R. Shen, 'Surface Properties Probed by Second-harmonic and Sum-frequency Generation', *Nature*, **337**, 519–525 (1989).
87. X. Su, P.S. Cremer, Y.R. Shen, G.A. Somorjai, 'Pressure Dependence (10^{-10}–700 Torr) of the Vibrational Spectra of Adsorbed CO on Pt(111) Studied by Sum Frequency Generation', *Phys. Rev. Lett.*, **77**, 3858–3860 (1996).
88. J.E.M. Goldsmith, 'Photochemical Effects in 205-nm, Two-photon-excited Fluorescence Detection of Atomic Hydrogen in Flames', *Opt. Lett.*, **11**, 416–418 (1986).
89. J. Seitzmann, J. Haumann, R.K. Hanson, 'Quantitative Two-photon LIF Imaging of Carbon Monoxide in Combustion Gases', *Appl. Opt.*, **26**, 2892–2899 (1987).
90. M. Aldén, H. Edner, S. Wallin, 'Simultaneous Spatially Resolved NO and NO_2 Measurements Using One- and Two-photon Laser-induced Fluorescence', *Opt. Lett.*, **10**, 529–531 (1985).
91. P.J.H. Tjossem, K.C. Smith, 'Multiphoton Excitation Spectroscopy of the $B^1\Sigma^+$ and $C^1\Sigma^+$ Rydberg States of CO', *J. Chem. Phys.*, **91**, 2041–2048 (1989).
92. D.R. Crosley, G.P. Smith, 'Two-photon Spectroscopy of the $A^2\Sigma^+$–$X^2\Pi_i$ System of OH', *J. Chem. Phys.*, **79**, 4764–4773 (1983).
93. K. Shimoda, 'Limits of Sensitivity of Laser Spectrometers', *Appl. Phys.*, **1**, 77–86 (1973).
94. B. Attal, D. Débarre, K. Müller-Dethlefs, J.P.E. Taran, 'Resonance-enhanced Coherent Anti-Stokes Raman Scattering in C_2', *Rev. Phys. Appl.*, **18**, 39–50 (1983).
95. T.N. Tanada, J. Velazques, N. Hemmi, T.A. Cool, 'Detection of Toxic Emissions from Incinerators', *Ber. Bunsenges. Phys. Chem.*, **97**, 1516–1526 (1993).
96. C. Weickhard, U. Boesl, 'Time Resolved Trace Analysis of Exhaust Gas by Means of Laser Mass Spectrometry', *Ber. Bunsenges. Phys. Chem.*, **97**, 1716–1719 (1993).
97. B. Attal-Trétout, P. Berlemont, J.P. Taran, 'Three-colour CARS Spectroscopy of the OH Radical at Triple Resonance', *Mol. Phys.*, **70**, 1–51 (1990).
98. M.S. Chou, A.M. Dean, D. Stern, 'Laser Absorption Measurements on OH, NH and NH_2 in NH_3/O_2 Flames: Determination of an Oscillator Strength for NH_2', *J. Chem. Phys.*, **76**, 5334–5340 (1982).
99. K. Kohse-Höinghaus, U. Meier, B. Attal-Trétout, 'Laser-induced Fluorescence Study of OH in Flat Flames of 1–10 bar Compared with Resonance CARS Experiments', *Appl. Opt.*, **29**, 1560–1569 (1990).
100. C.D. Carter, G.B. King, N.M. Laurendeau, 'Saturated Fluorescence Measurements of the Hydroxyl Radical in Laminar High-pressure $C_2H_6/O_2/N_2$ Flames', *Appl. Opt.*, **31**, 1511–1522 (1992).
101. T. Dreier, D. Rakestraw, 'Measurement of OH Rotational Temperatures in a Flame Using Degenerate Four Wave Mixing', *Opt. Lett.*, **15**, 72–74 (1990).
102. H. Bervas, B. Attal-Trétout, S. Le Boiteux, J.P. Taran, 'OH Detection and Spectroscopy by DFWM in Flames: Comparison with CARS', *J. Phys. B: At., Mol. Opt. Phys.*, **25**, 949–969 (1992).
103. T. Dreier, D.J. Rakestraw, 'Degenerate Four-wave Mixing Diagnostics on OH and NH Radicals in Flames', *Appl. Phys. B*, **50**, 479–485 (1990).
104. R.L. Van der Wal, B.E. Holmes, J.B. Jeffries, P.M. Danehy, R.L. Farrow, D.J. Rakestraw, 'Detection of HF Using Infrared Degenerate Four-wave Mixing', *Chem. Phys. Lett.*, **191**, 251–258 (1992).
105. C.F. Kaminski, I.G. Hughes, P. Ewart, 'Degenerate Four-wave Mixing Spectroscopy and Spectral Simulation of C_2 in an Atmospheric Pressure Oxy-acetylene Flame', *J. Chem. Phys.*, **106**, 5324–5332 (1997).
106. L.A. Rahn, P.L. Mattern, R.L. Farrow, 'A Comparison of Coherent and Spontaneous Raman Combustion Diagnostics', 18th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 1533–1542, 1981.
107. T. Dreier, J. Wolfrum, 'Detection of the free OH ($X^2\Pi$) radical by CARS Spectroscopy', *J. Chem. Phys.*, **80**, 975–976 (1984).
108. T. Dreier, J. Wolfrum, 'Detection of free NH_2 (X^2B_1) Radicals by CARS Spectroscopy', *Appl. Phys. B*, **33**, 213–218 (1984).
109. K. Nyholm, M. Kaivola, C.G. Aminoff, 'Polarization Spectroscopy Applied to C_2 Detection in a Flame', *Appl. Phys. B*, **60**, 5–10 (1995).
110. K. Nyholm, 'Measurements of OH Rotational Temperatures in Flames by Using Polarization Spectroscopy', *Opt. Comm.*, **111**, 66–70 (1994).
111. W. Juchmann, H. Latzel, D.I. Shin, G. Peiter, T. Dreier, H.-R. Volpp, J. Wolfrum, R.P. Lindstedt, K.M. Leung, 'Absolute Radical Concentration Measurements and Modeling of Low Pressure CH_4/O_2/NO Flames', 27th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 461–467, 1998.
112. M. Bachmann, J. Griesheimer, K.-H. Homann, 'Thermal and Chemical Influences on the Soot Mass Growth', Twenty-fifth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 635–643, 1994.
113. M. Hausmann, K.H. Homann, 'Scavenging of Hydrocarbon Radicals from Flames with Dimethyl–Disulfide II. Hydrocarbon Radicals in Fuel-rich Low-pressure Flames of Acetylene, Ethylene, 1,3-Butadiene and Methane with Oxygen', *Ber. Bunsenges. Phys. Chem.*, **101**, 651–667 (1997).

114. J. Ahrens, A. Keller, R. Kovacs, K.-H. Homann, 'Large Molecules, Radicals, Ions and Soot Particles in Fuel-rich Hydrocarbon Flames. Part III: REMPI Mass Spectrometry of Large Flame PAHs and Fullerenes and their Quantitative Calibration through Sublimation', *Ber. Bunsenges. Phys. Chem.*, **102**, 1823–1839 (1998).
115. M. Woyde, W. Stricker, 'The Application of CARS for Temperature Measurements in High Pressure Combustion Systems', *Appl. Phys. B*, **50**, 519–525 (1990).
116. T. Dreier, M. Ridder, G. Schiff, A. Saur, A.A. Suvernev, 'Determination of Temperature from N_2 and O_2 CARS Spectra at Very High Pressure', Twenty-fifth Symposium (International) on Combustion, The Combustion Institute, Pittsburg, 1727–1734, 1994.
117. L. Martinsson, P.-E. Bengtsson, M. Aldén, S. Kröll, 'Applications of Rotational CARS for Temperature Measurements at High Pressure and in Particle-laden Flames', in *Temperature: Its Measurements and Control in Science and Industry*, ed. J.F. Schooley, American Institute of Physics, New York, 679–684, Vol. 6, 1992.
118. R.L. Vander Wal, R.L. Farrow, D.J. Rakestraw, 'High-resolution Investigation of Degenerate Four-wave Mixing in the $\gamma(0, 0)$ Band of Nitric Oxide', Twenty-fourth Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 1653–1659, 1992.
119. S.J. Tsay, K.G. Owens, K.W. Aniolek, D.L. Miller, N.P. Cernansky, 'Detection of CN by Degenerate Four-wave Mixing', *Opt. Lett.*, **20**, 1725–1727 (1995).
120. A.P. Smith, G. Hall, B.J. Whitaker, A.G. Astill, D.W. Neyer, P.A. Delve, 'Effect of Inert Gases on the Degenerate Four-wave Mixing Spectrum of NO_2', *Appl. Phys. B*, **60**, 11–18 (1995).
121. G. Hall, A.G. Suits, B.J. Whitaker, 'Resonant Degenerate Four Wave Mixing Detection of HCO', *Chem. Phys. Lett.*, **203**, 277–282 (1993).
122. B. Löfstedt, R. Fritzon, M. Aldén, 'Investigation of NO Detection in Flames Using Polarization Spectroscopy', *Appl. Opt.*, **35**, 2140–2146 (1996).
123. K. Nyholm, R. Fritzon, M. Aldén, 'Two-dimensional Imaging of OH in Flames by Use of Polarization Spectroscopy', *Opt. Lett.*, **18**, 1672–1674 (1993).
124. K. Nyholm, R. Fritzon, M. Aldén, 'Single-pulse Two-dimensional Temperature Imaging in Flames by Degenerate Four-wave Mixing and Polarization Spectroscopy', *Appl. Phys. B*, **59**, 37–43 (1994).
125. A.A. Suvernev, A. Dreizler, T. Dreier, J. Wolfrum, 'Polarization-spectroscopic Measurement and Spectral Simulation of $OH(A^2\Sigma-X^2\Pi)$ and $NH(A^3\Pi-X^3\Sigma)$ Transitions in Atmospheric Pressure Flames', *Appl. Phys. B*, **61**, 421–427 (1995).
126. T. Reichardt, R.P. Lucht, 'Theoretical Calculation of Line Shapes and Saturation Effects in Polarization Spectroscopy', *J. Chem. Phys.*, **109**, 5830–5843 (1998).
127. Q.-V. Nguyen, P.H. Paul, 'The Time Evolution of a Vortex-flame Interaction Observed via Planar Laser-induced Fluorescence Imaging of CH and OH', 26th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 357–364, 1996.
128. J.B. Kelman, A.R. Masri, 'Reaction Zone Structure and Scalar Dissipation Rates in Turbulent Diffusion Flames', *Appl. Opt.*, **36**, 3506–3514 (1997).
129. A.R. Masri, R.W. Dibble, R.S. Barlow, 'The Structure of Turbulent Nonpremixed Flames Revealed by Raman–Rayleigh LIF Measurements', *Prog. Energy Combust. Sci.*, **22**, 307–362 (1996).
130. M. Decker, A. Schik, U.E. Meier, W. Stricker, 'Quantitative Raman Imaging Investigations of Mixing Phenomena in High-pressure Cryogenic Jets', *Appl. Opt.*, **37**, 5620–5627 (1998).
131. M. Péalat, P. Bouchardy, M. Lefebvre, J.P. Taran, 'Precision of Multiplex CARS Temperature Measurements', *Appl. Opt.*, **24**, 1012–1022 (1985).
132. R.P. Lucht, D. Dunn-Rankin, T. Walter, T. Dreier, S.C. Bopp, 'Heat Transfer in Engines: Comparison of CARS Thermal Boundary Layer Measurements and Heat Flux Measurements', Proceedings of International Congress, Detroit, SAE Technical Paper Series, 910722, 1–16, 1991.
133. A.C. Eckbreth, G.M. Dobbs, J.H. Stufflebeam, P.A. Tellex, 'CARS Temperature Measurements in Augmented Jet Engine Exhausts', *Appl. Opt.*, **23**(9), 1328 (1984).
134. A.C. Eckbreth, T.J. Anderson, J.A. Shirley, 'Laser Raman Diagnostics for Propulsion Systems Development', *Ber. Bunsenges. Phys. Chem.*, **97**, 1597–1608 (1993).
135. R. Bédué, R. Gastebois, R. Bailly, M. Pealat, J.P. Taran, 'CARS Measurements in a Simulated Turbomachine Combustor', *Comb. Flame*, **57**, 141–153 (1984).
136. B. Lavorel, G. Millot, J. Bonamy, D. Robert, 'Study of Rotational Relaxation Fitting Laws from Calculation of SRS N_2 Q-Branch', *Chem. Phys.*, **115**, 69–78 (1987).
137. G. Millot, R. Saint-Loup, S. Santos, R. Chaux, H. Berger, J. Bonamy, 'Collisional Effects in the Stimulated Raman Q-Branch of O_2 and O_2-N_2', *J. Chem. Phys.*, **96**, 961–971 (1992).
138. L.A. Rahn, R.L. Farrow, G.J. Rosasco, 'Measurement of the Self-broadening of the H_2 Q(0-5) Raman Transitions from 295 to 1000 K', *Phys. Rev. A*, **43**, 6075–6088 (1991).
139. J.P. Looney, G.J. Rosasco, L.A. Rahn, W.S. Hurst, J.W. Hahn, 'Comparison of Rotational Relaxation Rate Laws to Characterize the Raman Q-Branch Spectrum of CO at 295 K', *Chem. Phys. Lett.*, **161**, 232–238 (1989).
140. B. Lavorel, G. Millot, G. Fanjoux, R. Saint-Loup, 'Study of Collisional Effects on Bandshapes of the $\nu_1/2\nu_2$ Fermi Dyad in CO_2 Gas with Stimulated Raman Spectroscopy. III. Modeling of Collisional Narrowing and Study of Vibrational Shifting and Broadening at High Temperature', *J. Chem. Phys.*, **101**, 174–177 (1994).

141. W. Kiefer, in *Non-linear Raman Spectroscopy and Its Chemical Applications*, eds. W. Kiefer, D.A. Long, NATO Advanced Study Institutes Series C: Mathematical and Physical Sciences, D. Reidel, Dordrecht, 241–260, Vol. 93, 1982.
142. T.M. Brugman, R. Klein-Douwel, G. Huigen, E. van Walwijk, J.J. ter Meulen, 'Laser-induced Fluorescence Imaging of NO in an n-Heptane and Diesel Fuel Driven Diesel Engine', *Appl. Phys. B*, **57**, 405–410 (1993).
143. F. Hildenbrand, C. Schulz, V. Sick, E. Wagner, 'Investigation of Spatially Resolved Light Absorption in an SI Engine', *Appl. Opt.*, **38**, 1452–1458 (1999).
144. C. Schulz, B. Yip, V. Sick, J. Wolfrum, 'A Laser-induced Fluorescence Scheme for Measuring Nitric Oxide in Engines', *Chem. Phys. Lett.*, **242**, 259–264 (1995).
145. C. Schulz, V. Sick, J. Wolfrum, V. Drewes, M. Zahn, R. Maly, 'Quantitative 2D Single-shot Imaging of NO Concentrations and Temperatures in a Transparent SI Engine', 26th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 2597–2604, 1996.
146. M.D. DiRosa, R.K. Hanson, 'Collisional Broadening and -Shift of NO $\gamma(0,0)$ Absorption Lines by O_2 and H_2O at High Temperatures', *J. Quant. Spectrosc. Radiat. Transfer*, **52**, 515–529 (1994).
147. F. Hildenbrand, C. Schulz, V. Sick, G. Josefsson, I. Magnusson, Ö. Andersson, M. Aldén, 'Laser Spectroscopic Investigation of Flow Fields and NO-formation in a Realistic SI Engine', *J. Engines*, **107**, 205–214 (1998).
148. G. Grünefeld, V. Beushausen, P. Andresen, 'Spatially Resolved Raman Scattering for Multi-species and Temperature Analysis in Technically Applied Combustion Systems: Spray Flame and Four-cylinder In-line Engine', *Appl. Phys. B*, **58**, 333–342 (1994).
149. R.P. Lucht, D. Dunn-Rankin, T. Walter, T. Dreier, S.C. Bopp, 'Heat Transfer in Engines: Comparison of CARS Thermal Boundary Layer Measurements and Heat Flux Measurements', Proceedings of International Congress, Detroit, SAE Technical Paper Series, 910722, 1–16, 1991.
150. D. Klick, K.A. Marko, L. Rimai, 'Broadband Single-pulse CARS Spectra in a Fired Internal Combustion Engine', *Appl. Opt.*, **20**, 1178–1185 (1981).
151. R.M. Mihalcea, D.S. Baer, R.K. Hanson, 'Diode-laser Sensor for Measurements of CO, CO_2 and CH_4 in Combustion Flows', *Appl. Opt.*, **36**, 8745–8752 (1997).
152. M.P. Arroyo, S. Langlois, R.K. Hanson, 'Diode Laser Absorption Technique for Simultaneous Measurements of Multiple Gasdynamic Parameters in High-speed Flows Containing Water Vapor', *Appl. Opt.*, **33**, 3296–3307 (1994).
153. E.R. Furlong, R.M. Mihalcea, M.E. Webber, D.S. Baer, T.P. Parr, R.K. Hanson, 'Diode-laser Sensor System for Closed-loop Control of a 50-kW Incinerator', AIAA 33rd Joint Propulsion Conference, Seattle, AIAA, Reston, VA, Paper No. 97-2833, 1997.
154. V. Ebert, J. Fitzer, I. Gerstenberg, K.-U. Pleban, H. Pitz, J. Wolfrum, M. Jochem, J. Martin, 'Simultaneous Laser-based In situ Detection of Oxygen and Water in a Waste Incinerator for Active Combustion Control Purposes', 27th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 1301–1308, 1998.
155. W. Meienburg, H. Neckel, J. Wolfrum, 'In situ Measurement of Ammonia Concentration in Industrial Combustion Systems', Twenty-third Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 231–236, 1990.
156. F. Schuler, F. Rampp, J. Martin, J. Wolfrum, 'TACCOS – A thermography-assisted Combustion Control System for Waste Incinerators', *Combust. Flame*, **99**, 431–439 (1994).
157. V. Ebert, K.-U. Pleban, J. Wolfrum, 'In situ Oxygen Monitoring Using Near-infrared Diode Lasers and Wavelength Modulation Spectroscopy', in *Laser Applications to Chemical and Environmental Analysis*, Technical Digest (Optical Society of America, Washington DC), Paper No. LWB3, 206–209, 1998.
158. R.C. Oldenburg, S.L. Baughcum, 'Photofragment Fluorescence as an Analytical Technique: Application to Gas-phase Alkali Chlorides', *Anal. Chem.*, **58**, 1430–1436 (1986).
159. F. Greger, K.T. Hartinger, P.B. Monkhouse, J. Wolfrum, H. Baumann, B. Bonn, 'In Situ Alkali Concentration Measurements in a Pressurized Fluidized Bed Coal Combustor by Excimer Laser-induced Fragmentation Fluorescence', 26th Symposium (International) on Combustion, The Combustion Institute, Pittsburgh, 3301–3307, 1996.
160. K.T. Hartinger, S. Nord, P.B. Monkhouse, 'Quenching of Fluorescence from $Na(3^2P)$ and $K(4^2P)$ Atoms following Photodissociation of NaCl and KCl at 193 nm', *Appl. Phys. B*, **64**, 363–367 (1997).

Laser-induced Breakdown Spectroscopy, Elemental Analysis

Fang-Yu Yueh, Jagdish P. Singh, and Hansheng Zhang
Mississippi State University, Starkville, USA

1 Introduction 1
2 Fundamental Studies 2
2.1 Effect of Physical Variables 3
2.2 Analytical Characteristics 5
2.3 Interference 8
3 Experimental Techniques 9
3.1 Experimental Methods 9
3.2 Calibration Methods 12
4 Environmental Applications 13
4.1 Off-gas Emission 13
4.2 Metal and Glass 14
4.3 Soil, Concrete, and Paint 15
4.4 Air Sampling Filters 15
4.5 Radioactive Materials 16
5 Comparison with Other Spectroscopy Techniques 16
5.1 Sample Preparation 16
5.2 Sampling Rate 17
5.3 Detection Limit 17
5.4 Accuracy and Precision 18
6 Future Developments 19
Acknowledgments 19
Abbreviations and Acronyms 19
Related Articles 20
References 20

Laser-induced breakdown spectroscopy (LIBS) is a laser-based technique that can provide nonintrusive, qualitative and quantitative measurement of metals in various test environments. LIBS is an emission-type technology that has been successfully applied to gas, liquid, and solid samples. The major advantages of LIBS compared to other analytical techniques is that no time-consuming sample preparation is necessary. LIBS uses the plasma generated by a high-energy laser beam to prepare and excite the sample in one step. It has the ability to perform multielement real-time analysis. Its major disadvantage is that the excitation condition is sensitive to the fluctuation of environmental conditions as well as the laser energy, which can result in poor measurement precision. The small amount of sample material used in LIBS analysis also gives poorer sensitivity for some metals compared to inductively coupled plasma atomic emission spectrometry (ICPAES) and atomic absorption spectrometry (AAS).

The potential of LIBS to detect toxic metals in harsh environments was recognized in the early 1970s. Recent developments towards improving its analytical capability has led to more applications. This article reviews the analytical applications of LIBS with an emphasis on environmental monitoring. A brief review of some fundamental LIBS studies is also given. The analytical abilities of LIBS are compared with some spectroscopic techniques commonly used in the laboratory, such as AAS, ICPAES, and X-ray fluorescence spectroscopy (XFS).

1 INTRODUCTION

Laser-induced breakdown phenomena were observed just after the invention of the laser in 1963. Initial work was focused on understanding the various mechanisms involved in the process. As the material is atomized and excited at the high temperature of laser spark plasma, the potential for using LIBS for elemental analysis was soon recognized by many researchers. The application of LIBS to gases, liquid, and solid samples was explored by various groups from 1970 to 1985. Compared to other analytical techniques, LIBS uses very small amounts of samples, and no sample preparation is necessary. It has the ability to perform real-time analysis because it prepares and excites the sample in one step. The disadvantage of LIBS is that the plasma condition varies with the environmental condition as well as laser energy fluctuation, resulting in poor measurement precision. The microgram-sized samples used in LIBS also causes lower sensitivity for some elements. Since the late 1980s, many studies on LIBS characteristics for quantitative measurements have been initiated. Techniques to improve the precision and sensitivity of LIBS are also being tested. This ongoing development has improved the analytical capabilities of LIBS and led to more applications.

The application of LIBS to environmental monitoring has drawn significant attention. LIBS has been used to detect toxic metals in soil and paint, and has been explored as a multimetal continuous emission monitor (CEM). There are a few reports on the application of LIBS to detect radioactive species. LIBS has also being used to measure the composition of molten metals and glasses. This article reviews the application of LIBS to analytical problems with an emphasis on environmental applications. Detailed reviews on the LIBS principle and other

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

applications can be found.[1–5] Brief reviews of some fundamental and experimental application of LIBS are given in sections 2 and 3 of this article. Environmental monitoring applications are discussed in section 4. LIBS's quantitative capability is addressed in section 5. A brief discussion of future developments is given in the last section.

2 FUNDAMENTAL STUDIES

The physics of laser-induced breakdown has been studied extensively.[6,7] The minimum optical power density required to form a plasma is called the breakdown threshold; different types of laser, sample, and ambient conditions will have different breakdown thresholds. A laser power density greater than $10^{10}\,W\,cm^{-2}$ is generally required to generate a breakdown in air using a Q-switched Nd : YAG laser of 10–15 ns duration. Breakdown thresholds of solids and liquids are usually much lower than for gases. Table 1 lists some measured breakdown thresholds in different media.[8–11]

The two mechanisms responsible for electron generation and for growth in laser-induced breakdown are multiphoton absorption and collision-induced ionization. Multiphoton absorption involves simultaneous absorption of a certain number of photons by an atom or molecule, to cause its ionization. This mechanism produces the initial few free electrons in the focal volume. The electron density grows linearly with time. Generally, this mechanism is not the major process in the growth of electrons in laser plasma. It is more important for a short-wavelength laser ($\lambda < 1\,\mu m$) or low density ($P < 10$ torr in N_2).[12] In the collision-induced ionization process, free electrons in the focal volume are accelerated by the electric field of the laser and gain energy by colliding with neutral atoms. After the electrons have gained sufficient energy, they can ionize atoms by collision and this causes the electron density to grow exponentially with time. This process is dominant at high pressure ($P > 100$ torr in N_2) and long wavelength ($\lambda > 1\,\mu m$).[12] The breakdown threshold in this process has a λ^{-2} or λ^{-3} dependence.[12]

Table 1 Breakdown thresholds for various media

Test medium	Laser type	
	Ruby (0.69 μm) (10^{10} W cm^{-2})	Nd^{+3} (1.06 μm) (10^{10} W cm^{-2})
Gas[a]		
Air	20	7
Ar	8	2
He	15	4
N_2	21	9
Liquid		
Tap water[b]	3.8	10
Distilled water[c]		62
Benzene[c]		4.4
Solid[d]		
NaCl	15	14
KBr	5.7	5

[a] Pressure = 1 atm; ruby, $T_p = 40$ ns; Nd^{3+}, $T_p = 40$ ns.[8]
[b] Ruby, $T_p = 21$ ns, Nd^{3+}, $T_p = 50$ ns.[9]
[c] Nd^{3+}, $T_p = 50$ ns.[10]
[d] Ruby, $T_p = 10$ ns; Nd^{3+}, $T_p = 10$ ns.[11]
T_p is the laser pulse width.

The characterization of laser-induced plasma (LIP) has been carried out both in time and spatial evolution of the plasma. The evolution of laser plasma can be divided into several transient phases. The initial plasma (0–100 ns) is characterized by high electron and ion densities (10^{17}–10^{20} cm^{-3}), and temperatures around 20 000 K. The emission spectrum from the early stage of the plasma is characterized by a continuum background emission due to bremsstrahlung (strong collisions between the free electrons and the excited atoms and ions) and recombination of electrons with ions processes in the plasma. Emission lines from ions and atoms can be found after about 300 ns delay. These lines are superimposed on the strong continuum background. Owing to the high electron density, the emission lines are broadened by the Stark effect. As the plasma expands and cools, the electrons and ions recombine, and a characteristic loud sound can be heard due to the shock wave emitted from the plasma. The continuum background decays rapidly and the atomic emission lines become narrower and weaker. After the initial plasma (>10 μs), the atomic emission decays slowly, and emission from the simple molecule appears. Figure 1 shows LIBS spectra recorded at delay times of 0.25, 5, and 35 μs. As shown in the figure, a properly selected detection window can improve the signal-to-background ratio (S/B) of the data, which is very important to quantitative measurement.

The excitation characteristics of the laser plasma are determined by the plasma properties, which can be described by the temperature and electron density of the plasma. The analytical measurements are generally performed after the initial plasma, when a state of local thermal equilibrium (LTE) is achieved. At the LTE condition, the analyte intensity is proportional to the relative population of the level, and follows the Boltzmann distribution. The Griem criterion (Equation 1) is usually used to check the existence of LTE in a laser plasma[13]

$$N_e(\mathrm{cm}^{-3}) \gg 30\,545 \times 10^{17} \times T_e(\mathrm{K})^{1/2} \times \left(\frac{Z^2}{\lambda(\mathrm{nm})}\right)^3 \quad (1)$$

where N_e, Z, and λ are the critical electron density for LTE, the degree of ionization, and the transition wavelength, respectively.

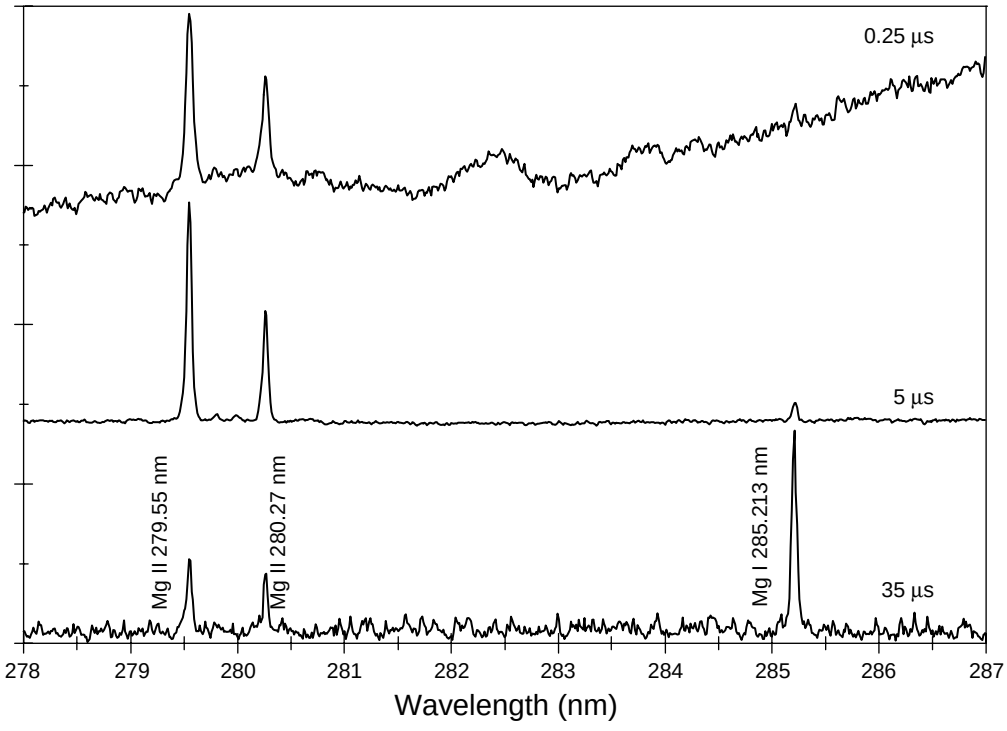

Figure 1 LIBS spectra recorded at different delay times.

Under LTE, the intensity of an emission line at v_{ji} can be expressed (Equation 2) as

$$I_{ji} = n(hv_{ji})\frac{A_{ji}g_j}{Q(T)}\,\mathrm{e}^{-hcE_j/kT} \qquad (2)$$

Here, n is the total number density of a neutral or ion of the element, g_j is the statistical weight of the excited state, A_{ji} is the transition probability, E_j is the energy of the excited state in units of cm^{-1}, h is Planck's constant, k is the Boltzmann constant, T is the plasma temperature, and $Q(T)$ is the partition function.

The details of LIP are described elsewhere. Here, only the characteristic and physics parameters are discussed, which are important to quantitative analysis of LIBS. At present, the physics of LIP is not fully understood but work in this area continues.

2.1 Effect of Physical Variables

Variables that can influence LIBS measurements are laser properties (i.e. wavelength, energy, pulse duration and shot-to-shot power fluctuation), focusing spot size, ambient conditions, physical properties of the sample, and the detection window (delay time, gate width). How these parameters affect the precision and accuracy of LIBS are addressed below.

2.1.1 Laser Parameters

In LIBS, a high-energy laser beam is used to form a plasma to prepare and excite the sample. As the laser plasma is the excitation source of the LIBS measurement, the properties of the laser can influence the analytical qualities of the LIBS measurement. Regardless of the type of laser used, a power density of the order of $10^9\ \mathrm{W\,cm^{-2}}$ is required. If the laser energy is very close to the breakdown threshold, the pulse-to-pulse fluctuation can cause the plasma condition to be unreproducible, which results in poor measurement precision. The LIBS signal is proportional to the laser energy while the laser plasma is in the optical thin region. Figure 2 shows the dependence of the analyte line intensity and background on laser energy. When the laser energy further increases, it produces a very dense and hot plasma that can absorb laser energy and cause self-absorption. This will lead to an increase in the continuum emission and a decrease in the signal intensity. Laser wavelength can also affect the plasma-formation process. Collision-induced ionization is the dominant process for LIP with a long-wavelength laser. As the laser wavelength decreases, the collision-induced ionization rate decreases, and the multiphoton process increases. A maximum breakdown threshold is found in the intermediate laser wavelength. Laser wavelength can also affect energy coupling between the laser and the sample. Studies have shown that the use of an ultraviolet (UV) wavelength can improve the energy

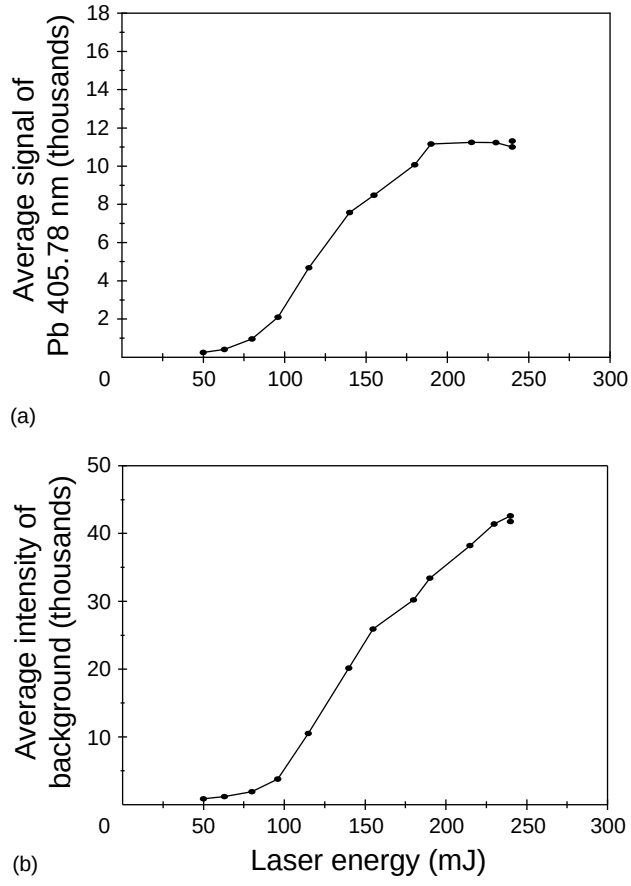

Figure 2 Dependences on the laser energy of (a) the LIBS signal and (b) the background intensity.

coupling efficiency due to its low reflectivity for most metals.[14]

Studies of laser wavelength on material removal from the sample surface have been performed by many researchers.[15,16] Dahmain[17] has also derived the mass-ablation rate with absorbed laser intensity I_a, laser wavelength λ, and atomic number Z based on experimental data for a laser energy lower than 10^{13} W cm^{-2} and $Z \leq 13$ (Equation 3):

$$m(\text{kg s}^{-1}\,\text{cm}^{-2}) = 65\left(\frac{I_a\,\text{W cm}^{-2}}{10^{13}}\right)^{5/9} \lambda^{-4/9}(\mu\text{m})Z^{1/4} \quad (3)$$

Besides laser wavelength and laser energy, the laser pulse duration and laser shot-to-shot fluctuation can also affect the signal reproducibility and, hence, LIBS precision. When the laser duration is short compared to electron energy loss time, multiphoton absorption dominates over collision-induced ionization. A study has shown that the threshold intensity increases with decreasing laser pulse duration for laser pulses shorter than 10^{-7} s at atmospheric pressure.[18] A 1000-fold higher breakdown threshold was found with an Nd : glass laser of about 10 ps than for a nanosecond width pulse in atmospheric air.[19]

2.1.2 Physical Properties of the Sample

All types of material can be sampled using LIBS. However, the physical properties (such as reflectivity of the sample surface, density, specific heat, and boiling point of the target) can have a considerable influence on the result. The surface reflectivity determines the fraction of the laser energy that can be absorbed by a sample, and hence that can affect the material ablation rate. Study has shown that if the laser energy is high enough, the laser energy can still effectively couple to a sample with high reflectivity. This is because, during a laser pulse, the laser energy causes rapid heating of the sample and this results in a sample phase change which significantly reduces the reflectivity.

The laser plasma heats the sample to cause the vaporization and atomization of the sample material. The amount of material vaporized depends on thermal properties of the sample (i.e. conductivity, specific heat, and boiling point). At a low laser power densities, thermal conductivity is the most important parameter for material vaporization. If the material has a high thermal conductivity, the absorbed heat is conducted away quickly and results in a lower amount of material being vaporized. At higher laser power densities, the heat provided is too quickly to be conducted away; therefore, the latent heat of vaporization of the sample becomes a more important factor in determining the amount of material vaporized during a laser pulse.

2.1.3 Focal Properties

High laser power densities can produce a LIP which has higher excitation temperatures with minimum matrix interferences due to more complete atomization. It is a condition preferable to quantitative measurement. The laser power density at the focal volume is inversely proportional to the focused spot size. For a laser beam with a Gaussian profile, the focused beam waist (w_0) is given by Equation (4)

$$w_0 = \frac{\lambda f}{\pi D} \quad (4)$$

where D is the radius of the unfocused beam, f is the focal length of the lens, and λ is the laser wavelength. A higher laser power density at the focal point can be achieved by reducing the focused beam waist using a shorter focal length lens or larger unfocused beam.

Effects of the laser-spot diameter on the sample ablation rates were studied by Wolff-Rottke et al.[20] They found the ablation rate increased with decreasing

spot diameter in the range 10–200 μm for a pulse laser. However, if the size of the spark is very small and the distribution of material within it is inhomogeneous, the precision and accuracy of the measurement is degraded.

A short focal length lens is preferred to produce a highly localized spark for a spatially resolved measurement. A long focal length lens provides a larger focal volume and is usually used when the focal lens cannot be kept close to the sample. This situation requires higher laser energy to produce the laser spark, and the larger focal volume associated with the long focal length can cause particle-induced sparks – formed at different locations along the focal axis – and this introduces noise into measurements.

The lens-to-surface distance (LTSD) is a critical parameter for the LIBS measurement of a solid sample. If the laser beam is tightly focused on the surface, a change of the LTSD of a few millimeters can affect the absolute analyte intensity. Studies have shown that by defocusing the laser beam on the sample surface a more reproducible LIBS signal can be obtained.[21] Therefore, this technique can be used to improve LIBS precision.

2.1.4 Detection Time Window

In LIBS, the desired atomic emission signal is always accompanied by a strong continuum background. As the continuum background and analyte signal decay with different rates, it is possible to use a time-resolved technique to discriminate against the strong continuum radiation and also avoid spectral interference between species that emit at different times during the plasma decay (see Figure 1). The detection window is selected to obtain an optimum S/B. As the background signal decays at a faster rate than the emission signal, the best S/B can be achieved at a long delay time. The long delay time also ensures that LTE is achieved. Generally, signal-to-noise ratio (S/N) and S/B values follow the same trend as the delay time changes. However, a low S/B can sometimes give a good S/N. Therefore, the selection of the best detection window should depend on both S/B and S/N.

2.1.5 Ambient Conditions

The hot plasma interacts with the surrounding gas by expanding the high-pressure vapor, driving a shock wave into the atmosphere and transferring the energy to the atmosphere via heat transfer and heating by the shock wave. The size and shape of the LIP are largely dependent on the ambient conditions such as pressure, gas composition, mass density of the gas, etc. The breakdown threshold of gas near atmospheric pressure (10^2–10^3 torr) has been examined.[6] A pressure dependence of P^{-m}, where m is close to 1, was found for pulse widths of 10–100 ns. This is consistent with the electron atom collision frequency's dependence on pressure. Kuzuya et al. have taken plasma images at different pressures.[22] They found that the spatial confinement of the plasma becomes stronger with increasing gas pressure. The confining effect produces a denser and hotter plasma which could cause an increase in both the emission period and the emission intensity. At low pressures, owing to the weak confining effect of the plasma by the surrounding gas, the laser-produced plasma expands rapidly and becomes thin, resulting in a decrease in the emission intensity. At moderately high pressures, high-temperature and high-density plasma is generated, and the intensity of the continuous emission spectrum is very high; also, the self-absorption effect takes place to a remarkable extent due to the increase in concentration of the absorbing species surrounding the hot plasma. As a result, one is faced with difficulty in terms of precision and sensitivity. To obtain intense emission spectra free from self-absorption, LIBS measurements should be performed at a moderate pressure range ($50 < P < 500$ torr in air).

The effects of various atmospheres including Ar, air, O_2, N_2, He on LIBS spectra have been studied.[22,23] The inert gas acts like a buffer gas to prevent rapid oxidation of the free atom in the plasma. Argon breakdown occurred in the plasma at a relatively high pressure. The argon atmosphere can enhance the analytical signal by re-excitation of atoms, resulting from collision with photon-excited argon. LIP in the Ar atmosphere has a higher plasma temperature and longer emission period due to the low thermal conductivity of Ar. The higher plasma temperature in the Ar atmosphere also produces the highest continuum background. The emission characteristics of LIP from the air atmosphere are similar to that from the Ar atmosphere. However, the continuum background is only half of that in the Ar atmosphere. Helium has higher thermal conductivity and higher ionization potential than nitrogen or argon. Hence, it has a high-energy coupling efficiency compared with Ar or air. Therefore, background from a He atmosphere is lower and less sensitive to changes in laser energy and pressure.

2.2 Analytical Characteristics

In LIBS, the material in the laser spark is immediately reduced to elemental form, and the resulting atoms are electronically excited due to the high plasma temperature. The characteristic atomic and ionic emissions from the excited species are detected and used for analysis. Owing to the rapid expansion and the subsequent cooling effect, the analytical signal from a laser plasma is time dependent. In the first several microseconds, the continuum background emission dominates. As the plasma cools down, the background emission decreases

significantly and ionic and atomic line signals become the dominant spectral feature. Most LIBS work is performed with time-resolved measurements to avoid the strong continuum background in the initial plasma and to improve the S/B. Kagawa and Yokoi found that the LIP consists of two distinct regions when using an N_2 laser (337.1 nm) – primary plasma acts as an initial explosion energy source and emits an intense continuous emission background for a short time just above the surface of the target; secondary plasma expands with time around the primary plasma.[24] It is formed by excitation from the shock wave and by emitting sharp atomic emission lines with a low background signal. The shock wave travels in the opposite direction of the laser beam, to heat and ionize the gas it passes through. The gas is then more capable of absorbing laser radiation and becomes the heat source to maintain the shock wave in the initial stage. Studies have shown that emissions from secondary plasma are suitable for quantitative analysis. Several researchers have proposed the use of time-integrated LIBS in the secondary plasma for quantitative elemental measurement. Time-integrated LIBS does not require the expensive gated detection system necessary in time-resolved LIBS. Therefore, the cost of the LIBS system can be greatly reduced.

The theoretical analysis of emission lines assumes LTE in the environment. Assuming LTE, the plasma temperature and electron density can be used to understand the atomization, ionization, and excitation processes occurring in the plasma. The plasma temperature can be determined from Equation (2) using the relative intensity of two lines of a given atomic species. If more than two lines of intensities are available, the temperature can also be extracted through the Boltzmann plot method.[25] The concentration ratio can be determined from the intensity of the analytical lines free from spectral interference. For this measurement, analyte lines with similar upper energies should be chosen to compensate for changes in the plasma. The degree of ionization for different elements can be quite large, therefore the degree of ionization will also need to be considered in the calculation of the concentration ratio.

The interaction of the laser pulse with the sample strongly influences plasma properties. In a gas, most laser energy is used to excite the sample. But in liquids and solids a lot of laser energy goes into material vaporization and less energy is left for excitation. Therefore, the optimum experimental conditions for different samples will be different. Some characteristics of LIBS in different phases of samples are discussed below.

2.2.1 Solid Sample

Both conducting and nonconducting solid samples can be analyzed with LIBS. As the laser beam is focused on the sample surface, the sample material absorbs the laser energy to melt and vaporize the target material. The vapor absorbs more energy and forms a high temperature plasma near the surface. The plasma expands into the atmosphere and transfers its energy to the atmosphere. Highly ionized emission lines can be found close to the target surface, whereas singlely ionized and neutral emissions appear further away from the surface. As the laser breakdown threshold is lower in solids than in gas, generally less laser energy is needed for solid-sample measurement. The analyte signal is proportional to the laser energy. Increasing the laser energy could cause ambient gas breakdown prior to the formation of metal vapor plasma. This will influence the laser energy coupling to the target and result in a weak analyte signal, because laser energy is used to vaporize and excite the sample rather than to produce air breakdown. A lower plasma temperature is expected for solid-sample measurement. The short-wavelength laser seems to give a greater ablated mass and better S/B. Study has also shown that the relatively colder plasma generated with a short-wavelength laser can result in incomplete atomization and, hence, a poorer analytical result.[26] The interaction of the laser spark with the sample surface causes nanograms of material to be ablated from the surface and form a small crater on the sample surface. The size of the crater is independent of the sample material, but the depth of the crater depends on the type of matrix[27] and is proportional to the thermal conductivity of the sample material.

Hwang et al. used an ArF excimer laser to study laser energy with emission intensity for four metals (Cu, Zn, Ni alloy, and iron alloy).[28] They compared the experimental data with calculated mass removal $m(t)$ which was derived based on a heat conduction mechanism (Equation 5):

$$m(t) = AIt + BI^2t^{3/2} \tag{5}$$

where I is the laser irradiance and A and B are the constants determined by the thermal properties of the metal.

Good correlation between the experimentally obtained emission intensity and theoretically calculated mass using the above equation was found for the four metals studied. Experiment shows that the LIBS signal also changes with time when the same location is sampled by the laser pulse (Figure 3). As the laser continuously samples at the same location, the crater size changes and results in a time-varied LIBS signal. To obtain reproducible results, LIBS measurements should be taken after a certain ablation time when the signal is at a more stable level. If possible, the sample should also be slowly translated to ensure that the measurement is performed at a new spot each time.

Surface properties can affect the material ablated from the sample, and therefore can affect LIBS precision and

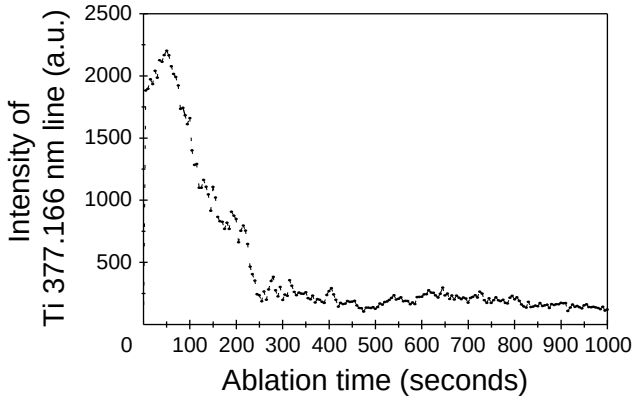

Figure 3 Dependence of the LIBS signal on the ablation time.

accuracy. As the laser spark only interacts with the sample surface, a clean surface that has the representative elements of the bulk sample is necessary. Surface contamination can greatly distort LIBS analysis. However, this problem can be solved by applying a few laser shots at the same spot to clean the sample surface, so that reliable LIBS signal can then be obtained. Sample homogeneity is very important for bulk analysis. For an inhomogeneous sample, averaging the data taken at various locations of the sample can improve the accuracy of the measurement. The sample roughness can change the lens-to-sample distance from shot to shot and, therefore, influence the plasma characteristics and also affect the LIBS precision. Therefore, choosing a smooth surface for LIBS analysis can also improve LIBS precision.

The presence of a buffer gas also affects the energy coupling to the target and, hence, the LIBS signal. Some spectral interference can also be deduced with the selected buffer gas. For example, the interference of C emission due to the dissociation of CO or CO_2 in air can affect the accuracy of carbon measurement in molten steel. Use of an Ar atmosphere can improve the C detection and minimize the interference from air atmosphere. At 1 atm, a high density and high temperature plasma is generated, emitting a strong continuum emission, and self-absorption takes place. This reduces the system sensitivity and the linearity of the calibration curve. One way to improve the detection sensitivity is to use gated detection. The other method is to reduce the pressure of the buffer gas. Lee et al. found the emission characteristics generated by different sample species were quite different.[29] For spectrochemical analysis, different detection windows are sometimes needed for different metals.[29]

Castle et al. conducted experiments to find the parameters which can be used to improve LIBS repeatability for solid samples.[30] The variables considered important to solids analysis are laser energy, degree of focusing, sample translation speed, buffer gas, surface contamination, and surface smoothness. They found the sample translation velocity, laser pulse energy, and the number of spectra averaged to be the most important factors to improve LIBS precision. Parameters such as gate delay, surface roughness, and background correction have the least effect on LIBS measurement precision.

2.2.2 Liquid Samples

The physical properties of the laser spark in liquids have been studied extensively.[31] The application of LIBS in liquid analysis was not demonstrated until 1984.[32] LIBS for liquid-phase analysis can be performed by focusing the laser beam in the bulk liquid or at the surface. Bulk liquid analysis does not require a stable free surface. It avoids the liquid splashing due to the plasma shock wave propagated through the liquid surface which can cause the contamination of the optical components and shot-to-shot signal fluctuation. However, it requires a liquid that is transparent at the wavelength of the laser and the analyte line. When plasma forms inside a liquid, the atomic and ionic emissions from the constituent elements have relatively short lifetimes compared to the spark in air. Also, emissions from the background continuum, neutral and ionized atoms, and simple molecules appear simultaneously. Therefore, the time-resolved measurement cannot effectively eliminate spectral interferences for liquid analysis. However, a significant improvement of S/B for liquid analysis has been demonstrated with a shorter laser wavelength.[33] Studies of the effect of laser wavelength on the analysis of liquid using LIBS found that a 532 nm laser beam produced a very strong continuum background interfering with the analyte signal, whereas a 193 nm laser beam produced several orders of magnitude higher analyte signal with a relatively dimmer plasma. The improvement on S/B with a short-wavelength laser is because the LIP process became less thermal, and the optical energy is effectively coupled into the target rather than heating up the plasma shield. This avoids violent ablation while the liquid is absorbing the laser energy, and makes single-shot measurement in liquids possible.

The LIBS detection limit in liquids is only sufficient for measurement of the minor and major species. To improve the LIBS detection limit in liquids, several techniques were tested. A technique utilizing two sequential laser pulses to sample the same volume of liquid has shown significant improvements in the detection limits of some elements in liquids. The two sparks are spatially overlapped and temporally separated. The first spark vaporizes the liquid, producing a gaseous cavity containing species characteristic of the liquid. The second spark excites the vaporized material. The detection limit for boron had been improved 15 times by using the sequential spark method.[32] However, the sensitivity of Li and

Na did not change with this double spark technique, and it is believed that these elements are fully excited by the single spark.

The detection limits of some elements were improved by forming the laser spark on the surface of the liquid, which produced a higher-temperature spark than the spark formed in the bulk liquid. Uranium could be detected in the bulk liquid at $300\,g\,L^{-1}$, whereas a detection limit of $0.1\,g\,L^{-1}$ can be achieved with the surface excitation technique.[34]

For LIBS measurements in liquids, it is also useful to send the laser beam through the side or the bottom of the mixture, to avoid optical interference from liquid splashing and the gas bubbles liberated as the liquid–gas surface.[32] Long spectral averaging should be used to improve LIBS shot-to-shot fluctuation in liquid application.

2.2.3 Gas Samples

LIP in gas has been studied extensively.[6] The breakdown in the atmospheric pressure gases were in the cascade ionization mechanism where the threshold was proportional to the ionization potential of the gas and inversely proportional to the collision frequency. The laser breakdown threshold depends on laser wavelength, gas pressure, and laser energy. The experimentally obtained breakdown threshold is much less than that predicted by theory, because the presence of aerosols causes aerosol-induced breakdown. The mechanism of aerosol breakdown is not fully understood, but experimental data shows that aerosols, acting as seeds, can significantly lower the breakdown threshold of clean gas. Smith and Brown have shown that aerosol-induced breakdown has a strong dependence on gas species, pressure, and laser wavelength, similar to that predicted for a collision-induced process.[35] They also demonstrated that, in the presence of micron-sized aerosols and impurity particles, the breakdown threshold in a gas is inversely proportional to the size of the focused laser beam. This is because the probability of finding a particle of sufficient size within the focal volume is higher.

Typically, laser-induced air breakdown has a plasma temperature of 20 000 K and an electron density of 10^{17}–$10^{18}\,cm^{-3}$ after the plasma is formed. Emission from CN can be observed from LIP in air. The CN is produced from the reaction between C and N, produced in the spark. The intensity of the CN bands depends on the concentration of a C-containing compound in the gas stream. The elements in the spectral region covered by the CN bond have less sensitivity due to this spectral interference.

LIBS can directly measure a gas sample, eliminating the need of sample preparation. Cremers and Radziemski detected Cl and F in air with detection limits of 80 ng and 2000 ng respectively.[36] They found that the absolute detection limit for Cl and F can be improved to 3 ng in an He atmosphere. Radziemski et al. used LIBS to detect Be, Na, P, As, and Hg in air.[37] Owing to the small sample volume and possible sample inhomogeneity, the measurement precision in a gas sample is generally poor. Variously sized particulates in the gas can cause breakdown to occur at different locations along the axis of the laser beam, and this leads to significant signal variation. The variables affecting the characteristics of LIBS spectra in a gas include particle size, gas pressure, temperature, and laser energy. For quantitative measurements, the calibration procedure should keep these variables as close to the measurement conditions as possible.

2.3 Interference

Three types of interferences – self-absorption, spectral overlap (spectral line interference, band interference), and matrix effect (chemical interferences) – are generally encountered in LIBS. Self-absorption occurs when the emission from the hotter region is absorbed by cool atoms surrounded by the high-temperature core of the laser plasma. To avoid the problems with self-absorption, resonance lines should only be used for the measurement of trace elements.

Spectral interference due to line overlap is very common, because the resolution depends upon the monochromator used. Care should be taken to avoid spectral interference lines in the analysis. Emission lines are often superimposed on bands emitted by oxides and other molecular species from air or samples. A background correction for band emission is necessary. The absolute intensity of an analyte line can be obtained by subtracting the peak height from the background signal near the analyte line.

The physical and chemical properties of the sample can affect the plasma composition, a phenomenon known as the matrix effect. The matrix effect can result in the sample being ablated differently from the target sample. This was found in measurements of Zn–Cu alloy foil, for which the Zn/Cu weight ratio was found to be as much as 30% greater than that of the sample (independent of the laser energy, number of laser pulses, and linear to the weight ratio of the sample).[4] The matrix effect may be due to the incomplete vaporization (atomization) in LIP. Some researchers suggest including a correction factor in the calibration to take into account the matrix effect.

Sometimes interferences from the same element, dissociated from the molecular species in the atmosphere, can cause a higher inferred concentration. In carbon measurement in air, atmospheric CO_2 near the laser

spark is dissociated, and the resulting carbon atoms are added to those arising from the sample. This leads to a higher inferred carbon concentration. To improve the carbon measurement, LIBS has to be performed in a CO_2-free environment. Aguilera et al. have shown that, for carbon measurement in steel, a nitrogen environment effectively eliminates the residual carbon emission.[38]

Other types of interference include signal fluctuations due to particles and detector saturation by a very intense signal. In addition to experimental precaution, certain data selection schematics are sometimes needed to exclude some distorted data. The resonances lines are not often the most suitable lines for quantitative analysis, because they are susceptible to self-absorption. Reducing the pressure of the surrounding atmosphere can minimize the occurrence of self-reversed emission.

As the plasma condition fluctuates from pulse to pulse, and it is difficult to control the variables associated with the excitation condition, an internal standard is generally employed to compensate for the effects of laser plasma variation. An internal standard is a major element in the sample. The concentration of the internal standard is known and always the same. An emission line of the internal standard is selected for measuring the relative concentration of the other elements in the samples using the intensity ratio of the analyte line to the standard line. The internal standard line is chosen so that the excitation energies and partition functions of the analyte and internal standard lines are similarly matched, so as to compensate for changes in the plasma. To minimize the effects of the spectral response of the detection system at different spectral regions, the analyte lines and the standard line should be chosen from the same spectral region. The matrix effect, which causes the sample to be ablated differently from the target sample, cannot be corrected with the internal standard. The requirements to perform a matrix-independent measurement by using internal standard are: complete atomization of particles and droplets in an LTE microplasma; and the intensity ratios of line pairs of different elements with approximately the same excitation energies must be dependent only on the concentration ratios in different matrices.

3 EXPERIMENTAL TECHNIQUES

3.1 Experimental Methods

Since the first report of a laser-induced spark with a ruby laser by Maker, Terhune, and Savage in 1963,[39] a variety of experimental methods have been developed to perform basic research and practical applications. The development of LIBS is closely related to the development of laser techniques. From the early ruby and CO_2 lasers to more recently and widely used Nd : YAG and excimer lasers, more powerful, reliable, and compact laser systems have been used to perform LIBS. LIBS development is also based on the development of detectors, fiber optics, and computer technology. New detectors such as the intensified charge coupled device (ICCD) have replaced the old photomultiplier tube (PMT) and photodiode array, and have achieved much faster time responses and higher S/N values. Recent computer technology advances have enabled LIBS to perform real-time operations, which have enhanced its capability as a method for process monitor and control.

In 1982, Sandia National Laboratories-Livermore reported a time-integrated LIBS system for combustion applications with a Q-switched CO_2 transversely excited atmospheric pressure (TEA) laser.[40] The laser generates a series of mode-locked pulses separated by 7.8 ns extending over about 1 μs with total energy about 1 J at a wavelength of 10.5 μm. The beam is expanded, then focused at the point of interest. The emission from the induced spark is collected in the direction perpendicular to the beam and coupled to a 1 m spectrometer with proper optics. An intensified diode array was used to record the LIBS spectrum.

Almost at the same time, L.J. Radziemski et al. introduced a time-resolved technique to record the LIBS signal.[37,41,42] As the emission characteristics of the induced spark vary with time, the time-resolved measurement can temporally separate the emissions from different species in the induced plasma. This technique has been successfully used to maximize the S/B and to minimize the spectral interferences between atoms of different ionization stages as well as between atoms and molecules. In the time-resolved detection, the detector was only exposed to light during a specified gate time which is controlled by a time gate pulser. The gate pulse signal was delayed with respect to the laser pulse with a width of 200 ns to several μs. It provides a high voltage to the detector intensifier to only record the signal during the time of the gate. Figure 4 is a diagram of their system for time-resolved LIBS measurement.[37] An Nd : YAG laser operated at 1.064 μm (pulse width 10 ns and pulse energy 100–300 mJ) was used to generate the laser spark. The array detector, consisting of 1024 diodes in 25.4 mm, was controlled by a multichannel analyzer to record LIBS spectra. The sample was delivered by a nebulizer/heat-chamber system into a six-armed crossed-sample chamber where the laser beam, observation direction, and the sample flow are perpendicular to each other. With a different gate time, the ionic and neutral lines of nitrogen were successfully observed. This time-resolved technique was later adopted by many researchers due to its ability to obtain a better S/B.

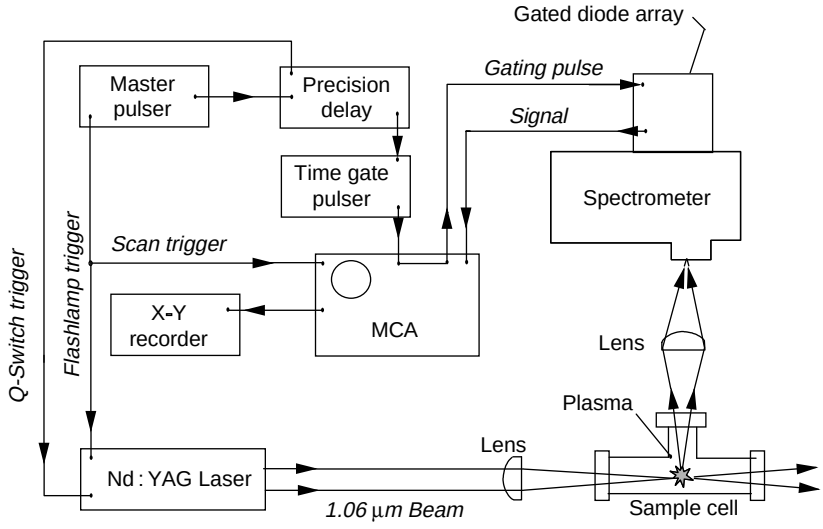

Figure 4 Diagram of the time gated diode array system for LIBS measurement. (Reprinted from L.J. Radziemski, T.R. Loree, D.A. Cremers, N.M. Hoffman,[37] Copyright (1983) American Chemical Society.)

In 1984, D.A. Cremers et al. reported two experimental methods for LIBS measurements of liquids.[32] The two methods are repetitive single spark (RSS) and repetitive spark pair (RSP). In RSS, an Nd : YAG laser operated at 1.064 μm was focused into the liquid by a pair of 5 cm focal length lenses. The beam was focused into the cell through a quartz window on the bottom side. The cell was constructed of Teflon® and its volume was 20 cm^3. A small, intense, reproducible spark was obtained when the beam was focused in the liquid very close to the input window (within 2–5 mm). A strong pressure generated by the spark was observed. The spark light was collected at right angles to the laser beam. Either a PMT with boxcar averager or a photodiode array with multichannel analyzer was used. The signal was temporally resolved and averaged over many shots. In RSP, two Nd : YAG lasers were employed to generate a pair of sparks separated in time by about 18 μs. The beams from the two lasers were spatially superimposed with a 50%T : 50%R beam splitter (R = reflection and T = transmission). The second spark is generated inside the gaseous cavity formed after the first spark. The signal intensity of some species was found to be enhanced significantly with the RSP method.

The RSP technique was applied to solid samples by Uebbing et al.[43] Two Nd : YAG lasers (both operated at 1.06 μm) were used. The first laser (output 13 mJ per pulse) was focused on the sample surface perpendicularly to vaporize the sample. The beam of the second laser (output 77 mJ per pulse) was directed parallel to the sample surface and 1.5 mm above the surface. It was focused 2 mm behind the middle of the first plasma. The samples were placed in a sample chamber filled with argon. There was a time delay between two laser shots. The detection system was a gatable optical multichannel analyzer. The intensity of the Zn line at 468.014 nm was observed to be more than 10 times stronger than the intensity with the vaporizing laser only.

Cremers and Radziemski carried out an experiment using the long-spark technique for direct detection of beryllium on the filter.[44] A long spark was generated by focusing the beam of a pulsed and Q-switched Nd : YAG laser on the surface with a cylindrical lens of 75 mm focal length. Laser energy of 125 mJ per pulse was used to ensure spark stability. The footprint of the spark on the surface was about 0.1 mm wide and 4–8 mm long, depending on the angle of incidence of the laser beam on the filter. The emission from the induced plasma was collected by another lens in the perpendicular direction and analyzed by a monochromator tuned to the wavelength of 313.042/313.107 nm which is the doublet of Be^+. The spectrally resolved light was detected by a PMT. The signal was time-resolved and averaged with a boxcar average.

Mason and Goldberg produced a laser spark in a pulsed magnetic field to characterize the effects of a magnetic field on plasma atomization, excitation, and ionization.[45–47] Both emission and absorption spectroscopy were performed. An eight-turn coil was wound on a 1 inch outside diameter polycarbonate tube and a high-voltage capacitive discharge source was used to generate the pulsed magnetic field. A pulsed hollow cathode lamp was used for the absorbance measurement. The hollow cathode lamp was triggered before the formation of the laser plasma and kept active for over 300 μs in order to output a stable emission. The light was collimated by a lens and, after passing through the coil,

was focused by another lens onto the entrance slit of the monochromator. The emission characteristics were found to have significant radial compression and axial expansion of the laser induced plasma.

In 1987 Cremers used a fused silica incoherent fiber bundle instead of a lens to collect the emission from the laser spark.[48] The cable was 1 m in length and 3 mm in diameter. One end of the cable was positioned next to the focusing lens (50 cm focal length) and the other end of the cable was positioned up against the entrance slit of a spectrograph. When the fiber optic cable was positioned 55 cm from the spark, the light collected by the fiber optic bundle was equivalent to that collected by a lens with an f number of 183. Because of the large acceptance angle of fiber optics, light from a wide angle range will be collected by the fiber and is not changed significantly at the output end. This fiber collection technique is suitable for field measurements because it is insensitive to the spark position changes. Schechter and Valery Bulatov have developed a fine technique using an optical fiber.[49] A multiple optic fiber system, with each fiber pointing at a different region of the spark, was used in their measurement. The light from a different fiber was fed into a computerized imaging spectrometer which was attached to an ICCD detector. The fluctuation in the spark was compensated for by observing different elements at a different region in the spark.

LIBS is a useful technique for in situ field measurements. However, most experimental methods introduced previously are laboratory systems and may not be well suited to field measurements. A field system usually requires minimal optical access to the test facility and minimal on-site alignment. In 1994, researchers from Sandia National Laboratory completed a field demonstration of their prototype system to monitor metals emissions at Clemson University's Vitrification Facility.[50] The main feature of the system was a mirror with a hole at the center to allow the laser beam to pass and to reflect the emission signal from the laser spark (Figure 5). The frequency-doubled Q-switched Nd : YAG laser beam (532 nm) was expanded and collimated to pass through the hole in the mirror. The beam was focused in the effluent stream and produced the spark. The same lens was also used to collect and collimate emission from the spark. The lens was mounted at the end of a probe inserted into the test facility through an optical port. The collected light was reflected by the mirror to turn 90° and forward to other optics coupled to the fiber optic bundle. The fiber optic bundle was connected to the spectrometer and detector. This configuration allowed excitation and detection to be performed collinearly and, therefore, only one optical port was used for access and measurement.

Another approach to backwards LIBS detection was developed at the Diagnostic Instrumentation and Analysis Laboratory of Mississippi State University (Figure 6).[51] The frequency-doubled Q-switched Nd–YAG laser beam was reflected at a harmonic separator to remove its fundamental beam. The 532 nm beam was then reflected to the probe lens through a dichroic mirror that reflects 532 nm and transmits other wavelengths. The beam was focused in the off-gas stream to produce the breakdown. The same lens was also used to collect and collimate emission from the plasma. The LIBS signal was transmitted through the dichroic mirror and coupled to the fiber optic bundle with lenses. The fiber optic bundle was connected to the spectrometer and detector.

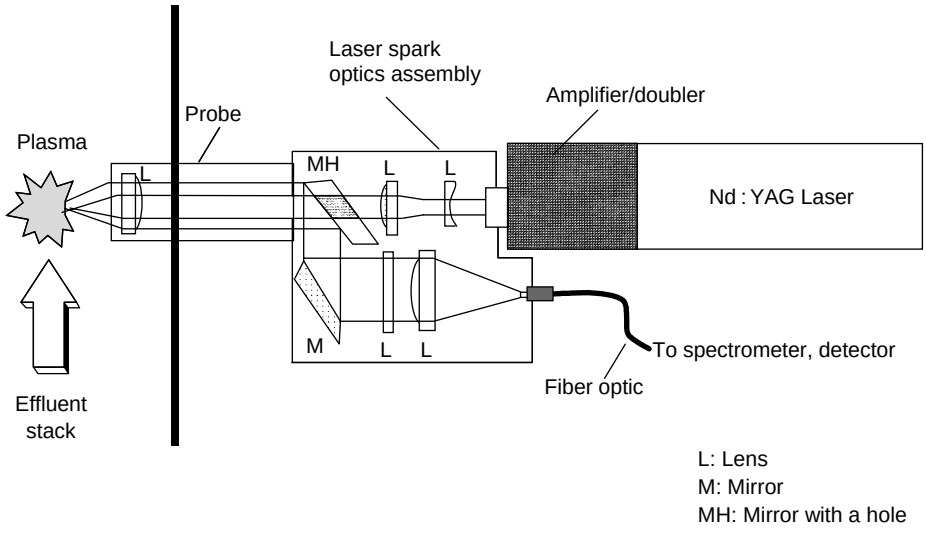

Figure 5 Diagram of a LIBS system using a mirror with a hole for backward detection. (Reproduced from L.W. Peng et al. *Process Control and Quality*, **7**, 39–49 (1995), by permission of VSP International Science Publishers.)

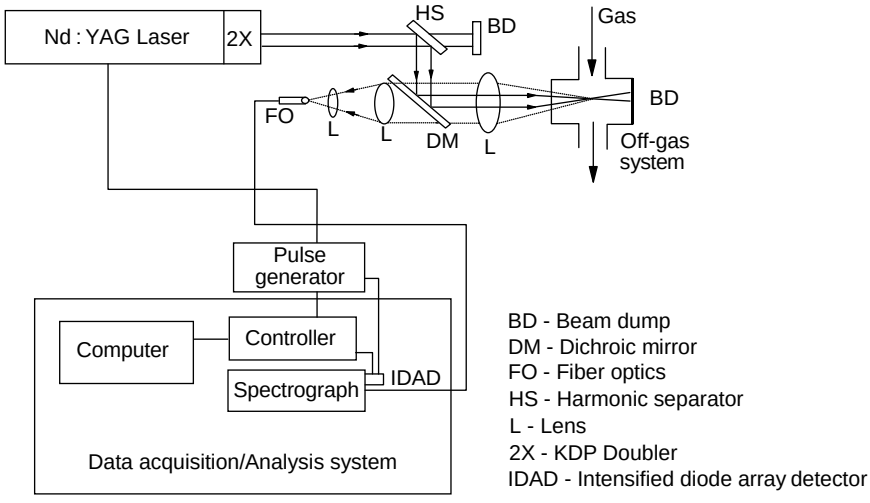

Figure 6 Diagram of a LIBS system using a dichroic mirror for backward detection.

The rectangular output of the fiber bundle worked as an entrance slit for the spectrometer. This optical configuration, which requires only one port, also allows collinear excitation and detection. It is suitable for real-time measurement in off-gases.

Due to the low breakdown threshold of the optical fiber, optical fibers in early LIBS work were limited to delivering the emission to the detection system. It is now possible to use optical fibers to deliver the laser beam in LIBS applications. Marquardt et al. have recently developed a microprobe for LIBS and Raman measurements using an imaging optical fiber.[52] The microprobe has five separate optical fibers. One 1 mm core diameter silica–silica-clad multimode fiber was used to deliver the Q-switched Nd : YAG laser beam (1.064 μm) for LIBS excitation. A 12 mm focal length lens near the fiber collimated the exited laser beam and a 6.4 mm focal length lens focused the laser beam on the sample. These two lenses were also used to collect spark light and couple the light to another 1 mm core diameter hard-clad fiber which coupled to the spectrograph at another end. The Nd : YAG laser beam delivery fiber was also used to deliver a He–Ne beam to indicate the region of interest on the sample. One image-guide fiber was coupled to a video camera to provide video monitoring while the probe was positioned and aligned for analyzing specific sites on the sample (granite). An ICCD detector was used and operated in gated mode.

3.2 Calibration Methods

For practical applications, the measured emission intensities need to be related to the relative or absolute elemental concentration. The system response must be calibrated for a certain measurement. Calibration is the most difficult issue in the development of LIBS, especially for field measurements. LIBS is an atomic emission spectroscopy. In addition to variables related to emission spectra, several other variables greatly affect the intensity of the LIBS signal: the fluctuation of incident laser energy; the size and density of particles and associated matrix effects in the gas flow; the location of the focus point for solid and liquid samples; and the surface feature and the history of ablation by laser shots for solid samples. For on-site measurement, LIBS is considered as a nonsampling technique. However, this implies an extra difficulty for calibration.

For solid sampling, a standard sample can usually be pressed from well-mixed powders. Ernst et al. determined the copper concentration in steel using LIBS.[53] They prepared cylindrical calibration samples using a hot isostatic press process. The flat surfaces of the sample were polished with a soft polishing wheel and cleaned. Eleven samples containing from 0.01% to 5.00% copper were used to obtain a quadratic calibration curve. To reduce the effect of some variables, the intensity ratio of Cu/Fe was used, instead of Cu intensity, to infer the calibration curve against the Cu concentration. For some solid samples, instead of pressing, well-mixed powders can be placed in a furnace to be melted and then frozen.

For gas-phase sampling, a nebulizer is widely used to produce a dry aerosol from a solution of standard reference materials. During a United States Environmental Protection Agency (USEPA) metal CEM test in 1996, a pneumatic nebulizer was used to produce an aerosol flow containing certain concentrations of metal.[54] The nebulizer was enclosed in a separate cell and was inserted into the optical path when the calibration was required.

During the same CEM test, an on-site calibration was performed at the Rotary Kiln Incinerator Simulator.[55,56] An ultrasonic nebulizer (USN) was used to generate a dry metal aerosol from the standard solution which was pumped into the USN with a peristaltic pump. The aerosol was brought to the gas stream through a sample injection probe. The probe tube was mounted on the opposing port against the optical probe across the gas stream. The laser beam was focused 2 cm above the end of the injection probe.

In some cases, if the absolute concentration calibration is too difficult to obtain, the relative concentration may be considered. According to Equation (2), relative concentration can be inferred by fit or by intensity comparisons of two or more different elements in one LIBS spectrum.[57] In addition to the spectroscopic constants, the plasma temperature and degree of ionization are required to perform this calculation. However, these two parameters are not easily determined accurately. An alternative approach to obtaining the relative concentration was the use of reference line intensity information from the National Institute of Standards and Technology (NIST) collections to perform spectral fitting.[58] However, the excitation condition need to be checked for this simple method because the intensities in these reference collections are obtained under conditions which differ considerably from induced plasma.

The techniques for LIBS calibration, especially for field calibration, lags behind the development of LIBS itself. Scientists are still working to solve this problem.

4 ENVIRONMENTAL APPLICATIONS

LIBS has the capability to perform rapid, on-site analysis in harsh environments. It can significantly reduce the time and costs associated with sample preparation required by convention analytical techniques and is therefore a promising technique for environmental monitoring and process control.[59] LIBS has been applied to many environmental situations. The major interests here are applications in environmental monitoring and in industrial process control. Other application areas, e.g. clinical, metals, etc., are discussed elsewhere.[60]

4.1 Off-gas Emission

The detection of trace metals in the off-gas from waste processing is very important for public health. Conventional analytical techniques require sending a sample to a laboratory for analysis. LIBS can perform in situ off-gas measurement by focusing the laser beam on the gas stream through a window and collecting the signal through an optical fiber. Radziemski and Cremers applied LIBS to the analysis of effluent gases from a prototype fixed-bed coal-gasifier at the Department of Energy Morgantown Energy Technology Center.[31] They demonstrated that LIBS had the capability for near real-time monitoring of the concentrations of major and minor species in the off-gas emission. Neuhauser et al. tested their on-line Pb aerosol detection system with aerosol diameters between 10 and 800 nm. A detection limit of $155\,\mu g\,m^{-3}$ was found.[61] Singh et al. have demonstrated LIBS as a process monitor and control tool for waste remediation.[62] They monitored the toxic metals from three plasma-torch test facilities and proved that LIBS can be integrated with a torch-control system to minimize toxic metal emission during plasma-torch waste remediation.

To use LIBS as a CEM requires the quantitative trace-level determination of toxic metals. The challenges here are to overcome the problems of this technique related to poor detection sensitivity for some elements, limited accuracy and precision, and difficulty of calibration for in situ analysis. With a sensitive ICCD detector and properly selected experimental set-up, the sensitivity and precision has been greatly improved. However, on-line calibration is still a big problem for CEM-type applications due to difficulties of introducing known amount samples into the off-gas system.

Singh et al. investigated the possibility of using metal hydrides to calibrate metals in off-gas emissions.[63] They used a static sample cell to perform LIBS experiments on metal hydrides and found the LIBS signal to be affected by gas composition, gas pressure, and laser intensity. They also found that the LIBS signal intensity for metal hydrides changes with time. The time at which the signal reaches its peak also depends on the concentration of the sample. Therefore the selection of different time periods for the LIBS measurement will be critical for LIBS calibration. They suggested that a flowing metal hydride sample cell system might solve some of the problems related to the static system they used.

Detection of Hg in off-gas emission has been reported by several researchers. Lazzari et al. performed LIBS to detect Hg in air in a laboratory.[64] The Hg vapor was successively diluted in a chamber, and a Nd : glass laser (1064 nm, 180 mJ) was used to produce plasma in the Hg vapor. They reported a very low detection limit of $42\,\mu g\,m^{-3}$ for Hg using the Hg 253.7 line. Singh et al. used an USN to generate a known concentration of metal aerosol in a sample cell and found the minimum detectable Hg level to be approximately one order higher than that reported by Lazzari et al.[64] The reported Hg detection limit with a pneumatic nebulizer is consistent with the result with an USN.[65] The reason for a higher Hg detection limit with a nebulizer system is unclear. One possible explanation is that the aerosol-induced

breakdown might cause incomplete atomization of larger particles in the colder region of the plasma and this results in a higher Hg detection limit.

The laboratory-tested calibration technique is successful for simulated effluent in a sample cell using an aerosol generator. However, it cannot be directly applied to samples from the practical environment due to difficulties in reproducing the plasma conditions under quite different sampling conditions (i.e. particle size, gas flow, temperature, pressure, etc.). As a result of the difficulties in injecting a known amount of sample into a gas stream, researchers are now exploring the possibility of using an internal reference for self-calibration. The internal reference has to be a major species in the gas medium. The concentrations of the internal reference must be known to an accuracy below 10%. This calibration method is intended for compensating plasma condition changes due to the variation of gas stream conditions or laser pulse-to-pulse fluctuations. Therefore, the plasma temperature has to be inferred from the reference lines. The intensity ratio of the analyte line and the reference line are corrected based on the inferred plasma temperature. Loge has tested this calibration procedure for LIBS CEM using N_2^+ lines to determine plasma temperature and to monitor the changes in the plasma condition.[66] The observed emission line intensities were adjusted based on the inferred plasma temperature, and the concentration was extracted from the plasma temperature-corrected calibration curve. However, the plasma condition is not only characterized by plasma temperature. Electron density can also affect the degree of ionization of the probed atoms and hence the signal of the neutral or ion line. Furthermore, this calibration relys on the plasma temperature extracted from the intensity ratio of two emission lines, which is known to be an inaccurate method for plasma-temperature determination. This calibration might be suitable for LTE conditions in which and electron density and plasma temperature are not significantly different from the calibration condition. However, the success of this internal calibration method will greatly depend on the plasma characteristics and matrix effect.

LIBS has low sensitivity for some elements in off-gas measurements because of its small sample volume. Hahn et al. developed a technique based on random sampling and conditional analysis to solve the problem associated with finite sample volume and size distribution of the target metal.[67] They compared Monte Carlo simulation with the experimental results and claimed considerable enhancement with this technique for low metal concentrations where only a few single-shot spectra contained the desired metal emission. This calibration technique seems promising; however, more tests are required to prove its validity for different sampling conditions.

LIBS has been tested as a multimetal CEM. Using concentrations as specified by the Resource Conservation and Recovery Act (RCRA), metals were measured and the results compared with the USEPA's reference method.[65] A relative precision of 50% and a relative accuracy (RA) of 78% or better were obtained from these tests. The rapid sampling rate and potential for metal CEM are demonstrated in this test. However, the sensitivity and accuracy for certain toxic metals will need to be further improved before LIBS can be accepted as a metal CEM by the USEPA.

4.2 Metal and Glass

Kurniawan et al. have applied LIBS to the elemental analysis of glass.[68] They found the detection limits are much lower than those usually required for glass with low surrounding pressure (1–5 torr). In addition, light elements (such as Li and B) that are usually difficult to observe by X-ray fluorescence were also successfully detected with a detection limit less than 10 ppm. Thiem et al. performed LIBS measurements on alloys in an ultrahigh vacuum ($<5 \times 10^{-7}$ torr).[69] They found the background signal is significantly reduced in comparison to that from an atmospheric environment. This makes a simultaneous, multielement analysis in very complex solid samples possible. They obtained calibrations for Al, Fe, Cu, Ni, and Zn using a standard reference material and found the detection limit to be matrix dependent. Absolute detection limits were estimated to be in the 20–200 $\mu g\, g^{-1}$ range for the above elements. Sabsabi and Cielo made quantitative analyses of aluminum alloys with LIBS.[70] To find the optimum condition for LIBS analysis, they first studied laser plasma characterization by extracting plasma temperature and electron density with Fe and Al^+ lines, respectively. They found plasma temperatures are close to a constant within experimental uncertainty at 10 μs delay and 10 μs gate width with different concentrations of Mg in aluminum alloys. Calibration curves for Mg, Mn, Cu, and Si were constructed under this optimum experimental condition. The Mg calibration based on the 285.2 nm resonance line shows problems of self-absorption at higher concentrations, whereas the calibration based on nonresonance 518.36 nm and 517.27 nm Mg lines is linear for the same concentration range. The precision and detection limits for LIBS analysis of aluminum alloy were calculated based on the calibration data. They found a relative standard deviation (RSD) of better than 6% and detection limits better than 15 ppm for all the elements studied.

The current process for molten steel analysis includes collecting a molten sample, cooling it, and then transporting it to a nearby analytical laboratory for analysis. It is a time- and labor-intensive process. The hot melt composition might change due to vaporization of the more

volatile components during this period. Also, compositional fluctuations cannot be effectively monitored with the current analysis method. LIBS has the capability of performing on-line measurements, which can reduce the data processing time and reduce the cost for the steel industry.

Argagon et al. have demonstrated the application of LIBS to molten steel.[71] To measure the carbon content they used an argon atmosphere to reduce carbon emission from dissociated CO or CO_2. As the composition of the molten surface may not represent the bulk melt, an Ar jet was applied on the surface of molten steel to measure the bulk concentration during the measurement. A C/Fe detection limit of 250 ppm with a RSD of 6% was found. Lorenzen et al. developed an on-line monitor for molten steel.[59] Elemental concentrations were determined based on calibration of the intensity ratio of an analyte line and a reference line versus elemental concentration. To ensure measurement accuracy, they also monitored the plasma condition using the intensity ratio of an atomic/ionic line pair. When a change in the plasma condition was detected, a proper correction factor was applied to the calibration curve. Paksy et al. also performed LIBS on the surface of a molten alloy using an Nd : YAG laser.[72] They compared the measurement with argon and air atmospheres and found that the Ar atmosphere prevents surface oxidation and gave a higher *S*/*B* with good precision for a relative intensity measurement. Detection limits for Si/Fe, Mn/Fe, Cr/Fe, Si/Al, and Mn/Al were found to be 10 ppm, 50 ppm, 40 ppm, 600 ppm, and 70 ppm respectively.

The application of LIBS to solid/molten materials is very promising. However, problems such as matrix effects (selective vaporization), sensitivity for all the element of interest, calibration, precision, and accuracy still need to be solved for quantitative measurement. Although it cannot provide the high accuracy and precision of laboratory-based methods, it has great potential as a process control monitor in the steel and glass industries by providing rapid, semiquantitative composition measurements in a remote location.

4.3 Soil, Concrete, and Paint

The detection of contaminated soil and concrete is another challenging area for LIBS. Yamamoto et al.[73] have used a portable LIBS system to detect toxic metals in soil, finding Ba, Be, Pb, and Sr in soil with detection limits of 265, 9.3, 298, and 42 ppm, respectively. Cremers et al. have detected Ba and Cr in soil using an optical fiber probe.[74] Limits of detection (LODs) of 26 ppm and 50 ppm were found for Ba and Cr, respectively. Eppler et al. have studied the matrix effect in soil. They found LODs for Pb and Ba in a sand matrix were 17 and 76 ppm (by weight), respectively, with a precision of 7% RSD or less.[75] In soil, the detection limits were 112 and 63 ppm (w/w) for Pb and Ba, respectively, with 10% RSD precision. The LIBS signal was affected by chemical speciation as well as matrix composition. LIBS accuracy can be degraded if calibrations are not compound and matrix specific. The accuracy of LIBS measurement in soil is affected by many factors. It is presently more suitable as an initial screening of highly contaminated soils to determine the remediation boundary of the contaminated area.

LIBS for detection of lead in contaminated concrete was explored by Pakhomov et al.[76] They used a Q-switched Nd : YAG laser (10 Hz) to perform time-resolved LIBS for quantitative measurement of the lead content in concrete. Pb calibrations were obtained for different delay times using the ratio of the integrated emission line of lead (405.78 nm) to an oxygen line (407.59 nm). They found the absolute Pb signal is independent of the laser pulse energy for laser energies between 250 and 400 mJ. Based on analysis of the sensitivity of lead measurements at different delay times, they derived a best Pb detection limit of 10 ppm in concrete at an optimum delay time of 3.0 μs.

Lead in paint is known to be a potential health threat, especially to children. The feasibility of using LIBS to determine Pb in the paint surface has been successfully demonstrated by Yamamoto et al. using a portable LIBS system.[73] The temporally integrated LIBS signal is recorded with an nongateable charge-coupled device (CCD) detector. Because of the problem of spectral interference associated with the Pb 405.78 nm line, they used a weaker Pb^+ line at 220.35 nm for calibration. They obtained a Pb detection limit of 0.8% (8000 ppm) in paint with an RSD between 23% and 47%. Marquardt et al. used a fiber optics LIBS system to determine Pb in paint.[52] They used a gate delay of 3.0 μs and gate width of 12.6 μs in the measurements. They reported a LOD of 140 ppm for Pb with a precision of 5–10% using the Pb 405.78 nm line.

4.4 Air Sampling Filters

Direct detection of trace element in air with LIBS is very difficult due to its insufficient sensitivity. For trace metals below the LIBS detection limits, an air sample can be collected on a filter which is then subsequently analyzed by LIBS. This method has a lower detection limit and provides a quasi-on-line measurement. Arnold and Cremers have used this technique to determine metal particles on an air damping filter.[77] They used a cylindrical lens to form a long spark on the filter to increase the sample volume and reduce the filter damage. Using the calibration curve for Tl line at 535.05 nm, a LOD of 40 ng cm^{-2} for Tl in filter paper was obtained. Later, Yamamoto et al.

determined LODs of 21 ng cm^{-2} and 5.6 μg cm^{-2} for Be and Pb on a filter.[73] They also noticed that particle size can affect the detection limit for filter analysis. Cremers and Radziemski have measured Be on a filter surface and found that Be particles greater than 10 μm on the filter were not completely vaporized. The particle size dependence of the LIBS signal puts a restriction on LIBS application to air sampling through a filter.[44]

4.5 Radioactive Materials

LIBS has been used to monitor the level of radioactive elements in a process stream. Wachter and Cremers found a detection limit of 100 ppm for uranium in solution.[34] To overcome poor shot-to-shot precision due to small variations in the lens-to-sample distance, they had to average 1600 laser shots for analysis. To investigate the feasibility of applying LIBS to detect radioactive elements, Singh et al. recorded LIBS spectra of U, Pu, and Np in a glove box.[78] Emission lines suitable for detection of these radioactive elements were identified. Their preliminary study showed that LIBS is suitable for the measurement of radioactive elements in waste streams. Ernst et al. used LIBS as a tool for detection of radiation embrittlement in a nuclear power plant by determining the Cu concentration in A533b steel.[79] As Cu is the key impurity contributing to radiation embrittlement, the Cu concentration in the steel is an indicator of radiation embrittlement and of expected material lifetime. Due to the hazardous environment, a fiber optic delivery system was initially used in the measurement. However, the optical fiber system could not deliver sufficient laser energy for sensitive Cu detection. Copper concentrations between 100 ppm and 5% were later detected using a conventional all-optic beam delivery system. Their measurements demonstrated that LIBS can provide quick, accurate, and remote detection for monitoring the reactor pressure vessel material.

5 COMPARISON WITH OTHER SPECTROSCOPY TECHNIQUES

The LIBS technique has several advantages compared to conventional emission methods. In LIBS, a high-energy pulsed laser beam is used to produce atomic emissions from the focal volume. Hence it provides time and spatially resolved measurement. It requires only a small amount of sample and minimal sample preparation. Gated detection with an intensified detection system discriminates the background emission and also improves the detection limit. With properly selected atomic lines, multiple species can be simultaneously monitored in a spectral region. Furthermore, as a laser spark can be generated in a remote location, it has remote monitoring capability. In this section the accuracy of LIBS is compared with some spectroscopic techniques commonly used in the laboratory, such as AAS, ICPAES, and XFS.

AAS is used to detect the presence and amount of metal atoms in very dilute solutions. It is widely used in laboratories that analyze the purity of water. The resonant wavelength of choice from a selected spectral lamp passes through the sample vapor. The atomized element absorbs the light energy over time. The concentration is determined through a calibration curve. The sample solution can be atomized by passing an electrical current through a graphite tube which contains the analyte (graphite furnace atomic absorption (AA)) or using a flame to atomize the sample (flame AA). The graphite furnace AA technique often offers about 1000-fold sensitivity improvement as compared with flame AA.

ICPAES measures the intensities of light emitted from each species as they are atomized and perhaps ionized in the plasma. As an AES (atomic emission spectrometry) technique, ICPAES has the capability of simultaneous multielement measurement. Atomization is more complete in the ICP (inductively coupled plasma) system than in the AAS system; therefore, LOD values are generally lower for ICPAES than for AAS. However, because the temperature of ICP is much greater than for AAS, spectral interference tends to be a greater problem for ICPAES than for AAS. The temperature cross-section of the plasma is relatively uniform; as a consequence, self-absorption and self-reversal effects are not encountered.

XFS is a spectroscopic method that is commonly used for solids in which secondary X-ray emission is generated by excitation of a sample with X-rays. The X-rays eject inner-shell electrons. Outer-shell electrons take their place and emit photons in the process. The photon wavelength depends on the energy difference between the outer-shell and inner-shell electron orbitals. The amount of fluorescence is very sample dependent and quantitative analysis requires calibration with standards with a similar matrix. XFS can provide rapid quantitative analysis for various samples. However, XFS is not applicable to light elements ($M < 10$) such as Be, B, and Li. As the XFS signal is generated from a significant thickness of sample within which absorption and scattering can occur. It is known to suffer significant disturbance from the matrix which is caused by either absorption from the matrix elements or enhancement by the emission of a matrix element.

5.1 Sample Preparation

Laser plasma is a direct sampling and excitation tool. Only small amounts of samples are used. The laser prepares

and excites the sample in one step. Therefore, minimal sample preparation is required compared to conventional techniques. It also avoids possible contamination by tedious and time-consuming sample preparation procedures. However, because a minimal amount of sample is ablated and analyzed, the precision and accuracy of measurements depends on the homogeneity of the sample.

Typical sample preparation in AAS involves dissolving a solid sample in an acidified solution. Aqueous dilution with the appropriate acid, ionization suppression agent(s), or matrix modifier(s) may be necessary. The diluted solution is aspirated into an atomizer (i.e. flame or furnace) to atomize the sample. Air samples are collected with a filter. The particles collected on the filter are dissolved in acid to produce a solution which is analyzed with conventional AA. Flame AA utilizes milliliters of sample. Graphite AA typically uses about 20 μL of sample. The sample preparation procedure for AAS is generally time-consuming which is not conducive to on-line and in situ measurement.

In ICPAES, the sample may be an aerosol, a thermally generated vapor, or a fine powder. The sample is carried into the hot plasma at the head of the tubes by argon flowing at about $1\,L\,min^{-1}$ through the central quartz tube of the ICP torch. Typically, the sample is reduced to a solution that is nebulized into the ICP system. An autosampler is frequently used to supply solutions (including blanks, standards, and sample) to the plasma sequentially. It is an attractive technique for the analysis of geological material due to its multielement capability. However, samples need to be dissolved prior to analysis. Various chemical schemes are necessary to dissolve as many elements into solution as possible. Thiem and Wolf have compared the analysis of Al ores from the NIST by ICPAES and LIBS.[80] In ICP measurements, the powder samples are first dried in an oven at 140 °C for 2 h. A specific weight of dried sample is then digested with a mixture of concentrated HCl and deionized water in a digestion flask at 90 °C overnight. The digested samples are then put into a volumetric flask for analysis. When using LIBS to analyze the same type of sample, a small amount of sample was compressed into a disk. The sample attached to a sample translation device is then placed into the sampling chamber. The LIBS measurement started after the sampling chamber was evacuated to about 1 mtorr. The required sample preparation time (including time required to pump down the sampling chamber) for LIBS analysis is less than one hour which is much shorter than that required for an ICP measurement.[80]

XFS can be used to analyze a variety of samples. Liquid samples, from water to oil, require essentially no preparation. Many solid samples, if homogeneous, can be machined to the proper shape and run directly. Inhomogeneous samples, such as ores and biological materials, must be pulverized or fused. Then the homogenized material can be dissolved for analysis as a liquid sample, pressed into pellet form, or cast as a glass-like solid. XFS can also be used for the analysis of atmosphere pollutants. The air sample is first collected through a stack consisting of micropore filters and is then analyzed by XFS directly. Although XFS has direct sampling capability, it suffers more severe matrix effects with direct sampling than does LIBS. Also, XFS does not have remote analysis capability, whereas LIBS can use optical fibers to perform remote measurements.

Comparing LIBS with ICP, AAS, and XFS, LIBS and XFS require minimum sample preparation. AA and ICP require destructive and relatively tedious and time-consuming pretreatment of the samples, which is likely to introduce uncontrollable error in the measurements.

5.2 Sampling Rate

LIBS is an optical emission based-technique. It can perform multielement measurements. Generally LIBS can produce useful results in 3–60 s. The sampling rate depends on the laser repetition rate and sample concentration. XFS spectra are relatively simple and the technique is nondestructive. The speed (1–10 min) and convenience of the procedure permits multielement analyses to be completed in a few minutes. ICP and AAS can have a sampling time of less than 1 min. However, sample preparation often takes several hours. It takes minutes to hours for complete sample analysis.

5.3 Detection Limit

The detection limit is the minimum concentration that can in principle be measured. It is generally defined as the concentration that produces a signal three times larger than the SD of the background. As multielement techniques, LIBS, ICPAES, and XFS all rely on being able to resolve spectral signals unique to each element. When the signals of some elements are very strong, they start to interfere with weaker signals from the other elements, and it may be difficult to achieve the optimum detection limits. LIBS has poorer sensitivity in gas measurements due to the low concentration and small laser microplasma. Typical LODs for LIBS are in the ppm to percentage range. For ICPAES, detection limits for most elements are of the order of 10 ppb.

XFS is generally less sensitive than the various optical methods. It offers LODs from a few ppm to the percentage range. In AAS, the amount of light that is absorbed by the sample is directly proportional to the concentration of the ground state element of interest in that aspirated sample. Some samples may have matrix, chemical, and/or ionization interferences. When recognized, each interference can usually be controlled to yield meaningful

sample analysis. Detection limits ranging from around 3×10^{-4} ppm to 20 ppm are observed for various metallic elements using flame AA. Graphite furnace atomic absorption spectrometry (GFAAS) often reduces this limit by a factor of 10–1000.

This direct comparison of LIBS detection limits with other analytical techniques is unrealistic because of the different samples used. In general, LIBS still cannot match the detection limits of most conventional analytical techniques in most cases. However, its ability to directly analyze samples without cumbersome sample preparation has led to its promising application in practical field determinations of elements for environment and industrial samples.

5.4 Accuracy and Precision

Precision and accuracy are generally used to determine the merit of a measurement technique. Precision is the measure of the degree of reproducibility of a measurement. It is generally measured as the RSD of a set of repeated measurements. Accuracy is used to determine the difference between the evaluated measurement method and a measurement by a standard method. It is generally expressed as the RA. The size and excitation characteristics of a LIP can affect the amount of sample vaporized and the degree of excitation, and hence affect the LIBS accuracy and precision. Typical LIBS precision is 5–20%. Laser shot-to-shot variation causes differences in the plasma properties and therefore affects the magnitude of the element signal and hence degrades the LIBS precision. To improve LIBS precision, spectra from several laser shots have to be averaged in order to reduce statistical error due to laser shot-to-shot fluctuation.

The accuracy of LIBS depends on the sample composition, homogeneity, surface condition, and particle size generated. The sample matrix can affect the amount of element ablated and hence the LIBS signal. As LIBS only analyzes a small amount of sample, the result might be biased if the composition of the ablated sample does not represent the composition of the bulk. Multipoints measurements can improve the LIBS accuracy for an inhomogeneous sample.

For LIBS measurements based on the theoretical analysis of emission line intensity ratios, the accuracy depends greatly on the available spectroscopic constants, inferred temperature, and electron density. The uncertainty in these parameters is generally larger than 25%. Therefore, this is not a very accurate analysis method for LIBS. The best estimated accuracy with this technique is about 10%.

If the LIP condition is reproducible, calibration by emission line intensity is a more promising analysis tool for LIBS. The accuracy of this technique relies on absolute intensity measurement and therefore is affected by laser shot-to-shot fluctuation and particle size. Also this type of calibration shows a loss of sensitivity at high analyte concentrations due to self-absorption in the laser plume.

The sensitivity of AAS is dependent on the element, but is usually in the ppm region for flame AA and ppb region for graphite AA. Sample concentration techniques enable detection at even lower levels. Its precision is a function of several preparative, elemental, and instrumental conditions. Typically, the RSD associated with a flame absorption analysis is of the order of 1–2%. With caution, this figure can be lowered to a few 10ths of 1%.

The measurement precision of ICPAES is typically 1–5%. Thiem and Wolf have analyzed mining ores with both LIBS and ICPAES.[80] The elements Al, Ca, Cu, Fe, K, Mg, Mn, Si, and Ti were determined in aluminum and manganese ore. In these measurements, LIBS obtained an RA of 0.5–20% and ICP an RA of 0.7–14%. Overall results show that ICP has better precision than LIBS (Table 2).

The potential of LIBS as a screening instrument in field work was recently evaluated, based on its accuracy and precision. Yamamoto et al. have compared the performance of a portable LIBS with a commercial portable XFS unit.[73] They used certified reference soil as the test sample to determine Sr concentration in soils. XFS measurements were obtained from a Spectrace™ 9000 portable XFS unit with 200 s integration time and LIBS was performed with 10 s integration time (10 laser sparks). The certified soils have Sr concentrations in the range 26–380 ppm. The test results are shown in Table 3. Sr concentrations were measured from all the samples with LIBS except the sample containing 26 ppm Sr, which is below its detection limit. The LIBS precision was found to be in the range 3–18% and the XFS precision was between 3% and 23%. The RA obtained from these measurements were 0–44% for LIBS and 9–44% for XFS. The portable LIBS therefore has comparable accuracy and precision to a portable XFS unit. However, the portable LIBS has slightly poorer sensitivity for Sr than the portable XFS unit. This study indicates that LIBS has the same capability as an in situ screening device for soil as XFS.

Table 2 Comparison of LIBS and ICP measurement of NIST Al ore samples[80]

Element	NIST certified value	ICP	LIBS
Al	25.52 ± 0.40	25.34 ± 0.39	24.3 ± 1.2
Ca	0.443 ± 0.020	0.404 ± 0.02	0.427 ± 0.010
Fe	13.71 ± 0.20	13.44 ± 0.1	13.63 ± 0.75
Mg	0.035 ± 0.008	0.038 ± 0.01	0.028 ± 0.008
Mn	0.294 ± 0.030	0.252 ± 0.01	0.25 ± 0.05
Si	0.323 ± 0.03	0.305 ± 0.05	0.275 ± 0.03
Ti	1.43 ± 0.07	1.52 ± 0.10	1.32 ± 0.15

Table 3 Comparisons of determination of Sr in soil using LIBS and the ICP.[73]

Certified soil	Actual concentration (ppm)	LIBS (ppm)	LIBS %RSD	XFS (ppm)	XFS %RDS
07401	155	87	5.3	200	5.9
07402	187	187	5.1	221	4.9
07403	380	295	4	434	3.2
07405	42	30	17.5	48	13.6
07406	39	28	12	56	14.4
07407	26	[a]	[a]	35	22.9
07408	236	204	3.2	257	4.5

[a] Sr signal was not detectable.

As a result of the small sample volume used the analytical capabilities of LIBS are generally poorer than most conventional methods. However, various studies have shown that LIBS is best for applications where standard laboratory instrumentation is not feasible, such as measurements in remote or harsh environments.

6 FUTURE DEVELOPMENTS

The development of LIBS has progressed rapidly since that late 1980s, with work on improving LIBS measurement of gas, liquid, and solid samples. The following research areas need to be strengthened before the technology reaches maturity for commercialization.

LIBS application as a CEM needs work on improving calibration methods, sensitivity and accuracy. An on-line calibration method needs to be developed for off-gas emission measurements. Sensitivity can be improved either by increasing the signal or reducing the background noise from the plasma and detector. Accuracy of the measurement can be improved by reducing pulse-to-pulse fluctuations of the laser-induced spark. Some toxic metals have analytical lines shorter than 200 nm, whereas optical fiber capability and detector sensitivity decrease considerably in this spectral region. Further work is therefore needed to improve measurement in this spectral region.

Further work is also required for the application of LIBS to the analysis of liquid samples. The main research areas of interest are studies of *S*/*N* for the laser spark produced with different liquid surfaces and bulk liquids, and techniques to measure concentrations at various depths. Quantitative measurement of solid samples with LIBS needs additional research to improve its analytical capabilities by improving the pulse-to-pulse spark reproducibility and laser ablation. Hybrid techniques such as combining LIBS with other discharge excitation sources need to be further tested to improve LIBS analytical capabilities. Matrix effects on concentration measurements also need further investigation. LIBS has great potential for molten metal/glass applications. However, more detailed studies of LIBS on molten materials are needed.

To develop LIBS as a mature environmental analytical technique, work to commercialize a field LIBS instrumentation is also needed. Instrument size and cost should be reduced. A spectrometer that can measure simultaneously all the elements of interest needs to be developed.

ACKNOWLEDGMENTS

The authors are thankful to Dr D.L. Monts and Ms L. Wade for a critical review of the manuscript. This work was supported by DOE contract DE-FG02-93CH-10575.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
CCD	Charge-coupled Device
CEM	Continuous Emission Monitor
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
ICCD	Intensified Charge Coupled Device
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
LIBS	Laser-induced Breakdown Spectroscopy
LIP	Laser-induced Plasma
LOD	Limit of Detection
LTE	Local Thermal Equilibrium
LTSD	Lens-to-Surface Distance
NIST	National Institute of Standards and Technology
PMT	Photomultiplier Tube
RA	Relative Accuracy
RCRA	Resource Conservation and Recovery Act
RSD	Relative Standard Deviation
RSP	Repetitive Spark Pair
RSS	Repetitive Single Spark
S/*B*	Signal-to-background Ratio
S/*N*	Signal-to-noise Ratio
TEA	Transversely Excited Atmospheric Pressure
USEPA	United States Environmental Protection Agency
USN	Ultrasonic Nebulizer
UV	Ultraviolet
XFS	X-ray Fluorescence Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Automotive Emissions Analysis with Spectroscopic Techniques

Environment: Water and Waste **(Volume 3)**
Heavy Metals Analysis in Seawater and Brines

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Laser Ablation in Atomic Spectroscopy • Laser Spectrometric Techniques in Analytical Atomic Spectrometry • Laser-induced Breakdown Spectroscopy • Microwave-induced Plasma Systems in Atomic Spectroscopy

REFERENCES

1. G. Bekefi (ed.), *Principles of Laser Plasma*, John Wiley & Sons, New York, 1976.
2. L.J. Radziemski, D.A. Cremers (eds.), *Laser Induced Plasma and Application*, Marcel Dekker, New York, 1989.
3. L.J. Radziemski, R.W. Solarz, J.A. Paisner (eds.), *Laser Spectroscopy and Its Applications*, Marcel Dekker, New York and Basel, 1987.
4. R.S. Adrain, J. Watson, 'Laser Microspectral Analysis: A Review of Principles and Application', *J. Phys. D: Appl. Phys.*, **17**, 1915–1940 (1984).
5. V. Majidi, M.R. Joseph, 'Spectroscopic Applications of Laser-induced Plasma', *Crit Rev. Anal. Chem.*, **23**(3), 143–162 (1992).
6. D.C. Smith, R.G. Meyerand, Jr, 'Laser Radiation Induced Gas Breakdown', in *Principles of Laser Plasma*, ed. G. Bekefi, John Wiley & Sons, New York, 457–507, 1976.
7. S.I. Anisimov, V.A. Khokhlov, *Instabilities in Laser-Matter Interaction*, CRC Press, Boca Raton, 1995.
8. R.G. Tomlinson, E.K. Damon, H.T. Buscher, 'The Breakdown of Noble and Atmospheric Gases by Ruby and Neodymium Laser Pulses', in *Physics of Quantum Electronics*, eds. P.L. Kelley, B. Lax, P.E. Tannenwals, McGraw-Hill, New York, 1966.
9. V.N. Pozhidaev, A.I. Fatievskii, 'Optical Breakdown Thresholds in Liquid Water and Micron Size Water Droplets Exposed to Single Laser Pulses', *Sov. J. Quantum Electron.*, **11**, 65–68 (1981).
10. J.F. Roach, J.M. Davies, 'Electric Strength of Some Liquid Dielectrics Subjected to a Q-switched Laser Pulse', *Proc. IEEE*, 1388–1390 (1970).
11. B.G. Gorshkov, Y.K. Danileiko, A.S. Lobachev, V.A. Lobachev, A.A. Manenkov, A.V. Sidorin, 'Laser Induced Breakdown in Alkali Halides', *Sov. Phys. JETP*, **45**, 612–618 (1977).
12. G.M. Weyl, 'Physics of Laser-induced Breakdown: An Updata', in *Laser Induced Plasma & Application*, eds. L.J. Radziemski, D.A. Cremers, Marcel Dekker, New York, 1989.
13. C.A. Bye, A. Scheeline, 'Saha–Boltzmann Statistics for Determination of Electron Temperature and Density in Spark Discharge Using an Echelle/CCD System', *Appl. Spectrosc.*, **47**, 2022–2039 (1993).
14. J.A. Woodroffe, J. Hsia, A. Ballantyne, 'Thermal and Impulse Coupling to an Aluminum Surface by a Pulsed KrF laser', *Appl. Phys. Lett.*, **36**, 14–15 (1980).
15. H.T. Buscher, R.G. Tomlinson, E.K. Damon, 'Frequency Dependence of Optically Induced Gas Breakdown', *Phys. Rev. Lett.*, **15**, 847–849 (1965).
16. R.A. Bingham, P.L. Salter, 'Analysis of Solid Materials by Laser Probe Mass Spectrometry', *Anal. Chem.*, **48**, 1735–1740 (1976).
17. F. Dahmain, 'Ablation Scaling in Laser-produced Plasma with Laser Intensity, Laser Wavelength, and Target Atomic Number', *Phys. Fluids B*, **4**, 1585–1588 (1992).
18. G.H. Canavan, W.A. Proctor, P.E. Nielsen, S.D. Rockwood, 'CO_2 Laser Air Breakdown Calculation', *IEEE J. Quantum Electron.*, **8**, 564 (1972).
19. Y.P. Raizer, *Laser-induced Discharge Phenomena*, Consultants Bureau, New York, 63–102, 1977.
20. B. Wolff-Rottke, J. Ihlemann, H. Schmidt, A. Scholl, 'Influence of the Laser-spot Diameter on Photo-ablation Rates', *Appl. Phys. A*, **60**, 13–17 (1995).
21. R.A. Multari, L.E. Foster, D.A. Cremers, M.J. Ferris, 'Effect of Sampling Geometry on Elemental Emissions in Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**(12), 1483–1499 (1996).
22. M. Kuzuya, H. Matsumoto, H. Takechi, O. Mikami, 'Effect of Laser Energy and Atmosphere on the Emission Characteristics of Laser-induced Plasma', *Appl. Spectrosc.*, **47**, 1659–1664 (1993).
23. Y.I. Lee, T.L. Thiem, G.-Y. Kim, Y.Y. Teng, J. Sneddon, 'Interaction of a Laser Beam With Metals. Part III: The Effect of a Controlled Atmosphere in Laser-ablated Plasma Emission', *Appl. Spectrosc.*, **46**, 1597–1604 (1992).

24. K. Kagawa, S. Yokoi, 'Application of the N_2 Laser to Laser Microprobe Spectrochemical Analysis', *Spectrochim. Acta, B*, **37**, 789–795 (1982).
25. D.C. McDonald, 'Temperature, Electron Densities and Degree of Ionization in a Boosted Glow Discharge Emission Lamp', *Spectrochim. Acta, B*, **37**(9), 747–758 (1982).
26. W. Sdorra, J. Brust, K. Niemax, 'Basic Investigation for Laser Microanalysis IV. The Dependence on the Laser Wavelength in Laser Ablation', *Mikrochim. Acta*, **108**, 1–10 (1992).
27. T. Ishizuka, 'Laser Emission Spectrography of Rare Earth Elements', *Anal. Chem.*, **45**, 538 (1973).
28. Z.W. Hwang, Y.Y. Teng, K.P. Li, J. Sneddon, 'Interaction of a Laser Beam with Metals. Part I: Quantitative Studies of Plasma Emission', *Appl. Spectrosc.*, **45**(3), 435–441 (1991).
29. Y.I. Lee, S.P. Sawan, T.L. Thiem, Y.Y. Teng, J. Sneddon, 'Interaction of a Laser Beam with Metals. Part II: Space-resolved Studies of Laser-ablated Plasma Emission', *Appl. Spectrosc.*, **46**(3), 436–441 (1992).
30. B.C. Castle, K. Talabardon, B.W. Smith, J.D. Winefordner, 'Variable Influencing the Precision of Laser-induced Breakdown Spectroscopy Measurement', *Appl. Spectrosc.*, **52**(5), 649–657 (1998).
31. L.J. Radziemski, D.A. Cremers (eds.), *Laser-induced plasma and Applications*, Marcel Dekker, New York, 1989.
32. D.A. Cremers, L.J. Radziemski, T.R. Loree, 'Spectrochemical Analysis of Liquids Using the Laser Spark', *Appl. Spectrosc.*, **38**(5), 721–729 (1984).
33. C.W. Ng, W.F. Ho, N.H. Cheung, 'Spectrochemical Analysis of Liquids Using Laser-induced Plasma Emissions: Effect of Laser Wavelength on Plasma Properties', *Appl. Spectrosc.*, **51**(7), 976–983 (1997).
34. J.R. Wachter, D.A. Cremers, 'Determination of Uranium in Solution Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **41**(6), 1042–1048 (1987).
35. D.C. Smith, R.T. Brown, 'Aerosol-induced air Breakdown with CO_2 Laser Radiation', *J. Appl. Phys.*, **46**, 1146–1154 (1975).
36. D.A. Cremers, L.J. Radziemski, 'Detection of Chlorine and Fluorine in Air by Laser-induced Breakdown Spectroscopy', *Anal. Chem.*, **55**, 1252–1256 (1983).
37. L.J. Radziemski, T.R. Loree, D.A. Cremers, N.M. Hoffman, 'Time-resolved Laser-induced Breakdown Spectrometry of Aerosols', *Anal. Chem.*, **55**, 1246–1252 (1983).
38. J.A. Aguilera, C. Aragon, J. Campos, 'Determination of Carbon Content in Steel Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **46**(9), 1382–1387 (1992).
39. P.D. Maker, R.W. Terhune, C.M. Savage, *Quantum Electronics (Proc. of Third International Congress, Paris, 1963)*, eds. P. Grivet, N. Bloembergen, Dunod, Paris, and Columbia University Press, New York, 1559, Vol. 2, 1964.
40. R.W. Schmieder, 'Combustion Applications of Laser Induced Breakdown Spectroscopy', Sandia Report SAND81-8886, Sandia National Laboratories, Livermore, CA, 1982.
41. L.J. Radziemski, T.R. Loree, 'Laser-induced Breakdown Spectroscopy: Time Resolved Behavior', LANL Report LA-UR-81-697, Los Alamos National Laboratory, Los Alamos, NM, 1981.
42. L.J. Radziemski, T.R. Loree, D.A. Cremers, N.M. Hoffman, 'Time-resolved Laser-induced Breakdown Spectroscopy (LIBS): A New Method for Spectrochemical Analysis', LANL Report LA-UR-82-487, Los Alamos National Laboratory, Los Alamos, NM, 1982.
43. J. Uebbing, J. Brust, W. Sdorra, F. Leis, K. Niemax, 'Reheating of a Laser-produced Plasma by a Second Pulse Laser', *Appl. Spectrosc.*, **45**(9), 1419–1423 (1991).
44. D.A. Cremers, L.J. Radziemski, 'Direct Detection of Beryllium on Filter Using the Laser Spark', *Appl. Spectrosc.*, **39**(1), 57–63 (1985).
45. K.J. Mason, J.M. Goldberg, 'Production and Initial Characterization of a Laser-induced Plasma in a Pulsed Magnetic Field for Atomic Spectrometry', *Anal. Chem.*, **59**, 1250–1255 (1987).
46. K.J. Mason, J.M. Goldberg, 'Characterization of a Laser Plasma in a Pulsed Magnetic Field. Part I: Spatially Resolved Emission Studies', *Appl. Spectrosc.*, **45**(3), 370–379 (1991).
47. K.J. Mason, J.M. Goldberg, 'Characterization of a Laser Plasma in a Pulsed Magnetic Field. Part II: Time-resolved Emission and Absorption Studies', *Appl. Spectrosc.*, **45**(9), 1444–1455 (1991).
48. D.A. Cremers, 'The Analysis of Metals at a Distance Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **41**(4), 572–579 (1987).
49. D. Bradley, 'News from The Electronic Analytical Chemistry Conference', *Anal. Chem. News & Features*, January 1 (1998).
50. L.W. Peng, W.L. Flower, K.R. Hencken, H.A. Johnsen, R.F. Renzi, N.B. French, 'A Laser-based Technique for Continuously Monitoring Metal Emissions From Thermal Treatment Units', *Process Control Quality*, **7**, 39–49 (1995).
51. J.P. Singh, F.Y. Yueh, H. Zhang, R.L. Cook, 'Analytical Method Using Laser-induced Breakdown Spectroscopy', US Patent 5,751,416.
52. B.J. Marquardt, D.N. Stratis, D.A. Cremers, S.M. Angel, 'Novel Probe for Laser-induced Breakdown Spectroscopy and Raman Measurements Using an Imaging Optical Fiber', *Appl. Spectrosc.*, **52**(9), 1148–1153 (1998).
53. W.E. Ernst, D.F. Farson, D.J. Sames, 'Determination of Copper in A533b Steel for the Assessment of Radiation Embrittlement Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**(3), 306–309 (1996).
54. B. Flower, H. Johnsen, K. Hencker, D. Hahn, 'Field Test of a Continuous Emissions Monitor for Metals Based on Laser Spark Spectroscopy', Technical report for continuous emission monitor (CEM) test at the Rotary Kiln Incinerator Simulator (RKIS) at the EPA Environmental Research Center, Research Triangle Park, North Carolina, April, 1996.

55. J.P. Singh, H. Zhang, F.Y. Yueh, 'DOE and EPA Continuous Emission Monitoring Test at EPA National Risk Management Research Laboratory (NRMRL)', Technical Report for Continuous Emission Monitor (CEM) Test at the Rotary Kiln Incinerator Simulator (RKIS) at the EPA Environmental Research Center, Research Triangle Park, North Carolina, April, 1996.
56. H. Zhang, F.Y. Yueh, J.P. Singh, 'Laser-induced Breakdown Spectrometry as a Multi-metal Continuous Emission Monitor', *Appl. Optics*, **38**(9), 1459–1466 (1999).
57. H. Zhang, J.P. Singh, F.Y. Yueh, R.L. Cook, 'Laser-induced Breakdown Spectra in a Coal-fired MHD Facility', *Appl. Spectrosc.*, **49**(11), 1617–1623 (1995).
58. D.K. Ottesen, J.C.F. Wang, L.J. Radziemski, 'Real-time Laser Spark Spectroscopy of Particulates in Combustion Environments', *Appl. Spectrosc.*, **43**(6), 967–976 (1989).
59. C.J. Lorenzen, C. Carlhoff, U. Hahn, M. Jogwich, 'Applications of Laser-induced Emission Spectral Analysis for Industrial Process and Quality Control', *J. Anal. At. Spectrom.*, **7**, 1029–1035 (1992).
60. K. Song, Y.I. Lee, J. Sneddon, 'Application of Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc. Rev.*, **32**(3), 183–235 (1997).
61. R.E. Neuhauser, U. Panne, R. Niessner, G.A. Petrucci, P. Vavalli, N. Omenetto, 'On-line and In-situ Detection of Lead Aerosols by Plasma-spectroscopy and Laser-excited Atomic Fluorescence Spectroscopy', *Anal. Chim. Acta*, **346**, 37–48 (1997).
62. J.P. Singh, F.Y. Yueh, H. Zhang, R.L. Cook, 'Study of Laser-induced Breakdown Spectroscopy as a Process Monitor and Control Tool for Hazardous Waste Remediation', *Process Control Quality*, **10**, 247–258 (1997).
63. J.P. Singh, H. Zhang, F.Y. Yueh, K.P. Carney, 'Investigation of the Effects of Atmospheric Conditions on the Quantification of Metal Hydrides Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**(6), 764–773 (1996).
64. C. Lazzari, M.D. Rosa, S. Rastelli, A. Ciucci, V. Palleschi, A. Salvetti, 'Detection of Mercury in Air by Time-resolved Laser-induced Breakdown Spectroscopy Technique', *Laser Particle Beams*, **12**, 525–530 (1994).
65. N.B. French, W.J. Haas, P.M. Lemieux, S.J. Priebe, 'Performance Testing of Multi-metal Continuous Emissions Monitors', US EPA, NRMRL, Raleigh, US Department of Energy, EM-Office of Science and Technology, 1998.
66. G.W. Loge, 'A Continuous Calibration Procedure for Reliable LIBS Monitoring of Metals', in *Proceedings of International Conference on Incineration and Thermal Treatment Technologies*, University of California, Irvine, CA, 227–230, 1997.
67. D.W. Hahn, W.L. Flower, K.R. Hencken, 'Discrete Particle Detection and Metal Emission Monitoring Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **51**(12), 1836–1844 (1997).
68. H. Kurniawan, S. Nakajima, J.E. Batubara, M. Marpaung, M. Okamoto, K. Kagawa, 'Laser-induced Shock Wave Plasma in Glass and its Application to Elemental Analysis', *Appl. Spectrosc.*, **49**(8), 1067–1072 (1995).
69. T.L. Thiem, R.H. Salter, J.A. Gardner, Y.I. Lee, J. Sneddon, 'Quantitative Simultaneous Elemental Determinations in Alloys Using Laser-induced Breakdown Spectroscopy (LIBS) in an UltraHigh Vacuum', *Appl. Spectrosc.*, **48**(1), 58–64 (1994).
70. M. Sabsabi, P. Cielo, 'Quantitative Analysis of Aluminum Alloys by Laser-induced Breakdown Spectroscopy and Plasma Characterization', *Appl. Spectrosc.*, **49**(4), 499–507 (1995).
71. C. Argagon, J.A. Aguilera, Campos, 'Determination of Carbon Content in Molten Steel Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **47**(5), 606–608 (1993).
72. L. Paksy, B. Nemet, A. Lengyel, I. Kozma, J. Czekkel, 'Production Control of Metal Alloys by Laser Spectroscopy of the Molten Metals. Part 1. Preliminary Investigations', *Spectrochim. Acta, B*, **51**, 279–290 (1996).
73. K.Y. Yamamoto, D.A. Cremers, M.J. Ferris, L.E. Foster, 'Detection of Metals in the Environment Using a Portable Laser-induced Breakdown Spectroscopy Instrument', *Appl. Spectrosc.*, **50**(2), 222–233 (1996).
74. D.A. Cremers, J.E. Barefield, A.C. Koskelo, 'Remote Elemental Analysis by Laser-induced Breakdown Spectroscopy Using a Fiber-optic Cable', *Appl. Spectrosc.*, **49**(6), 857–860 (1995).
75. A.S. Eppler, D.A. Cremers, D.D. Hickmott, M.J. Ferris, A.C. Koskelo, 'Matrix Effects in the Detection of Pb and Ba in Soils Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**(9), 1175–1181 (1996).
76. A.V. Pakhomov, W. Nichols, J. Borysow, 'Laser-induced Breakdown Spectroscopy for Detection of Lead in Concrete', *Appl. Spectrosc.*, **50**(7), 880–884 (1996).
77. S.D. Arnold, D.A. Cremers, 'Rapid Determination of Metal Particles on Air Sampling Filters Using Laser-induced Breakdown Spectroscopy', *AIHA J.*, **56**, 1180–1186 (1995).
78. J.P. Singh, F.Y. Yueh, H. Zhang, K.P. Carney, 'A Preliminary Study of the Determination of Uranium, Plutonium and Neptunium by Laser Induced Breakdown Spectroscopy', *Recent Res. Develop. Appl. Spectrosc.*, **2**, 59–67 (1999).
79. W.E. Ernst, D.F. Farson, D.J. Sames, 'Determination of Copper in A533b Steel for the Assessment of Radiation Embrittlement Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**(3), 306–309 (1996).
80. T.L. Thiem, P.J. Wolf, 'Analysis of National Institute of Standards and Technology Ore and United States Geological Survey Samples by Inductively Coupled Plasma Spectroscopy/Laser-induced Breakdown Spectroscopy', *Microchem. J.*, **50**, 244–252 (1994).

Matrix Isolation Spectroscopy in Atmospheric Chemistry

K.S. Viswanathan, K. Sankaran, and K. Sundararajan
Indira Gandhi Centre for Atomic Research, Kalpakkam, India

1 Introduction **1**
2 Need for a Technique such as Matrix Isolation **3**
2.1 Broadening Mechanisms 3
3 Matrix Isolation Spectroscopy **4**
3.1 Experimental Aspects 6
4 Applications **9**
4.1 Analytical Spectroscopy 9
4.2 Study of Reaction Intermediates 15
5 Conclusions **16**
Abbreviations and Acronyms **17**
Related Articles **17**
References **17**

There is a growing concern about our environment and it has been realized that quick and definite measures are required if we are to preserve our planet for posterity. The rising level of greenhouse gases in the atmosphere and its consequent effect on the global climatic conditions, the depletion of stratospheric ozone, acid rain, and photochemical smog are some of the issues that have been addressed by scientists and policy makers the world over. Solutions to most environmental problems can be obtained only through collective efforts and, to ensure that such efforts are effective, it is necessary that policies and legislation are made based on scientific data on our ever-changing environment. The data must provide information on the nature of trace constituents present in the different layers of the atmosphere, their concentrations, and their chemistry. A variety of experimental techniques have been used for this purpose, such as fluorescence spectroscopy, infrared (IR) and ultraviolet/visible (UV/VIS) absorption, Raman scattering, and photoacoustic spectroscopy. All these techniques exhibit some combination of the following features: multielement detection, low detection limits (DLs), specificity for an unequivocal identification of species, accuracy, and precision. One of the experimental techniques that has found important applications in the study of the atmosphere is matrix isolation (MI) spectroscopy. In this technique, the molecules of interest are diluted in a large excess of an inert gas and sprayed onto a cold substrate held at 10 K. Under these conditions, the molecules are trapped, isolated from each other and only surrounded by inert gas atoms. Such cold, isolated molecules yield spectra that have narrow spectral widths. Furthermore, if the trapped species are reactive molecules or radicals, they are likely to have extended lifetimes in the inert cage, thus allowing their chemistry and spectroscopy to be studied at leisure. These features of MI have made it particularly useful in the study of atmospheric chemistry. The high-resolution capability enables one to use this technique as an analytical tool, as closely lying spectral features of different molecules can now be resolved – a feature that has been employed to identify and estimate trace constituents in the upper tropospheric and stratospheric samples. In conjunction with gas chromatography (GC), MI and Fourier transform infrared (FTIR), referred to as GC/MI/FTIR, offers a powerful tool with which to resolve isomeric forms of environmentally hazardous chemicals such as polycyclic aromatic hydrocarbons (PAHs), dioxins, and polychlorinated biphenyls (PCBs). Such isomeric resolution is essential, as it is well known that only certain isomeric forms are biologically active, whereas the others are not. In this respect, GC/MI/FTIR even scores over GC mass spectroscopy (MS). Analytical spectroscopy of matrix-isolated species is also done using fluorescence spectroscopy, where again the advantage of high resolution enables one to resolve isomeric forms of compounds. Where the species of interest is a free radical, electron spin resonance (ESR) is the technique of choice to study the trapped species. This is a particularly powerful tool, as free radicals are known to play an important role in a number of atmospheric processes. Another aspect of MI spectroscopy is its ability to study reaction intermediates, a feature that has been employed to study the reaction mechanisms of atmospherically important chemical and photochemical reactions. All these aspects of the technique are discussed in this article, using examples.

1 INTRODUCTION

The remarkable technological growth that the human race has seen in recent decades is not without its influence on our environment. The increase in automobile exhaust emissions, effluents and emissions from industrial processes, and harmful and toxic effects of new chemicals have all contributed to changes in our atmosphere. There has been a growing concern over this deleterious effect of human ''progress'' and it has been realized for some time now that effective measures are imperative if we are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

to preserve this planet for posterity. Implementation of the Montreal Protocol in 1987 and a subsequent London Revisions in 1990 restricting the use of chlorofluorocarbon (CFC) is an example of the legislative measures that have been sought. The United Nations First Conference on Environment and Development held at Rio de Janeiro in 1992, is also testimony to the concern shown by nations world over with regard to our environment. Although legislation is undoubtedly important, the formulation of effective legislative measures requires reliable data about our environment. Information is needed on the changes in the composition of the atmosphere over a period of time, and the effect of such changes on global climatic conditions.

Our environment is continuously altered by a variety of forcing agents, both natural and man made. For example, the concentrations of greenhouse gases such as carbon dioxide, methane, CFCs, and oxides of nitrogen have been on the rise. Some of these gases are also new additions, such as the CFCs, whose role in atmospheric chemistry is a twentieth-century phenomenon. Another activity that has an impact on our environment is biomass burning, which is resorted to for clearing forests for agricultural purposes and burning postharvest residues. In developing countries, the use of wood as a fuel material also serves to contribute to this process. As a result, the atmosphere is loaded with a variety of trace gases and particles, that have a profound influence on the atmosphere, both locally and globally. For example, the emission of oxides of nitrogen and sulfur, can result in acid rain, affecting the local ecology. However, species such as CO and CH_4 have global impacts. These species can significantly lower the concentration of OH radicals, which controls the concentrations of many atmospheric trace constituents and provides a sink for some of the greenhouse gases. A reduced concentration of the OH radicals may consequently increase the concentration of these greenhouse gases. It is therefore imperative to be alert to changes which is only possible by conducting a continuous ''assay'' of the atmosphere.

Before discussing the experimental techniques used to identify the atmospheric constituents, it may be worthwhile understanding a little about the structure of the atmosphere itself. The earth's atmosphere is divided into many regions based on the temperature gradient, as shown in Figure 1.[1] The region up to a height of about 10 km above the surface of the earth is the troposphere, the height that we usually touch during air travel. The region from 10 to 50 km is the stratosphere, where a number of important photochemical reactions take place. The impact that chemical species can have on our environment depends on where they are located in the atmosphere and their reactivity towards other species present alongside. For example, ozone in the troposphere

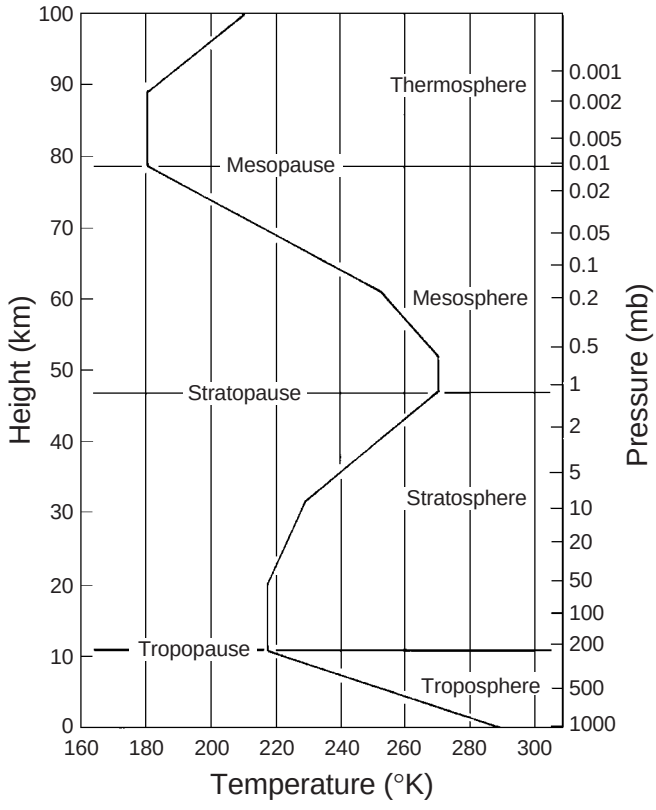

Figure 1 Vertical temperature profiles and the different regions of the atmosphere. (Reproduced by permission of Academic Press from Wallace and Hobbs[1].)

contributes to undesirable greenhouse warming and has adverse effects on humans and vegetation, whereas its presence in the stratosphere is essential to filter ultraviolet (UV) radiation from the sun. It is therefore necessary to map the concentration profiles of the various species in the atmosphere. Such mapping is done by analyzing air samples collected over different regions on the land and sea and at different heights in the atmosphere. Furthermore, the lifetimes of these species in the atmosphere is determined by their chemical reactivity and their transport dynamics. For a clear understanding of the ever-evolving atmosphere, it is necessary to understand its composition with respect to the major and trace constituents, their chemistry, and transport phenomena.

The experimental techniques used for the study of the atmosphere can be classified into two groups. The first of these are analytical methods to identify and estimate atmospheric constituents. A number of spectroscopic techniques are used for this purpose, such as fluorescence spectroscopy, IR and UV absorption, differential optical absorption spectroscopy (DOAS), photoacoustic spectroscopy, and laser-based methods involving both photon and ion detection.[2] These techniques exhibit all

or some combination of the following features: low DLs, specificity for an unequivocal identification of species, multispecies detection, accuracy, and precision. The second category of studies are aimed at understanding the kinetics and mechanisms of reactions and reaction intermediates, which play an important role in the photochemical processes of the atmosphere. These studies are generally conducted using laser-based pump-probe methods, IR and UV absorption, and ESR spectroscopy. The first category of experiments provide information on what species are present in the atmosphere, whereas the second describes their reactivity.

This article discusses the use of MI spectroscopy for atmospheric studies, both as an analytical tool and to study reaction mechanisms and intermediates, relevant to atmospheric processes. MI is a technique in which the molecules of interest are isolated in a cold, inert gas matrix and studied using a variety of spectroscopic techniques such as IR, fluorescence, ESR, and Mössbauer spectroscopy, to name a few. The technique has been widely used to study conformations, hydrogen bonded and van der Waals molecules, and high-temperature species; its use in atmospheric studies is yet another example of its utility.

2 NEED FOR A TECHNIQUE SUCH AS MATRIX ISOLATION

To appreciate the need for the technique of MI, it is worthwhile taking a look at spectra obtained with the commonly employed gas-phase and condensed-phase sampling methods. To build up this discussion, the IR spectra of a few compounds are considered.

Figure 2 shows the IR spectra of ethylene, benzene, and benzoic acid, serving as typical examples of gas, liquid, and solid samples, respectively. Ethylene at a pressure of 1 torr was used to fill a 6 in path-length gas cell, benzene was taken as a liquid film on a zinc selenide substrate, and benzoic acid was impregnated in a KBr pellet (0.5%). All these sampling methods are popularly used in IR spectroscopy. The peaks seen in the spectra reflect absorptions corresponding to vibrations of specific groups of atoms in the molecule.

What should be noted in these spectra are the *widths* of the spectral features. The absorption peaks in Figure 2 display a width that is typically about 20–30 cm^{-1}. In the case of gas-phase samples, the spectra consist of a large number of sharp lines with the contour extending to almost 100 cm^{-1}. Very often large spectral widths or contours are undesirable, particularly in experiments where it is required to identify the constituents in a multicomponent mixture. Two closely

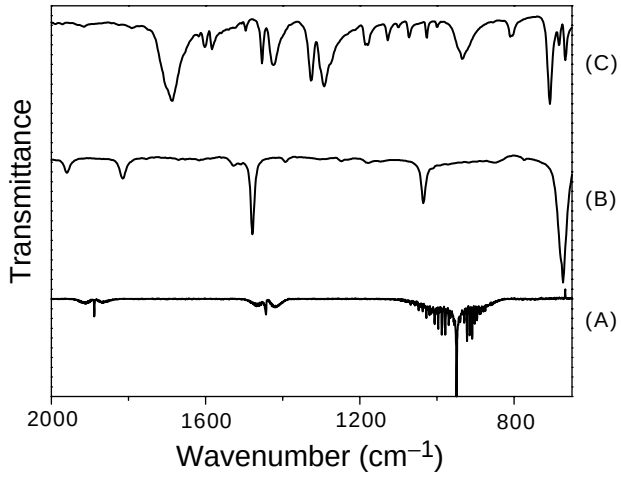

Figure 2 IR spectrum of (A) ethylene (gas phase), (B) benzene (liquid), and (C) benzoic acid (solid phase).

lying spectral features of two different compounds can be unambiguously resolved only if the peak width is small compared with the separation between the peaks. It is therefore necessary to be able to obtain spectra with small peak widths.

2.1 Broadening Mechanisms

To reduce the width of the spectral features, it is necessary to understand what broadens them in the first place. There are a number of reasons as to why spectral features are broadened,[3] including

- natural broadening;
- Doppler broadening;
- collisional broadening;
- distribution of molecules over a number of energy states;
- interactions between molecules.

2.1.1 Natural Broadening

This arises owing to the finite lifetimes of excited states and is a manifestation of the Heisenberg uncertainty principle. The shorter the lifetime of a state, the greater is the uncertainty in the energy of that state. Consequently, the spectral width of a transition associated with that state will be large. The experimentalist has hardly any control over this broadening mechanism.

2.1.2 Doppler Broadening

Doppler broadening is a consequence of molecular motion in a sample. The relative motion of the molecules and the source (or detector), causes the transition frequency to be shifted; the magnitude and direction of

the shift depend on the relative velocity of the molecule with respect to the source (or detector). (Recall the varying pitch of a hooting locomotive as it approaches and then recedes from you.) As the molecules in a sample have a distribution of molecular speeds given by Maxwell's distribution, the transition frequencies are also distributed over a range of values, leading to a broadened spectral profile. The higher the temperature, the greater is the width of the speed distribution and hence the larger the peak width due to Doppler broadening.

2.1.3 Collisions

The collisions that occur between molecules, both elastic and inelastic, are also responsible for broadened spectral features. Inelastic collisions tend to shorten the lifetimes of the excited states, which culminates in a broadened profile. However, elastic collisions may result in phase perturbations during absorption or emission which can result in peak broadening.

2.1.4 Energy States

The distribution of molecules over various energy states, corresponding to rotational and vibrational motion, is also a cause of broad spectral profiles. For example, the large number of sharp peaks seen in the gas-phase spectrum of Figure 2 (trace A), corresponds to transitions from a number of rotational levels of ethylene; such complexities in spectra are referred to as *spectral congestion*.

2.1.5 Intermolecular Interactions

These can also perturb energy levels, leading to spectral broadening.

Only some of the broadening mechanisms have been discussed here; there are others, such as pressure broadening, time-of-flight broadening, etc. A detailed treatment of these can be found in textbooks on spectroscopy.[3] All or some combination of these factors play a role in broadening the spectral features of a sample.

A number of these broadening mechanisms depend on the temperature of the sample. For example, if the temperature T of the sample is lowered, the width of the speed distribution decreases. As a result the Doppler width, which can be shown to be proportional to $\sqrt{T}$, also decreases. Lowering the temperature also restricts the population to only a few rotational and vibrational levels, which gives rise to spectra with fewer features. It may therefore appear that by simply lowering the temperature of the sample, one can obtain a reduction in the peak widths. However, this is not necessarily the case, as Figure 3 shows. The IR spectra of liquid trimethyl phosphate at room temperature is compared with the frozen compound at 12 K. In spite of the decrease

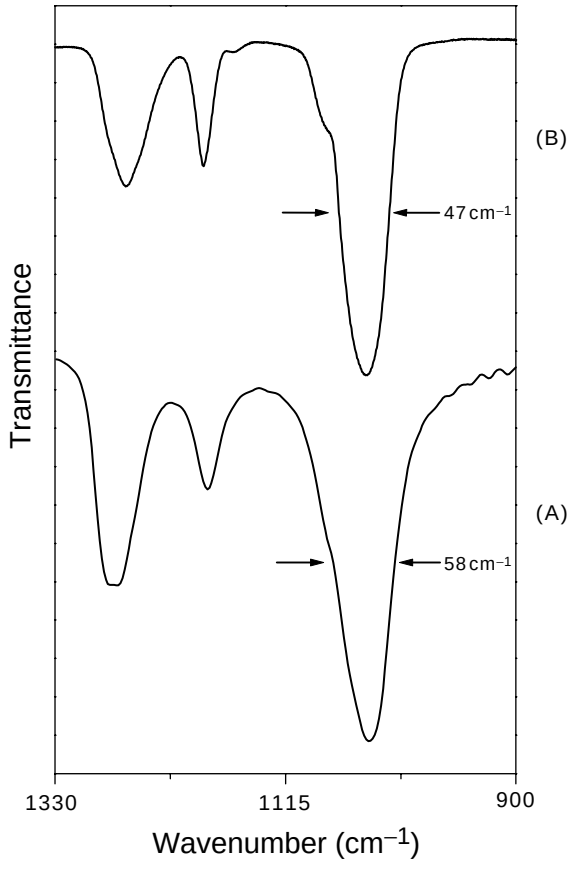

Figure 3 IR spectrum of trimethyl phosphate (A) liquid at room temperature and (B) solid at 12 K.

expected in the Doppler width and spectral congestion at low temperatures, almost no reduction in peak width is obtained. This is because intermolecular interactions continue to operate in the frozen sample, resulting in broad spectral features.

It is therefore necessary not only to cool the samples, but also to minimize the intermolecular interactions, by isolating the molecules from each other.

The common sampling methods mentioned earlier – gas, liquid, or solid – cannot combine low temperatures and isolated-molecule conditions. It is therefore necessary to look beyond the common sampling methods to produce *cold* and *isolated* molecules. MI spectroscopy is one such experimental technique.

Although the above discussion uses IR spectroscopy as an example, the arguments apply equally well to other branches of spectroscopy.

3 MATRIX ISOLATION SPECTROSCOPY

In an MI experiment, vapors of the sample of interest are mixed with a large excess of an inert gas, such as argon

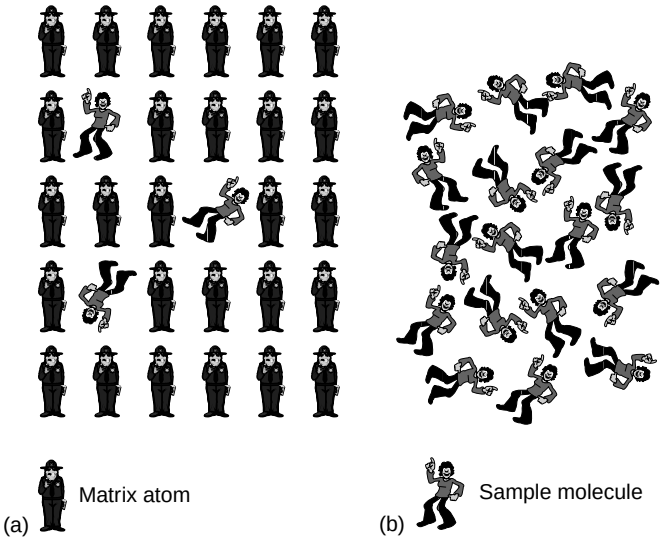

Figure 4 Cartoon depicting the environment in (a) a matrix-isolated sample and (b) a liquid sample.

or nitrogen, and this gas mixture is then sprayed onto a cold substrate (≈10 K). The inert gas to sample ratios are typically 1000 : 1. The large excess of an inert gas ensures that when the gas mixture freezes on the substrate, the sample molecules are mostly trapped isolated from each other and only surrounded by inert gas atoms, as depicted in the cartoon (Figure 4a). To a first approximation it can be assumed that the inert gas atoms do not interact with the trapped species, which therefore reside in an inert cage. This environment is devoid of strong intermolecular interactions, as opposed to the situation for a liquid sample (Figure 4b). The low temperatures of the matrix also ensure that only the low vibrational and rotational levels are populated. Furthermore, the low temperatures and the rigid matrix cage constrain molecular motions and lead to an elimination of the Doppler width. Figure 5 shows the IR spectra of gas phase acetylene, and acetylene trapped in an argon matrix at about 12 K. The extensive rotational contour that extends over 100 cm^{-1} in the gas phase is absent in the matrix-isolated spectra, which displays a sharp peak with a width of only about 2 cm^{-1}. Cooling and isolating the molecules in an inert environment has therefore done the trick!

Apart from sharp spectral features, another important feature of the MI technique is the interaction-free environment that it provides for the trapped species. If the trapped species are reactive, such species now have extended lifetimes, simply for want of a reaction partner. Without the isolation, the reactive species would most likely annihilate themselves by reacting with each other or with the surrounding species. Spectroscopic studies of reactive species trapped in an inert matrix can therefore be made at leisure; a feature which makes it possible to

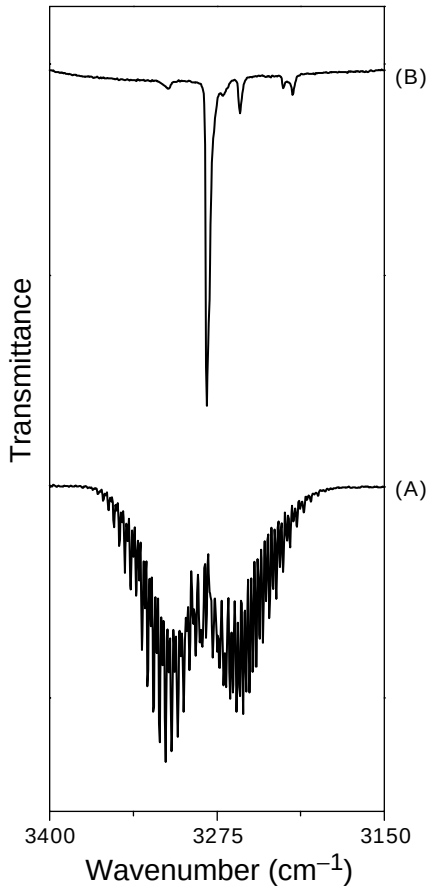

Figure 5 Comparison of IR spectra of acetylene: (A) in the gas phase, (B) trapped in an argon matrix at 12 K.

study intermediates in chemical reactions. Historically, MI was first developed for this very purpose of studying transient species, by trapping them in inert matrices.[4]

MI therefore presents a twofold advantage. First, it can be used as an analytical tool for the estimation of the constituents in an atmospheric sample. In these applications the small line widths afforded by the technique is taken advantage of. Figure 6 illustrates this aspect of MI spectroscopy, where the gas-phase and matrix isolation infrared (MI/IR) spectra of a five-component mixture consisting of ethylene and deuterated ethylene are presented.[5] It is clear that the peaks due to the different components in the mixture are clearly resolved in the MI spectra. Furthermore, the small peak widths also lead to larger peak heights, thereby rendering measurements of peak intensities more accurate. Consequently, low concentrations can be accurately measured in MI experiments. Second, the ability of the technique to trap reactive species can be used to study the structure and reactions of species involved in atmospheric chemical processes. Both aspects are highlighted in the discussions that follow.

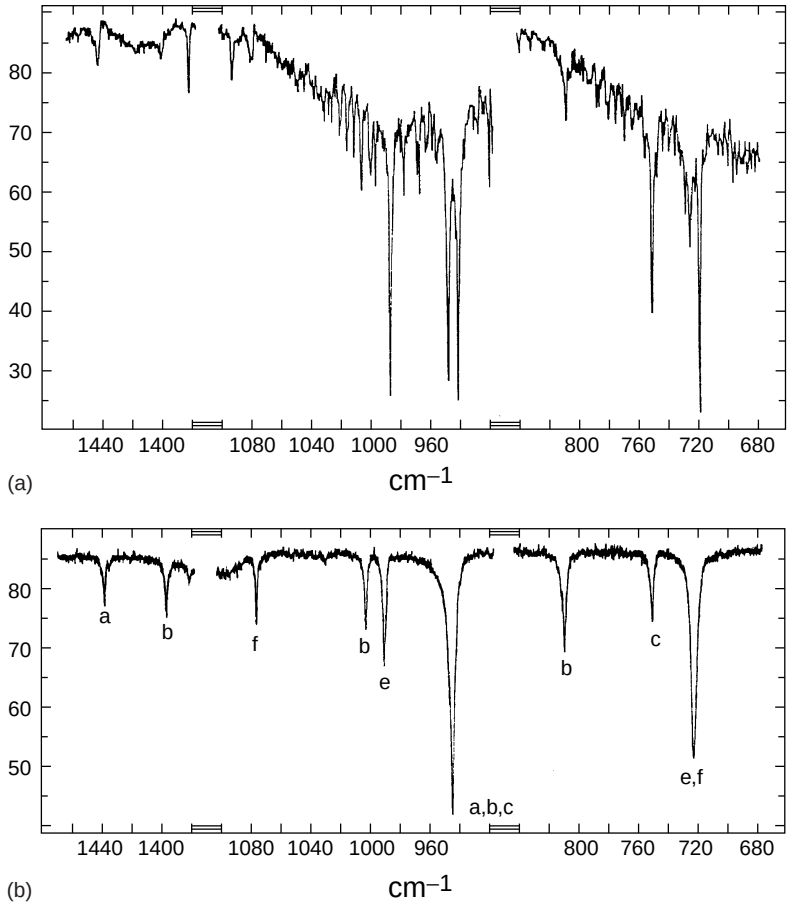

Figure 6 IR absorption spectra of a five-component mixture of ethylene and deuterated ethylenes in the gas phase (a) and matrix isolated in nitrogen (b): a = ethylene; b = ethylene-D_1; c = ethylene-D_2 (1,1); e = ethylene-D_2 (1,2-*trans*); f = ethylene-D_4. (Reproduced with permission from Rochkind.[5] Copyright 1968 American Chemical Society.)

MI is fundamentally a sample preparation technique. Once the species are trapped in the inert matrix, a variety of spectroscopic methods may be used, such as IR, UV/VIS, and ESR spectroscopy. The studies performed on the trapped species depends on the information being sought.

3.1 Experimental Aspects

A detailed treatment of the experimental aspects of MI is given in many text books on this subject.[6–8] Only a broad outline, detailing the important aspects of the technique is presented here. The MI equipment consists of:

- a cryostat to produce the low temperatures;
- a vacuum system to house the cold tip, where the sample is deposited;
- a sample handling system and deposition lines;
- a spectrometer to record the spectra of matrix-isolated species.

3.1.1 Cryostat

A cryostat is used to produce the low temperatures required to deposit the sample and the inert gas. Cryostats are of three types.

3.1.1.1 Liquid Refrigerant-based Cryostats In these cryostats, liquid refrigerants such as helium (4.2 K), liquid nitrogen (77 K) and liquid air (90 K) are used to attain low temperatures. The earliest designs involved a tubular reservoir, in which the liquid refrigerant is filled. To this reservoir is attached a sample window on which the matrix is deposited. Where liquid helium is used, a double-Dewar arrangement is essential to minimize the evaporation loss of the expensive refrigerant. In this design, the liquid helium reservoir is surrounded by a reservoir of liquid nitrogen, the secondary refrigerant acting as a heat shield.

In later modifications, liquid helium was passed through the reservoir to which the cold finger is attached. These continuous-flow cryostats do not require a secondary

refrigerant. They are also more economical than double-Dewar cryostats, and are compact and portable.

3.1.1.2 Joule–Thomson Open Cycle Cryostat In this type of cryostat a high-pressure gas is expanded through an orifice to produce low temperatures. This design does away with the need to handle liquid refrigerants and are extremely compact. Commercial models are available that cover the temperature range 4–200 K.

3.1.1.3 Closed Cycle Cryostats This type of cryostat is the most popular. In these cryostats, which work on the principle of the Gifford–McMahon cycle, cooling is achieved by the continuous compression and expansion of He, in an enclosed space. Helium is first compressed to a pressure of about 20 bar and the heat of compression is removed by circulating cold water. The helium is then delivered to the cold head, where it expands against a piston, and eventually cools the regenerator. The expanded helium is then returned to the compressor for the next cycle. Although the initial capital costs of these cryostats are high, they involve almost no running costs and are extremely easy to handle.

3.1.2 Vacuum System

The cryotip onto which the sample is deposited is housed in a vacuum system. A vacuum system is necessary to maintain the low temperature and also to provide a clean environment in which the matrix can be deposited. A vacuum of better than 10^{-6} mbar is essential for the successful conduct of an MI experiment. The presence of impurities, apart from producing impurity peaks, can also lead to complications by forming complexes with the molecules of interest. Hence the need to maintain a high vacuum in these experiments cannot be overemphasized. A diffusion pump backed by a rotary pump is popularly used to produce the necessary high vacuum. Although diffusion pumps are easy to handle and relatively inexpensive, problems due to oil backstreaming into the vacuum chamber can at times be annoying. This problem can be surmounted by using a liquid-nitrogen trap. However, wherever possible, it may be desirable to use oil-free high-vacuum pumps, such as a turbomolecular pump.

3.1.3 Deposition of the Matrix

A variety of deposition schemes is used to prepare the matrix-isolated sample. In the simplest case, the matrix gas and the sample(s) of interest are mixed to the desired concentrations in a mixing chamber using manometric procedures and this gas mixture is then sprayed onto a cold substrate through a nozzle. The flow through this nozzle, which is generally effusive, is controlled using fine flow-control valves. Typical deposition rates are about 1 mmol h^{-1} of matrix gas. Deposition rates should be controlled to ensure good optical quality of the matrices and also to prevent any excessive heat liberation during matrix condensation. Undue temperature rise of the cold finger can cause diffusion of the solute molecules, leading to undesirable aggregation. As a variation of the above deposition procedure, the matrix and the sample can be taken in two different containers and passed through two different nozzles, referred to as a twin-nozzle system. In this set-up, mixing of the matrix gas with the sample occurs soon after effusion out of the nozzle and the mixture then deposits on to the cold tip. In such experiments, the required matrix to sample ratio is achieved by controlling the relative flow rates of the sample and the matrix gas. As a logical extension, where reactions between two precursors are to be studied, a triple-jet assembly is used, where the matrix gas and the two samples effuse out through three separate nozzles. The different nozzle designs are shown in Figure 7.

The methods of sample deposition described above are employed where stable molecules are to be studied. For the study of reactive species, variations of the above procedures are employed.[9] A stable species can be first deposited, followed by photolysis to yield the reactive species. For example, photolysis of $W(CO)_6$ in an Ar matrix yielded the reactive $W(CO)_5$ and CO. Alternatively, the reactive species can be produced in the gas phase (prior to deposition) by passing the sample(s) through an electric discharge or by subjecting it to photolysis, and the reaction products are then

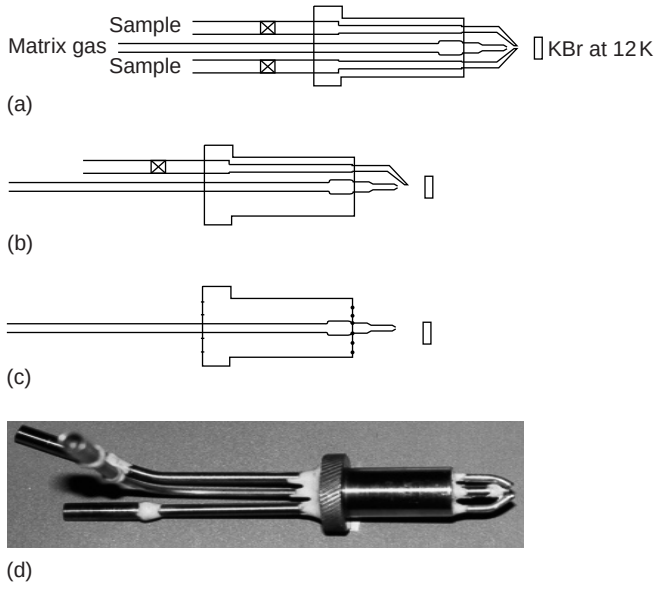

Figure 7 Different nozzle designs: (a) triple-jet nozzle; (b) twin-jet nozzle; (c) single-jet nozzle; (d) photograph of a triple-jet nozzle.

trapped in the inert matrix. For example, $XeCl_2$ can be prepared by passing Xe and Cl_2 through a discharge tube.[9] As mentioned earlier, these reactive species, which normally would have instantaneously indulged in a chemical reaction destroying themselves, now have extended lifetimes in the inert cage. This technique has been exploited for studies on atmospheric chemistry, as reactive molecules expected to participate in atmospheric processes can be trapped in the matrix for spectroscopic analysis and study of their reactions.

3.1.3.1 Annealing Once the sample and matrix are deposited, a spectrum of the matrix-isolated sample is recorded. Very often, a spectrum of the sample is also recorded after annealing the matrix. Annealing is a process where the temperature of the matrix is raised to about $0.4\,T_m$, where T_m is the melting point of the matrix material. The matrix is held at this temperature for some time, (typically about 15 min to 1 h) and then recooled to 12 K. Annealing of the matrix is done for two reasons. First, it removes unstable sites in the matrix in which the sample molecules may be trapped. The sites referred to are the different environments in which the sample molecules are trapped in the lattice formed by the inert gas during deposition. One such environment is formed by substituting the sample molecules for one or more of the matrix atoms; such a site is referred to as a substitutional site and is the most likely one for large molecules. Alternatively, the sample molecules may occupy either a dislocation or an interstitial site. The interstitial site is usually too small to accommodate most molecules other than small diatomics. Dislocation sites are large but nonuniform. Although some of these sites may be unstable, they may nevertheless be formed during matrix deposition owing to the rapid freezing. The presence of these different sites may be signaled by the appearance of multiplets in a spectra. Annealing the matrix generally simplifies such spectra, by removing sites that are unstable. Figure 8 shows the spectra of trimethyl phosphate recorded after deposition and after annealing the matrix. It can be seen that with the loss of the unstable sites, the spectral features sharpen on annealing. As discussed below, the role of sites in MI spectra is particularly relevant in the study of fluorescence of species trapped in hydrocarbon matrices. Another reason to anneal the matrix is to deliberately cause diffusion of the solute molecules in an effort to drive a reaction. The diffusion may cause two molecules, (which were initially isolated) to approach each other and form aggregates. Appearance of new peaks following annealing is generally a signature of the occurrence of such reactions.

3.1.3.2 Matrix Substances must possess certain properties for them to serve as matrices. They must be *inert* so that they do not perturb or react with the trapped species. At the temperatures of the MI experiments, the substance must form a *rigid* matrix, so that the molecules of interest may be held isolated. If this criterion is not met, the trapped species will diffuse, leading to undesirable reactions. At the same time, the substance must be *volatile* at room temperature, so that it may be conveniently deposited onto the cold substrate. It is also important that the matrix substance must be available with a high degree of *purity* and not have any spectral absorptions in the region of interest.

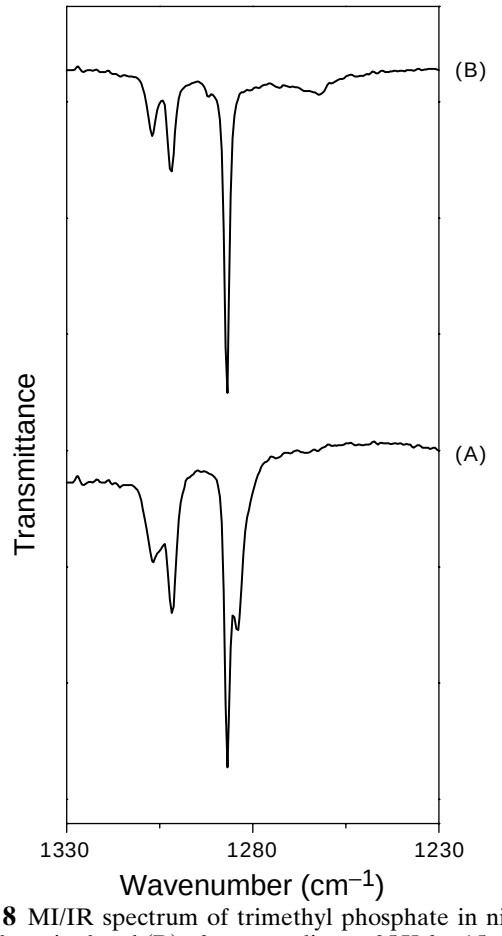

Figure 8 MI/IR spectrum of trimethyl phosphate in nitrogen (A) as deposited and (B) after annealing at 35 K for 15 min.

Inert gases and nitrogen generally meet the above criteria and are used as matrix gases. Among the inert gases, argon and to a lesser extent neon and xenon are commonly used. These gases are popular where IR spectroscopy is used to probe the trapped species, as these matrix materials do not have any IR absorptions. There have been specific instances where other materials, such as carbon dioxide, water (or D_2O), and hydrocarbons, have been used as matrices. These are discussed later.

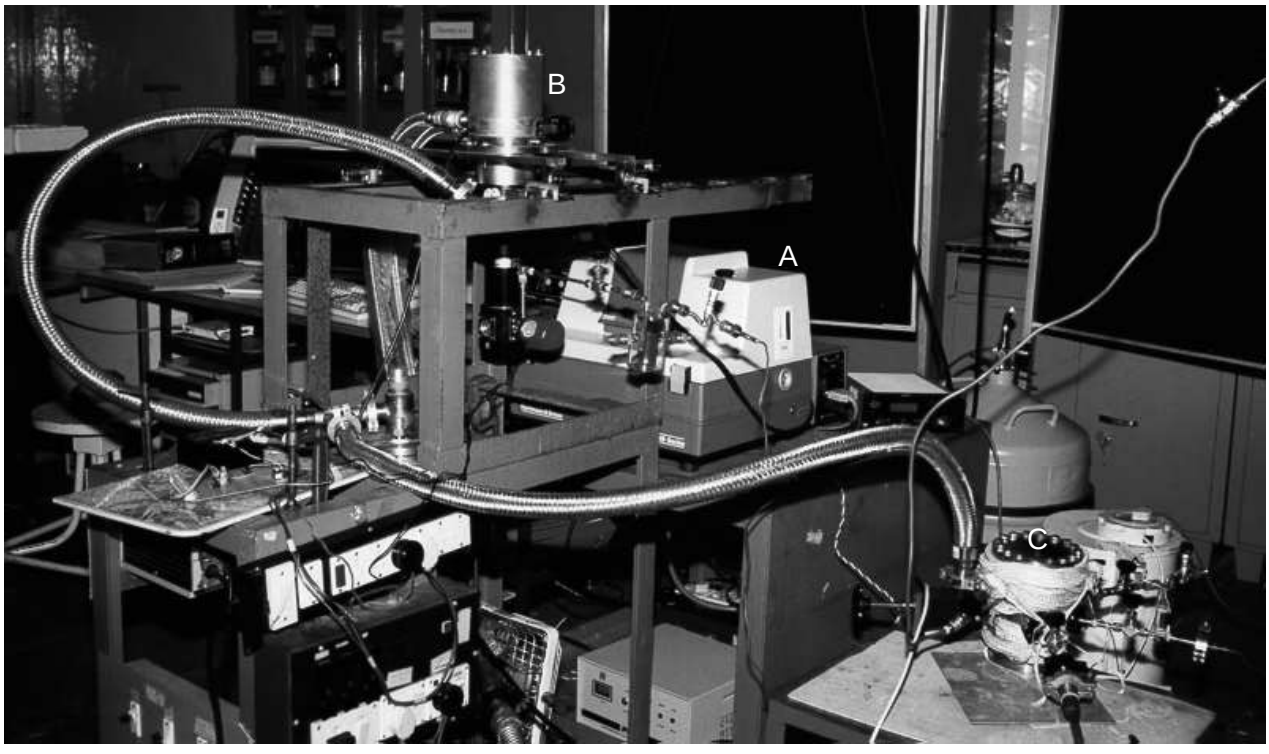

Figure 9 Photograph showing the MI set-up: (A) is the FTIR spectrometer, (B) the closed cycle helium cryostat, and (C) the mixing chamber.

3.1.4 Spectrometer

Once the sample molecules are isolated in the matrix, they are studied by using any of the spectroscopic methods, such as IR, fluorescence, or ESR spectroscopy. The choice of the spectroscopic technique is generally determined by the type of information one is looking for. For example, an analytical exercise is usually undertaken using IR and fluorescence spectroscopy, whereas studies on free radicals (which may be reactive intermediates in a photochemical reaction), are usually performed using ESR and IR spectroscopy.

A MI set-up is shown in Figure 9. Details of the experimental and theoretical aspects of the technique can be found in a number of treatments on this subject.[6–10]

4 APPLICATIONS

4.1 Analytical Spectroscopy

MI has been used as an analytical tool for the detection and quantitative estimations of trace constituents in the atmosphere. As mentioned earlier, this application stems from the fact that the spectral features in MI experiments have small line widths, which makes it possible to resolve closely lying peaks of the different constituents.

Analysis of air samples can be classified into two groups: first, those that monitor the occurrence and concentrations of species that have a long-term effect on the climatic conditions of the earth; second, those that serve to identify the presence of hazardous chemicals that pose a threat to the local environment. Examples of the first class are the study of air samples from the upper troposphere and stratosphere for species such as OH radicals, oxides of nitrogen, sulfur, and halogens, carbonyl halides, halomethanes, and peroxides. It is now realized that these species play an important role in the earth's atmospheric chemistry,[11] and data on such species serve as inputs for atmospheric modeling exercises. Examples of the second class are the study of air samples to monitor the presence of toxic substances such as PAHs and PCBs. In these analyses the air samples are collected from nearer to the earth's surface.

The use of MI spectroscopy for these analytical studies in conjunction with IR, fluorescence, and ESR spectroscopy is now reviewed.

4.1.1 Matrix Isolation Infrared Spectroscopy

First, the use of this technique for studies on upper tropospheric and stratospheric samples is considered, together with problems associated with such studies. Studies on near-ground-level samples are then discussed.

4.1.1.1 Samples from the Upper Troposphere and Stratosphere A number of species that dominate the chemistry in these regions of the atmosphere are labile; it can therefore be extremely difficult to retain the identity of these species till they are brought back to the laboratory for MI studies. Sample collection techniques that retained the identity of the species had therefore to be worked out. In one of the first attempts, a closed cycle refrigerator was transported in a military aircraft to collect air samples at altitudes of around 15 km. The cryostat was then brought back to the laboratory for analysis. Although these first attempts by Snelson,[12,13] did not particularly succeed in obtaining MI data, they nevertheless paved the way for future studies.

The work of Griffith has contributed significantly towards the development of reliable sample collection methods and pretreatment.[14] Associated with the problem of sample collection, the choice of a suitable matrix material is also an important question to be addressed, which is discussed first.

Choice of the Matrix Material. It was recognized that instead of the commonly employed matrix gases, such as Ar or N_2, there was a significant advantage to be derived by using the CO_2 present in the air sample itself as the matrix gas.[14] Carbon dioxide is present in air to the extent of 350 ppmv. Other trace gases are present in the air sample, at concentrations about 1000 times lower. For example, N_2O is present at levels around 300 ppbv. If the CO_2 present in the air is used as the matrix, the sample to matrix ratio, (such as N_2O/CO_2) would be about 1 : 1000, which is typical in MI experiments. If argon had been used as the matrix, as is usually done in MI experiments and the air to argon ratio were 1 : 1000, then the CO_2 to argon ratio would be about $1 : 3 \times 10^6$, and the N_2O to argon ratio would be about $1 : 3 \times 10^9$. These concentrations are extremely low. The choice of CO_2 as the matrix, can therefore be seen to preconcentrate the trace constituents, and is the key to the success of the experiment. The use of CO_2 as the matrix poses problems of interfering IR absorption bands, as CO_2 has strong IR absorptions. However, useful spectral windows are still available with this matrix for use in trace gas analysis. Interestingly, these interfering CO_2 bands have been used to advantage to monitor the thickness of a CO_2 matrix, as discussed later.

Method of Air Sample Collection and Pretreatment. Having decided on the matrix, the next question to be addressed is the method to be adopted to collect the atmospheric samples. It is necessary to be able to collect the air sample and transfer it to the cold finger along with the matrix gas, CO_2, with little loss of sample integrity. Obviously, the best method would be to deposit the air sample directly on to a cold substrate without any sample treatment. If the substrate temperature is maintained at 77 K, only CO_2 (to be used as the matrix) and other nonvolatile constituents would deposit. This procedure was not often successful, particularly for IR studies, because the water normally present in the air samples to the extent of about 1 ppmv led to poor-quality matrices. Furthermore, the condensation of CO_2 in the presence of large quantities of N_2 and O_2, the major constituents of air, also resulted in poor-quality matrices. Hence some amount of sample preparation was necessary before the air sample could be analyzed. After considerable research effort, a two-step process was devised, to prepare dry air samples.[14] The field samples were drawn into a cryosampler, and maintained at liquid nitrogen temperatures. In the laboratory, the sample was evaporated by raising the temperature of the cryosampler carefully to a predetermined value. This temperature of evaporation was determined by the amount of water expected to be present in the field sample. Where the water content was low, such as in upper tropospheric samples, the temperature of evaporation was maintained at about $-50\,^{\circ}C$. However, where the water content was expected to be high, such as in air samples collected nearer to the earth's surface, lower evaporation temperatures were employed or drying agents such as P_2O_5 were used. Care must be exercised during such drying procedures as chemical modifications of the sample species may occur, which is obviously undesirable.

Quantification Procedure. Once the matrix-isolated sample is prepared, quantification of the trace species is done using the celebrated Beer–Lambert law, although the application of the law in this case needs a little elaboration. The Beer–Lambert law can be written in the form (Equation 1)

$$A = \varepsilon \times c \times d \tag{1}$$

where A is the integrated absorbance of the analyte, ε is the integrated absorption coefficient expressed in appropriate units, c is the concentration of the analyte in the sample (which in this case is the concentration of the analyte in the CO_2 matrix), and d is the matrix thickness. In the usual application of the Beer–Lambert law, d represents the path length of the cell containing the sample. This quantity is generally known a priori, or at least remains a constant and can therefore be absorbed into a calibration plot. In the MI experiments, the matrix thickness d is a variable, depending on the rate and duration of deposition, and must therefore be measured for each experiment. The matrix thickness is usually measured using interference techniques with an He–Ne laser. Griffith and group devised a method that used the intensity of CO_2 (matrix) absorption line as a thickness monitor. A series of measurements was made of the

intensities of the different CO_2 absorption lines, together with a direct thickness measurement using the He–Ne interference method. This data enabled the absorption coefficient of the different lines in the CO_2 IR spectrum to be evaluated. Using this data, the thickness d of the CO_2 matrix can be calculated, simply by monitoring the integrated absorbance of any of the CO_2 peaks. A detailed discussion of this method is given by Griffith.[14]

For the estimation of the trace gas concentrations, calibration plots are obtained by taking known concentrations of the trace gas in question, in a CO_2 matrix and measuring the intensities of its absorption line. From the calibration plot, one evaluates the concentration of the trace gas with respect to CO_2 in the atmospheric sample. If the mixing ratio of CO_2 in air is known, the concentration of the trace gas in air can be estimated. The mixing ratio of CO_2 in air can be separately determined by taking the air sample in a gas cell for gas-phase spectral analysis. It must be remembered that the straightforward gas-phase method cannot be applied to trace gas estimation, because the low concentration precludes such a simplistic determination.

Figure 10 provides an example of the analysis of an air sample collected from the upper troposphere from a Lear Jet.[15] Species such as $COCl_2$, OCS, and freons were identified in this sample, in addition to other carbonyl halides such as COFCl and COF_2. These species were found to be present at mixing ratios of approximately 20 pptv in the upper troposphere and increased with height in the lower stratosphere. The identification of the carbonyl halides in the upper tropospheric samples raised the question as to where they were formed. It is now believed that the carbonyl halides are formed by the oxidation of chlorinated hydrocarbons in the troposphere, and then transported to the stratosphere. (A later section describes how the oxidation of halocarbons was studied using MI spectroscopy.) Once in the stratosphere, carbonyl halides dissociate photolytically to yield free chlorine, which plays an active role in ozone depletion. The point that this study drives home is that an oxidative species from the troposphere may have an impact on stratospheric ozone depletion – a useful input for ozone-depletion models. The same study[15] also pointed out that phosgene mixing ratios of 17 pptv in the upper troposphere and 22 pptv in the lower stratosphere were greater than those estimated by models which assumed CCl_4 as the only source. These experiments therefore implied that other sources of phosgene must be present, such as $CHCl_3$, CH_3CCl_3, C_2HCl_3, and C_2Cl_4. This study is a classic example of how detection and quantification of trace species in the atmosphere can serve as useful inputs for atmospheric models.

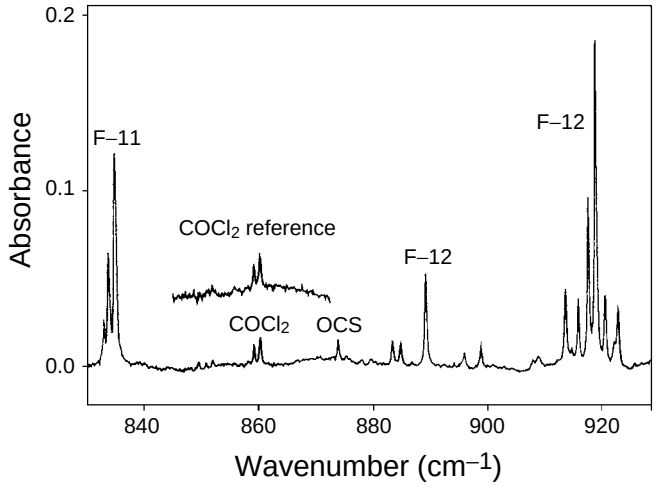

Figure 10 MI/IR spectrum of an air sample collected in the upper troposphere from a Lear Jet. A reference spectrum of $COCl_2$ is also shown for comparison. (D.W.T. Griffith, 'Matrix Isolation Spectroscopy in Atmospheric Chemistry', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, Copyright © 1994 John Wiley & Sons. Reprinted by permission of John Wiley & Sons, Inc.)

4.1.1.2 Samples from Near to the Earth's Surface In the second class of analysis of air samples, the problem is to identify and measure the concentrations of hazardous chemicals, such as PAHs, chlorinated dioxin and dibenzofuran derivatives, and PCBs, that pose a threat to the local environment. In these applications involving usually stable species, GC is used in conjunction with MI spectroscopy. The sample mixture is first subjected to a gas chromatographic separation and the material that is eluted is deposited on a rotating gold-plated cryogenic surface. The samples eluting at different times are deposited on different points on the cryogenic surface. In these experiments, helium–argon mixtures are used as the carrier gas for chromatography. The argon present in this mixture serves as the matrix gas to prepare the matrix-isolated sample. The deposited chromatogram is stored on the cryogenic surface and an IR spectrum then recorded. The technique is referred to as GC/MI/FTIR, where the matrix isolation/Fourier transform infrared (MI/FTIR) detection scheme replaces the light pipe probe used in conventional GC/FTIR. The GC/MI/FTIR method is more sensitive and has a higher spectral resolution than conventional GC/FTIR. Figure 11 compares a spectrum of *N*-nitrosodimethyl amine recorded using GC/MI/FTIR with that recorded using a light-pipe-based absorption.[16] The MI/FTIR detection clearly offers spectra with better signal-to-noise ratio (S/N) than the light-pipe detection. The superior S/N in the MI/FTIR spectra is due to the signal-averaging capabilities offered by this technique. Furthermore, light pipes are usually heated to prevent

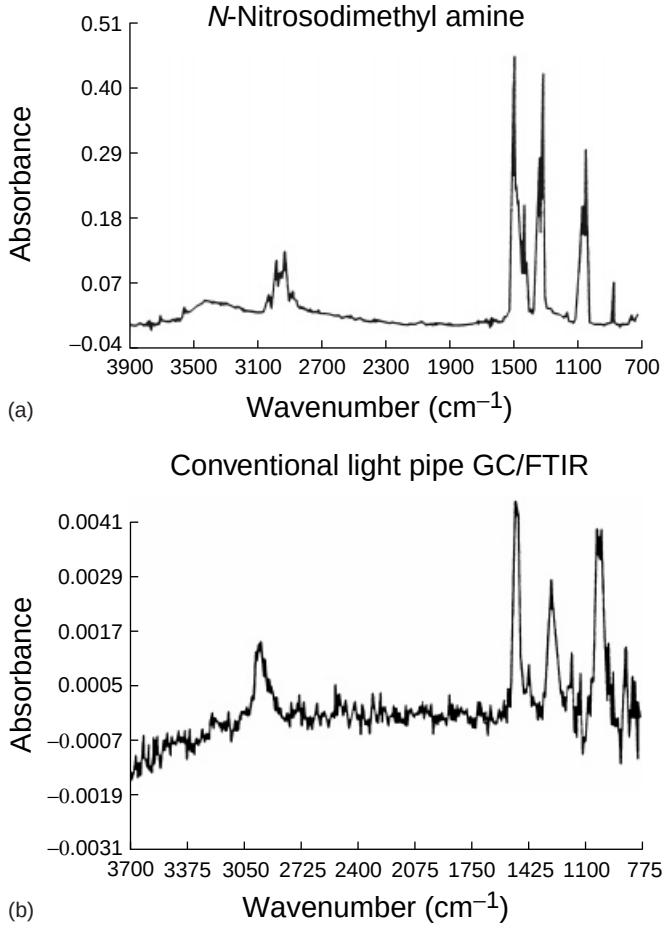

Figure 11 FTIR spectra of *N*-nitrosodimethylamine obtained by MI/GC/FTIR (a) and conventional light-pipe GC/FTIR (b). (Reproduced from the *Journal of Chromatographic Science* by permission of Preston Publications, a Division of Preston Industries, Inc. from Schneider et al.[16])

condensation of the eluted samples. The heating gives rise to an IR background in the light-pipe detection.

The single most attractive feature of GC/MI/FTIR is its ability to distinguish between isomers.[17,18] This advantage is of paramount importance, as it has been shown that biological toxicity is often highly isomer specific. For example, amongst the polychlorodibenzo-*p*-dioxins, the 2,3,7,8-tetrachloro isomer is the most toxic.[17] Similarly, benzo[*a*]pyrene (B[*a*]P) is a potent carcinogen, whereas its isomer benzo[*e*]pyrene is significantly less hazardous. From the point of view of environmental monitoring and waste management, it is therefore essential that the analytical technique possess the ability to resolve isomers. Figure 12 (trace A) shows the GC/MI/FTIR spectrum of a fraction that eluted out of the GC column at a certain time (31.58 min).[17] This spectrum did not find a match in the MI/IR spectral library, and it was suspected that the eluted fraction may contain a mixture of an isomer pair. It is clear from traces B and C, which

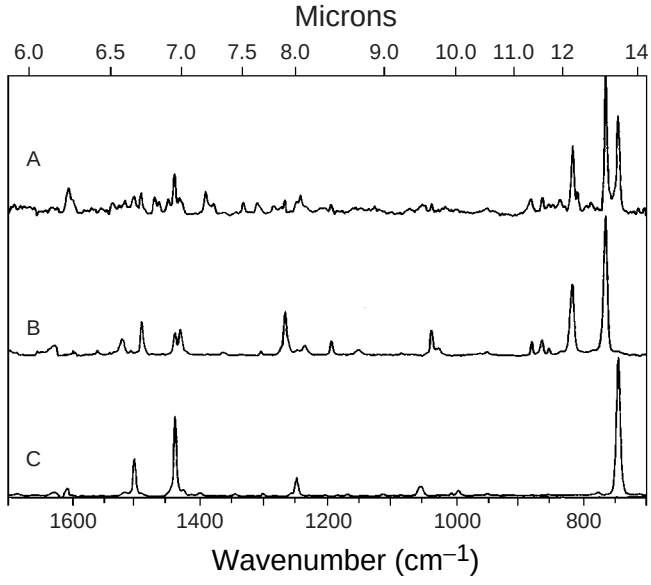

Figure 12 MI/IR spectrum (trace A) of component of a filter extract with a retention time of 31.58 min, (B) chrysene reference spectrum, (C) triphenylene reference spectrum. (Reproduced by permission of Society for Applied Spectroscopy from J.W. Childers, N.K. Wilson, R.K. Barbour.[17])

are reference spectra of chrysene and triphenylene, that the eluted fraction contained both these isomers. The example clearly demonstrates that compounds eluting simultaneously can still be differentiated using MI/FTIR. In this regard, GC/MI/FTIR even scores over mass spectroscopic detection, which enjoys superior DLs, but is not always capable of resolving isomeric species. For example, electron impact ionization mass spectra of chrysene and triphenylene yielded identical parent ions, similar fragmentation patterns and only subtle differences between the two spectra, making differentiation between the two isomers difficult. Similar resolving capabilities were also demonstrated for the B[*a*]P/benzo[*e*]pyrene pair, and benzofluoranthene isomers.[17]

4.1.2 Matrix Isolation Fluorescence Spectroscopy

Fluorescence spectroscopy has been extensively used as an analytical tool for condensed- and gas-phase samples. Compared with absorption spectroscopy, fluorescence spectroscopy has superior DLs, which has made it the technique of choice for trace level estimations.

Obtaining sharp absorption and emission lines is again a principal motivation for doing fluorescence spectroscopy of matrix-isolated samples. However, unlike in IR spectroscopy, where inert gases and nitrogen are excellent matrices, the choice of the matrix is not so straightforward in fluorescence. The same considerations as those discussed earlier for sharpening of spectral lines apply here too, except that the site effects are extremely important

in this case. Molecules trapped in different environments can display a multitude of spectral features. If a matrix can be chosen in which the molecules can be trapped in a few well-defined sites, spectral simplification may occur, leading to sharp spectral features, referred to as the Shpol'skii effect.[19] Hydrocarbons, such as hexane and octane, have often proved to be successful Shpol'skii matrices. It has been realized that the best matrices are those in which one of the dimensions of the solute molecules is similar to the length of the hydrocarbon chain. In such systems, the solute preferentially occupies a well-defined site, leading to line narrowing. Figure 13 shows the laser-induced fluorescence spectra of the environmentally important B[*a*]P trapped in heptane and octane matrices.[20] It can be seen that the spectrum in the octane matrix is sharper than that in the heptane matrix, which displays a multiplet structure. The sharp spectral feature in the octane matrix may be useful in characterizing the trapped species for analytical applications.

Furthermore, the trapped species exhibit sharp features in absorption too, which allows one to selectively excite a particular species in a well-defined site, using a narrow-band excitation source, such as a dye laser. This aspect was convincingly presented by D'Silva and Fassel.[21] They showed that each component in a multicomponent mixture can be selectively excited to obtain a fluorescence spectrum corresponding to each of the trapped species. Figure 14 shows the spectra of four different PAHs, obtained by a selective excitation of each of the species in the mixture, using dye laser excitation. The sample was a solvent-refined coal liquid sample trapped in an *n*-octane matrix. Occasionally, a partial overlap of the absorption lines may lead to spectra of more than one compound

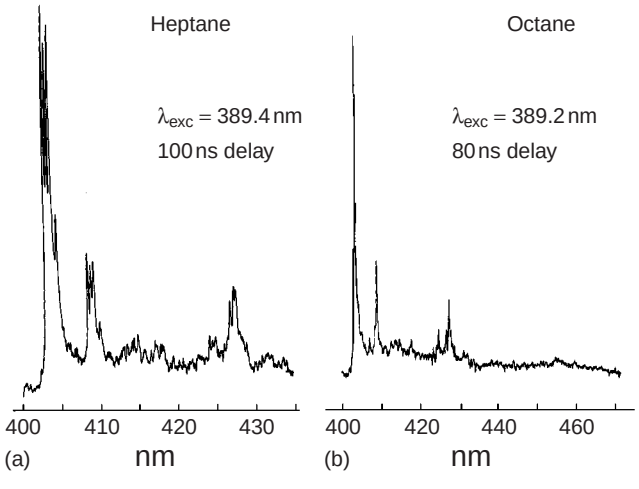

Figure 13 Laser-induced fluorescence spectra of B[*a*]P in a solid solvent-refined coal sample in (a) heptane and (b) octane matrices at 15 K. (Reproduced with permission from M.B. Perry, E.L. Wehry, G. Mamantov,[20] Copyright 1989 American Chemical Society.)

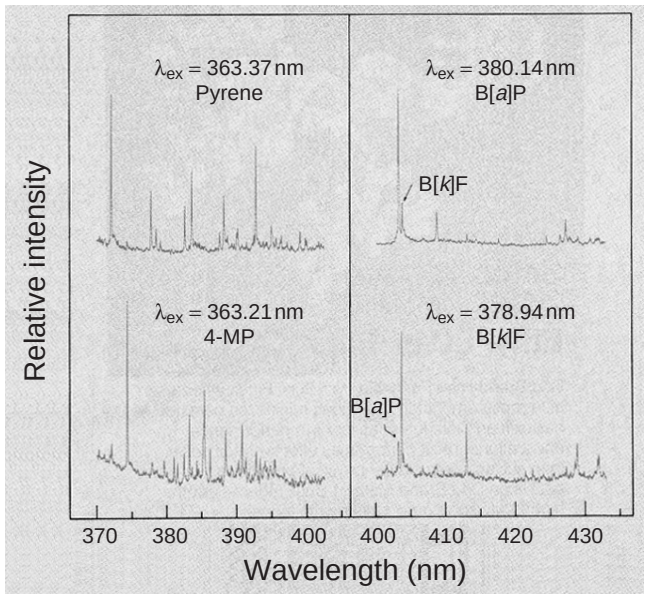

Figure 14 Site-selective laser-excited fluorescence spectra of pyrene, 4-methylpyrene (4-MP), B[*a*]P, and benzo[*k*]fluoranthene (B[*k*]F) in a solvent-refined coal sample. (Reproduced with permission from A.P. D'Silva, V.A. Fassel,[21] Copyright 1984 American Chemical Society.)

being observed. For example, at the excitation wavelength (380.14 nm) where B[*a*]P was excited, the feature for B[*k*]F was observed; likewise the B[*k*]F fluorescence spectra, excited at 378.94 nm showed features of B[*a*]P. However, as can be seen from the spectra, the emission lines are sufficiently sharp to allow for identification of the different components in the mixture. Similarly, all the components in a mixture consisting of six mono-, two di- and three tri-alkylated benz[*a*]anthracenes were shown to be unequivocally identified.[21] The technique can therefore be seen to yield isomer-specific detection of samples, essential in the characterization of hazardous wastes.

However, a notable disadvantage of the Shpol'skii experiment is that the choice of matrix is not always obvious. A given compound or a class of compounds may have only a few matrices where the Shpol'skii effect may be observed. Secondly, some compounds may photodecompose in a Shpol'skii matrix. However, wherever applicable, MI fluorescence spectroscopy, utilizing laser-excited site selection (LESS), has been extremely useful in toxic waste evaluation, prompting the remark that one can "Do more with LESS."[21]

4.1.3 Matrix Isolation/Electron Spin Resonance Spectroscopy

As free radicals play an important role in atmospheric processes, it can only be expected that ESR spectroscopy would be a useful tool for these studies. The reactive

free radicals can be trapped in the inert matrix and then studied using ESR.

The matrix is deposited on sapphire or copper block substrates and the cold finger then introduced into the microwave cavity for ESR measurements. The unstable radical species survive in the inert matrix, on which ESR studies are performed. Details of the matrix isolation/electron spin resonance (MI/ESR) set-up and experimental procedures can be found in treatments on this subject.[14,22] In experiments where MI/ESR is used for atmospheric studies, CO_2 again turns out to be a useful matrix. Water (or D_2O), which are avoided in IR experiments, serve as a useful matrix for ESR studies. Oxygen, being paramagnetic, must be avoided in the matrix, or else it can interact with the deposited free radicals causing significant line broadening, something which the MI technique set out to eliminate in the first place! The first attempts of Snelson[12] did not yield useful spectra, essentially because the sampling method adopted then, which was to simply freeze out the air sample on to the matrix substrate, incorporated O_2 in the matrix.

Mihelcic et al. obtained MI/ESR spectra of free radicals by trapping them in CO_2 and H_2O matrices.[23,24] Oxygen was eliminated by maintaining the cold finger at liquid-nitrogen temperatures. In these experiments, care was taken to minimize the collisions of the sampled radicals with any of the surfaces of the sample compartment other than, of course, the cold finger. This was achieved by having the sampled air form a gas jet that concentrates around the cold finger. In early experiments,[23] the CO_2 and H_2O present in the sampled air were used as the matrix. However, it was soon realized that the trapping efficiency was low in such experiments and depended on the H_2O concentration, which was variable in the troposphere. They ensured a stable and a large collection efficiency by deliberately adding water during the matrix deposition. In later experiments, they showed that the use of D_2O yielded ESR spectra with smaller linewidths than those obtained with an H_2O matrix.[24]

Figure 15 shows the ESR spectra of free radicals, NO_2, NO_3, HO_2, and CH_3COO_2, trapped in H_2O and D_2O matrices.[24] The spectra in the D_2O matrices are

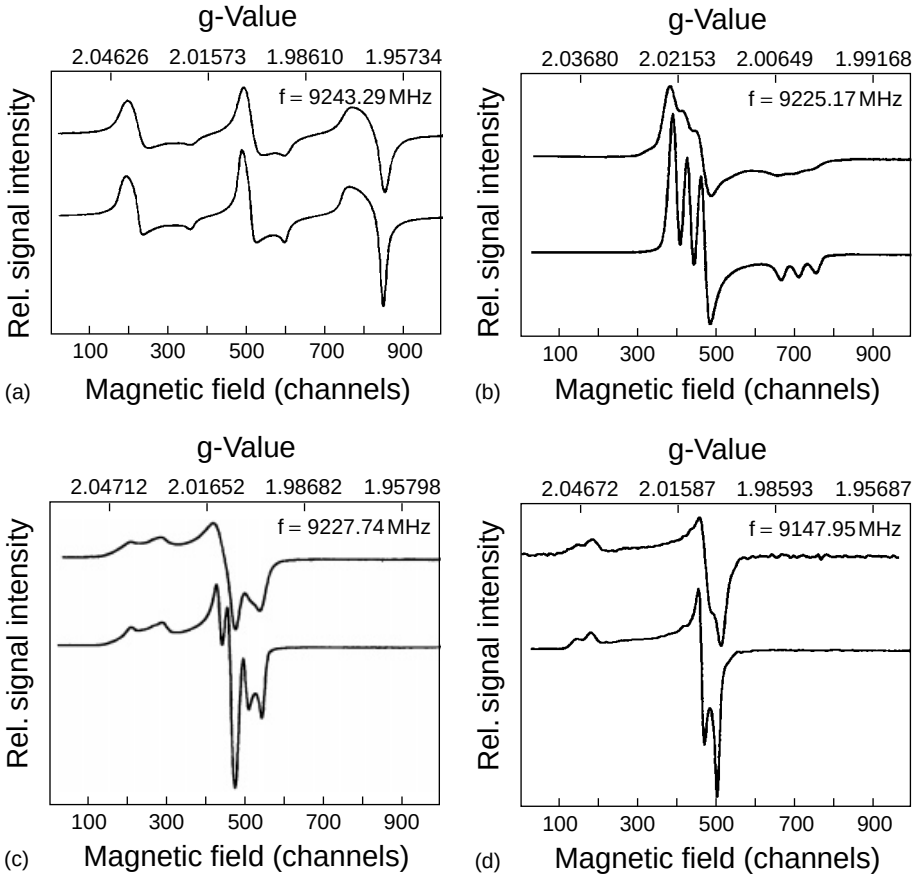

Figure 15 MI/ESR spectrum of common atmospheric radicals trapped in H_2O (upper trace) and D_2O (lower trace) matrices at 77 K: (a) NO_2, (b) NO_3, (c) HO_2, (d) CH_3COO_2. (D. Mihelcic, A. Volz-Thomas, H.W. Patz, D. Kley, M. Mihelcic, 'Numerical Analysis of Electron Spin Resonance Spectra from Atmospheric Samples', *Journal of Atmospheric Chemistry*, **11**, 275 (1990), Figure 1, with kind permission from Kluwer Academic Publishers.)

clearly sharper, resulting in greater sensitivity, and also allow for the identification of those radicals that show a characteristic hyperfine structure.

Using the experimental procedures described above, Mihelcic et al. studied the diurnal variation of NO_2 and RO_2 (total peroxide concentration). Their studies showed that the concentrations of NO_2 and RO_2 were anticorrelated and that the nighttime RO_2 concentration hardly differed from those obtained during the day. In another experiment, they also measured the vertical concentration profiles of NO_2 and RO_2 over altitudes ranging from 1 to 6 km. Using a least-squares spectral analysis algorithm, they showed that small amounts of HO_2 could be detected in the presence of a manyfold excess of NO_2 and other peroxy radicals. HO_2 and CH_3COO_2 radicals in tropospheric samples were shown to be detected at concentrations up to 40 pptv.

Being reactive species, free radicals play a very important role in the chemistry of the atmosphere and the MI/ESR technique provides a powerful technique with which to study them.

4.2 Study of Reaction Intermediates

To understand the chemistry and the various dynamic processes in the atmosphere, it is important not only to identify and quantify the different species present, but also to elucidate the important chemical reactions that they manifest. Many atmospheric reactions are photochemically driven processes involving reactive species. MI is a technique ideally suited for the study of such process as the following examples will show. The examples cited here are only representative and are not an exhaustive compilation.

As an example of the use of MI for the study of reactive species, some of the work pertaining to stratospheric ozone depletion is described. In this context, it is relevant to briefly review some of the chemistry involving ozone loss due to halocarbons.

Depletion of stratospheric ozone has been linked to the introduction of synthetic CFCs into the atmosphere, such as CF_2Cl_2 and $CFCl_3$. These CFCs are photolyzed in the upper regions of the atmosphere at altitudes of 20–40 km, leading to the formation of atomic chlorine (Equations 2 and 3):[25]

$$CF_2Cl_2 \longrightarrow CF_2Cl + Cl \quad (2)$$

$$CFCl_3 \longrightarrow CFCl_2 + Cl \quad (3)$$

An extensive catalytic chain reaction then occurs which depletes the O_3 concentration, (Equations 4 and 5):[26]

$$Cl + O_3 \longrightarrow ClO + O_2 \quad (4)$$

$$ClO + O \longrightarrow Cl + O_2 \quad (5)$$

Equation (5) regenerates atomic chlorine and the cycle is resumed. (The atomic oxygen, involved in Equation (5), is a trace constituent in the stratosphere and is produced by the photodecomposition of molecular oxygen.) The (ClO, Cl) chain therefore leads to a net ozone-loss reaction (Equation 6):

$$O_3 + O \longrightarrow 2O_2 \quad (6)$$

An alternative pathway is Equation (4) followed by Equation (7):

$$ClO + ClO \longrightarrow 2Cl + O_2 \quad (7)$$

Here again a (ClO,Cl) chain is involved and the net reaction is given by Equation (8):

$$2O_3 \longrightarrow 3O_2 \quad (8)$$

(Only the overall stoichiometry has been shown in Equation 7. As discussed later, the reaction actually proceeds through the intermediacy of Cl_2O_2 and ClO_2.) In this discussion, only reactions involving Cl have been shown; in reality a number of species such as OH, Br, and NO undergo reactions similar to Equations (4), (5), and (7), and play a significant role in ozone depletion.

From the above reaction scheme, it is clear that ClO is an important molecule in ozone depletion processes. Measurements have shown an increase in the ClO concentration over the polar stratospheric regions. It is therefore important to study the reactions of ClO with other species present in the atmosphere, so that the fate of the halogen monoxides in the atmosphere can be understood.

One such reaction is that between ClO and NO_2 to form chlorine nitrate. This reaction is important where the NO_2 concentration is high (Equation 9):

$$ClO + NO_2 + M \longrightarrow ClONO_2 \quad (9)$$

(In this reaction, M is a third body that removes the energy of the complexed species and stabilizes it.) The chlorine nitrate is expected to photolyze only slowly in the UV. This reaction would tie up the ClO, by converting it into the less active $ClONO_2$, referred to as a reservoir molecule. Equation (9) would therefore lead to a decrease in ozone depletion. However, there were speculations that this reaction may also produce a photolytically active *isomer* of chlorine nitrate, that may counter the reservoir effect. It was therefore important to study whether isomers of chlorine nitrate are produced by Equation (9). Burrows et al.[27] addressed this problem using MI/IR spectroscopy, where they studied the reaction between ClO and NO_2 by photolyzing flowing mixtures of Cl_2, Cl_2O, and NO_2 and trapping the reaction product in an

N_2 matrix. The following Equations (10–12) were driven during photolysis:

$$Cl_2 + h\nu \longrightarrow 2Cl \quad (10)$$

$$Cl + Cl_2O \longrightarrow ClO + Cl_2 \quad (11)$$

$$ClO + NO_2 + M \longrightarrow \text{product} \quad (12)$$

The find product formed in these reactions was then trapped in a nitrogen matrix and the IR spectrum of this product was compared with synthetically produced $ClONO_2$. It was observed that the product spectrum agreed well with the spectrum of synthetically produced $ClONO_2$ and no extra features were seen that could be assigned to an isomer of chlorine nitrate. It was therefore concluded that Equations (9) and (12) do not give rise to any isomer of $ClONO_2$. The hypothesis of the existence of a photochemically labile isomer of chlorine nitrate that could alter the ozone depletion scenario was therefore discarded.

In environments where NO_x abundances are low, the formation of chlorine nitrate (Equation 9) may be unimportant and alternate reaction pathways for ClO may dominate. Where the ClO concentrations are high, the self-reaction (Equation 7), may be significant. Such conditions appear to exist in the stratosphere over Antarctica. Although the net products are Cl and O_2, Equation (7) is believed to occur through the intermediacy of Cl_2O_2 and ClO_2 (Equations 13–15):

$$ClO + ClO + M \longrightarrow Cl_2O_2 \quad (13)$$

$$Cl_2O_2 + h\nu \longrightarrow Cl + ClOO \quad (14)$$

$$ClOO \longrightarrow Cl + O_2 \quad (15)$$

The species produced in the above reactions, Cl_2O_2 and ClO_2, have been the subject of a number of MI studies.

The structure, spectroscopy and reactions of the dimer, Cl_2O_2, have been studied in great detail. Following these studies, it was found that the ClO dimer exists in three isomeric forms (structures **(1–3)**):[28–31]

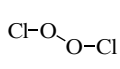

Dichlorine peroxide
(1)

Chloryl chloride
(2)

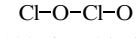

Chlorine chlorite
(3)

The most stable of these forms were structures **(1)** and **(2)**, which were of comparable energy, whereas structure **(3)** was the least stable. Johnsson et al.,[28] Jacobs et al.,[29] and Cheng and Lee[30] trapped the ClO dimers in an argon matrix and showed that these forms can photochemically interconvert before eventually decomposing to yield the products, O_2 and Cl as given in Equation (7). These studies also established the different photolysis channels of these species, information that was required to establish the fate of the ClO dimer in the atmosphere.

Likewise, ClO_2 has also been the subject of MI investigations.[32,33] Two different structures are possible for this species, OClO and ClOO. Both these isomers play an important role in stratospheric chemistry. The species ClOO is believed to be the primary photolysis product of ClOOCl, which then eventually dissociates to yield Cl and O_2, whereas OClO is suggested as an intermediate in reactions involving ClO and BrO. Muller and Willner[32] isolated both these isomers of ClO_2 in argon and neon matrices and studied their vibrational and electronic spectra. They also showed that OClO, when irradiated with a medium-pressure Hg vapor lamp ($\lambda \geq 395$ nm), converted to ClOO. The MI studies described above[28–33] therefore revealed the intricate details of Equation (7).

It may be recalled that an analysis of the upper troposphere using MI spectroscopy indicated the presence of carbonyl halides (section 4.1.1). The carbonyl halides were believed to be formed by the oxidation of halocarbons in the troposphere. In an effort to understand this oxidative process, MI studies were conducted to probe the reaction of oxygen atoms with halocarbons. Schriver et al.[34,35] have studied the reactions of compounds such as CCl_3F (Freon 11) and CH_3Cl with oxygen atoms produced by a photolytic decomposition of ozone. Among other products, they observed the formation of carbonyl halides such as COFCl, which was one of the species identified by Wilson et al.[15] in their analysis of the upper tropospheric samples. The experiments of Schriver et al. mimicked the oxidation of the halocarbons in the troposphere.

MI spectroscopy has been used in elucidating a number of these reactions to understand the atmospheric processes.

5 CONCLUSIONS

The technique of MI in conjunction with IR, fluorescence, and ESR has been used to address problems in atmospheric chemistry. Trace constituents in tropospheric and stratospheric samples have been measured, information that has proved useful in atmospheric modeling. Air samples from nearer to the earth's surface have also been analyzed to estimate chemically hazardous materials that can potentially affect the local ecology of a place. The technique of GC/MI/FTIR used for these analyses is particularly useful for isomer-specific detection of chemically hazardous materials. In these applications, GC/MI/FTIR even scores over other more sensitive techniques, such as MS. MI has also been used to study the structure,

spectroscopy, and chemistry of reactive species that play an important role in the chemistry of the atmosphere.

ABBREVIATIONS AND ACRONYMS

B[*a*]P	Benzo[*a*]pyrene
B[*k*]F	Benzo[*k*]fluoranthene
CFC	Chlorofluorocarbon
DL	Detection Limit
DOAS	Differential Optical Absorption Spectroscopy
ESR	Electron Spin Resonance
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
IR	Infrared
LESS	Laser-excited Site Selection
MI	Matrix Isolation
MI/ESR	Matrix Isolation/Electron Spin Resonance
MI/FTIR	Matrix Isolation/Fourier Transform Infrared
MI/IR	Matrix Isolation Infrared
MS	Mass Spectroscopy
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
S/N	Signal-to-noise
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
4-MP	4-Methylpyrene

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Differential Optical Absorption Spectroscopy, Air Monitoring by • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Laser-induced Breakdown Spectroscopy, Elemental Analysis • Laser Mass Spectrometry in Trace Analysis • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments • Photoacoustic Spectroscopy in Trace Gas Monitoring • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Fluorescence Lifetime Measurements, Applications of

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Gas Chromatography/Infrared Spectroscopy • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Electron Spin Resonance Spectroscopy

REFERENCES

1. J.M. Wallace, P.V. Hobbs, *Atmospheric Science–An Introductory Survey*, Academic Press, San Diego, 1977.
2. M.W. Sigrist (ed.), *Air Monitoring by Spectroscopic Techniques*, John Wiley & Sons, New York, 1994.
3. W. Demtroder, *Laser Spectroscopy, Basic Concepts and Instrumentation*, Springer-Verlag, Berlin, 1982.
4. E. Whittle, D.A. Dows, G.C. Pimentel, 'Matrix Isolation Method for the Experimental Study of Unstable Species', *J. Chem. Phys.*, **22**, 1943 (1954).
5. M.M. Rochkind, 'Quantitative Infrared Analysis of Mixtures of Isotopically Labelled Gases', *Anal. Chem.*, **40**, 762–768 (1968).
6. H.E. Hallam (ed.), *Vibrational Spectroscopy of Trapped Species*, John Wiley & Sons, London, 1973.
7. L. Andrews, M. Moskovits (eds.), *Chemistry and Physics of Matrix-isolated Species*, Elsevier Science, Amsterdam, 1989.
8. A.J. Barnes, W.J. Orville-Thomas, A. Muller, R. Gaufres (eds.), *Matrix Isolation Spectroscopy*, D. Reidel, Dordrecht, 1981.
9. A.J. Rest, 'Generation and Trapping of Unstable Solutes in Low Temperature Matrices', in *Matrix Isolation Spectroscopy*, eds. A.J. Barnes, W.J. Orville-Thomas, A. Muller, R. Gaufres, D. Reidel, Dordrecht, Chapter 9, 187–205, 1981.
10. E.L. Wehry, G. Mamantov, 'Matrix Isolation Molecular Spectrometry in Chemical Analysis', *Prog. Anal. Spectrosc.*, **10**, 507–527 (1987).
11. J.G. Calvert (ed.), *The Chemistry of the Atmosphere: Its Impact on Global Change*, Blackwell Scientific, London, 1994.
12. A. Snelson, 'Chemical Composition of the Atmosphere at 40 000–65 000 Feet using a Modified Matrix Isolation Technique', Air Force Off. Sci. Res. TR-74-1727. IIT Research Institute, Chicago, IL, 1974.
13. A. Snelson, 'Chemical Composition of the Atmosphere at 10 000–40 000 Feet Using a Modified Matrix Isolation Technique', Air Force Off. Sci. Res. TR-75-1171. IIT Research Institute, Chicago, IL, 1975.

14. D.W.T. Griffith, 'Matrix Isolation Spectroscopy in Atmospheric Chemistry', in *Air Monitoring by Spectroscopic Techniques*, ed. M.W. Sigrist, John Wiley & Sons, New York, Chapter 7, 471–514, 1994.
15. S.R. Wilson, P.J. Crutzen, G. Schuster, D.W.T. Griffith, G. Helas, 'Phosgene Measurements in the Upper Troposphere and Lower Stratosphere', *Nature*, **334**, 689–691 (1988).
16. J.F. Schneider, J.C. Demirgian, J.C. Stickler, 'A Comparison of GC/IR Interfaces: The Light Pipe Versus Matrix Isolation', *J.Chromotogr. Sci.*, **24**, 330–335 (1986).
17. J.W. Childers, N.K. Wilson, R.K. Barbour, 'Gas Chromatography/Matrix Isolation-Infrared Spectrometry for the Identification of Polycyclic Aromatic Hydrocarbons in Urban Air Particulate Matter', *Appl. Spectrosc.*, **43**, 1344–1349 (1989).
18. C.J. Wurrey, B.J. Fairless, H.E. Kimball, 'Gas Chromatographic/Matrix Isolation/Fourier Transform Infrared Spectra of the Laterally Chlorinated Dibenzo-*p*-Dioxins and Dibenzofurans', *Appl. Spectrosc.*, **43**, 1317–1324 (1989).
19. E.L. Wehry, G. Mamantov, 'Low Temperature Fluorimetric Techniques and their Application to Analytical Chemistry', in *Modern Fluorescence Spectroscopy*, ed. E.L. Wehry, Plenum Press, New York, 193–250, Vol. 4, 1979.
20. M.B. Perry, E.L. Wehry, G. Mamantov, 'Determination of Polycyclic Aromatic Hydrocarbons in Unfractionated Solid Solvent-refined Coal by Matrix Isolation Fluorescence Spectrometry', *Anal. Chem.*, **55**, 1893–1896 (1983).
21. A.P. D'Silva, V.A. Fassel, 'Laser Excited Shpol'skii Spectrometry', *Anal. Chem.*, **56**, 985A–1000A (1984).
22. L.B. Knight, Jr, 'Production Methods for ESR Studies of Neutral and Charged Radicals', in *Chemistry and Physics of Matrix Isolated Species*, eds. L. Andrews, M. Moskovits, Elsevier Science, Amsterdam, 167–201, 1989.
23. D. Mihelcic, P. Musgen, D.H. Ehhalt, 'An Improved Method of Measuring Tropospheric NO_2 and RO_2 by Matrix Isolation and Electron Spin Resonance', *J. Atmos. Chem.*, **3**, 341–361 (1985).
24. D. Mihelcic, A. Volz-Thomas, H.W. Patz, D. Kley, M. Mihelcic, 'Numerical Analysis of ESR Spectra from Atmospheric Samples', *J. Atmos. Chem.*, **11**, 271–297 (1990).
25. M.J. Molina, F.S. Rowland, 'Stratospheric Sink for Chlorofluoromethanes: Chlorine Atom Catalysed Destruction of Ozone', *Nature*, **249**, 810–812 (1974).
26. M.B. McElroy, R.J. Salawitch, S.C. Wofsy, J.A. Logan, 'Reductions of Antarctic Ozone due to Synergistic Interactions of Chlorine and Bromine', *Nature*, **321**, 759–762 (1986).
27. J.P. Burrows, D.W.T. Griffith, G.K. Moortgat, G.S. Tyndall, 'Matrix Isolation Fourier Transform Infrared Study of the Products of the Reaction between ClO and NO_2', *J. Phys. Chem.*, **89**, 266–271 (1985).
28. K. Johnsson, A. Engdahl, J. Kolm, J. Nieminen, B. Nelander, 'The ClOClO, BrOClO and IOClO Molecules and their Photoisomerization. A Matrix Isolation Study', *J. Phys. Chem.*, **99**, 3902–3904 (1995).
29. J. Jacobs, M. Kronberg, H.S.P. Muller, H. Willner, 'An Experimental Study on the Photochemistry and Vibrational Spectroscopy of Three Isomers of Cl_2O_2 Isolated in Cryogenic Matrices', *J. Am. Chem. Soc.*, **116**, 1106–1114 (1994).
30. B.-M. Cheng, Y.-P. Lee, 'Production and Trapping of Gaseous Dimeric ClO: The Infrared Spectrum of Chlorine Peroxide (ClOOCl) in Solid Argon', *J. Chem. Phys.*, **90**, 5930–5935 (1989).
31. T.J. Lee, C.M. Rohlfing, J.E. Rice, 'An Extensive ab initio Study of the Structures, Vibrational Spectra, Quadratic Force Fields, and Relative Energetics of Three Isomers of Cl_2O_2', *J. Chem. Phys.*, **97**, 6593–6605 (1992).
32. H.S.P. Muller, H. Willner, 'Vibrational and Electronic Spectra of Chlorine Dioxide, OClO, and Chlorine Superoxide, ClOO Isolated in Cryogenic Matrices', *J. Phys. Chem.*, **97**, 10589–10598 (1993).
33. A. Arkell, I. Schwawger, 'Matrix Infrared Study of the ClOO Radical', *J. Am. Chem. Soc.*, **89**, 5999–6006 (1967).
34. L. Schriver, O. Abdelaoui, A. Schriver, 'Atmospheric Cryochemistry: Oxygen Atom Reaction with the Fluorocarbon Freon 11 in Matrices. FTIR Spectra of Isolated COFCl and COFCl : Cl_2 Complex in Solid Argon', *J. Phys. Chem.*, **96**, 8069–8073 (1992).
35. L. Schriver-Mazzuoli, A. Schriver, Y. Hannachi, 'Matrix Infrared Study of O-Initiated Atomic Oxidation of CH_3Cl: Identification of the Triplet $CH_3Cl\cdots O$ Complex', *J. Phys. Chem. A*, **102**, 10221–10229 (1998).

Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

F. Rathgeb and G. Gauglitz
Eberhard-Karls-Universität Tübingen, Tübingen, Germany

1 Introduction **1**
1.1 General Properties of Optical Sensors 1
1.2 Classification of Optical Sensors 2
2 Fundamentals **3**
2.1 Aspects of Sensing in the Evanescent Field Region 3
2.2 Sensitive Layers 3
3 Reflectometric Interference Spectroscopy **4**
3.1 Introduction 4
3.2 Principle of Reflectometric Interference Spectroscopy 5
3.3 Applications 5
4 Fiber Optic Sensors **6**
4.1 Extrinsic Sensing Applications 6
4.2 Intrinsic Sensing Applications 7
5 Planar Waveguide Sensors **7**
5.1 Introduction 7
5.2 Applications 8
6 Surface Plasmon Resonance Spectroscopy **9**
6.1 Introduction 9
6.2 Principle of Surface Plasmon Resonance 9
6.3 Measurement Set-ups 9
6.4 Applications 10
7 Summary **11**
Abbreviations and Acronyms **11**
Related Articles **11**
References **11**

Current concepts for optical gas sensors are reviewed. Different approaches for gas sensing such as intrinsic and extrinsic, direct and indirect sensing were taken into consideration. For each sensing approach reviewed here, a short principle description and recent applications are given. Several aspects governing the enrichment of gaseous analytes in sensitive layers are also discussed briefly.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The requirements of industry and legislation take advantage of the qualities of chemical sensors and especially fiber optics and/or planar waveguide transducers, using their capability to continuously report changes of a physical parameter or the concentration of an analyte and the sensor's ability to work reversibly. Industry discovered that chemosensors could be used for applications in process control, work place monitoring, and to follow legislation which requires threshold values for air, water, and industrial production to be met. Thus, demand for sensors has increased. These devices combine the properties of physical sensors (transducer) with the advantages and problems of chemical sensitive layers, by being responsible for the special properties of chemical sensors.

In the following, the properties of such optical sensors are discussed, they are classified according to certain characteristics and then the peculiarities of chemical sensing layers are considered. Since the definitions and general descriptions of sensors have been given in many reviews and books in recent years, just some of the special considerations regarding the optical chemical sensors applied to detect gases are given here. The main topics are the discussion of the chemical sensing layers and their present status of suitability, recent applications, and some trends in development and application to real-life situations in addition to modern analytical tools.

1.1 General Properties of Optical Sensors

An ideal sensor is sensitive, allows remote sensing and analysis in real-time, is economically priced, robust, reliable, cheap in maintenance, miniaturized, and simple to handle. Furthermore an ideal sensor is selective, has a high operational stability, has a wide dynamic range, and shows no drift of the baseline. Whereas the first of the mentioned prerequisites of an ideal sensor nowadays can be fulfilled by several transduction principles, the latter represent the remaining challenges in the field of chemical sensing. Most of these named restrictions are caused by the limited properties of the sensing element or sensitive layer.

One approach to overcome the limited selectivity of chemosensors is the application of chemometrics or pattern recognition.[1,2] However, these techniques often require defined boundary conditions, which unfortunately are seldom found outside the researcher's laboratory. This explains why, besides the optimization of the transduction principles, great efforts are taken in order to find the ultimate sensitive layers.

Light interacting with matter can yield a lot of different information about conditions of matter at high speed. Absorbance, luminescence, polarization,

interference, time, and wavelength (frequency), to name some examples, are measurements which can be interrogated by optical techniques. Optical sensors are generating increased interest due to several advantages such as remote sensing capabilities, no interference by electric or magnetic fields, and use in a wide area of applications, e.g. medical diagnostics.[3–5] Due to this growing interest in optical sensors, a large number of publications and review articles have been published. Therefore it is not possible, or the intention, to give here a complete overview of this wide and growing area of fiber optic and waveguide gas sensors. Accordingly, this article is restricted to certain aspects and new developments have not been reviewed in detail up to now. Therefore we want to emphasize that nonquotation of a paper does not give any judgement about its quality. For a detailed survey, one should refer to the volumes of the monograph *Chemical and Biochemical Sensors* edited by Göpel et al.,[6] which also gives a detailed insight into optical sensors.[6] A review covering chemical sensors from 1 December 1993 to 30 November 1997, with more than 900 references, was published recently.[7] Fiber optic sensors (FOS) are well represented by several monographs[8–12] and reviews.[13–17] Additionally, a detailed overview of infrared (IR) FOS is given by Saito and Kikuchi.[18] The development of liquid core optical fibers is discussed by Altkorn et al.[19]

Thus we will focus on certain techniques of optical gas sensors such as devices like surface plasmon resonance (SPR), planar waveguide sensors, FOS, and reflectometric interference spectroscopy (RIfS). Only recent developments in gas sensing will be discussed here.

1.2 Classification of Optical Sensors

Besides the technique (SPR, FOS), optical sensors can be classified by their waveguide optics and the analytical way to detect the analyte and the modulated parameter of light leading to a three-dimensional classification.[16] Whereas two dimensions, namely the detected parameter of radiation and the waveguide optics, are defined by the transduction principle and set-up of the device, the third axis is defined by the interaction the analyte experiences in the sensitive layer (Figure 1). We will not follow the scheme represented here strictly, but we consider it as a useful tool to classify the different methods, approaches and optics in the overall picture of optical sensors. A more detailed discussion of the following explained terms, which are used to classify the different types of FOS, is given by Boisdé and Harmer[9] and Potyrailo et al.[16]

In general, fiber optical and planar waveguide sensors use the same principle; they guide electromagnetic radiation by internal reflectance. They differ in geometry and in fields of application. Therefore, in the following,

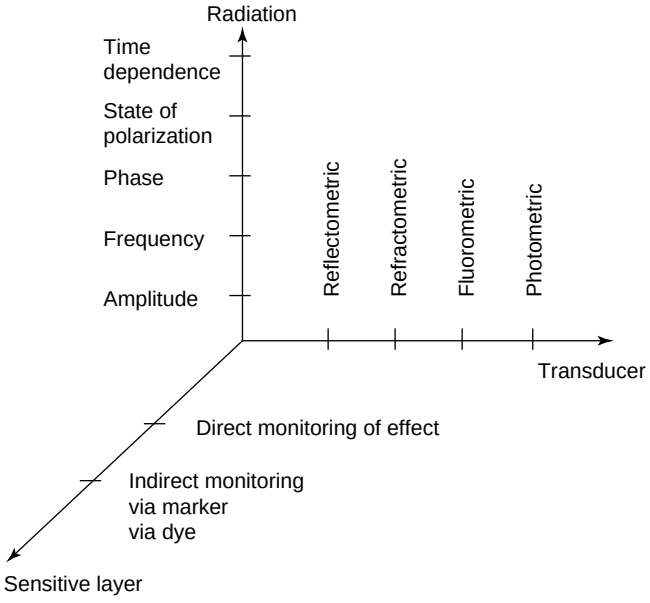

Figure 1 Classification of optical sensor systems with respect to the effect on or by radiation, the transduction principle and the properties of the sensitive layer.

fiber optics and planar waveguide structures will be used synonymously and called waveguides in general. The first are in many cases preferably used to just transport radiation. Both can be applied to monitor influences on this guided radiation, in addition. Accordingly, in the figure both types are marked on one of the axes. Additionally, optical sensors can further be distinguished by the terms *extrinsic* or *intrinsic*. In extrinsic sensors, radiation is just transported by the waveguide, is modulated outside the waveguide, and will re-enter the waveguide to be re-transported to the detector. The waveguide is simply used to transport the light to the sampling point and from there to the detector. In intrinsic optical sensors the light is not intended to leave the light-guiding structure to be modulated, rather the light-guiding structure itself represents the sensing element.

Influence on radiation can either cause a change in signal height, a shift in absorption maximum (frequency), a change in phase by shift in superposition of partial beams of radiation, a shift in the state of polarization, or a change in the time domain.

Direct detection is defined by the measuring of direct spectroscopic properties of the analyte or the sensitive layer influenced by the analyte, whereas *indirect* detection is defined by the detection of an analyte interacting with an additional reagent. The terms passive or reactive detection are used in the literature as well. Indirect detection often gives the opportunity to enhance the selectivity, but, on the other hand, further obstructions as effective immobilization or bleaching of a dye may

occur. The large amount of papers dealing with indirect detection is reflected by the efforts taken to find new selective chromophores,[20,21] a collection about several applications and frontiers in analytical chemistry, sensor technology and the importance of near IR dyes herein is given in Daehne et al.[12]

In the following text, first principles of operation of planar waveguide or FOS are given, then the principles of interaction between sensitive layer and analyte are discussed. Next, some applications are discussed for selected optical transduction principles.

2 FUNDAMENTALS

2.1 Aspects of Sensing in the Evanescent Field Region

Evanescent wave sensors are devices for the detection of refractive index changes of the sensitive layer n_{sl} in contact with the waveguide, thus frequently called microrefractometers. This changes of n_{sl} may be caused simply by sorption or by a chemical reaction. The sensitive layer is hereby a component of the light-guiding element, which can be a planar waveguide or fiber optics. The concept of intrinsic sensing is demonstrated in Figure 2 and will be discussed in a qualitative matter using FOS. Small diameter (<1 μm dependence on wavelength of guided radiation) of a waveguide results in monomode properties, larger diameters allowing multimode operation. For a more detailed discussion see Göpel et al.,[6] Wolfbeis[8] and Boisdé and Harmer.[9]

The properties of optical fibers depend on their numerical aperture and the critical angle of total reflection. Both

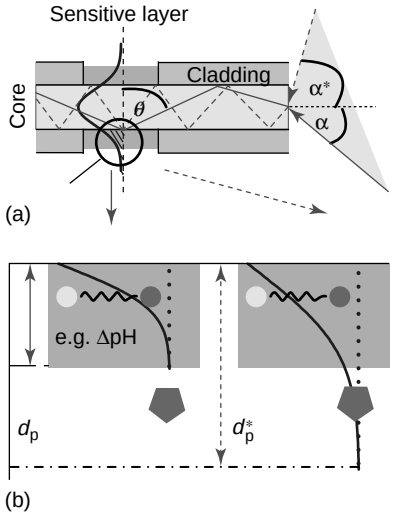

Figure 2 Principle of intrinsic sensing. (a) Evanescent field in the sensitive layer; (b) d_p depending on different experimental conditions (α, α^*, λ).

of these depend on the refractive index of the core n_{core} and cladding n_{clad} of the fiber. Additionally, the numerical aperture depends on the refractive index of the surrounding medium as well. At each spot, where the guided light meets the interface between the core and the cladding, an electromagnetic field will be present on the side of the cladding, respectively, in the sensitive layer. This field is the so-called evanescent field, which penetrates into the cladding and the sensitive layer, depending on the refractive indices of the core and the cladding, respectively, the sensitive layer (n_{clad}, n_{sl}), and on the incident angle θ at the core/cladding interface. The penetration depth d_p can be calculated with Equation (1):[9]

$$d_p = \lambda \quad 2\pi n_{core} \quad \sin^2\theta - \left(\frac{n_{core}}{n_{clad}}\right)^{2} \quad ^{-0.5} \qquad (1)$$

Analyte molecules sorbing into the sensitive layer will change n_{sl} (if $n_{analyte}$ and n_{sl} differ) and influence the evanescent field. This means that the sorbed analyte molecules modulate the light transmission properties of the fiber, which can be detected by a spectrometer or photodiodes.

Advantages of this intrinsic sensing concept are:

- no interrupted optical path, therefore an effective light budget;
- distributed sensing;
- many different transducers available, direct and indirect sensing possible;
- if $d_p < d_{sl}$, no cross-interference with substances outside the sensitive layer.

The parameters such as numerical aperture of the fiber, monomode or multimode fiber, wavelength, and thickness of the sensitive layer[22] must be chosen with care, in order to ensure that the penetration depth d_p will not become greater than the thickness of the sensitive layer. For applications in the middle IR region, this aspect is of special interest, since d_p is proportional to wavelength and d_p is increasing considerably in the region of the critical angle.[22] An increase of the sensitive layer's thickness would increase diffusion times of the target molecule and hence decrease the sensor's performance.

2.2 Sensitive Layers

In recent years a lot of different materials have been investigated in order to enhance sensitivity and selectivity of the sensitive layers. A new approach to mimic antibodies is imprinted polymers,[23,24] whereby the most promising examples achieving specific detection were demonstrated in liquid phases. Additionally, cyclohexapeptide,[25] porphyrins,[26,27] phthalocyanines,[28,29] tailored polymers,[21,30–32] sol-gels[33–36] and organically modified

silicates[37,38] (ormosils) were the focus of several publications.

2.2.1 Aspects Influencing Sorption

It will be of importance to obtain highly selective matrices, which are used to enrich the analyte molecules or to bind receptor molecules like dyes. For in doing so a profound understanding of the sorption of analytes in polymeric, sol-gel or other layers, which will be referred to as sensitive layers in the following text, is necessary. The sorption of an analyte molecule into a sensitive layer can be separated schematically in at least three processes:

1. adsorption onto the surface of the sensitive layer;
2. phase transition and diffusion from the outer surface into the sensitive layer's bulk;
3. subsequent physical interactions with the sensitive layer.

Another process will take place, if the analyte molecule has chemical interactions with certain reaction sites present in the sensitive layer (Figure 3).

2.2.2 Boiling Point Dependency of Sorption

Sometimes the dependence of sorption on boiling point is misinterpreted as size selectivity of the sensitive layer. However, the energy needed to cause the phase transition from the adsorbed state to the gas phase increases with lower vapor pressure. If no real size-selective properties of the matrix or sensitive layer exist, e.g. as found in zeolites[39] or polyimides,[40] the sorption of analytes for direct detection can be described in the following way.

A scale for the enrichment of a compound i is the partition coefficient f^i, which is the ratio of the concentration of the analyte molecule in the sensitive layer c^i_{sl} and the concentration in the gas phase c^i_g as shown in Equation (2):

$$f^i = \frac{c^i_{\mathrm{sl}}}{c^i_g} \tag{2}$$

Nonpolar analytes interact with the chosen sensitive layer in a similar way. Then f^i will increase with decreasing vapor pressure. Taking Picet–Trouton's rule, which is valid in a homologous series, the boiling point (T_{b}) dependence of f^i for the analyte i can be estimated[41] according to

$$\log f^i = 4.8\frac{T_{\mathrm{b}}}{T} - 4.8 + 0.43\frac{\Delta S^{0,\mathrm{m}}}{R} \tag{3}$$

where $\Delta S^{0,\mathrm{m}}$ is the standard entropy of mixing. Equation (3) shows that the vapor pressure of a substance is governing the enrichment of a substance within a sensitive layer, if no subsequent chemical interactions occur.

2.2.3 Further Factors Affecting Enrichment

Attractive forces between analyte molecules and the sensitive layer can cause additional effects. Besides Coulomb forces, permanent dipole–dipole, induced dipole–dipole, and dispersion forces are discussed. In addition, hydrogen bonds will increase the interaction between analyte and sensitive layer. For these kinds of interactions the chemist's rule of thumb "like dissolves like" may be used to choose which sensitive layer is an appropriate choice for a certain analyte, in order to achieve high partition coefficients.

Furthermore, the analyte molecule may undergo chemical reactions at reaction sites present in the sensitive layer (see steps 3 → 4 in Figure 3) enhancing the sensitivity and selectivity. A typical example of such a reaction is the reaction between an acidic or basic molecule and an indicator dye entrapped in the sensitive layer.

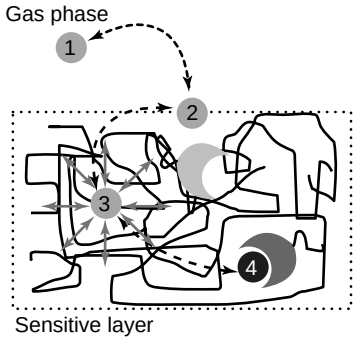

Figure 3 Sorption steps of an analyte molecule: (1) adsorption onto sensitive layer; (2) phase transition and diffusion from the surface into the bulk; (3) subsequent physical interactions; (4) diffusion to reaction site (physically or chemically trapped in the polymer) with chemical interaction.

3 REFLECTOMETRIC INTERFERENCE SPECTROSCOPY

3.1 Introduction

RIfS was described by Gauglitz et al.[42] and has similar sensing capabilities as quartz microbalances (QMBs). Both methods can detect the enrichment of volatile organic compounds (VOCs) or other analytes in polymer matrices. Whereas QMBs detect in principle the mass uptake of the sensitive layer, RIfS detects the change of the optical path length, which can be measured directly. With respect to the terminology of section 2.2, RIfS

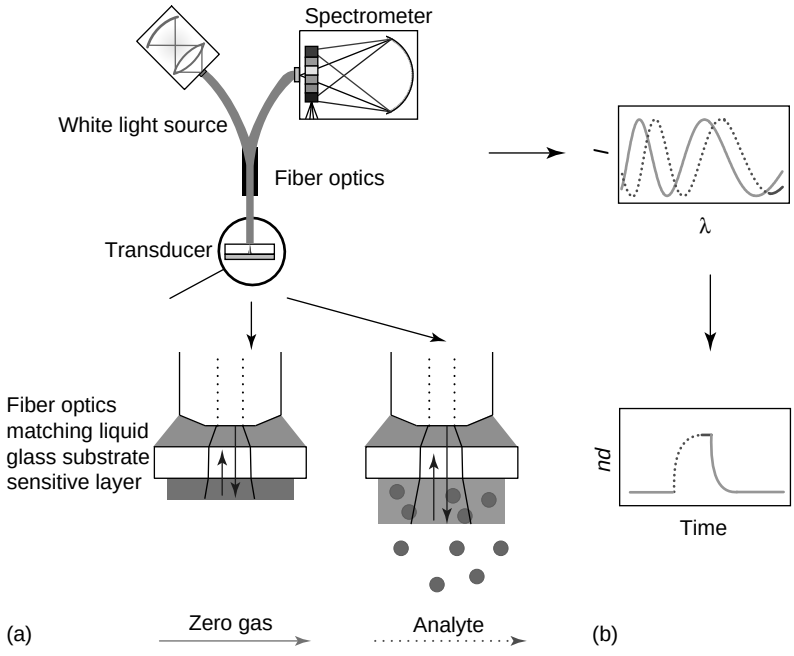

Figure 4 (a) Measurement set-up consisting of a white light source, fiber optics, transducer, and a diode array spectrometer; (b) typical reflectance spectra and the corresponding change of the optical pathlength *nd*.

is an extrinsic FOS system detecting modulations in the wavelength region, and allowing direct and indirect analyte detection.

3.2 Principle of Reflectometric Interference Spectroscopy

Spectral interferometry measures the interference pattern caused by the superposition of two reflected partial beams at parallel interfaces of a thin layer. Spectral interferometry depends on the refractive indices of the film, the thickness of the film, the wavelength (preferably in the optical region 400–700 nm), and the angle of incidence of radiation.

The principle set-up depicted in Figure 4 consists of a white light source, Y-fibers, the sensing element, and a diode array spectrometer. The signal is measured relative to a coated (noncoated) substrate I_r (I_{ref}), according to Equation (4). I_0 represents the blank or zero signal.

$$I = \frac{I_r - I_0}{I_{ref} - I_0} \tag{4}$$

If an analyte sorbs into a polymer film, the film will change its optical thickness *nd*, where *d* is the physical thickness and *n* the refractive index of the sensitive layer. A more detailed description of the principle and evaluation of RIfS is given by Kraus and Gauglitz.[43] The system was adapted to multianalyte sensing by Seemann et al.[44] A miniaturized RIfS set-up, where the white light source was replaced by four light emitting diodes, and the spectrometer was replaced by a simple photo diode, was presented recently.[45] Because of the coherence condition of white light and the interference condition, the film thickness of that layer, where the sensitive coating builds up the interference layer, has to range between 300 nm and 30 μm. However, if the sensitive coating is attached to a solid SiO_2 layer acting as basic interference layer, even sorption on to monolayers can be investigated.[25,46]

3.3 Applications

The response of a polymer like poly(dimethylsiloxane) (PDMS) versus several VOCs as *n*-pentane to *n*-octane or chlorinated hydrocarbons is described by Gauglitz et al.[46] Films thinner than 1 μm yield response times less than 1 s. For tetrachloroethene a detection limit in the ppm range was found and the sensor response was linear up to 10 000 ppm. Since the PDMS-based sensor showed high cross-sensitivities against the mentioned VOCs, it became obvious that a multisensor set-up and subsequent chemometrics[47] must be used, in order to overcome this limitation in the analysis of mixtures. Artificial neuronal networks[44] and partial least-squares regression methods[48] were applied to discriminate binary or ternary VOC mixtures. The sorption behavior of different VOCs in five different polymer-based sensitive layers with and without relative humidity was investigated.

Based on these considerations, great effort was made to find sensitive layers with higher selectivities. For

cyclohexapeptides, which proved to be highly selective for certain amino acids in aqueous phase,[49] it could be verified that the analyte's permittivity plays a major role besides the analyte's vapor pressure for the resulting sensitivity. Similar selectivities to those in the aqueous phase towards different VOCs,[25] could not be found in the gas phase.

The detection of basic or acidic substances with optic gas sensors is usually accomplished by physical or chemical entrapment of an indicator dye in a sol-gel or polymer matrix. As long as the optical properties of the indicator dye will not change in a unexpected way, covalent linking of the dye is advantageous, since diffusion or leaching of the dye can be prevented. One major drawback of indicator dyes is due to photobleaching,[13] which limits the operational stability of the sensor. pH-responsive polymers allow the dyeless optical detection of basic or acidic substances by covalent incorporation of acidic or basic functional groups in the polymer matrix. Analyte-induced deprotonation of these functional groups leads to additional uptake of water by the polymer. Using a modified polysiloxane[32] with carboxylic acid end-groups, detection limits below 200 ppb could be determined for ammonia. The signal decrease was 30% in 68 days. During this time period the sensor was continuously exposed to white light and synthetic air (20.5% O_2 in N_2). A dye doped sol-gel layer exhibited a 40% decrease of the signal under identical boundary conditions within 10 h. Drawbacks were the long response time and the influence of relative humidity on the sensor's signal.

In principle, achieving chiral discrimination by gas sensors is more difficult than by gas chromatography (GC), because the sensor uses only one ''theoretical plate'' (one adsorption/one desorption step). In contrast, using GC discrimination typically results from thousands of successive adsorption/desorption equilibria. Nevertheless Bodenhöfer et al.[50] successfully demonstrated the separation of optical enantiomers using QMB and RIfS. *S*- and *R*-octyl-Chiralsil-Val were used as selective coatings, and the enantiomers of *N*-trifluoroacetyl-alanine-methylester (TFA-Ala-OMe) were used as analytes. Both sensors, the *S*- and *R*-type, showed identical signals when exposed to the racemic mixture of TFA-Ala-OMe; however, there was little increase in signal height for the equivalent isomer.

4 FIBER OPTIC SENSORS

The term opt(r)ode is commonly used to name sensor systems, wherein optical fibers are taken. Both terms were defined by Lübbers. Optode[51] is derived from the Greek, meaning the ''optical path'', whereas optrode[52] stands for ''optical electrode''. Since the applications for FOS, especially indirect FOS, are extensively covered in the literature (see section 1), the theory will only be handled in brief. Only recent applications will be considered. A description of interferometric noise is given by Jin et al.[53,54] and may prove useful for the design of a particular fiber optic gas sensor, since interference between signal radiation and reflected radiation can cause signal fluctuation in the output, which can limit the performance of the sensing system. By the application of gas-permeable membranes,[55] FOS are easily adapted for liquid sensing.

4.1 Extrinsic Sensing Applications

4.1.1 Extrinsic Direct

Dasgupta et al.[56] presented a new long-path optical transmission cell for fiber optic absorbance measurements, which is a liquid core fiber optic waveguide. The cladding of this waveguide is simply a hollow polymer tube. Since the polymer's refractive index is <1.333, water can be used as the core of the fiber. The fluoropolymer used has a three times higher gas permeability with respect to poly(tetrafluoroethylene)-type Teflon®, therefore the fiber shows potential for chemical sensing in the liquid and gaseous phases. A FOS for smoke measurements was presented by Matias et al.[57] The sensor works based on the following principle: diesel engine particulates get stuck on a gluing surface of a white tape, changing the diffuse reflectance of the tape's surface. Response times in the range of 0.1 s were reported. Different types of extrinsic sensing configurations are schematically shown in Figure 5.

4.1.2 Extrinsic Indirect

Cardoso et al.[58] described a highly sensitive fluorescence-based extrinsic indirect FOS for the detection of atmospheric H_2S in the ppb_v range. A drop of

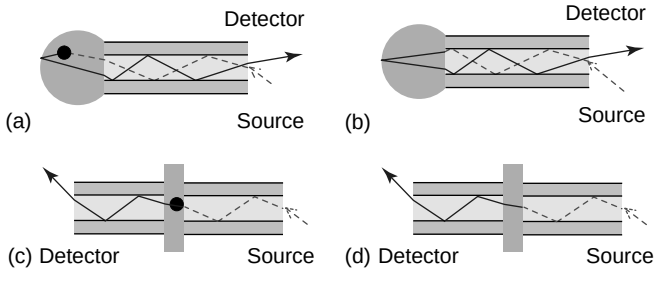

Figure 5 Different measurement configurations for extrinsic fiber optic sensing. (a) Indirect, end of fiber; (b) direct, end of fiber; (c) indirect, transmission; (d) extrinsic, direct, transmission.

alkaline fluorescein acetate solution is suspended in a flowing air sample stream. As H_2S in the sample is collected by the alkaline drop, it reacts rapidly with the fluorescence dye, resulting in a significant decrease in fluorescence. Tabacco et al.[59] modified an indirect CO_2 sensor with a gas permeable membrane for measurements in seawater. The sensor exhibited a reversible working range between 200 ppm and 1000 ppm PCO_2. Selective opt(r)ode membranes were presented by Spichiger-Keller[21] for the detection of ethanol and CO_2 in bioreactors.

4.2 Intrinsic Sensing Applications

The principle of sensing in the evanescent field region has already been shown in Figure 2. Additionally, typical sensing configurations are depicted on Figure 6. In the normal fiber arrangement, part of the cladding is removed to allow the evanescent field to be directly in contact with the sensitive layer either in a transmission mode (Figure 6a) or in reflection (Figure 6b). Another approach is to bend the fiber at the region coated with the sensitive layer (Figure 6c). This may lead to a significant increase of the sensitivity;[60,61] however, the system becomes sensitive to mechanical disturbance.

4.2.1 Direct Intrinsic

A porous sol-gel layer[36] was used in order to detect analyte-induced changes of the refractive index. The detection limits of different alkanes and halogenated hydrocarbons were in the range between 0.6% and 25%. An ozone sensor was described by Potyrailo et al.[62] with good stability over a 2-month period of investigation and a linear response between 0.02 vol.% O_3 and 0.35 vol.% O_3. The response proved to be reversible with a time constant of 1 min. It is noteworthy that these investigations were performed in the near ultraviolet (UV) region (at 254 nm), whereas most published investigations for evanescent field sensing are either in the visible or near and mid IR wavelength region. Attenuated total reflection IR/FOS[18] can be classified as direct intrinsic sensors as well.

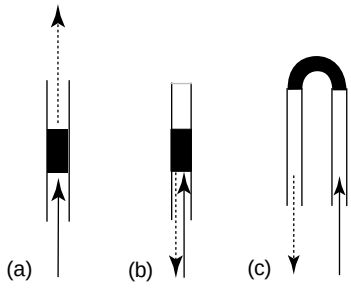

Figure 6 Different fiber optic evanescent field sensing configurations: (a) sensitive layer, side on fiber; (b) like (a) with mirror attached, end on fiber; (c) bent fiber.

A multipoint sensor system is described in Stewart et al.[63] for methane detection along a long fiber using optical time-domain reflectometry (OTDR).[64] A short light pulse is coupled into the fiber and transported by total internal reflection. A certain amount of the light will be back-scattered at the fiber's end and can be recorded by a fast detector resolved with time. This allows localization of any effect (laterally resolved) within the cladding of the fiber by the time delay between the excitation pulse and the received back-scattered signal (change in the effective refractive index at a point causes partial back-scattering of the guided radiation). This is a well-known technique to characterize fiber optic networks. An overview of the possible applications with direct and indirect signal modulation was recently given by Sensfelder et al.[65] One limitation of the OTDR technique is the loss of light guiding capabilities, if the refractive index change reaches such an extent that the out-coupling of radiation damps the propagating radiation. Thus no information can be obtained about effects beyond this point.

4.2.2 Indirect Intrinsic

An indirect intrinsic fiber optic evanescent field sensor based on near IR dyes was reported by Malins et al.[66] Methyl and ethyl derivatives of a 5-(4′-dialkylaminophenylimino)-quinolin-8-one copper(II) perchlorate complex were immobilized in a thin poly(vinyl chloride) matrix coated on an optical fiber with 600 μm diameter. Ammonia in the range between 10 ppm and 100 ppm was investigated. The sensor showed response (recovery times) of the order of 30 to 40 s.

UV-curable ormosils have been synthesized by the sol-gel method and used as sensitive layer for silica optical fibers.[31] The refractive index of the coating could be tailored in the range between 1.46 and 1.56. The sensitivity versus CO_2 could be enhanced by the introduction of amino groups into the polymer.

A humidity sensor based on a hydroxypropylcellulose film doped with Rhodamine B was presented by Taguchi et al.[60]

5 PLANAR WAVEGUIDE SENSORS

5.1 Introduction

An illustrative insight into the field of integrated optics (IO) and planar or slab waveguides is given by Göpel et al.[6] and Boisdé and Harmer.[9] IO comprises several benefits such as ruggedness, miniaturization, construction on a chip, high sensitivity, and remote sensing capabilities,

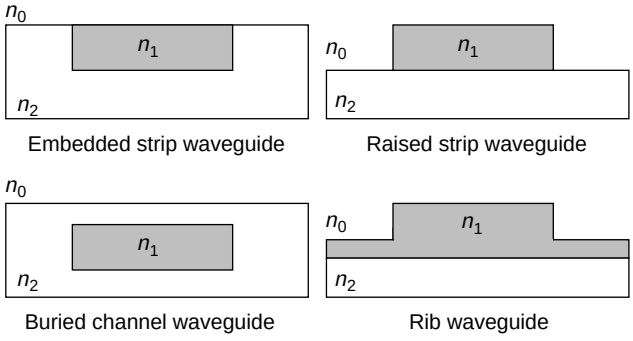

Figure 7 Different types of planar waveguides, where $n_1 > n_2, n_0$.

since IO parts can efficiently be linked to fiber optics. Planar waveguides are frequently used for intrinsic sensing in the evanescent field region by the use of grating couplers and polymers acting as sensitive layers. IO grating couplers were proposed by Tiefenthaler and Lukosz[67] in 1987.

Planar waveguides can be classified as a film waveguide (thin layer on top of a substrate) or a strip waveguide (often called a slab waveguide). A slab waveguide can be described in the most simple form as a dielectric layer with the refractive index n_1 matched in-between two layers, the substrate and the superstrate layer, with the refractive indices n_0, n_2, where $n_1 > n_0, n_2$. Some variations of the set-up of planar waveguide configurations are depicted in Figure 7.

Since a proper theoretical discussion of the influence of the waveguide structure is far beyond the scope of this review, we want to refer to the already mentioned monograph by Göpel et al.[6] and references therein. Shortly, the sorbed analyte will directly or additionally indirectly vary the refractive index of the sensitive layer, thereby either modulating the transmission of the guided light or reflectance properties of a grating. Usually the light source is a laser and will be coupled into the waveguide by a grating, a prism[68] or a gradient index fiber.[69] Chirped waveguide gratings[70] simplify the conditions to meet the coupling conditions. Several concepts of the IO refractive index sensing systems are described by Kunz.[71]

Another approach uses interference between two waveguide arms forming a Mach–Zehnder type arrangement.[72] One of the waveguide arms is coated by polymer or any other sensitive layer, the other is a reference arm. Reduction of propagation velocity in the first arm causes phase modulation of the signal monitored after reunification of both radiation pathways (see Figure 8).

5.2 Applications

An integrated optical absorbance sensor for the detection of trichloroethene and tetrachloroethene in the near IR wavelength region was described by Burck et al.[69]

As frequently applied in FOS, the incorporation of indicator dyes into a sensitive layer is used for the detection of acidic or basic analytes. An example is given by Brandenburg et al.[73] for the detection of ammonia. Bromocresol Green and Bromophenol Blue were immobilized in a silicone matrix. Concentrations less than 1 ppm were detectable with a dynamic range up to 20 ppm. Acidic gases were found to reduce the relative change in signal.

Iodine gas can be detected by the formation of charge-transfer complexes between iodine and phenyl groups, which were incorporated in a methylated glass film.[74] A linear response was obtained in the range between 100 ppb and 15 ppm, with response and recovery times less than 15 s.

A coupling wave optical waveguide was presented by Qing et al.,[33] which proved to be very sensitive versus H_2S. The feature of this sensor is that the one waveguide can be adjusted for light guiding and the other for sensing.

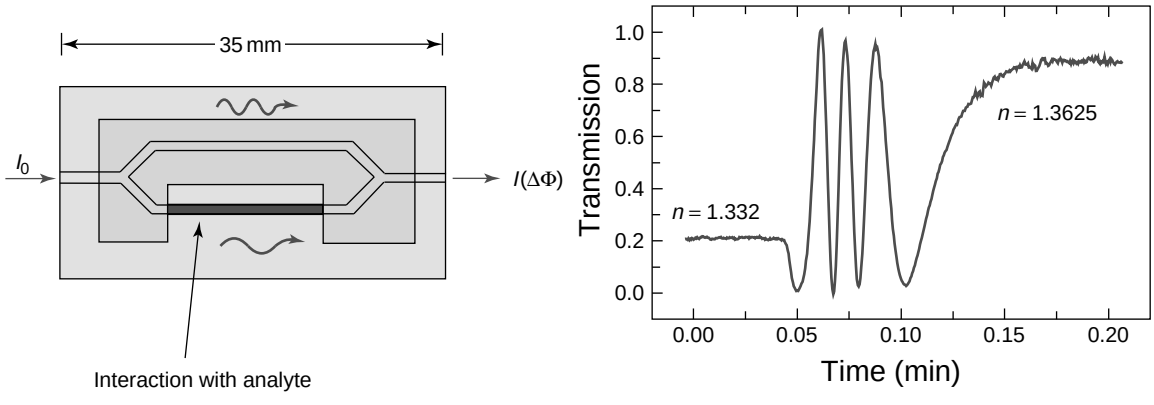

Figure 8 Mach–Zehnder arrangement with phase modulation of detected intensity.

Thin films of RbCl and AgCl were used as sensitive films on a planar optical waveguide to sense chlorine with a measurement range from 3 ppm to 1.5%.[75]

The properties of a nematic liquid crystal as a sensitive layer for VOCs were investigated by Drapp et al.[76] Therein it was demonstrated that the analyte-induced phase transition of liquid crystalline systems can lead to enhanced sensitivities and selectivities.

6 SURFACE PLASMON RESONANCE SPECTROSCOPY

6.1 Introduction

SPR spectroscopy is a frequently used technique in biosensor technology.[77] The great demand for a marker free transducer system is reflected by the appearance of commercially available systems like the Swedish Biacore AB, launching the first commercial SPR biosensor in the market in 1990. Another commercial SPR system was introduced by Texas Instruments (called TI-SPR-1) with bio- and chemosensing capabilities in liquid phases.[78] Liedberg et al.[79] pioneered SPR spectroscopy in the field of chemical sensors in 1983. However, the amount of publications concerning chemosensors is well below the amount of publications concerning the biosensor field. Forgenson and Yee[80] introduced the fiber optic evanescent field to SPR spectroscopy in 1993. Miniaturization and parallelization of the transducer can be achieved by integrated optic devices. The use of fiber optics and integrated optic devices would enhance the remote sensing capabilities, allow multiple component analysis and probably reduce the wide gap between the laboratory and field applications. A survey of trends in SPR was recently given by Homola et al.[81] With respect to the terminology discussed in section 1.2, SPR spectroscopy is a direct intrinsic sensing method.

6.2 Principle of Surface Plasmon Resonance

A plasmon wave is a collective excitation of the electron gas in a metal.[82] There exist two different kinds of surface plasmons depending on the excitation conditions, radiating and nonradiating surface plasmons. Whereas volume and radiating surface plasmons[83] have to be excited by UV light, nonradiating surface plasmons[84] can be excited by polarized visible light. The latter case exists for the applications given. No further distinction between the different kinds of the surface plasmons will be made in the following text.

The total reflection of light, occurring at a phase boundary (e.g. between the prism coupler and a metal film), is accompanied by the origin of an evanescent field, which penetrates the metal film. If the energy of the evanescent field is meeting the resonance condition for surface plasmon waves in the metal film, the reflected light will experience an attenuation. The metal film has usually a thickness of around 50 nm.[82,85] The surface plasmon wave is a resonant phenomenon that depends on the index of refraction and the thickness of the metal film, as well as on the index of refraction of the adjacent medium containing the analyte (and a sensitive layer, respectively). The resonance condition to excite a surface plasmon can be met by either varying the angle of incidence θ or wavelength λ of radiation.

SPR measurement techniques are based on the detection of modulation of the intensity,[79,86] the angle,[78,87,88] the wavelength, leaky modes[89] or by phase modulation of the reflected light.[90,91] The detection of the pure intensity, which is based on a measurement set-up with fixed angle and wavelength, is less desirable, since it cannot be determined whether the signal change is caused by a change of n or whether experimental errors are responsible. Common experimental methods are in principle based on angle or wavelength modulation. An elegant set-up for angle modulation, without any components in motion was accomplished by Melendez et al.[78] Wavelength modulation does not require moving instrumental components, simplifying the experimental conditions and enhancing reproducibility. A further method is based on the detection of the phase modulation.[90] Since surface plasmon waves can only be excited by transversal magnetic modes (plane of polarization parallel to the plane of incidence) excitation of surface plasmon waves will cause a detectable phase shift between reflected transversal electric (perpendicularly polarized) and transversal magnetic polarized light.

6.3 Measurement Set-ups

In contrast to the Otto type, which excites surface plasmons by visible light at the base of a prism with a short distance to a metal surface,[92] Kretschmann pursued another strategy.[93] The thin metal layer is directly coated on to the base of a dielectric. The plasmon is excited at the opposite surface of the metal film close to the analyte or a sensitive layer as demonstrated in Figure 9.

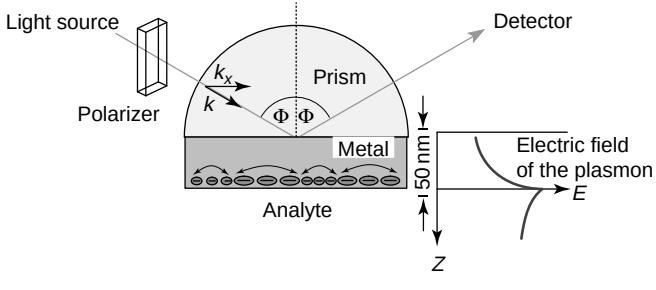

Figure 9 Typical Kretschmann configuration.

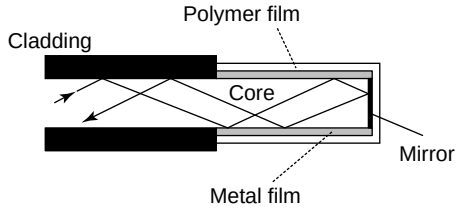

Figure 10 Set-up of a fiber optic SPR.

A modification of the Kretschmann configuration was described recently by Homola and Yee.[94] The use of two crossed polarizers with a 45° orientation to the plane of incidence and a phase retarder allow the detection of the surface plasmon as a peak, which may be advantageous if the usually detected resonance minimum is adversely affected by, say, an analyte-induced optical nonhomogeneity.

In the original set-up the high refractive index material was a prism, but it can also be the core of a fiber[80] or the window of a planar waveguide.[95] The fiber optic SPR set-up is simpler than the classical Kretschmann set-up. The prism and the ray trace is replaced by a Y-fiber optical waveguide as schematically shown in Figure 10. The reference spectra are recorded by measuring the reflectance spectra of the fiber with the bare metal film (without sensitive layer). No particular care needs to be taken wih regard to the coupling of light into the prisms. In contrast to the Kretschmann configuration, where just one reflection occurs, waveguide SPR increases the signal by taking advantage of multiple interactions along the length of the metal layer. Waveguide sensing devices are therefore simple, easy to handle, usable as probes or for remote sensing.

Integrated optical SPR sensors based on planar waveguides have several benefits: small size, ruggedness, the fact that multichannel sensors on a single chip are possible, with all this being combined with fiber optics remote sensing capabilities. A design tool for a multilayer set-up to calculate the transmission characteristics of planar waveguide SPR sensors is described by Harris and Wilkinson.[95] The device shown in Figure 11 demonstrates the multilayer approach.

A variant of the planar waveguide set-up was developed by Karlsen et al.[96] and Homola and Yee.[97] A measurement set-up with multicomponent chemical analysis

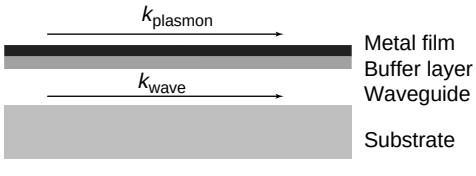

Figure 11 Scheme of a waveguide-based SPR configuration.

capabilities was proposed, based on a planar light pipe, which allows the reflectivity to be determined as a function of wavelength for a set of angles of incidence. The detection of phase shift between transversal magnetic and transversal electric polarized light was introduced by Nelson et al.[90] and has a high calculated resolution for Δn (0.5×10^{-6}), which could not be verified experimentally. Additionally, it must be noted that the experimental set-up for the detection of phase modulation amounts to as many as 21 components.

6.4 Applications

Since SPR spectroscopy detects changes of the refractive index, the sensitivity is not only governed by sorption effects, but in addition the difference between the refractive indices of the analyte and the sensitive matrix has to be taken into account.[98] Furthermore, the insufficient stability of the silver layer against oxidation or hydrogen sulfide[99] must be solved in order to achieve a high operational stability of the sensitive layer under real conditions. This is commonly achieved by the use of alkylthiols as self-assembled monolayers (SAMs). Niggemann et al.[100] obtained an operational stability of 3 months, with an identical sensitivity of the PDMS layer. The response time of the sensor was in the range of 2.5 s and the detection limit was estimated to be 15 ppm.

Abdelghani et al.[88,98] used a fluoroalkoxysiloxane polymer and alkythiol SAMs as well. The observed limits of detection for chlorinated hydrocarbons were still in the low percent range (C_2HCl_3, 0.3; CCl_4, 0.7; $CHCl_3$, 1; CH_2Cl_3, 2). However, coating the silver or gold layer will not solve all the problems, according to Schoenfisch and Pemberton.[101] They found that the sulfur–metal bonding is oxidized under ambient conditions. It was proposed that ozone is responsible for the observed oxidation process.

Besides the detection of VOCs with sensitive layers on polysiloxane bases, the detection of NO_x with porphyrin-based sensitive layers is described in several publications.[28,29,102] Zhu et al.[29] used a copper phthalocyanine Langmuir–Blodgett film to detect NO_x gas. The limit of detection was in the range of several ppm, with unacceptably long recovery times for a practicable sensor. A selective NO_2 was described by Ashwell and Roberts.[103] It exhibits few cross-interferences to 100 ppm SO_2, H_2S, Cl_2, CO, CO_2 and 10% deterioration after 1000 to 1500 adsorption/desorption cycles. Nikkitin et al.[102] described an experimental set-up that allows combination of surface-enhanced Raman scattering (SERS) and SPR to detect NO_2 (limit of detection < 4 ppm). The recovery time was reported to be longer than 300 s. Furthermore, it was noted that the

sensing effect of the pure gold film may exceed the signals of the phthalocyanine-coated layers. A recent publication describes the enhancement of the sensitivity using a dye (1,1′,3,3,3′,3′-hexamethylindodicarbocyanine iodide) which contains a strong absorption band centered in the dynamic range of the SPR coupling wavelength.[104]

Agbor et al.[105] investigated a polyaniline-based sensor versus NO_2 and H_2S with detection limits of 50 ppm. On to the silver layer a thin nickel layer was deposited (1 nm Ni on top of 50 nm Ag) in order to prevent other gas effects on the silver layer.

7 SUMMARY

The review intends to demonstrate that a large variety of optical transduction principles exist and these provide some solutions to different analytical problems. Most of them employ similar sensitive materials, proving that the quality of the sensing device in most cases depends on the properties of the sensitive layer rather than on the transduction principle chosen. Certainly in the future there will be some improvements in the optical properties of the different devices.

The main interest in recent years has been dedicated to sensing-material that is more stable, more selective, demonstrates better response times and is tailored to suit the application. In this area, new materials and new approaches will attract the most interest. Problem-solving that related especially to synthesized polymer films that combine stability and inertness with some selectivity by functional groups, while keeping their reversibility will be well worth the efforts of chemists. The first publications testing approaches with liquid crystals and polymer backbones, as well as using molecular imprinted polymers, demonstrate these trends.

Scanning the literature for new developments shows the need to combine knowledge in optics with skill in polymer chemistry in order to create improved sensor devices. This fact simply takes account of the fundamental idea of a chemical sensor containing (in addition to a physical sensor) a sensitive layer which takes care of the chemical interaction between the analyte and the sensor and determines the properties of the sensor system.

ABBREVIATIONS AND ACRONYMS

FOS	Fiber Optic Sensors
GC	Gas Chromatography
IO	Integrated Optics
IR	Infrared
OTDR	Optical Time-domain Reflectometry
PDMS	Poly(dimethylsiloxane)
QMB	Quartz Microbalance
RIfS	Reflectometric Interference Spectroscopy
SAM	Self-assembled Monolayer
SERS	Surface-enhanced Raman Scattering
SPR	Surface Plasmon Resonance
TFA-Ala-OMe	*N*-trifluoroacetyl-alanine-methylester
UV	Ultraviolet
VOC	Volatile Organic Compound

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Field-portable Instrumentation **(Volume 4)**
Field-based Analysis of Organic Vapors in Air

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

REFERENCES

1. H. Baltes, W. Göpel, J. Hesse (eds.), *Sensors Update*, VCH, Weinheim, Vol. 2, 1996.
2. M. Otto, *Chemometrie: Statistik und Compertereinsatz in der Analytik*, VCH, Weinheim, 1997.
3. B.A. McKinley, C.L. Parmley, 'Clinical Trial of an Ex Vivo Arterial Blood Gas Monitor', *J. Clin. Care*, **13**, 190–197 (1998).
4. I.K. Weiss, S. Fink, J.D. Feldmann, J.E. Brill, 'Clinical Use of Continuous Gas Monitoring in Pediatric Intensive Care Unit', *Pediatrics*, **103**(2), 440–445 (1999).
5. H.M. Heise, L. Küpper, L.N. Butvina, 'Attenuated Total Reflection Mid-infrared Spectroscopy for Clinical Chemistry Applications Using Silver Halide Fibers', *Sens. Actuators, B*, **51**, 84–91 (1998).
6. W. Göpel, J. Hesse, J.N. Zemel (eds.), *Sensors*, 3rd edition, VCH Publishers, New York, Vols. 2/3 and 6, 1993.
7. J. Janata, M. Josowicz, P. Vanýsek, 'Chemical Sensors', *Anal. Chem.*, **70**, 179R–208R (1998).
8. O.S. Wolfbeis, *Fiber Optic Chemical Sensors and Biosensors*, CRC Press, Boca Raton, Vols. I and II, 1991.
9. G. Boisdé, A. Harmer, *Chemical and Biochemical Sensing with Optical Fibers and Waveguides*, Artech House, London, 1996.

10. B. Culshaw, J. Dakin (eds.), *Optical Fiber Sensors: Systems and Applications*, Vol. 2, 1997.
11. B. Culshaw, J. Dakin (eds.), *Optical Fiber Sensors: Applications, Analysis and Future Trends*, Vol. 4, 1997.
12. S. Daehne, U. Resch-Genger, O.S. Wolfbeis (eds.), *Near Infrared Dyes for High Technology Applications*, NATO ASI Series, 3. High Technology, Kluwer Academic Publishers, Norwell, Vol. 52, 1998.
13. O.S. Wolfbeis, 'Analytical Chemistry with Optical Sensors', *Fresenius' J. Anal. Chem.*, **325**, 387–392 (1986).
14. K.R. Rogers, E.J. Poziomek, 'Fiber Optic Sensors for Environmental Monitoring', *Chemosphere*, **33**, 1151–1174 (1996).
15. V. Kumar, D. Chandra, 'Fiber Optic Methane Sensors for Mines', *Opt. Appl.*, **27**(4), 279–290 (1997).
16. R.A. Potyrailo, S.E. Hobbs, G.M. Hieftje, 'Optical Waveguide Sensors in Analytical Chemistry: Today's Instrumentation, Applications and Trends for Future Development', *Fresenius' J. Anal. Chem.*, **362**, 349–373 (1998).
17. B. Mizaikoff, R. Göbel, R. Krska, K. Taga, R. Kellner, M. Tacke, A. Katzir, 'Infrared Fiber-optical Chemical Sensors with Reactive Surface Coatings', *Sens. Actuators, B*, **29**, 58–63 (1995).
18. M. Saito, K. Kikuchi, 'Infrared Optical Fiber Sensors', *Optical Rev.*, **4**(5), 527–539 (1997).
19. R. Altkorn, I. Koev, A. Gottlieb, 'Waveguide Capillary Cell for Low-refractive-index Liquids', *Appl. Spectrosc.*, **51**, 1554–1558 (1997).
20. P. Bühlmann, E. Pretsch, E. Bakker, 'Carrier Based Ion-selective Electrodes and Bulk Optodes. 2. Ionophores for Potentiometric and Optical Sensors', *Chem. Rev.*, **98**, 1593–1687 (1997).
21. U.E. Spichiger-Keller, 'Ion- and Substrate-selective Optode Membranes and Optical Detection Modes', *Sens. Actuators, B*, **38–39**, 68–77 (1997).
22. C. Egami, K. Takeda, M. Isai, M. Ogita, 'Evanescent-wave Spectroscopic Fiber Optic pH Sensor', *Opt. Commun.*, **122**, 122–126 (1996).
23. G. Wulff, R. Schönfeld, 'Polymerizable Amidines – Adhesion Mediators and Binding Sites for Molecular Imprinting', *Adv. Mater.*, **12**, 957–959 (1998).
24. K. Haupt, A. Dzgoev, K. Mosbach, 'Assay System for the Herbicide 2,4-Dichlorophenoxyacetic Acid Using a Molecularly Imprinted Polymer as an Artificial Recognition Element', *Anal. Chem.*, **70**, 628–631 (1998).
25. D. Leipert, F. Rathgeb, M. Herold, J. Mack, G. Gauglitz, G. Jung, 'Interaction Between Volatile Organic Compounds and Cyclopeptides Detected with Reflectometric Interference Spectroscopy', *Anal. Chim. Acta*, **392**(2–3), 213–221 (1999).
26. D.P. Arnold, D. Manno, G. Micocci, A. Serra, A. Tepore, L. Valli, 'Porphyrin Dimers Linked by a Conjugated Alkyne Bridge: Novel Moieties for the Growth of Langmuir–Blodgett Films and their Applications in Gas Sensors', *Langmuir*, **13**(22), 5951–5956 (1997).
27. M.G. Baron, R. Narayanaswamy, S.C. Thorpe, 'Hydrophobic Membrane Sensors for the Optical Determination of Hydrogen Chloride Gas', *Sens. Actuators, B*, **34**(1–3), 511–515 (1996).
28. J.P. Lloyd, C. Pearson, M.C. Petty, 'Surface Plasmon Resonance Studies of Gas Effects in Phthalocyanine Langmuir–Blodgett Films', *Thin Solid Films*, **160**, 431–443 (1988).
29. D.G. Zhu, M.C. Patty, M. Harris, 'An Optical Sensor for Nitrogen Dioxide Based on a Copper Phthalocyanine Langmuir–Blodgett Film', *Sens. Actuators, B*, **2**, 265–269 (1990).
30. G.J. Mohr, N. Tirelli, C. Lohse, U.E. Spichiger-Keller, 'Development of Chromogenic Copolymers for Optical Detection of Amines', *Adv. Mater.*, **10**(16), 1353–1357 (1998).
31. K. Rose, V. Matejec, M. Hayer, M. Pospisilova, 'Organopolysiloxanes as Chemical Coatings for Optical Fibers', *J. Sol-gel Sci. Technol.*, **13**, 729–733 (1998).
32. F. Rathgeb, G. Gauglitz, 'Dyeless Optical Detection of Ammonia in the Gas Phase Using pH-responsive Polymers with Reflectometric Interference Spectroscopy', *Anal. Chim. Acta*, **372**, 333–340 (1998).
33. D.-K. Qing, K. Itoh, M. Murabayashi, 'A Coupling Wave Optical Waveguide Chemical Sensor Constructed by Combining the Ion-exchange and the Sol-gel Methods', *Chem. Lett.*, **8**, 623–624 (1996).
34. L. Han, T.M. Niemczyk, L. Yunfeng, G.P. Lopez, 'Chemical Sensors Based on Surface Modified Sol-gel-coated Infrared Waveguides', *Appl. Spectrosc.*, **52**(1), 119–122 (1998).
35. P.J. Skrdla, S.S. Saavedra, N.R. Armstrong, S.B. Mendes, N. Peyghambarian, 'Sol-gel-based, Planar Waveguide Sensor for Water Vapor', *Anal. Chem.*, **71**(7), 1332–1337 (1999).
36. A. Abdelghani, J.M. Chovelon, N. Jaffrezic-Renault, M. Lacroix, H. Gagnaire, C. Veillas, B. Berkova, M. Chomat, V. Matejec, 'Optical Fiber Sensor Coated with Porous Silica Layers for Gas and Chemical Vapor Detection', *Sens. Actuators, B*, **44**, 495–498 (1998).
37. P. Lavin, C.M. McDonagh, B.D. MacCraith, 'Optimization of Ormosil Films for Optical Sensor Applications', *J. Sol-gel Sci. Technol.*, **13**(1–3), 641–645 (1998).
38. C. Malins, B.D. MacCraith, 'Dye-doped Organically Modified Silica Glass for Fluorescence Based Carbon Dioxide Gas Detection', *Analyst*, **123**(11), 2373–2376 (1998).
39. D.W. Breck, *Zeolite Molecular Sieves*, Wiley & Sons, New York, 1974.
40. G. Maier, 'Polymermembranen zur Gastrennung', *Angew. Chem.*, **110**, 3128–3143 (1998).
41. R. Kherrat, N. Jaffrezic-Renault, P. Greco, H. Helmers, P. Benech, R. Rimet, 'Integrated Optical Enlarged Field Interferometers Used as a Chemical Sensor', *Sens. Actuators, B*, **37**, 7–11 (1996).

42. G. Gauglitz, W. Nahm, 'Observation of Spectral Interferences for the Determination of Volume and Surface Effects of Thin Polymer Films', *Fresenius' J. Anal. Chem.*, **341**, 279–283 (1988).
43. G. Kraus, G. Gauglitz, 'Application and Comparison of Algorithms for the Evaluation of Interferograms', *Fresenius' J. Anal. Chem.*, **344**, 153–157 (1992).
44. J. Seemann, F.-R. Rapp, A. Zell, G. Gauglitz, 'Classical and Modern Algorithms for the Evaluation of Data from Sensor Arrays', *Fresenius' J. Anal. Chem.*, **359**, 100–106 (1997).
45. D. Reichl, R. Krage, C. Krummel, G. Gauglitz, 'Sensing of Volatile Organic Compounds Using a Simplified Reflectometric Interference Set-up', *Appl. Spectrosc.*, (in press).
46. G. Gauglitz, A. Brecht, G. Kraus, W. Nahm, 'Chemical and Biochemical Sensors Based on Interferometry at Thin (Multi-)layers', *Sensors and Actuators, B*, **11**, 21–27 (1993).
47. G. Kraus, G. Gauglitz, 'Optical Reflectometric Gas Sensing: Classification of Hydrocarbon Vapors by Pattern Recognition Applied to RIfS Sensor Signals', *Chem. Int. Lab. Sys.*, **30**, 211–221 (1995).
48. S. Kaspar, F. Rathgeb, D. Nopper, G. Gauglitz, 'Influence of Relative Humidity in Sensing Halogenated Hydrocarbons with Reflectometric Interference Spectroscopy (RIfS)', *Fresenius' J. Anal. Chem.*, **363**, 193–196 (1999).
49. D. Leipert, D. Nopper, M. Bauser, G. Gauglitz, G. Jung, 'Investigation of the Molecular Recognition of Amino Acids by Cyclopeptides with Reflectometric Interference Spectroscopy', *Angew. Chem.*, **110**, 3503–3505 (1998).
50. K. Bodenhöfer, A. Hierlemann, J. Seemann, G. Gauglitz, B. Koppenhöfer, W. Göpel, 'Chiral Discrimination Using Piezoelectric and Optical Gas Sensors', *Nature*, **387**, 577–579 (1997).
51. D.W. Lübbers, N. Opitz, 'The pCO_2/pO_2 Optode: A New Probe for Measurement of pCO_2 and pO_2 of Gases and Liquids', *Z. Naturforsch.*, **30C**, 532 (1975).
52. N. Opitz, D.W. Lübbers, 'A New Fast-responding Optical Method to Measure pCO_2 and Solutions', *Eur. J. Physiol.*, **355**, R120 (1975).
53. W. Jin, M.S. Demokan, G. Stewart, W. Philip, B. Culshaw, 'Performance Limit of Fiber Optic Gas Sensors from Coherent Backscatter', *IEEE Proc. Optoelectron.*, **145**(3), 186–190 (1998).
54. W. Jin, Y.Z. Xu, M.S. Demokan, 'Investigation of Interferometric Noise in Fiber-optic Gas Sensors with use of Wavelength Modulation Spectroscopy', *Appl. Opt.*, **38**, 7239–7246 (1997).
55. S.K. Spear, S.L. Patterson, M.A. Arnold, 'Flow Through Fiber-optic Ammonia Sensor for Analysis of Hippocamus Slice Persulfates', *Anal. Chim. Acta*, **357**, 79–84 (1997).
56. P.K. Dasgupta, Z. Genfa, S.K. Porothoor, S. Caldwell, S. Dong, S.Y. Liu, 'High-sensitivity Gas Sensors Based on Gas Permeable Liquid Core Waveguides and Longpath Absorbance Detection', *Anal. Chem.*, **70**, 4661–4669 (1998).
57. F.A.A. Matias, W.A. de Oliveira, E. Moschim, 'Fiber Optic Sensor for Diesel Engine Smoke Measurement', *Sens. Actuators, B*, **41**, 159–162 (1997).
58. A.A. Cardoso, H. Liu, K. Dasgupta, 'Fluorometric Fiber Optic Drop Sensor for Atmospheric Hydrogensulfide', *Talanta*, **44**, 1099–1106 (1997).
59. M.B. Tabacco, M. Uttamal, M. McAllister, D.R. Walt, 'An Autonomous Sensor and Telemetry System for Low Level $p(CO_2)$ Measurements in Seawater', *Anal. Chem.*, **71**, 154–161 (1999).
60. S. Otsuki, K. Adachi, T. Taguchi, 'A Novel Fiber-optic Gas-sensing Configuration Using Extremely Curved Optical Fibers and an Attempt for Optical Humidity Detection', *Sens. Actuators, B*, **53**, 91–96 (1998).
61. G. Schwotzer, I. Latka, H. Lehmann, R. Willsch, 'Optical Sensing of Hydrocarbons in Air or Water Using UV Absorption in the Evanescent Field of Fibers', *Sens. Actuators, B*, **38–39**, 150–153 (1997).
62. R.A. Potyrailo, S.E. Hobbs, G.M. Hieftje, 'Near-ultraviolet Evanescent-wave Absorption Sensor Based on a Multimode Optical Fiber', *Anal. Chem.*, **70**, 1639–1645 (1998).
63. G. Stewart, C. Tandy, D. Moodie, M.A. Morante, F. Dong, 'Design of a Multi-point Sensor for Gas Detection', *Sens. Actuators, B*, **51**, 227–232 (1998).
64. M.K. Barnoski, S.M. Jensen, 'Fiber Waveguides: A Novel Technique for Investigating Attenuation Characteristics', *Appl. Opt.*, **15**, 2112–2115 (1976).
65. E. Sensfelder, J. Burck, H.-J. Ache, 'Characterization of a Fiber-optic System for the Distributed Measurement of Leakages in Tanks and Pipelines', *Appl. Spectrosc.*, **52**, 1283–1298 (1998).
66. C. Malins, M. Landl, P. Simon, B.D. MacCraith, 'Fiber Optic Ammonia Sensing Employing Novel Near Infrared Dyes', *Sens. Actuators, B*, **51**, 359–367 (1998).
67. K. Tiefenthaler, W. Lukosz, 'Integrated Optical Switches and Gas Sensors', *Opt. Lett.*, **9**, 137–139 (1987).
68. G. Hanisch, R.P. Podgorsek, H. Franke, 'Origin of Optical Anisotropy in Planar Waveguides', *Sens. Actuators, B*, **51**, 348–354 (1998).
69. J. Burck, B. Zimmermann, H.J. Ache, 'Integrated Optical Evanescent Wave Absorbance Sensor for Chemical Analysis', *Fresenius' J. Anal. Chem.*, **354**(3), 284–290 (1996).
70. R.E. Kunz, J. Edlinger, P. Sixt, M.T. Gale, 'Replicated Chirped Waveguide Gratings for Optical Sensing Applications', *Sens. Actuators, A*, **46–47**, 482–486 (1995).
71. R.E. Kunz, 'Gradient Effective Index Waveguide Sensors', *Sens. Actuators, B*, **11**, 167–176 (1993).
72. R. März, '*Integrated Optics, Design and Modeling*', Artech House, Boston, 205, 1994.

73. A. Brandenburg, R. Edelhauser, T. Werner, O.S. Wolfbeis, 'Ammonia Detection via Integrated Optical Evanescent Wave Sensors', *Mikrochim. Acta*, **121**, 95–105 (1996).
74. L. Yang, S.S. Saavedra, N.R. Armstrong, 'Sol-gel-based Planar Waveguide Sensor for Gaseous Iodine', *Anal. Chem.*, **68**, 1834–1841 (1996).
75. Z.A. Ansari, R.N. Karekar, R.C. Aiyer, 'Influence of Chlorine on Planar Optical Guide Cladded with RbCl and AgCl', *Appl. Surf. Sci.*, **125**(2), 149–156 (1998).
76. B. Drapp, D. Pauluth, J. Krause, G. Gauglitz, 'The Application of the Phase Transition in Nematic Liquid Crystals for the Optical Detection of Volatile Organic Compounds', *Fresenius' J. Anal. Chem.*, **364**, 121–127 (1999).
77. H. Lortat-Jacob, S. Ricard-Blum, 'Use of Plasmon Resonance (BIAcore) for the Analysis of Ligand–Receptor Interactions', in *Visualization of Receptors*, ed. G. Morel, CRC Press, Boca Raton, 161–180, 1997.
78. J. Melendez, R. Carr, D.U. Bartholomew, K. Kukansis, J. Elkind, S.S. Yee, C. Furlong, R. Woodbury, 'A Commercial Solution for Surface Plasmon Sensing', *Sens. Actuators, B*, **35–36**, 212–216 (1996).
79. B. Liedberg, C. Nylander, I. Lundström, 'Surface Plasmon Resonance for Gas Detection and Biosensing', *Sens. Actuators, B*, **4**, 299–304 (1983).
80. R.C. Forgenson, S.S. Yee, 'A Fiber-optic Chemical Sensor Based on Surface Plasmon Resonance', *Sens. Actuators, B*, **12**, 213–220 (1993).
81. J. Homola, S.S. Yee, G. Gauglitz, 'Surface Plasmon Resonance Sensors Review', *Sens. Actuators, B*, **54**(1–2), 3–15 (1999).
82. C.C. Jung, S.B. Saban, S.S. Yee, R.B. Darling, 'Chemical Electrode Surface Plasmon Resonance Sensor', *Sens. Actuators, B*, **32**, 143–147 (1996).
83. K.L. Kliewer, R. Fuchs, 'Collective Motion in a Metallic Slab', *Phys. Rev.*, **153/2**, 498–512 (1966).
84. H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer Verlag, Berlin, 1986.
85. I. Stemmler, A. Brecht, G. Gauglitz, 'Compact SPR-transducers with Spectral Readout for Biosensing Applications', *Sens. Actuators, B*, **54**, 98–105 (1999).
86. B. Chadwick, M. Gal, 'An Optical Temperature Sensor Using Surface Plasmons', *Jpn. J. Appl. Phys., Part 1*, **32**, 2716–2717 (1993).
87. B. Liedberg, I. Lundstrom, E. Stenberg, 'Principles of Biosensing with an Extended Coupling Matrix and Surface Plasmon Resonance', *Sens. Actuators, B*, **11**, 63–72 (1993).
88. A. Abdelghani, J.M. Chovelon, N. Jaffrezic-Renault, C. Ronot-Trioli, C. Veillas, H. Gagnaire, 'Surface Plasmon Resonance Fiber-optic Sensor for Gas Detection', *Sens. Actuators, B*, **38–39**, 407–410 (1997).
89. R.P. Podgorsek, H. Franke, J. Woods, S. Morill, 'Monitoring the Diffusion of Vapor Molecules in Polymer Films Using SP-leaky-mode Spectroscopy', *Sens. Actuators, B*, **51**, 145–151 (1998).
90. S.G. Nelson, K.S. Johnston, S.S. Yee, 'High Sensitivity Surface Plasmon Resonance Sensor Based on Phase Detection', *Sens. Actuators, B*, **35–36**, 187–191 (1996).
91. P.I. Nikitin, A.A. Beloglazov, A.V. Kabashin, M.V. Valeiki, V.E. Kochergin, 'Surface Plasmon Resonance Interferometry for Sensor Applications', 4th European Conference on Optical Chemical Sensors and Biosensors', Münster, Germany, book of abstracts, 145–146, 1997.
92. A. Otto, 'Excitation of Non-radiative Surface Plasma Waves in Silver by Method of Frustrated Total Reflection', *Z. Phys.*, **23a**, 398–410 (1968).
93. E. Kretschmann, 'Radiative Decay of Non-radiative Surface Plasmons Excited by Light', *Z. Naturforsch.*, **23a**, 2135–2136 (1968).
94. J. Homola, S.S. Yee, 'Novel Polarization Scheme for Spectral Surface Plasmon Resonance Sensors', *Sens. Actuators, B*, **51**, 331–339 (1998).
95. R.D. Harris, J. Wilkinson, 'Waveguide Surface Plasmon Resonance Sensors', *Sens. Actuators, B*, **29**, 261–267 (1995).
96. S.R. Karlsen, K.S. Johnston, S.S. Yee, C.C. Jung, 'First-order Surface Plasmon Resonance System Based on a Planar Light Pipe', *Sens. Actuators, B*, **32**, 137–141 (1996).
97. J. Homola, S.S. Yee, 'Surface Plasmon Resonance Sensor Based on Planar Light Pipe: Theoretical Optimization Analysis', *Sens. Actuators, B*, **37**, 145–150 (1996).
98. A. Abdelghani, J.M. Chovelon, N. Jaffrezic-Renault, C. Veilla, H. Gagnaire, 'Chemical Vapor Sensing by Surface Plasmon Resonance Optical Fiber Sensor Coated with Fluoropolymer', *Anal. Chim. Acta*, **337**, 225–232 (1997).
99. L. Vassilev, V. Vulchev, L. Ljutov, S. Russev, 'Sensitivity of Silver Films to H_2S Contamination: Ellipsometric and Conductivity Studies', *Sens. Actuators, B*, **28**, 223–236 (1995).
100. M. Niggemann, A. Katerkamp, M. Pellmann, P. Bolsmann, J. Reinbold, K. Camman, 'Remote Sensing of Tetrachloroethylene with a Micro-fiber Optical Gas Sensor Based on Surface Plasmon Resonance Spectroscopy', *Sens. Actuators, B*, **34**, 328–333 (1996).
101. M.H. Schoenfisch, J.E. Pemberton, 'Air Stability of Alkanethiol Self-assembled Monolayers on Silver and Gold Surfaces', *J. Am. Chem. Soc.*, **120**, 4502–4513 (1998).
102. P.I. Nikkitin, A.A. Beloglazof, M.V. Valeiko, J.A. Creighton, A.M. Smith, N.A.J.M. Sommerdijk, J.D. Wright, 'Silicon-based Surface Plasmon Resonance Chemical Sensors', *Sens. Actuators, B*, **38–39**, 53–57 (1997).

103. G.J. Ashwell, M.P.S. Roberts, 'Highly Selective Surface Plasmon Resonance Sensor for NO_2', *Electron. Lett.*, **32**(22), 2089–2091 (1996).
104. A. Hanning, J. Roeraade, J.J. Delrow, R.C. Jorgenson, 'Enhanced Sensitivity of Wavelength Modulated Surface Plasmon Resonance Devices Using Dispersion from a Dye Solution', *Sens. Actuators, B*, **54**(1–2), 25–36 (1999).
105. N.E. Agbor, J.P. Cresswell, M.C. Petty, A.P. Monkman, 'An Optical Gas Sensor Based on Polyaniline Langmuir–Blodgett Films', *Sens. Actuators, B*, **41**(1–3), 137–141 (1997).

Photoacoustic Spectroscopy in Trace Gas Monitoring

Frans J.M. Harren, Gina Cotti, Jos Oomens, and Sacco te Lintel Hekkert
University of Nijmegen, The Netherlands

1 Introduction **1**
2 History **1**
3 Devices and Equipment **2**
3.1 Light Sources 2
3.2 Photoacoustic Cells 4
3.3 Limitations, Selectivity, Interference, Detection Limits 6
4 Environmental Applications **11**
4.1 Stack Gas Emission 11
4.2 Car Exhaust Emissions 11
4.3 Ambient Air Monitoring 11
5 Biological and Medical Applications **13**
5.1 Postharvest Physiology 13
5.2 Plant Physiology 14
5.3 Microbiology 16
5.4 Human Health, Noninvasive Breath Analysis 16
5.5 Entomology 17
6 Comparison with other Spectroscopic Methods in Trace Gas Monitoring **18**
Acknowledgments **19**
Abbreviations and Acronyms **19**
Related Articles **19**
References **20**

Since ancient times people have searched for ways to understand processes occurring in the environment, atmosphere or living organisms. Study of the gaseous trace compounds present may shed new light on chemical reactions taking place in the atmosphere or biochemical reactions inside organisms such as plants, animals and human beings. This article presents photoacoustic spectroscopy as a sensitive, on-line and noninvasive tool for monitoring the concentration of trace gases. Following a short introduction and a historic overview, attention is focused on the description of devices and equipment which determine the detection limits and selectivity. An overview is given of the current detection limits for photoacoustic detection. Applications are discussed with emphasis on environmental monitoring (in ambient air, car exhaust and stack gas emission), on medical applications and on biological applications (in postharvest physiology, plant physiology, microbiology and entomology).

1 INTRODUCTION

A gaseous molecule that absorbs electromagnetic radiation is excited to a higher electronic, vibrational or rotational quantum state. Generally, depopulation of this quantum state to lower lying states occurs either via fluorescence or collisions, the latter giving rise to a temperature increase of the gas due to energy transfer to translation. This nonradiative relaxation process occurs when the relaxation time can compete with the radiative lifetime of the excited energy levels. Radiative decay has a characteristic lifetime of 10^{-7} s at visible wavelengths as compared with 10^{-2} s at 10 μm. For nonradiative decay these values depend on the pressure (decay time τ inversely proportional to the pressure) and can vary strongly at atmospheric pressures (10^{-3}–10^{-8} s).

By modulating the radiation source at an acoustic frequency the temperature changes periodically, giving rise to a periodical pressure change which can be observed as an acoustic signal; in the gas phase the effect can be detected with a sensitive microphone.

Laser-based photoacoustic detectors are able to monitor trace gas concentrations under atmospheric conditions with orders of magnitude better sensitivity than conventional scientific instrumentation; in addition they are able to monitor noninvasively and on-line under dynamic conditions.

2 HISTORY

The photoacoustic effect was first reported by Alexander Graham Bell in 1880;[1] he discovered that thin discs emitted sound when exposed to a rapidly interrupted beam of sunlight. In a later experiment,[2] he removed the eye-piece of a commercial spectroscope and placed absorbing substances at the focal point of the instrument. The substances were put in contact with the ear by means of a hearing tube (Figure 1a and b) and he found "good" sounds in all parts of the visible and invisible electromagnetic spectrum of the sun.

Other publications on this phenomenon followed this first work; we mention here the works of Röntgen,[3] Tyndall[4] and Preece.[5] However, due to the lack of a quantitative description and the lack of a sensitive microphone, interest in the photoacoustic effect soon declined.

In 1938 Viegerov refined the photoacoustic technique for the first spectroscopic gas analysis;[6] hereafter Luft[7]

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

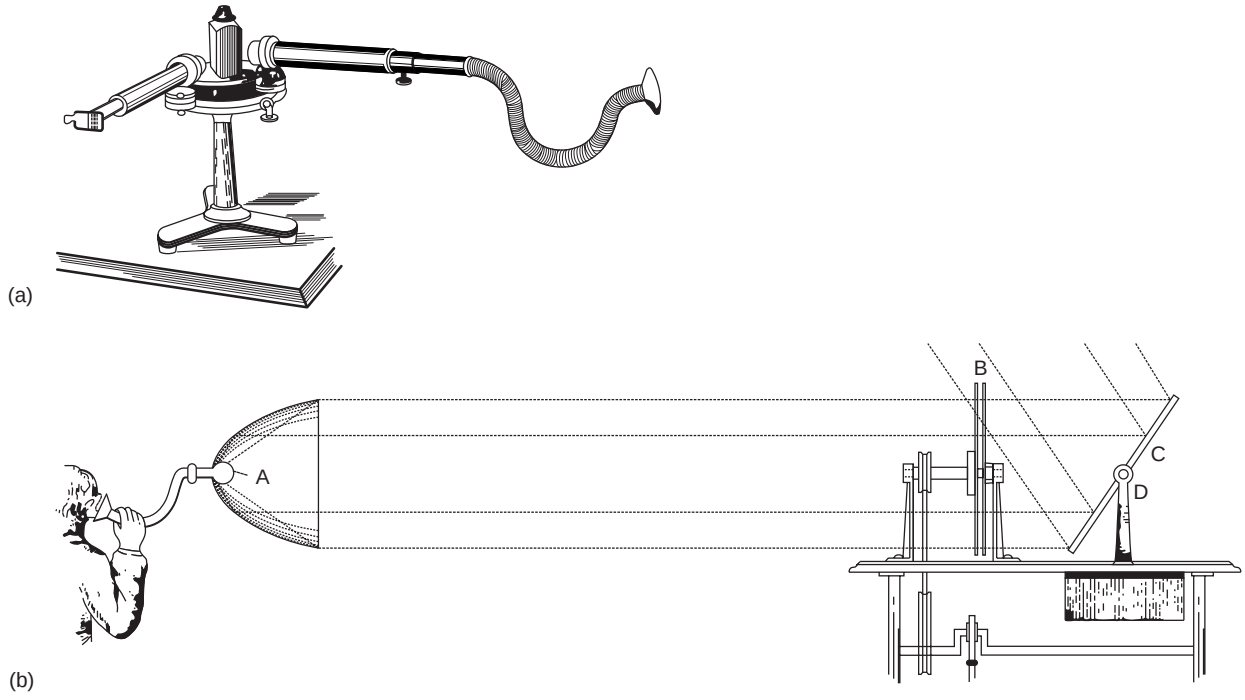

Figure 1 (a) The eye piece of a spectroscope is removed and substances are placed in the focal point of the instrument behind a slit. These substances are put in contact with the ear by means of a hearing tube.[2] (b) Sunlight is intercepted and aligned with a steering mirror (C, D), modulated with a chopper (B) and focused onto a glass bulb (A).[2]

measured trace gas absorption spectra with an infrared broadband light source down to the parts per million level.

By the end of the 1960s, after the invention of the laser, scientific interest increased once again. In 1968, Kerr and Atwood[8] utilized laser photoacoustic detection to obtain the absorption spectrum of small gaseous molecules. Due to the high spectral brightness of lasers and improved phase-sensitive lock-in techniques that amplified the acoustic signal, they were able to determine low concentrations of air pollutants. Kreuzer[9] demonstrated that it was possible to detect concentrations of e.g. 10^{-8} (10 parts per billion) of methane in nitrogen, using an intensity modulated infrared (3 μm) He-Ne laser. Patel et al. demonstrated the potential of the technique by measuring the NO and H_2O concentrations at an altitude of 28 km with a balloon-borne spin-flip Raman laser.[10] After this the photoacoustic effect was introduced into the field of trace gas detection with environmental, biological and medical applications.

3 DEVICES AND EQUIPMENT

3.1 Light Sources

With reference to the previous paragraph, lasers are not essential to operate photoacoustic gas detection systems. Although the spectral power density of broadband infrared light sources is orders of magnitude lower compared with lasers, their advantages are reliability and cost effectiveness. Infrared light sources in combination with various photoacoustic detection schemes are commercially available for trace gas detection at parts per million levels. Spectral selectivity is achieved by using FTIR (Fourier transformed infrared) spectroscopy in combination with spectral band filters in front of the photoacoustic cell; thus typically seven molecular gases can be detected simultaneously at the 1–100 ppm level.

The URAS gas analyzer of Hartmann & Braun uses a photoacoustic detection scheme which is able to detect a specific gas out of a multicomponent gas mixture avoiding cross interferences. In this instrument selectivity is achieved by comparing the direct absorption in a sample cell to that in a reference cell. After passing the sampling cells, each attenuated light beam enters a second detection cell filled only with the gas of interest (Figure 2); the detection cells are interconnected via a membrane connected to a capacitor. Since the dual beam is modulated the difference in acoustic energy reflects the difference in absorption and thus the concentration difference between sample cell and reference cell. The species under investigation enables the wavelength to be selected in such a way that all wavelengths at which absorption occurs are simultaneously active. When there is no spectral overlap from other gases, additional absorptions in the sample cell will not contribute to the acoustic signal; the light passes the

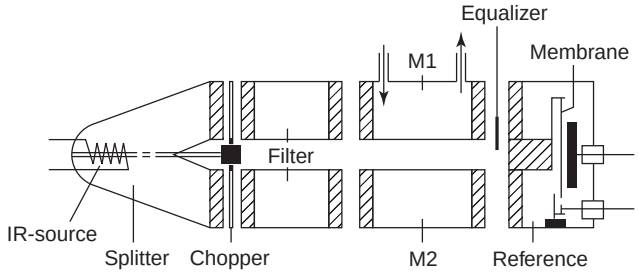

Figure 2 Infrared gas analyzer of Hartmann & Braun with photoacoustic detection scheme to detect a specific gas from a multicomponent gas mixture thereby avoiding cross-interference. Light from the infrared source is split into two paths. The chopper modulates the intensity for both paths. A filter volume in each path serves to filter out light of wavelengths not needed for the detection process; they can be filled with gases whose absorption spectra do not overlap with those of the species under scrutiny. M1 and M2 serve as measuring cell and reference cell respectively. With the help of the equalizer both light intensities become equal before entering the last cell. The last cell consists of two compartments with a membrane between them. Both compartments are filled with the gas under investigation so that all wavelength characteristics for this gas contribute to the signal. If the attenuation differs in M1 from that in M2 the membrane starts to oscillate with the frequency of the chopper. This oscillation (typically a few hertz) is detected capacitively.

detection cell unattenuated. When a specific compound, e.g. H_2O, causes spectral overlap, an extra cell can be placed in the light path filled with the interfering gas. This cell completely attenuates the wavelengths where this interfering molecule absorbs, including the spectral overlap regions. Thus, these wavelengths cannot contribute to the photoacoustic signal and a single component of a complex multicomponent gas mixture can be detected.

High spectral brightness renders CW (continuous wave) laser sources ideally suitable for photoacoustic trace gas detection. In contrast to direct absorption techniques, the photoacoustic signal is proportional to the laser power; from the Lambert–Beer law one finds for small absorptions (Equation 1)

$$P = P_0 e^{-\sigma N l} \longrightarrow (P_0 - P) \approx P_0 \sigma N l \qquad (1)$$

where P_0 and P are the laser power before and after the photoacoustic cell, respectively, σ is the absorption cross-section per molecule (in cm^2), N is the number of absorbing molecules per cubic centimeters and l the absorption path length (in cm). The absorbed energy $(P_0 - P)$ is converted into acoustic energy which is recorded by the microphone.

Saturation, due to nonlinear absorption of the laser power only occurs in focused high-power laser beams. The pumping rate to a higher rovibrational level is proportional to the laser light intensity; in the case of saturation it exceeds the collisional de-excitation rates.

Besides sensitivity, lasers achieve high selectivity. The spectral selectivity is only limited by the pressure broadened absorption profiles of the gases under investigation. The first practical lasers which were used to detect trace gases were CO_2 lasers.[11] These line tunable lasers cover the infrared 9–11 μm wavelength region with a laser line spacing of 0.5–2 cm^{-1}. Fingerprint absorption spectra can be achieved if we compare the laser line spacing to the pressure broadened absorption lines for the trace gas molecules (typically a few gigahertz at atmospheric pressures). Additionally, they are able to deliver high laser powers (1–100 W) from a relatively small gas discharge tube.

A new development is the application of the CO laser in its $\Delta v = 1$ (5.0–7.6 μm wavelength region) and $\Delta v = 2$ (2.5–3.8 μm) version. Although less powerful (typically 1 W) its performance can be improved by applying an intracavity set-up. CO lasers are line tunable with a line spacing between 0.5–1 cm^{-1}.

Other CW lasers in the visible and infrared have been applied to photoacoustic trace gas detection such as a spin flip Raman laser,[10] diode lasers[12,13] and dye lasers.[14] Although dye lasers and titanium sapphire do not cover the ideal wavelength region for trace gas detection they are very well suited for photoacoustic spectroscopy of weak absorption bands with their continuous tunability and a typical CW laser power of 1 W and they have proved their potential in molecular spectroscopy of highly vibrationally excited molecules.[15,16] However, for trace gas detection they are less applicable since the overtone molecular absorption cross-sections are weak, thereby raising the detection limits. For the same reason infrared diode lasers derived from telecommunication research (0.8 and 1.5 μm) are not very well suited.

Of periodically poled nonlinear crystals that have been developed, periodically poled lithium niobate (PPLN) is probably most well known.[17] Lithium niobate ($LiNbO_3$) can be used in combination with pulsed lasers to generate mid-infrared radiation (2–5 μm) by parametric oscillation. CW operation was limited to a few microwatts of laser power due to phase matching problems. Quasi-phase matching with periodically poled materials overcomes this problem resulting in laser powers up to a few watts in the infrared.[18]

The advantage of using pulsed lasers for photoacoustic detection is their wider infrared tunability and consequently better spectral overlap with interesting molecular gases. Examples can be given of the experiments performed with a pulsed CO_2 laser by Repond and Sigrist[19] and experiments with a pulsed optical parametric oscillator (OPO).[20] A disadvantage of pulsed lasers is their high peak powers (megawatt) in relation to their relatively low average energy (typically 1 W) which has to generate a relatively slow process as the gas phase photoacoustic

effect. Peak-to-peak fluctuation and nonlinear absorption disturb the reproducibility of the generated signal, thereby raising detection limits. Typical detection limits reported with these types of laser are in the parts per million range.

3.2 Photoacoustic Cells

To perform trace gas detection the ideal photoacoustic cell should amplify the generated sound originating from the molecular gas absorption meanwhile rejecting acoustic (and electric) noise and in-phase infrared absorption from other materials. Interfering gases should be distinguished by spectroscopic or physical methods (see section 3.3)

For gas phase measurements, mainly resonant cells are combined with modulated CW lasers and lock-in amplifiers; pulsed lasers are combined with piezoelectric detectors and boxcars. These selective amplifiers arise from the necessity to lower acoustic and electric noise levels thus improving the signal-to-noise ratio.

Other requirements for photoacoustic cells are a low gas consumption or a fast response; for this the active volume of the cell should be small so that no dilution can take place when the trace gas and its carrier flow through the acoustic cell.

If we consider a nonresonant, cylindrical cell, its performance can be expressed as its efficiency to convert absorbed photon energy into acoustic energy; this cell constant F (Pa cm W^{-1})[21,22] is given by Equation (2)

$$F_{\text{nonresonant}} = \frac{G(\gamma - 1)L}{wV} \tag{2}$$

where L and V are the length and the volume of the cell, respectively, γ is the specific heat constant, $w = 2\pi\nu$ the modulation frequency and G is a geometrical factor in the order of one. Within a nonresonant cell the gas absorption signal is independent of the cell length but inversely proportional to its diameter. However, photoacoustic signals are also generated by infrared window absorption. To reduce these signals inside the cell its length should be as long as possible to spread this locally generated sound over the total cell volume.

For a resonant cell, the above cell constant F must be amplified with the quality factor Q of the generated acoustic resonance, Equation (3)

$$F_{\text{resonant}} = QF_{\text{nonresonant}} \tag{3}$$

where Q is expressed by the ratio of the energy stored in the acoustical standing wave divided by the energy losses per cycle. This acoustical amplification process is limited by various dissipation processes which can be divided into surface and volume effects. Main surface losses are viscous and thermal losses at the resonator surface, microphone losses and acoustic wave scattering losses at obstacles in the cell. Volumetric losses are not as important as surface losses and are mainly due to free space viscous and thermal losses and $V-V$, $V-R$, $V-T$ relaxation losses of polyatomic gases.[23]

A cylindrical cavity can be a resonant cavity for sound waves. The resonance frequencies of such a cavity are given by Equation (4)

$$f_{mnp} = v_s \left(\left(\frac{\alpha_{mn}}{2R} \right)^2 + \left(\frac{p}{2L} \right)^2 \right)^{1/2} \tag{4}$$

where v_s is the sound velocity in the gas filling the cavity, R is the radius of the resonator, L is the length, $p = 0, 1, 2, 3 \ldots$ axial mode numbers and α_{mn} is a suitable solution of Bessel equations with $m =$ radial mode number and $n =$ azimuthal mode number (Figure 3).[24]

The cell constant F for all types of cylindrical resonant cells is proportional to $L^{1/2}/R$.[25] Due to their larger diameters resonances in the radial or azimuthal acoustic mode have high Q values and high resonance frequencies but low F values. A longitudinally excited resonator will have a low Q value but, due to its small diameter, a high F value.

In order to obtain an optimum signal-to-noise ratio, noise control and interfering signals have to be taken into account. Theoretically, the ultimately lowest acoustical noise results from random density fluctuations, i.e. Brownian motion, in the gas which is distributed over all frequencies in the sound spectrum. The total power of these density fluctuations is constant, but the frequency distribution is dependent upon local resonances and their Q values. Therefore, using a resonant cell the signal-to-noise ratio will not be improved by considering only these random thermal noise fluctuations.

However, these noise levels are far below other noise sources such as amplifier noise and acoustic disturbances from outside the cell. The power of the amplifier noise varies as $1/f$, where f is the modulation frequency of the light beam. Therefore, in contrast to Brownian noise, it is more advantageous to operate at a cell resonance to increase the generated acoustical signal above the $1/f$

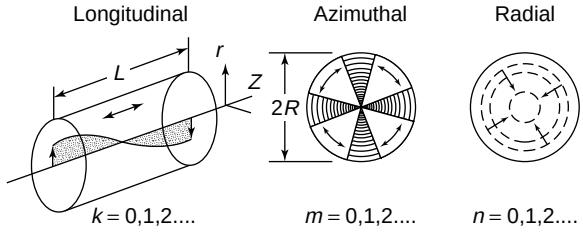

Figure 3 Resonant acoustic modes of a cylindrical closed chamber; the fundamental longitudinal, azimuthal and radial modes.

noise. A prerequisite is that external acoustic disturbances are shielded from the microphone by a proper cell wall construction, material choice and proper design of inlet and outlet ports.

Noise fluctuations do not have a fixed phase relation with the modulation of the light intensity. Other disturbing factors limiting the sensitivity do have a fixed phase relation. Directly generated coherent acoustical background signals, caused by the modulation process (e.g. chopper, current modulation) can be suppressed in the same way as external acoustical disturbances.

Other, more serious interferences are coherent photoacoustic background signals which are caused by absorption of the light beam in the window material or light scattered or reflected off the resonator wall. They are generated at the same frequency and in-phase with the modulated light beam. In resonant cells, window absorption signals can be suppressed by using large buffer volumes and $\lambda/4$-tubes next to the windows.[25] These one-end-open tubes (Figure 4), placed near the window, act as a notch filter for the window signal at the resonance frequency. The influence of the scattered light on the photoacoustic background signal can be minimized by using, for the resonator wall, a highly reflective polished material, with a thermally well conducting material as substrate, e.g. in the case of the CO_2 laser, a polished gold coated copper tube.[26]

In the past special designs have been developed for longitudinally, azimuthally and radially resonant photoacoustic cells,[27–30] even without windows[31,32] to improve sensitivity.

In order to improve the selectivity the combination of the Stark effect or Zeeman effect with photoacoustic detection represents an interesting solution for specific molecules such as ammonia (NH_3)[33,34] and nitric oxide (NO).[35] The change in absorption at a specific laser frequency depends on the shift and splitting of the molecular absorption lines of the species under investigation. Although the method does not suffer from interference problems within multicomponent gas mixtures, there are some limitations. The detection is limited to molecules with a permanent electric or magnetic dipole moment; in addition, the shift or splitting of the lines should be observable within the pressure broadened Lorentzian profile at the overlap with the laser frequency. For ammonia at the 10R8 transition of the $^{12}CO_2$ laser these requirements are satisfied; Thöny and Sigrist[33] showed that the absorption cross-section changed by a factor of two at atmospheric pressure by applying an electric field of 16.7 kV cm^{-1}.

Another way to improve the sensitivity for a specific gas is by temperature change. By increasing the temperature the vapor pressure of, for example, an oil with low vapor pressure is increased thereby improving the sensitivity. An elegant cell has been designed by Jalink and Bicanic,[36] who combined a heatpipe and a photoacoustic cell thereby keeping the microphone and windows at room temperature. In addition, a Helmholtz resonant photoacoustic cell (Figure 5) has been developed by Kästle and Sigrist for quantitative temperature-dependent absorption measurements of fatty acid vapors.[37] Helmholtz resonances (w_H) can be constructed

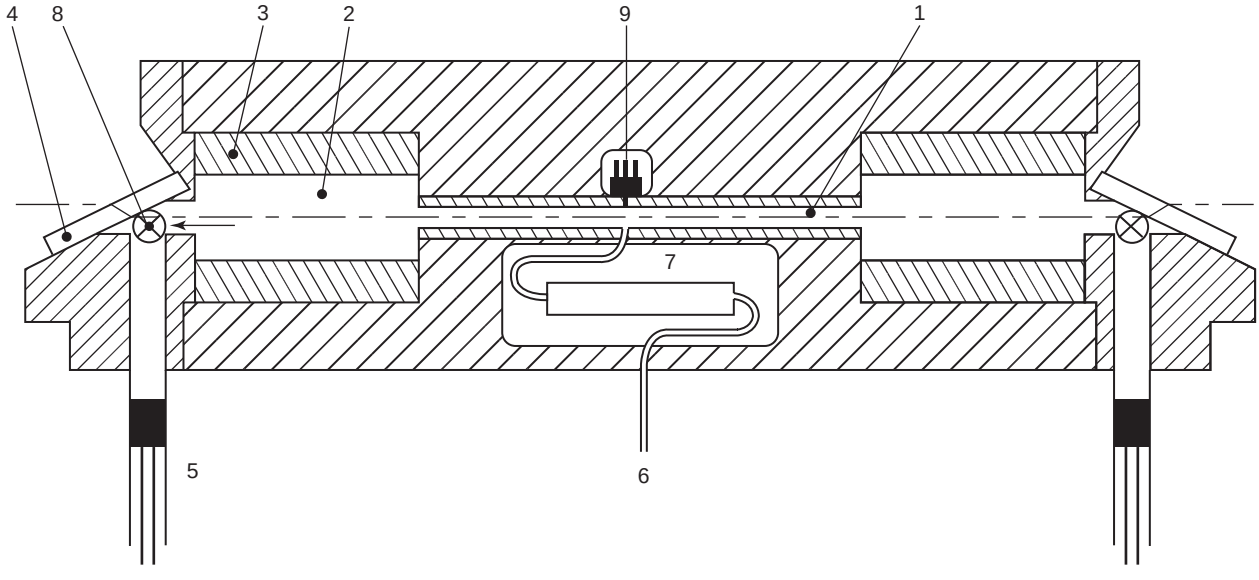

Figure 4 Resonant photoacoustic cell. 1, resonator; 2, buffer volume (maximum diameter 40 mm, length 50 mm); 3, buffer ring to decrease buffer radius; 4, ZnSe Brewster window; 5, adjustable $\lambda/4$ notch filter to suppress window signal; 6, inlet gas flow; 7, $\lambda/2$ notch filter to suppress flow noise; 8, outlet gas flow; 9, Knowless microphone.

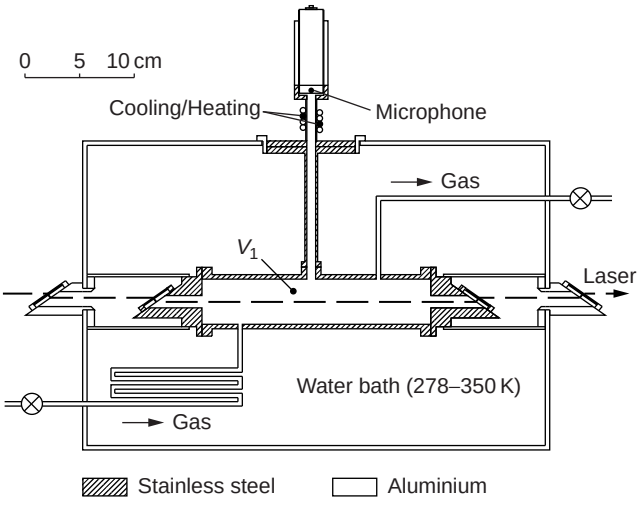

Figure 5 Photoacoustic cell applied for temperature-dependent investigations on fatty acids. The temperature of the water bath was varied between 278 and 350 K with a cold finger and two immersion heaters, while the temperature of the microphone was kept constant with a cooling/heating device. (Reproduced with permission from Kästle and Sigrist, *Appl. Phys. B*, **63**, 389–397 (1996), Copyright Springer Verlag.)

using a relatively large closed volume (V) connected to the outer world via a relatively long narrow duct (length L, cross-section S at a sound velocity c) Equation (5):[24]

$$w_{\mathrm{H}} = \overline{\frac{c^2 S}{LV}} \tag{5}$$

Amplification of sound waves can also be achieved with the thermoacoustic effect. The operating principle of such an amplifier connected to a photoacoustic resonator can be explained as follows. The gas inside the acoustical resonator is irradiated by an amplitude modulated laser beam. Trace gas absorption inside the resonator tube will produce a standing acoustical wave; this is the acoustic effect. The amplifier consists of a temperature gradient over a stack of thin sheets separated by a distance equal to two times the thermal boundary layer of the gas (Figure 6). If the gas is moving up and down over the temperature gradient inside the stack, heat energy is transferred to the gas and subsequently into the acoustic standing wave. With such a cell, Q values have been improved by a factor of 20–30.[38] By adjusting the amplifier just below the onset of self-oscillation, the acoustical wave in the resonator will put the amplifier in sustained oscillation, leading to a considerable enhancement of the photoacoustic signal. Noise, acoustic and photoacoustic background signals will also be enhanced. However, with careful design of the cell these levels can be suppressed thereby improving the signal-to-noise ratio.

3.3 Limitations, Selectivity, Interference, Detection Limits

Amplifying photoacoustic signals in resonant cells, reducing noise, using high laser powers, and so on are all used to gain an as high as possible sensitivity for molecular gas absorption. The sensitivity to trace a specific compound depends strongly on its spectroscopic properties. For example, closely spaced rotational absorption lines within a Q branch of a strong vibrational transition help to reach low detection limits of such a gas. One good example is the Q branch of the ν_7 band of C_2H_4 in the 10 μm region. The 10P14 CO_2 laser line is in exact resonance with this Q branch. This results in an extremely low detection limit for C_2H_4 of 6 parts per trillion (1 ppt = 1×10^{12}) in nitrogen. Some gases can be detected even better (e.g. SF_6) while others have higher limits due to less coincidence between available laser frequencies and molecular absorption features. In the future this problem may be overcome by powerful tunable infrared lasers such as the PPLN OPO[39] and quantum cascade lasers.[40] For detection limits see Table 1 and references therein.

Reactive gases such as ozone are very difficult to quantify in absolute concentrations, although ozone has a very strong absorption cross-section in the CO_2 laser wavelength region, resulting in sub-parts per billion detection limits. Ozone is highly reactive, thus interactions with tubing walls, sampling cuvettes, and the photoacoustic cell, cause rapidly decreasing ozone concentrations. This loss in concentration must be determined over the complete sampling system by calibrating the system with a specific ozone analyzer and checked for linearity over a large concentration range. For example, the breakdown of ozone over a 2-m sampling tube may be as high as 60%.[67] The degradation of ozone in a photoacoustic cell is determined to be 75% by inserting a second identical photoacoustic cell in the flow system. In spite of these limitations, successful studies have been performed to quantify atmospheric ozone concentrations.[68]

The first excited vibrational mode of N_2 has a long vibrational lifetime (≈1 ms at 1 atm (1 atm = 101.325 kPa) at $\nu_1 = 2200\,\mathrm{cm}^{-1}$). Using kilohertz modulation of the laser beam a phase shift may occur in the generated photoacoustic signal. A well known example is the kinetic cooling effect of CO_2 at CO_2 laser wavelengths (9–11 μm).[54,69,70] CO_2 has a hot-band absorption in this wavelength region and is excited to the ν_3 vibrational level which is almost resonant with the first vibrational level of N_2. Therefore, under atmospheric conditions the energy absorbed by CO_2 will only slowly be converted into a temperature and pressure increase of the gas.

N_2 is not unique. Another example is given in Figure 7 where CH_4 absorbs the CO laser radiation in

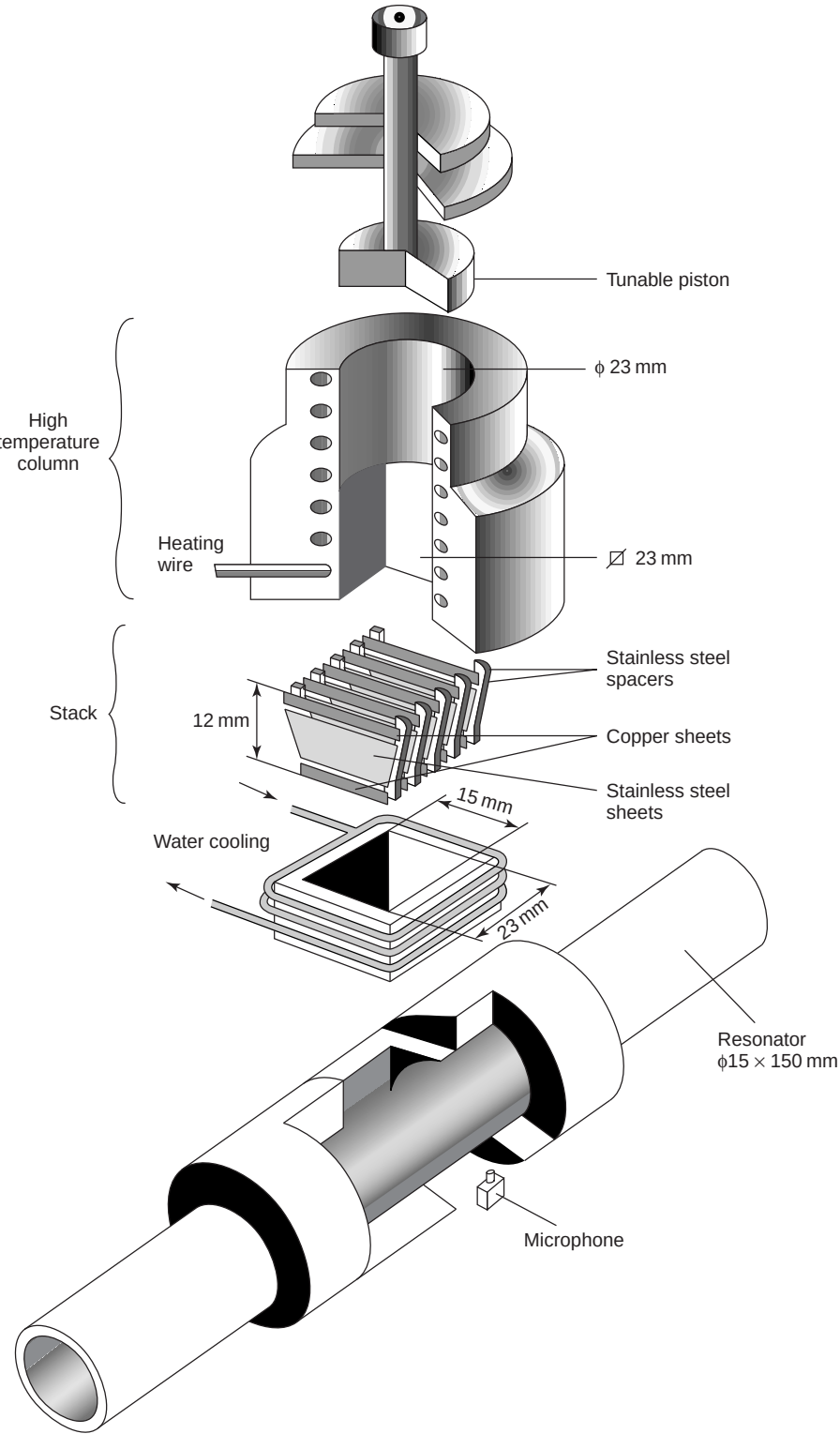

Figure 6 Expanded view of the thermoacoustic stack, dimensions are in millimeters. The tunable piston can slide in the upper part of the column which is cylindrically shaped (diameter 23 mm). The lower part (length 20 mm, square cross-section 23 × 23 mm) forms the connection to the stack. This part is heated by Inconel wire. The exploded view of the stack shows the stainless steel plates with the temperature gradient and copper plates to maintain the gradient; stainless steel spacers separate the plates. At the connection between thermoacoustic stack and the photoacoustic resonator, a water cooler removes the excess heat and keeps the lower end of the stack at constant temperature. (Reproduced with permission from Bijnen et al.)[38]

Table 1 A number of photoacoustic trace gas studies with laser type and molecules investigated[a]

(a) CO lasers

CO laser, 5–6.4 μm[41]a	ppb	CO laser, 5–8 μm[42]	ppb	CO laser[11]	ppb
Nitric oxide	15.4	Carbon disulfide	0.01	Ammonia	0.4
Nitrogen dioxide	0.88	Acetaldehyde	0.1	1,3-Butadiene	1
Nitrous oxide	50	Water vapor	0.1	1-Butadiene	2
Carbon dioxide	5×10^4	Carbon dioxide	1×10^3	Nitric oxide	0.4
Water vapor	65	Nitrogen dioxide	0.1	Nitrous oxide	0.1
Ethylene	42	Nitric oxide	0.3	Propylene	3
Vinyl chloride	10.6	Nitrous oxide	1	Water vapor	14
Propylene	47	Acetylene	1		
1,3-Butadiene	22	Ethylene	1		
Formaldehyde	7.7	Ethane	1		
Acetaldehyde	4.5	Methane	1		
Acrolein	2.1	Pentane	3		
Acetone	4.2	Trimethylamine	1		
Benzene	70	Carbonyl sulfide	1		
Toluene	40	Dimethyl sulfide	3		
m-Xylene	19	Sulfur dioxide	0.1		
o-Xylene	70	Methanethiol	3		
p-Xylene	97	Hydrogen sulfide	100		

(b) Various lasers

Laser	Species	c_{min} (ppb)	Ref.
Dye laser 290–310 nm	Sulfur dioxide	0.12	14
Dye laser 570–620 nm	Nitrogen dioxide	10	43
Dye laser 600 nm	Iodine	260	44
Dye laser 600 nm	Bromine	7.9×10^4	44
HeNe laser 3.39 μm	Methane	10	9
HeNe laser 3.39 μm	C_nH_{2n+2}	10^4–10^5	45
CO Spin flip Raman laser	Nitric oxide	0.1	46
CO laser, Zeeman effect	Nitric oxide	0.01	35
CO laser 4.75 μm	Carbon monoxide	150	47
CO laser 5.40 μm	Phosgene	Few	48
CO laser 5.42 μm	Nitric oxide	0.1–1	45
CO laser 6.13 μm	Nitrogen dioxide	0.01–0.1	45
DF laser 3.87 μm	Methane	1.6×10^3	49
$PbS_{1-x}Se_x$4.8 μm	Carbon monoxide	5×10^4	12
InP/InGaAsP 1.13 μm	Water vapor	1×10^4	13
Diode 10.5 μm	Ethene	200	50

(c) CO_2-lasers

	ppb		ppb
Ref. 51		Ref. 58	
Ethylene	46	Hydrazine	0.2
Benzene	2.05×10^3	Monomethylhydrazine	0.3
Ammonia	35	Unsymmetrical-dimethylhydrazine	0.3
Trichloroethylene	0.55×10^3	Dimethylamine	1.2
Freon-12	0.13×10^3	Trimethylamine	0.3
Ethanol	0.36×10^3	Methanol	0.05
Toluene	0.8×10^3	Ref. 61[b]	
Ref. 52		Acrolein	36
Ethylene	0.02	Styrene	36
Ozone	0.06	Ethyl acrylate	14
Hydrogen sulfide	170	Trichloroethylene	7
1,2-Dibromoethane	0.4	Vinyl bromide	25
Ethyl acetate	0.06	Vinylidene chloride	8

Table 1 (*continued*)

(c) CO_2-lasers (*continued*)

	ppb	Ref.
Ref. 62		
Nitroglycerine	0.28	
Ethylene glycol dinitrate	1.5×10^3	
Dinitrotoluene	16×10^3	
Ref. 63	ppb W^{-1}	
Ethylene glycol dinitrate	8.26	
Nitroglycerine	0.23	
2,4-Dinitrotoluene	0.5	
Cyclohexanone	20.2	
Methylamine	25.3	
Ethylene	0.3	53
Ammonia	1	54[c]
Nitrous oxide	2×10^4	55
Sulfur hexafluoride	0.01	56
Ethylene, Ammonia, Freons, etc.	≤ 4	57
Carbon dioxide	7.4×10^3	31
Ammonia	0.43	31
Ethylene	1.3	31
Ethylene	1–10	45
Ammonia	1–10	45
	Ref. 59[c] 10^{-20} cm^2	Ref. 60[c] ppb
Ethylene	130	0.7
Water vapor	0.005	2×10^4
Propylene	8.2	27
Vinyl chloride	29	2.6
Benzene	7.2	7.2
Toluene	4.5	20
Methanol	72	1.4
Ethanol	17	6
Carbon dioxide	0.015	
1,3-Butadiene	28	
o-Xylene	4.3	
m-Xylene	2.8	
p-Xylene	2.7	
Ethylbenzene	2.3	
Dichloromethane	0.17	
Chlorobenzene	13	
o-Dichlorobenzene	21	
m-Dichlorobenzene	11	
p-Dichlorobenzene	5.5	

	$atm^{-1}cm^{-1}$
Ref. 64	
Sulfur hexafluoride	646
Tungsten hexafluoride	0.091
Uranium hexafluoride	8.5×10^{-4}

	$atm^{-1}cm^{-1}$
Ref. 65	
Acetonenitrile	0.75
Ammonia	69.5
Arsine	5.3
Benzene	2.7
Butadiene	4.4
Ethrane	21.7
Ethylenechlorohydrin	9.3
Ethylene oxide	0.64
	3.4[b]
Formic acid	17.1
Freon-12	99.8
Germane	1.9
Hydrogen selenide	0.5
Perchloroethylene	25.9
Phosgene	1.1
Phosphine	4.0
Styrene	56
	6.0[b]
Sulfur dioxide	0.24
Sulfuryl fluoride	0.19
	53.3[b]
Tricloroethane 111	24.0
Trichloroethylene	15.5
Vinyl chloride	6.6
Ref. 66	
Freon-13	1.24[b]
Freon-22	0.74[b]
Freon-12	79.37[b]
Freon-11	4.23[b]
Ammonia	40.87[b]
Ref. 29	ppb
Acetonitrile	670
Benzene	48
Cyclohexane	250
1,2-Dichloroethane	450
Ethyl acetate	8.3
Freon-12	5.5
Freon-113	5.0
Furan	25
Isopropanol	29
Methylchloroform	11
Methyl ethyl ketone	83
t-Butanol	26
Vinyl chloride	12
Iodopropane	360
Acetylene	3
Ammonia	0.8
Ozone	9
Perchloroethylene	3.2
Trichloroethylene	13

(*continued overleaf*)

Table 1 (*continued*)

(c) CO_2-lasers (*continued*)

	$atm^{-1}cm^{-1}$		$atm^{-1}cm^{-1}$
Sulfur hexafluoride	0.2	Butane	200
1,1-difluoroethylene	5.3	Ethylene glycol dinitrate	38
Freon-11	18	Nitroglycerine	0.7
Freon-114	3.6	Methylamine	120
Ethylene	2.6		

[a] Unless indicated otherwise, detection limits are reported for single-component mixtures with N_2 at 1 atmosphere as buffer gas. Data are given in ppb (1 part per billion volume $= 1 \times 10^9$). In a number of references only absorption strengths are given: σ (in cm^2) or α (in $atm^{-1}cm^{-1}$), with $\alpha = N_{tot} \cdot \sigma$ ($N_{tot} = 2.5 \times 10^{19}$ molecules cm^{-3} at 1013 hPa and 20 °C). When the detection system has a sensitivity of typically $\alpha_{min} = 1 \times 10^{-8}$ the minimum detectable gas concentration can be calculated from $c_{min} = \alpha_{min}/(\sigma \cdot N_{tot})$.
[b] At $^{13}CO_2$ laser lines.
[c] Multicomponent study.

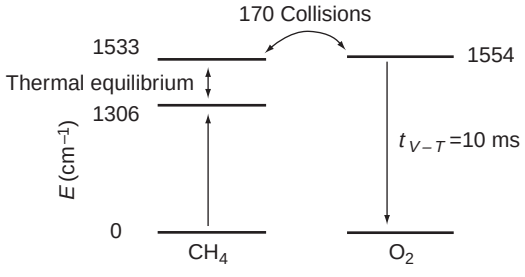

Figure 7 Vibrational energy transfer due to near-resonant energy levels and a slow V–T relexation process, in comparison with the modulation frequency (1010 Hz), may lead to an amplitude decrease and phase shift of the photoacoustic signal. This effect can, for instance, be observed when comparing mixtures of methane in nitrogen and in air, due to rapid vibrational energy transfer between methane and oxygen (see text for details).

the ν_4 vibrational mode centered around 1306.2 cm^{-1}. This is the lowest vibrational mode of the molecule; the collisional relaxation lifetime is therefore large compared with the lifetime of other molecules with a smaller energy gap between their lowest vibrational modes and the ground state. The adjacent ν_2 mode (1533.3 cm^{-1}) becomes thermally populated too (rate constant 13 μs^{-1} atm^{-1}).[71] Oxygen has a near resonant vibrational level (1554 cm^{-1}) and only 170 collisions are needed to transfer energy to the vibrational mode of O_2 (rate constant $\approx$28 μs^{-1} atm^{-1}).[72] Since this is a rather fast process most energy is deposited into O_2 and then relaxes slowly to the ground vibrational state; the average number of collisions required for relaxation to the ground state is 8.3×10^7 in pure O_2 (at standard temperature and pressure 5×10^9 collisions per second take place). The much larger number density of O_2 compared with CH_4 creates a buffer of vibrational energy leading to an effective transient cooling of the translational degrees of freedom. Due to our high modulation frequency (1010 Hz), we observed this transient cooling by O_2 as an amplitude decrease and a phase change relative to the signal caused by the same amount of CH_4 in N_2.

For specific molecules (such as NH_3, NO) spectroscopic interference can be avoided by combining the photoacoustic effect with the Stark or Zeeman effect; the selectivity is enhanced by intermodulating chopper and electric or magnetic fields. A more generally applicable method is to separate gases, either by gas chromatographic methods, selective trapping inside a cold trap,[26] or by a specific chemical reaction (e.g. CO_2 by KOH $\longrightarrow$ K_2CO_3 and water). In most cases, a small amount of the interfering compound remains present. To take this into account, a multicomponent analysis is necessary selecting a number of laser lines on which interfering species and the species of interest have both strong and weak absorptions.

For a multicomponent analysis of a gas mixture, the laser is subsequently tuned to L laser lines. For the analysis it is preferable to obtain an overdetermined system, $G < L$ with G being the number of gases; in this way, more information can be obtained from the gas mixture resulting in more accurate results.[60,73] For practical convenience the background signal generated by window absorption and/or resonator wall absorption is added as a fictive gas component. In the ideal case the measured spectrum is equal to the calculated spectrum. If the absorption coefficient $_{lg}$ of a gas g at laser line l is known, the total absorption at a laser line, α_l is given by Equation (6)

$$\alpha_l = \sum_{g=1}^{G} \Sigma_{lg} \cdot c_g \tag{6}$$

with c_g the concentration of component g. In matrix notation one writes Equation (7)

$$\vec{\alpha} = \vec{\vec{\Sigma}} \cdot \vec{c} \tag{7}$$

If no measurement errors are specified in the absorption coefficient matrix $\vec{\vec{\Sigma}}$ and the measured spectrum ($\tilde{\alpha_l}$), the gas concentration $\vec{c}$ can be calculated by[26] Equation (8)

$$\vec{c} = (\vec{\vec{\Psi}}^{\,t} \cdot \vec{\vec{\Psi}})^{-1} \cdot \vec{\vec{\Psi}}^{\,t} \cdot \vec{1} \tag{8}$$

where $\vec{\vec{\Psi}}^{\,t}$ is the transposed matrix of Ψ_{ij}, $\Psi_{ij} =$ $_{ij}/\alpha_i$ and $1_i = 1$

4 ENVIRONMENTAL APPLICATIONS

4.1 Stack Gas Emission

Several attempts have been made to measure stack gas emissions from power plants. Due to the high amount of nitrogen oxide (NO and NO_2) containing compounds in the emissions, they contribute significantly to photochemical smog formation and acidification of the soil. To reduce the total amount of NO_x in the stack gas, NH_3 is added in the exhaust gas towards the chimney. Using a voluminous catalyst, NO_x reduction within the stack gas takes place. In order to check the performance and to avoid an excess of ammonia injection, the ammonia concentration in the chimney is determined. Olafsson et al. have successfully applied a CO_2 laser-based photoacoustic detection system for in situ monitoring of ammonia concentrations.[74] Due to the difficult hostile environment halfway up the chimney (vibrations, temperature fluctuations, etc.), concessions have to be made as to the sensitivity of the apparatus. With a nonresonant photoacoustic cell at 125 °C, a detection limit of 1 ppm NH_3 was achieved in a multicomponent gas mixture containing 10–15% CO_2.

Gas detection using pulsed lasers in combination with photoacoustic spectroscopy in the UV (ultraviolet) and visible has been developed by Stenberg et al. to analyze gas concentrations in fluidized beds and other combustion environments.[75] The probe has been calibrated for typical combustion gases such as N_2O, NO, NO_2, NH_3, SO_2 and H_2S at temperatures between 20 and 910 °C.

Another example of the potential of the photoacoustic systems to control industrial processes is given by Sigrist[22] who was able to follow the ethanol and methanol emissions from a pharmaceutical production plant with a CO_2 laser-based set-up with a time resolution of 5 min. With this time resolution, strong concentration fluctuations were observed in the exhaust emission, whereas the gas chromatographic system which was used for comparison proved inadequate.

4.2 Car Exhaust Emissions

CO and CO_2 lasers were used to analyze the air polluting emissions from car exhausts. A thorough study was performed by Bernegger and Sigrist to unravel car exhaust mixtures quantifying CO_2, NO and 12 hydrocarbons (see Table 2).[76] To this end, absorption cross-sections of the individual species at each laser line were measured; certified trace gas mixtures in a buffer (mostly N_2) or other certified methods were used to quantify the absolute values. Once the absorption cross-sections are determined, care has to taken to avoid saturation effects (i.e. nonlinearity of the photoacoustic effect with the laser power[52]) and phase shift effects (see previous section). From the absorption cross-sections detection limits can be obtained. In the literature, extremely low detection limits are mentioned based on a signal-to-noise-ratio of one ($S/N = 1$, see Table 1), extrapolated from larger quantities of trace gases in a buffer gas. In reality the detection limits are higher due to multicomponent gas mixtures which induce cross-sensitivities in the absorption coefficients. This mathematical analysis of the photoacoustic spectra is based on the weighted least squares fit of the measured spectra with iteration steps. By choosing the best set of laser lines the error in the calculated concentrations can be reduced.[70,77]

Car exhaust gas mixtures have been investigated by collecting exhaust samples from different motorcycles, diesel and gasoline engines equipped with catalytic converter. The photoacoustic measurements were compared with conventional gas analytical methods. The engine exhaust stream was mixed with air and a fraction of the mixture was collected in a Tedlar sampling bag. Due to the strong water vapor absorption at the CO laser wavelengths and the high water vapor concentrations of 1–2% in the sampling bag, the gas was flowed over a dry-ice cold trap before analysis thereby reducing the water vapor concentration to 0.1%.[76]

4.3 Ambient Air Monitoring

Examples of laboratory photoacoustic studies on multicomponent studies can be found elsewhere.[22] Here we mention two examples of mobile systems which were developed and used in field campaigns. A complete laser photoacoustic system which fitted into two medium sized boxes (0.3 m^3 each) was developed by Rooth et al. to detect ammonia gradients above the heath in the Netherlands.[54] The potential danger of nitrification by NH_3 of such nutrition poor areas motivated the measuring campaign to follow the deposition/emission ratio of ammonia over the area. The system operated for several months, taking data for water vapor, CO_2 (every 40 min) and ammonia (every 6 min).

Another system was developed by Sigrist et al. Installing the whole set-up in a small trailer the stress on the equipment is less severe. Thanks to this approach the system has been operational for years and has been

Table 2 CO laser photoacoustic analyses of vehicle exhausts collected at idling operation of the engine[a] (after Bernegger and Sigrist)[76]

Vehicle	Dried	CO_2(%)	NO	Ethylene	Propylene	Benzene	Toluene	*m*-Xylene	*o*-Xylene	*p*-Xylene	Formal-dehyde	Acetal-dehyde	Acrolein
Volvo (leaded)	Yes	10.3 ± 2.5	105 ± 9	260 ± 20	191 ± 9	37 ± 5	26 ± 14	16 ± 10	42 ± 18	4 ± 19	5.5 ± 1.9	2.7 ± 0.8	2.5 ± 0.4
Jeep	Yes	9.1 ± 1.1	23 ± 2	191 ± 11	109 ± 3	48 ± 3	268 ± 9	67 ± 7	114 ± 10	7 ± 8	1.2 ± 0.6	2.6 ± 0.2	1.2 ± 0.1
(unleaded)	No	11.1 ± 1.3	40 ± 4	206 ± 13	107 ± 5	78 ± 4	196 ± 13	188 ± 12	103 ± 18	67 ± 13	0.1 ± 1.5	4.1 ± 0.3	2.4 ± 0.2
VW Bus	Yes	10.8 ± 1.1	11.8 ± 1.5	172 ± 10	109 ± 4	34 ± 3	149 ± 7	31 ± 7	59 ± 10	2.8 ± 8	0.1 ± 0.8	3.6 ± 0.3	1.7 ± 0.2
(cold catalytic convertor)	No	11.1 ± 1.5	27 ± 3	178 ± 14	98 ± 13	45 ± 5	87 ± 15	39 ± 14	33 ± 19	11 ± 11	1.2 ± 2.6	3.4 ± 0.9	1.9 ± 0.4
Renault Master	Yes	2.8 ± 0.3	10.4 ± 0.9	32 ± 3	18.2 ± 1.2	2.7 ± 0.7	18.5 ± 1.9	4.0 ± 1.7	7.7 ± 2.6	0.5 ± 2.5	8.1 ± 0.8	7.6 ± 0.4	3.2 ± 0.2
(diesel)	No	2.7 ± 0.6	24.5 ± 2.2	33 ± 6	14 ± 8	0.3 ± 2.8	16 ± 7	6.4 ± 6	5.6 ± 10	1.2 ± 6	7.1 ± 2.5	8.7 ± 1.0	4.4 ± 0.5
Unimog	Yes	2.6 ± 0.4	15.7 ± 2.2	5 ± 3	5 ± 2	2.0 ± 0.8	6 ± 2	2.5 ± 2	1 ± 3	5 ± 3	2.6 ± 0.7	2.5 ± 0.3	1.3 ± 0.1
(diesel)	No	2.2 ± 1.0	43 ± 5	7 ± 10	–	–	–	3.5 ± 4	–	8 ± 8	3 ± 7	0.9 ± 2	1.5 ± 1.2
FBW Truck	Yes	1.9 ± 0.4	37 ± 3	16 ± 3	17.5 ± 1.3	–	12 ± 2	1.1 ± 1.5	0.3 ± 3	–	4.2 ± 0.6	7.4 ± 0.2	3.4 ± 0.1
(diesel)	No	2.7 ± 0.9	78 ± 6	23 ± 10	23 ± 9	2 ± 5	20 ± 10	6 ± 7	11 ± 12	3 ± 10	3 ± 2	8.4 ± 0.5	6.9 ± 0.3
Puch motorcycle	Yes	7.9 ± 3	0.3 ± 5	220 ± 10	329 ± 7	203 ± 7	700 ± 20	47 ± 18	135 ± 15	15 ± 10	15.9 ± 1.8	4.1 ± 0.7	6.2 ± 0.3
(2-stroke)	No	8.5 ± 3	0.1 ± 2	234 ± 10	284 ± 7	217 ± 6	405 ± 16	95 ± 13	148 ± 15	20 ± 9	12.6 ± 2.2	5.2 ± 0.8	6.5 ± 0.3

[a] All data are given in parts per million (ppm) with their uncertainty, except for CO_2 (which is in %). The effect of drying of the gas could be observed.

applied in several field campaigns in urban and rural environments.[78] For example, multicomponent gas mixtures were analyzed during sunny days in the summer, using nine laser lines, for the concentration of H_2O, CO_2, NH_3, O_3 and C_2H_4 with a 5 min time resolution.

5 BIOLOGICAL AND MEDICAL APPLICATIONS

5.1 Postharvest Physiology

To delay the effects of aging and ripening, many harvested crops are stored under reduced oxygen concentrations thereby lowering the respiration rate and the rate at which metabolic processes take place. However, when the oxygen concentration falls below a certain limit, crops switch to fermentation, an alternative method to generate the energy needed to sustain their vital functions. This anaerobic metabolism, converts glucose, via several intermediate steps, into acetaldehyde which is then quickly reduced to ethanol. As accumulation of these products in the tissue may affect the quality of the crop, it is necessary to control the rate of fermentation during storage. The concentrations of the volatile in the headspace is an accurate quantitative indicator to assess the rate of fermentation.

Another gas of interest during crop storage is the plant hormone ethylene, which is mainly emitted during ripening of climacteric fruits, i.e. fruits which show a sudden rise in respiratory rate just prior to full ripening. Furthermore, carbon dioxide may be analyzed to assess both respiration and fermentation rates and ethane may be analyzed as the end product of lipid peroxidation causing breakdown of cell membranes.

The noninvasive nature of the photoacoustic trace detection technique renders it extremely suitable for applications in the field of postharvest physiology. Moreover, the high sensitivity combined with fast analysis allows study of an individual piece of fruit so that metabolic activity may be followed in real time. Here we give an overview of some recently performed experiments on single pieces of harvested crops. In general, the crop is placed in a glass container which is connected to the flow-through system leading the released metabolites to the photoacoustic cells. Storage conditions are simulated by supplying the sample with a premixed flow of oxygen and nitrogen (Figure 8); the storage temperature is controlled using a refrigerator.

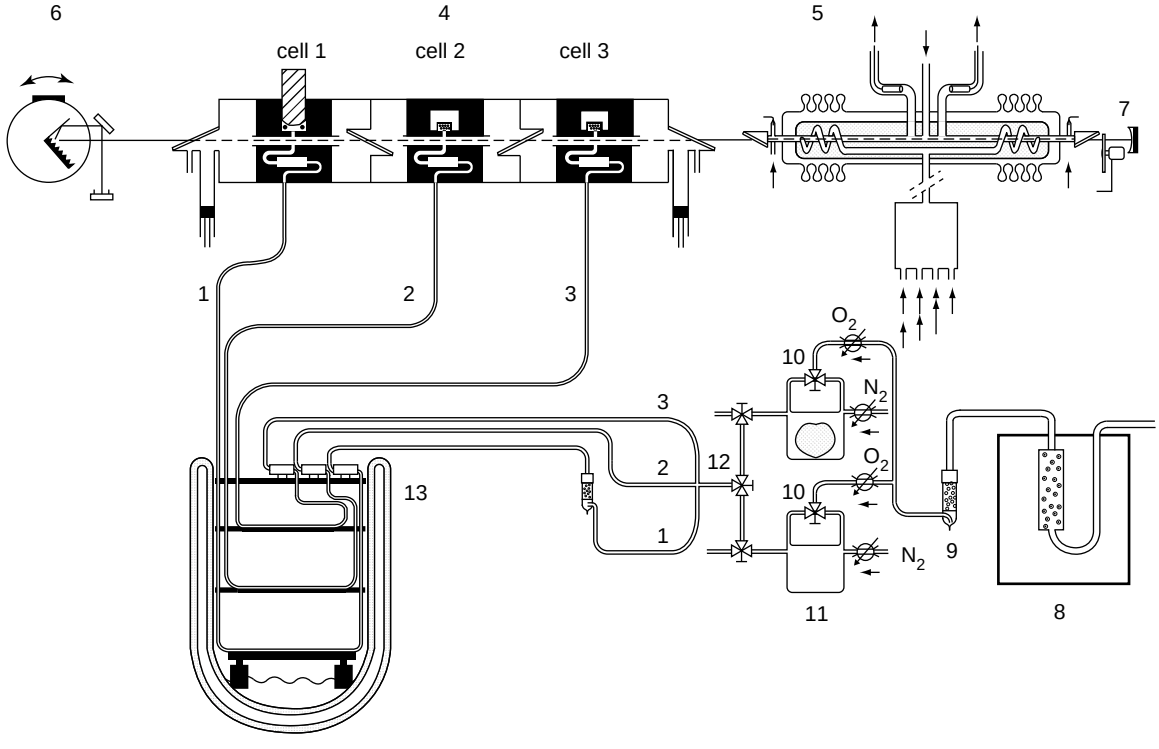

Figure 8 Laser photoacoustic detection set-up. 1, 2 and 3, trace gas flows; 4, triple photoacoustic cell; 5, liquid nitrogen cooled CO laser; 6, grating to select the appropriate transition; 7, chopper; 8, catalyst to remove hydrocarbons from the flow toward the fruit; 9, KOH scrubber to remove CO_2; 10, switching valve for (an)aerobic conditions; 11, cuvettes, one containing a tomato, the other empty as reference; 12, switching valve to select cuvette; 13, cold trap to freeze out interfering gases selectively.

The ethylene emission of climacteric fruit shows a distinct rise during ripening, reaching a maximum when the sample is at its best; subsequent degradation of the product accompanies a decline in the ethylene release. To localize the ethylene release of different crops, De Vries et al. applied the extreme sensitivity of CO_2-laser based photothermal and photoacoustic detectors.[79] Huge differences in emission pathways were found ranging from more than 99% through the skin for banana, apple and citrus fruits, to around 85% through the pedicel "stem" for tomato and bell pepper. In addition, the amount of ethylene released varied from a few tenths to a few thousands of nanoliters per hour per fruit.

When stored crops are returned from controlled atmosphere (CA, low oxygen, high CO_2, low temperature) conditions to normal atmospheric conditions, a rapid peak in the acetaldehyde release occurs due to oxidation of the ethanol accumulated in the tissue. This process was followed in real time for red bell peppers by Zuckermann et al. using a CO laser-based photoacoustic detector.[80] Returning a pepper from anaerobic conditions to normal oxygen conditions (21%) caused a peak in the acetaldehyde release with a maximum of about four times the anaerobic level occurring within 25 min after the switch. The high velocity with which the process takes place caused the authors to conclude that the effect is mainly due to the peroxidative action of the enzyme catalase.[81] Further investigations[82] have shown that the acetaldehyde peak occurs even at much lower postanaerobic O_2 concentrations with roughly equal intensity, though not as fast (Figure 9).

During CA storage of apples it is common practice to apply low concentrations of carbon dioxide externally as it is known to inhibit respiration. As both respiration and fermentation produce carbon dioxide, it is difficult to predict a priori how the fine balance that exists between respiration and fermentation at low O_2 concentrations, will be affected. For pears, at oxygen levels around 1% a substantial rise in the acetaldehyde release could be observed upon addition of a few percent of CO_2. The crop is no longer able to retrieve its energy from respiration and is forced to generate additional energy through fermentation.

5.2 Plant Physiology

One of the major environmental factors influencing crop productivity is flooding. Tolerant species often possess

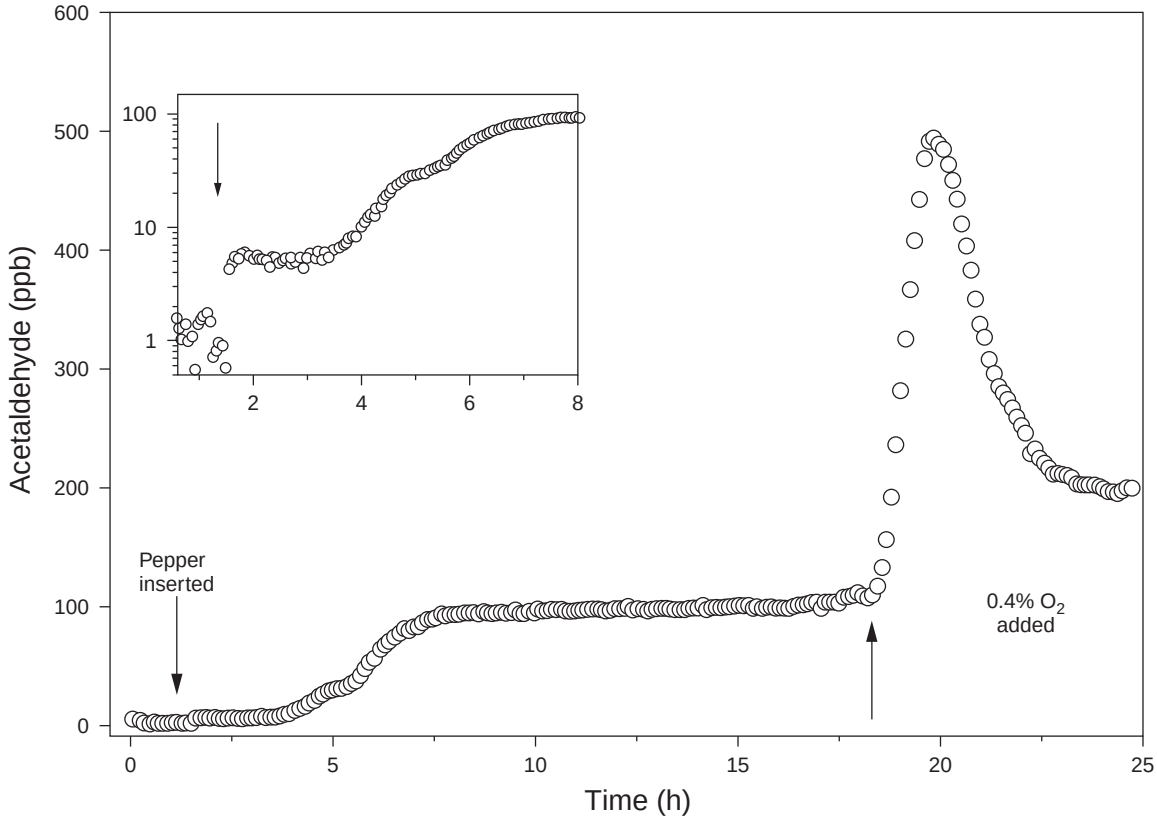

Figure 9 Emission of the precursor of ethanol in the fermentation process, acetaldehyde, is followed on-line while anaerobic conditions are imposed on a red bell pepper. Adding only 0.4% of O_2 results in a strong postanaerobic effect due to the oxidation of ethanol by transient free oxygen radicals in the tissue.

"escape" mechanisms such as fast underwater growth that minimizes the duration of total submergence.[83] However, at early vegetative stages, even in "deep-water" types of rice, such escape is precluded by a lack of sufficiently vigorous shoot elongation for resurfacing. Survival of submergence under these circumstances depends on intrinsic physiological and biochemical tolerance. Screening of cultivars is needed by plant breeders, agrotechnology and biotechnology to achieve sustainable improvements in productivity of rice farming in the rainfed lowland of southeast Asia.[84] Submergence is not an easily quantified stress compared with, for example, cold, heat, or pollution, since it cannot be assessed by one simple analysis. This is because it involves exclusion and entrapment of several key gases and interference with illumination. Normal rates of respiration and photosynthesis are affected; differences in submergence tolerance are also reflected in a different reaction to low oxygen concentrations. Precise measurements of fermentation rates can result in a powerful tool to shed some light on submergence stress. As mentioned in the previous chapter anaerobic fermentation in plant tissue can be studied by use of photoacoustic techniques, through detection of ethanol, acetaldehyde and CO_2 by means of a liquid nitrogen cooled CO laser. Experiments in the laboratory show that seed and seedlings of rice species with different degrees of tolerance toward submergence possess different fermentation rates under anaerobic conditions. Figure 10 shows that after a period of 5 h in which both species have similar ethanol production rates, the most tolerant species (FR13A) starts to slow down fermentation. After about 20 h the production rate for the less tolerant species (CT6241) is doubled with respect to the other; it probably runs out of storage material much faster than the resistant one.

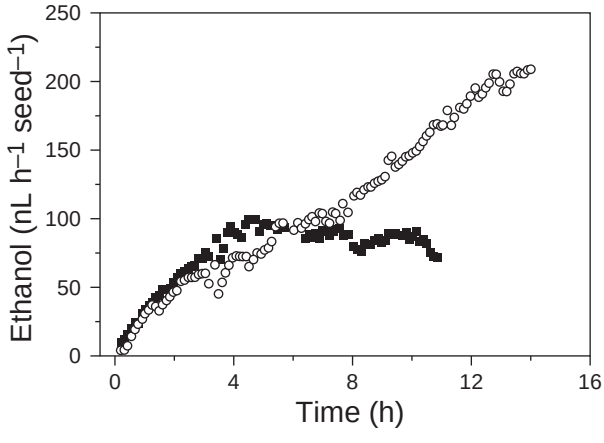

Figure 10 Production rate of ethanol plotted against time for rice seeds during germination under anaerobic conditions. The flooding resistant species (FR13A)(■) does not ferment as much as the intolerant species (CT6241)(○) indicating that it preserves its energy for the postanaerobic period.

Besides fermentation, the availability of oxygen in the roots is of major concern for the plant. In case of partial or complete submergence, rice plants must rely on a well-developed system for gas transport through the plant. The real path of the gas (especially in the shoots) and how this changes with long periods of submergence is unknown. Photoacoustic trace gas detection can also be very useful in this case: a known amount of the tracer gas SF_6 (biologically inactive; atmospheric lifetime 3200 years) is injected into the nutrient solution around the roots and detected coming out of the tip of the leaves, while the rest of the plant is completely submerged in water. We use SF_6 as the tracer gas because the CO_2 laser photoacoustic system is particularly sensitive to this gas (down to parts per trillion level). Figure 11 represents a typical output from the tip of a leaf (enclosed in a cap) after the injection of 4.5 mL of H_2O saturated with SF_6. The plant is kept under water during the experiment.

A question we can readily answer with our system is what kind of gas transport is predominant in the shoots of a young rice seedling. It could be either through aerenchyma (i.e. intercellular spaces present in roots) or through the air layer that is wrapped around the hydrophobic surface of the leaves. This air layer is destroyed by dissolving a few drops of detergent in the water surrounding the plant; because the SF_6 signal dropped to zero immediately afterwards, we can exclude the presence of aerenchyma in the shoots of the seedling.

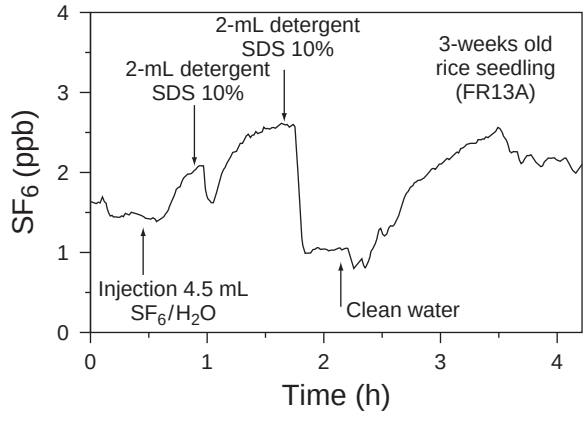

Figure 11 Depletion of the air layer around the submerged rice leaves by addition of detergent. The effect was monitored by SF_6 diffusion which was added into the root compartment (physically separated from the shoot compartment). SF_6 was collected from a small cap filled with air around the tip of a leaf. The air layer is completely restored by replacing the detergent with clean water. During the whole experiment the plant was kept completely submerged.

Another example of trace gas detection within plant physiology is the intriguing question of whether ethylene is a reaction to the growth of the pollen tube inside the pistil (i.e. a form of wounding) or if the pollen tube growth is a consequence of the hormone production by the flower (i.e. a real signal for the plant). It is known that a pollinated flower emits ethylene at the parts per billion level, with a characteristic double peak shaped curve.[85] To answer this question, incongruous pollination has been performed on tobacco flower (*Nicotiana tabacum*), i.e. by pollinating the flower with pollen from other species that show a certain degree of compatibility with the tobacco plant. For this we used pollen with a very low compatibility to the tobacco flower *Nicotiana repanda*; it enters the stigma and stops immediately thereafter. The production of ethylene, detected by photoacoustic spectroscopy, in these cases is very different, as shown in the Figure 12.

The first peak is always reproduced in both height and time and, comparing these results with microscopy work,[86] it corresponds to the entrance of the pollen tube in the stigma. The second peak has a more complicated shape which is never perfectly reproduced with incongruous pollination and which does not simply follow the growth of the pollen tube in the pistil. This might indicate that the second peak is a specific signal from the plant and not a simple wounding effect.

5.3 Microbiology

Biological systems cannot always use molecular nitrogen for their metabolic processes. To this end, dinitrogen

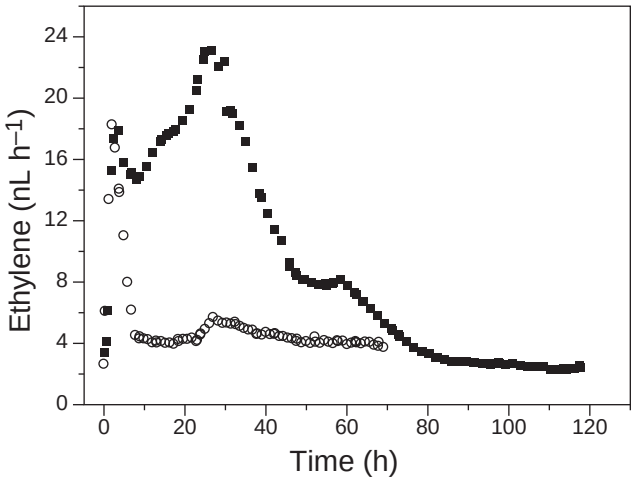

Figure 12 The in vivo production of ethylene during pollination of the *Nicotiana tabacum* flower with two types of pollen: *Nicotiana tabacum* and the incongruous pollen *Nicotiana repanda*. The latter enters the stigma and stops immediately thereafter, whereas the ''right'' pollen grows down the pistil into the ovary for fertilization. ■, *Nicotiana trigonofilla*; ○, *N. repanda*.

(N_2) must be reduced, i.e. the triple bond must be broken to convert the molecule into more appropriate compounds such as ammonium (NH_4^+), nitrate (NO_3^-) or amino acids. Only few organisms are capable of reducing dinitrogen; among them are the diazotrophic cyanobacteria.[87] The nitrogen fixation process is induced by the enzyme nitrogenase, which is very sensitive at low oxygen concentrations. However, like other organisms, cyanobacteria need energy which they derive from photosynthetically produced oxygen. In order to activate both photosynthesis and nitrogen fixation, these processes are separated either in time or in place.[88,89] *Nodularia spumigena* belong to the last category. They consist of large strings of cells formed mainly by vegetative cells in which photosynthesis takes place alternated every 10 to 20 cells by one heterocyst in which the nitrogen fixation takes place.[90]

The acetylene (C_2H_2) reduction method is commonly used to follow the process of nitrogen fixation.[91] As for N_2, the triple bond of acetylene is broken and reduced to ethylene (C_2H_4)[92] which can be detected very sensitively with laser photoacoustics.[52] In the past, gas chromatography utilizing flame ionization detection and photon ionization detection were applied to investigate the nitrogen fixation process.[91,93] Long incubation times, up to several hours, are necessary to reach gas concentrations observable for the equipment making dynamic measurements unfeasible.

Laser photoacoustic experiments have been performed to study the dynamic behavior of nitrogen fixation.[94] To avoid the buffering volume of the water, the algae *Nodularia spumigena* were put on filter paper and a gas mixture of O_2 and N_2 with different concentrations of acetylene was flushed over the sample. In this way the effect of parameters such as light intensity and temperature could be studied with a 20 s time resolution.

Utilizing the CO laser as radiation source instead of the CO_2 laser in the photoacoustic spectrometer, it is possible to detect other interesting gases. Methane and dimethyl sulfide emission in rice field paddies forms an important threat to the environment. Methane is, after CO_2, the second most efficient greenhouse gas whose biological production from the man-made rice cultures, needs further clarification.[95,96] Dimethyl sulfide is a climate-active natural marine emitter and responsible for sulfur transport into the atmosphere. The fact that both gases can be detected very efficiently by photoacoustic techniques, at the water surface as well as at various water depths, forms an important challenge for future activities at field sites.

5.4 Human Health, Noninvasive Breath Analysis

Since ancient times it has been known that the smell of exhaled air can be used as an indicator of several processes

taking place in the human body;[97,98] uncontrolled diabetes produces a sweet fruity odor, advanced liver diseases entail a musty fishy reek, failing kidneys bring about a urine-like smell and a lung abscess can be brought to light by its putrid stench. In the past, several attempts to use trace gas detection of exhaled air have been performed with variable success because of the nonsufficient detection limits of the detectors available. Gases produced in the body are transported to the lungs and are diluted (at rest, healthy persons exhale approximately $1000\,L\,h^{-1}$) before being exhaled. To measure them, it is necessary to enrich the air samples by adsorbing the gases on an agent and releasing the concentrated gas at a certain moment rendering the measurements less accurate and difficult to reproduce.[99]

Measurements on the effects of UV radiation on the human skin are presented to illustrate the possibilities of photoacoustic trace detection in this field. In these experiments lipid peroxidation of the skin is monitored.[100] UV radiation causes reactive oxygen species to be formed in the skin.[101] These radicals can damage the lipids in the cell membranes producing small hydrocarbons such as ethane, pentane and ethylene.[102] With the CO_2 laser set-up it is advantageous to study these latter molecules.[21,52]

A small amount of air is sampled from the exhaled air and cleaned from CO_2 (typically 5%), water vapor and other spectroscopically interfering gases like acetone and ethanol. The test persons are measured while resting under a solar bench. Under UV radiation a steady increase of the exhaled ethylene is observed with a 2 min time delay (Figure 13). The increase continues until the solarium is switched off after 15 min (the maximum exposure time advised). During the subsequent decrease of ethylene emission two decays can be distinguished; a fast and a slow decay. The first decay is caused by the washout of ethylene from the blood; the second decay results from ethylene stored in the body tissue. The dynamics of the system, i.e. transport of gases through the body, yields information for pharmakinetic research. In order to determine the locally induced damage, a specially designed cell was placed on the skin. Immediately after the start of the UV exposure a steady and constant production of ethylene in the skin was observed. Combining the two results, it is concluded that the slow increase in ethylene signal in the exhaled air is caused by the buffering effect of the human body.

Apart from the effect of UV radiation on the skin, other effects induced by lipid peroxidation can be studied: patients suffering from chronic obstructive pulmonary disease have increased cell membrane damage in the lungs; patients with pancreas problems cannot digest long-chain fatty acids, resulting in a lack of ethylene in the exhaled air after consumption of these fatty acids; patients suffering from cardiac insufficiency have increased ethylene levels in the exhaled air.

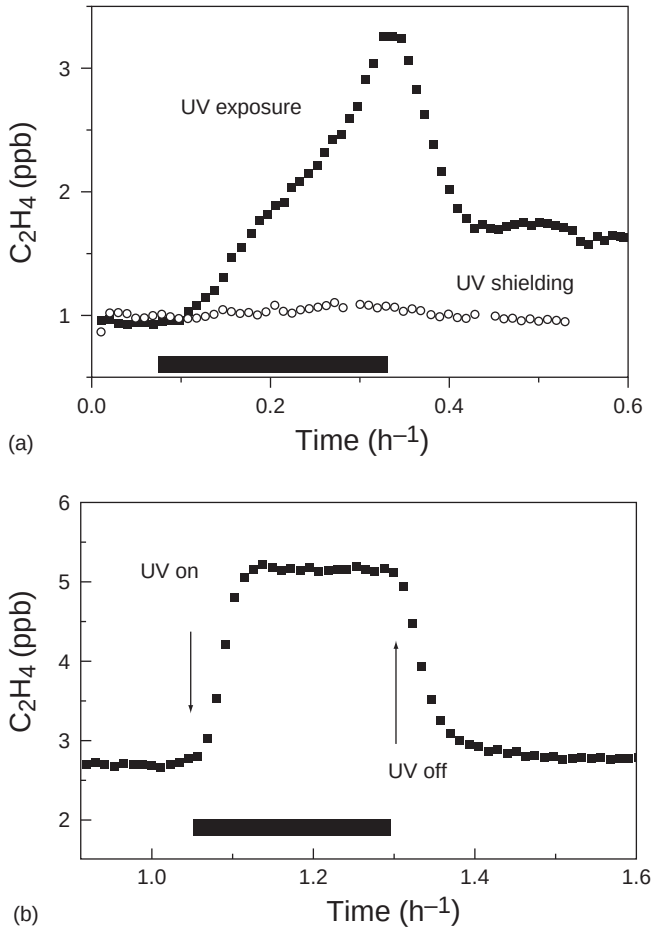

Figure 13 Ethylene emission from the human lungs induced by UV radiation from a solar bench. As a product of lipid peroxidation ethylene is found in low amounts in the exhaled air (a) and directly on the skin (b) utilizing a 5 cm diameter sample cell. The time during which the skin is irradiated (in total 15 min) is indicated by the black bar. Shielding the body from UV radiation using Lexan prevents generation of ethylene.

5.5 Entomology

Methane has been recognized as one of the principal greenhouse gases, second only to CO_2. Its estimated contribution to the enhanced greenhouse effect varies between 15–20% (Intergovernmental Panel on Climate Change, 1990) depending on the time window for which the calculation is made. Methane is not only a potent greenhouse gas, it is also chemically active in the atmosphere thus influencing concentrations of several important species, e.g. OH, O_3 and CO. Biogenic sources are responsible for the major part of the global methane flux (>70%).[103] Some arthropods (i.e. millipedes, cockroaches, termites and scarab beetles)

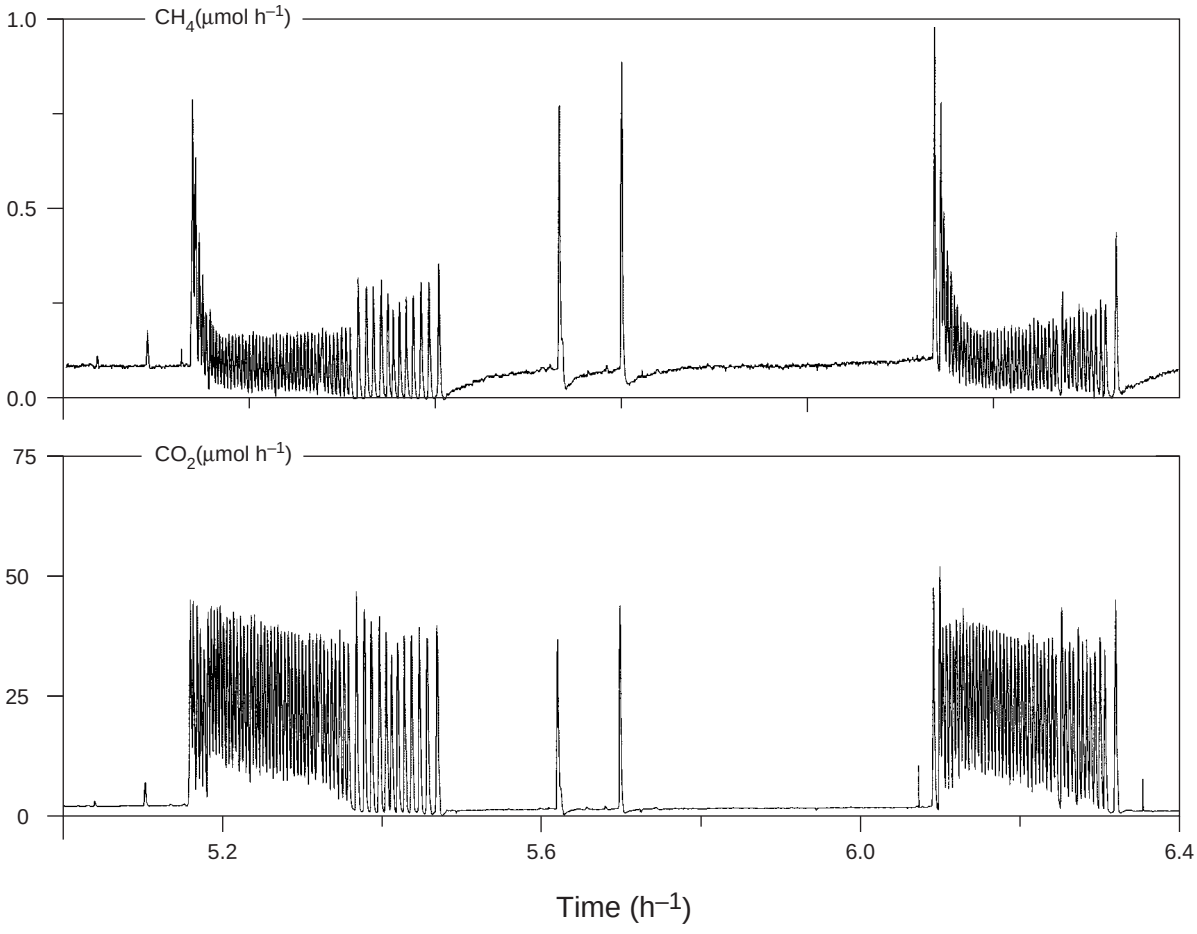

Figure 14 Methane release together with the CO_2 emission, during a complete respiration sequence of breathing (starting at 5.15 h and 6.08 h) and constriction (starting at 5.48 h and 6.32 h) for the cockroach species *Gromphadorhina portentosa*.

are believed to contribute up to 25% of the total budget.[104]

Gas chromatography allows measurement of the methane release of single animals only after incubation for several hours. Consequently these measurements cannot provide information about the dynamic character of the gas emission. Infrared absorption measurements featuring fast response time permitted observation of the dynamics of CO_2 emissions.

Besides CH_4 and CO_2 emissions from insects, water vapor release has been the subject of many studies. Information about water loss dynamics in insects is based upon theoretical modeling, sensitive weighing and relatively slow direct water vapor measurements.[105–108] With laser photoacoustic detection Bijnen et al. showed that cockroaches at rest show a regular breathing pattern that optimizes oxygen uptake and reduces water loss.[26] Respiration patterns of cockroaches and beetles have been recorded where microbiologically produced methane was found to be co-emitted with CO_2 and water vapor (Figure 14). This on-line observation is to be utilized in a study to establish conditions for hyperventilation of insects so that the minimum amount of insecticide will suffice to kill noxious insects which, for example, are known to transmit viral diseases.

6 COMPARISON WITH OTHER SPECTROSCOPIC METHODS IN TRACE GAS MONITORING

Fast and sensitive trace gas detection is not exclusively reserved for photoacoustics. In general, all spectroscopic techniques are able to monitor gas absorptions on a subsecond timescale. However, depending on the type of application not all techniques possess a high sensitivity.

Air pollution can be very well studied with long-path light absorption techniques rendering low detection limits; the sensitivity of the method relies strongly on the path length. A broadband light source (often UV) is directed towards a receiver over an open path in the atmosphere.[22] For stratospheric studies the sun has

also been used as a light source applying a $0.01\,cm^{-1}$ resolution Fourier transform spectrometer on board the space shuttle as part of the Spacelab 3 mission.[109] Laser-based long-path absorption techniques are used either with multipass arrangements, LIDAR (light detection and ranging) or the related DIAL (differential absorption lidar) methods.[22] In contrast to CW operation, using pulsed lasers has the advantage that the backscattered light gives spatial information about the atmospheric absorption pattern.[110]

All methods described above derive their sensitivity from long-path absorption thereby integrating over hundreds of meters. Laser-based photoacoustic detection has the advantage of tracing gases locally and can therefore also be used in laboratory studies. Other laser-based techniques which can also achieve a high sensitivity within a small volume of gas are intracavity laser absorption spectroscopy, which relies on very long effective absorption length within a high optical Q factor laser cavity;[111] cavity ring down spectroscopy combines very accurate decay time measurements of light pulses and a high Q optical cavity resulting in long effective absorption lengths.[112] In comparison with these techniques laser photoacoustics has the advantage that it is background free: it does not rely on a decrease of the transmitted light but on an increase from the zero baseline, i.e. on a collisional release of energy after absorption.

Other background-free and therefore strongly competitive techniques are laser-induced fluorescence (LIF) and multiphoton ionization (MPI). These techniques are both extremely sensitive due to the high collection efficiency of photomultipliers and ion counters; both can be applied locally within a small gas volume. LIF and laser photoacoustics are complementary in energy release. The advantage of LIF is that it can also detect free radicals (e.g. OH) with relatively low laser powers. MPI can be very selective and sensitive for molecules with a higher mass. An elegant example is the detection of OH with LIF in the upper atmospheric layers with an aircraft, avoiding spectroscopic cross interferences from other molecules, laser-induced generation of OH, depletion of OH during sampling, etc.[113] A restriction of LIF is the lifetime of the excited levels. As stated in the introduction, the competition between collisional relaxation and radiative decay restricts LIF mainly to the visible and the UV part of the electromagnetic spectrum.

ACKNOWLEDGMENTS

The authors wish to thank D.H. Parker and J. Reuss for their helpful discussions and continuous support. We are also grateful to H. Zuckermann (Open University of Israel), H.S.M. de Vries, J. Oosterhaven and R.H. Veltman (ATO-DLO) for their collaborations in postharvest research; M.B. Jackson (University of Bristol, UK) for his indispensable contribution in the rice plant research; C. Mariani (University of Nijmegen), D. de Martinis (ENEA, Italy) for their pollination research; M. Staal, L. Stal (NIOO) for their contributions in the field of microbiology; R. Berkelmans, K. Kuiper, P. Hollander (Free University, Amsterdam), P. Scheepers, R. Dekhuijzen (University of Nijmegen) for their help in breath analysis; and P. Kestler (University of Osnabrück, Germany) and J. Hackstein (University of Nijmegen) for their collaborations on entomological applications. Furthermore we thank F. Bijnen, S. Persijn, T. Groot, E. Santosa for their experimental support and C. Sikkens, C. Timmer, H. Schoutissen and L. Gerritsen for technical assistance. Finally we would like to thank the Dutch Technology Foundation (contracts no: NNS44.3404, NNS.4596 and NNS55.3921), the Dutch Royal Academy of Science (contract 95BTM04) and the European Union (contracts no: BRFSCI-CT91-0739, ERB3514-PL95-0708, PL95-0468 and ERB4062-PL970089) for their financial support.

ABBREVIATIONS AND ACRONYMS

CA	Controlled Atmosphere
CW	Continuous Wave
DIAL	Differential Absorption Lidar
FTIR	Fourier Transformed Infrared
LIDAR	Light Detection And Ranging
LIF	Laser-induced Fluorescence
MPI	Multiphoton Ionization
OPO	Optical Parametric Oscillator
PPLN	Periodically Poled Lithium Niobate
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Infrared Spectroscopy in Microbiology • Infrared Spectroscopy, Ex Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air • Spectroscopic Techniques in Industrial Hygiene

Remote Sensing **(Volume 10)**
Satellite and Sensor Systems for Environmental Monitoring

Atomic Spectroscopy **(Volume 11)**
Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Lifetime Measurements, Applications of

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Gas Chromatography/Infrared Spectroscopy • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. A.G. Bell, 'On the Production and Reproduction of Sound by Light', *Am. J. Sci.*, **XX**, 305–324 (1880).
2. A.G. Bell, 'Upon the Production of Sound by Radiant Energy', *Phil. Mag. J. Sci.*, **XI**, 510–528 (1881).
3. W.C. Röntgen, 'Über Töne, welche durch intermittierende Bestrahlung eines Gases entstehen', *Ann. der Phys. und Chem.*, **1**, 155–159 (1881).
4. J. Tyndall, 'Action of an Intermittent Beam of Radiant Heat upon Gaseous Matter', *Proc. R. Soc.*, **31**, 307–317 (1881).
5. W.H. Preece, 'On the Conversion of Radiant Energy into Sonorous Vibrations', *Proc. R. Soc.*, **31**, 506–520 (1881).
6. M.L. Viegerov, 'Eine Methode der Gasanalyse, Beruhend auf der Optisch-Akustischen Tyndall-Röntgenerscheinung', *Dokl. Akad. Nauk SSSR*, **19**, 687–688 (1938).
7. K.F. Luft, 'Über eine Neue Methode der Registrierenden Gasanalyse mit Hilfe der Absorption Ultraroter Strahlen ohne Spectrale Zerlegung', *Z. Tech. Phys.*, **5**, 97–104 (1943).
8. E.L. Kerr, J.G. Atwood, 'The Laser Illuminated Absorptivity Spectrophone: A Method for Measurement of Weak Absorptivity in Gases at Laser Wavelengths', *Appl. Opt.*, **7**, 915–921 (1968).
9. L.B. Kreuzer, 'Ultralow Gas Concentration Infrared Absorption Spectroscopy', *J. Appl. Phys.*, **42**, 2934–2943 (1971).
10. C.K.N. Patel, E.G. Burkhardt, C.A. Lambert, 'Spectroscopic Measurements of Stratospheric Nitric Oxide and Water Vapor', *Science*, **184**, 1173–1176 (1974).
11. L.B. Kreuzer, N.P. Kenyon, C.K.N. Patel, 'Air Pollution: Sensitive Detection of Ten Pollutant Gases by Carbon Monoxide and Carbon Dioxide Lasers', *Science*, **177**, 347–349 (1972).
12. T.H. Vansteenkiste, F.R. Faxvog, D.M. Roessler, 'Photoacoustic Measurement of Carbon Monoxide Using a Semiconductor Diode Laser', *Appl. Spectrosc.*, **35**, 194–196 (1981).
13. Z. Bozóki, J. Sneider, G. Szabó, A. Miklós, M. Serényi, G. Nagy, M. Fehér, 'Intracavity Photoacoustic Gas Detection With an External Cavity Diode Laser', *Appl. Phys. B*, **63**, 399–401 (1996).
14. K.P. Koch, W. Lahmann, 'Optoacoustic Detection of Sulfur Dioxide below the Parts per Billion Level', *Appl. Phys. Lett.*, **32**, 289–291 (1978).
15. G.J. Scherer, K.K. Lehmann, W. Klemperer, 'The High Resolution Visible Overtone Spectrum of CH_4 and CD_3H at 77 K', *J. Chem. Phys.*, **81**, 5319–5324 (1984).
16. O. Vaittinen, M. Saarinen, L. Halonen, I.M. Mills, 'High Resolution Photoacoustic Overtone Spectroscopy of Monofluoroacetylene, HCCF, with a Titanium : Sapphire Ring Laser', *J. Chem. Phys.*, **99**, 3277–3287 (1993).
17. L.E. Myers, W.R. Bosenberg, 'Periodically Poled Lithium Niobate and Quasi-phase-matched Optical Parametric Oscillators', *IEEE J. Quantum Electron.*, **33**, 1663–1672 (1997).
18. W.R. Bosenberg, A. Drobshoff, J.I. Alexander, L.E. Myers, R.L. Beyer, '93% Pump Depletion, 3.5 W Continuous-wave, Single Resonant Optical Parametric Oscillator', *Opt. Lett.*, **21**, 1336–1338 (1996).
19. Ph. Repond, M.W. Sigrist, 'Photoacoustic Spectroscopy on Trace Gases with a Continuously Tunable CO_2 Laser', *Appl. Opt.*, **35**, 4065–4085 (1996).
20. A. Bohren, M.W. Sigrist, 'Optical Parametric Oscillator Based Difference Frequency Laser Source for Photoacoustic Trace Gas Spectroscopy in the 3 μm mid-IR Range', *Infrared Phys. Technol.*, **38**, 423–435 (1997).
21. Frans Harren, Jörg Reuss, 'Photoacoustic Spectroscopy', in *Encyclopedia of Applied Physics*, ed. G.L. Trigg, VCH, Weinheim, 413–435, Vol. 19, 1997.
22. M.W. Sigrist, *Air Monitoring by Spectroscopic Techniques*, J. Wiley & Sons, New York, 1994.

23. A. Karbach, P. Hess, 'High Precision Acoustic Spectroscopy by Laser Excitation of Resonator Modes', *J. Chem. Phys.*, **83**, 1075–1084 (1985).
24. P.M. Morse, K.U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968.
25. F.G.C. Bijnen, J. Reuss, F.J.M. Harren, 'Geometrical Optimization of a Longitudinal Resonant Photoacoustic Cell for Sensitive and Fast Trace Gas Detection', *Rev. Sci. Instrum.*, **67**, 2914–2923 (1996).
26. F.G.C. Bijnen, F.J.M. Harren, J.H.P. Hackstein, J. Reuss, 'Intracavity CO Laser Photoacoustic Trace Gas Detection; Cyclic CH_4, H_2O and CO_2 Emission by Cockroaches and Scarab Beetles', *Appl. Opt.*, **35**, 5357–5368 (1996).
27. R.D. Kamm, 'Detection of Weakly Absorbing Gases Using a Resonant Optoacoustic Method', *J. Appl. Phys.*, **47**, 3550–3558 (1976).
28. A.C. Tam, 'Applications of Photoacoustic Sensing Techniques', *Rev. Mod. Phys.*, **58**, 381–431 (1986).
29. G.A. West, 'Photoacoustic Spectroscopy', *Rev. Sci. Instrum.*, **54**, 797–817 (1983).
30. V.P. Zharov, V.S. Letokhov, *Laser Optoacoustic Spectroscopy*, Springer Verlag, Heidelberg, 1986.
31. G.Z. Angeli, A.M. Solyom, A. Miklos, D.D. Bicanic, 'Calibration of a Windowless Photoacoustic Cell for Detection of Trace Gases', *Anal. Chem.*, **64**, 155–158 (1992).
32. R. Gerlach, N.M. Amer, 'Brewster Window and Windowless Resonant Spectrophones for Intracavity Operation', *Appl. Phys.*, **23**, 319–326 (1980).
33. A. Thöny, M.W. Sigrist, 'Novel Photoacoustic Stark Cell', *Rev. Sci. Instrum.*, **66**, 227–231 (1995).
34. H. Sauren, D. Bicanic, W. Hillen, H. Jalink, C. van Asselt, J. Quist, J. Reuss, 'Resonant Stark Spectrophone as an Enhanced Trace Level Ammonia Concentration Detector: Design and Performance at CO_2 Laser Frequencies', *Appl. Opt.*, **29**, 2679–2682 (1990).
35. M. Seaver, T.J. Manuccia, J.R. McDonald, 'Enhanced Species Selectivity in Opto-acoustic Detection', *Chem. Phys. Lett.*, **123**, 164–168 (1986).
36. H. Jalink, D. Bicanic, 'Concept, Design, and Use of the Photoacoustic Heat Pipe Cell', *Appl. Phys. Lett.*, **55**, 1507–1509 (1989).
37. R. Kästle, M.W. Sigrist, 'Temperature-Dependent Photoacoustic Spectroscopy with a Helmholtz Resonator', *Appl. Phys. B*, **63**, 389–397 (1996).
38. F.G.C. Bijnen, J.V. Dongen, J. Reuss, F.J.M. Harren, 'Thermoacoustic Amplification of Photoacoustic Signal', *Rev. Sci. Instrum.*, **67**, 2317–2324 (1996).
39. F. Kühnemann, K. Schneider, A. Hecker, A. Martis, W. Urban, S. Schiller, J. Mlynek, 'Photoacoustic Trace Gas Detection Using a CW Single Frequency Parametric Oscillator', *Appl. Phys. B*, **66**, 741–745 (1998).
40. J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, 'Quantum Cascade Laser', *Science*, **264**, 553–556 (1994).
41. S. Bernegger, 'CO-laser Photoacoustic Spectroscopy of Gases and Vapours for Trace Gas Analysis', Ph.D. Thesis, ETH, Zürich, 1988.
42. S. te Lintel Hekkert, M.J. Staal, R.H.M. Nabben, H. Zuckermann, S. Persijn, L.J. Stal, L.A.C.J. Voesenek, F.J.M. Harren, J. Reuss, D.H. Parker, 'Laser Photoacoustic Trace Gas Detection, an Extremely Sensitive Technique Applied in Biological Research', *Instrum. Sci. Technol.*, **26**, 157–175 (1998).
43. A.M. Angus, E.E. Marinero, J.J. Colles, 'Opto-acoustic Spectroscopy with a Visible CW Dye Laser', *Opt. Commun.*, **14**, 223–225 (1975).
44. R.A. Keller, N.S. Nogar, D.S. Bomse, 'Photoacoustic Detection of Intracavity Absorption', *Appl. Opt.*, **22**, 3331–3334 (1983).
45. A.L. Gandurin, S.B. Gerasimov, A.A. Zheltukhin, I.P. Konovalov, S.T. Kornilov, G.F. Melnik, Y.Y. Mikhalevich, D.D. Ogurok, V.A. Petrishchev, S.N. Chirikov, 'Optoacoustic Gas Analyzer for NO, NO_2, NH_3, C_2H_4, and Saturated Hydrocarbon Pollutants', *Appl. Spectrosc.*, **45**, 886–890 (1987).
46. C.K. Patel, R.J. Kerl, 'A New Optoacoustic Cell with Improved Performance', *Appl. Phys. Lett.*, **30**, 578–579 (1977).
47. R. Gerlach, N.M. Amer, 'Sensitive Optoacoustic Detection of Carbon Monoxide by Resonance Absorption', *Appl. Phys. Lett.*, **32**, 228–231 (1978).
48. X. Luo, F.Y. Shi, J.X. Lin, 'CO-laser Photoacoustic Detection of Phosgene ($COCl_2$)', *Int. J. Infrared Millimeter Waves*, **12**, 141–147 (1991).
49. D.H. Leslie, G.L. Trusty, 'Measurements of DF Laser Absorption by Methane Using an Intracavity Spectrophone', *Appl. Opt.*, **20**, 1941–1947 (1981).
50. F.J.M. Harren, F. Bijnen, C. Liedenbaum, J. Reuss, L.A.C.J. Voesenek, C.W.P.M. Blom, 'Sensitive Photoacoustic Trace Detection of Ethylene', in *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, eds. R. Grisar, G. Schmidtke, M. Tacke, G. Restelli, Kluwer Academic, Dordrecht, 1989.
51. N. Konjevic, S. Jovicevic, 'Spectrophone Measurements of Air Pollutants Absorption Coefficients at CO_2 Laser Wavelengths', *Spectrosc. Lett.*, **12**, 259–274 (1979).
52. F.J.M. Harren, F.G.C. Bijnen, J. Reuss, L.A.C.J. Voesenek, C.W.P.M. Blom, 'Sensitive Intracavity Photoacoustic Measurements with a CO_2 Waveguide Laser', *Appl. Phys. B*, **50**, 137–144 (1990).
53. E. Kritchman, S. Shtrikman, M. Slatkine, 'Resonant Optoacoustic Cells for Trace Gas Analysis', *J. Opt. Soc. Am.*, **68**, 1257–1271 (1978).
54. R.A. Rooth, A.J.L. Verhage, L.W. Wouters, 'Photoacoustic Measurement of Ammonia in the Atmosphere: Influence of Water Vapor and Carbon Dioxide', *Appl. Opt.*, **29**, 3643–3653 (1990).
55. D.D. Bicanic, A. Bizzarri, B.F.J. Zuidberg, G.P.A. Bot, T. de Jong, H.C.P. Wegh, 'CO_2 Laser Based N_2O Monitor', *Springer Ser. Opt. Sci.*, **58**, 143–144 (1988).

56. E. Nodov, 'Optimization of Resonant Cell Design for Optoacoustic Gas Spectroscopy (H-type)', *Appl. Opt.*, **17**, 1110–1119 (1978).
57. W. Schnell, G. Fischer, 'Carbon Dioxide Laser Absorption Coefficients of Various Air Pollutants', *Appl. Opt.*, **14**, 2058–2059 (1975).
58. G.L. Loper, A.R. Calloway, M.A. Stamps, J.A. Gelbwachs, 'Carbon Dioxide Laser Absorption Spectra and Low ppb Photoacoustic Detection of Hydrazine Fuels', *Appl. Opt.*, **19**, 2726–2734 (1980).
59. P.L. Meyer, M.W. Sigrist, 'Air-pollution Monitoring with a Mobile CO_2 Laser Photoacoustic System', Final Report Swiss National Science Foundation/NFP 14, Project No. 4.684.0.83.14, 1988.
60. P.L. Meyer, M.W. Sigrist, 'Atmospheric Pollution Monitoring Using CO_2-laser Photoacoustic Spectroscopy and Other Techniques', *Rev. Sci. Instrum.*, **61**, 1779–1807 (1990).
61. G.L. Loper, G.R. Sasaki, M.A. Stamps, 'Carbon Dioxide Laser Absorption Spectra of Toxic Industrial Compounds', *Appl. Opt.*, **21**, 1648–1653 (1982).
62. P.C. Claspy, C. Ha, Y.-H. Pao, 'Optoacoustic Detection of NO_2 Using a Pulsed Dye Laser', *Appl. Opt.*, **16**, 2972–2973 (1977).
63. R.A. Crane, 'Laser Optoacoustic Absorption Spectra for Various Explosive Vapours', *Appl. Opt.*, **17**, 2097–2102 (1978).
64. D.M. Cox, A. Gnauck, 'Continuous-wave CO_2 Laser Spectroscopy of SF_6, WF_6, and UF_6', *J. Mol. Spectrosc.*, **81**, 207–215 (1980).
65. M.H. Hubert, J.S. Ryan, R.A. Crane, 'Laser Optoacoustic Detector Measurement of Signatures of a Selection of Environmental Contaminants', Ultra Lasertech Inc., Mississauga, Canada, Rep. No 83-715-1, 1983.
66. Z. Zelinger, I. Jančik, P. Engst, 'Measurement of the NH_3, CCl_2F_2, $CHClF_2$, $CFCl_3$, and $CClF_3$ Absorption Coefficients at Isotopic $^{13}C^{16}O_2$ Laser Wavelengths by Photoacoustic Spectroscopy', *Appl. Opt.*, **31**, 6974–6975 (1992).
67. H.S.M. de Vries, 'Non-intrusive Fruit and Plant Analyses by Laser Photothermal Measurements of Ethylene Emission', in *Fruit and Nut Analyses*, ed. H.F. Linskens, Springer Verlag, Heidelberg, 1–18, 1996.
68. A. Thöny, M.W. Sigrist, 'New Developments in CO_2-laser Photoacoustic Monitoring of Trace Gas', *Infrared Phys. Technol.*, **36**, 585–615 (1995).
69. F.G. Gebhardt, D.C. Smith, 'Kinetic Cooling of a Gas by Absorption of CO_2 Laser Radiation', *Appl. Phys. Lett.*, **20**, 129–132 (1972).
70. M.A. Moeckly, C. Hilbes, M.W. Sigrist, 'Photoacoustic Multicomponent Gas Analysis Using a Levenberg-Marquardt Fit-algorithm', *Appl. Phys. B*, (1998), accepted.
71. E. Avramides, T.F. Hunter, 'Vibrational–Translational/Rotational and Vibrational–Vibrational Processes in Methane: Optoacoustic Measurements', *Chem. Phys.*, **57**, 441–451 (1981).
72. J.D. Lambert, *Vibrational and Rotational Relaxation in Gases*, Clarendon Press, Oxford, 1977.
73. S.B. Tilden, M. Bonner Denton, 'A Comparison of Data Reduction Techniques for Line Excited Optoacoustic Analysis of Mixtures', *Appl. Spectrosc.*, **39**, 1017–1022 (1985).
74. A. Olafsson, M. Hammerich, J. Bülow, J. Henningsen, 'Photoacoustic Detection of NH_3 in Power Plant Emission with a CO_2 Laser', *Appl. Phys. B*, **49**, 91–97 (1989).
75. J. Stenberg, R. Hernberg, J. Vattulainen, 'Analysis of Pollutant Chemistry in Combustion by in situ Pulsed Photoacoustic Laser Diagnostics', *Appl. Opt.*, **34**, 8400–8408 (1995).
76. S. Bernegger, M.W. Sigrist, 'CO-laser Photoacoustic Spectroscopy of Gases and Vapours for Trace Gas Analysis', *Infrared Phys.*, **30**, 375–429 (1990).
77. F.G.C. Bijnen, H. Zuckermann, F.J.M. Harren, J. Reuss, 'Multi-component Trace Gas Analysis by Three Intracavity Photoacoustic Cells in a CO-laser: Observation of Anaerobic and Post Anaerobic Emission of Acetaldehyde and Ethanol in Cherry Tomatoes', *Appl. Opt.*, **37**, 3345–3354 (1998).
78. M.A. Moeckli, M. Fierz, M.W. Sigrist, 'Emission Factors for Ethene and Ammonia from a Tunnel Study with a Photoacoustic Trace Gas Detection System', *Environ. Sci. Technol.*, **30**, 2864–2867 (1996).
79. H.S.M. De Vries, M.A.J. Wasono, F.J.M. Harren, E.J. Woltering, H.C.P.M. van der Valk, J. Reuss, 'Ethylene and CO_2 Emission Rates and Pathways in Harvested Fruits Investigated, in situ, by Laser Photothermal Deflection and Photoacoustic Techniques', *Postharvest Biol. Technol.*, **8**, 1–10 (1996).
80. H. Zuckermann, F.J.M. Harren, J. Reuss, D.H. Parker, 'Dynamics of Acetaldehyde Production During Anoxia and Post-anoxia in Red Bell Pepper Studied by Photoacoustic Techniques', *Plant Physiol.*, **113**, 925–932 (1997).
81. L.S. Monk, R. Brändle, R.M.M. Crawford, 'Catalase Activity and Post-anoxic Injury in Monocotyledonous Species', *J. Exp. Bot.*, **38**, 233–246 (1987).
82. J. Oomens, H. Zuckermann, S. Persijn, D.H. Parker, F.J.M. Harren, 'CO-laser Based Photoacoustic Trace Gas Detection; Applications in Postharvest Physiology', *Appl. Phys. B*, **67**, 459–466 (1998).
83. W. Armstrong, R. Brändle, M.B. Jackson, 'Mechanisms of Flood Tolerance in Plants', *Acta Bot. Neerl.*, **43**, 307–358 (1994).
84. M.B. Jackson, 'Rice for Life: A Physiological Basis for Improving Submergence Tolerance in Rainfed Lowland Rice', First Annual Progress Report, European Commission INCO-DC, Contract: IC18-CT96-00078, 1998.
85. A. Singh, K.B. Evensen, T. Kao, 'Ethylene Synthesis and Floral Senescence following Compatible and Incompatible Pollinations in *Petunia inflata*', *Plant Physiol.*, **99**, 38–45, 1992.

86. T. Kuboyama, C.S. Chung, G. Takeda, 'The Diversity of Interspecific Pollen-pistil Incongruity in *Nicotiana*', *Sex. Plant Reprod.*, **7**, 250–258 (1994).
87. J.M. Lepponen, A. Niemi, I. Rinne, 'Nitrogen Fixation of Cyanobacteria (Blue-Green Algae) and the Nitrogen Cycle of the Baltic Sea', *Symbiosis*, **6**, 181–194 (1988).
88. E. Flores, A. Herrero, 'Assimilatory Nitrogen Metabolism and Its Regulation', in *The Molecular Biology of Cyanobacteria*, ed. D.A. Bryant, Kluwer Academic Publishers, Dordrecht, 487–517, 1994.
89. B. Bergman, J.R. Gallon, A.N. Ray, L.J. Stal, 'N_2 Fixation by Non-heterocystous Cyanobacteria', *FEMS Microbiol. Rev.*, **19**, 139–185 (1997).
90. C.P. Wolk, A. Ernst, J. Elhai, 'Heterocyst Metabolism and Development' in *The Molecular Biology of Cyanobacteria*, ed. D.A. Bryant, Kluwer Academic Publishers, Dordrecht, 769–823, 1994.
91. R.W.F. Hardy, R.D. Holsten, E.K. Jackson, R.C. Burns, 'The Acetylene–Ethylene Assay for N_2 Fixation: Laboratory and Field Evaluation', *Plant Physiol.*, **43**, 1185–1207 (1968).
92. R.J. Flett, R.D. Hamilton, N.E.R. Campbell, 'Aquatic Acetylene-reduction Techniques: Solutions to Several Problems', *Can. J. Microbiol.*, **22**, 43–51 (1975).
93. A. Haystead, R. Robinson, W.D.P. Stewart, 'Nitrogenase Activity in Extracts of Heterocystous and Non-heterocystous Blue-Green Algae', *Arch. Mikrobiol.*, **74**, 235–243 (1970).
94. H. Zuckermann, M. Staal, L.J. Stal, J. Reuss, S. te Lintel Hekkert, F.J.M. Harren, D.H. Parker, 'On Line Monitoring of Nitrogenase Activity in Cyanobacteria by Sensitive Laser Photoacoustic Detection of Ethylene', *Appl. Environ. Microbiol.*, **63**, 4243–4251 (1997).
95. M.O. Andreae, 'Ocean–Atmosphere Interactions in the Global Biogeochemical Sulfur Cycle', *Mar. Chem.*, **30**, 1–29 (1990).
96. T.S. Bates, B.K. Lamb, 'Natural Sulfur Emissions to the Atmosphere of the Continental United States', *Glob. Biogeochem. Cyc.*, **6**, 431–436 (1992).
97. M. Philips, 'Breath Tests in Medicine', *Sci. Am.*, July, 52–57 (1992).
98. B. Krotozinsky, G. Gabriel, H. O'Niell, 'Characterization of Human Expired Air: A Promising Investigative and Diagnostic Technique', *J. Chromatogr. Sci.*, **15**, 239–244 (1977).
99. H. Esterbauer, 'Estimation of Peroxidative Damage. A Critical Review', *Pathol. Biol.*, **44**, 25–28 (1996).
100. C.M.F. Kneepkens, G. Lepage, C.C. Roy, 'The Potential of the Hydrocarbon Breath Test as a Measure of Lipid Peroxidation', *Free Radical Biol. & Med.*, **17**, 127–160 (1994).
101. F.R. de Gruijl, 'Health Effects from Solar UV Radiation', *Radiat. Prot. Dosim.*, **72**, 177–196 (1997).
102. A. van Gossum, J. Decuyper, 'Breath Alkanes as an Index of Lipid Peroxidation', *Eur. Respir. J.*, **2**, 787–791 (1989).
103. J.E. Rogers, W.B. Whitman, *Microbial Production and Consumption of Greenhouse Gases*, American Society for Microbiology, Washington DC, USA, 7–38, 1991.
104. J.H.P. Hackstein, C.K. Stumm, 'Methane Production in Terrestrial Arthropods', *Proc. Natl. Acad. Sci.*, USA, **91**, 5441–5445 (1994).
105. P. Kestler, *Comparative Biochemistry and Physiology of Insects*, ed. K.H. Hoffmann, Springer Verlag, Berlin, 137–183, 1985.
106. M.C. Quinlan, N.F. Hadley, 'New System for Concurrent Measurement of Respiration and Water Loss in Arthropods', *J. Exp. Zool.*, **222**, 255–263 (1982).
107. J.R.B. Lighton, D. Garrigan, F.D. Duncan, R.A. Johnson, 'Respiratory Water Loss During Discontinuous Ventilation in Queens of the Harvester Ant *Pogonomyrmex Rugosus*', *J. Exp Biol.*, **179**, 233–244 (1993).
108. E.B. Edney, *Water Balance in Land Arthropods*, Springer Verlag, New York, 227–238, 1977.
109. C.P. Rinsland, E. Mahieu, R. Zander, M.R. Gunson, R.J. Salawitch, A.Y. Chang, A. Goldman, M.C. Abrams, M.M. Abbas, M.J. Newchurch, F.W. Irion, 'Trends of OCS, HCN, SF_6, $CHClF_2$ (HCFC-22) in the Lower Stratosphere from 1985 and 1994 Atmospheric Trace Molecule Spectroscopy Experiment Measurements near 30°N Latitude', *Geophys. Res. Lett.*, **23**, 2349–2352 (1996).
110. S. Svanberg, 'Environmental Monitoring Using Optical Techniques', in *Applied Laser Spectroscopy*, eds. W. Demtröder, M. Inguscio, Plenum, New York, 417–434, 1991.
111 A. Campargue, F. Stoeckel, M. Chenevier, 'High-sensitivity Intracavity Laser Spectroscopy – Applications to the Study of Overtone Transitions in the Visible Range', *Spectrochim. Acta Rev.*, **13**, 69–88 (1990).
112. P. Zalicki, R.N. Zare, 'Cavity Ring-down Spectroscopy for Quantitative Absorption Measurements', *J. Chem. Phys.*, **102**, 2708–2717 (1990).
113. P.O. Wennberg, R.C. Cohen, N.L. Hazen, L.B. Lapson, N.T. Allen, T.F. Hanisco, J.F. Oliver, N.W. Lanham, J.N. Demusz, J.G. Anderson, 'Aircraft-borne, Laser-induced Fluorescence Instrument for the in situ Detection of Hydroxyl and Hydroperoxyl Radicals', *Rev. Sci. Instrum.*, **65**, 1858–1876 (1994).

Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

J.-P. Wolf
Université Claude-Bernard Lyon 1, Villeurbanne Cedex, France

1 Introduction 1
2 The Differential Absorption Light Detection and Ranging Technique 2
3 Differential Absorption Light Detection and Ranging Systems 3
4 Examples of Three-dimensional Mapping of Gaseous Pollutants 5
5 Detection of Aerosols and Dust Particles 13
6 Validation of Numerical Models 14
7 Comparison with Other Techniques and Calibration Procedures 15
8 New Light Detection and Ranging Standards 17
9 Conclusion 17
Acknowledgments 17
Abbreviations and Acronyms 17
Related Articles 18
References 18

Recent progress in light detection and ranging (LIDAR) and differential absorption light detection and ranging (DIAL) technology has allowed three-dimensional (3-D) mappings of concentrations of air pollutants to be obtained at the highest sensitivity (parts per billion) and over large distances (several kilometers). It is now possible to monitor on-line spatial distributions of nitrogen oxides, sulfur dioxide, ozone, benzene, toluene, and aerosols. With the advent of new tunable *all-solid-state laser systems, such as vibronic lasers and optical parametric oscillators (OPOs), LIDAR systems have furthermore reached the highest industrial standards in terms of reliability and ease of operation. An overview of results and capabilities of ''all-solid-state'' LIDAR/DIAL systems is presented. The examples of application include detailed urban studies performed in large cities (e.g. Berlin, Leipzig, Lyon, Stuttgart, Geneva, Athens, Seville), with special attention to the first direct access to smog formation processes and tropospheric ozone generation. Industrial applications are also numerous, such as site monitoring, identification of diffuse emissions, and evaluation of the impact of new emitters through plume dispersion studies. The use of LIDAR systems to validate numerical models is also discussed. The last section is dedicated to the calibration of the systems, intercomparison with standard analyzers, and legal aspects in terms of certification and new standards.*

1 INTRODUCTION

Air pollution is an extremely dynamic phenomenon, which makes its understanding and therefore its control difficult. This dynamic behavior appears in physical terms in the diffusion and transport of emitted pollutants and chemically through the many reactions which take place in the atmosphere. It is therefore of outstanding importance to be able to correlate emission and immission, and characterize the impact of different kinds of sources of pollution (industries, vehicles, domestic heaters) on the environment. Currently existing devices, however, although very sensitive, only provide spot measurements at the ground level. 3-D information, reflecting the dynamic character of pollution, has been for a long time sorely lacking.

These considerations have induced the development of the LIDAR technique, and in particular DIAL [also called differential absorption and scattering (DAS)]. These devices indeed allow selective measurements of pollutant concentrations over several kilometers, range-resolved like radio detection and ranging (RADAR), and in an interactive manner (i.e. without the usual need for samples). It becomes then possible to obtain ''3-D maps'' of concentrations, which show the propagation, spread, and chemical evolution of emitted pollution.

The DIAL technique was demonstrated for the first time in 1966 by Schotland,[1] who evaluated vertical profiles of water vapor contained in the atmosphere. The first mapping experiments were performed in 1974 by Rothe et al.[2] In the 1980s, the performances of the systems were widely extended by several research groups, essentially in Sweden,[3–7] England,[8,9] Germany,[10,11] France,[12–14] and the United States.[15–17]

However, LIDAR systems remained in the laboratory until very recently. The breakthrough to industrialization and standard use by the Environmental Protection Agencies (EPAs) has been allowed by different conjugated progress. First, the capabilities of today's DIAL systems almost perfectly match the needs and requirements of air enforcement regulations. In particular, LIDAR provides 3-D maps of concentrations of nitrogen oxides, sulfur dioxide, ozone, toluene, and benzene at the highest sensitivity (parts per billion range), over large distances, and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

this at much higher resolution and lower cost than standard networks. It is also, moreover, able to monitor the time evolution of pollutant concentration distributions, at high temporal resolution (some minutes) or in longer time frames.

Optical techniques are moreover often recognized as delivering more representative data than spot measurements: an integrated value over a defined optical path (ranging from few meters to several hundred meters) always contains more information than precise spot data at a wrong location! These arguments brought standardization commissions to introduce optical techniques into legislation. Pioneering work has been performed in this respect by the German commission Verein Deutscher Ingenieure (VDI)/Deutsche Institut für Normung (DIN).[18,19]

The most recent and decisive progress is a technological step forward, which makes the new DIAL systems very reliable devices, reaching the highest industrial standards. The advent of new *tunable* all-solid-state laser systems, such as vibronic lasers[20–22] or laser-pumped OPOs[23,24] has opened up a new era in the domain of user-friendly and fully automatic DIAL operation.

''All-solid-state'' DIAL systems, based on flashlamp-pumped Ti : sapphire lasers, have recently been very successful.[21,25,26] The development of the laser used was fully dedicated to LIDAR applications, in order to reach the specific requirements such as high energy, dual-wavelength operation, extreme reliability, and user-friendliness. The wavelength range accessible by the laser and its extension by non-linear optical devices make it an ideal tool for both DIAL and meteorological applications.

DIAL instruments can be classified in two different categories, depending on the application: stationary and mobile. The stationary systems, such as those installed in Berlin and Leipzig,[20] are dedicated to routine 3-D mappings over the city, whereas mobile units are mainly used for campaigns. Numerous campaigns have been performed in large cities such as Zürich,[27] Stuttgart, Lyons, Geneva,[28] Berlin,[29] Athens,[30,31] Seville,[32,33] and Mexico.

This article discusses the advantages offered by DIAL/LIDAR instruments in analyzing gaseous and particulate pollutants, examples of application, and norming activities.

2 THE DIFFERENTIAL ABSORPTION LIGHT DETECTION AND RANGING TECHNIQUE

When a laser beam is sent into the atmosphere, it is scattered in every direction by particles and molecules present in the air. This scattering is essentially caused by Rayleigh scattering on nitrogen and oxygen molecules and Mie scattering on aerosols (dusts, water droplets, etc.). In a LIDAR arrangement, the backscattered light is collected by a telescope, placed close to the laser emitter (see Figure 1). The signal is focused on to a photodetector through a spectral filter, adapted to the laser wavelength. Since a pulsed laser is used, the intensity of the backscattered light can be recorded as a function of time, which provides spatial resolution. The received signal reflects the aerosol and molecular concentrations as a function of distance, similar to an ''optical RADAR''. More precisely, the received number of photons from a distance R at a wavelength λ is given by Equation (1) (assuming that each photon is only scattered once):[34]

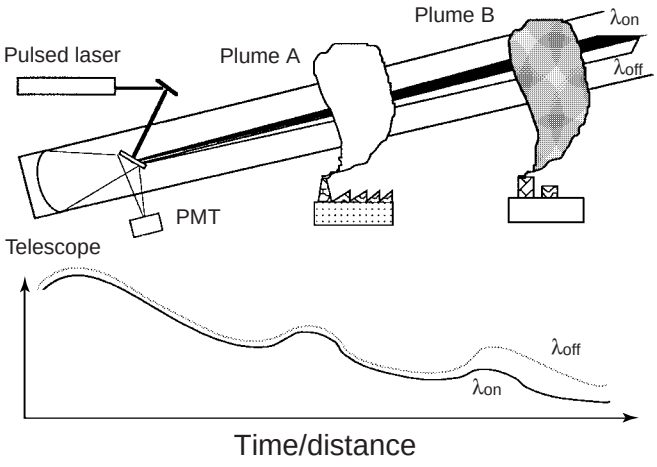

Figure 1 Principles of the LIDAR/DIAL technique.

$$M(R,\lambda) = M_0(\lambda)\,\frac{A_0}{R^2}\,\beta(R,\lambda)\,\Delta R\,\xi(R,\lambda) \times \exp\left[-2\int_0^R \alpha(R',\lambda)\,\mathrm{d}R'\right] \quad (1)$$

where $M_0(\lambda)$ is the number of photons emitted by the laser, A_0 the area of the receiver telescope, ΔR the spatial resolution of the system, essentially limited by the laser pulse duration and the dwell time of the acquisition electronics, $\xi(R,\lambda)$ the detection efficiency, $\beta(R,\lambda)$ the volume backscattering coefficient, and $\alpha(R',\lambda)$ the total atmospheric extinction coefficient.

The detection efficiency $\xi(R,\lambda)$ takes into account every geometrical and optical factor of the receiver arrangement. $\xi(R,\lambda)$ can be separated into two different parameters, one dedicated to the spectral characteristics $\xi(\lambda)$ of the detection channels (filters, monochromator, etc.) and of the sending and receiving optics, and the other, $\xi(R)$, to geometrical properties such as the overlap between the field illuminated by the laser and the telescope field of view. The latter parameter is often used to somewhat compress the large dynamic range of the signal.[35,36]

The signal dependence on the aerosol concentration, which allows their detection, is included in two scattering parameters: the volume backscattering coefficient β and the extinction α. However, the exponential factor $\exp[-2 \int \alpha(R', \lambda)\, dR']$ represents the total atmospheric attenuation of the laser beam, via the Beer–Lambert law. The extinction is caused by two different processes: the Rayleigh–Mie scattering α_{RM} and the specific molecular absorption α_A of the different species present in the atmosphere. The extinction coefficient α_{RM}, like β, is an average value of the size distribution, shape, and composition (refractive index) of the aerosols. In order to solve Equation (1), a standard method is to use a fitted value of the α_{RM}/β ratio.[37,38] This provides backscatter ratios as a function of distance, which *qualitatively* represent the spatial distribution of the atmospheric particle density.

The molecular absorption α_A allows a specific detection of a particular gaseous pollutant, using the DIAL technique. This technique is based on the use of a pair of wavelengths close to each other, with a large absorption coefficient difference (called λ_{on} and λ_{off}, for on-resonance and off-resonance wavelength, respectively).

Let us now assume that such a pair of wavelengths (λ_{on}, λ_{off}), chosen for the detection of a pollutant B, is sent into the atmosphere (Figure 1). It is important that λ_{on} and λ_{off} lie close enough to exhibit the same scattering properties. So, the first plume A in Figure 1 (which does not contain the detected species B) causes an increase of the backscattered signal, because the concentration of aerosols is larger than in the clear atmosphere, but the same amount for both pulses. Conversely, the second plume, which contains a certain quantity of the pollutant, absorbs the backscattered signal at the λ_{on} wavelength much more strongly than at the λ_{off} wavelength. From this difference, using the Beer–Lambert law, we obtain the specific concentration of the considered pollutant as a function of distance. More precisely, if we take the ratio of the two LIDAR returns (each represented by Equation 1), and then the derivative of its logarithm, we obtain Equation (2):

$$N_A(R) = \frac{1}{2[\sigma(\lambda_{on}) - \sigma(\lambda_{off})]} \frac{d}{dR} \ln \frac{M(R, \lambda_{off})}{M(R, \lambda_{on})} \qquad (2)$$

where $N_A(R)$ is the pollutant number density and $\sigma(\lambda)$ the absorption cross-section at the wavelength λ. Note that to obtain Equation (2), the detection efficiency $\xi(R, \lambda)$ must be the same for both laser beams, so that a perfect geometric alignment is necessary.[39] As mentioned before, both wavelengths must be taken close enough so that the backscattering and extinction coefficients are identical, i.e. $\beta(\lambda_{on}) = \beta(\lambda_{off})$ and $\alpha_{RM}(\lambda_{on}) = \alpha_{RM}(\lambda_{off})$.

The pollutant concentration is then measured range resolved. By scanning the measurement direction in azimuth or elevation, two-dimensional (2-D) or 3-D mappings are obtained, like a molecule-specific RADAR.

3 DIFFERENTIAL ABSORPTION LIGHT DETECTION AND RANGING SYSTEMS

In the 1980s, ultraviolet differential absorption light detection and ranging (UVDIAL) systems were mainly based on dye lasers, pumped by frequency doubled or tripled Nd : Yag or excimer lasers. The dye laser output was then frequency doubled to reach the 225–400 nm region,[63] where major pollutants (nitrogen oxides, sulfur dioxide, ozone, etc.) exhibit strong absorption features. The complexity and user unfriendliness of these laser systems definitely prevented the widespread of LIDAR systems at this time.

The advent of high-energy, tunable, all-solid-state lasers in the early 1990s completely changed this situation. DIAL systems were finally considered as reliable ''black boxes'', which could be widely used for air pollution monitoring applications. This also induced industrial developments of DIAL systems. The first ''all-solid-state'' UVDIAL system commercially available, based on a flashlamp-pumped Ti : sapphire laser, was for example presented in 1993.[41]

Since then, Ti : sapphire-based UVDIAL systems have been very successful. A typical example of such an instrument, manufactured by Elight Laser Systems, is presented below.

The laser is a flashlamp-pumped Ti : sapphire laser in an oscillator/amplifier configuration. A patented oscillator,[42] based on a dual prismatic dispersion arrangement, provides both reference and probe wavelengths (λ_{on} and λ_{off}), either simultaneously or alternately. The switching is performed by a rotating-face parallel prism allowing collinearity between both laser beams of better than 100 µrad. The use of a *single laser* drastically reduces maintenance.

Mode quality and frequency stability are assured by a graded reflectivity (super-Gaussian, 50% maximum) output coupler. With a Pockels cell as Q-switch, this laser system delivers 250 mJ in 20 ns pulses with a bandwidth of 0.5 cm^{-1} over the whole tuning range (720–900 nm).

The high peak power allows efficient nonlinear optical frequency conversion [second harmonic generation (SHG) ~20%, third harmonic generation (THG) ~7% of the fundamental] to reach the ultraviolet (UV) absorption bands of the detected pollutants. The laser beam is expanded 10× to meet eye safety requirements (according to VDI/DIN 0837).

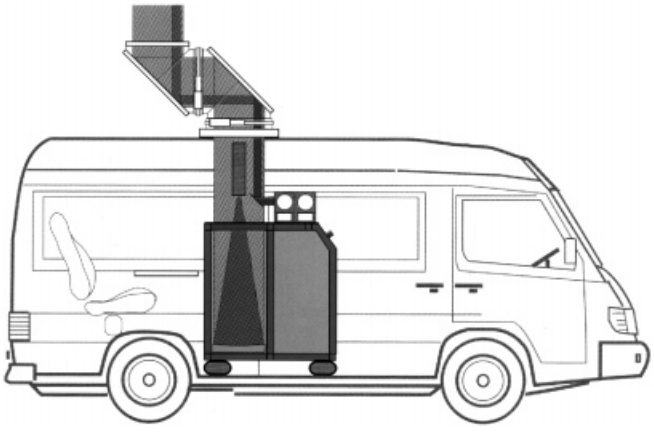

Figure 2 ''All-solid-state'' mobile LIDAR/DIAL system.

Note that for the eye safety reasons mentioned above, some weaker bands of pollutants have to be chosen instead of stronger ones, which lie above 400 nm and lower than 1400 nm. The laser beam is sent into the atmosphere by a dual scanning periscope in composite material which also directs the backscattered light back to the 400 mm/f3 parabolic mirror (Figure 2). This telescope focuses the light on to an iris to limit the field of view to about 0.5 mrad. The signal is detected by a photomultiplier, through a motorized filter wheel, in order to select automatically the proper spectral filter. The signal is digitized by a 12 bit, 40 MHz system, allowing a spatial resolution of 3.75 m in this case (for some emission measurements faster analog-to-digital converters are used). The data are handled by a PC microcomputer which controls every parameter: wavelength setting and calibration, receiver parameters [photomultiplier tube (PMT) high voltage, gain, spectral filter used], and the periscope steering. The software is Windows-based, and includes a sophisticated inversion algorithm optimising the signal-to-noise ratio (S/N), and modules for automatic operation, on-line evaluation of the data, and high-level graphic representation.

The described instrument is moreover of very compact design (110 × 100 × 120 cm), which allows its installation in a small van (e.g. a Mercedes MB 100). It is able to perform 3-D mappings of NO, NO_2, SO_2, O_3, benzene, and toluene at high sensitivity (a few parts per billion) and over large distances (several kilometers). Typical specifications are listed in Table 1.

An additional extension of the accessible wavelength range, in particular in view of hydrocarbon detection, is made by an OPO in a ''no-tracking'' configuration. The decisive advantage of pumping with a tunable laser such as Ti : sapphire is that no geometrical alignment of the OPO is necessary.[23,43] The device can even be monolithic with mirror coatings directly applied on the nonlinear crystal facets. With such a device, infrared (IR) radiation from 1 to 4 μm is generated with an efficiency of up to 10% just by tuning the pump wavelength from 730 to 860 nm.[43]

Recently, a new aspect was brought out by Douard et al.[44] concerning multi-DIALs. The idea is to send a broad-band laser (like a free-running OPO) in the atmosphere and to analyze spectrally the backscattered light by an optical multichannel analyzer (OMA) or a Fourier-transform infrared (FTIR) interferometer. This technique was demonstrated for the first time in 1995.[44] The decisive advantage of the method is the capability of detecting several species by DIAL at the same time. A very attractive application is thus the detection of volatile organic compounds (VOCs) in the mid-IR region (3.3–3.5 μm).

It should be noted at this point that taking into account the different sources of noise and error, the sensitivity is actually a value in parts per million times meters, because of the Beer–Lambert law: for a given S/N, the larger the range of integration, the lower the possible detected concentration. For example, if the detection limit is 1 ppm m, the lowest detectable level is 1 ppm for a range resolution ΔR of 1 m (emission conditions), but 5 ppb for a range resolution of 200 m (immission conditions).

The maximum distance of measurement is typically 3 km for standard meteorological conditions. The sensitivity and the absorption path can be increased in a large extent by using topographic targets down to about 500 ppt for a 10 km pathlength.[45] Nevertheless, the spatial resolution is then lost in one direction, and only average 2-D mappings are obtained.

Table 1 Specifications of a Ti : sapphire-based UVDIAL system[41]

Pollutant	λ_{on} (nm)	λ_{off} (nm)	K (cm^{-1} atm^{-1})	Energy (mJ)	Sensitivity (ppb)	Range (km)
SO_2	286.55	285.7	10.2	5	8	3
O_3	283.0	287.0	27.4	5	4	3
NO	226.80	226.83	105	1	2	1
NO_2	398.29	397.5	4.5	25	15	4
Toluene	266.9	266.1	29.5	7	7	3
Benzene	258.9	257.9	38.7	7	5	3

4 EXAMPLES OF THREE-DIMENSIONAL MAPPING OF GASEOUS POLLUTANTS

We present in this section several examples of DIAL measurements of gaseous pollutants. The first part is dedicated to the study of physical atmospheric dynamics, and in particular to transport processes and spread of primary pollutants in different topographic and meteorological conditions.

The first application consists in emission measurements in industrial environments. Figure 3 shows a typical example of a DIAL concentration profile. The detected pollutant is SO_2. The measurement was performed directly through the chimney plume of a power plant. This example shows the ultimate spatial resolution of the system,[43] of about 2 m, which allows the observation of the structure of the plume. The two concentration maxima are, indeed, due to the use of two separate burners in the combustion chamber. The measured SO_2 values are, furthermore, in good agreement (within $\pm 10\%$) with the standard spot measurements, performed by the company inside the chimney.[46] Moreover, remote sensing allows the plume concentration to be followed as a function of distance, and thus to characterize the dilution and its expected impact (see section 6).

Scanning the measurement field in azimuth or elevation yield the unique access to tridimensional maps of concentration. Figure 4 represents such a map of the NO_2 concentration emitted by a chemical factory. The high spatial resolution allows here to observe long-living macroscale turbulence inside the plume and gives a real picture of the plume dispersion and transport. A simultaneous air speed measurement leads to the emitted mass concentration flux.[26,47] This speed is often determined by correlation of video pictures of the aerosols

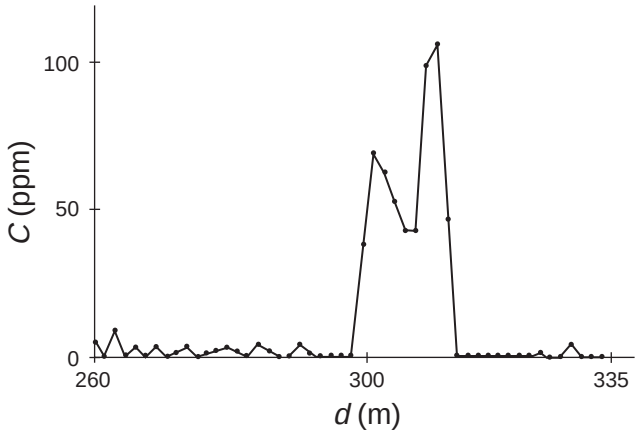

Figure 3 Remote SO_2 concentration measurement by LIDAR at the top of a chimney of a power plant.[46]

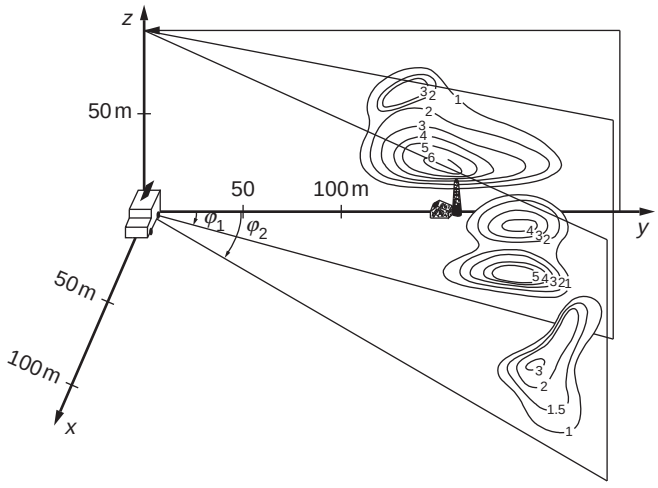

Figure 4 3-D mapping of NO_2 emitted by a chemical factory.

contained in the plume or by correlation of the LIDAR returns themselves.[6,48]

Note, however, that such 3-D mappings require relatively stable weather conditions, since the total measurement time in this case is 15 min. The atmospheric fluctuations in plume measurements by DIAL must therefore be carefully considered, so that the measurement time always fits the timescale of the monitored process.[11,49] Microscale, fast fluctuations may also induce errors in DIAL measurements of emitters, since the λ_{on} and λ_{off} wavelengths are usually sent alternately since they have to remain close to each other and thus cannot be spectrally separated in the detection chain. The switching time between the two wavelengths is critical because microscale fluctuations of concentration may be fast enough to induce errors.[50,51] In the UV and for DIAL measurements other than plume tracking, it has been demonstrated that atmospheric fluctuations do not induce significant errors in the results[49] and that shot noise is the major limitation.

If the characterization of identified emitters such as plumes is highly relevant in terms of impact on the environment, the use of DIAL systems in industrial environments is even more interesting for diffuse or unclassified emitters. An example of a toluene profile obtained over a petrochemical plant in Cubatao (Brazil) is presented in Figure 5. It clearly shows the diffuse emission of a storage tank of gasoline: the floating roof periodically exhausts some toluene bursts, which then move to higher altitudes. These emissions are usually very difficult to quantify or even to detect with standard techniques, although they may amount to as much as the classified emitters in some cases.[47]

The most important source of diffuse and ''moving'' emission of atmospheric pollution is obviously road

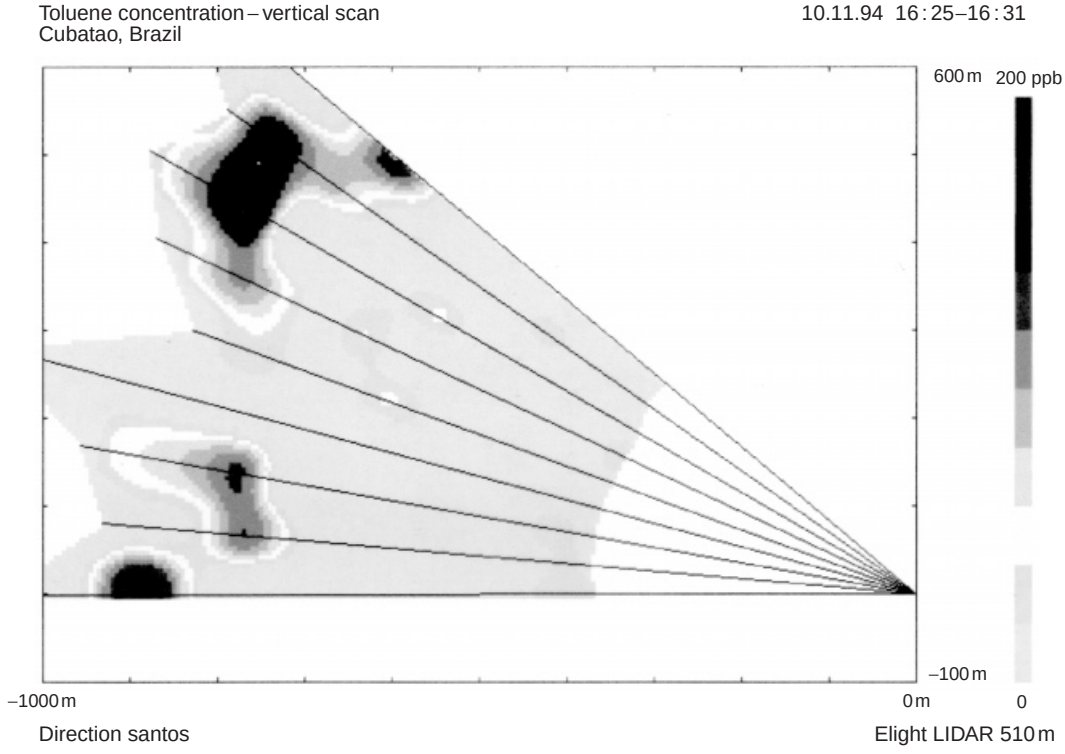

Figure 5 Diffuse toluene emission from a storage tank.[41]

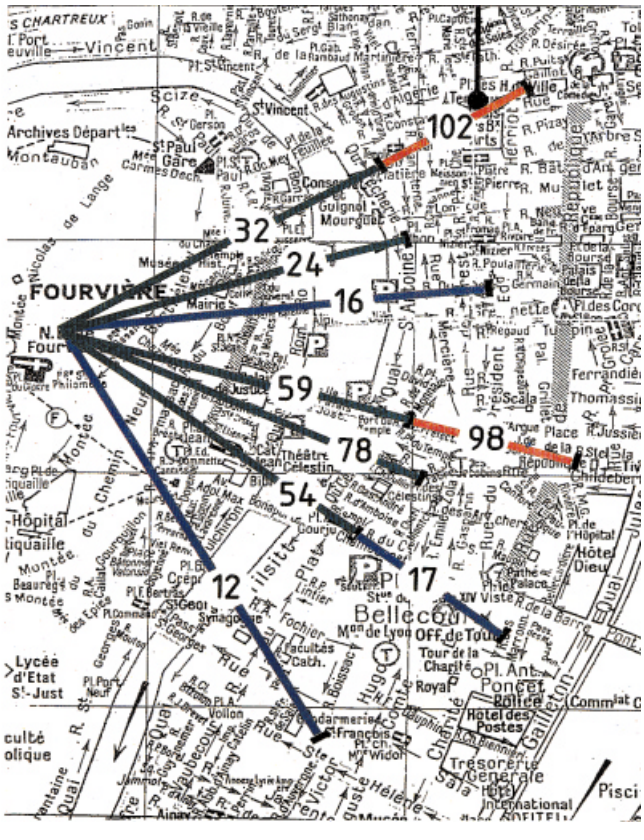

Figure 6 NO mapping over Lyons.[28]

traffic, so extensive studies have been performed in urban conditions. In particular, DIAL measurement campaigns in large European cities have provided very important information about the influence of urban topography.[28] Figure 6 presents a map of the NO concentration over the city of Lyons. Since Lyons is a relatively flat, well-ventilated city, the traffic-related pollution is efficiently spread, and the highest values (102 ppb) are recorded for narrow streets in the old part of the city, where a few vehicles are moving slowly, and not over the main avenues. Vertical gradients have been determined in these narrow street canyons: a dilution factor of 5–6 was usually recorded between the ground level and the canopy, attesting to the poor ventilation.

The topography of Stuttgart is opposite to that of Lyon, with a shape close to a ''Mexican'' hat, where the vertical air mass exchanges are strongly hindered. As a result, inversion layers, where air pollution is trapped, appear frequently, as shown in Figure 7. NO concentrations as high as 150 ppb have been recorded by DIAL inside this layer. Owing to the valley shape of the region, this layer reaches the outer residential quarters of Stuttgart more dramatically than the city center, although the primary NO pollutant is mainly emitted in the city center.

In addition to emission intensity and topography, meteorology also plays a crucial role in the distribution of air pollutants and in the formation of smog situations.

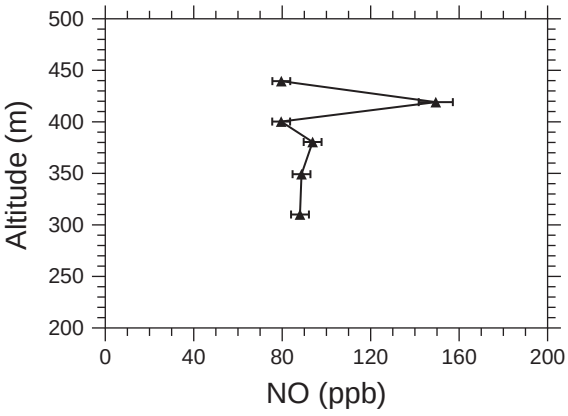

Figure 7 Inversion layer over Stuttgart.[28]

This is demonstrated in the case of Geneva in winter 1989. Figure 8(a) exhibits a typical smog situation, where high NO concentrations are recorded everywhere, without steep gradients, although NO is a primary, traffic-related pollutant. The concentrations increase almost without limit because of the lack of air motion. Two days later, a north wind destroyed this local equilibrium and blew the air masses out from the Geneva valley. This led to a return to a normal situation (Figure 8b), with steep gradients and lower concentrations. The Mont Blanc bridge, usually strongly subject to traffic jams, does not suffer too much, again because of good ventilation from Lake Geneva and the River Rhone. However it should be noted that the high value of 92 ppb remained over the English Garden (''Jardin Anglais''). It was found that this local maximum was due to the air conditioning system exhaust of the large underground car parking located under the Garden.

This example emphasizes the difficulty of obtaining a realistic picture of the air quality in a city with standard networks, and especially the difficulty of finding representative locations for spot analyzers. DIAL measurements provide very useful information in this respect.

An especially attractive application of DIAL is in tracing pollution in trans-border importation/exportation situations. An example of a study of this kind was performed during an intensive LIDAR campaign in Berlin in January 1989, in order to quantify the transport processes between East and West Berlin. At this time, the Berlin Wall still existed, and the West German pollution regulations were applied to West Berlin, although no sanitation actions could be taken in East Berlin. Even the concentration values of pollutants at ground level in East Berlin were not available. A typical result of this campaign is shown in Figure 9.

The directions in which the vertical profiles were recorded were chosen in the south-east (SE) wind direction and at 90° from it. In the SE direction (East

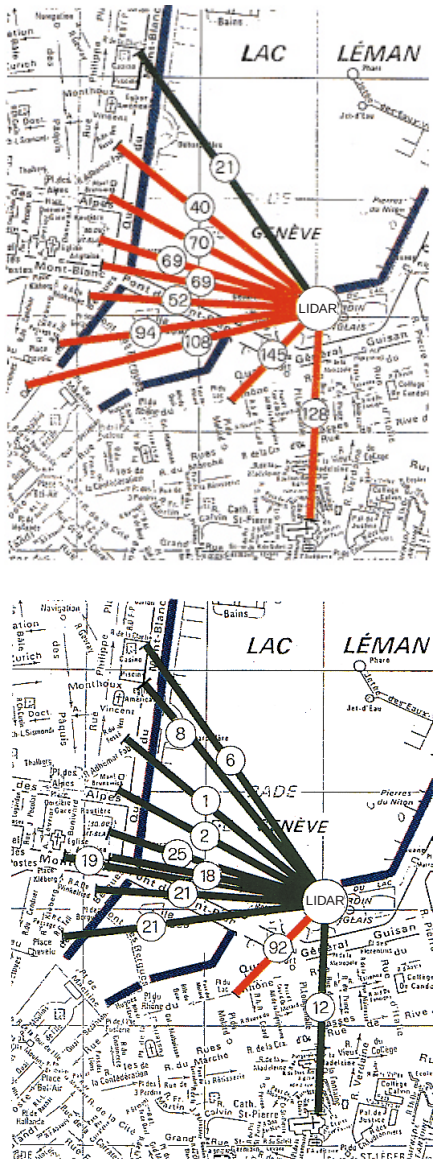

Figure 8 NO mappings over Geneva (a) during and (b) after a smog situation.[28]

Berlin) the SO_2 values reach 140 ppb at ground level and then smoothly decrease with altitude until a set of inversion layers between 400 and 600 m. The emitted concentrations at ground level are mainly due to coal burners. In the north-west (NW) direction (West Berlin), ground level values are much lower, while at 600 m a maximum is found, demonstrating pollution transport processes. At 90° south-west (SW), the maximum is located at lower altitude, confirming the interpretation of importation from SE to NW.

A more direct importation process has been recorded, related to the power plant ''Mitte'' located in the center of East Berlin. For this, the DIAL system was placed

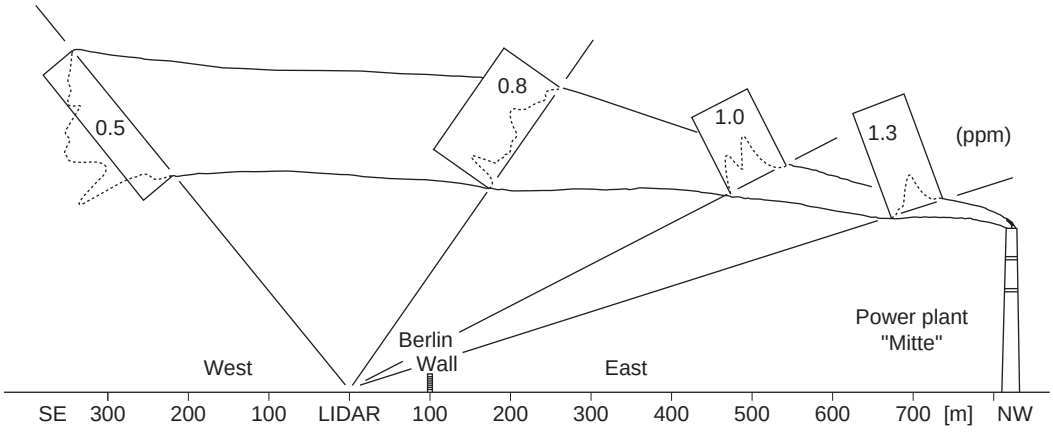

Figure 10 Importation in former West Berlin from a power plant located in former East Berlin.

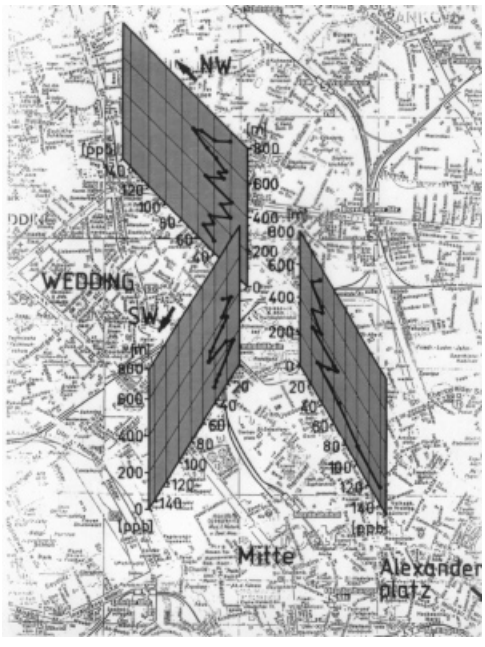

Figure 9 Vertical distributions of SO_2 over Berlin in January 1989.

close to the former Berlin Wall, about 1 km from the emitter (Figure 10). The SO_2 plume was followed for over 2 km, delivering very precise information. Besides the mass concentration fluxes determined from DIAL measurements just at the chimney output, it could be clearly observed that, again, two burners were used, and that their related plumes did not mix with the distance. It was also shown that more than 500 ppb of SO_2 were ''imported'' over West Berlin from this emitter at that time.

Routine monitoring is also a standard application of LIDARs. A stationary DIAL system has been implemented in the center of Leipzig since 1992. Figure 11(a) shows a typical map of SO_2 over the city. Such maps can be recorded hourly and give average information about air quality. However, if a higher concentration is observed in some direction, a ''zoom'' can be performed at higher spatial resolution (a few meters) and in three dimensions. The example of a power plant located 1.5 km from the station is shown in Figure 11(b). This allows us to determine the plume dimensions and dilution over more than 1 km, and then obtain a more accurate estimate of the effects of this particular emitter in the former low-resolution map, i.e. a direct correlation between emission and immission.

After these examples of accessing physical dynamics with DIAL, let us now focus on chemical and photochemical processes. As an example, we present here the results of a measurement campaign on ozone in Athens in summer 1994. Ozone is one of the most important pollutants responsible for the summer ''smog'' recorded in large urban areas. Its formation is particularly due to the emission of road traffic (nitrogen oxides, VOCs) combined with sunny periods. Photolysis is the dominant loss process for NO_2 in the troposphere and the major source of tropospheric O_3 (Equations 3–5):

$$NO_2 + h\nu \longrightarrow NO + O(^3P) \quad (3)$$

$$O(^3P) + O_2 + M \longrightarrow O_3 + M \quad (M = \text{air}) \quad (4)$$

$$O_3 + NO \longrightarrow O_2 + NO_2 \quad (5)$$

In the absence of other processes that convert NO to NO_2 and assuming steady-state conditions, one obtains the well-known Leighton relationship [Equation 6]:[53]

$$[O_3] = \frac{j_3[NO_2]}{k_5[NO]} \quad (6)$$

The ozone production depends in this very simple case on the solar irradiation-dependent rate j_3 and the rate constant of Equation (5), k_5. However, it has been shown

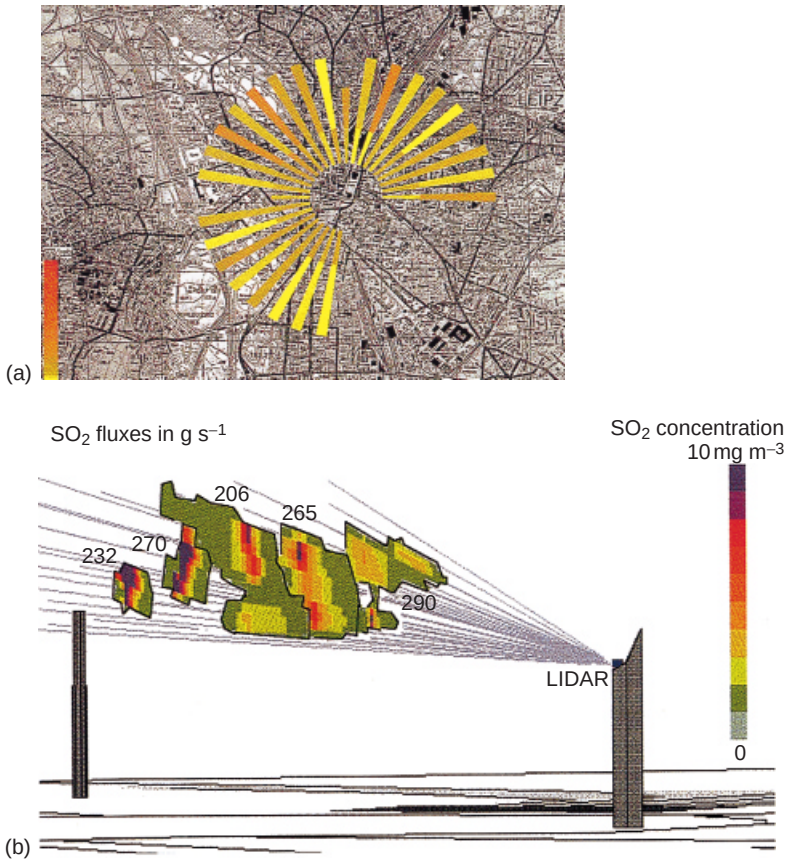

Figure 11 Routine measurements over Leipzig and 3-D "zooming" on to an emitter.[38]

that VOCs also play a key role in ozone formation, and in particular enhance the NO to NO_2 conversion and thus favor ozone production. A comprehensive description of the many reactions involving NO_x and VOCs can be found in a reference book published by the National Research Council.[54]

Ozone production is usually plotted in the form of isopleths as a function of the VOC and NO_x concentrations. The isopleths have roughly the shape of a hyperbola, and the diagram characteristically exhibits a diagonal ridge from lower to higher concentrations. This ridge is more or less a straight line of a given VOC/NO_x ratio. Depending on the value of the VOC/NO_x ratio encountered in the atmosphere, the ozone production can thus be either NO_x-limited (high VOC/NO_x ratios) or VOC-limited (low VOC/NO_x ratios). The location of the ozone maxima will thus be determined by the VOC/NO_x ratio and not by the isolated NO_x or VOC concentrations.

Athens is a very interesting example of ozone smog formation. Meteorological and photochemical dynamics are combined to form ozone in high concentrations. Athens is located in a basin, and thus subject to a back and forth motion of air masses from the sea to the ground. Hence cold wind flows from ground to sea in the night whereas a sea breeze moves the air masses back to the land inside during the day. The wind direction flips very fast from north to south at about 10 a.m. in summer, as shown by Grying.[55] Simultaneously, the dense traffic emits high NO_x concentrations which, together with the strong solar radiation and high temperature, favor ozone production.

DIAL measurements[26,56] were performed in summer 1994 within the large-frame MEDCAPHOT campaign, the LIDAR system being located in the city center (Pnyx Hill). In order to access the ozone formation dynamics, both NO_2 and O_3 mappings were recorded. Figure 12(b) shows the 2-D NO_2 vertical map over Athens at about 10 a.m. Since solar radiation had just started to increase, the NO emitted by the road traffic was oxidized during the early morning (Equations 5 and 6), yielding a thick NO_2 layer near ground level, with concentrations exceeding 100 ppb. Half an hour later, the effect of wind flipping is observed: the sea breeze induces vertical advection of the NO_2 concentration (Figure 12a). The increasing ground temperature also contributes to this rising motion of the air masses.

In Figure 12(a), a secondary local maximum of NO_2 at about 1000 m altitude is clearly observable. This NO_2 is

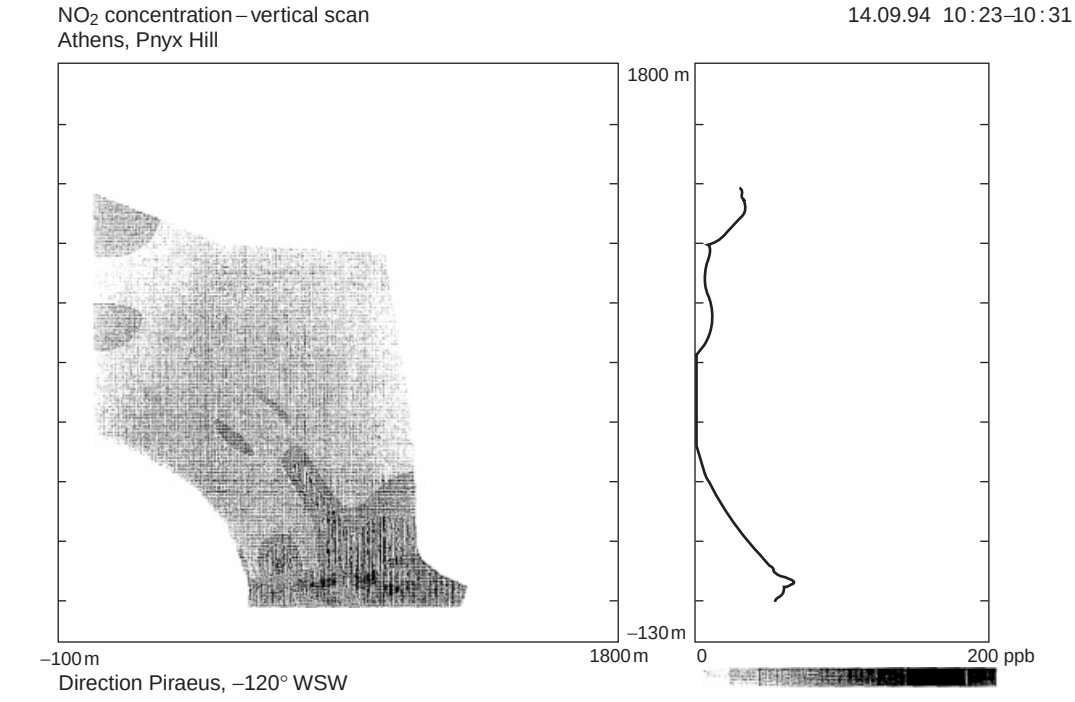

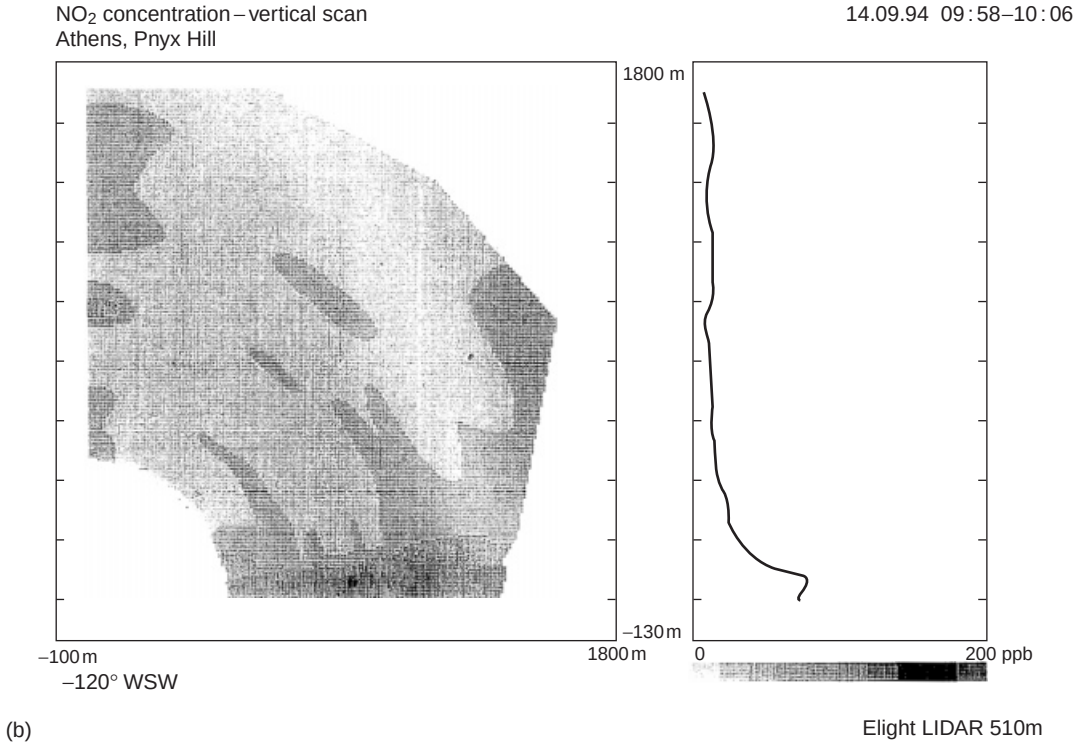

Figure 12 NO_2 mapping over Athens during a summer smog episode.[26,31]

then photodissociated by the solar radiation, generating ozone (Equations 3 and 4). This leads to the formation of an ozone storage layer, as presented in Figure 13. Ozone concentrations as high as 180 ppb are observed at around 1000 m altitude. Note that the concentration at ground level at this moment is only 50 ppb, owing to the local high NO concentration emitted by the dense traffic.

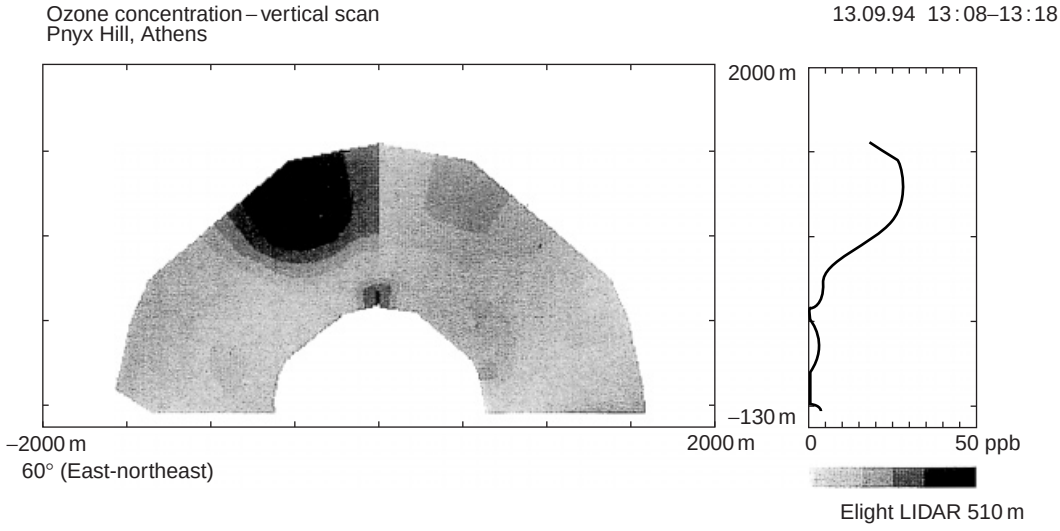

Figure 13 Ozone storage layer over Athens.[26,31]

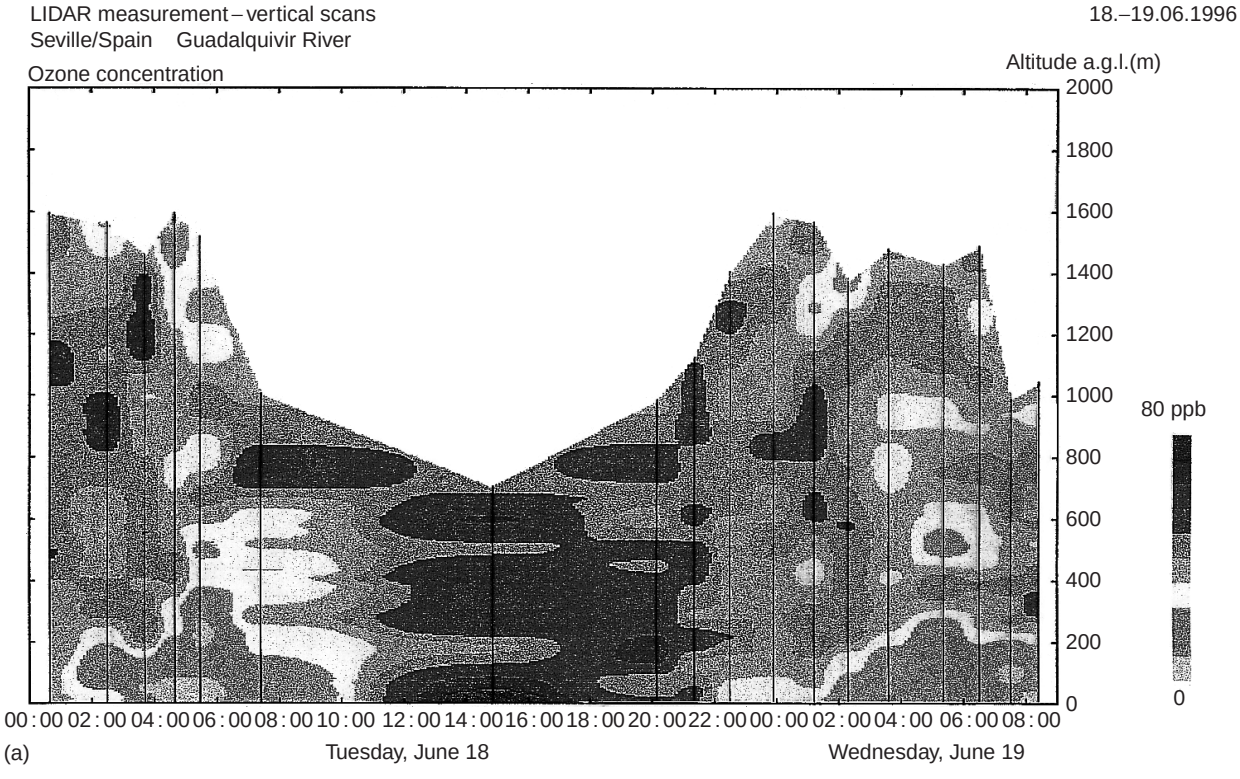

Figure 14 Ozone profiles obtained (a) at Seville center, (b) upwind from Seville and (c) 22 km downwind from Seville.[32,33]

The meteorological dynamics take over again at the end of the beginning of the night, when the air masses fall again and transport ozone back to the ground. A detailed study of these dynamics and a comparison with numerical models was described by Weidauer.[56]

The effects of vertical transport of the NO_2-producing ozone observed here can obviously also occur horizontally, in the surroundings of urban areas. The measurements in Seville within the frame of a campaign organised by the Joint Research Center Ispra[33] have been mainly performed at three particularly representative sites. The first series of mappings was recorded in the town center. Profiles were measured along a road with major traffic, interrupted with many traffic lights. Results are presented in Figure 14(a). They show a moderate average ozone concentration of about 50 ppb with peaks at 80 ppb

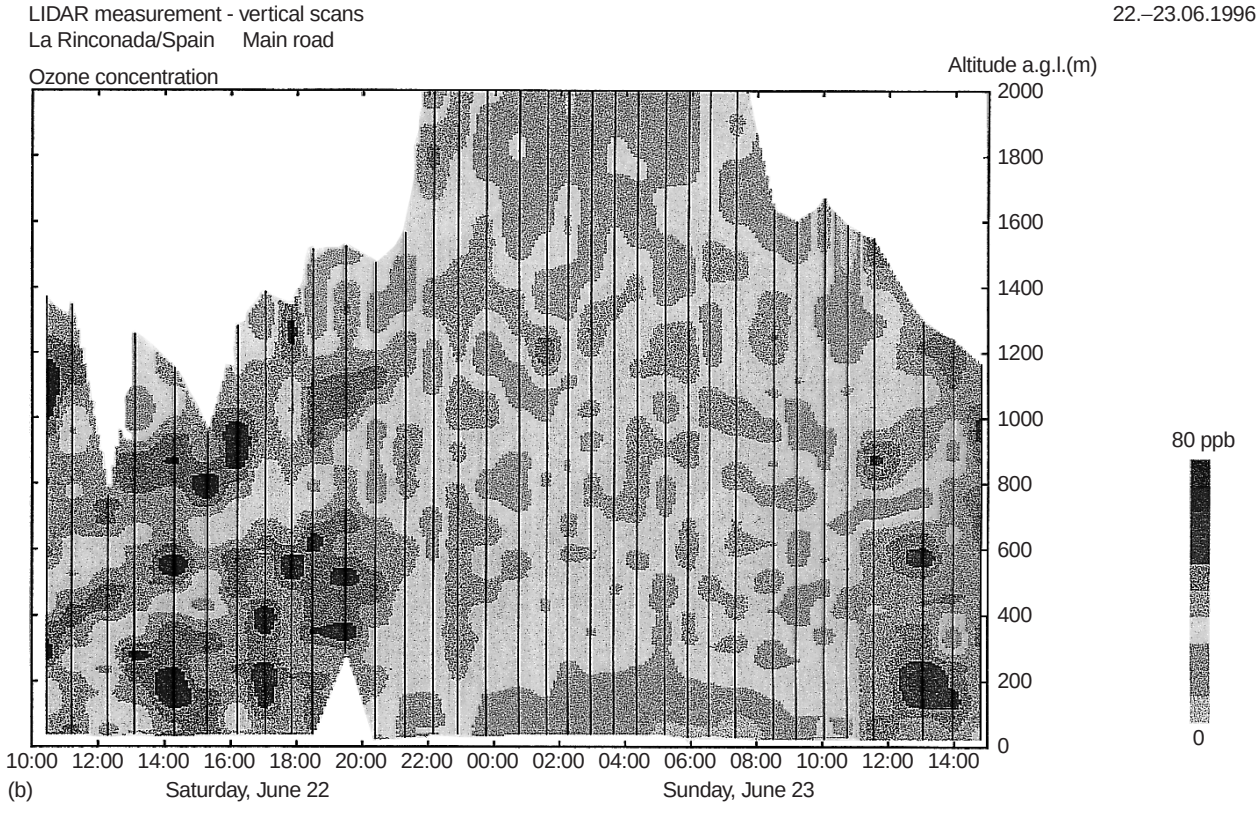

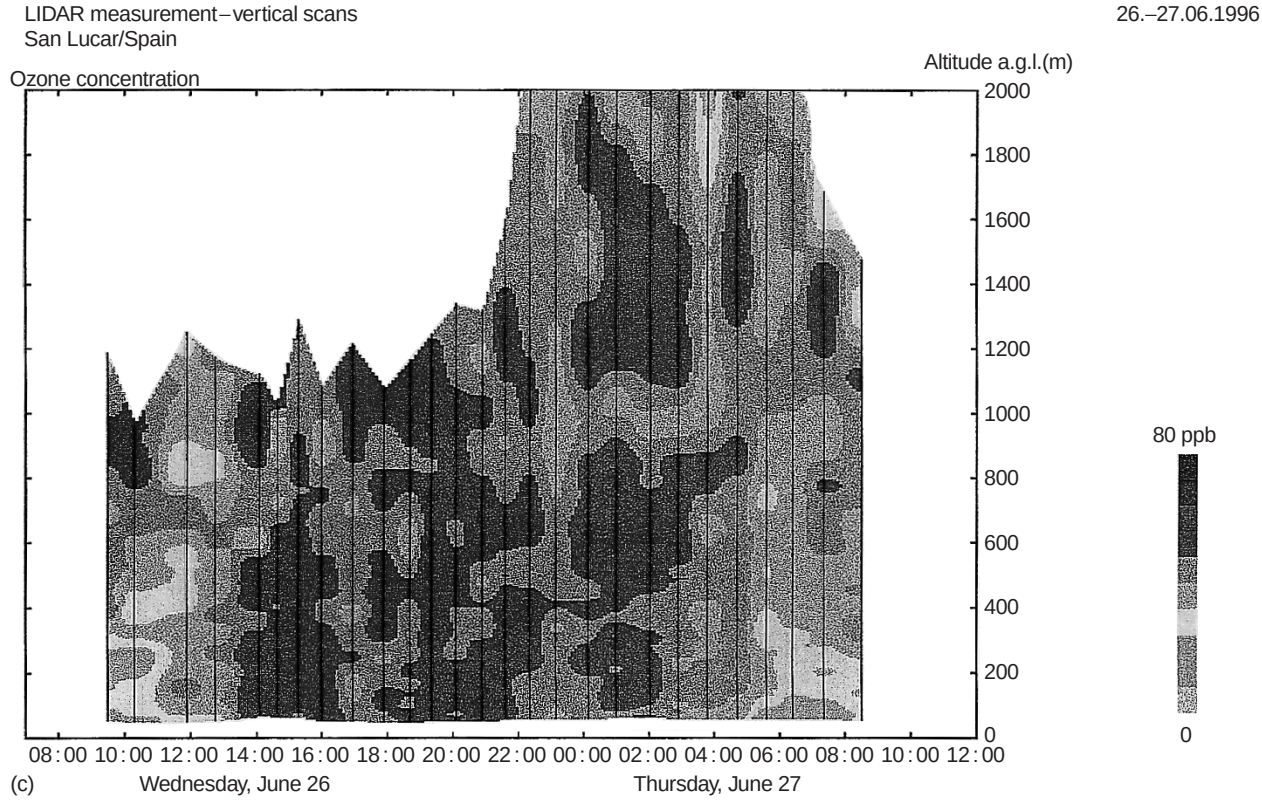

Figure 14 (*Continued*)

at the beginning of the afternoon when solar radiation is maximum. During the night, the concentration stabilizes around 35 ppb, although it remains important at high altitudes. Again, without solar photodissociation of NO_2, oxidation of nitrogen monoxide takes place, and yields a lower ozone concentration. In other words, as is often observed,[54] the VOC/NO_x in the center of large cities is too low to obtain maximum ozone production.

The second series of measurements was performed in a rural area, La Riconada, upwind from Seville during the measurements. Results are presented in Figure 14(b). Concentrations were low, from 30 ppb during the night to 50 ppb in the day. This can be easily interpreted, since the Seville city plume is propagating in the opposite direction and thus almost no nitrogen oxides (and also O_3) is imported. The population is weakly exposed compared with the population in the center of Seville.

The third site chosen for the LIDAR measurements was downwind from Seville at the measurement time, in a totally rural zone, 22 km from the city. This region is well known for its agricultural production. Long-distance transport of the Seville nitrogen oxides then encounter local biogenic production of VOCs (isoprene, terpene), which yields very favourable VOC/NO_x ratios for ozone production. Moreover, part of the ozone coming from Seville is also transported. This results in a high ozone concentration (Figure 14c) for such a rural site, relatively homogeneous in space and time: 50 ppb on average with peaks at 70 ppb. These concentrations remain high during the night owing to the fact that the local road traffic (and therefore NO) is low.

The results presented here show precisely on the one hand the influence of the meteorological conditions on ozone formation, and on the other hand the importance of transport processes of urban pollution. They also show that some legal reduction actions on nitrogen oxide or VOC emissions may lead to a geographical shift of the ozone maxima and not to an overall reduction.

In the whole discussion above, the effects of aerosols on the ozone production could not be taken into account, since their action is not fully understood. Quantitative measurements of aerosols by LIDAR are thus an important challenge, as described below.

5 DETECTION OF AEROSOLS AND DUST PARTICLES

The impact of atmospheric aerosols on public health and in heterogeneous chemistry is well recognized, although few conclusive studies have been reported so far. This is especially true for urban aerosols, the variety of which in size and composition makes complete characterization difficult. We present here a combined technique, using LIDAR measurements, scanning electron microscopy (SEM), and X-ray microanalysis.[57] Urban aerosols are first sampled using standard techniques on different filters. Their size is determined by SEM and counting and their composition by X-ray microanalysis. The size distribution exhibits mainly two modes: one around 0.1 μm constituted of soot and sulfates and the other around 1 μm composed half of larger soot aggregates and half of silica particles.

These data are then used to compute optical backscattering and extinction coefficients, leaving as the only unknown parameter the total number density of particles. The basis of the determination of the optical scattering coefficients α and β is a newly developed model, based on a fractal theory for small soot particles (from 50 nm to 0.5 μm) and on a modified Lorentz–Mie theory for larger particles such as silica.[58] The larger soot aggregates are considered as spheres made of lacunary media, with 50% carbon and 50% void.

Number-density profiles over Lyon were obtained by simultaneous LIDAR measurements using an eye-safe LIDAR system (frequency doubled Ti : sapphire laser-based system at 390 nm). The LIDAR data were inverted with the backscattering and extinction coefficients computed from the actual size distribution. These measurements were performed in the frame of an intensive campaign in summer 1996 in Lyon, in collaboration with the Departement d'Ecologie Urbaine of the City of Lyon.

LIDAR data were inverted using a Klett algorithm[37,38] with two reference points, one located close to the ground with the α and β values determined above, and the other at high altitude where a clear atmosphere is assumed. In the conditions encountered here, the assumption of an average shape of the size distribution is acceptable (warm, stable afternoon). This could be confirmed by a parallel epidemiological study, involving 30 volunteers with transportable samplers.

Quantitative 3-D distributions of aerosols over the city have been obtained with this method for the first time. The results show a fairly homogeneous mixing layer up to 1300 m. An interesting measurement is also the temporal dependence of the concentrations of these urban aerosols. Figure 15 shows the evolution of the vertical profile of aerosol concentration in Lyon over 24 h during a high-pollution episode, on June 27–28, 1996. It shows very clearly the motion of the mixing layer height. Owing to the temperature changes, the mixing layer top drops to 200 m during the night and rises to 1300 m in the early afternoon. Traffic jams also induce

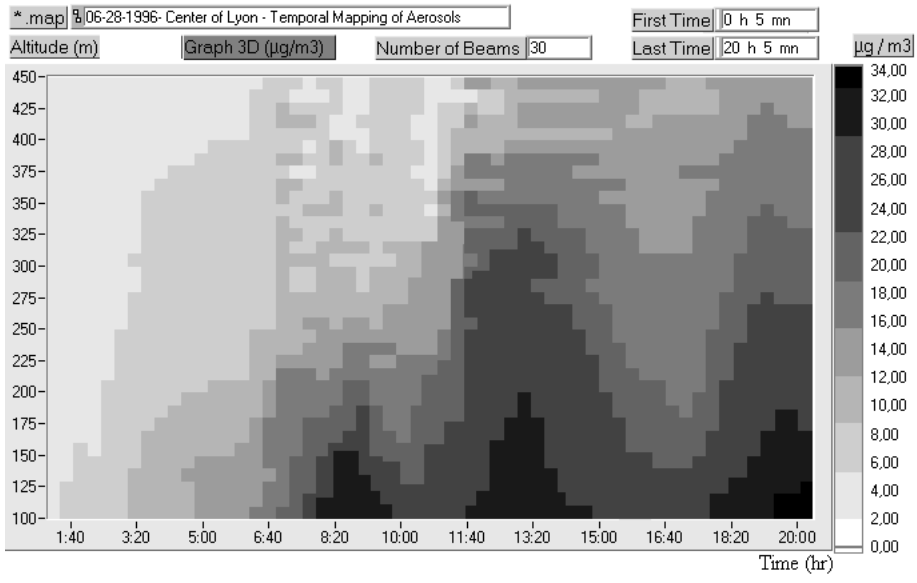

Figure 15 Temporal evolution of the aerosol mass concentration measured by LIDAR.

an increase in the aerosol density during rush hours, in the morning around 8 a.m., at noon, and in the evening around 6 p.m. Even 80 m above ground, concentrations as high as 40 $\mu g\,m^{-3}$ are observed during these rush hours.

6 VALIDATION OF NUMERICAL MODELS

A task of increasing importance for LIDAR systems is the validation of numerical models. This applies both to microscale dispersion models and to regional-scale codes, which provide output data averaged over grid cells several kilometers long.[54]

Verification of numerical model predictions has been performed both for microscale plume dispersion codes and for regional-scale Lagrangian/Eulerian models during a large frame campaign in the Swiss Rhone Valley.[46] Since both models tested did not include a chemical module, the pollutant chosen for the comparison was a relatively non-reactive species, SO_2.

The verification of the plume spread model was performed on the emission of a power plant, where 3-D mappings of the plume were recorded at high resolution and the concentration was averaged over sections perpendicular to the plume direction. This gave a good estimate of the plume dilution with distance (Figure 16). Owing to the windy meteorological conditions, the emitted SO_2 concentration is rapidly spread and transported in a rather flat plume of about 120 m width for only 30 m height. A typical dilution value is 30 at a distance of 100 m.

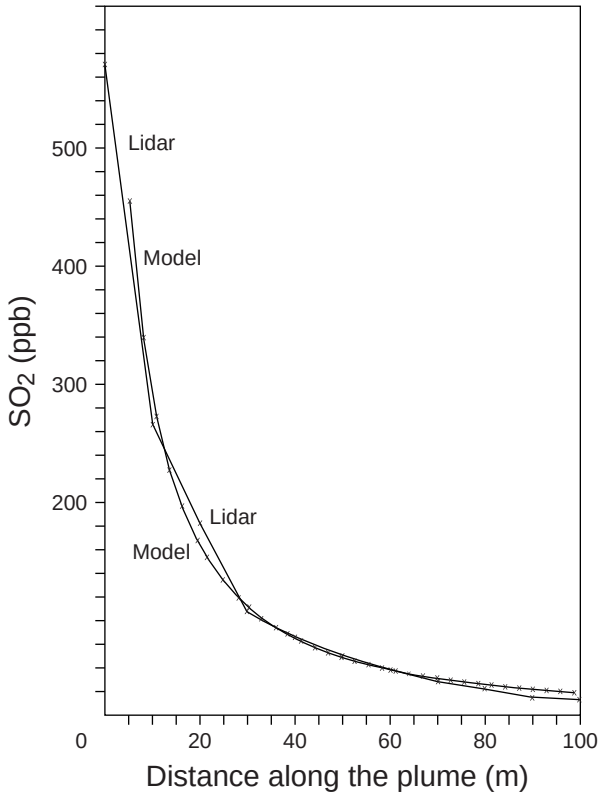

Figure 16 Verification of a plume spread model by DIAL measurements.[46]

The model (Micro-EL) is based on the dispersion of a given number of virtual particles and subsequent counting in the grid cells as a function of time. Together with the

meteorological conditions and the emitter characteristics (plume exhaust velocity, temperature, initial SO_2 concentration), one then retrieves the concentration in each grid cell. The agreement between DIAL data and numerical simulations is excellent. This is made possible only through the careful analysis of a very large number of Lagrangean particles (31^3) and the high density of the receptor grid ($1\,m^3$) located along the plume axis in order to estimate dilution. It can be seen, however, that the initial concentrations are overestimated by the numerical model in the vicinity of the chimney exhaust. This difference may be produced by the presence of numerous aerosols (especially water droplets), on which some SO_2 molecules are adsorbed. The SO_2 concentration, measured by DIAL, is then lowered at the benefit of other compounds, which are not measured.

Note also that if the model is verified for some section of the plume, it can be used in turn for the dispersion at longer distances, where DIAL data are no longer available. The LIDAR–model combination can thus be very useful and complementary for long-distance transport studies.

The verification of the large-scale model (MESOCONV[46]) was carried out in the Chablais Valley, a fairly wide channel, completed by several smaller and colder perpendicular valleys, which often leads to a multilayered stable situation. The emitted pollution can therefore be trapped in one of these sublayers without any mixing with the others. The pollution concentration is then highly dependent on altitude, and it is of great importance for vegetation damage to monitor whether the maxima always occur at the same levels above ground. Several SO_2 emitters are present at ground level, and also a highway which follows one of the edges of the valley.

On the day of the experiment, winds were light ($3–5\,m\,s^{-1}$) and blowing down-valley towards Lake Geneva. A very strong temperature inversion was located at about 450 m above the valley floor, at 850 m above sea level. This inversion plays a fundamental role, of course, in the distribution of pollutants within the valley. A comparison between DIAL data and results of the model is shown in Figure 17. For this, vertical profiles, averaged over the horizontal paths across the valley, were both calculated and measured. An excellent correlation was again found at the mesoscale level, under immission conditions. It can be seen that, even for a receptor located some 12 km from the southernmost sources, the value of the peak concentration and its height above the ground correspond very well with the observations. Above the peak value, which is located at the inversion level owing to trapping by thermal stability, as was explained previously, the DIAL system

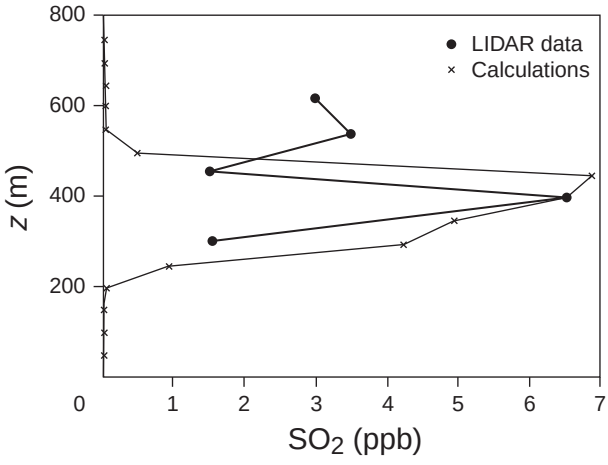

Figure 17 Comparison between LIDAR vertical profiles of SO_2 measured in the Swiss Rhone Valley and calculated data from a mesoscale model.[46]

detects nonzero pollution which is not computed by the numerical model. This upper-level pollution is in effect a feature of pollution from the previous day which was diffused at height but which remained at upper levels owing to alternating up- and down-valley flow which recirculates the pollutants. The numerical model computed only the contribution to pollution from emissions on the day chosen for this case study, and consequently contains no background pollution, which is a measure of the recent ''history'' of air quality in the region studied. This can be of advantage when attempting to identify the contribution of a particular source to global pollution of multiple origin. Conversely, it may be criticized that the absence of background pollution in the model leads to results which do not correspond to reality.

7 COMPARISON WITH OTHER TECHNIQUES AND CALIBRATION PROCEDURES

Calibration of LIDAR systems is an important but difficult issue, since path-averaged data (within the used spatial resolution) have to be compared with spot reference measurements. Note that the same problem applies to other optical techniques such as open-path FTIR spectroscopy and differential optical absorption spectroscopy (DOAS).

The success of comparative measurements between DIAL and spot reference data depends directly on the concentration gradients present in the atmosphere.[59] In homogeneous conditions, i.e. sites far from emitters, comparative studies can be performed without particular

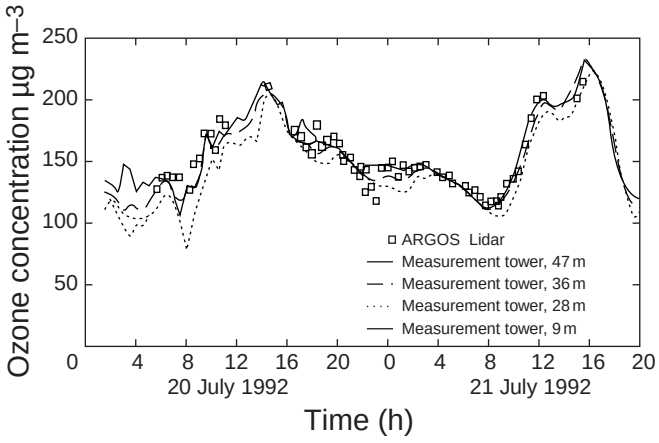

Figure 18 Comparison between DIAL ozone measurements and data from standard analyzers at different altitudes of a tower.[60]

difficulties. An example of such a study has been performed on ozone by Goers[60] in the forest of Albsfelde (Germany). The LIDAR data were compared with standard analyzer data (UV spectrophotometry) collected at different altitudes on a tower, located 1030 m from the LIDAR system. The laser beam was steered to pass near the tower at the altitude of the 47-m high spot measurement. The results, which are presented in Figure 18, show very good agreement (within $\pm 3\,\mu g\,m^{-3}$) between the LIDAR measurements and standard analyzer data.

Other favorable conditions are emission measurements. The high concentration (at chimney tops, for example) allows DIAL measurements at maximum spatial resolution, which again compare well with the in situ spot data.

Difficult cases are urban conditions. The steep gradients and the concentration levels necessarily require path averaging of fluctuations for the DIAL technique and extremely critical location for the reference analyzers. Comparison strategies therefore have to be carefully developed for these situations. The most critical case is the detection of primary pollutants emitted by road traffic, such as NO. An example of comparison strategies in the latter situation was performed in Stuttgart in 1989 (Figure 19). The spot NO analyzer was located at about 20 m altitude on the top of a building in the city center while the DIAL system beam was directed at it. The observed concentration evolutions as a function of time are remarkably close. However, the absolute concentration values from the DIAL system are still systematically lower by about 15% than the spot data. This is due to the spatial averaging mentioned above, since the LIDAR spatial resolution (50 m) includes lower concentrations from higher altitudes.

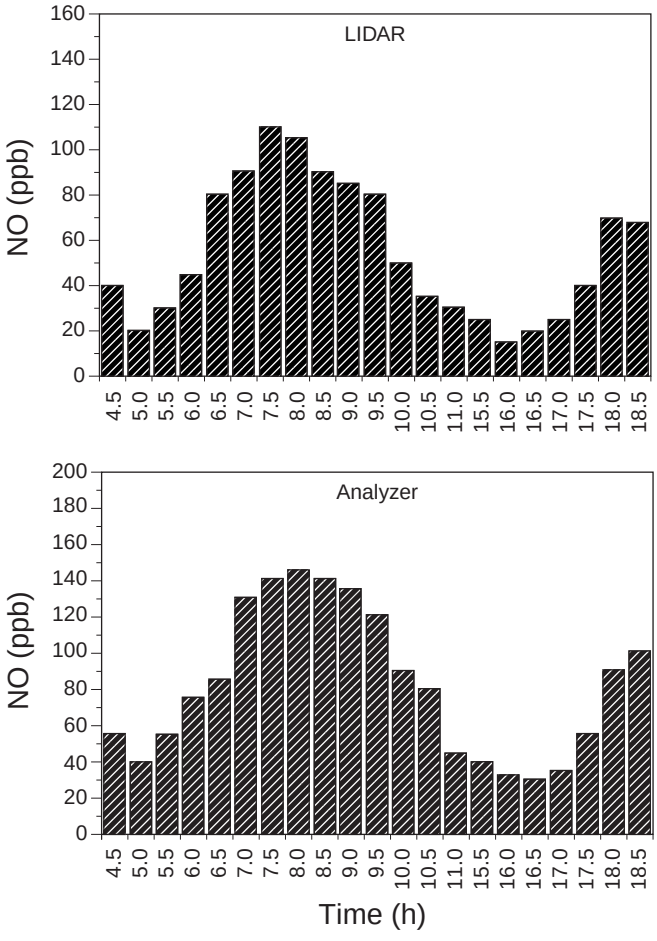

Figure 19 Comparison of LIDAR and a spot analyzer in urban conditions for a primary pollutant (NO).

Comparison with spot analyzers is a valuable calibration method but the tests performed on the system are limited to the concentration levels present in the atmosphere during the measurements. An attractive alternative technique is the use of reference gases. A cell is then located in the DIAL measurement path, in which calibrated concentrations of reference gases are introduced. This induces a supplementary absorption in the light path to the absorption from the atmospheric pollutant, and the difference from a second empty cell provides the reference gas concentration. This technique was used very successfully by the Institut National de l'Environnement et des Risques (INERIS) in Verneuil-en-Halatte for the evaluation of commercial DIAL systems.[61,62] The usual normative parameters[63,64] such as detection limit, linearity, repeatability, zero, cross-sensitivity, and interferences, could be determined with this method. As an example, Figure 20 shows the excellent results of the linearity measurement obtained on SO_2. It is important to note that

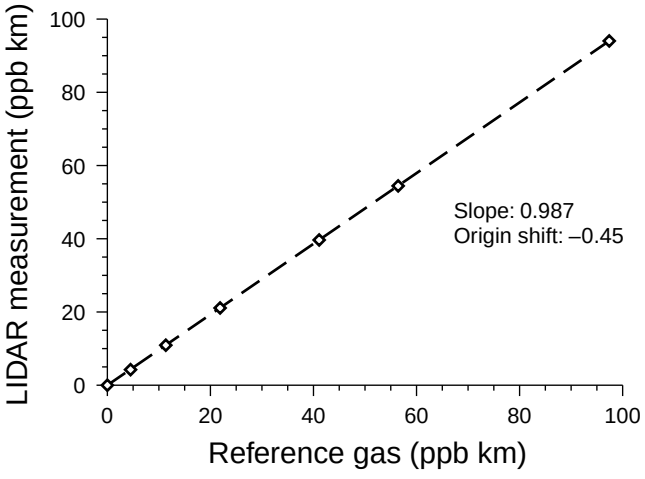

Figure 20 Calibration of LIDAR systems: results of linearity measurements using a reference gas (SO_2).[61]

similar techniques are used today for the calibration of DOAS and FTIR devices, so that a standard calibration procedure can be determined for all open-path optical methods.

8 NEW LIGHT DETECTION AND RANGING STANDARDS

Advances in technology and the advent of commercial systems recently accelerated the normative actions for DIAL systems. Prenormative activities such as the above-mentioned evaluation by INERIS or campaigns from the European Joint Research Center in Ispra[33] have also greatly supported the norms. Pioneering work was performed by the VDI/DIN commission in Germany from 1992 to 1997, which yielded the first normative text for DIAL systems.[18] A similar action is in progress at the European level for the Commission Européene de Normalisation (CEN). These actions are of crucial importance for the wide use by EPAs of this very attractive technique.

9 CONCLUSION

We have presented an overview of today's capabilities of LIDAR systems to provide 3-D profiles of concentrations of gaseous and particulate pollutants and temporal evolution. New normative actions are recognizing the unique advantages of this technique and show the way to operate and calibrate the available systems properly. More than 20 years after the first LIDAR mappings of gaseous species, the wide use of DIAL systems for monitoring air pollution has today been achieved.

ACKNOWLEDGMENTS

The author wishes especially to acknowledge the work of V. Boutou, M. Douard, B. Vezin, E. Frejafon, J. Kasparian, P. Rambaldi, and J. Yu at LASIM, the group of Prof. L. Wöste at the Free University Berlin, P. Rairoux at the AWI-Potsdam, D. Weidauer and M. Ulbricht at Elight Laser Systems GmbH (Teltow/Berlin), E. de Saeger and B. Ottobrini at the Joint Research Center ISPRA, and T. Ménard and E. Vindimian at INERIS, who all actively participated in one or other of the reported studies.

ABBREVIATIONS AND ACRONYMS

CEN	Commission Européene de Normalisation
DAS	Differential Absorption and Scattering
DIAL	Differential Absorption Light Detection and Ranging
DIN	Deutsche Institut für Normung
DOAS	Differential Optical Absorption Spectroscopy
EPA	Environmental Protection Agency
FTIR	Fourier-transform Infrared
INERIS	Institut National de l'Environnement et des Risques
IR	Infrared
LIDAR	Light Detection and Ranging
NW	North-west
OMA	Optical Multichannel Analyzer
OPO	Optical Parametric Oscillator
PMT	Photomultiplier Tube
RADAR	Radio Detection and Ranging
SE	South-east
SEM	Scanning Electron Microscopy
SHG	Second Harmonic Generation
S/N	Signal-to-noise Ratio
SW	South-west
THG	Third Harmonic Generation
UV	Ultraviolet
UVDIAL	Ultraviolet Differential Absorption Light Detection and Ranging
VDI	Verein Deutscher Ingenieure
VOC	Volatile Organic Compound
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Infrared LIDAR Applications in Atmospheric Monitoring

REFERENCES

1. R.M. Schotland, 'Some Observations of the Vertical Profile of Water Vapor by a Laser Optical Radar', in *Proceedings of 4th Symposium on Remote Sensing of the Environment, University of Michigan, Ann Arbor, MI*, 273, 1966.
2. K.W. Rothe, U. Brinkman, H. Walther, 'Remote Sensing of NO_2 Emission from a Chemical Factory by the Differential Absorption Technique', *Appl. Phys.*, **4**, 181–182 (1974).
3. M. Aldén, H. Edner, S. Svanberg, 'Laser Monitoring of Atmospheric NO Using Ultraviolet Differential-absorption Techniques', *Opt. Lett.*, **7**(11), 543–545 (1982).
4. M. Aldén, H. Edner, S. Svanberg, 'Remote Measurements of Atmospheric Mercury Using Differential Absorption LIDAR', *Opt. Lett.*, **7**, 221–223 (1982).
5. K. Fredriksson, B. Galle, K. Nystrom, S. Svanberg, 'Mobile LIDAR System for Environmental Probing', *Appl. Opt.*, **20**, 4181–4189 (1981).
6. K. Fredriksson, 'Differential Absorption LIDAR for Pollution Mapping', in *Laser Remote Chemical Analysis*, ed. R.M. Measures, J. Wiley & Sons, New York, 273–332, 1988.
7. B. Galle, A. Sunesson, W. Wendt, 'NO_2 Mapping Using Laser–Radar Techniques', *Atmos. Environ.*, **22**, 569–571 (1988).
8. B.W. Jolliffe, R.C. Felton, N.R. Swann, P.T. Woods, 'Field Measurement Studies Using a Differential Absorption LIDAR System', in *Proceedings of 12th International Laser Radar Conference, Aix-en-Provence*, 1984.
9. S. Adrian, D.J. Brassington, S. Sutton, R.H. Varey, 'The Measurement of SO_2 in Power Station Plumes with Differential LIDAR', *Opt. Quantum Electron.*, **11**, 253–264 (1979).
10. C. Weitkamp, 'The Distribution of Hydrogen Chloride in the Plume of Incineration Ships: Development of New Measurement Systems, Wastes in the Ocean', J. Wiley & Sons, New York, Vol. 3, 1981.
11. W. Staehr, W. Lahman, C. Weitkamp, 'Range Resolved Differential Absorption LIDAR: Optimization of Range and Sensitivity', *Appl. Opt.*, **24**, 1950–1956 (1985).
12. G. Ancellet, R. Capitini, D. Renaut, G. Mégie, J. Pelon, 'DIAL LIDAR Measurements of Atmospheric Pollutants (SO_2, O_3) during the Fos-Berre 83 Experiment', in *Proceedings of 12th International Laser Radar Conference, Aix en Provence*, 269, 1984.
13. J. Pelon, G. Mégie, 'Ozone Monitoring in the Troposphere and Lower Stratosphere: Evaluation and Operation of a Ground-based LIDAR Station', *J. Geophys. Res.*, **87**, 4947–4955 (1982).
14. G. Mégie, G. Ancellet, J. Pelon, 'LIDAR Measurements of Ozone Vertical Profiles', *Appl. Opt.*, **24**, 3454–3463 (1985).
15. J.G. Hawley, L.D. Fletcher, G.F. Wallace, 'Ground-based Ultraviolet Differential Absorption LIDAR (DIAL) System and Measurements', in *Optical and Laser Remote Sensing*, eds. D.K. Killinger, A. Mooradian, Springer Series in Optical Sciences, Springer, Berlin, 1983.
16. E.V. Browell, A.F. Carter, S.T. Shipley, R.J. Allen, C.F. Butler, M.N. Mayo, J.H. Siviter, W.M. Hall, 'NASA Multipurpose Airborne DIAL System and Measurements of Ozone and Aerosol Profiles', *Appl. Opt.*, **22**, 522–534 (1983).
17. W.B. Grant, R.D. Hake, E.M. Liston, R.C. Robbins, E.K. Proctor, 'Calibrated Remote Measurements of NO_2 Using Differential Absorption Backscatter Technique', *Appl. Phys. Lett.*, **24**, 550–552 (1974).
18. VDI–DIN 4210, 'Fernmessverfahren: Messungen in der Atmosphaere nach dem LIDAR-Prinzip; Messen Gasformiger Luftvereinigungen mit dem DAS-LIDAR', *VDI–DIN Handbuch Reinhaltung der Luft*, Band 5, Beuth, Berlin, 1997.
19. K.H. Eickel, P. Rairoux, M. Ulbricht, D. Weidauer, K. Weber, J.P. Wolf, 'Standardization of Differential Absorption LIDAR Measurements in Germany', in *Quality Assurance and Standardisation of Remote Sensing Methods*, ed. K. Weber, SPIE – The International Society for Optical Engineering, Bellingham, WA, Vol. 3106, 1997.
20. J.P. Wolf, M. Douard, K. Fritzsche, P. Rairoux, G. Schubert, M. Ulbricht, D. Weidauer, L. Wöste, '3D-Monitoring of Air Pollution Using All Solid State LIDAR Systems', in *Proceedings of SPIE Conference on Optical Sensing in Environmental Monitoring*, Atlanta, GA, 200–207, 1993.
21. J.P. Wolf, 'All Solid State DIAL Systems Take Pollution's Measure', *Photonics Spectra*, **4**, 97–98 (1994).
22. P. Rambaldi, M. Douard, J.P. Wolf, 'New UV-tunable Solid State Lasers for LIDAR Applications', *Appl. Phys. B*, **61**, 117–120 (1995).
23. M. Douard, P. Rambaldi, B. Vezin, J.P. Wolf, 'Tunable No-Tracking OPO–OPA Tandem in the Near Infrared Pumped by a Ti:sapphire Laser', *Appl. Phys. B*, **63**, 199–201 (1996).
24. M.J.T. Milton, T.D. Gardiner, F. Molero, J. Galech, 'Injection Seeded Optical Parametric Oscillator for Range Resolved DIAL Measurements of Atmospheric Methane', *Opt. Commun.*, **142**, 153–160 (1997).
25. J.P. Wolf, '3D Monitoring of Air Pollution Using Mobile 'All Solid State' LIDAR Systems', *Opt. Photonics News*, **6**(1), 27 (1995).

26. D. Weidauer, P. Rairoux, M. Ulbricht, J.P. Wolf, L. Wöste, 'Ozone, VOC, NO_2, and Aerosol Monitoring in Urban and Industrial Areas Using a Mobile DIAL System', in *Advances in Atmospheric Remote Sensing with LIDAR*, eds. A. Ansmann, R. Neuber, P. Rairoux, U. Wandinger, Springer, Heidelberg, 423–426, 1996.
27. H.J. Kölsch, P. Rairoux, D. Weidauer, J.P. Wolf, L. Wöste, 'Analysis of the Tropospheric Ozone Dynamics by LIDAR', *J. Phys. III, C4*, **4**, 643–646 (1994).
28. H.J. Kölsch, P. Rairoux, J.P. Wolf, L. Wöste, 'Comparative Study of Nitric Oxide Immission in Lyons, Geneva, and Stuttgart Using a Mobile DIAL System', *Appl. Phys. B*, **54**, 89–94 (1992).
29. M.J.T. Milton, P. Woods, R. Partridge, B. Jolliffe, 'Long Path Methods for Remote Sensing of Pollutants in the Boundary Layer', SPIE – The International Society for Optical Engineering, Bellingham, WA, 312, Vol. 1715, 1992.
30. D. Weidauer, H.D. Kambezidis, P. Rairoux, D. Melas, M. Ulbricht, 'Air Quality in the Athens Basin During Sea Breeze and Non Sea Breeze Days Using Laser Remote Sensing Techniques', *Atmos. Environ.*, **32**, 2173–2183 (1998).
31. J.P. Wolf, 'Optical Techniques for Air Pollution Monitoring', in *European Research Course on Atmospheres*, Book 3, ed. C. Boutron, Editions de Physique, Paris, 1998.
32. E. Frejafon, J. Kasparian, P. Rambaldi, B. Vezin, V. Boutou, J. Yu, M. Ulbricht, D. Weidauer, B. Kraemer, T. Leisner, P. Rairoux, L. Woeste, J.P. Wolf, 'Laser Applications for Atmospheric Pollution Monitoring', *Eur. Phys. J. D*, **4**, 231–238 (1998).
33. E. De Saeger, 'LIDAR Study on the Ozone Dynamics', *Report ERLAP 9*, Joint Research Center, Ispra, 1996.
34. R.M. Measures, *Laser Remote Sensing*, J. Wiley & Sons, New York, 1984.
35. J. Harms, W. Lahmann, C. Weitkamp, 'Geometrical Compression of LIDAR Return Signals', *Appl. Opt.*, **17**, 1131–1135 (1978).
36. J. Harms, 'LIDAR return Signals for Coaxial and Non-coaxial Systems with Central Obstruction', *Appl. Opt.*, **18**, 1559–1566 (1979).
37. J.D. Klett, 'Stable Analytical Inversion Solution for Processing LIDAR Returns', *Appl. Opt.*, **20**, 211–220 (1981).
38. J.D. Klett, 'LIDAR Inversion with Variable Backscatter/Extinction Ratio', *Appl. Opt.*, **24**, 1638–1643 (1985).
39. K. Sassen, G.C. Dodd, 'LIDAR Crossover Function and Misalignment Effects', *Appl. Opt.*, **21**, 3162–3165 (1982).
40. H.J. Kölsch, P. Rairoux, J.P. Wolf, L. Wöste, 'Simultaneous NO/NO_2 DIAL Measurement Using BBO Crystals', *Appl. Opt.*, **28**(11), 2052–2056 (1989).
41. Elight Laser Systems GmbH, *Documentation LIDAR 510 M*, Elight Laser Systems, Teltow/Berlin 1993; http://www.Elight.de.
42. Elight Laser Systems GmbH, 'Mehrwellenlänge-Laseroszillator mit geometrisch gekoppelten Resonatoren', *Ger. Pat.*, P 41 10 189.8-33, 1992; 'Laseroszillator', *Eur. Pat.*, 92 25 0114.3, 1992; 'Laser Oscillator', *US Pat.*, 07/887 347, 1992.
43. P. Rambaldi, M. Douard, B. Vezin, J.P. Wolf, 'Broadly Tunable $KNbO_3$ OPO Pumped by Ti : Sapphire Laser', *Opt. Commun.*, **142**(4–6), 262–264 (1997).
44. M. Douard, R. Bacis, G. Fabre, P. Rambaldi, A. Ross, R. Stringat, J.P. Wolf, 'Fourier Transform LIDAR', *Opt. Lett.*, **20**, 2140–2142 (1995).
45. H. Kölsch, P. Rairoux, J.P. Wolf, L. Wöste, 'New Perspectives in Remote Sensing Using Excimer-pumped Dye Lasers and BBO Crystals', in *Proceedings of 14th International Laser Radar Conference, Innichen San-Candido*, 1988.
46. M. Beniston, M. Beniston-Rebetez, H.J. Kölsch, P. Rairoux, J.P. Wolf, L. Wöste, 'Use of LIDAR Measurements and Numerical Models in Air Pollution Research', *J. Geophys. Res.*, **95**(D7), 9879–9894 (1990).
47. C. Weitkamp, U.B. Goers, J. Glauser, S. Köhnler, P. Rairoux, F. Immler, L. Wöste, M. Ulbricht, D. Weidauer, 'Laser Remote Sensing of Sulfur Dioxide, Nitrogen Dioxide, Toluene, Ozone, and Dust in the Industrial Area of Cubatao (Brazil)', in *Advances in Atmospheric Remote Sensing with LIDAR*, eds. A. Ansmann, R. Neuber, P. Rairoux, U. Wandinger, Springer, Heidelberg, 411, 1996.
48. L. Fiorani, B. Calpini, A. Clappier, L. Jaquet, F. Müller, H. van den Bergh, E. Durieux, 'Ozone Flux Measurements in the Troposphere with a Shot per Shot DIAL Instrument Operated in the UV Spectral Region during the MEDCAPHOT Campaign', in *Advances in Atmospheric Remote Sensing with LIDAR*, eds. A. Ansmann, R. Neuber, P. Rairoux, U. Wandinger, Springer, Heidelberg, 367, 1996.
49. C. Weitkamp, W. Lahmann, W. Staehr, 'Reichweite und Empfindlichkeitoptimiesierung beim DAS-LIDAR', *Laser Optoelektron.*, **4**, 375–380 (1987).
50. N. Menyuk, D.K. Killinger, 'Temporal Correlation Measurements of Pulsed Dual CO_2 LIDAR Returns', *Opt. Lett.*, **6**, 301–303 (1981).
51. N. Menyuk, D.K. Killinger, C.R. Menyuk, 'Error Reduction in Laser Remote Sensing: Combined Effects of Cross-correlation and Signal Averaging', *Appl. Opt.*, **24**, 118–131 (1985).
52. M.J.T. Milton, P. Woods, 'Pulse Averaging Methods for a Laser Remote Monitoring System Using Atmospheric Backscatter', *Appl. Opt.*, **26**, 2598–2603 (1987).
53. Ph.A. Leighton, *Photochemistry of Air Pollution*, Academic Press, New York, 1961.
54. National Research Council, 'Rethinking the Ozone Problem in Urban and Regional Air Pollution', National Academy Press, Washington, DC, 1992.
55. S.E. Grying, *MEDCAPHOT Report*, European Community, Brussels, 1994.

56. D. Weidauer, 'Analyze und Bewertung von Smog Situationen mittels LIDAR', PhD Thesis, Freie Universität Berlin, 1998.
57. J. Kasparian, E. Frejafon, P. Rambaldi, J. Yu, B. Vezin, P. Ritter, P. Viscardi, J.P. Wolf, 'Characterization of Urban Aerosols Using SEM-microscopy, X-ray Analysis and LIDAR Measurements', *Atmos. Environ.*, **32**, 2957–2967 (1998).
58. E. Frejafon, J. Kasparian, P. Rambaldi, J. Yu, B. Vezin, J.P. Wolf, '3D Analysis of Urban Aerosols Using A Combined LIDAR–SEM–X-ray Method', *Appl. Opt.*, **37**(12), 2231–2237 (1998).
59. M. Ulbricht, P. Rairoux, J. Reif, D. Weidauer, J.P. Wolf, 'Calibration Procedures of LIDAR Systems', in *Proceedings of SPIE Conference on Optical Sensing in Environmental Monitoring, Atlanta, GA*, 194–197, 1993.
60. U.B. Goers, 'Laser Remote Sensing of Sulfur Dioxide and Ozone with the Mobile Differential Absorption LIDAR Argos', *Opt. Eng.*, **34**(11), 3097–3102 (1995).
61. INERIS, 'Préévaluation du LIDAR 510 M Commercialisé par ELIGHT GmbH', T. Ménard, Y. Godet, E. Vindimian, M. Nominé, *Rapport INERIS No 526/96*, Institut National de l'Environnement Industriel et des Risques, Verneuil-en-Halatte, 1996.
62. T. Ménard, E. Vindimian, P. Rambaldi, M. Douard, B. Vezin, D. Weidauer, M. Ulbricht, J.P. Wolf, 'Calibration and Evaluation of LIDAR Systems', *Environ. Sci. Technol.*, submitted.
63. ISO 5725-1, 'Genauigkeit von Messverfahren und Messergebnissen', International Organization for Standardization, Geneva, 1996.
64. VDI 3950 Blatt 1, 'Kalibrierung Automatischer Emissioneinrichtungen', Beuth, Berlin, 1994.

Environmental Analysis of Water and Waste: Introduction

Marvin P. Miller
Dow Chemical, Midland, USA

The field of environmental analysis is driven by demand for development of analytical methods providing ever-increasing specificity and throughput, decreases in detection limits, and improvements in precision and accuracy. Typically, meeting these conflicting demands requires a compromise of some type. The methods with the lowest detection limits are seldom those that simultaneously provide the best specificity or most cost effective analysis. Concurrently, the number of analytes of interest in environmental samples continues to expand rapidly while the type and complexity of samples grows as well. Thirty years ago, most environmental analyses were conducted on water samples with perhaps 10–15 analytes per sample. Today samples of interest include water, waste water, soils, solid or liquid waste, and biological materials, and at times the list of analytes to be determined can number several hundred. For these reasons, the variety of analytical methodology utilized for environmental analysis is extensive. The number of approved methods by the United States Environmental Protection Agency (USEPA) alone exceeds several hundred. This section will provide a comprehensive overview of modern environmental methodology for the analysis of water and waste samples. Articles will describe both the application of a wide variety of instrumental methodology, as well as techniques for determination of an extensive array of parameters in many different matrices, including water, waste water, soils, hazardous wastes, and biological materials. Methods specific to field analysis are presented in the section entitled *Field-portable Instrumentation*.

Development of environmental analytical methodology, especially those methods in common usage, has frequently been driven by monitoring regulations and requirements. This is especially true in the United States, where the passage of numerous acts has led to the creation of a variety of target analyte lists. The Federal Water Pollution Control Act or 'The Clean Water Act' (1972) led to lists of analytes to be determined in wastewater discharge streams. As a result of a subsequent law suit, ''N.R.D.C. vs. Train, 8ERC2120'', an extensive list of 129 parameters was established and analysis of this entire list was required for all discharge streams. The Safe Drinking Water Act (1974) resulted in another specific list of 14 parameters that were to be determined in drinking water samples. The Resource Conservation and Recovery Act (1977) and subsequent amendments introduced the Appendix VIII and Appendix IX lists of target analytes for waste and groundwater samples. These list several hundred specific parameters or categories.

In some cases when these lists where first developed, the analytical methodology did not exist to make accurate determinations in the complex matrices of interest. Thus, the legally mandated analyte lists drove the development of more sensitive and selective methods in many areas. Furthermore, enforcement of these laws created tremendous numbers of samples to be analyzed, which created the demand for methods with faster throughput. Thus, the rapid advancements in the field of environmental analytical chemistry and the adoption of sophisticated methodology were in many ways the result of regulatory requirements. For example, early analysis of wastewater samples for many organic species was conducted by a variety of gas chromatographic techniques, requiring several separate analyses to complete one sample. The development of the gas chromatography/mass spectrometry (GC/MS) methods 624 and 625 provided more specific analysis of the organic parameters on the priority pollutant list with only two determinations. Initially, few laboratories had access to expensive, sophisticated GC/MS methods. However, the large volume of samples requiring analysis across the world quickly drove demand for improvements in the instrumentation. Today, GC/MS is relatively inexpensive and is part of every well-equipped environmental laboratory. A similar situation has occurred for the determination of elemental species. The use of inductively coupled plasma (ICP) and ICP coupled with a MS (mass spectrometry) detector (ICP/MS) has greatly enhanced the sensitivity and speed of analysis relative to the use of flame and flameless atomic absorption spectrophotometry.

As is the case in all fields of analytical chemistry, a critical component in environmental analysis is to conduct all determinations under the guidance of a carefully designed quality assurance (QA) program. In many cases, environmental analyses also must be conducted following strictly designed regulatory requirements, dictating such things as sampling and analysis methodology, quality control procedures, and complete documentation policies. Specific legally mandated QA requirements can vary significantly depending upon many parameters, including the country, the regulating agency, and the sample type. While the specific QA requirements can change, the fundamental concepts of QA do not. The introductory article in this section, **Quality Assurance in Environmental Analysis**, written by Malcolm Clark of the Canadian Ministry of Environment, provides a comprehensive discussion of the primary elements of QA. The author covers QA requirements from initial program design through sample collection, analysis, and final reporting. He also discusses

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

types of errors that can occur and means of determining detection limits. The use of certified reference materials, interlaboratory testing and legal QA requirements are discussed, and the author discusses the use of performance based methods, an area of particular emphasis in recent years for the USEPA. Robert Gibbons in **Detection and Quantification of Environmental Pollutants** provides a detailed discussion of a number of means to express detection and quantification data, including applications to limit both false positive and false negative results.

No matter how carefully the environmental analytical chemist conducts the laboratory determination, the quality of the data is totally dependent on the quality of the sample taken. The sample must be representative of the body it represents and it must be maintained in the appropriate manner to prevent contamination and/or deterioration of the analytes of interest. Once again, the vast array of sample matrices and analytes leads to the requirement for many different sampling protocols to meet the above requirements. The following series of articles covers the requirements for sampling in a variety of matrices of environmental interest.

Biological samples can provide an excellent indication of exposure to pollutants, as discussed in **Biological Samples in Environmental Analysis: Preparation and Cleanup** by Steven Barker. Barker further describes means for biological sample preparation and cleanup, including gel permeation, super critical fluid extraction, and solid phase extraction. Sampling for biological monitoring requires additional considerations in order to ensure the appropriate data are collected. **Sampling Considerations for Biomonitoring** are covered by Tim Koeckritz et al. who focus on three key axioms for sampling, i.e. chemical composition, distribution of organisms, and the economic aspects of the information content. While these axioms were developed for biological materials, they are nonetheless relevant with minor modifications to all types of sampling programs.

Soil Sampling for the Characterization of Hazardous Waste Sites is discussed by Brian Schumacher et al. who begin with a discussion of how to plan the sampling program, including definition of the data quality objectives for the program. They discuss considerations for selecting the appropriate spatial sampling strategy and ways to estimate variance. Site mapping, site characterization, and sample collection are described, as well as selection of sampling equipment. **Industrial Waste Dumps, Sampling and Analysis** is the topic of the next article, written by Winfried Rasemann. He discusses some of the basic problems encountered in sampling industrial waste dumps and examples of how these problems can be avoided by statistical evaluations.

Environmental analysis can be divided into three primary categories; organic, inorganic, and 'conventional'. The term conventional is used to describe a variety of determinations that include total carbon, biological oxygen demand, and in some cases ionic determinations by such techniques as ion selective electrode (ISE) or ion chromatography (IC). Asbestos analysis does not fit well in any of these categories, however the widespread use of asbestos in the past and the well-known hazards of exposure makes this a determination of particular interest around the globe.

Norber Hofert discusses the determination of airborne asbestos fibers and asbestos in bulk materials and drinking water in the article **Asbestos Analysis**. The use of scanning, transmission, and phase contrast microscopy is described, as well as the use of X-ray diffraction and infrared techniques.

The determination of total and dissolved organic carbon is one of the most commonly used methods for quickly assessing the overall water quality. Kenneth Mopper and Jian-guo Qian, compare the two primary methods, wet oxidation and high temperature combustion, and examine problems encountered when analyzing sea water in the article **Water Analysis: Organic Carbon Determinations**.

The use of ISEs is widespread in environmental determinations for a large number of analytes. **Ion-selective Electrodes in Environmental Analysis** are described by Robert Forster and Tia Keyes. The authors discuss the advantages of ISEs (portability, sensitivity, wide dynamic range, ease of use, continuous measurement, etc.) They provide a theoretical background for the technique, and a detailed discussion of the various types and applications of ISEs, including determination of metallic species as well as anions and cations. A comprehensive table of analytes and sensitivities that can be achieved with ISEs is provided.

The next series of articles in this section cover the determination of elemental species by a variety of techniques. The series begins with four articles on different sample preparation procedures used on various matrices. First, Sergio Marin discusses **Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices**. Discussions are provided on proper sample collection, handling, and storage, as well as appropriate laboratory conditions for minimizing the potential for contamination. A variety of digestion or extraction procedures are discussed, including wet digestion, dry ashing, sequential extraction, and pressurized digestion. **Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges)** is described by D. Kimbrough beginning with a discussion of ways to physically prepare the samples (milling, sieving, etc.). Kimbrough reviews several USEPA acid digestion procedures and discusses the application of each to different matrix types. He also describes methods for extraction of organic compounds from solids using solvent extraction.

The preparation of biological samples for elemental analysis requires special consideration and is described by Kunnath Subramanian in the article entitled **Sample Preparation for Elemental Analysis of Biological Samples in the Environment**. Protein precipitation, chelation, and solvent extraction are discussed, as well as a variety of procedures to remove the organic matter from the samples. Specific considerations are provided as a function of the element of interest, the sample matrix (e.g. urine, hair, bone, tissue), and the instrumentation to be used (e.g. ICP, ICP/MS). Mehdi Ashraf-Khorassani, Larry Taylor and Michael Combs, provide their insight on the use of **Supercritical Fluid Extraction of Inorganics in Environmental Analysis**. While direct extraction of cationic metal ions is not feasible with supercritical CO_2, in situ chelation of the metal ions can enhance the extraction efficiency. The authors examine the solubility of different ligands and metal chelates in supercritical fluids and assess the impact of the matrix as well.

The next series of articles covers a variety of instrumental techniques used for the determination of elemental species. Perhaps the most commonly used, and certainly the instrumental technique that has the longest history, is the use of atomic absorption spectroscopy (AAS), both by flame and graphite furnace. This topic is covered extensively by Michael Sperling, **Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis** starting with a comparison of detection limits for both flame and furnace AAS. Sperling provides a very useful summary of instrument fault recognition and means for optimization of the instrument. Spectral and nonspectral interferences and AAS specific QA are described. Sperling follows with a comprehensive discussion of the application of both flame and furnace AAS for analysis of waters, soils and other solid materials.

For certain elements with higher volatility (e.g. arsenic, selenium), conventional flame and furnace AAS techniques may not provide adequate sensitivity for some environmental concerns. One means to improve the sensitivity for these parameters is by use of hydride generation for sample introduction. Hisatake Narasaki discusses in **Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples**, including proper sample preparation, detection limits and application to not only AAS but also both optical emission ICP and ICP/MS.

Scott Baker and Nancy Miller-Ihli describe another sample introduction technique that can be used to enhance sensitivity, particularly for solid samples, **Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses**. The use of slurry sampling into the graphite furnace is described, including methods for producing and optimizing slurries of a variety of matrix types. Instrumental conditions are described and examples of environmental applications are given, as well as comparisons to other direct analysis methods.

The determination of mercury is of particular interest due to its toxicity and propensity to bio-accumulate. Unfortunately, neither flame nor graphite furnace AAS can provide a detection limit adequate for determination of mercury at levels of environmental concern, typically in concentrations of low to sub-part per billion. The most commonly used technique **Mercury Analysis in Environmental Samples by Cold Vapor Techniques**, is discussed here by Kim Anderson. The author describes the sample preparation technique, and the use of the specific instrumentation required, as well as sensitivities and common interferences. The sensitivity of cold vapor (CV) AAS is compared with other potential techniques, such as atomic fluorescence spectroscopy (AFS), and a summary of mercury regulatory standards around the world is given.

While AAS is an extremely valuable technique, the throughput is limited, particularly when using the graphite furnace. The introduction of inductively coupled plasma atomic emission spectroscopy (ICPAES) led to the rapid advancement of methodology to take advantage of the simultaneous multi-elemental analysis capability while providing detection limits equivalent or better than graphite furnace AAS. **Optical Emission Inductively Coupled Plasma in Environmental Analysis** is described by Mark Tatro. The author outlines available USEPA ICPAES methods and discusses means for overcoming common interferences. ICPAES specific QA requirements are also described.

In some cases, even ICP optical emission spectroscopy cannot provide sufficient sensitivity or specificity for certain elements in certain matrices. For example, optical ICP generally cannot be used for determination of arsenic, selenium, and mercury in drinking water samples because the detection limits are not adequate. The development of ICP coupled with a mass spectrometer as a detector provides an additional tool with enhanced sensitivities for many elements. **Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis** is described by S. Yamasaki who reviews the instrumentation and provides an overview of the types of sample introduction used with ICP/MS. The typical detection limits are provided for elemental analysis of environmental samples in water, soils, and biological materials. The author also describes the use of high resolution ICP/MS which can be used to provide enhanced selectivity.

The determination of specific elemental forms is described by Wang Xiaoru and Frank Sen-Chun Lee, in their contribution entitled **Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis** (CE/ICP/MS). The authors describe optimization of the capillary

electrophoresis (CE) separation procedure and compare the technique to conventional ICP/MS. A comparison is made of a number of nebulizers that have been used for CE/ICP/MS.

Conventional ICP or ICP/MS analysis requires the dissolution of the sample for introduction into the instrument as a solution. In addition to the time required for sample preparation, the dissolution can result in decreased sensitivity. **Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis**, as described by Steven Durrant, can provide a means for the direct vaporization of a solid sample and subsequent introduction into either an optical ICP or ICP/MS. The author describes several laser ablation systems and provides analysis procedures, including its use for samples of environmental interest, such as metals, geological, and biological materials.

In some cases, it is of interest to detect organic species containing specific elements. Stig Pedersen-Bjergaard has provided an article on **Gas Chromatography with Atomic Emission Detection in Environmental Analysis**. The column first separates the organic species and the elemental detector can be tuned to the parameter of interest. The author describes the selectivity and sensitivity of the method, as well as a compound independent calibration. He provides examples of several environmental applications for compounds containing specific elements such as tin, lead, and mercury, as well as the use for other compounds containing non-metallic elements.

Yong Cai discusses the use of another alternative technique for elemental analysis, **Atomic Fluorescence in Environmental Analysis**. AFS can provide enhanced detection limits for some elements, and is of particular interest for the ultratrace level determination of mercury. In AFS, the sensitivity is directly proportional to the intensity of the light source, thus the author describes the use of hollow cathode lamps, electrodeless discharge lamps, xenon arc lamps and lasers. The determination of mercury by AFS is covered in detail, including methodology for organometallic determination of mercury species. Hydride generation is also described as a means to further enhance sensitivity and selectivity for volatile elements such as arsenic and selenium.

Flow-injection Techniques in Environmental Analysis are described by J.L. Burguera and M. Burguera. The authors describe how flow injection spectrophotometric methods can be used as alternatives to AAS and ICP techniques. In addition, the authors describe how flow injection can be used in conjunction with other instrumental methods such as AAS or ICP. A comprehensive list of the use of flow injection in environmental analysis is provided.

A recent addition to the tools used in environmental analysis for separation of complex samples, CE, is discussed by Yang Sun Kim in the article **Inorganic Analysis in Environmental Samples by Capillary Electrophoresis**. The author outlines the use of this technique for environmental samples, specifically for the determination of anions, metal speciation, metal complexes, and organometallic compounds. Comparison is also made with alternative methods such as IC, ICP and ICP/MS.

Inorganic Environmental Analysis by Electrochemical Methods offers an alternative to the commonly applied spectroscopic methods for the determination of many elements in a variety of matrices. This technique is discussed by Guler Somer. Polarographic and voltametric methodology is covered, and application to air, drinking water, rivers, lakes, sea water, and wastewater, as well as for soil and sediment, is reviewed.

The determination of inorganic anions, cations, transition metals and low molecular weight organic acids and bases by IC is described by Peter Jackson in the article entitled **Ion Chromatography in Environmental Analysis**. The author provides a brief history of the technique and a comparison with alternative methodology. Sampling handling, preservation, preparation and clean-up are described, and the use of IC in regulatory methods is covered, including specifics for inorganic anions, hexavalent chromium, cyanide and others. The use of IC for a variety of matrix types such as water, brines, soils, and sludge, is discussed.

Neutron activation analysis (NAA) provides excellent sensitivity, multi-element capability, ease of sample preparation, and freedom from interferences. However, it also requires access to a reactor facility in order to irradiate the samples. This powerful technique is discussed by Keith Randle, in **Neutron Activation in Environmental Analysis**. The author provides a brief history of the technique, followed by a discussion of the use of NAA for environmental samples with a table of ideal sensitivities. Application to biological samples is also discussed. The focus of the article is on the analysis of water samples by NAA, as well as the analysis of soils and sediments.

Another alternative technique for the determination of elements in environmental is samples is **Proton-induced X-ray Emission in Environmental Analysis**, or PIXE. Like NAA, PIXE provides simultaneous determination of a large number of elements and high throughput. However, as described by Grazia Ghermandi, in order to achieve sensitivities of environmental interest, a pre-concentration technique is frequently required. The author describes the methodology and its application to environmental samples such as water, sediments, and atmospheric aerosols. Matrix effects and interferences are discussed, as well as typical detection limits and

ranges for a variety of elements. The author also outlines the procedure to follow to develop a method for environmental analysis using PIXE.

X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples is the title of the next article, written by Philip Russell. The author includes descriptions of energy dispersive (EDXRF), wavelength dispersive (WDXRF), and total reflection X-ray fluorescence (TRXRF) and their use in the elemental analysis of natural waters, waste waters, hazardous wastes, and samples of interest to the nuclear industry. A section is also included on development of both EDXRF and WDXRF methods and on comparison with other spectroscopic methods.

Heavy Metals Analysis in Seawater and Brines is of particular interest due to the desire to study bioaccumulation of these parameters. However, because of the high salt content of the matrix, special procedures have been developed to minimize the matrix affects. Osvaldo Troccoli reviews this topic. Preconcentration and separation are described, as well as the most commonly encountered interferences observed using atomic spectroscopic techniques. A comparison is given among the spectroscopic procedures and electrochemical methods.

The primary focus of the determination of elements in environmental samples continues to be on the determination of the total amount of the element present, regardless of the oxidation state or the specific compound present. However, as more information has become available on the significant variation in toxicity of different organometallic compounds containing the same element of interest, methods have been developed to determine specific organometallic compounds as they occur in the environment. **Organometallic Compound Analysis in Environmental Samples** is addressed by Les Ebdon and Andy Fisher. The authors describe specific issues that must be considered when collecting and storing samples for subsequent analysis for organometallic compounds. Sample preparation procedures are given to extract the analyte of interest from liquid or solid samples. In some cases, a derivatization technique may be required in order to ensure that the analyte is in the appropriate form. A variety of analysis techniques are described, including both GC (gas chromatography) and LC (liquid chromatography) coupled with atomic spectrometry, and CE coupled with atomic spectroscopy. QA requirements specific to organometallic analysis are also discussed.

The overall characteristics of soil quality and a summary of soil analysis methodology is provided by M.R. Carter et al., in their article entitled **Soil Instrumental Methods** The authors describe proper soil sampling procedures and statistical sampling methods, as well as discussing chemical, biochemical, and physical testing procedures.

As is the case for inorganic parameters, the determination of organic species of environmental interest frequently requires some means of specialized sample preparation in order to separate the analyte of interest from complex matrices. The first article in this section of sample preparation, **Soxhlet and Ultrasonic Extraction of Organics in Solids**, contributed by G.R. Barrie Webster, provides a comparison of two primary extraction procedures for solid materials, ultrasonic and Soxhlet extraction. A comparison is made between USEPA methods and alternate procedures and specific QA and trouble shooting considerations are discussed.

Another alternative extraction procedure for organic species is discussed in the article, **Solid-phase Microextraction in Environmental Analysis** is provided by J. Pawliszyn. Two distinct solid phase microextraction (SPME) coating types are discussed, and a theoretical description of their use is given. The use of SPME for field analysis with a portable GC to determine volatile organic compounds (VOCs) in groundwater, soil, and air. Field evaluations of recently developed prototype systems are described.

The use of microwave energy to assist in solvent extraction of organic species is the topic of two articles. First, Frank Smith and Guohua Xiong have contributed the article entitled **Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis**. The authors provide an extensive review of the use of microwave assisted extraction for a wide variety of organic parameters of environmental interest in soils, sediments, plant tissue, air and aqueous matrices. They also cover the use of focused microwave assisted Soxhlet extraction, microwave heating headspace analysis, and microwave-assisted derivatization. John Dean in his article entitled **Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis**, discusses the theoretical basis of microwave extraction of organic analytes and describes both atmospheric and pressurized microwave extraction techniques. Comparisons are made with Soxhlet, ultrasonic, and supercritical fluid extraction (SFE) for polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and pesticides in standard reference materials and other samples.

In the article, **Supercritical Fluid Extraction of Organics in Environmental Analysis** Bruce Benner provides a theoretical discussion of supercritical fluids and their use for extraction of organic species from complex matrices. Reviews of several models of the supercritical extraction process are given, followed by a description of the instrumentation commonly used for SFE. An extensive review of the use of SFE for a wide range of organic classes is provided, including PCBs, pesticides, PAHs,

and hydrocarbons. Studies of extractions conducted on standard reference materials are discussed, comparing SFE with other extraction procedures.

Membrane-based non-chromatographic continuous separation techniques can also be used in environmental analysis for increased selectivity. **Pervaporation, Analytical** by M.D. Luque de Castro and L. Gamiz-Gracia reviews the principles of pervaporation and its use in analytical separation. The laboratory scale pervaporator is described, including a discussion of various membranes that can be used for different analytes. The sensitivity and selectivity of pervaporation is compared with headspace analysis and systems for multi-determinations are described.

GC is likely the most commonly used methodology for the determination of organic species in environmental samples. The need to determine a variety of classes of organic compounds at very low concentrations in frequently very complex matrices has led to the development of many selective detectors, each with specific advantages for certain compound classes. In **Trace Organic Analysis by Gas Chromatography with Selective Detectors**, Erwin Rosenberg provides a review of selective GC detectors, including electron capture, nitrogen-phosphorus, flame photometric, photoionization, and electrolytic conductivity (Hall) detectors. The operating principles and applicability of each type are discussed, as well as relative sensitivities. The author also provides a review and comparison of mass spectrometric, Fourier transform infrared, and atomic emission detectors for GC. Hiroyuki Kataoka provides specific focus on the use of selective detectors for the determination of amines in **Gas Chromatography with Selective Detectors for Amines**. The use of thermionic, flame photometric, electron capture, chemiluminescence, electrolytic conductivity, and mass selective detectors for various amine compounds is reviewed, with comparisons of sensitivities and applicability to aliphatic, aromatic, heterocyclic, and nitrosamines.

Most methods for the determination of organic compounds in aqueous samples requires some means of pre-separation from the matrix, such as solvent extraction, headspace analysis, or solid phase extraction. However, some water soluble analytes cannot easily be separated by these means, or it may be desirable to limit alteration of the matrix. Marek Biziuk has contributed **Gas Chromatography by Direct Aqueous Injection in Environmental Analysis**, a method requiring no separation prior to analysis. The author reviews techniques for direct aqueous injection onto both packed and capillary columns, including discussions of selectivity, sensitivity, and applicability. On column injection is discussed, including application to determination of haloforms in water using electron capture detection. Data are provided for the use of direct aqueous injection in determining organohalogen parameters in tap, surface and swimming pool waters, and for the determination of semivolatile and nonvolatile compounds.

While GC can be used with a variety of selective detectors to provide excellent sensitivity, in some cases the selectivity is not sufficient for positive identification of specific analytes due to the potential for co-elution of multiple compounds, even with the excellent resolution of modern capillary columns. By using a mass spectrometer as a detector for GC, another dimension is provided, allowing identification based on unique mass fragmentation patterns. **Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry**, contributed by Matthew Booth and David Powell, provides a review of the history of the development and use of GC coupled to quadrupole MS. The two principle methods of ionization, electron and chemical, are described, and sample preparation techniques for volatile and semivolatile analytes are reviewed. Special considerations are given for the use of GC/MS instrumentation.

While GC/MS is a very powerful technique for the determination of volatile and semi-volatile compounds, many organic compounds of environmental interest cannot be determined because they are non-volatile, polar, thermally labile, or ionic. M. Careri discusses the coupling of LC to MS, in **Liquid Chromatography/Mass Spectrometry in Environmental Analysis**. The author reviews types of instrumentation used for LC/MS, LC/MS/MS (tandem mass spectrometry) and LC/ICP/MS and their application for the environmental determination of pesticides, PAHs, dyes, surfactants, and organometallic compounds.

Organic acids are wide spread in the environment, and cannot be determined at trace levels via GC methods. Sigrid Peldszus has provided **Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of**. The author reviews sources and samples of interest for organic acids, such as drinking water, waste water, and landfill leachates. The application of IC is discussed and sampling and preservation procedures are provided for various matrices. Specific sample preparation and analysis procedures are reviewed for air, drinking water, wastewater, and ground water samples, and comparison of IC with other methods or analysis of organic acids is given.

The next four articles describe alternate analytical methodologies that can be applied for the determination of organic analytes in environmental samples. Tom Viser has provided **Infrared Spectroscopy in Environmental Analysis**, where the author reviews application of infrared analysis to determination of a variety of VOCs in indoor air, atmospheric, and stack samples. In **Organic Analysis in Environmental Samples by Capillary Electrophoresis**,

Carmen Garcia-Ruiz et al. describe the use of CE for the separation of ionic compounds with low molecular weight and sufficient water solubility, compounds typically difficult to determine by conventional GC techniques. The use of CE for the determination of herbicides, phenols, amines, dyes, and aromatic sulfonic acids are described. In **Organic Analysis in Environmental Samples by Electrochemical Methods**, Ignacio Naranjo-Rodriguez and Jose Luis Hidalgo-Hidalgo de Cisneros describe the use of differential voltammetry, stripping methods and electrochemical detection. The authors outline the use of electrochemical detection in regulatory methods in the US and Europe, and provide specifics for determination of aromatic amines, aromatic hydrocarbons, phenols, PCBs, and other nitrogen containing compounds. Janet MacFall and Tony Riberio describe the use of nuclear magnetic resonance (NMR) spectrosopy in **Nuclear Magnetic Resonance for Environmental Monitoring**. NMR can be used for direct determination of samples, such as plants and soils, for analytes including pesticides, phosphorus, and water. Field application for study of soils is also described.

Analysis of organic compounds in the environment can also be categorized by compound class. The following series of articles reviews methodology for the determination of organic compound classes of particular environmental concern. In **Trihalomethanes in Water, Analysis of**, Louis Lepine and Roland Gilbert compare the primary analytical methods used for determining trihalomethanes in drinking water, key regulated analytes resulting from water disinfection with chlorine. The authors compare direct aqueous injection GC, liquid–liquid extraction with GC analysis, headspace GC, and purge and trap GC. An alternative disinfectant to chlorine for drinking water is the use of chloramine, frequently used to reduce the formation of trihalomethanes. However, chloramine can lead to the formation of cyanogen chloride, as discussed by Yuefeng Xie and Cordelia Hwang, in their article, **Cyanogen Chloride and Cyanogen Bromide Analysis in Drinking Water**. The authors outline the use of purge and trap GC/MS, headspace GC with electron capture detection, and liquid–liquid extraction methods.

Billions of pounds of dyes are used globally each year, as discussed by J. Riu and D. Barcelo in their submission entitled **Dyes, Environmental Analysis of**. Because of the nonvolatile nature of most dyes, LC/MS is commonly used for their analysis. The authors discuss thermospray LC/MS applications and the use of atmospheric pressure ionization interfaces. CE is also discussed for analysis of sulfonated azo dyes.

In his paper, **Explosives Analysis in the Environment**, Bruce Tomkins, describes the use of liquid–liquid extraction followed by high-performance liquid chromatography (HPLC) analysis for the determination of explosives in aqueous samples and reviews the use of USEPA Method 8330 for the determination of explosives in soils. Field screening techniques, such as colorimetric and antibody based tests, are evaluated and the metabolism of explosives in the environment are reviewed. Torsten Schmidt et al. review the **Nitroaromatics, Environmental Analysis of**. The relevant USEPA and other regulatory agency methods are cited and their application discussed. Sample preparation procedures for water, soil, and particulate samples are reviewed. In situ analytical methods including electrochemical, immunoassays and sensors are described, as well as laboratory analysis via GC, GC/MS, and HPLC.

Formaldehyde, Environmental Analysis of is reviewed by S.S. Srivastava and K. Maharaj Kumari. Colorimetric spot test, titrimetric, polarographic, and chromatographic methods, including GC, GC/MS, and HPLC, are discussed, as well as infrared and Fourier transform infrared analysis for formaldehyde in air samples.

Phenolic compounds are found both in nature and from industrial production, and phenol itself has been used as a disinfectant since 1865. Monika Moeder in **Phenols Analysis in Environmental Samples**, discusses the uses and toxocologic importance of phenols and regulations impacting their use and disposal. The commonly used analytical techniques for phenol determination are reviewed, including GC/FID (gas chromatography/flame ionization detector), GC/ECD (gas chromatography/electron capture detection), and GC/MS. The author also reviews the colorimetric method commonly used for determination of phenols as a group, normally referred to as 'total phenol'. The application of different sample preparation procedures to various matrices is discussed and conventional techniques are compared with new developments in biosensors and immunoassay.

Michael Fisher and Israel Schechter have contributed a review of **Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples**, beginning with a definition of PAHs and their importance in the environment. Various extraction procedures, including Soxhlet, ultrasonic, supercritical fluid, and microwave assisted are reviewed for use in PAH analysis, followed by solid phase extraction and gel permeation chromatography clean-up procedures. Analysis by GC, HPLC, GC/MS, and optical methods is discussed, as well as recent procedure utilizing immunoassay techniques. Andres D. Campiglia reviews the determination of PAHs by luminescence techniques, in the article entitled **Luminescence in Environmental Analysis**. The author compares fluorimetry, phosphorimetry, and Shpol'skii spectrometry and their use for the determination of PAHs both in the laboratory and using laser excitation for field and remote analysis.

Another group of organic compounds of common interest in environmental analysis is the PCBs, consisting of 209 separate isomers. This topic is reviewed in **Polychlorinated Biphenyls Analysis in Environmental Samples** by Bert van Bavel et al. The authors provide a short history of PCB analysis in environmental samples and discuss the two most commonly used analytical methods, GC with electron capture detection and GC/MS. Procedures are reviewed for sample clean-up and fat removal and for sample extraction from a variety of matrices. Recent developments in SFE coupled to LC and GC coupled to time of flight MS are also reviewed.

The ability to rapidly determine parameters of environmental interest, especially in the field, is of particular interest both to provide direction in sampling of potentially contaminated sites and to minimize the number of samples requiring more time consuming and costly analysis back in the laboratory. The next three articles provide insight into recent developments in techniques for rapid deployment in the field. First, in **Volatile Organic Compounds in Groundwater, Probes for the Analysis of**, B.M. Patterson and G.B. Davis describe the use of VOC probes for the determination of petroleum products and chlorinated solvents in groundwater in the field. The authors provide a review of alternative field methods for VOCs and comparison with the VOC probes, based on a coil of silicone tubing, are given for actual field samples. The advantages of the VOC probes include exclusion of high molecular weight organic compounds from the sample, elimination of the need to collect, transport, and store samples, and higher sensitivity vs. conventional methods. Disadvantages include a lack of specificity due to the nature of the sensor currently used.

The analysis of dioxin and similar compounds has been of primary interest to the environmental analyst for a number of years. Traditional methods of analysis require extensive sample clean up and pre-concentration techniques, followed by highly specific GC/MS analysis, in order to attain the detection limits required for this class of compounds. Nigel Bunce and John Petrulis describe an alternative screening technique in their article entitled **Dioxin-like Compounds, Screening Assays**. The use of bioassay methodology, such as receptor-ligand binding assays, saturation assays, gel retardation assays, competition assays, and others are described and their applications for the determination of dioxin like compounds are reviewed. While bioassays cannot replace direct analysis techniques due to the lack of chemical identification, they can provide rapid, expensive screening to help eliminate negative samples from requiring further analysis. Barry Lesnik in **Immunoassay Techniques in Environmental Analyses**, discusses the broader application of immunoassay techniques. The author describes the USEPA interest in these rapid screening methods and provides a background on the development of methods within the USEPA in general and for immunoassay methods in particular. The use of enzyme linked immunosorbent assay (ELISA) for the determination of PCB contamination is described, including a discussion on specificity. Alternative immunoassay techniques are described for determination of dioxins, furans, PAHs, and explosives. Specific considerations for immunoassay method development are presented as well as method validation criteria and a summary of current USEPA methods development programs.

Determination of whether or not a waste is hazardous is of key interest in determining appropriate disposal procedures. Several methods have been developed by the USEPA and others in order to classify wastes as hazardous. In **Waste Extraction Procedures**, Barton Simmons describes the history of the development of waste characterization procedures, including the Extraction Procedure and subsequent Toxicity Characteristic Leaching Procedure. Leachate preparation and analysis is described, including the use of the zero headspace extractor. The author also describes alternative waste characterization procedures, including the Synthetic Precipitation Leaching Procedure, batch, sequential, and multiple extraction techniques. Limitations encountered with various waste extract procedures are also discussed.

The final article of this section, entitled **Underground Fuel Spills, Source Identification**, was contributed by Barry K. Lavine et al. The widespread use of underground fuel storage tanks throughout the USA has lead to numerous problems with leakage and subsequent contamination of groundwater and drinking water supplies. The authors describe the pattern recognition process used to categorize spills, and its use for fuel spill identification, incorporating principal component analysis to reduce the number of independent variables. Case studies are provided to demonstrate the use of the pattern recognition techniques to identify the source of jet fuel spills, based on GC analysis of the samples.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AFS	Atomic Fluorescence Spectroscopy
CE	Capillary Electrophoresis
CE/ICP/MS	Capillary Electrophoresis Coupled to Inductively Coupled Plasma-mass Spectrometry
CV	Cold Vapor
EDXRF	Energy Dispersive X-ray Fluorescence
ELISA	Enzyme Linked Immunosorbent Assay
GC	Gas Chromatography

GC/ECD	Gas Chromatography/Electron Capture Detection
GC/FID	Gas Chromatography/Flame Ionization Detector
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ISE	Ion Selective Electrode
LC	Liquid Chromatography
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NAA	Neutron Activation Analysis
NMR	Nuclear Magnetic Resonance
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PIXE	Proton-induced X-ray Emission
QA	Quality Assurance
SFE	Supercritical Fluid Extraction
SPME	Solid Phase Microextraction
TRXRF	Total Reflection X-ray Fluorescence
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
WDXRF	Wavelength Dispersive X-ray Fluorescence

Asbestos Analysis

Norbert Höfert
Commission on Air Pollution Prevention of VDI and DIN, Düsseldorf, Germany

Reiner König
APC Analytische Produktions-, Steuerungs- und Controllgeräte GmbH, Kronberg, Germany

1 Introduction 1
2 Objectives of Measurement 2
3 Determination of Airborne Asbestos Fibers 3
3.1 General 3
3.2 Scanning Electron Microscopy 3
3.3 Transmission Electron Microscopy 6
3.4 Phase Contrast Optical Microscopy 8
4 Determination of Asbestos in Bulk Materials 8
4.1 General 8
4.2 Polarized Light Microscopy 9
4.3 X-ray Diffraction 9
4.4 Analytical Electron Microscopy 10
4.5 Infrared Absorption 12
5 Determination of Asbestos Fibers in Drinking Water 12
6 Quality Assurance 12
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 13
References 13

The comprehensive use of asbestos in approximately 3000 industrial applications in the past requires asbestos analysis at present and in the future. The different analysis methods range from simple and cost-effective procedures in order to decide whether asbestos is present up to the expensive and time-consuming determination of low fiber number concentrations of asbestos fibers including the identification of the fiber type using highly sophisticated measurement equipment.

The following survey discusses the most relevant objectives of asbestos analysis and the measurement techniques that are applied:

- *Ambient/indoor air measurements – low concentration range, fiber type usually not known, identification required – using transmission electron microscopy (TEM) in combination with energy-dispersive X-ray analysis (EDXA) and selected area electron diffraction (SAED), and using scanning electron microscopy (SEM) in combination with EDXA.*
- *Workplace air measurements – low to high concentration range, fiber type usually known, identification possibly required – using SEM, TEM, and phase contrast optical microscopy (PCOM).*
- *Stationary source emission measurements using SEM, TEM, and PCOM.*
- *Bulk material analysis – usually qualitative measurements; semiquantitative or quantitative results possible, if required – using SEM, TEM, infrared (IR) spectroscopy, PCOM, polarized light microscopy (PLM), and X-ray diffraction (XRD).*
- *Water analysis using TEM.*

1 INTRODUCTION

The combination of ideal technical properties (heat resistance, noncombustibility, high tensile strength, flexibility, resistance to chemical agents) has favored the use of asbestos in approximately 3000 industrial applications so that it plays an important role in modern civilization. These applications have been widely used because asbestos is an abundant, relatively inexpensive and nearly indestructible material, having unique properties that made it readily adaptable to a variety of physical forms and applications. Applications of asbestos comprise numerous building materials (for example loose-fill insulation, acoustic and thermal sprays, pipe and boiler wraps, flooring and roofing materials, cementitious products), friction plates, sealings, gaskets, heat-resistant clothing, adhesives, paints, coatings, etc.

Asbestos is the commercial term applied to the asbestiform varieties of six different silicate minerals. The asbestiform habit causes them to be easily separated

Table 1 Asbestos varieties, ideal composition and CAS registry numbers

Name	Ideal composition	CAS number
Chrysotile	$Mg_3Si_2O_5(OH)_4$	12001-29-5
Amosite	$(Fe,Mg)_7Si_8O_{22}(OH)_2$	12172-73-5
Anthophyllite	$(Mg,Fe)_7Si_8O_{22}(OH)_2$	77536-67-5
Crocidolite	$Na_2(Fe,Mg)_3Fe_2Si_8O_{22}(OH)_2$	12001-28-4
Actinolite	$Ca_2(Fe,Mg)_5Si_8O_{22}(OH)_2$	77536-66-4
Tremolite	$Ca_2(Mg,Fe)_5Si_8O_{22}(OH)_2$	77536-68-6

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

into long, thin, flexible, and strong fibers when crushed or processed. The names, ideal composition and Chemical Abstracts Service (CAS) numbers of these varieties are given in Table 1. Chrysotile is a fibrous mineral variety of serpentine. The five other asbestiform varieties belong to the amphibole group. The most important asbestos type is chrysotile, with about 94% of the worldwide production. Crocidolite has a share of 3–4% and amosite approximately 2% of the worldwide asbestos production. The other types play no significant role.

Industrial manufacturing of asbestos commenced in the middle of the nineteenth century. It was much later when medical research revealed the casual relationship between the inhalation of asbestos fibers and several severe diseases. Inhaled asbestos fibers can affect human health by causing conditions such as asbestosis and the growth of malignant tumors such as mesothelioma or lung cancer. Asbestos is now considered to be a costly mineral when all subsequent costs for its removal and occupational disease-related expenditure are taken into account. The health risks of asbestos fibers have highlighted the need for efficient measurement techniques.

Medical evidence indicates that the numerical fiber concentration, the fiber size, and the fiber type (biopersistence) are the relevant parameters for the evaluation of inhalation hazards. Fibers longer than 5 μm, narrower than 3 μm and with an aspect ratio greater than 3 : 1 are considered to be critical. A fiber counting and measuring technique is the logical approach, i.e. using microscopy methods. If positive identification of the fiber type is required, the only possible method is electron microscopy in combination with other identification systems, in order to discriminate between asbestos and nonasbestos fibers (for example naturally occurring or man-made mineral fibers (MMMFs)).

A variety of methods, using technical equipment of different levels of complexity and investment (operating) costs, have been developed to satisfy the recent requirements of science and legislation:

- comparison with exposure limits;
- monitoring airborne fiber concentrations (ambient air, indoor air, stationary source emissions);
- final clearance monitoring in connection with asbestos abatement in buildings;
- bulk material analysis.

In order to provide comparability between measurements, standard methods are required. This report only deals with methods that have undergone or will undergo future standardization by a national or international standards organisation (e.g. International Organization for Standardization (ISO)) or other institutions with regulation purposes (e.g. World Health Organization (WHO), United States Environmental Protection Agency (USEPA), Verein Deutscher Ingenieure (VDI)). Specific applications, as for example the determination of asbestos fibers in lung tissue, are not dealt with. Table 2 gives a survey of the proven and established methods.

2 OBJECTIVES OF MEASUREMENT

The following objectives are relevant for asbestos analysis:

Table 2 Proven and established methods for asbestos analysis

Method	Fiber specific	Method of identification	Limitations	Detection limit achieved
Microscopy methods				
PCOM	Yes	–	Visibility for fiber width ≥ 0.2 μm	About ≥10 000 fibers per cubic meter
PLM	Yes	Refractive index	Identification only for fiber widths ≥ 1 μm	About ≥10 000 fibers per cubic meter
SEM	Yes	Elemental composition by EDXA	Identification only for fiber widths ≥ 0.2 μm	≥300 fibers per cubic meter[a]
TEM	Yes	EDXA and crystallographic structure by SAED (electron diffraction, ED)	Theoretically no limits	For structures of length >0.5 μm about ≥2500 fibers per cubic meter[a] at magnification of 20 000 and indirect transfer
Other methods				
XRD	No	XRD	Only for bulk materials	≥1% in weight
IR	No	IR absorption	Mainly for bulk material samples, also possible for emissions from stationary sources	≥1% in weight

[a] The detection limit can be varied when varying the analysis conditions (e.g. the evaluated filter area).

1. Concentration (fibers/per cubic meter) of airborne asbestos fibers:
   (a) assessment of concentrations of airborne fibers in workplace atmosphere (comparison with occupational exposure limits; most commonly personal exposures);
   (b) assessment of concentrations of airborne fibers in ambient air (background concentrations);
   (c) assessment of concentrations of airborne fibers in emissions from stationary sources (assessing the effectiveness of process-control techniques and monitoring the effects of process modifications);
   (d) assessment of concentrations of airborne fibers in indoor air – sources of contamination, status quo measurement, assessment of the progress of abatement measures, final clearance monitoring;
   (e) epidemiology.
2. Mass concentration (milligrams per cubic meter) of asbestos fibers:
   (a) assessment of concentrations of airborne fibers in workplace atmospheres (most commonly personal exposures);
   (b) assessment of concentrations of airborne fibers in emissions from stationary sources.
3. Indicative or qualitative determination of asbestos in bulk materials.

Today the greater part of measurements for asbestos analysis concerns indoor air measurements, workplace measurements, and bulk sample analysis. Ambient air measurements no longer play a significant role in countries where the use of asbestos has been drastically restricted by law. The situation may be different in countries with lower restriction levels.

3 DETERMINATION OF AIRBORNE ASBESTOS FIBERS

3.1 General

Measurement tasks that can be solved by a pure fiber count commonly use PCOM. Such situations are frequently observed in workplace atmospheres of asbestos processing industries, where most fibers are asbestos.

In cases where identification of fiber type is needed the PLM technique can be used so long as the fibers are wider than about 1 μm.

Electron microscopy techniques (SEM, TEM) are more expensive but they provide additional information on fiber type. SEM, in conjunction with EDXA, can generally be used to determine the elemental composition of fibers wider than about 0.1 μm. The most expensive method, TEM, is generally acknowledged as the most accurate technique for characterizing crystalline inorganic fibers and can be used to provide chemical and structural information for fibers down to about 0.01 μm in width. The results of the different microscopical methods, which provide fiber concentrations, are not directly comparable, because the methods differ in sample preparation and evaluation, and partially in fiber counting rules.

Each measurement technique used gives an index for the numerical concentration of a certain fiber type. Some techniques may give lower counts, resulting from lower visibility caused by lower magnification or resolution; some may give higher counts due to false positive identification connected with the limits of the identification method used. Some problems related to the use of automatic image analysis systems for fiber counting purposes have yet to be resolved, so that this modern technique cannot yet replace the microscope operator.

The technical specifications and a detailed description of sampling, sample preparation, and sample evaluation are beyond the scope of this article and are given in the standards or other technical rules which are referenced.

3.2 Scanning Electron Microscopy

SEM, in combination with EDXA, enables the detection of fibers and the establishment of their elemental composition. Thus fibers may be classified, usually as chrysotile, amphibole asbestos, other inorganic fibers, or calcium sulfate fibers. This method is used to determine asbestos in air samples (ambient, indoor, and workplace air; section 3.2.1) and in bulk material samples (section 4.4).[1–4] Figure 1 shows a modern SEM set-up.

Figure 1 Modern SEM equipment. (Provided by R. König.)

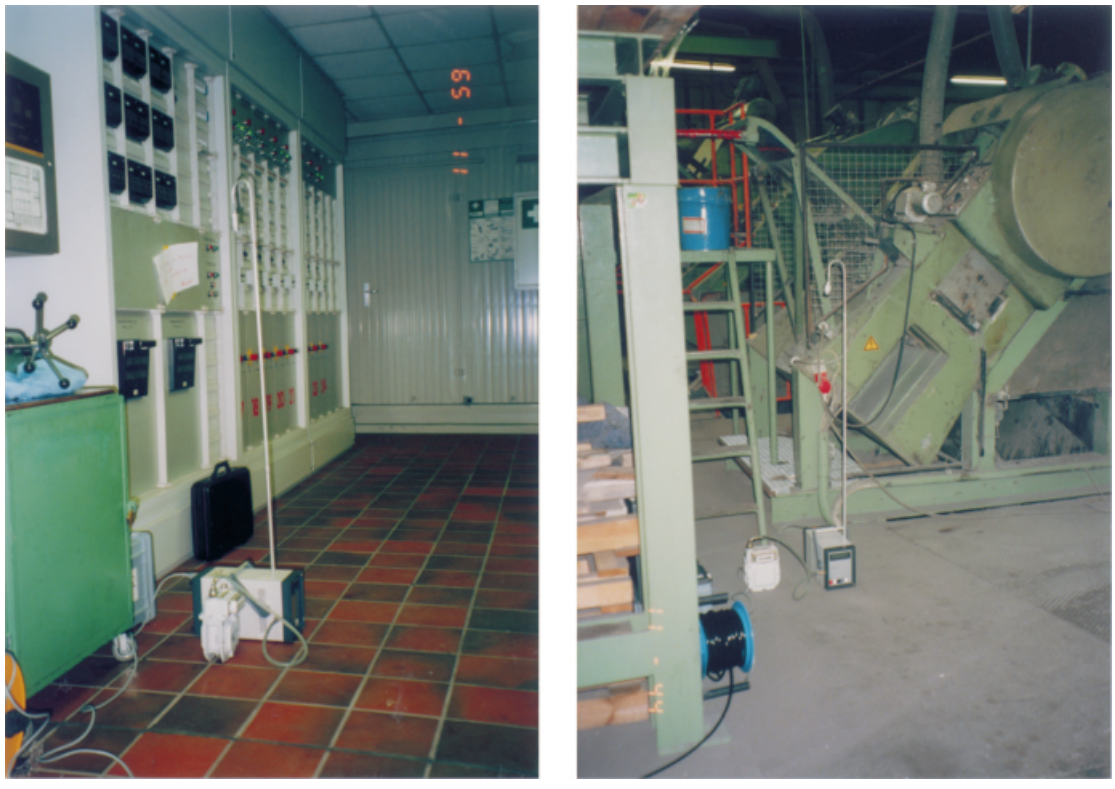

Figure 2 Sampling apparatus with sampling head in two different working environments. (Provided by U. Teichert.)

3.2.1 Principle

A sample of airborne particulate is collected by drawing a measured volume of air through a gold-coated capillary-pore membrane filter (polycarbonate filter, diameter usually 25 mm) with a nominal pore size of 0.8 µm. Figure 2 shows an appropriate sampling apparatus in two different working environments.

The sampled volume depends on the objective of measurement. Usually it is 2 L min^{-1} per square centimeter of filter area over a period of 8 h. The dust-loaded filter is treated in a plasma asher (oxygen plasma) to remove organic particles that could interfere with the detection and analysis of the asbestos fibers.

Figure 3 shows a plasma asher, whereas Figure 4 shows a filter loading before and after plasma ashing treatment of a sample of airborne dust.

The filter is then examined in the SEM. The individual fibrous particles and constituent fibers in a randomly selected area of the filter (the total area of the randomly distributed image fields is 1 mm^2) are then counted at a defined magnification (for example 2000×) according to defined counting rules. The fiber is then examined at a higher magnification (for example 10 000×) to measure its dimensions and to determine its elemental composition using EDXA. The fiber data are recorded and the fiber classified accordingly.

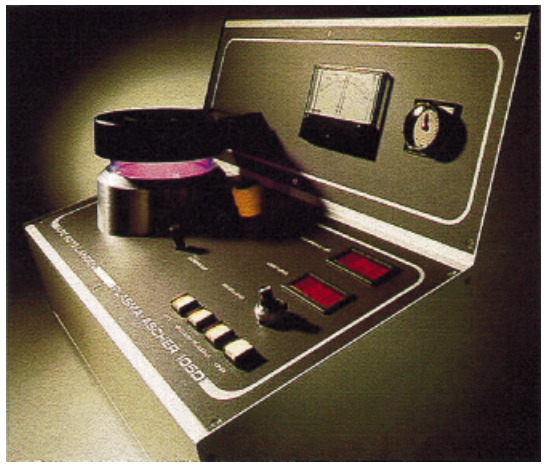

Figure 3 Plasma asher.

Figures 5 and 6 show some examples of asbestos fibers on the filter surface, indicating the geometric dimensions and the corresponding EDXA spectra.

3.2.2 Characteristics

The range of concentrations for fibers greater than 5 µm in length and within the range of diameters that can be detected and measured under standard

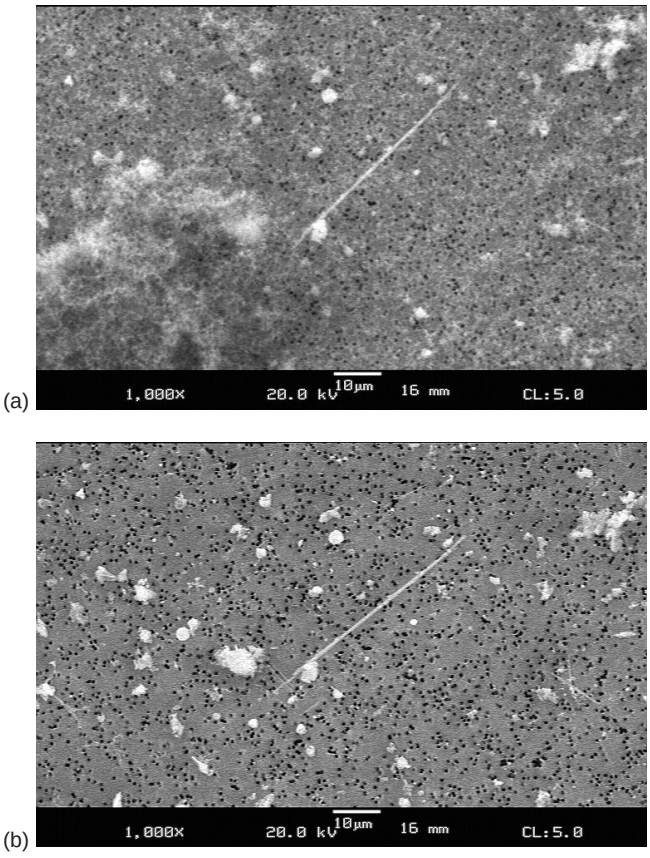

Figure 4 Dust sample (a) before and (b) after plasma ashing treatment.

measurement conditions, is approximately 3–200 fibers per square millimeter of filter area. The air concentrations represented by these values are a function of the volume of air sampled.

The ability of the method to detect and classify fibers less than 0.2 µm in diameter is limited. If airborne fibers are predominantly less than 0.2 µm in diameter, a TEM method, such as ISO 10312,*(5) can be used to determine the smaller fibers.

The detection limit is defined as the fiber concentration below which, with 95% confidence (Poisson distribution), no fibers are found during the SEM examination. The detection limit can be lowered by filtration of progressively larger volumes of air and by examination of a larger area of the specimen in the SEM. In practice, the lowest achievable detection limit for a particular area of SEM specimen

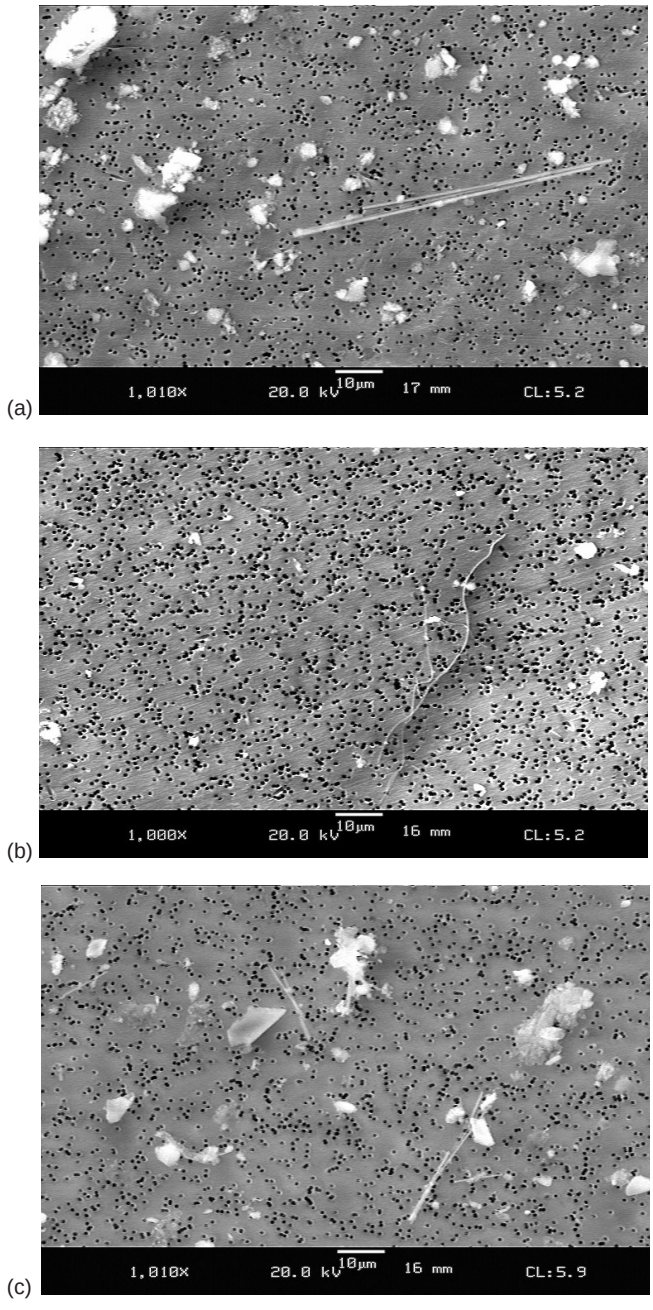

Figure 5 Air samples drawn through gold-coated polycarbonate filters. The photographs show SEM images of very thin (a) crocidolite fibers, (b) chrysotile fibers (bent) and crocidolite fibers (straight), and (c) amosite fibers.

examined is controlled by the total suspended particulate concentration remaining after the plasma ashing step.

A detection limit of approximately 300 fibers per cubic meter is obtained if an air volume of 1 m^3 per square centimeter of filter surface area passes through the filter, and an area of 1 mm^2 of the filter area is examined

* Extracts from Clause 4 taken from ISO 10312: 1995, Clause 1 from ISO 13794 and Clause 2.5 from ISO/DIS 14966 have been reproduced with the permission of the ISO. These standards can be obtained from DIN Deutsches Institut für Normung, Burggrafenstrasse 6, D-10787 Berlin or directly from the Central Secretariat, ISO, Case postal 56, 1211 Geneva 20, Switzerland. Copyright remains with ISO.

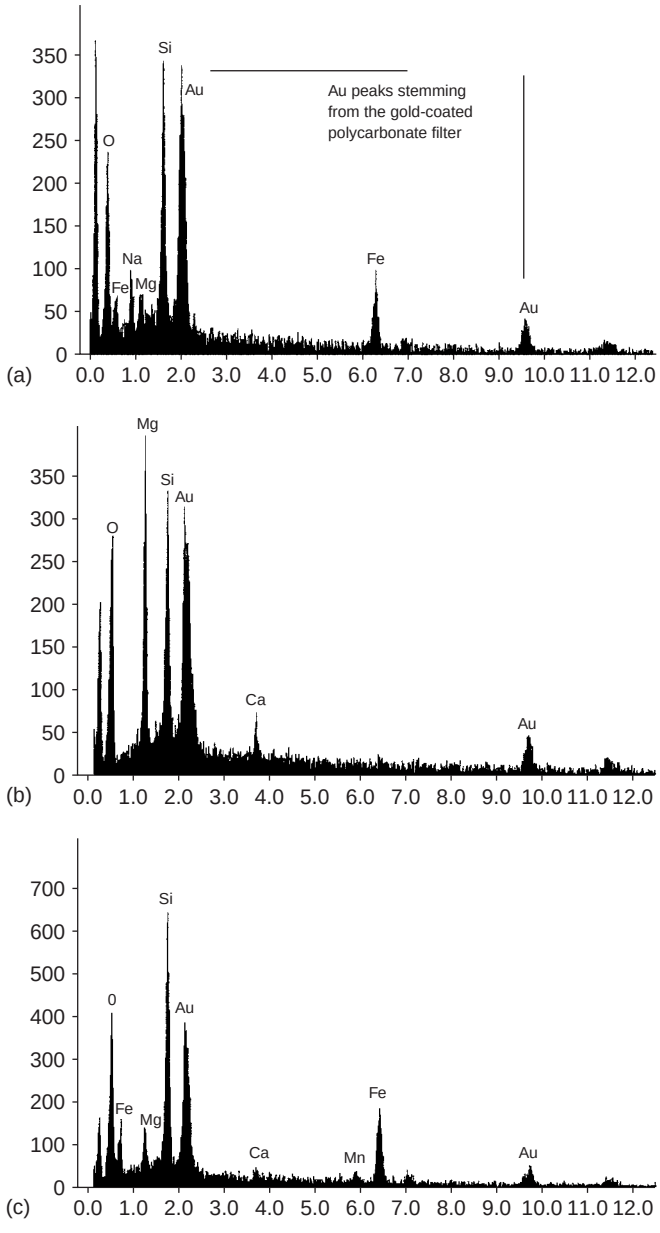

Figure 6 EDXA spectra of (a) crocidolite, (b) chrysotile, and (c) amosite.

in the SEM. This corresponds to an evaluated sample air volume of 0.01 m^3. Under these conditions a single detected fiber on the filter corresponds to 100 fibers per cubic meter.

This method had been developed and optimized as being cost-effective, economical to perform, and satisfying the needs of epidemiological research. The major goal concerning air quality characteristics was to provide a measurable result that reflects the concentration of airborne structures with the dimensions of the critical fiber (section 1).

3.3 Transmission Electron Microscopy

The TEM technique[5,6] is applicable to the determination of airborne asbestos in ambient and indoor air, and for detailed evaluation of any atmosphere in which asbestos structures are likely to be present.

TEM is currently the only technique capable of unequivocal identification of the majority of individual fibers of asbestos. However, it requires expensive and time-consuming measurement effort. Subjective and instrumental factors contribute to this measurement, and consequently a very precise definition is required of the procedure for identification and enumeration of asbestos. The sample preparation technique depends on the objective of measurement, i.e. whether the intention is to determine the concentration of airborne asbestos-containing structures or the asbestos fiber potential by disentangling the structures.

Airborne asbestos-containing structures comprise single fibers, fiber bundles, and even very complex and aggregated structures which may also be aggregated with other particles. The number of asbestos fibers and bundles incorporated in these complex structures can exceed the number of isolated fibers and bundles observed, and many of them may be completely obscured in *direct-transfer* sample preparations.[5] With the *indirect-transfer* preparation technique[6] organic and water-soluble materials are removed and the complex structures are dispersed via a liquid phase into their individual component fibers and bundles, hence giving a higher fiber count than direct transfer methods. However, the sample preparation effort is increased.

3.3.1 Direct-transfer Method

A sample of airborne particulate is collected by drawing a measured volume of air through either a capillary-pore polycarbonate membrane filter of maximum pore size of 0.4 μm or a cellulose-ester (either mixed esters of cellulose or cellulose nitrate) membrane filter of maximum pore size 0.45 μm by means of a pump. The choice of filter type determines the sampling pump required, as capillary-pore polycarbonate membrane filters produce a higher pressure drop than do cellulose-ester membrane filters.

TEM specimens are prepared from polycarbonate filters by applying a thin film of carbon to the filter surface by vacuum evaporation. Small areas are cut from the carbon-coated filter, supported on TEM specimen grids, and the filter medium is eliminated by a solvent extraction procedure (usually a Jaffe washer using chloroform (CAS 67-66-3) or *N*-methyl-2-pyrrolidone (CAS 872-50-4) as the solvent). This procedure leaves a thin film of carbon which bridges the openings in the TEM specimen grid, and which supports each particle from the original filter in its original position.

Cellulose ester filters are chemically treated (using a mixture of dimethylformamide (CAS 68-12-2), glacial acetic acid (CAS 64-19-7) and water) to collapse the pore structure of the filter. The surface of the collapsed filter is then etched in an oxygen plasma to ensure that all particles are exposed. A thin film of carbon is evaporated on to the filter surface and small areas are cut from the filter. These sections are supported on TEM specimen grids and the filter medium is dissolved away by a solvent extraction procedure (usually a Jaffe washer using dimethylformamide or acetone (CAS 67-64-1) as solvent).

The TEM specimen grids undergo a quantitative structure count on randomly selected grid openings (magnification 20 000×; lower magnifications can be used when larger objects, such as fibers longer than 5 μm, are to be determined). In the TEM analysis, ED is used to examine the crystal structure of a fiber, and its elemental composition is determined by EDXA. The analysis results are used to classify the fiber. The fiber classification procedure is based on successive inspection of the morphology, the ED pattern for a selected area, and the qualitative and quantitative EDXA. The identification of chrysotile is confirmed only by quantitative ED, and amphibole is confirmed only by quantitative EDXA and quantitative zone axis ED.

The lengths and widths of all classified structures are recorded. The number of asbestos structures found on a known area of the microscope sample, together with the equivalent volume of air filtered through this area, is used to calculate the airborne concentration in asbestos structures per liter of air.

3.3.2 Indirect-transfer Method

A sample of airborne particulate is collected by drawing a measured volume of air through either a capillary-pore polycarbonate membrane filter of maximum pore size 0.4 μm or a cellulose-ester (either mixed esters of cellulose or cellulose nitrate) membrane filter of maximum pore size 0.8 μm by means of a pump. The choice of the filter type determines the sampling pump required as capillary-pore polycarbonate membrane filters produce a higher pressure drop than do cellulose-ester membrane filters.

A portion of the filter is ashed in an oxygen plasma, and the residual ash is dispersed in distilled water with pH adjustment to between 3 and 4 using acetic acid. Analytical filters are then prepared by filtration of known volumes of this aqueous dispersion through either capillary-pore polycarbonate membrane filters of maximum pore size 0.2 μm or cellulose-ester membrane filters of maximum pore size 0.22 μm.

TEM specimens are prepared from polycarbonate analytical filters by applying a thin film of carbon to the filter surface by vacuum evaporation. Small areas are cut from the carbon-coated filter, supported on TEM specimen grids, and the filter medium is dissolved by a extraction procedure with an organic solvent (usually a Jaffe washer using chloroform or *N*-methyl-2-pyrrolidone as solvent). This procedure leaves a thin film of carbon which bridges the openings in the TEM specimen grid, and which supports each particle from the original filter in its original position.

Cellulose ester filters are chemically treated (using a mixture of dimethylformamide, glacial acetic acid, and water) to collapse the pore structure of the filter, and the surface of the collapsed filter is then etched in an oxygen plasma to ensure that all particles are exposed. A thin film of carbon is evaporated on to the filter surface and small areas are cut from the filter. These sections are supported on TEM specimen grids and the filter medium is dissolved away by a solvent extraction procedure (usually a Jaffe washer using dimethylformamide or acetone as solvent).

A quantitative fiber count on randomly selected grid openings is carried out at a magnification of approximately 20 000× (or less, depending on the fiber size/diameter spectrum to be analyzed). In TEM analysis, ED is used to examine the crystal structure of a fiber, and its elemental composition is determined by EDXA. The results are used to classify the fibers. The fiber classification procedure is based on successive inspection of the morphology, the selected area ED pattern, and qualitative and quantitative EDXA.

The lengths and widths of all classified structures are recorded. The number of asbestos structures found on a known area of the microscope sample, together with the volume of air filtered through the sample collection filter that was ashed, the proportion of the aqueous dispersion that was filtered, and the area of the analytical filter are used to calculate the airborne concentration of asbestos in asbestos structures per liter of air.

The examination consists of a count of the asbestos structures present on a specified number of grid openings. Fibers are classified into groups on the basis of morphological observations, ED patterns, and EDXA spectra. The total number of structures to be counted depends on the statistical precision desired. In the absence of asbestos structures, the area of the TEM specimen grids that must be examined depends on the analytical sensitivity required.

3.3.3 Characteristics

The upper limit for the range of concentrations that can be measured on the analytical filter is 7000 structures per square millimeter. The air concentrations represented by this value are a function of the volume of air sampled and the degree of dilution or concentration selected during the specimen preparation procedures. A minimum length of

0.5 μm has been proved to be a reasonable lower limit to the dimensions of asbestos fibers that should be recorded.

In theory the detection limit can be lowered indefinitely by filtration of progressively larger volumes of air, concentrating the sample during specimen preparation (indirect-transfer method), and by extending the area to be examined. In practice, the lowest achievable detection limit for a particular area of TEM specimen examined is controlled by the total suspended particulate concentration remaining after ashing and aqueous dispersal steps. In order to achieve lower detection limits for critical fibers (section 1), lower magnifications permit more rapid examinations of larger areas of the TEM specimens when the examination is limited to these dimensions of fiber.

3.4 Phase Contrast Optical Microscopy

The PCOM method,[7–9] despite its limitations concerning visibility and identification, is well-proven for widespread international use in the field of occupational hygiene and measurement of asbestos emissions from stationary sources.

*3.4.1 Principle**

A sample is collected by drawing a known volume of air through a membrane filter (membrane of mixed esters or cellulose nitrate, 0.8–1.2 μm pore size) by means of a pump. The filter is exposed to acetone vapor, which condenses on the filter, collapsing its pores and making it transparent. Different methods are described for this procedure.[7–9] The filter is then fixed to a glass slide, where it appears as a transparent and uniform plastic film with any fibers on the upper surface. A liquid must be added to provide optimal contrast – if the refractive index of the fibers exceeds 1.51 then triacetin (glycerol triacetate, CAS 102-76-1) is satisfactory; for inorganic fibers with refractive indices less than 1.51 the filter surface must be etched to expose the fibers using a plasma asher. Distilled water is then used as the contrast liquid. Fibers on a measured area of the filter are counted visually using PCOM with a magnification of 400–600, and the concentration of fibers in the volume of air sampled is calculated.

*3.4.2 Characteristics**

This method is usually used to measure the fiber concentration of airborne structures present with the critical dimensions length greater than 5 μm, diameter less than 3 μm, and aspect ratio greater than 3:1. The minimum visible width depends on the resolving power of the optical system, the difference in refractive index between the fiber and the contrast liquid and – as for all fiber counting methods – the visual acuity of the microscopist. Thus the limit of visibility can be as low as about 0.15 μm, but is usually higher in practice. The detection limit depends on the extent of sampling and evaluation, and especially the blank value of the analytical filters. The practical limit is 10 000 fibers per square meter or higher.

The method does not permit the determination of the chemical composition or crystallographic structure of fibers, and therefore cannot be used to discriminate between different fiber types. Supplementary information on fiber type may be obtained by using other methods, such as PLM, SEM, or TEM.

4 DETERMINATION OF ASBESTOS IN BULK MATERIALS

4.1 General

The comprehensive use of asbestos in the past has produced the present-day need to identify and possibly quantify asbestos in a variety of asbestos-containing industrial products. These include building materials such as loose-fill insulation, pipe and boiler wraps, plasters, paints, flooring products, roofing materials, and cementitious products. The diversity of these bulk materials necessitates the use of several different methods of sample preparation and analysis.[10] Figure 7 shows SEM images of different asbestos-containing products.

A precondition for a reliable analytical result is a representative sample of the materials of interest.[11] Many manufactured products containing asbestos, such as boards, sheets, cement pipes, textiles, ropes, friction products, plastics, and vinyls can be assumed to have the asbestos uniformly distributed throughout the material. Thus, a relatively small sample can be taken as being representative of the whole. Other products may be layered and the asbestos will be present in part of the cross-section. Insulation and spray materials tend to be more variable. Insulation was often wet-mixed on-site and subsequent repairs and patching may add to this variability. Samples should be taken that are representative of the whole material.

The sample may require mechanical and/or chemical treatment in order to release fibers from any matrix and to remove adhering particles prior to fiber analysis (microscopic methods). Any preparation technique must preserve the integrity of morphology, structure, and chemistry of the fibrous particles. IR and XRD require careful homogenization of the sample.

The simplest analytical method is the use of a stereomicroscope which provides preliminary information on the sample and the fiber type. Stereomicroscopic analysis

* Reproduced by permission of WHO.

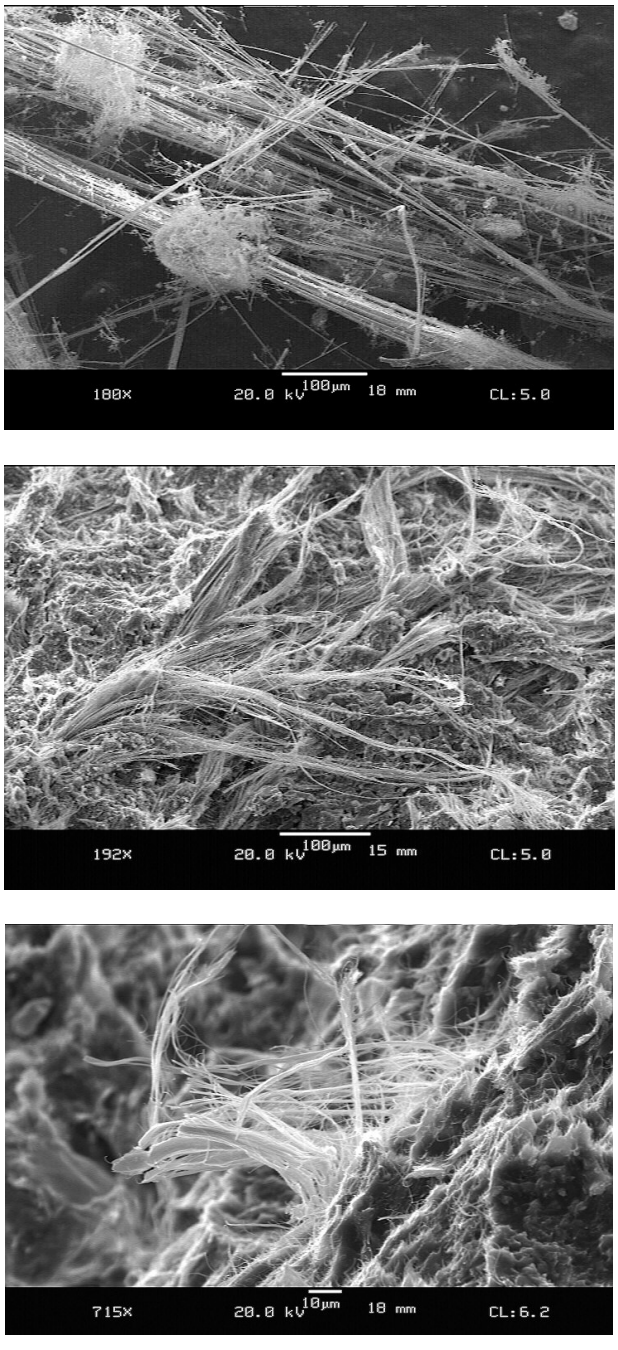

Figure 7 SEM images of different asbestos-containing products: (a) sprayed-on insulation material, (b) asbestos cement, (c) floor covering (vinyl floor tile).

determines the homogeneity, texture, friability, color, and the extent of fibrous components of the sample. This information is then used as a guide to the selection of further, more definitive qualitative and quantitative asbestos analysis methods.

Qualitative and semiquantitative results may be obtained by additional analyses, such as PLM,[10,11] XRD,[10] analytical electron microscopy (AEM),[10] or IR spectroscopy.[12,13]

The accuracy, precision, and detection limits of the different analytical techniques for bulk material analysis are dependent on the type of sample (matrix, components, texture, etc.), the preparation of the sample (homogeneity, grain size, etc.), and the specifics of the method.

4.2 Polarized Light Microscopy

Samples are examined using PLM to determine optical properties of constituents and to provide positive identification of suspect fibers. The major goal of the qualitative preparation is to mount easily visible representative fibers or fiber bundles in appropriate refractive index liquids on microscope slides for complete optical characterization. A tentative identification based on the stereomicroscopy evaluation or other preinformation is used to select the most appropriate mounting liquid. Identification of a single asbestos fiber as one of the six regulated asbestos types requires assessment of the following optical properties using PLM with magnifications from about 80 upwards, as appropriate:

- morphology
- color and pleochroism (if present)
- birefringence
- extinction characteristics
- sign of elongation
- refractive index assessment.

Identification is based on comparing the recorded observations (morphology, optical properties) on the analyzed fibers against the properties of asbestos reference standards. A close match between the optical properties of the sample fiber and the asbestos standard will normally be achieved.

Figures 8–10 show the application of PLM to asbestos identification. The semiquantitative estimation of the asbestos content in a sample is possible.[10]

4.3 X-ray Diffraction

In powder XRD analysis, a bulk sample of material is subjected to X-ray radiation and the angle of the diffracted radiation is measured. The solid powdered crystalline sample will diffract the incident beam of parallel, monochromatic X-rays whenever Bragg's law is satisfied for a particular set of planes in the crystal lattice. A diffraction pattern is generated which is a characteristic of the structure of the crystalline phases present.

This technique allows determination of the crystal structure of mineral compounds. However, XRD cannot determine crystal morphology. Therefore in asbestos

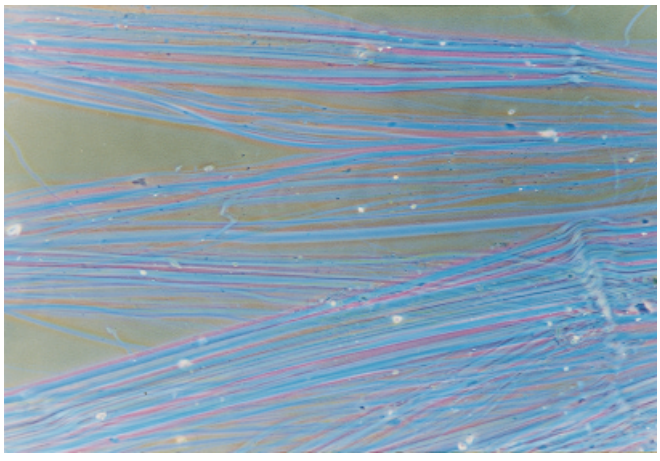

Figure 8 Chrysotile fiber bundles in cinnamic acid ethyl ester; positive phase contrast; fibers parallel to polarizer are blue with orange-red halo.

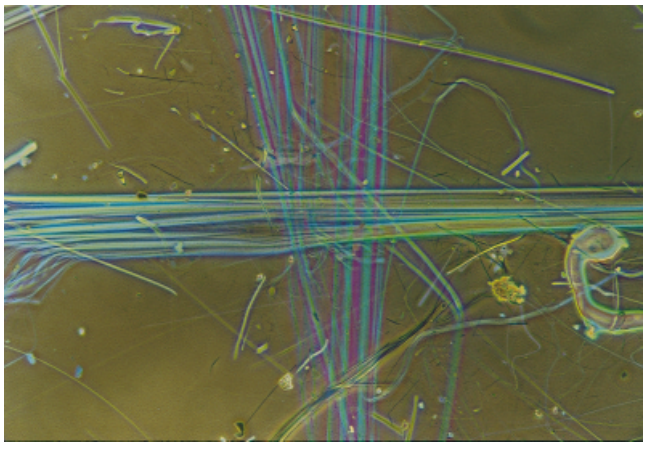

Figure 9 Chrysotile fiber bundles and MMMF (thick bent object) in cinnamic acid ethyl ester; positive phase contrast; fibers parallel to polarizer are blue with orange-red halo; the MMMF shows no optical effects.

analysis, it is not possible to distinguish between fibrous and nonfibrous forms of serpentine and amphibole minerals using XRD. XRD should be used in conjunction with other methods, such as PLM or electron microscopy, and thus can provide a reliable analytical method for the identification (diffraction pattern), characterization, and quantitation (signal intensities) of asbestiform minerals in bulk materials. In order to identify the minerals the diffraction patterns should be compared with powder diffraction patterns of well-characterized reference materials.

Quantitative results require calibration. Accurate quantitative analysis of asbestos in bulk samples by XRD is critically dependent on particle size distribution, crystallite size, preferred orientation and

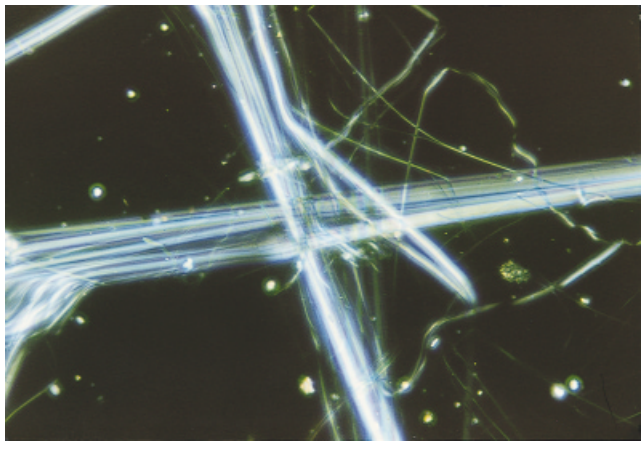

(a)

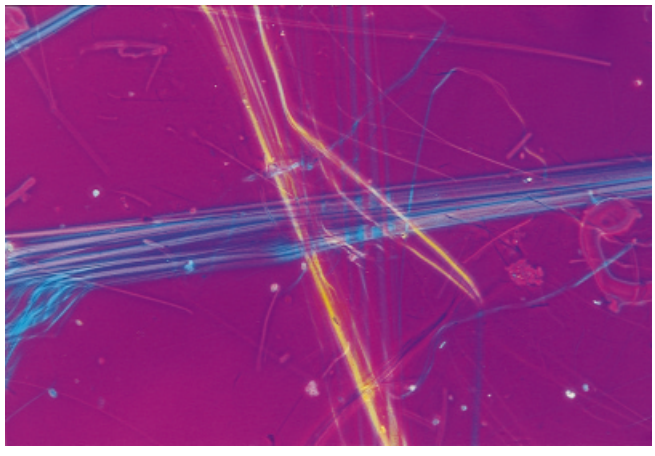

(b)

Figure 10 Birefringence: (a) chrysotile fibers are clearly visible between crossed polars, whereas the isotropic MMMF is not visible; (b) with the first-order red compensator the chrysotile fibers appear blue or yellow, depending on their orientation with respect to the polarizer, and the MMMF is visible.

matrix absorption effects, and comparability of standard reference and sample materials.[10]

Figure 11 shows the XRD spectra of the six different asbestos types. The samples were prepared for routine qualitative analysis.

4.4 Analytical Electron Microscopy

Both SEM[14] and TEM[10] can be used to examine bulk material samples, the latter being more expensive and time-consuming.

The preliminary visual inspection and examination of the material using the stereomicroscope served to identify representative or relevant portions which are prepared as samples for SEM. SEM allows a direct investigation of the surface of the sample, even if the surface is highly structured. The sample is mounted on

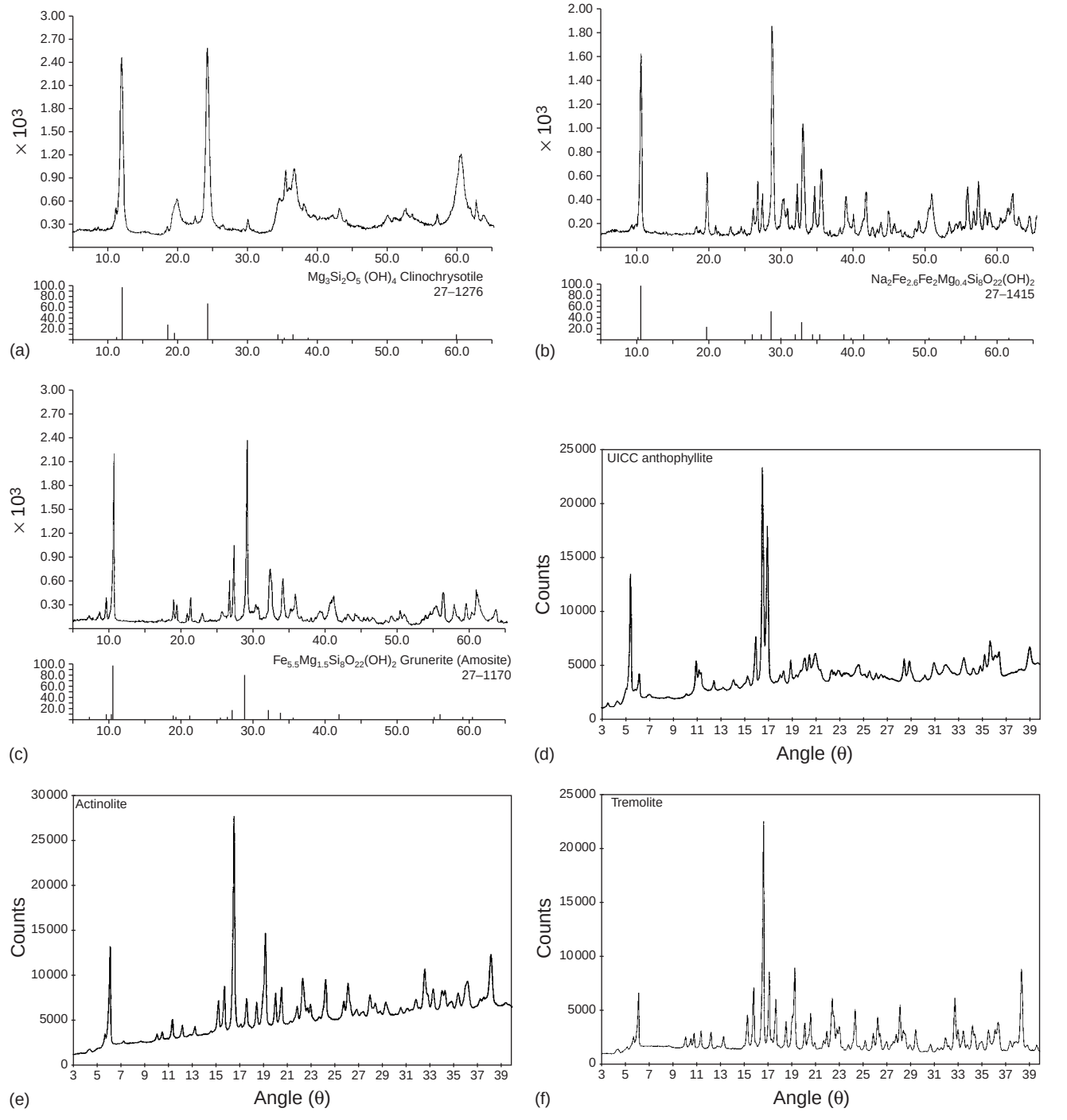

Figure 11 XRD spectra of the six asbestos types: (a) chrysotile 7 M 4 (Canada), (b) crocidolite KCBR (South Africa), (c) amosite S 11 (South Africa), (d) anthophyllite (International Union Against Cancer (UICC) reference standard), (e) actinolite, and (f) tremolite (Institute of Occupational Medicine (IOM) reference standard).

the sample holder or sputtered with carbon or gold. The sample is examined with respect to asbestos fibers initially with low (20–50), then with high (500–2000) magnification. Negative results have to be verified with a magnification of 5000–10 000. Detected fibers are assessed morphologically, and the elemental composition is determined using EDXA. Based on these analytical results the fibers are classified or even identified. SEM is very useful for observing surface features in complex particles, and for determining elemental compositions.

The application of TEM to bulk material analysis has been described in detail.[10] The sample material is treated by ashing or dissolution or, when this is not possible, at least by grinding or milling. The material is then dispersed in an organic solvent. A small volume of this suspension is transferred to the top of a carbon-coated TEM grid, the solvent is evaporated, and the grid examined for qualitative and semiquantitative analysis. Quantitative evaluation requires some additional procedural steps. A known weight of the bulk sample, after preparation, is dispersed in distilled water, filtered (see section 5) and TEM grids are prepared from appropriately loaded filters using the standard methods.

4.5 Infrared Absorption

This technique indicates the possible presence of certain functional groups in the components analyzed but cannot distinguish between fibers and other material with the same chemical properties. The resonant absorption of IR radiation is measured for peaks indicative of various functional groups.

Chrysotile shows a characteristic double band in the near IR at about 3660 and 3700 cm^{-1} (−OH stretching vibrations). The intensity of the latter is used for quantitation.

The IR spectra of the amphibole asbestos types are quite similar. A group of characteristic bands is found in the mid- and far IR in the wavenumber range between 300 and 1200 cm^{-1}. The absorption band at 780 cm^{-1} has proved best for the quantitation of crocidolite and amosite. Figure 12 shows the IR spectra of chrysotile, crocidolite, and amosite. The IR method is proven only for these three asbestos types.

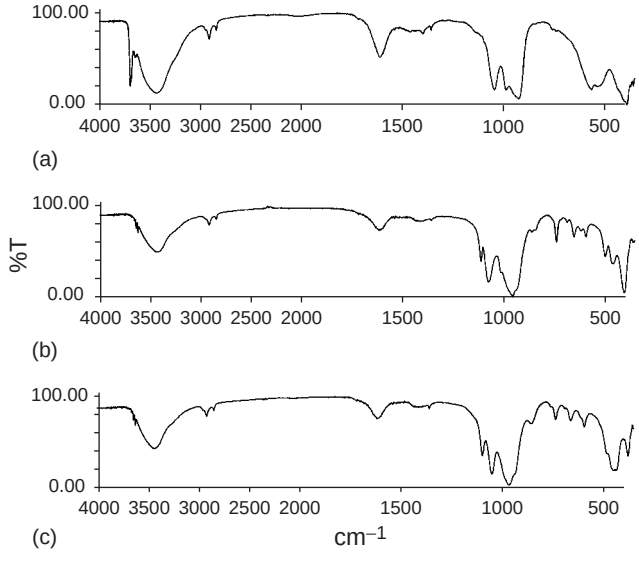

Figure 12 IR spectra of (a) chrysotile, (b) crocidolite, and (c) amosite.

Solid samples are prepared as follows. The powdered sample is homogeneously mixed with potassium bromide (CAS 7758-02-3) and pressed to form a transparent pellet. The IR spectrum of the sample is recorded in the range between about 250 and 4000 cm^{-1}.

As IR spectra do not contain any information about morphological structures, it is useful to verify the presence of fibrous structures using a light microscope, if IR indicates the presence of asbestos – especially amphibole – so as to avoid false positive identification as far as possible.

5 DETERMINATION OF ASBESTOS FIBERS IN DRINKING WATER

Asbestos cement pipes, used within drinking water supply networks, also need to be considered critically. Fiber release from this type of pipe depends on the condition of the material matrix and surface and also on the characteristics of the water, especially its saturation index. When the water is around saturation with respect to calcium carbonate, the fiber release tends to a minimum.

Due to the variety of fibrous particles in water, a method is required that allows identification of the fiber type and the visibility of fibrils at least 0.5 μm in length. Thus TEM, including EDXA and SAED, is the preferred method[15] (section 3.3). SAED allows the identification of particles of organic origin that cannot be ashed, such as specific parts of diatoms or excretions of iron-accumulating microorganisms. The recommended characteristics are:

- membrane filter pore size 0.2 μm;
- magnification 10 000;
- analytical sensitivity 3300 fibers per liter (corresponding to one detected fiber);
- detection limit 10 000 fibers per liter.

Health risks as a result of oral uptake of asbestos fibers from drinking water are not evident so far. Consequently WHO does not recommend a limiting value for asbestos fibers in drinking water.

6 QUALITY ASSURANCE

Usual and adequate intralaboratory quality assurance measures with respect to qualified personnel, laboratory equipment, reference material, etc. can be based on the standard EN 45001 or the draft standard ISO/IEC

DIS 17025.[16,17] In addition, an essential part of quality assurance for fiber-counting methods is the performance of counting checks because of the large differences in results obtained within and between laboratories using microscopical counting. In practice, the effects of chance superimposition on counting results are small compared with effects due to differences between individual microscopists. Laboratories using fiber-counting methods should participate in a national or international proficiency testing scheme in order to minimize interlaboratory variation. Interlaboratory exchange and verification of samples can further complement internal quality control.

The purpose of the quality assurance program is to minimize failures in the analysis of materials prior to submitting the results to the client. Failures in the analysis of asbestos materials include false positives, false negatives, and misidentification of asbestos types. False positives result from identification or quantitation errors. False negatives result from identification, detection, or quantitation errors.

ACKNOWLEDGMENTS

I am grateful to the following people for providing photographs and spectra: Friedrich Jaekel, Niedersächsisches Landesamt für Ökologie, Hannover, Germany; Jochen Krause, Landesamt für Arbeitsschutz des Landes Sachsen-Anhalt, Dessau, Germany; Markus Mattenklott, BIA Berufsgenossenschaftliches Institut für Arbeitssicherheit, Sankt Augustin, Germany; and Ullrich Teichert, GSA Gesellschaft für Schadstoffmessung und Auftragsanalytik GmbH, Neuss, Germany.

ABBREVIATIONS AND ACRONYMS

AEM	Analytical Electron Microscopy
CAS	Chemical Abstracts Service
ED	Electron Diffraction
EDXA	Energy-dispersive X-ray Analysis
IOM	Institute of Occupational Medicine
IR	Infrared
ISO	International Organization for Standardization
MMMF	Man-made Mineral Fiber
PCOM	Phase Contrast Optical Microscopy
PLM	Polarized Light Microscopy
SAED	Selected Area Electron Diffraction
SEM	Scanning Electron Microscopy
TEM	Transmission Electron Microscopy
UICC	International Union Against Cancer
USEPA	United States Environmental Protection Agency
VDI	Verein Deutscher Ingenieure
WHO	World Health Organization
XRD	X-ray Diffraction

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air • Carcinogens, Monitoring of Indoor Air

Particle Size Analysis **(Volume 6)**
Optical Particle Counting

Infrared Spectroscopy **(Volume 12)**
Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy • Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

X-ray Spectrometry **(Volume 15)**
Structure Determination, X-ray Diffraction for

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. International Organization for Standardization, *ISO/DIS 14966 Ambient Air – Determination of Numerical Concentration of Inorganic Fibrous Particles – Scanning Electron Microscopy Method*, ISO, Geneva, 1997.
2. *VDI 3492 Blatt 2 (Part 2) Messen von Innenraumluftverunreinigungen; Messen Anorganischer Faserförmiger Partikel; Meßplanung und Durchführung der Messung; Rasterelektronenmikroskopisches Verfahren* (Indoor Air Pollution Measurement; Measurement of Inorganic Fibrous Particles; Measurement Planning and Procedure; Scanning Electron Microscopy Method), Beuth Verlag, Berlin, 1994.
3. *ZH 1/120.46 Verfahren zur Getrennnten Bestimmung von Lungengängigen Asbestfasern und Anderen Anorganischen Fasern – Rasterelektronenmikroskopisches Verfahren* (Method for the Determination of Inhalable Asbestos Fibers and Other Inorganic Fibers – Scanning Electron

Microscopy Method), Carl Heymanns Verlag KG, Köln, 1991.
4. Asbestos International Association, *RTM 2 Method for the Determination of Airborne Asbestos Fibers and Other Inorganic Fibers by Scanning Electron Microscopy*, AIA Health and Safety Publication, Recommended Method No. 2, AI Association, London, 1984.
5. International Organization for Standardization, *ISO 10312 Ambient Air – Determination of Asbestos Fibers – Direct-transfer Transmission Electron Microscopy Method*, ISO, Geneva, 1995.
6. International Organization for Standardization, *ISO 13794 Ambient Air – Determination of Asbestos Fibers – Indirect-transfer Transmission Electron Microscopy Method*, ISO, Geneva, 1999.
7. International Organization for Standardization, *ISO 8672 Air Quality – Determination of the Number Concentration of Airborne Inorganic Fibers by Phase Contrast Optical Microscopy – Membrane Filter Method*, ISO, Geneva, 1993.
8. World Health Organization, *Determination of Airborne Fiber Number Concentrations: A Recommended Method, by Phase-contrast Optical Microscopy (Membrane Filter Method)*, WHO, Geneva, 1997.
9. International Organization for Standardization, ISO 10397 Stationary Source Emissions – Determination of Asbestos Plant Emissions – Method by Fiber Count Measurement, ISO, Geneva, 1993.
10. United States Environmental Protection Agency, *EPA Test Method: Method for the Determination of Asbestos in Bulk Building Materials*, EPA/600/R-93/116, EPA, Washington, DC, 1993.
11. Health and Safety Executive, *Identification of Asbestos in Bulk Materials by Polarised Light Microscopy (PLM)*, Final Report, Health and Safety Laboratory, Sheffield, UK, January 1996.
12. *VDI 3861 Blatt 1 (Part 1) Messen Faserförmiger Partikeln; Manuelle Asbeststaubmessung im Strömenden Reingas; IR-spektrographische Bestimmung der Asbeststaub-Massenkonzentration* (Measurement of Fibrous Particles; Manual Measurement of Asbestos in Flowing Clean Exhaust Gas; Determination of Asbestos Mass Concentration by IR-Spectroscopy), Beuth Verlag, Berlin, 1989.
13. *ZH 1.120/30 Verfahren zur Bestimmung der Massenanteile von Chrysotilasbest und Amphibolasbesten*, (Method for the Determination of the Mass Content of Chysotile and Amphibole Asbestos), Carl Heymanns Verlag KG, Köln, März 1991.
14. *VDI 3866 Blatt 5 (Part 5) Bestimmung von Asbest in Technischen Produkten; Rasterelektronenmikroskopisches Verfahren* (Determination of Asbestos in Technical Products; Scanning Electron Microscopy Method), Beuth Verlag, Berlin, in preparation.
15. United States Environmental Protection Agency, *EPA Analytical Method for the Determination of Asbestos Fibers in Water*, EPA, Washington, DC, 1982.
16. *EN 45001 General Criteria for the Operation of Testing Laboratories*, European Committee for Standardization, Brussels, 1989.
17. *ISO/IEC 17025 General Requirements for the Competence of Testing and Calibration Laboratories*, International Organization for Standardization, Geneva, 1999.

Atomic Fluorescence in Environmental Analysis

Yong Cai
Florida International University, Miami, USA

1 Theoretical Aspects **1**
1.1 Types of Atomic Fluorescence 1
1.2 Intensity of Fluorescence Radiance 1
1.3 Quenching 2
1.4 Interferences 3
2 Instrumentation **3**
2.1 Excitation Sources 3
2.2 Atomizers 5
2.3 Wavelength Selection and Detection Systems 6
3 Environmental Applications **7**
3.1 Mercury 7
3.2 Hydride-forming Elements 11
3.3 Non-hydride-generation Methods 15
4 Conclusions **16**
Acknowledgments **17**
Abbreviations and Acronyms **17**
Related Articles **17**
References **18**

Atomic fluorescence is a spectroscopic process which is based upon the absorption of radiation of a certain wavelength by an atomic vapor and subsequent radiational deactivation of the excited atoms toward the detection device. Both the absorption and the subsequent atomic emission processes occur at wavelengths which are characteristic of the atomic species present. Atomic fluorescence spectroscopy (AFS) is a very sensitive and selective method for the determination of a number of environmentally and biomedically important elements such as mercury, arsenic, selenium, bismuth, antimony, tellurium, lead, and cadmium. This technique has become one of the most important analytical tools for trace element analysis in environmental samples, such as mercury, owing to its advantages over other methods in terms of linearity and detection levels. Several books and a number of book chapters and review articles have been published dealing with the theory and instrumentation of AFS. This article will provide up-to-date AFS information regarding its application in environmental analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 THEORETICAL ASPECTS

1.1 Types of Atomic Fluorescence

The main types of atomic fluorescence process are illustrated in Figure 1(a–c), which shows diagrammatically the transitions involved. *Resonance fluorescence* occurs when atoms absorb and re-emit radiation of the same wavelength (Figure 1a). Many of the measurements made by analytical chemists employing AFS involve this type of fluorescence. The most analytically useful fluorescence line is the resonance fluorescence corresponding to the transition between an electronic excited state and the ground state of the atom.[1] The wavelengths of the absorption (λ_A) and fluorescence (λ_F) can also be different. If the photon energy of fluorescence is less than the photon energy of absorption, i.e. $\lambda_F > \lambda_A$, the process is called Stokes-type fluorescence. If the reverse is true, i.e. $\lambda_F < \lambda_A$, this process is called anti-Stokes type fluorescence.[1–3] *Direct line fluorescence* is nonresonance fluorescence, which is observed when a common upper level is involved for both excitation and fluorescence. Figure 1(b) shows an example of direct line fluorescence in which an atom is excited from the ground state to a higher excited electronic state and then undergoes a direct radiational transition to a metastable level above the ground state. *Stepwise line fluorescence* (Figure 1c) occurs when the upper energy levels of the exciting and of the fluorescence line are different. The excited atoms may undergo deactivation, usually by collision, to a lower excited state rather than return directly to the ground state. There are several more fluorescence lines which are not of much analytical use. Details about those fluorescence lines can be found in the literature.[1–6]

Generally, an atomic fluorescence spectrum consists of only a few lines because not all of these fluorescence-generating processes are active simultaneously.[7] Therefore, the atomic fluorescence spectrum is much simpler than the one obtained using atomic emission spectroscopy (AES).

1.2 Intensity of Fluorescence Radiance

It is beyond the scope of this article to provide details about the derivations of the intensity expressions. Only the results of the derivations will be given in order to understand the various dependences of the fluorescence signal. For a more comprehensive discussion and a complete derivation of the expressions, articles by Winefordner et al.[4,6,8] can be highly recommended.

The intensity of the fluorescence radiation produced in a transition depends on a number of factors, the most important of which are[5,7] (a) the intensity of the excitation source, (b) the concentration of the atom in

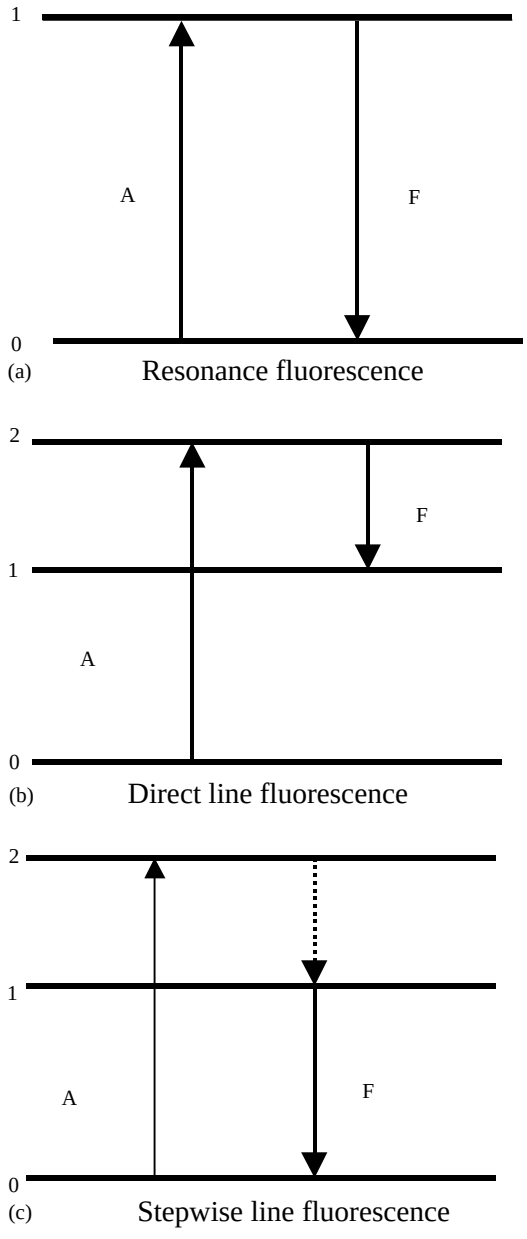

Figure 1 Basic types of atomic fluorescence lines.

the atomizer, (c) the quantum efficiency of the process (i.e. the ratio of the energy emitted in fluorescence to the energy absorbed, per unit of time), and (d) the extent of self-absorption in the atomizer.

In the derivation of intensity expression, it has been assumed that the atom under consideration has only two energy levels, ground state (1) and first excited state (2).[4] Other important assumptions are that the atoms are uniformly distributed in the atomizer and the concentration of the analyte atom is low, and the temperature of the atomizer and the radiation density of the source are spatially uniform. The fluorescence radiance[4,5,7] can be given by Equation (1):

$$B_F = \left(\frac{l}{4\pi}\right) Y_{21} E_{\nu_{12}} \int_0^\infty k_\nu \, d\nu \qquad (1)$$

where B_F is the absolute radiance; l is the pathlength in the direction of the detector; Y_{21} is the fluorescence power efficiency, $W_{\text{fluoresced}}/W_{\text{absorbed}}$; $E_{\nu_{12}}$ is the spectral irradiation of source at absorption line of frequency ν_{12}; and k_ν is the absorption coefficient.

The integration term gives the integrated absorption coefficient over the absorption line, which is a function of the concentration in states (1) and (2), the statistical weights of state (1) and state (2), and the Einstein B_{12} absorption coefficient.

The above equation gives some analytically important information. The fluorescence radiance B_F is linearly dependent on the source irradiance and the fluorescence quantum efficiency of the transition as long as $E_{\nu_{12}}$ is very much below the saturation value. Therefore, in order to obtain a large fluorescence signal it is necessary to use a highly irradiant light source. It can be further demonstrated that if the atomic concentration is low, B_F is linearly related to the total concentration of atoms in all states.[3]

1.3 Quenching

The intensity of the atomic fluorescence can be diminished by the collision between excited atoms and other molecules in the atomization sources. This process is called quenching. This problem is accounted for by the inclusion of the term Y_{21} in the fluorescence radiance equation given above. Since the quenching process is very important in atomic fluorescence, a few examples are given below. For a more comprehensive discussion on quenching processes, the reader is referred to several good reviews.[2,9]

1. The excess electronic energy of the excited atoms is converted into translational energy of the colliding species without involving the internal energy of the latter. For instance, this quenching can be processed by collision with free atoms (Equation 2):

$$A^* + B \longrightarrow A + B \qquad (2)$$

or with free electrons (Equation 3):

$$A^* + e^- \longrightarrow A + e^- \qquad (3)$$

2. The excess electronic energy of the excited atoms is released on collision, resulting in changes of the electronic or vibrational/rotational energy of the colliding species, for instance quenching by collision

with other atoms (Equation 4):

$$A^* + B \longrightarrow A + B^* \quad (4)$$

or with molecules (Equation 5):

$$A^* + BC \longrightarrow A + BC^* \quad (5)$$

Depending on the experimental conditions used, some of these quenching processes can significantly affect the fluorescence radiance.

1.4 Interferences

As with atomic absorption spectroscopy (AAS) and AES methods, interferences of two major types are encountered in atomic fluorescence.[10] Spectral interferences arise when lines in the source overlap lines of the matrix elements in the atomizer. Chemical interferences result from various chemical processes during atomization that reduce the population of free atoms.

Spectral interferences are seldom caused by line sources, but may be a problem with nondispersive instruments or in dispersive apparatus with continuum sources. Since generally there are several fluorescence lines available, it is sometimes necessary to select different lines to avoid the spectral interference. Chemical interferences caused by matrix components are one of the major problems in atomic fluorescence. This type of interference can be reduced by introducing different chemicals into the matrix, such as releasing agents and protective agents.[10] Coupling to vapor generation can also significantly reduce the problem.

2 INSTRUMENTATION

The basic layout of an AFS instrument is similar to those for AAS and AES except that the light source and the detector are located at a right-angle. A schematic diagram of an AFS instrument is shown in Figure 2. Although the configuration of an AFS instrument can be modified for different purposes, the basic apparatus consists of a radiation source, an atomizer, a wavelength-selection system, a signal detector, and an electronic readout system.

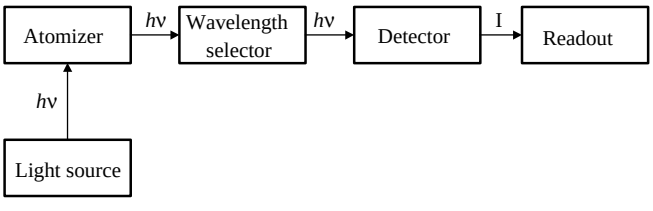

Figure 2 Schematic diagram of an AFS instrument.

2.1 Excitation Sources

A number of excitation sources have been used in AFS and can be categorized into two types, spectral line sources and continuum sources. Since the intensity of the fluorescence radiance is proportional to the intensity of the exciting radiation, sources with high radiance are required in order to achieve a good sensitivity and linear range. Indeed, a major part of the effort devoted to improvement of AFS instrumentation in recent years has been concerned with the excitation source. Winefordner discussed the criteria of importance in choosing excitation source for AFS and provided some general considerations.[3,4] Briefly, in addition to a high radiance over the absorption line (in order to obtain a good detection limit (DL) and linear range), the source should have good short- and long-term stability. In view of the need for convenience and ease of use, the source used should also require a minimum of maintenance and adjustment to obtain optimum performance. Only the most commonly used and recently developed sources are discussed here.

2.1.1 Hollow-cathode Lamps

Although the conventional, and commercially available, hollow-cathode lamp (HCL) is the most widely used line source used in AAS,[2,5] the early version of the HCL did not show impressive potential as a useful radiation source for AFS because it does not possess sufficient intensity. Different efforts have been made to improve the design of HCLs.[1,2,5,11,12] For instance, the lamps can be operated using short-duration pulses of current with a sufficient off-period between pulses to maintain the mean lamp current at a low level. The pulsed lamps provide an increase in peak intensity and a reduction of the total signal observation time.[1] Compared with regular HCLs, good results have been obtained with pulsed HCLs for a large number of elements.[1,5,13]

The boosted discharge hollow-cathode lamp (BDHCL), or boosted output HCL as it is sometimes called, produces intense spectra with narrow line widths and can be manufactured to provide an excellent source for many elements.[1,5,14] Since its introduction as an AFS excitation source by Sullivan and Walsh,[15] it has received a great deal of attention and a number of modifications to this type of source have been made.[1,11,16] Recently, BDHCLs have become commercially available and this contributed significantly to the availability of commercial AFS instruments.[12]

The operation principle of the BDHCL is illustrated in Figure 3.[14] The lamp consists of an anode mounted behind a cylindrical cathode. A primary discharge is struck between the cathode and anode to sputter atoms of the element of interest, as in a normal

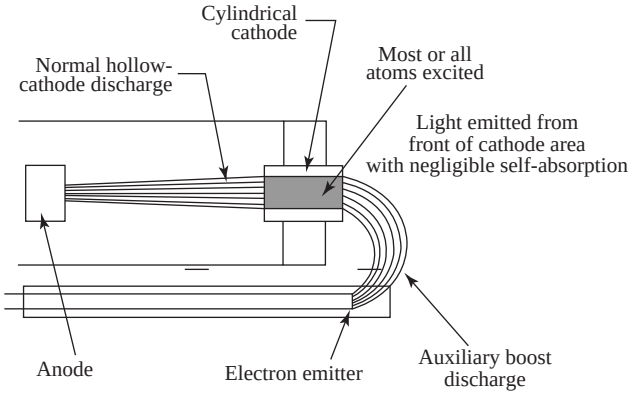

Figure 3 Schematic diagram of the boosted discharge HCL lamp. (Reproduced from *PSA 10.055 Millennium Excalibur, User Manual*, Issue 1.0, March 1998, by permission of P. S. Analytical.)

HCL. However, a secondary discharge (boost) is then struck between an efficient electron emitter and the anode, passing through the primary atom cloud. The use of a specifically developed high-efficiency electron emitter permits operation of the secondary discharge so that concentration broadening and line broadening are minimized.

2.1.2 Electrodeless Discharge Lamps

Electrodeless discharge lamps (EDL) are useful sources of atomic line spectra and provide radiant intensities that are usually one or two orders of magnitude greater than the normal HCLs.[10] This type of lamp has been widely employed for AFS. A typical lamp is constructed from a sealed quartz tube containing a few torr of an inert gas such as argon and a small quantity of the metal of interest (or its salt). The lamp does not contain an electrode but instead is energized by an intense field of radio frequency or microwave radiation. Radio and microwave frequency electromagnetic fields are very efficient for generating and accelerating electrons and thereby maintaining a gaseous glow discharge without internal electrode contact with the plasma.

The characteristics of EDLs have been studied and discussed in some detail.[5,17] Useful recommendations have been given by Van Loon[5] to predict lamp behavior and to select operating conditions for the EDLs for AFS. The main drawbacks of the EDLs are that their performance does not appear to be as reliable as that of the HCLs (signal instability with time) and they are commercially available only for some elements.

2.1.3 Xenon Arc Lamps

High-pressure xenon arc lamps, producing a stable and intense continuum source, have been commercially available for many years.[1,5] The lamp produces intense radiation by the passage of current through an atmosphere of xenon. The spectrum is continuous over the range 250–600 nm, with the peak intensity about 500 nm.[10] The advantage of this type of continuum source is that it can be readily employed for multielement analysis along with the use of a monochromator or interference filters. However, its intensity falls off severely below about 210 nm, thus making it unsatisfactory for the analysis of some environmentally important elements such as arsenic and selenium.

2.1.4 Laser Sources

Lasers have become important sources in analytical instrumentation because of their high intensity and narrow bandwidths and the coherent nature of their outputs. As a consequence of these unique properties, laser-induced fluorescence (LIF) offers a very sensitive and selective spectroscopic method, which has a low susceptibility to spectral interferences. Compared with conventional light sources, laser excitation also allows nonresonance transition lines to be used for many elements, which significantly reduces any laser radiation scattered off the vaporized sample.

LIF has been used for the determination of many elements in a variety of samples,[18–23] including environmental matrices.[21,24,25] The most sensitive DLs that have been reported have been as low as a few attograms.[21,26] The basic requirements for the laser include wavelength tunability across the ultraviolet (UV) and visible spectrum to allow for the determination of as many elements as possible. The most commonly used LIF is the pulsed excimer- or yttrium aluminum garnet (YAG)-pumped tunable dye laser because it provides sufficient peak energy and wavelength tunability. However, this type of source has some drawbacks, such as the need to replace the dry solution, which is often toxic and flammable, before a wavelength change of more than 20–30 nm, and dye degradation with time.[21] These disadvantages have significantly limited the application of this type of laser source to routine sequential multielement analysis. Research has been conducted to find alternatives to the pulsed dye laser. Recently, Zhou et al.[21] reported a pulsed (10 Hz) optical parametric oscillator (OPO) laser system based on β-barium borate crystals and equipped with a frequency-doubling option for use in LIF. This all-solid-state laser has a narrow spectral line width and wide spectral tuning range (220–2200 nm). It has been used for cobalt, copper, lead, manganese, and thallium analysis in Buffalo River sediment samples.

The production of far-UV radiation is required for excitation of some elements, such as arsenic at 193.70 nm and selenium at 196.03 nm. With the availability of β-barium

borate crystals, tunable UV radiation is now generated at wavelengths down to 205 nm by frequency-doubling. Sum frequency generation or stimulated Raman shifting techniques can further extend the UV range down to 187 nm with reasonable efficiency, making the determination of metalloids feasible.[22] Recently, Swart et al.[22] reported an LIF technique for selenium, arsenic, and antimony analysis. The far-UV radiation required was accomplished by stimulated Raman shifting of the output of a frequency-doubled dye laser operating near 230 nm.

2.1.5 Mercury Arc Lamps

The mercury arc lamp is a type of vapor-discharge lamp that forms simple, robust sources of line spectra.[10] The low-pressure mercury-vapor lamp equipped with a fused-silica window has been the most common source for filter fluorometers. This source produces intense lines at 254, 366, 405, 546, 577, 691, and 773 nm. Individual lines can be isolated with suitable absorption or interference filters. The mercury lamp has been used in commercially available instruments with a 254-nm filter for mercury analysis.[27]

2.2 Atomizers

Many atomizers used for AFS are similar to those used for AAS and AES. The design and properties of these atomizers have been discussed comprehensively in terms of their applications for AFS.[1–3,7] The basic requirements of an atomizer for AFS are an efficient and rapid production of free atoms with minimum background noise, long residence time of the analyte atoms in the optical path, low quenching properties, low cost of operation, and ease of handling. The quenching process is especially important for AFS and has been taken into consideration in the design of the atomizer. The order of quenching efficiency for some gases is $Ar < H_2 < H_2O < N_2 < CO < O_2 < CO_2$.[28] A number of atomizers have been developed and used for AFS, and some improvements have been made in recent years.

2.2.1 Flame Atomizers

Perhaps the most commonly used flame atomizer[5,7] in early studies was a premixed laminar flame with inert gas sheaths. Without using inert gas sheaths, the premixed laminar flame, in particular the acetylene-based flame, exhibits high background emission due to radicals present in the secondary reaction zone (OH, C_2, CH, CN) resulting in poor DLs. By preventing entrainment of the surrounding atmosphere into the flame gases, the secondary reaction zone may be shifted away from the primary reaction zone, thus extending the region of interconal zone, which is the most used part of the flame for spectroscopy. Separation of the secondary reaction zone can be effected using either nitrogen or argon. Argon is favored over nitrogen because of its greatly reduced quenching cross-section.

Flame emission and the associated flame flicker noise considerably increase the limits of detection in AFS. Flames with low background emission are therefore necessary. Diffusion flames, sometimes called cool flames, were developed and used for this purpose. The term diffusion flame is generally applied to flames in which the oxidant necessary for combustion is fully supplied by diffusion and/or entrainment from the surrounding atmosphere.[1] Argon is generally used as carrier gas and hydrogen is used as fuel gas. The temperature of these argon–hydrogen–air flames ranges from 280 to 850 °C, depending on the flame region selected. Despite the relatively low temperature achieved by these flames, they are still widely used as efficient atomizers for the analysis of some environmentally important elements. Indeed, a very simple atomizer based on an argon–hydrogen diffusion flame has recently been employed in a commercially available AFS system for hydride-forming elemental analysis.[12] The argon–hydrogen diffusion flame used emits very low background radiation over the wavelength region of interest, and the hydride compounds are easily decomposed in such a low-temperature flame. In addition, the quenching effect is relatively low in the argon-supported flame and in hydrogen.

Different designs of diffusion flame atomizers have been proposed and studied.[12,29–32] A simple glass or silica tube (i.d. 4–8 mm) has been used to support the argon–hydrogen flame.[31] Corns et al. recently investigated four different types of diffusion flame atomizers for arsenic analysis using hydride-generation/atomic fluorescence spectroscopy (HG/AFS) technique.[12] From this comprehensive study, they concluded that the AFS signals were observed only in the presence of argon–hydrogen flames. The attempt to achieve atomization using an electrically heated atomizer alone, without an argon–hydrogen diffusion flame, failed. These results provided further evidence to support the atomization mechanism proposed by several workers.[1,33–36] The atomization at such low temperature (about 800 °C) is not due to thermal decomposition, but to free radicals in the flame.

2.2.2 Electrothermal Atomizers

Although the flames discussed above are very convenient and have been widely used as atomization cells in AFS, they do have disadvantages for some analytical work. The high dilution of the sample by flame gases limits the atom concentration attainable. The large sample requirement with continuous nebulization makes

it unsuitable for small-volume sample analysis, and the flame cells are not convenient for the direct atomization of solid samples. Precise control over the chemical environment in the flame is impossible. These drawbacks of the flame atomizer can be overcome by using electrothermal atomization (ETA) techniques. These electrothermal atomizers are designed in different shapes and are generally made of graphite. Graphite atomizers are operated at high temperature and the atoms are contained in an inert gas atmosphere. This provides a high atomization efficiency and reduced quenching, consequently offering high sensitivity for small volumes of sample. It has been suggested that the most successful atomic atomizer for LIF spectrometry is the graphite furnace.[19]

LIF coupled with ETA has been one of the main research fields for many scientists in the past two decades.[19,20,22–24,37–40] The development of electrothermal atomizers used for AFS benefits greatly from the analytical knowledge accumulated over the years with the well-established atomic absorption technique. Generally, certain modifications of the conventional graphite furnaces are required if they are to be usable in LIF. The optical detection of fluorescence needs to be made at 90° with respect to the excitation axis. The following are the typical forms of the graphite furnace studied for AFS: graphite cup,[37] Massmann cup,[41] graphite rod,[42,43] and graphite tubes.[38,39] There seems to be a general consensus that graphite-tube atomization is preferable to other graphite atomization techniques,[39] and it has been used in most of the later studies.[19,20,22] With graphite-tube atomization, the atoms are contained in a hot environment while being excited by the laser, and therefore much better analytical performance can be achieved.

Besides examining a variety of designs and shapes for the furnace, other studies, such as those atomizing at a pressure-controlled[20] and pyrolytically coated[44] graphite atomizer, have also been conducted to improve the performance of these furnaces. The use of low-pressure vaporization to remove matrix interferences associated with the analysis of solid samples has been coupled with laser-induced fluorescence/electrothermal atomization (LIF/ETA).[45] A lower DL was found by using the low-pressure vaporization technique. Using a coating on graphite furnaces was initially conducted in ETA in the AAS technique to inhibit reaction between the analyte and the graphite.[46] Among the different coating methods used for the graphite atomizer in AFS, a pyrolytic coating has been the most popular.[19,20,38]

2.2.3 Cold-vapor Technique for Mercury

Mercury is the only metallic element that exhibits an appreciable atomic vapor pressure at ambient temperature. The mercury has a vapor pressure of 0.0016 mbar ($1\,\text{mbar} = 10^2\,\text{Pa}$) at 20 °C, corresponding to a concentration of approximately $14\,\text{mg}\,\text{m}^{-3}$ in air.[46] This unique property gives rise to the possibility of measuring mercury without the additional thermal energy supplied by a flame or electrothermal heating. The cold-vapor atomic absorption spectrometry (CVAAS) technique for mercury analysis has been studied extensively. This method is readily applicable to AFS [cold-vapor atomic fluorescence spectrometry (CVAFS)]. Briefly, the mercury vapor, produced by reduction of the metal from its compounds with suitable reductants [sodium borohydride or tin(II) chloride], is swept out of the reaction vessel using argon carrier gas and introduced into the optical beam where the mercury atoms can be excited with a suitable source.

2.2.4 Miscellaneous Atomization Techniques

The inductively coupled plasma (ICP) has been proposed as an atomizer for AFS.[47,48] Plasmas have better vaporization and atomization efficiency than flames and give less interference problems. Within an argon atmosphere, ICP is characterized by a high fluorescence yield because of the small quenching cross-section. Background emission is dependent on the viewing region but is generally higher than in flame.[28] Although the ICP shows some advantages as an atomizer, the high cost compared with flame and electrothermal atomizers has discouraged its widespread application to routine AFS.

The atomization of an analyte can be also carried out on the principle of an HCL in which the sample is vaporized by a glow discharge.[46] Recently, a microsecond pulsed glow discharge was studied as an analytical spectroscopic source for solid-sample analysis.[49] The results showed enhanced efficiency for analytical response of the sputtered sample atoms.

2.3 Wavelength Selection and Detection Systems

The simplicity of atomic fluorescence spectra makes the isolation and detection of the selected atomic fluorescence lines easier than in other atomic spectrometric techniques such as AAS and AES. Both a monochromator and an interference filter have been used as wavelength selectors with different advantages and limitations. The requirements for wavelength selection depend on the light source employed. With conventional line sources (HCL, EDL) or laser excitation, low-dispersion monochromators or nondispersive systems can be used. The nondispersive instruments are simple and cheap, and readily adaptable to multielemental analysis. They also have a high energy throughput and thus high sensitivity.[10] However, to realize these advantages, it is necessary that the output of the source is free of contaminating lines from other elements, and no significant background radiation should occur. In most of the instruments, a filter located between

the source and detector has been used to remove background radiation. With a continuum excitation source, a dispersive system is required to discriminate against the scattering of source radiation and the fluorescence radiation from other species over the entire ultraviolet/visible (UV/VIS) region.[28]

Fluorescence radiation is usually detected with a photomultiplier tube (PMT) as in many other spectrometers.[7,10] The PMTs are highly sensitive to UV and visible radiation and have extremely fast response times.

Since the wavelength selection and detection systems used in AFS are similar to those of AAS and AES, readers are referred to other comprehensive discussions in the literature.[10,46] Over the years, a number of AFS instruments have been developed, and have either been used exclusively in research laboratories or have only recently become commercially available. To the author's knowledge, AFS instruments for the analysis of mercury and some other elements are currently being manufactured by three companies: P. S. Analytical Ltd, Sevenoaks, Kent, UK, Brooks Rand Ltd, Seattle, WA, USA, and Tekran Inc., Ontario, Canada. Different features of these instruments will be discussed in the following sections along with their applications in environmental analysis.

3 ENVIRONMENTAL APPLICATIONS

3.1 Mercury

Mercury pollution has become a global problem because of its wide occurrence in the environment. The toxicological effects of mercury compounds on environmental systems have long been recognized. Therefore, mercury monitoring is a special concern in the field of heavy metal analysis. AFS-based techniques have been widely used in environmental analysis. The resonance fluorescence line at 253.7 nm ($^3P_1 - {}^1S_0$ transition) has been generally used in most atomic fluorescence studies on mercury. AFS instruments are now commercially available for both total mercury and organomercury species analysis.

3.1.1 Total Mercury

3.1.1.1 Instrumentation Traditionally, CVAAS has been the most commonly used method for determining mercury at trace levels. This technique has been well established and several commercial systems are available.[46] CVAFS was proposed almost three decades ago as a method for mercury analysis.[50–52] The advantages of AFS over AAS, in terms of sensitivity, linear range, and spectral interferences, have been demonstrated both theoretically[53] and experimentally.[50,54] However, it was only during the beginning of the 1990s that AFS became commercially available.[55,56] AFS detection, especially coupled with the vapor generation technique, is becoming popular and replacing AAS for mercury analysis in many research and service laboratories owing to its unique sensitivity, specificity, and simplicity.

Similarly to CVAAS, a general CVAFS system consists mainly of two parts, the mercury vapor generator and the AFS detector. A schematic diagram of a continuous-flow vapor generator is shown in Figure 4.[27] The reductant, blank (1% HCl), and sample solutions are delivered by variable-speed multichannel peristaltic pumps. An electronically controlled switching valve alternates between blank and sample solutions, and two of the liquid streams (reductant and sample or reductant and blank) are butt-mixed in the sample valve, where the reaction starts to occur. The steams and all gaseous products are continuously and rapidly pumped into a glass gas–liquid separator, from which the gaseous products are carried, by argon gas, through a dryer system, finally reaching the AFS detector.

The design of the AFS detector for mercury analysis is relatively simple owing to the absence of a thermally energized atomizer.[55,57] Generally, a UV mercury vapor lamp is used as excitation source and the fluorescence light is detected by a PMT, which is positioned perpendicular to the excitation source. Figure 5 is a schematic diagram illustrating the optical configuration of an AFS system.[27] A glass chimney is used for introduction of mercury vapor into the optical path. The chimney is shielded with a high flow rate of argon. A 253.7-nm interference filter is employed between the introduction chimney and the PMT to keep stray light away from the latter.

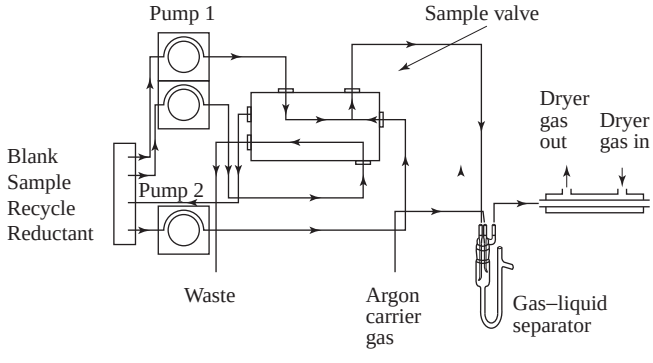

Figure 4 Schematic diagram of the continuous-flow vapor/hydride generator. (Reproduced from *PSA 10.055 Millennium Excalibur, User Manual*, Issue 1.0, March 1998, by permission of P. S. Analytical.)

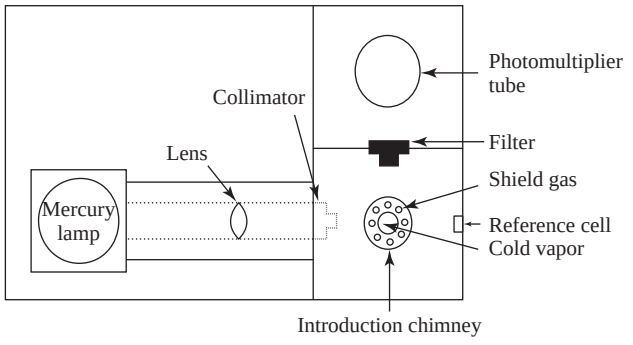

Figure 5 Schematic diagram illustrating the optical configuration of an AFS system for mercury analysis. (Reproduced from *PSA 10.025 Millennium Merlin, User Manual*, Issue 1.0, October 1997, by permission of P. S. Analytical.)

3.1.1.2 Liberation and Reduction of Mercury Since this system is used to determine total inorganic mercury (Hg^{2+}), the first requirement for performing analysis is to liberate all mercury compounds from the sample matrices and convert all organic forms of mercury to Hg^{2+} by various digestion/oxidation procedures.

For natural water sample analysis, the previously employed hot oxidizing method using permanganate–peroxodisulfate has been found unsuitable for low-level mercury determination because of the high blanks found in these reagents.[58] Bromine monochloride has been found to be an excellent oxidant and preservative for total mercury in water samples,[59] working faster and more efficiently on many organomercurials.[58,59] Hydroxylamine hydrochloride is added to destroy the excess bromine before analysis with CVAFS. This method has often been used for converting organic mercury to Hg^{2+}.[59–64]

Both open-vessel and closed-vessel digestion/oxidation techniques have been described for solid sample (soil, sediment, and tissue) analysis.[59,65–67] Open-vessel digestion results in considerable loss of organic mercury compounds by volatilization at elevated temperature or by incomplete oxidation of the sample matrix at ambient temperature. The open-vessel method is also complex and requires large amounts of glassware. These drawbacks can be overcome by using a closed decomposition system at elevated temperature. Microwave-assisted[65,66] and autoclave-assisted[67] closed-vessel digestion methods have been reported to achieve complete oxidation while minimizing losses of mercury. By using these closed-vessel methods, the preparation of the sample is effective and allows for a large number of samples to be prepared simultaneously and efficiently.

Once all forms of mercury have been converted to Hg^{2+}, the latter is reduced to elemental mercury (Hg^0) by using either acidified stannous chloride or sodium borohydride (Equations 6 and 7):

$$Hg^{2+} + Sn^{2+} \longrightarrow Hg^0 + Sn^{4+} \quad (6)$$

$$Hg^{2+} + 2NaBH_4 + 6H_2O \longrightarrow Hg^0 + 7H_2 + 2H_3BO_3 + 2Na^+ \quad (7)$$

The mercury vapor produced is carried by argon gas to the AFS instrument for detection. A typical curve obtained using this type of CVAFS system is shown in Figure 6. Peak area is generally used for quantification. This system has been successfully applied to a wide range of environmental samples, such as water,[56,60,63,66] soil,[60,66,67] and biological samples.[60,67]

3.1.1.3 Amalgamation The concentrations of mercury in natural waters are often found to be at sub-parts per trillion (nanograms per liter) levels. The determination of such low concentrations frequently requires a preconcentration step prior to AFS detection. Many of the enrichment techniques are based on the fact that mercury can be deposited easily on metals such as gold and copper.[46] The amalgam technique was used simply to improve the DLs of the cold-vapor technique. Normally, as shown in Figure 6, the mercury vapor is liberated slowly from solution and thus generates a broad signal. If instead the mercury vapor is collected by amalgamation,

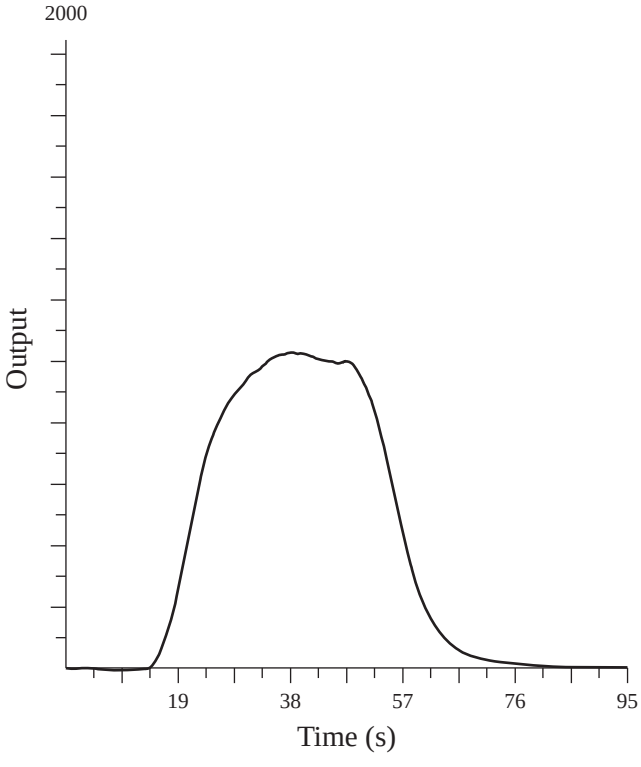

Figure 6 A typical response curve for mercury analysis using CVAFS.

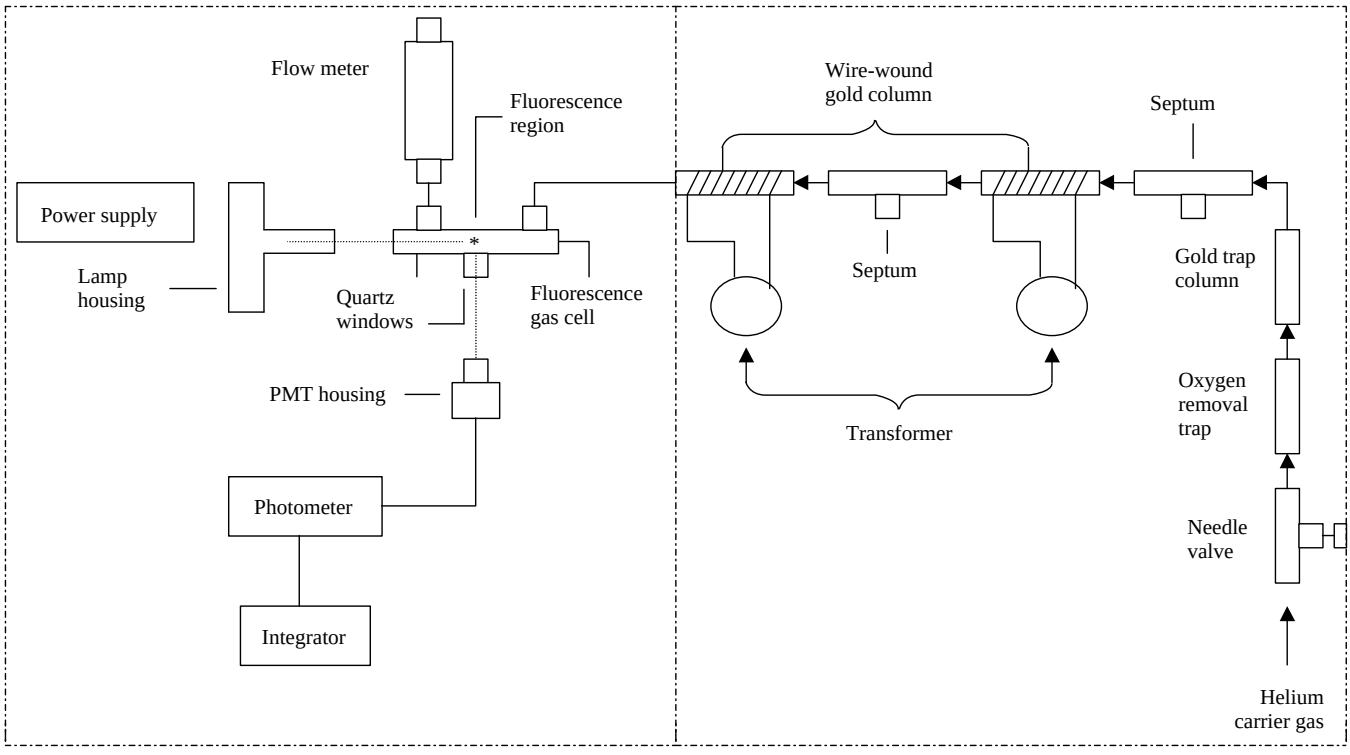

Figure 7 Schematic diagram of the Hg detection system. The boxed area on the left is the nondispersive resonance atomic fluorescence system. The boxed area on the right shows the two-stage gold amalgamation gas train used to introduce elemental Hg vapor (trapped on a gold column during the purging step) into the gas cell of the detection system. The carrier gas is high-purity helium, from which oxygen and Hg have been removed with trap columns. The flow rate is controlled by a needle valve positioned before the oxygen trap column and is monitored at the exit of the fluorescence cell with a flow meter. All connections are made with Teflon or silicone-rubber tubing. The gas flow direction is denoted by arrows. (Reprinted with permission from G.A. Gill, K.W. Bruland, *Environ. Sci. Technol.*, **24**, 1393 (1990). Copyright 1990 American Chemical Society.)

on a gold trap, for example, and then subsequently released by heating the trap rapidly to 500–700 °C, a higher signal-to-noise ratio (S/N) is obtained. A two-stage gold trap technique coupled with AAS for determining total mercury in environmental samples was described.[68] Gold-trap preconcentration of Hg^0 followed by AFS has also been reported.[57] Figure 7 shows a schematic diagram of the two-stage CVAFS system.[69] Because of the excellent sensitivity provided by this technique, gold-trap CVAFS is currently widely used in environmental analysis.[61,62,64,69–71] The United States Environmental Protection Agency (USEPA) has proposed, in the *Federal Register*, a method (Method 1631) for the determination of mercury in waters using oxidation, purge and trap, and CVAFS.[71] However, trace levels of mercury have been found in most reagents used. The application of this technique is often limited by the fact that mercury present as impurity in these reagents is also preconcentrated.

3.1.1.4 Miscellaneous Atomic Fluorescence Spectrometric Techniques for Total Mercury Determination LIF with ETA has been reported for total mercury determination with a 0.09-pg absolute DL. However, because of the small size of the sample to be used (only 10 µL), the concentration DL in water was only 9 ng L^{-1}.[72] This is not sufficient for total mercury determination in most natural waters, therefore limiting the application of this method.

Several flow-injection (FI) atomic fluorescence methods incorporating an on-line bromide–bromate oxidation step to determine mercury in water samples have been described.[73,74] The CVAFS-based technique uses a heated reaction coil in the FI manifold to increase the conversion of organic mercury into Hg^{2+}. However, the DL (6–25 ng L^{-1}) achieved by these techniques cannot, in many cases, meet the requirements for the analysis of environmental samples because only a small volume of sample can be used.

3.1.2 Mercury Speciation

Speciation is the identification and quantitation of the individual physicochemical forms of an element in a sample that together constitute its total concentration. It is well known that the various chemical forms (organic and inorganic) of mercury in the natural environment

behave differently, thereby affecting its biogeochemistry and toxicity to organisms. The development of analytical techniques able to distinguish different inorganic and organomercury species has been a main research area for scientists dealing with mercury in the environment. Basically, chromatography [gas chromatography (GC) and high-performance liquid chromatography (HPLC)]-based AFS, and purge and trap-based AFS are the two most widely used techniques for measuring inorganic and different forms of organomercury compounds, especially methylmercury.

3.1.2.1 Chromatographic Methods Gas chromatography/electron capture detection (GC/ECD) was traditionally used for the determination of MeHgCl or MeHgBr.[75] However, a major problem with GC/ECD is that the halogen-bearing compounds co-extracted with organic mercury interfere with the determination because of the nonspecificity of the ECD. Use of the very sensitive and selective AFS as a GC detector has successfully addressed this problem.[60,76–79] Gas chromatography/atomic fluorescence spectroscopy (GC/AFS) systems have been used for determination of methylmercury and ethylmercury in various environmental samples. A schematic diagram of the GC/AFS system is shown in Figure 8. A commercially available gas chromatograph, equipped with a megabore fused-silica column (DB-1 or BD-5), was used to separate organomercury chlorides or bromides. A pyrolyzer heated electrothermally at 800 °C was added between the column outlet and the AFS detector, converting organomercury compounds to Hg^0. The pyrolyzer was made of deactivated fused-silica tubing. The mercury atoms formed in the pyrolysis unit were transferred via a 0.5-mm i.d. Teflon tube into the AFS detector, which has a similar design to that shown in Figure 5. Argon gas was used as make-up and shield gases, and both helium and argon could be used as the GC carrier gas. A typical chromatogram obtained using this system for the determination of methylmercury and ethylmercury compounds is given in Figure 9. Column

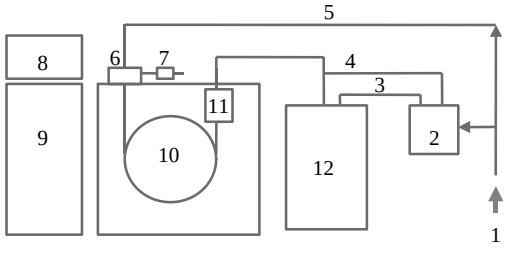

Figure 8 Block diagram of a GC/AFS system for the determination of organomercury compounds. (1) argon; (2) mass flow controller; (3) sheath gas; (4) make-up gas; (5) carrier gas; (6) injection port; (7) split value; (8) autosampler; (9) injector controller; (10) column; (11) pyrolyzer; (12) AFS detector.

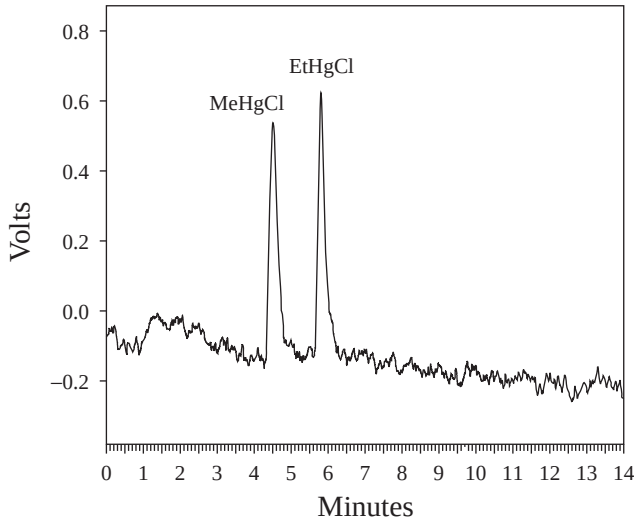

Figure 9 Typical chromatogram obtained using GC/AFS for the determination of methylmercury chloride (MeHgCl) and ethylmercury chloride (EtHgCl).

maintenance is always an important step when a GC method is used for organomercury halide determination. With regard to the GC/AFS system, the treatment and maintenance of the GC injection port, the column, and sample preparation, have been discussed.[77,79]

In order to improve GC performance, organomercury halides can also be converted to fully alkylated derivatives by different derivatization reactions before GC/AFS analysis. Two common derivatizing methods are ethylation using sodium tetraethylborate ($NaBEt_4$) and alkylation with Gringard reagents (e.g. butylmagnesium chloride, BuMgCl).[78,80,81] The application of $NaBEt_4$ is limited because this procedure does not distinguish between ethylmercury and inorganic mercury (Hg^{2+}). The method of Grignard reactions has to be carried out in completely dried organic solvents and therefore involves tedious workup steps, which inevitably introduce analytical errors. Other derivatization reagents, such as sodium tetraphenylborate ($NaBPh_4$)[82] and sodium tetrapropylborate ($NaBPr_4$),[83] have also been studied and used for organomercury speciation by GC/AFS.[84]

Sample preparation is one of the critical steps for MeHg determination in environmental samples. For water samples, liquid–liquid extraction (LLE) with an organic solvent has been a popular preconcentration method. However, the large volumes of both water sample and organic solvent required make it impractical for real sample analysis, especially when a large sample throughput is needed. The two currently used methods are based on steam distillation[85] and solid-phase extraction with sulfhydryl cotton fiber.[77] Isolation of organomercury compounds from soil and biological samples requires

special attention because mercury species cannot be alternated during the isolation procedure. Recently, evaluation of some isolation methods for organomercury determination in soil and fish samples by GC/AFS has been reported.[79]

3.1.2.2 Purge and Trap-based Gas Chromatographic/ Cold-vapor Atomic Fluorescence Spectrometric Methods Purge and trap methods for the determination of methylmercury requires conversion of ionic $MeHg^+$ to nonionic species. The generally used method is aqueous ethylation with $NaBEt_4$. The use of $NaBEt_4$ as an aqueous derivatization agent was first reported by Rapsomanikis et al. for lead determination using a purge and trap AAS system.[86] Bloom reported an analytical method for determination of MeHg in aqueous samples by aqueous-phase ethylation with $NaBEt_4$, followed by cryogenic gas chromatography/cold-vapor atomic fluorescence spectrometry (GC/CVAFS).[87] Several studies and modifications have been conducted in terms of the trap materials and methods. In typical room-temperature trapping experiments, the volatile derivatives are trapped on a Carbotrap® or Tenax-TA® trap, thermally desorbed, separated by isothermal packed-column GC, decomposed at 700–900 °C by a pyrolyzer, and then detected by CVAFS.[88] In the cryogenic trapping method, the derivatives produced are purged by helium gas, trapped directly on a packed GC column immersed in liquid nitrogen (−196 °C), and then separated by electrothermally heating the column.[89] The ethylation approach with room temperature trapping is at present the most popular method for speciation of mercury, but ethylation followed by cryogenic trapping is also effective.

Purge and trap CVAFS is an extremely sensitive method for mercury determination. A DL (3σ of the mean blank) of $0.016\,ng\,L^{-1}$ for MeHg determination in water samples has been reported.[61] However, this method cannot distinguish between inorganic mercury and ethylmercury because they both give diethylmercury as products. In addition, severe interferences are frequently encountered in the ethylation step and MeHg artifact has sometimes been observed when environmental samples with a high inorganic mercury content were analyzed.[90]

Hydride-generation with $NaBH_4$ for MeHg analysis was reported in 1992 by Filippelli et al.[91] and Craig et al.[92] MeHg was converted to volatile MeHgH, and Hg^{2+} was converted to Hg^0 (Equations 8 and 9):

$$MeHg^+ + NaBH_4 \longrightarrow MeHgH \quad (8)$$

$$Hg^{2+} + NaBH_4 \longrightarrow Hg^0 \quad (9)$$

Several research groups have demonstrated the presence of MeHgH in the gas phase by using different techniques.[91,92] Hydride-generation followed by CVAFS was reported for mercury determination in various environmental samples.[93]

3.1.2.3 Miscellaneous Atomic Fluorescence Spectrometric Techniques for Mercury Speciation AFS can also be used as an HPLC detector. Hintelmann and Wilken described a method for the determination of organomercury compounds using HPLC with on-line AFS detection.[94] The organomercury compounds separated by HPLC were converted to Hg^0 in a continuous-flow system by an oxidizing, and a subsequent reducing, solution. The Hg^0 generated was separated in a gas–liquid separator and swept into the cell of an AFS detector by a stream of argon. This system was successfully used to separate eight organomercury compounds. The absolute limit of detection (LOD) for MeHg was reported to be about 0.02 ng. This corresponds to a concentration LOD of $0.8\,\mu g\,L^{-1}$ for a 25-μL injected volume. Although HPLC has an advantage over GC in that formation of volatile derivatives is not necessary, the LOD achieved do not appear to meet the requirement for most environmental samples, especially for natural waters.

Recently, Cai et al. reported an analytical method involving ethylation using $NaBEt_4$ followed by solid-phase microextraction (SPME) and GC/AFS detection for the rapid determination of methylmercury in fish.[81] The procedure involves aqueous ethylation with $NaBEt_4$ and adsorption of Et_2Hg and MeHgEt from aqueous solution or headspace on a fiber coated with poly(dimethylsiloxane). After retraction of the fiber into a syringe needle, the analytes are desorbed in an injection port of a gas chromatograph, separated, and detected by AFS using the system shown in Figure 8. At present this procedure cannot be used for the determination of MeHg in natural waters owing to the limited sensitivity.

3.2 Hydride-forming Elements

Many environmentally important elements, such as arsenic, selenium, bismuth, antimony, tellurium, germanium, tin, lead, and mercury, can form volatile and covalent hydrides with ''nascent'' hydrogen. Hydride-generation is a preferred technique when hydride-forming elements are to be analyzed by atomic spectrometry. The advantage of volatilization as a gaseous hydride clearly lies in the separation and enrichment of the analyte element and thus in a reduction or even complete elimination of interferences.[46] Since many features of these hydride-generation techniques are similar between different elements in terms of the reactions and instrument designs, the analytical aspects of all these elements using

hydride-generation followed by AFS are discussed in this section. Of course, these elements can also be determined using other AFS techniques, which will be summarized in the next section along with the determination of other non-hydride-forming elements by AFS.

Since Holak in 1969 first applied hydride-generation for the determination of arsenic using AAS,[95] a number of papers have been published describing modifications and optimization of the techniques. Currently, the number of metals and metalloids that can be determined by the hydride-generation technique has increased to about nine (arsenic, selenium, bismuth, antimony, tellurium, germanium, tin, lead, and mercury). Although the major research was carried out using AAS for detection,[46] most of the elements have also been studied using hydride-generation followed by AFS detection. Tsujii and Kuga in 1974[96] were the first to describe hydride-generation coupled to nondispersive AFS. They reported a DL of 2 ng for arsenic analysis. Thompson in 1975[97] was the first to apply a dispersive AFS system for the determination of arsenic, selenium, antimony, and tellurium after hydride generation. DLs using this system ranged from 0.06 to 0.1 $\mu g\,L^{-1}$.

Although various metal–acid reactions (e.g. Zn–HCl) have been used as a means of producing hydride, $NaBH_4$ is currently used exclusively. By using $NaBH_4$ as the reductant, the technique is easy to automate, since only solutions are involved, so that a high sample throughput can be achieved. In contrast to the manual injection of $NaBH_4$ or sample solution into the reaction, all the solutions can be delivered by a continuous-flow hydride-generation system.[12] Such a system is currently commercially available.[14] The design of this system is similar to that of the cold-vapor generation system used in CVAFS for mercury analysis as shown in Figure 4.

Once the hydride has been formed and driven out of the solution, it can be directly delivered by an inert gas such as argon to the atomizer where it is excited by a fluorescence light source and measured by a detector such as solar blind PMT. The hydride produced can be first collected in a trap cooled in liquid nitrogen before it is warmed up to vaporize it for measurement. The configuration of this design is similar to that used in the purge and trap system for mercury determination discussed above. For routine analysis, the former (direct) method offers the advantage of a higher sample throughput and easy operation. However, the latter approach provides the highest yield and best sensitivity.[46] Both procedures have been widely used in HG/AFS.

When determination is needed for different ionic species of elements present in a sample, various types of HPLC are employed to separate these forms of compounds, which are then subjected to hydride generation, followed by direct or purge and trap AFS detection. Sometimes an on-line decomposition step is added between HPLC and hydride generation to convert non-hydride-forming species into hydride-forming species. A chemical reagent, such as potassium peroxodisulfate, assisted by UV photochemical oxidation or microwave digestion has been used for this purpose.[98–101]

Although occasionally an electrothermal atomizer has been used, an argon–hydrogen diffusion flame atomizer is the usual choice because of very low background radiation, easy decomposition of hydrides in such a low-temperature flame, and the low quenching effect in the argon-supported flame. The atomization mechanisms of hydrides in different atomizers have been studied extensively. Evidence has been presented showing that the atomization of hydrides in a cool argon–hydrogen diffusion flame is not the result of a simple thermal dissociation process.[12,35,46] Dedina and Rubeška concluded that atomization is brought about by free radicals produced in the primary reaction zone of the oxygen/hydrogen flame according to Equations (10–12):[35]

$$H + O_2 \longrightarrow OH + O \tag{10}$$

$$O + H_2 \longrightarrow OH + H \tag{11}$$

$$OH + H_2 \longrightarrow H_2O + H \tag{12}$$

It can be shown that the concentration of H radicals is several orders of magnitude higher than that of OH radicals. Taking selenium as an example, the atomization is via a two-step mechanism with the predominating H radicals (Equations 13 and 14):

$$SeH_2 + H \longrightarrow SeH + H_2 \tag{13}$$

$$SeH + H \longrightarrow Se + H_2 \tag{14}$$

It has already been shown that there is a large number of H radicals present in an argon–hydrogen diffusion flame, so the same reaction mechanisms can be expected.[46]

As with any other technique, HG/AFS suffers from interferences. Welz[46] gave an extensive discussion about various interferences encountered in hydride-generation methods. However, the sensitivity of the AFS system allows dilution of the sample so that the interferences are reduced significantly. Spectral interference is not a problem for HG/AFS because the analyte element passes into the atomizer as gaseous hydride, while concomitants normally remain in the reaction vessel. In most currently used nondispersive AFS detectors, an interference filter is located between the atomizer and the PMT, to reduce the flame background and emissions from the excitation source that do not produce intense fluorescence lines.[12] The flame background can also be minimized by choosing suitable flow rates of gases. Kinetic interferences are

caused by varying rates of development or liberation of the hydride from solution. These interferences only occur in the direct hydride-generation system, and not in the purge and trap system. A typical kinetic interference is the retardation of hydride liberation because of the sample matrix effect. Chemical interferences are perhaps the most frequently encountered problem. It appears that transition metals interfere significantly with hydride-generation reactions.

3.2.1 Arsenic

Arsenic was the first element to be determined using the HG/AFS technique.[96] A number of papers have been published on arsenic determination in environmental samples. Azad et al.[102] described a procedure for the determination of arsenic in soil digests in which arsine was generated with $NaBH_4$, passed to an argon–hydrogen air-entrained flame and an atomic fluorescence signal, excited by a modulated EDL, and measured using nondispersive AFS. The authors reported a DL of $3\,\mu g\,L^{-1}$ calculated as the mass of arsenic required to produce a S/N of 2 for the atomic fluorescence signal. Most papers published in the past utilized an EDL as the excitation source because this type of lamp provides intense and narrow lines. However, the EDL is often unpredictable in use and requires careful temperature control to achieve good stability. A BDHCL seems to be a reliable high-intensity excitation source and this type of lamp has been used in commercial AFS instruments. Corns et al.[12] described a fully automated continuous-flow hydrogen generation AFS system for the determination of hydride-forming elements, principally arsenic and selenium. A minature argon–hydrogen diffusion flame was used as the atomizer and a BDHCL was used as the excitation sources. The hydrogen for the flame was chemically generated as a byproduct of the $NaBH_4$ reduction. Fluorescence wavelengths of interest were selected using an interference filter. A solar blind PMT was used as the detector. A DL (3σ) of $0.1\,\mu g\,L^{-1}$ was reported for arsenic.

To date, about a dozen arsenic compounds have been found in environmental samples.[101] Extensive studies on the toxicity and biogeochemistry of arsenic have clearly depicted the importance of the chemical speciation of this element. To determine these different forms of arsenic in environmental samples, a powerful separation technique, such as HPLC, has to be used. A number of analytical systems based on high-performance liquid chromatography/hydride-generation/atomic fluorescence spectrometry (HPLC/HG/AFS) have been reported. Recently, Gomez-Ariza et al.[98] described an anion-exchange HPLC/HG/AFS system for the speciation of arsenite, arsenate, dimethylarsinic acid (DMAA), and monomethylarsonic acid (MMAA). The products of hydride-generation of arsenite, arsenate, DMAA, and MMAA are arsine (AsH_3), arsine, Me_2AsH, and $MeAsH_2$, respectively. DLs of 0.17, 0.38, 0.45 and $0.30\,\mu g\,L^{-1}$ were reported for arsenite, arsenate, DMAA, and MMAA, respectively. A schematic diagram of this system is shown in Figure 10. Arsenobetaine, a non-hydride-forming species of arsenic, was also determined by introducing an on-line photooxidation step after the chromatographic separation. Le et al.[100] described an on-line microwave derivatization coupled with HPLC/HG/AFS for arsenic speciation. They studied the separation of 11 arsenic compounds by using ion-pair chromatography at 30, 50, and 70 °C column temperature and found that the use of elevated column temperature improved the separation efficiency and dramatically reduced the chromatographic time for some arsenic species. This speciation technique was successfully applied to a study of metabolites of arseno sugars present in commercial seaweed products. A typical chromatogram obtained for speciation of seven arsenic compounds is shown in Figure 11.

Hydride-generation followed by purge and trap AFS can also be used for speciation of hydride-forming arsenic

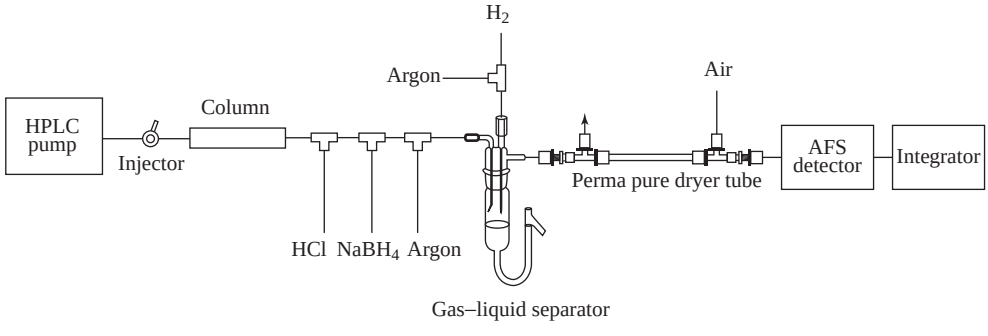

Figure 10 Operating scheme of the HPLC/HG/AFS system for arsenic speciation. (Adapted from J.L. Gomez-Ariza, D. Sanchez-Rodas, R. Beltran, W. Corns, P. Stockwell, *Applied Organomet. Chem.*, **12**, 1 (1998). Copyright 1998 John Wiley & Sons Limited. Reproduced with permission.)

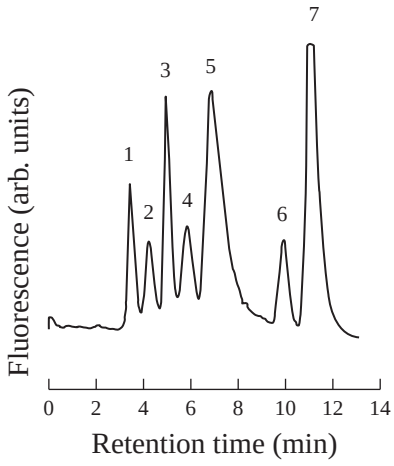

Figure 11 Chromatogram of seven arsenicals obtained on a reversed-phase C_{18} column (300 × 3.9 mm i.d., 10-μm particles) with 10 mM hexanesulfonate, 1 mM tetraethylammonium hydroxide, and 0.5% methanol as eluent (pH 4.0). Microwave digestion combined with hydride generation and AFS was used for detection. Column temperature, 70 °C. Peaks: 1, arsenate (10 ng); 2, arsenite (40 ng); 3, arsenobetaine (15 ng); 4, MMAA (10 ng); 5, DMAA (40 ng); 6, arsenocholine (15 ng); 7, tetramethylarsonium (40 ng). (Reprinted with permission from X.C. Le, M.H. Ma, N.A. Wong, *Anal. Chem.*, **68**, 4501 (1996). Copyright 1996 American Chemical Society.)

species such as arsenite, arsenate, DMAA, and MMAA. A pH-selective hydride-generation process should be used.[103] This process comprises two steps. First, arsenite is reduced to AsH_3 by $NaBH_4$ at pH 6. The product is swept out of the solution and collected with a cooled trap under liquid nitrogen. The trap is then allowed to warm to room temperature and the arsine is released into the AFS system for detection. Second, arsenate, DMAA, and MMAA in the sample solution are reduced to the corresponding arsines at pH 1, then followed by purge and trap AFS detection.

3.2.2 Selenium

The AFS determination of selenium was first reported by Dagnall et al.[104] using a dispersive spectrometer equipped with an air–propane flame giving a DL of 0.25 μg mL^{-1} of selenium on aspiration of aqueous solution using a pneumatic nebulizer. A dramatic improvement in DL (10 μg L^{-1}) was obtained by using a similar experimental arrangement, but using an argon–hydrogen diffusion flame and the hydride-generation technique.[105] This HG/AFS system was used for the determination of selenium in soil samples. A current commercially available HG/AFS instrument[12] utilizes a BDHCL as the excitation source and also an argon–hydrogen diffusion flame as the atomizer. A DL (3σ) of 0.05 μg L^{-1} was reported. A standard reference material (SRM) for trace elements in water, SRM 1643d from the National Institute of Standards and Technology (NIST) (Gaithersburg, MD, USA), was used to test the performance of this system.[84] The concentration of selenium was found to be 11.40 ± 0.07 μg L^{-1} ($n = 3$), which was in close agreement with the certified value of 11.43 ± 0.17 μg L^{-1}.

Selenium can also exist as different forms, both inorganic and organic, in environmental samples. Volatile organic selenium compounds can be directly determined using purge and trap followed by AFS without derivatization (see the next section). The determination of inorganic forms, Se(IV) and Se(VI), at low concentration levels is generally carried out using the hydride-generation technique. Since only Se(IV) is able to generate the hydride, if speciation data are required two procedures are usually used. In the first procedure, the sample has first to be analyzed without pretreatment, which provides the Se(IV) concentration. Then the sample is treated by heating it with 6 mol L^{-1} hydrochloric acid, which reduces any Se(VI) to Se(IV). The analysis is performed again, to give a total concentration of inorganic selenium present in the sample. The concentration of Se(VI) is then calculated by difference. The obvious disadvantage of this method is its indirect determination of Se(VI), which involves the errors induced by determination of Se(IV) and total inorganic selenium.[106] The speciation of Se(IV) and Se(VI) can also be carried out using HPLC separation followed by detection with different detectors.[106,107] Pitts et al.[99] recently described an on-line method for determination of inorganic selenium species in aqueous samples using HPLC/HG/AFS. Separation of these species was achieved using HPLC, after which the analyte was acidified, and passed through an on-line microwave system. The latter stage left the Se(IV) unaltered, but transformed Se(VI) to Se(IV). The selenium hydride was then determined using an AFS detector. DLs of 0.2 and 0.3 μg L^{-1} were reported for Se(IV) and Se(VI), respectively.

3.2.3 Other Hydride-forming Elements

In addition to arsenic and selenium, other elements such as antimony, bismuth, tellurium, germanium, tin, and lead have also been determined using HG/AFS. It should be noted that the arrangements of the instruments used for the analysis of these elements are identical with those for arsenic and selenium except for the utilization of a corresponding excitation source and wavelength-selection system. Both dispersive and nondispersive AFS have been used.[56,97] Kobayashi et al.[108] published the first paper on the determination of bismuth using an HG/AFS system. They used an argon–hydrogen diffusion flame as the atomizer and a microwave-excited EDL as the excitation source, and achieved a DL of 5 μg L^{-1} or 0.1 ng of bismuth. Nakahara et al.[109]

described an analytical method based on HG/AFS for the determination of tellurium using an argon–hydrogen diffusion flame and an EDL excitation source. The comparison of the Zn and $NaBH_4$ reduction methods was discussed. The best attainable DLs for tellurium were 0.1 μg L^{-1} (2 ng) and 1.5 μg L^{-1} (30 ng).

Nakahara et al. have also conducted a series of studies on the determination of other hydride-forming elements[110,111] using hydride generation coupled with nondispersive AFS. D'Ulivo and Papoff[112] described a method using nondispersive AFS combined with hydride generation for the determination of lead. They used a radio-frequency-excited EDL as the excitation source and a small argon–hydrogen flame as the atomizer. The DL was 0.06 μg L^{-1} and the calibration curve was linear up to 300 μg L^{-1}, superior to HG/AAS or inductively coupled plasma atomic emission spectrometry (ICP/AES). Using a similar HG/AFS system, D'Ulivo et al.[32] determined dialkyllead and trialkyllead. The DLs obtained for trimethyl, triethyl, dimethyl, and diethyllead (3σ) were 3–5 ng L^{-1}. Although HG/AFS is an appropriate analytical technique for the determination of these hydride-forming elements, its application in environmental analysis seems to be very limited. This is due, in part, to the fact that until recently no commercial HG/AFS instruments were available. Most of the research was carried out using laboratory-made systems. However, a continuous-flow HG/AFS system, recently made commercially available, provides a capability for arsenic, selenium, antimony, tellurium, and bismuth determination with a DL in the sub-parts per billion region.[12,14,56]

3.3 Non-hydride-generation Methods

3.3.1 Laser-induced Fluorescence

Determinations of hydride-forming elements in environmental samples are usually carried out using HG/AFS because of its high sensitivity and simplicity. Another very sensitive type of AFS, which has been used for the determination of many elements (both hydride-forming and non-hydride-forming), is LIF. LIF provides the lowest absolute DLs for a number of elements.[23]

Although different types of atomizers, such as flame, plasma, and electrothermal, can be coupled with LIF, it appears that the most sensitive practical measurements by LIF are performed in those ETA instruments, such as graphite-tube furnaces, that are currently used in AAS. With an ETA as atomizer, absolute DLs below femtogram level can be obtained experimentally for some elements. Omenetto[23] collected pertinent data for 22 elements investigated with LIF/ETA and the DLs were compared with those given by inductively coupled plasma mass spectrometry (ICP/MS). For most elements, the data obtained using LIF/ETA are superior to those obtained using ICP/MS. Recently, a DL of 10 fg mL^{-1} (100 ag absolute) was reported for lead determination in whole-blood samples using LIF/ETA.[19] A schematic diagram of the experimental design for LIF/ETA is shown in Figure 12.

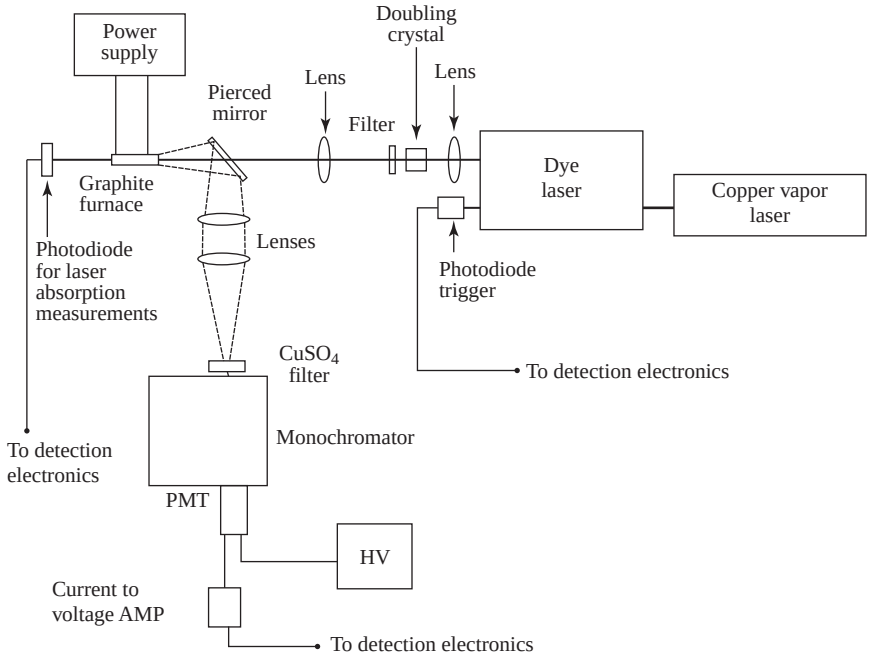

Figure 12 Experimental design for LIF/ETA. (Reprinted with permission from E.P. Wagner II, B.W. Smith, J.D. Winefordner, *Anal. Chem.*, **68**, 3199 (1996). Copyright 1996 American Chemical Society.)

LIF/ETA has been used for the determination of a number of elements in various environmental samples. Bolshov et al.[24] reported the results for the determination of cadmium in Antarctic and Greenland snow and ice by LIF/ETA. The DL was found to be 0.5 fg or 0.01 pg g^{-1} using a 50-μL sample volume. Silver was determined in seawater by LIF using novel diffusive graphite-tube ETA.[113] A 10-μL volume of sample was applied to a small graphite boat attached to one of the graphite electrodes and inserted into the graphite tube, which was then sealed by the electrodes and heated. The vaporized silver passed through the graphite wall and was excited by a laser beam a few millimeters above the surface of the tube. A DL of 90 fg for 1 : 1 diluted seawater was obtained. Other examples of recent applications of LIF/ETA include the determination of aluminum and lead[114] in atmospheric aerosol samples, antimony[40] in biological and environmental samples, bismuth,[115] cobalt,[116] and lead[25] in seawater, thallium[117] and lead[118] in natural water, and mercury[119] in soil.

LIF coupled with flame atomization has also been used for environmental sample analysis. Zhou et al.[21] described methods for the determination of cobalt, copper, lead, manganese, and thallium in Buffalo River sediment using flame LIF. A standard, long-path flame atomic absorption burner was used and DLs for cobalt (2 ng mL^{-1}), copper (2 ng mL^{-1}), lead (0.4 ng mL^{-1}), and thallium (0.9 ng mL^{-1}) were reported.

One of the major problems facing the future use of LIF instruments is that lasers are difficult to use. They are costly and require highly skilled operators. This certainly limits the development of commercially available instruments.

3.3.2 Atomic Fluorescence Spectrometry as a Chromatographic Detection Method

AFS can be conveniently used as a detection method for chromatography, such as GC and HPLC. Details can be found in an excellent paper by D'Ulivo,[7] who summarized the research regarding the utilization of AFS in chromatographic detection. Brief information is given below.

Van Loon et al.[120] first described a method for the simultaneous speciation of several elements using an atomic fluorescence spectrometer as an element-specific detector for chromatography. A three-channel nondispersive atomic fluorescence spectrometer with an HCL as the excitation source and a nitrogen-sheathed air–acetylene flame as the atomizer was used. It was interfaced to an HPLC system by direct connection of the column outflow to the nebulizer capillary of the burner. The column flow rate was compatible with the nebulizer flow rate. The technique was successfully applied to the simultaneous detection of several metals as their EDTA (copper, zinc, nickel), glycine (copper, zinc, nickel), and trien complexes. More research has been carried out to study metal speciation using different models of coupling of HPLC with AFS.[121]

Coupling AFS to GC greatly simplifies the task of the atomizer since the effluents supplied to the AFS system are already in the vapor phase.[7] Different types of atomizers have been designed and applied to the speciation of metals.[122] D'Ulivo and Papoff[123] described a technique using GC coupled with multichannel nondispersive AFS for the simultaneous determination of alkyllead, alkyltin, and alkylselenide compounds. Recently, Pécheyran et al.[124] described a GC/AFS method for the determination of ultratrace volatile selenium species in aqueous solutions. The method, optimized for field measurements of volatile selenium species in aquatic environments, involved a shipboad purge of the water samples collected under clean conditions, a cryogenic trapping stage followed by separation of the analytes by low-temperature chromatography. Under routine operating conditions, absolute DLs of 4 and 4.5 pg of Se for Me_2Se and Me_2Se_2 were obtained, respectively.

It is expected that recent developments will strengthen the role of commercial AFS in the determination of trace elements in environmental samples. Indeed, AFS has already been shown to be the most sensitive and important detection technique for mercury determination, especially for samples with low concentration levels.

4 CONCLUSIONS

Clearly, the most important and widely used AFS systems in environmental analysis have so far been CVAFS for mercury and HG/AFS for the determination of the hydride-forming elements. In practical terms, for environmental and other areas of analysis, the broader adoption of AFS will depend on the introduction of standardized commercial instrumentation and the ability to compete with other techniques that are commonly used, such as AAS, AES, and ICP/MS. Currently, several companies manufacture AFS detectors used either for mercury or hydride-forming elements, or both. These AFS detectors are of simple design and easy operation. Most importantly, these cost-effective instruments provide excellent DLs for some environmentally important elements, such as mercury. Important progress is, therefore, expected in using these instruments for environmental analysis.

Although LIF has practical limitations at present, such as high cost and difficult operation, the exceptional low DLs achieved using LIF for some elements are very attractive. This is especially true when only small sample amounts are available.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the encouragement given to me during preparation of this article by Dr Rudolf Jaffé. Special thanks are due to Anita Holloway for her assistance in the preparation of the manuscript. I wish to express my thanks to the many authors and publishers who have given permission for the reproduction of figures from the original literature. This is SERP contribution number 106.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AES	Atomic Emission Spectroscopy
AFS	Atomic Fluorescence Spectroscopy
BDHCL	Boosted Discharge Hollow-cathode Lamp
CVAAS	Cold-vapor Atomic Absorption Spectrometry
CVAFS	Cold-vapor Atomic Fluorescence Spectrometry
DL	Detection Limit
DMAA	Dimethylarsinic Acid
EDL	Electrodeless Discharge Lamps
ETA	Electrothermal Atomization
FI	Flow-injection
GC	Gas Chromatography
GC/AFS	Gas Chromatography/Atomic Fluorescence Spectroscopy
GC/CVAFS	Gas Chromatography/Cold-vapor Atomic Fluorescence Spectrometry
GC/ECD	Gas Chromatography/Electron Capture Detection
HCL	Hollow-cathode Lamp
HG/AFS	Hydride-generation/Atomic Fluorescence Spectroscopy
HPLC	High-performance Liquid Chromatography
HPLC/HG/AFS	High-performance Liquid Chromatography/Hydride-generation/Atomic Fluorescence Spectrometry
ICP	Inductively Coupled Plasma
ICP/AES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
LIF	Laser-induced Fluorescence
LIF/ETA	Laser-induced Fluorescence/ Electrothermal Atomization
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
MMAA	Monomethylarsonic Acid
NIST	National Institute of Standards and Technology
OPO	Optical Parametric Oscillator
PMT	Photomultiplier Tube
S/N	Signal-to-noise Ratio
SPME	Solid–phase Microextraction
SRM	Standard Reference Material
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
YAG	Yttrium Aluminum Garnet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Organometallic Compound Analysis in Environmental Samples • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Solid-phase Microextraction in Environmental Analysis

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis • Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

REFERENCES

1. V. Sychra, V. Svoboda, I. Rubeška, *Atomic Fluorescence Spectroscopy*, Van Nostrand Reinhold, London, 1975.
2. G.F. Kirkbright, M. Sargent, *Atomic Absorption and Fluorescence Spectroscopy*, Academic Press, London, 1974.
3. J.D. Winefordner, R.C. Elser, 'Atomic Fluorescence Spectrometry', *Anal. Chem.*, **43**(4), 24A–39A (1971).
4. J.D. Winefordner, 'Principles, Methodologies, and Applications of Atomic Fluorescence Spectrometry', *J. Chem. Educ.*, **55**(2), 72–78 (1981).
5. J.C. Van Loon, 'Atomic Fluorescence Spectrometry, Present Status and Future Prospects', *Anal. Chem.*, **53**(2), 332A–361A (1981).
6. J.D. Winefordner, T.J. Vickers, 'Atomic Fluorescence Spectrometry as a Means of Chemical Analysis', *Anal. Chem.*, **36**(1), 161–165 (1964).
7. A. D'Ulivo, 'Atomic Fluorescence Detectors', in *Environmental Analysis Using Chromatography Interfaced with Atomic Spectroscopy*, eds. R.M. Harrison, S. Rapsomanikis, Ellis Horwood, Chichester, 127–164, 1989.
8. N. Omenetto, J.D. Winefordner, 'Atomic Fluorescence Spectrometry, Basic Principles and Applications', *Prog. Anal. At. Spectrosc.*, **2**, 1–183 (1979).
9. C.Th.J. Alkemade, P.J.Th. Zeegers, 'Excitation and Deexcitation Processes in Flames,' in *Spectrochemical Methods of Analysis*, ed. J.D. Wineforder, Wiley-Interscience, New York, 3–125, 1971.
10. D.A. Skoog, F.J. Holler, T.A. Nieman, *Principles of Instrumental Analysis*, 5th edition, Saunders College Publishing, New York, 1998.
11. C.A. Watson, 'Use of Improved Design and Boosted-discharge Hollow-cathode Lamps for Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **3**, 407–414 (1988).
12. W.T. Corns, P.B. Stockwell, L. Ebdon, S.J. Hill, 'Development of an Atomic Fluorescence Spectrometer for the Hydride-forming Elements', *J. Anal. At. Spectrom.*, **8**, 71–77 (1993).
13. W.B. Barnett, H.L. Kahn, 'Comparison of Atomic Fluorescence with Atomic Absorption as an Analytical Technique', *Anal. Chem.*, **44**(6), 935–939 (1972).
14. *PSA 10.055 Millennium Excalibur, User Manual*, Issue 1.0, P.S. Analytical, Sevenoaks, March, 1998.
15. J.V. Sullivan, A. Walsh, 'High-intensity Hollow-cathode Lamps', *Spectrochim. Acta*, **21**, 721–726 (1965).
16. R.M. Lowe, 'A High-intensity Hollow-cathode Lamp for Atomic Fluorescence', *Spectrochim. Acta, Part B*, **26**, 201–205 (1971).
17. J.W. Novak, Jr, R.F. Browne, 'Pulsed Radio-frequency Electrodeless Discharge Lamps for Atomic Absorption and Atomic Fluorescence Spectrometry', *Anal. Chem.*, **50**, 1453–1457 (1978).
18. N. Omenetto, B.W. Smith, J.D. Winefordner, 'Laser-induced Fluorescence and Electrothermal Atomization: How Far from the Intrinsic Limit of Detection?', *Spectrochim. Acta, Part B*, **43**(9), 1111–1118 (1988).
19. E.P. Wagner II, B.W. Smith, J.D. Winefordner, 'Ultratrace Determination of Lead in Whole Blood Using Electrothermal Atomization Laser-excited Atomic Fluorescence Spectrometry', *Anal. Chem.*, **68**, 3199–3203 (1996).
20. R.F. Lonardo, A.I. Yuzefovsky, R.L. Irwin, R.G. Michel, 'Laser-excited Atomic Fluorescence Spectrometry in a Pressure-controlled Electrothermal Atomizer', *Anal. Chem.*, **68**(3), 514–521 (1996).
21. J.X. Zhou, X. Hou, S.-J.J. Tsai, K.X. Yang, R.G. Michel, 'Characterization of Tunable Optical Parametric Oscillator Laser System for Multielement Flame Laser Excited Atomic Fluorescence Spectrometry of Cobalt, Copper, Lead, Manganese, and Thallium in Buffalo River Sediment', *Anal. Chem.*, **69**(3), 490–499 (1997).
22. D.J. Swart, M. Ezer, H.L. Pacquette, J.B. Simeonsson, 'Laser-induced Fluorescence of Se, As, and Sb in an Electrothermal Atomizer', *Anal. Chem.*, **70**, 1324–1330 (1998).
23. N. Omenetto, 'Laser-induced Atomic Fluorescence Spectroscopy: Personal Viewpoint on Its Status, Needs and Perspectives', *Spectrochim. Acta, Part B*, **44**(2), 131–146 (1989).
24. M.A. Bolshov, C.F. Boutron, F.M. Ducroz, U. Gorlach, O.N. Kompanetz, S.N. Rudniev, B. Hutch, 'Direct Ultratrace Determination of Cadmium in Antarctic and Greenland Snow and Ice by Laser Atomic Fluorescence Spectrometry', *Anal. Chim. Acta*, **251**, 169–175 (1991).
25. V. Cheam, J. Lechner, I. Sekerka, R. Desrosiers, 'Direct Determination of Lead in Seawater by Laser-excited Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **9**, 315–320 (1994).
26. B.W. Smith, J.B. Womack, N. Omenetto, J.D. Winefordner, 'Approaching Single Atom Detection with Atomic Fluorescence in a Glow Discharge Atomic Reservoir', *Appl. Spectrosc.*, **43**(5), 873–876 (1989).
27. *PSA 10.025 Millennium Merlin, User Manual*, Issue 1.0, P.S. Analytical, Sevenoaks, October, 1997.
28. C. Vandecasteel, C.B. Block, *Modern Methods for Trace Element Determination*, John Wiley & Sons, Chichester, 1993.
29. K. Tsujii, K. Kuga, 'Improvement in the Nondispersive Atomic Fluorescence Spectrometric Detection of Arsenic and Antimony by a Hydride-generation Technique', *Anal. Chim. Acta*, **97**, 51–57 (1978).
30. K. Tsujii, E. Kitazume, K. Yagi, 'Determination of Antimony Depth Profile in Semiconductor Silicon by Chemical Etching and Nondispersive Atomic Fluorescence

Spectrometry with Hydride Generation', *Anal. Chim. Acta*, **128**, 229–234 (1981).
31. A. D'Ulivo, P. Papoff, C. Festa, 'A Simultaneous Multielement Nondispersive Atomic Fluorescence Spectrometer Using Modulated Sources and Frequency Discrimination of Fluorescence Signals', *Talanta*, **30**, 907–913 (1983).
32. A. D'Ulivo, R. Fuoco, P. Papoff, 'Behavior of Some Dialkyl- and Trialkyl-lead Compounds in the Hydride-generation Procedure Using a Nondispersive Atomic Fluorescence Detector', *Talanta*, **33**, 401–405 (1986).
33. B. Welz, M. Melcher, 'Investigations on Atomization Mechanisms of Volatile Hydride-forming Elements in a Heated Quartz Cell', *Analyst*, **108**, 213–224 (1983).
34. B. Welz, M. Schubert-Jacobs, 'Investigation on Atomization Mechanisms in Hydride-generation Atomic Absorption Spectrometry', *Fresenius' Z. Anal. Chem.*, **324**, 832–838 (1986).
35. J. Dedina, I. Rubeška, 'Hydride Atomization in a Cool Hydrogen–Oxygen Flame Burning in a Quartz Tube Atomizer', *Spectrochim. Acta, Part B*, **35**, 119–128 (1980).
36. J. Dedina, B. Welz, 'Quartz Tube Atomizers for Hydride-generation Atomic Absorption Spectrometry: Mechanism for Atomization of Arsine', *J. Anal. At. Spectrom.*, **7**, 307–314 (1992).
37. M.A. Bolshov, A.V. Zybin, I.I. Smirenkina, 'Atomic Fluorescence Spectrometry with Laser Excitation', *Spectrochim. Acta, Part B*, **36**, 1143–1152 (1981).
38. D. Goforth, J.D. Winefordner, 'Laser-excited Atomic Fluorescence of Atoms Produced in a Graphite Furnace', *Anal. Chem.*, **58**, 2598–2602 (1986).
39. M. Leong, J. Vera, B.W. Smith, N. Omenetto, J.D. Winefordner, 'Laser-induced Double Resonance Fluorescence of Lead with Graphite-tube Atomization', *Anal. Chem.*, **60**, 1605–1610 (1988).
40. J. Enger, A. Marunkov, N. Chekalin, O. Axner, 'Direct Detection of Antimony in Environmental and Biological Samples at Trace Concentrations by Laser-induced Fluorescence in Graphite Furnace with an Intensified Charge Coupled Device', *J. Anal. At. Spectrom.*, **10**, 539–549 (1994).
41. H. Massmann, 'Vergleich von Atomabsorption und Atomfluoreszenz in der Graphitküvette', *Spectrochim. Acta, Part B*, **23**, 215–226 (1968).
42. S. Neumann, M. Kriese, 'Sub-picogram Detection of Lead by Non-flame Atomic Fluorescence Spectrometry with Dye Laser Excitation', *Spectrochim. Acta, Part B*, **29**, 127–137 (1974).
43. H.G.C. Human, N. Omenetto, P. Cavalli, G. Rossi, 'Laser Excited Analytical Atomic and Ionic Fluorescence in Flames, Furnaces, and Inductively Coupled Plasmas – II. Fluorescence Characteristics and Detection Limits for Fourteen Elements', *Spectrochim. Acta, Part B*, **39**, 1345–1363 (1984).
44. R.E. Sturgeon, C.L. Chakrabarti, 'Evaluation of Pyrolytic-graphite-coated Tubes for Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chem.*, **49**, 90–97 (1977).
45. M.A. Bolshov, A.V. Zybin, V.G. Koloshnikov, I.I. Smirenkina, 'Analytical Applications of a LEAFS–ETA Method', *Spectrochim. Acta, Part B*, **43**, 519–528 (1988).
46. B. Welz, *Atomic Absorption Spectrometry*, 2nd edition, VCH, Weinheim, 1985.
47. A. Montaser, V.A. Fassel, 'Inductively Coupled Plasma as Atomization cells for Atomic Fluorescence Spectrometry', *Anal. Chem.*, **48**, 1490–1499 (1976).
48. M.B. Leong, A.P. D'Silva, V.A. Fassel, 'Evaluation of Stimulated Raman Scattering of Tunable Dye Laser Radiation as a Primary Excitation Source for Exciting Atomic Fluorescence in an Inductively Coupled Plasma', *Anal. Chem.*, **58**, 2594–2598 (1986).
49. W. Hang, W.O. Walden, W.W. Harrison, 'Microsecond Pulsed Glow Discharge as an Analytical Spectroscopic Source', *Anal. Chem.*, **68**, 1148–1152 (1996).
50. K.C. Thompson, G.D. Reynolds, 'The Atomic Fluorescence Determination of Mercury by the Cold-vapor Technique', *Analyst*, **96**, 771–775 (1971).
51. V.I. Muscat, T.J. Vickers, 'Determination of Nanogram Quantities of Mercury by the Reduction–Aeration Method and Atomic Fluorescence Spectrometry', *Anal. Chim. Acta*, **57**, 23–30 (1971).
52. V.I. Muscat, T.J. Vickers, A. Andren, 'A Simple and Versatile Atomic Fluorescence System for the Determination of Nanogram Quantities of Mercury', *Anal. Chem.*, **44**, 218–221 (1972).
53. C.D. West, 'Relative Effect of Molecular Absorption on Atomic Absorption and Atomic Fluorescence', *Anal. Chem.*, **46**, 797–799 (1974).
54. K.C. Thompson, R.G. Godden, 'Improvement in the Atomic Fluorescence Determination of Mercury by the Cold-vapor Technique', *Analyst*, **100**, 544–548 (1975).
55. R.G. Godden, P.B. Stockwell, 'Atomic Fluorescence Spectrometric Determination of Mercury Using a Filter Fluorimeter', *J. Anal. At. Spectrom.*, **4**, 301–303 (1989).
56. P.B. Stockwell, W.T. Corns, 'Environmental Sensors Based on Atomic Fluorescence', *Analyst*, **119**, 1641–1645 (1994).
57. N. Bloom, W.F. Fitzgerald, 'Determination of Volatile Mercury Species at the Picogram Level by Low-temperature Gas Chromatography with Cold-vapor Atomic Fluorescence Detection', *Anal. Chim. Acta*, **208**, 151–161 (1988).
58. J. Olafsson, 'A Report on the ICES Intercalibration of Mercury in Seawater for the Joint Monitoring Group of the Oslo and Paris Commissions', *Mar. Chem.*, **11**, 129–142 (1982).
59. O. Szakácss, A. Lásztity, Zs. Horváth, 'Breakdown of Organic Mercury Compounds by Hydrochloric Acid–Permanganate or Bromine Monochloride Solution for

the Determination of Mercury by Cold-vapor Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **121**, 219–224 (1980).
60. R.D. Jones, M.E. Jacobson, R. Jaffé, J. West-Thomas, C. Arfstrom, A. Alli, 'Method Development and Sample Processing of Water, Soil, and Tissue for the Analysis of Total and Organic Mercury by Cold-vapor Atomic Fluorescence Spectrometry', *Water Air Soil Pollut.*, **80**, 1285–1294 (1995).
61. J.P. Hurley, J.N. Benoit, C.L. Babiarz, M.M. Shafer, A.W. Andren, J.R. Sullivan, R. Hammond, D.A. Webb, 'Influences of Watershed Characteristics on Mercury Levels in Wisconsin Rivers', *Environ. Sci. Technol.*, **29**, 1867–1875 (1995).
62. C. Gagnon, E. Pelletier, A. Mucci, 'Behavior of Anthropogenic Mercury in Coastal Marine Sediments', *Mar. Chem.*, **59**, 159–176 (1997).
63. R. Jaffé, Y. Cai, J. West-Thomas, M. Morales, R.D. Jones, 'Occurrence of Methylmercury in Lake Valencia', *Bull. Environ. Contam. Toxicol.*, **59**, 99–105 (1997).
64. R.P. Mason, K.A. Sullivan, 'Mercury in Lake Michigan', *Environ. Sci. Technol.*, **31**, 942–947 (1997).
65. W. Van Delft, G. Vos, 'Comparison of Digestion Procedures for the Determination of Mercury in Soils by Cold-vapor Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **209**, 147–156 (1988).
66. R.P. Swift, J.E. Campbell, 'An Atomic Fluorescence-based Method for Ultratrace Mercury Detection in Environmental Samples', *Spectroscopy*, **8**(2), 38–46 (1993).
67. R.D. Jones, M.E. Jacobson, R. Jaffe, J. West-Thomas, C. Arfstrom, A. Alli, 'Method Development and Sample Processing of Water, Soil, and Tissue for the Analysis of Total and Organic Mercury by Cold-vapor Atomic Fluorescence Spectrometry', *Water Air Soil Pollut.*, **80**, 1285–1294 (1995).
68. W.F. Fitzgerald, G.W. Gill, 'Sub-nanogram Determination of Mercury by Two-stage Gold Amalgamation and Gas Phase Detection Applied to Atmospheric Analysis', *Anal. Chem.*, **15**, 1714–1720 (1979).
69. G.A. Gill, K.W. Bruland, 'Mercury Speciation in Surface Freshwater Systems in California and Other Areas', *Environ. Sci. Technol.*, **24**, 1392–1400 (1990).
70. W. Wang, C.T. Driscoll, 'Patterns of Total Mercury Concentration in Onondaga Lake, New York', *Environ. Sci. Technol.*, **29**, 2261–2266 (1995).
71. USEPA, 'Method 1631: Mercury in Water by Oxidation, Purge and Trap, and Cold-vapor Atomic Fluorescence Spectrometry', http/www.epa.gov/OST/Methods/mercury.html, 1998.
72. W. Resto, R.G. Badini, B.W. Smith, C.L. Stevenson, J.D. Winefordner, 'Two-step Laser Excited Atomic Fluorescence Spectrometry Determination of Mercury', *Spectrochim. Acta, Part B*, **48**, 627–632 (1993).
73. W. Jian, C.W. Mcleod, 'Rapid Sequential Determination of Inorganic Mercury and Methylmercury in Natural Waters by Flow-injection–Cold-vapor Atomic Fluorescence Spectrometry', *Talanta*, **39**, 1537–1542 (1992).
74. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Determination of Mercury in Filtered Sea-water by Flow-injection with On-line Oxidation and Atomic Fluorescence Spectrometric Detection', *J. Anal. At. Spectrom.*, **11**, 511–514 (1996).
75. G. Westöö, 'Determination of Methylmercury Compounds in Foodstuffs, II. Determination of Methylmercury in Fish, Egg, Meat, and Liver', *Acta Chem. Scand.*, **21**, 1790–1800 (1967).
76. A. Alli, R. Jaffe, R. Jones, 'Analysis of Organomercury Compounds in Sediments by Capillary GC with Atomic Fluorescence Detection', *J. High Resolut. Chromatogr.*, **17**, 745–748 (1994).
77. Y. Cai, R. Jaffé, A. Alli, R. Jones, 'Determination of Organomercury Compounds in Aqueous Samples by Capillary Gas Chromatography–Atomic Fluorescence Spectrometry following Solid-phase Extraction', *Anal. Chim. Acta*, **334**, 251–259 (1996).
78. Y. Cai, R. Jaffé, R. Jones, 'Ethylmercury in the Soils and Sediments of the Florida Everglades', *Environ. Sci. Technol.*, **31**, 302–305 (1997).
79. Y. Cai, G. Tang, R. Jaffé, R. Jones, 'Evaluation of Some Isolation Methods for Organomercury Determination in Soil and Fish Samples by Capillary Gas Chromatography–Atomic Fluorescence Spectrometry', *Int. J. Environ. Anal. Chem.*, **68**, 331–345 (1997).
80. S. Rapsomanikis, 'Derivatization by Ethylation with Sodium Tetraethylborate for the Speciation of Metals and Organometallics in Environmental Samples', *Analyst*, **119**, 1429–1439 (1994).
81. Y. Cai, S. Monsulad, K.G. Furton, R. Jaffé, R. Jones, 'Determination of Methylmercury in Fish and Aqueous Samples Using Solid-phase Microextraction Followed by Gas Chromatography–Atomic Fluorescence Spectrometry', *Appl. Organomet. Chem.*, **12**, 565–569 (1998).
82. G. Hu, X. Wang, Y. Wang, X. Chen, L. Jia, 'Determination of Methylmercury in Waters Using Sodium Tetraphenylborate Derivatization/Solvent Extraction and Gas Chromatography–Ion Trap Mass Spectrometry', *Anal. Lett.*, **30**(14), 2579–2594 (1997).
83. T. De Smaele, L. Moens, R. Dams, P. Sandra, J. Van Der Eycken, J. Vandyck, 'Sodium Tetra(*n*-propyl)borate: a Novel Aqueous In Situ Derivatization Reagent for the Simultaneous Determination of Organomercury, -lead, and -tin Compounds with Capillary Gas Chromatography–Inductively Coupled Plasma Mass Spectrometry', *J. Chromatogr. A*, **793**, 99–106 (1998).
84. Y. Cai, unpublished data.
85. M. Horvat, L. Liang, N.S. Bloom, 'Comparison of Distillation with Other Current Isolation Methods for the Determination of Methylmercury Compounds in Low Level Environmental Samples ', *Anal. Chim. Acta*, **282**, 153–168 (1993).

86. S. Rapsomanikis, O.F. Donard, J.H. Weber, 'Speciation of Lead and Methyllead Ions in Water by Chromatography/Atomic Absorption Spectrometry after Ethylation with Sodium Tetraethylborate', *Anal. Chem.*, **58**, 35–38 (1986).
87. N. Bloom, 'Determination of Picogram Levels of Methylmercury by Aqueous Phase Ethylation, Followed by Cryogenic Gas Chromatography with Cold-vapor Atomic Fluorescence Detection', *Can. J. Fish Aquat. Sci.*, **46**, 1131–1140 (1989).
88. L. Liang, M. Horvat, N.S. Bloom, 'An Improved Speciation Method for Mercury by GC/CVAFS after Aqueous Phase Ethylation and Room Temperature Precollection', *Talanta*, **41**, 371–379 (1994).
89. R. Fischer, S. Rapsomanikis, M.O. Andreae, 'Determination of Methylmercury in Fish Samples Using GC/AA and Sodium Tetraethylborate Derivatization', *Anal. Chem.*, **65**, 763–766 (1993).
90. L. Liang, M. Horvat, E. Cernichiari, B. Gelein, S. Balogh, 'Simple Solvent Extraction Technique for Elimination of Matrix Interferences in the Determination of Methylmercury in Environmental and Biological samples by Ethylation–Gas Chromatography–Cold-vapor Atomic Fluorescence Spectrometry', *Talanta*, **43**, 1883–1888 (1996).
91. M. Filippelli, F. Baldi, F.E. Brinckmann, G.J. Olson, 'Methylmercury Determination as Volatile Methylmercury Hydride by Purge and Trap Gas Chromatography in Line with Fourier Transform Infrared Spectroscopy', *Environ. Sci. Technol.*, **26**, 1457–1460 (1992).
92. P.J. Craig, D. Mennie, N. Ostah, O.F.X. Donard, F. Martin, 'Novel Methods for Derivatization of Mercury(II) and Methylmercury Compounds for Analysis', *Analyst*, **117**, 823–824 (1992).
93. M.A. Morrison, J.H. Weber, 'Comparison of Digestion Media for Speciation of Mercury in the Seagrass *Zostera marina* L. Followed by Quantitation by Hydride-generation–Atomic Fluorescence Spectrometry', *Environ. Sci. Technol.*, **31**, 3325–3329 (1997).
94. H. Hintelmann, R.D. Wilken, 'The Analysis of Organic Mercury Compounds Using Liquid Chromatography with On-line Atomic Fluorescence Spectrometric Detection', *Appl. Organomet. Chem.*, **7**, 173–180 (1993).
95. W. Holak, 'Gas-sample Technique for Arsenic Determination by Atomic Absorption Spectrometry', *Anal. Chem.*, **41**, 1712–1713 (1969).
96. K. Tsujii, K. Kuga, 'The Determination of Arsenic by Nondispersive Atomic Fluorescence Spectrometry with a Gas Sampling Technique', *Anal. Chim. Acta*, **72**, 85–90 (1974).
97. K.C. Thompson, 'The Atomic Fluorescence Determination of Antimony, Arsenic, Selenium and Tellurium by Using the Hydride-generation Technique', *Analyst*, **100**, 307–310 (1975).
98. J.L. Gomez-Ariza, D. Sanchez-Rodas, R. Beltran, W. Corns, P. Stockwell, 'Evaluation of Atomic Fluorescence Spectrometry as a Sensitive Detection Technique for Arsenic Speciation', *Appl. Organomet. Chem.*, **12**, 1–9 (1998).
99. L. Pitts, A. Fisher, P. Worsfold, S.J. Hill, 'Selenium Speciation Using High-performance Liquid Chromatography–Hydride-generation Atomic Fluorescence with On-line Microwave Reduction', *J. Anal. At. Spectrom.*, **10**, 519–520 (1995).
100. X.C. Le, M.H. Ma, N.A. Wong, 'Speciation of Arsenic Compounds Using High-performance Liquid Chromatography at Elevated Temperature and Selective Hydride-generation Atomic Fluorescence Detection', *Anal. Chem.*, **68**(24), 4501–4506 (1996).
101. Z. Šlejkovec, J.T. van Elteren, A.R. Byrne, 'A Dual Arsenic Speciation System Combining Liquid Chromatographic and Purge and Trap–Gas Chromatographic Separation with Atomic Fluorescence Spectrometric Detection', *Anal. Chim. Acta*, **358**, 51–60 (1998).
102. J. Azad, G.F. Kirkbright, R.D. Snook, 'Determination of Arsenic in Soil Digests by Nondispersive Atomic Fluorescence Spectrometry Using an Argon–Hydrogen Flame and the Hydride-generation Technique', *Analyst*, **105**, 79–83 (1980).
103. R.S. Braman, C.C. Foreback, 'Methylated Forms of Arsenic in the Environment', *Science*, **182**, 1247–1249 (1973).
104. R.M. Dagnall, K.C. Thompson, T.S. West, 'Studies in Atomic Fluorescence Spectroscopy – IV: The Atomic Fluorescence Spectroscopic Determination of Selenium and Tellurium', *Talanta*, **14**, 557–563 (1967).
105. J. Azad, G.F. Kirkbright, R.D. Snook, 'Determination of Selenium in Soil Digests by Nondispersive Atomic Fluorescence Spectrometry Using an Argon–Hydrogen Flame and the Hydride-generation Technique', *Analyst*, **104**, 232–240 (1980).
106. R. Muñoz, O.F.X. Donard, C. Cámara, Ph. Quevauviller, 'Analytical Techniques Applied to the Speciation of Selenium in Environmental Matrices', *Anal. Chim. Acta*, **286**, 357–370 (1994).
107. Y. Cai, M. Cabañas, J.L. Fernández-Turiel, M. Abalos, J.M. Bayona, 'On-line Preconcentration of Selenium(IV) and Se(VI) in Aqueous Matrices Followed by Liquid Chromatography–Inductively Coupled Plasma Mass Spectrometry Determination', *Anal. Chim. Acta*, **314**, 183–192 (1995).
108. S. Kobayashi, T. Nakahara, S. Musha, 'Non-dispersive Atomic Fluorescence Spectrometry of Trace Amounts of Bismuth by Introduction of Its Gaseous Hydride into a Premixed Argon (Entrained Air)–Hydrogen Flame', *Talanta*, **26**, 951–957 (1979).
109. T. Nakahara, T. Wakisaka, S. Musha, 'The Determination of Tellurium by Non-dispersive Atomic Fluorescence Spectrometry Using the Hydride Generation Technique', *Spectrochim. Acta, Part B*, **36**(7), 661–670 (1981).

110. T. Nakahara, T. Wasa, 'Non-dispersive Atomic Fluorescence Spectrometric Determination of Lead by Hydride Generation', *Anal. Sci.*, **1**, 291–292 (1985).
111. T. Nakahara, T. Wasa, 'Determination of Tin by Non-dispersive Atomic Fluorescence Spectrometry Coupled with a Hydride-generation Technique', *J. Anal. At. Spectrom.*, **1**, 473–477 (1986).
112. A. D'Ulivo, P. Papoff, 'Non-dispersive Atomic Fluorescence Spectrometric Determination of Lead by Hydride-generation Technique', *Talanta*, **32**(5), 383–386 (1985).
113. I.B. Gornushkin, B.W. Smith, J.D. Winefordner, 'Use of Laser-excited Atomic Fluorescence Spectrometry with a Novel Diffusive Graphite Tube Electrothermal Atomizer for the Direct determination of Silver in Sea Water and in Solid Reference Materials', *Spectrochim. Acta, Part B*, **51**(11), 1355–1370 (1996).
114. J. Enger, Y. Malmsten, P. Ljungberg, O. Axner, 'Laser-induced Fluorescence in a Graphite Furnace as a Sensitive Technique for Assessment of Traces in North Arctic Atmospheric Aerosol Samples', *Analyst*, **120**, 635–641 (1995).
115. M.A. Bolshov, S.N. Rudnev, J. Brust, 'Analytical Characterization of Laser Excited Atomic Fluorescence of Bismuth', *Spectrochim. Acta, Part B*, **49**, 1437–1444 (1994).
116. A. Yuzefovsky, R.F. Lonardo, M. Wang, R.G. Machel, 'Determination of Ultra-trace Amounts of Cobalt in Ocean Water by Laser-excited Atomic Fluorescence Spectrometry in a Graphite Electrothermal Atomizer with Semi On-line Flow-injection Preconcentration', *J. Anal. At. Spectrom.*, **9**, 1195–1202 (1994).
117. O. Axner, N. Chekalin, P. Ljungberg, Y. Malmsten, 'Direct Determination of Thallium in Natural Waters by Laser-induced Fluorescence in a Graphite Furnace', *Int. J. Environ. Anal. Chem.*, **53**, 185–193 (1993).
118. N. Cheam, J. Lechner, R. Desrosiers, I. Sekerka, J. Nriagu, G. Lawson, 'Application of Laser-excited Atomic Fluorescence Spectrometer to Study Lead Distribution in Great Lakes Waters', *Int. J. Environ. Anal. Chem.*, **53**, 13–27 (1993).
119. S.T. Pagano, B.W. Smith, J.D. Winefordner, 'Determination of Mercury in Microwave-digested Soil by Laser-excited Atomic Fluorescence Spectrometry with Electrothermal Atomization', *Talanta*, **41**, 2073–2078 (1994).
120. J.C. Van Loon, J. Lichwa, B. Radziuk, 'Non-dispersive Atomic Fluorescence Spectrometry, a New Detector for Chromatography', *J. Chromatogr.*, **136**, 301–305 (1977).
121. D.J. Mackey, 'The Adsorption of Simple Trace Metal Cations on Amberlite XAD-1 and XAD-2, a Study Using a Multichannel Non-dispersive Atomic Fluorescence Detector with Quantitation by Batch Measurements', *J. Chromatogr.*, **236**, 81–95 (1982).
122. B. Radziuk, Y. Thompson, L.R.P. Butler, J.C. van Loon, Y.K. Chau, 'A Study of Atomic Absorption and Atomic Fluorescence Atomization Systems as Detectors in the Gas Chromatographic Determination of Lead', *Anal. Chim. Acta*, **108**, 31–38 (1979).
123. A. D'Ulivo, P. Pappoff, 'Simultaneous Detection of Alkylselenide, Alkyllead and Alkyltin Compounds by Gas Chromatography Using a Multi-channel Non-dispersive Atomic Fluorescence Spectrometric Detector and a Miniature Flame as the Atomizer', *J. Anal. At. Spectrom.*, **1**, 479–484 (1986).
124. C. Pécheyran, D. Amouroux, O.F.X. Donard, 'Field Determination of Volatile Selenium Species at Ultra Trace Levels in Environmental Waters by On-line Purging, Cryofocusing and Detection by Atomic Fluorescence Spectroscopy', *J. Anal. At. Spectrom.*, **13**, 615–621 (1998).

Biological Samples in Environmental Analysis: Preparation and Cleanup

Steven A. Barker
Louisiana State University, Baton Rouge, USA

1 Introduction: the Examination of Biological samples to Determine the Exposure and Impact of Pollutants **1**
1.1 The Exposure of Flora and Fauna to Pollutants 2
1.2 Biological Sentinels for Monitoring Pollution 3
1.3 Analytical Chemistry's Role in Responding to Environmental Pollution 4
2 Environmental Regulations and Methods: The Response of Governments to the Problems of Pollution **4**
2.1 The Development of Laws, Regulations and Regulatory Agencies for Monitoring Environmental Pollution 4
2.2 Early Analytical Efforts to Respond to Pollution Events 6
2.3 The Evolution of Analytical Methods into Regulatory Methods 6
3 Methods and their Applications **7**
3.1 Sample Preparation 7
3.2 Extraction and Cleanup 9
4 New Approaches to Sample Preparation and Cleanup for Environmental Analyses **14**
4.1 Gel Permeation Chromatography 14
4.2 Supercritical-fluid Extraction 14
4.3 Accelerated Solvent Extraction 15
4.4 Microwave-assisted Extraction 15
4.5 Solid-phase Extraction 16
4.6 Solid-phase Microextraction 18
4.7 Matrix Solid-phase Dispersion 19
5 Methods for the Future **20**
Acknowledgments **21**
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

The most difficult samples to analyze are those derived from biological sources. The analysis of biological samples requires the ability to extract and isolate a wide range of environmental pollutants from an even wider range of matrices, from simple bacteria to tissues and fluids from humans, from fungi to the complex tissues and structures of the plants and products that compose the food chain. These matrices comprise a variety of cells, connective tissues and fibers, fluids and related materials that introduce literally thousands of different organic and inorganic compounds to every analysis effort. In order to recognize and assess the nature and severity of pollution and its impact on humans as well as the flora and fauna that constitute our environment, highly efficient and selective methods for the extraction and cleanup of such samples must be developed. New developments in analytical science have been brought to bear on this problem, speeding the extraction and analysis process while making the resulting analyses more sensitive and selective. The present article examines the historical approach to the extraction and cleanup of biological samples for the analysis of pollutants, and reviews recent advances in technology and methodology that have proven applicable to this field.

1 INTRODUCTION: THE EXAMINATION OF BIOLOGICAL SAMPLES TO DETERMINE THE EXPOSURE AND IMPACT OF POLLUTANTS

The composition of biological samples ranges in complexity from the simplest liquids to the most intricate mammalian tissues. It encompasses everything from bacteria and viruses to plant to animal, ranging in development from single cell, spore, seed or egg to adult. The compounds that constitute the matrix of each type in this variety of potential samples for environmental analysis also run the gamut from simple molecules and elements to the most elaborate of proteins, carbohydrates and, of course, nuclear DNA. It is of interest to identify and measure these compounds as such; assaying these naturally occurring molecules of which biological materials are composed is task enough. However, humankind has added to this task the need to assay all of the other compounds we introduce into the natural process; agrochemicals (herbicides, pesticides), pollutants (polychlorinated biphenyls, (PCBs) polyaromatic hydrocarbons, (PAHs) etc.) and, for food animals, residues of veterinary drugs, all needles in a sample matrix haystack.

The history of all analytical science is mirrored in environmental chemistry, with new technologies being applied as they are developed for each of the needs for analysis in the field. Often the new technology has

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

been developed by chemists already in the area, applied in an effort to solve the many difficult analytical tasks with which they were, and continue to be, presented in performing environmental analyses in biological samples. Many of the problems faced in this and other areas of the analytical sciences are the same, however, and involve the need to perform adequate sample preparation, homogenization and disruption, sample/target analyte(s) isolation, analyte(s) resolution from co-extracted sample components and analyte(s) detection and quantitation.

To begin to analyze a biological sample for a compound of environmental interest, certain steps must be taken to assure the quality and reliability of the result. The value of this analytical result is directly related to the procedures used for sample preparation. Sample preparation is not merely the steps required in the homogenization, extraction and cleanup prior to introduction to an instrument. It also requires consideration of the level of documentation, appropriate procedures for sample collection, storage and shipment, assuring the stability of the sample and the target analytes and the processing of the sample necessary before extraction begins.

The methods applied for these processes require the development and thorough validation of the procedures to be employed. As well as this being simply good laboratory practice (GLP), it must also be noted that the results from environmental analyses are often used in litigation and may prove essential to making decisions concerning threats to the public health. Thus, regulatory authorities at many levels of government have established, through law and regulation, analytical procedures and methods that have been proven to be highly reliable. These ''official methods'' have withstood the test of time, as well as the courts, but have been modified over the years to incorporate new technologies and approaches. Such modifications must, however, prove to be equivalent to or better than the techniques they replace.

The trend in sample preparation and cleanup is away from the use of procedures that require large volumes of extracting solvent, large sample size and complicated and expensive glassware and/or instrumentation that have characterized ''classical'' methods for environmental analysis. Such methods are being replaced by techniques that require smaller sample sizes, offer generic extraction techniques and that can be readily automated to speed obtaining the results of an analysis. The field is further evolving to include the use of immunoassay, affinity chromatography, biosensing and other analytical advances to even further simplify or completely eliminate the processes previously required for sample preparation and cleanup. Such advances are being assisted by the new generation of analytical instruments that provide greater sensitivity, selectivity and speed of analysis.

The present article will address the general processes and principles involved in the preparation and cleanup of biological samples prior to analysis for environmentally relevant pollutants. It will address these issues from a historical perspective, examining ''official'' methods for the analysis of pollutants and discussing newer techniques that are slowly replacing more classical approaches. It is noted here that it is not possible to cover every biological matrix or environmental pollutant of interest and that each combination of target analyte and sample can present unique complications. Indeed, the analysis of thousands of environmental pollutants, ranging from alar on apples to PCBs in the hepatopancreas of crayfish, and the variety of biota, encompassing all five kingdoms that are exposed to them, constitute a major challenge to the analytical chemist. The methods developed to isolate, detect, identify and quantify these compounds are, however, essential to maintaining the health of the food supply and our environment and provide us with the capacity to identify, regulate and correct environmental hazards.

Much of the information, references and examples presented will be based on analyses conducted in aquatic species for environmental pollutants such as chlorinated pesticides, herbicides, PAHs and PCBs. Attention to single compounds and specialized protocols will not be addressed here. Nevertheless, the principles involved do address both the problems and the solutions to be encountered in dealing with biological samples in general in environmental analyses.

1.1 The Exposure of Flora and Fauna to Pollutants

Pollution is a relative term. Natural processes, such as forest fires, thermal venting, leaching of metals and minerals by groundwater and rain, and photochemical reactions in the atmosphere, are all processes that can contribute to alteration of the natural chemical state of a given ecosystem. This alteration of the environment can be seen as a pollution of the native molecular scheme. Changes in this chemical scheme can be measured relative to its initial state or relative to other areas to assess the degree of change.

The evolution of human technologies and the ever-increasing population of the planet have, however, added significantly to these natural processes. Humankind has also introduced many molecular species never before seen in the natural environment.

The use of fossil fuels, the industrialization of many nations' economies, the industrialization of agriculture, our struggles to handle our manufacturing and human wastes have spiraled as we have sought a higher standard of living through increased production. These efforts for our well-being have often been at the expense of the flora and fauna whose ancestors readily predated our own on

this planet. It has also been at our own peril. It has become a vicious cycle of the interactions of the environment's natural processes with humankind's introduction of unnatural substances, polluting and contaminating the cycle of life.

It is perhaps ironic that the justification for the original production and use of many herbicides, pesticides and other manufactured pollutants, such as their environmental or thermal stability, intended toxicity toward a particular pest, etc., eventually led to their becoming a continuing threat to the environment. For example, while the chlorinated pesticides were valuable for crop and animal agriculture, they also persisted in the environment, being highly stable to most aspects of environmental degradation. From the producers' point of view this was advantageous, reducing the frequency of application of the pesticide and the costs inherent in the process. However, as the toxicity of many of these compounds became evident, this persistence became the very cause for the removal of many such products from the market in the USA. Unfortunately, large quantities of many of these compounds, as well as others, have continued to be used in developing countries around the world.

Of further concern has been the level of these compounds that has accumulated in the environment from spraying, from legal and illegal dumping, and from sequestration in soils and sediments from which continuing diffusion of these materials into various habitats and resources can occur. Since many of these compounds are highly lipophilic, they tend to accumulate in fatty as well as other tissues of animals that compose the food chain. For aquatic environments, even coastal and deep-sea species have been found to contain detectable levels of many such compounds on a decreasing concentration gradient as one moves farther from land and away from river-influenced bays, gulfs and ocean currents.

The persistence of many classes of pollutants, their magnification through bioaccumulation and the concerns for their known and unknown toxicity require that state, federal and international monitoring programs provide mechanisms for assessment and remediation. However, state, national and international programs have not always proven to be capable of being integrated. This is in part due to the variety of biological matrices that one encounters when dealing with analyses of pollutants. It has also led to a degree of confusion in the field and has underscored the need for new methodology that can rapidly assess a pollution problem. This demand is further heightened by the fact that adequate monitoring will require the analysis of more samples for more compounds in shorter time frames than existing methods can currently handle.

The role of the analytical chemist in this situation is to provide methodology to identify and measure the compounds that we define as pollutants. Identification and measurement are the first steps in recognition and amelioration of the problem.

1.2 Biological Sentinels for Monitoring Pollution

We can readily acknowledge the exposure of the environment to contaminants. We may then also note that these environments are filled with native plants and animals, which have unwittingly offered themselves as sentinels to their own, and to our, exposure. Thus, the practice of analyzing biological matrices from the field for environmental contaminants has been used extensively as a general environmental monitoring strategy, accompanied by measurements in soils, sediments, water and air. We have extended this strategy by actually placing particular species in an area to further allow assessment of exposure under more controlled conditions. The use of such an approach, taking advantage of the ability of plants and animals to bioconcentrate pollutants, is defensible in the sense that we are most concerned about the impact of pollution on living beings and, most selfishly, ourselves. Furthermore, such efforts may be justified based on the fact that the levels of environmental contaminants in water are often too low to detect or quantify by conventional analytical methods, requiring significant efforts at concentrating residues. The effort to measure contaminants in water can be further complicated if the contamination originates from a point source into a moving stream or river. As a result, the use of biomonitors or environmental sentinels has been proposed as a method to determine the status of contamination by chemicals at both the tissue residue and the genetic level.[1] This idea takes advantage of the bioaccumulating ability of some organisms for lipophilic organic chemicals, which is the category into which many of the environmental pollutants fall. For example, mussels (Mytilus species, for example) have been used in the past in the United States Environmental Protection Agency (USEPA) ''Mussel Watch'' program[2] as convenient, sedentary sentinels. The high lipid content and water-filtering function and capacity of both mussels and oysters make them ideal monitors of aquatic pollution. They may also be placed in baskets or pens to protect them from predators and strategically located in a grid pattern to obtain more controlled data.

Investigations have also used other species to monitor environmental contamination in water.[3,4] Catfish (*Ictalurus punctatus*, for example) are also a logical model species to investigate new analytical sample preparation techniques since they are filter feeders, bottom dwellers and an important food source.[5] Thus, any sedentary or relatively stationary species may serve as a living record of past exposure. Taken in combination with other data from soils, sediments, water, etc., this can provide an

interpretation of the pollution history and status of an area.

Vegetables, fruits, mushrooms, legumes, cereals, etc. are also exposed to a variety of herbicides, pesticides and water- or wind-borne pollutants. Such sentinels are completely stationary and can provide an exposure history when examined over a large area in the field or after collection and processing. Of most immediate concern, however, is their entry into the food chain.

Regardless of the species, all plants, animals and living organisms constitute the range of biological samples that can be analyzed to assess environmental pollution. They, as well as we, are all sentinels to our environment, all canaries in the mine.

1.3 Analytical Chemistry's Role in Responding to Environmental Pollution

How do we determine when a system has become polluted? How do we define the source and the nature of the pollutants involved? The ability to detect and assess the impacts of toxic wastes on various environments and the species that populate them is dependent on the ability to perform analyses of the water, sediments, soils and tissues of the affected biota. Such analyses must encompass the attributes of high efficiency of isolation, adequate sensitivity, specificity and linearity of response to whatever system of detection and identification is employed. Such methods must be of high quality, accurate, precise and reproducible, supplemented by a high degree of documentation, quality assurance (QA) and quality control (QC). Such methods must be thoroughly documented and validated as being able to be conducted in a number of different laboratories. In the USA this undertaking has been assumed by the government. In association with professional organizations, private industry and universities, there has been success in bringing some order to the chaos. This does not imply perfection, however.

While there has been significant progress in the recent past in the ability to extract and analyze pollutants in water, mainly due to solid-phase extraction (SPE)-disc technologies and advances in instrumentation ease of use and sensitivity, there has yet to be such dramatic progress in performing analyses in biological samples. The preparation and cleanup of such samples for submission to advanced analysis by our evolving instrumentation has remained a difficult and cumbersome task.

Most methods recommended by various governments for dealing with biological samples involve the classical approach of sample disruption, solvent/solvent extractions, back-extractions and column chromatographic cleanup steps. Analytical chemists have long recognized that such methods are limited in their turnaround or response times and expend large volumes of solvent and personnel hours, especially when they must be applied to large numbers of samples. The solvent use is of particular concern since many of the methods require the use of liters of solvent even for a single sample, solvent that is often evaporated into the atmosphere, polluting thousands of liters of air. Discarded solvent must be handled as hazardous waste, adding large volumes of solvent residues to already burdened disposal sites and greatly adding to the overall costs of solvent disposal, not to mention the initial costs of solvent purchase and the unknown costs of personnel exposure. Many of these methods may, in the final analysis, contribute more pollution to the environment than they ever detect. For these reasons future directions for the analyses of toxic wastes and pollutants in general will proceed through the development of simple, generic and environmentally friendly methods that require little or no solvent and that are capable of providing data in rapid response to inquiries concerning contamination and waste remediation.

2 ENVIRONMENTAL REGULATIONS AND METHODS: THE RESPONSE OF GOVERNMENTS TO THE PROBLEMS OF POLLUTION

The final expression of frustration in dealing with any serious problem is the establishment of a governmental agency to study and attempt to control the problem. Laws are passed and regulations are written. Once the regulations and laws are put into practice, bureaucrats are imbued with the authority to interpret them and to write further regulations to regulate the regulations already in place. Then come the judiciary to settle the disputes between the regulators and the regulated. On an international basis the situation can become even more complicated as the laws of one nation clash with those of its neighbor. However, pollution is at once local, national and international, with all nations sharing the same air, the same water and the same global environment in the end.

2.1 The Development of Laws, Regulations and Regulatory Agencies for Monitoring Environmental Pollution

A number of legislative events enacted in the USA since the late 1960s have initiated the development and standardization of analytical methods for environmental contaminants in general. An overview of early environmental legislation has been presented by Foster.[6] The first modern major legislative act approved in 1969 by the United States Congress was the National Environmental Policy Act (NEPA) and the Environmental Quality Improvement Act (EQIA) of 1970. Shortly thereafter,

Congress created the USEPA from which to coordinate government action and focus authority for protecting the environment. As a result, the USEPA implemented four major pieces of environmental legislation. Amendments to the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) of 1947 and the Federal Water Pollution Control Act (FWPCA) of 1948, in 1972 placed them under the control of the USEPA. Another piece of legislation in 1972 was the enactment of the Marine Protection Research and Sanctuaries Act (MPRSA) implemented to expand the United States protection of the open ocean to the extent of their authority. In 1976, the Toxic Substances Control Act (TSCA) gave the USEPA regulatory control over commercial chemical use. Other significant enactments of environmental legislation since then have been Resource Conservation and Recovery Act (RCRA) and Superfund.

Ultimately, three federal agencies participate in the responsibilities of regulating pesticides and related compounds for the United States. The USEPA (see above), the United States Department of Agriculture (USDA), and the Food and Drug Administration (FDA). The USEPA has the responsibility of approving the use of pesticides as well as establishing residue tolerances in foods.[7] The USDA enforces these tolerances in meat and poultry and the FDA is in charge of enforcing residue tolerances for any other food products shipped in interstate commerce[8] although the FDA does play a role in aiding the USEPA in establishing residue goals.[9]

Currently, the FDA conducts an extensive residue monitoring program each year involving the analysis of over 20 000 samples using multiple residue methods. An overview of this program and the results since 1978 has been described.[10–14] Data collection and dissemination programs have been implemented by the FDA.[15,16] Seafood and seafood products, although not required to pass unified, continual federal inspections by the FDA, remain a concern with the public because of the specter of environmental contaminants[17] occurring in the open ocean and off the coasts of other nations as well as the USA. In 1965, the United States Bureau of Commercial Fisheries began the National Pesticide Monitoring Program, which lasted until 1972, to monitor chlorinated pesticide residues in shellfish populations off the coast of 15 states.[18] The results demonstrated that 1,1,1-trichloro-2,2-bis[*p*-chlorophenyl]ethane (DDT) (CAS 50-29-3) was virtually ubiquitous. This program was continued to 1976 and indicated that pesticide concentrations were declining.[19] In 1977, the USEPA conducted a one-time sampling event of bivalves from the estuaries previously monitored under the National Pesticide Monitoring Program which demonstrated that the presence of pesticides had continued to decline dramatically.[20]

The Food Safety and Inspection Service (FSIS) of the USDA is responsible for monitoring the residues of chemicals that might be present in meat and poultry products. This responsibility is manifested into the National Residue Program (NRP). The NRP enables the sampling and testing of meat and meat products for residues from pesticides, environmental contaminants and animal drugs.[21,22]

The National Oceanic and Atmospheric Administration (NOAA) created the National Status and Trends (NS&T) Program to assess the effects of human activity on the quality of coastal and estuarine habitats throughout the USA. This program was designed to monitor chemical contamination and the biological effects of chemical contamination. Two efforts were begun in 1984 and 1986 to monitor the tissues of benthic fish (The Benthic Surveillance Program) and bivalve molluscs and sediments (The Mussel Watch Program), respectively. Results and some of the early trends from these programs have been reviewed.[23,24]

A problem arising from the existence of so many monitoring and data gathering programs is the lack of integration of the enormous amount of information that is generated. As a result, policy and decision makers seldom have easy access to this information from which to prioritize research or determine to what extent policy goals are being attained. From this concern, the USEPA has initiated an integrated program for monitoring ecological status and trends known as the Environmental Monitoring and Assessment Program (EMAP).[25] The problem of integrating various sources of environmental monitoring and residue data are even greater when one considers this issue globally.

Outside the USA, there are over 50 international organizations that are involved with pesticides and other pollutants. However, there are legitimate concerns pertaining to the ability of developing nations to possess the resources necessary to evaluate toxicity from contaminated food or perform residue monitoring. As a result, the United Nations has several organizations to address the problems of conflicting agencies and countries with limited resources. A review of these efforts and the problems that exist has been published.[26]

Even with the existence of substantial legislation and monitoring programs enacted to protect the consumer from harmful chemical residues in food, the public remains concerned over the issue. According to a review[27] on dietary pesticide risk assessment, 80% of shoppers surveyed in USA stated that pesticide residues in foods were a major concern. The subject of pesticides and food safety has been reviewed[28] as have residue concerns in seafoods.[29–31]

2.2 Early Analytical Efforts to Respond to Pollution Events

With the introduction of the chlorinated pesticides after the second World War and the general boom in industrial manufacturing and chemical synthesis, pollution of the environment proceeded almost unabated. It was not long before the toxicity of the by-products, wastes and the products themselves to the environment became evident. When the need to analyze for these materials was realized, the technology of the time was the first to be applied to their isolation and analysis. Most of this 1940s and 1950s analytical approach is still in use today. The same can be said of almost all recognized classical and/or official pollutant methods.

Most ''classical'' methods were based on homogenization of the sample matrix with an extracting solvent followed by extensive cleanup procedures to remove, for the most part, lipids and other co-extracting neutral compounds. Water and aqueous homogenate samples were extracted by liquid–liquid partitioning of the pesticides followed by a range of extraction protocols to remove interferences or clean up the sample from emulsions, tissue debris, etc.

Early determinative analyses were often conducted by non-chromatographic methods using measurement of total chlorine content, ultraviolet (UV), visible, and fluorescence spectroscopy. Biochemical techniques that took advantage of the activity of the compound on specific enzyme systems or organisms, such as acetylcholinesterase activity and the survival time of flies exposed to the extracts for certain pesticides were also used.[32]

Such methodologies were most often developed for single compounds. However, as a large number of such compounds became available, the extraction methodology had to become more nonspecific while the analytical methodology had to become more specific. The development of multiresidue methods became a necessity and was fostered by the growing need and concern to monitor for residues in agricultural crops and food animals that were exposed to such compounds. This led to the development of a variety of techniques that were based on different aspects of the nature of the pollutants and the sample matrix to be tested. Many of these procedures, particularly the ''Mills' method''[33,34] for chlorinated pesticides, were subsequently incorporated into what are now ''official'' methods of analysis for these compounds. Nevertheless, these methods have been regularly revised and still adhere to rather antiquated processes for sample homogenization, extraction and cleanup that are labor-, materials- and solvent-intensive. Furthermore, most of these methods have not been adequately examined for method performance in the variety of food-animal species and agricultural products available and must be validated anew for each application.

It became evident that the burgeoning publications of new methods and approaches would have to be centrally controlled and assessed and that particular government agencies should be appointed to control and regulate both the use of the methods and the use and production of these chemicals as potential pollutants. Enter the politicians and the accompanying bureaucrats.

2.3 The Evolution of Analytical Methods into Regulatory Methods

Historically, many of the analyses for residues (whether drugs, pesticides or other environmental contaminants) in biological matrices were generally developed and performed on the basis of individual analytes.[35] It was not until the advent of chromatographic techniques (paper chromatography, thin-layer chromatography, gas chromatography (GC) and high-pressure liquid chromatography (HPLC)) that multiresidue analyses began to be developed and routinely employed. The classical method developed for the analysis of chlorinated pesticide residues in foods using paper chromatography[33] was pivotal in terms of the evolution of multiresidue analyses. The Mills' method, which utilized solvent-partitioning techniques for compound isolation and columns (e.g. Florisil) for fractionation and sample cleanup, became the foundation of the FDA monitoring program for the analytical techniques utilized for multiple-residue analyses.[35]

Further refinement of the Mills' method was published in 1963 using GC and a halogen-specific detector[34] and became known as the MOG (from the authors' last names) method.[35] Subsequent collaborative laboratory studies following the Mills' method led to the publication of multiresidue methods in the Association of Official Analytical Chemists (AOAC) Official Methods of Analysis (1990) and are the basis of most multiresidue methods that are utilized today. A more thorough review of the evolution of these early multiresidue methods has been reported.[36] For example, most methods currently being used by monitoring agencies for pesticide analysis are based on five 'classical' multiresidue methods, some developed over thirty years ago. These methods are commonly called the nonfatty (MOG), fatty (Mills'), Luke, Storherr and Krause methods. Together they detect over 300 pesticides or pesticide-related compounds.[13] Most of these methods have undergone rigorous multilaboratory calibration studies, such as those needed to obtain official acceptance by the AOAC, a requirement in becoming an official method. These methods are the backbone of residue analysis protocols for the FDA,[14] USEPA,[15] and the NOAA.[16] These methods work well under

Table 1 Internet addresses for access to regulatory authorities and/or methods prescribed for environmental analysis

Organization	Internet address
AOAC International	www.aoac.org/
EPA	www.epa.gov/
FDA	www.fda.gov/
USDA	www.usda.gov/
USDA FSIS	www.usda.gov/agency/fsis/homepage.html
AACC	www.aaccnet.org/
AOCS	www.aocs.org/

AACC, American Association of Cereal Chemists; AOCS, American Oil Chemists Society.

certain conditions and for certain purposes. However, perhaps the greatest drawback to their continued use is their inefficiency as screening methods. Furthermore, these methods are sufficiently complex as to not allow the generation of relevant data in time to prevent contaminated foods from entering the marketplace before a pollution or contamination problem is noted. For example, the FDA is responsible for prohibiting interstate marketing of food containing illegal pesticide residues but, in many cases, food is sold before the FDA completes the analyses.

Table 1 provides the internet addresses of agencies involved in environmental analyses in the USA.

3 METHODS AND THEIR APPLICATIONS

3.1 Sample Preparation

As noted in the beginning of this article, analysis of any sample requires rigorous preparation and consideration of a wide range of factors before proceeding to extraction. These are addressed below.

3.1.1 Documentation

It is of absolute necessity in any analytical undertaking that the analyst provide and be provided with adequate documentation. This process begins with recognition that the data must be generated using validated official methods or their equivalent if the results are to be found acceptable by the environmental analytical community. Detailed protocols for sample collection and identification, chain of custody, transfers, handling, storage, access, etc. exist and are expected to be used in any environmental analytical process. Laboratories conducting such analyses are usually certified and regularly audited to assure that they meet QA and QC guidelines and regulations, of which much is documentation.

3.1.2 Collection

There is a considerable amount of information on sample collection, including statistically derived sampling protocols, in the technical literature.[37–39] Any method of sample collection attempts to ensure that samples represent their populations and that sample sizes are manageable and cost-effective for analysis. Typically, edible aquatic resources (i.e. mainly finfish, shellfish and molluscs) collected for regulatory purposes are either presented whole, as a raw agricultural commodity, or are presented as processed foods (e.g. in canned or breaded form). The same is true of other raw and/or processed commodities.

In conducting sample collection and preparation for analysis the analyst must keep in mind the many possible routes of exposure to the component parts of a target biological sample. The analyst must examine the environment surrounding the target as a whole to fully assess the contribution of air, water, soil, diet, etc. Assurances must be made that a further route of contamination does not occur; contamination by the analyst.

3.1.3 Preservation

Ideally, raw, whole samples should be analyzed directly after collection. Because analysis at the time of collection is often impossible, a few special precautions must be taken to preserve the integrity of the samples. For samples collected for environmental monitoring purposes, the USEPA recommends, for example, wrapping whole fish in aluminum foil and preserving by either icing down samples to be analyzed immediately (i.e. within 24 h) or by quick-freezing in dry ice any samples to be stored (see section 3.1.4).[39]

An alternative method of preservation described in the literature is tissue dehydration. Homogenized molluscan tissues that are first dehydrated with a desiccant mixture (90% Na_2SO_4 and 10% silica) can be wrapped in aluminum foil and held at room temperature for over 15 days without degradation or loss of chlorinated pesticide residues.[40–42] Whatever the biological matrix, consideration for the preservation of the sample itself and the possible pollutants contained therein must be given. Since most of the mechanisms for degradation involve heat, light, air and biological oxidation, enzyme activities and the like, freezing of the sample and wrapping the sample in a clean, non-transparent container will usually suffice to preserve the sample for transport.

3.1.4 Storage

Storage requirements vary according to how quickly residues will be extracted from the tissues. Tissues

to be extracted within 24 h may be stored at normal refrigeration temperatures (+2 to +4 °C); whereas tissues to be held for a longer period of time should be frozen (−12 to −18 °C).[43] Whole fish may be stored for up to six months if kept in a freezer that maintains an even temperature below −26 °C.[39] Samples may be homogenized to decrease storage space. However, if these homogenized samples are to be held for longer than a month, they should be monitored for stability of the pesticide residues of interest.

Storage containers (e.g. zip-lock freezer bags) should prevent contamination, moisture loss or gain, and oxidation, hydrolysis, isomerization, polymerization or other cause of decomposition of tissues or residues. Well-sealed containers are especially important to protect samples from desiccation in self-defrosting freezers.

In addition, samples should be individually packaged for storage to avoid unnecessary handling of tissues not immediately being analyzed. Repeated freezing and thawing disrupts cell membranes. As a result, lipids, and therefore lipophilic pesticides, such as the chlorinated hydrocarbons, may be lost from cells.

The USEPA proposes an alternative method for storing samples, which requires immediate solvent extraction of lipids and pesticides from tissues and then removal of all or most of the solvent. Extracts can then be stored at low temperatures. However, the USEPA recommends evaluating residue decomposition in this state by creating and storing spiked controls along with the sample extracts.[39]

Depending on the nature of the biological sample and the compound being targeted for analysis, different storage conditions may be required. It is of value to conduct storage stability analyses in complement to any analytical procedure. Fortification of control tissues with a range of concentrations of the target analyte(s) assayed over times that are relevant to the storage of actual samples provides the analyst with data concerning the accuracy of the final result. It cannot be assumed that simple freezing of a sample will assure stability, as many compounds will continue to degrade even when frozen at −80 °C.

3.1.5 Shipment

Samples for environmental analyses are often collected under field conditions that do not permit immediate freezing. Nevertheless, samples should be shipped frozen as described above. Samples subjected to elevated temperatures that may exist during travel could undergo rapid degradation. Such degradation may affect the chemical nature of the residues and will definitely affect lipid structure and content of tissues, possibly resulting in a loss of the residues. Changes in structure and content of samples may also adversely affect subsequent isolation procedures.

3.1.6 Stability

In general, most of the chlorinated pesticides and other compounds of interest, such as PCBs and PAHs, present in biological samples occur as relatively stable residues. For example, most DDT has been metabolized or has decomposed to 1,1-dichloro-2,2-bis[*p*-chlorophenyl]ethylene (DDE) (CAS 72-55-9) and 1,1-dichloro-2,2-bis[*p*-chlorophenyl]ethane (DDD) (CAS 72-54-8), both very stable compounds and hence, persistent residues. Most of the 16 chlorinated hydrocarbon pesticides that are regularly monitored by the USEPA have been banned for many years in the USA so are generally found in tissues as their most stable metabolite or decomposition product. Other examples include aldrin (CAS 309-00-2) and heptachlor (CAS 76-44-8), which occur as dieldrin (CAS 60-57-1) and heptachlor epoxide (CAS 1024-57-3), respectively, in tissues.[44]

Standards of all chlorinated pesticides are necessary for residue determination and confirmation. The FDA Pesticide Analytical Manual (PAM) states that these compounds are ''rather stable'' and not subject to decomposition in solution.[45] In a study to determine stability of several of these compounds (individually prepared in 2,2,4-trimethylpentane), results of GC analysis indicated no other substances other than the parent pesticides over a period of eight months. However, the manual recommends protecting all standards from UV radiation. Different recommendations will hold for the other classes of pollutants and individual compounds.

3.1.7 Processing Before Extraction

Samples generally must undergo some sort of processing procedure before the residues can be extracted. Most processing procedures are steps taken to isolate specific organs or tissue types. For some of the environmental monitoring protocols reviewed in the literature,[46] processing includes removal of any tissue parts having direct contact with the aluminum foil or container in which the samples were stored. This step is thought necessary to prevent surface contamination of tissues.

For example, the FDA PAM includes instructions for preparing the edible portion of fresh hard shell and soft shell crabs; oysters; clams; fish; and shrimp, crayfish and other shellfish:[45]

1. For hard shell crabs, a 'homogenous mixture of meat and fatty materials' is examined. This mixture is isolated by heating the crab in boiling water or steaming autoclave for one minute. The claws and other appendages are removed; meat is recovered

from them. After the back shell is removed, all meat from this location is conserved, but viscera and gills are discarded. The entire body of a soft shell crab is used for its analysis.

2. For oysters and clams, a homogenous mixture of meats and liquor is examined.
3. For fish, the following parts are discarded: heads, scales, tails, fins, guts, and inedible bones. Skin remains, and all muscle is filleted for inclusion in the sample. Muscle from very large fish may be sampled from various anatomical locations.
4. For shrimp, crayfish and other shellfish, only the edible meat is examined. All heads, tails and shells are discarded. The manual does not specify what to do with hepatopancreas ('backfat'), which is considered to be edible in crayfish and is often included in processed packages of frozen crayfish tail meat sold in commercial markets.

Homogenizing entire organisms or first isolating the edible portions before homogenizing is very time-consuming and labor-intensive. In most cases, large (20–50 g) subsamples of these materials are used for extraction with only a portion of the extract actually being used in the final steps of analysis. New methods efficient enough to extract and quantify the amount of residue present in a small aliquot of muscle, liver, fat or other relevant biological component could be easily adapted to these procedures and could eliminate some of these inconvenient steps.

3.1.8 Quality Assurance and Quality Control

Analytical data concerning pollutant residues must be easily interpretable, unambiguous, valid and reliable in order to support the decisions made by a regulatory agency to condemn contaminated products or consider assigning an area a designation as a public health hazard. Ideally, contaminated products intended for consumption are condemned and removed before they enter commercial markets in order to limit the amounts of chemical residues entering the food supply. The presence of chlorinated hydrocarbon pesticide residues in edible aquatic species, for example, continues to cause concern because of the as yet unknown health effects of many of these chemicals. Many of the chemicals are classified as suspected carcinogens.[47]

In order to obtain acceptable data, regulatory agencies on both state and federal levels that are currently involved in environmental monitoring must use one of the few 'approved' methods for residue analysis. These are the methods that have undergone the rigorous multi-laboratory calibration studies necessary to achieve the AOAC Official Method status. The methods in the FDA PAM are mostly AOAC approved and designated as such in the manual.[45] Often funds are not available for developing and testing new methods even though the existing methods often fail to be simple, rapid, safe to workers and cost-effective.

For its environmental monitoring programs, the USEPA has an interlaboratory QA program which functions for state and private laboratories working under contract for either the United States Department of Health and Human Services or the USEPA.[39] The program requires all participants to periodically analyze "check samples", samples prepared by a coordinating laboratory with an undisclosed level of residue, in order to "assess the continuing capability and relative performance" of the participant.[39] In addition, the program provides certain nonprofit laboratories with analytical-grade pesticide reference standards. All laboratories must follow the standard analytical methods in the USEPA manual[43] and use standard materials supplied by the USEPA. On-the-job training and assistance may be provided. The USEPA also runs an electronic facility for maintenance of laboratory instruments. The USEPA also utilizes a Laboratory Data Integrity Branch (LDIB) which is responsible for implementation and refinement of the agency's GLPs compliance monitoring program.

Part of the QA of such studies is the use of primary reference standards. Primary reference standards are analytical-grade standards that are greater than or equal to 99% pure. There are several commercial sources for primary standards of the chlorinated pesticides. The standards maintained by the USEPA QA Program are no longer available free of charge to any laboratory that requests them. The USEPA standards are now only supplied to specific laboratories.

3.2 Extraction and Cleanup

This section will provide an overview of the evolution of extraction and cleanup techniques in environmental analysis of biological samples. Table 2[48–50] is a descriptive outline of cleanup techniques. The outline is based on reviews by Walters[48] and Lott and Barker[30] on cleanup techniques for pesticides in fatty foods. The dates in the outline generally correspond to when particular techniques were first developed for cleanup of tissue extracts for pesticide analysis. Most of the techniques have been applied to clean up tissue extracts for subsequent isolation of chlorinated pesticides.

The initial approaches involved standard countercurrent, liquid–liquid extraction. These processes were complicated by the formation of emulsions, requiring lengthy procedures and addition of salts, etc., to prevent or resolve their formation. These extraction procedures also resulted in isolating both the lipophilic pollutants

Table 2 Descriptive outline of cleanup techniques used in methods for pesticide analysis of fatty foods

EARLY

I. Liquid–liquid partitioning
A. Hexane and ACN
1. Hexane – lipids, nonpolar constituents
2. ACN – pesticides, polar constituents
B. Petroleum ether (pet. ether) and ACN

II. Adsorption chromatography late 1950s
- All of the agents act as polar sorbents and retain polar lipids when organic solvents of lower polarity are used to elute analytes.
- used as an adjunct to liquid–liquid partitioning
- applications of alumina and silica gel studied more intensively in Europe

A. Magnesia
B. Florisil (synthetic magnesium silicate)
C. Alumina
D. Silica gel

III. Saponification late 1950s
- used as an adjunct to adsorption chromatography
- limited in use to analytes that are stable to the treatment (e.g. some OCs, such as PCBs, and DDT analogs)

A. Potassium hydroxide–alcohol treatment and subsequent partitioning of analytes into pet. ether
B. Sulfuric acid digestion when working with acid-stable pesticides

LATER

IV. Sweep co-distillation 1960s
- popular in Australia
- UNITREX is a commercial instrument
- can be used to isolate OCs
- OCs in beef fat[26]

A. Technique is volatilization of pesticides in fractionation tube and their subsequent collection in a Florisil trap
B. Relatively non-volatile lipids do not exit fractionation tube
C. Pesticides then eluted from trap and concentrated

V. Low-temperature precipitation[27] −78 °C dry ice/methanol bath
A. Lipids, waxes, H_2O – frozen out of solution (precipitation)
B. Pesticides (polar and apolar) left in supernatant acetone/benzene extraction solvent

VI. GPC 1970s
- official method in Germany
- commercial system available
- often requires additional cleanup with Florisil adsorption chromatography
- widely applicable (fish, OCs)

A. Polymer beads retain small pesticide molecules (M_r 200–400).
B. Lipids eluted from column first (M_r 600–1500)

VII. Semipreparative LC 1980s
- OCs and OPs

A. Uniform microparticulates:
1. Silica (normal phase)
2. ODS-bonded silica (reverse-phase)
   a. ODS-bonded silica retains long-chain lipids
   b. ACN is a common eluting solvent for pesticides

B. SPE cartridges

MORE RECENT

VIII. SFE
IX. MSPD
X. MAE
XI. ASE
XII. SPE
XIII. SPME

ACN, acetonitrile; ASE, accelerated solvent extraction; GPC, gel permeation chromatography; LC, liquid chromatography; MSPD, matrix solid-phase dispersion; MAE, microwave-assisted extraction; ODS, octadecylsilyl; OCs, organochlorine compounds; OPs, organophosphorus compounds; SPME, solid-phase microextraction; SFE, supercritical-fluid extraction.

and all the lipids of the sample simultaneously with little or no resolution. The introduction of sorbent materials such as Florisil and alumina that had a greater affinity for lipids over the relatively more polar analytes was a great advance in the field and greatly simplified the process of pesticide isolation from fatty samples. Removal of fats by saponification enjoyed a brief consideration but many of the pesticides were not stable to the conditions required. Steam distillation and sweep co-distillation which employed sweeping the distillate with a gas to carry the vapor into a collection tube was also found to be applicable but limited. The development of GPC greatly enhanced the ability to perform purification of biological samples for the isolation of pollutants. The procedure is discussed in detail later in this article. The evolution of other chromatographic methods, particularly SPE, also greatly enhanced the ability to perform environmental analyses in biological samples by providing more consistent materials for chromatography. The development of bonded phases that permitted a higher-resolution fractionation of extracts using a variety of phases and solvents was an immediate boon to the field. More recent developments, which will be discussed individually, are SFE, MAE, ASE, new SPE materials and devices, SPME and MSPD.

For illustrative purposes, Tables 3, 4 and 5 are outlines of the methods currently used by the FDA,[45] USEPA,[43] and NOAA,[51] respectively. These methods are very similar; extraction of lipids is done with solvents, lipids are removed by liquid–liquid partitioning and adsorption chromatography with Florisil, and further cleanup and fractionation of pesticide residues is usually performed with additional adsorption chromatography columns. The methods also include supplemental cleanup steps.

As mentioned previously, most analytical methods can be divided into four basic procedures: extraction of lipids and associated pesticides, preliminary cleanup of extract (bulk removal of lipids), isolation of pesticides and detection of pesticides. The separation of cleanup and isolation techniques into two different categories is not always applicable to a method. Some of the techniques, such as liquid–solid adsorption chromatography techniques, have been relied upon to both clean up and fractionate extracts with one elution.

As the preceding information indicates, there is sufficient methodological data concerning the analysis of edible aquatic species for chlorinated pesticides to formulate an opinion concerning their appropriateness and use in developing new or future directions for the field. As underscored by the General Accounting Office report of 1986,[52] there is an existing need to dramatically increase the numbers and types of samples analyzed for chlorinated pesticides as well as other compounds in the environment. A governmental monitoring program for aquacultured and wild-caught species is also proposed for the USA and is already under way in a number of other nations. Both of these needs will foster a demand for new multiresidue methodologies that are capable of handling the expected sample load and that will do so in a cost-efficient and rapid manner.

The ''official'' methods presented here are based on classical techniques that have evolved through and withstood the test of time and use. Methods become classical for a reason. They work. However, these methods are not capable of analyzing an order of magnitude or more samples without producing greatly increased costs for testing and a laboratory gridlock in the effort to impose such lengthy, labor-intensive methods on a task that requires speed and reduced costs. Nevertheless, such ''validated'' methodology has its place in the determinative or quantitative phase of pesticide analysis and in the final confirmation of results that will be taken before a court or regulatory body. The greatest failure is their use as a screening technique, for which they continue to be used today. These methods are not sufficiently fast to provide data on a time line that includes the prevention of contaminated foods from entering the marketplace. Presently, results are obtained after the fact and penalties are levied after the 'damage' has been done to the consumer. Further, these methods require large volumes of solvents, many of which are of greater environmental concern than the compounds they are used to isolate. Much of this solvent is evaporated into the atmosphere, contaminating the air, or may be improperly disposed of, leading to further contamination of aquifers and aquatic habits and resources. One of the greatest costs in analysis today, other than employee costs, is the purchase price and subsequent disposal costs of organic solvents and wastes. In many cases the disposal costs of a given solvent per liter now exceeds the original purchase price. For these reasons alone the use of these ''official'' methods should be severely curtailed and phased out as new ''official'' methods are developed and validated and as new methodology for the screening of samples begins to be developed and implemented. Chemists should avoid developing methods that simply mimic these classical approaches.

Most of the information for the variety of species examined to date involves the use of ''classical'' or modified classical methodology. There have been recent advances in the field, however, that offer a degree of optimism concerning our ability to address these analytical problems.

Table 3 Outline of the methods used by FDA for analysis of chlorinated pesticide residues in fish

I. Extraction of fats (AOAC)
   A. Sample: 25–50 g homogenized fish (approx. 3 g fat)
   B. Homogenization of tissues with anhydrous Na_2SO_4 which:
      • combines with water
      • disintegrates sample
   C. Multiple extractions with petroleum ether (pet. ether)
   D. Cleanup of extracts and rinses with a column of Na_2SO_4
   E. Concentration (Kuderna–Danish concentrator)
   F. Lipid determination
      Note: Fish containing <10% fat may be analyzed by methods for nonfatty foods.
      Total volume of solvent used: approximately 500 mL

II. Extraction of pesticides from isolated fat and oil (AOAC)
   A. Pet. ether/ACN partitioning
      Total volume of solvent used: approximately 365 mL
   B. Florisil column
   1. Florisil and Na_2SO_4 in column
   2. 6% ethyl ether/pet. ether
   3. 15% ethyl ether/pet. ether
   4. Concentration of each eluate
      Total volume of solvent used: approximately 460 mL

      Optional or alternative methods:

   C. ACN/pet. ether backwash (optional – to improve cleanup)
      Total volume of solvent used: approximately 275 mL
   D. Partition chromatography (alternative to pet. ether/ACN partitioning)
   1. Florisil column
      a. Sample
      b. 10% H_2O/ACN
   2. Separator
      a. Eluate
      b. Pet. ether
      c. Saturated NaCl
      d. H_2O
   3. Anhydrous Na_2SO_4
   4. Kuderna–Danish concentrator
      Total volume of solvent used: approximately 181 mL

III. Supplemental cleanup
   A. 6% eluate
   1. +/− GLC or thin-layer chromatography without further cleanup
   2. +/− Second Florisil column (AOAC)

      Total volume of solvent used: approximately 460 mL
   3. +/− Acid–Celite column (AOAC: DDT only)

      Total volume of solvent used: approximately 85 mL (more if sample contains more fat)
   B. 15% eluate
   1. +/− MgO–Celite (AOAC)

      Total volume of solvent used: approximately 140 mL
   2. +/− Alkaline hydrolysis (AOAC)

      Total volume of solvent used: approximately 50 mL
   3. +/− Alkaline hydrolysis followed by MgO–Celite

IV. Determination
   A. GLC/electron capture detector
   B. Confirmation:
   1. Initial test: thin-layer chromatography
   2. Others

Table 4 Outline of the methods used by the USEPA for analysis of chlorinated pesticide residues in human and environmental samples

I.	Extraction of fats
	A. Sample: 5 g
	B. Dry maceration with sand and anhydrous Na_2SO_4
	C. Addition of aldrin as a quantitative recovery check
	D. Multiple extractions with petroleum ether (pet. ether)
	E. Evaporation to dryness for lipid determination
	Total volume of solvent used: approximately 160 mL
II.	Liquid–liquid partitioning
	Pesticide residues are extracted from the fat with ACN and then partitioned back into pet. ether by aqueous dilution of the ACN extract.
	A. Solvents: ACN, 2% NaCl, pet. ether
	B. Drying with anhydrous Na_2SO_4
	C. Concentration (Kuderna–Danish concentrator)
	Total volume of solvent used: approximately 365 mL
III.	Florisil fractionation
	A. Florisil and anhydrous Na_2SO_4 in column
	B. 6% diethyl ether/pet. ether
	C. 15% diethyl ether/pet. ether (with addition of aldrin to this fraction) +/− MgO–Celite column chromatography afterwards
	D. Concentration of each eluate
	Total volume of solvent used: approximately 461 mL (MgO–Celite chromatography would require an additional 140 mL of solvent.)
IV.	GLC/electron capture detector determination

Table 5 Outline of the methods used by the NOAA for analysis of chlorinated pesticide residues in tissues

I.	Extraction
	A. Sample: 3 g
	B. Homogenization with internal standards and Na_2SO_4 and 3 extractions with DCM
	C. Centrifugation (repeated 3 times in conjunction with the extractions)
	D. Concentration with boiling chips
	Total volume of solvent used: approximately 80 mL
II.	Silica gel/alumina column chromatography
	A. Alumina, silica gel and sand in column
	B. Pentane
	C. 50% DCM/pentane
	D. Concentration of each eluate
	Total volume of solvent used: approximately 246 mL
III.	Sephadex LH-20 Chromatography
	A. Second fraction only (50% DCM/pentane)
	B. Concentration of eluate
	Total volume of solvent used: approximately 235 mL
IV.	Determination
	A. GLC/electron capture detector
	B. GLC/mass spectrometric detector confirmation

DCM, dichloromethane.

4 NEW APPROACHES TO SAMPLE PREPARATION AND CLEANUP FOR ENVIRONMENTAL ANALYSES

4.1 Gel Permeation Chromatography

A property of solids can be the presence of holes or pores in the surface, part of the very structure of the materials, that permit penetration or permeation of certain size molecules while excluding others. Thus, this size exclusion or permeation process can be used to fractionate materials by size, using smaller pores to trap or retard smaller molecules and increasingly larger pores to retain or slow increasingly larger, higher-molecular weight-species. Such materials can be said to possess molecular-weight 'cut-offs', retaining molecules below the stated value while permitting the elution of species above this value, acting, essentially, as molecular sieves. This process is not highly specific, however, and lacks resolving power, with overlap of a variety of similar molecular-weight species and, of course, much greater molecular-weight species. Nonetheless, as an approach to roughly separating lower- (such as pesticides and herbicides, MW range of 200–400 amu) and higher- (proteins, carbohydrates, triglycerides, etc. MW range of 600 or greater) molecular-weight species it can often prove to be a useful and powerful technique. This application has been enhanced by the development of automated systems.

Both natural and manufactured materials of defined pore sizes are available. These materials have the appearance of gels when in the presence of solvent, and the process has been given the general name of GPC. These gels are, as might be expected, subject to damage by high pressure and one must be careful to match the permeation material with the appropriate solvent. Newer materials are available for use with higher pressures and are beginning to incorporate surface chemistries that provide a further level of interaction with matrix and target analyte components that enhances the technique for certain applications. The most common materials employed are divinylbenzene-linked polystyrene beads of a 200–400 mesh size.

GPC, especially with the development of automated systems, has found renewed use for the fractionation and cleanup of biological samples for environmental analyses. It has also been used as an adjunct to the extraction and cleanup of samples following isolation of the target analytes by SFE as well as accelerated solvent and MAE techniques. Fractionation of target analytes from the large lipid band from high-fat samples is not always completely successful and additional cleanup steps are often required. These usually involve adsorptive or bonded-phase SPE techniques. Table 6 offers a list of applications and reviews for GPC in environmental analyses.[53–66]

Table 6 Applications of GPC to environmental analysis of biological samples

Analyte	Matrix	Ref.
Pesticides	Animal fats	53
Pesticides	Fatty foods	54
Pesticides	Fish and meat	55
Pesticides	Foods	56
Pesticides	Foods	57
Pesticides	Foods	58
Pesticides	Fruits, cereals, etc.	59
Pesticides	Fruits, vegetables	60
Pesticides	Olive oil	61
Pesticides	Plant materials	62
Pesticides	Plants, animals	63
Pesticides	Rice, carrots, etc.	64
Pesticides	Rice, fruit, etc.	65
Pesticides, drugs	Foods	66

4.2 Supercritical-fluid Extraction

If gases are maintained at specific conditions of pressure and temperature they will exist in a state known as the supercritical (SC) phase. Under these conditions the gases exist as liquids and can be used in the same manner as other solvents. Since most gases are nonpolar in character, the SC gas can be used as a nonpolar solvent to extract nonpolar analytes from samples. Removing the SC conditions, such as lowering the pressure or raising the temperature, causes the gas to return to its original state. Thus, if we use carbon dioxide (CO_2) from the atmosphere, compress it to a SC state and use it as an elution solvent to isolate compounds from a sample matrix, we can eventually ''dispose'' of this solvent by letting it go back whence it came. Modification of the polarity of the SC gas to obtain a better or more selective extraction can be accomplished with the addition of other gases or, more commonly, the addition of small amounts of ''normal'' solvents, such as methanol, to the SC phase. This provides a high degree of flexibility and more selectivity to the technique. Extraction efficiency in SFE also varies with temperature, pressure, SC fluid flow rate, extraction vessel dimensions, sample water content and a variety of other condition variables.

With the SFE technique carbon dioxide (SC-CO_2) is commonly used as the extraction medium in place of organic solvents. Chlorinated pesticides and associated lipids are soluble in SC-CO_2. They can be extracted from tissues and then collected once the pressurized CO_2 is brought back to atmospheric pressure. This may be accomplished by collection of the SC gas in a solvent or trapping of the analytes by passing the eluate through a solid surface support material. The main advantage of this

technique is that expensive, flammable, and potentially hazardous organic solvents are not used in quantity. However, because the extracts contain contaminating lipids, a cleanup step is usually needed before samples can be injected onto a GC apparatus. Cleanup can be performed with a variety of chromatographic techniques, such as GPC or adsorption chromatography with Florisil, for example.

Water in the sample can interfere with the extraction process and variable water content from sample to sample can confound analytical results. This problem has been overcome with the development of a protocol that calls for first mixing samples with diatomaceous earth and other dessicants (silica, for example) before extracting with SC-CO_2.[67] This step disperses the sample material and allows adsorption of water before extraction with SC-CO_2 or other suitable SC-solvent.

A drawback to the use of SFE is the initial capital outlay required for the equipment and the limits on the number of samples that can be processed in a batch. Newer equipment is becoming less expensive and more efficient for sample throughput and such concerns must be balanced against the savings incurred from not having to purchase and dispose of large volumes of organic solvents.

Table 7 is a list of applications of SFE to isolation of pollutants from biological samples.[62,68–91] An excellent general review of SFE has also been published.[92]

Table 7 Application of SFE to the isolation of pollutants from biological samples

Matrix	Pollutant	Ref.
Tomatoes	58 pesticides	68
Fruits and vegetables	71 pesticides	69
Plant materials	Pesticides	62, review
Foodstuffs	Pesticides	70
Eggs	Organochlorine pesticides	71
Fruits and vegetables	46 pesticides	72
Vegetables	Methamidophos	73
Vegetables	Pentachloronitrobenzene Pesticides	74
Adipose breast tissue	Chlorinated pesticides, PCBs	75
Grass	Chlorpyrifos	76
Meat products	Pesticides	77
Foodstuffs	Pesticides	78
Plant materials	Sulfonylurea herbicides	79
Mussels	Chlorinated pesticides, PCBs	80
Foods	Pesticides	81, review
Fats	Organochlorine and Organophosphorus pesticides	82
Foods	Pyrethroid pesticides	83
Eggs, fish	PCBs	84
Human adipose	PCBs	85
Fish	PCBs	86
Rice	Organophosphorus pesticides	87
Apples	Thiocarbamate pesticides	88
Milk	PCBs	89
Tissues	Chlorinated hydrocarbons	90, review
Tissues	Pesticides	

4.3 Accelerated Solvent Extraction

The chemist can increase the rate of interaction of one molecule with another by increasing temperature which is, after all, a measure of relative motion. This increased motion in a closed system results in increased pressure. Thus, if we place a sample in contact with an extracting solvent in a closed system and elevate the temperature and the pressure, we will obtain a greater degree of interaction of the solvent with the sample. This process may translate into more efficient extractions, requiring less solvent and time. This is the case and such procedures have been applied in the form of automated systems performing what has come to be known as ASE. This is also a drawback to ASE; the initial outlay of funds to purchase the equipment. The analyst must assess such cost-effectiveness against sample load, solvent costs, employee costs and savings and turnaround time.

References for the use of ASE on biological samples specifically are limited at this writing. Obana et al.[93] have, however, reported the isolation of up to 19 organophosphorus pesticides at 0.1 ppm or less from four different foods using an ASE instrument operating at 100 °C and 1500 psi in less than 20 minutes. As in SFE, samples were mixed with a drying agent prior to ASE and the extract was cleaned up using GPC prior to instrumental analysis.

4.4 Microwave-assisted Extraction

Microwave heating for sample dissolution in elemental analysis has been a recognized technique for over two decades. The application of microwave irradiation to the extraction of compounds from biological samples has been a more recent development. In general, samples are homogenized and mixed with a solvent and the suspension irradiated at greater than 2000 MHz for short periods of time. Heating is usually repeated several times with periods of cooling in between to prevent boiling. The efficiencies seen with this technique approach those of classical Soxhlet extraction but can be performed much more rapidly. Further modifications of the technique mimic sweep co-distillation or steam distillation with air sweeping the surface of the heated sample and being collected by a condenser that protrudes through the oven housing.Techniques have also been developed using

closed-system microwave heating. The major advantages of MAE are shorter extraction times and reduced solvent use.[94] Ganzler et al.[95] have reported the extraction of a variety of compounds from seed, foods and feeds, finding the process to be more efficient than classical Soxhlet or liquid–liquid partioning extraction. Ren and Desmarchelier[96] applied the technique to the isolation of fumigants from grain, releasing the fumigants into a sealed-system headspace by microwave heating and assaying the headspace. The technique has also found application in the isolation of biomolecules[97] and in conducting environmental analyses in soils, sludge and other non-biological matrices.

4.5 Solid-phase Extraction

The use of the inherent chemistry of the surface of particulates such as silica, Florisil, sodium sulfate, etc. to interact with and assist with the isolation and purification of a variety of pollutants in different matrices, or the process of SPE, has been the backbone of many of the classical methods for performing environmental analyses on biological samples. In the SPE process, a compound is isolated from a liquid sample based on its relative solubility in the liquid mobile phase compared to its solubility in a solid support or a modified solid-support/bound-liquid stationary phase of differing polarity. Isolation is accomplished by passing the analyte dissolved in solvent (organic or aqueous) through a column containing the stationary phase with subsequent elution using an appropriate solvent. Several SPE methods have been developed to facilitate the extraction and cleanup of biological liquid and tissue samples.

For liquid matrices, such as blood plasma or serum, acceptable residue recovery may be obtained using protein precipitation and direct injection without cleanup, but the many impurities present can affect the chromatogram and accumulate on the analytic column, thus resulting in increasing back pressure. SPE cleanup helps avoid these problems and works well with biological fluids such as plasma, urine and cerebral spinal fluid. In addition, SPE extraction and analysis can be automated and done on-line and/or with on-line dialysis and column switching.

Before SPE can be used with solid tissue (e.g. muscle and liver), a separate homogenization step and often multiple filtration, sonication, centrifugation and liquid–liquid cleanup steps are required. While SPE may improve cleanup of these solid tissue samples, the additional labor and materials costs make SPE less suitable, in some cases.

SPE methods published for solid biological samples are often combinations of SPE with other methods such as homogenization, liquid–liquid partition, filtration, sonication and centrifugation. Because choice of SPE column depends on the matrix and on the particular compound of interest, a wide range of solid-phase columns of differing polarities have been used for extraction in biological samples and include C_2, C_8, C_{18}, NH_2, amberlite resins, and polystyrene–divinylbenzene polymers as well as materials that constitute the category of materials for adsorption chromatography.

4.5.1 Adsorption Mode

This mode of solid-phase interaction relies on the inherent surface chemistry of a given material, the functional groups naturally available on the surface. Materials such as alumina (Al_2O_3), Florisil (an activated magnesium silicate), silica (SiO_2) or silicic acid (silica gel; H_2SiO_3), activated charcoal (carbon), etc. can provide such adsorption surfaces. Florisil is perhaps the most often used sorbent in environmental analyses because of its capacity to adsorb fats. Using solvents of low polarity for elution, relatively nonpolar residues can be isolated with high efficiency. Florisil must be activated, however, by maintaining a certain percentage of water. Variability in this value can lead to erratic results. In many cases Florisil can be replaced by alumina. However, basic alumina will lead to the decomposition of some organophosphates and can be highly retentive of more polar target analytes. The silicas have found less general use but have been employed to perform fractionation of more polar constituents in complex extracts from the less polar components.

Extracted samples in solution may be percolated or passed through columns of such materials to obtain a differential distribution of target analytes from other components between the immobile solid phase and the mobile liquid phase. In this manner, target analytes may be retained on the column while co-extracted interferences pass through or are washed from the column by addition of another volume of the same or different solvent. Conversely, the target analytes may be made to pass through while undesirable compounds are retained. Often one sees both and a bit of overlap.

A number of such materials are also manufactured, often in the form of resins, that possess ionic surface chemistry characteristics. These materials may contain an intrinsic negative charge (sulfate, phosphate, etc.) capable of bonding or exchanging positive-ion-containing species or may contain a positive charged surface (quaternary amines, for example) capable of ionic bonding or exchange with negatively charged compounds. The ability to exchange ions using such surface chemistry, thus leading to greater retention of compounds undergoing such exchange, is useful in isolating and resolving charged species from neutral or weakly ionized compounds.

We should also note here that, especially in food chemistry, a number of solid sorbents have been used

to assist in 'drying' the sample. This is accomplished essentially by adsorbing the water in the sample matrix into a solid material such as diatomaceous earth (silica) or sodium sulfate by blending the materials together. Such an approach is used today to conduct SFE, ASE and Soxhlet extraction of homogenized samples. The adsorptive properties of diatomaceous earth blended with the sample eliminates complications due to the presence of water in such extractions. It also provides a level of interaction of the sample and the target analytes with the surface chemistry of the dessicant. Further, solid materials such as sand (silicates) have been combined with samples to serve as an abrasive, using the materials to assist in the disruption of sample architecture through blending or grinding. These agents also have a surface chemistry which may interact with the sample to varying extents.

The sample matrix itself is always a significant contributor to loss of target analytes owing to electrostatic and/or covalent or ionic interactions, much as occurs in the use of sorbent technology. Indeed, the solid sample matrix is the surface chemistry that the analyst must initially deal with in terms of obtaining efficient isolation of analytes, breaking protein binding or ionic interactions of analytes with matrix components. Homogenization of a sample with solvent, extraction by Soxhlet solvent reflux, etc. are extraction/fractionation processes that are controlled by interactions of the analytes with the sample matrix components, the sample's own surface chemistries. These chemistries are far more complex than simple adsorption, however, and cover the entire range from affinity binding (receptor-bound target analytes) to size exclusion, to adsorption and phase distribution.

4.5.2 Bonded-phase Partition Mode

While the use of solids in the adsorption mode does provide a range of chemistries for a variety of applications it does not always provide sufficiently adequate interactions to resolve very similar chemical entities. Thus, alteration of the surface chemistry by derivatization of the functional groups present on the surface with a variety of chemically unique compounds offers the ability to provide an even greater range of interactions that can be brought to bear on a given analytical problem. Solid materials that are commonly used for such purposes have been silica, with available silanols for derivatization, and synthetic polymers such as polydivinylbenzene or polystyrene, which can be chemically modified or initially formed to provide other functionalities.

The derivatizing agents to be used to prepare such surfaces can be designed so as to provide a range of surface chemistries, highly polar to highly nonpolar, high carbon content, low carbon content, cross-linked, etc. that make such materials applicable to a variety of analytical problems. The covalent bonding of the agent to the solid material provides a phase that is immobile or stationary, one that cannot be readily eluted from the material. Thus, this stationary phase can be used, essentially, as an immobile solvent in which sample components can be partitioned. This provides a mechanism for distribution of analytes and other sample components between this stationary "solvent" and the solvent in which the sample is dissolved and applied to the material, and the solvent subsequently applied to the material in an effort to selectively elute analytes or potential interferences.

4.5.2.1 Normal Phase A polar stationary material (or phase). Application of a sample will lead to retention of polar sample components and analytes while nonpolar species will be readily eluted or will pass through the material without retention. Washing with nonpolar solvents will clean the column of such nonpolar sample components or, depending on whether one seeks retention of target analytes, permit further isolation of nonpolar analytes in the eluate. Removal of retained material will require solvents of increasing polarity and/or the use of changes in pH or ionic strength, depending on the nature of the stationary phase and analyte interactions. Examples of normal-phase materials are underivatized silica, and diol-, cyano- and amine-containing derivatives.

4.5.2.2 Reversed Phase A nonpolar stationary phase. Application of a sample will lead to retention of nonpolar sample components while polar species will be permitted to pass through. Washing the column with more polar solvents will further remove the more polar components while removal of the retained species will require less polar solvents, thus, a reversed condition compared to normal phase. Examples of reversed-phase materials are ethyl, butyl, octadecyl, octyl, cyclohexyl or phenyl derivatives.

4.5.2.3 Ion Pairing A nonpolar stationary phase in the presence of an ion that counters the charge of the ion(s) present on the analyte, thus making it neutral and more interactive with the stationary phase; e.g. using tetraethylammonium ions to counter negatively charged functional groups on a target analyte.

4.5.2.4 Ion Exchange A stationary phase containing ionizable functional groups, ranging from strongly to weakly ionized. Application of a sample containing target analyte cations (positive charges) will cause such compounds to be retained by a stationary phase containing anions (negative charges) and vice versa. The material may be washed with organic or aqueous solutions containing no ions or ions of lower ionic strength that will, thus, not exchange for the ions retained. Retained

compounds are eluted with organic solvents or aqueous solutions containing ions of greater ionic strength than those bound to the stationary phase of the column. Some examples of ion-exchange materials are those derivatized with amino groups that are or can be readily quaternized (triethylaminopropyl (SAX); diethylaminopropyl (DEA)) or species carrying a negative charge (benzenesulfonylpropyl (SCX); sulfonylpropyl (PRS); carboxymethyl (CBA)).

4.5.2.5 Mixed Modes All sorbent materials are capable of exhibiting some or all of the characteristics described above. As previously mentioned, underivatized silanols on silica-based sorbents give the material both an adsorptive character that is associated with a normal phase and the intended character of the existing derivative, whether normal phase, reversed phase, ion pairing or ion exchange. Furthermore, many of the derivatives that are used to provide the more polar functionalities are covalently bonded to the silica surface through a C–H-containing chain that may be 1–8 carbons in length. Thus, molecules may undergo ionic interactions, lipohilic or hydrophobic interactions and/or adsorption to underivatized functional groups on the solid support all at the same time. This depends, of course, on the structure of the analytes and the other myriad matrix components. Interaction of target analytes with retained matrix components which are interacting with the solid support/stationary phase is also part of the consideration of the overall mixed-mode interactions taking place.

In all forms of sorbent technology one seeks to obtain varying degrees of retention and elution for the target analytes and other matrix components relative to one another, to take advantage of the relative selectivity of the sorbent and the method applied. This is under the control of several variables, such as the chemical structure of the analytes, the characteristics of the sorbents, the composition of the sample matrix and the solvent-elution conditions or profile selected. Maximum retention or selectivity for a given analyte occurs when the structure and chemistry of the analyte most closely matches that of the sorbent while all other components are divergent from this condition.

The capacity of the sorbent material, its ability to retain a relative quantity of material per gram or milligram of sorbent, must also be considered. This may range from 1–5% of the mass of sorbent (1–5 mg per 100 mg, for example) but this quantity must also take into consideration the fact that matrix components may also be retained on the column. Retention of large quantities of matrix and analytes can greatly alter the character of a column and the subsequent result. Thus, the greater the selectivity of a column for a given analyte, or the more divergent matrix components are from the analytes, the higher the resolution of components in the final eluate.

In all cases the analyst should keep in mind that a large range of variables and possible compositions of materials (sorbents, solvents, pH conditions, etc.) are available for addressing an extraction/analysis problem through the use of sorbent technology and procedures. Several chemically opposite approaches may ultimately work for a given problem.

SPE has played a major role in the development of methodology for the analysis of pollutants in water, with many of the methods now prescribed by the USEPA involving the use of this technology. Table 8 illustrates many of the applications of this process for the isolation of pollutants from biological samples.[53,58,62,63,83,98–125]

4.6 Solid-phase Microextraction

Miniaturization of sorbent technology and the concomitant decrease in solvent (purchase, exposure, disposal)

Table 8 Application of SPE to the extraction and cleanup of target analytes from biological matrices in environmental analyses

Analyte	Matrix	Ref.
PAHs	Fish	98
PAHs	Seafood	99
PAHs	Seafood	100
PAHs	Grilled meat	101
Pesticides	Animal fats	53
Pesticides	Beef	102
Pesticides	Cereal flours	103
Pesticides	Cereals	104
Pesticides	Crops, foods	83
Pesticides	Eggs	105
Pesticides	Fish	106
Pesticides	Fish	107
Pesticides	Foods	58
Pesticides	Foods	109
Pesticides	Foods	110
Pesticides	Foods	111
Pesticides	Foods	112
Pesticides	Fresh produce	113
Pesticides	Fruits, vegetables	114
Pesticides	High-water foods	115
Pesticides	Legumes, fruits	116
Pesticides	Liver	117
Pesticides	Milk	118
Pesticides	Milk, formula	119
Pesticides	Oils and fats	120
Pesticides	Oils	121
Pesticides	Plant materials	62
Pesticides	Plants, animals	63
Pesticides	Produce	108
Pesticides	Seafood	122
Pesticides, PCBs	Fish	123
Pesticides, PCBs	Nonfatty foods	124
Pesticides, PCBs	Seafood	125

has also taken a further giant step with the development of SPME devices. In this technology a fused silica fiber coated with polyacrylate, polydimethylsiloxane, Carbowax or otherwise modified bonded phase (addition of other sorbents) is placed in contact with a liquid sample, exposed to the vapor above a sample (solid, liquid) or placed in the stream of a gaseous sample. In such a manner SPME has been used most frequently to analyze volatile and semi-volatile compounds and may be used to replace purge-and-trap procedures. In relation to food chemistry SPME has proven to be particularly applicable to flavor and aroma analyses. While the technique has yet to be extensively applied to analysis of pollutants in biological samples it does, nonetheless, have great potential for such applications.

SPME is based on a simple principle that applies to all sorbent technologies; the materials in the sample will establish an equilibrium with the solid phase, based on their relative distribution coefficients. This does not require complete extraction of the compounds present but rather a representive sample that can be related to samples of known concentrations. Such samples can be readily quantitated by conducting similar exposures to standards of known concentration. The most interesting aspect of this technology involves the ability to use the exposed fiber as both an extraction and a sample delivery device, using specially designed interfaces to place the fiber in-line for GC or LC desorption or elution of analytes. Phases for SPME are available in a range of polarities and properties for application to analyses of volatile and semivolatile compounds as well as for possible applications to extraction of analytes from liquid samples. However, direct application of the method to biological samples for environmental analyses is, at present, lacking.

One of the limitations of SPME is the low capacity of the fiber and the perturbation of equilibria that can occur in the presence of sample components or analytes at very high concentration versus those of lesser concentration. Dilution of the sample can overcome some of these problems but not all, as limits of detection for trace analytes are compromised.

4.7 Matrix Solid-phase Dispersion

Another technique recently developed and applied to biological samples is MSPD. This technique is basically a simple two-step procedure for the extraction, cleanup, and isolation of pesticide (and other chemical) residues from any type of sample matrix (e.g. whole oyster homogenate, fish muscle, milk, fruits, vegetables, etc.). In general terms, the process involves blending a tissue sample (0.1–1.0 g) with lipophilic polymer-derivatized silica particles (e.g. ODS-derivatized silica (C_{18}), which simultaneously disrupts and disperses the sample. This mixture of C_{18} and tissue becomes part of a potentially multiphasic column that possesses unique chromatographic character. Elution of the MSPD column with a solvent or solvent sequence can provide a high-resolution fractionation of target analytes that can be further purified by simultaneous use of co-columns of Florisil, silica or alumina. The final eluate can, in most cases, be directly analyzed or further concentrated or manipulated to meet the demands of the individual analysis. The extracts obtained from these methods are most often detected by HPLC (in the case with drugs) or gas–liquid chromatography (GLC) with electron capture detection (ECD) or mass spectrometry (MS) (in the case with pesticides). However, they can also be used in immuno- or receptor assays.

Additionally, the MSPD process can be modified for a particular application by

1. a change in the ratio of sample to stationary phase support;
2. use of a different polarity polymer or solid support; and
3. blending of the C_{18}/tissue in the presence of modifiers such as chelators, acids, bases, etc.

MSPD could also be used in conjunction with SFE. The water in biological matrices often interferes with the SFE extraction process and analysts have used samples blended with diatomaceous earth to remove water from the sample. However, blending samples first with polymer-coated silicas, as is done in the MSPD process, would remove water and provide an initial stage of fractionation at the point of elution of the analytes with SC fluid and modifiers.

In most cases, eluate samples obtained after MSPD may be directly analyzed by GC without further cleanup. This technique is rapid, simple and inexpensive because it eliminates supplies and separate steps needed for:

1. sample preparation (e.g. blenders or tissue homogenizers);
2. lipid and pesticide extraction (e.g. large volumes of solvents and filtration steps);
3. extract cleanup and pesticide isolation (e.g. more solvents and perhaps adsorption chromatography); and
4. extract concentration (e.g. N_2 and various types of evaporation apparatus).

Three of its main advantages are that

1. it considerably decreases solvent use compared to the ''official'' methods;
2. it allows for a rapid turnover of samples and hence, access to relevant data on the pesticide residue levels present in the samples; and
3. it is amenable to robotics automation.

Table 9 Applications of MSPD to biological sample extraction and cleanup in environmental analyses

Analyte	Matrix	Ref.
PCBs	Fish	126
PCBs, pesticides	Fish	127
Pesticides	Beef fat	128
Pesticides	Catfish muscle	129
Pesticides	Crayfish	130
Pesticides	Fish	131
Pesticides	Fruit, vegetables	132
Pesticides	Milk	133
Pesticides	Oranges	134
Pesticides	Oysters	135
Pesticides	Vegetables	136
Pesticides, PCBs	Fish	137
Pyrethroids	Vegetables	138
Pesticides, PAHs	Aquatic species	31

As seen in Table 9, MSPD has been used to provide for the multiresidue analysis of various environmental contaminants in a variety of matrices.[126–138]

5 METHODS FOR THE FUTURE

It should be evident from this discussion that methods development for residue determination should focus on rapid screening tests, multiresidue capabilities, metabolite detection and improved sensitivity. The development and use of determinative methods for isolating the compound(s) of interest from biological samples must be rapid and inexpensive, and not generate large volumes of solvents for disposal. Classical or official isolation methods using homogenization and/or liquid–liquid partitioning of biological samples may be sufficient for some applications, such as the generation of data that will withstand a court challenge, but are poor for screening purposes because they are often lengthy, involve multiple steps and use large volumes of solvents. Such methods should not be used, as they are at present, to screen samples to determine a problem.

Solvent disposal is becoming increasingly expensive and environmentally unsound. Therefore, methods using low solvent volumes are desirable. A main purpose of this article was to present a case for phasing out existing official methods in favor of newer technologies that require less sample, less solvent, less employee time and less cost per sample. Newer techniques such as SC fluid extraction, SPE and MSPD offer alternative isolation strategies. When compared to the classical methods, these new methods greatly reduce labor and solvents costs and improve throughput. There are a few drawbacks to the new methods and more work is needed to further develop SPE, SFE, ASE, MAE, SPME and MSPD for use with the many different types of matrices that may contain residues of chemical contaminants. Nonetheless, these methods show tremendous potential. For example, MSPD, when compared to classical methods, has been estimated to reduce solvent use by approximately 98% and analysis time by 97%. Furthermore, once the MSPD column is prepared, the process of solvent elution, collection and analysis can be automated by the use of robotics. Cost of analysis is decreased because less solvent is needed and fewer laboratory technicians need to undergo training. Safety and environmental protection are increased because less solvent is needed. Finally, data are generated more quickly because of the ease of the process and its potential to be automated. These features of MSPD make it a general and perhaps significantly useful method in designing future residue analysis screening programs.

By reducing the sample size, smaller quantities of solvents are needed and less total lipid and co-eluates are present. However, less total analyte is also present. The advantages are that it requires less manipulation, if any, to cleanup the sample and that the capacities of subsequent SPE or adsorption chromatographic materials are not overwhelmed by the quantities of co-extracting interferences that are therein removed. This allows the final extract to be taken up in a smaller final volume, which compensates for the smaller sample size as compared to classical methods. Given that the regulatory levels for the chlorinated pesticides and many other pollutants are relatively high, the use of a smaller sample does not compromise the analytical result. Similarly, larger sample sizes (10 g) of biological materials have been extracted for chlorinated pesticide analysis by blending with solvent and cleaned up using a sequence of SPE columns to give excellent results.[39] These approaches offer a potentially automatable mechanism for the more rapid extraction of the chlorinated pesticides while simultaneously reducing solvent use, overall costs and data turn around time. The extracts obtained from these methods can be screened by gas–liquid chromatography/electron capture detection (GLC/ECD) or gas–liquid chromatography/mass spectrometry (GLC/MS). However, they could also be used in screening formats based on immuno- or receptor assays.

Biotechnology will be the field that will most greatly advance and revolutionize analytical chemistry in the near future. Theoretically, the chlorinated pesticides could be assayed by molecular probes or enzyme/receptor-based electrodes inserted directly into the sample. However, until such technology is fully available, we may apply the growing field of immunology-based detection and quantitation systems to assist us in solving the problems presently inherent in conducting environmental analysis. As described here, there are

presently several commercially available immunoassay kits for the screening and/or quantitation of some of the chlorinated pesticides, certain herbicides and other pollutants. Others could be developed if the market for them is created. Such technology, coupled with rapid isolation procedures such as MSPD or SPE, would allow the analyst to determine the presence and the relative level of contamination of hundreds of samples in a short time frame, especially in comparison to similar screening conducted for the same number of samples by GLC/ECD, GLC/MS, supercritical-fluid extraction/mass spectrometry (SFE/MS), or other instrumental methods of analysis. However, the reliance on such technology will require a complete understanding of the assay's specificity, sensitivity and susceptibility to false positive and/or false negative results. This is not a trivial task but can be accomplished with the proper conjunction of academic, industry, and regulatory agency input and support. The development of these technologies could also produce immunoaffinity column materials that could be used to provide a solventless extraction system for the overall analysis.

Since there is a large variety of biological samples that will require monitoring, it is unlikely that any one approach will be completely applicable to all environmental analyses. Each will require a degree of modification for a given application. However, it is evident that existing official methods should be phased out in favor of newer technologies that require less sample, less solvent, less employee time, and less cost per sample. This article offers a summary of the existing methodology for the analysis of biological samples for compounds of environmental concern and offers an opinion on the future directions for the field. In this age of environmental concern and the recognition of our environmental difficulties and our culpability in them, it is hoped that this information will serve to further the effort to define and ameliorate our past environmental misdeeds.

ACKNOWLEDGMENTS

The author wishes to acknowledge the many contributions of his graduate students to research on MSPD and its applications and their reviews of the literature related to the content of this article; Dr Patricia Gaunt, Dr Michael Crouch, Dr Calvin Walker and Dr Heidi Lott (deceased). Special acknowledgments are extended to Dr Austin Long, the co-inventor of MSPD during his tenure as a post-doctoral fellow, and to Dr John O'Rangers of the Center for Veterinary Medicine of the US FDA, the agency monitor for the research support awarded to our laboratory and gratefully acknowledged here.

ABBREVIATIONS AND ACRONYMS

AACC	American Association of Cereal Chemists
ACN	Acetonitrile
AOAC	Association of Official Analytical Chemists
AOCS	American Oil Chemists Society
ASE	Accelerated Solvent Extraction
CBA	Carboxymethyl
DCM	Dichloromethane
DDD	1,1-Dichloro-2,2-bis[*p*-chlorophenyl]-ethane
DDE	1,1-Dichloro-2,2-bis[*p*-chlorophenyl]-ethylene
DDT	1,1,1-Trichloro-2,2-bis[*p*-chlorophenyl]-ethane
DEA	Diethylaminopropyl
ECD	Electron Capture Detection
EMAP	Environmental Monitoring and Assessment Program
EQIA	Environmental Quality Improvement Act
FDA	Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FSIS	Food Safety and Inspection Service
FWPCA	Federal Water Pollution Control Act
GC	Gas Chromatography
GLC	Gas–Liquid Chromatography
GLC/ECD	Gas–Liquid Chromatography/Electron Capture Detection
GLC/MS	Gas–Liquid Chromatography/Mass Spectrometry
GLP	Good Laboratory Practice
GPC	Gel Permeation Chromatography
HPLC	High-pressure Liquid Chromatography
LC	Liquid Chromatography
LDIB	Laboratory Data Integrity Branch
MAE	Microwave-assisted Extraction
MPRSA	Marine Protection Research and Sanctuaries Act
MS	Mass Spectrometry
MSPD	Matrix Solid-phase Dispersion
NEPA	National Environmental Policy Act
NOAA	National Oceanic and Atmospheric Administration
NRP	National Residue Program
NS&T	National Status and Trends
OC	Organochlorine Compound
ODS	Octadecylsilyl
OP	Organophosphorus Compound
PAH	Polyaromatic Hydrocarbon
PAM	Pesticide Analytical Manual

PCB	Polychlorinated Biphenyl
PRS	Sulfonylpropyl
QA	Quality Assurance
QC	Quality Control
RCRA	Resource Conservation and Recovery Act
SAX	Triethylaminopropyl
SC	Supercritical
SCX	Benzenesulfonylpropyl
SFE	Supercritical-fluid Extraction
SFE/MS	Supercritical-fluid Extraction/Mass Spectrometry
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TSCA	Toxic Substances Control Act
USDA	United States Department of Agriculture
USEPA	United States Environmental Protection Agency
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Nitroaromatics, Environmental Analysis of • Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods • Organometallic Compound Analysis in Environmental Samples • Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Quality Assurance in Environmental Analysis • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Herbicide Residues in Biota, Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of

Liquid Chromatography **(Volume 13)**
Supercritical Fluid Chromatography

REFERENCES

1. J.W. Farrington, E.D. Goldberg, R.W. Ricebrough, J.H. Martin, V.T. Bowen, 'U.S. "Mussel Watch" 1976–1978: Overview of the Trace Metal, DDE, PCB, Hydrocarbon and Artificial Radionuclide Data', *Environ. Sci. Technol.*, **17**, 490–496 (1983).
2. E.D. Goldberg, V.T. Bowen, J. Farrington, G. Harvey, J.H. Martin, P. Parker, R. Ricebrough, R. Robinson, E. Schneider, E. Gamble, *The Mussel Watch*, United States Enviromental Protection Agency, Washington, DC, 1977.
3. S.D. Kilikidis, J.E. Psomas, A.P. Kamarianos, A.G. Panetsos, 'Monitoring of DDT, PCBs and other Organochlorine Compounds in Marine Organisms of the North Aegean Sea', *Bull. Environ. Contam. Toxicol.*, **26**, 496–501 (1981).
4. T.D. Sabourin, W.B. Stickle, T.C. Michot, C.E. Villars, D.W. Garton, H.R. Mushinsky, 'Organochlorine Residue Levels in Mississippi River Water Snakes in Southern Louisiana', *Bull. Environ. Contam. Toxicol.*, **32**, 460–468 (1984).
5. M.C. Broussard, Jr, 'Aquaculture: Opportunities for the 90's', *J. Animal Sci.*, **69**(10), 4221–4228 (1991).
6. R.B. Foster, 'Environmental Legislation', in *Fundamentals of Aquatic Toxicology*, eds. G.M. Rand, S.R. Petrocelli, Hemisphere Publishing, Washington, DC, 587–600, 1985.
7. C.L. Trichilo, R.D. Schmitt, 'Tolerance Setting Process in the U.S. Environmental Protection Agency', *J. Assoc. Off. Anal. Chem.*, **72**(3), 536–538 (1989).
8. P. Lombardo, 'The FDA Pesticides Programs: Goals and New Approaches', *J. Assoc. Off. Anal. Chem.*, **72**(3), 518–520 (1989).
9. I.J. Boyer, C.J. Kokoski, P.M. Bolger, 'Role of FDA in Establishing Tolerable Levels of Dioxin and PCBs in Aquatic Organisms', *J. Toxicol. Environ. Health*, **33**(1), 93–101 (1991).
10. D.V. Reed, P. Lombardo, J.R. Wessel, J.A. Burke, B. McMahon, 'The FDA Pesticide Monitoring Program', *J. Assoc. Off. Anal. Chem.*, **70**(3), 591–595 (1987).

11. E.L. Gunderson, 'FDA Total Diet Study, April 1982–April 1984; Dietary Intakes of Pesticides, Selected Elements and other Chemicals', *J. Assoc. Off. Anal. Chem.*, **71**(6), 1200–1209 (1988).
12. J.R. Wessel, N.J. Yess, 'Pesticide Residues in Foods Imported into the United States', *Rev. Environ. Contam. Toxicol.*, **120**, 83–104 (1991).
13. N.J. Yess, 'FDA Monitoring Program. Residues in foods – 1990', *J. Assoc. Off. Anal. Chem.*, **74**(5), 121A–141A (1991).
14. N.J. Yess, M.G. Houston, E.L. Gunderson, 'Food and Drug Administration Pesticide Residue Monitoring of Foods: 1978–1982', *J. Assoc. Off. Anal. Chem.*, **74**(2), 265–272 (1991).
15. J.P.J. Minyard, W.E. Roberts, W.Y. Cobb, 'State Findings on Pesticide Residues in Food – 1988 and 1989', *J. Assoc. Off. Anal. Chem.*, **72**(3), 438–452 (1991).
16. J.P.J. Minyard, W.E. Roberts. 'State Programs for Pesticide Residues in Foods', *J. Assoc. Off. Anal. Chem.*, **72**(3), 525–533 (1989).
17. P.B. Johnsen, 'Aquaculture Product Quality Issues: Market Position Opportunities Under Mandatory Seafood Inspection Regulations', *J. Animal Sci.*, **69**, 4209–4215 (1991).
18. P. Butler, 'Residues in Fish, Wildlife and Estuaries', *Pest. Monitor. J.*, **6**(4), 238–362.
19. P. Butler, R. Schutzman, 'Fish, Wildlife and Estuaries', *Pest. Monitor. J.*, **12**(2), 51–59 (1978).
20. P. Butler, C. Kennedy, R. Schutzman, 'Fish, Wildlife and Estuaries', *Pest. Monitor. J.*, **12**(3), 99–101 (1978).
21. R.L. Ellis, 'Changing Pesticide Technology in Meat and Poultry Products', *J. Assoc. Off. Anal. Chem.*, **72**(3), 521–524 (1989).
22. M.A. Norcross, A.R. Post, 'New Food Safety Initiatives in The Food Safety and Inspection Service, U.S. Department of Agriculture', *J. Animal Sci.*, **68**(3), 863–869 (1990).
23. T.P. O'Connor, 'Concentrations of Organic Contaminants in Mollusks and Sediments at NOAA National Status and Trend Sites in the Coastal and Estuarine United States', *Environ. Health Perspec.*, **90**, 69–73 (1991).
24. T.P. O'Connor, C.N. Ehler, 'Results from the NOAA National Status and Trends Program on Distribution and Effects of Chemical Contamination in the Coastal and Estuarine United States', *Environ. Monitor. Assess.*, **17**, 33–49 (1994).
25. J.J. Messer, R.A. Linthurst, W.S. Overton, 'An EPA Program for Monitoring Ecological Status and Trends', *Environ. Monitor. Assess.*, **17**, 67–78 (1991).
26. G. Ekstrom, M. Akerblom, 'Pesticide Management in Food and Water Safety: International Contributions and National Approaches', *Rev. of Environ. Contam. Toxicol.*, **114**, 23–55 (1990).
27. C.K. Winter, 'Dietary Pesticide Risk Assessment', *Rev. Environ. Contam. Toxicol.*, **127**, 23–67 (1992).
28. A.M. Fan, R.J. Jackson, 'Pesticides and Food Safety. Institution California Department of Health Services, Berkeley 94704. Abstract', *J. Regul. Toxicol. Pharmacol.*, **9**(2), 158–174 (1989).
29. R.J. Price, 'Residue Concerns in Seafood', *Dairy Food Env. San.*, **12**(3), 139–143 (1992).
30. H.M Lott, S.A. Barker, 'Chlorinated Pesticides' in *Analysis of Contaminants in Edible Aquatic Resources*, Elsevier, New York, 379–410, 1994.
31. M.D. Crouch, S.A. Barker, 'Analysis of Toxic Wastes in Tissues from Aquatic Species: Applications of Matrix Solid Phase Dispersion', *J. Chromatogr. A*, **774**(1–2), 287–309 (1997).
32. J. Sherma, 'Pesticides', *Anal. Chem.*, **61**, 153R–165R (1989).
33. P.A. Mills, 'Detection and Semiquantitative Estimation of Chlorinated Organic Pesticide Residues in Foods by Paper Chromatography', *J. Assoc. Off. Anal. Chem.*, **42**, 734–740 (1959).
34. P.A. Mills, J.H. Onley, R.A. Gaither, 'Rapid Method for Chlorinated Pesticides Residues in Nonfatty Foods', *J. Assoc. Off. Anal. Chem.*, **46**, 186–190 (1963).
35. J. Sherma, 'Papers that made a Difference', *J. Assoc. Off. Anal. Chem.*, **73**(1), 2–3 (1990).
36. J.A. Burke, 'Development of the Food and Drug Administration's Method of Analysis for Multiple Residues of Organochlorine Pesticides in Foods and Feeds', in *Pesticide Reviews*, ed. F.A. Gunther, Springer-Verlag, New York, 59–90, 1971.
37. F.M. Garfield, 'Sampling in the Analytical Scheme', *J. Assoc. Off. Anal. Chem.*, **72**, 405–411 (1989).
38. D.L. Park, A.E. Pohland, 'Sampling and Sample Preparation for Detection and Quantitation of Natural Toxicants in Food and Feed', *J. Assoc. Off. Anal. Chem.*, **72**, 399–404 (1989).
39. J. Sherma, *Manual of Analytical Quality Control for Pesticides and Related Compounds in Human and Environmental Samples, 2nd revision*, US Environmental Protection Agency, Office of Research and Development, Health Effects Research Laboratory: Research Triangle Park, NC, 1981.
40. P.A. Butler, C.D. Kennedy, R.L. Schutzmann, 'Pesticide Residues in Estuarine Mollusks, 1977 versus 1972 – National Pesticide Monitoring Program', *Pestic. Monit. J.*, **12**, 99–101 (1978).
41. P.A. Butler, 'Organochlorine Residues in Estuarine Mollusks, 1965–72 – National Pesticide Monitoring Program', *Pest. Monitor J.*, **6**, 238–362 (1973).
42. P.A. Butler, R.L. Schutzmann, 'Residues of Pesticides and PCBs in Estuarine Fish, 1972–76 – National Pesticide Monitoring Program', *Pest. Monitor J.*, **12**, 51–59 (1978).
43. US Environmental Protection Agency, *Analysis of Pesticide Residues in Human and Environmental Samples*, US Environmental Protection Agency, Health Effects

Research Laboratory, Environmental Toxicology Division: Research Triangle Park, NC, 1977.
44. S.D. Murphy, 'Toxic Effects of Pesticides', in *Toxicology – The Basic Science of Poisons*, 3rd edition, eds. C.D. Klaassen, M.O. Amdur, J. Doull, Macmillan, New York, 519–581, 1986.
45. US Food and Drug Administration, *Pesticide Analytical Manual, Vol. I*, US Food and Drug Administration, Washington, DC, 1977.
46. W. Ernst, H. Goerke, G. Eder, R.G. Schaefer, 'Residues of Chlorinated Hydrocarbons in Marine Organisms in Relation to Size and Ecological Parameters I. PCB, DDT, DDE, and DDD in Fishes and Mollusks from the English Channel', *Bull. Environ. Contam. Toxicol.*, **15**, 55–65 (1976).
47. US Department of Health and Human Services, US Public Health Service, *Fourth Annual Report on Carcinogens: Summary 1985*, National Toxicology Program: Research Triangle Park, NC, 1985.
48. S.M. Walters, 'Cleanup Techniques for Pesticides in Fatty Foods', *Anal. Chim. Acta*, **236**, 77–82 (1990).
49. B.G. Luke, J.C. Richards, E.F. Dawes, 'Recent Advances in Cleanup of Fats by Sweep Co-distillation. Part 1: Organochlorine Residues', *J. Assoc. Off. Anal. Chem.*, **67**, 295–298 (1984).
50. H.A. McLeod, P.J. Wales, 'A Low Temperature Cleanup Procedure for Pesticides and their Metabolites in Biological Samples', *J. Agr. Food Chem.*, **20**, 624–627 (1972).
51. W.D. MacLeod, Jr, D.W. Brown, A.J. Friedman, D.G. Burrows, O. Maynes, R.W. Pearce, C.A. Wigren, R.G. Bogar, *Standard Analytical Procedures of the NOAA National Analytical Facility, 1985–1986, Extractable Toxic Organic Compounds*, 2nd edition, National Marine Fisheries Service, Northwest and Alaska Fisheries Center, Environmental Conservation Division: Seattle, WA, 1985.
52. General Accounting Office, *Pesticides. Need to Enhance FDA's Ability to Protect the Public from Illegal Residues*, Report to Congressional Requesters, US General Accounting Office: Washington, DC, 1986.
53. P. Armishaw, R.G. Millar, 'Comparison of Gel Permeation Chromatography, Sweep Codistillation, and Florisil Column Adsorption Chromatography as Sample Cleanup Techniques for the Determination of Organochlorine Pesticide Residues in Animal Fats', *J. AOAC Int.*, **76**, 1317–1322 (1993).
54. A. Sannino, P. Mambriani, M. Bandini, L. Bolzoni, 'Multiresidue Method for Determination of Organophosphorus Insecticide Residues in Fatty Processed Foods by Gel Permeation Chromatography', *J. AOAC Int.*, **78**, 1502–1512 (1995).
55. Y. Hirama, M. Nakano, 'Sample Pretreatment for Analysis of Organochlorine Pesticide Residues in Meat and Fish by Gel Permeation Chromatography', *Hokkaidoritsu Eisei Kenkyushoho*, **42**, 54–57 (1992).
56. J. Daft, M. Hopper, D. Hensley, R. Sisk, 'A Displacement Pump Procedure to Load Extracts for Automated Gel Permeation Chromatography', *J. Assoc. Off. Anal. Chem.*, **73**, 992–994 (1990).
57. W. Specht, M. Tillkes, 'Gas Chromatographic Determination of Pesticide Residues after Cleanup by Gel-permeation Chromatography and Mini Silica Gel-column Chromatography', *J. Anal. Chem.*, **322**, 443–455 (1985).
58. H. Obana, S. Hori, 'Latest Analytical Methods for the Residual Pesticides in Foods', *Jpn. J. Toxicol. Environ. Health*, **42**, 1–16 (1996).
59. G. Fuchsbichler, 'Gel Permeation Chromatography, a Universal Cleanup Method for Determination of Pesticides', *Landwirtsch. Forsch.*, **36**, 130–139 (1983).
60. I. Saito, S. Yamada, H. Oshima, J. Hayakawa, 'Multiresidue Method for Determination of Pesticides using Luke Method Extraction, Gel Permeation Chromatography, and Gas Chromatography', *Nippon Noyaku Gakkaishi*, **20**, 109–118 (1995).
61. J.J. Vreuls, R.J.J. Swen, V.P. Goudriaan, M.A.T. Kerkhoff, G.A. Jongenotter, U.A.Th. Brinkman, 'Automated Online Gel Permeation Chromatography/Gas Chromatography for the Determination of Organophosphorus Pesticides in Olive Oil', *J. Chromatogr.*, **750**, 275–286 (1996).
62. J. Tekel', S. Hatrick, 'Pesticide Residue Analyses in Plant Material by Chromatographic Methods: Cleanup Procedures and Selective Detectors', *J. Chromatogr.*, **754**, 397–410 (1996).
63. D.M. Holstege, D.L. Scharberg, E.R. Tor, L.C. Hart, F.D. Galey, 'A Rapid Multiresidue Screen for Organophosphorus, Organochlorine, and *N*-Methyl Carbamate Insecticides in Plant and Animal Tissues', *J. AOAC Int*, **77**, 1263–1274 (1994).
64. Y. Nakamura, Y. Tsumura, Y. Tonogai, T. Shibata, 'Elution Patterns of Pesticides, Lipids, Chlorophyll, and Carotene in Gel Permeation Chromatography', *Shokuhin Eiseigaku Zasshi*, **37**, 329–336 (1996).
65. W. Specht, S. Pelz, W. Gilsbach, 'Gas-chromatographic Determination of Pesticide Residues after Clean-up by Gel-permeation Chromatography and Mini-silica Gel-column Chromatography', *Fresenius' J. Anal. Chem.*, **353**, 183–190 (1995).
66. M. Baldi, S. Coppi, M.C. Cristofori, M.L. Davi, S. Benedetti, A. Bovolenta, L. Penazzi, M.P. Previati, 'Application of Gel Permeation Chromatography in Purification of Organic Extracts for Determination of Trace Contaminants', *Boll. Chim. Ig., Parte Sci.*, **41**, 295–324 (1990).
67. M.L. Hopper, J.W. King, 'Enhanced Supercritical Fluid Carbon Dioxide Extraction of Pesticides from Foods using Pelletized Diatomaceous Earth', *J. Assoc. Off. Anal. Chem.*, **74**, 661–666 (1991).
68. S.J. Lehotay, C.H. Lee, 'Evaluation of a Fibrous Cellulose Drying Agent in Supercritical Fluid Extraction and Pressurized Liquid Extraction of Diverse Pesticides', *J. Chromatogr. A*, **785**(1–2), 313–327 (1997).

69. K.I. Eller, S.J. Lehotay, 'Evaluation of Hydromatrix and Magnesium Sulfate Drying Agents for Supercritical Fluid Extraction of Multiple Pesticides in Produce', *Analyst*, **122**(5), 429–435 (1997).
70. M.L. Hopper, J.W. King, J.H. Johnson, A.A. Serino, R.J. Butler, 'Multivessel Supercritical Fluid Extraction of Food Items in Total Diet Study', *J. AOAC Int.*, **78**(4), 1072–1079 (1995).
71. Y.Y. Wigfield, J. Selwyn, S. Khan, R. McDowell, 'Comparison of Supercritical Fluid Extraction and Solvent Extraction of Twenty-two Organochlorine Pesticides from Eggs', *Chemosphere*, **32**(5), 841–847 (1996).
72. S.J. Lehotay, K.I. Eller, 'Development of a Method of Analysis for 46 Pesticides in Fruits and Vegetables by Supercritical Fluid Extraction and Gas Chromatography/Ion Trap Mass Spectrometry', *J. AOAC Int.*, **78**(3), 821–830 (1995).
73. A. Valverde-Garcia, A.R. Fernandez-Alba, A. Aguera, M. Contreras, 'Extraction of Methamidophos Residues from Vegetables with Supercritical Fluid Carbon Dioxide', *J. AOAC Int.*, **78**(3), 867–873 (1995).
74. S.J. Lehotay, M.A. Ibrahim, 'Supercritical Fluid Extraction and Gas Chromatography/Ion Trap Mass Spectrometry of Pentachloronitrobenzene Pesticides in Vegetables', *J. AOAC Int.*, **78**(2), 445–452 (1995).
75. M.V. Djordjevic, D. Hoffmann, J. Fan, B. Prokopczyk, M.L. Citron, S.D. Stellman, 'Assessment of Chlorinated Pesticides and Polychlorinated Biphenyls in Adipose Breast Tissue using Supercritical Fluid Extraction Method', *Carcinogenesis*, **15**(11), 2581–2585 (1994).
76. H.J. Cortes, L.S. Green, R.M. Campbell, 'On-Line Coupling of Supercritical Fluid Extraction with Multidimensional Microcolumn Liquid Chromatography/Gas Chromatography', *Anal. Chem.*, **63**(23), 2719–2724 (1991).
77. J.E. France, J.W. King, 'Supercritical Fluid Extraction/Enzyme Assay: a Novel Technique to Screen for Pesticide Residues in Meat Products', *J. AOAC Int.*, **74**(6), 1013–1016 (1991).
78. S.J. Lehotay, C.-H. Lee, 'Evaluation of a Fibrous Cellulose Drying Agent in Supercritical Fluid Extraction of Diverse Pesticides', in *Supercritical Fluids in Chromatography and Extraction*, eds. R.M. Smith, S.B. Hawthorne, Elsevier, Amsterdam, 313–327, 1997.
79. M.E. McNally, J.R. Wheeler, 'Supercritical Fluid Extraction Coupled with Supercritical Fluid Chromatography for the Separation of Sulfonylurea Herbicides and their Metabolites from Complex Matrices', *J. Chromatogr.*, **435**(1), 63–71 (1988).
80. Y.C. Ling, H.C. Teng, 'Supercrtical Fluid Extraction and Clean-up of Organochlorine Pesticides and Polychlorinated Biphenyls in Mussels', *J. Chromatogr. A*, **790**(1–2), 153–160 (1997).
81. S.J. Lehotay, 'Supercritical Fluid Extraction of Pesticides in Foods', *J. Chromatogr. A*, **185**(1–2), 289–312 (1997).
82. M.L. Hopper, 'Extraction and Cleanup of Organochlorine and Organophosphorus Pesticide Residues in Fats by Supercritical Fluid Techniques', *J. AOAC Int.*, **80**(3), 639–646 (1997).
83. Z.M. Chen, Y.H. Wang, 'Chromatographic Methods for the Determination of Pyrethrin and Pyrethroid Pesticide Residues in Crops, Foods and Environmental Samples', *J. Chromatogr. A*, **754**(1–2), 367–395 (1996).
84. E.G. Alley, G. Lu, 'Analysis of Polychlorinated Biphenyls in Fatty Biological Matrices by On-line Supercritical Fluid Extraction and Supercritical Fluid Cleanup', *J. AOAC Int.*, **78**(4), 1051–1054 (1995).
85. B. Van Bavel, P. Dahl, L. Karlsson, L. Hardell, C. Rappe, G. Lindstom, 'Supercritical Fluid Extraction of PCBs from Human Adipose Tissue for HRGC/LRMS Analysis', *Chemosphere*, **30**(7), 1229–1236 (1995).
86. S. Bowadt, B. Johansson, P. Fruekilde, M. Hansen, D. Zilli, B. Larsen, J. De Boer, 'Supercritical fluid Extraction of Polychlorinated Biphenyls from Lyophilized Fish Tissue', *J. Chromatogr. A*, **675**(1–2), 189–204 (1994).
87. Z.V. Skopec, R. Clark, P.M. Harvey, R.J. Wells, 'Analysis of Organophosphorus Pesticides in Rice by Supercritical Fluid Extraction and Quantitation using an Atomic Emission Detector', *J. Chromatogr. Sci.*, **31**(11), 445–449 (1993).
88. A.L. Howard, C. Braue, L.T. Taylor, 'Feasibility of Thiocarbamate Analysis in Apples by Supercritical Fluid Extraction and High Performance Liquid Chromatography', *J. Chromatogr. Sci.*, **31**(8), 323–329 (1993).
89. A.G. Mills, T.M. Jefferies, 'Rapid Isolation of Polychlorinated Biphenyls from Milk by a Combination of Supercritical Fluid Extraction and Supercritical Fluid Chromatography', *J. Chromatogr.*, **643**(1–2), 409–418 (1993).
90. R.C. Hale, J. Greaves, 'Methods for the Analysis of Persistent Chlorinated Hydrocarbons in Tissues', *J. Chromatogr.*, **580**(1–2), 257–278 (1992).
91. D.E. Games, H. Hernandez, J.R. Perkins, E.D. Ramsey, D. Thomas, 'Supercritical Fluid Chromatography, Supercritical Fluid Extraction and their Combination with Mass Spectrometry for the Analysis of Pesticides and Veterinary Drugs', *Acta Vet. Scand. Suppl.*, **87**, 471–472 (1991).
92. T.L. Chester, J.D. Pinkston, D.E. Raynie, 'Supercritical Fluid Chromatography and Extraction', *Anal. Chem*, **66**, 106R–130R (1994).
93. H. Obana, K. Kikuchi, M. Okihashi, S. Hori, 'Determination of Organophosphorus Pesticides in Foods Using an Accelerated Solvent Extraction System', *Analyst*, **122**(3), 217–220 (1997).
94. A. Zlotorzynski, 'The Application of Microwave Radiation to Analytical and Environmental Chemistry', *Crit. Rev. Anal. Chem.*, **25**(1), 43–76 (1995).
95. K. Ganzler, A. Salgo, K. Valko, 'Microwave Extraction. A Novel Sample Preparation Method for Chromatography', *J. Chromatogr.*, **371**, 299–306 (1986).

96. Y.L. Ren, J.M. Desmarchelier, 'Release of Fumigant Residues from Grain by Microwave Irradiation', *J. AOAC Int.*, **81**(3), 673–678 (1998).
97. R.D. McDowall, 'Sample Preparation in Biomedical Analysis', *J. Chromatogr.*, **492**, 3–58 (1989).
98. E. Jaervenpaeae, R. Huopalahti, P. Tapanainen, 'Use of Supercritical Fluid Extraction/High Performance Liquid Chromatography in the Determination of Polynuclear Aromatic Hydrocarbons from Smoked and Broiled Fish', *J. Liq. Chromatogr. Rel. Technol.*, **19**, 1473–1482 (1996).
99. P.J. Nyman, G.A. Perfetti, F.L. Joe, G.W. Diachenko, 'Comparison of Two Clean-up Methodologies for the Gas Chromatographic/Mass Spectrometric Determination of Low Nanogram/Gram Levels of Polynuclear Aromatic Hydrocarbons in Seafood', *Food Addit. Contam.*, **10**, 489–501 (1993).
100. G.A. Perfetti, P.J. Nyman, S. Fisher, F.L. Joe, G.W. Diachenko, 'Determination of Polynuclear Aromatic Hydrocarbons in Seafood by Liquid Chromatography with Fluorescence Detection', *J. AOAC Int.*, **75**, 872–877 (1992).
101. J. Rivera, M.J.C. Curto, P. Pais, M.T. Galceran, L. Puignou, 'Solid-phase for the Selective Isolation of Polycyclic Aromatic Hydrocarbons, Azaarenes and Heterocyclic Aromatic Amines in Charcoal-grilled Meat', *J. Chromatogr. A*, **731**, 85–94 (1996).
102. K. Ono, Y. Yamamoto, T. Kosaka, O. Takeda, 'A Simple Method for Pesticide Residue Estimation in Beef, using a Solid-phase Extraction Minicolumn Cartridge', *Miyazaki-ken Eisei Kankyo Kenkyusho Nenpo*, **5**, 51–57 (1994).
103. G. Pavoni, E. Errani, 'Rapid Solid-phase Extraction/Gas Chromatography (SPE/GC) Method for Determination of Organophosphorus Pesticide Residues in Cereal Flours', *Boll. Chim. Ig., Parte Sci.*, **47**, 143–145 (1996).
104. S. Mouratidis, H.-P. Thier, 'Solid Phase Extraction for the Confirmation of Results in Polar Pesticides Residue Analysis by HPTLC', *Z. Lebensm.-Unters. Forsch.*, **201**, 327–330 (1995).
105. F.J. Schenck, R. Wagner, M.K. Hennessy, J.L. Okrasinski, 'Screening Procedure for Organochlorine and Organophosphorus Pesticide Residues in Eggs using a Solid-phase Extraction Cleanup and Gas Chromatographic Detection', *J. AOAC Int*, **77**, 1036–1040 (1994).
106. E.B. Asafu-Adjaye, P.W. Kohler, S.Y. Su, 'Retention Behavior and Solvent Optimization of Organochlorine Pesticides in Solid-phase Extraction', *Chromatographia*, **21**, 435–438 (1986).
107. F.J. Schenck, L. Calderon, L.V. Podhorniak, 'Determination of Organochlorine Pesticide and Polychlorinated Biphenyl Residues in Fatty Fish by Tandem Solid-phase Extraction Cleanup', *J. AOAC Int.*, **79**, 1209–1214 (1996).
108. R.C. Hsu, I. Biggs, N.K. Saini, 'Solid-phase Extraction Cleanup of Halogenated Organic Pesticides', *J. Agric. Food Chem.*, **39**, 1658–1666 (1991).
109. Y. Akiyama, M. Yano, T. Mitsuhashi, N. Takeda, M. Tsuji, 'Simultaneous Determination of Pesticides in Agricultural Products by Solid-phase Chromatography–Mass Spectrometry', *Shokuhin Eiseigaku Zasshi*, **37**, 351–362 (1996).
110. C. Chen, 'Application of Solid Phase Extraction to Pesticide Determination', *Shanghai Huanjing Kexue*, **11**, 31–32 (1992).
111. F.J. Schenck, L. Calderon, D.E. Saudarg, 'Florisil Solid-phase Extraction Cartridges for Cleanup of Organochlorine Pesticide Residues in Foods', *J. AOAC Int.*, **79**, 1454–1458 (1996).
112. W.H. Weber, E. Klein, 'Solid-phase Extraction in Food and Environmental Analysis', *Lebensmittelchem. Gerichtl. Chem.*, **43**, 75–77 (1989).
113. K.-C. Ting, G.S. Tamashiro, 'Off-line High-performance Liquid Chromatography and Solid-phase Extraction Clean-up for Confirmation of Pesticide Residues in Fresh Produce by Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **754**, 455–462 (1996).
114. A. De Kok, M. Hiemstra, 'Optimization, Automation, and Validation of the Solid-phase Extraction Cleanup and On-line Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides in Fruits and Vegetables', *J. AOAC Int.*, **75**, 1063–1072 (1992).
115. J.A. Casanova, 'Use of Solid-phase Extraction Disks for Analysis of Moderately Polar and Nonpolar Pesticides in High-moisture Foods', *J. AOAC Int.*, **79**, 936–940 (1996).
116. D. Cardinali, 'Use of SPE (Solid-phase Extraction) in the Determination of Fungicides, Insecticides, Chlorinated Acaricides and Pyrethroids, in Legumes and Fruits, using Capillary GLC', *Riv. Soc. Ital. Sci. Aliment.*, **21**, 461–469 (1992).
117. R.S. Bakowski, M.S. Ali, J.D. White, E.T. Phillippo, R.L. Ellis, 'Gas Chromatographic/Mass Spectrometric Confirmation of Ten *N*-Methylcarbamate Insecticides in Liver', *J. AOAC Int.*, **77**, 1568–1574 (1994).
118. M.A.L. Zapata, R.M.M. Santamaria, I.M. Alvarez, V.S. Salazar, U. Muller, 'Organochlorine Pesticide Residues in Cow's Milk', *Bol. Oficina. Sanit. Panam.*, **120**, 483–490 (1996).
119. Y. Pico, E. Viana, G. Font, J. Manes, 'Determination of Organochlorine Pesticide Content in Human Milk and Infant Formulas using Solid Phase Extraction', *J. Agric. Food Chem.*, **43**, 1610–1615 (1995).
120. A.M. Gillespie, S.L. Daly, D.M. Gilvydis, F. Schneider, S.M. Walters, 'Multicolumn Solid-phase Extraction Cleanup of Organophosphorus and Organochlorine Pesticide Residues in Vegetable Oils and Butterfat', *J. AOAC Int.*, **78**, 431–437 (1995).
121. A.M. Gillespie, S.M. Walters, 'Rapid Cleanup of Fat Extracts for Organophosphorus Pesticide Residue Determination Using C_{18} Solid-phase Extraction Cartridges', *Anal. Chim. Acta*, **245**, 259–265 (1991).

122. P.W. Kohler, S.Y. Su, 'Determination of Some Organochlorine Pesticides in Seafood Samples by Solid-phase Extraction–Gas Chromatography', *Chromatographia*, **21**, 531–537 (1986).
123. F.J. Schenck, 'Screening of Nonfatty Fish for Organochlorine Pesticide Residues by Solid-phase Extraction Cleanup', *J. AOAC Int.*, **79**, 1215–1219 (1996).
124. S. Noegrohati, W.E. Hammers, 'Cleanup by Solid-phase Extraction and HPLC of Organochlorine Insecticides and PCBs in Various Nonfatty and Fatty Samples', *Toxicol. Environ. Chem.*, **34**, 219–235 (1992).
125. F.J. Schenck, R. Wagner, M.K. Hennessy, J.L. Okrasinski, 'Screening of Organochlorine Pesticide and Polychlorinated Biphenyl Residues in Nonfatty Seafood Products by Tandem Solid-phase Extraction Cleanup', *J. AOAC Int.*, **77**, 102–106 (1994).
126. Y.-C. Ling, M.-Y. Chang, I.-P. Huang, 'Matrix Solid-phase Dispersion Extraction and Gas Chromatographic Screening of Polychlorinated Biphenyls in Fish', *J. Chromatogr. A*, **669**, 119–124 (1994).
127. Y.-C. Ling, I.-P. Huang, 'Multiresidue Matrix Solid Phase Dispersion Method for Determining 16 Organochlorine Pesticides and Polychlorinated Biphenyls in Fish', *Chromatographia*, **40**, 259–266 (1995).
128. A.R. Long, L. Hsieh, C.R. Short, S.A. Barker, 'Matrix Solid Phase Dispersion (MSPD) Extraction and Gas Chromatographic Screening of Nine Chlorinated Pesticides in Beef Fat', *J. Assoc. Off. Analyt. Chemists*, **74**, 493–496 (1991).
129. A.R. Long, L. Hsieh, C.R. Short, S.A. Barker, 'Multiresidue Matrix Solid Phase Dispersion (MSPD) Extraction and Gas Chromatographic Screening of Nine Chlorinated Pesticides in Catfish (*Ictalurus punctatus*) muscle tissue', *J. Assoc. Off. Analyt. Chemists.*, **74**, 667–670 (1991).
130. H.M. Lott, S.A. Barker, 'Extraction and Gas Chromatographic Screening of 14 Chlorinated Pesticides in Crayfish (*Procambarus clarkii*) Hepatopancreas', *J. AOAC Int.*, **76**, 663–668 (1993).
131. H.M. Lott, S.A. Barker, 'Comparison of a Matrix Solid Phase Dispersion and a Classical Method for the Determination of Chlorinated Pesticides in Fish Muscle', *Environ. Monit. Assessment*, **28**, 109–116 (1993).
132. C.M. Torres, Y. Pico, J. Manes, 'Analysis of Pesticide Residues in Fruit and Vegetables by Matrix Solid Phase Dispersion', *Chromatographia*, **41**, 685–692 (1995).
133. F.J. Schenck, R. Wagner, 'Screening Procedure for Organochlorine and Organophosphorus Pesticide Residues in Milk using Matrix Solid Phase Dispersion Extraction and Gas Chromatographic Determination', *Food Add. Contam.*, **12**, 535–541 (1995).
134. C.M. Torres, Y. Pico, M.J. Redondo, J. Manes, 'Matrix Solid-phase Dispersion Extraction Procedure for Multiresidue Pesticide Analysis in Oranges', *J. Chromatogr.*, **719**, 95–103 (1996).
135. H.M. Lott, S.A. Barker, 'Matrix Solid Phase Dispersion (MSPD) Extraction and Gas Chromatographic Screening of 14 Chlorinated Pesticides in Oysters (*Crassostrea virginica*)', *J. AOAC Int.*, **76**, 67–72 (1993).
136. E. Viana, J.C. Molt, G. Font, 'Optimization of a Matrix Solid-phase Dispersion Method for the Analysis of Pesticide Residues in Vegetables', *J. Chromatogr.*, **754**, 437–444 (1996).
137. R. Yang, C. Fu, 'Matrix Solid Phase Dispersion and High Performance Liquid Chromatographic Determination of Trace Pirimicarb and Amitraz in Fruits and Vegetables', *Hebei Daxue Xuebao, Ziran Kexueban*, **14**, 29–33 (1994).
138. Y.-C. Ling, I.-P. Huang, 'Multiresidue Matrix Solid Phase Dispersion Method for the Determination of Six Synthetic Pyrethroids in Vegetables Followed by Gas Chromatography with Electron Capture Detection', *J. Chromatogr.*, **695**, 75–82 (1995).

Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis

Xiaoru Wang
Xiamen University, Xiamen, China

Frank Sen-Chun Lee
Hong Kong Baptist University, Hong Kong, China

1 Introduction 1
2 Capillary Electrophoresis in Speciation Analysis 2
3 Optimization of Capillary Electrophoresis Separation 4
4 Detection Systems for Capillary Electrophoresis 6
5 Capillary Electrophoresis/Inductively Coupled Plasma Mass Spectrometry Systems 7
5.1 General Consideration in Interface Design 8
5.2 Glass Concentric Nebulizer 9
5.3 Direct Injection Nebulizer 10
5.4 Ultrasonic Nebulizer 10
5.5 Oscillating Capillary Nebulizer 10
5.6 Hydride Generation 11
Acknowledgments 12
Abbreviations and Acronyms 12
Related Articles 12
References 12

Speciation analysis is a term commonly used to describe the analytical processes of identifying and measuring the different chemical forms of trace elements in the sample. Starting from the mid-1980s, analytical techniques for elemental speciation analysis have been advancing at an accelerated pace owing to the rapid advancement in analytical instrumentation. Such advancement is especially evident in the development of hyphenated techniques, which involve the on-line coupling of a species-specific separation device with an elemental-specific detection system. They are now well accepted as the methods of choice for speciation analysis because of their speed, sensitivity, specificity and resolution. Capillary electrophoresis/inductively coupled plasma mass spectrometry (CE/ICPMS) is one of the newest members of this family of hyphenated techniques which has found increasing applications in speciation analysis in recent years. Until today, chromatography has been the dominant separation technique in hyphenated techniques and the subject has been extensively reviewed. However, literature information on capillary electrophoresis (CE)-based hyphenated techniques is much less available. In this chapter, the current status of CE/ICPMS will be reviewed with emphasis being placed on the interface design, analytical performance and applications of the different versions of the technique in speciation analysis.

1 INTRODUCTION

Trace elements play a key role in the functioning of life on earth. They can be either highly toxic or essential to life depending not only on their total concentration levels, but just as important, also on the particular chemical forms in which they exist in the system. The latter information is of additional significance in environmental studies since, besides biological and toxicological concern, the transport and fate of trace elements in the environment is also highly 'species' dependent.

Speciation analysis is a term commonly used to describe the analytical processes of identifying and measuring the different chemical forms of trace elements in the sample. Starting from the mid-1980s, analytical techniques for elemental speciation analysis have been advancing at an accelerated pace owing to the rapid advancement in analytical instrumentation. Such advancement is especially evident in the development of hyphenated techniques, which involve the on-line coupling of a species-specific separation device with an elemental-specific detection system. They are now well accepted as the methods of choice for speciation analysis because of their speed, sensitivity, specificity and resolution. CE/ICPMS is one of the newest members of this family of hyphenated techniques which has found increasing applications in speciation analysis in recent years. Until today, chromatography has been the dominant separation technique in hyphenated techniques and the subject has been extensively reviewed. However, literature information on CE-based hyphenated techniques is much less available.

In this chapter, we attempt to provide a capsule review of the current development of CE/ICPMS technique. The emphasis will be placed on the interface design, analytical performance and applications of the different versions of the technique in speciation analysis. The principles and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

instrumental details of the individual CE and inductively coupled plasma mass spectrometry (ICPMS) instruments themselves will not be covered since the subjects can be readily found in many books or review articles.

2 CAPILLARY ELECTROPHORESIS IN SPECIATION ANALYSIS

The species of principle interest in speciation analysis include organometallic species, biological macromolecules, and organic or inorganic complexes of metals (Pb, Hg, Cr, Cu) or non-C elements (As, Ge, S). In environmental samples, these species are usually present at trace levels in highly complex matrices. For such analysis, the use of hyphenated techniques in an on-line fashion is preferred because the closed systems employed minimize the risk of contamination and losses of analysts during analysis. Among the variety of hyphenated techniques developed, the coupling of chromatography-based separation with atomic spectrometric detection are the most widely used ones.

Chromatography in its various forms including gas chromatography (GC), ion chromatography (IC) or high-performance liquid chromatography (HPLC) has been used extensively as the separation technique in speciation studies.[1–3] In chromatographic separation, however, the stationary phases often interact with the analytes, resulting in the alteration of the original distribution of species in the sample. This problem is especially serious in the speciation of inorganic, metal-containing analytes because most of them are not amenable to chromatographic columns. In CE, the electrophoretic separation is applied in free solutions. There is minimum interaction of the analytes with the capillary wall or the stationary phases, since the separation capillaries are mostly made of inert fused silica tubing. Thus, the preservation of the original bondings of the species is often achievable with the proper selection of buffer solutions. Since the separation of analytes is carried out in the aqueous phase, it allows the analysis of labile species by direct injection of the sample without prior derivatization, as is often needed in GC or HPLC separations. This is particularly useful for the analysis of aqueous samples since derivatization could result in sample losses due to incomplete reactions or handling.

The separation mechanisms of CE and chromatography are completely different; the two are therefore complementary to each other in speciation applications. The versatility of CE is demonstrated by its ability to separate a wide range of ionic or neutral species including small inorganic cations or anions, organometallic compounds, organic molecules as well as large biological molecules such as peptide and proteins.[4] The ability of CE to analyze large biological molecules is another crucial factor stimulating our interest in CE because the bindings of metals to peptides and protein enzymes in plant, animal and human tissues have important ecotoxicological implications in environmental studies.

The other attractive features of CE are that the technique is straightforward in principle, easy to operate and simple in experimental set-up.[5] In a typical experiment, the components of a mixture are transported through a horizontal capillary tube by applying a high d.c. potential. A buffer-filled capillary is placed between two containers filled with the same buffer. Typically the tubing is fashioned from fused silica or inert polymeric tubing with a length of 50 to 100 cm and an inside diameter of 25 to 100 μm. Platinum foil electrodes in the two buffer vessels are connected to a d.c. power supply capable of developing a potential of 20 to 30 kV. Electroosmotic flow (EOF) occurs in the capillary where the solvents move from the vessel containing the positive electrode to the one containing the negative. The cause of EOF flow is the electric double layer that develops at the silica capillary surface. The fixed negative charges on the capillary surface arise from the dissociation of functional groups on the surface of the capillary. The negative surface charge attracts positive ions from the buffer solution, thus giving a typical double layer structure. The mobile positive ions that ring the interior surface of the capillary tubing are attracted to the negative electrode, carrying with them the solvent molecules. The analytes are separated by their differences in electrophoretic mobility, with the positive species moving towards negative electrode at velocities faster than EOF; while the negative ones moving slower than EOF. The neutral species, which move with the EOF flow, can also be made to separate by the selective addition of surfactants or other modifiers to alter the surface charge of the analytes.[5,6]

Of the different modes of CE separation techniques, CZE (capillary zone electrophoresis) is the most widely used in elemental speciation analysis. Molecular separation in CZE is based on the difference in the electrophoretic mobilities of the species, which is proportional to the charge of the analyte, and inversely proportional to the frictional forces encountered by the analytes during elution. The latter are determined by the hydrodynamic size of the analytes and viscosity of the medium. Solvent properties such as pH, ionic strength, viscosity are all important parameters because they affect the effective charge on the solute; and for large molecules, also their effective hydrodynamic sizes. CZE has been used widely for the separation of inorganic cations, anions as well as organometallic species. Since cations with the same charge and hydrodynamic radius are often inseparable, a weakly complexing ligand can be added to enhance

Table 1 Speciation analysis by CE with conventional ultraviolet/visible (UV/VIS) detectors

Analyte	Operation mode	Capillary (i.d. × effective length)	Buffer solution	Applied voltage	Sample	Reference (Author)
Hg(II), MMHg, MEHg, MPhHg	CZE	75 μm × 45 cm	Cysteine, 100 mM sodium borate (pH 8.35), 10% methanol	15	Mussel	25 (Medina)
MMHg	CZE	75 μm × 56 cm polymide-coated	Cysteine, 200 mM borate buffer (pH 8.11)	15	Tuna mussel	26 (Diaz)
As(III), As(V), Se(IV), Se(VI)		75 μm × 50 cm polyimide-clad	5 mM chomate (pH 10.0), 0.25 mM HTAB	−15	Standard	27 (Vogt)
Gd-DTPA	CZE	50 μm × 50 cm polyimide-clad	25 mM phosphate (pH 6.8), 10% acetonitrile	25	Standard	27 (Vogt)
Cu-BSA, Cu-MYO		25 μm × 20 cm polyimide-coated	0.01 mol L^{-1} phosphate (pH 2.5)	6	Standard	28 (O'Keeffe)
As(III), As(V), DMA, MMA, BzAs	CZE	50 μm × 52 cm	25 mM phosphate (pH 8)	25	Standard	29 (Rocca)
Se(IV), Se(VI), SeCys, SeMet	CZE	50 μm × 63 cm	80 mM phosphate (pH 8.5), 2 mM TTAB	−12	Standard	29 (Rocca)
As(V), MMA, DMA, As(V)	CZE	50 μm × 75 cm	6 mM chromate solution, pH10, NICE-Pak OFM Anion-BT (1 : 40 v/v)	−28	Coal fly ash	30 (Caruso)
As(III), DMA, Se(IV), AS(V)	CZE	75 μm × 50 cm	15 mM phosphate	25	Standard	31 (Schlegel)
Se(VI), AsF6(V), Se(IV), As(V), PhAs, ABAs, DMA	CZE	75 μm × 70 cm	1 mM CTAB, 50 mM CHES, Triton X-100 0.03 wt%, LiOH 20 mM, pH 9.4	−30	Standard	31 (Schlegel)
As(III), As(V), DMA, MMA	CZE	50 μm × 72 cm	25 mM phosphate buffer, pH 5.6	30	Standard	32 (Leroy)
Fe(II), Fe(III)	CZE	50 μm × 30 cm	100 mM borate, pH 9.2, 10 mM *o*-phenanthroline and EDTA	30	Standard	33 (Schaffer)
Ferrocenecarbozaldehyde, ferrocene, butylferrocene	MEKC	75 μm × 70 cm	10 mM phosphate pH 6.7, 30 mM SDS, 35% acetonitrile	25	Standard	27 (Vogt)
SO_4^{2-}, $S_2O_3^{2-}$, $S_4O_6^{2-}$, HS^-	CZE	75 μm × 60 cm	1.5 mM PM, 10 mM TRIS, 0.5 mM DETA, pH 8.0	−20	Standard	34 (Motellier)
Metallothinoeins	CZE	75 μm × 94 cm polyimide-clad	50 mM Tris-GCl pH 9.1		Rabbit liver	35 (Richards)
Ca- and Zn-binding proteins	CZE	50 μm × 100 cm	100 mM Tris-100 mM tricine (pH 8.3), or 50 mM Tris-384 mM glycine pH 8.3 containing 2 mM $CaCl_2$, $ZnCl_2$ and EDTA	30	Proteins	36 (Kajiwara)
Ca- and Zn-binding proteins	MEKC	50 μm × 100 cm	100 mM Tris-100 mM tricine (pH 8.3), 0.1% SDS	30	Standard	36 (Kajiwara)
Cu, Zn haemeto- and protoporphyrin	MEKC	50 μm × 50 cm	40 mM SDS in CAPS buffer (pH 11)-DMF (10 : 2, v/v)	30	Standard	37 (Kiyohara)
TML, TEL, DPhSe, MphSe	MEKC	50 μm × 44 cm	50 mM SDS and 5 mM β-cyclodextrin in 25 mM phosphate borate buffer, pH 6.0	15	Standard	38 (Ng)
Fe(II)- phenols, tanning agents	CZE	(25–100) μm × (12–100) cm polymer-coated	Borate/HCl (pH 8.5)	12	Beverage	39 (Harms)

DMA, dimethylarsenic acid; MMA, monomethylarsenic acid.

selectivity and separation efficiency.[7,8] Many organic compounds with electron-donating properties can be used as weak charge transfer complexes with metals for such applications. Significant advancement has been made to exploit various chemical techniques of complexation, ion-pairing, and micellization interactions between the metal analytes and electrolyte additives to enhance separation or alter selectivity.[9]

Conventional reversed-phase liquid chromatography (RPLC) ion-pairing reagents, ionic polymers and dicarboxylic acids have all been effectively used to enhance CZE separation. A number of reviews are available discussing the use of CZE for speciation analysis of metals in different oxidation state, organometallic compounds, and metal complexes with different organic or inorganic ligands.[7–15] In these applications, the proper selection of operational electrolytes to control acid–base and complex formation equilibria is the most important in controlling the effective electrophoretic mobilities of the analytes. Such applications are especially useful for the determination of transition metals, heavy metals and rare earth elements.

In addition to cations, CZE has also been used extensively for the separation and speciation of inorganic anions of arsenic, selenium, phosphorus, sulfur, halogen and phosphorus compounds. The mechanisms of anion separation by CE were reviewed by Kaniansky et al.,[13] and Haddad et al.[14] Comparison between CE and RPLC techniques for metal speciations was reviewed by Timerbaev et al.[11] The application of CE for the separation of inorganic metals and organometallic species in environmental samples including air particulates, soils and water samples was discussed in several recent reviews.[16,17] The main emphasis of these reviews was placed on the discussion of matrix effects in complex environmental samples; and the different approaches one can take to minimize such matrix interference.

For the separation of organometallic or uncharged organic or biological molecules, MEKC (micellar electrokinetic chromatography) is the other CE technique commonly used.[18–22] A recent review on the use of micelles as a separation media in both HPLC and CE applications was given by Khaledi.[23] The set-up of the apparatus in MEKC is the same as those of CZE. The chief advantage of MEKC is that both neutral and charged compounds can be separated. Terabe et al.[18] first reported the use of micelles in buffer solutions for CE in 1984. The charged micelles are incorporated in the buffer solution to act as a pseudo-stationary phase. The separation of the analytes is based on the differences in their partition coefficients between the buffer solution and the micelles. Such differences result in the separation of species, which by themselves would have the same electrophoretic mobilities, because of the electrophoretic migration of the micelles along the capillary. Furthermore, with the inclusion of suitable additives in the buffer, unique separation or selectivity may be exploited in MEKC.[20,23]

MEKC has found wide applications for the separation of metal binding proteins, peptides or porphyrins. For the separation of these biopolymers, conditioning or shielding of the fused silica capillary wall is often necessary in order to avoid undesirable wall interactions of the analytes. Two approaches can be taken. The first is to apply dynamic coatings, with deactivator compounds added in the solution to continuously condition the capillary during separation. Alternatively, a permanently coated or organic polymeric capillaries can be used to replace the fused silica capillary commonly used.[24]

The application of CE for speciation studies started to grow rapidly in the early 1990s. Most of these studies used on-line spectrophotometric or fluorescence detectors. Limited by detection sensitivity, this earlier work involved primarily the analysis of standard solutions or samples with higher metal concentrations. Table 1 lists selectively some of these reports to illustrate the range of application of CE techniques. The primary CE modes used in these studies were CZE and MEKC. All the listed studies were based on CE with UV/VIS detectors except the one by Schlegel, in which conductivity detector was used. The operating conditions, buffer systems used for each of these studies are included in the table along with sample and analysts information. The versatility of CE separation can be seen from its wide applicability in systems consisting of a variety of metals, anions and organometellic species in different types of sample matrices. The capability of CE is further enhanced when it is coupled with sensitive and elemental specific detectors such as plasma-based atomic spectrometry or mass spectrometry (MS), as will be discussed in later sections.

3 OPTIMIZATION OF CAPILLARY ELECTROPHORESIS SEPARATION

To achieve high efficiency in CE separation requires the careful control and systematic optimization of the various operating parameters in the system. The injection mode, pH control of the buffer solution, stability of the applied voltage, conductivity and electric current of the buffer system are all important factors contributing to the resolution and reproducibility of the separation.

The type of injection and injection duration are important in CE as it may bias the results and lead to poor separation and irreproducibility of migration times.[40] The two basic sample injection techniques in CE are hydrodynamic injection and electrokinetic injection. The former is to introduce the sample by hydrodynamic flow through hydraulic pressure, while the latter is by

electrophoretic flow with applied d.c. potential. The volume of sample injected is a compromise between separation resolution and sensitivity requirement. Quite often in speciation studies, the concentration-based detection limits in CE have to be maximized through the injection of large sample volumes. However, increasing the injection volume risks the possibility of degradation in resolution because of overloading.

For large sample volumes, using electrokinetic injection will give a better resolution than hydrodynamic injection as the sample is continuously being separated during the entire injection period. Magnuson et al.[41] reported enhanced sensitivity with detection limits in the ppt range for arsenic species of As (III) and As (V) by CE/ICPMS using electrokinetic injection. However, such injection often complicates the quantitation process as the amounts of individual analytes introduced during the sample injection period may vary depending on their respective electrophoretic mobilities and conductivities.[40] Generally speaking, electrokinetic injection is a viable sample introduction technique which is more suitable for the analysis of samples requiring on-line preconcentration. But its precision and accuracy are strongly affected by experimental conditions and thus require tighter control compared with that of hydrodynamic injection.

Reproducible sample injection is crucial to obtain quantitative data in applying CE for speciation studies. Recently, a number of modified CE injection techniques based on flow-injection principles or automated devices have been developed.[42–44] These injectors offer the opportunity for fully automated sample injection with high precision, and thus are preferred for applications involving CE hyphenated techniques for speciation studies.

The conditioning of the surface of the capillary is also important in achieving reproducible CE separation. The pH of the buffer system has a direct impact on the EOF because buffer pH affects the charge of the inner capillary tube surface. This is because the degree of protonation of the silanol groups increases with decreasing pH. Phosphate or other anions commonly used in buffer systems can also reduce EOF through the conversion of acidic silanols on surface to more easily protonated silica–phosphate complexes.[45] Preconditioning of the capillary is therefore important to control the equilibrium of the surface charge.[45,46] The use of CZE for the separation of arsenic species of As (III), As (V), MMA, DMA and arsenilic acid (ANA) was found to depend significantly on capillary preconditioning and pH.[47–49]

The separation efficiency of CE is related greatly to pH and the choice of buffer because they affect the electrophoretic mobility and separation efficiency of the analytes. During separation, charged species migrate in the capillary under the influence of an applied electric field. Ideally, the buffer should not interact chemically with the species of interest, maintain the pH and minimize ion loss. The choice of buffer often involves trade-offs between separation efficiency and peak resolution. In cation analysis, for example, cations with a high charge will be more easily trapped in the electrical double layer than cations with smaller charge. Thus, using a buffer containing a cation with a high charge will reduce analyte cation sample loss caused by trapping in the double layer. On the other hand, the higher charged ions are likely to interact with anionic ligands in the buffer, which affects resolution. The pH of the buffer should also be optimized since selectivity and resolution may occur at different pH range, and thus the two parameters need to be investigated individually.

The choice of buffer and its concentration also determines the conductivity of the solution within the capillary.[50] Conductivity of the sample analytes and electrolytes is a main concern because it will affect the electrophoretic resolution as well as the amount of heat generated in the capillary. When the conductivity of the sample approaches or exceeds that of the electrolytes in the buffer solution, CE resolution is degraded. Increasing the conductivity of the electrolyte solution leads to increased Joule heating which produces peak broadening. On the other hand, if the conductivity of electrolyte is too high, heating may be large enough to increase the temperature of the electrolyte, and produces radial temperature gradients in the capillary. Consequently, convection currents will lead to mixing and the degradation of resolution. If the conductivity is too low, the voltage gradient in the injected sample plug and nearby analyst zone in the capillary will always be smaller than the voltage gradient in the electrolyte solution. As a result, electrophoresis peaks will also be distorted. It is generally recommended that the concentration of the electrolysts should be at least 1000 times the concentration of sample.

Applied voltage is another factor affecting CE separation. As the applied voltage is increased, the separation time is reduced. However, such increase causes an increase in electric current, resulting in Joule heating in the capillary. Consequently, the EOF and electrophoretic mobility of the analytes will also be affected. Thus, while high electric field often improve resolution and efficiency, the effects of Joule heating essentially sets the limit on such increase in real practice. For efficient separation, a stable and reliable electrical current along the capillary is necessary. In addition, the capillary should also be cooled actively during analysis to prevent local heating.

4 DETECTION SYSTEMS FOR CAPILLARY ELECTROPHORESIS

As the total sample intake by CE analysis is extremely low (5–100 nanoliter range), the concentration-based sensitivity of the detection system must be sufficiently high in order to be applicable in speciation analysis. The development of highly sensitive and selective detectors for CE has always been a subject of intense interest. Until today, ultraviolet and fluorescence detectors are the most commonly used ones for CE in speciation analysis. As was shown in Table 1 the detection limits of these detectors are generally in the order of 10^{-5} to 10^{-6} M range. Such sensitivity is not sufficient for the speciation analysis of trace quantities of analytes often faced in environmental samples. Other detection techniques such as MS, laser-induced-fluorescence, amperometry and conductivity have all been successfully applied as CE detectors with high sensitivity.[51–55] However, these detectors lack the needed 'elemental specificity' to selectively detect the specific elemental species of interest in the presence of complex matrix interference.

As was stated earlier, the use of atomic spectrometry as an elemental specific detector in hyphenated techniques has been a popular approach in speciation analysis. Such detection may be single element in nature, e.g. atomic absorption spectroscopy (AAS), or multi-element, e.g. inductively coupled plasma optical emission spectroscopy (ICPOES). The interface of chromatography with these detection devices is relatively straightforward due to the continuous flow characteristics and the compatible flow rates (mL min^{-1} range) of the two partner devices. The early interface problems in these systems have now been mostly alleviated. Most of these interface systems such as high-performance liquid chromatography/inductively coupled plasma atomic emission spectroscopy (HPLC/ICPAES) or gas chromatography/microwave induced plasma (GC/MIP) currently in use are relatively cheap and easy to construct, and can be readily constructed and demounted when not in use. The instrumental design and their applications in speciation analysis have been the subject of many reviews as was given in several articles included in the earlier references.[1–3]

Before the wide availability of ICPMS, inductively coupled plasma atomic emission spectroscopy (ICPAES) was the most popular detection systems used in hyphenated CE techniques. Olesik et al.[56] was the first to report the application of CE/ICPOES for elemental speciation of chromium and iron. The system was shown to be a viable technique for the rapid analysis of free ions of different charge states and organometallic species with sensitivities in the low ppb range. A similar system developed by Lu et al.[57] was applied successfully for the speciation of Cu^{2+}, $Cu(EDTA)_2^{2-}$, and Cr^{3+} and $Cr_2O_7^{2-}$ with sensitivities also in the ppb range. These and other earlier studies provided valuable information on the design of the interface and nebulizer to couple CE with the plasma systems. On the other hand, they also demonstrated clearly that further improvement in detection sensitivity is needed before hyphenated CE techniques can be utilized effectively for routine speciation analysis of real world samples.

ICPMS represents the newest generation of multi-elemental detection instrument. The state-of-the-art of the instrument can be found in many reviews including the more recent ones by Caruso et al.[58,59] Compared to ICPOES, ICPMS is 10–1000 times more sensitive, and offers simpler spectral interpretation with less interference. The analytical capability of MS to monitor ionic species from ICP has been well known since the mid-1980s by the work of Houk et al.[60] Different versions of the instruments are now available commercially. In these instruments, the plasma torch is in a horizontal position under normal pressure. An interface cone is placed between the plasma source and the mass spectrometer. The ions produced in the ICP are transported into the MS through a small hole in the cone through a differentially pumped interface linked to a quadruple spectrometer. Quadruple MS are the most commonly used MS instrument as it is more compact, less expensive and more rugged than magnetic sector MS. The quadruple MS covers a mass range of 3 to 300, with unit mass resolution, and has a dynamic range of 6 orders of magnitude. Over 90% of the elements in the periodic table can be determined by ICPMS with detection limits in the sub-ppb range for most elements. The spectra so produced, which consist of a simple series of isotope peaks, are much simpler compared with conventional ICP emission spectra. These spectra can be used for quantitative measurement based on calibration curves, often with an internal standard. The fast scan rate and peak jumping routines of the quadruple MS result in an essentially simultaneous multi-element detection when it is interfaced on-line to different sample introduction systems.

ICPMS is an unique technique to discriminate between the masses of the different isotopes of an element where more than one stable isotope occurs. Consequently an enriched isotope that normally is present in relatively low abundance can be used as a tracer in metabolic studies to compare to a reference whose abundance is constant. Another major advantage is the application of isotope dilution analysis. Isotope ratio for two selected isotopes of an element of interest is measured in a solution after the addition of a known quantity of a spike that contains enrichment of one of the isotopes,

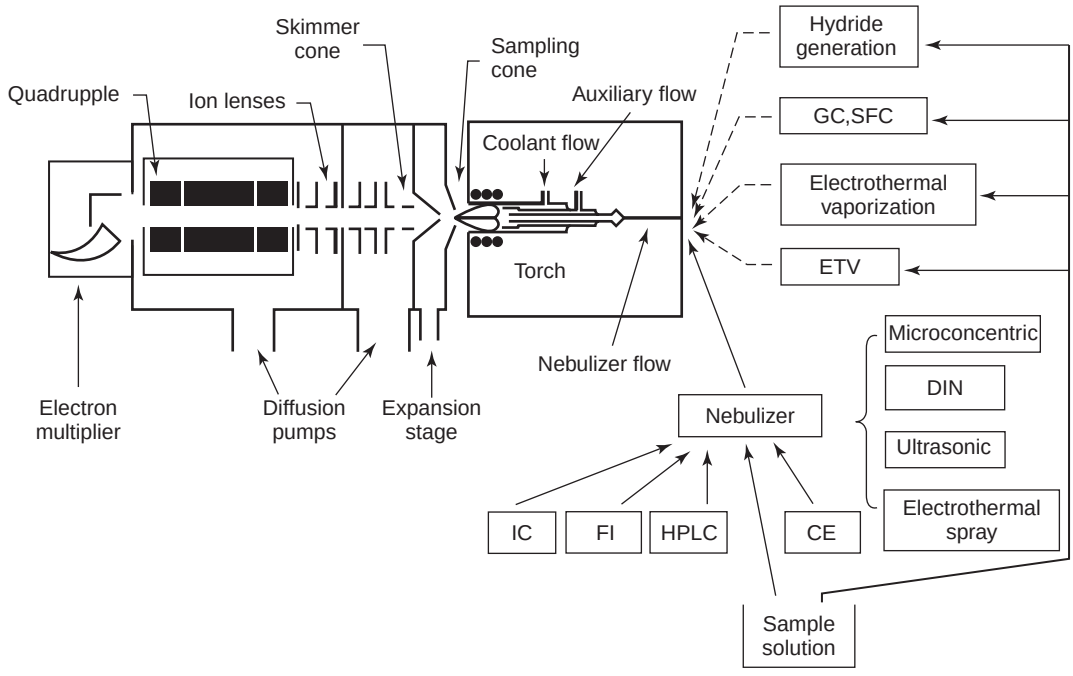

Figure 1 Schematics showing the components of an ICPMS instrument and its interface to different sample introduction systems.

permitting calculation of the original concentration of the element. The use of isotope dilution saves the time required to obtain calibration graphs and compensates for matrix effects as was illustrated in the recent articles by Heumann et al.[61,62] Using ICPMS coupled with SEC (size exclusion chromatography), these authors were able to determine the extent of complexation of Cu, Mo and Fe with different fractions of dissolved organic materials or humic substances in aqueous samples. The speciation of iodine compounds of iodate, iodide and organoiodine in fresh and seawaters were also successfully accomplished by the coupling of ICPMS with reversed-phase HPLC. The sensitivities of ICPMS detection for these species were down to the pg mL^{-1} range.

Since the late 1980s, a variety of ICPMS based hyphenated techniques have been reported in the literature. Thompson and Houk[63] first reported the use of plasma MS as detector for HPLC for the speciation of As and Se. The study demonstrated the superior capability of ICPMS for multi-element detection and isotope ratio analysis with high precision and accuracy. A variety of sample introduction devices have since been interfaced to ICPMS for speciation applications. Figure 1 gives a schematic representation of the basic components of ICPMS; and the wide variety of sample introduction or chromatographic separation techniques with which ICPMS has been successfully interfaced.

The coupling of CE and ICPMS provides an ideal tool for speciation analysis because it combines the high separation efficiency of CE with high elemental sensitivity and specificity of ICPMS. The analytical performance and instrumentation of the variety of ICPMS-based hyphenated technique can be found in some recent reviews.[64]

Table 2 Attributes of CE and ICPMS

Advantages of CE	Advantages of ICPMS
High efficiency and resolution	High sensitivity
Less sample requirement	Multi-element determination
Ionic and large molecules	Isotope dilution analysis
Less reagents consumption	Wide dynamic range
High separation speed	Less mass spectral interference

Table 2 outlines the major attributes of CE and ICPMS which make the coupling of the two an attractive approach in speciation studies.

5 CAPILLARY ELECTROPHORESIS/ INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY SYSTEMS

When used as a stand-alone instrument, ICPMS can generally use the same or closely similar sample introduction techniques that had been developed for the more well-established ICPOES systems. Many high efficiency nebulizers have been developed to efficiently transport the liquid samples or chromatographic eluent to the

Table 3 Speciation analysis by CE/ICPMS

Nebulizer	Spray chamber	Analyst	Detection limits			Reference (Author)
Modified Meinhard high-efficiency nebulizer	Home-made spray chamber	Fe^{2+}, Fe^{3+}, $Fe(CH_3COO)_2$, $Cr(OH)_3$, Cr(VI), As(III), As(V), Sn^{4+}, Sn^{2+}	Sr Cr As Sn Co Li	0.06 μg mL^{-1} 2 μg mL^{-1} 1 μg mL^{-1} 0.5 μg mL^{-1} 0.09 μg mL^{-1} 0.2 μg mL^{-1}		56, 65 (Olesik)
Glass frit DIN	None	$PbCl_2$, $C_4H_9Pb(O_2CH_3)$, Li^+, K^+, Mn^{2+}, Co_2^+, Sr^{2+}, Cd^{2+}, T^{1+}, Pb^{2+}, As^{3+}, As^{5+}, DMA, MMA, Se(IV), Se(VI)	Tl Pb As(III) As(V) Se(IV) Se(VI)	7 ng mL^{-1} 600 ng mL^{-1} 100 ng mL^{-1} 20 ng mL^{-1} 300 ng mL^{-1} 100 ng mL^{-1}		58, 59, 66 (Caruso)
Meinhard TR-30 A	Conical spray Chamber	Metallothionein: Fe, Cu, Zn, Cd, Pb	Fe Cd	2.5 μg g^{-1} 0.055 μg g^{-1}		57 (Barnes)
Ultrasonic nebulizer		Metallothioneins (Cd and Zn)	Cd Zn	0.082 μg g^{-1} 1.18 μg g^{-1}		67 (Barnes)
Modified Meinhard nebulizer	Home-made quartz tubing	Pt-Methioine, Se (IV), Se (VI)				68 (Michalke)
Pneumatic concentric nebulizer (CPCN), microconcentric nebulizer (MCN)	Home-made conical spray chamber	Cr(III)m Cr(VI), Cu^{2+}, $Cu(EDTA)^{2-}$	Mo Cr Zn Ni	7.4 μg mL^{-1}, Co 7.4 μg mL^{-1}, Mn 1.7 μg mL^{-1}, Cu 2.7 μg mL^{-1}, Fe	3.9 μg mL^{-1} 0.9 μg mL^{-1} 3.3 μg mL^{-1} 2.0 μg mL^{-1}	69 (Mei)
Modified concentric nebulizer	Low volume cyclone spray chamber	Cd in rabbit metallothionein	Cd	2.36 μg mL^{-1}		70 (Majidi)
Standard cross-flow nebulizer	Commercial Scott spray chamber	Metallothioneins (Cd, Zn and Cu)	Cd	0.21 μg mL^{-1}		70 (Majidi)
HG	None	As(III), As(V), DMA, MMA	As(III) 25 ng mL^{-1}, As(V) 6 ng mL^{-1} DMA 58 ng mL^{-1}, MMA 9 ng mL^{-1}			76 (Magnuson)
HG by movable reduction bed	None	As(III), As(V), DMA, MMA	As (III) 3.0 ng mL^{-1}, As (V) 6.2 ng mL^{-1}			49 (Wang)

DIN, direct injection nebulizer; HG, hydride generation.

plasma in aerosol forms so that they can be effectively vaporized, atomized, ionized or excited. However, different sets of problems occur when these interfaces are applied to CE. In the following text, general considerations on the design of the interface will be introduced. Some of the successful interfaces reported in the literature along with their applications will be described. A summary of the different systems studied in these reports is given in Table 3. The types of metal compounds which have been speciated and the corresponding detection limits for each of these species are included in Table 3. Details of these systems will be discussed further in later text.

The technique of CE/ICPMS is still in the early stage of development and thus only limited numbers of papers with experimental details have been published so far. The information summarized in Table 3 is meant to give an overview of the status of development of the technique rather than an exhaustive literature review.

5.1 General Consideration in Interface Design

There are three main challenges in the design of an interface between CE and ICPMS. First, in CE/ICPMS, different from those of conventional CE set-up, the exit end of CE capillary is no longer placed inside a grounded buffer reservoir. Thus, one has to find an alternative way to complete the electric connection in order to apply the high voltage across the entire CE capillary. Secondly, to achieve high concentration-based sensitivity, the transport efficiency of the interface needs to be as high as possible because of the small volumes of samples introduced in CE (5–50 nL). Lastly, along with high transport efficiency, the interface should also be able to minimize electrophoretic band broadening caused by problems of dead volumes or flow incompatibilities. To solve these problems requires a substantial modification of the conventional nebulizer/spray-chamber designs

commonly used for sample introduction in atomic spectrometry. Using these conventional nebulizers, which are built to handle flow rate in the mL min^{-1} range, would not only lower the sensitivity due to sample lose or dilution, but also degrade the electrophoretic peak shape because of dispersion in the laminar flow of the conduits.

Two general approaches have been taken to achieve proper electric connection of the CE capillary in CE/ICPMS. In the first approach, the electric connection to the capillary was made by conducting silver paint on the outside of the capillary. The grounded cathode end is then placed directly into the nebulizer for sample transport to the plasma.[56] The second approach is to mix an electrically conductive sheath buffer flow to combine with the CE flow at the tip of the CE capillary before entering the nebulizer.[57,70] The sheath flow serves two purposes. Besides serving as an electrical connection to the cathodic end of the separation capillary, it also serves as a make-up flow to match the liquid flow requirement of the nebulizer. All the connections and tubings used in these interfaces are micro-volume connectors in order to minimize the dead volume. The operating conditions such as buffer composition, flow compatibility and nebulizer gas flow rate also have to be carefully optimized to obtain satisfactory performance. Interfaces based on the use of sheath flow design have been found to give minimum band broadening with aerosol transport efficiency in the range of 20–80%, depending on the total liquid flow rate delivered to the nebulizer.

Many different new designs of nebulizers for CE interface have been reported in the literature, and these include DIN without the use of a spray chamber, and low-flow versions of concentric nebulizers based on Meinhard concentric nebulizer design coupled with low volume spray chambers. The transport efficiency in general is close to 100% for the direct injection techniques, and better than 50% for the low-flow nebulizers.

Although both CE and ICPMS are now well established techniques by themselves, the coupling of CE with ICPMS is relatively new and interface designs are still being developed and improved. Up to the present, there has not yet been a fully commercialized CE/ICPMS in the market which is of universal utility to all types of analytes of interest. Nevertheless, the use of direct injection nebulization and low-flow nebulizers appears to hold the best potential for routine applications.

5.2 Glass Concentric Nebulizer

Olesik first described the application of CE/ICPOES/MS for elemental speciation.[56,65] A commercial Mainhard concentric nebulizer was modified to couple CE capillary to ICP. The electrophoresis capillary was fitted into the inner tube of the nebulizer. The outside 4–5 cm of the capillary was coated with silver paint, which serves as an electric connection between the solution at the tip of the capillary and the ground. Under some conditions, contamination of the capillary effluent by Ag was observed, especially when ligands capable of complexing strongly with Ag were present in the buffer solution. A close to zero liquid dead volume and up to 100% sample transfer efficiency was obtained with the interface.

Using this interface, a series of Cr and Fe compounds were separated and detected using CE/ICPOES. Satisfactory electrophorograms of a series of separations of tin (Sn^{4+} and Sn^{2+}), strontium (Sr^{2+}), arsenic (As (III) and As (V)) and chromium (Cr (III) and Cr (VI)) species presented each at 1 ppm concentration was reported. The separation of free ions from metal ligand complexes such as the mixture of Cu^{2+}, $Cu(EDTA)_2{}^{2+}$ was also accomplished. The detection limit for Sr using CE/ICPOES with an injection volume of 0.14 μL was 0.3 ppb, compared with 0.03 ppb obtained with direct ICPOES using conventional sample introduction techniques. The interface was also used to couple CE with ICPMS and the detection limits found in these coupled systems were in the range of 0.06–2 ppb for the elements investigated. Detection limits for Sr, Cr, As and Sn by CE/ICPMS were a factor of 6–200 times worse than that of direct ICPMS analysis using standard nebulizers and spray chambers. These initial studies suggested the promising potential of CE/ICPMS technique, but also the need for sensitivity improvement in order to be used for routine speciation analysis of real world samples.

Another version of Meinhard type nebulizer design was developed by Lu et al.[57] The unique feature of this design is the variable positioning of the capillary in the concentric nebulizer. The CE capillary was inserted into the central glass tube of the nebulizer, about 2 cm away from the tip. Another silica capillary was also inserted coaxially into the central tube around the CE capillary. The maximum distance between the end of the CE capillary and the outer tube was 18 mm. A coaxial liquid sheath using 10 mmol L^{-1} NaCl solution was supplied through the outer capillary used to complete the electric circuit of CE. The signal was found to decrease as the make-up sheath flow increases because of sample dilution by the make-up flow.

The system was used for the analysis of metal binding proteins, metallothionein and ferritin. The effect of CE capillary positioning on separation resolution was studied. Because of suction induced by the nebulizer, the farther the capillary was placed inside the nebulizer, the poorer was the resolution. The problem can be overcome by supplying negative pressure to the sample vial, but a compromise has to be made between resolution and signal intensity, which is affected by pressure applications.[57] The detection limits reported were 2.5 μg of $^{57}Fe\ g^{-1}$ and

0.055 μg of $^{114}Cd\,g^{-1}$. The absolute detection limits for 73.6 nL sample injection were 184 fg of ^{57}Fe and 4 fg of ^{114}Cd.

Mei et al.[69] reported a new version of the modified concentric nebulizer interface. The interface used a simple Y quartz tube. The first arm was connected with the outer end of the CE capillary. The second arm was connected with the auxiliary capillary used to draw up the make-up flow. The other end of the auxiliary tubing was dipped into the 0.05 M HNO_3 solution. The third arm was connected directly to the sampling tube of the nebulizer. The solutions were individually sucked up into their respective capillaries by self-aspiration and then mixed in the junction before being aspired into the nebulizer. While the make-up solution was contacting with the ground of the d.c. supply, the buffer electrode solution was connected to the negative polarity of the d.c. supply. The negative high voltage was used to form an opposite EOF against the self-aspirating flow. Because the interface was placed outside and independent of the nebulizer, it is more flexible and could be easily connected to various types of nebulizers. High efficiency separation of Cu^{2+}, $Cu(EDTA)_2^{2-}$, and Cr^{3+} and $Cr_2O_7^{2-}$ pairs was observed.

5.3 Direct Injection Nebulizer

DIN is a micro-concentric pneumatic nebulizer without the use of a spray chamber. Instead, the liquid sample is nebulized directly into the central channel of the plasma torch.[72] A commercial DIN unit is now available in the market. DIN has been used widely as an interface for micro-bore HPLC/ICPMS systems because of its high transport efficiency, low dead volume, and flow compatibility with the chromatographic systems.

Liu and coworkers[73] used a Microneb 2000 DIN nebulizer to interface CE with ICPMS. The CE capillary is inserted concentrically inside the fused-silica DIN sample introduction capillary. A liquid make-up flow in the DIN sample introduction capillary around the CE capillary was used to establish the electric contact. The design has an extremely small internal volume of $<2\,\mu L\,min^{-1}$ and can accommodate low rate in the range of $10–100\,\mu L\,min^{-1}$. Besides serving as the electric contact, the make-up flow design in DIN is also used to ensure that the DIN system could be operated independently of the CE system, and thus independently of the EOF generated by the eletrophoresis process. This is advantageous because it allows the use of different buffers and applied voltages to optimize separation

It was found that the DIN interface did not cause any noticeable suction and back pressure problems during CE analysis. Baseline separation of Li^+, K^+, Sr^{2+} and Tl^+ was achieved in less than 210 s. The detection limits ranged from $7\,pg\,mL^{-1}$ for Tl^+ and $1000\,pg\,mL^{-1}$ for K^+. Satisfactory results were also reported for the speciation analysis of Se and As.

5.4 Ultrasonic Nebulizer

Ultrasonic nebulizers (USN) are used widely in ICPOES to improve transport efficiency and detection capability for elemental analysis. Its improvement over conventional pneumatic nebulizer systems is the dense aerosol and smaller droplet produced, and the resulting 3–8 times improvement in detection limits.[74]

In a series of studies, Lu and Barnes[67] have made a systematic comparison of the efficiency and resolution between USN and conventional concentric nebulizers as an interface in CE/ICPMS. A commercial pneumatic-based USN was used in the study. The USN interface was found to provide considerable improvement in separation resolution and better signal sensitivity. It also showed low sensitivity to mechanical parameters. The systems were used for the separation of metallothioneins. The signal intensity of the USN system showed an improvement of 3.6–8.4 fold compared to those with the concentric nebulizer interface. The gain in sensitivity, however, was compromised by the high background noise. Caruso et al.[75] has recently improved the USN interface design which showed noticeable improvement in resolution and signal/noise ratio.

5.5 Oscillating Capillary Nebulizer

Caruso et al.[76] recently reported the use of a pneumatic-based Oscillating Capillary Nebulizer (OCN) as an interface for CE/ICPMS. One attractive feature of the interface is its compatibility with a wide range of flow rates from $1\,\mu L\,min^{-1}$ to $2\,mL\,min^{-1}$. The transport efficiency was close to 100% under low flow rate conditions. In OCN, the aerosol was generated by both the flow of the nebulizer gas and the induced oscillation of the transport tubes. The concerted action enhanced the breakup of liquid samples into small droplets. The design used by Caruso is a modified version of the original model. The original OCN was constructed from two concentrically mounted flexible fused-silica tubes used each for the liquid and nebulizer gas flows. In the modified version, a three tube system was used and the additional tube was installed to accommodate a makeup buffer flow for ground connection.

Detection limits in the low ppbs were reported for the analysis of both inorganic cations and anions using the system. However, noticeable broadening and tailing of the peaks were also observed. It was suggested that both the turbulence in the spray chamber and the action of nebulizer suction could be responsible for the problems observed.

5.6 Hydride Generation

HG provides a different approach to interface CE with ICPMS.[71,77] It circumvents some of the problems in nebulizer-based interfaces in which the CE buffer often causes the build up of deposits on the sampling interface, and thus influences the long term stability of the system. On the other hand, the gaseous hydride can be nearly quantitatively introduced into the plasma without the problem of high dissolved solids commonly associated with CE buffers. The common problem of chloride interference (m/z of 75) in HG systems can be eliminated by using a microporous membrane as the separator.

The HG system is particularly useful for the speciation of hydride-forming species such as As, as was demonstrated in the work by Mathew et al.[77] In their system, the CE unit is interfaced to a microporous membrane as the gas–liquid separator before ICPMS detection. As (III), As (V), DMA and MMA can be speciated in real water systems in less than 5 min.

Magnuson et al. described the use of hydrodynamic modified EOF CE (HMEOF) with HG and ICPMS detection for As speciation.[41,71] A microporous PTFE tube was used as a gas–liquid separator. In HMEOF, the EOF in CE was modified by hydrodynamic pressure opposite to the direction of the EOF. HMEOF permits the injection of increased quantities of analytes for sensitivity enhancement. Because of the low flow from the capillary, a makeup buffer flow was used. The capillary ground was also completed through this makeup flow. The CE/HG/ICPMS system was able to speciate arsenic species in real water samples with detection limits down to parts per trillions for As (III) and As (V).

An interesting alternative to the conventional hydrogenation techniques recently reported is the movable reduction bed hydride generation system (MRBHG) developed by Wang and co-workers.[49,78] They have used the MRBHG as an interface in both ICPOES and ICPMS systems. The unique feature of the MRBHG device is its capability to carry out HG reactions in a liquid/solid system at microliter sample volumes, and without the use of gas–liquid separators or the addition of acid solutions to the sample. Instead, the hydride forming reactions take place on the surface of a moving tape which has been precoated with a mixture of dry $NaBH_4$ and solid tartaric acid powders at appropriate ratios. During a CE run, the eluent from the capillary, combined with a makeup flow, is made to drip continuously onto the tape surface. Hydride forming reactions occur upon the contact of the aqueous eluent from CE with the reagents on the surface. The 'reduction-bed' on the tape surface is continuously refreshed by the constant movement of the tape during a run.

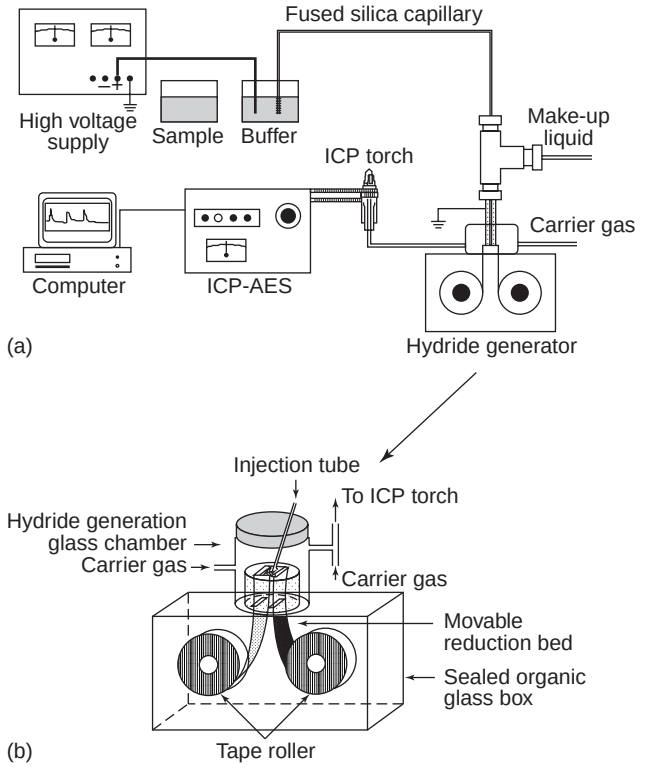

Figure 2 Scheme showing the CE/MRBHG/ICPMS system (a) and the MRBHG interface (b).

The construction of the system is shown in Figure 2. In the set-up, the reagent-coated rolling tape is situated inside a sealed reaction chamber. The tape is made to move at a specified rate by turning the two tape rollers that are placed in a sealed organic glass cassette. The tape then crosses the two holes on the stopper into and out of a reaction chamber. The hydrides produced in the chamber are swept into the ICP torch by argon carrier gas.

Compared with conventional HG techniques, MRBHG provides some unique advantages. The major ones include: (1) its ability to handle very small sample sizes down to the μL volumes, (2) it suffers less chemical interference from other co-existing hydride forming elements in the sample matrix, (3) its efficiency in HG is insensitive to solution pH over a wide pH range, (4) its general applicability to all hydride forming elements of different valence state and chemical forms, and finally (5) its low cost of operation since no complicated apparatus is needed. The advantage of the system as a CE/ICP interface is its low dead volume because of the elimination of the use of complex plumbing and the gas–liquid separator.

The above authors have applied the MRBHG interface in both CE/ICPOES and CE/ICPMS systems for the speciation of four arsenic species including As (III), disodium methylarsonate, dimethylarsinic acid and As (V).

The detection limits reported for the optical emission spectroscopy (OES) system were in the range of $0.3\,\mu g\,mL^{-1}$.[78] Much lower detection limits were obtained in the ICPMS system, and the sensitivities observed for the same arsenic species were better than $10\,ng\,mL^{-1}$.[49]

ACKNOWLEDGMENTS

The authors would like to acknowledge the contribution of Dr Tian Xiaodan of Xiamen University. Part of the information reported here was extracted from her thesis work which she did under the supervision of Professor Wang Xiaoru.

Grant support from the key project of National Natural Science Foundation of China (X.R. Wang) and HK UGC (F.S.-C. Lee) is acknowledged.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
ANA	Arsenilic Acid
CE	Capillary Electrophoresis
CE/ICPMS	Capillary Electrophoresis/ Inductively Coupled Plasma Mass Spectrometry
CZE	Capillary Zone Electrophoresis
DIN	Direct Injection Nebulizer
DMA	Dimethylarsenic Acid
EOF	Electroosmotic Flow
GC	Gas Chromatography
GC/MIP	Gas Chromatography/Microwave Induced Plasma
HG	Hydride Generation
HPLC	High-performance Liquid Chromatography
HPLC/ICPAES	High-performance Liquid Chromatography/Inductively Coupled Plasma Atomic Emission Spectroscopy
IC	Ion Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICPOES	Inductively Coupled Plasma Optical Emission Spectroscopy
MEKC	Micellar Electrokinetic Chromatography
MMA	Monomethylarsenic Acid
MRBHG	Movable Reduction Bed Hydride Generation System
MS	Mass Spectrometry
OCN	Oscillating Capillary Nebulizer
OES	Optical Emission Spectroscopy
RPLC	Reversed-phase Liquid Chromatography
SEC	Size Exclusion Chromatography
USN	Ultrasonic Nebulizers
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis

REFERENCES

1. I.S. Krull (ed.), 'Trace Metal Analysis and Speciation', *J. Chromatography Library*, Elsevier Science Publishing Company Inc., New York, NY, Vol. 47, 1991.
2. P.C. Uden (ed.), 'Elemental-specific Chromatographic Detection by Atomic Emission Spectroscopy', *ACS Symposium Series 479*, American Chemical Society, Washington, DC, 1992.
3. R. Lobinski, 'Elemental Speciation and Coupled Techniques', *Appl. Spectrosc.*, **51**, 260–278A (1997).
4. H. Poppe, *Theory of Capillary Zone Electrophoresis*, Series: Advances in Chromatography, 233–300, Vol. 38, 1998.
5. P. Camilleri (ed.), *Capillary Electrophoresis Theory and Practice*, 2nd edition, CRC Press, Boca Raton, FL, 1998.
6. A.G. Ewing, R.A. Wallingford, T.M. Olefirowicz, 'Capillary Electrophoresis', *Anal. Chem.*, **61**(4), 292A (1989).
7. E. Dabekzlotorzynska, E.P.C. Lai, A.R. Timerbaev, 'Capillary Electrophoresis: The State-of-the-art in Metal Speciation Studies', *Anal. Chim. Acta*, **359**, 1–26 (1998).
8. A.R. Timerbaev, 'Strategies for Selectivity Control in Capillary Electrophoresis of Metal Species', *J. Chromatogr. A*, **792**(1–2), 495–518 (1997).
9. P. Janos, 'Role of Chemical Equilibrium in the Capillary Electrophoresis of Inorganic Substances', *J. Chromatogr. A*, **834**(1–2), 3–20 (1999).

10. T. Okada, 'Polyethers in Inorganic Capillary Electrophoresis', *J. Chromatogr. A*, **834**(1–2), 173–187 (1999).
11. A.R. Timerbaev, W. Buchberger, 'Prospects for Detection and Sensitivity Enhancement of Inorganic Ions in Capillary Electrophoresis', *J. Chromatogr. A*, **834**(1–2), 117–132 (1999).
12. M. Chiari, 'Enhancement of Selectivity in Capillary Electrophoretic Separations of Metals and Ligands through Complex Formation', *J. Chromatogr. A*, **805**(1–2), 1–15 (1998).
13. V. Kaniansky, M. Masar, J. Marak, R. Boder, 'Capillary Electrophoresis of Inorganic Anions', *J. Chromatogr. A*, **834**(1–2), 133–178 (1999).
14. A.H. Harakuwe, P.R. Haddad, 'Control of Separation Selectivity in Capillary Zone Electrophoresis of Inorganic Anions', *J. Chromatogr. A*, **834**(1–2), 213–232 (1999).
15. V. Pacakova, P. Coufal, K. Stulik, 'Capillary Electrophoresis of Inorganic Cations', *J. Chromatogr. A*, **834**(1–2), 257–275 (1999).
16. K. Fukushi, S. Takeda, K. Chayama, S. Wakida, 'Application of Capillary Electrophoresis to the Analysis of Inorganic Ions in Environmental Samples', *J. Chromatogr. A*, **834**(1–2), 349–362 (1999).
17. S.M. Valsccehi, S. Polesello, 'Analysis of Inorganic Ions in Environmental Samples by Capillary Electrophoresis', *J. Chromatogr. A*, **834**(1–2), 363–385 (1999).
18. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, *Anal. Chem.*, **56** (1984) and **57** (1985).
19. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open-Tubular Capillaries', *Anal. Chem.*, **56**, 111 (1984).
20. K. Otsuka, S. Terabe, T. Ando, 'Electrokinetic Chromatography with Micellar Solutions Separation of Phenylthiohydantoin–Amino Acids', *J. Chromatogr.*, **332**, 219 (1985).
21. L.N. Amankwa, W.G. Kuhr, *Anal. Chem.*, **63**, 1733 (1991).
22. T. Lee, E.S. Yeung, M. Sharma, *J. Chromatogr.*, **565**, 197 (1991).
23. M.G. Khaledi, 'Micelles as Separation Media in High Performance Liquid Chromatography and High Performance Capillary Electrophoresis: Overview and Perspective', *J. Chromatogr. A*, **780**(1–2), 3–40 (1997).
24. I. Rodriquez, S.F.Y. Li, 'Surface Deactivation in Protein and Peptide Analysis by Capillary Electrophoresis', *Anal. Chim. Acta*, **383**(1–2), 1–26 (1999).
25. I. Medina, E. Rubí, M.C. Mejuto, R. Cela, 'Speciation of Organomercurials in Marine Samples Using Capillary Electrophoresis', *Talanta*, **40**(11), 1631 (1993).
26. A.M.C. Diaz, R.A. Lorenzo-Ferreira, R. Cela-Torrijos, 'Validation and Quality Control of Methylmercury Determinations by Means of Capillary Electrophoresis', *Mikrochim. Acta*, **123**, 73 (1996).
27. C. Vogt, G. Werner, 'Speciation of Heavy Metals by Capillary Electrophoresis', *J. Chromatogr.*, **686**, 325 (1994).
28. M. O'Keeffe, L. Dunemann, A. Theobald, G. Svehla, 'Capillary Electrophoresis in Speciation Analysis. Investigations of Metal–Polyaminopolycarboxylate Complexes and Effects of Metals on Proteins', *Anal. Chim. Acta*, **306**, 91 (1995).
29. M. Albert, C. Demesmay, J.L. Rocca, 'Analysis of Organic and Non-organic Arsenious or Selenious Compounds by Capillary Electrophoresis', *Fresenius J. Anal. Chem.*, **351**, 426 (1995).
30. L. Lin, J. Wang, J. Caruso, 'Arsenic Speciation Using Capillary Zone Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr. Sci.*, **33**, 177 (1995).
31. D. Schlegel, J. Mattusch, R. Wennrich, 'Speciation Analysis of Arsenic and Selenium Compounds by Capillary Electrophoresis', *26th International Symposium on Environ. Anal. Chem.*, Wien, Austria, 1996.
32. J.F. Lopez-Sanchez, M.B. Amram, M.D. Lakkis, F. Lagarde, G. Rauret, M.J.F. Leroy, 'Quantitative Aspects of the Separation of Arsenical Species by Free Solution Capillary Electrophoresis', *Fresenius J. Anal. Chem.*, **348**, 810 (1994).
33. S. Schaffer, D.P. Gareil, C.D. Richard, *J. Chromatogr.*, **740**, 151 (1996).
34. S. Motellier, K. Gurdale, H. Pitsch, 'Sulfur Speciation by Capillary Electrophoresis with Indirect Spectrophotometric Detection', *J. Chromatogr.*, **770**, 311 (1997).
35. M.P. Richards, J.H. Beattie, R. Self, 'Application of Capillary Zone Electrophoresis to the Separation of Metallothionein Isoforms', *J. Liq. Chromatogr.*, **16**, 2113 (1993).
36. H. Kajiwara, 'Application of High-performance Capillary Electrophoresis to the Analysis of Conformation and Interaction of Metal-binding Proteins', *J. Chromatogr.*, **559**, 345 (1991).
37. K. Kiyohara, K. Saitoh, N. Suzuki, 'Micellar Electrokinetic Capillary Chromatography of Haematoporphyrin, Protoporphyrin and Their Copper and Zinc Complexes', *J. Chromatogr.*, **646**, 397 (1993).
38. C.L. Ng, H.K. Lee, S.F.Y. Li, 'Determination of Organolead and Organoselenium Compounds of Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **652**, 547 (1993).
39. J. Harms, G. Schwedt, 'Applications of Capillary Electrophoresis in Element Speciation Analysis of Plant and Food Extracts', *Fresenius J. Anal. Chem.*, **350**, 93 (1994).
40. V. Majidi, N.J. Miller-Ihli, 'Potential Sources of Error in Capillary Electrophoresis-Inductively Coupled Plasma Mass Spectrometry for Chemical Speciation', *Analyst*, **123**, 809–814 (1998).
41. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, *Analyst*, **122**, 1057 (1997).
42. C.E. Evans, *Anal. Chem.*, **69**, 2952–2954 (1997).
43. P. Kuban, A. Engstrøm, J.C. Olsson, G. Thorsøn, R. Tryzell, B. Karlberg, 'New Interface for Coupling Flow-injection and Capillary Electrophoresis', *Anal. Chim. Acta*, **337**, 117–124 (1997).

44. P. Kuban, B. Karlberg, *Anal Chem.*, **69**, 1169–1173 (1997).
45. R.M. McCormick, *Anal. Chem.*, **60**, 2322 (1990).
46. J.W. Lambert, D. Middleton, *Anal. Chem.*, **60**, 2322 (1990).
47. X. Tian, X. Zhuang, X.R. Wang, 'Application of Capillary Zone Electrophoresis in Arsenic Speciation Analysis', *Chemical Journal of Chinese Universities* (in Chinese), **1**, 27–30 (1998).
48. X. Tian, 'The Hyphenation of Capillary Electrophoresis and Inductively Coupled Plasma-Atomic Emission Spectrometry/Mass Spectrometry for Arsenic Speciation Analysis', Ph.D. thesis, Xiamen University, Xiamen University Publishings, Xiamen, China, 1998.
49. X.D. Tian, Z.X. Zhuang, B. Chen, X.R. Wang, 'Determination of Arsenic Speciation by Capillary Electrophoresis and ICP-MS Using a Movable Reduction Bed Hydride Generation System', *Atomic Spectroscopy*, **20**(4), 127–133 (1999).
50. J.W. Olesik, J.A. Kinzer, E.J. Grunwald, K.K. Thaxton, S.V. Olesik, 'The Potential and Challenges of Elemental Speciation by Capillary Electrophoresis-inductively Coupled Plasma Mass Spectrometry and Electrospray or Ion Spray Mass Spectrometry', *Spectrochim. Acta*, Part B, **53**, 239 (1998).
51. R.D. Smith, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, 'Capillary Electrophoresis/Mass Spectrometry', *Anal. Chem.*, **65**, 574A (1993).
52. J.A. Olivares, N.T. Nguyen, C.R. Yonker, R.D. Smith, 'On-line Mass Spectrometric Detection for Capillary Zone Electrophoresis', *Anal. Chem.*, **59**, 1230 (1987).
53. E. Gassman, J.E. Kuo, R.N. Zare, 'Electrokinetic Separation of Chiral Compounds', *Science*, **230**, 813 (1985).
54. P.K. Dasgupta, L. Bao, 'Suppressed Conductometric Capillary Electrophoresis Separation Systems', *Anal. Chem.*, **65**, 1003 (1993).
55. M. Albin, P.D. Grossman, S.E. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 489A (1993).
56. J.W. Olesik, J.A. Kinzer, S.V. Olesik, 'Capillary Electrophoresis Inductively Coupled Plasma Spectrometry for Rapid Elemental Speciation', *Anal. Chem.*, **67**, 1 (1995).
57. Q. Lu, S.M. Bird, R.M. Barnes, 'Interface for Capillary Electrophoresis and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**, 2949–2956 (1995).
58. K. Sutton, R.M.C. Sutton, J.A. Caruso, 'Inductively Coupled Plasma Mass Spectrometric Detection for Chromatography and Capillary Electrophoresis – Review', *J. Chromatogr. A*, **789**, 85–126 (1997).
59. G.K. Zoorob, J.M. Kiernan, J.A. Caruso, 'ICP-MS for Elemental Speciation Studies', *Mikrochim. Acta*, **128**(3–4), 145–168 (1998).
60. R.S. Houk, V.A. Fassel, G.D. Flesch, H.J. Svec, A.L. Gray, C.E. Taylor, 'Inductively Coupled Argon Plasma as an Ion Source for Mass Spectrometric Determination of Trace Elements', *Anal. Chem.*, **52**, 2283 (1980).
61. K.G. Heumann, in *Inorganic Mass Spectrometry*, eds. R. Adams, R. Gijbels, van Grieken, Wiley, New York, 301–376, 1988.
62. K.G. Hermann, L. Rottmann, J. Vogl, *J. Anal. Spectrom.*, **9**, 1351–1355 (1994).
63. J.J. Thompson, R.S. Houk, 'Inductively Coupled Plasma Mass Spectrometric Detection for Multielement Flow Injection Analysis and Elemental Speciation by Reversed-phase Liquid Chromatography', *Anal. Chem.*, **58**, 2541–2548 (1986).
64. For recent review articles on the subject, see papers and references quoted in Proceedings of the 1998 Winter Conference on Plasma Spectrochemistry, Scottsdale, Arizona, ICP Information Newsletter, Inc., University of Massachusetts, Box 34510, Amherst, MA 01003-4510, USA.
65. J.W. Olesik, L.C. Bates, 'Characterization of Aerosols Produced by Pneumatic Nebulizers for Inductively Coupled Plasma Sample Introduction; Effect of Liquid and Gas Flow Rates on Volume Based Drop Size Distributions', *Spectrochim. Acta*, **50B**, 285 (1995).
66. M.J. Tomlinson, L.L. Lin, J.A. Caruso, 'Plasma Mass Spectrometry as a Detector for Chemical Speciation Studies', *Analyst*, **120**, 583 (1995).
67. Q. Lu, R.M. Barnes, 'Evaluation of an Ultrasonic Nebulizer Interface for Capillary Electrophoresis and Inductively Coupled Plasma Mass Spectrometry', *J. Microchem.*, **54**, 129–143 (1996).
68. B. Michalke, 'Capillary Electrophoresis–A Useful Tool in Speciation Investigations', *Fresenius J. Anal. Chem.*, **354**, 557 (1996).
69. E. Mei, H. Ichihashi, W. Gu, S. Yamasaki, 'Interface for Coupling Capillary Electrophoresis to Inductively Coupled Plasma and On-column Concentration Technique', *Anal. Chem.*, **69**, 2187 (1997).
70. V. Majidi, Miller-Ihli, 'Two Simple Interface Designs for Capillary Electrophoresis-Inductively Coupled Plasma Mass Spectrometry', *Analyst*, **123**, 803–808 (1998).
71. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, *J. Anal. Atomic Spectrom.*, **12**, 689 (1997).
72. S.C.K. Shum, R.S. Houk, 'Elemental Speciation by Anion Exchange and Size Exclusion Chromatography with Detection by Inductively Coupled Plasma Mass Spectrometry with Direct Injection Nebulization', *Anal. Chem.*, **65**, 2972 (1993).
73. Y. Liu, V. Lopez-Avila, J.J. Zhu, D.R. Wiederin, W.F. Beckert, *Anal. Chem.*, **67**, 2020 (1995).
74. A. Montaser, H. Tan, L. Ishii, S.-H. Nam, M. Cai, 'Argon Inductively Coupled Plasma Mass Spectrometry with Thermospray, Ultrasonic and Pneumatic Nebulization', *Anal. Chem.*, **63**, 2660 (1991).
75. P.W. Kirlew, M.T.M. Castillano, J.A. Caruso, 'An Evaluation of Ultrasonic Nebulizers as Interfaces for Capillary Electrophoresis of Inorganic Anions and Cations with Inductively Coupled Plasma Mass Spectrometric Detection', *Spectrochim. Acta*, **53B**, 221 (1998).

76. P.W. Kirlew, J.A. Caruso, 'Investigation of a Modified Oscillating Capillary Nebulizer Design as an Interface for CE/ICPMS', *Appl. Spectrosc.*, **52**(5), 770 (1998).
77. L.M. Matthew, T.C. John, A.B. Carol, *J. Anal. Atom. Spectrom.*, **12**, 689 (1997).
78. X.-D. Tian, Z.-X. Zhuang, B. Chen, X.R. Wang, 'Movable Reduction Bed Hydride Generation System as an Interface for Capillary Zone Electrophoresis and Inductively Coupled Plasma Atomic Emission Spectrometry for Arsenic Speciation Analysis', *Analyst*, **123**, 899–903 (1998).

Cyanogen Chloride and Cyanogen Bromide Analysis in Drinking Water

Yuefeng Xie
The Pennsylvania State University at Harrisburg Middletown, USA

Cordelia J. Hwang
Metropolitan Water District of Southern California, La Verne, USA

1 Introduction **1**
2 Standards Preparation **2**
2.1 Preparation of CNCl and CNBr Standards with Neat Chemicals 2
2.2 In Situ Synthesis of CNCl Standards 3
2.3 Commercially Prepared Stock Standards 4
3 Sample Collection, Preservation, and Storage **4**
3.1 Sample Collection 4
3.2 Sample Preservation 4
3.3 Sample Storage 4
4 Analytical Methods **4**
4.1 Purge-and-trap Gas Chromatography/Mass Spectrometry 4
4.2 Headspace Gas Chromatography/Electron-capture Detection 6
4.3 Micro Liquid–Liquid Extraction Gas Chromatography/Electron-capture Detection 8
5 Summary **10**
Abbreviations and Acronyms **10**
Related Articles **10**
References **10**

Cyanogen chloride (CNCl) (CAS 506-77-4) is a common disinfection by-product (DBP) formed in chloraminated waters. Cyanogen bromide (CNBr) (CAS 506-68-3), the brominated analog of CNCl, is formed in chloraminated or ozonated waters that are high in bromide. Both CNCl and CNBr are labile in drinking water. They undergo degradation reactions at high pHs or in the presence of free chlorine or sulfite. Because of its potential health effects, CNCl was included in the United States Environmental Protection Agency (USEPA) Information Collection Rule (ICR). Aqueous standards for CNCl are generally prepared with pure CNCl gas. They can also be prepared with chlorinated or chloraminated cyanide (CN^-) solutions. Aqueous standards for CNBr are prepared with commercially available CNBr. For sample preservation, ascorbic acid is commonly used to quench the residual chlorine or chloramines. Acid is often added to lower the pH to prevent possible hydrolysis degradation. There are three methods for analyzing CNCl and CNBr in drinking water samples: purge-and-trap gas chromatography/mass spectrometry (PT GC/MS), headspace gas chromatography/electron-capture detection (GC/ECD), and micro liquid–liquid extraction gas chromatography/electron-capture detection (LLE GC/ECD). All three methods are capable of analyzing microgram per liter levels of CNCl and CNBr in drinking water, and all three give similar method performance, including accuracy, precision, spiking recovery, and method detection limit. The PT GC/MS method is a modification of the USEPA Method 524.2, which requires the use of specialized instrumentation. Because of CNBr's low vapor pressure and high water solubility, the analysis of CNBr also requires a heated purge. Headspace GC/ECD method is a relatively simple and rapid analytical method. As the samples are prepared individually, the method requires a full-time analyst to prepare and inject each sample. However, this analytical procedure can be automated using a commercially available headspace autosampler. In general, micro LLE GC/ECD is a well developed and common procedure for DBP analysis. The micro LLE GC/ECD method, originally developed for CNBr analysis, has been further developed for both CNCl and CNBr analysis in drinking water. The sample injection generally is automated with a common gas chromatography (GC) autosampler.

1 INTRODUCTION

Since chloroform was first reported in chlorinated drinking waters in the 1970s, many DBPs have been identified, including trihalomethanes (THMs), haloacetic acids (HAAs), and other miscellaneous DBPs. To address the concern over the presence of DBPs in finished drinking waters, the USEPA promulgated two DBP regulations, the ICR[1] and the Disinfectants and Disinfection Byproducts (D-DBP) Rule.[2] New maximum contaminant levels (MCLs) have been set by the USEPA for total THMs and five HAAs under the D-DBP rule.[2] Chloramines, which are inexpensive and persistent disinfectants, have been used by many water utilities to control the formation of DBPs. A 1989–90 survey indicated that 20% of the utilities reported using chloramines as part of their THM

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

minimization program.[3] The percentage could be significantly increased as a result of the D-DBP regulation.

CNCl is a product of the reaction between chloramines and natural organic matter (such as fulvic acid or amino acids). In a survey of 35 water utilities reported by Krasner et al.,[4] CNCl was detected typically in chloraminated water but not in waters containing a free chlorine residual. The concentration of CNCl in chloraminated waters ranged from 1 to 12 $\mu g\,L^{-1}$. CNBr, the brominated analog of CNCl, is formed in chloraminated or ozonated waters that are high in bromide.[5,6] In a water containing 0.11–0.39 $mg\,L^{-1}$ of bromide, 1.5–3.7 $\mu g\,L^{-1}$ of CNBr was reported, whereas concentrations of CNCl were between 9 and 18 $\mu g\,L^{-1}$.[7] In another study, increasing bromide levels increased the concentration of CNBr and decreased the formation of CNCl but had little effect on the sum of the molar concentration of CNCl and CNBr, as shown in Figure 1.[8] At bromide levels of 0.3 $mg\,L^{-1}$ or higher, CNBr became the dominant form.

Both CNCl and CNBr are labile in drinking waters and undergo degradation reactions at high pHs, or in the presence of free chlorine or sulfite. The degradation rates were well studied by two research groups.[9–11] The two studies reported similar CNCl degradation rates at various experimental conditions. At high pHs, both CNCl and CNBr undergo hydrolysis degradation to form cyanate. At pH 10, the half-life of CNCl has been estimated at 43 min,[11] and at 21 min.[9] The half-life of CNBr was reported as 34 min at pH 10.[11] The hydrolysis degradation may significantly affect the stability of CNCl and CNBr in distribution systems or during sample storage. In the presence of free chlorine, CNCl undergoes much faster degradation reactions. At pH 7 and a free chlorine concentration of 0.5 $mg\,L^{-1}$, the half-life of CNCl was 60 min.[11] At pH 9.2 and a free chlorine concentration of 5.68 $mg\,L^{-1}$, the half-life of CNCl was 1.9 min.[9] These results may well explain why high levels of CNCl or CNBr are rarely found in waters containing a free chlorine residual. In the presence of sulfite, CNCl also undergoes degradation reactions. In the presence of 25 μM sulfite and at pH 7, the half-life of CNCl was 76 min, whereas CNBr decomposed completely in 5 min.[11] In another study, the half-life of CNCl was 22 min at pH 7 and in the presence of 78.9 μM sulfite.[10] Therefore, there is a need to be extremely cautious when using sulfite to quench residual chlorine or other oxidants.

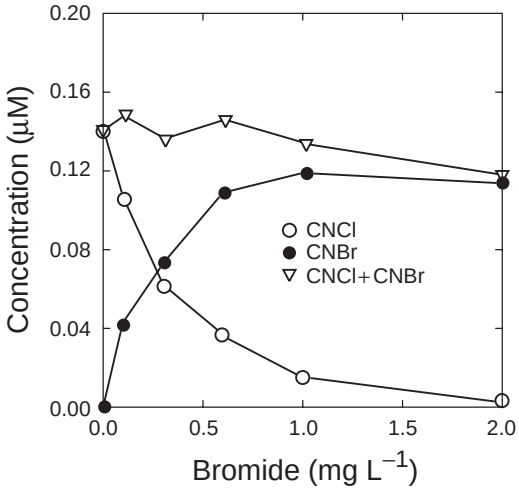

Figure 1 Effects of bromide on the formation of CNCl and CNBr. [Reprinted from *Water Research*, Vol. 27, Xie and Reckhow, 'A Rapid and Simple Analytical Method for Cyanogen Chloride and Cyanogen Bromide in Drinking Water', 511, Copyright (1993), with permission from Elsevier Science.]

Because of its hazardous nature and possible health effects, CNCl was listed in the USEPA's first drinking water priority list.[12] CNCl was also included in the USEPA ICR which required water utilities using chloramines to monitor CNCl in their finished water. CNCl may also become regulated under the USEPA Stage 2 D-DBP regulation. Because there are greater health concerns over brominated DBPs, CNBr, a brominated DBP in ozonated and chloraminated waters that are high in bromide, may also be included in future regulations.

2 STANDARDS PREPARATION

Accurate and stable standards are essential for quantitative analyses of environmental contaminants. Standards are needed to prepare calibration curves and determine other method performance parameters. For CNBr standards, the pure chemical can be purchased commercially (e.g. Aldrich Chemical Company, Inc., Milwaukee, WI, USA). For CNCl, pure CNCl gas (Island Pyrochemical Industries, Mineola, NY, USA) or CNCl diluted standards (Protocol Analytical Supplies, Inc., Middlesex, NJ, USA) may be purchased. The instability of CNCl diluted stocks and the hazardous nature and high cost of pure CNCl gas, however, have significantly limited the capability of many laboratories to analyze CNCl samples.[13,14] Under the ICR, as the USEPA could not provide performance evaluation samples or calibration standards, no laboratory was approved for CNCl analysis except the USEPA laboratory. All participating water utilities were requested to send CNCl samples to the USEPA laboratory for analysis.

2.1 Preparation of CNCl and CNBr Standards with Neat Chemicals

CNBr primary methanolic standards are prepared gravimetrically. Using an analytical balance, weigh an exact mass of pure CNBr into 10–50 mL of methanol. CNCl

primary methanolic standards are also prepared gravimetrically. Because of its low boiling point, CNCl is in gas form at room temperature. Prepare CNCl solution as follows.[7,8] Using a glass syringe, equipped with a luer lock adapter and a two-way syringe valve, transfer approximately 15 mL of the pure CNCl to a preweighed 40-mL septum-capped vial containing 10 mL of methanol. Weigh the vial again and use the weight difference to determine the exact amount of CNCl added. This primary standard yields a concentration of approximately 2500 mg L^{-1}. CNCl primary standards can also be prepared by exhausting 10–50 mL of CNCl gas over 10 mL methanol in a volumetric flask.[13]

Secondary or tertiary standard solutions are diluted in methanol from the primary stock. Store both CNBr and CNCl methanolic standard solutions in a −10 °C freezer. These standards are stable for 4–5 weeks.[13] The stability of standards can be improved by using different solvents (e.g. acetone or acetonitrile), as suggested by Protocol Analytical Supplies, Inc. Aqueous standards are prepared by diluting the secondary or tertiary standard with reagent water. Because of the hydrolysis degradation of both CNCl and CNBr, aqueous solutions should be freshly prepared. Aqueous standards should be used to prepare the calibration curves.

CNCl is a highly toxic and volatile chemical. Because of its hazardous nature, CNCl gas is very expensive to purchase, transport and dispose of. The price of a bottle of gaseous CNCl ranges from US $2200 for 0.7 lb to US $700 for 0.5 lb.[14] *Be extremely cautious when using CNCl gas*. Commercial CNCl gas may contain a high percentage of impurities (e.g. hydrogen cyanide) or stabilizers (e.g. tetrasodium pyrophosphate) which may significantly affect the accuracy of the analytical results.

2.2 In Situ Synthesis of CNCl Standards

In situ synthesis of CNCl standards was reported by Wu et al.[14] The method consists of quantitatively converting cyanide to CNCl in the presence of excess free chlorine or chloramine T (sodium *N*-chloro-*p*-toluenesulfonamide). As the CNCl synthesized standards are aqueous standards, there is no need to prepare primary, secondary or tertiary methanolic standard solutions. All chemicals used for CNCl synthesis should be reagent grade or better.

First, using an analytical balance, weigh and dissolve 0.212 g of KCN (CAS 151-50-8) into 100 mL of reagent water.[14] This cyanide stock solution is equivalent to 847 mg L^{-1} of CN^- or 2000 mg L^{-1} of CNCl solution. Take 1 mL of the cyanide stock solution and dilute it into 1000 mL using reagent water. The diluted cyanide solution is the primary cyanide solution. Take 5 mL of primary cyanide solution and dilute it with approximately 30 mL reagent water. Then add 4 mL of phosphate buffer (0.5 M), followed by the addition of NaOCl solution at a chlorine dose between 0.5 and 5 mg L^{-1}. Alternatively, 2 mL of chloramine T solution may be used instead of NaOCl solution. Immediately bring the mixture to 50 mL with reagent water. After exactly 2 min of reaction, add 25 mg of ascorbic acid to quench residual chlorine. After 30–40 s reaction, this primary CNCl standard, with a CNCl concentration of 200 µg L^{-1}, is ready for further dilution to various calibration standard concentrations. Cap the calibration standards headspace-free and store them at 4 °C in the dark.

The free chlorine-generated CNCl was stable for at least 2 days, and 1 h for chloramine T-generated CNCl, as shown in Figure 2.[14] The free chlorine-generated CNCl could be used as a calibration standard or a quality control check sample without repeated synthesis. Because of its instability, chloramine T-generated CNCl needs to be freshly synthesized; however, the synthesis is easily done in less than 10 min.

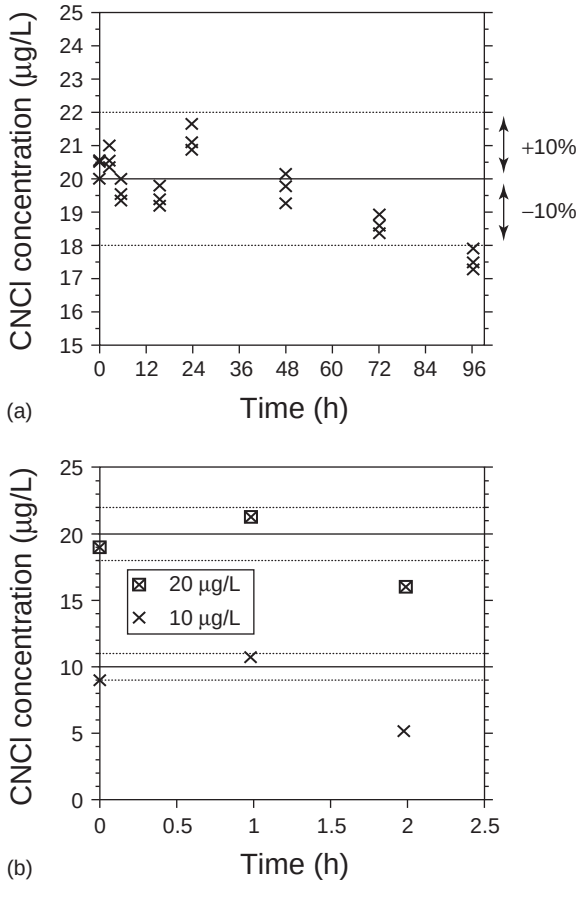

Figure 2 Stability tests of synthesized standards: (a) generated with hypochlorite; (b) generated with chloramine T. [Reprinted from *Water Research*, Vol. 32, Wu, Chadik and Schmidt, 'An In Situ Synthesis of Cyanogen Chloride as a Safe and Economical Aqueous Standard', 2867, Copyright (1998), with permission from Elsevier Science.]

As reported by Wu et al.,[14] the synthesized CNCl standards are comparable with standards prepared with pure CNCl gas. Microgram quantities of synthesized CNCl are easily disposed of by conversion to innocuous cyanate by simply raising the pH. The in situ synthetic procedure is a safe and economical way to prepare CNCl standards. This procedure is especially useful for laboratories in areas where CNCl is difficult to acquire.

2.3 Commercially Prepared Stock Standards

Commercial CNCl standards at 100, 200, 1000, 2000, or 5000 $\mu g\,L^{-1}$ in acetonitrile are available from Protocol Analytical Supplies, Inc. CNCl standards in methanol or acetone can also be obtained from them upon special request. The standards in methanol are least stable and those in acetonitrile are most stable. Sealed ampules and solvent-diluted standards should be stored at $-10\,^{\circ}C$ until use.

3 SAMPLE COLLECTION, PRESERVATION, AND STORAGE

3.1 Sample Collection

It is recommended that CNCl and CNBr samples are collected in duplicate or triplicate in nominal 40-mL vials with poly(tetrafluoroethylene) (PTFE)-faced septa and threaded polypropylene caps. If the sample is to be taken from a tap, flush the line for approximately 5 min to allow the water temperature to stabilize and to avoid sampling stagnant water. Adjust the flow to a stream that is gentle enough to prevent aeration of the sample and fill the vials. If a preservative is present do not rinse the bottles before filling or allow them to overflow. Seal the sample vials headspace free.

3.2 Sample Preservation

CNCl and CNBr degrade at high pHs or in the presence of free chlorine residual. Formation of CNCl and CNBr may continue if both natural organic matter and a combined chlorine residual are present. Dechlorination and pH adjustment are recommended in combination with cold storage to preserve the samples.

3.2.1 Dechlorination

Any free or combined chlorine residual must be quenched. Approximately 0.1 mL of 25 $mg\,mL^{-1}$ freshly prepared ascorbic acid solution per 40-mL sample is sufficient to remove a chlorine residual of 1 $mg\,L^{-1}$. Add an appropriate amount of ascorbic acid, but avoiding an excess which could degrade CNCl and CNBr. Be extremely cautious when using sulfite to quench free or combined chlorine residual. Both CNCl and CNBr degrade rapidly in the presence of sulfite.[10,11] Trace amounts of sulfite carried over into the purge-and-trap (PT) or GC (e.g. injector or column) system may also significantly impair instrumental sensitivity for CNBr. Ammonium chloride is not a true chlorine quenching reagent. It converts free chlorine to combined chlorine. Addition of ammonium chloride may not stop the continuous formation of CNCl and CNBr in water samples.[15]

3.2.2 pH Adjustment

To prevent the possible hydrolysis degradation of CNCl and CNBr, the pH of the samples must be lowered to 3.0–3.5. For typical drinking waters (i.e. not highly buffered and at pH 7–8), add approximately 0.2 mL of 1 M sulfuric acid to each 40-mL vial. Check the pH of unknown samples on an extra vial.

3.3 Sample Storage

Store samples iced or in a refrigerator at $4\,^{\circ}C$. Typical water samples, even with dechlorination, pH adjustment, and refrigeration, should be analyzed within 48 h for CNBr and within 2 weeks for CNCl.[7] It is best to analyze samples for CNCl and CNBr as soon as possible after collection.

4 ANALYTICAL METHODS

4.1 Purge-and-trap Gas Chromatography/Mass Spectrometry

4.1.1 General Discussion

The PT technique permits introduction of a high analyte concentration into the GC or GC/MS without any accompanying liquid (water or solvent). A high concentration factor is achieved because the analyte from a 25-mL water sample is injected into the GC. Volatile analytes are stripped from the water sample by a stream of nonreactive gas (helium or nitrogen) and trapped on a sorbent (resin or activated carbon). Analytes are desorbed by rapid heating of the trap and are carried into the GC by the carrier gas for separation and detection. With narrow-bore GC columns, a cryo-trap is used to refocus the compounds and reduce peak width. The PT technique has been used for many years for the analysis of volatile organic compounds including solvents and aromatic gasoline components, such as in USEPA methods 524.2[16] and 503.1.[17]

The PT GC/MS method for the analysis of CNCl was first developed by Flesch and Fair[13] using USEPA method 524.2.[16] The PT GC/MS method was the approved method used to collect occurrence data from

U.S. utilities using chloramine residuals under the ICR. Krasner et al. modified the method to include the analysis of CNBr in addition to CNCl.[5] Because CNBr is less volatile than CNCl, the sample had to be heated to achieve effective transfer to the trap. Detection is by mass spectrometry (MS), with spectral identification and quantitation based on a single ion chromatogram.

4.1.2 Sample Preparation and Analysis

Commercial PT systems are available or a system can be built from a fritted gas washing bottle plumbed to a stainless steel trap with deactivated fused silica transfer lines to the GC/MS. First bring samples to room temperature. Pour the sample into a 25-mL gas-tight syringe, insert the plunger, adjust the volume to 25 mL and add fluorobenzene (CAS 462-06-6) internal standard to give a concentration of 0.5 μg L^{-1}. Transfer the sample to the purge vessel and initiate the gas flow. Purge, trap, and desorption and GC/MS operating conditions are given in Table 1.[13,16,17] After purging is complete, switch the gas flow to carry the analytes from the sorbent trap to the GC and heat the trap. Simultaneously, a short piece of deactivated fused silica precolumn is cooled with liquid nitrogen to refocus the analytes. Then ballastically heat the cryotrap to inject the sample onto the analytical GC column. The GC injector is not used. Hold the GC oven at −10 °C from the start of sorbent trap desorption until the start of the GC temperature program at the end of the cryofocusing process.

Aqueous standards of CNCl and CNBr analyzed by the same method should be used to establish a calibration curve based on the ratio of the area of analyte divided by the area of the internal standard. The masses used for quantitation are 61 *m*/*z* for CNCl, 105 *m*/*z* for CNBr, and 96 *m*/*z* for the internal standard, fluorobenzene. Typical spectra for CNCl, CNBr, and fluorobenzene, are shown in Figure 3.

4.1.3 Discussion

Typical gas chromatograms of a PT analysis of CNCl and CNBr calibration standard are shown in Figure 4. The GC run time was less than 9 min. A typical calibration curve for CNCl over a concentration range of 0.2–20 μg L^{-1} was linear with an R^2 value of 0.997, and for CNBr over

Table 1 Operating conditions for PT GC/MS

Parameters	CNCl-only method	CNBr and CNCl method
PT parameters		
Sample preheat	Ambient	60 °C for 1.5 min
Purge conditions	Ambient temperature for 4 min	60 °C for 12 min
Purge gas	Helium at 40 ml min^{-1}	Helium at 40 ml min^{-1}
Trap type	Tenax™	Tenax/charcoal™
Desorption parameters	180 °C for 1.5 min	180 °C for 1.5 min
Cryofocusing parameters		
Cooldown (liquid N_2)	−150 °C for 1.5 min	−150 °C for 1.5 min
Thermal injection	to 220 °C in 0.85 min	to 200 °C in 0.75 min
GC parameters for analytical column		
Type	DB-5	DB-5
Length	30 m	30 m
Internal diameter	0.25 mm	0.25 mm
Film thickness	1 μm	1 μm
Injector temperature	200 °C	150 °C
Oven temperature program		
Initial temperature	10 °C	10 °C
Initial holding time	4 min	2 min
Temperature ramp	20 °C min^{-1}	15 °C min^{-1}
Final temperature	184 °C	130 °C
Final holding time	3 min	6 min
MS parameters		
GC/MS transfer line	200 °C	180 °C
Source temperature	180 °C	180 °C
MS resolution	500 or unit mass	500 or unit mass
Ionization voltage	70 eV	70 eV
Scan range	40–200 *m*/*z*	40–200 *m*/*z*
Scan time	0.6 s	0.6 s

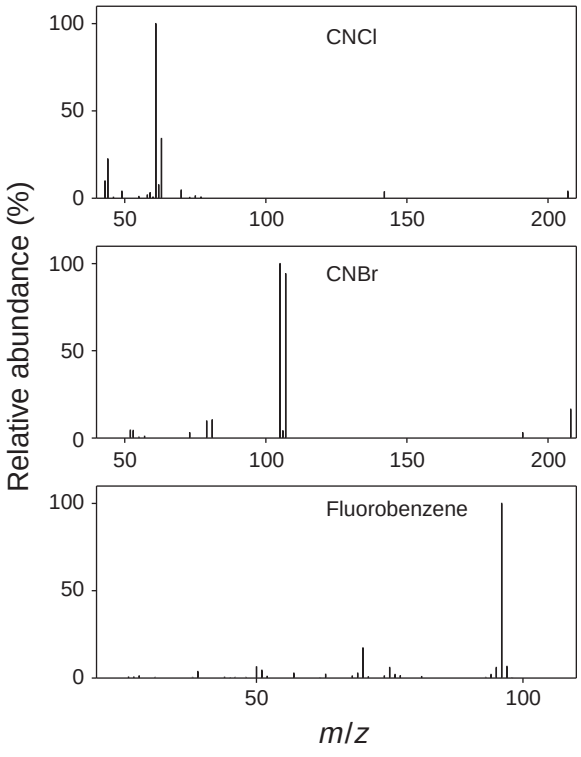

Figure 3 Mass spectra of CNCl, CNBr, and fluorobenzene.

a concentration range of 0.5–20 $\mu g\,L^{-1}$ a quadratic fit with an R^2 value of 0.997.[5] The method detection limits, based on seven replicate analyses, were 0.13 $\mu g\,L^{-1}$ for both CNCl and CNBr. However, the CNBr calibration curve does not pass through zero, indicating a threshold concentration. Therefore a minimum reporting limit of 0.5 $\mu g\,L^{-1}$ is used for CNBr. By analyzing 12 treated surface water samples spiked with 5 or 10 $\mu g\,L^{-1}$ cyanogen chloride and cyanogen bromide, the mean recovery was determined to be $96 \pm 18\%$ for CNCl and $88.5 \pm 20.9\%$ for CNBr.[5]

The heated PT GC/MS method was the first method to analyze CNBr in drinking water and provided valuable information about its formation and biodegradation.[5] A good transfer efficiency (25%) was achieved for CNBr. However, the heated PT method is difficult to run continuously because water vapor is carried over to the trapping system. When both CNBr and CNCl analyzes are required, headspace GC/ECD and micro LLE GC/ECD are more cost effective and precise methods. Ambient temperature PT GC/MS is an efficient method for the analysis of CNCl and can be run in conjunction with the determination of other volatile organic compounds. It also provides the advantage of mass-spectral compound identification and can eliminate interference from unresolved GC peaks.

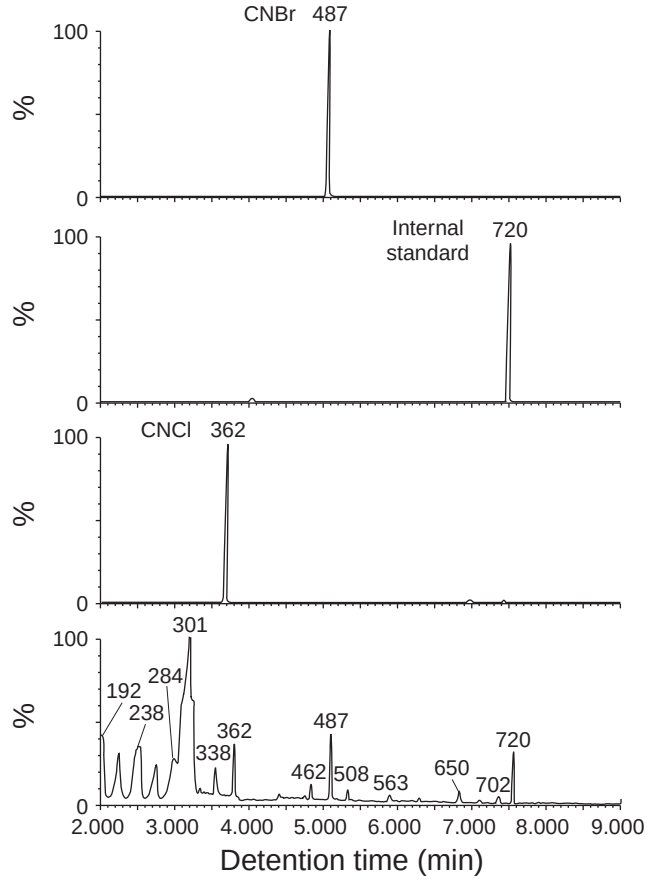

Figure 4 Typical selected ion chromatograms with the PT GC/MS method.

4.2 Headspace Gas Chromatography/Electron-capture Detection

4.2.1 General Discussion

Headspace GC method is a common procedure for analyzing volatile organic compounds in aqueous samples. The analytical procedure includes headspace preparation and GC analysis. The headspace preparation involves transferring volatile substances from the liquid phase into the gas phase (or headspace). Increasing the interfacial area between the gas and liquid phases by shaking the samples increases the efficiency of mass transfer and reduces the time required for reaching the gas–liquid equilibrium. At equilibrium, the concentration of a substance in gas phase is proportional to its concentration in the liquid phase, according to Henry's law. The concentration of the substance in the gas phase is analyzed by GC. The calibration curves are prepared by analyzing aqueous standards at various concentrations. As most of the volatile substances occur in water samples at microgram per liter levels, linear calibration curves are easily obtained.

The CNCl has a boiling point of 13.1 °C at 1 atm. At 20 °C, the Henry's law constant was estimated as 0.9 atm L mol^{-1}.[8] CNBr has a boiling point of 61–62 °C at 1 atm and a Henry's law constant of 0.16 atm L mol^{-1} at 20 °C.[8] These properties indicate that CNCl and CNBr can be easily analyzed by headspace GC/ECD.

4.2.2 Sample Preparation and Analysis

Pour water samples into 40-mL vials without aeration and cap the vials headspace free with PTFE-faced septa and screw caps. Prior to headspace preparation, bring the water samples to ambient temperature (20 °C). An internal standard, 1,2-dibromopropane (CAS 78-75-1), may be added at this point to monitor method performance.[18] The recommended concentration for the internal standard is 50 μg L^{-1}. A bare syringe needle and a syringe equipped with a stainless steel needle are needed to prepare the headspace. First, fill the syringe with 10 mL of nitrogen (99.999%). Force both needles through the septa on the sample vial. The tip of the bare needle should be placed close to the septa and the tip of the syringe needle should be placed in the middle of the sample vial. While keeping the sample vial upside down, gently depress the plunger on the nitrogen-containing syringe. The displaced sample volume is expelled through the bare needle. Remove the needles and vigorously shake the sample vial for 1 min by hand. After settling for 10–20 s, the gas phase is ready for sampling. Withdraw 400 μL of headspace sample using a 1 mL or 500 μL gas-tight syringe, and inject the headspace sample into the GC in the split mode.

A gas chromatograph equipped with a split/splitless injector and an electron capture detector (ECD) is required for CNCl and CNBr analysis. Typical GC operating conditions for headspace GC/ECD are shown in Table 2.[8]

4.2.3 Discussion

Typical gas chromatograms obtained with headspace GC/ECD are shown in Figure 5.[8] Both CNCl and CNBr are well separated. A small negative peak was observed just prior to the elution of CNCl.[8] This caused a non-zero intercept in the calibration curves. This phenomenon was not observed by Sclimenti et al.[18] The calibration curve for CNBr always passed through the origin.[8]

The method detection limits were estimated to be 0.04 μg L^{-1} for CNCl and 0.2 μg L^{-1} for CNBr.[8] By analyzing 11 fulvic acid solution samples (4 mg L^{-1} dissolved organic carbon) spiked with 5 μg L^{-1} CNCl and 10 μg L^{-1} CNBr, the spiking recovery was determined

Table 2 GC operating conditions for headspace GC/ECD

Parameter	Setting
Analytical column	
Type	DB-1701
Length	30 m
Internal diameter	0.32 mm
Film thickness	1 μm
Carrier gas	nitrogen (99.999%)
Carrier gas flow	2.3 mL min^{-1}
Oven temperature	
Initial temperature	35 °C
Initial holding time	2 min
Temperature ramp	15 °C min^{-1}
Final temperature	110 °C
Final holding time	3 min
Injector parameters	
Type	Split
Temperature	100 °C
Split flow	20 mL min^{-1}
Detector	
Type	ECD
Temperature	250 °C
Make-up gas flow	20 mL min^{-1}

Reprinted from *Water Research*, Vol. 27, Xie and Reckhow, 'A Rapid and Simple Analytical Method for Cyanogen Chloride and Cyanogen Bromide in Drinking Water', 508, Copyright (1993), with permission from Elsevier Science.

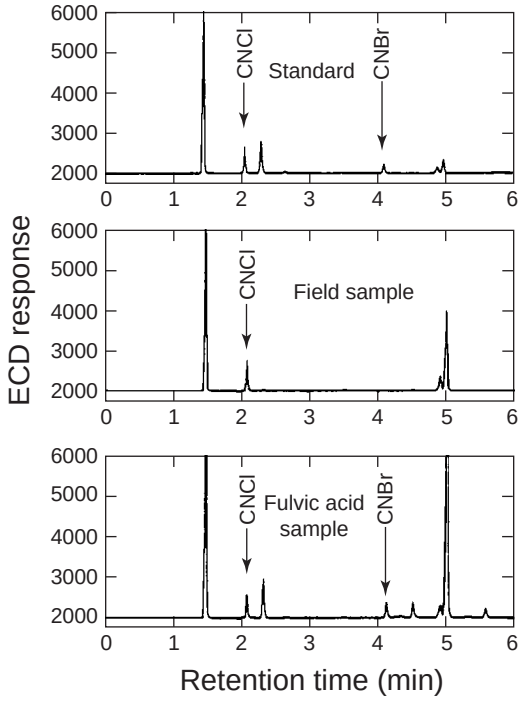

Figure 5 Typical gas chromatograms with headspace GC/ECD. [Reprinted from *Water Research*, Vol. 27, Xie and Reckhow, 'A Rapid and Simple Analytical Method for Cyanogen Chloride and Cyanogen Bromide in Drinking Water', 508, Copyright (1993), with permission from Elsevier Science.]

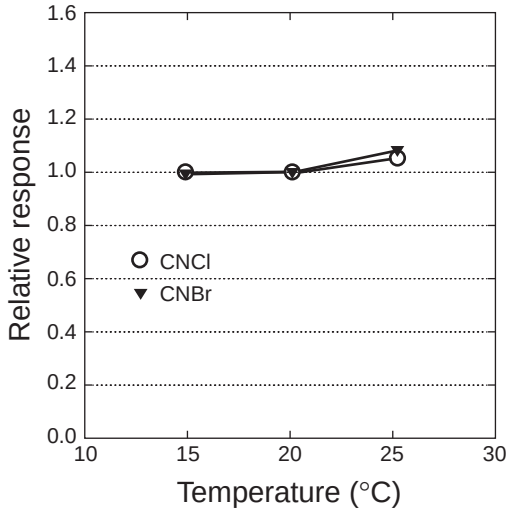

Figure 6 Effects of extraction temperature on GC peak areas. [Reprinted from *Water Research*, Vol. 27, Xie and Reckhow, 'A Rapid and Simple Analytical Method for Cyanogen Chloride and Cyanogen Bromide in Drinking Water', 509, Copyright (1993), with permission from Elsevier Science.]

to be 91.1% for CNCl and 74.8% for CNBr. The precision was 6.6% for CNCl and 6.2% for CNBr. Sclimenti et al. estimated the quantitation limit for CNCl at approximately $0.5\,\mu g\,L^{-1}$.[18] By analyzing 13 water samples spiked with 5 or $10\,\mu g\,L^{-1}$ CNCl, the mean matrix-spike recovery for CNCl was determined at $107 \pm 14\%$ and its precision was $14.2 \pm 9.5\%$.

As shown in Figures 6 and 7, the variations of normal room temperature and drinking water sample salinity

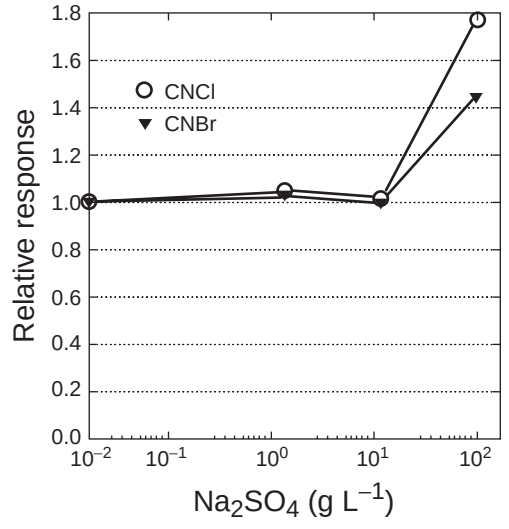

Figure 7 Effects of salt addition on GC peak areas. [Reprinted from *Water Research*, Vol. 27, Xie and Reckhow, 'A Rapid and Simple Analytical Method for Cyanogen Chloride and Cyanogen Bromide in Drinking Water', 509, Copyright, (1993) with permission from Elsevier Science.]

had little impact on the analytical results.[8] Adding supplemental salt at $100\,g\,L^{-1}$ and/or increasing temperature beyond 25 °C might improve the detection limit for both CNCl and CNBr. However, addition of such large amounts of any salt or heating samples at high temperature would present a new opportunity for contamination or sample degradation, especially for CNBr.

The headspace GC/ECD method is a simple and rapid method for analyzing CNCl and CNBr. The total time for headspace preparation and GC analysis is less than 12 min. Having analyzed more than 250 samples with this method,[8] it was concluded that this method is rapid, simple, and robust.

4.3 Micro Liquid–Liquid Extraction Gas Chromatography/Electron-capture Detection

4.3.1 General Discussion

Liquid–liquid extraction (LLE) is a common concentration procedure for organic analysis. Traditionally a separatory funnel or a continuous extraction apparatus is used to facilitate the extraction of liter-size samples. Micro LLE uses a small volume of water sample and extraction solvent. Generally, the extraction is conducted in a 40-mL sample vial and there is no need to concentrate the extract. Addition of a salt (e.g. sodium sulfate) is used to increase the extraction efficiency and improve the separation of solvent and aqueous phases. This extraction technique has been used for analyzing many DBPs, including THMs (USEPA method 551.1), HAAs (USEPA Method 552.2 and Standard Method 6251), aldehydes (Standard Method 6252), and ketoacids.[19–22]

The micro LLE GC/ECD method was first developed for the analysis of CNBr in ozonated waters.[6] The method used methyl *tert*-butyl ether (MTBE) (CAS 1634-04-4) as the extraction solvent, sodium sulfate as the supplemental salt to increase the extraction efficiency, and GC/ECD for separation and detection. This method was further developed and expanded for analysis of both CNCl and CNBr in chloraminated waters.[7] A similar method was also developed for CNCl analysis,[23] where cyclohexane was used as the extraction solvent.

4.3.2 Sample Preparation and Analysis

Prior to sample extraction, bring the samples to room temperature. Use 40-mL glass vials with PTFE-faced septa and screw caps for the sample extraction. To a vial containing a 30-mL aliquot of sample, add 10 g of Na_2SO_4 and 4 mL of MTBE spiked with $100\,\mu g\,L^{-1}$ internal standard, 1,2-dibromopropane. Shake once or twice to disperse the salt and then proceed to the next

sample. Shake all the vials for 10 min using a mechanical shaker or 2 min by hand. After the solvent and water phases are well separated, transfer the MTBE layer to two 1.5-mL autosampler vials for GC/ECD analysis. The extracts can be stored at 4 °C for 90 days.

A gas chromatograph installed with a septum-equipped, temperature-programmable or split/splitless injector, ECD, and autosampler is needed for CNCl and CNBr analysis. A 30 m × 0.32 mm, 1.8 μm film thickness DB-624 fused silica capillary column or equivalent is needed to obtain baseline resolution of all the analytes. Other operating conditions are listed in Table 3.[7]

4.3.3 Discussion

A typical gas chromatogram of 10 μg L^{-1} of extracted CNCl, CNBr, and calibration standard (1,2-dibromopropane) is shown in Figure 8.[7] All three components are well separated. The GC run time was less than 14 min. A linear calibration curve ranging from 0.5 to 24 μg L^{-1} was obtained for CNCl with an R^2 value of 0.9993, and

Table 3 GC operating conditions for the micro LLE GC/ECD

Parameter	Setting
Analytical column	
Type	DB-624
Length	30 m
Internal diameter	0.32 mm
Film thickness	1.8 μm
Carrier gas	Helium (99.999%)
Carrier gas flow	3.5 mL min^{-1}
Oven temperature program	
Initial temperature	25 °C
Initial holding time	1 min
First temperature ramp	10 °C min^{-1}
Second temperature	120 °C
Second holding time	0 min
Second temperature ramp	35 °C min^{-1}
Final temperature	190 °C
Final holding time	1 min
Injector parameters	
Type	Split
Initial temperature	35 °C
Initial holding time	0 min
Temperature ramp rate	180 °C min^{-1}
Final temperature	200 °C
Final holding time	12.59 min
Split flow	20 mL min^{-1}
Detector	
Type	ECD
Temperature	300 °C
Make-up gas	nitrogen (99.999%)
Make-up gas flow	27 mL min^{-1}

Reprinted from Proceedings of the 1994 Water Quality Technology Conference,[7] with permission. Copyright 1995. American Water Works Association.

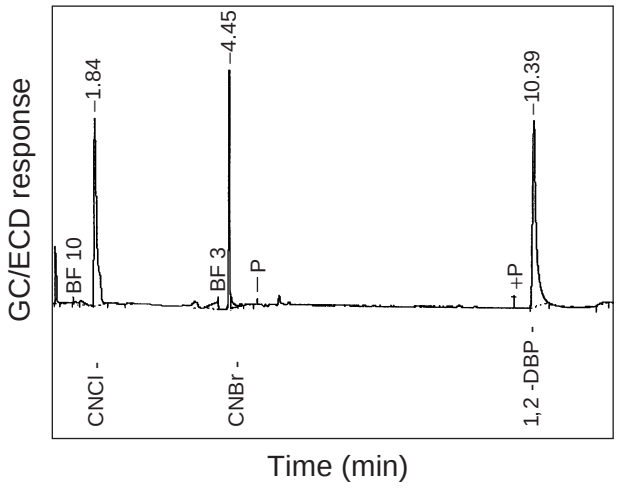

Figure 8 A typical gas chromatogram with the micro LLE GC/ECD method. [Reprinted from *Proceedings of the 1994 Water Quality Technology Conference*,[7] with permission. Copyright 1995. American Water Works Association.]

second-order polynomial calibration curves ranged from 0.5 to 24 μg L^{-1} for CNBr with an R^2 value of 0.9993. The method detection limits for CNCl and CNBr were 0.13 and 0.26 μg L^{-1}. By analyzing 10 treated surface water samples spiked with 5–7 μg L^{-1} of CNCl or CNBr the mean recovery for CNCl was 98 ± 6.3%, whereas the mean recovery for CNBr was 100 ± 10%. In another study, the quantitation limit for CNCl was estimated at 0.5 μg L^{-1}.[18]

Four organic solvents were studied by Sclimenti et al.[7] Both hexane and isooctane co-eluted with CNCl and could not be used as the extraction solvents. Pentane extracted both CNCl and CNBr well. In comparison with pentane, however, MTBE gave a better extraction for CNBr, giving a similar extraction efficiency for CNCl. Therefore, the micro LLE GC/ECD method uses MTBE as the extraction solvent.

The effects of a salt, sodium sulfate, and shaking time on the extraction efficiency were also studied.[7] Addition of Na_2SO_4 at a concentration range from 5 g to 12 g per 30-mL aliquot increased the extraction efficiency of CNCl. However, CNBr recoveries began to diminish at the same salt concentration range. This may be due to trace levels of nucleophiles which adversely affect the stability of CNBr.[8] For the LLE GC/ECD method, a salt concentration of 10 g of Na_2SO_4 per 30-mL of aliquot and a shaking time of 10 min were chosen. These parameters represent a compromise among the conditions that were most suitable for each analyte separately.

The micro LLE GC/ECD method is efficient for simultaneously determining both CNCl and CNBr in chloraminated waters. The method performance study showed a high degree of precision and accuracy. Compared to the

PT GC/MS system, the GC/ECD is a much more common system in most water-testing laboratories. The downtime for GC/ECD is significantly less than for GC/MS. Compared to the headspace GC/ECD method, the LLE sample injection could be easily automated with a liquid autosampler.

5 SUMMARY

Both LLE GC/ECD and headspace GC/ECD are viable methods for determining CNCl and CNBr in drinking water. Selection may depend upon the resources of the particular laboratory. The PT GC/MS method (modified USEPA 524.2) can be efficiently used to analyze for CNCl, especially when other volatile solvents are also being determined and when mass spectral identification is desired.

ABBREVIATIONS AND ACRONYMS

chloramine T	Sodium *N*-Chloro-*p*-toluene-sulfonamide
DBP	Disinfection By-product
D-DBP	Disinfectants and Disinfection Byproduct
ECD	Electron Capture Detector
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron-capture Detection
HAA	Haloacetic Acid
ICR	Information Collection Rule
LLE	Liquid–Liquid Extraction
LLE GC/ECD	Liquid–Liquid Extraction Gas Chromatography/Electron-capture Detection
MCL	Maximum Contaminant Level
MS	Mass Spectrometry
MTBE	Methyl *tert*-butyl Ether
PT	Purge-and-trap
PTFE	Poly(tetrafluoroethylene)
PT GC/MS	Purge-and-trap Gas Chromatography/Mass Spectrometry
THM	Trihalomethane
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Gas Chromatography by Direct Aqueous Injection in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Food **(Volume 5)**
Sample Preparation, Headspace Techniques

Gas Chromatography **(Volume 12)**
Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

REFERENCES

1. US Environmental Protection Agency, 'National Primary Drinking Water Regulations: Monitoring Requirements for Public Drinking Water Supplies; Final Rule', *Fed. Regist.*, **61**(94), 24354–24388 (1996).
2. US Environmental Protection Agency, 'National Primary Drinking Water Regulations: Disinfectants and Disinfection Byproducts; Final Rule', *Fed. Regist.*, **63**(241), 69390–69476 (1998).
3. Water Quality Division Disinfection Committee, 'Survey of Water Utility Disinfection Practices', *J. Am. Water Works Assoc.*, **84**(9), 121–128 (1992).
4. S.W. Krasner, M.J. McGuire, J.G. Jacangelo, N.L. Patania, K.M. Reagan, E.M. Aieta, 'The Occurrence of Disinfection By-products in US Drinking Water', *J. Am. Water Works Assoc.*, **81**(8), 41–53 (1989).
5. S.W. Krasner, R. Chinn, C.J. Hwang, S.E. Barrett, 'Analytical Methods for Brominated Organic Disinfection By-products', *Proceedings of 1990 AWWA Water Quality Technology Conference*, AWWA, Denver, CO, USA, 1991.
6. M.J. West, T.K. Lieu, S.W. Krasner, 'Analytical Determination of Cyanogen Bromide by Electron-capture Gas Chromatography', *Proceedings of 1991 AWWA Water Quality Technology Conference*, AWWA, Denver, CO, USA, 1992.
7. M.J. Sclimenti, C.J. Hwang, G.E. Speitel, A.C. Diehl, 'The Simultaneous Determination of Cyanogen Chloride and Cyanogen Bromide in Chloraminated Waters by a Simplified Microextraction GC/ECD Technique', *Proceedings of 1994 AWWA Water Quality Technology Conference*, AWWA, Denver, CO, USA, 1995.
8. Y. Xie, D.A. Reckhow, 'A Rapid Simple Analytic Method for Cyanogen Chloride and Cyanogen Bromide in Drinking Water', *Water Res.*, **27**(3), 507–511 (1993).
9. P.L. Bailey, E. Bishop, 'Hydrolysis of Cyanogen Chloride', *J. Chem. Soc.*, **9**, 912–916 (1973).
10. P.L. Bailey, E. Bishop, 'Reaction of Sulphite Ions with Cyanogen Chloride', *J. Chem. Soc.*, **9**, 917–921 (1973).
11. Y. Xie, D.A. Reckhow, 'Stability of Cyanogen Chloride in the Presence of Sulfite and Chloride', *Proceedings of 1992 AWWA Water Quality Technology Conference*, AWWA, Denver, CO, USA, 1993.

12. US Environmental Protection Agency, 'Substitution of Contaminants and Drinking Water Priority List', *Fed. Regist.*, **53**, 1892 (1988).
13. J.J. Flesch, P.S. Fair, 'The Analysis of Cyanogen Chloride in Drinking Water', *Proceedings of 1988 AWWA Water Quality Technology Conference*, AWWA, Denver, CO, USA, 1989.
14. W.W. Wu, P.L. Chadik, C.J. Schmidt, 'An In Situ Synthesis of Cyanogen Chloride as a Safe and Economical Aqueous Standard', *Water Res.*, **32**(9), 2865–2869 (1998).
15. S.W. Krasner, M.J. Sclimenti, C.J. Hwang, 'Experiences with Implementing a Laboratory Program to Sample and Analyze for Disinfection By-products in a National Study', *Disinfection By-products: Current Perspectives*, AWWA, Denver, CO, USA, 1989.
16. US Environmental Protection Agency, *Method 524.2: Volatile Organic Compounds in Water by Purge-and-trap Capillary Gas Chromatography/Mass Spectrometry, Environmental Monitoring and Support Laboratory*, EPA, Cincinnati, OH, USA, 1986.
17. US Environmental Protection Agency, *Method 503.1: Volatile Aromatic and Unsaturated Organic Compounds in Water by Purge and Trap Gas Chromatography, Environmental Monitoring Systems Laboratory*, EPA, Cincinnati, OH, USA, 1989.
18. M.J. Sclimenti, C.J. Hwang, S.W. Krasner, 'A Comparison of Analytical Techniques for Determining Cyanogen Chloride in Chloraminated Drinking Water', in *Water Disinfection and Natural Organic Matter Characterization and Control*, eds. R.A. Minear, G.L. Amy, American Chemical Society, Washington, DC, USA, 1996.
19. US Environmental Protection Agency, *Method 551.1: Determination of Chlorination Disinfection Byproducts, Chlorinated Solvents, and Halogenated Pesticides/Herbicides in Drinking Water by Liquid–Liquid Extraction and Gas Chromatography with Electron-capture Detection*. National Exposure Research Laboratory, Cincinnati, OH, USA, 1995.
20. US Environmental Protection Agency, *Method 552.2: Determination of Haloacetic Acids and Dalapon in Drinking Water by Liquid–Liquid Extraction, Derivatization and Gas Chromatography with Electron Capture Detection*, National Exposure Research Laboratory, Cincinnati, OH, USA, 1995.
21. American Public Health Association, American Water Works Association, Water Environmental Association, *Standard Methods for the Examination of Water and Wastewater*, 20th edition, APHA, AWWA, and WEF, Washington, DC, USA, 1998.
22. Y. Xie, D.A. Reckhow, D.C. Springborg, 'Analyzing HAAs and Ketoacids without Diazomethane', *J. Am. Water Works Assoc.*, **90**(4), 131–138 (1998).
23. S.A. Andrews, P.M. Huck, 'Analysis of Haloacetic Acids, Cyanogen Chloride and Chloral Hydrate in Drinking Water Using the Simulated Distribution System', *Proceedings of 1992 AWWA Annual Conference – Water Quality*, AWWA, Denver, CO, USA, 1992.

Detection and Quantification of Environmental Pollutants[1]

Robert D. Gibbons
University of Illinois, Chicago, USA

1 Introduction 1
2 Detection Limits for Single Concentration Designs 2
2.1 Kaiser–Currie Method 3
2.2 United States Environmental Protection Agency 40CFR 136 3
2.3 Prediction Limits 4
2.4 Tolerance Limits 4
2.5 Simultaneous Control of False Positive and False Negative Rates 5
2.6 Limitations 5
3 Detection Limits for Calibration Designs 6
3.1 Illustration 7
4 Experimental Design of Detection Limit Studies 9
5 Quantification Limits 10
5.1 The United States Environmental Protection Agency Minimum Level 10
5.2 Alternative Minimum Level 11
5.3 Illustration 12
5.4 Alternative Formulations 12
5.5 The Estimator According to Gibbons et al. 12
6 Discussion 13
Abbreviations and Acronyms 14
Related Articles 14
Appendix 1 14
References 16

With increasing concern over chemicals that are potential health hazards at low levels, determination of limits of detection and quantification have undergone considerable scrutiny. Most traditional detection limit (DL) estimators suffer from extensive statistical and/or conceptual limitations. In this article, traditional detection and quantification limit estimators are described and critically evaluated. Methods are categorized into decision limits, DLs and quantification limits. The methods are further categorized into single concentration design versus calibration design methodologies. While the single concentration design methods are useful for fixing ideas and clarifying definitions they are shown to be extremely limited in practice since dependence of variability on concentration can be neither estimated nor incorporated. Calibration-based detection and quantification limit estimators are described, compared and contrasted. Discussion of nonconstant variance, multiple future detection decisions and simultaneous control of Type I and II errors is provided. The various methods are illustrated using real data and experimental design issues for DL studies discussed.

1 INTRODUCTION

Of critical concern in protection of human health and the environment is the earliest possible detection of hazardous environmental constituents (i.e. air, surface water, groundwater, drinking water). The very presence of anthropogenic compounds in environmental media may lead to assessment and potential remediation of industrial activity. As an example, the presence of vinyl chloride, a known carcinogenic initiator, in a groundwater sample obtained from a monitoring well located downgradient of a hazardous or municipal waste-disposal facility may result in an expensive site assessment, possible corrective action or closure of the faculty. In some cases, these consequences may be the result of a single detection decision in a single environmental sample. In other cases, the measured concentration of a constituent is compared with a concentration limit, possibly health-based, and if the measured concentration exceeds the limit, remedial action is taken. Unfortunately, in many cases the limit of detection for the constituent in a particular medium (e.g. surface water) is reported to be considerably larger than the health-based concentration limit. There is public concern regarding protection of human health and the environment in that potentially harmful levels of a chemical may exist in the environment but since they are undetectable no regulatory impact decision can be made.

In response to these concerns there has been tremendous motivation to develop new analytical methods capable of detecting these compounds at lower levels. Furthermore, these concerns have led to a proliferation of new definitions and statistical approaches to estimating limits of detection. These new definitions and statistical estimators often lead to further confusion and suffer from statistical and conceptual errors.[1–3] New methods proposed for setting limits on detection may result in lower limits regardless of the actual analytical capabilities of the method.

[1] Portions reprinted from *Environmental and Ecological Statistics* and the *Encyclopedia of Environmental Analysis and Remediation*, with permission.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Unfortunately, virtually all practical applications of environmental detection decisions in the United States are based on the most severely limited of the definitions.[4,5] They are carried out by the United States Environmental Protection Agency (USEPA). Not only are the statistical problems with USEPA's estimator, termed the method detection limit (MDL), insurmountable but the use of this estimator has led to the routine practice of setting DL at lower and lower levels regardless of the true analytical capabilities of the method and/or laboratory. The purpose of this article is to illustrate these statistical limitations and to provide a review of less problem-prone alternative estimators. To do so, this article reviews a variety of methods from both statistical and conceptual frameworks. The intended audience is both the statistician interested in environmental applications and the environmental scientist interested in the conceptual and statistical foundations of limits of detection. Public policy in these areas has not been adequately examined in either the statistical or chemometric literatures.

There are a myriad of terms for limits of detection: MDL, DL, instrument detection limit (IDL), practical method detection limit (PMDL) etc. However, their conceptual foundations as well as implementation algorithms are often quite diverse, leading to tremendous inconsistency in practice. To provide a framework to compare and contrast these various methodologies the pioneering definitions of Currie[6] are used, since they form the basis of most approaches. Currie defined two levels; the ''decision limit'' and the ''DL''. The decision limit is the measured concentration ''at which one may decide whether or not the result of an analysis indicates detection''.[6] The DL is the true concentration ''at which a given analytical procedure may be relied upon to lead to a detection''.[6] From a statistical perspective, the decision limit is the critical value for a test of the null hypothesis ''analyte absent'' versus the alternative hypothesis ''analyte present''. Note that the DL is a direct function of the selection of the critical value.

Operationally, the decision limit or critical level L_C, is the measurement above which the response signal significantly differs from zero. When a measurement exceeds L_C we can make the binary decision, ''detected''. Note that when the true concentration is zero, the probability of the correct decision, ''not detected'' is $1-\alpha$ where α is the Type I error rate or false positive rate of the statistical test (e.g. $\alpha = 0.01$). Assuming a symmetric distribution of measurement errors, when the true concentration is equal to L_C the probability of reporting it below L_C is 50%. This is termed the Type II error rate or false negative rate (β). To control both Type I and II errors Currie then developed the DL L_D. When the true concentration is L_D, then the Type II error rate (β) using L_C as the decision limit is small. For example, if we assume that $\alpha = \beta = 0.01$, then another way of stating this is that 99% of the measurements for samples not containing the analyte will be less than the decision limit L_C and 99% of the measurements with true concentration at the DL L_D will exceed the decision limit L_C.[7]

Investigators have defined the limit of detection to include control of both false positive and false negative rates, varying levels of confidence, uncertainty in the concentration mean and variance and in the calibration function that relates instrument response to concentration, variance components related to multiple instruments and analysts, application to multiple future detection decisions, and nonconstant variance and baseline instrument response.[1,6,8–12] However, these operational definitions may cloud the distinction between decision limits and DLs (as an example see Glaser et al.[5])

An alternative approach and review of many of these issues can be found in Lambert et al.[7] and Hastie and Tibshirani[13] The method introduced by Lambert et al. provides much information regarding the statistical properties of low-level analytical measurements conditional on a fixed and pre-existing method of making binary detection decisions, but does not provide an independent estimate of the limit of detection. In a sense, Lambert's method models the detection criterion of the analyst but not the actual capabilities of the analytical method itself.

In the following sections we examine the strengths and weaknesses of a variety of approaches to the problem of estimating limits of detection in use or proposed for widespread application. These approaches fall into two general categories; single concentration-based methods and calibration-based methods. Methods based on single concentration designs are useful for fixing ideas and definitions, particularly the critical distinction between decision limits and DLs. In general single point designs lead to problems which severely limit their use in practice which can be eliminated through use of calibration-based methods. These methods are fully described and classified as decision limits or DLs. Generalization of calibration-based methods to problems of nonconstant variance, multiple future detection decisions and alternative approaches to controlling Type I and II error rates are presented. Application of the various methodologies are illustrated using real data and experimental design considerations are discussed. Following discussion of DLs, various estimators of the quantification limit are described and illustrated.

2 DETECTION LIMITS FOR SINGLE CONCENTRATION DESIGNS

Single concentration designs involve spiking a series of n samples with a fixed concentration of a compound

and using variability in instrument response at that concentration to derive a decision limit, DL or both. In the following, the major strategies for estimating decision limits and DLs from single concentration designs are described.

2.1 Kaiser–Currie Method

Based on developments due to Kaiser[14–16], Currie[6] described a two-stage procedure for calculating the DL L_D. At the first level of analysis, Currie computes the decision limit L_C (Equation 1) as:

$$L_C = z_{1-\alpha}\sigma_0 \tag{1}$$

where σ_0 is the population standard deviation of the response signal when the true concentration (x) is zero (i.e. the standard deviation of the net signal found in the population of blank samples),
and $z_{1-\alpha}$ is a multiplication factor based on the $(1-\alpha)100$ percentage point of the standardized normal distribution.

For example, the one-sided 99% point of the normal distribution is 2.33; therefore, the decision limit (Equation 2) for a 1% false positive rate is defined as:

$$L_C = z_{1-\alpha}\sigma_0 = 2.33\sigma_0 \tag{2}$$

As discussed, when the true concentration is equal to L_C and the measurement error distribution is symmetric, the Type II error rate at the decision limit is 50%. That is, we have a 50% chance of declaring that the analyte is not present when the analyte in fact is present at the decision limit. To provide an acceptable Type II error rate, Currie defined the DL (L_D) (Equation 3) as

$$L_D = L_C + z_{1-\beta}\sigma_D \tag{3}$$

where σ_D is the population standard deviation of the response signal at L_D (or net response signal after subtracting the background signal),
and β is the acceptable Type II error rate (i.e. false negative rate).

Currie points out that if we make the simplifying assumption that $\sigma_0 = \sigma_D$ (i.e. the variability of the signal is constant in the range of L_C to L_D) and that the risk of false positive and false negative rates is equivalent (i.e. $z_{1-\alpha} = z_{1-\beta} = z$) then the DL is simply (Equation 4):

$$L_D = L_C + z_\beta\sigma_D = z(\sigma_0 + \sigma_D) = 2L_C \tag{4}$$

or twice the decision limit. For $\alpha = \beta = 0.01$, the DL is therefore $4.66\sigma_0$. If the true concentration is L_D then the probability of a measured valued below L_C is 1%.

In reviewing Currie's method it is critically important to note that the only case considered is the one in which the population values σ_0 and σ_D are known. In practice, however population values are never known and the methods described in following sections are required to incorporate uncertainty in sample-based estimates of these statistics in decision-limit and DL estimates.

2.2 United States Environmental Protection Agency 40CFR 136

Glaser et al[5] defined the MDL as "the minimum concentration of a substance that can be identified, measured and reported with 99% confidence that the analyte concentration is greater than zero", and that "on average, 99% of the trials measuring the analyte concentration of the MDL must be significantly different from zero". The first definition is in terms of true concentration and could therefore be interpreted as L_D. Similarly, if one interprets "significantly different from zero" as greater than L_C, which is reasonable, then the MDL as conceptually defined is also consistent with L_D. Unfortunately, their derivation converges asymptotically to an estimate of the 99th percentile of the distribution of measurements at concentration $x = 0$, which is, of course, L_C. These definitions have been adopted by USEPA[4] and the MDL is used in establishing all published DLs under USEPA regulations for environmental monitoring of drinking water, groundwater and surface water. In addition, all laboratories performing analyses under these regulations must be certified on the basis of meeting published MDLs using this method.

In constructing a model for the MDL, Glaser et al.[5] initially assume that variability is a linear function of concentration, (Equation 5) such that:

$$s_x = b_0 + b_1 x \tag{5}$$

where s_x is the standard deviation of n replicate analyses at concentration x and b_0 and b_1 are the intercept and the slope of the linear regression of standard deviation on concentration. They define the MDL (Equation 6) to be that concentration x for which:

$$t_x = \frac{x}{s_x/\sqrt{n}} \tag{6}$$

holds when t is set equal to the 99th percentile of Student's t-distribution with $n-1$ degrees of freedom. At this point Glaser et al. make two simplifying assumptions. First, they set b_1 to zero, therefore (Equation 7):

$$\text{MDL} = \frac{t_{[0.01,n-1]}b_0}{\sqrt{n}} \tag{7}$$

Second, they assume that $s_x = b_0/\sqrt{n}$, therefore (Equation 8):

$$\text{MDL} = t_{[0.01,n-1]}s_x \tag{8}$$

where s_x is defined as the standard deviation of n analytical replicates spiked at a single concentration x.

There are numerous problems with this model. First, having started with a conceptual definition that sounds promisingly like L_D they derive a mathematical definition that corresponds to L_C. This highlights the major distinction between the MDL and the other methods reviewed here. Second, having started with the useful generalization, variance allowed to be a function of concentration, they remove it by assuming $b_1 = 0$ and using a single concentration to estimate s_x. To estimate b_0 and b_1 would require multiple samples at each of several concentrations (i.e. a calibration design). Third, there are inexplicable inconsistencies about whether $s_x = b_0$ (per Equation 6 after you set $b_1 = 0$) or $b_0/\sqrt{n}$ as asserted in the sidebar in Glaser et al.[5] Of course, $b_0/\sqrt{n}$ is the standard error of measurement at $x = 0$, not the standard deviation at concentration x.

It seems odd that the initial assumption of this model is that variability is dependent on concentration, yet in practice the observed standard deviation for seven replicates is used regardless of the concentration at which the sample is spiked. In practice, laboratories demonstrate they can achieve low state and federally mandated MDLs (used as decision limits, not DLs) by spiking at progressively lower concentrations. Glaser et al. suggest that the spiking concentration should be close to the "true MDL"; however, since the true MDL is unknown, the spiking concentration is often set at risk-based standards (e.g. a maximum contaminant level – MCL) and can be well below the true MDL. Since Glaser et al. assume that the s_x is a linear function of concentration, this practice can lead to a gross underestimate of the true MDL.

It should be noted that despite the obvious limitations of the statistical derivation of the MDL, the underlying conceptual framework foreshadows some of the ideas of calibration designs. Conceptually, the underlying model assumes that the measured value is normally distributed with mean equal to the true concentration x and standard deviation $b_0 + b_1x$. This conceptual model leads directly to many of the statistical results for calibration designs but is intractable for single concentration designs since b_0 and b_1 are unknown and cannot be estimated from a single concentration. Unfortunately, USEPA continues to use the MDL as the basis for regulatory detection decisions. The MDL also forms the basis for the new approach USEPA has taken to quantification, the minimum level (ML) which is computed as 3.18 MDL.[17]

2.3 Prediction Limits

Prediction limits were apparently not considered in the context of single concentration designs. Nonetheless they are more relevant to making detection decisions than either the MDL or the Kaiser–Currie method. A prediction limit provides a certain level of confidence (e.g. 99%) of including the next single measurement based on a previous sample of n measurements. If we consider a single future detection decision, then the decision limit (Equation 9) based on a prediction limit[18] for a normally distributed instrument response is

$$L_C = t_{(n-1,\alpha)}s_0\sqrt{1 + \frac{1}{n}} \tag{9}$$

where s_0 is an estimate of σ_0 (i.e. the standard deviation in blank samples). For the seven replicate samples required by USEPA, the decision limit (Equation 10) becomes

$$L_C = 3.14s_0\sqrt{1 + \frac{1}{7}} = 3.36s \tag{10}$$

where $\alpha = 0.01$. Assuming $s_0 = s_D = s$ and $\alpha = \beta$ the DL (Equation 11) can be approximated by

$$L_D = 2t_{(n-1,\alpha)}s\sqrt{1 + \frac{1}{n}} = 2L_C \tag{11}$$

however, this is an invalid use of the central t distribution since the statistic $\sqrt{n}x/s$ has a noncentral t distribution when $x > 0$ (i.e. in computing L_D, $x = L_C$).

2.4 Tolerance Limits

In practice, a limit of detection is computed and applied to a large and potentially unknown number of future detection determinations. As such, the prediction limit described in the previous section may not be relevant to many situations because the error rate pertains to a single future detection decision. An alternative is to base the decision limit on a statistical tolerance limit that will contain a proportion of all future measurements (e.g. 99%) with a specified level of confidence (e.g. 99%). Tolerance limits are well known in the statistical literature and have been described in detail by Guttman[19] with corresponding tables relevant to this application. In this context (Equation 12), the decision limit becomes

$$L_C = K_{(P,\alpha)}s_0 \tag{12}$$

where values of K for varying coverage proportions (P) and confidence levels $(1 - \alpha)$ are available in tabular form.[11,18,19]

Following USEPA guidance using seven replicate samples, the decision limit would be $6.41s_0$, and would provide 99% confidence of including 99% of all future

measurements when the true concentration is zero. Of course, as the number of background samples n increases, the size of the multiplier decreases. Use of tolerance limits to compute the DL is not straightforward; however, an approximate method has been developed.[10]

2.5 Simultaneous Control of False Positive and False Negative Rates

The distinction between decision limits and DLs is based on simultaneous control of both false positive and false negative rates. When the population parameters σ_0 and σ_D are known, as was assumed by Currie, repeated application of the normal multiplier $z_{1-\alpha}$ followed by $z_{1-\beta}$ is appropriate. In practice however, σ_0 and σ_D are unknown and are replaced by their sample-based estimates s_0 and s_D provided by a sample of n spiked measurements. In this case repeated application of Student's t distribution (i.e. $t_{1-\alpha}$ and $t_{1-\beta}$) is incorrect. In computing L_C, the true concentration $x = 0$ and the false positive rate of the decision limit is based on Student's t distribution. In computing L_D the true concentration $x = L_C$ and the statistic $\sqrt{n}x/s_0$ has a noncentral t distribution on $n - 1$ degrees of freedom.[1,10] Relevant tabled values of the noncentrality parameters $\phi(n-1, \alpha, \beta)$ are also available.[1,10] For the case of $n = 7$ measurements and false positive and false negative rates of 1%, the limit of detection is $\phi(6, 0.01, 0.01)s_0 = 6.64s_0$. Although this limit of detection considers both false positive and false negative rates it pertains only to a single future detection decision. Gibbons et al.[10] have further generalized the result to include coverage of a specified proportion of all future values and have provided a corresponding table of multipliers (also see Gibbons[11]).

2.6 Limitations

The critical assumption underlying single concentration designs is that variability is homogeneous in the range of possible spiking concentrations and the true limit of detection. This assumption is implicit in the use of observed spiking concentration variance as an estimator of the true variance at the limit of detection. This assumption is rarely realized in practice. As a typical example, consider the calibration function presented in Figure 1.

For the purpose of illustration, actual concentrations are presented on the x-axis and measured concentrations are presented on the y-axis. In practice, the y-axis should represent actual peak area ion counts prior to calibration. Ignoring this step will lead to ignoring uncertainty in the calibration function, never known with certainty.

Figure 1 displays calibration data for silver using USEPA Method 1638[20] in distilled water. These data were obtained under controlled nonblind conditions that are not representative of routine practice. Spiking concentrations were 0, 10, 20, 50, 100, 200, 500, 1000 ppt. The center line that passes through these data represents the weighted least squares (WLS) estimate of the linear calibration regression line. The curvilinear functions that provide an interval surrounding the data are WLS prediction bands and provide 99% confidence of including the next single measurement at a given concentration.

Inspection of Figure 1 reveals that variability increases with increasing concentration. In fact, this same pattern

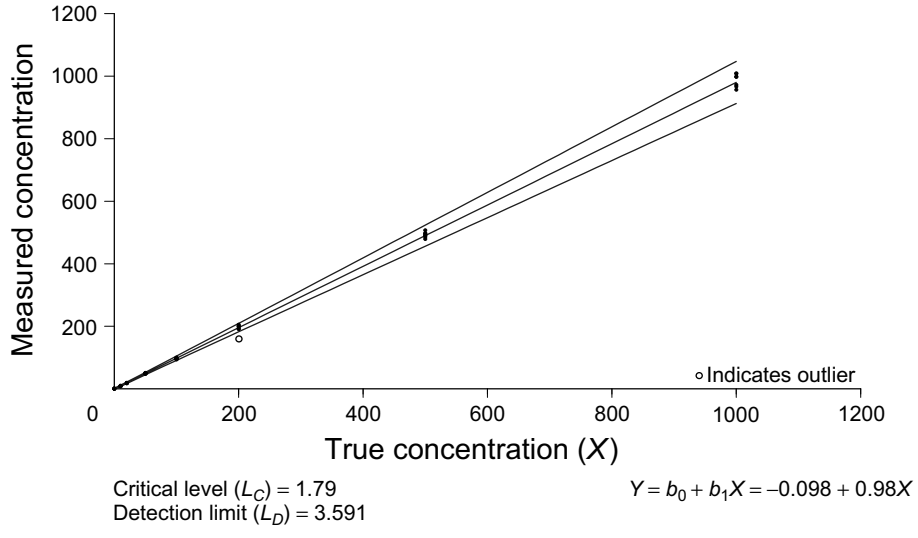

Figure 1 Linear calibration assuming increasing variance based on Rocke and Lorenzato Model: Method 1638 Silver (ppt) – prediction interval.

was observed for each of 8 other compounds. The prediction bands, which are based on the assumption that variability increases with concentration, fit the observed measurements extremely well. Of the 56 measurements (i.e. 7 replicates of each of the 8 spiking concentrations) one measurement is beyond the prediction bands, and therefore consistent with the 99% confidence level. The implication of this finding is that the driving force behind limits of detection from single concentration designs is the concentration at which the samples are spiked. To illustrate this point, Table 1 displays the USEPA/Glaser et al. MDL estimator applied to the data in Figure 1 at various concentrations.

Table 1 reveals that USEPA's MDL estimator will yield values ranging from 67.29 ppt to 1.47 ppt from exactly the same data. If a new analytical method with lower limits of detection were developed, presumably it would be tested by spiking at a correspondingly lower concentration. Thus, regardless of the true analytical properties of the method a lower MDL will be found.

Anticipating this problem to some extent, Glaser et al. suggest both a single-step and an iterative approach to computing the MDL (also described in USEPA's adoption of the procedure). For the single-step approach they recommend spiking samples at a concentration between one and five times the best estimate of the true MDL. They note that it is possible that an incorrect MDL could be obtained even if the spiking concentration were less than five times the computed MDL, illustrated in the higher values in Table 1. This logic requires that we know the actual MDL, which never happens in practice. In this case they recommend an iterative approach in which successively lower concentrations are computed until the pooled variance stabilizes. They note that convergence of the iterative procedure depends on the closeness of the estimated MDL to the calculated MDL. They also note that if the calculated MDL is lower than the background level of the analyte, the iterative procedure cannot be used.

Table 1 USEPA 40CFR MDL at various spiking concentrations of silver in distilled water using method 1638 in ppt

Concentration	SD	MDL
0	0.47	1.47
10	0.67	2.09[a]
20	0.66	2.06
50	1.08	3.40
100	2.23	7.00
200	15.32	48.10[a]
500	9.05	28.43
1000	21.43	67.29

[a] Meets USEPA's 5 : 1 criterion.

Inspection of Table 1 reveals that spiking concentrations of 200 ppt and 10 ppt both meet USEPA's criterion of a spike to MDL ratio of 5 : 1 or less. Presumably, if we had started at higher concentrations and iteratively decreased the spiking concentration we would obtain an MDL of 48 ppt. Alternatively if we started at zero and worked up, we would have obtained an MDL of 2 ppt. This level of arbitrariness is unacceptable in a DL estimator and can lead to both false positive and false negative regulatory impact decisions based on samples spiked at higher or lower concentrations. For this reason, single concentration decision limits and DLs are of little practical value despite their widespread use.

3 DETECTION LIMITS FOR CALIBRATION DESIGNS

An alternative method for estimating DLs is a calibration design. In this case, a series of samples are spiked at known concentrations in the range of the hypothesized DL, and variability is determined by examining the deviations of the actual response signals from the fitted regression line of response signal on known concentration. In these designs, it is generally assumed that the deviations from the fitted regression line are normally distributed.

There are two major advantages of calibration designs over single concentration designs. First, single concentration designs ignore uncertainty in the calibration function that relates measured to true concentration. Second, when variability in measured concentration is a function of true concentration, the estimated DL, based on a single concentration design, will depend in large part on the spiking concentration selected by the analyst. If different laboratories select different spiking concentrations, it will appear that they have different analytical capabilities when they may, in fact, be identical. This is not a problem for calibration-based DLs since the relationship between variability and concentration can be explicitly modeled and incorporated into the DL estimator.

Proponents of single concentration designs suggest that nonconstant variance can only lead to overestimates of the true DL since variability at a concentration greater than zero will be larger than the variability when the analyte is absent. In many cases this is true, and artificially high DLs can result. However, if a very low spiking concentration is selected, gross underestimation of the DL is also possible. Modern analytical instruments are often designed to censor measured concentrations less than zero. If the spiking concentration is close to zero, then we would expect up to 50% of the measured concentrations to be negative. However, if the instrument assigns a concentration of zero to these samples, the estimated

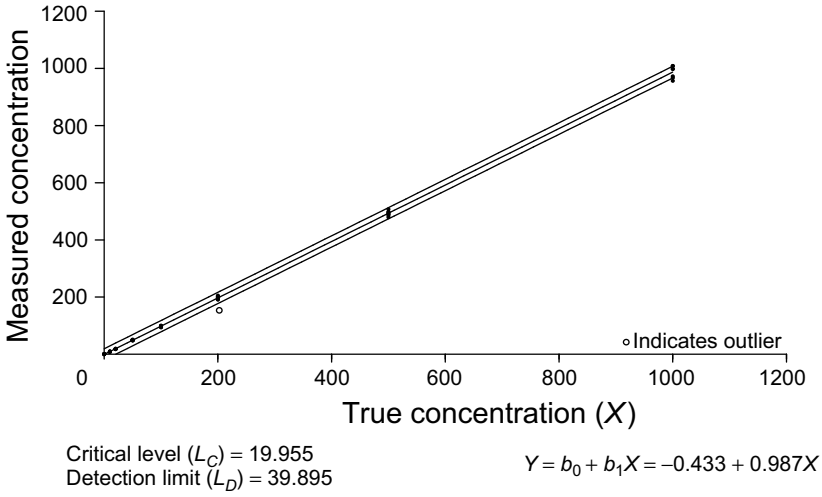

Figure 2 Linear calibration assuming constant variance: Method 1638 Silver (ppt). Prediction interval with Hubaux–Vos type estimates of L_C and L_D graphically depicted.

standard deviation will be a gross underestimate of the true population standard deviation. To make matters worse, many instruments use peak area rejection criteria so that even positive concentrations of low magnitude will also be assigned a measured concentration of zero. Again, this practice will lead to even further underestimates of variability and corresponding DLs. Calibration designs are immune to such problems since concentrations both above and below the true DL are included and nonconstant variance accommodated. The primary disadvantage of calibration based DL estimators is that they are more complicated to compute and typically will require a computer program capable of unweighted least squares (ULS) regression and/or WLS regression. Fortunately computer programs for these computations are now becoming readily available.[21,22]

Hubaux and Vos[8] were the first to describe a procedure for calibration based DL estimation. Beginning from a calibration design in which response signals are determined for analyte-containing samples with concentrations throughout the range of L_C to L_D, they constructed a 99% prediction interval for the calibration regression line (see Figure 2). The prediction interval is given by Equation (13):

$$\hat{y}_x \pm t_{[1-\alpha/2,n-2]} s(\hat{y}_x) \tag{13}$$

where $s(\hat{y}_x)$ is the estimated uncertainty of a predicted instrument response (or measured concentration) $\hat{y} = b_0 + b_1 x$ at true concentration x (see Appendix). The DL is defined as the point at which we can have 99% confidence that the response signal is greater than L_C; therefore, Hubaux and Vos suggest that the response signal be obtained graphically by locating the abscissa corresponding to L_C on the lower prediction limit (see Figure 2). A more direct solution for L_D is provided in the technical Appendix.

The method of Hubaux and Vos assumes that variability is constant throughout the range of concentrations used in the calibration design. If this assumption is violated, then the DL will be overestimated because variability at high concentrations are given equal weight as those at lower concentrations. Clayton[1] suggested a variance-stabilizing square-root transformation which helps to some extent but does not eliminate the problem. Oppenheimer[23] proposed an estimator of L_D based on WLS regression which provides a general solution to this problem but requires an iterative solution (see Appendix 1). Gibbons et al.[10] provide a noniterative computing approximate and further generalize this result to the case of multiple future detection decisions by substituting tolerance limits for prediction limits. See Gibbons[3,11] for a review of this literature.

3.1 Illustration

Figures 1–5 display actual and measured concentrations of silver in seven eight-point concentration calibrations. Figure 2 represents the same data as in Figure 1, but assumes constant variance and provides corresponding Hubaux–Vos estimates of L_C and L_D. Figure 2 reveals that if we assume constant variance, in order to incorporate larger variability at higher concentrations, variability is overestimated at the lower concentrations. In contrast, the WLS approach described in Figure 1 provides excellent fit to the observed data for both high and low true concentrations. Figure 3 represents the same data as Figure 1 but shows error bands based on a statistical tolerance interval with 95% confidence and 99% coverage. Figure 3 reveals that substitution of tolerance

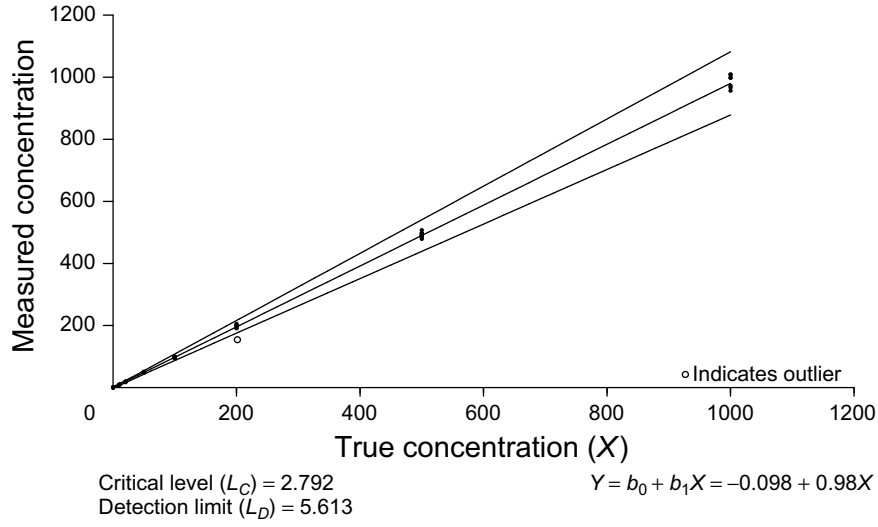

Figure 3 Linear calibration assuming increasing variance based on Rocke and Lorenzato Model: Method 1638 Silver (ppt) – tolerance interval.

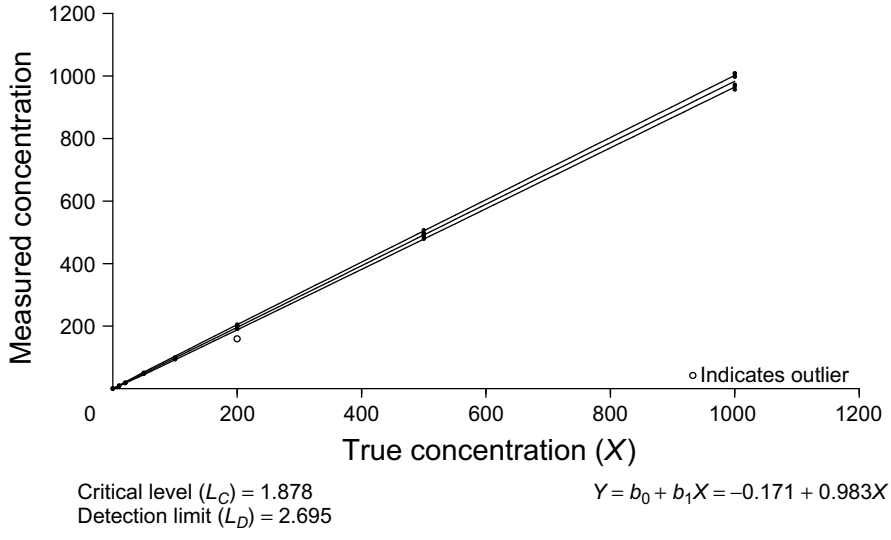

Figure 4 Linear calibration assuming proportional variance: Method 1638 Silver (ppt) – prediction interval. $L_D = 0.545$.

intervals for prediction intervals provides slightly wider intervals but yields DLs that will include 99% of all future detection decisions whereas prediction limits only provide this level of confidence for a single future detection decision. Figure 4 displays the WLS approximate prediction limits described by Gibbons et al. based on the assumption of proportional variance and concentration. Figure 4 reveals quite similar results to Figure 1 where variance is explicitly modeled. Figure 5 displays the WLS approximate tolerance intervals based on the assumption of proportional variance and concentration. Again, the approximation appears to provide similar results to the case in which variance is explicitly modeled as in Figure 3.

Table 2 presents estimated DLs for several of the previously described approaches.

Inspection of Table 2 reveals that, as expected, the Hubaux–Vos ULS approach assuming constant variance yields a much higher DL than the WLS approaches which assume that variability is an increasing function of true concentration. Once nonconstant variability is accounted for, we pay a very modest price for generalizing the limits to apply to 99% of all future detection decisions as opposed to the next single detection decision (e.g. 5.6 ppt versus 3.6 ppt). The approximate noniterative solution suggested by Gibbons et al. works well in this example. Note that the calibration-based estimates of L_C and L_D

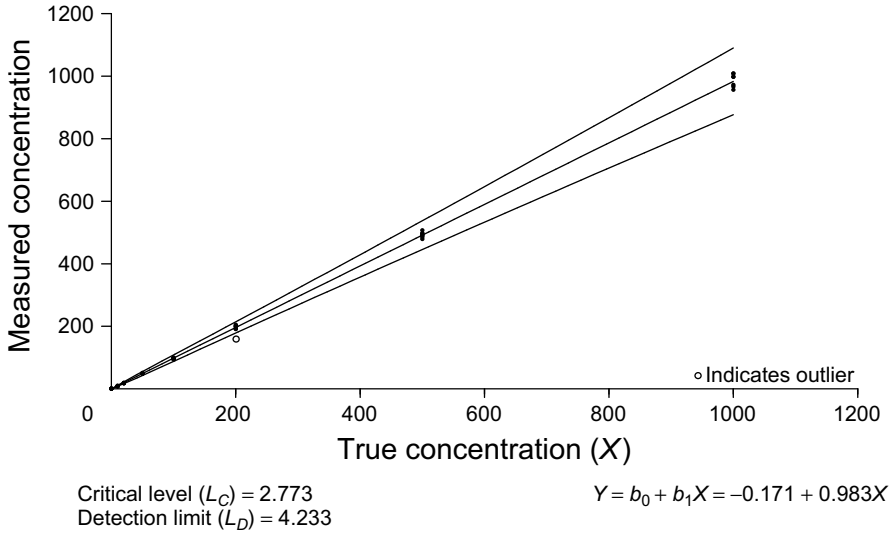

Figure 5 Linear calibration assuming proportional variance: Method 1638 Silver (ppt) – tolerance interval.

Table 2 Comparison of calibration-based DL estimators for constant (ULS) and nonconstant (WLS) variance silver via method 1638 in ppt

Model	L_C	L_D	Figure
Hubaux–Vos (ULS)	19.955	39.895	2
WLS PL[a] – $WT = x$	1.878	2.695	4
WLS PL[b] – $WT = 1/s_x^2$	1.790	3.591	1
WLS TL[a] – $WT = x$	2.773	4.233	5
WLS TL[b] – $WT = 1/s_x^2$	2.792	5.613	3

[a] Gibbons[22] noniterative solution.
[b] Oppenheimer[23] iterative solution.

appear to be quite reasonable given the data displayed in Table 1. Note that these data were obtained under highly controlled ''clean lab'' research conditions in distilled water and the DLs obtained here may be considerably higher in routine practice. Nevertheless, these results clearly reveal that calibration-based methods do not necessarily lead to higher DL estimates than their single concentration counterparts (e.g. the MDL).

4 EXPERIMENTAL DESIGN OF DETECTION LIMIT STUDIES

A detailed review of the principles of experimental design of DL studies would easily require a chapter to itself and has been reviewed in some detail by others.[2,24] There are several guiding concepts critical for producing unbiased DL estimates of practical relevance.

First, in analyte present studies, the analysts must be blind to both the number of compounds in the sample and to their spiking concentrations. To achieve this goal, the number of compounds must vary (perhaps randomly) from sample to sample. Furthermore, the concentration of each constituent should vary both within and across samples. Without ensuring that the analyst is blind to both presence and concentration of the analyte under study, the resulting DL simply cannot be applied to routine practice in which such uncertainty must always exist. In practice, it is often impossible to execute such studies since numerous samples would have to be prepared at widely varying concentrations. In the absence of this level of experimental control, standard calibration data in which the analysts are unaware that they are being tested may have to suffice. The critical issue is that the analysts must *not* go back and retest samples that appear to be anomalous relative to the known spiking concentration.

Second, two or more instruments and analysts must be used and the assignment of samples to analysts and instruments must also be random. Since in large production laboratories, any one of a number of analysts and/or instruments may be called upon to analyze a test sample, this same component of variability must be included in determining the DL.

Third, if multiple laboratories are used or a regulatory agency analyzes split-samples or additional samples from the facility, then the entire DL study must be replicated in multiple laboratories. Data from a single laboratory should only be used when it is technically unfeasible to provide common calibration standards, or to split common standard samples, or a dedicated laboratory is used for all relevant analyses.

Fourth, the number of samples selected should be based on statistical power criteria, such that a reasonable

balance of false positive and false negative rates is achieved. For example, if we estimate σ by computing s on seven samples, our uncertainty in σ will be extremely large and our resulting DL estimate L_D will also be quite large. By increasing the number of samples to say, 25, we achieve a much more reasonable estimate of σ, and resulting L_D are greatly reduced. The cost of running a few additional samples is far cheaper than dealing with the drawbacks of using DLs incapable of detecting anything but the largest signals.

An additional note regarding analyte-absent experiments (i.e. blank samples): rather than running a series of blank samples at once, they should be randomly entered into the analysts' work load throughout the course of the day. The purpose of this approach is to ensure that the analysts are blind to sample composition. The broader question is whether analyte-absent experiments are relevant to establishing DLs. It can be argued that the properties of the *method* can only be evaluated when the analyte is present, at least in some of the samples. The general recommendation of calibration designs over fixed-concentration designs allows for a mixture of samples in which the analyte is present and absent.

There are several experimental designs that can fulfill the preceding requirements. When the number of samples is large, say $n = 30$, one possibility is to give each compound in the study a 0.5 probability of being in any given sample, and once selected, its concentration could also be randomly selected from a uniform distribution covering the range of 0 to $2L'_D$, where L'_D is the presumed DL. This design may be optimal to ensure blindness, but it may not be necessary to maximize the signal-to-noise ratio which increases the amount of information gained in such a study. For example, Hubaux and Vos[8] suggest a three values repartition, in which n_1 replicate samples with concentrations at the lowest permissible content (x_1), $n_2 = n - n_1 - 1$ samples at the highest "permissible content" (x_2) and a single sample at $(x_1 + x_2)/2$ are selected. They find that this design minimizes the number of required standards for a fixed level of sensitivity.

Liteanu and Rica,[24] review a wide variety of sampling designs for DL studies, including response surface designs, fractional factorial designs and rotatable designs. Youden pair-type designs are excellent candidates for maintaining blind and unbiased DL studies.

Finally, in any design in which there are multiple components of variability (e.g. analyst, instrument, laboratory), appropriate steps must be taken to obtain unbiased estimates of variability. The naive estimator of σ which is obtained by computing s in the usual way with all measurements should be replaced by the unbiased variance estimate based on the appropriate variance components model for the problem at hand.[25,26]

5 QUANTIFICATION LIMITS

Quantitative determination in analytical chemistry is a two-stage process. First, we must make the binary decision of whether or not the compound is present in the sample. Second, in the event the compound is present, we must determine if the estimated concentration supports quantitative determination. The first determination is governed by the DL as discussed in the previous sections. In contrast, very little literature exists on the second stage of the decision process, i.e. quantification. Currie[6] originally described the "determination limit" (L_Q) as the concentration at which the relative standard deviation (RSD) is 10% (i.e. a signal-to-noise ratio of 10 to 1). Gibbons et al.[27] have provided a more statistically rigorous estimator of L_Q that accommodates effects of nonconstant variance, uncertainty in the parameters of the recovery curve and uncertainty in the sample-based estimator of the true population variance. More recently, USEPA[17] has defined the ML as the method DL (defined in 40CFR Part 136 App. B[4]) multiplied by a factor of 3.18. USEPA is using the ML as a quantification limit above which measured concentrations can be directly compared to regulatory standards and compliance decisions can be made. In an attempt to provide an estimator that is conceptually similar to the ML, but with fewer statistical limitations, Gibbons et al.[28] have developed an "Alternative Minimum Level" (AML). In the following, strengths and weaknesses of these various alternatives are explored.

5.1 The United States Environmental Protection Agency Minimum Level

USEPA[17] has defined the ML (Equation 14) as:

$$\mathrm{ML} = 3.18\ (\mathrm{MDL}) = 3.18(3.14)s_x = 10s_x \qquad (14)$$

where s_x is the standard deviation of 7 replicate measurements spiked at true concentration x. All limitations previously described for the USEPA MDL apply equally to the ML since it forms the basis for this estimator, i.e. the MDL and hence the ML is a direct function of the choice of spiking concentration x. In addition, it is of critical importance to understand that the ML is not based on the permittee's MDL, but rather on a single MDL developed by a single USEPA laboratory under ideal conditions. This further reduces generalizability of the ML to application under routine conditions in which analysts are blind to the true concentration and any one sample can be analyzed by different analysts, instruments and even laboratories.

5.2 Alternative Minimum Level

The fundamental problem with the ML is not its definition (i.e. $10s_x$) but the fact that the sample standard deviation s_x is a moving target and depends completely on the choice of spiking concentration x. As previously noted, the ML suffers from numerous other problems in that it completely ignores interinstrument, interanalyst and interlaboratory components of variation, but this has more to do with the way in which USEPA has applied the method and not the statistical derivation itself. It is the statistical derivation that we focus on in developing the AML.

As a conceptual foundation, Currie defined the determination limit L_Q (Equation 15) as

$$L_Q = 10\sigma_Q \tag{15}$$

where σ_Q represents the true population standard deviation at the determination limit L_Q. The L_Q is based on the idea that quantitative determinations can be confidently made when the signal-to-noise ratio is ten to one (i.e. a RSD of 10%).[6] Unfortunately, in the real world, we never know the population standard deviation (i.e. it must be estimated from a sample of n measurements) and in many cases, the proportionality constant between the standard deviation and concentration is larger than 10% in the practical calibration range and therefore a 10% RSD may not be achievable or only achievable at unacceptably high levels.

A reasonable alternative to both approaches is to compute the standard deviation at the lowest concentration that is differentiable from zero and use that standard deviation in computing the ML. In this way, the spiking concentration is no longer arbitrary and the effect of nonconstant variance is incorporated up to the lowest nonzero concentration. Of course, this alternative ML (AML) will not guarantee a 10%RSD at the AML; however, we can compute the actual %RSD at the ML by modeling the association between s and x. The algorithm is as follows:

1. Using calibration data, model the relationship between s and x, for example an exponential function (Equation 16)

$$s_x = a_0 e^{a_1(x)} \tag{16}$$

   The exponential model $s_x = a_0 e^{a_1 x}$ generally provides excellent fit to the observed data and it will be used in the following discussion. However, the model suggested by Rocke and Lorenzato[29] (see Appendix) is also an excellent choice. The Rocke and Lorenzato model was used in generating the prediction intervals in Figure 1. The exponential model can be fitted using nonlinear least squares (e.g. Gauss–Newton) or by ULS regression of natural log transformed measured concentration on actual concentration (Snedecor and Cochran, p. 397[30]).

2. Compute the WLS regression of measured concentration or instrument response (y) on true concentration (x) for the linear model (Equation 17)

$$\hat{y} = b_{0w} + b_{1w} x_i \tag{17}$$

   as described in Appendix 1.

3. Iteratively compute the L_C using the method originally proposed by Scott et al.[12] which is the upper 95% confidence 99% coverage tolerance limit for measured concentrations when the true concentration is zero, (Equation 18):

$$\begin{aligned} y_C &= K_{0.95,0.99} s_{L_C} + b_{0w} \\ &= K_{0.95,0.99} a_0 e^{(a_1(y_c - b_{0w})/b_{1w})} + b_{0w} \end{aligned} \tag{18}$$

   where $K_{0.95,0.99}$ is the 95% confidence 99% coverage one-sided normal tolerance limit factor for n observations, where, in this case, n is the total number of measurements (see Table 4.2 in Gibbons[11]). L_C is the corresponding true concentration (Equation 19):

$$L_C = \frac{y_C - b_{0w}}{b_{1w}} \tag{19}$$

   Note that although it is not required for computing the AML, a Currie-type estimate of the DL (L_D) can also be obtained by iteratively computing Equation (20):

$$\begin{aligned} y_D &= y_C + K_{0.95,0.99} s_{L_D} \\ &= y_C + K_{0.95,0.99} a_0 e^{(a_1(y_D - b_{0w})/b_{1w})} \end{aligned} \tag{20}$$

   and L_D is the corresponding true concentration (Equation 21):

$$L_D = \frac{y_D - b_{0w}}{b_{1w}} \tag{21}$$

4. Compute the standard deviation at the L_C from Step 1 (i.e. s_{L_C}), for example (Equation 22):

$$s_{L_C} = a_0 e^{a_1(L_C)} \tag{22}$$

5. Compute the measured concentration at ten times the standard deviation at the L_C (Equation 23):

$$y_Q = 10 s_{L_C} + b_{0w} \tag{23}$$

6. The AML is then computed as Equation (24):

$$x_Q + \left(\frac{t}{b_{1w}} \quad \sqrt{V(y_Q)} \right. \tag{24}$$

where $x_Q = (y_Q - b_{0w})/b_{1w}$. The value t is the upper 99th percentile of Student's t distribution on $n - 2 - p$ degrees of freedom, where p is the number of unknown parameters in the standard deviation model. Alternatively when n is large (e.g. $n > 25$) the AML can be approximated by Equation (25):

$$\text{AML} \sim \left(\frac{x_Q + ts_{y_Q}}{b_{1w}} \right. \tag{25}$$

5.3 Illustration

To illustrate the computation of the AML, return to the data in Figure 1. Inspection of Figure 6 reveals that the Rocke and Lorenzato model fits the observed data quite well. The various steps in the computation are displayed in Table 3. Inspection of Table 3 reveals that the estimated AML is 7.276 ppt and this provides a RSD of 7.78%. Note that a wide range of MLs are possible for these data. For example, spiking at 200 ppt the ML is 153 ppt whereas spiking at 10 ppt yields 6.66 ppt. Both spiking concentrations meet USEPA's 5 : 1 criterion and would therefore represent valid estimates of the ML.

5.4 Alternative Formulations

An alternative that departs somewhat from the AML is to estimate the true concentration that leads to a measured RSD of 10% or 20% or whatever is achievable. Requiring 10% will not always lead to a solution since the rate of change in standard deviation relative to concentration may be greater than 0.1. The solution for the case of nonconstant variance Equation (26) involves iterative solution of

$$y_Q = \frac{1}{r}\sqrt{V(\hat{y}_Q)} \tag{26}$$

where r represents the RSD (e.g. $r = 0.1$ for a 10% RSD). The estimate, referred to as L_Q based on Currie's "determination limit", is then Equation (27):

$$L_Q = \frac{y_Q - b_{0w}}{b_{1w}} \tag{27}$$

Note that here, the RSD is of the predicted measured concentration relative to the standard deviation of the predicted measured concentration, which includes uncertainty in the calibration function. In contrast, for the previous estimators, including the AML, the RSD is computed post hoc in terms of true concentration and as such, will generally be somewhat smaller. When applied to the silver data application of Equations (28) and (29) yields $L_Q = 7.836$ for an RSD of 10% which is quite similar to the AML.

5.5 The Estimator According to Gibbons et al.

Gibbons et al.[27] have provided a direct estimator of L_Q under the assumption that square root transformation of both x and y homogenizes variance.[1] This leads (Equation 28) to the predicted concentration x^*_α for which

$$\hat{y} = \frac{100}{\alpha} s(\hat{y}) \tag{28}$$

the predicted response ($\hat{y}$) that is $100/\alpha$ times its estimated standard deviation ($s(\hat{y})$). Let y^*_α denote the

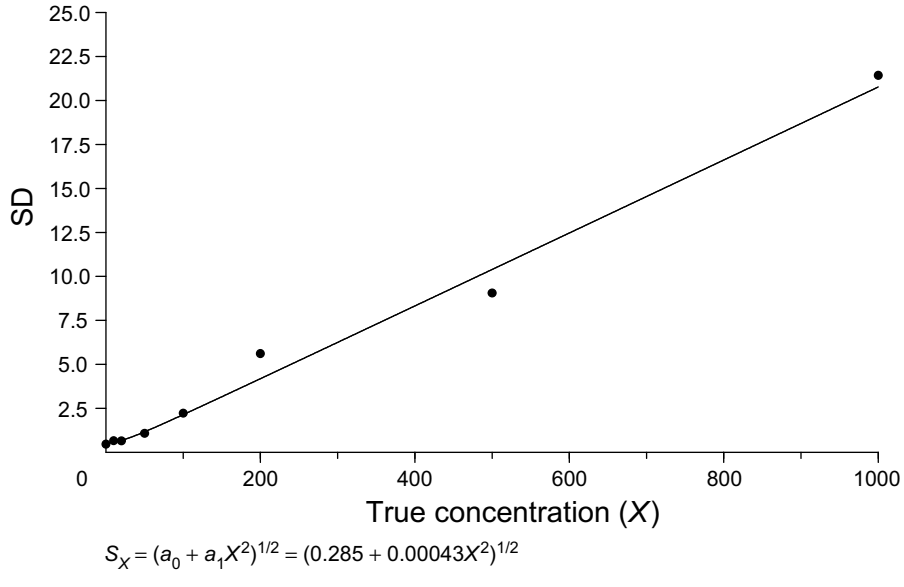

Figure 6 Relationship between concentration and variability – Rocke and Lorenzato Model.

Table 3 AML model – Silver using Method 1638

Step	Equation	Description
1	$S_X = (a_0 + a_1X^2)^{\frac{1}{2}} = (0.285 + 0.00043X^2)^{\frac{1}{2}}$	R−L regression model for standard deviation versus concentration (X); see Figure 6
2	$L_C = KS_X + b_0 = 2.833\ \ S_X - 0.098 = 1.546$	Result of iterative solution for Currie (1968) type critical level or EPRI CMDL
3	$S_{L_C} = (a_0 + a_1L_C^2)^{\frac{1}{2}} = (0.285 + 0.00043 \times 2.389)^{\frac{1}{2}} = 0.535$	Estimated SD at the critical level L_C
4	$L_D = L_C + KS_X = 1.546 + 2.833S_X = 3.1$	Result of iterative solution for Currie (1968) type DL
5	$S_{L_D} = (a_0 + a_1L_D^2)^{\frac{1}{2}} = (0.285 + 0.00043 \times 9.609)^{\frac{1}{2}} = 0.538$	Estimated SD at the DL L_D
6	$Y_Q = 10S_{L_C} + b_0 = 10 \times 0.535 - 0.98 = 5.25$	Averaged measured concentration (or instrument response) at 10 times S_{L_C}
7	$X_Q = (Y_Q - b_0)/b_1 = (5.25 + 0.098)/0.98 = 5.456$	True concentration corresponding to Y_Q, where b_0 and b_1 are the intercept and slope of the recovery equation; see Figure 1
8	$S_{X_Q} = (a_0 + a_1X_Q^2)^{\frac{1}{2}} = (0.285 + 0.00043 \times 29.768)^{\frac{1}{2}} = 0.546$	estimated SD at X_Q
9	$\text{AML} = X_Q + [tS_W/b_1] \times [S_{X_Q}^2 + 1 \quad [1/k] + (X_Q - \overline{X})^2 \quad [(X - \overline{X})^2/K]^{\frac{1}{2}} = 5.456 + [2.402 \times 1.328/0.98] \times [0.298 + 1/71.562 + (5.456 - 14.258)^2/66812.09]^{\frac{1}{2}} = 7.276$	The AML
10	$S_{\text{AML}} = (a_0 + a_1\text{AML}^2)^{\frac{1}{2}} = (0.285 + 0.00043 \times 52.946)^{\frac{1}{2}} = 0.555$	Estimated SD AML
11	$\%\text{RSD} = 100(S_{\text{AML}}/\text{AML})/b_1 = 100(0.555/7.276)/0.98 = 7.78\%$	The RSD at the AML

value of $\hat{y}$ that satisfies Equation (28). Gibbons et al.[27] solved Equation (28) for y_α^* and showed that a good approximation can be obtained as (Equation 29)

$$x_\alpha^* = \bar{x} + \frac{(100/\alpha)s_{y\cdot x} - \bar{y}}{b} \tag{29}$$

$$L_Q = (x_\alpha^*)^2 + 0.632456(x_\alpha^*) \tag{30}$$

also see Gibbons.[11] Gibbons et al. make several observations regarding this estimator. First, they suggest that with small numbers of measurements, $s_{y\cdot x}$ may be a poor estimator of the true population standard deviation $\sigma_{y\cdot x}$ and we may wish to replace $s_{y\cdot x}$ with its upper confidence limit.

Second, the previous definition of $\alpha/100$ times the standard deviation applies to the metric in which the L_Q is estimated. In the present context, a square-root transformation is used to stabilize variance across the calibration line while preserving linearity of the function. An RSD of 10% in this transformed response metric actually corresponds to a 20% RSD in the raw concentration metric. Therefore, obtaining a 10% RSD in the metric of measured concentrations would require a 5% RSD in the transformed response scale.

Application of this estimator to the silver data yields an L_Q of 22.159 $\mu g\,L^{-1}$ for a 10% RSD in the transformed metric. This value is somewhat higher than the previous estimators indicating that in this example, the square-root transformation failed to bring about constant variance at the lowest concentrations.

6 DISCUSSION

Perhaps the most significant contribution of this work is to demonstrate that the most critical source of bias in

estimating limits of detection and quantification is the effect of nonconstant variance. Results presented here reveal that a myriad of DL estimates can be obtained from exactly the same data and analytic method solely as a function of the concentration at which samples are spiked. The only solution in this case is to base limits of detection on multiple-concentration calibration data and either directly model association between variability and concentration or use a variance-stabilizing transformation.

In the calibration setting, four general DL estimators are described and the four methods appear to give reasonably similar results, at least in the example considered here. This finding indicates that we pay a very small price for considerable generality by adopting a tolerance interval-based approach that will cover 99% of all future detection decisions compared to only a single future detection decision. Similar results were found for calibration-based quantification limit estimators.

Despite the guidance provided by this developing literature on analytical detection decisions, USEPA continues to base regulations on the MDL or, for cases in which quantifiable measurements are required, a simple multiple of the MDL (i.e. the ML). As illustrated here, this practice can lead to bias, most notably an extreme dependence on the selection of spiking concentration. Despite some guidance on selection of spiking concentration, actual practice is guided by demonstrating compliance with published standards rather than using the true analytical capabilities of the laboratory. For example, MDLs are routinely determined in laboratories by spiking as many as 70 constituents at a fixed concentration in each of seven replicate samples. In this way it is impossible to iteratively adjust each constituent's concentration in the way suggested by USEPA. In contrast, if calibration-based methods were used, USEPA could designate the range of concentrations in advance to insure that the computed MDL is not adversely affected by a poor selection of initial spiking concentration. In future, USEPA should consider calibration-based approaches for determining limits of detection and quantification.

ABBREVIATIONS AND ACRONYMS

AML	Alternative Minimum Level
DL	Detection Limit
IDL	Instrument Detection Limit
MCL	Maximum Contaminant Level
MDL	Method Detection Limit
ML	Minimum Level
PMDL	Practical Method Detection Limit
RSD	Relative Standard Deviation
ULS	Unweighted Least Squares
USEPA	United States Environmental Protection Agency
WLS	Weighted Least Squares

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration

APPENDIX 1

Details of the WLS prediction interval approach described by Oppenheimer[23] are provided in the following. To obtain the Hubaux–Vos result set all weights $k_i = 1$. To obtain a DL based on tolerance intervals see equation 5.3.2 in Gibbons[11] and Zorn, Gibbons and Sonzogni.[31]

Compute the WLS regression[32] of measured concentration or instrument response (y) on true concentration (x) for the linear model

$$\hat{y}_{wi} = b_{0w} + b_{1w}x_i \tag{31}$$

where

$$b_{1w} = \frac{\sum_{i=1}^{n}[(x_i - \bar{x}_w)y_i/k_i]}{\sum_{i=1}^{n}[(x_i - \bar{x}_w)^2/k_i]} \tag{32}$$

$$b_{0w} = \bar{y}_w - b_{1w}\bar{x}_w \tag{33}$$

$$\bar{y}_w = \frac{\sum_{i=1}^{n}[y_i/k_i]}{\sum_{i=1}^{n}[1/k_i]} \tag{34}$$

$$\bar{x}_w = \frac{\sum_{i=1}^{n}[x_i/k_i]}{\sum_{i=1}^{n}[1/k_i]} \tag{35}$$

and the weight $k_i = s^2_{x_i}$ is the variance for sample i (Equations 31–35). The weighted residual variance is (Equation 36):

$$s^2_w = \frac{\sum_{i=1}^{n}[(y_i - \hat{y}_{wi})^2/k_i]}{n-2} \tag{36}$$

and the estimated variance for a predicted value $\hat{y}_{wj}$ is (Equation 37):

$$V(\hat{y}_{wj}) = s^2_w \quad k_j + \frac{1}{\sum_{i=1}^{n}(1/k_i)} + \frac{(x_j - \bar{x}_w)^2}{\sum_{i=1}^{n}[(x_i - \bar{x}_w)^2/k_i]} \tag{37}$$

where k_j is the estimated variance at concentration x_j. An upper $(1-\alpha)100\%$ confidence interval for $\hat{y}_{wj}$ (i.e. a prediction interval for a new measured concentration or instrument response at true concentration x_j) is (Equation 38):

$$\hat{y}_{wj} + t\sqrt{V(\hat{y}_{wj})} \tag{38}$$

where t is the upper $(1-\alpha)100$ percentage point of Student's t distribution on $n-2$ degrees of freedom[12,18]. The WLS estimate of L_C is therefore (Equation 39):

$$L_C = \frac{ts_w}{b_{1w}} \quad \overline{s^2_{L_C} + \frac{1}{\sum_{i=1}^{n}(1/k_i)} + \frac{(L_C - \bar{x}_w)^2}{\sum_{i=1}^{n}[(x_i - \bar{x}_w)^2/k_i]}} \tag{39}$$

and the WLS estimate of L_D is (Equation 40):

$$L_D = L_C + \frac{ts_w}{b_{1w}} \quad \overline{s^2_{L_D} + \frac{1}{\sum_{i=1}^{n}(1/k_i)} + \frac{(L_D - \bar{x}_w)^2}{\sum_{i=1}^{n}[(x_i - \bar{x}_w)^2/k_i]}} \tag{40}$$

Note that in order to compute L_C and L_D we must have estimates of $s^2_{L_C}$ and $s^2_{L_D}$ which are often unavailable and must be estimated using a model of standard deviation versus concentration, for example an exponential model[28] of the form (Equation 41):

$$s_x = a_0 e^{a_1 x} \tag{41}$$

or the model suggested by Rocke and Lorenzato[29] which reduces to (Equation 42):

$$s_x = \sqrt{a_0 + a_1 x^2} \tag{42}$$

both of which can be fitted by nonlinear least squares and require iterative solution. Gibbons et al.[10] suggest a noniterative approximation based on the assumption that variance is proportional to concentration. This assumption leads to the WLS estimates Equations (43–45):

$$b_{0w} = \frac{\sum_{i=1}^{n}(y_i/x_i)\sum_{i=1}^{n}x_i - n\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}(1/x_i)\sum_{i=1}^{n}x_i - n^2} \tag{43}$$

$$b_{1w} = \frac{\sum_{i=1}^{n}(1/x_i)\sum_{i=1}^{n}y_i - n\sum_{i=1}^{n}(y_i/x_i)}{\sum_{i=1}^{n}(1/x_i)\sum_{i=1}^{n}x_i - n^2} \tag{44}$$

and

$$s^2_w = \frac{\sum_{i=1}^{n}(y_i^2/x_i) - b_0\sum_{i=1}^{n}(y_i/x_i) - b_1\sum_{i=1}^{n}y_i}{n-2} \tag{45}$$

To compute L_C we can set the weight equal to the lowest nonzero concentration $x_{\min}$ (Equation 46):

$$L_C = \frac{ts_w}{b_{1w}} \quad \overline{x_{\min} + \frac{1}{n} + \frac{(x_{\min} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{46}$$

To compute L_D we can set the weight equal to L_C (Equation 47):

$$L_D = L_C + \frac{ts_w}{b_{1w}} \quad \overline{L_C + \frac{1}{n} + \frac{(L_C - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{47}$$

Given current computing environments, however, there is no reason why the iterative solution should not be performed i.e. simple repeated substitution beginning from $L_C = 0$ and $L_D = L_C$ until convergence (i.e. change of less than 10^{-4} in estimates of L_C and L_D on successive iterations).

REFERENCES

1. C.A. Clayton, J.W. Hines, P.D. Elkins, 'Detection Limits with Specified Assurance Probabilities', *Anal. Chem.*, **59**, 2506–2514 (1987).
2. C.B. Davis, 'Environmental Regulatory Statistics', in *Handbook of Statistics: Environmental Statistics*, eds. G.P. Patil, C.R. Rao, Elsevier, Amsterdam, Chap. 26, Vol. 12, 1994.
3. R.D. Gibbons, 'Some Conceptual and Statistical Issues in the Detection of Low Level Environmental Pollutants', *J. Environ. Ecological Stat.*, **2**, 125–167 (1995).
4. USEPA, 'Definition and Procedure for Determination of the Method Detection Limit – revision 1.11', Appendix B to Part 136, *Federal Register*, **49**(209), 43430, October 26 1984.
5. J.A. Glaser, D.L. Foerst, G.D. McKee, S.A. Quane, W.L. Budde, 'Trace Analyses for Wastewaters', *Environ. Sci. Tech.*, **15**, 1426–1435 (1981).
6. L.A. Currie, 'Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry', *Anal. Chem.*, **40**, 586–593 (1968).
7. D. Lambert, B. Peterson, I. Terpenning, 'Nondetects, Detection Limits and the Probability of Detection', *J. Am. Stat. Assoc.*, **86**, 266–277 (1991).
8. A. Hubaux, G. Vos, 'Decision and Detection Limits for Linear Calibration Curves', *Anal. Chem.*, **42**, 849–855 (1970).
9. R.D. Gibbons, F.H. Jarke, K.P. Stoub, 'Method Detection Limits', *Proc. Fifth Ann. USEPA Waste Testing and Quality Assurance Symp.*, **2**, 292–319 (1988).
10. R.D. Gibbons, F.H. Jarke, K.P. Stoub, 'Detection Limits for Linear Calibration Curves with Increasing Variance and Multiple Future Detection Decisions', in *Waste Testing and Quality Assurance*, ed. D. Friedman, ASTM STP 1075, American Society for Testing and Materials, Philadelphia, PA, 377–390, Vol. 3, 1991.
11. R.D. Gibbons, *Statistical Methods for Groundwater Monitoring*, John Wiley & Sons, New York, 1994.
12. J.W. Scott, N.T. Whiddon, R.F. Maddalone, *Compliance Monitoring Detection and Quantitation Levels for Utility Aqueous Discharges*, EPRI Final Report TR-103205, Palo Alto, 1993.
13. T. Hastie, R. Tibshirani, 'Generalized Additive Models', *Statistical Sci.*, **1**, 297–318 (1986).
14. H. Kaiser, H. Specker, 'Bewertung und Vergleich von Analyzenverfahren', *Fresenius' Z. Ann. Chem.*, **149**, 46 (1956).
15. H. Kaiser, 'Zum Problem der Nachweisgrenze', *Fresenius' Z. Ann. Chem.*, **209**, 1 (1965).
16. H. Kaiser, 'Zur Definition der Nachweisgrenze der Garantiegrenze und der Dabei Benutzen Begriffe: Fragen und Ergebnisse der Diskussion', *Fresenius' Z. Ann. Chem.*, **216**, 80 (1966).
17. USEPA, *Guidance on Evaluation, Resolution, and Documentation of Analytical Problems Associated with Compliance Monitoring*, 821-B-93-001, Washington DC, June, 1993.
18. G.J. Hahn, W.Q. Meeker, *Statistical Intervals: A Guide for Practitioners*, John Wiley & Sons, New York, 1991.
19. I. Guttman, *Statistical Tolerance Regions: Classical and Bayesian*, Hafner, Darien, CT, 1970.
20. USEPA, *Test Methods for Evaluating Solid Waste: Physical and Chemical Methods*, SW846, 3rd edition, Washington DC, July, 1992.
21. R.D. Gibbons, AML: A computer program for computing the Alternative Minimum Level. Scientific Software International, Chicago, 1996.
22. R.D. Gibbons, DETECT: A computer program for computing detection limits. Scientific Software International, Chicago, 1996.
23. L. Oppenheimer, T.P. Capizi, R.M. Weppelman, H. Mehta, 'Determining the Lowest Limit of Reliable Assay Measurement', *Anal. Chem.*, **55**, 638–643 (1983).
24. C. Liteanu, I. Rica, *Statistical Theory and Methodology of Trace Analysis*. Halsted Press, Chicago, 1980.
25. R.D. Gibbons, 'Statistical Prediction Intervals for the Evaluation of Ground-water Quality', *Ground Water*, **25**, 455–465 (1987).
26. S.R. Searle, *Variance Components Models*, John Wiley & Sons, New York, 1992.
27. R.D. Gibbons, N.E. Grams, F.H. Jarke, K.P. Stoub, 'Practical Quantitation Limits', *Chemometrics and Intelligent Laboratory Systems*, **12**, 225–235 (1992).
28. R.D. Gibbons, D.E. Coleman, R.F. Maddalone, 'An Alternative Minimum Level Definition for Analytical Quantification', *Environ. Sci. Tech.*, **31**, 2071–2077 (1997).
29. D.M. Rocke, S. Lorenzato, 'A Two-component Model for Measurement Error in Analytical Chemistry', *Technometrics*, **37**, 176–184 (1995).
30. G.W. Snedecor, W.G. Cochran, *Statistical Methods*, Iowa State University Press, Ames, IA, 1980.
31. M.E. Zorn, R.D. Gibbons, W.C. Sonzogni, 'Weighted Least Squares Approach to Calculating Limits of Detection and Quantification by Modeling Variability as a Function of Concentration', *Anal. Chem.*, **69**, 3069–3075 (1997).
32. R. Caulcutt, R. Boddy, *Statistics for Analytical Chemists*, Chapman & Hall, New York, 1983.

Dioxin-like Compounds, Screening Assays

Nigel J. Bunce and John R. Petrulis
University of Guelph, Ontario, Canada

1 Introduction 1
1.1 Dioxin-like Compounds Defined 1
1.2 Chemical Assay: Toxic Equivalents 2
1.3 Bioassays 2
1.4 Toxicity and Mechanism of Action of Dioxin-like Compounds 5
2 Mechanism-based Assays 8
2.1 Receptor–Ligand Binding Assays 8
2.2 Post-transformation Assays 12
2.3 Markers of Abnormal Metabolism 18
3 Immunoassays 19
4 Early Life Stage Bioassays 22
5 Cell Proliferation Assays 23
6 Conclusions 23
Abbreviations and Acronyms 24
Related Articles 24
References 24

*A dioxin-like compound (DLC) is a halogenated aromatic compound that has toxicological properties similar to those of the reference toxicant 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). The DLCs include other halogenated dibenzo-*p*-dioxins and dibenzofurans (especially those that are chlorinated in the 2-,3-,7-, and 8-positions), and coplanar polychlorinated biphenyls (PCBs). They associate with the aryl hydrocarbon receptor (AhR) protein, and share some biological end-points with TCDD, notably the induction of Phase 1 monooxygenase enzymes. Because DLCs normally occur as complex mixtures in environmental and biological samples, it is common to refer to the TCDD equivalent concentration [toxic equivalence (TEQ)], which is obtained by summing for each congener its actual amount or concentration by an empirical toxic equivalency factor (TEF). Conventional analysis of DLCs by gas chromatography/mass spectrometry (GC/MS) is cumbersome and expensive, and much attention has been given to developing bioassays that will yield a measure of the TEQ in a single determination. Many such bioassays are mechanism-based, meaning that the assay end-point is one of the steps along the pathway of the mechanism of action.*

This article begins with an overview of bioassay methods in general, pointing out similarities and differences between chemical assays and bioassays, before describing particular assays that have been developed for DLCs. Among mechanism-based assays, AhR binding assays are well established, and are useful because they include the dioxin-like activity of both productive and unproductive compounds. In assays based on subsequent stages of the mechanism of action, unproductive substances can antagonize the responses of productive compounds, a phenomenon that is explicable in terms of target molecule antagonism, in which the AhR protein is the target molecule.

Immunoassays for DLCs have recently received much attention; although their spectrum of cross-reactivity does not always correlate well with TEFs, there has been important recent progress in terms of the sensitivity and detection limit (DL) of these assays. Finally, early life stage bioassays are becoming increasingly important as research has revealed the toxicological effects of DLCs during development.

1 INTRODUCTION

1.1 Dioxin-like Compounds Defined

The term ''DLC'' is applied to compounds having toxicological properties similar to the reference toxicant TCDD. DLCs are halogenated aromatic compounds that bind to the AhR protein, or that cause overt toxic and biochemical end-points similar to those of TCDD, as will be discussed in section 1.4. DLCs include other polychlorinated dibenzo-*p*-dioxin (PCDD) compounds (especially those that are chlorinated in the 2-,3-,7-, and 8-positions: Figure 1), the corresponding polychlorinated dibenzofuran (PCDF) compounds, bromo- and bromochloro-substituted dibenzo-*p*-dioxins and dibenzofurans, and coplanar PCBs. Many polycyclic aromatic hydrocarbons (PAHs) also associate with the AhR protein, and share some biological end-points with TCDD, notably the induction of Phase I monooxygenase enzymes.

The organization of this article is as follows. The Introduction describes bioassay methods in general, and includes a brief discussion of the mechanism of toxic action of TCDD, on account of our bias towards mechanism-based assay methods. Subsequent sections address sequential steps along the mechanism of toxic action that have been used to develop assays for DLCs. These are followed by a brief discussion of immunoassays (which are not directly mechanism-based) and assays based on overtly toxic end-points such as early life stage and cell proliferation bioassays.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

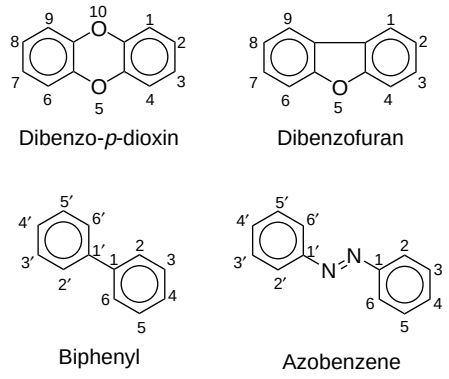

Figure 1 General chemical structures and numbering system for several DLCs.

1.2 Chemical Assay: Toxic Equivalents

The standard chemical method of analyzing DLCs is labor intensive, and therefore expensive. After spiking with an isotopically labeled recovery standard such as [$^{13}C_{12}$]TCDD, sample preparation involves solvent extraction and multiple chromatographic procedures to isolate the PCDD/PCDF fraction. Quantitation of individual PCDD and PCDF congeners is achieved by capillary gas chromatography/high-resolution mass spectrometry (MS),[1,2] giving the concentration of each congener in units such as nanograms of congener per gram of sample.

Chemical analysis cannot provide biological information such as an estimate of a sample's toxic potency. One such measure, the TEQ, is obtained from analytical data by the use of a TEF, that relates the potency of DLC X to that of the reference toxicant TCDD, whose TEF = 1. TEFs are not precisely defined parameters, but are based on a consensus of relative potency, considered over a wide range of end-points and species; the size of the available database varies from congener to congener. The TEQ of a sample is given by [(concentration of X) × (TEF of X)]. In the commonly used I-TEF scheme,[3] TEFs are available for the 17 ''most toxic'' PCDD and PCDF congeners that are chlorinated in the 2-,3-,7-, and 8-positions, and provisional TEFs have been proposed for the coplanar PCB congeners.[4] The potencies of different DLC congeners vary greatly, as do their relative concentrations in environmental samples. For example, the concentrations of coplanar PCBs in many animal tissues have been found to be orders of magnitude higher than those of PCDD/PCDFs,[5,6] making the PCBs major contributors to the TEQ of a sample, despite their small TEF values.

The TEQ approach is inherently cumbersome, because a unique TEF must be assigned to each DLC, in order to include its contribution to the calculated potency. The TEQ will be underestimated if the sample contains DLCs for which no TEF has been established, and the more congeners to be assessed, the greater are the complexity of analytical separation and the cost of analysis. Another problem is that the TEQ obtained by GC/MS may not afford an optimal measure of the potency of a sample towards a particular end-point: potency rankings of DLC congeners vary with the end-point and species examined, and not all DLCs produce all end-points. A third problem is that the TEF approach explicitly takes the TEQ of a complex sample to be the sum of the contributions of its components, assuming an additive response to a complex mixture of DLCs (no synergism or antagonism). Putzrath[7] has questioned the assumption of additivity in the TEF approach from a different perspective. He argues that for parallel log dose–response curves (as is common for DLCs) the relative concentrations of toxicants X and Y needed to produce equal large responses are different from the relative concentrations needed to produce equal small responses. The apparent TEF of a given DLC relative to TCDD will therefore vary according to the position on the dose–response curve, and so considerable inaccuracies might result when TEFs derived from high-dose laboratory studies are used to extrapolate to low-dose environmental exposure.

1.3 Bioassays

We define a bioassay as a technique whose end-point is attained through the use of either living organisms or biological materials (e.g. tissues, enzymes, or antibodies). Detection of the end-point signals qualitatively the presence of the analyte, while the intensity of the response is used to quantitate the analyte, as in conventional chemical analysis. In either case, the assay end-point must be calibrated against the concentration of the analyte. In vivo bioassays employ living animals or plants. In vitro methods can utilize living cells, such as bacteria or cultures of animal or plant cells, or they can use isolated tissues, such as liver slices excised from a whole animal. They may also employ subcellular fractions, such as microsomes, chloroplasts, or cytosol, or preparations containing or comprising specific biomolecules such as receptors, enzymes, or nucleic acids. The latter bioassays are hardly different from chemical assays, and many of them use conventional chemical detection techniques such as absorption or fluorescence spectroscopy. One recently developed assay for DLCs, for example, employs the firefly luciferase gene that has been genetically engineered into bacterial and yeast cells (section 2.2.2). In this assay, the measured parameter is light emission catalyzed by luciferase, detected by a photomultiplier tube that records and quantitates light intensity. Likewise, an enzyme-linked immunosorbent assay (ELISA) relies on an enzyme that catalyzes the formation of a product

that can be quantitated by techniques such as absorption or fluorescence spectroscopy (section 3). The quantity of the product of the enzymatic reaction is proportional to the amount of the antigen–antibody (Ab) complex that was formed.

The cost of analysis generally decreases in the order in vivo assays > in vitro assays with whole cells > in vitro assays with subcellular fractions, but so does the biological relevance of the assay end-point in terms of human health or environmental risk assessment. Animal assays increase in cost with the type of organism (invertebrates < fish < rodents < primates) and with the length of time that is needed for the results to manifest themselves (acute toxicity tests < carcinogenicity tests < multigenerational reproductive tests). Although the use of in vitro assays with bacteria or immortalized cell lines avoids the use of live animals, biological relevance requires some in vivo animal testing when the final objective is human health or environmental risk assessment. Even if an in vivo bioassay replicates the actual toxic end-point, laboratory animals must be used as surrogates when humans are the target organism of concern. Extrapolation of results obtained with laboratory animals to humans is difficult because of the varying susceptibilities of different species and strains towards intoxication by DLCs.[8]

Unlike chemical assays, many bioassays monitor *function*, and have little or no diagnostic capability for identifying the agent responsible for causing the effect. For example, the acute toxicity of industrial effluents may be monitored by means of the 96 h rainbow trout test or the 24 h *Daphnia magna* test, both of which employ single organisms as surrogates for aquatic species in general. From a regulatory perspective, the lack of diagnostic information is immaterial, since what is required is knowledge of whether or not the effluent is acutely toxic; the identity of the toxicant can be ascertained in a follow-up study, if it is not already known. Other effects such as developmental toxicity in aquatic organisms can also be conveniently studied in this way (see section 4). Similar considerations apply to testing new drugs or pesticides for effects such as carcinogenicity or teratogenicity; in such cases, the identity of the substance is known, and what is required is quantitative information about its possible adverse biological effects.

Bioassays with live organisms or cell cultures, enzyme assays, and immunoassays monitor overt end-points that are different from the end-point of interest to environmental or human health, and a correlation must then be drawn between the surrogate end-point and the end-point of interest. Acute toxicity tests such as the LD_{50} have often been used, incorrectly, as a measure of the toxic potencies of substances towards chronic exposures. This is particularly inappropriate for DLCs, which cause lethality by mechanisms different from those involved in other end-points.[9]

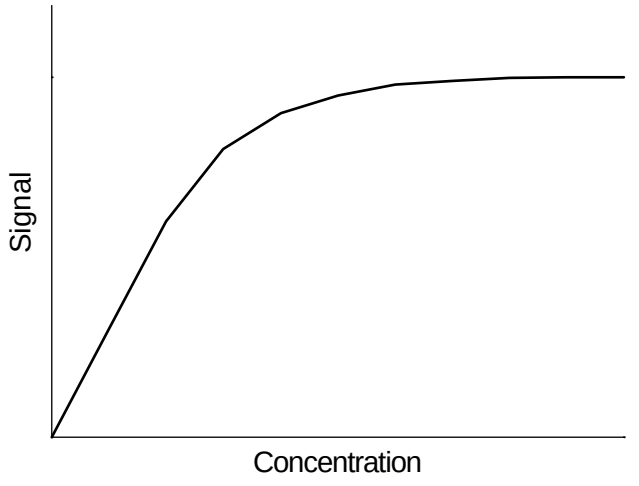

Figure 2 Plot of response intensity vs analyte concentration demonstrating the nonlinear curve obtained when the response is saturable, as is the case in most bioassays.

Many issues concerning the applicability of bioassays are fundamental questions associated with any analytical procedure, such as the shape of the dose–response curve and its limits of quantitation. Consider a generic calibration curve such as that shown in Figure 2. Parameters of analytical importance include its shape (linear or nonlinear) and slope (calibration sensitivity), the range of analyte concentrations over which the calibration curve is valid, reproducibility, and the DL. In chemical analysis, the DL is frequently defined as the response that is three standard deviations away from that of the "blank".[10] The analyte concentration at the DL is given by $c_{DL} = 3 \times s_{bl}/S$, where s_{bl} is the standard deviation of the blank and S is the calibration sensitivity. The ideal analytical procedure has a linear calibration curve over a wide dynamic range and has a large slope, permitting a low DL. The foregoing parameters are mutually dependent: linear responses have inherently wider dynamic ranges than responses whose calibration sensitivities fall progressively with increasing analyte concentration. Low DLs require techniques that have large calibration sensitivities (i.e. a small amount of analyte produces a response that is significantly different from a blank) and/or high precision, implying reproducibility and a small standard deviation of the measurements. Chemical analysts also seek procedures that have high specificity for the analyte of interest, implying no interference by other analytes but, as noted already, this is often not possible with bioassays.

A wide dynamic range is rare with in vitro bioassays, and is impossible for in vivo bioassays, because the response is saturable. In acute lethality or teratogenicity assays, for example, there will be some concentration

of analyte above which essentially all the individuals die or are born with birth defects. Consequently, it is impossible for the response to increase further, and so $S \rightarrow 0$ at high analyte concentration. Moreover, the experimental uncertainties are usually larger than in physical techniques, because of biological variability among individuals in the groups being tested, even when using inbred strains of organisms of the same sex, age, and weight. At the lower limit of quantitation, a relatively high concentration of analyte may be needed to achieve a response (the lowest observed effect level) that differs in a statistically significant way from that of the unexposed group.

A particular difficulty arises with respect to the carcinogenicity of DLCs. TCDD is a weak carcinogen in laboratory animals such as rats[11] and a disputed human carcinogen.[12] The linear models for extrapolating cancer incidence from experimentally observable doses to the low doses associated with occupational or environmental exposure assume no threshold for the response.[13] Regulators may define the dose corresponding to a standard such as one excess cancer in 10^6 lifetimes [e.g. Sielken[14]], as corresponding to risks that might be acceptable to most members of society. Such a response intensity corresponds to concentrations of DLCs that are far below the lowest observed effect level in a bioassay, and a long extrapolation is inevitable.

The foregoing factors mean that quantitative data can rarely be obtained from in vivo bioassays over a range of more than about one order of magnitude. Calibration curves such as Figure 2 are therefore not generally used to evaluate in vivo bioassays; instead an approach such as the analysis of variance (ANOVA) may often be used to give information about whether the various treatment groups are statistically different from each other and from controls. A more appropriate presentation of the data is a histogram such as Figure 3 rather than the continuous function implied by Figure 2.

Another analytical question is whether an assay affords *diagnostic* information as well as quantitative information; in other words, whether the parameter being monitored identifies the analyte. Bioassays usually have little diagnostic capability in comparison with chemical methods of analysis such as nuclear magnetic resonance (NMR) and MS, although, as already noted, this may be irrelevant for a functional bioassay.

Important applications for DLC bioassays include environmental remediation programs, which usually require the screening of large numbers of samples to determine which sites require remediation, or to monitor the progress of cleanup. In these samples, the analytes must be extracted from matrices such as soil, sediment, and biota. The extent of pretreatment (cleanup) of the sample that is needed before the assay can be applied

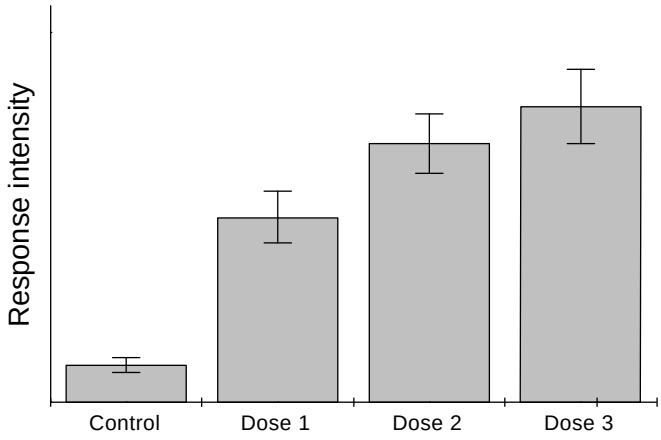

Figure 3 An example of bioassay data presented as a histogram.

is a major consideration, because much of the cost of conventional chemical analysis for DLCs is associated with the lengthy cleanup of the sample prior to analysis by GC/MS. Monitoring or screening assays are of limited use unless a major portion of the cleanup effort can be avoided (e.g. Höckel et al.;[15] Schwirzer et al.;[16] Hu et al.[17]). This issue is often not addressed by the advocates of new bioassays.

Mixtures of compounds present further difficulties in bioassays compared with chemical analysis. A major problem in chemical assays is interference between the responses of different analytes. For example, the mass, infrared (IR), and NMR spectra of mixtures are generally uninterpretable, because the analyst cannot confidently assign the peaks to the correct analytes. This difficulty is overcome by the use of linked methods such as GC/MS that combine the resolving power of chromatography with the diagnostic capability of spectroscopic methods. By contrast, the response in a bioassay is due to the whole sample rather than individual analytes, and the activity can be calibrated against the response of a standard substance. Assays for DLCs are normally calibrated against TCDD, and the response of the sample is reported in TCDD equivalents, i.e. the dose of TCDD that would cause the same response intensity. This approach inherently assumes additive behavior of the mixture's components, without synergism or antagonism of the biological response. Synergism leads to an overestimation of the concentration of the analyte, but is rarely encountered. Antagonism, which is fairly common, can underestimate the concentration of the analyte to the point of a false negative (the sample mistakenly appears to be uncontaminated).

In vivo antagonism can occur by numerous mechanisms that are difficult to differentiate; the situation in vitro is usually less complicated. A common situation is that the

analyte and the antagonist compete for a limited amount of a target biomolecule, such as a receptor or an enzyme; the analyte is processed and leads to the normal biological response or enzyme product, but the antagonist does not get processed. An agonist is a substance that elicits the toxic or biochemical end-point under study, whereas an antagonist blocks or masks the biological activity of an agonist. When administered on its own, such a substance is better described as unproductive rather than as an antagonist.

Antagonism of this sort (target molecule antagonism) is seen only at high concentrations of analytes (agonist with or without antagonist), at which most of the active sites on the target molecule are occupied. At low total occupancy of the target molecule, the unproductive molecule can occupy vacant target molecules without displacing the agonist, and no inhibition of the assay end-point is seen; conversely, at high total occupancy, the unproductive occupation of target molecules by the antagonist can occur only by displacing the agonist. This phenomenon has been seen in teratogenicity studies in which pregnant mice are administered TCDD, with and without coadministration of PCBs.[18] At high doses of TCDD, coadministration of PCBs antagonizes the teratogenic response, owing to occupation of the AhR target molecule by the nonteratogenic PCBs; antagonism is not seen at low doses of TCDD, at which receptor occupancy is incomplete.

Although a bioassay for DLCs can give a numerical estimation of the TEQ of the sample as a whole, relative to the reference toxicant TCDD, it is unlikely that regulatory decisions could ever be based solely on the results of such assays owing to the lack of diagnostic information about the analyte(s). Lesnik[19] has identified several regulatory issues in the context of using immunoassays, but his comments apply equally to other bioassays. The concerns include extraction efficiency and comparison of assay results with those from previously validated procedures, the occurrence of false negatives and false positives, and cross-reactivity with other analytes. These will be discussed in turn.

Extraction efficiency is difficult to monitor even when a recovery standard is used, because this added surrogate may be more easily removed from the matrix than endogenous DLCs. The practice of adding a surrogate, such as [$^{13}C_{12}$]TCDD in chemical analysis by GC/MS, is impractical in bioassays, because the surrogate would itself be a DLC and thus contribute to the assay. In principle, the surrogate could be used in the standard addition mode,[20] but this seems not to have been tried in practice. As for validation against other procedures, few proposed assays for DLCs have been rigorously tested in this way (e.g. Zacharewski et al.[21]). Ideally, the assay should require the minimum of extraction and other work-up, to save on analysis time (for a discussion and example, see Hu et al.[17])

False negatives and false positives: a major application of dioxin bioassays involves screening samples for activity and determining which are below and which above a consensus cut-off TEQ, so that only those samples assaying above the cut-off are taken to detailed chemical analysis. False positives occur when an interfering, but not dioxin-like, substance can evoke the end-point being monitored; to this extent, false positives make the bioassay conservative because the sample's contamination is overestimated. False positives increase the number of samples that must be subjected to confirmatory analysis; a high rate of false positives negates the benefit of a low-cost bioassay. False negatives, on the other hand, suggest that contaminated samples are clean; from a regulatory viewpoint, any significant occurrence of false negatives is unacceptable. False negatives are most likely to occur if a relatively nontoxic congener masks or antagonizes the activity of a more potent congener.

Cross-reactivity in bioassays would be a problem if one wished to monitor the concentration of a single DLC (e.g. TCDD). Realistically, bioassays for DLCs will be applied to assessing the TEQ of a sample; cross-reactivity is advantageous if each DLC behaves additively in the assay and if the responses of different DLCs match their TEFs reasonably well. Immunoassays for dioxins in environmental samples exhibit cross-reactivity; it is usually possible to predict the types of structures will cross-react, but the real question is whether the TEQ determined by a particular assay reflects the toxic potency of the sample. Bearing this in mind, we state here our prejudice for *mechanism-based bioassays*: those that use a point along or minimally divergent from the established metabolic pathway leading to toxicity as the indicator of potency.

1.4 Toxicity and Mechanism of Action of Dioxin-like Compounds

Most of our knowledge of the toxicity of DLCs is based on that of TCDD, which elicits numerous toxic responses. In addition to the widely reported acute lethality, which occurs at very low doses and proceeds via a wasting syndrome, other effects include reduced immunocompetence, hyperplasia, carcinogenicity, developmental toxicity (including teratogenicity and errors in sexual differentiation), and altered porphyrin metabolism. TCDD does not induce all effects in all species, and evidence is lacking that all effects are exhibited by all DLCs (for reviews, see Landers and Bunce;[22] Whitlock;[23] Okey et al.;[24] Mukerjee;[25] DeVito and Birnbaum[26]).

The complexity of biological responses to TCDD, and extremely limited knowledge about other DLCs and DLC mixtures, make it difficult to assess the toxic potential of

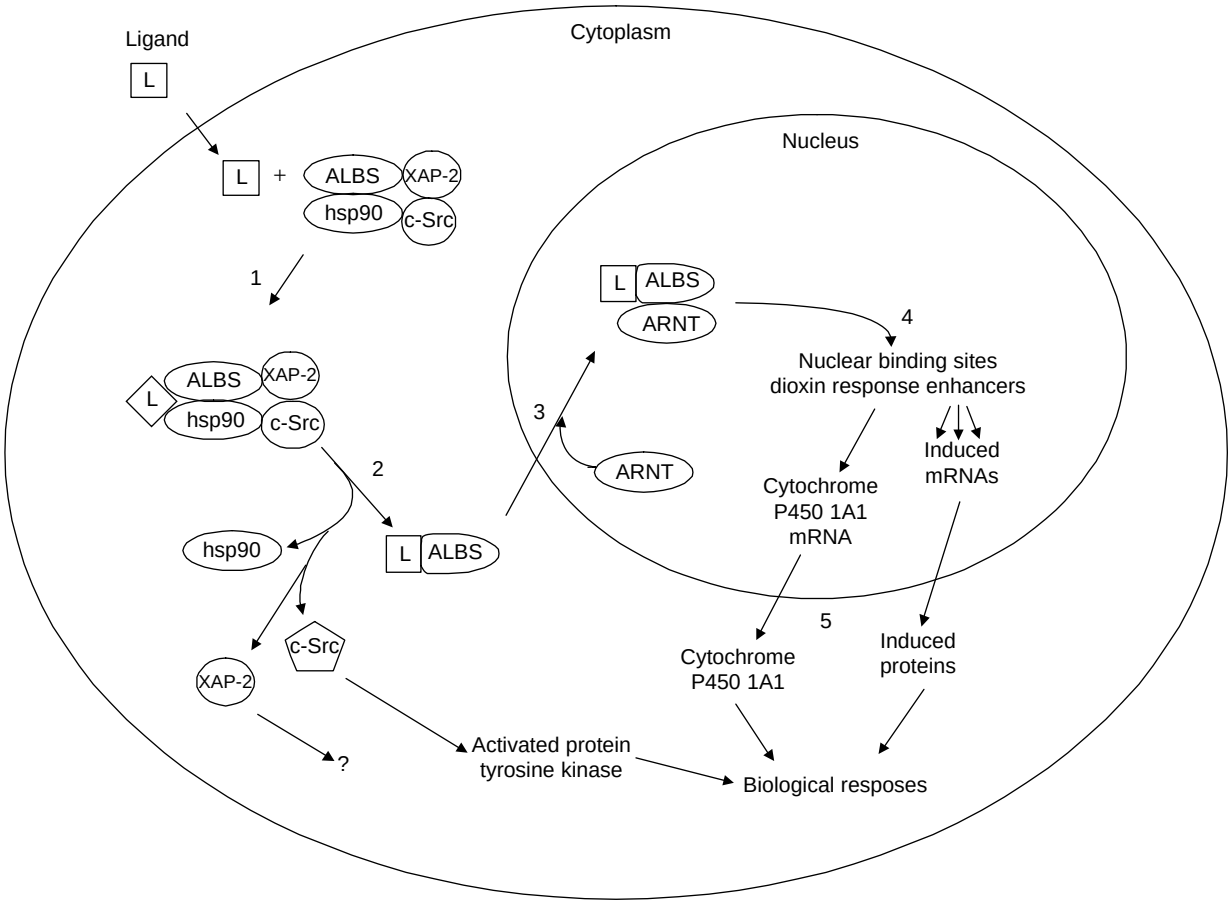

Figure 4 Schematic representation of the mechanism of action of DLCs. A ligand (L) enters the cell and binds to the aryl hydrocarbon receptor ligand binding subunit (ALBS) of the cytosolic AhR. The receptor undergoes transformation/activation, nuclear translocation, and dimerization with the aryl hydrocarbon receptor nuclear translocator protein (ARNT), following which it binds to dioxin response enhancers (DREs), resulting in the induction of cytochrome P450 1A1 and a variety of other proteins. The numbers indicate the stages in this mechanism on which bioassays can be based: (1) receptor–ligand binding; (2) receptor transformation and activation; (3) nuclear translocation; (4) DNA binding; and (5) enzyme induction.

environmental samples, which usually contain complex mixtures of DLCs. Combustion by-products such as fly ash contain a wide range of PCDDs and PCDFs,[27] while biota such as fish, meat, and milk contain a large number of organochlorines. Among the latter, PCDDs and PCDFs are usually minor constituents in comparison with PCBs, at least some of which are dioxin-like, and organochlorine pesticides such as DDT, that in general are not. Among DLCs, PCDDs, PCDFs, and coplanar PCBs combine toxicity with environmental persistence, and their lack of chemical and metabolic reactivity allows them to bioaccumulate through food webs. Other compounds such as partly methylated PCDDs and PCDFs[28] and the chlorinated azobenzenes are of less concern because they are metabolized more rapidly.

Mechanism-based bioassays utilize current knowledge about the mechanism of action of the toxicant, by exploiting a biochemical event in the mechanism of action as an assay end-point from which to estimate the TEQ. The closer the relationship between the biochemical event being monitored and the toxic end-point of interest, the greater is the biological relevance of the assay. The mechanism of action of TCDD as it is currently known is summarized in Figure 4. As already noted, the behavior of other DLCs is generally assumed to be similar.[22,24,29]

Step 1. TCDD is believed to enter cells by passive diffusion: as a lipophilic substance it has a high affinity for the lipid bilayers constituting the plasma membrane, and bioconcentrates in adipose tissue.[30] The biochemical target of TCDD in vertebrates is the AhR protein, so named because it was first detected as a binding protein for certain aromatic hydrocarbons. Most, if not all, biochemical effects of TCDD are AhR-mediated i.e. association of TCDD with the receptor is a necessary precondition for toxicity. The AhR is an oligomeric ~280 kDa protein complex found in the cytosolic fraction

of tissue preparations. It contains an ALBS with molar mass 90–120 kDa, depending on the species,[31] and a molecule of the 90 kDa heat shock protein (hsp90), which is thought to mask the DNA-binding site[32] and keep the receptor in a conformation that can bind the ligand.[33] Association of the AhR with hsp90 is required to maintain functional signal transduction.[34] The remaining composition of the cytosolic AhR has not been elucidated, but evidence exists that it may contain a molecule of the tyrosine kinase c-src[35] and/or the 38 kDa hepatitis B virus X-associated protein 2 (XAP-2).[36]

The susceptibility of animal strains to intoxication by TCDD correlates with the affinity of TCDD for the AhR. In highly susceptible animals such as the C57BL/6 mouse, estimates of the association constant are $>10^{11}\,M^{-1}$;[37,38] less susceptible strains such as the DBA/2 mouse have a lower affinity form of the AhR.[39,40] The human form of the AhR also has low affinity for TCDD.[41–43] Binding affinity to the AhR has long been used to identify substances that might exhibit dioxin-like activity.[44] Numerous correlations, beginning with those published by Poland et al.,[45] and continuing with quantitative structure–activity relationship (QSAR) studies through the 1980s, showed an approximate rank ordering between receptor binding affinity and enzymatic or overtly toxic end-points. These provided the basis for the accepted paradigm that the toxicity of DLCs is AhR mediated.[29,46,47]

The AhR preferentially binds lipophilic molecules of the size and shape of TCDD [planar, molecular dimensions ca. 1.0×0.3 nm[45]]; the poorer the fit between the DLC and the receptor binding site, the lower is the affinity (Figure 5). Much thought has been given to the natural role of the AhR and to finding endogenous ligands for it. One possibility is binding to foreign substances such as PAHs and indole-3-carbinol (present in cruciferous vegetables),[48] leading to induction of elevated levels of monooxygenase enzymes that metabolize the foreign substances. Other endogenous compounds that have been demonstrated to bind and activate the AhR are tryptophan and tryptophan metabolites[49] and the heme degradation products bilirubin and biliverdin.[50] It is a reasonable but unproven supposition that many of the toxic responses of DLCs may result from altering the intracellular levels of monooxygenases, which are powerful oxidants. The AhR may also be a regulatory protein that controls the expression of many unrelated genes,[24]

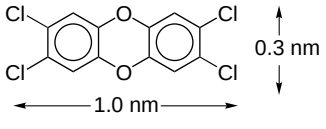

Figure 5 The dimensions of TCDD.

which changes the focus of the mechanism of intoxication from monooxygenase enzyme induction to the activation of c-src and the activation of a kinase cascade.

Step 2. Following association with a ligand such as TCDD, the AhR–ligand complex undergoes a poorly understood process termed transformation, which is temperature- and species-dependent. Immunoprecipitation experiments have shown that AhR subunits are released at this stage,[34,51,52] leaving behind free ALBS in the form of its TCDD complex. Although the initially formed AhR–TCDD complex can release its ligand, TCDD is bound almost irreversibly, although not covalently, following transformation. This suggests that a significant conformational change in ALBS accompanies transformation, and that the role of hsp90 may be to act as a molecular chaperone, maintaining ALBS in a conformation that can bind ligand.[52] An ambiguity exists in older literature reports of ligand binding affinity to the AhR, because it is unclear whether the untransformed or transformed AhR–ligand complex was the species present after incubation of the ligand and the receptor preparation.

Step 3. The transformed ALBS–TCDD complex penetrates the nuclear membrane, where it heterodimerizes with the ARNT through basic helix–loop–helix motifs.[53,54] Phosphorylation may occur at this stage, following which heterodimers bind to recognition elements on DNA known as DREs.[55] The term *enhancer* reflects the *increase* in the expression of gene products following exposure to substances such as TCDD. DRE binding with TCDD occurs with high affinity.[56,57] Several DREs have been identified, and are characterized by similarity of their nucleotide sequences. In vitro binding experiments have been carried out with oligonucleotides containing the consensus sequence.

Steps 4 and 5. The DREs appear to act as genomic switches, initiating the transcription of mRNAs, which in turn yield the appropriate protein gene products.[23] Transcription of *CYP 1A1* (cytochrome P-450 1A1 mRNA) begins within a short time of exposure to TCDD, and all necessary components appear to be present constitutively, provided that the cells possess functional AhR and ARNT. The exact mechanism of regulation is not yet completely understood; a very recent suggestion is that the liganded AhR complex activates gene expression of a repressor protein that competes with ALBS for binding ARNT, thereby permitting down-regulation of AhR response subsequent to activation.[58] Past speculation has focused on the induction of a plethora of gene products that lead to different toxicological responses; the only well-characterized pathway, however, is monooxygenase enzyme induction. The recognition of the tyrosine kinase pathway suggests that the latter may be responsible for other biological responses to DLC exposure.

One use of mechanism-based bioassays is to follow a particular DLC stepwise through its mechanism of action to determine whether it acts as an agonist or as an antagonist at each step. Consider the case of two substances that both bind the AhR, only one of which induces CYP 1A1. An agonist not only binds the AhR, but successfully surmounts all subsequent hurdles leading to induction of the enzyme; its antagonist may bind the AhR, thereby reducing the stock of AhR molecules available to bind the agonist, but will not necessarily give a positive result at each subsequent step. The use of a series of assays, step by step in the mechanism, can allow the investigator to probe the exact mechanism of action of such compounds. For example, in the case of AhR binding assays, not all substances that bind to the AhR are necessarily toxic, as is clearly demonstrated by indole-3-carbinol[59,60] and certain nonplanar PCBs such as 2,2′,4,4′,5,5′-hexachlorobiphenyl (PCB 153). These compounds are active in an AhR binding assays, but do not cause dioxin-like toxicity because they undergo the subsequent association with DRE only inefficiently.[61]

2 MECHANISM-BASED ASSAYS

2.1 Receptor–Ligand Binding Assays

Receptor–ligand binding assays are employed in mechanistic biochemistry and toxicology to assess the relative binding affinity (RBA) of ligands, in order to define parallels between RBA and biological potency. They can also be used to assess TEQ, by determining the amount of TCDD with the same affinity for a known aliquot of AhR as the test substance or mixture. Binding affinity can be determined by means of a saturation assay or a competition assay.

2.1.1 Saturation Assays

In a saturation assay, increasing concentrations of the test substance (ligand) are incubated with a fixed amount of AhR, giving increasing amounts of bound ligand. An experimental complication in AhR–ligand binding assays is that the concentration of AhR in cells is so small [about 10^5 receptors per cell[29]] that the concentrations of the liganded and unliganded protein cannot be measured directly. The traditional approach is to use a radiolabeled ligand such as [^{3}H]TCDD; the parameter that is actually measured is the concentration of radiolabeled receptor–ligand complex obtained following incubation of the receptor preparation and the ligand(s).

For saturation binding assays using radioligands, the desired measurement is the concentration of AhR bound to the radioligand. The principle is that the ligand can bind specifically to saturable, high-affinity sites on the receptor, and nonspecifically but with low affinity to other sites on the receptor or on other proteins. Two determinations must be made. The first is to incubate a fixed amount of the receptor preparation with increasing concentrations of radiolabeled ligand, and measure the concentration of receptor–ligand complex (''total binding''). In the second determination, the experiment is repeated with a large (e.g. 200-fold) excess of unlabeled ligand in addition to the radioligand to measure nonspecific binding. The specific binding is the difference between the two measurements. For the assay to be accurate the ligands must be present in excess over the saturable sites (at the highest concentrations) but not in excess over the nonspecific sites.

Saturation assays involving radioligands are usually analyzed by the method of Scatchard,[62] which assumes that the concentration of unliganded receptor (R) is at all times represented by the difference between the initial, total receptor concentration (R_0) and the concentration of specifically bound receptor–ligand complex (RL) [Equation 1]:

$$[R] = [R]_0 - [RL] \tag{1}$$

Under these conditions, Equation (2) gives the relationship between K_d, the dissociation constant of the receptor–ligand complex, and the concentration of the radioligand (L):

$$\frac{[RL]}{[L]} = \frac{[R]_0}{K_d} - \frac{[RL]}{K_d} \tag{2}$$

From a plot of [RL]/[L] vs [RL], the slope gives K_d and the intercept affords the total concentration of specific binding sites.

Unfortunately, the Scatchard method is inapplicable to the AhR assay under most conditions, because the unliganded receptor is very susceptible to thermal inactivation of its binding capability. This means that $[R] < ([R]_0 - [RL])$ and, as a result, K_d is overestimated, often by several orders of magnitude.[63] Previous experiments in our laboratory[38] have shown that it is possible to estimate K_d from kinetic experiments, even in the presence of thermal inactivation of the receptor. The method is, however, cumbersome and inaccurate, because the rate constants for both receptor–ligand association and complex dissociation must be fitted to the time course of complex formation.[64]

The whole-cell binding assay reported by Dold and Greenlee[65] has been developed to screen cell culture lines for their AhR content, rather than screening different samples for their DLC content. In this assay, the cells are incubated with radiolabeled TCDD (with

and without unlabeled competitor), and after appropriate manipulations the cells are counted for their radioactivity. Uptake of radioligand is observed only for cell lines having a functional high-affinity AhR.

2.1.2 Competition Assays

Competitive binding assays involve incubating the receptor preparation with a mixture of ligands, one of which is radioactive (the reference radioligand, usually [^{3}H]TCDD in studies involving DLCs). The reference radioligand should bind more avidly than the unlabeled competitor(s)[66] and be present in sufficient concentration to saturate the receptor's binding sites. In the absence of any competitor, the maximum concentration of specifically bound receptor–ligand complex is formed; increasing the concentration of competitor, at a fixed concentration of the reference radioligand, reduces the level of bound radioactivity, ultimately to near zero. The usual objective is to determine the EC_{50}, the concentration of competitor that displaces 50% of the radioligand. At the EC_{50}, the radioligand and the competitor have equal occupancies of the AhR, and so their RBAs are in the ratio of their concentrations. Alternatively, binding ligands might be distinguished from nonbinding ligands by selecting a cut-off RBA as the limit of a positive response. For example, suppose that TCDD is the reference radioligand and that 10^{-5} is the cut-off RBA. For a concentration of radiolabeled TCDD of 1 nM (a reasonable value), a series of commercial chemicals to be tested could be prepared at 100 μM. The RBA will be $<10^{-5}$ if the bound radioactivity for radioligand + competitor is >50% of that for the radioligand alone. Relative EC_{50}s may reflect kinetic affinities or equilibrium affinities, depending on whether the association between receptor and ligands is freely reversible.[67]

The common graphical presentation of a competition assay is a sigmoidal curve of percent specific binding vs log (competitor concentration). The EC_{50} may be obtained either graphically (Figure 6) or, better, using a statistical technique such as the probit or logit method.[68] We have stressed elsewhere that the statistical uncertainty in the EC_{50} will be high unless a large number of competitor concentrations is used to define the sigmoidal curve.[69] Care must be used in interpreting the sigmoidal curve, because of the incorrect impression that at low ligand concentrations there is a threshold below which the level of specific binding (and by implication the toxic potency) of the DLC is insignificant [e.g. Roberts[70]]. This is an artifact of using a logarithmic concentration scale (Figure 7a and b).[71]

An advantage of the competitive method is that more than one unlabeled competitor can be present; an occupation ratio is established for the specific binding sites on the receptor, based on the relative concentrations and affinities of all the competitors. For a given competitor,

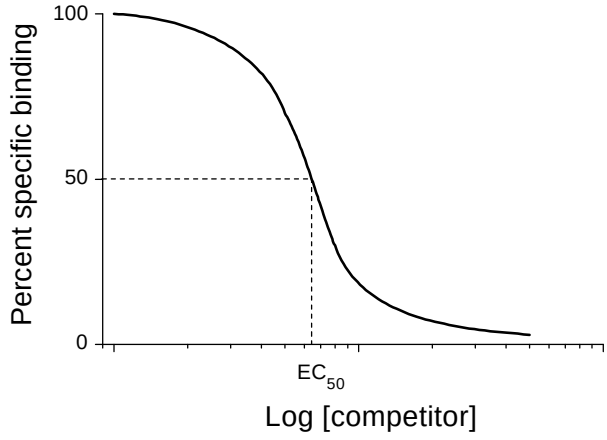

Figure 6 Example of a competition curve, demonstrating how an EC_{50} value is determined graphically.

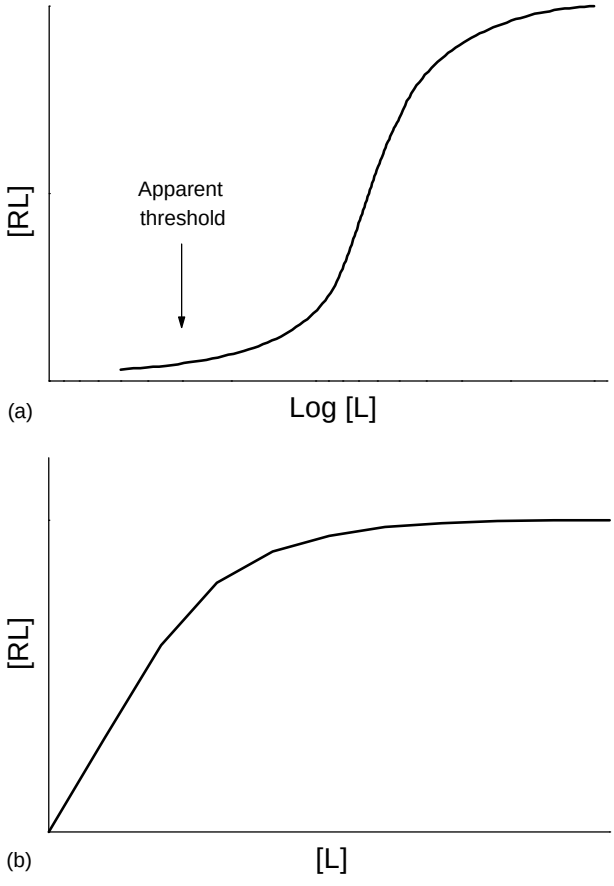

Figure 7 The relationship between the receptor–ligand complex formed and the concentration of free ligand. When a logarithmic concentration scale is used (a), there is an apparent threshold below which the level of specific binding is insignificant. However, this is only an artifact of the logarithmic scale, as seen when a linear scale is used (b).

or mixture of competitors, the EC_{50} value is by definition equivalent to the concentration of the reference radioligand. In the case of the AhR, with 1.0 nM [^{3}H]TCDD as the reference radioligand, whatever the actual concentration of the competitor(s) at the EC_{50}, this will occupy the same fraction of the receptors as 1.0 nM TCDD. Therefore, a measure of the TEQ is obtained from a single experiment.

An important check on this methodology is to use the unlabeled form of the reference radioligand as the unlabeled competitor; the EC_{50} value for the unlabeled ligand should equal the concentration of the reference radioligand. In consequence, the EC_{50} concentration obtained from a competition experiment depends upon the concentration of the reference radioligand; therefore in comparing literature values of EC_{50}s one must take care to note the concentration of the reference radioligand. RBAs, however, are independent of the concentration of the reference radioligand.

In the analysis of an extract from an environmental matrix, the analyst prepares a series of dilutions of the initial extract and plots dilution (abscissa) vs bound radioactivity (ordinate). Aliquots of these solutions are incubated at a fixed temperature for a predetermined time with fixed concentrations of the AhR preparation and the reference radioligand. The bound radioactivity is conveniently expressed as the percentage of the maximum specific binding (%SB) obtained with the reference radioligand in the absence of competitor. The %SB is determined for each dilution, from which one can estimate the extent by which the initial extract must be diluted to produce 50% of the maximum specific binding. At this dilution of the environmental extract, the DLCs in the extract occupy half the receptor or other recognition sites (EC_{50}); the TEQ of the sample is then equal to the TEQ of the reference radioligand, from which one can work back to the concentration in the undiluted sample.

A complication for interpreting RBAs, whether they represent kinetic or equilibrium affinities, is that the receptor–ligand complex may be untransformed, partly transformed, or wholly transformed, depending on the conditions of the experiment and the source of the AhR.[51] This issue was not appreciated before transformation was recognized as a step in the mechanism of toxic action of TCDD. One clear-cut study is that of Bradfield et al.[37] using mouse hepatic AhR at 4 °C; their results refer to equilibrium, because mouse hepatic AhR neither inactivates[72] nor transforms[73] at a significant rate at this temperature, and so the Scatchard method of analysis is justified. In contrast, it is likely that most studies involving rat hepatic receptor near room temperature have been conducted under thermally inactivating, and hence kinetic, conditions.[67,69]

Competition assays all suffer from the disadvantage that the signal (in this case, the intensity of radioactivity) is maximum when there is zero competitor and declines asymptotically towards zero as the concentration of the competitor increases. The DL for the competition assay is therefore difficult to define. Schneider et al.[69] took a 20% reduction in specifically bound radioactivity to be distinguishable from the 100% level, which is equivalent saying that the TEQ of the unlabeled competitor must be at least 25% that of the reference radioligand. For 1 nM [^{3}H]TCDD, this corresponds to 0.25 nM TEQ. A typical experimental protocol employs 10-μL aliquots of competitor solution in a final 1 mL of incubation mixture, so the DL is 250 fmol TEQ per incubation mixture.

It is difficult to account for nonspecific binding quantitatively in a competition assay, unlike a saturation assay (see above). The usual approach is to determine nonspecific binding under conditions that no unlabeled competitor is present, and to assume constant nonspecific binding for the entire series of dilutions of the unlabeled competitor.

The competition assay was first applied to the analysis of environmental samples by Hutzinger et al.[74] Municipal fly ash was analyzed by incubating a fixed amount of [^{3}H]TCDD and a fixed amount of rat hepatic cytosol (containing AhR) with increasing amounts of an extract prepared from the fly ash at 0–5 °C. The amount of bound radioactivity was determined by sucrose density gradient centrifugation analysis (SDGC) (see Tsui and Okey;[75] Okey et al.[76]). This method involved treating the incubated mixture with dextran-coated charcoal to remove unbound and loosely bound radioligand, separating an aliquot by SDGC, collecting the fractions, and assaying each for its radioactivity, a very time-consuming procedure. The DL was in the range 100–200 pg of unlabeled TCDD. The authors recognized that the assay ''assesses the total biologic activity of the PCDDs/PCDFs in a sample and does not focus on only one specific congener'', and also that the DL of any such competition assay would be lower the higher is the specific activity of the radioligand. The method was later extended to a range of individual PCDDs and their brominated analogs,[77] and to the analysis of 3,3′,4,4′-tetrachloroazobenzene (TCAB) in a series of chloroaniline-based herbicides.[78] In these last two reports the receptor–ligand binding assay was compared with aryl hydrocarbon hydroxylase (AHH) induction; for PCDD, PCDF, and PCB analytes, the sensitivities of the two assays were similar, but with TCAB the receptor-binding assay was much more sensitive than AHH induction, probably because TCAB is more readily metabolized.

Bradfield and Poland[79] made several modifications to the receptor–ligand binding assay. Their approach differed from that of Hutzinger et al.[74] in the following

respects: (1) the reference radioligand was [^{125}I]-2-iodo-7,8-dibromodibenzo-*p*-dioxin, which can be obtained at much higher specific activity than [^{3}H]TCDD, permitting lower DLs to be achieved (3 pg of unlabeled TCDD); (2) they carried out the assay by removing excess ligands from the incubation mixture with dextran-coated charcoal followed by γ-counting, without using SDGC; (3) they used hepatic cytosol from the C57BL/6 mouse rather than the rat, having found that the mouse receptor is thermally more stable with respect to loss of binding ability;[37] and (4) they used a 16-h incubation at 4 °C. The assay was applied to a series of halogenated dibenzo-*p*-dioxins and dibenzofurans, with the recognition that other DLCs could in principle be analytes for the method. Besides its higher specific activity, 2-[^{125}I]-iodo-7,8-dibromodibenzo-*p*-dioxin is a good choice as the reference radioligand because it binds to the AhR almost as strongly as TCDD [cf. Goldstein and Barrett[66]]. Its disadvantage is its relatively short half-life of 60 days, which necessitates resynthesizing the ligand fairly frequently, compared with a tritiated ligand whose radiochemical half-life is 12.3 years.

The introduction of the hydroxyapatite (HAP) assay[80] greatly improved the convenience of the receptor–ligand binding technique. In this assay, proteins in the incubated mixture are adsorbed on HAP, excess loosely bound ligands are removed using buffer containing a mild detergent (Triton X-100), and the radioactivity is counted after suspending the protein–HAP pellet in scintillation fluid. Using [^{3}H]TCDD as the reference radioligand, the HAP version of the receptor–ligand assay was shown to be applicable to a range of substituted azobenzenes[81] and chlorinated stilbenes[82] in addition to PCDDs, PCDFs, and PCBs. Chlorinated stilbenes were unusual in that both AhR binding affinity and the induction of monooxygenase enzyme activity in rat H4IIE hepatoma cells showed no obvious dependence on ligand structure. The HAP assay is inapplicable, however, to AhRs that bind only weakly to DLCs, for example the human AhR, because the detergent washing step strips the DLCs from the receptor.[83]

Schneider et al.[69] carried out a detailed examination of parameters such as incubation time for a successful AhR ligand binding assay. Using [^{3}H]TCDD as the radioligand, C57BL/6 mouse cytosol as the source of AhR, and 23 °C as the incubation temperature, they showed that mixtures of ligands of widely varying structure behaved additively with respect to receptor binding activity, once their concentrations were expressed in terms of TCDD equivalents. In common with previous workers they found that RBAs varied in the order PCDDs > coplanar PCBs > nonplanar PCBs [but note that low RBA towards the AhR does not absolve nonplanar PCBs of suspicion of non-AhR-mediated toxic effects[84]]. It is important to add the cytosol preparation to the premixed ligands; if the ligands are added sequentially, the first ligand added occupies a disproportionate fraction of AhR binding sites.[67,69] This observation indicates that association between the ligand and the AhR is essentially irreversible at this temperature, signifying that most of the ligand–AhR complex had already undergone transformation by the time it was analyzed, so that RBAs determined under these conditions represent kinetic rather than equilibrium affinities for the AhR.

Hu et al.[17] compared the HAP AhR ligand binding assay and chemical analysis by GC/MS for the analysis of DLCs in environmental matrices, with emphasis on how the results were affected by the extent of sample cleanup. Matrices analyzed included fly ash, fish tissue, soil, pulp-mill sludge, and petrochemical catalyst wash water. Fish tissue and contaminated soil required extensive cleanup before meaningful assay results could be obtained. The bioassay was considered conservative in that the apparent TEQ detected was always as great as (and often much greater than) that found by GC/MS, even after including the contributions of coplanar PCBs. Further analysis by GC/MS identified much of the missing material as PAHs, which also bind to the AhR but which do not evoke dioxin-like toxicity. Although AhR binding assays can be used to screen unknown environmental samples for possible DLC contamination, they cannot provide quantitative information unless interfering analytes have been shown separately to be absent. In appropriate situations they would be useful for monitoring the progress of an environmental remediation program fairly rapidly and inexpensively, always remembering that AhR binding assays do not provide chemical identification of what actually binds the AhR and also that agonists and antagonist both give a positive response. In addition, the observed radiation intensity does not depend linearly upon sample TEQ. It approaches zero asymptotically when the TEQ of the sample greatly exceeds that of the radioligand, and so a series of serially diluted samples must be assayed in order to establish the numerical factor by which the original extract must be diluted in order to reach the EC_{50}.

Rao and Unger[85] suggested using the AhR affinities of different ligands to estimate risk of exposure to mixtures of congeners. Competitive affinities of the DLCs for rat AhR were used to estimate the fraction of each DLC congener in human adipose tissue that would be bound to AhR. Drawbacks to this approach include the need for congener-specific analysis, and assumptions about the equivalence of human and rodent AhR towards different DLCs, equivalence among DLCs with respect to bioavailability, and whether receptor occupancy is a sufficient criterion for toxicity. Among these factors, that

of equivalency of human and rodent AhRs is especially suspect, because they differ substantially in their thermal stability in vitro[72] and their affinity towards TCDD.[24,40]

2.2 Post-transformation Assays

This group of assays includes tests for the production of the transformed AhR–ligand complex and the production of nuclear AhR–ligand complexes, and downstream end-points such as monooxygenase enzyme induction and reporter gene assays in which the reporter gene product is under the control of a DRE. Transformation in vitro is temperature dependent; its rate varies with species (guinea pig > rat > human).[73,86] Transformation is accompanied by an increase in the strength of ligand binding[51] and can be monitored by following the loss of the ligand's ability to dissociate from the complex.[67] Other parameters that are altered by transformation are the change of molecular mass of the AhR resulting from the dissociation of protein subunits[87] and the properties of the dissociated subunits, such as the binding affinity of ALBS–ARNT to DNA[73,88] and the phosphorylating ability of c-src.[35] Transformation is difficult to follow because no simple methods yet exist to demonstrate the production of specifically the transformed or the untransformed TCDD–AhR complex, or to show that one of them has been produced to the exclusion of the other.

Petrulis and Bunce[89] studied the first step beyond ligand binding in experiments in which competing ligands were added sequentially to hepatic cytosol. At low temperature (0–4 °C), transformation is sufficiently slow in some species that the reversibly formed AhR–ligand complex can be distinguished from the transformed AhR in which the ligand is essentially irreversibly bound. Provided that binding occurs at all, even weakly binding ligands such as PCB 153 appear to induce transformation as effectively as TCDD. This was taken as evidence that the antagonism of TCDD-induced end-points by PCB 153 does not occur at the stage of dissociation of the multimeric AhR complex.

2.2.1 Gel Retardation Assay

The gel retardation assay can be used to show binding to a synthetic oligonucleotide containing the DRE consensus sequence. In this method, [^{32}P]adenosine triphosphate (ATP) is used to phosphorylate the oligonucleotide, which is then allowed to bind to a deficit of cytosolic AhR–ligand complex. The bound and unbound oligonucleotides are separated by nondenaturing polyacrylamide gel electrophoresis. Binding to the protein reduces the charge-to-mass ratio of the oligonucleotide, and it runs more slowly through the gel, hence the name gel retardation analysis. The retarded and unretarded bands

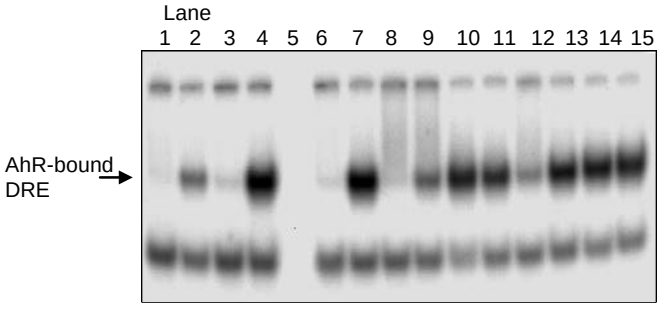

Figure 8 Autoradiogram of a gel shift assay experiment showing the AhR-bound [^{32}P]-DRE bands. Assay used guinea pig hepatic cytosol, treatments at 23 °C unless indicated. Lanes: 1, dimethyl sulfoxide (DMSO) (0 °C); 2, 10 nM TCDD (0 °C); 3, DMSO (30 °C); 4, 10 nM TCDD (30 °C); 5, empty; 6, DMSO; 7, 10 nM TCDD; 8, 100 μM PCB 153; 9, 100 μM PCB 153 + 10 nM TCDD; 10, 100 μM 3-methylcholanthrene (3-MC); 11, 100 μM 3-MC + 10 nM TCDD; 12, 1 nM TCDD; 13, 100 nM 3-MC + 1 nM TCDD; 14, 1 μM 3-MC + 1 nM TCDD; 15, 10 μM 3-MC + 1 nM TCDD.[90]

are visualized using autoradiography (Figure 8). A key advantage of this technique is that it does not require a radiolabeled ligand, as it is the oligonucleotide that is labeled. It can therefore be applied to any AhR-binding compound.

Safe's group studied the production of the nuclear complex by incubating radioligand with AhR and monitoring the accumulation of radioactivity in the nucleus.[91,92] The rates of transformation of the AhR–ligand complexes were parallel when two end-points were compared: translocation to the nucleus and formation of a shifted band in the gel retardation assay.[93] In principle, formation of the nuclear AhR–ligand complex could be used as an environmental assay, although a competition assay would be required to monitor an environmental sample because the latter is unlabeled. As with receptor binding, the experiment would consist of measuring the reduction in bound radioactivity when the environmental sample was incubated with a fixed aliquot of receptor and a fixed aliquot of radioligand.

Santostefano et al.[94] attempted to use the gel retardation assay to examine the relative potencies for AhR activation of different ligands. By the use of Woolf plots, they found that transformed complexes of the rat AhR with different ligands had similar affinities for oligo-DRE. Equilibrium binding between AhR–ligand complex and the oligo-DRE was demonstrated by showing that unlabeled DRE added later could displace the ^{32}P-labeled DRE from the AhR–ligand complexes. Ligands such as 1,2,7,8-tetrachlorodibenzofuran (TCDF), which binds weakly to rat AhR, produced less transformed AhR complex; this was proposed to be due to inactivation of the unliganded AhR competitively with binding. In principle, an assay for environmental samples could be developed

using a fixed amount of AhR, followed by incubation with a fixed amount of DRE to probe the environmental samples' ability to transform the AhR.

Petrulis and Bunce[61] studied the relative transforming ability of AhR ligands both singly and in combination, using cytosolic extracts from rat and guinea pig. Neglecting minor interspecies difference, ligands with high affinity for AhR [TCDD, 3-MC and 3,3′,4,4′-tetrachlorobiphenyl (PCB 77)] all produced comparable intensities of the shifted band upon gel electrophoresis. Weakly binding ligands such as PCB 153 and 2,3,3′,4,4′,5-hexachlorobiphenyl (PCB 156) gave only limited intensities of the shifted band. Combinations of productive ligands (those that exhibit high affinity for AhR and high toxic potency) at saturating concentrations yielded intensities similar to those resulting from a single high-affinity ligand; unproductive ligands at high concentration outcompeted lower concentrations of productive ligands. This behavior shows clearly that the mechanism of action of the unproductive ligands diverges at or before the DRE binding step in the mechanism, and that target molecule antagonism is a plausible explanation for the antagonism of the toxicity of productive ligands by unproductive ligands. At high total receptor occupancy, the ligands compete for a limited amount of the target molecule AhR, whereas at low total receptor occupancy, AhR sites are plentiful for occupancy by both ligands.

The biological relevance of the gel retardation assay may be limited because it employs an oligonucleotide which is only a short DNA fragment, precluding the possibility of reproducing multiple DRE occupancy which may be associated with toxicity in actual organisms. Although binding to native DREs occurs only after the AhR–ligand complex has dimerized with ARNT,[55] in vitro experiments with oligonucleotides are carried out without adding ARNT. It is not clear whether the in vitro system does not require ARNT or whether sufficient ARNT is present in broken-cell preparations of cytosol.

A current point of interest is whether antagonistic ligands that bind AhR fail to undergo transformation. Petrulis and Bunce[61] indicate that unproductive ligands such as PCB 153 induce dissociation of liganded AhR complexes as effectively as TCDD, but bind much less strongly to synthetic oligonucleotide. This phenomenon may also be species-dependent, since α-naphthoflavone–AhR complexes transformed much more efficiently (judged by gel retardation assays) when guinea pig was the source of cytosolic AhR compared with rat.[94,95] The possibility that mechanism-based bioassays may give species-dependent results led Denison et al.[96] to suggest that bioassays should be run in multiple species to assess toxic potency thoroughly, but that is impractical if the goal of bioassays is to achieve high throughput and low cost. The alternative might be to standardize the strain of animal or the cell line that is to be used in the assay, but at the risk of overlooking some of the DLCs present in environmental samples.

Other possible assays that could be devised include one based on detecting the formation of the ALBS–ARNT heterodimer, possibly using an Ab-based method. This would require an Ab that could recognize the ligand-bound ALBS–ARNT complex, rather than the individual components, and may not be practical. An assay could also be imagined in which the formation of the mRNAs following transcriptional activation was monitored, but this does not yet seem to have been attempted.

With respect to cancer as a toxic end-point, Büsser and Lutz[97] investigated an assay based on the ability of hepatocarcinogens, including TCDD, to stimulate unscheduled DNA synthesis. The protocol involved oral administration of the test substances to rats or mice, followed 24 h later by radiolabeled thymidine. Four hours later the animals were killed and the extent of radiothymidine incorporation into hepatic DNA was measured in comparison with untreated controls. The female rat and male mouse (but not male rat) showed increased DNA synthesis at TCDD exposure levels of 1–2 nmol kg^{-1}. The procedure is relatively fast, but a screening program would require large numbers of animals, in addition to which, no information is yet available about the responsiveness of other DLCs in the assay.

2.2.2 *Reporter Gene Assays*

Reporter genes are DNA sequences that encode for proteins or enzymes with characteristic activities that are unique to the cells used for the test and that are easily measured using conventional biochemical assays. As noted by Lewis et al.,[98] reporter genes were first employed for studying the mechanisms of transcriptional activation in cells, but have now become valuable tools in bioassay systems. Techniques are now available for transfecting new genetic material into the host cells so that the change is heritable over the long term (stable transfection), although the selection of the stable transfectants is very time consuming.[99] The alternative, transient transfection, does not provide indefinite stability, but allows more copies of the foreign DNA to be introduced into the host cell; transient transfection must always be used when the host cells are primary cultures, which are inherently viable only for short periods.

An early reporter gene to be developed was chloramphenicol acetyl transferase (CAT), an enzyme from *E. coli* that is foreign to eukaryotic cells. It is assayed by quantitating the acetylation products of the antibiotic chloramphenicol; the method carries an internal check

because only the cells that remain transfected survive in the presence of the antibiotic. The original CAT methods required the use of either radiolabeled chloramphenicol or radiolabeled acetyl coenzyme A in order to assay the enzyme product; more recently, fluorescent chloramphenicol derivatives have been developed to allow fluorescence as the assay method.[98] Many cell lines used for bioassays today are transfected with CAT or another antibiotic-resistance gene, in addition to the reporter gene whose gene product is the primary objective of the method. The antibiotic-resistance gene is present to monitor that the cells, which are grown in a medium supplemented with the antibiotic, retain their viability, and so that contaminant microorganisms will not survive.

In the context of analysis for DLCs, assay methods so far developed involve introducing DNA that encodes for an abnormal response into cells that already contain functional AhR machinery, usually immortal hepatoma cells. The foreign DNA is introduced downstream from the DRE; ligand binding followed by interaction with the DRE causes the translation of the reporter gene product in the bioengineered cells. Reporter gene assays require the DLC to undergo three stages in the mechanism of toxic action (ligand binding, transformation, and DRE binding) in order to observe a positive response; as such, they are not strictly receptor–ligand binding assays.

El-Fouly et al.[100] first reported linking AhR binding to the expression of an easily monitored gene product through the use of recombinant cell lines. They constructed, in AhR-containing mouse hepatoma cell lines, a stable recombinant vector that expressed human placental alkaline phosphatase (AP) under the control of a section of CYP 1A1 promoter that contains several DREs. Exposure to TCDD led to receptor binding, transformation, DRE binding, and dose-dependent expression of AP. The significance of the use of human placental AP is that this particular enzyme is unusually stable to heat. Other native phosphatases can be thermally denatured before monitoring the culture's phosphatase activity by measuring the rate of hydrolysis of *p*-nitrophenyl phosphate (a substrate for the enzyme) to *p*-nitrophenolate anion, which is easily monitored spectroscopically. In the presence of TCDD, AP activity was induced in a saturable, dose-dependent manner ($ED_{50} \approx 0.3$ nM). Other DLCs, including TCDF, 3-MC, PCB 77, and 3,3′,4,4′,5-pentachlorobiphenyl (PCB 126) (but not PCB 153) also induced near-saturating levels of AP activity. The rate of production of *p*-nitrophenolate was taken to be proportional to receptor occupancy by DLCs, although this implies the questionable assumption that all DLCs undergo stages subsequent to AhR binding with equal efficiency.

Several groups have recently introduced elegant variants of the AhR reporter gene assay in which AhR binding has been linked to activation of the LUC gene, which encodes for luciferase, the enzyme that is responsible for light emission in the common firefly. The significance of this technique is that the experimentally observed parameter is light intensity, which can be monitored easily and with high precision using a luminometer. Balaguer et al.[101] transiently transfected the luciferase reporter gene into Hepa 1c1c7 cells so that the luciferase gene was under the control of DREs. Ligand binding, followed by transformation and DRE binding, led to the translation of luciferase gene, leading to the dose-dependent emission of light when luciferin was added. This system was able to assay for TCDD itself, and for TCDD-like activity in black liquor from pulp and paper mills.

Murk et al.[102] obtained stable transfectants of the luciferase reporter gene in rat hepatoma H4IIE cells, and were able to analyze both reference compounds and PCB mixtures by means of this assay, which they termed chemically activated luciferase gene expression (CALUX). The reaction was carried out in 24-well plates, and with TCDD as a reference, the EC_{50} was reached at 10 pM. The DL, defined as three times the standard deviation of the DMSO blank, corresponded to only 0.5 fmol of TCDD per well. The maximum light emission, however, differed among analytes, presumably reflecting differences in the efficiency of events subsequent to AhR binding (transformation and/or nuclear translocation and/or DRE binding). The assay was applicable to both single DLCs and environmental samples, including sediment porewater[102] and blood plasma,[103] and it was suggested that the CALUX assay would provide a rapid and relatively noninvasive test for exposure to DLCs in wild-life. The same group has also developed a trout hepatoma cell line with the luciferase reporter gene; luminescence was measured by lysing the cells after 4 days of exposure to the DLC and adding standard luciferin reagents. TCDD had an EC_{50} for light emission of 0.06 nM (maximum intensity 100%, assumed); 1,2,3,7,8-pentachlorodibenzo-*p*-dioxin, 2,3,4,7,8-pentachlorodibenzofuran, PCB 126 and PCB 77 had EC_{50} (maximum intensity) 0.3 nM (120%), 0.2 nM (114%), 10 nM (110%), and 11 nM (43%), respectively. 2,3,3′,4,4′-pentachlorobiphenyl (PCB 105) and 2,3′,4,4′,5-pentachlorobiphenyl (PCB 118) were unresponsive in the assay.[104]

The CALUX system and the related P450 reporter gene system (P450RGS) have been commercialized by contract analytical laboratories. P450RGS utilizes human HepG2 hepatoma cells modified by the introduction of the luciferase gene. The P450RGS assay screens for a wide variety of planar aromatic xenobiotics,[92] prospectively allowing those environmental samples of the greatest potential concern to be identified for more detailed testing. PAHs are more responsive at short incubation times (6 h) and DLCs at longer times (16 h). The

relative light emission varied significantly among DLCs (TCDD, 27-fold induction; PCB 77, 15-fold; 2,3,4,4′,5-pentachlorobiphenyl (PCB 114), 19-fold; PCB 126, 111-fold; PCB 156, 6-fold; 3,3′,4,4′,5,5′-hexachlorobiphenyl (PCB 169), 29-fold), but it is not clear from the reported data whether these represent saturated responses. TEFs calculated from the P450RGS system were substantially different from the consensus values.[105] Although the P450RGS assay has been applied to the determination of the TEQ of environmental samples, it does not yet appear to have been established that mixtures of DLCs or PAHs respond additively in the assay. Consequently, TEQs obtained in the assay must so far be considered to be only semiquantitative.

The development of reporter gene assays is clearly more demanding than conventional ligand binding assays, but the potential payoff is a simpler assay protocol. The limitation of conventional saturation assays to radiolabeled ligands is overcome by using reporter gene assays, because the assay end-point is the phenomenon under the control of the reporter gene, such as light emission from luciferase. Reporter gene assays are always carried out as saturation assays, because all competitors for the same AhR binding sites act as potential agonists for the reporter gene. The data cited above[92,106] make it clear, however, that the intensity of light emission from different ligands is not necessarily equal even under conditions of saturation of AhR binding sites. This is consistent with unequal transforming or DRE binding ability of different DLCs. Experiments with 2,2′,5,5′-tetrachlorobiphenyl (PCB 52) in the recombinant assay involving firefly luciferase as the reporter gene product[107] indicate that the extent of DRE activation by less productive DLCs is also cell-line dependent.

2.2.3 Enzyme Induction Assays

Several assays have been developed in which the induction of a particular gene product has been monitored in vivo or in cell culture following exposure to TCDD. The most common end-point is the induction of a monooxygenase activity such as AHH or 7-ethoxyresorufin-*O*-deethylase (EROD). These are markers for the CYP 1A1 gene product (cytochrome P450 1A1), which is the principal form of cytochrome P450 induced by TCDD and PAHs such as 3-MC. CYP 1A2 is inducible by nonplanar PCB congeners[108] and also by TCDD;[109] its activity can be monitored by following the marker enzyme acetanilide-4-hydroxylase.[110] In addition, 7-pentoxyresorufin-*O*-deethylase (PROD) is a marker for the phenobarbital-inducible CYP 2B family of cytochrome P450s[111] and 7-ethoxycoumarin-*O*-deethylase (ECOD) is a marker for both phenobarbital- and TCDD-type inducers.[112] All of these enzyme assays have some degree of cross-reactivity among substrates; no substrate is known that has complete specificity for a single monooxygenase isozyme [reviewed by Hu and Bunce[113]]. Unlike the usual method of carrying out enzyme kinetics which employs constant enzyme and variable substrate concentrations, these assays employ a constant concentration of substrate that is optimized so that the reaction rate $V \approx V_{max}$. Under these conditions the rate of product formation is proportional to the enzyme concentration.

The usual protocols for the AHH induction assay[114] involve incubating a preparation of lysed induced cells (and uninduced controls) with a standard solution of benzo[*a*]pyrene and reduced nicotinamide adenine dinucleotide (NADPH) in the presence of oxygen. The reaction involves the hydroxylation of benzo[*a*]pyrene, one of whose products, 3-hydroxybenzo[*a*]pyrene, is monitored by its characteristic fluorescence at 522 nm.[115,116] AHH activity is reported as picomoles of 3-hydroxybenzo[*a*]-pyrene equivalents formed per unit time per milligram of total protein. Studies of AHH activity frequently employ the rat hepatoma cell line H4IIE, which has low basal activity, but whose activity is readily induced by 3-MC[117] and TCDD.[118] The protocol was refined by Sawyer and Safe[119] as a rapid screening tool for the detection of planar halogenated aromatic compounds. The EROD assay is carried out in a similar fashion,[120] in which the lysed cells are incubated with NADPH and 7-ethoxyresorufin, and the reaction is followed by monitoring the fluorescence of the product resorufin at 582 nm. The assay results are reported as picomoles of resorufin formed per unit time per milligram of total protein. The EROD assay is greatly preferred over AHH, because it avoids the use of the highly carcinogenic benzo[*a*]pyrene.

A key advantage of the AHH and EROD assays is that the parameter detected is fluorescent light intensity, which allows the possibility of low DLs, automation and easy data handling. The protocol for the analysis of unknowns involves preparing a calibration curve of fluorescence intensity vs TCDD concentration and reading the TCDD equivalent of the unknown directly from the curve. The introduction of fluorescence plate readers has permitted the EROD assay to be carried out in 48- or 96-well plates.[121]

Both AHH and EROD assays can be carried out using animal tissue or, more commonly, with cells in culture. When whole animals are used, the CYP 1A activity of microsomal preparations of the livers of killed animals is compared with that of hepatic microsomes from control (unexposed) animals. Examples include the work of Parrott et al.[122] on the accumulation of PCDDs in rainbow trout and of Diliberto et al.[123] on tissue distributions of TCDD in mice. Machala et al.[124] used chick hepatocytes for EROD analysis, with injection of

test substances on day 14 of incubation and killing and removal of the livers 24 h later. The assay was applied to eggs treated with individual DLCs but not to mixtures or environmental samples.

The rat hepatoma cell line H4IIE has been extensively used as a convenient standard procedure for assaying environmental samples using cell cultures. Many other hepatoma lines, from humans, rodents, and fish, have also been developed. Primary cultures of hepatocytes are also used for EROD analysis [e.g. Kennedy et al.;[121] Bosveld et al.;[125] Dubois et al.[126]]. Primary hepatocytes are readily prepared, allow the examination of any species of interest, and, it has been argued, are more representative of the metabolism of the normal animal than immortal cell lines.

Because enzyme assay activities are reported in units such as femtomoles per minute per milligram of protein, the total protein content of the biological preparation must be measured. This is usually done using either the method of Lowry et al.[127] or Bradford;[128] more recently, Lorenzen and Kennedy[129] introduced a method based on fluorescamine association that can be carried out by fluorescence assay in the same 96-well plates that are used for EROD analysis.

TEQ values are estimated from enzyme induction experiments by preparing a calibration curve of enzyme activity vs dose of TCDD incubated with the cells. The ability of the environmental sample to induce AHH or EROD is then compared with the dose of TCDD needed to induce the enzyme to the same extent. With respect to DL, the EROD assay is able to detect low concentrations (typically $<10\,\text{pmol}\,\text{L}^{-1}$), because fluorescence is the experimentally observed parameter.

Monooxygenase enzyme induction is regarded as a normal, adaptive metabolic response to exposure to AhR-binding xenobiotics rather than a toxic response[130,131] [although recent work suggests that the induction of CYP 1A1 by TCDD causes oxidative damage to DNA[132]]. The justification for using enzyme assays to estimate TEQs is the well-established rank-order correlation among DLCs between the doses required for enzyme induction in cell culture and in vivo effects such as thymic atrophy and body weight loss in whole animals.[47,133] In common with overtly toxic end-points, enzyme induction requires the toxicant to initiate all of the biochemical steps from receptor binding through to DRE binding and mRNA transcription. Clemons et al.[134] have shown a good correlation in wild fish between TEQs determined by GC/MS and those obtained by EROD assay using either H4IIE cells or a rainbow trout cell line designated RTL-W1. This represents an important validation of the enzyme induction assay with real samples (see Figure 9a and b).

Considerable effort has recently been invested in comparing different monooxygenase enzyme assays because

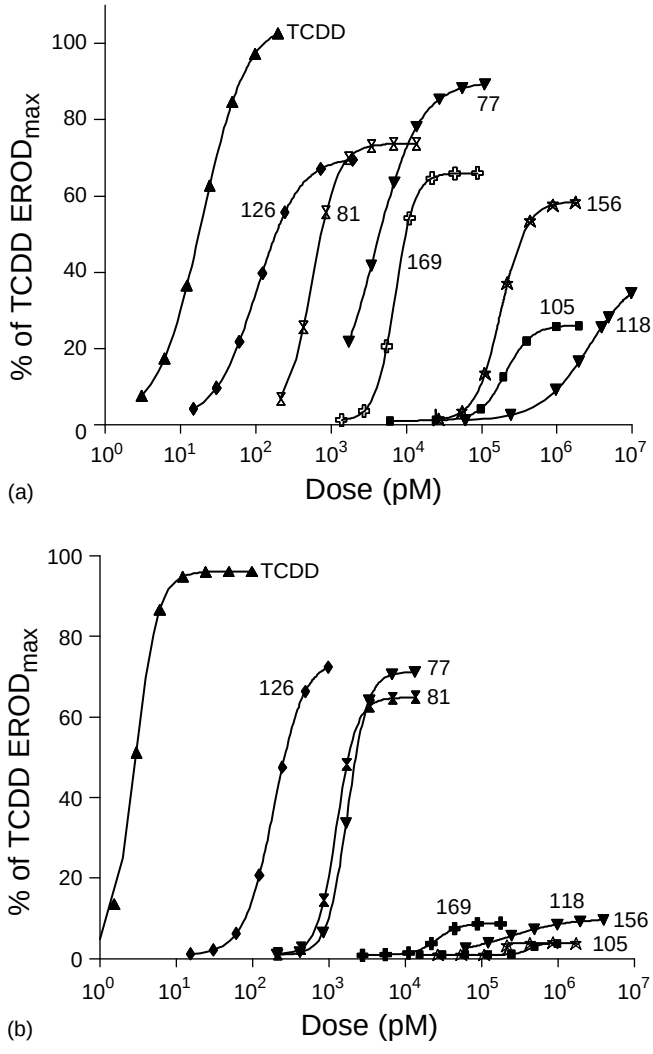

Figure 9 EROD induction by four nonortho coplanar PCBs (PCB 77, PCB 81, PCB 126, and PCB 169) and three mono-ortho coplanar PCBs (PCB 105, PCB 118, and PCB 156) in (a) H4IIE and (b) RTL-W1 ($n \geq 3$). [Reproduced by permission from J.H. Clemons, L.E.J. Lee, C.R. Myers, D.G. Dixon, N.C. Bols, 'Bioassays for the Analysis of Dioxin-like Compounds', *Can. J. Fish. Aquat. Sci.*, **53**, 1177–1185 (1996).[134]]

of their important application of screening industrial effluents, particularly in the context of the toxicity of pulp-mill effluents towards fish in their receiving waters.[135–137] This effort followed the observation in the 1980s that chlorine-bleaching mills emitted small quantities of PCDDs and PCDFs. The toxicity and monooxygenase enzyme activity of pulp-mill effluents did not correlate with DLC concentration: both responses were also observed downstream of mills that used no chlorine. These observations reinforce the point that monooxygenase enzyme activity can be induced by many AhR-binding compounds, not all of which are DLCs, and not all of which are necessarily toxic.

Sanderson et al.[138] compared the DL obtained from EROD analysis using H4IIE cells with the value obtained based on light emission from H4IIE cells that had been genetically modified to contain the luciferase reporter gene. The latter technique gave a threefold lower DL on exposure to TCDD (0.8 pmol L^{-1}). Hodson et al.[137] compared the work of several groups with regard to DLs and ranges of quantitation in monooxygenase assays. Most of the studies reviewed report the induced EROD activity vs log[TCDD], and so "ranges of linear response" actually refer to the almost linear central portion of the sigmoidal dose–response curves [see Figure 7a]. This accounts for the similarity between the reported EC_{50} values and the DLs of the assay. For the commonly used rat hepatoma H4IIE cell line, typical results are a log-linear range about 10–200 pM TCDD, with an EC_{50} of ~20 pM TCDD and a DL of ~10 pM TCDD.[138] Primary chicken embryo hepatocytes show a log-linear range of about 1–100 pM TCDD, $EC_{50} \approx 20$ pM TCDD, and DLs of ~1 pM TCDD.[124] In the rainbow trout liver cell line RTL-W1, a log-linear range of about 2–20 pM TCDD, $EC_{50} \approx 15$ pM TCDD, and DLs of 2 pM TCDD were reported.[134] Luciferase expression in reporter-gene assays in HepaG2 and Hepa1c1c7 cells typically demonstrated a wider log-linear range of about 10–1000 pM, but the EC_{50} values and DLs were increased to approximately 100 pM TCDD and 50 pM TCDD, respectively.

Parenthetically, Hodson et al.[137] also reviewed the use of passive devices based on semipermeable membranes for preconcentrating natural and effluent water samples in order to increase the concentration of DLCs prior to EROD analysis. By the choice of triolein inside the membrane, the deployment device is able to mimic, to some extent, the behavior of an aquatic organism in bioconcentrating these xenobiotics from water.

A limitation to the use of monooxygenase enzyme assays is that their activity may be species-dependent. Dubois et al.[126] observed different patterns of dose vs EROD activity in quail hepatocytes, rat hepatocytes, and the human hepatoma line HepG2 when comparing enzyme induction by PCB 77 and the commercial PCB mixture Aroclor 1254. In experiments with rats, DeVito et al.[110] found that the relative EROD induction potencies of different DLCs varied with the tissue tested, and they cautioned against the universal applicability of hepatic EROD induction potencies.

The curves of fluorescence intensity vs concentration for different inducers are displaced along the abscissa (Figure 9a and b), as would be expected for compounds with different TEFs, and so the abscissa of the curve for a mixture represents TEQ. EROD activity is a saturable response, because beyond a certain concentration of DLC all AhR sites have been occupied and CYP 1A1 production has been maximized. There are, however, two consistent findings that complicate the interpretation of data obtained from EROD assays: (1) different DLCs yield dissimilar values of the maximum (saturable) fluorescence intensity; and (2) for many DLCs, EROD activity declines when the concentration of the inducer exceeds the saturating value (Figure 10). This phenomenon has been the subject of some controversy.[139] Recently it has been shown that DLCs not only induce CYP 1A1, but also act as competitive inhibitors of the deethylation of the enzyme substrate, 7-ethoxyresorufin.[61,140] This leads to biphasic curves when cells in culture are treated with increasing amounts of the DLC: the EROD activity increases at low [DLC], as more enzyme is induced, but falls again once the induction response (which is AhR

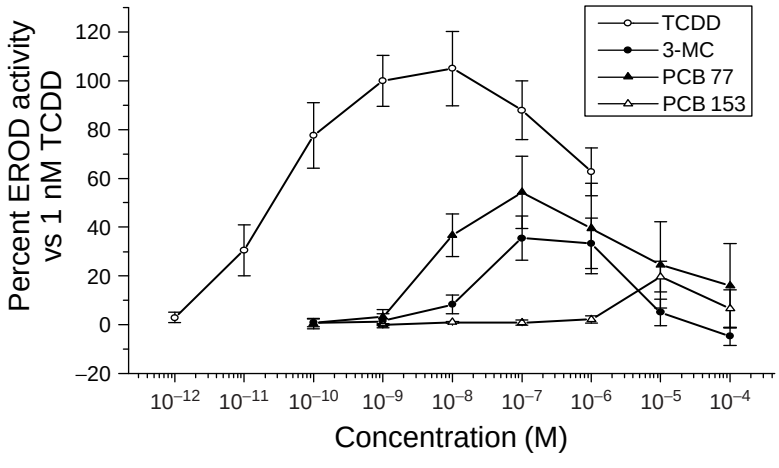

Figure 10 EROD activity in primary rat hepatocytes incubated for 24 h with TCDD, 3-MC, PCB 77, or PCB 153. Activities are relative to the activity obtained following treatment with 1 nM TCDD. Error bars represent the standard deviation of the calculated EROD activities.[89] (Redrawn from J. Petrulis, PhD Thesis, University of Guelph, 1998.)

mediated) has become saturated, because additional DLC inhibits the enzyme. The quantitative EROD response becomes further complicated when mixtures of inducing agents are present or when the inducer is metabolized concurrently with enzyme induction.[61] This has also been implied by experiments in which the length of exposure of cells to the inducing agents is varied: the apparent EC_{50}s of some DLCs increased at longer exposure times, on account of differential rates of metabolism.[141] Furthermore, Aarts et al.[95] disclosed that the extent of antagonism between TCDD and either PCB 77 or 2,2′,3,3′,4,4′-hexachlorobiphenyl (PCB 128) was species and cell-line dependent in luciferase reporter gene assays. For all these reasons, caution must be employed when using EROD and related enzyme assays in the mode of inducing enzyme activity by incubating cells with an environmental extract.

Competitive inhibition does not apply to the use of monooxygenase assays for assaying the enzyme activity of hepatic microsomes, when detecting prior exposure of wild-life to DLCs. In that application, there is no carryover of the DLCs into the microsomal preparation, and so competitive inhibition of the enzyme does not occur.

Another potential problem is the antagonism, or masking, of the effects of one DLC by another. This has been observed in the EROD assay by Hasspieler et al.,[142] who tested coplanar PCBs together with TCDD as inducers of EROD activity in the human hepatoma cell line HepG2. PCB 77 and PCB 169 had similar behavior: neither induced EROD activity alone, but each inhibited EROD induction by TCDD. PCB 126 was a weak EROD inducer on its own; it weakly antagonized induction by a high dose of TCDD, but had little effect at lower doses of TCDD. Similarly, Janz and Metcalfe[143] found less than additive induction of AHH in intact rainbow trout treated with TCDD and PCB 77. In a 13-week, low-dose feeding study in rats, Van Birgelen et al.[144] found evidence for nonadditive behavior between TCDD and PCB 156 in several end-points. These observations are consistent with target molecule antagonism, when high concentrations of DLCs compete for a limited stock of AhR,[67] or with competitive enzyme inhibition. The combination of target molecule antagonism and competitive enzyme inhibition makes it clear that enzyme activity assays for CYP 1A1 may underestimate the TEQ of a complex environmental mixture.

Both the foregoing effects come into play in the interpretation of the study by Bosveld et al.,[125] who studied the effects of the interaction between TCDD and the nonplanar PCB 153 on EROD activity in chicken embryo hepatocytes. The ratio [TCDD] : [PCB] was kept constant at $1 : 10^4$ for every dose of the toxicants, reflecting their typical ratio in environmental samples. EROD activity was found to peak at [TCDD] $\approx$ 1 nM, followed by a fall in activity as the TCDD concentration was increased further. Viability assays on the hepatocytes demonstrated that the decrease in activity was not due to cytotoxicity at the high doses of TCDD.

2.3 Markers of Abnormal Metabolism

A well-studied end-point following DLC exposure is the hyperaccumulation of hepatic uroporphyrin III (UP-III). As in all cases of abnormal metabolism, the question arises as to whether the end-point observed can be more than a marker of exposure, i.e. does it really parallel toxicity?

The overproduction of UP-III in the liver, and the related human disease porphyria cutanea tarda, are observed in animals following exposure to many halogenated aromatic compounds. The link to an AhR-mediated pathway was established by showing that inhibitors of cytochrome P450 1A1 suppressed the overproduction of UP-III [reviewed by Bonkovsky et al.[145]] and that UP-III accumulation paralleled AHH activity in AhR-responsive and -nonresponsive mice.[146]

Later work revealed that the link between porphyrin metabolism and the induction of cytochrome P450 enzymes is an indirect one. UP-III is principally formed from uroporphyrinogen III by the action of cytochrome P450 1A2; when cytochrome P450 1A2 is induced [e.g. by nonplanar PCBs such as PCB 153[147]], uroporphyrinogen III is diverted from its normal route in heme biosynthesis. Knockout mice lacking CYP 1A2 activity had less uroporphyrinogen oxidation activity than wild-type mice;[148] in human hepatoma cells, CYP 1A1 and CYP 3A4 also contributed to uroporphyrinogen oxidation activity.[148] Normally, heme suppresses the formation of uroporphyrinogen III through feedback inhibition, thus the

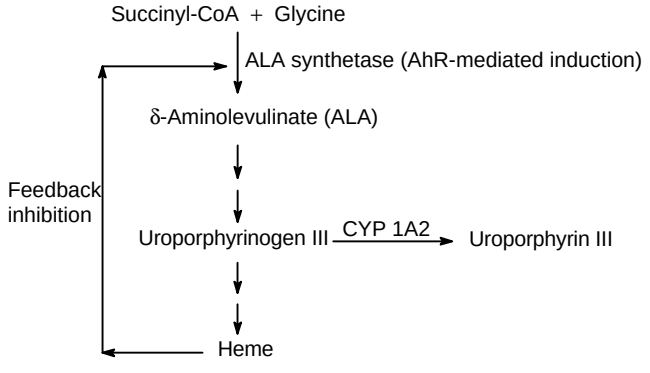

Figure 11 The pathway of porphyrin metabolism, showing the proposed mechanism of hepatic porphyrin accumulation. (Redrawn from A.P.J.M. van Birgelen, J. Van der Kolk, K.M. Fase, I. Bol, H. Poiger, M. Van den Berg, A. Brouwer, *Toxicol. Appl. Pharmacol.*, **126**, 202–213 (1994)[144].)

diversion of uroporphyrinogen III causes a build-up of UP-III (Figure 11).

Studies on rats dosed with mixtures of TCDD and PCB 153 showed that the effect of the two compounds was highly synergistic. Neither TCDD alone nor PCB 153 alone promoted significant accumulation of UP-III, which most likely involves the induction of both CYP 1A1 and 1A2. In addition to monooxygenase induction,[146] TCDD induces δ-aminolevulinic acid synthetase,[150] which is required for the further synthesis of uroporphyrinogen III.[151] The 1A2 isozyme is induced by the nonplanar PCB, thereby promoting the oxidation of uroporphyrinogen III to UP-III. Although porphyrin accumulation can be regarded as an AhR-mediated process, the interactive effects among DLCs are clearly complex, and not completely understood.

The accumulation of hepatic porphyrin in chicken embryos[151,152] has been suggested as a useful marker for the exposure of avian species to DLCs. Individual DLC congeners had broadly similar relative potencies with respect to EROD induction and porphyrin accumulation in primary hepatocytes from chicken and pheasant.[153] Interactive effects between TCDD and several PCB congeners have been studied in primary chicken hepatocytes.[154] For both planar and nonplanar PCBs it appeared that significant porphyrin accumulation only occurred at doses higher than those at which EROD induction was maximum (where the occupancy of the AhR is also maximum). One of the combinations examined by Tysklind et al.[154] was TCDD + PCB 153. At a fixed concentration of TCDD that caused maximum EROD induction but no porphyrin accumulation, increasing levels of the PCB depressed EROD activity and led to porphyrin build-up. We can explain the depression of EROD activity (primarily CYP 1A1) in terms of target molecule antagonism (PCB outcompeting TCDD from AhR binding sites) and competitive enzyme inhibition; this must occur at a PCB concentration that does not induce maximum CYP 1A2. These interactive effects among DLCs make it difficult to calibrate porphyrin accumulation as a quantitative assay for evaluating the toxic potency of environmental samples. Complicating the matter further is the observation that the extent of porphyrin production in the liver following exposure to PCBs is apparently related to the concentration of iron in the liver,[155] an additional parameter that would need to be addressed in the development of a quantitative assay.

3 IMMUNOASSAYS

Immunoassays depend upon the recognition of an antigen (or foreign molecule) by an Ab, which is a soluble protein that has been produced by B-lymphocytes in a live animal. Exposure to a particular antigen causes the proliferation of the B-cell clone that produces the antigen-specific Ab, such that an immune response can be mounted in response to a subsequent exposure to the antigen. Immunoassays are based on exploiting the high affinity of an Ab for the antigen to which it was raised. The actual assay requires both the recognition and quantitation of free analyte in a test solution; the latter is usually accomplished using a competition assay, although a wide variety of assay methods has been described [reviewed by Sherry[156]].

Immunoassays can be used to detect and quantify large biomolecules such as specific proteins or parts of proteins, polysaccharides, or nucleic acids at low concentrations, e.g. <10 fmol of the N-terminal region of the AhR itself.[157] They have become standard methods of analysis in clinical laboratories, where they have been adapted for use in the rapid, automated analysis of large numbers of samples.

Antibodies are normally produced in response to molecules having molecular mass >1000. Small molecules such as TCDD therefore do not elicit an immune response on their own, and so the antibodies must be raised by covalently binding the small molecule (termed a hapten) to a carrier protein that can produce a response and then injecting the conjugate into a live animal. For TCDD analysis, an extensively used hapten is 1-(*N*-adipylamino)-3,7,8-trichlorodibenzo-*p*-dioxin, which is a TCDD molecule one of whose chlorine substituents is replaced by a linking moiety that connects it covalently to the carrier protein (Figure 12).[158,159] If such a protein–hapten conjugate is injected into a live animal, some of the antibodies produced may subsequently be able to recognize the hapten–protein conjugate while others may recognize only the unattached hapten.

Antibodies produced as just described are termed polyclonal antibodies: they are produced by multiple clones of cells, each of which contributes one type of Ab. The antibodies bind to the antigen by way of many different specific recognition elements or epitopes that are functions of the size, shape, and hydrophobicity of different regions of the protein–hapten conjugate. Most of the polyclonal antibodies will interact with epitopes on the carrier protein itself, rather than recognizing the hapten. Even if the hapten is part of the recognition site,

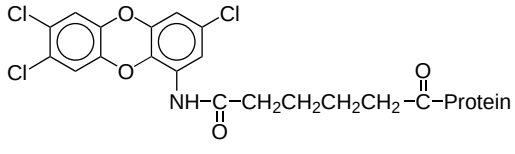

Figure 12 Structure of the protein conjugate of 1-(*N*-adipylamino)-3,7,8-trichlorodibenzo-*p*-dioxin.

it is unlikely that an Ab will recognize the hapten alone, rather than the hapten plus part of the carrier, unless the hapten has a molecular mass above 1000.

Polyclonal Ab assays for DLCs were reported around 1980 [for TCDD, Albro et al.;[159] for TCDF, Luster et al.[160]]. In these assays the DLC molecule was bound covalently to the carrier protein by an *N*-adipylamido chain, and the antibodies were produced in rabbit. The assay was carried out by a radioimmunological method in which the unlabeled DLC and the appropriate *N*-adipylaminotrichlorodibenzofuran or -dibenzo-*p*-dioxin, labeled with ^{125}I in the adipyl chain, competed for the Ab. The complexed antibodies were precipitated with goat anti-rabbit γ-globulin and γ-counted for ^{125}I. The greater the concentration of unlabeled DLC, or in the case of cross-reactivity experiments the higher its affinity for the Ab, the less radioactivity was detected in the pellet. For pure TCDD, as little as 0.08 pmol could be detected by this method. Competitive radioimmunoassays suffer from the same disadvantages as competitive ligand binding assays: it is the labeled competitor rather than the analyte that is actually detected, and the response is saturable, with the signal (bound radioactivity) declining to zero as the concentration of the analyte increases.

A general problem with polyclonal antibodies is poor recognition of the hapten and interference by those antibodies that recognize different structural motifs. In addition, the particular range of antibodies formed varies with time in a given preparation, and from animal to animal among different preparations, making it impossible to standardize methods between different laboratories. These limitations can be avoided by the use of a monoclonal Ab, which is a unique protein derived from a single cloned cell line, and which can recognize the analyte specifically. Rather than attempt to separate a single protein from the polyclonal Ab mixture, a more successful approach is to separate the individual cells that produce the specific monoclonal Ab. This is done by fusing cells from an immortal cell line to spleen cells that have been previously exposed to the protein-linked hapten. The resulting hybridomas are then screened to select a particular cell that secretes an Ab that recognizes the hapten. With luck, a hybridoma can be selected with the ability to form antibodies that recognize specifically the free ligand; this cell line can then be grown up indefinitely in culture.

The monoclonal Ab is uniform and can be distributed to different laboratories, allowing the development of standardized assay procedures. Among the advantages of an optimally selected monoclonal Ab are ease of use, minimal sample preparation, low cost per sample, reproducibility, high selectivity for the analyte, small sample size, and potential for both automation and use in the field.[155,156,161] Disadvantages include the long development time and high cost of obtaining the monoclonal Ab, including hapten synthesis and screening thousands of hybridomas, and problems of cross-reactivity from related and unrelated analytes. The latter problem means that independent (chemical) confirmation of the analyte responsible for the response is necessary. In addition, a single immunoassay cannot respond to analytes of different structural types as can a chemical assay such as GC/MS.

In the case of TCDD analysis, monoclonal antibodies were first developed by Kennel et al.[162] and Stanker et al.[158] Kennel et al.[162] used the same assay approach as Albro et al.,[159] namely radioimmunoassay in which the DLC competes with a ^{125}I-labeled reference ligand. The hybridomas were obtained by the fusion of murine spleen cells with murine myeloma cells, and the hybridomas were screened for their ability to recognize TCDD by identifying those that could bind to bovine serum albumin that was covalently coupled to trichlorodibenzo-*p*-dioxin. As stated by Gilman et al.,[163] ''(The) ideal patterns for PCDD analysis are either high selectivity for only one specific PCDD congener or a pattern of crossreactivity for a wide range of PCDD-like compounds that mimics (their) relative toxicity''. These authors reported the development of immunoassays based on haptens that are structurally closer to TCDD than the original adipylaminotrichlorodibenzo-*p*-dioxin used by Stanker et al.,[158] in an attempt to improve specificity. In recent work, this group reported haptens based on 2-alkenyl-3,7,8-trichloro- (or -7,8-dichloro)-dibenzo-*p*-dioxin that have a working range of 2–240 pg per well for the TCDD surrogate 2-methyl-3,7,8-trichlorodibenzo-*p*-dioxin.[164]

Stanker et al.[158] developed an ELISA for PCDDs. In an ELISA, an enzyme is covalently linked to the monoclonal Ab, and the final assay method involves detection of the product of the enzyme–substrate reaction by some convenient chemical or physical method of analysis. A standard aliquot of the hapten or a simple hapten derivative is initially adsorbed to a plastic surface, such as those on a 96-well plate. The Ab and the solution to be analyzed are placed into the well, where the Ab complexes competitively to the plastic-bound hapten and the free analyte in solution. After equilibration, the solution phase is removed. The plastic-bound Ab complex is quantitated by adding the substrate for the Ab-linked enzyme; the enzyme product is detected by a standard method such as fluorescence or colorimetry, and so the concentration of enzyme product is a measure of the adsorbed Ab–hapten complex. In this type of competition assay the maximum signal is obtained when the solution being tested contains zero hapten; the greater the concentration of hapten in the test solution, the smaller is the concentration of enzyme product detected.

An alternative protocol involves immobilizing the Ab on plastic in 96-well plates. This is incubated with the test DLC mixture and a fixed aliquot of a competitor–enzyme conjugate. After removing the solution phase, the enzyme substrate is added and the amount of enzyme product is determined by how much competitor–enzyme conjugate is bound to the Ab (again, the less DLC in the sample, the more enzyme product is formed). DLs in the range of 0.1 ng of TCDD have been reported,[165] using a mouse monoclonal Ab having specificity for tetra- and pentachlorinated dibenzo-*p*-dioxins that roughly parallels their I-TEFs. A similar approach was used to prepare monoclonal antibodies that recognized coplanar PCBs in the presence of larger amounts of the nonplanar congeners,[166,167] using as the hapten a 4-alkoxy-3,3′,4′-trichlorobiphenyl linked to the protein through a terminal carboxyl on the 4-alkoxy substituent. However, various ''PCB fragments'' (i.e. substituted dichlorobenzenes) were more reactive with the Ab than PCB 77 which severely limited the usefulness of this approach. A recently reported electrochemical immunosensor is claimed to overcome this problem;[168] the Ab is immobilized on an electrode surface by entrapment in an electrically conducting polypyrrole matrix. The signal in this case was the current produced in response to a pulsed voltage waveform on the sensing electrode. Linear calibration curves for Aroclor mixtures were obtained up to (at least) $100\,ng\,L^{-1}$, and DLs of $0.4–3\,ng\,L^{-1}$, depending on the Aroclor mixture, with low cross-reactivity with non-PCB chlorinated compounds.

In principle, monoclonal antibodies could be developed as probes to recognize individual DLCs, by screening appropriate hybridomas. In practice, this has not been very successful because of problems of cross-reactivity.[158] The recognition sites on monoclonal antibodies are fairly large (~2–3 nM in diameter) compared with the size of the analyte of interest (0.3×1.0 nM in the case of TCDD), and so other molecules of similar size and hydrophobicity can be accommodated almost equally well.[169] This makes the monoclonal Ab assay more applicable to DLCs as a class than to the quantitation of individual DLC congeners. A calibration curve can be constructed using test solutions containing known concentrations of TCDD, in which case the assay measures the total concentration of DLCs as TCDD equivalents.

An ongoing criticism of Ab methods for assaying DLCs is their higher DLs compared with GC/MS. The group at Cape Technologies has successfully validated their enzyme immunoassay down to the picograms per gram level for both fly ash[170] and soil[171] samples, with excellent linearity between enzyme immunoassay and GC/MS methodologies. Sugawara et al.[172] have recently introduced a polyclonal ELISA for TCDD with a DL of <1 pg per well, based on horseradish peroxidase activity as the enzymatic detection system. Hapten–protein conjugates were based on 2,3-dichloro-7-(3′-propenoic acid) derivatives. A point of interest was the use of 2-methyl-3,7,8-trichlorodibenzodioxin as a less toxic surrogate for TCDD; the two compounds behaved almost identically in the assay.

To a limited extent, an assay for DLCs based on monoclonal antibodies resembles the competitive AhR ligand binding assay (section 2.1.2). In each case a protein is used to complex the DLC, and a reference ligand (in this case the immobilized hapten) and the DLCs in the test solution compete for a limited amount of protein. Because the monoclonal Ab recognizes the ligand, its binding site is a model for the binding site of the AhR: how accurate a model depends on the similarities between the binding sites, which is a matter of chance. Cross-reactivity between DLCs is observed because each DLC in the sample binds to the Ab according to its relative affinity and its concentration. Therefore, the TEQs obtained by immunoassay may differ from those from the receptor binding assay, because of different relative affinities of DLCs for the two proteins. The immunoassay is not mechanism based because the DLCs do not interact with the Ab as part of the biological mechanism of action.

Like the AhR binding assay, immunoassays tend to be conservative in recognizing more analytes than are toxicologically relevant. This has already been noted in the case of recognition of ''PCB fragments'',[166] and was also observed by Stanker et al.[158] and Vanderlaan et al.,[169] who compared immunoassays of several environmental matrices to analyses by GC/MS. Immunoassays consistently showed more of the ''toxic'' PCDDs and PCDFs than GC/MS, and in some cases the discrepancy was substantial.

Another aspect of DLC toxicology that has been exploited as a bioassay is immunosuppression. Spleen cells from a test species are combined with sheep red blood cells (SRBCs) and DLCs. After 5 days, the spleen cells are resuspended and plaqued. The number of plaques in the red blood cell lawn are counted and reported relative to the number of viable cells. DLCs have been demonstrated to cause a concentration-dependent suppression of the splenic plaque-forming cell (SPFC) response to SRBCs.[173] The utility of this technique as a screening assay is questionable as recent studies have shown that the response to mixtures of DLCs is not additive. The assay is very sensitive to TCDD, causing significant suppression at doses as low as $1\,\mu g\,kg^{-1}$, but PCB 153 has actually been found to enhance the response.[174]

4 EARLY LIFE STAGE BIOASSAYS

Interest in this research area has recently been high because of the recognition that the AhR may be necessary for proper development in a wide range of tissues.[175] In addition, AhR binding substances may alter the actions of endogenous hormones indirectly; for example, exposure to TCDD increases the rate of metabolism of estradiol, presumably by upregulating CYP 1A enzymes.[176] In utero exposure to DLCs causes clear reproductive abnormalities in rodents, and there is limited evidence for such effects also in humans.[177] Willey et al.[178] have shown that the AhR is transcriptionally activated during development in tissues known to be adversely affected by exposure to DLCs. To date, no endogenous ligand for activating AhR during development has been identified.[175]

TCDD shows dose-dependent AhR-mediated teratogenicity in mice, with the most prominent effect being cleft palate.[179–182] The dams are exposed to TCDD around days 7–10 of gestation and killed on day 18 for examination of the fetuses. Besides being inconveniently time consuming and requiring large numbers of animals, this end-point shows marked interactive effects between PCBs and TCDD. Certain nonplanar PCB congeners are able to antagonize TCDD-induced teratogenicity when administered at doses in the range of 10^4 times the teratogenic dose of TCDD, while planar PCB congeners are themselves teratogenic.[179,181,183] The rationalization of these observations is that PCBs and TCDD compete for a finite quantity of AhR. The complexes formed between the AhR and TCDD, and between the AhR and the toxic PCBs go on to cause the teratogenic response, whereas the complexes formed with the protective PCBs do not. When the total receptor occupancy is high, a nontoxic or antagonistic substance can protect against the effects of a toxic substance by binding unproductively to the AhR, thus in a sense displacing the toxic ligand.[67] This effect is not seen at low total receptor occupancy when there are sufficient receptor molecules available to bind all ligands.

Recent interest in environmental estrogens, and the more general phenomenon of endocrine disruption,[184] have focused attention on the role of TCDD (and potentially other DLCs) in developmental and reproductive toxicology. Although the mechanisms of action are not yet clear, it is already apparent that TCDD has strong interactive effects with steroid hormones (although there is no cross-reactivity between TCDD and steroid hormone receptors). In one recent paper, TCDD was shown to block the binding of the estradiol–estrogen receptor complex to estrogen response elements on DNA, while conversely estradiol blocked the association of the TCDD–AhR complex to DREs.[185] TCDD and other DLCs, notably 6-methyl-3,7,8-trichlorodibenzofuran, inhibit the growth of the MCF-7 human breast cancer line.[186,187] TCDD is also developmentally toxic to laboratory rats, causing demasculinization of male pups when administered at sublethal doses to the dams during gestation.[188] These effects, which persist into maturity, may be due to TCDD-induced elevation of the levels of mixed-function oxidase (MFO), thereby enhancing the metabolism of steroid hormones, but this has not yet been demonstrated experimentally. These recent observations have led to interest in developing assays to test for adverse effects of DLCs in early stages of life.

The interest in developmental toxicity is not really new, because some of the earliest concerns of TCDD exposure were "toxic fat syndrome" in young chickens in the 1960s, and the debate in the 1970s over whether forest spraying programs could cause human birth defects and miscarriages. Both of these situations involved the use of the herbicide 2,4,5-trichlorophenoxyacetic acid (2,4,5-T), which contained parts per million levels of TCDD as a contaminant introduced during manufacture.[189]

In the case of toxic fat syndrome, a characteristic finding was the development of pericardial edema in newborn chicks, whose early death caused substantial economic loss in the broiler chicken industry. Higginbotham et al.[190] suggested this as a bioassay end-point. In related work, Flick et al.[191] showed a dose-dependent reduction in hatchability when eggs were treated with TCDD. Rifkind et al.[152] correlated histopathological changes including hepatic edema in chick embryos with the induction of AHH and EROD activity, and Brunström[192] correlated hatchability with EROD induction, in each case after dosing eggs with various PCB congeners.

Helder[193,194] demonstrated the toxicity of TCDD towards fish and their embryos, and developed an assay based on the lethality of DLCs towards yolk sac fry of *Salmo gairdneri*.[195] Similar assays have been introduced for rainbow trout, fathead minnow, and Japanese medaka. In more recent studies, the toxicants have been injected into the developing fish, rather than exposure through the water, to control the dose more precisely [for the technique, see Metcalfe et al.;[196] Walker et al.[197]]. Zabel et al.[198] found that pairs of PCDD and PCB congeners behaved nearly additively (no deviations greater than 2–3-fold) in an early life stage assay in rainbow trout. The same group assessed a complex PCDD–PCDF–PCB mixture in lake trout as having close to the additive lethality of its components.[199] LD_{50}s were in the range of 200 pg of TEQ per egg. A disadvantage of these protocols is the relatively long assay time (several weeks).

The Japanese medaka (*Oryzias latipes*) is a small fish that is convenient for laboratory assays because it is easily reared in captivity, it can be induced to lay eggs in the laboratory all year round, and its larval envelope is transparent, permitting easy evaluation of developmental toxicity in addition to mortality. Wisk and Cooper[200]

studied the survival of Japanese medaka embryos to 3 days post-hatch, after exposing the embryos to the test chemicals within 1–2 h after fertilization. The LC_{50} for TCDD was 13 ng L^{-1} (40 pM) nominal concentration and, as in mammalian systems, the congeners that were fully chlorinated in the 2-,3-,7-, and 8-positions were substantially more toxic. An AhR-mediated mechanism of toxicity is apparent in that neither lethality nor morphological change is seen until after the formation of the liver,[201] and lethality occurs in the same concentration range as AHH induction.[202] In addition to synthetic DLC mixtures, Harris et al.[203] have demonstrated that lipophilic extracts from rainbow trout taken from Lake Ontario are embryotoxic to the Japanese medaka.

The effect of DLCs on the immune systems of intact animals has been known for many years; a dose-dependent reduction in thymic weight in rodents is observed. This was shown to be an AhR-mediated response by correlating the immunotoxic activities of different DLCs with their ability to induce monooxygenase enzymes.[204,205] Numerous interactive effects on these end-points involving TCDD and PCB congeners have been reported,[183,205–208] yielding evidence that certain PCB congeners can mask the activity of TCDD. Smialowicz et al.[209] treated mice with TCDD and PCB 153, alone or in combination, and examined both thymus weights and the ability of their spleen cells to form plaques upon treatment with SRBCs. PCB 153 alone had little effect on either end-point, but depending on its dose it could either enhance or suppress the responses of TCDD. Such complex interactive effects make it unlikely that a broadly useful bioassay for DLCs could be developed, using these end-points.

5 CELL PROLIFERATION ASSAYS

Knutson and Poland[210] developed an assay system for TCDD based on culturing the XB keratinocyte cell line, which is derived from a mouse teratoma. When cultured at low cell density in the presence of lethally irradiated 3T3 fibroblasts as feeder cells, the XB cells replicate and differentiate to give characteristic flattened, keratinized cells that are identifiable both by their shape and because they stain red with the dye Rhodanile Blue. At high cell density, the XB cells replicate but do not keratinize. However, low concentrations of TCDD cause the high-density cultures to keratinize in a dose-dependent fashion; therefore, the extent of keratinization can be correlated with the intensity of color development when the cells are stained. Keratinization requires a minimum concentration of TCDD of 50 pM; additionally, the efficacy of keratinization by a PCDD congener roughly parallels its binding affinity for the AhR, and thus suggests that this is a mechanism-based assay. This assay was shown to be applicable to other DLCs such as TCAB. A drawback to the assay is the relatively long incubation time that is required (nearly 2 weeks), which militates against the use of the assay for following the progress of a remediation ''on the fly''.

A variant of the keratinization assay, introduced by Gierthy and Crane,[211] employed a strain of XB cells denoted XBF. These do not keratinize in the presence of TCDD, but during a 14-day incubation form ''cobblestone-like'' cells (flat cells) that can be visualized by staining. The DL in this assay is 10 pM TCDD in the culture medium, with PAHs having 10^3–10^4-fold less response than TCDD. The flat-cell assay correlated linearly with chemical analysis by GC/MS for a series of soot samples that were contaminated with PCDFs.[121] It showed semiquantitative agreement with chemical analysis for the DLC content of a range of fish tissue and sediment samples[212] and of several transformer fluids.[213]

As a related end-point, one of the most characteristic and well-documented responses of halogenated aromatic compounds in humans is chloracne, caused by hyperkeratosis of squamous skin cells.[214] This end-point can also be seen in experimental animals such as rabbits and hairless mice. Indeed, an early example of a bioassay for DLCs was the production of chloracne in the rabbit pinna.[215]

One of the characteristic responses of DLCs is an increase in cell proliferation, which is linked mechanistically to a change in the binding between the epidermal growth factor and its receptor.[216,217] Cancers induced by TCDD also appear to be principally those of epithelial cells.[210,216]

6 CONCLUSIONS

Present regulations establish allowable concentrations of DLCs either on the concentration of TCDD or the total TEQ of named congeners, based in each case on congener-specific analysis. Under these circumstances bioassays can never replace an ''absolute'' analytical method such as GC/MS because they do not provide chemical identification. However, an inexpensive and rapid bioassay can usefully act as a screening tool based on a ''cut-off'' TEQ, to eliminate negative samples from further analysis. This could greatly reduce costs, especially if most samples are expected to have low levels of contamination and the assay has a low rate of false positives.

We have noted that AhR binding assays and immunoassays tend to have high false-positive rates (a ''conservative'' feature) because they respond to substances that bind to the receptor (or Ab) without causing toxicity. A

critical limitation occurs when screening assays have a significant rate of false negatives, which imply incorrectly that a contaminated sample is ''clean''. Enzyme induction and porphyrin accumulation are examples of assays in which one DLC may mask (antagonize) the dioxin-like activity of another, and an analyst cannot be confident that the TEQ obtained from the assay is a true reflection of the TEQ of the sample. Such assays are of limited value in screening programs unless interacting analytes are independently known to be absent; however, they may remain viable options for monitoring the progress of a remediation project.

The question of biological relevance always arises with an assay based on a biological principle. Whereas human health concerns might be addressed by an enzyme induction assay, preferably using a human cell culture, protection of fish in an aquatic environment might suggest an early life stage assay to monitor the potential for reproductive or developmental abnormalities. It is unrealistic to imagine that a single procedure will be applicable to all situations.

ABBREVIATIONS AND ACRONYMS

Ab	Antibody
AHH	Aryl Hydrocarbon Hydroxylase
AhR	Aryl Hydrocarbon Receptor
ALBS	Aryl Hydrocarbon Receptor Ligand Binding Subunit
ANOVA	Analysis of Variance
AP	Alkaline Phosphatase
ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator Protein
ATP	Adenosine Triphosphate
CALUX	Chemically Activated Luciferase Gene Expression
CAT	Chloramphenicol Acetyl Transferase
DL	Detection Limit
DLC	Dioxin-like Compound
DMSO	Dimethyl Sulfoxide
DRE	Dioxin Response Enhancer
ECOD	7-Ethoxycoumarin-*O*-deethylase
ELISA	Enzyme-linked Immunosorbent Assay
EROD	7-Ethoxyresorufin-*O*-deethylase
GC/MS	Gas Chromatography/Mass Spectrometry
HAP	Hydroxyapatite
IR	Infrared
MFO	Mixed-function Oxidase
MS	Mass Spectrometry
NADPH	Reduced Nicotinamide Adenine Dinucleotide
NMR	Nuclear Magnetic Resonance
P450RGS	P450 Reporter Gene System
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PCB 52	2,2′,5,5′-Tetrachlorobiphenyl
PCB 77	3,3′,4,4′-Tetrachlorobiphenyl
PCB 105	2,3,3′,4,4′-Pentachlorobiphenyl
PCB 114	2,3,4,4′,5-Pentachlorobiphenyl
PCB 118	2,3′,4,4′,5-Pentachlorobiphenyl
PCB 126	3,3′,4,4′,5-Pentachlorobiphenyl
PCB 128	2,2′,3,3′,4,4′-Hexachlorobiphenyl
PCB 153	2,2′,4,4′,5,5′-Hexachlorobiphenyl
PCB 156	2,3,3′,4,4′,5-Hexachlorobiphenyl
PCB 169	3,3′,4,4′,5,5′-Hexachlorobiphenyl
PCDD	Polychlorinated Dibenzo-*p*-dioxin
PCDF	Polychlorinated Dibenzofuran
PROD	7-Pentoxyresorufin-*O*-deethylase
QSAR	Quantitative Structure–Activity Relationship
RBA	Relative Binding Affinity
SDGC	Sucrose Density Gradient Centrifugation Analysis
SPFC	Splenic Plaque-forming Cell
SRBCs	Sheep Red Blood Cells
TCAB	3,3′,4,4′-Tetrachloroazobenzene
TCDD	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin
TCDF	1,2,7,8-Tetrachlorodibenzofuran
TEF	Toxic Equivalency Factor
TEQ	Toxic Equivalence
UP-III	Uroporphyrin III
2,4,5-T	2,4,5-Trichlorophenoxyacetic acid
3-MC	3-Methylcholanthrene

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples

Pesticides **(Volume 7)**
Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. US EPA Office of Water, Method 1613, Revision B, EPA 821-B-94-005, October 1994.

2. D. Firestone, *J. Assoc. Off. Anal. Chem.*, **74**, 375–384 (1991).
3. 'International Toxicity Equivalency Factor (I-TEF) Method of Risk Assessment for Complex Mixtures of Dioxins and Related Compounds', NATO–CCMS Report No. 176, 1988.
4. S. Safe, *J. Anim. Sci.*, **76**, 134–141 (1998).
5. S. Tanabe, N. Kannan, An. Subramanian, S. Watanabe, R. Tatsukawa, *Environ. Pollut.*, **47**, 147–163 (1987).
6. D.L. Stalling, R.J. Norstrom, L.M. Smith, M. Simon, *Chemosphere*, **14**, 627–643 (1985).
7. R.M. Putzrath, *Organohalogen Compd.*, **25**, 111–115 (1995).
8. E.K. Silbergeld, T.A. Gasiewicz, *Am. J. Ind. Med.*, **16**, 455–474 (1989).
9. K. Rozman, *Exp. Toxicol. Pathol.*, **44**, 473–480 (1992).
10. D.A. Skoog, F.J. Holler, T.A. Nieman, *Principles of Instrumental Analysis*, 5th edition, Harcourt Brace College Publishers, Philadelphia, PA, 1998.
11. R.J. Kociba, D.G. Keyes, J.E. Beyer, R.M. Carreon, C.E. Wade, D.A. Dittenber, R.P. Kalnins, L.E. Frauson, C.N. Park, S.D. Barnard, R.A. Hummel, C.G. Humiston, *Toxicol. Appl. Pharmacol.*, **46**, 279–303 (1978).
12. E.S. Johnson, *Environ. Health Perspect.*, **99**, 383–390 (1993).
13. D.J. Paustenbach, M.W. Layard, R.J. Wenning, R.E. Keenan, *J. Toxicol. Environ. Health*, **34**, 11–26 (1991).
14. R.L. Sielken, Jr., *Food Chem. Toxicol.*, **25**, 257–267 (1987).
15. J. Hockel, L. Dusterhoft, W. Korner, H. Hagenmaier, *Organohalogen Compd.*, **23**, 135–138 (1995).
16. S.M.G. Schwirzer, F.J. Wiebel, K.-W. Schramm, *Organohalogen Compd.*, **23**, 85–88 (1995).
17. K. Hu, N.J. Bunce, B.G. Chittim, C.H.M. Tashiro, B.R. Yeo, B.J. Sharrat, F.J. Campbell, D.W. Potter, *Environ. Sci. Technol.*, **29**, 2603–2609 (1995).
18. L. Biegel, L. Howie, S. Safe, *Chemosphere*, **19**, 955–958 (1989).
19. B. Lesnik, *Organohalogen Compd.*, **23**, 203–208 (1995).
20. M. Bader, *J. Chem. Educ.*, **57**, 703 (1980).
21. T. Zacharewski, L. Safe, S. Safe, *Environ. Sci. Technol.*, **23**, 730–735 (1989).
22. J.P. Landers, N.J. Bunce, *Biochem. J.*, **276**, 273–287 (1991).
23. J.P. Whitlock, Jr, *Chem. Res. Toxicol.*, **6**, 754–763 (1993).
24. A.B. Okey, D.S. Riddick, P.A. Harper, *Toxicol. Lett.*, **70**, 1–22 (1994).
25. D. Mukerjee, *J. Air Waste Manage. Assoc.*, **48**, 157–165 (1998).
26. M.J. DeVito, L.S. Birnbaum, *Toxicology*, **102**, 115–123 (1995).
27. H. Fiedler, O. Hutzinger, C.W. Timms, *Toxicol. Environ. Chem.*, **29**, 157–234 (1990).
28. S.D. Gilman, J.R. Sanborn, S.J. Gee, M.S. Denison, L.H. Stanker, A.D. Jones, B.D. Hammock, *Organohalogen Compd.*, **23**, 231–235 (1995).
29. A. Poland, J.C. Knutson, *Annu. Rev. Pharmacol. Toxicol.*, **22**, 517–554 (1982).
30. L. Marple, R. Brunck, L. Throop, *Environ. Sci. Technol.*, **20**, 180–182 (1988).
31. A. Poland, E. Glover, *Biochem. Biophys. Res. Commun.*, **146**, 1439–1449 (1987).
32. A. Wilhelmsson, S. Cuthill, M. Denis, A.C. Wikström, J.Å. Gustafsson, L. Poellinger, *EMBO J.*, **9**, 69–76 (1990).
33. I. Pongratz, G.G.F. Mason, L. Poellinger, *J. Biol. Chem.*, **267**, 13728–13734 (1992).
34. W.B. Pratt, *Proc. Soc. Exp. Biol. Med.*, **217**, 420–431, 1998.
35. E. Enan, F. Matsumura, *Biochem. Pharmacol.*, **52**, 1599–1612 (1996).
36. B.K. Meyer, M.G. Pray-Grant, J.P. Van Den Heuvel, G.H. Perdew, *Mol. Cell. Biol.*, **18**, 978–988 (1998).
37. C.A. Bradfield, A.S. Kende, A. Poland, *Mol. Pharmacol.*, **34**, 229–237 (1988).
38. N.J. Bunce, J.P. Landers, S.H. Safe, *Arch. Biochem. Biophys.*, **267**, 384–397 (1988).
39. A.B. Okey, L.M. Vella, P.A. Harper, *Mol. Pharmacol.*, **35**, 823–830 (1989).
40. A. Poland, D. Palen, E. Glover, *Mol. Pharmacol.*, **46**, 915–921 (1994).
41. D.K. Manchester, S.K. Gordon, C.L. Golas, E.A. Roberts, A.B. Okey, *Cancer Res.*, **47**, 4861–4868 (1987).
42. P.A. Harper, R.D. Prokipcak, L.E. Bush, C.L. Golas, A.B. Okey, *Arch. Biochem. Biophys.*, **290**, 27–36 (1991).
43. A. Lorenzen, A.B. Okey, *Toxicol. Appl. Pharmacol.*, **107**, 203–214 (1991).
44. A. Poland, E. Glover, A.S. Kende, *J. Biol. Chem.*, **251**, 4936–4946 (1976).
45. A. Poland, W.F. Greenlee, A.S. Kende, *Ann. N.Y. Acad. Sci.*, **320**, 214–230 (1979).
46. S.H. Safe, *Annu. Rev. Pharmacol. Toxicol.*, **26**, 371–399 (1986).
47. M. Harris, T. Zacharewski, J. Piskorska-Pliszczynska, R. Rosengren, S. Safe, *Toxicol. Appl. Pharmacol.*, **105**, 243–253 (1990).
48. L.F. Bjeldanes, J.-Y. Kim, K.R. Grose, J.C. Bartholomew, C.A. Bradfield, *Proc. Natl. Acad. Sci. USA*, **88**, 9543–9547 (1991).
49. S. Heath-Pagliuso, W.J. Rogers, K. Tullis, S.D. Seidel, P.H. Cenijn, A. Brouwer, M.S. Denison, *Biochemistry*, **37**, 11508–11515 (1998).
50. D. Phelan, G.M. Winter, W.J. Rogers, J.C. Lam, M.S. Denison, *Arch. Biochem. Biophys.*, **357**, 155–163 (1998).
51. D.S. Riddick, P.A. Harper, A.B. Okey, *FASEB J.*, **6**, A1201 (1992) (abstract).
52. E.C. Henry, T.A. Gasiewicz, *Biochem. J.*, **294**, 95–101 (1993).

53. E.C. Hoffman, H. Reyes, F.-F. Chu, F. Sander, L.H. Conley, B.A. Brooks, O. Hankinson, *Science.*, **252**, 954–958 (1991).
54. R.S. Pollenz, C.A. Sattler, A. Poland, *Mol. Pharmacol.*, **45**, 428–438 (1994).
55. H. Reyes, S. Reisz-Porszasz, O. Hankinson, *Science.*, **256**, 1193–1195 (1992).
56. E.F. Yao, M.S. Denison, *Biochemistry.*, **31**, 5060–5067 (1992).
57. L. Wu, J.P. Whitlock, Jr, *Nucleic Acids Res.*, **21**, 119–125 (1993).
58. J. Mimura, M. Ema, K. Sogawa, Y. Fujii-Kuriyama, *Genes Dev.*, **13**, 20–25 (1999).
59. S. Safe, *Organohalogen Compd.*, **26**, 7–13 (1995).
60. I. Chen, N. Harper, S. Safe, *Organohalogen Compd.*, **25**, 57–60 (1995).
61. J.R. Petrulis, N.J. Bunce, *Toxicol. Lett.*, **105**, 251–260 (1999).
62. G. Scatchard, *Ann. N.Y. Acad. Sci.*, **51**, 660–672 (1949).
63. J.S. Beck, H.J. Goren, *J. Receptor Res.*, **3**, 561–577 (1983).
64. J.S. Nakai, M.J. Winhall, N.J. Bunce, R. Rosengren, S. Safe, *J. Biochem. Toxicol.*, **9**, 199–209 (1994).
65. K.M. Dold, W.F. Greenlee, *Anal. Biochem.*, **184**, 67–73 (1990).
66. A. Goldstein, R.W. Barrett, *Mol. Pharmacol.*, **31**, 603–609 (1987).
67. M.M. Brown, U.A. Schneider, J.R. Petrulis, N.J. Bunce, *Toxicol. Appl. Pharmacol.*, **129**, 243–251 (1994).
68. J.J. Hubert, *Bioassay*, Kendall/Hunt Publishing Co., Dubuque, IA, 1984.
69. U.A. Schneider, M.M. Brown, R.A. Logan, L.C. Millar, N.J. Bunce, *Environ. Sci. Technol.*, **29**, 2595–2602 (1995).
70. L. Roberts, *Science.*, **251**, 624–626 (1991).
71. M.M. Brown, T.L. McCready, N.J. Bunce, *Toxicol. Lett.*, **61**, 141–147 (1992).
72. N.J. Bunce, J.P. Landers, J.S. Nakai, M.J. Winhall, S. Safe, *Toxicol. In Vitro*, **4**, 87–92 (1990).
73. P.A. Harper, J.V. Giannone, A.B. Okey, M.S. Denison, *Mol. Pharmacol.*, **42**, 603–612 (1992).
74. O. Hutzinger, K. Olie, J.W.A. Lustenhouwer, A.B. Okey, S. Bandiera, S. Safe, *Chemosphere*, **10**, 19–25 (1981).
75. H.O. Tsui, A.B. Okey, *Can. J. Physiol. Pharmacol.*, **59**, 927–931 (1981).
76. A.B. Okey, G.P. Bondy, M.E. Mason, G.F. Kahl, H.J. Eisen, T.M. Guenthner, D.W. Nebert, *J. Biol. Chem.*, **254**, 11636–11648 (1979).
77. G. Mason, M.A. Denomme, L. Safe, S. Safe, *Chemosphere*, **16**, 1729–1731 (1987).
78. G. Sundstrom, T. Sawyer, O. Hutzinger, S. Safe, *Chemosphere*, **15**, 2105–2107 (1986).
79. C.A. Bradfield, A. Poland, *Mol. Pharmacol.*, **34**, 682–688 (1988).
80. T.A. Gasiewicz, R.A. Neal, *Anal. Biochem.*, **124**, 1–11 (1982).
81. N.J. Bunce, J.P. Landers, S.H. Safe, T.R. Zacharewski, *Environ. Toxicol. Chem.*, **8**, 25–30 (1989).
82. N.J. Bunce, J.P. Landers, S.H. Safe, U.A. Schneider, T.R. Zacharewski, *Toxicol. Environ. Chem.*, **28**, 217–229 (1990).
83. J.S. Nakai, N.J. Bunce, *J. Biochem. Toxicol.*, **10**, 151–159 (1995).
84. L.G. Hansen, *Environ. Health Perspect.*, **106**, 171–189 (1998).
85. V.R. Rao, A. Unger, *Environ. Carcinogen. Ecotoxicol. Rev.*, **C13**, 53–74 (1995).
86. M.S. Denison, L.M. Vella, *Arch. Biochem. Biophys.*, **277**, 382–388 (1990).
87. E.C. Henry, G. Rucci, T.A. Gasiewicz, *Biochemistry.*, **28**, 6430–6440 (1989).
88. T.A. Gasiewicz, C.J. Elferink, E.C. Henry, *Biochemistry.*, **30**, 2909–2916 (1991).
89. J.R. Petrulis, PhD Thesis, University of Guelph, 1998.
90. J.R. Petrulis, N.J. Bunce, *Organohalogen Compd.*, **34**, 317–322 (1997).
91. T. Zacharewski, M. Harris, S. Safe, *Arch. Biochem. Biophys.*, **272**, 344–355 (1989).
92. J.W. Anderson, S.S. Rossi, R.H. Tukey, T. Vu, L.C. Quattrochi, *Environ. Toxicol. Chem.*, **14**, 1159–1169 (1995).
93. X. Wang, S. Safe, *Arch. Biochem. Biophys.*, **315**, 285–292 (1994).
94. M. Santostefano, H. Liu, X. Wang, K. Chaloupka, S. Safe, *Chem. Res. Toxicol.*, **7**, 544–550 (1994).
95. J.M.M.J.G. Aarts, M.S. Denison, M.A. Cox, M.A.C. Schalk, P.M. Garrison, K. Tullis, L.H.J. de Haan, A. Brouwer, *Eur. J. Pharmacol.*, **293**, 463–474 (1995).
96. M.S. Denison, P.M. Garrison, J.M.M.J.G. Aarts, K. Tullis, J.A.C. Schalk, M.A. Cox, A. Brouwer, *Organohalogen Compd.*, **23**, 225–229 (1995).
97. M.-T. Busser, W.K. Lutz, *Carcinogenesis.*, **8**, 1433–1437 (1995).
98. J.C. Lewis, A. Feltus, C.M. Ensor, S. Ramanathan, S. Daunert, *Anal. Chem.*, **70**, 579A–585A (1998).
99. J.H. Clemons, T.R. Zacharewski, in *Molecular Biology of the Toxic Response*, eds. A. Puga, K.B. Wallace, Taylor & Francis, London, 1998.
100. M.H. El-Fouly, C. Richter, J.P. Giesy, M.S. Denison, *Environ. Toxicol. Chem.*, **13**, 1581–1588 (1994).
101. P. Balaguer, A. Joyeux, M.S. Denison, R. Vincent, B.E. Gillesby, T. Zacharewski, *Can. J. Physiol. Pharmacol.*, **74**, 216–222 (1996).
102. A.J. Murk, J. Legler, M.S. Denison, J.P. Giesy, C. Van De Guchte, A. Brouwer, *Fundam. Appl. Toxicol.*, **33**, 149–160 (1996).
103. A.J. Murk, P.E.G. Leonards, A.S. Bulder, A.S. Jonas, M.J.C. Rozemeijer, M.S. Denison, J.H. Koeman, A. Brouwer, *Environ. Toxicol. Chem.*, **16**, 1583–1589 (1997).

104. C.A. Richter, V.L. Tieber, M.S. Denison, J.P. Giesy, *Environ. Toxicol. Chem.*, **16**, 543–550 (1997).
105. J.W. Anderson, K. Bothner, D. Edelman, S. Vincent, T.P. Vu, R.H. Tukey, in *ACS Symposium Series 643: Biomarkers for Agrochemicals and Toxic Substances*, eds. J.N. Blancato, R.N. Brown, C.C. Dary, M.A. Saleh, American Chemical Society, Washington, DC, 150–168, 1996.
106. G.B. Kim, J.W. Anderson, K. Bothner, J.-L. Lee, C.-H. Koh, S. Tanabe, *Biomarkers.*, **2**, 181–188 (1997).
107. P.M. Garrison, K. Tullis, J.M.M.J.G. Aarts, A. Brouwer, J.P. Giesy, M.S. Denison, *Fundam. Appl. Toxicol.*, **30**, 194–203 (1996).
108. M.A. Denomme, S. Bandiera, I. Lambert, L. Copp, L. Safe, S. Safe, *Biochem. Pharmacol.*, **32**, 2955–2963 (1983).
109. D.K. Strom, H. Postlind, R.H. Tukey, *Arch. Biochem. Biophys.*, **294**, 707–716 (1992).
110. M.J. DeVito, W.E. Maier, J.J. Diliberto, L.S. Birnbaum, *Fundam. Appl. Toxicol.*, **20**, 125–130 (1993).
111. R.A. Lubet, R.T. Mayer, J.W. Cameron, R.W. Nims, M.D. Burke, T. Wolff, F.P. Guengerich, *Arch. Biochem. Biophys.*, **238**, 43–48 (1985).
112. F.P. Guengerich, G.A. Dannan, S.T. Wright, M.V. Martin, L.S. Kaminsky, *Biochemistry*, **21**, 6019–6030 (1982).
113. K. Hu, N.J. Bunce, *J. Toxicol. Environ. Health(B)*, **2**, 183–210 (1999).
114. A. Poland, A.S. Kende, in *Origins of Human Cancer (Book B): Mechanisms of Carcinogenesis*, eds. H.H. Hiatt, J.D. Watson, J.A. Winsten, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 847–867, 1977.
115. J.A. Bradlaw, J.L. Casterline, *Anal. Chem.*, **62**, 904–916 (1979).
116. R.L. Jansing, W. Shain, *Fundam. Appl. Toxicol.*, **5**, 713–720 (1985).
117. W.F. Benedict, J.E. Gielen, I.S. Owens, A. Niwa, D.W. Nebert, *Biochem. Pharmacol.*, **22**, 2766–2769 (1973).
118. A. Niwa, K. Kumaki, D.W. Nebert, *Mol. Pharmacol.*, **11**, 399–408 (1975).
119. T.W. Sawyer, S. Safe, *Chemosphere*, **14**, 79–84 (1985).
120. M.D. Burke, R.T. Mayer, *Drug Metab. Dispos.*, **2**, 583 (1974).
121. S.W. Kennedy, A. Lorenzen, C.A. James, B.T. Collins, *Anal. Biochem.*, **211**, 102–112 (1993).
122. J.L. Parrott, P.V. Hodson, D.G. Dixon, *J. Toxicol. Environ. Health*, **46**, 301–316 (1995).
123. J.J. Diliberto, P.I. Akubue, R.W. Luebke, L.S. Birnbaum, *Toxicol. Appl. Pharmacol.*, **130**, 197–208 (1995).
124. M. Machala, L. Matlova, I. Svoboda, K. Nezveda, *Arch. Toxicol.*, **70**, 362–367 (1996).
125. A.T.C. Bosveld, E. Verhallen, W. Seinen, M. Van den Berg, *Organohalogen Compd.*, **25**, 309–312 (1995).
126. M. Dubois, I. De Waziers, J.P. Thome, P. Kremers, *Comp. Biochem. Physiol.*, **113C**, 51–59 (1996).
127. O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, *J. Biol. Chem.*, **193**, 265 (1951).
128. M.M. Bradford, *Anal. Biochem.*, **72**, 248 (1976).
129. A. Lorenzen, S.W. Kennedy, *Anal. Biochem.*, **214**, 346–348 (1993).
130. J.V. Schmidt, C.A. Bradfield, *Annu. Rev. Cell. Dev. Biol.*, **12**, 55–89 (1996).
131. J.P. Whitlock, S.T. Okino, L. Dong, H.P. Ko, R. Clarke-Katzenberg, Q. Ma, H. Li, *FASEB J.*, **10**, 809–818 (1996).
132. J.-Y.K. Park, M.K. Shigenaga, B.N. Ames, *Proc. Natl. Acad. Sci. USA*, **93**, 2322–2327 (1996).
133. S. Safe, D. Davis, M. Romkes, C. Yao, B. Keyes, J. Piskorska-Plisczynska, K. Farrel, G. Mason, M.A. Denomme, L. Safe, B. Zmudzka, M. Holcomb, *Chemosphere*, **19**, 853–860 (1989).
134. J.H. Clemons, L.E.J. Lee, C.R. Myers, D.G. Dixon, N.C. Bols, *Can. J. Fish. Aquat. Sci.*, **53**, 1177–1185 (1996).
135. P. Lindström-Seppä, A. Oikari, *Aquat. Toxicol.*, **16**, 187–204 (1990).
136. R. Soimasuo, I. Jokinen, J. Kukkonen, T. Petänen, T. Ristola, A. Oikari, *Aquat. Toxicol.*, **31**, 329–345 (1995).
137. P.V. Hodson, J. Sherry, J. Parrott, in *Microscale Testing in Aquatic Toxicology*, eds. P.G. Wells, K. Lee, C. Blaise, CRC Press, Boca Raton, FL, 53–75, 1998.
138. J.T. Sanderson, J.M.M.J.G. Aarts, A. Brouwer, K.L. Froese, M.S. Denison, J.P. Giesy, *Toxicol. Appl. Pharmacol.*, **137**, 316–325 (1996).
139. E.Y. Verhallen, M. Van Den Berg, A.T.C. Bosveld, *Environ. Toxicol. Chem.*, **16**, 277–282 (1997).
140. H.T. Besselink, M.S. Denison, M.E. Hahn, S.I. Karchner, A.D. Vethaak, J.H. Koeman, A. Brouwer, *Fundam. Appl. Toxicol.*, **43**, 161–171 (1998).
141. J.H. Clemons, D.G. Dixon, N.C. Bols, *Chemosphere*, **34**, 1105–1119 (1997).
142. B.M. Hasspieler, K. Gillian, A. Niimi, G.D. Haffner, K. Adeli, *Organohalogen Compd.*, **25**, 69–73 (1995).
143. D.M. Janz, C.D. Metcalfe, *Chemosphere*, **23**, 467–472 (1991).
144. A.P.J.M. Van Birgelen, J. Van der Kolk, K.M. Fase, I. Bol, H. Poiger, M. Van den Berg, A. Brouwer, *Toxicol. Appl. Pharmacol.*, **126**, 202–213 (1994).
145. H.L. Bonkovsky, P.R. Sinclair, W.J. Bement, R.W. Lambrecht, J.F. Sinclair, *Ann. N.Y. Acad. Sci.*, **514**, 96–112 (1987).
146. K.G. Jones, G.D. Sweeney, *Toxicol. Appl. Pharmacol.*, **53**, 42–49 (1980).
147. A.P.J.M. Van Birgelen, M.J. DeVito, J.M. Akins, D.G. Ross, J.J. Diliberto, L.S. Birnbaum, *Toxicol. Appl. Pharmacol.*, **138**, 98–109 (1996).
148. P.R. Sinclair, N. Gorman, T. Dalton, H.S. Walton, W.J. Bement, J.F. Sinclair, A.G. Smith, D.W. Nebert, *Biochem. J.*, **330**, 149–153 (1998).
149. P.R. Sinclair, N. Gorman, I/B. Tsyrlov, U. Fuhr, H.S. Walton, J.F. Sinclair, *Drug Metab. Dispos.*, **26**, 1019–1025 (1998).

150. A. Poland, E. Glover, *Mol. Pharmacol.*, **9**, 736–747 (1973).
151. A.B. Rifkind, S. Sassa, J. Reyes, H. Muschick, *Toxicol. Appl. Pharmacol.*, **78**, 268–279 (1985).
152. A.B. Rifkind, A. Firpo, Jr, D.R. Alonso, *Toxicol. Appl. Pharmacol.*, **72**, 343–354 (1984).
153. A. Lorenzen, S.W. Kennedy, *Organohalogen Compd.*, **25**, 65–68 (1995).
154. M. Tysklind, B. Bosveld, P. Andersson, E. Verhallen, T. Sinnige, W. Seinen, C. Rappe, M. Van den Berg, *Organohalogen Compd.*, **25**, 49–52 (1995).
155. A.G. Smith, P. Carthew, B. Clothier, D. Constantin, J.E. Francis, S. Madra, *Toxicol. Lett.*, **82/83**, 945–950 (1995).
156. J.P. Sherry, *Chemosphere*, **34**, 1011–1025 (1997).
157. G.H. Perdew, B. Abbott, L.H. Stanker, *Hybridoma*, **14**, 279–283 (1995).
158. L.H. Stanker, B. Watkins, N. Rogers, M. Vanderlaan, *Toxicology*, **45**, 229–243 (1987).
159. P.W. Albro, M.I. Luster, S.K. Chaudhary, G. Clark, L.D. Lawson, J.T. Corbett, J.D. McKinney, *Toxicol. Appl. Pharmacol.*, **50**, 137–146 (1979).
160. M.I. Luster, P.W. Albro, K. Chae, L.D. Lawson, J.T. Corbett, J.D. McKinney, *Anal. Chem.*, **52**, 1497–1500 (1980).
161. J. Sherry, *Organohalogen Compd.*, **23**, 181–186 (1995).
162. S.J. Kennel, C. Jason, P.W. Albro, G. Mason, S.H. Safe, *Toxicol. Appl. Pharmacol.*, **82**, 256–263 (1986).
163. S.D. Gilman, J.R. Sanborn, S.J. Gee, M.S. Denison, L.H. Stanker, A.D. Jones, B.D. Hammock, *Organohalogen Compd.*, **23**, 231–235 (1995).
164. Y. Sugawara, S.J. Gee, J.R. Sanborn, S.D. Gilman, B.D. Hammock, *Anal. Chem.*, **70**, 1092–1099 (1998).
165. R.O. Harrison, R.E. Carlson, H. Shirkhan, *Organohalogen Compd.*, **23**, 187–192 (1995).
166. R.E. Carlson, Y.-W. Chiu, K.L. Marcus, A.E. Karu, *Organohalogen Compd.*, **23**, 193–198 (1995).
167. Y.-W. Chiu, R.E. Carlson, K.L. Marcus, A.E. Karu, *Anal. Chem.*, **67**, 3829–3839 (1995).
168. S. Bender, O.A. Sadik, *Environ. Sci. Technol.*, **32**, 788–797 (1998).
169. M. Vanderlaan, L.H. Stanker, B.E. Watkins, P. Petrovic, S. Gorbach, *Environ. Toxicol. Chem.*, **7**, 859–870 (1988).
170. M. Zennegg, R.O. Harrison, P. Schmid, *Organohalogen Compd.*, **35**, 213–215 (1998).
171. R.O. Harrison, R.E. Carlson, *Organohalogen Compd.*, **35**, 43–46 (1998).
172. Y. Sugawara, S.J. Gee, J.R. Sanborn, S.D. Gilman, B.D. Hammock, *Anal. Chem.*, **70**, 1092–1099 (1998).
173. D. Davis, S. Safe, *Immunopharmacology*, **21**, 183–190 (1991).
174. R.J. Smialkowicz, M.J. DeVito, M.M. Riddle, W.C. Williams, L.S. Birnbaum, *Fundam. Appl. Toxicol.*, **37**, 141–149 (1997).
175. T.A. Gasiewicz, *Neurotoxicology*, **18**, 393–414 (1997).
176. B.D. Abbott, L.S. Birnbaum, G.H. Perdew, *Dev. Dyn.*, **204**, 133–143 (1995).
177. Y.L. Guo, T.J. Lai, S.H. Ju, Y.C. Chen, C.C. Hsu, *Chemosphere*, **14**, 235–238 (1993).
178. J.J. Willey, B.R. Stripp, R.B. Baggs, T.A. Gasiewicz, *Toxicol. Appl. Pharmacol.*, **151**, 33–44 (1998).
179. L.S. Birnbaum, H. Weber, M.W. Harris, J.C. Lamb, IV, J.D. McKinney, *Toxicol. Appl. Pharmacol.*, **77**, 292–302 (1985).
180. R. Bannister, S. Safe, *Toxicology*, **44**, 159–169 (1987).
181. J.M. Haake, S. Safe, K. Mayura, T.D. Phillips, *Toxicol. Lett.*, **38**, 299–306 (1987).
182. R.E. Morrissey, M.W. Harris, J.J. Diliberto, L.S. Birnbaum, *Toxicol. Lett.*, **60**, 19–25 (1992).
183. L. Biegel, M. Harris, D. Davis, R. Rosengren, L. Safe, S. Safe, *Toxicol. Appl. Pharmacol.*, **97**, 561–571 (1989).
184. B. Hileman, *Chem. Eng. News.*, **72**(January 31), 19–23 (1994).
185. I. Kharat, F. Saatcioglu, *J. Biol. Chem.*, **271**, 10 533–10 537 (1996).
186. J.F. Gierthy, J.A. Bennett, L.M. Bradley, D.S. Cutler, *Cancer Res.*, **53**, 3149–3153 (1993).
187. S. Safe, B. Astroff, M. Harris, T. Zacharewski, R. Dickerson, M. Romkes, L. Biegel, *Pharmacol. Toxicol.*, **69**, 400–409 (1991).
188. D.L. Bjerke, T.J. Brown, N.J. MacLusky, R.B. Hochberg, R.E. Peterson, *Toxicol. Appl. Pharmacol.*, **127**, 258–267 (1994).
189. J.R. Long, D.J. Hanson, *Chem. Eng. News.*, **61** (June 6), 23–36 (1983).
190. G.R. Higginbotham, A. Huang, D. Firestone, J. Verrett, J. Ress, A.D. Campbell, *Nature*, **220**, 702–703 (1968).
191. D.F. Flick, D. Firestone, J. Ress, J.R. Allen, *Poult. Sci.*, **52**, 1637–1641 (1973).
192. B. Brunström, *Arch. Toxicol.*, **64**, 188–192 (1990).
193. T. Helder, *Sci. Total Environ.*, **14**, 255–264 (1980).
194. T. Helder, *Toxicology*, **19**, 101–112 (1981).
195. T. Helder, W. Seinen, *Chemosphere*, **14**, 183–193 (1985).
196. C.D. Metcalfe, V.W. Cairns, J.D. Fitzsimons, *Aquat. Toxicol.*, **13**, 347–356 (1988).
197. M.K. Walker, L.C. Hufnagle, M.K. Clayton, R.E. Peterson, *Aquat. Toxicol.*, **22**, 15–38 (1992).
198. E.W. Zabel, M.K. Walker, M.W. Hornung, M.K. Clayton, R.E. Peterson, *Toxicol. Appl. Pharmacol.*, **134**, 204–213 (1995).
199. M.K. Walker, P.M. Cook, B.C. Butterworth, E.W. Zabel, R.E. Peterson, *Fundam. Appl. Toxicol.*, **30**, 178–186 (1996).
200. J.D. Wisk, K.R. Cooper, *Chemosphere*, **20**, 361–377 (1990).
201. J.D. Wisk, K.R. Cooper, *Environ. Toxicol. Chem.*, **9**, 1159–1169 (1990).
202. J.D. Wisk, K.R. Cooper, *Arch. Toxicol.*, **66**, 245–249 (1992).
203. G.E. Harris, T.L. Metcalfe, C.D. Metcalfe, S.Y. Huestis, *Environ. Toxicol. Chem.*, **13**, 1393–1403 (1994).

204. J.B. Silkworth, L. Antrim, L.S. Kaminsky, *Toxicol. Appl. Pharmacol.*, **75**, 156–165 (1984).
205. D. Davis, S. Safe, *Toxicology*, **63**, 97–111 (1990).
206. N. Harper, K. Connor, S. Safe, *Toxicology*, **80**, 217–227 (1993).
207. R. Bannister, D. Davis, T. Zacharewski, I. Tizard, S. Safe, *Toxicology*, **46**, 29–42 (1987).
208. D. Davis, S. Safe, *Toxicol. Lett.*, **48**, 35–43 (1989).
209. R.J. Smialowicz, M.J. DeVito, M.M. Riddle, W.C. Williams, L.S. Birnbaum, *Organohalogen Compd.*, **25**, 299–302 (1995).
210. J.C. Knutson, A. Poland, *Cell*, **22**, 27–36 (1980).
211. J.F. Gierthy, D. Crane, *Fundam. Appl. Toxicol.*, **5**, 754–759 (1985).
212. J.F. Gierthy, D.W. Lincoln, P. O'Keefe, R. Smith, C. Meyer, D. Hilker, K. Aldous, *Chemosphere*, **18**, 793–800 (1989).
213. J.F. Gierthy, K. Swami, R. Narang, A. Narang, G. Eadon, *Chemosphere*, **26**, 1225–1235 (1993).
214. S. Safe, *CRC Crit. Rev. Toxicol.*, **13**, 319–395 (1984).
215. E. Jones, H. Krizek, *J. Invest. Dermatol.*, **39**, 511–517 (1962).
216. R. Osborne, W.F. Greenlee, *Toxicol. Appl. Pharmacol.*, **77**, 434–443 (1985).
217. B.V. Madhukar, D.W. Brewster, F. Matsumura, *Proc. Natl. Acad. Sci. USA*, **81**, 7407–7411 (1984).

Dyes, Environmental Analysis of

J. Riu and D. Barceló
IIQAB–CSIC, Barcelona, Spain

1 Introduction 1
2 Liquid Chromatography/Mass Spectrometry 3
2.1 Thermospray 4
2.2 Atmospheric Pressure Ionization Interfaces 5
3 Capillary Electrophoresis 7
3.1 Capillary Electrophoresis/Ultraviolet Detection 8
3.2 Capillary Electrophoresis/Mass Spectrometry 9
4 Sample Handling 10
5 Conclusions 12
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 13
References 14

An overview of current analytical methodologies for the determination of dyes in environmental samples is presented. Conventional analytical methods involving liquid chromatography (LC) and mass spectrometry (MS) with thermospray (TSP) interface are reviewed as well as the newly developed methods using atmospheric pressure ionization (API) interfaces. The combination of capillary electrophoresis (CE) with MS as a new alternative for the analysis and confirmation of polar compounds (and particularly the sulfonated azo dyes) in environmental samples is also discussed. Finally, the extraction and preconcentration of dyes from water samples involving various solid-phase extraction (SPE) cartridges are commented upon.

1 INTRODUCTION

Large quantities of dyes are produced for diverse applications, including textiles, paper, paint pigments, printing inks, and food coloring. According to recent information, nearly 20% of the world dye production, which rose more than 10% annually to 2.2 billion lb in 1994, is manufactured in Western Europe.[1] The textile industry is the largest consumer of these products, accounting for two-thirds of the dyestuff market. Since intermediates and degradation products of synthetic dyes can be potential hazards owing to their toxicity or carcinogenicity, they are of major environmental interest.

The analysis of dyes poses special problems because these products do not belong to a single group of chemical compounds, but encompass many chemical functionalities with large differences in solubility, volatility, ionization efficiency, and other properties. Additional complications in the analysis of sulfonated azo dyes exist because some of the manufacturing precursors to dyes are carried over to, and are not removed from, the final product. The result is a complex mixture containing the dye itself and several other interfering compounds.

In the nomenclature of dyes it is important to differentiate between dye classification and dye use. Dye classification is based on the major chromophore functionality of the dye, the main classes being azo (including the sulfonated azo dyes), anthraquinone, polymethine, phthalocyanine, sulfur, arylmethane, stilbene, and coumarin dyes. The use of a dye generally refers to the manner in which it is applied. Some of the most common applications are in acidic or basic media, as mordants, lakes, pigments, solvents, or dispersants. Figure 1 shows some typical structures of the main groups of dyes.

Dyes which are widely used in the textile industry to color natural fibers, may possess acid groups in their chemical structure. Especially, sulfonic acid groups are often present as sulfonate anions and provide very good water solubility. Recent estimates indicate that approximately 12% of the synthetic textile dyes used each year are lost to waste streams during manufacturing and processing operations and that 20% of those losses will enter the environment through effluents from wastewater treatment plants.[2] These compounds are difficult to remove in water-treatment procedures and, because of their high water solubility, they can be transported from municipal sewers through to rivers. In addition, dyes have been shown to undergo reduction in natural waterways and the environment and the degradation products include amines, which are known to be carcinogenic. Their presence in effluents and industrial wastewaters is of considerable interest because of their potential to contaminate groundwater and drinking water supplies.[3] Therefore, the detection, identification, and quantification of dyes at low levels in the aquatic environment is important for the protection of natural waters.

Most dyes are nonvolatile and thermally unstable, and consequently cannot be analyzed by gas chromatography (GC). For this reason, LC/MS was applied to the analysis and confirmation of the dyes. Current official analytical

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

methods for disperse azo dyes, e.g. United States Environmental Protection Agency (USEPA) 8321A,[4] are based on preconcentration methods followed by TSP/LC/MS, but are not applicable for sulfonated azo dyes (see Table 1). Problems were encountered with this interface for this kind of compound.[5]

With the development of API LC/MS interfaces, good sensitivity and structural information can be achieved as compared with other LC/MS interfacing devices such as TSP or particle beam (PB).

CE is a powerful alternative to classical chromatographic techniques for the separation of polar and ionic analytes because of its high separation efficiency and low solvent costs. It can be useful for overcoming the problems of LC in the separation of sulfonated azo dyes. In this respect, a major objective of the analytical chemistry

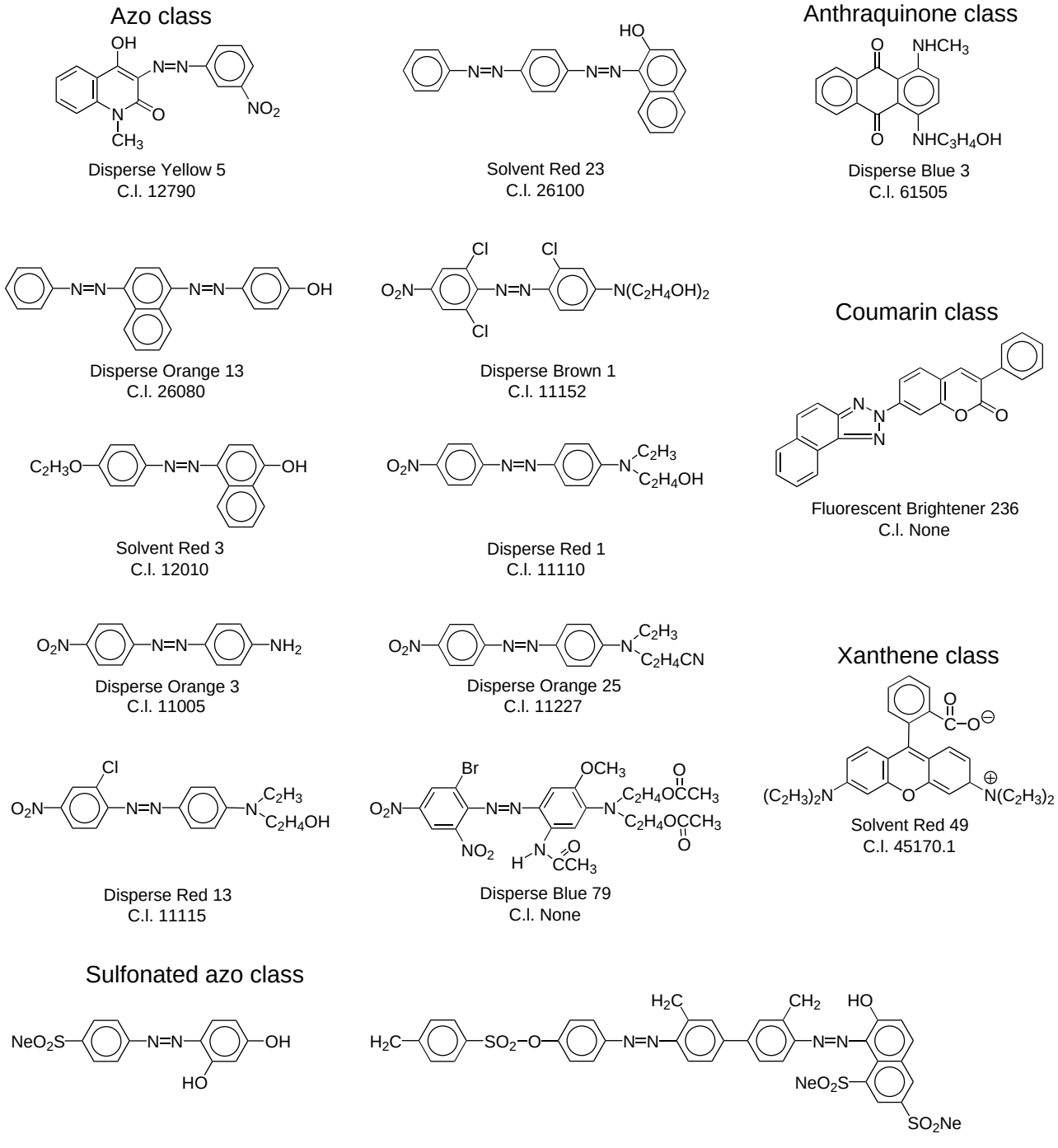

Figure 1 Structures of some examples of the most important groups of dyes.

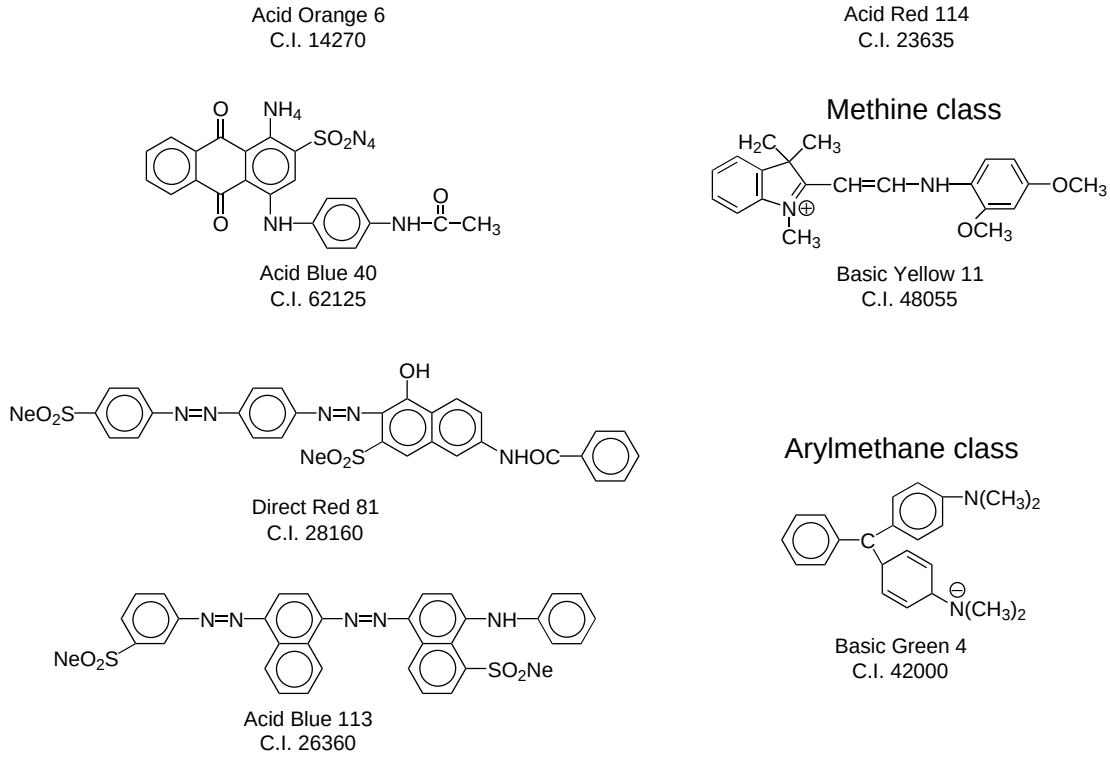

Figure 1 (*Continued*)

Table 1 Compounds included in USEPA Method 8321[4]

Dye type	Compound name	CAS no.
Azo dyes	Disperse Red 1	2872-52-8
	Disperse Red 5	3769-57-1
	Disperse Red 13	126038-78-6
	Disperse Yellow 5	6439-53-8
	Disperse Orange 3	730-40-5
	Disperse Orange 30	5261-31-4
	Disperse Brown 1	1764-91-4
	Solvent Red 3	6535-42-8
	Solvent Red 23	85-86-9
Anthraquinone dyes	Disperse Blue 3	2475-46-9
	Disperse Blue 14	2475-44-7
	Disperse Red 60	17418-58-5
	Coumarin Dyes	

program at USEPA's National Exposure Research Laboratory (Las Vegas) is to foster and advance the adoption of new environmental monitoring technologies, especially those suited to pollutants for which conventional analytical approaches fail. An analytical separation technique that has enormous potential for environmental analysis is CE.[6] One of the broad areas of great potential for CE in environmental monitoring is outlined in the goals adopted by USEPA and set forth in its Environmental Technology Initiative (ETI) formulated in 1993. Under ETI, environmental monitoring technologies are sought that offer continuous monitoring capabilities, are inexpensive, are exportable to foreign countries, are low generators of waste, minimize exposure of personnel to hazardous chemicals, and address a broad range of analytes. These widely divergent goals could be easily addressed by CE in its various formats.

An overview of the current analytical methods for the determination of dyes in environmental matrices is presented, and aspects such as sample extraction and preconcentration strategies are also discussed.

2 LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY

Most azo dyes are nonvolatile and thermally unstable compounds and consequently they cannot be analyzed directly by GC. Therefore, USEPA has validated a method[4] for several azo dyes and anthraquinone dyes, not including sulfonated azo dyes. The method is based on LC with ultraviolet (UV) and/or TSP/MS detection. Prior to analysis by high-performance liquid chromatography (HPLC) the target analytes are extracted by liquid–liquid extraction with methylene chloride for water samples or methylene chloride–acetone (1 : 1) for solid samples.

Some problems exist in the determination of sulfonated azo dyes by LC/TSP/MS and are discussed below.

Other LC/MS interfaces for the determination of sulfonated dyes have been tested such as the PB. Straub et al.[7] studied the mass spectra of some azo dyes including four solvent dyes, nine disperse dyes (one of which was a diazo dye), and one acid dye (the disulfonated azo dye Acid Orange 10).

With an HPLC/MS system using a PB interface it is possible to record the electron ionization (EI) mass spectra and gain useful structural information on these dyes. In addition, in some instances, additional organic impurities derived from decomposition, the coupling reaction process, or from external sources during the handling of commercial products, could be identified by HPLC/PB/MS. Characteristic fragmentation in azo dyes included cleavage of the C–N bonds on either side of the azo linkage, and cleavage of the N=N double bond with transfer of one or two hydrogen atoms, to form an imine or amine, respectively. Most azo dyes showed a molecular ion, and fragmentation from ring cleavage was observed. Naphthol-containing azo dyes presented $[M - CO]^+$ and $[M - HCO]^+$ fragment ions, which are not very intense but characteristic of a hydroxyl group on an aromatic ring. The acid sulfonated azo dye showed a spectra which indicated that complete desulfonation (m/z 246) was thermally accomplished in the mass spectrometer, in addition to characteristic fragment ions of naphthol-containing azo dyes. Detection limits of the compounds studied were in the range 500 ng–5 μg but the exact detection limit for Acid Orange 10 was not reported. LC/PB/MS was found to be two or three orders of magnitude less sensitive than TSP. The reduction products of Acid Orange 10 were identified by Voyksner et al.[8] These products are formed in industrial sludge and can be more harmful than their precursors.

Several desorption ionization methods have also been used for the analysis of sulfonated azo dyes. Ventura et al.[9] studied the spectra of some sulfonated dyes used in the industrial area of Barcelona (Spain) by fast atom bombardment (FAB) MS, including nitro, monoazo, diazo, trisazo, oxazine, and anthraquinone dyes. The behavior of the studied dyes varied widely. Cationized molecular ions, cleavage of azo groups, and loss of sulfonate from protonated or cationized molecular ions were the major common features. HPLC fractions of real water samples were analyzed with quantification limits that ranged from several nanograms to the low microgram level. However, because of low sample purity, modern dyes cannot be analyzed using the above-mentioned direct-probe techniques. This problem can be solved with the combination of a powerful separation technique, such as LC/MS.

2.1 Thermospray

TSP is a widely effective LC/MS technique because it can handle conventional LC solvents and chromatographic flow rates into the MS system and provides ionization for nonvolatile or thermally unstable samples. Therefore, TSP has been applied to the determination of dyes. Ballard and Betowski[5] reported detection limits between 15 and 200 ng for several dyes belonging to six different classes with a TSP interface coupled to a triple-stage quadrupole. This technique worked well for some representative dyes of the azo, methine, arylmethane, anthraquinone, coumarin, and xanthene classes, but it was less effective for the sulfonated azo dyes.

The negative-ion TSP mode of operation was found to be less sensitive than the positive-ion mode by a factor of 10. Switching the filament on can be a way to improve the structural information of the TSP ionization, as shown by Voyksner[10] with azo, diazo, and anthraquinone dyes in wastewater and soil. Detection limits of 10 ppt and 100 ppb in wastewater and soil, respectively, were achieved. The commercial diazo and anthraquinone dyes proved to be very complex mixtures of nearly 40 alkyl-substituted dye components, making monitoring and identification of a particular dyestuff difficult.

Flory et al.[11] investigated various factors that could affect the TSP response for nine sulfonated azo dyes. They found that too high concentrations of ammonium acetate buffer suppresses the ionization of these anionic dyes since the more volatile acetate anion will eject instead of evaporation of the dye anion.

The sensitivity problem of the TSP interface in the analysis of sulfonated azo dyes has been investigated by several authors and different solutions were given. Yinon et al.[12,13] recommended the use of a wire repeller placed in the opposite side of the ion extraction funnel. An increase of about two orders of magnitude was obtained with the wire repeller and the spectra of several sulfonated azo dyes could be recorded. In some instances, losses of $NaSO_3$ and $2NaSO_3$ were observed in many sulfonated dyes with the replacement of these groups by hydrogen atoms. Rafols and Barceló[14] found that for Acid Red 14, which has two $NaSO_3$ groups attached to different rings, both groups were lost. For Acid Blue 113, Acid Red 1, Mordant Red 9, and Acid Yellow 23, with two sulfonated groups in different rings, loss of only one group could be found. Owing to the above-mentioned controversy, it was suggested that, apart from the fact that the SO_3 groups are attached to different rings, other reasons related to their chemical structure make it difficult to lose the second $NaSO_3$ group. Losses of Na^+ or $2Na^+$ are common in all sulfonated dyes with replacement of these groups by hydrogen atoms. As an example, the TSP spectra of Acid Red 73 are shown in Figure 2.

The repeller-activated collisionally induced dissociation (CID) mass spectra of some dyes[15] was also compared with those obtained by TSP/MS with CID and

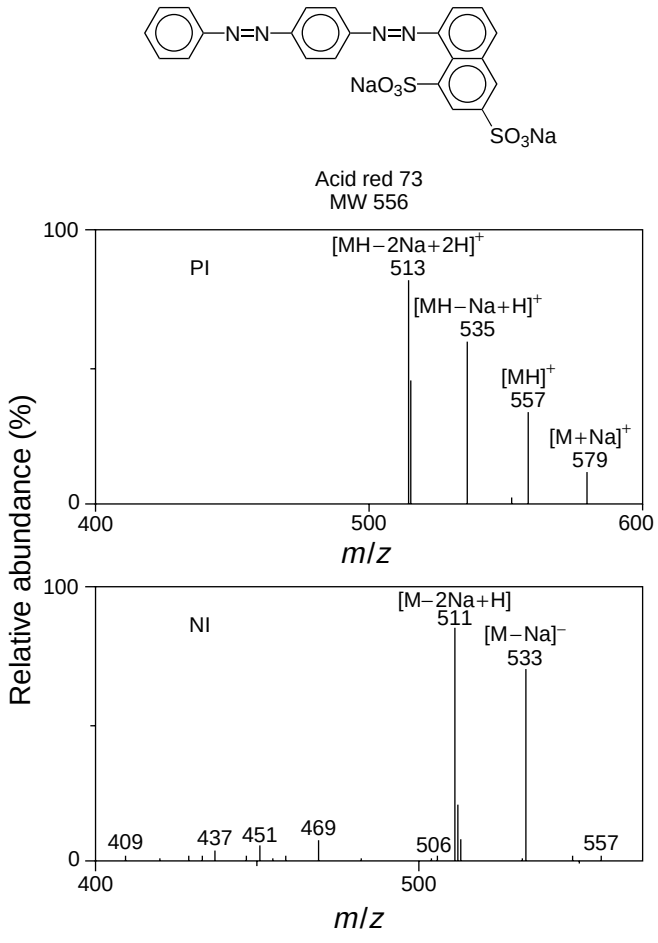

Figure 2 Spectra of Acid Red 73 with TSP/MS positive ionization (PI) and negative ionization (NI).

by EI. Common ions were found between these ionization modes.

McLean and Freas[16] modified a TSP/LC/MS system in order to improve the sensitivity for the analysis of some dyes by restricting the vaporizer exit orifice and adding a needle-tip repeller to the ion source. An increase in signal response was achieved for disulfonated azo dyes in the negative-ion mode. The TSP mass spectra contained mainly $[M - Na]^-$, $[M - 2Na]^{2-}$ and $[M - 2Na + H]^-$ ions.

Quantitative determination of disperse azo dyes has been carried out in a municipal wastewater treatment plant. The investigation of the fate of disperse azo dyes in the activated sludge process was the main objective of the study. Although the identification of two major reduction products of Disperse Red 1 by TSP/MS/MS could be achieved by Betowski et al.,[17] the mechanism of this reduction was unclear. Limits of detection of 180 pg for Disperse Red 1 in selected reaction monitoring could be achieved.

In summary, we can conclude that TSP gives, in most cases, poor structural information (tandem mass spectrometry (MS/MS) can be a solution for this problem) apart from molecular weight and some losses, e.g. SO_3^- from sulfonated azo dyes.

With reference to sensitivity, although most classes of dyes can be analyzed, good detection limits for monosulfonated and specially for polysulfonated azo dyes are difficult to achieve.

2.2 Atmospheric Pressure Ionization Interfaces

There are several interfaces which are classified as API interfaces, and that are given different names, but all of them have common features.[18]

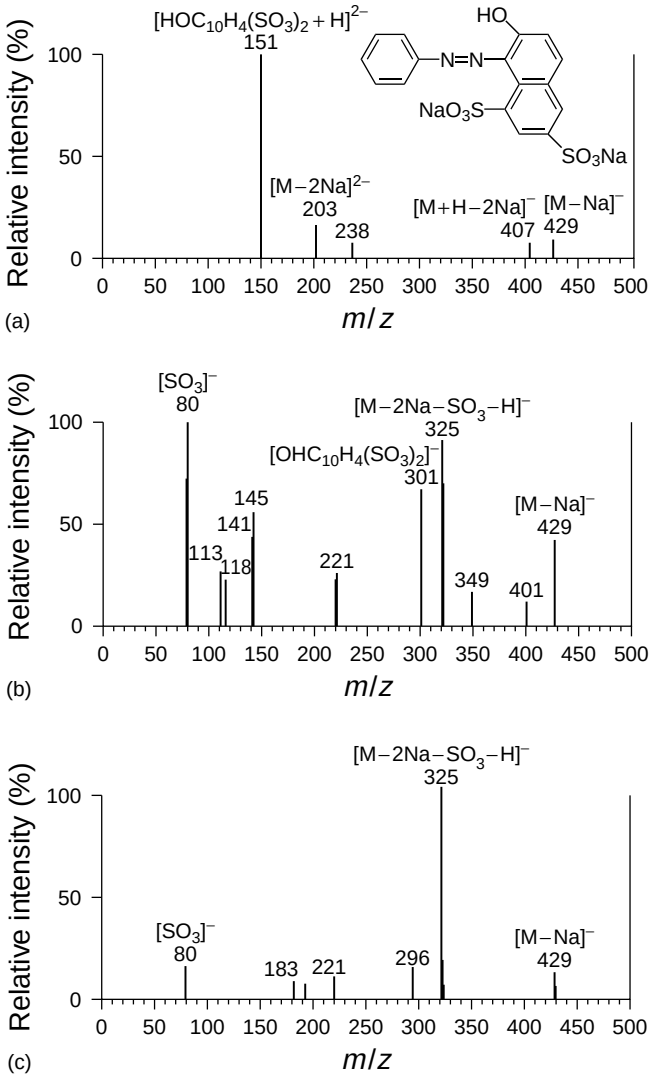

Figure 3 Flow injection/ESP mass spectra of 100 ng μL^{-1} of Acid Orange 10 at collision activated dissociation (CAD) voltage of (a) −100, (b) −200 and (c) −300 V.

API techniques of pneumatically assisted ionspray (ISP) and electrospray (ESP) have been used for the analysis of dyes.[19,20] ESP achieves the best sensitivity for compounds that are already charged in solution.[21] For this reason the initial work with ESP/MS focused on ionic dyes such as sulfonated azo dyes which were difficult to analyze by PB or TSP, as was shown in the work of Straub et al.[7] While the disulfonated azo dye Acid Orange 10 showed a weak $[MH - 2Na + 2H]^+$ ion in TSP ionization, and underwent considerable thermal degradation in EI with a PB interface, ESP in the negative-ion mode was able to detect characteristic anions formed in the solution. The negative-ion ESP spectra of this dye were recorded under three capillary–skimmer potential differences of −100, −220, and −300 V (see Figure 3). Lower potential differences resulted in the formation of singly charged ions $[M - Na]^-$, $[M + H - 2Na]^-$, and a doubly charged $[M - 2Na]^{2-}$ ion. Higher potential differences resulted in more fragmentation with some losses in sensitivity. The base peak at m/z 151 $[HOC_{10}H_4(SO_3)_2 + H]^{2-}$ disappeared and $[M - 2Na - SO_3 + H]^-$ and SO_3^- ions at m/z 325 and 80, respectively, appeared from cleavage of a sulfonic group. The ESP/MS detection limit in negative-ion operation under full-scan MS conditions was 10 ng. In contrast, positive ESP ionization of Acid Orange 10 showed an almost 50-fold lower sensitivity. Only a weak $[M + Na]^+$ ion was identified at positive potentials of 100, 240, and 300 V.

Rafols and Barceló[14] studied the spectra of 11 sulfonated azo dyes by negative ISP ionization. They found an optimum extraction cone voltage of 20 V to obtain the maximum sensitivity and enough fragmentation to have structural information. The detection limits for the compounds studied varied between 1 and 80 ng. All the

Table 2 Calibration data obtained with LC/ESP/MS in time-scheduled SIM/NI mode for the studied dyes

Compound	m/z	Calibration equation	r^2	Linear range (ng)	LOD (ng)	LOQ (ng)
Acid Red 1	464	$y = -6128.8 + 4336.6x$	0.989	20–600	20	60
	232	$y = -818.7 + 1700.4x$	0.966	60–600	60	150
Acid Red 13	228	$y = 6040 + 1700.4x$	0.996	3–900	3	11
Acid Red 14	228	$y = -14257.8 + 3810.2x$	0.983	3–800	3	11
Acid Red 73	511	$y = -29605.2 + 20762.9x$	0.984	12–800	12	50
Mordant Yellow 8	401	$y = -35766.5 + 24110.3x$	0.991	12–800	12	45
Acid Yellow 23	211	$y = -3752.5 + 1873.9x$	0.995	76–800	76	180
Direct Yellow 28	317	$y = -23062.1 + 5517.3x$	0.987	1.4–800	1	5
	318	$y = -6043.6 + 1657.7x$	0.988	2.6–800	3	10
Mordant Black 11	266	$y = -2392.5 + 5286.9x$	0.997	13–800	13	30
	294	$y = -2691.1 + 971.1x$	0.992	26–800	26	80
Mordant Black 17	221	$y = -10599.1 + 4267.3x$	0.991	70–700	70	160
	249	$y = -1926.4 + 989x$	0.996	70–700	70	160
Acid Blue 113	317	$y = -1638 + 18650.8x$	0.997	2–700	2	5
	318	$y = -336.8 + 24575.8x$	0.995	2–700	2	5

LOD, Limit of detection; LOQ, limit of quantification; SIM/NI, Selected ion monitoring/negative ionization.

Table 3 Calibration data obtained with LC/APCI/MS in time-scheduled SIM/PI mode for the studied dyes

Compound	m/z	Calibration equation	r^2	Linear range (ng)	LOD (ng)	LOQ (ng)
Mordant Red 9	247	$y = 5409.5 + 218.5x$	0.986	600–2600	600	1800
Acid Red 13	158	$y = -886.6 + 665.9x$	0.918	300–2000	300	1000
	297	$y = 3211.6 + 231.2x$	0.973	800–2000	800	2400
Acid Red 14	158	$y = 20822.3 + 1185.8x$	0.987	130–3200	130	800
	297		0.757	70–3200	70	210
	299	$y = 25893.3 + 837.6x$	0.999	800–3200	800	2400
		$y = 17869.6 + 112.4x$				
Acid Red 73	198	$y = 87152.8 + 20634x$	0.988	30–2000	30	200
Acid Yellow 23	218	$y = -5415.5 + 735.1x$	0.981	240–2000	240	720
Direct Yellow 28	241	$y = -33352.3 + 4269x$	0.996	240–2200	240	1000
	321	$y = 7313.1 + 213.75x$	0.987	800–2200	800	>2000
Acid Blue 113	231	$y = 4284 + 2473x$	0.988	100–2300	100	400
	159	$y = 20442.4 + 931.5x$	0.946	600–2300	600	1800

SIM/PI, Selected ion monitoring/positive ionization.

dyes in this study were acid compounds and only the negative-ion mode of detection was used. Additionally, calibration graphs with acceptable correlation coefficients were obtained, demonstrating that quantification is feasible with LC/API/MS techniques. Calibration equations for ISP interfaces in the range 1–80 ng and 600–800 ng, depending on the compound, were calculated. The authors also studied the calibration graphs by using a heated pneumatic nebulizer interface (or atmospheric pressure chemical ionization (APCI) interface). This interface provides more structural information than ion evaporation, and can handle chromatographic flow rates up to 1–2 mL min^{-1}. The technique was found to be less sensitive than the ISP interface with the exception of Mordant Red 9, a sulfonated dye that exhibits COONa and OH groups that give the molecule a much more basic character and showed better sensitivity in the PI mode. Tables 2 and 3 show the calibration data and detection limits for the studied dyes obtained with LC/ESP/MS and LC/APCI/MS in the PI mode. Figure 4 shows the LC/MS traces using LC/APCI/MS (in PI and NI mode) and LC/ISP/MS interfaces. Poor sensitivity, lower than 20–50 times, in less volatile disulfonated azo dyes versus monosulfonated dyes were reported in previous studies.[22] A better nebulization in the interface described by Rafols and Barceló[14] was attributed to the differences. The sensitivities of disulfonated dyes were lower than those of monosulfonated ones by a factor of 10 for ISP in both studies.[14,22]

Ballantine et al.[23] found that amine bases sensitize sulfonated azo salts towards ESP/MS. This has the effect of increasing the intensity of these ions by one to two orders of magnitude. It is proposed that the base enhances the removal of associated cations. This combines any ion peak series (from the different number of removed cations) into a single acid ion peak for each charge state (z), i.e. only the $[M - t\mathrm{Na} + (t + z)\mathrm{H}]/|z|$ ions would be observed. Once all cations have been removed, protons may associate with the compound to balance the removal of positive charges. The number of protons observed to associate with an ion depends on the final overall ion charge (z). The cation removal is not fully understood, but is thought to be influenced by the gas-phase proton affinity.

A relevant study was published by Lee and Henion[24] They used thermally assisted ESP to detect a hexasulfonated dye (Direct Red 80) of molecular weight 1240. The mass spectrum showed only multiply charged ions ranging from $[M - 6H]^{6-}$ to $[M - 2H]^{2-}$. They also worked on the determination of several polysulfonated azo dyes[25] by LC/ISP/MS which produced abundant $[M - Na]^-$ and $[M - 2Na]^{2-}$ ions with good sensitivity according to the number of sulfonic acid substituents.

In summary, the use of ESP in the NI mode offers higher sensitivity for sulfonated azo dyes than other techniques, e.g. APCI. Although APCI achieves more fragmentation, negative ESP ionization offers enough structural information which can be obtained either by increasing the extraction voltage or by using MS/MS.

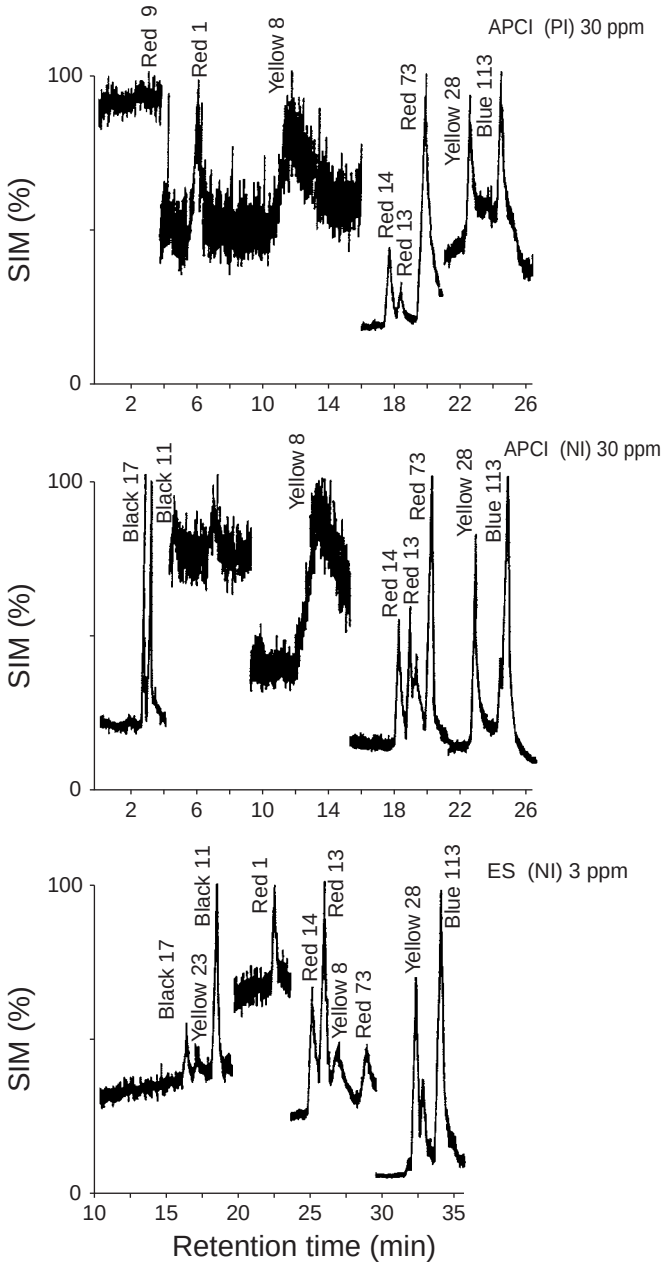

Figure 4 LC/APCI/MS (PI and NI) and LC/ESP/MS (NI) and time-scheduled selected ion monitoring (SIM) conditions of a mixture of dyes at 30 and 3 ppm, respectively.

3 CAPILLARY ELECTROPHORESIS

Analytes for CE include ionic compounds and nonvolatile analytes such as dyes, especially those that are not

amenable to GC and that offer problems with LC separations, such as certain dyes. Nonvolatiles of interest in the environment include ecological indicators of stress or health, biomarkers of exposure, and surfactant molecules whose metabolites act as hormone modulators that can affect wildlife and humans. A closed-circle dilemma derives from the knowledge that many of these compounds are not regulated, and therefore there is no urgent need for their monitoring. Conversely, such compounds are not monitored only because effective methods for their identification and quantitation do not exist. A partial resolution to this circle is to develop research methods for these problematic compounds to assess their occurrence, action, and fate in the environment. Such data would ultimately serve as the basis for regulation and risk reduction. Sulfonated dyes are not regulated and the use of CE can be of help in developing a routine analytical technique for monitoring them.

Recent applications of CE in the analysis of dyes are discussed here.

3.1 Capillary Electrophoresis/Ultraviolet Detection

In general, CE separation is based on the different electrophoretic mobilities of ions, and can be influenced by several parameters, e.g. pH and electrolyte concentration. In the case of dyes, the separation is often difficult to accomplish by varying the pH because they may have moieties that are strong acids. Therefore, two methods have been applied to adjust the electrophoretic mobility for the separation of six sulfonated azo dyes and other related compounds.[26] Complexation by bis-tris-propane (BTP) and interaction with linear polymers added to the buffer and acting as pseudophases were performed. A buffer system based on BTP containing poly(ethylene glycol) (PEG) and polyvinylpyrrolidone permitted the separation of all analytes. Retention of the dyes caused by the polymeric additives was related to the analyte structure. It was demonstrated that the relative decrease in the electrophoretic mobility of the dyes correlates with the number of benzoaromatic rings in the molecules. Other buffers with different micellar agents, e.g. borate at pH 8.3 with cholic acid, were employed for the determination of several sulfonated azo dyes and other classes of compounds.[27] Analysis using SPE of spiked water samples and recovery studies were also carried out. Burkinshaw et al.[28] used a micellar buffer system consisting of 10 mM $Na_2B_4O_7$ and 40 mM sodium dodecyl sulfate (SDS) at pH 9 to separate two sulfonated azo dyes with very similar structures and relative molecular weights, and also several anionic and cationic dye intermediates which could not be separated by a conventional buffer. Suzuki et al.[29] also used 10 mM SDS as micellar agent in a mixture of 25 mM sodium phosphate buffer and 25 mM sodium borate buffer (1 : 1) at pH 8.0 for the analysis of sulfonated azo dyes and xanthene dyes used as food additives in Japan. However, the isomeric dyes R-2 and R-102 were not separated with good resolution, and β-cyclodextrin instead of SDS was added to achieve a good separation. Although most of the authors reported that separation of dyes by free capillary zone electrophoresis (CZE) is not possible and micellar electrokinetic capillary chromatography (MEKC) has to be applied, in some cases these compounds have been successfully separated by CZE.[30]

Quantification studies were also done by Liu et al.[31] using a CZE method for the separation of six synthetic food colorants including four sulfonated azo dyes and one arylmethane dye. A linear relationship between the standard concentration and peak area (PA) of each of these dyes was obtained in the concentration range 2–50 ppm, with a correlation coefficient >0.995. The relative standard deviation (RSD) of the method was about 3% and the minimum detectable amount at a signal-to-noise ratio (S/N) of 3 for all six synthetic colorants was 3 ppm.

Calibration graphs for standard solutions of eight sulfonated azo dyes and for extracted groundwater samples spiked with the same compounds were also obtained by Schönsee et al.[32] Additionally, to evaluate the performance of the automated off-line SPE/CE/UV method, groundwater samples were spiked with mixtures of the eight dyes in the range 0.05–0.25 mg L^{-1}. The preconcentration procedure was carried out with a volume of 300 mL in order to achieve a calibration range from 10 to 50 mg L^{-1}. The LODs for the dyes ranged from 11 to 300 ppb. The regression equations were characterized by correlation coefficients > 0.95, and better LODs could be calculated, for a S/N of 3, for concentrations between 10 and 150 μg L^{-1}. Figure 5 shows the separation of the target compounds in spiked groundwater (3 mg L^{-1}); after off-line SPE, this allows the determination of all the dyes.

Detection limits between 11 and 300 ppb for 11 synthetic food colorants (including some sulfonated azo dyes) were achieved using CZE with the host–guest complexation effects of β-cyclodextrin.[33] A 300 μm i.d. capillary tube made of fluorinated ethylene–propylene copolymer in a hydrodynamically closed separation compartment was used for the analysis, accommodating 90-nL sample injection volumes, thus providing the above-mentioned detection limits. This CZE procedure was applied to several samples including a soft drink concentrate.

CE is especially useful for analyzing certain reactive dyes that HPLC often has difficulty in analyzing, such as the bifunctional β-sulfatoethylsulfone reactive dyes

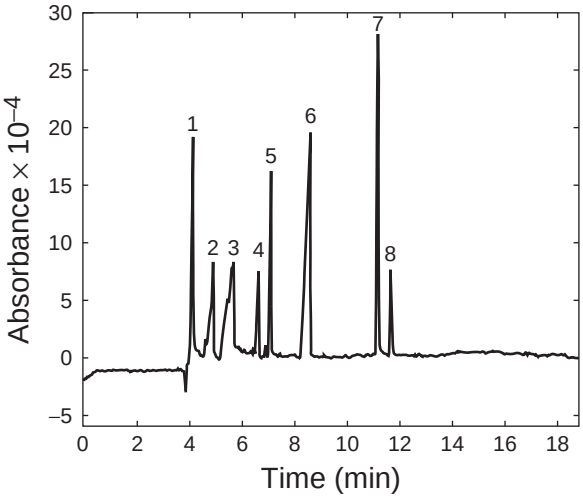

Figure 5 Electropherogram (214 nm) of the separation of eight sulfonated azo dyes spiked in groundwater (3 mg L^{-1}) after off-line SPE with Isolute ENV+ cartridges. Separation was carried out with a buffer solution of 9.2 mM ammonium acetate and 0.05% Brij 35. Peaks: (1) Acid Blue 113; (2) Acid Red 73; (3) Acid Red 13; (4) Mordant Yellow 8; (5) Acid Red 1; (6) Acid Red 14; (7) Mordant Red 9; (8) Acid Yellow 23.

and the phthalocyanine-based dyes. CE was employed in testing the purity of reactive dye samples, monitoring the reactions (kinetic studies) of reactive dyes with nucleophiles, or analyzing colored effluents.[34] A variety of buffers were investigated and the optimum was found to be acetonitrile at a ratio of 1 : 9 in a micellar buffer system (10 mM SDS, 10 mM sodium tetraborate, and 6 mM potassium dihydrogenphosphate at pH 9.0). Detection was effected with an on-line ultraviolet/visible (UV/VIS) detector positioned at the cathode. A CZE method with UV detection was developed for screening black reactive dyes and black acid dyes, isolated from cotton and wool materials.[35] The dye components were identified by using a newly developed marker technique, employing phenylacetic acid, benzoic acid, and *meso*-2,3-diphenylsuccinic acid. With this marker technique the RSDs of the migration indices and the electrophoretic mobilities for the analytes were improved and were below 0.6%.

Other reactive textile dyes such as chlorotriazine and vinylsulfone dyes, and their hydrolysis products, have also been analyzed in several commercial dyebath samples and sewage effluents from a water-treatment plant.[36] In these samples the majority of the peaks could not be assigned to any known compound and it was concluded that it would be necessary to couple CE with MS.

3.2 Capillary Electrophoresis/Mass Spectrometry

A comparative study of capillary liquid chromatography (CLC) and CE on the separation of some sulfonated and other classes of dyes has been carried out.[37] MEKC provided excellent separation and quantification of synthetic dyes and was found to be complementary to reversed-phase CLC in the separation and in their ability to quantify and identify analytes. Recovery data for water and soil extraction were obtained by MEKC/UV detection. Although CLC was coupled with continuous-flow liquid secondary ion mass spectrometry (CFLSI/MS) for confirmation purposes, interfacing CE with the MS instrument using a coaxial arrangement as used in CLC was more difficult and only UV detection with CE was performed.

Ion evaporation appears to be the preferred ionization mechanism for on-line CZE/MS because separated components already exist as charged species in the CZE buffer. Furthermore, this ionization process is best-suited for species which exist as ions in solution, such as sulfonated azo dyes.

Continuous-flow (CF) FAB has been coupled to CE[38] with a coaxial interface to analyze aromatic sulfonic acids and some azo dyes. However, excessive band broadening and loss of separation efficiency occurred owing to the long transfer line to the FAB ion source and the high-vacuum requirements. The newly developed atmospheric pressure interfaces such as ISP produce a mild form of ionization described by Thomson and Iribarne;[39] it was first coupled to CE by Henion et al.[40,41] with a liquid-junction interface for determining sulfonated azo dyes. Detection limits range from high parts per billion to low parts per million. The correlation coefficient for the linear regression through the means of four levels of concentration of the dye injected was found to be 0.990 and the reproducibility for automatic injection could routinely be within 2–4%. The mass spectra of the azo dyes gave peaks $[M - nH]^{n-}$, depending on the number of sulfonic groups. CE tandem mass spectra of the deprotonated molecular parent ion showed a daughter ion at m/z 80 corresponding to the sulfonated ion. Three sulfonated azo dyes were detected at low parts per million in spiked wastewater extract monitoring the sulfonate ion at m/z 80.

Eight mono- and disulfonated azo dyes have recently been analyzed with a Beckman P/ACE CE system connected with a Micromass interface to a VG Platform MS instrument.[42] Spiked and extracted water samples were separated and determined for confirmation of the studied compounds and a typical chromatogram of the analysis of an extracted sample is shown in Figure 6. Water samples were analyzed at the 3 ppm level by monitoring one ion for each studied compound in order to achieve maximum sensitivity. Detection limits between 100 and 800 ppb were achieved, with the exception of Acid Red 73, which exhibited low response and could not be quantified. Compounds which co-eluted and were not resolved could be determined because ions with different m/z ratios were obtained. The spectra obtained

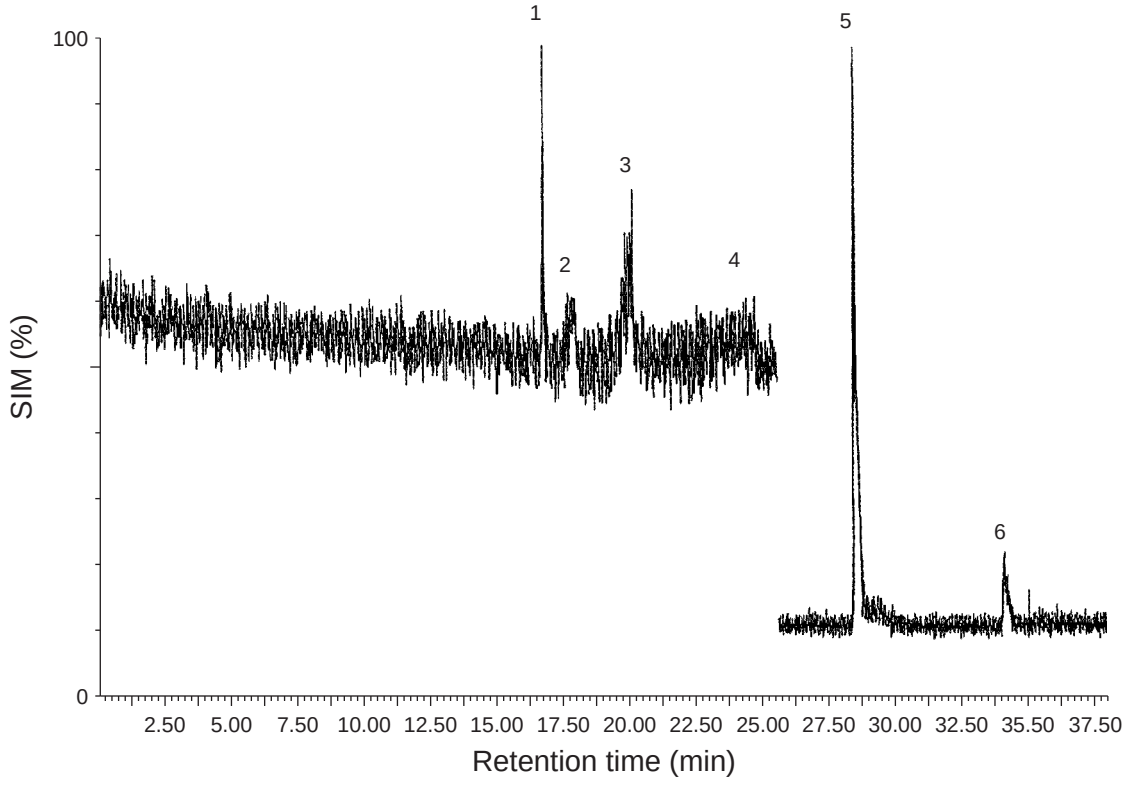

Figure 6 Total ion chromatogram (TIC) of a CE/MS electropherogram for an extracted spiked (3 ppm) water sample. Peaks: (1) internal standard; (2) Acid Red 73; (3) Acid Red 1 + Mordant Yellow 8; (4) Acid Red 13; (5) Acid Red 14; (6) Mordant Red 9.

by CE/MS were similar to those obtained with ISP ($[M - tNa + (t + z)H]$ ions, depending on the number of sulfonated groups (n) and the charge (z), and some losses of SO_3^- and COO^- groups). In a previous study in the same laboratory,[14] the spectra of one of the studied dyes were compared using both interfaces (Figure 7).

In summary, CE offers good resolution and better efficiencies than LC; the use of MEKC gives a wide variety of separations of components. This enables one to analyze neutral compounds in addition to strong acids and very polar compounds that are not amenable to GC or LC. However, further optimization is required when coupling CE with MS. Detection limits are less satisfactory than those obtained using LC/MS interfaces because of the low CE injection volumes.

4 SAMPLE HANDLING

Nielen has reviewed sample handling which included SPE from aqueous matrices containing acidic herbicides for analysis by CE.[43]

Because extremely small volumes of solutions are injected in CE (nominally picoliters to tens of microliters), preconcentration, cleanup, and on-line sample concentration must be addressed in the context of the analysis. By means of field amplification techniques, relatively large volumes can be injected without introducing intolerable peak broadening.

SPE has been used for preconcentration and cleanup of sulfonated azo dyes by several authors. Brumley and Brownrigg[27] extracted two sulfonated azo dyes (mono- and polysulfonated) and several aromatic-containing organic acids by means of Empore poly(styrene–divinylbenzene) (SDB) extraction disks and SPE cartridges (SCX cartridges). Spiked water samples were extracted by SDB disks, eluted in two ways (with organic aqueous solvent and with an ion-pairing agent) and the fractions obtained, were passed through the SCX cartridge. Extractions of solid samples were carried out by either sonification or Soxhlet methods; subsequently, the extracts were treated like the water samples. The recovery for the polysulfonated dye Triptan Blue from soil was only about 28% and was attributed to the complex structure of the dye. From water, this compound showed higher recoveries. In contrast, the monosulfonated dye Orange II could be extracted with good recoveries from both water and soil samples.

Other types of solid phases have been employed for sulfonated dyes, such as C_{18} cartridges and columns[22,25] that were eluted with aqueous organic solvents. C_{18} cartridges

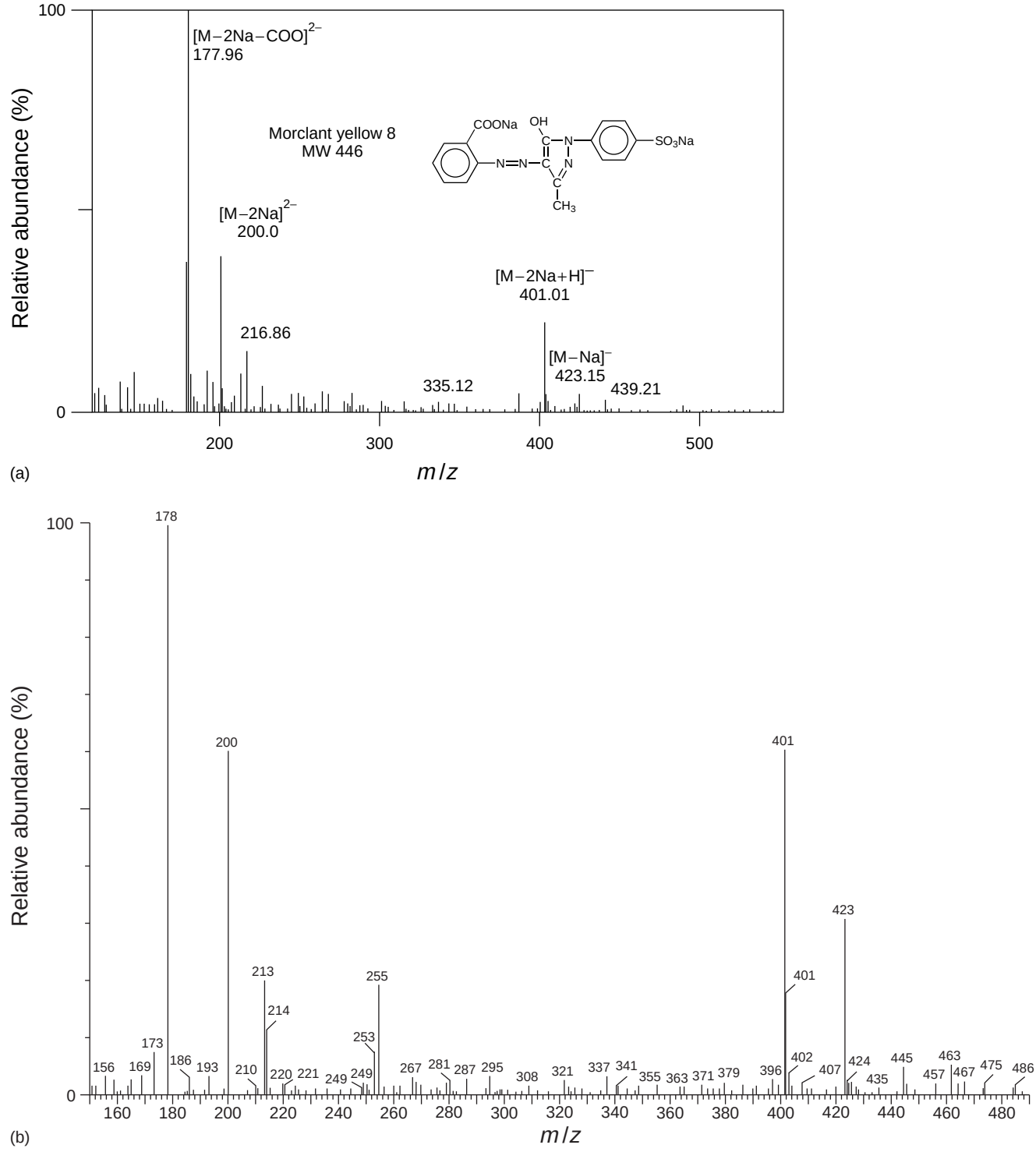

Figure 7 Mass spectra of Mordant Yellow 8 with (a) LC/ISP/MS interface and (b) CZE/MS interface.

from J.T. Baker, which were eluted with solvent mixtures of methanol–water–triethylamine (TEA), gave recoveries that ranged from 87 to 103% for monosulfonated azo dyes and from 65 to 105% for disulfonated azo dyes.[22] With the C_{18} columns from J.T. Baker, eluted with methanol after preconditioning with ammonium acetate buffer, recoveries that ranged from 70 to 122%[25] were achieved.

The application of copolymers such as SDB was found to be most suitable for the SPE of very polar

Table 4 Recoveries obtained from 300 and 500 mL preconcentrated spiked groundwater samples (3 mg L^{-1}) by SPE with LiChrolut EN and Isolute ENV+ sorbents

Compound	LiChrolut EN			Isolute ENV+	
	100% MeOH 0.01% TEA	90% MeOH 10% H_2O 0.01% TEA	70% MeOH 30% H_2O 0.01% TEA	100% MeOH 0.01% TEA	80% MeOH 20% H_2O 0.2% TEA[c]
Acid Blue 113	N.D.[a]	N.D.	N.D.	N.D.	34 ± 7
Acid Red 73	39 ± 2	39 ± 1	32 ± 2	55 ± 4	83 ± 14
Acid Red 13	36 ± 2	40 ± 1	32 ± 1	71 ± 7	80 ± 11
Mordant Yellow 8	47 ± 3	49 ± 1	42 ± 4	64 ± 3	69 ± 7
Acid Red 1	19 ± 1	17 ± 1	20 ± 1	53 ± 4	69 ± 2
Acid Red 14	31 ± 3	34 ± 3	27 ± 1	68 ± 2	74 ± 7
Acid Red 9	16 ± 2	18 ± 1	17 ± 1	51 ± 5	64 ± 1
Acid Yellow 23	n.d.[b]	n.d.	n.d.	20 ± 3	15 ± 2

[a] N.D. = not determined.
[b] n.d. = not detected.
[c] Preconcentration of 500 mL of sample.

compounds.[44] Cross-linked polymeric sorbents with sulfonic functions, such as Isolute ENV+ and LiChrolut EN, were examined for their suitability for off-line preconcentration of the studied sulfonated azo dyes.[32] The recovery and standard deviation data obtained on processing 300 and 500 mL of spiked groundwater samples are reported in Table 4. Preconcentration with Isolute ENV+ using 100% MeOH (0.01%TEA) as eluent achieved recoveries up to 71%. The application of Isolute ENV+ for further optimizations was found to be more effective. Although the recoveries of Acid Yellow 23 and Acid Blue 113 were below 50%, for six dyes the recoveries could be improved to 64–83% by eluting with a mixture of 80% MeOH, 20% water, and 0.2% TEA.

More sorbents should be tested for extraction and preconcentration of dyes in various environmental water matrices. Polymeric sorbents with different polymer functionalities, e.g. sulfonation, seem to work better than silica-based phases. The development of immunoaffinity sorbents may lead to a solution to the problem of low recoveries for some of the more polar and ionic dyes. In general, it would be necessary to find sorbents that allow the percolation of large volumes of up to 1–2 L in order to reduce the LODs of CE techniques.

5 CONCLUSIONS

LC/MS has been shown to be a good method for the determination of dyes. In this respect TSP has worked well for several classes of these compounds, but it has shown poor sensitivity for sulfonated azo dyes. Some techniques were applied to improve the detection limits of these compounds, such as the use of a wire repeller or restricting the vaporizer exit, but the development of LC/API/MS interfaces, which provide more structural information and high sensitivity, can lead to more extensive use in environmental analysis,

Table 5 LODs of sulfonated dyes using different interfaces

PB	TSP	ISP	APCI	CE/UV	CE/MS	Ref.
	10 ng[a,c]					14
	5–20 ng[b,c]					17
	4000 ng[a]	1–80 ng[b]	PI 30–800 ng[b]			15
			NI 50–700 ng[b]			
500 ng–5 μg[a]			10 ng[a]			6
		10–500 ng[a]	100 ng–2 μg[a]			22
				1–6 ppm		33
				11–300 ppb[d]		34
					High ppb–low ppm[d]	41
					High ppb–low ppm[d]	43

[a] Full scan.
[b] SIM.
[c] Modification of the source or the probe.
[d] Combined with preconcentration.

replacing classical LC/MS interfaces. In particular, ISP can provide, with the manipulation of the cone voltage or combination with MS/MS, enough structural information and selectivity.

CE, with its different modes, shows several advantages for polar analytes and can be easily coupled with classical ESP interfaces because of the restricted flow rates. Sensitivity problems result from the small injection volumes and, therefore, preconcentration techniques should be applied. Limits of detection for the various techniques reported in this paper are shown in Table 5. More work needs to be done in the future and the focus will likely be towards the application of CE/MS combined with SPE. Owing to the high efficiency and diversity of separation in CE, and the fact that it is inexpensive and a low generator of waste, the technique will permit the required analytical performance for determining sulfonated azo dyes in environmental waters to be achieved.

ACKNOWLEDGMENTS

This work was supported by the Commission of the European Communities, Environment and Climate Program 1994–98 (Contract Number ENV4-CT97-0475) and by CICYT (AMB98-913-CO3).

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
BTP	Bis-tris-propane
CAD	Collision Activated Dissociation
CE	Capillary Electrophoresis
CF	Continuous-flow
CFLSI/MS	Continuous-flow Liquid Secondary Ion Mass Spectrometry
CID	Collisionally Induced Dissociation
CLC	Capillary Liquid Chromatography
CZE	Capillary Zone Electrophoresis
EI	Electron Ionization
ESP	Electrospray
ETI	Environmental Technology Initiative
FAB	Fast Atom Bombardment
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
ISP	Ionspray
LC	Liquid Chromatography
LOD	Limit of Detection
LOQ	Limit of Quantification
MEKC	Micellar Electrokinetic Capillary Chromatography
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NI	Negative Ionization
PA	Peak Area
PB	Particle Beam
PEG	Poly(ethylene Glycol)
PI	Positive Ionization
RSD	Relative Standard Deviation
SDB	poly(styrene–divinylbenzene)
SDS	Sodium Dodecyl Sulfate
SIM	Selected Ion Monitoring
SIM/NI	Selected Ion Monitoring/ Negative Ionization
SIM/PI	Selected Ion Monitoring/ Positive Ionization
S/N	Signal-to-noise Ratio
SPE	Solid-phase Extraction
TEA	Triethylamine
TIC	Total Ion Chromatogram
TSP	Thermospray
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Gas Chromatography with Selective Detectors for Amines • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-assisted Techniques for

Sample Preparation in Organic Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Solid-phase Microextraction in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Food **(Volume 5)**
Electrophoresis and Isoelectric Focusing in Food Analysis

Pesticides **(Volume 7)**
Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Micellar Electrokinetic Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. M.S. Reich, 'Asian Textile Dye Markes are a Growing Power in a Changing Market', *Chem. Eng. News*, Jan. 15, 10–12 (1996).
2. E.J. Weber, R.L. Adams, 'Chemical-mediated and Sediment-mediated Reduction of the Azo-dye Disperse-blue-79', *Environ. Sci. Technol.*, **29**, 1163–1170 (1995).
3. E. Rindle, W.J. Troll, 'Metabolic Reduction of Benzidine Azo Dyes to Benzidine in the Rhesus Monkey', *J. Natl. Cancer Inst.*, **55**, 181 (1975).
4. USEPA Method 8321A, 'Solvent Extractable Nonvolatile Compounds by HPLC/TS/MS or UV Detection', USEPA, Washington DC, 1–50, 1995.
5. J.M. Ballard, L.D. Betowski, 'Thermospray Ionization and Tandem Mass Spectrometry of Dyes', *Org. Mass Spectrom.*, **21**, 575–578 (1986).
6. W.C. Brumley, W. Witnnik, in *Applications of LC/MS in Environmental Chemistry*, ed. D. Barceló, Elsevier, Amsterdam, 481–527, 1996.
7. R. Straub, R.D. Voyksner, J.T. Keever, 'Thermospray, Particle Beam and Electrospray Liquid Chromatography/Mass Spectrometry of Azo Dyes', *J. Chromatogr.*, **627**, 173–186 (1992).
8. R.D. Voyksner, R. Straub, J.T. Keever, H.S. Freeman, W.-N. Hsu, 'Determination of Aromatic-amines Originating from Azo Dyes by Chemical-reduction Combined with Liquid Chromatography/Mass Spectrometry', *Environ. Sci. Technol.*, **27**, 1665–1672 (1993).
9. F. Ventura, A. Figueras, J. Caixach, D. Fraisse, J. Rivera, 'FAB Mass Spectrometry of Sulphonated Dyes. Application to Water Analysis', *Fresenius' J. Anal. Chem.*, **335**, 272–278 (1989).
10. R.D. Voyksner, 'Characterization of Dyes by Thermospray High-performance Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **57**, 2600–2605 (1985).
11. D.A. Flory, M.M. McLean, M.L. Vestal, L.D. Betowski, 'Environmental Applications of Thermospray LCMS: Qualitative Analysis of Sulfonated Azo Dyes', *Rapid Commun. Mass Spectrom.*, **1**, 48–50 (1987).
12. J. Yinon, T.L. Jones, L.D. Betowski, 'Enhanced Sensitivity in Liquid Chromatography/Thermospray Mass Spectrometry of Dyes using a Wire Repeller', *Rapid Commun. Mass Spectrom.*, **3**, 38 (1989).
13. J. Yinon, T.L. Jones, L.D. Betowski, 'High-sensitivity Thermospray Ionization Mass Spectrometry of Dyes', *Biomed. Environ. Mass Spectrom.*, **18**, 445–449 (1989).
14. C. Rafols, D. Barceló, 'Determination of Mono and Disulphonated Azo Dyes by Liquid Chromatography/Atmospheric Pressure Ionization/Mass Spectrometry', *J. Chromatogr., A*, **777**, 177–192 (1997).
15. J. Yinon, T.L. Jones, L.D. Betowski, 'Thermospray Ionization with Repeller-induced Collisionally Activated Dissociation of Dyes', *Rapid Commun. Mass Spectrom.*, **4**, 245 (1990).
16. M.A. McLean, R.B. Freas, 'Enhanced Analysis of Sulfonated Azo Dyes using Liquid Chromatography/Thermospray Mass Spectrometry', *Anal. Chem.*, **61**, 2054–2058 (1989).
17. L.D. Betowski, S.M. Pyle, J.M. Ballard, G.M. Shaul, 'Thermospray LC/MS/MS Analysis of Wastewater for Disperse Azo Dyes', *Biomed. Environ. Mass Spectrom.*, **14**, 343–354 (1987).
18. J.F. Garcia, D. Barceló, 'An Overview of LC/MS Interfacing Systems with Selected Applications', *J. High. Resolut. Chromatogr.*, **16**, 633–641 (1993).
19. W.M.A. Niessen, U.R. Tjaden, J. Van der Greef, 'Strategies in Developing Interfaces for Coupling Liquid Chromatography and Mass Spectrometry', *J. Chromatogr.*, **554**, 3–26 (1991).

20. E.C. Huang, T. Wachs, J.J. Conboy, J.D. Henion, 'Atmospheric Pressure-ionization Mass Spectrometric-detection for the Separation Sciences', *Anal. Chem.*, **62**, 713 (1990).
21. C.M. Whitehouse, R.N. Dreyes, M. Yamashita, J.B. Fenn, 'Electrospray Interface for Liquid Chromatographs and Mass Spectrometers', *Anal. Chem.*, **57**, 675–679 (1985).
22. A.P. Bruins, L.O.G. Weidolf, J.D. Henion, W.L. Budde, 'Ion Spray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 2647–2652 (1987).
23. J.A. Ballantine, D.E. Games, P.S. Slater, 'The Use of Amine Base to Enhance the Sensitivity of Electrospray Mass Spectrometry towards Complex Polysulphonated Azo Dyes', *Rapid Commun. Mass Spectrom.*, **9**, 1403–1410 (1995).
24. E.D. Lee, J.D. Henion, 'Thermally Assisted Electrospray Interface for Liquid Chromatography/Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 727–733 (1992).
25. P.O. Edlund, E.D. Lee, J.D. Henion, W.L. Budde, 'The Determination of Sulfonated Azo Dyes in Municipal Wastewater by Ion Spray Liquid Chromatography Tandem Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **18**, 233–240 (1989).
26. P. Blanty, C.H. Fisher, A. Rizzi, E. Kendler, 'Linear Polymers Applied as Pseudophases in Capillary Zone Electrophoresis of Azo Compounds used as Textile Dyes', *J. Chromatogr., A*, **717**, 157–166 (1995).
27. W.C. Brumley, C.M. Brownrigg, 'Electrophoretic Behavior of Aromatic-containing Organic Acids and the Determination of Selected Compounds in Water Soil by Capillary Electrophoresis', *J. Chromatogr.*, **646**, 377–389 (1993).
28. S.M. Burkinshaw, D. Hinks, D.M. Lewis, 'Capillary Zone Electrophoresis in the Analysis of Dyes and Other Compounds Employed in the Dye-manufacturing and Dye-using Industries', *J. Chromatogr.*, **640**, 413–417 (1993).
29. S. Suzuki, M. Shirao, M. Aizawa, H. Nakazawa, K. Sasa, H. Sasagawa, 'Determination of Synthetic Food Dyes by Capillary Electrophoresis', *J. Chromatogr.*, **680**, 541–547 (1994).
30. S.N. Croft, D. Hinks, 'Analysis of Dyes by Capillary Electrophoresis', *JSDC*, **108**, 546–551 (1992).
31. H. Liu, T. Zhu, Y. Zhang, S. Qi, A. Huang, Y. Sun, 'Determination of Synthetic Colourant Food Additives by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **718**, 448–453 (1995).
32. I. Schönsee, J. Riu, D. Barceló, 'Determination of Sulfonated Azo Dyes in Water by Automated Solid Phase Extraction Followed by MEKC/UV', *Quim. Anal.*, **16**(Suppl. 2), S243–S249 (1997).
33. M. Masár, D. Kaniasky, V. Madajová, 'Separation of Synthetic Food Colourants by Capillary Electrophoresis in a Hydrodynamically Closed Separation Compartment', *J. Chromatogr., A*, **724**, 327–336 (1996).
34. K.N. Tapley, 'Capillary Electrophoretic Analysis of the Reactions of Bifunctional Reactive Dyes under Various Conditions Including a Study of the Analysis of the Traditionally Difficult to Analyse Phthalonine Dyes', *J. Chromatogr., A*, **706**, 555–562 (1995).
35. H. Sirén, R. Sulkava, 'Determination of Black Dyes from Cotton and Wool Fibres by Capillary Zone Electrophoresis with UV Detection Application of Marker Technique', *J. Chromatogr., A*, **717**, 149–155 (1995).
36. T. Poiger, G.L. Baughman, 'Separation and Identification of Reactive Dyes in Sewage Effluent' in preparation.
37. W.C. Brumley, C.M. Brownrigg, A.H. Grange, 'Capillary Liquid Chromatography/Mass Spectrometry and Micellar Electrokinetic Chromatography as Complementary Techniques in Environmental Analysis', *J. Chromatogr., A*, **680**, 635–644 (1994).
38. W.C. Brumley, 'Qualitative Analysis of Environmental Samples for Aromatic Sulfonic Acids by High-performance Capillary Electrophoresis', *J. Chromatogr.*, **603**, 267–272 (1992).
39. B.A. Thomson, J.V. Iribarne, 'Field Induced Ion Evaporation from Liquid Surfaces at Atmospheric Pressure', *J. Chem. Phys.*, **71**, 4451–4463 (1979).
40. E.D. Lee, W. Mück, T.R. Covey, J.D. Henion, 'Capillary Zone Electrophoresis/Tandem Mass Spectrometry for the Determination of Sulfonated Azo Dyes', *Biomed. Environ. Mass Spectrom.*, **18**, 253–257 (1989).
41. E.D. Lee, W. Mück, T.R. Covey, J.D. Henion, 'Liquid Junction Coupling for Capillary Zone Electrophoresis/Ion Spray Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **18**, 844–850 (1989).
42. J. Riu, I. Schönsee, D. Barceló, 'Determination of Sulfonated Azo Dyes in Groundwater and Industrial Effluents by Automated Solid-phase Extraction Followed by Capillary Electrospray/Mass Spectrometry', *J. Mass Spectrom.*, **33**, 653–663 (1998).
43. M.W.F. Nielen, 'Trace Enrichment of Environmental Samples', *Trends Anal. Chem.*, **12**, 345–356 (1993).
44. D. Puig, D. Barceló, 'Comparison of Different Sorbent Materials for Online Liquid–Solid Extraction Followed by Liquid Chromatographic Determination of Priority Phenolic-compounds in Environmental Waters', *J. Chromatogr.*, **733**, 371–381 (1996).

Explosives Analysis in the Environment

Bruce A. Tomkins
Oak Ridge National Laboratory, Oak Ridge, USA

1 Introduction **2**
1.1 Chemical Classes of Explosives 2
1.2 The Origin of Explosives in Contaminated Soil 2
1.3 The Origin of Explosives in Contaminated Groundwaters 4
1.4 Regulatory Limits for Explosives in Groundwater and Soil 4
2 Sampling Considerations **5**
2.1 Special Considerations for Soil Sampling 5
2.2 Maximum Holding Times for Soil and Aqueous Samples 7
3 Determination of Explosives in Aqueous Samples **8**
3.1 Procedures Employing "Salting-out" Liquid–Liquid Extraction Followed by High-pressure Liquid Chromatography 8
3.2 Procedures Employing Solid-phase Extraction 9
3.3 Determination of Explosives in Seawater 11
3.4 Additional Procedures 13
4 Determination of Explosives in Soil Samples **14**
4.1 Extraction Techniques for Explosives Residues in Soil: Initial Considerations 14
4.2 Experimental Considerations and Statistical Validation of the Proposed Standard Method 14
4.3 Performance of the Standard United States Environmental Protection Agency Method 8330 16
4.4 Additional Procedures 17
5 Determination of Explosives using Field Screening Procedures **18**
5.1 Introduction and Motivation 18
5.2 Description and Evaluation of Typical Colorimetric Field Screening Procedures 19
5.3 Description of Antibody-based Field Test Kits 20
5.4 Evaluation of Commercially Available Field Screening Kits 22
5.5 Field Test Procedures for Picric Acid and Ammonium Picrate 23
5.6 Field Screening Procedures for Explosives in Soils using Ion Mobility Spectrometry 23
6 Transformation and Metabolism of Explosives in the Environment **24**
6.1 Expected Reaction Pathways 24
6.2 Analytical Procedures for the Transformation Products 24
6.3 Experimental Evidence for the Three Transformation Pathways 25
7 Detection of Explosives in Air **26**
7.1 The Challenge of Detecting Pure Explosive Vapors in Air 26
7.2 Canine Detection of Explosives in Air 26
7.3 Mass Spectrometric Detection of Explosives in Air 27
7.4 Additional Procedures for Explosives in Air 28
8 Additional Procedures for the Determination of Explosives **28**
8.1 Procedures Featuring Chemiluminescent Nitrogen Detection of Explosives 29
8.2 Procedures Featuring Mass Spectrometric Detection of Explosives 29
8.3 Procedures Featuring Nontraditional Analytical Methods 31
9 Summary and Conclusions **31**
Acknowledgments **32**
List of Symbols **32**
Abbreviations and Acronyms **32**
Related Articles **33**
References **34**

The interest in analytical procedures for explosives residues in environmental media steadily increases as more and more formerly utilized military sites are certified and converted to full- or part-time civilian use. Careful statistical studies have clearly demonstrated that analyses based on randomly collected "grab" samples of soils frequently produce a false picture of the extent of explosives contamination. Well-designed sampling plans, which

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

employ multiple samples taken from small areas and which consider the distribution of the soil particles, are mandatory for understanding the true extent of contamination. Because site remediation is such a costly venture, there is a strong need for analytical procedures which are capable of quantitating explosives in soil and groundwater samples quickly and accurately, using equipment and methods which may be operated reliably by nontechnical personnel. For that reason, scientific interest is gradually moving away from the completely laboratory-based, statistically validated, and highly technical standardized methods, such as high-pressure liquid chromatography (HPLC) and gas chromatography (GC) and their mass spectroscopic analogs, towards the many fieldable analytical techniques. Some of these, such as ion mobility spectrometry and the various magnetic resonance techniques, are still in their infancy; their potential and utility are still to be fully developed. Others, such as the colorimetric and enzyme-based field test kits, have been fully tested and, in some cases, have been adopted as standard methods themselves. Still others, such as the various mass spectrometric methods, involve highly technical equipment which must be simplified before it can be used routinely in the field. However, the extreme sensitivities characteristic of these state-of-the-art instruments partially offset the need for extensive operator training. A variety of mass spectrometric techniques may yet be used routinely for explosives residue analysis in the field. Some of these new methods may be used in, or even be derived from, forensic applications.

1 INTRODUCTION

1.1 Chemical Classes of Explosives

Nitrogen-containing explosives and their transformation products detected in the environment are almost always of military origin and fall into three distinct classes. The *nitroaromatics* are related to 2,4,6-trinitrotoluene (TNT, CAS 118-96-7), and include the following species: 1,3,5-trinitrobenzene (1,3,5-TNB, CAS 99-35-4); 1,3-dinitrobenzene (1,3-DNB, CAS 99-65-0); methyl-2,4,6-trinitrophenyl nitramine (tetryl, CAS 479-45-8); 2,4-dinitrotoluene (2,4-DNT, CAS 121-14-2); 2,6-DNT, (CAS 606-20-2); 2-amino-4,6-dinitrotoluene (2-Am-DNT, CAS 33572-78-2); and 4-amino-2,6-dinitrotoluene (4-Am-DNT, CAS 19406-51-0). The *nitramines*, which are typically more energetic than the nitroaromatics, include the two heterocyclic compounds hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX, CAS 121-82-4) and octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX, CAS 2691-41-0). The *nitrate esters* are represented by nitroglycerine (NG, CAS 55-63-0) and pentaerythritol tetranitrate (PETN, CAS 78-11-5). The physical properties of these species are summarized in Table 1[1,2,3] while their structures are presented in Figure 1. The chemical names and common acronyms of additional explosives have been tabulated in Beveridge.[4]

Ammonium picrate (ammonium 2,4,6-trinitrophenoxide, CAS 131-74-8 AP) was the principal explosive used in a number of armor-piercing shells, bombs, and rocket warheads from the early 1900s until after World War II. While AP is no longer manufactured, wastes containing this component are still encountered at military installations in the United States. Picric acid (2,4,6-trinitrophenol, CAS 88-89-1 PA) is both an explosive and a common industrial chemical with uses that include dyeing fabric and leather, glass coloration, and metal etching.[5]

The scope of this review focuses on the analysis of *organic* explosives in the environment. A discussion of the properties and determination of the inorganic explosives, such as water gels and ammonium nitrate fuel oil, has been omitted intentionally.

1.2 The Origin of Explosives in Contaminated Soil

The contamination of soil with explosives frequently resulted from historical – and now both outdated and unapproved – disposal practices from manufacturing, spills, ordnance demilitarization, lagoon disposal of explosives-contaminated wastewater, and open burn/open detonation (OB/OD) of sludges from explosives processing, waste explosives, excess propellants, and unexploded ordnance. Facilities that may be contaminated with explosives include active and former manufacturing plants, ordnance works, army ammunition plants, naval ordnance plants, army depots, naval ammunition depots, army and naval proving grounds, burning grounds, artillery impact ranges, explosive ordnance disposal (EOD) sites, bombing ranges, firing ranges, ordnance test and evaluation facilities, etc. A number of these facilities have high levels of soil and groundwater contamination, although waste disposal was discontinued 20 to 50 years ago. Common munitions fillers and their associated secondary explosives include Amatol (ammonium nitrate/TNT), Baratol (barium nitrate/TNT), Cyclonite or Hexogen (RDX), Cyclotols (RDX/TNT), Composition A-3 (RDX), Composition B (TNT/RDX), Composition C-4 (RDX), Explosive D or Yellow D (AP/PA), Octogen (HMX), Octols (HMX/TNT), Pentolite (PETN/TNT), Picratol (AP/TNT), Tritonal (TNT), Tetrytols (tetryl/TNT), and Torpex (RDX/TNT).[6] The United States has characterized 30 large and medium sized TNT-contaminated sites at federal facilities; together, they contain over one million cubic yards of contaminated soil. The total number of smaller sites contaminated with TNT is not yet known,

Table 1 Physical and chemical properties of the major nitroaromatics and nitramines

Analyte	CAS reg. no.	Molecular weight	Melting point (°C)	Boiling point (°C)	Water solubility, (mg L^{-1})	Vapor pressure at 20 °C (torr)	Log K_{ow}^{a}	Henry's Law constant H_c (torr M^{-1})
TNT	118-96-7	227.13	80.1–81.6	240 (explodes)	130 @ 20 °C	1.1 E – 06	1.86	0.18
RDX	121-82-4	222.26	204.1	Decomposes	42 @ 20 °C	4.1 E – 09	0.86	2 E – 05
HMX	2691-41-0	296.16	276–280	Decomposes	5 @ 20 °C	3.3 E – 14	0.061	–
1,3,5-TNB	99-35-4	213.11	122.5	315	278 @ 15 °C	2.2 E – 04	1.18	1.5
1,3-DNB	99-65-0	168.11	89.6	300–303	460 @ 15 °C	3.9 E – 03	1.49	1.8
Tetryl	479-45-8	278.14	129.5	Decomposes	80	5.7 E – 09[b]	1.65	–
2,4-DNT	121-14-2	182.15	70	300 (decomposes)	270 @ 22 °C	2.2 E – 04[b]	1.98	3.4
2,6-DNT	606-20-2	182.15	64–66	–	206 @ 25 °C	5.7 E – 09	2.02	18
2-Am-DNT	35572-78-2	197.17	176	–	2800 @ 20 °C	4.0 E – 05	1.94	3 E – 03
4-Am-DNT	19406-51-0	197.17	171	–	2800 @ 20 °C	2.0 E – 05	1.91	1 E – 03

(Reproduced by permission of the Cold Regions Research and Engineering Laboratory, M.E. Walsh, T.F. Jenkins, P.S. Schnitker, J.W. Elwell, M.H. Stutz, *Evaluation of SW846 Method 8330 for Characterization of Sites Contaminated with Residues of High Explosives*, CRREL Report 93-5, 4, 1993.)

[a] Octanol–water partition coefficient.

[b] Measured at 25 °C.

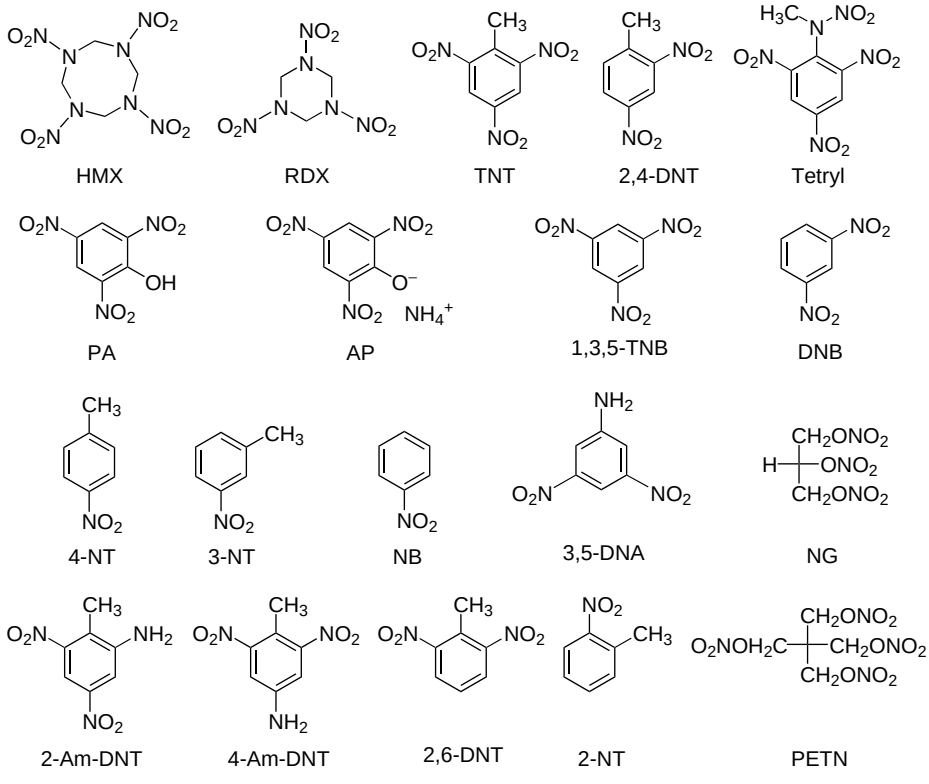

Figure 1 Structures of the nitramine, nitroaromatic, nitrate esters, and picrate explosives analyzed in the environment.

but it is reported that the army alone has on the order of 2000 explosive-contaminated sites that may require remediation,[7] and more exists at navy and air force facilities. Similar patterns of contamination are found throughout Europe and the former Soviet Union.[8] Evaluating sites potentially contaminated with explosives is necessary to carry out US Department of Defense, United States Environmental Protection Agency (USEPA), and US Department of Energy policies on-site characterization and remediation under the Superfund, Resource Conservation and Recovery Act (RCRA), Installation Restoration, Base Realignment and Closure (BRAC), and Formerly Utilized Defense Site (FUDS) environmental programs.

1.3 The Origin of Explosives in Contaminated Groundwaters

Groundwater may be contaminated with explosives in two distinct ways. The first is a direct consequence of the disposal of munitions wastes during World War I, World War II, the Korean War, and the Viet Nam War. Unlined lagoons, which could not contain species such as munition charges, were frequently employed. Leached mobile contaminants migrated with groundwater at several installations. The second involved the disposal of washwaters used to clean equipment and interior surfaces at munition manufacturing and demilitarization facilities. It has been estimated that up to 2×10^6 L of this type of wastewater is generated daily from a single production line. Current practice is to collect washwater from these processing operations in a holding tank and periodically pump it through a series of carbon adsorption columns prior to discharge to surface streams. Standard practice allows monitoring of explosives concentrations between columns to prevent contamination downstream. These point discharges are subject to state and federal National Pollutant Discharge Elimination System (NPDES) permits that generally limit the acceptable concentrations of TNT and RDX. Although carbon adsorption technology can reduce TNT and RDX to low part per billion levels, which meet present discharge limitations, these carbon columns have finite lifetimes. Eventually, breakthrough occurs and regeneration or replacement is necessary. Ground- and surface-water contamination may therefore occur immediately prior to replacement of the carbon adsorption water treatment columns.[9]

1.4 Regulatory Limits for Explosives in Groundwater and Soil

The USEPA, in collaboration with the Oak Ridge National Laboratory, has published a series of health advisories and recommended water criteria for several explosives-related chemicals, as summarized in Table 2.[2,3,10] These values for the nitroaromatics range over two orders of magnitude between 2,6-DNT ($0.0068\,\mu g\,L^{-1}$ water)[11] to TNT ($1.0\,\mu g\,L^{-1}$ water).[12] If the criterion for HMX ($400\,\mu g\,L^{-1}$ water)[13] is also included, the range spans four orders of magnitude. The criteria for 2,4-DNT ($0.17\,\mu g\,L^{-1}$ water),[11] RDX ($2.0\,\mu g\,L^{-1}$ water),[14] and 1,3,5-TNB ($1.0\,\mu g\,L^{-1}$ water)[15] are intermediate. No general recommendations have been issued for contaminant levels in soil. Instead, soil levels have been evaluated on a site-by-site basis, depending on such factors as the proximity of the contaminated soil to locations of groundwater use. For example, at Cornhusker Army Ammunition Plant, cleanup criteria of $5\,\mu g$ TNT g^{-1} soil, $10\,\mu g$ RDX g^{-1} soil, and $15\,\mu g$ TNB g^{-1} soil were established.[16]

Table 2 Proposed drinking water criteria for nitroaromatics and nitramines

Compound	Proposed drinking water limit ($\mu g\,L^{-1}$)
HMX	400[a,b]
RDX	2.0[a,c]
TNT	1.0[a,d]
2,4-DNT	0.17[e,f]
2,6-DNT	0.0068[e,f]
1,3,5-TNB	1.0[a,g]

(Reproduced by permission of the Cold Regions Research and Engineering Laboratory, M.E. Walsh, T.F. Jenkins, P.S. Schnitker, J.W. Elwell, M.H. Stutz, *Evaluation of SW846 Method 8330 for Characterization of Sites Contaminated with Residues of High Explosives*, CRREL Report 93-5, 3, 1993.)
[a] Lifetime exposure cancer risk level 10^{-6}.
[b] According to USEPA.[13]
[c] According to USEPA.[14]
[d] According to USEPA.[12]
[e] Recommended criterion for increased cancer risk of 1.0×10^{-6}.
[f] According to USEPA.[11]
[g] According to Etnier.[15]

Table 3 Half-lives of explosives present in synthetically prepared contaminated soil samples at 0.1% w/w

Explosive	Half-life (years)
TNT	1
RDX	36
HMX	39
PETN	92
PBX 9010[a]	108
PBX 9011	224

(Reprinted by permission of Sandia National Laboratories, P.J. Rodacy, P.K. Leslie, S.D. Reber, S.E. Klassen, G.T. Cordes, *The Development of Ion Mobility Spectroscopy for the Detection of Explosives in Environmental Samples*, SAND97-2215, 9, © 1997.)
[a] Plastic-bounded explosive.

To compound the problem of toxicity, explosives are remarkably persistent in the environment when they are initially present at elevated concentrations. A study was performed at Los Alamos National Laboratory to determine the half-lives of various explosives in the soil in a temperate climate.[17] The samples contained 0.1% w/w of one of twelve explosives mixed with soil in a blender. Samples were then placed in transite cylinders with fine-mesh stainless steel screen

on the bottom and hardware cloth over the top, then buried in the test area outdoors with their tops flush with the surrounding surface. The explosives were therefore subjected to naturally occurring temperature fluctuations, rainfall, and bacterial decomposition over a period of twenty years. Estimates of the half-lives were made from normalized ultraviolet (UV) chart data by applying first-order rate equations. The results of these tests are summarized in Table 3,[17,18] and demonstrate that the half-life of various explosives in soil can range between 1 year (for TNT) to 92 years (for PETN). The half-lives of plastic-bonded explosives exceed 100 years.

It should be emphasized that the Los Alamos results applied to (a) very substantial concentrations of explosives which were (b) mixed artificially with soils to create samples. Other investigators have emphasized that the degradation behavior of low concentrations of explosives in contaminated soils found in the environment frequently exhibits a strikingly different behavior. For example, Rodacy et al.[19] exposed small masses of authentic contaminated soil samples containing 10 ppb TNT or RDX in shallow dishes to sunlight outdoors. The concentration of TNT was undetectable after 60 days, while that of RDX was essentially unchanged. Observations such as these suggest that the degradation of low concentrations of explosives such as TNT in soil may proceed by mechanisms other than first-order kinetics. Special handling may be needed to assure that such concentrations in soil remain stable until the sample is actually analyzed. The entire issue of determining ''holding times'' for explosives in soil samples is discussed in section 2.2, entitled ''Maximum Holding Times (MHT) for Soil and Aqueous Samples''.

2 SAMPLING CONSIDERATIONS

2.1 Special Considerations for Soil Sampling

The traditional approach for sampling soils which may have been contaminated with explosives is to use ''grab'' samples, taken from a wide area. These samples would then be analyzed using appropriate field or laboratory-based procedures. The resulting data would be used to evaluate the level of explosives contamination and to choose the appropriate cleanup procedure(s). Careful thought would suggest that random grab sampling might not be the most appropriate strategy. An implicit assumption in this approach is that the distribution of contaminants is sufficiently homogeneous that the results from the analysis of these discrete samples can be used to represent average site conditions within the grid. In general, the number of discrete samples collected is insufficient to address the short-range distribution of contaminants and, practically speaking, all samples are automatically assumed to be representative of average site conditions within the grid from which they were collected.[20] Furthermore, the environmental characteristics of explosives-related compounds in soil indicate that they are extremely heterogeneous in spatial distribution. Concentrations range from nondetectable levels (<0.5 ppm) to percent levels (>10 000 ppm) for samples collected within several feet of each other. In addition, the waste disposal practices at these sites, such as OB/OD, exacerbate the problem and may result in conditions ranging from no soil contamination up to solid ''chunks'' of bulk high-explosives, such as TNT or RDX. Secondary explosives concentrations above 10% (>100 000 ppm) in soil are of concern because of potential reactivity that affects sampling and materials handling processes during remediation.[21] Often, 70–90% of soil samples analyzed during an explosives site investigation do not contain detectable levels of contamination.[22] Jenkins et al. observed that:

> Accurate chemical characterization of a hazardous-waste site requires the development and implementation of a well-designed sampling plan. After defining the area of interest [target population(s)], which might be an entire site or several defined areas within a site, samples are collected according to one of several possible schemes. Distributions of contaminants are very site specific, depending on the manner in which the contamination occurred, the physical and chemical properties of the contaminants involved, soil type, and the geology and hydrogeology of the site. Because of these site-specific characteristics, many references recommend that one perform a preliminary study before devising a sampling plan. [Reprinted by permission of John Wiley & Sons, Inc., © 1997 from T.F. Jenkins, C.L. Grant, G.S. Brar, P.G. Thorne, P.W. Schumacher, T.A. Ranney, *Field Analyt. Chem. Technol*, **1**, 151.]

Beginning in 1996, Jenkins et al.[23] systematically investigated the degree of short-range heterogeneity in analyte concentrations present at explosives-contaminated sites. Soil was sampled at nine locations on three installations. At each location, seven discrete grab samples were collected in a wheel pattern of radius 61 cm (one sample from the center and six equally spaced around the perimeter). Each of the seven samples was homogenized in the field and duplicates were analyzed by both an on-site colorimetric method (USEPA SW-846, Method 8515, ''Colorimetric Screening Method for TNT in Soil'')[24] as well as an off-site standard method (USEPA SW-846 Method 8330 ''Nitroaromatics and Nitramines by High Performance Liquid Chromatography''),[25] both described elsewhere in this article. Portions of the seven discrete samples were

also used to form a composite sample which was analyzed by both procedures. TNT concentrations obtained from both types of analysis varied by as much as two and a half orders of magnitude within a sampling wheel. Partitioning of overall variances for data indicated that sampling error dominated over analytical error. Therefore, the probability of collecting discrete samples that represent average analyte concentrations is very unlikely with these levels of heterogeneity. Results from analysis of composite samples, however, exhibited mean concentrations representative of the average condition, as well as low relative standard deviations (RSDs), indicating that field homogenization was adequate for characterization.

These results[20] exhibit several unifying themes that can be applied in designing future investigations of explosives-contaminated sites. First, it is clear that there was extreme heterogeneity at all sampling locations. A single sample from any of the 122-centimeter diameter circles could differ by orders of magnitude or more from the mean concentration of the small area sampled. RSDs for the seven discrete samples were often greater than 100%. A second consistent finding was that composite samples of the seven discrete samples could be reliably homogenized and subsampled on-site. This also opens the possibility of compositing discrete samples representing a larger area if concentration variations suggest that this approach would be desirable. Most importantly, it permits on-site processing without elaborate apparatus.

Jenkins et al.[20] summarized the shortcomings of the grab sampling approach as follows:

> Although random grab sampling is appealing from a cost perspective, it may be totally inadequate for decisions about the need for, or adequacy of, remediation. To provide data that can satisfy this need with a high level of confidence, the total uncertainty associated with site characterization must be understood and reduced to acceptable levels. ... For these four locations [as described in reference 20], standard deviations due to analysis, whether on-site or laboratory, were always much lower than the corresponding deviations due to sampling, and hence total error was dominated by sampling error. For the other locations, sampling error was even greater and so overwhelmed analytical error that this type of fractionation would only be possible with the use of asymmetric (here, logarithmic) limits. Clearly, if we want to make a significant improvement in the quality of site-characterization data, the major effort should be placed on reducing sampling error. Single grab samples are totally inadequate. [Reprinted by permission of John Wiley & Sons, Inc., © 1997, from T.F. Jenkins, C.L. Grant, G.S. Brar, P.G. Thorne, P.W. Schumacher, T.A. Ranney, *Field Anal. Chem. Technol*, **1**, 159.]

The results presented above needed to be expanded by evaluating additional explosives, sites, and sampling strategies. Jenkins et al.[26] described such additional work performed on soil samples at an active firing range located at Canadian Force Base (CFB)-Valcartier, Quebec. The principal contaminant, HMX, was present at concentrations about two orders of magnitude greater than TNT, even though the munition fired at this site had a 70 : 30 ratio of HMX to TNT. The authors again observed that sampling error frequently exceeded analytical method error by at least tenfold, and that more precise analytical results would be obtained only when more representative samples were taken. Several approaches were recommended to address this problem, such as on-site homogenization, compositing of discrete samples, and on-site analysis with appropriate confirmation results at an off-site environmental laboratory.

Additional work which evaluated the effect of sampling depth and novel sampling patterns was performed at the Ft. Ord, CA, firing range.[27] The contamination at Ft. Ord, like that at CFB-Valcartier, originated from antitank rockets. Hence, the contamination was not random. Depth profiling was performed using manual excavation for 0–15, 15–30, 30–45, and 45–60 cm, and a post-hole digger for 105–120 cm. Analysis of these soils showed that by far the greatest concentration of HMX was in the 0–15 cm depth range, with minimal penetration to lower levels. The authors also noted that the vast majority of explosives contamination was centered near the tank targets, rather than being distributed uniformly or near-uniformly throughout the site. The investigators observed:

> Since much of the site is free of significant explosives residue, it is inappropriate to estimate a mean concentration for a large area. Instead, sampling should concentrate on surface soils around target locations, to define the boundaries of contamination. To demonstrate this point, the concentrations of HMX in surface soils were plotted as a function of distance from the target along with similar data from CFB-Valcartier. With concentrations decreasing as distance from the target increases, a logical sampling strategy would use concentric rings with the target at the center. Such rings might employ radii varying by 3 to 5 m. Eight or more sample points would be located equidistant from each other around each ring, which would allow two composite surface samples to be formed for each ring. Each discrete sample should contain about a kilogram (±200 g) of soil and be homogenized thoroughly prior to subsampling for formation of composites. Each composite would contain one-half of the sampling points chosen in an alternating pattern. Results of this work demonstrate that EOD personnel can efficiently collect soil samples for site characterization at the same time that they are clearing the site for unexploded ordnance. Since the major portion of the explosives residues at the inland ranges are in the surface soils, samples for additional characterization should be relatively easy to collect. The depth of contamination in specific areas can be determined iteratively after the extent of surface contamination is mapped. [Reprinted by permission of the Cold Regions Research and Engineering Laboratory from T.F. Jenkins,

M.E. Walsh, P.G. Thorne, P.H. Miyares, T.A. Ranney, C.L. Grant, J.R. Esparza, *Site Characterization for Explosives Contamination at a Military Firing Range Impact Area*, Special Report 98-9, 21–22, 1998.]

An unexpected observation from the Ft. Ord study was the effect of very small particles on bulk measured concentrations of HMX. A large amount of fine dust was noted when very dry bulk samples were homogenized and subsampled. Analysis of two samples of dust showed HMX concentrations two to three times higher than that for the bulk soil. Apparently, a substantial portion of the explosives residues are associated with very fine particles that are not bound to the bulk sands. Clearly, there is danger of loss of these fine particles during mixing and subsampling.

2.2 Maximum Holding Times for Soil and Aqueous Samples

The ''MHT'' is the longest interval that a sample, aqueous or soil, may be stored prior to analysis before its analyte concentration is reduced by an unacceptable level or percentage. USEPA procedures specify a maximum pre-extraction holding time of seven days at 4 °C in the dark, to be consistent with the MHT values of other organics in soil.[28] One study[29] described the determination of the MHT in uncontaminated soils and aqueous matrices. The soils were taken from three totally unrelated sources, viz. (a) standard soil obtained through the US Army Toxic and Hazardous Materials Agency; (b) Captina silt loam from Roane County, TN; and (c) McLaurin sandy loam from Stone County, Mississippi (MS). Similarly, the aqueous samples were also drawn from three independent sources, viz.: (a) reagent-grade water; (b) groundwater from the Oak Ridge National Laboratory Aquatic Ecology Facility; and (c) surfacewater from the headwaters of White Oak Creek on the Oak Ridge Department of Energy Reservation. Aliquots of soil or water were spiked with explosives as required, then stored and analyzed as described in the reference.

The two statistical definitions used to determine the MHT were specified by the American Society for Testing and Materials (ASTM)[30] and Environmental Science and Engineering, Inc. (ESE) for a holding time study conducted in co-operation with the USEPA.[31] Both definitions are based on an approximating model for predicting concentration with time. The ASTM defines the MHT as the time the predicted concentration falls below the lower two-sided 99% confidence interval on the initial concentration. The ESE defines the MHT as the time the one-sided 90% confidence interval on the predicted concentration falls below a 10% change in the initial concentration. The difference between the two definitions is the method of placing a lower bound on the initial concentration. The ESE MHTs are usually longer than the ASTM MHTs because decreasing the initial concentration by 10% is usually a larger reduction than the lower two-sided 99% confidence limit. The ASTM MHT definition is recommended for analytical methods with precision such that the lower bound on 99% confidence limit for an analyte concentration is less than 10% of the initial concentration. Otherwise, using the ESE MHT definition would be more conservative. The results of the study on MHT for four explosives in both aqueous and soil matrices are summarized in Table 4; the calculated MHT are all considerably longer than those recommended by the USEPA.

Table 4 Recommended MHT for four munitions in soils and water

Munition	Storage condition (°C)	Matrix type	Recommended MHT (days)
HMX/RDX	4	Groundwater	50
HMX/RDX	4	Surface water	30
TNT	4	Groundwater	16
DNT	4	Groundwater	4
TNT/DNT	4	Surface water	14
HMX	4	Soil	52
RDX	4	Soil	63
TNT[a]	−20	Soil	233
DNT	4	Soil	107

(Reprinted by permission of Oak Ridge National Laboratory from M.P. Maskarinec, C.K. Bayne, L.H. Johnson, S.K. Holladay, R.A. Jenkins, B.A. Tomkins, *Stability of Explosives in Environmental Water and Soil Samples*, ORNL/TM-11770, 42, © 1991.)
[a] Immediate freezing recommended.

Grant et al.[32] provided additional MHT values not only from dry spiked soil, but also authentic contaminated soils from the Rockeye site at the Naval Surface Warfare Center (NSWC) at Crane, IN. The latter soils contained measurable concentrations of HMX, RDX, TNT, TNB, two isomeric microbiological transformation products of TNT, viz., 2-Am-DNT and 4-Am-DNT, and 3,5-dinitroaniline (3,5-DNA), a microbiological transformation product of TNB. The blank soils employed in this study were obtained from Windsor (VT), Charlton (NH), and Ft. Edwards (NY). All soils were spiked with explosives dissolved in water to eliminate any acetonitrile effects. The results on spiked soils showed that RDX and HMX are stable for at least eight weeks when refrigerated (2 °C) or frozen (−15 °C) but that significant transformation of TNT and TNB can occur within two hours without preservation. Freezing provides adequate preservation of spiked 2,4-DNT for eight weeks or longer. The results on field-contaminated soils did not show the rapid transformation of TNT and TNB that was observed in the spiked soils, and refrigeration appeared satisfactory.

Presumably, the explosives still present in the field soil after many years of exposure are less biologically available than in the spiked soils.[6] An earlier related study[33] demonstrated that explosives in spiked, air-dried soils are stable for a 62-day period under refrigeration. Data from the Grant et al.[32] study indicate that air drying of field-contaminated soils may not result in significant losses of explosives contaminants. Explosives in air-dried soils are stable at room temperature if they are kept in the dark.

Most of the analytical procedures for explosives in soils employ acetonitrile extracts, which are expected to be stable for at least six months under refrigeration. Acetone extracts are also thought to be stable if they are stored in the dark under refrigeration. The latter solvent is normally not recommended because it enhances the photodegradation of explosives.[6]

3 DETERMINATION OF EXPLOSIVES IN AQUEOUS SAMPLES

3.1 Procedures Employing "Salting-out" Liquid–Liquid Extraction Followed by High-pressure Liquid Chromatography

The most traditional extraction procedure for isolating explosives from aqueous samples involves the liquid–liquid extraction of the sample with a nonpolar solvent such as dichloromethane. Leggett et al.[34] evaluated this approach critically and proposed a superior alternative based on "salting-out" the explosives from a sample. Briefly, 130 g of sodium chloride was dissolved in 400 mL aqueous sample, to which 100 mL acetonitrile was then added. After the mixture was shaken vigorously in a separatory funnel, the layers were allowed to separate. Under these conditions, the final volume of immiscible acetonitrile was 22 mL, allowing a preconcentration factor >16. Individual component recoveries for a wide variety of nitroaromatics, nitroesters, and nitramines ranged between 93 and 100%. The authors confirmed that recoveries were not dependent on analyte concentration over at least a 200-fold range in concentration, and the presence of chloride in the acetonitrile phase had no effect on analyte retention times (see below).

The most widely-used method for separating and detecting extracted explosives employs HPLC with UV absorbance detection at 254 nm. The analytes are separated on an octyl (LC-8, 25 cm × 4.6 mm, 5 μm particles, product of Supelco, Inc., Bellefonte, PA) column using a water–methanol–acetonitrile (50 : 38 : 12 v/v/v) eluent (1.5 mL min^{-1} flow rate). Such a configuration permitted the major target analytes, viz. HMX, RDX, TNT, and 2,4-DNT, to be separated from the following species: TNB; SEX (1-acetyloctahydro-3,5,7-trinitro-1,3,5,7-tetrazocine, CAS 13980-00-2); TAX (1-acetylhexahydro-3,5-dinitro-1,3,5-triazine, CAS 14168-42-4); 2,6-DNT; 2-Am-DNT; 4-Am-DNT; 2,6-diamino-4-nitrotoluene (2,6-Dam-NT, CAS 59229-75-3); 2,4-diamino-6-nitrotoluene (2,4-Dam-NT, CAS 6629-29-4); and 2,4,5-TNT.[10] When a typical 100 μL sample injection was employed, the detection limits were 26, 22, 14, and 10 μg L^{-1} for HMX, RDX, TNT, and 2,4-DNT, respectively. These values were initially calculated using three times the standard deviation of the detector baseline noise, and were later confirmed using formal calculations based on the method of Hubaux and Vos.[35] Such calculations employ not only replication but also user-specified estimates of both false-positive and false negative errors (typically 5% for each). This analytical procedure was later modified to employ a methanol–water (1 : 1) eluent with the same flow rate. A related confirmation procedure employed a bonded cyano HPLC phase (LC-CN) column (25 cm × 4.6 mm, 5 μm particles, Supelco) with identical mobile phase, flow rate, detector, and monitoring wavelength. The two aforementioned separations provide almost exactly opposite analyte retention orders, resulting in a powerful tool for analyte identification.

The HPLC-based-procedures described above were validated using a collaborative study involving nine independent laboratories.[36,37] The four test matrices employed represented uncontaminated, treated, fortified, and authentic contaminated matrices. Based on the overall excellent performance of the HPLC procedures, they were ultimately adopted as standard methods for the analysis of nitroaromatic and nitramine explosives in water by the Association of Official Analytical Chemists (AOAC) and the USEPA, and in soil by AOAC, ASTM, and the USEPA.[1,28]

The HPLC methods described above may be modified slightly to enable either improved separation of specific analytes or improved reliability of identification. For example, two pairs of aminodinitrotoluenes (2-Am-DNT and 4-Am-DNT) and diaminonitrotoluenes (2,4-Dam-NT and 2,6-Dam-NT) cannot be resolved using the LC-CN column alone. However, all constituents may be baseline resolved by using two columns, viz., a short octadecyl column (LC-18, 3.3 cm × 4.6 mm i.d., 3 μm particle size, Supelco) and a traditional cyano column (LC-CN, 25 cm × 4.6 mm i.d., 5 μm particle size), placed in series. The separation employs a ternary eluent (water–methanol–tetrahydrofuran, 60.5 : 25 : 14.5 v/v/v) with a flow rate of 1.5 mL min^{-1}.[38] The use of HPLC in combination with photodiode-array detection is advantageous, because peak identification is more reliable with additional information from the UV spectrum. Further, by comparison of the various UV spectra at the peak front edge, peak apex, and peak end edge, one

can determine whether a single component is, indeed, present, or whether contaminants are co-eluting with the analyte of interest. Emmrich et al.[39] studied the separation of 21 nitroaromatics and nitramines in a single run using HPLC and reported both absorbance maxima λ_{max} and the molar absorptivities ε_{max} for each analyte. The authors used a Merck LiChrospher 100 RP-18 column (25 cm × 4 mm i.d., 5 μm particle diameter) and mobile phases employing (a) methanol–water (1 : 1 v/v, isocratic elution), (b) methanol–water gradient elution, or (c) methanol–water containing 2% v/v tetrahydrofuran, gradient elution. All analytes were eluted and resolved using the gradients, but TNT and nitrobenzene (NB) almost co-eluted. The presence of a small volume of tetrahydrofuran enabled these two particular analytes to be resolved cleanly. The HPLC method of Caton and Griest[40] was developed to quantitate the transformation products of TNT which would be observed during composting. This procedure uses an RP-C_{18}/Anion Mixed-Mode reverse phase/anion exchange column and three eluents: (a) 90 : 10 water–methanol v/v solution that was 0.015 M in potassium phosphate; (b) methanol; (c) acetonitrile. Two independent gradient programs were developed and evaluated. Quantitations were performed employing the UV absorbance at 254 nm. The aminodinitrotoluenes were baseline-resolved successfully using a single column.

3.2 Procedures Employing Solid-phase Extraction

The "salting-out extraction" (SOE) procedures described above were clearly successful for removing and concentrating explosives from aqueous samples, but they also exhibited several distinct disadvantages. First, each extraction generated a considerable volume of chemically hazardous waste. Second, liquid–liquid extractions in general may produce emulsions which may or may not clear easily with time. Third, concentration factors greater than approximately 25 would be difficult to achieve without either substantial volume reduction of the acetonitrile extract or the use of very large (>2 L) aqueous sample volumes. The extraction of such large sample volumes using traditional liquid–liquid extraction glassware becomes unwieldy. Taken together, it is worthwhile considering procedures based on solid-phase extraction (SPE), where the aqueous sample passes through a small cartridge or disk containing an appropriate medium which has been conditioned with organic solvent. After the disk or cartridge is dried, the explosives are eluted with a small volume, typically not more than 5 to 10 mL, of organic solvent.

Maskarinec et al.[41] evaluated the extraction properties of three SPE media, viz. XAD-4, Porapak® R (vinyl pyrrolidone, Waters Corporation), and Porapak® S (vinyl pyridine, Waters), which were packed into small columns and conditioned initially using acetone and water. Fortified aqueous samples (500 mL) were pumped through each column, and explosives were removed using 10 mL acetone under gravity flow. Acetone was exchanged for ethanol prior to HPLC analysis.

The presence of explosives was monitored using both UV and liquid chromatography/electrochemical detectors (LC/ELCD). The sensitivity of the former was approximately 1 μg mL^{-1} ethanol extract for HMX, RDX, TNT, 2,4-DNT, and 2,6-DNT when the monitoring wavelength was 254 nm. The LC/ELCD (−1.0 V vs Ag/AgCl reference electrode) had several notable advantages and disadvantages compared to the complementary UV detector. The LC/ELCD exhibited approximately tenfold better sensitivity for explosives than the UV detector.[42] Furthermore, the detector response would be proportional to the number of nitro groups present on a given analyte. Hence, the LC/ELCD was ideal for detecting potentially unidentified by-products and transformation products of explosives with at least some selectivity. In addition, LC/ELCD was superior to the UV detector for monitoring species such as NG and PETN. These compounds absorb weakly at 254 nm and strongly at 210 nm, where substantial interferences are well-known. As a worst-case example, the estimated detection limits of NG using the UV detector (254 nm) and LC/ELCD were 160 ng and 0.38 ng, respectively.[43] On the other hand, the LC/ELCD was considerably less robust than the UV detector. Oxygen had to be removed scrupulously from both the final ethanol extract and the eluent. The gold substrate and mercury-film electrode had to be prepared carefully and refreshed frequently due to contamination from irreversibly adsorbed species.

Overall, the two Porapak® resins outperformed XAD-4 as an extraction medium for explosives. They behaved in a manner consistent with specific adsorption of nitro compounds, in that the breakthrough volume was nearly equivalent for the test analytes HMX, RDX, TNT, and DNT. Even SEX and TAX, two very polar explosives by-products, were retained quantitatively.

Jenkins et al.[10] compared the characteristics of the SOE with two SPE media, viz. cartridge-based Porapak®-R and Empore® styrene–divinylbenzene disks, both conditioned with acetonitrile and water. The data collected over a four-day evaluation period were used to calculate certified reporting limits (CRL) according to a statistically unbiased procedure.[44] The CRL values are tabulated in Table 5. In general, none of the procedures were consistently superior to the others in low concentration detection capability. The only CRL value that appears out of line is that for tetryl determined using membrane-SPE, 0.83 μg L^{-1}. Inspection of the data

Table 5 CRL for various preconcentration techniques

Analyte	CRL ($\mu g\,L^{-1}$)		
	Salting-out (SOE)	Cartridge SPE[a]	Membrane SPE[b]
HMX	0.19	0.21	0.33
RDX	0.13	0.27	0.12
TNB	0.052	0.042	0.051
DNB	0.081	0.032	0.036
Tetryl	0.20	0.24	0.83
TNT	0.086	0.068	0.13
2-Am-DNT	0.10	0.046	0.055
2,4-DNT	0.083	0.085	0.044
o-NT	0.13	0.10	0.20
p-NT	0.22	0.12	0.23
m-NT	0.21	0.13	0.37

(Reprinted with permission of Elsevier Science B.V. © 1994 from T.F. Jenkins, P.H. Miyares, K.F. Myers, E.F. McCormick, A.B. Strong, *Anal. Chim. Acta*, **289**, 72.)

[a] Using cartridges containing Porapak® R conditioned with acetonitrile, then water.

[b] Using Empore® styrene–divinylbenzene disks conditioned with acetonitrile, then water.

NT, nitrotoluene.

indicated that this high CRL was caused by low recovery on only one of the four certification days. Elevated CRL values for tetryl were noted by other workers who attributed this behavior to the instability of the analyte in water. The present authors suggested that this instability is due to photochemical decomposition. The study also included the evaluation of 58 groundwater samples taken for method comparison from the Rockeye site at the Naval Surface Warfare Center (NSWC) in Martin County, IN.

Overall, the results may be summarized as follows:

- The three methods were comparable with respect to low-concentration detection capability, ranging from 0.05 to 0.30 $\mu g\,L^{-1}$.
- Percent recoveries generally exceeded 80% except for HMX and RDX by membrane-SPE, where recoveries were lower.
- Large interferences were found on about half of the groundwater samples from the NSWC, using the two SPE methods, but none were found by SOE.
- The SPE interferences were traced to a matrix interaction of the SPE polymers with low pH groundwaters which apparently caused the release of unreacted monomers or other contaminants from the interior of the polymer.

At least partly in response to the problems identified above, several manufacturers of SPE materials sought to improve the retention of very polar organics such as HMX and RDX, and experimented with new cleaning procedures to better remove interferences from these materials. As a result, Waters Corporation released a new ultra clean SPE material for use in cartridge SPE under the name Porapak® RDX (ultraclean divinylbenzene-vinylpyrrolidone), and 3 M Corporation developed a new surface modified styrene divinylbenzene membrane which had also been cleaned more extensively (Empore® styrene–divinylbenzene ''reverse-phase sulfonated'', or SDB-RPS).[45,46] The latter is a modified phase which provided additional retention for polar organics such as HMX.

Porapak® RDX cartridges were evaluated extensively, as described in Bouvier and Oehrle.[47] High-purity water samples fortified to 1 or 10 ppb in each of 14 explosives or explosives-related products were extracted using cartridges which had been preconditioned with acetonitrile and water. The concentrated explosives were eluted with a small volume of acetonitrile, which was subsequently diluted to 40% with high-purity water, vortexed, and analyzed by HPLC (variable wavelength detection). The recoveries of all 14 analytes, ranging from the polar HMX and RDX to the nonpolar 2,4-DNT and 2,6-DNT, exceeded 89% with RSDs of less than 7%.

Walsh and Ranney[48] recently and independently compared the performance of the Waters Sep-Pak Vac Porapak® RDX cartridges with Empore® (SDB-RPS) extraction membranes. The cartridges were conditioned using acetonitrile followed by water; the membranes were conditioned with acetone, isopropanol, methanol, acetonitrile, and finally water. After passage of each water sample through the solid-phase, air was drawn through the solid phase for 15–20 min to remove as much residual water as possible. The explosives were eluted from the solid phases with 4–5 mL acetonitrile. Final separation and quantitation of each analyte was performed using a gas chromatograph equipped with an electron capture detector (ECD) using the specific modifications noted in the reference. The authors specifically noted that calibration curves employing linear models were not appropriate for quantitating explosives when the ECD was employed. Instead, a quadratic model with all terms statistically significant, not only was more satisfactory but also allowed a greater concentration range to be covered. Furthermore, low concentration trinitroaromatic standards were unstable even when left in amber vials at room temperature. When the autosampler tray was cooled, these concentrations were stable during a typical 12-hour shift. The average recoveries observed from the two extraction media were entirely comparable for water samples spiked to 5 and 10 $\mu g\,L^{-1}$ in each of nine explosives, regardless of whether gas chromatography with electron capture detection (GC/ECD) or HPLC was used as the final quantitating method, and ranged between 85 and 136% for all analytes and both quantitating methods.

Table 6 MDL for target analytes using GC/ECD[a]

Analyte	MDL values using GC/ECD ($ng\,L^{-1}$)	
	Porapak® RDX Cartridge	Empore® SDB/RPS membrane
DNB	4	3
2,6-DNT	3	3
2,4-DNT	9	10
TNB	7	3
TNT	10	20[b]
RDX	4	20[b]
4-Am-DNT	3	3
2-Am-DNT	3	4
Tetryl	9	10
HMX	4	8

(Reproduced by permission of Preston Publications from M.E. Walsh, T. Ranney, *J. Chromatogr. Sci.*, **36**, 415 © 1998.)

[a] 1-L water samples fortified to $10\,ng\,L^{-1}$ preconcentrated to 4.0 mL acetonitrile using either cartridge or membrane SPE.

[b] MDL greater than the spiked concentration and therefore not valid.

The method detection limit (MDL)[49] for ten analytes, using the two SPE media and the GC/ECD method described, are presented in Table 6, and typically range between 3 and $20\,ng\,L^{-1}$ for each analyte, regardless of medium considered.

The same authors also quantitated the explosives present in authentic groundwaters drawn from wells at several military installations using both membrane and cartridge SPE media and GC/ECD with HPLC confirmation. The concentrations of each analyte ranged over five orders of magnitude; in the specific case of TNT, the concentrations spanned 0.1 and $10\,500\,\mu g\,L^{-1}$. In all cases, virtually identical values were obtained using either extraction procedure and either method of determination.

3.3 Determination of Explosives in Seawater

The determination of explosives in seawater provides a challenging extension to those methods applicable to groundwater. For many years, the disposal of obsolete munitions was performed by merely dumping them in the ocean. In 1963, the United States Navy initiated a program in which obsolete Maritime Administration hulks were used for this purpose. These old Liberty ships were loaded with munitions and scuttled at sea.

One of the first procedures for determining explosives in this matrix involved the extraction of 100 mL seawater with benzene followed by evaporation of the resulting extract under reduced pressure. Quantitation was performed by GC/ECD, employing injection of 8–9 μL of concentrate onto a packed column. The limits of detection for TNT, RDX, and tetryl observed were estimated to be 2, 5, and 20 parts per trillion, respectively.[50] This procedure was used to determine TNT, RDX, and tetryl in seawater, sediment, and ocean floor fauna samples taken from an explosives dumping area 85 miles west of Cape Flattery, WA, and from similar samples taken 172 miles south-southeast of Charleston, SC, at a nominal depth of 2500 meters. No explosive contaminants were found in any of the samples examined.[51] Sediment samples were initially extracted with high-purity benzene. The solvent was subsequently evaporated completely, and the residues dissolved in 50 μL fresh benzene. Half of the concentrate was spotted along a ChromAr® 500 thin-layer chromatography sheet (glass fiber thin-layer chromatographic paper consisting of 70% Silicar TLC 7F containing a short wave, 254 nm, UV phosphor on 30% glass fiber), and developed with hexane–acetone (45 : 55 v/v) as eluent. The appropriate disks corresponding to the explosives were removed and re-eluted with benzene. Later, gas chromatographic analyses were performed on the final benzene concentrates. The limits of detection for TNT, RDX, and tetryl in tissues were determined by injecting 4.1×10^{-7} g TNT, 11×10^{-7} g RDX, and 17×10^{-7} g tetryl into a 51.4 g sea cucumber. After the usual subsequent workup and analysis, the figures of merit were estimated to be 47, 123, and 740 parts per trillion, respectively. The procedure for determining the detection limit was not given.

A very recent paper described the determination of explosives in seawater using solid-phase microextraction (SPME) fibers as the extraction medium, followed by quantitation using either HPLC or gas chromatography/ion-trap mass spectroscopy (GC/ITMS).[52] The test matrix initially evaluated was synthetic ocean water, prepared by dissolving an appropriate quantity of ''Instant Ocean'' (a commercially available product) into high-purity water. This test matrix was then spiked with 50 ppb in each of TNT, RDX, 2-Am-DNT, 4-Am-DNT, and 2,4-DNT. Four SPME fibers, which exhibited widely differing polarities, were evaluated: poly(dimethylsiloxane) (PDMS, 100- and 65-μm thickness, nonpolar); polyacrylate (85 μm film thickness, polar); and Carbowax–divinylbenzene (65-μm thickness, polar). All determinations were performed with a Finnigan MAT GCQ GC/ITMS operated in the negative chemical ionization (CI) mode, with methane as the reagent gas.

The study demonstrated clearly that no one fiber was suitable for all of the analytes; however, the most suitable fiber for general use was Carbowax–divinylbenzene. Two critical figures of merit were calculated and reported, viz., the limit of detection (LOD) and the limit of quantitation (LOQ). Both of these quantities are defined in terms of X, the mean blank signal, and σ, the standard deviation of the blank signal. The LOD values, calculated as $X + 3\sigma$, were in the low (5–10) part-per-trillion range for all of the target NT analytes, but greater for the nitramines.

The LOD and LOQ ($X + 10\sigma$) for RDX were 325 and 790 parts-per-trillion, respectively.

To test the utility of the developed method, both SPE and SPME were used for the analysis of several real oceanwater samples. Several pairs of samples, consisting of one-liter samples for SPE and 20-mL samples for SPME, were collected. The one-liter oceanwater samples were extracted through SDB-RPS SPE disks, which had been conditioned as described previously. The disks were washed with water, then ultrasonically extracted overnight with 4–5 mL acetonitrile. The recovered acetonitrile extract was then concentrated via nitrogen ''blowdown'' over a period of several hours to a final volume of 0.5 mL for HPLC and GC/ITMS analysis. The 20-milliliter oceanwater samples were separated into 5-milliliter aliquots for direct SPME sampling and GC/ITMS analysis.

The HPLC and GC/ITMS results using SPE generally agreed quite well for most of the samples. The SPME results, however, were often high compared to the SPE results. This behavior may be explained in several ways:

- The most probable explanation is that the SPE results may appear low due both to decreased extraction efficiency and losses in the lengthy analyte concentration/solvent evaporation process.
- The SPME results could appear high because SPME, with its fewer steps and sample transfers, may recover a greater portion of the sample.
- The accuracy of the SPME results may be biased high because the calibration curve was prepared using simulated oceanwater, while the samples contained real oceanwater. This is unlikely because ''Instant Ocean'' salt is designed to be an accurate surrogate matrix for seawater. The residual solvent ''tail'' in the gas chromatograms was significantly smaller when SPME, rather than disk SPE, was employed as the extraction method, as expected.

Caton and Barshick[53] examined various ways to improve the qualitative and quantitative determination of TNT and its reduction products in SPE extracts of seawater using the Finnigan GCQ instrument described previously. The authors were able to detect 10 pg of TNT, 2,4-DNT, RDX, 4-Am-DNT, and 2-Am-DNT with a signal-to-noise ratio exceeding 6 using a standard 1 μL injection when the instrument was operated in the negative ion chemical ionization (NICI) mode.

Darrach et al.[54] also investigated the use of SPME for ocean-related samples, including sediments which were contaminated with unexploded World War II-vintage undersea ordnance. This material was scattered in Halifax Harbor, Nova Scotia, Canada, after a minor fire in 1945 caused the detonation of a small nearby munitions bunker complex. The PDMS-divinylbenzene fiber exhibited the highest relative efficiency of the commercially available fibers for TNT extraction. After each extraction, the fiber was dried in a vacuum desiccator for at least 15 min. The analytes collected on the fiber were then desorbed and swept into a quadrupole mass spectrometer. Explosives detection was achieved using the reversal electron attachment detection (READ) method, which takes advantage of the NTs' extremely large cross-section for attaching zero-energy electrons. The explosives detected in this work were TNT and DNT in sediments, at concentrations ranging from the low parts-per-billion to high parts-per-trillion ($\mu g\,kg^{-1}$ to $ng\,kg^{-1}$).

Rodacy et al.[18] described the determination of explosives in raw seawater or acetonitrile extracts (from solid-phase disk extractions) of seawater using the ion mobility spectrometer (IMS). Approximately 100 samples containing 10 ppb TNT or 20 ppb RDX (0.3 μL injections) can be analyzed at the rate of three per minute before excess water must be vented from the system. Scatter gradually increases from ±10% of the nominal value at the beginning of the series to ±25% at the end of the series. The scatter is presumed to be due to the increased concentration of water in the drift region. With direct injection of seawater samples, a large deposit of salt builds up in the injection liner. If this deposit remains, it sorbs some of the analyte, resulting in broad ''stubby'' peaks.

A better method of analyzing seawater samples is to deposit the sample aliquot onto a quartz sample tube or filter paper, followed by evaporation of the water. The residual explosive is then thermally desorbed in the standard manner. This method eliminates the problem of salt build-up in the injector, permitting samples up to several hundred microliters to be analyzed routinely. One obvious disadvantage, however, is that water evaporation can be relatively slow (several minutes), which decreases the sample throughput. To partially overcome this limitation, several samples may be prepared simultaneously. Alternatively, one can use a more volatile solvent, such as acetonitrile and acetone, and significantly decrease the amount of time required for the solvent to evaporate.

Rodacy et al.[18] noted that the acetone or acetonitrile extracts of seawater were frequently colored, from light to dark yellow, and even further to a deep green. For the most part, the compounds responsible for the color did not interfere with the analysis. On one occasion, the authors reported that the water surface was covered with what appeared to be millions of small ($\frac{1}{4}$ inch and smaller diameter) jellyfish. On that one day, a chemical species in the extract, probably a jellyfish metabolite, was problematic.

The agreement between IMS and HPLC concentrations for explosives determined in the same extract was frequently excellent. At concentrations between 100 and 1000 parts per trillion (ppt), the IMS values averaged

42.2% less than the HPLC values determined using the standard Method 8330 (described in detail below). Between 1 and 100 parts per billion (ppb), the IMS values were 2.4% higher than the HPLC values. Between 100 and 800 ppb, the IMS data was 20.0% higher than the HPLC values. Overall, the IMS results agreed with the HPLC analysis within 7.3%, based on a data set representing nearly 200 samples.[55] IMS values appeared to be biased low compared to those generated in the HPLC analyses, possibly due to nonlinearity of the calibration curve. The major problems involved with the IMS included severe sensitivity to vibration and the lack of robust structural construction, leading to frequent instrument failures.

3.4 Additional Procedures

The foregoing discussion has emphasized the determination of explosives in water using HPLC because the analytes are not degraded thermally, as they would typically be if gas chromatographic methods were employed. The thermal lability of explosives places a potentially severe limitation on the use of GC-based methods, which would offer far greater potential sensitivity and selectivity than their HPLC counterparts. To address this problem, Yinon et al.[56] evaluated a gas chromatograph equipped with a septum-programmable injector, which permitted cool on-column injections. The mass spectrometric detector was deployed in both its electron impact (EI) and CI modes (reagent gas isobutane). Two groups of explosives were evaluated: (A) TNT, tetryl, RDX, HMX, and PETN; and (B) 2,3-, 2,4-, 2,5-, 2,6-, and 3,4-DNT. The test samples were prepared by spiking water with a mixture of explosives at the desired concentration, then extracting them with dichloromethane. The extracts were taken to dryness; the residues were dissolved in methanol and injected into the gas chromatograph. The authors noted that the nitramines RDX and HMX did not degrade in the injection port. Tetryl, a chemically unstable analyte, formed *N*-methylpicramide in the ionization source prior to ionization as expected,[57] but did not degrade further. The method enabled detection of 10 ppb TNT, RDX, HMX, and tetryl in water.

Feltes et al.[58] quantitated explosives in aqueous samples using a variety of gas chromatographic detectors, including mass spectrometric detectors in three different modes, viz. EI; positive-ion chemical ionization (PICI); NICI; ECD; and the chemiluminescent nitrogen detector (CLND), also known as the Thermal Energy Analyzer (TEA). SPEs of fortified one-liter water samples were performed with cartridges packed with resins (XAD-2, XAD-4, XAD-8) or silica-based SPE media (octadecyl, phenyl, cyano). The explosive analytes were eluted with a small volume of dichloromethane, which was later dried over anhydrous sodium sulfate and reduced in volume to about 1 mL in a rotary evaporator after exchange of the solvent to methanol.

The authors noted that while nitroaromatics can be determined successfully by ECD as a result of their high electron affinity, the response for nitroaromatics depends strongly on the individual compound. On the other hand, the CLND response is more similar for all the compounds because the nitrogen content of each analyte does not vary as sharply as the electron affinity. For example, the CLND response ratio of 3,4-DNT to 2-NT is 3.4, while the corresponding ECD response ratio is 61. The CLND exhibits excellent selectivity towards nitrogen-containing analytes, and is therefore better suited for the determination of explosives in complex environmental samples than the ECD, which is less selective. On the other hand, the ECD is approximately three orders of magnitude more sensitive than the CLND. The authors compare and contrast the merits of the various detectors as follows:

> As a result of their high electron affinities, nitroaromatic compounds can be determined with high sensitivity by GC using ECD. If, however, highly polluted surface water is to be investigated, ECD is not sufficiently selective for an unambiguous compound assignment. In contrast, a chemiluminescence detector is much more selective for nitro compounds, allowing a reliable peak assignment if reference compounds are available. The major disadvantage of the chemiluminescence detector is its lower sensitivity (compared with the ECD), which precludes the analysis of weakly polluted water. Definite compound assignment can usually be achieved with GC/MS (gas chromatography/mass spectrometry) where NICI is more sensitive and selective than EI, but gives fewer structure-specific fragments. [Reproduced by permission of Elsevier Science Publishers B.V. from J. Feltes, K. Levsen, D. Volmer, M. Spikermann, *J. Chromatogr.*, **518**, 39 © 1990.]

One highly successful, but seldom-used, approach for improving the sensitivity of HPLC-based quantitations of explosives is to derivatize the analytes in a post-column reactor, such as that described in Engelhart et al.[59] Initially, the explosives were photolyzed by elution through a photo reactor consisting of 20 meters of crocheted Tefzel™ capillary tubing coiled around a low-pressure mercury lamp (254 nm). In the second step, the nitrite ion so generated is reacted with colorimetric reagents, such as sulfanilamide and [napthyl-(1)]-ethylene diammonium chloride, to form an azo dye whose absorbance can be monitored at 540 nm (via the Griess reaction, discussed in detail elsewhere in this chapter). Using this approach and nominal 5 to 10 μL injections, it was possible to detect selectively 25–50 ppb nitramines and 30–100 ppb of nitrate esters, even in complex mixtures. The detection limit for esters and nitramines was approximately 100 pg using the derivatization, compared to approximately

100 μg unreacted. Only for the nitroaromatic explosives is a UV detection system twice as sensitive as the detection system described here.

4 DETERMINATION OF EXPLOSIVES IN SOIL SAMPLES

4.1 Extraction Techniques for Explosives Residues in Soil: Initial Considerations

Many of the analytical procedures employed for the determination of explosives in aqueous samples may also be used for their analysis in soils. The most important difference between the overall methods is, of course, the exact procedure(s) employed to extract the analytes from the matrix. The explosives range in polarity from the moderately polar DNT and TNT to the highly polar HMX and RDX; hence, the choice of extraction solvent becomes an important variable. In addition, there are a variety of generally accepted procedures available for extracting solid samples, including Soxhlet extractors, ultrasonic baths, mechanical wrist-action shakers, and homogenizer–sonicators. Two very recent additions to this list are supercritical fluid extraction (SFE) and accelerated solvent extraction (ASE™).

Jenkins and Grant[60] performed an extensive optimization study for the extraction of explosives in soil. The test matrix, taken from the Iowa Army Ammunition Plant, was the surface of an old ordnance burning area that had not been used for several years prior to sampling. The extraction behavior of two soils and four analytes, viz., TNB, RDX, HMX, and TNT, were evaluated during this study. The authors noted that the solubilities of RDX and HMX were over twenty times higher in acetonitrile than in methanol; hence, acetonitrile was the preferred extraction solvent. The performance of acetonitrile was equivalent to that of methanol for TNT and TNB. Both solvents were suitable for the HPLC determinations which would be employed subsequently. The extraction techniques evaluated were Soxhlet, ultrasonic bath, mechanical wrist-action shaker, and homogenizer–sonicator. Both the Soxhlet extractor and ultrasonic bath recovered more than the homogenizer or shaker. The ultrasonic bath and shaker offered additional advantages to Soxhlet and homogenizer–sonicator in that several samples could be processed simultaneously.

The chromatographic separations of the analytes were performed using the HPLC methods described previously for groundwater samples. The explosives were eluted from the ''primary'' column (Supelco LC-8, 250 × 4.6 mm, 5 μm particle size) using a water–methanol–acetonitrile eluent (50 : 38 : 12 v/v/v, 1.5 mL min^{-1}). The ''confirmation'' procedures employed either a Supelco LC-CN or LC-18 column (both 250 × 4.6 mm) using 50 : 50 v/v water–methanol eluent.[61] LC-CN, which is frequently used as a ''normal'' phase column, was particularly effective for confirmation because the elution order was almost completely inverted from that observed for LC-8 and LC-18, which are both strictly ''reversed'' phase columns. Since few commercial laboratories are equipped to use either the electrochemical detector or TEA with LC (liquid chromatography), and the intent is to develop a method for general use, the authors concentrated on UV detection. A wavelength of 254 nm was chosen because a variety of manufacturers offer a fixed wavelength detector based on the mercury lamp emission line at 254 nm. All analytes of interest absorbed either strongly (nitroaromatics) or modestly (nitramines) at this wavelength, while potential interferences such as chlorinated organic compounds do not.

Overall, the sonic bath method is recommended because it is inexpensive and can be used to process a number of samples simultaneously and unattended. The gentle heating that occurs with the sonic bath (≈39 °C) also seemed desirable to increase the rate at which equilibrium is attained. The sonic bath does not suffer from the problem experienced with the Soxhlet device of co-extracting significant amounts of interfering substances that have a low solvent–soil partition coefficient. It is also worth noting that the Soxhlet extraction of thermally labile analytes such as tetryl frequently presents unexpected difficulties. On the other hand, there were some disadvantages and/or precautions to be noted in conjunction with sonic agitation. Analyte extraction efficiency may depend on the vigor of agitation, which depends on many variables for sonic devices. It had been observed that the extent of particle disruption varies inversely with solid/solution ratio; consequently this ratio should be constant in comparison analyses. Solvent variables such as viscosity, vapor pressure, and dissolved air concentration affect acoustic impedance. Not surprisingly, temperature increases allow cavitation bubbles to collapse with less force; consequently, the increased solubility may be offset by reduced agitation. However, as long as conditions are held constant, the sonic bath represents an excellent way to perform such extractions.

4.2 Experimental Considerations and Statistical Validation of the Proposed Standard Method

The procedure sketched above was subjected to a rigorous statistically constructed collaborative study involving eight independent laboratories.[33,62,63] An open-literature paper describing this study[64] thoroughly explained many of the critical experimental details. For example, air-drying of the soil samples away from direct sunlight for at least 24 hours prior to extraction minimized

concentration changes to due to photodegradation. It was recommended, but not required, that the soil sample be ground to pass a 30 mesh sieve prior to extraction. Weisberg and Ellickson[65] recommended that soil samples be inspected carefully prior to extraction. Extreme caution must be exercised if the samples arrive with a dry appearance. Raw munitions will appear as a grayish-white powder or pellets, mixed in with the solid matrix. Samples containing greater than 2% explosives should be neither ground nor sieved.

The authors carefully evaluated the use of the ultrasonic bath and the relevant extraction kinetics, which varied both with the specific analyte and soil considered. In general, extraction periods of less than 4 hours often recovered less than 80% of the total extractable analyte. An extraction of 18 hours, on the other hand, recovered greater than 90% of all the extractable analytes in a variety of field-contaminated soils. An 18-hour period was also convenient because the extraction could be started in the late afternoon and the extracts processed the following morning. The authors noted that the rate of analyte extraction from spiked soils was quite different than from field-contaminated soils. Despite the fact that the spiked soils were aged for two days to allow solvent evaporation and analyte interaction with the soils, only 15 min of extraction time was required for greater than 90% analyte recovery. Clearly it was unrealistic to develop extraction methodology based solely on spiked material. For these reasons, an 18-hour ultrasonic period with acetonitrile was employed. The recovery of tetryl at a bath temperature of 45 °C was significantly less than that at 11 °C; hence, the cooling and temperature of the ultrasonic bath demand particular attention.

Various soil-to-extraction solvent ratios were investigated, with the initial method evaluation performed using 2 g of soil to 50 mL acetonitrile. A ratio of 2 g soil to 10 mL acetonitrile was selected because it gave the best low-level detection capability. Higher soil-to-solvent ratios were not tested because sufficient volumes of processed extract were needed to allow both primary analysis and secondary analyte confirmation.

One of the clear potential advantages of using the ultrasonic bath was the ability to process large numbers of samples simultaneously. Such a practice, would, however, introduce the possibility of reduced sonic dispersion and extraction recovery. To test this possibility more carefully, 4 replicate, 2.0 g portions of contaminated soil were extracted for 18 hours as usual with no other tubes in the bath, and another 4 portions were processed in an identical fashion along with 32 additional samples in the bath. No significant differences were found between the two treatments at the 95% confidence level for any of the analytes.

Prior to analysis by LC, the soil extracts had to be filtered to protect expensive reverse-phase columns. Because of the long period of sonication required for maximum analyte recovery, the soil aggregates were dispersed into very fine particles with long settling times. These fine particles made filtration of the extracts extremely difficult, even after the extracts were centrifuged at 5000 rpm for extended periods of time. Rupturing of the membranes sometimes occurred due to the excessive pressure required. For this reason, additional steps had to be introduced to enable facile filtration of the acetonitrile extracts. Adding aqueous calcium chloride solution (optimized concentration 5 g calcium chloride L^{-1}) produced flocculation of the fine particles and rapid settling for most soils. After a 15-min settling time, the clarified solutions were readily filtered through 0.5 μm filters.

Stock standards of explosives dissolved in acetonitrile can be safely used for periods up to one year if they are stored in glass, at 4 °C, in the dark, and with precautions to minimize solvent evaporation. Working standards can be prepared and used over a 28-day period if refrigerated and kept in the dark when not in use. In general, extracts can be held for up to two months without large adverse effect.

The HPLC methods employed the columns and methodology described previously. The target analytes were all resolved when eluted from a Supelco LC-18 column (25 cm × 4.6 mm i.d., 5 μm particles) with methanol–water (1 : 1 v/v). These explosives were also eluted and resolved using the same eluent from the "confirmation" column, which was a Supelco LC-CN (25 cm × 4.6 mm, 5 μm particles), but in a nearly inverted order. UV detection at 254 nm was employed, as discussed previously. Weisberg and Ellickson[65] further recommended a complete solvent change to methanol with a 5-min hold to flush the column, followed by a gradual return to the initial program conditions over a 5-min period. This practice elutes nonpolar nontarget analytes that will otherwise elute in subsequent samples and interfere with the usual determination method. The authors observed that the benefits of greater baseline stability outweighed the longer analysis turnaround time.

Overall, the method described above was sufficiently robust and reliable to be approved for both "Official First Action" and "Final Action", first by the AOAC, then by the USEPA (Method 8330, "Determination of Concentration of Nitroaromatics and Nitramines by High-performance Liquid Chromatography"). The MDL[66] and CRL values were calculated for thirteen target compounds, as shown in Table 7. MDL values ranged between 0.03 μg g^{-1} soil for 2,4-DNT and 2-Am-DNT, to 1.27 μg g^{-1} for HMX; however, HMX was the only species where the MDL exceeded 1 μg g^{-1}. Recoveries ranged between 80% (for HMX) to 119% (for TNB).

Table 7 Detection capability estimates and overall analyte recovery for the standard analytical method for munitions in soil

Analyte	MDL[a] ($\mu g\,g^{-1}$)	CRL[b] ($\mu g\,g^{-1}$)	Recovery (%)
HMX	1.27	2.15	80
Tetryl	0.12	0.65	83
RDX	0.74	1.03	84
2-Am-DNT	0.03	0.11	91
O-NT	0.07	0.24	92
NB	0.08	0.11	100
M-NT	0.07	0.25	101
TNT	0.08	0.24	102
2,4-DNT	0.03	0.07	102
P-NT	0.07	0.22	103
DNB	0.11	0.12	105
2,6-DNT	0.07	0.16	107
TNB	0.29	0.24	119

(Reproduced from T.F. Jenkins, M.E. Walsh, P.W. Schumacher, P.H. Miyares, C.F. Bauer, C.L. Grant, *J. Assoc. Off. Anal. Chem.*, **72**, 897 (1989) © 1989 AOAC International.)
[a] As defined in Federal Register.[66]
[b] As defined in US Army.[44]

4.3 Performance of the Standard United States Environmental Protection Agency Method 8330

Soil samples obtained from two independent EOD sites were analyzed by staff members of the Cold Regions Research and Engineering Laboratory (CRREL) and Missouri River Division Laboratory (MRD). At both sites, 2,4-DNT was detected in all samples with detectable analytes. The 2,4-DNT was present at much higher concentrations than TNT, the reverse of what is found at other types of sites. The source of this contamination was probably the improper demolition of excess propellant (i.e. it was detonated, not burned). In fact, whole propellant grains were found scattered about each EOD site.

GC/MS analysis of acetonitrile extracts of soil samples and propellant grains confirmed the presence of diphenylamine and dibutylphthalate, which, along with nitrocellulose (NC) and 2,4-DNT, are the ingredients of M1 propellant.[3] Most analytes could be confirmed using the cyano (LC-CN) confirmation column eluted with $1.5\,mL\,min^{-1}$ 1:1 methanol–water, as specified in Method 8330. HMX and RDX, which elute several minutes before TNT on the analytical column (LC-18), elute after TNT on the confirmation separation. This dramatic shift in retention order makes the confirmation of nitramines certain. However, confirmation of DNT is difficult in many cases. The isomers of DNT elute close to TNT on the confirmation separation, and since they are often present at much lower concentrations than TNT, their confirmation may be ambiguous. However, some improvement in the resolution of 2,4-DNT and TNT can be achieved using a slower flow rate ($1.2\,mL\,min^{-1}$) and a weaker eluent (35:65 methanol–water). Additionally, this flow rate and eluent greatly improve the separation of 2-Am-DNT and TNT.

Another problem associated with the confirmatory separation for some samples is the presence of many more peaks in the confirmation chromatogram than in the

Table 8 Concentration ranges observed for various explosives in soil and water. Samples analyzed by staff members from CRREL and MRD

Analyte	CRREL[a]		MRD[b]		Median concentration for combined data sets	
	Soil ($\mu g\,g^{-1}$)	Water ($\mu g\,L^{-1}$)	Soil ($\mu g\,g^{-1}$)	Water ($\mu g\,L^{-1}$)	Soil ($\mu g\,g^{-1}$)	Water ($\mu g\,L^{-1}$)
HMX	t[c]-5700	0.13–673	0.13–365	2.45[d]	3.7	76
RDX	t-13 900	0.02–1400	0.19–105	0.1–162	3.6	3.0
TNB	0.3–550	1.0–46	0.08–1790	0.10–36.1	5.0	1.5
DNB	0.2–45	0.15–1.4	0.11–61	0.06–8.7	0.66	0.78
Tetryl	t-1260	0.18–0.4	0.36–171	0.07–11.6	3.0	0.92
TNT	t-102 000	0.07–981	0.13–31 000	0.08–125	5.5	3.5
2-Am-DNT	t-37	0.02–218	0.32–373	0.86–216	0.62	11.2
4-Am-DNT	t-3.9	0.06–217	0.15–10.6	1.09–2.58	0.27	4.6
2,4-DNT	t-84	0.05–4.6	0.22–318	0.12–6.74	0.65	1.2
2,6-DNT	0.08–4.5	0.02–29	1.23[d]	1.5[d]	0.53	0.10

(Reproduced by permission of the Cold Regions Research and Engineering Laboratory, M.E. Walsh, T.F. Jenkins, P.S. Schnitker, J.W. Elwell, M.H. Stutz, *Evaluation of SW846 Method 8330 for Characterization of Sites Contaminated with Residues of High Explosives*, CRREL Report 93-5, 9, 1993.)
[a] Cold Regions Research and Engineering Laboratory.
[b] Missouri River District Laboratory.
[c] Analyte was detected, but concentration below the reporting limit.
[d] Only one sample where analyte was detected.

analytical chromatogram. The cyano function on the LC-CN ''confirmation'' column is less retentive for aromatic compounds than the hydrocarbon-based phases of the LC-18 or LC-8 ''primary'' analytical columns. Because the confirmation column is less retentive, it is more prone to interference from nontarget analytes that have long retention times on the LC-18.

The concentration ranges observed in this study for various explosives and related products in soils and water ranged from ''nondetectable'' to thousands of parts-per-million in soil or water, as shown in Table 8.

4.4 Additional Procedures

Several new approaches to the extraction and quantitation of explosives in soil have given the analytical chemist more choices for determining these species both rapidly and reliably. For example, ASE™ employs both elevated temperature and pressure to extract a solid sample in a sealed vessel rapidly. A manufacturer's application note[67] describes the rapid extraction (14 min) of small soil samples using either acetone or methanol (17–50 mL sample^{-1}). This approach was tested with clean soil spiked with explosives, authentic contaminated soils from Germany, and heavily contaminated wet soil. For the latter, 2.5 g soil were mixed with 0.8 g diatomaceous earth prior to extraction. In all cases, the recoveries from the ASE™ were greater than those with sonication. Parallel results obtained using Soxhlet extraction were not reported. Extractions from wet heavily contaminated soils were reproducible (6% RSD at 3500 mg TNT kg^{-1}).

Griest et al.[68] described the separation of explosives using packed-column supercritical fluid chromatography (SFC). The eluent employed was SFC-grade carbon dioxide without modifier. The analytes were eluted from microbore Deltabond® Cyano columns (2.50 or 1.50 m × 1 mm i.d.). The instrument employed both a variable-wavelength UV detector (12 μL flowcell) and a flame ionization detector (FID) in series, enabling facile detection of both the nitroaromatic and nitramines. Eight explosives eluted in the order 2,6-DNT < 2,4-DNT < NG < TNT < PETN < tetryl < RDX (some tailing) ≪ HMX (was not observed). The authors postulated that the addition of a polar modifier such as methanol would have allowed HMX to elute from the column and be detected.

Miyares[69] explored the possibilities of SFC further to enable the facile elution of HMX, which was not possible using the conditions described in Griest et al.[68] The author employed short narrow-bore SB-Cyanopropyl columns (2.5 m or 5.0 m × 0.05 mm i.d.). HMX was finally eluted as a clean sharp peak with supercritical fluid grade carbon dioxide using the SB-Cyanopropyl-25 capillary column (5.0 m × 0.05 mm i.d.); however, resolution of the earlier-eluting explosives was sacrificed compared to that observed in Griest et al.[68] Miyares also compared the instrumentation and ease of method development between SFC and HPLC, and noted the following:

- SFC columns are smaller and more flexible than GC columns, and are much harder to install than HPLC columns. HPLC columns may be installed or removed in less than one minute, compared to up to 15 minutes for SFC columns.
- Method development involving reversed-phase HPLC columns is fairly straightforward because retention times are strongly correlated to octanol–water partition coefficients. It becomes reasonably easy for an experienced operator to choose the column phase, mobile phase, and gradient (if required). In SFC, the partitioning of analytes between the mobile and stationary phases is part of the partition mechanism for separation, but it may also include specific effects from density, temperature, linear velocity, and volatility of analytes. The temperature, density, and pressure of supercritical fluids are interrelated in such a way that for a given density, there are many temperature and pressure combinations. Consequently, considerably more trial-and-error is required for SFC method development than for HPLC method development.
- The standard HPLC method employs simple isocratic elution, which may be executed using even very unsophisticated pumping systems. In contrast, the SFC separation reported in Miyares required an intricate combination of temperature and density ramps which demanded a highly sophisticated pumping system.

Taken together, Miyares concluded that for routine environmental monitoring and site characterization required by the military, SFC offered no substantial advantages over the standard HPLC/UV methods.

Rodacy et al.[18] discussed the application of IMS both to aqueous samples, as discussed earlier, and soils. Explosives are desorbed thermally from soil samples prior to introduction into the IMS. Each explosive examined to date, including TNT, RDX, PETN, ethylene glycol dinitrate (EGDN), NG, HMX, and DNT have unique and repeatable delay times for any given temperature up to the point where sample decomposition occurs. The type of soil being analyzed does not affect the relative delay times because explosives desorbed from soils containing high sand content, high clay content, high humic matter, and pure quartz surfaces show the same relative delays. However, if one compares the length of time it takes for an explosive to be desorbed from, say, a sandy soil and a soil with high humic content, a slight difference in desorption times can be noted. The differences, however, are minor

and of no importance to the analysis. This combination of desorption profile with ion mobility allows accurate, unambiguous differentiation of the explosive species and eliminates many potential interferences. A prototype system was developed and used for the analysis of RDX, TNT, DNTs, and PETN in both water and soil samples. The system was challenged with 700–800 soil samples contaminated with explosives ranging in concentration from ''not detected'' to 50–60 ppb. Approximately 10% of these samples were ''split'' and analyzed both ''in house'' using the prototype method and by an independent laboratory using USEPA Method 8330. In all cases, there was excellent agreement between data generated for a given sample by the two procedures.[70]

Interferences to the IMS method found to date are extremely limited, and include such diverse sources as wintergreen-flavored snuff, Brut™ deodorant, some musk perfumes, and some freshly cut foams and fresh adhesive tapes. After having been exposed to the atmosphere for approximately two hours, the foams cannot be detected; after 6 to 12 hours, the tape can no longer be detected. These materials may be possible contaminants, but not interferences, because the drift times and desorption lag time allows unambiguous separation of these materials from the explosives. To date, the authors have not been able to detect true interferences which are native to soils.

In analyzing soil samples, the explosive vapors are detected as they are desorbed. Using a 200 milligram soil sample (optimal for the desorption furnace), the following detection limits were obtained: 0.5 ppb TNT; 3.0 ppb RDX; 0.5 ppb NG; 2.0 ppb PETN; 4.0 ppb EGDN. (The criterion for ''detection limit'' was not specified.)

The IMS shows a linear calibration curve from approximately one to 100 picograms of TNT and RDX. Larger quantities of explosives produce a nonlinear asymptotic response. This falloff could be attributed to one or more of the following phenomena: insufficient reactant ions in the source; space charging; and the increasing likelihood of ion–ion collisions to form neutral species.

At least two authors have noted difficulties in detecting TNT, in particular using a direct thermal-desorption inlet to an IMS, and have reported a much higher detection limit than Rodacy.[18] Some of these discrepancies may be explained by differences in both sample and instrument characteristics, as follows:[71,72]

- Explosives may be detected easily at low concentrations if they are artificially spiked onto a soil sample. The analyte does not have sufficient time to diffuse into the tiny pores on the surface, and is therefore removed easily.
- Explosives may be detected at concentrations exceeding hundreds of ppm (i.e. >0.01%) in authentic contaminated soil, but not at lower levels.
- The sensitivity of the IMS instrument is critically dependent upon the design of the desorption furnace (short vs long heated zones). The transfer line must be heated to avoid any possible cold spots which would promote analyte condensation prior to introduction into the ion source.
- TNT is detected as TNT, rather than as a degradation product.

Jenkins and Walsh[73] extended previously reported work concerning the application of GC/ECD in aqueous samples to include soils. Use of this approach permitted detection limits of about $1\,\mu g\,kg^{-1}$ for the important signature chemicals, without additional preconcentration. This method employs short (6 m) Megabore (0.53 mm i.d.) fused silica columns with deactivated injector liners and linear velocities of about $100\,cm\,s^{-1}$. The authors resolved 20 important nitroaromatic, nitramine, and aminonitroaromatic signature chemicals within a 15-minute analysis time using a 6-m HP-5 column. This method has been given preliminary approval by the USEPA as SW-846 Method 8095.

Yinon et al.[74] described the application of thermal pyrolysis/GC/MS for several explosives, including RDX, HMX, tetryl, monoaminotrinitrobenzene (MATB), diaminotrinitrobenzene (DATB), and triaminotrinitrobenzene (TATB). The final pyrolysis temperature was 400–1000 °C, held for 2 seconds. The nitramines RDX and HMX showed similar mass spectra upon pyrolysis, both of which were very different from those of the three nitroaromatics.

5 DETERMINATION OF EXPLOSIVES USING FIELD SCREENING PROCEDURES

5.1 Introduction and Motivation

The preceding discussions all tacitly assumed that the determination of explosives in either aqueous or soil samples would be performed (a) using standard well-validated methods (b) using highly instrumented analytical methods (c) in a well-equipped laboratory (d) in a timely manner. Such scenarios frequently are not helpful for routine site clean-up operations. ''Traditional'' analytical laboratories are rarely available at a clean-up site. Further, reliable analytical data needs to be available *immediately* so that remediation personnel may make correct and timely decisions. Lengthy delays imply that expensive excavation equipment and skilled operators must remain idle for indefinite periods of time. Crockett et al.[6] described additional complications and considerations as follows:

Reliance on laboratory analyses only for site characterization may result in a large percentage of samples (up to 80%, depending upon the site) with nondetectable levels. The remaining samples may indicate concentrations within a range of four orders of magnitude. Analyzing a small number of samples at an off-site laboratory may result in inadequate site characterization for estimating soil quantities for remediation and may miss potentially reactive material. ... Laboratory analytical costs vary depending on required turnaround time. Typical costs for USEPA Method 8330 range from $250 to $350 per sample for 30-day turnaround, $500 to $600 for 7-day turnaround, and approximately $1000 per sample for 3-day turnaround, if it is available. [Reproduced by permission of the Cold Regions Research and Engineering Laboratory from A.B. Crockett, T.F. Jenkins, H.D. Craig, W.E. Sisk, *Overview of On-Site Analytical Methods for Explosives in Soil*, Special Report 98-4, 3, 1988.]

Taken together, there is an urgent demand for analytical procedures which are:

- well-documented and validated;
- fieldable at clean-up sites;
- easily operated by nonchemical professionals;
- rapid;
- capable of producing valid analytical data quickly;
- inexpensive.

5.2 Description and Evaluation of Typical Colorimetric Field Screening Procedures

Many of the existing and successful analytical field test kits for explosives in environmental samples are based upon two well-documented chemical reactions. As early as 1891, Janowsky[75] observed that colored reaction products were observed when polynitroaromatic compounds reacted with alkali such as potassium hydroxide in the presence of acetone. Meisenheimer[76] and Jackson and Earle[77] independently proposed a final end product with a quinoidal structure to explain this phenomenon. In general, Jackson–Meisenheimer anions for dinitroaromatics are blue to purple in color and those from trinitroaromatics are red.[78] The formation of Jackson–Meisenheimer anions has historically been the basis of a reliable field test for nitroaromatics. The reactions are summarized in Scheme 1.

Colorimetric chemical methods have also been developed for RDX and the nitramines for forensic applications.[79] These procedures generally rely on sequential reactions where zinc metal converts RDX, for example, to nitrous acid (Franchimont reaction). The resulting nitrous acid is used to nitrosate an aniline derivative such as sulfanilic acid; the resulting diazo cation is then coupled to a naphthylamine to form a highly colored azo dye (Griess[80] reaction). This reaction sequence is summarized in Scheme 2. Several pairs of reagents may be

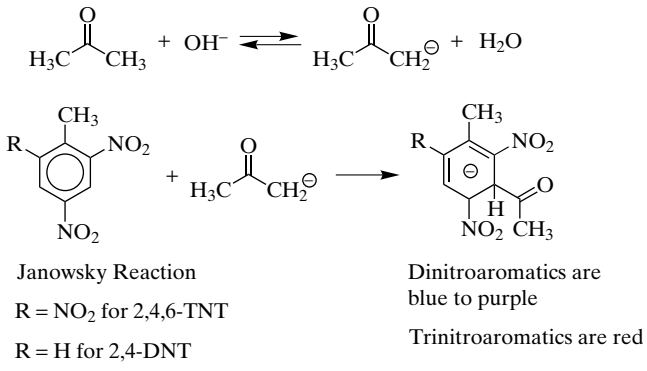

Scheme 1 Reactions used for the colorimetric determination of TNT and DNT. [Reprinted from T.F. Jenkins, M.E. Walsh, *Talanta*, **39**, 420 © 1992 with permission from Elsevier Science.]

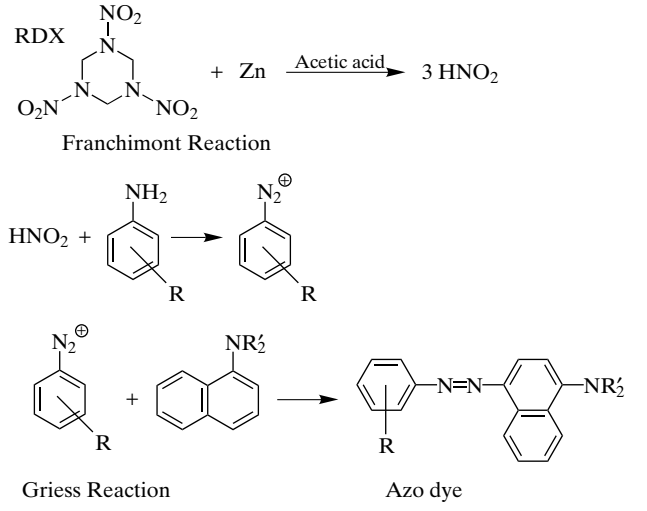

Scheme 2 Reactions used for the colorimetric determination of RDX. [Reprinted from T.F. Jenkins, M.E. Walsh, *Talanta*, **39**, 420 © 1992 with permission from Elsevier Science.]

used to produce azo dyes. A reagent containing procaine and *N*,*N*-dimethylnaphthylamine was initially used for the procedure described in Jenkins and Walsh.[81] This choice was based on the work of Wyant, who tested several reagents and found this combination best in terms of detection capability and shelf life.[82] However, this liquid reagent is light-sensitive, and therefore cumbersome to handle in the field. In addition, there was some concern about the possible cancer risk associated with *N*,*N*-dimethylnaphthylamine. Further work demonstrated that a Hach NitriVer 3 ''powder pillow'', which is specifically designed for the determination of nitrite in the field, and distilled water, can be substituted for the liquid Griess reagent. The powder pillow contains sodium sulfanilate, 4,5-dihydroxy-2,7-naphthalene-disulfonic acid disodium sulfate, potassium phosphate monobasic, potassium pyrosulfate, and *trans*-1,2-diaminocyclohexane tetra-acetic

acid trisodium salt. Abramovich-Bar et al.[83] described an additional set of workable Griess reactants. Here, a diazonium salt is formed from sulfanilamide and nitrite in the presence of 10% phosphoric acid. This product, in turn, reacted with *n*-naphthylethylenediamine to produce a red azo dye compound.

Jenkins and Walsh[81] described colorimetric field screening procedures (the ''CRREL Methods'') for explosives. Approximately 20 g wet soil is shaken with 100 mL acetone, and the extract filtered with a disposable syringe filter. The first test is designed to detect nitroaromatics. Portions of the extract are reacted with a strong base and sodium sulfite. If only TNT is present, a reddish Janowsky complex is produced. If only DNT is present, a bluish-purple complex is produced. If a mixture is present and TNT dominates, DNT will not be detected. The second test is designed to detect nitramines, primarily RDX. A second portion of acetone extract, which is first passed through a disposable anion exchange cartridge to remove native nitrate or nitrate, is acidified and reacted with powdered zinc. The nitrous acid so produced is reacted with the contents of a Hach NitriVer3 powder pillow and distilled water, as described above. The development of a pink Griess reaction product is indicative of the presence of RDX or one of several other military explosives which are potential interferents (HMX, NG, PETN, or nitrocellulose (NC)). A battery-operated spectrophotometer may be used to quantitate the products so formed at 540 nm (for TNT and nitroaromatics) and 507 nm (for RDX and the nitramines). While absorbance of the Janowsky product is greater at 460 nm than at 540 nm, the latter wavelength was selected to reduce interferences from yellowish humic materials that are also extracted into acetone and absorb strongly at 460 nm.[1] The detection limit was 1 $\mu g\, g^{-1}$ for TNT and RDX, with linear range extending to 50 $\mu g\, g^{-1}$ for TNT and 20 $\mu g\, g^{-1}$ for RDX.

Medary[84] reported a related screening procedure for TNT in which methanol, not acetone, was the extraction solvent. TNT was estimated for the absorbance of the colored Meisenheimer anion produced at 516 nm, and a detection limit of 4–8 $\mu g\, g^{-1}$ is estimated. The major advantages of the procedure described in Jenkins and Walsh[81] include a four-fold greater absorbance of the colored product and better extraction of the nitramines when acetone, not methanol, was employed. The same authors investigated the extraction recovery of the proposed field method by both shaking aliquots of soil manually (3 min) and ultrasonically extracting identical portions for 18 hr with acetone. The average recovery after 3 min of manual shaking with acetone for TNT was 96% and for RDX was 98% of that obtained with the more exhaustive procedure, indicating that acetone is an excellent extraction solvent for these two analytes over a wide concentration range. Subsequent studies indicated that longer extraction times (up to 30 minutes) were useful for soils with a ''heavy'' clay content.

The color-forming reactions used for these field screening methods are not specific for TNT, 2,4-DNT, and RDX. Other polynitroaromatics such as 1,3-DNB and 1,3,5-TNB and polynitrophenols such as picric acid (PA) also give colored anions when reacted with strong base. During site clean-up activities, however, the ability to detect these other compounds, as well as TNT and 2,4-DNT, would be quite useful. Similarly, the same azo dye produced from the RDX test is also produced when other nitramines such as HMX and tetryl or nitrate esters such as NG, PETN and NC are treated under similar conditions. Heavy metal cations such as copper were found to interfere with the 2,4-DNT determinations. These cations could form complexes with either the unreacted DNT or the Janowsky complexes.

Absorbance varied little for the TNT and RDX solutions over the range of moisture contents (10–75%) expected for the large majority of surface soils from potentially contaminated sites. However, the absorbance for the 2,4-DNT standard significantly declined as the water content exceeded 10%. Based on this variability, determinations for 2,4-DNT will be semi-quantitative, while the corresponding procedures for TNT and RDX may be used quantitatively.

The extraction procedure described in Jenkins and Walsh[81] was field-tested extensively with authentic contaminated soil samples. The concentrations obtained in the field for TNT, RDX, and 2,4-DNT were compared with those using USEPA Method 8330, which was described previously. In all cases, there was excellent agreement obtained between the concentrations obtained using the two procedures; the typical recovery using the field procedure exceeded 90%. These data are summarized in Table 9. The success of the colorimetric method for TNT led to its adoption as a standard analytical method, entitled USEPA Method 8515, ''Colorimetric Screening Method for Trinitrotoluene (TNT) in Soil''. The RDX method has also undergone standardization as SW-846 draft method 8510.

5.3 Description of Antibody-based Field Test Kits

Recent alternatives to the classical colorimetric assays for explosives, such as those described above, are field tests employing enzyme-linked immunosorbent assays (ELISA).[85] One such assay for TNT involves several steps. In step 1, the plate is precoated with a TNT-protein conjugate that adheres to the well surface. In step 2, competitive inhibition occurs when samples containing TNT are incubated with mouse anti-TNT antibodies. The antibodies can either bind to the free TNT in the

Table 9 Comparison of extraction efficiency of field and standard laboratory procedure

Sample origin	Concentration ($\mu g\,g^{-1}$)			
	Analyte	Field procedure[a]	Lab procedure[b]	Recovery, %, by field procedure[c]
Nebraska ordnance plant	TNT	0.065	0.071	91.5
	TNT	340	349	97.4
	TNT	63.5	67.9	93.5
	TNT	0.32	0.32	122
Hawthorne AAP (NV)	TNT	4.53	4.75	95.4
	TNT	5.79	5.65	102
	TNT	0.79	0.90	87.3
Weldon Springs (MO)	TNT	0.96	1.26	76.2
	TNT	163	176	92.6
	TNT	0.075	0.077	97.4
Vigo Chem. Plant (IN)	TNT	11.7	13.4	87.3
Hastings East Ind Park (NE)	TNT	67.6	68.8	98.3
Sangamon ordnance pt (IL)	TNT	21.5	23.2	98.2
Raritan arsenal (NJ)	TNT	71.7	80.6	98.0
Lexington–Bluegrass Depot (KY)	TNT	5.90	7.11	83.0
Chicksaw ordnance works (IN)	TNT	0.21	0.16	131
Nebraska ordnance plant	RDX	13.6	14.1	98.3
	RDX	60.2	65.9	95.5
	RDX	1073	1080	99.7
	RDX	9001	10 455	92.6
Hawthorne AAP (NV)	RDX	1.97	2.01	99.0
	RDX	3.32	2.96	105
Lexington Bluegrass Depot (KY)	RDX	9.10	9.37	98.5
Camp Shelby (MS)	2,4-DNT	3.4	4.2	80.9
	2,4-DNT	226	563	40.1
	2,4-DNT	6.7	7.3	91.8
Eagle River Flats (AK)	2,4-DNT	12.7	13.6	93.4
	2,4-DNT	7.4	7.7	96.1

(Reprinted from T.F. Jenkins, M.E. Walsh, *Talanta*, **39**, 425 by permission of Elsevier Science © 1992.)
[a] 20 g soil shaken with acetone for 3 min.
[b] 20 g soil extracted with acetone for 18 hr in sonic bath.
[c] Value calculated relative to the laboratory procedure.

samples or to the TNT bound to the surface of the well. After an equilibrium is reached, the plate is washed and the antibodies bound to the free TNT are washed away, leaving the antibody bound to the immobilized conjugate. The greater the concentration of TNT in the sample, the less monoclonal antibody is available to the immobilized TNT conjugate. Therefore, the amount of antibody bound to the well is inversely proportional to the TNT concentration present in the sample. In step 3, the plate is incubated with goat anti-mouse (μ-chain specific)-alkaline phosphatase conjugate which bonds only to the surface-bound mouse monoclonal anti-TNT antibodies. The plate is washed again. In step 4 the enzyme substrate is added. The surface-bound enzyme–antibody conjugate converts the colorless substrate to a yellow product. The color is directly proportional to the amount of TNT monoclonal antibody bound in each well, and, therefore, is inversely proportional to the amount of TNT in the sample. Mouse monoclonal anti-TNT antibody for TNT is specific for TNT and shows no cross-reactivity with those explosives or interferent tested in this study. The assay was capable of detecting as little as 50 pg TNT.

Musks, as well as a number of other chemicals, are known to present potential interferences with chemical or instrumental means of explosive detection. For example, synthetic nitro musks are similar in chemical structure to TNT and are common fragrance additives in toiletries. To determine the effects on the assay, 100 μL of 2 undiluted musks perfumes – an extreme quantity – were added directly to the microtiter plates in triplicate. In these cases, some inhibition of color development was observed. This may be due to partial stripping of the coating antigen from the plate or binding to the monoclonal anti-TNT antibody.

Westinghouse's Law Enforcement Products Division has developed an antibody-based method for PETN using a latex-bead agglutination technology.[85] Agglutination reactions are widely used throughout the biochemical community for the detection of small quantities of

proteins and drugs in physiological fluids. The time required to complete a PETN assay was reduced to 30 seconds. The sample collection device and assay are being incorporated into a disposable item which is analyzed using a reusable hand-held spectrophotometer. In step 1 of the assay, agglutinated latex particles form when the latex particles (<1μm particles) which have covalently bound to their surface the analyte of interest, are mixed with the specific antibody. Because of the multiple binding sites on both the latex particles and the antibody, a crosslinking or polymerization of many particles occurs. In the presence of the analyte of interest, this agglutination is inhibited. This agglutination reaction can be monitored both visually and electronically in about 30 seconds. PETN antibodies from one particular rabbit showed detection as low as 30 parts per trillion using an ELISA assay similar to the TNT assay. Recent preliminary results of the agglutination assay have demonstrated a detection limit of 25 ng of PETN.

Bart et al.[86] described a continuous-flow immunosensor (CFI) which would be capable of fulfilling the mission of an explosives detector for both environmental analysis and bomb detection. The CFI is a semi-automated system in which the analyte-containing medium (generally an aqueous sample which is spiked with a concentrated solution of buffer, organic co-solvent, and a surfactant) is pumped through a column containing antibodies immobilized on beads. These antibodies are saturated with a fluorescent analog of the analyte before the analysis begins. As the analyte passes through the antibody matrix, some of the analyte binds to the antibody, displacing the labeled analog. The effluent from the column is thus a mixture of analyte which did not bind to the antibodies and dye-labeled antigen which was displaced by the analyte; the fluorescent analog is then detected in a fluorometer located downstream from the column. The amount of dye-labeled antigen displaced is directly related to the amount of analyte in the sample. A prototype system was assembled and evaluated for the evaluation of RDX in aqueous samples over the range 15–1200 ng RDX mL^{-1}. The lowest detectable limit was 20 ng RDX mL^{-1}. The cross-reactivity of other explosives with the anti-RDX antibody was modest for HMX (15% response vs RDX), and minimal for TNT and its derivatives (typically <1%). This behavior was expected because of the structural similarities of RDX and HMX, and dissimilarities of RDX and TNT. When both the CFI and standard HPLC methods were used to analyze several groundwater samples for RDX, the agreement between the two methods was excellent. Kusterbeck and Charles[87] described the extensive field demonstration of a similar device, the FAST 2000, for the determination of TNT at two Department of Defense Superfund cleanup sites. The procedure was cost-effective and capable of screening a number of samples in a short period of time at a project cost of approximately $3–$10/sample, compared to $250–$1000/sample using the standard HPLC method.

5.4 Evaluation of Commercially Available Field Screening Kits

The frequency of occurrence of specific explosives in soils was assessed using analytical data on soils collected from 44 army ammunition plants, arsenals, and depots, and two EOD sites.[3] Of the 1,155 samples, a total of 319 samples (28%) contained detectable levels of explosives. TNT was detected in 66% of the samples and 80% of the samples if the two EOD sites are excluded. Overall, TNT or RDX or both were detected in 82% of the samples containing explosive residues, and 94% if the ordnance sites are excluded. Thus, by screening for TNT and RDX at these facilities, 94% of the contaminated areas could be identified (80% if only TNT was determined). This demonstrates the feasibility of screening for one or two compounds. Screening for all of the explosives listed in USEPA Method 8330 was not necessary to evaluate the degree of contamination in a particular site.[1,20,88] For that reason, several manufacturers now offer commercially available test kits which focus on the determination of TNT and RDX. These kits have been evaluated extensively by many independent authors, with their performance summarized[6,88,89] in the following general comments:

- A USEPA study[90] compared the CRREL, EnSys (colorimetric), D Tech, Quantix, and Ohmicron methods for TNT. Overall, no single method significantly out-performed other methods, and the accuracies for all the field screening methods were comparable.
- Myers et al.[91] compared the EnSys colorimetric test to a commercial enzyme immunoassay (EIA) procedure that utilizes an immunofiltration approach (D Tech) using 99 soil samples obtained from the NSWC. The results from both methods were compared with analyses conducted using the standard reversed-phase high-performance liquid chromatography (RPHPLC) laboratory procedure (SW-846 Method 8330). Quantitative agreement between the field screening results and the RPHPLC results was somewhat better for the colorimetric method; however, a portion of this advantage may be due to larger sample sizes utilized in the EnSys protocol, which results in reduced sampling error.

- Haas and Simmons[92] evaluated immunoassay kits for TNT (EM Science/D Tech, EnSys/EnviroGard Tube and Plate, Idetek/Quantix, and Ohmicron/RaPID Assay.) They concluded that for semiquantitative screening, all kits have the potential to accurately screen soil samples for contamination at risk-based levels. The correlation between the immunoassay kits and HPLC was generally good above the 1 ppm level.
- A study conducted by the USEPA Region 10[93] and reported in Jenkins et al.[1] compared two versions of the colorimetric test (CRREL[81] and the EnSys commercial version) and two commercial immunoassay methods (D Tech and Quantix) during Phase 1 of the explosives washout lagoon soil remediation project at the Umatilla Army Depot. Samples were also analyzed using RPHPLC Method 8330; these results served as the basis for comparison. Overall, the CRREL colorimetric method was judged the least expensive for large sampling efforts and the most accurate for concentrations above $30\,\mu g\,g^{-1}$, while the D Tech EIA procedure was judged the most accurate below $30\,\mu g\,g^{-1}$. The Quantix method, which utilizes an EIA approach in the plate format, was judged the most efficient method for large numbers of samples. Overall, this study confirmed the value of the use of field screening for explosives-contaminated sites.
- A major field demonstration and evaluation of on-site analytical methods was performed during full-scale remediation at the Umatilla Chemical Depot (Hermiston, OR) and US Naval Submarine Base (Bangor, WA) Superfund sites. An additional evaluation involved pilot scale composting at the NSWC (Crane, IN) RCRA Corrective Action site.[94] Compost samples were analyzed by each of the on-site methods and results compared to standard laboratory analyses employing HPLC. The on-site methods evaluated included EnSys TNT and RDX colorimetric methods (USEPA SW-846 Methods 8515 and 8510) with and without an organic matrix cleanup step, and D Tech TNT and RDX immunoassay methods (USEPA SW-846 Methods 4050 and 4051). This study indicated that both colorimetric and immunoassay methods may be used for determination of explosives concentrations in compost residues during bioremediation treatment.

5.5 Field Test Procedures for Picric Acid and Ammonium Picrate

Ammonium picrate (AP; Explosive D; Yellow D) was used in armor-piercing shells, bombs, and rocket warheads by the US military from the turn of the century until after World War II. Although it is no longer manufactured, AP represents 8% of the demilitarization inventory. Picric acid (PA) was used in grenades and mine fillings. Unlike many of the other high explosives that are no longer manufactured and present environmental clean-up problems unique to the military, PA is a common industrial chemical, used widely for metal etching and as a feedstock in many processes in the dye, leather, and glass industries.[95]

Thorne and Jenkins[95] described the optimization of a field analytical method for AP/PA in considerable detail, with the final procedure sketched as follows: soils are extracted with acetone in a manner similar to the colorimetric screening methods for TNT and RDX. A 30-mL filtered aliquot is mixed with 30 mL of deionized water and passed through a 3-mL SPE-Alumina-A (Supelco) cartridge. Picrate ion is retained on the alumina. Interferences are removed by passing a 3-mL aliquot of methanol through the cartridge, followed by a 3-mL aliquot of acetone. Picrate is converted to PA with a solution of 10 mL acetone containing four drops of sulfuric acid and eluted from the cartridge. The solution is then diluted with 5 mL of deionized water and 15 mL of acetone. A change from the colorless (PA) to yellow (picrate) is indicative of the presence of PA/AP in the soil at a concentration above $3\,\mu g\,g^{-1}$. The picrate concentration can be estimated from the absorbance of the solution at 400 nm with a MDL of $1.3\,\mu g\,g^{-1}$. In the same study, concentrations of picrate were measured in 49 contaminated soil samples using both the field method described and RPHPLC. The results obtained exhibited an excellent correlation over three orders of magnitude. The field method did not produce false positives.

5.6 Field Screening Procedures for Explosives in Soils using Ion Mobility Spectrometry

IMS was evaluated as a potential field screening procedure for explosives in soil.[96] The US Army provided a large number of soil samples that had been collected from three locations at each of three explosive contaminated installations, dried, ground, homogenized, and previously analyzed in duplicate using Method 8330. Duplicate two-gram aliquots of these samples were extracted with 10 mL of acetone by shaking for three minutes. After solids had been allowed to settle, the quantities of ''typical'' explosives were determined using IMS. For field application, the extraction procedure should be changed to 20 g moist soil, 100 mL acetone. The chief limitation of the IMS method was the initial cost of instrumentation ($\sim$\$50 K).

6 TRANSFORMATION AND METABOLISM OF EXPLOSIVES IN THE ENVIRONMENT

6.1 Expected Reaction Pathways

The presence of species such as TNB and the various isomers of poly-amino poly-NT suggest that TNT residues in soil may be transformed by a variety of photochemical and/or microbiological processes. While the transformation pathways of some explosives have been studied in cell cultures, composting systems, and water, little research has been conducted to define what by-products are present in explosives-contaminated soil. Griest et al.[97] have reported the aquatic toxicity of TNT and its metabolites using the *Ceriodaphnia dubia* assay. The same source discussed the mutagenic and genetic activity of the same compounds using the Ames test, Chinese Hamster Ovary (CHO) assay, and the Sister Chromatid Exchange (SCE) assay. Potential transformation products of TNT are numerous. Major et al.[8] outlined the following two major and one minor transformation pathways:

- An *oxidation* pathway, in which the methyl group of TNT is oxidized stepwise to the corresponding alcohol, aldehyde, and acid. The latter is ultimately decarboxylated to form 2,4,6-TNB.
- A *reductive* pathway, in which either the 2- or 4-nitro group of TNT is reduced stepwise to the corresponding hydroxylamine and amine. The reduction of the remaining nitro groups would proceed similarly, ultimately forming a 2,4,6-triaminotoluene (TAT) end product. Many bacteria and fungi are capable of performing this reduction.[39] The reduction only proceeds to the monoamine stage under aerobic conditions, and may occur abiotically.[98] If the medium considered is composted soil, which contains a significant quantity of humic organics, a significant fraction of the reduced TNT products will become covalently bound to the matrix.[99]
- A *condensation* pathway, in which monoamine products from the reductive pathway combine to form a variety of azoxy dimers, such as 2,2′,6,6′-tetranitro-4,4′-azoxytoluene (4,4′-Az) and in trace amounts 4,4′,6,6′-tetranitro-2,2′-azoxytoluene (2,2′-Az).[39]

The transformation of TNT does not necessarily mean a removal of its deleterious characteristics; the primary biotransformation products of TNT, which are the dinitromonoamines 2-Am-DNT and 4-Am-DNT, have also been shown to have toxic and mutagenic effects on certain biological species. The toxicity and mutagenicity of secondary and subsequent biodegradation intermediates of TNT, such as the diaminomononitroamines 2,6-Dam-NT, 2,4-Dam-NT, and the trinitroamine TAT, have yet to be explored. Therefore, simple and accurate methods for the identification and quantitation of TNT and its biotransformation products are needed so that acceptable mass balances can be established during biodegradation.[100]

6.2 Analytical Procedures for the Transformation Products

Several investigators have proposed and evaluated a variety of analytical methods to separate and quantitate these various products. Many of these employ either UV absorbance measurements or manipulation of the sample pH. In a very early work, Glover et al.[101] reported both pK_a values and molar absorptivities for the two monoamine and two diamine isomers (total four compounds). Ahmad and Roberts[100] used an Alltech Alltima RPHPLC column (RPHPLC C_{18}, 25 cm × 2.1 mm i.d., 5 μm particles) with a pH gradient to resolve the analytes. A diode array detector was programmed to allow both dual-wavelength detection (at 210 and 254 nm) and compound characterization (scanning between 200–600 nm). Baseline separation of aminodinitrotoluenes could not be achieved using the method presented, which used a strict reversed-phase column, but could be achieved using either a single mixed-mode column[40] or an LC-CN column in series with an RP C_{18} column.[38] Ahmad and Roberts[100] also studied the separation of the various tetranitroazoxy-toluenes, such as 4,4′-Az and 2,2′-Az. Mobile phases consisting of aqueous buffers adjusted to three different pH values in a gradient with acetonitrile were examined for their efficiency in separating the intermediate compounds and for the minimization of speciation of the ionizable intermediates. Solvent consumption was minimized by the use of a narrow-bore column (2.1 mm i.d.). Liquid chromatography/mass spectrometry (LC/MS) supplemented with a UV detector (214 nm) has been used for detecting both the monoaminodinitrotoluenes and 2,4-Dam-NT in the urine of rats fed TNT and the same analytes plus 2,6-Dam-NT in human TNT workers.[102] TNT and its principal metabolites were studied by direct injection electrospray ionization/mass spectrometry (ESI/MS).[103]

Several investigators have reported procedures in which the transformation products of TNT are analyzed using either GC/ECD or GC/MS, rather than HPLC. For example, Huang et al.[104] described a procedure for quantitating 2,6-Dam-NT in urine. Samples were initially neutralized with sodium bicarbonate, then extracted with toluene which was subsequently dried with anhydrous sodium sulfate. Quantitation of the underivatized analyte was performed using GC/ECD. Mußmann et al.[105] described a special procedure for the diaminonitrotoluenes in which the

pH of an aqueous sample was initially adjusted to 12. After batch extraction with dichloromethane and extract concentration, the analytes were derivatized with heptafluorobutyric anhydride (HFBA) or *N*-methyl-bis(trifluoroacetamide) (MBTFA), then separated and detected using GC/ECD. Schoene et al.[106] observed that derivatizations of the TNT metabolites and transformation products performed with *N*-(*tert*-butyldimethylsilyl)-*N*-methyl-trifluoroacetamide (MTBSTFA) were frequently incomplete. More satisfactory results were obtained when MTBSTFA derivatizations were supplemented with MBTFA, as noted above. Detection and quantitation of the derivatized analytes were performed using a gas chromatograph equipped with either a mass-selective detector (MSD) or atomic emission detector (AED).

6.3 Experimental Evidence for the Three Transformation Pathways

Caton and Griest[40] demonstrated clearly that the microbial transformation of TNT in composting material is strongly time-dependent; products both ''grow in'' and ''die out'' with time. The two monoaminodinitrotoluenes were observed in both the fresh composts and their leachates. The two diaminonitrotoluenes ''grew in'' as the monoaminonitrotoluenes decreased in abundance during the course of composting, and then decreased further as composting continued. Except for one of the azoxy dimers (4,4′-Az) in leachates prepared from soils in the early stages of composting, no other transformation products or reduction/oxidation metabolites of TNT (or HMX and RDX) were observed.

Jenkins and Walsh[2,107] evaluated eleven samples of authentic contaminated soils taken from six independent remediation sites for the presence of TNT transformation products. The concentration of the parent TNT in soil ranged from $1\,\mu g\,g^{-1}$ to $14\,mg\,g^{-1}$. Subsamples were extracted with acetone and analyzed by GC/MS. The most common transformation products were 2-Am-DNT and 4-Am-DNT, the initial microbiological reduction products of TNT. TNB, a photodecomposition product of TNT, was identified in 5 out of the 11 soils. Other transformation products identified in 4 out of the 11 soils were trinitrobenzaldehyde (TNBA), and 3,5-DNA. TNBA, like TNB, is a photodecomposition product of TNT, and converts to TNB by decarbonylation (loss of carbon monoxide from the aldehyde). 3,5-DNA is a microbiological reduction product of TNB. The conversion of 3,5-DNA from TNB in soil is consistent with the formation of 2-Am-DNT and 4-Am-DNT from TNT in soil.[3]

Caton et al.[108] investigated the transformation of explosives by common marine micro-organisms. Cultures of *Vibrio*, *Amphora*, or *Skeltonema* were exposed to carbon-14 labeled or unlabeled TNT or RDX coated on steel plates in micro-aquariums to simulate the leaching and transformations of explosives from unexploded ordnance, and the reactions and partitioning of the explosives were followed using HPLC and liquid scintillation counting. TNT was transformed into the same metabolites as have been observed in the composting studies. The water-soluble mono- and diamino-metabolites of TNT were observed in the aqueous phase, while the sparingly soluble (ca. 1 ppm) azoxydimers preferentially accumulated in the particulate phase. In contrast, RDX did not transform appreciably during the experiment, and it accumulated in the aqueous phase.

Hawari et al.[109] evaluated the biotransformation of TNT in anaerobic sludge supplemented with molasses using a combination of SPME/GC/MS and LC/MS. This work confirmed the existence of the reductive pathway, in which the end product is TAT. Under biotic conditions at pH 7.0, TAT is converted according to the condensation pathway first to azoxy dimers and later to polymers. In contrast, under abiotic conditions at pH 2–3, TAT is converted to stepwise to a variety of hydroxydiamine and dihydroxyamine isomers, ultimately yielding 2,4,6-trihydroxytoluene. The biotransformation of TNT to the various reduction products is very rapid, yielding almost complete disappearance of TNT within three hours. The production of TAT increases monotonically for at least seven hours.

Thorne and Leggett[99] hypothesized that reduced TNT products were covalently bound to matrices containing considerable quantities of organic matter, such as composted soil. They exhaustively extracted samples of this matrix contaminated with TNT under three distinct conditions, viz. (a) acetonitrile (removal of all unbound TNT and its reduction products); (b) 50% aqueous sulfuric acid; and (c) 0.5 M sodium hydroxide followed by concentrated sulfuric acid. The authors noted that after 40 days of composting, very little of the amino and diamino reduction products were released by acid hydrolysis. The acid hydrolysis did degrade the matrix substantially, but did not release additional TNT, even though the analyte is stable under the extraction conditions. When the composted soil samples were subjected to an initial base hydrolysis, the quantities of amino and diamino compounds released were roughly equal to those released by the acid-only hydrolysis. However, the subsequent acid hydrolysis of these base-treated samples released considerably more of the reduction products. The authors postulated that the transformation of TNT to solvent nonextractable reduction products appeared to go through two stages of covalent binding to the compost matrix. In the first stage, a significant percentage of the bonds are hydrolyzable by

acid alone or a combination of base and acid. As composting continues, a second stage occurs in which either the bonds are altered to form different functional groups that are nonhydrolyzable or additional bonds form as the bound transformation products are further reduced. This process may result in the potential for long-term releases of TNT metabolites that might not occur if composting is continued into the second stage of nonhydrolyzable binding.

7 DETECTION OF EXPLOSIVES IN AIR

7.1 The Challenge of Detecting Pure Explosive Vapors in Air

The challenge of detecting explosives in the environment is compounded substantially when the sample matrix is air, rather than soil or groundwater. The equilibrium vapor phase concentration of most explosives at 25 °C ranges between 6 and 18 parts per trillion for RDX and PETN, respectively, to 9.4 and 580 parts-per-billion for TNT and NG, respectively – a range of nearly eight orders of magnitude.[110]

Fetterolf[110] noted that attempts to detect pure explosive vapors, particularly in a forensic setting, often fall short of expectations.

> Consider the scenario of a briefcase containing a small amount of plastic explosive such as C-4 (RDX). A simple calculation, using Boyle's gas laws and the equilibrium vapor pressure of pure explosives at standard temperature and pressure, shows that a standard-sized empty briefcase (10.6 L) would contain only 6.3×10^{-11} g or 5.94×10^{-14} g mL^{-1} RDX vapor. This is, however, the best that one can expect. In reality, the amount of explosive vapor available for analysis is much less. These simple calculations fail to take into account Raoult's Law, which states that the true vapor pressure of the explosive will be reduced because it is a component of a mixture. Another factor to be considered is the kinetics or time dependence associated with vaporization and the absorption of explosives on surfaces. TNT has been shown to vapor deposit on a variety of surfaces. Griffy's kinetic model of explosive vapor concentration estimated that until a monolayer of TNT was deposited on the surrounding walls of a room, the vapor pressure was 6×10^{-4} less than the equilibrium vapor pressure. The practical implication of this result is that explosive detectors which function by collecting air samples are unlikely to be effective unless their sensitivity is orders of magnitude greater than that required to detect equilibrium amounts of the vapor. Finally, the emission rate of explosives from the concealment device must be taken into account. Lucero has estimated that this is approximately 1 nL/min. The emission rate of the explosive is also affected by any barrier material, (e.g. plastic wrap or foil) wrapped around it. [Reproduced by permission of the CRC Press © 1995 from D.D. Fetterolf, 'Detection and Identification of Explosives by Mass Spectrometry' in *Forensic Applications of Mass Spectrometry*, ed. J. Yinon, 240–241, 1995.]

Taken together, the physical limitations of the problem dictate that only the most sensitive analytical detection processes will be capable of detecting explosive compounds in air. Some of these are described in the following sections.

The ability to generate explosive vapors of known concentration is a critical concern for calibrating and testing any device used in the sampling of air for the presence of nearby explosives. For example, the Independent Validation and Verification (IV&V) facility (Explosives Detection Systems, a branch of the Federal Aviation Administration at Idaho National Engineering Laboratory) required pulses of explosive vapors between 1 and 1000 pg in less than 5 seconds with a high degree of accuracy. Furthermore, the mass output had to be varied during a test with only a few minutes between pulses. Davies et al.[111] designed and tested a device which employed a heated coil packed with a support containing the desired explosive. The generator was capable of producing pulses of TNT, RDX, or PETN with well-defined concentrations. A typical vapor generator calibration was performed at a fixed coil temperature and pulse width (5 seconds in this study) with the mass output varied by changing the carrier gas flow rate and/or pulse width. Varying the coil temperature to adjust the vapor generator mass output was not feasible in the IV&V tests because of the lengthy equilibration time required to reach a steady-state saturated vapor concentration.

7.2 Canine Detection of Explosives in Air

The traditional method for detecting explosives in air has been the use of canines trained to recognize the ''scent'' of these compounds. A major advantage of dogs is their mobility, a significant improvement over fixed detector installations, which need only be circumvented by a terrorist placing a bomb or smuggling explosives.[112] The use of canines, while widely practiced, exhibits several major drawbacks and concerns. The greatest of these is that there is no objective unbiased understanding of *exactly* how the dog senses the presence of explosives. Unlike machines, dogs do not come with precise specifications, including detection limit, probability of detection, mean time between failure, etc. Many internal and external factors can affect canine olfaction, including heat stress, toxins, injury, and even disease at pre-symptomatic stages. There's no ''press-to-test'' button for a dog to check if the ''system'' is working effectively – a major disadvantage.[113] In spite of these apparent disadvantages, trained canines may fulfill many of the missions assigned to instruments both

more effectively and at lower cost. A major advantage of dogs is that they actually locate suspicious items; at present, machines can only verify the presence or absence of explosives.

A dog's supposed ''supersensitivity'' is a myth. Rather, the dog's accurate and discriminating detection ability make it invaluable. It has never been demonstrated conclusively exactly what causes a trained dog to respond – and, similarly, what causes a false-positive alarm in a dog. A clear understanding of odor signatures should help explain false-positive responses. People have reported bomb-sniffing dogs ''alerting'' on materials unrelated to explosives. In such a situation, more dogs are brought in to sniff the material. If all the other dogs also alert, one concludes that it might be something which, to the dogs, ''smells'' like explosives. It doesn't mean the dog is wrong. The dog may have detected the odor it was trained to smell, but the ''scent'' may be coming from some other material that has the same odor signature as the target substance. For example, a dog trained to look for explosives based on ammonium nitrate may alert at the ''scent'' of ammonium nitrate in garden fertilizer.

Strobel et al.[112] estimated that there are over 19 000 known explosives compositions worldwide. For that reason, it is impractical to condition a dog to alert correctly for each. Nevertheless, the vast majority of explosives are composed of a relatively small number of chemical compounds or ingredients. For that reason, the Bureau of Alcohol, Tobacco, and Firearms have evaluated a new program in which dogs are trained only with pure explosives.[112] In this way, a trained dog should recognize any formulation containing that particular component.

7.3 Mass Spectrometric Detection of Explosives in Air

There has been a long-standing interest in an instrumental approach for trace explosives detection both as a complement to canine olfaction and as a means for explosives detection in scenarios in which canine olfaction is either impractical or inappropriate. For the detection of explosives vapors, gas-phase ionization techniques are the methods of choice. A variety of such methods has been applied to the analysis and detection of explosives vapors, including electron ionization, CI, field ionization, atmospheric pressure ionization, low-pressure electron capture, and atmospheric sampling glow discharge ionization (ASGDI). In general, ASGDI[114] is a simple and rugged means for ionizing organic species. The response is essentially instantaneous, with very little memory effect arising from analyte species that enter the ionization region. Chemical noise arising from the analyte deposited within the ionization volume, which ordinarily requires extensive and time-consuming cleaning with other sources, poses few problems even after many months of nearly continuous operation.

The analysis of negative ions in the explosives detection application is employed because, unlike most compounds in nature, explosives readily form anions. Several other approaches to explosives vapor detection take advantage of this characteristic, including IMS, atmospheric pressure ionization mass spectrometry, and GC/ECD.

Certain physical properties of the ASGDI in particular make it the ideal mass spectrometric source for the detection of explosives in air:

- It exhibits a simple and rugged design which provides a direct interface between the atmosphere and the vacuum system of the mass spectrometer. The ASGDI has no filaments or discharge needles to replace and can be operated continuously for months without maintenance. Despite rather extensive ion burning, source cleaning is not important. Some of the ion sources have been operated over the course of several years without cleaning and show no loss in performance.
- Its high throughput leads to rapid response and pump-out of analyte and minimized memory effects.

The ion-trap mass spectrometer (ITMS) can accumulate selected ions prior to detection; hence, it has the capability of detecting ultratrace concentrations of explosives in air. McLuckey et al.[115] coupled an ASGDI with an ITMS and challenged the new instrument with air containing low parts-per-billion concentrations of TNT. In order to improve the detection capabilities of the instrument still further, both NICI and tandem mass spectrometry were employed. The signal observed at m/z 227 (negative ion mass spectrum) corresponded to the molecular anion from ca. 100 fg TNT. The instrument described has proved to be capable of detecting organic explosives in air at the parts per 10^{12} (trillion) level.

The sensitivity and selectivity of ion-trap mass spectrometric measurements of explosives can be improved substantially by including a filtered noise field (FNF) as part of the ion collection process. FNF is a form of mass-selective ejection, in which undesired masses are removed from the mass spectrometer, and the desired masses are retained, collected, and amplified. When a FNF is applied during ion accumulation, interferences in the desired mass spectra are reduced significantly, and the desired ions, defined by frequency ''notches'', stand out very clearly and cleanly.[116,117] In FNF, a basic waveform having evenly spaced (typically 250, 500, or 1000 Hz) frequency components spanning the frequency spectrum of the entire ion-trap mass/charge range is synthesized digitally. Subsequent digital processing permits the waveform to be shaped so as to effect mass-selective accumulation

(via mass-selective ejection). Moreover, the user can optimize the FNF signal characteristics interactively. Two literature examples illustrate the power of FNF technology: (a) the clean determination of TNT, RDX, and PETN in air at sub parts-per-billion concentrations;[114] and (b) the determination of TNT and the polysulfur ion clusters (S_2•), (S_3•), and (S_4•) in black powder.[116] McLuckey et al.[118] have also applied random noise intentionally to the end caps of an ion-trap as a means of improving the overall selectivity of the device.

Fetterolf[119] discussed the application of IMS to the detection of explosives in air; previous sections in this chapter have described its application to water and soil matrices. In general, IMS possesses many of the desirable analytical figures of merit of conventional laboratory-based instruments such as good sensitivity, selectivity, and speed of analysis. They are often small and simple to operate. This ease of use and portability permit operation in real world scenarios not only as a forensic tool but also as an investigative tool. The Barringer IONSCAN Model 200 was evaluated as a ''typical'' IMS instrument. In the reaction region, atmospheric pressure carrier gas (purified air), hexachloroethane (C_2Cl_6), the reactant gas, and 4-nitrobenzonitrile, an internal calibrant are ionized by a ^{63}Ni beta emitter to form Cl-ions. The use of modified ion chemistry to improve the sensitivity and selectivity of the detection by IMS is quite common. The reactant ions can then undergo one or more ion/molecule reactions with an explosive molecule. In general, negative ionization provides a high degree of selectivity, as discussed earlier. The formation of multiple species greatly increases the specificity and reduces the possibility of false positives since the peak detection algorithm must recognize two of the three peaks for RDX, PETN, and NG before the audible alarm will sound. In one set of experiments, a subject deliberately touched C-4, an RDX-containing explosive, then subsequently touched various parts of an automobile. Samples were collected before and after the contact transfer from each of the touched areas of the car and from the subject's hands by vacuum onto a Teflon® filter disk. An alternative sampling method which has also proven successful is to simply wipe the suspected surface with the sample collector. Careful procedures were followed to ensure that no cross-contamination occurred by verifying that both the sample disks and the IMS were clear of explosives before sample collection. The car surfaces and hands were negative for RDX prior to touching the explosive, but showed easily detectable RDX residue after contact with C-4. In a separate experiment, after handling C-4, eight consecutive hand washes with soap and water were required before the IMS could no longer detect the RDX. Experiments involving the IMS frequently demonstrate presence or absence of a given explosive rather than providing quantitative information. Fetterolf[110] has reported the detection limits (criterion not specified) for TNT, RDX, PETN, and NG as 200, 200, 80, and 50 pg, respectively for at least one of the ions formed in the IMS.

Gieray et al.[120] and Reilly et al.[121] have reported a mass spectrometric method capable of characterizing individual airborne particulates of uranium oxide or diesel engine smoke. The same system was evaluated for real-time mass spectrometry of individual airborne and waterborne explosives particles.[122] Samples of TNT, RDX, and PETN were analyzed on individual particles generated from suspensions of micrometer-sized particles in water, or from solution in acetone. Suspensions and solutions were nebulized and the resulting droplets dried and introduced into the ITMS. When individual particles arrived at the center of the ion-trap, a focused XeCl (308 nm) or ArF (193 nm) laser was used to both ablate and ionize material from the particle. Once the particulate matter was ionized, it was analyzed using a variety of ion-trap-based mass spectrometric methods. Negative ion mass measurements of the three explosives clearly show that acetone, but not water, is involved in the observed spectra obtained using the XeCl excimer laser. Furthermore, PETN particles apparently decomposed when ablated by the XeCl laser pulses, yielding mostly nitrate ions at m/z 62. PETN was also studied with an ArF laser at 193 nm, where the analyte absorbs strongly with little absorption at 308 nm. The work demonstrated that it was possible to generate mass spectra of individual micrometer-sized particles of TNT, RDX, and PETN, either as aqueous suspensions, as dry airborne particles, or as particles nebulized from solution. Preliminary estimates of method sensitivity were not presented.

7.4 Additional Procedures for Explosives in Air

Jenkins and Walsh[73] discussed the determination of explosives in air using SPME fibers as collection devices. Polydimethylsiloxane/divinylbenzene fibers were particularly efficient at collecting vapors of the isomers of DNT, DNB, and TNT. The use of this technique, combined with GC using a new micro ECD, has resulted in the ability to detect many of these energetic chemicals below 1 picogram, with concentration detection limits as low as 10^{-14} g mL^{-1} of air.

8 ADDITIONAL PROCEDURES FOR THE DETERMINATION OF EXPLOSIVES

The procedures which have been discussed in this chapter have been developed using bona fide environmental

matrices, viz. air, groundwater, soil, and compost. A large body of literature exists which describes the detection of explosives in solutions or unusual samples such as hand swabs used in forensics analysis. Some of these procedures will be mentioned here to indicate additional analytical tools available for detecting these compounds in the environment.

8.1 Procedures Featuring Chemiluminescent Nitrogen Detection of Explosives

The use of CLND for the gas- or liquid-chromatographic detection of explosives was mentioned earlier. (For many years, the only CLND device available was the ''TEA'', a product of Thermo Electron Corp. For this reason, the two acronyms are frequently, but inaccurately, used synonymously.) Several investigators have expressed a preference for the ECD because of the latter's greater sensitivity. The CLND exhibits greater selectivity towards nitro-containing compounds and is more tolerant of matrix interferences than the ECD. For that reason, the degree of extract cleanup required is frequently reduced. The explanation for both of these features lies in the three-step operating mechanism of the CLND: (a) combustion of the analyte at 900–1100 °C to form NO, carbon dioxide, water, and other oxides; (b) reaction of NO with ozone to form ''excited state'' NO_2^*; and (c) relaxation of NO_2^* back to ground state NO_2 with the emission of a photon of light in the near-infrared.[123] In general, GC of explosives requires short (typically 6 to 12 m)[124] narrow-bore (250 μm diameter) columns, compared to more commonly used dimensions of 30 to 60 m and 320 to 530 μm, respectively. Thin films are preferred to reduce peak broadening caused by low diffusion velocities.[125] The TEA must be operated in its ''total nitrogen'' mode, in which the furnace temperature is frequently 900 to 1000 °C, to combust the analytes and convert all nitrogen present to nitrogen oxides.[126–129] The dead volume of the TEA or CLND is typically greater than that of a contemporary ECD; hence, some additional peak broadening should be expected when the chemiluminescence detectors are employed.

Douse[130] described the characteristics of a capillary supercritical fluid chromatograph equipped with a CLND. The analytes were eluted through a short nonpolar column (6.8 m × 0.05 mm i.d. SB Octyl 50 Superbond cross-linked flexible silica capillary, 0.25 μm film thickness) with supercritical carbon dioxide (no modifier). Minimum detectable levels (not defined) ranged from 20 to 60 pg for a representative group of compounds: NG (23 pg); PETN (40 pg); tetryl (60 pg); HNBB, hexanitrobibenzyl (55 pg). Compounds containing very polar functional groups, e.g. amino (polynitrodiphenylamines and nitroguanidine) and hydroxy (PA, styphnic acid, and NC) could not be analyzed. In addition, HMX exhibited poor peak shape, probably due to its low solubility in supercritical carbon dioxide. The use of organic modifiers or a more polar supercritical fluid might improve the chromatographic analyses of these compounds. The author noted that some explosives which cannot be analyzed by capillary GC, due to their involatility or instability, were readily analysed by capillary SFC. In addition, capillary SFC/TEA appeared to be more resistant than capillary GC/TEA to contamination by coextractives.

Kolla et al.[131] also noted that the efficiency of SFE, which may be used to extract explosives from soils and related samples, depends not only upon the chemical nature of the analyte but also on the nature of the supporting surface. The influence of the matrix was studied by spiking soils of different carbon content with PETN and NG. The recovery of NG drops steeply if it is mixed into soil of a high carbon content. The recovery of PETN is less influenced than NG by the carbon content of the matrix. If soils with a high humus content are tested, higher pressures and longer extraction times than usual are recommended. As the polarity of the explosive analyte decreases, its retention time from reversed-phase columns increases and its recovery rate in SFE improves. This effect is explained by the increased solubility of nonpolar explosives in the nonpolar supercritical carbon dioxide medium.

In a very recent paper, Bailey et al.[132] described the separation of explosives using capillary electrochromatography (CEC). This can be viewed as a hybrid technique between HPLC and capillary electrophoresis. In CEC, neutral species are separated on the basis of differential partitioning between the mobile and stationary phases, while charged species are additionally separated according to differences in electrophoretic mobility. Under isocratic conditions, using capillaries packed with ~20 cm of 1.5 μm nonporous octyldecylsilica particles and moderate electric fields ($<1000\ V\ cm^{-1}$), baseline separations of all 14 of the Method 8330 components were achieved in under 7 min. Using shorter columns and stronger fields, 13 of the 14 components were separated in less than 2 min. These analysis times represent major improvements in turnaround time compared to the traditional HPLC methods, which typically require 30 min per determination or more. The addition of modifiers such as sodium dodecyl sulfate and/or 2-(*N*-morpholino)ethanesulfonic acid monohydrate reduced tailing in the analyte peaks.

8.2 Procedures Featuring Mass Spectrometric Detection of Explosives

The reason for the outstanding sensitivity of the ECD towards certain explosives is the electron-capturing

nitro groups (up to four) on a given analyte. These same electron-capturing properties also make explosives amenable to highly-sensitive and selective mass spectrometric detection featuring NICI, which can permit greater selectivity and sensitivity than the more traditional and general EI ionization. The mass spectrometer in a given ionization mode is usually coupled with a gas, liquid, or supercritical fluid chromatograph both to separate co-extracted interferences from the analytes and to improve resolution of the explosives from each other and matrix components.

Mass spectra were obtained for several explosives (TNT, tetryl, NG, PETN, RDX) using ''direct exposure CI mass spectrometry'', a form of NICI, with isobutane as the reagent gas.[133] Similar data appears for TNT, NG, PETN, RDX, and HMX using NICI with isobutane as the reagent gas in two additional references.[134,135] A detailed discussion of the EI and CI spectra of explosives is given elsewhere.[136,137]

Huang et al.[104] discussed the application of SFC and Mass Spectroscopy (MS) to the determination of explosives. The application of SFC/MS under EI conditions has been reported; however, it is more difficult to achieve EI conditions in an SFC/MS, compared to a GC/MS, system. CI, therefore, is the most common ionization method for SFC/MS. Methane, ammonia, isobutane, and the supercritical mobile phase itself frequently have been used as the primary reagent gases to generate positive-ion CI mass spectra. The columns employed are frequently short (3 and 6 m), narrow bore (50 and 25 μm internal diameters), with a thin film (0.25 μm thickness 25% biphenyl polysiloxane or SE-54 stationary phases). The supercritical mobile phase evaluated was carbon dioxide doped with either methane or ammonia. Four nitroaromatics, viz. 2,6-DNT, 2,4-DNT, TNT, and TNB were evaluated; the molecular ion was the most abundant for all of them except TNT. No further information about sensitivity or selectivity was provided.

HPLC is frequently preferred to GC as a sample injection method for the mass spectrometer because thermal transformation of the analyte prior to ionization is minimized or eliminated. Typical mobile phases for explosives such as TNT, RDX, PETN, tetryl, NG, and diethylene glycol dinitrate (DEGN) are acetonitrile or methanol–water (50 : 50 v/v).[138] For example, Yinon and Laschever[139] determined explosives in water using direct injection CI with water as the reagent gas. The detection limits (typical signal-to-noise ratio = 5) for six representative explosives ranged between 2 (p-NT) and 40 (RDX, tetryl) ng μL injected. A thermospray (TSP) LC/MS system has been used for the separation and identification of a series of standard and technical explosive mixtures, using positive and negative ion detection, with and without filament CI. The TSP interface appears to offer the greatest capability for analyzing explosives separated by HPLC with high sensitivity. The interface allows high flow-rates (near 2.0 mL min^{-1}) of aqueous solvents and volatile buffers into the system and provides the gentle ionization necessary to obtain molecular ions. The system was tested with individual compounds (TNT, RDX, HMX, PETN, and ammonium nitrate), a technical mixture of RDX/PETN, and a mixture containing TNT, NG, PA, ammonium nitrate, and 2,4-DNT.[140] Positive-ion chromatograms showed only RDX, HMX, and tetryl; PETN, AP, and TNT were not amenable to TSP ionization in the positive mode. On the other hand, when the instrument was operated in the NICI mode, all six components of the standard mixture were readily detected. Detection limits observed in the NICI mode were 200 pg (TNT), 1 ng (RDX, HMX, PETN), 5 ng (tetryl), and 50 ng (AP).

ESI (electrospray ionization) is a sample introduction procedure for HPLC which produces charged droplets in the ion source capable of ionizing species such as explosives. Yinon et al.[141] used such a system in the study of TNT, RDX, HMX, tetryl, PETN, NG, EGDN, and Semtex. The test compounds were eluted isocratically from a short minibore column (Keystone ODS Hypersil column, 15 cm × 2 mm i.d., 3 μm particle size) with either methanol–water (1 : 1) or acetonitrile–water (70 : 30) at 150 or 200 μL min^{-1}, respectively. Mass spectra were obtained in the NICI mode using a Finnigan LCQ ITMS. Standard solutions containing TNT, tetryl, RDX, HMX, and PETN were prepared, and aliquots were injected into the HPLC/MS system. The resulting responses were used to calculate the LODs. The mass spectrometer was operated in the selected ion monitoring (SIM) mode, using characteristic ion or ions for each compound, to provide enhanced sensitivity. The criterion for estimating the LODs was the smallest amount injected that yielded a signal-to-noise ratio greater than 3 for the mass chromatogram of the characteristic ions. LODs were found to be between 5 to 10 pg, corresponding to 1–2 ng mL^{-1} in solution.

Tamiri and Zitrin[142] employed short thin-film fused silica capillary columns in their measurement methods, which featured mass spectroscopic detection using both EI and CI (reagent gas methane). NG, 2,4-DNT, TNT, PETN, RDX, and tetryl emerged as well-resolved peaks. PETN and RDX occasionally failed to elute from longer (30 m) columns; hence, the shorter 15 m columns were preferred. Tetryl routinely degraded partially in the ion source to the stable well-characterized product *N*-methylpicramide, which was frequently used to confirm the presence of tetryl.

8.3 Procedures Featuring Nontraditional Analytical Methods

Some of the most promising new analytical methodologies for detecting explosives in environmental or forensic materials are various magnetic resonance methods which are still in their infancy. For example, hydrogen transient nuclear magnetic resonance[143] (HTNMR) has been found to be particularly useful for the practical detection of most explosives concealed in letters, parcels and baggage. The technique might therefore be applicable for soils. Explosives exhibit HTNMR signal relaxation properties that are different from most other common materials. This feature is absolutely essential to the successful application of this method for inspection since the vast majority of nonmetallic materials contain hydrogen and produce strong hydrogen nuclear magnetic resonance (NMR) signals. Additionally, selectivity is provided in HTNMR by the ^{1}H-NMR to ^{14}N-nuclear quadrupole resonance (NQR) level crossing effect. This provides useful selective signature data for some explosives and is also effective in drastically reducing the analysis time. The unique HTNMR signal relaxation properties have been found to be characteristic for almost all explosives of concern. This includes the high-energy military explosives as well as the more common commercial explosives such as dynamites and water gels used in civilian construction, excavation, and mining activities. Detection is accomplished through magnetic and electromagnetic fields used to sense the hydrogen contained within the materials of the item to be inspected. By analysis and processing of this total hydrogen NMR signal, any contribution to the response produced by hydrogen in explosives can be separated from that produced by the hydrogen in most other materials. Through this procedure the HTNMR response has been found to be useful for rapidly and selectively detecting NG-based dynamites, water gel explosives, C-4 (plastic explosives), RDX, TNT, PETN, smokeless powder, and most other high-energy and commercial explosives and propellants. Results of several thousand tests have demonstrated the general suitability of the HTNMR method for baggage, letter mail, and parcel inspection and have shown a high detection probability and a low rate of false alarms.

Buess et al.[144] discussed the application of ^{14}N pure NQR to the detection of RDX in various samples. RDX was an ideal target compound because it is not only an ideal NQR model compound, but it is also widely used in military applications. NQR is a particularly attractive approach to the problem of explosives detection, because the ^{14}N NQR absorption frequencies from crystalline materials are virtually unique. Hence, by looking for the nitrogen signal at, for example, the NQR frequency of RDX, only those nitrogens in RDX will be detected. Therefore, interferences are not expected with an NQR explosives detector. A literature search of the 10 000 compounds that have been examined by NQR failed to show any other NQR resonances sufficiently near to those of RDX that would constitute a plausible interference. This specificity feature should be contrasted with nuclear-based techniques, for which all nitrogen cross-sections are the same, independent of the chemical composition of the material. Subsequently, NQR would determine the presence of other classes of explosives, (e.g. PETN) by examination at the corresponding NQR frequency for PETN. In principle, such examination could be performed essentially simultaneously with the RDX inspection. NQR is the examination of nuclei with nuclear quadrupolar moments in the absence of a static magnetic field. Hence, in addition to its specificity, a second advantage of NQR is that no magnet is required, in contrast to NMR methods. For that reason, potential damage to magnetically recorded material is avoided; there is no exposure of large, static magnetic fields to personnel. Because a magnet is not required, the costs are lower than for an NMR system. Furthermore, NQR uses nonionizing radiofrequency (rf) radiation, whereas nuclear techniques employ ionizing radiation. There are some inherently unattractive features of NQR which must be addressed before an NQR explosives detector can be marketed commercially. For example, NQR signals are weak and are generally below the level determined by the random electrical noise in the rf coil which detects the signal. The signal-to-noise ratio is increased by repeating the experiment as rapidly as possible and adding up the results; the signal-to-noise ratio improves by the square root of the number of repetitions.

9 SUMMARY AND CONCLUSIONS

The interest in analytical procedures for explosives residues in environmental media steadily increases as more and more formerly utilized military sites are certified and converted to full- or part-time civilian use. Careful statistical studies have clearly demonstrated that analyses based on randomly-collected ''grab'' samples of soils frequently produce a false picture of the extent of explosives contamination. Well-designed sampling plans, which employ multiple samples taken from small areas and which consider the distribution of the soil particles, are mandatory for understanding the true extent of contamination. Because site remediation is such a costly venture, there is a strong need for analytical procedures which are capable of quantitating explosives

in soil and groundwater samples quickly and accurately, using equipment and methods which may be operated reliably by nontechnical personnel. For that reason, scientific interest is gradually moving away from the completely laboratory-based, statistically validated, and highly technical standardized methods, such as HPLC and GC and their mass spectroscopic analogs, towards the many fieldable analytical techniques. Some of these, such as ion mobility spectrometry and the various magnetic resonance techniques, are still in their infancy; their potential and utility are still to be fully developed. Others, such as the colorimetric and enzyme-based field test kits, have been fully tested and, in some cases, have been adopted as standard methods themselves. Still others, such as the various mass spectrometric methods, involve highly technical equipment which must be simplified before it can be used routinely in the field. However, the extreme sensitivities characteristic of these state-of-the-art instruments partially offset the need for extensive operator training. A variety of mass spectrometric techniques may yet be used routinely for explosives residue analysis in the field. Some of these new methods may be used in, or even be derived from, forensic applications.

ACKNOWLEDGMENTS

The author gratefully acknowledges the assistance and suggestions of Dr. Wayne H. Griest (Chemical and Analytical Sciences Division) and Drs. Sylvia S. Talmage and Annetta P. Watson (Life Sciences Division), all of the Oak Ridge National Laboratory, during the preparation of this manuscript. The author also acknowledges Dr. Thomas F. Jenkins, US Army CRREL, for kindly providing many of the CRREL documents which were cited in this chapter. Research was sponsored by the US Army's Rocky Mountain Arsenal pursuant to an Interagency Agreement with the US Department of Energy (DOE), DOE No. 1989-H077-A1. This article was prepared by Lockheed Martin Energy Research Corporation as the Management and Operating Contractor for DOE's Oak Ridge National Laboratory, under federal contract no. DE-AC05-96OR22464.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Lockheed Martin Energy Research Corporation. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

LIST OF SYMBOLS

H_c	Henry's Law Constant
K_{ow}	Octanol–water partition coefficient
X	Mean blank signal
σ	Standard deviation of the blank signal

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detector
AOAC	Association of Official Analytical Chemists
AP	Ammonium 2,4,6-trinitrophenoxide
ASE	Accelerated Solvent Extraction
ASGDI	Atmospheric Sampling Glow Discharge Ionization
ASTM	American Society for Testing and Materials
CEC	Capillary Electrochromatography
CFI	Continuous-flow Immunosensor
CHO	Chinese Hamster Ovary
CI	Chemical Ionization
CLND	Chemiluminescent Nitrogen Detector
CRL	Certified Reporting Limits
CRREL	Cold Regions Research and Engineering Laboratory
DATB	Diaminotrinitrobenzene
DEGN	Diethylene Glycol Dinitrate
DNB	Dinitrobenzene
DNT	Dinitrotoluene
ECD	Electron Capture Detector
EGDN	Ethylene Glycol Dinitrate
EI	Electron Impact
EIA	Enzyme Immunoassay
ELISA	Enzyme-linked Immunosorbent Assays
EOD	Explosive Ordnance Disposal
ESE	Environmental Science and Engineering, Inc.
ESI	Electrospray Ionization
ESI/MS	Electrospray Ionization/Mass Spectrometry
FID	Flame Ionization Detector
FNF	Filtered Noise Field
GC	Gas Chromatography
GC/ECD	Gas Chromatography Electron Capture Detection
GC/ITMS	Gas Chromatography/Ion-trap Mass Spectroscopy
GC/MS	Gas Chromatography/Mass Spectrometry
HFBA	Heptafluorobutyric Anhydride
HMX	Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine

HNBB	Hexanitrobibenzyl
HPLC	High-pressure Liquid Chromatography
HTNMR	Hydrogen Transient Nuclear Magnetic Resonance
IMS	Ion Mobility Spectrometer
ITMS	Ion-trap Mass Spectrometer
IV&V	Independent Validation and Verification
LC	Liquid Chromatography
LC/ELCD	Liquid Chromatography/ Electrochemical Detectors
LC/MS	Liquid Chromatography/Mass Spectrometry
LOD	Limit of Detection
LOQ	Limit of Quantitation
MATB	Monoaminotrinitrobenzene
MBTFA	*N*-Methyl-bis(trifluoroacetamide)
MDL	Method Detection Limit
MHT	Maximum Holding Times
MRD	Missouri River Division Laboratory
MS	Mass Spectroscopy
MSD	Mass-selective Detector
MTBSTFA	*N*-(*tert*-butyldimethylsilyl)-*N*-methyl-trifluoroacetamide
NB	Nitrobenzene
NC	Nitrocellulose
NG	Nitroglycerine
NICI	Negative Ion Chemical Ionization
NMR	Nuclear Magnetic Resonance
NQR	Nuclear Quadrupole Resonance
NSWC	Naval Surface Warfare Center
NT	Nitrotoluene
OB/OD	Open Burn/Open Detonation
PA	2,4,6-Trinitrophenol
PDMS	Poly(dimethylsiloxane)
PETN	Pentaerythritol Tetranitrate
PICI	Positive-ion Chemical Ionization
RCRA	Resource Conservation and Recovery Act
RDX	Hexahydro-1,3,5-trinitro-1,3,5-triazine
READ	Reversal Electron Attachment Detection
rf	Radiofrequency
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RSD	Relative Standard Deviation
SCE	Sister Chromatid Exchange
SDB-RPS	Styrene–Divinylbenzene 'Reverse-phase Sulfonated'
SEX	1-Acetyloctahydro-3,5,7-trinitro-1,3,5,7-tetrazocine
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SOE	Salting-out Extraction
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TAT	2,4,6-Triaminotoluene
TATB	Triaminotrinitrobenzene
TAX	1-Acetylhexahydro-3,5-dinitro-1,3,5-triazine
TEA	Thermal Energy Analyzer
tetryl	Methyl-2,4,6-trinitrophenyl Nitramine
TNB	Trinitrobenzene
TNBA	Trinitrobenzaldehyde
TNT	Trinitrotoluene
TSP	Thermospray
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
2-Am-DNT	2-Amino-4,6-dinitrotoluene
2,2′-Az	4,4′,6,6′-Tetranitro-2,2′-azoxytoluene
2,4-Dam-NT	2,4-Diamino-6-nitrotoluene
2,6-Dam-NT	2,6-Diamino-4-nitrotoluene
3,5-DNA	3,5-Dinitroaniline
4-Am-DNT	4-Amino-2,6-dinitrotoluene
4,4′-Az	2,2′,6,6′-Tetranitro-4,4′-azoxytoluene

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Nitroaromatics, Environmental Analysis of • Nuclear Magnetic Resonance for Environmental Monitoring • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods • Sample Preparation for Environmental Analysis in Solids

(Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Gas Chromatography **(Volume 12)**
Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. T.F. Jenkins, M.E. Walsh, P.G. Thorne, 'Analytical Methods for Characterization of Explosives-contaminated Sites on US Army Installations', Proc. SPIE-Int. Soc. Opt. Eng. 2504 (Environmental Monitoring and Hazardous Waste-site Remediation), 342–349, 1995.
2. M.E. Walsh, T.F. Jenkins, P.G. Thorne, 'Laboratory and Analytical Methods for Explosives Residues in Soil', *J. Energetic Materials*, **13**, 357–383 (1995).
3. M.E. Walsh, T.F. Jenkins, P.S. Schnitker, J.W. Elwell, M.H. Stutz, 'Evaluation of SW846 Method 8330 for Characterization of Sites Contaminated with Residues of High Explosives', Hanover, NH, US Army Cold Regions Research and Engineering Laboratory, CRREL Report 93-5, 1993.
4. A.D. Beveridge, 'Development in the Detection and Identification of Explosive Residues', *Forensic Sci. Rev.*, **4**, 18–49 (1992).
5. T.F. Jenkins, M.E. Walsh, P.W. Schumacher, P.G. Thorne, 'Development of Colorimetric Field Screening Methods for Munitions Compounds in Soil', Proc. SPIE-Int. Soc. Opt. Eng. 2504 (Environmental Monitoring and Hazardous Waste Site Remediation), 324–333, 1995.
6. A.B. Crockett, T.F. Jenkins, H.D. Craig, W.E. Sisk, 'Overview of On-site Analytical Methods for Explosives in Soil', Hanover, NH, US Army Cold Regions Research and Engineering Laboratory. Special Report 98-4, 1988.
7. 'Framework for Action. Outcome of the Bioremediation of Explosives-contaminated Sites Working Meeting', Louisiana State University, Rice University, Georgia Tech Research Institute of Technology, Atlanta, GA, 29–30 March, 1995.
8. M.A. Major, W.H. Griest, J.C. Amos, W.G. Palmer, 'Toxicological Study No. 87-3012-95. Evidence for the Chemical Reduction and Binding of TNT During the Composting of Contaminated Soils. March 1995–January 1996', Aberdeen Proving Ground, MD, US Army Center for Health Promotion and Preventive Medicine, 1997.
9. T.F. Jenkins, D.C. Leggett, C.L. Grant, C.F. Bauer, 'Reversed-phase High-performance Liquid Chromatographic Determination of Nitroorganics in Munitions Wastewater', *Anal. Chem.*, **58**, 170–175 (1986).
10. T.F. Jenkins, P.H. Miyares, K.F. Meyers, E.F. McCormick, A.B. Strong, 'Comparison of Solid-phase Extraction with Salting-out Solvent Extraction for Preconcentration of Nitroaromatic and Nitramine Explosives from Water', *Anal. Chim. Acta.*, **289**, 69–78 (1994).
11. 'Ambient Water Quality Criteria for Dinitrotoluene', Washington, D.C., US Environmental Protection Agency, Office of Water Regulations and Standards, Criteria and Standards Division, EPA 440/5-80-045, 1980.
12. 'Trinitrotoluene Health Advisory', Washington, D.C., US Environmental Protection Agency, Criteria and Standards Division, Office of Drinking Water, 1989.
13. 'Health Advisory for HMX', Washington, D.C., US Environmental Protection Agency, Criteria and Standards Division, Office of Drinking Water, 1988.
14. 'Health Advisory for RDX', Washington, D.C., US Environmental Protection Agency, Criteria and Standards Division, Office of Drinking Water, 1988.
15. E.L. Etnier, 'Water Quality Criteria for 2,4-Dinitrotoluene and 2,6-Dinitrotoluene', Oak Ridge, TN, Oak Ridge National Laboratory, Report AD-ORNL-6312, 1987.
16. D.H. Rosenblatt, 'Contaminated Soil Cleanup Objectives for Cornhusker Army Ammunition Plant', Ft. Detrick, MD, US Army Bioengineering Research and Development Laboratory, Technical Report 8601, 1986.
17. F.W. DuBois, J.F. Bayton, 'Weathering of Explosives for Twenty Years', Los Alamos, NM, Los Alamos National Laboratory, LA-11931, 1991.
18. P.J. Rodacy, P.K. Leslie, S.D. Reber, S.E. Klassen, G.T. Cordes, 'The Development of Ion Mobility Spectroscopy for the Detection of Explosives in Environmental Samples', Albuquerque, NM, Sandia National Laboratories. SAND97-2215, 1997.
19. P.J. Rodacy, Sandia National Laboratory, Albuquerque, NM, private communication, July, 1999.
20. T.F. Jenkins, C.L. Grant, G.S. Brar, P.G. Thorne, P.W. Schumacher, T.A. Ranney, 'Sampling Error Associated with Collection and Analysis of Soil Samples at TNT-Contaminated Sites', *Field Anal. Chem. Technol.*, **1**, 151–163 (1997).

21. F.T. Kristoff, T.W. Ewing, D.E. Johnson, 'Testing to Determine Relationship Between Explosive Contaminated Sludge Components and Reactivity', Aberdeen Proving Ground, MD, US Army Toxic and Hazardous Materials Agency, Report No. AMXTH-TE-CR-86096, 1987.
22. A.B. Crockett, H.D. Craig, T.F. Jenkins, W.E. Sisk, 'Guidance for Characterizing Explosives Containing Soils: Sampling and Selecting On-site Analytical Methods', Proc. Twelfth Annual Waste Testing & Quality Assurance Symposium, Washington, D.C., 37–44, 1996.
23. T.F. Jenkins, C.L. Grant, G.S. Brar, P.G. Thorne, P.W. Schumacher, 'Sample Representativeness: A Necessary Element in Explosives Site Characterization', Proc. Twelfth Annual Waste Testing & Quality Assurance Symposium, Washington, D.C., 30–36, 1996.
24. Method 8515, 'Colorimetric Screening Method for Trinitrotoluene (TNT) in Soil', Revision 0, December 1996. http://www.epa.gov/SW-846/8515.pdf.
25. Methods Manual for Detection of PEP Compounds. USEPA Method 8330. Determination of Concentration of Nitroaromatics and Nitramines by High-performance Liquid Chromatography (HPLC). Revision 0, September 1994. http://www.crrel.usace.army.mil/techtransfer/products/pepmanual/Method8330.html. May also be found at http://www.epa.gov/SW-846/8330.pdf.
26. T.F. Jenkins, M.E. Walsh, P.G. Thorne, S. Thiboutot, G. Ampleman, T.A. Ranney, C.L. Grant, 'Assessment of Sampling Error Associated with Collection and Analysis of Soil Samples at a Firing Range Contaminated with HMX', Hanover, N.H. US Army Cold Regions Research and Engineering Laboratory. Special Report 97-22, 1997.
27. T.F. Jenkins, M.E. Walsh, P.G. Thorne, P.H. Miyares, T.A. Ranney, C.L. Grant, J.R. Esparza, 'Site Characterization for Explosives Contamination at a Military Firing Range Impact Area', Hanover, N.H. US Army Cold Regions Research and Engineering Laboratory. Special Report 98-9, 1998.
28. 'Nitroaromatics and Nitramines by HPLC. Second Update to SW846, Method 8330', Washington, D.C., US Environmental Protection Agency, Office of Solid Waste, Washington, D.C., 1994.
29. M.P. Maskarinec, C.K. Bayne, L.H. Johnson, S.K. Holladay, R.A. Jenkins, B.A. Tomkins, 'Stability of Explosives in Environmental Water and Soil Samples', ORNL/TM-11770, Oak Ridge National Laboratory, Oak Ridge, TN, 1991.
30. 1986 Annual Book of ASTM Standards, Volume 11.02 Water (II), Philadelphia, PA, American Society for Testing and Materials, 21–27, 1986.
31. H.S. Prentice, D.F. Bender, 'Project Summary: Development of Preservation Techniques and Establishment of Maximum Holding Times: Inorganic Constituents of the National Pollutant Discharge Elimination System and Safe Drinking Water Act, Research and Development', Washington, D.C., US Environmental Protection Agency, EPA/600/S4-86/043, March, 1987.
32. C.L. Grant, T.F. Jenkins, K.F. Myers, E.F. McCormick, 'Holding-time Estimates for Soils Containing Explosives Residues: Comparison of Fortification vs Field Contamination', *Environ. Tox. Chem.*, **14**, 1865–1874 (1995).
33. C.F. Bauer, S.M. Koza, T.F. Jenkins, 'Liquid Chromatographic Method for Determination of Explosives Residues in Soil: Collaborative Study', *J. AOAC*, **73**, 541–552 (1990).
34. D.C. Leggett, T.F. Jenkins, P.H. Miyares, 'Salting-out Solvent Extraction for Preconcentration of Neutral Polar Organic Solutes from Water', *Anal. Chem.*, **62**, 1355–1356 (1990).
35. A. Hubaux, G. Vos, 'Decision and Detection Limits for Linear Calibration Curves', *Anal. Chem.*, **42**, 840 (1970).
36. C.F. Bauer, C.L. Grant, T.F. Jenkins, 'Interlaboratory Evaluation of High-performance Liquid Chromatographic Determination of Nitroorganics in Munition Plant Wastewater', *Anal. Chem.*, **58**, 176–182 (1986).
37. T.F. Jenkins, C.F. Bauer, D.C. Leggett, C.L. Grant, 'Reverse Phase HPLC Method for Analysis of TNT, RDX, HMX, and 2,4-DNT in Munitions Wastewater', CRREL Report 84-29. Hanover, NH US Army Cold Regions Research and Engineering Laboratory, December, 1984.
38. M.E. Walsh, T.F. Jenkins, 'Liquid Chromatographic Separation of 2,4,6-Trinitrotoluene and its Principal Reduction Products', *Anal. Chim. Acta*, **231**, 313–315 (1990).
39. M. Emmrich, M. Kaiser, H. Rüden, S. Sollinger, 'Determination of RDX, 2,4,6-Trinitrotoluene, and Other Nitroaromatic Compounds by High-performance Liquid Chromatography with Photodiode-array Detection', *J. Chromatogr.*, **645**, 89–94 (1993).
40. J.E. Caton, W.H. Griest, 'Determination of Explosives and Some Metabolites of TNT in Biological and Environmental Samples by Liquid Chromatography on a Mixed-mode C_{18}-Anion Column', *J. Liq. Chrom. Rel. Technol.*, **19**, 661–677 (1996).
41. M.P. Maskarinec, D.L. Manning, R.W. Harvey, W.H. Griest, B.A. Tomkins, 'Determination of Munitions Components in Water by Resin Adsorption and High-performance Liquid Chromatography Electrochemical Detection', *J. Chromatogr.*, **302**, 51 (1984).
42. K. Bratin, P.T. Kissinger, C.S. Bruntlett, 'Reductive Mode Thin-layer Amperometric Detector for Liquid Chromatography', *J. Liq. Chromatogr.*, **4**, 1777–1795 (1981).
43. K. Bratin, P.T. Kissinger, R.C. Briner, C.S. Bruntlett, 'Determination of Nitroaromatic, Nitramine, and Nitrate Ester Explosive Compounds in Explosive Mixtures and Gunshot Residue by Liquid Chromatography and Reductive Electrochemical Detection', *Anal. Chim. Acta*, **130**, 295 (1981).
44. United States Army Toxic and Hazardous Materials Agency Quality Assurance Program, Second Edition.

Aberdeen Proving Ground, MD, US Army Toxic and Hazardous Materials Agency, 1987.
45. T.F. Jenkins, P.G. Thorne, K.F. Myers, E.F. McCormick, D.E. Parker, B.L. Escalon, 'Evaluation of Clean Solid Phases for Extraction of Nitroaromatics and Nitramines from Water', Hanover, NH, US Army Cold Regions Research & Engineering Laboratory. Special Report 95-22, 1995.
46. T.F. Jenkins, P.G. Thorne, K.F. Myers, E.F. McCormick, 'Evaluation of the New Clean Solid-phases for Extraction of Nitroaromatics and Nitramines from Water', Proc. Eleventh Annual Waste Testing & Quality Assurance Symposium, Washington, D.C., 128–142, 1995.
47. E.S.P. Bouvier, S.A. Oehrle, 'Analysis and Identification of Nitroaromatic and Nitramine Explosives in Water Using HPLC and Photodiode-array Detection', *LC–GC*, **13**, 120–127 (1995).
48. M.E. Walsh, T. Ranney, 'Determination of Nitroaromatic, Nitramine, and Nitrate Ester Explosives in Water Using Solid-phase Extraction and Gas Chromatography/Electron Capture Detection: Comparison with High-performance Liquid Chromatography', *J. Chromatogr. Sci.*, **36**, 406–416 (1998).
49. Appendix B to Part 136 – Definition and Procedure for the Determination of the Method Detection Limit – Revision 1.11, in Code of Federal Regulations: Protection of the Environment, 40, Parts 100 to 146, revised as of July 1, 1990.
50. J.C. Hoffsommer, J.M. Rosen, 'Analysis of Explosives in Seawater', *Bull. Environ. Contam. Toxicol.*, **7**, 177–181 (1972).
51. J.C. Hoffsommer, D.J. Glover, J.M. Rosen, 'Analysis of Explosives in Seawater and in Ocean Floor Sediment and Fauna', White Oak, Silver Spring, MD, Naval Ordnance Laboratory, Report Number NOLTR 72–215, 1972.
52. S.-A. Barshick, W.H. Griest, 'Trace Analysis of Explosives in Seawater Using Solid-phase Microextraction and Gas Chromatography/Ion-trap Mass Spectrometry', *Anal. Chem.*, **70**, 3015–3020 (1998).
53. J.E. Caton, S.A. Barshick, private communication, 1999.
54. M.R. Darrach, A. Chutjian, G.A. Plett, 'Trace Explosives Signatures from World War II Unexploded Undersea Ordnance', *Environ. Sci. Technol.*, **32**, 1354–1358 (1998).
55. P. Rodacy, P. Leslie, S. Reber, W. Chambers, 'Field Monitoring of Explosives in the Marine Environment', *J. Proc. Analyt. Chem.*, **4**, 137–141 (1998).
56. J. Yinon, 'Trace Analysis of Explosives in Water by Gas Chromatography–Mass Spectrometry with a Temperature-programmed Injector', *J. Chromatogr. A.*, **742**, 205–209 (1996).
57. T.F. Jenkins, M.E. Walsh, 'Instability of Tetryl to Soxhlet Extraction', *J. Chromatogr. A.*, **662**, 178–184 (1994).
58. J. Feltes, K. Levsen, D. Volmer, M. Spikermann, 'Gas Chromatographic and Mass Spectrometric Determination of Nitroaromatics in Water', *J. Chromatogr.*, **518**, 21–40 (1990).
59. H. Engelhardt, J. Meister, P. Kolla, 'Optimization of Post-column Reaction Detector for HPLC of Explosives', *Chromatographia*, **35**, 5–12 (1993).
60. T.F. Jenkins, C.L. Grant, 'Comparison of Extraction Techniques for Munitions Residues in Soil', *Anal. Chem.*, **59**, 1326–1331 (1987).
61. T.F. Jenkins, M.E. Walsh, 'Development of an Analytical Method for Explosive Residues in Soil', CRREL Report 87-7. Hanover, NH, US Army Cold Regions Research and Engineering Laboratory, 1987.
62. T.F. Jenkins, P.W. Schumacher, M.E. Walsh, C.F. Bauer, 'Development of an Analytical Method for Explosive Residues in Soil. Part II. Additional Development and Ruggedness Testing', CRREL Report 88-8. Hanover, NH, US Army Cold Regions Research and Engineering Laboratory, 1988.
63. C.F. Bauer, T.F. Jenkins, S.M. Koza, P.W. Schumacher, P.H. Miyares, M.E. Walsh, 'Development of an Analytical Method for Explosive Residues in Soil. Part III. Collaborative Test Results and Final Performance Evaluation', CRREL Report 89-9. Hanover, NH, US Army Cold Regions Research and Engineering Laboratory, 1989.
64. T.F. Jenkins, M.E. Walsh, P.W. Schumacher, P.H. Miyares, C.F. Bauer, C.L. Grant, 'Liquid Chromatographic Method for Determination of Extractable Nitroaromatic and Nitramine Residues in Soil', *J. AOAC*, **72**, 890–899 (1989).
65. C.A. Weisberg, M.L. Ellickson, 'Practical Modifications to USEPA Method 8330 for the Analysis of Explosives by HPLC', *Amer. Lab.*, **30**, 32N–32V (1998).
66. Federal Register. October 26, Rules and Regulations, 49(209), 43430–43431, 1984.
67. Extraction of Explosives from Soils by Accelerated Solvent Extraction (ASE™), Dionex Corporation, Sunnyvale, CA. Application Note 328, 1997.
68. W.H. Griest, C. Guzman, M. Dekker, 'Packed Column Supercritical Fluid Chromatographic Separation of Highly Explosive Compounds', *J. Chromatogr.*, **476**, 423 (1989).
69. P.H. Miyares, 'Supercritical Fluid Chromatography for the Analysis of Nitroaromatics, Nitramines, and Nitrate Esters', Hanover, NH, US Army Cold Regions Research and Engineering Laboratory, Report 92-21, 1992.
70. P.J. Rodacy, Sandia National Laboratory, Albuquerque, NM, private communication, July, 1999.
71. P.J. Rodacy, Sandia National Laboratory, Albuquerque, NM, private communication, July, 1999.
72. J. Avolio, Barringer Instruments, Inc., Warren, NJ, private communication, July, 1999.
73. T.F. Jenkins, M.E. Walsh, 'Determination of Ultra-trace Levels of Nitroaromatic Explosives and Manufacturing

Impurities in Air and Soil Using Gas Chromatography/Electron Capture Detection', Presented to PITTCON99™, Orange County Convention Center, Orlando, FL. Paper number 375, 1999.
74. J. Yinon, R.A. Yost, S. Bulusu, 'Thermal Decomposition Characterization of Explosives by Pyrolysis–Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A.*, **688**, 231–242 (1994).
75. J.V. Janowsky, 'Ueber eine Reaction der Dinitrokörper', *Berichte der Deutschen Chemischen Gesellschaft*, **24**, 971–972 (1891).
76. J. von Meisenheimer, 'Ueber Reactionen Aromatischer Nitrokörper', *Leibig's Annalen der Chemie*, **323**, 205 (1902).
77. C.L. Jackson, R.B. Earle, 'On Certain Colored Substances Derived from Nitro Compounds', *Am. Chem. J.*, **29**, 89–120 (1903).
78. R.W. Bost, F. Nicholson, 'Color Test for the Identification of Mono-, Di-, and Trinitro Compounds', *Ind. Eng. Chem., Anal. Ed.*, **7**, 190–191 (1935).
79. J. Yinon, S. Zitrin, *The Analysis of Explosives*, Pergamon Press, Oxford, 1981.
80. P. Griess, 'Bermerkungen zu der Abhandlung der HH Weselsky und Benedikt ''Ueber einige Azoverbindungen'' ', *Chem. Ber.*, **12**, 426–428 (1879).
81. T.F. Jenkins, M.E. Walsh, 'Development of Field Screening Methods for TNT, 2,4-DNT, and RDX in Soil', *Talanta*, **39**, 419–428 (1992).
82. R.E. Wyant, Technical Report TR-185, Indian Head, MD, Naval Explosive Ordnance Disposal Facility, 1977.
83. S. Abramovich-Bar, Y. Bamberger, M. Ravreby, S. Levy, 'Applications of Ion Chromatography for Determination and Identification of Chlorate, Nitrite, and Nitrate in Explosives Residues', in J. Yinon, ed., *Advances in Analysis and Detection of Explosives*; Proc. Fourth Intern. Symp. Analysis and Detection of Explosives, Jerusalem, Israel, Kluwer Academic Publishers, 41–54, 1993.
84. R.T. Medary, 'Inexpensive, Rapid Field Screening Test for 2,4,6-Trinitrotoluene in Soil', *Anal. Chim. Acta*, **258**, 341–346, 1992.
85. D.D. Fetterolf, 'Antibody-based Field Test Kits for Explosives', in J. Yinon, ed., *Advances in Analysis and Detection of Explosives*; Proc. Fourth Intern. Symp. on Analysis and Detection of Explosives. Jerusalem, Israel, Kluwer Academic Publishers, 19–27, 1993.
86. J.C. Bart, L.L. Judd, A.W. Kusterbeck, 'Environmental Immunoassay for the Explosive RDX Using a Fluorescent Dye-labeled Antigen and the Continuous-flow Immunosensor', Sensors and Actuators B, 38-39, 411–418, 1997.
87. A.W. Kusterbeck, P.T. Charles, 'Field Demonstration of a Portable Flow Immunosensor', *Field Analyt. Chem. Technol.*, **2**, 341–350 (1998).
88. A.B. Crockett, H.D. Craig, T.F. Jenkins, 'On-site Analytical Methods for Explosives in Soils', *Amer. Environ. Labor.*, **9**, 27–29 (1997).
89. A.B. Crockett, H.D. Craig, T.F. Jenkins, 'Field Sampling and Selecting On-site Analytical Methods for Explosives in Water. EPA Federal Facilities Forum Issue', Washington, D.C., US Environmental Protection Agency, Office of Research and Development, Office of Solid Waste and Emergency Response Report, EPA/600/S-99/002, 1999.
90. 'Field Screening Technologies Umatilla Explosive Washout Lagoon Soils', Prepared by Black & Veatch Waste Science, Inc., Tacoma, WA, for US Environmental Protection Agency Region 10, Seattle, WA, draft, 1995.
91. K.F. Myers, E.F. McCormick, A.B. Strong, P.G. Thorne, T.F. Jenkins, 'Comparison of Commercial Colorimetric and Enzyme Immunoassay Field Screening Methods for TNT in Soil', TR-IRRP-94-4. Vicksburg, MS, US Army Corps of Engineers – Waterways Experiment Station, 1994.
92. R.A. Haas, B.P. Simmons, 'Measurement of TNT and RDX in Soil by Enzyme Immunoassays and High Performance Liquid Chromatography (EPA Method 8330)', California EPA, Hazardous Materials Laboratory, 1995.
93. Black & Veatch Waste Science, 'Comparison of Field Screening Technologies Implemented During Phase 1 Remediation of Explosives Washout Lagoon Soils. Final Report for US Environmental Protection Agency Region 10', Seattle WA, 1995.
94. H.D. Craig, G.K. Ferguson, A.G. Markos, P.G. Thorne, T.F. Jenkins, C.A. Witt-Smith, 'Evaluation of On-site Analytical Methods for Explosives in Compost Residues', in G.B. Wickramanayake, R.B. Hinchee, eds., Risk, Resour. Regul. Issues, Int. Conf. Rem. Chlorinated Recalcitrant Comp., 1st, Columbus, OH, Battelle Press, 13–18, 1998.
95. P.G. Thorne, T.F. Jenkins, 'A Field Method for Quantifying Ammonium Picrate and Picric Acid in Soil', *Field Anal. Chem. Technol.*, **1**, 165–170 (1997).
96. A.B. Crockett, D.A. Atkinson, T.F. Jenkins, 'Field Screening of Soils Contaminated with Explosives Using Ion Mobility Spectrometry', Proc. Twelfth Annual Waste Testing & Quality Assurance Symposium, Washington, D.C., 46, 1996.
97. W.H. Griest, A.J. Stewart, A.A. Vass, C.-h. Ho, 'Chemical and Toxicological Characterization of Slurry Reactor Biotreatment of Explosives-contaminated Soil', ORNL/TM-13384, Oak Ridge National Laboratory, Oak Ridge, TN (March, 1997, publ. Sept., 1998).
98. J.M. Brannon, C.B. Price, C.A. Hayes, 'Abiotic and Biotic TNT Transformations', Vicksburg, MS, US Army Engineering Waterways Experiment Station, Misc. Paper IRRP-97-3, 1997.
99. P.G. Thorne, D.C. Leggett, 'Hydrolytic Release of Bound Residues from Composted Soil Contaminated with 2,4,6-Trinitrotoluene', *Environ. Toxicol. Chem.*, **16**, 1132–1134 (1997).
100. F. Ahmad, D.J. Roberts, 'Use of Narrow-bore High-performance Liquid Chromatography–Diode Array Detection for the Analysis of Intermediates of the

Biological Degradation of 2,4,6-Trinitrotoluene', *J. Chromatogr. A*, **693**, 167–175 (1995).
101. D.J. Glover, J.C. Hoffsommer, D.A. Kubose, 'Analysis of Mixtures of 2-Amino-4,6-Dinitrotoluene, 4-Amino-2,6-Dinitrotoluene, and 2,4-Diamino-6-Nitrotoluene, and 2,6-Diamino-4-Nitrotoluene', *Anal. Chim. Acta*, **88**, 381–384 (1977).
102. J. Yinon, D.-G. Hwang, 'Detection of TNT and its Metabolites in Body Fluids of Laboratory Animals and in Occupational Exposed Humans', *J. Energetic Mater.*, **4**, 305–313 (1986).
103. J. Carlson, P. Vouros, D. Kaplan, 'The Detection of TNT and TNT Metabolites by Electrospray Mass Spectrometry', Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics, 193, 1995.
104. E.C. Huang, B.J. Jackson, K.E. Markides, M.L. Lee, 'Application of Capillary Supercritical Fluid Chromatography – Double Focusing Mass Spectrometry Under Negative Ion Chemical Ionization Conditions', *J. Microcol. Sep.*, **2**, 88–89 (1990).
105. P. Mußmann, R. Eisert, K. Levsen, G. Wünsch, 'Determination of Nitrophenols, Diaminotoluenes, and Chloroaromatics in Ammunition Wastewater', *Acta Hydrochim. Hydrobiol.*, **23**, 13–19 (1995).
106. K. Schoene, H.J. Bruckert, J. Steinhanses, A. König, 'Two-stage Derivatization with *N*-(*tert*-butyldimethylsilyl)-*N*-methyl-trifluoroacetamide (MTBSTFA) and *N*-methyl-bis-(trifluoroacetamide) (MBTFA) for the Gaschromatographic Analysis of OH-, SH-, and NH-Compounds', *Fresenius' J. Anal. Chem.*, **348**, 364–370 (1994).
107. T.F. Jenkins, M.E. Walsh, 'An Overview of Explosives Contamination at DoD Facilities: An Analytical Perspective', Proc., Ann. Meet. – Air Waste Manage. Assoc. (88th), 12, 1995.
108. J.E. Caton, W.H. Griest, D.P. Schrum, personal communications, 1999.
109. J. Hawari, A. Halasz, L. Paquet, E. Zhou, B. Spencer, G. Ampleman, S. Thiboutot, 'Characterization of Metabolites in the Biotransformation of 2,4,6-Trinitrotoluene with Anaerobic Sludge: Role of Triaminotoluene', *Appl. Environ. Microbiol.*, **64**, 2200–2206 (1998).
110. D.D. Fetterolf, 'Detection and Identification of Explosives by Mass Spectrometry', in J. Yinon, ed., Forensic Applications of Mass Spectrometry. Boca Raton, FL, CRC Press, 215–261, 1995.
111. J.P. Davies, L.G. Blackwood, S.G. Davis, L.D. Goodrich, R.A. Larson, 'Design and Calibration of Pulsed Vapor Generators for TNT, RDX, and PETN', in J. Yinon, ed., *Advances in Analysis and Detection of Explosives*; Proc. Fourth Intern. Symp. Analysis and Detection of Explosives. Jerusalem, Israel, Kluwer Academic Publishers, 511–532, 1993.
112. R.A. Strobel, R. Noll, C.R. Midkiff, Jr, 'The ATF Canine Explosives Detection Program', in J. Yinon, ed., *Advances in Analysis and Detection of Explosives*; Proc. Fourth Intern. Symp. Analysis and Detection of Explosives. Jerusalem, Israel, Kluwer Academic Publishers, 455–461, 1993.
113. A.M. Rouhi, 'Detecting Illegal Substances', *Chem. Eng. News*, **75**, 24–27 (1997).
114. S.A. McLuckey, D.E. Goeringer, K.G. Asano, G. Vaidyanathan, J.L. Stephenson, Jr, 'High Explosives Vapor Detection by Glow Discharge–Ion-trap Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 287–298 (1996).
115. S.A. McLuckey, G.L. Glish, K.G. Asano, 'Coupling of an Atmospheric-sampling Ion Source with an Ion-trap Mass Spectrometer', *Anal. Chim. Acta*, **225**, 25–35 (1989).
116. D.E. Goeringer, K.G. Asano, S.A. McLuckey, D. Hoekman, S.W. Stiller, 'Filtered Noise Field Signals for Mass-selective Accumulation of Externally Formed Ions in a Quadrupole Ion Trap', *Anal. Chem.*, **66**, 313–318 (1994).
117. K.G. Asano, D.E. Goeringer, S.A. McLuckey, 'Parallel Monitoring for Multiple Targeted Compounds by Ion Trap Mass Spectrometry', *Anal. Chem.*, **67**, 2739–2742 (1995).
118. S.A. McLuckey, D.E. Goeringer, G.L. Glish, 'Collisional Activation with Random Noise in Ion Trap Mass Spectrometry', *Anal. Chem.*, **64**, 1455–1460 (1992).
119. D.D. Fetterolf, 'Detection of Trace Explosive Evidence by Ion Mobility Spectrometry', in J. Yinon, ed., *Advances in Analysis and Detection of Explosives*; Proc. Fourth Intern. Symp. Analysis and Detection of Explosives. Jerusalem, Israel, Kluwer Academic Publishers, 117–131, 1993.
120. R.A. Gieray, P.T.A. Reilly, M. Yang, W.B. Whitten, J.M. Ramsey, 'Tandem Mass Spectrometry of Uranium and Uranium Oxides in Airborne Particulates', *Anal. Chem.*, **70**, 117–120 (1998).
121. P.T.A. Reilly, R.A. Gieray, W.B. Whitten, J.M. Ramsey, 'Real-time Characterization of the Organic Composition and Size of Individual Diesel Engine Smoke Particles', *Environ. Sci. Technol.*, **32**, 2672–2679 (1998).
122. W.B. Whitten, P.T.A. Reilly, A. Lazar, J.M. Ramsey, 'Real-time Mass Spectrometry of Individual Airborne and Waterborne Explosives Particles', Final Report FY1998 to the US Federal Aviation Administration, 1998.
123. M.E. Homan, J.-F.A. Borny, 'The Analysis of Single and Double Based Propellants by HPLC–CLND', Presented at PITTCON '98. Summary provided by Antek® Instruments, Inc., Houston, TX.
124. P. Kolla, A. Sprunkel, 'Identification of Dynamite Explosives in Post Explosion Residues', *J. Forensic Sci.*, **40**, 406–411 (1995).
125. P. Kolla, 'Gas Chromatography, Liquid Chromatography, and Ion Chromatography Adapted to the Trace

Analysis of Explosives', *J. Chromatogr. A.*, **674**, 309–318 (1994).
126. P. Kolla, 'Trace Analysis of Explosives from Complex Mixtures with Sample Pretreatment and Selective Detection', *J. Forensic Sci.*, **36**, 1342–1359 (1991).
127. J.M.F. Douse, 'Trace Analysis of Explosives at the Low Nanogram Level in Handswab Extracts Using Columns of Amberlite XAD-7 Porous Polymer Beads and Silica Capillary Column Gas Chromatography with Thermal Energy Analysis and Electron-capture Detection', *J. Chromatogr.*, **328**, 155–165 (1985).
128. J.M.F. Douse, 'Trace Analysis of Explosives at the Low Picogram Level Using Silica Capillary Column Gas Chromatography with Thermal Energy Analyser Detection', *J. Chromatogr.*, **256**, 359–362 (1983).
129. J.M.F. Douse, 'Dynamic Headspace Method for the Improved Clean-up of Gunshot Residues Prior to the Detection of Nitroglycerine by Capillary Column Gas Chromatography with Thermal Energy Analysis Detection', *J. Chromatogr.*, **464**, 178–185 (1989).
130. J.M.F. Douse, 'Trace Analysis of Explosives by Capillary Supercritical Fluid Chromatography with Thermal Energy Analysis Detection', *J. Chromatogr.*, **445**, 244–250 (1988).
131. P. Kolla, H. Engelhardt, J. Zapp, 'Sample Preparation by Supercritical Fluid Extraction in Explosives Trace Analysis', in J. Yinon, ed., *Advances in Analysis and Detection of Explosives*; Proc. Fourth Intern. Symp. Analysis and Detection of Explosives, Jerusalem, Israel, Kluwer Academic Publishers, 55–65, 1993.
132. C. Bailey, C. Yan, 'Separation of Explosives Using Capillary Electrochromatography', *Anal. Chem.*, **70**, 3275–3279 (1998).
133. J. Yinon, 'Direct Exposure Chemical Ionization Mass Spectra of Explosives', *Organ. Mass Spectrom.*, **15**, 637–639 (1980).
134. J. Yinon, 'Analysis of Explosives by Negative Ion Chemical Ionization Mass Spectrometry', *J. Forensic Sci.*, **25**, 401–7 (1980).
135. J. Yinon, 'Identification of Explosives Mixtures by Tandem Mass Spectrometry (MS/MS)', *Can. Soc. Forens. Sci. J.*, **21**, 46–53 (1988).
136. C.E. Parker, R.D. Voyksner, Y. Tondeur, J.D. Henion, D.J. Harvan, J.R. Hass, J. Yinon, 'Analysis of Explosives by Liquid Chromatography/Negative Ion Chemical Ionization Mass Spectrometry', *J. Forens. Sci.*, **27**, 495–505 (1982).
137. J. Yinon, 'Mass Spectrometry of Explosives: Nitro Compounds, Nitrate Esters, and Nitramines', *Mass Spectrometry Reviews*, **1**, 257–307 (1982).
138. J. Yinon, 'Analysis of Explosives by LC/MS', Proc. Int. Symp. Anal. Detect. Explos. Washington, D.C., Fed. Bur. Invest, 227–34, 1983.
139. J. Yinon, M. Laschever, 'Direct-injection Chemical Ionization Mass Spectrometry of Explosives in Water', *Eur. J. Mass Spectrom. Biochem., Med. Environ. Res.*, **2**, 101–3 (1982).
140. R.D. Voyksner, J. Yinon, 'Trace Analysis of Explosives by Thermospray High-performance Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **354**, 393–405 (1986).
141. J. Yinon, J.E. McClellan, R.A. Yost, 'Electrospray Ionization Tandem Mass Spectrometry Collision-induced Dissociation Study of Explosives in an Ion-trap Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **11**, 1961–1970 (1997).
142. I. Tamiri, S. Zitrin, 'Capillary Column Gas Chromatography/Mass Spectrometry of Explosives', *J. Energetic Materials*, **4**, 215–237 (1986).
143. J.D. King, 'Explosives Detection–The Case for Magnetic Resonance', in *Advances in Analysis and Detection of Explosives*, ed. J. Yinon, Proc. Fourth Intern. Symp. Analysis and Detection of Explosives, Jerusalem, Israel, Kluwer Academic Publishers, 351–359, 1993.
144. M.L. Buess, A.N. Garroway, J.B. Miller, J.P. Yesinowski, 'Explosives Detection by ^{14}N Pure NQR+', in *Advances in Analysis and Detection of Explosives*; ed. J. Yinon, Proc. Fourth Intern. Symp. Analysis and Detection of Explosives, Jerusalem, Israel, Kluwer Academic Publishers, 361–368, 1993.

Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis

Michael Sperling
PerkinElmer Bodenseewerk GmbH, Überlingen, Germany

1 Introduction **1**
2 Atomic Absorption Techniques and Their Capabilities for Environmental Analysis **2**
2.1 Flame Atomic Absorption Spectrometry 2
2.2 Graphite Furnace Atomic Absorption Spectrometry 3
2.3 Other Atomic Absorption Spectrometry Techniques 4
3 Method Development and Validation **5**
3.1 The Analytical Task 5
3.2 Sampling and Sample Pretreatment 6
3.3 Atomic Absorption Analysis and Methods 7
3.4 Quality Assurance 12
4 Environmental Regulations and Methods **13**
4.1 Water Analysis 14
4.2 Soils, Sediments, Sludges, Coal, Dust and Other Solid Materials 22
4.3 Biological Materials 29
Abbreviations and Acronyms **36**
Related Articles **37**
References **38**

Atomic absorption spectrometry (AAS) is one of the most often used techniques for the quantitative determination of elements in environmental materials at trace and ultratrace levels. AAS is an optical atomic spectrometric technique based on the measurement of the specific absorption originating from free nonionized atoms in the gas phase. To transfer the analyte to free atoms, different types of atomizer are in use, the flame and the graphite furnace types being the most often used. Typical detection limits of flame atomic absorption spectrometry (FAAS) are of the order of 1–100 μg L^{-1}, making it a perfect tool for the determination of minor and trace elements, at least for contaminated samples. Graphite furnace atomic absorption spectrometry (GFAAS), offering detection limits which are about a factor of 20–200 lower than for FAAS, is the standard method for many trace elements, especially for background values, and for unpolluted samples, such as fresh water and biological materials. AAS in its conventional configuration is a single-element technique, which has to be used in a sequential mode when more than one element has to be determined. However, there are commercial instruments available that can be used for the determination of 6–8 elements simultaneously.

1 INTRODUCTION

AAS is one of the most often used techniques for the quantitative determination of elements (up to 60–70 analytes) in environmental materials at trace and ultratrace levels down to nanograms (ng, 10^{-9} g) and picograms (pg, 10^{-12} g). It is a very well established, reliable and cost-effective analytical tool in thousands of laboratories throughout the world. Its high degree of maturity is reflected in the declining interest of researchers in fundamental research devoted to AAS on one side, while on the other maintaining intensive application of this technique in various areas of analytical chemistry, which is traceable through about 1000 publications per year.

AAS is an optical atomic spectrometric technique based on the measurement of the specific absorption originating from free nonionized atoms in the gas phase. To transfer the analyte to free atoms and provide the hostile environment for these atoms at least for a short moment, different types of atomizer are in use. A scheme of the principle components of an AAS instrument is given in Figure 1.

In FAAS, flames with different fuel–oxidant combinations are used for atomization, whereas in electrothermal atomic absorption spectrometry (ETAAS) different types of electrothermal atomizers are used, such as the graphite tube furnace in GFAAS or the quartz tube atomizer (QTA) in hydride generation atomic absorption spectrometry (HGAAS). These atomizers are meant to provide just the necessary energy via a chain of reactions to transfer the sample into atoms. Most of these atoms are in their electronic ground state with energy, E_0, and are able to absorb energy ($\Delta E = E_1 - E_0 = h\nu = hc/\lambda$) from a radiation source (lamp), where E_1 is the energy of the (first) excited state, h is Planck's constant, c is the velocity of light and ν and λ are the frequency and the wavelength of the atomic spectral line, respectively. The probing beam from the radiation source passes through the observation volume of the atomizer, wherein the radiation is absorbed by the cloud of atoms, which are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

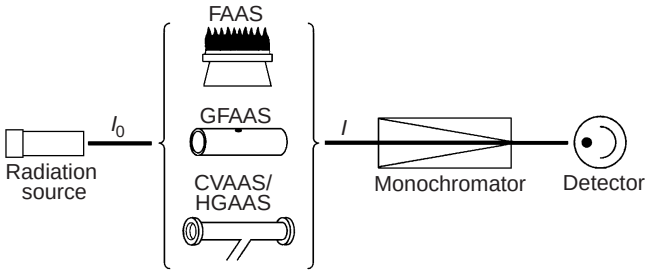

Figure 1 Schematic view of atomic absorption (AA) spectrometric techniques. CVAAS, cold vapor atomic absorption spectrometry.

excited to one of their permitted upper energy levels (E_1, E_2, etc.). According to Beer's law, the radiation absorbed is proportional to the number of absorbers in the absorption volume. The ratio of the incident radiant power, I_0, and the power transmitted through the absorption volume, I, is then determined by the detection system of the spectrometer and processed to yield the analyte signal, absorbance, A [Equations 1 and 2]:

$$A = \log_{10} \frac{I_0}{I} = -\log_{10} T \qquad (1)$$

$$A = abc \qquad \text{(Beer's law)} \qquad (2)$$

where T is the transmittance, a is the absorptivity, b is the thickness of the absorption volume and c is the concentration of the absorbing analyte atoms. The relationship between the concentration of the analyte in the sample, c_{sample} and the concentration of the analyte in the absorption volume, c, is normally established via a calibration graph [Equation 3]:

$$c_{\text{sample}} = f(A) \qquad (3)$$

The very high selectivity of AAS originates from the simplicity of the absorption spectra, where only a limited number of lines limits the chance of cases of spectral interference, and the use of line sources that emit just the radiation specific to the analyte, such as the hollow-cathode lamp (HCL) or the electrodeless discharge lamp (EDL). Multielement HCLs are available, but should be used with care, because of the reduction in selectivity due to the possibility of spectral interferences.

The role of the spectrometer in this configuration is limited to the separation of the analytical line from other lines emitted by the radiation source and from emission originating from other sources than the lamp, e.g. the atomizer itself and from emitting species inside the absorption volume. For this reason, the radiation of the lamp is modulated electronically or mechanically with a certain frequency of 50–300 Hz and the detection system is synchronized with this modulation. In order to distinguish the specific AA from other nonspecific absorption due to, for example, molecular absorption or scattering, the AA spectrometer in general uses a background correction system. Commercial instruments are equipped with at least one of the following systems: (a) deuterium background corrector, (b) halogen lamp background corrector, (c) Smith–Hieftje background corrector or (d) Zeeman-effect background corrector, the last one being the most accurate and powerful technique.[1]

2 ATOMIC ABSORPTION TECHNIQUES AND THEIR CAPABILITIES FOR ENVIRONMENTAL ANALYSIS

2.1 Flame Atomic Absorption Spectrometry

Flames with different fuel–oxidant combinations are used for atomization in FAAS, the air–acetylene and the nitrous oxide–acetylene flames being the most popular. The air–acetylene flame generates a flame temperature of about 2525 K, is transparent over a wide spectral range and only starts to absorb radiation below 230 nm. Also, the emission of the air–acetylene flame is very low, so ideal conditions are given for most elements. Normally, this flame is operated stoichiometrically or slightly oxidizing, but it can be operated also in a reducing mode. However, in the reducing mode the lower partial pressure of oxygen is responsible for a reduced flame temperature, enhancing the risk of interference problems. Interference through ionization occurs only in a few cases, for example, for alkali metals, and can be easily avoided by the addition of an easily ionizable element as an ionization buffer, such as sodium, potassium or cesium. There are about 30 elements that in effect cannot be determined with the air–acetylene flame since they form refractory oxides, such as aluminum and silicon. Such elements require a hotter flame in order to break the element–oxygen bond, such as the nitrous oxide–acetylene flame; this flame is operated with a slight excess of fuel, providing reducing conditions and a flame temperature of about 2975 K. Both flames can be safely operated, and modern instruments contain sensors and interlocks for fully automated operation.

Samples are introduced into flames normally in the liquid form by a pneumatic nebulizer and mixed in a spray chamber with the flame gases before these reach the head of the (premixed) burner, where they are ignited. The burner has a single slot of appropriate length providing a certain depth for the absorption volume (comparable to a cuvette length), that is essential for the sensitivity of the method. Absorption of light coming from a primary light source is directed through this absorption volume and is measured at an observation height in the flame

Table 1 Comparison of relative instrumental detection limits ($\mu g\,L^{-1}$) for FAAS and GFAAS (after Welz and Sperling[2])

Element	FAAS	GFAAS	Element	FAAS	GFAAS	Element	FAAS	GFAAS
Ag	1.5	0.02	Ho	60		Ru	100	1.5
Al	45	0.1	In	30		Sb	45	0.15
As	150	0.2	Ir	900	3	Sc	30	
Au	9	0.15	K	3	0.008	Se	100	0.25
B	1000	20	La	3000		Si	90	1
Ba	15	0.35	Li	0.8	0.06	Sm	3000	
Be	1.5	0.01	Lu	1000		Sn	150	0.2
Bi	30	0.25	Mg	0.15	0.004	Sr	3	0.025
Ca	1.5	0.01	Mn	1.5	0.01	Ta	1500	
Cd	0.8	0.003	Mo	45	0.03	Tb	900	
Co	9	0.15	Na	0.3	0.02	Te	30	0.4
Cr	3	0.01	Nb	1500		Ti	75	0.35
Cs	15		Nd	1500		Tl	15	0.25
Cu	1.5	0.1	Ni	6	0.3	Tm	15	
Dy	50		Os	120		U	15 000	
Er	60		P	75 000	130	V	60	0.06
Eu	30		Pb	15	0.1	W	1500	
Fe	5	0.1	Pd	30	0.8	Y	75	
Ga	75		Pr	7500		Yb	8	
Gd	1800		Pt	60	2	Zn	1.5	0.1
Ge	300		Rb	3	0.04	Zr	450	
Hf	300		Re	750				
Hg	300	1.0	Rh	6				

some millimeters above the burner exit, where the atom concentration is at its maximum. Operational parameters to control flame atomization are very few (e.g. the flame stoichiometry, the nebulizer uptake and the observation height). Typical detection limits of FAAS are of the order of $1–100\,\mu g\,L^{-1}$ (see Table 1) and are similar to those obtained by radially viewed inductively coupled plasma optical emission spectrometry (ICPOES). The precision of measurements can be better than 0.3% using optimum conditions.

FAAS has found wide application in environmental analysis and is often still used as standard method. However, owing to its lack in sensitivity, its application is limited to the determination of the main and minor elements and some trace elements, especially for contaminated samples (sewage sludges, wastewater, sediments, etc.).

2.2 Graphite Furnace Atomic Absorption Spectrometry

GFAAS offers detection limits which are about a factor of 20–200 lower than for FAAS (see Table 1). This significant improvement is the result of the improved residence time of the atoms in the absorption volume, which is mainly dictated by diffusional losses from the tube. In principle, the much longer residence time should also lead to greater freedom from interferences, provided that atomization takes place under isothermal conditions. To come as close as possible to the ideal state of isothermal atomization, different efforts have been made, of which platform atomization is the most prominent. Platform atomization together with the use of a powerful Zeeman effect background correction (ZBC) system, fast signal evaluation and the application of matrix (chemical) modification, forms the basis of the stabilized temperature platform furnace (STPF) concept.[3] A step even further in this direction is the transversely heated graphite atomizer (THGA), which heats uniformly over the whole tube length avoiding any cold spots,[4] which are observed in longitudinally heated furnaces, owing to the necessity for cooling the tube ends for reasons of reproducible electrical contact.[5]

The disadvantage of GFAAS compared with FAAS is the lower sample throughput. In a graphite furnace system the measurement cycle starts with the injection of liquid samples (digest, extract, suspension, standard solution) of about $5–50\,\mu L$ into the graphite tube or into the cavity of a platform. Thereafter a temperature program is initiated, including drying, pyrolysis (thermal pretreatment for the decomposition of organic matrix or the volatilization of other matrix components), atomization (with temperatures up to 2800 °C), and an additional cleaning step to remove remaining matrix components from the furnace. The whole cycle takes typically about 50–100 s (i.e. about 10 times a regular FAAS measurement) and must be optimized for a particular analyte–matrix combination. During the steps of this cycle the furnace is flushed with an inert purge gas stream (argon is recommended) in order to protect hot graphite

parts against oxidation and to remove volatile reaction products. Other alternative gases such as air, oxygen and hydrogen can be added during the first steps, in order to facilitate thermal pretreatment. GFAAS is called a mature technique, not only with respect to instrumentation, but also with respect to methodology. Using STPF conditions together with integrated absorbance for signal evaluation, it can be shown that the characteristic mass (i.e. the mass of element in picograms that produces an integrated signal equivalent to 0.0004 A s) for various elements in different matrices differs by only around ±15% between different instruments,[6] offering a way towards standardless analysis (sometimes even called 'absolute analysis'). At least such possibilities can be used for semiquantitative analysis for screening purposes and for system diagnosis. With respect to environmental analysis, GFAAS is the standard method for many trace elements, especially for background values, and for unpolluted samples, such as fresh water and biological materials.

2.3 Other Atomic Absorption Spectrometry Techniques

Since neither FAAS nor GFAAS offers a particular detection power for mercury and hydride-forming metalloids such as antimony, arsenic, selenium, and tin, special vapor generation techniques to measure these elements after volatilization in the gas phase were developed and are commercially available. Mercury is commonly reduced, after digestion of solid samples or directly in liquids, to elemental mercury by the addition of stannous chloride or sodium tetrahydroborate and analyzed at ambient or slightly increased temperature by the cold vapor technique with detection limits at or below the micrograms per liter level. In a similar manner, the hydride-forming elements are measured by the hydride generation technique, in which the analyte is volatilized by reduction with sodium tetrahydroborate and analyzed by atomization in a heated QTA or after in situ trapping in the graphite furnace. These different vapor generation techniques often provide detection limits improved by a factor of about 10 in comparison with GFAAS of liquid samples. Disadvantages of vapor generation techniques are the additional sources of interference during the process of vapor generation and atomization, which for the case of QTAs takes place at temperatures limited to about 1000 °C. In order to reduce the problems with vapor generation, special attention has to be paid to sample preparation, for which the requirements are the most stringent in comparison with other AAS techniques. An extensive discussion of the hydride generation technique can be found in a monograph by Dedina and Tsalev[7] and a detailed discussion of both vapor generation techniques can be found in a monograph devoted to AAS by Welz and Sperling.[2]

The call for automation of the vapor generation techniques was one of the driving forces behind the development of using flow injection (FI) as a sample managing system for AAS.[8] The basic principle of using FI for the transport of sample solution for AAS is shown in Figure 2.

Nowadays the coupling of FI has been realized with all AAS techniques, including the GFAAS technique, with great benefits for AAS in many respects. FI can be used just for sample introduction in FAAS, improving the tolerance for high-salt matrices, but also can be used for sample dilution, calibration, separation, preconcentration and speciation in combination with all AAS techniques. The main advantages of FI for AAS can be summarized as its high potential for automation and its high repeatability and reproducibility of all the process. Further, FI provides a closed system, which reduces external contamination to very low levels and is operated under thermodynamically nonequilibrated conditions, which allow for high sampling rates. The ease of automation comes from the generally simple system, which normally consists of one or two independent multichannel peristaltic pumps, a multichannel injection valve, a manifold made from polymeric tubing, flow conjunctions, mixers, reactors, separators and columns, with all the active parts under computer control. Commercial systems are available and integration of FI into the AAS software has been realized.[9] Environmental analysis can greatly benefit from the coupling of FI with AAS in all cases where the automation reduces costly manpower or the sensitivity or selectivity of AAS alone is not sufficient for the analytical task. Examples of such applications will be briefly discussed in section 4.

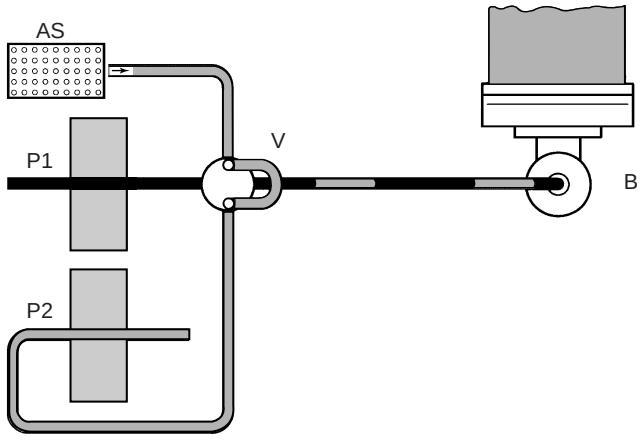

Figure 2 Simplest principle using FI for the transport of measurement solution for FAAS. P1 = pump for carrier solution; P2 = pump for measurement solution; AS = autosampler; V = injection valve; B = AAS burner. (Reproduced by kind permission of Perkin-Elmer.)

3 METHOD DEVELOPMENT AND VALIDATION

3.1 The Analytical Task

Environmental chemical analysis very often is the basis of other environmental studies, such as risk assessments, or court actions, such as the withdrawal of allowances, involving decisions which have an enormous economic impact. In order for analytical results to provide the information necessary to support such decisions, confidence in the quality of the analysis must be sufficiently high. Those parameters that will influence the decisions, the data quality objectives (DQO) should be defined, in order to fit the analysis for the purpose. Two basic parameters should be considered when discussing analytical results: *accuracy* (closeness of the agreement between the result of a measurement and a true value of the measurable quantity of an analyte) and *uncertainty* (expressed as the confidence of variation or the confidence interval) due to random errors and random variations in the procedure.[10] While accuracy is of primary importance, an unacceptably high uncertainty renders the result useless. Apart from the requirements on the quality of the results, the whole analytical procedure must be designed to answer the questions of the analytical task in an economical way (i.e. it should be fit for purpose). Errors can occur in every step of an analysis and negatively influence its accuracy to a degree corresponding to their share of the entire method. In order to reduce such errors, it is of major importance that all steps of an analytical method are documented and that special attention is paid to steps that are particularly critical. Table 2 presents a general scheme of the analytical process for trace metal analysis, where the actual analytical procedure covers steps III–XI. When comparing the estimation of possible errors in each step, it becomes obvious that the most often discussed step, 'instrumental determination', is normally the most accurate and precise step of the analytical procedure, while the most critical step, the sampling, often is not even under the control of the analyst. It cannot be overemphasized that the analyst should have all the steps of a method under control since only then can possible problems be recognized, evaluated and, if necessary, remedied.

Table 2 General scheme of a trace element analysis

Step	Topic	Associated problems	Influence on final results[a]
I	Defining the analytical task	Defining the scientific problem: general association (degree of difficulty), extent (number of samples, frequency of sampling), urgency	++++++
II	Planning	Orientation of task on the potential of the laboratory (instrumentation, methods and staff; organizational structure and time schedule)	+++++
III	Sampling	Representativeness of the sample with respect to the objective of sampling; contamination during sampling, subsampling or homogenization	+++++
IV	Sample preservation	Contamination; losses; disturbance of speciation	++++
V	Sample storage	Stability of the constituents for short-, medium- or long-term storage (specimen banking)	++++
VI	Sample digestion or decomposition	Analyte losses due to volatilization or insufficient digestion; contamination from reagents and containers	++++
VII	Trace–matrix separation	Efficiency of separation; completeness of preconcentration; contamination from reagents and containers	++++
VIII	Instrumental determination	Suitability and performance of the selected procedure for the given analytical task; conditions of the instruments used	+/++
IX	Calibration	Selection of a suitable calibration technique; quality of the calibration samples and calibration solutions employed; adequate weighting of the blank values	++/+++
X	Data evaluation	Selection of all relevant values and statistics; sample identifiers	++
XI	QC	Adequate statistical control of all analytical steps; selection of appropriate QC samples; interlaboratory comparison	++
XII	Analytical report	Adequate assessment of the analytical results and QC	++
XIII	Interpretation	Adequate assessment of the analytical results with respect to the environmental task (ecochemical, ecological, geological, oceanographical, ecotoxicological, toxicological, occupational, epidemiological, etc.)	++++++

[a] + = 2–20%, ++ = 20–50%, +++ = 50–100%, ++++ = 100–300%, +++++ = 300–1000%, ++++++ = production of 'data cemetery'.
QC, quality control.

A number of rules for the conductance and documentation of an analysis, meant to facilitate achievement of such a goal, are generally summarized under the term good laboratory practice (GLP). Further, it is of equal importance to use methods which are fit for the purpose and are well known to the analyst. If such methods are available and validated, their performance has to be controlled throughout the whole duration of the analysis. The installation of a suitable QC program is often a part of the contract or, as many other aspects of environmental analysis, is part of national and international regulations. General guidelines for the installation of such programs are available.[11] In case of lack of validated methods, effective method development must take place, which has the role of detecting possible sources of errors and eliminating or reducing them to an acceptable level. Normally such method development starts by modifying existing methods rather than starting from scratch.

3.2 Sampling and Sample Pretreatment

3.2.1 Sampling

The analytical procedure starts with the removal of a portion from the material to be examined; this is the laboratory sample that subsequently will be analyzed.[12] During the sampling procedure, care must be taken to ensure that the laboratory sample is representative of the material from which the sample was taken or of the situation which is to be examined. Incorrect sampling renders even the most accurate analysis ad absurdum, a fact often underestimated, since the largest errors take place during sampling and cannot subsequently be corrected. In general, a differentiation can be made between two procedures for representative sampling: (a) random sampling and (b) systematic sampling. Random sampling means the random collection of individual specimens from a total parent population frequently aiming at the determination of an average value (e.g. collection of birch leaves from a birch forest to determine average calcium content at time x). Systematic sampling, on the other hand, mainly refers to problems of changeable values as a function of space and time (e.g. the collection of spruce needles of a certain age or from various levels of a spruce tree to determine the exposure to air pollutants), thereby coupling a specific systematic study to a random sample. Regardless, the following three fundamental sampling rules are valid for both types of sampling:

1. The sample taken from the system should have exactly the same chemical composition as the original material.
2. The probability of being taken as a sample from a total population or total mass of material must be equally great for each individual samples.
3. The greater the variability between individual samples and the greater the number of individuals or the greater the mass of material to be evaluated, the greater are the efforts required for sampling.

Generally, different sources of error prevent the first two fundamental rules from being fulfilled, such as problems with the sampling strategy (place, time, duration and method of sampling), the selection of the apparatus (contamination from the materials used, sorption losses), the execution (sampling, subdividing, homogenization, filtering, preservation) and delivery (transport, identification, protocol). The particular problems depend very much on the type of material or situation to be investigated, and therefore will be discussed later (see section 4).

3.2.2 Sample Preservation

If an analysis is to have any meaning, care must be taken to ensure that the laboratory sample is unadulterated during transportation to and storage at the laboratory while awaiting further sample pretreatment. The preservation of a sample in all of its complexity is impossible and therefore can only be replaced by a 'freeze' of selected properties of the sample, such as pH and pE values, as well as the contents of main and trace constituents, matching the needs of the given task. Microbiological activity must be suppressed just as much as analyte concentration changes caused by adsorption losses or contamination. Subdivision of the sample into a number of portions may be necessary if a number of constituents must be determined that require different stabilization. If a number of analytes are in a dynamic equilibrium (e.g. redox equilibrium), such as in speciation analysis, preservation is exceedingly difficult, if not one of the most difficult steps of the whole procedure.[13–15]

3.2.3 Sample Pretreatment

Since the risk of systematic errors increases with the extent of sample pretreatment, the number of steps and the mass of reagents used during such treatment should be kept to a minimum. Similarly, direct methods are preferred to methods which require extensive preparation such as separation and preconcentration. However, solid samples can generally only be directly analyzed relatively free of problems by GFAAS, so that a digestion is usually required prior to determination. Sample decomposition is prone to three different types of error sources: (a) incomplete dissolution, (b) loss of analytes by volatilization and (c) contamination by the reagents

(acids, etc.). Fusion and dry ashing are the most seriously affected with respect to volatilization losses owing to the higher temperatures employed, but acid digestions are not without problems. All digestion procedures performed in open systems are prone to external contamination, which can be best avoided by operation in closed vessels under pressure.

The risk of errors also increases with decreasing analyte concentration. For those cases in which a low analyte concentration requires separation and preconcentration, all steps should be either carried out under clean working conditions (clean room or clean bench) or performed in a closed system, such as provided by FI manifolds.

3.3 Atomic Absorption Analysis and Methods

It has been pointed out already that normally the determination step is the most accurate and precise step of the whole analytical procedure. However, such a general statement can be valid only if the determination step is following a well-developed method, using an appropriate technique and instrumentation that is well calibrated and maintained in good condition. Each of the AAS techniques (e.g. FAAS, GFAAS, HGAAS) has an optimum working range with respect to concentrations, analytes and matrices and should not be used outside this range. AAS is a relative technique in which quantitative determinations are performed by comparison with calibration samples (calibration solutions). Since changes in the instrument parameters, the condition of the instrument, and the conditions of measurement are reflected in the calibration and are thus eliminated (except drift), it is still possible to obtain quantitative results even when the sensitivity is far from optimum. Despite this possibility, working in this manner cannot be recommended since then substantial control of the condition of the instrument is lost. Apart from the fact that the performance characteristics of an analysis are degraded under these conditions, judgment of the analytical data is questionable since significant changes in the sensitivity or noise are indications of malfunctions of the instrument or errors in the method, which may lead to measurement errors.

3.3.1 Instrument Fault Recognition

The absence of instrumental faults is a prerequisite for a meaningful analysis. For this reason a protocol of the performance characteristics of every analysis should be carefully maintained, and any larger deviations from specified values or values in earlier protocols should be a reason to search for the cause.

Frequent causes of a lack of sensitivity in AAS are in general as follows:

1. dilution errors when making up the calibration solutions;
2. wrong or faulty lamp; too high a lamp current;
3. analyte losses in the stock or calibration solutions due to insufficient stabilization or unsuitable containers;
4. incorrect wavelength or too large a slit width;
5. matrix effects due to the use of either the wrong buffer or modifier, or the absence of either.

The following causes can be additionally present in FAAS:

6. wrong flame type or nonoptimized flame stoichiometry;
7. nonoptimized horizontal or vertical burner position;
8. clogged sampling capillary, nebulizer, or burner slot; incorrect aspiration rate;
9. unsuitable inserts in the spray chamber; misaligned impact bead.

Frequent causes of reduced sensitivity in GFAAS (enhanced characteristic mass) are as follows:

10. erroneous sample dispensing;
11. preatomization analyte losses; these can be due to a pyrolysis temperature that is too high or to either an unsuitable or missing modifier;
12. unsuitable (excessive) atomization temperature;
13. erroneous blank value or baseline correction;
14. too high a gas flow rate during atomization or the use of an unsuitable purge gas;
15. too low a heating rate; this can be caused by an unsuitable power supply or poor electrical contact of the graphite tube;
16. too low a magnetic field strength for ZBC (e.g. insufficient voltage at the electromagnet).

Much more seldom than decreased sensitivity is enhanced sensitivity (reduced characteristic mass). This can be traced almost without exception, especially in trace analysis, to accidental introduction of the analyte. If this occurs a systematic search should be made to determine the cause. All analytical results generated under such conditions, and also all corresponding publications, should be regarded with considerable scepticism.

Insufficient precision or excessive noise generally has one of the following causes:

1. low lamp energy; poor quality lamp;
2. wrong lamp current or poor lamp alignment;
3. wrong wavelength or too narrow a slit width;
4. contaminated optical components (windows, lenses, mirrors, etc.);
5. emission noise;

6. poor alignment of the line source and the continuum source for background correction;
7. inhomogeneous sample solution or slurry, the presence of suspended matter in the measurement solution, nugget effect for solids analysis.

For FAAS the following specific causes in addition to those above may be considered:

8. partially clogged nebulizer or burner; defective nebulizer;
9. contaminated spray chamber or inserts (large droplets on the surface are an indication of wetting problems);
10. irregular drain flow from the spray chamber;
11. irregular gas flow (e.g. too low a gas pressure);
12. unsteady flame caused by drafts, strong ventilation system, change of solvent, etc.;
13. an excessively short integration time or insufficient damping.

Specific causes of insufficient precision or excessive noise in GFAAS include the following:

14. irreproducible dispensing of the test sample solution;
15. spattering of the solution caused by either too rapid or incomplete drying;
16. irregular spreading of the measurement solution over the atomization surface (wetting problem);
17. aged graphite tube or contacts;
18. an excessive integration time or incorrect integration window;
19. baseline offset correction (BOC) too short or erroneous.

Finally, drift is often caused by one of the following problems:

1. changes in the radiant intensity of the lamp (or lamps) or of the line profile during the warming-up phase or in the event of a defect; changes in the ratio of the radiant intensities of the line source and the continuum source for background correction;
2. evaporation of the solvent from the measurement solution;
3. for FAAS, warming or cooling of the burner or the formation of encrustations on the burner slot;
4. for FAAS, changes in the rate of aspiration or in the efficiency of nebulization due to physical changes in the sample capillary or the nebulizer;
5. for GFAAS, aging of the graphite tube or the accumulation of residues (e.g. carbon from biological materials).

This list of frequent causes of marked deviations in the most important performance characteristics can only serve as a starting point for troubleshooting. Nevertheless, deviations of this sort are nearly always an indication of a malfunction or an operational error and must therefore be taken seriously. The important performance characteristics such as sensitivity (or characteristic mass or concentration) and the signal-to-noise ratio (S/N) should always be documented with the analytical results so that malfunctions or operational errors can be traced or eliminated. In modern instruments, functions such as monitoring the condition of the instrument and maintaining a protocol are performed automatically with the corresponding algorithms being implemented in the instrument software.

A convenient procedure for monitoring further instrumental functions in GFAAS, in addition to the sensitivity and S/N, has been described by Slavin et al.[16] Using a certified reference material (CRM) (Trace Elements in Water, NIST 1643b), the characteristic masses and the Zeeman factors for chromium, copper and silver are determined without a chemical modifier and under easily reproducible conditions; additionally the lamp energy is recorded. Copper exhibits the greatest loss in sensitivity at the 324.8-nm line when ZBC is applied and is very sensitive to changes in the magnetic field strength. This element can thus serve as an indicator for an insufficiently strong magnetic field. Chromium is a difficult to volatilize element and requires a relatively high heating rate; at the 257.9-nm line its determination is sensitive to emission interferences, which can indicate a poorly aligned graphite tube atomizer. Silver, on the other hand, is not noticeably influenced by the magnetic field strength, the heating rate or the alignment of the graphite tube. If the sensitivity of all three elements is influenced to the same degree, this is an indication of dispensing problems or that the actual temperature deviates markedly from the rated value. If this test is performed at regular intervals, it serves as a powerful tool to document the state of the instrument.

3.3.2 Instrument Optimization

Various strategies have been applied for optimizing the instrument parameters of AAS techniques, such as methods of experimental design,[17] simplex optimization[18] or complete response-surface mapping.[18] The use of such chemometric methods for optimization is useful, since running simple series of experiments changing one parameter at a time may lead to nonoptimum conditions. The situation is further complicated, since different performance characteristics can exhibit varying optima, so that compromise conditions may represent the practical optimum. For example, often it is the case that the conditions for optimum sensitivity are not the same as the

conditions for the best possible tolerance to constituents of the matrix. The required task must therefore be defined before optimization can be undertaken. Frequently, the task is to optimize analytical performance characteristics such as detection capability, precision, working range or tolerance to matrix constituents, but occasionally considerations of economics, such as high sample throughput, minimal manual effort, low sample or reagent consumption or low costs, are more important.

In FAAS we can observe that a burner position or a nebulizer which is optimized for maximum sensitivity can exhibit very low tolerance to matrix constituents. In the case that high sensitivity is not the primary objective, the nebulizer/burner should be optimized for high tolerance against matrix constituents. This is normally done by adjusting the nebulizer and its impact bead (if used) for generating mostly very fine droplets rather than for maximum aerosol density. Also, the observation height for the best sensitivity and the lowest interference effects might be different.

In GFAAS, the stabilizing effect of a modifier (e.g. palladium) can lead to a lower rate of atomization, so that broader signals are obtained than if a modifier were not used. Such signal broadening may lead to a deterioration of the S/N and therefore to a poorer limit of detection (LOD). On the other hand, the linear working range can be extended upward since the limiting absorbance is only reached at higher analyte masses, and the tolerance against concomitants is significantly improved owing to more effective separation during the pyrolysis step and the higher effective atomization temperature.

For samples that do not require the best detection limits achievable by GFAAS, the sample throughput can be enhanced and the analysis costs reduced.[19] This is possible by the reduction of the furnace cycle time. A set of strategies can be combined to achieve fast furnace programs: (a) reduction of the sample volume to the minimum for the required sensitivity allows for (b) shortening of the drying step and (c) if the sample matrix is not complex, for the omission of the pyrolysis step. The drying step can be further shortened if started already during sample dispensing (hot injection). For maximum throughput, the addition of modifiers can be either omitted totally or done in advance directly into the sample cups rather than the graphite tube.

A different strategy leads to improved detection power. Since the measurement signal of GFAAS is proportional to the analyte mass and not to the concentration, the detection capability can be improved by dispensing larger volumes of the measurement solution. When the upper limit for a single dispensing is reached, which is dependent on the particular design about 40–100 µL, multiple injections each followed by a drying step can be used with modern autosamplers to enhance the total sample mass further. The additional time required for these steps can be reduced by injecting into a preheated graphite tube.

The greatest improvement in sample throughput can be achieved by using simultaneous multielement analysis. However, multielement methods require particular attention to be paid to the optimization of the FAAS[20] or GFAAS[21] operational parameters. The conditions normally chosen for pyrolysis and atomization of various analytes in graphite furnaces differ significantly. This behavior appears to stand in the way of attaining acceptable compromise conditions. However, the selection of compromise conditions is relatively simple, since the most volatile analyte dictates the maximum pyrolysis temperature and the least volatile analyte dictates the minimum atomization temperature. Nevertheless, the use of a transversely heated atomizer with an integrated platform is decisive, since it allows one to apply STPF conditions for both volatile and nonvolatile elements. Losses in sensitivity and selectivity under such compromise conditions are much smaller than assumed when a modern instrument concept is applied. For example, the loss in sensitivity because of an increase in the atomization temperature above the optimum due to increasing diffusional losses[22] is partially compensated for by an increase in the absorption coefficient with increasing temperature. Furthermore, the signal narrows with increasing temperature, improving the LOD since shorter integration times lead to an improvement in the S/N.[22] All in all, only small losses in the detection capability by a factor of 2–3 are expected for multielement determinations, which can be tolerated in many cases.

3.3.3 Recognizing and Overcoming Interferences

As a result of a lack of selectivity, the presence of concomitants in the test sample can cause interferences in the determination of the analyte. An interference can influence both the mean value of the analytical results and the attainable precision. A false mean value may lead to measurement errors if it is not eliminated or if it is not taken into consideration in the evaluation report. Measures for the elimination can be instrumental, such as appropriate background correction, or the choice of a suitable calibration technique, such as the analyte addition method, or the use of chemical additives. According to the International Union of Pure and Applied Chemistry (IUPAC), two types of interference can be distinguished, spectral and nonspectral interferences. Spectral interferences are due to the incomplete isolation of the radiation absorbed by the analyte from other absorption or radiation sources. Nonspectral interferences directly affect the number of free analyte atoms in the absorption volume and can be further classified according to the place or time of their occurrence.

3.3.3.1 Spectral Interferences Since AA spectra are much simpler than those of optical emission spectrometry (OES), spectral interferences are encountered far less frequently than in OES. The following cases can be distinguished:

1. direct overlap of the analytical line with an absorption line of another element;
2. absorption of the radiation of the analytical line by gaseous molecules;
3. absorption by concomitants of other radiation from the radiation source (or sources) that is not separated by the monochromator;
4. radiation scattering caused by particles in the absorption volume;
5. thermal emission of concomitants in the bandwidth of the monochromator.

While the thermal emission of concomitants is generally corrected for in AAS by modulating the primary radiation source, the light still reaches the detector. In the case that the amount of light from this interfering source is significant with respect to the modulated light from the primary source, increased noise is detected, lowering the attainable precision.

Only a limited number of direct overlaps occur in AAS, and of the 38 cases reported[2] only 10 are related to primary analytical lines. In most cases the interfering element must be present in a large excess to cause false measurement, which has not been reported in environmental analysis.

Spectral interferences of type 1 and 3 can be observed more often and also have been reported for environmental analysis. When a continuum source is used for background correction, any concomitant that exhibits an analytical line within the bandwidth of the monochromator can absorb radiation from the continuum source to a degree corresponding to its concentration. This supposed background absorption is then falsely subtracted from the AA, leading to overcompensation and thus to a measurement error. In order to absorb significant amounts of the radiation from the deuterium source, the interferent must be present in high concentrations. Normally this type of interference can be observed in the determination of trace elements in metallurgical samples such as alloys. However, a few cases have also been reported for environmental samples, where the matrix element iron interferes with the determination of Pb 217.0 nm and Cd 228.8 nm[23] and Se 196.0 nm and Sb 217.6 nm[24] and aluminum interferes with the determination of As 193.7 nm[24,25] and Sb 217.6 nm.[24] Molecular absorption also can lead to interference, either directly by overlapping with the analytical line (type 2) or by absorbing radiation from the background correction channel (type 3). Particularly cumbersome are the interferences from P–O for the determination of arsenic, selenium and tellurium and from S–O for the determination of selenium.

Generally, problems arising from background absorption (interference type 3/4) are more frequent with GFAAS than with FAAS.[26] In flames, background absorption is rarely significant at wavelengths higher than about 240 nm, rarely exceeding 0.05 absorbance. In furnace analysis, background absorption can be significant at wavelengths up to 500 nm, exceeding signal levels of 2 absorbance units. Type 4 spectral interferences (light scattering) is caused by the condensation of the sample matrix (or the modifier!), by forming a smoke or mist. This event occurs when a huge amount of volatilized matrix reaches cooler regions in the absorption volume, caused by temperature gradients inside the atomizer.

The prerequisite to counter such interferences is to recognize them. Spectral interferences can be detected by changing to another alternative analytical line, by changing the slit width of the monochromator and by measuring a matrix blank solution that is identical to the test solution but does not contain the analyte. In GFAAS, careful inspection of the time-resolved absorption signal may give indications of spectral interferences, for example, when the analyte signal falls below the baseline or when the analyte signals shows minima coinciding with maxima in the background channel. To counter such interference problems, there are some complementary measures available to the analyst:

1. selection of the appropriate background correction;
2. optimization of the temperature programming;
3. use of chemical modifiers.

Interferences of type 2 and 3 can in general be compensated by ZBC with the exception of molecules showing a Zeeman effect.[27] Such behavior has been reported for the molecular species N–O in the determination of Zn 213.9 nm[28] and Se 196.0 nm[29] and P–O in the determination of Ag 328.1 nm, Cd 326.1 nm, Fe 248.3 nm, Hg 253.6 nm and Pd 244.8 nm.[30,31] For such cases the optimization of the method is the only measure to avoid these interferences. The goal of method optimization is either to avoid the interferent species (do not use phosphate modifier, phosphoric acid, in digestions or extractions) or to separate the analyte and the interferent species by optimizing the thermal programming of the atomizer and the use of appropriate modifiers.

3.3.3.2 Nonspectral Interferences Nonspectral interferences are those which have an influence on the number of analyte atoms in the absorption volume. In the case where the mechanism of the interference is known, the interferent can be classified by naming the place where it

is acting or the process in which it is involved, for example, condensed-phase interferences concern all processes with the liquid and solid sample until it vaporizes, while vapor-phase interferences occur thereafter. For example, a condensed-phase interference might occur in FAAS when the presence of a particular matrix influences the transport of analytes from the solution to the absorption volume. This would cause a transport interference by influencing the nebulizer uptake rate, its nebulization efficiency (due to changing viscosity and/or surface tension) or the transport of the aerosol from the spray chamber to the flame. Volatilization interferences can be observed in both FAAS and GFAAS. In FAAS this type of interference occurs when the analyte forms a specific compound together with the interferent, which shows a different thermal stability (most often higher stability) in the flame or is occluded in a matrix particle of a difficult to melt or volatilize substance. In GFAAS, the most common form is premature loss of the analyte during thermal pretreatment by forming a volatile species with the analyte. Vapor-phase interferences can occur in all AAS techniques when the fraction of analyte that is dissociated, ionized or excited in the vapor-phase is changed owing to the presence of an interferent.

The presence of such nonspectral interferences can be diagnosed by comparing the slope of the calibration graph for a standard solution with that obtained from the complex sample solution. Nonspectral interferences can generally be eliminated by matching the test sample solutions and the calibration solutions as closely as possible. A practical way of doing this is to modify the sample composition by adding a spectrochemical buffer to all solutions (in the case of FAAS, called ionization buffers or releasers depending on the type of interference to be overcome) or a modifier (in the case of GFAAS), counteracting the interference effect. Another practical way of matching samples and standards is the application of the standard addition technique, where the calibration graph is directly obtained from the sample by spiking it with known quantities of the analyte.

3.3.4 Method Development and Validation

A suitable analytical method must have been developed and adequately characterized before analysis can take place, that is, measures must be taken to ensure that major problems such as interferences or instrument malfunctions are recognized and eliminated, and that the method generates reasonably accurate results. In particular the following tasks have to be performed during methods development and validation:

1. check for contamination;
2. determination of the precision within a series and between series;
3. determination of the LOD and limit of quantification (LOQ);
4. determination of the linear working range;
5. check for possible interferences;
6. comparison of the analytical results with those from alternative methods;
7. comparative analyses with other experienced laboratories;
8. analysis of CRMs closely matching the composition of the target sample.

Since a contract laboratory can only select methods that have been adequately characterized, in principle it cannot accept chance analytical tasks.

3.3.5 Measurement, Calibration, Evaluation

The material obtained after suitable treatment or preparation of the laboratory sample is the test sample (or, if only analytical chemistry is involved, the analytical sample). A test (or analytical) portion is removed from this for analysis. If the sample is a solid material, it may be analyzed directly by solid sampling GFAAS. In most cases it is transferred into a test sample solution, by performing a digestion, extraction or dissolution. The test sample solution can be used directly or after further treatment steps, such as dilution, addition of buffers or modifiers, and so on, for the measurement. After having sorted out all the possible error sources during method development and instrument maintenance, only a few errors sources remain that may influence the final measurement. The measurement solution actually used for the measurement by AAS must be stable for the duration of the analysis. This is especially important if an autosampler with numerous sample cups is used, since the measurement solutions can stand for several hours before being analyzed. While solvent extracts with volatile solvents or complexes with unstable metal chelates are particularly critical, even aqueous solutions become more concentrated with time; this effect becomes greater the smaller is the volume of the sample container. The evaporation rate from 2-mL standard GFAAS autosampler cups at normal room temperature is about $35\,\mu L\,h^{-1}$ and can be reduced to about $10\,\mu L\,h^{-1}$ when the autosampler is covered, and additional containers filled with water for humidification are placed below the cover.[16]

Another significant source of errors for the instrumental determination is the calibration.[32] Errors occur mainly during the preparation of calibration solutions by dilution of stock solutions. The dilution error increases with both the number of dilutions and the ratio of the dilution step.[33] In order to reduce such errors, only calibrated volumetric devices of quality class A should be used or, better, the dilutions should be performed gravimetrically;

dilution factors of 1 : 100 per dilution step should not be exceeded. Another error source in calibration is the purity and stability of the stock solution, which especially for speciation analysis cannot always be guaranteed. Also problematic is the handling of any blank solutions, which often are misused as the zeroing solution for evaluation of the signal baseline of the instrument, but for which the pure solvent should be used.[34] Blank solutions must be analyzed in the same manner as other sample solutions, and the result for the analyte concentration of samples should be corrected accordingly. Further, calibration can be significantly in error if the wrong calibration function is used. This is especially the case when a linear calibration is assumed in situations where the calibration function is nonlinear.

3.4 Quality Assurance

The accuracy of a determination depends not only on the selection of a suitable procedure, but also on the stability of the performance characteristics of that procedure. In principle, two different strategies are applied to verify that both goals are met. QC is the mechanism of verifying the absence of errors, while quality assessment is the mechanism of verifying that the system is operating within known and acceptable limits.[35] Quality assurance (QA) for environmental analysis was a growing feature of the 1990s, as illustrated by the number of QA guidelines and systems which are currently being implemented. The increasing interest is also reflected in the launch of the journal *Accreditation and Quality Assurance* in 1996.

3.4.1 Internal Laboratory Operations

The quality of laboratory operations, in the most general sense, is measured by the dedication of higher management to emphasizing quality in a delicate balance with the ever-present demands of productivity. QC always requires a high level of effort and must therefore be accommodated to the DQO (i.e. the higher the requirements on the accuracy of a method, the greater is the effort required for QC). Nevertheless, higher costs for QC can be justified when costs in other areas are reduced, such as those caused by repeated measurements or by false results.

The first requirement for implementing a QA program is the organization of events in the laboratory following certain rules that are generally summarized as GLP and good analytical practice (GAP). GLP provides guidance in the organization of laboratory events and for the conditions under which laboratory investigations are planned, conducted and assessed. By maintaining a protocol and issuing a report of the investigation, the aim of GLP is to make procedures in the laboratory so transparent that they can be verified by third parties (e.g. regulatory authorities, clients). The scope of these rules are laid down in the chemical legislation in force in most industrial nations. The installation of such rules is most often part of the contract between the client and the laboratory. A detailed discussion on this topic is beyond the scope of this article and the interested reader is therefore referred to introductory monographs.[36] Maintaining all the rules of GLP is both expensive and requires a considerable increase in personnel. Some typical violations of GLP rules are as follows:

1. balances remaining uncalibrated;
2. failure in recertifying mechanical pipets;
3. use of pH buffers of unknown age or origin;
4. use of unsuitable purity standards;
5. application of software beyond its purpose.

On the other hand, GLP alone cannot guarantee the trueness and quality of analytical data. The final responsibility for QA still remains with the analyst. QC and quality assessment programs must be implemented and the rules of GAP must be followed in a laboratory to reach the DQO. The rules of GAP include the following:

1. performing the analysis with an adequate number of adequately trained staff;
2. the use of validated methods;
3. analytical instruments that are calibrated and maintained in good condition;
4. the use of CRMs for calibration;
5. an effective internal QA program.

As far as possible, effective QC should be incorporated directly into the analytical program. Analytical results cannot first be generated and then deliberations subsequently made about their quality. Typically, the more complex the sample, the greater the effort the laboratory should make with respect to QC. Depending on the DQO, the QC program must be designed either to maximize the probability of error recognition or to minimize the probability of a false warning. By trapping and excluding error sources such as drift, the QC program can lead to an improvement in the method performance.

In order to be effective, at least the following procedures should be part of the internal quality assessment:

1. characterization and optimization of the precision and accuracy of the calibration function;[37]
2. characterization of the blank values;
3. characterization of the instrumental repeatability (precision on replicates);
4. characterization of the method repeatability (precision on duplicate samples);

5. characterization of the method performance by analyzing a control sample of known analyte content (recovery);
6. characterization of the accuracy of the method by analyzing a CRM (recovery).

To assure that the method is under statistical control, the results of such performance checks are documented in the form of control charts. Control charts serve multiple functions,[38,39] including:

1. recognition of drift or malfunctions;
2. to uncover errors in sampling or sample preparation;
3. to decide, whether the analysis is fit for purpose;
4. to support the analyst with respect to his/her clients by assuring his/her results.

3.4.2 External Quality Control

Outside evaluation of the laboratory provides an opportunity for the laboratory to view quality from another perspective. The GAP rules include the following external QC:

1. regular participation in interlaboratory ('round-robin') trials;
2. quality assessment by an independent QA unit;
3. accreditation or other certification of the laboratory by regulatory authorities.

Regular interlaboratory trials are organized by several national and international quality programs, sending performance evaluation samples to the participants before an audit or as part of ongoing performance checks. An on-site audit by a third party group can be very effective in reviewing the installed quality programs.

4 ENVIRONMENTAL REGULATIONS AND METHODS

Environmental studies cover a broad range of disciplines, for example, analytical chemistry, geology and biology, and include several aims such as monitoring (routine analysis), research (studies of environmental pathways) and modeling. In many cases chemical analyses are the basis of these studies. Environmental analysis traditionally is divided into the analysis of water, soils and the atmosphere. Such 'classical' division cannot be strictly followed here in view of the multitude of different environmental samples originating from the different compartments of the 'environment' and the mutual interactions that can be expected between them (see Figure 3).

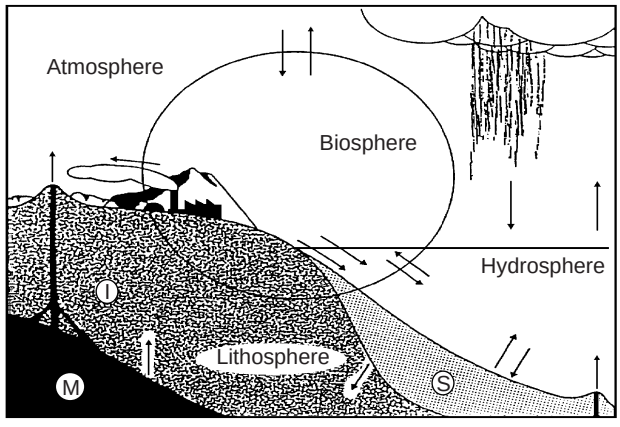

Figure 3 Schematic representation of our environment and the interactions that can be expected. S = sediments; I = soils and rocks; M = magma. (Reproduced by permission from Mitra.[105])

Nearly all types of sample materials can become the target of environmental analysis, at least when the material is disposed of as waste. In order to assign the most widely varying materials into suitable categories, three categories were chosen. The categories are based on the analytical methods used for their analysis, rather than the environmental compartment from which they are taken; the categories are (a) water samples, (b) solid materials of mainly mineral phases and (c) biological materials.

A particular characteristic of environmental analysis is that most aspects of the analysis, (i.e. the frequency of the analyses, the strategy and methods of sampling, the pretreatment and instrumental determination, the QC of analytical results and the data presentation) are governed to a large extent in the finest detail by laws, regulations and guidelines. An overview of prescribed analytical methods and of international and German standards was given by Hein and Kunze.[40] Since the analyst must procure and follow the original method exactly, in order to be compliant with these rules, details of such methods will not be presented here. Instead, information will be given that will point to the most important work.

Introductory information on environmental analysis is available in the form of monographs[40–45] and review articles.[46–52] Publications on the determination of individual elements in environmental samples are also available, for instance, for aluminum,[53–55] arsenic,[56–61] cadmium,[62–65] chromium,[66–71] cobalt,[72] copper,[73] lead,[74–78] manganese,[79] mercury,[80–85] molybdenum,[86] nickel,[72,87,88] platinum,[89,90] selenium,[91–95] tellurium,[92] thallium,[96,97] tin,[98,99] vanadium[100] and zinc.[101] Regular reviews of environmental analytical chemistry appear annually in the *Journal of Analytical Atomic Spectrometry*[102] and every second year in *Analytical Chemistry*.[103] The application of AAS in

environmental analysis has been recently reviewed by Sturgeon.[104]

4.1 Water Analysis

Although the compound H_2O is clearly defined for the chemist, the matrix 'water' offers an unbelievable multitude of forms for the analyst, reflected in the number of methods for its analysis. Water samples can be classified by legislators, analysts or users by the degree of its anthropogenic contamination (natural waters, processed water, polluted water, wastewater), by its salinity (fresh water, briny water, seawater), by its environmental compartment (atmospheric precipitation, surface water, groundwater, glacial ice) or by its intended use (drinking water, cooling water, bathing water, process water). Correspondingly, the palette ranges from highly pure waters to waters containing extremely high contents of impurities, with respect to both the analyte concentration and the matrix content. The parameters analyte concentration, total salt concentration and organic content are decisive for the procedure and the selection of the analytical technique.

Even though most natural waters contain over 99.9% of H_2O, accurate determination of elements in this simple matrix is frequently not as straight forward as would first appear. Most unpolluted natural waters contain less than $10\,\mu g\,L^{-1}$ of the commonly determined toxic metals and it is easy to obtain significantly biased results when measuring such low concentrations. The problem is obvious from the history of published concentrations of trace elements in seawater, which had decreased by around an order of magnitude within a decade. This is primarily due to (a) the improvements in the control of contamination during sampling, storage and analysis, and (b) the improved separation techniques and advances in the development of the instrumentation.[106] Therefore, the majority of data on the determination of the base concentrations of metals in seawater published before 1986 is practically worthless.[107] The same is probably true for other non-polluted waters and rainwater. Toxic metal levels in sewage and effluents are normally significantly higher than those of natural waters. However, interlaboratory comparison studies have demonstrated that inaccurate results at these higher concentrations can still occur unless meticulous attention to detail is observed in this case particularly to the control of interferences.

For an overview on water analysis, special monographs[108–112] and review articles[113] on this topic or more specific publications on seawater analysis[114–117] should be consulted. Review articles on the determination of trace elements in water are available for aluminum,[118–120] antimony,[121,122] iron[123] and mercury.[124] The journals *Water Research* and *Vom Wasser* (in German) deal exclusively with this subject.

4.1.1 Water Analysis and Legislation

As mentioned previously, environmental analysis is regulated to a large extent by national and international laws and regulations. Since water is a prerequisite for life, it is of primary concern. The different types of water meant to be consumed or used by humans or being discharged into the environment are extensively regulated and controlled. It is beyond the scope of this article to discuss such regulations in detail. However, some limits for a range of elements in treated (potable) water are given in Table 3 in order to discuss the applicability of AAS for the control of water quality.

In order to monitor for these elements at concentrations close to the limits set by regulations and to determine compliance, the LOD of the analytical technique should ideally be at least 10 times lower than the actual limit. By comparing the achievable LODs of different AAS techniques with these control limits, it is clear that FAAS can only be used to monitor directly seven of the 16 elements listed in Table 3.

4.1.2 Sampling and Sample Conservation

Appropriate sampling and subsampling techniques are a crucial prerequisite of the generation of meaningful

Table 3 Some international standards for treated potable waters (all concentrations in $\mu g\,L^{-1}$)[a]

Element	USA limit	EEC guideline value	EEC limit	WHO guideline value	German limit	Preferred method
Ag	50		10		10	FAAS, ETAAS
Al		50	200		200	ETAAS
As	50		50	50	40	ETAAS, HGAAS
Ba	1000	100			–	FAAS
Cd	10		5	5	5	ETAAS
Cr	50		50	50	50	ETAAS
Cu	1000	100[b] 3000[c]			–	FAAS
Fe	300	50	200		200	FAAS
Hg	2		1	1	1	CVAAS
Mn	50	20	50		50	FAAS
Na		20	150			FAAS
Ni		50			50	ETAAS
Pb	50		50	50	40	ETAAS
Sb			10			ETAAS, HGAAS
Se	10		10	10		ETAAS, HGAAS
Zn	5000	100[b] 5000[c]				FAAS

[a] EEC, European Economic Community; WHO, World Health Organization.
[b] At the water treatment plant.
[c] After 12 h in the pipe.

environmental data. The technique of sampling has to be selected in accordance with the requirements determined by the type of water that is to be sampled, such as piped water (drinking and industrial waters), springs and wells, groundwater (bore holes, seepage waters), flowing surface waters (canals, creeks, rivers, streams), still surface waters (ponds, lakes, barrages), marine waters and precipitation. The majority of still and flowing waters exhibit substantial variations in their element concentrations over time and in their spatial distribution. It is therefore relatively difficult to take a 'representative' sample. The required task must therefore clearly be established before sampling is undertaken. The simplest case is to take random samples, for example, at an outflow location at a given time. However, sampling of this type is only meaningful when the place and time of sampling are clearly defined, based on prior detailed information, and refer to a special situation. Much better information can be obtained for one or several analytes by determining the change in concentration with time, either continuously or at regular time intervals. For example, such sampling could be meaningful for the inflows and outflows of a sewage treatment plant. The degree of information can be further enhanced by a network of sampling points (e.g. along the length of a river or over the depth profile of a lake), and changes at each sampling point over the short term (days) or over the long term (months, years) are monitored. Multivariate statistical procedures can aid in the interpretation of such complex data.[125] Sampling strategies are described in numerous national and international standards,[40] some of which are given in Table 4, and also in a number of review articles.[106,113,126]

The collection of wet precipitation (rain, snow, cloud aerosols, etc.) is especially challenging,[127] particularly when the wet precipitate must be kept separate from the dry precipitate. For this purpose, rainwater collectors have been designed using a device that covers the collector during dry periods and opens only at the onset of rain[128–131] or by changing the exposed collection funnel by moving the cover at that moment.[132] Materials for such precipitation collectors have to be selected with care, since substantial errors can occur owing to adsorption on the walls of the collection funnel and sample container.[133] Collection time is also important, especially if an appropriate preservative has not been added.[134]

Also problematic is the collection of seawater samples from different depths or from the water surface, which contains high concentrations of surface-active compounds (biofilm). The ship itself and all the equipment used for sampling (winches, wires, bottom weights, samplers and messengers) are potential sources of contamination either due to their material or by their action (mixing different layers). These problems are discussed in more detail in review papers on seawater sampling.[106,135]

The sample pretreatment following the sampling of water samples depends on whether the total concentration or the concentration of dissolved analyte should be determined. In the first case, the water sample may have to be digested depending on its properties and the selected analytical technique. In the second case, the water sample is filtered after sampling, usually through a membrane filter of pore size 0.45 μm. Even such a simple task might not be without problems. Dissolved analytes might be lost during filtration owing to adsorption on the filter material or apparatus,[136] or the sample might become contaminated during filtration.[137] Various parts of colloidal forms of the analytes might pass through the filter, depending on the state of the filter with respect to loading with solid material.[138–140] For these reasons, those variations introduced during sampling of water are

Table 4 National and international standards for water sampling

Topic	International	USA	UK	Germany
Sampling strategy	ISO 5667-1/1991		BS 6068-6.2	DIN EN 25 667-1
Sampling techniques	ISO 5667-2/1991			DIN EN 25 667-2
Terminology	ISO 6107-1-9/1989		BS 6068-1.2	
Frequency of sampling		EPA 0004.3		DIN 38 402-6
QA for water sampling	ISO 5667-14/1998			
Marine water	ISO 5667-9/1992		BS 6068-6.9	DIN 38 402-16
Wastewater	ISO 5667-10/1992	EPA 0006.8	BS 6069-6.10	DIN 38 402-11
Precipitation	ISO 5667-8/1993		BS 6068-6.8	DIN 38 402-17
River and stream water	ISO 5667-6/1990		BS 6068-6.6	DIN 38 402-15
Drinking water	ISO 5667-5/1991		BS 5667-6.5	DIN 38 402-14
Lakes, barrages	ISO 5667-4/1987		BS 6068-6.4	DIN 38 402-12
Aquifiers	ISO 5667-11/1993	EPA 600/2-81-160 EPA 625/R-93-003a		DIN 38 402-13
Mineral and medical springs				DIN 38 402-18
Swimming pools and baths				DIN 38 402-19
Tidal waters				DIN 38 402-20
Cooling water, boiler water	ISO 5667-7/1993			DIN 38 402-22

often significantly higher than the instrumental variation of the measurement.[141] Immediately following filtration, or sampling if the sample is not filtered, the sample must be stabilized, usually through the addition of 10 mL of nitric acid per liter of water sample. The requirements for preservation are different for the different water samples (e.g. high salinity acts in a stabilizing manner), and also depend very much on the elements to be determined. In general, the pH value must be <1.5 in order to minimize the risk of analyte concentration changes due to precipitation or wall reactions. Different procedures may be required if individual element species are to be determined (see section 4.1.5). If the sample has been filtered, it may be of interest to examine the residue separately, following a method often very similar to those used for the analysis of sludges or sediments (see section 4.2).

Preconcentration at the site of sampling has been proposed as a strategy for sample preservation;[142] the claimed advantages are (a) large volumes of sample do not need to be transported, (b) concentrated samples are more stable and (c) the risk of contamination during transport and storage is reduced. The disadvantages are that preconcentration is a selective procedure, including only (a) selected elements, (b) mostly ionic species and often (c) only one oxidation state of the analyte, unless a comprehensive sample preparation procedure is performed at the sampling site. If interferences occur during preconcentration, it is very difficult to recognize them subsequently, since the original sample is no longer available. This strategy has not received wide acceptance and has only been used for special cases such as the determination of gold[143] and the determination of species of chromium[144] and mercury.[145]

Samples should be analyzed soon after sampling, especially when species determination is desired. Stabilized samples should be kept in appropriate containers, whereby the most suitable material depends on the target analytes and the concentration levels [details can be found in an overview published by Sturgeon and Berman[126]]. The best materials for the storage of samples with very low concentrations of analytes are fluoropolymers with a very smooth surface,[146] such as perfluoroalkoxy (PFA) [polyfluoroethylene (Teflon®PFA)][147–149] or polished polytetrafluoroethylene (PTFE) (Teflon®)[150] with the exception of mercury, for which glass is the best material.

4.1.3 Methods for Flame Atomic Absorption Spectrometry

According to most official methods for water analysis by FAAS, a digestion can be omitted if the analyte can be completely detected without such pretreatment. Normally, this is the case for filtered and stabilized water samples and often even for unfiltered water samples, provided that they do not contain large amounts of suspended matter that could block the sample introduction system or cannot be fully atomized in the flame. However, the direct analysis of water samples by FAAS is limited by its lack in sensitivity. Apart from the main elements sodium, potassium, calcium and magnesium, and under given conditions strontium, none of the trace elements can be determined in unpolluted surface waters, rainwater and seawater directly by FAAS. If the task is to monitor given limiting values, such as in the drinking water regulations or other corresponding international guidelines (see Table 3), then it is also possible in some cases to control for barium, copper, iron, manganese and zinc, and under given conditions for cobalt and chromium. For the analysis of some wastewaters, seepage waters or more strongly polluted surface waters in general, it is also often possible to determine elements such as cadmium, lithium, nickel and lead satisfactorily by FAAS. However, for such types of sample, a digestion is often indispensable.

By far the most often used digestion procedure for the determination of heavy metals in polluted waters by FAAS is digestion in nitric acid–hydrogen peroxide. Nitric acid and hydrogen peroxide are added to the water sample in a beaker and the volume is reduced to a wet residue, if required under repeated addition of hydrogen peroxide. The residue is taken up in nitric acid and water and the digestate is analyzed by FAAS.

In order to avoid potential interferences, to which especially the analysis of strongly polluted waters might be exposed, the following points should be particularly observed:

1. Background correction should be used for all determinations at wavelengths < 350 nm.
2. The nitrous oxide–acetylene flame should be used for the determination of chromium and is also strongly recommended for the determination of iron and manganese in water samples with complex matrices (e.g. wastewater), even if the sensitivity with this flame is poorer than that with the air–acetylene flame.
3. The addition of at least 2 g L^{-1} La as the chloride is required if calcium and magnesium are determined with the air–acetylene flame.
4. The addition of 1 g L^{-1} Cs as an ionization buffer is recommended for all determinations with the nitrous oxide–acetylene flame and also for the determination of sodium and potassium with the air–acetylene flame.
5. The operating conditions should be optimized for minimum interferences rather than for maximum sensitivity, particularly for strongly polluted waters.

If the sensitivity is then inadequate, a more sensitive AAS technique should be chosen (e.g. GFAAS).

Before GFAAS was introduced as a more sensitive technique in AAS, preconcentration by solvent extraction in combination with FAAS was used for the determination of trace elements in fresh waters. In this case complete digestion of the sample is essential. Nowadays, such methods are seldomly used routinely even though large numbers of these have been and continue to be published. The arguments against such methods are numerous (e.g. lack of sample throughput, involvement of too much manual work, contamination problems and the presence of interferences are only some examples). As a general rule, direct determination by other more sensitive techniques is generally preferred. A possible exception is on-line preconcentration,[151] for example, by sorption on packed microcolumns,[152–156] and this aspect is discussed in more detail together with GFAAS. Table 5 lists methods for the analysis of water samples by FAAS.

4.1.4 Methods for Graphite Furnace Atomic Absorption Spectrometry

With a sensitivity of about 2–3 orders of magnitude higher than that for FAAS, GFAAS offers the best prerequirements for the direct determination of a large number of trace elements in water. Also, digestion is not required for the majority of water samples, since in situ pretreatment such as oxygen ashing can be made part of the GFAAS heating procedure. Even water samples with a high content of suspended particles, which call for a homogenization directly prior to dispensing, can be directly analyzed with the aid of a slurry sampler.[170] The quantities of reagents and external

Table 5 Methods for the analysis of water samples by FAAS

Element	International	USA	Germany	Other methods and applications (see refs.)
Ag		EPA 272.1 EPA 7760A	DIN 38 406-18	
Ba		EPA 208.1 EPA 7080A		157–159
Ca	ISO 7980-1986	EPA 215.1 EPA 7140		157, 158, 160–165
Cd	ISO 8288-1986 ISO 5961-1995	EPA 213.1 EPA 3170		159, 166
Co	ISO 8288-1986	EPA 219.1 EPA 7200	DIN 38 406-24	160
Cr	ISO 9174-1990	EPA 218.1 EPA 7190		159, 160, 166, 167
Cu	ISO 8288-1986	EPA 220.1 EPA 7210	DIN 38 406-7	158–160, 166, 167
Fe		EPA 236.1 EPA 7380		158, 160, 167, 168
K	ISO 9964-2/1993	EPA 258.1 EPA 7610	DIN 38 406-13	158, 160, 162–165
Li		EPA 7430		160, 169
Mg	ISO 7980-1986	EPA 242.1 EPA 7450	DIN 38 406-3	157, 158, 160, 162–165
Mn		EPA 243.1 EPA 7460		158–160, 167, 168
Na	ISO 9964-1/1993	EPA 273.1	DIN 38 406-14	158, 160, 162–165
Ni	ISO 8288-1986	EPA 249.1 EPA 7520	DIN 38 406-11	159, 166
Pb	ISO 8288-1986	EPA 239.1 EPA 7420	DIN 38 406-6	166
Sr		EPA 7780		157, 160
Zn	ISO 8288-1986	EPA 7950	DIN 38 406-8	158–160, 166, 167

pretreatment steps can be thus reduced to a minimum for GFAAS, which is a prerequirement for accurate results in the extreme trace range. This reduction of sample pretreatment compensates for the often-cited disadvantage of GFAAS that the measurement time is relatively long and the sample throughput is low, which can be largely overcome by the use of state-of-the-art instrumentation, providing fully automated sequential or even simultaneous multielement analysis.

Another often cited disadvantage of GFAAS is its susceptibility to interferences. The relevance of such interferences, especially those reported in the early history of the development of the GFAAS methodology, has been greatly reduced with the introduction of the STPF concept by Manning and Slavin in 1983.[171] These and later authors showed that numerous trace elements in surface waters could be determined free of interferences against matrix-free calibration solutions by applying the STPF concept, including platform atomization, the use of matrix modification and ZBC.[172–174] Nowadays, many standard methods are based on a direct determination, such as EPA 200.9 for the determination of Ag, Al, As, Be, Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Sb, Se, Sn and Tl and EPA 7010 for the determination of Ag, As, Ba, Be, Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pb, Se, Tl, V and Zn. Similar methods exist in Europe for Ag (DIN 38 406-E18), Cd (ISO 5961-1995, DIN 38 406-E19), Co (DIN 38 406-E24), Cr (ISO 9174-1990, DIN 38 406-E10), Ni (DIN 38 406-E11), Pb (DIN 38 406-E6), Se (DIN 38 405-E23), Tl (DIN 38 406-E21) and Zn (DIN 38 405-E8), and STPF conditions are recommended, even if they are not directly prescribed. Publications of direct determinations of trace elements in natural water samples are very numerous, and only selected papers are given in Table 6.

The probability of interferences increases with the complexity of the sample matrix and the requirements for low detection limits. Strongly polluted water samples usually contain a higher analyte content and therefore allow for substantial dilution, which greatly reduces interference effects during analysis by GFAAS. Mineral and therapeutic waters can nevertheless cause considerable analytical problems on occasion if they have a high matrix content, e.g. of sulfur compounds,[183,206] or when the analyte concentration is very low. Samples of such complexity require very careful method development, method optimization, validation and QC.

If the sensitivity of GFAAS is inadequate for the direct determination of trace elements in very pure water, the sensitivity can be markedly increased by using multiple injections[207–209] or by carefully reducing the sample volume by evaporation.[210–213] Nevertheless, while such procedures can easily be used for waters with low concentrations of inorganic or organic concomitants, more problems are expected when applied for wastewater, high-salt mineral water or seawater, so that the application for such samples is limited to cases where the analyte can be easily separated from the large amounts of matrix by thermal pretreatment.[214–218] Often, separation and preconcentration steps cannot be avoided for such complex samples.

The trace element concentrations in unpolluted seawater are in the lower micrograms to nanograms per liter range. From a comparison of the best published LOD for GFAAS with the trace element concentrations of two unpolluted seawater samples of different origins, one could conclude that it should be possible to determine a number of trace elements directly by GFAAS (see Table 7). However, an additional difficulty with seawater samples is the total salt concentration of $35\,g\,L^{-1}$ on average, which can lead to spectral and nonspectral interferences.

State-of-the-art systems providing high-performance ZBC together with STPF conditions are able to overcome these problems in some cases, as is demonstrated by selected publications on the direct determination of trace elements in seawater compiled in Table 8.

However, scrupulous optimization of the instrumentation and methods is necessary in order to reduce interferences to an acceptable level and to achieve the necessary S/N level. For this purpose the following points should be taken into consideration:

1. The STPF concept must be applied consistently to minimize spectral and nonspectral interferences. A THGA with integrated platform provides the best results.
2. As much of the chloride matrix as possible must be removed prior to atomization. The addition of ammonium nitrate or nitric acid as a modifier aids in vaporizing the chloride as NH_4Cl or HCl.
3. The thermal stability of volatile elements should be enhanced by the addition of modifiers (e.g. Pd) and/or the addition of hydrogen to the purge gas during the pyrolysis.
4. Since high background attenuation is to be expected, even when the matrix is largely removed, the application of ZBC is recommended.
5. A polarizer is not required for the longitudinal configuration of the magnet for ZBC. This configuration thus provides a better S/N than other configurations.

Even then, only a limited number of elements can be determined directly in seawater. For the others, as with the analysis of fresh water, preconcentration is indispensable. Great care is required to avoid errors at such a low concentration range $< 1\,\mu g\,L^{-1}$[113,126] owing

Table 6 Selected publications on the direct determination of trace elements in natural waters by GFAAS

Analyte	Test sample	Remarks[a]	Ref.
Ag, Al, As, Ba, Be, Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Se, V, Zn	Spring water	STPF conditions	175
Ag, Al, Ba, Cd, Co, Cr, Cu, Fe, Li, Mn, Mo, Ni, Pb, Zn	Surface water, drinking water		173
Ag, Al, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, Pb, Zn	High-altitude Alpine snow	Clean room conditions	176
Ag, Cd, Pb, Sb	Drinking water	STPF conditions, THGA, simultaneous determination	177
Al	Mineral water	Mg modifier	178
		PD–Mg modifier	179
Al, As, Cd, Pb	Mineral water		180
Al, As, Be, Cd, Co, Cr, Mn, Ni, Pb, Se, V	Surface water	STPF conditions	3
Al, As, Be, Cd, Cr, Pb, Sb, Se, Tl	Drinking water, mineral water	STPF conditions	158
Al, Cu, Fe, Mn, Pb, Zn	Cloud aerosol, Alpine precipitation		163
As	Drinking water	STPF conditions	181
As, Cd, Ge, Sb, Se	Thermal waters rich in Fe	Platform, Mo modifier, ZBC	182
As, Se	Mineral waters rich in sulfates	STPF conditions, ZBC	183
B	Surface water	Ca–Mg modifier	184
	Surface water, wastewater	Ni modifier, Zr-coated tube	185
Ba	Drinking water	ZBC	186
Be	Drinking water		187
	Natural water		188
Bi	Wastewater	Pd modifier	189
Cd, Cr, Cu, Ni, Zn	Rainwater	Influence through seawater aerosols	190
Cd, Cr, Pb	Wastewater	STPF conditions, ZBC	174
Cd, Cu, Fe, Mn, Pb, Zn	Rainwater	Comparison with PIXE	191
Cd, Cu, Pb	Rainwater	Comparison with ASV	192
Cd, Cu, Pb, Zn	Precipitation	Comparison with ICPOES	193
Cd, Fe, Mn, Pb	Surface water	Platform, HNO_3 modifier	194
Cd, Pb	Surface water	Platform, La modifier	195
Cr	Surface water	V–Mo modifier	196
Cu, Pb	Drinking water	STPF conditions, fast program	197
Mo	Surface water	Ca modifier	198
	Surface water	Pd^0 modifier, ZBC	199
Sb	Wastewater	Comparison with HGAAS	200
	Drinking water		201
Se	Surface water	STPF conditions, Pd^0 modifier	202
Si	Drinking water	Co modifier, Ti–C-coated tube	203
Sn	Tap water	STPF conditions, Mg or Pd–Mg modifier	204
V	Water	Ascorbic acid modifier	205

[a] PIXE, particle-induced X-ray emission; ASV, anodic stripping voltammetry.

to contamination or analyte losses, and the following guidelines should be considered:

1. The number of necessary steps in the procedure, the number of containers, funnels, etc. and the mass of added reagents should be kept to a minimum.
2. All containers should be used exclusively for trace analysis and must be cleaned thoroughly prior to use, preferably by fuming out with nitric acid[240–242] or in the case of very clean PFA by leaching with dilute aqueous solutions of H_2O_2, ammonia and triethanolamine, which reveals comparable cleaning effects for this material.[243]
3. All reagents must be of the highest purity. If necessary, they must be further purified, e.g. by subboiling distillation.
4. All procedures should be performed in a clean room, clean work-bench or at least in a dust-free environment in a closed system, to minimize contamination from the laboratory atmosphere.

Table 7 Trace element concentrations ($\mu g\,L^{-1}$) in unpolluted seawater reference materials compared with the published LODs for GFAAS.[219] (Reproduced with permission from B. Welz, M. Sperling, 'Atomic Absorption Spectrometry', Wiley-VCH, Weinheim, 1999)

Element	NASS-1[a]	ICES-5[b]	LOD
As	1.65	1.4	0.2
Cd	0.029	0.020	0.003
Co	0.004	0.004	0.15
Cr	0.184	0.08	0.01
Cu	0.099	0.123	0.1
Fe	0.192	0.376	0.1
Mn	0.022	0.240	0.01
Mo	11.5	–	0.03
Ni	0.257	0.202	0.3
Pb	0.039	0.049	0.1
Zn	0.159	0.392	0.1

[a] National Research Council Canada.[220]
[b] International Council for the Exploration of the Sea.[106]

5. Polymer containers, funnels, etc. become aged with usage, developing active sites on the surface,[149] being able to exchange ions,[244] and therefore must be exchanged regularly for new ones.
6. On-line procedures in closed systems[151] are superior to all manual batch procedures owing to the significantly lower risk of contamination and to their much better repeatability and reproducibility.

For preconcentration, different techniques, such as liquid–liquid extraction,[245,246] coprecipitation,[247] ion exchange and solid sorbent extraction[48] electrodeposition,[249] might be considered, but some specialities have to be taken into account with respect to their combination with AAS techniques.[250] For the combination of solvent extraction techniques with GFAAS, the following facts should be considered:

1. Although organic solvents have the advantage of enhancing the sensitivity in FAAS, they tend to be a disadvantage in GFAAS since they often wet graphite and can spread uncontrollably.[251]
2. Organic solvents containing halogens are a potential source of gas-phase interferences.[252,253]
3. Organic solvents tend to evaporate from the sample cups, and some of the complexes extracted are unstable, calling for a back extraction into aqueous media.

The most often used methods for solvent extraction of trace metals from seawater are based on the use of dithiocarbamates as chelating agents.[254] A standard procedure is based on the use of equal amounts of ammonium pyrrolidine dithiocarbamate (APDC) and diethylammonium diethyldithiocarbamate (DDDC) at pH 4–5, which is adjusted with an acetate buffer.[255–257] Freon TF (1,1,2-trichloro-1,2,2-trifluoroethane) is chosen as an organic solvent because of its low toxicity, its

Table 8 Selected publications on the direct determination of trace elements in seawater by GFAAS

Element	Modifier	Background correction	Refs.
Ag	Pd–ascorbic acid	Deuterium	221
As	Pd	Deuterium	222
As, Cd, Pb	Pd	Deuterium	223
Ba	V–Si	ZBC	224
Cd	NaOH	ZBC	225
	$(NH_4)_2HPO_4$–HNO_3	ZBC	218
Cd, Mn, Pb	Pd–ammonium oxalate	ZBC	226
	Pd–$NH_4H_2PO_4$–$Mg(NO_3)_2$	Deuterium	227
Cr	W	Deuterium	228, 215
Cr, Cu, Mn	HNO_3 or oxalic acid	ZBC	229
Cr, Mn, Mo	Pd–$NH_2OH{\cdot}H_2O$ for Cr, $Mg(NO_3)_2$ for Mn, Pd for Mo	ZBC	230
Cu	HNO_3	ZBC	231, 232
	NH_4NO_3	ZBC/deuterium	233
Cu, Mn	Pd, Pd–Mg	ZBC, THGA	234
Mn	NaOH	ZBC	235
Mo, V	Ascorbic acid	Deuterium	236
Ni	Pd	ZBC, THGA	217
Pb	Pd–Mg or oxalic acid	ZBC, THGA	237
Sn	Pd	Deuterium	238
Zn	NH_4VO_3	ZBC	239
	HNO_3 or ascorbic acid	ZBC	28

very low solubility in water (0.017% m/m at 21 °C), its quick separation from the aqueous phase and its low metal blank even at technical-grade quality.[256] Alternatively, the metal chelates may also be extracted into chloroform.[255,258] In order to overcome stability problems of the extract, the analytes are transferred back into acidic aqueous solution. This can be performed by either back extraction into dilute nitric acid[259] or by smooth evaporation of the solvent and acidic redissolution of the metals. Very high preconcentration factors of up to 6000 have been reported by such a combination of extraction and back extraction.[260]

Coprecipitation as a method for preconcentration has also often been combined with GFAAS determination, having its own advantages and disadvantages:

1. Carriers for coprecipitation are a potential source of contamination, since they are added in excess.
2. Coprecipitation carriers that can be easily vaporized or decomposed during the pyrolysis step are particularly suitable; these include elements such as indium, mercury, selenium[261] and tellurium[262] and organic reagents.[263–266]
3. Coprecipitation carriers, that act as modifiers, such as magnesium,[267,268] nickel[269] and palladium,[270] can also be advantageous.
4. The use of hydroxy precipitates[267,271–275] is preferred to that of sulfides,[276] which tend to create interference effects.
5. An interesting aspect of preconcentration by coprecipitation in combination with GFAAS is the direct analysis of the precipitate as the solid or a slurry.[263,277,278]

The use of ion exchangers and solid sorbents for preconcentration has found particular attention recently:

1. Column-based methods are particularly suitable for automation and even on-line methods are possible.
2. The sorbent can be used repeatedly, and often the amount of added reagent is reduced compared with other preconcentration methods.
3. Closed systems in general are easier to control for contamination. However, some ion-exchange resins and solid sorbent materials produce persistent and irreproducible blanks, owing to contamination of the polymeric material during its production.[279]
4. Organically bound metals might be lost during preconcentration, especially when using ion-exchange or solid-sorbent extraction, calling for a digestion procedure prior to preconcentration [e.g. ultraviolet (UV) photolysis].

Since ion exchangers seldom have the desired selectivity for the trace analytes, other solid-sorbent techniques have been developed, based on the use of more selective functional groups. At least two different approaches can be differentiated:

1. A chelating compound is added to the sample or preloaded on to the column. The column is often a hydrophobic (reversed-phase) material, which retains the chelating agent or the metal–chelate complex by sorption. Often used sorbents for this approach are silica C-18,[280,281] activated carbon[282] and macroporous resins such as Amberlite XAD-4.[283]
2. A functional group is permanently bound to the base material by a coupling reaction during the synthesis of the sorbent. Typical sorbents of this type are Chelex-100,[283] Muromac A-1[283,284] and 8-hydroxyquinoline bound to controlled-pore glass (CPG) or silica.[285]

A standard method using macroporous iminodiacetate chelating resin is available [EPA Method 200.13[286]] and can be used for the preconcentration of Cd, Co, Cu, Pb and Ni. Further details on solid-sorbent extraction techniques can be found in some review articles.[248,283]

On-line procedures especially for sorbent extraction with GFAAS detection, mostly based on FI, have been increasingly applied, using the same or similar sorbents as for off-line techniques. These on-line procedures offer a number of advantages:

1. The analysis time may be substantially reduced, since an equilibrium state is not needed.
2. There is a much lower risk of contamination since preconcentration and elution are performed in a closed flow system. Only inert materials and small surface areas are used, and the reagents can be purified on-line.
3. Since reactions such as acidification, complexing and sorption take place on-line and can follow each other in rapid sequence in a reproducible way, it is also possible to work with unstable complexes or intermediate products.
4. All processes can be easily automated.

On-line preconcentration procedures have made possible the determination of trace elements in the lower nanograms per liter range in seawater on a more or less routine basis. Commercial instrumentation including the necessary software, allowing for the coupling between FI and GFAAS, is available.[287] Proposed methods are based on the use of the reversed-phase sorption of chelate complexes with

dithiocarbamates[288–290] or dithiophosphates,[291] immobilized 8-hydroxyquinoline[292] or Muromac A-1.[293] An overview of such techniques was given by Fang and Tao[284] and progress has been reviewed by Fang.[294]

4.1.5 Methods for Speciation Analysis

Speciation analysis is an area which has found increasing interest in the 1990s. Element species can be distinguished by their oxidation state (redox species), by their molecular form (organometallic compound), by coordination (complex), by the phase in which the analyte exists (dissolved, colloidal, suspended) or by its solubility (leachability, bioavailability). In this broad sense, the analysis of filtered water samples is in principle already a form of speciation analysis. This physical separation can be refined to a fractionation of particulate and colloidal species via multistage filtration and ultrafiltration[295] and can be automated by means of FI-type manifolds[296] or field-flow fractionation. Methods used for such a type of speciation for water samples have been reviewed by De Mora and Harrison.[297] Considerably more detailed deliberations on the term 'species in aquatic systems' can be found in the monographs by Batley,[298] Krull,[299] Ure and Davidson[300] and Tessier and Turner.[301] In this section, however, the focus is on chemical species such as redox species and organometallic compounds. Very few direct methods exists for the determination of such species. More often, separation techniques have to be combined with suitable detection techniques. As far as speciation analysis is concerned, the FAAS technique lacks sensitivity, whereas the more sensitive ETAAS is not particularly suitable for detection in gas chromatography (GC), high-performance liquid chromatography (HPLC), and other chromatographic techniques most often used for the separation of trace element species. Therefore, AAS techniques can be ranked in the following decreasing order of their potential for speciation: HGAAS > CVAAS $\gg$ ETAAS > FAAS.

Since vapor generation techniques (HGAAS and CVAAS) are not the topic of this article, only speciation by ETAAS and FAAS techniques will be discussed briefly. Both techniques have been used in the past as detection methods for chromatographic separation techniques, as discussed by Ebdon and Hill,[302] but have been overtaken by more powerful detection techniques such as inductively coupled plasma mass spectrometry (ICPMS). Although FAAS and GFAAS are not very suitable to be coupled directly with chromatographic separation techniques, their combination with nonchromatographic separation methods is more straightforward, as we have discussed already for the case of preconcentration. Techniques such as solvent extraction or solid-sorbent extraction used for preconcentration can also be used for speciation purposes, for example, when the separation process is selective for one oxidation state of the analyte. Such combined methods, using GFAAS for speciation analysis, have been reviewed by Das and Chakraborty.[303] Often such nonchromatographic separation methods can be automated by using FI manifolds. When used for separation, FI provides only limited separation power, which mostly is exploited for binary separation of the type retained/not retained, for example, for preconcentration. While such a technique is not very well suited to separate a great number of analytes, it seems well fitted for screening purposes requiring high sample throughput. Further, for cases where the separation of species can be reduced to two, FI can be successfully used for speciation analysis. This is actually true for many cases of speciation of oxidation states of metal ions and metalloid ions, where the chemistry of these elements normally allows only for the presence of two different oxidation states in the presence of each other, at least under normal environmental conditions. Methods have been developed for the determination of As(III)–As(V),[304] Cr(III)–Cr(VI),[305–309] Fe(II)–Fe(III),[310] Sb(III)–Sb(V),[311] Se(IV)–Se(VI)[312] and V(IV)–V(VI).[313] The approach of binary separation can also be used for those cases where interest in the degree of speciation can be reduced to a binary differentiation. For example, such limited separation is useful for the differentiation of inorganic and organic species, labile and stable complexes, bound and free element species or the particulate and dissolved part of the analyte. This approach can also be applied for cases when other characteristics of the species can be grouped into two different categories, such as toxic/nontoxic. In this area of binary speciation, methods have been developed for the determination of free and complexed Cu[314] and of free and complexed Al.[315]

4.2 Soils, Sediments, Sludges, Coal, Dust and Other Solid Materials

A variety of widely differing solid materials will be discussed here together, simply because they have some common steps in their analytical procedure while differing in others. Solid materials such as arable soils, contaminated soils, sediments from rivers, lakes and seas, atmospheric dust and waste materials such as sewage sludge, ashes and other solid wastes are collected as samples from different compartments of the environment in order to gain information about the status of that compartment or to learn about the fluxes between them. The emphasis of environmental analysis is on the anthropogenic part of these fluxes and on the impact of such fluxes on the biosphere.

With respect to soils, in addition to the native content, various amounts of metals and metalloids are supplied

to the soil by dry and wet atmospheric deposition, by agronomic practices via fertilizers, sewage sludges or irrigation water, and so on and by waste discharges or waste disposal. For certain metals the anthropogenic proportion constitutes only a fraction of the natural amount, for others it can be the exclusive source. These metals and metalloids enter the food chain mainly via plant uptake.

Human activities are connected with the 'production' of waste materials including municipal solid wastes, industrial by-products, sewage sludge, dredged material, wastes from mining and smelting operations, filter residues from wastewater treatment and atmospheric emission control, ashes and slags from burning of coal and oil and from waste incinerators. All these wastes pose most challenging problems (e.g. when disposed of on land), creating a potential source for contamination of the soil and the groundwater.

The materials on the lake, stream and ocean floor, known as bottom sediments, are both a sink and a source of contaminants. Sediment-associated contaminants can be transported by resuspension of sediment particles, may accumulate in the food chain or may affect the health of biota and water quality in aquatic ecosystems. Knowledge of the character of the sediments and the fluxes involved is necessary for an understanding of environmental cycles but also in the development of techniques for remediation of aquatic ecosystems or the disposal of the sediments in the dredging of harbors and waterways.

With respect to atmospheric emission, monitoring frequently starts at the raw material stage. For example, emission monitoring of coal-fired power stations, coke works, incineration plants or cement furnaces starts with the analysis of the coal or waste materials to be burned. For the mass balance for such thermal reactors the different streams must be analyzed. While the mass carried with the scrubbed gases is important for the actual emission, the dusts collected in cyclones and electrostatic separators, ashes and other combustion residues such as slags must also be analyzed to characterize the complete system.

Air pollutants can occur in gaseous, fluid (dissolved or suspended) or solid form. Air must therefore be considered as an inhomogeneous, multiphase aerosol, whose composition is subjected to considerable spatial and temporal variations.[316] With respect to metals and metalloids, the gas phase is only of limited interest, since only mercury and a small number of volatile species occur as gases or vapors, especially close to hot emission sources.[317] The liquid phase (rain, snow, etc.), also termed wet deposition, belongs analytically to the area of water, which has been discussed in section 4.1. The analysis of air, with the restrictions mentioned already, is thus largely limited to the analysis of airborne dusts. Looked at globally, the bulk of airborne particles (approximately 90%) are derived from a number of natural sources (terrestrial, marine, volcanic, biogenic),[318] whereby salt from the oceans makes the largest contribution with about $10^9\,t\,a^{-1}$. Looking more specifically into metal emission, the view changes drastically, with a clear dominance of anthropogenic sources of the most important trace elements such as As, Cd, Cu, Cr, Pb and Zn on the global scale.[319] The anthropogenic sources of particle emissions (combustion, industrial, automobile) are by no means evenly distributed, but are related to the typical conurbations in which human activity is the most pronounced. Owing to the relatively high population and industrial density, such sources are concentrated in the northern hemisphere, but because of long-range transport globally these sources have an impact on a large proportion of the population in the form of pollution.[320]

4.2.1 Sampling of Solid Environmental Materials

Obviously, the procedures for sampling, sample homogenization and preparation of such widely differing materials are correspondingly manifold, and cannot be discussed in great detail here. Before one even starts to discuss analytical details, the exact purpose of the analysis should be clarified, since the type of information needed is decisive for the selection of the method. Three general procedures can be distinguished for these samples, determination of (a) the total concentration of an element, (b) the extractable portion and (c) individual species, especially organometallic compounds; the boundaries between these procedures are naturally flexible. The decision about which method to select is driven by the emphasis given to topics such as risk assessment, evaluation of bioavailability or toxicity, pollution control, baseline mapping and transport phenomena. The interested reader is referred to introductory monographs and review articles on the analysis of soils,[321–324] sediments,[325,326] sludges and other waste products,[327,328] and airborne particulates,[329] and on speciation analysis in soils and sediments[330–332] and to general treatises on environmental analytical chemistry, discussing such questions in more detail. The selection of the determination technique (FAAS, GFAAS, etc.) is simple in comparison. It is largely determined by the concentration of the analyte and also by the nature of the matrix.

4.2.1.1 Sampling of Soil, Sediments, Sludges and Solid Waste Materials The problem for sampling soils,[333–337] sediments,[338–341] sludges[328,335] and other waste materials[328,342,343] is the removal of a few grams that should be representative of the situation being investigated from a material that is present to the extent of tons or cubic meters. Meaningful sampling of such

heterogeneous media requires an understanding of the spatial and temporal variations, which is a prerequisite for answering the basic questions of how many samples are needed and where they should be taken. In the case of soils and sediments the planning of sampling has to start with a site investigation making use of available (e.g. historical) information, which could help to build hypotheses and conceptual models. A site characterization follows, consisting of two major steps, namely site observation and subsequent site mapping. Such site characterization is a multimedia, multidimensional, multidisciplinary effort, that requires vigorous communication among all of the principal parties involved and often includes preliminary sampling. Unfortunately, no checklists are capable of addressing all of the important issues and factors that influence the sampling and analysis of such materials. However, field guides and handbooks are available for soils[334] and sediments[344] that are helpful in defining the task. From all this information a sampling strategy is built, leading to a sampling network design. Depending on the required information, the sampling strategy must consider the local inhomogeneities in the horizontal plane of the sampling site, the dependence of the element distribution on the depth of sampling and possible seasonal variations.[167,345,346] A number of statistical procedures have been elaborated to facilitate representative sampling or to estimate the analytical uncertainty due to sampling errors, for example, in order to characterize contaminated areas.[347–352] Usually a large number of individual samples of the material being investigated are taken from different places (e.g. 25 samples from an agricultural field not bigger than 5 ha) and then combined into a composite sample. The advantages of composite sampling are that it reduces the cost of analysis and provides an estimate of the mean concentration. Its disadvantages are a loss of information about variation and the time required for such pretreatment. In order to avoid sampling errors[353] due to the heterogeneity of the material and the effect of particle size on element distribution,[354,355] one has to follow certain rules dictated by the particulate sampling theory, such as that developed by Gy.[356] In order to create an analytical subsample, the composite sample has to be homogenized and reduced in size. Different techniques for such homogenization exist, such as cone and quartering, sectorial splitting, riffling, milling and sieving. Sample preservation then depends on the analysis. In the case of total element determination the material is often dried and for other cases the samples are simply cooled to 4 °C or frozen (usually at −20 °C). The problems of sampling and sample preservation naturally become even more complex when element species are to be determined.[357,358] In any case, the uncertainty introduced by sampling of such heterogeneous materials supersedes all other following contributions,[339,359,360] and therefore needs careful evaluation in order to create meaningful data. It should be mentioned once again that this area is regulated by national and international rules, for example, for the sampling of soil (ISO-TC 190/SC2), sediments (DIN 38 414-11/1987) and sludges (DIN 38 414-1/1986).

4.2.1.2 Sampling Dusts Collecting the sample is the most critical step in the analysis of dust, as it is for other samples in environmental analytical chemistry. For the analyst, air is a practically infinitely expanded aerosol on which random investigations must be performed according to defined spatial and temporal regulations. The number and size of the random samples, as well as the place, time and sequence of sampling, are decisive for a representative statement on the relationship between quantity of material, space and time. A number of authors have treated the subject of sampling dusts very thoroughly in review articles and book contributions. This has been done either in general[361] or in selected areas such as the free atmosphere[316] or at the workplace.[362–364]

For the investigation of trace substances in the free atmosphere, the temporal and spatial distribution of sample collection in an area to be investigated must be matched to the meteorological conditions, such as wind direction, wind speed, air temperature and humidity. The length of an investigation program, the measurement frequency and the density of monitoring stations have to be carefully selected. Frequently, the concentration of constituents determined by a measurement program is influenced by the geographical location of the monitoring station. This is typical in areas where dust is whirled up strongly or close to the coast where the constituents of the sea spray become clearly noticeable. The extent to which meteorological conditions can influence sampling can be seen when the wind velocity is high, causing turbulence at the sampling port or even causing an under pressure so that the collected sample material can be transported back to the outside.

Dust cannot be sufficiently characterized by only its general composition, but physical and chemical forms (species) have to be differentiated;[365,366] the morphological composition of dust can be very heterogeneous depending on the origin of the particles[364,367] and the various forms can have markedly varying physiological activity. Smooth globular grains have a much smaller surface area, and are thus less reactive than needle-shaped, fine crystalline particles. Apart from the morphology, the size of particles, ranging between 0.001 and 150 μm, also has a decisive influence on the physiological activity. Particles smaller than 10 μm can enter the lungs of humans or animals, while larger particles can be filtered out. Dust samples often show a particle size distribution with characteristic maxima at two or three different particle sizes (bimodal, trimodal size distribution) depending on

sources and history. A further criterion is the solubility, which decides whether an element can directly enter the nutritional chain, or whether it is initially sedimented.

For sampling aerosols, two different approaches are available: procedures that use suction and those that do not. For the latter, various collectors for dust precipitates are available, which again are often regulated by national standards. In general, such devices do collect preferentially larger particles since these fall out more easily than smaller particles.

For sampling under suction, measurement and maintenance of a constant volume stream by the air pump are essential components of sampling. The sampling method most widely used is the high-volume sampler, which has been standardized in many countries, and operates at sampling rates between 40 and 100 $m^3 h^{-1}$. For sampling gaseous constituents, wash bottles with a suitable absorber solution or tubes being coated with a specific agent, so-called denuders, are frequently used. To separate dust from the air, flat membrane filters made of cellulose materials are most suitable owing to their low and constant element blank values. In high-volume samplers large filters can be used, which later can be cut to several aliquots that can be used for (a) various types of analysis, (b) replicates, (c) sharing samples with other groups or (d) storing them as 'backup' samples. Fiberglass filters do allow a larger volume of air to be taken through, but are less useful for trace analysis of airborne dust because of their high and often markedly varying element blank values.[368] Preseparating inlet systems are available for such samplers, which exclude particles larger than 10 μm, for example, in a cyclone. Cascade impactors most often are used for size separating sampling, which are available with 4–9 stages having size ranges between 0.02 and 20 μm.[361,369,370]

Sampling for emission measurements is usually performed directly in the stack or chimney, for example, of a power station or a waste incineration plant. The difficulties to be expected in taking a sample representative of the total mass flow under isokinetic conditions can be seen from the example of a stack gas stream of several 100 000 $m^3 h^{-1}$ at a temperature of up to 300 °C and with a varying dust content from milligrams to grams per cubic meter. For trace analysis, the problems of the level and scatter of the blank value and also contamination during sampling can additionally be detrimental to the results of the examination. This is more so the case for emission measurements since cellulose filters cannot be used, owing to the high temperatures at the sampling point, and quartz wool must preferentially be used.

For the measurement of dust at the workplace two different strategies exist, using either stationary or personal dust measuring instruments. Stationary measuring instruments are similar to those used for measuring in the free air and have the advantage that the air throughput is relatively high at about 20 $m^3 h^{-1}$, permitting a sampling period of only a few hours. The relatively small personal measuring instruments, on the other hand, have an air throughput of only about 120 $L h^{-1}$ (corresponding approximately to the volume of air breathed), so that sampling is required for the entire length of a working day to collect sufficient material for the subsequent chemical analysis.[362,364,371] Membrane filters made of cellulose materials are used preferentially for dust measurements at the workplace.

In addition to the collectors mentioned so far, which are largely standardized, there is a wide range of collectors designed for special applications, such as the investigation of automobile exhaust gases[372] or soot from Diesel engines,[373] or those that are specific for the subsequent analytical technique. Among these in particular are those developed for GFAAS procedures in which a probe[374] made of porous graphite is used as a filter and inserted directly into the atomizer, or the inner wall of the graphite tube itself is used for collection of aerosols either by impaction or by electrostatic deposition.[375,376] Owing to the high sensitivity of such in situ sampling techniques, very short collection times are possible so that the measurement can be performed virtually in real time. The disadvantages are that only one element can be determined per sample collection, except in simultaneous spectrometers, and that no sample is available for further evaluation. In general, procedures of this nature must therefore be seen as complementary to the standardized collection procedures and not as alternatives. A list of methods approved by United States Environmental Protection Agency (USEPA) was given by Kao,[377] but those for metals are few, however.

4.2.2 Determining the Total Concentration of Analytes

4.2.2.1 Total Dissolution For the determination of the total analyte concentration in a test sample, very harsh decomposition procedures are required, which are aimed at the total dissolution of the sample. Soil and dust samples exhibit considerable variations in mineralogical and chemical composition, including very resistant materials, making the task of total dissolution a difficult one. The total dissolution procedures available for such samples are either fusion methods using sodium carbonate[378–380] or lithium metaborate,[379–383] or acid digestions in the presence of hydrofluoric acid and frequently also perchloric acid.[380,384–386] Although such harsh procedures are time-consuming and the use of such dangerous acids demands special safety measures, they cannot ascertain the total dissolution of a number of minerals, such as cassiterite, chromite, zircon, rutile, monazite and corundum.[387–389] Apart from incomplete

dissolution, formation of volatile fluorides is another possibility for analyte loss for a number of elements such as antimony, arsenic, boron, germanium and silicon (which should be removed in this way).

Since fusion methods lead to high total salt concentration with a corresponding risk of contamination and interferences in GFAAS due to the high background signal, wet digestion methods are more popular in connection with AAS. However, often the difficult matrix cannot be dissolved by a single-stage dissolution, but calls for a combined multistage digestion using oxidizing acids (nitric acid, perchloric acid, sulfuric acid, hydrogen peroxide) and hydrofluoric acid.[386,390] If necessary, dry ashing can be performed in advance[391,392] and if insoluble, fluoride-containing residues are produced during digestion, they can be taken back into solution by the addition of boric acid.[390] A comparison study concluded that the combination of an acid decomposition with final fusion of insoluble residues is the best method for the hard to extract elements Cr, Y and Zr.[378]

Membrane filters normally used nowadays in dust collectors dissolve in nitric acid and also in organic solvents such as benzene.[371] On the other hand, quartz wool used as a filter medium for emission measurements usually requires digestion in nitric acid or hydrofluoric acid under the addition of hydrogen peroxide.[393]

For the digestion of coal and coal ash often a pressure digestion with hydrofluoric acid–nitric acid[394,395] is used, preferentially with the addition of a strong oxidant such as perchloric acid or ammonium peroxodisulfate.[392,394] Less common is a digestion with perchloric acid (72%) alone, which requires a good deal of attention.[396] The combustion of coal in an oxygen bomb[397] or oxygen combustion[398] and fusion with lithium metaborate[399] have also been described. In a review article, Mills and Belcher[400] discussed in detail the problems of sample pretreatment and digestion for the analysis of coal and ash.

4.2.2.2 Partial Dissolution In general, total dissolution is a difficult task for many samples and, once achieved, leads to high salt concentrations. The latter may pose problems with nebulization in FAAS or create strong background attenuation in GFAAS. The high matrix concentrations require regular blank controls to be performed, and frequently the calibration solutions must be matched to the matrix contents. Owing to such problems, it is very common to avoid 'total dissolution' in environmental analysis. Instead, an extraction with a strong mineral acid such as boiling aqua regia under reflux is preferred.[401–406] This is justified on the assumption that heavy metals insoluble in aqua regia will probably not be relevant for environmental chemistry, at least not in the short term. With the exception of a few cases, a valid approach seems to be that the aqua regia-resistant compounds will not be taken up by plants and cannot be dissolved by water or bacteria.[405]

4.2.2.3 The Microwave Oven Option Microwave-assisted digestions (MWDs) have found rapid acceptance in the environmental analysis of solid samples.[407–409] The major advantages of MWDs are reported to be the substantially shorter digestion times,[410,411] the lower risks of contamination, the reduced reagent requirements[412] and the lower risk of losses.[413] Different acids and acid mixtures have been investigated, mostly using nitric acid or mixtures of hydrochloric acid and nitric acid[414–416] or aqua regia,[417–419] but also including hydrofluoric acid[415,420–423] and perchloric acid.[424] While the careful selection of the acid mixture and the operational parameters was a prerequisite for a successful digestion until recently,[425] the development of systems with automatic pressure and/or temperature control[426–428] have significantly reduced such requirements. Owing to the high effectiveness of MWDs it has also been possible to incorporate them on-line in a flow system with FAAS[429–431] or GFAAS[432] detection.

For the actual determination by AAS, we should always remember the general rule that the best results can be expected when the optimum concentration range is used. This means that FAAS is only suitable for the determination of higher analyte contents, and not for trace analysis. Owing to the frequently complex and refractory matrix, particularly when the digestion is complete, the hottest possible flame should be used in general and the measurement should be performed as high above the burner slot as possible. In other words, the measurement should be optimized for minimum interferences and not for maximum sensitivity. Aqua regia extracts can be analyzed by FAAS relatively free of interferences, since the element concentrations normally to be expected in soil, sediment and dust samples are so high for numerous analytes that the sensitivity of FAAS is fully adequate. Some selected publications of such applications are given in Table 9.

The alternative way of avoiding interferences is higher dilution followed by GFAAS determination. The concentrations of elements in soils, sediments and sludges which are considered to be tolerable depend strongly on regional regulations, but in general they are between 1 and 300 $mg\,kg^{-1}$ for agricultural soils and about a factor of 10 higher for sludges.[455] Owing to the dilution introduced by the acid extraction, such soil concentrations produce solution concentrations between 10 and 3000 $\mu g\,L^{-1}$. Even one tenth of the tolerable concentration is still above the detection limit of GFAAS for most elements. The partial dissolution and the high dilution factor ensure that the matrix concentrations in the sample solution are often

Table 9 Selected publications on the determination of metals and metalloids in soils, sediments, sludges, dust and solid waste by FAAS

Analyte	Test sample	Remarks[a]	Refs.
Ag	Soils		433
Al, Ca, Fe, K, Mg, Na, Si	Coal, fly ash	$Li_2B_4O_7$ fusion	434
Ba, Be, Cr, Cu, Li, Mn, Ni, Pb, Sr, V, Zn	Coal, coke, fly ash	Australian Standard Method AS-1038	435
Be, Ca, Cd, Co, Cu, K, Li, Mg, Mn, Ni	Coal	Pressure digestion, $HNO_3-HF-H_3BO_3$	436
Ca	Soils	Addition of lanthanum	437
Cd, Co, Li, Pb	Coal, ash	Dry ashing/pressure digestion, $HNO_3-HF-HClO_4$	392
Cd, Cu, Fe, Mn, Pb, Zn	Sewage sludge	Microwave-assisted extraction	438
Cd, Cu, Pb	Soils, sediments	FI on-line sorbent extraction with DDTC/silica C-18	439
Cd, Cu, Pb, Zn	Sediments (lake)	Extraction overnight with $HCl-HNO_3$	440
Cd, Pb	Soils	Background interference by iron	23
Cd, Pb, Hg, As	Soils, waste	FI on-line dilution	441
Co	Sewage sludge	German standard method DIN 38 406-E24	442
Cr	Soils, sediments	Interferences, optimization of flame conditions	443–445
Cr	Sediments	Digestion problems	446
Cu	Sewage sludge	German standard method DIN 38 406-E7	447
Cu, Cr, Ni, Pb, Zn	Sewage sludge	Dry ashing at 450 °C	448
Cu, Mn, Pb	Sewage sludge	FI microwave-assisted on-line digestion	429, 449, 450
Cu, Zn	Soils	Microwave-assisted pressure digestion	411
K	Sewage sludge	German standard method DIN 38 406-E13	451
Mn	Soils, silicate materials	Digestion, $HF-HClO_4$	452
Na, K, Ca, Mg	Soils	FI on-line dilution	453
Ni	Sewage sludge	German standard method DIN 38 406-E11	454

[a] DDTC, diethyldithiocarbamate.

very low, so that the matrix remaining in solution might be considered relatively simple. Because of this, method development can often be focused towards fast analysis and high sample throughput.[466,467] Nevertheless, STPF conditions and ZBC frequently have to be used.[423,456–465] In any case, the absence of interferences should be checked, for example, by recovery studies using reference materials. In the meantime, reference materials have been produced for which, in addition to the certified total concentration, information is also provided on the fraction that can be leached by aqua regia, for example, soil samples NIST SRM-2709[468] and SRM-2711,[469] Bureau Communautaire de Référence (BCR) CRM 141R,[470] 142R[471] and 143R[472] and sewage sludge samples BCR CRM 144R,[473] 145R[474] and 146R.[475]

4.2.2.4 Direct Analysis of Solid Samples and Slurries

As has already been mentioned, the graphite furnace can be used as a thermochemical reactor, performing in situ digestion. Procedures have been developed that allow for the determination of the total concentration of an element, either by direct solids analysis[476,477] or by slurry sampling.[478,479] In addition to the benefit of not having to perform a digestion (especially when the digestion is really problematic, e.g. for coal), the major reasons for the application of these techniques are the much lower sample requirement and the increased sensitivity. The application of the STPF concept and ZBC are quasi prerequisites for accurate analyses,[476,480,481] because the matrix often has a substantial influence on the appearance time and the signal profile.[482] A number of authors have attempted to simplify the introduction of the solid dust samples by punching small disks from the filter and analyzing directly.[481,483] The precision of this procedure is nevertheless poorer than when the filter is homogenized in a mill.[484]

In contrast to direct solids analysis, the analysis of slurries is more flexible, and precision and trueness are better for the analysis of slurries in comparison with solids,[485,486] especially when they are homogenized by an ultrasonic probe directly mounted to the autosampler of the GFAAS instrument.[487–491] Further improvements, such as a reduction of the background attenuation, can be obtained by in situ ashing with air or oxygen[492,493] or the addition of hydrogen to the purge gas.[494] Aqueous calibration solutions can often be employed in combination with matrix modification. Special attention has to be paid to particle size and density effects, which might call for particle size reduction, which could then reduce the benefit of reduced sample preparation owing

to the absence of a predigestion. For the same reason, the atomization of slurry samples by FAAS, which has been performed with sample introduction by FI,[450,495,496] has not attracted great attention owing to the restriction to very fine particles or analytes that can be easily extracted into the aqueous phase during the preparation of the suspension.

4.2.3 Selective Extraction of Individual Constituents and Bonding Forms

The determination of the total concentrations of analytes cannot give the right answer to many environmental questions, such as evaluation of the exposure of plants and animals to pollutants, evaluation of the fertility of agricultural soil or risk assessment for a waste dump site. Rather than the total concentrations of elements, the part which is bioavailable to plants and animals, the part which is leachable and mobile and thus can move away from the site and reach the aquifier and so on is of interest for such studies. For this type of element speciation, leaching procedures using a variety of differing reagents, either in a single step or in a sequential extraction scheme, have been used successfully for decades. The choice of extractant will depend not only on the element in question but also on the soil type, the crop or animal species, the weather and the technique for determination. The list of extractants which have been used is correspondingly long and sequential extraction schemes using a set of extractants are also numerous, varying between three and nine steps for differentiation. Descriptions and comparisons of such procedures can be found in review articles which have been published recently for soils and sediments by Rauret,[497] Jones[498] and Ure et al.[499] and for waste materials by Van der Sloot.[500] The method developed by Tessier et al.[501] has been applied to all the materials discussed here and is designed to differentiate between five fractions:

1. *Exchangeable fraction*. Trace metals adsorbed at the surface of particles, which can be desorbed by simply changing the ionic composition.
2. *Carbonate fraction*. Trace metals associated with carbonates, which are susceptible to pH changes.
3. *Fraction bound to Fe/Mn oxides*. Iron and manganese oxides are present as nodules, concretions, cement between particles or as a coating on particles. These oxides are excellent scavengers for trace metals and are thermodynamically unstable under anoxic conditions.
4. *Fraction bound to organic matter*. Trace metals bound to organic matter, such as organisms, detritus, coatings on mineral particles, etc. through complexation, peptization or bioaccumulation. These metals can be released by degradation of the organic matter under oxidizing conditions.
5. *Residual fraction*. The residual solid, having resisted the removal of the previous fractions, contains trace metals in the crystal lattice structure. It is believed that normal environmental conditions do not release these metals.

In sequential extraction, the bioavailability of metals is meant to decrease with each step of the scheme, but it is generally recognized that most extraction schemes are less than perfect. Few extractants can be relied on to release elements solely from a particular phase[502–504] and redistribution between phases can occur during the sequential procedure.[505,506] Also, the results will depend on the thermal pretreatment of samples.[507] Also, such operational defined methods do not help any fundamental understanding of the chemistry involved but offer only empirical guidance.[508]

Despite these limitations, they are often specific enough to provide the information necessary for making diagnostic or managerial decisions. However, the different extractants and the numerous extraction schemes used significantly hinder the comparability of results between different studies. In order to overcome such problems, harmonized procedures have been developed, such as those under the auspices of the BCR,[509–511] for the single extraction with ethylenediaminetetraacetic acid (EDTA) or acetic acid and a sequential extraction scheme using three steps, which is given in Table 10. The use of microwave heating for speeding up the extraction procedure[512] can also lead to improved reproducibility. In addition, reference materials are now available for which not only the total element content but also extractable portions are certified.[513–516]

For the determination of the element by AAS, it is of no great consequence whether the sample material is a soil, a sediment, a sludge or a waste material. Since

Table 10 BCR sequential extraction scheme

Extraction	Extractant	(Nominal) phase extracted
Step 1	0.11 mol L^{-1} acetic acid	Exchangeable
Step 2	0.1 mol L^{-1} $NH_2OH{\cdot}HCl$, pH 2	Reducible, Fe/Mn oxide
Step 3	8.8 mol L^{-1} H_2O_2, then 1 mol L^{-1} CH_3COONH_4, pH 2	Oxidizable – organic and sulfide

only relatively small element concentrations are extracted during the leaching, especially with weak extractants, FAAS is only sufficiently sensitive for a few elements, such as Al, Ca, Co, Cu, Fe, K, Li, Mg, Mn, Na, Sr, Zn.[517–522] This situation might be different for airborne particulates, for which the extractable concentration of some more metals (e.g. Cd, Cr, Ni, Pb) might be sufficient for their determination by FAAS.[365,523–525] However, in many cases for elements such as Ag, As, Cd, Cr, Mo, Ni, Pb, Se, Sb, Sn, Te, Tl, V, it is usually necessary to apply GFAAS[519,520,526–531] or other more sensitive techniques. Owing to the considerably lower total salt concentration, the matrix influences are far less pronounced than with a complete digestion, at least for those extractants that do not add significantly to the total salt concentration.

4.2.4 Speciation Analysis in Soils, Sediments, Sludges, Dust and Other Solid Samples

The determination of redox species and organometallic compounds in solid materials, such as soils, sediments and sludges, is preceded by an extraction procedure, and in this respect is similar to the determination of bioavailable species. The determination of the individual species from the extract is hardly different from the analysis of water. In rare cases where FAAS is sensitive enough for speciation analysis, the separation procedure for the species is coupled with preconcentration, for example, for the determination of Cr(VI).[532–534] The use of GFAAS for speciation analysis has been reviewed by Das and Chakraborty.[303] As has already been pointed out, GFAAS is not a very suitable technique to be coupled with chromatographic separation since all attempts to run GFAAS in a continuous mode were unsuccessful,[535] and coupling between these two techniques[536] has to be done via either a fraction collector or a flow-through subsampling interface.[537] These interfaces require careful synchronization[538] and optimization of the flow rates and sampling intervals,[539] which significantly reduces its applicability. Other techniques, which can be operated continuously, such as vapor generation techniques (CVAAS, HGAAS) or plasma techniques, such as ICPMS, are much better suited.

For those cases where GFAAS has been used for speciation analysis of soils, sediments, sludges and dust samples, mostly nonchromatographic separation procedures have been used. These include the extraction of Cr(VI) in a weakly alkaline medium,[67,540–542] the extraction of tributyltin oxide with *n*-hexane–dichloromethane[543,544] or isooctane[545] and its back extraction with nitric acid.[543,544] The selective extraction of chelates has been applied for the species determination of As(III),[546] Sb(III)[547] and Cr(VI).[548]

The determination of Cr(VI) in welding fumes is one of the most important industrial hygiene examinations owing to the toxicity of this species.[362,549] The influence of the collecting procedure, the welding fumes themselves, the storage period and the analytical procedure have been investigated in detail.[550] While the Cr(VI) content of fiberglass filters did not appear to change over several weeks of storage, Rohling and Neidhart[551] found that during collection Cr(VI) is reduced by SO_2, which can lead to severe losses of Cr(VI), depending on the length of the collection period. Girard and Hubert[552] showed that the Cr(VI) can be quantitatively recovered from the filter medium by extraction with acetic acid–sodium acetate. Brescianini et al.[553] eliminated the interferences caused by large quantities of calcium, iron and sodium during the determination of chromium by GFAAS by separating the analyte on an ion exchanger.

Moreover, ion exchangers, chelating resins and other solid sorbents have been utilized in batch or column procedures to distinguish between different arsenic species such as As(III) and As(V),[554] Cr(III) and Cr(VI)[555] and organotin compounds.[556]

4.3 Biological Materials

Biological materials are investigated not only for environmental analysis, but also for reasons of biological, biochemical or biomedical analysis, toxicological, pharmaceutical, nutritional and clinical analysis, food analysis or biogeochemical analysis; the information sources available about sampling and analysis of such materials are correspondingly spread over a wide range of disciplines. For the purpose of environmental analysis, it is seldom the biological material itself which is the primary target of interest (as in toxicological or biomedical analysis), but its function as an indicator for the environmental compartment from which it was taken. In this function, biological materials are collected from different compartments of the environment. For the same reasons, other inorganic materials are collected, to gain information about the conditions of that particular compartment with respect to being part of a hostile biosphere or to learn about the exchange and fluxes of materials (pollutants) between different compartments or different parts of a compartment. With respect to biological materials, the primary concern is the entrance of pollutants into the food chain and the concentration of such pollutants by bioaccumulation. For plants and vegetable products, sources of pollutants are the soil on which they are grown, fertilizers, crop protection agents, insecticides and pesticides used and atmospheric precipitation, especially due to the proximity of roads or industrial installations. Animals (and humans) are influenced by the environmental conditions either directly (e.g. aquatic environment) or indirectly by

taking up pollutants (e.g. with their nutrition or drinking water) from that environment.

Those biological objects and systems which respond very sensitively to anthropogenic influences on the environment can be used as bioindicators for the detection of the presence of pollutants and biomonitoring can be used as a more quantitative method for the determination of the pollutants present. For example, low-level plants such as lichens and mosses accumulate large amounts of heavy metals, because their unprotected surfaces provide ion-exchange facilities which allow the effective adsorption of metals. The complexes then formed between the metals and the negatively charged organic groups are fairly stable. The accumulation of heavy metals on such plants is the basis for their use as bioindicators of atmospheric pollution. The problem with biological materials as samples for environmental analysis is biological variation, which hampers the assessment of baseline concentrations and the interpretation of analytical results.

The analysis of biological materials by AAS has a long history. While the reasons for analyzing biological materials are numerous, the different objectives are secondary for the selection of methods and techniques applied to the determination of elements in these materials by AAS. Review articles have been published by Slavin on the use of FAAS[557] and GFAAS[558] for the analysis of biological materials in general, by Hanlon[559] on plant analysis, by Rains[560] on the subject of foodstuffs analysis, by Pritchard and Lee[561] on agricultural samples and by Hoenig and Guns[48] on environmental and biological samples. Apart from review articles on single elements already mentioned (see section 4), numerous other minerals and trace elements are treated in the monographs *Quantitative Trace Analysis of Biological Materials* edited by McKenzie and Smythe,[562] *Trace Minerals in Foods* edited by Smith,[563] *Trace Element Analysis of Biological Specimens* edited by Herber[564] and *AAS in Occupational and Environmental Health Practice* written by Tsalev.[565] Markert gives an overview of plant analysis[566] and the use of bioindicators[567] and Kiceniuk and Ray treat the analysis of contaminants in edible aquatic resources.[568] Annual reviews on the determination of elements in biological materials appear in the *Journal of Analytical Atomic Spectrometry*.[569]

4.3.1 Sampling and Sample Pretreatment of Biological Materials

Again, an analyst should not waste time in investigating a laboratory sample whose history is not known in complete detail and is not documented. Sampling is of decisive significance for the accuracy of the analyses of biological materials. Some problems are common with the sampling of other sample materials, for example, selection of appropriate locations, techniques and designs, errors due to the limited numbers of samples and contamination due to the tools used. However, the main problems are the high biological variability and biotic and abiotic influences on the availability of the sample material and on the reproducibility of its analysis. For example, such influences are as follows:

1. the exposure of the biological species and its behavior with respect to accumulation;
2. the time of sampling with respect to the season and to the phenological state of development of the biological species;
3. fluctuations or trends in the abundance or composition depending on age, gender or physiological state or morphological and genetic variability;
4. fluctuations between sampling locations with respect to edaphic or microclimate factors and exposure to emission;
5. errors due to a selective sampling procedure, favoring individuals or groups, having special characteristics.

For the sampling of plant materials, this means that one has to decide which parts of the plant to sample (roots, leaves, needles, etc.), their location with respect to the soil (distance from the ground) and their exposure (free-standing plant, plant from a forest ecosystem).[570] Furthermore, the time of exposure depends on the age of that particular sample, which must be known (e.g. to create comparable results when sampling tree needles of different ages).[571] Depending on the questions to be answered by the investigation, one has to decide whether external contamination of the plant material by soil and precipitation has to be removed before analysis or not.[572,573] If it has to be removed, the cleaning procedure must be designed to remove only the external contamination, and to avoid leaching the plant components.[574] If external contaminants should not be removed, the sampling procedure has to assure the retention of those components.

When sampling biological materials of animal origin, a practical method of capture, not prone to contamination, and in accordance with ethical rules, is a prerequisite for using that species for environmental analysis. Depending on the type of sample (body fluid, soft tissue, hard tissue) and whether it can be collected without killing the animal (e.g. bird feathers, egg shells, body fluids), one has to account for changes in element distributions due to stress or postmortem changes (autolysis, etc.).[575] Contamination risks are especially high when internal organs have to be taken by dissection or when blood samples have to be analyzed.

These and other sampling and presampling factors cannot be discussed in detail here. Comprehensive information on this subject has been published by Jones and Case,[576] Keith[577] and Markert.[578] In the laboratory, the sample must be homogenized, conserved and typically stored.[579] Naturally, all precautions to prevent contamination from tools, containers, the laboratory atmosphere and losses due to volatilization or adsorption must be scrupulously observed. For the homogenization of plant and animal tissue samples, cutting and mixing tools that have stainless-steel parts are often used. Contamination can be expected from such tools, especially for Fe, Cr and Ni.[580]

4.3.2 Digestion Procedures

Plants and biological tissues are usually decomposed for the determination of total element concentrations by AAS techniques, although they can be determined also by direct techniques using solid-sample introduction or slurry sampling. The main reason for using routine decomposition techniques is concerns about the homogeneity of sample masses as small as a few micrograms. In addition, the samples often have to be prepared as powders and such pretreatment may take as long as a fast decomposition technique.

For the decomposition of various biological materials, a wide selection of methods have been developed, including dry ashing, wet digestion, pressure digestion and MWD. The selection of an appropriate digestion procedure depends on the analyte, the sample matrix and the method of determination. In general, atomic spectrometric techniques do have a much higher tolerance against residual carbon after the digestion than other instrumental techniques such as electrochemical methods, since residual matrix components can be destroyed during atomization. The amount of residual carbon which can be tolerated by different AAS techniques decreases in the order FAAS > GFAAS $\gg$ HGAAS/CVAAS. Digestion solutions, which are clear and free from particles in general, fulfill the requirements for FAAS.

In the 1960s and 1970s, dry ashing was widely used, since relatively large sample aliquots can be ashed (up to 20 g) and taken up in a minimum volume of acids. In this way, many trace elements could be determined by FAAS, which was the only technique available at that time. Dry ashing can be performed directly or with the addition of ashing aids such as magnesium nitrate or small amounts of nitric acid and sulfuric acid, usually at 450–550 °C. Accurate temperature control, often not available with simple muffle furnaces, is a prerequisite for complete destruction of the organic material without loss of analytes.[581] The ash is then leached with dilute hydrochloric acid or nitric acid. Care is needed to make sure that losses due to siliceous ash residue or by adsorption on the wall of the crucible are avoided. The addition of hydrofluoric acid might be helpful to avoid such residues[582] while the choice of an adequate crucible material is of prime importance to avoid adsorption losses. Since the most important retention losses have been reported for silica and porcelain vessels, platinum vessels are generally recommended. Losses have been observed also for volatile elements, such as Hg, As, Se and Cd, but often can be circumvented by the addition of ashing aids. High blank values, which are difficult to control, at least when standard muffle furnaces are used instead of special furnaces (e.g. equipped with quartz walls), often limit the applicability for samples with low trace metal contents. Nowadays dry ashing is only applied sporadically, since other digestion procedures are faster and better.

Oxidative wet digestions are used far more frequently than dry ashing or fusions for subsequent analysis by FAAS or GFAAS. These can be performed in open vessels, with reflux condensation if required, or in closed containers under pressure (autoclaves). Open vessel digestions have a number of disadvantages, which should exclude them from the list of recommended procedures, especially (a) low sample throughput and (b) incomplete digestion due to low reaction temperatures, (c) high consumption of reagents creating high costs, large amounts of waste and environmental pollution by fumes, (d) high risk of contamination due to the amount of reagents and the exposure to the laboratory environment and (e) the possibility of analyte losses due to volatilization. In order to reach the necessary oxidation power with open-vessel digestions, strongly oxidizing, high-boiling acids such as perchloric acid, sulfuric acid and phosphoric acid are often used as digestants. These acids are more difficult to clean up than lower boiling acids and are thus more strongly contaminated. Further, the use of such acids cannot be recommended in combination with determination by GFAAS for many reasons: these acids (a) penetrate the graphite lattice and often remain in the tube after pyrolysis, leading to (b) reduction of the lifetime of the graphite tube and to (c) background interferences by molecular absorption. In the case of perchloric acid, volatile chlorides can be formed, leading to analyte loss during pyrolysis and/or atomization especially for main group III elements (aluminum, gallium, thallium).[583–585] Even in FAAS these acids create some problems, either by reducing the uptake rate of the nebulizer because of their high viscosity or by other interference effects.[586] By using low-boiling mineral acids, complete destruction of the organic material is not often possible by open digestion. However, it has frequently been shown that many elements can be extracted from biological materials, even with diluted mineral acid, without complete destruction of the organic material.[587] In any case, such methods

cannot be generalized and have to be verified for the matrix being investigated. In general, animal materials cause fewer problems than plant materials, and elements such as Ca, Cd, Cu, K, Mg, Mn, Na, Pb and Zn can be more easily extracted than Al, Cr, Fe, P and Si.

While the application of high-boiling acids for AAS is problematic, it is the oxidation power which is needed to decompose totally the complex organic matrix of some of the environmental samples. Of the mineral acids that are compatible with AAS (e.g. HNO_3, H_2O_2, HCl and HF), only the first two develop sufficient oxidation power but, because of their low boiling points, the oxidation power is not enough for complex matrix-containing substances, which are difficult to digest, such as fat. In order to raise the boiling point of these acids and therefore their oxidation power, the wet digestion procedure can be performed under pressure. There are many advantages of a pressure dissolution compared with open acid dissolution.[588] Since the pressurized decomposition vessel represents a closed system, the loss of volatile elements is avoided, in addition to the risk of contamination from the laboratory environment. Depending on the pressure which the system can withstand, either temperatures in the region of 200 °C (medium pressure, ≈80 bar) or >300 °C (high pressure, ≈130 bar) are possible. Under such harsh conditions, the total decomposition of complex matrices is possible, and organic materials can be decomposed with very little residual carbon remaining in the digest (<1%). The disadvantage of pressure digestion in autoclaves ('bombs') is the small amount of organic matter which can be treated. The special problem with biological materials is that large amounts of gaseous products (CO_2) are formed during the digestion. As a result, the maximum sample mass that can be loaded into such pressure vessels is limited to about 200–500 mg for a vessel volume of about 100 mL for medium-pressure systems and about 1 g for high-pressure systems. Using a medium-pressure system, operated at 170–180 °C, the optimum digestion time for biological materials with nitric acid is about 3 h.[589] For a high-pressure digestion system, the effective digestion time can often be reduced to about 1.5 h. About 2 mL of nitric acid are generally sufficient for a pure carbon content of 100 mg, independently of the type of biological sample. For determining the maximum sample mass, one has to consider that the carbon content of dried biological material varies between 35 and 50% for plant materials and between about 45 and 55% for animal materials with the exception of oils and fats that have carbon contents between 74 and 78%. Nitric acid alone is sufficient for complete digestion under high pressure for most elements and biological sample types.[589] Some samples (e.g. some plant samples, mussels) contain high amounts of silicate, requiring the addition of hydrofluoric acid. For more details, introductory papers and review articles on pressure digestion should be consulted.[588,590]

Since the end of the 1980s, these various digestion techniques have been increasingly replaced by microwave-assisted acid digestions which are frequently performed in closed containers and under low (up to 40 bar) to medium pressure (up to 80 bar).[408,591] The major advantages compared with conventional digestion techniques include the reduced time and reagent requirements, the lower blank values, the reduced risk of analyte losses, the possibility of automation and the ability to digest a number of samples at the same time. When comparing the time required for a conventional pressure digestion (CPD) and for a MWD, one has to consider that the digestion time is drastically reduced in a microwave system. This is shown in Table 11.

The most important difference is the fast heating for MWD. In addition to the time advantage for MWD, there is also an advantage with respect to the wear of materials. The digestion vessels in CPD are under heat and pressure stress for a much longer time than the vessels in MWD systems. Therefore, the vessels in CPD systems wear out earlier, and contamination due to sorption/desorption processes are more pronounced than in MWD systems.

The time gain with MWD is further reduced when the digestion cannot be performed in a single step. This is the case when either the gases produced during the digestion must be vented or reagents have to be added after the first step in order to complete the digestion. Especially under low-pressure conditions, the oxidation power of nitric acid is not sufficient for all types of material. Materials rich in fat are especially difficult to decompose[592] and often require the addition of additional oxidants such as hydrogen peroxide.[593,594] Stronger oxidants such as sulfuric acid and perchloric acid are (besides their incompatibility with GFAAS) also not very useful in MWD. Typically, they develop their oxidizing power at such high temperatures (above 250 °C) that the polymeric material often used for the vessels breaks down. When a single-step digestion is the target for method development, the method requires careful optimization

Table 11 Comparison of the time involved in MWD and CPD (in minutes)

Preparation step	MWD	CPD
Preparation of the digestion (weighing of the sample material, adding the acids, capping the vessels)	10	10
Digestion period (heating)	10–20	180–210
Cooling period	5–10	20
Preparation of the determination	5–10	5–10
Cleaning of vessels	5–15	10–20
Total time	35–65	225–270

with respect to the matrix. Chemometric techniques have been applied for this purpose.[595] State-of-the-art MWD systems having temperature and/or pressure sensors for each vessel require much less method development and optimization, since the digestion can be automatically operated under optimum conditions.[426,427] Some systems use quartz vessels that can be used at temperatures above 200 °C and allow pressures up to 80 bar.

In addition to these widely used digestion procedures a number of techniques have been proposed that have not come into general use, but can occasionally be of interest as comparative techniques or for a special combination of analytes and matrices. A comparison of such techniques with respect to their characteristics is given in Table 12. Since FAAS and GFAAS methods do not require the highest digestion quality with respect to completeness of the destruction of organic materials, even the incorporation of MWDs into flow systems has been successfully performed for some biological materials. A review of such methods was presented by Burguera and Burguera.[596] Solubilization of materials with tetramethylammonium hydroxide (TMAH) has been used successfully, sometimes under microwave assistance,[597] especially in combination with speciation. Other less frequently used techniques are based on combustion in a stream of oxygen in a closed system[592,598] or on intense infrared (IR) irradiation. Details of these and other techniques can be found in treatises and review papers on sample digestion.[599,600]

4.3.3 Methods for Flame Atomic Absorption Spectrometry

In general, digestion of solid material is always combined with dilution. Under normal conditions, when using acid digestion, 200–500 mg of biological material are usually decomposed in about 5 mL of acid, the digest being then diluted to 10–20 mL. Hence the dilution factor is about 40 for such decomposition, if not higher. Following such dilution and remembering that a technique should only be applied in its optimum working range, FAAS is suitable without reservation only for the direct determination of major components such as potassium, sodium, calcium and magnesium and minor components such as copper, iron and zinc. In some cases manganese can also be determined by FAAS in biological materials. Under given conditions it is also possible to determine aluminum, chromium, nickel and silicon. The use of the nitrous oxide–acetylene flame is a prerequisite for most of these elements to obtain correct results. Table 13 presents a number of recent applications.

For other trace elements in biological materials, the sensitivity of FAAS is generally not sufficient. For such cases, preconcentration has often been used in combination with FAAS, especially before 1980. Since today more sensitive direct methods, using for example GFAAS, are available, such preconcentration methods cannot be recommended any longer.

These classical preconcentration techniques take on a fully new dimension when they are performed automatically and on-line in a closed FI system. The most often used methods of preconcentration with such FI/FAAS

Table 12 Digestion methods for biological materials[a]

Method	Contamination	Completeness	Throughput	Costs	Comment
Dry ashing (450–550 °C)	+++	++/+++	++	+	Usually less applicable for low contents and high precision
Combustion	++/+	+++	++	++	Low sample throughput
Open acid digestion, automated system	++/+ ++	++ ++	+++ ++++	++ ++	Medium to high sample throughput
Open microwave-assisted	++	++	++++	++	High sample throughput
Pressure digestion PTFE < 200 °C	++/+	++	++	++	Digestion quality mostly sufficient (at least for AAS)
Microwave-assisted pressure digestion, <200 °C	++/+	++/+	+++/++++	++	Digestion quality not always sufficient
Microwave-assisted pressure digestion, <250 °C, quartz vessel	+	++/+++	++	++	Digestion quality mostly sufficient (at least for AAS)
High-pressure digestion, >300 °C	+	++++	++	++++	Optimum digestion quality but costly
UV-photolysis	+	++	++	++	Restricted applicability, but possibility of on-line digestion in FI systems

[a] +, Small; ++, medium; +++, high; ++++, very high.

Table 13 Selected publications on the analysis of biological materials by FAAS

Analyte	Sample material	Remarks	Ref.
Ca, Cr, Cu, Fe, Mg, Mn, Zn	Biological reference materials	Comparison with ICPOES	601
Ca, Fe, K, Mg, Zn	Biological reference materials	Comparison of CPD and MWD	602
Ca, Fe, K, Mg, Mn, Na, Zn	Biological reference materials	Solubilization with TMAH	597
Ca, K, Mg, Mn	Biological reference materials	MWD	603
Ca, Mg	Various	Interlaboratory trial, dry ashing	604
Ca, Mg	Foliage	Dry ashing	605
Ca, Mg, K	Pine needles	MWD	606
Ca, Mg, K, Na, Cu, Fe, Mn, Zn	Tree rings	Open acid digestion	607
Ca, Na	Fruits, vegetables	La + K as buffer	608
Cd, Co, Cu, Fe, Mn, Ni, Pb, Zn	Medicinal plants	Comparison of digestion methods	609
Cd, Cr, Fe, Mn, Pb, Zn,	Moss and bark	Polluted bioindicators	610
Cd, Pb	Seaweed	Comparison of digestion methods	611
Fe	Potatoes	Speciation study	612
Rb	Biological materials	Overview of Rb in the food chain	613
Zn	Marine organisms	Reference materials	614

techniques for the analysis of biological materials are sorbent extraction on packed microcolumns,[615–618] precipitation/coprecipitation[619,620] and sorption in knotted reactors.[621,622] More important than high enrichment factors is the high rate of sample throughput that can be achieved with these on-line techniques, making them very suitable for FAAS. Enrichment factors of 16–40 with a sample throughput of 120–180 h^{-1}[622,623] appear to be much more sensible than enrichment factors of 90–180 at a sample throughput of only 13 h^{-1}.[624] In addition, if the enrichment time is too long, the susceptibility to interferences and the risk of errors increase, such as the breakthrough of the column.

4.3.4 Methods for Graphite Furnace Atomic Absorption Spectrometry

Digests of biological materials contain about 5–50 mg of the original matrix per milliliter of solution. Since about 50% of the biological matrix is hydrocarbons, which were oxidized and removed as gases during the digestion, the remaining matrix is of the order of 2–30 mg mL^{-1}. This salt content is much higher than for fresh water and the highest concentrations may reach those of seawater. This means that for those cases where the analyte concentration is low, as in seawater, the analysis of biological materials approaches the complexity of seawater analysis. For this reason, the direct analysis of biological materials has greatly benefited from the development of the STPF concept and the instrumentation associated with it. It is accepted nowadays that accurate results can be obtained for the direct analysis of many biological materials by strictly applying this concept and by excluding all forms of contamination. For example, As, Cd, Cu, Fe, Pb, Sb, Se and Zn have been determined by using platform atomization and applying Pd–Mg as modifier and B, Cr, Mn by using Mg as a modifier. The vast amount of literature in this area has been reviewed in the form of a monograph by Tsalev[565] covering the publications up to 1993 and by Welz and Sperling[2] covering publications until 1997.

The number of publications on extraction techniques has decreased markedly owing to the clear advantage of the direct determination of trace elements by GFAAS with respect to simplicity, sample throughput and accuracy. The application of such techniques can be restricted to those cases where the concentration of the element is not sufficiently high for a direct determination. Preconcentration is often required for noble metals such as gold,[625,626] palladium and platinum,[598,627–630] but also for a number of other trace elements such as beryllium,[631] bismuth,[632] cobalt[633,634] and nickel[635] under certain conditions.

Although biological materials are relatively easily dissolved or digested, direct solids analysis has frequently been described (about150 publications during 1975–99). Problems with atomization interferences and background correction, reported in early publications since 1975, have been largely overcome by ZBC and the STPF concept. Introduction of air or oxygen during pyrolysis allows for in situ ashing and helps to avoid carbon build-up in the graphite furnace. Intercomparison studies have shown that direct solid analysis of biological materials is often able to produce accurate results, even when using only samples masses in the range 1–5 mg. When samples are properly homogenized, relative standard deviations in the range <10–20% are typical. When 5–10 subsamples are analyzed by direct solid sampling, confidence intervals for the mean result are in general within about ±10% as precise as the determination of a digest with 2–3 subsamples.

Of special interest is solid sampling for the distribution analysis of biological structures (e.g. birds feathers, human hair, special regions in organs), very small samples (e.g. insects), or for the investigation of the homogeneity

of samples. A disadvantage of solid sampling GFAAS is that not all modern furnaces allow for the introduction of solid samples and often special equipment is needed.

This and other disadvantages can be overcome by the use of suspensions of powdered samples (slurry) in a solvent such as water or dilute acid. Using an automated slurry sampler incorporating an ultrasonic probe for homogenization just prior to dispensing the sample volume into the furnace, fully automated analysis is possible with characteristics very similar to those in solution analysis. Homogenization by ultrasonic agitation is very effective, so that practically all types of biological samples can be suspended in dilute nitric acid with the addition of Triton X-100 as a wetting agent. By using a slightly wider sampler capillary (0.8 mm i.d. instead of the standard 0.4 mm i.d.), particles $< 600\,\mu m$ can be dispensed without major problems. Table 14 gives an overview of elements determined in biological materials by either solid- or slurry-sampling GFAAS.

A complete overview of these techniques can be found in a monograph edited by Kurfürst.[478]

4.3.5 Methods of Speciation Analysis for Biological Materials

Being part of the biosphere, living organisms take part in the environmental cycling of elements. Owing to the metabolism of living organisms, trace elements can be present in biological materials in a large variety of chemical forms. To identify the physiologically active forms, to store and transport of species and to find sensitive indices of the status of trace elements are the main purposes of most environmental investigations into speciation in biological systems. In general, in the investigation of toxic effects, the speciation of small molecules is of concern (e.g. methylmercury), whereas in the investigation of biological functions the determination of large molecules has priority (e.g. vitamin B_{12}).

As in other areas of speciation analysis, the methods of sample collection, pretreatment and storage of biological materials for speciation studies are critical. Not only must a representative sample be collected without contamination, but also the distribution and integrity of the chemical species has to be maintained. Changes in parameters such as temperature, ionic strength, pH, pE, oxygen level, irradiation with UV light, to which the sample is exposed during and after sampling, can influence the speciation of elements. In general, it is essential to choose those analytical conditions which do not differ markedly from those found in the original system, in order to minimize such influences.

Most of the speciation studies on biological materials have been carried out with wet tissue, which is either extracted as fresh material or has previously been dissected and rapidly frozen. Extraction can be performed by blending the sample as a paste or slurry with the selected solvent or solvent mixtures[637] using an Ultra Turrax homogenizer or ultrasonication. Other techniques, such as supercritical fluid extraction (SFE) or microwave-assisted leaching have been applied for this purpose. Using selective extraction procedures for separation, 'first-order speciation' is sometimes possible, especially when the number of species present is limited or when a group parameter is sufficient for the characterization. The most often performed speciation analysis of biological materials is the determination of methylmercury. However, this determination is only rarely done by using GFAAS. The most frequently described technique is extraction of methylmercury chloride or methylmercury bromide from acidic solution with toluene (e.g. fish tissue). The extract can then be analyzed directly by GFAAS after addition of dithizone or thiosulfate acting as a modifier.[638]

The determination of inorganic and organic arsenic compounds is also of great interest. A review of methods for arsenic speciation in environmental and biological samples has been published by IUPAC.[59] Again, since the GFAAS technique cannot be easily coupled with separation techniques such as HPLC, arsenic speciation analysis by using GFAAS has not found wide application. By using extraction with KI–toluene, the more toxic species As(III), As(V), monomethylarsenic

Table 14 Biological matrices and elements investigated by solid-sampling and slurry-sampling GFAAS [information from the literature[2,478,636]]

Matrix	Elements determined by solid-sampling GFAAS	Elements determined by slurry-sampling GFAAS
Animal tissues	Cd, Cs, Hg, Ni, Pb	B, Cd, Cr, Cu, Fe, Mn, Ni, Pb, Se, Si, Sn
Biological reference materials	Ag, As, Cd, Co, Cr, Cs, Cu, Fe, Hg, Mn, Mo, Rb, Se, Tl, Zn	Al, As, B, Ca, Cd, Co, Cr, Cu, Fe, Hg, Mg, Mn, Mo, Ni, Pb, Se, Sn, Zn
Human materials (hair, nails)	Ag, Al, As, Ba, Cd, Co, Cr, Cu, Fe, Hg, La, Mg, Mn, Ni, Pb, Pt, Se, Zn	Al, Ag, B, Cd, Co, Cr, Mn, Ni, Pb
Plant tissues	Cd, Co, Cu, Fe, Mg, Ni, Pb, Tl	Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Tl, Zn
Marine and aquatic samples	As, Cr, Fe, Mn	As, Cd, Co, Cr, Cu, Mn, Ni, Pb, Se

acid and dimethylarsenic acid can be separated from the nontoxic arsenobetaine. Determining arsenic in the extract by GFAAS then gives the total toxic arsenic concentration.[639,640]

Organotin compounds are of interest because of their toxicity. This applies in particular to tributyltin oxide, which is used as an antifouling paint. The separation of this compound from mussels and other marine organisms starts with the extraction from the homogenized tissue mixed with hydrochloric acid by using an organic solvent such as *n*-hexane. Mono- and dibutyltin compounds can be removed with dilute sodium hydroxide solution, and the tributyltin oxide is back extracted in nitric acid. The tin concentration in the extract is determined by GFAAS.[543,544,641,642] Another tin compound of interest is the pesticide tricyclohexyltin hydroxide (Cyhexatin). Residues in apples can be determined directly by GFAAS after extraction with chloroform.[643]

Bioavailable chromium has been determined in cereal products, vegetables and oil seeds by extraction with ethanol and determination by GFAAS after a digestion of the extract.[644] When the extraction of the material alone is not selective for the species, a separation technique such as liquid chromatography (LC) has to follow. Of the LC techniques available, fractionation according to size [SEC or fast protein liquid chromatography (FPLC)] has been most frequently used.

For selenium speciation, the determination of the organic compounds and the distinction between inorganic selenite and selenate are of interest. Ion chromatography has been applied to separate selenite and selenate, and FAAS and GFAAS were used for on-line detection, depending on the required sensitivity for the analysis of feedstuffs and selenium additives in animal feeds.[645] A greater number of inorganic and organic selenium species can be separated by coupling FPLC with GFAAS via an ultralow volume fraction collector. By using this approach for the determination of selenium species in plant tissues, a sensitivity was obtained that is comparable to that obtained by LC/ICPMS.[646,647] Again, extraction of such Se species from the biological tissues is a critical step.[648]

Chromium speciation, which is of great importance because of the high toxicity of Cr(VI), has not been widely applied to biological materials. Metal-free HPLC equipment and fast procedures are a prerequirement for obtaining undisturbed speciation; FPLC was successfully used to separate different Cr complexes extracted from cabbage xylem.[649] Positively charged chromium(III) species and kinetically labile complexes are not retained on the column, while the relatively stable chromium(III)–EDTA and –oxalate complexes and chromium(V) are retained and separated. Determination of chromium in the individual fractions was performed off-line by GFAAS.

For the separation of various lead species in biological materials, fractionated wash procedures using organic solvents and acids of varying strengths have been applied.[650,651] The investigation of roots, foliage and other plant components, and the dissection of leaves into subcellular fractions with the corresponding compartmentalization of the lead content, have been described.[651] In many cases GFAAS is used as the determination technique for the sum determination of volatile lead, that is, tetraalkyllead compounds in marine organisms.[650]

Metallothioneins and other high molecular weight species are often determined by LC separation with various AAS detectors. A review of coupled techniques used for such analysis has been presented by Lobinski et al.[652] The major interest is in cadmium,[653–661] but other elements such as copper,[655,658,661–663] zinc,[658,661] lead,[655] magnesium,[663] nickel, silver[661] and thallium[660] are also determined. It is of particular interest to determine to which protein the metal is bound; gel filtration and gel permeation chromatography are especially suitable for separating various metal-binding proteins. The determination of the metal concentrations in the eluate is carried out either by on-line coupling with FAAS or by GFAAS. Likewise for this application, the difficulty of coupling HPLC with AAS detectors is apparent. On-line coupling with FAAS is possible when the nebulizer aspiration rate is reduced to about $2\,mL\,min^{-1}$, but the sensitivity is relatively poor.[654,658,664] Sensitivity can be improved by using high-efficiency nebulizers for interfacing, such as the thermospray nebulizer.[653,655,663] GFAAS offers the required sensitivity but it can only be applied off-line[656,657] by using a fraction collector or quasi-on-line by a flow-through interface,[537] both with limitations with respect to resolution and versatility as already mentioned (see section 4.1.5).

Other techniques which have been used in combination with AAS detection are ultrafiltration,[665,666] ultracentrifugation[667,668] and supercritical fluid chromatography (SFC).[659]

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
AAS	Atomic Absorption Spectrometry
APDC	Ammonium Pyrrolidine Dithiocarbamate
ASV	Anodic Stripping Voltammetry
BCR	Bureau Communautaire de Référence
BOC	Baseline Offset Correction
CPD	Conventional Pressure Digestion
CPG	Controlled-pore Glass
CRM	Certified Reference Material
CVAAS	Cold Vapor Atomic Absorption Spectrometry

DDDC	Diethylammonium Diethyl-dithiocarbamate
DDTC	Diethyldithiocarbamate
DQO	Data Quality Objectives
EDL	Electrodeless Discharge Lamp
EDTA	Ethylenediaminetetraacetic acid
EEC	European Economic Community
ETAAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Flame Atomic Absorption Spectrometry
FI	Flow Injection
FPLC	Fast Protein Liquid Chromatography
GAP	Good Analytical Practice
GC	Gas Chromatography
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
GLP	Good Laboratory Practice
HCL	Hollow-cathode Lamp
HGAAS	Hydride Generation Atomic Absorption Spectrometry
HPLC	High-performance Liquid Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICPOES	Inductively Coupled Plasma Optical Emission Spectrometry
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LOD	Limit of Detection
LOQ	Limit of Quantification
MWD	Microwave-assisted Digestion
OES	Optical Emission Spectrometry
PFA	Perfluoroalkoxy
PIXE	Particle-induced X-ray Emission
PTFE	Polytetrafluoroethylene
QA	Quality Assurance
QC	Quality Control
QTA	Quartz Tube Atomizer
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
S/N	Signal-to-noise Ratio
STPF	Stabilized Temperature Platform Furnace
THGA	Transversely Heated Graphite Atomizer
TMAH	Tetramethylammonium Hydroxide
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
WHO	World Health Organization
ZBC	Zeeman Effect Background Correction

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Airborne Instrumentation for Aerosol Measurements • Automotive Emissions Analysis with Spectroscopic Techniques

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Biological Samples in Environmental Analysis: Preparation and Cleanup • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Flow-injection Techniques in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Organometallic Compound Analysis in Environmental Samples • Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses • Soil Instrumental Methods • Soil Sampling for the Characterization of Hazardous Waste Sites • Underground Fuel Spills, Source Identification • Waste Extraction Procedures

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Atomic Spectroscopy in Food Analysis • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air • Dust, Measurement of Trace Elements in

Industrial Hygiene cont'd **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration

REFERENCES

1. G.R. Dulude, 'Background Correction Techniques in AAS', in *Advances in Atomic Spectroscopy*, ed. J. Sneddon, JAI Press, Greenwich, CT, 126–160, 1992.
2. B. Welz, M. Sperling, *Atomic Absorption Spectrometry*, Wiley-VCH, Weinheim, 614–655, 1998.
3. D.C. Manning, W. Slavin, 'Determination of Trace Elements in Natural Waters Using STPF', *Appl. Spectrosc.*, **37**(1), 1–11 (1983).
4. M. Sperling, B. Welz, J. Hertzberg, C. Rieck, G. Marowsky, 'Temporal and Spatial Temperature Distributions in THGAs and Their Analytical Characteristics for AAS', *Spectrochim. Acta, Part B*, **51**(9–10), 897–930 (1996).
5. B. Welz, M. Sperling, G. Schlemmer, N. Wenzel, G. Marowsky, 'Spatially and Temporally Resolved Gas Phase Temperature Measurements in a Massmann-type Graphite Tube Furnace Using CARS', *Spectrochim. Acta, Part B*, **43**(9–11), 1187–1207 (1988).
6. W. Slavin, G.R. Carnrick, 'The Possibility of Standardless Furnace AAS', *Spectrochim. Acta, Part B*, **39**(2–3), 271–282 (1984).
7. J. Dedina, D.L. Tsalev, *Hydride Generation Atomic Absorption Spectrometry*, John Wiley & Sons, Chichester, 1995.
8. Z.-l. Fang, *Flow Injection AAS*, John Wiley & Sons, Chichester, 1995.
9. B. Welz, M. Sperling, 'FI: The Ultimate Approach to Automation in Analytical Atomic Spectroscopy', *Pure Appl. Chem.*, **65**(12), 2465–2472 (1993).
10. International Standardization Organization, *Accuracy (Trueness and Precision) of Measurement Methods and Results, Parts 1–6*, ISO 5725, ISO, Geneva, 1993.
11. E. Prichard, G. Mackay, J. Points (eds.), *Trace Analysis – a Structured Approach to Obtaining Reliable Results*, Royal Society of Chemistry, Cambridge, 1996.
12. A.M. Ure, L.R.P. Butler, R.G. Scott, R. Jenkins, 'Nomenclature, Symbols, Units and Their Usage in Spectrochemical Analysis. Part X. Preparation of Materials for Analytical Atomic Spectroscopy and Other Related Techniques', *Pure Appl. Chem.*, **60**(9), 1461–1472 (1988).
13. P.E. Gardiner, 'Factors That Influence the Distribution of Trace Element-containing Chemical Species in Living Systems Before and After Sample Collection', *J. Trace Elem. Electrol. Health Dis.*, **7**, 1–8 (1993).
14. B. Michalke, 'Quality Control in Speciation Investigations', *Fresenius' J. Anal. Chem.*, **350**(1–2), 2–6 (1994).
15. Ph. Quevauviller, M.B. de la Calle-Guntiñas, E.A. Maier, C. Cámara, 'Stability of Chemical Species in Solution During Storage: A Review', *Mikrochim. Acta*, **118**, 131–141 (1995).
16. W. Slavin, D.C. Manning, G.R. Carnrick, 'Quality-assurance Procedures for GFAAS', *Talanta*, **36**(1+2), 171–178 (1989).
17. H. Le Garrec, P. Giamarchi, J.-Y. Cabon, A. Le Bihan, 'Electrothermic Factors Optimization in ETAAS via an Optimal Experimental Design Matrix', *Anal. Chim. Acta*, **368**(1–2), 59–70 (1998).
18. I. Koch, C.F. Harrington, K.J. Reimer, W.R. Cullen, 'Simplex Optimization of Conditions for the Determination of Sb in Environmental Samples by Using ETAAS', *Talanta*, **44**(5), 771–780 (1997).
19. D.J. Halls, 'Analytical Minimalism Applied to the Determination of Trace Elements by Atomic Spectrometry', *J. Anal. At. Spectrom.*, **10**(3), 169–175 (1995).
20. K.S. Farah, J. Sneddon, 'Optimization of a Simultaneous Multi-element AA Spectrometer', *Talanta*, **40**(6), 879–882 (1993).
21. J.M. Harnly, J.S. Kane, 'Optimization of Electrothermal Atomization Parameters for Simultaneous Multielement AAS', *Anal. Chem.*, **56**(1), 48–54 (1984).
22. J.M. Harnly, B. Radziuk, 'Effect of Furnace Atomization Temperatures on Simultaneous Multielement Atomic Absorption Measurement Using a THGA', *J. Anal. At. Spectrom.*, **10**(3), 197–206 (1995).
23. J.J. van der Lee, E. Temminghoff, V.J.G. Houba, I. Novozamsky, 'Background Corrections in the Determination of Cd and Pb by FAAS in Plant and Soil Samples with High Fe Levels', *Appl. Spectrosc.*, **41**(3), 388–390 (1987).
24. V.A. Letourneau, B.M. Joshi, L.C. Butler, 'Comparison Between Zeeman and Continuum Background Correction for GFAAS on Environmental Samples', *At. Spectrosc.*, **8**(5), 145–149 (1987).
25. D.C. Baxter, W. Frech, I. Berglund, 'Use of Partial Least Squares Modeling to Compensate for Spectral Interference in ETAAS with Continuum Source Background Correction. I. Determination of As in Marine Sediments', *J. Anal. At. Spectrom.*, **6**(2), 109–113 (1991).
26. W. Slavin, G.R. Carnrick, 'Background Correction in AAS', *CRC Crit. Rev. Anal. Chem.*, **19**(2), 95–134 (1988).
27. U. Heitmann, M. Schütz, H. Becker-Ross, S. Florek, 'Measurement on the Zeeman-splitting of Analytical

Lines by Means of a Continuum Source GFAA Spectrometer with a Linear CCD Array', *Spectrochim. Acta, Part B*, **51**(9–10), 1095–1105 (1996).
28. J.-Y. Cabon, A. Le Bihan, 'Direct Determination of Zn in Seawater Using ETAAS with ZBC: Effects of Chemical and Spectral Interferences', *J. Anal. At. Spectrom.*, **9**(3), 477–481 (1994).
29. A. Le Bihan, J.Y. Cabon, C. Elleouet, 'Spectral Interference Related to the Presence of Nitrate and Nitrogen Mono-oxide in Zeeman-corrected AAS with Electrothermal Atomization', *Analusis*, **20**, 601–604 (1992).
30. K.E. Anders Ohlsson, W. Frech, 'Photographic Observation of Molecular Spectra in Inverse Zeeman-effect GFAAS', *J. Anal. At. Spectrom.*, **4**(5), 379–385 (1989).
31. G. Wibetoe, F.J. Langmyhr, 'Interferences in Inverse Zeeman-corrected AAS Caused by Zeeman Splitting of Molecules', *Anal. Chim. Acta*, **198**, 81–86 (1987).
32. P. Quevauviller, 'Quality Assurance for Environmental Analysis', in *Quality Assurance for Environmental Analysis*, eds. P. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 1–25, 1995.
33. B.D. Zehr, M.A. Maryott, 'Accuracy and Precision of Dilutions Prior to AAS Analysis', *Spectrochim. Acta, Part B*, **48**(10), 1275–1280 (1993).
34. D. Erber, J. Bettmer, K. Cammann, 'Influence of Blank Value Correction on the Accuracy of Determination in Routine Analysis', *GIT Fachz. Lab.*, **39**(4), 340–349 (1995).
35. J.K. Taylor, *Quality Assurance of Chemical Measurements*, Lewis, Chelsea, MI, 1989.
36. W.Y. Garner, M.S. Barge, J.P. Ussary, *Good Laboratory Practice Standards. Application for Field and Laboratory*, American Chemical Society, Washington, DC, 571, 1992.
37. U. Hillebrand, 'Quality Assurance in Water Analysis: The Calibration Function', *GIT Fachz. Lab.*, **35**, 1001–1007 (1991).
38. E. Mullins, 'Getting More from Your Laboratory Control Charts', *Analyst*, **124**(4), 433–442 (1999).
39. E. Mullins, 'Introduction to Control Charts in the Analytical Laboratory', *Analyst*, **119**(3), 369–375 (1994).
40. H. Hein, W. Kunze, *Environmental Analysis by Spectrometry and Chromatography – from Laboratory Design to Data Interpretation*, Wiley-VCH, Weinheim, 250, 1999.
41. R.N. Reeve, J.D. Barnes, *Environmental Analysis, Analytical Chemistry by Open Learning*, ACOL Series, John Wiley & Sons, Chichester, Vol. 32, 1994.
42. E. Merian, *Metals and Their Compounds in the Environment: Occurence, Analysis and Biological Relevance*, VCH, Weinheim, 1991.
43. M. Stoeppler, *Hazardous Metals in the Environment*, Techniques and Instrumentation in Analytical Chemistry, Elsevier, Amsterdam, Vol. 12, 1992.
44. F.W. Fifield, P.J. Haines, *Environmental Analytical Chemistry*, Blackie Academic and Professional, Glasgow, 1995.
45. M. Radojevic, V. Bashkin, *Practical Environmental Analysis*, Royal Society of Chemistry, Cambridge, 1999.
46. T. Braun, S. Zsindely, 'Some Recent Trends in Instrumental Analysis of Environmental Materials', *Anal. Proc.*, **28**, 283–286 (1991).
47. M.S. Cresser, 'Atomic Spectrometry in Environmental Analysis: Jam Today or Jam Tomorrow?', *J. Anal. At. Spectrom.*, **8**(2), 269–272 (1993).
48. M. Hoenig, M.F. Guns, 'Analysis of Environmental and Biological Samples by Atomic Spectroscopic Methods', in *Quality Assurance for Environmental Analysis*, eds. Ph. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 63–88, 1995.
49. L.H. Keith, W. Crummett, J. Deegan, R.A. Libby, J.K. Taylor, G. Wentler, 'Principles of Environmental Analysis', *Anal. Chem.*, **55**(14), 2210–2218 (1983).
50. A. Kumar, M.Z. Hasan, B.T. Deshmukh, 'Flameless AAS Methods for the Determination of Trace Metals of Environmental Significance: Review', *Indian J. Pure Appl. Phys.*, **25**(2), 49–57 (1987).
51. P. Schramel, 'Trends and Demands in and on Trace-element Analysis in the Biomedical and Environmental Field', *J. Radioanal. Nucl. Chem.*, **168**(1), 215–221 (1993).
52. G.M. Pierzynski, 'Past, Present and Future Approaches for Testing Metals for Environmental Concerns and Regulatory Approaches', *Commun. Soil. Sci. Plant Anal.*, **29**(11–14), 1523–1536 (1998).
53. W. Frech, A. Cedergren, 'Aluminum', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, 451–473, Vol. 12, 1992.
54. R.C. Massey, D. Taylor (eds.), *Aluminium in Food and the Environment*, Royal Society of Chemistry, Cambridge, 1991.
55. J. Savory, R.L. Bertholf, S. Brown, M.R. Wills, 'Aluminum', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 273–289, 1994.
56. K.J. Irgolic, 'Arsenic', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, 287–350, Vol. 12, 1992.
57. M. Burguera, J.L. Burguera, 'Analytical Methodology for Speciation of As in Environmental and Biological Samples', *Talanta*, **44**(9), 1581–1604 (1997).
58. M.B. Amran, F. Lagarde, M.J.F. Leroy, A. Lamotte, C. Demesmay, M. Ollé, M. Albert, G. Rauret, J.F. López-Sánchez, 'As Speciation in Environmental Matrices', in *Quality Assurance for Environmental Analysis*, eds. Ph. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 285–304, 1995.
59. M. Morita, J.S. Edmonds, 'Determination of As Species in Environmental and Biological Samples', *Pure Appl. Chem.*, **64**(4), 575–590 (1992).

60. M.L. Cervera, R. Montoro, 'Critical Review of the Atomic Spectrometric Analysis of As in Foods', *Fresenius' J. Anal. Chem.*, **348**(5–6), 331–340 (1994).
61. M. Stoeppler, M. Vahter, 'Arsenic', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 291–320, 1994.
62. K. Robards, P.J. Worsfold, 'Analytical Approaches to the Monitoring of Toxicological Levels of Cd in Environmental Samples', *Trends Anal. Chem. (Pers. Ed.)*, **9**(7), 228–231 (1990).
63. K. Robards, P.J. Worsfold, 'Cd: Toxicology and Analysis – a Review', *Analyst*, **116**(6), 549–568 (1991).
64. M. Stoeppler, 'Cadmium', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, Vol. 12, Chapter 8, 1992.
65. R.F.M. Herber, 'Cadmium', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, Chapter 15, 1994.
66. N. Miller-Ihli, 'Chromium', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, 373–400, Vol. 12, 1992.
67. R. Rubio, A. Sahuquillo, G. Rauret, P. Quevauviller, 'Determination of Cr in Environmental and Biological Samples by AAS: A Review', *Int. J. Environ. Anal. Chem.*, **47**, 99–128 (1992).
68. S.A. Katz, 'Chromium, Environmental Analytical Chemistry', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 1128–1141, Vol. 2, 1998.
69. M. Sperling, 'Chromium – Elemental Determination', in *Encyclopedia of Analytical Science*, ed. A. Townshend, Academic Press, London, 729–736, 1995.
70. M. Sperling, 'Chromium – Compound Determination', in *Encyclopedia of Analytical Science*, ed. A. Townshend, Academic Press, London, 736–743, 1995.
71. R. Cornelis, 'Chromium', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 339–358, 1994.
72. M. Stoeppler, P. Ostapczuk, 'Nickel and Cobalt', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, Vol. 12, Chapter 14, 1992.
73. H.T. Delves, M. Stoeppler, 'Copper', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 359–370, 1994.
74. S.J. Hill, 'Lead', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, Vol. 12, Chapter 9, 1992.
75. A.R. Flegal, D.R. Smith, 'Measurements of Environmental Pb Contamination and Human Exposure', *Rev. Environ. Contam. Toxicol.*, **143**, 1 (1995).
76. K. Pyrzynska, 'Organolead Speciation in Environmental Samples: A Review', *Mikrochim. Acta*, **122**(3–4), 279–293 (1996).
77. D.C. Baxter, W. Frech, 'Speciation of Pb in Environmental and Biological Samples', *Pure Appl. Chem.*, **67**(4), 615–648 (1995).
78. U. Ewers, M. Turfield, E. Jermann, 'Lead', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 371–384, 1994.
79. D.J. Halls, 'Manganese', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 385–402, 1994.
80. I. Drabæk, Å. Iverfeldt, 'Mercury', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, Vol. 12, Chapter 10, 1992.
81. B.E. Rood, C.L. Sanford, 'Mercury, Determination in the Environment', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 2737–2744, Vol. 5, 1998.
82. M. Morita, J. Yoshinaga, J.S. Edmonds, 'The Determination of Hg Species in Environmental and Biological Samples', *Pure Appl. Chem.*, **70**(8), 1585–1615 (1998).
83. J.E. Sánchez-Uría, A. Sanz-Medel, 'Inorganic and Methylmercury Speciation in Environmental Samples', *Talanta*, **47**(3), 509–524 (1998).
84. W. Baeyens, 'Speciation of Hg in Different Compartments of the Environment', *Trends Anal. Chem. (Pers. Ed.)*, **11**(7), 245–254 (1992).
85. A. Schütz, G. Skarping, S. Skerfving, K. Malmquist, 'Mercury', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 403–467, 1994.
86. Z. Marczenko, R. Lobinski, 'Determination of Mo in Biological Materials', *Pure Appl. Chem.*, **63**(11), 1627–1636 (1991).
87. J. Angerer, R. Heinrich-Ramm, J. Begerova, 'Bestimmung von Ni in Biologischem Material und Umweltrelevanten Matrices', in *Analytiker Taschenbuch*, eds. H. Günzler, R. Borsdorf, W. Fresenius, W. Huber, H. Kelker, I. Lüderwald, Springer, Berlin, 398–433, Vol. 10, 1991.
88. D.M. Templeton, 'Nickel', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 469–487, 1994.
89. M. Balcerzak, 'Analytical Methods for the Determination of Pt in Biological and Environmental Materials – a Review', *Analyst*, **122**(5), 67R–74R (1997).
90. R.R. Barefoot, 'Determination of Pt at Trace Levels in Environmental and Biological Materials', *Environ. Sci. Technol.*, **31**(2), 309–314 (1997).
91. M. Ihnat, 'Selenium', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, 475, Vol. 12, 1992.

92. A. D'Ulivo, 'Determination of Se and Te in Environmental Samples', *Analyst*, **122**(12), 117R–144R (1997).
93. G. Kölbl, K. Kalcher, K.I. Irgolic, 'Identification and Determination of Se Compounds in Samples of Environmental Importance', in *Main Group Elements and Their Compounds*, ed. V.G. Kumar, Narosa Publishing House, New Delhi, 161–172, 1996.
94. R. Muñoz-Olivas, O.F.X. Donard, C. Cámara, P. Quevauviller, 'Analytical Techniques Applied to the Speciation of Se in Environmental Matrices', *Anal. Chim. Acta*, **286**(3), 357–370 (1994).
95. Y. Thomassen, S.A. Lewis, C. Veillon, 'Selenium', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 489–500, 1994.
96. M. Sager, 'Thallium', in *Hazardous Metals in the Environment*, ed. M. Stoeppler, Techniques in Instrumental Analytical Chemistry, Elsevier, Amsterdam, Vol. 12, Chapter 12, 1992.
97. M. Sager, 'Thallium', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, Chapter 23, 1994.
98. International Union of Pure and Applied Chemistry, 'Determination of Sn Species in Environmental Samples (Technical Report)', *Pure Appl. Chem.*, **70**(10), 2051–2064 (1998).
99. P. Quevauviller, 'The Analysis of Butylated Sn Compounds in the Environment and in Biological Materials', in *Tributyltin: Case Study of an Environmental Contaminant*, ed. S. De Mora, Cambridge University Press, Cambridge, 62–93, 1996.
100. K.-H. Schaller, 'Vanadium', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, Chapter 24, 1994.
101. G.S. Fell, T.D.B. Lyon, 'Zinc', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 541–561, 1994.
102. J.R. Dean, O. Butler, A. Fisher, L.M. Garden, M.S. Cresser, P. Watkins, M. Cave, 'Atomic Spectrometry Update – Environmental Analysis', *J. Anal. At. Spectrom.*, **13**(2), 1R–56R (1998).
103. R.E. Clement, P.W. Yang, C.J. Koester, 'Environmental Analysis', *Anal. Chem.*, **69**(12), 251R–287R (1997).
104. R.E. Sturgeon, 'GFAAS and Environmental Challenges at the Ultratrace Level – a Review', *Spectrochim. Acta, Part B*, **52**(9–10), 1451–1457 (1997).
105. S. Mitra, *Mercury in the Ecosystem: Its Dispersion and Pollution Today*, Trans Tech Publications, Aedermannsdorf, 1986.
106. S.S. Berman, P.A. Yeats, 'Sampling of Seawater for Trace Metals', *CRC Crit. Rev. Anal. Chem.*, **16**(1), 1–14 (1985).
107. G. Topping, 'Quality of Data: With Special Reference to the Measurement of Trace Metals in Marine Samples', *Sci. Total Environ.*, **49**, 9–25 (1986).
108. T.R. Crompton, *Comprehensive Water Analysis. Volume One: Natural Waters*, Elsevier Applied Science, New York, 1992.
109. T.R. Crompton, *Comprehensive Water Analysis. Volume Two: Treated Waters*, Elsevier Applied Science, New York, 1992.
110. D.T.E. Hunt, A.L. Wilson, *The Chemical Analysis of Water: General Principles and Techniques*, Royal Society of Chemistry, Cambridge, 1995.
111. H.H. Rump, H. Krist, *Laboratory Manual for the Examination of Water, Wastewater and Soil*, 2nd edition, VCH, Weinheim, 1992.
112. T.S. West, H.W. Nürnberg, *The Determination of Trace Metals in Natural Waters*, Blackwell Scientific, Oxford, 1988.
113. M. Blankley, A. Henson, K.C. Thompson, 'Waters, Sewage and Effluent', in *AAS – Theory, Design and Application*, ed. S.J. Haswell, Elsevier, Amsterdam, 79–123, 1991.
114. G. Gillain, 'Seawater Samples', in *Quantitative Trace Analysis of Biological Materials*, eds. H.A. McKenzie, L.E. Smythe, Elsevier, Amsterdam, 389–399, 1988.
115. K. Grasshoff, K. Kremling, M. Ehrhardt (eds.), *Methods of Seawater Analysis*, 3rd edition, Wiley-VCH, Weinheim, 1999.
116. H. Haraguchi, T. Akagi, 'Application of AAS to Marine Analysis', in *AAS – Theory, Design and Applications*, ed. S.J. Haswell, Elsevier, Amsterdam, 125–157, 1991.
117. M. Sadiq, *Toxic Metals in Marine Environments*, Marcel Dekker, New York, 400, 1992.
118. B. Fairman, A. Sanz-Medel, 'Determination of Al Species in Natural Waters', in *Quality Assurance for Environmental Analysis*, eds. Ph. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 215–233, 1995.
119. N. Clarke, L.-G. Danielsson, A. Sparén, 'Analytical Methodology for the Determination of Al Fractions in Natural Fresh Waters', *Pure Appl. Chem.*, **68**(8), 1597–1638 (1996).
120. K. Pyrzynska, E. Bulska, S. Güçer, A. Hulanicki, 'Al Speciation in Natural Waters', *Chem. Anal. (Warsaw)*, **44**(1), 1–14 (1999).
121. M.B. de la Calle-Guntiñas, Y. Madrid, C. Cámara, 'Sb Speciation in Water', in *Quality Assurance for Environmental Analysis*, eds. Ph. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 263–283, 1995.
122. P. Smichowski, Y. Madrid, C. Cámara, 'Analytical Methods for Sb Speciation in Waters at Trace and Ultratrace Levels. A Review', *Fresenius' J. Anal. Chem.*, **360**(6), 623–629 (1998).
123. F.B. Salem, 'Determination of Fe in Water Samples', *Rev. Anal. Chem.*, **14**(1), 59 (1995).
124. M. Meili, 'Hg in Lakes and Rivers', in *Mercury and its Effects on Environment and Biology*, eds. A. Sigel, H. Sigel, Marcel Dekker, New York, Chapter 2, 1997.
125. S. Geiss, J. Einax, K. Danzer, 'Multivariate-statistical Examination of the Sampling Strategy in the Analysis of

River Water', *Fresenius' Z. Anal. Chem.*, **333**(2), 97–101 (1989).
126. R.E. Sturgeon, S.S. Berman, 'Sampling and Storage of Natural Water for Trace Metals', *CRC Crit. Rev. Anal. Chem.*, **18**(3), 209–244 (1987).
127. P. Ostapczuk, 'Precipitation – Rain and Snow', in *Probennahme und Aufschluss*, ed. M. Stoeppler, Springer, Berlin, 9–17, 1994.
128. M. Mongoni, R. Udisti, G. Piccardi, 'Sequential Sampling of Rain: Construction and Operation of an Automatic Wet-only Apparatus', *Int. J. Environ. Anal. Chem.*, **69**(1), 53–66 (1998).
129. J.R. Kramer, 'Analysis of Precipitation', in *Handbook of Air Pollution Analysis*, eds. R.M. Harrison, R. Perry, Chapman and Hall, New York, 535–561, 1986.
130. J.R. Scudlark, T.M. Church, K.M. Conko, S.M. Moore, 'Method for the Automated Collection, Proper Handling and Accurate Analysis of Trace Metals in Precipitation', *Gen. Tech. Rep. NC (North Cent. For. Exp. Stn.)*, **150**, 57 (1992).
131. M.S. Landis, G.J. Keeler, 'Critical Evaluation of a Modified Automatic Wet-only Precipitation Collector for Hg and Trace Element Determinations', *Environ. Sci. Technol.*, **31**(9), 2610 (1997).
132. E. Rohbock, H.W. Georgii, C. Perseke, L. Kins, 'Determination of Heavy Metals in Rain Water', in *Atomspektrometrische Spurenanalytik*, ed. B. Welz, Verlag Chemie, Weinheim, 295–304, 1982.
133. P. Winkler, M. Schulz, W. Dannecker, 'Problems with the Determination of Heavy Metals in Precipitation', *Fresenius' J. Anal. Chem.*, **340**(9), 575–579 (1991).
134. T.B. Ridder, T.A. Buishand, H.F.R. Reijnders, M.J. 't Hart, J. Slanina, 'Effects of Storage on the Composition of Main Components in Rainwater Samples', *Atmos. Environ.*, **19**, 759–762 (1985).
135. L. Brügmann, K. Kremling, 'Sampling [of Seawater]', in *Methods of Seawater Analysis*, eds. K. Grasshoff, K. Kremling, M. Ehrhardt, 3rd edition, Wiley-VCH, Weinheim, 1–25, 1999.
136. P.E. McGuire, B. Lowery, P.A. Helmke, 'Potential Sampling Error: Trace Metal Absorption on Vacuum Porous Cup Samplers', *Soil Sci. Soc. Am. J.*, **56**(1), 744–782 (1992).
137. M.J. Gardner, S. Comber, 'Sample Filtration as a Source of Error in the Determination of Trace Metals in Marine Waters', *Analyst*, **122**(10), 1029–1032 (1997).
138. A.J. Horowitz, K.R. Lum, J.R. Garbarino, G.E.M. Hall, C. Lemieux, C.R. Demas, 'Problems Associated with Using Filtration to Define Dissolved Trace Element Concentrations in Natural Water Samples', *Environ. Sci. Technol.*, **30**(3), 954–963 (1996).
139. A.M. Shiller, H.E. Taylor, 'Comment on 'Problems Associated with Using Filtration to Define Dissolved Trace Element Concentrations in Natural Water Samples'', *Environ. Sci. Technol.*, **30**(11), 3397–3398 (1996).
140. A.J. Horowitz, K.R. Lum, J.R. Garbarino, G.E.M. Hall, C. Lemieux, C.R. Demas, 'Response to Comments on Problems Associated with Using Filtration to Define Dissolved Trace Element Concentrations in Natural Water Samples', *Environ. Sci. Technol.*, **30**(11), 3398–3400 (1996).
141. A.J. Horowitz, 'Some Thoughts on Problems Associated with Various Sampling Media Used for Environmental Monitoring', *Analyst*, **122**(11), 1193–1200 (1997).
142. J.C. Méranger, K.S. Subramanian, C.H. Langford, 'On-site Sampling with Pre-concentration for Trace-metal Analysis of Natural Waters', *Rev. Anal. Chem.*, **5**(1–2), 29–51 (1980).
143. Z.-h. Hou, 'Au in Tuanjiegou Natural Waters: A Method of Determination by Preconcentration with Pretreated Polyurethane Foam in the Field', *J. Geochem. Explor.*, **33**, 255 (1989).
144. A.G. Cox, C.W. McLeod, 'Field Sampling Technique for Speciation of Inorganic Cr in Rivers', *Mikrochim. Acta*, **109**(1–4), 161–164 (1992).
145. J. Wie, C.W. McLeod, 'Field Sampling Techniques for Mercury Speciation', *Anal. Proc.*, **28**(9), 293–294 (1991).
146. H.M. Ortner, H.H. Xu, J. Dahmen, K. Englert, H. Opfermann, W. Görtz, 'Surface Characterization of Fluorinated Polymers (PTFE, PVDF, PFA) for Use in Ultratrace Analysis', *Fresenius' J. Anal. Chem.*, **355**(5–6), 657–664 (1996).
147. G. Giebenhain, H.J. Rath, 'Application of Fluorpolymer in Trace Analysis', *GIT Fachz. Lab.*, **39**(3), 247–250 (1995).
148. H.M. Ortner, C. Berger, O. Boese, P. Hoffmann, A. Dedik, K. Englert, U. Sengutta, H.-J. Staerck, 'Significance of Vessel Materials in Ultra-trace Analysis. Part 1', *GIT Fachz. Lab.*, **42**(3), 186–192 (1998).
149. H.M. Ortner, 'Importance of Vessel Materials for Ultratrace Analysis (Part 2)', *GIT Fachz. Lab.*, **42**(4), 390–394 (1998).
150. M. Takenaka, M. Hayashi, I. Suzuki, Y. Yamada, K. Takamatsu, M. Kageyama, 'Evaluation of a Mirror-polishing Technique for Fluorocarbon Polymer Surfaces for Reduction of Contamination from Containers Used in Ultratrace Analysis', *Anal. Chem.*, **69**(5), 972–976 (1997).
151. Z.-l. Fang, *Flow-injection Separation and Pre-concentration*, VCH, New York, 1993.
152. Z.-l. Fang, S.-k. Xu, S.-c. Zhang, 'The Determination of Trace Amounts of Heavy Metals in Waters by a FI System Including Ion-exchange Preconcentration and FAAS Detection', *Anal. Chim. Acta*, **164**, 41–50 (1984).
153. Z.-l. Fang, B. Welz, 'High Efficiency Low Sample Consumption On-line Ion-exchange Preconcentration System for FI FAAS', *J. Anal. At. Spectrom.*, **4**(6), 543–546 (1989).
154. Y. Liu, J.D. Ingle, 'Automated On-line Ion-exchange Trace Enrichment System with FAAS Detection', *Anal. Chem.*, **61**(6), 520–524 (1989).

155. M.A. Memon, Z.X. Zhuang, Z.L. Fang, 'Preconcentration of Trace Cu in Water Samples with FI–Solvent Extraction and AAS', *At. Spectrosc.*, **14**(2), 50–54 (1993).
156. B. Mohammad, A.M. Ure, D. Littlejohn, 'On-line Preconcentration of Al, Ga and In with Quinolin-8-ol for Determination by AAS', *J. Anal. At. Spectrom.*, **8**(2), 325–331 (1993).
157. Department of the Environment (UK), 'Li, Mg, Ca, Sr and Ba in Waters and Sewage Effluents by AAS 1987', *Methods Exam. Waters Assoc. Mater.*, 38 (1987).
158. C.P. Bosnak, Z.A. Grosser, 'The Analysis of Drinking Water and Bottled Water by FAAS and GFAAS', *At. Spectrosc.*, **17**(6), 218–224 (1996).
159. J.C. Latino, Z.A. Grosser, 'Modern FAAS Analysis of Wastewater', *At. Spectrosc.*, **17**(6), 215–217 (1996).
160. H. Schinkel, 'Bestimmung von Ca, Mg, Sr, K, Na, Li, Fe, Mn, Cr, Ni, Cu, Co, Zn und Cd. Eine Universalvorschrift Zur Untersuchung von Wässern, Kohlen, Aschen, Schlacken, Erzen, Gesteinen, Baustoffen, Metallen und ähnlichen Proben', *Fresenius' Z. Anal. Chem.*, **317**(1), 10–26 (1984).
161. A. Batho, C. Phillips, 'Simultaneous Determination of Ca and Mg by Dual-channel AAS', *Lab. Pract.*, **38**(3), 70 (1989).
162. M. Abbas, R. Bruns, I. Scarminio, 'A Multivariate Statistical Analysis of the Composition of Rainwater Near Cubatao, Sp, Brazil', *Environ. Pollut.*, **79**, 225 (1992).
163. M. Grasserbauer, S. Paleczek, J. Rendl, A. Kasper, H. Puxbaum, 'Inorganic Constituents in Aerosols, Cloud Water and Precipitation Collected at the High Alpine Measurement Station Sonnblick: Sampling, Analysis and Exemplary Results', *Fresenius' J. Anal. Chem.*, **350**(7–8), 431–439 (1994).
164. M. de la Guardia, V. Carbonell, A. Morales-Rubio, A. Salvador, 'Direct Determination of Ca, Mg, Na and K in Water by FI Flame Atomic Spectroscopy, Using a Dilution Chamber', *Fresenius' Z. Anal. Chem.*, **335**(8), 975–979 (1989).
165. D. Maier, H.W. Sinemus, E. Wiedeking, 'AAS-Bestimmung Gelöster Spurenelemente im Bodenseewasser des Überlinger Sees', *Fresenius' Z. Anal. Chem.*, **296**, 114–124 (1979).
166. R.G. Zytner, N. Biswas, J.K. Bewtra, 'Comparison of D.C. Plasma AES and AAS Analysis in Determining Cu, Cd, Cr, Pb, Ni and Zn Concentration in Municipal Wastewater', *Int. J. Environ. Stud.*, **33**(4), 299–306 (1989).
167. M.A. Zayed, A.F. Nour El-Dien, K.A. Rabie, 'Comparative Study of Seasonal Variation in Metal Concentrations in River Nile Sediment, Fish, and Water by AAS', *Microchem. J.*, **49**, 27–35 (1994).
168. Fe and Mn in Potable Waters by AAS: Tentative Method, 1983, Methods for the Examination of Waters and Associated Materials, Department of the Environment and National Water Council, London, 1983.
169. K.C. Thompson, P.M. Cummings, 'Some Problems Observed in the Determination of Li in Wastewaters by AAS', *Analyst*, **109**(4), 511–514 (1984).
170. N.J. Miller-Ihli, 'Automated Ultrasonic Mixing Assessory for Slurry Sampling into a GFAAS', *J. Anal. At. Spectrom.*, **4**(3), 295–297 (1989).
171. D.C. Manning, W. Slavin, 'Determination of Trace Elements in Natural Waters Using STPF', *Appl. Spectrosc.*, **37**(1), 1–11 (1983).
172. B. Welz, G. Schlemmer, J.R. Mudakavi, 'Pd Nitrate–Mg Nitrate Modifier for GFAAS Part 2. Determination of As, Cd, Cu, Mn, Pb, Sb, Se and Tl in Water', *J. Anal. At. Spectrom.*, **3**(5), 695–701 (1988).
173. H.-W. Sinemus, D. Maier, B. Radziuk, 'Zehn Jahre Spurenelementbestimmung im Bodenseewasser', in *4. Colloquium Atomspektrometrische Spurenanalytik*, ed. B. Welz, Bodenseewerk Perkin-Elmer, Überlingen, 607–615, 1987.
174. U. Völlkopf, Z. Grobenski, B. Welz, 'The STPF with ZBC for Trace Analysis in Wastewater', *At. Spectrosc.*, **4**(5), 165–170 (1983).
175. C. Minoia, L. Vescovi, S. Canedoli, A. Ronchi, P. Apostoli, L. Pozzoli, E. Sabbioni, 'Direct Determination of Trace Elements in Mineral Waters by GFAAS Using Zeeman Correction', in *Applications of Zeeman Graphite Furnace Atomic Absorption Spectrometry in the Chemical Laboratory and in Toxicology*, eds. C. Minoia, S. Caroli, Pergamon Press, Oxford, 209–226, 1992.
176. F.M. Batifol, C.F. Boutron, 'Atmospheric Heavy Metals in High Altitude Surface Snows from Mont Blanc, French Alps', *Atmos. Environ.*, **18**(11), 2507–2515 (1984).
177. J.C. Latino, D.C. Sears, F. Portala, I.L. Shuttler, 'The Simultaneous Determination of Dissolved Ag, Cd, Pb, and Sb in Potable Waters by ETAAS', *At. Spectrosc.*, **16**(3), 121–126 (1995).
178. M. Plessi, A. Monzani, 'Al Determination in Bottled Mineral Waters by ETAAS', *J. Food Compos. Anal.*, **8**(1), 21–26 (1995).
179. P. Bermejo-Barrera, E. Beceiro-González, A. Bermejo-Barrera, 'Use of Pd–Mg Nitrate as Matrix Modifier for the Determination of Al in Water by ETAAS', *Microchem. J.*, **45**(1), 90–96 (1992).
180. R.W. Dabeka, H.B.S. Conacher, J. Salminen, G.R. Nixon, G. Riedel, R. Crocker, G. Dube, 'Survey of Bottled Drinking Water Sold in Canada. I. Pb, Cd, As, Al, and Fluoride', *J. AOAC Int.*, **75**(6), 949–953 (1992).
181. A.D. Eaton, 'Determining the Practical Quantitation Level for As', *J. Am. Water Works Assoc.*, **86**, 100–114 (1994).
182. A. Criaud, C. Foulliac, 'Use of the L'vov Platform and Mo Coating for the Determination of Volatile Elements in Thermomineral Waters by AAS', *Anal. Chim. Acta*, **167**(1), 257–267 (1985).
183. G. Bozsai, Z. Kárpáti, 'Interference Study for the Examination of As and Se in Natural Water by STPF–Zeeman

AAS Using Pd as a Matrix Modifier', *Acta Chim. Hung.*, **126**(3), 377–383 (1989).
184. R.P. van der Geugten, 'Determination of B in River Waters with Flameless AAS (Graphite Furnace Technique)', *Fresenius' Z. Anal. Chem.*, **306**(1), 13–14 (1981).
185. M. Luguera, Y. Madrid, C. Cámara-Rica, 'Combination of Chemical Modifiers and Graphite Tube Pretreatment to Determine B by ETAAS', *J. Anal. At. Spectrom.*, **6**(8), 669–671 (1991).
186. C. Minoia, S. Caroli, 'Study of Ba Distribution in Waters for Human Consumption by Zeeman GFAAS', in *Applications of Zeeman Graphite Furnace Atomic Absorption Spectrometry in the Chemical Laboratory and in Toxicology*, eds. C. Minoia, S. Canedoli, L. Vescovi, E. Rizzio, E. Sabbioni, R. Pietra, L. Manzo, Pergamon Press, Oxford, 179–207, 1992.
187. D.A. Lytle, M.R. Schock, N.R. Dues, J.U. Doerger, 'Determining Be in Drinking Water by GFAAS', *J. Am. Water Works Assoc.*, **85**, 77–83 (1993).
188. J. Koreckova-Sysalova, 'Determination of Be in Natural Waters Using AAS with Ta-coated Graphite Tube', *Int. J. Environ. Anal. Chem.*, **68**(4), 397–404 (1997).
189. L.-Z. Jin, Z.-M. Ni, 'Matrix Modification for the Determination of Trace Amounts of Bi in Wastewater, Seawater and Urine by GFAAS', *Can. J. Spectrosc.*, **26**(5), 219–223 (1981).
190. J.-Y. Cabon, A. Le Bihan, 'Influence of Seawater Salts on the Determination of Trace Metals in Rainwater by ETAAS', *Environ. Technol.*, **12**, 769–776 (1991).
191. H.-C. Hansson, A.-K. Ekholm, H.B. Ross, 'Rain Water Analysis: Comparison Between PIXE and GFAAS', *Environ. Sci. Technol.*, **22**(5), 527–531 (1988).
192. S.R. Marin, S.G. Olave, O.E. Andonie, O.G. Arlegui, 'Comparison of Analytical Techniques for the Determination of Trace Elements in Rain Water', *Int. J. Environ. Anal. Chem.*, **52**, 127–136 (1993).
193. M.M. Reddy, M.A. Benefiel, H.C. Claassen, 'Cd, Cu, Pb and Zn Determination in Precipitation: A Comparison of ICP-AES and GFAAS', *Mikrochim. Acta*, **1**, 159–170 (1986).
194. S.J. Nagourney, M. Heit, D.C. Bogen, 'ETAAS Analysis of Lake Waters for Mn, Fe, Pb, and Cd', *Talanta*, **34**(5), 465–472 (1987).
195. D.T.E. Hunt, D.A. Winnard, 'Appraisal of Selected Techniques for the Determination of Pb and Cd in Waters by GFAAS', *Analyst*, **111**(7), 785–789 (1986).
196. J.L. Manzoori, A. Saleemi, 'Determination of Cr in Serum and Lake Water by ETAAS Using V and Mo Modifier', *J. Anal. At. Spectrom.*, **9**(3), 337–339 (1994).
197. C.A. Schneider, J.T. McCaffrey, 'Using Fast Furnace Techniques to Meet the Pb and Cu Rule for Drinking Water', *At. Spectrosc.*, **14**(3), 65–70 (1993).
198. R.J. Emerick, 'Calcium Chloride Matrix Modification for Preventing Sulfate Interference in the Determination of Mo by GFAAS', *At. Spectrosc.*, **8**(2), 69–71 (1987).
199. E.W. Loya, 'Graphite Furnace Determination of Mo by Pd–Hydroxylamine Hydrochloride Matrix Modification', *At. Spectrosc.*, **10**(2), 61–65 (1989).
200. B. Pohl, G. Weichbrodt, S. Fraunhofer, 'New Realizations in Determination of Sb in Water and Wastewater Samples', *GIT Fachz. Lab.*, **39**(12), 1134–1135 (1995).
201. K.S. Subramanian, 'Sb in Drinking Water, Red Blood Cells, and Serum: Development of Analytical Methodology Using THGA/AAS', *Arch. Environ. Contam. Toxicol.*, **32**(4), 431–435 (1997).
202. L.-C. Tian, F. Peng, X.-R. Wang, 'The Determination of Se Environmental Background Value in Natural River Water Using Mixed Matrix Modifier–STPF', *Anal. Sci.*, **7**(Suppl.), 1155–1158 (1991).
203. K. Matsusaki, T. Yamaguchi, T. Sata, 'Matrix Modification for the Determination of Si by AAS Using a Graphite Furnace Coated with Ti Carbide', *Anal. Sci.*, **12**(5), 797–800 (1996).
204. P. Bermejo-Barrera, R.M. Soto-Ferreiro, A. Bermejo-Barrera, 'Direct Determination of Sn in Tap Waters by ETAAS', *Fresenius' J. Anal. Chem.*, **345**(1), 60–62 (1993).
205. N.S. Thomaidis, E.A. Piperaki, 'Comparison of Chemical Modifiers for the Determination of V in Water and Oil Samples by ETAAS', *Analyst*, **121**(2), 111–117 (1996).
206. B. Welz, G. Bozsai, M. Sperling, B. Radziuk, 'Pd Nitrate–Mg Nitrate Modifier for ETAAS Part 4. Interference of Sulfate in the Determination of Se', *J. Anal. At. Spectrom.*, **7**(3), 505–509 (1992).
207. F. Fehse, 'The Determination of Si in Deionized Process Water by GFAAS', *Spectrochim. Acta, Part B*, **39**(4), 597–598 (1984).
208. P. Bermejo-Barrera, E. Beceiro-González, A. Bermejo-Barrera, 'Determination of V in Water by AAS with Electrothermal Atomization and Using Hot Injection and Preconcentration on the Graphite Tube', *Anal. Chim. Acta*, **236**, 475–477 (1990).
209. N.-W. Kuo, C.-R. Lan, Z.B. Alfassi, 'In Situ Preconcentration of Trace Metals in High Purity Water onto Graphite Tube by Multiple Injections Followed by GFAAS', *J. Radioanal. Nucl. Chem.*, **172**(1), 117–123 (1993).
210. J.-P. Candelone, S.-M. Hong, C.F. Boutron, 'An Improved Method for Decontaminating Polar Snow or Ice Cores for Heavy Metal Analysis', *Anal. Chim. Acta*, **299**(1), 9–16 (1994).
211. U. Görlach, C.F. Boutron, 'Preconcentration of Pb, Cd, Cu and Zn in Water at the pg g^{-1} Level by Nonboiling Evaporation', *Anal. Chim. Acta*, **236**(2), 391–398 (1990).
212. C.F. Boutron, U. Görlach, J.-P. Candelone, M.A. Bolshov, R.J. Delmas, 'Decrease in Anthropogenic Pb, Cd and Zn in Greenland Snows Since the Late 1960s', *Nature (London)*, **353**, 153–156 (1991).

213. E. Nakamura, M. Watanabe, H. Namiki, 'Preconcentration for the Determination of Trace Metals in Highly Purified Water', *Bunseki Kagaku*, **46**(8), 669–675 (1997).
214. D. Baralkiewicz, J. Siepak, 'Determination of Trace Amounts of Mo in Water Samples by GFAAS with Multiple Injections and Cool Down Step', *Anal. Chim. Acta*, **353**(1), 85–89 (1997).
215. P. Bermejo-Barrera, J. Moreda-Piñeiro, A. Moreda-Piñeiro, A. Bermejo-Barrera, 'Cr Determination in Seawater by ETAAS Using ZBC and a Multi-injection Technique', *Fresenius' J. Anal. Chem.*, **360**(2), 208–212 (1998).
216. P. Bermejo-Barrera, J. Moreda-Piñeiro, A. Moreda-Piñeiro, A. Bermejo-Barrera, 'Usefulness of the Chemical Modification and the Multi-injection Technique Approaches in the ETAAS Determination of Ag, As, Cd, Cr, Hg, Ni and Pb in Seawater', *J. Anal. At. Spectrom.*, **13**(8), 777–786 (1998).
217. P. Bermejo-Barrera, J. Moreda-Piñeiro, A. Moreda-Piñeiro, A. Bermejo-Barrera, 'Direct ETAAS Determination of Ni in Seawater Using Multiple Hot Injection and Zeeman Correction', *Talanta*, **45**(5), 807–815 (1998).
218. C. Han, S.-D. Huang, 'Direct Determination of Cd in Seawater with a GFAA Spectrometer', *Spectrochim. Acta, Part B*, **49**(3), 283–288 (1994).
219. I.L. Shuttler, H. Schulze, 'Multi-element ETAAS Becomes a Reality', *Anal. Eur.*, **1**(1), 44–47 (1994).
220. 'Seawater Reference Material for Trace Metals, NASS-1', Information Sheet, National Research Council Canada, Ottawa, May 1982.
221. P. Bermejo-Barrera, J. Moreda-Piñeiro, A. Moreda-Piñeiro, A. Bermejo-Barrera, 'Study of Chemical Modifiers for Direct Determination of Ag in Seawater by ETAAS with Deuterium Background Correction', *Talanta*, **43**(1), 35–44 (1996).
222. P. Bermejo-Barrera, J. Moreda-Piñeiro, A. Moreda-Piñeiro, A. Bermejo-Barrera, 'Comparison of Different Chemical Modifiers for the Direct Determination of As in Seawater by ETAAS', *Fresenius' J. Anal. Chem.*, **355**(2), 174–179 (1996).
223. E. Alvarez-Cabal Cimadevilla, K. Wróbel, A. Sanz-Medel, 'Capabilities and Limitations of Different Techniques in ETAAS for Direct Monitoring of As, Cd and Pb Contamination of Seawater', *J. Anal. At. Spectrom.*, **10**(2), 149–154 (1995).
224. J.K.B. Bishop, 'Determination of Ba in Seawater Using V/Silicon Modifier and Direct Injection GFAAS', *Anal. Chem.*, **62**(6), 553–557 (1990).
225. C.-R. Lan, 'Direct Determination of Cd in Seawater by ETAAS with NaOH as a Chemical Modifier', *Analyst*, **118**(2), 189–192 (1993).
226. S. Sachsenberg, T. Klenke, W.E. Krumbein, H.J. Schellnhuber, E. Zeeck, 'Direct GFAAS Determination of Metals in Seawater: Application of Pd Modifiers and a Fractal Approach to Their Analytical Support', *Anal. Chim. Acta*, **279**, 241–251 (1993).
227. M. Hoenig, E. Puskaric, P. Choisy, M. Wartel, 'Direct Determination of High- and Mid-volatile Elements (Cd, Pb and Mn) in Seawater by ETAAS: Existing Approaches and Critical Parameters', *Analusis*, **19**(9), 285–291 (1991).
228. S.C. Apte, S.D.W. Comber, M.J. Gardner, A.M. Gunn, 'Direct Determination of Cr in Estuarine and Coastal Waters by ETAAS', *J. Anal. At. Spectrom.*, **6**(2), 169–172 (1991).
229. J.-Y. Cabon, A. Le Bihan, 'The Determination of Cr, Cu and Mn in Seawater with Transversely Heated GFAAS', *Spectrochim. Acta, Part B*, **50**(13), 1703–1716 (1995).
230. S.-D. Huang, W.-R. Lai, K.-Y. Shih, 'Direct Determination of Mo, Cr and Mn in Seawater by GFAAS', *Spectrochim. Acta, Part B*, **50**, 1237–1246 (1995).
231. J.-Y. Cabon, A. Le Bihan, 'Cu Signals from Seawater Matrices in ETAAS. Part 1. Study of the Effects of Principal Inorganic Ions', *Anal. Chim. Acta*, **198**, 87–101 (1987).
232. J.-Y. Cabon, A. Le Bihan, 'Cu Signals from Seawater Matrices in ETAAS. Part 2. Model for Phenomena Observed with Seawater', *Anal. Chim. Acta*, **198**, 103–111 (1987).
233. S.-D. Huang, K.-Y. Shih, 'Direct Determination of Cu in Seawater with a GFAA Spectrometer', *Spectrochim. Acta, Part B*, **48**(12), 1451–1460 (1993).
234. P.-g. Su, S.-d. Huang, 'Direct and Simultaneous Determination of Cu and Mn in Seawater with a Multielement GFAA Spectrometer', *Spectrochim. Acta, Part B*, **53**(5), 699–708 (1998).
235. C.-R. Lan, Z.B. Alfassi, 'Direct Determination of Mn in Seawater by ETAAS with NaOH as Chemical Modifier for Interference Removal', *Analyst*, **119**(5), 1033–1035 (1994).
236. M. Tominaga, K. Bansho, Y. Umezaki, 'ETAAS Determination of Pb, Mn, V and Mo in Seawater with Ascorbic Acid to Reduce Matrix Effects', *Anal. Chim. Acta*, **169**, 171–177 (1985).
237. J.Y. Cabon, A. Le Bihan, 'Determination of Pb in Seawater by ETAAS with Transversely Heated Furnace by Using Oxalic Acid or Pd/Mg Modifiers', *Spectrochim. Acta, Part B*, **51**(9–10), 1245–1251 (1996).
238. E. Iwamoto, H. Shimazu, K. Yokata, T. Kumamaru, 'Atomization of Sn in Saline Water Media in GFAAS with a Tungsten-coated Tube Using Pd as a Chemical Modifier', *Anal. Chim. Acta*, **274**(2), 231–235 (1993).
239. S.-D. Huang, K.Y. Shih, 'Direct Determination of Zn in Seawater by GFAAS', *Spectrochim. Acta, Part B*, **50**(8), 837–846 (1995).
240. P. Tschöpel, L. Kotz, W. Schulz, M. Veber, G. Tölg, 'Causes and Elimination of Systematic Errors in the Determination of Elements in Aqueous Solutions in the ng/ml and pg/ml Range', *Fresenius' Z. Anal. Chem.*, **302**(1), 1–14 (1980).
241. K. Gretzinger, L. Kotz, P. Tschöpel, G. Tölg, 'Causes and Elimination of Systematic Errors in Determination

of Fe and Co in Aqueous Solutions in the ng/ml and pg/ml Range', *Talanta*, **29**(11), 1011–1018 (1982).
242. P. Tschöpel, 'Modern Strategies in the Determination of Very Low Concentrations of Elements in Inorganic and Organic Materials', *Pure Appl. Chem.*, **54**(4), 913–925 (1982).
243. M. Schuster, S. Ringmann, R. Gärtner, X. Lin, J. Dahmen, 'Studies on a New Cleaning Procedure for PFA Surfaces', *Fresenius' J. Anal. Chem.*, **357**(3), 258–265 (1997).
244. F.L.L. Muller, J.D. Burton, P.J. Statham, 'Long-term Changes in the Adsorptive Properties of FEP Separating Funnels Used in a Mixed Dithiocarbamate–Freon–TF Extraction System', *Anal. Chim. Acta*, **245**(1), 21–25 (1991).
245. J.C. Godfrey, M.J. Slater (eds.), *Liquid–Liquid Extraction Equipment*, John Wiley & Sons, New York, 1995.
246. J. Rydberg, C. Musikas, G.R. Choppin (eds.), *Principles and Practices of Solvent Extraction*, Marcel Dekker, New York, 1992.
247. A. Townshend, E. Jackwerth, 'Precipitation of Major Constituents for Trace Preconcentration: Potential and Problems', *Pure Appl. Chem.*, **61**(9), 1643–1656 (1989).
248. C.J. Kantipuly, S. Katragadda, A. Chow, H.D. Gesser, 'Chelating Polymers and Related Supports for Separation and Preconcentration of Trace Metals', *Talanta*, **37**(5), 491–517 (1990).
249. R.E. Sioda, G.E. Batley, W. Lund, J. Wang, S.C. Leach, 'Electrolytic Preconcentration in Instrumental Analysis', *Talanta*, **33**(5), 421–428 (1986).
250. A.B. Volynskii, B.Ya. Spivakov, Y.A. Zolotov, 'Solvent Extraction–ETAAS Analysis', *Talanta*, **31**(6), 449–458 (1984).
251. M.J. Dudas, 'Effect of Drying Parameters on the Heated Graphite Furnace Determination of Trace Metals in MIBK–APDC Solutions', *At. Absorpt. Newsl.*, **13**(3), 67–69 (1974).
252. A. Hulanicki, E. Bulska, 'Effect of Organic Solvents on Atomization of Ni, Cu and Pb in GFAAS', *Can. J. Spectrosc.*, **29**(6), 148–152 (1984).
253. K.-R. Sperling, 'Determination of Heavy Metals in Seawater and Marine Organisms by Flameless AAS. XIV: Comments on the Usefulness of Organohalides as Solvents for the Extraction of Heavy Metal (Cd) Complexes', *Fresenius' Z. Anal. Chem.*, **310**(3–4), 254–256 (1982).
254. R.J. Magee, J.O. Hill, 'The Analytical Chemistry of Metal Dithiocarbamate Complexes', *Rev. Anal. Chem.*, **8**(1–2), 5–72 (1985).
255. K.W. Bruland, R.P. Franks, G.A. Knauer, J.H. Martin, 'Sampling and Analytical Methods for the Determination of Cu, Cd, Zn, and Ni at the Nanogram per Liter Level in Seawater', *Anal. Chim. Acta*, **105**, 233–245 (1979).
256. L.-G. Danielsson, B. Magnusson, S. Westerlund, 'An Improved Metal Extraction Procedure for the Determination of Trace Metals in Seawater by AAS with Electrothermal Atomization', *Anal. Chim. Acta*, **98**, 47–57 (1978).
257. L. Rasmussen, 'Determination of Trace Metals in Seawater by CHELEX-100 or Solvent Extraction Techniques and AAS', *Anal. Chim. Acta*, **125**, 117–130 (1981).
258. K.W. Bruland, K.H. Coale, L. Mart, 'Analysis of Seawater for Dissolved Cd, Cu and Pb: An Inter-comparison of Voltammetric and AA Methods', *Mar. Chem.*, **17**(4), 285–300 (1985).
259. J.-M. Lo, J.C. Yu, F.I. Hutchison, C.M. Wai, 'Solvent Extraction of Dithiocarbamate Complexes and Back-extraction with Hg(II) for Determination of Trace Metals in Seawater by AAS', *Anal. Chem.*, **54**(14), 2536–2539 (1982).
260. Y. Shijo, M. Suzuki, T. Shimizu, S. Aratake, N. Uehara, 'Determination of Trace Amounts of Mo in Rainwater and Snow by GFAAS After Solvent Extraction and Micro-volume Back-extraction', *Anal. Sci.*, **12**(6), 953–957 (1996).
261. R. Eller, F. Alt, G. Tölg, H.J. Tobschall, 'An Efficient Combined Procedure for the Extreme Trace Analysis of Au, Pt, Pd and Rh with the Aid of GFAAS and TRXRF Analysis', *Fresenius' Z. Anal. Chem.*, **334**, 723–739 (1989).
262. J.B. McHugh, 'Determination of Au in Water in the Nanogram Range by Electrothermal Atomization After Coprecipitation with Te', *At. Spectrosc.*, **4**(2), 66–68 (1993).
263. K. Akatsuka, N. Nobuyama, I. Atsuya, 'AAS of Nanogram Amounts of Cd, Pb and Zn After Precipitation with 8-Quinolinol', *Anal. Sci.*, **4**(3), 281–285 (1988).
264. K. Akatsuka, N. Nobuyama, I. Atsuya, 'Preconcentration of Submicrogram Co with 8-Quinolinol from Natural Waters for ETAAS', *Anal. Sci.*, **5**(4), 475–479 (1989).
265. K. Akatsuka, I. Atsuya, 'Preconcentration by Coprecipitation of Submicrogram Amounts of Cu and Mn with 8-Quinolinol and Direct ETAAS of the Precipitates', *Anal. Chim. Acta*, **202**, 223–230 (1987).
266. M. Kan, T. Nasu, M. Taga, 'Enrichment of Cu in Environmental Water Samples on a Membrane Filter and Determination by GFAAS', *Anal. Sci.*, **7**(Suppl.), 1115–1116 (1991).
267. R. Boniforti, R. Ferraroli, P. Frigieri, D. Heltai, G. Queirazza, 'Intercomparison of Five Methods for the Determination of Trace Metals in Seawater', *Anal. Chim. Acta*, **162**, 33–46 (1984).
268. Z.-s. Chen, M. Hiraide, H. Kawaguchi, 'Preconcentration of Trace Heavy Metals in Water by Coprecipitation with Mg Oxinate for ETAAS', *Mikrochim. Acta*, **124**(1–2), 27–34 (1996).
269. F. Sugimoto, K. Yoshikawa, Y. Maeda, 'Determination of Trace Amounts of Pb in Water by GF/AAS After Ni Diethyldithiocarbamate Coprecipitation and

Dissolution in Ethyl Acetate', *Chem. Express*, **6**(7), 467–470 (1991).

270. Z.-x. Zhuang, C.-l. Yang, X.-r. Wang, P.-y. Yang, B.-l. Huang, 'Preconcentration of Trace Elements from Natural Water with Pd Precipitation', *Fresenius' J. Anal. Chem.*, **355**(3–4), 277–280 (1996).
271. O. Fujino, S. Nishida, H. Togawa, K. Hiraki, 'Determination of Yb in Seawater by GFAAS After Coprecipitation and Solvent Extraction', *Anal. Sci.*, **7**(6), 889–892 (1991).
272. M.N. Quigley, F. Vernon, 'Comparison of Coprecipitation and Chelating Ion Exchange for the Preconcentration of Selected Heavy Metals from Seawater', *Anal. Proc.*, **28**(6), 175–176 (1991).
273. C.P. Weisel, R.A. Duce, J.L. Fasching, 'Determination of Al, Pb and V in North Atlantic Seawater After Coprecipitation with Ferric Hydroxide', *Anal. Chem.*, **56**, 1050–1052 (1984).
274. G. Ciceri, R. Ferraroli, L. Guzzi, 'Determination of Total Cr in Natural Waters by Co-precipitation with Iron Hydroxides and AAS', in *Heavy Metals in the Hydrological Cycle, Conference, Lisbon*, Selper, London, 545–551, 1988.
275. S. Welch, W.B. Lyons, C.A. Kling, 'A Coprecipitation Technique for Determining Trace Metal Concentrations in Iron-rich Saline Solutions', *Environ. Technol.*, **11**(2), 141–144 (1990).
276. M.N. Quigley, F. Vernon, 'Coprecipitation of Ag from Seawater with Hg(II) Sulfide and Determination by ETAAS', *Anal. Proc.*, **29**(1), 1–3 (1992).
277. I. Atsuya, K. Itoh, 'Fundamental Studies on the Coprecipitation of Nanogram Quantities of Some Metals with the Dimethylglyoxime/Ni/1-(2-Pyridylazo)-2-naphthol Complex and Their Direct Determination by AAS Using an Inner Miniature Cup for the Solid Sampling Technique', *Fresenius' Z. Anal. Chem.*, **329**(7), 750–755 (1988).
278. T. Nakamura, H. Oka, M. Ishii, J. Sato, 'Direct Atomization AAS Determination of Be, Cr, Fe, Co, Ni, Cu, Cd, and Pb in Water with Zr Hydroxide Coprecipitation', *Analyst*, **119**(6), 1397–1401 (1994).
279. S. Strassburg, D. Wollenweber, G. Wünsch, 'Contamination Caused by Ion-exchange Resin !? Consequences for Ultra-trace Analysis', *Fresenius' J. Anal. Chem.*, **360**(7–8), 792–794 (1998).
280. M. Schuster, M. Schwarzer, 'A New On-line Column Separation and Preconcentration System for the Selective Determination of Trace and Ultratrace Levels of Pd', *At. Spectrosc.*, **19**(4), 121–128 (1998).
281. M. Sperling, X.-f. Yin, B. Welz, 'Determination of Ultratrace Concentrations of Elements by Means of On-line Solid Sorbent Extraction GFAAS', *Fresenius' J. Anal. Chem.*, **343**, 754–755 (1992).
282. J.B.B. da Silva, M.B.O. Giacomelli, A.J. Curtius, 'Determination of Bi in Al and in Steels by ETAAS After On-line Separation Using a Minicolumn of Activated Carbon', *Analyst*, **124**(8), 1249–1253 (1999).
283. R.A. Nickson, S.J. Hill, P.l J. Worsfold, 'Solid Phase Techniques for the Preconcentration of Trace Metals from Natural Waters', *Anal. Proc.*, **32**(9), 387–395 (1995).
284. Z.-l. Fang, G.-h. Tao, 'New Developments in FI Separation and Preconcentration Techniques for ETAAS', *Fresenius' J. Anal. Chem.*, **355**(5–6), 576–580 (1996).
285. N.-K. Djane, F. Malcus, E. Martins, G. Sawula, G. Johansson, 'Hollow Fiber Cartridges for Removal of Particulate Matter from Natural Waters Prior to Matrix Isolation and Trace Metal Enrichment Using an 8-Quinolinol Chelating Ion Exchanger in a Flow System', *Anal. Chim. Acta*, **316**(3), 305–311 (1995).
286. J.T. Creed, T.D. Martin, 'Determination of Trace Elements in Marine Waters by Off-line Chelation Preconcentration with GFAAS', EPA Method 200.13, National Exposure Research Laboratory, Cincinnati, OH, 1–19, 1997.
287. M. Sperling, X.-p. Yan, B. Welz, 'Investigation of On-line Coupling ETAAS with FI Sorption Preconcentration Using a Knotted Reactor for Totally Automatic Determination of Pb in Water Samples', *Spectrochim. Acta, Part B*, **51**(14), 1891–1908 (1996).
288. M. Sperling, X.-f. Yin, B. Welz, 'FI On-line Separation and Preconcentration for ETAAS Part 1. Determination of Ultratrace Amount of Cd, Cu, Pb and Ni in Water Samples', *J. Anal. At. Spectrom.*, **6**(4), 295–300 (1991).
289. V. Porta, O. Abollino, E. Mentasti, C. Sarzanini, 'Determination of Ultra-trace Levels of Metal Ions in Seawater with On-line Pre-concentration and ETAAS', *J. Anal. At. Spectrom.*, **6**(2), 119–122 (1991).
290. M. Sperling, X.-f. Yin, B. Welz, 'FI On-line Separation and Preconcentration for ETAAS Part 2. Determination of Ultra-trace Amounts of Co in Water', *J. Anal. At. Spectrom.*, **6**(8), 615–621 (1991).
291. R.-l. Ma, W. Van Mol, F. Adams, 'Selective FI Sorbent Extraction for Determination of Cd, Cu and Pb in Biological and Environmental Samples by GFAAS', *Anal. Chim. Acta*, **293**(3), 251–260 (1994).
292. L.C. Azeredo, R.E. Sturgeon, A.J. Curtius, 'Determination of Trace Metals in Seawater by GAAS Following On-line Separation and Preconcentration', *Spectrochim. Acta, Part B*, **48**(1), 91–98 (1993).
293. Y.-H. Sung, Z.-S. Liu, S.-D. Huang, 'Use of Muromac A-1 Chelating Resin for Determination of Cu and Mo in Seawater by AAS with On-line Preconcentration', *Spectrochim. Acta, Part B*, **52**(6), 755–764 (1997).
294. Z.-l. Fang, 'Trends and Potentials in FI On-line Separation and Preconcentration Techniques for ETAAS', *Spectrochim. Acta, Part B*, **53**(10), 1371–1379 (1998).
295. J.Y. Lu, C.L. Chakrabarti, M.H. Back, A.L.R. Sekaly, D.C. Gregoire, W.H. Schroeder, 'Speciation of Some Metals in River Surface Water, Rain and Snow, and the Interactions of These Metals with Selected Soil

Matrices', *J. Anal. At. Spectrom.*, **11**(12), 1189–1201 (1996).
296. P. Burba, B. Aster, T. Nifant'eva, V. Shkinev, B.Ya. Spivakov, 'Membrane Filtration Studies of Aquatic Humic Substances and Their Metal Species: A Concise Overview. Part 1. Analytical Fractionation by Means of Sequential-stage Ultrafiltration', *Talanta*, **45**(5), 977–988 (1998).
297. S.J. De Mora, R.M. Harrison, 'The Use of Physical Separation Techniques in Trace-metal Speciation Studies', *Water Res.*, **17**(7), 723–733 (1983).
298. G.E. Batley, *Trace Element Speciation: Analytical Methods and Problems*, CRC Press, Boca Raton, FL, 1989.
299. I.S. Krull, *Trace Metal Analysis and Speciation*, Elsevier, Amsterdam, 1991.
300. A.M. Ure, C.M. Davidson, *Chemical Speciation in the Environment*, Blackie, Glasgow, 1995.
301. A. Tessier, D.R. Turner, *Metal Speciation and Bioavailability in Aquatic Systems*, John Wiley & Sons, Chichester, 1995.
302. L. Ebdon, S. Hill, 'Interface Between LC and AA', in *Environmental Analysis Using Chromatography Interfaced with Atomic Spectrometry*, eds. R.M. Harrison, S. Rapsomanikis, Ellis Horwood, Chichester, 165–188, 1989.
303. A.K. Das, R. Chakraborty, 'ETAAS in the Study of Metal Ion Speciation', *Fresenius' J. Anal. Chem.*, **357**(1), 1–17 (1997).
304. M. Sperling, X.-f. Yin, B. Welz, 'Differential Determination of As(III) and Total As Using FI On-line Separation and Preconcentration for GFAAS', *Spectrochim. Acta, Part B*, **46**(14), 1789–1801 (1991).
305. E.B. Milosavljevic, L. Solujic, J.H. Nelson, J.L. Hendrix, 'Simultaneous Determination of Cr(VI) and Cr(III) by FAAS with a Chelating Ion-exchange FI System', *Mikrochim. Acta*, **3**, 353–360 (1985).
306. M. Sperling, S.-k. Xu, B. Welz, 'Determination of Cr(III) and Cr(VI) in Water Using FI On-line Preconcentration with Selective Adsorption on Activated Alumina and FAAS Detection', *Anal. Chem.*, **64**(24), 3101–3108 (1992).
307. P.A. Sule, J.D. Ingle, 'Determination of the Speciation of Cr with an Automated Two-column Ion-exchange System', *Anal. Chim. Acta*, **326**, 85–93 (1996).
308. R.M. Cespón-Romero, M.C. Yebra-Biurrun, M.P. Bermejo-Barrera, 'Preconcentration and Speciation of Cr by the Determination of Total Cr and Cr(III) in Natural Waters by FAAS with a Chelating Ion-exchange FI System', *Anal. Chim. Acta*, **327**(3), 37–45 (1996).
309. M. Sperling, X.-F. Yin, B. Welz, 'Differential Determination of Cr(VI) and Total Cr in Natural Waters Using FI On-line Separation and Preconcentration ETAAS', *Analyst*, **117**(3), 629–635 (1992).
310. S. Krekler, W. Frenzel, G. Schulze, 'Simultaneous Determination of Fe(II)/Fe(III) by Sorbent Extraction with FI–AAS Detection', *Anal. Chim. Acta*, **296**(2), 115–117 (1994).
311. X.-p. Yan, W. Van Mol, F. Adams, 'Determination of (Ultra)trace Amounts of Sb(III) in Water by FI On-line Sorption Preconcentration in a Knotted Reactor Coupled with ETAAS', *Analyst*, **121**(8), 1061–1067 (1996).
312. X.-p. Yan, M. Sperling, B. Welz, 'On-line Coupling of FI Microcolumn Separation and Preconcentration to ETAAS for Determination of (Ultra)trace Selenite and Selenare in Water', *Anal. Chem.*, **71**(19), 4353–4360 (1999).
313. B. Patel, P.G. Riby, S.J. Haswell, R. Grzeskowiak, 'FI-FAAS System for the Pre-concentration of V(V) and Characterisation of V(IV) and -(V) Species', *J. Anal. At. Spectrom.*, **4**(2), 195–198 (1989).
314. E.B. Milosavljevic, J. Ruzicka, E.H. Hansen, 'Simultaneous Determination of Free and EDTA-complexed Cu Ions by FAAS with an Ion Exchange FI System', *Anal. Chim. Acta*, **169**(3), 321–324 (1985).
315. L .-G. Danielsson, A. Sparén, 'A Mechanized System for the Determination of Low Levels of Quickly Reacting Al in Natural Waters', *Anal. Chim. Acta*, **306**(2–3), 173–181 (1995).
316. D. Klockow, 'Zum Gegenwärtigen Stand der Probennahme von Spurenstoffen in der freien Atmosphäre', *Fresenius' Z. Anal. Chem.*, **326**, 5–24 (1987).
317. M. Demange, B. Hervé-Bazin, B. Carton, 'Sampling Metals and Metal Compounds in the Vapour Phase', *Analusis*, **19**, 73–78 (1991).
318. J.O. Nriagu, 'A Global Assessment of Natural Sources of Atmospheric Trace Metals', *Nature (London)*, **338**, 47–49 (1989).
319. H. Puxbaum, 'Metal Compounds in the Atmosphere', in *Metals and Their Compounds in the Environment*, ed. E. Merian, VCH, Weinheim, 257–286, 1991.
320. J.O. Nriagu, C.I. Davidson, *Toxic Metals in the Atmosphere*, John Wiley & Sons, New York, 1986.
321. M.S. Cresser, 'Analysis of Polluted Soils', in *AAS – Theory, Design and Applications*, ed. S.J. Haswell, Elsevier, Amsterdam, 515–526, 1991.
322. A. Kabata-Pendias, H. Pendias, *Trace Elements in Soils and Plants*, 2nd edition, CRC Press, Boca Raton, FL, 1992.
323. A.M. Ure, 'Trace Elements in Soil: Their Determination and Speciation', *Fresenius' J. Anal. Chem.*, **337**(5), 577–581 (1990).
324. B.J. Alloway, *Heavy Metals in Soils*, Blackie, 2nd edition, London, 1995.
325. A.J. Horowitz, *A Primer on Sediment-trace Element Chemistry*, 2nd edition, Lewis, Chelsea, MI, 1991.
326. T.R. Crompton, *Analysis of Solids in Natural Waters*, Springer, Berlin, 1996.
327. H. Heinrichs, H.-J. Brumsack, W. Schultz, 'Analysis of Wastes', *LaborPraxis*, **15**(9), 709–715 (1991).
328. U. Osberghaus, 'Waste', in *Probennahme und Aufschluß*, ed. M. Stoeppler, Springer, Berlin, 41–53, 1994.

329. T. Kouimtzis, C. Samara (eds.), *Airborne Particulate Matter*, Springer, Berlin, 1995.
330. M. Kersten, U. Förstner, 'Speciation of Trace Metals in Sediments and Combustion Waste', in *Chemical Speciation in the Environment*, eds. A.M. Ure, C.M. Davidson, Blackie, Glasgow, 234–275, 1995.
331. L. Landner, *Speciation of Metals in Water, Sediment and Soil Systems*, Springer, Berlin, 1987.
332. G.S.P. Ritchie, G. Sposito, 'Speciation in Soils', in *Chemical Speciation in the Environment*, eds. A.M. Ure, C.M. Davidson, Blackie, Glasgow, 201–233, 1995.
333. M.L. Berrow, 'Sampling of Soils and Plants for Trace-element Analysis', *Anal. Proc.*, **25**(4), 116–118 (1988).
334. J.R. Boulding, *Description and Sampling of Contaminated Soils, a Field Guide*, 2nd edition, Lewis, Boca Raton, FL, 1994.
335. A. Gomez, R. Lechber, P.L. Hermite, *Sampling Problems for the Chemical Analysis of Sludge, Soils and Plants*, Elsevier, Amsterdam, 1986.
336. K.G. Liphard, 'Probenahme und Analytik von Fremdstoffen im Boden', *Staub. Reinhalt. Luft*, **47**(1–2), M126 (1987).
337. B. Schumacher, J. Jeffrey Van Ee, E.J. Englund, 'Soil Sampling at Hazardous Waste Sites', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 4498–4517, 1998.
338. J. Kraft, D. Truckenbrodt, J. Einax, 'Investigations of Sampling Representativity and Distribution Homogeneity of River Sediments', *GIT Fachz. Lab.*, **40**(1), 33–36 (1996).
339. D. Truckenbrodt, J. Einax, 'Sampling Representativity and Homogeneity of River Sediments', *Fresenius' J. Anal. Chem.*, **352**(5), 437–443 (1995).
340. A. Mudroch, J.M. Azcue, *Manual of Aquatic Sediment Sampling*, Lewis, Ann Arbor, MI, 1995.
341. A. Mudroch, J.M. Azcue, 'Aquatic Sediment Sampling', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 377–402, 1998.
342. H.H. Rump, B. Scholz, *Untersuchung von Abfällen, Reststoffen und Altlasten-Praktische Anleitung für Chemische, Physikalische und Biologische Methoden*, VCH, Weinheim, 1995.
343. L.P. Provost, E.G. Schweitzer, J.A. Santolucito, *Environmental Sampling for Hazardous Wastes*, American Chemical Society, Washington, DC, 79, 1984.
344. A. Mudroch, *Handbook of Techniques for Aquatic Sediment Sampling*, 2nd edition, Lewis, Ann Arbor, MI, 1994.
345. C.W. Holmes, 'Trace Metal Seasonal Variations in Texas Marine Sediments', *Mar. Chem.*, **20**, 13–27 (1986).
346. H. Hintelmann, R.D. Wilken, 'Levels of Total Mercury and Methylmercury Compounds in Sediments of the Polluted Elbe River: Influence of Seasonally and Spatially Varying Environmental Factors', *Sci. Total Environ.*, **166**, 1–10 (1995).
347. J. Einax, K. Oswald, K. Danzer, 'Analytical Investigations and Chemometrical Characterization of Polluted Soils', *Fresenius' J. Anal. Chem.*, **336**(5), 394–399 (1990).
348. J. Einax, B. Machelett, S. Geiss, K. Danzer, 'Chemometric Investigations on the Representativity of Soil Sampling', *Fresenius' J. Anal. Chem.*, **342**(4–5), 267–272 (1992).
349. M. Krieg, J. Einax, 'Chemometric Characterization of Soil Depth Profiles', *Fresenius' J. Anal. Chem.*, **348**(8–9), 490–495 (1994).
350. A. Lepretre, S. Martin, 'Sampling Strategy of Soil Quality', *Analusis*, **22**(3), M40–M43 (1994).
351. M.H. Ramsey, A. Argyraki, M. Thompson, 'Estimation of Sampling Bias Between Different Sampling Protocols on Contaminated Land', *Analyst*, **120**(5), 1353–1356 (1995).
352. M. Thompson, M. Maguire, 'Estimating and Using Sampling Precision in Surveys of Trace Constituents of Soils', *Analyst*, **118**(9), 1107–1110 (1993).
353. P.M. Gy, 'Sampling Errors. They May Render the Analytical Result Meaningless', *Analusis*, **11**(9), 413–440 (1983).
354. G.M. Kolesov, V. Anikiev, S.K. Prasad, E.M. Sedykh, 'Determination of Trace Elements in Samples of Bottom Sediments, Suspended Matter and Aerosols', *Chem. Anal. (Warsaw)*, **38**(5), 625–637 (1993).
355. B. Daus, H.W. Zwanziger, 'Heavy Metals in River Sediments – a Chemometrical Evaluation', *Fresenius' J. Anal. Chem.*, **352**, 444–450 (1995).
356. P. Gy, *Sampling for Analytical Purposes*, John Wiley & Sons, New York, 1998.
357. G.E. Batley, 'Collection, Preparation, and Storage of Samples for Speciation Analysis', in *Trace Element Speciation: Analytical Methods and Problems*, ed. G.E. Batley, CRC Press, Boca Raton, FL, 1–24, 1989.
358. R. Rubio, A.M. Ure, 'Approaches to Sampling and Sample Pretreatments for Metal Speciation in Soils and Sediments', *Int. J. Environ. Anal. Chem.*, **51**(1–4), 205–217 (1993).
359. M.H. Ramsey, 'Measurement Uncertainty Arising from Sampling: Implications for the Objectives of Geoanalysis', *Analyst*, **122**(11), 1255–1260 (1997).
360. J.W.A. Lustenhouwer, J.A. Hin, F.J.M.J. Maessen, G. Den Boef, G. Kateman, 'Characterization of Compost with Respect to its Content of Heavy Metals. Part III. Precision of the Total Analytical Procedure', *Int. J. Environ. Anal. Chem.*, **44**(2), 103–115 (1991).
361. C.A. Pio, 'General Sampling Techniques', in *Handbook of Air Pollution Analysis*, eds. R.M. Harrison, R. Perry, 2nd edition, Chapman & Hall, New York, 1–93, 1986.
362. J.C. Septon, 'Analysis of Airborne Particles in Workplace Atmospheres', in *AAS – Theory, Design and Applications*, ed. S.J. Haswell, Elsevier, Amsterdam, 159–190, 1991.

363. L.C. Kenny, 'Developments in Workplace Aerosol Sampling. A Review', *Analyst*, **121**(9), 1233–1239 (1996).
364. S. Hetland, Y. Thomassen, 'Sampling and Chemical Characterization of Aerosols in Workplace Air', *Pure Appl. Chem.*, **65**(12), 2417–2422 (1993).
365. K. Polyák, I. Bódog, J. Hlavay, 'Determination of Chemical Species of Selected Trace Elements in Fly Ash', *Talanta*, **41**(7), 1151–1159 (1994).
366. K.R. Spurny, *Physical and Chemical Characterization of Individual Airborne Particles*, Ellis Horwood, Chichester, 1986.
367. J.T. Karlsen, G. Farrants, T. Torgrimsen, A. Reith, 'Chemical Composition and Morphology of Welding Fume Particles and Grinding Dusts', *Am. Ind. Hyg. Assoc. J.*, **53**, 290 (1992).
368. T. Berg, O. Røyset, E. Steinnes, 'Blank Values of Trace Elements in Aerosol Filters Determined by ICP MS', *Atmos. Environ.*, **A27**(15), 2435–2439 (1993).
369. D.A. Dolske, J. Schneider, H. Sievering, 'Trace-element Pass-through for Cellulose Filters When Used for Aerosol Collection', *Atmos. Environ.*, **18**(11), 2557–2558 (1984).
370. J.H. Vincent, *Aerosol Sampling – Science and Practice*, John Wiley & Sons, New York, 1989.
371. J. Hlavay, 'Analysis of Respirable Dust', *Microchem. J.*, **46**(1), 121–129 (1992).
372. H.P. König, R.F. Hertel, W. Koch, G. Rosner, 'Determination of Pt Emissions from a Three-way Catalyst-equipped Gasoline Engine', *Atmos. Environ.*, **A26**(5), 741–745 (1992).
373. B. Docekal, V. Krivan, N. Pelz, 'Trace and Minor Element Characterization of Diesel Soot', *Fresenius' J. Anal. Chem.*, **343**, 873–878 (1992).
374. C.L. Chakrabarti, X.-r. He, S.-l. Wu, W.H. Schroeder, 'Direct Determination of Metals Associated with Airborne Particulates Using a Graphite Probe Collection Technique and Graphite Probe AAS Analysis', *Spectrochim. Acta, Part B*, **42**, 1227–1233 (1987).
375. J. Sneddon, 'Impaction/Electrostatic Precipitation', in *Sample Introduction in Atomic Spectroscopy*, ed. J. Sneddon, Elsevier, Amsterdam, 329–351, 1990.
376. J. Sneddon, M.V. Smith, S. Indurthy, Y.-i. Lee, 'Direct and Near-real-time Multielement Determination of Metals in Aerosols by Impaction-GFAAS', *Spectroscopy (Eugene, Oreg.)*, **10**(1), 26–32 (1995).
377. A.S. Kao, 'Air Sampling Instruments', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 270–285, 1998.
378. J. Medved, V. Stresko, J. Kubova, J. Polakovicova, 'Efficiency of Decomposition Procedures for the Determination of Some Elements in Soils by Atomic Spectroscopic Methods', *Fresenius' J. Anal. Chem.*, **360**(2), 219–224 (1998).
379. J.B. Fournier, V. Vignier, P. Renaud, G.J. Martin, 'Element Analysis Carried Out on Reference Samples. Comparison Between Five Digestion Techniques Using Wet and Dry Processes', *Analusis*, **25**(6), 196–201 (1997).
380. T. Yamashige, M. Yamamoto, H. Sunahara, 'Comparison of Decomposition Methods for the Analysis of Atmospheric Particulates by AAS', *Analyst*, **114**(9), 1071–1077 (1989).
381. D.C. Bartenfelder, A.D. Karathanasis, 'A Comparison Between X-ray Fluorescence and Dissolution Methods Employing Lithium Metaborate Fusion for Elemental Analysis of Soil Clays', *Commun. Soil. Sci. Plant Anal.*, **19**, 471–492 (1988).
382. M. Bettinelli, 'Determination of Trace Metals in Siliceous Standard Reference Materials by ETAAS After Lithium Tetraborate Fusion', *Anal. Chim. Acta*, **148**, 193–201 (1983).
383. A.A. Verbeek, M.C. Mitchell, A.M. Ure, 'Analysis of Small Samples of Rock and Soil by AA and Emission Spectrometry After a Lithium Metaborate Fusion/Nitric Acid Dissolution Procedure', *Anal. Chim. Acta*, **135**, 215–228 (1982).
384. H. Agemian, E. Bedek, 'A Semi-automated Method for the Determination of Total As and Se in Soils and Sediments', *Anal. Chim. Acta*, **119**, 323–330 (1980).
385. P. Woitke, A. Thiesies, G. Henrion, G. Steppuhn, 'Digestion of Sediment Samples for Determination of Heavy Metals by FAAS', *Z. Chem.*, **29**(5), 180–181 (1989).
386. C. Nerín, R. Zufiaurre, J.C. Palomar, 'Critical Study of Several Procedures for Digestion of Fly Ash for Analysis of As, Fe, and Mn', *Analusis*, **20**, 597–600 (1992).
387. G.E.M. Hall, 'Are your REE Results Total?', *Explore*, **68**, 18–20 (1990).
388. A. Hall, 'Determination of Total Sn Content of Some Geological Materials by AAS', *Chem. Geol.*, **30**(1–2), 135–142 (1980).
389. P.J. Lamothe, T.L. Fries, J.J. Consul, 'Evaluation of a Microwave-oven System for the Dissolution of Geological Samples', *Anal. Chem.*, **58**(8), 1881–1886 (1986).
390. K. Naumann, J. Bergmann, W. Dannecker, 'Analytik von Schwebstäuben der Luft mit Hilfe Atomspektrometrischer Methoden', in *Fortschritte in der Atomspektrometrischen Spurenanalytik*, ed. B. Welz, Verlag Chemie, Weinheim, 543–555, 1984.
391. A.J. Aller, D. Bonilla, 'A Digestion Procedure for Determination of Major and Minor Elements in Fly Ash by AAS', *Anal. Sci.*, **6**(2), 309–311 (1990).
392. M.W. Doughten, J.R. Gillison, 'Determination of Selected Elements in Whole Coal and in Coal Ash from the Eight Argonne Premium Coal Samples by AAS–AES and Ion-selective Electrode', *Energy Fuels*, **4**(5), 426–430 (1990).
393. Verein Deutscher Ingenieure eV, 'Stoffbestimmung an Partikeln; Bestimmung der Elemente Ba, Be, Cd, Co, Cr, Cu, Ni, Pb, Sr, V, Zn in emittierten Stäuben Mittels Atomspektrometrischer Methoden', Richtlinie VDI-2268/1, *VDI-Handbuch Reinhaltung der Luft*, VDI, Beuth Verlag, Berlin, 1–14, 1987.

394. H.-J. Abraham, C. Schlums, 'Determination of Metals and Metalloids in Homogenized Lignite Ash', *Staub, Reinhalt, Luft*, **52**(9), 353–360 (1992).
395. M. Sager, 'Determination of As, Cd, Hg, Sb, Tl and Zn in Coal and Coal Fly-ash', *Fuel*, **72**(9), 1327–1330 (1993).
396. L. Ebdon, J.R. Wilkinson, 'The Determination of As and Se in Coal by Continuous Flow HGAAS and AFS', *Anal. Chim. Acta*, **194**, 177–187 (1987).
397. P.C. Lindahl, A.M. Bischop, 'Determination of Trace Elements in Coal by an Oxygen Bomb Combustion AAS Method', *Fuel*, **61**, 658–662 (1982).
398. W. Mniszek, 'CVAAS for Total Hg Determination in Coal Sample after Oxygen Combustion', *Chem. Anal. (Warsaw)*, **41**(2), 269 (1996).
399. M. Bettinelli, 'Fusion Procedure for the Trace Metal Analysis of Coal by AA', *At. Spectrosc.*, **4**(1), 5–9 (1983).
400. J.C. Mills, C.B. Belcher, 'Analysis of Coal, Coke, Ash and Mineral Matter by Atomic Spectroscopy', *Prog. Anal. Spectrosc.*, **4**, 49–88 (1981).
401. M.L. Berrow, W.M. Stein, 'Extraction of Metals from Soils and Sewage Sludges by Refluxing with Aqua Regia', *Analyst*, **108**, 277–285 (1983).
402. T.H. Christensen, 'Comparison of Methods for Preparation of Municipal Compost for Analysis by AAS', *Int. J. Environ. Anal. Chem.*, **12**(3–4), 211–221 (1982).
403. T.H. Christensen, L.I. Pedersen, J.C. Tjell, 'Comparison of Four Methods of Treating Sewage Sludge Samples for Analysis of Metals by AAS', *Int. J. Environ. Anal. Chem.*, **12**(1), 41–50 (1982).
404. E. Macalalad, R. Bayoran, B. Ebarvia, I. Rubeska, 'A Concise Analytical Scheme for 16 Trace Elements in Geochemical Exploration Samples Using Exclusively AAS', *J. Geochem. Explor.*, **30**, 167–177 (1988).
405. I.L. Marr, P. Kluge, L. Main, V. Margerin, C. Lescop, 'Digests or Extracts? – Some Interesting but Conflicting Results for Three Widely Differing Polluted Sediment Samples', *Mikrochim. Acta*, **119**(3–4), 219–232 (1995).
406. J.J. McAlister, K.R. Hirons, 'Total Elemental Analysis of Lake Sediments in Palaeoecology: An Investigation of Sample Preparation Techniques for AAS', *Microchem. J.*, **30**, 79–91 (1984).
407. F.E. Smith, E.A. Arsenault, 'Microwave-assisted Sample Preparation in Analytical Chemistry', *Talanta*, **43**(8), 1207–1268 (1996).
408. K.J. Lamble, S.J. Hill, 'Microwave Digestion Procedures for Environmental Matrices', *Analyst*, **123**(7), 103R–133R (1998).
409. H.M. Kingston, P.J. Walter, S.J. Chalk, E. Loretzen, D.D. Link, 'Environmental Microwave Sample Preparation: Fundamentals, Methods, and Applications', in *Microwave Enhanced Chemistry*, eds. H.M. Kingston, S. Haswell, American Chemical Society, Washington, DC, 223–349, 1997.
410. H.M. Kingston, P.J. Walter, 'Comparison of Microwave Versus Conventional Dissolution for Environmental Applications', *Spectroscopy (Eugene, Oreg.)*, **7**(9), 20–27 (1992).
411. B.G. Kratochvil, S. Mamba, 'Microwave Acid Dissolution of Soil Samples for Elemental Analysis', *Can. J. Chem.*, **68**(2), 360–362 (1990).
412. A.D. Hewitt, C.M. Reynolds, 'Dissolution of Metals from Soils and Sediments with a Microwave–Nitric Acid Digestion Technique', *At. Spectrosc.*, **11**(5), 187–192 (1990).
413. J. Huang, D. Goltz, F. Smith, 'A Microwave Dissolution Technique for the Determination of As in Soils', *Talanta*, **35**(11), 907–908 (1988).
414. J. Alvarado, L.E. León, F. López, C. Lima, 'Comparison of Conventional and Microwave Wet Acid Digestion Procedures for the Determination of Fe, Ni and V in Coal by ETAAS', *J. Anal. At. Spectrom.*, **3**(1), 135–138 (1988).
415. Y. Feng, R.S. Barratt, 'Digestion of Dust Samples in a Microwave Oven', *Sci. Total Environ.*, **143**, 157–161 (1994).
416. A.D. Hewitt, J.H. Cragin, 'Comment on 'Acid Digestion for Sediments, Sludges, Soils, and Solid Wastes. A Proposed Alternative to EPA SW 846 Method 3050' ', *Environ. Sci. Technol.*, **25**(5), 985–986 (1991).
417. C.G. Millward, P.D. Kluckner, 'Microwave Digestion Technique for the Extraction of Minerals from Environmental Marine Sediments for Analysis by ICP-AES and AAS', *J. Anal. At. Spectrom.*, **4**(9), 709–713 (1989).
418. J. Nieuwenhuize, C.H. Poley-Vos, A.H. van den Akker, W. van Delft, 'Comparison of Microwave and Conventional Extraction Techniques for the Determination of Metals in Soil, Sediment and Sludge Samples by Atomic Spectrometry', *Analyst*, **116**(4), 347–351 (1991).
419. M. Mogg, J. Rudolf, 'Microwave or Classical Digestion?', *LaborPraxis*, **18**(10), 22–27 (1994).
420. M. Bettinelli, U. Baroni, N. Pastorelli, 'Microwave Oven Sample Dissolution for the Analysis of Environmental and Biological Materials', *Anal. Chim. Acta*, **225**(1), 159–174 (1989).
421. M. Bettinelli, U. Baroni, 'A Microwave Oven Digestion Method for the Determination of Metals in Sewage Sludges by ICP-AES and GFAAS', *Int. J. Environ. Anal. Chem.*, **43**(1), 33–40 (1990).
422. H. Matusiewicz, R.E. Sturgeon, S.S. Berman, 'Vapour-phase Acid Digestion of Inorganic and Organic Matrices for Trace Element Analyses Using a Microwave Heated Bomb', *J. Anal. At. Spectrom.*, **6**(4), 283–287 (1991).
423. M. Bettinelli, U. Baroni, N. Pastorelli, 'Determination of As, Cd, Pb, Sb, Se and Tl in Coal Fly Ash Using the STPF and ZBC', *J. Anal. At. Spectrom.*, **3**(7), 1005–1011 (1988).
424. S. Nakashima, R.E. Sturgeon, S.N. Willie, S.S. Berman, 'Acid Digestion of Marine Samples for Trace Element Analysis Using Microwave Heating', *Analyst*, **113**(1), 159–163 (1988).

425. C.-y. Zhou, M.-k. Wong, L.-l. Koh, Y.-c. Wee, 'Orthogonal Array Design for the Optimization of Closed-vessel Microwave Digestion Parameters for the Determination of Trace Metals in Sediments', *Anal. Chim. Acta*, **314**(1–2), 121–130 (1995).
426. M. Zischka, P. Kettisch, A. Schalk, G. Knapp, 'Closed Vessel Microwave-assisted Wet Digestion with Simultaneous Control of Pressure and Temperature in All Vessels', *Fresenius' J. Anal. Chem.*, **361**(2), 90–95 (1998).
427. G. Knapp, F. Panholzer, A. Schalk, P. Kettisch, 'Pressure-controlled Microwave-assisted Wet Digestion Systems', in *Microwave-enhanced Chemistry*, eds. H.M. Kingston, S.J. Haswell, American Chemical Society, Washington, DC, 423–451, 1997.
428. J. Baasner, H. Schulze, P. Kainrath, P. Kettisch, 'A New Generation of High-pressure Microwave Sample Digestion Systems', *Spectroscopy (Eugene, Oreg.)*, **14**(6), 42–44 (1999).
429. M. de La Guardia, V. Carbonell, A. Morales-Rubio, A. Salvador, 'On-line Microwave-assisted Digestion of Solid Samples for Their FAAS Analysis', *Talanta*, **40**(11), 1609–1617 (1993).
430. T.J. Gluodenis, J.F. Tyson, 'FI Systems for Directly Coupling On-line Digestions with Analytical Atomic Spectrometry. Part 1. Dissolution of Cocoa Under Stopped-flow, High-pressure Conditions', *J. Anal. At. Spectrom.*, **7**(2), 301–306 (1992).
431. T.J. Gluodenis, J.F. Tyson, 'FI Systems for Directly Coupling On-line Digestions with Analytical Atomic Spectrometry Part 2. Reactions in a Microwave Field', *J. Anal. At. Spectrom.*, **8**(5), 697–704 (1993).
432. J.L. Burguera, M. Burguera, 'Determination of Pb in Biological Materials by Microwave-assisted Mineralization and FI–ETAAS', *J. Anal. At. Spectrom.*, **8**(2), 235–241 (1993).
433. S. Saeki, M. Kubota, T. Asami, 'Determination of Ag in Soils by AAS', *Water, Air, Soil Pollut.*, **83**(3–4), 253 (1995).
434. R.A. Nadkarni, R.I. Botto, S.E. Smith, 'Comparison of Two Spectroscopic Methods for Elemental Analysis of Geological Materials', *At. Spectrosc.*, **3**(6), 180–184 (1982).
435. 'Methods for the Analysis and Testing of Coal and Coke. Part 10.1. Determination of Trace Elements', Australian Standard AS-1038-1986, Standards Association of Australia, Sydney, 1986.
436. A.M. Hartstein, R.W. Freedman, D.W. Platter, ' Novel Wet-digestion Procedure for Trace-metal Analysis of Coal by Atomic Absorption', *Anal. Chem.*, **45**(3), 611–614 (1973).
437. C. Maqueda, E. Morillo, 'Determination of Ca by AAS in Samples Dissolved by Acid Mixtures', *Fresenius' J. Anal. Chem.*, **338**(3), 253–254 (1990).
438. A. Morales-Rubio, F. Pomares, M. de la Guardia, A. Salvador, 'Determination of Cd, Cu, Fe, Mn, Pb and Zn in Sewage Sludges with Prior Acid Digestion in a Microwave Oven and Slurry Introduction', *J. Anal. At. Spectrom.*, **4**(4), 329–332 (1989).
439. R.-l. Ma, W. Van Mol, F. Adams, 'Determination of Cd, Cu and Pb in Environmental Samples. An Evaluation of FI On-line Sorbent Extraction for FAAS', *Anal. Chim. Acta*, **285**(1–2), 33–43 (1994).
440. E. De Vevey, G. Bitton, D. Rossel, L.D. Ramos, L. Mungia Guerrero, J. Tarradellas, 'Concentration and Bioavailability of Heavy Metals in Sediments in Lake Yojoa (Honduras)', *Bull. Environ. Contam. Toxicol.*, **50**(2), 253–259 (1993).
441. O. Elsholz, 'Determination of Cd, Pb, Hg and As in Soil and Waste by FI/AAS', in *CANAS'93, Colloquium Analytische Atomspektroskopie, Proc. Conf. 1993*, eds. K. Dittrich, B. Welz, University of Leipzig, 451–456, 1993.
442. Deutsches Institut für Normung eV, 'German Standard Methods for the Examination of Water, Wastewater and Sludge – Cations (Group E); Determination of Co by AAS (E24), DIN 38406-E24, 1993.
443. R. Ajlec, M. Cop, J. Stupar, 'Interference in the Determination of Cr in Plant Materials and Soil Samples by FAAS', *Analyst*, **113**(4), 585–590 (1988).
444. D. Nägele, W. Brunn, 'The Improvement of Chromium Determination in Soil Extracts by Means of AAS (Flame Technique)', *Z. Pflanzenernähr. Bodenkd.*, **151**(1), 55–56 (1988).
445. A. Sahuquillo, R. Rubio, G. Rauret, 'Detection of Sources of Error in the Determination of Cr in Environmental Matrices by FAAS and ETAAS', in *Quality Assurance for Environmental Analysis*, eds. Ph. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 39–62, 1995.
446. J.-h. Liu, R.E. Sturgeon, V.J. Boyko, S.N. Willie, 'Determination of Total Cr in Marine Sediment Reference Material BCSS-1', *Fresenius' J. Anal. Chem.*, **356**, 416–419 (1996).
447. Deutsches Institut für Normung eV, 'German Standard Methods for the Examination of Water, Wastewater and sludge – Cations (Group E): Determination of Cu by AAS', DIN 38406/7, 1991.
448. U. Hornig, C. Krause, G. Sbieschni, 'Determination of Heavy Metal Contaminants in Municipal Sewage Systems by Analysis of Sewage Sludge', *Acta Hydrochim. Hydrobiol.*, **19**(3), 285–293 (1991).
449. V. Carbonell, A. Morales-Rubio, A. Salvador, M. de la Guardia, J.L. Burguera, 'AAS Analysis of Solids with On-line Microwave-assisted Digestion', *J. Anal. At. Spectrom.*, **7**(7), 1085–1089 (1992).
450. R. Martinez-Avila, V. Carbonell, M. de la Guardia, A. Salvador, 'Slurries Introduction in FI/AAS. Analysis of Sewage Sludges', *J. Assoc. Off. Anal. Chem.*, **73**(3), 389–393 (1990).
451. Deutsches Institut für Normung eV, 'German Standard Methods for the Examination of Water, Wastewater and Sludge – Cations (Group E); Determination of K by

AAS Using an Air–Acetylene Flame', DIN 38406/13, 1990.
452. P. Chattopadhyay, B.N. Sahoo, 'Direct FAAS Determination of Mn in Rocks, Nodules, Soils and Related Materials Using Tartaric Acid–KCl–$SrCl_2$ Mixture', *Talanta*, **40**(5), 701–706 (1993).
453. A.M.R. Ferreira, A.O.S.S. Rangel, J.L.F.C. Lima, 'FI Systems with a Stream Splitting and a Dialysis Unit for the Soil Analysis of Na and K by Flame Emission Spectrometry, and Ca and Mg by AAS', *Commun. Soil. Sci. Plant Anal.*, **26**(1–2), 183–195 (1995).
454. Deutsches Institut für Normung eV, 'German Standard Methods for the Examination of Water, Wastewater and Sludge – Cations (Group E): Determination of Ni by AAS', DIN 38406/11, 1991.
455. H. Hein, G. Schwedt, *Richtlinien und Grenzwerte, Luft, Wasser, Boden, Abfall*, 4th edition, Vogel Verlag, Würzburg, 1998.
456. N.S. Bloom, 'Determination of Ag in Marine Sediments by Zeeman-corrected GFAAS', *At. Spectrosc.*, **4**(6), 204–207 (1983).
457. A. Brzezinska-Paudyn, J.C. Van Loon, R. Haneock, 'Comparison of the Determination of As in Environmental Samples by Different Analytical Techniques', *At. Spectrosc.*, **7**(3), 72–75 (1986).
458. J.A.H. Desaulniers, R.E. Sturgeon, S.S. Berman, 'Atomic-absorption Determination of Trace Metals in Marine Sediments and Biological Tissues Using a STPF', *At. Spectrosc.*, **6**(5), 125–127 (1985).
459. J.P. Dubois, 'Measurement of Cd Speciation in Soil by ETAAS with ZBC', *J. Trace Microprobe Tech.*, **9**, 149–163 (1991).
460. E. Lundberg, B. Bergmark, 'Determination of Total Sn in Geological Materials by ETAAS', *Anal. Chim. Acta*, **188**, 111–118 (1986).
461. E. Pruszkowska, P. Barrett, 'Determination of As, Se, Cr, Co and Ni in Geochemical Samples Using the STPF and ZBC', *Spectrochim. Acta, Part B*, **39**(2–3), 485–491 (1984).
462. D.C. Barron, B.W. Haynes, 'Determination of Ga in Phosphorus Flue Dust and Other Materials by GFAAS', *Analyst*, **111**(1), 19–21 (1986).
463. M. Bettinelli, N. Pastorelli, U. Baroni, 'STPF Determination of Trace Metals in Fly Ash Samples', *At. Spectrosc.*, **7**(1), 45–48 (1986).
464. M. Bettinelli, U. Baroni, N. Pastorelli, 'Determination of Trace Elements in Environmental Matrices by Zeeman GFAAS and Matrix Modification', in *Applications of Zeeman Graphite Furnace Atomic Absorption Spectrometry in the Chemical Laboratory and in Toxicology*, eds. C. Minoia, S. Caroli, Pergamon Press, Oxford, 47–77, 1992.
465. B. Welz, G. Schlemmer, J.R. Mudakavi, 'Pd Nitrate–Mg Nitrate Modifier for GFAAS. Part 1. Determination of As, Sb, Se and Tl in Airborne Particulate Matter', *J. Anal. At. Spectrom.*, **3**(1), 93–97 (1988).
466. L. Zhang, G. Carnrick, W. Slavin, 'Rapid Analysis with Transversely Heated GFAAS', *Spectrochim. Acta, Part B*, **48**(11), 1435–1443 (1993).
467. M. Hoenig, A. Cilissen, 'ETAAS: Fast or Conventional Programs?', *Spectrochim. Acta, Part B*, **48**(8), 1003–1012 (1993).
468. 'Standard Reference Material 2709 (San Joaquin Soil), Certificate of Analysis, National Institute of Standards and Technology, Gaithersburg, MD, August, 1993.
469. 'Standard Reference Material 2711, Montana Soil Moderately Elevated Trace Element Concentrations', Certificate of Analysis, National Institute of Standards and Technology, Gaithersburg, MD, August, 1993.
470. P. Quevauviller, H. Muntau, U. Fortunati, K. Vercoutere, 'The Certification of the Total Contents (Mass Fractions) of Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn and the Aqua Regia Soluble Contents (Mass Fractions) of Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn in a Calcareous Loam Soil, CRM 141R', EUR Report 16890 EN, CEC, Brussels, 1996.
471. E.A. Maier, B. Griepink, H. Muntau, K. Vercoutere, 'Certification of the Total Contents (Mass Fractions) of Cd, Co, Cu, Pb, Mn, Hg, Ni and Zn and the Aqua Regia Soluble Contents (Mass Fractions) of Cr, Cu, Ni and Zn in a Light Sandy Soil (CRM 142R)', EUR Report 15283 EN, CEC, Brussels, 1994.
472. E.A. Maier, B. Griepink, H. Muntau, K. Vercoutere, 'Certification of the Total Contents (Mass Fractions) of Cd, Co, Cu, Pb, Mn, Hg, Ni and Zn and the Aqua Regia Soluble Contents (Mass Fractions) of Cr, Cu, Ni and Zn in a Sewage Sludge Amended Soil (CRM 143R)', EUR Report 15284 EN, CEC, Brussels, 1994.
473. P. Quevauviller, H. Muntau, U. Fortunati, K. Vercoutere, 'The Certification of the Total Contents (Mass Fractions) of Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn and the Aqua Regia Soluble Contents (Mass Fractions) of Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn in a Sewage Sludge from Domestic Origin, CRM 144 R', EUR Report 16891 EN, CEC Brussels, 1996.
474. E.A. Maier, B. Griepink, H. Muntau, K. Vercoutere, 'Certification of the Total Contents (Mass Fractions) of Cd, Co, Cu, Pb, Mn, Hg, Ni and Zn and the Aqua Regia Soluble Contents (Mass Fractions) of Cr, Cu, Ni and Zn in a Sewage Sludge, CRM 145R', Eur Report 15282 EN, CEC, Brussels, 1993.
475. P. Quevauviller, H. Muntau, U. Fortunati, K. Vercoutere, 'The Certification of the Total Contents (Mass Fractions) of Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn and the Aqua Regia Soluble Contents (Mass Fractions) of Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn in a Sewage Sludge from Industrial Origin, CRM 146 R', EUR Report 16892 EN, CEC Brussels, 1996.
476. P. Tittarelli, C. Biffi, 'Direct Determination of Trace Elements in Solid Samples by GFAAS with Zeeman Correction', in *Applications of Zeeman Graphite Furnace Atomic Absorption Spectrometry in the Chemical*

Laboratory and in Toxicology, eds. C. Minoia, S. Caroli, Pergamon Press, Oxford, 79–94, 1992.
477. U. Völlkopf, Z. Grobenski, R. Tamm, B. Welz, 'Solid Sampling in GFAAS Using the Cup-in-tube Technique', *Analyst*, **110**(6), 573–577 (1985).
478. U. Kurfürst, *Solid Sample Analysis – Direct and Slurry Sampling Using ETAAS and ETV/ICP-OES/MS*, Springer, Heidelberg, 1998.
479. N.J. Miller-Ihli, 'Slurry Sampling GFAAS: A Preliminary Examination of Results from an International Collaborative Study', *Spectrochim. Acta, Part B*, **50**(4–7), 477–488 (1995).
480. R. Dürnberger, P. Esser, A. Janssen, 'Determination of Se with Solid Sampling and Direct Zeeman-AAS. Part 1', *Fresenius' Z. Anal. Chem.*, **327**(3+4), 343–346 (1987).
481. P.S. Low, G.J. Hsu, 'Direct Determination of Atmospheric Pb by Zeeman Solid Sampling GFAAS', *Fresenius' J. Anal. Chem.*, **337**(3), 299–305 (1990).
482. G. Schlemmer, B. Welz, 'Determination of Heavy Metals in Environmental Reference Materials Using Solid Sampling GFAAS', *Fresenius' Z. Anal. Chem.*, **328**(4+5), 405–409 (1987).
483. M. José, M.P. Moura, M. Teresa, S.D. Vasconcelos, A.A. S.C. Machado, 'Determination of Pb in Atmospheric Aerosols by ETAAS with Direct Introduction of Filters into the Graphite Furnace', *J. Anal. At. Spectrom.*, **2**, 451–454 (1987).
484. R.C. Schothorst, H.M.A. Géron, D. Spitsbergen, R.F.M. Herber, 'Determination of Heavy Metals on Filter Material by Solid Sampling Direct Zeeman-AAS', *Fresenius' Z. Anal. Chem.*, **328**(4–5), 393–395 (1987).
485. F. Dolinsek, J. Stupar, V. Vrscaj, 'Direct Determination of Cadmium and Lead in Geological and Plant Materials by ETAAS', *J. Anal. At. Spectrom.*, **6**(8), 653–660 (1991).
486. M. Hoenig, P. van Hoeyweghen, 'Alternative to Solid Sampling for Trace Metal Determination by Platform ETAAS: Direct Dispensing of Powdered Samples Suspended in Liquid Medium', *Anal. Chem.*, **58**(13), 2614–2617 (1986).
487. G.R. Carnrick, G. Daley, A. Fotinopoulos, 'Design and Use of a New Automated Ultrasonic Slurry Sampler for GFAAS', *At. Spectrosc.*, **10**(6), 170–174 (1989).
488. N.J. Miller-Ihli, 'A Systematic Approach to Ultrasonic Slurry GFAAS', *At. Spectrosc.*, **13**(1), 1–6 (1992).
489. C. Bendicho, A. Sancho, 'Analysis of Solid Samples by AAS with Electrothermal Atomization by Automatic Introduction of Suspensions', *Quim. Ind. (Madrid)*, **38**, 526–529 (1992).
490. D.K. Bradshaw, W. Slavin, 'Rapid Slurry Analysis of Solid Coal and Fly Ash Samples', *Spectrochim. Acta, Part B*, **44**(12), 1245–1256 (1989).
491. M.C. Carneiro, R.C. Campos, A.J. Curtius, 'Determination of Sb, Ni and V in Slurry from Airborne Particulate Material Collected on Filter by GFAAS', *Talanta*, **40**(12), 1815–1822 (1993).
492. L. Ebdon, A.S. Fisher, H.G.M. Parry, A.A. Brown, 'Direct Atomic Spectrometric Analysis by Slurry Atomisation. Part 10. Use of An Air-ashing Stage in ETAAS', *J. Anal. At. Spectrom.*, **5**(4), 321–324 (1990).
493. L. Ebdon, H.G.M. Parry, 'Direct Atomic Spectrometric Analysis by Slurry Atomisation. Part 4. Determination of Se in Coal by ETAAS', *J. Anal. At. Spectrom.*, **3**(1), 131–134 (1988).
494. L. Ebdon, A.S. Fisher, S.J. Hill, 'Use of Hydrogen in ETAAS to Decrease the Background Signal Arising from Environmental Slurry Samples', *Anal. Chim. Acta*, **282**, 433–436 (1993).
495. A. Fernández, C.R. Fernández, N. Carrión, D. Loreto, Z. Benzo, R. Fraile, 'Metals Determination in Atmospheric Particulates by AAS with Slurry Sample Introduction', *At. Spectrosc.*, **12**(4), 111–118 (1991).
496. I. López-García, J.A. Cortéz, M.H. Córdoba, 'FI FAAS for Slurry Atomization. Determination of Fe, Ca and Mg in Samples with High Silica', *Talanta*, **40**(11), 1677–1685 (1993).
497. G. Rauret, 'Extraction Procedures for the Determination of Heavy Metals in Contaminated Soil and Sediments', *Talanta*, **46**(3), 449–455 (1998).
498. J.B. Jones, Jr, 'Soil Test Methods: Past, Present and Future Use of Soil Extractants', *Commun. Soil. Sci. Plant Anal.*, **29**(11–14), 1543–1552 (1998).
499. A.M. Ure, C.M. Davidson, R.P. Thomas, 'Single and Sequential Extraction Schemes for Trace Metal Speciation in Soil and Sediment', in *Quality Assurance for Environmental Analysis*, eds. P. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 505–523, 1995.
500. H.A. van der Sloot, 'Quick Techniques for Evaluating the Leaching Properties of Waste Materials: Their Relation to Decisions on Utilization and Disposal', *Trends Anal. Chem. (Pers. Ed.)*, **17**(5), 298–310 (1998).
501. A. Tessier, P.G.C. Campbell, M. Bisson, 'Sequential Extraction Procedure for the Speciation of Particulate Trace Metals', *Anal. Chem.*, **51**, 844–851 (1979).
502. C. Kheboian, C.F. Bauer, 'Accuracy of Selective Extraction Procedures for Metal Speciation in Model Aquatic Sediments', *Anal. Chem.*, **59**(10), 1417–1423 (1987).
503. A. Tessier, P.G.C. Campbell, 'Comments on the Testing of the Accuracy of An Extraction Procedure for Determining the Partitioning of Trace Metals in Sediments', *Anal. Chem.*, **60**(14), 1475–1476 (1988).
504. K. Wallmann, M. Kersten, J. Gruber, U. Förstner, 'Artifacts in the Determination of Trace-metal Binding Forms in Anoxic Sediments by Sequential Extraction', *Int. J. Environ. Anal. Chem.*, **51**(1–4), 187–200 (1993).
505. M. Raksasataya, A.G. Langdon, N.D. Kim, 'Assessment of the Extent of Pb Redistribution During Sequential Extraction by Two Different Methods', *Anal. Chim. Acta*, **332**(1), 1–14 (1996).
506. S.O. Ajayi, G.W. Vanloon, 'Studies on Redistribution During the Analytical Fractionation of Metals in Sediments', *Sci. Total Environ.*, **87–88**, 171–187 (1989).

507. F.M.G. Tack, M.G. Verloo, 'Leaching Behaviour and Solid-phase Fractionation of Selected Metals as Affected by Thermal Treatment of a Polluted Sediment', *Int. J. Environ. Anal. Chem.*, **51**(1–4), 167–175 (1993).
508. C.F. Bauer, C. Kheboian, 'Comments on the Testing of the Accuracy of An Extraction Procedure for Determining the Partitioning of Trace Metals in Sediments – Response to Comments', *Anal. Chem.*, **60**(14), 1477 (1988).
509. G. Rauret, J.F. Lopez-Sanchez, A. Sahuquillo, R. Rubio, C. Davidson, A. Ure, P. Quevauviller, 'Improvement of the BCR Three Step Sequential Extraction Procedure Prior to the Certification of New Sediment and Soil Reference Materials', *J. Environ. Monit.*, **1**(1), 57–61 (1999).
510. A.M. Ure, P. Quevauviller, H. Muntau, B. Griepink, 'Speciation of Heavy Metals in Soils and Sediments. An Account of the Improvement and Harmonization of Extraction Techniques Undertaken Under the Auspices of the BCR of the CEC', *Int. J. Environ. Anal. Chem.*, **51**(1–4), 135–151 (1993).
511. H.A. van der Sloot, L. Heasman, Ph. Quevauviller (eds.), *Harmonization of Leaching/Extraction Test*, Elsevier, Amsterdam, 1997.
512. E. Campos, E. Barahona, M. Lachica, M.D. Mingorance, 'A Study of the Analytical Parameters Important for the Sequential Extraction Procedure Using Microwave Heating for Pb, Zn and Cu in Calcareous Soils', *Anal. Chim. Acta*, **369**(3), 235–243 (1998).
513. P. Quevauviller, 'Operationally Defined Extraction Procedures for Soil and Sediment Analysis. II. Certified Reference Materials', *Trends Anal. Chem. (Pers. Ed.)*, **17**(10), 632–642 (1998).
514. S.J. Nagourney, N.J. Tummillo, J. Birri, K. Peist, J.S. Kane, 'Development of Reference Materials for the Analysis of Acid-extractable Metals in Sludge', *Talanta*, **44**(2), 189–196 (1997).
515. P. Quevauviller, G. Rauret, J.F. López-Sánchez, R. Rubio, A. Ure, H. Muntau, 'The Certification of the Extractable Contents (Mass Fractions) of Cd, Cr, Cu, Ni and Zn in Sediment Following a Three-step Sequential Extraction Procedure, CRM 601', EUR Report 17554 EN, CEC, Brussels, 1997.
516. P. Quevauviller, M. Lachica, E. Barahoba, G. Rauret, A. Gomez, H. Muntau, 'The Certification of the EDTA and DTPA Extractable Contents (Mass Fractions) of Cd, Cr, Cu, Ni, Pb and Zn in Calcareous Soil, CRM 600', EUR Report 17555 EN, CEC, Brussels, 1997.
517. C.P. Bosnak, Z.A. Grosser, 'The Determination of Soil Cations by FAAS', *At. Spectrosc.*, **17**(6), 211–214 (1996).
518. B. Pérez-Cid, I. Lavilla, C. Bendicho, 'Analytical Assessment of Two Sequential Extraction Schemes for Metal Partitioning in Sewage Sludges', *Analyst*, **121**(10), 1479–1484 (1996).
519. M.L. Räisänen, L. Hämäläinen, L.M. Westerberg, 'Selective Extraction and Determination of Metals in Organic Stream Sediments', *Analyst*, **117**(3), 623–627 (1992).
520. E. Ruiz, F. Romero, G. Besga, 'Selective Solubilization of Heavy Metals in Torrential River Sediments', *Toxicol. Environ. Chem.*, **33**, 1 (1991).
521. R.P. Thomas, A.M. Ure, C.M. Davidson, D. Littlejohn, G. Rauret, R. Rubio, J.F. López-Sánchez, 'Three-stage Sequential Extraction Procedure for the Determination of Metals in River Sediments', *Anal. Chim. Acta*, **286**(3), 423–429 (1994).
522. L. Campanella, D. D'Orazio, B.M. Petronio, E. Pietrantonio, 'Proposal for a Metal Speciation Study in Sediments', *Anal. Chim. Acta*, **309**(1–3), 387–393 (1995).
523. J. Hlavay, K. Polyák, I. Bódog, A. Molnár, E. Mészáros, 'Distribution of Trace Elements in Filter-collected Aerosol Samples', *Fresenius' J. Anal. Chem.*, **354**(2), 227–232 (1996).
524. A. Profumo, G. Spini, L. Cucca, E. Zecca, 'Sequential Extraction Procedures for Speciation of Inorganic Cadmium in Emissions and Working Areas', *Talanta*, **47**(3), 605–612 (1998).
525. J. Hlavay, K. Polyak, A. Molnar, E. Meszaros, 'Determination of the Distribution of Elements as a Function of Particle Size in Aerosol Samples by Sequential Leaching', *Analyst*, **123**(5), 859–863 (1998).
526. E.M. Donaldson, 'Determination of As in Ores, Concentrates and Related Materials by GFAAS After Separation by Xanthate Extraction', *Talanta*, **35**(1), 47–53 (1988).
527. E.M. Donaldson, M.E. Leaver, 'Determination of Te in Ores, Concentrates and Related Materials by GFAAS After Separations by Iron Collection and Xanthate Extraction', *Talanta*, **37**(2), 173–183 (1990).
528. M. Legret, 'Speciation of Heavy Metals in Sewage Sludge and Sludge-amended Soil', *Int. J. Environ. Anal. Chem.*, **51**, 161–165 (1993).
529. L. Orsini, A. Bermond, 'Application of a Sequential Extraction Procedure to Calcareous Soil Samples: Preliminary Studies', *Int. J. Environ. Anal. Chem.*, **51**(1–4), 97–108 (1993).
530. S. Pai, F. Lin, C. Tseng, D. Sheu, 'Optimization of Heating Programs of GFAAS for the Determination of Cd, Cu, Ni and Pb in Sediments Using Sequential Extraction Technique', *Int. J. Environ. Anal. Chem.*, **50**(3), 193–205 (1993).
531. S. Poole, 'Determination of Extractable Plant-available Se from Soils by ETAAS', *Commun. Soil. Sci. Plant Anal.*, **19**(15), 1681–1691 (1988).
532. J. Mocak, M. Vanickova, J. Labuda, 'Determination of Extractable Cr(VI) in the Presence of Large Excess of Cr(III) in Solid Materials', *Mikrochim. Acta*, **2**(2), 231–246 (1985).
533. J. Posta, H. Berndt, S.-K. Luo, G. Schaldach, 'High-performance Flow FAAS for Automated On-line Separation and Determination of Cr(III)/Cr(VI) and Preconcentration of Cr(VI)', *Anal. Chem.*, **65**(19), 2590–2595 (1993).

534. P. Rychlovsky, M. Krenzelok, R. Volhejnova, 'Online Simultaneous Sorption Preconcentration and Determination of Cr(III) and Cr(VI) with AAS Detection', *Collect. Czech. Chem. Commun.*, **63**(12), 2015–2026 (1998).
535. J. Stupar, W. Frech, 'Ultrasonic Nebulizer Interface System for Coupling LC and ETAAS', *J. Chromatogr.*, **541**, 243–255 (1991).
536. W.D. Marshall, J.S. Blais, F.C. Adams, 'HPLC–AAS Interface for the Determination of Ionic Alkyllead, Alkyltin, Arsonium and Selenonium Compounds', in *Metal Speciation in the Environment*, eds. J.A.C. Broekaert, S. Güçer, F. Adams, Springer, Berlin, 253–273, 1990.
537. F. Laborda, M.V. Vicente, J.M. Mir, J.R. Castillo, 'Evaluation of On-line Coupling Size Exclusion Chromatography ETAAS for Se Speciation', *Fresenius' J. Anal. Chem.*, **357**(7), 837–843 (1997).
538. G. Kölbl, K. Kalcher, K.J. Irgolic, 'Computerized Data Treatment for an HPLC–GFAAS System for the Identification and Quantification of Trace Element Compounds', *J. Autom. Chem.*, **15**(2), 37–45 (1993).
539. A. Astruc, R. Pinel, M. Astruc, 'On-line GFAAS for Liquid Chromatographic Detection. Theoretical and Experimental Study', *Anal. Chim. Acta*, **228**(1), 129–137 (1990).
540. K. Furtmann, D. Seifert, 'On the Routine Determination of Cr(VI) in Soils', *Fresenius' J. Anal. Chem.*, **338**(1), 73–74 (1990).
541. R. Milacic, J. Stupar, 'Fractionation and Oxidation of Cr in Tannery Waste- and Sewage Sludge-amended Soils', *Environ. Sci. Technol.*, **29**, 506–514 (1995).
542. N. Kozuh, J. Stupar, R. Milacic, B. Gorenc, 'Optimization of Extraction Procedure for Determination of Total Water-soluble Cr and Cr(VI) in Various Soils', *Int. J. Environ. Anal. Chem.*, **56**, 207–217 (1994).
543. N. Cardellicchio, S. Geraci, C. Marra, P. Paterno, 'Determination of Tributyltin Oxide in Coastal Marine Sediments and Mussels by ETAAS', *Appl. Organomet. Chem.*, **6**, 241–246 (1992).
544. M.D. Stephenson, D.R. Smith, 'Determination of Tributyltin in Tissues and Sediments by GFAAS', *Anal. Chem.*, **60**(7), 696–698 (1988).
545. G. Mortensen, B. Pedersen, G. Pritzl, 'Determination of Organotin Compounds in Marine Sediments Using GFAAS', *Appl. Organomet. Chem.*, **9**, 65–73 (1995).
546. P. Bermejo-Barrera, M.C. Barciela-Alonso, M. Ferrón-Novais, A. Bermejo-Barrera, 'Speciation of As by the Determination of Total As and As(III) in Marine Sediment Samples by ETAAS', *J. Anal. At. Spectrom.*, **10**(3), 247–252 (1995).
547. M.B. de la Calle-Guntiñas, Y. Madrid, C. Cámara, 'Speciation of Sb by AAS. Applicability to Selective Determination of Sb(III) and Sb(V) in Liquid Samples and of Bioavailable Sb in Sediments and Soil Samples', *Mikrochim. Acta*, **109**(1–4), 149–155 (1992).
548. S. Herwald, U. Backes, B. Straka-Emden, B. Neidhart, 'Kopplung von Extraktion und ETAAS zur Chromspeciesanalyse', in *5. Colloquium Atomspektrometrische Spurenanalytik*, ed. B. Welz, Perkin-Elmer, Überlingen, 419–428, 1989.
549. H.-D. Bauer, D. Dahmann, H.-H. Fricke, S. Herwald, B. Neidhart, 'Quality Control of Measuring Procedures by Interlaboratory Tests. Cr(VI)-species in Welding Fumes', *Staub Reinhalt Luft*, **51**, 125–128 (1991).
550. B. Pedersen, E. Thomsen, R.M. Stern, 'Some Problems in Sampling, Analysis and Evaluation of Welding Fumes Containing Cr(VI)', *Ann. Occup. Hyg.*, **31**, 325–338 (1987).
551. O. Rohling, B. Neidhart, 'Application of the Denuder Technique in Filter Sampling of Airborne Chromates at Ground Level Concentrations', *Fresenius' J. Anal. Chem.*, **351**(1), 33–40 (1995).
552. L. Girard, J. Hubert, 'Speciation of Cr(VI) and Total Cr Determination on Welding Dust Samples by FI Analysis Coupled to AAS', *Talanta*, **43**(11), 1965–1974 (1996).
553. C. Brescianini, A. Mazzucotelli, F. Valerio, R. Frache, G. Scarponi, 'Determination of Hexavalent Cr in Welding Fumes by GFAAS After Liquid Anion-exchange Separation. Investigation of Interferences', *Fresenius' Z. Anal. Chem.*, **332**(1), 34–36 (1988).
554. W.H. Ficklin, 'Extraction and Speciation of As in Lacustrine Sediments', *Talanta*, **37**(8), 831–834 (1990).
555. M. Kersten, B. Schulz-Dobrick, T. Lichtensteiger, C.A. Johnson, 'Speciation of Cr in Leachates of a MSWI Bottom Ash Landfill', *Environ. Sci. Technol.*, **32**, 1398–1403 (1998).
556. P. Bermejo-Barrera, G. González-Campos, M. Ferrón-Novais, A. Bermejo-Barrera, 'Column Preconcentration of Organotin Compounds with Tropolone-immobilized and Their Determination by ETAAS', *Talanta*, **46**(6), 1479–1484 (1998).
557. W. Slavin, 'AAS. Flame AAS', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 68, 1994.
558. W. Slavin, 'GFAAS', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 51–90, 1994.
559. E.A. Hanlon, 'Elemental Determination by AAS', in *Handbook of Reference Methods for Plant Analysis*, ed. Y.P. Kalra, CRC Press, Boca Raton, FL, 157–164, 1998.
560. T.C. Rains, 'Application of AAS to the Analysis of Foods', in *AAS–Theory, Design and Application*, ed. S.J. Haswell, Elsevier, Amsterdam, 191–226, 1991.
561. M.W. Pritchard, J. Lee, 'Agriculture Samples', in *Quantitative Trace Analysis of Biological Materials*, eds. H.A. McKenzie, L.E. Smythe, Elsevier, Amsterdam, 367–387, 1988.
562. H.A. McKenzie, L.E. Smythe (eds.), *Quantitative Trace Analysis of Biological Materials – Principles and Methods for Determination of Trace Elements and Trace*

Amounts of Some Macro Elements, Elsevier, Amsterdam, 1988.
563. K.T. Smith, *Trace Minerals in Foods*, Marcel Dekker, New York, 1988.
564. R.F.M. Herber (ed.), *Trace Element Analysis in Biological Specimens*, Elsevier, Amsterdam, 1994.
565. D.L. Tsalev, *AAS in Occupational and Environmental Health Practice. Vol. III: Progress in Analytical Methodology*, CRC Press, Boca Raton, FL, 1995.
566. B. Markert, *Instrumental Element and Multi-element Analysis of Plant Samples*, John Wiley & Sons, Chichester, 1996.
567. B. Markert (ed.), *Plants as Biomonitors*, VCH, Weinheim, 677, 1993.
568. J.W. Kiceniuk, S. Ray, *Analysis of Contaminants in Edible Aquatic Resources*, VCH, New York, 1994.
569. A. Taylor, S. Branch, D.J. Halls, L.M.W. Owen, M. White, 'Atomic Spectroscopy Update – Clinical and Biological Materials, Foods and Beverages', *J. Anal. At. Spectrom.*, **14**(4), 717–781 (1999).
570. G. Wagner, 'Variability of Element Concentrations in Tree Leaves on Sampling Parameters', in Element Concentration Cadasters in Ecosystems (ECCE), eds. H. Lieth, B. Markert, Verlag Chemie, Weinheim, 41–54, 1990.
571. W. Maurer, G. Schaldach, W. Wagener, J. Peters, 'Determination of Essential and Toxic Cations in Needles of Conifers', *Fresenius' Z. Anal. Chem.*, **322**(3), 359–364 (1985).
572. B. Markert, 'Multi-element Analysis in Ecosystems: Basic Conditions for Representative Sampling of Plant Materials', *Fresenius' Z. Anal. Chem.*, **335**(6), 562–565 (1989).
573. B. Markert, 'Aspects of Cleaning Environmental Materials for Multi-element Analysis, e.g. Plant Samples', *Fresenius' J. Anal. Chem.*, **342**(4–5), 409–412 (1992).
574. V. Krivan, G. Schaldach, 'Sampling and Sample Pretreatment of Tree Needles for Element Analysis', *Fresenius' Z. Anal. Chem.*, **324**, 158–167 (1986).
575. G.V. Iyengar, 'Presampling Factors in the Elemental Composition of Biological Systems', *Anal. Chem.*, **54**, 554A–558A (1982).
576. J.B. Jones, V.W. Case, 'Sampling, Handling, and Analyzing Plant Tissue Samples', in *Soil Testing and Plant Analysis*, 3rd edition, ed. R.L. Westerman, Soil Science Society of America, Madison, WI, 389–427, 1990.
577. L.H. Keith, *Principles of Environmental Sampling*, 2nd edition, American Chemical Society, Washington, DC, 1996.
578. B. Markert (ed.), *Environmental Sampling for Trace Analysis*, VCH, Weinheim, 1994.
579. E.I. Hamilton, 'State of the Art of Trace Element Determination in Plant Matrices: Determination of the Chemical Elements in Plant Matrices, an Overview', *Sci. Total Environ.*, **176**(1–3), 3 (1995).
580. M. Sager, J. Mittendorfer, 'Influence of Milling or Cutting Procedures on Trace Element Contents of Plant Samples', *Int. J. Environ. Anal. Chem.*, **67**(1–4), 59–71 (1997).
581. P. Mader, J. Száková, E. Curdová, 'Combination of Classical Dry Ashing with Stripping Voltammetry in Trace Element Analysis of Biological Materials: Review of the Literature Published After 1978', *Talanta*, **43**(4), 521–534 (1996).
582. G.H. Llimous, P. Fallavier, 'Comparison Between Two Dry-mineralization Methods (With and Without Hydrofluoric Acid) of Palm Leaves for Determination of K – Influence of Silica', *Commun. Soil. Sci. Plant Anal.*, **27**(5–8), 1623 (1996).
583. H.A. Hamid, M.I. Al Jaboury, A.K. Mohammed, 'Determination of Tl by Furnace AAS', *Anal. Chim. Acta*, **243**(2), 239–245 (1991).
584. X.-Q. Shan, Z.-N. Yuan, Z.-M. Ni, 'Determination of Ga in Sediment, Coal, Coal Fly Ash and Botanical Samples by GFAAS Using Nickel Matrix Modification', *Anal. Chem.*, **57**(4), 857–861 (1985).
585. W. Slavin, G.R. Carnrick, D.C. Manning, 'Graphite-tube Effects on Perchloric Acid Interferences on Al and Tl in the STPF', *Anal. Chim. Acta*, **138**, 103–110 (1982).
586. M.M. Emara, M.M. Ali, A. El-Aziz Gharib, 'Effect of Acids on the Determination of Ni by AAS', *Anal. Chim. Acta*, **102**, 181–184 (1978).
587. E. Wieteska, A. Ziótek, A. Drzewinska, 'Extraction as a Method for Preparation of Vegetable Samples for the Determination of Trace Metals by AAS', *Anal. Chim. Acta*, **330**(2–3), 251–257 (1996).
588. E. Jackwerth, M. Würfels, 'Pressure Digestion, Possibilities, Problems and Applications', in *Sampling and Sample Preparation*, ed. M. Stoeppler, Springer, New York, Chapter 11, 1997.
589. M. Würfels, E. Jackwerth, 'Investigations on the Carbon Balance in Decomposition of Biological Materials with Nitric Acid', *Fresenius' Z. Anal. Chem.*, **322**(3), 354–358 (1985).
590. E. Jackwerth, S. Gomiscek, 'General Aspects of Trace Analytical Methods – VI. Acid Pressure Decomposition in Trace Element Analysis', *Pure Appl. Chem.*, **56**, 479–489 (1984).
591. L. Dunemann, 'Microwave Digestion', in *Sampling and Sample Preparation*, ed. M. Stoeppler, Springer, New York, Chapter 12, 1997.
592. L. Dunemann, M. Meinerling, 'Comparison of Different Microwave-based Digestion Techniques in View of Their Application to Fat-rich Foods', *Fresenius' J. Anal. Chem.*, **342**(9), 714–718 (1992).
593. M.D. Mingorance, M.L. Pérez-Vazquez, M. Lachica, 'Microwave Digestion Methods for the Atomic Spectrometric Determination of Some Elements in Biological Samples', *J. Anal. At. Spectrom.*, **8**(6), 853–858 (1993).

594. L.H.J. Lajunen, J. Piispanen, E. Saari, 'Microwave Dissolution of Plant Samples for AAS Analysis', *At. Spectrosc.*, **13**(4), 127–131 (1992).
595. A.A. Mohd, J.R. Dean, W.R. Tomlinson, 'Factorial Design Approach to Microwave Dissolution', *Analyst*, **117**(11), 1743–1748 (1992).
596. M. Burguera, J.L. Burguera, 'Microwave-assisted Sample Decomposition in Flow Analysis', *Anal. Chim. Acta*, **366**(1–3), 63–80 (1998).
597. C.-y. Zhou, M.-k. Wong, L.-l. Koh, Y.-c. Wee, 'Microwave Digestion of Biological Samples with Tetraethylammonium Hydroxide and EDTA for Elemental Determination', *Talanta*, **43**(7), 1061–1068 (1996).
598. R. Bader, A.M. Reichlmayr-Lais, M. Kirchgessner, 'AAS Determination of Pt in Animal Tissues', *Agribiol. Res.*, **44**, 212–218 (1991).
599. M. Stoeppler (ed.), *Sampling and Sample Preparation*, Springer, New York, 1997.
600. B. Griepink, G. Tölg, 'Sample Digestion for the Determination of Elemental Traces in Matrices of Environmental Concern', *Pure Appl. Chem.*, **61**(6), 1139–1146 (1989).
601. N.J. Miller-Ihli, 'Trace Element Determinations in Foods and Biological Samples Using ICP-AES and FAAS', *J. Agric. Food Chem.*, **44**, 2675–2679 (1996).
602. V.A. Granadillo, H.S. Cubillán, J.M. Sánchez, J.E. Tahán, E.S. Márquez, R.A. Romero, 'Three Pressurized Mineralization Procedures that Permit Subsequent FAAS Determination of Ca, Fe, K, Mg and Zn in Bovine Blood Plasma-containing Cookies and in Standard Reference Materials', *Anal. Chim. Acta*, **306**(1), 139–147 (1995).
603. H. Lippo, A. Särkelä, 'Microwave Dissolution Method for the Determination of Heavy Metals in Biomonitors Using GFAAS and FAAS', *At. Spectrosc.*, **16**(4), 154–157 (1995).
604. H.A.M.G. Vaessen, C.G. van de Kamp, 'Reference-material-based Collaborative Test of FAAS Determination of Ca and Mg in Foods and Biological Materials', *Z. Lebensm.-Unters. Forsch.*, **190**, 199–204 (1990).
605. M.R. Shepard, C.E. Lee, R.S. Woosley, D.J. Butcher, 'Determination of Ca and Mg by FAAS in Fraser Fir (*Abies fraseri*) and Red Spruce (*Picea rubens*) Foliage from Richland Balsam Mountain, North Carolina', *Microchem. J.*, **52**(1), 118–126 (1995).
606. T. Staszewski, A. Uzieblo, J. Szdzuj, 'Characteristics of Pine Needles from Trees of Different Age Growing in the Protective Zone of ''Konin'' Aluminum Smelter', *Chemosphere*, **36**(4–5), 1013–1018 (1998).
607. M. Ferretti, R. Udisti, E. Barbolani, 'Mineral Nutrients and Trace Metals in Tree Rings', *Fresenius' J. Anal. Chem.*, **347**(10), 467–470 (1993).
608. M.C. Yebra-Biurrun, A. Bermejo-Barrera, M.P. Bermejo-Barrera, 'Determination of Na and Ca in Fruits and Vegetables by AAS', *Alimentaria (Madrid)*, **28**(222), 53–55 (1991).
609. O.D. Diaz, M.C. Portuondo, M.G. Valdes, R.R. Chang, C.E. Handal, 'Comparison of Various Digestion Methods for the Determination of Different Metals in *Petiveria alliacea* Lynn (Anamu) by FAAS', *Quim. Anal. (Barcelona)*, **15**(2), 123 (1996).
610. I. Türkan, E. Henden, Ü. Celik, S. Kivilcim, 'Comparison of Moss and Bark Samples as Biomonitors of Heavy Metals in a Highly Industrialized Area in Izmir, Turkey', *Sci. Total Environ.*, **166**, 61–67 (1995).
611. J. Fleurence, C. Le Coeur, 'Influence of Digestion Procedures on the Determination of Pb and Cd Levels in the Laminariale Undaria Pinnatifida (Wakame) by FAAS', *Bot. Mar.*, **37**, 555 (1994).
612. A.B. Tawali, G. Schwedt, 'Iron Species and Enzymatic Browning in Potatoes (*Solanum tuberosum*)', *Dtsch. Lebensm. Rundsch.*, **93**(12), 387–390 (1997).
613. M. Anke, L. Angelow, 'Rb in the Food Chain', *Fresenius' J. Anal. Chem.*, **352**(1–2), 236–239 (1995).
614. S.R. Aston, B. Oregioni, A. Veglia, 'Homogeneity Testing of a Marine Biological Intercomparison Material by AAS', *At. Spectrosc.*, **8**(5), 155–158 (1987).
615. S.-k. Xu, M. Sperling, B. Welz, 'FAAS Determination of Cd and Cu in Biological Reference Materials Using On-line Sorbent Extraction Preconcentration', *Fresenius' J. Anal. Chem.*, **344**(12), 535–540 (1992).
616. M.F. Enriquez-Dominguez, M.C. Yebra-Biurrun, M.P. Bermejo-Barrera, 'Determination of Cd in Mussels by FAAS with Preconcentration on a Chelating Resin in a FI System', *Analyst*, **123**(1), 105–108 (1998).
617. X.-p. Yan, M. Sperling, B. Welz, 'Application of a Macrocycle Immobilized Silica Gel Sorbent to FI On-line Microcolumn Preconcentration and Separation Coupled with FAAS for Interference-free Determination of Trace Pb in Biological and Environmental Samples', *Anal. Chem.*, **71**(19), 4216–4222 (1999).
618. G.-h. Tao, Z.-l. Fang, 'FAAS Determination of Pb in Biological Materials Using a FI On-line Separation and Preconcentration System Based on Ion-pair Sorbent Extraction', *At. Spectrosc.*, **17**(1), 22–26 (1996).
619. C. Dittfurth, E. Ballesteros, M. Gallego, M. Valcárcel, 'On-line Precipitation/Dissolution System for the Preconcentration and Determination of Mn Traces by AAS', *Spectrochim. Acta, Part B*, **51**(14), 1935–1941 (1996).
620. B. Welz, S.-k. Xu, M. Sperling, 'FAAS Determination of Cd, Co, and Ni in Biological Samples Using a FI System with On-line Preconcentration by Co-precipitation without Filtration', *Appl. Spectrosc.*, **45**(9), 1433–1443 (1991).
621. Z.-l. Fang, S.-k. Xu, L.-p. Dong, W.-q. Li, 'Determination of Cd in Biological Materials by FAAS with FI On-line Sorption Preconcentration', *Talanta*, **41**, 2165–2172 (1994).
622. H.-w. Chen, S.-k. Xu, Z.-l. Fang, 'Determination of Cu in Water and Rice Samples by FAAS with FI On-line Adsorption Preconcentration Using a Knotted Reactor', *Anal. Chim. Acta*, **298**(2), 167–173 (1994).

623. M. Anwar Memon, X.-r. Wang, B.-l. Huang, 'FI Extraction with Mixed Solvents: Its Application to Cu Determination with FAAS', *At. Spectrosc.*, **14**(4), 99–102 (1993).
624. S. Hirata, K. Honda, T. Kumamaru, 'Trace Metal Enrichment by Automated On-line Column Preconcentration for FI–AAS', *Anal. Chim. Acta*, **221**(1), 65–76 (1989).
625. Z.-h. Hou, 'Influence of the Siliceous Component in Determination of Gold in Plant Materials', *J. Geochem. Explor.*, **27**, 323–328 (1987).
626. S. Itoh, J. Ogawa, S. Kaneco, K. Ohta, T. Mizuno, 'Magnesium–Tungsten Cell–ETAAS – Determination of Au in Environmental Samples', *Int. J. Environ. Anal. Chem.*, **69**(2), 181–189 (1998).
627. F. Alt, U. Jerono, J. Messerschmidt, G. Tölg, 'Determination of Pt in Biotic and Environmental Materials. I. µg/kg to g/kg Range', *Mikrochim. Acta*, **3**, 299–304 (1988).
628. E.L. Kothny, 'Note on the Interpretation of Pt and Pd Values in Vegetation Near the Limit of Determination with a Graphite Furnace Technique', *Appl. Spectrosc.*, **41**(4), 700–702 (1987).
629. G.E.M. Hall, J.-C. Pelchat, C.E. Dunn, 'The Determination of Au, Pd, and Pt in Ashed Vegetation by ICP-MS and GFAAS', *J. Geochem. Explor.*, **37**(1), 1–23 (1990).
630. E. Beinrohr, M.-L. Lee, P. Tschöpel, G. Tölg, 'Determination of Pt in Biotic and Environmental Samples by GFAAS After its Electrodeposition into a Graphite Tube Packed with Reticulated Vitreous Carbon', *Fresenius' J. Anal. Chem.*, **346**(6–9), 689–692 (1993).
631. Y. Hayashibe, F. Watai, Y. Enoki, K. Oguma, R. Kur, 'Determination of Trace Be in Biological Materials by a Combined Ion-exchange GFAAS', *Anal. Sci.*, **5**(5), 531–533 (1989).
632. R. Kuroda, K. Oguma, Y. Mori, H. Okabe, T. Takizawa, 'Anion-exchange Pre-concentration and Determination of Bi in Environmental Materials', *Chromatographia*, **32**(11), 583–585 (1991).
633. G. Collado-Gomez, C. Bosch Ojeda, A. Garcia de Torres, J.M. Cano Pavon, 'Determination of Co in Biological Samples by ETAAS After Extraction with 1,5-Bis(di-2-pyridylmethylene) Thiocarbohydrazide', *Analusis*, **23**(5), 224–227 (1995).
634. W.J. Blanchflower, A. Cannavan, D.G. Kennedy, 'Determination of Trace Amounts of Co in Feed Grains and Forages by Solvent Extraction and GFAAS', *Analyst*, **115**(10), 1323–1325 (1990).
635. J.M. Cano Pavón, E. Vereda Alonso, C. Bosch Ojeda, A. García de Torres, 'Determination of Ni in Biological Samples by ETAAS Involving a Prior Extraction with 1,5-Bis-(di-2-pyridylmethylene) Thiocarbonohydrazide', *Anal. Lett.*, **24**(1), 153–166 (1991).
636. R.F.M. Herber, 'Use of Solid Sampling Analysis for the Determination of Trace Elements in Tissues', *Microchem. J.*, **51**(1–2), 46–52 (1995).
637. D.E. Wells, 'Extraction and Pre-concentration of Organometallic Species from Environmental Samples', *Mikrochim. Acta*, **109**(1–4), 13–21 (1992).
638. M. Schintu, F. Jean-Caurant, J.-C. Amiard, 'Organomercury Determination in Biological Reference Materials: Application to a Study on Hg Speciation in Marine Mammals off the Faröe Islands', *Ecotoxicol. Environ. Safety*, **24**, 95–101 (1992).
639. L. Benramdane, M. Accominotti, J.J. Vallon, 'Validated Determination of Total As Species of Toxicological Interest (As(III), As(V) and Their Metabolites) by AAS After Separation from Dietary As by Liquid Extraction: Toxicological Applications', *Analyst*, **123**(8), 1711–1715 (1998).
640. P. Bavazzano, A. Perico, K. Rosendahl, P. Apostoli, 'Determination of Urinary As by Solvent Extraction and ETAAS. A Comparison with Directly Coupled HPLC–ICPMS', *J. Anal. At. Spectrom.*, **11**(7), 521–524 (1996).
641. J.L. Kacprzak, 'Determination of Tributyltin in Oysters by ETAAS', *Int. J. Environ. Anal. Chem.*, **38**(4), 561–564 (1990).
642. F.Y. Pang, Y.L. Ng, S.M. Phang, S.L. Tong, 'Determination of Total Sn and Tributyltin in Biological Tissues', *Int. J. Environ. Anal. Chem.*, **53**, 53–61 (1993).
643. R. Giordano, L. Ciarelli, G. Gattorta, M. Ciprotti, S. Costantini, 'Analysis of Tricyclohexyltin Hydroxide [Cyhexatin] Using Zeeman GFAAS', *Microchem. J.*, **49**(1), 69–77 (1994).
644. M. Plessi, A. Monzani, 'Survey of Total and Bioavailable Cr in Grain and Cereal Products by AAS', *J. Assoc. Off. Anal. Chem.*, **73**(5), 798–800 (1990).
645. G. Kölbl, K. Kalcher, K.J. Irgolic, 'Ion Chromatographic Determination of Selenite and Selenate with Se-specific Detection by Flame and GFAAS', *Anal. Chim. Acta*, **284**(2), 301–310 (1993).
646. H. Emteborg, G. Bordin, A.R. Rodriguez, 'Speciation of Se in Aqueous and Biological Matrices by Microbore Ion Chromatography Coupled with ETAAS via Ultra Low Volume Fraction Collection', *Analyst*, **123**(5), 893–898 (1998).
647. H. Emteborg, G. Bordin, A.R. Rodriguez, 'Speciation of Organic and Inorganic Se in a Biological CRM Based on Microbore Ion-exchange Chromatography Coupled to ICPAES via a Direct Injection Nebulizer or Coupled to ETAAS', *Analyst*, **123**, 245–253 (1998).
648. M. Potin-Gautier, N. Gilon, M. Astruc, I. De Gregori, H. Pinochet, 'Comparison of Se Extraction Procedures for Its Speciation in Biological Materials', *Int. J. Environ. Anal. Chem.*, **67**(1–4), 15–25 (1997).
649. R. Milacic, J. Stupar, 'Simultaneous Determination of Cr(III) Complexes and Cr(VI) by Fast Protein Anion-exchange LC–AAS', *Analyst*, **119**(4), 627–632 (1994).
650. R.B. Cruz, C. Lorouso, S. George, Y. Thomassen, J.D. Kinrade, L.R.P. Butler, 'Determination of Total, Organic Solvent Extractable, Volatile and Tetraalkyllead

in Fish, Vegetation, Sediment and Water Samples', *Spectrochim. Acta, Part B*, **35**(11–12), 775–783 (1980).
651. G. Schwedt, G. Jahns, 'Analysenschema Zur Untersuchung der Bleiverteilung in und auf Kontaminierten Pflanzen', *Fresenius' Z. Anal. Chem.*, **328**(1+2), 85–88 (1987).
652. R. Lobinski, H. Chassaigne, J. Szpunar, 'Analysis for Metallothioneins Using Coupled Techniques', *Talanta*, **46**(2), 271–289 (1998).
653. K.A. High, R. Azani, A.F. Fazekas, Z.A. Chee, J.S. Blais, 'Thermospray–Microatomizer Interface for the Determination of Trace Cd and Cd-metallothioneins in Biological Samples with FI and HPLC–AAS', *Anal. Chem.*, **64**(24), 3197–3201 (1992).
654. C.D. Klaassen, L.D. Lehman-McKeeman, 'Separation and Quantification of Isometallothioneins by HPLC/AAS', *Methods Enzymol.*, **205**, 190–198 (1991).
655. E.H. Larsen, J.-S. Blais, 'Performance of a Modular Thermospray Interface for Signal Enhancement in FAAS Coupled On-line to FI or LC', *J. Anal. At. Spectrom.*, **8**(4), 659–664 (1993).
656. P. Sun, X. Shan, Y. Zheng, L. Jin, W. Xu, 'Determination of Dietary Cd-induced Metallothioneins in Rabbit Kidneys and Cd in Metallothioneins by Anion-exchange HPLC Coupled with GFAAS', *J. Chromatogr. B. Biomed. Appl.*, **110**, 73–84 (1991).
657. Y. Zheng, X.-q. Shan, P. Sun, L.-Z. Jin, 'Metallothionein Separation and Analysis by Reversed-phase HPLC Coupled with GFAAS', *Chem. Speciation Bioavailability*, **3**(1), 30–36 (1991).
658. H. Van Beek, A.J. Baars, 'Isolation and Quantitation of Cd-, Zn- and Cu-metallothioneins by HPLC–AAS', *J. Chromatogr.*, **442**, 345–352 (1988).
659. G. Ji, K. Günther, 'Beiträge zur Charakterisierung von Cd-spezies in Kontaminierten Pflanzlichen Lebensmitteln', *Lebensmittelchemie*, **49**(3), 62–63 (1995).
660. K. Günther, F. Umland, 'Speziesanalytik von Cd und Tl in Nativen Rapspflanzen (*Brassica napus*)', *Fresenius' Z. Anal. Chem.*, **331**(3+4), 302–309 (1988).
661. C.L. Chou, J.F. Uthe, R.D. Guy, 'Determination of Free and Bound Cd, Zn, Cu, and Ag Ions in Lobster (*Homarus americanus*) Digestive Gland Extracts by Gel Chromatography Followed by AAS and Polarography', *J. AOAC Int.*, **76**(4), 794–798 (1993).
662. K. Wedepohl, G. Schwedt, 'Analysis of Copper/Flavonoid Complexes as Element Species in Vegetables', *Anal. Chim. Acta*, **203**, 23–34 (1987).
663. S.C. Mo, D.S. Choi, J.W. Robinson, 'Speciation Studies of Cu and Mg in Duckweed Extract Using Interfaced HPLC and Thermospray Nebulizer FAAS', *J. Environ. Sci. Health, Part A*, **23**(5), 441–451 (1988).
664. K.T. Suzuki, H. Sunaga, T. Yajima, 'Separation of Metallothionein into Isoforms by Column Switching on Gel Permeation and Ion Exchange Columns with HPLC–AAS', *J. Chromatogr.*, **303**, 131–136 (1984).
665. K. Lange-Hesse, L. Dunemann, G. Schwedt, 'Properties and Binding Forms of Cd and Ni in Protein Extracts from Bean Seeds (*Phaseolus vulgaris* L.)', *Fresenius' J. Anal. Chem.*, **349**(6), 460–464 (1994).
666. K. Lange-Hesse, 'Investigation of Properties and Binding Forms of Cd and Ni in Protein Extracts by Ultra/Diafiltration, Photometry and AAS', *Fresenius' J. Anal. Chem.*, **350**(1–2), 68–73 (1994).
667. W. Henze, F. Umland, 'Isolation and Quantification of Cd and Cu-species in Lettuce Plants', *Anal. Sci.*, **3**, 225–227 (1987).
668. K. Günther, H. Waldner, 'Speciation of Zn and Cd in Ordinary Vegetable Foodstuffs', *Anal. Chim. Acta*, **259**(1), 165–173 (1992).

Flow-injection Techniques in Environmental Analysis

J.L. Burguera and M. Burguera
University of Los Andes, Mérida, Venezuela

1 Introduction 1
2 History 1
3 Developing Methods for Sample Preparation 2
4 Manifold Considerations 2
5 Manifold Designs with Spectrometric Detectors 2
5.1 Spectrophotometry 2
5.2 Turbidimetry 5
5.3 Atomic Spectrometry 5
5.4 Luminescence 8
6 Manifold Designs with Electrochemical Detectors 8
6.1 Amperometry 8
6.2 Ion-selective Electrodes 9
6.3 Voltammetry 9
7 Determination of Pollutants in Environmental Samples 9
8 The Present and the Future 9
Abbreviations and Acronyms 10
Related Articles 10
References 10

An overview is given on the most important applications of flow-injection (FI) systems developed for environmental materials, highlighting their main basic components and modes. Also, the article systematically describes the different FI manifolds coupled to spectrometric or electrochemical (EQ) detectors used for the determination of analyte species to offer diverse solutions in the analysis of various samples.

1 INTRODUCTION

The rapid growth in agriculture and industrial activity has resulted in chemical pollution. Industrial waste water, toxic metals, toxic organic substances, sewage sludge and soil waste materials have been discharged into the environment. The complexity of the real samples and the diversity of analytes typically encountered in environmental studies hampers the direct analysis of key components by use of a specific detection technique. To supply required data rapidly, efficiently, with the demanded accuracy and precision and with the least human attention, it would be an advantage to count on reliable and easily automatable systems.

Undoubtedly, any FI system can be considered to have a certain degree of automation mainly of the preliminary operations of the analytical process. On the other hand, it can be readily adapted to virtually every type of analytical problem, it allows direct control by the researcher and can be interfaced with a variety of possible configurations with different optical or EQ instruments typically used in routine laboratories.[1] In this respect, FI systems have proven useful for handling contaminated complex samples and for direct pretreatment of solids and liquids with harsh chemical matrices. For these types of analytes/samples the primary advantage of FI is economic. A secondary advantage is the ease of development or refinement of method.

2 HISTORY

From the birth of FI important applications have been developed for agricultural, metallurgical, food, biological and environmental materials. Ruzicka et al. initiated the FI-spectrophotometric and potentiometric approach for the determination of phosphorus[2] and total nitrogen content in plant materials.[3] The same research team demonstrated for the first time the feasibility of using ion-selective electrodes for the determination of potassium, sodium and nitrate in soil extracts.[4] The Brazilian group of Piracicaba initiated the FI-turbidimetric[5] method (for the determination of sulfate in natural waters and plant digests) and atomic absorption spectrometry (AAS)[6] (for the determination of calcium, magnesium and potassium in plant materials) approaches. The first two papers dealing with the determination of air pollutants (chlorine, bromine and SO_2) were published by Mottola et al.[7,8] based on photometric detection. Since these pioneering works were published, many books[1,9–13] and reviews[14–18] have provided an overview of the developments of FI interfaced with different kinds of detectors, highlighting the different applications in the various fields, including environmental analysis. Undoubtedly, the largest effort for the development of FI techniques for environmental analysis has been focused in the use of spectrophotometric, atomic spectrometry (AS) and some EQ detectors. The control of pollutants in waters (cationic, anionic and organic pollutants), sludges and sediments (metallic species) has received most of the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

attention, while the determination of air pollutants has been more restricted.

3 DEVELOPING METHODS FOR SAMPLE PREPARATION

The organic matter often present in environmental samples, in most cases interferes with the subsequent instrumental measurement. This serious problem can be eliminated through processes capable of converting the organic constituents into inorganic, which are free of this type of interference effect or by removing the analyte from the interfering matrix through any conventional separation technique. Transformation of the analyte into a form that can be analyzed by a specific measurement technique often requires unpleasant, tedious and time-consuming procedures such as dilution, separation, dissolution or digestion. Depending on the type of matrix, the analyte and its expected concentration in the sample and the technique to be used for measurement, one pretreatment procedure or various are to be followed. These appear to be the reason for the rather late development of FI digestion procedures followed by on-line detection of the analyte. Ultraviolet (UV), ultrasonic and microwave (MW) radiation emerged in the analytical field to dramatically improve the digestion process. Additionally, the development of on-line decomposition systems proved to be simple; they are relatively safe to use, provide a decrease in the blank values, reduce the contamination risk, are applicable to samples of different nature and are completely fit for automation. The automated procedures enable on-line coupling of the pretreatment procedures to powerful determination methods such as spectrophotometry, turbidimetry, flame atomic absorption spectrometry (FAAS), hydride generation (HG) and electrothermal atomic absorption spectrometry (ETAAS), inductively coupled plasma/atomic emission spectrometry (ICP/AES), inductively coupled plasma/mass spectrometry (ICP/MS), fluorimetry and EQ methods or to hybrid techniques such as high-pressure liquid chromatography/atomic absorption spectrometry (HPLC/AAS) or HPLC/ICP/MS, which widened the application fields to speciation studies.[19]

4 MANIFOLD CONSIDERATIONS

The complete FI system would contain a sampler, the FI unit, a detector and a recording device that is capable of sample identification and evaluation of the detector signal in terms of analyte concentration. In a typical FI procedure, the entire dynamic process may take place in open tubular reactors and, therefore, a certain degree of mixing between the injectate and reagent streams is necessary before the reaction zone reaches the detector. Thus, it is necessary to emphasize the applicability of different injection modes, pumping and manifold design, which are important considerations in the field of environmental analysis. Despite the intrinsic characteristics of a given detection technique, it does not always meet the demand for specificity, sensitivity and/or selectivity for environmental analysis. The feasibility of employing in-line sample preparation further enhances the utility of FI techniques for the analysis of many samples of environmental origin, and improves the performance of the detector. In this way, (a) the reagent/analyte chemistry can be carried out under the required conditions, (b) interferents can be minimized or eliminated resulting in selectivity enhancements, (c) preconcentration of the analyte allows for lower detection limits (DLs), (d) the analyte can be placed into some detectable form and (e) speciation studies can be carried on.

The final design of a FI system is dependent on the particular requirements of the user and on the conditions defined by the method. Since sample dispersion is being used to affect specific functions, including sample dilution, mixing, reaction, preconcentration steps, etc., it should be controlled as meticulously as possible. In seeking to exploit the chemistry fully, it is desirable that manifold modifications can be made rapidly and easily to save time. In particular, the interface of the FI manifold with the detector unit must be described in detail to allow its use for routine purposes by other researchers.

5 MANIFOLD DESIGNS WITH SPECTROMETRIC DETECTORS

5.1 Spectrophotometry

The simplicity of the reactions and the ease of adaptation to continuously flowing systems enhanced the popularity of FI spectrophotometric methods for the determination of many environmentally important species, especially in water samples. The first applications were carried out with the aim of demonstrating the principles of FI techniques. Thus a number of FI manifolds have been adapted to well-known spectrophotometric methods to determine a number of anionic[20–22] and cationic (calcium, iron, aluminum, lead, manganese, mercury, etc.)[22] as well as organic species.[19–23] With few exceptions, the manifolds are single line and permit the control of the pH and color formation, dilution of samples or preconcentration of the analyte with enhanced sensitivity and without loss of reproducibility.

Most of the spectrophotometric methods developed for the determination of sulfate are based on the removal of barium by sulfate, from its complexes with different color indicators followed by the measurement of the free ligand: methylthymol blue,[24] dimethylsulphonazo-III[25–27] or chloranilate.[28] Barium chromate and barium chloranilate[29] reaction columns were used as components of FI manifolds for the determination of microamounts of sulfate in sea and surface water and in rainwater samples, respectively. Other methods, indirectly determined sulfate by measuring the absorbance of thiocyanate[30] or of the $FeSO_4^+$ ion.[31] The catalytic effect of sulfate on the reaction between zirconyl ions and methylthymol blue was also used for the indirect determination of sulfate.[32] DLs in the range 0.1 to 0.6 mg L^{-1} with relative standard deviation (RSD) below 4% but low sample throughput (up to 30 samples per hour) and narrow working range (up to 14 mg L^{-1}) have been achieved, although in most of these cases, removal of interfering cations by ion-exchange was necessary.

The majority of the FI procedures developed for the determination of different forms of organic and dissolved phosphorus so far are based on the complexation of orthophosphate with acidic molybdate, although strong interferences from silicates have been reported. Most of the published methods require a preliminary digestion step to convert all the phosphorus compounds present in the sample to orthophosphate. The early procedures involve hydrolysis and thermal oxidation in long digestion coils (10 m) with consequent lengthy residence times at high temperatures and elevated pressure necessary to mineralize the organic and condensed phosphates. Less drastic conditions have been needed for a heated reactor containing platinum wire as catalyst in the presence of persulfate.[33] Speciation of dissolved phosphorus in environmental samples was possible by gel-filtration in a FI manifold.[34] UV photooxidation has been successfully used to perform in-line determination of organic phosphorus in natural waters with good sensitivity and precision.[35] Different materials and designs of photoreactors were tested in the development of the method. Teflon photoreactors have been preferred – some of their advantages being: robustness, low cost, suitability for field applications and ease of manufacture. Samples in the range 0.1 to 4 mg P L^{-1} can be analyzed at rates of 72 to 50 samples h^{-1}, respectively, with a DL of 0.01 mg L^{-1}.

An on-line MW digestion of wastewater was carried out by Haswell et al.[36] The MW pretreatment of samples with the inorganic enzyme pyrophosphatase ensured the determination of total phosphate, as straight-chain polyphosphates and ring metaphosphates, probably present in the samples, were all converted to orthophosphate. The system design allowed all the operations, including sample injection, digestion, de-bubbling, cooling, pressure control, mixing with reagents and signal measurement to be carried out on-line in about 2 min, compared to 25 min when using traditional digestion methods. A similar, low-cost and robust FI on-line MW-induced sample digestion for total phosphorus in natural waters and wastewater has also been described in the literature.[37] A 6 m poly(tetrafluoroethylene) (PTFE) flow-through reactor located in a domestic MW oven was used to produce the optimal oxidation and hydrolysis of the organic and condensed species suspected to be present in the real samples, with the aid of peroxodisulfate dissolved in perchloric acid. The orthophosphate formed was also detected as phosphomolybdenum blue after on-line filtration of the digest.

Free cyanide has been spectrophotometrically determined in FI systems by slowly (over 30 min) developing the color with a buffered chloramine-T solution used as carrier, or by injecting, in a reversed-mode system, pyridine/barbituric acid reagent into the buffered chloramine-T solution.[38] For total cyanide determination, which includes complexed cyanide, the procedure must be preceded by a tedious distillation which, if incorporated in the manifold might result in a rather complicated system with low sampling frequency. However, the combination of a fast reaction, based on the rapid formation of an unstable red intermediate product with isonicotinic acid/3-methyl-1-phenyl-2-pyrazolin-5-one, with the on-line generation and separation of hydrogen cyanide resulted in a successful system for preconcentration and sampling of total cyanide.[39] A sampling frequency of 40 h^{-1}, a DL of 6 μg L^{-1} and a preconcentration factor of 3.5 for a 2 mL sample have been achieved.

FI spectrophotometric methods developed so far for the determination of free chlorine in such samples are based on the formation of colored complexes with: *o*-tolidine (yellow),[40] *N*,*N*-diethyl-*p*-phenylene diamine in phosphate buffer (pink)[41] or 4,4′-tetramethyldiamino-thio-benzophenone (blue).[42] Only the latest method is sensitive enough (0.2–1.0 mg L^{-1}) to measure free chlorine in tap water.

Another anion very often determined in water samples is chloride. The most common procedure is based on the measurement of the red-colored complex between iron(III) and thiocyanate. The amount of chloride is proportional to the absorbance measured at 480 nm. Single-line manifolds permitted the application of a simple methodology to different types of samples (surface, ground, rain or seawater and the expansion of the working range by using the stream splitting principle[43] or performing on-line dilution.[44] However, the sampling frequency is rather slow (up to 120 h^{-1}) and the precision poor (1–8%) taking into account that the sample volume was large (up to 400 μL). Better sensitivity has been achieved (0.01–10 mg L^{-1}) for the determination

of chloride in air and water, based on the Fe(II)/Hg-tripyridyl-s-triazine flow system.[45]

Although nitrogen is well known as an essential element to life, some of its compounds (urea, ammonia, nitrite, nitrate, amines, etc.) could be hazardous to health if present in high concentrations in water supplies owing to their toxicity or carcinogenity. Simultaneous determination of nitrate and nitrite is possible by controlling two consecutive steps: reduction of nitrate to nitrite; the efficiency of this reaction depends on the metal used, the conditioning of the column and the pH and diazotation, which will depend on the reagents used. Zinc(II), copper(II) or copperized cadmium columns have been used in the reduction step.[46] The nitrite is diazotized with sulfanilamide and the product coupled to *N*-(1-naphthyl)ethylenediammonium dichloride to form an azo dye which is measured at 520 nm.[47] A more sensitive method used *p*-aminoacetophenone for diazotization, coupled the product with *m*-phenylenediamine at pH 1.5–3.0 and measured the absorbance at 460 nm.[48] Many modifications and adaptations of the original procedure have been published, all aimed to increase sensitivity, column lifetime, sampling rate and selectivity.

The commonly used methods for ammonia determination are based on the Nessler and Berthelot reaction in which ammonium ion is transformed into ammonia, separated through gas-permeable membranes and the gas absorbed on-line in a specific reagent and the color intensity measured at 640–800 nm.[49]

There are several spectrophotometric methods for the determination of boron in environmental samples that require a prior separation step owing to lack of sensitivity.[22] One of the most sensitive methods applied for the determination of boron in this type of sample is based on the formation of a complex with azomethine-H using ultrasonic leaching or a time-based injector[50] to avoid time-consuming sample treatment and reagent consumption, respectively.

Calcium in natural waters, soil extracts and plant digests has been spectrophotometrically determined in a FI system using the principle of merging zones.[51] Glyoxal bis (2-hydroxyanil), known as a sensitive and selective reagent for calcium, was used as the color-forming reagent. Interference effects from co-existing elements in the matrixes studied were minimized by the addition of triethanolamine and potassium cyanide to the carrier stream. A sampling rate of 180 determinations h^{-1} with a RSD of less than 1% have been accomplished. Chlorphosphonazo-III in hydrochloric acid (pH 2.2) was used as a chromogenic reagent for the determination of calcium in seawater in a FI system with a sampling throughput of 200 to 250 injections h^{-1} and insignificant effects from other alkaline-earth elements.[52]

An indirect FI spectrophotometric method based on the exchange reaction between calcium and the zinc complex of ethylene glycol-bis(2-aminoethylether)tetraacetic acid in the presence of 4-(2-pyridylazo)resorcinol has been described.[53] The sampling rate was 80 samples h^{-1} and the results for pond and tap water samples were in good agreement with those obtained by FAAS.

Dissolved chromium concentrations in natural waters is below 1 μg L^{-1} and its determination is dependent on the preconcentration method used and changes between species [Cr(VI)/Cr(III)] during sampling and/or storage. Ion-exchange coupled on-line with the standard method based on Cr(VI) color formation with 1,5-diphenylcarbazide is an example of a successful application to seawater samples.[21]

A simple FI set-up based on the absorbance of the orange-red iron(II)-1,10-phenanthroline complex at 512 nm allowed the determination of iron(II) and total iron in a large number of waters and plant digests. However, FI systems need serious optimization procedures to be followed in terms of sensitivity, linearity, sampling rate, etc. Natural computation techniques, using jointly the neural network and genetic algorithms have been applied in the modeling and optimization of a FI manifold for the colorimetric determination of iron(III) in water samples.[54] Thiocyanate was used as a color developing reagent. A comparison with AAS showed the high quality of the results. However, discrimination between Fe(II) and Fe(III) species has certain relevance in aquatic environments where they are active in many redox processes. The Fe(III)−SCN complex was selectively retained on a flow cell packed with an exchange resin while the oxidation of Fe(II) to Fe(III) took place in the loop of a previous valve.[55] The method presents a determination limit of 80 μg L^{-1}, a linear range of 80–500 μg L^{-1} for Fe(II) + Fe(III) and a precision of 5%.

Several automated FI systems have been developed for the photometric determination of lead after on-line column preconcentration, based on its property of forming a colored complex with 2-pyridylazo resorcinol,[22] and with 5, 10, 15, 20-tetra(4-*N*-sulfoethylpyridinium)porphyrin[56] or the extraction with dicyclohexyl-18-crown-6 into chloroform from acidic medium and addition of dithizone as chromogenic agent.[22]

Eriochrome Cyanine R in the presence of cationic surfactants proved to be a very sensitive photometric reagent for the determination of aluminum in natural waters.[57] A strict on-line control of the pH and interference effects from some concomitant species permitted the determination of aluminum in the range 0.8 to 4 mg L^{-1} with a precision RSD of 0.7% at 1 mg L^{-1} level and a DL of 1 μg L^{-1}.

Maximova et al. optimized an on-line preconcentration of some heavy metals (Pb, Co, Mn and Fe) on a capillary

column with a hydrophobic inner surface followed by MW sample processing for photometric detection in swamp water.[58] The procedure eliminates the noise caused by the organic matrix and allows the immobilization of the photometric reagents on the sorption column.

In order to avoid costly instrumentation such as ICP/AES or AAS, a fast, automated and inexpensive screening-test for recognizing heavy-metal contamination of drinking, surface and wastewater was implemented.[59] The test was based on the reaction of Cd, Co, Cu, Ni, Pb and Zn with dithizone at pH 6 followed by the on-line extraction of the complexes in carbon tetrachloride and photometric measurement at 500 nm.

In many environmental studies related to water quality monitoring the analysis of a large number of samples for chemical oxygen demand (COD) is often required. However, certain steps of the reactions involved prior to the spectrophotometric determination of such species are very slow. MW irradiation has been used to accelerate the oxidation of the organic matter with a potassium dichromate–sulfuric acid mixture for COD determination in natural waters.[60] This reaction takes about 2 h in a conventional system and only a few minutes in a flow-injection microwave (FI/MW) irradiated on-line arrangement. The determinations were performed up to 100 mg L^{-1} with a precision RSD of 2.12% at 40 mg L^{-1} and a DL of 1.5 mg L^{-1}.

On-line MW treatment of water samples proved successful for the reduction of urea to ammonium ions, which on reaction with NaOH release gaseous ammonia. The gas diffuses through a hydrophobic membrane and causes a pH change in a bromothymol blue indicator stream which is colorimetrically measured at 635 nm.[61] Calibration was linear up to 45 μmol L^{-1}, with a DL of 2.4 μmol L^{-1} (ten times higher than in the conventional method), a precision RSD (4%, $n = 4$) and a sampling rate, including sample preparation, of 14 min per sample. Although the sensitivity was worse, there are many advantages (better precision, better recoveries, rapid analysis and ease of automation) over the batch procedures.

A similar procedure, based on color change of bromothymol blue was used for the indirect determination of trimethylamine oxide in seawater at micromolar levels.[62] The analyte was converted to the volatile trimethylamine through a titanium-mediated deoxygenating reaction; the method features a wide linear operating range (0–200 μmol L^{-1}) and high sensitivity with cheaper instrumentation and lower maintenance.

5.2 Turbidimetry

The turbidimetric detection of barium sulfate is almost universally used for the determination of sulfate ions in waters. The sample containing sulfate is injected in a carrier stream of barium chloride and allowed to mix in the manifold in order to form a uniform suspension. The turbidity is measured at 420 nm, but the accuracy and precision strongly depend on the crystalline form and size distribution of the precipitate particles in the suspension. Therefore, strict control of the experimental conditions must be undertaken at any time; this perfectly suits the FI concept because every standard and sample can be treated under exactly the same conditions. There are some problems associated with this procedure, namely the stability of the suspension, accumulation of precipitate on the tubing lines as well as the presence of suspended organic matter or colored components in the sample. Polyvinyl alcohol[63] and gelatin[64] have been used to stabilize the barium sulfate to improve the precision. The unfavorable build-up of precipitates on the transfer lines as well as their tendency to settle in the flow cell led to blocking of the manifold. Addition of hydrochloric acid to the carrier stream prevents the formation of barium carbonate, chromate, phosphate or oxalate and the alternating injection of an alkaline buffer–ethylene diamine tetraacetic acid (EDTA) solution redissolves the accumulated barium sulfate precipitate and keeps the system clean.[65,66] The use of pH gradients and alkaline EDTA provided a stable suspension, which obviates the use of protective colloids. EDTA was also found useful for reducing interferences from metal ions.[67] In order to avoid interferences from turbid samples, Van Staden incorporated in the manifold a prevalve filtration accessory consisting of a packed activated carbon tube[68] or an active carbon filter paper.[69] The incorporation of an anion exchange resin in a FI manifold permitted the preconcentration of sulfate for its sensitive turbidimetric determination in water.[70]

A silver nitrate turbidimetric procedure has been applied to the determination of chloride in river water in a single-line flow system.[71] The turbidity of silver chloride suspensions formed in nitric acid is measured at 440 nm in the range 0 to 14 mg L^{-1} with a sampling rate of only 15 h^{-1}.

5.3 Atomic Spectrometry

Initial reports dealt with the use of the FI system as a mere alternative to manual introduction for FAAS. In such systems, the noise problems associated with the air bubbles in segmented sample insertion systems are eliminated, the analyst can use sample volumes in the microlitre range per measurement, and the sample flow rate is optimized to allow higher atomization with a net increase in sample analysis throughput. Olsen et al.[72] determined cadmium, copper, lead and zinc in polluted seawater at the part per million level at a

rate of 180–250 samples per hour; however, the sampling frequency depended on the preconcentration required. They developed an on-line FI preconcentration method using a microcolumn of Chelex-100 resin, allowing the determination of lead at concentrations as low as $10\,\mu g\,L^{-1}$ and $1\,\mu g\,L^{-1}$ for cadmium and zinc.

Martínez-Jimémez et al.[73] have developed a continuous precipitation and filtration flow system coupled on-line with the atomic absorption spectrometer. This technique allows the quantitative separation and preconcentration (by a factor of 700) of trace amounts of lead from tap water samples. Ammonia is used as a precipitating reagent and no collector was required. Several calibration graphs suited to the degree of enrichment were proposed for the determination of lead(II) in the range $1.2–1500\,\mu g\,L^{-1}$.

On-line coprecipitation–preconcentration systems can also greatly improve the DLs down to 0.23, 3.2 and $1.4\,\mu g\,L^{-1}$ for determination of metal species such as cadmium, lead and nickel, respectively, in various water samples and soils. The analytes are precipitated and collected on the inner wall of a knotted reactor with diethyldithiocarbamate-copper(II) and eluted with isobutyl methyl ketone.[74]

The advantages of FI solvent extraction coupled with AAS include high preconcentration factors, improved nebulization efficiency, reduced reagent consumption and the reduction of laboratory waste.[75] Gallego et al.[76] reported the indirect determination of anionic surfactants with an automatic continuous liquid–liquid extractor. The method involves the formation of the detergent-[1,10-phenanthroline-copper(II)] ion pair and extraction into methyl isobutyl ketone. The overall concentration can be determined indirectly by measurement of copper present in the organic layer by AAS. A phase separator fitted with a PTFE porous membrane was specially designed for the determination of detergents in the concentration range $0.1–5.0\,\mu g\,mL^{-1}$.

Ma et al.[77] described a FI on-line sorbent extraction system to determine cadmium, copper and lead in digest solutions of environmental samples using octadecyl functional groups (C_{18}) bonded silica gel as sorbent with diethylammonium-*N*,*N*-diethyldithiocarbamate or ammonium diethyldithiophosphate as complexing agent and methanol as eluent. DLs were 0.8, 1.4 and $10.0\,\mu g\,L^{-1}$ respectively for cadmium, copper and lead. These elements were determined in coal fly ash, calcareous loam soil and lake and estuarine sediments.

Interesting work presented by Bysouth et al.[78] used masking agents in the determination of lead in tap water by FAAS with FI preconcentration. The selectivity of immobilized 8-hydroxyquinoline for lead was improved by the use of masking agents during preconcentration, prior to its determination. Interference by iron, copper, aluminum and zinc is suppressed by including triethanolamine, thiourea, fluoride, acetylacetone or cyanide in the buffer as masking agents. On the other hand, Sperling et al.[79] determined chromium(III) and chromium(VI) in fresh water samples using on-line preconcentration with selective adsorption on activated alumina. Linear calibration for both species was established over the concentration range $10–200\,\mu g\,L^{-1}$ with DLs of 1.0 and $0.8\,\mu g\,L^{-1}$ for Cr(III) and Cr(VI) respectively.

The research team of Perkin-Elmer at Uberlingen has introduced some routine applications of on-line ion-exchange preconcentration systems for AS.[12] In particular, enrichment factors from 25 to 31 were obtained for copper, cadmium and lead from natural and sea water samples by FI coupled with AAS. Conical ion-exchange columns were found to be superior to uniform diameter columns in reducing dispersion during elution so that higher enrichment factors could be achieved using smaller sample volumes (1.6 mL per determination). The system can be easily also adaptable to an ICP spectrometric system.

FI sample introduction for ETAAS systems is less straightforward than FAAS or hydride generation atomic absorption spectrometry (HGAAS) owing to the discontinuity of ETAAS operation, this obviously being the cause of the slower development of the related techniques.[16] Historically, the main obstacles for efficiently and harmoniously combining FI on-line with ETAAS have been overcome by the research groups of the Department of Applied Research of Perkin-Elmer in Uberlingen and Mérida, Venezuela.[16,80] FI and ETAAS by on-line column preconcentration of metal species (such as: As, Pb, Cd, Co, Cu, Fe, Ni and Zn) from natural and seawater samples, has provided further impetus to the development in this field. Various sample introduction modes have been proposed by different workers. The conventional off-line mode is performed using pipettes for the deposition of the concentrate into the furnace after its pretreatment and collection in small vials. With on-line systems, the concentrates are dispensed automatically into the furnace by a pump via a sampling probe. FI also proved to be an efficient interface to ETAAS for: (1) sample deposition in the furnace by hydride sequestration, thermospray and high pressure nebulization; (2) selective sorbent extraction.[81]

Without any question, ICP at its current stage is the analytical technique that provides the most adequate simultaneous multielement analysis capacity. ICP allows the simultaneous assay of several elements at major, minor, trace and ultratrace levels without changing any experimental parameter in aqueous or organic solvents. However, the formation of polyatomic ions (particularly below $m/z = 80$) can cause serious interferences. Such

species can be introduced via precursor atoms in atmospheric gases, with the sample matrix or impurities in the argon support gas, and cause a systematic error. The first applications of FI/ICP can be traced in the review published by Christian and Ruzica in 1987.[14] Up to then, FI was used as an adjunct of conventional ICP instruments, which had bulky nebulizers. Therefore, FI systems are coupled to ICP or to ICP/MS for: (1) the on-line removal of matrix interferences;[82,83] (2) trace enrichment of analytes;[84] (3) on-line dilution[85] and (4) on-line sample digestion procedures.[86] In this context, trace metals in seawater,[82,87] sewage sludge,[86] waste water[88] and natural waters[88] have been determined with good precision and accuracy. The introduction of samples by capillary gas chromatography allowed the speciation and determination (down to $0.1\,ng\,L^{-1}$) of organotin compounds in river-water samples.[89]

Tsalev et al. have initiated the on-line MW sample pretreatment for the determination of Hg in water.[90] A reaction coil and a ballast-load coil are placed and oriented vertically within the cavity and chimney of a focused MW oven in a manner that provides a reliable long-term operation with increased reaction and irradiation times. Samples are mixed ''in batch'' with an appropriate oxidation reagent and loaded on an autosampler. Thereafter, all further operations are performed automatically within 90–402 s (depending on the sample volume and the instrument program used for a given determination). When the amalgamation accessory was used, a PTFE filter holder with a PTFE micropore membrane filter was placed between the gas–liquid separator and the quartz absorption cell to avoid aerosol droplets entering the amalgamation accessory. The limits of detection were 0.01 and $0.2\,\mu g\,L^{-1}$ for mercury, with and without amalgamation, respectively.

A technical note reports on the evaluation of a Perkin-Elmer flow-injection mercury system (FIMS) for the determination of Hg in sewage sludge, sediment and soil samples.[91] The DLs were as low as $5\,\mu g\,L^{-1}$ for sample volumes of 500 µL and RSD of 2% ($n = 3$) at $10\,\mu g\,L^{-1}$ Hg level. The same approach has been successfully tested for the determination of total mercury in environmental samples using MW digestion coupled to FIMS.[92]

A similar system was described by Lamble and Hill[93] for the determination of total mercury in environmental solid samples. The slurried samples were injected into a carrier of hydrochloric acid and brominated mixture before passing through the coil situated in the MW cavity. After mixing with hydroxylammonium chloride to remove excess bromine, mercury was determined in an on-line generation system. Each sample was processed in 50 s with a DL of $13\,ng\,L^{-1}$.

Morales-Rubio et al.[94] have described a two-line system for the determination of mercury in environmental materials (sewage sludge, polluted farmland soil and lake sediment) using on-line MW digestion and atomic fluorescence spectrometry (AFS) as detector. The acidified slurry samples (400 µL) were injected and carried through the PTFE digestion coil (4 m) located inside the oven (at 10% power level) for 50 s and through an ice/water bath. The reductant stream merged with the digest and mixed in a reduction coil before entry to the gas liquid separator. This system allowed accurate determination of mercury in certified environmental reference materials with a precision ranging between 0.5 and 1.5% RSD, a DL of $0.09\,ng\,g^{-1}$ and a throughput of 15 samples per hour.

Another interesting approach has been developed for the determination of total arsenic in geothermal water samples.[95] An appropriate volume of sample (0.1–2.0 mL) mixed with acids (H_2SO_4 and HNO_3) and an oxidizing agent ($K_2S_2O_8$) is subjected to MW heating. The cooled digest was mixed with a prereductant (KI and ascorbic acid) and sodium borohydride to form volatile arsine which was detected by FAAS in a quartz cell. The method allowed the determination of total arsenic with a DL of 0.2 ng of arsenic, a precision RSD of 2.5–4.5% and recovery values in the range 101–102%.

López-Gozálvez et al.[96] have described an on-line HPLC/MW oxidation-HGAAS four channel system for the selective determination of arsenite, arsenate, dimethylarsinate (DMA), monomethylarsonate (MMA), arsenobetaine (AsB) and arsenocholine (AsC) in environmental samples (mineral water, sewage water, harbor seawater, synthetic fish extract and sediment extract). Samples and multistandard solutions were injected into an anionic cartridge leading to a 100 µL sample loop, placed before an HPLC anionic column. Arsenite, arsenate, MMA and DMA were quantitatively retained in the anionic cartridge while AsB and AsC passed through it into the sample loop and separated in the chromatographic column. The column effluent mixed with the oxidizing solution flowed to the digestion coil in the oven for decomposition. The digest was cooled in an ice-bath and a T-junction was used to acidify the sample. In a second T-junction the digested sample was mixed with the reductant and the separated volatile arsine entered the quartz atomization cell. The anionic species can be separated and determined by removing the anionic cartridge from the system and introducing water instead of the oxidizing agent. After optimization of the chromatographic and MW-oxidation parameters, the RSD was 3–5% with DLs between 0.3 and 0.9 ng for all species (for 100 µL sample volume) and conversion efficiency close to 100% in all cases.

Le et al.[97] also speciated arsenic except that they used a dual system to avoid removal of the cartridge during the procedure. Comparable resolution was obtained by using HPLC-separation/MW-digestion with HGAAS or

ICP/MS detection. Complete separation of five arsenicals was achieved on a reversed-phase C_{18} column by using sodium heptanesulfonate as ion pair reagent and the DLs were for example $10\,\mu g\,L^{-1}$ for arsenite, DMA and AsB, $15\,\mu g\,L^{-1}$ for MMA and $20\,\mu g\,L^{-1}$ for arsenate.

An automatable and on-line cryogenic trapping system and MW heating was used for the determination and speciation of arsenic in aqueous samples by FI/HGAAS using sodium tetrahydroborate as reductor.[98] The separation of the species was based on a pH selective procedure: arsine from As(III) alone was generated in the presence of citric acid while nitric acid was used to generate the corresponding arsines from total inorganic arsenic, MMA and DMA. These were cryogenically trapped in a PTFE coil knotted and sealed inside another wider diameter tube located in the oven and through which liquid nitrogen was suctioned by negative pressure. Based on their different dielectric constants, the arsines were selectively liberated by using a heating cycle of MW radiation with volatilization times of 90, 180 and 235 s for dimethylarsine, monomethylarsine and arsine, respectively. The DLs were in the range 20 to 60 ng As for 10 mL sample volume, with the possibility of improving it by increasing sample size or running several consecutive reactions.

5.4 Luminescence

The first applications of FI coupled with chemiluminescence (CL) detection were reviewed by Burguera et al. in 1982.[99] Up to then, some metal species, tertiary amines, hydrogen peroxide and sulfite[100] were the species of environmental interest which were determined in natural waters. This initial trend has almost been followed up to date, but using different CL reactions, e.g. (1) the determination of tertiary amines in water and seawater over the range 1×10^{-5}–$1 \times 10^{-2}\,mol\,L^{-1}$ based on the oxidation of the analyte by sodium hypochlorite in presence of Rhodamine B as a sensitizer;[101] (2) the determination of aqueous ozone for potable water treatment;[102] (3) the determination of a nonionic surfactant in aqueous environmental samples;[103] (4) the determination of ultratrace concentrations of nitrite in water;[104] (5) the simultaneous determination of nitrate and nitrite in water by gas-phase CL.[105] One interesting application of FI is the in situ shipboard determination of species such as hydrogen peroxide,[106] copper and manganese using CL reactions.[107,108]

The analytical signals are dramatically affected by environmental conditions such as ionic strength, pH and interferents which can be stabilized in a FI system. The advantage of using FI for fluorescence-based measurements is that reproducibility can be significantly improved.[109] FI sample processing has been used for the determination of: 0–$10^{-5}\,mol\,L^{-1}$ total primary amines in seawater,[110] 0.2–300 $ng\,mL^{-1}$ copper(II) in natural waters;[111] 10^{-8}–$10^{-4}\,mol\,L^{-1}$ of residual peroxides in water;[112,113] and $\mu mol\,L^{-1}$ levels of sulfur species (sulfite, sulfide and methanethiol) in fogwaters[114] with good reproducibility, precision and rapidity using the *o*-phthalaldehyde/mercapto-ethanol, 2,2′-dipyridylketone hydrazone/hydrogen peroxide, *p*-hydroxyphenylacetate/peroxide/peroxidase and *N*-acridinylmaleimide/*N*,*N*-dimethylformamide reactions, respectively.

6 MANIFOLD DESIGNS WITH ELECTROCHEMICAL DETECTORS

The use of EQ methods in environmental analysis depends on the availability of instrumentation, trained personnel and practical procedures for complex matrixes. They are suitable for on-line field applications although EQ detection suffers from memory effects and electrostatic interferences. The design of flow-through EQ detectors for on-line couplings needs special consideration. Some novel sensors have been tried, including wall-coated (or open) tubular, packed bed (or porous) tubular, wire, cascade, wall-jet and thin layer designs. They monitor anionic species such as chloride, bromide, cyanide, nitrate and ammonium and to a lesser extent toxic metals, such as Pb, Hg, Cd, Cu, Zn and Ni, which substantially contribute to the pollution of land and water, and are readily determined by spectroscopic methods.

6.1 Amperometry

An amperometric flow-through detector was used for the determination of cyanide in different kinds of surface and groundwaters[115] by injecting the sample into an alkaline electrolytic carrier stream with good precision (2%) and high sampling frequency ($100\,h^{-1}$).

Arsenic was determined in contaminated soil samples by a novel technique based on gas diffusion through porous membranes.[116] Arsine generated in a donor stream of sodium borohydride passes quickly across a membrane into an acceptor stream of sulfuric acid, and the resulting solution is carried to a flow-through amperometric cell equipped with a dual gold electrode and Ag/AgCl as reference electrode. The detector provided enhanced selectivity, high sensitivity and low DL and affords the possibility of the development of field instrumentation for on-site determinations. A closed loop FI system was used for the determination of sulfur dioxide in air samples.[117] The manifold permitted the direct intercalation of air samples into a carrier solution containing the complex Fe(III)-1,10-phenanthroline as a regenerable chemical form. The extent of reduction of the complex to ferroin at the gas/liquid interface was

amperometrically monitored with an Ag/AgCl–carbon-paste electrode system. The composition and mode of preparation of the paste was found to considerably affect the performance of the system. The analyte was determined in the range 0.3–14.0 ppm (v/v) with RSD of about 3% at a rate of 25 samples per hour.

6.2 Ion-selective Electrodes

A home-made coated tubular solid-state chloride selective electrode was used by Van Staden to determine chloride up to 5000 mg L^{-1} with a precision of less than 1.7%.[118] The drawback of this procedure was the strong interference from mercury(II) ions. A similar approach was manufactured for the determination of inorganic bromide in soil in the range 1–5000 mg L^{-1}, at a rate of 80 samples per hour and with a standard deviation (SD) of 1.6%.[119]

Okumoto et al. interfaced a flow system with a gas-diffusion separator and a cyanide-selective electrode to screen wastes from metal plating processes.[120] They claimed a DL of 0.1 mg cyanide L^{-1} and a sampling frequency of 40 h^{-1}.

Improved DLs, compared to spectrophotometric methods, have been reported for the potentiometric determination of ammonium in seawater using a liquid-membrane ammonium-sensitive electrode[121] or a gas diffusion/dialyser unit.[35] On-line preconcentration on ion-exchange microcolumns and incorporation of gas-diffusion units proved successful but the final choice will depend on the particular analytical requirement.

6.3 Voltammetry

Voltammetric techniques, in particular absorptive cathodic stripping voltammetry are suitable for trace metal determination and speciation as they combine high sensitivity with speciation capabilities. Their combination with FI techniques facilitates in situ monitoring of various species in environmental samples.[122] For this purpose, a flow-through cell (4 μL) mounted in an automated voltammetric system monitored traces of cobalt in uncontaminated seawater.[123] A hanging mercury drop was used as electrode, while the chloride present in the seawater was the counter-ion for the Ag/AgCl reference electrode. The cell was combined with on-line UV digestion, on-line de-oxygenation and on-line standard addition for on-board determination of cobalt distribution in coastal waters.

Traces of metals can also be determined using chemically modified electrodes designed to enable both, preconcentration (binding) and surface interaction for subsequent quantification of the analyte accumulated on the surface. Carbon-paste electrodes for instance, are simple to prepare and have an easily adjustable fresh surface, which is generated in a reproducible manner. Such an electrode, modified with 1,2-bismethyl(2-aminocyclopentene-carbohydride)ethane was incorporated in a flow system for the preconcentration and determination of copper in water.[124] As little as 1.4×10^{-6} mol L^{-1} of Cu(II) can be determined in water with a precision of 2.9% for a preconcentration time of 4 min. Cu, Pb, Cd and Hg were determined in seawater, speciated as total particulate, particulate sorbed to inorganic matter, total dissolved and labile by differential pulse anodic stripping voltammetry.[125] The total-particulate-containing filters were MW digested and the organic matter in water samples was destroyed by UV photolysis. The elements in the filtered water were on-line pre-concentrated on a Chelamine microcolumn, eluted and determined by the EQ method and also by FAAS (Cu, Pb and Cd) or cold vapor atomic absorption spectrometry (CVAAS) (Hg) for quality control (QC).

7 DETERMINATION OF POLLUTANTS IN ENVIRONMENTAL SAMPLES

The main development in FI environmental analysis have been highlighted in the following reviews:

1982 Water pollutants.[99]
1985 Cations, anions, COD, halogens, sulfur dioxide, organic and other polluting species determination.[19,126]
1987 Anions determination.[11,14,20]
1988 Environmental samples.[1]
1989 Environmental samples.[9,10]
1990 Cations and anions and other polluting species determination.[15,127]
1991 Water quality monitoring.[21,128,132]
1994 Different pollutants in water.[23,129,130]
1995 Cations and anions determination in different kinds of samples.[12]
1996 Soil pollutants.[131]
1997 Different polluting species determination in different kinds of samples.[80]
1998 Different polluting species determination in different kinds of samples.[133]

8 THE PRESENT AND THE FUTURE

At present FI provides simple, handy, automable and versatile instrumentation which makes it suitable to meet the requirements of any laboratory wanting to make this type of analysis. The volume of papers involving the

analysis of water is substantial, and in a more restricted way the determination of different air and soil pollutants is well documented.

As realization spreads that the determination of the ''total'' concentration of an analyte is insufficient to yield an overall measure of toxicity, continued concern over environmental pollution will be reflected in the development and optimization of multielemental and speciation studies with imaginative FI setups. Analysis of biological materials as biomonitors of environmental pollution, airborne particulate, soils and waters will become of interest with on-line sample pretreatment, analyte separation from the matrix and preconcentration techniques to improve the sensitivity for the determination of individual species of the material under study.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AFS	Atomic Fluorescence Spectrometry
AS	Atomic Spectrometry
AsB	Arsenobetaine
AsC	Arsenocholine
CL	Chemiluminescence
COD	Chemical Oxygen Demand
CVAAS	Cold Vapor Atomic Absorption Spectrometry
DL	Detection Limit
DMA	Dimethylarsinate
EDTA	Ethylene Diamine Tetraacetic Acid
EQ	Electrochemical
ETAAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Flame Atomic Absorption Spectrometry
FI	Flow-injection
FIMS	Flow-injection Mercury System
FI/MW	Flow-injection Microwave
HG	Hydride Generation
HGAAS	Hydride Generation Atomic Absorption Spectrometry
HPLC/AAS	High-pressure Liquid Chromatography/Atomic Absorption Spectrometry
ICP/AES	Inductively Coupled Plasma/Atomic Emission Spectrometry
ICP/MS	Inductively Coupled Plasma/Mass Spectrometry
MMA	Monomethylarsonate
MW	Microwave
PTFE	Poly(tetrafluoroethylene)
QC	Quality Control
RSD	Relative Standard Deviation
SD	Standard Deviation
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Infrared Spectroscopy in Environmental Analysis • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis

REFERENCES

1. J. Ruzicka, E.H. Hansen, *Flow Injection Analysis*, 2nd edition, John Wiley & Sons, New York, 1988.
2. J. Ruzicka, J.W.B. Stewart, 'Flow Injection Analysis. Part II. Ultrafast Determination of Phosphorus in Plant Materials by Continuous Flow Spectrophotometry', *Anal. Chim. Acta*, **79**, 79–91 (1975).
3. J.W.B. Stewart, J. Ruzicka, H. Bergamin Filho, E.A.G. Zagatto, 'Flow Injection Analysis. Part III. Comparison of Continuous Flow Spectrophotometry and Potentiometry for Rapid Determination of the Total Nitrogen Content in Plant Digests', *Anal. Chim. Acta*, **81**, 371–386 (1976).
4. J. Ruzicka, E.H. Hansen, E.A.G. Zagatto, 'Flow Injection Analysis. Part VIII. Use of Ion-selective Electrodes for Rapid Analysis of Soil Extracts and Blood Serum. Determination of Potassium, Sodium and Nitrate', *Anal. Chim. Acta*, **88**, 1–16 (1977).
5. F.J. Krug, H. Bergamin Filho, E.A.G. Zagatto, S.S. Jorgensen, 'Rapid Determination of Sulphate in Natural Waters and Plant Digest by Continuous Flow Injection Turbidimetry', *Analyst*, **102**, 705–713 (1977).
6. E.A.G. Zagatto, F.J. Krug, H. Bergamin Filho, S.S. Jorgensen, B.F. Reis, 'Merging Zones in FIA. Part 2. Determination of Calcium, Magnesium and Potassium in Plant Materials by Continuous Flow Injection Atomic Absorption and Flame Emission Spectrometry', *Anal. Chim. Acta*, **105**, 279–284 (1979).

7. S.M. Ramasamy, M.S.A. Jabbar, H.A. Mottola, 'Flow Injection Analysis Based on Two Consecutive Reactions at a Gas–Solid Interface for Determination of Bromine and Chlorine', *Anal. Chem.*, **52**, 2062–2066 (1980).
8. S.M. Ramasamy, H.A. Mottola, 'Repetitive Determinations of Sulfur Dioxide Gas in Air Samples by Flow Injection and Chemical Reaction at a Gas–Liquid Interface', *Anal. Chem.*, **54**, 283–286 (1982).
9. B. Karlberg, G.E. Pacey, *Flow Injection Analysis. A Practical Guide*, Elsevier, Amsterdam, 1989.
10. J.L. Burguera (ed.), *Flow Injection Atomic Spectroscopy*, Marcel Dekker, New York, 1989.
11. M. Valcárcel, M.D. Luque de Castro, *Flow Injection Analysis: Principles and Applications*, Ellis Horwood, Chichester, 1981.
12. Z. Fang, *Flow Injection Atomic Absorption Spectrometry*, John Wiley & Sons, Chichester, 1995; Z. Fang, *Flow Injection Separation and Preconcentration*, John Wiley & Sons, New York, 1993.
13. J. Martínez Calatayud, *Flow Injection Analysis of Pharmaceuticals*, Taylor & Francis, London, 1996.
14. G.D. Christian, J. Ruzicka, 'Flow Injection Analysis: A Novel Tool for Plasma Spectroscopy', *Spectrochim. Acta*, **42B**, 157–167 (1987).
15. M.D. Luque de Castro, M. Valcárcel, 'Novel Flow-injection Approaches to Environmental Analysis', *Int. J. Environ. Anal. Chem.*, **38**, 171–183 (1990).
16. Z. Fang, S. Xu, G. Tao, 'Developments and Trends in Flow Injection Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1–24 (1996).
17. J.L. Burguera, M. Burguera, 'On-line Sample Pretreatment Systems Interfaced to Electrothermal Atomic Absorption Spectrometry', *Analyst*, **123**, 561–569 (1998).
18. J.R. Dean, O. Butler, A. Fisher, L.M. Graden, M.S. Cresser, P. Watkins, M. Cave, 'Atomic Spectrometry Update – Environmental Analysis', *J. Anal. Atom. Spectrom.*, **13**, 1R–56R (1998).
19. F. Lázaro, M.D. Luque de Castro, M. Valcárcel, 'Flow Injection Environmental Analysis: A Review', *Analusis*, **13**, 147–159 (1985).
20. J.F. van Staden, 'Flow-injection Analysis of Substances in Water: Part I. Anions. A Critical Review', *Water SA*, **13**, 197–208 (1987).
21. J. Atienza, M.A. Herrero, A. Maquiera, R. Purchades, 'Flow Injection Analysis of Seawater: Anionic and Organic Species', *Crit. Rev. Anal. Chem.*, **22**, 331–334 (1991).
22. J. Atienza, M.A. Herrero, A. Maquiera, R. Purchades, 'Flow Injection Analysis of Seawater: Cationic Species', *Crit. Rev. Anal. Chem.*, **23**, 1–14 (1992).
23. K.N. Andrew, N.J. Blundell, D. Price, P.J. Worsfold, 'Flow Injection Techniques for Water Monitoring', *Anal. Chem.*, **66**, 916A–922A (1994).
24. B.C. Madsen, R.J. Murphy, 'Flow-injection and Photometric Determination of Sulfate in Rain Water with Methylthymol Blue', *Anal. Chem.*, **53**, 1924–1926 (1981).
25. H.F.R. Reijnders, J.J. Van Staden, B. Griepink, 'Flow-through Determination of Sulfate in Water Using Different Methods: A Comparison', *Fresenius' J. Anal. Chem.*, **300**, 273–276 (1980).
26. O. Kondo, H. Miyata, K. Tóei, 'Determination of Sulfate in River Water by Flow-injection Analysis', *Anal. Chim. Acta*, **134**, 353–358 (1982).
27. S. Nakashima, M. Yagi, M. Zenki, M. Doi, K. Tóei, 'Determination of Sulfate in Natural Water by Flow-injection Analysis', *Fresenius' J. Anal. Chem.*, **317**, 29–31 (1984).
28. A. Sakuragawa, S. Nakayama, T. Okutani, 'Flow Injection Spectrophotometric Determination of Microamounts of Sulfate Ion in Surface Water and Sea Water Samples with a Barium Chromate Reaction Column', *Anal. Sci.*, **10**, 77–81 (1994).
29. K. Yakata, F. Sagara, I. Yoshida, K. Ueno, 'Determination of the Sulfate Ion Concentration in Rain Water by Flow Injection Analysis Incorporated with a Barium Chloranilate Reaction Column', *Anal. Sci.*, **6**, 711–714 (1990).
30. A.T. Faizullah, A. Townshend, 'Applications of Ion-exchange Mini-columns in a Flow Injection System for the Spectrophotometric Determination of Anions', *Anal. Chim. Acta*, **179**, 233–244 (1986).
31. A. Kojlo, J. Michalowski, M. Trajanowicz, 'Flow-injection Ultraviolet Spectrophotometric Determination of Sulfate in Natural Waters', *Anal. Chim. Acta*, **228**, 287–292 (1990).
32. E.A. Jones, 'Spectrophotometric Determination of Sulfate in Sodium Hydroxide Solutions by Flow Injection Analysis', *Anal. Chim. Acta*, **156**, 313–317 (1984).
33. M. Aoyagi, Y. Yasumasa, A. Nishida, 'Rapid Spectrophotometric Determination of Total Phosphorus in Industrial Wastewaters by Flow-injection Analysis Including a Capillary Digestor', *Anal. Chim. Acta*, **214**, 229–238 (1988).
34. I.D. McKelvie, B.T. Hart, J. Cardwell, R.W. Cattrall, 'Speciation of Dissolved Phosphorus in Environmental Samples by Gel-filtration and Flow Injection Analysis', *Talanta*, **40**, 1981–1993 (1993).
35. I.D. McKelvie, B.T. Hart, J. Cardwell, R.W. Cattrall, 'Spectrophotometric Determination of Dissolved Organic Phosphorus in Natural Waters Using In-line Photooxidation in Flow Injection', *Analyst*, **114**, 1459–1463 (1989).
36. K.E. Williams, S.J. Haswell, D.A. Barkley, G. Preston, 'Determination of Total Phosphate in Wastewaters by On-line Microwave Digestion Incorporating Colorimetric Detection', *Analyst*, **118**, 245–248 (1993).
37. R.L. Benson, I.D. McKelvie, B.T. Hart, I.C. Hamilton, 'Determination of Total Phosphorus in Waters and Wastewaters by On-line Microwave-induced Digestion and Flow-injection Analysis', *Anal. Chim. Acta*, **291**, 233–242 (1994).

38. A. Ríos, M.D. Luque de Castro, M. Valcarcel, 'Spectrophotometric Determination of Cyanide by Unsegmented Flow Methods', *Talanta*, **31**, 673–678 (1984).
39. Z. Zhu, Z. Fang, 'Spectrophotometric Determination of Total Cyanide in Wastewater in a Flow-injection System with Gas-diffusion Separation and Pre-concentration', *Anal. Chim. Acta*, **198**, 25–36 (1987).
40. D.J. Leggett, N.H. Chen, D.S. Mahadevappa, 'Rapid Determination of Residual Chlorine by Flow-injection Analysis', *Analyst*, **107**, 433–441 (1982).
41. D.J. Leggett, N.H. Chen, D.S. Mahadevappa, 'A Flow-injection Method for Analysis of Residual Chlorine by the DPD Procedure', *Fresenius' J. Anal. Chem.*, **315**, 47–50 (1983).
42. M. Zenki, H. Komatsubara, K. Tóei, 'Determination of Residual Chlorine in Tap Water by Flow-injection Spectrophotometry', *Anal. Chim. Acta*, **208**, 317–320 (1988).
43. J. Ruzicka, J.W.B. Stewart, E.A.G. Zagatto, 'Flow-injection Analysis. Part IV. Stream Sample Splitting and its Application to the Continuous Spectrophotometric Determination of Chloride in Brackish Waters', *Anal. Chim. Acta*, **81**, 387–396 (1976).
44. J.F. Van Staden, 'Automated Pre-valve Dilution in Flow-injection Analysis. The Automated Determination of Chloride in Surface, Ground and Domestic Water', *Fresenius' J. Anal. Chem.*, **322**, 36–41 (1985).
45. B. Rossner, G. Schwedt, 'Determination of Inorganic Anions. I. Photometric Analysis of Chloride in Air and Water', *Fresenius' J. Anal. Chem.*, **315**, 197–200 (1983).
46. B.C. Madsen, 'Utilization of Flow Injection with Hydrazine Reduction and Photometric Detection for the Determination of Nitrate in Rain-water', *Anal. Chim. Acta*, **53**, 437–441 (1981).
47. L. Anderson, 'Simultaneous Spectrophotometric Determination of Nitrite and Nitrate by Flow Injection Analysis', *Anal. Chim. Acta*, **110**, 123–128 (1979).
48. S. Nakashima, M. Yagi, M. Zenki, A. Takahashi, K. Tóei, 'Spectrophotometric Determination of Nitrite in Natural Waters by Flow Injection Analysis', *Anal. Chim. Acta*, **155**, 263–268 (1983).
49. A. Cerdá, M.T. Oms, R. Forteza, V. Cerdá, 'Evaluation of Flow-injection Methods for Ammonium Determination in Wastewater Samples', *Anal. Chim. Acta*, **311**, 165–173 (1995).
50. P. Carrero, J.L. Burguera, M. Burguera, C. Rivas, 'A Time-based Injector Applied to the Flow Injection Spectrophotometric Determination of Boron in Plant Materials and Soils', *Talanta*, **40**, 1967–1974, (1993).
51. A.O. Jacintho, E.A.G. Zagatto, B.F. Reis, L.C.R. Pessenda, F.J. Krug, 'Merging Zones in Flow Injection Analysis. Determination of Calcium in Natural Waters, Soil and Plant Materials with Glyoxal bis(2-hydroxyanil)', *Anal. Chim. Acta*, **130**, 361–368 (1981).
52. Y. Yuan, 'Simultaneous Spectrophotometric Determination of Calcium and Magnesium with Chlorophosphonazo-III by Flow Injection Analysis', *Anal. Chim. Acta*, **212**, 291–295 (1988).
53. G. Nakagawa, H. Wada, C. Wei, 'Spectrophotometric Determination of Calcium with a Flow Injection System', *Anal. Chim. Acta*, **145**, 135–141 (1983).
54. J.De García, M.L.M.F.S. Saravia, A.N. Araujo, J.L.F.C. Lima, M. del Valle, M. Poch. 'Evaluation of Natural Computation Techniques in the Modelling and Optimization of a Sequential Injection Flow Injection System for Colorimetric Iron(III) Determination', *Anal. Chim. Acta*, **348**, 143–150 (1997).
55. A.C. López da Conceicao, M.T. Tena, M.M. Correia dos Santos, M.L. Simoes Gonsalves, M.D. Luque de Castro, 'Flow Injection-assisted Optical Sensor for the Determination of Iron(II) and Iron(III) in Natural Waters', *Anal. Chim. Acta*, **343**, 191–197 (1997).
56. J.A. Schneider, J.F. Hornig, 'Spectrophotometric Determination of Lead in Tap Water with 5,10,15,20-tetra(4-*N*-sulfoethylpyridinium)porphyrin using Merging Zones Flow Injection', *Analyst*, **118**, 933–936 (1993).
57. O. Royset, 'Flow Injection Spectrophotometric Determination of Aluminum in Natural Water using Eriochrome Cyanine R and Cationic Surfactants', *Anal. Chem.*, **59**, 899–903 (1987).
58. I.M. Maximova, E.I. Morosanova, N.M. Kuz'min, Yu.A. Zolotov, 'Segmented Flow Analysis. On-line Microwave Sample Processing for the Analysis of Natural Waters', *Fresenius' J. Anal. Chem.*, **357**, 946–949 (1997).
59. B. Romberg, H. Muller, 'Photometric Screening-test for Heavy Metals Under Flow Injection Conditions using Extractive Determination with Dithizone', *Anal. Chim. Acta*, **353**, 165–172 (1997).
60. M.L. Balconi, M. Borgarello, R. Ferraroli, F. Realini, 'Chemical Oxygen Demand Determination in Well and River Waters by Flow Injection Analysis using a Microwave Oven During the Oxidation Step', *Anal. Chim. Acta*, **261**, 295–299 (1992).
61. A. Schmitt, L. Buttle, R. Uglow, K. Williams, S. Haswell, 'An Integrated Reduction Method for the Determination of Urea as Ammonia in Fresh Water Samples', *Anal. Chim. Acta*, **284**, 249–255 (1993).
62. S. Sadok, R.F. Uglow, S.J. Haswell, 'Determination of Trimethylamine Oxide in Seawater and Small Volumes of Seafood Tissue Extract using Flow Injection Analysis', *Anal. Chim. Acta*, **334**, 279–285 (1996).
63. F.J. Krug, F.H. Bergamin, E.A.G. Zagatto, S.S. Jorgensen, 'Rapid Determination of Sulfate in Natural Waters and Plant Digests by Continuous Flow-injection Turbidimetry', *Analyst*, **102**, 503–508 (1977).
64. W.B. Basson, J.F. Van Staden, 'Turbidimetric Determination of Sulfate with a Miniaturized Manifold on a Non-segmented High Speed Continuous Flow System', *Lab. Practice*, **25**, 863–865 (1980).

65. J.F. Van Staden, 'Automated Turbidimetric Determination of Sulfate in Surface, Ground and Domestic Water by Flow-injection Analysis', *Fresenius' J. Anal. Chem.*, **310**, 239–242 (1982).
66. F.J. Krug, E.A.G. Zagatto, B.F. Reis, F.O. Bahia, A.O. Jacintho, S.S. Jorgensen, 'Turbidimetric Determination of Sulfate in Plant Digests and Natural Waters by Flow Injection Analysis with Alternating Streams', *Anal. Chim. Acta*, **145**, 179–187 (1983).
67. S. Baban, D. Beetlestone, D. Betteridge, P. Sweet, 'The Determination of Sulfate by Flow-injection Analysis with Exploitation of pH Gradients and EDTA', *Anal. Chim. Acta*, **114**, 319–323 (1980).
68. J.F. Van Staden, 'Automated Pre-valve Sample Filtration in Flow-injection Analysis: Determination of Sulfate in Water Removing Suspended Solids and Color Before Sampling', *Fresenius' J. Anal. Chem.*, **312**, 438–440 (1982).
69. J.F. Van Staden, 'Flow-injection Turbidimetric Analysis of Sulfate in Water', *Water SA*, **12**, 43–50 (1986).
70. M.M.S. Filha, B.F. Reis, F. Krug, C.H. Collins, N. Baccan, 'Sulfate Preconcentration by Anion Exchange Resin in Flow Injection and its Turbidimetric Determination in Water', *Talanta*, **40**, 1529–1534 (1993).
71. T. Zaitsu, M. Maehara, K. Tóei, 'Flow-injection Analysis by using Turbidimetry for Chloride in River Water', *Bunseki Kagaku*, **33**, 149–153 (1984).
72. S. Olsen, L.C.R. Pessenda, J. Ruzicka, E.H. Hansen, 'Combination of Flow Injection Analysis with Flame Atomic Absorption Spectrophotometry: Determination of Trace Amounts of Heavy Metals in Polluted Sea Water', *Analyst*, **108**, 905–917 (1983).
73. P. Martínez-Jiménez, M. Gallego, M. Valcárcel, 'Preconcentration and Determination of Trace Amounts of Lead in Water by Continuous Precipitation in an Unsegmented-flow Atomic Absorption Spectrometric System', *Analyst*, **112**, 1233–1236 (1987).
74. H. Chen, J. Jin, Y. Wang, 'Flow Injection On-line Coprecipitation–Preconcentration System using Copper-(II) Diethyldithiocarbamate as Carrier for Flame Atomic Absorption Spectrometric Determination of Cadmium, Lead and Nickel in Environmental Samples', *Anal. Chim. Acta*, **353**, 181–188 (1997).
75. L. Nord, B. Karlberg, 'Sample Preconcentration by Continuous Flow Extraction with a Flow Injection Atomic Absorption Detection System', *Anal. Chim. Acta*, **145**, 151–158 (1983).
76. M. Gallego, M. Silva, M. Valcárcel, 'Indirect Atomic Absorption Determination of Anionic Surfactants in Waste Waters by Flow Injection Continuous Liquid–Liquid Extraction', *Anal. Chem.*, **58**, 2265–2269 (1986).
77. R. Ma, W. Van Mol, F. Adams, 'Determination of Cadmium, Copper and Lead in Environmental Samples. An Evaluation of Flow Injection On-line Sorbent Extraction for Flame Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **285**, 33–43 (1994).
78. S.R. Bysouth, J.F. Tyson, P.B. Stockwell, 'Use of Masking Agents in the Determination of Lead in Tap Water by Flame Atomic Absorption Spectrometry with Flow Injection Preconcentration', *Analyst*, **115**, 571–573 (1990).
79. M. Sperling, S. Xu, B. Welz, 'Determination of Chromium(III) and Chromium(VI) in Water using Flow-injection On-line Preconcentration and Flame Atomic Absorption Spectrometric Detection', *Anal. Chem.*, **64**, 3101–3108 (1992).
80. J.L. Burguera, M. Burguera, 'Flow Injection for Automation in Atomic Spectrometry', *J. Anal. At. Spectrom.*, **12**, 643–651 (1997).
81. R.L. Ma, W. Van Mol, F. Adams, 'Selective Flow-injection Sorbent Extraction for Determination of Cadmium, Copper and Lead in Biological and Environmental Samples by Graphite-furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **293**, 251–260 (1994).
82. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Determination of Trace Metals in Seawater and the On-line Removal of Matrix Interferences by Flow Injection with Inductively Coupled Plasma Mass Spectrometric Determination', *J. Anal. Atom. Spectrom.*, **9**, 935–938 (1994).
83. R.A. Nickson, S.J. Hill, P.J. Worsfold, 'Behavior of Matrix Cations (Ca^{2+}, K^{+}, Mg^{2+} and Na^{+}) During On-line Preconcentration and Atomic Spectrometric Detection of Trace Metals in Natural Waters', *Anal. Chim. Acta*, **351**, 311–317 (1997).
84. M.M. Hidalgo, M.M. Gómez, M.A. Palacios, 'Trace Enrichment and Measurement of Platinum by Flow Injection Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **354**, 420–423 (1996).
85. P.L. Lu, K.S. Huang, S.J. Jiang, 'Determination of Traces of Copper, Cadmium and Lead in Biological and Environmental Samples by Flow Injection Isotope Dilution Inductively Coupled Plasma/Mass Spectrometry', *Anal. Chim. Acta*, **284**, 181–188 (1993).
86. L. Bordera, V. Hernandis, A. Canals, 'Automatic Flow Injection System for the Determination of Heavy Metals in Sewage Sludge by Microwave Digestion and Detection by Inductively Coupled Plasma/Atomic Emission Spectrometry (MW/ICP/AES)', *Fresenius' J. Anal. Chem.*, **355**, 112–119 (1996).
87. A. Stroh, U. Vollkopf, 'Optimization and Use of Flow-injection Vapor Generation Inductively Coupled Plasma Mass-spectrometry for the Determination of Arsenic, Antimony and Mercury in Water and Sea-water at Ultratrace Levels', *J. Anal. At. Spectrom.*, **8**, 35–40 (1993).
88. J.A.C. Broekaert, F. Leis, 'An Injection Method for the Sequential Determination of Boron and Several Metals in Waste-waters Samples by Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Chim. Acta*, **109**, 73–83 (1979).
89. J. Szpunar-Lobinska, M. Ceulemans, R. Lobinski, F.C. Adams, 'Flow Injection Sample Preparation for Organotin Speciation Analysis of Water by Capillary Gas

Chromatography-microwave Induced Plasma Atomic Emission Spectrometry', *Anal. Chim. Acta*, **278**, 99–113 (1993).
90. D.L. Tsalev, M. Sperling, B. Welz, 'On-line Microwave Sample Pre-treatment for Hydride Generation and Cold Vapour Atomic Absorption Spectrometry', *Analyst*, **117**, 1729–1733 (1992).
91. T. Guo, J. Baasner, 'Technical Note: Using FIMS to Determine Mercury Content in Sewage Sludge, Sediment and Soil Samples', *J. Autom. Chem.*, **18**, 221–223 (1996).
92. C.P. Hanna, S. McIntosh, 'Determination of Total Hg in Environmental Samples with On-line Microwave Digestion Coupled to a Flow Injection Mercury System (FIMS)', *At. Spectrosc.*, **16**, 106–114 (1995).
93. K.J. Lamble, S.J. Hill, 'Determination of Mercury in Slurried Samples by Both Batch and On-line Microwave Digestion–Cold Vapor Atomic Fluorescence Spectrometry', *J. Anal. Atom. Spectrom.*, **11**, 1099–1103 (1996).
94. A. Morales-Rubio, M.L. Mena, C.E. Mcloed, 'Rapid Determination of Mercury in Environmental Materials Using On-line Microwave Digestion and Atomic Fluorescence Spectrometry', *Anal. Chim. Acta*, **308**, 364–370 (1995).
95. J.L. Burguera, M. Burguera, A.F. Danet, 'On-line Total Arsenic Determination in Urine and Geothermal Fluids by Flow Injection-Hydride Generation Atomic Absorption Spectrometry with Microwave Oven Sample Mineralization/Oxidation', *Rev. Roum. Chim.*, **43**, 301–306 (1998).
96. M.A. López-Gonzálvez, M.M. Gómez, C. Cámara, M.A. Palacios, 'On-line Microwave Oxidation for the Determination of Organoarsenic Compounds by High-performance Liquid Chromatography-Hydride Generation Atomic Absorption Spectrometry', *J. Anal. Atom. Spectrom.*, **9**, 291–295 (1994).
97. X.Ch. Le, W.R. Cullen, J. Kenneth, 'Effect of Cysteine on the Speciation of Arsenic by Using Hydride Generation Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **285**, 277–285 (1994).
98. J.L. Burguera, M. Burguera, C. Rivas, P. Carero, 'On-line Cryogenic Trapping with Microwave Heating for the Determination and Speciation of Arsenic by Flow Injection/Hydride Generation/Atomic Absorption Spectrometry', *Talanta*, **45**, 531–542 (1998).
99. J.L. Burguera, M. Burguera, A. Townshend, 'The Principles, Applications and Trends of Flow Injection Analysis for Monitoring Chemiluminescent Reactions', *Rev. Roum. Chim.*, **27**, 879–883 (1982).
100. M. Yamada, T. Nakada, S. Suzuki, 'The Determination of Sulfite in a Flow Injection System with Chemiluminescence Detection', *Anal. Chim. Acta*, **147**, 401–404 (1983).
101. J.S. Lancaster, P.J. Worsfold, A. Lynes, 'Flow Injection Procedure for the Determination of Tertiary Amines in Water and Sea Water Using Chemiluminescence Detection', *Analyst*, **114**, 1659–1661 (1989).
102. H.K. Chung, H.S. Bellamy, P.K. Dasgupta, 'Determination of Aqueous Ozone for Potable Water Treatment Applications by Chemiluminescence Flow Injection Analysis – A Feasibility Study', *Talanta*, **39**, 593–598 (1992).
103. J.S. Lancaster, P.J. Worsfold, A. Lynes, 'Determination of a Nonionic Surfactant in Aqueous Environmental Samples by Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **239**, 189–194 (1990).
104. P. Mikuska, Z. Vecera, Z. Zdrahal, 'Flow-injection Chemiluminescence Determination of Ultra-low Concentrations of Nitrite in Water', *Anal. Chim. Acta*, **316**, 261–268 (1995).
105. T. Aoki, M. Wakabayashi, 'Simultaneous Flow-injection Determination of Nitrate and Nitrite in Water by Gas-phase Chemiluminescence', *Anal. Chim. Acta*, **308**, 308–312 (1995).
106. D. Price, P.J. Worsfold, R.F.C. Mantoura, 'Determination of Hydrogen Peroxide in Sea-water by Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **298**, 121–128 (1994).
107. T.P. Chapin, K.S. Johnson, K.H. Coale, 'Rapid Determination of Manganese in Sea-water by Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **249**, 469–478 (1991).
108. K.H. Coale, K.S. Jonson, P.M. Stout, C.M. Sakamoto, 'Determination of Copper in Seawater Using a Flow Injection Method with Chemiluminescence Detection', *Anal. Chim. Acta*, **266**, 345–351 (1992).
109. F. Carrillo, C. Perez, C. Camara, 'Sensitive Flow-injection Spectrofluorometric Method to Determine Aluminium(III) in Water', *Anal. Chim. Acta*, **262**, 91–96 (1992).
110. R. Petty, W.C. Michel, J.P. Snow, K.S. Johnson, 'Determination of Total Primary Amines in Seawater and Plant Nectar with Flow Injection Sample Processing and Fluorescence Detection', *Anal. Chim. Acta*, **142**, 299–304 (1982).
111. F. Lázaro, M.D. Luque de Castro, M. Valcárcel, 'Stopped-Flow Injection Determination of Copper(II) at the ng ml^{-1} Level', *Anal. Chim. Acta*, **165**, 177–185 (1984).
112. H. Hwang, K. Dasgupta, 'Fluorometric Flow Injection Determination of Aqueous Peroxides at Nanomolar Level Using Membrane Reactors', *Anal. Chem.*, **58**, 1521–1524 (1986).
113. H.K. Chung, H.S. Bellamy, P.K. Dasgupta, 'Determination of Aqueous Ozone for Potable Water Treatment Applications by Chemiluminescence Flow Injection Analysis – A Feasibility Study', *Talanta*, **39**, 593–598 (1992).
114. K. Dasgupta, H.C. Yang, 'Trace Determination of Aqueous Sulfite, Sulfide and Methanethiol by Fluorometric Flow-injection Analysis', *Anal. Chem.*, **58**, 2839–2844 (1986).
115. H. Ma, L. Jin, H. Yan, 'Flow-injection Analysis of Traces of Free Cyanide in Surface and Ground Waters with an

Amperometric Flow-through Detector', *Kexue Tongbao (Science Bulletin)*, **8**, 1145–1148 (1981).
116. J.R. Farell, P.J. Iles, Y.J. Yuan, 'Determination of Arsenic by Hydride Generation Gas Diffusion Flow Injection Analysis with Electrochemical Detection', *Anal. Chim. Acta*, **334**, 193–197 (1996).
117. A. Ríos, M.D. Luque de Castro, M. Valcarcel, H.A. Mottola, 'Electrochemical Determination of Sulfur Dioxide in Air Samples in Closed-loop Flow Injection Systems', *Anal. Chem.*, **59**, 666–670 (1987).
118. J.F. Van Staden, 'A Coated Tubular Solid-state Chloride-selective Electrode in Flow-injection Analysis', *Anal. Chim. Acta*, **179**, 407–417 (1986).
119. J.F. Van Staden, 'Flow-injection Determination of Inorganic Bromide in Soils with a Coated Tubular Solid-state Bromide-selective Electrode', *Analyst*, **112**, 595–599 (1987).
120. C. Okumoto, M. Nagashima, S. Mizoiri, M. Kazama, K. Akyiama, 'Flow-injection Analysis of Cyanide in Wastewater from Metal Plating Processes', *Eisei Kagaku*, **30**, 7–13 (1984).
121. G. Schulze, C.Y. Liu, M. Brodowski, O. Elsholz, W. Frenzel, J. Moller, 'Different Approaches to the Determination of Ammonium Ions at Low Levels by Flow Injection Analysis', *Anal. Chim. Acta*, **241**, 121–136 (1988).
122. J. Wang, D. Larson, N. Foster, S. Armalis, J. Lu, X. Rong-Rong, K. Olsen, A. Zirino, 'Remote Electrochemical Sensor for Trace Metal Contaminants', *Anal. Chem.*, **67**, 1481–1485 (1995).
123. C. Colombo, C.M.G. van den Berg, A. Daniel, 'A Flow Cell for On-line Monitoring of Metals in Natural Waters by Voltammetry with a Mercury Drop Electrode', *Anal. Chim. Acta*, **346**, 101–111 (1997).
124. A. Safavi, M. Pakniat, N. Maleki, 'Design and Construction of a Flow System for Determination of Cu(II) Ions in Water by Means of a Chemically Modified Carbon Paste Electrode', *Anal. Chim. Acta*, **335**, 275–282 (1996).
125. M.T.S.D. Vasconcelos, M.F.C. Leal, 'Speciation of Cu, Pb, Cd and Hg in Waters of the Oporto Coast in Portugal, Using Pre-concentration in a Chelamine Resin Column', *Anal. Chim. Acta*, **353**, 189–198 (1997).
126. F. Lázaro, M.D. Luque de Castro, M. Valcárcel, 'Flow Injection Environmental Analysis: Review', *Analusis*, **13**, 147–159 (1985).
127. M.D. Luque de Castro, M. Valcárcel, 'Novel Flow-injection Approaches to Environmental Analysis', *Internat. J. Environ. Anal. Chem.*, **38**, 171–183 (1990).
128. M. Trojanowicz, R.L. Benson, P.J. Worsfold, 'Recent Developments in Water-quality Monitoring by Flow Injection Analysis', *Trends Anal. Chem.*, **10**, 11–17 (1991).
129. M. Aguso, A. Rios, M. Valcárcel, 'Assessment of Analytical Quality in Water Analysis by Flow-injection Methods', *Trends Anal. Chem.*, **13**, 409–414 (1994).
130. K.N. Andrews, N.J. Blundell, D. Prices, P.J. Worsfold, 'Flow-injection Techniques for Water Monitoring', *Anal Chem.*, **66**, A916–A922 (1994).
131. Z.L. Zhi, A. Rios, M. Valcárcel, 'Direct Processing and Analysis of Solid and Other Complex Samples with Automatic Flow Injection Systems', *Crit. Rev. Anal. Chem.*, **26**, 239–260 (1996).
132. J.F. Tyson, 'Flow Injection Atomic Spectrometry', *Spectrochim. Acta Rev.*, **14**, 169–233 (1991).
133. M. Burguera, J.L. Burguera, 'Microwave-assisted Sample Decomposition in Flow Analysis', *Anal. Chim. Acta*, **366**, 63–80 (1998).

Formaldehyde, Environmental Analysis of

S.S. Srivastava and K. Maharaj Kumari
Dayalbagh Educational Institute, Dayalbagh Agra, India

1 Introduction **1**
2 Properties **4**
2.1 Physical Properties 4
2.2 Chemical Properties 4
3 Methods of Manufacture **4**
4 Uses **5**
5 Extraction of Formaldehyde from Different Matrices **5**
5.1 Air 5
5.2 Rainwater/Drinking Water/ Wastewater 5
5.3 Toothpaste 5
5.4 Tablet 5
5.5 Formal 5
5.6 Crystalline Formal 5
5.7 Cellulose 5
5.8 Milk 5
5.9 Urine 5
6 Analysis **5**
6.1 Qualitative Analysis (Snell) 5
6.2 Quantitative Analysis 6
Acknowledgments **31**
Abbreviations and Acronyms **31**
Related Articles **32**
References **32**

Formaldehyde is a ubiquitous component of both the remote atmosphere and polluted urban atmospheres. Interest in ambient concentrations of formaldehyde derives both from concerns over health effects and from the primary role that formaldehyde plays in the tropospheric chemistry cycle. It is an important intermediate product occurring in all regions of the atmosphere, and arises from the oxidation of methane and other biogenic and anthropogenic hydrocarbons. In spite of the increasing interest in the levels and fate of formaldehyde in the environment, there is at present no simple method available for the determination of trace amounts (i.e. nanograms) of carbonyls in air, water, and other environmental samples. This article reviews all the available methods (qualitative and quantitative) for the analysis of formaldehyde in various matrices, including sampling of atmospheric formaldehyde. The various methods discussed include colorimetric, titrimetric, polarographic, and chromatographic methods. Fourier transform infrared spectroscopy (FTIR) and tunable diode laser absorption spectroscopy (TDLAS) have been reviewed, and fluorimetric methods of analysis have also been dealt with. The article includes a comparison of the various methods with respect to sensitivity, detection limits and interference.

1 INTRODUCTION

Formaldehyde (HCHO) is a ubiquitous component of both the remote atmosphere and polluted urban atmospheres.[1] This, together with the uniqueness of the chemistry it is involved in, make it a particularly important compound. Interest in ambient concentrations of formaldehyde derives both from concerns over health effects[2] and from the primary role that formaldehyde plays in the tropospheric chemistry cycle.[3] Formaldehyde is unique in that it acts as a funnel through which much of the carbon flux passes during the course of oxidation to CO and CO_2.[4]

In spite of the increasing interest in the levels and fate of aldehydes and other carbonyl compounds in the environment, there is at present no simple method available for the determination of trace amounts (i.e. nanograms) of carbonyls in air, water, and other environmental samples. In the field of atmospheric chemistry alone, measurements of parts per billion (ppb) levels of carbonyls in ambient air or simulated atmospheres (''smog chamber studies'') are critical to the development of a better understanding of hydrocarbon photochemistry, oxidant formation and removal processes including scavenging by precipitation, and atmospheric aerosols. Specifically, accurate formaldehyde measurements are important in constraining and validating photochemical models of the troposphere. Other important areas requiring measurement of trace levels of carbonyls in air include source and near-source sampling (engine exhausts, industrial emissions, etc.), and health related studies (e.g. indoor pollution, workplace environments etc.), all of which have important regulatory implications. In addition, formaldehyde is listed as a carcinogen or suspected carcinogen and its workplace exposure needs to be monitored.

Typical concentrations observed range from 0.5 ppb in remote areas to 75 ppb during severe air pollution episodes in urban areas.[5] In the late 1960s formaldehyde levels greater than 120 ppbv were observed by unambiguous spectroscopic techniques FTIR in polluted areas

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

near Los Angeles.[6] Between 1980 and 1985 the few data available show that the average concentrations in major cities were between 3 and 16 ppbv with maxima as high as 68 ppbv in Downey, California.[7] Average concentrations for the more remote Colorado mountains are around 1 ppbv during the summer and fall months,[8] while values between 50 and 500 pptv have been recorded for the Mauna Loa Observatory, Hawaii.[9,10]

About 24 million metric tons (Mt) of 37% formaldehyde solution are produced worldwide per year, of which 4.2 million Mt are in North America, 6.2 million Mt are in Europe and 6.3 million Mt are in Asia.[11] The end-uses of formaldehyde in the USA and in Europe are shown in Figure 1. The formaldehyde solution is incorporated into urea–formaldehyde and phenol–formaldehyde resins, which are widely used in industry for the production of foam insulation, particle board and plywood.[12] The particularly simple structure of formaldehyde results in extreme reactivity. Formaldehyde and ammonia yield methenamine, or hexamethylene tetramine, which is used as a urinary antiseptic. Nitration of methenamine gives the explosive cyclonite, also known as research development explosive (RDX). Formaldehyde reacts in the presence of $Ca(OH)_2$ to give pentaerythritol, the tetranitrate of which is the explosive pentaerythritol tetranitrate (PETN). The ready reaction of formaldehyde with protein leads to its use in the tanning industry and in treating various vegetable proteins to render them fibrous. The reactivity with proteins is also the basis for the use of formaldehyde as a disinfectant and embalming agent and as a soil sterilant.

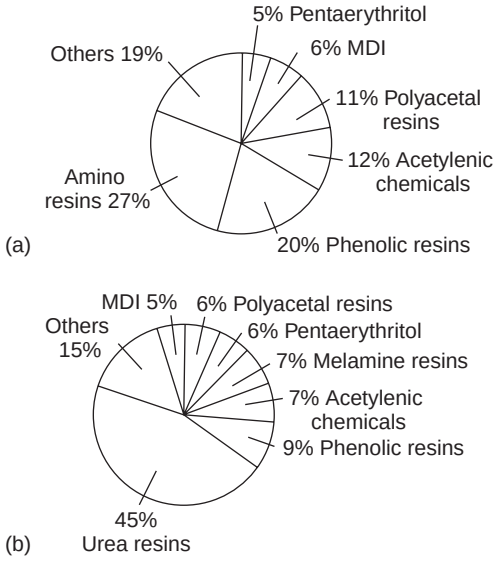

Figure 1 End-uses of formaldehyde: (a) United States, 1997 total 4.2 million Mt, (b) Europe, 1997 total 6.2 million Mt.[11] MDI, methane diphenyl di-isocyanate.

Formaldehyde is an important intermediate product occurring in all regions of the atmosphere, and arises from the oxidation of methane and other biogenic and anthropogenic hydrocarbons. Formaldehyde serves both as a major source of the hydroperoxyl radical ($HO_2^{\bullet}$) and as a means of sequestering the hydroxyl radical ($^{\bullet}OH$) in the form of HO_2. Formaldehyde is also a major source of carbon monoxide in air masses not influenced by anthropogenic sources. In the troposphere, where the concentration of methane is considerably higher than that of non-methane hydrocarbons (NMHCs), methane becomes the dominant formaldehyde precursor. At the earth's surface, local sources of NMHCs also become important in producing formaldehyde. Formaldehyde is generated from the oxidation of biogenic hydrocarbons such as isoprene and the terpenes.[13–15] Formaldehyde is also anthropogenically generated directly from incomplete combustion processes and through secondary processes.[16] The predominant source of urban formaldehyde is believed to be the photooxidation of many hydrocarbons that are present in the atmosphere.[17] In rural areas of dense vegetation biogenic sources are often the predominant precursor. For example, isoprene oxidation (indicated by reactions with either OH or O_3) efficiently forms formaldehyde along with several other key atmospheric species.

Many foundries use furan resins as binding agents for sand, resulting in emissions of phenol, furfuryl alcohol and formaldehyde. In indoor environments, formaldehyde is emitted in substantial amounts as phenol and urea resins are incorporated into foam insulation, plywood and particle board. Growing concerns over the toxicity and the resultant properties of formaldehyde has created a significant demand for new methods to monitor exposures.

The use of titrimetric methods to measure formaldehyde concentrations is widespread for industrial hygiene applications and source measurements but the methods are not sensitive enough for ambient analysis. With colorimetric methods formaldehyde is converted into a chromogen by the addition of such compounds as chromotropic acid, pararosaniline, acetylacetone or 3-methyl-2-benzothiazolone hydrazone (MBTH) and the amount of chromogen is measured spectrophotometrically. They are among the earliest known of the refined versions,[18] are reasonably sensitive, have low detection limits and can be automated[19] but color development is relatively slow and sensitivity is limited.

Most of the present techniques for sampling atmospheric formaldehyde involve running a stream of atmospheric air through either a liquid or a solid sorbent. As the air passes through the sorbent it is scrubbed free of aldehydes. Liquid sorbents are normally placed in a small vessel called an impinger. Impingers must normally be

refrigerated if they are to be operated for long periods of time, otherwise the media will evaporate. The most common liquid medium for trapping formaldehyde is water. Solid sorbents alone do not have a high collection efficiency unless they are coated with a scavenger compound such as acidified 2,4-dinitrophenylhydrazine (DNPH), 2-hydroxymethylpiperidine or *N*-benzylene-ethanolamine. Low sampling rates of 1–1.5 L min^{-1} of air must be used, with typical sampling times of 20 min or several hours. Thus determinations of low concentrations of formaldehyde require rather long sampling times followed by rather slow laboratory analysis.

Chromatographic methods, including ion chromatography (IC) as formate, gas chromatography (GC), and high-performance liquid chromatography (HPLC), have been used for formaldehyde analysis. Direct analysis in urban air using GC has been less successful due to poor sensitivity.[20] However, derivatization of formaldehyde either in the sampling device or immediately after collection has improved stability of samples and allows quantitation of low levels of formaldehyde. GC of oxazolidine or the dinitrophenyl hydrazone derivative and especially HPLC of the DNPH derivative has been shown to provide adequate sensitivity.[21] Capillary GC has also been used

Table 1 Sensitivity and interferences of different quantitative methods

Method	Sensitivity	Interferences	Where used	Reference
Titrimetric methods				
Sodium sulfite	1.6×10^{-7} M	Aldehydes, methyl ketone	Field tests, air	Reynolds and Irwin[35]
Alkaline peroxide	–	Large amount of methanol		
Iodimetric	2.0×10^{-5} M	Organic impurities, ethyl alcohol, acetone	Small quantity of pure HCHO	Reynolds and Irwin[35]
Mercurimetric	–	Organic peroxides		
Potassium cyanide	–	Acetaldehyde (>5%) acetone (>5%)		
Hydroxylamine	3.3×10^{-6} M	Ferric salts decompose the reagent		Reynolds and Irwin[35]
Colorimetric methods				
Chromotropic acid	0.3×10^{-7} M	Acrolein, glyceraldehyde, β-hydroxy propionaldehyde, acetaldehyde		
MBTH	30 ppb	Aromatic amine azo dye, amino heterocycles	Air, rainwater, drinking water, wastewater	
Nash	0.05 ppb	Acetaldehyde, acrolein, propionaldehyde		
Pararosaniline	–			Miksch et al.[18]
Polarographic method	1.1 mg m^{-3}		Air, incomplete combustion	Septon and Ku[36]
Chromatographic methods				
Direct GC (FID detector)	0.3–0.8 ppm (in air)		HCHO in solution and high purity gas	Bombaugh and Bull[37]
GC using mass spectrophotometer	0.3 ppb		Air and rainwater	Yokouchi et al.[38]
HPLC, pre-column derivatization (DNPH)	2.5 g		Air and rainwater	Beasley et al.[39]
HPLC, post-column derivatization (MBTH)	4.8×10^{-8} M		Air and rainwater	Kim et al.[40]
IEC	–	Any substance that has a retention time coinciding with that of the formate	Air and rainwater	
FTIR spectroscopy	6 ppb		Air, vehicular emissions	Tuazon et al.[24]
Fluorimetric methods	120 ppt		Air	Lazrus et al.[28]
TDLAS	0.3 ppb		Air, tropospheric measurements	Mackay et al.[27]

FID, flame ionization detection; IEC, ion exchange chromatography.

to analyze formaldehyde derivatives.[22] Several of these methods suffer from stability problems and the need for expensive and specialized equipment for analysis. Faster real-time measurements are possible with spectroscopic methods. Spectroscopic techniques have included differential ultraviolet (UV) absorption,[23] FTIR.[24] and TDLAS.[25] These techniques generally employ long atmospheric paths or use multiple pass cells to attain adequate sensitivity. Sawicki et al.[26] studied a very large number of spectrometric methods for formaldehyde analysis and concluded that spectrofluorimetric determination with 2,4-pentanediene should be the method of choice when sensitivity, specificity and mild reaction conditions are desired.

Current techniques[8] employed in measuring ambient formaldehyde concentrations include TDLAS,[27] continuous scrubbing fluorimetric detection (CSFD),[28] DNPH derivatization techniques employing cartridges[29] and coils,[30] FTIR[31] and differential optical absorption spectroscopy (DOAS).[32] The first three classes of techniques were compared at a rural location in North Carolina in a study published by Kleindienst et al.[33] and again in the Los Angeles basin by Lawson et al.[34] (this study also included FTIR and DOAS measurements). In relatively clean rural air (5–15 ppbv) all methods appear to provide reasonably good agreement to within approximately 35%.[8] Table 1 shows different methods used for the estimation of formaldehyde for different matrices with respect to sensitivity and interferences.

2 PROPERTIES

2.1 Physical Properties

Formaldehyde (methanal) is a gas (bp $-21\,^{\circ}C$). Monomeric formaldehyde gas is characterized by a pungent odor and is extremely irritating to the mucous membrane of the eye, nose and throat even when present in concentrations as low as 20 ppm.[41] Formaldehyde is handled either as an aqueous solution (formalin) or as one of the solid polymers paraformaldehyde ($(HCHO)_2$) or trioxane ($(CH_2O)_3$). It is extremely soluble in water.

Monomeric formaldehyde is available commercially only in solutions stabilized with methanol. In most cases, these are aqueous solution with methanol contents ranging from 1 to 15%.

2.2 Chemical Properties

The chemical properties of formaldehyde are given in Table 2.

Table 2 Chemical properties of formaldehyde

Reaction conditions	Product
At high temperature	$CO + H_2$
Catalytic hydrogenation, $2H^+$	CH_3OH (Methyl alcohol)
$4NH_3$	$(CH_2)_6N_4 + 6H_2O$ (80%) Hexamethylene tetramine
$C_2H_5NH_2 + HCOOH$, $100\,^{\circ}C$	$C_2H_5NHCH_3 + CO_2 + H_2O$ Ethylmethylamine
H_2N-NH_2	$H_2C=N-NH_2H_2O$ Hydrazone
$C_6H_5NHNH_2$	$C_6H_5NHN=CH_2$ Phenyl hydrazones
NH_2OH	$OH_2=NOH$ Oxime
$H_2C=N-NH_2$	$H_2C=NN=CH_2$
$NHNH_2CONH-NH_2$	$NH_2NHCONHN=CH_2 + H_2O$ Semicarbazones
RSH	$H_2C(SR)_2 + H_2O$ Mercaptan
RMgX, H^+	RCH_2OH Alcohol
$LiAlH_4$	$CH_3OH + AlCl_3 + LiCl$ Methyl alcohol
PCl_5	$H_2CCl_2 + POCl_3$ Dichloromethane
NaOH	$HCO_2Na + CH_3OH$ Sodium formate
CH_3CHO, OH^-	CH_2OHCH_2CHO Glycolaldehyde
$C_6H_5COCH_3(CH_3)_2NH$	$C_6H_5COCH_2N(CH_3)_2$ Mannich base
C_6H_5OH	Bakelites
In dilute solution	$CH_2(OH)_2$ Methylene glycol
Evaporation (solid)	$CH_2OH(CH_2O)_nCH_2OH$ ($n = 6–50$) Paraformaldehyde
Conc. H_2SO_4	$(CH_2O)_n \cdot 2H_2O$ ($n > 100$) POMs
At room temperature	$(CH_2O)_3$ Trioxane
$Ca(OH)_2$	$C_6H_{12}O_6$ Formose

POM, polyoxymethylene.

3 METHODS OF MANUFACTURE

Formaldehyde is prepared industrially in three different ways. The first method involves the dehydrogenation of methanol. Methanol vapors are passed over a heated silver catalyst at 300 °C, as shown in Scheme 1.

$$CH_3OH \xrightarrow[300\,^{\circ}C]{Ag} HCHO + H_2O$$

Scheme 1

In the second method methanol is oxidized in air. Methanol vapors and a limited amount of air are passed

over a heated silver catalyst at 450 °C, as shown in Scheme 2. In this method, the condensate obtained is a mixture of formaldehyde, methanol and water. The excess methanol is removed by distillation and the resulting mixture is known as formalin (40% HCHO, 8% CH_3OH and 52% H_2O).

$$2CH_3OH + O_2 \xrightarrow[450\,^\circ C]{Ag} 2HCHO + 2H_2O$$

Scheme 2

Thirdly methane or propane–butane mixtures (from natural gas) can be oxidized in the presence of various metallic oxides, as shown in Scheme 3.

$$CH_4 + O_2 \xrightarrow{\text{metallic oxides}} HCHO + H_2O$$

Scheme 3

4 USES

Formaldehyde is a powerful disinfectant and antiseptic and is used for preserving anatomical specimens. Its main uses are in making plastics. Formaldehyde is condensed with urea and phenols to make plastics, with ammonia to give hexamethylenetetramine, a urinary antiseptic, and is also used to make RDX. Formaldehyde binds to the amino group of proteins, thus rendering toxic proteins harmless, but does not change their immunological properties. This is the basis for the production of the polio vaccine, and the tetanus and diphtheria toxoids.

5 EXTRACTION OF FORMALDEHYDE FROM DIFFERENT MATRICES

5.1 Air

Formaldehyde can be sampled from air using bubblers, impingers, mist chambers, sorbent cartridges, etc., in different media. The more commonly used media include water, MBTH, and chromotropic acid. Details of specific sampling methods are covered in the appropriate section.

5.2 Rainwater/Drinking Water/Wastewater

These samples can be analyzed directly by any colorimetric or chromatographic method.

5.3 Toothpaste

1 g of the paste is dissolved in 5 mL of 0.1 N NaOH diluted to 100 mL with distilled water. An aliquot is used to estimate the formaldehyde concentration.

5.4 Tablet

The tablet is weighed and 200 mL water is added. The solution is refluxed for 30 min, then the condenser is cooled and rinsed out. The solution is diluted to 500 mL and filtered. An aliquot of the filtrate is used to estimate the formaldehyde concentration.

5.5 Formal

The sample is dissolved in water and is diluted to a volume such that the equivalence of formaldehyde is approximately 0.05 mg mL^{-1}. This solution is used to estimate the formaldehyde concentration.

5.6 Crystalline Formal

The sample (20–40 mg) is dissolved in 1 : 2 H_2SO_4 and is diluted to 1 L. The solution is heated and incubated for 2 h and then cooled. An aliquot of the solution is used to estimate the formaldehyde concentration.

5.7 Cellulose

The sample (0.1 g) is added to 100 mL of 1 : 2 H_2SO_4. The solution is sealed and incubated at 90 °C for 2 h and then cooled. An aliquot of the solution is used for estimation.

5.8 Milk

Milk samples can be analyzed directly using any colorimetric method or any standard technique.

5.9 Urine

A few drops of 50% NaOH solution are added to the sample to precipitate the phosphate, and the mixture is then shaken well and filtered. An aliquot of the solution is used to estimate the formaldehyde concentration.

6 ANALYSIS

6.1 Qualitative Analysis (Snell[42])

6.1.1 Colorimetric Spot Test Method

Reagent	Color	Sensitivity	Interferences
1. Modified Schiff's reagent	Blue	<1 ppm	Acrolein, glyoxylic acid
2. Chromotropic acid	Violet	3 ppm	Glyceraldehyde, furfural
3. Potassium ferricyanide+ phenylhydrazine	Red	<15 µg	Higher aldehydes

hydrochloride +sodium hydroxide			
4. Acetyl acetone +ammonium acetate + acetic acid	Yellow	–	Large amount of acetaldehyde
5. Phenylhydrazine	Red	1 μg	Not specified
6. Ammoniacal silver nitrate (Tollens' reagent)	Black ppt	2 ppm	Not specified
7. 2,7-Dihydronaphthalene	Violet	0.5–2.0 μg mL^{-1}	Acrolein
8. Morphine hydrochloride +sulfuric acid	Violet ring at liquid junction	<5%	Not specified
9. 2-Hydroxycarbazole sulfuric acid	Dark blue	0.004 μg	Pyruvaldehyde
10. 2-Methyl-3-benzohydrazine hydrochloride +ferric chloride	Deep blue	0.1 μg	Not specified

6.1.2 Derivatives

6.1.2.1 Methylene Bismethone A few drops of a 5–10% solution of methone or dimedon (dimethyl dihydro resorcinol), and 5,5-dimethylcyclohexanediene in ethanol are added to the sample which has been made slightly acidic with acetic acid. A precipitate of methylene bismethone is formed, and, recrystallized from methanol and water (mp 189 °C).

6.1.2.2 Formaldehyde p-*nitrophenyl Hydrazone*
A few drops of *p*-nitrophenyl hydrazine hydrochloride in mildly acidic ethanol are added to a solution containing formaldehyde. A precipitate of formaldehyde *p*-nitrophenyl hydrazone is formed, and recrystallized from ethanol (mp 181–182 °C).

6.1.2.3 Formaldehyde 2,4-dinitrophenyl hydrazone
A few drops of 2,4-DNPH hydrochloride in mildly acidified ethanol are added to a solution containing formaldehyde. A precipitate of formaldehyde 2,4-dinitrophenyl hydrazone is formed, and, recrystallized from ethanol (mp 166–167 °C).

6.1.2.4 Methylene Dinaphthol A few drops of a formaldehyde-containing solution are treated with 33% alcohol, β-naphthol and a little hydrochloride. A precipitate of methylene dinaphthol is formed. On gently boiling small white needles are formed, which turn brown at 180 °C, and melt with decomposition to a brownish-red liquid at 189–192 °C.

6.2 Quantitative Analysis

6.2.1 Titrimetric Methods

Titrimetric methods, because of their simplicity and accuracy, are widely employed for the determination of formaldehyde concentration when present as a major component. However, there is a limit for aldehyde concentration below which it is inadvisable to expect great accuracy. An assessment of these methods has been carried out by Buchi,[43] and by Reynolds and Irwin.[35] Reynolds and Irwin[35] reviewed the different titrimetric methods available and concluded that for measuring small quantities of pure formaldehyde, the iodimetric method is most suitable. The potassium cyanide method is the best volumetric method for the estimation of formaldehyde concentration in the presence of acetone, acetaldehyde, benzaldehyde, higher aldehydes and ketones. The sodium sulfite method is very simple and adaptable to field tests. The hydroxylamine hydrochloride method has the disadvantage of being insufficiently sensitive for low concentrations and is rather slow, requiring several hours for completion. For colored solutions the pH titration method has a distinct advantage, but it is no better than ordinary titration in the case of colorless solutions.

6.2.1.1 Sodium Sulfite Method

Principle. When formaldehyde reacts with sodium sulfite, it forms the formaldehyde bisulfite addition product, and sodium hydroxide is liberated quantitatively according to Scheme 4.

$$CH_2O + Na_2SO_3 + H_2O \longrightarrow NaOH + CH_2(NaSO_3)OH$$

Scheme 4

The amount of base produced is determined titrimetrically.

Reagents. The following reagents are used: (a) Na_2SO_3 (125 g of anhydrous sodium sulfite or 250 g of hydrated sodium sulfite is dissolved in water and diluted to 1 L); (b) 0.25 N H_2SO_4.

Procedure. To 100 mL sodium sulfite solution in a Erlenmeyer flask, *W* g (about 2 g) of the sample is added and the mixture is swirled to effect complete reaction. Then 3–5 drops of 1% ethanolic thymolphthalein indicator solution are added and the solution is titrated with 0.25 N H_2SO_4 to the colorless end-point. A blank determination is also carried out.

Calculation.

$$\text{HCHO (wt\%)} = \frac{(A - B)(N)(MW)(100)}{(W)(1000)}$$

where A = volume of sulfuric acid solution required for sample titration (mL), B = volume of sulfuric acid solution required for blank titration (mL), N = normality of the sulfuric acid solution, MW = molecular weight of formaldehyde (30.03 g) and W = sample weight (g).

6.2.1.2 Alkaline Peroxide Method

Principle. Formaldehyde is oxidized with H_2O_2 in the presence of excess base according to Scheme 5.

$$2HCHO + 2\,NaOH + H_2O_2 \longrightarrow 2\,HCOONa + 2H_2O + H_2$$

Scheme 5

Excess NaOH is determined by titration with standard acid.

Interferences. Acetone (less than 7 wt%) and acetaldehyde (less than 0.5 wt%) do not interfere.[43]

Reagents. The following reagents are used: (a) 1 N NaOH; (b) neutral H_2O_2 (10 mL of 30% H_2O_2 is neutralized with 1 N NaOH using bromothymol blue as an indicator and diluted to 100 mL); (c) 1 N H_2SO_4.

Procedure. Water (10 mL) is placed in a stoppered 250 mL Erlenmeyer flask and weighed, then 3 mL of sample is added and the flask is weighed again. To this 50 mL of 1 N NaOH solution is added and then immediately but slowly 50 mL of neutral H_2O_2 solution is added. The mixture is heated on a water bath for 15 min, while shaking cautiously with a general rotatory motion. It is cooled to room temperature with cold water. The excess NaOH is titrated with 1 N H_2SO_4 using 2–3 drops of bromothymol blue as indicator to a bluish-green endpoint. A blank determination is also made.

Calculation.

$$\text{HCHO (wt\%)} = \frac{(B - A)(N)(MW)(100)}{(W)(1000)}$$

where A = volume of sulfuric acid solution required for sample titration (mL), B = volume of sulfuric acid solution required for blank titration (mL), N = normality of the sulfuric acid solution, MW = molecular weight of formaldehyde (30.03 g), and W = sample weight (g).

6.2.1.3 Iodimetric Method

Principle. The reaction of formaldehyde with hypoiodite reagent serves as a method for determination, according to Scheme 6.

$$HCHO + NaOI + NaOH \longrightarrow HCOONa + NaI + H_2O$$

Scheme 6

In acidic media iodide self-oxidizes into iodine, as shown in Scheme 7.

$$4I^- + O_2 + 4H^+ \longrightarrow 2I_2 + H_2O$$

Scheme 7

Many other aldehydes and ketones also react with the iodine reagent and the method described is basically useful only with known systems. The recovery of formaldehyde is approximately 98.5%.

Interferences. The interference of methanol, a known constituent in commercial formaldehyde solutions, has been disputed. This iodimetric method is particularly useful for the determination of small concentrations of formaldehyde in aqueous solution.

Reagents. The following reagents are used: (a) 3 N NaOH; (b) 0.2 N NaOI; (c) 0.1 N $Na_2S_2O_3$; (d) H_2SO_4.

Procedure. A sample containing up to 0.16 g of formaldehyde is weighed into an iodine flask and diluted with water to about 100 mL. Then 30 mL of 3 N sodium hydroxide solution and 75 mL of 0.2 N sodium hypoiodite are added and the flask is stoppered. The mixture is allowed to stand at room temperature for 30 min. Approximately 95 mL of H_2SO_4 is added and the liberated iodine is titrated immediately with 0.1 N sodium thiosulfate solution using starch indicator. At least one blank determination is carried out with each set of samples.

Calculation.

$$\text{HCHO (wt\%)} = \frac{(B - A)(N)(MW)(100)}{(W)(1000)}$$

where A = volume of sodium thiosulfate solution required for sample titration (mL), B = volume of sodium thiosulfate solution required for blank titration (mL), N = normality of the sodium thiosulfate solution, MW = molecular weight of formaldehyde (30.03 g), and W = sample weight (g).

6.2.1.4 Potassium Cyanide Method

Principle. Formaldehyde reacts with potassium cyanide, unreacted cyanide is converted into silver cyanide by the addition of a measured amount of standard silver nitrate solution and excess silver nitrate is determined by titration with thiocyanate as shown in Schemes 8 and 9.

$$HCHO + KCN \longrightarrow CH_2(OK)CN$$

Scheme 8

$$KCN + AgNO_3 \longrightarrow AgCN + KNO_3$$

Scheme 9

Reagents. The following reagents are used: (a) 0.1 N KCN; (b) 0.1 N $AgNO_3$; (c) 0.1 N KCNS.

Procedure. A sample containing approximately 0.15–0.25 g of dissolved formaldehyde is weighed and added to a 200 mL glass-stoppered Erlenmeyer flask containing 100 mL of 0.1 N KCN. After shaking, the mixture is washed into a volumetric flask with 40 mL of 0.1 N $AgNO_3$ and 10 mL of 1 : 9 nitric acid. The volume is made up to 200 mL. The solution is filtered and 100 mL of the filtrate is titrated with 0.1 N KCNS solution using ferric alum as an indicator.

Calculation.

$$\text{HCHO (wt\%)} = \frac{[100 - (40 - 2A)](N)(MW)(100)}{(W)(1000)}$$

where A = volume of thiocyanate solution required for titration (mL), N = normality of the thiocyanate solution, MW = molecular weight of formaldehyde (30.03 g) and W = sample weight (g).

6.2.1.5 Ammonium Chloride Method

Principle. The reaction is shown in Scheme 10.

$$6HCHO + 4NH_3 \longrightarrow (CH_2)_6N_4 + 6H_2O$$

Scheme 10

This reaction is the basis for a quantitative determination, with the excess ammonia being determined by steam distillation or by titration. Buchi[43] modified the technique for the use of neutral NH_4Cl.

Interferences. Ethanol and acetone have no significant effect on the analysis, but acetaldehyde in concentrations higher than 5% leads to high values.

Reagents. The following reagents are used: (a) 1 N NaOH; (b) neutral NH_4Cl; (c) 1 N HCl.

Procedure. A 6 g sample of approximately 37% formaldehyde solution is weighed into a 100 mL flask and diluted to the volume of the flask. Then 25 mL of this solution is added to an Erlenmeyer flask together with 3 drops of 0.1% bromothymol blue indicator and neutralized with NaOH solution until the indicator color changes to blue. Neutral NH_4Cl (1.5 g) and NaOH solution (25 mL) are added rapidly, the flask is closed tightly using silicon grease to ensure a perfect seal and the reaction mixture is allowed to stand for 1.5 h. The excess NH_3 is titrated with HCl until the color changes to green. A blank determination is also carried out.

Calculation.

$$\text{HCHO (wt\%)} = \frac{(B - A)(N)(180.2)(100)}{W(1000)}$$

where A = volume of hydrochloric acid required for sample titration (mL), B = volume of hydrochloric acid required for blank titration (mL), N = normality of hydrochloric acid and W = sample weight (g).

6.2.1.6 Mercurimetric Method

Principle. The reaction is shown in Scheme 11.

$$K_2HgI_4 + HCHO + 3NaOH \longrightarrow Hg + HCOONa + 2KI + 2NaI + 2H_2O$$

Scheme 11

The precipitated Hg is determined iodimetrically.

Interferences. Organic hydroperoxide and acetaldehyde in concentrations as low as 0.5% can cause interference. Acetone and methyl and ethyl alcohols have no appreciable effect.

Reagents. The following reagents are used: (a) mercuric chloride; (b) KI solution; (c) 2 N NaOH; (d) 2 N acetic acid; (e) 0.1 N iodine solution; (f) 0.1 N sodium thiosulfate.

Procedure. A 1.75 g sample of 37% formaldehyde solution or its equivalent is weighed into a 500-mL volumetric flask and diluted to the volume of the flask. A 25-mL portion of this solution is added to a 200-mL glass-stoppered Erlenmeyer flask, containing a solution of 1 g of mercuric chloride and 2.5 mL of KI in 35 mL of water. Then 20 mL of 2 N NaOH is added. The mixture is shaken vigorously for 5 min then acidified with 25 mL of 2 N acetic acid. Immediately 25 mL of 0.1 N iodine solution is added and the mixture is agitated until the precipitate is completely dissolved. The excess iodine is back-titrated with 0.1 N sodium thiosulfate solution using starch solution as the indicator. A blank determination is also carried out.

Calculation.

$$\text{HCHO (wt\%)} = \frac{(B - A)(N)(MW)(100)}{W(2)(1000)}$$

where B = volume of sodium thiosulfate solution required for blank titration (mL), A = volume of sodium thiosulfate solution required for sample titration (mL), N = normality of the sodium thiosulfate solution,

MW = molecular weight of formaldehyde (30.03 g) and W = sample weight (g) in 25 mL.

6.2.1.7 Hydroxylamine Method The hydroxylamine method is widely used for the determination of carbonyl compounds in the presence of alkaline-sensitive products and impurities, such as phenol, cresol and resinous materials. Formaldehyde reacts with hydroxylamine hydrochloride forming an oxime and liberating hydrochloric acid as shown in Scheme 12. Using a suitable acid–base indicator (or preferably by pH titration when the solutions are colored) the reaction can serve as a basis for quantitative estimation.

$$HCHO + H_2NOH{\cdot}HCl \longrightarrow HCN{=}NOH + HCl + H_2O$$

Scheme 12

Reagents. The following reagents are used: (a) hydroxylamine hydrochloride; (b) 0.1 N sodium hydroxide.

Procedure. To 10 mL 1 N hydroxylamine hydrochloride solution in an Erlenmeyer flask, W g (about 2 g) of sample is added and agitated. After agitation, the mixed solutions are neutralized with 0.1 N sodium hydroxide solution using methyl orange as indicator. It is essential to match the end-point (which is red to orange and not yellow) against a blank. The method is reported as being sensitive to 0.1 mg of formaldehyde. A blank determination is also carried out.

Calculation.

$$\text{HCHO (wt\%)} = \frac{(A - B)(N)(MW)(100)}{W(1000)}$$

where A = volume of sodium hydroxide solution required for sample titration (mL), B = volume of sodium hydroxide solution required for blank titration (mL), N = normality of sodium hydroxide solution, MW = molecular weight of formaldehyde (30.03 g) and W = sample weight (g).

6.2.2 Colorimetric Methods

Absorption spectrophotometry forms the basis for many of the aldehyde monitoring techniques. The chromotropic acid method is used widely and is approved protocol for the measurement of formaldehyde.[44,45] The method is specific for formaldehyde but is not very sensitive. Another commonly used colorimetric method utilizes pararosaniline as the chromogen since it gives an enhancement in the sensitivity by approximately a factor of 2 over the chromotropic acid method. However, the technique is prone to interference from carbonyl compounds and other low-molecular-weight aldehydes. The modified pararosaniline method[18] is not affected by phenol : formaldehyde ratios ranging from 1 to 100, whereas such ratios cause total inhibition of the chromotropic acid method. Miksch et al.[18] concluded that the modified pararosaniline method was more sensitive, more reproducible and easier to use than the chromotropic acid method. Another important method uses MBTH for formaldehyde measurement. This method's sensitivity has been reported as 30 ppb formaldehyde based on a sampled air volume of 25 L in 35 mL of absorbing solution. Several other colorimetric methods for formaldehyde have also been reported. These include the 7-amino-5-hydroxy-2-naphthalene sulfamic acid and the *p*-phenylenediamine methods.

To perform large numbers of analyses or for semicontinuous analysis of water or air, colorimetric analyses can be adapted for use with autoanalyzers. Autoanalyzers are usually implemented by one of two methods: either continuous or discrete (batch). In a continuous flow analysis, samples are analyzed from a flowing stream with any necessary operations (such as filtering and reagent addition) performed prior to the measurement of chemical and physical properties. Actual determinations are made using flow-through sample cells. In discrete analysis, samples are placed in individual containers for the duration of the analysis. Preliminary operations of dilution, reagent addition and mixing are performed on each sample at different locations within the analyzer. Each treated sample is then presented in sequence to the sensing device. A batch of samples is usually preloaded for processing by this technique. While both technique are used in process control and laboratory analyses, continuous analysis is more often utilized in automated process control applications because of its faster response time.[46]

All the colorimetric methods covered here relate to formaldehyde analysis in air and require specific sampling methods, although these methods could also be used directly for estimation of formaldehyde in solutions.

6.2.2.1 Chromotropic Acid Method

Principle. Formaldehyde reacts with chromotropic acid (1,8-dihydroxynaphthalene 3,6-disulfonic acid) in concentrated sulfuric acid solution to form a purple colored dye, which is determined spectrophotometrically at 580 nm. The formaldehyde may be collected either in water, in aqueous bisulfite or directly in the chromotropic acid–sulfuric acid solution. By collecting in water 0.1–2.0 $\mu g\,mL^{-1}$ formaldehyde can be measured. This corresponds to a minimum detectable quantity of 0.1 ppm in a 40-L air sample collected in 20 mL water. The sensitivity can be increased 10-fold by sampling directly into the chromotropic acid–sulfuric acid solution. The reproducibility of the method is within ±5%.[47]

Interferences. The chromotropic acid method has very little interference from other aldehydes (<0.01% from saturated aldehydes and a few percent from acrolein). Alcohols, phenols, aromatic hydrocarbons and olefins all show negative interference but the concentrations of these classes of compounds in ambient air are too low to lead to serious errors.

Apparatus. Two coarse-fritted bubblers in series are used to collect the samples. A pump capable of maintaining a metered flow of 1 L min^{-1} through the sampling train for up to 24 h is required.

Reagents. The following reagents are used: (a) chromotropic acid reagent (0.10 g of chromotropic acid disodium salt is dissolved in water and diluted to 10 mL, filtered if necessary and stored in the dark. Fresh solution is made weekly); (b) concentrated H_2SO_4 (specific gravity = 1.86); (c) stock formaldehyde solution (1 mg mL^{-1}) (2.7 mL of 37% formalin solution is diluted to 1 L with distilled water. This solution is standardized by the addition of excess sodium bisulfite to an aliquot of the solution, followed by iodimetric titration of the formaldehyde bisulfite addition product after the excess bisulfite has been oxidized by I_2. Alternatively, sodium formaldehyde bisulfite can be used as a primary standard, in which case 4.470 g of formaldehyde bisulfite is dissolved in distilled water and diluted to 1 L. This solution is stable for at least three months); (d) dilute standard formaldehyde solution (10 μg mL^{-1}) (1 mL of standard stock solution is diluted to 100 mL with distilled water. Fresh solution is prepared daily).

Procedure. Distilled water (20 mL) is placed in each bubbler, and the sampling train is connected and aspirated at 1 L min^{-1} for a suitable time to collect enough formaldehyde for analysis (24 h). The collection efficiency in a single bubbler is approximately 80%, i.e. 95% of the formaldehyde is trapped using two bubblers. After sampling, the volume is adjusted to 20 mL and a 4-mL aliquot of the sample is pipetted to the stoppered test tube. A blank is also prepared using 4 mL distilled water. A 0.1 mL portion of 1% chromotropic acid solution is added and mixed, then 6 mL concentrated sulfuric acid is cautiously added from a pipette. The solution is allowed to cool and the absorbance is read at 580 nm in a 1-cm cell. The amount of formaldehyde present in each aliquot taken is determined from the calibration curve obtained as described next.

Calibration. A freshly prepared standard formaldehyde solution (10 μg formaldehyde mL^{-1}) is used to prepare the calibration solution. Different amounts (0, 0.1, 0.3, 0.5, 0.7, 1.0 and 2.0 mL) of this solution are pipetted into a series of glass-stoppered graduated tubes. Each sample is diluted to 4 mL with distilled water and the color is developed as described above. The absorbance is plotted against formaldehyde concentration (μg mL^{-1}). The air concentration of formaldehyde is calculated as follows:

$$\text{ppm (volume)} = \frac{\text{total HCHO}(\mu g) \times 24.47}{VM}$$

where total formaldehyde is the sum of the amounts collected in each bubbler, V is the volume of air sampled (at 101.325 kPa and 25 °C) and M is the molecular weight of formaldehyde (30.03 g).

6.2.2.2 3-Methyl-2-benzothiazole Hydrazone Method

Principle. The formaldehyde in ambient air is collected in a 0.05% aqueous solution of 3-methyl 2-benzothiozolone hydrazone hydrochloride. The resulting azine is then oxidized by ferric chloride/sulfamic acid solution to form a blue cationic dye in acid solution which can be measured at 628 nm after 12 min (Figure 2). The concentration of total aldehydes is calculated in terms of formaldehyde, and formaldehyde in the range 0.03–0.7 μg mL^{-1} can be measured. This corresponds to a minimum detectable concentration of 0.03 ppm in a 25-L air sample. The reproducibility of the method is within ±5%.[47]

Apparatus. The samples are collected in all-glass bubblers with a coarse fritted inlet. It is normally necessary to sample ambient air for quite long periods to obtain sufficient aldehyde for analysis and a pump capable of drawing at least 0.5 L min^{-1} for 24 h is required. A critical orifice, a rotameter or a gas meter can be used to meter the flow.

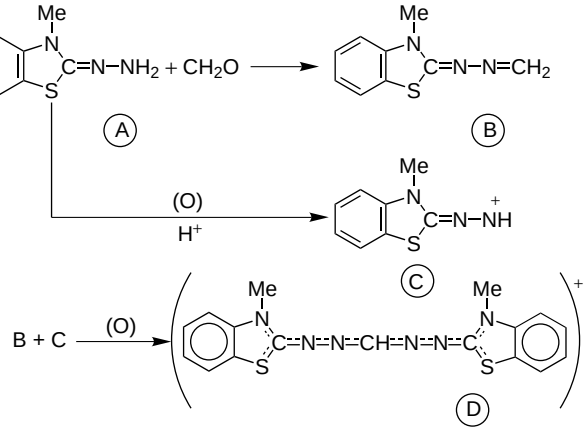

Figure 2 Sequence of reactions involved in the MBTH method. A = 3-methyl-2-benzothiazolone hydrazone, B = azine, C = reactive cation, D = blue cationic dye. (Reprinted with permission from Sawicki et al.[26] Copyright 1961 American Chemical Society.)

Reagents. The following reagents are used: (a) MBTH (0.05%) (0.5 g of MBTH is dissolved in distilled water and diluted to 1 L. The reagent becomes turbid either on storage or during sampling. If this occurs, it is filtered by gravity. The solution is stable for a week or more if stored in the cold in a dark bottle); (b) oxidizing reagent (1.6 g sulfamic acid and 1.0 g ferric chloride is dissolved in distilled water and diluted to 1 L); (c) stock formaldehyde solutions (1.0 mg mL^{-1}, prepared as in chromotropic acid method); (d) dilute standard formaldehyde solution (10 μg mL^{-1}, prepared as in chromotropic acid method).

Procedure. A measured volume of ambient air at a rate of 0.5 L min^{-1} is sampled through 35 mL of MBTH solution in the absorber until sufficient aldehyde for analysis has been collected. An average collection efficiency of 84% has been determined for sampling under these conditions although higher efficiencies have been reported for slightly modified conditions. The efficiency is between 90 and 95% using 10 mL of 0.2% aqueous MBTH in a bubbler aspirated at 1 L min^{-1}.

After sampling the volume of absorbing reagent solution is made up to exactly 35 mL with distilled water (to compensate for evaporation losses) and allowed to stand for 1 h. Then 10 mL of the sample solution is pipetted into a glass-stoppered tube and an equal volume of reagent is pipetted into a second tube to serve as a blank. A 2 mL portion of oxidizing solution is added to each tube and mixed well. After allowing to stand for at least 12 min the absorbance of the sample is read at 628 nm against the reagent blank in 1-cm cells. The aldehyde content (expressed as μg formaldehyde mL^{-1}) is determined from the calibration plot prepared as described.

Calibration. Freshly prepared dilute standard formaldehyde solution is used to calibrate this method. Different amounts (0, 0.5, 1.0, 3.0, 5.0 and 7.0 mL) of this solution are pipetted into a series of 100 mL volumetric flasks and diluted to the volume of the flask with 0.05% MBTH solution. After allowing to stand for 1 h, 10 mL aliquots of each of these solutions are transferred to stoppered test tubes. Oxidizing solution (2 mL) is added and mixed well. After 12 min the absorbance of each solution is determined at 628 nm against the blank in 1-cm cells. Absorbance is plotted against μg formaldehyde per mL of solution. The air concentration of total aliphatic aldehydes (as formaldehyde) is given by:

$$\text{ppm (volume)} = \frac{\mu\text{g mL}^{-1}\text{HCHO} \times 35 \times 24.45}{VME}$$

where V = volume of air sampled (at 101.325 kPa and 25 °C), M is molecular weight of formaldehyde (30.03 g) and E is the collection efficiency of the bubbler which can be estimated by using two bubblers in series. If a coarse-fritted bubbler is used, E may be taken to be 0.84%.

6.2.2.3 Nash Method

Principle. Formaldehyde reacts with ammonium acetate, acetic acid and acetylacetone (Nash reagent) and forms the blue compound diacetyldihydrolutidine.[48] The detection limit and precision of the method are 0.05 μg mL^{-1} and 1.43%.[49]

Apparatus. A mist chamber, filter holder, pump and rotameter/gas meter are required.

Reagents. The following reagents are used: (a) Nash reagent (75 g of ammonium acetate, 1.5 mL acetic acid, and 1.0 mL acetylacetone are placed in a volumetric flask and made up to 500 mL with distilled water); (b) stock formaldehyde solution (as described for chromotropic acid method); (c) standard formaldehyde solution (as described for chromotropic acid method).

Procedure. Formaldehyde is sampled using the mist chamber technique. The air is first passed through a 1.2-μm pore size poly(tetrafluoroethylene) (PTFE) filter to remove particles and then exposed to a fine spray of deionized water (25 mL) with a flow rate of 12 L min^{-1}.[49,50] After sampling, the volume of absorbing reagent solution is made up to exactly 25 mL with distilled water (to compensate for evaporation losses). Then 5 mL of Nash reagent and 5 mL of sample solution are placed in a test tube. After allowing to stand for at least 10 min, the absorbance of the sample is determined at 412 nm against the reagent blank (water) in a 1-cm cell. Formaldehyde can be determined from the calibration plot prepared as described next.

Calibration. A freshly prepared, dilute standard formaldehyde solution containing 10 μg formaldehyde mL^{-1} is used to prepare calibration solutions. Different amounts (0, 2, 4, 6 and 8 mL) of this solution are added into a series of 10-mL volumetric flasks and diluted to the volume of the flask with water. A 5-mL aliquot of each solution is transferred to a stoppered test tube and 5 mL of Nash reagent is added and mixed well. After 10 min the absorbance of each solution is determined at 412 nm against the blank in 1-cm cells. The absorbance is plotted against μg formaldehyde mL^{-1} of solution.

6.2.2.4 Modified Pararosaniline Method

Principle. Acid-bleached pararosaniline and formaldehyde lead to the formation of intensely colored pararosaniline methyl sulfonic acid which is read at

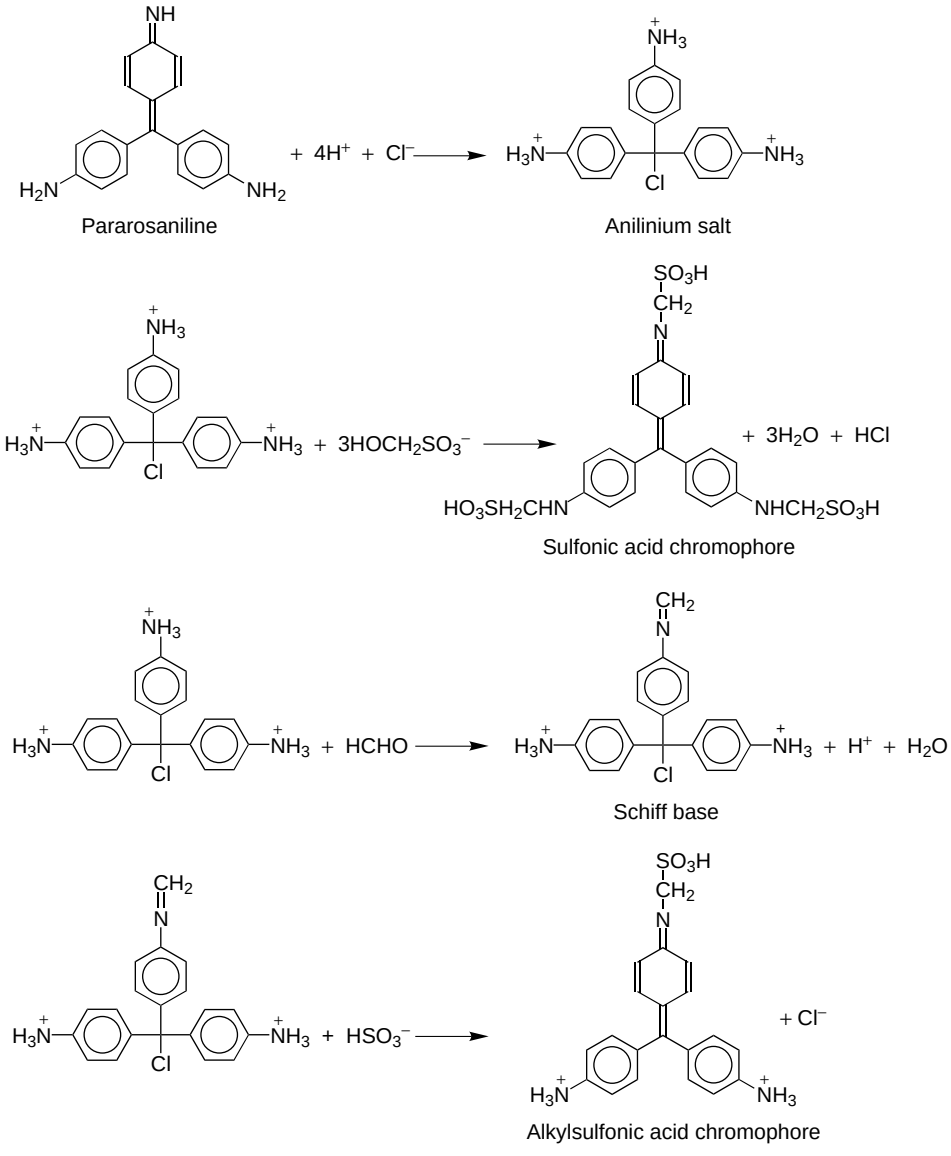

Figure 3 Reaction pathway for pararosaniline method. (Reprinted with permission from Miksch et al.[18] Copyright 1981 American Chemical Society.)

570 nm (Figure 3). The reproducibility of the slope is $0.52 \pm 0.37\ mL^{-1}$.

Only low-molecular-weight aldehydes give positive interferences, and only when present in a large excess compared with formaldehyde. Negative interferences could be caused by compounds that react with either pararosaniline or formaldehyde. Cyanide, sulfite and hydroxylamine all form stable, well-characterized adducts with formaldehyde.[18]

Reagents. The reagents used are formaldehyde stock solution, pararosaniline solution and sodium sulfite. Details are given below. Methanol-free formaldehyde solutions are prepared by refluxing approximately 20 g of paraformaldehyde in water, filtering the solution and diluting to 1 L. The solution is standardized by adding the formaldehyde solution to a 1 M solution of sodium sulfite to quantitatively produce sodium formaldehyde bisulfite and hydroxyl ions. It is back-titrated to its original pH using 1.0 M HCl to determine the amount of formaldehyde added. The stock solution is stable for several months.

Pararosaniline (0.16 g) is dissolved in 20 mL of concentrated HCl and the mixture is diluted to 100 mL with deionized water. This reagent remains stable for several months when stored in an amber bottle.

Sodium sulfite (8 mM) is prepared by dissolving 0.10 g of anhydrous sodium sulfite in deionized water and

diluting the solution to 100 mL with deionized water. It is prepared daily.

Procedure. A 2.5-mL sample of formaldehyde solution is placed in a cuvette and 250 µL of acidified pararosaniline reagent is added. After mixing thoroughly, 250 µL of the sodium sulfite reagent is added. The cuvette is capped and placed in a water bath at 25 °C. The color is allowed to develop for 60 min and read at 570 nm. The blank and the sample are read against deionized water.

Calibration. Freshly prepared standard formaldehyde solution (0.06 mg mL^{-1}), is used to calibrate this method. Different amounts (0, 0.5, 1.0, 3.0, 5.0 and 7.0 mL) of this solution are pipetted into cuvettes and 250 µL of acidified pararosaniline reagent is added. After mixing thoroughly, 250 µL of the sodium sulfite reagent is added. The cuvettes are capped and placed in a water bath at 25 °C. The color is allowed to develop for 60 min and read at 570 nm. The blank and the sample are read against deionized water.

6.2.3 Polarographic Method

The polarographic technique has gained favor as a practical method for the determination of formaldehyde and other aldehydes.[51] The increase in popularity of the polarographic technique may be attributed to improved electrode design and automated microprocessor-based differential pulse polarography. A pulse technique that is capable of good signal-to-noise ratios employs the pulse train shown in Figure 4. The formaldehyde–hydrazone derivative could be determined in acetate buffer solution by differential pulse polarography. An aliquot of the derivatized solution is then analyzed by differential pulse polarography with a dropping mercury electrode.[52]

6.2.3.1 Principle Polarography involves the study of current–voltage relationships at a dropping mercury electrode under certain controlled conditions. It consists

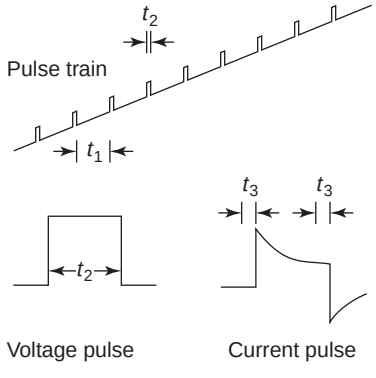

Figure 4 Pulse train, voltage and current pulses for differential pulse polarography.[52]

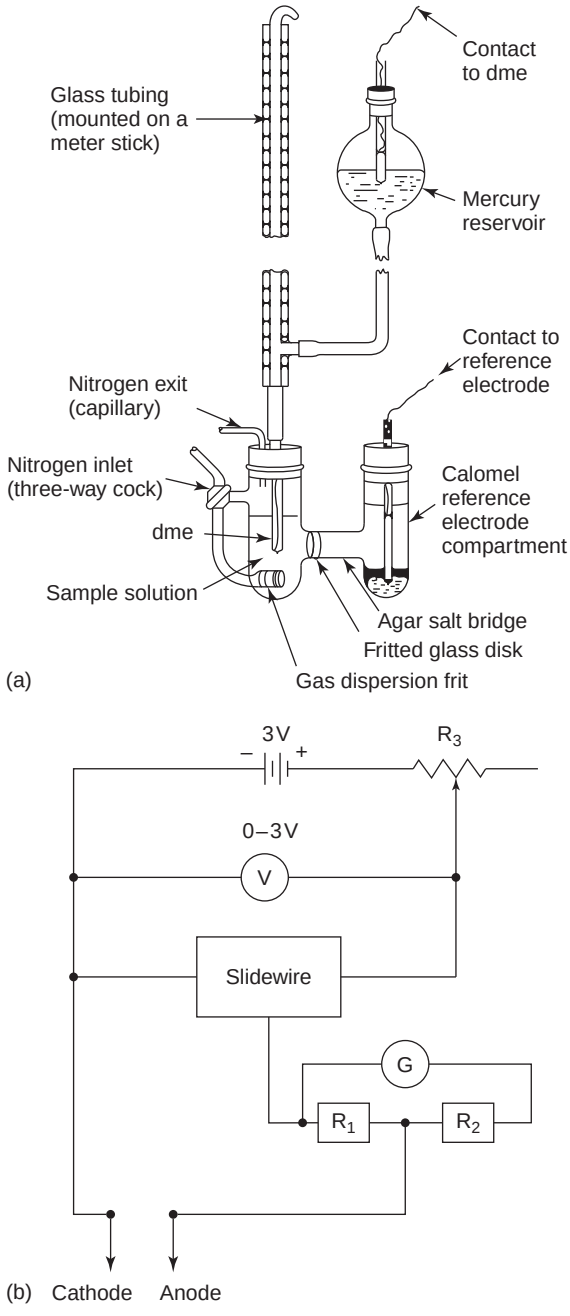

Figure 5 (a) Schematic diagram of polarographic cell; (b) circuit for obtaining polarograms. (Reproduced by permission from Willard et al.[46])

in electrolyzing a solution of an electroactive substance between a dropping mercury electrode functioning as a cathode, and some reference electrode (a pool of mercury) which acts as an anode. The area of the anode is correspondingly large so that it may be regarded as incapable of becoming polarized and the potential of such an electrode remains fairly constant. A potential is applied between these electrodes and increased in a

stepwise manner. At first only a small current flows (the so-called residual current), and this continues until the decomposition of the reducible ionic species is reached. Figure 5 shows the polarographic cell and a basic circuit for obtaining a polarogram.

A steep rise in the current is observed, which continues to rise with increasing potential until it reaches a maximum value when the concentration of ions at the electrode surface approaches zero. Now the current no longer increases with applied potential but approaches a steady limiting value. The particle reduced at the mercury surface diffuses away from the surface either by forming an amalgam with the mercury, which subsequently detaches from the capillary top or by simply diffusing back into the bulk of the solution.

Septon and Ku[36] have described a method for polarographic determination of formaldehyde in air samples. This polarographic method is free from interferences from acetaldehyde, acrolein and benzaldehyde as their potential differences are different from that of formaldehyde. Therefore, this technique can be used to monitor formaldehyde where incomplete combustion occurs since it is possible to resolve formaldehyde from other aldehydes. Septon and Ku[36] determined the collection efficiency by connecting two sampling devices in series. The efficiency was determined from the amount of vapor that passed through the first bubbler and was absorbed in the backup. The efficiency was $98.6 \pm 0.3\%$ for six samples when 10 ppm formaldehyde was collected at a flow rate of $1\,L\,min^{-1}$ for 30 min in a bubbler containing 10% methanol–water solution.

This method was validated by generating samples with concentration of $5.8–17.7\,mg\,m^{-3}$. Average recovery was 103%. The pooled coefficient of variation or relative standard was 0.08. A linear regression of formaldehyde taken vs formaldehyde found showed correlation with a slope of 1.1 and an intercept of $-1.4\,mg\,m^{-3}$. The quantitative determination limit is 50 ppb, and the relative standard deviation of replicate determinations at this level was 0.04.

6.2.3.2 Sampling Air (30 L) is sampled in 10% v/v methanol/deionized water in a bubbler connected to a sampling pump at a flow rate of $0.5–1\,L\,min^{-1}$. A prefilter of mixed cellulose acetate membrane (0.8-μm pore size and 37 mm diameter) with a 37 mm polystyrene filter holder is used. The maximum volume sampled should not exceed 160 L at $1\,L\,min^{-1}$. The bubbler is connected to the sampling pump with rubber tubing (poly(vinyl chloride) (PVC) tubing should not be used before the bubbler). If evaporation losses occur during the sampling, additional solution is added to the bubbler to maintain the required concentration of methanol solution. With each batch of ten samples at least one blank is subjected to the same handling except that no air is drawn through it. The bubbler solution is transferred to a 20-mL glass vial.

6.2.3.3 Reagents Details of reagents to be used are as follows. Methanol solution (10% v/v) is prepared by diluting 100 mL of reagent grade methanol to 1 L with deionized water. The acetate buffer (pH 4) is an equimolar mixture of acetic acid and sodium acetate (0.1 M) in deionized water. For the hydrazine reagent (2% w/v) 2 g of hydrazine sulfate ($H_2NNH_2 \cdot H_2SO_4$) is dissolved in deionized water and diluted to 100 mL with deionized water. It is prepared fresh every week and kept tightly closed. The supporting electrolyte is prepared by mixing 10% methanol solution, acetate buffer and 2% hydrazine reagent in a ratio of 5 : 4 : 1, respectively. A 100-mL portion of this solution is required for each analysis and is prepared fresh daily. The formaldehyde stock solution (1000 ppm) is prepared by dissolving 2.7 g (about 3 mL) of 37% formaldehyde solution in 1 L of deionized water. The solution is stable for 3 months. Appropriate serial dilutions of the formaldehyde stock solution are made with 10% methanol to prepare 100 ppm, 10 ppm and 1 ppm formaldehyde solutions.

Reagents used for standardization of formaldehyde stock solutions are: sodium carbonate (certified 99.9% minimum purity, dried at 120 °C for 2 h, then transferred to a desiccator, dried to a constant weight and used as a primary standard); concentrated (98%) sulfuric acid; 0.1 N sulfuric acid (standardized against sodium carbonate solution); and 12.5% (w/v) sodium sulfite.

Standardization of formaldehyde stock solutions involves weighing 50 ± 0.5 g of formaldehyde solution and transferring it to a 250-mL beaker, then adding 25 mL of 12.5% w/v sodium sulfite (whose pH is adjusted to 9.6 with 0.1 N H_2SO_4) and titrating to pH 9.6 with 0.1 N H_2SO_4. The concentration of formaldehyde (in ppm) is calculated as follows:

$$\text{Formaldehyde (ppm)} = \frac{(A - B)(C)(D)}{E}$$

where A = sulfuric acid solution required to titrate the sample (mL), B = sulfuric acid solution required to titrate the blank (mL), C = normality of the sulfuric acid solution (mequiv mL^{-1}), D = (30 mg mequiv^{-1} of HCHO) × (1000 μg mg^{-1}) or 30×10^3 μg mequiv^{-1} of formaldehyde, and E = amount of sample used (g).

6.2.3.4 Sample Preparation The total volume of each sample is measured and recorded and the volume is made up with 10% methanol. Then 5 mL of the sample solution, 4 mL of acetate buffer and 1 mL of 2% hydrazine reagent are added to the polarographic cell.

6.2.3.5 Analysis All glassware and polarographic cells are soaked in 6 M nitric acid, rinsed thoroughly

with deionized water and air-dried prior to use. The polarographic analyzer and the automated cell sequencer are warmed up for at least 30 min. Each standard and sample are purged for 5 min with pre-purified nitrogen. A calibration curve is made by analyzing standards and blanks (supporting electrolyte). The standard is reanalyzed after every five or six samples. The peak current (nA) for each standard and sample is recorded. The differential pulse polarogram of formaldehyde hydrazone at a dropping mercury electrode yields a peak at approximately −0.850 V. A least squares regression program is used to plot a calibration curve of peak current (nA) vs concentration (total μg) of standards.

6.2.3.6 Calculations The weight in μg corresponding to each response (nA) is recorded from the standard curve. No volume corrections are needed if the standard curve is based upon μg per 10 mL formaldehyde and the volume of the sample determined is identical to the volume of the standards determined.

Corrections for the blanks must be made for each sample as follows:

$$\mu g\ mL^{-1} HCHO\ (\text{blank corrected}) = \frac{\mu g\ HCHO}{\text{sample aliquot (mL)}} - \frac{\mu g\ \text{blank}}{\text{blank aliquot (mL)}}$$

The concentration of formaldehyde in the air sampled can be expressed in $mg\,m^{-3}$, which is numerically equivalent to $\mu g\,L^{-1}$.

$$mg\ m^{-3} = \frac{(\text{corrected } \mu g\ mL^{-1})(\text{sample volume})}{(\text{air volume sampled, liters})}$$

Another method of expressing concentration is in ppm (corrected to standard conditions of 25 °C and 101.325 kPa):

$$\text{ppm} = mg\ m^{-3} \times \frac{24.45}{MW}$$

where 24.45 = molar volume (L mol^{-1}) of analyte, at 25 °C and 101.325 kPa, and MW = molecular weight (g mol^{-1}) of analyte.

6.2.4 Chromatographic Methods

6.2.4.1 Direct Gas Chromatographic Method

Principle. A direct gas chromatographic procedure is suitable for the quantitative determination of formaldehyde. In GC a mobile phase (a carrier gas) and a stationary phase (column packing) are used to separate individual compounds. The carrier gas is nitrogen or helium. The stationary phase is a liquid that has been coated on an inert granular solid, called the column packing, that is held in borosilicate glass tubing. The column is installed in an oven with the inlet attached to a heated injector block and the outlet attached to a detector. Precise and constant temperature control of the injector block, oven and detector is maintained. Stationary phase material and concentration, column length and diameter, oven temperature, carrier gas flow and detector type are the controlled variables. The sample solution is injected through a tetrafluoroethylene-coated silicone rubber septum onto the column with a microsyringe. The samples are vaporized and moved through the column at different rates, depending on differences in partition coefficients between the mobile and stationary phases. As each component passes through the detector a quantitatively proportional change in electrical signal is measured on a strip-chart recorder. The retention time is indicative of the particular component and the peak height/peak area is proportional to its concentration. The detector system may be selected on the basis of specificity and sensitivity needed.[53]

Apparatus. Numerous instruments varying in sophistication and price are available for GC. The basic components are shown in Figure 6. The different parts of the instrument are as follows:

- Carrier gas supply. Carrier gases, which must be chemically inert, include He, Ar, N_2 and H_2. He is the most widely used. Flow rates are controlled by a pressure regulator. Inlet pressure usually ranges from 69 to 345 kPa which leads to flow rates of 25–50 mL min^{-1}.
- Sample injection system. Column efficiency requires that the sample be of a suitable size and be introduced as a ''plug'' of vapor. A microsyringe is used to inject liquid samples through a rubber or silicone diaphragm into a heated sample port located at the head of the column. For ordinary analytical columns sample sizes vary from 0.1 to 10 μL. Capillary columns require much smaller samples ($\sim 10^{-3}$ μL).
- Columns. Two types of columns are employed. One type is fabricated from capillary tubing (0.3–0.5 mm i.d.), the bore of which is coated with a very thin film (1 μm) of the liquid phase, and is 10–100 m long. Packed columns are fabricated from glass or metal tubes (1–8 mm i.d.). They are designed to hold solid packing and range from 2 to 20 m in length.
- Solid support for packed columns. Most common supports are derived from diatomaceous earths such as firebrick and kieselguhr. Firebrick, which is sold under trade names such as Chromosorb P™, C22 and Sterchamol, has the better strength and larger specific surface area ($\sim 4\ m^2\ g^{-1}$). Its disadvantage lies in the fact that it is more active and hence cannot

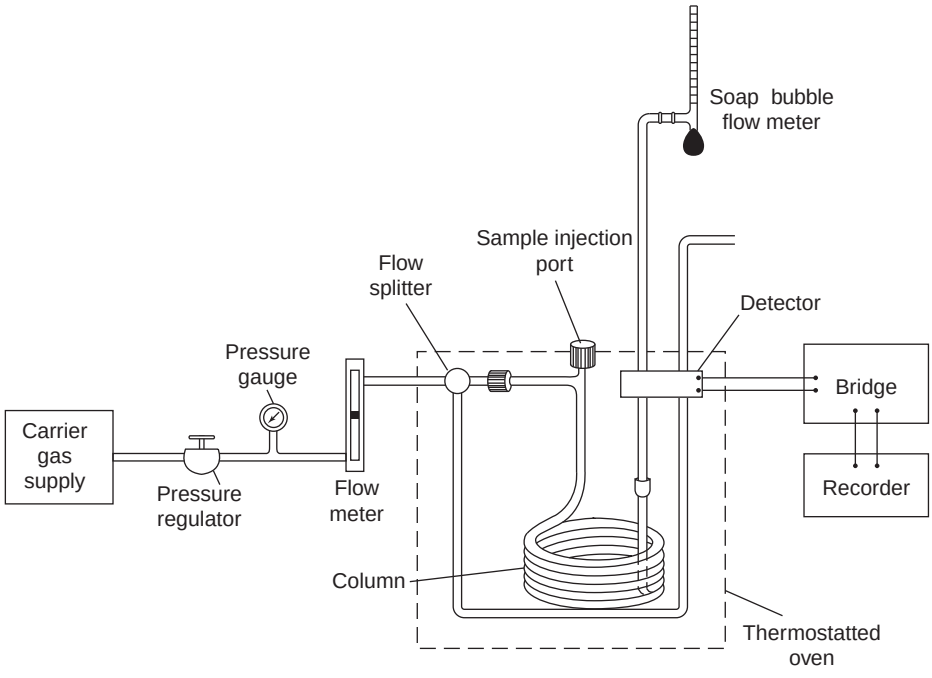

Figure 6 Schematic diagram of GC system.[53]

be employed with polar compounds. Kieselguhr is more fragile than firebrick and has a smaller surface area ($\sim 1\,m^2\,g^{-1}$) but is less reactive. It is sold under trade names as Chromosorb W™, Celite™, Embacel™, Celatom™, Fluoropak 80™, Teflon 6™, Columpak T™, etc.

- Liquid phase. The different compounds used for analysis are Ethofat 60/25, a polyoxyethylene monostearate containing an average of 15 ethylene oxide units *MW* 938). Other substances investigated are sorbitol, diethylene glycol, mannitol hexaacetate, sucrose octaacetate, Carbowax 6000, Apiezon N and Lac 296 (polyethylene glycol succinate). All substrates are loaded to 10 wt% on the fluorocarbon supports and to 20 wt% on chromosorb supports.

Detectors.

- Thermal conductivity detection (TCD). TCD is based upon changes in the thermal conductivity of the gas stream, measured using a device called a katharometer. TCD detectors are simple, rugged, inexpensive, nonselective, accurate and nondestructive of the sample.
- FID. When compounds are pyrolyzed at the temperature of a H_2/air flame they produce ionic intermediates that provide a mechanism by which electricity can be carried through the flame. The electrical resistance of a flame is very high ($10^{12}\ \Omega$) and the resulting currents are, therefore, minuscule, so an electrometer must be employed for their measurement.
- Electron capture detection (ECD). With electron capture detectors the effluent from the column is passed over a β-emitter such as Ni-63 or tritium (adsorbed on platinum or titanium foil). An electron from the emitter causes ionization of the carrier gas (N_2) and production of a burst of electrons. In the absence of organic species a constant current between a pair of electrodes results from this ionization process. The current decreases, however, in the presence of organic molecules that tend to capture electrons. The response is nonlinear unless the potential across the detector is pulsed. ECD devices are highly sensitive and possess the advantage of not altering the sample significantly.

Reagents. A test solution is prepared from paraformaldehyde, formalin or formalin plus butanol.[37] A formaldehyde standard solution containing 23.83% formaldehyde is prepared by boiling paraformaldehyde in distilled water. The solution is filtered and its formaldehyde content is determined by the standard peroxide method. This solution is used to determine the sensitivity for formaldehyde.

Formalin solution used for the calibration is prepared by diluting previously analyzed formalin with distilled water. A standard reference solution is prepared containing 15 wt% methanol, 47.4 wt% water and 37.6 wt% butanol for routine use as reference.

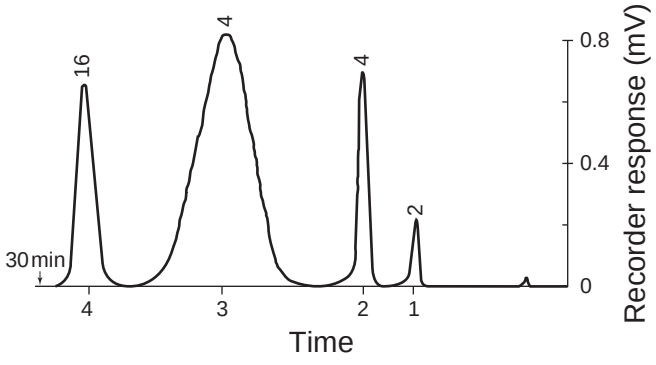

Figure 7 Gas chromatograms of the test and reference solutions. Conditions: column, 5 meter Ethofat 60/25, 10 wt% on columnpak T, column temperature 115 °C, injector temperature 250 °C. flow 43 mL min^{-1}, bridge current 230 mA, sample size 5 µL. peaks: (1) methanol, (2) water, (3) formaldehyde, (4) butanol. (Reprinted with permission from Bombaugh and Bull.[37] Copyright 1962 American Chemical Society.)

Analytical Procedure. Chromatograms of the test and of the reference solutions are prepared under the operating conditions described in Figure 7, and the areas of the peaks of interest are determined with a polar planimeter or integrator. The peak areas are corrected using Equation (1):

$$A_c = A_m \frac{S_0}{S_m} \tag{1}$$

where A_c = corrected peak area, A_m = measured peak area, S_0 = peak area of reference at time of calibration, and S_m = peak area of reference at the time of analysis.

The corrected peak areas are used to determine the mass G of each component in the test solution according to Equation (2):

$$G = \frac{A_c - b}{m} \tag{2}$$

where b = intercept determined at the time of calibration, and m = slope of calibration line. Equations (1) and (2) can be combined to give Equation (3):

$$\% \text{ of component present} = \frac{[(A_m(S_0/S_m)) - b]100}{m/g} \tag{3}$$

where g = grams injected.

6.2.4.2 Derivatization Approaches Requiring Chromatographic Separation A major concern for formaldehyde analysis is the stability of the sample on the sampling media.[54] Derivatization of the compounds either in the sampling device or immediately after collection has improved the stability of samples for many methods. The most commonly used derivatives are 2,4-DNPH and pentafluorophenyl hydrazine (PEPH).

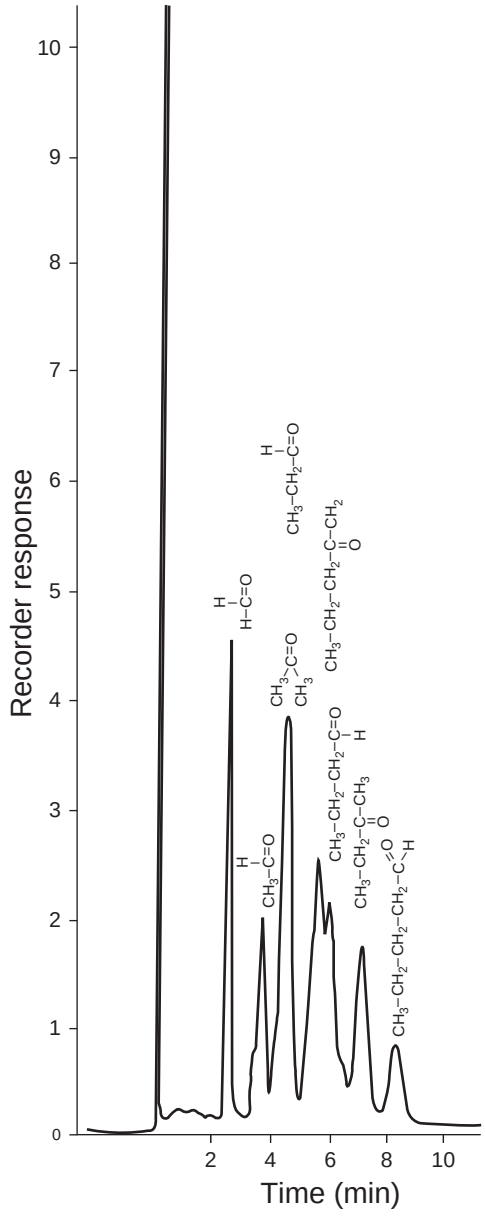

Figure 8 Gas chromatogram of a mixture of 2,4-dinitrophenyl hydrazones of carbonyl compounds. Conditions: column 4 ft × 6 mm o.d. × 2 mm i.d. glass column, packed with 2% OV liquid phase on chromosorb WAW DMCS (80–100 mesh). Flow rate of argon : methane (95 : 5) mixture 65 mL min^{-1}. (Reprinted from *J. Chromatogr.*, **136**, S. Selim, 'Separation and Quantitative Determination of Traces of Carbonyl Compounds as their 2,4-Dinitrophenylhydrazones by High-pressure Liquid Chromatography', 271–277, Copyright 1977, with permission from Elsevier Science.)

The principle and analysis procedure for formaldehyde using derivatization are same as in the direct chromatography method. In derivatization methods FID is used as the detection method.[22] The preparation of derivatives is described here.

Table 3 Operating conditions for analyses (Reprinted from *J. Chromatogr.*, **120**, Y. Hoshika, Y. Takata, 'Gas Chromatographic Separation of Carbonyl Compounds as their 2,4-Dinitrophenyl hydrazones Using Glass Capillary Columns', 379–389, Copyright 1976, with permission from Elsevier Science.)

Conditions	I	II-1	II-2	II-3
Apparatus	Hitachi model 163	Shimadzu model GC $5AP_5$	Shimadzu model GC $5AP_5$	Shimadzu model GC $5AP_5$
	Glass capillary column (Hitachi Chemi-column) 20 m × 0.25 mm i.d.	Glass capillary column (G-SCOT column), 20 m × 0.28 mm i.d.	Glass capillary column (G-SCOT column), 20 m × 0.28 mm i.d.	Glass capillary column (G-SCOT column), 20 m × 0.28 mm i.d.
Stationary phase	SE-96	SE-96	OV-17	OV-101
Column temperature	200–240 °C	200–240 °C	180–230 °C	185–225 °C
Injector and detector temperature	280–290 °C	280 °C	280 °C	280 °C
Carrier gas and flow rate	He, 1.0–1.2 mL min^{-1}	N_2, 1.0–1.2 mL min^{-1}	N_2, 1.0–1.2 mL min^{-1}	N_2, 1.0–1.2 mL min^{-1}
Sample volume	1–6 μL	1–6 μL	1–6 μL	1–6 μL
Splitting ratio	1:80–1:50	1:110	1:60	1:60
Detector	FID	FID	FID	FID

Preparation of 2,4-Dinitrophenyl Hydrazone Derivatives. 2,4-Dinitrophenyl hydrazones of formaldehyde are prepared by shaking 100 μL of formaldehyde with 100 mL of a saturated solution of 2,4-DNPH in aqueous 2 N HCl and allowing the mixture to stand at room temperature overnight. The precipitate formed is isolated, washed with 2 N HCl and water and dried over silica gel in a vacuum desiccator. The derivative formed is pure enough for GC analysis and typical gas chromatograms (Figure 8) were recorded using the conditions listed in Table 3.[22]

Preparation of Pentafluorophenyl Hydrazine Derivatives. PEPH derivatives are prepared from 0.5 mL of formaldehyde and 0.5 mL of PEPH (1.5 mg mL^{-1}) and one drop of 0.6 M phosphate buffer (pH 7.0) in 10-mL centrifuge tubes by mixing and allowing to stand for 20 min at room temperature.[56] The reaction mixture is then saturated with NaCl and one drop of 18 N H_2SO_4 and extracted with 0.2 mL of ethyl acetate containing 50 μg of *p*-xylene dichloride as internal standard. The extraction is carried out at room temperature. The excess NaCl and the aqueous layer are removed with the aid of syringe with a long needle and dried by adding a small amount of sodium sulfate. An aliquot in ethyl acetate is then applied onto the GC column. The GC conditions are as follows: 2-m glass column packed with 3% XE-60 on 80–100 mesh Celite™ 545, column temperature 120 °C, detector FID temperature 150 °C, injection temperature 150 °C and chart speed 0.25 cm min^{-1} (Figure 9).

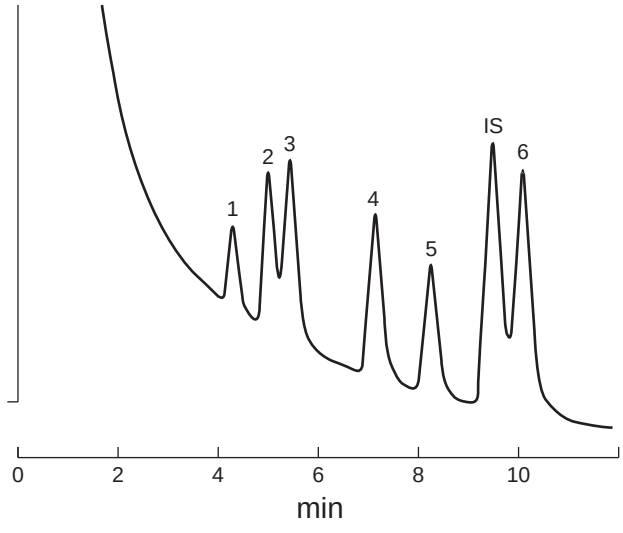

Figure 9 Gas chromatogram of some carbonyl compounds as their pentafluorophenyl hydrazones. Conditions: 3% XE-60, 2.0 m glass column, temperature programmed from 105 to 130 °C at 2 °C min^{-1}, FID. Peaks: (1) formaldehyde, (2) acetaldehyde, (3) acetone, (4) isobutylaldehyde, (5) diethyl ketone, (IS) *p*-xylene dichloride, (6) methyl isobutyl ketone. (Reprinted from *J. Chromatogr.*, **176**, K. Kobayashi, M. Tanaka, S. Kawai, T. Ohno, 'Gas Chromatographic Determination of Low-Molecular Weight Carbonyl Compounds in Aqueous Solution as their Pentafluorophenylhydrazones', 118–122, Copyright 1979, with permission from Elsevier Science.)

6.2.4.3 Gas Chromatography/Mass Spectrometry

The gas chromatographic analysis of derivatives is tedious, requiring chemical reaction, extraction and concentration steps. An interesting approach is the gas chromatography/mass spectrometry (GC/MS) method for the analysis of formaldehyde in air. This method has the advantage of sensitivity and selectivity.[38]

Principle. GC/MS is based upon mass fragmentation techniques, with an initial trapping of formaldehyde on an adsorption tube filled with the molecular sieve 13X. The collected samples are thermally desorbed onto the analytical column (Porapak T) for separation, and then

quantified by mass fragmentography (*m*/*z* 29 and 30). Advantages of the technique include parts-per-billion sensitivity and selectivity and quantitative recovery.

Apparatus. A GC/MS instrument[38] equipped with a mass fragmentation function is used to perform the analysis. A schematic diagram of the GC/MS instrument is shown in Figure 10. The system contains a two-way valve to switch over the carrier gas and block heater for desorption of the sample from the adsorption tube.

- Adsorption tube. A 1 g amount of the molecular sieve 13X is packed into a Pyrex™ tube (15 cm × 4.5 mm i.d.) to form a bed ca. 10 cm long and the tube is secured at both ends with glass wool plugs. Prior to use, the tubes are conditioned by heating at 330 °C overnight and passing nitrogen gas through at a flow rate of 30 mL min^{-1}. The nitrogen gas is purified carefully by passing it through a molecular sieve trap in dry ice–ethanol (−72 °C) in order to prevent adsorption of trace formaldehyde from the nitrogen gas during the cooling process.
- Reagents. A stock of formaldehyde is prepared by mixing 1.3 mL formalin (about 35% w/w formaldehyde) and 1 L distilled water. The accurate concentration of formaldehyde in the solution is determined by titration before use. Standards are made up by adding appropriate amounts of this solution to distilled water.

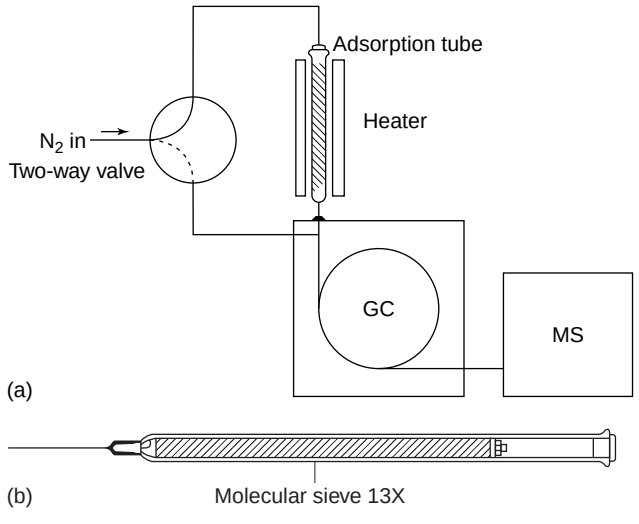

Figure 10 (a) Schematic diagram of GC/MS system; (b) absorption tube.[38] (Reprinted from *J. Chromatogr.*, **180**, T. Yokouchi, T. Fujii, Y. Ambe, K. Fuwa, 'Gas Chromatographic Mass Spectrometric Analysis of Formaldehyde in Ambient Air Using a Sampling Tube', 133–138, Copyright 1979, with permission from Elsevier Science.)

Analytical Procedure. Air (1 L) is drawn through an adsorption tube packed with the molecular sieve 13X by a portable pump (Charles Austen Pumps). The sampling flow rate of 100 mL min^{-1} is controlled by a needle valve. The sampled tubes are sealed with silicon rubber and GC/MS analysis is performed as soon as possible after sampling.

Chromatographic separations are accomplished on Porapak T (80–100 mesh) packed in a glass column (1.8 m × 2 mm i.d.). Desorption of the sample from the adsorbent is carried out at 240 °C with the high-purity helium carrier gas flowing at a rate of 20 mL min^{-1} for 2 min. During thermal desorption, the formaldehyde transferred onto the analytical column is temporarily trapped at the top for 2 min at a temperature of 30 °C and the carrier gas flow is then switched to the bypass. The mass fragmentograms of the *m*/*z* 29 and 30 ions are obtained by temperature programming at 16–150 °C min^{-1}. The concentration of formaldehyde in the samples is determined from the peak heights of the mass fragmentograms of the *m*/*z* 30 ion, using a calibration curve prepared from the standard solutions (Figure 11).

Measurement of Breakthrough Volume. Breakthrough volumes of adsorption tubes are measured in the following way. The elution volumes of the tubes are monitored for formaldehyde with the GC/MS system. When helium gas is flowing through the tube at room temperature (25 °C), formaldehyde vapor is injected (ca. 100 μg) onto the top of the tube, and the elution volume is taken as the product of the flow rate (20 mL min^{-1}) and the elution time from the tube (which is obtained by subtracting the retention time of formaldehyde on the GC column from the total elution time). The breakthrough

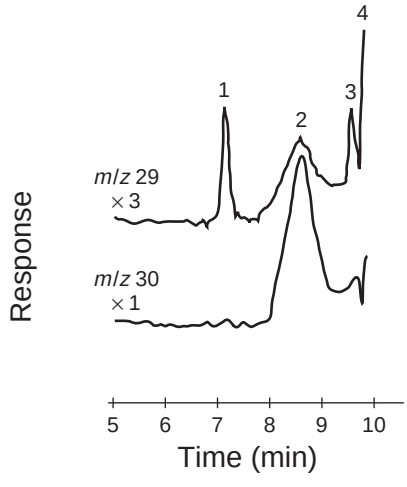

Figure 11 Mass fragmentograms of air sample. Peaks: (1) propane, (2) formaldehyde, (3) *n*-butane, (4) water peak due to excess source pressure.[38]

volume is taken as that volume of air which releases a detectable vapor (0.2 ng) from the adsorbent tube.

Calibration Curve. The calibration curve for formaldehyde is linear in the range 0–50.0 ng. The detection limit of the method is 0.3 ppb at a signal-to-noise ratio of >3.

Recovery. Recoveries for three runs with a high loading of 10 ng formaldehyde per 5 μL aqueous solution and a low loading of 1 ng formaldehyde per 5 μL solution are 101%, with standard deviations of 2 and 14%, respectively.

6.2.4.4 High-performance Liquid Chromatography

Principle. The limitations of insufficient volatility or thermal stability faced in GC analysis are not present in liquid column chromatography. A sample is introduced into the system via an injector from which it is forced by a flowing stream of solvent, often called the mobile phase, through a narrow bore transport tube to the column. The column is a large tube (2–8 mm i.d.) containing small particles (5–125 μm) known as the stationary phase. The particles alone, when held within the confines of the tubes, comprise the chromatographic bed, and a sample mixture separates as a result of different components adhering to or diffusing into the packing particles. Thus, as the mobile phase is forced through the chromatographic bed, a sample is separated into various zones of sample components, referred to as bands. The bands continue to migrate through the bed, eventually pass out of the column (elution), and pass through one or more detectors. The recorder tracing from the elution of a single band is called a peak. Peaks are usually identified by their retention time, which is the time required to elute the corresponding band from the column.

Apparatus. For isocratic operation, a standard HPLC instrument has the following components (Figure 12):[46] (1) a solvent reservoir for the mobile phase; (2) a solvent pump (or pumps) equipped with a damping unit, if a pulsating action results, to force the mobile phase through the chromatographic system; (3) a precolumn (except for bonded phases) and a guard column, the former to presaturate the mobile phase with the stationary phase and latter to prevent contamination of the separation

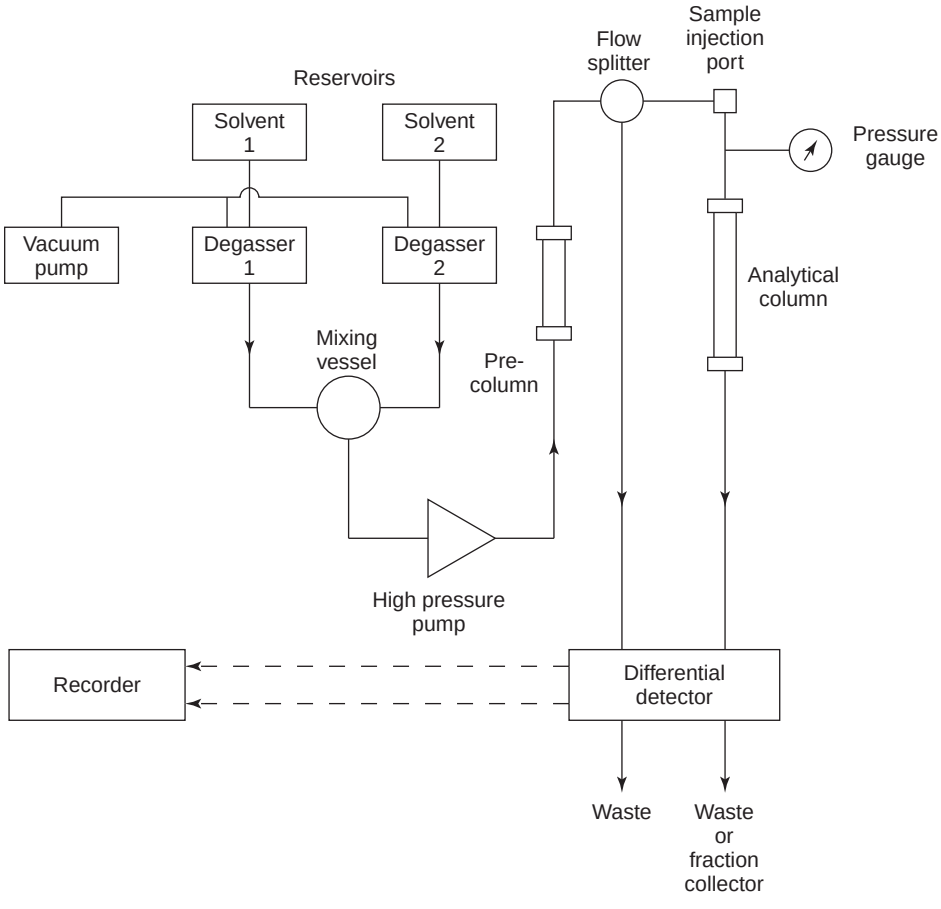

Figure 12 Schematic diagram of HPLC system.[53]

column; (4) a pressure gauge inserted close to the column to measure column inlet pressure; (5) a sampling or injection device to introduce the sample into the column; (6) the separation column; and (7) a detector with recorder readout or other data handling device.

Derivatives that absorb strongly in UV light have been prepared for HPLC. The method commonly used to determine traces of formaldehyde is to form the 2,4-dinitrophenyl hydrazone and determine the derivative by HPLC. However, for C_1-C_5 aldehydes 2,4-dinitrophenyl hydrazones have been synthesized and successfully separated by adsorption chromatography. Selim[55] reported the separation of 2,4-dinitrophenyl hydrazones of aldehydes by using reversed-phase HPLC. Beasley et al.[39] reported a new solid sorbent silica gel coated with 2,4-DNPH for formaldehyde. Analysis of the resulting hydrazone derivative is done by HPLC with a UV detector.

Reagents. Reagents include formaldehyde (37%), dimethylformamide (DMF), hydrochloric acid, acetonitrile and silica gel. Paraformaldehyde permeation tubes can be purchased or α-POM permeation tubes can be prepared from α-POM.

The α-POM tubes or the purchased paraformaldehyde permeation tubes can be used successfully to generate a dynamic formaldehyde standard in air.

The 2,4-DNPH reagent is prepared by adding 0.25 g of 2,4-DNPH to 100 mL of 6 N HCl.

The 2,4-DNPH coated silica gel is prepared as follows. The mixture of 2,4-DNPH in DMF is unstable and must be used quickly. DMF is used because of its superior solvent properties. Silica gel (12.5 g) is placed in a 100-mL round-bottom flask with by 12.5 mL of 6 N HCl. DMF (40 mL) is added to a 100-mL volumetric flask containing 5.0 g 2,4-DNPH. This mixture is swirled for 10–15 s, quickly diluted to the mark with additional DMF and immediately poured through a glass wool plug into the flask containing the silica gel and HCl. The volumetric flask and glass wool are rinsed with an additional 5 mL of DMF, which is also added to the round-bottom flask. The mixture is allowed to stand for 30 min, with occasional swirling, before isolation by vacuum filtration. The coated silica gel may be rinsed sparingly with DMF (2–3 mL). Excessive washing will result in increased capacity for formaldehyde. It is left on the filter paper with continued suction for 1–2 min and then transferred quickly to a 100-mL round-bottom flask. It is dried under vacuum at 55 °C for 1 h with brief turning every 10 min.

The collection device is a glass tube (10 cm × 4 mm i.d. × 6.2 mm o.d.) containing a 300-mg front section and 75-mg backup section of the coated silica gel. The two sections are separated by small plugs of glass wool. Tubes prepared as described here are found to have a pressure drop of 52 kPa in H_2O at a flow rate of 115 mL min^{-1}. They have a capacity of ~123 μg formaldehyde (5 ppm for a 20-L air sample) before significant breakthrough occurs, regardless of humidity.

Air Sampling. Sampling is done by connecting the sampling tube to the Dupont pumps. Formaldehyde loadings are varied by time of collection with constant permeation and flow rates. Typically, the output of the permeation device is set for 1.0 ppm and is sampled at 100–200 mL min^{-1}.

Analysis. The front section of the collection tube, including the front glass wool plug, is transferred to a 1.8-g vial containing 2.0 mL acetonitrile. The backup section and the glass wool plug that separates the two section are treated similarly. The vials are stoppered with polypropylene caps (no Bakelite) and allowed to stand overnight.

For analysis 0.2 mL of the desorbing solution is injected in the loop filter port to provide flushing and loading of the 50 μL injection volume. The system also is flushed with acetonitrile between injections. The mobile phase for the analysis is 65 : 35 (v/v) water : acetonitrile, pumped at 2.0 mL min^{-1}. A typical chromatogram is shown in Figure 13. Quantitation of this work is done by peak height/peak area using a calibration curve generated from a series of the standards containing 10–400 μg mL^{-1} authentic formaldehyde 2,4-dinitrophenyl hydrazone in acetonitrile (corresponding to 0.1–4.6 ppm formaldehyde in air for a 20-L sample). Beasley et al.[39] validated this method over the range 2.5–95 μg of formaldehyde with 94% recovery and relative standard deviation ±0.04.

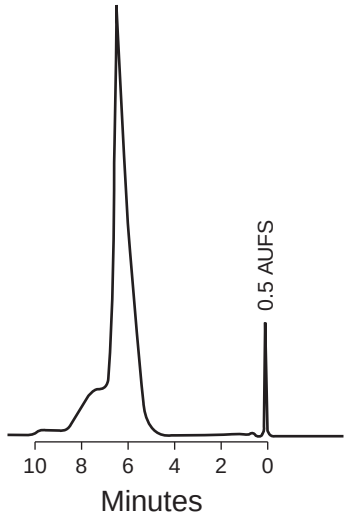

Figure 13 HPLC chromatogram for injection of 4.6 μg of formaldehyde-2,4-dinitrophenyl hydrazone (equivalent to 1.08 ppm formaldehyde in air, for a 20-L air sample). (Reprinted with permission from Beasley et al.[39] Copyright 1980 American Chemical Society.)

DNPH derivatization followed by HPLC analysis involves pre-column reactions which may be affected by other components in the sample. In addition the resultant signals, absorptivities and conductivities are sensitive to matrix effects. Igawa et al.[57] have developed an analytical procedure that couples HPLC with a segmented flow analyzer which involves a post-column reaction. Formaldehyde was separated on a reversed-phase C_{18} column derivatized with MBTH and detected at 640 nm.

3-Methyl-2-benzothiazole Hydrazone Derivative. The aldehydes which were separated on the reversed-phase C_{18} column are derivatized with MBTH and detected at 640 nm.

The eluent is an acetonitrile–water mixture prepared from HPLC-grade acetonitrile and purified deionized water.

The reagents used in the reaction detector are MBTH (0.5% w/v), Fe^{3+} solution and Triton X-100. MBTH is prepared by dissolving 3-methyl-2-benzothiazolidone hydrazone hydrochloride monohydrate in water to form a 0.5% (w/v) solution. $FeCl_3$ is prepared by dissolving $FeCl_3 \cdot 6H_2O$ in 0.01 M HCl (10% w/v). Triton X-100 is prepared by diluting with water to 0.1% (v/v).

The HPLC apparatus is coupled with a rapid flow analyzer. The HPLC apparatus is equipped with a reversed-phase C_{18} analytical column and an adsorbosphere C_{18}. A guard column is connected just before the analytical column, and there is a stream tubular reactor.

The sample is injected by a variable volume injector. The separation of aldehyde is performed with isocratic elution using a 40:60% or 1:99% by volume acetonitrile:water mixture. The flow rate and the column temperature are maintained at 0.5 mL min^{-1} and 30 °C, respectively. Post-column reaction occurs at the stream tubular reactor. Absorbance of the colored solution is measured intermittently at 640 nm using the bubble gate of the analyzer. The pathlength is 1 cm. The signal is recorded on a chart and concentration is measured from the peak height or peak area, as determined by the integrator (Figure 14).

Calibration is based on comparison of the sample peak height with those obtained by injecting a mixture containing a known concentration of formaldehyde with a 1% acetonitrile–water eluent.

6.2.4.5 Ion Exchange Chromatography IEC is an analytical method combining formaldehyde collection on an oxidizer impregnated with charcoal with subsequent IC analysis of the formaldehyde as formate ions. It has been developed by Kim et al.[40]

Principle. The essential principle of this technique is an ion exchange process between the mobile phase and the exchange group covalently bound to the stationary phase. The exchange function in anion chromatography is generally a quaternary ammonium base. During the chromatographic process, the counter-ion of the exchange function is exchanged for the sample ion, and for a short time the latter becomes attached to the fixed charge on the resin. Due to variations in the affinities of the species being chromatographed for the stationary phase, a separation of the various sample components is possible.[58]

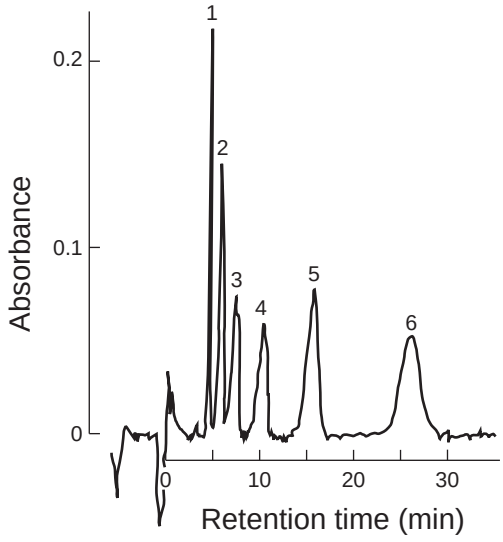

Figure 14 HPLC chromatogram of C_1–C_6 carbonyls. Conditions: Spherisorb ODS-2 (5 μm), 4.6 × 250 mm, injection volume 20 μL, eluent 40% acetonitrile + 60% water. Peaks: (1) 25 μM HCHO, (2) 50 μM CH_3CHO, (3) 75 μM C_2H_5CHO, (4) 100 μM C_3H_7CHO, (5) 150 μM C_4H_9CHO, (6) 200 μM $C_5H_{11}CHO$. (Reprinted with permission from Igawa et al.[57] Copyright 1989 American Chemical Society.)

Any substance that has a retention time coinciding with that of formate will interfere. A high concentration of any one ion also interferes with the resolution of others but sample dilution overcomes many interferences.

Apparatus. An ion chromatograph (Figure 15) includes an injection valve, a sample loop, an AG4A guard column, an AS4A separator, a self-regenerating suppressor column, a temperature-compensated small volume conductivity cell (4 μL) and a recorder capable of full scale response of 2 s or less. An electronic peak integrator is optional. The ion chromatograph should be capable of delivering 2–5 mL eluent min^{-1} at a pressure of 9.65–13.8 MPa.

Reagents. Deionized or distilled water is free from interferences at the minimum detection limit of each constituent. The eluent solution is 5 mM $Na_2B_4O_7 \cdot 10 H_2O$ (borax). Standard solution (100 mg L^{-1}) contains 1.410 g sodium formate dissolved in 1 L water and stored

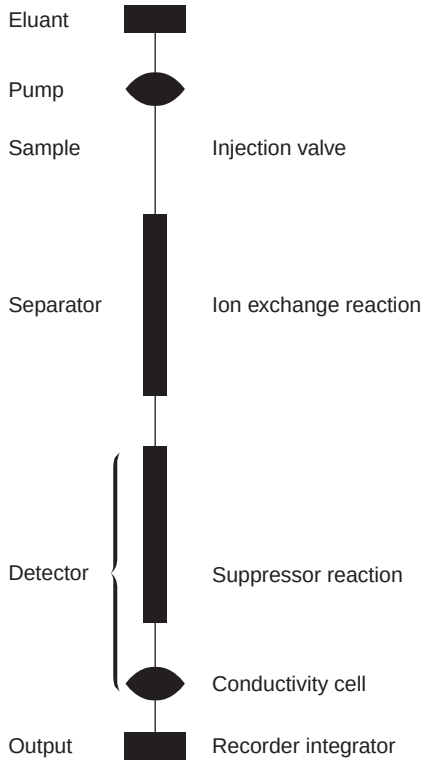

Figure 15 Schematic diagram of ion exchange chromatograph.[58]

in plastic bottles in a refrigerator. The working solution (1 ppm) contains 1 mL of standard solution diluted to 100 mL.

Procedure. The air is passed through a solid sorbent tube (4 mm i.d. × 7 cm long) containing 150 mg of impregnated charcoal at 50–200 mL min^{-1} for 3–5 h. The charcoal is removed to a tube containing 15.0 mL of 0.1% H_2O_2 and shaken for 1 h, then sonicated for 20 min. The solution is filtered through a 0.45-μm membrane filter and the sample is injected into an ion chromatograph.

After turning on the ion chromatograph the eluent flow rate is adjusted to 2 mL min^{-1} and the system is allowed to equilibrate for 15–20 min. A stable baseline indicates equilibrium conditions.

Calibration. The approximate retention time of the standard is determined (Figure 16). Concentrations are calculated based upon peak areas of standard that are prepared daily from 100 ppm stock solution. The stock solutions are prepared weekly from the sodium salt of formic acid, spiked with $CHCl_3$ and stored at 4 °C. Typically standard concentrations are in the range 0.1–1 ppm. The standards are run in triplicate. After every five samples the standard is run to check peak response, and if deviation is more than 2% the system should be recalibrated.

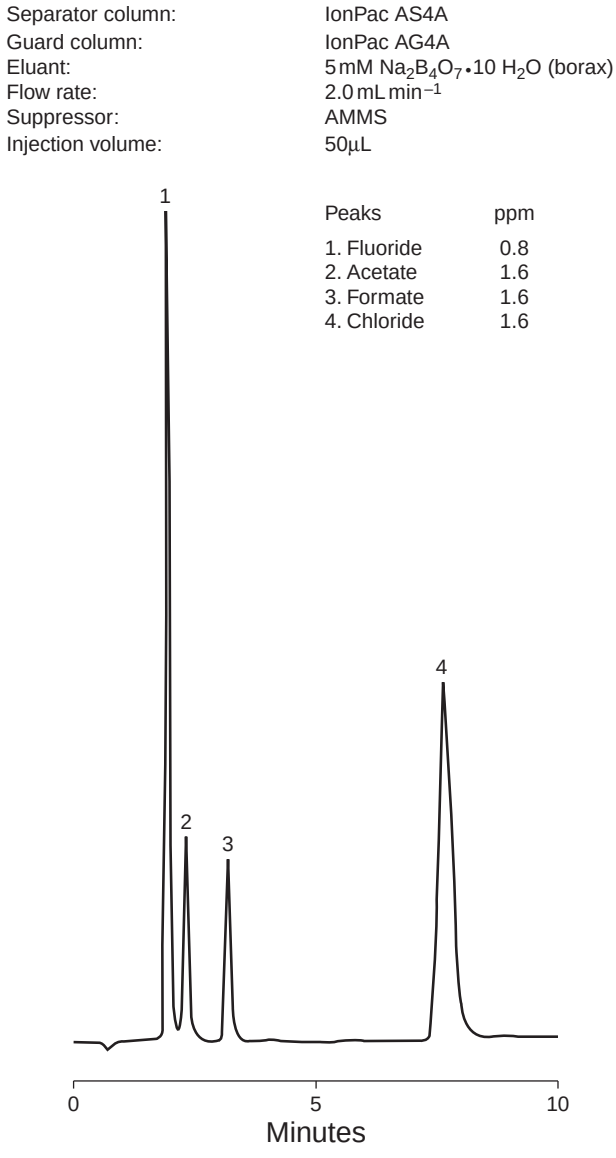

Figure 16 Chromatogram showing determination of formaldehyde as formate ion.[40]

The analytical precision is determined by running the 1 ppm standards nine times, and is found to be 1.64%.[49]

Sample Analysis. The sample is injected into the ion chromatograph and the peak height/peak area and retention time are recorded.

Calculation.

$$C = HFD$$

where C = mg formaldehyde L^{-1}; H = peak height/area, F = response factor = concentration of standard/height (area) and D = dilution factor for those samples requiring dilution.

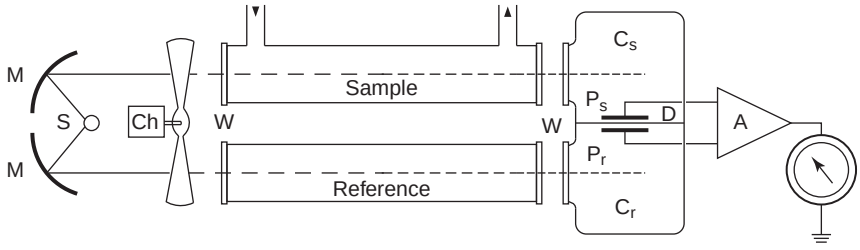

Figure 17 Nondispersive Luft type IR gas analyzer. A = amplifier, C_s, C_r = Luft detector cells for sample and reference beams, Ch = chopper, D = thin metal diaphragm, M = mirrors curved to focus beam from source S on to windows W of suitable IR-transparent material, P_s, P_r = fixed plates of double capacitor P_s–D–P_r.[61]

6.2.5 Infrared Spectroscopy

Infrared (IR) gas analyzers depend for their operation upon the fact that some gases and vapors absorb specific wavelengths of IR radiation. The technique uses a conventional double beam IR spectrometer system with a pair of matched gas cells in the two beams. One cell is filled with a reference gas, which is a non-absorbing gas like nitrogen, while the measuring cell contains the sample. The difference in optical absorption detected between the two cells is a measure of the absorption of the sample at the particular wavelength. The selectivity of this technique is high because of characteristic absorption spectra corresponding to the difference in vibrational energy levels of the bands. Thomson and Keller[59] and Zeller[60] have compiled and published IR spectra of hazardous gases and vapors. Aliphatic aldehydes show strong absorption near 1730 cm^{-1} (5.75 μm) (1740–1720 cm^{-1} region) due to C=O stretching. C–H stretching absorption [2825, 2717 (3.54, 3.68 μm)] is seen in the 2830–2695 cm^{-1} (3.53–3.71 μm) region. Two moderately intense bands are frequently observed in this region. The appearance of two bands is attributed to Fermi resonance between the fundamental aldehyde C–H stretching and the first overtone of the aldehyde C–H bending vibration that usually appears near 1390 cm^{-1} (7.20 μm).

6.2.5.1 Nondispersive Infrared Analysis Long-path IR techniques usually involve long cylindrical glass cells (about 5 cm diameter and 100 cm length) filled with gases and end windows of NaCl (transparent down to 650 cm^{-1}), KBr (transparent down to 400 cm^{-1}) or CaF_2 (transparent down to 1250 cm^{-1}). CaF_2 is the most resistant to attack by water but is also much more easily chipped or cracked. In a nondispersive infrared (NDIR) spectrometer the IR radiation from the source is not dispersed according to wavelength by a prism or grating as in a standard IR spectrometer. It is made very specific for a given compound or type of compound by using the material under investigation as part of the detector. The detector is efficient in that it ''sees'' all the wavelengths of interest simultaneously. It is also selective in that interference by another gas at one particular wavelength will be minimized because it is not present at most other wavelengths.[61]

This is the basis of the Luft nondispersive detector shown in Figure 17. Light is split along two paths: in one is a reference cell filled with clean air, in the other is the sample of contaminated air. The detector consists of two sealed chambers filled with the same (pure) gas which is to be determined in the air sample. As less radiation reaches the area C_s, the pressure increase is less than in C_r, so the flexible metal diaphragm D is bent towards the plate P_s of the differential capacitor P_r–D–P_s. The chopper then cuts off radiation from both halves and the pressure differential falls off. The output is a series of pulses, the height of which increases with increasing formaldehyde concentration.

6.2.5.2 Fourier Transform Infrared Spectroscopy In FTIR instruments, the IR spectra are more accurately and more readily obtained in a matter of seconds using the

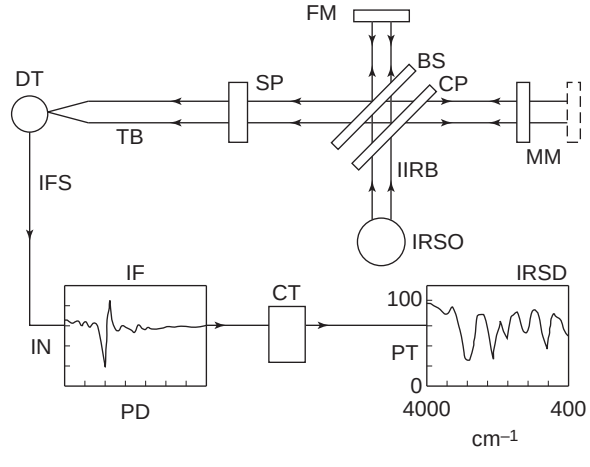

Figure 18 Schematic diagram of FTIR spectrometer. IRSO = IR source and optics, IIRB = incident IR beam, MM = movable mirror, CP = compensator, BS = beam splitter, FM = fixed mirror, SP = sample, TB = transmitted beam, DT = detector, IFS = interferogram signal, IF = interferogram, IN = intensity, PD = path difference, CT = computer, PT = %transmittance, IRSD = IR special display.[62]

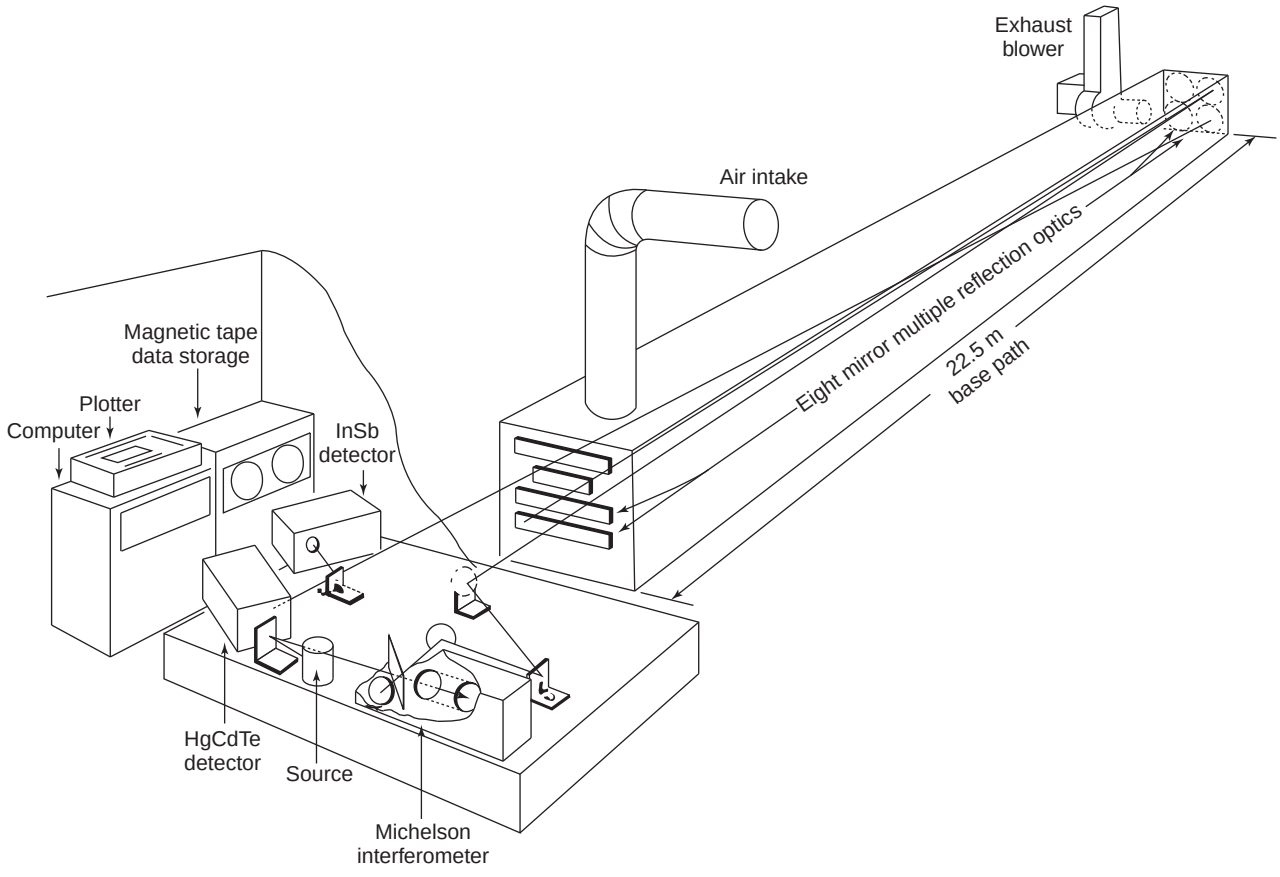

Figure 19 Schematic diagram of an FTIR system used in ambient air studies.[24]

mathematical technique called Fourier transformation. Figure 18 shows the schematic diagram of an FTIR instrument.[62] An interferogram is generated as the mirror moves, and the path difference between the two beams changes. The Fourier transformation of the interferogram is done by a computer to decode it and produce the IR spectrum.

Since all wavelengths of IR spectra are measured simultaneously, the method is several times faster than with dispersive instruments. As slits and gratings are not used, the full IR energy is available to be sent through the sample (high energy throughout) so that even a trace of the sample is adequate. This technique is sensitive down to 20 ppb formaldehyde in air and is ideal for determining vehicle emissions. Measurement of formaldehyde by FTIR has been reported by Scott Research Laboratories,[6] Tuazon et al.,[24] Haack et al.,[63] and Lawson et al.[34] Figure 19 is a schematic diagram of an FTIR system used to monitor pollutants in ambient air. The eight mirror system shown in Figure 19 was exploited by Tuazon et al.[24] in measurements over a period of five years in the Los Angeles basin. The 22.5-m path multiple reflection cell shown afforded a total pathlength in excess of 1 km.

6.2.5.3 Fourier Transform Infrared Spectroscopy for the Measurement of Formaldehyde in Vehicle Exhaust

Measurement of formaldehyde in vehicles exhausts is described here following the method developed by Haack et al.[63]

Apparatus. The general set-up used for vehicle exhaust analysis is depicted in Figure 20. The constituents are a vehicle chassis dynamometer, dilution tunnel and FTIR apparatus connected as shown. The vehicle tailpipe is connected to a 7.6 m × 30.5 cm i.d. dilution tunnel via an insulated pipe. Pure, dry air is used as the diluent for vehicle exhaust. Dilution ratios from 7 : 1 to 10 : 1 are used depending upon the anticipated levels of formaldehyde generated by each test vehicle. The dilution of the exhaust ensures that condensation of water does not occur in the unheated FTIR apparatus.

Fourier Transform Infrared Sampling System. The FTIR sampling system is shown in Figure 21. The sample is drawn from the dilution tunnel through an unheated 45.7 m × 0.95 cm i.d. line made of Teflon™ and hence through the Wilks cell of the FTIR apparatus by means

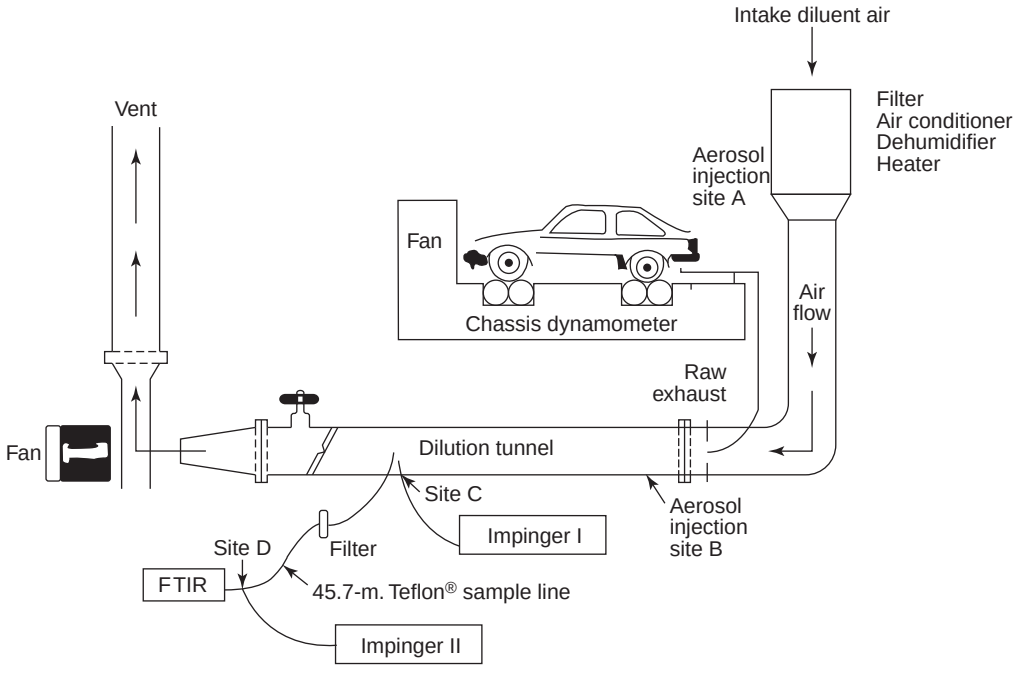

Figure 20 Sampling apparatus for vehicle exhaust analysis. (Reprinted with permission from Haack et al.[63] Copyright 1986 American Chemical Society.)

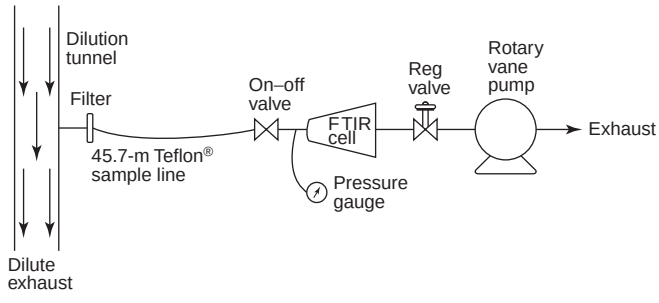

Figure 21 FTIR sampling system. (Reprinted with permission from Haack et al.[63] Copyright 1986 American Chemical Society.)

of a rotary vane pump at a flow rate of 25 L min^{-1} (10 s lag time). The sample in the cell remained at room temperature, while the pressure was maintained at 93 kPa by using a metering valve between the cell and the rotary vane pump. The sample is filtered at the dilution tunnel through a 142-mm diameter quartz fiber filter in a stainless-steel holder at room temperature. The filter holder is heated to 55 °C for diesel-fueled vehicle experiments. The purpose of the filter is to prevent particulate matter from entering and contaminating the FTIR sample line and cell.

Fourier Transform Infrared Spectrometer. As described previously, FTIR spectrometers differ from conventional grating or prism instruments in that wavelength determination is accomplished by modulating each wavelength at its own unique audio-range frequency via a scanning Michelson interferometer. The digital interferogram is frequency analyzed (Fourier transformed) to produce the IR spectrum. The sample cell used is a custom gold-coated Wilks 20-m variable path cell, used in 14th order with an effective pathlength of 21.75 m. A HgCdTe detector exhibiting negligible nonlinearity is used. An absorbance spectrum is generated by calculating the logarithm of the ratio of the corrected spectrum of the tunnel dilution air to the corrected sample spectrum.

The formaldehyde IR signal is measured by integrating the absorbance of selected vibration–rotation lines in the C–H stretching region from 2739 to 2827 cm^{-1}. Figure 22 displays the FTIR formaldehyde spectra of a standard and a methanol-fueled vehicle exhaust sample. The exhaust spectrum is generated from one interferogram. By using this standard the formaldehyde diluted exhaust concentration is determined.

6.2.6 Fluorimetric Method

Most of the in situ techniques described earlier require several minutes to several hours to collect sufficient amounts of the samples, depending on the ambient formaldehyde concentration. These techniques have higher detection limits (ppm or ppb) and therefore cannot be used for in situ measurements where formaldehyde occurs in parts per trillion by volume concentration ranges. Fluorimetric determination of formaldehyde in

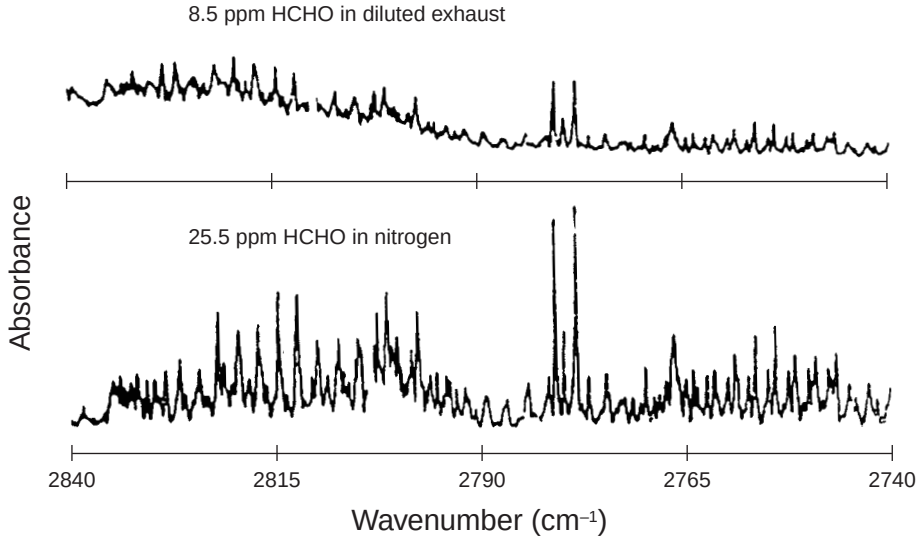

Figure 22 FTIR formaldehyde spectra in nitrogen and in diluted methanol-fueled vehicle exhaust. (Reprinted with permission from Haack et al.[63] Copyright 1986 American Chemical Society.)

air has been discussed by Lazrus et al.,[28] Heikes,[10] Heikes et al.,[64] and Dong and Dasgupta.[65] A fluorimetric method based on diffusion of atmospheric formaldehyde vapor to an enzyme-coated piezoelectric crystal detector[28] has been developed for automated fluorimetric determination of formaldehyde in air. The unique advantage of enzyme catalysis for atmospheric measurements is the ability of many enzymes to react selectively with a single species within a complex mixture of compounds. The automated instrument consists of device for continuous rapid transfer of formaldehyde vapor from the sample air to water, a peristaltic Technicon Auto Analyzer pump and a fluorometer fitted with a flow cell to analyze the aqueous formaldehyde.

The dissolved formaldehyde reacts reversibly with H_2O to form the dihydroxymethane adduct which is the dominant form in solution. The $CH_2(OH)_2$ reduces nicotinamide adenine dinucleotide (NAD^+) to its fluorescent form (NADH), a reaction which is catalyzed by formaldehyde dehydrogenase (FDH). The suggested overall reaction mechanism is shown in Schemes 13 and 14.

$$CH_2(OH)_2 + FDH \longrightarrow HCOO^- + H^+ + H_2FDH$$

Scheme 13

$$H_2FDH + NAD^+ \longrightarrow NADH + FDH + H^+$$

Scheme 14

The automated technique is free of interferences from common atmospheric constituents, sensitive and relatively fast. The detection limit is 120 pptv. The calibration curve is linear to approximately 100 ppbv. The coefficient of variation is 2% at 20 ppbv formaldehyde. The rise time of the signal from 10 to 90% is 80 s.

6.2.6.1 Reagents Stock standard solution (0.01 N) is made by dissolving 0.30 g of paraformaldehyde in water (1 L), which is standardized volumetrically as described previously.

Formaldehyde vapor standards are generated by equilibrating air with aqueous formaldehyde standard. The standard generator apparatus (Figure 23) consists of a tube lined with glass-fiber filter paper. A liquid formaldehyde standard runs down the filter while air

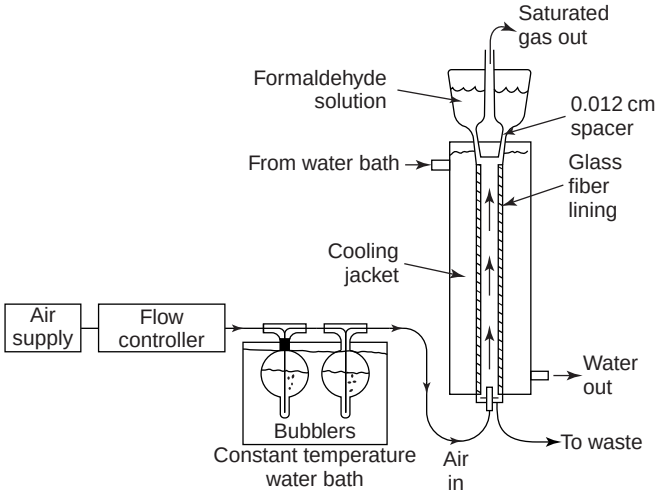

Figure 23 Vapor phase standard generator. Typical air flow rate is $<1\,L\,min^{-1}$. (Reprinted with permission from Lazrus et al.[28] Copyright 1988 American Chemical Society.)

supplied by a pure air generator runs in a counter current flow. The tube is thermostatted (usually at 18 °C). Dilution of this air to yield the desired concentration of formaldehyde occurs in the test manifold. The coefficient of variation for the difference between calculated and observed formaldehyde in vapor phase standards ranging between 1.8 and 4.4 ppbv is 10.3% ($n = 11$) resulting principally from errors in gas flow control and dilution. No wall losses are detected when using an inlet of PFA (polyfluoroalkene) tubing, with 0.64 cm i.d. and 3 m length. To set the instrument at zero, formaldehyde is initially removed from incoming air by passing it through a molecular sieve trap which is 25.4 mm i.d. and 152.4 mm long. The molecular sieve is 13X (0.32 cm pellets). The efficiency of the molecular sieve trap appears to decline too quickly for it to be useful as a zeroing device in this application. To irreversibly bind formaldehyde, a trap constructed of PFA Teflon® (4 cm i.d., 6 cm long) was filled with silica gel coated with 2,4-DNPH[39] The collection efficiency, at 1.0 L min^{-1}, is 97% for 7.0 ppbv formaldehyde. When clean air is sampled (0.1–2.0 ppbv formaldehyde), two additional techniques can be used for stabilizing the instrument to zero. The first is simply to use ''zero'' air free of formaldehyde. The second is to use the reagent blank as the zero signal. The reagent blank is obtained by allowing the scrubbing solution to bypass the air scrubbing coil. The use of ''zero'' air and the reagent blank yield identical signals and are believed to be the most reliable means of zeroing the instrument at a very low concentration of formaldehyde.

To prepare the FDH reagent, FDH from *Pseudomonas putido* in the form of lyophilized powder is dissolved in water (16 enzyme units per 100 mL) (one enzyme unit oxidizes 1.0 μmol of formaldehyde to formic acid per minute at pH 7 and 30 °C). This reagent is buffered at pH 8 with 0.01 M Na_2HPO_4.

To prepare the NAD^+ reagent, 35 mg of 99+% crystalline β-NAD is dissolved in 100 mL water (approximately 5×10^{-4} M). The pH is adjusted to 4–5 with HCl for stabilization.

6.2.6.2 Experimental Formaldehyde vapor is stripped from the atmosphere by means of concurrently pulling the air sample and scrubbing water through a 28-turn 3 mm i.d. Technicon Auto Analyzer mixing coil (Figure 24). The scrubbing water is adjusted to pH 2.0 with hydrochloric acid and nominally flows at 0.42 mL min^{-1} through the coil. Air is sampled at 1.0 L min^{-1} giving a residence time of about 0.3 s. The scrubbing water, which is impelled by the air as it flows through the vertical coil, forms a thin film on the glass surface, providing a large surface for gas exchange. For maximum accuracy, the delivery rate of the pump tube introducing water into the scrubbing coil is calibrated.

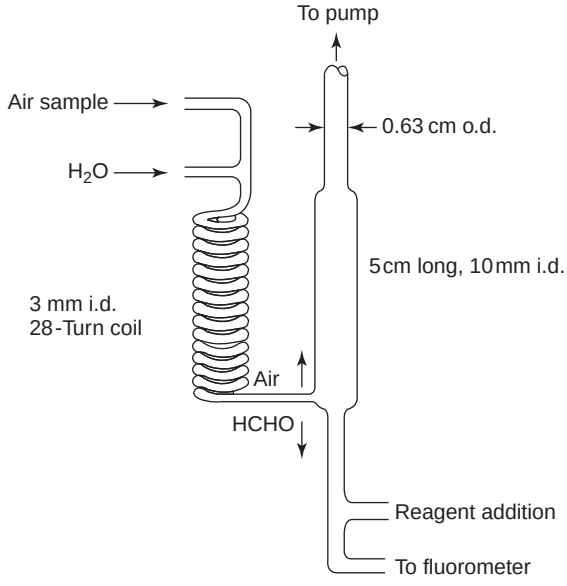

Figure 24 Pyrex™ glass stripping coil and separator. Typical flow rates are 1.0 L min^{-1} for air and 0.42 mL min^{-1} for scrubbing water. (Reprinted with permission from Lazrus et al.[28] Copyright 1988 American Chemical Society.)

The aqueous formaldehyde is isolated from the scrubbing air in a vertical separator column by gravity. It is then mixed with buffer (0.2 M Na_2HPO_4 adjusted to pH 8 with HCl) and led into a five-turn mixing coil. The FDH and NAD^+ are then added to the reaction stream, followed by a second five-turn mixing coil. The reaction mixture is pumped through a reaction time delay coil (13.2-min delay) to allow production of NADH to completion (Figure 25). A filter is placed in the enzyme line downstream of the pump to remove particles occasionally seen in this reagent. The polycarbonate filter (8.0 μm pore size and 13 mm diameter) is contained in an in-line polycarbonate filter holder. The bubbles are removed from the flow stream by a Technicon debubbler placed immediately upstream of the fluorometer flow cell.

6.2.7 Tunable Diode Laser Absorption Spectroscopy

TDLAS is a technique based on IR absorption spectroscopy. Unlike FTIR this technique relies on measuring the absorbance at specific wavelengths due to absorption of IR radiation by various pollutants. However, rather than using a continuous wavelength light source and scanning the entire IR spectrum, TDLAS employs a laser light source of very narrow linewidth which is tunable over a smaller (300 cm^{-1}) wavelength range.[5]

The very high spectral resolution of the lasers allows the measurement of single rotational vibrational lines in the spectrum of the target molecule which leads to excellent specificity. Moreover, the method is highly

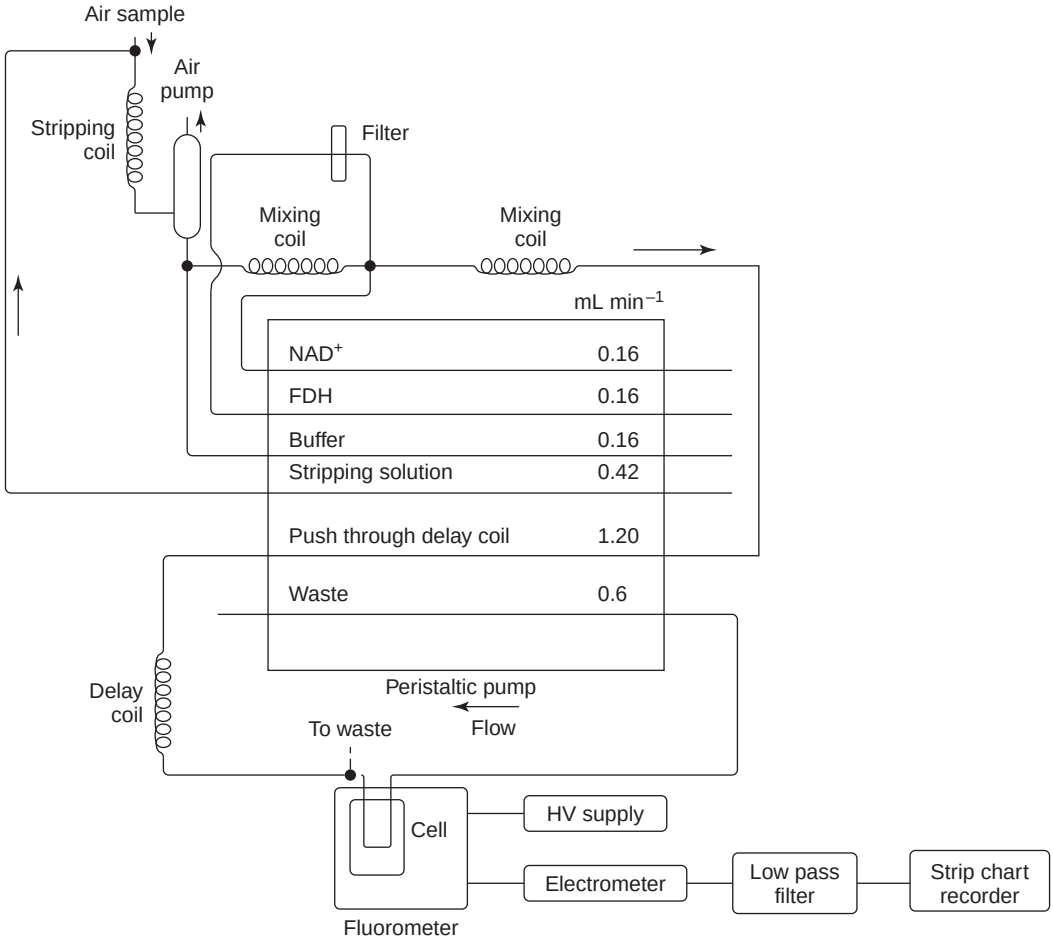

Figure 25 Schematic diagram of the automated fluorometer. (Reprinted with permission from Lazrus et al.[28] Copyright 1988 American Chemical Society.)

sensitive and allows for a very good time resolution. The method is therefore suitable for sampling from aircraft or for the detailed study of temporal variations of photochemically related species and for validating photochemical models of the troposphere in the remote troposphere, particularly where accurate measurements are needed. TDLAS is more useful for following specific pollutants known to be present than for searching for previously unidentified species. Many workers[25,66–69] have discussed different types of tunable IR laser sources used in absorption spectroscopy for monitoring of pollutants. Tuning of the emitted wavelength of the diode can be accomplished through variation of one of three possible parameters: applied magnetic field strength, diode temperature and hydrostatic pressure. Temperature, which can be controlled by changing the current through the diode, is the most commonly used.

Several different modulation techniques can be used to increase the signal-to-noise ratio. For example, the laser beam can be mechanically chopped and detected using phase-sensitive detection with a lock-in amplifier. A more commonly used method for accurately measuring small absorbances is to modulate the frequency output of the laser by modulating the current and thus the temperature of the diode.[70] Absorbances down to 10^{-5} M can be measured, corresponding to ppb to sub-ppb concentrations at a pathlength of 40 m (e.g. using white cell optics) for many pollutants of atmospheric interest.

The Doppler limited C–H stretching spectrum in the 2700–3000 cm^{-1} region (3.3–3.5 μm) or functional group (carbonyl group) absorption in the 1740 cm^{-1} (5.6 μm) region is used for monitoring of formaldehyde. A continuous wave (CW) difference frequency spectrometer, which operates by mixing on argon ion laser with a tunable dye laser in a non-linear optical crystal ($LiNbO_3$), is used for monitoring formaldehyde in the 3.5-μm region. Pine[67] constructed this type of laser source, which operates in the 2.2–4.2 μm region and achieved a tunable IR output with microwatt power at a resolution of 0.001 cm^{-1}.[71–73]

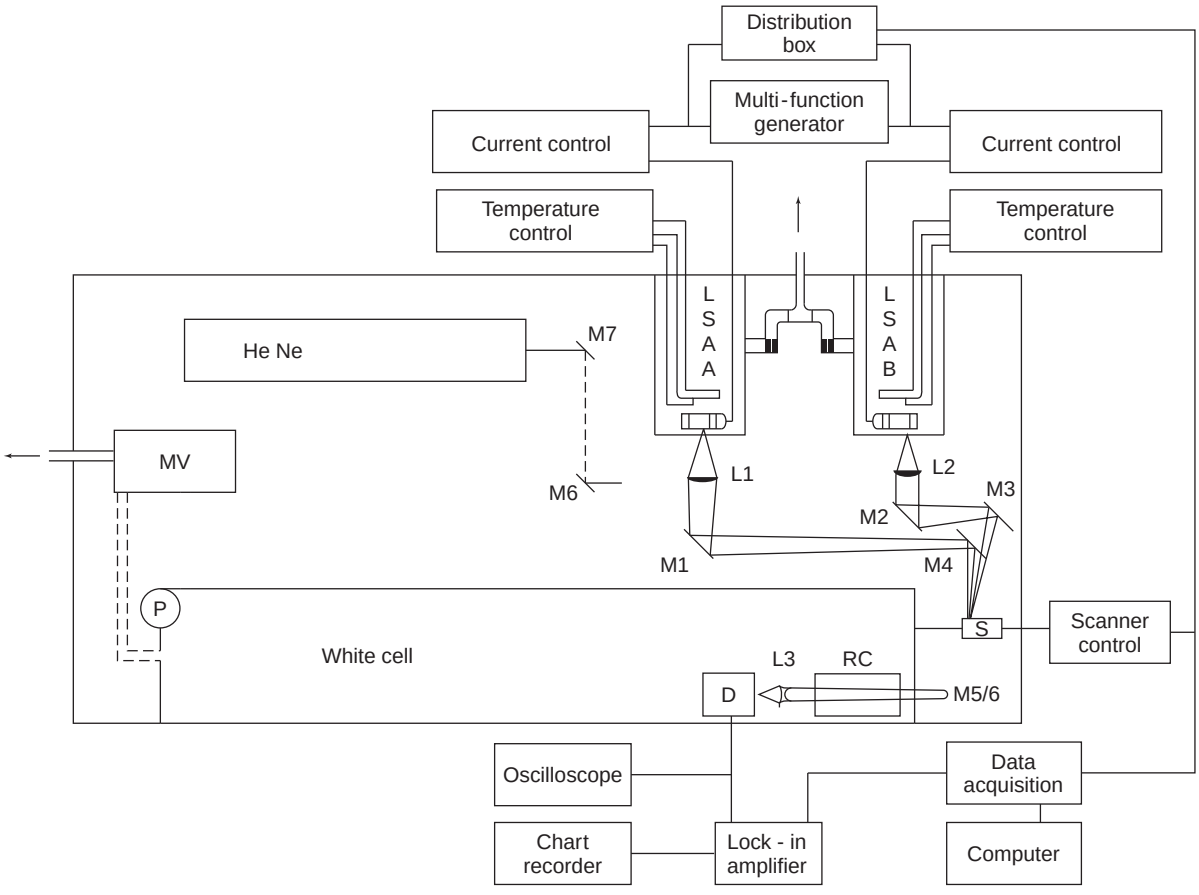

Figure 26 Schematic diagram of the TDLAS instrument. LSA = laser source assembly, M = mirror, D = detector, S = switching mirror, MV = motivated valve, RC = reference cell, P = pressure gauge, L = lens. (Reproduced from Harris et al.[25] with kind permission from Kluwer Academic Publishers.)

A white cell is used in the TDLAS system to provide the long pathlength required to attain the desired sensitivity. The sample is pumped at reduced pressure (3.33 kPa) through the white cell to reduce the overlap between absorption lines of atmospheric constituents. There is also a provision to fit the observed spectra to the known spectral properties of various gases that absorb in the spectral region of interest. A background spectrum is taken in the ambient sample after removing only the formaldehyde. The difference between the ambient and background spectra is due to formaldehyde alone. When the curve is fitted to a database of known spectra the resulting concentration is tied directly to the physical properties of the gas of interest. A calibration with known concentrations is only needed as a check on whether the sampling system efficiently passes the formaldehyde to the optical cell. Formaldehyde has been estimated using TDLAS by Harris et al.,[25] Mackay et al.,[27] Harder et al.,[8] and NASA.[74]

6.2.7.1 Instrumentation A schematic diagram of the instrument developed by Harris et al.[25] is shown in Figure 26. The instrument incorporates two laser source assemblies to enable simultaneous monitoring of two species in the same air sample, and uses line locking of the linear output frequency to the selected absorption features by microprocessor control of the laser operating temperatures. It also incorporates a data acquisition and instrument control system which permits automated operation. The laser output is frequency modulated at 7.5 kHz and synchronously detected at 15 kHz. A switching mirror, S, is first aligned to direct radiation from laser A into the white cell. The computer generates the laser output as a 2*f* spectrum.

The entrance window on the white cell is mounted at an angle of about 45° with respect to the laser beam and serves as a beam splitter. About 5% of the laser power is reflected through cells (RC1 and RC2) containing high concentrations of the monitored species to detector D. The signal from this detector is demodulated at 7.5 kHz using a phase sensitive detector whose output is compared with a dc voltage to register a reproducible point near the absorption line center on each scan. At each step

of the scan, the software checks to see if the line center has been reached. This approach allows the system to remain locked to the absorption feature while retaining the advantages of scanning the laser output across the whole absorption feature.

Data are accumulated in this manner for a period of for example 3 min, providing an average value of the 2*f* line profiles for the two species during that time. The raw data are analyzed by the program at the end of each averaging period. The analysis program first digitally smooths the data using a least squares algorithm,[75] which is very effective at removing high frequency noise without distorting the 2*f* lineshape if the smoothing window is correctly chosen. After smoothing of the ambient spectrum, either a smoothed background (zero air) spectrum or a least squares fitted low order polynomial, or both, is subtracted to suppress frequency-dependent background structures and/or residual etalon signals in the spectrum.

6.2.7.2 Linearity, Response Time, Sampling Integrity and Detection Limits To assess the relationship between the 2*f* signal and the mixing ratio, signals are measured as a function of formaldehyde mixing ratio determined by adding the output of the permeation device to various flows of clean air. No deviations from linearity are observed up to 50 ppbv.

The response time of the instrument toward changes in formaldehyde mixing ratios is investigated by setting the laser frequency to line center and recording the 2*f* peak signal amplitude as a function of time when calibration gas is suddenly added or removed from the flow at the air inlet. The results showed that the response time for a 90% change in the formaldehyde mixing ratio is a few seconds, corresponding to the residence time of the sampled air in the white cell. At these concentrations, polymerization of formaldehyde on surfaces does not occur and the monomer does not appear to interact strongly with the Teflon® surfaces of the sampling system.

The sampling integrity of the inlet system is confirmed by the observation that no differences in signal could be detected when calibration gas is introduced at the inlet of the sampling line (5 m of 6.25 mm o.d. Teflon® tubing) or directly into the white cell. Further tests compared the signals when a calibration spike is added to dry nitrogen carrier gas, to zero air or to ambient air streams. No effect on the throughput is detected. The laser is operated on individual lines in the 1740-cm^{-1} region of the formaldehyde absorption spectrum. The calibration method removes the need to assign the chosen line(s) spectroscopically or to measure their integrated line strengths. The methods used for determining the system detection limits and avoiding spectroscopic interferences are discussed by Slemr et al.[72] The minimum detectable mixing ratios are derived from reproducibility of the background spectra and low ppbv level calibration mixtures.

The detection limits achieved for formaldehyde are 0.25 ppbv using a 33.5 m optical pathlength and an averaging time of 3 min. The factors governing detection limits have been found to be a combination of random (high frequency) noise arising from the laser and the detector and slow changes (drifts) in the wavelength-dependent background structure. This structure arises because the laser power output varies when the driving current is modulated and swept across the absorption feature. Detection limits could be reduced to ~0.15 ppbv by co-addition of the 3 min spectra to form 30 min averages.

6.2.7.3 Measurement Accuracy The accuracy of the measurements is estimated to be ±20%. Determination of the output of the permeation device produces the largest source of error with the uncertainty in the colorimetric procedure assessed at ±10%. Additional uncertainties of ±2% are caused by temperature variations of the permeation device and of ±5% from the measurement of the air flow into which the calibration gas is introduced.

ACKNOWLEDGMENTS

The authors thank Puja Khare for involvement in preparing the manuscript. The help of U.C. Kulshrestha, S.P. Singh and Ranjit Kumar is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

CSFD	Continuous Scrubbing Fluorimetric Detection
CW	Continuous Wave
DMF	Dimethylformamide
DNPH	Dinitrophenylhydrazine
DOAS	Differential Optical Absorption Spectroscopy
ECD	Electron Capture Detection
FDH	Formaldehyde Dehydrogenase
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared Spectroscopy
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
IEC	Ion Exchange Chromatography
IR	Infrared
MBTH	3-Methyl-2-benzothiazolone Hydrazone
MDI	Methane Diphenyl Di-isocyanate

NAD	Nicotinamide Adenine Dinucleotide
NDIR	Nondispersive Infrared
NMHC	Non-methane Hydrocarbon
PEPH	Pentafluorophenyl Hydrazine
PETN	Pentaerythritol Tetranitrate
PFA	Polyfluoroalkene
POM	Polyoxymethylene
ppb	Parts Per Billion
PTFE	Poly(tetrafluoroethylene)
PVC	Poly(vinyl chloride)
RDX	Research Development Explosive
TCD	Thermal Conductivity Detection
TDLAS	Tunable Diode Laser Absorption Spectroscopy
UV	Ultraviolet

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Differential Optical Absorption Spectroscopy, Air Monitoring by • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Infrared Spectroscopy in Environmental Analysis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy

Infrared Spectroscopy **(Volume 12)**
Quantitative Analysis, Infrared

Liquid Chromatography **(Volume 13)**
Ion Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. T. Gilpin, E. Apel, A. Fried, B. Wert, J. Calvert, Z. Genfa, P. Dasgupta, J.W. Harder, B. Heikes, B. Hopkins, H. Westberg, T. Kleindienst, Y.N. Lee, X. Zhou, W. Lonneman, S. Sewell, 'Intercomparison of Six Ambient HCHO Measurement Techniques', *J. Geophys. Res.*, **102**(D17), 21 161–21 188 (1997).
2. *Committee on Aldehydes in Formaldehyde and Other Aldehydes*, National Academy Press, Washington, DC, 1981.
3. J.G. Calvert, 'The Homogeneous Chemistry of Formaldehyde Generation and Destruction Within the Atmosphere', in *Proceedings of the NATO Advanced Study Institute on Atmospheric Ozone*, ed. A.C. Aikin, US Department of Transport, FAA Office of Environ and Energy, High Altitude Program, Washington, DC, Report FAA-EE-80-20, 153–190, 1980.
4. G.A. Dawson, J.C. Farmer, 'Soluble Trace Gases in the South Western United States 2. Organic Species HCHO, HCOOH, CH_3COOH', *J. Geophys. Res.*, **93**, 5200–5206 (1988).
5. B.J. Finlayson-Pitts, J.N. Pitts, Jr, *Handbook of Atmospheric Chemistry: Fundamentals and Experimental Techniques*, John Wiley and Sons, Inc., Toronto, Canada, 1986.
6. Scott Research Laboratories, Inc., 'Atmospheric Reaction Studies in Los Angeles Basin', Phase I, US Department of Health, Education, and Welfare, Public Health Service, National Air Pollution Control Administration, Washington, DC, 142–149, 310–314, Vol. II, 1969.
7. E. Grosjean, E.L. Williams, II, D. Grosjean, 'Ambient Levels of Formaldehyde and Acetaldehyde in Atlanta, Georgia', *J. Air Waste Manage. Assoc.*, **43**, 469–474 (1993).
8. J.W. Harder, A. Fried, S. Sewell, B. Henry, 'Comparison of Tunable Diode Laser and Long Path Ultraviolet/Visible Spectroscopic Measurements of Ambient Formaldehyde Concentrations During the 1993 OH Photochemistry Experiment', *J. Geophys. Res.*, **102**, 6267–6287 (1993).
9. X. Zhou, Y.N. Lee, L. Newman, X. Chen, K. Mopper, 'Tropospheric Formaldehyde Concentration at Mauna Loa Observatory During the Mauna Loa Observatory Photochemical Experiment 2', *J. Geophys. Res.*, **101**, 14 711–14 719 (1996).

10. B.G. Heikes, 'Formaldehyde and Hydroperoxides at Mauna Loa Observatory', *J. Geophys. Res.*, **97**, 18 001–18 013 (1992).
11. 'Markets and Economics, Product Focus 'Formaldehyde'', *Chemical Week*, **8**, 30 (May 6, 1998).
12. C.S. Woo, S.E. Barry, S. Zaromb, 'Detection and Estimation of Part Per Billion Levels of Formaldehyde Using a Portable High-throughput Liquid Absorption Air Sampler', *Environ. Sci. Technol.*, **32**, 169–176 (1998).
13. P.R. Zimmerman, R. Chatfield, J. Fishman, P.J. Crutzen, P.L. Hanst, 'Estimates on the Production of CO and H_2 from the Oxidation of Hydrocarbon Emissions from Vegetation', *Geophys. Res. Lett.*, **5**, 679–682 (1978).
14. R.A. Duce, V.A. Mohnen, P.R. Zimmerman, D. Grosjean, W. Cautreels, R. Chatfield, R. Jaenicke, J.A. Ogren, E.D. Pellizzari, G.T. Wallace, 'Organic Material in Global Troposphere', *Rev. Geophys.*, **21**, 921–952 (1983).
15. J.S. Levine, *Water and the Photochemistry of the Troposphere, in Remote Sensing of a Cloudy Atmosphere*, ed. A.H. Sellers, Taylor and Francis, Philadelphia, 123–166, Chapter 4, 1984.
16. C. de Serves, 'Gas Phase Formaldehyde and Peroxide Measurements in the Arctic Atmosphere', *J. Geophys. Res.*, **99**, 25 391–25 398 (1994).
17. A.P. Altshuller, D.L. Miller, S.F. Sleva, 'Determination of Formaldehyde in Gas Mixture by the Chromotropic Acid Method', *Anal. Chem.*, **33**, 621–625 (1961).
18. R.R. Miksch, D.W. Anthon, L.Z. Fanning, C.D. Hollowell, K. Revzan, J. Glanville, 'Modified Pararosaniline Method for the Determination of Formaldehyde in Air', *Anal. Chem.*, **53**, 2118–2123 (1981).
19. R. Walters, 'Automated Determination of Formaldehyde in Air Without the Use of Tetrachloromercurate(II)', *Am. Ind. Hyg. Assoc. J.*, **44**, 659–661 (1983).
20. T. Dumas, 'Determination of Formaldehyde in Air by Gas Chromatography', *J. Chromatogr.*, **247**, 289–295 (1982).
21. K. Fung, D. Grosjean, 'Determination of Nanogram Amounts of Carbonyls as 2,4-Dinitrophenyl Hydrazones by HPLC', *Anal. Chem.*, **53**, 168–171 (1981).
22. Y. Hoshika, Y. Takata, 'Gas Chromatographic Separation of Carbonyl Compounds as their 2,4-Dinitrophenylhydrazones Using Glass Capillary Columns', *J. Chromatogr.*, **120**, 379–389 (1976).
23. U. Platt, D. Perner, H. Patz, 'Simultaneous Measurement of Atmospheric CH_2O, O_3 and NO_2 by Differential Optical Absorption', *J. Geophys. Res.*, **84**, 6329–6335 (1979).
24. E.C. Tuazon, R.A. Grahm, A.M. Winer, R.R. Easton, J.N. Pitts, Jr, P.L. Hanst, 'Kilometer-pathlength Fourier Transform Infrared System for the Study of Trace Pollutants in Ambient and Synthetic Atmospheres', *Atmos. Environ.*, **12**, 865–875 (1979).
25. G.W. Harris, G.I. Mackay, T. Iguchi, L.K. Mayne, H.I. Schiff, 'Measurements of Formaldehyde in the Troposphere by Tunable Diode Laser Absorption Spectroscopy', *J. Atmos. Chem.*, **8**, 119–137 (1989).
26. E. Sawicki, T.R. Hauser, T.W. Stanley, W. Elbert, 'The 3-Methyl 2-Benzothiazolone Hydrazone Test', *Anal. Chem.*, **33**(1), 93–96 (1961).
27. G.L. Mackay, D.R. Karecki, H.I. Schiff, 'Tunable Diode Laser Absorption Measurements of H_2O_2 and HCHO During the Mauna Loa Observatory Photochemistry Experiment', *J. Geophys. Res.*, **101**, 14 721–14 728 (1996).
28. A.L. Lazrus, K.L. Fong, J.A. Lind, 'Automated Fluorometric Determination of Formaldehyde in Air', *Anal. Chem.*, **60**, 1074–1078 (1988).
29. S.B. Tajada, 'Evaluation of Silica Gel Coated In Situ with Acidified 2,4-Dinitrophenylhydrazine for Sampling Aldehydes and Ketones in Air', *Int. J. Environ. Anal. Chem.*, **26**, 167–185 (1986).
30. Y.N. Lee, X. Zhou, 'Method for the Determination of Some Soluble Atmospheric Carbonyl Compounds', *Environ. Sci. Technol.*, **27**, 749–756 (1993).
31. A.M. Winer, R. Atkinson, J. Arey, H.W. Biermann, W.P. Harger, E.C. Tuazon, B. Zielinska, 'The Role of Nitrogeneous Pollutants in the Formation of Atmospheric Mutagens and Acid Deposition', Final Report, California Air Resources Board, Contract A4-081-32, 1987.
32. R. Atkinson, A.M. Winer, J. Arey, H.W. Biermann, T. Dinoff, W.P. Harger, P.A. McElroy, E.C. Tuazon, B. Zielinska, 'Measurements of NO_2, HONO, NO_3, CH_2O, PAN, Nitroarenes and Particulate Mutagenic Activities During the Carbonaceous Species Methods Comparison Study', Final Report, California Air Resources Board, Contract A5-150-32, 1988.
33. T.E. Kleindienst, P.B. Shepson, C.M. Nero, R.R. Arnts, S.B. Tejada, G.I. Mackay, L.K. Mayne, H.I. Schiff, J.A. Lind, G.L. Kok, A.L. Lazrus, P.K. Dasgupta, S. Dong, 'An Intercomparison of Formaldehyde Measurement Techniques at Ambient Concentration', *Atmos. Environ.*, **22**, 1931–1939 (1988).
34. D.R. Lawson, H.W. Biermann, E.C. Tuazon, A.M. Winer, G.I. Mackay, H.I. Schiff, G.L. Kok, P.K. Dasgupta, K. Fung, 'Formaldehyde Measurement Methods Evaluation and Ambient Concentrations During the Carbonaceous Species Methods Comparison Study', *Aerosol Sci. Technol.*, **12**, 64–76 (1990).
35. J.G. Reynolds, M. Irwin, 'Determination of CH_2O and Other Aldehydes', *Chem. Ind. (London)*, 419–424 (1948).
36. J.C. Septon, J.C. Ku, 'Workplace Air Sampling and Polarographic Determination of Formaldehyde', *Am. Ind. Hyg. Assoc. J.*, **43**(11), 845–852 (1982).
37. K.J. Bombaugh, W.C. Bull, 'Gas Chromatographic Determination of Formaldehyde in Solution and Purity Gas', *Anal. Chem.*, **34**, 1237–1241 (1962).

38. T. Yokouchi, T. Fujii, Y. Ambe, K. Fuwa, 'Gas Chromatographic Mass Spectrometric Analysis of Formaldehyde in Ambient Air Using a Sampling Tube', *J. Chromatogr.*, **180**, 133–138 (1979).
39. R.K. Beasley, C.E. Hoffmann, M.L. Rueppel, J.W. Worley, 'Sampling of Formaldehyde in Air with Coated Solid Sorbent and Determination by High Performance Liquid Chromatography', *Anal. Chem.*, **52**, 1110–1114 (1980).
40. W.S. Kim, Jr, C.L. Gerace, R.E. Kupel, 'Sampling and Analysis of Formaldehyde in the Industrial Atmosphere', US Department of Health, Education and Welfare, NIOSH, Cincinnati, OH 452, 26 Oct., 1978.
41. B. Coyne, R.E. Cook, J.R. Mann, S. Bouyoucos, O.I. McDonald, C.L. Baldwin, 'Formaldehyde: A Comparative Evaluation of Four Monitoring Methods', *Am. Ind. Hyg. Assoc. J.*, **46**(10), 609–619 (1985).
42. F.D. Snell, *Encyclopedia of Industrial Chemical Analysis*, Interscience Publishers, London, 100–118, Vol. 13, 1971.
43. J. Buchi, 'A Critical Review of the Methods for the Determination of Formaldehyde and a Rapid and Accurate Method for its Determination', *Pharm. Acta Helv.*, **6**, 1–54 (1931).
44. *Manual of Analytical Methods*, ed. D.G. Taylor, National Institute for Occupational Safety and Health Government Printing Office, Cincinnati, OH, DHHS/NIOSH Pub. No 77-157A, 1977.
45. *Methods of Air Sampling and Analysis*, 2nd edition, ed. M. Katz, American Public Health Association, Washington, DC, 1977.
46. H.H. Willard, L.L. Merritt, Jr, J.A. Dean, F.A. Sittle, Jr, *Instrumental Methods of Analysis*, Wadsworth Publishing Company, Belmont, CA, 1986.
47. R.M. Harrison, 'Secondary Pollutants', in *Handbook of Air Pollution Analysis*, eds. R.M. Harrison, R. Perry, Chapman and Hall, London, 279–634, 1986.
48. T. Nash, 'The Colorimetric Estimation of Formaldehyde by Means of the Hantzsch Reaction', *Biochem. J.*, **55**, 416–421 (1953).
49. P. Khare, G.S. Satsangi, N. Kumar, K.M. Kumari, S.S. Srivastava, 'Surface Measurement of Formaldehyde, Formic Acid and Acetic Acid at a Subtropical Semi-arid Site in India', *J. Geophys. Res.*, **102**(D15), 18 997–19 007 (1997).
50. P. Khare, G.S. Satsangi, N. Kumar, K.M. Kumari, S.S. Srivastava, 'HCHO, HCOOH and CH_3COOH in Air and Rain Water at a Rural Tropical Site in North Central India', *Atmos. Environ.*, **31**(25), 3867–3875 (1997).
51. J.D. Mclean, J.F. Holland, 'Development of a Portable Polarograph for Determination of Aldehydes in Automotive Exhaust and Production Plant Samples', *Environ. Sci. Technol.*, **9**, 127–131 (1975).
52. P.H. Rieger, *Electrochemistry*, Prentice-Hall International Editions, New Jersey, 233, 1987.
53. D.A. Skoog, D.M. West, *Fundamentals of Analytical Chemistry*, Saunders College Publishing, New York, 671, 684, 1982.
54. F.R. Kennedy, R.H. Hill, Jr, 'Determination of Formaldehyde in Air as an Oxazolidine Derivative by Capillary Gas Chromatography', *Anal. Chem.*, **54**, 1739–1742 (1982).
55. S. Selim, 'Separation and Quantitative Determination of Traces of Carbonyl Compounds as their 2,4-Dinitrophenylhydrazones by High-pressure Liquid Chromatography', *J. Chromatogr.*, **136**, 271–277 (1977).
56. K. Kobayashi, M. Tanaka, S. Kawai, T. Ohno, 'Gas Chromatographic Determination of Low-molecular Weight Carbonyl Compounds in Aqueous Solution as their Pentafluorophenylhydrazones', *J. Chromatogr.*, **176**, 118–122 (1979).
57. M. Igawa, J.W. Munger, M.R. Hoffmann, 'Analysis of Aldehydes in Cloud and Fog Water Samples by HPLC with a Post Column Reaction Detector', *Environ. Sci. Technol.*, **23**, 556–561 (1989).
58. *Handbook of Ion Chromatography*, ed. E.L. Johnson, Dionex Corporation, Sunnyvale, CA, 3, 1989.
59. B.C. Thompson, C.W. Keller, 'Infrared Spectra of Hazardous Gases and Vapors', Beckman Instruments Inc., Fullerton, CA, Industrial Technical Report TR-590, 1973.
60. M.V. Zeller, 'Infrared Methods in Air Analysis', Perkin-Elmer Corporation, Norwalk, CT, Part No. 993-9236, 1976.
61. I.L. Marr, M.S. Cresser, *Environmental Chemical Analysis*, International Text Book Co., New York, 1983.
62. G. Thyagrajan, 'Fourier Transform Infrared Spectroscopy', Indian Society of Analytical Scientists, Newsletter, **2**, September, 7–12, 1984.
63. L.P. Haack, D.L. Lacourse, T.J. Korniski, 'Comparison of Fourier Transform Infrared Spectrometry and 2,4-Dinitrophenylhydrazine Impinger Techniques for the Measurement of Formaldehyde in Vehicle Exhaust', *Anal. Chem.*, **58**, 68–72 (1986).
64. B. Heikes, et al., 'Formaldehyde Methods Comparison in the Remote Lower Troposphere During the Mauna Loa Photochemistry Experiment 2', *J. Geophys. Res.*, **101**, 14 741–14 755 (1996).
65. S. Dong, P.K. Dasgupta, 'Fast Fluorometric Flow Injection Analysis of Formaldehyde in Atmospheric Water', *Environ. Sci. Technol.*, **21**, 581–588 (1987).
66. E.D. Hinkley, 'High Resolution Infrared Spectroscopy with a Tunable Diode Laser', *Appl. Phys. Lett.*, **16**, 351 (1970).
67. A.S. Pine, 'High Resolution Methane ν_3-Band Spectra Using a Stabilized Tunable Difference-frequency Laser System', *J. Opt. Soc. Am.*, **66**, 97 (1976).
68. R.L. Byer, *Laser Spectroscopy*, eds. R.G. Brewer, A. Mooradin, Plenum, New York, 1974.
69. G.W. Harris, D. Klemp, T. Zenker, J.P. Burrows, B. Mathieu, 'Tunable Diode Laser Measurements of Trace Gases During the 1988 Polarstern Cruise and Intercomparisons with Other Methods', *J. Atmos. Chem.*, **15**, 315–326 (1992).
70. J. Reid, M. El-Sherbiny, B.K. Garside, E.A. Ballik, 'Sensitivity Limits of a Tunable Diode Laser Spectrophotometer

with Application to the Detection of NO_2 at the 100 pptv Level', *Appl. Optics.*, **19**, 3349–3353 (1980).
71. D.R. Hastie, G.I. Mackay, T. Iguchi, B.A. Ridley, H.I. Schiff, 'Tunable Diode Laser Systems for Measuring Trace Gases in Tropospheric Air', *Environ. Sci. Technol.*, **17**, 352A–364A (1983).
72. F. Slemr, G.W. Harris, D.R. Hastie, G.I. Mackay, H.I. Schiff, 'Measurement of Gas Phase Hydrogen Peroxide in Air by Tunable Diode Laser Spectroscopy', *J. Geophys. Res.*, **91**, 5371–5376 (1986).
73. H.I. Schiff, G.W. Harris, G.I. Mackay, 'Measurement of Atmospheric Gases by Light Absorption Spectrometry', American Chemical Society Monograph, **349**, 274–288, 1987.
74. 'Research Needs in Heterogeneous Tropospheric Chemistry', National Aeronautics and Space Administration, NASA Conference Publication 2320, 1984.
75. A. Savitzky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Anal. Chem.*, **36**, 1460–1464 (1964).

Gas Chromatography by Direct Aqueous Injection in Environmental Analysis

Marek Biziuk
Technical University of Gdańsk, Gdańsk, Poland

1 Introduction **1**
2 Classification of Techniques for the Isolation and Determination of Organic Compounds in Water **3**
3 Determination of Organic Compounds by Direct Aqueous Injection **4**
3.1 Determination of Organic Compounds by Direct Aqueous Injection into Packed Columns 5
3.2 Determination of Organic Compounds by Direct Aqueous Injection into a Capillary Column 7
3.3 Determination of Organic Compounds by Direct Large-volume Aqueous Injection Capillary Gas Chromatography 12
4 Application of Direct Aqueous Injection for Determination of Haloorganic Water Pollutants in the Gdańsk District **14**
4.1 Determination of Volatile Haloorganic Compounds in the Tap, Surface and Swimming Pool Waters from the Gdańsk District 14
4.2 Application of Direct Aqueous Injection for Determination of Semivolatile and Nonvolatile Haloorganic Water Pollutants 18
4.3 The Combination of Thin-layer Headspace Technique and Direct Aqueous Injection for Determination of Haloorganic Water Pollutants 18
5 Conclusions **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **21**

Sample pretreatment, the initial stages of an analytical procedure, becomes the essential stage in environmental analysis. The only method of determination of organic compounds in water which avoids the isolation and preconcentration step uses direct injection of an aqueous sample on to a gas chromatographic column. Direct aqueous injection (DAI) is the simplest method for the determination of organic water pollutants. This method is also very fast and easy to automate. The method avoids the problems associated with using solvents (toxic and expensive) and incomplete recovery of the analytes during their isolation from the aqueous phase, the effect of potential contaminants when using solvent or solid-phase extraction (SPE), and losses of the analytes during the enrichment step.

The idea of DAI on to a chromatographic column is very old and for liquid chromatography and ion chromatography was widely applied from the beginning of these techniques. In the beginning of the application of this technique in gas chromatography (GC), packed columns were used. The real issue for the development of GC and the application of DAI is connected with introducing to practical use a capillary column with a polar or nonpolar stationary thin liquid film or thin layer. The next, very important step was to introduce a direct injection oncolumn, which eliminates a lot of problems connected with injection in the split or splitless mode.

The determination of haloforms in water using DAI on to a capillary column and electron capture detection (ECD) was the most spectacular application of this method. The DAI/GC/ECD method was used for determination of haloorganic compounds in the different aqueous samples (tap, surface, swimming pool, and seawaters, beers, juices, soft drinks, and physiological liquids). This method was also used for the determination of semivolatile and nonvolatile haloorganic and other organic compounds in the different aqueous samples.

1 INTRODUCTION

Water, the substance so vital to humans and their well-being, the source and basis of life on earth, has recently become a serious danger to our health and life and to the life and health of flora and fauna. Furthermore, desertification of large areas of the world is now observed, and water, particularly clean water, is rapidly becoming a highly desirable and increasingly expensive commodity. Intensive industrialization of the world has resulted in an increased input of toxic inorganic and organic compounds to waters, drastically reducing the quality of surface waters which are a source of drinking water for a large part of the world population. Groundwater supply, which is basically nonrenewable, is diminishing and, in addition, anthropogenic groundwater pollution has become a fact of life. A majority of anthropogenic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

water pollutants are toxic compounds, dangerous not only to humans but also to animals and plants. Consequently, there is a need for continuous monitoring of the degree of pollution of potable and surface waters by anthropogenic inorganic and organic compounds. Excessive concentrations of pollutants, above maximum contaminant levels, not only warn consumers but also signal the existence of uncontrolled discharge of wastes, improperly operating treatment plant, lack of enforcement of legislation dealing with water management, or other violations of environmental laws. In order to be able to monitor environmental pollutants properly and effectively, analysts should have a variety of methods of isolation and determination of these contaminants, taking into account both physico-chemical properties of individual compounds and groups of compounds in addition to characteristics of matrices in which these compounds occur.

A majority of organic compounds in water are of natural origin. This is a large and diversified group of mostly unidentified compounds such as humic and fulvic acids, tannins, peptides, and amino acids. Their total concentration varies from 0.1 mg L^{-1} for groundwater, through 1–5 mg L^{-1} for most surface waters, to 20–25 mg L^{-1} for some waters with the highest concentrations of organic matter.[1,2] These compounds are predominantly nontoxic, but can be precursors of toxic compounds in the process of water treatment.

Anthropogenic organic compounds present the main hazard to the life and health of humans and flora and fauna. The number of known organic compounds is now estimated to be about 16 million, 2 million of which are compounds produced by synthesis alone. Every year, approximately 250 000 new compounds are synthesized and about 1000 of these are manufactured on an industrial scale. At presently, ca. 70 000 organic compounds are commercially available, with an annual global production of 100–200 million tons. Approximately one-third of all organic compounds produced end up in the environment, including water. Over 700 chemical compounds, including more than 600 organic compounds, many of which are biologically active, have been detected in drinking water samples.

Volatile organohalogen compounds and pesticides are particularly important pollutants among organic compounds as a result of their common use, persistence in the environment, and toxicity. They are primarily anthropogenic. Volatile organohalogen compounds are used mainly as solvents, cleaning and degreasing agents, blowing agents, polymerization modifiers, and heat-exchange fluids. As wastes, they find their way into lakes and rivers, and then into seas and oceans. Their concentrations in water and air are very variable and depend upon atmospheric conditions due to the washing by rain and evaporation from water during long periods of warm weather. It is estimated that the annual global production of organohalogen solvents alone amounts to several million tons.[3]

One of the most important sources of organohalogen compounds, particularly volatile ones, is water disinfection by chlorination.[1,3–10] The actual disinfecting agent is hypochlorous acid formed in the course of a disproportionation reaction taking place when chlorine dissolves in water. During chlorination, humic and fulvic compounds (so-called precursors), harmless and naturally occurring in water, are converted into organohalogen compounds which are dangerous to human life and health. The largest group of compounds formed during chlorination is the trihalomethane (THM) group. These are chloroform, the most abundant compound, bromodichloromethane, dibromochloromethane, and tribromomethane.[1,3–10] Organobromine compounds are formed when the water being chlorinated contains a large amount of bromides or when the chlorine used for disinfection is contaminated with bromine. Hypobromous acid formed in the reaction of bromide ions with hypochlorous acid reacts with an organic matrix about 200 times faster than does hypochlorous acid.[8] The amount and kind of organohalogen compounds formed depend upon the water pH, the amount of chlorine used and the content of organic matrix (total organic carbon (TOC)) in chlorinated water.[1,3–10]

THMs are not the only organohalogen compounds formed in the course of chlorination. Other volatile organochlorine compounds, such as tetrachloromethane, chloroethylene, 1,1-dichloroethylene, 1,1,2-trichloroethylene, tetrachloroethylene, 1,1,1-trichloroethane, and 1,2-dichloroethane, are also commonly found in chlorinated water. In addition, chlorination of humic substances and organic water pollutants yields a variety of other derivatives, of which over 100 have already been identified, including chlorinated acetone, chlorinated acetonitrile, chloropicrin, chloral, chloroacetic acids, chlorinated ethers, chlorophenols, and chlorinated ketones.[3–10] Koch and Krasner[10] estimated that among organohalogen compounds formed during the chlorination of water, 77% are THMs, 15% haloacetic acids, 3% halonitriles, 4% trichloroacetaldehyde hydrate and 1% the remaining compounds.

Pesticides are another group of organic compounds hazardous to human health and life.[1,11–13] They increase crop yields by reducing the amount of a crop that is lost to pests, and control diseases transmitted by insects. Pesticides are necessarily toxic, as they are used in agriculture, industry, and households to kill weeds (herbicides), insects (insecticides), fungi (fungicides), rodents (rodenticides), molluscs (molluscicides), mites (acaricides), roundworms (nematicides), aphids (aphicides), and eggs (ovicides). They are also used as

fumigants, attractants, and repellents.[1,11–13] The chemical structures of pesticides are very diversified, including organohalogen compounds, organophosphates, carbamates, triazines, and phenol derivatives.

2 CLASSIFICATION OF TECHNIQUES FOR THE ISOLATION AND DETERMINATION OF ORGANIC COMPOUNDS IN WATER

Anthropogenic organic compounds occur in water at relatively low concentration levels, but even at this level are dangerous to the life and health of humans and flora and fauna. This low concentration causes, in the majority of cases, the need for isolation and preconcentration of the analytes from the complex water matrix. Sample pretreatment, the initial stage of an analytical procedure, becomes the essential stage in environmental analysis. According to the participants in a survey,[14] among analytical laboratories the sample pretreatment takes up 61% of total time of analysis and is a source of 30% of the errors generated during sample analysis. The above data clearly indicate that sample pretreatment, including isolation and preconcentration of the analytes, is an essential step in trace analysis. The methods of isolation and preconcentration of organic compounds from water are closely associated with the kind of analytes, their volatility, polarity, stability, water solubility, solubility in organic solvents, etc. Many techniques for the isolation and preconcentration of the analytes from water have been developed.[4,5,11,15–23] The most popular are solvent or liquid–liquid extraction (LLE), SPE, solid-phase microextraction (SPME) and techniques utilizing the distribution of solute among the liquid and the gaseous phase (headspace (HS), purging). Less commonly used methods are supercritical fluid extraction (SFE), freezing out, lyophilization, vacuum distillation, steam distillation, and membrane processes (reverse osmosis, ultrafiltration, dialysis).[4,5,11,15–23] For final determination, GC with specific detection methods such as ECD, nitrogen–phosphorus detection (NPD) flame photometric detection (FPD), atomic emission detection (AED) and electrolytic conductivity detection (ELCD) or universal methods, such as flame ionization detection (FID) or mass spectrometry (MS) are used. Liquid chromatographic methods with detection methods such as ultraviolet (UV) detection, diode-array detection (DAD), electrochemical detection or fluorescence detection are also often used. The characteristics of typical detectors used in GC are listed in Table 1. A schematic diagram of utilization of various isolation techniques for the determination of organic compounds in water is shown in Figure 1.[17] In Figure 1 are shown two main ways for the determination of organic compounds in water: individual compound or total parameter determination. Total parameters such as TOC, dissolved organic carbon (DOC) and suspended organic carbon (SOC) are used to characterize the content of organic compounds in water. Other parameters are defined in terms of the method of isolation of an organic fraction from water: VOC or purgeable organic carbon (POC), EOC and AOC. However, total parameters measuring the carbon content in an organic fraction are not particularly suitable as an estimate of anthropogenic water pollutants and their hazard to human health, since a decisive majority of organic compounds in water are biogenic.

Two main methods for the isolation and preconcentration of organic contaminants from water are LLE and SPE but, although simple and not requiring sophisticated apparatus, they have a number of drawbacks. More volatile compounds can be evaporated during these operations (e.g. THMs, other volatile halocarbons, volatile hydrocarbons) with a consequent decrease in recovery and low results being obtained. Some compounds are poorly

Table 1 Characteristics of typical detection methods used in GC

Detection method	Type	Classification	Applicability	Detection limit (pg)	Linear dynamic range
FID		Universal	Hydrocarbons	100	10^6
ECD	Cl_2, Br	Selective	Haloorganic compounds	5	10^4
	Polyfluoride			0.05	
TCD, NPD	Nitrogen	Selective	Nitrogen and phosphorus	0.2	10^5
	Phosphorus		organic compounds	0.1	10^4
FPD	Phosphorus	Specific	Phosphorus and sulfur organic	0.5	5×10^4
	Sulfur		compounds	5	5×10^2
ELCD	F, Cl, Br	Specific	Halo- and nitroorganic	1	10^6
	N		compounds	1	10^4
MS	SCAN	Universal/specific	All compounds	100	10^3
	SIM			1	10^3

SCAN, scanning ion; SIM, selected ion monitoring; TCD, thermal conductivity detection.

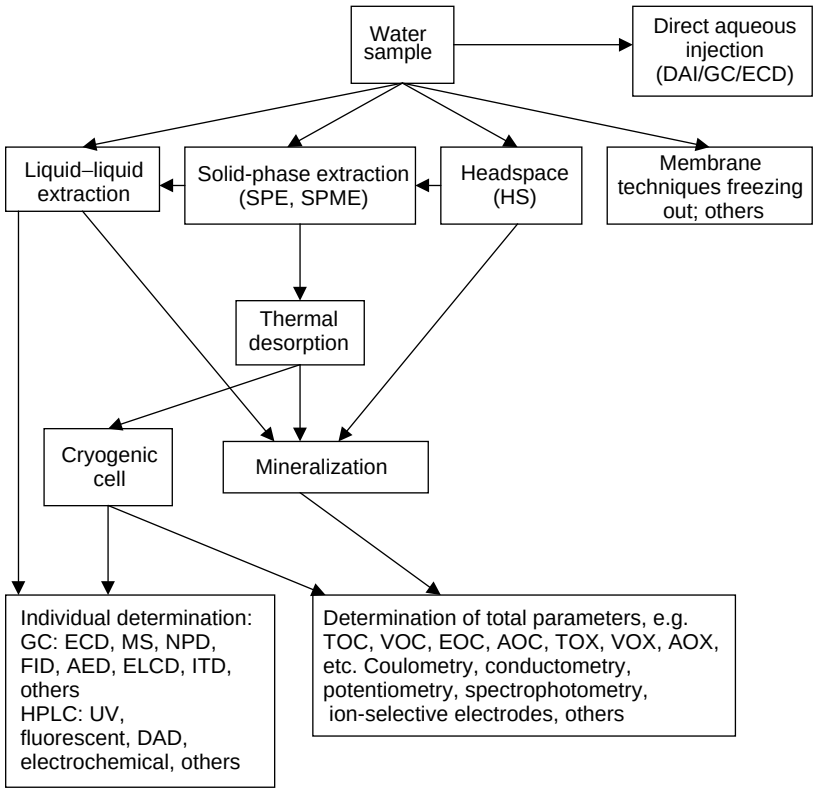

Figure 1 Classification of techniques for the isolation and determination of organic compounds in water. HPLC, high-performance liquid chromatography; TOC, total organic carbon; VOC, volatile organic carbon; EOC, extractable organic carbon; AOC, adsorbable organic carbon; TOX, total organic halogen; VOX, volatile organic halogen; AOX, adsorbable organic halogen.

extracted from the water phase (e.g. ketones, alcohols) or the partition coefficient is strongly dependent on the matrix (unknown). The LLE method requires the use of large volumes of expensive, and possibly toxic, solvents. After SPE (except methods using thermal desorption for volatile compounds), solvents are also used, although not in such large quantities. The solvents are subsequently evaporated and after analysis create a problem of storage of hazardous waste. Working with these solvents may require the use of personal protection devices, such as respirators. The formation of emulsions is yet another problem with LLE. The solvents and sorbents used for extraction should be of very high purity, free from traces of the analytes and other potential interferents. Otherwise, a large background is obtained, precluding the determination.

HS techniques[4,11,15–17] are limited to volatile compounds and the partition coefficient is also strongly dependent on the matrix (unknown) and affects the accuracy. Additionally, some compounds, belonging to priority pollutants, are not purgeable from the water phase (e.g. acrylonitrile, propionitrile).

The only method for the determination of organic compounds in water which avoids the isolation and preconcentration step uses direct injection of an aqueous sample on to a GC column.

3 DETERMINATION OF ORGANIC COMPOUNDS BY DIRECT AQUEOUS INJECTION

DAI[4,16,24–81] is the simplest method for avoiding the often very complicated steps of isolation and preconcentration of analytes. For analysts the best idea is to introduce the sample directly to the GC column without any pretreatment. This technique also avoids all problems related to the hot vaporizing injection technique by introduction of the sample directly into the oven-thermostated column inlet. However, the aqueous samples, particularly environmental, have a very complex nature of the matrix and a large content of inorganic salts and organic compounds (humic acids, other contaminants). These compounds can rest on the top of column and cause additional retention or deactivation of the column. Some low-volatility compounds can be eluted from the column during a long period of time. After direct introduction of aqueous samples, during evaporation, a

large volume of water vapor is formed. At 20 °C from 1 μL of water, 60 mL of carrier gas saturated with water vapor is produced. A large volume of water can deactivate the stationary phase and the layer and change the chromatographic condition of the column. Another problem is retention of water on the chromatographic column. This is, particularly in a packed column, a tedious and time-consuming process. Water also has a high surface tension (poor wetting properties) and poor properties concerning solvent effects.

Most of the detectors used in chromatographic determinations are not compatible with water. This concerns, especially, flame-based detectors or infrared (IR) or other spectroscopic detectors. The flame in an FID or thermionic NPD instrument can be blown out by water. On the other hand, in the case of an IR detector the measurement cell, often made from KBr, can be diluted in water. Relatively resistant to the presence of water are ECD and ion trap detection (ITD) instruments.

3.1 Determination of Organic Compounds by Direct Aqueous Injection into Packed Columns

The idea of DAI on to a chromatographic column is very old and for liquid and ion chromatography has been widely applied since the beginning of these techniques, but here water is used as one of the components of the mobile phase. A reversed-phase PRP-1 column, sodium form of Aminex A-5, and a Waters Pak cation-exchange column with a Dionex normal cation suppressor have been used. The application of this technique for GC is more difficult. In the beginning of the application of this technique in GC, packed columns were used.

One of the most popular column packings in the early days was graphitized carbon black (GCB). This sorbent was used by Di Corcia et al.[24,25] for determination of aliphatic amines (C_1–C_{16}) in aqueous solution. The application of GC for the determination of strong organic bases, such as amines, was always a difficult analytical problem, because of strong solid–gas interactions, the loss of sample, ghosting phenomena and badly tailed elution. It was very important to prepare the sorbent for packing of the column in a very special way. They had used modifications of Sterling FT-G (Carbopack A) and Vulcan GCB (Carbopack B) deactivated with suitable amounts of KOH and covered with polyethylene glycol (PEG), e.g. PEG-20M or PEG-1500. The GC apparatus used (Carlo Erba Model GI) equipped with FID allowed one to inject an aqueous sample (1.5 μL) directly into the column packing. Water was eluted after analytes. It was found[25] that the use of Carbopack B (60–80 mesh), modified with 0.3% potassium hydroxide and 4.8% PEG 20M, gave the best results with regard to the following: reducing the instability of the selectivity and the retention times of amines on prolonged use of the column, the intolerance of the packing to injections of water, baseline disturbances, and interferences caused by alcohols and peak tailing. The detection limit for this method was about 1 mg L^{-1}.

FID does not produce sufficient information for the identification of analyte compounds among the variety of different compounds that could be present in a water sample. Computerized gas chromatography/mass spectrometry (GC/MS) overcomes this difficulty. In 1974, Harris et al.[26] described a method for the direct analysis of water samples by GC/MS. A 1–10 μL sample of water sample was injected on-column into a Varian Model 1400 GC interfaced with a Finnigan Model 1015 C quadruple mass spectrometer. The different types of packed column used were as follows:

1. stainless-steel tube packed with Gas Chrom Q (60–80 mesh) coated with 5% Carbowax 20M;
2. coiled glass column (2 mm i.d.) packed with acid-washed Chromosorb W (60–80 mesh) coated with 10% free fatty acid phase (FFAP);
3. coiled glass column (2 mm i.d.) packed with Chromosorb 101 (50–60 mesh).

From the column with Gas Chrom Q (1), water was eluted much later than the analytes and so the ion source potential and multiplier were applied immediately after injection and switched off during elution of water. From the column with Chromosorb 101 (3), water was eluted very quickly, so the source and multiplier were not applied and data acquisition was not begun until the water had eluted.

This method can be applied to amines, ketones, alcohols, and aliphatic hydrocarbons, but only for domestic sewage and waste effluent water samples owing to very high detection limits (1–50 mg L^{-1}), depending on a variety of factors.

A much better concentration range (from 0.1 to 50 μg L^{-1}) was achieved by Fujii,[27] who introduced much larger samples (up to 100 μL). His method, also based on DAI and GC/MS analysis, was dedicated to haloorganic compounds present in tap water. Two columns were used: the precolumn contained Chromosorb W NAW (60–80 mesh) coated with 10% diglycerol and the main column packed with Chromosorb W AW DMCS (60–80 mesh) coated with 5% SE-30. The elution time of the overload water peak with the diglycerol precolumn was very long and all organohalogen compounds were eluted much earlier. This method can be applied to all compounds which elute before water, e.g. phenols, alcohols, ketones, hydrocarbons, and amines. The longer diglycerol precolumn should be selected in order to elute the higher boiling compounds before the water.

The analytes should not be eluted on the tail of the overload water peak. The main column makes it possible to achieve the required separation. The column temperature was maintained at 100 °C isothermally. On the outlet of the main column a three-way valve was positioned to prevent water (eluting after analytes) from entering the MS detector (possibility of damage to the filament or electron multiplier). High sensitivity and precision were achieved by operation in the mass fragmentography mode. Identification was supported not only by known retention times but also from the selectivity afforded by SIM. Detection limits depended upon the kind of compounds and varied from 0.1 μg L^{-1} for trichloromethane, dibromomethane, and dichloroethene to 0.8 μg L^{-1} for tetrachloromethane. The time of analysis for organohalogen compounds was about 6.5 min, but after analysis the column had to be maintained at 100 °C for at least 50 min to remove water and other less volatile compounds before the next analysis.

For the determination of haloorganic compounds, highly specific selective detection methods, such as ECD or ELCD (Hall detector) are more suitable. Nicholson et al.[28] described a method for the determination of haloforms in drinking water using DAI into a GC column (Varian 2400 GC equipped with ECD) packed with Chromosorb 101 (60–80 mesh). Also very interesting is the comparison of the DAI/GC/ECD method with the gas purging GC/ELCD method. Excellent agreement between the two methods was demonstrated. The gas purging method is very popular for the determination of volatile compounds in water but needs additional equipment.

In the DAI/GC/ECD method, 9 μL of aqueous sample volume was injected into the column operated under isothermal conditions (either 130 or 150 °C) and a 20-min analysis time was used. The column oven must be operated isothermally to prevent excessive baseline drift. The temperature depended on the range of compounds analyzed. This technique when automated and working with an autosampler allows analyses of up to 60 samples per day. The detection limit for this method for chloroform, bromodichloromethane, chlorodibromomethane, and carbon tetrachloride was at or below 1 μg L^{-1}. The authors observed that the results obtained by DAI/GC/ECD can be treated as the maximum haloform concentration which can be reached while the water is in the distribution system, because of reactions of the precursors and the other haloforms at high injector temperatures. The difference was attributed to nonpurgeable chlorinated compounds, formed as a result of the chlorination process, that supposedly decomposed in the heated injection port of the gas chromatograph to yield chloroform.

Similar investigations concerning the determination of THMs were also conducted by Pfaender et al.[29] For analysis by DAI they used a Perkin-Elmer Model 900 gas chromatograph equipped with two stainless-steel columns (column for analysis and column for makeup) packed with 50–80 mesh Porapak Q operating at 150 °C and with ECD. The novelty introduced by them in the construction of the apparatus (except the column for makeup) was a four-way valve installed in the manifold between the column outlet and the detector inlet. The valve permitted the effluent from the analytical column to be vented while the water peak was emerging and then redirected to the detector. During venting, the effluent from the makeup column was connected to the detector. It had been observed that injection of 1 μL of water into the electron capture detector caused no appreciable changes in its performance, but injection of larger quantities of water caused a decrease in detector sensitivity. Installation of the valve permitted a larger volume of water to be injected without a decrease in detector sensitivity. This made it possible to inject on to the column 20 μL of the sample and the detection limit reached for chloroform was 0.5 μg L^{-1}.

Another possibility for removing water during the determination of haloforms in water after DAI was described by Simonds and Kerns,[30] water being selectively removed by diffusion across a permselective membrane prior to the sample entering the chromatographic column. Nafion® (a perfluorosulfonic acid polymeric material made by DuPont) with strong adsorption properties for water, was used as a membrane. A volume of 1–20 μL of water was injected into a separate heated (150–200 °C) injection port, vaporized and the water vapor permeated through the walls of Nafion® tubing, positioned ahead of the chromatographic column, to a stream of dry nitrogen gas sweeping the outer surface of tubing. Any traces of organic compounds were directed to the chromatographic column and were chromatographed. Hydrocarbons, most aldehydes, esters, and ethers are not removed by the Nafion® tube, whereas highly polar compounds such as alcohols, amines, and certain ketones are partially or wholly removed with the water. Since water is removed before the sample enters the chromatographic column, any appropriate column can be used. Two separate types of apparatus were used during investigations. The first used during preliminary experiments was a Pye 104 (Pye Instruments) gas chromatograph equipped with aluminum tubing maintained isothermally at 60 °C and packed with Carbowax 400/Porasil C, Low K (Waters) and an electron capture detector (^{63}Ni ECD). The second apparatus used was a Varian 2100 gas chromatograph equipped with a glass column maintained isothermally at 75 °C and packed with 3% Carbowax 20M coated on Carbopack B (60–80 mesh) and an electron capture detector (^{3}H-Sc ECD). The

detection limits for haloorganic compounds varied from 0.05 $\mu g\,L^{-1}$ for carbon tetrachloride, 0.1 $\mu g\,L^{-1}$ for trichloroethylene and 1,1,1-trichloroethane, 0.3 $\mu g\,L^{-1}$ for chloroform, 35.0 $\mu g\,L^{-1}$ for 1,2-dichlorobenzene and 1,3-dichlorobenzene to 100.0 $\mu g\,L^{-1}$ for dichloromethane and 1500 $\mu g\,L^{-1}$ for chlorobenzene.

A comparison of the DAI method with a simple direct HS technique was also investigated. Both methods gave similar results.

Loung et al.[31] used DAI techniques to investigate drinking water in London. Water samples were collected in a glass vial equipped with a Mininert valve set into a screw cap which allowed the introduction of 5 µL of water sample by a syringe directly into the chromatographic column. This analysis was carried out using a Hewlett-Packard 5713 A gas chromatograph equipped with a glass column packed with Chromosorb 101 (100–120 mesh) operated isothermally at 150 °C and an electron capture detector. Detection limits for THMs under these conditions were about 1 $\mu g\,L^{-1}$.

Hou and Tang[32] used a special column for the rapid elution of water during determination of halocarbons in drinking water. Volumes of 0.4–1 µL of water sample were introduced directly on to the GC column (Hewlett-Packard 5890) packed with GDX-103 (a Chinese sorbent similar to Porapak) coated with 1% of SE-30. The final determinations were carried out with an electron capture detector (^{63}Ni ECD). The detection limits varied from 0.4 $\mu g\,L^{-1}$ for tetrachloromethane, ~1 $\mu g\,L^{-1}$ for chloroform, bromochloromethane and dibromochloromethane to 2.6 $\mu g\,L^{-1}$ for tribromomethane.

DAI was proposed also in the United States Environmental Protection Agency (USEPA) standard method (Method 8010),[33] but only in very limited applications (e.g. aqueous process wastes). In this method first of all the purge-and-trap (P&T) technique had been used for the determination of 39 volatile haloorganic compounds in water. The detection limit for the P&T method varied from 0.02 $\mu g\,L^{-1}$ for 1,1,2-trichloroethane through 0.05 $\mu g\,L^{-1}$ for chloroform up to 0.52 $\mu g\,L^{-1}$ for chloroethane, whereas for DAI it was very approximately 10^4 $\mu g\,L^{-1}$. There are two columns which can be used in this method, a stainless-steel column packed with Carbopack-B (60–80 mesh) coated with 1% SP-1000 or a stainless-steel column packed with chemically bonded *n*-octane on Porasil-C (100–120 mesh). Detection was achieved with a halogen-specific detector.

DAI has also been applied to the determination of other compounds. Fussell and McCalley[34] described a method for the determination of volatile fatty acids (C_2–C_5) and lactic acid in aqueous samples of silage based on direct injection of samples (1–2 µL) on to a glass GC column (Pye Unicam) which was packed with GCB coated with Carbowax 20M. This packing gave improved results compared with porous polymer column materials (Porapak Q, Porapak QS, Chromosorb 101). The final determination was carried out using FID. This method was more selective, more precise, and less time-consuming than a classical distillation procedure. The method gave an excellent linear calibration response for volatile fatty acids from 0 to 500 mg L^{-1} and for lactic acids from 250 to 10 000 mg L^{-1}. The detection limit varied from 2 mg L^{-1} for fatty acids to 60 mg L^{-1} for lactic acid. To protect the column, the quartz-wool insert positioned on the top of the column should be renewed regularly if the samples contain non-volatile compounds. Also, the first few centimeters of packing can be replaced.

Malaiyandi and Goddard[35] used DAI for the determination of amines in steam and boiling industrial waters. The addition of amines to boiling industrial water prevents the corrosion caused by CO_2. However, amines are strongly irritant to the skin, mucous membranes, and respiratory tract, so it is necessary to control their content. Amines are highly hydrophilic, so it is very difficult to isolate them from the aqueous sample without loss of analytes. Additionally, the partition coefficient for amines during LLE is not favorable, so DAI is the best approach for the determination of amines in water. The method was employed for the determination of morpholine, cyclohexylamine, and diethylaminoethanol. Amine analysis under the usual GC conditions suffers from several drawbacks: adsorption or even decomposition in the column, bad tailing and ghosting. In the described method, from 2 to 5 µL of aqueous sample were injected directly on to the GC column packed with GCB (60–80 mesh) coated with 4.8% PEG 20M and 0.4% potassium hydroxide. Temperature programming of the column and FID for final determination were used. To prolong the life of the column packing, a guard insert was placed before the column. This guard insert containing a wad of silanized and ammonia solution-treated glass wool could be removed and reused after cleaning. It protected the column from the accumulation of alkali metal carbonates and hydrogen-carbonates and nonvolatile organic compounds on the top of the column, which can plug the column. The detection limits for morpholine, cyclohexylamine, and diethylaminoethanol varied from 52 to 55 ng for one injection.

3.2 Determination of Organic Compounds by Direct Aqueous Injection into a Capillary Column

3.2.1 Capillary Columns in Gas Chromatography

The real issue for the development of GC and the application of DAI is connected with introducing to practical use a capillary column with a polar or non-polar stationary thin-liquid film or thin layer. Different

groups of researchers have dealt with these problems and their applications. The initial work connected with the application of a capillary column for GC was performed by Desty,[36] Schomburg et al.[37,38] Novotny et al.[39,40] and van Rijswick and Tesařik.[41] This method was improved and developed by Grob et al.[42–49]

Initially there were many problems with the preparation of columns, specifically with their repeatability and reproducibility, thermal stability, and fragility. For a long time capillary columns were not widely used. However, capillary columns have many advantages.[37–49] Compared with packed columns they have a much higher separation efficiency in terms of effective plate numbers per meter and the total effective plate numbers that can be achieved in long columns. Also, tailing is much lower. It is now possible without any problems to prepare or to buy capillary columns which can provide high separation efficiency, defined polarity, and temperature stability. Low-volatility compounds can be eluted without decomposition at low temperatures with moderate retention times, because only small amounts of stationary liquids are contained in long columns. Capillary columns also have very low pressure drops, so long columns and high carrier gas flow-rates can be used.

In the studies by Novotny et al.[39,40] and Rijswick and Tesařik,[41] special attention was focused on the role of selective surface monolayers in wettability, the deactivation procedure, and the deposition of stationary phases on the capillary column, which are the most important parameters for column efficiency.

Schomburg et al.[37,38] made and etched a glass capillary column with dry hydrogen chloride and coated the etched surface with a stationary liquid phase (OV-101, LAC-3R-728, Carbowax 20M). The tailing of polar compounds can be reduced by coating the surfaces with a thin film of deactivating agent. Carbowaxes were found to be the best for this purpose. For the determination of aqueous samples, two procedures were used and double-column chromatography was proposed. The capillary columns must be protected against a large volume of water (highly polar solvent), which can change the separation efficiency by destroying the thin-film properties. Water can be trapped in the first column (precolumn) packed with a strongly polar stationary phase, such as Carbowax, and less polar analytes eluted before the water can be chromatographed on the main capillary column. After analysis, water can be back-flushed from the precolumn. The second procedure is to elute water, without passing through the main capillary column before the analytes, using a nonpolar Porapak precolumn, which traps analytes in the very first part. After elution of water, analytes can be back-flushed from the precolumn and with or without trapping (in the mantling tube made of stainless-steel and cooled or heated with a stream of nitrogen) directed to the main capillary column. There are two separate injection ports for each method. In this way, only small, negligible amounts of water are introduced to the capillary column. The other solution is to use two capillary columns,[14] the first as the trap and the second for the main separation.

3.2.2 Injection On-column in Gas Chromatography

The next, very important step was to introduce direct injection on-column. Conventional injection had been carried out via an evaporation step: split or splitless injection, the falling needle, or a precolumn. During evaporation, the sample is diluted in carrier gas, causing band broadening and a degradation of the detection limits. All vaporizing injectors possess various dead volumes from which the last traces of water vapor are eluted with a delay. Analytes can be absorbed in the injector, or the high temperature during evaporation can destroy some compounds, causing problems with quantitative determinations. Another problem is connected with the septum used on the top of the injection port. The septum can absorb analytes or emit some volatile components, causing tailing, artifacts, ghost peaks, and baseline drift. The syringe needle can introduce to the injector some trace of organic contaminants and in cutting the material of the septum it can add contaminants of rubber or plastic. The latter can plug the injector or the top of the column. This effect is especially dangerous for a capillary column. On-column injection eliminates many of these problems.

3.2.3 Injection On-capillary Column in Gas Chromatography

As far as we know, the first application of direct injection on a large-bore steel capillary column (0.03–0.06 mm i.d.) was described by Zlatkis and Walker.[50] The volume of sample introduced was up to 0.3 μL and the efficiency deteriorated rapidly for larger samples. In the next paper, written by Desty,[36] the introduction of the sample as a layer a few tenths of a micrometer thick on a few millimeters of wire of small enough diameter to enter the capillary column was proposed. Analytes were dissolved in the stationary phase on the top of the column at low temperature without decomposing it.

Application of on-capillary column injection for aqueous samples was realized by Grob et al.[42–49] Their work was fundamental for the development of the DAI technique. There were many problems to resolve for on-column injection: temperature, pressure, and flow conditions of the injector, kind of syringes used, design and functioning of the on-column injector, and the type of capillary columns and detectors. A syringe is the simplest and safest way to introduce a sample into a capillary column. The volume of sample introduced is well defined,

all the sample volume is introduced to the column and the syringe is easily cleaned. The outer diameter of the syringe needle should be at least 0.05 mm less than the inner diameter of the column. The needle, introduced into the column over a length of several centimeters, should not plug the column and the flow conditions of carrier gas. Some problems are caused by partial sample vaporization out of the hot syringe needle and the difficulty with stabilization of the pressure of the carrier gas at the column inlet during injection, because we have to work without a septum. The sample starts to evaporate at the injection point after deposition on the column wall (after leaving the outlet of the needle in the capillary column), not in the syringe needle (including its tip), so the injector has to be cooled. The injection point has to be sufficiently far from the oven wall inside the oven to reach the controlled oven temperature, but not so far as to prevent evaporation of the sample in the needle. It depends upon the design of the oven, but a 10-mm average distance from the wall of oven to the end of the needle is advised by Grob and Grob.[42] The temperature of injection should be such as to prevent explosive evaporation with a sudden pressure increase, which can flush back part of the sample into the cool injector. The vapor pressure of the water should be below the carrier gas inlet pressure. On-column injection produces a much lower relative standard deviation for all analytes than the conventional injection procedure and no breakdown products were observed.

3.2.4 *Determination of Haloforms in Water Using Direct Aqueous Injection on to Capillary Columns*

The determination of haloforms in water using DAI on to a capillary column and ECD was the most spectacular application proposed by Grob et al.[43–49] An apolar, extremely inert column was applied to elute water before the volatile haloorganic compounds. Water should be eluted even before much more volatile compounds such as methylene chloride, but there is a need for strong retention of the apolar halocarbons without broadening of the water peak. Non-volatile compounds and inorganic salts are left in the inlet of column. Water is eluted for a longer time during the analysis of natural samples containing inorganic salts than for model samples prepared from deionized water (the dry inorganic salts retain traces of water). The inlet of the column can be periodically broken off or the precolumn consisting of an empty, 1–1.5-m long fused, deactivated (usually silylated) but uncoated silica capillary column should be applied. The precolumn, which retains nonvolatile compounds, should after some period be shortened or replaced. A capillary column coated with an extremely thick film (about 5 μm) of methylsilicone phase (SE-30) or a slightly more polar, phenyl-containing phase (SE-54) was applied. This film was more suitable than the thick films commonly used (1–2 μm). This thick film causes even very volatile organic compounds to have a relatively long retention time, but it also causes a decrease in column efficiency. For the analysis of high-boiling organic compounds a thinner film of nonpolar stationary phase is recommended.

The column length depends on the compounds to be analyzed, but longer columns result in a decrease in sensitivity. An increase in retention time for halocarbons can be achieved by increasing the column length or the film thickness. Volumes of 0.5–2.0 μL of the aqueous samples were injected on to a capillary column (Carlo Erba Model 4 160 gas chromatograph) operating isothermally (103–104 °C) and connected to an electron capture detector (Carlo Erba ^{63}Ni ECD) heated to 350 °C. Cooler columns can accumulate some residual water, causing baseline drift and poorer resolution of high-volatility halocarbons. At higher column temperatures, water vapor can damage the column and cause peak splitting. The optimal length of the column was established as 25–30 m. On-column injection produced much sharper halocarbon peaks than splitless injection. During on-column injection, analytes are evaporated from the water inside the column, whereas during splitless injection analytes are diluted with the carrier gas in the injector, forming a broad initial band and causing peak broadening and even distortion. The detection limits obtained varied from 0.025 μg L^{-1} for tetrachloroethylene, through 0.03 μg L^{-1} for chloroform, up to 0.6 μg L^{-1} for trichloroethane. For the analysis of heavier halocarbons (e.g. chlorobenzenes), a capillary column coated with a thicker film (0.8 μm) is recommended.

3.2.5 *Commercially Available Methods for the Determination of Haloforms in Water Using Direct Aqueous Injection on to Capillary Columns*

Application sheets for the determination of halocarbons in drinking water using DAI on to a capillary column are available for Carlo Erba and Hewlett-Packard gas chromatographs equipped with ECD.[51–53] These are elaborations on the basic work of Grob et al.[42–49]

In the case of the Carlo Erba application sheet,[53] cold on-column injection with secondary cooling, a fused-silica capillary column (FSCC) (30 m × 0.32 mm i.d.) PS 255 (coated with a 5-μm film) and a gas chromatograph [Carlo Erba 4160 high-resolution gas chromatography (HRGC) system] equipped with an electron capture detector (ECD 40/400) were used. The injection volume of the water sample was 2 μL. The column was operated isothermally at 104 °C. The injection temperature was higher than the boiling point of water, so cold-on-column injection with secondary cooling was necessary to prevent sample back-flushing.

To prevent contamination of the analytical column by nonvolatile organic and inorganic compounds, an easily replaceable, uncoated and deactivated precolumn placed before the analytical column was used as a retention gap. The precolumn was connected with the analytical column by a zero dead volume butt connector. The other recommended chromatographic conditions are: detector ECD, operated at 350 °C in the constant-current mode; makeup gas, nitrogen (50 mL min^{-1}); carrier gas, hydrogen (2 mL min^{-1}). The use of an autosampler (AS 550) for routine application was also advised. The dynamic filling method (using pressurized inert gas) utilized by the automatic sampler avoids gas bubble formation in the syringe, which is responsible for the loss of volatiles and irreproducible injection. Detection limits for chosen compounds were as follows: 0.6 μg L^{-1} for methylene chloride, 0.03 μg L^{-1} for 1,1,1-trichloroethane, 0.02 μg L^{-1} for chloroform, 0.015 μg L^{-1} for tetrachloromethane, 0.03 μg L^{-1} for trichloroethylene, and 0.025 μg L^{-1} for tetrachloroethylene.

Temmerman et al.[51] described a similar method, but with some changes of the chromatographic conditions. They also used a cold on-column injector with secondary cooling and a Carlo Erba 5300 HRGC gas chromatograph equipped with ECD operating at 350 °C. In this method, 1 μL of water sample was injected on-column and used a 1.5 m × 0.53 mm i.d. uncoated and undeactivated FSCC as a retention gap. This was connected by a press-fit connector to a 30 m × 0.53 mm i.d. FSCC coated with a 5-μm film of methylsilicone (DB-1) operating isothermally at 104 °C. The undeactivated retention gap slightly improved the elution of the water peak. As the carrier gas, helium at a flow-rate of 1–2 mL min^{-1} was used, and, as the makeup gas, Ar–CH_4 at a flow-rate of 60 mL min^{-1} was used. In comparison with solvent extraction (with pentane), the DAI method gave better recoveries.

Temmerman et al. elaborated on the basic work of Hewlett-Packard Application Note 228-135[52,54,55] which was applied as USEPA Method 502.1–3.[55] This method used cool on-column injection, an HP 7673A automatic sampler, an HP 5890 Series II gas chromatograph equipped with a retention gap (1.5 m × 0.53 mm i.d. uncoated, undeactivated FSCC), a 30 m × 0.53 mm i.d. capillary column coated with a 2.65-μm film of HP-1 and an electron capture detector heated at 300 °C. The other chromatographic conditions were as follows: injection of 1 μL of water sample at 50 °C; oven temperature program, 50–150 °C at 10 °C min^{-1}; detector make-up gas, Ar–CH_4 at a flow-rate of 60 mL min^{-1}; carrier gas, hydrogen at an inlet pressure of 10 kPa. The detection limit was 1 μg L^{-1} for chloroform and 0.5 μg L^{-1} for other THMs. The linearity for THM determination was very good up to a concentration of 500 μg L^{-1}.

Hewlett-Packard later introduced some changes.[54] They noticed that the retention gap should have a polar or medium-polarity surface and so should be deactivated with PEG or with a medium-polarity reagent, not with dimethylsilicones or an apolar silylating reagent. The result was that they recommended as the retention gap a 5 m × 0.53 mm i.d. deactivated capillary column, and as the analytical column a 30 m × 0.53 mm i.d. capillary column covered with a 1-μm film of HP-1.

A similar solution for the determination of THMs by DAI using ECD was elaborated by Carmichael and Holmes[56] for Varian Instrument Business Manufacturing Facilities.[57,58] A septum-equipped programmable detector (special construction) with an on-column glass insert was used for cold on-column injection (without secondary cooling), an uncoated deactivated fused-silica column (4 m × 0.53 mm i.d.) as the retention gap and a 0.53 to 0.32 mm press-fit connector, an analytical column (30 m × 0.32 mm i.d.) coated with a film of DB-624 cyanopropyl stationary phase (1.8 μm) and a sheath foil electron capture detector (^{63}Ni ECD) were used to determine THMs. After injection of 1–5 μL of the sample (the injector was held at 60 °C for 0.04 min and then heated to 150 °C at 75 °C min^{-1}), the column was held for 3 min at 104 °C and then heated to 150 °C at 10 °C min^{-1}. The carrier gas used was helium (0.2 m s^{-1} or 10 psi head pressure) and the makeup gas used was nitrogen (50 mL min^{-1} for a Varian 3400 and 25 mL min^{-1} for a Varian 3500 GC). The DB-624 analytical column gives effective resolution of haloforms and has a low retention of water and hypochloric acid. The retention gap should be replaced after a few analyses or solvent-washed. The detection limit was lowered to 0.043 μg L^{-1} for chloroform, 0.016 μg L^{-1} for bromodichloromethane, 0.017 μg L^{-1} for dibromochloromethane and 0.060 μg L^{-1} for bromoform when a smaller cell volume in ECD was applied. This method was introduced on to the market with success and is used in many laboratories.[56–59]

3.2.6 Determination of Other Organic Compounds in Water Using Direct Aqueous Injection on to Capillary Columns

DAI on to a capillary column was successfully introduced for the determination of volatile organohalogen compounds in water. The success was due to the very high sensitivity of the electron capture detector routinely used in this method. With other compounds, which are not so volatile and do not contain halogen, it is more difficult. Ellington and Trusty[60] described such a method for the determination of trimethyl and triethyl phosphates in water. The widespread use of these compounds as plasticizers, pesticides, and flame retardants causes

their release into the environment. These compounds are very soluble in water, so their isolation using LLE or SPE is very difficult. They used a Hewlett-Packard 5890 gas chromatograph equipped with an automated cool on-column aqueous injector, with a retention gap of 1–1.5 m × 0.53 mm i.d. uncoated, deactivated FSCC. The analytical column used was a 15 m × 0.53 mm i.d. DB-5 column coated with a 1.0-μm film or a 13 m × 0.55 mm i.d. DB-Wax column coated with a 1.0-μm film. FID at 250 °C was used. The column temperature was held initially at 90 °C for 1 min, then programmed to 115 °C at 5 °C min^{-1} and to 160 °C at 70 °C min^{-1} and held at 160 °C for 2 min. The carrier gas used was helium at a flow-rate of 10–12 mL min^{-1}, and nitrogen acted as the makeup gas at a flow-rate of 25 mL min^{-1}. The flow-rates of hydrogen and air were 32 and 330 mL min^{-1}, respectively. The injection volume was 1 μL. The samples were buffered at pH 3 for phthalate and pH 7 or 11 for phosphate. The main problem encountered is the buildup of the buffer salts which necessitates rinsing of the column or periodic replacement of the retention gap. Good linearity and precision in determination were obtained.

A method for the determination of volatile fatty acids using on-column injection of an aqueous solution was described by McCalley.[61] A Hewlett-Packard 5890 gas chromatograph was used with on-column injection (0.5 μL of analyte water or aqueous solution of oxalic acid, 0.03 mol L^{-1}). The analytical column was either a 10 m × 0.53 mm i.d. fused-silica column containing cross-linked FFAP, film thickness 1.0 μm, or a 15 m × 0.55 mm i.d. DB-Wax column coated with a 1.0 μm film. A single flame ionization detector operating at 200 °C was used. An electronically stored blank chromatogram was used to eliminate the effect of stationary phase bleeding during the temperature program. The carrier gas was helium at a flow-rate of 31 cm s^{-1}.

DAI with GC coupled to Fourier transform infrared (FTIR) spectroscopy was used by Payne and Collette[62] for identification of bis(2-chloroethyl) ether hydrolysis products: 2-(2-chloroethoxy)ethanol (2CEE), diethylene glycol (DEG), *p*-dioxane (*p*DOX), 2-chloroethanol (CE) and ethylene glycol (EG). They also used a Hewlett-Packard 5890 gas chromatograph equipped with a cool on-column injection system, comprising a 30 m DB-Wax (a PEG phase), 1.0-μm Megabore™ column and FTIR with a Digilab FTS-60 spectrometer. This was equipped with a mercury–cadmium–telluride detector using a light pipe-based interface operated at 220 °C. The FTIR detector was connected in series with a thermal conductivity detection (TCD) instrument. The column temperature was programmed from 80 to 220 °C at 15 °C min^{-1} and held at 220 °C for 10 min. Helium was used as the carrier gas. Between the column outlet and FTIR interface, a Valco six-port valve was installed. During water elution from the column (before the analytes on this column), the outlet of the column was connected to a TCD instrument and, during elution of analytes, to the FTIR interface. The injection volume was 1 μL.

Gurka et al.[63] applied DAI coupled with GC/FTIR or GC/ion trap spectrometry (ITS) for the determination of a wide range (13) of organic compounds, which were water soluble, low-molecular-weight, and poorly or not purgeable efficiently from water. These included alcohols, ketones, nitriles, and esters. They used for this purpose a Hewlett-Packard Model 5965 A IR detector (containing a narrow-band mercury–cadmium–telluride detector) and a Finnigan Model ITS 40 (equipped with a septum programmable injector (SPI) and scanned from 29 to 300 u). Both detectors were interfaced to a gas chromatograph equipped with an on-column injector. In the case of GC/FTIR analysis, a Hewlett-Packard Model 5890 gas chromatograph equipped with a fused-silica capillary column (FSCC) (27.5 m × 0.32 mm i.d.), coated with a 10-μm Pora Plot Q film, was used. After 5 min at 60 °C, the GC column was heated to 200 °C at 10 °C min^{-1} and held at 200 °C for 8 min. The light pipe and transfer lines were also heated to 220 °C. As the carrier gas, helium was used at a flow-rate of 33 cm s^{-1}. No makeup gas was used. Helium was dried by passing it through a cartridge drier. The detection limits varied from 5 to 100 mg L^{-1} so this method is not useful for routine analysis of drinking water, but can be used for industrial process streams and effluents.

In the case of GC/ITS analysis, the Varian Model 3400 gas chromatograph equipped with an FSCC (1.5 m × 0.32 mm i.d.) coated with a 0.5-μm polysiloxane film with 5% phenyl groups (PTE-5) and a 30 cm × 0.05 mm i.d. fused-silica restrictor connected between the analytical column and the SPI to limit the flow of carrier gas to the ITS 40 manifold (it should be less than 1.5 mL min^{-1}) was used. After 1 min at 60 °C, the GC column was heated to 240 °C at 16 °C min^{-1}. As the carrier gas, helium at a flow-rate of 45 cm s^{-1} was used. The injection volume was 2 μL. The detection limits varied from 4 to 200 μg L^{-1}.

The GC/ITS method was improved by Pyle and Gurka[64] for the determination of 27 volatile organic compounds. They had noticed that injections greater than 0.2 μL provided imprecise results caused by residual water eluted together with the analytes, so 0.2 μL was routinely injected on-column. Using similar equipment to before[63] with a Restec column (30 m × 0.53 mm i.d. XTI-5 Megabore™ FSCC coated with a 1.5-μm film of bonded 5% phenyl and 95% dimethyl polysiloxane liquid phase), they changed the working conditions of the detector, specifying the rf voltage of the ion trap to exclude water ions from the manifold. Single-ion and triple-ion quantitation methods were tested. Better precision

and detection limits (about twofold) for the triple-ion method were obtained. The detection limits varied from 1.56 μg L^{-1} for ethylbenzene, through 2.35 μg L^{-1} for chlorobenzene, 3.19 μg L^{-1} for benzene and 4.83 μg L^{-1} for trichloromethane, up to 20.37 μg L^{-1} for 2-chloroethyl vinyl ether. The total time of analysis was 12 min. In comparison with the P&T method, DAI/GC/ITS is faster, simpler to apply, and has a large dynamic range (from 20 μg L^{-1} to 20 mg L^{-1}). DAI also shows possibilities for the determination of higher boiling semivolatile compounds.

3.3 Determination of Organic Compounds by Direct Large-volume Aqueous Injection Capillary Gas Chromatography

The typical injection volume for GC is about 1 μL of sample and such a volume was used in the previously described methods. This volume is too small for DAI (without any preconcentration steps) to reach required detection limits for most of the organic pollutants of water dictated by European Union (EU) regulations. The main solution for direct water analysis is to inject much larger volumes of samples, despite water being a very difficult solvent for GC, as was previously described. Water affects the peak shape and resolution, can deactivate analytical columns and can damage the detector, causing a reduction in detection limits. To avoid these difficulties, water should be released after injection from the GC system via an early water vapor exit, before the chromatographic column and detector. The analytes are trapped on some sorbent or liquid layer inside the chromatographic system. This is the basic idea of on-line enrichment GC analysis after direct large-volume injection of aqueous samples.[65]

Zlatkis et al.[66] described a method for the quantitative determination of the trace organic compounds (2-octanone, 5-nonanone, 2-decanone, 2-undecanone, benzaldehyde, and lindane) in aqueous solutions at parts per billion (micrograms per liter) and parts per trillion (nanograms per liter) levels. Up to 400 μL of water sample was injected directly on to a bonded hydrophobic phase capillary column connected to a postcolumn [50 cm × 0.32 mm i.d. empty fused-silica open-tubular capillary column (FSOT)]. The postcolumn was not connected to the detector. The analytes were adsorbed in the main column while water was released by the open exit of the postcolumn. After water had been removed from the main column and the postcolumn, the main column was connected to the detector and adsorbed analytes were thermally desorbed, trapped in the cold trap, and analyzed by GC using the same column operating in reverse or one in series. Less than 0.1 μL of water was trapped in the cold trap. For the final determination FID or ECD (for lindane) was used.

In the case of GC/FID analysis, the Hewlett-Packard gas chromatograph used (HP 5830A) was equipped with a 50 m × 0.32 mm i.d., 0.5-μm film, bonded-phase methylsilica FSOT column, a 50 cm × 0.32 mm i.d. empty FSOT postcolumn and a cold trap (50 cm × 0.37 mm i.d. empty stainless-steel tubing in liquid nitrogen) and FID was used. Injection was performed at room temperature. After removing water (about 15 s), the main column was connected to the cold trap and heated at 200 °C for 20 min, then the oven was cooled, the outlet of the cold trap was connected to the injector and the column was reversed (the end of the column formerly connected to the injector was connected to the detector). Once the cooling from the cold trap ceased, the analysis started. The column was held isothermal for 2 min and then heated at 2 °C min^{-1}. Helium was used as the carrier gas.

In the case of GC/ECD analysis (for lindane) a Varian 3700 GC gas chromatograph equipped with an 11 m × 0.32 mm i.d., 1-μL film DB-5 (equivalent to SE-54) FSOT column as the adsorbing column, a postcolumn and a cold trap as before and a 50 m × 0.32 mm i.d., 0.52-μm film, cross-linked 5% phenyl–methylsilicone FSOT column as the analytical column. ECD was used. Injection was performed at room temperature. After removing water (about 15 s), the adsorbing column was connected to the cold trap and heated at 200 °C for 20 min, then the cold trap was disconnected from the adsorbing column and connected to the analytical column. Once cooling of the cold trap ceased, the analysis started. The temperature of the analytical column was programmed from 50 °C (initial temperature) to 240 °C at 5 °C min^{-1}. Helium was used as the carrier gas and argon with 5% methane (30 mL min^{-1}) as the makeup gas. Two columns were used to decrease the time for ECD stabilization.

Both methods required considerable manual transfer operations and were therefore laborious and time-consuming, but represented the first feasible method based on direct injection of large amounts of aqueous samples on to a capillary column. These ideas were used by other workers to realize on-line enrichment of analytes during direct analysis of large water samples.

Gerhardt and Cortes[67] improved the method for the determination of pesticides using large-volume DAI capillary GC for the determination of chlorpyrifos in well water. A Hewlett-Packard 5890 gas chromatograph was used, equipped with a 10-port valve, and a 20 m × 0.25 mm i.d. nondeactivated FSCC as the inlet (retention gap) connected via a low-dead-volume union with the analytical column (30 m × 0.25 mm i.d., 5% phenyl–methylsilicone capillary with a 0.25-μm film thickness). ECD was used. A 20-μL volume of water sample was injected via the

10-port injection valve equipped with a stainless-steel external sample loop on to the uncoated inlet. The inlet allowed complete solvent vaporization prior to reaching the analytical column. All water passed the analytical column and detector, but collecting and processing of the detector signals started after passing all water vapor. The column was heated for 15 min at 130 °C then increased to 280 °C at 20 °C min^{-1}. Helium was used as the carrier gas (6.4 mL min^{-1} at 130 °C) and argon–methane (10 : 90) at a flow-rate 32 mL min^{-1} as the makeup gas. The detection limit for chlorpyrifos was 0.22 μg L^{-1} with a linearity range from 0.9 to 18 μg L^{-1} and with high accuracy and precision. They noted that relatively large volumes of water injected into the capillary column do not affect the peak shape and resolution, even after 150 injections (3 mL of water).

Much higher volumes of water can be introduced using splitless injection with vapor overflow, as proposed by Grob et al.[68,69] for the determination of triazines (atrazine, prometryn, propazine, simazine, terbutylazine, desisopropylatrazine, and desethylatrazine). Water vapor was removed by a vapor outlet beyond the analytical column. A Carlo Erba gas chromatograph (Model 4160) was equipped with an injector containing an insert with a vaporizing chamber (80 mm × 5 mm i.d.) packed with Tenax™ TA (20–35 mesh), a 2 m × 0.32 mm i.d. precolumn coated with a 0.6-μm film of an apolar stationary phase (PEG–Superox 0.6), a 40 cm × 0.32 mm i.d. raw fused-silica column as the vapor outlet, and a 15 m × 0.32 mm i.d. FSCC coated with a 0.4-μm film of Superox and 5% Irganox 1010, and alkali flame ionization detection (AFID)/NPD was used. Up to 400 μL of water sample was injected using a syringe to the injector insert heated at 310 °C. Water vapor and volatile compounds were removed from the system by the precolumn and the open vapor outlet while the compound of interest remained in the precolumn. The temperature of the column during injection was 130 °C. After 2 min the vapor outlet was closed and the released precolumn gas was directed to the analytical column and detector. The temperature program of the column (from 130 to 240 °C at 8 °C min^{-1}) was started 5 min after injection. Hydrogen was used as the carrier gas. The detection limit for triazines was about 0.5 μg L^{-1}, so it was higher than required for routine analysis, where EU standards dictate 0.1 μg L^{-1} as the maximum acceptable concentration in drinking water.

The idea of direct injection of large amounts of water sample in an adsorbent-packed programmed temperature vaporization (PTV) injector was developed by Engewald et al.[70–73] Water was removed through the split vent while analytes were trapped on an adsorbent inside the PTV injector. Adsorbents used for such a purpose should have a large breakthrough volume for the analytes and a very low one for water, and they should also be a mechanically and thermally stable and should easily release analytes after heating. They tested Tenax™ TA, Tenax™ GR, Chemipack C18 and GCB from the point of view of breakthrough volumes for pesticides and nitroaromatic compounds. Tenax™ TA was chosen as the best sorbent with regard to the breakthrough volumes for the chosen analytes. They used a Hewlett-Packard HP 5890 Series II gas chromatograph equipped with a PTV injector (cold injection system), a glass insert packed with a 2-cm layer of Tenax™ TA (4.2 mg), a 2 m × 0.32 mm i.d. retention gap, a 25 m × 0.32 mm i.d. column with a 0.2-μm film of CP Sil 19 CB as the analytical column and ECD or NPD. ECD was used for determination of organochlorine and NPD for nitrogen- or phosphorus-containing pesticides and nitrotoluenes. Up to 500 μL or of aqueous sample were injected into the injection insert at 50 °C with a split flow of 600 mL min^{-1}. Water was removed from the insert over a 40-min period as the analytes remained on it. Elimination of water can be controlled by means of a thermal conductivity cell placed after the split vent. Desorption was realized at 280 °C for 5 min (splitless) and the analyte compounds were focused by 'cold trapping' at a column temperature of 40 °C. The column was then heated to 270 °C at 10 °C min^{-1}. The detection limits varied from 0.01 μg L^{-1} (dieldrin with ECD) to 0.5 μg L^{-1} (aldimorph with NPD). Even with extremely water-sensitive NPD, no sensitivity loss was observed after numerous analyses.

This method was applied to the determination of target and non-target pollutants in water from the Elbe River,[71] using for the final determination an MS detector in the SIM mode (for target analysis) or SCAN mode (for non-target analysis). A volume of 1 mL or more of aqueous sample was slowly injected (CIS 3 injector) into the insert with sorbent. It is very important in this technique to control the rate of sample injection.[72] Water should be completely evaporated in front of the adsorbent bed in the insert to ensure quantitative enrichment of the analytes from the gas phase. Very important also were parameters such as temperature and structure of the insert and the kind and flow-rate of the carrier gas. A simple back-flush device was developed and installed between the injector and the analytical column to avoid possible damage to the column or detector by water vapor. The detection limit of approximately 0.01 μg L^{-1} was comparable to that with conventional isolation and preconcentration techniques such as SPE or LLE connected with selective detection (ECD or NPD).

A small change to the chromatographic conditions was used by the same authors for the determination of 12 triazines and their decomposition products in water from the Elbe River.[73] A 500-μL volume of water sample was injected using the CSI 3 cooled injection system directly

into the Tenax™-packed insert of a PTV injector heated at 50 °C with an injection speed of 10 μL min^{-1} and a split flow of 600 mL min^{-1} (including 15 mL min^{-1} backflushing). Sample introduction lasted 50 min and then the split vent was closed and the analytes thermally desorbed in the splitless mode to the analytical column. For the chromatographic determination a Hewlett-Packard 5890 Series II gas chromatograph equipped with an SPB-5 column (30 m × 0.25 mm i.d. with a 0.1-μm film thickness) and helium as the carrier gas with an inlet pressure of 1.7 bar was used, with NPD. The injector was heated at 12 °C min^{-1} to 300 °C and the column was maintained at an initial temperature of 70 °C for 3.5 min, then heated to 150 °C at 15 °C min^{-1}, to 175 °C at 3 °C min^{-1}, to 280 °C at 20 °C min^{-1} and held for 2 min at 280 °C. After analysis of samples containing large amounts of nonvolatile organic and inorganic compounds, the insert must be changed frequently. This method can also be used for the determination of polychlorinated biphenyls and polycyclic aromatic hydrocarbons at parts per trillion (nanograms per liter) levels.

4 APPLICATION OF DIRECT AQUEOUS INJECTION FOR DETERMINATION OF HALOORGANIC WATER POLLUTANTS IN THE GDAŃSK DISTRICT

DAI on to capillary chromatographic columns connected with a selective electron capture detector was used for the determination of haloorganic compounds in different aqueous samples (tap, surface, swimming pool, and seawater, beers, juices, soft drinks, and physiological liquids).

4.1 Determination of Volatile Haloorganic Compounds in the Tap, Surface and Swimming Pool Waters from the Gdańsk District

The presence of volatile organohalogen compounds in tap and swimming pool water mainly originates from disinfection of the water, most often accomplished by chlorination. Even when other methods are used, such as ozonation, UV irradiation, or application of hydrogen peroxide, chlorination is necessary for water preservation in the water-mains system. About 70% of Gdańsk water is supplied from underground water intakes but 30% comes from treated water. The latter is surface water intake from Straszyn. This water, taken from the Radunia River, is treated by ozonation and at the outlet to the water-mains system by chlorination (to protect it from secondary bacteriological pollution). The water in swimming pools is also treated by chlorination. Chlorination causes very toxic, mutagenic, and carcinogenic volatile haloorganic compounds at relatively high levels (up to 100 μg L^{-1} and more) to be present in tap and swimming pool water. The World Health Organization (WHO), the EU and most individual countries, including Poland, have introduced a maximum admissible concentration (MAC) for some organohalogen compounds present in the tap water. There are first of all THMs (trichloromethane, bromodichloromethane, dibromochloromethane, tribromomethane), but also tetrachloromethane, 1,2,2-trichloroethene, 1,1-dichloroethene, 1,1,2,2-tetrachloroethene, 1,2-dichloroethane and 1,1,1-trichloroethane. The MAC varied from 0.3 μg L^{-1} for 1,1-dichloroethene (WHO, Norway) to 100 μg L^{-1} (EU, WHO, USEPA, UK) for total trihalomethanes (TTHMs) and 350 μg L^{-1} for TTHMs (Canada). Typical concentrations of trichloromethane (chloroform), the compound most frequently present in the water treated by chlorination, varied from 1 to 30 μg L^{-1}; 30 μg L^{-1} is the MAC in the directives of the WHO, Poland and UK for chloroform in tap water. DAI on to a capillary GC column with ECD is a suitable method for the determination of organohalogen compounds in this range of concentrations. This method requires a special injector allowing cold on-column injection and special capillary columns. A nonvaporizing, septumless, cold on-column injector with secondary cooling (Grob type) fulfilled nearly all requirements for the good introduction of the sample on to the chromatographic column.

The construction of such an injector, produced by Carlo Erba, is shown in Figure 2.[17] The aqueous sample (1–2 μL) was introduced directly in to the oven-thermostatted capillary GC column inlet using a syringe with an outer diameter of the needle smaller then the inner diameter of the column. Secondary air cooling of the inlet of the column prevented sample evaporation in the syringe needle (including its tip). The evaporation started after deposition of the sample on the column wall. The sample was mechanically transferred instead of being evaporated, so high-boiling components did not remain in the syringe needle. The secondary cooling was switched on only before and during the sample injection. The primary cooling of the injector body worked permanently. The secondary cooling permitted rapid cooling of the bottom part of the injector attached to the oven, preventing discrimination of compounds at the syringe tip. It also allowed the introduction of a larger volume of sample, which permitted a decrease in the detection limits. The sample was injected at an injector temperature lower, than the solvent boiling point, but at an oven temperature higher than the solvent boiling point.

Nonvolatile inorganic and organic compounds can gradually accumulate at the column inlet which will

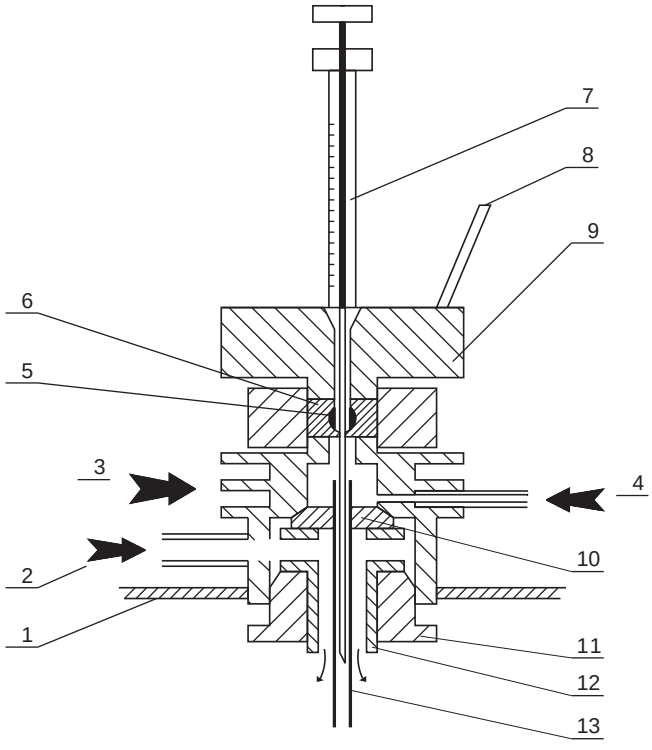

Figure 2 Schematic diagram of nonvaporizing septumless, cold on-column injector: 1 = oven wall; 2 = secondary air cooling; 3 = principal air cooling system; 4 = carrier gas inlet; 5 = stainless-steel rotating valve; 6 = valve seal; 7 = syringe; 8 = valve lever; 9 = upper part of injector; 10 = column seal; 11 = bottom part of injector; 12 = cooling jacket; 13 = capillary column.

make the resolving power of the column deteriorate. To prevent this, readily exchanged, deactivated, noncoated precolumns (2 m × 0.32 mm i.d.) connected to the analytical column through a zero-dead-volume fitting were used. Following a specified number of analyses, the precolumn can be completely exchanged or part of it can be broken off. The precolumns play another important role: they serve as a so-called retention trap, narrowing the chromatographic bands of the analytes. The sample volume during fast injection spreads in the column inlet, flooding it and causing 'band broadening in space', so it is necessary to reconcentrate the broad inlet band. The band broadening in space depends on the temperature, the speed of injection, and the volume injected. In the precolumn all compounds present in the gaseous phase migrate with the same speed owing to the lack of any interaction. In the main column coated with a stationary film of nonpolar phase, the analytes, interacting with the stationary phase, are focused and water, which does not react with the stationary phase, is washed out freely. To prevent retention of water the precolumn and the analytical column should be deactivated. The film of the stationary phase should also be immobilized by cross-linking.

During the chromatographic process, water was eluted before the analytes owing to the use of an inert column coated with a thick layer (ca. 5 μm) of nonpolar stationary phase which resulted in rapid and complete elution of water and sufficient retention of the analyte organohalogen compounds. Methylsilicones (DB-1) were used as nonpolar stationary phases. To improve separation, methylsilicone stationary phases modified with vinyl groups (PS-255, 1% vinyl groups) were applied. Capillary columns 30 m long with i.d. ca. 0.5 mm were used. Longer columns provided improved resolution but at the expense of sensitivity. Isothermal runs at about 102–105 °C were employed. As a result of the high sensitivity of the ECD instrument employed, the stationary phase in the analytical column was immobilized through cross-linking and chemical binding to the column wall.

The electron capture detector was operated at 350 °C, which minimizes the residence time of water in the detector, thus improving the sensitivity of the detector. For the identification of the analytes and for quantitative analysis of other organic compounds, GC/MS was used. For the determination of volatile organohalogen compounds (trichloromethane, bromodichloromethane, dibromochloromethane, tribromomethane, tetrachloromethane, trichloroethene, tetrachloroethene, dichloromethane, 1,1,1-trichloroethane, 1,1,2,2-tetrachloroethane), DAI into a capillary column and ECD were used.[75–81] A Carlo Erba Vega 6180 gas chromatograph was applied with ECD (ECD 40/400). The chromatographic conditions were as follows: a 30 m × 0.32 mm i.d. FSCC, coated with bonded 5-μm apolar DB-1 phase (J&W Scientific) or slightly more polar PS-255 (with 1% vinyl groups); a 2 m × 0.32 mm i.d. fused-silica precolumn; temperature program, 103 °C isothermal; injection system, cold on-column with secondary cooling (20 s before injection); detection, ECD at 350 °C with pure nitrogen (99.999%) as the makeup gas (30 mL min^{-1}); carrier gas, hydrogen at 0.4 m s^{-1}; and injection volume, 2 μL. The retention times, standard deviation ($P = 95\%$; $n = 7$, concentration about 20 μg L^{-1}), and detection limits of volatile organohalogen compounds for the applied columns DB-1 and PS-255 are listed in Table 2. The detection limits were calculated as three times the baseline random noise level.

The DAI/ECD technique has been successfully used for the determination of volatile organohalogen compounds containing one or two carbon atoms (the products of chlorination of humic substances; THMs) in tap water,[75,76] swimming pool water,[77] and surface water.[76,78–81] Typical results of analysis are listed in Tables 3, 4 and 5. An example of a chromatogram of volatile organohalogen compounds obtained using this method is presented in

Table 2 Retention times, standard deviations ($P = 95\%$; $n = 7$) and detection limits of volatile organohalogen compounds for different columns

Compound	Retention time (min)		Relative standard deviation (%)		Detection limit ($\mu g\,L^{-1}$)
	DB-1 column	PS-255 column	DB-1 column	PS-255 column	
Dichloromethane	3.08	3.39	5.6	4.2	1
Trichloromethane	3.40	4.57	2.4	1.7	0.01
Tetrachloromethane	4.45	5.88	2.2	1.8	0.002
1,1,1-Trichloroethane	5.20	–	4.0	–	0.03
Trichloroethene	6.45	6.47	2.9	2.2	0.03
Bromodichloromethane	6.45	6.68	2.9	2.4	0.015
1,1,2,2-Tetrachloroethane	8.59	–	5.8	–	0.03
Dibromochloromethane	9.93	10.94	3.2	3.2	0.03
Tetrachloroethene	11.05	12.04	3.8	4.2	0.015
Tribromomethane	15.97	16.71	3.6	3.7	0.1

Table 3 Determination of volatile organohalogen compounds ($\mu g\,L^{-1}$) in tap water from the Gdańsk district (DB-1 column)

Sampling site	Sampling date	$CHCl_3$	$CHBrCl_2$ + C_2HCl_3	$CHBr_2Cl$	TTHM	CCl_4	CH_2Cl_2	C_2Cl_4
Gdańsk Wrzeszcz (tap water)[a]	6.05.93	16.8	8.1	1.2	26.1	–	–	–
	01.03.95	8.3	4.7	1.6	14.6	0.7	0.8	n.d.[d]
	30.03.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Gdańsk Przymorze (tap water)[b]	6.05.93	0.5	1.4	n.d.	1.9	–	–	–
	01.03.95	2.3	1.8	0.6	4.7	n.d.	0.1	0.1
	31.03.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Gdańsk Suchanino (tap water)[c]	6.05.93	32.0	10.6	0.8	43.4	–	–	–
	01.03.95	9.0	5.6	2.0	16.6	0.7	0.4	0.1
	30.03.99	0.03	n.d.	0.6	0.6	0.08	0.03	n.d.
Gdańsk Morena (tap water)[c]	6.05.93	23.0	5.9	0.4	29.3	–	–	–
	01.03.95	11.1	6.5	2.3	19.9	1.1	0.5	n.d.
	30.03.99	n.d.	n.d.	0.5	0.5	0.08	n.d.	n.d.
Gdańsk Żabianka (tap water)[b]	25.06.93	0.5	12.7	n.d.	13.2	n.d.	n.d.	–
	20.03.95	0.9	4.0	n.d.	4.9	n.d.	n.d.	0.3
	30.03.99	n.d.	n.d.	n.d.	n.d.	0.2	n.d.	n.d.
Straszyn (water intake)[c]	18.03.93	24.0	10.1	n.d.	34.1	n.d.	–	–
	02.03.95	8.4	5.7	2.4	16.5	1.0	0.3	0.1
	31.03.99	n.d.	n.d.	0.5	0.5	0.02	n.d.	n.d.
Pregowo (water intake)[c]	2.12.93	2.3	2.2	1.3	5.8	n.d.	–	–
	02.03.95	3.1	2.1	0.8	6.0	n.d.	0.2	0.1
	31.03.99	n.d.	n.d.	0.08	0.08	0.17	n.d.	n.d.
Lipce (water intake)[b]	20.07.93	0.07	0.03	n.d.	0.1	n.d.	0.76	0.84
	02.03.95	0.9	0.2	0.5	1.6	n.d.	n.d.	0.1
	31.03.99	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

[a] Mixed chlorinated and unchlorinated water.
[b] Unchlorinated water.
[c] Chlorinated water.
[d] n.d., not detected. Tribromomethane, 1,1,1-trichloroethane, 1,2-dichloroethane and 1,1-dichloroethene were not detected at any of the sampling sites.

Figure 3. The changes in the chloroform concentration in the tap water of Gdańsk during the period 1993–97 are shown in Figure 4. In Figure 4, data from three site points are presented: Straszyn, outlet of water intake (surface water treated by ozonation, filtration with a carbon filter, and final chlorination); Morena, district of Gdańsk supplied by water intake from Straszyn; and Przymorze, district of Gdańsk supplied by underground water intake.

It is clearly evident, from the results of the determination of organohalogen compounds in the tap

Table 4 Determination of volatile organohalogen compounds ($\mu g\,L^{-1}$) in swimming pool water from the Gdańsk district (PS-255 column)

Sampling date	$CHCl_3$	C_2HCl_3	$CHBrCl_2$	$CHBr_2Cl$	CCl_4	CH_2Cl_2	CH_3CCl_3
14.03.91	92.4	23.3	11.5	n.d.[a]	6.9	27.0	n.d.
23.09.91	47.3	15.2	n.d.	n.d.	6.5	–	2.6
19.12.91	52.8	2.4	2.8	0.8	5.7	–	13.3

[a] n.d., not detected, Tribromomethane, 1,2-dichloroethane and 1,1-dichloroethene were not detected.

Table 5 Determination of volatile organohalogen compounds ($\mu g\,L^{-1}$) in surface water from the Gdańsk district (DB-1 column)

Sampling site	Sampling date	$CHCl_3$	$CHBrCl_2$ + C_2HCl_3	$CHBr_2Cl$	$CHBr_3$	CCl_4	C_2Cl_4
Gdynia Orlowo (seawater)	7.03.94	2.2	1.44	0.55	n.d.[a]	0.11	n.d.
	24.11.94	0.6	0.49	0.22	n.d.	0.02	n.d.
Kacza River	7.03.94	1.28	0.86	0.36	n.d.	0.01	n.d.
	21.11.94	0.74	n.d.	n.d.	0.2	0.04	n.d.
Vistula River	7.03.94	1.44	0.08	0.14	n.d.	0.03	0.03
	21.11.94	0.17	0.03	n.d.	n.d.	n.d.	0.01
Motława River	7.03.94	0.26	0.02	n.d.	n.d.	0.03	0.02
	21.11.94	0.04	0.02	n.d.	n.d.	0.01	n.d.
Rozwójka River	7.03.94	0.65	0.06	n.d.	n.d.	0.02	0.02
	21.11.94	0.04	0.02	n.d.	n.d.	n.d.	n.d.

[a] n.d., not detected. Dichloromethane, 1,1,1-trichloroethane, 1,2-dichloroethane and 1,1-dichloroethene were not detected at any of the sampling sites.

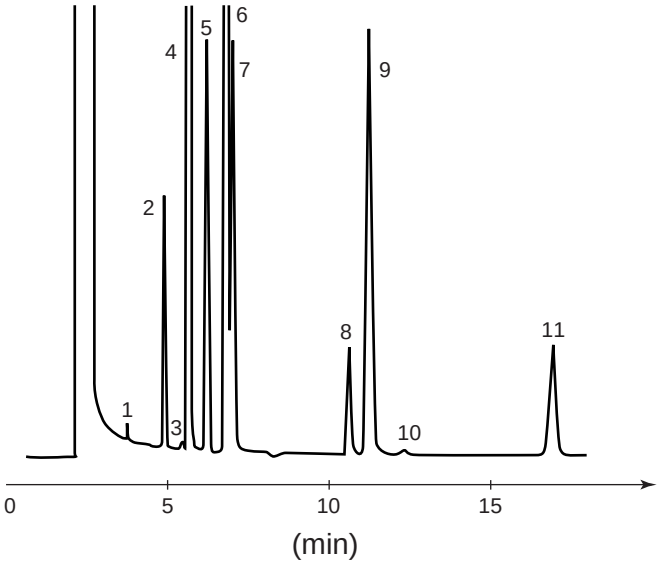

Figure 3 Typical chromatogram of volatile organohalogen compounds (DAI/GC/ECD, isothermal). Peaks: 1 = dichloromethane; 2 = trichloromethane; 3 = 1,2-dichloroethane; 4 = tetrachloromethane; 5 = 1,1,1-trichloroethane; 6 = bromodichloromethane; 7 = trichloroethene; 8 = 1,1,2,2-tetrachloroethane; 9 = dibromochloromethane; 10 = tetrachloroethene; 11 = tribromomethane.

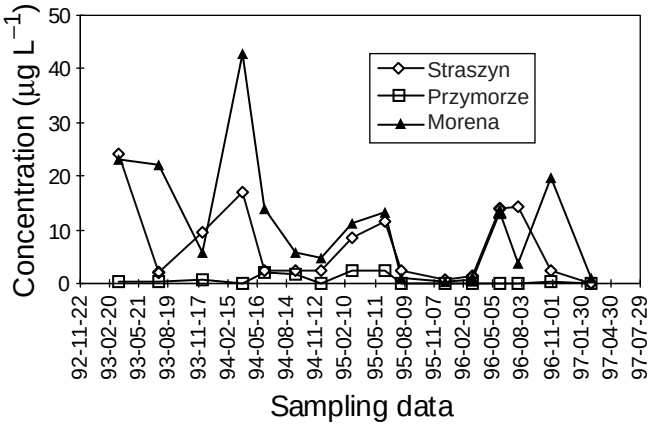

Figure 4 Changes in chloroform concentration in the tap water from different sampling sites in the Gdańsk area during the period 1993–97.

water, which of the districts of Gdańsk were supplied with water treated by chlorination. Only in 1993, in some cases (Gdańsk Suchanino), did the content of chloroform exceed the Polish and WHO standards. At the same time the concentrations of THMs in all districts supplied by surface water intake from Straszyn were relatively high. After changing the water treatment in this water intake from chlorination to ozonation and introducing water filtration with a carbon filter, these concentrations decreased significantly.

The investigations on tap and surface waters were carried out using a DB-1 chromatographic column, on which, unfortunately, bromodichloromethane and trichloroethene were determined as the sum. Bromodichloromethane is a typical product of humic acid chlorination, whereas trichloroethene originates from industrial sewage. It was confirmed by GC/MS that in the tap water in the Gdańsk Żabianka district trichloroethene was present at relatively high concentration (but not exceeding any standards). A similar high concentration of trichloroethene were found in the underground water intake in Sopot, from which this district of Gdańsk was supplied. Unfortunately, some underground water intakes are polluted by industrial sewage, in this case probably from laundries. Trichloroethene was found also in the Kacza River, flowing across the densely populated district of Gdynia, and in the seawater at Gdynia Orłowo, near the mouth of the Kacza River.

The highest concentration of volatile haloorganic compounds was found in the water from the swimming pool (swimming pool of the Technical University of Gdańsk in Gdańsk Wrzeszcz). This water is intensely chlorinated for sanitary purposes. Here a wide range of haloorganic compounds was found: not only THMs, but also dichloromethane, tetrachloromethane, 1,1,1-trichloroethane, etc. The occurrence of bromoorganic compounds was due to the fact that the raw water contained small amounts of bromides. During chlorination, free bromine is formed, which reacts with organic compounds to form organobromine compounds.

The DAI/GC/ECD technique proved to be a good and very useful method for the determination of volatile haloorganic compounds in tap and swimming pool water owing to its simplicity (no isolation and preconcentration steps are necessary), repeatability, reduction of the possibility of sample contamination by the solvent or sorbent used, reduction of possible losses of volatile compounds during additional steps, and low detection limits (0.002–0.1 $\mu g\,L^{-1}$, only for dichloromethane 1 $\mu g\,L^{-1}$) depending on the percentage of halogen in the compound. The detection limit of the method is related to the amount of the analyte in the sample injected on to the column and the volatility of the analytes. On the basis of presented investigations, Biziuk and Czerwiński[82] elaborated a proposal for a Polish Standard Method (PrPN-C-04549-1) for the determination of dichloromethane, trichloromethane, tetrachloromethane, bromodichloromethane, dibromochloromethane, tribromomethane, trichloroethene, tetrachloroethene, 1,1,1-trichloroethane, and 1,1,2,2-tetrachloroethane in water by GC using direct injection of the sample.

4.2 Application of Direct Aqueous Injection for Determination of Semivolatile and Nonvolatile Haloorganic Water Pollutants

DAI on to capillary chromatographic columns connected with selective ECD has also been used for the determination of semivolatile and nonvolatile haloorganic compounds in different aqueous samples.

In the case of the simultaneous determination of volatile and semivolatile compounds, a longer analytical column with an appropriate oven temperature program and a thinner film layer in the analytical column have to be used. Wolska et al.[79] used for the determination of 12 volatile and semivolatile organohalogen compounds (from 1,2-dichloroethane to dichlorobenzenes, trichlorobenzenes and hexachlorobutadiene) cold on-column injection (2 μL), a precolumn (2 m × 0.32 mm i.d.) and an Rtx-624 (60 m × 0.32 mm i.d.) column coated with a 1.8-μm film of stationary phase (6% cyanopropyl–94% dimethylpolysiloxane). The oven temperature program was as follows: 104 °C for 4 min, then heated to 110 °C at 3 °C min^{-1} and to 200 °C at 5 °C min^{-1} and held for 1 min at 200 °C. The detection limit varied from 0.02 $\mu g\,L^{-1}$ for tetrachloromethane and trichloroethene, through 0.04 $\mu g\,L^{-1}$ for trichloromethane to 1.2 $\mu g\,L^{-1}$ for 1,2-dichlorobenzene. The analysis was carried out on water from the Vistula River and Reda River.

In the method routinely used in the laboratory of SAUR–Neptun Gdańsk (water supply and sewage system for Gdańsk) for the determination of volatile and semivolatile haloorganic compounds in tap water, an Rtx-Volatiles column (60 m × 0.32 mm i.d.) coated with a 3-μm film and an oven temperature program (104 °C for 1 min, then heated to 160 °C at 2.5 °C min^{-1}) were used. A typical chromatogram of 15 compounds obtained in this laboratory is shown in Figure 5.

It was also possible to use the DAI/GC/ECD method for the determination of pesticides but at high concentrations. This method was performed for the determination of the pesticide destruction ratio during their chemical oxidation in the aqueous phase (our work). Cold on-column injection (2 μL) and an Rtx-5 capillary column (30 m × 0.25 mm i.d.) coated with a 0.25-μm film (5% diphenyl–95% dimethylpolysiloxane) were used. The oven temperature program was as follows: from 60 to 100 °C at 30 °C min^{-1} and to 280 °C at 10 °C min^{-1}, then held for 10 min at 280 °C. For lindane, *p*,*p*′-DDT and methoxychlor the detection limit was about 0.1 $mg\,L^{-1}$.

4.3 The Combination of Thin-layer Headspace Technique and Direct Aqueous Injection for Determination of Haloorganic Water Pollutants

The application of the DAI/GC/ECD method to samples with very complex matrices containing numerous

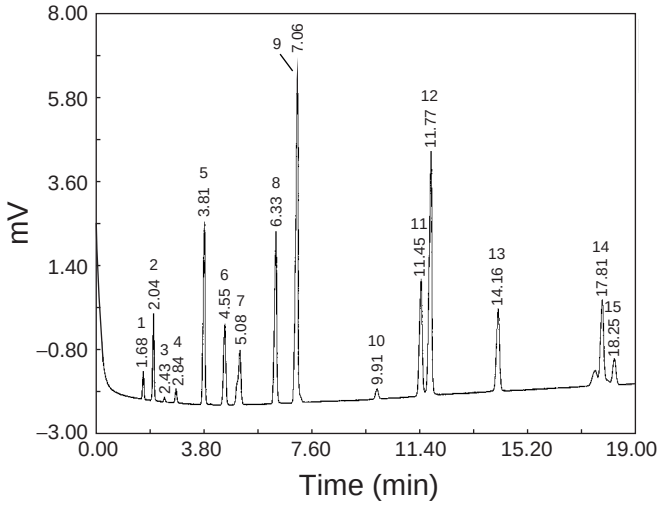

Figure 5 Typical chromatogram of volatile organohalogen compounds (DAI/GC/ECD, temperature programming). Peaks: 1 = 1,1-dichloroethene; 2 = dichloromethane; 3 = 1,2-dichloroethene; 4 = 1,1-dichloroethane; 5 = trichloromethane; 6 = 1,1,1-trichloroethane; 7 = tetrachloromethane; 8 = trichloroethene; 9 = bromodichloromethane; 10 = 1,1,2-trichloroethane; 11 = tetrachloroethene; 12 = dibromochloromethane; 13 = 1, 1, 1, 2-tetrachloroethane; 14 = tribromomethane; 15 = 1,1,2,2-tetrachloroethane.

inorganic salts or nonvolatile organic compounds and also for low concentrations of analytes creates many problems. Only small amounts of the samples (2 μL) are introduced into the column, which causes a high detection limit. The detection limit of the DAI/GC/ECD method for volatile haloorganic compounds is sufficient for tap and swimming pool water treated by chlorination and for polluted surface waters, but insufficient for unpolluted underground or surface waters. Direct injection of aqueous samples into a GC column cannot be performed for complex matrices such as seawater, wastewater, surface water rich in humic acids, body fluids, mineral waters, soft drinks, beers, and juices. Inorganic salts and nonvolatile organic compounds deposited in the precolumn retain water, causing broadening of the water peak and an increase in the retention time for the analytes. These contaminants can be slowly transferred to the main column and to the detector, affecting the peak shape, deactivating the column, and changing the chromatographic conditions of the column and detector. These drawbacks can be eliminated by application of a combination of the thin-layer HS technique with the DAI method. This combination allows the preconcentration of analytes in the aqueous phase and the matrix to be changed from a complex mixture of inorganic and organic compounds to a solution of analytes in pure water. The thin-layer countercurrent headspace (TLHS) method introduced by Kozłowski et al.[83–96] is based on passing volatile compounds from the thin layer of aqueous sample to a countercurrently flowing stream of purified air.

The main part of the apparatus for TLHS is a vertically positioned, spirally wound, thermostated glass tube (2 m × 7 mm i.d.) shown in Figure 6. The sample was introduced on the top of the tube and flowed down in the form of a thin film, from which analytes together with water vapor were transferred to the gaseous phase and transported to the final determination. This approach can be used as a different method for the determination of volatile organic compounds, for the determination of total parameters and individual compounds, and also for preparing the sample for DAI analysis.[90–96] In such a case two TLHS columns have to be used. In the first column, thermostated at a high temperature (up to 90 °C), the analyte compounds were transferred, together with water vapor to a stream of air, and in the second column, cooled at 10–11 °C, water vapor was condensed together with the analytes. The condensate leaving the second column was taken for DAI/GC/ECD analysis. There are very important advantages of such a procedure: the compounds of interest are isolated from the complex aqueous matrix and preconcentrated in pure water, and it allows determination of volatile organohalogen compounds in beverages[94] (mineral waters, juices, soft drinks) and biological fluids (urine).

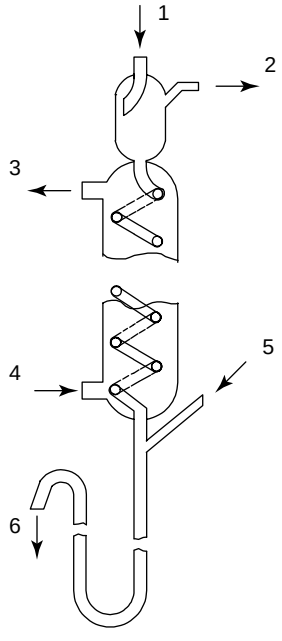

Figure 6 Column for countercurrent TLHS in a spiral tube. 1 = sample inlet; 2 = outlet of the gas stream containing the liberated analytes (to the combustion tube); 3, 4 = outlet and inlet of water from thermostat, respectively; 5 = inlet of purified air; 6 = sample outlet.

5 CONCLUSIONS

DAI is the simplest method for the determination of organic water pollutants that avoids isolation and preconcentration steps.[4,16,17,74] This method is also very fast and easy to automate. The DAI/GC/ECD method is very useful for the determination of volatile and semivolatile organohalogen compounds in tap, swimming pool and surface waters and also for the determination of pesticides and other compounds which are present in water at sufficient concentrations for detection. The method avoids the problems associated with using solvents (toxic and expensive) and incomplete recovery of the analytes during their isolation from the aqueous phase, the effect of potential contaminants when using solvent or SPE, and losses of the analytes during the enrichment step.

The application of the DAI/GC/ECD method to samples with very complex matrices containing numerous inorganic salts or nonvolatile organic compounds, and for low concentrations of analytes, creates many problems. These drawbacks can be eliminated by the application of a combination of the thin-layer HS technique and the DAI method. This combination allows the preconcentration of analytes in the aqueous phase and changes the matrix from a complex mixture of inorganic and organic compounds to a solution of analytes in pure water.

The best method for the determination of pesticides and other nonvolatile compounds is the direct injection of large amounts of water samples in an adsorbent-packed PTV injector with removal of water vapor from the GC system via an early water vapor exit, before the chromatographic column and detector.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
AFID	Alkali Flame Ionization Detection
AOC	Adsorbable Organic Carbon
AOX	Adsorbable Organic Halogen
CE	2-Chloroethanol
DAD	Diode-array Detection
DAI	Direct Aqueous Injection
DEG	Diethylene Glycol
DOC	Dissolved Organic Carbon
ECD	Electron Capture Detection
EG	Ethylene Glycol
ELCD	Electrolytic Conductivity Detection
EOC	Extractable Organic Carbon
EU	European Union
FFAP	Free Fatty Acid Phase
FID	Flame Ionization Detection
FPD	Flame Photometric Detection
FSCC	Fused-silica Capillary Column
FSOT	Fused-silica Open-tubular Capillary Column
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HRGC	High-resolution Gas Chromatography
HS	Headspace
IR	Infrared
ITD	Ion Trap Detection
ITS	Ion Trap Spectrometry
LLE	Liquid–Liquid Extraction
MAC	Maximum Admissible Concentration
MS	Mass Spectrometry
NPD	Nitrogen–Phosphorus Detection
*p*DOX	*p*-Dioxane
PEG	Polyethylene Glycol
POC	Purgeable Organic Carbon
PTV	Programmed Temperature Vaporization
P&T	Purge-and-trap
SCAN	Scanning Ion
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SOC	Suspended Organic Carbon
SPE	Solid-phase Extraction
SPI	Septum Programmable Injector
SPME	Solid-phase Microextraction
TCD	Thermal Conductivity Detection
THM	Trihalomethane
TLHS	Thin-layer Countercurrent Headspace
TOC	Total Organic Carbon
TOX	Total Organic Halogen
TTHM	Total Trihalomethane
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
VOC	Volatile Organic Carbon
VOX	Volatile Organic Halogen
WHO	World Health Organization
2CEE	2-(2-Chloroethoxy)Ethanol

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Trihalomethanes in Water, Analysis of • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of

Pesticides cont'd **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. J. Dojlido, *Chemia Wody*, Arkady, Warsaw, 1987.
2. H. Hellmann, *Analysis of Surface Waters*, Ellis Horwood, New York, 1987.
3. H.A.M. de Kruijf, J.H. Kool (eds.), *Organic Micropollutants in Drinking Water and Health*, Elsevier, Amsterdam, 1985.
4. M. Biziuk, A. Przyjazny, 'Methods of Isolation and Determination of Volatile Organohalogen Compounds in Natural and Treated Waters', *J. Chromatogr.*, **733**, 417–448 (1996).
5. W. Fresenius, K.E. Quentin, W. Schneider (eds.), *Water Analysis*, Springer, Berlin, 1988.
6. E.W.B. de Leer, *Aqueous Chlorination Products. The Origin of Organochlorine Compounds in Drinking and Surface Waters*, Delft University Press, Delft, 1987.
7. R.L. Jolley, W.A. Brungs, R.B. Cumming, V.A. Jacobs (eds.), *Water Chlorination. Environmental Impact and Health Effects*, Ann Arbor Science Publishers, Ann Arbor, MI, 1980.
8. M. Rebhun, J. Manka, A. Zilberman, 'Trihalomethane Formation in High-bromide Lake Galilee Water', *J. Am. Water Works Assoc.*, June, 84–89 (1988).
9. J.M. Montgomery (ed.), *Water Treatment. Principles and Design*, John Wiley & Sons, New York, 1985.
10. B. Koch, S.W. Krasner, 'Occurrence of Disinfection Byproducts in a Distribution System', in *Proceedings of Water Quality Technology Conference, AWWA, Los Angeles, June 18–22, 1989*, American Water Works Association, Denver, CO, 1203–1230, 1989.
11. M. Biziuk, A. Przyjazny, J. Czerwiński, M. Wiergowski, 'Occurrence and Determination of Pesticides in Natural and Treated Waters', *J. Chromatogr.*, **754**, 103–124 (1996).
12. J.W. Moore, S. Ramamoorthy, *Organic Chemicals in Natural Waters. Applied Monitoring and Impact Assessment*, Springer, New York, 1984.
13. N.L. Nemerow, *Stream, Lake, Estuary and Ocean Pollution*, Van Nostrand Reinhold, New York, 1985.
14. R.E. Majors, 'An Overview of Sample Preparation', *LC/GC Int.*, **4**(2), 10–14 (1991).
15. J. Namieśnik, T. Górecki, M. Biziuk, L. Torres, 'Isolation and Preconcentration of Volatile Organic Compounds from Water', *Anal. Chim. Acta*, **237**, 1–60 (1990).
16. J. Namieśnik, Z. Jamrógiewicz (eds.), *Fizykochemiczne Metody Kontroli Zanieczyszczeń Środowiska*, WN-T, Warsaw, 1998.
17. M. Biziuk, 'Methods to Isolate and Determine Organochlorine Compounds and Selected Pesticides in Natural and Treated Waters', *Zesz. Nauk. Politech. Gdańsk.*, No. 513, Chemia No. 31, 3–147 (1994).
18. M. Dressler, 'Extraction of Trace Amounts of Organic Compounds from Water with Porous Organic Polymers', *J. Chromatogr.*, **165**, 167–206 (1979).
19. S.K. Poole, T.A. Dean, J.W. Oudsema, C.F. Poole, 'Sample Preparation for Chromatographic Separation: An Overview', *Anal. Chim. Acta*, **236**, 3–42 (1990).
20. J.A. Leenheer, *Water Analysis*, Academic Press, Oxford, 1984.
21. L. Starostin, Z. Witkiewicz, 'Environmental Water Samples Preparation for Chemical Analysis', *Chem. Anal. (Warsaw)*, **39**, 263–279 (1994).
22. H. Hühnerfuss, H. Kallenborn, 'Chromatographic Separation of Marine Organic Pollutants', *J. Chromatogr.*, **580**, 191–214 (1992).
23. J. Pawliszyn, *Solid Phase Microextraction – Theory and Practice*, Wiley-VCH, New York, 1997.
24. A. Di Corcia, R. Samperi, C. Severini, 'Improvement in the Gas Chromatographic Determination of Trace Amounts of Aliphatic Amines in Aqueous Solution', *J. Chromatogr.*, **170**, 325–329 (1979).
25. A. Di Corcia, R. Samperi, 'Gas Chromatographic Determination at the Parts-per-million Level of Aliphatic Amines in Aqueous Solution', *Anal. Chem.*, **46**(8), 977–981 (1974).
26. L.E. Harris, W.L. Budde, J.W. Eichelberger, 'Direct Analysis of Water Samples for Organic Pollutants with Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **46**, 1912–1917 (1974).
27. T. Fujii, 'Direct Aqueous Injection Gas Chromatography/Mass Spectrometry for Analysis of Organohalides in Water at Concentrations Below the Parts per Billion Level', *J. Chromatogr.*, **139**, 297–302 (1977).

28. A.A. Nicholson, O. Maresz, B. Lemyk, 'Determination of Free Total Potential Haloforms in Drinking Water', *Anal. Chem.*, **49**, 814–819 (1977).
29. F.K. Pfaender, R.B. Jonas, A.A. Stevens, L. Moore, 'Evaluation of Direct Aqueous Injection Method for Analysis of Chloroform in Drinking Water', *Environ. Sci. Technol.*, **12**, 438–441 (1978).
30. P.G. Simonds, E. Kerns, 'Direct Aqueous Injection Gas Chromatography for the Analysis of Trace Organics in Water', *J. Chromatogr.*, **186**, 785–794 (1979).
31. T. Luong, C.J. Peters, R.J. Young, R. Perry, 'Bromide and Trihalomethanes in Water Supply', *Environ. Technol. Lett.*, **1**, 299–310 (1980).
32. D. Hou, J. Tang, 'Determination of Halocarbons in Drinking Water by Direct Aqueous Injection Gas Chromatography', *Anal. Chem.*, **63**, 2078–2080 (1991).
33. *Method 8010 EPA, Halogenated Volatile Organics*, Environmental Protection Agency, Washington, DC, 1986.
34. R. Fussell, D.V. McCalley, 'Determination of Volatile Fatty Acids (C_2–C_5) and Lactic Acid in Silage by Gas Chromatography, *Analyst*, **112**, 1213–1216 (1987).
35. M. Malaiyandi, M.J. Goddard, 'Development of Analytical Methodology and a Report on Collaborative Study on the Determination of Morpholine, Cyclohexylamine, and Diethylaminoethanol in Aqueous Samples by Direct Aqueous Injection Gas Chromatography', *J. Test. Eval.*, March, 87–97 (1990).
36. D.H. Desty, 'Capillary Columns: Trials, Tribulations, and Triumphs', in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, 199–228, Vol. 1, 1965.
37. G. Schomburg, H. Husman, F. Weeke, 'Preparation, Performance and Special Applications of Glass Capillary Columns', *J. Chromatogr.*, **99**, 63–79 (1974).
38. G. Schomburg, H. Husman, F. Weeke, 'Aspects of Double-column Gas Chromatography with Glass Capillaries Involving Intermediate Trapping', *J. Chromatogr.*, **112**, 205–217 (1975).
39. M. Novotny, K.D. Bartle, 'Surface Chemistry on Glass Open Tubular (Capillary) Columns Used in Gas–Liquid Chromatography', *Chromatographia*, **7**(3), 122–127 (1974).
40. M. Novotny, K. Tesařik, 'Surface Treatment of Glass Capillaries for Gas Chromatography', *Chromatographia*, **1**(7/8), 332–333 (1968).
41. M.H.J. van Rijswick, K. Tesařik, 'The Effect of Surface Treatments of Glass Capillaries on Column Performance', *Chromatographia*, **7**(3), 135–137 (1974).
42. K. Grob, K. Grob, Jr, 'On-column Injection on to Glass Capillary Columns', *J. Chromatogr.*, **151**, 311–320 (1978).
43. K. Grob, A. Habich, 'Trace Analysis of Halocarbons in Water; Direct Aqueous Injection with Electron Capture Detection', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **6**, 11–15 (1983).
44. K.J. Grob, 'Further Development of Direct Aqueous Injection with Electron-capture Detection in Gas Chromatography', *J. Chromatogr.*, **299**, 1–11 (1984).
45. K. Grob, Jr, Z. Li, 'Introduction of Water and Water-containing Solvent Mixtures in Capillary Gas Chromatography. I. Failure to Produce Water-wettable Precolumns (Retention Gaps)', *J. Chromatogr.*, **473**, 381–390 (1989).
46. K. Grob, Jr, Z. Li, 'Introduction of Water and Water-containing Solvent Mixtures in Capillary Gas Chromatography. II. Wettability of Precolumns by Mixtures of Organic Solvents and Water: Retention Gap Technique', *J. Chromatogr.*, **473**, 391–400 (1989).
47. K. Grob, Jr, H. Neukom, Z. Li, 'Introduction of Water and Water-containing Solvent Mixtures in Capillary Gas Chromatography. III. Water Resistant Deactivation of Uncoated Precolumns', *J. Chromatogr.*, **473**, 401–409 (1989).
48. K. Grob, E. Muller, 'Introduction of Water and Water-containing Solvent Mixtures in Capillary Gas Chromatography. IV. Principles of Concurrent Solvent Evaporation with Cosolvent Trapping', *J. Chromatogr.*, **473**, 411–422 (1989).
49. K. Grob, F. Küffer, 'Analysis of Volatile Halocarbons in Water by GC/ECD Involving Direct On-column Injection of the Sample', *J. High Resolut. Chromatogr.*, **13**, 561–564 (1990).
50. A. Zlatkis, J.Q. Walker, 'Direct Sample Introduction for Large Bore Capillary Columns in Gas Chromatography', *J. Gas Chromatogr.*, **1**(5), 9–11 (1963).
51. I.F.M.M. Temmerman, D.J.M. Quaghebeur, 'Analysis of Trihalomethanes by Direct Aqueous Injection (THM/DAI)', *J. High Resolut. Chromatogr.*, **13**, 379–381 (1990).
52. I. Tammerman, F. David, P. Sandra, R. Soniassy, 'GC Analysis of Trihalomethanes by Direct Aqueous Injection Using Automatic On-column Injection and Electron Capture Detection', Application Note 228–135, Publication No. (43) 5091-1675E, Hewlett-Packard, Waldbronn, 1991.
53. 'Determination of Halocarbons in Drinking Water by Direct Aqueous Injection and Electron Capture Detector (DAI/ECD)', Short Notes, Application Sheets, SN/32-E-A.G.F. Ili Zonca, Carlo Erba, Milan, 1987.
54. R. Soniassy, P. Sandra, C. Schlett, 'Direct Aqueous Injection/CGC/ECD', in 'Water Analysis. Organic Micropollutants', Publication No. 5962-6216E, Hewlett-Packard, Waldbronn, 69–72, 1994.
55. W. Sanders, 'Volatiles. Trihalomethanes (THMs), EPA Methods 501.1, 501.2, 501.3', in 'Hewlett-packard Environmental Catalog', Publication No. (43) 5091–5909EUS, Hewlett Packard, Avondale, PA, 16–17, 1993.
56. D. Carmichael, W. Holmes, 'Screening of Trihalomethanes by Direct Aqueous Injection Using Electron Capture Detection', *J. High Resolut. Chromatogr.*, **13**, 267–269 (1990).
57. 'Gas Chromatography. 1079 Universal Capillary Injector', Paper No. 802723-002 SEP-8697B/207, Varian, Palo Alto, CA, 1998.

58. '3800 GC', Paper No. 802810-002 SEP-5805E/121, Varian, Palo Alto, CA, 1997.
59. M. De Fatima Alpendurada, M.R. Mourao, 'Determinacao de Trihalometanos na Áqua Distribuida no Concelho do Porto Pelo Método de Injeccao Aquosa Directa', *Rev. Port. Farm.*, **4**, 183–187 (1994).
60. J.L. Ellington, C.D. Trusty, 'Quantitative Analysis of Alkyl Phosphates Using Automated Cool On-column Aqueous Injection', *J. High Resolut. Chromatogr.*, **12**, 470–473 (1989).
61. D.V. McCalley, 'Analysis of Volatile Fatty Acids by Capillary Gas Chromatography Using On-column Injection of Aqueous Solutions', *J. High Resolut. Chromatogr.*, **12**, 465–467 (1989).
62. W.D. Payne, T.W. Collette, 'Identification of Bis(2-chlorethyl) Ether Hydrolysis Products by Direct Aqueous Injection GC/FT/IR', *J. High Resolut. Chromatogr.*, **12**, 693–696 (1989).
63. D.F. Gurka, S.M. Pyle, R. Titus, 'Environmental Analysis by Direct Aqueous Injection', *Anal. Chem.*, **64**, 1749–1754 (1992).
64. S.M. Pyle, D.F. Gurka, 'Volatile Organic Analysis by Direct Aqueous Injection', *Talanta*, **41**, 1845–1852 (1994).
65. H.G.J. Mol, H.-G.M. Janssen, C.A. Cramers, J.J. Vreuls, U.A.Th. Brinkman, 'Trace Level Analysis of Micropollutants in Aqueous Samples Using Gas Chromatography with On-line Sample Enrichment and Large Volume Injection', *J. Chromatogr.*, **703**, 277–307 (1995).
66. A. Zlatkis, F.-S. Wang, H. Shanfield, 'Direct Gas Chromatographic Analysis of Aqueous Samples at the Part-per-billion and Part-per-trillion Levels', *Anal. Chem.*, **55**, 1848–1852 (1983).
67. B.B. Gerhardt, H.J. Cortes, 'Determination of Chlorpyrifos in Water by Large-volume Direct Aqueous Injection Capillary Gas Chromatography', *J. Chromatogr.*, **503**, 377–383 (1990).
68. K. Grob, D. Fröhlich, 'Splitless Injection of Large Volumes of Aqueous Samples: Further Experience. Experiments on Analyzing Triazines by Direct Injection of Drinking Water', *J. High. Resolut. Chromatogr.*, **17**, 792–794 (1994).
69. K. Grob, M. Buedermann, 'Vaporising Systems for Large Volume Injection or On-line Transfer into Gas Chromatography: Classification, Critical Remarks and Suggestions', *J. Chromatogr.*, **750**, 11–23 (1996).
70. S. Müller, J. Efer, W. Engewald, 'Gas Chromatographic Water Analysis by Direct Injection of Large Sample Volumes in an Adsorbent-packed PTV Injector', *Chromatographia*, **38**, 694–700 (1994).
71. S. Müller, J. Efer, W. Engewald, 'Water Pollution Screening by Large-volume Injection of Aqueous Samples and Application to GC/MS Analysis of a River Elbe Sample', *Fresenius' J. Anal. Chem.*, **357**, 558–560 (1997).
72. J. Teske, J. Efer, W. Engewald, 'Large-volume PTV Injection: New Results on Direct Injection of Water Samples in GC Analysis', *Chromatographia*, **46**, 580–586 (1997).
73. J. Teske, J. Efer, W. Engewald, 'Large-volume PTV Injection: Comparison of Direct Water Injection and In-vial Extraction for GC Analysis of Triazines', *Chromatographia*, **47**, 35–41 (1998).
74. M. Turska, L. Wolska, B. Zygmunt, J. Namieśnik, 'Direct Injection of Aqueous Samples into Gas Chromatographic Columns', *Chem. Anal. (Warsaw)*, **42**, 787–797 (1997).
75. M. Biziuk, Ż. Połkowska, D. Gorlo, W. Janicki, J. Namieśnik, 'Determination of Volatile Organic Compounds in Water Intakes and Tap Water by Purge and Trap and Direct Aqueous Injection–Electron Capture Detection Techniques', *Chem. Anal. (Warsaw)*, **40**, 243–251 (1995).
76. M. Biziuk, J. Namieśnik, J. Czerwiński, D. Gorlo, B. Makuch, W. Janicki, Ż. Połkowska, L. Wolska, 'Occurrence and Determination of Organic Pollutants in Tap and Surface Waters of the Gdańsk District', *J. Chromatogr.*, **733**, 171–184 (1996).
77. M. Biziuk, J. Czerwiński, E. Kozłowski, 'Identification and Determination of Organohalogen Compounds in Swimming Pool Water', *Int. J. Environ. Anal. Chem.*, **50**, 109–115 (1993).
78. J. Czerwiński, D. Gorlo, Ż. Połkowska, M. Biziuk, J. Namieśnik, 'Seasons Changes in Concentration Level of Volatile Hydrocarbon Organohalogen Compounds and Pesticides in the Surface Waters and Seawater in the Vincinity of Gdańsk', *Ecol. Chem.*, **4**, 231–238 (1995).
79. L. Wolska, C. Olszewska, M. Turska, B. Zygmunt, J. Namieśnik, 'Volatile and Semivolatile Organohalogen Trace Analysis in Surface Water by Direct Aqueous Injection GC/ECD', *Chemosphere*, **37**, 2645–2651 (1998).
80. M. Biziuk, Ż. Połkowska, D. Gorlo, W. Janicki, J. Namieśnik, 'Monitoring of Volatile Organic Compounds (VOC's) in the Surface and Seawaters in the Gdaśsk District', in *International River Water Quality*, eds. G.A. Best, T. Bogacka, E. Niemirycz, E&FN Spon, London, 127–131, 1997.
81. B. Zygmunt, M. Biziuk, W. Wardencki, J. Namieśnik, 'Surface Water Pollutants and Problems with Their Analysis', in *International River Water Quality*, eds. G.A. Best, T. Bogacka, E. Niemirycz, E&FN Spon, London, 95–103, 1997.
82. M. Biziuk, J. Czerwiński, Polish Standard Method – PrPN-C-04549–1, 'Determination of Dichloromethane, Trichloromethane, Tetrachloromethane, Bromodichloromethane, Dibromochloromethane, Tribromomethane, Trichloroethylene, Tetrachloroethylene, 1,1,1-Trichloroethane, 1,1,2,2-Tetrachloroethane in Water by Gas Chromatography Using Direct Injection of the Sample', Polish Committee for Standardization, Warsaw, 1998.
83. E. Kozłowski, E. Sieńkowska-Zyskowska, M. Biziuk, 'Countercurrent Thin-layer Headspace as a New Approach to Continuous Analysis of Volatile Organic

Compounds in Water', *Chem. Anal. (Warsaw)*, **28**, 817–821 (1983).
84. E. Kozłowski, E. Sieńkowska-Zyskowska, T. Górecki, 'Continuous Flow Thin-layer Headspace (TLHS) Analysis. I. Conductometric Determination of Volatile Organic Halogens (VOX) in Tap Water', *Fresenius' J. Anal. Chem.*, **339**, 19–25 (1991).
85. E. Kozłowski, T. Górecki, E. Sieńkowska-Zyskowska, 'Continuous Flow Thin-layer Headspace (TLHS) Analysis. II. Spectrophotometric Determination of Volatile Organic Halogens (VOX) in Water', *Fresenius' J. Anal. Chem.*, **339**, 882–885 (1991).
86. E. Kozłowski, T. Górecki, E. Sieńkowska-Zyskowska, 'Continuous Flow Thin-layer Head-space (TLHS) Analysis. III. Potentiometric Determination of Volatile Organic Halogens (VOX) in water', *Fresenius' J. Anal. Chem.*, **340**, 454–458 (1991).
87. E. Kozłowski, T. Górecki, E. Sieńkowska-Zyskowska, 'Continuous Flow Thin-layer Headspace (TLHS) Analysis. IV. Indirect Potentiometric Determination of Volatile Organic Halogens (VOX) in Water', *Fresenius' J. Anal. Chem.*, **340**, 773–776 (1991).
88. E. Kozłowski, E. Sieńkowska-Zyskowska, T. Górecki, 'Continuous Flow Thin-layer Head-space (TLHS) Analysis. V. Comparison of the Conductometric and Indirect Potentiometric Methods of Volatile Organic Halogens (VOX) Determination in Tap Water', *Fresenius' J. Anal. Chem.*, **342**, 20–25 (1992).
89. E. Kozłowski, T. Górecki, E. Sieńkowska-Zyskowska, 'Continuous Flow Thin-layer Headspace (TLHS) Analysis. VI. Comparison of the Results of Determination of Volatile Organohalogen Compounds in Tap Water in the Form of a Group Parameter (VOX) with Partial Speciation by DAI–ECD Capillary Gas Chromatography $(GC)^2$', *Fresenius' J. Anal. Chem.*, **342**, 401–404 (1992).
90. E. Kozłowski, A. Kołodziejczyk, J. Kozłowski, 'Continuous Flow Thin-layer Headspace (TLHS) Analysis. VII. General Equation of Simultaneous Separation and Preconcentration of Volatile Analytes in Liquids', *Chem. Anal. (Warsaw)*, **38**, 315–321 (1993).
91. E. Kozłowski, A. Kołodziejczyk, J. Kozłowski, 'Continuous Flow Thin-layer Head-space (TLHS) Analysis. VIII. Partial Solutions of the General Equation, Depending on the Flow Rate of the Liquid Phases', *Chem. Anal. (Warsaw)*, **38**, 699–710 (1993).
92. E. Kozłowski, Ż. Połkowska, 'Continuous Flow Thin-layer Head-space (TLHS) Analysis. IX. Isolation of Trihalomethanes from Complex Mixtures with Autogenous Generation of a Liquid Sorbent Stream', *Chem. Anal. (Warsaw)*, **41**, 173–181 (1996).
93. Ż. Połkowska, E. Kozłowski, 'Continuous Flow Thin-layer Headspace (TLHS) Analysis. X. Preconcentration of Volatile Analytes in Autogenously Generated Stream of Liquid Sorbent', *Chem. Anal (Warsaw)*, **41**, 183–191 (1996).
94. Ż. Połkowska, J. Namieśnik, J. Czerwiński, B. Zygmunt, 'Volatile Organo-halogen Trace Analysis in Beverages by Thin-layer Headspace Enrichment and Electron-capture Gas Chromatography', *Int. J. Food Sci. Technol.*, **31**, 387–398 (1996).
95. Ż. Połkowska, M. Turska, T. Górecki, J. Namieśnik, 'Comparison of the New Version of Thin Layer Headspace Analysis with Autogenous Generation of the Liquid Sorbent (TLHS/DAI/GC/ECD) with Liquid–Liquid Extraction for the Determination of Volatile Organohalogen Compounds in Liquid Samples', *Process Control Qual.*, **11**, 1–8 (1998).
96. Ż. Połkowska, E. Kozłowski, T. Górecki, J. Namieśnik, 'Theoretical Principles of Thin Layer Headspace Analysis (TLHS)', *Toxicol. Environ. Chem.*, **68**, 1–42 (1999).

Gas Chromatography with Atomic Emission Detection in Environmental Analysis

Stig Pedersen-Bjergaard
University of Oslo, Oslo, Norway

1 Introduction 1
2 Selectivity and Sensitivity Considerations 2
3 Compound Independent Calibration for Quantitative Analysis 4
4 Estimation of Molecular Formulas 5
5 Environmental Applications 7
5.1 Tin-containing Compounds 7
5.2 Lead-containing Compounds 9
5.3 Mercury-containing Compounds 9
5.4 Pesticides 10
5.5 Chlorinated Biphenyls and Phenols 11
5.6 Other Compounds Containing Nonmetallic Elements 12
6 Current Status and Future Directions 12
Abbreviations and Acronyms 12
Related Articles 13
References 13

When capillary gas chromatography (GC) is coupled with atomic emission detection (AED), mixtures of volatile and semivolatile organic compounds are first separated based on different retention in a capillary GC column, whereafter the separated compounds are introduced directly into a helium (or argon) plasma. The high temperature of this plasma results in atomization of the separated compounds followed by excitation of the constituent atoms. From the excited states, the atoms emit light of characteristic wavelengths which are monitored continuously during the chromatographic separation by an optical spectrometer. This results in element-selective chromatograms, where all elements from the periodic table (except helium or argon) may be monitored selectively with a simple change of wavelength on the spectrometer. With the use of multichannel spectrometers, several elements may be monitored simultaneously resulting in multielemental detection. Both nonmetallic and metallic elements are detected with a very high selectivity. This enables environmental contaminants containing heteroatoms to be determined in dirty extracts without extensive sample cleanup. However, since detection limits (DLs) in GC/AED range from 1–100 pg s^{-1}, extensive analyte enrichment is often required prior to environmental analysis. Because the separated compounds are atomized within the plasma, only minor structural variations are observed for the detector responses. This, in combination with the multielemental capability, enables estimation of empirical formulas for each of the separated compounds, which may be utilized to support identification of environmental micropollutants. In addition, the structure-independent detector responses enable simplified calibration for quantitative purposes because different compounds may be quantified based on a single elemental calibration curve.

1 INTRODUCTION

When capillary GC is coupled with AED, high separation power and simultaneous multielemental detectability are combined into an attractive (hyphenated) analytical technique GC/AED. Volatile and semivolatile organic compounds present in complex mixtures are separated in the GC and subsequently introduced into a small helium (or argon) plasma placed in continuation of the GC column (Figure 1). The high temperature of this discharge (3000–10 000 K) results in atomization of each of the separated compounds followed by excitation of the constituent atoms to higher electronic states. Light of wavelengths characteristic of the elements introduced are emitted by the excited atoms as they undergo transitions to lower energy levels. With an optical spectrometer, emission light for the elements of interest is selected and measured continuously resulting in element-selective chromatograms. Since atomic lines are narrow and often intense, GC/AED provides high elemental selectivity and acceptable sensitivity. Theoretically, every element from the periodic table (except helium or argon) may be detected by a simple change of wavelength. Since all emission lines are present in the plasma simultaneously, either single- or multichannel detection is possible. Compared with the array of selective detectors responding to one or two elements (e.g. flame photometric detection (FPD), nitrogen–phosphorus detection (NPD), sulfur chemiluminescence detection (SCD), oxygen-flame ionization detection (O-FID)), this multielemental capability represents a major improvement in terms of versatility.

The attractive features of GC/AED were reported for the first time in 1965 by McCormick et al. with an atmospheric pressure argon microwave induced plasma (MIP) coupled to packed column GC.[1] Although the period 1965–1975 also saw research into direct current discharges,[2] most GC/AED research was carried out

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

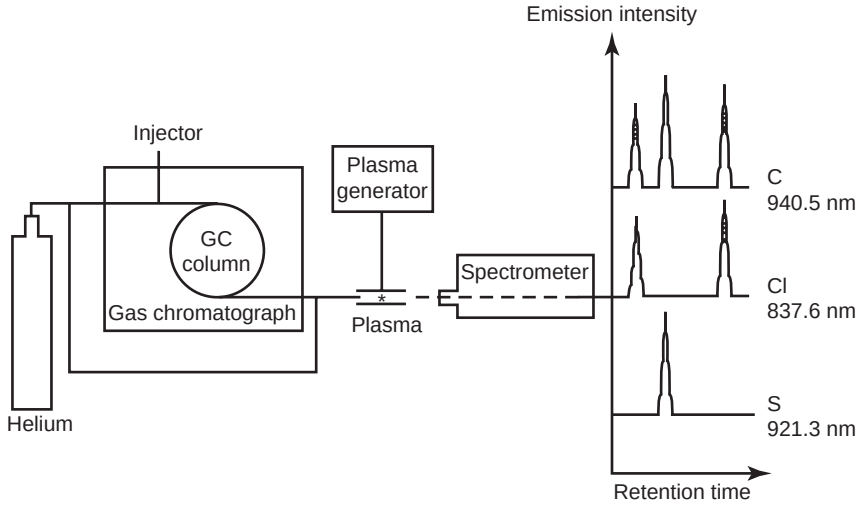

Figure 1 Principle of AED in capillary GC.

with systems utilizing argon MIPs or reduced-pressure helium MIPs. Unfortunately, the excitation potential of the argon plasmas was insufficient for the detection of several nonmetallic elements, and the evacuation equipment required to produce low-pressure helium plasmas was complex. Therefore, the introduction of the TM_{010} cavity by Beenakker in 1976 represented a distinct technical improvement in the field of GC/AED because this permitted helium MIPs to be sustained at atmospheric pressure.[3] Implementation of the TM_{010} cavity by other workers was reported from 1978–1980, which almost terminated the interest for reduced-pressure MIP systems.

Although most effort has been focused on MIP systems, after 1980, successful GC detection was accomplished with several alternative discharges, including a 25 kHz helium after-glow device,[4] a 350 kHz radio frequency helium plasma,[5,6] a 60 Hz alternating current helium plasma,[7] a 27.18 MHz capacitively coupled helium plasma,[8] and a 27.12 MHz stabilized capacitively coupled helium plasma.[9] In addition, with argon at atmospheric pressure, direct current plasmas (DCPs)[10] and inductively coupled plasmas (ICPs)[11] have been evaluated, but both plasma systems suffered from poor excitation characteristics for several important nonmetallic elements and from a large consumption of high-purity argon.

GC/AED was carried out solely with laboratory-built instrumentation until the late 1970s, when the first commercial atomic emission detector for GC was introduced by Applied Chromatography Systems. This equipment was based on a reduced-pressure helium MIP and a Rowland circle spectrometer with 12 fixed channels for simultaneous multielement detection.[12] However, the popularity was low, and even a modified version based on an atmospheric-pressure helium MIP sustained in a TM_{010} cavity was implemented only in a limited number of research laboratories. In 1989, a second attempt was made by Hewlett-Packard to commercialize an atomic emission detector for GC. A totally automated benchtop system was introduced, which utilized an atmospheric-pressure helium MIP for analyte excitation and a photodiode array for the monitoring of atomic emission.[13,14] The photodiode array, which covered the wavelength range from 160–800 nm in 25 nm portions, provided simultaneous multielement detection, multipoint spectral background correction, and elemental confirmation by atomic emission spectra. From the same company, an improved version of this atomic emission detector was recently introduced as a less expensive second-generation instrument with enhanced sensitivity. The two commercially available GC/AED systems from Hewlett-Packard have increased the interest for the technique substantially, and several environmental applications of GC/AED have emerged in recent years. Below, advantages and limitations of GC/AED for environmental analysis are discussed and some of the most interesting environmental applications are reviewed. Principal attention is focused on results obtained with the commercial system, while only a few data from home-built systems are included owing to the difficulties of reproducing this type of complex instrumentation.

2 SELECTIVITY AND SENSITIVITY CONSIDERATIONS

Environmental samples are typically very complex with a large number of matrix components, among which

the analytes of interest are present at low concentration levels. Therefore, both selectivity and DLs are important parameters for the successful determination of most environmental micropollutants. In this section, attention will be focused on both of these parameters in order to understand the advantages and the limitations of GC/AED for environmental analysis.

As atomic emission lines are very narrow, overlap from different elements is normally not a problem in GC/AED. However, emission spectra from helium plasmas utilized for GC/AED normally also contain molecular bands (originating from incomplete atomization or from atomic recombinations) which are generally relatively broad. Unfortunately, these may interfere with the atomic lines of interest causing false positive signals (interferences) or negative baseline excursions. Therefore, in order to realize the full potential of the narrow atomic emission lines and to ensure high elemental selectivity, spectral background correction is of great importance in GC/AED.[15] With commercial instrumentation, this is accomplished utilizing a photodiode array for the optical measurements, where continuous multipoint data correction results in very high selectivity for the most important elements found in GC amenable compounds. This is illustrated in Table 1, where the selectivity for several metallic and nonmetallic elements has been determined relative to carbon. With chlorine as an example (selectivity value of 27 000), 27 000 ng carbon (from a hydrocarbon) is required to produce the same signal on the chlorine channel as obtained from 1 ng of chlorine. For practical work with environmental samples, the high elemental selectivities obtained by GC/AED is obviously a major advantage. Organic micropollutants containing heteroatoms (such as Cl, Br, S, Sn, and Hg) may be selectively detected even in very complicated extracts where the resolving power of GC is insufficient. This is illustrated in Figure 2, where GC/AED was utilized for the detection of polychlorinated biphenyls (PCBs) in a highly contaminated marine sediment.[17] In spite of the high background from a crude oil matrix, the PCBs emerged as distinct peaks in the highly selective chlorine trace. Thus, the PCBs were easily located in the chlorine trace and quantitative analysis was rapidly performed because no interferences distorted the peak areas. The highly contaminated sediment sample was also exposed to analysis by GC/MS,[17] but this technique suffered from serious interference when operated both in the low- and high-resolution modes (Figure 2). Thus, although selected ion monitoring was utilized, localization and quantitation of the PCBs were almost impossible based on the GC/MS results.

Table 1 Selectivities and DLs for the second generation of the commercial GC/AED system from Hewlett-Packard[16]

Element	Wavelength (nm)	Selectivity[a]	DL ($pg\,s^{-1}$)[b]
C	193	–	0.6
H	656	–	0.2
F	690	166 000	10.7
Cl	837	27 000	13.0
Br	827	15 000	15.3
N	174	7 000	4.6
O	171	18 000	43.4
S	181	37 000	0.6
P	186	14 000	0.5
Sn	326	3 300 000	0.4
Hg	254	258 000	0.6
Pb	406	286 000	0.2

[a] Selectivities were defined as the ratio of the response per nanogram of element to the response per nanogram of carbon.
[b] DLs were defined as the amount of element required to produce a peak twice the noise level, divided by the full width at half-height of the peak in seconds.

In addition to selectivity considerations, DLs of GC/AED are of great importance in evaluating the potential for environmental analysis. As illustrated in Table 1 for the commercial system, DLs of GC/AED are generally at the low picograms per second level. Unfortunately, values for the different elements vary by a factor of 200 depending on several factors such as atomic line intensity, excitation potential, and natural background in the plasma. For oxygen, an important element in environmental micropollutants, the DL is relatively high and therefore the practical utility of O-selective detection by GC/AED is strongly limited for environmental applications. DLs for chlorine and bromine, which are important elements within environmental monitoring, were relatively high for the first generation of the Hewlett-Packard GC/AED, but significantly lower values have been reported especially for chlorine with the new version.[17] Thus, especially with the second generation of GC/AED, the technique may be an alternative for the determination of pesticides and polyhalogenated compounds such as PCBs (discussed in detail in sections 5.4 and 5.5). For sulfur, phosphorus, and most of the metallic elements, DLs are at the low picograms per second level supporting the idea that compounds containing these elements may be monitored at relatively low concentration levels by GC/AED.

From the discussion above, high elemental selectivity is the major advantage of GC/AED for environmental analysis. Thus, organic micropollutants containing heteroatoms may be selectively detected in complicated samples without serious interference, and time-consuming sample cleanup may be reduced. However, for some elements DLs are relatively high, which may limit the applicability of the technique or which may call for extensive analyte preconcentration in order to match

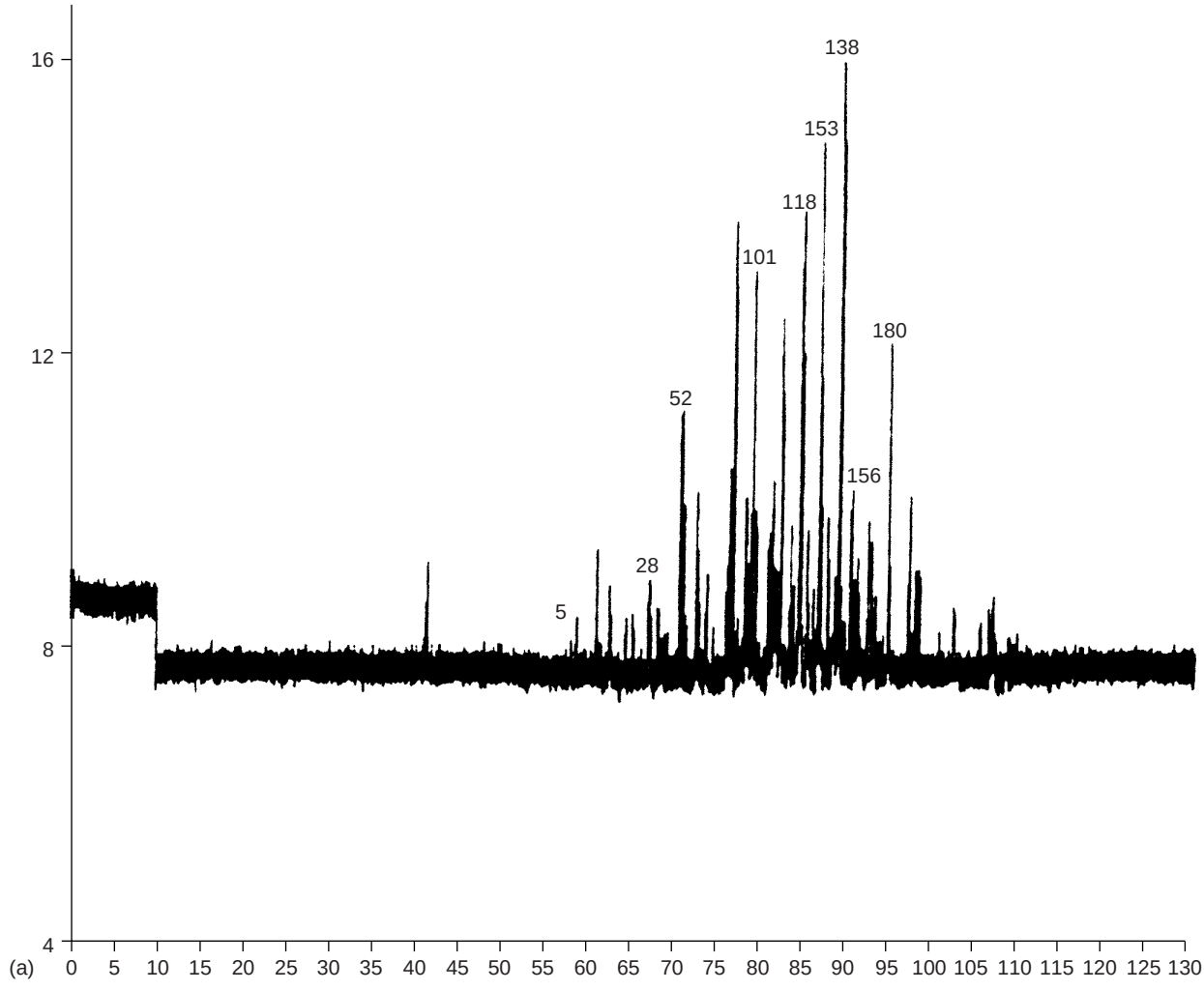

Figure 2 GC analysis of an extract of a highly contaminated marine sediment by (a) GC/AED in the Cl-selective mode and by GC/MS in the (b) low- and (c) high-resolution modes (summarized ion selective chromatograms for molecular ions of PCBs).[16] (Reproduced from *Chromatographia*, **43**, 44–52, 1996, with permission from Vieweg.)

the low concentration levels typical for environmental samples.

3 COMPOUND INDEPENDENT CALIBRATION FOR QUANTITATIVE ANALYSIS

During detection, the compounds separated by GC are atomized within the AED and emission is measured from excited atoms. Therefore, in theory, the detector response per element (area counts per nanogram of element) should be independent of the molecular structure from which it originated. For quantitative purposes, this should enable universal calibration where a large number of different compounds are quantified based on a single elemental calibration curve (e.g. for chlorine) obtained for a randomly selected reference compound. Unfortunately, for practical work, this type of universal calibration is somewhat complicated by discrimination during GC injection and by chemical reactions/spectral interferences occurring within the plasma. As illustrated in Table 2 for some PCBs, deviations up to 10% occurred when the chlorine signal from a standard solution of PCB no. 128 was utilized for the quantitation of the other PCBs included in Table 2.[17] Nevertheless, for the determination of organic micropollutants, purchase of reference material is expensive and in several cases, reference standards for the compounds are not commercially available. Thus, for applications where very high accuracy is not required, analysis by GC/AED and universal calibration may be a very fast and simple concept to obtain quantitative data.

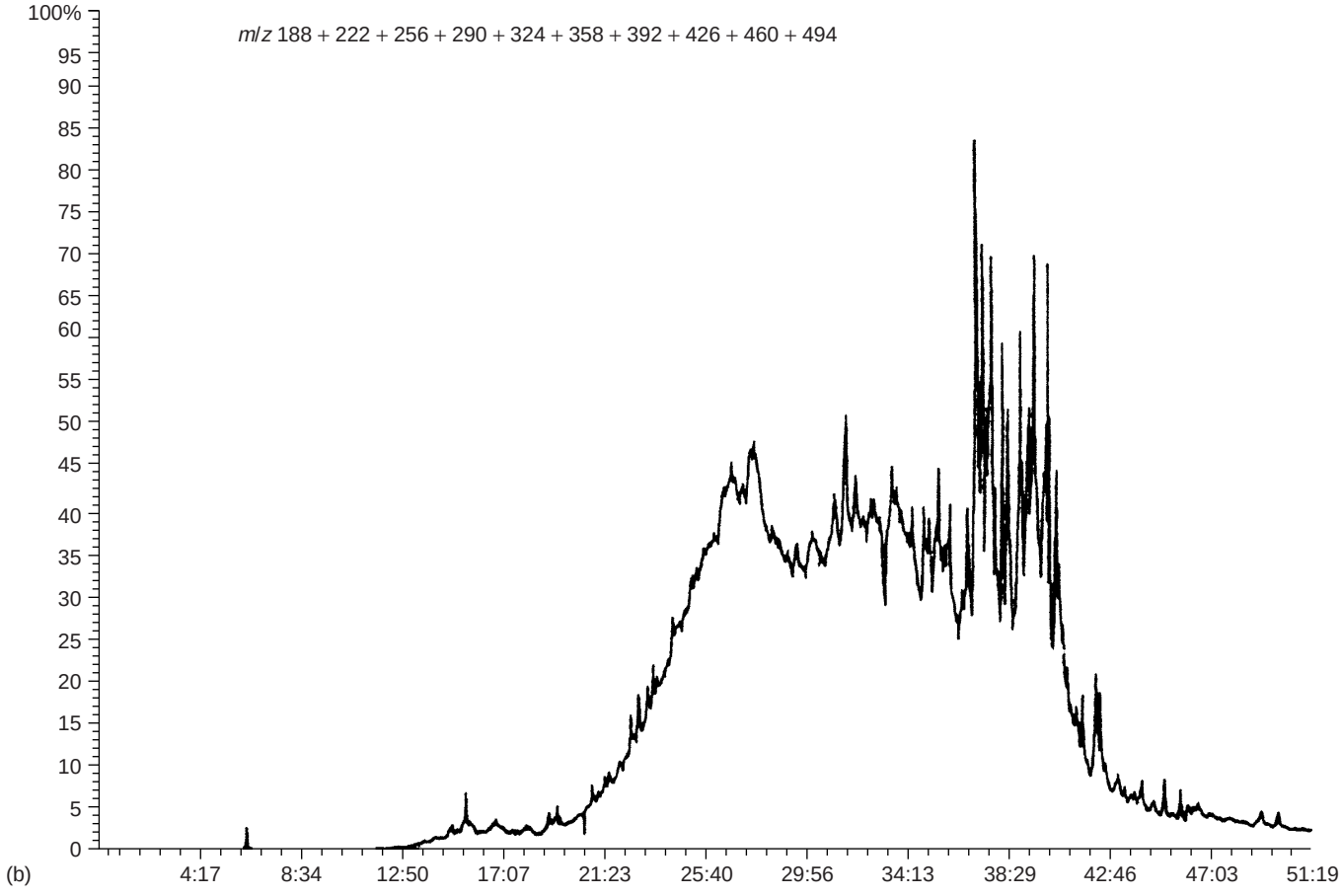

Figure 2 (*Continued*)

4 ESTIMATION OF MOLECULAR FORMULAS

With GC/AED operated in the multielement selective mode, where several different elements are monitored simultaneously, the resulting element-selective chromatograms directly give qualitative information about the constituent elements for each of the separated compounds. With peak no. 5 in Figure 3 as an example, where halogenated alkylbenzenes were detected in nickel industry wastewater, the element-selective chromatograms proved that C, H, and Cl were all present in the compound, while Br was not a constituent of this particular compound. In addition to this, since elemental responses are almost independent of molecular structures (as discussed in section 3), the element-selective chromatograms also contain quantitative information about the elemental composition for each of the separated compounds. Thus, based on simple calibration and multielemental detection, empirical/molecular formulas may be calculated based on GC/AED. This is illustrated in Table 3, where the molecular formula for several halogenated alkylbenzenes present in nickel refinery wastewater (Figure 3) was calculated based on dichlorobenzene as reference.[18] Although deviations up to 30% from correct values were observed in some cases, the approximate empirical formulas were utilized to support rapid analyte identification. For target compound analysis, where analyte identification principally is based on retention time data obtained with standard solutions of the components of interest, empirical formulas calculated by GC/AED may significantly improve the reliability of analyte identifications. For nontarget purposes, however, where retention time information normally is not available, identification based solely on GC/AED is unreliable and should be based principally on GC with mass spectrometric or infrared detection.[19,20] Although identification is based on GC/MS, results from GC/AED may be of great interest to complement mass spectral data. This is illustrated in Figure 4, where chromatograms from both GC/MS and GC/AED analysis of an extract of a pulp mill effluent are shown.[19] The compound corresponding to peak no. 4 was identified as tetrachlorothiophene (C_4Cl_4S) based on

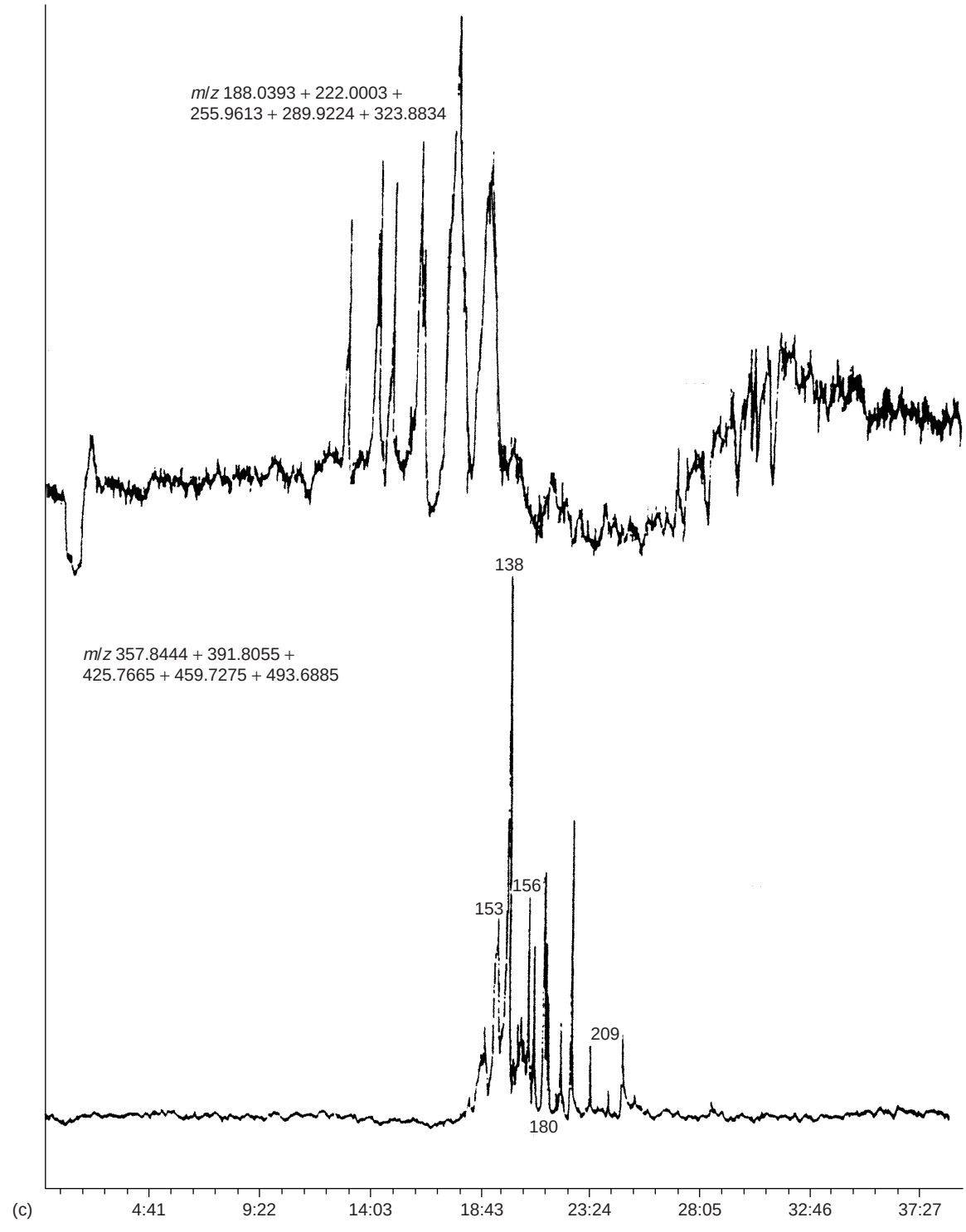

Figure 2 (*Continued*)

library search on the GC/MS system. This conclusion was supported by the GC/AED results proving that C, Cl, and S were present in the compound while H and O were absent. For peak no. 5, no library mass spectrum was available and manual interpretation was necessary. The mass spectrum (Figure 4) revealed the presence of three chlorine atoms and a molecular mass of 216. Unfortunately, the mass spectrum also contained fragments from co-eluting compounds, and further interpretation was therefore difficult. However, the GC/AED results

Table 2 Normalized chlorine response factors (area counts per nanogram) for selected PCB congeners

Congener number	No. of chlorine atoms	Normalized chlorine response factor[a]
31	3	0.974
53	4	1.049
105	5	1.000
128	6	0.966
185	7	1.018
194	8	1.050
209	10	1.059

[a] Chlorine response factors normalized against congener no. 105.

demonstrated that C, H, S, Cl, and O were all present in the compound. Based on this additional information, and assuming also that peak no. 5 was a thiophene, the compound contained one S atom, three Cl atoms, and at least four C atoms (C_4Cl_3S corresponds to 185 mass units). The remaining mass difference up to 216 was 31, and for this part of the molecule, the GC/AED results proved the presence of a single O atom. The 31 mass units probably corresponded to a methoxy group ($-OCH_3$), and the compound was suspected to be an isomer of trichloromethoxythiophene.

5 ENVIRONMENTAL APPLICATIONS

Based on the discussion in sections 2, 3, and 4, the major limitation of GC/AED is related to relatively high DLs for some of the nonmetallic elements, while the major advantages are related to the high selectivity, the multielemental capability, and the possibilities of performing universal calibration (quantitative analysis) as well as estimating empirical/molecular formulas. Based on these characteristics, important environmental applications will be reviewed in this section (section 5).

5.1 Tin-containing Compounds

Substantial research has been reported on the GC/AED determination of organotin compounds originally used as antifouling paints on ships, slimicides, pesticides, and as stabilizers in polyvinyl chloride. Thus, organotin compounds have been determined in sediments,[21–27] soil,[23] water,[22,26–31] and biological samples[21,24,25] by GC/AED. An example is illustrated in Figure 5, where tin-organic compounds were detected in mussel tissue.[24] Excellent selectivity has been reported for GC/AED operated in the Sn-selective mode. For complex environmental samples, the selectivity of GC/AED significantly exceeded both GC/MS[26] and GC/FPD.[21] In addition to the high selectivity, the presence of

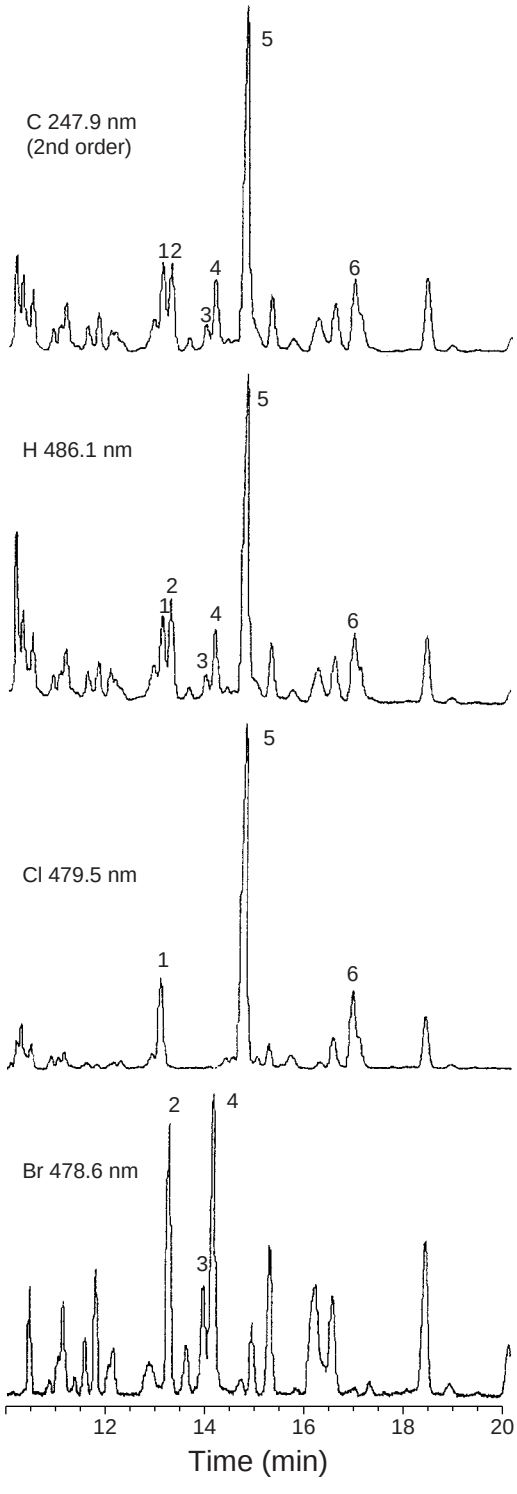

Figure 3 GC/AED analysis of an extract of nickel industry wastewater in the C-, H-, Cl-, and Br-selective modes.[18] (Reproduced by permission of 'MicroSeparations. Molecular Formula Determination of Halogenated Compound in Environmental Samples Using GC and AED', S. Pedersen-Bjergaard, T.N. Asp, J. Vedde, T. Greibrokk, *J. Microcolumn. Sep.* Copyright©. Reprinted by permission of John Wiley & Sons, Inc.)

Table 3 Molecular formulas determined by GC/AED for six halogenated alkylbenzenes present in nickel refinery waste water[18]

Peak no.	Correct formula	Formula determined by GC/AED
1	$C_9H_9Cl_3$	$C_{8.8}H_{8.3}Cl_3$
2	$C_9H_{10}Br_2$	$C_{11.3}H_{13.2}Br_2$
3	$C_9H_{10}Br_2$	$C_{8.1}H_{7.9}Br_2$
4	$C_9H_{10}Br_2$	$C_{7.3}H_{7.2}Br_2$
5	$C_9H_9Cl_3$	$C_{9.0}H_{8.6}Cl_3$
6	$C_9H_9Cl_3$	$C_{8.2}H_{7.9}Cl_3$

tin was confirmed in the case of GC/AED utilizing atomic emission spectra obtained by the photodiode array and stored during each GC analysis.[21] The DL of the commercial GC/AED has been reported to be comparable with GC/FPD,[27] which is the most popular detector for Sn-selective detection. GC with atomic absorption spectrometry has also been utilized for Sn-selective detection, but this technique was 1–2 orders of magnitude less sensitive than GC/AED.[24,27] When operated in the selected ion monitoring mode, GC/MS was found to provide only slightly better DLs than GC/AED[26] supporting the idea that GC/AED is highly attractive for Sn-selective detection. With GC/AED, organotin compounds have been detected down to the 0.1-ng L^{-1} level in water samples[28–31] and to the 0.2-ng g^{-1} level in sediments.[23,25] While quantification by GC/MS required references for each of the compounds of interest, several tin-organic compounds

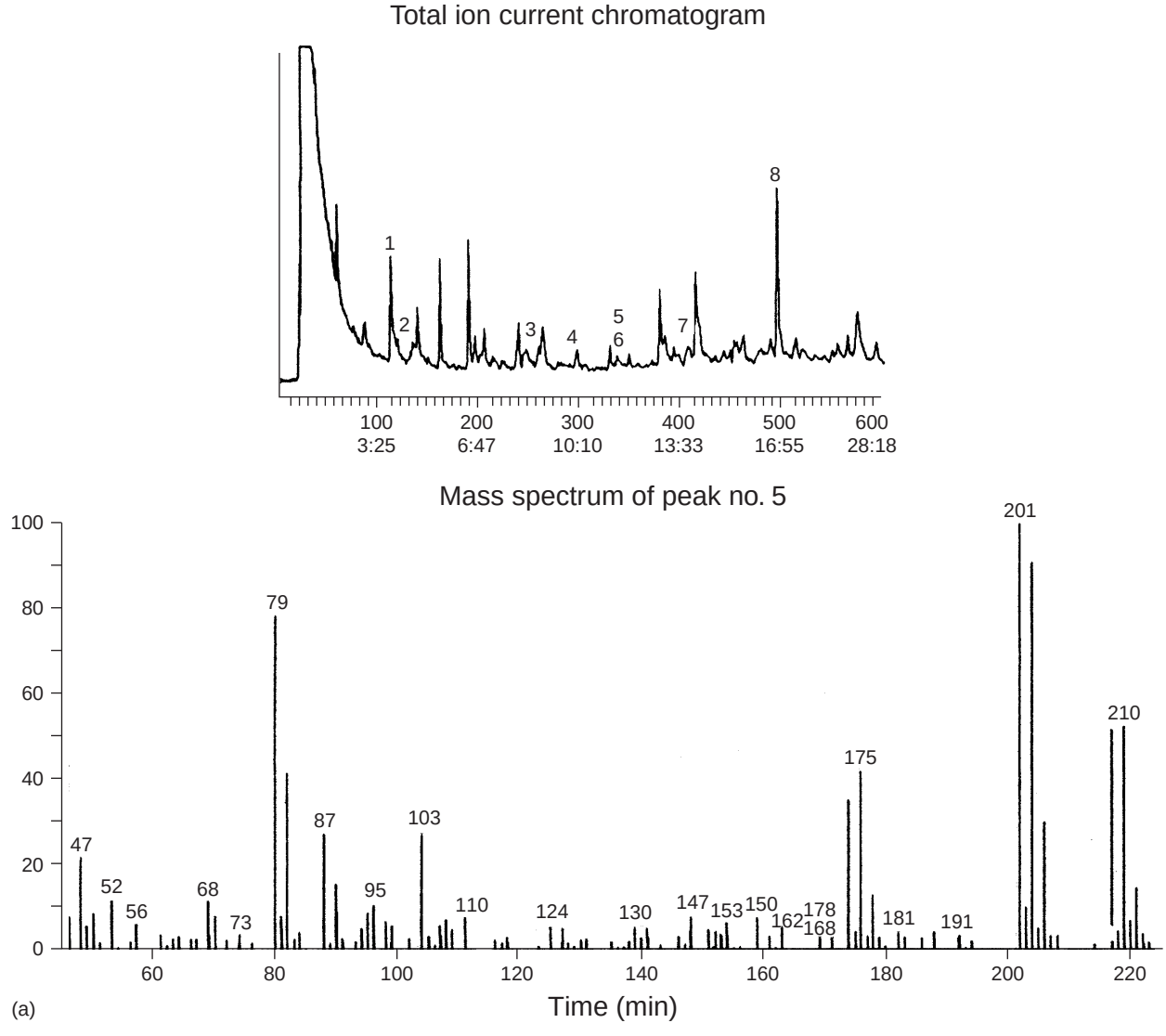

Figure 4 GC analysis of an extract of a pulp mill effluent by (a) GC/MS and by (b) GC/AED in the C-, H-, Cl-, S-, and O-selective modes.[19] (Reproduced from *Chromatographia*, **35**, 193–198, 1993, with permission from Vieweg.)

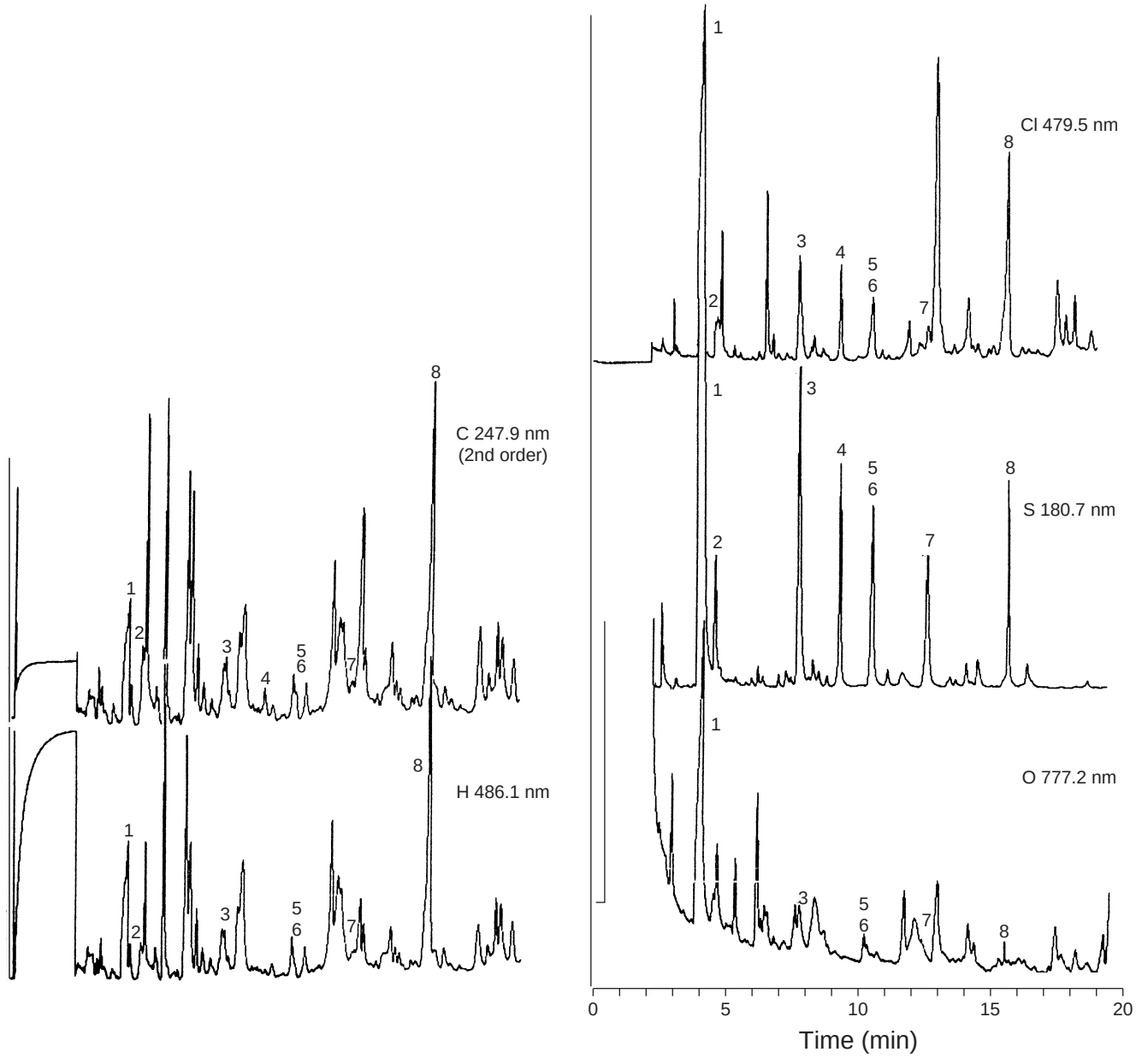

(b)

Figure 4 (*Continued*)

have been quantified by GC/AED utilizing universal calibration.[26,28] In addition, tin-organic compounds have been identified based on empirical formulas calculated from GC/AED results.[22]

5.2 Lead-containing Compounds

In addition to tin-organic compounds, GC/AED has been successfully applied for the determination of lead-containing compounds in water,[32–34] road dust,[34] and snow.[35] With emphasis on organolead compounds used as antiknock additives in gasoline, GC/AED was found to be substantially more sensitive than GC coupled with atomic absorption spectrometry,[32] the latter of which has been the most popular technique for Pb-selective detection in capillary GC. Thus, with GC/AED, organolead compounds were detected at the 0.1–2-ng L^{-1} level in water[32,33] and down to the femtogram per gram level in snow.[35] As for tin, GC/AED is very attractive also for lead compounds because of excellent selectivity and sensitivity.

5.3 Mercury-containing Compounds

Organomercury compounds have been utilized in paints, wood preservatives, paper, and pesticides, and this class of

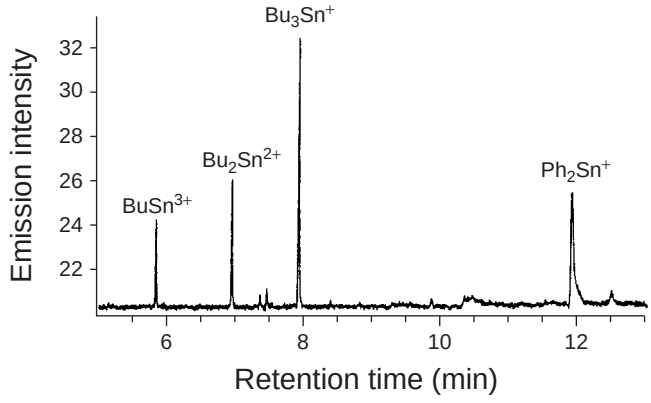

Figure 5 GC/AED analysis of an extract of mussel tissue in the Sn-selective mode.[24] (Reproduced by permission LC–GC International, Advanstar Communications Inc.)

compounds is of great importance within environmental monitoring. GC/AED operated in the Hg-selective mode has been used for the determination of organomercury compounds present in biological samples[36,37] and in water.[38] Compared with gas chromatography/electron capture detection (GC/ECD), which is in frequent use for the detection of organomercurals, the selectivity was superior utilizing GC/AED while DLs of the two detectors were comparable.[36] Thus, with GC/AED, organomercury compounds were detected down to the low nanogram per liter level in water.[38]

Owing to the simultaneous multielement capability of GC/AED, several reports have emerged where tin, lead, and mercury compounds have been determined simultaneously in biological material,[39] soil,[40] and water.[41] An example of this is illustrated in Figure 6, where both organotin and organolead compounds were found in river water.[41] This, in addition to excellent selectivity and sensitivity, suggests that GC/AED is a very attractive technique for the determination of organometallic compounds in environmental samples.

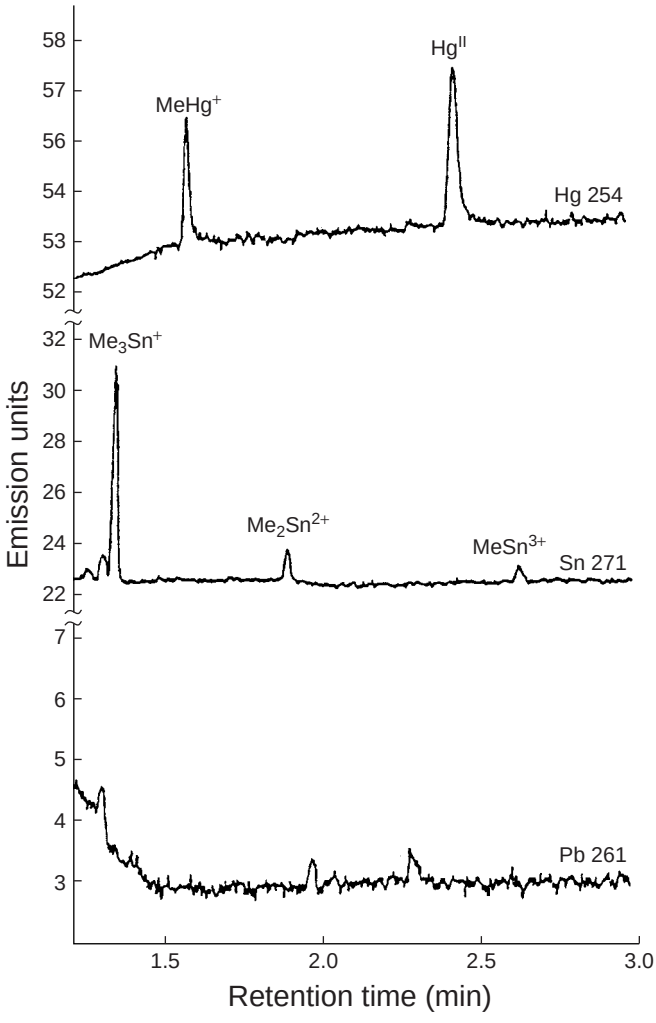

Figure 6 GC/AED analysis of an extract of river water in the Hg-, Sn-, and Pb-selective modes.[41] (Reproduced by permission of The Royal Society of Chemistry.)

5.4 Pesticides

Several publications have focused on the GC/AED determination of pesticides in environmental and agricultural samples. This broad class of compounds, which are produced industrially as agricultural chemicals, are particularly good candidates for GC/AED analysis since they are rich in heteroatoms such as chlorine, bromine, fluorine, phosphorus, sulfur, nitrogen, and oxygen. Thus, pesticides have been detected in different vegetables,[42–46] soil,[47,48] and water[48–53] utilizing GC/AED. An example is illustrated in Figure 7, where several pesticides were detected in an onion sample by GC/AED.[43] Excellent selectivity has been reported in the chlorine-selective mode for different pesticides present in agricultural products, while the established ECD (electron capture detection) and electrolytic conductivity detection (ELCD) devices suffered from matrix-related interferences.[42,43] Thus, a GC/AED screening method has been tentatively approved by the US Environmental Protection Agency for inclusion in the Update IV of the SW-846 methods.[48] In addition, GC/AED was found to provide accurate quantitation based on a single chlorine calibration curve (universal calibration). Unfortunately, the pesticide DL in the chlorine-selective mode of the first generation of the commercial GC/AED system was found to be a factor of 10^2–10^4 higher than for GC/ECD.[54] Thus, extensive analyte preconcentration was required to detect organochlorine pesticides by GC/AED at the 0.01 ppm level in agricultural matrices.[43] With the second generation of the commercial GC/AED system, the sensitivity

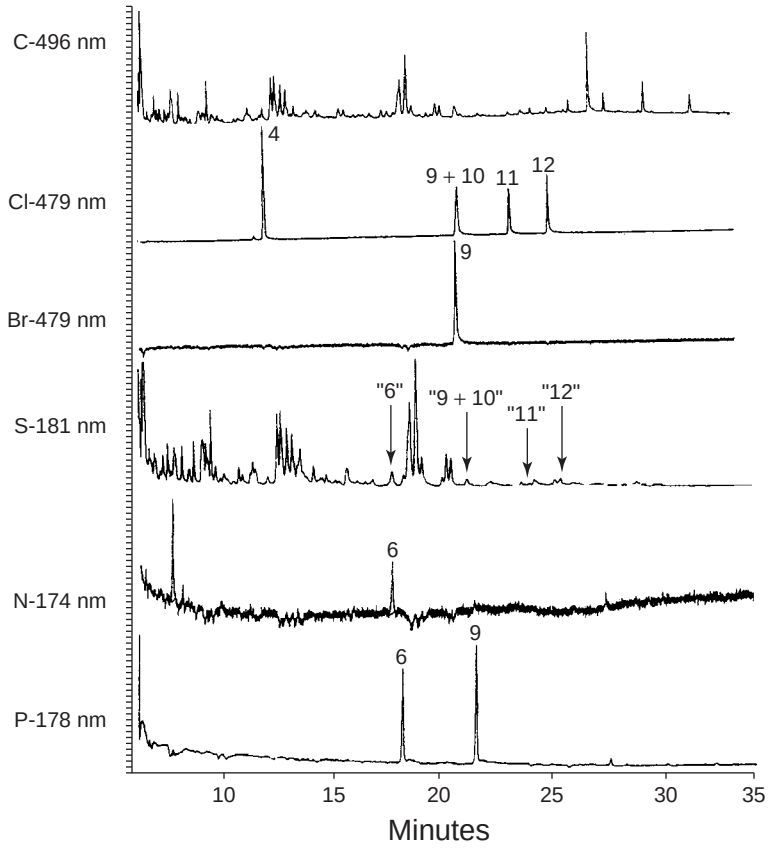

Figure 7 GC/AED analysis of an extract of onion in the C-, Cl-, Br-, S-, N-, and P-selective modes.[43] 4 = γ-HCH, 6 = parathion, 9 = bromophos ethyl, 10 = α-endosulfan, 11 = β-endosulfan, and 12 = endosulfan sulfate. (Reproduced from the *Journal of High Resolution Chromatography*, **16**, 539–548, 1993, with permission from Hüthig-GmbH.)

problem is expected to decrease, and GC/AED pesticide DLs (concentration units) have been improved by a factor of 100 compared with 1-ml injections by performing large volume injections (100 ml) with a solvent vapor exit in front of the analytical column.[49,52]

In addition to chlorine-selective detection, GC/AED in the bromine-, phosphorus-, sulfur-, and nitrogen-selective modes has been reported for the determination of pesticides. While bromine-selective detection is used only occasionally, several papers have emerged utilizing phosphorus-selective detection. The selectivity has been reported to be excellent in different agricultural matrices, and superior to results with both NPD and FPD detection.[42,43] The phosphorus DL of GC/AED was found to be comparable with FPD,[43] but significantly higher than for NPD.[54] Nevertheless, organophosphorus pesticides have been detected down to the 0.01-ppm level in plant foodstuffs by GC/AED utilizing 1 ml injections.[42,43] With 100 ml injections, even lower DLs have been obtained in the phosphorus-selective mode utilizing a solvent vapor exit in front of the analytical column.[51,52] Quantitative results from GC/AED were in excellent accord with data from NPD, and GC/AED was also found to enable accurate quantitation based on universal calibration in the phosphorus-selective mode.[48]

Some of the commercial pesticides also contain sulfur, and these may be selectively detected by GC/AED operated in the sulfur-selective mode. With the commercial GC/AED system, which is relatively sensitive towards sulfur (Table 1), DLs for sulfur-containing pesticides were lower than with FPD,[43] and the linearity problems of the FPD were eliminated.[55] Some pesticides also contain nitrogen and oxygen, but GC/AED is relatively insensitive towards these elements owing to atmospheric contamination of the plasma. Nevertheless, pesticides have been detected by GC/AED in several environmental samples utilizing nitrogen-selective detection. The method was limited to pesticides present above the 1.5–25-ng level in the final extracts, and the GC/AED DLs were found to be a factor of 10^3–10^4 higher than for NPD.[54]

5.5 Chlorinated Biphenyls and Phenols

Among the large number of polyhalogenated organic micropollutants included in environmental monitoring

programs, most attention has been focused on the GC/AED determination of PCBs. These compounds, which were used as, for example, electrical insulators and flame retardants until the 1980s, have been detected by GC/AED in human blood,[56] seal blubber,[56] beef tallow,[57] fish oil,[57] fish liver,[58] and marine sediments.[16,58] In addition to PCBs, their methylsulfonyl metabolites have also been detected in seal blubber by GC/AED.[56] With GC/AED, excellent selectivity has been reported for PCBs in the chlorine-selective mode (Figure 2).[16] Thus, GC/AED has been successfully applied for the determination of PCBs in highly contaminated marine sediments, while the established techniques of electron capture and mass spectrometric detection suffered from matrix interferences.[16] Additionally, quantification was significantly simplified by GC/AED because responses were dependent solely on the moles of chlorine present[16,57,58] (see also Table 2). Thus, individual PCB congeners and total PCB were quantified with a single chlorine calibration curve (universal calibration). Unfortunately, with the first generation of the commercial GC/AED instrument, individual PCBs were not detected below the 250–400-pg level.[16,57,58] This was a factor of 10^4 higher than reported with ECD, and a factor of 10^2 higher than with electron ionization mass spectrometric detection in the selected ion monitoring mode. Therefore, extensive analyte preconcentration was required to obtain DLs at the 0.15-ppm level for individual PCB congeners in fat.[57] Minor technical modifications of the commercial GC/AED instrument have been found to improve PCB DLs by a factor of approximately 5,[56] but even these values are relatively high for PCB applications. Thus, while GC/AED may be utilized for contaminated samples, the technique is currently not an alternative for the monitoring of PCB background levels.

In addition to PCBs, some GC/AED interest has been focused on the determination of chlorophenols (CPs) which originate as industrial byproducts. With principal attention on water, GC/AED was found to provide excellent selectivity for CPs in the chlorine-selective mode, and CPs were determined in water below the 0.5-ppb level.[59,60]

5.6 Other Compounds Containing Nonmetallic Elements

Above, attention was focused on organic micropollutants containing metallic elements (Sn, Pb, Hg), pesticides, PCBs, and CPs. For all of these contaminants, substantial research has been documented within the literature utilizing the commercial GC/AED system, and GC/AED has been demonstrated to be an attractive alternative to existing methods. In addition to the above mentioned components, several other organic micropollutants have been detected by GC/AED, including polychlorinated dibenzodioxins (PCDDs), volatile organochlorines, trichloroacetic acid, chlorinated byproducts of humic and fulvic acids, chlorinated aromatics, polybrominated biphenyls (PBBs), polybrominated diphenylethers (PBDEs), iodated herbicides, volatile sulfur compounds, nitrosubstituted aromatics, and nitro musks. More information on these applications as well as others can be found in a review article concerned with environmental applications of GC/AED.[61]

6 CURRENT STATUS AND FUTURE DIRECTIONS

The literature reviewed in the present paper has demonstrated the great potential of GC/AED for analysis of organic micropollutants. Owing to the elemental capability, GC/AED enables rapid screening by elemental content of environmental samples, provides important structural information complementary to mass spectral data, and enables simplified quantitation based on universal calibration with a single elemental calibration curve. Particularly for environmental analysis, which frequently is complicated by matrix constituents, the very high elemental selectivity of GC/AED may reduce both interference and the demands of sample cleanup. However, as pointed out above, DLs of GC/AED are relatively high for some elements and extensive analyte preconcentration may be crucial. This, in combination with the lack of low-price commercially available instrumentation, currently limits the propagation of GC/AED in environmental laboratories.

For the future, therefore, it is hoped that the number of companies committed to producing GC/AED instrumentation will increase to reduce the price level of this highly versatile technology. With some further effort to improve DLs and to include the concept in standard procedures for the monitoring of organic micropollutants, GC/AED may rapidly become a ''work horse'' in laboratories concerned with environmental chemistry.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
CP	Chlorophenol
DCP	Direct Current Plasma
DL	Detection Limit
ECD	Electron Capture Detection
ELCD	Electrolytic Conductivity Detection
FPD	Flame Photometric Detection
GC	Gas Chromatography

GC/ECD	Gas Chromatography/Electron Capture Detection
ICP	Inductively Coupled Plasma
MIP	Microwave Induced Plasma
MS	Mass Spectrometry
NPD	Nitrogen–Phosphorus Detection
O-FID	Oxygen-flame Ionization Detection
PBBs	Polybrominated Biphenyls
PBDEs	Polybrominated Diphenylethers
PCB	Polychlorinated Biphenyl
PCDD	Polychlorinated Dibenzodioxin
SCD	Sulfur Chemiluminescence Detection

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Organometallic Compound Analysis in Environmental Samples • Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis

Atomic Spectroscopy **(Volume 11)**
Microwave-induced Plasma Systems in Atomic Spectroscopy

REFERENCES

1. A.J. McCormick, S.C. Tong, W.D. Cooke, 'Sensitive Selective Gas Chromatography Detector Based on Emission Spectrometry of Organic Compounds', *Anal. Chem.*, **37**, 1470–1476 (1965).
2. B.J. Richardson, C. Narain, 'Some Notes on the Direct Current Discharge Detector', *J. Chromatogr.*, **26**, 248–251 (1967).
3. C.I.M. Beenakker, 'A Cavity for Microwave-induced Plasmas Operated in Helium and Argon at Atmospheric Pressure', *Spectrochim. Acta, Part B*, **31**, 483–486 (1976).
4. G.W. Rice, A.P. D'Silva, V.A. Fassel, 'A New He Discharge-afterglow and its Application as a Gas Chromatographic Detector', *Spectrochim. Acta, Part B*, **40**, 1573–1584 (1985).
5. R.J. Skelton, K.E. Markides, P.B. Farnsworth, M.L. Lee, 'Multi-element Selective Radio Frequency Plasma Detector for Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **11**, 75–81 (1988).
6. S. Pedersen-Bjergaard, T. Greibrokk, 'On-column Bromine- and Chlorine-selective Detection for Capillary Gas Chromatography Using a Radio Frequency Plasma', *Anal. Chem.*, **65**, 1998–2002 (1993).
7. R.B. Constanzo, E.F. Barry, 'Alternating Current Plasma Detector for Selective Mercury Detection in Gas Chromatography', *Anal. Chem.*, **60**, 826–829 (1988).
8. D. Huang, D.C. Liang, M.W. Blades, 'Capacitively Coupled Plasma Detector for Gas Chromatography', *J. Anal. At. Spectrom.*, **4**, 789–791 (1989).
9. G. Knapp, E. Leitner, M. Michaelis, B. Platzer, A. Schalk, 'Element Specific GC-detection by Plasma Atomic Emission Spectroscopy – A Powerful Tool in Environmental Analysis', *Int. J. Environ. Anal. Chem.*, **38**, 369–378 (1990).
10. R.J. Loyd, R.M. Barnes, P.C. Uden, W.G. Elliott, 'Direct Current Atmospheric Pressure Argon Plasma Emission Echelle Spectrometer as a Specific Metal Gas Chromatographic Detector', *Anal. Chem.*, **50**, 2025–2029 (1978).
11. D.L. Windsor, M.B. Denton, 'Evaluation of Inductively Coupled Plasma Optical Emission Spectrometry as a Method for the Elemental Analysis of Organic Compounds', *Appl. Spectrosc.*, **32**, 366–371 (1978).
12. J. Bonnekessel, M. Klier, 'Daten-Erfassung und Verarbeitung Eines Mikro-Welleplasmadetektors fur die Gas-Chromatographie', *Anal. Chim. Acta*, **103**, 29–42 (1978).
13. B.D. Quimby, J.J. Sullivan, 'Evaluation of a Microwave Cavity, Discharge Tube, and Gas Flow System for Combined Gas Chromatography/Atomic Emission Detection', *Anal. Chem.*, **62**, 1027–1033 (1990).
14. J.J. Sullivan, B.D. Quimby, 'Characterization of a Computerized Photodiode Array Spectrometer for Gas Chromatography/Atomic Emission Spectrometry', *Anal. Chem.*, **62**, 1034–1043 (1990).
15. S. Pedersen-Bjergaard, T. Greibrokk, 'Evaluation of a Low-resolution Near-IR Monochromator for Atomic Emission Detection in Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **18**, 9–14 (1995).
16. B.D. Quimby, P.A. Larson, P.C. Dryden, 'A Comparison of the HP G2350A AED vs. HP 5921A AED for Average Values of MDL and Selectivity for Selected Elements', Application Note 228–363, Hewlett-Packard Company, Wilmington, DE, 1996.
17. S. Pedersen-Bjergaard, S.I. Semb, J. Vedde, E.M. Brevik, T. Greibrokk, 'Comparison of GC–ECD, GC–MS and GC–AED for the Determination of Polychlorinated Biphenyls in Highly Contaminated Marine Sediments', *Chromatographia*, **43**, 44–52 (1996).

18. S. Pedersen-Bjergaard, T.N. Asp, J. Vedde, T. Greibrokk, 'Molecular Formula Determination of Halogenated Compounds in Environmental Samples Using Gas Chromatography and Atomic Emission Detection', *J. Microcolumn Sep.*, **4**, 163–170 (1992).
19. S. Pedersen-Bjergaard, T.N. Asp, J. Vedde, G.E. Carlberg, T. Greibrokk, 'Identification of Chlorinated Sulfur Compounds in Pulp Mill Effluents by GC–MS and GC–AED', *Chromatographia*, **35**, 193–198 (1993).
20. S. Pedersen-Bjergaard, S.I. Semb, J. Vedde, E.M. Brevik, T. Greibrokk, 'Environmental Screening by Capillary Gas Chromatography Combined with Mass Spectrometry and Atomic Emission Spectroscopy', *Chemosphere*, **32**, 1103–1115 (1996).
21. B.F. Scott, Y.K. Chau, A. Rais-Firouz, 'Determination of Butyltin Species by GC/Atomic Emission Spectroscopy', *Appl. Organomet. Chem.*, **5**, 151–157 (1991).
22. R. Lobinski, W.M.R. Dirkx, M. Ceulemans, F. Adams, 'Optimization of Comprehensive Speciation of Organotin Compounds in Environmental Samples by Capillary Gas Chromatography Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Chem.*, **64**, 159–165 (1992).
23. Y. Liu, V. Lopez-Avila, M. Alcaraz, W.F. Beckert, 'Off-line Complexation/Supercritical Fluid Extraction and Gas Chromatography with Atomic Emission Detection for the Determination and Speciation of Organotin Compounds in Soil and Sediments', *Anal. Chem.*, **66**, 3788–3796 (1994).
24. M. Ceulemans, F. Adams, 'GC-based Hyphenated Techniques Applied to the Speciation Analysis of Organometals', *LC-GC Int.*, **7**, 694–697 (1994).
25. J. Szpunar, V.O. Schmitt, R. Lobinski, J.-L. Monod, 'Rapid Speciation of Butyltin Compounds in Sediments and Biomaterials by Capillary Gas Chromatography/Microwave-induced Plasma Atomic Emission Spectrometry After Microwave-assisted Leaching/Digestion', *J. Anal. At. Spectrom.*, **11**, 193–199 (1996).
26. J.A. Stab, W.P. Cofino, B. Van Hattum, U.A.T. Brinkman, 'Comparison of GC/MSD and GC/AED for the Determination of Organotin Compounds in the Environment', *Fresenius' J. Anal. Chem.*, **347**, 247–255 (1993).
27. W.M.R. Dirkx, R. Lobinski, F. Adams, 'Speciation Analysis of Organotin in Water and Sediments by Gas Chromatography with Optical Spectrometric Detection after Extraction Separation', *Anal. Chim. Acta*, **286**, 309–318 (1994).
28. T.J. Gremm, F.H. Frimmel, 'Determination of Organotin Compounds in Aqueous Samples by means of HPGC–AED', *Water Res.*, **26**, 1163–1169 (1992).
29. T.M. Dowling, P.C. Uden, 'Alkyltin Speciation in Sea Water with On-line Hydride Conversion and Gas Chromatography/Atomic Emission Detection', *J. Chromatogr.*, **644**, 153–160 (1993).
30. M. Ceulemans, R. Lobinski, W.M.R. Dirkx, F. Adams, 'Rapid Sensitive Speciation Analysis of Butyl- and Phenyltin Compounds in Water by Capillary Gas Chromatography Atomic Emission Spectrometry (GC/AES) After In Situ Ethylation and In-liner Preconcentration', *Fresenius' J. Anal. Chem.*, **347**, 256–262 (1993).
31. M. Ceulemans, J. Szpunar-Lobinska, W.M.R. Dirkx, R. Lobinski, F. Adams, 'Speciation Analysis of Organotin in the River Scheldt by Capillary Gas Chromatography Atomic Emission Spectrometry (GC/AES)', *Int. J. Environ. Anal. Chem.*, **52**, 113–125 (1993).
32. R. Lobinski, F. Adams, 'Sensitive Speciation Analysis of Lead in Environmental Waters by Capillary Gas Chromatography Microwave-induced Plasma Atomic Emission Spectrometry', *Anal. Chim. Acta*, **262**, 285–297 (1992).
33. R. Lobinski, F. Adams, 'Ultratrace Speciation Analysis of Organolead in Water by Gas Chromatography/Atomic Emission Spectrometry After In-liner Preconcentration', *J. Anal. At. Spectrom.*, **7**, 987–992 (1992).
34. C. Witte, J. Szpunar-Lobinska, R. Lobinski, F. Adams, 'Determination of Trimethyllead in Rainwater and Road Dust by Capillary GC MIP–AE Spectrometry after In Situ Ethylation and Extraction', *Appl. Organomet. Chem.*, **8**, 621–627 (1994).
35. R. Lobinski, C.F. Boutron, J.-P. Candelone, S. Hong, J. Szpunar-Lobinska, F. Adams, 'Speciation Analysis of Organolead Compounds in Greenland Snow at the Femtogram-per-gram Level by Capillary Gas Chromatography/Atomic Emission Spectrometry', *Anal. Chem.*, **65**, 2510–2515 (1993).
36. A.M. Carro-Diaz, R.A. Lorenzo-Ferreira, R. Cela-Torrijos, 'Speciation of Organomercurals in Biological and Environmental Samples by Gas Chromatography with Microwave-induced Plasma Atomic Emission Detection', *J. Chromatogr., Part A*, **683**, 245–252 (1994).
37. M.K. Donais, P.C. Uden, M.M. Schantz, S.A. Wise, 'Development, Validation, and Application of a Method for Quantification of Methylmercury in Biological Marine Materials Using Gas Chromatography Atomic Emission Detection', *Anal. Chem.*, **68**, 3859–3866 (1996).
38. M.L. Mena, C.W. McLeod, P. Jones, A. Withers, V. Minganti, R. Capelli, P. Quevauviller, 'Microcolumn Preconcentration and Gas Chromatography/Microwave Induced Plasma/Atomic Emission Spectrometry (GC/MIP/AES) for Mercury Speciation in Waters', *Fresenius' J. Anal. Chem.*, **351**, 456–460 (1995).
39. V. Minganti, R. Capelli, R. De Pellegrini, 'Evaluation of Different Derivatization Methods for the Multielement Detection of Hg, Pb and Sn Compounds by Gas Chromatography/Microwave Induced Plasma/Atomic Emission Spectrometry in Environmental Samples', *Fresenius' J. Anal. Chem.*, **351**, 471–477 (1995).
40. Y. Liu, V. Lopez-Avila, M. Alcaraz, 'Simultaneous Determination of Organotin, Organolead, and Organomercury Compounds in Environmental Samples Using Capillary Gas Chromatography with Atomic Emission Detection', *J. High. Resolut. Chromatogr.*, **17**, 527–536 (1994).

41. M. Ceulemans, F. Adams, 'Integrated Sample Preparation and Speciation Analysis for the Simultaneous Determination of Methylated Species of Tin, Lead and Mercury in Water by Purge-and-trap Injection-capillary Gas Chromatography/Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **11**, 201–206 (1996).
42. S. Mark Lee, P.L. Wylie, 'Comparison of the Atomic Emission Detector to Other Element-selective Detectors for the Gas Chromatographic Analysis of Pesticide Residues', *J. Agric. Food Chem.*, **39**, 2192–2199 (1991).
43. H.-J. Stan, M. Linkerhagner, 'Capillary Gas Chromatography – Atomic Emission Detection: A Useful Instrumental Method in Pesticide Residue Analysis of Plant Foodstuffs', *J. High Resolut. Chromatogr.*, **16**, 539–548 (1993).
44. K.-C. Ting, P. Kho, 'GC/MIP/AED Method for Pesticide Residue Determination in Fruits and Vegetables', *J.-Assoc. Off. Anal. Chem.*, **74**, 991–998 (1991).
45. P.L. Wylie, R. Oguchi, 'Pesticide Analysis by Gas Chromatography with a Novel Atomic Emission Detector', *J. Chromatogr.*, **517**, 131–142 (1990).
46. Z.V. Skopec, R. Clark, P.M.A. Harvey, R.J. Wells, 'Analysis of Organophosphorus Pesticides in Rice by Supercritical Fluid Extraction and Quantitation Using an Atomic Emission Detector', *J. Chromatogr. Sci.*, **31**, 445–449 (1993).
47. J.L. Bernal, M.J. Del Nozal, M.T. Martin, J.J. Jimènez, 'Possibilities of Gas Chromatography/Atomic Emission Detection in Pesticide Multiresidue Analysis – Application to Herbicide Analysis in Soils', *J. Chromatogr., Part A*, **754**, 245–256 (1996).
48. N.L. Olson, R. Carrell, R.K. Cummings, R. Rieck, 'Gas Chromatography with Atomic Emission Detection for Pesticide Screening and Confirmation', *LC-GC Int.*, **12**, 142–154 (1994).
49. E.C. Goosens, D. De Jong, G.J. De Jong, F.D. Rinkema, U.A.T. Brinkman, 'Continuous Liquid–Liquid Extraction Combined On-line with Capillary Gas Chromatography/Atomic Emission Detection for Environmental Analysis', *J. High Resolut. Chromatogr.*, **18**, 38–44 (1995).
50. H.-J. Stan, M. Linkerhagner, 'Multimethod Applying GC/AED to Water Samples', in *Analysis of Pesticides in Ground and Surface Water II – Latest Developments and State-of-the-art of Multiple Methods*, ed. H.-J. Stan, Chemistry of Plant Protection, Springer, Berlin, 59–90, Vol. 12, 1995.
51. T. Hankemeier, A.J.H. Louter, F.D. Rinkema, U.A.T. Brinkman, 'On-line Coupling of Solid-phase Extraction and Gas Chromatography with Atomic Emission Detection for Analysis of Trace Pollutants in Aqueous Samples', *Chromatographia*, **40**, 119–124 (1995).
52. F.D. Rinkema, A.J.H. Louter, U.A.T. Brinkman, 'Large-volume Injections in Gas Chromatography/Atomic Emission Detection: An Approach for Trace-level Detection in Water Analysis', *J. Chromatogr., Part A*, **678**, 289–297 (1994).
53. R. Eisert, K. Levsen, G. Wunsch, 'Element-selective Detection of Pesticides by Gas Chromatography/Atomic Emission Detection and Solid-phase Microextraction', *J. Chromatogr., Part A*, **683**, 175–183 (1994).
54. R. Eisert, K. Levsen, G. Wunsch, 'Analysis of Polar Thermally Labile Pesticides Using Different Solid-phase Extraction (SPE) Materials with GC and HPLC Techniques', *Int. J. Environ. Anal. Chem.*, **58**, 103–120 (1994).
55. F. David, P. Sandra, 'Comparison of the Sensitivity of the FPD and the AED for the Analysis of Thiazole', in *Proc. 13th Internat. Symp. Capillary Chromatogr.*, ed. P. Sandra, Riva del Garda, Italy, May 1991, Huthig, Heidelberg, 1479–1484, 1991.
56. K. Janak, E. Grimvall, C. Ostman, A. Colmsjo, 'Gas Chromatography–Atomic Emission Detection (GC–AED) Set-up for Biomonitoring of PCBs and Methylsulfonyl-PCBs', *J. Microcolumn Sep.*, **6**, 605–616 (1994).
57. J. Hajslova, P. Cuhra, M. Kempny, J. Poustka, K. Holadova, V. Kocourek, 'Determination of Polychlorinated Biphenyls in Biotic Matrices using Gas Chromatography/Microwave Induced Plasma Atomic Emission Spectrometry', *J. Chromatogr., Part A*, **699**, 231–239 (1995).
58. S. Pedersen-Bjergaard, S.I. Semb, E.M. Brevik, T. Greibrokk, 'Capillary Gas Chromatography Combined with Atomic Emission Detection for the Analysis of Polychlorinated Biphenyls', *J. Chromatogr., Part A*, **723**, 337–347 (1996).
59. M.I. Turnes, I. Rodriguez, M.C. Mejuto, R. Cela, 'Determination of Chlorophenols in Drinking Water Samples at the Subnanogram per Millilitre Level by Gas Chromatography with Atomic Emission Detection', *J. Chromatogr., Part A*, **683**, 21–29 (1994).
60. I. Rodriguez, M.I. Turnes, M.J. Mejuto, R. Cela, 'Determination of Chlorophenols at the Sub-ppb Level in Tap Water Using Derivatization, Solid-phase Extraction and Gas Chromatography with Plasma Atomic Emission Detection', *J. Chromatogr., Part A*, **721**, 297–304 (1996).
61. S. Pedersen-Bjergaard, T. Greibrokk, 'Environmental Applications of Capillary Gas Chromatography Coupled with Atomic Emission Detection – A Review', *J. High Resolut. Chromatogr.*, **19**, 597–607 (1996).

Gas Chromatography with Selective Detectors for Amines

Hiroyuki Kataoka
Okayama University, Okayama, Japan

1 Introduction **1**
1.1 Hazardous Amines from Environmental Sources 1
1.2 Objective and Scope 2
2 Selective Detectors for Analysis of Amines by Gas Chromatography **3**
2.1 Amine Analysis by Gas Chromatography 3
2.2 Thermionic Detector 3
2.3 Flame Photometric Detector 6
2.4 Electron-capture Detector 7
2.5 Chemiluminescence Detector 9
2.6 Mass-selective Detector 10
2.7 Electrolytic Conductivity Detector 13
3 Application in Environmental Analysis **14**
3.1 General Techniques for Sample Preparation 14
3.2 Aliphatic Amines 15
3.3 Aromatic Amines 17
3.4 *N*-Nitrosamines 17
3.5 Heterocyclic Amines 18
4 Conclusions **19**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **23**

The presence, identity and quantity of amines in the environment should be established to evaluate possible health hazards. In many cases, environmental amines are present at very low concentration and are often found among a myriad of other compounds from which they must be separated and identified. Therefore, analytical methodology for the isolation and quantification of these amines needs selective and sensitive detection. The present article is concerned with the utilization of selective gas chromatographic detectors for amine analysis and their application to the determination of amines in environmental samples.

1 INTRODUCTION

1.1 Hazardous Amines from Environmental Sources

As environmental issues and global environmental change are generating an increasing amount of attention worldwide, the occurrence and toxicity of hazardous chemicals in the biosphere have received a great deal of attention in recent years. As technology progressed and man's appetite for a carefree life expanded, the problems of environmental pollution became more complex and difficult to alleviate.[1] The three main categories of pollution are water pollution, air pollution and soil pollution, and knowledge of these types of pollution has become an urgent requirement for the protection of the terrestrial environment. The environment contains a variety of naturally occurring and anthropogenic pollutants to which humans are exposed. Moreover, every year many new substances are synthesized that differ radically from the natural products that exist in biosystems. Many of these compounds are not biodegradable and will thus progressively pollute the environment. Among the many environmental pollutants, the occurrence and determination of amines have recently received a great deal of attention. These amines occur in a number of ambient environments such as air, water, soil and foods, and have become a source of serious social and hygienic problems.[2,3]

Aliphatic and aromatic mono-, di- and polyamines are naturally occurring compounds formed as metabolic products in microorganisms, plants and animals, in which the principal routes of amine formation include the decarboxylation of amino acids, amination of carbonyl compounds and degradation of nitrogen-containing compounds. These amines are considered to play a significant role as storage sources of nitrogen and precursors for the synthesis of hormones, alkaloids, nucleic acids and proteins. Amines are also widely used as raw materials or at an intermediate stage in the manufacture of industrial chemicals.[2–6] In particular, aromatic amines such as aniline and its substituted analogs, phenylenediamine and diphenylamine, have been widely used as industrial intermediates[2,7–9] in the manufacture of carbamate and urethane pesticides, dyestuffs, cosmetics and medicines. These amines have also been employed in the rubber industry as antioxidants and antiozonants[5] and as components in epoxy and polyurethane polymers.[4,10] An additional source for amines in the environment is the abiotic and biotic degradation of animal waste, domestic waste and industrial products.[9,11,12] Many of these amines have been discharged into the atmosphere[10,13–30] and water[18,31–48] from anthropogenic sources such as cattle feedlots and near livestock buildings,[13] waste

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

incineration, sewage treatment, automobile exhaust, cigarette smoke,[11,25,49–54] soil,[22,23,38,41,55] foods [56–61] and various industries.[2,3,6,7,62] These amines have also been detected in biological samples.[12,35,63–66] Many amines have an unpleasant smell and are hazardous to health, i.e. as sensitizers and irritants to the skin, eye, mucous membrane and respiratory tract.[67] Some amines are also suspected to be allergenic and mutagenic or carcinogenic substances owing to their adsorption tendency in tissues.[68,69] In particular, the toxicity of aromatic amines to mammals and fish is well established,[69,70] and many of these amines are known to be highly mutagenic and carcinogenic and to form adducts with proteins and DNA.[7,68,71–79] Several polycyclic aromatic amines such as benzidine, 4-aminobiphenyl and 2-naphthylamine have been also classified by the International Agency for Research on Cancer (IARC) as known human carcinogens.

Aliphatic and aromatic amines not only are toxic in themselves but also become toxic *N*-nitrosamines through chemical reactions with nitrosating agents such as nitrite or nitrate.[80–83] In general, the nitrosation of amines occurs in the human diet, the environment and in vivo in the stomach or small intestine of experimental animals.[84–87] The reaction of nitrosating agents with primary amines produces short-lived alkylating species that react with other compounds in the matrix to generate products (mainly alcohols) devoid of toxic activity in the relevant concentrations. The nitrosation of secondary amines leads to the formation of stable *N*-nitrosamines whereas that of tertiary amines produces a range of labile *N*-nitroso products. Although the reaction pathways are uncertain, it has also been demonstrated in model experiments that amines react with NO_x and OH radicals in air to form nitrosamines and nitramines.[62,88] Moreover, it is considered that volatile *N*-nitrosamines in cigarette smoke are formed from the decarboxylation of *N*-nitrosoamino acids, pyrolytic nitrosation of secondary amines and concentrated decarboxylation/nitrosation of amino acids during combustion processes.[82] *N*-Nitrosamines are widely distributed in foodstuffs[61,81,89–93] and various human environments such as indoor and outdoor air,[85,94,95] combustion smoke,[52,82,96–103] water,[104–108] sediment,[109] household dishwashing liquids,[110] foods,[81,89,91–93,111] rubber products,[111–119] metalworking fluids,[120] drug formulations[121,122] and agricultural chemical formulations.[110,123,124] Furthermore, some *N*-nitrosamines have been detected in human gastric juice, saliva and cervical mucus.[125,126] *N*-Nitrosamines represent a major class of important chemical mutagens, carcinogens, teratogens and immunotoxic agents, which have been described as a serious hazard to human health.[68,127–130] Recent developments in environmental carcinogenesis have demonstrated that *N*-nitrosamines lead to a wide variety of tumors in many animals, which are likely to result in the formation of DNA adducts.[68,78,131,132] The toxicity of *N*-nitrosamines can be manifested even at the micrograms per kilogram level.

A new series of heterocyclic amines formed during heating of amino acids, proteins, creatinine and sugars are potent mutagens in the Ames/*Salmonella* assay.[133–140] Many of these mutagenic heterocyclic amines have been isolated and identified not only from various proteinaceous foods [141–155] including cooked meats and fish, but also from environmental components such as airborne particles and diesel-exhaust particles,[156,157] combustion smoke[158–164] indoor air,[165–167] cooking fumes,[168–170] rainwater[165,166,171] and river water.[172] Moreover, some mutagenic heterocyclic amines have been detected in biological samples such as urine, plasma, bile and feces.[150,173–176] These facts suggest that heterocyclic amines may be emitted into the atmosphere through combustion of various materials such as foods, wood, grass, garbage and petroleum, and discharged into the water through domestic waste and human waste, although their mechanisms have not been determined. Some heterocyclic amines have much higher mutagenic activity than typical mutagens/carcinogens such as aflatoxin B1, AF-2 and benzo[*a*]pyrene and have been verified to be carcinogenic in rats and mice.[133–137,139,177,178] These compounds induced tumors in the liver, small and large intestine, Zymbal gland, clitoral gland, skin, oral cavity and mammary gland in rats, and the liver, forestomach, lung, hematopoietic system, lymphoid tissue and blood vessels in mice. 2-Amino-3-methylimidazo[4,5-*f*]quinoline (IQ) was also found to be carcinogenic in the monkey, inducing hepatocellular carcinomas.[179,180] Moreover, recent investigations revealed that heterocyclic amines also possess cardiotoxic effects[178] and various pharmacotoxicological activities such as convulsant activity[181,182] and potent inhibitory effects on platelet function and dopamine metabolism. [182–186] However, significant risks to human health and the environment through long-term exposure and bioaccumulation of heterocyclic amines are scientifically unclear.

1.2 Objective and Scope

The presence, identity and quantity of amines in the environment should be established to evaluate possible health hazards. In many cases, environmental amines are present at very low concentration and are often found among a myriad of other compounds from which they must be separated and identified. Therefore, analytical methodology for the isolation and quantification of these amines needs selective and sensitive detection. The present article is concerned with the utilization of the

selective gas chromatography (GC) detectors for amine analysis and their application for the determination of amines in environmental samples. The article consists of two parts. In the first part (section 2), general aspects of the selective detection of amines by GC are surveyed according to the detector types. In the second part (section 3), the applications of selective GC methods in environmental analysis are considered according to the amine types. The article covers not only aliphatic and aromatic primary, secondary and tertiary amines but also *N*-nitrosamines and heterocyclic amines that often occur in various environments as highly toxic compounds. References that have appeared in the past two decades are cited. General aspects of GC detectors have been reviewed in several books.[187–190] General aspects of environmental analysis by GC have also been detailed in several books[190–192] and reviews.[2,3]

2 SELECTIVE DETECTORS FOR ANALYSIS OF AMINES BY GAS CHROMATOGRAPHY

2.1 Amine Analysis by Gas Chromatography

GC has been widely used for amine analysis because of its inherent advantages of simplicity, high resolving power, high sensitivity, short analysis time and low cost. However, GC analysis of free amines generally has some inherent problems related to the difficulty in handling low-molecular-mass amines because of their high water solubility and high volatility. Therefore, these amines are difficult to extract from water and are not easily chromatographed owing to their polarity. Furthermore, amines are tend to be adsorbed and decomposed on the columns, and readily give tailed elution peaks, ghosting phenomena and low detector sensitivity. The adsorption tendency in the analytical systems, i.e. in sample vessels, injection system, glass wool and GC column, decreases in the order primary > secondary > tertiary amines, and it is generally more difficult to chromatograph aliphatic than aromatic amines. A common method of overcoming these problems is to convert such polar compounds into relatively nonpolar derivatives more suitable for GC analysis. Derivatization methods have been employed to reduce the polarity of the amino group and improve GC properties. Derivatization reactions, often selective for amine type (primary, secondary, tertiary), have also been used to improve the detection and separation of these amines. A number of derivatives such as acyl, silyl, dinitrophenyl, permethyl, Schiff base, carbamate, sulfonamide and phosphonamide compounds have been used for this purpose. For example, introducing halogen- and phosphorus-containing groups in the molecule enhances the response of the electron-capture detector (ECD) and flame photometric detector (FPD), respectively. These derivatization reactions were described in detail in previous review.[3] For amine analysis, a wide variety of detectors such as the thermionic detector, FPD, ECD, electrolytic conductivity detector (ELCD) and chemiluminescent detector (CLD) can be used, and offer increased selectivity for specific amines. Furthermore, the combined technique of gas chromatography/mass spectrometry (GC/MS) can provide structure information for the unequivocal identification of amines and these amines can be determined with mass-selective detector (MSD) based on a selected-ion monitoring (SIM). By using these detectors, subnanogram detection limits can be achieved. GC detectors for environmental amine analysis are summarized in Table 1. Although most detectors respond directly to amines, some detectors, such as the FPD and ECD, need conversion of amines into suitable derivatives prior to detection.

2.2 Thermionic Detector

The thermionic detector, which evolved from the flame ionization detector (FID), is particularly suitable for the analysis of compounds containing nitrogen and phosphorus. There are two types of thermionic detectors depending on whether a flame is present in or absent in the detector system. In general, a flame-type thermionic detector is called a flame thermionic detector (FTD) or alkali flame ionization detector (AFID), and a flameless-type detector is called a thermionic specific detector (TSD) or nitrogen–phosphorus detector (NPD). The flame type can also detect halogen-containing compounds, but the flameless type selectively respond only to N- and P-containing compounds. A diagram of a Hewlett-Packard NPD is shown in Figure 1.

Although the mechanism of the response of the thermionic detector for N- and P-containing compounds is not fully understood, a widely accepted explanation is as follows. The normal FID ion collection assembly is replaced with a collector that has a small alumina cylinder coated with a rubidium salt (active element), which is heated electrically, creating a thermionic source. In this environment, N- and P-containing organic compounds are ionized. The detector collects the ions and measures the resulting current with an electrometer. The flame in the NPD is not ignited, but rather is adjusted to give a low-temperature plasma. This plasma is not used to heat the collector, but only for ion production and dissociation as the organic compounds eluted from the column. For N-containing compounds, pyrolysis takes place, producing intermediate stable CN radicals and then the radicals formed take electrons from the excited atom of the alkali

Table 1 Selective GC detectors for environmental amine analysis

GC detector	Amines	Matrix	References
Thermionic detector	Aliphatic amines	Air	Kuwata et al.[13]
	Aliphatic and aromatic amines	Working atmosphere	Audunsson and Mathiasson,[14,15] Skarping et al.[16]
	Aliphatic amines	Ambient air, rain water	Gronberg et al.[17,18]
	Aromatic amines	Workplace atmosphere	Becher[19]
	Anilines	Wastewater	Riggin et al.[31]
	Aromatic amines	Combustion smoke	Dalene and Skarping ,[49] Kataoka et al.[50]
	Dinitroanilines	Soil, plant, air	Garcia-Valcarcel et al.[55]
	N-Nitrosodimethylamine	Herbicide formulation	Scharfe and McLenagham[123]
	Heterocyclic tertiary amines	Air	Andersson and Andersson[20]
	Heterocyclic amines	Combustion smoke	Kataoka et al.[158]
FPD	Aliphatic amines	Cigarette smoke	Kataoka et al.[52]
	Aromatic amines	Cigarette smoke	Kijima et al.[51]
	N-Nitrosamines	Combustion smoke	Kataoka et al.[52,96]
ECD	Aromatic amines	Surface water	Wegman and De Korte[32]
	Aromatic amines	Groundwater, wastewater, sewage	Schmidt et al.[33,34]
	Dinitroanilines	River water	Guan et al.[35]
	N-Nitrosamines	Metalworking fluids	Fadlallah et al.[120]
CLD	Trimethylamine	Atmosphere	Kashihira et al.[21]
	Aliphatic amines	Marine sediments and atmosphere	Lee et al.[22,23]
	N-Nitrosamines	Airborne	Fine et al.[94]
	N-Nitrosamines	Ambient air	Fadlallah et al.[95]
	N-Nitrosamines and tobacco-specific nitrosamines	Cigarette and tobacco smoke, tobacco	Brunnemann et al.,[97–100] Truhaut et al.,[101] Tricker et al.[102]
	N-Nitrosamines	Drinking water, tap water	Kimoto et al.,[104] Sen et al.[105]
	N-Nitrosodimethylamine	Drinking water, groundwater	Tomkins et al.[106,107]
	N-Nitrosamines	Household dishwashing liquid	Morrison and Hecht [110]
	N-Nitrosamines	Rubber nipples and pacifiers	Havery and Fazio,[112] Osterdahl,[113] Thompson et al.,[114,115] Gray and Stachiw,[116] Sen et al.[117]
	N-Nitrosodibutyl- and *N*-nitrosobenzylamines	Elastic rubber netting	Fiddler et al.[111,118]
	N-Nitrosamines	Drug formulation	Dawson and Lawrence,[121] Nielsen and Lings[122]
GC/MS	Aliphatic tertiary amine	Airborne	Seeber et al.[11]
	Aliphatic amines	River water, sediment, wastewater	Koga et al.,[36] Avery and Junk,[37] Terashi et al.,[38] Pietsch et al.,[39] Sacher et al.[40]
	Aliphatic tertiary amines	Lake water, sediment	Tsukioka et al.[41]
	Aromatic amines	Airborne	Meichini et al.[24]
	Aromatic amines	Combustion smoke, indoor air	Pieraccini et al.[25,53]
	Aromatic amines	Groundwater, river water, sediment	Okumura et al.,[45] Longo and Cavallaro,[46] Muller et al.[47,48]
	N-Nitrosamines	Cigarette smoke	Arrendale et al.[103]
	N-Nitrosodimethylamine	Drinking water	Sen et al.[105]
	N-Nitrosamines	Herbicides	Frassanito et al.[124]
	Heterocyclic amines	Cooking fumes	Vainiotalo et al.[168]

metal, resulting in a cyanide ion. A positive alkali metal ion migrates to the collector electrode and again liberates an electron. Therefore, compounds not containing a C–N bond, such as nitrogen gas, nitrogen oxides and ammonia, cannot respond to the NPD.

Thermionic detectors require hydrogen and air as with the FID, but at lower flow rates. Therefore, normal FID-type ionizations are minimal, and the selectivity ratios for detection of N- and P-containing compounds versus carbon detection are $3.5 \times 10^4 : 1$ and $7.5 \times 10^4 : 1$,

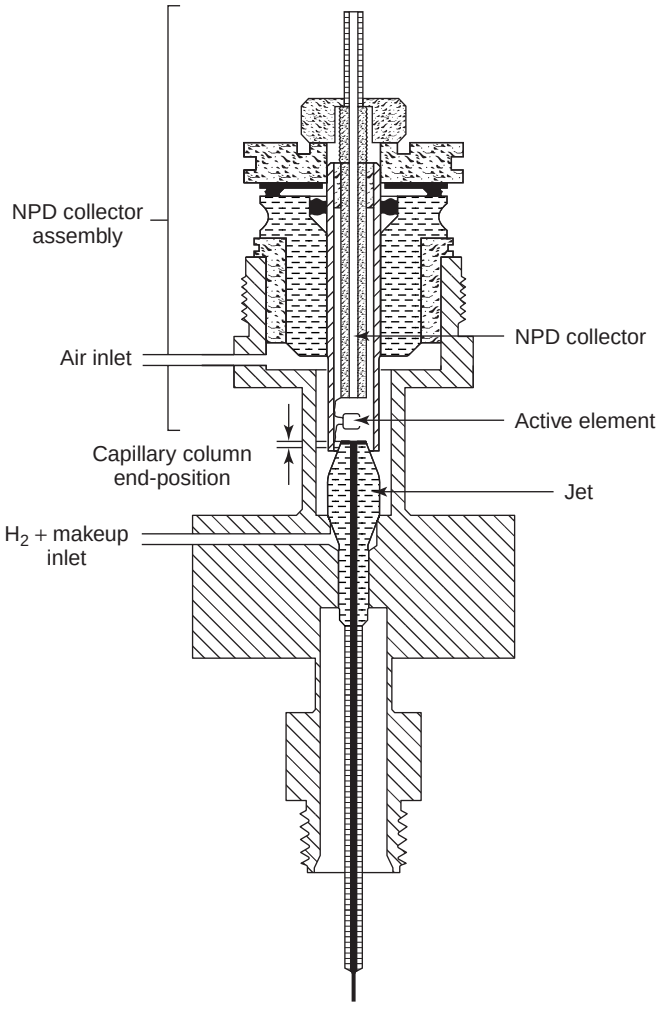

Figure 1 Diagram of a NPD designed for optimal performance with capillary columns. (Reproduced by permission of Hewlett-Packard from K.J. Hyver, P. Sandra, *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard, Avondale, PA, Chapters 4–19, 1989.)

respectively. The sensitivity of a thermionic detector is about 0.4–10 pg for N-containing compounds and 0.1–1 pg for P-containing compounds, and it is about 50–500 times more sensitive for nitrogen and phosphorus, respectively, than the FID. The main disadvantage of this detector is that its performance deteriorates with time. The alkali metal salt employed as the bead is continuously lost during the operation of the detector. Furthermore, the thermionic detector is not usable with columns of liquid phases containing halogen, phosphorus or nitrogen such as XE-60 and OV-210. In addition, the use of column, supports, glass wool treated with phosphoric acid and leak-detection fluids should be avoided to preserve the lifetime of the active element.

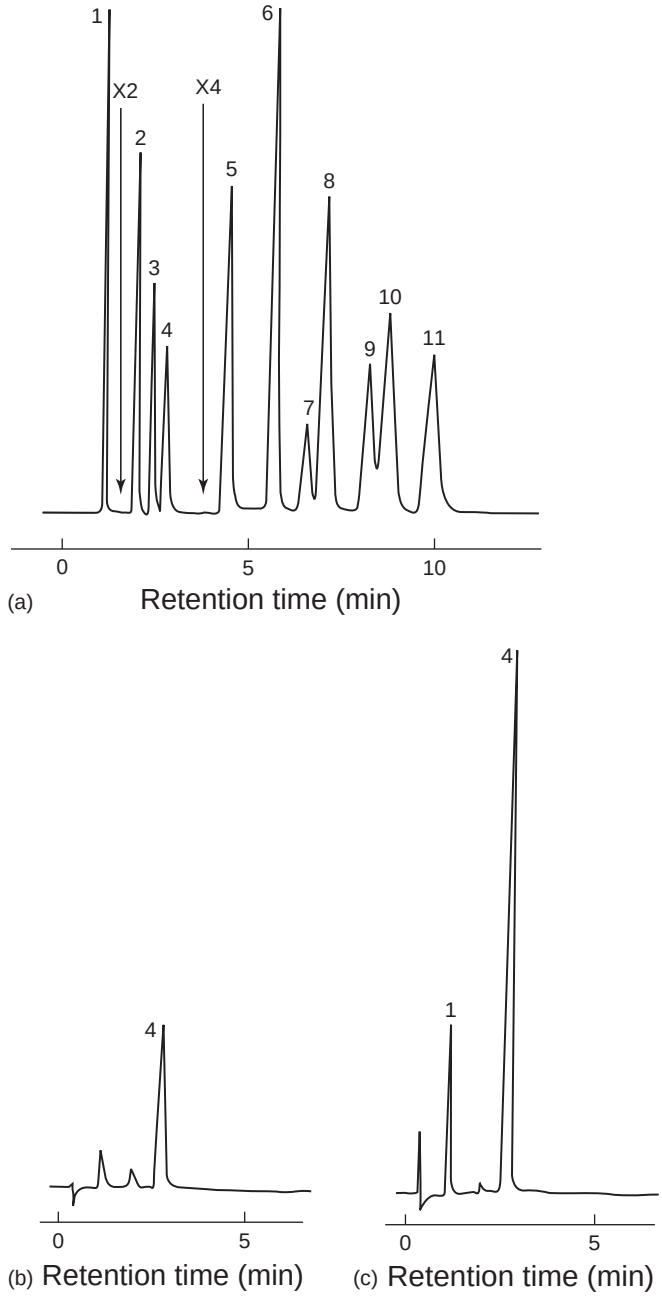

Figure 2 Gas chromatograms obtained from standard C_1–C_4 aliphatic amines and polluted sources. (a) Standard amines, 6 ng; (b) air in a cowshed; (c) air around a fermentation system of poultry wastes. GC conditions: column, alkalinized GHP-1 (2 m × 2 mm i.d.); column temperature, isothermal at 140 °C for 3 min and then programmed from 140 to 170 °C at 10 °C min^{-1} and isothermal at 170 °C for 10 min; injection temperature, 180 °C; detector temperature, 250 °C; carrier gas, nitrogen at 45 mL min^{-1}. Peaks: 1 = methylamine; 2 = dimethylamine; 3 = ethylamine; 4 = trimethylamine; 5 = isopropylamine; 6 = *n*-propylamine; 7 = *tert*-butylamine; 8 = diethylamine; 9 = *sec*-butylamine; 10 = isobutylamine; 11 = *n*-butylamine. (Reprinted by permission from K. Kuwata, E. Akiyama, Y. Yamazaki, H. Yamasaki, Y. Kuge, Y. Kiso, *Anal. Chem.*, **55**, 2199–2201 (1983), Copyright 1983, American Chemical Society.)

The thermionic detector is one of the most popular selective detectors for the GC analysis of amines and is used for the analyses of various amines such as aliphatic[13–18,56,63,193–196] and aromatic[19,26,55,56,193,197] amines, *N*-nitrosamines[89,123] and heterocyclic amines.[20,158,198] Kuwata et al.[13] reported that 11 low-molecular-mass aliphatic amines could be determined at the parts per billion level by flame-type thermionic detector (Figure 2a). Aliphatic amines were also detected with TSD after acylation,[16,194] permethylation[195] and isobutoxycarbonylation.[196] Skarping et al.[26,64,194,197] reported that some aliphatic and aromatic amines can be detected at the femtomole level with a TSD. Kataoka et al.[50] separated 22 aromatic amines using an HP-5 capillary column within 15 min (Figure 3a). This method is based on derivatization with *n*-propyl chloroformate and subsequent NPD analysis, and can detect picogram levels of aromatic amines. Furthermore, Kataoka and Kijima[198] developed a simple and rapid derivatization method for the GC analysis of heterocyclic amines. Ten heterocyclic amines were determined as their *N*-dimethylaminomethylene derivatives with an NPD. The detection limits of these compounds ranged from 2 to 15 pg.

2.3 Flame Photometric Detector

The FPD, based on a chemiluminescent species produced during the burning process in a hydrogen-rich flame, is a selective detector for compounds containing sulfur and phosphorus. The emission of S-containing compounds ranges from 390 to 480 nm and that of P-containing compounds from 480 to 580 nm. By choosing a filter having a transmission window of wavelength 390–410 nm for S-containing compounds or 520–560 nm for P-containing compounds, the maximum emission from S- or P-containing compounds can be sensed by the photomultiplier. The emission of hydrocarbons is between 390 and 520 nm, and therefore the emission from the background hydrocarbons will be almost negligible.

A diagram of a Hewlett-Packard FPD is shown in Figure 4. The detector is essentially an FID equipped with a filter that can be changed according to the desired wavelength for sulfur or phosphorus. The end of the capillary column is led into the flame jet where the column eluate mixes with hydrogen flow and is burned. The light emitted passes through two heat filters and then through the selective filter and finally on to the photomultiplier. The chemical reactions that occur in the flame and emission zone are complex depending on the gas flows and temperature. Sulfur and phosphorus are detected as S=S and HPO molecules, respectively. These molecules are created in a metastable state and, when they decay, they release energy in the form of a photon of a specific wavelength. The light emitted from S-containing compounds is not linear

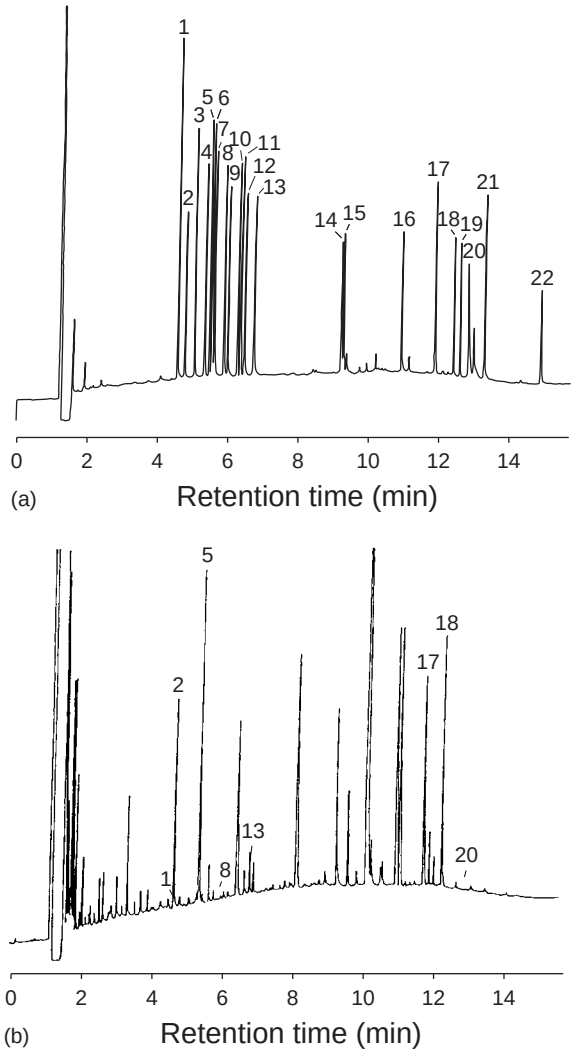

Figure 3 Gas chromatograms obtained from *n*-propoxycarbonyl derivatives of aromatic amines and cigarette smoke sample. (a) Standard amine derivatives (each peak corresponding to 1–5 ng of amine); (b) cigarette smoke sample. GC conditions: column, HP-5 (30 m × 0.32 mm i.d.); column temperature, programmed from 120 to 220 °C at 10 °C min^{-1} and then from 220 to 320 °C at 20 °C min^{-1} and isothermal at 320 °C for 1 min; injection and detector temperatures, 330 °C; carrier gas, helium, programmed from 80 to 120 kPa at 4 kPa min^{-1} and then from 120 to 270 kPa at 30 kPa min^{-1} and held at 270 kPa for 1 min; detector, NPD. Peaks: 1 = aniline; 2 = internal standard (*N*-methyl-*m*-toluidine); 3 = 2-toluidine; 4 = 2, 6-xylidine; 5 = 3-toluidine; 6 = 4-toluidine; 7 = 2-ethylaniline; 8 = 2,5-xylidine; 9 = 2,4-xylidine; 10 = 2,3-xylidine; 11 = 3,5-xylidine; 12 = 4-ethylaniline; 13 = 3,4-xylidine; 14 = α-naphthylamine; 15 = *o*-aminobiphenyl; 16 = *N*-phenylnaphthylamine; 17 = internal standard (2,4-diaminotoluene); 18 = 2-aminofluorene; 19 = *p*-aminoazobenzene; 20 = benzidine; 21 = *p*,*p*′-diaminodiphenylmethane; 22 = 2-aminoanthracene.

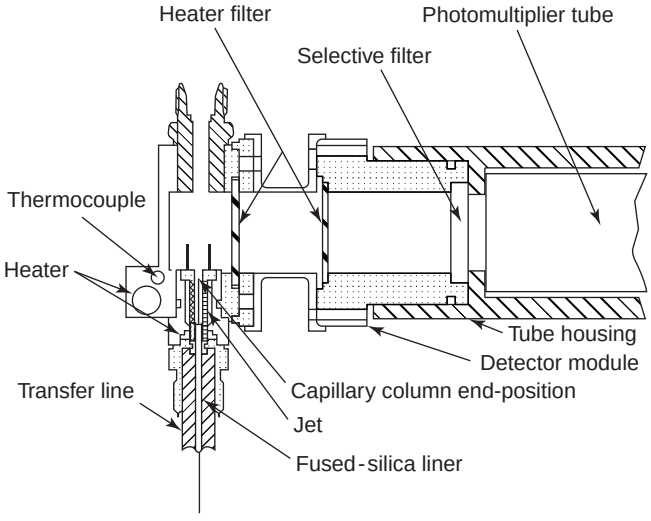

Figure 4 Diagram of a FPD optimized for use with capillary columns. (Reproduced by permission of Hewlett-Packard from K.J. Hyver, P. Sandra, *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard, Avondale, PA, Chapters 4–20, 1989.)

with concentration owing to the reactions in the flame, but is approximately proportional to the square of the sulfur atom concentration. The selectivity ratio for detection of S- or P-containing compounds versus carbon detection is $10^4:1$. The sensitivity of the FPD is about 20 pg for S-containing compounds and 1 pg for P-containing compounds. One serious problem with the FPD is the quenching or reabsorption of the light emitted by the selected species. Hydrocarbon quenching occurs from collisions when the peak of S-containing compounds is co-eluted with another hydrocarbon in relatively high concentration. Self-quenching, which is due to collisional energy absorption, competing chemical reactions or the reabsorption of the photon by an inactivated species, can also occur at high concentrations of the heteroatom species. Precise control of the gas flow to the detector virtually eliminates quenching of the detector response.

Although the FPD is widely used for S- and P-containing compounds such as agricultural chemicals and SO_2 in the environment, it can also detect amino compounds by conversion into their S- or P-containing derivatives.[51,52,57,58,65,96,199,200] Hamano et al.[57] and Kataoka et al.[65,199] reported that primary and secondary aliphatic amines could be detected as their *N*-benzenesulfonyl derivatives by FPD in the S mode. As shown in Figure 5(a) and (b), 14 primary and 9 secondary amines were separately determined. The detection limits of these amines ranged from 10 to 100 pg. Furthermore, Kataoka et al. reported that aliphatic[58,200] and aromatic[51] amines and *N*-nitrosamines[52,96] could be analyzed as their *N*-dialkylthiophosphoryl derivatives with an FPD in the P mode. In particular, secondary amines can be selectively converted into their *N*-diethylthiophosphoryl derivatives after treatment with *o*-phthaldialdehyde (OPA), because OPA reacts only with primary amino groups. As shown in Figure 6(a–d), primary amines are also derivatized without OPA treatment and detected with an FPD, but they are not detected at all with OPA pretreatment. On the other hand, secondary amines are detected irrespective of OPA pretreatment, because these amines do not react with OPA. This technique is also applied to the determination of *N*-nitrosamines, which are denitrosated with hydrobromic acid to produce the corresponding secondary amines and then derivatized.

2.4 Electron-capture Detector

The ECD, which consists of a radioactive source such as ^{63}Ni, is selective detector for halogenated or nitrated compounds, organometallics and conjugated carbonyls. Electron-capture detection involves the following three steps: generation of thermal energy electrons, capture of some of these electrons by an electrophilic compound and collection and measurement of the unreacted electrons. In order to produce capturable (low-energy) thermal electrons, the carrier gas is ionized by β-particles from a radioactive source placed in the cell. This electron flow produces a small current, which is collected and measured by a suitable amplifier. When the sample molecule is introduced into the cell, electrons are captured by the sample, resulting in a decreased current. The detector can be used in two ways: either with a constant potential applied across the cell (dc mode) or with a pulse potential across the cell (pulse mode). In the dc mode, hydrogen or nitrogen can be used as the carrier gas and a small potential (a few volts) is applied across the cell just sufficient to collect all the electrons available and provide a small standing current. When an electron-capturing molecule enters the cell, the electrons are captured by the molecule and the molecules become charged. The mobility of the captured electrons is much reduced compared with the free electrons and therefore the electrode current falls dramatically. On the other hand, in the pulse mode, a mixture of 10% methane in argon is usually used as the carrier gas and the electron-capturing environment is different. The period of the pulsed potential is adjusted such that relatively few of the slow negatively charged molecules reach the anode, but the faster moving electrons are all collected. A diagram of a Hewlett-Packard ECD is shown in Figure 7. The anode and cavity are in a ''pi and cup'' geometry, and the radioactive source is ^{63}Ni plated on the cup. Although the selectivity

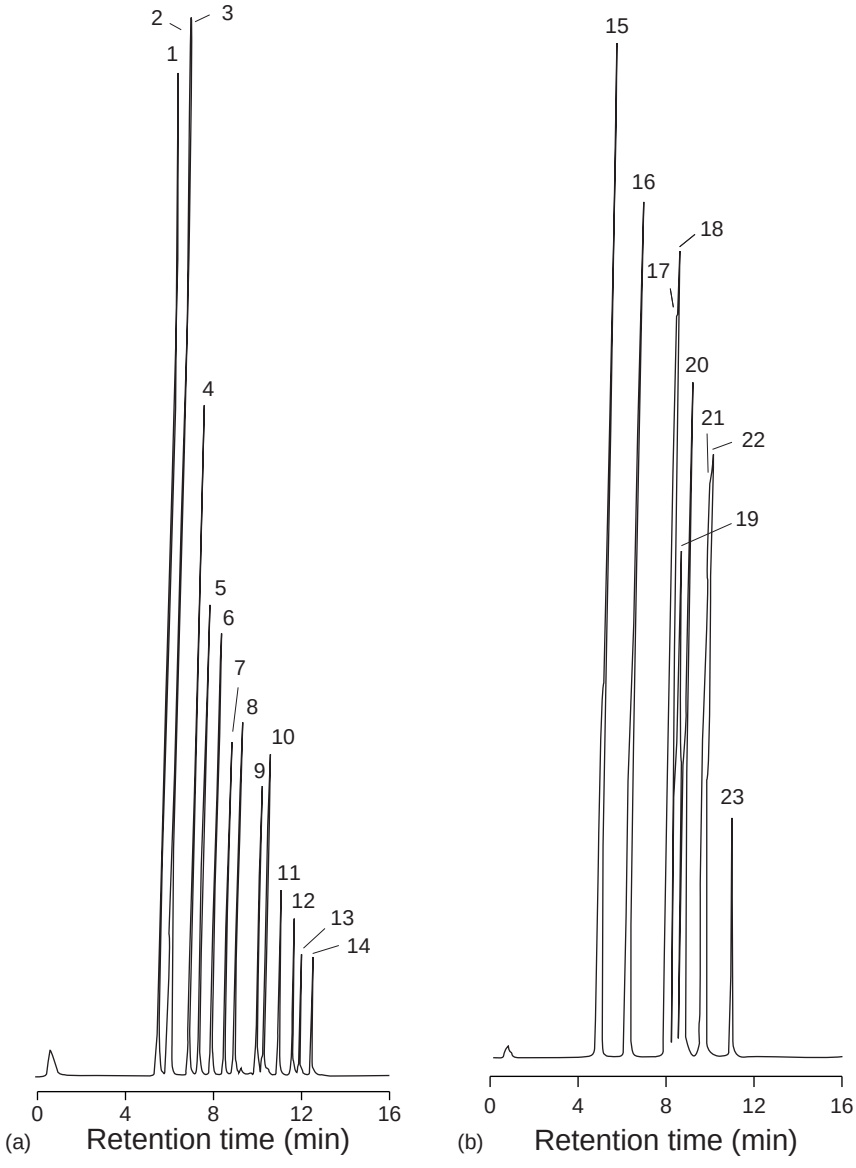

Figure 5 Gas chromatograms obtained from benzenesulfonyl derivatives of (a) primary and (b) secondary amines. GC conditions: column, DB-1 (15 m × 0.53 mm i.d.); column temperature, programmed from 120 to 280 °C at 10 °C min^{-1}; injection and detector temperatures, 290 °C; nitrogen flow-rate, 10 mL min^{-1}; detector, FPD (S mode). Each peak corresponds to 2 ng of amine. Peaks: 1 = methylamine; 2 = ethylamine; 3 = isopropylamine; 4 = *n*-propylamine; 5 = isobutylamine; 6 = *n*-butylamine; 7 = isoamylamine; 8 = *n*-amylamine; 9 = hexylamine; 10 = cyclohexylamine; 11 = heptylamine; 12 = benzylamine; 13 = octylamine; 14 = β-phenylethylamine; 15 = dimethylamine; 16 = diethylamine; 17 = dipropylamine; 18 = pyrrolidine; 19 = morpholine; 20 = piperidine; 21 = dibutylamine; 22 = hexamethyleneimine; 23 = *N*-methylbenzylamine. (Reprinted from H. Kataoka, *J. Chromatogr. A*, **733**, 19–34 (1996), Copyright 1996, with permission from Elsevier Science.)

depends on the affinity of compounds for the electrode, the selectivity ratio for detection of polyhalogenated compounds versus carbon detection is 10^5–10^6 : 1. The ECD is extremely sensitive, probably the most sensitive GC detector available. Although the sensitivity of the ECD depends on the molecular structure, that of chlorinated compounds is about 0.05–1 pg. For successful electron-capture detection, it is important that carrier and purge gases are very clean and dry (99.9995%).

Although the ECD is widely used in the detection and analysis of halogenated compounds, in particular pesticides, it can also detect amino compounds by conversion into halogen-containing derivatives.[32–34,120,194,201–205] Miyamoto et al.[201] determined selectively some primary amines as their *N*-benzenesulfonyl-*N*-trifluoroacetyl derivatives by gas chromatography/electron-capture detector (GC/ECD). Skarping et al.[204] reported that aromatic amines could be separated as their perfluoro fatty acid

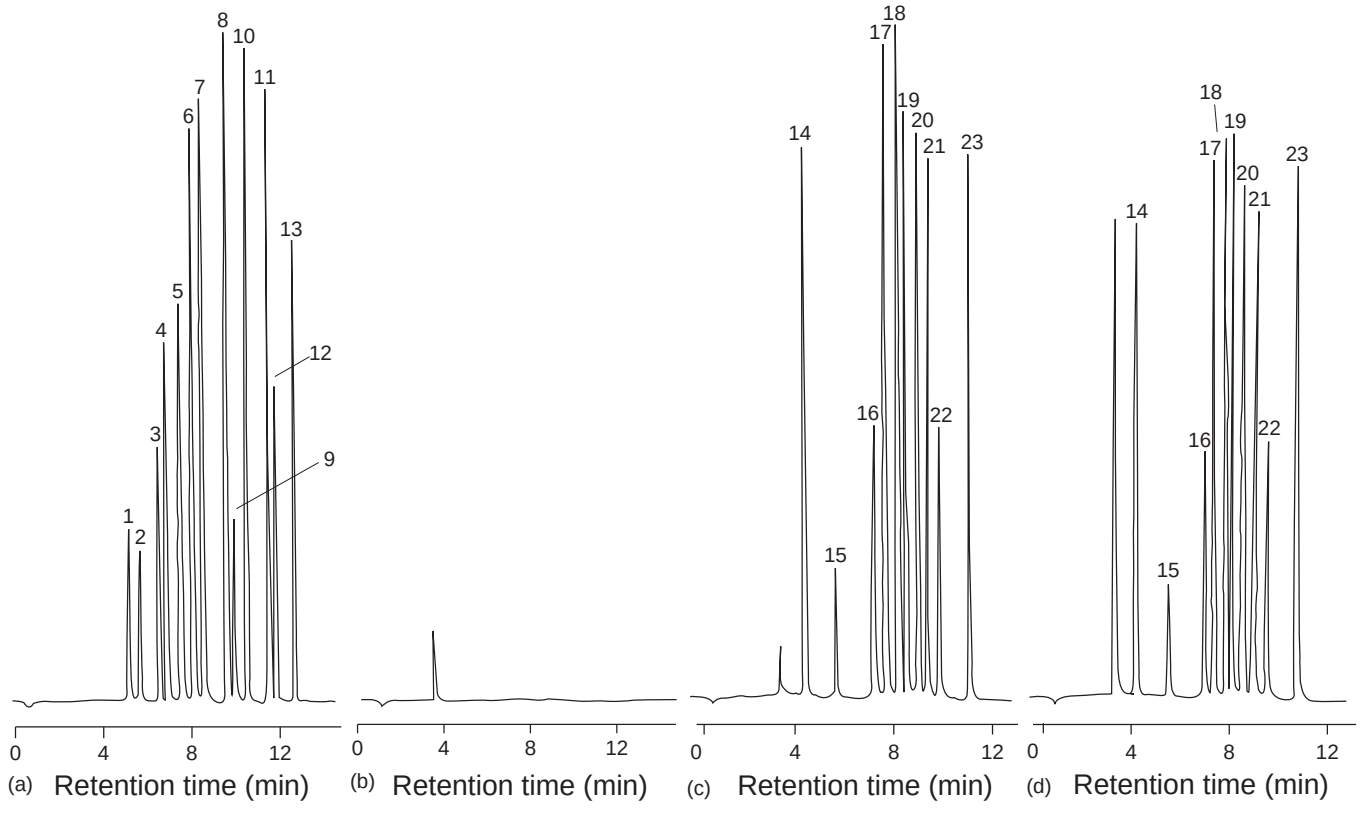

Figure 6 Gas chromatograms obtained from *N*-diethylthiophosphoryl derivatives of primary and secondary amines. (a) Primary amines; (b) primary amines pretreated with OPA; (c) secondary amines; (d) secondary amines pretreated with OPA. GC conditions: column, DB-1701 (15 m × 0.53 mm i.d.); column temperature, programmed from 100 to 260 °C at 10 °C min^{-1}; injection and detector temperatures, 280 °C; nitrogen flow-rate, 10 mL min^{-1}; detector, FPD (P mode). Each peak corresponds to 20 pmol of amine. Peaks: 1 = methylamine; 2 = ethylamine; 3 = *n*-propylamine; 4 = isobutylamine; 5 = *n*−butylamine; 6 = isoamylamine; 7 = *n*-amylamine; 8 = hexylamine; 9 = cyclohexylamine; 10 = heptylamine; 11 = benzylamine; 12 = octylamine; 13 = β-phenylethylamine; 14 = dimethylamine; 15 = diethylamine; 16 = dipropylamine; 17 = pyrrolidine; 18 = piperidine; 19 = morpholine; 20 = dibutylamine; 21 = hexamethyleneimine; 22 = *N*-methylcyclohexylamine; 23 = *N*-methylbenzylamine. (Reprinted from H. Kataoka, M. Eda, M. Makita, *Biomed. Chromatogr.*, **7**, 129–133 (1993), Copyright 1993, with permission from John Wiley & Sons.)

amides with a glass capillary column and determined using on-column injection and ECD (Figure 8a and b). The detection limits of these amines were ca. 1 pg. Furthermore, Schmidt et al.[33,34] reported a selective and sensitive method for the determination of aromatic amines based on derivatization to the corresponding iodobenzenes and subsequent GC/ECD analysis. Separation of at least 30 compounds in a single chromatographic run in 30 min is possible (Figure 9) and 56 aromatic amines were investigated. On the other hand, dinitroaniline herbicides could be directly analyzed by GC/ECD, because they possess a nitro group in the molecule.[35]

2.5 Chemiluminescence Detector

The CLD or thermal energy analyzer (TEA), based on emission spectroscopy, is a selective detector for nitroso and nitroaromatic compounds. The response of the TEA to *N*-nitroso compounds is based on the reaction of nitrogen oxide with ozone combined with the preceding pyrolysis of nitroso compounds. A schematic diagram of a TEA is shown in Figure 10. The eluate from the GC column enters the pyrolyzer where the selective catalytic decomposition of *N*-nitroso compounds takes place, resulting in a nitrosyl radical and an organic radical. The N−NO bond is the weakest in these compounds. The pyrolyzer eluent expands into the evacuated reaction chamber in which the nitrogen oxide reacts with ozone, resulting in excited nitrogen dioxide. The excited nitrogen dioxide rapidly decays back to its ground state with the characteristic emission of light. Then the emitted radiation is detected by a photomultiplier through a red optical filter. The filter is used to block possible light emission resulting from sources other than NO chemiluminescence. The intensity of the emission is proportional to the nitrosyl radical concentration, and hence to the *N*-nitroso compound concentration. The decomposition of substances with a catalyst at lower temperatures of

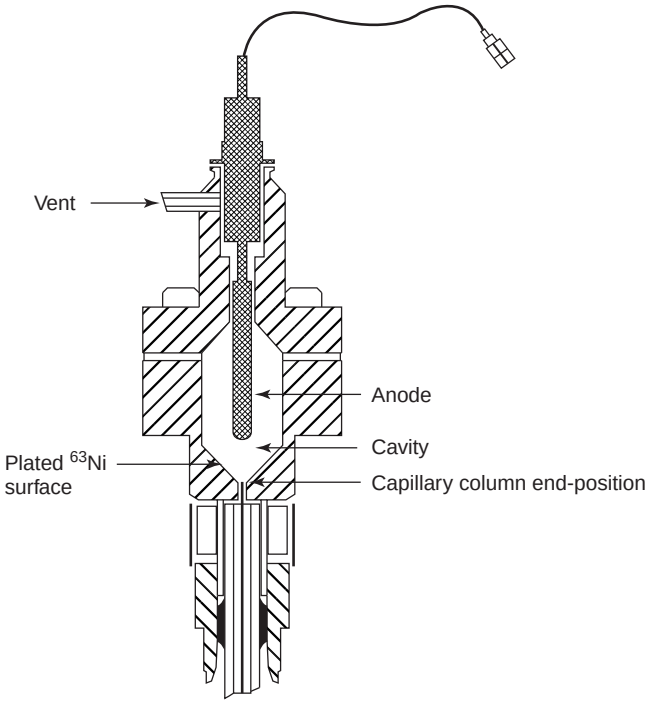

Figure 7 Diagram of an ECD design showing the capillary column interface. (Reproduced by permission of Hewlett-Packard from K.J. Hyver, P. Sandra, *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard, Avondale, PA, Chapters 4–18, 1989.)

about 300 °C is more advantageous than pyrolysis at elevated temperatures, because the decomposition of the N–NO bond is more selective, even though the detector response to nitrosamines is diminished at lower temperature. The response of the CLD to nitroaromatic compounds increases rapidly with increasing temperature at 800–900 °C of the pyrolyzer. Nitrogen-containing compounds are also detected with CLD, based on the catalytic decomposition of these compounds in the presence of oxygen at temperatures of 900–1000 °C and on the subsequent ozone oxidation of the nitrogen oxide formed, as with nitroso compounds. The sensitivity of CLD for *N*-nitrosodimethylamine is about 50 pg.

The CLD has been used as a selective detector for *N*-nitrosamines.[94,95,97–118,121,122] Fadlallah et al.[95] separated volatile *N*-nitrosamines within 12 min (Figure 11a). Adams et al.[98] analyzed tobacco-specific *N*-nitrosamines by GC/TEA. The CLD can be used for the selective analysis of not only *N*-nitrosamines but also aliphatic[21–23,59] and aromatic[202] amines by modification of the pyrolyzer temperature. Pfundstein et al.[59] reported that primary and secondary amines could be analyzed as their *N*-benzenesulfonyl derivatives by gas chromatography/chemiluminescent detection (GC/CLD) with a modified TEA. Thermal pyrolysis of amines at 720 °C with a nickel catalyst is used to produce NO radicals.

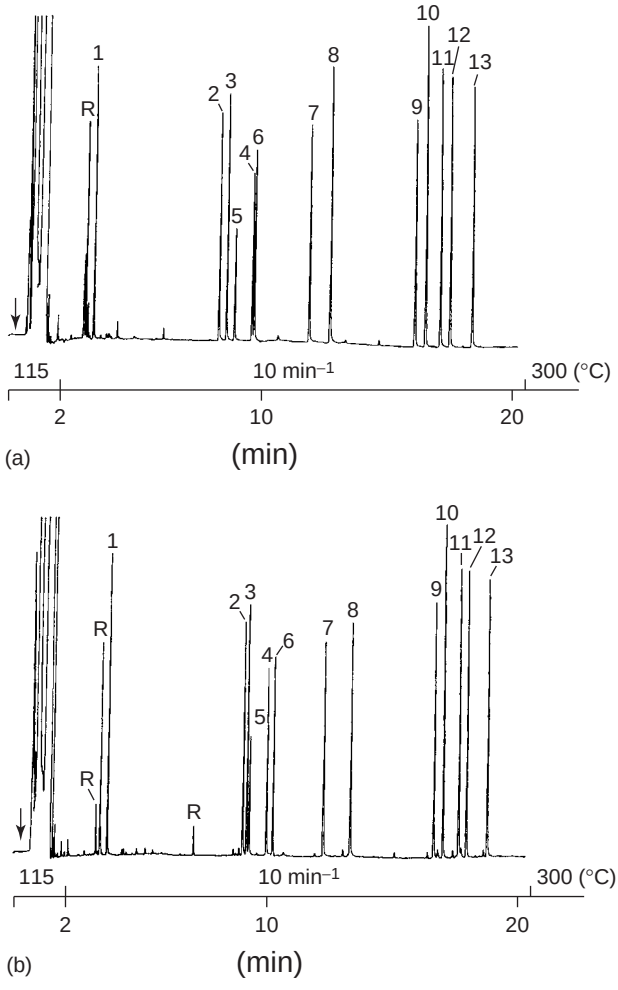

Figure 8 Chromatograms of amine derivatives. (a) pentafluoropropionic acid (PFPA) derivatives; (b) heptafluorobutyric anhydride (HFBA) derivatives. GC conditions: column, OV-73 (25 m × 0.32 mm i.d.); on-column injection of 2 µL of toluene solution at 115 °C; column temperature, isothermal at 115 °C for 2 min and then programmed from 115 to 300 °C at 10 °C min^{-1} and held at 300 °C for 5 min; carrier gas, helium at 0.8 kg cm^{-2}; detector, ECD (constant-current mode); make-up gas, argon–methane (95 : 5) at 55 mL min^{-1}. Each peak corresponding to 20–80 pg of amine. Peaks: 1 = aniline; 2 = 2,6-toluidine; 3 = 2,4-toluidine; 4 = 2,4-diaminoanisole; 5 = α-naphthylamine; 6 = β-naphthylamine; 7 = 4-aminobiphenyl; 8 = 1,5-naphthalenediamine; 9 = benzidine; 10 = 4,4′-methylenedianiline; 11 = *o*-tolidine; 12 = 4,4′-methylenebis-(*o*-chloroaniline); 13 = 4,4′-diamino-3,3′-dimethoxybiphenyl(*o*-dianisidine). Peaks denoted R refer to reagent and solvent impurities. (Reprinted from G. Skarping, L. Renman, B.E.F. Smith, *J. Chromatogr.*, **267**, 315–327 (1983), Copyright 1983, with permission from Elsevier Science.)

2.6 Mass-selective Detector

MSD has been recognized as an excellent detector for GC, because the spectral data provide the qualitative information lacking in other GC detectors. When a molecule in the vapor state is bombarded with electrons,

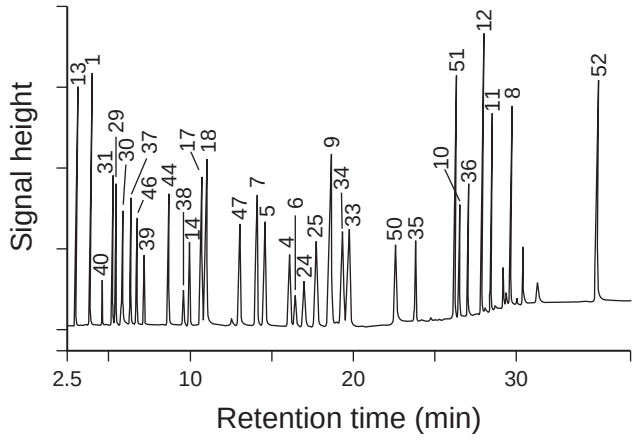

Figure 9 GC separation of 33 iodinated derivatives of aromatic amines after enrichment on HR-P phase. Concentration of aromatic amines, 10–50 ng mL^{-1}. GC conditions: column, 5% phenylmethylpolysiloxane (30 m × 0.25 mm i.d.); column temperature, isothermal at 135 °C for 20.5 min, then programmed from 135 to 235 °C at 12.5 °C min^{-1} and held at 235 °C for 8.5 min; injection and detector temperatures, 250 and 300 °C, respectively; carrier gas, nitrogen at 100 kPa; split ratio, 1:120; detector, ECD. Peaks: 1 = 4-aminotoluene; 4 = 2,4-diaminotoluene; 5 = 4-amino-2-nitrotoluene; 6 = 2,6-diaminotoluene; 7 = 2-amino-6-nitrotoluene; 8 = 2,6-diamino-4-nitrotoluene; 9 = 2-amino-4-nitrotoluene; 10 = 4-amino-2,6-dinitrotoluene; 11 = 2,4-diamino-6-nitrotoluene; 12 = 2-amino-4,6-dinitrotoluene; 13 = aniline; 14 = 1,3-phenylenediamine; 17 = 3-nitroaniline; 18 = 4-nitroaniline; 24 = 2-amino-3-nitrotoluene; 25 = 2-amino-5-nitrotoluene; 29 = 2,6-dimethylaniline; 30 = 3,4-dimethylaniline; 31 = 3,5-dimethylaniline; 33 = 1-naphthylamine; 34 = 2-naphthylamine; 35 = 2-aminobiphenyl; 36 = 4-aminobiphenyl; 37 = 4-isopropylaniline; 38 = 2,6-diethylaniline; 39 = 2-ethyl-6-methylaniline; 40 = 4-chloro*N*-methylaniline; 44 = 3,4-dichloroaniline; 46 = 4-chloro-2-methylaniline; 47 = 3-chloro-4-methoxyaniline; 50 = 2,6-dinitroaniline; 51 = 3,5-dinitroaniline; 52 = benzidine. (Reprinted from T.C. Schmidt, M. Less, R. Haas, E. von Low, K. Steinbach, G. Stork, *J. Chromatogr. A*, **810**, 161–172 (1998), Copyright 1998, with permission from Elsevier Science.)

bonds are ruptured and the molecle ionizes. The kind and amount of fragments obtained are characteristic of the molecule. In a magnetic sector MSD, when the positively charged fragments are accelerated into a magnetic field, they travel curved paths, the radii of which are proportional to the square roots of the masses of the ions. By varying the magnetic field, the segregated ion beams containing ions of identical mass-to-charge ratio are swept successively past a collector. Here the ions discharged, producing currents proportional to the relative abundances of ions of each mass. These currents are recorded successively to give the mass spectrum. On the other hand, a quadrupole MSD performs a different mass separation, using four rods with an electric field between them. As the ions enter the field, they interact with it in such a way that only ions of a specific mass-to-charge ratio will pass through to the detector. A scan of many ions is achieved by varying the electric field. An ion trap mass detector (ITD) can store the ions of a certain mass in stable paths for many seconds. The ion trap is scanned over a mass range whereby the ions are detected by a conventional electron multiplier. The two most common techniques used to produce ions in an MS are electron impact ionization (EI) and chemical ionization (CI). A schematic diagram of a typical electron impact ionization mass spectrometry (EIMS) system is shown in Figure 12.

An MSD consists of four main parts: ion source, mass analyzer, detector and instrument control and data-handling computer system. The first step in a GC/MS analysis is qualitative identification, using a complete scan across the entire mass range. Identification can be done with the help of spectral libraries, usually kept in a database on a computer with the detector-controlling software. The molecular ion and characteristic ions are used for the identification of compounds. The second step is quantitative analysis using SIM. By using a filter to select only a few ions, the sensitivity can be increased. The final step is the total ion chromatogram (TIC), based on the plot of total response (sum of all ions) over time by scanning repeatedly. For quantitative SIM analysis, it is necessary first to accomplish the optimum chromatography, select ions which are compound specific (molecular ion or fragment ion), give a large signal and have no interferences. Next it is necessary to

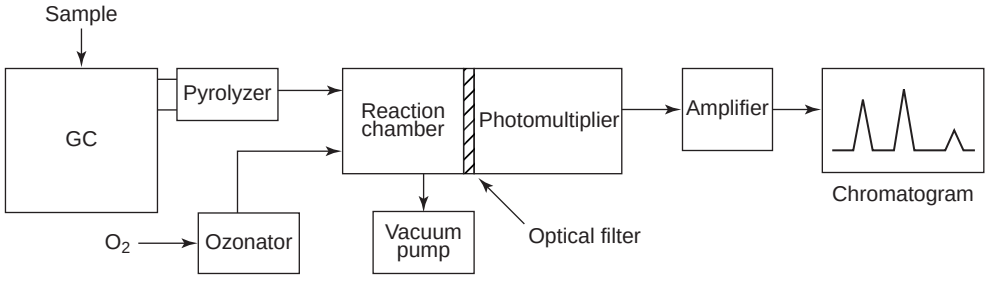

Figure 10 Schematic diagram of a TEA.

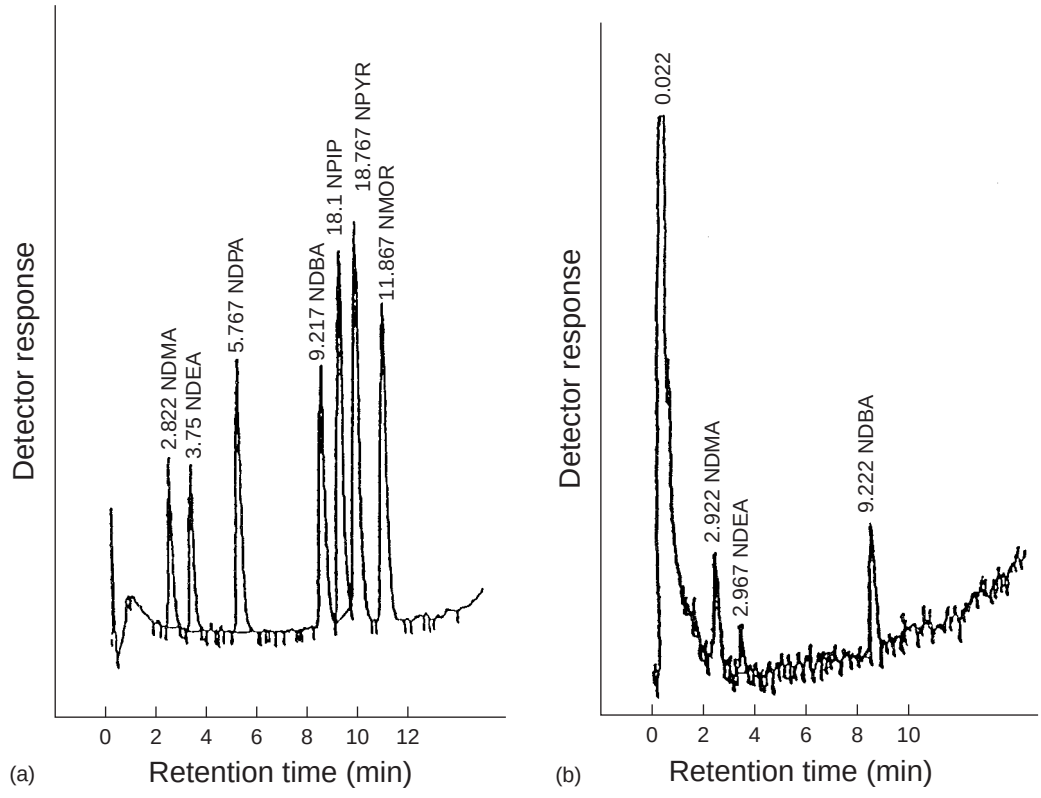

Figure 11 Chromatograms obtained from (a) volatile nitrosamine standards and (b) an air sample using gas chromatography/thermal energy analysis (GC/TEA). GC conditions: column, 10% Carbowax™ + 2% KOH on Chromosorb® W (2 m × 2 mm i.d.), column temperature, programmed from 130 to 210 °C at 4 °C min^{-1} and held at 210 °C for 5 min; injector temperature, 200 °C, carrier gas, nitrogen at 15 mL min^{-1}; detector, TEA; detector furnace, 550 °C; cold trap, −196 °C; vacuum in the reaction chamber, 0.65 mmHg. Peaks: NDMA = *N*-nitrosodimethylamine; NDEA = *N*-nitrosodiethylamine; NDPA = *N*-nitrosodipropylamine; NDBA = *N*-nitrosodibutylamine; NPIP = *N*-nitrosopiperidine; NPYR = *N*-nitrosopyrrolidine; NMOR = *N*-nitrosomorpholine. (Reproduced by permission of Springer-Verlag from S. Fadlallah, S.F. Cooper, G. Perrault, G. Truchon, J. Lesage, *Bull. Environ. Contam. Toxicol.*, **57**, 867–874 (1996), Copyright Springer-Verlag.)

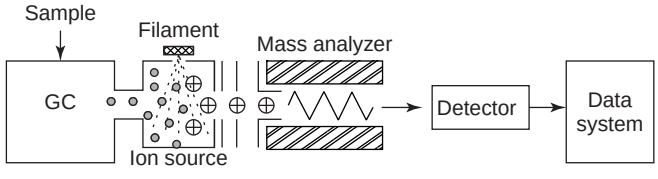

Figure 12 Schematic diagram of a typical mass spectrometer with an electron-impact ion source.

optimize the MSD response to the monitored ions, calibrate each ion mass and assign the dwell time for each ion using GC peak width and ion signal ratios.

GC/MS can provide structural information for the unequivocal identification of amines and can be applied to the analyses of various amines such as aliphatic [36–41] and aromatic[8,11,12–25,45–48,59,205] amines, *N*-nitrosamines[103,124–126] and heterocyclic amines.[141–145,168] Aliphatic amines are analyzed after derivatization with 2,4-dinitrofluorobenzene, benzenesulfonyl chloride and pentafluorobenzaldehyde (PFBA). Sacher et al.[40] reported that aliphatic primary and secondary amines could be determined as their *N*-benzenesulfonyl and 2,4-dinitrophenyl derivatives (Figure 13a and b) and the method using the former derivatives was more selective and sensitive. Longo and Cavallaro[46] developed a method for the simultaneous identification of 73 primary and secondary aromatic amines as their heptafluorobutyramides. (Figure 14) The electrophoretic derivatives were analyzed by GC combined with an ECNICI MSD. Detection limits of these amines were in the range 0.3–66.3 pg injected in the full-scan mode and 0.01–0.57 pg injected in the SIM mode. Furthermore, Skog et al.[145] analyzed six nonpolar heterocyclic amines by GC/MS in the SIM mode. Murray et al.[141–143] developed a sensitive and selective method for the determination of several heterocyclic amines as their 3,5-bistrifluoromethylbenzyl derivatives by negative-ion chemical ionization mass spectrometry (NICIMS) in the SIM mode. The detection limits were 1 pg.

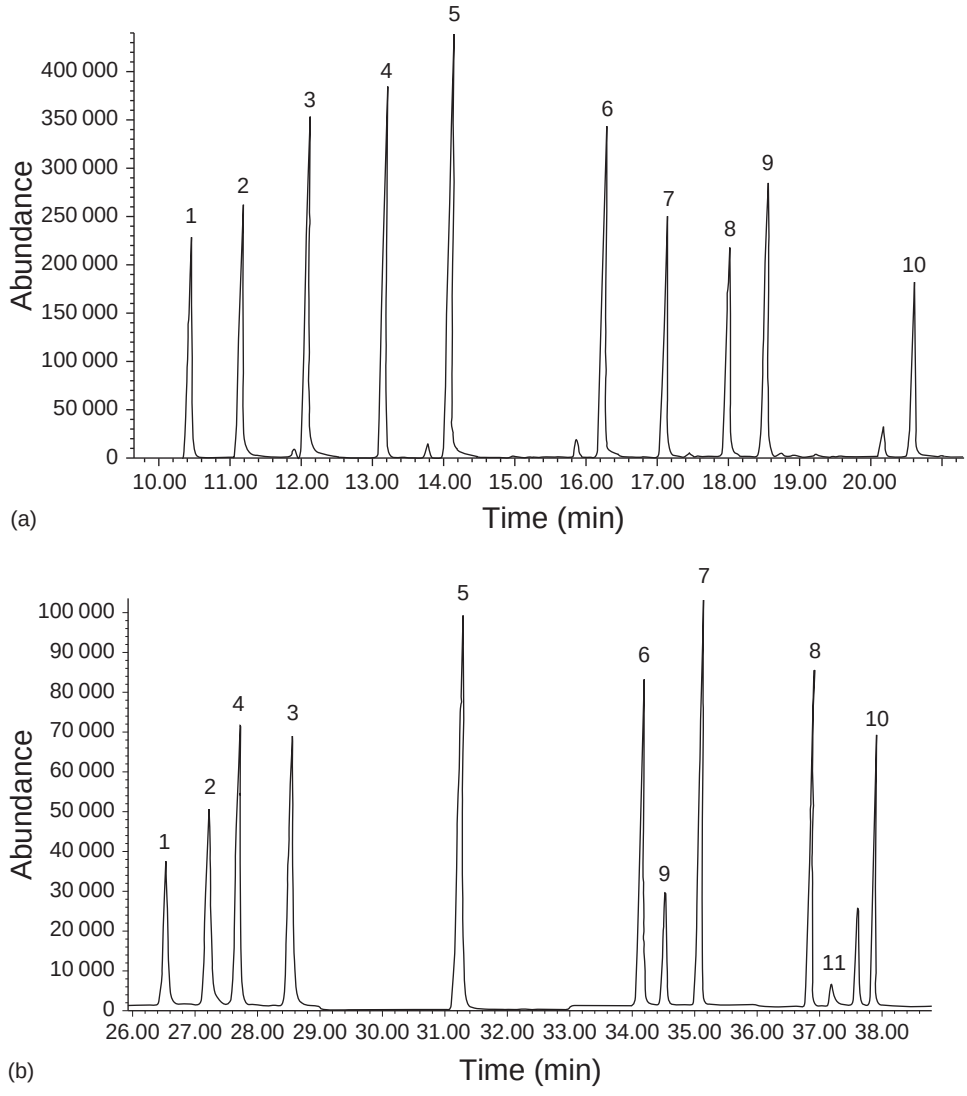

Figure 13 Gas chromatograms of aliphatic primary and secondary amines. (a) Benzenesulfonyl derivatives; (b) 2,4-dinitrophenyl derivatives. GC conditions: DB-5 (35 m × 0.25 mm i.d.); (a) column temperature, isothermal at 120 °C for 3 min, then programmed from 120 to 220 °C at 5 °C min^{-1}, from 220 to 290 °C at 10 °C min^{-1} and held at 290 °C for 5 min; (b) column temperature, isothermal at 140 °C for 3 min, then programmed from 140 to 210 °C at 3 °C min^{-1}, from 210 to 290 °C at 10 °C min^{-1} and held at 290 °C for 5 min; injector and detector temperatures, 290 °C; carrier gas, helium at 1 mL min^{-1}; detector, MSD. Peaks: 1 = dimethylamine; 2 = methylamine; 3 = ethylamine; 4 = diethylamine; 5 = *n*-propylamine; 6 = *n*-butylamine; 7 = pyrrolidine; 8 = morpholine; 9 = piperidine; 10 = *n*-hexylamine internal standard; 11 = ethanolamine. Concentrations, 25 ng mL^{-1} each. (Reprinted from F. Sacher, S. Lenz, H.-J. Brauch, *J. Chromatogr., A*, **764**, 85–93 (1997), Copyright 1997, with permission from Elsevier Science.)

2.7 Electrolytic Conductivity Detector

The ELCD is a useful detector in environmental analysis owing to its selectivity for halogen-, sulfur- and nitrogen-containing compounds. This is done by mixing the effluent with a reaction gas (oxidizing or reducing) in a reaction catalyst (usually nickel). The resulting gas-phase reaction products are then combined with a stream of deionized solvent, producing a conductive solution. The electrical conductivity of the liquid is measured and recorded. In the case of nitrogen-mode analysis, the column eluate is reduced with H_2 at 800–900 °C in a nickel reaction tube, producing ammonia by the decomposition of organic nitrogen compounds. Under these conditions, the organic halogen, sulfur, oxygen and phosphorus compounds are converted into HX, H_2S, H_2O, PH_3 and lower alkanes. Although H_2S and lower alkanes (mainly methane) give a low response owing to the low ionization, the acidic products can be removed with a basic scrubber placed between the reaction chamber and the conductivity cell.

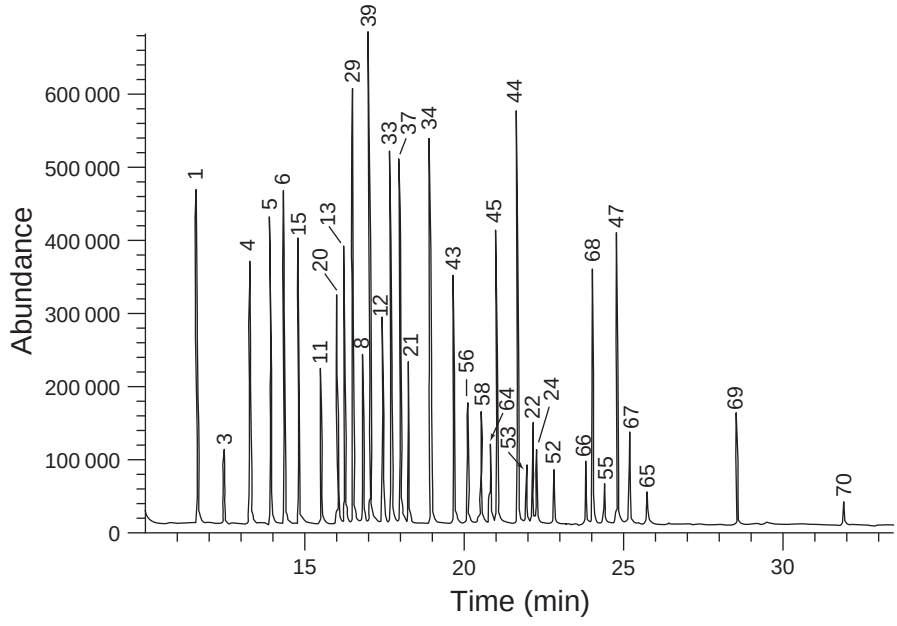

Figure 14 TIC of a mixture of HFBA derivatives under electron-capture negative-ion chemical ionization (ECNICI) conditions. GC/MS conditions: GC column, HP-5 (50 m × 0.2 mm i.d.); column temperature, initially at 50 °C for 0.5 min, heated to 110 °C at 50 °C min^{-1}, to 225 °C at 5 °C min^{-1}, to 280 °C at 20 °C min^{-1} and held at 280 °C for 15 min; injector and transfer line temperatures, 250 and 280 °C, respectively; carrier gas, helium at 250 kPa; ion-source temperature, 200 °C; electron energy, 150 eV. Peaks: 1 = aniline; 3 = *N*-methylaniline; 4 = 2-methylaniline; 5 = 3-methylaniline; 6 = 4-methylaniline; 8 = 2,3-dimethylaniline; 11 = 2,6-dimethylaniline; 12 = 3,4-dimethylaniline; 13 = 3,5-dimethylaniline; 15 = 2-ethylaniline; 20 = 2-methoxyaniline; 21 = 4-methoxyaniline; 22 = 2,4-dimethoxyaniline; 24 = 3,4-dimethoxyaniline; 29 = 4-chloroaniline; 33 = 3-chloro-2-methylaniline; 34 = 3-chloro-4-methylaniline; 37 = 5-chloro-2-methylaniline; 39 = 2,4-dichloroaniline; 43 = 3,5-dichloroaniline; 44 = 2,3,4-trichloroaniline; 45 = 2,4,5-trichloroaniline; 47 = 3,4,5-trichloroaniline; 52 = *N*-methyl-4-nitroaniline; 53 = 2-methyl-3-nitroaniline; 55 = 2-methyl-5-nitroaniline; 56 = 2-methyl-6-nitroaniline; 58 = 4-methyl-2-nitroaniline; 64 = 5-chloro-2-nitroaniline; 65 = 4-chloro-3-nitroaniline; 66 = 1-aminonaphthalene; 67 = 2-aminonaphthalene; 68 = 2-aminobiphenyl; 69 = 4-aminobiphenyl; 70 = 4-phenylazoaniline. Each peak corresponds to 50 pg of amine. (Reprinted from M. Longo, A. Cavallaro, *J. Chromatogr. A*, **753**, 91–100 (1996), Copyright 1996, with permission of Elsevier Science.)

Ammonia is the only base formed during the reductive catalytic decomposition of the organic compounds. This gaseous product is then dissolved in HCl–ethanol and the change in solvent conductivity is measured. Therefore, other reduction products do not cause an increase in conductivity in this solvent. The carrier gas and hydrogen reducing gas must be very pure for noise-free, sensitive operation. The solvent is conditioned in an ion-exchange resin to maintain conductivity. Although the selectivity depends on the compound and element, the selectivity ratios for the detection of nitrogen and halogen compounds versus carbon detection are 10^4–10^6 : 1 and 10^6 : 1, respectively. The sensitivity of the ELCD to chlorinated compounds is about 0.1–1 pg. Whereas the solvent used in analysis with an ELCD can be of liquid chromatographic grade, the gas flow system must be kept very clean. Carbon particles in the conductivity cell can cause spiking and noise, as can a used or dirty nickel reaction tube.

The ELCD has been applied to the analyses of several amines.[31,206] Riggin et al.[31] evaluated some selective GC detectors, such as the NPD, ELCD and photoionization detector (PID), for the analysis of aniline compounds. PID can detect components which are ionized by an ultraviolet (UV) light source. The NPD and ELCD gave similar detection limits, although the NPD gave a more stable baseline and less day-to-day variation in response. On the other hand, the PID gave a lower response, although the detection limits were improved by using a higher energy source.

3 APPLICATION IN ENVIRONMENTAL ANALYSIS

3.1 General Techniques for Sample Preparation

Amines are present in the environment at low parts per billion levels or less. When environmental samples are analyzed by GC with a nonselective detector such as an FID, many peaks with the same retention times as those of amines are often present in the chromatograms. Although GC analyses with selective detectors can

save sample preparation, a cleanup procedure for the complex sample matrix is necessary for the reliable and accurate analysis of amines. In order to remove co-eluting interferences and to preconcentrate amines, the extraction and cleanup of the sample have been performed using a number of different purification techniques, such as distillation,[22,23,38,41,46,207] liquid–liquid extraction,[25,39,45,49–55,95–103,111–118,121] adsorption with various adsorbent tubes or columns,[11,19,21,31,36,123,168–170] preparative high-performance liquid chromatography (HPLC),[156,157,162,163,165–167] solid-phase extraction (SPE) with various cartridges[13,33,34] and solid-phase microextraction (SPME).[27,35,47,48,93] Some amines can be separated by acid–base partitioning, because these compounds can be extracted with organic solvents not at low pH (<1), but at high pH (>10). Adsorption with an adsorbent tube or column is effective for the separation of amines, but it is time-consuming. Blue-cotton,[161,208] blue-rayon[158,165,166,171,208] or blue-chitin,[209] cotton, rayon or chitin bearing covalently linked copper phthalocyanine trisulfonate as a ligand can selectively adsorb heterocyclic amines and other mutagens/carcinogens having polycyclic planar molecular structures. Normal- or reversed-phase HPLC fractionations are also useful for sample preparation. SPE is simple and rapid, and good recoveries of amines are obtained. SPME, recently developed by Pawliszyn,[210] is an extraction technique using a fused-silica fiber that is coated outside with an appropriate stationary phase, and saves preparation time, solvent purchase and disposal costs, and can improve the detection limits. It has been used routinely in combination with GC and GC/MS and has been successfully applied to a wide variety of compounds, including amines.

3.2 Aliphatic Amines

GC methods for the determination of aliphatic amines using selective detectors have been applied to various environmental samples such as air,[13–18,21–23] cigarette smoke,[52] water[18,36–41] and soil.[22,23,38,41] Many of these methods have been used for free amines,[13,15,17,18,21,41,44,205] and the trace analysis of low-molecular-mass aliphatic amines in air has been performed with nitrogen-selective detectors such as NPD, CLD and GC/MS types by direct injection or the headspace technique. Derivatization methods have also been used with water and soil samples, [36–38] because these samples cannot be analyzed directly. Kuwata et al.[13] analyzed low-molecular-mass aliphatic amines in air samples in and around livestock houses by Sep-Pak C_{18} cartridge treatment and subsequent GC with thermionic detection, and detected methylamine (0–12.4 ppb) and trimethylamine (0.28–69.7 ppb) (Figure 2b and c). Gronberg et al.[17,18] developed a method for the determination of short-chain aliphatic amines in ambient air based on impinger sampling in dilute H_2SO_4, selective enrichment across a polytetrafluoroethylene (PTFE) gas membrane and quantification by gas chromatography/thermionic specific detection (GC/TSD). The enrichment step was carried out in a flow system directly connected to the GC column and methyl-, dimethyl-, trimethyl-, diethyl- and triethylamine were detected at the parts per trillion level. Kataoka et al.[52] reported the determination of secondary amines as their *N*-diethylthiophosphoryl derivatives in cigarette smoke by gas chromatography/flame photometric detector (GC/FPD). This method is selective and sensitive to secondary amines, and the detection limits of amines are 0.05–0.2 pmol. By using this method, it was confirmed that dimethylamine, pyrrolidine, piperidine and morpholine occur in main- and side-stream smokes of cigarettes, and the contents of these amines in side-stream smoke are very high compared with those in main-stream smoke. Kashihira et al.[21] analyzed trimethylamine in freezer air by GC/CLD after preconcentration with a Tenax-GC® tube, and detected 2.4 ppb of trimethylamine. Koga et al.[36] reported the determination of trace amounts of 12 aliphatic primary and secondary amines as their dinitrophenyl derivatives in waters from sewage and river by GC/MS. By using this method, low-molecular-mass aliphatic amines in water samples can be determined in the range 1–3 ppb. Avery and Junk[37] reported the determination of trace amounts of seven aliphatic primary amines in waters from tap, river and oil shale process by GC/MS based on Schiff base formation with PFBA. This method is specific for primary amines and the detection limit is 10 ppb with no sample transfers or manipulations being necessary for sample volumes of 0.5 mL. Terashi et al.[38] reported the determination of eight primary and secondary amines in river water, seawater and sea sediment by GC/MS in the SIM mode based on distillation of the sample under alkaline conditions and subsequent benzenesulfonylation. In this method, the detection limits of amines in water and sediment are 0.02–2 and 0.5–50 ppm, respectively. Pietsch et al.[39] analyzed aliphatic and alicyclic amines in wastewater samples by GC/MS after derivatization with trichloroethyl chloroformate and liquid–liquid extraction, and detected morpholine and piperidine. Sacher et al.[40] analyzed primary and secondary amines in industrial wastewater and surface water samples by GC/MS as their *N*-benzenesulfonyl derivatives, and detected methylamine, dimethylamine, morpholine and ethanolamine in concentrations up to 100 $\mu g\,L^{-1}$. Furthermore, ethylamine and diethylamine were found in concentrations up to 30 $\mu g\,L^{-1}$. Tsukioka et al.[41] reported the determination of lower aliphatic tertiary amines in pond water and bottom sediments (Figure 15a–c). These amines were determined by headspace GC/MS in the SIM mode

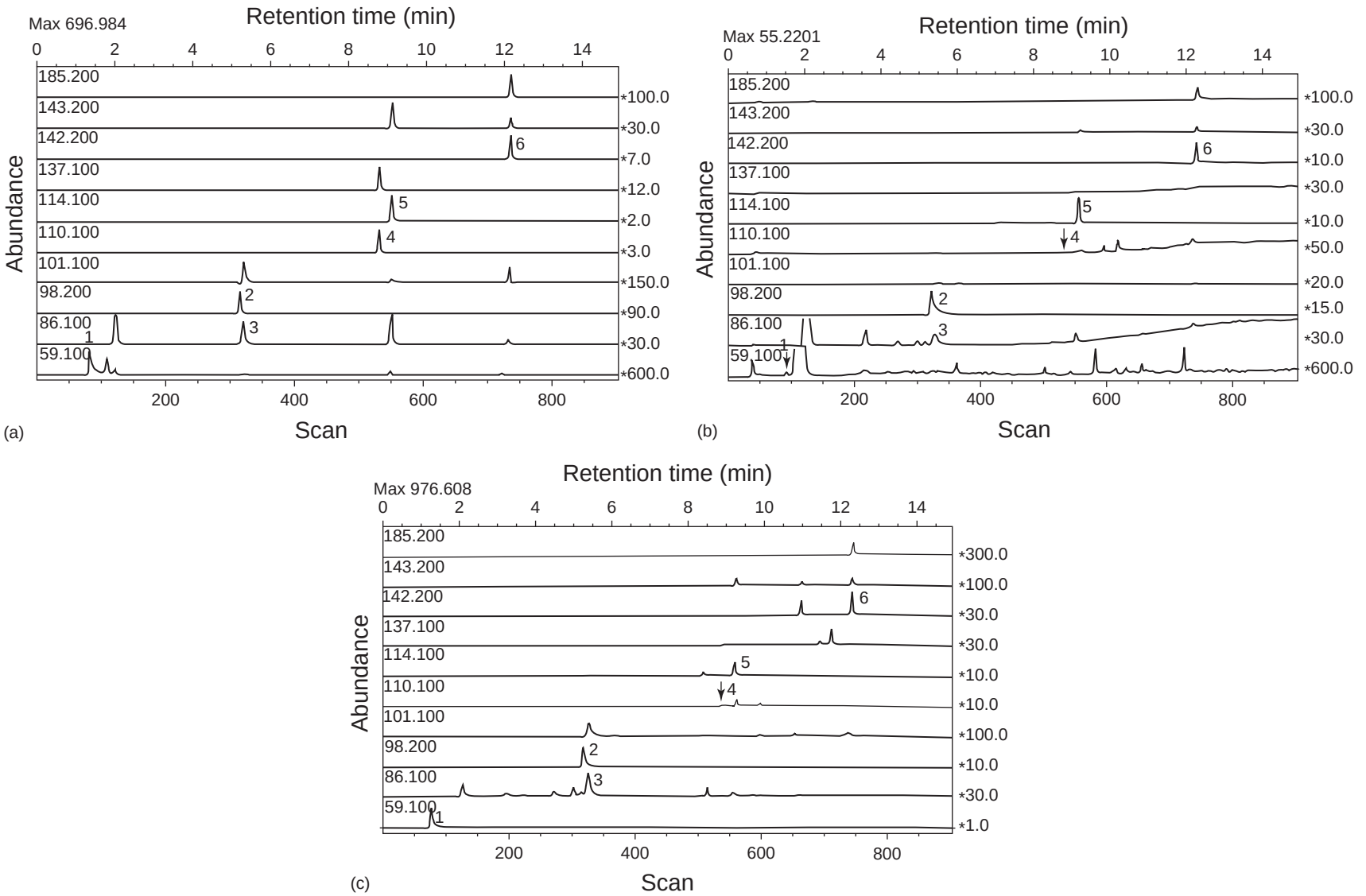

Figure 15 SIM chromatograms of standard aliphatic amines and environmental samples. (a) Standard aliphatic amines; (b) pond water from Suwa Lake; (c) bottom sediment from Tokyo Bay. GC/MS conditions: GC column, TC-1 (30 m × 0.53 mm i.d.); column temperature, initially 40 °C for 2 min, increased to 180 °C at 20 °C min^{-1} and held at 180 °C for 5 min; injection temperature, 180 °C; separator and ion source temperatures, 250 °C; ionization volatage, 70 eV; carrier gas, helium at 75 cm^3 s^{-1}. Peaks: 1 = trimethylamine; 2 = triethylamine-d_{15}, 3 = triethylamine; 4 = triallylamine; 5 = tripropylamine; 6 = tributylamine. (Reprinted from T. Tsukioka, H. Ozawa, T. Murakami, *J. Chromatogr.*, **642**, 395–400 (1993), Copyright 1993, with permission from Elsevier Science.)

after distillation. The detection limits for the final 40-mL samples were 5–50 ng and this method showed excellent sensitivity and selectivity.

3.3 Aromatic Amines

GC methods for the determination of aromatic amines using selective detectors have been applied to various environmental samples such as air,[19,20,24,25] cigarette smoke,[11,25,49–51,53] wastewater [31–34] and soil.[55] Becher[19] reported the determination of several aromatic amines in workplace atmosphere by gas chromatography/nitrogen–phosphorus detector (GC/NPD). Air samples are collected in three-section silica gel tubes and the amines on silica gel are eluted with 2-butanone. Garcia-Valcarcel et al.[55] determined several dinitroaniline herbicides in environmental samples by GC/NPD after solvent extraction. Schmidt et al.[33] reported a method for the selective determination of aromatic amines in water samples. This method is based on the SPE at pH 9, subsequent derivatization to the corresponding iodobenzenes and GC/ECD analysis. Aniline and nitroaniline compounds were detected in groundwater samples (Figure 16). Pieraccini et al.[25,53] reported the determination of 17 primary aromatic amines as their pentafluoropropionamides in cigarette smoke and indoor air by GC/MS in the SIM mode. Cigarettes are smoked in a laboratory-made smoking machine and the amines in the main- and side-stream smokes are trapped in dilute hydrochloric acid. It is confirmed that side-stream smoke contains total levels of aromatic amines about 50–60 times higher than those in main-stream smoke, and some aromatic amines in ambient air such as offices and houses may be derived from considerable contamination of aromatic amines in side-stream smoke. Kataoka et al.[50] analyzed aromatic amines in the cigarette smoke samples as their *N*-*n*-propoxycarbonyl derivatives (Figure 3b) and their *N*-dimethylthiophsphoryl derivatives by GC/NPD and GC/FPD (P mode), respectively, after collection of smoke samples in the same manner. Okumura et al.[45] determined aniline and related aromatic amines in river water and sediment samples by GC/MS in the SIM mode after liquid–liquid extraction and steam distillation. The detection limits of the anilines in water and sediment samples were 4.2–31 pg mL^{-1} and 1.2–4.0 ng mL^{-1}, respectively. Longo and Cavallaro[46] analyzed aromatic amines in groundwater samples as their heptafluorobutyryl derivatives by gas chromatography/electron-capture negative-ion chemical ionization mass spectrometry (GC/ECNICIMS) after concentration with a Kuderna–Danish evaporator. Muller et al.[47,48] reported the determination of aromatic amines by SPME and GC/MS. This method was simple, rapid, precise and sensitive and various groundwater samples could be directly analyzed (Figure 17a and b). Guan et al.[35] also analyzed nitroaniline herbicides in river water by SPME and GC/ECD.

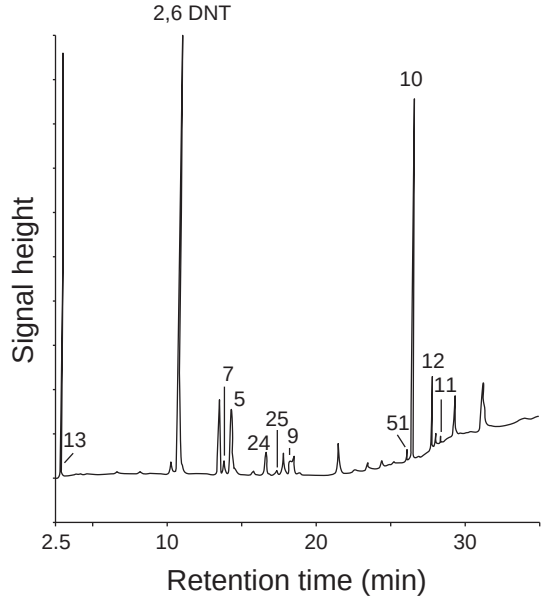

Figure 16 GC/ECD of a groundwater from Stadtallendorf. GC conditions and peak numbers as in Figure 9. (Reprinted from T.C. Schmidt, M. Less, R. Haas, E. von Low, K. Steinbach, G. Stork, *J. Chromatogr. A*, **810**, 161–172 (1998), Copyright 1998, with permission from Elsevier Science.)

3.4 *N*-Nitrosamines

GC determinations of *N*-nitrosamines in environmental samples have been carried out in indoor and outdoor air,[94,95] combustion smoke,[52,82,96–103] water,[104–107,110] rubber products,[111–118] metalworking fluids,[120] drug formulations[121,122] and agricultural chemical formulations.[110,123,124] In most of them, *N*-nitrosamines are directly analyzed as the free form by GC/TEA, based on the detection of the chemiluminescence emitted from a reaction between releasing NO radicals and ozone after thermal cleavage of N–NO bond in *N*-nitroso compounds. Fadlallah et al.[95] determined volatile nitrosamines in the factory environment by GC/TEA (Figure 11b). Air samples were collected by drawing air through a Thermosorb/N cartridge, and then the sorbent was eluted with methanol–dichloromethane (1 : 3). Brunnemann et al.[97–100] analyzed volatile and tobacco-specific *N*-nitrosamines in main- and side-stream smoke of cigarette and tobacco by GC/TEA. The smoke sample was collected in citrate buffer (pH 4.5) containing 20 mM ascorbic acid and then extracted with dichloromethane. *N*-Nitrosodimethylamine, *N*-nitrosopyrrolidine, *N*-nitrosomethylethylamine and some tobacco-specific nitrosamines such as *N*-nitrosonornicotine, 4-(methyl-

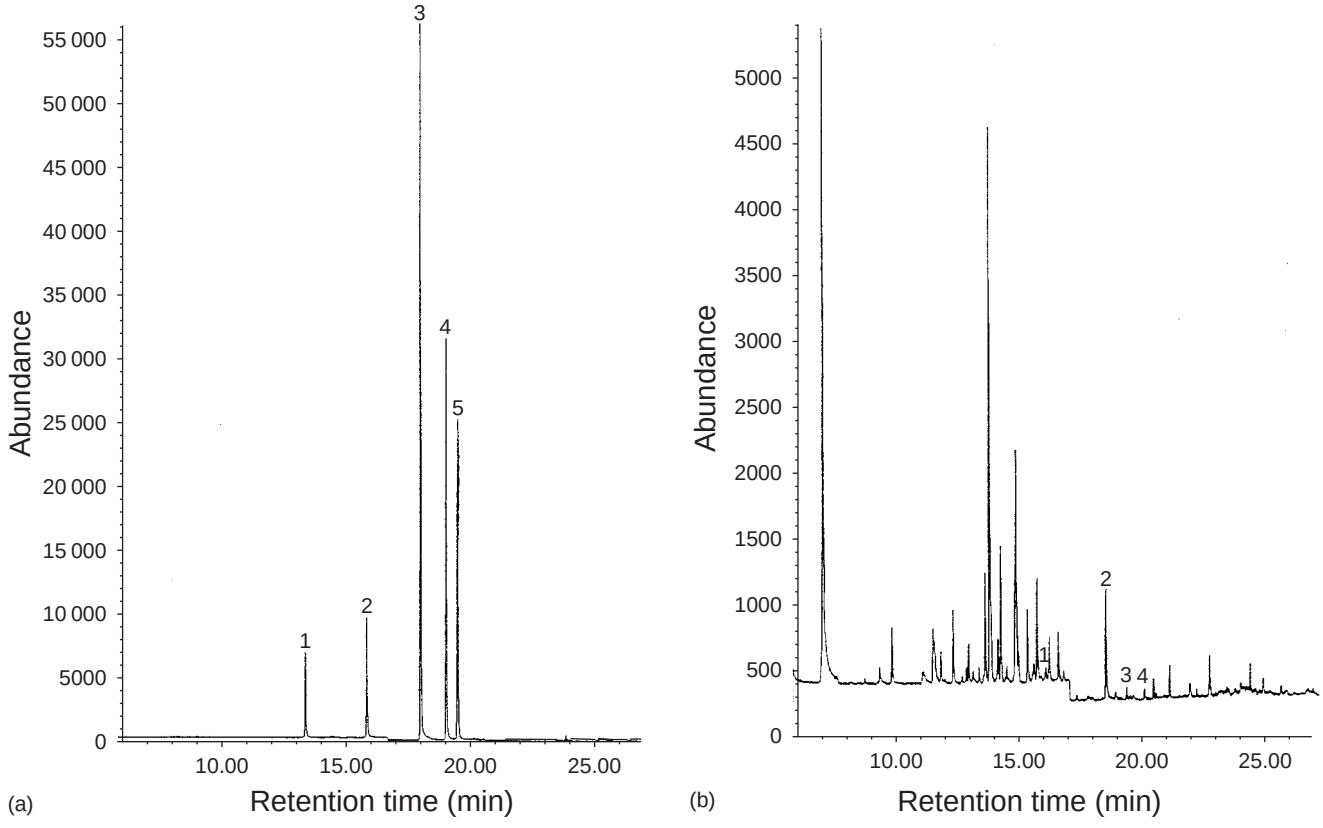

Figure 17 TICs of spiked water and groundwater samples extracted by SPME. (a) Water solution spiked with 10 ng mL^{-1} aromatic amines; (b) groundwater collected near an industrial area north of Milan, Italy. GC/MS conditions: GC column, base-deactivated PTA-5 (30 m × 0.25 mm i.d.); column temperature, initially at 60 °C for 5 min, increased to 250 °C at 10 °C min^{-1} and held at 250 °C for 3 min; injector and detector temperatures, 260 and 280 °C, respectively; carrier gas, helium at 40 kPa; detection, SIM mode. Peaks: 1 = *o*-toluidine; 2 = *p*-chloroaniline; 3 = 2,4- and 2,5-dichloroaniline; 4 = 3,5-dichloroaniline; 5 = 3,4-dichloroaniline. (Reprinted from L. Muller, E. Fattore, E. Benfenati, *J. Chromatogr. A*, **791**, 221–230 (1997), Copyright 1997, with permission from Elsevier Science.)

nitrosamino)-4-(3-pyridyl)butanal and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone were detected. The *N*-nitrosamines in environmental tobacco smoke, to which both smokers and nonsmokers are exposed, have received a great deal of attention as a source of indoor pollution. Tomkins et al.[106,107] determined *N*-nitrosodimethylamine in drinking water and groundwater at the parts per trillion level by GC/CLD (nitrogen mode) (Figure 18a–c). Furthermore, they developed new sample preparation procedure for the extraction of *N*-nitrosamines using a C_{18} membrane extraction disk. Thompson et al.[114,115] reported the determination of *N*-nitrosamines in rubber nipples and pacifiers at the parts per billion level by GC/TEA (Figure 19a–c). In these samples, *N*-nitrosodibutylamine was the principal *N*-nitrosamine found, along with trace amounts of *N*-nitrosodimethylamine, *N*-nitrosodiethylamine and *N*-nitrosopiperidine. The occurrence of *N*-nitrosamines in baby bottle rubber nipples and pacifiers is of special concern because traces of these amines may migrate to infant saliva during sucking, and then be ingested. Although GC/TEA can be used as a sensitive and specific method for the analysis of *N*-nitroso compounds, it is too expensive to use in many laboratories. As an alternative method, Kataoka et al.[52,96] reported the determination of seven *N*-nitrosamines by GC/FPD (P mode). The method is based on denitrosation with hydrobromic acid to produce the corresponding secondary amines and subsequent diethylthiophosphorylation of secondary amines. By using this method, it was confirmed that *N*-nitrosodimethylamine, *N*-nitrosopyrrolidine and *N*-nitrosopiperidine occur in main- and side-stream smoke of cigarettes. In addition, some *N*-nitrosamines in various environmental samples were determined by GC/NPD,[123] GC/ECD[120] and GC/MS.[103,105,124]

3.5 Heterocyclic Amines

Most heterocyclic amines are polar and of low volatility, and tend to elute as broad and tailing peaks owing to strong adsorption on the column and injector during

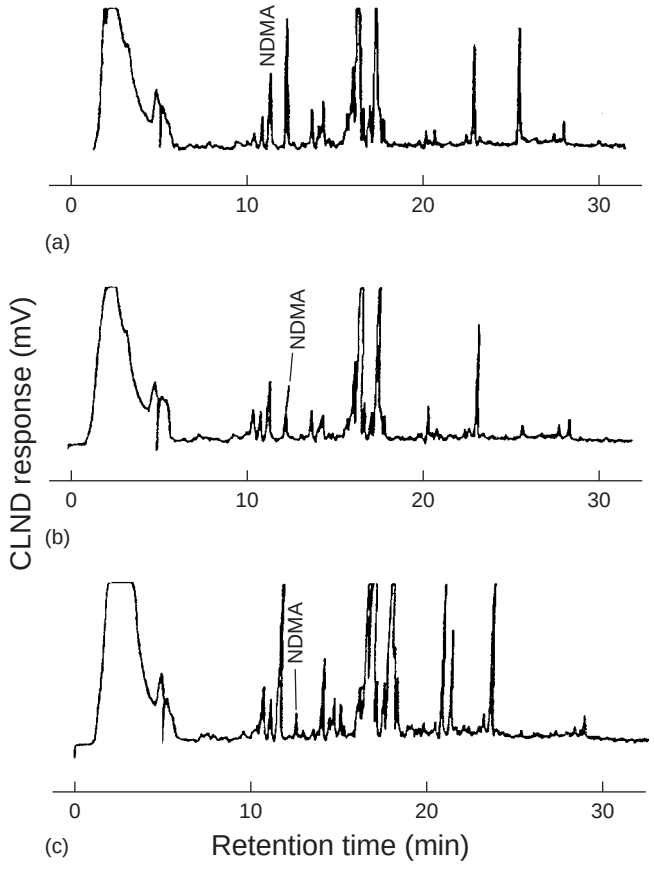

Figure 18 Detection of *N*-nitrosodimethylamine in two sources of drinking water, (a) and (b). The corresponding synthetic groundwater blank (c) exhibits a smaller concentration of *N*-nitrosodimethylamine than in either source. GC conditions: column, Rtx-200 (cross-bond trifluoropropylmethyl, 30 m × 0.53 mm i.d.); column temperature, initially isothermal at 35 °C for 5 min, programmed from 35 to 175 °C at 6 °C min^{-1} and held at 175 °C for 5 min; injector temperature, 150 °C, carrier gas, helium at 4 mL min^{-1}; detector, CLD (nitrosamine-selective mode); detector furnace, 250 °C. NDMA = *N*-nitrosodimethylamine. (Reprinted by permission from B.A. Tomkins, W.H. Griest, C.E. Higgins, *Anal. Chem.*, **67**, 4387–4395 (1995), Copyright 1995, American Chemical Society.)

GC analysis, when they are analyzed without derivatization. Therefore, the analysis of these amines has been generally carried out by HPLC.[211] For the GC analysis of heterocyclic amines, several derivatizations using acetic anhydride, trifluoroacetic anhydride, HFBA, pentafluorobenzyl bromide, 3,5-bistrifluoromethylbenzyl bromide and 3,5-bistrifluoromethylbenzoyl chloride have been tested.[141–144,146,168,173,176] However, acylation with acid anhydrides yielded derivatives with very poor GC properties. The alkylation products with pentafluorobenzyl bromide, 3,5-bistrifluoromethylbenzyl bromide and 3,5-bistrifluoromethylbenzoyl chloride had good GC properties for some heterocyclic amines. However, these methods gave a mixture of mono- and dialkylated forms and were used for GC/MS analysis but not for GC analysis. Recently, Kataoka and Kijima[198] developed a simple and rapid derivatization method for the GC analysis of mutagenic heterocyclic amines. Ten heterocyclic amines were converted into their *N*-dimethylaminomethylene derivatives with *N*,*N*-dimethylformamide dimethyl acetal and measured by GC/NPD. As shown in Figure 20(a), these heterocyclic amines were separated within 7 min, although AαC and 2-aminodipyrido[1,2-*a*:3′,2′-*d*]imidazole (Glu-P-2) co-eluted. The detection limits of these compounds ranged from 2 to 15 pg. By using this method, AαC, 3-amino-1,4-dimethyl-5*H*-pyrido[3,4-*b*]indole (Trp-P-1), 3-amino-1-methyl-5*H*-pyrido[3,4-*B*]indole (Trp-P-2), 2-amino-6-methyldipyrido[1,2-*a*:3′,2′-*d*]imidazole (Glu-P-1), IQ, 2-amino-3,4-dimethylimidazo[4,5-*f*]quinoline (MeIQ), and 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]-pyridine (PhIP) were determined in several combustion smoke samples[158] (Figure 20b and c). On the other hand, few GC/MS (SIM mode) data using positive-ion electron ionization and negative-ion CI on the analysis of heterocyclic amines after derivatization are available. Vainiotalo et al.[168] analyzed MeIQx and DiMeIQx in cooking fumes as their 3,5-bistrifluoromethylbenzyl derivatives by GC/MS in the SIM mode (Figure 21a and b). The cooking fumes were collected through a glass-fiber funnel into a sampler which consisted of two glass-fiber filters and an XAD-2 sorbent tube. Although these heterocyclic amines give a mixture of mono- and dibistrifluoromethylbenzyl derivatives, the spectra of the dibistrifluoromethylbenzyl derivatives possess high-mass fragment ions suitable for SIM work. When ions at *m*/*z* 438 MeIQx and *m*/*z* 452 DiMeIQx were specially monitored, the detection limit of these compounds was 2 pg.

4 CONCLUSIONS

Environmental pollution is recognized as an important problem for public health, because it is very close to the daily life of people throughout the world. The occurrence and toxicity of hazadous chemicals in the biosphere have received a great deal of attention in recent years. Neglect of environmental pollution caused widespread contamination of air, soil and water, unhealthy conditions for workers and the loss of many precious natural resources. Therefore, environmental measurements are essential as a significant index of risk assessment not only in establishing the quantitative relationship between exposure and response to these pollutants, but also in establishing the natural baseline conditions in the

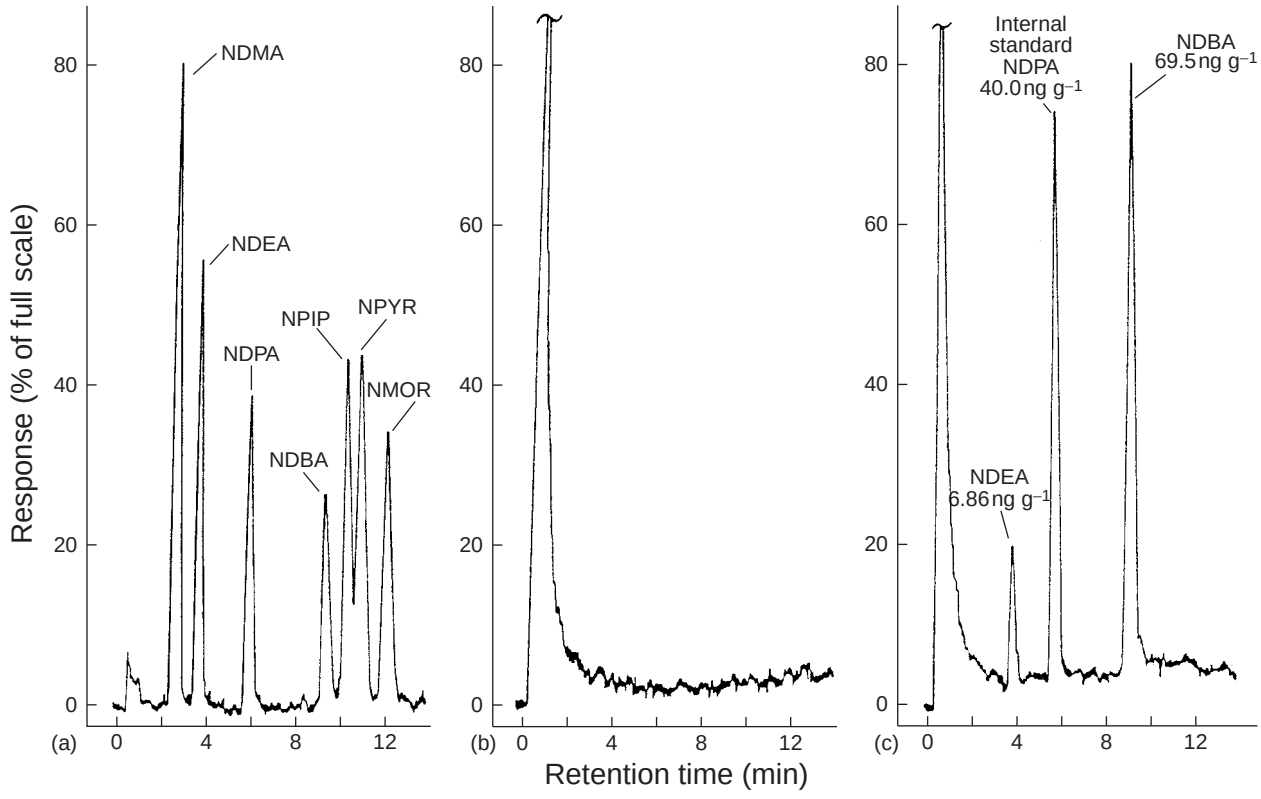

Figure 19 GC/TEA of standard *N*-nitrosamines and pacifier sample. (a) Standard mixture containing 50 ng mL^{-1} each of *N*-nitrosamine; (b) reagent blank; (c) pacifier sample. GC conditions: column, 10% Carbowax™–2% KOH on Chromosorb® W AW (6 ft × 4 mm i.d.); column temperature, programmed from 150 to 190 °C at 4 °C min^{-1}; injector temperature, 250 °C, carrier gas, nitrogen at 40 mL min^{-1}; detector, TEA; detector furnace, 500 °C; cold trap, −150 °C. Peaks: NDMA = *N*-nitrosodimethylamine; NDEA = *N*-nitrosodiethylamine; NDPA = *N*-nitrosodipropylamine; NDBA = *N*-nitrosodibutylamine; NPIP = *N*-nitrosopiperidine; NPYR = *N*-nitrosopyrrolidine; NMOR = *N*-nitrosomorpholine. (Reprinted from S.M. Billedeau, H.C. Thompson, B.J. Miller, M.L. Wind, *J. Assoc. Off. Anal. Chem.*, **69**, 31–34 (1986), Copyright 1986, by permission of AOAC International.)

environment. Toxic amines, described in this article, are widely distributed in a number of ambient environmental components such as air, diesel exhaust, cigarette smoke, cooking fumes, river water, sewage water, soil and chemical products. The presence of these amines in the environment may be more extensive than previously thought and humans are continually exposed to these compounds in normal daily life. Therefore, monitoring of amines in the environment is very important to preserve human health and the natural environment, and practical and reliable methods for determining accurate exposure levels of these amines need to be established. For this purpose, various analytical methods have been developed. In particular, GC is one of the most useful techniques for the simultaneous determination of most amines. However, the amines in environmental samples are generally present at very low concentrations and are often to be found among complex matrices containing a number of co-existing substances. Therefore, GC detectors for environmental analysis must be not only sensitive to minute amounts of analytes, but also selective enough to discriminate against reasonable amounts of co-existing substances.

The nitrogen-selective thermionic detectors, the nitrosamine-specific detector TEA and GC/MS have been widely used for obtaining high sensitivity and selectivity. In addition, conversions of amines into fluorinated derivatives to introduce an ECD response and into sulfur- and phosphorus containing derivatives to introduce an FPD response have been devised for these purposes.

GC with nitrogen-selective thermionic detection is highly sensitive and selective for amine analysis, and available in most laboratories. Although the selectivity and sensitivity of thermionic detectors depend on the flow rates of hydrogen, air, make-up gas and carrier gas, a relatively stable response is realized using a flameless type of thermionic detector. GC/MS, capable of measuring simultaneously retention time and molecular mass, is a powerful technique for the identification and quantification of amines in complex matrix samples. The TEA is also selective and sensitive for *N*-nitroso compounds, but its response depends on the pyrolysis temperature. Although

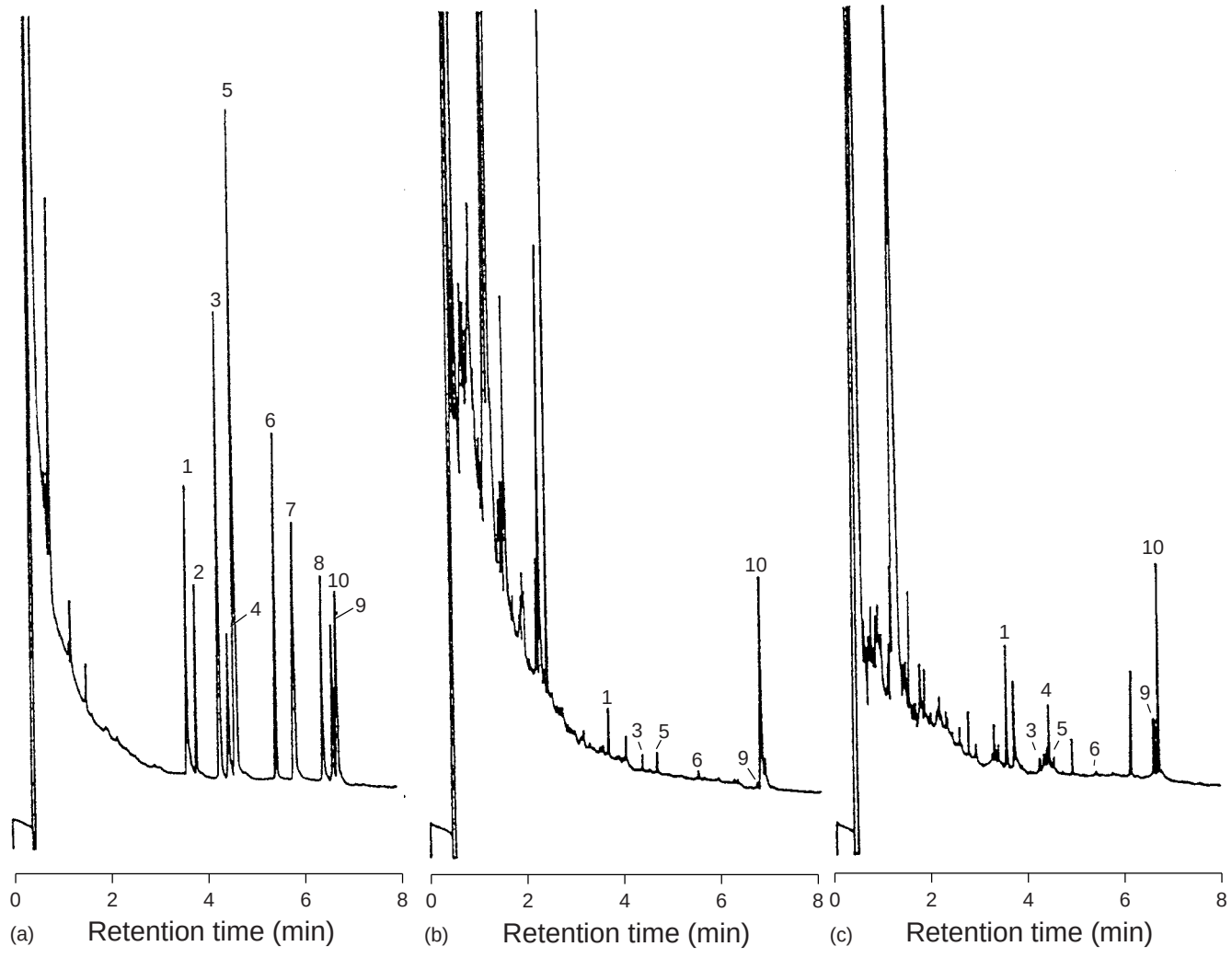

Figure 20 Gas chromatograms obtained from (a) standard heterocyclic amines, (b) cigarette side-stream smoke and (c) combustion smoke of semidried fish. GC conditions: column, connected DB-1 (10 m × 0.25 mm i.d.) and DB-17ht (10 m × 0.25 mm i.d.); column temperature, programmed from 230 to 280 °C at 10 °C min^{-1}, from 280 to 330 °C at 25 °C min^{-1} and held at 330 °C for 1 min; injection and detector temperatures, 340 °C; carrier gas, helium, programmed from 180 to 230 kPa at 10 kPa min^{-1}, from 230 to 280 kPa at 25 kPa min^{-1} and held at 280 kPa for 1 min; split ratio, 10 : 1; detector, NPD. Peaks: 1 = AαC (2-amino-9*H*-pyrido[2,3-*b*]indole); 2 = Glu-P-1; 3 = Trp-P-1; 4 = Trp-P-2; 5 = IQ; 6 = MeIQ; 7 = MeIQx (2-amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline); 8 = DiMeIQx (2-amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline); 9 = PhIP; 10 = TriMeIQx (2-amino-3,4,7,8-tetramethylimidazo[4,5-*f*]quinoxaline) (internal standard). (Reproduced by permission of Springer-Verlag, from H. Kataoka, K. Kijima, G. Maruo, *Bull. Environ. Contam. Toxicol.*, **60**, 60–67 (1998), Copyright Springer-Verlag.)

GC/MS in the SIM mode and GC/TEA are most effective for the analysis of a specified amine, they require sophisticated and expensive equipment that is beyond the reach of many laboratories. GC/FPD and GC/ECD require the derivatization of amines to the corresponding halogen- and sulfur- or phosphorus-containing compounds, respectively, but thereby offer increased selectivity and sensitivity for specific amines. GC methods with selective detectors, described in this article, have advantages and disadvantages and so far there are no all-powerful analytical methods.

The choice of an analytical method depends on the presence of amines in the environment at low parts per billion levels or less and the variety and complexity of the sample. In order to achieve an efficient isolation and preconcentration of amines, several methods of sample preparation for the analysis of various environmental samples have been developed using a number of different purification techniques, such as distillation, liquid–liquid extraction, column chromatography, SPE, extraction with mutagen specific adsorbent and preparative HPLC. The combination of these techniques is used successfully for

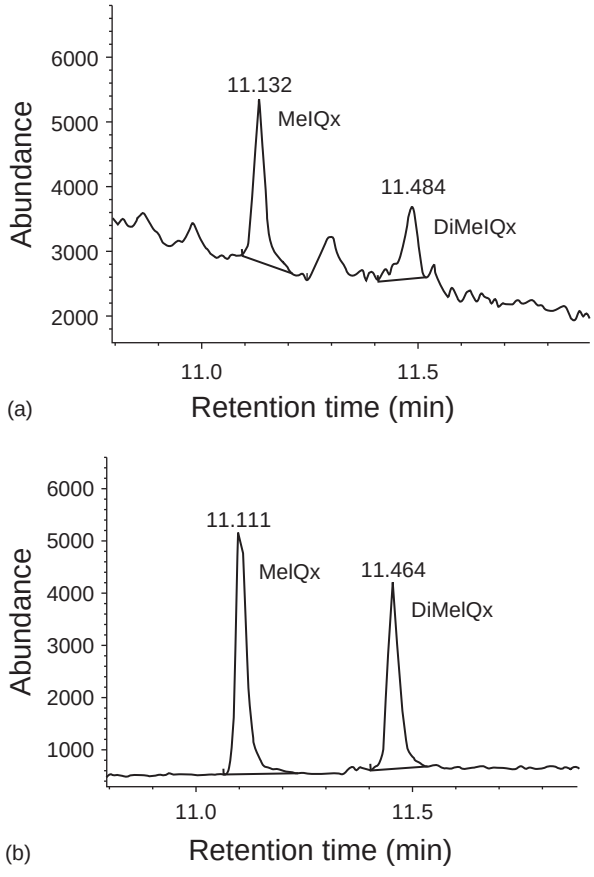

Figure 21 GC/MS (SIM mode) traces of (a) fumes from fried beef/pork and (b) a standard sample containing 5 ng of MeIQx, 1.5 ng of 4,8-DiMeIQx and 2.5 ng of 7,8-DiMeIQx. The base peaks of the di-bis-TFMB (trifluoromethylbenzyl) derivatives were monitored (MeIQx, *m*/*z* 438; DiMeIQx, *m*/*z* 452). GC/MS conditions: column, HP-5 (20 m × 0.2 mm i.d.); column temperature, initially isothermal at 70 °C for 1 min, increased to 280 °C at 25 °C min^{-1} and held at 280 °C for 5 min; injector temperature, 250 °C; carrier gas, helium at 1 mL min^{-1}; ionization voltage, 70 eV; emission current, 300 μA. (Reproduced by permission of Springer-Verlag from S. Vainiotalo, K. Matveinen, A. Reunanen, *Fresenius' J. Anal. Chem.*, **345**, 462–466 (1993), Copyright Springer-Verlag.)

the extraction and purification of amines. However, many of these methods are time-consuming, laborious and give low recoveries for some specific amines. On the other hand, the SPME method is simple and rapid, and is directly applicable to the sample.

Environmental analysis has benefited from the recent advances in GC, including high-resolution columns for complex mixtures and improved sensitivity and selectivity in detection. Powerful selective detection tools coupled with the many technological advances in capillary columns, software and hardware over the past decade have added a new dimension. Measurement techniques in environmental analysis will become increasingly more sensitive and selective in the future. For this purpose, further developments in selective GC detectors and in capillary columns, e.g. shorter inactive columns with smaller internal diameters giving ultrahigh column efficiency and speed, higher temperature phases and exterior coating for the fused-silica tubing, will permit the analysis of both high-temperature and highly volatile amines. Furthermore, simple, rapid and automated separation and analysis of compounds in various environmental samples will be achieved by combination with convenient sample preparation techniques such as SPME.

ACKNOWLEDGMENTS

I thank Hewlett-Packard, Elsevier Science, Springer-Verlag, the American Chemical Society and the AOAC International for permission to reproduce many of the figures. I also thank Sarah Brier of the University of Waterloo for correcting my rambling English writing style.

ABBREVIATIONS AND ACRONYMS

AαC	2-Amino-9*H*-pyrido[2,3-*b*]indole
AFID	Alkali Flame Ionization Detector
CI	Chemical Ionization
CLD	Chemiluminescent Detector
DiMeIQx	2-Amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline
ECD	Electron-capture Detector
ECNICI	Electron-capture Negative-ion Chemical Ionization
EI	Electron Impact Ionization
EIMS	Electron Impact Ionization Mass Spectrometry
ELCD	Electrolytic Conductivity Detector
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
FTD	Flame Thermionic Detector
GC	Gas Chromatography
GC/CLD	Gas Chromatography/Chemiluminescent Detection
GC/ECD	Gas Chromatography/Electron-capture Detector
GC/ECNICIMS	Gas Chromatography/Electron-capture Negative-ion Chemical Ionization Mass Spectrometry
GC/FPD	Gas Chromatography/Flame Photometric Detector
GC/MS	Gas Chromatography/Mass Spectrometry

GC/NPD	Gas Chromatography/Nitrogen–Phosphorus Detector
GC/TEA	Gas Chromatography/Thermal Energy Analysis
GC/TSD	Gas Chromatography/Thermionic Specific Detection
Glu-P-1	2-Amino-6-methyldipyrido[1,2-*a*:3′,2′-*d*]imidazole
Glu-P-2	2-Aminodipyrido[1,2-*a*:3′,2′-*d*]-imidazole
HFBA	Heptafluorobutyric Anhydride
HPLC	High-performance Liquid Chromatography
IARC	International Agency for Research on Cancer
IQ	2-Amino-3-methylimidazo[4,5-*f*]quinoline
ITD	Ion Trap Mass Detector
MeIQ	2-Amino-3,4-dimethylimidazo[4,5-*f*]quinoline
MeIQx	2-Amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline
MSD	Mass-selective Detector
NICIMS	Negative-ion Chemical Ionization Mass Spectrometry
NPD	Nitrogen–Phosphorus Detector
OPA	*o*-Phthaldialdehyde
PFBA	Pentafluorobenzaldehyde
PFPA	Pentafluoropropionic acid
PhIP	2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine
PID	Photoionization Detector
PTFE	Polytetrafluoroethylene
SIM	Selected-ion Monitoring
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TEA	Thermal Energy Analyzer
TFMB	Trifluoromethylbenzyl
TIC	Total Ion Chromatogram
TriMeIQx	2-Amino-3,4,7,8-tetramethylimidazo[4,5-*f*]quinoxaline
Trp-P-1	3-Amino-1,4-dimethyl-5*H*-pyrido[3,4-*b*]indole
Trp-P-2	3-Amino-1-methyl-5*H*-pyrido[3,4-*b*]indole
TSD	Thermionic Specific Detector
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

REFERENCES

1. J.J. Peirce, R.F. Weiner, P.A. Vesilind, *Environmental Pollution and Control*, 4th edition, Butterworth-Heinemann, Woburn, MA, 1998.
2. H.A.H. Billiet, 'Amines from Environmental Sources', *J. Chromatogr. Libr.*, **51B**, 583–595 (1992).
3. H. Kataoka, 'Derivatization Reactions for the Determination of Amines by Gas Chromatography and Their Applications in Environmental Analysis', *J. Chromatogr. A*, **733**, 19–34 (1996).
4. G.D. Clayton, F.E. Clayton (eds.), *Patty's Industrial Hygiene and Toxicology*, Wiley-Interscience, New York, Vol. 28, 1981.
5. L. Parmeggiani (ed.), *Encyclopedia of Occupational Health and Safety*, International Labour Office, Geneva, Vol. 1, 1983.
6. P.F. Vogt, J.J. Geralis, *Ullman's Encyclopedia of Industrial Chemistry*, 5th edition, VCH, Weinheim, Vol. A2, 1985.
7. S.E. Manahan, *Environmental Chemistry*, 4th edition, Lewis, Chelsea, MI, 1990.
8. R.F. Straub, R.D. Voyksner, J.T. Keever, 'Determination of Aromatic Amines Originating from Azo Dyes by Hydrogen–Palladium Reduction Combined with Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **65**, 2131–2136 (1993).
9. G.I. Baughman, E.J. Weber, 'Transformation of Dyes and Related Compounds in Anoxic Sediment: Kinetics and Products', *Environ. Sci. Technol.*, **28**, 267–276 (1994).
10. L. Hansen, B. Aakesson, J. Sollenberg, T. Lundh, 'Determination of *N*-Methylmorphorine in Air Samples from a Polyurethane Foam Factory. Comparison Between Two Methods', *Scand. J. Work Environ. Health*, **12**, 66–69 (1986).
11. G. Seeber, M.R. Buchmeiser, G.K. Bonn, T. Bertsch, 'Determination of Airbone, Volatile Amines from Polyurethane Foams by Sorption onto a High-capacity Cation-exchange Resin Based on Poly(succinic acid)', *J. Chromatogr. A*, **809**, 121–129 (1998).
12. G. Skarping, M. Dalene, P. Lind, 'Determination of Toluenediamine Isomers by Capillary Gas Chromatography and Chemical Ionization Mass Spectrometry with Special Reference to the Biological Monitoring of 2,4- and 2,6-Toluenediisocyanate', *J. Chromatogr. A*, **663**, 199–210 (1994).

13. K. Kuwata, E. Akiyama, Y. Yamazaki, H. Yamasaki, Y. Kuge, Y. Kiso, 'Trace Determination of Low Molecular Weight Aliphatic Amines in Air by Gas Chromatography', *Anal. Chem.*, **55**, 2199–2201 (1983).
14. G. Audunsson, L. Mathiasson, 'Simultaneous Determination of Amines and Isocyanates in Working Atomspheres by Gas–Liquid Chromatography', *J. Chromatogr.*, **261**, 253–264 (1983).
15. G. Audunsson, L. Mathiasson, 'Direct Analysis of Free Amines in Salt Solutions at Sub-ppm Levels by Gas–Liquid Chromatography', *J. Chromatogr.*, **315**, 299–312 (1984).
16. G. Skarping, T. Bellander, L. Mathiasson, 'Determination of Piperazine in Working Atmosphere and Human Urine Using Derivatization and Capillary Gas Chromatography with Nitrogen- and Mass-selective Detection', *J. Chromatogr.*, **370**, 245–258 (1986).
17. L. Gronberg, P. Loukvist, J.A. Jonsson, 'Determination of Aliphatic Amines in Air by Membrane Enrichment Directly Coupled to Gas Chromatography', *Chromatographia*, **33**, 77–82 (1992).
18. L. Gronberg, P. Loukvist, J.A. Jonsson, 'Measurement of Aliphatic Amines in Ambient Air and Rain Water', *Chemosphere*, **24**, 1533–1540 (1992).
19. G. Becher, 'Glass Capillary Columns in the Gas Chromatographic Separation of Aromatic Amines. II. Application to Samples from Workplace Atmospheres Using Nitrogen-selective Detection', *J. Chromatogr.*, **211**, 103–110 (1981).
20. B. Andersson, K. Andersson, 'Determination of Heterocyclic Tertiary Amines in Air', *Appl. Occup. Environ. Hyg.*, **6**(1), 40–43 (1991).
21. N. Kashihira, K. Makino, K. Kirita, Y. Watanabe, 'Chemiluminescent Nitrogen Detector–Gas Chromatography and its Application to Measurement of Atmospheric Ammonia and Amines', *J. Chromatogr.*, **239**, 617–624 (1982).
22. C. Lee, B.L. Olson, 'Dissolved, Exchangeable and Bound Aliphatic Amines in Marine Sediments: Initial Results', *Org. Geochem.*, **6**, 259–264 (1984).
23. A. Van Nesta, R.A. Duce, C. Lee, 'Methylamines in the Marine Atmosphere', *Geophys. Res. Lett.*, **14**, 711–714 (1987).
24. E. Meichini, L. Boniforti, S.D. Marzio, 'Occupational Exposure to Airborne Aromatic Amines in Rubber Manufacturing: Determination by Gas Chromatography/Mass Spectrometry', *Toxicol. Environ. Chem.*, **20**, 9–16 (1989).
25. F. Luceri, G. Pieraccini, G. Moreti, P. Dolara, 'Primary Aromatic Amines from Side-stream Cigarette Smoke are Common Contaminants of Indoor Air', *Toxicol. Ind. Health*, **9**(3), 405–413 (1993).
26. S. Skarping, L. Renman, C. Sango, L. Mathiasson, M. Dalene, 'Capillary Gas Chromatographic Method for the Determination of Complex Mixture of Isocyanates and Amines', *J. Chromatogr.*, **346**, 191–204 (1985).
27. L. Pan, J.M. Chong, J. Pawliszyn, 'Determination of Amines in Air and Water Using Derivatization Combined with Solid-phase Microextraction', *J. Chromatogr. A*, **773**, 249–260 (1997).
28. P. Simon, C. Lemacon, 'Determination of Aliphatic Primary and Secondary Amines and Polyamines in Air by High-performance Liquid Chromatography', *Anal. Chem.*, **59**, 480–484 (1987).
29. C-.X. Gao, I.S. Krull, T. Trainor, 'Determination of Aliphatic Amines in Air by On-line Solid-phase Derivatization with HPLC/UV/FL', *J. Chromatogr. Sci.*, **28**, 102–108 (1990).
30. E.D. Zlotorzynska, W. Maruszak, 'Determination of Dimethylamine and Other Low-molecular-mass Amines Using Capillary Electrophoresis with Laser-induced Fluorescence Detection', *J. Chromatogr. B*, **714**, 77–85 (1998).
31. R.M. Riggin, T.F. Cole, S. Billets, 'Determination of Aniline and Substituted Derivatives in Wastewater by Gas and Liquid Chromatography', *Anal. Chem.*, **55**, 1862–1869 (1983).
32. R.C.C. Wegman, G.A.L. De Korte, 'Gas Chromatographic Determination of Aromatic Amines After Bromination in Surface Waters', *Int. J. Environ. Anal. Chem.*, **9**, 1–6 (1981).
33. T.C. Schmidt, M. Less, R. Haas, E. von Low, K. Steinbach, G. Stork, 'Gas Chromatographic Determination of Aromatic Amines in Water Samples After Solid-phase Extraction and Derivatization with Iodine. I. Derivatization', *J. Chromatogr. A*, **810**, 161–172 (1998).
34. M. Less, T.C. Schmidt, E. von Low, G. Stork, 'Gas Chromatographic Determination of Aromatic Amines in Water Samples After Solid-phase Extraction and Derivatization with Iodine. II. Enrichment', *J. Chromatogr. A*, **810**, 173–182 (1998).
35. F. Guan, K. Watanabe, A. Ishii, H. Seno, T. Kumazawa, H. Hattori, O. Suzuki, 'Headspace Solid-phase Microextraction and Gas Chromatographic Determination of Dinitroaniline Herbicides in Human Blood, Urine and Environmental Water', *J. Chromatogr. B*, **714**, 205–213 (1998).
36. M. Koga, T. Akiyama, R. Shinohara, 'Determination of Trace Aliphatic Amines in Water with 2,4-Dinitrophenylamine Derivatives by Gas Chromatography/Mass Spectrometry', *Bunseki Kagaku*, **30**, 745–750 (1981).
37. M.J. Avery, G.A. Junk, 'Gas Chromatography/Mass Spectrometry Determination of Water-soluble Primary Amines as Their Pentafluorobenzaldehyde Imides', *Anal. Chem.*, **57**, 790–792 (1985).
38. A. Terashi, Y. Hanada, A. Kido, R. Shinohara, 'Determination of Primary and Secondary Aliphatic Amines in the Environment as Sulphonamide Derivatives by Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **503**, 369–375 (1990).
39. J. Pietsch, S. Hampel, W. Schmidt, H.-J. Brauch, E. Worch, 'Determination of Aliphatic and Alicyclic Amines

in Water by Gas and Liquid Chromatography After Derivatization by Chloroformates', *Fresenius' J. Anal. Chem.*, **355**, 164–173 (1996).
40. F. Sacher, S. Lenz, H.-J. Brauch, 'Analysis of Primary and Secondary Aliphatic Amines in Waste Water and Surface Water by Gas Chromatography/Mass Spectrometry After Derivatization with 2,4-Dinitrofluorobenzene or Benzenesulfonyl Chloride', *J. Chromatogr. A*, **764**, 85–93 (1997).
41. T. Tsukioka, H. Ozawa, T. Murakami, 'Gas Chromatographic – Mass Spectrometric Determination of Lower Aliphatic Tertiary Amines in Environmental Samples', *J. Chromatogr.*, **642**, 395–400 (1993).
42. W.S. Gardner, P.A. St. John, 'High-performance Liquid Chromatographic Method to Determine Ammonium Ion and Primary Amines in Seawaters', *Anal. Chem.*, **63**, 537–540 (1991).
43. J. Lehotay, V. Rattay, E. Brandsteterova, D. Oktavec, 'Determination of Methyl- and Dimethylamine in Waste Water by HPLC', *J. Liq. Chromatogr.*, **15**, 307–317 (1992).
44. X.-H. Yang, C. Lee, M.I. Scranton, 'Determination of Nanomolar Concentrations of Individual Dissolved Low Molecular Weight Amines and Organic Acids in Seawater', *Anal. Chem.*, **65**, 572–576 (1993).
45. T. Okumura, K. Imamura, Y. Nishikawa, 'Determination of Anilines in River Water, Sediment, and Fish Samples by Gas Chromatography/Mass Spectrometry', *J. Chromatogr. Sci.*, **34**, 190–198 (1996).
46. M. Longo, A. Cavallaro, 'Determination of Aromatic Amines at Trace Levels by Derivatization with Heptafluorobutyric Anhydride and Gas Chromatography/Electron-capture Negative-ion Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **753**, 91–100 (1996).
47. L. Muller, E. Fattore, E. Benfenati, 'Determination of Aromatic Amines by Solid-phase Microextraction and Gas Chromatography/Mass Spectrometry in Water Sample', *J. Chromatogr. A*, **791**, 221–230 (1997).
48. E. Fattore, L. Muller, E. Davoli, D. Castelli, E. Benfenati, 'Industrial Pollutants in Ground Waters from Northern Milan', *Chemosphere*, **36**(9), 2007–2017 (1998).
49. M. Dalene, G. Skarping, 'Trace Analysis of Amines and Isocyanates Using Glass Capillary Gas Chromatography and Selective Detector. IV. Determination of Free Aromatic Amines Using Nitrogen-selective Detection', *J. Chromatogr.*, **331**, 321–330 (1985).
50. H. Kataoka, G. Maruo, K. Kijima, S. Yamamoto, S. Narimatsu, 'Determination of Aromatic Amines in Combustion Smoke Samples', in preparation.
51. K. Kijima, H. Kataoka, M. Makita, 'Determination of Aromatic Amines as their *N*-Dimethylthiophosphoryl Derivatives by Gas Chromatography with Flame Photometric Detection', *J. Chromatogr. A*, **738**, 83–90 (1996).
52. H. Kataoka, S. Shindoh, M. Makita, 'Selective Determination of Volatile *N*-Nitrosamines by Derivatization with Diethyl Chlorothiophosphate and Gas Chromatography with Flame Photometric Detection', *J. Chromatogr. A*, **723**, 93–99 (1996).
53. G. Pieraccini, F. Luceri, G. Moneti, 'New Gas Chromatographic/Mass Spectrometric Method for the Quantitative Analysis of Primary Aromatic Amines in Main- and Side-stream Cigarette Smoke', *Rapid Commun. Mass Spectrom.*, **6**, 406–409 (1992).
54. G. Grimmer, K.W. Noujack, G. Dettbarn, 'Gas Chromatographic Determination of Polycyclic Aromatic Hydrocarbons, Aza-arenes, Aromatic Amines in the Particle and Vapor Phase of Mainstream and Sidestream Smoke of Cigarettes', *Toxicol. Lett.*, **35**, 117–124 (1987).
55. A.I. Garcia-Valcarcel, C. Sanchez-Brunete, L. Martinez, J.L. Tadeo, 'Determination of Dinitroaniline Herbicides in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **719**, 113–119 (1996).
56. R.I.P. Martin, J.M. Franco, P. Molist, J.M. Gallardo, 'Gas Chromatographic Method for the Determination of Volatile Amines in Seafoods', *Int. J. Food Sci. Technol.*, **22**, 509–514 (1987).
57. T. Hamano, Y. Mitsuhashi, Y. Matsuki, 'Glass Capillary Gas Chromatography of Secondary Amines in Foods with Flame Photometric Detection After Derivatization with Benzenesulfonyl Chloride', *Agric. Biol. Chem.*, **45**, 2237–2243 (1981).
58. H. Kataoka, S. Shindoh, M. Makita, 'Determination of Secondary Amines in Various Foods by Gas Chromatography with Flame Photometric Detection', *J. Chromatogr.*, **695**, 142–148 (1995).
59. B. Pfundstein, A.R. Tricker, R. Preussmann, 'Determination of Primary and Secondary Amines in Foodstuffs Using Gas Chromatography and Chemiluminescence Detection with a Modified Thermal Energy Analyser', *J. Chromatogr.*, **539**, 141–148 (1991).
60. B. Pfundstein, A.R. Tricker, E. Theobald, B. Spiegelhalder, R. Preussmann, 'Mean Daily Intake of Primary and Secondary Amines from Foods and Beverages in West Germany in 1989–1990', *Food Chem. Toxicol.*, **29**, 733–739 (1991).
61. R.A. Scanlan, J.F. Barbour, C.I. Chappel, 'A Survey of *N*-Nitrosodimethylamine in US and Canadian Beers', *J. Agric. Food Chem.*, **38**, 442–443 (1990).
62. B.J. Finlayson-Pitts, J.N. Pitts, *Atmospheric Chemistry*, Wiley-Interscience, New York, 1986.
63. S. Lowis, M.A. Eastwood, W.G. Brydon, 'The Measurement of Methylamines in Biological Materials Using Gas Chromatographic Headspace Techniques', *J. Chromatogr.*, **278**, 139–143 (1983).
64. G. Skarping, M. Dalene, T. Brorson, J.F. Sandstrom, C. Sango, A. Tiljander, 'Chromatographic Determination of Amines in Biological Fluids with Special Reference to the Biological Monitoring of Isocyanates and Amines. I. Determination of 1,6-Hexamethylenediamine Using Glass Capillary Gas Chromatography and

Thermionic Specific Detection', *J. Chromatogr.*, **479**, 125–133 (1989).
65. H. Kataoka, S. Ohrui, Y. Miyamoto, M. Makita, 'Determination of Low Molecular Weight Aliphatic Primary Amines in Urine as Their Benzenesulphonyl Derivatives by Gas Chromatography with Flame Photometric Detection', *Biomed. Chromatogr.*, **6**, 251–254 (1992).
66. A.Q. Zhang, S.C. Mitchell, R. Ayesh, R.L. Smith, 'Determination of Trimethylamine and Related Aliphatic Amines in Human Urine by Headspace Gas Chromatography', *J. Chromatogr.*, **584**, 141–145 (1992).
67. L. Belin, U. Wass, G. Audunsson, L. Mathiasson, 'Amines: Possible Causative Agents in the Development of Bronchial Hyperreactivity in Workers Manufacturing Polyurethanes from Isocyanates', *Br. J. Ind. Med.*, **40**, 251–257 (1983).
68. J.S. Wishnok, 'Environmental Carcinogens: In Vivo Monitoring Using GC/MS', *Anal. Chem.*, **64**, 1126A–1135A (1992).
69. J.M. Sontag (ed.), *Carcinogens in Industry and the Environment*, Marcel Dekker, New York, 1981.
70. F.S.H. Abram, I.R. Sims, 'The Toxicity of Aniline to Rainbow Trout', *Water Res.*, **16**, 1309–1312 (1982).
71. G. Birner, H.G. Neumann, 'Biomonitoring of Aromatic Amines II. Hemoglobin Binding of Some Monocyclic Aromatic Amines', *Arch. Toxicol.*, **62**, 110–115 (1988).
72. E. Ward, A. Carpenter, S. Markowitz, D. Roberts, W. Halperin, 'Excess Number of Bladder Cancers in Workers Exposed to *ortho*-Toluidine and Aniline', *J. Natl. Cancer Inst.*, **83**, 501–506 (1991).
73. P.L. Skipper, X. Peng, C.K. Soohoo, S.R. Tannenbaum, 'Protein Adducts as Biomarkers of Human Carcinogen Exposure', *Drug Metab. Rev.*, **26**, 111–124 (1994).
74. M. Riffelmann, G. Muller, W. Schmieding, W. Popp, K. Norpoth, 'Biomonitoring of Urinary Aromatic Amines and Arylamine Hemoglobin Adducts in Exposed Workers and Nonexposed Control Persons', *Int. Arch. Occup. Environ. Health*, **68**, 36–43 (1995).
75. M. Dalene, G. Skarping, P. Brunmark, 'Assessment of Occupational Exposure to 4,4′-Methylenedianiline by the Analysis of Urine and Blood Samples', *Int. Arch. Occup. Environ. Health*, **67**, 67–72 (1995).
76. E.M. Ward, G. Sabbioni, D.G. DeBord, A.W. Teass, K.K. Brown, G.G. Talaska, D.R. Roberts, A.M. Ruder, R.P. Streicher, 'Monitoring of Aromatic Amine Exposures in Workers at a Chemical Plant with a Known Bladder Cancer Excess', *J. Natl. Cancer Inst.*, **88**, 1046–1052 (1996).
77. B. Stahlbom, B. Akesson, B. Jonsson, 'Experimental Human Exposure to *N*,*N*-Dimethylbenzylamine: Generation of a Controlled Atmosphere and Biological Monitoring', *Int. Arch. Occup. Environ. Health*, **70**, 393–398 (1997).
78. C. Kutzer, B. Branner, W. Zwickenpflug, E. Richter, 'Simultaneous Solid-phase Extraction and Gas Chromatographic – Mass Spectrometric Detection of Hemoglobin Adducts from Tobacco-specific Nitrosamines and Aromatic Amines', *J. Chromatogr. Sci.*, **35**, 1–6 (1997).
79. H.-G. Neumann, I.Z. Baier, C. van Dorp, 'Markers of Exposure to Aromatic Amines and Nitro-PAH', *Arch. Toxicol., Suppl.*, **20**, 179–187 (1998).
80. K.D. Brunnemann, S.S. Hecht, D. Hoffmann, '*N*-Nitrosamines: Environmental Occurrence, In Vivo Formation and Metabolism', *J. Toxicol. Clin. Toxicol.*, **19**(6–7), 661–688 (1982).
81. A.R. Tricker, R. Preussmann, 'Carcinogenic *N*-Nitrosamines in the Diet: Occurrence, Formation, Mechanisms and Carcinogenic Potential', *Mutat. Res.*, **259**, 277–289 (1991).
82. A.R. Tricker, R. Preussmann, 'Volatile *N*-Nitrosamines in Mainstream Cigarette Smoke: Occurrence and Formation', *Clin. Invest.*, **70**, 283–289 (1992).
83. K.D. Brunnemann, B. Prokopczyk, M.V. Djordjevic, D. Hoffmann, 'Formation and Analysis of Tobacco-specific *N*-Nitrosamines', *Crit. Rev. Toxicol.*, **26**(2), 121–137 (1996).
84. W. Lijinsky, 'Significance of In Vivo Formation of *N*-Nitroso Compounds', *Oncology*, **37**, 223–226 (1980).
85. B. Spiegelhalder, R. Preussmann, 'In Vivo Nitrosation of Amidopyridine in Humans: Use of 'Ethanol Effect' for Biological Monitoring of *N*-Nitrosodimethylamine in Urine', *Carcinogenesis*, **6**, 545–548 (1985).
86. W.R. Licht, M.W. Deen, 'Theoretical Model for Predicting Rates of Nitrosamine and Nitrosamide Formation in the Human Stomach', *Carcinogenesis*, **9**, 2227–2237 (1988).
87. E.W. Van Stee, R.A. Sloane, J.E. Simmons, M.P. Moorman, K.D. Brunnemann, 'Endogeneous Formation of *N*-Nitrosomorpholine in Mice from $^{15}NO_2$ by Inhalation and Morpholine by Gavage', *Carcinogenesis*, **16**, 89–92 (1995).
88. M.C. Bowman, *Handbook of Carcinogens and Hazardous Substances*, Marcel Dekker, New York, 1982.
89. K. Takatsuki, T. Kikuchi, 'Determination of *N*-Nitrosodimethylamine in Fish Products Using Gas Chromatography with Nitrogen–Phosphorous Detection', *J. Chromatogr.*, **508**, 357–362 (1990).
90. M.H. Li, C. Ji, S.J. Cheng, 'Occurrence of Nitroso Compounds in Fungi-contaminated Foods: A Review', *Nutr. Cancer*, **8**, 63–69 (1986).
91. J.H. Hotchkiss, 'Preformed *N*-Nitroso Compounds in Foods and Bevarages', *Cancer Surv.*, **8**, 295–321 (1989).
92. A.R. Tricker, B. Pfundstein, E. Theobald, R. Preussmann, 'Mean Daily Intake of Volatile *N*-Nitrosamines from Foods and Bevarages in West Germany in 1989–1990', *Food Chem. Toxicol.*, **29**, 729–732 (1991).
93. N.P. Sen, S.W. Seaman, B.D. Page, 'Rapid Semiquantitative Estimation of *N*-Nitrosodibutylamine and *N*-Nitrosodibenzylamine in Smoked Ham by Solid-phase

Microextraction Followed by Gas Chromatography/Thermal Energy Analysis', *J. Chromatogr. A*, **788**, 131–140 (1997).
94. D.H. Fine, D.P. Rounbehler, U. Goff, 'Determination of Airborne Nitrosamines by Means of Air-sampling Cartridge and GC/TEA', *IARC Sci. Publ.*, **109**, 269–274 (1993).
95. S. Fadlallah, S.F. Cooper, G. Perrault, G. Truchon, J. Lesage, '*N*-Nitroso Compounds in the Ambient Air of Metal Factories Using Metal-working Fluids', *Bull. Environ. Contam. Toxicol.*, **57**, 867–874 (1996).
96. H. Kataoka, M. Kurisu, S. Shindoh, 'Determination of Volatile *N*-Nitrosamines in Combustion Smoke Samples', *Bull. Environ. Contam. Toxicol.*, **59**, 570–576 (1997).
97. K.D. Brunnemann, W. Fink, F. Moser, 'Analysis of Volatile *N*-Nitrosamines in Mainstream and Sidestream Smoke from Cigarettes by GLC/TEA', *Oncology*, **37**, 217–222 (1980).
98. J.D. Adams, K.D. Brunnemann, D. Hoffman, 'Chemical Studies on Tobacco Smoke, LXXV. Rapid Method for the Analysis of Tobacco-specific *N*-Nitrosamines by Gas–Liquid Chromatography with a Thermal Energy Analyzer', *J. Chromatogr.*, **256**, 347–351 (1983).
99. K.D. Brunnemann, L. Genoble, D. Hoffman, '*N*-Nitrosamines in Chewing Tobacco: An International Comparison', *J. Agric. Food Chem.*, **33**, 1178–1181 (1985).
100. K.D. Brunnemann, D. Hoffmann, 'Analytical Studies on Tobacco-specific *N*-Nitrosamines in Tobacco and Tobacco Smoke', *Crit. Rev. Toxicol.*, **21**(4), 235–240 (1991).
101. R. Truhaut, N.P. Lich, M. Castegnaro, M.C. Bourgade, C. Martin, 'Volatile Nitrosamines in the Main Stream Smoke of Black Tobacco', *J. Cancer Res. Oncol.*, **108**, 157–159 (1984).
102. A.R. Tricker, C. Ditrich, R. Preussmann, '*N*-Nitroso Compounds in Cigarette Tobacco and Their Occurrence in Mainstream Tobacco Smoke', *Carcinogenesis*, **12**, 257–261 (1991).
103. R.F. Arrendale, W.J. Chamberlain, O.T. Chortyk, J.L. Baker, 'Determination of Tobacco-specific *N*-Nitrosamines by Capillary Gas Chromatography/Selected Ion Monitoring Mass Spectrometry', *Anal. Chem.*, **58**, 565–568 (1986).
104. W.I. Kimoto, C.J. Dooley, W. Fiddler, 'Nitrosamines in Tap Water After Concentration by a Carbonaceous Adsorbent', *Water Res.*, **15**, 1099–1106 (1981).
105. N.P. Sen, P.A. Baddoo, D. Weber, M. Boyle, 'A Sensitive and Specific Method for the Determination of *N*-Nitrosodimethylamine in Drinking Water and Fruit Drinks', *Int. J. Environ. Anal. Chem.*, **66**, 149–163 (1994).
106. B.A. Tomkins, W.H. Griest, C.E. Higgins, 'Determination of *N*-Nitrosodimethylamine at Part-per-trillion Levels in Drinking Waters and Contaminated Groundwaters', *Anal. Chem.*, **67**, 4387–4395 (1995).
107. B.A. Tomkins, W.H. Griest, 'Determinations of *N*-Nitrosodimethylamine at Part-per-trillion Concentrations in Contaminated Groundwaters and Drinking Waters', *Anal. Chem.*, **68**, 2533–2540 (1996).
108. J.S. Ho, T.A. Bellar, J.W. Eichelberger, W.L. Budde, 'Distinguishing Diphenylamine and *N*-Nitrosodiphenylamine in Hazadous Waste Samples by Using High-performance Liquid Chromatography/Thermospray/Mass Spectrometry', *Environ. Sci. Technol.*, **24**, 1748–1751 (1990).
109. W.S. Brewer, A.C. Draper, S.S. Wey, 'The Detection of Dimethylnitrosamine and Diethylnitrosamine in Municipal Sewage Sludge Applied to Agricultural Soil', *Environ. Pollut., Ser. B, Chem. Phys.*, **1**, 37–43 (1980).
110. J.B. Morrison, S.S. Hecht, '*N*-Nitroso-*N*-Methyldodecylamine and *N*-Nitroso-*N*-Methyltetradecylamine in Household Dishwashing Liquids', *Food Chem. Toxicol.*, **20**, 583–586 (1982).
111. J.W. Pensabene, W. Fiddler, 'Gas Chromatographic/Thermal Energy Analyzer Method for *N*-Nitrosodibenzylamine in Hams Processed in Elastic Rubber Netting', *J. Assoc. Off. Anal. Chem.*, **77**, 981–984 (1994).
112. D.C. Havery, T. Fazio, 'Estimation of Volatile *N*-Nitrosamines in Rubber Nipples for Babies' Bottles', *Food Chem. Toxicol.*, **20**, 939–944 (1982).
113. B.-G. Osterdahl, '*N*-Nitrosamines and Nitrosable Compounds in Rubber Nipples and Pacifiers', *Food Chem. Toxicol.*, **21**, 755–757 (1983).
114. H.C. Thompson, S.M. Billedeau, B.J. Miller, E.B. Hansen, J.P. Freeman, 'Determination of *N*-Nitrosamines and *N*-Nitrosamine Precursors in Rubber Nipples for Baby Pacifiers by Gas Chromatography/Thermal Energy Analysis', *J. Toxicol. Environ. Health*, **13**, 615–632 (1984).
115. S.M. Billedeau, H.C. Thompson, B.J. Miller, M.L. Wind, 'Volatile *N*-Nitrosamines in Infant Pacifiers Sold in the United States as Determined by Gas Chromatography/Thermal Energy Analysis', *J. Assoc. Off. Anal. Chem.*, **69**(1), 31–34 (1986).
116. J.I. Gray, M.A. Stachiw, 'Gas Chromatographic/Thermal Energy Analysis Method for Determination of Volatile *N*-Nitrosamines in Baby Bottle Rubber Nipples: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **70**(1), 64–68 (1987).
117. N.P. Sen, S.W. Seamen, S.C. Kushwaha, 'Improved Method for Determination of Volatile Nitrosamines in Baby Bottle Rubber Nipples and Pacifiers', *J. Assoc. Off. Anal. Chem.*, **70**, 434–438 (1987).
118. J.W. Rensaber, W. Fiddler, R.A. Gates, 'Nitrosamines Formation and Penetration in Hams Processed in Elastic Rubber Nettings: *N*-Nitrosodibutylamine and *N*-Nitrosodibenzylamine', *J. Agric. Food Chem.*, **43**, 1919–1922 (1995).
119. J.M. Fajen, G.A. Carson, D.P. Roinbehler, T.Y. Fan, R. Vita, U.E. Goff, M.H. Wolf, G.S. Edwards, D.H. Fine,

V. Reinhold, K. Biemann, '*N*-Nitrosamines in the Rubber and Tire Industry', *Science*, **205**, 1262–1264 (1979).
120. S. Fadlallah, S.F. Cooper, M. Fournier, D. Drolet, G. Perrault, 'Determination of *N*-Nitroso Compounds in Metalworking Fluids', *J. Chromatogr. Sci.*, **28**, 517–523 (1990).
121. B.A. Dawson, R.C. Lawrence, 'Analysis of Selected Drug Formulations for Volatile Nitrosamines', *J. Assoc. Off. Anal. Chem.*, **70**(3), 554–556 (1987).
122. L.B. Nielsen, S. Lings, 'Nitrosamine Formation via Nonprescription Drugs ?', *Med. Hypoth.*, **42**, 265–268 (1994).
123. R.R. Scharfe, C.C. McLenagham, 'Rapid Gas Chromatographic Method Using Nitrogen–Phosphorus Detection for *N*-Nitrosodimethylamine in 2,4-D and MCPA Herbicide Formulations', *J. Assoc. Off. Anal. Chem.*, **72**(3), 508–512 (1989).
124. R. Frassanito, E. Benfenati, G. Ciotti, R. Fanelli, 'A Simple Method for Determination of *N*-Nitrosamine Traces in Trifluralin Samples by Gas Chromatography/Mass Spectrometry', *Toxicol. Environ. Chem.*, **45**, 199–204 (1994).
125. B. Prokopczyk, J.E. Cox, D. Hoffmann, S.E. Waggoner, 'Identification of Tobacco-specific Carcinogen in the Cervical Mucus of Smokers and Nonsmokers', *J. Natl. Cancer Inst.*, **89**, 868–873 (1997).
126. J.W. Dallinga, D.M.F.A. Pachen, A.H.P. Lousberg, J.A.M. van Geel, G.M.P. Houben, R.W. Stockbrugger, J.M.S. van Maanen, J.C.S. Kleinjans, 'Volatile *N*-Nitrosamines in Gastric Juice of Patients with Various Conditions of the Gastrointestinal Tract Determined by Gas Chromatography/Mass Spectrometry and Related to Intragastric pH and Nitrate and Nitrite Levels', *Cancer Lett.*, **124**, 119–125 (1998).
127. R. Preussmann, 'Carcinogenic *N*-Nitroso Compounds and their Environmental Significance', *Naturwissenschaften*, **71**, 25–30 (1984).
128. A.R. Tricker, B. Spiegelhalder, R. Preussmann, 'Environmental Exposure to Preformed Nitroso Compounds', *Cancer Serv.*, **8**, 251–272 (1989).
129. R. Desjardins, M. Fournier, F. Denizeau, K. Krzystyniak, 'Immunosuppression by Chronic Exposure to *N*-Nitrosodimethylamine (NDMA) in Mice', *J. Toxicol. Environ. Health*, **39**, 281–288 (1992).
130. E. De Stefani, P. Boffetta, M. Mendilaharsu, J. Carzoglio, H.D. Pellergrini, 'Dietary Nitrosamines, Heterocyclic Amines, and Risk of Gastric Cancer: A Case-control Study in Uruguay', *Nutr. Cancer*, **30**, 158–162 (1998).
131. D.E. Shuker, 'Detection of Adducts Arising from Human Exposure to *N*-Nitroso Compounds', *Cancer Surv.*, **8**, 475–487 (1989).
132. S. Fadlallah, S.F. Cooper, M. Denizeau, S. Mansour, F. Guertin, K. Krzystyniak, M. Fournier, 'Assessment of *N*-Nitrosodimethylamine DNA Adducts in Rat Hepatocytes by High-performance Liquid Chromatography and Immunosorbent Assay', *Int. J. Environ. Anal. Chem.*, **56**, 165–174 (1994).
133. T. Sugimura, K. Wakabayashi, in *Mutagens and Carcinogens in the Diet*, eds. M.W. Pariza, H.-U. Aeschbacher, J.S. Felton, S. Sato, Wiley-Liss, New York, 1–18, 1990.
134. J.S. Felton, M.G. Knize, 'Occurrence, Identification, and Bacterial Mutagenicity of Heterocyclic Amines in Cooked Food', *Mutat. Res.*, **259**, 205–217 (1991).
135. M. Jagerstad, K. Skog, S. Grivas, K. Olsson, 'Formation of Heterocyclic Amines Using Model Systems', *Mutat. Res.*, **259**, 219–233 (1991).
136. K. Wakabayashi, M. Nagao, H. Esumi, T. Sugimura, 'Food-derived Mutagens and Carcinogens', *Cancer Res. (Suppl.)*, **52**, 2092s–2098s (1992).
137. G. Eisenbrand, W. Tang, 'Food-borne Heterocyclic Amines. Chemistry, Formation, Occurrence and Biological Activities. A Literature Review', *Toxicology*, **84**, 1–82 (1993).
138. K. Skog, 'Cooking Procedures and Food Mutagens: A Literature Review', *Food Chem. Toxicol.*, **31**, 655–675 (1993).
139. B. Stavric, 'Biological Significance of Trace Levels of Mutagenic Heterocyclic Aromatic Amines in Human Diet: A Critical Review', *Food Chem. Toxicol.*, **32**, 977–994 (1994).
140. T. Sugimura, M. Nagao, K. Wakabayashi, 'Carcinogenicities of Food Mutagens', *Environ. Health Perspect.*, **104**(Suppl. 3), 429–433 (1996).
141. S. Murray, N.J. Gooderham, V.F. Barner, A.R. Boobis, D.S. Davies, 'Trp-P-2 is Not Detectable in Cooked Meat and Fish', *Carcinogenesis*, **8**, 937–940 (1987).
142. S. Murray, N.J. Gooderham, A.R. Boobis, D.S. Davies, 'Measurement of MeIQx and DiMeIQx in Fried Beef by Capillary Column Gas Chromatography Electron Capture Negative Ion Chemical Ionisation Mass Spectrometry', *Carcinogenesis*, **9**, 321–325 (1988).
143. S. Murray, A.M. Lynch, M.G. Knize, N.J. Gooderham, 'Quantification of the Carcinogens 2-Amino-3,8-dimethyl- and 2-Amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline and 2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine in Food Using a Combined Assay Based on Gas Chromatography–Negative Ion Mass Spectrometry', *J. Chromatogr.*, **616**, 211–219 (1993).
144. L.M. Tikkanen, T.M. Sauri, K.J. Latva-Kala, 'Screening of Heat Processed Finished Foods for the Mutagens 2-Amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline, 2-Amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline and 2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine', *Food Chem. Toxicol.*, **31**, 717–721 (1993).
145. K. Skog, A. Solyakov, P. Arvidsson, M. Jagerstad, 'Analysis of Nonpolar Heterocyclic Amines in Cooked Foods and Meat Extracts Using Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **803**, 227–233 (1998).
146. M. Takahashi, K. Wakabayashi, M. Nagao, M. Yamamoto, T. Masui, T. Goto, N. Kinae, I. Tomita,

T. Sugimura, 'Quantification of 2-Amino-3-methylimidazo[4,5-*f*]quinoline (IQ) and 2-Amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline (MeIQx) in Beef Extracts by Liquid Chromatography with Electrochemical Detection (LCEC)', *Carcinogenesis*, **6**, 1195–1199 (1985).

147. J.S. Felton, M.G. Knize, M. Roper, E. Fultz, N.H. Shen, K.W. Turteltaub, 'Chemical Analysis, Prevention and Low-level Dosimetry of Heterocyclic Amines from Cooked Food', *Cancer Res. (Suppl.)*, **52**, 2103s–2107s (1992).
148. M.G. Knize, J.S. Felton, G.A. Gross, 'Chromatographic Methods for the Analysis of Heterocyclic Amine Food Mutagens/Carcinogens', *J. Chromatogr.*, **624**, 253–265 (1992).
149. G.A. Gross, L. Fay, in *Heterocyclic Amines in Cooked Foods: Possible Human Carcinogens*, eds. R.H. Adamson, J.-A. Gustafsson, N. Ito, M. Nagao, T. Sugimura, K. Wakabayashi, Y. Yamazoe, Princeton Scientific Publishing, Princeton, NJ, 20–49, 1995.
150. K. Wakabayashi, H. Ushiyama, M. Takahashi, H. Nukaya, S-B. Kim, M. Hirose, M. Ochiai, T. Sugimura, M. Nagao, 'Exposure to Heterocyclic Amines', *Environ. Health Perspect.*, **99**, 129–133 (1993).
151. D.W. Layton, K.T. Bogen, M.G. Knize, F.T. Hatch, V.M. Johnson, J.S. Felton, 'Cancer Risk of Heterocyclic Amines in Cooked Foods: An Analysis and Implications for Research', *Carcinogenesis*, **16**, 39–52 (1995).
152. S. Robbana-Barnat, M. Rabache, E. Rialland, J. Fradin, 'Heterocyclic Amines: Occurrence and Prevention in Cooked Food', *Environ. Health Perspect.*, **104**, 280–288 (1996).
153. N.K.J. Rich, K. Zhao, B.P. Murray, S. Bhadresa, S.J. Crosbie, A.R. Boobis, D.S. Davies, 'Heterocyclic Amines: Evaluation of Their Role in Diet Associated Human Cancer', *Br. J. Clin. Pharmacol.*, **42**, 91–98 (1996).
154. M.T. Galceran, P. Pais, L. Puignon, 'Isolation by Solid-phase Extraction and Liquid Chromatographic Determination of Mutagenic Amines in Beef Extracts', *J. Chromatogr. A*, **719**, 203–212 (1996).
155. G.A. Gross, A. Gruter, 'Quantitation of Mutagenic/Carcinogenic Heterocyclic Aromatic Amines in Food Products', *J. Chromatogr.*, **592**, 271–278 (1992).
156. S. Manabe, N. Kurihara, O. Wada, S. Izumikawa, K. Asakuno, M. Morita, 'Detection of a Carcinogen, 2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine, in Airborne Particles and Diesel-exhaust Particles', *Environ. Pollut.*, **80**, 281–286 (1993).
157. S. Manabe, S. Izumikawa, K. Asakuno, O. Wada, Y. Kanai, 'Detection of Carcinogenic Amino-α-Carbolines and Amino-γ-carbolines in Diesel-exhaust Particles', *Environ. Pollut.*, **70**, 255–265 (1991).
158. H. Kataoka, K. Kijima, G. Maruo, 'Determination of Mutagenic Heterocyclic Amines in Combustion Smoke Samples', *Bull. Environ. Contam. Toxicol.*, **60**, 60–67 (1998).
159. D. Yoshida, T. Matsumoto, 'Amino-α-carbolines as Mutagenic Agents in Cigarette Smoke Condensate', *Cancer Lett.*, **10**, 141–149 (1980).
160. T. Matsumoto, D. Yoshida, H. Tomita, 'Determination of Mutagens, Amino-α-carbolines in Grilled Foods and Cigarette Smoke Condensate', *Cancer Lett.*, **12**, 105–110 (1981).
161. M. Yamashita, K. Wakabayashi, M. Nagao, S. Sato, Z. Yamaizumi, M. Takahashi, N. Kinae, I. Tomita, T. Sugimura, 'Detection of 2-Amino-3-methylimidazo[4,5-*f*]quinoline in Cigarette Smoke Condensate', *Jpn. J. Cancer Res.*, **77**, 419–422 (1986).
162. S. Manabe, O. Wada, Y. Kanai, 'Simultaneous Determination of Amino-α-carbolines and Amino-γ-carbolines in Cigarette Smoke Condensate by High-performance Liquid Chromatography', *J. Chromatogr.*, **529**, 125–133 (1990).
163. Y. Kanai, O. Wada, S. Manabe, 'Detection of Carcinogenic Glutamic Acid Pyrolysis Products in Cigarette Smoke Condensate', *Carcinogenesis*, **11**, 1001–1003 (1990).
164. S. Manabe, K. Tohyama, O. Wada, T. Aramaki, 'Detection of Carcinogenic 2-Amino-1-methyl-6-phenylimidazo[4,5-*f*]pyridine (PhIP) in Cigarette Smoke Condensate', *Carcinogenesis*, **12**, 1945–1947 (1991).
165. S. Manabe, E. Uchino, O. Wada, 'Carcinogenic Tryptophan Pyrolysis Products in Airborne Particles and Rain Water', *Mutat. Res.*, **226**, 215–221 (1989).
166. S. Manabe, O. Wada, M. Morita, S. Izumikawa, K. Asakuno, H. Suzuki, 'Occurrence of Carcinogenic Amino-α-carbolines in Some Environmental Samples', *Environ. Pollut.*, **75**, 301–305 (1992).
167. S. Manabe, O. Wada, 'Carcinogenic Tryptophan Pyrolysis Products in Cigarette Smoke Condensate and Cigarette Smoke-polluted Indoor Air', *Environ. Pollut.*, **64**, 121–132 (1990).
168. S. Vainiotalo, K. Matveinen, A. Reunanen, 'GC/MS Determination of the Mutagenic Heterocyclic Amines MeIQx and DiMeIQx in Cooking Fumes', *Fresenius' J. Anal. Chem.*, **345**, 462–466 (1993).
169. H.P. Thiebaud, M.G. Knize, P.A. Kuzmicky, J.S. Felton, D.P. Hsieh, 'Mutagenicity and Chemical Analysis of Fumes from Cooking Meat', *J. Agric. Food Chem.*, **42**, 1502–1510 (1994).
170. H.P. Thiebaud, M.G. Knize, P.A. Kuzmicky, D.P. Hsieh, J.S. Felton, 'Airborne Mutagens Produced by Frying Beef, Pork and a Soy-based Food', *Food Chem. Toxicol.*, **33**, 821–828 (1995).
171. J. Wu, M.K. Wong, H.K. Lee, C.N. Ong, 'Capillary Zone Electrophoretic Determination of Heterocyclic Aromatic Amines in Rain', *J. Chromatogr. Sci.*, **33**, 712–716 (1995).
172. T. Ohe, 'Quantitation of Mutagenic/Carcinogenic Heterocyclic Amines, MeIQx, Trp-P-1, Trp-P-2 and PhIP, Contributing Highly to Genotoxicity of River Water', *Mutat. Res.*, **393**, 73–79 (1997).

173. S. Murray, N.J. Gooderham, A.R. Boobis, D.S. Davies, 'Detection and Measurement of MeIQx in Human Urine After Ingestion of a Cooked Meat Meal', *Carcinogenesis*, **10**, 763–765 (1989).
174. S. Manabe, O. Wada, 'Identification of Carcinogenic Tryptophane Pyrolytic Products in Human Bile by High Performance Liquid Chromatography', *Environ. Mol. Mutagen.*, **15**, 229–235 (1990).
175. H. Ushiyama, K. Wakabayashi, M. Hirose, H. Itoh, T. Sugimura, M. Nagao, 'Presence of Carcinogenic Heterocyclic Amines in Urine of Healthy Volunteers Eating Normal Diet, but Not of Inpatients Receiving Parenteral Alimentation', *Carcinogenesis*, **12**, 1417–1422 (1991).
176. M.D. Friesen, L. Garren, J.-C. Bereziat, F. Kadlubar, D. Lin, 'Gas Chromatography/Mass Spectrometry Analysis of 2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine in Urine and Feces', *Environ. Health Perspect.*, **99**, 179–181 (1993).
177. H. Ohgaki, S. Takayama, T. Sugimura, 'Carcinogenicities of Heterocyclic Amines in Cooked Food', *Mutat. Res.*, **259**, 399–410 (1991).
178. R.H. Adamson, U.P. Thorgeirsson, 'Carcinogens in Foods: Heterocyclic Amines and Cancer and Heart Disease', *Adv. Exp. Med. Biol.*, **369**, 211–220 (1995).
179. R.H. Adamson, U.P. Thorgeirsson, E.G. Snyderwine, S.S. Thorgeirrson, J. Reeves, D.W. Dalgard, S. Takayama, T. Sugimura, 'Carcinogenicity of 2-Amino-3-methylimidazo[4,5-*f*]quinoline in Nonhuman Primates: Induction of Tumors in Three Macaques', *Jpn. J. Cancer Res.*, **81**, 10–14 (1990).
180. R.H. Adamson, U.P. Thorgeirsson, T. Sugimura, 'Extrapolation of Heterocyclic Amine Carcinogenesis Data from Rodents and Nonhuman Primates to Humans', *Arch. Toxicol. (Suppl.)*, **18**, 303–318 (1996).
181. T. Mukai, Y. Yamazoe, R. Kato, 'Inhibition of Benzodiazepin and GABA Receptor Binding by Amino-γ-carbolines and Other Amino Acid Pyrolysate Mutagens', *Eur. J. Pharmacol.*, **98**, 35–44 (1984).
182. Y. Kanai, O. Wada, S. Manabe, 'Antagonism of γ-Aminobutyric acid: A Receptor-mediated Response by Amino-γ-carbolines', *J. Pharmacol. Exp. Ther.*, **252**, 1269–1276 (1990).
183. S. Manabe, Y. Kanai, S. Ishikawa, O. Wada, 'Carcinogenic Tryptophan Pyrolysis Products: Potent Inhibitors of type A Monoamine Oxidase and the Platelet Response to 5-Hydroxytryptamine', *J. Clin. Chem. Biochem.*, **26**, 265–270 (1988).
184. H. Ichinose, N. Ozaki, D. Nakahara, M. Naoi, K. Wakabayashi, T. Sugimura, T. Nagatsu, 'Effects of Heterocyclic Amines in Food on Dopamine Metabolism in Nigro-strial Dopaminergic Neurons', *Biochem. Pharmacol.*, **37**, 3289–3295 (1988).
185. M. Naoi, T. Takahashi, H. Ichinose, K. Wakabayashi, T. Sugimura, T. Nagatsu, 'Reduction of Enzyme Activity of Tyrosine Hydroxylase and Aromatic L-Amino acid Decarboxylase in Clonal Phenochromocytoma PL12h Cells by Carcinogenic Heterocyclic Amines', *Biochem. Biophys. Res. Commun.*, **157**, 494–499 (1988).
186. R. Hasegawa, H. Tanaka, S. Tamano, T. Shirai, M. Nagao, T. Sugimura, N. Ito, 'Synergistic Enhancement of Small and Large Intestinal Carcinogenesis by Combined Treatment of Rats with Five Heterocyclic Amines on a Medium-term Multi-organ Bioassay', *Carcinogenesis*, **15**, 2567–2573 (1994).
187. M. Dressler, *Selective Gas Chromatographic Detectors*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 36, 1986.
188. R. Buffington, M.K. Wilson, *Detectors for Gas Chromatography – A Practical Primer*, Hewlett-Packard, Avondale, PA, 1991.
189. R.P.W. Scott, *Chromatographic Detectors: Design, Function and Operation*, Chromatographic Science Series, Marcel Dekker, New York, Vol. 73, 1996.
190. R.L. Grob, M.A. Kaiser, *Environmental Problem Solving Using Gas and Liquid Chromatography*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 21, 1982.
191. F.I. Onuska, F.W. Karasek, *Open Tubular Column Gas Chromatography in Environmental Sciences*, Plenum Press, New York, 1984.
192. F. Bruner, *Gas Chromatographic Environmental Analysis: Principles, Techniques, Instrumentation*, VCH, New York, 1993.
193. M. Dalene, L. Mathiasson, J.A. Jonsson, 'Trace Analysis of Free Amines by Gas–Liquid Chromatography', *J. Chromatogr.*, **207**, 37–46 (1981).
194. S. Skarping, B.E.F. Smith, M. Dalene, 'Trace Analysis of Amines and Isocyanates Using Glass Capillary Gas Chromatography and Selective Detection. III. Determination of Aliphatic and Alicyclic Amines as Perfluoro Fatty Acid Amides Using Electron-capture and Nitrogen-selective Detection', *J. Chromatogr.*, **303**, 89–98 (1984).
195. M. Dalene, T. Lundh, L. Mathiasson, '*N*-Permethylation of Polyamines at Trace Levels for Gas Chromatographic Analysis', *J. Chromatogr.*, **322**, 169–176 (1985).
196. T. Lundh, B. Akesson, 'Gas Chromatographic Determination of Primary and Secondary Low-molecular-mass Aliphatic Amines in Urine Using Derivatization with Isobutyl Chloroformate', *J. Chromatogr.*, **617**, 191–196 (1993).
197. S. Skarping, L. Renman, M. Dalene, 'Trace Analysis of Amines and Isocyanates Using Glass Capillary Gas Chromatography and Selective Detection. II. Determination of Aromatic Amines as Perfluoro Fatty Acid Amides Using Nitrogen-selective Detection', *J. Chromatogr.*, **270**, 207–218 (1983).
198. H. Kataoka, K. Kijima, 'Analysis of Heterocyclic Amines as their *N*-Dimethylaminomethylene Derivatives by Gas Chromatography with Nitrogen–Phosphorus Selective Detection', *J. Chromatogr. A*, **767**, 187–194 (1997).

199. S. Ohrui, H. Kataoka, Y. Miyamoto, K. Ohtsuka, M. Makita, 'Determination of Primary Amines by Benzenesulfonylation/GC with Flame Photometric Detection', *Bunseki Kagaku*, **40**, 119–123 (1991).
200. H. Kataoka, M. Eda, M. Makita, 'Selective Determination of Secondary Amines as their *N*-Diethylthiophosphoryl Derivatives by Gas Chromatography with Flame Photometric Detection', *Biomed. Chromatogr.*, **7**, 129–133 (1993).
201. Y. Miyamoto, H. Kataoka, S. Ohrui, M. Makita, 'Determination of Primary Amines as their *N*-Benzenesulfonyl-*N*-trifluoroacetyl Derivatives by GC with Electron Capture Detection', *Bunseki Kagaku*, **43**, 1113–1118 (1994).
202. G.W. Diachenko, 'Determination of Several Industrial Aromatic Amines in Fish', *Environ. Sci. Technol.*, **13**, 329–333 (1979).
203. R.T. Coutts, E.E. Hargesheimer, F.M. Pasutto, G.B. Baker, 'Analysis of Aniline and Aminophenols in Aqueous Solutions Using Electron Capture Gas Chromatography', *J. Chromatogr. Sci.*, **19**, 151–155 (1981).
204. S. Skarping, L. Renman, B.E.F. Smith, 'Trace Analysis of Amines and Isocyanates Using Glass Capillary Gas Chromatography and Selective Detection. I. Determination of Aromatic Amines as Perfluoro Fatty Acid Amides Using Electron-capture Detection', *J. Chromatogr.*, **267**, 315–327 (1983).
205. H.-B. Lee, 'Perfluoro and Chloro Amide Derivatives of Aniline and Chloroanilines: A Comparison of their Formation and Gas Chromatographic Determination by Mass Selective and Electron-capture Detectors', *J. Chromatogr.*, **457**, 267–278 (1988).
206. M. Mohnke, B. Schmidt, R. Schmidt, J.C. Buijten, Ph. Mussche, 'Application of a Fused-silica Column to the Determination of Very Volatile Amines by Gas–Solid Chromatography', *J. Chromatogr.*, **667**, 334–339 (1994) .
207. J.J. Richard, G.A. Junk, 'Steam Distillation, Solvent Extraction, and Ion Exchange for Determining Polar Organics in Shale Process Waters', *Anal. Chem.*, **56**, 1625–1628 (1984).
208. H. Hayatsu, 'Cellulose Bearing Covalently Linked Copper Phthalocyanine Trisphosphate as an Adsorbent Selective for Polycyclic Compounds and its Use in Studies of Environmental Mutagens and Carcinogens', *J. Chromatogr.*, **597**, 37–56 (1992).
209. H. Hayatsu, T. Hayatsu, S. Arimoto, H. Sakamoto, 'A Short-column Technique for Concentrating Mutagens/Carcinogens Having Polycyclic Structures', *Anal. Biochem.*, **235**, 185–190 (1996).
210. J. Pawliszyn, *Solid-phase Microextraction: Theory and Practice*, Wiley-VCH, New York, 1997.
211. H. Kataoka, 'Methods for the Determination of Mutagenic Heterocyclic Amines and Their Applications in Environmental Analysis', *J. Chromatogr. A*, **774**, 121–142 (1997).

Heavy Metals Analysis in Seawater and Brines

Osvaldo E. Troccoli
University of Buenos Aires, Argentina

1 Introduction **1**
2 Analytical Chemistry of Heavy Metals **2**
2.1 Challenges Posed by the Ever-increasing Need to Improve Detection Limits 2
2.2 Preconcentration and Speciation 4
2.3 Interferences Most Commonly Encountered in Seawater and Brines 7
3 Spectroscopic Techniques Applied to Seawater Analysis **7**
3.1 Atomic Spectroscopies 7
3.2 Other Atomic Spectrometries 10
4 Electroanalytical Techniques **12**
4.1 Stripping Techniques 13
5 Comparison of Different Approaches **14**
Acknowledgments **15**
Abbreviations and Acronyms **15**
Related Articles **16**
References **16**

There has been growing concern during and since the 1990s about the effect that the toxicity and persistence of heavy metals will have on the biosphere. Consequently, these new environmental pressures have challenged analysts to find better experimental procedures for handling, recognition and evaluation of traces of heavy metals.

Moreover, the determination of low levels of heavy metals in seawater has received particular attention as a way of assessing the early impact of human activities on the marine environment. It is necessary to protect the marine and continental waters against contamination so as to allow suitable conditions for the maintenance of aquatic life and the various uses derived from aquatic life.

It is not the metal concentration alone that enables an environmental diagnosis to be made but the way metal combines with other elements to form species. These species are responsible for the biological behavior of the heavy metal under consideration, as its bioavailability will determine the amount of harm it may cause. Thus, ''speciation'' studies should be conducted to assess the potential environmental impact of human activities.

From the analytical point of view, seawater and brines may be considered to be closely related systems in terms of matrix effects and analytical difficulties. Hence these two systems will be included in this article although the aims of heavy metal detection and determination might be diverse in each case.

Analytical inaccuracies that arise from the high salt content of the seawater or brine solutions are exacerbated by the low levels of trace metals, which require careful control of contamination during sample collection and analysis and very sensitive analytical techniques.

Different techniques for the determination of trace metals in seawater, such as spectroscopic and electroanalytical methods, will be discussed here in terms of their performance in dealing with the problem of trace metal detection and determination.

Even the term ''trace'' has had different meanings over the years, normally being associated with the detection limit of the technique employed (''traces'', ''footprints'' or ''trails'' have a lot to do with the way they are observed). Amounts that would have been considered traces in the 1960s are probably no longer considered as such in the 1990s, as detection limits for most heavy metals have been lowered by three to six orders of magnitude (using ''state of the art'' techniques) since the late 1960s.

Special attention has been paid to the preconcentration of the analyte and its isolation from the matrix constituents which might obscure its accurate determination. A variety of separation–preconcentration methods will be presented, with a strong bias towards those procedures leading to some kind of automation, because pretreatment of the samples is often time-consuming and prone to contamination and may lead to an incomplete recovery of the analytes.

Flow injection analysis (FIA) has appeared as a solution to time-consuming and complicated separations and it will also be discussed in relation to alternative forms of analysis, especially in terms of its ability to perform speciation studies.

1 INTRODUCTION

Interest in the determination of heavy metals at trace levels in seawater derives from the importance that has been placed on the study of the fate of accumulative toxins in nature.

Marine chemistry, marine biology, geochemistry, and environmental science researchers are interested in the horizontal, vertical, and seasonal distribution of trace metals in seawater, as well as in trace metal biochemical and geochemical cycles in the ocean. Moreover, the determination of low levels of heavy metals in seawater has received particular attention as a way of assessing the early impact of human activities on

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the marine environment, as a necessary preliminary to protecting marine and continental waters against contamination so as to allow suitable conditions for the maintenance of aquatic life and the various uses derived from it.

Most coastal nations have passed regulations in order to assess and control the contamination of their marine waters. To illustrate the importance that has been given to this issue, European Community Directives on the quality of continental and marine waters should be consulted: for marine ecosystems European Directive 79/923/CEE[1] limits the levels of Ag, As, Cd, Cr, Cu, Hg, Ni, Pb and Zn in seawater used to keep molluscs.

Knowledge of the metal concentration, together with the way it is combined with other elements to form species, enables an environmental diagnosis to be made. These species are responsible for the biological behavior of the heavy metal under consideration, as its bioavailability will determine the amount of harm it may cause. Thus, ''speciation'' studies should be conducted to assess the potential environmental impact of human activities.

From the analytical point of view, seawater and brines may be considered to be closely related systems in terms of matrix effects and analytical difficulties. Hence these two systems will be included in this article although the aims of heavy metal detection and determination might be diverse in each case.

Analytical inaccuracies that arise from the high salt content of the seawater or brine solutions are exacerbated by the low levels of trace metals, which require careful control of contamination during sample collection and analysis and very sensitive analytical techniques.

Different techniques for the determination of trace metals in seawater, such as spectroscopic and electroanalytical methods will be discussed here in terms of their performance in dealing with the problem of trace metal detection and determination.

High salinity and complex chemical composition of seawater and brines, together with a very low concentration level for the heavy metals, are the main obstacles to their proper analysis. Special attention has been paid to the preconcentration of the analyte and its isolation from the matrix constituents which might obscure its accurate determination. A variety of separation–preconcentration methods will be presented, with a strong bias towards those procedures leading to some kind of automation, because pretreatment of the samples is often time-consuming and prone to contamination and may lead to an incomplete recovery of the analytes.

FIA has appeared as a solution to time-consuming and complicated separations[2] and it will also be discussed in relation to alternative forms of analysis, especially in terms of its ability to perform speciation studies.

2 ANALYTICAL CHEMISTRY OF HEAVY METALS

2.1 Challenges Posed by the Ever-increasing Need to Improve Detection Limits

The identification and quantitation of heavy metals at trace levels has been widely studied and several books are available for consultation.[3–5]

Concern about the effects of the toxicity and persistence of heavy metals on the biosphere is growing. These new environmental concerns have challenged analysts to find better experimental procedures for handling, recognition and evaluation of traces of heavy metals.

What is meant by ''better experimental procedures''? This is a difficult question to answer, because it depends on the purpose of the trace determination, which might be classified by the following:

- precise concentration determination
- concentration profiles assessment
- bioavailability determination.

Precise concentration determination should be understood as a means of selecting those operations that pursue the highest accuracy and reproducibility for a single set of measurements. This case is the most frequently used in the development of new techniques, where interest is focused on finding novel ways of improving the features of merit of the technique.

Concentration profiles assessment here refers to operations that obtain information through analysis of a large number of sets of samples taken at different locations at different times. In this case, a new feature of merit, analysis frequency, which will reflect the quantity of reliable data obtained per unit time should be added.

Finally, bioavailability determination is one of the most recent requirements of the analytical chemist involving knowledge of the chemical form in which the analyte is present in the sample.

Here, the chemical ''species'' is the object of the analysis, and the features of merit will be referred to each species involved, because the toxicity of the heavy metals is related to the way that they could be introduced into living organisms (i.e. as the ''chemical species'').

Each of these alternatives may present the analyst with the need to use a different approach, from the evaluation of a large number of samples in a short period of time, to the accurate measurement of a single set of samples or the establishment of the chemical form in which the trace metal is encountered. Even the term ''trace'' has had different meanings over the years, normally being associated with the detection limit of the technique employed (''traces'', ''footprints'' or ''trails'' have a lot to do with the way they are observed). Amounts that would

have been considered traces in the 1960s are probably no longer considered as such in the 1990s, as detection limits for most heavy metals have been lowered by three to six orders of magnitude (using ''state of the art'' techniques) since the late 1960s.

In the early 1960s, the rise in the use of atomic absorption spectrometry (AAS) and the decline of atomic emission spectrometry (AES) produced a variety of studies using flames as ''atom cells'' to reveal the potential of this novel technique when compared with prevailing trace analysis techniques.

The average limit of detection (LOD) by that time was in the order of the parts per million for the refractory elements and around the parts per billion range for the easily atomized metals (Zn, Cd). It should be emphasized that the term ''LOD'' used then corresponded strictly to an instrumental LOD, while this concept was later changed to the expression ''methodological limit of detection'' (MLD). The MLD reflects more closely the environment of ''real'' samples than does the LOD, which reflects the instrument performance, usually expressed against a solvent blank and tending to enhance the value of the instrumental technique. (According to IUPAC (International Union of Pure and Applied Chemistry) definitions, LOD is the concentration that produces a signal equal to three times the standard deviation of the blank signal, after a number of determinations of that blank. This LOD is essentially instrumental and should not be confused with the MLD which is defined as the minimum concentration that can be reported as different from zero at a 99% confidence level. The sample containing the analyte whose MLD is being assesed should be compared to a blank that should be clearly defined and that has undergone the same analytical procedure as the sample.)

In the field of the electroanalytical techniques, stripping analysis over Hg (mainly anodic) appeared to be a promising technique for the easily reducible metals, even considering that anodic stripping voltammetry (ASV) is prone to interference when high ionic strength solutions (the case of seawater and brines) are used.

Later, spectroscopists developed a new atom cell, the electrothermal atomizer (ETA), which improved detection limits several times when background interferences were handled properly, while new electroanalytical procedures derived from stripping analysis become more popular as most of the inconveniences of the former ASV diminished or were eliminated.

The use of inductively coupled plasmas (ICPs) as atom cells introduced a new dimension to metal trace analysis, as detection limits were lowered by two or three orders according to the detection mode selected: inductively coupled plasma/optical emission spectrometry (ICP/OES) or inductively coupled plasma mass spectrometry (ICPMS).

The evolution of the detection limits for some heavy metals may be summarized as follows.

2.1.1 Spectroscopic Methods

In most cases, improvements in detection limits can be associated with the development of new instruments that enhance the signal-to-noise ratio via digital handling of the output signal or by coupling the atomic cell to newly developed detectors like mass spectrometers, or through the development of new on-line analytical procedures that improve the MLD. This becomes evident when observing the evolution of the detection limit for the same metal and technique over the years. FAES (flame atomic emission spectrometry) limits were enhanced when ICP excitation was available, although the excited species might not be the same. Flame atomic absorption spectrometry (FAAS) limits remained almost constant or, in other words, they did not follow the evolution of the electronics and data handling procedures as it would have been reasonable to predict, the explanation being that AAS signals are relative and do not depend so strongly on sensitivity and noise reduction procedures.

The real improvement in AAS MLDs was obtained when the flame atomic cell, continuously fed by spray chamber nebulizers, was replaced by discrete atomizers (mainly ETA) where the atomic cloud is confined giving rise to transient signals which are easily amplified and integrated.

Several important factors should be mentioned when discussing the enhanced sensitivity obtained over the years for electrothermal atomic absorption spectrometry (ETAAS):

1. The use of chemical modifiers: the introduction of a chemical reagent (called the modifier) in excess with respect to the solution to be measured by graphite furnace atomic absorption spectrometry (GFAAS), has proved efficient in converting the analyte element into a phase of higher volatility and/or in increasing the volatility of concomitants.[6,7]
2. Atomizer development has taken place through the introduction of new atomizing platforms as already mentioned,[13–15] new atomizer materials like W or Mo,[8,9] and new ways of heating as transverse heating.[10–12]
3. New background correction systems[13–15] have been used.

Finally some of the improvements found in ICP/OES could be ascribed to new sample introduction systems, together with enhanced data handling capabilities.

2.1.2 Electroanalytical Methods

Although acquisition of data for these methods is hindered by the different nomenclatures, it could be said that the enhancement in LODs produced over the years is a consequence of improved data acquisition and data treatment devices. However, it is important to point out that stripping techniques allow better MLDs because the electrolysis time can be managed.

2.2 Preconcentration and Speciation

2.2.1 Preconcentration

The nature of the matrix in which the trace metal is embedded might, however, also condition the analytical performance of the procedure developed. As mentioned earlier, what makes heavy metals analysis in seawater or brines so difficult is the low level in which heavy metals are usually found in these media and the highly saline matrix that contributes to a worsening of the detection performance.

Matrix elimination, sample pretreatment, or simply sample introduction, have been problematic issues in most heavy metals determination techniques, where accurate analyses are sought. No matter which instrumental device is used for the detection, saline samples will generally need special treatment before the analyte reaches the detection stage.

In any case, even if the lowest detection limits were achieved, some of the heavy metals present in seawater or brines cannot be quantified because their concentration levels in those matrices are below the LOD attainable, and some way of increasing their concentrations needs to be used in order to obtain a signal that is at least ten times greater than the standard deviation of ten replicates of the blank.

The set of operations performed with the aim of increasing the analyte concentration is usually referred to as ''preconcentration'', and it has become a discipline of interest among the analytical procedures for trace analysis. Again, available ''batch'' preconcentration methods have evolved from macro- or semimicroscale solvent extraction or ionic exchange to microscale ''on-line'' operations.

Preconcentration steps could be classified according to the phase equilibria involved:

1. solvent extraction
2. ion exchange
3. adsorption methods.

These procedures, or a combination of them, have been applied to seawater and brines. Some of the published studies are summarized in Table 1.

Usually, highly saline matrices require a conditioning step to eliminate or diminish the concentration of the salts that cause interference. This conditioning step might also act as a preconcentration step, and the detection device will dictate the extent of conditioning needed, as shown later.

As can be observed in Table 1, preconcentration is often performed using ion-exchange resins, including the so-called ''chelating'' resins in which the functional group complexes the analyte. These have been included in this group as the final process will be ionic exchange in both resin types. Chelating resins will behave more selectively than plain ionic exchange resins, as a covalent complex is formed between the metal atom and the donor atoms of the ligand (i.e. O, N, S); furthermore, these complexes will be less prone to the influence of ionic strength which seriously affects electrostatic interactions like those observed in ordinary ionic exchange resins. This allows the use of deionized water prepared standards even when measuring seawater samples.[46] Specially prepared fillings combining an inert support like silica with an organic ligand, like oxine derivatives, have also been employed, their selectivity and rigidity (they do not swell like the 100% organic polymers) being their main advantage. However, commercial chelating resins like Chelex 100 might be used with minor mechanical adjustments to compensate for the characteristic swelling. More recently strongly cross-linked resins like Muromac have been employed with good results in terms of selectivity and diminished swelling characteristics, but on-line operation with these resins might be severely impaired due to the low diffusion rate of the metal ions into the resin which in terms affects the chelating velocity.[47]

The choice between batch and on-line operation will depend on several practical considerations, with cost being one of the major issues to be taken into account. Most operations can be performed off-line and the samples can then be processed rapidly, making this approach competitive with on-line processing in terms of time employed, although the probability of contamination or analyte losses is still a serious drawback.

Commercial instruments can be fitted with automatic sampling devices capable of handling hundreds of vials in which samples, standards and blanks may be placed. Later, associated software will perform the necessary operations to calibrate and measure the analyte concentration in each sample.

This automation is particularly useful when multielemental detection systems (typically ICP/OES or ICPMS) that present such advantage over single-element on-line systems (typically HG/AAS, or GFAAS) are used.

Since it was first introduced by Ruzicka and Hansen in 1975[48] the FIA, which involves the introduction

Table 1 Preconcentration procedures

Element	Characteristics	Ref.
Solvent extraction		
Bioactive	Use of solid-phase extraction and high-resolution ICPMS	16
Cd	Preconcentration by solvent extraction	17
Fe, Cu	Use of 2-hydroxy-1-naphthaldehyde-4-phenyl-3-thiosemicarbazone extraction	18
Sn (tributyl)	Use of solvent extraction and HG/ETAAS	19
Yb	Coprecipitation and solvent extraction	20
Zn	Solvent extraction using a new chelating agent	21
Ion exchange		
Cd	Use of a dynamically coated column of quaternary ammonium salt on C_{18}-bonded silica gel	22
Cd, Cu, Pb	Chelating resin preconcentration and Hg(II)-assisted elution	23
Cd, Pb	Fully automatic on-line separation preconcentration system, using a Chelex 100 Microcolumn	24
Cu, Mo	Use of Muromac A-1 chelating resin	25
Cu, Ni, Cd	Synthesis, properties and applications of silica-immobilized 8-quinolinol	26
Pb	Use of a chelating resin	27
Pb	Use of Chelex-100 for isotope ratio determination	28
Rare earths	Use of poly (acrylaminophosphonic dithiocarbamate) chelating fiber	29
Several	Use of a cellulose–zinc hydroxide system	30
Several	Chelating resin preconcentration	31
Trace elements	Use of immobilized quinolin-8-ol silica gel	32
Trace elements	Ion chromatographic separation system: use of nitrilotriacetate chelating resin	33
Zn	Use of a dynamically coated column of methyltricaprylylammonium chloride on octadecylsilyl silica	34
Sorption/adsorption		
Ag	Coprecipitation mercury(II) sulfide	35
Cr(III)	Adsorption of silica	36
Cu, Mn, Ni	FI on-line sorption in a knotted reactor	37
Fe(III)	Sorption of the 3-(2-pyridyl)-5,6-bis (4-phenylsulfonic acid)-1,2,4-triazine complex with poly(chlorotrifluoroethylene) resin	38
Hg, Hg (methyl)	Sorbent extraction	39
Mo, V, Zr, Cr	Sorption preconcentration	40
Pb, Cu	On-line Co-APDC coprecipitation	41
Rare earths	Adsorption on activated carbon	42
Trace elements	Formation of water-soluble complexes and their adsorption on C_{18}-bonded silica gel	43
Trace elements	Use of cobalt diethyldithiocarbamate coprecipitation	44
Trace elements	Sorption ICP/OES	45

FI, flow injection; HG, hydride generation; APDC, ammonium pyrrolidine dithiocarbamate.

of a known volume of a liquid sample into a flowing unsegmented liquid carrier, where it is transported to a detector for measurement, has gone a long way in analytical chemistry towards becoming one of the ''established'' techniques of sample analysis and on-line sample treatment. FIA has gained acceptance as a powerful tool for rapid, accurate and economic performance of repetitive determinations. In this sense, and as will be shown later, most analytical procedures needed in the determination of heavy metals in seawater and brine require preconditioning and preconcentration steps which can be achieved using FIA.

FI operations may be performed with the aid of a very simple and economic device, the FI manifold, which consists typically of a propulsion unit (a peristaltic pump), a sample introduction device (a rotary injection valve), a reaction unit (a piece of tubing and/or complementary device), and a detector.

Continuous-flow processes like FI enjoy a flexibility that makes them suitable for performing most of the tasks associated with automation in the laboratory and obtaining the benefits of automation: an increase in the reliability (precision) of analytical operations; a reduction in the analysis time dedicated to routine operations; and a more efficient utilization of chemicals.

One of the major goals of FIA has been redefinition of the operational analytical scale on a micro basis, where volumes in the range of microliters are dealt with and consequently there are important savings in chemicals and waste is reduced to a minimum. Moreover, the miniaturization of FI systems has led to a sustantial reduction in the ''idle'' time between consecutive sample

plugs, speeding up on-line systems to make them compatible with detection devices that are very expensive when run idle (typically ICPMS or ICP/OES).

2.2.2 *Speciation*

Metal speciation could be broadly defined as the study of the ''distribution of the metals over distinct molecular binding forms'',[49] regardless of the effects on the biota that these molecular bindings might have, or it can be focused specifically on the impact on living organisms of these molecular forms. In either case there is a need to isolate the signals produced by those forms in order to establish the composition of the system. This isolation may be performed in situ on-line and continuously or before the detection step.

In natural systems the ''molecular form'' involved must be defined or characterized because natural molecules are too complex to be defined exactly. It is generally accepted that molecular size together with compound solubility are the most important parameters to be taken into account where the bioavailability of metals is concerned, and an operational or behavioral classification is often preferred to the plain chemical classification.

Del Castilho[49] clearly illustrates the operational definition through its application to dissolved and particulated matter in seawater, referring to the passage of the sample through a 0.45-μm pore filter to distinguish between dissolved (the fraction that goes through the filter) and particulate bound matter (the fraction deposited on the filter), leaving colloidal complexes in the soluble fraction.

Easily bioavailable metals will probably be found in the soluble fraction, while the insoluble fraction contains particles that will encounter a certain resistance to passing through cellular membranes. Furthermore, even the soluble fraction may need metal to dissociate from its binding particles before going through a successful diffusion process in order to enter the cell. Thus, molecular size determination will be needed on the insoluble fraction and some kind of dissociation behavior should be determined on the soluble fraction.

Bearing in mind both needs, speciation studies should lead to the determination of either the size of the substituents and/or the dissociation rate of the compounds likely to exist, together with characterization of the diffusion pattern of the heavy metals into the cellular membrane.

Several approaches have been used to achieve this purpose. These include previous separation of different molecular forms with the aid of on-line separative devices, like high-performance liquid chromatography (HPLC) and gas–liquid chromatography (GLC) in all their varieties, and operational characterization by means of differential pulse anodic stripping voltammetry (DPASV) and the time of contact concept. Redox speciation might be considered to be a special case of the soluble fraction speciation, provided the species belong to the soluble fraction. However, detailed knowledge of the distribution of oxidation states for a given metal may provide a useful probe of the redox conditions prevailing in the environment.

Some speciation studies performed on seawater are summarized in Table 2.

It must be emphasized that when ''total metal'' detectors are used (mainly spectroscopic) a previous separation stage is mandatory to achieve some kind of speciation. On the other hand, electroanalytical techniques, under special conditions, might perform speciation without previous separation, as most reduction

Table 2 Speciation studies

Element	Species	Separation/detection	Ref.
Cr	Cr(III), Cr(VI)	CCSV	50
Cr	Cr(III), Cr (total)	Seagoing method with detection by electron-capture gas chromatography	51
Cr	Cr(III)	Preconcentration by adsorption of silica, voltammetric detection	36
Cr	Cr(VI), Cr (total)	FI on-line separation and preconcentration with ETAAS detection	52
Cr	Cr(III), Cr (total)	Preconcentration in a flow-through electrochemical sorption cell and FAAS determination	53
Cr	Cr(III), Cr(VI)	ETAAS	54
Fe	Fe(II), Fe(III)	Selective on-line preconcentration and spectrophotometric determination	55
Fe	Fe(II), Fe(III)	CCSV	56
Fe	Organoferric	CSV	57
Mo	Mo(VI)	Ion pair formation and solvent extraction	58
Se	Se(IV), Se(VI)	Direct determination by continuous flow HG and AFS	59
Sn	Organotin	Microcolumn field sampling	60
U	U(VI)	Fluorometric determination using separation on an activated silica gel column	61
U	U(VI)	Direct determination by CSV	62

CCSV, catalytic cathodic stripping voltammetry; CSV, cathodic stripping voltammetry; AFS, atomic fluorescence spectrometry.

potentials are modified according to the environmental conditions of the ionic species.

2.3 Interferences Most Commonly Encountered in Seawater and Brines

Tables 3 and 4 show the average composition for seawater with a salinity of 35 (inorganic constituents) and the most common organic compounds found for this seawater.

Taking into account the average seawater composition, one may predict that the most common interferences will be the chlorides and sulfates of the alkali and alkaline earth metals, and although it would not be evident at a glance, the organics that might form complexes with the heavy metals. However, these components will not always affect the measurements to the same extent, as interferences are closely related to the detection device employed.

Some facts about the chemistry of highly saline solutions must be remembered, particularly the tendency to form aggregates or compounds with the alkaline chlorides, or even the presence of the alkaline chloride, which may hinder the behavior of the trace metal, or physicochemical phenomena associated with high ionic strength solutions. Most natural or artificial brines will fall into the same category in terms of interference. Organic matter will affect the available heavy metal concentration, giving rise to the speciation studies. The interferences associated with each detection technique will be discussed later.

Sources of interference other than the matrix itself are the reactives that have to be used in the conditioning or preconcentration steps. This is particularly true for preconcentration systems in which the preconcentration factor is high and the reactives were used prior to this stage and are difficult to purify (i.e. conditioning with ethylenediamine tetraacetic acid (EDTA) or sodium salts). It is often necessary to pass the reagents through purification steps to ensure an appropriate level of impurities in the sample under consideration.

Table 3 Major composition of seawater[63]

Constituent	Seawater at $S = 35$
Na^+	10.77
Mg^{2+}	1.29
Ca^{2+}	0.4121
K^+	0.399
Sr^{2+}	0.0079
Cl^-	19.354
SO_4^{2-}	2.712
HCO_3^-	0.1424
Br^-	0.0673
F^-	0.0013
B	0.0045

S, salinity, is the weight in grams of the dissolved inorganic matter in 1 kg of seawater after all Br^- and I^- ions have been replaced by the equivalent quantity of Cl^-, and all HCO_3^- and CO_3^{2-} converted to oxide.

Table 4 Average concentrations of organics in seawater[63]

Components	Concentration in seawater ($\mu g\,C\,L^{-1}$)
Free amino acids	20
Combined amino acids	50 (to 100)
Free sugars	20
Fatty acids	10
Phenols	2
Sterols	0.2
Vitamins	0.006
Ketones	10
Aldehydes	5
Hydrocarbons	5
Urea	20
Uronic acids	18
Approximate total	340

3 SPECTROSCOPIC TECHNIQUES APPLIED TO SEAWATER ANALYSIS

3.1 Atomic Spectroscopies

Atomic spectra are unique in that, given the appropriate conditioning and instrumentation, the signals can be assigned univocally to a single element, making the spectroscopic techniques highly specific. Interelement interferences are seldom encountered provided the operation is conducted properly. This, together with the economy of operation, is responsible for the vast usage and popularity of atomic spectroscopies in their two main modes of operation, emission and absorption, and also, but to a lesser extent, fluorescence.

The use of flames as the traditional atomic source has receded with the introduction of novel ways of producing atomic vapors such as ETA, cold vapor generators, laser excitation probes or atmospheric pressure plasmas (ICP or direct current (DC) plasma) and so on, but although these new atom sources have many advantages, there are still thousands of flame-based spectrometers in operation.

3.1.1 Atomic Absorption Spectrometry

During the 1990s, ETA and to a lesser extent cold vapor generators, hydride generators and flames (in that order) have been used as atomic sources in atomic absorption studies on seawater and brines.

Table 5 FAAS

Element	Characteristics	Ref.
Cd, Co, Cu, Mn, Ni, Pb, Zn	On-line preconcentration and solvent extraction by FI system	64
Cr, Cu, Mn	Preconcentration on pretreated solid materials	65
Cu, Mo	Use of Muromac A-1 chelating resin with on-line preconcentration	25
Fe, Cu	2-hydroxy-1-naphthaldehyde-4-phenyl-3-thio semicarbazone extraction	18
Ni(II)	Preconcentration by flotation	66
Sn (organic)	Use of capillary column gas chromatography–AAS	67
Zn	Solvent extraction using a new chelating agent	21
Several	Preconcentration with modified silicas	68
Several	Karachi coastal waters	69
Several	Cobalt–diethyldithiocarbamate coprecipitation	44

Some of the latest studies using flame atomization are summarized in Table 5.

A quick look at Table 5 indicates that preconcentration procedures together with conditioning of the samples have been the objective of most FAAS studies related to heavy metal determination in seawater. No direct determinations are mentioned as a result of the poor LODs obtained.

The spectral interference caused by sodium chloride absorption bands is one of the most important drawbacks of ETAAS applied to seawater analysis. Background correction has proved useful but it still is not a solution to the problem for all the elements. Special devices such as Zeeman effect or Smith–Hiftje background correctors are needed when a deuterium lamp background corrector does not suffice, owing to the structured nature of the sodium chloride absorption spectra.[70–72] Moreover, vaporization of the analyte is also impaired by the massive amounts of salts present in the sample.

Table 6 shows similar compendia for ETA, mainly GF and related techniques.

Again, it can be seen that preconditioning of the sample has a major role in most of the studies presented, although several papers[9,74,75,81,90,92,94,99,103–105] are devoted to direct determination of the heavy metals, indicating that for some of them the actual LOD are good enough to render direct determination possible or that the metal concentration in those seawater samples is high enough to reach the existing LOD.

Several papers show the interest in coupling different automatic or semi-automatic sampling devices in order to speed up analysis and to enhance its reproducibility,[24,41,76,81,87,96,99,105,106,108,110] but the discrete nature of the atomization step has made achievement of this quite difficult.

Metals forming volatile compounds, such as hydrides, have been successfully analyzed by AAS or ETAAS. ETAAS is the most sensitive but, at the same time, the most difficult to couple with the HG step due to the discrete nature of both procedures, although hydride trapping has been used to overcome this difficulty. Several studies on this subject are shown in Table 7.

3.1.2 Inductively Coupled Plasma as an Atomic Cell, with Emission or Mass Spectrometry Detection

Since their commercial appearance in the 1970s, ICPs as atomic cells have continuously been improved in terms of performance and ease of operation. At the same time they have become available to more and more laboratories thanks to the decrease in their cost. As a result, ICP is now one of the most promising atomic cells in terms of its efficiency in producing and exciting atomic (or ionic) vapor, which may later be observed by its optical emission (ICP/OES) or by analyzing its mass spectrum (ICPMS).

Table 8 collects together several studies performed using ICP as an atomic cell and different detection devices.

Direct analysis of seawater samples poses a difficult task for ICP atom cells, because the elevated salt content interferes with the proper functioning of the sample introduction system. In this sense, a distinction between sample introduction in ICP/OES and ICPMS should be made. While optical detectors proceed at atmospheric pressure, mass spectrometers work under high vacuum and hence an interface should be provided for operation. Although solutions with a relatively high saline concentration (up to 1% solids) may be employed when using optical detection, in mass spectrometry (MS) instruments these solutions may produce deposition of material on the sampler and skimmer cones and subsequent restriction of the orifices, which in turn will cause a reduction in sensitivity and long-term stability.

Taking into account the nature of the final detection system (photomultiplier tube in ICP/OES and electron multiplier in ICPMS), it could be concluded that the latter would be more efficient than the former as no ''photon to electron'' transducer interface is needed. The increase in instrumental sensitivity may be over 1000 times for the electron multiplier when compared with the photomultiplier.

The extent and nature of interferences are also a function of the detection system: it was thought that the simplicity of the elemental mass spectra relative to the atomic emission spectra of the elements was an important advantage of ICPMS over ICP/OES. Although

Table 6 ETAAS

Element	Characteristics	Ref.
Ag	Coprecipitation of silver from seawater with mercury(II) sulfide	35
Ag, As, Cd, Cr, Hg, Ni, Pb	Use of chemical modification and multi-injection technique approaches	73
As, Cd, Hg, Pb	Direct determination by vapor generation in iridium-coated graphite tubes	74
As	Direct determination using D2 and Zeeman background correction	75
Cd	Automated on-line preconcentration and direct injection	76
Cd	Preconcentration using a dynamically coated column of quaternary ammonium salt on C_{18}-bonded silica gel	22
Cd, Pb, Mn	Existing approaches and critical parameters	77
Cd	Sodium hydroxide as chemical modifier	78
Cd, Pb	Fully automatic on-line separation preconcentration system	24
Cd	Zeeman-effect background correction and oxalic acid as a chemical modifier	79
Cd	Preconcentration by solvent extraction	17
Cd	Direct determination	80
Co, Cu, Ni, Pb, Cd, Fe	Mechanized continuous flow system for the preconcentration and determination	81
Cr(III), Cr(VI)	Determination of dissolved chromium(III) and chromium(VI)	54
Cr, Cu, Mn	Use of a transversely heated GF	82
Cr	Direct determination with multiple injections	83
Cr	Zeeman effect background correction and a multi-injection technique	84
Cr	Direct determination in estuarine and coastal waters	85
Cu, Cd	Preconcentration and ETAAS	86
Cu, Mn, Ni	FI on-line sorption preconcentration in a knotted reactor	37
Cu, Mo	Automated on-line preconcentration system	87
Cu, Cd	FI semi-on-line preconcentration	88
Cu, Mn	Direct and simultaneous determination	89
Hg	Comparative study of magnesium nitrate, palladium nitrate and reduced palladium as modifiers	90
Mn	Sodium hydroxide as chemical modifier for removal of interference	91
Mn	Comparative study of different procedures for the determination of total manganese	92
Mn	Use of a tungsten atomizer	93
Mo, V	Direct and simultaneous determination	94
Mo, Cr, Mn	Direct determination	95
Pb, Cu	FI on-line Co-APDC coprecipitation preconcentration	41
Pb	Improved automatic on-line preconcentration system	96
Pb	Use of a transversely heated furnace and oxalic acid or Pd/Mg as modifiers	97
Pb (labile)	Electrodeposition on pyrolytic graphite platforms	98
Sn	Preconcentration and speciation by HPLC	99
Sn (organic)	Organotin compounds in coastal waters of Greece	100
Yb	Preconcentration by coprecipitation and solvent extraction	20
Zn	Effects of chemical and spectral interferences	101
Zn	Direct determination	102
Several	Direct determination with Zeeman-effect background correction	103
Several	Application of palladium modifiers and a fractal approach	104
Several	Automatic on-line preconcentration system	105
Several	Use of FI/GFAAS	106
Several	Determination of trace metals in rainwater subject to marine influence	107
Several	On-line preconcentration	108
Several	Determination of trace elements with and without a preconcentration step	109
Several	On-line separation and preconcentration	110

GF, graphite furnace.

not as important as in OES, numerous serious overlaps have been found in ICPMS. It must be emphasized that optical interferences are a result of the complexity of the atomic/ionic spectra, while mass/charge overlappings arise primarily from the generation of molecular species from sample, plasma, and matrix constituents. We cannot alter the atomic spectral complexity of a plasma constituent (we can use high resolution for solving some overlappings), but by changing plasma conditions and plasma chemistry, we can alter the nature of plasma constituents, controlling or eliminating certain spectral interferences in ICPMS.

Several specific sources of spectral interferences can be identified in ICPMS, including basic background species, isobaric or interelement spectral overlaps, analyte species spectral overlaps, and matrix-induced molecular species.

Table 7 HG/AAS

Element	Characteristics	Ref.
As	Automated hydride AAS with HPLC	111
As, Se	Electrochemical HG with in situ concentration	112
Cd	HG with in situ concentration	113
Sb	Electrochemical HG with in situ concentration	114
Se	Selenium species by FI HG in situ trapping	115
Sn	Separation by solvent extraction and HG/ETAAS	19

Interelement formation is one of the major drawbacks in MS detection: it is well known that matrix ions generated in the plasma deposit on the interface cones, changing the ionization potential of the plasma and reducing the transport efficiency of the analyte ions into the ion–lens region due to space charge effects. In each case signal suppression results.The combination of these ions with other matrix or plasma ions to form polyatomic interferences has also to be considered when the masses of those compounds lie close to the analyte mass.

This is the case for most heavy metals and the compounds formed between Cl, Br, O, Na, Ca, K, and H present in seawater and N or Ar ions present in the plasma, as can be seen in Table 9.

For these reasons, most of the procedures for seawater analysis begin with a pretreatment step that is used as a preconcentration and matrix-conditioning step at the same time. There are, however, a number of papers dealing with direct determination of heavy metals in seawater[125,137,142] by plain dilution of the samples, taking advantage of the very low LODs obtained with this technique and the relatively high contents of metals in the samples under consideration. It should be mentioned that in some direct determinations ETV was used for the same purpose: the reduction of salt content.

Isotopic dilution techniques, usually employed in radiochemical analysis, have been adapted to MS detection, giving rise to ICP/IDMS systems, that rely for their success on the existence of well-defined and carefully calibrated isotopic mixtures and the possibility of finding noninterferred lines for those analyte isotopes that do not appear in the sample.

3.1.3 Atomic Fluorescence Spectrometry

The major drawbacks of AFS for the analysis of heavy metals are the interferences, spectral or chemical, that can exist when analyzing a particular matrix. The former (spectral) are mainly caused by the atom cell design and the presence of molecular fluorescence of matrix concomitants and the latter (chemical) are associated with interactions between the analyte atoms and the matrix compounds either in condensed phase (loss of sensitivity caused by poor atomization) or in vapor phase, where quenching effects are very strong.

Then, obtaining a good LOD means that the concentration of the fluorescence quenching species in the vapor phase is eliminated or diminished and the interelement compounds that might be formed in the atomization step are broken down.

Flames and ICPs provide enough energy to atomize even the refractory elements, but at the same time the population of quenching compounds in those atom cells is high enough to eliminate the fluorescence signal. ETAs, on the other hand, will be free from fluorescence quenching species because no combustion gases will be present, but matrix gases within the furnace atmosphere will present a serious quenching problem.

From the point of view of the spectral source, it should be considered that because fluorescence signals are directly proportional to the spectral line source light intensity, conventional excitation sources like hollow cathode lamps, high-intensity hollow cathode lamps or electrodeless discharge lamps will produce poor excitation and consequently poor LODs while posing geometric limitations to the instrument developer. Lately, laser excitation has been the choice for obtaining the best LOD.

Heavy metal analysis by atomic fluorescence has been confined to a few cases in which the atomization conditions allow good sensitivity. These are described in Table 10.

As can be seen from Table 10, most of the studies performed on seawater by AFS employ a preliminary vaporization step: cold vaporization (Hg), HG (As, Se) or ETV (Ag) in order to ''clean'' the atomic vapor of interfering compounds and to facilitate the atomization step, avoiding the use of flames. Laser excitation is used when possible.

3.2 Other Atomic Spectrometries

3.2.1 X-ray Fluorescence Spectrometry

Few studies have been reported on heavy metal analysis on seawater or brine by X-ray fluorescence spectrometry (XRFS). The reasons for this could be found in the poor sensitivity of this technique and/or the need for cumbersome preconcentration and pretreatment steps. However, some laboratories have succeeded in certification experiments on seawater samples. Table 11 shows the reports found since 1989.

3.2.2 Nuclear Spectrometries

Neutron activation analysis has also scarcely been used for heavy metals determination in seawater or brine. The need for a neutron source or adequate facilities for

Table 8 Use of ICP as atomic source

Element	Detection	Characteristics	Ref.
Fe	IDMS	High-resolution isotope dilution after $Mg(OH)_2$ coprecipitation	116
Ni, Cu, Zn, Ag, Cd, Pb	IDMS	Simultaneous determination	117
Pb, Cu, Cd	IDMS	Low blank preconcentration technique small-volume samples	118
V	IDMS	ETV	119
As, Sb, Hg	MS	Use of FI vapor generation	120
Cu, Cd, Pb, Bi, Se(IV)	MS	Use of ETV on-line separation	121
Cu, Ni, Zn, Mn, Co, Pb, Cd, V	MS	FI	122
Fe	MS	Open ocean seawater reference material NASS-3	123
Hg	MS	Use of liquid chromatography as preconcentration/separation step	124
Mn, Mo, U, Cd, Pb	MS	Direct determination, using FI and ETV	125
Mo	MS	Candidate reference material	126
Pb	MS	Determination of isotope ratios after preconcentration using Chelex-100	28
Pb	MS	On-line determination isotope ratios	127
Rare earths	MS	Preconcentration with poly(acrylaminophosphonic dithiocarbamate) chelating fiber	29
Se	MS	Marine-certified reference materials by HG	128
Se(IV), As (total), Sb (total), Ge (total)	MS	Automated continuous-flow HG	129
Several	MS	Use of fluorinated metal alkoxide glass-immobilized 8-hydroxy quinoline preconcentration	130
Several	MS	Preconcentration by gallium coprecipitation	131
Several	MS	Related techniques coupled with ICPMS	132
Several	MS	ETV of dithiocarbamate complexes in methylisobutyl ketone	133
Several	MS	Sequential determination after ETV of dithiocarbamate complexes in methylisobutyl ketone	134
Several	MS	On-line preconcentration and matrix elimination with FI	135
Several	MS	Off-line chromatographic separation and preconcentration	136
Several	MS	Direct determination using ETV	137
Several	MS	Estuarine and seawater reference materials (HRMS)	138
Several	MS	Off-line dithiocarbamate solvent extraction	139
Several	MS	Preconcentration by formation of water-soluble complexes and adsorption on C_{18}-bonded silica gel	43
Several	MS	Nitrilotriacetate chelating resin as stationary phase for preconcentration and elimination of matrix effects	33
Several	MS	Ion chromatography separation and preconcentration	140
Several	MS	Rapid preconcentration by solid-phase extraction	141
Several	MS	Direct determination by ETV $Pd-HNO_3$ modifier	142
Several	MS	Capabilities and limits for direct determination	143
Several	MS	Immobilization of 8-hydroxyquinoline on to silicone tubing for FI	144
Silica	MS	On-line determination by ion exclusion chromatography	145
Several	MS/OES	Multielement major-to-trace elements after chelating resin preconcentration	31
Several	MS/OES	Multielement determination after aluminum coprecipitation associated with magnesium	146
Co, Cu, Mo, Mn, Fe, Ti, V	OES	Reference river water and seawater samples with the third phase preconcentration	147
Cu, Ni, Cd	OES	Silica-immobilized 8-quinolinol, on-line column preconcentration	26
Hg	OES	Use of FI cold vapor generation and preconcentration on silica functionalized with methyl thiosalicylate	148
Mo, V, Zr, Cr	OES	Sorption preconcentration	40
Pb	OES	Chelating resin preconcentration and HG	27
Several	OES	Determination in marine media	149
Several	OES	Determination in Arima spring water after coprecipitation with iron(III) hydroxide	150
Several	OES	Preconcentration using a cellulose–zinc hydroxide system	30
Several	OES	HPLC	151
Several	OES	Sorption preconcentration	45

IDMS, isotopic dilution mass spectrometry; HRMS, high-resolution mass spectrometry; OES, optical emission spectrometry; ETV, electrothermal vaporization.

Table 9 Some interferent compounds found in seawater (MS detection)

Element	Mass	Interfering ion
Ti	48	$^{48}Ca^{+}$
V	51	$^{35}ClO^{+}$
Cr	52	$^{35}ClOH^{+}$, $^{40}ArC^{+}$
Fe	54	$^{37}ClOH^{+}$, $^{40}ArN^{+}$
Cu	63	$^{40}ArNa^{+}$
As	75	$^{40}Ar^{37}Cl^{+}$

Table 10 AFS

Element	Characteristics	Ref.
Ag	Direct determination using laser-excited graphite tube electrothermal atomization	152
As	Direct determination continuous-flow HG	153
As	HG	154
Hg	FI with on-line oxidation	155
Pb	Laser-excited direct determination	156
Se	Nondispersive hydride determination with electrothermally heated quartz tube–flame atomizer	157
Se(IV), Se(VI)	Direct determination by continuous flow HG	59
Se(IV), Se(T)	Integrated on-line FI HG	158

Table 11 XRFS

Element	Characteristics	Ref.
Ge, As, Sn, Sb	Simultaneous determination with total-reflection XRFS using the HG	159
Several	Total reflection XRFS of trace metals in the North Sea	160
Several	ZISCH and PRISMA North Sea research projects	161
Several	Reference materials: performance of total reflection XRFS in the intercomparison and certification stages	162

handling radioisotopes might be some of the reasons for the lack of applications of this technique in the field of seawater analysis. Some published papers are presented in Table 12.

Sample conditioning and preconcentration steps are dominant in the studies presented, while LODs are similar to those obtained with other techniques, meaning that little is gained by the use of nuclear techniques, except for those laboratories that already have facilities for such procedures.

Table 12 Neutron activation analysis

Element	Characteristics	Ref.
Cd	Coprecipitation and neutron activation analysis	163
Hg	Neutron activation	164
Se(IV)	Neutron activation analysis after coprecipitation of dibenzyl dithiocarbamate with phenolphthalein	165
Se(IV), Se(VI)	Neutron activation analysis after chemical precollection	166

4 ELECTROANALYTICAL TECHNIQUES

Electroanalytical procedures have been used for the analysis of heavy metals since the development of conventional (DC) polarography, and they have proved to be reliable methods for water analysis. The ever-increasing need to improve the LOD, as has been stated earlier, has been the driving force in the development of new electroanalytical technologies to improve the sensitivity and selectivity of the detecting devices and new ways to enhance the electroactivity of the sample species.

Technological improvements, focused mainly on the field of sampling and measuring current intensities and potential differences, have changed traditional electrolysis over a mercury drop to more sophisticated ways of observing the same or related phenomena. In this field, the major role played by computer-assisted techniques, under continuous development, that have set new milestones in electrochemical LODs should be stressed.

Voltammetry, based on the current–potential relationship in an electrochemical cell, and particularly with the current–time response of an electrode at a controlled potential, uses the current intensity at that potential as a means of measuring the concentration of the electroactive species.

Considering that the measured current is related to the rate of charge transfer through the electrode–solution interface, and that from an analytical point of view only Faradaic processes contribute to the overall sensitivity of the technique, then all the mechanisms which might enhance either the charge transference or the Faradaic to capacitive current ratio will be welcome. In the same direction, current sampling procedures will provide another means of increasing the sensitivity of the technique. It should be mentioned here that one of the major drawbacks of the electroanalytical techniques is the need to eliminate oxygen from the solutions in order to avoid its interference either as an electroactive substance

or as a competitor for the reduction of the metals at the electrode surface.

Stripping voltammetry, one of the electroanalytical techniques designed to deal with traces of metals, relies on the preconcentration of traces by their electrodeposition over a convenient electrode material (usually mercury) and then the release of the electrodeposited material, with the aid of a potential sweep, while measuring the curve i against V.

As the LOD will be a function of the ability to detect the difference between the current obtained with the analyte and the residual current (or blank current), then better results could be expected when reducing the residual current and increasing the sensitivity of the detecting device. Careful selection of the supporting electrolyte is mandatory in order not to add impurities or simply to contaminate the sample, giving rise to a signal superimposed on the analyte's signal, particularly when measuring current intensities in the nanoampere order.

Voltammetric analysis has been used scarcely in the determination of heavy metals in seawater or brine solutions. The reasons for this can be found in the widespread use of stripping techniques, which are more sensitive than classical voltammetric analysis.

4.1 Stripping Techniques

According to the nature of the species accumulated on the surface of the electrode material and consequently the nature of the potential sweep, different techniques may be described:

Table 13 Electroanalytical techniques

Element	Technique	Characteristics	Ref.
As(III), As(V)	ASV	Gold-film electrode preparation for ASV	168
Cu	ASV	Slopes of standard addition plots by ASV	169
Cu, Cd, Pb	ASV	Chelating resin preconcentration and Hg(II)-assisted elution DPASV	23
Pb, Cu	ASV	Square wave stripping voltammetry at a rotating thin-film mercury electrode	170
Several	ASV	ASV	171
Several	ASV	FI/ASV preceded by immobilized quinolin-8-ol silica gel preconcentration	32
As	CSV	CSV in the presence of pyrrolidine dithiocarbamate	172
As(III)	CSV	Square wave CSV	173
Cd	CSV	CSV of complexes with 5-fluorouracil	174
Co	CSV	Direct determination by high-speed adsorptive CSV	175
Co	CSV	CAdSV	176
Cr	CSV	CAdSV with TTHA and nitrate	177
Cr	CSV	Speciation using CCSV	50
Cu, Pb, Cd	CSV	Rapid determination by AdSV	178
Fe	CSV	Adsorptive CSV	179
Fe	CSV	AdSV	180
Fe	CSV	Fast AdSV	181
Fe	CSV	Redox speciation by CCSV	56
Fe	CSV	CCSV	182
Fe(III)	CSV	Complexation with natural organic complexing ligands using CSV	57
Mn(II)	CSV	Differential pulse CSV	183
Ni	CSV	AdSV with phenyldithiocarbamate	184
Pb	CSV	AdSV with xylenol orange	185
Se	CSV	Adsorptive CSV	186
Several	CSV	AdSV in brine. Part 1. Batch voltammetric measurements	187
Several	CSV	AdSV in brine. Part 2. Development and evaluation of an FI system	188
Several	CSV	Simultaneous determination using CSV with mixed ligands	189
U	CSV	Adsorptive differential pulse voltammetry	190
U(VI)	CSV	Direct determination by CSV	62
Mn	PSV	Rapid determination by stripping potentiometry	191
Zn	PSV	Elimination of zinc/copper intermetallic interference by PSA	192
Cr(III)	V	Preconcentration by adsorption of silica and voltammetric determination	36
Cu	V	Influence of ethylenediamine in the voltammetric determination at a rotating mercury film electrode	193
Hg	V	Voltammetry at glassy carbon electrodes modified with polycationic ionomers	194
Several	V	Comparison of alternating current and differential pulse voltammetry	195
Several	V	Shipboard automated voltammetric system	196
Zn	V	Complexation parameters in marine and estuarine waters	197

V, voltammetry; TTHA, triethylene tetranitrile hexa acetic acid.

1. ASV: the ''naked'' metals are reduced at the surface of the electrode (and hence accumulated) and then they are reoxidized by a potential sweep and the resulting current is measured.
2. CSV: the metals are complexed with proper ligands and adsorbed on to the electrode surface at a prefixed potential; the complexed metals are then reduced by a potential sweep and the reduction current is measured. This procedure is also known as adsorptive stripping voltammetry'' (AdSV). The addition of oxidant species may lead to a reoxidation of the complexes during the lifetime of the cycle, giving rise to catalytic adsorptive (cathodic) stripping voltammetry (CAdSV).

It can be said that the principal variables affecting the performance of a voltammetric technique are the electrolysis time, the stirring rate, the electrode area to volume ratio and the mode of the voltage sweep.[167] While the first three account for an improvement in mass transport to the electrode, the last takes into consideration the way the metal ions can be oxidized out of the electrode surface. Novel electrode surface materials, shapes and movements are then added to the previous classification, broadening the panorama of electroanalytical techniques available.

One of the latest developments in the field of the stripping procedures has been potentiometric stripping analysis or potentiometric stripping voltammetry (PSA or PSV): in this case, the system is treated as in ASV until the deposition has been completed; it is then disconnected from the potentiostat and the curve V versus i is measured.

Highly saline solutions may be analyzed directly by ASV or CSV provided that the electrolysis time is long enough to accumulate a quantity of analyte great enough to be measured accurately afterwards.

Each of the techniques briefly described above has different capabilities for seawater or brine analysis associated with them. Its high sensitivity, speciation ability, and the use of the matrix (NaCl) as support electrolyte are advantages, while the need for sophisticated equipment or mercury as electrode material (in recession nowadays) is a disadvantage.

Table 13 shows some of the latest studies performed on seawater and brine using electroanalytical procedures.

5 COMPARISON OF DIFFERENT APPROACHES

So far we have tried to give a broad panorama of the possible alternatives for the determination of heavy metals in seawater and in brine. Most of the analytical techniques available have been discussed, and a comparison of these alternatives is imperative in order to establish some useful guidelines at the time of choosing the appropriate technique of analysis.

The aims of heavy metal determination are of paramount importance in analyzing alternative procedures. At the beginning of this chapter (section 2.1) we classified the reasons for heavy metal determination into three broad categories, and the procedures for these determinations should be selected in consequence.

Modern analytical chemistry tends to keep the analyte in a ''clean environment'' reducing handling to a minimum, giving rise to the new breed of hyphenated techniques – those that include a separation step before the detection step. FI techniques have already been used for the determination of most heavy metals in seawater and brine, being employed not only with some of the previously described detection devices, but also with traditional detection systems like fluorescemetric and spectrophotometric devices.

Table 14 shows some of the latest studies on seawater and brine in which FI techniques have played an important role. This list is by no means exhaustive, as many FI studies have been mentioned before and will not appear in this table; furthermore, with one exception it only covers the 1990s.

The fact that UV/VIS[211–216] and chemiluminiscence[199–205] detection devices are used for the analysis of seawater samples (see Table 14) demonstrates that FI techniques are suitable for concentrating heavy metals to such a level as to allow the use of less sensitive detectors, without contamination by handling, and in some cases performing reagent purification in the same system.

Selecting the analytical procedure will be then a very demanding task, especially when starting from scratch, that is when one has to decide which equipment to purchase and which determination method to use. The reader is referred to the first paragraphs of this article in order to determine her or his needs.

There is a long-standing competition between analytical spectroscopists and analytical electrochemists about the efficacy of each technique for dealing with traces of metals. We will not enter in this discussion, but most of the literature consulted is devoted to spectroscopies (83%) and thus, unavoidably, there will be some bias towards spectroscopic methods for trace analysis.

With regard to the metal involved, and taking into account the usual concentration level in which the metal is found in seawater or brine, two broad metal groups could be mentioned: those forming some kind of volatile compounds, and the rest.

- ***Forming volatile compounds***: As, Se, Te, Pb, Sn, Sb, Ge, Cd and Hg belong to this category. They

Table 14 FI applied to seawater and brine analysis

Element	Detection	Characteristics	Ref.
Several	Atomic Spectroscopy	Cationic species	198
Co	Chemiluminiscence	On-line preconcentration	199
Cu	Chemiluminiscence	Chemiluminiscence detection	200
Fe	Chemiluminiscence	Subnanomolar levels	201
Fe(II), Fe(T)	Chemiluminiscence	Subnanomolar levels of iron(II) and total dissolved iron	202
Fe(II)	Chemiluminiscence	Stopped flow at subnanomolar levels	203
Mn	Chemiluminiscence	Rapid determination	204
Mn	Chemiluminiscence	Selective and sensitive determination	205
Cd	ETAAS	Cold vapor generation and preconcentration on coated graphite tubes	206
Hg	ETAAS	Cold vapor generation and preconcentration on coated graphite tubes	207
Cd	FAAS	Use of 4-(2-pyridylazo) resorcinol or 2-(2-pyridylazo)-5-dimethyl aminophenol as chelating agent for sorbent extraction	208
Zn	Fluorescence	Fluorometric detection	209
Several	ICPMS	Determination of Cu, Ni, Zn, Mn, Co, Pb, Cd, and V	210
Co	UV/VIS	Shipboard determination with catalytic spectrophotometric detection	211
Fe	UV/VIS	Determination of dissolved iron in seawater	212
Fe	UV/VIS	On-line preconcentration and spectrophotometric detection	213
Mn	UV/VIS	On-line preconcentration and spectrophotometric detection	214
Mn	UV/VIS	Shipboard method using in-valve preconcentration and catalytic spectrophotometric detection	215
Mn	UV/VIS	Determination of dissolved manganese with colorimetric detection	216

UV/VIS, ultraviolet/visible.

may all form some kind of volatile compound, mainly their hydrides (except for Hg, which is volatile as the element). Volatile compounds will be decomposed at high temperatures to yield an atomic cloud, which in turn will yield adequate conditions for atomic spectroscopic methods, free from matrix interferences. For these reasons, HG or cold vapor generation are preferred when analyzing this set of metals, leaving electrochemical methods as a second choice.

- ***Metals that remain in solution (the rest)***: for those metals that do not form volatile compounds easily, a similar criterion for choosing the appropriate technique could be used. Those that will be easily volatilized (not as compounds but as elements) should be analyzed by spectroscopic techniques, as sensitivity will be high if they are easily volatilized. Nevertheless, electrochemical techniques may be preferred for some easily reducible ions or ions which form easily reducible complexes (Co, Cr, Cu, Fe, Pb, Zn).

A new concept has been added to the usual features of merit in analytical procedures: the number of determinations that can be performed per unit time. This is a measure of the capabilities of the technique to deal with a heavy workload, and is used to evaluate its automation potential. Taking into account that most of the sample preparation work might be performed using FI techniques, detection will be the determining step. In this field, multielemental detection devices, such as those that are usually coupled with ICP atom cells, are one of the major improvements in terms of efficiency and operation economy.

Almost all detection devices show instant response to the analyte except for stripping techniques, which require a certain electrodeposition time before reaching the measuring step. Thus, stripping methods may have a poor performance in terms of analysis throughput.

In summary, all techniques presented here will yield the best results when used by trained professionals who have mastered the analytical procedures, and hence the choice of technique will be dictated by common sense rather than for technological reasons.

ACKNOWLEDGMENTS

The author wishes especially to thank Dra Mabel Tudino for her wise advice and her collaboration on the manuscript, and to the people of Laboratorio de Análisis de Trazas who have made this work possible.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AdSV	Adsorptive Stripping Voltammetry
AES	Atomic Emission Spectrometry

AFS	Atomic Fluorescence Spectrometry
APDC	Ammonium Pyrrolidine Dithiocarbamate
ASV	Anodic Stripping Voltammetry
CAdSV	Catalytic Adsorptive Stripping Voltammetry
CCSV	Catalytic Cathodic Stripping Voltammetry
CSV	Cathodic Stripping Voltammetry
DC	Direct Current
DPASV	Differential Pulse Anodic Stripping Voltammetry
EDTA	Ethylenediamine Tetraacetic Acid
ETA	Electrothermal Atomizer
ETAAS	Electrothermal Atomic Absorption Spectrometry
ETV	Electrothermal Vaporization
FAAS	Flame Atomic Absorption Spectrometry
FAES	Flame Atomic Emission Spectrometry
FI	Flow Injection
FIA	Flow Injection Analysis
GF	Graphite Furnace
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
GLC	Gas–Liquid Chromatography
HG	Hydride Generation
HPLC	High-performance Liquid Chromatography
HRMS	High-resolution Mass Spectrometry
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICP/OES	Inductively Coupled Plasma/Optical Emission Spectrometry
IDMS	Isotopic Dilution Mass Spectrometry
IUPAC	International Union of Pure and Applied Chemistry
LOD	Limit of Detection
MLD	Methodological Limit of Detection
MS	Mass Spectrometry
OES	Optical Emission Spectrometry
PSA	Potentiometric Stripping Analysis
PSV	Potentiometric Stripping Voltammetry
TTHA	Triethylene Tetranitrile Hexa Acetic Acid
UV/VIS	Ultraviolet/Visible
XRFS	X-ray Fluorescence Spectrometry

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by Electrochemical Methods • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Luminescence in Environmental Analysis • Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Neutron Activation in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

REFERENCES

1. European Community, 'Council Directive of 30 October 1979 on the Quality Required of Shellfish Waters (79/923/EEC)', *Off. J. Eur. Commun.*, NL 281, 47–52, 1979.
2. D.G. Cunningham, 'Laboratory Robotic Automation Through Intelligent System Control and Instrument/Data Networking', *Scientific Computing & Automation*, **21**, 26 December (1989).
3. T.R. Crompton, *Analysis of Seawater*, Butterworth, Guildford, 1989.
4. R.W. Furness, in *Heavy Metals in the Marine Environment*, ed. P.S. Rainbow, CRC Press, Boca Raton, FL, 1988.
5. G.F. Kirkbright, L. Ranson, 'Use of Nitrous Oxide–Acetylene Flame for the Determination of Arsenic and Selenium by Atomic Absorption Spectrometry', *Anal. Chem.*, **43**, 1238–1241 (1971).
6. R.D. Ediger, 'Atomic Absorption Analysis with the Graphite Furnace Using Matrix Modification', *At. Absorpt. Newslett.*, **14**, 127–130 (1975).
7. G. Schlemmer, B. Welz, 'Palladium and Magnesium Nitrates, a More Universal Modifier for Graphite Furnace Atomic Absorption Spectrometry' *Spectrochim. Acta, Part B*, **41**(11), 1157–1165 (1986).
8. E.J. Krakovska, 'Utilization of WETA 82 Tungsten Atomizer for the Determination of Rare Earth Elements', *Anal. At. Spectrom.*, **5**, 205–207 (1990).

9. K. Ohto, I. Sugiyama, T. Mizuno, 'Determination of Phosphorus in Biological Materials Using Electrothermal Atomization Atomic Absorption Spectrometry with a Molybdenum Tube Atomizer', *Analyst*, **115**, 279–282 (1990).
10. M. Berglund, W. Frech, D.C. Baxter, 'Achieving Efficient Multielement Atomization Conditions for Atomic Absorption Spectrometry Using a Platform-equipped Integrated-contact Furnace and a Palladium Modifier', *Spectrochim. Acta, Part B*, **46**, 1767–1777 (1991).
11. W. Frech, B. L'vov, 'Matrix Vapors and Physical Interference Effects in Graphite Furnace Atomic Absorption Spectrometry II', *Spectrochim. Acta, Part B*, **48**, 1381–1389 (1993).
12. P.J. Parsons, W. Slavin, 'A Rapid Zeeman Graphite Furnace Atomic Absorption Spectrometric Method for the Determination of Lead in Blood', *Spectrochim. Acta, Part B*, **48**, 925–939 (1993).
13. B.V. L'vov, L.A. Pelieva, A. Sharnopolsky, 'Decrease in the Effect of the Base During the Atomic Absorption Analysis of Solutions in Tube Furnaces by Evaporation of Samples from a Graphite Substrate', *Zh. Prikl. Spectroskopiya*, **27**, 395–399 (1978).
14. W. Slavin, D.C. Manning, G.R. Carnrick, 'The Stabilized Temperature Platform Furnace', *At. Spectrosc.*, **2**, 137–145 (1981).
15. W. Slavin, 'Atomic Absorption Spectroscopy. The Present and Future', *Anal. Chem.*, **54**, 689A–694A (1982).
16. M.L. Wells, K.W. Bruland, 'An Improved Method for Rapid Preconcentration and Determination of Bioactive Trace Metals in Seawater Using Solid-phase Extraction and High Resolution Inductively Coupled Plasma Mass Spectrometry', *Mar. Chem.*, **63**(1–2), 145–153 (1998).
17. J.S. Lee, W.H. Choi, K. Uesugi, 'Determination of Cadmium in Seawaters by Graphite Furnace Atomic Absorption Spectrophotometry After Preconcentration by Solvent Extraction', *Nippon Kaisui Gakkaishi*, **49**(1), 22–26 (1995).
18. S. Yamaguchi, K. Uesugi, 'Determination of Iron and Copper in Seawater and Common Salt by the Atomic Absorption Spectrophotometric Method with 2-hydroxy-1-naphthaldehyde-4-phenyl-3-thiosemicarbazone Extraction', *Nippon Kaisui Gakkaishi*, **49**(6), 352–355 (1995).
19. S. Ni, H. Hang, A. Li, B. He, F. Xu, 'Determination of Tributyltin and Inorganic Tin in Seawater by Solvent Extraction and Hydride Generation Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**(5), 385–387 (1991).
20. O. Fujino, S. Nishida, H. Togawa, K. Hiraki, 'Determination of Ytterbium in Seawater by Graphite Furnace Atomic Absorption Spectrometry After Coprecipitation and Solvent Extraction', *Anal. Sci.*, **7**(6), 889–892 (1991).
21. J.S. Lee, K. Uesugi, T. Nagahiro, W.H. Choi, D. Kim, 'Determination of Zinc in Seawaters by Flame Atomic Absorption Spectrophotometry After Solvent Extraction Using a New Chelating Agent', *Nippon Kaisui Gakkaishi*, **50**(5), 344–348 (1996).
22. K. Akatsuka, Y. Yoshida, N. Nobuyama, S. Hoshi, S. Nakamura, T. Kato, 'Preconcentration of Trace Cadmium from Seawater Using a Dynamically Coated Column of Quaternary Ammonium Salt on C_{18}-Bonded Silica Gel and Determination by Graphite Furnace Atomic Absorption Spectrometry', *Anal. Sci.*, **14**(3), 529–533 (1998).
23. C. Yuh, Y.L. Tu, J. Mierzwa, 'Chelating Resin Preconcentration and Hg(II)-assisted Elution Followed by Differential-pulse Anodic Stripping Voltammetry for the Determination of Cd, Cu and Pb in Seawater', *Fresenius' J. Anal. Chem.*, **360**(5), 550–555 (1998).
24. F.M. Fernandez, J.D. Stripeikis, M.B. Tudino, O.E. Troccoli, 'Fully Automatic On-line Separation Preconcentration System for Electrothermal Atomic Absorption Spectrometry: Determination of Cadmium and Lead in Seawater', *Analyst (Cambridge, UK)*, **122**(7), 679–684 (1997).
25. Y. Sung, Z. Liu, S. Huang, 'Use of Muromac A-1 Chelating Resin for Determination of Copper and Molybdenum in Seawater by Atomic Absorption with On-line Preconcentration', *Spectrochim. Acta, Part B*, **52**, 755–764 (1997).
26. C. Lan, M. Yang, 'Synthesis, Properties and Applications of Silica-immobilized 8-quinolinol. Part 2. On-line Column Preconcentration of Copper, Nickel and Cadmium from Seawater and Determination by Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Chim. Acta*, **287**(1–2), 111–117 (1994).
27. R.A. Reimer, A. Miyazaki, 'Determination of Lead in Seawater by Inductively Coupled Plasma Atomic Emission Spectrometry Combined with Chelating Resin Preconcentration and Hydride Generation', *J. Anal. At. Spectrom.*, **7**(8), 1238–1244 (1992).
28. A. Miyazaki, R.A. Reimer, 'Determination of Lead Isotope Ratios and Concentrations in Seawater by Inductively Coupled Plasma Mass Spectrometry After Preconcentration Using Chelex-100', *J. Anal. At. Spectrom.*, **8**(3), 449–452 (1993).
29. T. Zhang, X. Shan, R. Liu, H. Tang, S. Zhang, 'Preconcentration of Rare Earth Elements in Seawater with Poly(acrylaminophosphonic dithiocarbamate) Chelating Fiber Prior to Determination by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **70**, 3964–3968 (1998).
30. I.M.M. Kenawy, 'Preconcentration, Separation and Multielements Determination in Seawater Using a Cellulose–Zinc Hydroxide System with Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Sci.*, **8**(1), 45–50 (1992).
31. S. Ji, C. Kimata, T. Yabutani, A. Itoh, E. Fujimori, K. Chiba, H. Haraguchi, 'Multielement Determination of Major-to-trace Elements in Natural Solar Salt by

ICP/AES and ICP/MS After Chelating Resin Preconcentration', *Biomed. Res. Trace Elem.*, **8**(1), 37–46 (1997).
32. B.J. Daih, H.J. Huang, 'Determination of Trace Elements in Seawater by Flow Injection Anodic Stripping Voltammetry Preceded by Immobilized Quinolin-8-ol Silica Gel Preconcentration', *Anal. Chim. Acta*, **258**(2), 245–252 (1992).
33. H. Kumagai, M. Yamanaka, T. Sakai, T. Yokoyama, T.M. Suzuki, T. Suzuki, 'Determination of Trace Metals in Seawater by Inductively Coupled Plasma Mass Spectrometry Interfaced with an Ion Chromatographic Separation System: Effectiveness of Nitrilo Triacetate Chelating Resin as the Column Stationary Phase for Preconcentration and Elimination of Matrix Effects', *J. Anal. At. Spectrom.*, **13**(6), 579–582 (1998).
34. K. Akatsuka, T. Katoh, N. Nobuyama, T. Okanaka, M. Okumura, S. Hoshi, 'Preconcentration of Trace Amounts of Zinc in Seawater Using a Dynamically Coated Column of Methyl Tricaprylylammonium Chloride on Octadecylsilyl Silica and Determination by Graphite Furnace AAS', *Anal. Sci.*, **12**(2), 109–113 (1996).
35. M.N. Quigley, F. Vernon, 'Coprecipitation of Silver from Seawater with Mercury(II) Sulfide and Determination by Electrothermal Atomic Absorption Spectrophotometry', *Anal. Proc.*, (London), **29**(1), 1–3 (1992).
36. M. Boussemart, C.M.G. Van den Berg, 'Preconcentration of Chromium(III) from Seawater by Adsorption of Silica and Voltammetric Determination', *Analyst (Cambridge, UK)*, **119**(6), 1349–1353 (1994).
37. E. Ivanova, K. Benkhedda, F. Adams, 'Determination of Copper, Manganese and Nickel in Biological Samples and Seawater by Flow Injection On-line Sorption Preconcentration in a Knotted Reactor Coupled with Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **13**(6), 527–531 (1998).
38. L. Zhang, K. Terada, 'Spectrophotometric Determination of Iron(II) in Seawater After Preconcentration by Sorption of its 3-(2-Pyridyl)-5,6-bis(4-phenylsulfonic acid)-1,2,4-triazine Complex with Poly(chlorotrifluoroethylene) Resin', *Anal. Chim. Acta*, **293**(3), 311–318 (1994).
39. M. Fernandez Garcia, R. Pereiro Garcia, N. Bordel Garcia, A. Sanz-Medel, 'On-line Preconcentration of Inorganic Mercury and Methylmercury in Seawater by Sorbent-extraction and Total Mercury Determination by Cold Vapor Atomic Absorption Spectrometry', *Talanta*, **41**(11), 1833–1839 (1994).
40. N.L. Shcherbinina, G.R. Ishmiyarova, G.V. Myasoedova, G.P. Panteleev, N.P. Starshinova, E.M. Sedykh, 'Sorption Preconcentration and Inductively Coupled Plasma Atomic Emission Determination of Molybdenum, Vanadium, Zirconium, and Chromium in Seawater', *Zh. Anal. Khim.*, **46**(8), 1585–1588 (1991).
41. Z. Zhuang, X. Wang, P. Yang, C. Yang, B. Huang, 'Flow Injection On-line Co-APDC Coprecipitation Preconcentration and Determination of Lead and Copper in Seawater with Graphite Furnace Atomic Absorption Spectrometry', *Can. J. Appl. Spectrosc.*, **39**(4), 101–107 (1994).
42. D.S. Murty, G. Chakrapani, 'Preconcentration of Rare Earth Elements on Activated Carbon and its Application to Groundwater and Seawater Analysis', *J. Anal. At. Spectrom.*, **11**(9), 815–820 (1996).
43. K. Akatsuka, T. Suzuki, N. Takahiro, N. Nobuyama, S. Hoshi, K. Haraguchi, K. Nakagawa, T. Ogata, T. Kato, 'Determination of Trace Elements in Seawater by Inductively Coupled Plasma Mass Spectrometry After Preconcentration by Formation of Water-soluble Complexes and their Adsorption on C_{18}-bonded Silica Gel', *J. Anal. At. Spectrom.*, **13**(4), 271–275 (1998).
44. L. Elci, U. Sahin, S. Oztas, 'Determination of Trace Amounts of Some Metals in Samples with High Salt Content by Atomic Absorption Spectrometry After Cobalt–diethyldithiocarbamate Coprecipitation', *Talanta*, **44**(6), 1017–1023 (1997).
45. G.P. Panteleev, G.I. Tsizin, A.A. Formanovskii, N.P. Starshinova, E.M. Sedykh, N.M. Kuz'min, Y.A. Zolotov, 'Sorption Inductively Coupled Plasma Atomic Emission Determination of Metals in Highly Mineralized Natural Waters', *Zh. Anal. Khim.*, **46**(2), 355–360 (1991).
46. W. Stumm, J.J. Morgan, *Aquatic Chemistry*, 3rd edition, J. Wiley & Sons, New York, 1996.
47. A. Warshawsky, in *Synthesis and Separation Using Functional Polymers*, eds. D. Sherrington, P. Hodge, J. Wiley & Sons, New York, 1988.
48. J. Ruzicka, E.H. Hansen, 'Flow Injection Analysis I. New Concept of Fast Continuous Flow Analysis', *Anal. Chim. Acta*, **78**, 145–157 (1975).
49. P. Del Castilho, 'Determination and Speciation of Heavy Metals in Seawater: An Overview of Methods Available and Results Achievable', *Anal. Proc.*, **28**, 253–254 (1991).
50. M. Boussemart, C.M.G. Van den Berg, M. Ghaddaf, 'The Determination of the Chromium Speciation in Seawater Using Catalytic Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **262**(1), 103–115 (1992).
51. R.K. Mugo, K.J. Orians, 'Seagoing Method for the Determination of Chromium(III) and Total Chromium in Seawater by Electron Capture Detection Gas Chromatography', *Anal. Chim. Acta*, **271**(1), 1–9 (1992).
52. M. Sperling, X. Yin, B. Welz, 'Differential Determination of Chromium(VI) and Total Chromium in Natural Waters Using Flow Injection On-line Separation and Preconcentration Electrothermal Atomic Absorption Spectrometry', *Analyst (London)*, **117**(3), 629–635 (1992).
53. E. Beinrohr, A. Manova, J. Dzurov, 'Preconcentration of Cr(III) and Total Cr in Waters in a Flow-through Electrochemical/Sorption Cell for Flame AAS Determination', *Fresenius' J. Anal. Chem.*, **355**, 528–531 (1996).

54. A. Boughriet, L. Deram, M. Wartel, 'Determination of Dissolved Chromium(III) and Chromium(VI) in Seawater by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **9**(10), 1135–1142 (1994).
55. S. Blain, P. Treguer, 'Iron(II) and Iron(III) Determination in Seawater at the Nanomolar Level with Selective On-line Preconcentration and Spectrophotometric Determination', *Anal. Chim. Acta*, **308**(1–3), 425–432 (1995).
56. A.P. Aldrich, C.M.G. Van den Berg, 'Determination of Iron and its Redox Speciation in Seawater Using Catalytic Cathodic Stripping Voltammetry', *Electroanalysis*, **10**(6), 369–373 (1998).
57. M. Gledhill, C.M.G. Van den Berg, 'Determination of Complexation of Iron(III) with Natural Organic Complexing Ligands in Seawater Using Cathodic Stripping Voltammetry', *Mar. Chem.*, **47**(1), 41–54 (1994).
58. Y.S. Kim, S.G. Nho, J.M. Choi, 'Determination of Trace Molybdenum(VI) in Seawater Samples by Ion Pair Formation and Solvent Extraction', *Anal. Sci. Technol.*, **6**(3), 329–334 (1993).
59. Y. He, J. Moreda-Pineiro, M.L. Cervera, M. de la Guardia, 'Direct Determination of Dissolved Selenium(IV) and Selenium(VI) in Seawater by Continuous Flow Hydride Generation Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **13**(4), 289–293 (1998).
60. G.D. Woods, C.W. McLeod, K. Kawabata, S. Nagaoka, H. Okochi, A. Otsuki, 'Microcolumn Field Sampling for Determination and Speciation of Organo-tin in Seawater', *Spec. Publ. – R. Soc. Chem.*, **202**(Plasma Source Mass Spectrometry), 159–165 (1997).
61. J. De Pablo, L. Duro, J. Gimenez, J. Havel, M.E. Torrero, I. Cases, 'Fluorimetric Determination of Traces of Uranium(VI) in Brines and Iron(III) Oxides Using Separation on an Activated Silica Gel Column', *Anal. Chim. Acta*, **264**(1), 115–119 (1992).
62. R. Djogic, M. Branica, 'Direct Determination of Dissolved Uranyl(VI) in Seawater by Cathodic Stripping Voltammetry', *Analyst*, **120**(7), 1989–1993 (1995).
63. W. Stumm, J.J. Morgan, *Aquatic Chemistry: An Introduction Emphasizing Chemical Equilibria in Natural Waters*, J. Wiley & Sons, New York, 1981.
64. A. Bortoli, M. Gerotto, M. Marchiori, F. Mariconti, M. Palonta, A. Troncon, 'Determination of Cd, Co, Cu, Mn, Ni, Pb, and Zn by Inductively Coupled Plasma Mass Spectroscopy or Flame Atomic Absorption Spectrometry After On-line Preconcentration and Solvent Extraction by Flow Injection System', *Microchem. J.*, **54**(4), 402–411 (1996).
65. F. Baffi, A.M. Cardinale, R. Bruzzone, 'Preconcentration of Chromium, Copper and Manganese from Seawater on Pretreated Solid Materials for Determination by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **270**(1), 79–86 (1992).
66. M.A. Kabil, S.E. Ghazy, M.R. Lasheen, M.A. Shallaby, N.S. Amar, 'Spectrophotometric and Atomic Absorption Determination of Ni(II) in Fresh and Seawaters After Preconcentration by Flotation', *Fresenius' J. Anal. Chem.*, **354**, 371–373 (1996).
67. H. Narasaki, J. Wang, 'Determination of Organotin Compounds in Seawater by Capillary Column Gas Chromatography–Atomic Absorption Spectrometry', *Anal. Sci.*, **14**(4), 857–861 (1998).
68. S. Akman, H. Ince, U. Koklu, 'Determination of Some Trace Elements in Seawater by Atomic Absorption Spectrometry After Concentration with Modified Silicas', *J. Anal. At. Spectrom.*, **7**(2), 187–190 (1992).
69. I.I. Naqvi, Z. Gul, S. Habib, I. Zehra, A.A. Mahmood, 'Determination of Trace Metals in Karachi Coastal Waters Using Atomic Absorption Spectrometry', *J. Chem. Soc. Pak.*, **15**(1), 1–6 (1993).
70. S.C. Pai, S. Huang, K.Y. Shih, G.C. Gong, C.C. Hung, 'Effect of Optical Background Correction Systems on the Direct Determination of Heavy Metals in Seawater Using Flame Atomic Absorption Spectrometry', *Hua Hsueh*, **48**(3), 182–200 (1990).
71. G. Damianelli, D.A. Katskov, P.J.J.G. Marais, P. Tittarelli, 'Characterization of the Vapor-phase Molecular and Atomic Absorption from Seawater Matrices in Electrothermal Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **53**, 945–964 (1998).
72. G. Absalan, C.L. Chakrabarti, K.L. Headrick, 'A Study of NaCl Interference with the Determination of Gold and Iron in Solution Residues by Atomic Absorption Spectrometry Using dc and rf Glow Discharges', *Can. J. Appl. Spectrosc.*, **41**(3), 51–60 (1996).
73. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Usefulness of the Chemical Modification and the Multi-injection Technique Approaches in the Electrothermal Atomic Absorption Spectrometric Determination of Silver, Arsenic, Cadmium, Chromium, Mercury, Nickel and Lead in Seawater', *J. Anal. At. Spectrom.*, **13**(8), 777–786 (1998).
74. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Iridium-coated Graphite Tubes for the Direct Determination of As, Cd, Hg, and Pb in Seawater by Vapor Generation ETAAS (Electrothermal Atomic Absorption Spectrometry)', *At. Spectrosc.*, **19**(3), 100–106 (1998).
75. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Direct Determination of Arsenic in Seawater by Electrothermal Atomization Atomic Absorption Spectrometry Using D2 and Zeeman Background Correction', *Mikrochim. Acta*, **128**(3/4), 215–221 (1998).
76. D. Colbert, K.S. Johnson, K.H. Coale, 'Determination of Cadmium in Seawater Using Automated On-line Preconcentration and Direct Injection Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **377**(2–3), 255–262 (1998).

77. M. Hoenig, E. Puskaric, P. Choisy, M. Wartel, 'Direct Determination of High- and Mid-volatile Metals (Cadmium, Lead, Manganese) in Seawater by Electrothermal Atomic Absorption Spectrometry: Existing Approaches and Critical Parameters', *Analusis*, **19**(9), 285–291 (1991).
78. C.R. Lan, 'Direct Determination of Cadmium in Seawater by Electrothermal Atomic Absorption Spectrometry with Sodium Hydroxide as a Chemical Modifier', *Analyst*, **118**(2), 189–192 (1993).
79. J.Y. Cabon, A. Le Bihan, 'Direct Determination of Cadmium in Seawater Using Electrothermal Atomization Atomic Absorption Spectrometry with Zeeman-effect Background Correction and Oxalic Acid as a Chemical Modifier', *J. Anal. At. Spectrom.*, **7**(2), 383–388 (1992).
80. H. Chuang, S. Huang, 'Direct Determination of Cadmium in Seawater with a Graphite Furnace Atomic Absorption Spectrometer', *Spectrochim. Acta, Part B*, **49**(3), 283–288 (1994).
81. K. Baeckstroem, L.G. Danielsson, 'A Mechanized Continuous Flow System for the Concentration and Determination of Cobalt, Copper, Nickel, Lead, Cadmium and Iron in Seawater Using Graphite Furnace Atomic Absorption', *Mar. Chem.*, **29**(1), 33–46 (1990).
82. J.Y. Cabon, A. Le Bihan, 'The Determination of Cr, Cu and Mn in Seawater with Transversely Heated Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **50**(13), 1703–1716 (1995).
83. L.L. Hsieh, Y.Y. Hung, C.R. Lan, 'Direct Determination of Total Chromium in Seawater by Graphite Furnace Atomic Absorption Spectrometry with Multiple Injections', *J. Chin. Chem. Soc.*, **40**(3), 241–244 (1993).
84. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Chromium Determination in Seawater by Electrothermal Atomic Absorption Spectrometry Using Zeeman Effect Background Correction and a Multi-injection Technique', *Fresenius' J. Anal. Chem.*, **360**(2), 208–212 (1998).
85. S.C. Apte, S.D.W. Comber, M.J. Gardner, A.M. Gunn, 'Direct Determination of Chromium in Estuarine and Coastal Waters by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**(2), 169–172 (1991).
86. Z.S. Liu, S. Huang, 'Determination of Copper and Cadmium in Seawater by Preconcentration and Electrothermal Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **267**(1), 31–37 (1992).
87. Y. Sung, Z. Liu, S. Huang, 'Automated On-line Preconcentration System for Electrothermal Atomic Absorption Spectrometry for the Determination of Copper and Molybdenum in Seawater', *J. Anal. At. Spectrom.*, **12**(8), 841–847 (1997).
88. M. Wang, A.I. Yuzefovsky, R.G. Michel, 'Determination of Ultratrace Amounts of Copper and Cadmium in Seawater by Graphite Furnace Atomic Absorption Spectrometry with Flow Injection Semi On-line Preconcentration', *Microchem. J.*, **48**(3), 326–342 (1993).
89. P. Su, S. Huang, 'Direct and Simultaneous Determination of Copper and Manganese in Seawater with a Multielement Graphite Furnace Atomic Absorption Spectrometer', *Spectrochim. Acta, Part B*, **53**(5), 699–708 (1998).
90. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Comparative Study of Magnesium Nitrate, Palladium Nitrate and Reduced Palladium for the Direct Determination of Mercury in Seawater by Electrothermal Atomization Atomic Absorption Spectrometry', *Mikrochim. Acta*, **124**(1–2), 111–122 (1996).
91. C.R. Lan, Z.B. Alfassi, 'Direct Determination of Manganese in Seawater by Electrothermal Atomic Absorption Spectrometry with Sodium Hydroxide as Chemical Modifier for Interference Removal', *Analyst*, **119**(5), 1033–1035 (1994).
92. A.T. Rangel, R.C. De Campos, I. Moreira, 'Comparative Study of Different Procedures for the Determination of Total Manganese in Coastal Waters by Graphite Furnace Atomic Absorption Spectrometry', *At. Spectrosc.*, **19**(4), 133–136 (1998).
93. X. Shan, B. Radziuk, B. Welz, O. Vyskocilova, 'Determination of Manganese in River and Seawater Samples by Electrothermal Atomic Absorption Spectrometry with a Tungsten Atomizer', *J. Anal. At. Spectrom.*, **8**(3), 409–413 (1993).
94. P. Su, S. Huang, 'Direct and Simultaneous Determination of Molybdenum and Vanadium in Seawater Using a Multielement Electrothermal Atomic Absorption Spectrometer', *J. Anal. At. Spectrom.*, **13**(7), 641–645 (1998).
95. S. Huang, W. Lai, K. Shih, 'Direct Determination of Molybdenum, Chromium and Manganese in Seawater by Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **50**(10), 1237–1246 (1995).
96. Z. Liu, S. Huang, 'Determination of Lead in Seawater with a Graphite Furnace Atomic Absorption Spectrometer and An Improved Automatic On-line Preconcentration System', *Spectrochim. Acta, Part B*, **50**(2), 197–203 (1995).
97. J.Y. Cabon, A. Le Bihan, 'Determination of Lead in Seawater by Electrothermal Atomic Absorption Spectrometry with Transversely Heated Furnace by Using Oxalic Acid Or Pd/Mg as Modifiers', *Spectrochim. Acta, Part B*, **51**(9–10), 1245–1251 (1996).
98. J. Shiowatana, J.P. Matousek, 'Electrodeposition on Pyrolytic Graphite Platforms for Electrothermal Atomic Absorption Spectroscopic Determination of Labile Lead in Saline Water', *Talanta*, **38**(4), 375–383 (1991).
99. P. Bermejo-Barrera, R.M. Soto-Ferreiro, R. Dominguez-Gonzalez, A. Bermejo-Barrera, 'Determination of Tributyltin, Dibutyltin and Total Tin in the Galicia Estuaries by HPLC/ETAAS', *Ann. Chim.*, **86**(9–10), 495–508 (1996).

100. K. Fytianos, V. Samanidou, 'Determination of Trace Quantities of Organotin Compounds in Coastal Waters of Greece by Graphite Furnace Atomic Absorption Spectrometry', *Sci. Total Environ.*, **92**, 265–268 (1990).
101. J.Y. Cabon, A. Le Bihan, 'Direct Determination of Zinc in Seawater Using Electrothermal Atomic Absorption Spectrometry with Zeeman-effect Background Correction: Effects of Chemical and Spectral Interferences', *J. Anal. At. Spectrom.*, **9**(3), 477–481 (1994).
102. S. Huang, K. Shih, 'Direct Determination of Zinc in Seawater by Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **50**(8), 837–846 (1995).
103. J.G. Sen Gupta, J.L. Bouvier, 'Direct Determination of Traces of Ag, Cd, Pb, Bi, Cr, Mn, Co, Ni, Li, Be, Cu and Sb in Environmental Waters and Geological Materials by Simultaneous Multielement Graphite Furnace Atomic Absorption Spectrometry with Zeeman-effect Background Correction', *Talanta*, **42**(2), 269–281 (1995).
104. S. Sachsenberg, T. Klenke, W.E. Krumbein, H.J. Schellnhuber, E. Zeeck, 'Direct Graphite Furnace Atomic Absorption Spectrometric Determination of Metals in Seawater: Application of Palladium Modifiers and a Fractal Approach to their Analytical Support', *Anal. Chim. Acta*, **279**(2), 241–251 (1993).
105. Z. Liu, S. Huang, 'Automatic On-line Preconcentration System for Graphite Furnace Atomic Absorption Spectrometry for the Determination of Trace Metals in Seawater', *Anal. Chim. Acta*, **281**(1), 185–190 (1993).
106. U. Bredthauer, J. Agger, W. Dannecker, 'Determination of Heavy Metal Concentrations in the Trace and Ultratrace Regions in Rain, Seawater, and Fish Liver Sample Matrices by Flow-injection Graphite Furnace Atomic Absorption Spectrometry (FI/GFAAS)', CANAS '95, Colloq. Anal. Atomspektrosk. (1996), Meeting Date1995, 299–305, eds. B. Welz, Bodenseewerk Perkin-Elmer, Ueberlingen, Germany, 1995.
107. J.Y. Cabon, A. Le Bihan, 'Determination of Trace Metals in Rainwater Subject to Marine Influence by Atomic Absorption Spectrometry with Electrothermal Atomization', *Environ. Technol.*, **12**(9), 769–776 (1991).
108. V. Porta, O. Abollino, E. Mentasti, C. Sarzanini, 'Determination of Ultratrace Levels of Metal Ions in Seawater with On-line Preconcentration and Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**(2), 119–122 (1991).
109. A. Lopez Garcia, E. Blanco Gonzalez, A. Sanz-Medel, 'Determination of Trace Elements in Seawater by Electrothermal Atomic Absorption Spectrometry with and without a Preconcentration Step', *Mikrochim. Acta*, **112**(1–4), 19–29 (1993).
110. L.C. Azeredo, R.E. Sturgeon, A.J. Curtius, 'Determination of Trace Metals in Seawater by Graphite Furnace Atomic Absorption Following On-line Separation and Preconcentration', *Spectrochim. Acta, Part B*, **48**(1), 91–98 (1993).
111. C. Arenas, J. Victor, 'Contributions to the Environmental Sample Bank. 7. Development and Use of an Automated Hydride-AAS (Atomic Absorption Spectrometry) Method with Arsine Enrichment for the Determination of Arsenic Compounds in Algae, Mussels and Seawater with HPLC. Investigation of Various Arsenic Compounds in Sea Organisms of the German North and East Sea Coasts and for Comparison in Samples from Antofagasta, Chile', *Forschungszent. Juelich: Ber., Juel*, **2512**, 186 (1991).
112. W.W. Ding, R.E. Sturgeon, 'Evaluation of Electrochemical Hydride Generation for the Determination of Arsenic and Selenium in Seawater by Graphite Furnace Atomic Absorption with in situ Concentration', *Spectrochim. Acta, Part B*, **51**(11), 1325–1334 (1996).
113. H. Matusiewicz, M. Kopras, R.E. Sturgeon, 'Determination of Cadmium in Environmental Samples by Hydride Generation with in situ Concentration and Atomic Absorption Detection', *Analyst*, **122**(4), 331–336 (1997).
114. W.W. Ding, R.E. Sturgeon, 'Evaluation of Electrochemical Hydride Generation for the Determination of Total Antimony in Natural Waters by Electrothermal Atomic Absorption Spectrometry with in situ Concentration', *J. Anal. At. Spectrom.*, **11**(3), 225–230 (1996).
115. J.Y. Cabon, W. Erler, 'Determination of Selenium Species in Seawater by Flow Injection Hydride Generation In Situ Trapping followed by Electrothermal Atomic Absorption Spectrometry', *Analyst*, **123**(7), 1565–1569 (1998).
116. J. Wu, E.A. Boyle, 'Determination of Iron in Seawater by High-resolution Isotope Dilution Inductively Coupled Plasma Mass Spectrometry After $Mg(Oh)_2$ Coprecipitation', *Anal. Chim. Acta*, **367**(1–3), 183–191 (1998).
117. T. Kato, S. Nakamura, M. Morita, 'Determination of Nickel, Copper, Zinc, Silver, Cadmium and Lead in Seawater by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry', *Anal. Sci.*, **6**(4), 623–626 (1990).
118. J. Wu, E.A. Boyle, 'Low Blank Preconcentration Technique for the Determination of Lead, Copper, and Cadmium in Small Volume Seawater Samples by Isotope Dilution ICPMS', *Anal. Chem.*, **69**(13), 2464–2470 (1997).
119. D.W. Hastings, S.R. Emerson, B.K. Nelson, 'Determination of Picogram Quantities of Vanadium in Calcite and Seawater by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry with Electrothermal Vaporization', *Anal. Chem.*, **68**, 371–377 (1996).
120. A. Stroh, U. Voellkopf, 'Optimization and Use of Flow Injection Vapor Generation Inductively Coupled Plasma Mass Spectrometry for the Determination of Arsenic, Antimony and Mercury in Water and Seawater at Ultratrace Levels', *J. Anal. At. Spectrom.*, **8**(1), 35–40 (1993).
121. D. Pozebon, V.L. Dressler, A.J. Curtius, 'Determination of Copper, Cadmium, Lead, Bismuth and Selenium(IV) in Seawater by Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry After On-line

Separation', *J. Anal. At. Spectrom.*, **13**(5), 363–369 (1998).
122. S.N. Willie, Y. Iida, J.W. McLaren, 'Determination of Cu, Ni, Zn, Mn, Co, Pb, Cd, and V in Seawater Using Flow Injection ICP/MS', *At. Spectrosc.*, **19**(3), 67–72 (1998).
123. K. Akatsuka, J.W. McLaren, J.W. Lam, S.S. Berman, 'Determination of Iron and Ten Other Trace Elements in the Open Ocean Seawater Reference Material NASS-3 by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **7**(6), 889–894 (1992).
124. M.J. Bloxham, A. Gachanja, S.J. Hill, P.J. Worsfold, 'Determination of Mercury Species in Seawater by Liquid Chromatography with Inductively Coupled Plasma Mass Spectrometric Detection', *J. Anal. At. Spectrom.*, **11**(2), 145–148 (1996).
125. E. Rosland, W. Lund, 'Direct Determination of Trace Metals in Seawater by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**(11), 1239–1244 (1998).
126. F. Vanhaecke, J. Goosens, R. Dams, C. Vandecasteele, 'The Determination of Molybdenum in a Seawater Candidate Reference Material by Inductively Coupled Plasma/Mass Spectrometry', *Talanta*, **40**(7), 975–979 (1993).
127. L. Halicz, J.W.H. Lam, J.W. McLaren, 'An On-line Method for the Determination of Lead and Lead Isotope Ratios in Fresh and Saline Waters by Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta, Part B*, **49**(7), 637–647 (1994).
128. H. Tao, J.W.H. Lam, J.W. McLaren, 'Determination of Selenium in Marine Certified Reference Materials by Hydride Generation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**(8), 1067–1073 (1993).
129. S.J. Santosa, H. Mokudai, S. Tanaka, 'Automated Continuous-flow Hydride Generation with Inductively Coupled Plasma Mass Spectrometric Detection for the Determination of Trace Amounts of Selenium(IV), and Total Antimony, Arsenic and Germanium in Seawater', *J. Anal. At. Spectrom.*, **12**(4), 409–415 (1997).
130. C. Veillon, in *Trace Analysis, Spectroscopic Methods for the Elements*, ed. J.D. Winefordner, J. Wiley & Sons, New York, Chapter 6, 1976.
131. H. Sawatari, E. Fujimori, H. Haraguchi, 'Multielement Determination of Trace Elements in Seawater by Gallium Coprecipitation and Inductively Coupled Plasma Mass Spectrometry', *Anal. Sci.*, **11**(3), 369–374 (1995).
132. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Analysis of Trace Metals in Seawater by Inductively Coupled Plasma Mass Spectrometry and Related Techniques', *Anal. Proc.*, **30**(3), 159–161 (1993).
133. S.J. Santosa, S. Tanaka, K. Yamanaka, 'Sequential Determination of Trace Metals in Seawater by Inductively Coupled Plasma Mass Spectrometry After Electrothermal Vaporization of their Dithiocarbamate Complexes in Methylisobutyl Ketone', *Environ. Monit. Assess.*, **44**(1–3), 515–528 (1997).
134. S.J. Santosa, S. Tanaka, K. Yamanaka, 'Inductively Coupled Plasma Mass Spectrometry for the Sequential Determination of Trace Metals in Seawater After Electrothermal Vaporization of their Dithiocarbamate Complexes in Methyl Isobutyl Ketone', *Fresenius' J. Anal. Chem.*, **357**(8), 1122–1127 (1997).
135. L. Ebdon, A. Fisher, H. Handley, P. Jones, 'Determination of Trace Metals in Concentrated Brines Using Inductively Coupled Plasma Mass Spectrometry On-line Preconcentration and Matrix Elimination with Flow Injection', *J. Anal. At. Spectrom.*, **8**(7), 977–981 (1993).
136. K.E. Jarvis, J.G. Williams, E. Alcantara, J.D. Wills, 'Determination of Trace and Ultra-trace Elements in Saline Waters by Inductively Coupled Plasma Mass Spectrometry After Off-line Chromatographic Separation and Preconcentration', *J. Anal. At. Spectrom.*, **11**(10), 917–922 (1996).
137. H.E. Allen, R.A. Minear, in *A Reference Handbook Published on Behalf of the World Health Organization Regional Office for Europe*, ed. M.J. Suez, Pergamon Press, Oxford, 43, Vol. 2, 1982.
138. I. Rodushkin, T. Ruth, 'Determination of Trace Metals in Estuarine and Seawater Reference Materials by High Resolution Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**(10), 1181–1185 (1997).
139. G.J. Batterham, N.C. Munksgaard, D.L. Parry, 'Determination of Trace Metals in Seawater by Inductively Coupled Plasma Mass Spectrometry After Off-line Dithiocarbamate Solvent Extraction', *J. Anal. At. Spectrom.*, **12**(11), 1277–1280 (1997).
140. M. Bettinelli, S. Spezia, 'Determination of Trace Elements in Seawater by Ion Chromatography–Inductively Coupled Plasma Mass Spectrometry', *J. Chromatogr., A*, **709**(2), 275–281 (1995).
141. M.L. Wells, K.W. Bruland, 'An Improved Method for Rapid Preconcentration and Determination of Bioactive Trace Metals in Seawater Using Solid-phase Extraction and High Resolution Inductively Coupled Plasma Mass Spectrometry', *Mar. Chem.*, **63**(1–2), 145–153 (1998).
142. J.P. Byrne, G. Chapple, 'Direct Determination of Trace Metals in Seawater by Electrothermal Vaporization ICP/MS with Pd-HNO_3 Modifier', *At. Spectrosc.*, **19**(4), 116–120 (1998).
143. H. Falk, R. Geerling, B. Hattendorf, K. Krengel-Rothensee, K.P. Schmidt, 'Capabilities and Limits of ICP/MS for Direct Determination of Element Traces in Saline Solutions', *Fresenius' J. Anal. Chem.*, **359**(4–5), 352–356 (1997).
144. S.N. Willie, H. Tekgul, R.E. Sturgeon, 'Immobilization of 8-Hydroxyquinoline Onto Silicone Tubing for the Determination of Trace Elements in Seawater Using Flow Injection ICP/MS', *Talanta*, **47**(2), 439–445 (1998).

145. A. Hioki, J.W.H. Lam, J.W. McLaren, 'On-line Determination of Dissolved Silica in Seawater by Ion Exclusion Chromatography in Combination with Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **69**, 21–24 (1997).
146. H. Sawatari, T. Hayashi, E. Fujimori, A. Hirose, H. Haraguchi, 'Multielement Determination of Trace Elements in Coastal Seawater by ICP/MS and ICP/AES After Aluminum Coprecipitation Associated with Magnesium', *Bull. Chem. Soc. Jpn.*, **69**(7), 1925–1931 (1996).
147. X.J. Li, P. Schramel, H.Z. Wang, W.P. Grill, A. Kettrup, 'Determination of Trace Metal Ions Co, Cu, Mo, Mn, Fe, Ti, V in Reference River Water and Reference Seawater Samples by Inductively Coupled Plasma Emission Spectrometry Combined with the Third Phase Preconcentration', *Fresenius' J. Anal. Chem.*, **356**(1), 52–56 (1996).
148. P.C. Rudner, J.M.C. Pavon, F.S. Rojas, A.G. de Torres, 'Use of Flow Injection Cold Vapor Generation and Preconcentration on Silica Functionalized with Methyl Thiosalicylate for the Determination of Mercury in Biological Samples and Seawater by Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **13**(10), 1167–1171 (1998).
149. S. Birghila, 'Determination of Trace Elements in a Marine Medium by Inductively Coupled Plasma Atomic Emission Spectrometry (ICP/AES)', *Rev. Chim.*, **48**(2), 152–153 (1997).
150. H. Tsuji, N. Awano, Y. Tamari, K. Chayama, K. Teranishi, K. Isomura, 'Determination of Trace Elements in Arima Spring Water by ICP/AES After Coprecipitation with Iron(III) Hydroxide', *Bunseki Kagaku*, **44**(6), 471–476 (1995).
151. A. Mazzucotelli, P. Rivaro, 'Determination of Trace Amounts of Metallorganics in Marine Ecosystem by High-pressure Liquid Chromatography Followed by Inductively Coupled Plasma Atomic Emission Spectrometric Detection', *Microchem. J.*, **51**(1/2), 231–237 (1995).
152. I. Gornushkin, B.W. Smith, J.D. Winefordner, 'Use of Laser-exited Atomic Fluorescence Spectrometry with a Novel Diffusive Graphite Tube Electrothermal Atomizer for the Direct Determination of Silver in Seawater and in Solid Reference Materials', *Spectrochim. Acta, Part B*, **51**, 1355–1370 (1996).
153. J. Moreda-Pineiro, M.L. Cervera, M. De La Guardia, 'Direct Determination of Arsenic in Seawater by Continuous-flow Hydride Generation Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **12**(12), 1377–1380 (1997).
154. A.M. Featherstone, E.C.V. Butler, B.V. O'Grady, P. Michel, 'Determination of Arsenic Species in Seawater by Hydride Generation Atomic Fluorescence Spectroscopy', *J. Anal. At. Spectrom.*, **13**(12), 1355–1360 (1998).
155. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Determination of Mercury in Filtered Seawater by Flow Injection with On-line Oxidation and Atomic Fluorescence Spectrometric Detection', *J. Anal. At. Spectrom.*, **11**(7), 511–514 (1996).
156. V. Cheam, J. Lechner, I. Sekerka, R. Desrosiers, 'Direct Determination of Lead in Seawaters by Laser-excited Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **9**(3), 315–320 (1994).
157. A. D'Ulivo, L. Lampugnani, R. Zamboni, 'Combined Electrothermally Heated Quartz Tube – Flame Atomizer for the Hydride Generation Non Dispersive Atomic Fluorescence Determination of Selenium', *J. Anal. At. Spectrom.*, **5**(3), 225–229 (1990).
158. Y. He, H. El Azouzi, M.L. Cervera, M. de la Guardia, 'Completely Integrated On-line Determination of Dissolved Selenium(IV) and Total Inorganic Selenium in Seawater by Flow Injection Hydride Generation Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **13**(11), 1291–1296 (1998).
159. E. Haffer, D. Schmidt, P. Freimann, W. Gerwinski, 'Simultaneous Determination of Germanium, Arsenic, Tin and Antimony with Total-reflection X-ray Fluorescence Spectrometry Using the Hydride Generation Technique for Matrix Separation – First Steps in the Development of a New Application', *Spectrochim. Acta, Part B*, **52**(7), 935–944 (1997).
160. P. Freimann, D. Schmidt, 'Application of Total Reflection X-ray Fluorescence Analysis for the Determination of Trace Metals in the North Sea', *Spectrochim. Acta, Part B*, **44**, 505–510 (1989).
161. M. Haarich, D. Schmidt, P. Freimann, A. Jacobsen, 'North Sea Research Projects ZISCH and PRISMA: Application of Total-reflection X-ray Spectrometry in Seawater Analysis', *Spectrochim. Acta, Part B*, **48**, 183–192 (1993).
162. P. Freimann, D. Schmidt, A. Neubauer-Ziebarth, 'Reference Materials for Quality Assurance in Seawater Analysis: Performance of Total Reflection X-ray Fluorescence in the Intercomparison and Certification Stages', *Spectrochim. Acta, Part B*, **48**, 193–198 (1993).
163. A.I. Saleh, S.W. Remail, F.M. Milad, 'Determination of Cadmium in Water Samples by Coprecipitation and Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **168**(1), 23–27 (1993).
164. G.R. Doshi, S.N. Joshi, K.C. Pillai, 'Determination of Mercury in Environmental Samples by Neutron Activation', *Indian J. Environ. Health*, **33**(4), 492–497 (1991).
165. A.I. Saleh, J.R.W. Woittiez, G.D. Wals, H.A. Das, 'Determination of Selenium(IV) in Water by Neutron Activation Analysis After Coprecipitation of Metal Dibenzyl Dithiocarbamate with Phenolphthalein', *J. Radioanal. Nucl. Chem.*, **140**(2), 357–363 (1990).
166. Y. Sakai, K. Tomura, K. Ohshita, 'Determination of Selenium(IV) and Selenium(VI) in Natural Water

Samples by Neutron Activation Analysis After Chemical Precollection', *J. Radioanal. Nucl. Chem.*, **187**(6), 441–450 (1994).
167. A. Sánchez Misiego, R. Garcìa Monco-Carrá, A. Zirino, 'Variability of the Slopes of Standard Addition Plots in the Analysis of Copper in Seawater by Anodic Stripping Voltammetry', *Electroanalysis*, **8**(6), 534–538 (1996).
168. Y. Sun, J. Mierzwa, M. Yang, 'New Method of Gold-film Electrode Preparation for Anodic Stripping Voltammetric Determination of Arsenic(III and V) in Seawater', *Talanta*, **44**(8), 1379–1387 (1997).
169. A. Sánchez-Misiego, R. Garcìa-Monco Carrá, A. Zirino, 'Variability of the Slopes of Standard Addition Plots in the Analysis of Copper in Seawater by Anodic Stripping Voltammetry', *Electroanalysis*, **8**(6), 534–538 (1996).
170. C. Bing, L. Kryger, 'Optimization of Square Wave Stripping Voltammetry at a Rotating Thin Film Mercury Electrode for the Determination of Traces of Lead and Copper in Seawater', *Bull. Singapore Natl. Inst. Chem.*, **23**, 91–98 (1995).
171. H. Keis, U. Piibar, 'Determination of Heavy Metals in Sea- and Freshwaters by Anodic Stripping Voltammetry', *Tartu Ulik. Toim.*, **975**, 119–127 (1994).
172. J. Zima, C.M.G. Van den Berg, 'Determination of Arsenic in Seawater by Cathodic Stripping Voltammetry in the Presence of Pyrrolidine Dithiocarbamate', *Anal. Chim. Acta*, **289**(3), 291–298 (1994).
173. H. Li, R.B. Smart, 'Determination of Sub-nanomolar Concentration of Arsenic(III) in Natural Waters by Square Wave Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **325**(1–2), 25–32 (1996).
174. M. Khodari, 'Determination of Cadmium in Seawater by Cathodic Stripping Voltammetry of Complexes with 5-Fluorouracil', *Electroanalysis*, **10**(15), 1061–1063 (1998).
175. J.A. Herrera-Melian, J. Hernandez-Brito, M.D. Gelado-Caballero, J. Perez-Pena, 'Direct Determination of Cobalt in Unpurged Oceanic Seawater by High Speed Adsorptive Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **299**(1), 59–67 (1994).
176. M. Vega, C.M.G. van der Berg, 'Determination of Cobalt in Seawater by Catalytic Adsorptive Cathodic Stripping Voltammetry', *Anal. Chem.*, **69**, 874–881 (1997).
177. M. Paneli, A.V. Voulgaropoulos, K. Kalcher, 'Catalytic Adsorptive Stripping Voltammetric Determination of Chromium with TTHA and Nitrate', *Mikrochim. Acta*, **110**(4–6), 205–215 (1993).
178. C. Collado-Sanchez, J. Perez-Pena, M.D. Gelado-Caballero, J.A. Herrera-Melian, J.J. Hernandez-Brito, 'Rapid Determination of Copper, Lead and Cadmium in Unpurged Seawater by Adsorptive Stripping Voltammetry', *Anal. Chim. Acta*, **320**(1), 19–30 (1996).
179. C.M.G. Van den Berg, M. Nimmo, O. Abollino, E. Mentasti, 'The Determination of Trace Levels of Iron in Seawater Using Adsorptive Cathodic Stripping Voltammetry', *Electroanalysis*, **3**(6), 477–484 (1991).
180. M. Czae, K. Kim, Y. Kwon, 'Determination of Iron in Seawater by Adsorptive Stripping Voltammetry', *J. Korean Chem. Soc.*, **39**(3), 186–190 (1995).
181. M.D. Gelado-Caballero, J.J. Hernández-Brito, J.A. Herrera-Melián, C. Collado-Sánchez, J. Pérez-Peña, 'Fast Adsorptive Stripping Voltammetry of Iron in Oxygenated Seawater', *Electroanalysis*, **8**(11), 1065–1071 (1996).
182. K. Yokoi, C.M.G. Van den Berg, 'The Determination of Iron in Seawater Using Catalytic Cathodic Stripping Voltammetry', *Electroanalysis*, **4**(1), 65–69 (1992).
183. J.S. Roitz, K.W. Bruland, 'Determination of Dissolved Manganese(II) in Coastal and Estuarine Waters by Differential Pulse Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **344**(3), 175–180 (1997).
184. A.L.B. Marques, G.O. Chierice, 'Trace Nickel Determination with Phenyldithiocarbamate in Seawater by Adsorptive Stripping Voltammetry', *J. Braz. Chem. Soc.*, **4**(1), 17–19 (1994).
185. Q. Wu, G.E. Batley, 'Determination of Sub-nanomolar Concentrations of Lead in Seawater by Adsorptive Stripping Voltammetry with Xylenol Orange', *Anal. Chim. Acta*, **309**(1–3), 95–101 (1995).
186. C.M.G. Van den Berg, S.H. Khan, 'Determination of Selenium in Seawater by Adsorptive Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **231**(2), 221–229 (1990).
187. A. Romanus, H. Muller, D. Kirsch, 'Application of Adsorptive Stripping Voltammetry for the Determination of Trace Metals in Brine. Part 1. Batch Voltammetric Measurements', *Fresenius' J. Anal. Chem.*, **340**(6), 363–370 (1991).
188. A. Romanus, H. Mueller, D. Kirsch, 'Application of Adsorptive Stripping Voltammetry for the Determination of Trace Metals in Brine. Part 2. Development and Evaluation of a Flow Injection System', *Fresenius' J. Anal. Chem.*, **340**(6), 371–376 (1991).
189. C. Colombo, C.M.G. Van den Berg, 'Simultaneous Determination of Several Trace Metals in Seawater Using Cathodic Stripping Voltammetry with Mixed Ligands', *Anal. Chim. Acta*, **337**(1), 29–40 (1997).
190. R.N. Khandekar, R. Raghunath, 'Adsorptive Differential Pulse Voltammetry for Determination of Uranium in Seawater', *Bull. Electrochem.*, **6**(9), 799–801 (1990).
191. G.R. Scollary, G.N. Chen, T.J. Cardwell, V.A. Vincente-Beckett, 'Rapid Determination of Manganese in Wastewaters and Seawater by Stripping Potentiometry', *Electroanalysis*, **7**(4), 386–389 (1995).
192. G.R. Scollary, T.J. Cardwell, R.W. Cattrall, C.G. Nan, V.A. Vicente-Beckett, I.C. Hamilton, S. Roden, 'Elimination of Zinc/Copper Intermetallic Interference in the Determination of Zinc in Wastewaters and Seawater by Potentiometric Stripping Analysis', *Electroanalysis*, **5**(8), 685–689 (1993).
193. G. Scarano, C. Romei, A. Seritti, A. Zirino, 'Ethylenediamine in the Voltammetric Determination of Copper in

Seawater at a Rotating Mercury Film Electrode', *Anal. Chim. Acta*, **245**(2), 177–81 (1991).
194. P. Ugo, L.M. Moretto, G.A. Mazzocchin, 'Voltammetric Determination of Trace Mercury in Chloride Media at Glassy Carbon Electrodes Modified with Polycationic Ionomers', *Anal. Chim. Acta*, **305**(1–3), 74–82 (1995).
195. C. Locatelli, T. Garai, F. Fagioli, 'Critical Comparison of Alternating Current and Differential Pulse Voltammetry in the Determination of Some Heavy Metals in Seawater', *Mikrochim. Acta*, **107**(1–2), 41–48 (1992).
196. E.P. Achterberg, C.M.G. Van den Berg, 'Automated Voltammetric System for Shipboard Determination of Metal Speciation in Seawater', *Anal. Chim. Acta*, **284**(3), 463–471 (1993).
197. F.L.L. Muller, D.R. Kester, 'Voltammetric Determination of the Complexation Parameters of Zinc in Marine and Estuarine Waters', *Mar. Chem.*, **33**(1–2), 71–90 (1991).
198. J. Atienza, M.A. Herrero, A. Maquieira, R. Puchades, 'Flow Injection Analysis of Seawater. Part II. Cationic Species', *Crit. Rev. Anal. Chem.*, **23**(1), 1–14 (1992).
199. S. Hirata, Y. Hashimoto, M. Aihara, G.V. Mallika, 'On-line Preconcentration for the Determination of Cobalt in Seawater by Flow Injection Chemiluminiscence Detection', *Fresenius' J. Anal. Chem.*, **355**, 676–679 (1996).
200. K.H. Coale, K.S. Johnson, P.M. Stout, C.M. Sakamoto, 'Determination of Copper in Seawater Using a Flow Injection Method with Chemiluminiscence Detection', *Anal. Chim. Acta*, **266**, 345–351 (1992).
201. A.R. Bowie, E.P. Achterberg, R.F.C. Mantoura, P.J. Worsfold, 'Determination of Sub-nanomolar Levels of Iron in Seawater Using Flow Injection with Chemiluminescence Detection', *Anal. Chim. Acta*, **361**(3), 189–200 (1998).
202. V.A. Elrod, K.S. Johnson, K.H. Coale, 'Determination of Subnanomolar Levels of Iron(II) and Total Dissolved Iron in Seawater by Flow Injection and Analysis with Chemiluminescence Detection', *Anal. Chem.*, **63**(9), 893–898 (1991).
203. D.W. O'Sullivan, A.K. Hanson, Jr, D.R. Kester, 'Stopped Flow Luminol Chemiluminescence Determination of Fe(II) and Reducible Iron in Seawater at Subnanomolar Levels', *Mar. Chem.*, **49**(1), 65–77 (1995).
204. T.P. Chapin, K.S. Johnson, K.H. Coale, 'Rapid Determination of Manganese in Seawater by Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **249**(2), 469–478 (1991).
205. K. Okamura, T. Gamo, H. Obata, E. Nakayama, H. Karatani, Y. Nozaki, 'Selective and Sensitive Determination of Trace Manganese in Seawater by Flow Through Technique Using Luminol–hydrogen Peroxide Chemiluminescence Detection', *Anal. Chim. Acta*, **377**(2–3), 125–131 (1998).
206. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Use of Flow Injection Cold Vapor Generation and Preconcentration on Coated Graphite Tubes for the Determination of Cadmium in Seawater by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **11**(11), 1081–1086 (1996).
207. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Use of Flow Injection Cold Vapor Generation and Preconcentration on Coated Graphite Tubes for the Determination of Mercury in Polluted Seawaters by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **12**(3), 317–321 (1997).
208. P. Su, S. Huang, 'Use of 4-(2-Pyridylazo)resorcinol or 2-(2-Pyridylazo)-5-dimethylamino Phenol as Chelating Agent for Determination of Cadmium in Seawater by Atomic Absorption Spectrometry with On-line Flow Injection Sorbent Extraction', *Anal. Chim. Acta*, **376**(3), 305–311 (1998).
209. J.L. Nowicki, K.S. Johnson, K.H. Coale, V.A. Elrod, S.H. Lieberman, 'Determination of Zinc in Seawater Using Flow Injection Analysis with Fluorometric Detection', *Anal. Chem.*, **66**(17), 2732–2738 (1994).
210. S.N. Willie, Y. Iida, J.W. McLaren, 'Determination of Cu, Ni, Zn, Mn, Co, Pb, Cd, and V in Seawater Using Flow Injection ICP/MS', *At. Spectrosc.*, **19**(3), 67–72 (1998).
211. A. Malahoff, I.Y. Kolotyrkina, L.K. Shpigun, 'Shipboard Determination of Dissolved Cobalt in Seawater Using Flow Injection with Catalytic Spectrophotometric Detection', *Analyst*, **121**(8), 1037–1041 (1966).
212. I.Y. Kolotyrkina, L.K. Shpigun, Y.A. Zolotov, A. Malahoff, 'Application of Flow Injection Spectrophotometry to the Determination of Dissolved Iron in Seawater', *Analyst (Cambridge, UK)*, **120**(1), 201–206 (1995).
213. C.I. Measures, J. Yuan, J.A. Resing, 'Determination of Iron in Seawater by Flow Injection Analysis Using In-line Preconcentration and Spectrophotometric Detection', *Mar. Chem.*, **50**(1–4), 3–12 (1995).
214. J.A. Resing, M.J. Mottl, 'Determination of Manganese in Seawater Using Flow Injection Analysis with On-line Preconcentration and Spectrophotometric Detection', *Anal. Chem.*, **64**(22), 2682–2687 (1992).
215. I.Y. Kolotyrkina, L.K. Shpigun, Y.A. Zolotov, G.I. Tsysin, 'Shipboard Flow Injection Method for the Determination of Manganese in Seawater Using In-valve Preconcentration and Catalytic Spectrophotometric Detection', *Analyst (London)*, **116**(7), 707–710 (1991).
216. L.J. Mallini, A.M. Shiller, 'Determination of Dissolved Manganese in Seawater by Flow Injection Analysis with Colorimetric Detection', *Limnol. Oceanogr.*, **38**(6), 1290–1295 (1993).

Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples

Hisatake Narasaki
Saitama University, Shimo-Okubo, Urawa, Japan

1 Historical Review of Hydride Generation Spectroscopic Analysis 1
2 Hydride Generation for Spectroscopic Analysis 2
2.1 Hydride Generation for Atomic Absorption Spectrometry 2
2.2 Hydride Generation for Inductively Coupled Plasma Atomic Emission Spectrometry 5
2.3 Hydride Generation for Inductively Coupled Plasma Mass Spectrometry 6
Acknowledgments 9
Abbreviations and Acronyms 9
Related Articles 9
References 9

The hydride generation (HG) technique using sodium tetrahydroborate(III) (SBH) is applied to atomic absorption spectrometry (AAS), inductively coupled plasma atomic emission spectrometry (ICP/AES), and inductively coupled plasma mass spectrometry (ICP/MS). All the detection limits using the HG technique are less than those for solution nebulization. Chemical speciation in the HG technique is effected by changing the pH of the solution, the cold trap method and various hyphenated techniques. The author used a gas–liquid separator containing a spiral tube to separate efficiently the hydrides from solution. This minimized the consumption of the sample and SBH solutions and of the hydrogen evolution compared with the continuous-flow method. The author used a high-resolution ICP/MS instrument which allowed the use of hydrochloric acid in the sample preparation of arsenic and the determination of selenium without its isotopes.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 HISTORICAL REVIEW OF HYDRIDE GENERATION SPECTROSCOPIC ANALYSIS

Since AAS was founded by Walsh,[1] most metals have been determined rapidly with high sensitivity. However, the sensitivity of metalloid elements such as arsenic (CAS 7440-38-2) and selenium (7782-49-2) were low in AAS, because these elements are difficult to atomize directly. In 1969, Holak[2] generated arsine (7784-42-1) by reaction of zinc (7440-66-6) with hydrochloric acid (10035-10-6) and collected it in a U-tube in liquid nitrogen (7727-37-9). The hydride was then warmed and passed into a flame in an atomic absorption spectrometer. In his method, arsine was separated from solution containing several interferents and determined with high sensitivity. As shown in Equations (1–4) arsenate (15584-04-0) is reduced to arsenite (15502-74-6) and then reduced to arsine, which is decomposed to arsenic (7440-38-2) in the flame.

$$Zn + 2HCl \longrightarrow ZnCl_2 + 2H \quad (1)$$

$$H_3AsO_4 + 2H \longrightarrow H_3AsO_3 + 3H_2O \quad (2)$$

$$H_3AsO_3 + 6H \longrightarrow AsH_3 + 3H_2O \quad (3)$$

$$AsH_3 \longrightarrow As + 3H \quad (4)$$

Schmidt and Royer[3] reduced the metalloids with SBH (16940-66-2) instead of zinc. By using SBH with acid the reduction yield is increased, the time required is shortened and the reagent blank is lowered. The reaction schemes for arsenic and selenium are shown in Equations (5) and (6). Arsenate is reduced to arsenite and then reduced to arsine, which is decomposed to arsenic thermally.

$$4H_3AsO_4 + BH_4^- + H^+ \longrightarrow 4H_3AsO_3 + H_3BO_3 + H_2O \quad (5)$$

$$4H_3AsO_3 + 3BH_4^- + 3H^+ \longrightarrow 4AsH_3 + 3H_3BO_3 + 3H_2O \quad (6)$$

Selenite (14124-67-5) is reduced to hydrogen selenide (7783-07-5), which is decomposed to selenium, as shown in Equations (7) and (8).

$$4H_3SeO_3 + 3BH_4^- + 3H^+ \longrightarrow 4H_2Se + 3H_3BO_3 + 3H_2O \quad (7)$$

$$H_2Se \longrightarrow Se + 2H \quad (8)$$

Elements that form gaseous hydrides (12184-88-2) at ambient temperature are called generically hydride-forming elements. Table 1 summarizes the properties of hydride-forming elements.[4]

Table 1 Properties of hydride-forming elements

Analyte	Hydride		Melting point (°C)	Boiling point (°C)	Solubility in water ($\mu g\,mL^{-1}$)
	Formula	Name			
As	AsH_3 (7784-42-1)	Arsine	−116.3	−62.4	696
Bi	BiH_3 (18288-22-7)	Bismuthine	−67	16.8	–
Ge	GeH_4 (7782-65-2)	Germane	−164.8	−88.1	Insoluble
Pb	PbH_4 (15875-18-0)	Plumbane	−135	−13	–
Sb	SbH_3 (7803-52-3)	Stibine	−88	−18.4	4100
Se	H_2Se (7783-07-5)	Hydrogen selenide	−65.7	−41.3	37 700–68 000
Sn	SnH_4 (2406-52-2)	Stannane	−146	−52.5	–
Te	H_2Te (7783-09-7)	Hydrogen telluride	−51	−4	Very soluble

Reproduced by permission from J. Dedina, D.L. Tsalev, *Hydride Generation Atomic Absorption Spectrometry*, Wiley, Chichester, 1995, p. 19. Copyright John Wiley & Sons Limited.

ICP/AES was established between 1964 and 1965 by Greenfield et al.[5] and Wendt and Fassel.[6] Thompson et al.[7] first applied the HG technique to ICP/AES and determined simultaneously arsenic, antimony (7440-36-0), bismuth (7440-69-9), selenium and tellurium (13494-80-9). Houk et al.[8] combined the inductively coupled plasma (ICP) with mass spectrometry (MS) and constructed an ICP mass spectrometer. Powell et al.[9] determined arsenic, selenium, antimony, bismuth, tellurium and mercury (7439-97-6) by ICP/MS using a commercial HG system. There are several review articles on HG spectroscopic analysis. Nakahara reviewed HG/AAS[10] and HG/ICP/AES,[11] and Campbell reviewed HG/AAS and HG/ICP/AES.[12]

2 HYDRIDE GENERATION FOR SPECTROSCOPIC ANALYSIS

2.1 Hydride Generation for Atomic Absorption Spectrometry

2.1.1 Apparatus for Hydride Generation Atomic Absorption Spectrometry

There are several methods to separate the hydrides and to sweep them into a quartz cell in an atomic absorption spectrometer. These include the cold trap method, the batch method and the continuous-flow method. In the cold trap method, the hydrides are trapped in a U-tube in liquid nitrogen and then released thermally into a heated cell in an atomic absorption spectrometer.[2] The cold trap method is used for chemical speciation of the hydride-forming elements, since the sensitivity can be enhanced by enrichment of the hydrides in the cell. In the batch method, the hydrides are purged with carrier gas to the spectrometer as shown in Figure 1.[13] In the continuous-flow method, the hydrides are separated from solution and swept continuously into the cell. Since the former procedure is tedious, the latter method is used for the total determination.

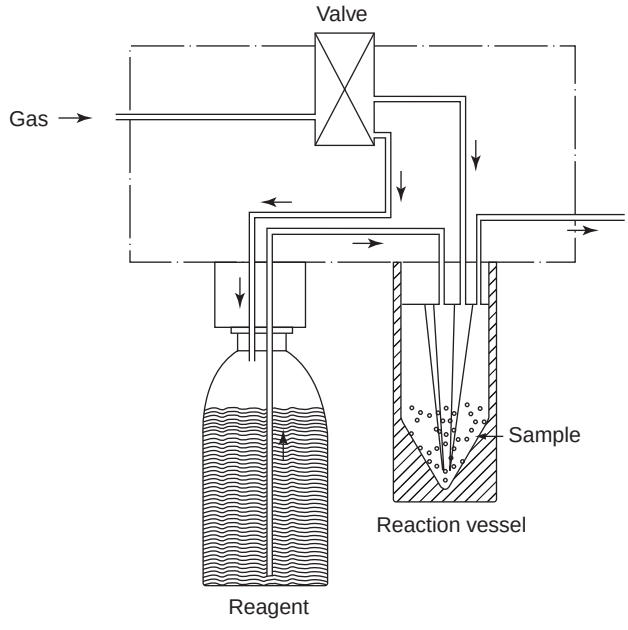

Figure 1 Schematic diagram of the batch method. (Reprinted from B. Welz, M. Melcher, *Anal. Chim. Acta*, **131**, 17–25 (1981). Copyright 1981, with permission from Elsevier Science.[13])

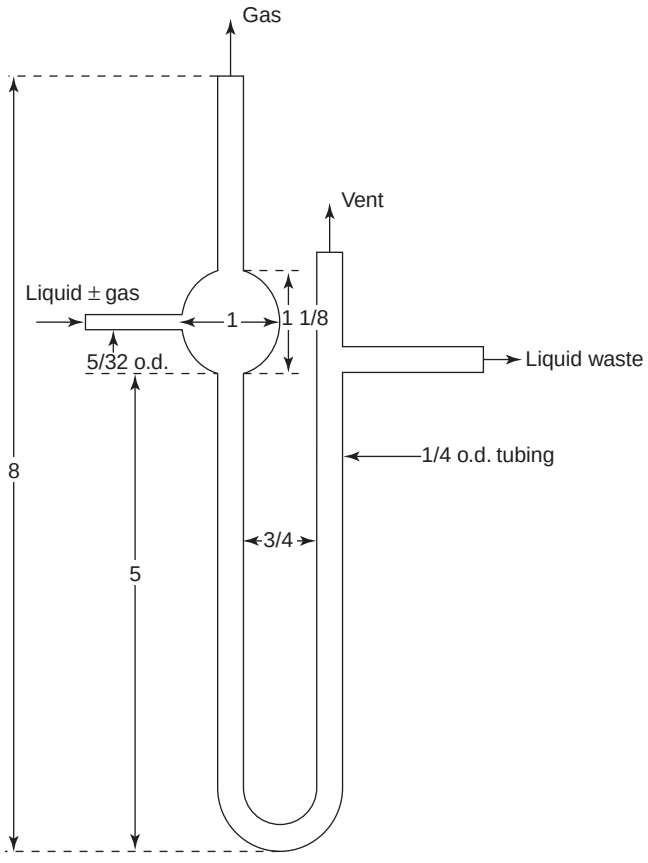

Figure 2 Schematic diagram of the gas–liquid separator. Dimensions in inches. (Reproduced from P.N. Vijan, G.R. Wood, *At. Abs. Newsl.*, **13**(2), 33–37 (1974) by permission of The Perkin-Elmer Corporation.[14])

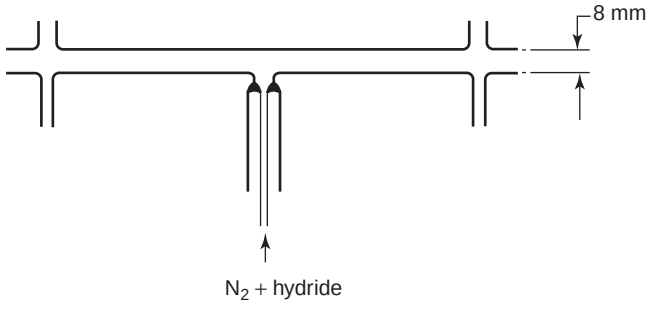

Figure 3 Schematic diagram of the fused-silica furnace. (Reproduced from K.C. Thompson, D. R. Thomerson, *Analyst*, **99**, 595–601 (1974) by permission of The Royal Society of Chemistry.[15])

Figures 2 and 3 show typical dimensions of a gas–liquid separator[14] and a quartz cell,[15] respectively. The sample and SBH solutions are separately conveyed with each peristaltic pump and mixed with a mixing coil. Tygon R-3603 (Norton) tubings are the best for the conveyance of gases and liquids. The hydrides formed are separated from solution in the separator and then swept into the cell. Usually nitrogen or argon is used as the carrier gas. The separator and tubes should be replaced with new ones after several uses, since some hydrides such as hydrogen selenide and plumbane are unstable and deteriorate with contaminants. The cell is heated electrically or with a combustible gas. The electric heating controls the cell temperature. The inner wall of the cell should be cleaned with sandpaper or mineral acids after several uses.

A graphite furnace in which the hydrides are trapped and atomized in situ has also become popular.[16]

2.1.2 Detection Limits and Ranges

Table 2 shows the detection limits of the hydride-forming elements using AAS, ICP/AES and ICP/MS.[17] The detection limits with HG are lower than those obtained by solution nebulization in each type of spectrometry and the limits in AAS lie at the parts per billion level. Since the ranges in AAS are regulated by the Lambert–Beer law, the ranges are narrow. Therefore, sample solutions must be concentrated or diluted depending upon the concentration.

2.1.3 Interferences

The HG process is retarded with transition metals such as copper (7440-50-8) and nickel (7440-02-0)[18] and the other coexisting hydride-forming elements.[13] The transition metals are eliminated with Chelex 100 chelating resin for the determination of arsenic, selenium and antimony.[19] Various masking agents are added to complex the transition metals, as shown in Table 3.

2.1.4 Environmental Sample Preparation for Hydride Generation Atomic Absorption Spectrometry

When the total amount of hydride-forming elements is to be determined, the sample should be decomposed. River waters are concentrated with nitric acid (7697-32-2) and sulfuric acid (7664-93-9) including boiling tips. Organic materials are digested with nitric, sulfuric and perchloric acid (7601-90-3). Siliceous residues deposited during the digestion are dissolved with hydrofluoric acid (7664-39-3).[19] Since selenium and antimony are oxidized during the digestion, selenate is reduced to hydride, forming selenite, by boiling with concentrated hydrochloric acid before hydridization (Equation 9).

$$SeO_3^{2-} + 2HCl \longrightarrow SeO_3^{2-} + Cl_2 + H_2O \quad (9)$$

Antimonate (20175-28-4) is reduced to antimonite with potassium iodide (7681-11-0) solution (Equation 10).

$$SbO_4^{3-} + 3I^- + 2H^+ \longrightarrow SbO_3^{3-} + I_3^- + H_2O \quad (10)$$

Table 2 Detection limits (ng mL^{-1}) for the hydride-forming elements

Analyte	FAAS		ICP/AES		ICP/MS	
	Solution nebulization	Hydride generation	Solution nebulization	Hydride generation	Solution nebulization	Hydride generation
As	630	0.8	40	0.02	0.031	0.0015
Bi	44	0.2	50	0.3	0.004	0.0007
Ge	20	3.8	150	0.2	0.013	0.00001
Pb	17	0.6	8	1.0	0.01	0.01
Sb	60	0.5	200	0.08	0.012	0.00003
Se	230	1.8	30	0.03	0.37	0.001
Sn	150	0.5	300	0.05	0.01	0.0009
Te	44	1.5	80	0.7	0.032	0.001

[a] FAAS, Flame atomic absorption spectrometry. (Revised by permission of JAI Press from T. Nakahara, *Advances in Atomic Spectroscopy*, JAI Press, Stanford, CT, 156 ,Vol. 2, 1995.)

Table 3 Masking agents in HG/AAS

Element	Masking agent	Interfering element	Ref.
As	1,10-Phenanthroline (66-71-7)	Cu(II), Ni(II), Pt(IV), Pd(II)	20
As	Thiosemicarbazide (79-16-6)	Cu(II), Ni(II), Pt(IV), Pd(II)	20
As	Thiourea (62-56-6)	Co(II), Cu(II), Ni(II)	21
As	Cysteine (52-90-4)	Co(II), Ni(II)	22
Sb	Potassium iodide (7681-11-0)	Ag(I), Cr(III), Cu(II), Sn(II)	23
Ge	Cysteine	Co(II), Ni(II), Cu(II)	24

Lead(II) (14280-50-3) should be oxidized to hydride forming lead(IV) (15158-12-0) with potassium peroxodisulfate (7727-21-1) or potassium dichromate (7778-50-9) just before hydridization.

SBH powder should be weighed out as quickly as possible and away from heat sources, since it is hygroscopic and inflammable. SBH solutions must be stabilized with small amounts of sodium hydroxide (1310-73-2) or potassium hydroxide (1310-58-3) and filtered with a filter paper before use.[25] The most commonly used acid in HG is hydrochloric acid, but sulfuric acid is used for the determination of antimony and tin (7440-31-5).

2.1.5 Chemical Speciation by Hydride Generation Atomic Absorption Spectrometry

HG depends upon the oxidation state of the hydride-forming elements and the acidity of a solution containing the elements. Inorganic nonhydride-forming species are determined differentially by subtracting the concentration of the relevant hydride-forming species from the total concentration. Agett and Aspell[26] determined arsenic(III) by HG, maintaining the pH between 4 and 5, and total arsenic by evolution from 5 M hydrochloric acid by AAS. Yamamoto et al. determined antimony(III) at pH 2 and total antimony with 1 M hydrochloric acid.[27]

Organic hydride-forming species are difficult to determine by controlling the acidity. As for arsenic, arsenite forms arsine, methylarsonate forms monomethylarsine $[(CH_3)AsH_3]$, dimethylarsinate forms dimethylarsine $[(CH_3)_2AsH_2]$ and phenylarsonate forms phenylarsine $[(C_6H_5)AsH_2]$. Hydrides trapped in the U-tube are released sequentially on standing at ambient temperature and atomized with an oxyhydrogen flame. The cold trapping methods used for chemical speciation are summarized in Table 4.

Chromatographic methods are the most reliable techniques for the separation of chemical species and are coupled to spectroscopic analyses to enhance the sensitivity, but this article is restricted to methods using HG. The cold trapping method is coupled with gas chromatography (GC) to refine its resolution. Andreae[34] speciated arsenite, arsenate, monomethylarsonate and dimethylarsinate by cold trapping GC/AAS after hydridization. HG/GC/AAS is exclusively used for the speciation of organotin compounds.[35] Since organotin chlorides are liable to adsorb on the packing materials in GC columns, the chlorides are hydridized with SBH before injection. Effluents from the GC columns are ignited with an oxyhydrogen flame to atomize tin in the effluents. Peaks obtained with capillary columns are better resolved than those obtained with packed columns, but the injection volumes with capillary columns are limited. Samples to be speciated by GC are extracted with organic solvents such as hexane (110-54-3) and purified by cleanup. GC cannot be used to speciate thermally unstable hydrides and nonvolatile species.

Table 4 Chemical speciation by the cold trapping method in AAS

Element	Chemical species	Ref.
As	As(III), As(V), monomethylarsine (593-52-2), dimethylarsine (593-57-7), monoethylarsine (593-59-9)	28
Se	Se(IV), Se(VI), dimethyl selenide (593-79-3), dimethyl diselenide (7101-31-7)	29
Sb	Sb(III), Sb(V), methylstibine, dimethylstibine	30
Sn	Stannane (2406-52-2), monomethyltin hydride (1631-78-3), dimethyltin hydride (2067-76-7), trimethyltin hydride (1631-73-8), diethyltin hydride, triethyltin hydride (997-50-2), dibutyltin hydride (1002-53-5), tributyltin hydride (686-73-3), phenyltin hydride (2406-68-0)	31
Pb	Plumbane (15875-18-0), trimethyllead hydride (7442-13-9), monoethyllead hydride (30691-92-0), diethyllead hydride, triethyllead hydride, tetraethyllead	32
Ge	Germane (7482-65-2), monomethylgermanium hydride (1449-65-6), dimethylgermanium hydride (1449-64-5), trimethylgermanium hydride (1449-63-4)	33

Table 5 Chemical speciation by HPLC coupled to AAS using HG

Element	Chemical species	Ref.
As	Arsenous acid (13464-58-9), arsenic acid (7778-39-4), monomethylarsonic acid (124-58-3), dimethylarsinic acid (76-60-5), *o*-arsanilic acid (2045-00-3), *p*-arsanilic acid (98-50-0), arsenobetaine (64436-13-1), arsenocholine (39895-81-3)	36
Se	Selenous acid (7783-00-8), selenic acid (7783-08-6), selenomethionine (1464-42-2), selenoethionine (6810-64-6), selenocysteine (3614-08-2)	37
Sb	Antimonous acid (13453-11-7), antimonic acid (12712-36-6)	38

High-performance liquid chromatography (HPLC) can be coupled to AAS, ICP/AES and ICP/MS to separate both volatile and nonvolatile species. Samples to be speciated by HPLC are extracted with water-soluble solvents such as methanol (67-56-1) and purified by centrifugation. The effuluents from HPLC are decomposed and hydridized with SBH. Table 5 shows some examples of chemical speciation by HPLC coupled with AAS using HG.

There are several review articles on chemical speciation by HG in spectroscopic analysis. Szpunar-Lobinska et al. included the HG technique in their speciation review.[39] Nakahara described chemical speciation by HG spectroscopic analysis[40] and Tao included the HG technique in his ICP/MS speciation review.[41] Ethylation with sodium tetraethylborate is also used for the speciation[42] of metals and organometallics.

2.2 Hydride Generation for Inductively Coupled Plasma Atomic Emission Spectrometry

2.2.1 Apparatus for Hydride Generation Inductively Coupled Plasma Atomic Emission Spectrometry

In ICP/AES, the hydrides are separated from the solution and swept continuously into an argon (8007-14-5) plasma in an ICP atomic emission spectrometer in a similar way to the continuous-flow method in AAS. Since the temperature of the plasma is much higher than that of the quartz cell in AAS, the plasma can be extinguished by ignition of hydrogen (1333-74-0) evolved as a by-product. Hence the instrumental parameters in ICP/AES should be optimized to prevent the plasma from being extinguished. Figure 4 shows a typical HG apparatus for ICP/AES.[43] ICP torches should be cleaned with concentrated nitric acid or aqua regia in a tall beaker after several uses.

2.2.2 Detection Limits and Ranges

The detection limits for ICP/AES range from parts per billion to parts per trillion as shown in Table 2. The dynamic ranges are broad.

2.2.3 Interferences

Interferences in ICP/AES[44] are similar to those in AAS. Therefore, the interferents are treated in a similar manner to those in AAS.

2.2.4 Environmental Sample Preparation for Hydride Generation Inductively Coupled Plasma Atomic Emission Spectrometry

Samples in ICP/AES are treated in a similar way to those in AAS.

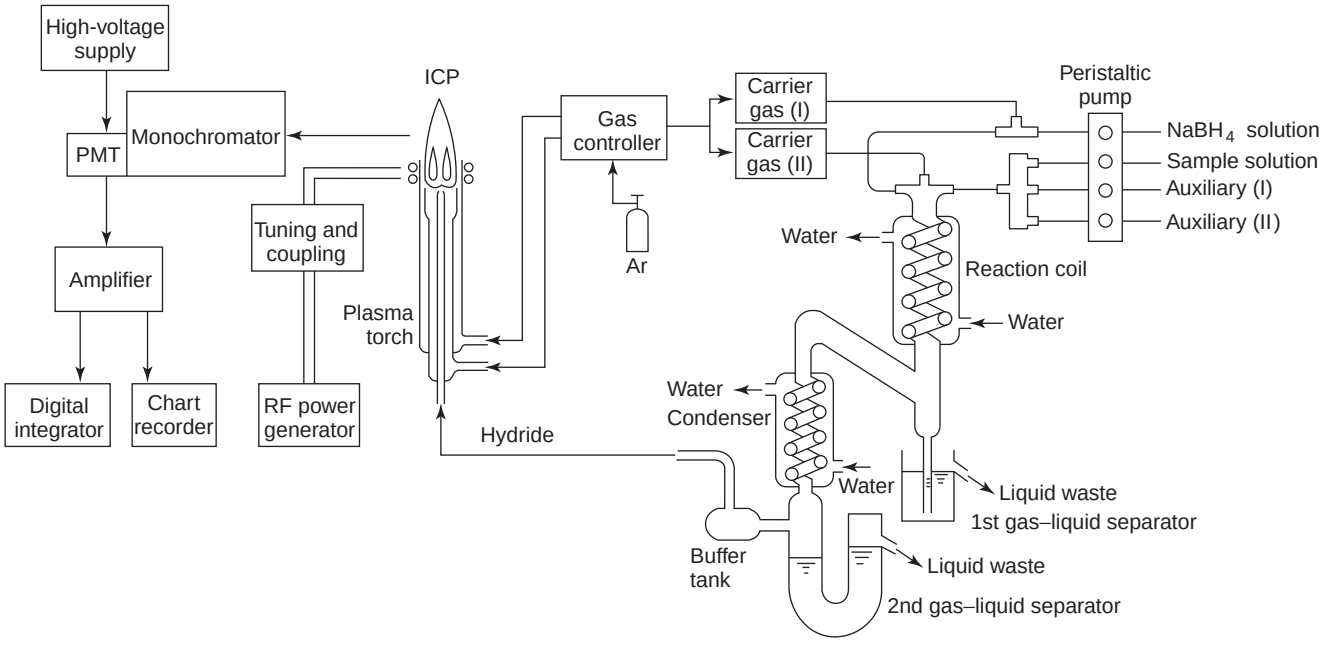

Figure 4 Schematic diagram of the continuous-flow hydride generator in ICP/AES. (Reproduced from T. Nakahara, *Appl. Spectrosc*, **37**(6), 539–545 (1983) by permission of the Society of Applied Spectroscopy.[43])

2.2.5 Chemical Speciation by Inductively Coupled Plasma Atomic Emission Spectrometry

HPLC is also coupled to ICP/AES. Table 6 shows some examples of HPLC coupled to ICP/AES using HG. Figure 5 shows a schematic diagram of the HPLC system in HG/ICP/AES.[45] Effluents from a high-performance liquid chromatograph are decomposed, hydridized and swept into the torch of an ICP atomic emission spectrometer.

2.3 Hydride Generation for Inductively Coupled Plasma Mass Spectrometry

2.3.1 Apparatus for Hydride Generation Inductively Coupled Plasma Mass Spectrometry

The HG apparatus for ICP/MS is similar to that for ICP/AES. Figure 6 shows a typical hydride introduction system to an ICP mass spectrometer.[48] The ICP torch positions are optimized by sweeping a hydride-forming element not to be determined. The torches should be cleaned in the same way as described for those in ICP/AES. Sampling and skimmer cones are cleaned by warming in 0.1% nitric acid after several uses.

2.3.2 Detection Limits and Ranges

The detection limits for ICP/MS are the lowest and range from parts per trillion to parts per quadrillion, as shown in Table 2. However, the sensitivity is restricted by contamination from the reagent blanks during the preparation of standard and sample solutions. Ultrahigh-purity reagents should be used to minimize the reagent blank.

2.3.3 Interferences

Chemical interferences in ICP/MS are the same as those in AAS and ICP/AES.

Hydrochloric acid is usually used to stabilize the hydride-forming elements in dilute standard solutions and to prepare sample solutions. However, the signal of

Table 6 Chemical speciation by HPLC coupled to ICP/AES using HG

Element	Chemical species	Ref.
As	Arsenite (15502-74-6), arsenate (15584-04-0), monomethylarsonate, dimethylarsinate, arsenobetaine, arsenocholine	45
Se	Selenite (14124-67-5), selenate (14124-68-6), selenomethionine, selenoethionine	46
Sn	Monobutyltin chloride (1118-46-3), dibutyltin chloride (683-18-1), tributyltin chloride (1461-22-9), triphenyltin chloride (639-58-7)	47

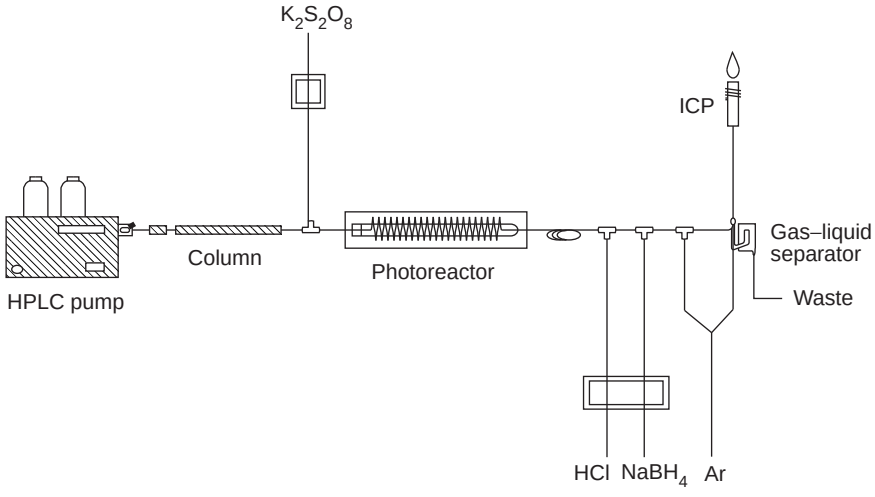

Figure 5 Schematic diagram of the HPLC system in HG/ICP/AES. (Reproduced from *Fresenius' J. Anal. Chem.*, **351**(4–5), 415–419 (1995) by permission of Springer-Verlag.[39])

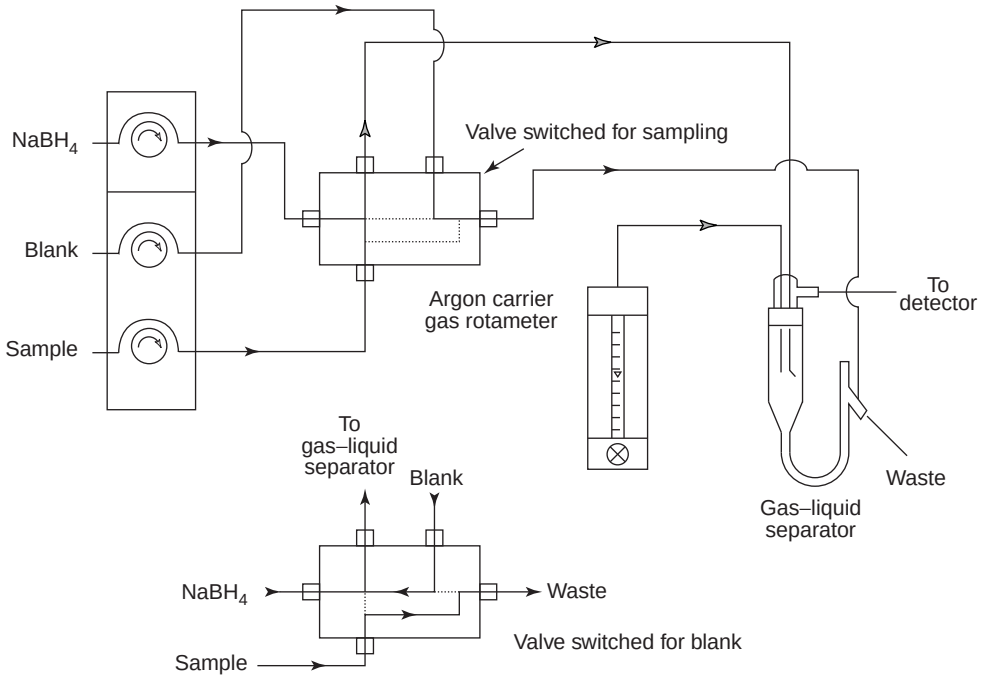

Figure 6 Schematic diagram of the continuous-flow hydride generator in ICP/MS. (Reproduced from P. Hitchen, R. Hutton, C. Tye, *J. Autom. Chem.*, **14**(1), 17–23 (1992)[48] by permission of Taylor and Francis Limited.)

^{75}As has an isobaric interference with $^{40}Ar^{35}Cl$, as shown in Table 7. Nitric acid is used instead of hydrochloric acid to avoid the overlap of signals in a quadrupole mass spectrometer. Selenium is determined by measuring the intensity ratio of its isotopes in a quadrupole mass spectrometer, since it has several isotopes and isobaric interferences, as shown in Table 8. The isobaric interference of ^{75}As with $^{40}Ar^{35}Cl$ can be resolved in a high-resolution mass spectrometer. Further, the signal of

Table 7 Interferences of arsenic with main isobars in ICP/MS

Isobar	Mass number	Abundance ratio[a]	Resolution
^{75}As	74.921 596	100	
$^{40}Ar^{35}Cl$	74.931 236	75.7	−7 772.45
$^{38}Ar^{37}Cl$	74.928 635	0.02	−10 644.3

[a] The abundance ratios of molecules were calculated by multiplying both atom percentages in elements from IUPAC.[49]

Table 8 Interferences of selenium with main isobars in ICP/MS

Isobar	Mass number	Abundance ratio[a]	Resolution
^{74}Se	73.922 477	0.90	
$^{40}Ar^{34}S$	73.930 251	4.19	−9 509.44
^{76}Se	75.919 207	9.00	
$^{36}Ar^{40}Ar$	75.929 929	0.67	−7 081.19
^{77}Se	76.919 908	7.60	
$^{40}Ar^{37}Cl$	76.928 286	24.13	−9 181.68
^{78}Se	77.917 304	23.60	
$^{38}Ar^{40}Ar$	77.925 115	0.12	−9 975.83
^{80}Se	79.916 521	49.7	
$^{40}Ar_2$	79.924 766	99.2	−9 693.23
^{82}Se	81.916 709	9.2	
$^{65}Cu^{17}O$	81.926 923	0.01	−8 020.54
$^{40}Ar_2H_2$	81.940 146	–	3 455.38
^{82}Kr	81.913 483	11.6	25 392.9

[a] The abundance ratios of molecules were calculated by multiplying both atom percentages in elements from IUPAC.[49]

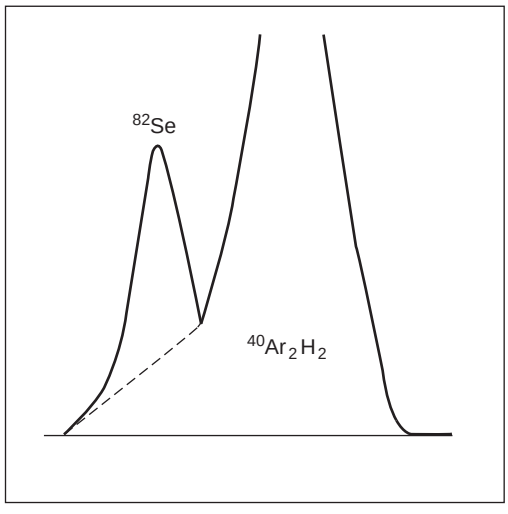

Figure 7 Separation of the ^{82}Se signal from a $^{40}Ar_2H_2$ band. (Reproduced from H. Narasaki, J.-Y. Cao, *Anal. Sci.*, **12**(8), 623–627 (1996) by permission of The Japan Society for Analytical Chemistry.)

^{82}Se can be resolved graphically from a $^{40}Ar_2H_2$ broad band in a high-resolution mass spectrometer as shown in Figure 7.[50] A silicone-rubber tubular membrane gas–liquid separator has been introduced to avoid signal overlaps as shown in Figure 8.[51] The separator removes any vapor of chlorides (16887-00-6) and allows arsenic and selenium of masses 75 and 77 to be separated.

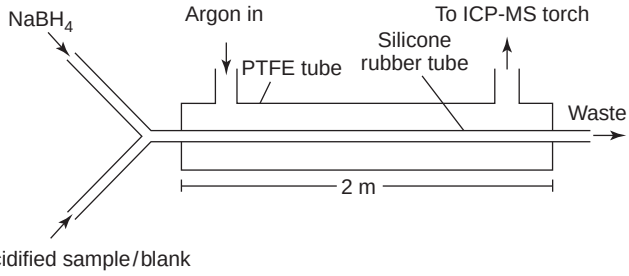

Figure 8 Schematic diagram of the membrane gas–liquid separator. (Reproduced from S. Branch, W.T. Corns, L. Ebdon, S. Hill, P. O'Neill, *J. Anal. At. Spectrom*, **6**(3), 155–158 (1991) by permission of The Royal Society of Chemistry.)

2.3.4 Environmental Sample Preparation for Hydride Generation Inductively Coupled Plasma Mass Spectrometry

The concentrations of acid and SBH solutions used in HG/ICP/MS should be diluted to approximately one tenth of those in HG/ICP/AES. The introduction of large amounts of water vapor and nonvolatile salts into the vacuum system should be avoided. The ion multiplier (detector) of an ICP mass spectrometer deteriorates as a result of the accumulation of salts. Evolution of hydrogen by HG should be minimized to maintain a stable plasma, since combustion of large amounts of hydrogen gas damages the plasma torch by its heat.

2.3.5 Chemical Speciation by Hydride Generation Inductively Coupled Plasma Mass Spectrometry

HPLC is coupled either directly or via hydridization to ICP/MS. The latter technique is very sensitive. Table 9 shows some examples of chemical speciation by HPLC coupled to ICP/MS using HG. Effluents from a high-performance liquid chromatograph are decomposed, hydridized and swept into the torch of an ICP mass spectrometer. Capillary electrophoresis (CE) is combined with ICP/MS owing to its sensitivity. HG/CE/ICP/MS is used for speciation of arsenic with a microporous membrane,[53] since the volume of effluents from CE is in the nanoliter range.

Table 9 Chemical speciation by HPLC coupled to ICP/MS using HG

Element	Chemical species	Ref.
Se	Se(IV), Se(VI), selenomethionine, selenoethionine	46
Sb	Sb(III), Sb(V)	38
Pb	Pb(II), tetramethyllead chloride (75-74-1), tetraethyllead chloride (78-00-2)	52

ACKNOWLEDGMENTS

I express my gratitude to the following persons who made valuable comments: Masahiko Tachibana, Department of Chemistry, Faculty of Science, Saitama University, Taketoshi Nakahara, Faculty of Engineering, Osaka Prefectural University and Kiyotaka Ishii and Kouichiro Outsuka, Japan Electron Optics Laboratory, Tokyo, Japan.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
CE	Capillary Electrophoresis
FAAS	Flame Atomic Absorption Spectrometry
GC	Gas Chromatography
HG	Hydride Generation
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
ICP/AES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
MS	Mass Spectrometry
SBH	Sodium Tetrahydroborate(III)

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Sample Preparation for Food Analysis, General

Atomic Spectroscopy **(Volume 11)**
Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications

REFERENCES

1. A. Walsh, 'The Application of Atomic Absorption Spectra to Chemical Analysis', *Spectrochim. Acta*, **7**(2), 108–117 (1955).
2. W. Holak, 'Gas-sampling Technique for Arsenic Determination by Atomic Absorption Spectrophotometry', *Anal. Chem.*, **41**(12), 1712–1713 (1969).
3. F.J. Schmidt, J.L. Royer, 'Submicrogram Determination of Arsenic, Selenium, Antimony and Bismuth by Atomic Absorption Utilizing Sodium Borohydride Reduction', *Anal. Lett.*, **6**(1), 17–23 (1973).
4. J. Dedina, D.L. Tsalev, *Hydride Generation Atomic Absorption Spectrometry*, John Wiley & Sons, Chichester, 19, 1995.
5. S. Greenfield, I.L. Jones, C.T. Berry, 'High-pressure Plasmas as Spectroscopic Emission Sources', *Analyst*, **89**(11), 713–720 (1964).
6. R.H. Wendt, V.A. Fassel, 'Induction-coupled Plasma Spectrometric Excitation Source', *Anal. Chem.*, **37**(7), 920–922 (1965).
7. M. Thompson, B. Pahlavanpour, S.J. Walton, G.F. Kirkbright, 'Simultaneous Determination of Trace Concentrations of Arsenic, Antimony, Bismuth, Selenium and Tellurium in Aqueous Solution by Introduction of the Gaseous Hydrides into an Inductively Coupled Plasma Source for Emission Spectrometry. Part I. Preliminary Studies', *Analyst*, **103**(6), 568–579 (1978).
8. R. Houk, V.A. Fassel, G.D. Flesch, H.J. Svec, A.L. Gray, C.E. Taylor, 'Inductively Coupled Argon as an Ion Source for Mass Spectrometric Determination of Trace Elements', *Anal. Chem.*, **52**(14), 2283–2289 (1980).
9. M.J. Powell, D.W. Boomer, R.J. McVicars, 'Introduction of Gaseous Hydrides into an Inductively Coupled Plasma Mass Spectrometer', *Anal. Chem.*, **58**(13), 2864–2867 (1986).
10. T. Nakahara, 'Applications of Hydride Generation Techniques in Atomic Absorption, Atomic Fluorescence and Plasma Atomic Emission Spectroscopy', *Prog. Anal. At. Spectrosc.*, **6**, 163–223 (1983).
11. T. Nakahara, 'Hydride Generation Techniques and Their Applications in Inductively Coupled Plasma/Atomic Emission Spectrometry', *Spectrochim. Acta Rev.*, **14**(1–2), 95–109 (1991).

12. A.D. Campbell, 'A Critical Survey of Hydride Generation Techniques in Atomic Spectroscopy (Technical Report)', *Pure Appl. Chem.*, **64**(2), 227–244 (1992).
13. B. Welz, M. Melcher, 'Mutual Interactions of Elements in the Hydride Technique in Atomic Absorption Spectrometry. Part 1. Influence of Selenium on Arsenic Determination', *Anal. Chim. Acta*, **131**, 17–25 (1981).
14. P.N. Vijan, G.R. Wood, 'An Automated Submicrogram Determination of Arsenic in Atmospheric Particulate Matter by Flameless Atomic Absorption Spectrophotometry', *At. Absorpt. Newsl.*, **13**(2), 33–37 (1974).
15. K.C. Thompson, D.R. Thomerson, 'Atomic-absorption Studies on the Determination of Antimony, Arsenic, Bismuth, Germanium, Lead, Selenium, Tellurium and Tin by Utilizing the Generation of Covalent Hydrides', *Analyst*, **99**(9), 595–601 (1974).
16. H. Matusiewicz, R.E. Sturgeon, 'Review of Atomic Spectrometric Detection of Hydride Forming Elements Following In Situ Trapping Within a Graphite Furnace', *Spectrochim. Acta, Part B*, **51**(4), 377–397 (1996).
17. T. Nakahara, *Advances in Atomic Spectroscopy*, JAI Press, Stanford, CT, 156, Vol. 2, 1995.
18. B. Welz, M. Melcher, 'Mechanisms of Transition Metal Interferences in Hydride Generation Atomic-absorption Spectrometry Part 1. Influence of Cobalt, Copper, Iron and Nickel on Selenium Determination', *Analyst*, **109**(5), 569–572 (1984).
19. J.W. Jones, S.G. Capar, T.C. O'Haver, 'Critical Evaluation of a Multi-element Scheme Using Plasma Emission and Hydride Evolution Atomic-absorption Spectrometry for the Analysis of Plant and Animal Tissues', *Analyst*, **107**(4), 353–377 (1982).
20. G.F. Kirkbright, M. Taddia, 'Application of Masking Agents in Minimizing Interferences from Some Metal Ions in the Determination of Arsenic by Atomic Absorption Spectrometry with the Hydride Generation Technique', *Anal. Chim. Acta*, **100**, 145–150 (1978).
21. J. Aggett, Y. Hayashi, 'Observations on the Interference by Copper(II), Cobalt(II) and Nickel(II) on the Determination of Arsenic by Arsine Generation Atomic Absorption Spectrometry', *Analyst*, **112**(3), 277–282 (1987).
22. B. Welz, M. Sucmanova, 'L-Cysteine as a Reducing and Releasing Agent for the Determination of Antimony and Arsenic Using Flow Injection Hydride Generation Atomic Absorption Spectrometry, Part 2. Interference Studies and the Analysis of Copper and Steel', *Analyst*, **118**(11), 1425–1432 (1993).
23. M. Yamamoto, T. Shohji, T. Kumamaru, Y. Yamamoto, 'Masking Effect of Potassium Iodide on the Interferences in the Atomic-absorption Spectrophotometric Determination of Antimony Utilizing Stibine Generation by Sodium Borohydride Tablet Reduction', *Fresenius Z. Anal. Chem.*, **305**(1), 11–14 (1981).
24. I. Bridle, X.-C. Le, X.-F. Li, 'Convenient Method for the Determination of Trace Amounts of Germanium by Hydride Generation Direct Current Plasma Atomic Emission Spectrometry: Interference Reduction by L-Cystine and L-Cysteine', *J. Anal. At. Spectrom.*, **4**(3), 227–232 (1989).
25. J.R. Knechtel, J.L. Fraser, 'Preparation of a Stable Borohydride Solution for Use in Atomic-absorption Studies', *Analyst*, **103**(1), 104–105 (1978).
26. J. Aggett, A.C. Aspell, 'The Determination of Arsenic(III) and Total Arsenic by Atomic-absorption Spectroscopy', *Analyst*, **101**(5), 341–347 (1976).
27. M. Yamamoto, K. Urata, K. Murashige, Y. Yamamoto, 'Differential Determination of Arsenic(III) and Arsenic(V), and Antimony(III) and Antimony(V) by Hydride Generation/Atomic Absorption Spectrometry, and its Application to the Determination of These Species in Sea Water', *Spectrochim. Acta, Part B*, **36**(7), 671–677 (1981).
28. R.J.A.V. Cleuvenbergen, W.E.V. Mol, F.C. Adams, 'Arsenic Speciation in Water by Hydride Cold Trapping–Quartz Furnace Atomic Absorption Spectrometry: an Evaluation', *J. Anal. At. Spectrom.*, **3**(1), 169–176 (1988).
29. G.A. Cutter, 'Species Determination of Selenium in Natural Waters', *Anal. Chim. Acta*, **98**, 59–66 (1978).
30. M.O. Andreae, J.-F. Asmode, P. Foster, L. Van't dack, 'Determination of Antimony(III), Antimony(V), and Methylantimony Species in Natural Waters by Atomic Absorption Spectrometry with Hydride Generation', *Anal. Chem.*, **53**(12), 1766–1771 (1981).
31. V.F. Hodge, S.L. Seidel, E.D. Goldberg, 'Determination of Tin(IV) and Organotin Compounds in Natural Waters, Coastal Sediments and Macro Algae by Atomic Absorption Spectrometry', *Anal. Chem.*, **51**(8), 1256–1259 (1979).
32. P. Foster, M. Laffond, R. Perraud, P. Baussand, V. Jacob, 'Analysis of Atmospheric Pollutants by Gas Chromatography–Atomic Absorption or Thermodesorption–Gas Chromatography', *Int. Environ. Anal. Chem.*, **28**(1–2), 105–120 (1987).
33. G.A. Hambrick III, P.N. Froelich, Jr, M.O. Andreae, B.L. Lewis, 'Determination of Methylgermanium Species in Natural Waters by Graphite Furnace Atomic Absorption Spectrometry with Hydride Generation', *Anal. Chem.*, **56**(3), 421–424 (1984).
34. M.O. Andreae, 'Determination of Arsenic Species in Natural Waters', *Anal. Chem.*, **49**(6), 820–823 (1977).
35. P.M. Sarradin, F. Leguille, A.A. Astruc, R. Pinel, M. Astruc, 'Optimization of Atomization Parameters in the Speciation of Organotin Compounds by Hydride Generation–Gas Chromatography–Electrothermal Atomic Absorption Spectrometry', *Analyst*, **120**(1), 79–83 (1995).
36. A.G. Howard, L.E. Hunt, 'Coupled Photooxidation–Hydride AAS Detector for the HPLC of Arsenic Compounds', *Anal. Chem.*, **65**(21), 2995–2998 (1993).
37. J.M. Marchante-Gayon, J.M. Gonzalez, M.L. Fernandez, E. Blanco, A. Sanz-Medel, 'Selenium Speciation by Coupling Vesicle Mediated HPLC with Off-line ETAAS

and On-line Focused Microwave Digestion HG-AAS Detection', *Fresenius' J. Anal. Chem.*, **355**(5–6), 615–622 (1996).

38. P. Smichowski, Y. Madrid, M. Beatriz de la Calle Guntinas, C. Camara, 'Separation and Determination of Antimony(III) and Antimony(V) Species by High-performance Liquid Chromatography with Hydride Generation Atomic Absorption Spectrometric and Inductively Coupled Plasma Mass Spectrometric Detection', *J. Anal. At. Spectrom.*, **10**(10), 815–821 (1995).
39. J. Szpunar-Lobinska, C. Witte, R. Lobinski, F.C. Adams, 'Separation Techniques in Speciation Analysis for Organometallic Species', *Fresenius' J. Anal. Chem.*, **351**(4–5), 351–377 (1995).
40. T. Nakahara, 'Chemical Speciation by Hydride Generation–Atomic Spectrometry', *Bunseki Kagaku*, **46**(7), 513–536 (1997).
41. H. Tao, 'Trace Elemental Speciation by Chromatography/Inductively Coupled Plasma Mass Spectrometry', *Bunseki Kagaku*, **46**(4), 239–263 (1997).
42. S. Raspomanikis, 'Derivatization by Ethylation With Sodium Tetraethylborate for the Speciation of Metals and Organomatallics in Environmental Samples. A Review', *Analyst*, **119**(7), 1429–1439 (1994).
43. T. Nakahara, 'The Determination of Trace Amounts of Tin by Inductively Coupled Argon Plasma Atomic Emission Spectrometry with Volatile Hydride Method', *Appl. Spectrosc.*, **37**(6), 539–545 (1983).
44. J.W. Hershey, P.N. Keliher, 'Some Hydride Generation Inter-element Interference Studies Utilizing Atomic Absorption and Inductively Coupled Plasma Emission Spectrometry', *Spectrochim. Acta, Part B*, **41**(7), 713–723 (1986).
45. J. Alberti, R. Rubio, G. Rauret, 'Arsenic Speciation in Marine Biological Materials by LC–UV–HG–ICP/OES', *Fresenius' J. Anal. Chem.*, **351**(4–5), 415–419 (1995).
46. J.M. Gonzalez Lafuente, M.L. Fernandez Sanchez, A. Sanz-Medel, 'Speciation of Inorganic Selenium and Selenoaminoacids by On-line Reversed-phase High-performance Liquid Chromatography–Focused Microwave Digestion–Hydride Generation–Atomic Detection', *J. Anal. At. Spectrom.*, **11**(12), 1163–1169 (1996).
47. P. Rivaro, L. Zaratin, R. Frache, A. Mazzucotelli, 'Determination of Organotin Compounds in Marine Mussel Samples by Using High-performance Liquid Chromatography–Hydride Generation Inductively Coupled Plasma Atomic Emission Spectrometry', *Analyst*, **120**(7), 1937–1939 (1995).
48. P. Hitchen, R. Hutton, C. Tye, 'The Application of a Commercial Gas/Liquid Separator Coupled with an Inductively Coupled Argon Plasma/Mass Spectrometer', *J. Autom. Chem.*, **14**(1), 17–23 (1992).
49. IUPAC, Inorganic Chemistry Division, Commission on Atomic Weights and Isotopic Abundances, 'Elements by Element Review of Their Atomic Weights', *Pure Appl. Chem.*, **56**(6), 695–768 (1984).
50. H. Narasaki, J.-Y. Cao, 'Determination of Arsenic and Selenium in River Water by Hydride Generation Inductively Coupled Plasma/Mass Spectrometry with High Resolution', *Anal. Sci.*, **12**(8), 623–627 (1996).
51. S. Branch, W.T. Corns, L. Ebdon, S. Hill, P. O'Neill, 'Determination of Arsenic by Hydride Generation Inductively Coupled Plasma Mass Spectrometry Using a Tubler Membrane Gas–Liquid Separator', *J. Anal. At. Spectrom.*, **6**(3), 155–158 (1991).
52. H.-J. Yang, S.-J. Jiang, 'Hydride Generation Inductively Coupled Plasma Mass Spectrometric Detection of Lead Compounds Separated by Liquid Chromatography', *J. Anal. At. Spectrom.*, **10**(11), 963–967 (1995).
53. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, 'Speciation of Arsenic Compounds in Drinking Water by Capillary Electrophoresis with Hydrodynamically Modified Electroosmotic Flow Detected Through Hydride Generation Inductively Coupled Plasma/Mass Spectrometry with a Membrane Gas–Liquid Separator', *J. Anal. At. Spectrom.*, **12**(7), 689–695 (1997).

Immunoassay Techniques in Environmental Analyses

Barry Lesnik
United States Environmental Protection Agency, Washington, USA

1 Introduction 1
2 Background and History 2
2.1 Overview of United States Environmental Protection Agency Regulatory Programs 2
2.2 Developmental History of Immunoassay Methods in United States Environmental Protection Agency Programs 3
3 Types of Immunoassay Technologies Employed in Environmental Programs 4
3.1 Enzyme-linked Immunosorbent Assay 4
3.2 Reporter Gene on a Human Cell Line 5
3.3 Immunosensors 6
3.4 Affinity Chromatography 6
3.5 Dissociation-enhancement Lanthanide Fluoroimmunoassay 6
4 Method Development 6
4.1 General Guidelines for Development of Screening Methods 6
4.2 Validation Criteria for Immunoassay Methods 7
5 Status of the United States Environmental Protection Agency Immunoassay Methods Development Program 8
5.1 Current Validated Methods 8
5.2 Methods under Development 9
5.3 List of Validated Immunoassay Test Products 11
6 Environmental Applications 11
6.1 Advantages and Disadvantages of Immunoassay Methods 11
6.2 Implementation Issues 12
6.3 Examples of Performance Based Measurement Systems 13
6.4 Applications as Screening Methods 13
6.5 Applications as Quantitative Methods 14
7 Future Directions 15
Acknowledgments 15
Abbreviations and Acronyms 15
Related Articles 16
References 16

In the 1980s, several Program Offices of the United States Environmental Protection Agency (USEPA) investigated the use of immunoassay methods for environmental applications with limited success. Most of the problems were with method ruggedness, i.e. they worked well in clean spiked matrices, but not very well on real world samples. The situation changed significantly in January of 1992, when the Office of Solid Waste (OSW) received its first rugged immunoassay method (for pentachlorophenol (PCP)) that worked on field samples. Several more followed, and through the early and mid-1990s, the USEPA initiated use of immunoassay methodology in its hazardous waste program. The first group of methods incorporated into the OSW's methods compendium, Test Methods for Evaluating Solid Waste*, or SW-846, between 1993 and 1995 were a group of 10 enzyme-linked immunosorbent assay (ELISA)-based screening methods for various environmental analyte classes. These methods were officially 'approved' by the USEPA in July, 1997. We provided a great deal of training on these methods to raise the comfort level of the environmental community for their use.*

This chapter addresses several major topics including an overview of the USEPA's major Regulatory Programs; how analytical methods are used in Regulatory Programs; general guidelines for the development of screening methods; specific validation criteria for immunoassay methods; the current status of the USEPA immunoassay method development program; current and potential environmental applications for immunoassay technology; barriers to implementation of immunoassay methods and the steps being taken to overcome them; and future directions and new developments in immunoassay technology for regulatory programs.

The overall future of the technology for environmental monitoring and analysis looks very bright. It offers a cost effective way to generate reliable information upon which to base sound environmental decisions.

1 INTRODUCTION

Immunoassay technology has several attributes which make it a useful tool for environmental monitoring, e.g. selectivity, sensitivity, portability, and rapid turnaround time. Immunoassay kits can be tailored to target specific analytes or classes of analytes, thus eliminating the need for cleanup methods in most cases to remove interferences. They also have the capability of detecting

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

target analytes at very low levels, which are needed in many environmental applications. The portability of immunoassay test kits and speed of analysis allows for rapid analyses to be run on a site in the field. This capability can be especially useful in lowering the costs of cleanup projects because equipment does not have to lay idle while awaiting the results of laboratory analyses.

The USEPA has been looking at the potential use of immunoassay technology for environmental monitoring for several years. The early methods development efforts were unsuccessful because the immunoassay chemistry utilized in the methods was not sufficiently rugged for use on real world environmental matrices. The methods performed well on clean water matrices and spiked samples, but did not perform effectively on natural environmental samples. Because of this poor initial performance on real samples, USEPA Program Office interest in the technology declined.

In January 1992, EnSys, Inc. demonstrated a viable immunoassay test kit for PCP in both soil and water matrices to USEPA's OSW. Since that time, OSW has been working with several manufacturers to develop and validate a whole battery of immunoassay test kits both for individual analytes and for classes of analytes. The first series of immunoassay methods adopted by OSW for inclusion in its methods manual, *Test Methods for Evaluating Solid Waste*, (SW-846),[1] used ELISA as the technique of choice. Currently, OSW has issued more than fifteen immunoassay methods using several different techniques, which can be used for analyses performed under the Resource Conservation and Recovery Act (RCRA).[2]

2 BACKGROUND AND HISTORY

2.1 Overview of United States Environmental Protection Agency Regulatory Programs

2.1.1 Organization

The USEPA is not a single entity. It consists of several regulatory Program Offices which have the responsibility for implementing the major environmental laws passed by the US Congress, as well as several administrative and technical offices which support these regulatory Program Offices. These regulatory Program Offices include the Office of Solid Waste and Emergency Response (OSWER), the Office of Water (OW), the Office of Prevention, Pesticides and Toxic Substances (OPPTS), and the Office of Air and Radiation (OAR).

The various regulatory Program Offices were established at different times in response to Congressional mandates establishing the major environmental laws. These laws are very general in nature, and the USEPA regulatory program offices were formed to set-up the specific regulations needed to administer and enforce these environmental laws. Approaches to regulatory requirements and philosophy vary greatly among the regulatory Program Offices, because of the significant differences in their areas of responsibility. For example, regulations governing air and water, dealing with single media, are necessarily different than those dealing with the management of solid and hazardous waste, which involve a wide variety of media.

2.1.2 Programmatic Regulatory Responsibilities

OSWER is responsible for administering the RCRA through its OSW and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), also known as Superfund, through its Office of Emergency Response and Remediation (OERR). The OW administers the Clean Water Act (CWA) and the Safe Drinking Water Act (SDWA). The OPPTS administers the Toxic Substances Control Act (TSCA) and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), while the OAR administers the Clean Air Act (CAA). The Office of Administration and Resource Management (OARM) provides administrative support while the Office of Research and Development (ORD), along with several other non-regulatory program offices provide technical support to the regulatory Program Offices.

2.1.3 Regulatory Analytical Methods

The administration of these USEPA regulatory programs involves compliance monitoring, which utilizes a wide variety of analytical methods and techniques. This section of the chapter will focus on the similarities and differences only between the OSWER and OW monitoring programs.

The OSW regulates the management of solid and hazardous waste under RCRA. This includes regulations on generation, storage, and treatment of wastes and waste residuals from active management sites. Most RCRA applications involve site-specific analyses for compliance with permits or other regulations, unlike the general maximum contaminant level (MCL) requirements common to compliance with drinking water regulations. Since RCRA applications encompass a wide variety of media, including, soils, solids, sludges, organic liquids, water, stack emissions, and ambient air, OSW has historically used the performance based measurement system (PBMS) approach to sampling and analysis. This allows for the needed flexibility in matrix-specific and project-specific methods selection for RCRA applications.

The OSW Methods Team publishes a manual of analytical methods applicable for use in the RCRA

Program, *Test Methods for Evaluating Solid Waste*, SW-846. SW-846 functions primarily as a guidance document. In most cases under RCRA, analyses may be performed using either SW-846 methods or any other method that may be appropriate, since RCRA regulations usually only specify what the analytical requirements are, and not specifically how to do them. Whichever methods are selected, the analyst must demonstrate their applicability for their intended use. This can be done by using a two-tiered demonstration of proficiency: 1) the analyst must demonstrate the ability to perform the method and obtain acceptable results in a 'clean' matrix, such as reagent water or Ottawa sand; and 2) the analyst must demonstrate that the method will give acceptable performance in the actual matrix of concern.

However, there are a very few instances in the RCRA regulations where SW-846 analytical methods, in general, must be used to characterize waste. Because of these few cases which require the use of SW-846 methods, all new and revised methods which are included in SW-846 must be published as regulations. OSW is currently working on a regulation under the 'Reinvention of Government' guidelines to eliminate this requirement and to publish SW-846 methods as guidance as was the original intent of the manual. Methods which will not become guidance are the 'method-defined parameters', or methods which directly define regulatory requirements, e.g. flash point, corrosivity.

The OERR, on the other hand, deals with the cleanup of abandoned waste sites under CERCLA (Superfund). Most Superfund analyses are performed using the same multimedia PBMS approach as is used for RCRA. Some Superfund analyses are performed under specific contracts which can use any method deemed appropriate for the application. However, when Superfund methods are written into their contract format, they are very detailed and must be performed as written. Changes or modifications are generally not permitted.

The OW, for compliance with both the CWA and SDWA, is required to publish lists of both regulated and unregulated target analytes and to publish approved methods for the analyses of these analytes. Use of alternative methods is not permitted, unless a method equivalency petition is requested by a petitioner and granted by USEPA or other designated authority.

Thus, of the four major USEPA Program Offices for which substantial numbers of analyses are performed, RCRA offers the most flexibility as to choice and application of analytical methods. RCRA only specifies what the analytical requirements are, and leaves the decision as to how the analyses are to be performed to the analyst.

2.2 Developmental History of Immunoassay Methods in United States Environmental Protection Agency Programs

While several USEPA Program Offices investigated the potential applicability of immunoassay methods to their programs, OSW is the first USEPA Program Office to formally incorporate these techniques into its methods program. OSW has taken the Agency lead in developing new immunoassay techniques for its own methods program and is actively working with other USEPA Program Offices to assist them in developing immunoassay methods for their programs.

OSW began evaluation of its first immunoassay method (for PCP) in January 1992, followed by three others in rapid succession. By July 1995, OSW had completed validation of a general overview method and ten individual immunoassay methods (with more than 25 different kits) utilizing ELISA, and published them in the proposed Update III of SW-846. For a USEPA regulatory Program Office, this is very rapid progress. These immunoassay methods covered the analysis of PCP, polychlorinated biphenyls (PCBs), polynuclear aromatic hydrocarbons (PAHs), petroleum hydrocarbons, pesticides and explosives. See Table 1 for a list of these methods. OSW formally approved the first immunoassay method, for PCP, as part of a January 1994, regulation. The rest of the methods in Table 1 were formally approved as part of Update III to SW-846 in June, 1997. All of these methods can be used for any RCRA application for which they can be demonstrated to work.

Between late 1995 and late 1997, the environmental immunoassay industry fell on hard times due to severe cutbacks in remediation projects for which there was a major potential market for use of these environmental immunoassays. This resulted in a consolidation of the

Table 1 Immunoassay methods in update III of SW-846

Method 4000:	Immunoassay[3–10]
Method 4010:	PCP in Water and Soils by Immunoassay[11–14]
Method 4015:	2,4-D in Water and Soils by Immunoassay[15–17]
Method 4020:	PCBs in Soil by Immunoassay[18–26]
Method 4030:	TPH in Soil by Immunoassay[27–29]
Method 4035:	Soil Screening for PAHs by Immunoassay[30–32,80]
Method 4040:	Toxaphene in Soils by Immunoassay[33,34]
Method 4041:	Chlordane in Soils by Immunoassay[35]
Method 4042:	DDT in Soils by Immunoassay[36]
Method 4050:	TNT Explosives in Water and Soils by Immunoassay[37–40]
Method 4051:	RDX Explosives in Water and Soils by Immunoassay[40,41]

2,4-D, 2,4-dichlorophenoxyacetic acid; TPH, total petroleum hydrocarbons; TNT, 2,4,6-trinitrotoluene; RDX, hexahydro-1,3,5-trinitro-1,3,5-triazine.

four major manufacturers into a single entity. While the industry was busy sorting itself out during this period, very little was happening in the area of new product and methods development. However, by the end of 1997, several new manufacturers began operating and several new products have either come on the market or are in the final stages of testing, including second generation ELISA kits for PAHs and PCBs, the first immunoassay method (ELISA for mercury) for a metal, immunosensors for explosives, and, finally a series of screening methods for dioxins and coplanar PCBs using a variety of techniques.

Some of the other USEPA Program Offices are also looking at immunoassay methods to address some of their analytical requirements. OPPTS is considering using immunoassay methods in its Pesticide Registration Program. OW is beginning to look at using the technology in both the Drinking Water and Wastewater Programs. However, they may have to revise some of their regulations to allow for the use of 'less than' values in reporting MCLs or look toward developing quantitative immunoassays.

3 TYPES OF IMMUNOASSAY TECHNOLOGIES EMPLOYED IN ENVIRONMENTAL PROGRAMS

3.1 Enzyme-linked Immunosorbent Assay

ELISA[3–10] is the most commonly used immunoassay technique for environmental analysis. The immunoassay test products available will often vary in both format and chemistry. The characteristics of a specific product are described in the package insert provided by the manufacturer. This summary is, therefore, general in scope, and is intended to provide a general description of the more common elements of these methods.

Immunoassay test products use an antibody molecule to detect and quantitate a substance in a test sample. These testing products combine the specific binding characteristics of an antibody molecule with a detection chemistry that produces a detectable response used for interpretation. In general, antibody molecules specific for the method's intended target are provided at a predefined concentration. A reporter (i.e. signal generating) reagent, composed of the target compound conjugated to a signal producing compound or molecule (e.g. enzymes, chromophores, fluorophores, luminescent compounds, etc.), is also provided. The concentration, affinity, and specificity of the products' antibody influences performance, as does the chemistry of the reporter reagent.

The reporter reagent and antibody molecules of a given product are binding partners, and form a complex in solution. The addition of a positive sample containing the target substance to this solution results in a competitive binding reaction between the target analyte and the reporter reagent for the antibody sites. The antibody concentration, and therefore binding capacity, is limited to prevent the simultaneous binding of both the reporter and target molecules. The concentration of reporter reagent that can bind to the antibody is inversely proportional to the concentration of substance in the test sample.

Immunoassay methods may be heterogeneous (i.e. requiring a wash or separation step), or homogeneous (i.e. not requiring a separation step). In commonly available heterogeneous testing products, the antibody is immobilized to a solid support such as a disposable test tube, and the bound reporter reagent will be retained after removing the unbound contents of the tube by washing. Therefore, a negative sample results in the retention of more reporter molecules than a positive sample. The analysis of a standard containing a known concentration results in the immobilization of a proportional concentration of reporter reagent. A positive sample (i.e. containing a higher concentration than the standard) results in the immobilization of fewer reporter molecules than the standard, and a negative sample (i.e. containing less than the standard) will immobilize more.

A chemistry of the detection of the immobilized reporter is used for interpretation of results. The reporter molecule may be a conjugate of the target molecule and a directly detectable chromophore, fluorophore, or other specie, or conjugated to an enzyme that will act upon a substrate to produce the detectable response. Immunoassay testing products have a quantitative basis, and will produce a signal that is dependant on the concentration of analyte present in the sample. For environmental immunoassay methods, the signal produced is exponentially related to the concentration of the compounds present. Many immunoassay methods use enzymes to develop a chromogenic response, and are termed enzyme immunoassays. Assays that generate a chromogenic response are analyzed photometrically, and use the principles of Beer's law (absorbance = extinction coefficient $\times$ concentration $\times$ pathlength) to determine the concentration of analyte in a sample.

Immunoassay methods can provide quantitative data when configured with a series of reference standards that are analyzed and used to construct a standard curve. The signal generated from the analysis of a test sample is used to determine concentration by interpolation from the standard curve. Alternatively, these testing products can be configured to determine if a sample is positive or negative relative to a single standard.

Individual immunoassay testing products are reviewed and accepted by the USEPA OSW for the detection of sample analytes in specified matrices. A variety of testing products, produced by several different developers, may

be available for the same compound(s) and matrices. Each of these methods have been formulated using independently developed reagents that may result in significantly different performance characteristics and limitations.

The performance of the immunoassay testing products ultimately relates to the characteristics of the antibody, reporter molecule, and sample processing chemistry. The dose-response characteristics of a method, the position of the standard relative to the claimed action level, and the stated cross-reactivity characteristics of the selected test product, provide relevant information regarding the performance and recognition profile of the selected test product.

The precision, and ultimately the sensitivity of an immunoassay method, is a function of the signal-to-noise characteristics of its dose-response curve, and its operational consistency. Methods having a high slope and low non-specific signal generation produce the most sensitive and precise methods. Signal imprecision applied to a dose-response curve having a shallow slope exhibits proportionally greater imprecision in the calculated concentration than would a method having a steeper slope. In an action level testing product, this would cause the reference standard to be positioned further from the action level, increasing the incidence of false positive results. Similarly, a method having less non-specific signal generation (higher signal-to-noise ratio) will be more sensitive and precise when other characteristics (i.e. dose-response slope) are held constant.

Immunoassay methods are used to detect contamination at a specific concentration below the claimed detection level for the test product. For example, an immunoassay used to detect PCB contamination in soil at 1 ppm will include a standard preparation containing less than 1 ppm. The reference preparation concentration is positioned to minimize the incidence of false negative results at the claimed detection level. For remediation and monitoring applications, where action levels of interest are defined, immunoassay methods should exhibit a negligible incidence of false negative results, and minimal false positives.

For a single point action level test, the concentration of analyte relative to the action level is selected by the developer, and is influenced by the precision (i.e. intra-assay, inter-person, inter-lot, inter-day, etc.), sample matrix interferences and other performance characteristics and limitations of the basic method. The concentration of analyte in the reference materials should be less than, but close to, the claimed action level. The concentration selected for the standard defines the concentration that will produce a 50% incidence of false positive results by the test product. While this issue is one representing limited liability to the operator, it is a practical issue that often requires attention. An immunoassay method for the detection of 1 ppm of PCB using a standard containing 0.8 ppm of PCB will experience a 50% false positive incidence in samples containing 0.8 ppm of PCB, and some incidence of false positive results in a sample containing between 0.8 and 1 ppm. A similar immunoassay that uses a standard containing 0.4 ppm will experience a 50% false positive incidence in samples containing 0.4 ppm of PCB, and some incidence of false positive results in a sample containing between 0.4 and 1 ppm. The closer the standard concentration is to the action level, the better the overall performance.

Cross reactivity characteristics illustrate the specificity of the underlying immunochemistry. The antibody molecules used by a test product bind to a target compound and then participate in the process of generating the signal used for interpretation. Antibody molecules bind by conformational complimentarity. These molecules can be exquisitely specific, and can differentiate subtle differences in the structure of a compound. The binding characteristics of reagents in different test products can vary, and influence the recognition profile and incidence of false results obtained by the method. Immunoassay methods should detect the target analytes claimed by the test product and exhibit limited recognition for compounds and substances not specified.

3.2 Reporter Gene on a Human Cell Line

The reporter gene system[42–50] utilizes a human cell line (101L) into which a plasmid containing a human CYP1A1 promoter and 5′-flanking sequences fused to a reporter gene, firefly luciferase, have been stably integrated. In the presence of CYP1A1-inducing compounds, the enzyme luciferase is produced, and its reaction with luciferin can be detected by measuring relative light units (RLUs) in a luminometer. CYP1A1-inducing environmental contaminants include polychlorinated dibenzo-*p*-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), coplanar PCB congeners, and high molecular weight PAHs.

To quantify the inducing compounds in the sample, the mean response, in RLUs, of the three sample replicates is divided by the mean response of three replicates of a solvent blank, yielding a 'fold induction', which is a measure of the increase of the sample response over the background response. Fold induction may be converted to toxic equivalents (TEQ) for PCBs and PCDDs/PCDFs, or benzo[*a*]pyrene equivalents (B[*a*]PEqs) for PAHs, based on the fold induction responses to standards containing a mixture of PCDDs/PCDFs, or benzo[*a*]pyrene, respectively.

This method can be used to estimate the concentrations of PAHs and/or coplanar PCBs in soils over a range from 0.1 to over 100 mg kg^{-1}. In addition, the presence

of PCDDs/PCDFs in the extract will be indicated at concentrations ranging from 100 ng kg^{-1} to over 1 mg kg^{-1}. These sensitivity estimates are based on the extraction of 40-g solid samples and evaporation of the extract to 1 mL, with the application of 20 μL to one million cells covered by 2 mL of medium. The sensitivity of the method for water samples is approximately 25 times greater, assuming that a 1-L water sample is extracted and the extract is concentrated to the same 1-mL volume (e.g. 4 μg L^{-1} to 4 mg L^{-1} for PAHs and PCBs, and 4 ng L^{-1} to 40 μg L^{-1} for PCDDs/PCDFs).

The method contains an optional procedure in which exposures are conducted over two specific time periods (6 and 16 hours). This option allows the test to distinguish between PAHs in the samples and chlorinated compounds, since the PAHs reach maximum induction at 6 hours, while the peak in induction from chlorinated planar compounds (PCBs, PCDDs/PCDFs) is not until 16 hours.

This method is a screening procedure, and depending on project needs. A subset of the samples should be confirmed using quantitative analytical techniques.

3.3 Immunosensors

Two types of immunosensor techniques have been evaluated at this time,[51–68] both of them for TNT and RDX explosives. The first, a flow cell technique is based on performing a fluorescent displacement immunoassay on a membrane or other solid support in flow. Antibodies immobilized on a solid support (i.e. membrane) are saturated with a fluorescent analog of either TNT or RDX. A test sample is injected into the flow stream. If the appropriate analyte (TNT or RDX) is present in the test sample, it will displace a proportional amount of fluorescent analog. The displaced analog is detected downstream. The fluorescent signal is proportional to the concentration of the analyte in the test sample. This technique can be used as a screening tool to determine if TNT or RDX is present above a critical limit or for quantitation of the explosive in the test sample. To quantitate, the area of the fluorescent peak is compared to standards tested pre and/or post test sample. Results obtained with this method can be used to identify samples with TNT and RDX concentrations between 10 and 1200 ppb.

The second, a fiberoptic technique, is based on performing a competitive fluorescent immunoassay on the surface of a fiber optic probe. A fluorescent analog of either TNT or RDX is added to the test sample. The sample is then passed over optical probes which have antibodies immobilized on the surface which are specific for either TNT or RDX. A decrease in maximum fluorescence proportional to the concentration of TNT or RDX in the sample is observed. This technique can be used as a screening tool to determine if TNT or RDX is present above a critical limit or for quantitation of the explosive in the test sample. To quantitate, the percent inhibition is compared to a standard curve. Results obtained with this method can be used to identify samples with TNT and RDX concentrations between 5 and 200 ppb.

3.4 Affinity Chromatography

Affinity chromatography is a potential sample preparation technique whereby the selective nature of immunoassay methods, particularly ELISA, can be used to extract specific compounds or compound classes out of a difficult matrix. The bound analytes can then be desorbed and analyzed by an appropriate chromatographic or other determinative technique, e.g. high-performance liquid chromatography (HPLC).

3.5 Dissociation-enhancement Lanthanide Fluoroimmunoassay

The DELFIA (dissociation-enhancement lanthanide fluoroimmunoassay) methodology was developed by scientists at EG&G Wallac, Turku, Finland and the Department of Molecular Endocrinology of Middlesex Hospital, London.[84] This method is clinically proven and is highly reliable and adaptive. The DELFIA system is based on time-resolved fluorometry of lanthanide compounds such as europium. Lanthanide ions exhibit a unique fluorescence that is characterized by narrowband emission lines, a long decay time, and large Stokes shift. The specific fluorescence of the lanthanide label is measured after a certain time delay following an activation pulse allowing all of the nonspecific background to expire. The DELFIA system and instrumentation allows the detection of four different toxicants in one sample by using four different lanthanide labels each having a distinctive fluorescence spectrum.

4 METHOD DEVELOPMENT

4.1 General Guidelines for Development of Screening Methods

The primary applicability that we, in the RCRA Program, see for immunoassay methods is for quantitative screening purposes.[69] By quantitative screening, we mean setting a quantitative action level (usually the regulatory action level), where a positive response means that the analyte is present at or above the action level, and a negative response tells us that the analyte is either absent or present below the level of regulatory concern. Analyses can be run at multiple action levels giving a useful range of concentrations for specific target analytes.

For example, if we are mapping a site contaminated with PCBs to determine the extent to which it needs to be cleaned up, knowing where PCB levels are <10 ppm, between 10 and 100 ppm, and >100 ppm can be useful in planning and expediting the cleanup.

The OSW Methods Team provides a guidance document (available on-line at http://www.epa.gov/SW-846) to potential developers of screening methods providing guidance on what general validation criteria should be applied to a screening method that will potentially be included in SW-846. While screening procedures need not be fully quantitative, they should measure the presence or absence of target analytes at or below regulatory action levels. Therefore, initial demonstration of method performance involves measuring the percentage of false negatives and false positives generated using the procedure for a single sample matrix. Data should be submitted for split samples analyzed using the developer's technique and an appropriate SW-846 quantitative method. A candidate procedure should ideally produce no false negatives. Definition of a false negative is a negative response for a sample that contains the stated detection level of the target analyte(s). A candidate procedure should produce no more than 10% false positives. Definition of a false positive is a positive response for a sample that does not contain analytes at the detection level. Between 20 and 50 samples spiked at the detection level should be tested to establish the percentage of false positives. Between 20 and 50 samples spiked at the detection level should also be tested to establish the percentage of false negatives. It is recommended that a sufficient volume of each spiked sample be prepared to complete each test with one lot of material. Sufficient randomly selected aliquots of each spiked matrix should be analyzed by appropriate SW-846 methods to demonstrate sample homogeneity and to characterize the sample in terms of target analytes and potential interferences.

A separate study should also be conducted to establish the effect of non-target interferences. A screening procedure should produce no more than 10% false positives for a set of 20 samples that contains a 100 fold excess of interferences. Positive interferences should be selected that are chemically related to the target analytes and are environmentally relevant. Negative interferences (i.e. masking agents) should also be investigated whenever they are suspected.

Developers should also analyze three different types of samples to provide matrix-specific performance data. These samples should either be characterized reference materials or spiked matrices containing known amounts of target analytes. In either case, bulk samples should be carefully homogenized to reduce sub-sampling errors. The sample matrices should be selected to represent what is regulated under RCRA (e.g. soil, oily waste or wastewaters), not to provide the best performance data. Blanks should be analyzed with each set of samples.

Matrix-specific performance data, including detection limits and dynamic range, are gathered by analyzing ten replicate aliquots of three different sample matrices spiked at two concentrations. If spiked samples are used, suggested spiking levels are the matrix-specific detection limit and 50 times the detection limit. Positive or negative results should be reported for the low concentration samples. Results for high concentration samples should be reported as either semi-quantitative results or as positive/negative with the dilution factor used for the samples. As an alternative to establishing matrix-specific detection limits, specific spiking concentrations are provided for selected target analytes in the guidance document. The low values are normal reporting limits for routine analyses and the high value is 50 times the low concentrations. The Methods Team recognizes that it may not be appropriate to spike all of the target analytes listed within a chemical class.

If the developer has field data, the Methods Team would welcome the opportunity to compare the results obtained using the screening procedure with sample concentrations determined in a laboratory using SW-846 or other appropriate methods.

To summarize, the Methods Team does not require an unreasonable body of data for the initial evaluation of new techniques. Data will need to be submitted on the percentage of false negatives, percentage of false positives, sensitivity to method interferences, and matrix-specific performance data. In addition to these data, the developer should also provide a description of the procedure and a copy of any instructions provided with the test kits.

4.2 Validation Criteria for Immunoassay Methods

4.2.1 Screening Methods

In addition to the guidelines for developing screening methods in general, OSW, based on its own experience, has generated some validation criteria specifically applicable to immunoassay methods.[70,71] These validation criteria, based on the United States Food and Drug Administration (USFDA) 510 K guidelines, are required to be submitted to OSW for review for all immunoassay test kits, whether the kits are to be the basis for a new method or as an alternative kit being added to existing methods. The data needed for validation of immunoassay methods that will be included directly in the method is as follows:

1. cross reactivity with similar analytes;
2. cross reactivity with dissimilar analytes which may be reasonably expected to be found at waste sites;
3. false negative/false positive rates;

4. extraction efficiency (for soil test kits);
5. performance data on spiked samples in environmental matrices validated against standard SW-846 analytical methods;
6. performance data on actual environmental field samples validated against standard SW-846 analytical methods.

Since interferences can be a major problem in environmental analyses, it is important to demonstrate that the analytes of concern can be identified in the presence of similar analytes or dissimilar analytes which may be present in environmental samples. In many instances, substantial cross reactivity with other analytes is a desirable situation. Examples of desirable cross reactivity include sensitivity to esters of 2,4-D as well as the 2,4-D, and for other 3-, 4-, and 5-membered PAHs when testing for phenanthrene in a PAH screening method.

The false negative/false positive rate for a particular immunoassay kit is very important. OSW screening methods are designed to generate 0% false negatives and up to 10% false positives at the regulatory action level. Slightly higher false positive rates are tolerable, e.g. up to 25%. High false positive rates, i.e. >25%, negate the cost effectiveness of the technique because of the excessive numbers of confirmatory tests that would need to be performed. High false negative rates, i.e. >5% at the regulatory action level eliminate the potential use of the method for regulatory purposes.

The extraction efficiency data are important for setting the appropriate action level for a soil analysis. Recoveries are the primary determining factor for making sure that the analyte of concern can be detected at the regulatory action level and for minimizing false negative/false positive rates. Extraction efficiencies need not be quantitative, because of the sensitivity of the antibodies used. However, they must be consistent and reproducible.

The performance data generated from environmental samples spiked with the target analytes give a good indication as to whether or not an immunoassay method will work. However, the performance generated in the field on real environmental samples is the key determining factor on whether or not the immunoassay method is sufficiently rugged to be included in SW-846 as an analytical method.

Additional data that OSW requests, but does not include in the method and treats as confidential business information (CBI), includes dosage curves and the manufacturer's internal validation and quality control criteria. The slope of the dosage curve can be a good indication of whether or not an immunoassay method will exhibit a high rate of false positives.

Up to this time, all of the immunoassay test kits (>25) that the OSW has evaluated have been extensively tested and validated by the manufacturers. USEPA validation has primarily consisted of confirmation of the manufacturers' results and performing some additional testing on well-characterized environmental samples, which are more easily available to USEPA regional laboratories.

4.2.2 Quantitative Methods

OSW has issued a guidance document for the validation of quantitative methods, ''Guidance for Methods Development and Methods Validation for the RCRA Program'',[69] describing the key elements that need to be met from the Proof of Concept stage through single and multilaboratory validation. The document is also available on the Internet from the OSW Methods Team Homepage at ''http://www.epa.gov/SW-846''. Quantitative immunoassay methods would be validated using the same criteria as any other quantitative method.

4.2.3 Revalidation Issues

Based on the successful incorporation of ten immunoassay methods involving more than 25 validated individual test products, the issue has come up as to what type of revalidation of a previously validated method is necessary if that method is changed. Once again, we defer to the USFDA for guidance on this issue. The guidelines that will be included in the revised Immunoassay Screening Methods Development Guidance Document are that for non-substantive changes, i.e. changes that do not affect the basic chemistry of the method, no revalidation will be necessary. An example of this type of change would be change of an inert washing solution in an ELISA kit.

If the manufacturer decided to change the format of the product, e.g. from a tube reaction to a home pregnancy test kit type format without changing the chemistry, the manufacturer would need to do a partial revalidation. This partial revalidation would need to be sufficient to demonstrate that the format change has not adversely affected or changed the chemistry and that the new kit gives the same results as the old one on split samples.

If the manufacturer changes the chemistry in an existing product, this would be considered as a new product and would require a new validation.

5 STATUS OF THE UNITED STATES ENVIRONMENTAL PROTECTION AGENCY IMMUNOASSAY METHODS DEVELOPMENT PROGRAM

5.1 Current Validated Methods

The first eleven immunoassay methods (see Table 1) were formally approved for incorporation into SW-846

by regulation in June, 1997. The generic Immunoassay method (Method 4000) and the ten individual ELISA methods may be used for any analytical application for which they can be demonstrated to be appropriate. No regulatory barriers remain to prevent the use of these methods.

The five new methods included in Table 2 have been formally validated and accepted by OSW's Technical Workgroup. However, until the regulations removing the requirements to adopt new methods through regulatory action are removed (expected to be completed by the end of 2000), the use of these methods may be limited by some regulatory restrictions. However, for most of the applications for which they are intended to be used, there are no actual regulatory barriers, only perceptions.

Method4425, developed by Columbia Analytical Services (CAS),[42–50] is a gross screening method using a reporter gene on a human liver cell line to screen for PAHs, PCBs and dioxins. The method can distinguish between PAHs and chlorinated compounds by differences in reaction time. The method can give useful semiquantitative information only on well characterized sites in the monitoring mode.

Method 4500, developed by BioNebraska,[72,73,92] is the first USEPA immunoassay method for a metal, mercury. It provides semiquantitative screening data for mercury extracted from soils with acid and followed by a colorimetric determination using ELISA.

Methods 4655 and 4656 are immunosensor probe methods for TNT and RDX explosives developed by the Naval Research Laboratory (NRL)[51–68,81,82] using previously developed ELISA antibodies used in the earlier test kit methods (see Methods 4050 and 4051 in Table 1) and detection using a flow cell technique based on performing a fluorescent displacement immunoassay on a membrane or other solid support in flow or a fiberoptic technique based on performing a competitive fluorescent immunoassay on the surface of a fiber optic probe.

Table 2 Next generation of methods validated for use in RCRA programs

Method 4425:	Screening Extracts of Environmental Samples for Planar Organic Compounds (PAHs, PCBs, Dioxins/Furans) by a Reporter Gene on a Human Cell Line[a]
Method 4500:	Mercury in Soil by Immunoassay[a]
Method 4655:	Explosives Analysis in Soil and Water Using Environmental Immunosensors[a]
Method 4656:	Explosives Analysis in Soil and Water Using Fiber-optic Immunosensors[a]
Method 4670:	Triazines as Atrazine in Water by Immunoassay[a]

[a] Draft methods. Can be used for applications where the use of promulgated SW-846 methods is not necessary.

Method 4670[74–78,80] developed by Ohmicron (SDI, Inc.) is a competitive ELISA immunoassay method for the quantitative determination of triazine herbicides as atrazine in drinking water. This method will be discussed in more detail in section 6.5.

Copies of the methods in Table 2 are available now or will be shortly on the OSW Methods Team Homepage at "http://www.epa.gov/SW-846".

5.2 Methods under Development

The six methods listed in Table 3 are in various stages of development. Methods 4025 for dioxin and 4026 for coplanar PCBs are ELISA methods from Cape Technologies and should have their validations completed by the end of 1999. The validation of the DELFIA method for dioxin TEQ, Method 4430, from Hybrizyme should also be completed around the same time. The other three methods, Method 36xx for affinity chromatography, Method 4016 for 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) and Method 4060 for TCE are all only in the beginning stages of development at this time.

Dioxin is a general term that describes a group of highly toxic chemicals that are extremely persistent in the environment. These compounds have been of major interest to immunoassay method developers. There are 210 different dioxin and furan congeners, 17 of which have been determined to be toxic. Each toxic congener has been assigned a toxic equivalency factor (TEF) which expresses how biologically active it is as compared to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), the most toxic congener. Studies have demonstrated that the toxicity of the individual congeners is additive. The additive effects of these compounds are expressed as total TEQ in comparison to TCDD.[85,86]

The scientific community has now identified a common pathway responsible for most, if not all, of the harmful effects associated with exposure to dioxin and related compounds. The interaction of dioxin-like compounds with a cellular protein known as the "Aryl hydrocarbon receptor" or "Ah receptor" represents the first step in a series of events that ultimately alters normal physiological processes.[87] Exposure to dioxin-like molecules has been

Table 3 New methods under development

Method 36xx:	Affinity Chromatography
Method 4016:	2,4,5-T in Water by Immunoassay
Method 4025:	Dioxin in Water and Soil by Immunoassay
Method 4026:	Coplanar PCBs in Water and Soil by Immunoassay
Method 4060:	TCE in Soil by Immunoassay
Method 4430:	Dioxin TEQ in Water and Soil by DELFIA

TCE, trichloroethylene.

linked to a variety of illnesses including cancer, harmful reproductive and developmental effects, immunotoxicity, diabetes, and endometriosis.[88]

Discovery of the Ah receptor provided a key to understanding the molecular mechanism of dioxin toxicity. The Ah receptor resembles the broad category of steroid hormone receptors and like those receptors, it regulates gene transcription. The receptor is capable of a high degree of structural discrimination between molecules and of transducing signals at very low concentrations. The toxicity of the 17 dioxins and furans results from changes in the expression of critical genes following binding and activation of the Ah receptor. Therefore, biological systems respond to the cumulative exposure of Ah receptor-mediated chemicals rather than to the exposure to any single dioxin-like compound, thus providing the molecular basis for the empirically derived TEF and TEQ value system.[89]

Several different approaches have been used for screening of toxic dioxins and furans by immunoassay. These include the reporter gene approach (Method 4425) described in the previous section, ELISA (Method 4025) and DELFIA (Method 4430).

Cape Technologies' Method 4025 (ELISA) is a screening method that measures TEQ of dioxins and furans in soil and water extracts. It does involve a detailed sample preparation procedure, unlike the test kits for simpler target analytes. Therefore it is more suitable for laboratory use than for field use. The method is readily sensitive to 500 ppt TEQ in soils.[90]

Hybrizyme's dioxin assay (Method 4430) is based upon a recombinant version of the Ah receptor gene which is used in the assay to accurately measure dioxin-like compounds. The DNA sequence adjacent to the Ah receptor gene has been modified so that the recombinant protein can be manufactured in large quantities and easily incorporated into an immunoassay format. Accordingly, the assay embodies 'nature's perfect device' for detecting the presence of these compounds and provides the user with a 'risk-based' approach to TEQ analysis. In conjunction with the Ah receptor-based assay system, Hybrizyme has developed sample processing procedures that will allow the test to selectively detect dioxins and furans, co-planar PCBs, or carcinogenic PAHs such as benzo[*a*]pyrene.

The Method 4430 dioxin assay is noncompetitive by design. Almost all immunoassays that detect small molecules are competitive in nature. The sensitivity of noncompetitive assays increases with lower

Table 4 Immunoassay test products validated by OSW

	Manufacturer						
Method #	**Beacon**	**BioNebraska**	**Columbia**	**EnSys**	**Millipore**	**Ohmicron**	**SDI**
Analyte	(SDI)		**Anal. Serv.**	(SDI)	(SDI)	(SDI)	
4010				Soil & Water	Soil & Water	Soil	
PCP							
4015					Water & Soil	Water & Soil	
2,4-D							
4020	Soil			Soil	Soil	Soil	Soil
PCB				Oil			
4030				Soil	Soil	Soil (SCHC)	
TPH					Soil (SCHC)		
4035	Soil (A/CPAH)			Soil (APAH)	Soil (APAH)	Soil (APAH)	
PAH						Soil (CPAH)	
4040					Soil		
Toxaphene							
4041					Soil		
Chlordane							
4042					Soil		
DDT							
4050					Soil	Soil	Soil
TNT							
4051							Soil
RDX							
4425			Soil				
PAH, PCB, Dioxins			Gross Screen				
4500		Soil					
Mercury							
4670						Water	
Triazines							

APAH, anthracene polynuclear aromatic hydrocarbons; CPAH, carcinogenic polynuclear aromatic hydrocarbons; SCHC, short-chain hydrocarbons.

immunoreactant concentrations since, at low concentrations, a variation in the number of competing molecules has a larger impact on the interaction with the labeled species. However, according to the law of mass action, the low antigen and antibody concentrations considerably reduce the rate of complex formation and, at very low concentrations, the accuracy tends to be poor. As a non-competitive assay, the Ah receptor test will utilize a large excess of receptor to obtain a maximum signal for the compound being tested. Even at very low concentrations of molecules, a high fraction will react if the receptor is added in excess. Although not practically achievable, the theoretical limit of the detection for this system is one molecule.[91]

The Hybrizyme DELFIA assay provides an ultra-sensitive measurement system for detecting dioxin-like molecules that interact with the Ah receptor. By utilizing molecular cloning methods and clinical instrumentation, the method provides the analytical laboratory with a product containing built-in QA/QC routines, retained standard curves, LIMS compatible data-reduction, and the ease-of-use demanded in state-of-the-art clinical laboratories.

Cape Technologies has also developed an ELISA-based immunoassay for coplanar PCBs (Method 4026)[93] similar to its dioxin immunoassay (Method 4025). The method is sensitive to the 14 dioxin-like PCB congeners and is in the process of field testing to complete the final validation for inclusion in SW-846.

5.3 List of Validated Immunoassay Test Products

Table 4 provides a list of individual immunoassay test products that have been validated by OSW and were commercially available. Due to the merger of the four major manufacturers in 1997 and a reorganization of product lines, some of the listed ELISA kits may have been discontinued.

6 ENVIRONMENTAL APPLICATIONS

6.1 Advantages and Disadvantages of Immunoassay Methods

The use of immunoassay methods provides many advantages to the analyst in solving environmental problems. Some of the significant advantages derived from using this technology include selectivity, sensitivity, portability, rapid turnaround time, improved data quality, and overall cost effectiveness. Immunoassay methods by their very nature and design are very selective in their scope of applicability, i.e. they are designed to be very specific for the analysis of either an individual compound, e.g. 2,4-D or toxaphene, or an individual class of compounds, e.g. PAHs or PCBs. Compounds other than the target analytes which may be present on a potential site do not react under the conditions and reagents used for the particular methods.

Sensitivity is another advantage of immunoassay technology. Many target analytes can be detected in the ppm range or low ppb range, if necessary. Many commercial products can be obtained or prepared to operate in several ranges, e.g. 1, 10 and 100 ppm, allowing for useful site mapping to determine which areas of a contaminated site are in need of cleanup and which are not. Some kits can operate both at the low ppb range and with appropriate dilution in the ppm range. An example is the 2,4-D kit (Method 4015) developed by Ohmicron which was designed to be sensitive to drinking water MCLs of 18 ppb, and could also be used with appropriate dilution for RCRA toxicity characteristic (TC) analysis at a 10-ppm regulatory limit.

Portability is another advantage of immunoassay technology. Most immunoassay products are sold in kit form and utilize very small apparatus, both for sample handling and in the colorimeter used for detection. Thus, it is very easy to take these kits directly to a field site for use or to an on-site mobile laboratory. However, when working directly on a field site with immunoassay methods, the operator must be very careful to follow the manufacturer's instructions on the package insert as to operating parameters, particularly with respect to effective temperature range. If the method is performed outside of the manufacturer's specified temperature range, i.e. if performed at 10 °C when the specified range is 15 °C to 30 °C, the method will not work and will give erroneous data. Most kits that are operated outside of their designated temperature range are usually designed to give positive results, i.e. potential false positives.

Rapid turnaround time is another advantage of immunoassay technology. Most samples can be run in less than an hour directly on-site. This is particularly useful in a remediation activity where soil is being removed and the site manager needs to know when the cleanup level has been achieved. At this point the excavation can be demonstrated to have been completed. This can result in a considerable cost savings over the conventional approach of sending a few samples out to a laboratory and waiting several days for the analytical results to come back. By this time, a great deal of unnecessary expenditure for excavation or equipment demurrage may have occurred.

Use of immunoassay technology can also improve the overall data quality of a remediation project. The rapid turnaround time and significantly lower cost per sample over conventional laboratory methods, e.g. gas chromatography/mass spectrometry (GC/MS) or gas chromatography/electron capture detection (GC/ECD),

can result in the ability to run a much larger number of samples during the remediation project. The data from these additional samples provide a much better characterization of the site and the progress of the remediation than would the data from the relatively few conventional samples that may normally be analyzed.

The use of immunoassay technology can be very cost effective compared to conventional technologies for all of the reasons previously mentioned in this section. However, since immunoassay methods are used primarily as quantitative screening methods, one still needs to do appropriate confirmation analysis using conventional analytical methods to support the immunoassay data. OSW recommends that approximately 5 to 10% of the immunoassay samples that generate negative results be confirmed by conventional laboratory techniques, depending on the number of total samples that are run for the particular project. Confirmation of positive samples may be done on a routine basis or as needed.

However, despite the many advantages of using immunoassay methods, there are a few disadvantages as well. The primary disadvantage is that, because they are very selective in nature, they are not effective tools for characterization of an unknown site. The most effective use of immunoassay methods is in the monitoring mode for a well-characterized site, rather than in the survey mode when one is trying to determine what compounds are the primary contaminants of the site.

Another disadvantage is that for class-specific methods, the immunoassay kit does not differentiate between the individual analytes present. Also depending on the nature of the antibody employed, the specific method may not be sensitive to all compounds within that particular class. Some of the early PAH kits (Method 4035) who had antibodies designed on phenanthrene were not particularly sensitive to PAHs containing more than four rings. An additional kit sensitive to benzo[*a*]pyrene was needed to analyze for PAHs containing 5 to 7 rings. A second generation kit with an antibody targeted to a four-membered ring (chrysene) now covers the entire range. In addition, some of the early PCB kits (Method 4020) were designed around the congeners present in Aroclor 1248 or Aroclor 1254 and did not respond well to the less chlorinated congeners found in Aroclor 1242. Thus, it is important to have a good site characterization prior to employing immunoassay methods in the monitoring mode.

Another factor of which an operator needs to be aware is the type of potential matrix interferences that may be present on a site to prevent an immunoassay method from performing. Five per cent of an oil could prevent most soil PCB kits from working, except for the one product that was specifically designed to perform in an oily matrix. Moisture levels and pH can also prevent some kits from performing as expected by the user or claimed by the manufacturer. The precautions necessary to overcome these problems are usually included on the package inserts from the manufacturers.

A peripheral issue is the extent of training necessary for an operator to become proficient in the use of the immunoassay kits. Early on, the sales push was that a minimally trained operator could take a kit, maybe read the instructions, and then go out into the field and start performing analyses. In reality, this was not the case at all. Operators need to be trained well in the nuances of handling immunoassay kits. A non-scientist, in particular, needs to be educated very intensely as to which steps in the procedure are critical and must not be changed. In the validation studies for the first 10 ELISA methods, OSW found that analytical chemists, by far, obtained the best analytical results.

6.2 Implementation Issues

There have been some initial barriers to getting immunoassay methods accepted for routine use in the environmental community. These barriers have been both technical and cultural in nature. The technical barriers include lack of knowledge about analytical options; use of expensive time-consuming methodology when more efficient methodology is available; poor planning of the initial analytical scheme; failure to identify proper questions to be answered which can result in generation of data inappropriate to address the problem at hand. Cultural barriers include inappropriate or excessive regulatory restrictions on use of new methods, e.g. requiring the use of only promulgated methods for program applications that do not have these requirements, and requiring the use of expensive broad-scope methods, e.g. GC/MS, for limited monitoring applications for only a few known and well-characterized analytes.

An additional issue of concern was whether the Regulatory Program Offices could live with analytical values that were not a specified number, i.e. a less than value (usually the regulatory action level) vs. a definite number (0.1 ppm) or a range of values (>5 and <50). We, in the RCRA Program, decided that we could indeed use these values to answer the basic questions for which these analyses were performed, i.e. Have we attained our cleanup criteria? Where do we have to focus our cleanup efforts? We decided that our normal operating procedures for confirming quantitative screening results would be to use the standard reference method to confirm positives and to spot check a certain percentage (usually 10%) of negative results.

Other Program Offices in USEPA, such as the OW may have some restrictions in their current regulations which require them to generate a definite analytical value. The

Office of Groundwater and Drinking Water (OGWDW) has decided to include quantitative immunoassay screening methods in their regulations, the first of which will be Method 4670 for triazines as atrazine. This potential application is described in section 6.5.

OSW has initiated a major effort to educate USEPA permit writers, enforcement people, and others who deal with analytical methods in their jobs as to what the regulatory aspects of using RCRA methods really are. Many of these people were not aware that RCRA regulations did not require the use of only promulgated methods for most applications. The Methods Team has developed a formal training program for RCRA personnel in the Regions and at Headquarters to make them aware of what methods the regulations really allow them to use for RCRA applications, and how to plan an efficient, cost effective sampling and analysis plan.

State programs are a little more difficult. Since RCRA is a Federal Program which has been passed down to most States to administer, the State regulations can be more restrictive and tend to vary greatly. Some States mandate the use of SW-846 methods for all RCRA analytical applications within the State. Flexibility within State Programs varies from allowing only the use of promulgated methods to using any method that may be appropriate for an application. Through dialogue with the USEPA Regions and Headquarters, some of the States are beginning to take an interest in utilizing immunoassay methods. TPHs analysis is the major focus right now in State Programs, since it is not regulated at the Federal level. Several States are beginning to adopt Method 4030 for use in their Underground Storage Tank (UST) Programs, e.g. Georgia and California.

6.3 Examples of Performance Based Measurement Systems

The USEPA has established a policy that, where possible, the Agency will use the PBMSs approach in its monitoring programs for generating environmental data.[83] The Agency defines PBMS as a set of processes wherein the data quality needs, mandates or limitations of a program or project are specified, and serve as criteria for selecting appropriate methods to meet those needs in a cost-effective manner.[94]

The PBMS approach has been the basis of the RCRA monitoring program since its inception back in 1980. The inherent flexibility of this approach is an absolute necessity for a program that has responsibility for overseeing analyses in a wide variety of diverse media. The approach is very simple in nature. Rather than focusing on a prescriptive method approach, where following a published method exactly as written is the main requirement, the PBMS approach focuses on a series of questions to be answered about a specific application. These questions are based on the scientific method and focus on whether a selected method or methods can provide performance appropriate to address the data quality requirements of a particular project. The questions include:

1. What is the purpose of this analysis? (Why are we doing this analysis?)
2. How (for what action) are the data generated from this analysis to be used?
3. What are the data quality needs for this project, i.e. how good does the data have to be for it to be useful for its intended purpose? (Including regulatory drivers, target analytes, matrices, concentration levels, statistical confidence levels, etc.)

Immunoassay methods are the ultimate tools for demonstrating and using this PBMS approach. They are designed and manufactured around specific performance criteria, i.e. methods appropriate for use in Method 4020 are capable of determining whether a soil sample contains PCBs above a concentration of 5 ppm to a 95% upper confidence limit around the mean. There are currently four validated immunoassay products in Method 4020 that can meet these stated manufacturers' performance criteria, all of which utilize different chemistries.

6.4 Applications as Screening Methods

6.4.1 In the Field

OSW decided to take a cautious approach to the introduction of a new technology to the environmental field, with which most analytical practitioners were unfamiliar, and limit the initial applications of immunoassay methods to quantitative screening. We were aware that the technique had been used in Clinical Laboratories for many years in both screening and determinative applications. Since Regulatory Agencies tend to be slow to accept new and different approaches to analysis, anyway, we decided to take a ''walk before you run'' approach to introducing the new methodology to the people actually doing site assessments and cleanups.

The two primary applications of immunoassay methods in the RCRA Program are mapping of contamination at well-characterized sites slated for cleanup and monitoring the effectiveness of the cleanup activities. Immunoassay lends itself very well to these two particular applications. It is not particularly applicable to the identification and characterization of unknown contaminants at waste sites when compared to much more comprehensive techniques such as GC/MS. However, the reporter gene method (Method 4425) for planar aromatic compounds can be

used to determine the presence of and approximate concentrations of PAHs and chlorinated planar organics and distinguish between them. For monitoring applications of known contaminants, the specificity, sensitivity, and cost effectiveness of quantitative screening immunoassay methods are excellent.

Over the past few years, the general acceptability and willingness to use immunoassay methods within the USEPA Regions for RCRA and Superfund applications has increased exponentially. A significant factor in this change of attitude, in addition to OSW's attempts to educate users in the applicability of the technique, is the specter of shrinking budgets. Field people who are charged with actually doing cleanups are looking for more cost effective ways to do their jobs with less available money. A technique, such as immunoassay methodology, which can generate results of known quality in real-time, and can keep the bulldozers rolling can contribute significantly to reducing the costs of cleanups, and is being looked upon more favorably.

The initial application of immunoassay technology in the RCRA Program was for determining compliance at wood surface treating facilities with PCP regulatory limits. The selectivity and sensitivity of the immunoassay method easily met the regulatory action limit of 0.1 ppm. Use of the PCP immunoassay method (Method 4010) for compliance monitoring was encouraged by OSW and the method was added to SW-846 as a part of the Wood Surface Treatment Rule in 1994.

The major applications for which immunoassay methods are currently used in the RCRA Program are site mapping and monitoring cleanups at sites contaminated with PCBs. Use of the PCB method (Method 4020) has resulted in cost savings at many sites in several Regions. The speed and low cost of the test allows for more extensive mapping of contamination at a site, because many more samples can be analyzed on-site, thus generating a more detailed map of the site. This results in lower cleanup costs, since the cleanup efforts can be directed only at the places that need to be cleaned up, instead of to a broader area. The design of the method allows for rapid determination of whether or not the site cleanup level has been met, thus reducing costs of cleanup in both time and equipment. With the availability of the PAH method (Method 4035) and several types of kits with various sensitivities, the technique is now beginning to be used to monitor sites contaminated with PAHs.

Another major application within OSWER is for mapping and cleanup of sites contaminated with petroleum hydrocarbons from leaking storage tanks partially for the Office of Underground Storage Tanks (OUST) and primarily for the States. The TPH method (Method 4030) is effective for determining gasoline, diesel, kerosene, and jet fuel at most State-required cleanup levels.

The recent Department of Defense cutbacks set into motion a large number of military base closures in the United States, resulting in a plethora of site cleanup projects. The primary analytes of concern for these base closure cleanups were explosives, petroleum hydrocarbons and PCBs, all of which could be monitored using immunoassays and other screening techniques. The immunoassay ELISA methods for TNT (Method 4050) and RDX (Method 4051) have been used extensively along with the immunosensor methods (Methods 4655 and 4656) and the colorimetric screening Methods 8010 (for RDX) and 8015 (for TNT).

The ELISA method for mercury (Method 4500) has been used effectively to monitor the progress of the soil cleanups of several Superfund sites contaminated with mercury.

6.4.2 In the Laboratory

Immunoassay methods have been used in many laboratories to screen samples for high levels of contaminants to prevent instrument downtime. The dioxin screening method (Method 4025) was primarily designed for use in a laboratory, rather than in the field, because of the detailed sample preparation procedures involved. An effective dioxin screening method can result in significant cost savings for monitoring dioxin-contaminated sites. A negative immunoassay screening result can eliminate the need for a high resolution GC/MS analysis costing between $1500 and $2000 each. The method has been validated on a set of 56 real world soil samples and gave results that were 91% correct with 9% false positives. No false negatives were reported. It has also been used to screen river sediment samples and food samples potentially contaminated with dioxin.

6.5 Applications as Quantitative Methods

At the time of writing, there are very few quantitative applications for immunoassay methods. There are a few being developed. USEPA's Drinking Water Program is planning to incorporate Method 4670 for triazines as atrazine into its screening regulations. Current regulations require that drinking water suppliers test their supplies quarterly for atrazine at an MCL of 3 ppb. Method 4670 is appropriate for this measurement requirement, because it can measure these compounds sufficiently below the 3 ppb MCL to be effective. Since it also has a positive crossreactivity to other triazines, e.g. simizine and the like, the data generated as atrazine would be an overestimation and thus would be a conservative measurement.

Some of the more selective screening methods, e.g. TNT, RDX, and 2,4-D were originally submitted as screening methods for reasons described previously.

However, they are sufficiently selective to allow their use as quantitative methods. The TNT methods have been used for some quantitative applications on military base cleanups. They tend to generate high results with respect to TNT, because they also exhibit significant (and desirable) crossreactivity to TNT breakdown products such as DNT and DNB.

7 FUTURE DIRECTIONS

The environmental immunoassay method program has come a very long way, from nonexistent in early 1992 to a significant and viable program with a variety of available methods utilizing several different immunoassay technologies today. However, we still intend to continue to push ahead and keep the new methods coming and remain current with the state-of-the-art.

The primary focus right now is on completing the validations of the dioxins and coplanar PCB methods in Table 3, which is expected to be done by early 2000. We are also interested in continuing with the development of additional quantitative immunoassay methods for specific target analytes. We are continuing to investigate the applicability of immunoassay in the area of affinity chromatography.

Another area of interest for continued development is that of immunosensors. Several Federal Agencies are interested in further developing this technology beyond the realm of explosives. We are in the process of planning some preliminary projects, beginning in 2000, to complete the development of new immunosensors for PAHs, PCBs, and TCE.

Now that the industry has sorted itself out after the mergers of 1997, there are a number of small, energetic companies with a broad base of expertise in a wide variety of immunoassay techniques. We, in OSW, plan on continuing our productive relationship with these industrial partners to bring the best of immunoassay technology to bear on future environmental problems.

ACKNOWLEDGMENTS

The author would like to acknowledge Gail Hansen of the OSW Methods Team for her help in proofreading the manuscript and to several contributors of technical documents that helped him a great deal by providing him with the information that he needed to finish certain sections of this chapter. These contributors include Bob Harrison of Cape Technologies, Randy Allen of Hybrizyme, Jack Anderson of Columbia Analytical Services, and the other developers of immunoassay methods with whom I have worked over the years, particularly Steve Friedman of Sylvanus Laboratories, who has been my primary mentor in immunoassay technology over the years.

ABBREVIATIONS AND ACRONYMS

APAH	Anthracene Polynuclear Aromatic Hydrocarbons
B[*a*]PEqs	Benzo[*a*]pyrene Equivalents
CAA	Clean Air Act
CBI	Confidential Business Information
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CPAH	Carcinogenic Polynuclear Aromatic Hydrocarbons
CWA	Clean Water Act
DELFIA	Dissociation-enhancement Lanthanide Fluoroimmunoassay
ELISA	Enzyme-linked Immunosorbent Assay
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
MCL	Maximum Contaminant Level
OAR	Office of Air and Radiation
OARM	Office of Administration and Resource Management
OERR	Office of Emergency Response and Remediation
OGWDW	Office of Groundwater and Drinking Water
OPPTS	Office of Prevention, Pesticides and Toxic Substances
ORD	Office of Research and Development
OSW	Office of Solid Waste
OSWER	Office of Solid Waste and Emergency Response
OUST	Office of Underground Storage Tanks
OW	Office of Water
PAH	Polynuclear Aromatic Hydrocarbon
PBMS	Performance Based Measurement System
PCB	Polychlorinated Biphenyl
PCDDs	Polychlorinated Dibenzo-*p*-dioxins
PCDFs	Polychlorinated Dibenzofurans
PCP	Pentachlorophenol
RCRA	Resource Conservation and Recovery Act

RDX	Hexahydro-1,3,5-trinitro-1,3,5-triazine
RLUs	Relative Light Units
SCHC	Short-chain Hydrocarbons
SDWA	Safe Drinking Water Act
TC	Toxicity Characteristic
TCDD	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin
TCE	Trichloroethylene
TEF	Toxic Equivalency Factor
TEQ	Toxic Equivalents
TNT	2,4,6-Trinitrotoluene
TPH	Total Petroleum Hydrocarbon
TSCA	Toxic Substances Control Act
USEPA	United States Environmental Protection Agency
USFDA	United States Food and Drug Administration
UST	Underground Storage Tank
2,4-D	2,4-Dichlorophenoxyacetic Acid
2,4,5-T	2,4,5-Trichlorophenoxyacetic Acid

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Immunochemistry

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment

Environment: Water and Waste cont'd **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Quality Assurance in Environmental Analysis • Sampling Considerations for Biomonitoring

Food **(Volume 5)**
Enzyme Analysis and Bioassays in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food

Forensic Science **(Volume 5)**
Immunoassays in Forensic Toxicology

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis

Liquid Chromatography **(Volume 13)**
Affinity Chromatography

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry

REFERENCES

1. *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, 3rd edition*, (SW-846) including Updates, US Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC, 1997.
2. B. Lesnik, 'Immunoassay Methods: Development and Implementation Programme at the USEPA', *Food & Agricultural Immunology*, **6**, 251–259 (1994).
3. S.B. Friedman, 'Doing Immunoassays in the Field', *Chemtech.*, 732–737 December (1992).
4. L. Roitt, J. Brosstoff, M. Male (eds.), *Immunology*, J.B. Lippincott Co., Philadelphia, Pennsylvania, 1989.
5. D.P. Stites, A.I. Terr (eds.), *Basic and Clinical Immunology*, Appleton and Lange, Norwalk, Connecticut, 1991.
6. W.D. Odell, W.H. Daughaday, *Principles of Competitive Protein-binding Assays*, J.B. Lippincott Co., Philadelphia, Pennsylvania, 1971.
7. E. Ishikawa, T. Kawai, K. Miyai (eds.), *Enzyme Immunoassay*, Igaku-Shoin, Tokyo, Japan, 1981.
8. P. Tijssen (ed.), *Practice and Theory of Enzyme Immunoassays*, Elsevier, NY, Vol. 15, 1985.
9. J.E. Butler (ed.), *Immunochemistry of Solid-phase Immunoassay*, CRC Press, Boca Raton, Florida, 1991.
10. T.T. Ngo, H.M. Lenhoff, *Enzyme-mediated Immunoassay*, Plenum Press, New York, 1985.
11. J.P. Mapes, K.D. McKenzie, L.R. McClelland, S. Movassaghi, R.A. Reddy, R.L. Allen, S.B. Friedman, *Rapid, On-site Screening Test for Pentachlorophenol in Soil and Water – PENTA-RISc™*, Ensys Inc., Research Triangle Park, NC 27709, 1992.
12. J.P. Mapes, K.D. McKenzie, L.R. McClelland, S. Movassaghi, R.A. Reddy, R.L. Allen, S.B. Friedman, *PENTA-RISc™ – An On-site Immunoassay for Pentachlorophenol in Soil*, Ensys Inc., Research Triangle Park, NC 27709, 1992.
13. *PENTA-RISc™ Instructions for Use*, Ensys Inc., 1992.
14. *EnviroGard™ PCP in Soil Test Kit Guide*, Millipore, Inc., 1993.
15. *2,4-D RaPID™ Assay kit Users Guide*, Ohmicron, 1993.
16. *EnviroGard™ 2,4-D in Soil Test Kit Guide*, Millipore, Inc., 1993.
17. M.C. Hayes, S.W. Jourdan, T.S. Lawruk, D.P. Herzog, 'Screening of TCLP Extracts of Soil and Wastewater for 2,4-D by Immunoassay', USEPA Ninth Annual Waste Testing and Quality Assurance Symposium, 1993.
18. J.P. Mapes, T.N. Stewart, K.D. McKenzie, L.R. McClelland, R.L. Mudd, W.B. Manning, W.B. Studabaker, S.B. Friedman, 'PCB-RISc™ – An On-site Immunoassay for Detecting PCB in Soil', *Bull. Environ. Contam. Toxicol.*, **50**, 219–225 (1993).
19. *PCB RISc™ Users Guide*, Ensys Inc., 1993.
20. R.W. Counts, R.R. Smith, J.H. Stewart, R.A. Jenkins, 'Evaluation of PCB Rapid Immunoassay Screen Test System', Oak Ridge National Laboratory, Oak Ridge, TN 37831, April 1992, unpublished.

21. *EnviroGard PCB in Soil Package Insert*, Millipore Corp., 1993.
22. Technical Evaluation Report on the Demonstration of PCB Field Screening Technologies, SITE Program. USEPA Contract Number 68-CO-0047, 1993.
23. *D TECH™ PCB Users Guide*, SDI/Em Sciences, 1993.
24. J.M. Melby, B.S. Finlin, A.B. McQuillin, H.G. Rovira, J.W. Stave, 'PCB Analysis by Enzyme Immunoassay', Strategic Diagnostics Incorporated, Newark, Delaware, 1993.
25. J.M. Melby, B.S. Finlin, A.B. McQuillin, H.G. Rovira, 'Competitive Enzyme Immunoassay (EIA) Field Screening System for the Detection of PCB', 1993 PCB Seminar, EPRI, 1993.
26. *PCB RISc™ Liquid Waste Test System*, User's Guide, EnSys Environmental Products, Inc., 1993.
27. *PETRO RISc™ Users Guide*, Ensys Inc., 1993.
28. P.J. Marsden, S.-F. Tsang, N. Chau, *Evaluation of the PETRO RISc™ kit Immunoassay Screen Test System*, Science Applications International Corporation under contract to EnSys Inc., June 1992, unpublished.
29. *EnviroGard™ Petroleum Fuels in Soil Test Kit Guide*, Millipore, Inc., 1993.
30. *PAH-RISc™ Users Guide*, EnSys Inc., 1993.
31. P.P. McDonald, R.E. Almond, J.P. Mapes, S.B. Friedman, 'PAH-RISc™ Soil Test – A Rapid, On-site Screening Test for Polynuclear Aromatic Hydrocarbons in Soil', *JAOAC Intern.* (accepted for publication document #92263).
32. R.P. Swift, J.R. Leavell, C.W. Brandenburg, 'Evaluation of the EnSys PAH-RISc™ Test Kit', Proceedings, USEPA Ninth Annual Waste Testing and Quality Assurance Symposium, 1993.
33. *EnviroGard™ Toxaphene Soil Test Kit Guide*, Millipore, Inc., 1994.
34. P.J. Marsden, S.-F. Tsang, V. Frank, N. Chau, M. Roby, 'Comparison of the Millipore Immunoassay for Toxaphene with Soxhlet Extraction and Method 8081', Science Applications International Corporation, under contract to Millipore Inc., May 1994, unpublished.
35. *EnviroGard™ Chlordane in Soil Test Kit Guide*, Millipore, Inc., 1994.
36. *EnviroGard™ DDT in Soil Test Kit Guide*, Millipore, Inc., 1994.
37. *D TECH™ TNT Users Guide*, SDI/EM Sciences, 1994.
38. L. Hutter, G. Teaney, J.W. Stave, 'A Novel Field Screening System for TNT Using EIA', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Proceedings of the 1993 USEPA/A&WMA International Symposium, 472, Vol. 1, 1993.
39. G. Teaney, J. Melby, L. Hutter, J. Stave, 'A Novel Field Analytical Method for TNT', Proceedings of the American Association of Analytical Chemists, 1993.
40. R.J. Haas, B.P. Simmons, 'Measurement of Trinitrotoluene (TNT) and Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) in Soil by Enzyme Immunoassay and High Performance Liquid Chromatography (USEPA Method 8330)', California Environmental Protection Agency, Department of Toxic Substances Control, Hazardous Materials Laboratory, March, 1995.
41. *D TECH™ TNT Users Guide*, SDI/Em Sciences, 1994.
42. J.W. Anderson, S.S. Rossi, R.H. Tukey, T. Vu, L.C. Quattrochi, 'A Biomarker, 450 RGS, for Assessing the Potential Toxicity of Organic Compounds in Environmental Samples', *Environ. Toxicol. Chem.*, **14**(7), 1159–1169 (1995).
43. J.W. Anderson, K. Bothner, T. Vu, R.H. Tukey, 'Using a Biomarker (P450 RGS) Test Method on Environmental Samples, in *Techniques in Aquatic Toxicology*, ed. G.K. Ostrander, Lewis Publishers, Boca Raton, FL, Chapter 15, 277–286, 1996.
44. J.W. Anderson, K. Bothner, D. Edelman, S. Vincent, T. Vu, R.H. Tukey, 'A Biomarker, P450 RGS, for Assessing the Potential Risk of Environmental Samples, in *Field Applications of Biomarkers for Agrochemicals and Toxic Substances*, eds. J. Blancato, R. Brown, C. Dary, M. Saleh, American Chemical Society, Washington, DC, Chapter 12, 150–168, 1996.
45. J.W. Anderson, F.C. Newton, J. Hardin, R.H. Tukey, K.E. Richter, 'Chemistry and Toxicity of Sediments from San Diego Bay, Including a Biomarker (P450 RGS) Response', in *Environmental Toxicology and Risk Assessment: Biomarkers and Risk Assessment*, 5th Volume, ASTM STP 1306, eds. D.A. Bengtson, D.S. Henshel, American Society for Testing and Materials, West Conshohocken, PA, 53–78, 1996.
46. J.W. Anderson, J.M. Jones, S. Steinert, B. Sanders, J. Means, D. McMillin, T. Vu, R. Tukey, 'Correlation of CYP1A1 Induction, as Measured by the P450 RGS Biomarker Assay, with High Molecular Weight PAHs in Mussels Deployed at Various Sites in San Diego Bay in 1993 and 1995', *Mar. Environ. Res.*, **48**(4–5), 389–405 (1999).
47. APHA. P450 Reporter Gene Response to Dioxin-like Organics. Method 8070, in *Standard Methods for the Examination of Water and Wastewater*, 19th edition Supplement, American Public Health Association, Washington, DC, 24–25, 1996.
48. ASTM. Standard Guide E 1853–96 for Measuring the Presence of Planar Organic Compounds which Induce CYP1A, Reporter Gene Test Systems, in Volume 11.05, Biological Effects and Environmental Fate; Biotechnology; Pesticides, *1997 Annual Book of ASTM Standards*, section 11 Water and Environmental Technology, American Society for Testing and Materials, West Conshohocken, PA, 1392–1397, August 1997.
49. G.B. Kim, J.W. Anderson, K. Bothner, J.-H. Lee, C.-H. Koh, S. Tanabe, 'Application of P450RGS (Reporter Gene System) as a Bioindicator of Sediment PAH Contamination in the Vicinity of Inchon Harbor, Korea', *Biomarkers*, **2**, 181–188 (1997).

50. H. Postlind, T.P. Vu, R.H. Tukey, L.C. Quattrochi, 'Response of Human CYP1-luciferase Plasmids to 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin and Polycyclic Aromatic Hydrocarbons', *Toxicol. Appl. Pharmacol.*, **118**, 255–262 (1993).
51. J.C. Bart, L.L. Judd, A.W. Kusterbeck, 'Environmental Immunoassay for the Explosive RDX Using a Fluorescent Dye-labeled Antigen and the Continuous Flow Immunosensor', *Sensors and Actuators B*, **38–39**, 411–418 (1997).
52. J.C. Bart, L.L. Judd, K.E. Hoffman, A.M. Wilkins, P.T. Charles, A.W. Kusterbeck, 'Application of a Portable Immunosensor to Detect the Explosives TNT and RDX in Groundwater Samples', *Environ. Sci. Technol.*, **31**, 1505–1511 (1997).
53. H. Craig, G. Furguson, A. Markos, A. Kusterbeck, L. Shriver-Lake, T. Jenkins, P. Thorne, *Field demonstration of On-site Analytical Methods for TNT and RDX in Groundwater*. Proceedings of the Great Plains – Rocky Mountain Hazardous Substance Research Center (HSRC)/Waste Education and Research Consortium (WERC) Joint Conference on the Environment, Alberquerque, NM, May 21–23, 204–219, 1996.
54. J.C. Bart, L.L. Judd, K.E. Hoffman, A.M. Wilkins, P.T. Charles, A.W. Kusterbeck, 'Detection and Quantitation of the Explosives TNT and RDX in Groundwater Using a Continuous Flow Immunosensor', in *Immunochemical Technology for Environmental Applications*, ACS Symposium Series, ed. D.S. Aga, American Chemical Society, Washington, DC, 210–220, 1996.
55. L.L. Judd, A.W. Kusterbeck, D.W. Conrad, H. Yu, H.L. Myles, Jr, F.S. Ligler, 'An Antibody-based Fluorometric Assay for Detection of the Explosives TNT and PETN', *SPIE*, **2388**, 198–204 (1995).
56. J.P. Whelan, A.W. Kusterbeck, G.A. Wemhoff, R. Bredehorst, F.S. Ligler, 'Continuous-flow Immunosensor for Detection of Explosives', *Anal. Chem.*, **65**, 3561–3565 (1993).
57. F.S. Ligler, B. Gaber, A.W. Kusterbeck, G. Wemhoff, 'Flow Immunosensor: Method and Apparatus', US Patent No. 5,183,740, 1993.
58. L.C. Shriver-Lake, N.A. Naz, F.S. Ligler, 'A Fiber Optic Biosensor for the Detection of TNT/RDX in Environmental Samples', in *Current Protocols in Field Analytical Chemistry*, ed. V. Lopez-Avila, John Wiley & Sons, New York, 1998.
59. R.M. Wadkins, F.S. Ligler, 'Immuno-biosensors based on Evanescent Wave Excitation', in *Methods in Biotechnology*, Humana Press, Totowa, NJ, 77–87, Vol. 7, 1998.
60. J.P. Golden, E. Saaski, L.C. Shriver-Lake, G. Anderson, F.S. Ligler, 'A Portable Multichannel Fiber Optic Biosensor for Field Detection', *Opt. Eng.*, **36**(4), 1008–1013 (1997).
61. B.L. Donner, L.C. Shriver-Lake, F.S. Ligler, 'Transition from Laboratory to On-site Environmental Monitoring of TNT Using a Portable Fiber Optic Biosensor', in *Immunochemical Technology for Environmental Applications*, ACS Symposium Series 657, eds. D. Aga, E. Thurman, American Chemical Society, Washington, DC, 198–209, 1997.
62. L.C. Shriver-Lake, B.L. Donner, F.S. Ligler, 'On-site Monitoring of TNT with a Portable Fiber Optic Biosensor', *Environ. Sci. Technol.*, **31**(3), 837–841 (1997).
63. K.R. Rogers, C.L. Gerlach, 'Environmental Biosensors, A Status Report', *Environ. Sci. Technol.*, **30**(11), 486A–491A (1996).
64. L.C. Shriver-Lake, K.A. Breslin, J.P. Golden, F.S. Ligler, 'Detection of TNT in Water Using an Evanescent Wave Fiber Optic Biosensor', *Anal. Chem.*, **67**, 2431–2435 (1995).
65. J.P. Golden, L.C. Shriver-Lake, G.P. Anderson, R.B. Thompson, F.S. Ligler, 'Fluorimeter and Tapered Fiber Optic Probes for Sensing in the Evanescent Wave', *Opt. Eng.*, **31**(7), 1458–1462 (1992).
66. M.D. Pease, L.C. Shriver-Lake, F.S. Ligler, 'Adaptation of a Fiber-optic Biosensor for Use in Environmental Monitoring', in *Biosensor and Chemical Sensor Technology: Process Monitoring and Control*, ACS Symposium Series 613, eds. K.R. Rogers, A. Mulchandani, W. Zhou, American Chemical Society, Washington, DC, 33–43, 1995.
67. F.S. Ligler, J.H. Georger, S.K. Bhatia, J. Calvert, L.C. Shriver-Lake, R. Bredehorst, 'Immobilization of Active Agents on Substrates with a Silane and Heterobifunctional Crosslinking Agent', US Patent #5,077,210 issued December 31, 1991.
68. S.K. Bhatia, L.C. Shriver-Lake, J. Georger, J. Calvert, K. Prior, R. Bredehorst, F. Ligler, 'Use of Thiol-terminal Silanes and Heterobifunctional Crosslinkers for Immobilization of Antibodies on Silica Surfaces', *Anal. Biochem.*, **178**, 408–413 (1989).
69. OSW Guidance Document: 'Guidance for Methods Development and Methods Validation for the RCRA Program', October, 1995, http://www.epa.gov/SW-846.
70. OSW Guidance Document: 'Immunoassay Methods in SW-846: Recommended Format and Content for Documentation Supporting New Submittals', July, 1995, http://www.epa.gov/SW-846.
71. 510 K of the Federal Food, Drug and Cosmetics Act, section 21, CFR 807.87.
72. BiMelyze® Mercury Assay Kit and BiMelyze® Mercury Assay Soil Extraction Kit, BioNebraska, Inc., 1995.
73. California Environmental Protection Agency, Department of Toxic Substances Control, Environmental Technology Certification Program, 'BiMelyze® Field Screening Assay for Mercury (''Tube Assay'') and Soil Extraction Kit, with a Partial Evaluation of BiMelyze® Laboratory Screening Test for Mercury (''Plate Assay'')', BioNebraska, Inc. Lincoln, NB; Contract No. 93-T0470, June 1995.

74. W.O. Dell, P. Franchimont, *Principles of Competitive Protein-binding Assays*, John Wiley and Sons, New York, 1983.
75. 'Immunoassay Analysis and GC/MS Confirmation for Residues of Atrazine in Water Samples from a Field Study Conducted by the State of Wisconsin', Project No. 101174, Report No. ABR-91069, CIBA-GEIGY Corporation, April 6, 1992.
76. M.C. Hayes, S.W. Jourdan, D.P. Herzog, 'Determination of Atrazine in Water by a Magnetic Particle Immunoassay: Collaborative Study', *JAOAC*, **79**(2), 530–538 (1996).
77. F.M. Rubio, J.M. Itak, A.M. Scutellaro, M.Y. Selisker, D.P. Herzog, 'Performance Characteristics of a Novel Magnetic-particle-based Enzyme-linked Immunosorbent Assay for the Quantitative Analysis of Atrazine and Related Triazines in Water Samples', *Food Agric. Immunol.*, **3**, 113–125 (1991).
78. B. Gruessner, N.C. Shambaugh, M.C. Watzin, 'Comparison of an Enzyme Immunoassay and Gas Chromatography/Mass Spectrometry for the Detection of Atrazine in Surface Waters', *Environ. Sci. Technol.*, **29**, 251–254 (1995).
79. B. Lesnik, 'Method-at-a-Glance: Method 4035-Polynuclear Aromatic Hydrocarbons in Soil by Immunoassay', *Environ. Lab.*, April/May (1994).
80. B. Lesnik, 'Method-at-a-Glance: Method 4670: A Quantitative Immunoassay Method for the Determination of Triazine Herbicides as Atrazine', *Environ. Lab.*, November (1996).
81. B. Lesnik, 'USEPA-OSW Methods Update: Method 4655: Determination of TNT and RDX with the Continuous Flow Immunosensor', *Environ. Testing Anal.*, July/August (1999).
82. B. Lesnik, 'USEPA-OSW Methods Update: Method 4656: Determination of TNT and RDX with the Fiber Optic Immunosensor', *Environ. Testing Anal.*, September/October (1999).
83. B. Lesnik, O. Fordham, 'An Update of the Current Status of the RCRA Methods Development Program', *Environ. Testing Anal.*, December (1998).
84. I. Hemmilä, 'Fluoroimmunoassay and Immunofluorometric Assays', *Clin. Chem.*, **31**, 359–370 (1985).
85. E.K. Silbergeld, 'Understanding Risk: The Case of Dioxin', *Scientific America Science and Medicine*, **2**(6), 48–57 (1995).
86. USEPA Update of Toxicity Equivalency Factors (TEFs) for Estimating Risks Associated with Exposures to Mixtures of Chlorinated Dibenzo-p-dioxins and Dibenzofurans (CDDs/CDFs). USEPA/625/3-89/016, March 1989, Risk Assessment Forum, Washington, DC, 1989.
87. A. Poland, E. Glover, A.S. Kende, 'Stereospecific, High Affinity Binding of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin by Hepatic Cytosol; Evidence that the Binding Species is Receptor for Induction of Aryl Hydrocarbon Hydroxylase', *J. Biol. Chem.*, **251**(16), 4936–4946 (1976).
88. A. Schecter (ed.), *Dioxins and Health*, Plenum Press, New York, NY, 1994.
89. L.S. Birnbaum, 'The Importance of Molecular Mechanisms in Assessing the Risk of Dioxin', *Environ. Health Perspectives*, **102**(9), 157–167 (1994).
90. R.O. Harrison, 'Application Note AN-007: Analysis of PCDDs/Fs in soil at 500 ppt Using Rapid Extraction and Rapid One-step Cleanup', Cape Technologies, August, 1999, http://www.cape-tech.com.
91. R.P. Ekins, 'Immunoassay Design and Optimisation', in *Principles and Practice of Immunoassay*, 2nd edition, eds. C.P. Price, D.J. Newman, Stockton Press, New York, NY, 173–207, 1997.
92. O. Fordham, 'Method-at-a-Glance: SW-846 Method 4500: Mercury in Soils by Immunoassay', *Environ. Lab.*, September/October (1997).
93. R.E. Carlson, R.O. Harrison, 'Demonstration of a TEQ Selective PCB Immunoassay', *Dioxin '99*, Venice, Italy, 1999.
94. *Federal Register*, 62 *FR* 52098, (October 6, 1997).

Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Shin-ichi Yamasaki
Tohoku University, Sendai, Japan

1 Introduction 1
2 History 1
3 Instrumentation 2
3.1 Instrument Overview 2
3.2 Inductively Coupled Plasma 2
3.3 Sample Introduction Systems 2
4 Sample Preparation 6
4.1 Methods of Dissolution 6
5 Environmental Applications 8
5.1 Overview 8
5.2 Detection Limits and Range 8
5.3 Water Analysis 9
5.4 Soils and Related Materials 9
5.5 Biological Samples 10
6 Method Development 10
6.1 High-resolution Inductively Coupled Plasma Mass Spectrometry 10
6.2 Multiple-collector Instruments 12
6.3 Time-of-flight Instruments 13
6.4 Collision Cell Interface Inductively Coupled Plasma Mass Spectrometry 13
6.5 Low Sample Consumption Nebulizers 13
Abbreviations and Acronyms 14
Related Articles 14
References 14

The first sections of this article describe a brief history, the basic principles and advantages as well as disadvantages of inductively coupled plasma mass spectrometry (ICPMS) over other analytical techniques commonly used for environmental analysis (see Figure 1). Section 4 is devoted to the discussion of the relative merits and demerits of a wide variety of dissolution techniques and their limitation in the pretreatment of various environmental samples. In section 5, efforts are concentrated on the novel application of ICPMS to such typical environmental samples as water, soil, and biological materials. In the last sections of the article, future perspectives of new ICPMS and related devices are discussed. The topics chosen are high resolution ICPMS, multiple-collector instruments, time-of-flight (TOF) ICPMS, collision cell interface ICPMS, and low-sample consumption nebulizers.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

In recent years, there has been ever increasing concern about the concentration levels of various trace and ultra-trace elements in environmental samples. In spite of this growing interest, the efforts of most environmental scientists appeared to be centered around limited numbers of typical pollutants. Phenomena occurring in the ''real world'' are for the most part so complicated that usually no clear cut conclusions can be reached when research work is limited only to the element in question. But because elements within the same Group and/or sub-Group of the periodic table possess similar chemical as well as physical properties, it can be expected that more comprehensive understanding will be obtained when all the elements are studied comparatively in relation to their position in the periodic table. In addition to this, various kinds of elements, which were not commonly used in the past, are now increasingly consumed by modern industries in the production of numerous new materials. There is a strong possibility that these industrial products will be disposed of in the environment and, hence will cause new types of environmental pollution. In order to cope with this situation, it seems essential to employ analytical techniques that possess rapid simultaneous multielement capabilities like ICPMS.

2 HISTORY

ICPMS is a relatively new technique that combines two well-established analytical tools (inductively coupled plasma (ICP) and mass spectrometry (MS)) to produce an instrument with powerful potential in the field of multielement trace analysis. ICP is used as a source of ions rather than as a source for optical emission measurement (as in inductively coupled plasma atomic emission spectrometry, ICPAES). Since the introduction of the first commercial ICPMS instruments in 1983, the technique has gained increasing attention from analytical chemists,[1,2] and the total number of instruments installed worldwide by the end of 1998 was estimated to be about 3000.

The major advantages of ICPMS over other analytical methods are:

1. increased sensitivity and wide dynamic range

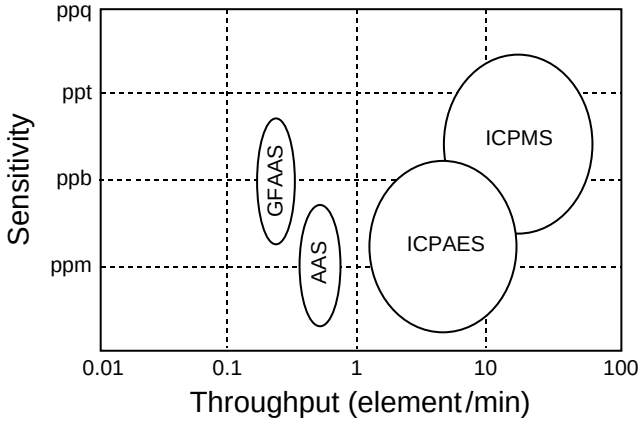

Figure 1 Comparison of techniques used for environmental analysis (see text for abbreviations).

2. high selectivity
3. rapid multielement capability
4. simplicity in the spectra
5. low and uniform background
6. the ability to measure specific isotopes.

It should be noted, however, that there are also several disadvantages. These are as follows:

1. lower precision compared with ICPAES and atomic absorption spectrometry (AAS)
2. the total dissolved salts should be less than 1000 ppm
3. severe matrix effects
4. higher cost of the instrument.

3 INSTRUMENTATION

3.1 Instrument Overview

A schematic representation of the main components of a typical ICPMS instrument is shown in Figure 2. Although ICPs are basically similar to those that have been used in emission spectrometry (ICPAES) for many years, the torch is mounted horizontally rather than vertically (torches have also been mounted horizontally in ICPAES). The interface between the plasma and the mass spectrometer consists of two cones: the first is called the sampling cone and the other one that is installed behind the sampling cone is known as the skimmer cone.

Transfer from the skimmer to the mass analyzer is effected by a system of electrostatic lenses, which extract the positively charged ions and focus the ion beam in a form suitable for transmission through the quadrupole mass filter. The mass filter transmits the ions of a particular selected mass to charge ratio (m/z) to an ion detector (channel electron multiplier) that registers the ions. Each

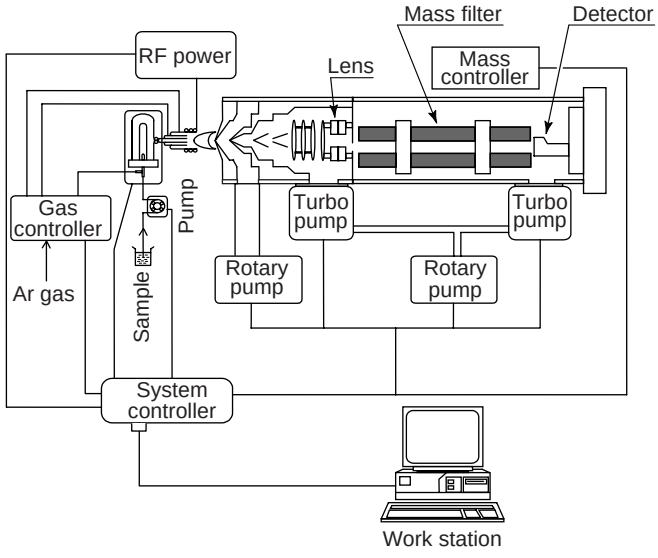

Figure 2 Schematic diagram of a typical ICPMS instrument. (Modified from the original drawing provided by Yokogawa Analytical Systems, Japan.)

naturally occurring element has a unique and simple pattern of near-integer mass to charge ratio corresponding to its stable isotopes. Accordingly, the identification of an element is very simple and straightforward. The number of registered ions from a given isotope depends, of course, on the concentration of the relevant elements in the sample. More detailed descriptions of the instrumentation of ICPMS have been presented elsewhere.[3–5]

3.2 Inductively Coupled Plasma

The ICP is produced by passing ionized Ar gas ''seeded'' with a few electrons from a Tesla Coil through a quartz torch located inside a Cu coil that is connected to a radiofrequency (RF) generator. The high-frequency currents from the generator flow in the Cu coil and thus generate oscillating magnetic fields. Electrons and ions passing through the oscillating electromagnetic field flow at high-acceleration rates in a closed annular path inside the quartz tube. The direction and strength of the induced magnetic fields varies with time, resulting in electron acceleration at each half cycle. Collisions between accelerated electrons (and ions) and unionized Ar gas cause further ionization. However, other Ar atoms impede the movement of this current of charged particles and cause ohmic heating of the gas.

3.3 Sample Introduction Systems

A key advantage of ICPMS over other types of MS is that the ion source (ICP) is located outside the vacuum system and allows much simpler sample introduction.

For analysis by the ICPMS technique, samples must be in a form which can be carried in a stream of Ar gas (nebulizer gas). The most convenient method for sample introduction into the gas stream is the generation of a fine aerosol using a nebulizer and a spray chamber, as illustrated in Figure 2, because most of analyses are carried out on liquid samples.

3.3.1 Pneumatic Nebulization

The basic principle of pneumatic nebulization is that the stream of sample solution is disrupted by a high velocity gas flow to produce a spray of droplets. The construction of a pneumatic nebulizer suitable for ICPMS demands much greater skill because of the low gas flow rates. A spray chamber is attached to the nebulizer to remove most of larger droplets and to ensure that only the smallest solution droplets (less than 10-μm diameter) are carried into the plasma to avoid excessive cooling of the plasma. There are essentially three types of pneumatic nebulizer commercially available for ICPMS: concentric flow, crossflow, and Babington-type nebulizers. For those interested in more detailed information on this subject, reviews by Sharp[6,7] can be recommended.

The most widely used nebulizer is the Meinhard concentric glass nebulizer,[8] shown in Figure 3. This all-glass device is constructed by standard glassblowing techniques. Nebulizer gas (Ar) at a flow rate of around $1\,L\,min^{-1}$ is passed through the outer glass tube. The pressure drop generated (Venturi effect) causes the sample solution to be drawn up into the axial capillary and simultaneously discharged into aerosol droplets. The fragility and fineness of the sample inlet capillary and of the annular gas orifice are major drawbacks of this type of nebulizer. Blockage caused by the accumulation of salt particles (salting-up) and/or suspended solids is common, although a high salt tolerant type is also available.

The operation of the crossflow nebulizer is based on the well-known ''scent-spray'' principle.[9] A horizontal jet of gas passes across the top of a vertical tube. The reduced pressure so generated sucks the sample solution up the vertical tube, and at the top is dispersed into fine droplets (Figure 4). Although the analytical performance offered by the crossflow nebulizer is similar to the concentric type, it can be constructed from various materials other than glass. It is possible, therefore, to introduce solutions containing high concentrations of HF.

In the Babington-type nebulizer (Figure 5), the sample solution flows freely over an open slot or V-groove.[10,11] Gas forced through a small aperture in the bottom of the groove ruptures the fed solution to produce an aerosol. As the solution is simply flowing over a slot, rather than passing through a fine capillary, this type of nebulizer shows high tolerance towards samples with large amounts of dissolved solid, and even to slurries.[12,13]

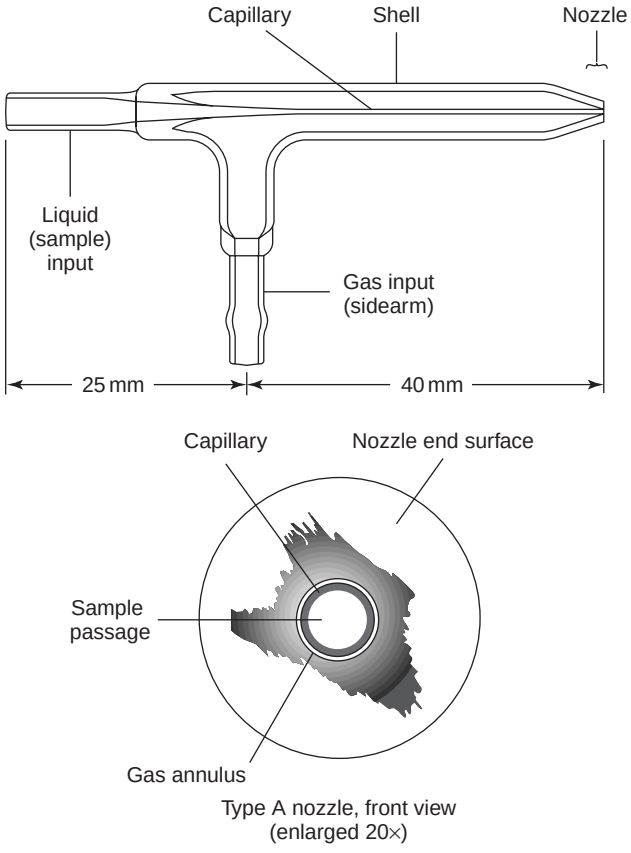

Figure 3 Meinhard concentric glass nebulizer. (Reproduced with permission from J.E. Meinhard Associates Inc., USA.)

3.3.2 Ultrasonic Nebulization

The pneumatic nebulizer has been commonly used for sample introduction into the ICP. However, transport efficiency is extremely low, owing to the very wide range of droplet sizes produced. Ultrasonic nebulization greatly increases the transport efficiency and hence results in better analytical sensitivity and/or detection limits.[14–16] A schematic diagram of the ultrasonic nebulizer (USN) used in the author's laboratory is shown in Figure 6. A peristaltic pump is used to deliver the sample solution to the oscillating surface of the quartz plate. Aerosols with smaller and more uniform particle sizes are produced and transported first to the heated tube, and then to the condenser by the nebulizer Ar gas. Most of the water in the aerosol is removed by this process and only the dry aerosol is introduced into the plasma. The temperature of the heated tube is around 120 °C (slightly higher than the boiling point of water) and that of the condenser is usually set around 1 °C (near freezing point), and the sample introduction rate is typically $2–4\,mL\,min^{-1}$. CETAC Technologies has introduced a membrane desolvator to couple with the USN. The aim was to preserve the high

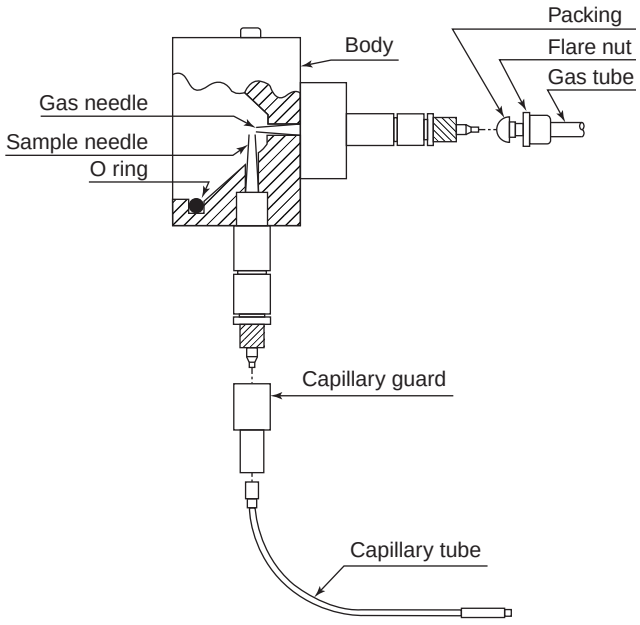

Figure 4 Schematic diagram of the crossflow nebulizer. (Modified from the original drawing provided by Nippon Jarrell-Ash, Japan.)

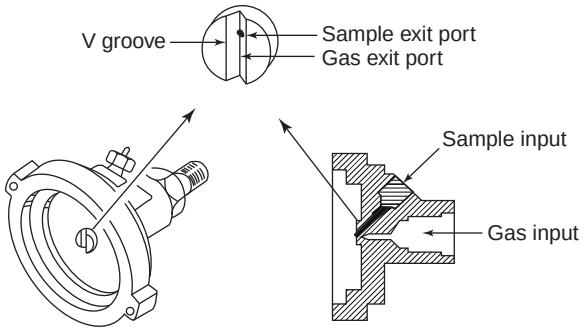

Figure 5 Simplified diagram of Babington-type nebulizer. (Modified from the original drawing provided by Perkin-Elmer, Japan.)

efficiency of USN and significantly reduce solvent loading into ICP.

3.3.3 Direct Injection Nebulizer

In the direct injection nebulizer (DIN), the aerosol is also generated upon the interaction of a high-velocity gas stream with the sample liquid as in many other pneumatic nebulizers. However, this newly developed nebulizer is unique in that the sample is delivered through a narrow thin-walled highly inert capillary tube by a high-pressure gas-displacement pump ($>40\,\mathrm{kg\,cm^{-2}}$). Argon gas passes into the nebulizer tip through the narrow annular space between the sample capillary and the nebulizer tip, where a fine mist with a narrow particle size distribution is

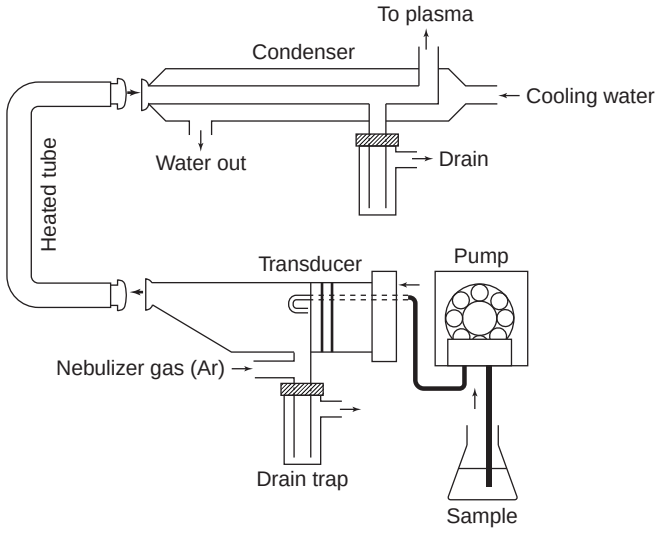

Figure 6 Schematic diagram of a USN.

created by the interaction between the nebulizer gas and the liquid sample. The absence of large droplets eliminates the need for a spray chamber and thus the analyte is directly injected into the central channel of the ICP. The DIN is an ideal sample introduction system for samples with limited volume, as the typical consumption of the sample is only $50–100\,\mu\mathrm{L\,min^{-1}}$. In addition, analysis of samples for elements with a greater memory effect, such as Hg and B, can be successfully carried out by DIN because the spray chamber is eliminated and the internal volume of the nebulizer is only a few microliters.

3.3.4 Chemical Vapor Generation

Several elements belong to groups IVB, VB, and VIB in the periodic table, namely Ge, As, Sn, Pb, Se, Sb, Bi, and Te form hydrides, which are gaseous at room temperature. These can be generated easily from aqueous solutions in a reducing environment. The most commonly used method for hydride generation is the acid–borohydride reaction. The capability of this technique has been fully exploited in AAS, and the AAS–hydride method is now widely used on routine basis.[17,18] Although less complicated batch processes are possible for hydride generation, continuous methods are generally preferred in ICPMS. This is because signals of shorter duration (transient response) are not suitable for multielement determination in ICPMS.

A schematic diagram of a continuous hydride generator is shown in Figure 7. The acidified sample solution and the reducing agent ($NaBH_4$ solution) streams are pumped through narrow-bore tubing by a peristaltic pump, and combine at a capillary T-junction, where a vigorous reaction begins. The solution passes into the phase

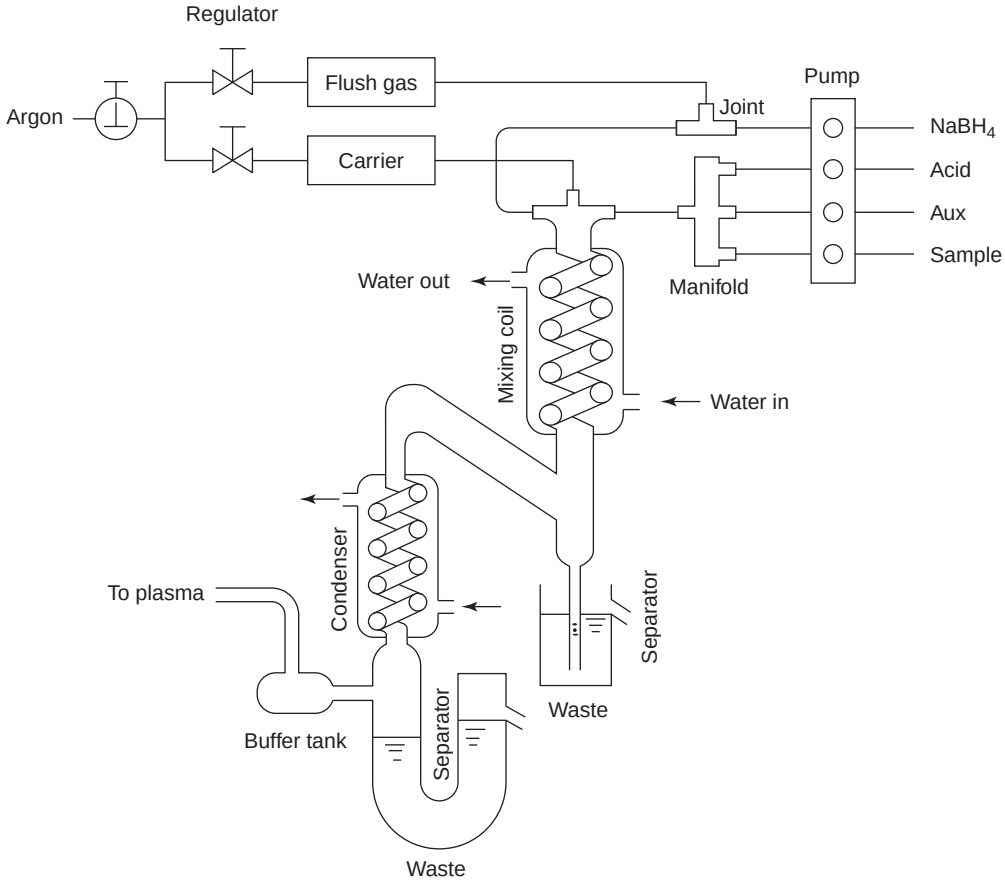

Figure 7 Schematic diagram of the hydride generation system. (Modified from the original drawing provided by Yanagimoto Mfg., Kyoto, Japan.)

separator and the gaseous products alone are carried to the plasma by the nebulizer gas (Ar).

3.3.5 Electrothermal Vaporization

Electrothermal vaporization (ETV), also developed in AAS (typically known as GFAAS, graphite furnace atomic absorption spectrometry), has come into widespread use for the determination of ultratrace levels of elements with considerable success.[19,20] The technique involves evaporation of a liquid sample directly into the plasma from an electrically heated graphite cuvette. Whereas the full atomization of the sample solution in the cuvette is the major concern in AAS, the evaporated samples could be further transported effectively (nebulizer and spray chamber are removed) into the plasma in ICPMS. Therefore, the ETV device for ICPMS and/or ICPAES[21,22] is (Figure 8) considerably different from that used in AAS.[19,20]

The major advantage of ETV is the very much improved sample introduction efficiency compared with the pneumatic nebulizer. The improvement in sensitivity is, therefore, an order of magnitude or more. In addition, the level of oxide ions is dramatically decreased because only the dry aerosol is introduced into the plasma. For example, considerable reduction in detection limit for ^{56}Fe results from reduced spectral overlap from ^{40}ArO.

3.3.6 Laser Ablation

In this technique, a laser is used as a method of volatilizing sample materials. The system is illustrated schematically in Figure 9.[23] The sample is located in a quartz or glass cell of small volume, typically 30–100 mL, to minimize dilution of the sample vapor. The laser beam passes through the optical window and vaporizes a small part of the sample. The vapor ablated from the sample surface is then transported directly to ICP by the carrier (nebulizer) gas. When the ablated material reaches the ICP, a pulsed signal of short duration is observed similar to the peak obtained using ETV. This method allows direct analysis of the solid sample with minimum preparation. In addition, microprobe analysis can be developed when laser ablation is used with a device that controls an X-Y motion of the sample.[24]

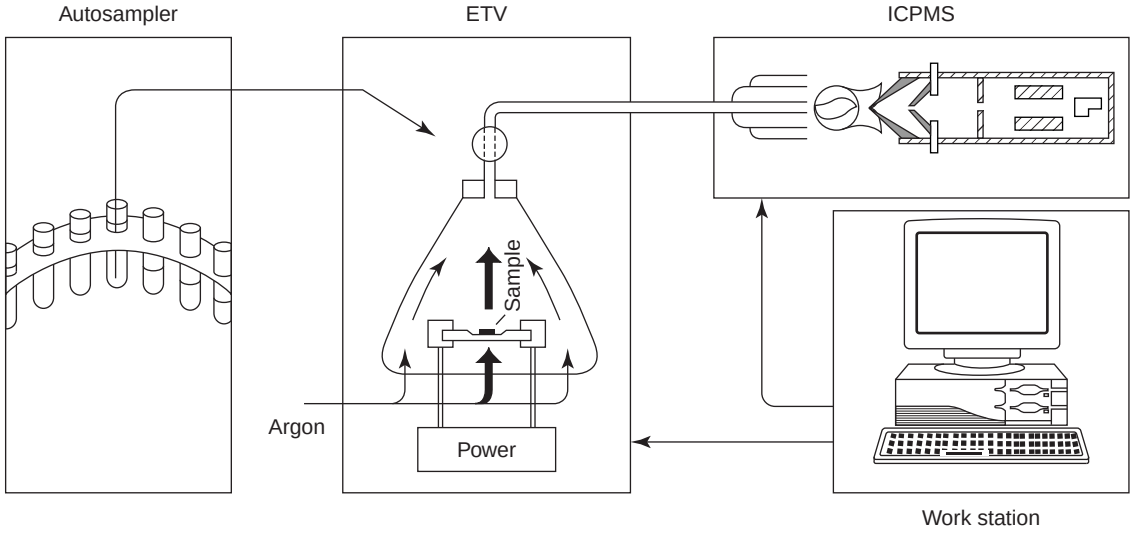

Figure 8 Schematic diagram of ETV for ICPMS. (Modified from the original drawing provided by Yokogawa Analytical Systems, Japan.)

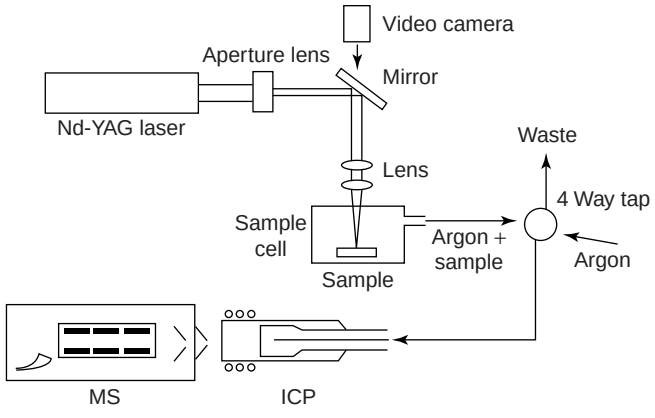

Figure 9 General arrangement of laser, ablation cell, and gas flow.

3.3.7 Other Techniques

One of the unique features of the ICP is its ability to accept a wide range of sample introduction systems. Modifications of ICPMS are, therefore, relatively simple and hence, hyphenated systems for specific purposes will be easily constructed. Introduction systems with greater potential are liquid chromatography (LC), and flow injection (FI).

The former (LC/ICPMS) provides information about the chemical forms, or elemental species, rather than about total amount of elements alone. In addition, on-line separation from the matrix elements and/or preconcentration are easily attainable when combined with appropriate LC systems. FI/ICPMS systems are advantageous for solutions with high dissolved salts content because only a limited amount of sample, usually 50–100 μL, is injected into a continuously flowing stream of a carrier solution of very low or no salts content.

4 SAMPLE PREPARATION

4.1 Methods of Dissolution

For the total analysis of solid samples, all the elements have to be brought into solution, although such techniques as laser ablation[23,24] and slurry introduction[12,13] are known to be available. Classical thermal decomposition methods (dry ashing) are not applicable for ICPMS because loss of certain volatile elements and/or contamination during the ashing procedures are rather common in these methods. As solutions of high salts content are not suitable for ICPMS, fusion methods are also not acceptable unless the analytes are subsequently separated by applying appropriate chemical techniques. A further problem associated with this method is the difficulty of obtaining high-purity fusion reagents (fluxes), such as $NaCO_3$ and $Na_2B_4O_7$. As a consequence, digestion with strong acids (wet digestion) is exclusively used in ICPMS, although the fusion techniques are often required to dissolve the refractory substances mentioned below. Acids of ultrahigh purity are commercially available now.

4.1.1 Open Digestion Systems

Digestion methods for plant and tissues and related materials for ICPMS are essentially similar to those already developed for other types of analytical methods, especially for ICPAES and AAS. Typically, 0.5–1 g of finely ground samples are treated with HNO_3–$HClO_4$

on a hot plate, followed by dissolution of the residue and dilution to 50–100 mL. The use of H_2SO_4 is not recommended because of the adverse effects of this acid in the subsequent ICPMS measurements.

The use of HF, in combination with $HClO_4$ and HNO_3, is a well-established method for the dissolution of samples to determine the total element content of silicate minerals and related materials, and has been extensively applied in AAS and ICPAES.[25–27] Although Si is lost by evaporation of gaseous SiF_4 when an open system is used, no problem will arise because ICPMS is mostly applied to trace and ultratrace analysis. A problem can arise, however, because some minerals, such as chromite ($FeCr_2O_4$), garnet ($R''_3\ R'''_2(SiO_4)_3$, where R″ may be Ca, Mg, Fe, and Mn, and R‴ may be Al, Fe, Cr and Mn), magnetite (Fe_3O_4), and zircon ($ZrSiO_4$), resist acid attack and remain undissolved even after repeated treatment. As these minerals almost always contain a wide variety of elements as ''impurities'' (a typical example is Hf and heavy rare earth elements in zircon), special dissolution techniques should be applied for samples containing considerable amounts of these minerals. Care is also required when the $HClO_4$–HF mixture is evaporated to dryness, because acid insoluble oxides will be formed again if too much heat is applied.

The advantages of the open digestion systems are as follows:

1. Silicon and the excess HF are easily removed by heating.
2. The final solutions are stable (silica-bearing solutions tend to hydrolyze and precipitate on standing).
3. The removal of corrosive HF allows the solution to be handled in conventional glassware for periods of at least a couple of days.
4. The salts content of the resultant solutions is low.

4.1.2 Closed Digestion Systems

An alternative dissolution procedure using HF is to digest the samples in a sealed Teflon vessel or ''Teflon bomb'' placed in a stainless steel casing.[28–30] Bombs with different designs are now commercially available (Figure 10). After the addition of sample and acid, the bomb is sealed and heated in an electric oven.

The advantages of closed digestion systems over the open vessel techniques are as follows:

1. Digestion at higher temperatures is possible because the boiling point of the reagents is raised by the pressure generated in the vessel. This may shorten the time needed for decomposition and also enable the dissolution of some acid-tolerant minerals.

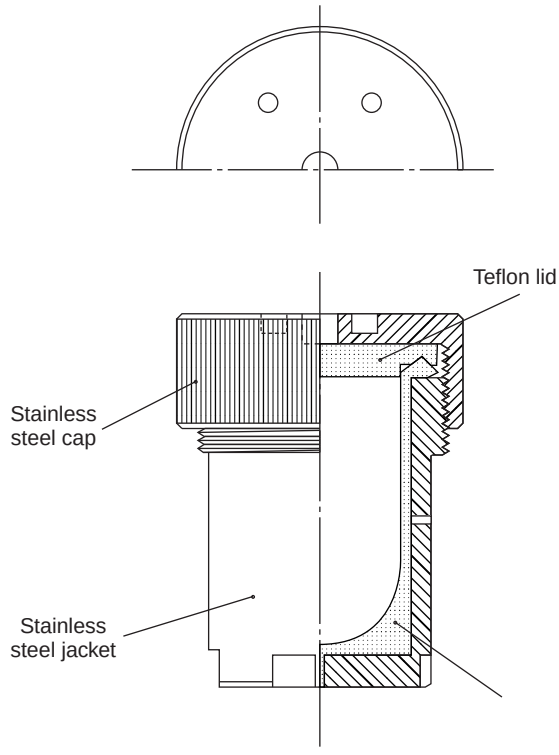

Figure 10 Diagrammatic cross-section of a ''bomb''. (Modified from the original drawing provided by Yanagimoto Mfg., Kyoto, Japan.)

2. Volatile elements, such as B, Cr, As, Se, Sn and Hg, will remain within the vessel.
3. Smaller amounts of acid are sufficient for the digestion because evaporation does not occur. This lowers the reagent blank level considerably.
4. Contamination is significantly reduced because the sealed system prevents the introduction of airborne particles during decomposition.

The disadvantages are the following:

1. The maximum amount of sample that can be processed is rather limited, usually less than 0.5 g. This is especially the case when the content of organic matter in the sample is high.
2. Large amounts of H_3BO_3 should be added to complex the excess HF. This results in high salts content in the analyzing solutions.
3. Subsequent open vessel procedures are needed when the removal of Si and HF is required.

Heating needed for the acid decomposition of samples is traditionally provided externally via chemical flames or electrically on hot plates. However, methods using microwave energy for heating have been

widely accepted.[31] This technique is applied most suitably to closed digestion systems because the sample solutions can be heated directly in microwave ''transparent'' vessels. Although domestic microwave ovens can be used successfully, purpose-built microwave digestion systems are commercially available. Very rapid heating is possible in comparison with open digestion and bomb methods because the heat is generated from inside rather than supplied from outside. A comprehensive review on this subject has been published.[32]

5 ENVIRONMENTAL APPLICATIONS

5.1 Overview

The definition of environmental analysis seems to be not very clear. But when samples are analyzed in connection with problems associated with environmental deterioration, these activities are usually termed ''environmental analysis''. Therefore, the sample types considered in this section cover various different materials with diversely different natures, such as water, soils and sediments, biological samples including foods and animal tissue, and geochemical samples. Literature retrieval has shown that more than 150 reports related to environmental analysis by ICPMS were published between 1996 and 1998. The types of sample examined in the literature and their relative ratios (%) are shown in Table 1.

Extensive annual reviews of environmental analysis have been published from 1985 by the Royal Society of Chemistry (UK) in the *Journal of Analytical Atomic Spectrometry*. In addition to these, there also have been a large number of reviews on specific types of samples and topics, as well as reviews on environmental analysis in general published in several different languages. These are soil samples,[34] double-focusing sector field ICPMS,[35–37] sediment, river water and seawater (in Japanese[38,39]), trace radioactivity,[40] boron in plants and soil,[41] trace element species,[42] speciation by gas chromatography (GC) ICPMS (Japanese,[43] Chinese[44]), speciation by LC/ICPMS,[45] speciation by HPLC ICPMS,[46] speciation by capillary electrophoresis ICPMS (Chinese[47]),[48] ETV/ICPMS (German[49]), isotope ratio analysis of environmental and biological samples (Japanese[50]), Pu in environment[51] and environmental samples in general[52–55] (Japanese[56,57,83,105]).

Therefore, it is not the author's intention to prepare a comprehensive review in this article but rather try to introduce several resent research works as examples.

Table 1 Applications of ICPMS to water analysis

Sample, techniques and elements	Ref.
Volatile metal(loid), GC/ICPMS	97
River water, comparison with neutron activation analysis	98
Rainwater, heavy metals	99
Rainwater, Pb isotope ratios	100
Seawater and sediment, Pb isotope ratios	101
As and Se, HPLC/ICPMS	102
Tap water and river water, ^{99}Tc	103
Seawater, flow injection, Pb, Cd	104
Wet deposition, Mn, Co, Ni, Cu, Zn, Cd, Pb, Tl	105
Seawater, ETV/ICPMS, Se, Mo, Pt, Cd, Pb	106
Certified reference materials, online solid chelation, Cd, Co, Cu, Mn, Ni, Pb, U, Zn	107
Groundwater, flow injection	108
Drinking and vadose water, on-site water analyses	109
Rain and dry fallout, ^{99}Tc	110
Coastal seawater, ^{99}Tc	111

HPLC, high-performance liquid chromatography.

5.2 Detection Limits and Range

The periodic table in Figure 11 shows 70 or so elements that can be determined by ICPMS and their approximate detection limits. In ICPMS, detection limits are defined as the analyte concentration equivalent to three times the standard deviation (3d) of the blank response (background) measured, in most cases, in a single ion monitoring mode for an integration time of 10 s. The standard deviation of the blank solution is calculated from 10 consecutive measurements carried out within a relatively short period (usually less than 10 min). It should be noted, however, that 5–10 times the detection limits are usually required for good qualitative measurements.

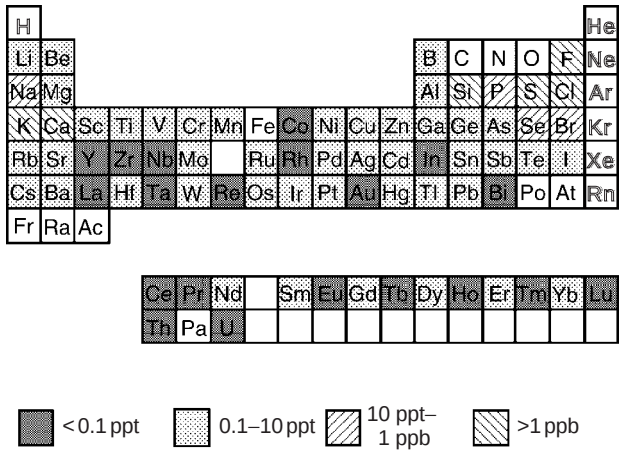

Figure 11 Periodic table showing the detectable elements and their approximate detection limits. Unhatched elements have not been examined yet.

The detection limits for a considerable number of elements are about 1 part per trillion (ppt). The elements not suitable for ICPMS measurements are the halides, the inert gases and atmospheric gases. Elements with a low sensitivity and/or higher detection limits fall within these groups. The so-called nonmetallic elements are not easily converted to positively charged ions.[33] The higher detection limits observed for Si, K, Ca, and Fe are mainly due to the higher background signals caused by spectral overlaps of various molecular species originating from the matrix and/or gaseous components in the plasma (e.g. Ar, N, H and O).

5.3 Water Analysis

The types of water samples in environmental studies include lake water, river water, seawater, groundwater, and rainwater. However, assessment of water quality often necessitates analysis of related samples such as sediments, suspended materials, and aquatic biota, including algae, higher aquatic plants, insects and fish. For such samples, analytical procedures are essentially similar to those for soils and biological samples described below. Suspended materials are removed by filtration through a 0.4–0.45-μm pore size membrane filter or centrifugation.

The detection limits of ICPMS are, in general, low enough to allow the analysis of natural water samples without using preconcentration procedures.[58–73] If separation and/or preconcentration steps are needed to reduce or eliminate matrix interferences (for example, seawater), wide varieties of techniques including ion exchange,[74–78] coprecipitation,[79] solvent extraction,[80–82] solid-phase extraction,[83] and chromatographic resin[84] are available.

Miyazaki and Reimer[85] determined Pb concentrations and isotope ratios after preconcentration with a Chelex-100. A similar FI procedure has been applied to the analysis of concentrated brines[86] and seawater.[87] Seubert et al.[88] described synthesis of an inert type of 8-hydroxyquinoline-based chelating ion exchanger to be used for seawater analysis. It was possible to determine several elements in seawater in the low-ppt range by on-line ICPMS detection. An FI hydride generation technique has been described for the determination of As, Bi, Sb, Se, Te, and Hg in water at ultratrace concentration levels.[89–91] McLaren et al.[92] described an on-line method for the determination of trace elements in seawater. Shabani and Masuda,[93] and Aggarwal et al.[94] developed on-line methods for trace rhenium and sub-ppt levels of rare earth elements in seawater, respectively. Speciation of arsenic compounds in drinking water by capillary electrophoresis and hydride generation ICPMS has been reported by Magnuson et al.[95] It was possible to speciate four environmentally significant toxic forms of As: arsenic, arsenate, monomethylarsonic acid, and dimethylarsinic acid. Ion chromatography ICPMS utilizing hydride generation with a membrane separator was also investigated by the same authors for this purpose.[96] Interesting examples other than those described above are summarized in Table 1.

5.4 Soils and Related Materials

The first report on ICPMS appeared in 1980.[112] Although the total number of publications on ICPMS reached more than 3500 by April 1999, rather fewer applications of ICPMS for the analysis of soil samples and related materials have been reported.[113–134] These are summarized in Table 2. The high cost of the instrument appears to be the major reason preventing its widespread use among soil scientists. The price however continues to fall dramatically with increased annual production, so it might be expected that ICPMS will become one of the most widely used techniques for the determination of trace and ultratrace element concentrations, as well as for isotope ratio measurements in soil samples in the near future.

Yamasaki et al.[134] selected about 100 soil samples in such a way as to cover a wide range of chemical as

Table 2 Applications of ICPMS to soil analysis

Samples, techniques and elements	Ref.
Microwave-assisted leaching (EPA 3051)	113
Pt, road dust and soil	114
^{99}Tc, HPLC/ICPMS	115
As species in soil and sediment, HPLC/ICPMS	116
Radionuclides	117
Global fallout ^{99}Tc, paddy soils	118
Polluted soils, site characterization	119
^{230}Th, ^{234}U, ^{239}Pu, ^{240}Pu, FI preconcentration	120
Mn, Cu, Zn, Cd, Hg and Pb in 3 soils and sludges	121
Rocks, soils, sediments, chalcophile elements, laser ablation	122
As and Se in soils, continuous hydride generation	123
7 Chinese reference soils, laser ablation, 30 elements	124
Plants and soil, tracer technique, ^{10}B	125
Molten borax or pressed polyethylene disk, laser ablation, 41 elements	126
30 soil samples, feasibility test of laser ablation, As, Ba, Cr, Pb	127
Soil colloids, chemical and mineralogical properties of soil particles up to 0.5 μm	128
Review in general	129
Soil leachates, combined with chromatography, Al, Mn, Fe, Ni, Cu, Zn, Cd, La	130
Plants and soil, Mn, Co, Ni, Cu, Zn, As, Mo, Sr	131
Reference soils, HF–$HClO_4$ digestion, comparison with photon activation analysis, 42 elements	132
Various soils, HF–$HClO_4$ digestion, comparison with ICPAES, Na, Mg, Al, Ca, Ti, Mn, Fe, Ni, Cu, Zn, Pb	133
Various soils and standard reference materials, HF–$HClO_4$ digestion, 47 trace and ultratrace elements	134

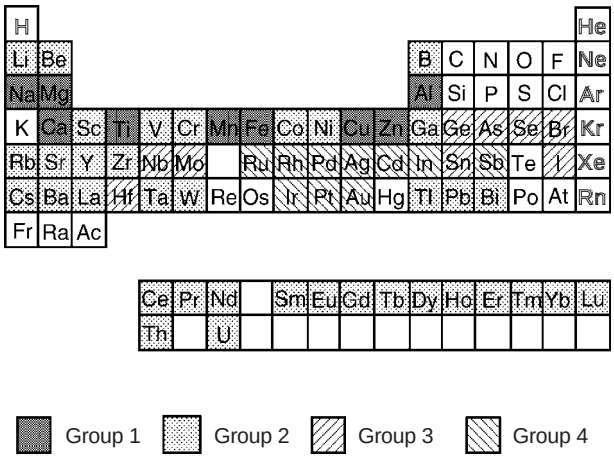

Figure 12 Periodic table showing the elements detectable in soil samples. The identified percentages for Groups 1, 2, 3, and 4 are respectively 100%, more than 90%, more than 70%, and less than 50%.

well as mineralogical compositions and examined the detectability of elements. A large number of elements were found (levels 10 times greater than the blank) in most soil samples. These elements are conveniently classified into four Groups, as shown in Figure 12 using the periodic table.

The elements classified in Group 1 were identified for all the samples examined. The problem associated with determination of these elements (except for Cu and Zn) is that high and variable dilution ratios (100–1000 fold) must be employed to adjust the concentration of each element to the optimum range. This not only requires time-consuming manipulation but also greatly reduces the multielement capability of ICPMS.

The elements belonging to Group 2 were detected in more than 90% of the samples. The unparalleled detection power of ICPMS is clearly demonstrated by the fact that it was possible to confirm the occurrence of more than 40 elements (Group 1 + Group 2) within several minutes. In addition to these elements, those classified in Group 3 were also found in about 70% of the samples. The elements in Group 4 were detected in only less than 50% of the samples. This indicates that the contents of these elements in soil samples are generally too low ($<0.1\,mg\,kg^{-1}$) for detection even with extremely sensitive ICPMS. It is apparent, therefore, that there are still a considerable number of elements for which we need adequate preconcentration and/or separation techniques to obtain reliable data on their content in soils.

5.5 Biological Samples

ICPMS is increasingly used for the analysis of biological samples mainly because of its high sensitivity, high selectivity and rapid multielement capability along with the ability to perform isotopic ratio measurements. Sample preparation procedures for ICPMS measurements are essentially similar to those for other types of environmental samples, although the use of HF is seldom needed for biological samples. Although the amount of the decomposable sample is usually limited to less than 1.0 g, microwave digestion (see section 4.1.2) is most suitable because samples can be decomposed using only HNO_3 by this technique.[135] This makes it possible to eliminate the use of $HClO_4$ and/or H_2SO_4 which are known to produce various molecular species which overlap with many important isotopes.[136] Table 3 shows some recent applications of ICPMS to biological samples. Reviews by Barnes[160] and Mukai and Yoshinaga[161] can be used for more detailed information.

Table 3 Applications of ICPMS to biological samples

Sample, techniques and elements	Ref.
Brown rice, microanalysis (<20 mg)	137
Brown rice, B isotope ratios	138
Cannabis, sourcing the provenance, laser ablation	139
Whole blood, HR/ICPMS, Pb isotope ratios	140
Biological reference materials, I	141
Seaweed, extraction with chromatographic resin, ^{99}Tc	142
Urine, HR/ICPMS, Pt group metals	143
Breast milk, toxic metals and radionuclides	144
Human placenta, microwave digestion	145
Brain, bone, kidney, liver, lung, As, Be, Cd, Cr, Hg, Mn, Ni, Pb, Sn, Tl, V, Zn	146
Human serum, HR/ICPMS	147
Human blood, HR/ICPMS, Pd, Pt, Ir, Au	148
Lichens and vegetable	149
Hair, liver, algae, monomethylmercury	150
Rice flour, isotope dilution, trace elements	151
Human hair, isotope dilution, Hg	152
Mussel tissue, hydride generation-cold trap-GC, butyltin	153
Urine, Sb, Bi, Pb, Cd, Hg, Pd, Te, Sn, Tl, W	154
Human serum and urine, HPLC/ICPMS, organic selenium	155
Human plasma and serum, LC/ICPMS, Se	156
Human blood, isotope dilution, Cd and Pb	157
Urine, Th and U	158
Blood serum, rare earth elements	159

6 METHOD DEVELOPMENT

6.1 High-resolution Inductively Coupled Plasma Mass Spectrometry

6.1.1 Background of Development

Although ICPMS has various advantages over other analytical techniques as mentioned previously, it has become evident that high selectivity and simplicity of spectra are not always the case. This is mainly because

various polyatomic molecular species, originating from gaseous components (i.e. H, C, N, O, and Ar), matrix elements in samples, and components in acids or alkalis used for sample dissolution are formed in spite of the high temperature of the plasma. It is often the case that spectra appear simple only because the commonly used quadrupole-type mass spectrometers in ICPMS are incapable of separating chemically different ions at the same nominal m/z value.

The driving force behind the development of a new type of high resolution instrument that uses a double-focusing mass spectrometer has been the separation of the analyte ions from the overlapping polyatomic ions which is especially noticeable in the region between mass 40–80, that is, from ^{40}Ar to $^{40}Ar_2$.[162,163] Typical examples of such overlapping peaks and the resolution needed for the separation of the analyte ions are shown in Table 4.

6.1.2 Instrumentation

Commercial high-resolution ICPMS was first launched by VG Elemental (now Micromass UK Ltd.) in 1988. Similar systems were also introduced into the marketplace by two other manufactures (JEOL in Japan and Finnigan MAT in Germany). In 1998, the total number of this type of instrument working in the whole world was estimated to be more than 200 provided by four manufacturers. A schematic representation of the main components of the ELEMENT, an instrument provided by Finnigan MAT, is shown in Figure 13.

The main reason for developing the high-resolution ICPMS was to resolve molecular overlaps caused by matrix species. However, it has also become evident that the detection limits of most elements are significantly improved when this instrument is operated in the low resolution mode because of the much lower background signal and higher ion transmission. The low background is because the long and curved ion flight path together with the complex focusing elements prevent photons from reaching the detector. The higher ion transmission is the result of ion acceleration and efficient focusing.

Table 4 Resolution required to separate interferences

(m/z)	Analyte	Interferent	Resolution
28	Si	$^{14}N_2$	960
32	S	$^{16}O_2$	1800
51	V	^{35}ClO	2600
52	Cr	$^{40}Ar^{12}C$	2400
55	Mn	ArNH	1600
56	Fe	ArO	3000
75	As	ArCl	7800
80	Se	Ar_2	9700

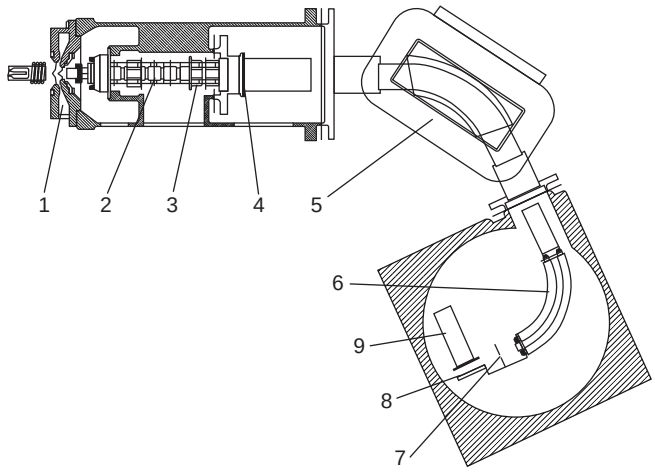

Figure 13 Schematic diagram of a high-resolution (double focusing) ICPMS, ELEMENT: 1, plasma interface; 2, transfer and focusing optics; 3, acceleration and beam focusing; 4, entrance slit; 5, electromagnet; 6, electric sector; 7, exit slit; 8, conversion dynode; 9, dual mode SEM. (Reproduced with permission from Finnigan MAT, Germany.)

6.1.3 Resolution and Sensitivity

Figure 14 illustrates how the signal intensity changes with resolution.[58] Data was obtained by aspirating the 1-ppb (parts per billion) indium (In) solution using a USN. Signal intensity is expressed in terms of counts per second (Hz) per parts per million (ppm) of In. The analyzing system used in this experiment has such excellent detection power that a signal intensity as high as 2 GHz ppm^{-1} is reached when the instrument operates in the lowest resolution mode (i.e. with both slits opened wide). The resolution in this mode is estimated to be around 400. When the resolution is increased to 1500 to separate, for example, Si from N_2, the sensitivity decreases to 60% of that obtained in the low-resolution mode. If the resolution is further increased to 3500 for the first transition elements, the signal intensity is around 400 MHz ppm^{-1}. At a resolution of 8500, the sensitivity is only 2% of that attained in the low-resolution mode.

6.1.4 Typical Detection Limits

Detection limits of a high-resolution instrument are summarized in Figure 15. These values were obtained under essentially similar conditions to those in quadrupole-type instruments, although sample solutions were introduced through a USN.[58]

As mentioned above, spectral overlaps caused by polyatomic molecular species originating from gaseous components and/or matrix elements are common in the mass range of 40–80, so measurement of the first transition elements must be carried out in the high-resolution mode. Because sensitivity decreases sharply

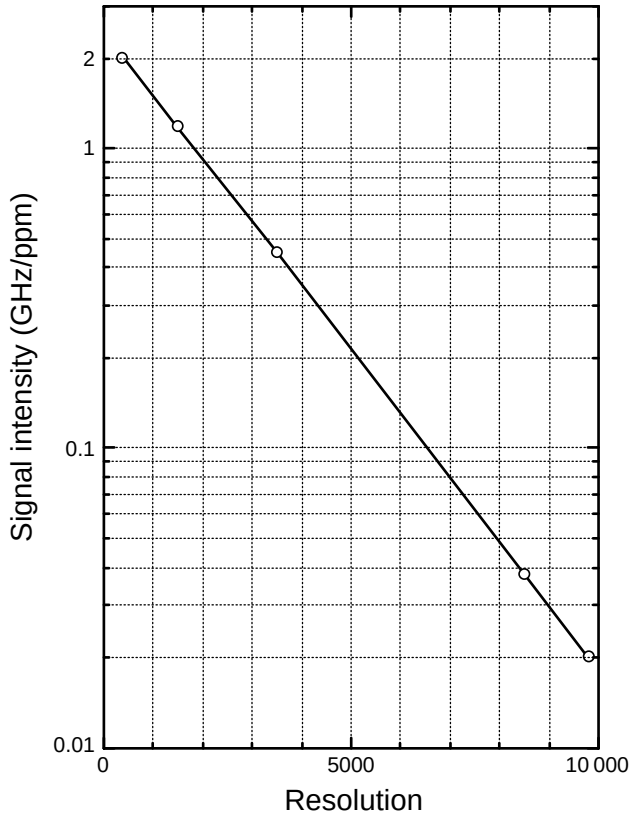

Figure 14 Sensitivity versus resolution curve. Data was obtained using 1 ppb of indium (In) solution. Signal intensity is expressed in terms of counts per second (Hz) per ppm of In.

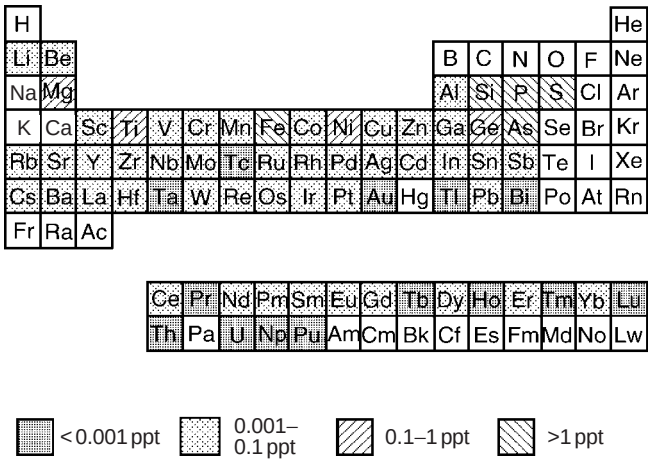

Figure 15 Typical detection limits obtained by a high resolution ICPMS. The total peak dwell times (integration times) were 5 s. Reliable values were not obtained for those elements shown in italics (Na, K, and Ca) owing to higher blank levels. Other unhatched elements have not been examined yet.

when the instrument operates in the high resolution mode, as shown in Figure 14, the detection limits of these elements were expected to be considerably higher than those of the elements for which measurements can be done in the low-resolution mode. It can be seen, however, that the detection limits of the first transition elements, except Fe, are all well below 1 ppt even in the high-resolution mode. This high value for ^{56}Fe can be attributed to the higher total counts of the blank, presumably owing to the effect of $^{40}Ar^{16}O$. It was not possible to obtain reliable values for Na and Ca because of a very high blank value possibly originating in the sample introduction system and/or owing to the memory effects.

With the exception of Mg, Si, P, S, Ge and As, detection limits of elements other than the first transition elements were below 0.1 ppt, and reached as low as less than 1 ppq (0.001 ppt) for some elements.

Because the instrumental conditions for the measurement of each element are different, it seems impossible to relate the detection limits to the nature of the element based on comparison of these values between different elements. There is a general tendency, however, for the detection limits of monoisotopic elements to be lower than those of other elements because the signal intensities for each isotope are, in principle, proportional to its abundance. A much higher detection limit was observed for As because it must be measured in a very high-resolution mode because of $ArCl^+$, and the high value for Si is due to a high blank value possibly originating from the sample introduction glassware and/or a torch made from quartz. The disappointing detection limits for Ir and Pt appear to be due to higher blank signals presumably originating from the sample introduction tube made of Ir–Pt alloy. In contrast to this, the higher detection limits for P and S apparently resulted from a combination of the lower degree of ionization for these elements[4] and the requirement for a higher resolution (i.e. around 1000) mode to be used.

6.2 Multiple-collector Instruments

The precision of isotope ratio measurements attainable by ICPMS are much poorer than those by thermal ionization mass spectrometry (TIMS), which is widely used as the benchmark method for the very precise determination of isotope ratios in such fields as isotope geology and nuclear sciences. Apparently, this limited precision is a consequence of the unstable nature of ICP. Theoretically, considerable numbers of errors caused by the instability of the ion source can be eliminated if each isotope ion is measured simultaneously by using multiple-collector systems. This type of instrument was introduced into the marketplace by VG Elemental in 1994 (Figure 16). Precision of the order of 0.03–0.003% has been reported for these instruments.[164,165] At present, this type of instrument is provided by three different manufacturers

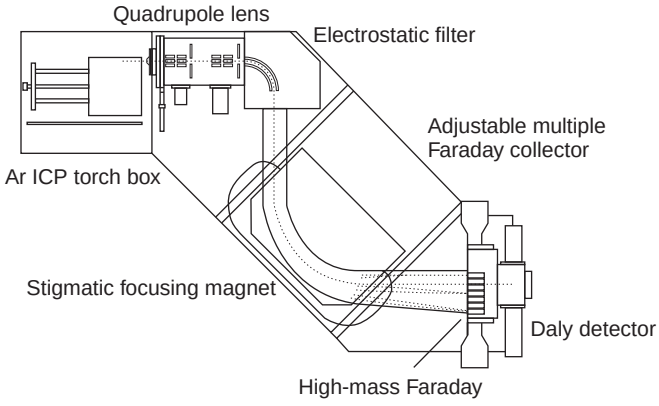

Figure 16 Schematic diagram of a double-focusing multicollector ICPMS, P54. (Reproduced with permission from VG Elemental, UK.)

and more than 40 sets of machines are working all over the world.[166]

The advantages of this instrument over TIMS are as follows:

1. Sample preparation procedures are simplified.
2. Measurement time is much shorter (several minutes) than TIMS (typically several hours).
3. Mass bias correction can be made using separate elements.
4. Precise ratios can be measured even for transient signals originating from such techniques as ETV and laser ablation.

Another type of multiple-collector instrument with the collision cell interface (see section 6.4) has been introduced into the marketplace by Micromass UK Ltd.

6.3 Time-of-flight Instruments

Myers et al.[167–169] described a novel type of ICPMS combined with a TOF mass spectrometer (ICP/TOF/MS). In the TOF instrument, separation of ions is based on the measurement of the TOF of ions of known energy over a known distance. As it is possible to measure the flight times of all the ions, TOF/MS is very efficient compared with other types of mass spectrometer, where only a very small fraction of the ions are measured in the collector system. Detection limits were from 0.4 to 2 ppb.[169] ICP/TOF/MS is currently provided by two manufacturers (GBC Scientific Equipment Inc. and LECO Corporation).

6.4 Collision Cell Interface Inductively Coupled Plasma Mass Spectrometry

The usefulness of collision cells to attenuate polyatomic ion interferences in ICPMS has been studied by Douglas,[170] Rowan and Houk[171] and Bricker et al.[172] A new type of quadrupole ICPMS equipped with a hexapole RF collision cell (Platform ICP) was launched by Micromass UK Ltd. The collision cell is placed between the skimmer cone and the quadrupole mass analyzer to reduce interference resulting from Ar-based molecular ions. The detection limits for easily ionizable elements such as Pb, In, Sr, were less than 1 ppt, and those for difficult elements including Fe, Ca, K, Cr, As and Se were around 10 ppt.[173] A similar type of instrument (Dynamic Reaction Cell) is also provided by Perlin Elmer/Sicex Corporation. However, further work on more difficult samples seems to be necessary before the usefulness of this very new technique can be evaluated fully.

6.5 Low Sample Consumption Nebulizers

Two types of nebulizer with much lower sample uptake rate have been introduced into the marketplace. One is a high-efficiency nebulizer (HEN) provided by J E Meinhard Associates,[174] and the other is a microconcentric nebulizer (MCN) (Figure 17) from CETAC Technologies.[175] The MCN comprises a nebulizer body with polyimide capillary and spray chamber end-cap adapter. The latter is used to connect the sample tubing to standard Scott-type spray chambers.

These kinds of devices including DIN (section 3.3.3), and several other prototypes not yet commercially available have clear analytical advantages for environmental samples. First, they will often be useful where the amounts of samples are limited, such as animal tissues and body fluids. Second, the reduction in sample amounts immediately results in the reduction of the analytical cost needed for sample collection and storage.

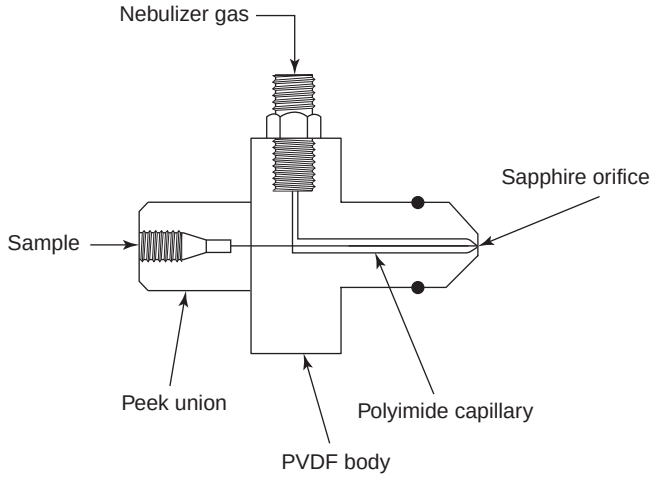

Figure 17 Schematic diagram of an MCN. (Modified from the original drawing provided by CETAC Technologies.)

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
DIN	Direct Injection Nebulizer
ETV	Electrothermal Vaporization
FI	Flow Injection
GC	Gas Chromatography
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HEN	High-efficiency Nebulizer
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LC	Liquid Chromatography
MCN	Microconcentric Nebulizer
MS	Mass Spectrometry
RF	Radiofrequency
TIMS	Thermal Ionization Mass Spectrometry
TOF	Time-of-flight
USN	Ultrasonic Nebulizer

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Flow-injection Techniques in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Soil Instrumental Methods

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications • Time-of-flight Mass Spectrometry

REFERENCES

1. D.W. Koppenaal, 'Atomic Mass Spectrometry', *Anal. Chem.*, **64**, 320R–342R (1992).
2. J.R. Bacon, A.T. Ellis, A.W. McMahon, P.J. Potts, J.G. Williams, 'Atomic Spectrometry Update – Atomic Mass Spectrometry and X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **9**, 267R–305R (1994).
3. A. Montaser, J.A. McLean, H. Liu, J.-M. Mermet, 'An Introduction to ICP Spectrometries for Elemental Analysis', in *Inductively Coupled Plasma Mass Spectrometry*, ed. A. Montaser, Wiley-VCH, New York, 1–31, 1998.
4. K.E. Jarvis, A.L. Gray, R.S. Houk (eds.), *Handbook of Inductively Coupled Plasma Mass Spectrometry*, Blackie, Glasgow, 1992.
5. A.R. Date, A.L. Gray (eds.), *Applications of Inductively Coupled Plasma Mass Spectrometry*, Blackie, Glasgow, 1989.
6. B.L. Sharp, 'Pneumatic Nebulisers and Spray Chambers for Inductively Coupled Plasma Spectrometry, A Review, Part 1. Nebulisers', *J. Anal. At. Spectrom.*, **3**, 613–652 (1988).
7. B.L. Sharp, 'Pneumatic Nebulisers and Spray Chambers for Inductively Coupled Plasma Spectrometry, A Review, Part 2. Spray Chambers', *J. Anal. At. Spectrom.*, **3**, 939–963 (1988).
8. J.E. Meinhard, 'The Concentric Glass Nebulizer', *ICP Inform. Newslett.*, **2**, 163–165 (1976).
9. R.H. Knisely, H. Amensen, C.C. Bulter, V.A. Fassel, 'An Improved Pneumatic Nebulizer for Use at Lower Nebulizing Gas Flow', *Appl. Spectrosc.*, **28**, 285–286 (1974).
10. R.S. Babington, 'It's Superspray', *Popular Sci.*, May, 102–104 (1973).
11. R.F. Suddendorf, K.W. Boyer, 'Nebulizer for Analysis for High Salt Content Samples with Inductively Coupled Plasma Emission Spectrometry', *Anal. Chem.*, **50**, 1769–1771 (1978).
12. J.G. Williams, A.L. Gray, P. Norman, L. Ebdon, 'Feasibility of Solid Sample Introduction by Slurry Nebulisation for Inductively Coupled Plasma-Mass Spectrometry', *J. Anal. At. Spectrom.*, **2**, 469–472 (1987).
13. L.M. Ebdon, E. Foulkers, H.G.M. Parry, C.T. Tye, 'Direct Atomic Spectrometric Analysis by Slurry Atomization. Part 7. Analysis of Coals Using Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **3**, 753–761 (1989).
14. R.H. Wendt, V.A. Fassel, 'Induction-coupled Plasma Spectrometric Excitation Source', *Anal. Chem.*, **37**, 920–922 (1965).
15. K.W. Olson, W.J. Haas, V.A. Fassel, 'Multielement Detection Limits and Sample Nebulization Efficiencies of an Improved Ultrasonic Nebulizer and a Conventional Pneumatic Nebulizer in Inductively Coupled Plasma-atomic Emission Spectroscopy', *Anal. Chem.*, **49**, 632–637 (1977).

16. V.A. Fassel, B.R. Bear, 'Ultrasonic Nebulization of Liquid Sample for Analytical Inductively Coupled Plasma-atomic Spectroscopy: An Update', *Spectrochim. Acta*, **41B**, 1089–1113 (1986).
17. W.B. Robbins, J.A. Caruso, 'Development of Hydride Generation Method for Atomic Spectroscopic Analysis', *Anal. Chem.*, **51**, 889A–899A (1979).
18. T. Nakahara, 'Applications of Hydride Generation Techniques in Atomic Absorption, Atomic Fluorescence, and Plasma Atomic Emission Spectroscopy', *Prog. Anal. Atom. Spectrosc.*, **6**, 163–223 (1983).
19. D.E. Baker, N.H. Suhr, 'Atomic Absorption and Flame Emission Spectrometry', in *Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties*, 2nd edition, eds. A.L. Page, R.H. Miller, D.R. Keeney, American Society of Agronomy, and Soil Science Society of America, Madison, Wisconsin, 13–27, 1982.
20. A.M. Ure, 'Atomic Absorption and Flame Emission Spectrometry', in *Soil Analysis: Modern Instrumental Techniques*, 2nd edition, ed. K.A. Smith, Marcel Dekker, New York, 1–62, 1991.
21. A.M. Gunn, D.L. Millard, G.F. Kirkbright, 'Optical Emission Spectrometry with an Inductively Coupled Radio-frequency Argon Plasma Source and Sample Introduction with a Graphite Electrothermal Vaporisation Device. Part 1. Instrumentation Assembly and Performance Characteristics', *Analyst*, **103**, 1066–1078 (1978).
22. C.J. Park, J.C. Van Loon, P. Arrowsmith, J.B. French, 'Design and Optimization of an Electrothermal Vaporizer for Use in Plasma Source Mass Spectrometry', *Can. J. Spectrosc.*, **32**, 29–36 (1987).
23. A.L. Gray, 'Solid Sample Introduction by Laser Ablation for Inductively Coupled Plasma Mass Spectrometry', *Analyst*, **110**, 551–556 (1985).
24. M. Thompson, J.E. Goulter, F. Sieper, 'Laser Ablation for the Introduction of Solid Samples into an Inductively Coupled Plasma for Atomic-emission Spectrometry', *Analyst*, **106**, 32–39 (1981).
25. P.R. Hesse, *A Textbook of Soil Chemical Analysis*, John Murray, London, 1971.
26. C.H. Lim, M.L. Jackson, 'Dissolution for Total Elemental Analysis', in *Method of Soil Analysis. Part 2. Chemical and Microbiological Properties*, 2nd edition, eds. A.L. Page, R.H. Miller, D.R. Keeney, American Society of Agronomy, and Soil Science Society of America, Madison, Wisconsin, 1–12, 1982.
27. P.J. Potts, *A Handbook of Silicate Rock Analysis*, Blackie, Glasgow, 1987.
28. B. Bernas, 'A New Method for Decomposition and Comprehensive Analysis of Silicates by Atomic Absorption Spectrometry', *Anal. Chem.*, **42**, 1682–1686 (1968).
29. F.J. Langmyhr, P.E. Pause, 'The Analysis of Siliceous Materials by Atomic Absorption Spectrophotometry and the Hydrofluoric Acid Decomposition Technique. Part 1: The Analysis of Silicates Rocks', *Anal. Chim. Acta*, **3**, 397–408 (1968).
30. R. Uhrberg, 'Acid Digestion Bomb for Biological Samples', *Anal. Chem.*, **54**, 1906–1908 (1982).
31. H. Matsusiewicz, R.E. Sturgeon, 'Present Status of Microwave Sample Dissolution and Decomposition for Elemental Analysis', *Prog. Anal. Spectrosc.*, **12**, 21–39 (1989).
32. H.M. Kingston, P.J. Walter, 'The Art and Science of Microwave Sample Preparations for Trace and Ultratrace Elemental Analysis', in *Inductively Coupled Plasma Mass Spectrometry*, ed. A. Montaser, Wiley-VCH, New York, 33–81, 1998.
33. R.S. Houk, 'Mass Spectrometry of Inductively Coupled Plasmas', *Anal. Chem.*, **58**, 97A–105A (1986).
34. G.M. Pierzynski, 'Past, Present, and Future Approaches for Testing Metals for Environmental Concerns and Regulatory Approaches', *Commun. Soil Sci. Plant Anal.*, **29**, 1523–1536 (1998).
35. J.S. Becker, H.-J. Dietze, 'Ultratrace and Precise Isotope Analysis by Double-focusing Sector Field Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1057–1063 (1998).
36. J.S. Becker, H.-J. Dietze, 'Double-focusing Sector Field Inductively Coupled Mass Spectrometry for Highly Sensitive Multi-element and Isotopic Analysis. Invited Lecture', *J. Anal. At. Spectrom.*, **12**, 881–889 (1997).
37. A. Parriaux, M. Bensimon, M. Looser, 'Contributions and Use of the High Resolution Plasma Source Mass Spectrometry Technique in Environmental Geology', *Analusis*, **24**, M16–M18 (1996).
38. H. Haraguchi, K. Inagaki, 'Analysis for Environmental Monitoring: Analysis of Sediment, River Water and Seawater Samples by ICP-MS and ICP-AES', *Bunseki*, 494–503 (1998).
39. H. Haraguchi, A. Itoh, 'The Global Environment and Analyses. Analyses of the Aquatic Environment. Heavy Metals', *Kikai Kagaku Sosetsu*, **29**, 35–57 (1996).
40. F. Sus, J. Krtil, Z. Malek, M. Beran, 'Determination of Trace Radioactivity of Thorium, Uranium, Neptunium, Plutonium, Americium, and Curium in Swipes and Environmental Samples', *Nucleon*, **1**, 13–22 (1998).
41. R.N. Sah, P.H. Brown, 'Techniques for Boron Determination and their Application to the Analysis of Plant and Soil Samples', *Plant Soil*, **193**, 15–33 (1997).
42. G.E.M. Hall, 'Analytical Perspective on Trace Element Species of Interest in Exploration', *J. Geochem. Explor.*, **61**, 1–19 (1998).
43. H. Tao, 'Trace Elemental Speciation by Chromatography/Inductively Coupled Plasma Mass Spectrometry', *Bunseki Kagaku*, **46**, 239–263 (1997).
44. Z. Ni, 'Coupled Chromatography-atomic Spectrometry for the Determination of Trace Organometallic Compounds in Environmental and Biological Samples', *Guangpuxue Yu Guangupu Fenxi*, **15**, 87–94 (1995).

45. M.K. Donais, 'How to Interface a Liquid Chromatograph to an Inductively Coupled Plasma-mass Spectrometer for Elemental Speciation Studies', *Spectroscopy*, **13**, 32–35 (1998).
46. C. Rivas, L. Ebdon, S.J. Hill, 'Speciation of Organotin Compounds Utilizing HPLC-ICP-MS: Application to the Measurements and Testing (BCR) Certification Program', *Quim. Anal.*, **14**, 142–147 (1995).
47. L. Jia, X. Chen, X.X. Wang, Y.P. Musheng, 'Application of Capillary Electrophoresis in Speciation Analysis', *Sepu*, **16**, 402–405 (1998).
48. R.M. Barnes, 'Capillary Electrophoresis and Inductively Coupled Plasma Spectrometry (CE-ICP-MS). Status Report', *Fresenius' J. Anal. Chem.*, **361**, 246–251 (1998).
49. R. Wennrich, K. Grunke, A. Walther, 'Environmental Analyses with Electrothermal-vaporization Inductively Coupled Plasma Mass Spectrometry (ETV-ICP-MS)', *Nachr. Chem. Tech. Lab.*, **45**, 291–295 (1997).
50. H. Mukai, J. Yoshinaga, 'Isotope Ratio Analysis of Environmental and Biological Samples by ICP-MS', *Kankyo Kagaku*, **5**, 789–798 (1995).
51. S.M. Jerome, D. Smith, M.J. Woods, S.A. Wood, 'Metrology of Plutonium for Environmental Measurements', *Appl. Radiat. Isot.*, **46**, 1145–1150 (1995).
52. R.E. Wolf, E. Denoyer, Z. Grosser, 'Overview and Comparison of ICP-MS Methods for Environmental Analyses', *Proc. Water Qual. Technol. Conf.*, 1297–1304 (1996).
53. R.E. Wolf, Z.A. Grosser, 'Overview and Comparison of ICP-MS Methods for Environmental Analyses', *At. Spectrosc.*, **18**, 145–151 (1997).
54. E. Hoffmann, C. Luedke, H. Stephanowitz, 'Application of Laser-ICP-MS in Environmental Analysis', *Fresenius' J. Anal. Chem.*, **355**, 900–903 (1996).
55. J.S. Crain, 'Applications of Inductively Coupled Plasma-mass Spectrometry in Environmental Radiochemistry', *Spectroscopy*, **11**, 30–39 (1996).
56. Y. Muramatsu, S. Yoshida, 'Analysis of Multielement Environmental Samples by ICP-MS', *Hoshasen Kagaku*, **40**, 164–170 (1997).
57. A. Miyazaki, 'Subjects and Prospect Regarding Environmental Measurements', *Shigen to Kankyo*, **4**, 261–265 (1995).
58. S. Yamasaki, A. Tsumura, Y. Takaku, 'Ultra-trace Elements in Terrestrial Water as Determined by High Resolution ICP-MS', *Microchem. J.*, **49**, 305–318 (1994).
59. K. Akatuka, J.W. McLaren, S.S. Berman, 'Determination of 7 Trace Elements in Environmental Water Samples by Inductively Coupled Plasma Mass Spectrometry with On-line Preconcentration', *Bunseki Kagaku*, **42**, 423–428 (1993).
60. H. Klinkenberg, T. Beeren, W. Vanborm, F. Vanderlinden, M. Raets, 'The Use of an Enriched Isotope as an On-line Internal Standard in Inductively Coupled Plasma Mass Spectrometry: A Reference Method for a Proposed Determination of Tellurium in Industrial Waste by Means of Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **48**, 649–661 (1993).
61. G.E.M. Hall, J.C. Pelchat, 'Determination of Palladium and Platinum in Fresh Water by Inductively Coupled Plasma Mass Spectrometry and Activated Charcoal Preconcentration', *J. Anal. At. Spectrom.*, **8**, 1059–1063 (1993).
62. M.J. Powell, E.S.K. Quan, D.W. Boomer, D.R. Wiederin, 'Inductively Coupled Plasma Mass Spectrometry with Direct Injection Nebulization for Mercury Analysis of Drinking Water', *Anal. Chem.*, **64**, 2253–2257 (1992).
63. E.S. Hall, E. Murphy, 'Determination of Sources of Lead in Tap Water by Inductively Coupled Plasma Mass Spectrometry (ICP-MS)', *J. Radioanal. Nucl. Chem. Lett.*, **175**, 129–138 (1993).
64. P.L. Trentini, M. Ascanelli, B. Zanforlini, F. Venturini, G. Bucchi, F. Fagioli, 'Determination of Aluminum by Inductively Coupled Plasma Mass Spectrometry in Serum of Patients Treated by Haemodialysis, Dialysis Solutions and Tap Water, and a Comparison with Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **8**, 905–909 (1993).
65. Y.V. Yi, A. Masuda, 'Simultaneous Determination of Ruthenium, Palladium, Iridium, and Platinum at Ultra-trace Levels by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry in Geological Samples', *Anal. Chem.*, **68**, 1444–1450 (1996).
66. E.H. Bailey, A.J. Kemp, K.V. Ragnarsdottir, 'Determination of Uranium and Thorium in Basalts and Uranium in Aqueous Solution by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**, 551–556 (1993).
67. J.S. Crain, B.L. Mikesell, 'Detection of Sub ng/L Actinides in Industrial Wastewater Mixtures by Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **46**, 1498–1502 (1992).
68. I.B. Brenner, H.E. Taylor, 'A Critical Review of Inductively Coupled Plasma Mass Spectrometry for Geoanalysis, Geochemistry, and Hydrology. Part I. Analytical Performance', *Crit. Rev. Anal. Chem.*, **23**, 355–367 (1992).
69. K. Shiraishi, T. Nakajima, Y. Takaku, A. Tsumura, S. Yamasaki, I.P. Los, I.Y. Kamarikov, M.G. Buzinny, A.V. Zelensky, 'Elemental Analysis of Fresh Water Samples Collected in the Former USSR by Inductively Coupled Plasma Mass Spectrometry', *J. Radioanal. Nucl. Chem. Articles*, **173**, 313–321 (1993).
70. K.S. Huang, S.J. Jiang, 'Determination of Trace Levels of Metal Ions in Water Samples by Inductively Coupled Plasma Mass Spectrometry After On-line Preconcentration on SO_3-oxine CM-Cellulose', *Fresenius' J. Anal. Chem.*, **347**, 238–242 (1993).
71. H.J. Yang, K.S. Huang, S.J. Jiang, C.C. Wu, C.H. Chou, 'Determination of Trace Metal Ions in Water Samples

by On-line Preconcentration and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem. Acta*, **282**, 437–443 (1993).

72. V.S. Burakov, A.V. Isaevich, P.Y. Misakov, P.A. Naumenkov, S.N. Raikov, 'Intracavity Laser Spectroscopic Method for Determining Trace Amounts of Iodine and Barium in Water and Biological Samples', *J. Anal. At. Spectrom.*, **9**, 307–309 (1994).
73. D. Beauchemin, J.W.P. McLaren, P. Mykytiuk, S.S. Berman, 'Determination of Trace Metals in a River Water Reference Material by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **39**, 778–783 (1987).
74. K. Akatuka, J.W. McLaren, J.W. Lam, S.S. Berman, 'Determination of Iron and 10 Other Trace Elements in the Open Ocean Seawater Reference Material NASS-3 by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **7**, 889–894 (1992).
75. R.J. Rosenberg, R. Zillacus, P.K.G. Manninen, 'Determination of Transition Metals in the Primary Water of Pressurized Water Reactors by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **9**, 713–717 (1994).
76. J.I. Garcia Alonso, A. Sanzmedel, L. Ebdon, 'Determination of Butyltin Ion Species by Ion-exchange Chromatography with Inductively Coupled Plasma Mass Spectrometry and Spectrofluometric Detection', *Anal. Chim. Acta*, **283**, 261–271 (1993).
77. G. Klinkhammer, C.R. German, H. Elderfield, M.J. Greaves, A. Mitra, 'Rare Earth Elements in Hydrothermal Fluids and Plume Particulates by Inductively Coupled Plasma Mass Spectrometry', *Mar. Chem.*, **45**, 179–186 (1994).
78. W. Zhu, E.W. De Leer, M. Kennedy, G.J.F.R. Alaerts, 'Study of the Preconcentration and Determination of Ultratrace Rare Earth Elements in Environmental Samples with an Ion Exchange Micro-column', *Fresenius' J. Anal. Chem.*, **360**, 74–80 (1998).
79. H. Sawatari, E. Hujimori, H. Haraguchi, 'Multi-element Determination of Trace Elements in Seawater by Gallium Coprecipitation and Inductively Coupled Plasma Mass Spectrometry', *Anal. Sci.*, **11**, 369–373 (1995).
80. M.B. Shabani, S.K. Sahoo, A. Masuda, 'Determination of Bismuth at Ultratrace Levels in Seawater by Inductively Coupled Plasma Mass Spectrometry After Preconcentration with Solvent Extraction and Back-extraction', *Analyst*, **117**, 1477–1480 (1992).
81. S. Morita, V. Tobita, M. Kurabayashi, 'Determination of Technetium-99 in Environmental Samples by Inductively Coupled Plasma Mass Spectrometry', *Radiochem. Acta*, **63**, 63–67 (1993).
82. S. Sumiya, S. Morita, K. Tobita, M. Kurabayashi, 'Determination of Technetium-99 and Neptunium-237 in Environmental Samples by Inductively Coupled Plasma Mass Spectrometry', *J. Radioanal. Nucl. Chem. Articles*, **177**, 149–159 (1994).
83. M.L. Wells, K.W. Bruland, 'An Improved Method for Rapid Preconcentration and Determination of Bioactive Trace Metals in Seawater Using Solid Phase Extraction and High Resolution Inductively Coupled Plasma Mass Spectrometry', *Mar. Chem.*, **63**, 145–153 (1998).
84. S. Uchida, K. Tagami, 'Improvement of Tc Separation Procedure Using a Chromatographic Resin for Direct Measurement by ICP-MS', *Anal. Chim. Acta*, **357**, 1–3 (1997).
85. A. Miyazaki, R.A. Reimer, 'Determination of Lead Isotope Ratios and Concentrations in Sea-water by Inductively Coupled Plasma Mass Spectrometry After Preconcentration Using Chelex-100', *J. Anal. At. Spectrom.*, **8**, 449–452 (1993).
86. L. Ebdon, A. Fisher, H. Handley, P. Jones, 'Determination of Trace Metals in Concentrated Brines Using Inductively Coupled Plasma Mass Spectrometry On-line Preconcentration and Matrix Elimination with Flow Injection', *J. Anal. At. Spectrom.*, **8**, 979–981 (1993).
87. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Determination of Trace Metals in Sea-water and the On-line Removal of Matrix Interferences by Flow Injection with Inductively Coupled Plasma Mass Spectrometric Detection', *J. Anal. At. Spectrom.*, **9**, 935–938 (1994).
88. A. Seubert, G. Petzold, J.W. McLaren, 'Synthesis and Application of an Inert Type of 8-hydroxyquinoline-based Chelating Ion Exchanger for Sea-water Analysis Using On-line Inductively Coupled Plasma Mass Spectrometry Detection', *J. Anal. At. Spectrom.*, **10**, 371–379 (1995).
89. A. Stroh, U. Vollkopf, 'Optimization and Use of Flow Injection Vapor Generation Inductively Coupled Plasma Mass Spectrometry for the Determination of Arsenic, Antimony and Mercury in Water and Sea-water at Ultratrace Levels', *J. Anal. At. Spectrom.*, **8**, 35–40 (1993).
90. J.T. Creed, M.L. Magnuson, C.A. Brockhoff, I. Chamberlain, M. Siveganesan, 'Arsenic Determination in Saline Waters Utilizing a Tubular Membrane as a Gas-liquid Separator for Hydride Generation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 505–509 (1996).
91. C.J. Hwang, S.J. Jiang, 'Determination of Arsenic Compounds in Water Samples by Liquid Chromatography Inductively Coupled Plasma Mass Spectrometry with an In Situ Nebulizer-hydride Generator', *Anal. Chim. Acta*, **289**, 205–213 (1994).
92. J.W. McLaren, J.W.H. Lam, S.S. Berman, K. Akatuka, M.A. Azeredo, 'On-line Method for the Analysis of Seawater for Trace Elements by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**, 279–286 (1993).
93. M.B. Shabani, A. Masuda, 'Determination of Trace Rhenium in Sea Water by Inductively Coupled Plasma Mass Spectrometry with On-line Preconcentration', *Anal. Chim. Acta*, **261**, 315–321 (1992).

94. J.K. Aggarwal, M.B. Shabani, M.R. Palmer, K.V. Ragnarsdottir, 'Determination of the Rare Earth Elements in Aqueous Samples at Sub-ppt Levels by Inductively Coupled Plasma Mass Spectrometry and Flow Injection ICP-MS', *Anal. Chem.*, **68**, 4418–4423 (1996).
95. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, 'Speciation of Arsenic Compounds in Drinking Water by Capillary Electrophoresis with Hydrodynamically Modified Electroosmotic Flow Detected through Hydride Generation Inductively Coupled Plasma Mass Spectrometry with a Membrane Gas-liquid Separator', *J. Anal. At. Spectrom.*, **12**, 689–695 (1997).
96. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, 'Speciation of Arsenic Compounds by Ion Chromatography with Inductively Coupled Plasma Mass Spectrometry Detection Utilizing Hydride Generation with a Membrane Separator', *J. Anal. At. Spectrom.*, **11**, 893–898 (1996).
97. D. Amouroux, E. Tessier, C. Pecheyran, O.F.X. Donald, 'Sampling and Probing Volatile Metal(loid) Species in Natural Waters by In-situ Purge and Cryogenic Trapping Followed by Gas Chromatography and Inductively Coupled Plasma Mass Spectrometry (P-CT-GC-ICP/MS)', *Anal. Chim. Acta*, **377**, 241–254 (1998).
98. G. Pinte, M.A.R. Veado, A.H. Oliveira, M. Khalis, S. Ayrault, G. Revel, 'Comparison of Neutron Activation Analysis and ICP-MS Used for River Water Pollution Control', *Hydrobiologia*, **373/374**, 61–73 (1998).
99. V. Riganti, A. Lombardi, O. Conio, F. Bonfiglioli, 'Application of the ICP-MS [Inductively Coupled Plasma-Mass Spectrometry] Technique for the Analysis for Heavy Metals in Samples of Rain Water', *Riv. Merceol.*, **36**, 201–216 (1997).
100. A. Cocherie, P. Negrel, S. Roy, C. Guerrot, 'Direct Determination of Lead Isotope Ratios in Rainwater Using Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1069–1073 (1998).
101. N.C. Munksgaard, G.J. Batterham, D.L. Parry, 'Lead Isotope Ratios Determined by ICP-MS: Investigation of Anthropogenic Lead in Seawater and Sediment from the Gulf of Carpentaria, Australia', *Mar. Pollut. Bull.*, **36**, 527–534 (1998).
102. P. Thomas, 'Significance of Combined HPLC-ICP-MS – Study of Chemical Forms of Arsenic and Selenium in Water – Toward Use in Routine Analysis', *Analusis*, **25**, M35–M36 (1997).
103. R.C. Richter, S.R. Koirtyohann, S.S. Jurisson, 'Determination of Technetium-99 in Aqueous Solutions by Inductively Coupled Plasma Mass Spectrometry: Effects of Chemical Form and Memory', *J. Anal. At. Spectrom.*, **12**, 557–562 (1997).
104. M. Kikuchi, T. Kato, 'Determination of Lead and Cadmium in Sea Water by Inductively Coupled Plasma Mass Spectrometry', *Sendai-shi Eisei Kenkyushoho*, **25**, 223–228 (1997).
105. A.J.H. Groemping, P. Ostapczuk, H. Emons, 'Wet Deposition in Germany: Long-term Trends and the Contribution of Heavy Metals', *Chemosphere*, **34**, 2227–2236 (1997).
106. P. Schramel, I. Wendler, 'Inductively Coupled Plasma Mass Spectrometry (ICP-MS) with Electrothermal Vaporization (ETV)', *CANAS '95*, 137–142 (1996).
107. D.B. Taylor, H.M. Kingston, D.J. Nogay, D. Koller, R. Hutton, 'Online Solid-phase Chelation for the Determination of Eight Metals in Environmental Waters by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 187–191 (1996).
108. J.H. Aldstadt, J.M. Kuo, L.L. Smith, M.D. Erickson, 'Determination of Uranium by Flow Injection Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **319**, 135–143 (1996).
109. M.J. Duane, S. Facchetti, 'On-site Environmental Water Analyses by ICP-MS', *Sci. Total Environ.*, **172**, 133–143 (1995).
110. K. Tagami, S. Uchida, 'Determination of ^{99}Tc in Rain and Dry Fallout by ICP-MS', *J. Radioanal. Nucl. Chem.*, **197**, 409–416 (1995).
111. N. Momoshima, M. Sayad, T. Takashima, 'Determination of ^{99}Tc in Coastal Seawater Collected in Fukuoka, Japan', *J. Radioanal. Nucl. Chem.*, **197**, 245–251 (1995).
112. R.S. Houk, V.A. Fassel, G.D. Flesch, H.J. Svec., A.L. Gray, C.E. Taylor, 'Inductively Coupled Argon Plasma as an Ion Source for Mass Spectrometric Determination of Trace Elements', *Anal. Chem.*, **52**, 2283 (1980).
113. D. Florian, R.M. Barnes, G. Knapp, 'Comparison of Microwave-assisted Leaching Techniques for the Determination of Heavy Metals in Sediments, Soils, and Sludges', *Fresenius' J. Anal. Chem.*, **362**, 558–565 (1998).
114. M.E. Farago, P. Kavanagh, R. Blanks, J. Kelly, G. Kazantzis, D.H. Trevor, G.E.M. Hall, 'Platinum Concentrations in Urban Road Dust and Soil, and in Blood and Urine in the United Kingdom', *Analyst*, **123**, 451–454 (1998).
115. T. Muto, T. Shimogawara, 'Determination of Technetium-99 in Soil Samples by High Performance Liquid Chromatography Coupled to Inductively Coupled Plasma Mass Spectrometry', *Tokyo-toritsu Aisotopu Sogo Kenkyusho Kenkyu Hokoku*, **14**, 79–88 (1997).
116. P. Thomas, J.K. Finnie, J.G. Williams, 'Feasibility of Identification and Monitoring of Arsenic Species in Soil and Sediment Samples by Coupled HPLC-inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 1367–1372 (1997).
117. S.Y. Lee, D.R. Watkins, B.L. Jackson, R.L. Schmoyer, D.A. Lietzke, B.B. Burgoa, J.T. Branson, J.T. Ammons, 'Evaluation of Metal and Radionuclide Data from Neutron Activation and Acid-digestion-based Spectrometry Analyses of Background Soils. Significance in Environmental Restoration', *J. Radioanal. Nucl. Chem.*, **217**, 57–64 (1997).

118. K. Tagami, S. Uchida, 'Concentration of Global Fallout ^{99}Tc in Rice Paddy Soils Collected in Japan', *Environ. Pollut.*, **95**, 151–154 (1997).
119. M.J. Duane, S. Facchetti, G. Pigozzi, 'Site Characterization of Polluted Soils and Comparison of Screening Techniques for Heavy Metals by Mobile ICP-MS, GFAAS/ICP-AES (Fixed Laboratory) and EDXRF (Fixed Laboratory)', *Sci. Total Environ.*, **177**, 195–214 (1996).
120. M. Hollenbach, J. Grohs, M. Kroft, S. Mamich, 'Determination of ^{230}Th, ^{234}U, ^{239}Pu, and ^{240}Pu in Soils by ICP-MS Using Flow-injection Preconcentration', *ASTM Spec. Tech. Publ.*, **STP 1291**, 99–115 (1995).
121. J. Goossens, L. Moens, R. Dams, 'Inductively Coupled Plasma Mass Spectrometric Determination of Heavy Metals in Soil and Sludge Candidate Reference Materials', *Anal. Chim. Acta*, **304**, 307–315 (1995).
122. X. Guo, F.E. Lichte, 'Analysis of Rocks, Soils and Sediments for the Chalcophile Elements by Laser Ablation-inductively Coupled Plasma Mass Spectrometry', *Analyst*, **120**, 2707–2711 (1995).
123. S.T.G. Anderson, R.V.D. Robert, H.N. Farrer, 'Determination of Total and Leachable Arsenic and Selenium in Soils by Continuous Hydride Generation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **9**, 1107–1110 (1994).
124. S.F. Durrant, N.I. Ward, 'Rapid Multielemental Analysis of Chinese Reference Soils by Laser Ablation Inductively Coupled Plasma-source Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **345**, 512 (1993).
125. P.T. Brown, G. Picchioni, M. Jenkin, H. Mu, 'Use of ICP-MS and Boron-10 to Trace the Movement of Boron in Plants and Soil', *Commun. Soil Sci. Plant Anal.*, **23**, 2781 (1992).
126. B. Magyar, H. Cousin, B. Aeschlimann, 'Combined Use of X-ray Fluorescence and Laser Ablation Inductively Coupled Plasma Mass Spectrometry in the Analysis of Soils', *Anal. Proc.*, **29**, 282 (1992).
127. C. Jeffrey, J. Hansel, J.E. Troxel, 'Determining Regulated Metals in Soil', *Spectroscopy*, **7**, 40–46 (1992).
128. D.J. Chittleborough, D.M. Hotchinn, R. Beckett, 'Sedimentation Field-flow Fractionation: A New Technique for the Fractionation of Soil Colloids', *Soil Sci.*, **153**, 341–348 (1992).
129. S. Yamasaki, 'Modern Instrumental Techniques for Soil Analysis. 1. Inductively Coupled Plasma Mass Spectrometry', *Nippon Dojo Hiryogaku Zasshi*, **63**, 245–250 (1992).
130. D. Beauchemin, R.K. Micklethwaite, G.W. VanLoon, G.W. Hay, 'Determination of Metal-organic Associations in Soil Leachates by Inductively Coupled Plasma Mass Spectrometry', *Chem. Geol.*, **95**, 187–198 (1992).
131. X. Jin, Y.Y. Cheung, 'Determination of Trace Manganese, Cobalt, Nickel, Copper, Zinc, Arsenic, Molybdenum and Strontium in Cabbage, Turnip, Soyrlan and Soil by Inductively Coupled Plasma Mass Spectrometry', *Fenxi Huaxue*, **19**, 430–432 (1991).
132. T. Yoshida, S. Yamasaki, A. Tsumura, I. Yamada, K. Aoki, 'Chemical Analysis of Micro- to Trace-elements in the Reference Standard Specimens of Soil by Photon Activation Analysis and Inductively Coupled Plasma Mass Spectrometry', *Kakuriken Kenkyu Hokoku (Tohoku Daigaku)*, **23**, 251–270 (1990).
133. S. Yamasaki, Y. Tamura, 'Total Analysis of Soils by ICP-AES and ICP-MS (Mass Spectrometry)', *Commun. Soil Sci. Plant. Anal.*, **21**, 2017–2028 (1990).
134. S. Yamasaki, T. Yoshida, A. Tsumura, G. Murase, 'ICP-MS: One of the Most Promising Techniques for Comprehensive Studies of Heavy Metals in Soils', 14th International Congress of Soil Science, Kyoto, Japan, *Trans. 14th Cong. Soil Sci.*, II-108–II-113 (1990).
135. S. Yamasaki, A. Tsumura, D. Cai, 'Microwave Dissolution and ICP-MS for the Determination of Trace and Ultra-trace Elements in Plant Tissue', in *Application of Plasma Source Mass Spectrometry*, eds. G. Holland, A.N. Eaton, Royal Society of Chemistry, Cambridge, 110–118, 1991.
136. M.A. Vaughan, G. Horlick, 'Oxide, Hydroxide and Doubly Charged Species in Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **40**, 434–445 (1986).
137. S. Yamasaki, A. Takeda, M. Nanzyo, A. Yasui, 'Single Grain Analysis for Determining the Authencity of ''High Quality'' Rice', in *Plasma Source Mass Spectrometry: New Developments and Applications*, eds. G. Holland, S.D. Tanner, Royal Society of Chemistry, Cambridge, 182–190, 1999.
138. A. Kawasaki, H. Watanabe-Oda, 'Distribution of Boron Isotope Ratios in Brown Rice and Soil Extracts Determined by ICP-MS', in *Plasma Source Mass Spectrometry: New Developments and Applications*, eds. G. Holland, S.D. Tanner, Royal Society of Chemistry, Cambridge, 173–181, 1999.
139. R.J. Watling, 'Sourcing the Provenance of Cannabis Crops Using Inter-element Association Patterns ''Fingerprinting'' and Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 917–926 (1998).
140. R. Gwiazda, D. Woolard, D. Smith, 'Improved Lead Isotope Ratio Measurements in Environmental and Biological Samples with a Double Focussing Magnetic Sector Inductively Coupled Plasma Mass Spectrometer (ICP-MS)', *J. Anal. At. Spectrom.*, **13**, 1233–1238 (1998).
141. Y. Gelinas, G.V. Iyengar, R.M. Barnes, 'Total Iodine in Nutritional and Biological Reference Materials Using Neutron Activation Analysis and Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **362**, 483–488 (1998).
142. K. Tagami, S. Uchida, M. Garcia-Leon, 'Comparison of a Radiation Counting Method and ICP-MS for the Determination of ^{99}Tc in Environmental Samples', *J. Radioanal. Nucl. Chem.*, **234**, 147–151 (1998).

143. M. Krachler, A. Alimonti, F. Petrucci, K.J. Irgolic, F. Forastiere, S. Caroli, 'Analytical Problems in the Determination of Platinum-group Metals in Urine by Quadrupole and Magnetic Sector Field Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **363**, 1–10 (1998).
144. C. Lutter, G.V. Iyengar, R. Barnes, D. Amarasiriwardena, M. Kotrebai, A. Krushevska, T. Kazbekova, T. Sharmanov, 'Concentrations of Toxic Metals and Radionuclides in Kazakhstani Breast Milk', *Harmonization Health Relat. Environment Meas. using Nuclear Isotope Technology Proceedings International Symposium 1996*, International Atomic Energy Agency, Vienna, Austria, 405–419, 1997.
145. P.W. Kirlew, 'Development of Interfaces for Capillary Electrophoresis with Inductively Coupled Plasma Mass Spectrometric Detection and Supercritical Fluid Extraction of Metal-based Species from Environmental Samples', *Diss. Abst. Int.*, **58**, 5926 (1997).
146. J.M. Llobet, S. Granero, M. Schuhmacher, J. Corbella, J.L. Domingo, 'Biological Monitoring of Environmental Pollution and Human Exposure to Metals in Tarragona, Spain. Part 2. Levels in Autopsy Tissues', *Trace Elem. Electrolytes*, **15**, 44–49 (1998).
147. J. Riondato, F. Vanhaecke, L. Moens, R. Dams, 'Determination and Trace and Ultratrace Elements in Human Serum with a Double Focusing Magnetic Sector Inductively Coupled Plasma Mass Spectrometer', *J. Anal. At. Spectrom.*, **12**, 933–937 (1997).
148. J. Begerow, M. Turfeld, L. Dunemann, 'Determination of Physiological Palladium, Platinum, Iridium and Gold Levels in Human Blood Using Double Focusing Magnetic Sector Field Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 1095–1098 (1997).
149. M. Bettinelli, S. Spezia, U. Baroni, G. Bizzari, 'Multielement Characterization of Lichens and Vegetable Materials by ICP-MS', *Spec. Publ. – R. Soc. Chem.*, **202**, 211–222 (1997).
150. H. Hintelmann, R. Falter, G. Ilgen, R.D. Evans, 'Determination of Artifactual Formation of Monomethylmercury ($MeHg^+$) in Environmental Samples Using Stable Hg^{2+} Isotopes with ICP-MS Detection. Calculation of Contents Applying Species Specific Isotope Addition', *Fresenius' J. Anal. Chem.*, **358**, 363–370 (1997).
151. C.J. Park, J.K. Suh, 'Determination of Trace Elements in Rice Flour by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 573–577 (1997).
152. J. Yoshinga, M. Morita, 'Determination of Mercury in Biological and Environmental Samples by Inductively Coupled Plasma Mass Spectrometry with the Isotope Dilution Technique', *J. Anal. At. Spectrom.*, **12**, 417–420 (1997).
153. G.E. Segovia, A.J.I. Garcia, A. Sanz-Medel, 'Determination of Butyltin Compounds by Means of Hydride Generation-cold Trapping-gas Chromatography Coupled to ICP-MS Detection', *Spec. Publ. – R. Soc. Chem.*, **202**, 182–191 (1997).
154. P. Schramel, I. Wendler, J. Angerer, 'Determination of Metals (Antimony, Bismuth, Lead, Cadmium, Mercury, Palladium, Platinum, Tellurium, Thallium, Tin, and Tungsten) in Urine Samples by Inductively Coupled Plasma-mass Spectrometry', *Int. Arch. Occup. Environ. Health*, **69**, 219–223 (1997).
155. R.M. Olivas, O.F.X. Donard, N. Gilson, M. Potin-Gautier, 'Speciation of Organic Selenium Compounds by High-performance Liquid Chromatography–Inductively Coupled Plasma Mass Spectrometry in Natural Samples', *J. Anal. At. Spectrom.*, **11**, 1171–1176 (1996).
156. K. Honda, H. Koyama, 'Analysis of Selenium Biological Constituents Using LC-ICP-MS', *Nippon Iyo Masu Supekutoru Gakkai Koenshu*, **21**, 119–120 (1996).
157. C.J. Park, J.K. Suh, S.H. Lee, 'Determination of Cd and Pb in Human Blood by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry. International Comparison', *Anal. Sci. Technol.*, **9**, 145–160 (1996).
158. B.G. Ting, D.C. Daniel, K.L. Caldwell, 'Determination of Thorium and Uranium in Urine with Inductively Coupled Argon Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 339–342 (1996).
159. E. Fujimoto, Y. Tomosue, H. Haraguchi, 'Determination of Rare Earth Elements in Blood Serum Reference Sample by Chelating Resin Preconcentration and Inductively Coupled Plasma Mass Spectrometry', *Tohoku J. Exp. Med.*, **178**, 63–74 (1996).
160. R.M. Barnes, 'Analytical Plasma Source Mass Spectrometry in Biomedical Research', *Fresenius' J. Anal. Chem.*, **355**, 433–441 (1996).
161. H. Mukai, J. Yoshinaga, 'Isotope Ratio Analysis of Environmental and Biological Samples by ICP-MS', *J. Environ. Chem.*, **5**, 789–798 (1995).
162. N. Bradshaw, E.F.H. Hall, N.E. Sanderson, 'Inductively Coupled Plasma as Ion Source for High-resolution Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **4**, 801–803 (1989).
163. M. Morita, H. Itoh, T. Uehiro, K. Otsuka, 'High Resolution MS with an ICP Ion Source', *Anal. Sci. (Jpn)*, **5**, 609–610 (1989).
164. A.J. Walder, D. Koller, N.M. Reed, R.C. Hutton, P.A. Freedman, 'Isotope Ratio Measurement by Inductively Coupled Plasma Multiple Collector Mass Spectrometry Incorporating a High Efficiency Nebulization System', *J. Anal. Atom. Spectrom.*, **8**, 1037–1041 (1993).
165. H. Hirata, 'Lead Isotopic Analyses of NIST Standard Reference Materials Using Multiple Collector Inductively Coupled Plasma Mass Spectrometry Coupled with a Modified External Correction Method for Mass Discrimination Effect', *Analyst*, **121**, 1407–1411 (1996).
166. C.B. Douthitt, 'Magnetic Sector ICP-MS: Comprehensive Bibliographies of HR-ICP-MS and MC-ICP-MS', *ICP Newslett.*, **25**, 87–90 (1999).

167. D.P. Myers, G. Li, P. Yang, G.M. Hieftje, 'An Inductively Coupled Plasma-time-of-flight Mass Spectrometer (ICP-TOFMS) for Elemental Analysis. Part I. Optimization and Characteristics', *J. Anal. At. Spectrom.*, **5**, 1008–1016 (1994).
168. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An Inductively Coupled Plasma-time-of-flight Mass Spectrometer (ICP-TOFMS) for Elemental Analysis. Part II. Optimization and Characteristics', *J. Anal. At. Spectrom.*, **6**, 400–410 (1995).
169. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An Inductively Coupled Plasma-time-of-flight Mass Spectrometer (ICP-TOFMS) for Elemental Analysis. Part III. Analytical Performance', *J. Anal. At. Spectrom.*, **6**, 411–420 (1995).
170. D.J. Douglas, 'Some Recent Perspectives on ICP-MS', *Can. J. Spectrosc.*, **34**, 38–49 (1989).
171. J.T. Rowan, R.S. Houk, 'Attenuation of Polyatomic Ion Interferences in Inductively Coupled Mass Spectrometry by Gas-phase Collisions', *Appl. Spectrosc.*, **43**, 976–980 (1989).
172. T.M. Bricker, F.G. Smith, R.S. Houk, 'Charge Transfer Reactions Between Xenon Ions and Iron Atoms in an Argon–Xenon Inductively Coupled Plasma', *Spectrochim. Acta, Part B*, **50**, 1325–1335 (1995).
173. P.J. Turner, T. Merren, J. Speakman, C. Haines, 'Interface Studies in the ICP-MS Spectrometer', in *Plasma Source Mass Spectrometry: Development and Applications*, eds. G. Holland, S.D. Tanner, Royal Society of Chemistry, Cambridge, 28–34, 1997.
174. H. Tan, B.A. Meinhard, J.E. Meinhard, 'Recent Investigation of Meinhard Concentic Nebulizers', Presented at the 19th Annual Meeting of the Federation of Analytical Chemistry and Spectroscopy Societies, Philadelphia, Pennsylvania, USA, September, 1992.
175. A.F. Vanhaecke, M. Van Holderbeke, L. Moens, R. Dams, 'Evaluation of a Commercially Available Microconcentric Nebulizer for Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **11**, 543–548 (1996).

Industrial Waste Dumps, Sampling and Analysis

Winfried Rasemann
Freiberg University of Mining and Technology, Freiberg, Germany

1 Introduction 1
2 Basic Problems of Sampling and Analysis of Industrial Wastes and Waste Dumps 2
3 Some Experiences 4
3.1 Geostatistical and Statistical Evaluation of a Heterogeneous Industrial Site 4
3.2 Evaluation of a Mercury-contaminated Site – A Problem of Sampling and Sample Pretreatment 18
Acknowledgments 24
List of Symbols 24
Abbreviations and Acronyms 25
Related Articles 25
References 25

Industrial sites where residues and wastes such as slags, ashes, dust and sludges have been dumped are essential parts of the environment and of the economic structure. The amount of waste produced, distributed and deposited is constantly increasing. The wastes can contain hazardous components that pollute and endanger the environment, but they can also consist of valuable materials, which are a source of secondary raw materials. To assess the environmental risk caused by the waste, or to calculate the economic benefit of dumped material, a reliable knowledge of the waste composition is required. Waste management experience has regularly shown that conflicts and lawsuits are the result if the composition of the waste materials is difficult to determine reliably. Investigations carried out by different institutions and persons, or by the same personnel under varying conditions, will often have very different results. The measuring technology and the measuring methods cannot be the only reasons for that. Nowadays, it is possible to accurately determine chemical components in any natural concentration, and there is no problem in distinguishing the size and shape of particles down to the nanometer scale. The problems are created because the wastes are mixtures of particles and lumps which vary in size and shape as well as in chemical composition and physical properties. A waste dump with a varied production history and dumping conditions, with chemical reactions or physical changes occurring after dumping, will be heterogeneous as a rule. Therefore, the evaluation of any dump of industrial waste materials is quite difficult. To ensure that the results of evaluation are comparable, certain regulations of investigation must be followed. According to the delivery, the wastes are classified into material streams (stationary, moving or free falling), heaps delivered within containers and vehicles, and free-standing heaps. As it is economically unjustifiable to investigate the entire waste dump, subsets of material (called samples) must be taken from the stream, the container or the heap in order to determine the measurements of interest. In doing this, rather a lot of errors can be made. Therefore, the errors that exist are frequently caused by unobjective sampling and improper handling of the sample material. However, a false estimation can also be made by evaluating reliable data using unsuitable statistical methods.

The present contribution takes up this problem. An industrial waste dump and a contaminated site were chosen as examples, and proven sampling regulations were applied. The applicability of known statistical and geostatistical methods developed for homogeneous granular bulk solids and uniformly contaminated soils to the evaluation of extremely heterogeneous waste dumps and nonuniformly contaminated industrial sites is shown.

1 INTRODUCTION

The evaluation of heterogeneous granular and lumpy industrial wastes by sampling and statistical data analysis is still a problem worldwide. First of all the sample must be obtained objectively, and secondly the sample material must be properly stored and preserved.

Sampling is objective under the following conditions:

- the sampler (person taking the sample) knows what material is being presented;
- the sampler knows the aim of the investigation;
- the sampler acts strictly according to certain regulations to reach the aim of the investigation;
- the sampler justifies these regulations to himself and to the client, and considers that the method of sampling prescribes how to evaluate the results.

For homogeneous granular mixtures and mining goods such as powders, sand, gravel, salt, iron ore and solid fuels, these problems have been thoroughly investigated. There are detailed instructions for sampling, for common procedures between the sampler and the laboratory, and for agreement between contracting parties.[1–3] The

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

evaluation of the results depending on the sampling scheme is controlled by mathematical models.[4,5] The handling of instructions and methods is justified by national and international standards, the applications of which are obligatory within the framework of the respective economic and legal systems.[6,7] The regulations, mathematical models, instructions and standards for sampling are based on the assumption of a small grain size range and a uniform particle shape. The methods of data analysis are based on mathematical statistics and geostatistics. They suppose a homogeneous isotropic material with uniform variation of the characteristic under consideration.

In contrast, industrial wastes can be extremely heterogeneous with respect to their material composition and to the degree of mixing of the individual portions. The concentration of harmful substances or the content of valuable material can vary considerably depending on the origin of the waste. Waste can consist of particles and lumps, with extreme distributions of particle size and shape. Characterization of the size and shape of lumps is not yet regulated by specific measures and dimensions.[8] Within an individual lump, several components can be intermixed, melted, screwed, welded or stuck together. Recycled products, wastes and scraps can also contain lubricants, cooling agents, brake fluids and other substances which make the material sticky, lumpy and greasy.

Pollutants in contaminated sites and soils can occur in complicated, partly volatile compounds which demand a lot from the handling and analysis of the samples.[9,10] Plants and animals can also serve as indicators of pollution. They require specific goals for sampling and special equipment for trace analysis[11,12] as well as particular care in sample preparation[13,14] and sample pretreatment.[15]

As a consequence, the assumptions underlying the geostatistics are not always correctly fulfilled by industrial waste dumps. There are still no special regulations expressly valid for the sampling of heterogeneous waste materials, and no statistical models specifically evolved for the evaluation of data obtained from wastes. Further research work is necessary in this field.[16] Previously, recommendations and statistical methods which are satisfactory or promising for other solids systems that are structurally or materially similar to the respective waste dump have been used.[17]

In the present article, a slag heap from Saxony, Germany, is used as an example in the application of sampling regulations and of known statistical methods for homogeneous bulk solids to the evaluation of extremely heterogeneous coarse grained and lumpy solids systems. The article also deals with the problem of the composition of a material being changed by improper handling after sampling. Soil samples taken from a mercury-contaminated industrial site are used as an example. It is shown that the mercury content detected by analysis is systematically decreased by improper preservation of the samples. Frequently, soil pollution is influenced by the geogenic background; it varies significantly as the location within the site is changed and as the sampling depth is varied. The investigations are planned in order to quantify the systematic error of sample preservation and to simultaneously distinguish the sample value from the geogenic background. For this reason, a statistical approach is proposed.

2 BASIC PROBLEMS OF SAMPLING AND ANALYSIS OF INDUSTRIAL WASTES AND WASTE DUMPS

The aim of sampling and analysis is to obtain a representative evaluation of the waste material. In doing sampling and chemical or granulometric analysis, two opposing principles have to be brought into line. On the one hand, the sample should be as large as possible, but this is not economical or practical. As a compromise, a sample having at least a certain minimum mass is required. The minimum sample mass is established according to certain mathematical models or empirically founded rules. This quantity increases as the particle size increases and as the reliability, which is required to describe the composition of the entire waste material by sample, enlarges. On the other hand, the amount of material needed for chemical or granulometric analysis is very very low. To get a final sample for analysis, the composition of which substantially represents the properties of the original sample material, very expensive sample preparation procedures must be applied such as comminution, homogenization (by mixing or melting), and dividing and reducing of the material. With these procedures, some errors are unavoidable. But sample preparation is not the only source of errors which influence the results of the investigation. All procedures, from sample selection up to statistical data analysis, cause unavoidable errors. Their effects on the evaluation are estimated as follows.

1. *Sampling* is the selection of samples from the waste material. Error at this stage is caused by global and local variations of the waste material and can amount to up to 1000% of the true value.
2. *Sample pretreatment* includes all events which can occur as a result of the handling of the samples, from their selection until preparation for analysis, and which cause the original properties to be changed. Volatilization or adsorption of components are examples of processes by which the material is

changed; pretreatment errors can amount up to 100% of the true value.

3. *Sample preparation* includes the processes of comminution, homogenization, division and reduction of the original material to establish the final sample for analysis. The maximum error of sample preparation is between 100% and 300% of the true value.
4. *Instrumental measurement*; this error is vanishingly small compared with those from the preceeding steps, and amounts to a maximum of 2–20% of the true value.
5. *Data analysis* is applied to solve the scientific problem using all the data available. If an incorrect mathematical model is used for this purpose it can cause an error up to 50%.

A simplified scheme of the unavoidable errors caused by the different steps in the material evaluation is shown in Figure 1.

Consequently, false sampling and gross defects in handling the material may influence the results of the investigation more seriously than the instrumental measurements. Generally, the following principles are valid:

- if the sampling is false, everything is false;
- if the measured values are right, but the method of data evaluation is unsuited for the intended purpose, the results of the evaluation may be doubtful.

To limit the errors of evaluation, every step from sampling up to data analysis must be carried out strictly according to certain regulations and specific mathematical methods. To plan the investigation, the following questions are asked.

1. What is the special purpose of the investigation and what is already known about the waste dump or the industrial site?
2. What additional information is striven for?

To answer these questions, the following parameters are established:

1. The sampling design, which defines the time, the location and other conditions of sampling that are essential for the characterization of the waste material (dump, stream, site). When planning the sampling conditions, the sampler has to imagine an appropriate structure for the object of investigation.
2. The number of samples.
3. The minimum mass of an individual sample and the gross sample mass.
4. The sample handling procedures, which are necessary to ensure that the data are of the required quality.

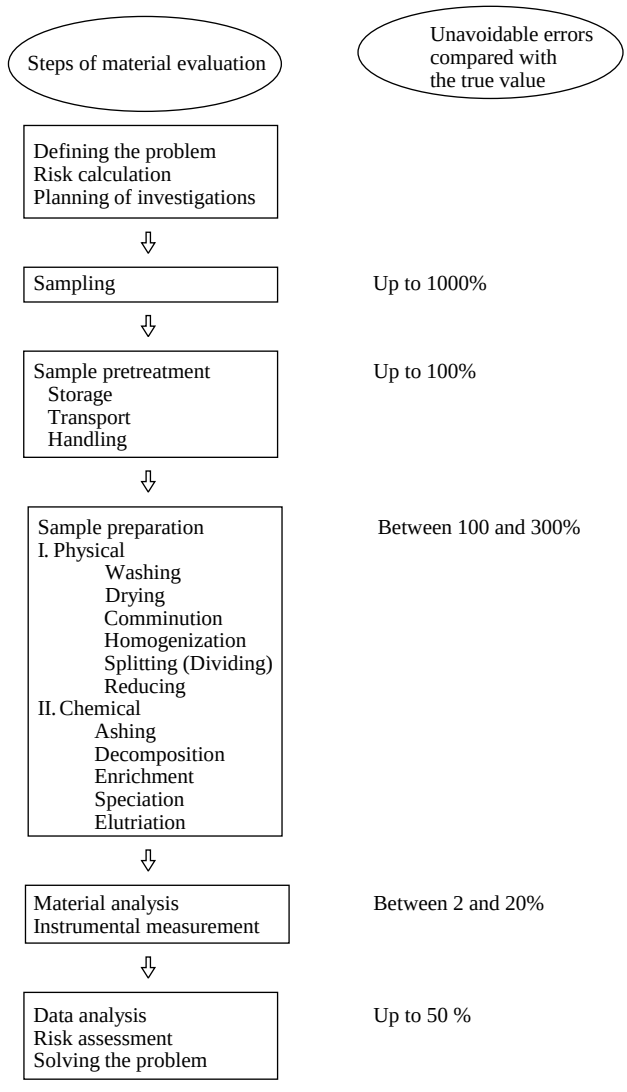

Figure 1 Simplified scheme of the steps in the evaluation of material systems.

Handling includes all the procedures for transport, preservation, preparation, decomposition for chemical analysis, etc.

5. The quantity to be measured.
6. The methods of statistical data analysis. These methods have to be suitable for the sampling design.

The success of sampling and data analysis is measured by

- the accuracy (precision)
- the confidence (trustworthiness) and
- the reliability

of the waste characterization.

The gist of these criteria is explained as follows.

1. Accuracy or precision: Let μ be the true value of a characteristic of interest (e.g. the content of a hazardous component). This value is not known and can only be estimated from samples. An error $|\overline{Y}_n - \mu|$ is made by using the mean value $\overline{Y}_n$ obtained from n samples instead of the true value μ. This error is called the *accuracy of estimation* (accuracy of prediction) of μ by the mean $\overline{Y}_n$. In practice, an upper limit ε_0 of $|\overline{Y}_n - \mu|$ is prescribed. The limit ε_0 is termed the *maximum admissible error of the mean*.
2. Confidence: Because the accuracy of estimation $|\overline{Y}_n - \mu|$ is a random variable, a probability β must be given, where the relation $|\overline{Y}_n - \mu| \leq \varepsilon_0$ is valid. The value β is called the *confidence probability* of the admissible error of the mean. In practice, a β value between 0.90 and 0.99 is recommended.
3. Reliability: When using inaccurate data, a certain risk (e.g. expected loss of confidence, additional costs) must be taken into account. If the risk associated with a data set is acceptable, it is called a *reliable* data set.

In the following section, some experiences in sampling and sample pretreatment are presented.

3 SOME EXPERIENCES

3.1 Geostatistical and Statistical Evaluation of a Heterogeneous Industrial Site

3.1.1 Object Investigated

The slag heap of a smelting plant at Freiberg in Saxony, Germany, was investigated. Further studies have also been carried out.[14,18,19] Rasemann and Markert[14] have explained the methodical concept of investigation and the statistical methods used for evaluating the results of sampling and analysis. In this article some of the practical conclusions reached in the papers mentioned with the help of additional data have been proven to be true. The slag was a by-product from the thermal processing of zinc blende (ZnS) ore concentrates. The concentrates were roasted in a fluidized-bed roasting furnace which produced calcined zinc oxide and sulfur dioxide in the roaster gas. The roaster gas was used to produce sulfuric acid, which was used to leach the calcine. The leaching was followed by separation into solid low-grade leaching tailings and high-grade leachate. The leachate was an electrolyte, from which the zinc was cathodically precipitated. The leaching tailings consisted of the remaining portions of zinc and other valuable components such as lead, cadmium and iron. The tailings were washed and dried, and then thermally processed in a rotary furnace to extract the metals. Fine coke was added to reach a high furnace temperature and to reduce the metal oxides. The residue of this process, the rotary furnace slag, was continuously discharged, then cooled, granulated and transported to a slag pit. Finally, the slag was moved by truck from the pit to the dump.

The leaching tailings processed were delivered from different factories so they varied in composition over the entire operating period. Consequently, the rotary furnace slag was composed of widely varying portions of SiO_2 (10–20%), FeO (15–30%), CaO (10–15%), Al_2O_3 (about 6%), C (10–45%), S (about 6%), Zn (1–5%), Pb (0.1–2.5%) and Cu (0.3–0.6%). Cadmium and arsenic also occurred in low concentrations. The high carbon content resulted from unburnt coke. After the slag was dumped, the fine coke in the heap was carbonized at low temperatures, parts of the normally grained slags were incinerated, the heap caved in and the slags were sintered to blocks here and there. The very heterogeneous slag heap thus consisted of blocks and lumps of slag and of coarse-grained bulk material. The chemical composition was extremely varied. After the first inspection, the heap surface was distinguished by three regions (Figure 2):

- normally grained slag which was partly burnt
- very incinerated regions
- solid blocks and lumps of sintered slag.

The heap extended over an area of 1.6 ha and it was between 4 and 12 m high. The total mass was about 80 kt.

3.1.2 Sampling Strategy

3.1.2.1 Sampling Pattern, Number of Samples and Mass of an Individual Sample For evaluation, the area was covered by a regular grid consisting of 60 points

Figure 2 View at the rotary furnace slag heap of the smelting plant at Freiberg in Saxony.

forming equilateral triangles with sides of 10 m. The efficiency of such a grid is evaluated as follows.[20] If there is no information about the spatial distribution of contaminants or valuable components, the largest uninvestigated subsection of the grid may be a criterion for the probability of incorrect evaluation. To compare different grids, such a subsection is defined as the area of a circle drawn to connect adjacent grid points. A grid with minimal uninvestigated subsections compared with any other geometric pattern is assumed to be optimal. This condition is found to be fulfilled by a grid of equilateral triangles. A square and an equilateral triangle, as respective units of a regular grid, are compared with each other in Figure 3. The squared grid completely neglects a circular area of 157 m^2 and can only recognize circular contaminations exceeding a radius of 7.07 m with certainty. In contrast, only 104.7 m^2 remains unrecognizable, and circular contaminations having a minimum radius of 5.77 m can be completely identified by a triangular grid. The proposed sampling pattern and other geometric schemes were officially introduced into the field of soil survey and investigation of contaminated sites by the International Organization for Standardization (ISO) with their International Standard for Soil Quality Sampling, ISO 10381.[21] This standard serves as the design for sampling programs, for agreement between the customer and the contractor, or appropriate authorities, and for common action between the sampler and the analytical laboratory. The main emphases of the different parts of ISO 10381, and their relationship to the German standards for soil investigation, have been explained.[22] Additionally, some general remarks must be made about practical situations where it is not possible or desirable to cover the entire project area or parts of it with a regular grid. This can result in an irregular geometric arrangement of sampling points because of special conditions, such as morphology (rivers, lakes, mountains, railroad tracks), the distribution of property, curvilinear boundaries between estates, and so on. Some useful irregular patterns have been proposed.[23]

One surface sample each of grained slag was taken from the slagheap at each of 56 grid points of the equilateral triangle pattern at distances 10 m apart. Sampling from the remaining four points of the 60-point pattern was impossible because of sintered and crusted blocks and lumps. Additionally, eight individual surface samples were taken from each of ten different points of the pattern to accurately determine the chemical composition and physical properties of the slagheap and the error of sampling. Subsurface samples were obtained by percussion coring at 18 points of an equilateral triangular grid, each side of which was $dx = 20$ m long (see Figure 3). A percussion core 1 m long was considered to be an individual sample. Taking into account the

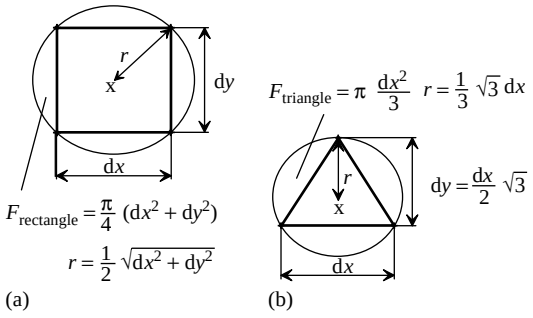

Figure 3 Area of the circumscribed circle and the corresponding radius belonging to the cells of a rectangular grid (a) and of a grid of equilateral triangles (b).

inside diameter of the percussion coring tube (40 mm) and an average bulk density of 1.1 g cm^{-3}, the mass of an individual sample was about 1.4 kg. In practice, a lower sample mass caused by unavoidable core loss during boring has been taken into account. Because of the high iron content and the predominantly grained structure of the slag, the masses and the numbers of samples were compared with the International Sampling Guidelines for Iron Ores, ISO 3081,[24] and with the German standard for sieve analysis from the Deutsches Institut Für Normung (DIN) 18123.[25] According to these standards, the required sample mass and the minimum number of samples depend on the maximum grain size, the grain size of the sieve curve at 5% oversize.

In the case of the granular rotary furnace slag, a maximum grain size of 2.5–25 mm was obtained using sieve analyses of the eight surface samples from each of the ten sampling points. Taking these data as a basis, DIN 18123 requires a minimum mass for an individual sample of 0.2–3 kg, (see Table 1). The ISO 3081 standard calls for a mass of 0.3–4 kg under these conditions (see Table 2). In addition, this standard prescribes 180 individual samples for a test material mass of 70–100 kt. Following these recommendations, six or seven samples

Table 1 Minimum sample masses required for different grain sizes, according to the German guidelines DIN 18123[25] for the determination of the sieve curve of granular materials

Maximum grain size (mm)	Minimum sample mass (kg)
2	0.15
5	0.30
10	0.70
20	2.00
30	4.00
40	7.00
50	12.00
60	18.00

Table 2 Minimum sample mass and number of samples required according to the International Standard for Iron Ores, ISO 3081[24]

Maximum grain size (mm)	Minimum sample mass for every consignment (kg)
>10	0.30
10–22.4	0.80
22.4–50	4.00
50–100	12.00
100–150	40.00
150–200	190.00
Mass of consignment (t)	**Number of individual samples**
45 000–70 000	160
70 000–100 000	180
100 000–150 000	200
150 000–210 000	220

from each of the 18 grid points of the percussion coring investigation are necessary, in addition to the 56 surface samples. Therefore, the sampling strategy chosen by us did not contradict the regulations now valid for bulk solids handling.

Appropriate rules are lacking for sampling extremely heterogeneous wastes, as has been repeatedly emphasized.[26,27] General guidelines are being developed for sampling wastes but at most they contain only rough information. An overview of the European standards in this field is given by Knoop.[28] These standards require chemical analyses, which unfortunately have not yet been controlled by corresponding regulations and European standards. Until now, regulations for the chemical analysis of wastes have been used, which have been developed only for water, soil, sludge and sediments or for materials handled in the fuel and mineral oils industry. However, there are exact regulations regarding the chemical analyses of contaminated soils, such as the German Deutsche Einheitsverfahren (DEV) standards for water, sludges and sewage, but rules for the sample mass and the number of samples at contaminated sites are not included. The guidelines only contain general information. Consequently, the legal status of a person doing sampling (a sampler) is quite a problem. For a detailed consideration of this aspect, consult Krutz and Niessing.[29] The judicial consequences resulting from incorrect decisions by the sampler have been discussed in detail.[30,31]

For the sample mass, the German guideline and PN 2/78K of the Länderarbeitsgemeinschaft Abfall (LAGA)[32] require that the minimum mass of an individual sample (in kg) is 0.06 times the maximum grain size (in mm). In the present situation (a maximum grain size between 2.5 and 25 mm), a minimum sample mass of 0.15–1.5 kg is recommended. This guideline also contains instructions for the number of samples required depending on the range of maximum particle sizes, the mass of the consignment and the delivery of the waste (see Table 3). Following these instructions, a maximum of 15 individual samples should be taken. Of course, this number is too small in the present case. However, the requirements of the German TA Abfall (Technical Instructions for Wastes)[33] are even more unrealistic. The TA Abfall modified and simplified the sampling instructions of the LAGA guideline PN 2/78K as follows. First, the waste classes ''delivered within containers'' and ''delivered without containers'' are distinguished. Second, waste delivered without containers is classified by appearance, as homogeneous or heterogeneous. This places the rotary furnace slag in the class of ''heterogeneous wastes being delivered to the dump without containers'', and therefore one individual sample of 1 kg mass per 5 t waste must be taken. For a slagheap of 80 kt total mass, 16 000 samples and a total sample mass of 16 000 kg would be necessary. This is a completely unreasonable recommendation.

3.1.3 Sample Preparation

3.1.3.1 Chemical Analysis and Granulometric Analyses Each individual sample was prepared in its final form by homogenization, splitting and reducing. The splitting and reduction involved a three-fold coning of the sample material and subsequent quartering of the final cone using a splitting cross. The two opposite quarters of the cone were used as final samples for chemical and granulometric analyses. The quarters were weighed, and the concentrations of certain chemical components and their elution behaviors were obtained using the DEV standards for water, sludges and sewage.[34] To determine the ignition loss, a German standard for foundation testing was applied. An overview of the DIN standards and the related United States Environmental Protection Agency (USEPA) standards applied for the investigation of waste is given in Table 4. Further details have been provided by Hein and Kunze.[35] The concentrations of the heavy metals Fe, Zn, Pb, Cu and Cd and the As content were analyzed by acid decomposition and subsequent AAS. The S content was determined by an oxidizing melting decomposition with $Na_2CO_3/K_2CO_3 + Na_2O_2$ according to DIN 38405, part 5.[36] The sulfur compounds were converted to water-soluble sulphates, and the barium chloride added to precipitate the sulphates was quantitatively measured by weighing. The ignition loss, a measure of the combustible matter, was determined at temperatures of 550 °C and 1000 °C, according to DIN 18128.[37] Additionally, the grain size distribution was obtained by sieve analysis following DIN 18123.[25]

Different strategies were used for sample preparation and analysis depending on the available sample mass

Table 3 Instructions relating to sample masses from German (LAGA) sampling guidelines valid for wastes[32]

Range of maximum particle sizes	Minimum number of individual samples					
	From moved waste		From stationary waste			
			In trucks	In deposits		
	Mass of consignment			Mass of consignment		
	up to 50 t	>50 t	3	up to 50 t	50–150 t	>150 t
<20 mm	3	3 every 50 t	per	3	3 every 50 t	8
>20 mm	5	1 every 10 t	truck	5	1 every 10 t	15

Minimum sample mass = 0.06 × maximum particle size.

Table 4 Overview of the different analytical chemistry methods used to investigate wastes and of the respective German (DIN) and American (USEPA) standards

Pollutant	Concentration range	Analytical method	Standard applied for waste
Heavy metals	Normal	ICPAES[a]	DIN 38406, part 22 USEPA 3050
Heavy metals	Low or traces	AAS[b]	DIN 38406 USEPA 3050
Heavy metals	Traces	ICPMS[c]	USEPA 6020 USEPA 200.8
Anions	Normal	Several methods	DIN 38405, part 5
Aliphatic hydrocarbons	Low	GC/MS[d]	DIN 38409, part 18
	High	IR[e]	USEPA 413.2 USEPA 418.1 USEPA 8440
Aromatic hydrocarbons		GC[f]	DIN 38407
PAHs[g]		HPLC[h]	DIN 38407
		GC/MS	USEPA 8310
High-volatility halogenic hydrocarbons		GC/MS	
Low-volatility halogenic hydrocarbons		GC	DIN 38407
PCBs[i]		GC/MS	DIN 38414
Dioxins and furans		GC/MS	
Nitrophenols		HPLC	USEPA 8333
Aromatic amino-compounds		HPLC	USEPA 8330
Phosphoric compounds		UV/VIS[j] spectrometry	DIN 38414
Herbicides and pesticides	High	HPLC	DIN 38407
	Low	(thin-layer chromatography)	USEPA 8325 DIN 38407
Surfactants		Potentiometry	
Ignition loss (sludges)		Combustion	DIN 38414, part 3
Ignition loss (granular materials)		Combustion	DIN 18128 for foundation testing
Elution behavior		Elution methods	DIN 38414, part 4

[a] ICPAES, inductively coupled plasma atomic emission spectrometry.
[b] AAS, atomic absorption spectrometry.
[c] ICPMS, inductively coupled plasma mass spectrometry.
[d] GC/MS, gas chromatography/mass spectrometry.
[e] IR, infrared.
[f] GC, gas chromatography.
[g] PAHs, polycyclic aromatic hydrocarbons.
[h] HPLC, high-performance liquid chromatography.
[i] PCBs, polychlorinated biphenyls.
[j] UV/VIS, ultraviolet/visible.

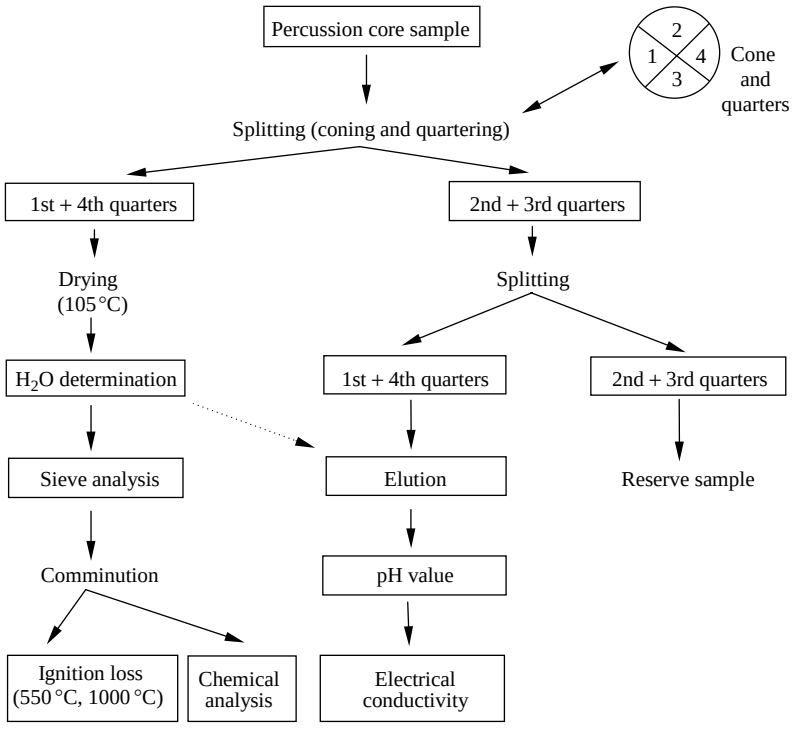

Figure 4 Scheme of preparation and analysis of percussion core samples with a mass from 400 g to 900 g.

in order to provide all the chemical and physical characteristics needed. The strategy used for a percussion core sample with a mass between 400 and 900 g is shown in Figure 4. Another variant for samples with a mass greater than 900 g was covered previously.[14] The preparation and analysis schemes used for repeated surface sampling and for a single sampling of surface points, respectively, are shown in Figure 5. According to these schemes, the grain size distribution and the ignition loss were determined for the respective first and fourth quarters of the eight individual samples from ten surface points, as a first step. Afterwards, the residuals of the handled first and fourth quarters and the untouched second and third quarters of these eight samples were mixed, and a composite sample was obtained.[14] Finally, the composite sample was prepared and analyzed as a single surface sample. In this way, 10 composite samples and 46 single surface samples were investigated.

3.1.4 Statistical Analysis of Data

3.1.4.1 Geostatistics The chemical composition of the rotary furnace slag was evaluated from the original surface sampling data and from the percussion coring data obtained at fixed depths. To explain the spatial variability of the chemical components, variograms were constructed. The technique of point kriging was used to predict additional values of these quantities at points inaccessible for sampling and observation.

Let G be the set of all possible points of the slagheap. Both the construction of the variogram and the technique of kriging are based on the proposition[38] that for any measuring quantity Y_t, $t \in G$ at a point t in G is a regionalized variable.[39] Following this theory of the regionalized variable, Y_t is locally a random variable, and the set of all measuring quantities $\{Y_t\}_{t \in G}$ globally describes a random field. The random field is characterized by its expected values EY_t at points t and by the covariances $\mathrm{cov}(Y_t, Y_{t+\tau})$ which describe the correlation between two measuring quantities obtained in the different locations t and $t+\tau$ of the heap G. This random field is assumed to have second order homogeneity, i.e. the expected value of any measuring quantity Y_t in the total heap G is equal to a constant μ. The covariance cov(.,.) observed in a particular direction does not depend on the absolute locations t and $t+\tau$, but is a function only of the distance τ between the locations. This is mathematically expressed by Equations (1) and (2).

$$EY_t = \mu \qquad \text{for all } t \in G \tag{1}$$

$$\begin{aligned} \mathrm{cov}(Y_t, Y_{t+\tau}) &= E[(Y_{t+\tau} - \mu)(Y_t - \mu)] \\ &= C_Y(\tau) \quad \text{for all } t \in G \end{aligned} \tag{2}$$

C_Y is called the *covariance function* of the random field.

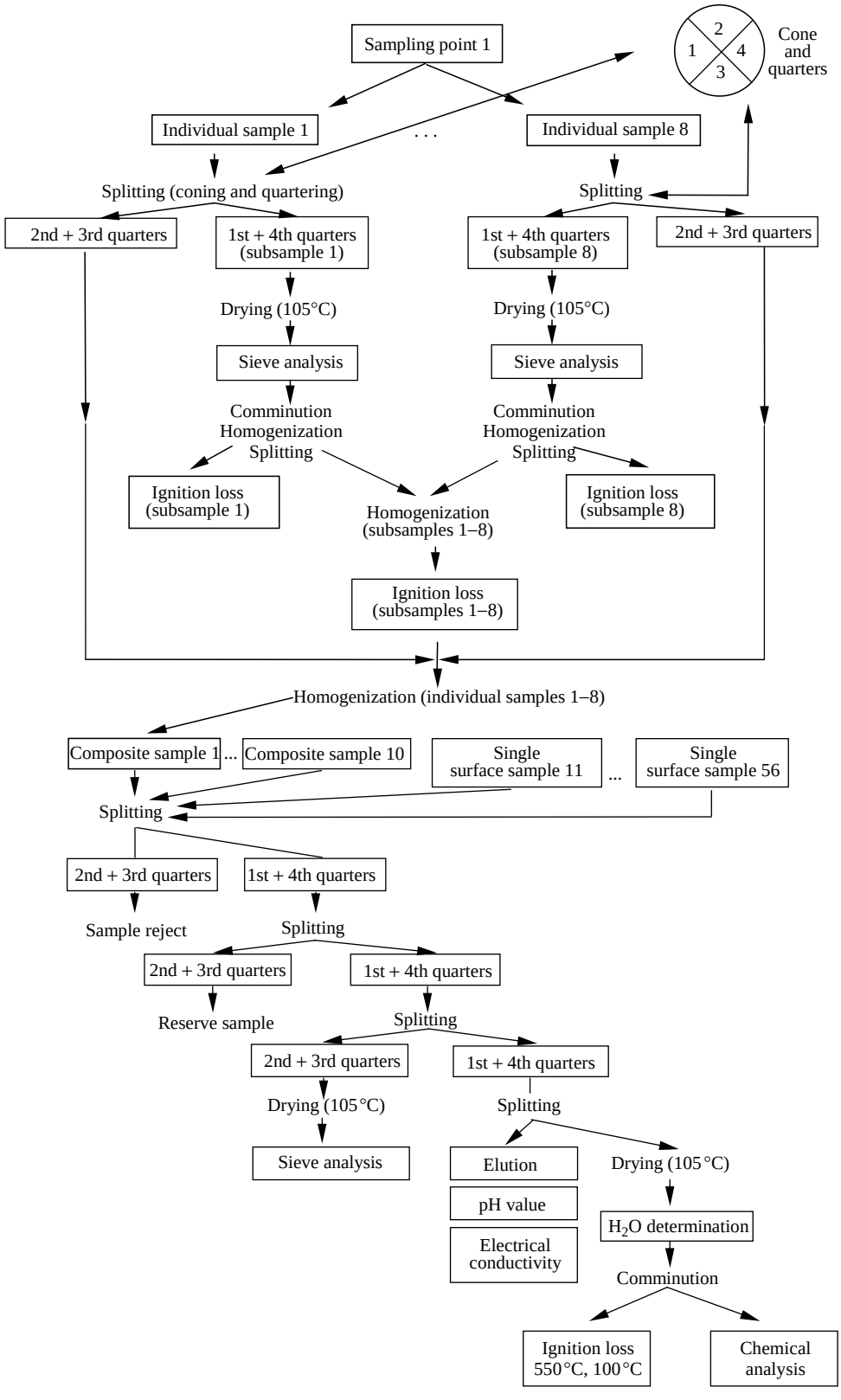

Figure 5 Scheme of preparation and analysis in repeated surface sampling and single sampling of surface points.

3.1.4.2 The Variogram Because the variances D^2Y_t und $D^2Y_{t+\tau}$ are equal to the covariance $C_Y(0)$ at a distance 0, the assumptions of homogeneity in Equations (1) and (2) respectively imply the following relationship between the variance and the covariance (Equation 3):

$$\begin{aligned} V(\tau) &= E[Y_{t+\tau} - Y_t]^2 = E[(Y_{t+\tau} - \mu) - (Y_t - \mu)]^2 \\ &= E[Y_{t+\tau} - \mu]^2 + E[Y_t - \mu]^2 - 2E[(Y_{t+\tau} - \mu)(Y_t - \mu)] \\ &= 2[C_Y(0) - C_Y(\tau)] \end{aligned} \tag{3}$$

The function V is called the *variogram function* of the random field, and it is twice the *semivariogram function* γ, as shown in Equation (4).

$$\gamma(\tau) = 0.5V(\tau) = C_Y(0) - C_Y(\tau) \tag{4}$$

In the following, the variance $D^2Y_t = C_Y(0)$ is represented as the simplified term S_Y^2.

The variogram function is a continuous one. Usually, this function is not immediately known and only its estimation at discrete distances is available (called the *empirical variogram function*). This estimation is obtained as follows. Let j and $j + \tau$ be any locations and Y_j and $Y_{j+\tau}$ the corresponding measurable values in these locations, respectively. Then, the value of the empirical variogram function at the distance τ is given by Equation (5).

$$\hat{V}(\tau) = \frac{1}{n(\tau)} \sum_{i=1}^{n(\tau)} [Y_{i+\tau} - Y_i]^2 \tag{5}$$

Here, $n(\tau)$ is the number of pairs of measuring values obtained at a distance τ from each other. Sufficiently accurate estimations of the variogram function must be based on a large number of data pairs. At least 50 data pairs are necessary as a general rule.[40] For example, in a pattern of n_0 points arranged as k equilateral triangles, whose sides are $\tau = \tau_{\min}$ long ($\tau_{\min} = 10\,\text{m}$ for the surface sampling and $\tau_{\min} = 20\,\text{m}$ for the percussion coring) exactly $n(\tau_{\min}) = n_0 + k - 1$ pairs of points are available for the variogram estimation at the minimum distance $\tau = \tau_{\min}$.

The empirical variogram function was fitted to an appropriate theoretical model to evaluate the spatial variability. The appearance of a variogram is characterized by the following structural parameters of the model:

V_{total} = the total variance, calculated as the mean squared deviation of all the data from the overall mean; C_0 = the nugget effect describing the maximum variance of immediately adjacent measuring quantities (it is caused by the error in sampling and by the spatial variability in the microscale region of the site to be investigated); a = the range (correlation length, radius of influence), i.e. the minimum spatial distance between uncorrelated measuring quantities; S_Y^2 = the sill, i.e. the maximum variance of uncorrelated measuring quantities; τ = the lag, i.e. the spatial distance between subsequent values of the variogram function.

These parameters occur in different models.[41] With chemical data from contaminated soils and waste sites, the spherical model (cubic model, Matheron's model[42]) is frequently used successfully (Equation 6).

$$\gamma(\tau) = 0.5V(\tau) = \begin{cases} 0 & \text{for } \tau = 0 \\ C\left(\frac{3}{2}\frac{|\tau|}{a} - \frac{1}{2}\left(\frac{|\tau|}{a}\right)^3\right) + C_0 & \text{for } 0 < \tau \le a \\ C + C_0 & \text{for } \tau > a \end{cases} \tag{6}$$

In addition to the parameters explained above, $S_Y^2 = C + C_0$ is the sill and C is the partial sill. In Equation (6), $\gamma(0) = 0$ is explained for continuous dependence of the measuring quantity on the lag τ. Mathematically, it follows from Equations (3) and (4). If $\gamma(\tau) \to C_0 > 0$, as $\tau \to 0$, then C_0 is the *nugget effect* (named by Matheron).[42] This is because it is believed that a variation at very small distances (microscale variation) caused by small nuggets generates a discontinuity at the origin of the data. But the only possible reason for such a phenomenon is an error of observation in both sampling and chemical analysis. Hence, the nugget effect is unverifiable without samples that are very close together. Following this idea, eight surface samples were taken from each of ten points of the slagheap. The mean squared deviation of the replicate values at a sampling point from the respective mean is an estimation of the variance of the sampling error. The error in chemical analysis is mostly assumed to be negligible without investigating the measuring process. In geostatistics, the nugget effect is determined by extrapolating variogram estimates from lags closest to zero. Its calculation is implemented in common geostatistical software following the recommendations of Cressie.[41] A more rigorous presentation of the nugget effect and its partitioning is available.[41]

An appropriate approximation of the sill $S_Y^2 = C + C_0$ is given by the sample variance in Equation (7),

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 \tag{7}$$

which is calculated on the basis of the total amount of n data of the series of measuring values, where $\overline{Y}_n$ is the overall mean.

Besides the spherical model, various parametric variogram models are applied in environmental science and pollution research. The following two basic isotropic models[43] are considered in terms of the semivariogram:

linear and exponential. They also include a possible nugget effect, as in the spherical model. A linear or an exponential variogram is recommended if there is a trend in the spatial behavior of the measuring quantity and if there are local anomalies in the chemical composition, or if there are polymodally or asymmetrically distributed measuring values.

The linear model is given by Equation (8)

$$\gamma(\tau) = \left\{ \begin{array}{ll} 0 & \text{for } \tau = 0 \\ C_0 + b_l|\tau| & \text{for } \tau > 0 \end{array} \right\} \tag{8}$$

where $b_l \geq 0$ is the slope of the function which holds when $C_0 \geq 0$.

The exponential model is given by Equation (9)

$$\gamma(\tau) = \left\{ \begin{array}{ll} 0 & \text{for } \tau = 0 \\ C_0 + C_e \left(1 - \exp\left(-\dfrac{|\tau|}{a_e}\right)\right) & \text{for } \tau > 0 \end{array} \right\} \tag{9}$$

with parameters $C_0 \geq 0$, $C_e \geq 0$, and $a_e \geq 0$.

The isotropic variogram models described above (spherical, linear and exponential) have been successfully used to obtain an informative cartography of spatial heterogeneities of chemical pollutants in agricultural soils near industrial sites.[44–46] The influence of the sampling distance on the estimation of the nugget effect has been investigated.[47] The graph of the variogram function, the *variogram*, is used for practical conclusions. It gives more detailed information regarding the range of correlations and the magnitude of the local and global variability in the directions of investigation. For an anisotropic object of investigation, there are different variograms in different directions. Typical effects of anisotropy on the ranges found in the Saxon slagheap example have been presented.[18,19]

3.1.4.3 Point Kriging Geostatistics uses the technique of point kriging to calculate additional data at points inaccessible for sampling. The main principle of point kriging is to estimate the value $Y(t_0) = Y_{t_0}$ of the regionalized variable at an uninvestigated point t_0 as the weighted mean $\overline{Y}(t_0)$ of n measuring values Y_{t_i} observed at points t_i, where $i = 1, 2, \ldots, n$ (Equation 10)

$$\overline{Y}(t_0) = \sum_{i=1}^{n} w_i Y_{t_i} \tag{10}$$

The weights w_i are determined by side conditions which state that the expected value of $\overline{Y}(t_0)$ equals the unknown $Y(t_0)$, and that the variance of prediction of $Y(t_0)$ by $\overline{Y}(t_0)$ is as small as possible. The estimate $\overline{Y}(t_0)$ of $Y(t_0)$ is unbiased, as shown in Equation (11), because

$$E\left[\overline{Y}(t_0) - Y(t_0)\right] = 0 \tag{11}$$

Equation (12) then holds because of Equation (10),

$$E\overline{Y}(t_0) = E\left[\sum_{i=1}^{n} w_i Y_{t_i}\right] = \sum_{i=1}^{n} w_i E Y_{t_i} \tag{12}$$

and Equation (13) follows because of Equation (1).

$$E\overline{Y}(t_0) = \mu \sum_{i=1}^{n} w_i \tag{13}$$

Because $E\overline{Y}(t_0) = \mu$ holds, the weights have to fulfill the side condition shown in Equation (14).

$$\sum_{i} w_i = 1 \tag{14}$$

The variance associated with predicting unknown values using the weighted mean of known values is given by Equation (15).

$$\begin{aligned} \sigma^2_{\text{prediction}} &= E\left[\overline{Y}(t_0) - Y(t_0)\right]^2 = D^2\overline{Y}(t_0) \\ &= E\left[\sum_{i} w_i Y_{t_i} - Y(t_0)\right]^2 \end{aligned} \tag{15}$$

This variance can also be written in the form shown in Equation (16).

$$\begin{aligned} \sigma^2_{\text{prediction}} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \operatorname{cov}\left(Y_{t_i}, Y_{t_j}\right) - 2\sum_{i=1}^{n} w_i \\ \times \operatorname{cov}\left(Y_{t_i}, Y(t_0)\right) + E\left[Y(t_0)\right]^2 \end{aligned} \tag{16}$$

Equation (17) follows from Equation (16), because of the requirement of homogeneity (Equation 2),

$$\begin{aligned} \sigma^2_{\text{prediction}} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j C_Y\left(|t_i - t_j|\right) - 2\sum_{i=1}^{n} w_i C_Y \\ \times \left(|t_i - t_0|\right) + C_Y(0) \end{aligned} \tag{17}$$

(where C_Y is the covariance function).

The requirement to minimize the prediction variance under the side conditions in Equation (14) provides a system of $n + 1$ linear equations in the weights w_i desired (Equation 18).

$$\begin{array}{cccccc} C_Y(|t_1 - t_1|) & C_Y(|t_1 - t_2|) & \ldots & C_Y(|t_1 - t_n|) & 1 & \\ C_Y(|t_2 - t_1|) & C_Y(|t_2 - t_2|) & \ldots & C_Y(|t_2 - t_n|) & 1 & \\ \vdots & \vdots & & \vdots & 1 & \\ C_Y(|t_n - t_1|) & C_Y(|t_n - t_2|) & \ldots & C_Y(|t_n - t_n|) & 1 & \\ 1 & 1 & \ldots & 1 & 0 & \end{array} \times \begin{array}{c} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{array} = \begin{array}{c} C_Y(|t_1 - t_0|) \\ C_Y(|t_2 - t_0|) \\ \vdots \\ C_Y(|t_n - t_0|) \\ 1 \end{array} \tag{18}$$

This equation system has a unique solution if the determinant of the system matrix on the left of Equation (18) is nonzero. The complementary quantity λ is the Lagrange parameter, associating the weights w_i, (where $i = 1, 2, \ldots, n$), with the side condition that they total 1. The prediction variance has the value given by Equation (19).

$$\hat{\sigma}^2_{\text{prediction}} = C_Y(|t_0 - t_0|) - \sum_{i=1}^{n} w_i C_Y(|t_i - t_0|) - \lambda \quad (19)$$

Because of Equation (4), the covariance function C_Y is expressed by either the semivariogram function γ or the variogram function V. Therefore, the kriging technique which generates additional data in inaccessible points serves a double purpose: (1) to provide the data necessary to model the spatial variability of the measurable quantities in the object investigated; and (2) to formalize the estimated variogram function.

The step-by-step realization of the strategy ''estimation of the variogram function, calculation of new values by point kriging'' leads to a theoretically sufficient number of data, but only the empirical variogram allows us to conclude whether or not the results of the kriging are useful. In the case of an abnormal variogram function contradicting the theoretical assumptions of a homogeneous random field, the geostatistical theory is insufficient. The variogram function is abnormal if its values decrease as the distance increases after an increasing period or if there is generally no recognizable structure. An interpretation of the calculated results is irresponsible in such cases.

Frequently, no distinction is made between the terms variogram and semivariogram because the connection between them is clearly defined by Equation (4).

For a detailed derivation and discussion of kriging techniques, consult Journel and Huijbregts[43] and Cressie.[48] A concise presentation of geostatistics in general and kriging in particular is given by Journel,[49] and its application to site investigations is covered by Rocklin and Jauzein.[50]

3.1.4.4 Results

Surface Sampling. The iron, cadmium, lead, zinc and sulfur levels obtained in eight different directions of the slagheap were evaluated. The numerical calculations were carried out by using the computer program Surfer.[51]

The variograms of the copper content and the iron content vary considerably in different directions. Useful results were calculated in the west–east extension. Because of the finely-dispersed spatial distribution of copper within the slag there was no nugget effect, i.e. no microscale variation caused a discontinuity of the copper content at a sampling point. The sills and the ranges are of the same order of magnitude. This shows the isotropic behavior of the copper content in this extension.

The variogram and the three-dimensional graph of the iron content, calculated on the basis of the original data in the direction mentioned and the three-dimensional graph after kriging, are shown in Figure 6, as an example. A partial sill of $C = 17.0\%^2$ and a range of $a = 30\,\text{m}$ were obtained. The fitted spherical model has a nugget effect of about $C_0 = 3.0\%^2$. The figures written near the function points of the variogram represent the respective numbers of pairs of measuring values used to calculate the corresponding function value at the respective distance (Figure 6a). The three-dimensional graph of the data visualizes the spatial distribution of the iron content on the heap surface (Figure 6b). The local anomalies characterize extreme iron contents and they clearly stand out against the globally inherent iron distribution on the heap surface. The anomalies reflect the original data obtained at the sampling points, but the overall spatial variability is easily recognizable.

In contrast, useful empirical variograms were obtained for both the sulfur content and the zinc content in all directions by a stepwise change of the directions investigated. In principle, the three-dimensional graphs of both the sulfur and zinc contents are similar to that of the iron content, in some directions. This may be caused by the mineralogical composition of the thermally processed ore concentrates, the residue of which is the rotary furnace slag.

Furthermore, the following effect was observed in the spatial variability of the sulfur and the zinc contents. All the variograms have one and the same sill but different ranges depending on the direction searched. The set of ranges forms the boundary points of an ellipse in the given coordinate system. This effect is called a *geometric anisotropy*. The largest range was observed in the east-north-east direction, with the ratio of the largest range to the smallest one amounting to 2.6. The direction of the largest range and the ratio were used to transform the coordinates of the investigated area so that the transformed area was considered isotropic and consequently fulfilled the assumption of the kriging technique. Convincing three-dimensional graphs of the original data and the kriging values were obtained.

As for the lead content, semivariograms with no valuable structure or at least an untypical appearance were calculated. It was proved that the empirical variograms are fitted by a set of spherical model functions whose parameters may vary over wide intervals without any significant change in the goodness of fit. For a detailed discussion of this fact, see Rasemann and Herbst.[19]

The arsenic and cadmium contents also proved to be untypical of the chemical composition of the slagheap. The semivariograms calculated were either difficult to

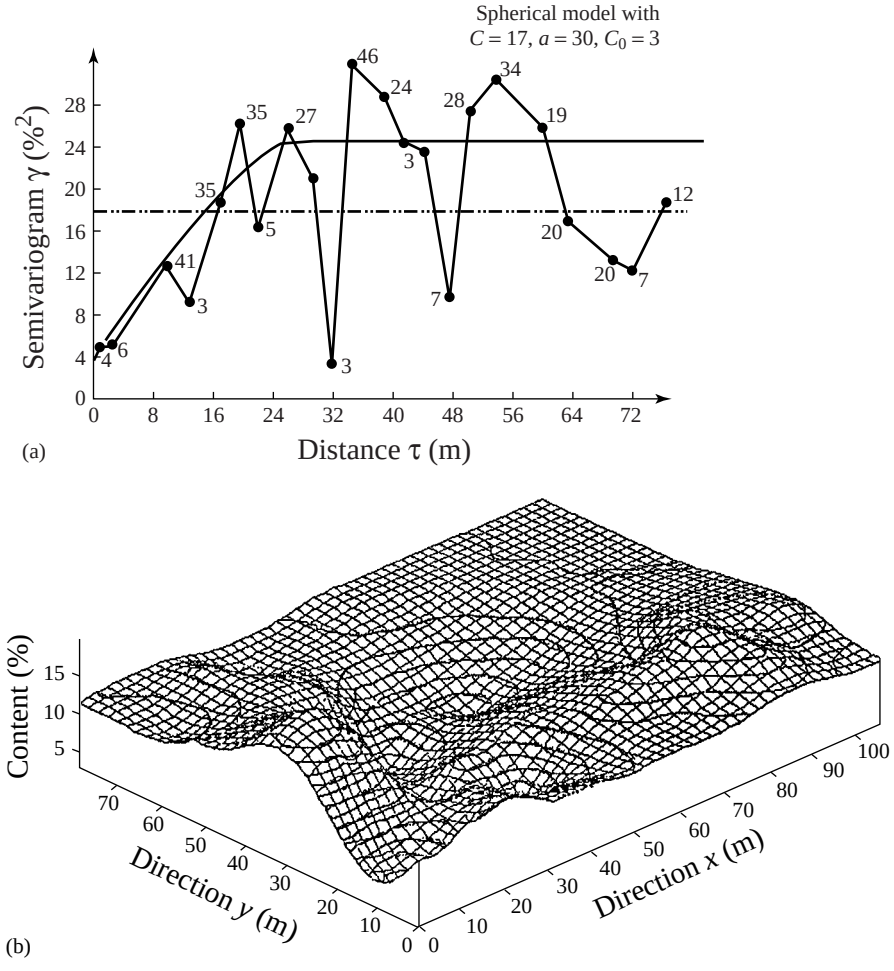

Figure 6 Geostatistical evaluation of the iron content using the original data from surface sampling in the west–east extension of the slagheap. (a) Empirical semivariogram and corresponding spherical model fitted with a nugget variance of 3.0%^{2}, a sill of 20.0%^{2}, and a range of 30 m, respectively; (b) three-dimensional graph obtained after point kriging.

evaluate or were useless. There was neither a preferred direction of variability nor any geostatistical similarity between different directions. The parameters of the fitted semivariograms varied considerably from distance to distance and from direction to direction without a recognizable systematic structure. The anomalous variograms of the arsenic and lead contents can be explained by local nested concentrations of these pollutants on the heap surface. Anomalies formed by environmental influences occurred after dumping of the slags. Thus, because of the high concentration of unburnt carbon, the slag was carbonized at a low temperature and volatile components such as arsenic and lead sublimated on the surface and formed corresponding chemical compounds. The frequency distributions are also polymodal and asymmetric. The lead content distribution of a normal slag is shown in Figure 7, as an example.

The ranges of the semivariograms were mostly larger than the minimum sampling distance of 10 m, i.e. more

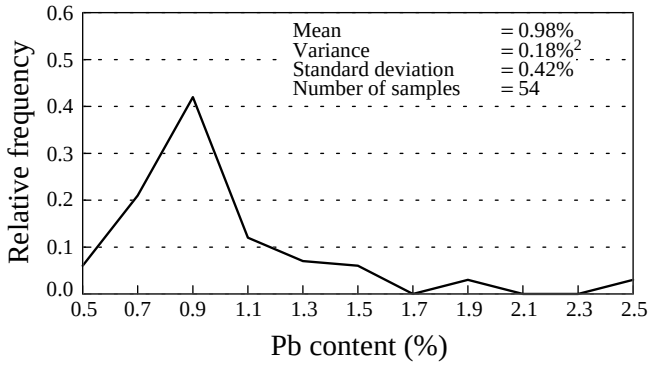

Figure 7 Frequency distribution of the lead content of a normal slag obtained from all surface samples.

samples than needed for a simple random sampling analysis using correlated data were taken. The available data generally gave grounds for a geostatistical analysis, but the number of values was insufficiently small in some

cases, and the errors expected were then too large for a reliable interpretation.

Percussion Coring. One sample was taken at every meter depth from percussion coring wells at 18 points of a regular pattern of equilateral triangles whose sides were 20 m long. The chemical components of samples from the same depth interval were geostatistically evaluated. The iron, copper, sulfur, cadmium, arsenic, and lead contents and the ignition loss were investigated. In general, both variograms and kriging structures obtained were similar to those for the surface sampling. In the following, some results from the 1–2 m depth interval are discussed.

The variograms and three-dimensional graphs obtained after kriging were especially valuable for the iron, sulfur and zinc contents in the west–east extension. In this orientation, structures of the spatial variability were traced even from the heap surface. But there are also directions in which the variograms are completely useless. To demonstrate this, the results of geostatistical evaluations of the zinc content obtained in the northerly direction are discussed. In Figure 8(a), an empirical semivariogram is presented. It was calculated only on the basis of the original data obtained from percussion core samples taken from 18 pattern points at a depth of 1–2 m. Between 2 and 17 pairs of measuring values were available to construct the points of the empirical

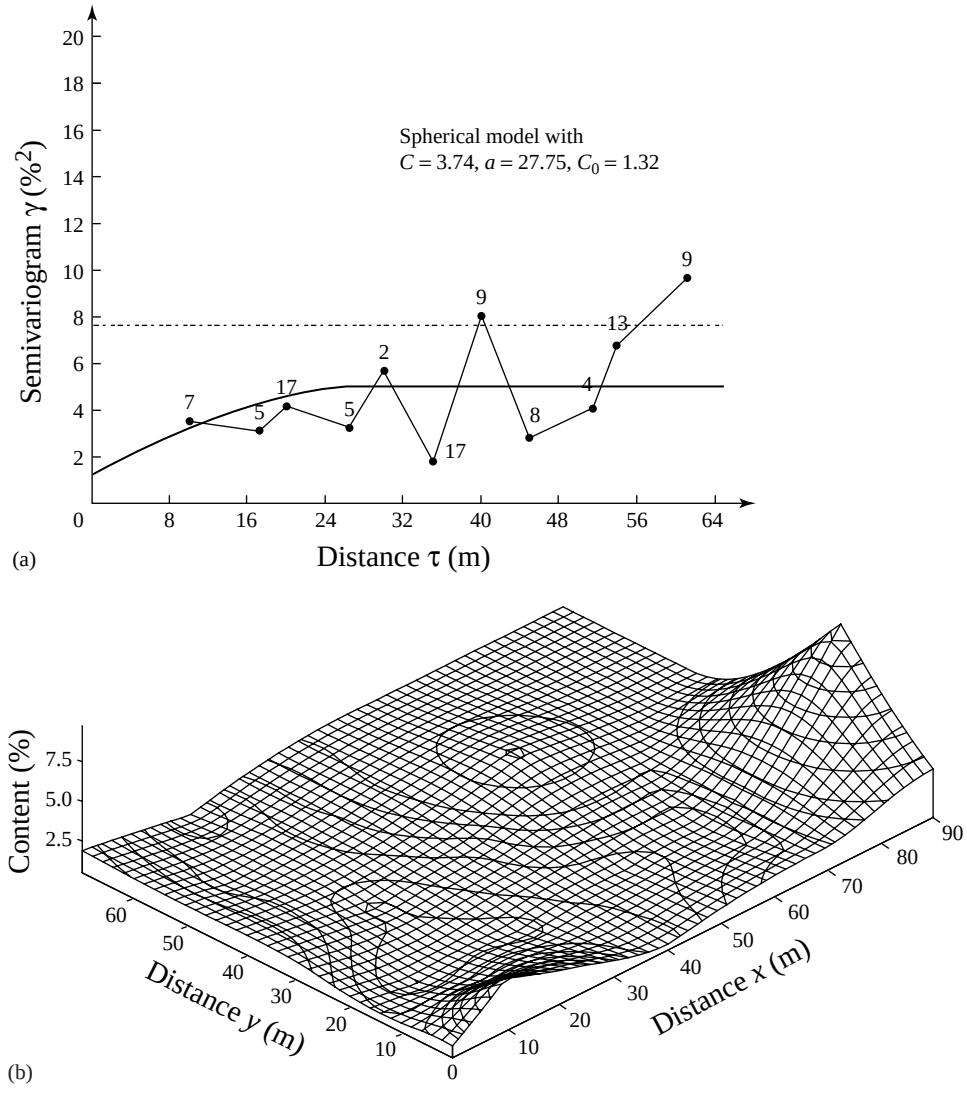

Figure 8 Geostatistical evaluation of the zinc content using the original data from percussion coring in the north extension of the slagheap. (a) Empirical semivariogram and corresponding spherical model fitted with the nugget variance $C_0 = 1.32\%^2$, the partial sill $C = 3.74\%^2$, and the range $a = 27.75\,\text{m}\%^2$; (b) three-dimensional graph on the basis of the original data and the calculated kriging values.

variogram. Because of the relatively small number of original data points, the error in the calculated results was correspondingly large. A spherical model with a nugget effect $C_0 = 1.32\%^2$, a partial sill $C = 3.74\%^2$, and a range $a = 27.75$ m was fitted. Using this model, point kriging was carried out. Figure 8(b) represents the corresponding three-dimensional plot. Obviously, this plot reflects only the original data and those in the neighbourhood of the sampling points, in principle. However, the data from this neighbourhood were inaccurate, because the distances between the original points were so large that the kriging procedure only provided an interpolation between the original data and the overall mean. This conjecture has been examined previously for a similar situation involving the evaluation of the iron content.[14,19] Therefore, for a mainly heterogeneous slagheap with wide distances between adjacent sampling points, the technique of point kriging must not be used to model the spatial variability of the chemical composition.

The spatial variability of the sulfur content was accurately represented by geostatistical methods. Structures obtained were similar to those of the surface sampling. No systematic dependence of the original data on their spatial allocation was observed for the cadmium content. Consequently, empirical (semi)variograms were obtained which could not be fitted by a single-value theoretical model function.

The use of percussion coring to evaluate the spatial distribution of the lead content provided no additional knowledge as compared with surface sampling. Empirical semivariograms calculated on the basis of 2–16 pairs of original data sets were found which are better fitted by an exponential function than by a spherical model (as an example see Figure 9a). Because of the finely-dispersed

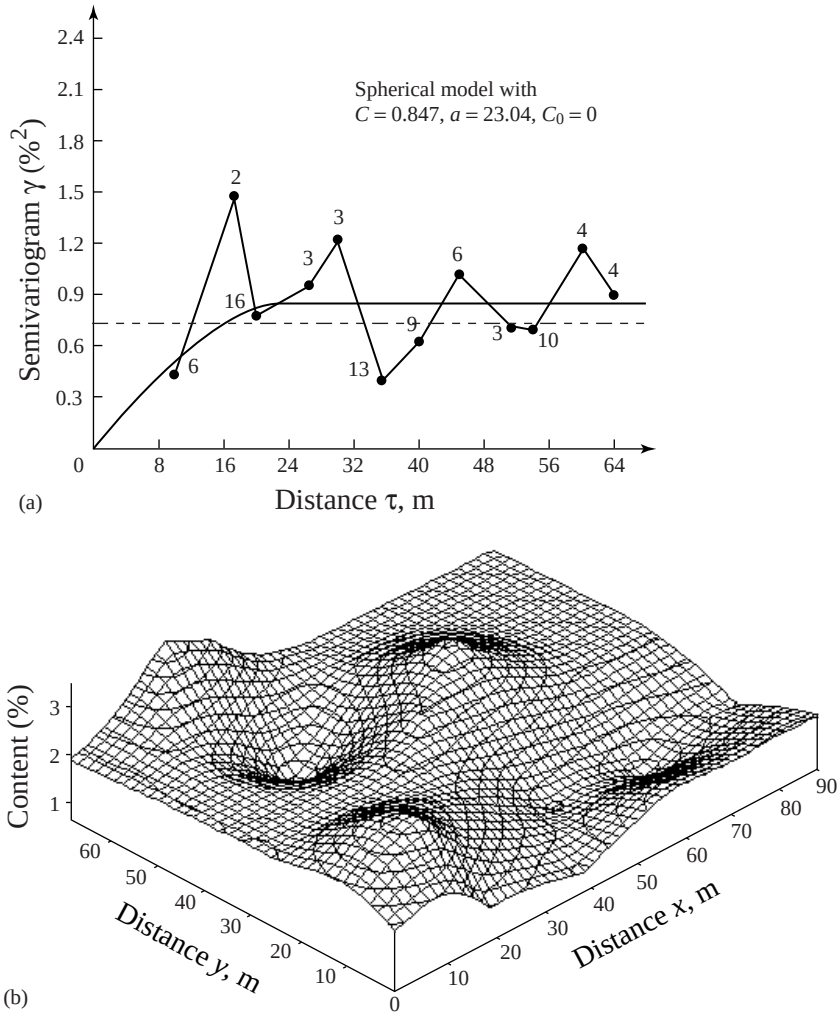

Figure 9 Geostatistical evaluation of the lead content obtained by percussion coring in the depth interval of 1–2 m in the west–east extension of the slagheap. (a) Empirical semivariogram and fitted spherical model with the nugget effect $C_0 = 0\%^2$, the sill $C = 0.847\%^2$, and the range $a = 23.04$ m; (b) three-dimensional graph of the original data and the values calculated by kriging on the basis of the fitted model in (a).

distribution of lead within the slag, a nugget variance was not observed. Theoretically the kriging procedure with a fitted spherical function provided additional data, on the basis of which a three-dimensional picture of the spatial variability was obtained (Figure 9b). But because of the small number of original data, the large distances between the neighboring sampling locations and the unavoidable errors of variograms and kriging, the validity and usefulness of the geostatistical conclusions were questionable.

Random Sampling Analysis. It was shown that the spatial variability of a chemical component of the slagheap cannot be successfully evaluated by geostatistical methods if this component is heterogeneously distributed over the heap. To avoid this, and to guarantee that the geostatistical assumptions (i.e. a homogeneous isotropic random field), are more likely to be fulfilled, the entire slagheap can be partitioned into several relatively small subregions. However, this procedure requires a more close-meshed sampling pattern to provide the data necessary for a reliable geostatistical evaluation.

If a detailed characterization of the slagheap is too expensive compared with the benefit expected, and the localization of highly contaminated subregions is impossible or unnecessary, the spatial location of any datum point obtained is not of interest. In such cases, random sampling analysis is recommended. Using random sampling analysis, the original data are assumed to be randomly located and stochastically independent. Independent variables are also uncorrelated. Then, the slagheap is globally evaluated by estimating the overall true content (expectation value) and the variance of the original data. The true content is estimated by the overall mean of all data available. The absolute difference between the mean and the (unknown) true value defines the absolute error (accuracy) of estimation. If an upper bound of this error and a lower bound of its confidence probability are prescribed, the minimum number of randomly located samples necessary to fulfill these limitations can be calculated.

For example, let μ be the true content of a component. Furthermore, let $\overline{Y}_n$ be the mean and S_n^2 be the empirical variance obtained by n samples, respectively. $\overline{Y}_n$ and S_n^2 are random variables. Here and in the following, the random variables are represented as capital letters, and the corresponding actual values are written as lower-case letters. Hence s_n^2 is the actual value of the empirical variance S_n^2, and $\bar{y}_n$ is the actual value of the mean $\overline{Y}_n$. The true content μ is not known, as a rule, but it can be replaced by the mean $\overline{Y}_n$. In doing so, an error is made. The difference between the mean and the true value is called the *absolute error of estimation* or the *accuracy of prediction*, and this error is denoted as $|\overline{Y}_n - \mu|$. This error is a random variable with an expected value of zero and a standard deviation $S_n/\sqrt{n}$. As the standard deviation S_n is the mean error of any individual observation, the quantity $S_n/\sqrt{n}$ is the mean error of the mean $\overline{Y}_n$. The mean error of the mean is abbreviated as $S_{\overline{Y}_n}$. A random interval which covers μ at the confidence probability (confidence level) β is defined by Equation (20).

$$\overline{Y}_n - \frac{S_n}{\sqrt{n}} t_{n-1;(1+\beta)/2} < \mu < \overline{Y}_n + \frac{S_n}{\sqrt{n}} t_{n-1;(1+\beta)/2} \quad (20)$$

Here, $t_{n-1;(1+\beta)/2}$ is the quantile of the order $(1+\beta)/2$ of the Student's t-distribution at the degree of freedom $n-1$; this value can be found in tables of mathematical statistics.[4] Furthermore, by rearranging Equation (20), an upper bound of the absolute error of estimation $|\overline{Y}_n - \mu|$ at the confidence probability β is obtained (Equation 21).

$$\varepsilon_n = \frac{S_n}{\sqrt{n}} t_{n-1;(1+\beta)/2} \quad (21)$$

The quantity ε_n is called the *probable error of estimation* of μ by $\overline{Y}_n$ at the confidence probability β. The confidence probability β is taken to be 90–95%, as mentioned above. In practice, an upper bound of the probable error of estimation, say ε_0, at a certain probability β is prescribed. That is, a quantity ε_0 is required which fulfills the condition $|\overline{Y}_n - \mu| \leq \varepsilon_0$ at the probability β. This bound is called the *maximally admissible error of estimation*. Substituting ε_0 for ε_n and n_0 for n in Equation (21), the inequality shown in Equation (22) is obtained.

$$\varepsilon_0 \leq \frac{S_{n_0}}{\sqrt{n_0}} t_{n_0-1;(1+\beta)/2} \quad (22)$$

If an estimated value s_n of S_{n_0} is known, and if for $t_{n_0-1;(1+\beta)/2}$ an approximate value of 2.00 at confidence probability 95% is taken, the minimum number of samples n_0 necessary to fulfill the condition in Equation (22) is obtained according to Equation (23).

$$n_0 \geq 4\frac{s_n^2}{\varepsilon_0^2} \quad (23)$$

For the special case where the admissible error of the mean must not be greater than the mean error of any individual observation ($\varepsilon_0 = s_n$), at least $n = 4$ samples are needed. To be sure that this requirement is fulfilled even with one outlier, five samples are taken in practice. If the stronger condition $\varepsilon_0 = s_n/2$ is required, $n = 16$ samples are needed.

The procedure explained above was applied to globally evaluate the lead content of the slagheap. A number $n = 54$ of surface samples were considered. The mean value $\bar{y}_n = 0.98\%$, the empirical variance $s_n^2 = 0.18\%^2$ (see Equation 7) and the standard deviation $s_n = 0.42\%$

were calculated. The two-sided quantile of the Student's t-distribution with a degree of freedom $n-1=53$ at a confidence level of 95% is $t_{53;0.975}=2.0057$. A quantile of a distribution function at the confidence level β is called two-sided or two-tailed if the relation $t_{(1-\beta)/2}=t_{(1+\beta)/2}$ holds. This is fulfilled when the distribution is symmetrical to the expectation value zero. By putting the t-value, the actual mean $\bar{y}_n=0.98\%$ and the empirical variance $s_n^2=0.18\%^2$ into Equation (20), the interval (0.86%, 1.10%) was calculated. This interval covered the true lead content at a probability of 95%. This estimate corresponds to an accuracy of prediction of $\varepsilon_n=0.12\%$. When $\varepsilon_0=0.10\%, n_0\geq 4\times(0.18/0.01)=72$ samples are needed. To guarantee this accuracy, the 56 surface samples and one sample from each of the 18 percussion coring wells would have been sufficient if the standard deviation s_n had not been changed considerably.

If the situation allows gradual sampling, chemical analysis and statistical evaluation of the data the number of samples needed for a prescribed precision of estimating ε_0 at a given confidence probability β can be sequentially determined in the following way. Let n be the minimum number of samples available at the beginning of the investigation, say $n=5$ as an example. Then, the mean $\bar{y}_n$, the error of the mean $s_{\bar{y}_n}$ and the probable error of estimation ε_n are calculated. If $\varepsilon_n<\varepsilon_0$ holds for a prescribed ε_0 at a given confidence probability β, the investigation is finished. If it is not, additional samples are taken, the statistical parameters $\bar{y}_n$ and $s_{\bar{y}_n}$ are calculated and the actual probable error of estimation, ε_n, is compared with the given maximum admissible error of estimation ε_0. The investigation is continued until the condition $\varepsilon_n<\varepsilon_0$ with the actual number of samples n is fulfilled.

Figure 10 shows a graphical presentation of the mean lead content $\bar{y}_n$ and the lower and upper confidence limits $\bar{y}_n\pm\varepsilon_n$ for the true content μ at a confidence probability

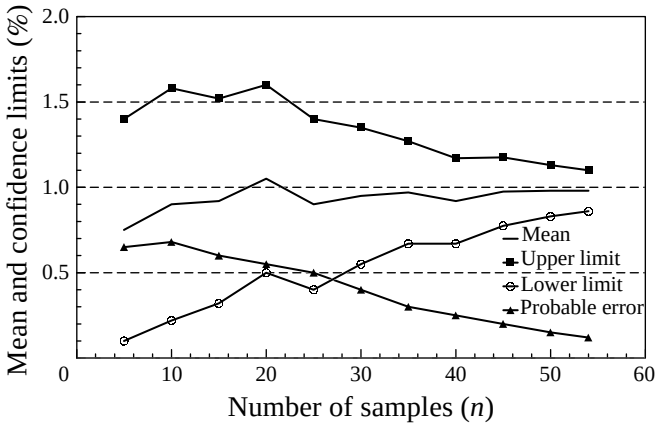

Figure 10 Graphical presentation of the sequential estimation of the expected lead content assuming uncorrelated observations.

95% against the number of samples n. The influence of possible outliers and extreme values on the estimation process obviously decreases as the number of samples increases, as expected.

Finally, let us compare the benefit of a random sampling method with a systematic method. As mentioned above, the error of the mean $s_{\bar{y}_n}$ measures the efficiency of a sampling method. In contrast to random sampling analysis, the systematic method takes into account the allocation of the data in space. For a data series consisting of n subsequently allocated data points, the squared error of the mean is calculated according to Equation (24).

$$S_{\bar{Y}_n}^2=\frac{S_Y^2}{n}\left[1+\frac{2}{n}\sum_{\tau=1}^{n-1}(n-\tau)K(\tau)\right] \tag{24}$$

In this equation, besides the quantities explained above, $K(\tau)$ is the value of the correlation function K describing the degree of correlation (correlation coefficient) of each of two measuring variables observed at a spatial distance of τ from each other.

Obviously, the error of the mean increases as the correlation coefficient between the variables at different sampling locations increases. The correlation function K is immediately related to the semivariogram function γ or the variogram function V. Because of Equation (4), and $S_Y^2=C_Y(0)$, the value of the semivariogram function at the lag τ is expressed as shown in Equation (25).

$$\gamma(\tau)=S_Y^2\left[1-\frac{C_Y(\tau)}{S_Y^2}\right] \tag{25}$$

Here, $C_Y(\tau)/S_Y^2$ is the value of the correlation function K mentioned, and this quantity is abbreviated as $K(\tau)$. Equation (26) thus results from Equation (25) as follows.

$$K(\tau)=1-\frac{\gamma(\tau)}{S_Y^2} \tag{26}$$

Substitution of Equation (26) for $K(\tau)$ in Equation (24) gives Equation (27).

$$S_{\bar{Y}_n}^2=\frac{S_Y^2}{n}\left[1+\frac{2}{n}\sum_{\tau=1}^{n-1}(n-\tau)\left[1-\frac{\gamma(\tau)}{S_Y^2}\right]\right] \tag{27}$$

This equation emphasizes the connections between random sampling analysis and geostatistical analysis. In practice, the mostly unknown variance S_Y^2 in Equations (24) and (27) is replaced by the respective estimate S_n^2 obtained according to Equation (7). In the special case of no correlations at any lag ($K(\tau)\equiv 0$ for any τ), the semivariogram function γ already has its maximum, actually S_Y^2, at the minimal lag $\tau=1$. Substituting S_Y^2 for $\gamma(\tau)$ at any lag τ in Equation (27), and considering that $\sum_{\tau=1}^{n-1}(n-\tau)=n(n-1)/2$, the expression

$S^2_{\overline{Y}_n} = S^2_Y/n$ describing the squared error of the mean of n randomly chosen variables results from Equation (27). This error is lower than any other error calculated by Equation (27) on the basis of the same number of correlated data. Thus, time-consuming and costly geostatistical planning and sampling experiments to investigate extremely heterogeneous heaps and wastes is justified only for compelling reasons, for example if there is a particularly great interest in evaluating the spatial variability of certain regions. Otherwise, random sampling analysis is preferred. In the case of the Saxon slagheap, a sampling distance in the same order of magnitude as the range (correlation length) should be chosen.

3.1.5 Summary and Conclusions

In some cases sampling according to a systematic pattern followed by geostatistical evaluation can be successfully used to obtain a realistic picture of the spatial variability of the slagheap. Empirical (semi)variogram functions were calculated using original data in definite directions of the field of investigation and in fixed levels of the heap. The empirical variogram functions were fitted by a spherical model. Nugget effect, sill and range were easily interpretable parameters of this model. The technique of point kriging was applied to calculate data for additional points on the basis of the fitted variogram functions. Using the original data and the calculated values, three-dimensional presentations show the spatial variability of essential chemical components such as iron, zinc and lead. Useful results were obtained for the sulfur content in all directions and all levels. The geostatistical evaluation of the iron content only provided valuable variograms and kriging data in some directions of the heap. However, the iron content varied so widely in other directions that geostatistical methods did not work there either. The cadmium and the arsenic contents were untypical of the composition of the slagheap. The semivariograms calculated were either difficult or impossible to evaluate and showed neither preferred directions of variability nor any geostatistical regularity.

To sum up, it can be concluded that the highly expensive sampling, chemical analysis and geostatistical evaluation of extremely heterogeneous heaps of slags and industrial wastes are justified only by compelling reasons. However, reasonable results are obtained even with the wrong assumptions, because many geostatistical procedures are fairly robust. On the other hand, sampling regulations based on relationships between grain size and sample mass fail for contaminated sites and wastes where the structure has been markedly changed by physical and chemical processes, such as leaching, carbonization at low temperatures, precipitation, etc. If a global evaluation of the heap by mean values and variances of certain chemical components has priority, however, the usual methods of sampling and statistical data analysis recommended and developed for bulk materials and soils are promising tools.

The methodical results obtained from evaluating extremely heterogeneous contaminated sites and wastes in the Saxon slagheap example clearly show the possibilities and limitations of known sampling methods and geostatistical models. On the one hand, probably the greatest disadvantage in applying geostatistics is the difficulty in knowing whether the assumptions, for example a homogeneous isotropic random field, are acceptable or not in an individual case. On the other hand, such a close-meshed sampling and percussion coring pattern as applied in the case of the Saxon slag heap cannot be realized under commercial conditions. In practice, a compromise must be made. A promising strategy is to lay a regular grid over the region so that the coordinate origin and axis orientation are random. Then, samples are taken either systematically at each grid point or randomly at each subsection of the pattern. As shown in the example of the slagheap, one advantage of systematic sampling using a randomly placed grid is that the data obtained are very suitable for analysis by random sampling and geostatistical methods.

3.2 Evaluation of a Mercury-contaminated Site – A Problem of Sampling and Sample Pretreatment

3.2.1 Introduction

To make decisions concerning contaminated sites, a reliable estimation of the environmental risk caused by hazardous components is necessary. This estimation can be performed by means of soil analysis, for instance. In this case, errors resulting from sampling, sample preparation and chemical analysis are unavoidable. To restrict these errors and to achieve comparable and legally valid criteria for the pollution levels, certain regulations must be considered.[52] Until recently such regulations have been available only for limited fields of application. The rules commonly used for the evaluation of contaminated sites were originally developed for soils, water, sewage and sludges. The DEV are methods relating to the investigation of water, sewage and sludges[34] and to sewage sludge regulation.[52] A detailed overview of the German, European and International standards used for the determination of physical properties and the composition of inorganic pollutants in soils is given in Tables 5, 6 and 7.[53] To apply these rules in soil quality estimation, it is assumed that:

1. the soil is in an area (potentially) suitable for agricultural or horticultural use;

Table 5 German (DIN) and International (ISO) standards for the determination of physical properties of soils[52]

Property	Standard
Dry mass	DIN ISO 11465
Organic carbon and total carbon by dry combustion	DIN ISO 10694
pH value ($CaCl_2$)	DIN ISO 10390
Grain size distribution	DIN ISO 11277/DIN 18123
Compactness	DIN ISO 11272

Table 6 German (DIN), European (EN) and International (ISO) standards for the determination of inorganic pollutants in soils[52]

Chemical components	Standard
As, Cd, Cr, Cu, Ni, Pb, Zn	DIN EN ISO 11885
Cd, Cr, Cu, Ni, Pb, Zn	DIN ISO 11047
As	DIN EN ISO 11969
Cr(VI)	DIN 38405-24/DIN 19734
Hg	DIN EN 1483/DIN 38414-7

Table 7 German (DIN), European (EN) and International (ISO) standards for the determination of pollutants in elutriates[52]

Pollutant	Standard
As, Cd, Cr, Cu, Mo, Ni, Pb, Sb, Se, Sn, Zn	DIN 38406-22
Sb, As	DIN EN ISO 11969
Pb	DIN 38406-6
Cd	DIN EN ISO 5961
Cr, total	DIN EN 1233
Cr(IV)	DIN EN ISO 10304-3
Co	DIN 38406-24
Cu	DIN 38406-7
Ni	DIN 38406-11
Hg	DIN EN 1483/DIN 38414-7
Se	DIN 39405-23
Zn	DIN 38406-8
CN^-, total or easily released	DIN 38405-13
F^-	DIN 38405-4/DIN EN ISO 10304-1

2. the inorganic contaminants in particular are distributed relatively uniformly over the whole area investigated;
3. the concentrations are low and of the same order of magnitude as the geologically originated background containing mainly nonvolatile compounds.

In contrast, the hazardous components of contaminated sites can occur in high concentrations, be distributed nonuniformly over the area of investigation and be bound in analytically relevant amounts to volatile species. If there is improper handling of the soil samples after sampling, chemical analysis can result in concentrations that are too small. This problem is known and can be controlled by regulations existing for sample preparation and for analysis of volatile compounds such as halogenated hydrocarbons, chlorinated hydrocarbons, hydrocarbons and phenols.

The fact that heavy metals such as mercury can escape from the soil sample is so far not reflected in the regulations for sample preparation. Mercury is known to exist in several forms or species in the environment. For the evaluation of contaminated sites, volatile species are of special interest.[54]

1. All kinds of volatile mercury species such as elemental metallic Hg(0), methylmercury, dimethylmercury and phenylmercury, could be emitted by anthropogenic sources.[53]
2. Bacteria are able to transform Hg^{2+} into volatile methylmercury. This process is described for various environmental matrixes and conditions. Bacteria are also able to reduce Hg^{2+} to volatile Hg(0). In such cases, a high vapour pressure of metallic mercury is responsible for its volatilization as the soil sample drys up.
3. Methylmercury can be transformed into volatile Hg(0), methane and other organic compounds by microbial activity and by abiotic processes.
4. Hg^{2+} can be transformed into Hg(0) by abiotic reduction.
5. In addition, mercury and its species adsorb to different types of materials (e.g. plastic bags for soil samples), and are released in an uncontrollable manner.

The aim of our investigations was to find out, using an example, to what extent the analytically detectable mercury content of soil samples taken from contaminated sites is influenced by the handling of samples between sampling and sample preparation. Handling covers all the means by which the soil sample can be exposed, up to its preparation for chemical analysis. These procedures are termed *sample pretreatment* in order to distinguish them from *sample preparation* by homogenization, splitting and reduction. Thus, methods of storing the sample material under different natural conditions or even active procedures that change the physical or chemical structure of the sample material (such as thermal evaporation at different temperatures) also define sample pretreatment measures. However, the measured results can be influenced not only by sample pretreatment but

also by errors of sampling, sample preparation and chemical analysis. Finally, we must take into account the fact that, in practice, the measured mercury content should reflect the intrinsic properties of the soil sample. These properties depend on the sampling location and are subject to global and local variations. To obtain general statements, it was necessary to plan the investigation program such that the different possible causes for the variation in measuring results could be evaluated independently. Other chemical components besides mercury were not investigated.

3.2.2 Object Investigated, Experimental Design

An industrial mercury-contaminated site was chosen as an example. Three locations of the site were investigated, named as A, B and C. At each point, *primary samples* of about 2 kg were taken from depths of 0–10 cm and 10–25 cm, respectively. Each primary sample was homogenized and split into three *subsamples*. The subsamples were separated again, and each portion was submitted to different pretreatment conditions before chemical analysis. The material obtained is referred to as a *pretreated sample*. The pretreatment methods were:

(1) unchanged original material;
(2) air drying for 2–4 weeks;
(3) storing in a plastic bag for 2–4 weeks; and
(4) thermal evaporation at 40, 60, 80, 105 and 120 °C.

For air drying for 2–4 weeks, about 100 g of the wet soil samples were spread on porcelain plates in a mercury-free atmosphere in normal light and at room temperature (20 °C).

For storing in plastic bags for 2–4 weeks, about 100 g of the wet soil samples were placed into commercial plastic bags and stored in a mercury-poor atmosphere in normal light and at room temperature (20 °C).

For thermal evaporation at 40, 60, 80, 105 and 120 °C about 20 g of the wet soil samples were placed in porcelain pans. The pans were placed on a sand bath at the chosen temperature for 24 h in a mercury-poor atmosphere. The mercury content of the laboratory atmosphere was determined once a week and proved to be below 80 ng m^{-3}.

Two or three *final samples* of 3 g each were selected from each pretreated sample and digested by 10 ml HNO_3/HCl (Merck) for 4 h at 90 °C according to DIN 38414-7 (see Tables 6 and 7). All chemicals were of analytical grade. Water from a Milli-Q purification system was used throughout. A 20% m/V solution of tin(II) chloride in 15% HCl (Merck) was applied in the reduction step. The calibration was carried out using mercury(II) solution (Merck). A Perkin-Elmer (model 430) atomic absorption spectrometer, equipped with a MHS 20 mercury hydride system with amalgam attachment and a model 056 recorder (Spectraphysics) were used for the total mercury determinations.

The measuring results can be arranged in a hierarchical scheme as shown in Figure 11.

The total amount of samples from a sampling location is referred to as the *gross sample*. If the mean values are formed to each branch of a hierarchical level in the scheme shown in Figure 11, for a gross sample, a primary sample within a gross sample, a subsample within a primary sample, etc., the following questions arise.

1. Are the mean values of the gross samples significantly different from each other, i.e. does the sampling location significantly influence the mercury content?
2. Are the mean values of the primary samples of a particular gross sample significantly different from each other, i.e. does the sampling depth influence the results at a sampling location?
3. Are the mean values of the subsamples from one primary sample significantly different from each other, i.e. is sample preparation a statistically important source of error?
4. Given that pretreated samples from one subsample produce significantly different results, does the method of sample pretreatment influence the mercury content detected by chemical analysis?

The first question is easily answered: the means of the three sampling locations are significantly different, as the frequency distribution of the mercury content clearly shows (see Figure 12).

3.2.3 Statistical Analysis of Data

3.2.3.1 Variance Analysis The questions mentioned above are mathematically and statistically treated by a model of variance analysis of four-way hierarchical classification (because of the four classes of hierarchically ordered test conditions) with random effects (because the experimental conditions are chosen at random from a large number of possibilities) and with a balanced experimental design (each experimental condition of a certain hierarchical level is combined with the same number of experimental conditions of other hierarchical levels in every case). The principle of variance analysis is to reduce the comparison of mean values to the comparison of variances. These variances are obtained by splitting the *total variance* of the measured results into individual portions produced independently of each other by differences between sampling locations (global variations), between sampling depths of a sampling location (local heterogeneities), and by errors made in sample preparation, sample handling and chemical analysis. The following mathematical model can be given.

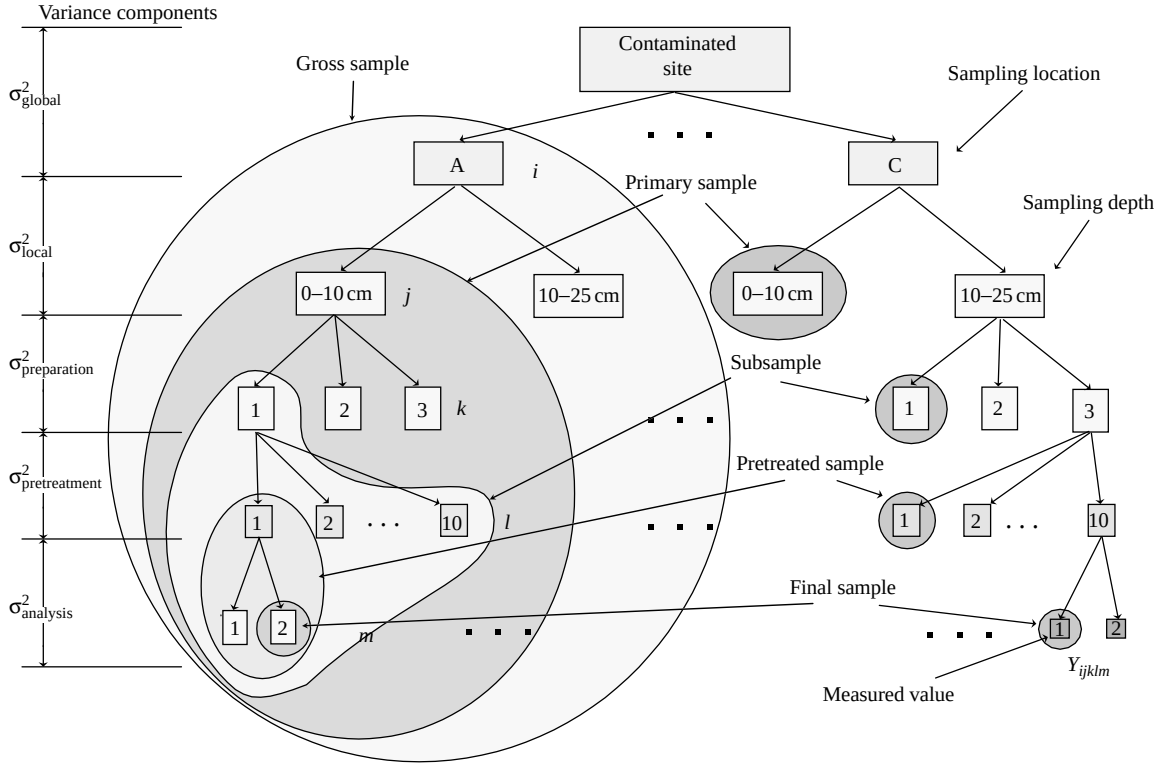

Figure 11 Hierarchical scheme of sampling conditions and measured results.

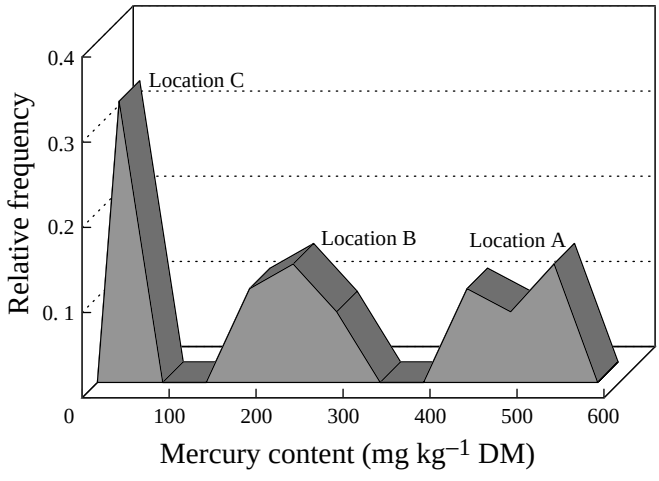

Figure 12 Frequency distribution of the mercury content at the sampling locations.

Two measured values were obtained for each pretreated sample. Let Y_{ijklm} be the mth measured value belonging to the lth pretreated sample of the kth subsample from the jth primary sample at the ith sampling location. Y_{ijklm} is thought to be decomposable as shown in Equation (28)

$$Y_{ijklm} = \mu + \alpha_i + \beta_{ij} + \gamma_{ijk} + \delta_{ijkl} + Z_{ijklm} \qquad (28)$$

with $i = 1, 2$; $j = 1, 2$; $k = 1, 2, 3$; $l = 1, \ldots 10$; $m = 1, 2$. μ is the overall expected value, estimated from the mean of all measuring results; α_i is the random effect caused by global fluctuations of the mercury content within the contaminated site, measured by the difference between the mean obtained from the values at the ith sampling location and the overall expected value; β_{ij} is the random effect caused by local heterogeneities of the contaminated site, described by the deviation between the ith gross sample mean and the respective jth primary sample mean, and is a measure of the random error of sampling; γ_{ijk} is the random effect of sample preparation, which describes the random deviation between the mean of the jth primary sample at the ith sampling location and the mean of the respective kth subsample obtained by homogenization and splitting of the material of the primary sample; δ_{ijkl} is the random effect of sample pretreatment, measured by the deviation between the mean of the lth pretreated sample and the mean of the corresponding kth subsample from which the pretreated material was separated; and Z_{ijklm} is the random error of analysis, which defines the deviation between the measured value Y_{ijklm} and the respective mean of the $ijkl$th pretreated sample.

The following assumptions are made: that the measured values Y_{ijklm} are normally distributed random variables with an expected value μ and a total variance σ^2_{total}, and that the effects α_i, β_{ij}, γ_{ijk}, δ_{ijkl} and the measuring error

Z_{ijklm} are normally distributed, mutually independent random variables with the expected value zero and respective variances σ_α^2, σ_β^2, σ_γ^2, σ_δ^2 and σ_Z^2 for all i, j, k, l and m.[55] Then, the total variance can be expressed as the sum of these individual variances as shown in Equation (29).

$$\sigma_{\text{total}}^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_\delta^2 + \sigma_Z^2 \tag{29}$$

The components in Equation (29) are the *variance components*. The following designations are used: $\sigma_\alpha^2 = \sigma_{\text{global}}^2$, $\sigma_\beta^2 = \sigma_{\text{local}}^2$, $\sigma_\gamma^2 = \sigma_{\text{preparation}}^2$, $\sigma_\delta^2 = \sigma_{\text{pretreatment}}^2$, and $\sigma_Z^2 = \sigma_{\text{analysis}}^2$. The quantity σ_{global}^2 represents the variance caused by global fluctuations of the mercury content over the site; σ_{local}^2 designates the variance by local heterogeneities and corresponds with the sampling error variance, in principle; $\sigma_{\text{preparation}}^2$ indicates particularly the variability between the mercury content of different subsamples obtained by homogenization and splitting of a primary sample; $\sigma_{\text{pretreatment}}^2$ describes the variance of the mercury content caused by the different methods of handling the subsamples before chemical analysis; and finally $\sigma_{\text{analysis}}^2$ is the variance of the error of chemical analysis.

The questions mentioned above regarding the influence of sampling conditions and handling of samples on the mercury content detectable by chemical analysis are reduced to significance tests of the respective variances. The statistical treatment of these questions can be done using analysis of variance (ANOVA).[55] The following results were obtained.[54,56]

1. The global fluctuations between the sampling locations of the contaminated site are highly significant. The variance σ_{global}^2 contributes about 88% to the total variance of the mercury content.
2. The local heterogeneity also significantly influences the mercury content, and its variance σ_{local}^2 amounts to 6% of the total variance. That is, the mean content at a location essentially depends on the depth from which the soil sample is taken. Generally, the mercury content near the soil surface is lower than that from deeper levels. This may be due to microbiotic changes of nonvolatile compounds into volatile species such as elemental metallic mercury, phenylmercury, methylmercury and others, and the leaching of soluble mercury salts.
3. The sample preparation variance $\sigma_{\text{preparation}}^2$ is nonsignificant. That is, the primary samples were split into statistically equivalent subsamples.
4. The sample pretreatment variance $\sigma_{\text{pretreatment}}^2$ is highly significant and of the same order of magnitude as the local heterogeneity variance. The error of sample pretreatment is three to five times greater than the error of the chemical analysis at a confidence probability of 95%. The chemical analysis variance $\sigma_{\text{analysis}}^2$ contributes less than 0.5% to the total variance.

3.2.3.2 Systematic Error of Pretreatment Variance analysis was used to *qualitatively* prove that handling of the soil samples significantly influences the mercury content detectable by chemical analysis. To decide *quantitatively* to what extent the mercury content of the original material is changed by a particular method of sample pretreatment, further statistical comparisons were carried out. More precisely, it was necessary to assess the error made if pretreated samples are used instead of the original material. Additionally, the probability of this error was evaluated. To find out the appropriate test procedure, the statistical properties of the mercury content were checked. Since the mercury content was normally distributed, pairwise comparisons of mean values seems to be reasonable. Because there was no significant error of sample preparation, the subsample variances of the measured values obtained for a fixed pretreatment method were summed up to the respective pretreated sample variance. As shown by Bartlett's test and Fisher's F-test,[4,57] the variances of different pretreated samples were proved to be statistically identical. That enabled us to use the two-sample t-test to compare the mean values obtained carrying out a fixed pretreatment method, with the respective mean of the unchanged original material of a primary sample. The comparisons were based on the following ideas.[56,58]

First, the pretreatment methods were arranged in a certain order to formally designate them by symbols or numbers. Now, let μ_1 be the true mercury content at a definite location and a definite sampling depth of the contaminated site, and μ_i the expected content after the ith method of pretreatment was applied to the original sample material ($i = 2, 3, \ldots$). The difference $|\mu_1 - \mu_i|$ is the systematic error which is made if the original material is not known and only the ith pretreated sample is available. Normally this error is not detectable, but the corresponding mean values $\bar{y}_1$ and $\bar{y}_i$ as well as the empirical variances s_1^2 and s_i^2 obtained by n_1 and n_i independent samples, respectively, are known. The quantities $\bar{y}_1$, $\bar{y}_i$, s_1^2, and s_i^2 are realizations of the respective random variables $\overline{Y}_1$, $\overline{Y}_i$, S_1^2, and S_i^2. Using them, a random interval can be given which covers the systematic error $|\mu_1 - \mu_i|$ at a prescribed confidence probability β (Equation 30).

$$I_\beta = |\overline{Y}_1 - \overline{Y}_i| \pm S\left(|\overline{Y}_1 - \overline{Y}_i|\right) t_{m;(1+\beta)/2} \tag{30}$$

Here, $t_{m;(1+\beta)/2}$ is the corresponding quantile of Student's t-distribution at the degree of freedom $m = n_1 + n_i - 2$,

and (as shown in Equation 31)

$$S\left(|\overline{Y}_1 - \overline{Y}_i|\right) = \sqrt{\frac{(n_1 + n_i)\ (n_1 - 1)S_1^2 + (n_i - 1)S_i^2}{n_1 n_i (n_1 + n_i - 2)}} \tag{31}$$

is the standard deviation (the mean error) of the mean value difference $|\overline{Y}_1 - \overline{Y}_i|$. The length $|I_\beta|$ of the interval I_β is termed the *probable systematic error at the confidence probability* β. Equation (32)

$$\varepsilon_\beta = \frac{|I_\beta|}{\mu_1} \tag{32}$$

is the corresponding *probable relative systematic error*. This criterion seems to be well suited to illustrate the change of the expected mercury content of a soil material after handling by the *i*th pretreatment method. Since the content of the original material μ_1 is mostly unknown, it must be replaced by an appropriate estimation such as the lower limit $\tilde{\mu}_{1,\beta}$ of the confidence interval for μ_1 at the confidence probability β obtained as shown in Equation (33)

$$\tilde{\mu}_{1,\beta} = \overline{Y}_1 - t_{n_1-1;(1+\beta)/2} \sqrt{\frac{S_1^2}{n_1}} \tag{33}$$

where $t_{n_1-1;(1+\beta)/2}$ is the quantile of Student's *t*-distribution at the degree of freedom $n_1 - 1$. The systematic error ε_β in Equation (32) is dimensionless and may be also given as a percentage.

Figure 13 compares the probable relative systematic errors caused by storing the sample in a plastic bag for 2–4 weeks, for samples from locations A and B at a sampling depth of 0–10 cm. A probable relative systematic error up to 40% may occur at a confidence probability of 95%. Storage of the soil samples in plastic bags for 2–4 weeks had the same effect as thermal evaporation at 40 °C. Generally, the absolute loss of mercury in a soil sample depends only on the method of handling and not on the original concentration or the sampling location. Therefore, the lower the original concentration, the greater the relative systematic error. Furthermore, this error increases as the confidence probability increases. Thus, the accuracy in determining the influence of sample preteatment on the mercury content detectable by chemical analysis at a given probability, and under fixed conditions of sampling and analysis, can be improved by decreasing the mean error of the mean value difference $S(\overline{Y}_1 - \overline{Y}_i)$ in Equation (31). This can be done by increasing the degree of homogenization of the samples and by increasing the numbers of samples, n_1 and n_i. Interlaboratory tests and intralaboratory tests must be done to improve the repeatability and the reproducibility of the instrumental measurements.[59]

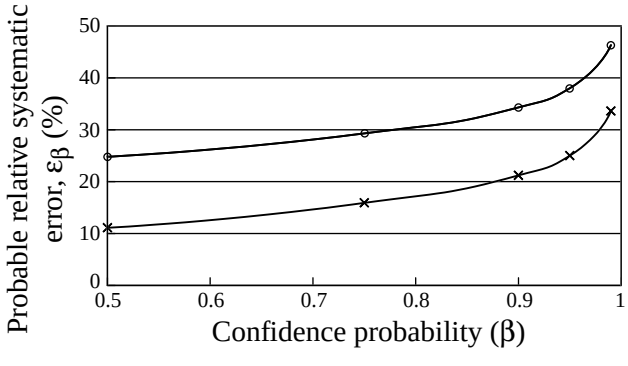

Figure 13 Graphic representation of the *probable relative systematic error* ε_β *at the confidence probability* β caused by storing mercury-contaminated soil samples in plastic bags over 2–4 weeks taken at the locations A and B in the depth interval 0–10 cm each. ×, mean values for location A, original 446 mg kg^{-1}, plastic bag 425 mg kg^{-1}; ∘, Mean values for location B, original 189 mg kg^{-1}, plastic bag 153 mg kg^{-1}.

To qualify the results represented, it must be emphasized that the definite errors mentioned are valid only for the object investigated and must not be immediately transferred to any other contaminated site. However, they may serve as a basis for developing regulations for the handling of soil samples contaminated by volatile compounds and for introducing instruments to control the influence of the handling on the mercury content detectable by chemical analysis.

3.2.4 Summary and Conclusions

Using the example of a mercury-contaminated site, it was demonstrated that the handling of soil samples can affect the mercury content detectable by chemical analysis. To distinguish this effect from the errors of sampling, sample preparation and chemical analysis, as well as from the global and local effects of the geogenic background, the following sampling design was used. One soil sample (called the primary sample) was taken at two different depth intervals from each of three locations, and each primary sample was then split into three more portions, called subsamples. The subsamples were again separated into several parts, each of which was submitted to a particular method of handling before analysis. Every method of handling defines a method of sample pretreatment. One portion was kept as the unchanged original material, and its mercury content was immediately determined. The following methods of sample pretreatment were applied: air drying for 2–4 weeks; storing in plastic bags for 2–4 weeks; and thermal evaporation at different temperatures from 40 °C up to 120 °C. The pretreated samples were analyzed, and the mercury content detected was compared with the content of the respective original material.

First, variance analysis was applied to globally quantify by variances the effects of sample preteatment, sample preparation, global fluctuations, local heterogeneities and chemical analysis. The dominant source of variability was the global fluctuations of the mercury content. ''Global fluctuations'' means that the mercury content varies significantly from location to location. These fluctuations were caused by the geogenic structure of the soil. As it turned out, the variability caused by sample pretreatment was of the same order of magnitude as the local heterogeneity variance. The local heterogeneity variance corresponds to the sampling error variance. A significant error of sample preparation did not actually exist.

Second, the qualitative statement that sample pretreatment influenced the mercury content determined by laboratory measurement was quantified by the concept of the probable systematic error of pretreatment. Following this concept, the difference between the mean mercury content of the pretreated sample and the mean value of the respective original sample material describes the systematic error which is made if the original soil material is not available, but only material handled by a certain method can be investigated. For example, if storing the soil material in a plastic bag for at least two weeks, a probable relative systematic error of up to 40% at a confidence probability of 95% must be taken into account. The concrete results obtained cannot be immediately transferred to other objects, but they may serve as a basis for developing regulations for the handling of soil samples containing volatile compounds and for introducing controlling instruments.

ACKNOWLEDGMENTS

The geostatistical calculations reported here were carried out by M. Herbst using the Windows-based commercial computer program ''Surfer'', which is frequently used for geostatistical applications. The author wishes to thank him for providing the calculations and for helpful discussions regarding the interpretation of the results.

The list of analytical chemistry methods used for the investigation of wastes was compiled by K. Volke. He also compiled the German, European, American and International standards for this field as well as for the determination of pollutants in soils and elutriates. The author wishes to thank him for his careful work and for some useful advice.

LIST OF SYMBOLS

a	Range, Correlation Length, Minimum Distance Between Uncorrelated Data
b_l	Parameter of a Linear Variogram Model Function
C	Partial Sill of a Variogram Model
C_e, a_e	Parameters of an Exponential Variogram Model Function
cov	Covariance Operator of Two Random Variables
C_Y	Covariance Function
C_0	Nugget Effect Obtained by a Variogram Model Function
D^2	Variance Operator of a Random Variable
D^2Y_t	Variance of a Random Variable Y_t
E	Operator of the Mathematical Expectation of a Random Variable
EY_t	Expectation Value of the Measuring Variable Y_t at the Point t
G	Set of all Possible Points of the Slag Heap
I_β	Probable Systematic Error at the Confidence Probability β
K	Correlation Function
m	Degree of Freedom in Statistical Comparisons
n, n_1, n_i	Number of Samples
S_n	Standard Deviation of a Measuring Variable Y, Mean Error of an Individual Value
S_n^2, S_Y^2, S_1^2	Empirical Variance Obtained by n Samples, Maximum Variance of a Variable Y, Sill
$S^2_{\overline{Y}_n}$	Variance of the Mean Obtained by n Samples
$S\left(\lvert\overline{Y}_1 - \overline{Y}_i\rvert\right)$	Standard Deviation of the Mean Value Difference $\lvert\overline{Y}_1 - \overline{Y}_i\rvert$
$t_{m;(1+\beta)/2}$	Quantile of the Order $(1+\beta)/2$ of the t-Distribution at the Degree of Freedom m
$t, t+\tau$	Points, Sampling Locations
V	Variogram Function
V_{total}	Total Variance
w_i	Weight of a Measuring Value at a Point t_i
$\overline{Y}_n$	Mean Value Obtained by n Samples
$\lvert\overline{Y}_n - \mu\rvert$	Absolute Error of the Mean $\overline{Y}_n$
$\overline{Y}(t_0)$	Mean of a Random Variable Y at the Point t_0
Y, Y_t	Random Variable, Measured Variable at a Location t
Z_{ijklm}	Random Error of Analysis
α_i	Random Effect of Global Fluctuations Within a Model of Variance Analysis
β	Confidence Probability, Confidence Level
β_{ij}	Random Effect of Local Heterogeneities Within a Model of Variance Analysis

ε_n	Probable Error of the Mean Obtained by n Samples
ε_β	Probable Systematic Error in Statistical Comparisons at the Confidence Probability β
ε_0	Maximum Admissible Error of the Mean
γ	Semivariogram Function
γ_{ijk}	Random Effect of Sample Preparation
δ_{ijkl}	Random Effect of Sample Pretreatment
λ	Lagrange Parameter in Systems of Equations with Side Conditions
μ	Expectation Value of a Random Variable, True Content, Overall Expected Value
$\tilde{\mu}_{1,\beta}$	Confidence Limit of the Mercury Content of the Unchanged Original Material
$\lvert\mu_1 - \mu_i\rvert$	Systematic Error in the Detection of the Mercury Content Caused by Handling the Soil Sample According to the ith Method of Pretreatment
$\sigma^2_{\text{analysis}}$	Variance of the Error of Instrumental Measurement, Variance of the Error of Analysis
σ^2_{global}	Variance of a Measured Variable Caused by Global Fluctuations
σ^2_{local}	Variance of a Measured Variable Caused by Local Heterogeneities of the Site
$\sigma^2_{\text{prediction}}$	Variance of Prediction (Estimation) of an Unknown Value by Kriging
$\sigma^2_{\text{preparation}}$	Variance of Sample Preparation
$\sigma^2_{\text{pretreatment}}$	Variance of Sample Pretreatment (Handling)
σ^2_{total}	Total Variance of a Measured Variable
τ	Lag of Each of Two Measuring Variables, Spatial Distance Between Sampling Points
τ_{min}	Minimal Lag of Each of Two Measured Variables in a Sampling Scheme

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
ANOVA	Analysis of Variance
DEV	Deutsche Einheitsverfahren
DIN	Deutsches Institut Für Normung
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
ISO	International Organization for Standardization
LAGA	Länderarbeitsgemeinschaft Abfall
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
USEPA	United States Environmental Protection Agency
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites

Food **(Volume 5)**
Sample Preparation for Food Analysis, General • Sample Preparation, Headspace Techniques

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Sieving in Particle Size Analysis

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Steel and Related Materials **(Volume 10)**
Iron Ore, Sample Preparation and Analysis of

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. P. Gy, *Sampling of Particulate Materials – Theory and Practice*, 2nd edition, Elsevier Publishing Company, Amsterdam, 1982.
2. K. Sommer, *Sampling of Powders and Bulk Materials*, Springer-Verlag, Berlin, 1986.
3. J.W. Merks, *Sampling and Weighing of Bulk Solids*, Trans Tech Publications, Clausthal-Zellerfeld, 1985.
4. R. Storm, *Wahrscheinlichkeitsrechnung, Mathematische Statistik und Statistische Qualitätskontrolle*, 9th edition, Fachbuchverlag, Leipzig, 1988.
5. P.H. Müller, (ed.), *Lexikon der Stochastik*, 5th edition Akademie-Verlag, Berlin, 1991.
6. H. Vogt, *Methoden der Statistischen Qualitätskontrolle*, B.G. Teubner-Verlag, Stuttgart, 1988.

7. *Qualitätssicherung und Angewandte Statistik. Verfahren 2: Probenahme und Annahmestichprobenprüfung*, DIN-Taschenbuch 225, Beuth Verlag, Berlin, 1993.
8. G. Schubert, 'Kennzeichnung der Festdispersen Sekundärrohstoffe und Abfälle', Preprint der 3 Tagung der Gesellschaft für Aufbereitungstechnik und Recycling, Freiberg, 31–44, November 9–10, 1994.
9. L.H. Keith (ed.), *Principles of Environmental Sampling*, American Chemical Society, Washington, 1988.
10. A. Gomez, R. Leschber, P. L'Hermite (eds.), *Sampling Problems for the Chemical Analysis of Sludge, Soils, and Plants*, Elsevier Applied Science Publishers, London, 1986.
11. B. Markert, *Instrumental Element and Multi-element Analysis of Plant Samples, Methods and Applications*, John Wiley & Sons, Chichester, 1996.
12. B. Markert (ed.), *Environmental Sampling for Trace Analysis*, VCH Publishers, Weinheim, 1994.
13. M. Stoeppler (ed.), *Sampling and Sample Preparation – Practical Guide for Analytical Chemists*, Springer-Verlag, Berlin, 1994.
14. W. Rasemann, B. Markert, 'Industrial Waste Dumps, Sampling and Analysis', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 2356–2373, Vol. 4, 1998.
15. P. Quevauviller (ed.), *Quality Assurance in Environmental Monitoring Sampling and Sample Pretreatment*, VCH Publishers, Weinheim, 1995.
16. W. Rasemann, 'Brauchen wir eine Probenahmeforschung für die Abfall- und Entsorgungswirtschaft?', *Müll und Abfall*, **25**(6), 460–469 (1993).
17. W. Rasemann (ed.), *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, Trans Tech Publications, Clausthal-Zellerfeld, 1999.
18. W. Rasemann, M. Herbst, 'Probenahme und Geostatistische Datenanalyse zur Bewertung von Altablagerungen – Möglichkeiten und Grenzen am Beispiel einer Freiberger Schlackenhalde', Preprint der 2 Tagung des Arbeitskreises ''Probenahme'' des IQS Freiberg, 25–40, October 25–26, 1996.
19. W. Rasemann, M. Herbst, 'Bewertung Heterogener Altlasten durch Probenahme, Statistische und Geostatistische Modellierung', *Umweltwissenschaften und Schadstoff-Forschung, Zeitschrift für Umweltchemie und Ökotoxikologie, Landsberg*, **10**(3), 81–89 (1998).
20. N. Nothbaum, R.W. Scholz, Th. May, *Probenplanung und Datenanalyse bei Kontaminierten Böden*, Erich Schmidt Verlag, Berlin, 1994.
21. *Soil Quality Sampling, Part 1: Guidance on the Design of Sampling Programs*, ISO/DIS (Draft Industrial Standard) 10381, International Organization for Standardization, Geneva, 1995.
22. A. Paetz, 'ISO 10382: Probenahme von Böden, Stand der Bearbeitung der Internationalen Normen', Preprint der 1 Tagung des Arbeitskreises ''Probenahme'' des IQS Freiberg, 119–127, November 17–18, 1995.
23. G. Woede, 'Meßpunktraster im Kleinen und im Großen Maßstab', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 97–104, 1999.
24. *Iron Ores, Increment Sampling and Sample Preparation, Manual Method*, ISO 3081, 2nd edition, International Organization for Standardization, Geneva, 1986.
25. 'Baugrund, Untersuchung von Bodenproben; Bestimmung der Korngrößenverteilung', DIN 18123, April 1983.
26. A. Rabich, 'Repräsentative Probenahme bei Heterogenen Feststoffen – Zur Untersuchung von Schredder-Rückständen auf PCB-Gehalt', *Abfallwirtschafts J.*, **3**(9), 501–510 (1991).
27. A. Rabich, 'Zur Bestimmungsunsicherheit von Stoffinhaltbestandteilen bei Heterogenen Massen', Preprint der 2 Tagung des Arbeitskreises ''Probenahme'' am IQS Freiberg, 105–117, October 25–26, 1996.
28. G. Knoop, 'Charakterisierung von Abfällen: Normen für die Probenahme und für die Stoffliche Analyse', Preprint der 1 Tagung des Arbeitskreises ''Probenahme'' am IQS Freiberg, 133–142, November 17–18, 1995.
29. H. Krutz, J. Niessing, 'Zur Stellung des Probenehmers bei der Bewertung von Abfällen und Altlasten', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 189–192, 1999.
30. L. Gramlich, 'Altlasten aus Umweltrechtlicher Sicht', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 193–199, 1999.
31. E.-M. Habersbrunner, 'Probenahme und Gutachten in der Umweltstrafrechtlichen Praxis', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 201–204, 1999.
32. PN 2/78K, Grundregeln für die Entnahme von Proben aus Abfällen und abgelagerten Stoffen, in *Richtlinie für das Vorgehen bei Physikalischen und Chemischen Untersuchungen im Zusammenhang mit der Beseitigung von Abfällen*, Erich Schmidt Verlag, Berlin, 12–24, 1985.
33. W. Schmeken, *TA Abfall, TA Siedlungsabfall – Textausgabe der Zweiten und Dritten Allgemeinen Verwaltungsvorschrift zum Abfallgesetz*, 3rd edition, Deutscher Gemeindeverlag und Verlag W. Kohlhammer, Cologne, 1993.
34. *Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung: Physikalische, Chemische, Biologische und Bakterielle Verfahren*, Part 41 Wiley-VCH Verlag, Weinheim, 1998.

35. H. Hein, W. Kunze, *Umweltanalytik*, 2nd edition, VCH Verlag, Weinheim, 1995.
36. 'Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung; Anionen (Gruppe D); Bestimmung der Sulfat-Ionen (D5)', DIN 38405, Part 5, January 1985.
37. 'Baugrund, Versuche und Versuchsgeräte, Bestimmung des Glühverlustes, DIN 18128, November 1990.
38. E.H. Isaaks, R.M. Srivastava, *Applied Geostatistics*, Oxford University Press, Oxford, 1989.
39. G. Matheron, *Les Variables Regionalisée et Leur Estimation*, Masson, Paris, 1965.
40. L. Ries, 'Geostatistik zur Gefährdungsabschätzung von Altlasten: Möglichkeiten und Grenzen der Methodik', in *Altlastenbewertung – Datenanalyse und Gefahrenbewertung*, eds. S. Schulte-Hostede, R. Freitag, A. Kettrup, W. Fresenius, Ecomed Verlagsgesellschaft, Landsberg, 113–138, 1996.
41. N.A.C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, 1993.
42. G. Matherton, *Traité de Geostatique Appliquée*, Editions Techniq, Paris, Vol. 1, 1962, Vol. 2, 1963.
43. A.G. Journel, Ch.J. Huijbregts, *Mining Geostatistics*, Academic Press, London, 1978.
44. L. Ries, 'Evaluation of Spatial Data Sets, Mathematical and Statistical Tools', in *Quality Control and Assurance of Applied Soil Microbiology and Biochemistry*, Academic Press, London, Chapter 2, 1995.
45. J.W. Einax, U. Soldt, 'Möglichkeiten und Grenzen der Anwendung Geostatistischer und Multivariat-statistischer Methoden zur Bewertung Belasteter Böden', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 73–79, 1999.
46. H. Geiler, H.S. Dengel, A. Donsbach, K. Knoblich, F. Aschenbrenner, R. Ostermann, W. Maurer, *Geostatistische Analyse von Schwermetallverteilungen in Böden*, GIT Labor-Fachzeitschrift, 1138–1143, Vol. 11, 1998.
47. G.T. Flatman, E.J. Englund, A.A. Yfantis, 'Geostatistical Approaches to the Design of Sampling Regimes', in *Sampling Problems for the Chemical Analysis of Sludge, Soils, and Plants*, eds. A. Gomez, R. Leschber, P. L'Hermite, Elsevier Applied Science Publishers, London, 73–84, 1986.
48. N.A.C. Cressie, 'Statistics for Spatial Data', *Mathematical Geol.*, **22**, 239–252 (1990).
49. A.G. Journel, *Fundamentals of Geostatistics in Five Lessons*, American Geophysical Union, Washington, DC, 1989.
50. P. Rocklin, M. Jauzein, 'An Optimized Interpolation-kriging Method for the Study of Spatially Distributed Parameters from Contaminated Sites', in *Contaminated Soil '95, Proceedings of the Fifth International FZK/TNO Conference on Contaminated Soil, 30 Oct.–3 Nov. 1995, Maastricht*, eds. W.J. Van den Brink, R. Bosman, F. Arendt, Kluwer Academic Publishers, Dordrecht, 255–256, 1995.
51. D. Keckler, *SURFER for Windows, User's Guide*, Golden Software, Inc., Golden, CO, 1995.
52. Klärschlammverordnung AbfKlärV vom 15.4.1992, Bundesgesetzblatt, Germany, BGB1, III 2129-6-6.
53. A. Paetz, 'Einführung in die im untergesetzlichen Regelwerk zum Bundesbodenschutzgesetz Aufgeführten DIN-Normen', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 117–120, 1999.
54. W. Rasemann, U. Seltmann, M. Hempel, 'Statistical Evaluation of Several Sample Pretreatment Methods on the Mercury Content Detectable by Chemical Analysis of Contaminated Soil Samples under Practical Conditions', *Fresenius' J. Anal. Chem.*, **351**, 632–641 (1995).
55. R.G. Miller, *Beyond ANOVA, Basics of Applied Statistics*, John Wiley & Sons, New York, 1986.
56. W. Rasemann, U. Seltmann, M. Hempel, 'Einflußder Probenahme und Probenvorbehandlung auf die Analyse von Quecksilber-kontaminierten Böden', Preprint der 3 Tagung der Gesellschaft für Aufbereitungstechnik und Recycling Freiberg, 261–268, November 9–10, 1994.
57. E.S. Pearson, H.O. Hartley, *Biometrika Tables for Statisticians*, 3rd edition, Cambridge University Press, Cambridge, Vol. 1, 1976.
58. W. Rasemann, 'Qualitätssicherung von Stoffsystemen durch Probenahme und Datenanalyse – Probleme, Erfahrungen, Tendenzen', in *Qualitätssicherung von Stoffsystemen im Abfall- und Umweltbereich – Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 9–21, 1999.
59. 'Precision of Test Methods; Determination of Repeatability and Reproducibility for a Standard Test Method by Inter-laboratory Tests', DIN ISO 5725, April 1988.

Infrared Spectroscopy in Environmental Analysis

Tom Visser
National Institute of Public Health and The Environment, Bilthoven, The Netherlands

1 Introduction 1
2 History 2
3 Sampling and Measurement Techniques 3
4 Environmental Applications 5
4.1 Indoor Air 5
4.2 Atmosphere and Stratosphere 5
4.3 Smoke Stacks and Exhaust Gases 7
4.4 Terpenes and Biogenic Aerosols 8
4.5 Volcanic Emissions 8
4.6 Biomass Burning 8
4.7 Volatile Organic Compounds 9
4.8 Industrial Contaminants 10
4.9 Pesticides 10
4.10 Phenols 12
4.11 Polyaromatic Hydrocarbons 13
4.12 Dioxins, Dibenzofurans and Polychlorinated Biphenyls 13
4.13 Oils and Greases 14
4.14 Soil Characterization 15
4.15 Inorganics and Asbestos 15
4.16 Polymer Characterization 15
4.17 Quality Assurance 16
4.18 Chemometrics and Model Studies 16
5 Current Position and Future Developments 16
Abbreviations and Acronyms 17
Related Articles 17
References 18

Infrared (IR) spectroscopy is a universal, versatile analytical technique for the structure elucidation and quantification of a large variety of organic, inorganic and biological samples. The technique is based on the measurement of IR radiation, absorbed or emitted by a sample, as a function of the wavelength in the region 2.5–100 μm, or 4000–100 cm^{-1} in wavenumbers. The absorption or emission of IR radiation is related to discrete vibrational and rotational transitions in a molecule and, therefore, represents structural information. The analytical value of IR spectroscopy is based on the fact that the IR bands occur at more or less localized positions in the spectrum which are correlated to the presence of characteristic structural features of the sample under study. This similarity and transferability of spectral characteristics and the corresponding structural features, makes IR a powerful tool for functional group analysis. On the other hand vibrations, and thus the exact peak positions, are sensitive to the local environment which also makes IR spectra unique molecular fingerprints, highly suited for the unambiguous identification of a sample by comparison with reference spectra. Finally, IR spectroscopy obeys Beer's Law and can thus be used for quantitative purposes too.

The major advantage of IR over other spectroscopic techniques is that practically all compounds show absorption/emission and can thus be analyzed both quantitatively and qualitatively. Besides, IR spectroscopy is nondestructive and admits in situ and remote measurement of almost any sample, irrespective the physical state and without elaborate preparations. A broad variety of advanced accessories and technologies, such as IR microscopy, optical fibers, remote sensing, reflection and emission equipment and the combination with gas chromatography (GC), liquid chromatography (LC) and supercritical fluid chromatography (SFC) sampling and measurement techniques are available to tackle a diversity of problems. The absolute sensitivity is about 50–500 picograms, which is sufficient to solve many problems in trace analysis.

IR spectroscopy is used in a broad range of environmental applications. The technique is a viable tool in remote measurements of gases, liquids and solids and as a fast and reliable method for the characterization and indicative structure classification of a wide variety of samples. Besides, IR is widely applied for identification purposes either to confirm results of other techniques or for reasons of quality assurance (QA).

1 INTRODUCTION

The identification of molecular structures and the determination of their concentration in the sample is an essential part of environmental analysis. Among other selective techniques such as ultraviolet (UV), fluorescence, atomic emission and mass spectrometry, IR spectroscopy has gained wide acceptance as a viable technique for this purpose. The principle of IR spectroscopy is the measurement of the amount of IR radiation absorbed or emitted by a sample as a function of the wavelength. The wavelength of IR ranges from about 0.8 to 100 μm (12 500–100 cm^{-1} in wavenumbers) and is divided into a near-infrared (NIR) (12 000–4000 cm^{-1}) and a mid-IR (mid-infrared) (4000–100 cm^{-1}) part. This

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

chapter deals with the spectroscopy related to the mid-IR radiation. NIR spectrometry is discussed in the article **Liquid Chromatography/Infrared Spectroscopy** in this encyclopedia.

The absorption or emission of IR photons is related to rotational and vibrational transitions in a molecule and therefore represents structural information. The corresponding IR bands occur at discrete frequencies in the IR spectrum and this relation between the molecular structure of a sample and the IR spectrum is the most important analytical value of IR spectroscopy.

The potential of IR spectroscopy in the elucidation of molecular structures is very high. The number of fundamental vibrations in a molecule of N free-moving atoms is $3N - 6$. In principle, each vibration may give rise to an absorption or emission band but in practice the number is smaller as vibrations may be inactive as result of symmetry and degeneracy. On the other hand, much more than $3N - 6$ absorption bands may be present in the spectra of vapor-phase molecules as result of rotational transitions. Besides, vibrations and rotations are sensitive to a number of physical parameters, such as temperature and pressure, and to intra- and intermolecular interactions. The large number of vibrations and rotations, and the diversity of variables affecting these, make IR spectra unique molecular fingerprints, highly suited for the identification of a sample by comparison with reference spectra.

Valuable structural information can also be derived from the interpretation of spectra, e.g. in case a matching reference spectrum is absent. Contrary to identification, based on the discriminative properties of the spectrum, interpretation relates to the similarity and transferability of spectral characteristics and the corresponding structural features of molecules, i.e. functional groups. For that reason IR spectroscopy is also a useful technique for the characterization and structural classification of samples. However, as result of the large number of variables affecting a spectrum, it is extremely difficult to arrive at a complete molecular structure from a single spectrum without additional knowledge.

Finally, IR spectroscopy is an absorption technique obeying Beer's Law. It can thus be used for quantitative purposes too. A drawback compared to, for instance, UV, visible and fluorescence spectrometry is the small extinction coefficient of IR absorption bands resulting in relatively poor detection limits. Nevertheless, the current absolute limit of detection of about 10–50 picograms, is sufficient to solve many problems in environmental trace analysis.

The major advantage of IR over other techniques is that, apart from homonuclear diatomic molecules such as H_2, N_2 and O_2, all compounds show IR absorbance and can thus be analyzed qualitatively as well as quantitatively. Besides, IR is a nondestructive, noncontact technique, offering the possibilities to perform in situ and remote measurements. From this practical point of view, IR is probably the most versatile of all analytical techniques, as a spectrum can be recorded of almost any sample, irrespective of the physical state and without elaborate preparations within tenths of a second. This versatility has made IR spectroscopy a valuable analytical tool in a variety of environmental applications.

2 HISTORY

Since the development of IR spectroscopy in the early 1940s the instrumentation and environmental applications have developed through a number of stages. The first generation IR spectrometers were single-beam instruments with alkali halide prisms as the dispersive element. The sensitivity and the resolution of these instruments were poor, limiting the applications to functional group analysis and indicative quantification of milligrams of material.

In the 1960s, the instruments became double-beam, and Echelette gratings, combined with optical filters for order separation, replaced the prisms. As a consequence, the optical resolution and the stability and sensitivity of IR spectrometers were considerably enhanced and the detection limits extended to micrograms. A further development, the coupling to chromatographic techniques, via stopped-flow and flow-through cells, led to new applications, particularly in the structure elucidation of constituents in complex matrices.

A further positive impulse was given by the incorporation of computers into the IR spectrometers. Advanced data acquisition and manipulation techniques such as spectrum subtraction, programmes to search in large libraries of digitized spectra, and chemometrical methods were brought to the bench of IR spectroscopists.

Finally, the introduction of the Fourier transform (FT) principle in IR spectrometers in the 1970s dramatically improved the possibilities of IR spectrometry in a wide range of analytical problems. The high scanning speed and sensitivity of Fourier transform/infrared (FT/IR), combined with extended computer facilities for data manipulation and the development of a series of advanced accessories and technologies, such as IR microscopy, high resolution interferometers ($>0.001\,cm^{-1}$), tunable lasers, specific detectors, reflection and emission equipment and the construction of mobile and portable instruments, have brought IR to its current state of versatility and popularity.

3 SAMPLING AND MEASUREMENT TECHNIQUES

As result of the unique versatility of IR spectrometry a broad variety of sampling and measurement techniques are available to tackle a diversity of problems. The development of several of these methods originate from an environmental demand. An illustrative example is the IR remote sensing technique for the qualitative and quantitative in situ determination of gases.[1] The need to study emissions and chemical processes of greenhouse gases largely contributed to the development of this open-path technique to its current state of maturity. Obviously, the fact that global warming is assumed to be caused by the increasing release of exogenic IR-absorbing gases into the atmosphere, plays a role in this.

The principle of IR remote sensing is shown in Figure 1. In the passive mode, the sun is used as a light source or the spectrometer is operated as a detector to determine the IR radiation emitted by the subject of study. In the active mode the instrument operates as a classical source–detector spectrometer, recording absorption along the beam path to a flat mirror or a retroreflector and back. The IR source and the detector are either incorporated in one instrument or separated into modules.

Absorption measurement in a closed gas cell is a good alternative to carry out model studies and in case the analysis has to be performed on previously collected samples. Internal multi-reflection options offer the possibility of enlarging the effective pathlength to 1–100 m within the relatively small dimensions of a 10–30 cm gas cell. This method is particularly attractive in case extremely low detection limits are not crucial. Gas cells are used both separately as an accessory, and fixed, incorporated into a (mobile) spectrometer. An example of the latter is the commercially available mobile FT/IR spectrometer GASMET™ (Temet Instruments Ltd, Finland) as shown in Figure 2.[2]

Remote measurements of gases, liquids and solids can also be carried out with the help of optical fibers.[3] Fibers are used as a transmission wave guide to transfer the IR signal to and from a sensing probe, and as a sensing element itself. The fibers are made of materials such as chalcogenides that combine a certain physical flexibility with a high IR transparency. Optical fibers are particularly useful to carry out in situ measurements and to study compounds with possible toxic or corrosive activity. At present, distances of 5–50 m can be covered, depending on the desired wavenumber region to be scanned.

The advantages of optical fibers and solid-phase micro extraction (SPME) have been combined to a fiber optic

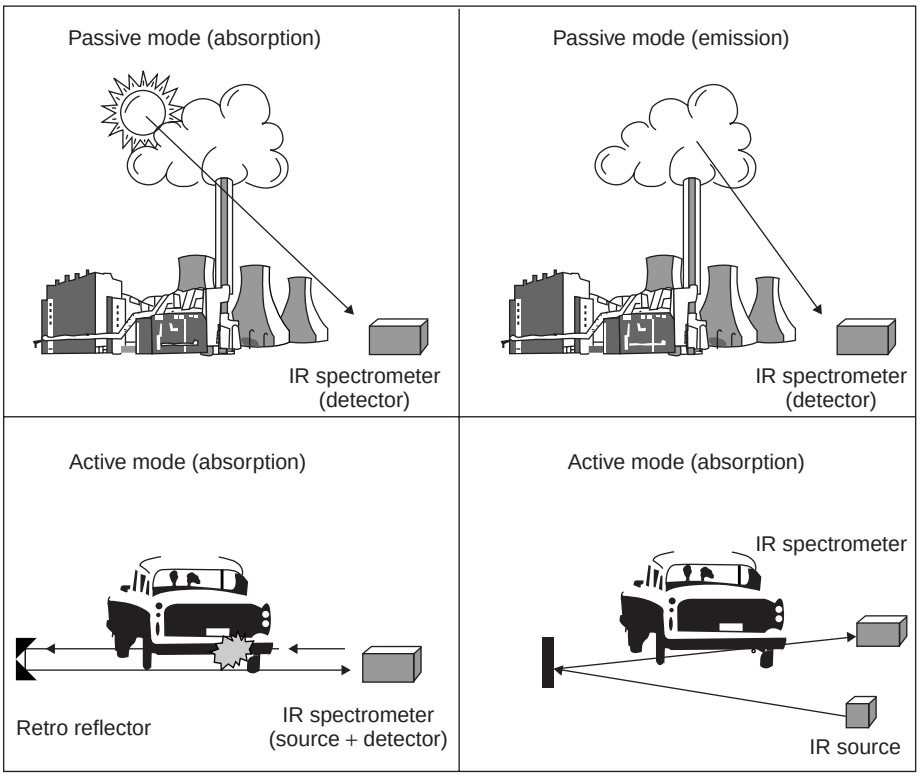

Figure 1 Schematic representation of the principle of active and passive remote sensing IR measurements.

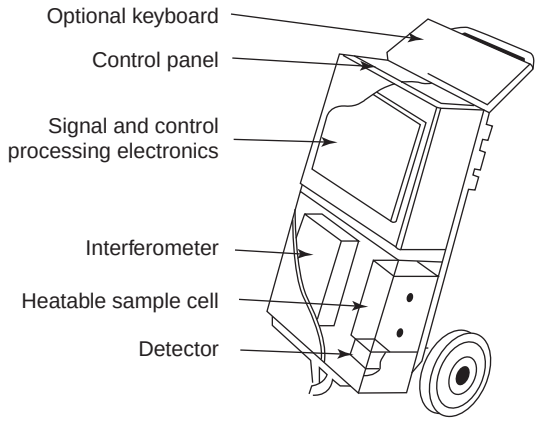

Figure 2 Schematic diagram of a mobile IR gas analyzer (Reproduced by permission of The Royal Society of Chemistry from I. Ahonen, H. Riipinen, A. Roos, 'Portable Fourier Transform Infrared Spectrometer for Use as a Gas Analyzer in Industrial Hygiene', *Analyst*, **121**, 1253–1255 (1996).)

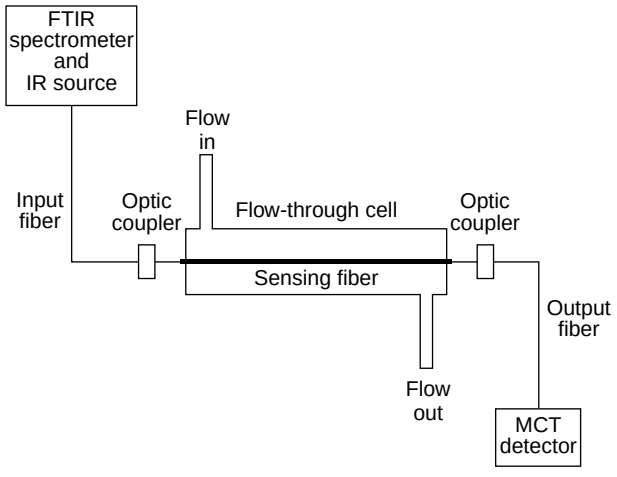

Figure 3 Schematic representation of optical fiber sensing for aqueous samples. (Reproduced by permission of The Royal Society of Chemistry from J.E. Walsh, B.D. MacCraith, M. Meaney, J.G. Vos, F. Regan, A. Lancia, S. Artjushenko, 'Sensing of Chlorinated Hydrocarbons and Pesticides in Water Using Polymer Coated Mid-infrared Optical Fibers', *Analyst*, **121**, 789–792 (1996).)

sensor system for the analysis of microcontaminants in aqueous samples.[4] The experimental set-up of this system is schematically depicted in Figure 3. The actual sensor consists of a silver halide fiber coated with a thin layer of polymer. Similar to SPME, the organic analytes are extracted from the water into the polymer film and scanned with a (portable) IR spectrometer after a certain equilibration time.

Conventional sampling methods such as the liquid cell and the KBr-pelleting technique, as well as more advanced methods such as diffuse reflectance infrared (DRIFT) and attenuated total reflection (ATR), are being used for the characterization and identification of samples in case the available amount is not the limiting factor. For quantitative purposes, liquid cells of variable pathlength, cell volume and window material are available.

IR microscopy is being employed for the analysis of small particles and in combination with the on- and off-line coupling to separation techniques for the determination of trace contaminants. The principle and applications of IR microscopy have been described by Messerschmidt and Harthcock.[5]

The application of IR microscopy as a detector in GC, LC and SFC has largely enhanced the sensitivity of these combined techniques. The current limit of detection is in the (sub)nanogram range which makes these hyphenated techniques useful alternatives or additives to commonly applied detectors such as flame ionization, electron-capture and mass spectrometry.

The interfacing of GC and IR is accomplished in three different ways which are all commercially available. The first one is on-line by means of a heated flow-through cell with an optical pathlength of 10–30 cm. This so called 'light pipe' interface allows continuous monitoring of the GC effluent.

The second one is off-line and comprises the trapping of the GC-separated analytes in an argon matrix on a rotating gold coated cylinder at 4K. After completion of the GC-run, the immobilized chromatogram can be scanned at any desired speed and resolution as long as the cryogenic conditions are maintained. This technique is called matrix isolation gas chromatography/infrared (GC/IR).

The third one is also a cryogenic sample storage technique and is called cryotrapping or direct deposition GC/IR. The analytes are crystallized directly on a moving IR transparent window at liquid nitrogen temperature (77K). Scanning can be carried out both on-line, a few seconds after deposition, or post-run, after completion of the GC separation.

The most important advantage of the sample storage methods is the sensitivity which is about two orders of magnitude better than that of a light-pipe instrument as result of the application of IR microscopy and the possibility of long-time signal averaging.[6]

A series of papers on the environmental applications of GC/IR have been published by Gurka et al.[7] The analytical value of multidimensional gas chromatography/infrared/mass spectrometry (GC/IR/MS) for the investigation of environmental samples has been summarized by Krock and Wilkins,[8] while the synergetic value of GC/IR in combination with other spectrometric techniques has been reviewed by Ragunathan et al.[9]

IR detection in LC and SFC is accomplished on-line by passing the effluent through a liquid flow-through cell, or off-line by elimination of the solvent prior

to detection of the deposited analytes. Flow cells for any type of mobile phase, including supercritical fluids, and solvent elimination liquid chromatography/infrared (LC/IR) interfaces can be purchased from several companies. Off-line detection is highly favorable since important parts of the spectral region may be obscured by absorption bands of the eluents. Particularly in reversed phase, the strong absorption of water is a major limiting factor not only in terms of poor spectral information but also in sensitivity as a feasible pathlength should not exceed 10–20 microns. An extensive review of the merits and limitations of on- and off-line coupled LC/IR has been reviewed by Somsen et al.[10]

4 ENVIRONMENTAL APPLICATIONS

Numerous papers on the application of IR spectrometry in environmental analysis have been published in analytical, environmental and spectroscopic journals. State-of-the-art surveys are frequently published of which the comprehensive biannual reviews of *Analytical Chemistry* on IR spectroscopy[11] and environmental analysis[12] are recommended as good starting points.

Papers on specific IR sampling and detection techniques also contain valuable information on environmental applications. Representative examples are the previously mentioned reviews on IR microscopy,[5] GC/IR[7–9] and LC/IR.[10]

The diversity in applications is extremely large. The analytes vary from organic and bio-organic molecules to polymers, inorganics and even microorganisms. Next to air, soil and water, IR analysis is employed in all sorts of matrices at concentration levels between bulk and trace. Real-time and in situ measurements are performed for dynamic studies while various on- and off-line methods are being used for characterization, classification, structure elucidation and quantitative analysis. A summary of the IR spectroscopic applications is presented in Table 1. A selection of the most relevant and illustrative applications is described into more detail in the following sections.

4.1 Indoor Air

The concern for good quality of indoor air has urged the development of techniques that are able to monitor the presence and release of harmful gases. Not only in the laboratory and industrial environment but also in, for instance cleaning and printing shops and during painting, relatively high concentrations of volatile organic compounds (VOCs) may be released. In principle, IR spectrometry is an attractive technique for the screening and monitoring of VOCs in the gas phase. All compounds show IR absorption and can thus be determined, while measurements can be performed remotely and continuously without any sampling. The sensitivity of IR is in the low ppm range, which is relatively low compared to some other techniques but sufficiently high for many applications. Therefore, several systems to monitor emissions of VOCs are based on IR detection.

Three types of IR–gas measurement systems can be distinguished:

1. low cost non-scanning detectors that are specifically used to determine the presence of one or two target-components;
2. portable instruments with filters or an interferometer for wavelength selection;[2]
3. remote open-path IR spectrometers.[13]

An instrument, specially designed to monitor indoor air concentrations, is the Infrared Gas Cloud (IGC) scanner developed by ter Kuile et al.[14] This apparatus is used for the selective analysis of gases in the ppm range to visualize the spatial distribution of the analyte in the workplace. Specific applications are the testing of ventilation systems and the remote detection of gas leaks.

Dispersive IR analysis of gases, released by the smoking of cigarettes, proves to be a better method for quantification than most other spectrometric techniques. The yield for carbon monoxide, hydrogen cyanide and nitric oxide turns out to be about 45% higher when detected with IR.[15]

In case direct IR measurement is hampered, for instance as result of weak absorbing properties of the analyte, indirect detection after conversion into a stronger absorbing compound is a good alternative. An example is the analysis of H_2S and O_2 which can be detected at much lower concentrations after oxidation to SO_2 and ozone respectively, by UV irradiation over a certain period of time.[16]

Preconcentration techniques, occasionally combined with a derivatization or modification, are used as well. Fillipi [17] for instance, developed a purge and trap method combined with a chemical reduction step prior to GC/IR and atomic absorption spectrometry (AAS) to determine methyl and dimethyl mercury in air at a 50 pg level.

4.2 Atmosphere and Stratosphere

The detection of exogenic and endogenic gases in the outdoor atmosphere relies to a large extent on the same methods as used for the analysis of indoor air. Similar specific single-compound detectors are used for continuous monitoring or occasional field measurements to determine and quantify a selection of predefined harmful gases and VOCs at, for example, industrial

Table 1 Applications of infrared spectroscopy in environmental analysis

Matrix type	Aim	Methods	Analytes
Ambient atmosphere	Structure elucidation, quantification, monitoring, screening, mapping, gas release	Nonscanning IR sensors, gas cells, active remote sensing, portable instruments	CFCs, VOCs, BTEX, PAHs, nitro-PAHs CO_x, NO_x, HCN, H_2S, O_2, Hg-alkanes
Stratosphere	Structure elucidation, quantification, monitoring, local and global environment, greenhouse effects, volcanic emission studies, biomass burning	Active and passive remote sensing, multipass gas cells, GC/IR mobile and airborne spectrometers	CFCs, VOCs, BTEX, PAHs, CO_x, NO_x, HCN, H_2S, O_3, ClO_x, SO_x, HCl, HF, SiF_4, $ClNO_x$, H_2O, hydrocarbons, etc.
Aerosols	Structure elucidation, quantification, combustion products and processes, photooxidation, air particulate matter	Active and passive remote sensing, multipass gas cells, GC/IR, LC/IR	(Metabolites of) terpenes, pheromones, PAHs, nitro-PAHs, nitro-cresols, phenols photodissociation and oxidation products
Smoke and exhausts	Structure analysis, quantification, monitoring, motor cars, aircraft, incinerators, power plants	Active and passive remote sensing, GC/IR mobile spectrometers, multipass gas cells	PAHs, hydrocarbons, H_2O, CO_x, NO_x, SO_x, HCN, H_2S, O_2
Water	Structure elucidation, quantification, confirmation, trace analysis	SPE, SPME, LLE/GC/IR and LC/IR, SFE/IR, sparging-IR, optical fibers	CFCs, pesticides, industrial contaminants, pharmaceuticals, oils, greases, VOCs, PAHs, biomolecules (microorganisms)
Soil	Soil characterization, trace analysis, structure elucidation, quantification	Films and pellets, DRIFT, PAS, GC/IR, LC/IR, SFE/IR, optical fibers	HA and FA fractions, pesticides, inorganics, nitrates, silicates (asbestos), oils and greases, VOCs, CFCs, PAHs
Sediments	Structure elucidation, quantification, trace analysis, confirmation	GC/IR, LC/IR, thermal desorption IR, multipass gas cells, SPME/IR	Pesticides, PAHs, oils and greases
Reference materials	QC, identity control geometrical characterization	KBr, liquid film, solutions, DRIFT, reflection, GC/IR, LC/IR	Pesticides, PAHs, PCBs, PCDFs, PCDDs
Waste	Characterization, structure analysis, quantification, recycling purposes, polymer sorting	Active remote reflection, ATR, PAS	Industrial and domestic contaminants, oil, polymers, plasticizers, detergents

PAS, photo-acoustic spectrometry; QC, quality control.

sites and waste incinerators. Multi-pass gas cells and open-path equipment are applied for the detection of other, non-target analytes.[18]

Remote sensing open-path IR spectrometry has substantial advantages over the other methods. It can be used to measure absorption as well as emission and the technique is capable of covering almost infinite distances. This offers the possibility of detecting fluxes of trace gases over a much wider area than can be achieved by any other technique. For that reason, active and passive remote sensing IR are the methods of choice for monitoring gases and gaseous emissions, not only in the lower atmosphere but at any altitude and even in space.

A review on atmospheric and stratospheric research is given by Farmer.[19] Most studies focus on the detection and the quantification of a large variety of trace gases in order to monitor possible compositional changes as a result of human activities and to get insight into the underlying physical and chemical processes. Subjects of research are the fluxes and effects of fertilizers, pesticides, industrial waste contamination, energy production, etc. Analogous studies are performed on environmental burdening by compounds emitted from natural sources, such as plants, animals, volcanic activity and biomass burning. As noted before, special interest is given to the greenhouse effect and the depletion of the stratospheric ozone layer over the Arctic and Antarctic areas.

The IR spectra of gases in space and in the stratosphere exhibit extremely narrow vibrational and rotational bands as result of the very low temperature at site. In order to extract maximum information from the spectra, high-resolution facilities are essential and for that reason, measurements are preferably carried out with instruments with a data point resolution of 0.05 to 0.002 cm^{-1}. As

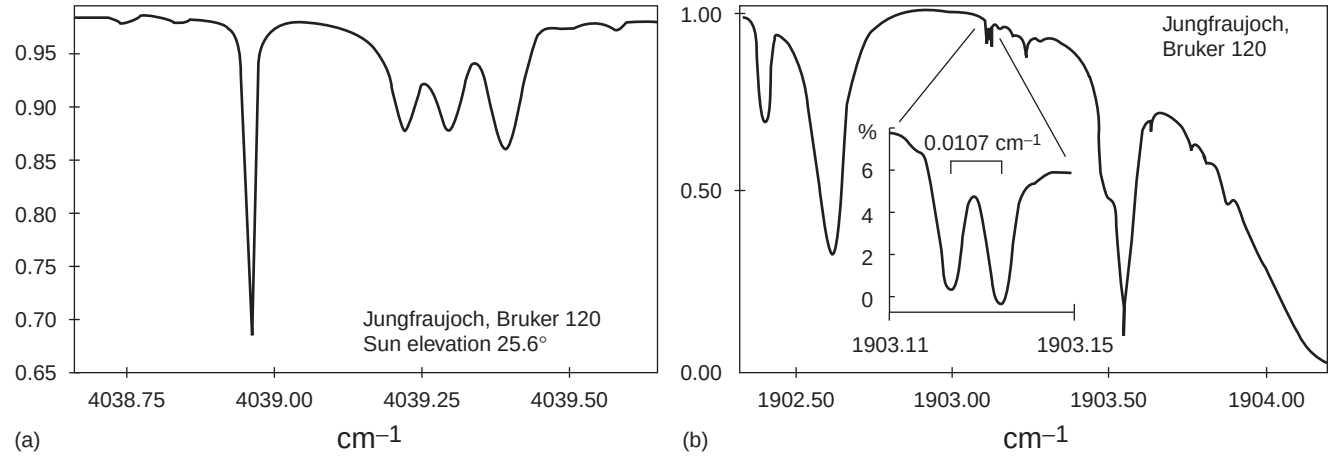

Figure 4 High-resolution IR spectra recorded at the Jungfraujoch observatory; (a) the HF line at 4038.9625 cm^{-1} and (b) the NO doublet around 1903 cm^{-1}. (Reproduced by permission of The Society of Photo-optical Instrumentation Engineers from L. Delbouille, 'The Role of Very High Resolution FTS in Recent Determinations of the Quantitative Composition of the Stratosphere', *Proc. SPIE-Int. Soc. Opt. Eng.*, **2089**, 32–37 (1993).)

well as molecules involved in the photodissociation of chlorofluorocarbons (CFCs), a diversity of other minor constituents such as CO, CH_4, HF, HCl, HCN, O_3, NO_2, NO, SF_6, N_2O, N_2O_5, H_2O_2, OCS and $ClONO_2$ can be determined with this type of IR-spectrometric technique.

Measurements are performed:

1. directly by air-borne spectrometers in space and aircraft or in balloons;
2. directly by ground-based remote sensing techniques;
3. indirectly by laboratory analysis of samples collected by aircraft or balloon.

In addition to these techniques, kinetic laboratory studies are undertaken to mimic atmospheric and stratospheric conditions and processes. For this type of research, pressure- and temperature-controlled gas cells are used.

Air-borne instruments are in general of special design such as the Atmospheric Trace Molecule Spectroscopy (ATMOS) instrument that is used in Space Shuttle flights. Measurements from aircraft are usually preferred to observations from balloons, since the latter are relatively expensive and limited to balloon-accessible altitudes of about 45 km. Yet it is a viable and frequently used method for direct analysis and for the validation of data obtained from other techniques.[20]

Examples of ground-based measurements are the observations at the Jungfraujoch mountain station in the Swiss Alps, where IR spectra of the global atmosphere and stratosphere have been acquired over a period of almost five decades,[21] and the measurements in the Arctic area where the column densities of several atmospheric gases have been determined during summer and winter seasons.[22] In these observations, the sun is used as the IR light source, either directly at daytime or indirectly by reflection via the moon at night. Long-term scanning is required to get sufficient signal-to-noise ratio which implies that a tracker is necessary to follow the course of the emitting object. Examples of the information obtained from such high-resolution data are depicted in Figure 4 showing the extremely narrow absorption lines of HF and NO recorded at the Jungfraujoch observatory.

4.3 Smoke Stacks and Exhaust Gases

Many studies have been dedicated to the on- and off-line measurement of smoke stacks, and flue gases released from traffic and power plants.[23] The popularity of the remote methods is rapidly growing due to the fact that it is a noncontact real-time technique. Besides, it enables the performance of multicomponent analysis in seconds.

A remote-sensing IR method to determine the constituents of the exhaust gas of a coal-fired power plant has been developed by E. Lindermeir.[24] The principle of this method is the measurement of the emission lines of released gases at medium resolution (0.5 cm^{-1}). Accurate concentrations of H_2O, CO_2, CO, NO and SO_2 in smoke stacks at heights up to 200 m were calculated from the IR spectra using special algorithms to compensate for background fluctuations.

A commercial remote sensing FT/IR instrument RSD 1000, (Remote Sensing Technologies Inc. Maryland, USA) is used by Sjödin[25] to determine the instantaneous concentrations of CO and hydrocarbons in the exhaust of motor cars. The measurements are used to select major polluters for subsequent roadside inspection. It has proven to be a useful tool to increase the efficiency of

roadside inspections to detect cars violating governmental idle test standards.

Roadside and laboratory inspections can be readily performed by flow-through and sampled gas cells as well. The portable mobile IR spectrometer GASMET™ for instance, equipped with a 1.8 m multipass gas cell, can detect a variety of exhaust gases at levels of 0.2–20 ppm.[2]

Specific spectrometers have been developed for FT/IR emission measurements of aircraft engine exhausts under active conditions. It enables the determination of emission rates of gases like NO, CO, CO_2 and H_2O with a resolution of 0.2 cm^{-1} during the flight at cruising altitudes up to 25 000 ft.[26]

A mobile air pollution search system has been developed by the Kayser–Threde Company (Munich, Germany). The system is based on a K300 double pendulum spectrometer that operates at a spectral resolution of 0.08 cm^{-1}. In the passive emission mode, the instrument can determine gas concentrations of CO, CO_2, N_2O, NO, SO_2 and HCl in smoke-stack effluents of power plants and municipal incinerators with high accuracy.[27]

4.4 Terpenes and Biogenic Aerosols

Environmental risk-assessment studies generally comprise the release of compounds of anthropogenic origin. From the viewpoint of health care, however, the emission of molecules from natural sources are of almost equal interest. The release of gases from a single volcanic eruption, for instance, can be in the order of billions of tons. Other sources that are known to have large impacts on the local or global environment are bush fires, aqueous microorganisms and emissions from terrestrial plants (terpenes) and animals (pheromones).

Terpenes, released from plants and trees, are known to be important in the formation of photochemical oxidants and secondary organic aerosols. To determine the qualitative and quantitative composition of terpene emissions and its oxidation products in the related aerosols, an IR spectroscopic method in addition to gas chromatography/mass spectrometry (GC/MS), has been developed by Schrader et al.[28] It is demonstrated that cryotrapping GC/IR supplies important complementary information for the structure elucidation of mono- and sesquiterpenes as well as other ethereal compounds, at the high picogram level.

Biogenic aerosols from photooxidation of terpenes have been analyzed by Palen et al.[29] too, using IR microscopy. The quantitative possibilities of this method are limited but valuable information can be obtained concerning the presence of compounds with oxygen-containing functional groups such as hydroxyls, ethers, ketones, aldehydes, carboxylic acids, lactones and esters.

Finally, IR spectrometry is also used to determine and characterize waterborne microorganisms in drinking water.[30]

4.5 Volcanic Emissions

Passive and active open-path IR spectrometry are the principal methods of monitoring emissions from active volcanoes. The IR spectra are primarily used to determine the identity and flux of gases released during eruption. Other information that is extracted from the IR data deals with the eruptive history of volcanoes and the effects of vulcanism on the local and global climate. A case study on this subject is given by Chaffin et al.[31] Measurements over pathlengths of 500 m to nearly 2 km have been carried out with a mobile spectrometer operating at a resolution of 0.5 cm^{-1}. Large amounts of CO_2 and SO_2 in the volcanic plume hamper the identification of compounds that absorb in the same regions as these two components; however, several other constituents such as HCl, SO_2 and SiF_4 can be determined at ppm level in very short measurement times (less than one minute is typical). As an example of the performance of these measurements, the field-recorded spectrum of HCl and the vapor-phase reference spectrum of this compound, are depicted in Figure 5.

4.6 Biomass Burning

According to Griffith,[32] the burning of biomass is a highly underestimated source of air pollution. The annual release of carbon into the atmosphere as result of wild fires and the prescribed combustion of bush, forest, savanna and agricultural waste is about 3–6 gigaton which is of the same order as the annual combustion of fossil fuels. FT/IR studies are used to determine the composition of the smoke gases in order to distinguish between the different types of biomass burning and to assess the impact of this type of process on the atmosphere. The IR analysis is performed both directly by field measurements using long-open-path spectrometers, and by laboratory studies of smoke samples that have been taken by cold trapping in canisters at the ground and at higher altitudes using light aircraft. The in situ measurements are carried out at high resolution (0.05 cm^{-1}) with a typical pathlength of 20–200 m. The samples are analyzed by matrix isolation FT/IR.

Next to the well known carbon, sulfur and nitrogen oxides, tens of other trace gases including CS_2, CSO, HCN, CH_3CN, and hydrocarbons have been identified and quantified. A spectrum, typical for a smoke sample, is shown in Figure 6.

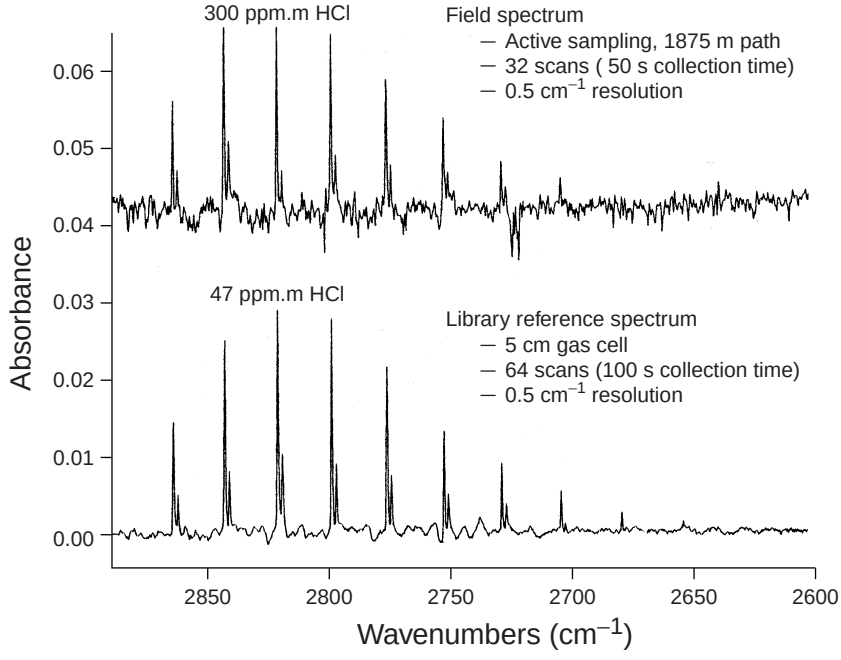

Figure 5 Reference IR spectrum and active field spectrum of HCl from a volcanic plume. Pathlength 1875 m, resolution 0.5 cm^{-1}, scanning time 50 s. (Reproduced by permission of Wiley-VCH from C.T. Chaffin, T.L. Marshall, W.G. Fateley, R.M. Hamaker, 'IR Analysis of Volcanic Plumes: A Case Study in the Application of Open Path FT/IR Monitoring Techniques', *Spectrosc. Eur.*, **7**, 20–24 (1995).)

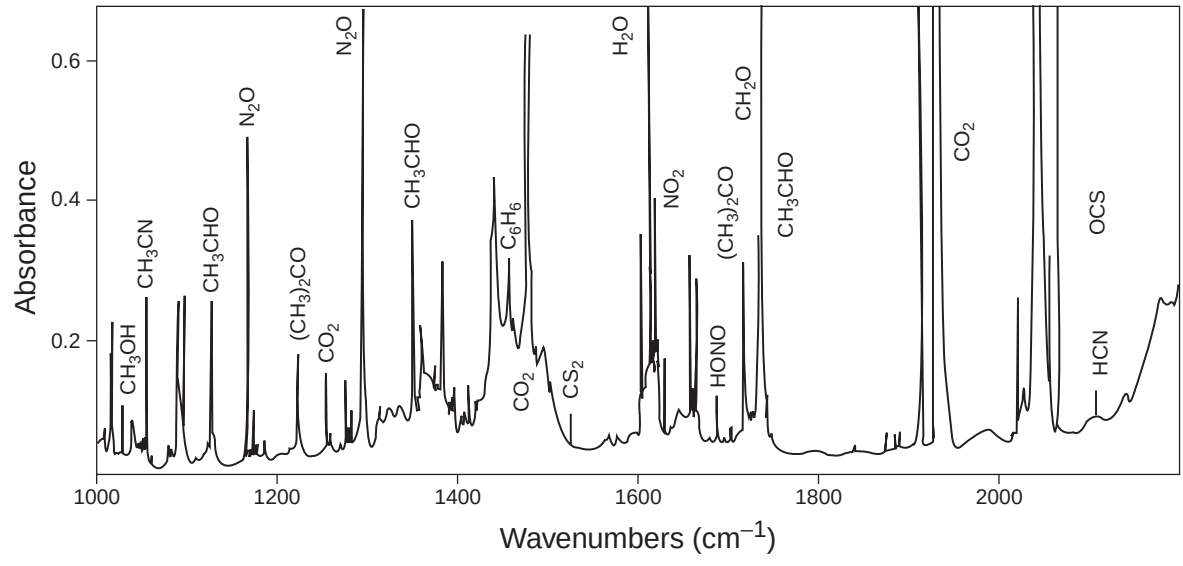

Figure 6 Matrix isolation FT/IR spectrum of a smoke sample from burning savanna grassland. (Reproduced by permission of The Society of Photo-optical Instrumentation Engineers from D.W.T. Griffith, 'FT/IR, Bushfires and Atmospheric Chemistry', *Proc. SPIE-Int. Soc. Opt. Eng.*, **1575**, 59–69 (1991).)

4.7 Volatile Organic Compounds

The relevance of the determination of VOCs in the indoor and outdoor atmosphere, and the IR spectroscopic techniques that can be used for this purpose, have already been described in sections 4.1 and 4.2. Obviously, the analysis of VOCs in other matrices is of similar importance. A rapid screening method to determine and quantify VOCs in soil has been developed by Clapper et al.[33] The principle of this method is the thermal desorption of the analytes from the soil sample with subsequent analysis of the vapor in a multipass gas cell

with a pathlength of 13.2 m. The sensitivity is at microgram level.

Fiber-optic analysis is a useful alternative when direct analysis of high concentrations of organic solvents in soil is required. As shown by Ewing et al.,[34] this technique is very selective for the identification of benzene, toluene, ethylbenzenes and xylenes (BTEX) as well as for the remote detection of high concentrations of trichloroethylene.[35]

Optical-fiber sensing is also used for the analysis of chlorinated hydrocarbons in water samples. The sensitivity of the sensor is increased by application of a polymer coating on the fiber surface. This type of analyte enrichment is analogous to the principle of SPME and enables the detection of chlorinated solvents such as tetrachloroethylene at a concentration level of 300 ppb.[36]

SPME has also been used for direct detection of VOCs in water by extraction of the organic compounds into a 130 μm thick parafilm.[37] After a certain equilibration time, the film is measured by IR transmission spectroscopy. Reproducible extraction and detection is achieved for VOCs such as benzene, toluene and chlorotoluene at concentration levels of 0.1–1 ppb.

Another technique for the determination of volatile pollutants in surface and wastewater is sparging-IR. This technique is based on the favorable partitioning of volatile and semivolatile organic molecules for the vapor phase over the liquid aqueous phase. Contaminants are forced into the vapor phase by a stream of nitrogen bubbles passing through the sample. Subsequently, the vapor is directed into a multipass gas cell for measurement of the target components. Sparging-IR is a multicomponent method with detection limits at the ppb level.[38]

4.8 Industrial Contaminants

Industrial processes and their products are a major source of environmental burdening. Production processes are generally accompanied by the emission of considerable amounts of waste chemicals, solvents and by-products. Indeed, the final products, especially chemical ones such as pesticides, pharmaceuticals, dyes, oils, plastics and paints, are obvious potential polluters. Finally, several of these compounds may give rise to the formation of metabolites and oxidation and degradation products.

Notorious examples of hazardous waste compounds are the polychlorinated benzenes (PCBs), dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs) because of their high toxicity and persistency. Other well-known industrial contaminants are VOCs, polyaromatic hydrocarbons (PAHs), phenols and (metabolites of) pesticides.

Many IR spectroscopic methods are available for the analysis of this large variety of compounds. The determination of VOCs has been discussed in the previous section. A selection of the most important other compound classes is presented in the following sections.

4.9 Pesticides

Several IR spectroscopic methods have been developed for the determination of pesticides and their metabolites in aqueous matrices. Because of the limited sensitivity of IR in relation to the required limits of detection (0.1 to 1 $\mu g\,L^{-1}$ typical), the methods usually comprise a pretreatment such as liquid–liquid extraction (LLE), solid-phase extraction (SPE), thermal desorption and derivatization.

The potentials of GC/IR for the determination of pesticides have been summarized by Kalasinsky.[39] Recent work on the combined GC/IR and GC/MS analysis of pesticides in surface and groundwaters from different locations in Lithuania has been carried out by Ellington et al.[40] Organic constituents were extracted from the water samples by SPE on cartridges. Subsequently, the cartridges were transported to the laboratory, eluted by an organic solvent and injected on the GC. As well as triazines and Alachlor, several chlorinated and aromatic carboxylic acids were determined.

A similar method for the analysis of microcontaminants in tap and surface water has been developed by Hankemeier et al.[41] After SPE extraction of 20 mL samples of water at site, the SPE cartridges are dried and brought to the lab. Next, the cartridges are desorbed on-line into the GC with 100 μL of ethylacetate using on-column large-volume injection interfacing. The IR detection is performed with a cryogenic sample storage GC/IR instrument. As can be seen from Figure 7(a), the Gram–Schmidt GC/IR chromatogram, reflecting the total IR absorption in the range 4000–6000 cm^{-1} as a function of time, permits detection and identification of all eluting components. Functional group GC/IR chromatograms (Figure 7b), show elution of compounds that absorb in a preselected wavenumber region and can be used to monitor the presence of compounds with a specific structural element.

A variety of pollutants have been identified at 0.1 to 1 ppb level, the current alert and alarm levels for micropollutants. As an example of the performance of this system, the IR spectra of the pesticides Sulfotep and Diazinon, extracted from the chromatograms recorded from 1 ppb in Figure 7, are shown in Figure 8.

The utility of solvent elimination LC/IR for the determination of pesticides in aqueous matrices has been studied extensively by Somsen et al.[42–44] Three different methods have been investigated for application to reversed-phase separations. The first method is an on-line post-column extraction of the analytes

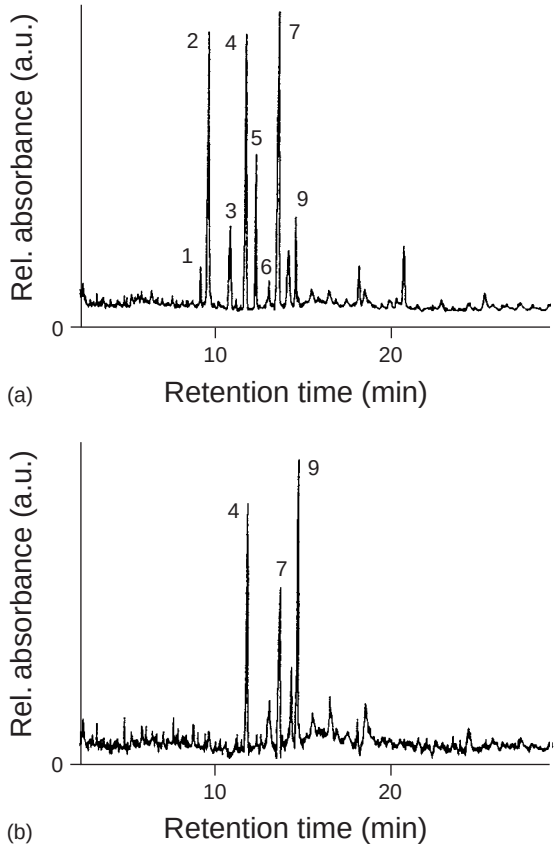

Figure 7 Gram–Schmidt GC/IR chromatogram (a) and functional group chromatogram of the region 1580–1520 cm^{-1} (b) obtained from on-line desorption of an SPE cartridge sampled with 20 mL tap water spiked at 1 $\mu g L^{-1}$. Peak 4: Atrazine, peak 7: Simetryn, peak 9: Metolachlor. (Reproduced from *J. Mol. Struct.*, **408/409**, T. Visser, M.J. Vredenbregt, A.P.J.M. de Jong, G.W. Somsen, Th. Hankemeier, N.H. Velthorst, U.A.Th. Brinkman, 'Recent Improvements in Environmental Trace Analysis by GC/IR and LC/IR', 97–103 (1997) by permission of Elsevier Science Ltd.)

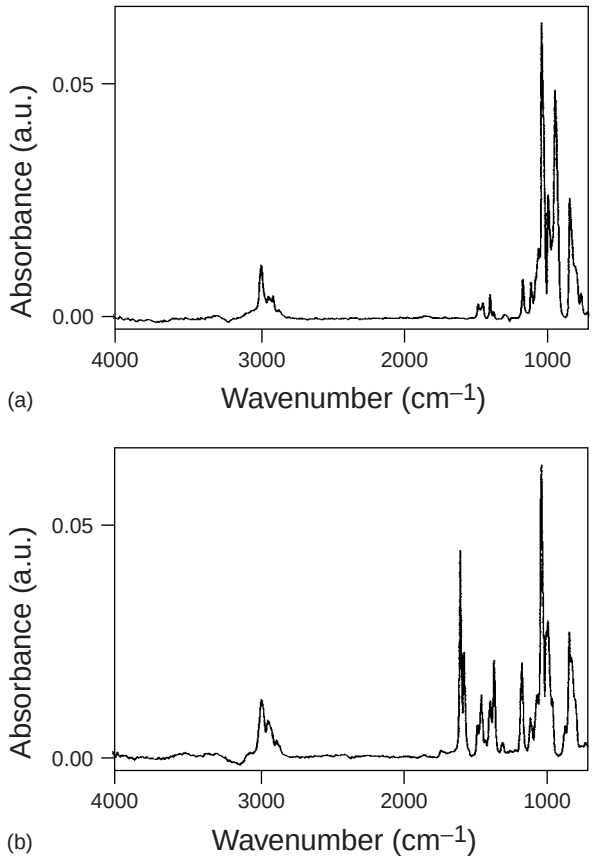

Figure 8 Cryotrapped GC/IR spectra of the insecticides (a) Sulfotep and (b) Diazinon obtained from SPE-large volume cryotrapping GC/IR of 20 mL of tap water spiked at 1 $\mu g L^{-1}$. (Reproduced from *J. Mol. Struct.*, **408/409**, T. Visser, M.J. Vredenbregt, A.P.J.M. de Jong, G.W. Somsen, Th. Hankemeier, N.H. Velthorst, U.A.Th. Brinkman, 'Recent Improvements in Environmental Trace Analysis by GC/IR and LC/IR', 97–103 (1997) by permission of Elsevier Science Ltd.)

from the buffered aqueous mobile phase into the more volatile dichloromethane to improve the performance of the spray-jet LC/IR interface.[42] This technique allows identification of phenylureas and quinones at a level of 50 $\mu g L^{-1}$. Improvement has been achieved by incorporation of a trace enrichment step using on-line SPE. With this method triazines and phenylurea herbicides can be identified at the alert level of 1 $\mu g L^{-1}$. The quality of the spectra obtained from the LC/IR analysis of a 100 mL sample of river Meuse water, spiked at 2 $\mu g L^{-1}$, is shown in Figure 9. For 20 mL injections, the analyte recovery is about 90% which indicates that this method can be used not only qualitatively but also for quantitative purposes.[43]

On-line trace enrichment combined with capillary column switching LC has proven to be a useful alternative for reversed-phase LC/IR when gradient elution separation is required. Post-column extraction cannot be applied because of the continuously changing composition of the mobile phase. However, these fluctuations are masked by addition of an excess of a volatile make-up liquid. The limit of identification of this method is about 20 $\mu g L^{-1}$ for triazines.[44]

Optical-fiber sensing is suited to the analysis of high concentrations of chlorinated hydrocarbons and pesticides in water samples. Analogous to SPME, analyte enrichment is accomplished by a thin polymer film on the fiber surface. The detection limit is about 2 ppm for the pesticides Alachlor and Atrazine.[45]

Finally, the same method of thermal desorption with subsequent measurement of the vapor phase, as applied for the determination of VOCs in soil, is applied for the determination of (semi)volatile pesticides when present in high concentrations.[33]

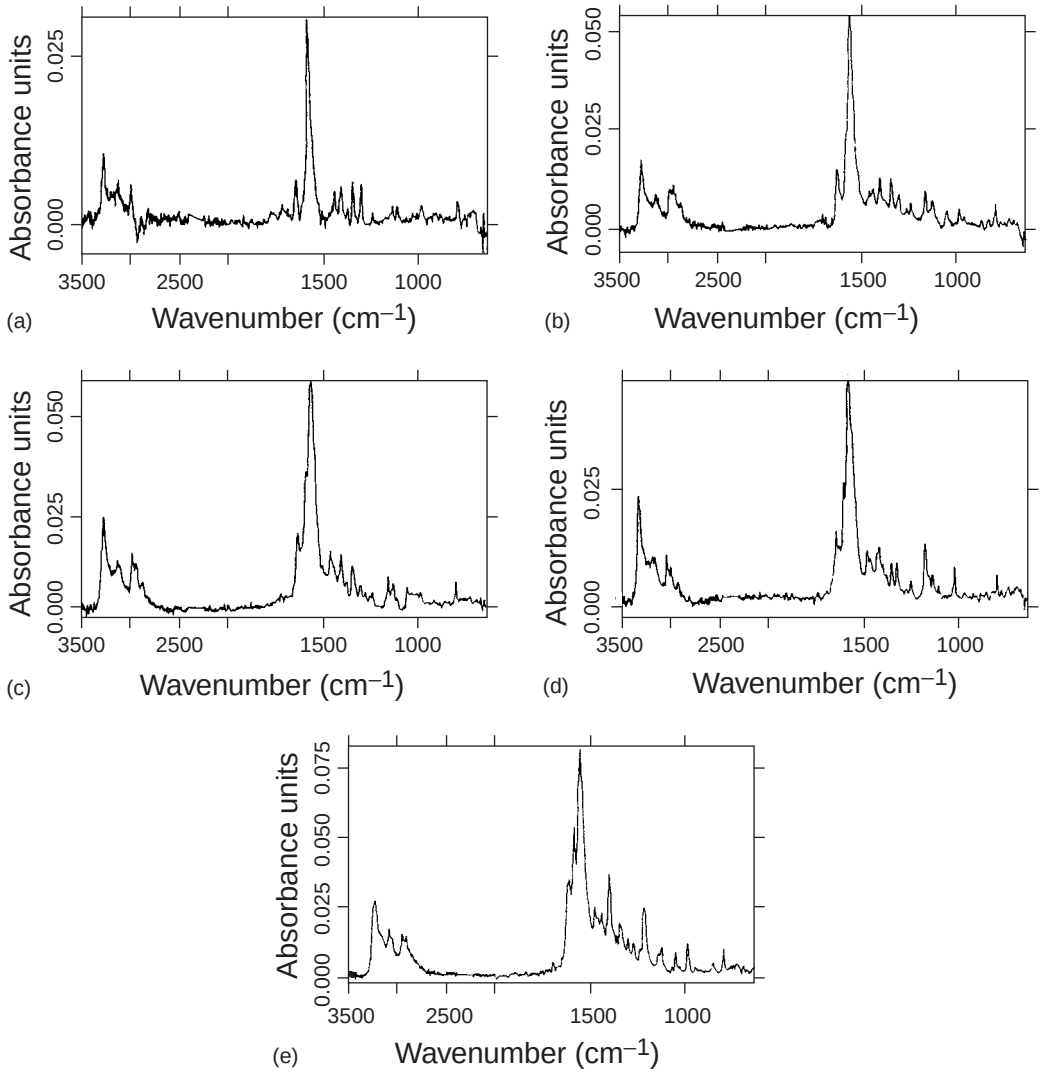

Figure 9 IR spectra of the triazines (a) Simazine, (b) Atrazine, (c) Sebutylazine, (d) Propazine and (e) Terbutylazine detected in river water by SPE on-line trace enrichment LC/IR analysis. Water sample: 100 mL, concentration level 2 μg L^{-1}. (Reproduced from *J. Chromatogr. A*, **756**, G.W. Somsen, I. Jagt, C. Gooijer, N.H. Velthorst, U.A.Th. Brinkman, T. Visser, 'Identification of Herbicides in River Water Using On-line Trace Enrichment Combined with Column Liquid Chromatography-Fourier-transform Infrared Spectroscopy', 145–157 (1996) by permission of Elsevier Science Ltd.)

4.10 Phenols

Phenols and, particularly, nitrated and chlorinated phenols are an important class of contaminants in aqueous matrices. The compounds are released by the chemical industry as by-products, but are also formed as result of degradation of insecticides and herbicides. As result of exposure to UV radiation from the sun, oxidation and isomerization of phenols easily occurs which implies that a large number of nitro- and chlorophenol isomers is found in air, particulate matter, aerosols, wastewater effluents, surface water and groundwater. Several of these phenols are considered as high-priority pollutants and for that reason unequivocal identification is demanded.

Since IR spectroscopy is extremely powerful in the differentiation between isomers, IR spectroscopic methods have been used for this purpose on several occasions. Vredenbregt et al.,[46] for example, applied cryotrapping GC/IR in addition to GC/MS, to distinguish between the ten molecular isomers of chloronitrophenol in surface water. Despite the absence of several reference standards, all isomers have been identified by interpretation of the complementary IR and MS data. The same sample storage GC/IR technique combined with GC/MS was used by Rodriguez et al.[47] to identify structural isomers of (poly)chlorinated phenol. After acetylation and preconcentration on graphitized carbon cartridges, 17 isomers have been identified.

Childers et al.[48] employed matrix isolation GC/IR for the characterization of nitrocresols in air sample extracts. These products are formed by photooxidation of NO_x and toluene in the atmosphere. A number of specific isomers have been identified by comparison with the GC/IR spectra of nitrocresol standards.

4.11 Polyaromatic Hydrocarbons

PAHs and their nitro derivatives are well established environmental contaminants which are invariably generated during combustion processes of fossil fuels. Unambiguous identification of PAHs and nitro-PAHs is important as the mutagenity and carcinogenity among these molecules may differ considerably. The identification and structure elucidation relies primarily on GC/MS. However, especially in the absence of reference compounds, MS is insufficiently specific to distinguish between the relatively large number of isomers. GC/IR is a useful method to discriminate between PAH congeners that are difficult to distinguish by other methods. The relevance of such analysis has been demonstrated by Childers et al.[49] and Kalasinsky et al.[50] in methods for detecting and identifying PAHs and nitro-PAHs in urban air particulate matter.

Other applications of GC/IR for the analysis of PAHs are the characterization of aromatic isomers in marine sediments from mildly polluted sites in the Arabian Gulf area[51] to endorse and confirm the results from GC/MS analysis and the confirmational analysis of PAHs in soil extracts in addition to a standard high-pressure liquid chromatography/ultraviolet (HPLC/UV)/fluorescence procedure.[52]

Raynor et al.[53] demonstrated that PAHs can also be determined adequately by capillary supercritical fluid extraction (SFE) combined with off-line FT/IR microspectrometry. Several PAHs have been identified in coal tar pitch samples after evaporation of the supercritical eluent and deposition of the separated analytes on a KBr disc.

Among the PAHs, the nonalterant cyclopentafused PAHs represent a unique class, being responsible for most of the genotoxicity of PAHs exhausts. Hence, their identification and the elucidation of their build-up mechanism is of major relevance. GC/IR has been used in addition to GC/MS for the identification of these PAHs and their intermediate PAHs. A cryogenic sample storage GC/IR interface has been applied to obtain sufficient sensitivity for subnanogram detection. Experiments have been carried out on standard compounds and on samples generated by in situ pyrolysis ($T = 600–1100\,^{\circ}C$) of ethynyl PAHs as model compounds for mimicking combustion processes. Conclusions from GC/MS have been confirmed and, additionally, the molecular structures of previously unidentified components have been elucidated.[54]

4.12 Dioxins, Dibenzofurans and Polychlorinated Biphenyls

PCDDs, PCDFs and PCBs are subject to considerable environmental interest due to the combination of their high toxicity and persistency. The three compound classes count 75, 135 and 209 congeners respectively, differing over six orders of magnitude in toxicity. It follows that highly sensitive univocal identification is required for these substances. For that reason, high-resolution GC/MS is the method of choice for this type of analysis.

Occasionally, however, additional IR spectrometric information is necessary to distinguish between the large number of possible molecular isomers, either for confirmational purposes in addition to GC/MS analysis, or for the identification of compounds that are to be used as analytical standards. As a consequence, many IR studies have been dedicated to the analysis of PCDDs, PCDFs and PCBs. Also, IR data are used to obtain information on the relationship between the toxicity and the geometrical structure of the molecules. Grainger and co-workers,[55,56] for instance, demonstrated that the IR vapor-phase spectra point to a near planar geometry for the highly toxic laterally substituted isomers, and a more tetrahedral type of geometry for the less toxic isomers in which the aromatic rings have lower electron withdrawing capacity. These conclusions are consistent with the data obtained from X-ray crystallography.

The importance of the identity conformation of standard materials for reference purposes has also been commonly acknowledged. Examples are the characterization of PCBs[57,58] and chlorinated and brominated dibenzodioxins and dibenzofuranes,[59] by matrix-isolation GC/IR, and the recording of the GC/IR reference spectra of all 209 individual PCB congeners.[60] The discriminative properties of IR spectrometry for this type of analyte are well illustrated by the spectra of 5 congeners of chlorinated and brominated dibenzodioxins in Figure 10.

Applications of IR spectrometry to the analysis of PCDDs, PCDFs and PCBs in real-life samples are scarce and restricted to samples with high concentrations. A recently published example is the GC/IR analysis of extracts of fly ash from a municipal waste incinerator to confirm the results from GC/MS.[61] PCDD- and PCDF-congeners have been identified unambiguously at high picogram level as result of the complementary information from both techniques.

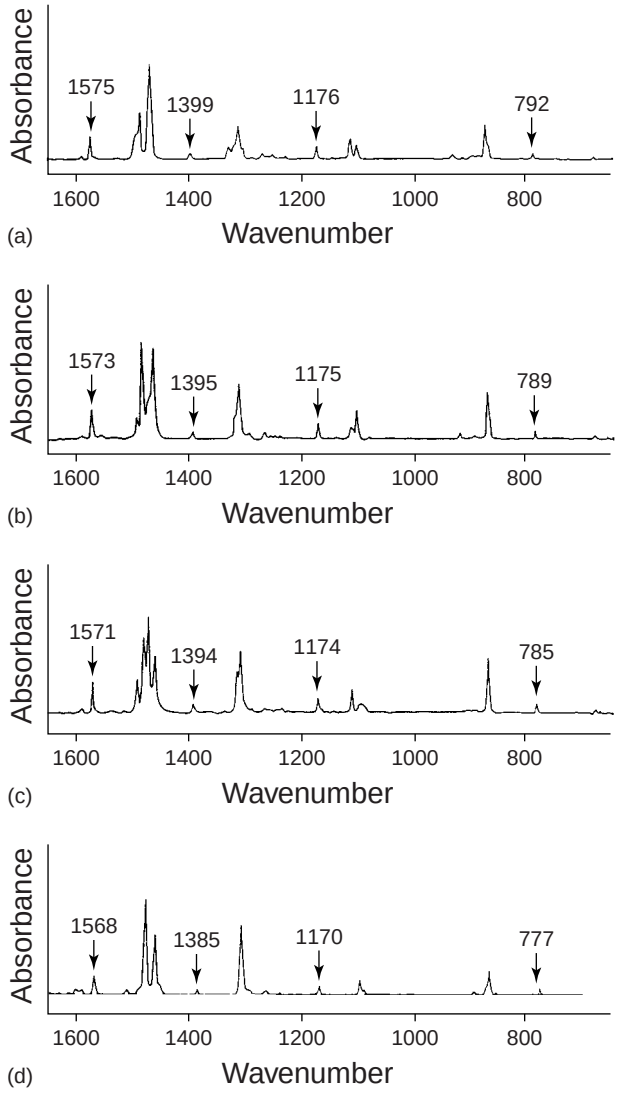

Figure 10 Matrix isolation GC/IR spectra of 2,3,7,8-substituted dibenzo-*p*-dioxins. (a) 2,3,7,8-tetrachloro-dibenzo-*p*-dioxin, (b) 2-bromo-3,7,8-trichloro-dibenzo-*p*-dioxin, (c) 2,3-dibromo-7,8-dichloro-dibenzo-*p*-dioxin and (d) 2,3,7,8-tetrabromo-dibenzo-*p*-dioxin. (Reproduced from *Chemosphere*, **25**, J.W. Childers, N.K. Wilson, R.L. Harless, R.K. Barbour, 'Characterization of Brominated and Bromo/chloro Dibenzo-*p*-dioxins and Dibenzofurans by Gas Chromatography/Matrix Isolation-Infrared Spectrometry', 1285–1290 (1992) by permission of Elsevier Science Ltd.)

4.13 Oils and Greases

For many years, the quantitative determination of mineral oil in water, soil and sediment samples by IR spectrometry has been a widely applied standard protocol for environmental and QC. The methods are based on the extraction of the hydrocarbons by an apolar solvent, followed by elimination of possibly co-extracted polar substances by means of Florisil, and subsequent measurement of the absorption maxima of the C–H

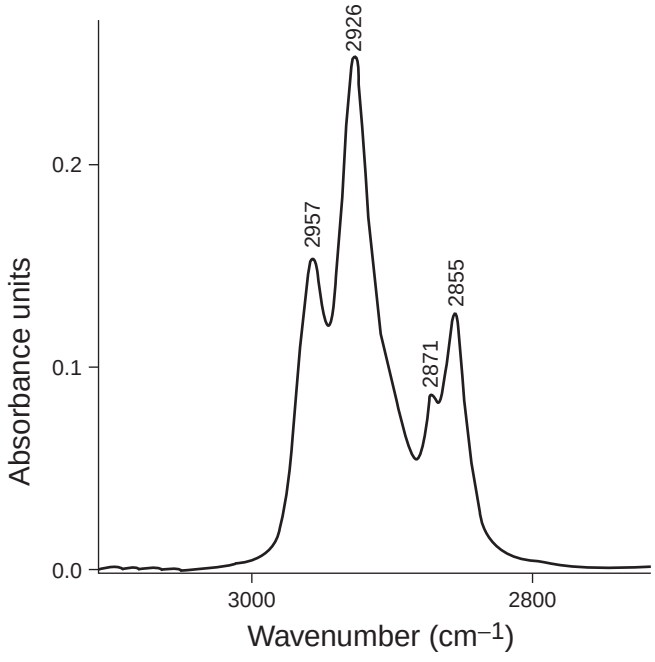

Figure 11 Characteristic absorption pattern of mineral oil extracted from a soil sample.

stretching bands in the extract. The limit of detection is in the order of 0.05 to 1 mg L^{-1}, depending on the extraction solvent and the pathlength of the liquid cell (1–5 cm typical). An example of the usual absorption pattern of a mineral oil is shown in Figure 11.

Up till now the extraction, prescribed in the protocols, has been carried out with halogenated hydrocarbons such as carbon tetrachloride and freon-113 as these solvents are IR transparent in the relevant wavenumber region 3200–2700 cm^{-1}. However, as result of the ozone-depleting effects of chlorinated fluorocarbons (CFCs), research is going on to alternative extraction methods without the use of these solvents. Several excellent alternatives for extraction and preconcentration have been proposed, such as adsorption filtration,[62] SPE, SFE, microwave-assisted extraction (MAE) and LLE in strongly reduced liquid volumes or in nonhalogenated solvents.[63] In parallel, ATR and DRIFT have been tested as alternative methods for the extinction measurement of the alkyl bands after evaporation of the solvent. The performances of several of the new methods are comparable or even better than the conventional ones with detection limits in the order of 0.005 ppb and a recovery of 90%.

Another alternative for the relatively elaborate extraction methods is direct screening by optical-fiber sensing. This in situ method is very fast but can only be applied for highly contaminated wet soils.[64]

4.14 Soil Characterization

IR spectrometry is commonly applied for the characterization of soil humic acid (HA) and fulvic acid (FA) fractions. HA and FA substances are complex organic macromolecules formed by degradation of biotic material such as cellulose, lignins, peptides, saccharides, etc. The large variety in the precursor structure combined with the many different parameters affecting the degradation process, results in an enormous diversity of the chemical and physical properties of humic substances. These properties play an important role in environmental analysis of xenobiotic compounds and for that reason the characterization is an important task. IR spectroscopy is a useful method for this purpose, particularly in the determination of the oxygen contents of humic and FAs. Information on the oxygen contents of both untreated as well as fractionized HAs and FAs can be derived from the position and the band pattern of the OH stretching bands around 3300 cm^{-1}, the strong absorbing symmetric and asymmetric C(=O)−O− stretching vibrations at 1600 and 1400 cm^{-1} and the −C−O−C bands in the region 1200–1000 cm^{-1}, which are correlated to the presence of hydroxylic, carboxylic and ether groups, respectively.[65,66] Examples of the IR spectra of peat HA fractions are given in Figure 12.

An IR method, closely related to characterization of HA fractions, is the analysis of residual fractions that remain after microbiological degradation of contaminated soil.[67] In this application, IR spectrometry is used to determine oxidation products of the original contaminants.

Another way to classify humic substances is by pyrolysis GC/IR.[68] With this technique, HA and FA fractions are characterized by the identity of the different pyrolysis products such as acids, lactones, alcohols and phenols, and their relative abundance.

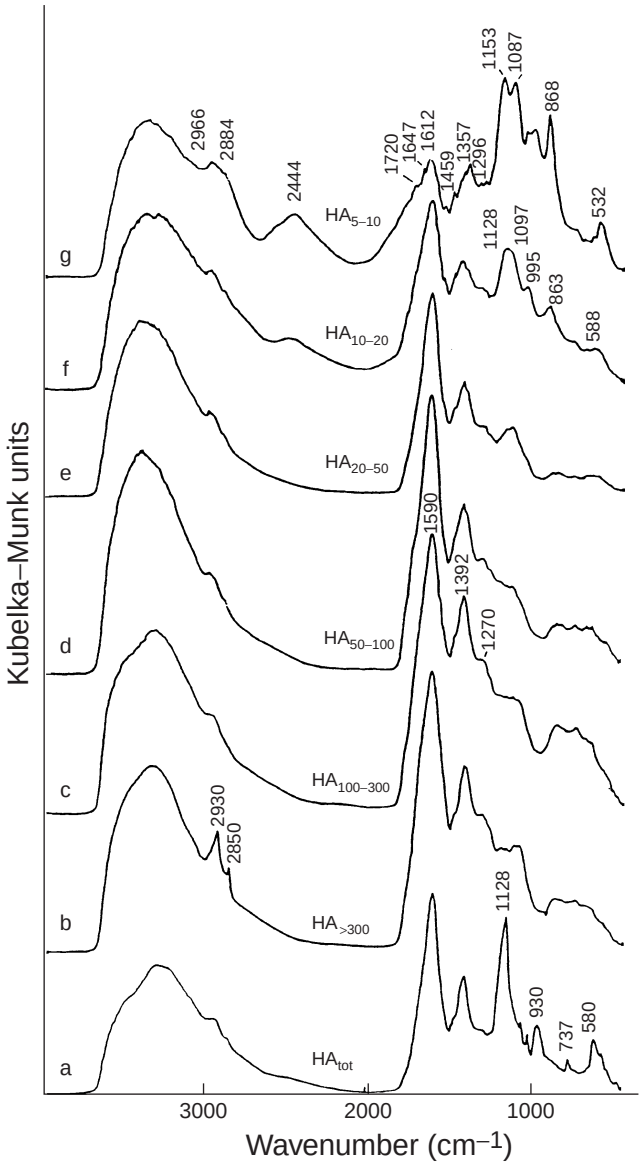

Figure 12 DRIFT spectra of peat HA. (Reproduced by permission of The Society for Applied Spectroscopy from O. Francioso, S. Sanchez-Cortes, V. Tugnoli, C. Ciavatta, L. Sitti, C. Gessa, 'IR, Raman and Nuclear Magnetic Resonance (^{1}H, ^{13}C and ^{31}P) Spectroscopy in the Study of Fractions of Peat HAs', *Appl. Spectrosc.*, **50**, 1165–1174 (1996).)

4.15 Inorganics and Asbestos

Occasionally, IR is used to determine the quantity of specific inorganics such as minerals and nitrates by means of conventional KBr measurements, DRIFT or photoacoustic IR.

A specific application is the analysis of asbestos (hydrated silica fibers) which is of concern owing to its assumed lung cancer causing properties. Although microscopy is usually the principal method for the determination of the type of asbestos, IR is often used for additional identification and confirmation. As demonstrated by Lang et al.,[69] IR can be used for the quantification as well. A method has been developed to quantify asbestos down to 0.01% by weight using an ashing technique to eliminate the cellulose matrix prior to absorption measurement.

4.16 Polymer Characterization

For obvious reasons, the recycling of sorted plastics has a number of advantages over unsorted regeneration. IR is highly suited for the identification of polymers and for that reason the application of IR for polymer recognition in waste industry is logical. At present several methods for automated plastic-waste identification are being employed. All applications are based on a fast, remote recognition of polymers in a stream of waste.[70]

Examples of IR classification of (co)polymers for recycling purposes are found in the automobile industry, the regeneration of electronic equipment,[71] the identification of waste carpet materials and in sorting plastics from municipal solid waste.[72] The latter combines the spectral IR and NIR information with additional geometric information by means of a focal-plane IR imaging system.

4.17 Quality Assurance

As already noted, IR spectrometry is commonly used for confirmational purposes and identity control of analytical standards prior to performing qualitative and quantitative analysis. Confirmational analysis is important since unambiguous identification of substances with suspected toxic properties is subject to stringent forensic and regulatory requirements. Identity control is relevant as the growing importance of good laboratory practice (GLP) has increased the prescription of IR for the QC of analytical standards.

In general, the analytical procedure and the criterion for confirmation are described into detail in a Standard Operating Procedure (SOP) in accordance with requirements as proposed by the International Organization for Standardization (ISO). Several criteria and algorithms have been developed to express the similarity between two spectra in a reliable and useful number. Nowadays, the software of virtually all IR instruments offers this possibility while also less stringent peak-matching requirements[73] as well as criteria based on neural networks,[74] have been proposed as decisive criteria to confirm the identity of a standard or an analyte.

4.18 Chemometrics and Model Studies

Nowadays, chemometrics plays a very important role in IR spectroscopy. The fact that FT/IR instruments are inherently equipped with fast processors and large storage capacity for data acquisition, processing and manipulation has made it considerably easier to use chemometric techniques like discriminant analysis (DA), principal component analysis (PCA), partial least squares (PLS) regression and artificial neural networks (ANN). On the other hand, the need for application of these techniques has been urged by two developments. Firstly, the current instrumentation produces increasing amounts of spectral data in shorter time and spectra are currently obtained within a tenth of a second. Secondly, the development of two- and three-dimensional mapping facilities and the coupling to GC and LC, has manifolded the number of spectra to be handled. In order to facilitate or even admit the interpretation of this enormous amount of information, chemometrics are practically indispensable and have, therefore, become common sense in modern IR spectroscopy. Data reduction and data storage methods are applied to reduce the large amount of data to feasible proportions in terms of desk and memory space while multivariate calibration techniques have been incorporated into the standard instrument software for data manipulation and simplification of the interpretation. Besides, expert systems have been developed to enable unskilled spectroscopists to take advantage of unknown possibilities of IR spectrometry. Examples are the chemometrical systems that are used to assist in the structure elucidation of unknowns.[75,76]

Model studies are a valuable method to predict the course of processes when only a limited number of data is available or in case extended studies cannot be performed for practical reasons. Particularly, in large-scale processes such as environmental exposure and risk assessment, computerized models are indispensable. Several models, partly or fully based on IR spectroscopic data, are currently used for this purpose. An important additional advantage of IR spectrometry is the possibility to perform real-time and in situ measurements which enables the development and the validation of a model not only in the laboratory but also in real-life situations.

A good example are the computerized models that are used in studies of the greenhouse effect. These models take into account a variety of parameters such as the atmospheric and earth temperature, cloud cover, geographic coordinates, weather conditions and the composition of the atmosphere as a function of the altitude. Obviously, IR remote sensing is very well suited to provide the required qualitative and quantitative data of greenhouse and other gases. Moreover, the same technique can be used for the validation by monitoring the actual atmospheric composition and (photo)chemical processes.

Specific examples of model studies based on IR data are of those of Collete[77] and Nelson et al.[78] The first one applied PLS modeling of IR spectral data, to predict chemical and physical constants and distribution coefficients of a number of pollutants in order to estimate their environmental fate. The second one carried out a model study to determine the parameters that affect the formation of toxic constituents during incineration of wastes.

5 CURRENT POSITION AND FUTURE DEVELOPMENTS

At present, IR spectroscopy is a viable method for many applications in environmental analysis. The merits of the technique lie in its versatility and in the unique fingerprinting properties of the spectra. The first makes

IR the principal method for remote sensing and in situ measurements as illustrated by the applications in atmospheric studies and the classification of polymers. The second makes it, in principle, the method of choice for unambiguous identification. In trace analysis, however, mass spectrometry is superior in sensitivity and ease of operation and for that reason the role of IR is mainly restricted to QC and confirmational purposes.

Future applications should be seen in relation to developments in hard and software. The recent introduction of diode array detection in IR spectrometry is a good example of the ongoing attempts at innovation. Further development of this type of detection can be expected. Research will also focus on improvement of the sensitivity by means of dedicated detectors and tunable IR laser technology. At short notice, the ongoing demand for fast and robust methods will lead to the further development and optimization of dedicated, low-cost sensor type of instruments, particularly suited for routine and field measurements. As such, the growing interest in NIR spectrometry for routine tasks, currently performed with mid-IR techniques, will most likely continue.

Finally, it should be noted that the improvement of IR technology will be attended by similar developments of other analytical techniques. For that reason, the choice for application of an IR spectroscopic method has always to be judged in relation to the potential of alternatives which includes economical and practical considerations such as the availability of instrumentation and expertise.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
ANN	Artificial Neural Networks
ATMOS	Atmospheric Trace Molecule Spectroscopy
ATR	Attenuated Total Reflection
BTEX	Benzene, Toluene, Ethylbenzenes, and Xylenes
CFC	Chlorinated Fluorocarbon
CFC	Chlorofluorocarbon
DA	Discriminant Analysis
DRIFT	Diffuse Reflectance Infrared
FA	Fulvic Acid
FT	Fourier Transform
FT/IR	Fourier Transform/Infrared
GC	Gas Chromatography
GC/IR	Gas Chromatography/Infrared
GC/IR/MS	Gas Chromatography/Infrared/Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
GLP	Good Laboratory Practice
HA	Humic Acid
HPLC/UV	High-pressure Liquid Chromatography/Ultraviolet
IGC	Infrared Gas Cloud
IR	Infrared
ISO	International Organization for Standardization
LC	Liquid Chromatography
LC/IR	Liquid Chromatography/Infrared
LLE	Liquid–Liquid Extraction
MAE	Microwave-assisted Extraction
mid-IR	Mid-infrared
NIR	Near-infrared
PAH	Polyaromatic Hydrocarbon
PAS	Photo-acoustic Spectrometry
PCA	Principal Component Analysis
PCB	Polychlorinated Benzene
PCDD	Dibenzo-*p*-dioxin
PCDF	Dibenzofuran
PLS	Partial Least Squares
QA	Quality Assurance
QC	Quality Control
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SOP	Standard Operating Procedure
SPE	Solid Phase Extraction
SPME	Solid Phase Micro Extraction
UV	Ultraviolet
VOC	Volatile Organic Compound

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Chemical Weapons Chemicals Analysis **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Automotive Emissions Analysis with Spectroscopic Techniques • Differential Optical Absorption Spectroscopy, Air Monitoring by • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by

Tunable Mid-infrared Diode • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments • Photoacoustic Spectroscopy in Trace Gas Monitoring

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Flow-injection Techniques in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Industrial Hygiene **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Fuel Performance Specifications, Mid-infrared Analysis of

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymers and Rubbers

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Satellite and Sensor Systems for Environmental Monitoring

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Gas Chromatography/Infrared Spectroscopy • Liquid Chromatography/Infrared Spectroscopy • Microspectroscopy

REFERENCES

1. M. Simonds, H. Xiao, S.P. Levine, 'Optical Remote Sensing for Air Pollutants – Review', *Am. Ind. Hyg. Assoc. J.*, **55**, 953 – 965 (1994).
2. I. Ahonen, H. Riipinen, A. Roos, 'Portable Fourier Transform Infrared Spectrometer for Use as a Gas Analyzer in Industrial Hygiene', *Analyst*, **121**, 1253 – 1255 (1996).
3. M. Saito, K. Kikuchi, 'Infrared Optical Fiber Sensors', *Opt. Rev.*, **4**, 527 – 538 (1997).
4. J.E. Walsh, B.D. MacCraith, M. Meaney, J.G. Vos, F. Regan, A. Lancia, S. Artjushenko, 'Sensing of Chlorinated Hydrocarbons and Pesticides in Water using Polymer Coated Mid-infrared Optical Fibers', *Analyst*, **121**, 789 – 792 (1996).
5. R.G. Messerschmidt, M.A. Harthcock, 'Infrared Microscopy, Theory and Applications', (Practical spectroscopy), Marcel Dekker, New York, Vol. 6, 1988.
6. P. Jackson, G. Dent, D. Carter, D. Schofield, J. Chalmers, T. Visser, M.J. Vredenbregt, 'Investigation of High Sensitivity GC/FT/IR as an Analytical Tool for Structural Identification', *J. High Resolut. Chromatogr.*, **16**, 515 (1993).
7. D.F. Gurka, S.M. Pyle, I. Farnham, R. Titus, 'Application of Hyphenated Fourier Transform Infrared Techniques to Environmental Analysis', *J. Chromatogr. Sci.*, **29**, 339 – 344 (1991).
8. K.A. Krock, C.L. Wilkins, 'Qualitative Analysis of Contaminated Environmental Extracts by Multidimensional Gas Chromatography with Infrared and Mass Spectral Detection (MDGC/IR/MS)', *J. Chromatogr. A*, **726**, 167 – 178 (1996).
9. N. Ragunathan, K.A. Krock, C. Clawun, T.A. Sasaki, C.L. Wilkins, 'Review – Multispectral Detection for Gas Chromatography', *J. Chromatogr. A*, **703**, 393 – 416 (1995).
10. G.W. Somsen, C. Gooijer, N.H. Velthorst, U.A.Th. Brinkman, 'Coupling of Column Liquid Chromatography and Fourier-transform Infrared Spectrometry: An Overview', *J. Chromatogr. A*, **811**, 1 – 34 (1998).
11. M.L. McKelvy, T.R. Britt, B.L. Davies, J.K. Gillie, L.A. Lentz, A. Leugers, R.A. Nyquist, C.L. Putzig, 'Infrared Spectroscopy', *Anal. Chem.*, **68**, 93R – 160R (1996).
12. R.E. Clement, P.W. Yang, 'Environmental Analysis', *Anal. Chem.*, **69**, 251R – 287R (1997).
13. V.D. Makepeace, C.W. Chase, C.T. Chaffin, T.L. Marshall, P.T. Jaakkola, R.M. Hoffman, R.M. Hamaker, W.G. Fateley, 'Monitoring Indoor Air Quality by Extractive FT/IR Spectrometry', *Mikrochim. Acta*, [Suppl.] **14**, 563 – 564 (1997).
14. W.M. ter Kuile, J.J.F. van Veen, B. Knoll, 'Design and Use of the Infrared Gas Cloud Scanner for Measurement and Imaging of the Spatial Distribution of Gases at Workplaces', *Proc. SPIE-Int. Soc. Opt. Eng.*, **2366**, 135 – 146 (1995).
15. S.K. Cole, P. Martin, 'Determination of Gas-phase Side Stream Cigarette Smoke Components Using Fourier Transform Infrared Spectrometry', *Analyst*, **121**, 495 – 500 (1996).
16. E.S. Larsen, W.H. Hong, M.L. Spartz, 'Hydrogen Sulfide Detection by UV-assisted Infrared Spectrometry', *Appl. Spectrosc.*, **51**, 1656 – 1663 (1997).
17. M. Fillipelli, 'Methylmercury Determination by Purge and Trap-GC/FT/IR/AAS after $NaBH_4$ Derivatization of an Environmental Thiosulfate Extract', *Appl. Organomet. Chem.*, **8**, 687 – 691 (1994).

18. S.P. Levine, G.M. Russwurm, 'Fourier Transform Infrared Optical Remote Sensing for Monitoring Airborne Gas and Vapor Contaminants in the Field.', *Trends Anal. Chem.*, **13**, 258–262 (1994).
19. C.B. Farmer, 'Stratospheric FTS', *Mikrochim. Acta*, [Suppl.] **14**, 89–101 (1997).
20. C. Camy-Peyret, 'Balloon Borne Infrared Fourier Transform Spectroscopy for Measurement of Atmospheric Trace Species', *Spectrochim. Acta*, **51A**, 1143–1152 (1995).
21. E. Mahieu, R. Zander, L. Delbouille, P. Demoulin, 'Observed Trends in Total Vertical Column Abundances of Atmospheric Gases from IR Solar Spectra Recorded at the Jungfraujoch', *J. Atmosph. Chem.*, **28**, 227–243 (1997).
22. J. Notholt, G. Toon, F. Stordal, S. Solberg, N. Schmidtbauer, A. Meier, E. Becker, B. Sen, 'Seasonal Variations of Atmospheric Trace Gases in the High Arctic at 79 °N', *J. Geophys. Res.*, **102**, 12855–12861 (1997).
23. N. Davies, M. Hilton, A.H. Lettington, 'Vehicle Exhaust Gas Monitoring using Remote Absorption and Emission FT/IR Spectroscopy', *Mikrochim. Acta*, [Suppl.] **14**, 551–553 (1997).
24. E. Lindermeir, 'Emission Mode Fourier Spectroscopy of Smokestack Exhaust Gas', *Mikrochim. Acta*, [Suppl.] **14**, 559–561 (1997).
25. A. Sjödin, M. Lenner, 'On-road Measurements of Single Vehicle Pollutants Emissions, Speed and Acceleration for Large Fleets of Vehicles in Different Traffic Environments', *Sc. Total Environm.*, **169**, 157–165 (1995).
26. P. Haschberger, 'Airborne FT/IR Emission Spectrometer for Trace Gas Monitoring', *Mikrochim. Acta*, [Suppl.] **14**, 793–795 (1997).
27. K. Shafer, R. Haus, J. Heland, A. Haak, 'Measurements of Atmospheric Trace Gases by Emission and Absorption Spectroscopy With FT/IR', *Ber. Bunsen-Ges. Phys. Chem.*, **99**, 405–411 (1995).
28. J. Steiger, T. Hoffmann, J. Kjahl, D. Klockow, E.H. Korte, W. Schrader, 'Using Thermodesorption-GC/Cryocondensation FT/IR for the Measurement of Biogenic VOC Emissions', *Fresenius'. J. Anal. Chem.*, **362**, 148–154 (1998).
29. S.N. Pandis, S.E. Paulson, J.H. Seinfeld, R.C. Flagan, 'Fourier Transform Infrared Analysis of Aerosol Formed in the Photo-oxidation of Isoprene and β-Pinene', *Atmosph. Environm.*, **26**, 1239–1251 (1992).
30. J. Schmitt, H.-C. Flemming, F. Mertens, 'FT/IR-Spectroscopy – a New Technique for Drinking Water Microbial Analysis'. *Biospectrum*, **2**, 48–49 (1996).
31. C.T. Chaffin, T.L. Marshall, W.G. Fateley, R.M. Hamaker, 'Infrared Analysis of Volcanic Plumes: A Case Study in the Application of Open Path FT/IR Monitoring Techniques', *Spectrosc. Eur.*, **7**, 20–24 (1995).
32. D.W.T. Griffith, 'FT/IR, Bushfires and Atmospheric Chemistry', *Proc. SPIE-Int. Soc. Opt. Eng.*, **1575**, 59–69 (1991).
33. M. Clapper, J. Demergian, G. Robitaille, 'A Quantitative Method Using FT/IR to Detect Explosives and Selected Semi-volatiles in Soil Samples', *Spectrosc.*, **10**, 44–49 (1995).
34. K.J. Ewing, G. Nau, F. Bucholtz, I.D. Aggerwal, G. Robertille, 'Fiber Optic Infrared Reflectance Sensor for Detection of Solvent in Soil, Field Screening Methods Hazard'. *Wastes Toxic Chem., Proc. Int. Symp.*, **1**, 144–150 (1995).
35. S.T. Vohra, F. Bucholtz, G.M. Nau, K.J. Ewing, I.D. Aggerwal, 'Remote Detection of Trichloro-ethylene in Soil by a Fiber-optic Infrared Reflectance Probe', *Appl. Spectrosc.*, **50**, 985–990 (1996).
36. M. Jakusch, B. Miziakoff, R. Kellner, A. Katzir, 'Towards a Remote IR-fiber Optic Sensor System for the Determination of Chlorinated Hydrocarbons in Water', *Sens. Actuators B*, **B38**, 83–87 (1997).
37. D.L. Heglund, D.C. Tilotta, 'Determination of Volatile Organic Compounds in Water by Solid Phase Microextraction and Infrared Spectroscopy', *Environ. Sci. Technol.*, **30**, 1212–1219 (1996).
38. W.M. Doyle, 'Two Dimensional Sparging-IR Analysis of Trace Organics is Water', *Proc. SPIE-Int. Soc. Opt. Eng.*, **2089**, 456–457 (1995).
39. K.S. Kalasinsky, 'Pesticide Determination by GC/FT/IR', in *Analytical Methods for Pesticides and Plant Growth Regulators*, Vol. XVII, Chap. 4 Academic Press, New York, 1989.
40. J.J. Ellington, R.V. Thurston, J. Sukyte, K. Kvietkus, 'Hazardous Chemicals in Waters of Lithuania', *Trends Anal. Chem.*, **15**, 215–224 (1996).
41. Th. Hankemeier, E. Hooijschuur, R. Vreuls, U.A.Th. Brinkman, T. Visser, 'On-line SPE-GC-cryotrapping-IR for the Trace Level Determination of Microcontaminants in Aqueous Samples', *J. High Resol. Chromatogr.*, **21**, 341–346 (1998).
42. G.W. Somsen, E.W.J. Hooijschuur, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, T. Visser, 'Coupling of Reversed-phase Liquid Chromatography and Fourier Transform Infrared Spectrometry Using Post-column On-line Extraction and Solvent Elimination', *Anal. Chem.*, **68**, 746–752 (1996).
43. G.W. Somsen, I. Jagt, C. Gooijer, N.H. Velthorst, U.A. Th. Brinkman, T. Visser, 'Identification of Herbicides In River Water Using On-line Trace Enrichment Combined with Column Liquid Chromatography-Fourier-Transform Infrared Spectroscopy', *J. Chromatogr. A*, **756**, 145–157 (1996).
44. T. Visser, M.J. Vredenbregt, G.J. ten Hove, A.P.J.M. de Jong, G.W. Somsen, 'Gradient Elution Liquid Chromatography/Infrared Spectrometry at $\mu g\ L^{-1}$ Level using Capillary Column Switching and Addition of a Make-up Liquid. A Preliminary Study', *Anal. Chim. Acta*, **342**, 151 (1997).
45. F. Regan, M. Meaney, J.G. Vos, B.D. MacCraith, J.E. Walsh, 'Determination of Pesticides in Water Using

ATR/FT/IR Spectroscopy on PVC/Chloroparaffin Coatings', *Anal. Chim. Acta*, **334**, 85–92 (1996).
46. M.J. Vredenbregt, H.A. den Hollander T. Visser, 'Isomer Differentiation and Identification of Chloro-nitrophenols by Means of Cryotrapping GC/FT/IR', *Proc. SPIE-Int. Soc. Opt. Eng.*, **1575**, 474–475 (1991).
47. I. Rodriguez, M.H. Bollain, C.M. Garcia, R. Cela, 'Analysis of Structural Isomers of Polychlorinated Phenols in Water by Liquid Nitrogen-trapping Gas Chromatography-Fourier Transform Infrared Spectroscopy', *J. Chromatogr. A*, **733**, 405–416 (1996).
48. J.W. Childers, D.F. Smith, N.K. Wilson, R.K. Barbour, 'Analysis of Nitrocresols and Related Compounds by Gas Chromatography/Matrix Isolation Infrared Spectrometry', *Appl. Spectrosc.*, **49**, 286–294 (1995).
49. J. Childers, N.K. Wilson, R.K. Barbour, 'Gas Chromatography/Matrix Isolation Infrared Spectrometry for the Identification of Polycyclic Aromatic Hydrocarbons in Urban Particulate Matter', *Appl. Spectrosc.*, **43**, 1344–1349 (1989).
50. V.F. Kalasinsky, C. Saiwan, K.G. Whitehead, 'GC/FT/IR Study of Nitrated Polycyclic Aromatic Hydrocarbons', *J. Chrom. Science*, **26**, 584–587 (1988).
51. M. Guilliano, P. Doumenq, A. Jawad, G. Mille, 'Structural Characterization of Aromatic Sediments by Combined Gas Chromatography/Fourier Transform Infrared Spectroscopy', *Appl. Spectrosc.*, **43**, 571–573 (1989).
52. T. Visser, M.J. Vredenbergt, A.P.M.J. de Jong, 'Confirmational Analysis of Polycyclic Aromatic Hydrocarbons in Soil Extracts by Cryotrapping GC/FT/IR', *J. Chromatogr. A*, **687**, 303–313 (1994).
53. M.W. Raynor, I.L. Davies, K.D. Bartle, A.A. Clifford, A. Williams, J.M. Chalmers, B.W. Cook, 'Supercritical Fluid Extraction/Capillary Supercritical Fluid Chromatography/Fourier Transform Infrared Microspectrometry of Polycyclic Aromatic Compounds in a Coal Tar Pitch', *J. High. Resol. Chromatogr.*, **11**, 766–775 (1988).
54. T. Visser, M. Sarobe, L.W. Jenneskens, J.W. Wesseling, 'Identification of Isomeric Polycyclic Aromatic Hydrocarbons (PAH) in Pyrolysates from Ethynylated PAH by Gas Chromatography/Fourier Transform Infrared Spectroscopy. Their Relevance for the Understanding of PAH Rearrangement and Interconversion Processes During Combustion', Fuel, **9/10**, 913–920 (1998).
55. J. Grainger, V.V. Reddy, D.G. Patterson, 'Molecular Geometry by Approximations for Chlorinated Dibenzodioxins by Fourier Transform Infrared Spectroscopy', *Appl. Spectrosc.*, **42**, 643–655 (1988).
56. J. Grainger, V.V. Reddy, D.G. Patterson, 'Hexachlorodibenzo-*p*-dioxin Isomer Differentiation by Capillary Gas Chromatography Fourier Transform Infrared Spectroscopy', *Appl. Spectrosc.*, **44**, 41–46 (1990).
57. D.M. Hembree, N.R. Smyrl, W.E. Davis, D.M. Williams, 'Isomeric Characterization of Polychlorinated Biphenyls Using Gas Chromatography/Fourier Transform Infrared/Gas Chromatography/Mass Spectrometry', *Analyst*, **118**, 249–252 (1993).
58. R. Fuoco, P.R. Griffiths, 'Analysis of Polychlorobiphenyls (PCBs) in Contaminated Soil Samples by Supercritical Fluid Extraction/Supercritical Fluid Chromatography Coupled With FT/IR Spectrometry', *Ann. Chim. (Rome)*, **82**, 235–245 (1992).
59. J.W. Childers, N.K. Wilson, R.L. Harless, R.K. Barbour, 'Characterization of Brominated and Bromo/chloro dibenzo-*p*-dioxins and Dibenzofurans by Gas Chromatography/Matrix Isolation-Infrared Spectrometry', *Chemosphere*, **25**, 1285–1290 (1992).
60. B. Bush, E.L. Barnard, 'Gas Phase Infrared Spectra of 209 Polychlorinated Biphenyl Congeners Using Gas Chromatography with Fourier Transform Infrared Detection: Internal Standardization with a ^{13}C-labelled Congener', *Arch. Environ. Contam. Toxicol.*, **29**, 322–326 (1995).
61. S. Sommer, R. Kamps, S. Schumm, K. Kleinermanns, 'GC/FT-IR/MS Spectroscopy of Native Polychlorinated Dibenzo-*p*-dioxins and Dibenzofurans Extracted from Municipal Fly-ash', *Anal. Chem.*, **69**, 1113–1118 (1997).
62. H. Zou, Q. Jiang, W. Liang, Z. Zhang, S. Zou, 'Determination of Trace Amounts of Water by Adsorption Filtration', *Analyst*, **122**, 945–948 (1997).
63. Y. Daghbouche, S. Garrigues, A. Morales-Rubio, M. de la Guardia, 'Evaluation of Extraction Alternatives for Fourier Transform Infrared Spectrometric Determination of Oil and Greases in Water', *Anal. Chim. Acta*, **345**, 161–171 (1997).
64. G. Hazel, F. Bucholtz, I.D. Aggarwal, G. Nau, K.J. Ewing, 'Multivariate Analysis of Mid IR FT/IR Spectra of Hydrocarbon Contaminated Wet Soils', *Appl. Spectrosc.*, **51**, 984–989 (1997).
65. O. Francioso, S. Sanchez-Cortes, V. Tugnoli, C. Ciavatta, L. Sitti, C. Gessa, 'Infrared, Raman and Nuclear Magnetic Resonance (^{1}H, ^{13}C and ^{31}P) Spectroscopy in the Study of Fractions of Peat Humic Acids', *Appl. Spectrosc.*, **50**, 1165–1174 (1996).
66. O. Francioso, S. Sanchez-Cortes, V. Tugnoli, C. Ciavatta, C. Gessa, 'Characterization of Peat Fulvic Acid Fractions by means of FT/IR, SERS and ^{1}H, ^{13}C NMR Spectroscopy', *Appl. Spectroc.*, **52**, 270–277 (1998).
67. V. Riis, D. Miethe, M. Moder, 'Analytical characterization of the Persistent Residues after Microbial Degradation of Mineral Oils', *Fres. Zeitschr. Anal. Chem.*, **356**, 378–384 (1996).
68. R. Kuckuk, W. Hill, P. Burba, A.N. Davies, *Fresenius' J. Anal. Chem.*, **350**, 528–532 (1994).
69. P.L. Lang, V.V. Chu, K. McCune, S. Franssen, M. Goodnight–Schmidt, R. Mendenhal, W. Coner, 'Quantitation of Asbestos Using Diffuse Reflection Infrared Spectroscopy', *Appl. Spectrosc.*, **52**, 212–217 (1998).
70. G. Zachmann, 'A Rapid and Dependable Identification System for Black Polymer Materials', *J. Mol. Struct.*, **348**, 453–456 (1995).

71. J. Jansen, C. van Hastenberg, 'Identifikation von Kunststofteilen', *Kunststoffe*, **84**, 51–54 (1994).
72. W.H.A.M. van de Broek, D. Wienke, W.J. Melssen, R. Feldhoff, T. Huth-Fehre, T. Kantimm, L.M.C. Buydens, 'Application of a Spectroscopic Infrared Focal Plane Array Sensor for On-line-Identification of Plastic Waste', *Appl. Spectrosc.*, **51**, 856–865 (1996).
73. W.G. de Ruig, R.W. Stephany, G. Dijkstra, 'Criteria for the Detection of Analytes in Test Samples', *J. Assoc. Off. Anal. Chem.*, **72**, 487–492 (1989).
74. H.J. Luinge, E.D. Leussink, T. Visser, 'Trace Level Identity Confirmation from Infrared Spectra by Library Searching and Artificial Neural Networks', *Anal. Chim. Acta*, **345**, 173–184 (1997).
75. M. Cadisch, E. Pretsch, 'Spec-TOOL: A Knowledge-based Hypermedia System for Interpreting Molecular Spectra', *Fresenius'. J. Anal. Chem.*, **344**, 173–183 (1992).
76. G.N. Andreev, O.K. Argirov, P.N. Penchev, 'Expert System for the Interpretation of Infrared Spectra', *Anal. Chim. Acta*, **284**, 131–140 (1993).
77. T.W. Collete, 'Predicting Environmental Fate Parameters with Infrared Spectroscopy', *Trends Anal. Chem.*, **16**, 24–36 (1997).
78. C.M. Nelson, J.D. Smith, R.D. VanDell, A.S. Bonnano, P.R. Solomon, 'Comparison of Extractive and In-situ FT/IR Measurements during Incineration of Chlorinated Hydrocarbons', *Proc. SPIE-Int. Soc. Opt. Eng.*, **2883**, 628–639 (1996).

Inorganic Analysis in Environmental Samples by Capillary Electrophoresis

Yangsun Kim
Semyoung University, South Korea

1 Introduction **1**
2 Basic Principles **2**
2.1 Capillary Electrophoresis System and Operating Parameters 2
2.2 Electrolyte Condition 3
2.3 Detection System 4
3 Selectivity Control by Complex Formation **6**
3.1 Partial Complexation 6
3.2 Formation of Inclusion Complexes 7
3.3 Complete Complexation 8
4 Manipulating Sensitivity **8**
4.1 Use of Detection Cells with Extended Optical Pathlength 8
4.2 On-column Preconcentration by Sample Stacking 9
4.3 On-line Preconcentration by Isotachophoresis 9
4.4 Preconcentration by Solid Phase Extraction 10
4.5 Use of Alternative Detection Techniques 10
5 Applications **10**
5.1 Determination of Metal Cations 10
5.2 Metal Speciation 10
5.3 Inorganic Anion Analysis 12
5.4 Simultaneous Determination of Cations and Anions 13
6 Conclusions and Future Trends **14**
Abbreviations and Acronyms **14**
Related Articles **15**
References **15**

Capillary electrophoresis (CE) is a very useful technique which involves the separation of charged species (molecules) on the basis of their movement under the influence of an applied electric field in a capillary format. CE has been developed into rapid, highly effective techniques with extremely high separation power (10^5–10^7 theoretical plates) and high mass sensitivity (femtomole to zeptomole amounts of substances in nanoliter to picoliter sample volumes) with high potential for separation in the analysis of environmental sample. CE has been compared with ion chromatography (IC) in the inorganic analysis of environmental samples since inorganic ions are usually separated by IC with suppressed conductivity detection. The advantages of CE compared to IC are speed, resolution, and the lack of a need for gradient elution. Operating costs are considerably lower, since ion-exchange columns are usually expensive. In application, however, limitations imposed by the lack of detection sensitivity complicate the technique's use and impede its wide acceptance. However, as instrumental development continues (especially in detection mode and other techniques for manipulating sensitivity), this limitation will most likely be overcome and a diversity of inorganic species in real samples has already been proved to be amenable to the analysis by CE, and more can be expected in the years ahead.

In this article, to understand how CE can be applied to separation and detection of inorganic ions of environmental sample, the basic principles underlying this technique and the operating parameters of CE for inorganic analysis of environmental samples are presented. For trace analysis of environmental samples, methods for manipulating the sensitivity and selectivity control for metal ions are discussed. Techniques such as the coupling of CE with inductively coupled plasma mass spectrometry (ICPMS) are also discussed. Specific applications in real world samples which have complex matrices as well as artificial samples or reference materials are offered to demonstrate the potential of CE in the analysis of environmental pollutants.

1 INTRODUCTION

Inorganic pollutants of environmental samples comprise a variety of metal species and small anions. Since at trace level they are usually in unknown and variable matrices as various chemical states, a sensitive and selective technique is always required for their determination.

CE is an emerging technique in this area. CE possesses many unique advantages for the analysis of environmental pollutants, which include small sample volumes, high mass sensitivity, high separation efficiency, low cost columns, easy rinsing of the column, low consumption of reagents, simple methodology, short analysis time, etc. CE, which was initially developed for the separation and determination of biopolymers such as proteins, peptides and organic compounds, has been rapidly introduced into the analysis of inorganic ions since the early 1990s and has become a promising technique for the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

inorganic analysis of environmental samples. The active research in CE reflects the expectation of this method for the replacement or supplement to high-performance liquid chromatography (HPLC) or IC with suppressed conductivity detection, which is the most popular method at present in this area. However, the use of CE as a routine quantitative technique in environment analysis is still limited because of the low concentration sensitivity of CE due to the small sample volume. However, as instrumental development continues (especially in detection mode and other techniques for manipulating sensitivity), this limitation will most likely be overcome.

In order to give a better understanding of how CE can be applied to the separation and detection of inorganic ions of environmental samples, this article will present the basic principles underlying this technique and operating parameters of CE for the inorganic analysis of environmental samples. Methods for manipulating the sensitivity (important for trace analysis) and selectivity control for metal ions are also introduced. Specific applications in real world samples as well as artificial samples or reference materials will be offered to demonstrate the potential of CE in the analysis of environmental pollutants.

The basic principles and instrumentation relevant to the separation of inorganic ions and applications to environmental analysis are discussed in several books[1–6] and reviews.[7–13] There are over 100 reviews about other specific issues related to CE, such as detection technique, derivatization, mathematical modeling, etc. Among them, some reviews on environmental application are recommended to the interested reader.[7–24]

CE can be performed in various modes with differing mechanisms of separation, such as capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE), micellar electrokinetic capillary chromatography (MEKC), capillary electrochromatography (CEC), capillary isoelectric focusing (CIEF) and capillary isotachophoresis (CITP). In this article, CE represents CZE with an on-line ultraviolet/visible (UV/VIS) absorbance detector unless specifically defined.

2 BASIC PRINCIPLES

2.1 Capillary Electrophoresis System and Operating Parameters

The general mechanism for CE separation is based upon differences in the mobility of solutes transported along a capillary tube by a high dc voltage. In the simple system shown in Figure 1(a), both ends of the fused silica capillary, filled with a suitable electrolyte, are immersed into two reservoirs of the same electrolyte.

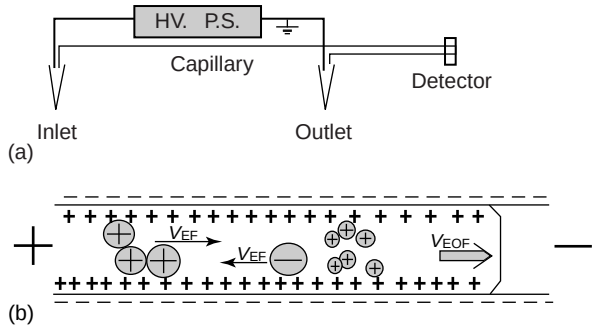

Figure 1 (a) Schematic of CE system; (b) mechanism of CE.

The capillary is normally 50–100 cm long with an internal diameter of 25–100 μm. A high potential difference of 20–30 kV is applied across two platinum electrodes in the reservoirs.

The sample (several nanoliters) is injected into the capillary and the components of the sample migrate towards the negative electrode. Because the cross-section of the capillary is small, a high surface to volume ratio is achieved. This dissipates heat, generated by joule heating, to the surroundings. Convective mixing within the capillary is, therefore, not significant and band broadening is minimized, resulting in efficiencies of several hundred thousand plates.

Electroosmotic flow (EOF) of the solvent occurs in the capillary because an electric double layer is formed at the capillary surface (Figure 1b). The fused silica capillary surface has a net negative charge owing to the dissociation of the functional groups on the capillary surface. Positive ions are attracted from the buffer solution and an electrical double layer is formed at capillary surface. When an electric field is applied, mobile positive ions, present at the capillary surface, are attracted to the negative electrode, and solvent molecules are transported concurrently. A unique feature of EOF in the capillary is the flat profile of the flow, which is beneficial because it does not directly contribute to the dispersion of solute zones.

The movement of analyte ions is caused by the combination of electrophoretic mobility of the analyte ion and the EOF of the solvent. Therefore, depending on whether or not the electrophoretic mobility has a similar direction to the electroosmosis, the separation mode in CE is divided into coelectroosmotic or counterelectroosmotic modes. Within a bare fused-silica capillary, small cations which are attracted to the negative electrode migrate at a velocity greater than the EOF (coelectroosmotic flow). Conversely, anions are repelled by the cathode and migrate more slowly than the EOF (counterelectroosmotic flow). Neutral species move with the EOF and are generally unseparated unless some modification to the surface charge of the analyte is made by

the addition of a surfactant (MEKC). The direction of the EOF can be decreased or reversed to achieve reasonable analysis times for small anions. The polarity of capillary wall can be modified by the addition of EOF modifiers.

A traditional CE detector that is commercially available is the on-column UV/VIS spectrophotometer. The small capillary diameter of CE creates a thin beam path for on-column UV/VIS detection. This creates a sensitivity problem in the study of environmental samples which usually require the detection of species in trace or ultratrace levels. To overcome this problem, many detection techniques have been developed for CE. Among them laser-induced fluorescence (LIF) and amperometric detectors are commercially available.

Since most small cations and anions lack the necessary chromophore properties allowing a straightforward detection at any wavelength above 200 nm, indirect detection is often employed for the detection of ultraviolet (UV) transparent ions. A UV-absorbing co-ion (or probe co-ion) is present in the background electrolytes (BGE) and detection is accomplished as a result of displacement of the probe ion by analyte ions. Commonly used electrolytes for anion determination are chromate[25] and pyrromellitate[26] and for cation determination, 4-aminopyridine,[27] imidazole[28] and creatinine.[29]

2.2 Electrolyte Condition

For many ions, the mere combination of electrophoresis and electroosmosis does not offer a sufficient selectivity for a satisfactory resolution from other peaks in the sample mixture. It is required frequently to refine CE separations by utilizing analyte interactions with suitable components in the electrolyte. Optimization of the CE separation can be difficult due to the wide array of parameters and variables that must be controlled. However, it is worth understanding several possible factors contributing to the selectivity of CE and its trend for the separation of inorganic ions. The pH level, concentration, component of electrolyte, additives, and EOF are regarded as the most important factors. Their influence on the selectivity is quite different depending on the sample status, especially on the sample polarity.

2.2.1 pH

The pH of the separation electrolyte must be carefully selected. While cationic separations are mostly performed at pH 3–5, anionic separations are seldom carried out using an electrolyte with a pH value below 7. This fact can be attributed to the different behaviors and properties of cations and anions in aqueous solutions. Cations, especially those of transition metals and some of alkaline earth metals, readily precipitate at high pH. In addition, the electrolyte constituents for providing an indirect detection mode for cations mostly have pK_a values in the low pH range. To match the mobility of the sample cations, the pH has to be sufficiently low so that almost all the electrolyte constituents are present as cations. On the other hand, for the determination of anions, the electrolyte pH has some effect only on the anions of weak acids. In acidic conditions, serious tailing and unacceptably long migration time occurs for the anions of weak acids in coelectroosmotic mode, e.g. acetate and carbonate. The long migration time results from the effect of EOF due to the change of pH. The EOF rate increases with the pH because of a large degree of dissociation of silanol groups on the capillary wall.

2.2.2 Concentration

The concentration of carrier electrolytes affects the selectivity. In general, a longer migration time can be expected with increasing concentrations of electrolyte. This is explained by increases in separation currents and the corresponding weakening of the electric field in the capillary filled with carrier electrolytes. Sample stacking effects can be obtained at an electrolyte concentration 10–100 times higher than the sample concentration. However, a high concentration of electrolyte will give the deteriorating joule heating effect which makes bubbles inside the capillary and sometimes results in a breakdown of voltage.

2.2.3 Selection of Electrolyte Component

Depending on the detection mode and interested ions, the background electrolyte is selected. When indirect UV/VIS detection is used, the electrolyte ion must be chromophoric as discussed above. A shift from one electrolyte to another is more commonly dictated by one of the conditions of CE, prescribing the match of mobilities between analyte ions and electrolyte co-ions. When the electrolyte anion has a lower mobility than the solute anions, an asymmetrical, fronting peak may result. This is because the high mobility solute anions tends to migrate ahead of the sample zone. Conversely, when the electrolyte anion has a higher mobility than the solute anion, a tailing peak results. Distinct tailing, usually observed for most carboxylates in chromate, can be avoided by choosing a phthalate-containing electrolyte. The selectivity of the separation of low mobility anions can be modified even to the complete reversal of order of migration by switching from coelectroosmotic to the counterelectroosmotic mode by addition of EOF modifier.

2.2.4 Additives

The use of additives in CE lead to dramatic changes in separation power. Micellar, complexing reagents, EOF modifiers, and organic solvents are various types of additives that have been used to control the selectivity of the separation.

The addition of an organic solvent to the electrolyte will change migration times and selectivity. The manipulation of selectivity use of organic solvents offers these advantages due primarily to an effect on the solvation of cations and on the ion pairing of anions. The addition of organic solvents to carrier electrolytes can lead to longer migration times with increasing concentrations of a solvent and the change of migration order of anions. It is assumed that organic solvents tend to shift ion pairing equilibrium to the right for anions. For cations, the solvation effect is accomplished by the changes in charge density, since in comparison to aqueous electrolytes, polar organic solvents affect the actual size of ions via forming ion solvent dipole bonds. The mobilities of solvated cations or ion paired anions are also influenced by changes in the viscosity of water–organic mixtures. One particular advantage of organic solvents, especially with regard to ionic complexing additives, is that the ionic strength of the capillary electrolyte is not increased, which prevents possible thermal effects.[30] Among a variety of organic solvents that have been studied with regard to their suitability for CE, methanol, *N*-methylformamide, dimethylformamide, dimethyl sulfoxide and acetonitrile[31] seem to be particularly useful. More detailed information on the important properties of organic solvents for use in nonaqueous CE separations were summarized in a review.[32]

The effect of micelles, complexing reagents, and EOF modifiers will be discussed in other sections.

2.2.5 Electroosmotic Flow

The regulation of the rate of EOF is another way to enhance the separation in CE. Although the EOF does not contribute to the selectivity changes (at least for the analytes of the same charge), it has been demonstrated to be useful in manipulation of the resolution, in particular, in micellar-mediated CE systems. Since the migration velocities of small, charged ions exceed the velocity of the EOF, the coelectroosmotic separation principle is usually preferred for inorganic analysis. In uncoated capillaries, the direction of the EOF is from the anode to cathode, which means that only cations can be separated without modification of the capillary wall. The direction of the EOF has to be decreased or reversed to achieve reasonable analysis times for small anions. EOF modifiers such as cetyltrimethylammonium bromide (CTAB),[33] tetradecyltrimethylammonium bromide (TTAB),[34] or hexamethonium hydroxide (HMH)[26] have been used for this purpose.

2.3 Detection System

2.3.1 Photometric Detection

Currently, most commercially available instruments for CZE are equipped with spectrophotometric detectors, in particular UV-absorbance detectors. This detector is preferred because it offers minimal loss in separation efficiency and easy set-up. As most small cations and anions lack the necessary intrinsic chromophore properties allowing a straightforward and sensitive detection at any wavelength above 200 nm, their detection generally requires derivatization using a suitable chromophore prior to separation. Otherwise indirect UV detection should be applied. For analysis of metal ions, derivatization can be also used to enhance the selectivity. Details of the method will be presented in the following section.

As the key to the latter approach is the displacement of a highly absorbing electrolyte co-ion, a close mobility match to the analyte ions is required; otherwise asymmetrical peak shapes are generated. Commonly used electrolytes for anion determination are chromate and pyrromellitate and for cation determination, 4-aminopyridine, imidazole and creatinine. While UV/VIS absorbance detection is the most common, straightforward, means of accomplishing on-column detection for CE, achievable sensitivity is limited due to short beam pathlength. Usually the limit of detection (LOD) is below the order of 10 mM. It is not sensitive enough for environmental trace analysis. Even so, CE still has an impressive mass sensitivity, which can compete with the best HPLC system because of a very low injected sample volume. To enhance concentration sensitivity for CE, various approaches have been developed. These approaches include preconcentration by solid phase extraction (SPE), on-line preconcentration by isotachophoresis (ITP), on-column preconcentration by sample stacking, and use of detection cells with extended optical pathlength. Albin et al.[24] have presented a review in this field.

As UV detectors, photomultipliers or diode array detector (DAD) is used. Photomultiplier has a higher sensitivity than DAD while DAD offers a full spectrum. Since inorganic ions do not have a unique UV/VIS spectrum, photomultiplier is desirable since it usually gives more sensitive results.

As more sensitive optical detection methods have been pursued, fluorescence has become the most popular, as demonstrated by the numerous applications described in a recent review.[18] LIF can provide LOD several orders lower than UV absorbance. Using direct fluorescence, detection limits in the attomole to zeptomole range and

even single molecule detection has been reported for polyaromatic hydrocarbons and proteins. In the field of study of metal ions, direct fluorescence detection of complexed metal ions was accomplished by Swaile and Sepaniak. Using 8-hydroxyquinoline-5-sulfonic acid (HQS), which fluoresces only when complexed, Mg^{2+}, Ca^{2+}, and Zn^{2+} can be detected in the ppb range.[35] More commonly, indirect fluorescence detection techniques were used for analyses of metal ions, although the number of studies has been rather limited. Using a quinine sulfate electrolyte buffer system for indirect fluorescence detection and electrokinetic detection, alkali and alkaline earth metals were detected at femtomole levels. Indirect LIF detection for capillary separation was reviewed by Yeung and Kuhr.[36]

Indirect LIF detection is also possible with MEKC. The sensitivity relies upon the different quantum efficiency of the fluorescent reagent in the two phases. Generally, the quantum efficiency of the fluorescent reagents is significantly enhanced in the micellar phase, and the displacement of it from the micellar phase to aqueous phase by the analytes caused a concomitant reduction in the fluorescence signal.

Chemiluminescence (CL) detection involves the reaction of the analyte, or some species produced by the chemical or enzymatic reaction of the analytes with a CL reagent to produce light in the detection cell. The most common CL detection system utilizes the hydrogen peroxide–luminol reaction, in which the hydrogen peroxide is a by-product of a chemical enzymatic reaction of the analyte. Alternatively, a variety of metal ions catalyze the reaction of hydrogen peroxide and luminol in alkaline solution and can thus form the basis of direct[37] and indirect[38,39] CL detection. For high sensitivity detection of metal ions separated by CE, Huang et al.[37] combined the incorporation of luminol into the separation buffer with post-column mixing of the hydrogen peroxide solution immediately before passing through the detection region. They reported the best system detection limits for Co(II) (20 zmol, $5 \times 10-13$ M) while other metals gave mass detection limits ranging from 2 amol to 100 fmol. Average reported plate counts were over 400 000, indicating no compromise in separation performance due to the post-capillary mixing of CL reagents.

2.3.2 *Conductivity Detection*

In addition to detection methods based on spectrophotometric principles, CZE combined with a conductometric detector has been presented in a number of published papers.[40–48] Different detector configurations suitable for both on-column and end-column conductivity detection have been reported but only the latter technology is commercially available to date.

Compared to CE with UV detection, devices equipped with conductometric detection offer significant enhancement in performance and sensitivity especially for the analysis of low molecular mass ionic compounds. Whereas in CE with UV detection the suitability of a certain carrier electrolyte strongly depends on its UV absorptivity, low-conductivity buffers or buffers suppressible with common methods are used in the case of CE devices equipped with non-suppressed or suppressed conductivity detection respectively. In non-suppressed conductivity detection, the so-called Good's buffer components[47,48] proved to be suitable carrier electrolytes for a number of separation problems.[42–45,49,50] Because the first commercially available CE device offering the possibility of conductivity detection was introduced recently,[45] only a small number of papers dealing with the applicability of this system for the analysis of real samples has been published. Since the high sensitivity inherent to electrochemical detection (ED) does not depend on the scale of measurement, ED could become an important alternative detection mode for CE.

Conductivity detection for CE was addressed during the covered period.[51,52] A simple conductivity detector with detection limits in the nanomolar range was demonstrated.[51] Suppressed conductivity detection for CE was discussed in terms of its compatibility with various electrolyte systems.[52]

2.3.3 *Inductively Coupled Plasma Mass Spectrometry*

ICPMS has great potential usefulness owing to a very low detection limit obtainable without preconcentration and a large dynamic range. It is noted that there is no need to separate species electrophoretically that contain different elements from each other as the ICPMS detector serves to identify the metals present individually. Speciation studies are of primary interest for CE coupled with ICPMS as only ions, complexes or molecules containing the same element need to be separated. Therefore, for different element ions of similar mobilities, electrophoretic resolution will be poor but the ions will be adequately identified by the ICPMS element specific detector. The coupling of CE with ICPMS is relatively new and interface designs are still being developed and evaluated. The use of direct injection nebulization and other low flow nebulizers are, without a doubt, the way forward if electropherographic resolution is to rival that obtained with UV detection.

The coupling of CE with ICPMS for elemental analysis has been reported by several groups.[53–56] Two major challenges for the coupling of this technique need to be addressed. They have been described by Tomlinson et al.[57] The flow-rate (typically of the order of nL min^{-1}) and low sample volume of a typical CE separation is the

first potential problem, as most nebulizers for ICPMS are designed to operate at flow-rates in the mL min^{-1} range. Glass frit,[55] direct injection, oscillating capillary, ultrasonic[54] and Meinhard nebulizers[56] are suitable for such low flows. Another factor to consider is that the end of the capillary will no longer be immersed in a buffer reservoir upon coupling to ICPMS and a method of ''grounding'' the electrode must be achieved. The construction of such an interface is an important consideration so that high transport efficiency is achieved.

3 SELECTIVITY CONTROL BY COMPLEX FORMATION

The electrophoretic mobility in CE, μ_{eff}, is originally determined by the intrinsic solute parameters, effective charge to size shown in Equation (1):

$$\mu_{eff} = \frac{q_i}{6\pi\eta r_i} \quad (1)$$

where q_i is the charge of hydrated or complexed ion, r_i is its size (the hydrodynamic radius) and η is the viscosity of the electrolyte medium. In the separation of metal species such as alkaline metals, alkaline earth metals, transition metals or lanthanides, the electrophoretic mobilities of most metal ions or metal complexes are similar in magnitude since they possess nearly identical charge and size. To improve separation efficiency of metal ions in CE, altering either the effective charge or size of a metal species or both is required. This can be done by the complex formation of the sample during (on-capillary) or before electrophoresis (pre-capillary), as well as routine practice of compositional changes of the carrier electrolytes resulting in ion dissociation or change of solvation effect, etc. Therefore a distinct feature of the separations of metal ions by CE is the frequent use of ligands to form complexes with the metal ions.

Classifications of complexation reaction are based on either their place in the final analytical set-up: before (pre-capillary), or during (on-capillary) the electrophoretic separation; or the degree of metal ion complexation due to the thermodynamic stability of the metal complexes: partial or total complexation.

In the partial complexation approach, usually on-capillary reaction with weakly complexing ligands such as the variety of carboxylic and hydroxycarboxylic acids takes place. In this case, the complexation reaction should be fast and quantitative to be detected within the time of typical CE run, and the complexing agent should be stable. Separation selectivity is most easily manipulated in this mode coupled with indirect photometric detection. The simultaneous separations of the greatest number of metal ions have been achieved using this approach.

On the other hand, in the total complex approach, metal ions are completely transformed into moderate to strong complexes usually in pre-capillary mode or rarely on-capillary mode. Adapting pre-capillary complexations, there are some advantages such as (i) the large flexibility in optimizing the reaction conditions, e.g. the possibility of selecting solvents which are not compatible with the CE system, or (ii) no limitation to the reaction kinetics, provided a single derivative is formed with the analyte, (iii) the conditions during complexation do not have to be compatible with the electrophoretic buffer, and (iv) the availability of a wide variety of complexing agents. However, in this case, control of separation selectivity and maintenance of good peak shapes is an intriguing task. Use of the total complexation mode with ligands such as ethylenediaminetetraacetic acid (EDTA) or cyclohexane-1,2-diaminetetraacetic acid (CDTA) has made possible the separation of some higher-valent metal ions which cannot be separated in other systems. However, the generally low sensitivity of direct photometric detection remains a weak point of this approach. Therefore, new metallochromic ligands with high absorbing characteristics, which react with as many metal ions as possible and form a single complex only with each metal, should be applied.

3.1 Partial Complexation

When a complex-forming equilibrium is established by introducing a weak complexing reagent into the electrolyte, partial complexation is set up. As a result, several different complexes of metal ions may exist simultaneously due to stepwise complexation. For an analyte species present in several forms which are in equilibrium with rapid kinetics of interchange between the forms, the effective mobility of the analyte is defined as the weighted average of the mobilities of each of the forms[58,59] (Equation 2):

$$\mu_{eff} = \sum \mu_i \alpha_i = \frac{\sum \mu_i [ML_i]}{\sum [ML_i]} = \frac{\sum \mu_i \beta_i [L]^i}{\sum \beta_i [L]^i} \quad (2)$$

where μ_i is mobility and α_i is the distribution coefficient of the form i, $[ML_i]$ is the concentration of the complex ML_i, [L] is the concentration of the form of ligand forming the complex and β_i is the overall stability constant ($ML_0 = M$, $\beta_0 = 1$). Therefore, based on the nature of the complexing reagent and the concentration of free ligand, usually dependent on the pH of the electrolyte, effective mobility of complexed metal ions can be controlled.

Affecting the degree of complexation and resulting migration velocity is the concentration of the complex-forming reagent, [L]. The concentration effects on the electrophoretic mobilities of metal ions have been studied in several studies.[60–65] The majority of organic acids used as partial complexing reagents are weak acids and hence undergo acid–base dissociation equilibria in the electrolyte buffer. In the case of a monoprotic acid, Equations (3) and (4) apply:

$$HL \rightleftharpoons H^+ + L^- \quad (3)$$

$$K_a = \frac{[H^+][L^-]}{[HL]} \quad (4)$$

in which K_a is the acid dissociation constant of HL. Consequently, it is not the total concentration of a given reagent added to the electrolyte but the concentration of free ligand which determines the ability to differentially reduce the mobilities of metal ions. Also the pH of the electrolyte is another important factor since the concentration of free ligand is dependent on it.

A variety of carboxylic and hydroxycarboxylic acids are found to be suitable as partial complexing reagents for the separation of metal cations. Of various complexing reagents tested, α-hydroxy-isobutyric acid (HIBA) has been most popularly used.[66–71] Remarkable separation results for alkali, alkaline earth, transition and rare earth metal ions are shown in Figure 2. Lactic acid also appears to be advantageous in terms of separation selectivity,[72,73] and citric,[74] malonic,[75] glycilic,[76] tartaric[77] and oxalic acids[78] have been used to give good selectivity. A comparative study of 10 carboxylic acids as complexing agents showed that oxalate and citrate exhibited the largest selectivity changes, and the optimum pH for separating metal ions in the presence of the complexing carboxylic acid was close to the pK_{a1} of the acid.[79] A noticeable improvement of the separation can be obtained by the combination of two ligands which possess different abilities for the complexation, for example, HIBA and acetate,[80] or lactate and HIBA.[81] Using a lactic acid and 18-crown-6 buffer with imidazole for indirect UV detection, alkali, alkaline earth, and transition metal cations have been simultaneously separated by CE.[82]

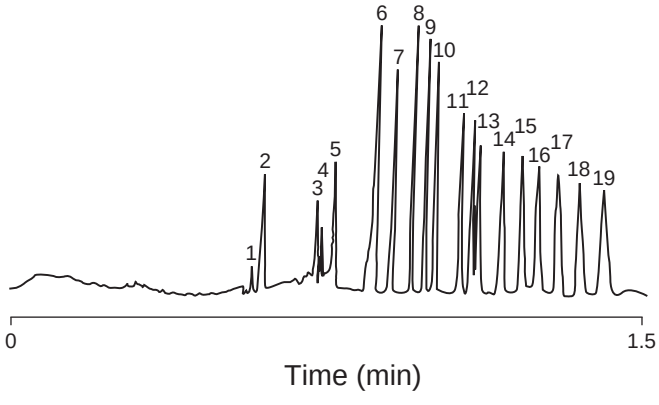

Figure 2 Simultaneous separation of alkali metals, alkaline earths, and lanthanides. Carrier electrolyte: 10 mM Waters UVCat-1, 4.0 mM HIBA, pH 4.4, fused silica capillary 36.5 × 75 µm, 30 kV, hydrostatic sampling: 20 s per 10 cm, indirect UV detection at 214 nm. Peak identities: 1: Rubidium 2 ppm; 2: Potassium 5 ppm; 3: Calcium 2 ppm; 4: Sodium 1 ppm; 5: Magnesium 1 ppm; 6: Lithium 1 ppm; 7: Lanthanum 5 ppm; 8: Cerium 5 ppm; 9: Praseodymium 5 ppm; 10: Neodymium 5 ppm; 11: Samarium 5 ppm; 12: Europium 5 ppm; 13: Gadolinium 5 ppm; 14: Terbium 5 ppm; 15: Dysprosium 5 ppm; 16: Holmium 5 ppm; 17: Erbium 5 ppm; 18: Thulium 5 ppm; 19: Ytterbium 5 ppm. (Reproduced with permission of VCH from Jandik and Bonn[3] © 1993.)

3.2 Formation of Inclusion Complexes

Enhanced selectivity for the separation of alkali and alkaline earth cations can be obtained from a somewhat different partial complexation: inclusion complexation. The formation of inclusion complexes is observed in molecules having a well-defined cavity in their structure such as the crown ether. The dimension of the cavity in crown ethers makes those compounds ideally suitable for selective interactions with alkali and alkaline earth cations. According to this complexation principle, macrocyclic ligands magnify differences in mobility of only certain cations like K^+, Ba^{2+}, Sr^{2+} which interact specifically with 18-crown-6. As the charge of a metal ion remains unchanged during the complexation, the retardation effect on the migration velocity is owing to the charge density alterations for much bulkier crown-complexed cations. Apart from the resolution of ammonium/potassium and calcium/strontium pairs, other changes in selectivity, most dramatically the large change in migration time of barium, are noted.

Since the crown ether ligands are electro-neutral, free ligand concentration and thereby complexation is virtually unaffected by the pH and the only electrolyte variable to improve the selectivity is the total concentration of crown ether. In general, higher electrolyte concentrations of crown ether are favorable as they widen the range of effective mobilities of the cations.

Macrocyclic polyethers, mostly 18-crown-6, are commonly used in combination with a weak complexing acid in order to separate potassium/ammonium or sodium/lead(II) in the presence of other group cations. An impressive example of separation selectivity gained by the addition of 18-crown-6 to the lactate system is the complete resolution of Mg, Ca, Sr and Ba against a 1000-fold excess of sodium.[83] Clearly seen from the electropherograms, enhanced cation selectivity can be

achieved, for example, by an addition of a suitable crown ether to carrier electrolytes.

3.3 Complete Complexation

Separation of metal ions in the pre-capillary-complexed form presents a viable supplementary approach of improving selectivity. Pre-capillary complexation usually leads to the formation of complete complexes with the metal ions. Since most of the metal ion is in the completely complexed state under all electrolyte conditions, manipulating selectivity is not as easy as a partial complexation. However, depending on their charge and size, the strong metal complexes can migrate towards different electrodes, so several migration modes can be applied and more flexibility with regard to the choice of separation conditions can be obtained. When complexes of metal ions with a chromogenic reagent are used, very sensitive detection can be obtained by a direct detection method. Other benefits of fully complexed metal ions are the possibility of largely reduced interferences from complex sample matrices, the feasibility of analyzing metal species incompatible with carrier electrolyte or those that possess a slow rate of complexation, and good tolerance of sample pretreatment conditions.[84]

In general, the complexing reagent for complete complexation is required to be non-selective and to form a highly stable metal complex not to be dissociated in the CE system. Another requirement for complexing reagents in CE, unless the micellar-mediated CE systems are concerned, is the formation of charged complexes. To satisfy this requirement the ligand must contain at least one ionizable group, such as sulfonate, carboxylic or a quaternary ammonium group, which does not participate in complexation. A variety of complexing reagents has been previously reported. Among the most commonly used reagents, polyaminocarboxylic acids such as EDTA, CDTA and analogs appear to be most advantageous for the separation of different metal groups. These reagents are generally considered to form only 1:1 metal complexes with various metals. Even though a polydentate ligand forms very stable complexes, it can be used in the partial complexation mode if a competing ligand is added to the electrolyte to decrease the conditional stability constant of the primary ligand.[85]

A number of divalent metals and Fe(III) as complexes with EDTA have been separated.[86] A linear relationship between the electrophoretic mobility and the reciprocal of the square root of molecular mass of alkaline earth metal complexes has been obtained while other metals have some deviations. This was explained by differences in the degree of hydration of the complexes. CDTA has also been used for the separation of 32 divalent and higher valent metal ions.[87] Among them 23 metal ions have been separated in the baseline within 10 min. Numerous high valent metal ions such as Fe(III), Sn(IV), or U(VI) have been separated. The selectivity of the separation of lanthanide complexes was studied with EDTA, CDTA and two other polyaminocarboxylic acids.[88] The electrophoretic mobility of the complexed lanthanides was observed to be strongly dependent on the pH of the electrolyte. This was explained by the formation of mixed hydroxo complexes. Monodentate inorganic ligands such as cyanide or chloride are also regarded as good complexing reagents for the separation of precious metals, even though complexation conditions are difficult to control.[89] Most metal complexes in this category are negatively charged. Therefore control of EOF is required for coelectroosmotic mode. More applications of complete complexation have been presented in other reviews.[8,9]

4 MANIPULATING SENSITIVITY

A number of steps have been taken to overcome the limitations imposed by the lack of detection sensitivity of CE. Development of new sensitive detection systems is one of the most active research areas in CE in recent years. Alternative capillary geometries and improved optical design for the photometric detector is another approach toward sensitivity enhancement.[3] Using commercial systems, different on-column preconcentration approaches were developed, including sample stacking,[90] SPE,[91] on-line ITP,[92] and transient ITP.[93] Among them, sample stacking is one of the most promising techniques from the practical and economical point of view. In this section, some techniques used only in environmental analysis will be introduced. Reviews of methods manipulating sensitivity have been presented in other papers.[3,22]

4.1 Use of Detection Cells with Extended Optical Pathlength

In commercial CE with photometric detectors, the optical pathlength for detection is limited by the narrow diameter of the capillary column due to on-column detection. To extend the optical pathlength, various special detection cells have been designed, such as Z-cells, rectangular cells, multireflection cells, along the column end cells, bubble cells, etc.[3] Although these cells have enhanced sensitivity, they have also posed difficulty in manufacture and manipulation. They can even decrease separation efficiency. The bubble cell has been used in environmental analysis with an enhanced sensitivity of more than three-fold without degradation of resolution of closely migration peaks[94,95] and is now commercially available.

4.2 On-column Preconcentration by Sample Stacking

Sample stacking occurs as an on-capillary concentration technique. Commonly, when the injected sample has a lower ionic strength than the running buffer, sample stacking occurs. It can also occur during the electrokinetic injection process, or separation process in case of hydrodynamic injection. Since the conductivity in the sample is lower than that of BGE, the applied voltage is not distributed uniformly along the capillary between the two ends. The field strength is higher in the sample zone, which results in higher migration velocity of an analyte ion, whereas the field strength is lower in the electrolyte, which results in lower migration velocity of the analyte ion. The analyte ion will then stack up at the stationary boundary between the two zones, thereby increasing its concentration. Sample stacking in CE was reviewed by Chien and Burgi[90] and Song et al.[91]

In the simplest form of sample stacking, a plug of water or low concentration buffer containing the analyte(s) is injected onto the column filled with a relatively high ionic strength run buffer. On application of a voltage of 20–30 kV across the capillary, a high electric field is developed across the sample plug which drives its ionic species towards the higher ionic strength run buffer regions. Positive species such as metal chelates and metal ions will stack at the front of the sample plug in the direction of the CZE cathode. This leads to a decrease in the LOD of order of a factor of three[96] by the time the species has migrated to the detector. This stacking is limited to the size of the sample plug that can be injected.

A large plug of the sample can be injected. However, in this case, if the sample zone is too long, laminar flow, which originates from the different local EOF caused by the nonuniform distribution of the applied voltage, will occur and de-stack the stacked analyte ion. Therefore, the relatively large sample matrix zone during or after the stacking stage has been removed by taking advantage of EOF, which is called large volume sample stacking.[90,91] For anionic analytes, this involves field polarity reversion after the capillary is filled with sample. Then the sample matrix is pushed out of the capillary from the injection end by EOF, and anionic analytes move with a very high velocity towards the boundary between the sample and the buffer. The field polarity is switched back when the boundary is approaching the injection end of the capillary, which is indicated by the current approaching 95% of its highest value, i.e. the current when the capillary is solely filled with buffer. The LOD can then be improved by a factor of 200. The preconcentration of cationic analytes is similar, but demands a reversed EOF. However, in this approach, close monitoring of the variation on current is required for the just in time polarity switching for complete removal of the sample plug and the avoidance of sample loss. This can be solved by adding an EOF modifier, such as CTAB,[96] diethylenetriamine (DETA)[92] or ethanolamine[93] in the BGE to suppress the EOF, therefore the electrophoretic mobility of anions becomes higher than that of EOF.

Sample stacking can also be performed by electrokinetic injection with LOD improved by a factor of 1000; however, the reproducibility is very poor.

The reproducibility of large volume sample stacking with matrix removal as an on-column preconcentration technique was compared with that of normal injection,[97] giving a result of typically 4.0 vs 1.2% for peak area and no differences for migration time. It was indicated that this preconcentration technique requires samples having a relatively low and reproducible conductivity.

In normal sample stacking with hydrodynamic injection, the ionic strength of samples has a significant effect on sensitivity.[98,99] Higher ionic strength of a sample led to lower LOD. Recently, large volume sample stacking with matrix removal has been applied to the analysis of small organic and inorganic anions without polarity switching.[100] The pH is adjusted to the acidic range so that the EOF can be adjusted to be slower than the electrophoretic mobilities of small anions and more than 300-fold enrichment was readily achieved. In the analysis of methylmercury, the achieved LOD was 12 ppb with large volume sample stacking with matrix removal.[101]

Sample stacking can also be used in MEKC. For ionic analytes, the procedure is the same as in CZE. However, for neutral analytes the procedure has to be modified a little. Liu et al.[102] reported the accomplishment of sample stacking of neutral analytes by dissolving them in a micellar solution with a low concentration just above critical micelle concentration (CMC). The micelles were stacked up under applied voltage, and caused stacking of neutral analytes. Normal and reversed field polarity was used and produced a 75–85-fold increase in sensitivity for 1,2,4,7- and 1,2,4,8-tetrachlorodibenzo-*p*-dioxin (TCDD). Susse and Muller[103] used sample stacking with reversed field polarity as an on-column preconcentration technique for the analysis of pesticides by MEKC. LODs between 0.01 and 0.1 ppb were achieved, where 30-fold preconcentration was attributed to sample stacking and 250-fold to SPE.

4.3 On-line Preconcentration by Isotachophoresis

ITP can be used as a preconcentration technique for diluted samples in CE. In this procedure a small quantity of the sample is introduced at the interface of a discontinuous buffer system, consisting of a leading and a terminating electrolyte. The analyte ions migrate between a leading ion having the highest mobility and a terminating ion having the lowest mobility in the

buffer system. Under applied voltage, the analyte ions stack up and migrate as consecutive zones, according to their mobility between the leading and terminating ions. Establishing ITP conditions at the beginning of a run provides a method for on-column preconcentration; separation then proceeds by zone electrophoresis. From the point of view of trace analysis, ITP can be considered a good preconcentration technique for CE but coupling ITP and CE instrumentation can be rather complex. Recently, ITP coupled with CZE has been shown to be capable of being performed in a single capillary by taking advantage of a counterflow. Moreover, ITP can also be performed before CZE separation as a transient preconcentration step (TITP) in commercial CE instruments by carefully selecting the CZE electrolyte as a leading or terminating electrolyte to form an ITP buffer system with an additional electrolyte. There is a review on these two techniques.[104]

4.4 Preconcentration by Solid Phase Extraction

SPE inside the capillary is another preconcentration technique capable of providing a concentration factor of $10–10^6$, effective especially for aqueous samples. This solid phase can be incorporated as a short plug at the beginning of the capillary, or as a hydrophobic stationary phase bonded to the inner wall such as an on-line coupled C_{18}-coated column.[105] However, in the latter case, because of the low amount of the wall-coated stationary phase, the enrichment factors were limited to 10–35-fold. When in-column SPE is applied, the sample must be cleaned before it is introduced into the capillary and cannot be introduced directly because of the risk of blocking the system.

4.5 Use of Alternative Detection Techniques

To improve the detection sensitivity of CE, a range of other detectors has been developed. Among them, indirect LIF, CL, ED and ICPMS received much attention as popular choices for high-sensitivity detection for inorganic ions. The advantages, disadvantages and limitations of each technique are presented in section 2.3.

5 APPLICATIONS

This section will give a general overview of CE procedures suitable for practical application. Therefore, some selected results of both basic and applied research will be presented. The following is divided into sections dealing with metal cations, metal speciation, inorganic anions and simultaneous determinations of cations and anions.

5.1 Determination of Metal Cations

The main strategy for the simultaneous determination of metal cations by CE is complexation. Many metal cations have nearly identical charges and identical hydrated ionic radii. Therefore, all complexations, whether weak, strong or inclusive, need to be applied for proper separations.

The most significant agent is HIBA, widely used for the separation of alkali, alkaline earth metals, transition metals, and lanthanide. Using the on-capillary complexation principle, separation of 19 cations in only a few minutes[3] has been accomplished as shown in Figure 2. A quinine sulfate buffer system with HIBA has been used for the indirect fluorescence detection of several alkali and alkaline earth metals.[106] Reagents such as HQS[107] and 1,10-phenanthroline[108] have been employed for the quantification of zinc(II) and iron(III), respectively, in tap water samples without preconcentration. The latter can lead to a loss of analyte and peak broadening. There have been several examples indicating that this complexation technique makes it possible to reduce interference from complex sample matrices (electroplating solutions,[109,110] ores,[111] catalysts,[107] ceramics,[107] etc.).

Complete complexation with pre-capillary mode has been applied for the separation of a number of alkali and alkaline earth metal ions.[112] EDTA was used as complexing reagent and pyridine as UV chromophore for indirect UV detection. CDTA also was used for the separation of lanthanides as anionic complex.[113] For the determination of precious metals, complexation with chlorine or cyanide is a popular method, even though more rigid control of complexation conditions is required. Os(IV), Pt(IV), Ir(III), Pd(II), Au(III), and Rh(III) were separated efficiently within 5–6 min as chloro complexes with detection limits of 0.1–0.6 ppm.[114]

Enhanced selectivity for alkali and alkaline earth metals can also be obtained with a suitable macrocyclic polyether (like crown ethers)[11] or polyethylene glycol[12] incorporated into the electrolyte. As an additional separation mechanism with complexation function, stronger complexing reagents, e.g. EDTA, but under conditions of incomplete complexation,[13] are of certain use to manipulate differences in effective mobilities of alkaline earths and magnesium as well. The more interested reader should refer to the reviews listing the experimental conditions of determination of metal ions by CE from almost all publications before 1997.[8,9]

5.2 Metal Speciation

The ability to resolve different oxidation states of a metal is a distinct benefit of CE. Since it causes a minor disturbance on the existing equilibrium between different species, CE gives a unique promise for speciation purposes. Most of the studies performed on speciation

have been focused on the optimization of separation parameters in order to separate various metal species in environmental, biological, and industrial samples. Since a recent review has been published,[7] some recent examples of the application of CE to the metal speciation will be presented in this section.

Complexation is the most valuable approach for performing metal speciation on different oxidation states. Separations of metal ions in a free, uncomplexed form are comparatively rare in the practice of CE for metal speciation. Complete conversion of metals into charged complexes, which takes place upon the addition of a complex reagent to a sample before introducing to a capillary, yields differences in charge densities while the initial concentration ratio of different oxidation states is preserved. For speciation involving metal oxidation forms of opposite charges, pre-capillary complexation is a straightforward strategy to impart the same charge and electrophoretic direction on the species.[115,116]

Among the numerous metal cations that exhibit toxic effects on humans and plants or aquatic organisms, the most frequently studied ones are aluminum, iron, copper, manganese, nickel, and zinc. Inorganic Al(III) and its complex forms appear to be much more toxic than organically bound aluminum. To study the distribution of aluminum and its inorganic and organic complexes in aqueous solutions by CE, indirect UV detection was performed with imidazole as a UV-absorbing component of the electrolyte.[117] Good resolution was obtained for the analysis of a sample solution containing four aluminum species: Al^{3+}, AlF^{2+}, AlF_2^+, and Al (oxalate)$^+$ during a period of less than 5 min. As a certain disadvantage for practical purposes, high detection limits (10 μM) shall be noted; these are of 1–2 orders of magnitude above those needed for use in natural waters.

With CE separation of Fe(II) and Fe(III) ions the major problem is the selection of complexing agents suitable for the derivatization of both iron species. There are some reports which show the possibility of separating the Fe(II) and Fe(III) ions by CE after pre-capillary complexation with EDTA and CDTA.[118]

The utility of micellar-mediated CE to differentiate Fe(II) and Fe(III) following derivatization with 4-(2-pyridylazo)resorcinol (PAR) was reported[119] (Figure 3). The limits of detection accomplished in the visible spectral region (500 nm) were in the submicromolar range. Later, in an attempt to enhance the sensitivity, on-column complexation with PAR, in combination with sample stacking as an enrichment step, provided a 100-fold improvement in the detection limit for iron(II) over the precolumn complexation method.[120] This allowed the authors to apply the procedure to the development of the analysis of a pond water sample. Iron species, while they are stable in the carrier electrolyte solution,

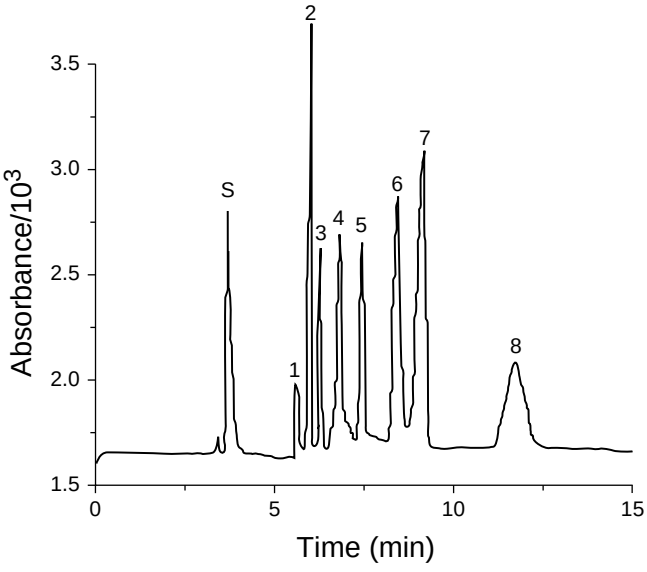

Figure 3 Electropherogram of Fe(II)-PAR and Fe(III)-PAR complexes. 75 μm × 50 cm capillary, 10 mM phosphate buffer (pH 8), 75 mM SDS and 1×10^{-4} M PAR. Peaks: S: acetone; 1: Cr(III); 2: Co(II); 3: Cu(II); 4: Pb(II); 5: Fe(II); 6: Zn(II); 7: Fe(III); 8: Cd(II). (Reprinted from A.R. Timerbaev, O.P. Semenova, P. Jandik, G.K. Bonn, 'Metal Ion Capillary Electrophoresis with Direct Ultraviolet Detection Effect of a Charged Surfactant on the Migration Behavior of Metal Chelates', *J. Chromatogr. A*, **671**, 419–427 © 1994, with permission from Elsevier Science.)

can be determined without derivatization by means of an element-specific detector such as inductively coupled plasma atomic emission spectroscopy (ICPAES) or ICPMS, as was demonstrated by Olesik et al.[171] Since the analytes exist in a highly mobile cationic form in this case, an analysis time of 1 min or less is attainable. Furthermore, separation of different metal ions from each other was not needed due to the high selectivity of ICP spectrometry.

There is a continuing interest in the determination of chromate ion, the most toxic of chromium species, motivated by the fact that chromium is commonly used in various industries and a considerable amount may be released into the environment. The first successful speciation of both oxidation states of chromium through CE separation was achieved by converting Cr(III) into a negatively charged complex with CDTA in order to impart the same charge and similar mobility to chromium species.[115] The separated anionic species were monitored by direct UV measurements at 214 nm. The procedure was applied to rinse water and processing solution samples and was found to be free of interference from transition metal impurities. Similar results have been obtained with EDTA instead of CDTA for the separation of the Cr(III) and Cr(VI) species by Jung et al.[116] with the same separation strategy.

Speciation of arsenic is of particular interest because of the wide range of toxicity exhibited by the different chemical forms. CE has been widely used for the separation of arsenic compounds.[121–123] Direct UV absorption detection at a wavelength between 190 and 200 nm is the most commonly employed. However, the arsenic concentrations to be detected are generally at ppm levels in aqueous solutions. Therefore on-column preconcentration such as field amplified injection[8] has been applied or coupling of CE with ICPMS has been actively developed.[124–128]

Selenium can be a toxic or an essential element depending on its chemical state. Several studies have focused on applying CE for the separation of selenium species using various electrolyte composition and detectors.[124,126]

For the speciation of oxoanions of selenium, or arsenic, the coelectroosmotic migration mode by reversing the direction of EOF towards the anode is suitable for resolving oxoanions and organic anionic forms of selenium.[122,123,129] Counterelectroosmotic migration modes have also been applied for the speciation of oxoanions of arsenics, where variations of the electrolyte pH offer an effective means for separation efficiency control over metal oxoanions. As anions of polybasic acids, these species are pH-sensitive and show a strong dependence of ionic mobility on the carrier electrolyte acidity. The anionic analytes can be transported towards the cathode by the EOF, if their charge is reduced by using slightly acidic or neutral buffers. This combination of cathodic EOF and detection is favorable for the separation of arsenic compounds.[122,130] If the electrophoretic mobilities of oxoanions are greater than that of EOF, they can migrate counterelectroosmotically towards the anode.[131] This is the case for arsenates and selanates which exist as doubly-charged anions over a wide pH range. For their fast separation from the slowly migrating arsenite and selenite species, the pH of the electrolytes should be higher than 10 where all anions are essentially ionized. To shorten the separation time further, an EOF modifier like a cationic surfactant can be incorporated into the electrophoretic buffer in order to suppress or reverse EOF.[132]

In practice, however, applications of the procedures developed to date are rather scanty, evoking problems with real sample matrices. Few exceptions deal with the analysis of waters and soil extracts,[133] coal fly ash,[134] water samples from a tailing of tin mine processing,[135] spiked water[109,130] and urine[110] samples.

The toxicity of many organometallics often exceeds that of the same metal inorganic species, which makes the speciation of paramount importance. Separation of cationic organomercurials and inorganic mercury(II), facilitated by the formation of their complexes with cysteine, was examined by Medina et al.[136] For methylmercury, which is recognized as one of the most dangerous chemical species in the environment, direct UV detection of the cysteine complex provided a detection limit as low as 10 pg. The CE procedure developed has proved to be efficient for rapid and simple determination of methylmercury (at the level of low ppb) in biological materials of marine origin. CH_3Hg^+ was extracted from the sample (as chloride derivative) with toluene, then back-extracted with a 0.1% cysteine solution and, finally, the aqueous extract was subjected to CE analysis. A unique possibility of MEKC to separate both neutral and charged analytes was exploited for simultaneous separation of organoselenium and organolead compounds.[137] The complete resolution of a test mixture of trialkyllead and phenylselenium compounds has been accomplished. The competitive interactions of the analytes with electrolyte components, sodium dodecyl sulfate and β-cyclodextrin, taking place via hydrophobic and host–guest complexation mechanism, respectively, appear to govern the migration pattern observed. For organotin species,[138] the migration behavior in MEKC was found to be dominated by the distribution ratios of the solutes between the aqueous electrophoretic and micellar phase, i.e. by their hydrophobicity.

Note that a majority of the reports involving water or biological matrices have only demonstrated the analysis of spiked samples, as for many metalloorganic contaminants the concentration levels in these objects are beyond the detection sensitivity of CE analysis. For example, Ng et al.[137] developed a scheme for the determination of trimethyl- and triethyllead, which includes preconcentration by means of liquid–liquid extraction and evaporation and then MEKC. However, the compounds under investigation were not even detected in a water sample collected at a heavily used car park, since they were present at levels below the detection limits. Neither allowed the enrichment by more advantageous membrane SPE on a C_{18} disk to determine organolead and organotin species in non-spiked water samples using MEKC.[138] A combination of subcritical fluid extraction, using chlorodifluoromethane as solvent, and MEKC has been applied for the determination of alkyllead and alkyltin compounds in solid samples.[139] To improve the recoveries of two less hydrophobic trimethyl species, sodium diethyldithiocarbamate was added as a complexing agent to the soil sample before extraction. However, direct analysis was only possible for a rather high content of dibutyltin stabilizer (1.2–1.5%) in poly(vinyl chloride) plastic.

5.3 Inorganic Anion Analysis

The determination of common inorganic anions such as fluoride, chloride, bromide, nitrite, nitrate, phosphate,

and sulfate, is a significant factor in the characterization of the quality and the extent of contamination of water. Since they do not have enough chromophoric characteristics and the direction of electrophoretic mobility of anions is opposite to that of EOF, in determining them by CE with untreated fused silica capillaries, the common procedure includes a reverse of EOF from cathode to anode and indirect detection. Many of studies performed have been focused on the optimization of separation parameters of electrolyte based on above two strategies.[140–150] Jandik and Jones systematically developed the separation procedure for anion analysis,[3,25] and accomplished spectacular separations of inorganic anions, as well as small organic anions in a very short time. CE has been applied to the anion analysis of a variety of environmental samples, such as natural water,[151] drinking water,[152] tap water, water from the space shuttle and space station,[153] wastewater,[154] seawater,[155] soil solution,[133,156] silicon product,[157,158] toothpaste,[159,160] single rain, fog, and cloud drops,[161] aerosols[161] etc. A specially developed sampling procedure has been applied in the analysis of a single rain, fog, or cloud drop. This shows that CE can be applied to the quantitative, multicomponent chemical analysis of ultrasmall volumes, like a single rain or fog drop. CE has been compared with IC in anion analysis.[153,158,162,163] It was reported that CE has been proved to be a good alternative to IC since it gives usually good agreement with IC in their results even though their detection limit is usually a little inferior to IC. A comparison for the results of CE and IC is shown in Figure 4.

Speciation of sulfur, nitrogen, or chlorine in various water samples has been performed.[164–166] Numerous chlorine-containing anions with different oxidation states, such as chloride, chlorite, chlorate, and perchlorate in tap water, swimming pool water, and bleaching preparation were separated by CE and IC.[166] CE has advantages for the quantitative determination of chlorine-containing anions in several hundreds ppb level in the presence of a large excess of chloride or nitrate and sulfate.

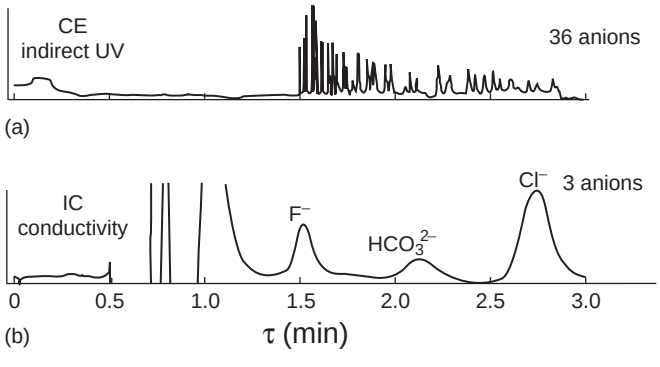

Figure 4 Comparison of first 3.1 min of an IC separation vs a CE separation. IC chromatogram (b) separates only 3 anions: fluoride, carbonate, and chloride using a Waters IC-Pak A, borate-glutamate eluent at 1.2 mL min^{-1} and Waters 431 conductivity detection. The electropherogram (a) separates 36 anions using a Waters Accusep fused silica capillary with indirect detection method. (Reprinted from W.R. Jones, P. Jandik, 'Controlled Changes of Selectivity in the Separation of Ions by Capillary Electrophoresis', *J. Chromatogr.*, **546**, 445–458 © 1991, with permission from Elsevier Science.)

5.4 Simultaneous Determination of Cations and Anions

A simultaneous determination of inorganic cations and anions in a single analysis is both advantageous and necessary. However, no specific technique is available for this purpose up to now. CE can be developed for simultaneous determination of cations and anions but it still remains a challenging problem. Cations and anions were usually separated using different BGEs under different conditions in CE. Following are some reports about the simultaneous determination of cations and anions. A common strategy among them is taking dual electrolyte for indirect detection of cation and anion and finding the optimum EOF.

A simultaneous determination of small cations and anions has been performed in one single electrophoretic run using one capillary and one detector.[167] In this

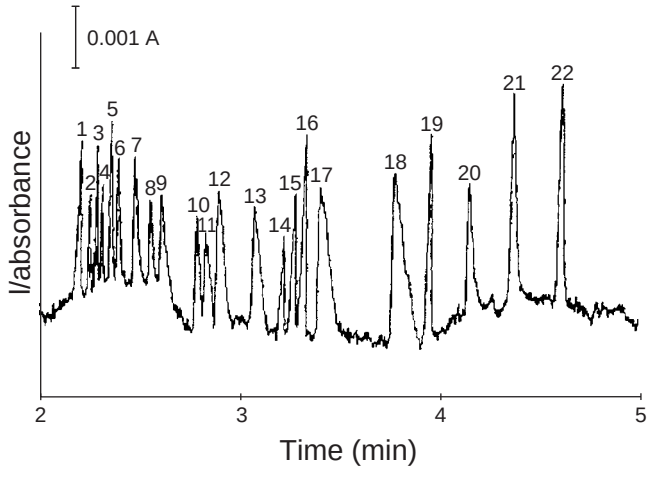

Figure 5 Separation of 22 anions and cations. Electrolyte: 6 mM 4-aminopyridine, 2 mM H_2CrO_4, 30 μM CTAB, pH 8. Injection conditions: hydrodynamic injection; capillary ends elevated to a height of 10 cm for 40 s (cathode end) and to 5 cm for 20 s (anodic end). Time between the injections, 80 s. Detection principle: indirect UV at 262 nm, peak reversal. The linear baseline drift observed during the first 3 min was corrected for. Peak identification: 1: $S_2O_3^{2-}$; 2: Br; 3: Cl; 4: SO_4^{2-}; 5: NO_2; 6: NO_3^-; 7: WO_4^{2-}; 8: MoO_4^{2-}; 9: citrate; 10: maleate; 11: fumarate; 12: F^-; 13: HPO_4^{2-}; 14: Cs^+; 15: K^+; 16: NH_4^+; 17: HCO_3^-; 18: acetate; 19: Na^+; 20: Ca^{2+}; 21: Mg^{2+}; 22: Li^+. (Reprinted from P. Kuban, B. Karlberg, 'Simultaneous Determination of Small Cations and Anions by Capillary Electrophoresis', *Anal. Chem.*, **70**, 360–365 © 1998, with permission of the American Chemical Society.)

case the sample is injected into the one end of the capillary and subsequently into the other end of the capillary. Cations and anions migrate against each other toward the center of the capillary where the detection window is placed. 4-Aminopyridine and chromate with CTAB have been used as electrolytes. 22 small inorganic and organic anions and alkaline and alkaline earth metal cations were separated within 5 min as shown in Figure 5. Adapting the same sample injection strategy with different electrolyte composition, chloride, sulfate, hydrogen carbonate, potassium, ammonium, calcium, sodium and magnesium ions were simultaneously separated.[168] Two electrolyte systems based on imidazole-nitrate and copper(II)-ethylenediamine-nitrate have been applied for indirect UV detection at 214 nm.

Introducing an ion-exchange electrokinetic chromatography mechanism, simple simultaneous determination of inorganic anions and kinetically stable metal-PAR chelates has been obtained.[143] The addition of cationic polymers to the background electrolyte reverses EOF and decreases electrophoretic mobility of anions according to their abilities to associate with polymeric cations. Chromate was chosen as a background electrolyte anion because it enables indirect UV detection of transparent anions at 254 nm and does not affect the direct detection sensitivity of chelates at 490 nm. Rapid (4 min) separation of Co(II), Ni(II) and Fe(II)-PAR chelates and of 11 inorganic anions was obtained.

Benzylamine, imidazole, benzenesulfonic acid, sulfosalicylic acid, and pyromellitic acid were tested as the components of the BGE for simultaneous separation of cations and anions.[169] The electrolyte components and control of the migration velocity were found to be the main factors for the separation.

6 CONCLUSIONS AND FUTURE TRENDS

As discussed in the previous sections, CE is an analytical separation technique that has great potential for environmental analyses. This is reflected in the rapidly increasing number of publications on CE. A diversity of inorganic species in real samples has already been proved to be amenable to the analysis by CE, and more can be expected in the years ahead. CE procedures provide high separation efficiency and high-speed separation, and are simple enough to be used on a routine basis. In application, however, limitations imposed by the lack of detection sensitivity complicate the technique's use and impede its wide acceptance. To overcome sensitivity problems in the determination of metal ions, the coupling of CE with ICPMS is now being attempted. This is done as an alternative to the two research techniques competing with each other. By doing this, not only the sensitive determination of elements but also particular speciation information of each element can be obtained. Interface design, however, is still being developed.

The advantages of CE compared to IC are speed, resolution, and the lack of a need for gradient elution. Operating costs are considerably lower, since ion-exchange columns are usually expensive. However, when a comparison on the data was made, the precision, linearity, and concentration LOD of CE were usually inferior to those of IC for the same analysis.[170] However, CE has the capability of analysis in three different areas:

1. analysis in complex matrices,
2. simultaneous determination of inorganic and organic pollutants and
3. metal speciation with simple pretreatment of sample and simple procedure.

As such, this technique remains as a complementary technique for IC or HPLC.

Another future of CE is in creating a miniaturized (or micro-) total analysis system. CE contains the qualities of speed, sensitivity, and small sample volumes that are desirable for a micro-total analysis system.[170] Using microlithography on glass structures, CE has been carried out on a chip. These chips allow for low sample volumes, fast analysis times and good separation efficiencies and has a possibility to be developed for a portable system that would be particularly useful in field analysis of environmental samples.

ABBREVIATIONS AND ACRONYMS

BGE	Background Electrolytes
CDTA	Cyclohexane-1,2-diaminetetraacetic Acid
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CGE	Capillary Gel Electrophoresis
CIEF	Capillary Isoelectric Focusing
CITP	Capillary Isotachophoresis
CL	Chemiluminescence
CMC	Critical Micelle Concentration
CTAB	Cetyltrimethylammonium Bromide
CZE	Capillary Zone Electrophoresis
DAD	Diode Array Detector
DETA	Diethylenetriamine
ED	Electrochemical Detection
EDTA	Ethylenediaminetetraacetic Acid
EOF	Electroosmotic Flow
HIBA	α-Hydroxy-isobutyric Acid
HMH	Hexamethonium Hydroxide

HPLC	High-performance Liquid Chromatography
HQS	8-Hydroxyquinoline-5-sulfonic Acid
IC	Ion Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ITP	Isotachophoresis
LIF	Laser-induced Fluorescence
LOD	Limit of Detection
MEKC	Micellar Electrokinetic Capillary Chromatography
PAR	4-(2-pyridylazo)resorcinol
SPE	Solid Phase Extraction
TCDD	Tetrachlorodibenzo-*p*-dioxin
TITP	Transient Preconcentration Step
TTAB	Tetradecyltrimethylammonium Bromide
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Capillary Electrophoresis in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Cyanogen Chloride and Cyanogen Bromide Analysis in Drinking Water • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Atomic Spectroscopy **(Volume 11)**
Inductively Coupled Plasma/Optical Emission Spectrometry

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Ion Chromatography

REFERENCES

1. H. Shintani, J. Polansky, *Handbook of Capillary Electrophoresis Application*, Blackie Academic & Professional, London, 1997.
2. D.R. Baker, *Capillary Electrophoresis*, John Wiley & Sons, New York, 1995.
3. P. Jandik, G. Bonn, *Capillary Electrophoresis of Small Molecules and Ions*, VCH, New York, 1993.
4. S.F.Y. Li, *Capillary Electrophoresis*, Elsevier Science, New York, 1997.
5. J.O. Landers, *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, FL, 1996.
6. F. Foret, L. Krivankova, P. Bocek, *Capillary Zone Electrophoresis*, VCH, 1993.
7. E. Dabek-Zlotorzynska, E.P.C. Lai, A.R. Timerbaev, 'Capillary Electrophoresis: the State-of-the Art in Metal Speciation Studies', *Anal. Chim. Acta*, **356**, 1–26 (1998).
8. M. Macka, P.R. Haddad, 'Determination of Metal Ions by Capillary Electrophoresis', *Electrophoresis*, **18**, 2482–2501 (1997).
9. A.R. Timerbaev, 'Analysis of Inorganic Pollutants by Capillary Electrophoresis', *Electrophoresis*, **18**, 185–195 (1997).
10. B.G. Belen'kii, V. Yu, G.E. Belov, Kasalaiene, 'High-performance Capillary Electrophoresis in Environmental Monitoring', *J. Anal. Chem.*, **51**(8), 753–769 (1996).
11. A.R. Timerbaev, 'Strategies for Selectivity Control in Capillary Electrophoresis of Metal Species', *J. Chromatogr. A*, **792**, 495–518 (1997).
12. K. Sarmini, E. Kenndler, 'Influence of Organic Solvents on the Separation Selectivity in Capillary Electrophoresis', *J. Chromatogr. A*, **792**, 3–11 (1997).
13. E. Dabek-Zlotorzynska, 'Capillary Electrophoresis in the Determination of Pollutants', *Electrophoresis*, **18**, 2453–2464 (1997).
14. L. Song, Z. Xu, J. Kang, J. Cheng, 'Analysis of Environmental Pollutants by Capillary Electrophoresis with Emphasis on Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **780**, 297–328 (1997).
15. S.J. Kok, N.H. Velthorst, C. Gooijer, U.A.T. Brinkman, 'Analyte Identification in Capillary Electrophoretic Separation Techniques', *Electrophoresis*, **19**, 2753–2776 (1998).

16. P.D. Voegel, R.P. Baldwin, 'Electrochemical Detection in Capillary Electrophoresis', *Electrophoresis*, **18**, 2267–2278 (1997).
17. T.D. Staller, M.J. Sepaniak, 'Chemiluminescence Detection in Capillary Electrophoresis', *Electrophoresis*, **18**, 2291–2296 (1997).
18. C.E. MacTaylor, A.G. Ewing, 'Critical Review of Recent Developments in Fluorescence Detection for Capillary Electrophoresis', *Electrophoresis*, **18**, 2279–2290 (1997).
19. K. Sutton, R.M.C. Sutton, J.A. Caruso, 'Inductively Coupled Plasma Mass Spectrometric Detection for Chromatography and Capillary Electrophoresis', *J. Chromatogr. A*, **789**, 85–126 (1997).
20. S.C. Beale, 'Capillary Electrophoresis', *Anal. Chem.*, **70**, 279R–300R (1998).
21. R.L. St. Claire III, 'Capillary Electrophoresis', *Anal. Chem.*, **68**, 569R–586R (1996).
22. M. Mazereeuw, U.R. Tjaden, N.J. Reinhoud, 'Single Capillary Isotachophoresis–Zone Electrophoresis: Current Practice and Prospects, A Review', *J. Chromatogr. Sci.*, **33**, 686–697 (1995).
23. P.R. Haddad, 'Comparison of Ion Chromatography and Capillary Electrophoresis for the Determination of Inorganic Ions', *J. Chromatogr. A*, **770**, 281–290 (1997).
24. M. Albin, P.D. Grossman, S.E. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 489A–497A (1993).
25. W.R. Jones, P. Jandik, 'Controlled Changes of Selectivity in the Separation of Ions by Capillary Electrophoresis', *J. Chromatogr.*, **546**, 445–458 (1991).
26. M.P. Harrold, M.J. Wojtusik, J. Riviello, P. Henson, 'Parameters Influencing Separation and Detection of Anions by Capillary Electrophoresis', *J. Chromatogr.*, **640**, 463–471 (1993).
27. W. Beck, H. Engelhardt, 'Separation of non UV-absorbing Cations by Capillary Electrophoresis', *Fresenius' J. Anal. Chem.*, **346**, 618–621 (1993).
28. W. Beck, H. Engelhardt, 'Capillary Electrophoresis of Organic and Inorganic Cations with Indirect UV Detection', *Chromatographia*, **33**, 313–316 (1992).
29. F. Foret, S. Fanali, A. Nardi, P. Bocek, 'Capillary Zone Electrophoresis of Rare Earth Metals with Indirect UV Absorbance Detection', *Electrophoresis*, **11**, 780–783 (1990).
30. I.E. Valko, S.P. Porras, M.L. Riekkola, 'Capillary Electrophoresis with Wide-bore Capillaries and Non-aqueous media', *J. Chromatogr. A*, **813**, 179–186 (1998).
31. F.M. Matysik, 'Non-aqueous Capillary Electrophoresis with Electrochemical Detection', *J. Chromatogr. A*, **802**, 349–354 (1998).
32. K. Sarmini, E. Kenndler, 'Influence of Organic Solvents on the Separation Selectivity in Capillary Electrophoresis', *J. Chromatogr. A*, **792**, 3–11 (1997).
33. K. Altria, C. Simpson, 'Measurement of Electroendo-osmotic Flows in High Voltage Capillary Zone Electrophoresis', *Anal. Proc.*, **23**, 453–454 (1986).
34. X. Huang, J.A. Luckey, M.J. Gordon, R.Z. Zare, 'Quantitative Analysis of Low Molecular Weight Carboxylic Acids by Capillary Zone Electrophoresis/Conductivity Detection', *Anal. Chem.*, **61**, 766–770 (1989).
35. D.F. Swaile, C.L. Copper, M.J. Sepaniak, D.E. Burton, L.L. Powell, 'The Use of Fluorescent Probe Molecules for Detection in Capillary Electrokinetic Separations', *Talanta*, **41**, 1499 (1994).
36. E.S. Yeung, W.G. Kuhr, 'Indirect Detection for Capillary Separations', *Anal. Chem.*, **63**, 275A (1991).
37. B. Huang, J. Li, J.-K. Cheng, 'In-line Chemiluminescence Detection for Capillary Ion Analysis', *Anal. Chem.*, **68**, 2366 (1996).
38. S.Y. Liao, Y.C. Chao, C.W. Whang, 'Indirect Chemiluminescence Detection of Amino Acids Separated by Capillary Electrophoresis', *J. Chromatogr. A*, **736**, 247–254 (1996).
39. Y. Zhang, B. Huang, J.-K. Cheng, 'On-line Indirect Chemiluminescence Detection in Capillary Electrophoresis', *Anal. Chim. Acta*, **363**, 157 (1998).
40. D.F. Swaile, C.L. Copper, M.J. Sepaniak, D.E. Burton, L.L. Powell, 'The Use of Fluorescent Probe Molecules for Detection in Electrokinetic Separations', *Talanta*, **41**, 1499 (1994).
41. A. Durgbanshi, W.T. Kok, 'Capillary Electrophoresis and Electrochemical Detection with a Conventional Detector Cell', *J. Chromatogr. A*, **798**, 289–296 (1998).
42. S. Hu, Z. Wang, P. Li, J. Cheng, 'Amperometric Detection in Capillary Electrophoresis with an Etched Joint', *Anal. Chem.*, **69**, 264 (1997).
43. M. Harrold, J. Stillian, L. Bao, R. Rocklin, N. Adalovic, 'Capillary Electrophoresis of Inorganic Anions and Organic Acids Using Suppressed Conductivity Detection Strategies for Selectivity Control', *J. Chromatogr. A*, **717**, 371 (1995).
44. W.R. Jones, 'Capillary Ion Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, Chapter 6, 1997.
45. S. Valsecchi, G. Tartari, S. Polesello, 'Determination of Anions in Rainwater by Capillary Electrophoresis with Conductivity Detection', *J. Chromatogr. A*, **760**, 326–332 (1997).
46. K. Govindaraju, E.A. Cowley, D.H. Eidelmann, D.K. Lloyd, 'Microanalysis of Lung Airway Surface Fluid by Capillary Electrophoresis with Conductivity Detection', *Anal. Chem.*, **69**, 2793–2979 (1997).
47. C.W. Klampfl, M.U. Katzmayr, W. Buchberger, N. Basener, 'Determination of Low-molecular-mass Ionic Compounds in Electrodeposition Coatings by Capillary Electrophoresis with Conductivity Detection', *J. Chromatogr. A*, **804**, 357–362 (1998).

48. X. Huang, M.J. Gordon, R.N. Zare, 'Effect of Electrolyte and Sample Concentration on the Relationship between Sensitivity and Resolution in Capillary Zone Electrophoresis Using Conductivity Detection', *J. Chromatogr.*, **480**, 285 (1989).
49. B.L. Lin, L.A. Colon, R.N. Zare, 'Dual Electrochemical Detection of Cysteine and Cystine in Capillary Zone Electrophoresis', *J. Chromatogr.*, **680**, 1263 (1994).
50. T. Kappes, P. Schnierle, P.C. Hauser, 'Potentiometric Detection of Inorganic Anions and Cations in Capillary Electrophoresis with Coated-wire Ion-selective Electrodes', *Anal. Chem. Acta*, **350**, 141–147 (1997).
51. D. Mueller, I. Jelinek, F. Opekar, K. Stulik, 'A Conductometric Detector for Capillary Separations', *Electroanalysis*, **8**, 722–725 (1996).
52. M. Harrold, J. Stillian, L. Bao, R. Rocklin, N. Avdalovic, 'Capillary Electrophoresis of Inorganic Anions and Organic Acids Using Suppressed Conductivity Detection Strategies for Selectivity Control', *J. Chromatogr. A*, **717**, 371–383 (1995).
53. E. Mei, H. Ichihashi, W. Gu, S. Yamasaki, 'Interface for Coupling Capillary Electrophoresis to Inductively Coupled Plasma and On-column Concentration Technique', *Anal. Chem.*, **69**, 2187–2192 (1997).
54. Q. Lu, R.M. Barnes, 'Evaluation of an Ultrasonic Nebulizer Interface for Capillary Electrophoresis and Inductively Coupled Plasma Mass Spectrometry', *Microchem. J.*, **54**, 129–143 (1996).
55. J.A. Kinzer, J.W. Olesik, S.V. Olesik, 'Effect of Laminar Flow in Capillary Electrophoresis Model and Experimental Results on Controlling Analysis Time and Resolution with Inductively Coupled Plasma Mass Spectrometry Detection', *Anal. Chem.*, **68**, 3250–3257 (1996).
56. B. Michalke, P. Schramel, 'Coupling of Capillary Electrophoresis with ICP-MS for Speciation Investigations', *Fresenius' J. Anal. Chem.*, **357**, 594–599 (1997).
57. M.J. Tomlinson, L. Lin, J.A. Caruso, 'Plasma Mass Spectrometry as a Detector for Chemical Speciation Studies', *Analyst*, **120**, 283 (1995).
58. T. Hirokawa, Y. Hashimoto, 'Simultaneous Separation of Yttrium and Lanthanide Ions by Isotachophoresis', *J. Chromatogr. A*, **772**, 357–367 (1997).
59. T. Hirokawa, Q. Mao, 'New Method to Determine Stability Constants of Lanthanide–EDTA Complexes by Isotachophoresis', *J. Chromatogr. A*, **786**, 377–381 (1997).
60. T. Wang, S.F.Y. Li, 'Migration Behaviour of Alkali and Alkaline-earth Metal Ion–EDTA Complexes and Quantitative Analysis of Magnesium in Real Samples by Capillary Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr. A*, **707**, 343 (1995).
61. X. Xu, W.Th. Kok, J.C. Kraak, H. Poppe, 'Simultaneous Determination of Urinary Creatinine, Calcium and Other Inorganic Cations by Capillary Zone Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr. B*, **661**, 35 (1994).
62. P. Jandik, W.R. Jones, 'Optimization of Detection Sensitivity in the Capillary Electrophoresis of Inorganic Anions', *J. Chromatogr.*, **546**, 431–443 (1991).
63. M. Jimidar, T. Hamoir, W. Degezelle, D.L. Massart, S. Soykene, P. van de Winkel, 'Method Development and Optimization for the Determination of Rare Earth Metal Ions by Capillary Zone Electrophoresis', *Anal. Chim. Acta*, **284**, 217 (1993).
64. Q. Yang, J. Smeyers-Verbeke, W. Wu, M.S. Khots, D.L. Massart, 'Simultaneous Separation of Ammonium and Alkali, Alkaline Earth and Transition Metal Ions in Aqueous-organic Media by Capillary Ion Analysis', *J. Chromatogr. A*, **688**, 339 (1994).
65. Q. Yang, Y. Zhuang, J. Smeyers-Verbeke, D.L. Massart, 'Interpretation of Migration Behaviour of Inorganic Cations in Capillary Ion Electrophoresis Based on an Equilibrium Model', *J. Chromatogr. A*, **706**, 503 (1995).
66. F. Foret, S. Fanali, A. Nardi, P. Bocek, 'Capillary Zone Electrophoresis of Rare Earth Metals with Indirect UV absorbance detection', *Electrophoresis*, **11**, 780 (1990).
67. M. Koberda, M. Konkowski, P. Youngberg, W.R. Jones, A. Weston, 'Capillary Electrophoretic Determination of Alkali and Alkaline-earth Cations in Various Multiple Electrolyte Solutions for Parenteral Use', *J. Chromatogr.*, **602**, 235 (1992).
68. T.K. Creson, P.J. Monaco, E.M. Rasch, A.H. Hagardorn, K.E. Ferslew, 'Capillary Ion Analysis of Lithium Concentrations in Biological Fluids and Tissues of Poecilia (teleost)', *Electrophoresis*, **19**, 3018–3021 (1998).
69. M. Chen, R.M. Cassidy, 'Separation of Metal Ions by Capillary Electrophoresis', *J. Chromatogr.*, **640**, 425 (1993).
70. H.A.H. Billiet, P.E. Andersson, P.R. Haddad, 'Computer Optimization of Background Electrolyte Composition in the Separation on Metal Ions by Capillary', *Electrophoresis*, **17**, 801 (1996).
71. K.R. Cooper, R.G. Kelly, 'Sources of Variability in Cation Analysis for a Quantitative Exfoliation Corrosion Resistance Test', *J. Chromatogr. A*, **739**, 183–190 (1996).
72. Y. Shi, J.S. Fritz, 'New Electrolyte Systems for the Determination of Metal Cations by Capillary Zone Electrophoresis', *J. Chromatogr.*, **671**, 429 (1994).
73. Y. Shi, J.S. Fritz, 'Separation of Metal Ions by Capillary Electrophoresis with a Complexing Electrolyte', *J. Chromatogr.*, **640**, 473 (1993).
74. A. Weston, P.R. Brown, P. Jandik, W.R. Jones, A.L. Heckenberg, 'Effect of Electrolyte Composition on the Separation of Inorganic Metal Cations by Capillary Ion Electrophoresis', *J. Chromatogr.*, **193**, 289 (1992).
75. T.I. Lin, Y.H. Lee, Y.C. Chen, 'Capillary Electrophoretic Analysis of Inorganic Cations. Role of Complexing Agent and Buffer pH', *J. Chromatogr. A*, **654**, 167–176 (1993).

76. Y.H. Lee, T.I. Lin, 'Determination of Metal Cations by Capillary Electrophoresis. Effect of Background Carrier and Complexing Agents', *J. Chromatogr. A*, **675**, 227 (1994).
77. H. Chen, Y. Xu, Vanlente, M.P.C. Ip, 'Indirect Ultraviolet Detection of Biologically Relevant Organic Acids by Capillary Electrophoresis', *J. Chromatogr. B*, **679**, 426 (1996).
78. C. Haber, W.R. Jones, J. Soglia, M.A. Surve, M. McGlynn, A. Caplan, J.R. Reineck, C. Krstanovic, 'Conductivity Detection in Capillary Electrophoresis – A Powerful Tool in Ion Analysis', *J. Cap. Electrophor.*, **3**, 1 (1996).
79. T.I. Lin, Y.H. Lee, Y.C. Chen, 'Capillary Electrophoretic Analysis of Inorganic Cations. Role of Complexing Agent and Buffer pH', *J. Chromatogr. A*, **654**, 167–176 (1993).
80. F. Hang, J.L. Fasching, P.R. Brown, 'Speciation of Organotin Compounds by Capillary Electrophoresis Using Indirect Ultraviolet Absorbance Detection', *J. Chromatogr. B*, **669**, 103 (1995).
81. C. Vogt, S. Conradi, 'Complex Equilibria in Capillary Zone Electrophoresis and their Use for the Separation of Rare Earth Metal Ions', *Anal. Chim. Acta*, **294**, 145–153 (1994).
82. C. Francois, P. Morin, M. Dreux, 'Separation of Transition Metal Cations by Capillary Electrophoresis Optimization of Complexing Agent Concentrations (Lactic Acid and 18-crown-6)', *J. Chromatogr. A*, **717**, 393–408 (1995).
83. A. Tangen, W. Lund, R.B. Frederiken, 'Determination of Na^+, K^+, Mg^{+2}, Ca^{+2} in Mixtures of Seawater and Formation Water by Capillary Electrophoresis', *J. Chromatogr. A*, **767**, 311–317 (1997).
84. A.R. Timerbaev, 'Separation of Metal Ions by Capillary Electrophoresis: an Understanding of the Basic Principles', *J. Capillary Electrophor.*, **2**, 165–174 (1995).
85. M. Macka, P. Nesterenko, P. Andersson, P.R. Haddad, 'Separation of Uranium(VI) and Lanthanides by Capillary Electrophoresis Using On-capillary Complexation with Arsenazo III', *J. Chromatogr A*, **803**, 279–290 (1998).
86. S. Motomizu, M. Oshima, S. Matsuda, Y. Obata, H. Tanaka, 'Separation an Determination of Alkaline-earth Metal Ions as UV-absorbing Chelates with EDTA by Capillary Electrophoresis. Determination of Calcium and Magnesium in Water and Serum Samples', *Anal. Sci.*, **8**, 619–625 (1992).
87. A.R. Timerbaev, O.P. Semenova, J.S. Fritz, 'Advanced Possibilities on Multi-element Separation and Detection of Metal Ions by Capillary Zone Electrophoresis Using Precapillary Complexation. I. Separation Aspects', *J. Chromatogr. A*, **756**, 300–306 (1996).
88. A.R. Timerbaev, O.P. Semenova, 'Theoretical Estimation of Capillary Zone Electrophoresis Behaviour of Metal Complexes Using Multivariate Regression Analysis', *J. Chromatogr. A*, **690**, 141–148 (1995).
89. A.V. Pirogov, J. Havel, 'Determination of Platinum, Palladium, Osmium, Iridium, Rhodium, and Gold as Chloro-complexes by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **772**, 347–355 (1997).
90. R.L. Chien, D.S. Burgi, 'On-column Sample Concentration Using Field-amplified CZE', *Anal. Chem.*, **64**, 489A (1992).
91. L. Song, H. Chen, J. Cheng, J. Chin, 'Analysis of Environmental Pollutants by Capillary Electrophoresis with Emphasis on Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **780**, 297 (1997).
92. R.L. Chien, D.S. Burgi, 'Sample Stacking of Extremely Large Injection Volume in High-performance Capillary Electrophoresis', *Anal. Chem.*, **64**, 1046–1050 (1992).
93. M. Louwagie, T. Rabilloud, J. Garin, 'Use of Ethanolamine for Sample Stacking in Capillary Electrophoresis', *Electrophoresis*, **19**, 2440–2444 (1998).
94. S. Brown, J.H.T. Luong, O.H.J. Szolar, A. Halasz, J. Hawari, 'Cyclodextrin-modified Capillary Electrophoresis: Determination of Polycyclic Aromatic Hydrocarbons in Contaminated Soils', *Anal. Chem.*, **68**, 287 (1996).
95. M.I. Turnes, M.C. Mejuto, R. Cela, 'Determination of Pentachlorophenol in Water Samples by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **733**, 395 (1996).
96. G.B. Harland, G. McGrath, S. McClean, W.F. Smyth, 'Use of Large Volume Sample Stacking for Improving Limits of Detection in the Capillary Electrophoretic Determination of Selected Cations', *Anal. Commun.*, **34**, 9–11 (1997).
97. M.W.F. Nielen, 'Trace Enrichment of Environmental Samples in Capillary Zone Electrophoresis', *Trends Anal. Chem.*, **12**, 345 (1993).
98. L. Song, Q. Ou, W. Yu, G. Xu, 'Effect of High Concentrations of Salts in Samples on Capillary Electrophoresis of Anions', *J. Chromatogr. A*, **696**, 317 (1995).
99. L. Song, Q. Ou, W. Yu, L. Fang, Y. Yin, 'Determination of the Petroleum Tracers Nitrate and Thiocyanate in Subterranean Waters by Capillary Ion Electrophoresis', *J. Chromatogr. A*, **715**, 376 (1995).
100. Y. He, H.K. Lee, 'Large Volume Sample Stacking in Acidic Buffer for Analysis of Small Organic and Inorganic Anions by Capillary Electrophoresis', *Anal. Chem.*, **71**, 995–1001 (1999).
101. A.M. Carro-Diaz, R.A. Lorenzo-Ferreira, R. Cela-Torrijos, 'Capillary Electrophoresis of Methylmercury with Injection by Sample Stacking', *J. Chromatogr. A*, **730**, 345 (1996).
102. Z. Liu, P. Sam, S.R. Sirimanne, P.C. McClure, J. Grainger, D.G. Patterson, Jr., 'Field-amplified Sample Stacking in Micellar Electrokinetic Chromatography for On-column Sample Concentration of Neutral Molecules', *J. Chromatogr. A*, **673**, 125 (1994).

103. H. Susse, H. Muller, 'Pesticide Analysis by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **730**, 337 (1996).
104. M. Mazereeuw, U.R. Tjaden, N.J. Reinhound, 'Single Capillary Isotachophoresis–Zone Electrophoresis: Current Practice and Prospects, A Review', *J. Chromatogr. Sci.*, **33**, 686 (1995).
105. J. Cai, Z.E. Rassi, 'On-line Preconcentration of Triazine Herbicides with Tandem Octadecyl Capillaries–Capillary Zone Electrophoresis', *J. Liq. Chromatogr.*, **15**, 1179 (1992).
106. L. Gross, E.S. Yeung, 'Indirect Fluorometric Detection of Cations in Capillary Zone Electrophoresis', *Anal. Chem.*, **62**, 427–431 (1990).
107. A.R. Timerbaev, W. Buchberger, O.P. Semenova, G.K. Bonn, 'Metal Ion Capillary Zone Electrophoresis with Direct UV Detection of Transition Metals Using an 8-hydroxyquinoline-5-sulfonic Acid Chelating System', *J. Chromatogr.*, **630**, 379–389 (1993).
108. J. Xu, Y. Ma, 'Determination of Iron in Rainwater, Lakewater, and Tapwater by High-performance Capillary Electrophoresis after Precolumn Complexation with 1,10-Phenanthroline', *J. Microcol.*, **8**, 137–143 (1996).
109. M. Aguilar, X. Huang, R.N. Zare, 'Determination of Metal Ion Complexes in Electroplating Solutions Using Capillary Zone Electrophoresis with UV Detection', *J. Chromatogr.*, **480**, 427–431 (1989).
110. W. Buchberger, O.P. Semenova, A.R. Timerbaev, 'Metal Ion Capillary Zone Electrophoresis with Direct UV Detection: Separation of Metal Cyanide Complexes', *J. High Resolut. Chromatogr.*, **16**, 153–156 (1993).
111. M. Aguilar, A. Farran, M. Martinez, K. Spahiu, 'Determination of Gold(I) and Silver(I) Cyanide in Ores by Capillary Zone Electrophoresis', *J. Chromatogr.*, **635**, 127–131 (1993).
112. T. Wang, S.F.Y. Li, 'Migration Behaviour of Alkali and Alkaline Earth Metal Ion–EDTA Complexes and Quantitative Analysis of Magnesium in Real Samples by Capillary Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr. A*, **707**, 343–353 (1995).
113. A.R. Timerbaev, O.S. Semenova, G.K. Bonn, 'Capillary Zone Electrophoresis of Lanthanide Elements After Complexation with Aminopolycarboxylic Acids', *Analyst*, **119**, 2795–2799 (1994).
114. A.V. Pirogov, J. Havel, 'Determination of Platinum, Palladium, Osmium, Iridium, Rhodium, and Gold as Chloro-complexes by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **772**, 347–355 (1997).
115. O.P. Semenova, A.R. Timerbaev, R. Gagstadter, G.K. Bonn, 'Speciation Studies by Capillary Electrophoresis – Simultaneous Determination of Chromium(III) and Chromium(VI)', *J. High Resolut. Chromatogr.*, **19**, 177–179 (1996).
116. G.Y. Jung, Y.S. Kim, H.B. Lim, 'Simultaneous Determination of Chromium(III) and Chromium(VI) in Aqueous Solution by Capillary Electrophoresis with On-Column UV/VIS Detection', *Anal. Sci. Int. J.*, **13**, 3463 (1997).
117. N. Wu, W.J. Horvath, P. Sun, C.W. Huie, 'Speciation of Aluminum Using Capillary Zone Electrophoresis with Indirect UV Detection', *J. Chromatogr. A*, **635**, 307–312 (1993).
118. S. Conradi, C. Bvogt, H. Wittrisch, G. Knobloch, G. Werner, 'Capillary Electrophoretic Separation of Metal Ions Using Complex Forming Equilibria of Different Stabilities', *J. Chromatogr. A*, **745**, 103 (1996).
119. A.R. Timerbaev, O.P. Semenova, P. Jandik, G.K. Bonn, 'Metal Ion Capillary Electrophoresis with Direct UV Detection Effect of a Charged Surfactant on the Migration Behaviour of Metal Chelates', *J. Chromatogr. A*, **671**, 419–427 (1994).
120. F.B. Regan, M.P. Meaney, S.M. Lunte, 'Determination of Metal Ions by Capillary Electrophoresis Using On-column Complexation with 4-(2-pyridylazo)resorcinol Following Trace Enrichment by Peak Stacking', *J. Chromatogr. B*, **657**, 409–417 (1994).
121. B. Michalke, P. Schramel, 'Application of Capillary Zone Electrophoresis–Inductively Coupled Plasma Mass Spectrometry and Capillary Isoelectric Focusing–Inductively Coupled Plasma Mass Spectrometry for Selenium Speciation', *J. Chromatogr. A*, **807**, 71–80 (1998).
122. Y.M. Huang, C.W. Whang, 'Capillary Electrophoresis of Arsenic Compounds with Indirect Fluorescence Detection', *Electrophoresis*, **19**, 2140–2144 (1998).
123. M.M. Albert, C. Demesmay, J.L. Rocca, 'Analysis of Organic and Non-organic Arsenious or Selenious Compound by Capillary Electrophoresis', *Fresenius' J. Anal. Chem.*, **351**, 426 (1995).
124. B. Michalke, P. Schramel, 'Application of Capillary Zone Electrophoresis–Inductively Coupled Plasma Mass Spectrometry and Capillary Isoelectric Focusing–Inductively Coupled Plasma Mass Spectrometry for Selenium Speciation', *J. Chromatogr. A*, **807**, 71–80 (1998).
125. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, 'Speciation Arsenic Compounds in Drinking Water by Capillary Electrophoresis with Hydrodynamically Modified Electroosmotic Flow Detected through Hydride Generation Inductively Coupled Plasma Mass Spectrometry with a Membrane Gas-Liquid Separator', *J. Anal. At. Spectrom.*, **12**, 689–695 (1997).
126. M.L. Magnuson, J.T. Creed, C.A. Brockhoff, 'Speciation of Selenium and Arsenic Compounds by Capillary Electrophoresis with Hydrodynamically Modified Electroosmotic Flow and On-line Reduction of Selenium(VI) to Selenium(V) with Hydride Generation Inductively Coupled Plasma Mass Spectrometric Detection', *Analyst*, **122**, 1057–1061 (1997).
127. X.D. Tian, Z.X. Zhuang, B. Chen, X.R. Wang, 'Movable Reduction Bed Hydride Generation System as an

Interface for Capillary Zone Electrophoresis and Inductively Coupled Plasma Atomic Emission Spectrometry for Arsenic Speciation Analysis', *Analyst*, **123**, 899–903 (1998).
128. V. Majidi, N.J. Miller-Ihli, 'Two Simple Interface Designs for Capillary Electrophoresis–Inductively Coupled Plasma Mass Spectrometry', *Analyst*, **123**, 803–808 (1998).
129. Y. Liu, J.V. Lopez-Avila, J.J. Zhu, D.R. Wiederin, W.F. Beckert, 'Capillary Electrophoresis Coupled On-line with Inductively Coupled Plasma Mass Spectrometry for Elemental Speciation', *Anal. Chem.*, **67**, 2020 (1995).
130. K. Li, S.F.Y. Li, 'Speciation of Selenium and Arsenic Compounds in Natural Waters by Capillary Zone Electrophoresis After On-column Preconcentration with Field-amplified Injection', *Analyst*, **120**, 361–366 (1995).
131. C. Vogt, G. Werner, 'Speciation of Heavy Metals by Capillary Electrophoresis', *J. Chromatogr. A*, **686**, 325 (1994).
132. L. Lin, J. Wang, J. Caruso, 'Arsenic Speciation Using Capillary Zone Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr. Sci.*, **33**, 177–180 (1995).
133. C. Casiot, M.C.B. Alonso, J. Boisson, O.F.X. Donald, M.P. Gautier, 'Simultaneous Speciation of Arsenic, Selenium, Antimony, and Tellurium Species in Waters and Soil Extracts by Capillary Electrophoresis and UV Detection', *Analyst*, **123**, 1887–2893 (1998).
134. D. Schlegel, J. Mattusch, R. Wennrich, 'Speciation Analysis of Arsenic and Selenium Compounds by Capillary Electrophoresis', *Fresenius' J. Anal. Chem.*, **354**, 535–539 (1996).
135. J.F. Lopez-Sanchez, M.B. Amran, M.D. Lakkis, F. Lagarde, G. Rauret, M.J.F. Leroy, 'Quantitative Aspects of the Separation of Arsenical Species by Free Solution Capillary Electrophoresis', *Fresenius' J. Anal. Chem.*, **348**, 810–814 (1994).
136. I. Medina, E. Rubi, M.C. Mejuto, R. Cela, 'Speciation of Organomercurials in Marine Samples Using Capillary Electrophoresis', *Talanta*, **40**, 1631–1636 (1993).
137. C.L. Ng, H.K. Lee, S.F.Y. Li, 'Determination of Organolead and Organotin Compounds by Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **652**, 547–553 (1993).
138. K. Li, S.F.Y. Li, H.K. Lee, 'Determination of Organolead and Organotin Compounds in Water Samples by Micellar Electrokinetic Chromatography', *J. Liq. Chromatogr.*, **18**, 1325–1347 (1995).
139. K. Li, S.F.Y. Li, 'Determination of Alkyllead and Alkyltin Compounds in Solid Samples by Subcritical Fluid Extraction and MEKC', *J. Chromatogr.*, **33**, 309–315 (1995).
140. T. Soga, G.A. Ross, 'Capillary Electrophoretic Determination of Inorganic and Organic Anions Using 2,6-Pyridinedicarboxylic Acid: Effect of Electrolyte's Complexing Ability', *J. Chromatogr. A*, **767**, 223–230 (1997).
141. W. Ding, M.J. Thornton, J.S. Fritz, 'Capillary Electrophoresis of Anions at High Salt Concentrations', *Electrophoresis*, **19**, 2133–2139 (1998).
142. S.W. Park, K.H. Jin, J.H. You, T.J. Kim, K.J. Paeng, K.H. Kong, 'Analysis of Inorganic Anions by Capillary Zone Electrophoresis', *Anal. Sci.*, **13**(Suppl.), 243–246 (1997).
143. O.V. Krokhin, H. Hoshino, O.A. Shpigun, T. Yotsuyanagi, 'Use of Cationic Polymers for the Simultaneous Determination of Inorganic Anions and Metal-4-(2-pyridylazo)resorcinolato Chelates in Kinetic Differentiation-mode Capillary Electrophoresis', *J. Chromatogr. A*, **776**, 329–336 (1997).
144. T. Meissner, F. Eisenbeiss, B. Jastorff, 'Determination of Anionic Trace Impurities in Glycerol by Capillary Isotachophoresis with Enlarged Sample Load', *J. Chromatogr. A*, **810**, 201–208 (1998).
145. H. Siren, A. Maattanen, M.L. Riekkola, 'Determination of Small Anions by Capillary Electrophoresis Using Indirect UV Detection with Sulfonated Nitrosonaphthol Dyes', *J. Chromatogr. A*, **767**, 293–301 (1997).
146. A.J. Zemann, 'Sub-minute Separations of Organic and Inorganic Anions with Co-electroosmotic Capillary Electrophoresis', *J. Chromatogr. A*, **787**, 243–251 (1997).
147. C. Finkler, H. Charrel, H. Engelhardt, 'Permanent Coated Capillaries with Reversed Electroosmotic Flow for Anion Analysis', *J. Chromatogr. A*, **822**, 191–206 (1998).
148. T. Ehmann, K. Bachmann, L. Fabry, H. Rufer, M. Serwe, G. Ross, S. Pahlke, L. Kotz, 'Capillary Preconditioning for Analysis of Anions Using Indirect UV Detection in Capillary Zone Electrophoresis Systematic Investigation of Alkaline and Acid Prerinsing Techniques by Designed Experiments', *J. Chromatogr. A*, **816**, 261–275 (1998).
149. P. Doble, M. Macka, P.R. Haddad, 'Use of Dye as Indirect Detection Probes for the High Sensitivity Determination of Anion by Capillary Electrophoresis,' *J. Chromatogr. A*, **804**, 327–336 (1998).
150. P. Doble, P. Andersson, P.R. Haddad, 'Determination and Prediction of Transfer Rations for Anions in Capillary Zone Electrophoresis with Indirect UV Detection', *J. Chromatogr. A*, **770**, 291–300 (1997).
151. M.A.G.T. van den Hoop, J.J. van Staden, 'Determination of Phosphate in Natural Waters by Capillary Electrophoresis Influence of Solution Composition on Migration Time and Response', *J. Chromatogr. A*, **770**, 321–328 (1997).
152. S.A. Ochrle, 'Analysis of Anions in Drinking Water by Capillary Ion Electrophoresis', *J. Chromatogr. A*, **733**, 101–104 (1996).
153. D. Orta, P.D. Mudgett, L. Ding, M. Drybread, J.R. Schultz, R.L. Sauer, 'Analysis of Water from the Space Shuttle and Mir Space Station by Ion Chromatography and Capillary Electrophoresis', *J. Chromatogr. A*, **804**, 295–304 (1998).

154. J. Font, J. Gutierrez, J. Lalueza, X. Perez, 'Determination of Sulfide in the Leather Industry by Capillary Electrophoresis', *J. Chromatogr. A*, **740**, 125–132 (1996).
155. K. Fukushi, K. Watanabe, S. Takeda, S.I. Wakida, M. Yamane, K. Higashi, K. Hiiro, 'Determination of Bromide Ions in Seawater by Capillary Zone Electrophoresis Using Diluted Artificial Seawater as the Buffer Solution', *J. Chromatogr. A*, **802**, 211–217 (1998).
156. B. Westergaard, H.C.B. Hansen, O.K. Borggaard, 'Determination of Anions in Soil Solutions by Capillary Zone Electrophoresis', *Analyst*, **123**, 721–724 (1998).
157. V. Verhelst, J.P. Mollie, F. Campeol, 'Capillary Zone Electrophoretic Determination of Ionic Impurities in Silicon Products Used for Electronic Applications', *J. Chromatogr. A*, **770**, 337–344 (1997).
158. Th. Ehmann, K. Bachmann, L. Fabry, H. Rufer, S. Pahlke, L. Kotz, 'Optimization of the Electrokinetic Sample Electrophoresis for the Ultra Trace Determination of Anions on Silicon Wafer Surfaces', *Chromatographia*, **45**, 301–311 (1997).
159. A.H. Harakuwe, P.R. Haddad, 'Analytical Performance Characteristics of the Separation of Fluoride in Toothpaste by Capillary Zone Electrophoresis', *Anal. Commun.*, **34**, 67–68 (1997).
160. P. Wang, S.F.Y. Li, H.K. Lee, 'Simultaneous Determination of Monofluorphosphate and Fluoride in Toothpaste by Capillary Electrophoresis', *J. Chromatogr. A*, **765**, 353–359 (1997).
161. B. Tenberken, K. Bachmann, 'Development of New Methods for the Analysis of Single Cloud, Fog, and Raindrops by Capillary Electrophoresis', *J. Chromatogr. A*, **775**, 372–377 (1997).
162. E. Dabek-Zlotorzynska, M. Piechowski, F. Liu, S. Kennedy, J.F. Dlouhy, 'Routine Determination of Major Anions in Atmospheric Aerosols by Capillary Electrophoresis', *J. Chromatogr. A*, **770**, 349–359 (1997).
163. R.E. Lewis, E.S. Ahuja, J.P. Foley, 'Ion Analysis by Capillary Zone Electrophoresis with Indirect Injection: Applications in the Nuclear Power Industry', *Analyst*, **123**, 1465–1469 (1998).
164. M.A. Friedberg, M.E. Hinsdale, Z.K. Shihabi, 'Analysis of Nitrate in Biological Fluids by Capillary Electrophoresis', *J. Chromatogr. A*, **781**, 491–496 (1997).
165. S. Motellir, K. Gurdale, H. Pitsch, 'Sulfur Speciation by Capillary Electrophoresis with Indirect Spectrophotometric Detection: In Search of a Suitable Carrier Electrolyte to Maximize Sensitivity', *J. Chromatogr. A*, **770**, 311–319 (1997).
166. M. Biesaga, M. Kwiatkowska, M. Trojanowicz, 'Separation of Chlorine-containing Anions by Ion Chromatography and Capillary Electrophoresis', *J. Chromatogr. A*, **777**, 375–381 (1997).
167. P. Kuban, B. Karlberg, 'Simultaneous Determination of Small Cations and Anions by Capillary Electrophoresis', *Anal. Chem.*, **70**, 360–365 (1998).
168. A. Padaraukas, V. Olsauskaite, G. Schwedt, 'Simultaneous Separation of Inorganic Anions and Cations by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **800**, 369–375 (1998).
169. X. Xiong, S.F.Y. Li, 'Dual UV-absorbing Background Electrolytes for Simultaneous Separation and Detection of Small Cations and Anions by Capillary Zone Electrophoresis', *Electrophoresis*, **19**, 2243–2251 (1998).
170. C.S. Effenhauser, 'Integrated Chip-Based Microcolumn Separation Systems,' *Top. Curr. Chem.*, **194**, 51–82 (1998).
171. J.W. Olesik, J.A. Klnzer, S.V. Olesik, 'Capillary Electrophoresis Inductively Coupled Plasma Spectrometry for Rapid Elemental Speciation', *Anal. Chem.*, **67**, 1 (1995).

Inorganic Environmental Analysis by Electrochemical Methods

Güler Somer
Gazi University, Ankara, Turkey

1 Introduction **1**
2 Electrochemical Methods **2**
3 Air Analysis **4**
3.1 Reviews and Monographs 5
3.2 Sample Collection 5
3.3 Techniques 5
3.4 Sources 6
3.5 Accidents and Emergencies 7
3.6 Atmospheric Chemistry, Transport and Deposition 7
4 Water Analysis **7**
4.1 Reviews and Monographs 7
4.2 Sample Collection 8
4.3 Techniques 8
4.4 Sources 9
5 Soil and Sediment Analysis **13**
5.1 Reviews 13
5.2 Sampling 14
5.3 Techniques Used for Soil and Sediment Analysis 14
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

After the industrial revolution considerable emphasis was placed on monitoring organic, inorganic and organometallic substances in the environment. Environmental contaminants include products of combustion, vapors, gases, industrial solvents, dusts, agricultural fertilizers, pesticides, radioisotopes, noise, ultraviolet light and certain microwave electromagnetic radiations.

Contaminants that are present in strong enough concentrations to pose an immediate threat to life or to cause unmistakable cases of disease or acute poisoning are usually readily apparent, and experience has led to the development of adequate, but by no means perfect, environmental control measures. The effects of environmental contaminants are related to time as well as concentration or intensity. Time factors become increasingly important when low concentrations of contaminants do not produce ill effects until after months or years of exposure. Trace analytical methods, therefore, play a very important role in environmental analyses. They are required for the detection and identification of chemical pollutants present in the environment. It is essential that the analytical methods applied are suitable and that they yield reliable data. These methods have to have high sensitivity, as many toxic chemicals are present in environmental samples at microgram per liter levels. Furthermore they have to cover a large dynamic range, and most important they have to be inexpensive. Voltammetry fulfills all of these features and it can be used both in trace metal analyses and in the determination of organic compounds in environmental samples. This article will cover the determination of inorganic pollutants in air, water and soil, using electrochemical (EC) methods.

1 INTRODUCTION

In recent years considerable emphasis has been placed on monitoring chemical pollutants in the environment. Most people are exposed to a large number of chemicals emitted from different sources. This increased exposure is viewed, by some, as an inevitable side effect of technological advancement. For example, the inhabitants of cities are exposed to large concentrations of lead caused by the combustion of leaded petrol in cars. Although many countries have taken measures to prevent the accumulation of lead in the atmosphere, soils, waters, and so on, by abolishing or restricting the use of tetraethyllead as a gasoline additive, developing countries still suffer from a steady increase in lead levels. Even at relatively low concentrations, lead acts as a metabolic poison and accumulates in bones by replacing calcium. It inhibits the enzymes needed for the synthesis of hemoglobin owing to its strong interaction with SH groups. One of the toxic manifestations associated with an increase of lead level in the body in young children is a reduction in their learning ability. Despite this knowledge, it is only in the recent past that there has been any attempt to reduce the concentration of lead in petrol.

The contamination of the environment is widespread. Nriagu and Pacyna[1] have presented a worldwide inventory of industrial and municipal discharge of metals into soils and the aquatic ecosystem. The annual discharge exceeds the combined output of all radioactive and organic wastes generated each year as measured by the amount of water needed to dilute such wastes to drinking water standards. One of the important features that distinguishes metals from other pollutants is that they are not biodegradable. Once they have entered the environment, their potential toxicity is controlled to a large extent by their chemical form.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Toxic metals, such as Cd, Cu, Ni, Pb, Zn, and so on, and acid rain contribute substantially to the pollution of land and waters and cause increasing damage to the various components of both, especially forests. Determination of these ions in wood samples using inverse differential pulse (DP) voltammetry showed that Pb and Cd concentrations were higher in wood from polluted regions than in nonpolluted regions. The difference between natural changes and the effects of industrial development during the last 100 years is the amount and speed of alteration, which deprives the biological systems of any chance of adapting, albeit slowly, to the newly created conditions. Today the average adult human body contains about 10 times the amount of lead found in mummified Egyptian bodies.

Increasing environmental pollution by toxic metals and also by organometallic and organic pollutants calls for reliable analytical procedures for their control in air, water and in soil. There has been an increasing public awareness of the dangers of exposure to toxic substances. Obviously there is a need to control and manage the environmental burden caused by ''chemical pollutants''. Hence research to ascertain the levels and pathways of these environmental pollutants has become more important. Trace analytical methods play a very important role in environmental analysis. They are required for the detection and identification of ''chemical pollutants'' present in the environment. Trace analytical determinations, subsequent to sampling and sample preparation steps, are usually carried out using instrumental methods. However, in order to be suitable for trace analysis, these methods must fulfill certain basic requirements. They need to have high sensitivity, as many toxic chemicals are present in environmental samples at microgram per liter levels. Furthermore, the dynamic range should be large, covering a concentration range typically from several hundred microgram per liter down to microgram or even picogram per liter levels. In situations where environmental levels of hazardous chemicals are to be established, an extension of the detection limit down to nanogram concentrations is essential. A further requirement of any instrumental method is that it has the necessary selectivity to deal with the problem in hand. The most important consideration in trace analysis is accuracy; that is, the closeness of agreement between the observed result and the known or true value. There now exist a large number of instrumental methods, but not all incorporate the aforementioned requirements. Among the most suitable are the EC methods of polarography and voltammetry. Voltammetry in particular has proved its usefulness both in the fields of trace metal analysis and in the determination of organic compounds in environmental samples.

2 ELECTROCHEMICAL METHODS

For environmental analysis, like every other kind of analysis, the need is for simple, reliable and viable techniques. Analyses are costly, and running costs particularly can be very high. The task of the analytical chemist is to choose the most appropriate procedure in order that the desired information about the particular material of interest can be provided. Next to advanced electroanalytical techniques, modern atomic absorption spectrometry (AAS), inductively coupled plasma atomic emission spectrometric and inductively coupled plasma mass spectrometric methods are commonly used in industrial, technical and control laboratories. Neutron activation analysis (NAA) is a powerful detection method but costly both in terms of investment and operation expenditure. The features that make electroanalytical methods competitive techniques are the very large useful concentration range ($1-10^{-10}$ M), the larger linear ranges available in many commonly used instrumental techniques, the ability to assay dissolved and solid samples, the ability to speciate on the basis of complex liability and to distinguish between different states of valency, and the immunity of the matrix to samples with high ionic content, and so on. Perhaps the most important point with respect to environmental analysis is that polarography and voltammetry are among the very few techniques that are equally suitable for analysis of inorganic, organometallic and organic compounds to concentrations below $10^{-9}-10^{-10}$ M, and surfactants using tensammetric techniques. Inductively coupled plasma mass spectrometry (ICPMS), for example, has a limited capability for the determination of organic pollutants. Modern electroanalytical methods are applicable to all toxic metals and metalloids of environmental and toxicological relevance. Metals of particular importance in freshwaters that can be easily determined by EC techniques are Cd, Cr, Cu, Hg, Ni, Pb and Zn.

Polarography and voltammetry rank today as low-cost techniques. The instrumental costs, degree of automation and the time required per determination are related in a complex way to the economics of the laboratory. Costs and expertise virtually exclude installation of an ICPMS system in a smaller routine laboratory. If the prices for purchase are compared (ICPMS costs 400 000–600 000 DM, AAS costs 25 000–150 000 DM, polarography and voltammetry cost 10 000–50 000 DM), it is clear that voltammetric instruments are much cheaper. The operating costs are very high for ICPMS due to the large amounts of Ar ($13-18\,L\,min^{-1}$) consumed; the costs are medium for AAS but again very low for voltammetric analyses.

Polarography and voltammetry are well-known EC methods which can be used to study solution composition through current–potential relationship. Polarography is

the branch of voltammetry that is used to investigate solution composition by the reduction or oxidation of ions or molecules at a dropping mercury electrode (DME) under the influence of an applied potential.[2] Direct current (DC) polarography involves measurement of the current variation at a DME as the applied potential ramp is linearly changed with time. As a result of the charging current at the DME a significant background current is observed, which limits the sensitivity of DC polarography to concentrations in the 100 $\mu g\,L^{-1}$ region.

An increase in the ratio of faradaic current to charging current has been accomplished by pulse polarography.[3] One of the most widely used pulse modes in trace analysis is DP polarography. In this mode, small potential pulses are superimposed over a conventional DC voltage ramp and applied to the DME near the end of the drop lifetime. The current is sampled just before application of the pulse and again at the pulse when the charging current has decayed. It is the difference between these two current measurements that is displayed. Hence the ability of DP polarography to discriminate against the charging current lowers its limit of detection over DC polarography. Because DP polarography is an extremely sensitive method, it is applicable to both inorganic and organic analysis, yielding peaks for concentrations in the microgram per liter to milligram per liter range.

There are at least three types of catalytic current which should not be intermixed:[4]

1. Catalytic reduction of hydrogen ions caused by a decrease in overpotential, resulting usually from interaction of hydrogen ions or other proton donors with the oxidized or reduced form of a catalyst adsorbed at the electrode surface. Numerous organic compounds containing sulfur or nitrogen (both acyclic and heterocyclic) are examples of such catalysts.
2. Catalysis of hydrogen evolution by platinum group metals, either in the solution or at the electrode surface.
3. Catalytic currents caused by a homogeneous chemical reaction, e.g. to a reoxidation of a metal ion, reduced electrochemically to a lower oxidation state. Such catalytic currents increase with increased concentration of the oxidizing agent and were observed in particular in the presence of Fe(III)/Fe(II) and Ti(IV)/Ti(III) couples as well as some complexes of Mo(VI) and Mo(V).

To increase the sensitivity above that offered by pulse and square-wave techniques, an accumulation of the species to be determined at the electrode surface is carried out. For this purpose voltammetric methods such as anodic stripping voltammetry (ASV), cathodic stripping voltammetry (CSV) or adsorptive stripping voltammetry (ADSV) employ stationary mercury electrodes or solid electrodes such as glassy carbon or gold. These methods are among the most sensitive voltammetric methods used in environmental analysis. Stripping voltammetry involves two steps.[5] In the first step, the analytical species of interest are deposited either as amalgams, as mercury salts or as adsorbates onto a stationary electrode such as a hanging mercury drop electrode (HMDE) or mercury film electrode (MFE). The time required for the deposition step depends on the concentration of the analyte present. It is important that the conditions used in the deposition step, such as deposition potential, deposition time and stirring rate, are strictly controlled to ensure reproducible results. The second step involves stripping the deposited species from the electrode back into solution. The most commonly used waveforms in the stripping step are the linear sweep and DP modes. The DP waveform has become more popular because it is more effective in discriminating against charging currents and permits lower detection limits. The magnitude of the stripping current is proportional over a certain range of concentrations and deposition times to the concentration of the species of interest.

So far, stripping voltammetry has proved to be a very useful technique in the determination at the subnanomolar concentration level of some 25 metals that deposit electrolytically and form stable amalgams with mercury. However, some heavy metals, i.e. Ni, Co, and Ti, do not possess these properties and, consequently, cannot be determined by conventional stripping voltammetry. Their determination has then been accomplished by ADSV.[6] In this technique, the preconcentration step is carried out by adsorbing a metal chelate at the surface of the electrode. The reduction current of either the metal ion or the ligand in the adsorbed complex is used to quantitate the surface bound species and this current is directly related to the metal concentration in the bulk of the solution. Since most metal–chelate complexes have a hydrophobic nature and show a marked tendency to adsorb onto the mercury electrode, this preconcentration scheme does not necessarily require that a preconcentration potential be applied to the electrode. This fact enables the avoidance of many interferences caused by electroactive impurities present in the sample. On the other hand, the presence of surfactants should be avoided for they might compete for the adsorption sites of the electrode, thus decreasing the stripping signal. ADSV has been applied to the simultaneous determination of Ni and Co in a variety of real samples including plants, fruits, body fluids[7] and rainwater,[8,9] using dimethylglyoxime (DMG) as chelator. Samples are adjusted to pH 9.2 with ammonia buffer followed by addition of DMG. A potential of

-0.7 V is then applied to the working electrode while the solution is stirred for several minutes. Subsequently, the adsorbed $Ni(DMG)_2$ and $Co(DMG)_2$ are reduced by scanning the potential in the negative direction. ASV could also be applied for the determination of selenium(IV) by using the adsorptive properties of selenium–TGA (thioglycolic acid) complex in 4 M HCl.[10] Here, the deposition was carried out for 2 min at -0.1 V and the response was recorded by scanning the potential from -0.1 to -0.7 V. A reduction peak at -0.46 V was observed and 2×10^{-8} M Se(IV) could be determined. Linear scan voltammetry can be applied to analyte concentrations of 1 $\mu g\,L^{-1}$, whereas the DP mode is used for concentrations down to 1 $ng\,L^{-1}$. The sensitivity of ADSV is greatly enhanced by the fact that the stripping step proceeds from the adsorbed layer rather than from the whole of the electrode. For this reason the use of this technique has been extended to the analysis of metals that traditionally had been determined by ASV.

In stripping voltammetry, the sensitivity of the method depends very much on the type of working electrode used. An HMDE may be used for the analysis of trace metals down to several micrograms per liter, an MFE on the other hand may be employed for even lower concentrations.[11] When mercury electrodes are used in ASV, about 20 amalgam-forming metals may be determined.[12] By contrast, metal ions such as selenium, tellurium, arsenic, mercury and silver may be determined using solid electrodes such as the glassy carbon electrode (GCE) or gold.[12] A chemically modified GCE in conjunction with differential pulse anodic stripping voltammetry (DPASV) has been used to measure Cu at concentrations as low as 100 $ng\,L^{-1}$ in laboratory prepared solutions.[13] The properties of the dialysis membrane and the advantages of the MFE have been brought together in a rotating glassy carbon membrane covered MFE.[14] The use of this electrode enabled detection of 97 $ng\,L^{-1}$ of Cd in aqueous solutions when the DPASV mode was employed, the response being linear over three orders of magnitude.

Carbon paste chemically modified electrodes are very simple to prepare and have been used to determine Ni(II) in a variety of reference materials, including water, coals and sediments.[15] Here Ni(II) was chemically deposited on the modified electrode and the DP voltammetric determination was carried out after transferring the electrode to a separate buffer. After a 240-s preconcentration period, up to 0.050 μM Ni could be detected. A novel strategy for electrode modification has been the incorporation of a microorganism as a modifier in a carbon paste electrode.[16] The common brown alga shows a strong affinity towards copper and advantage was taken of this fact to bioaccumulate Cu(II). A further example is the development of a gold film microelectrode formed by plasma-sputtering gold onto carbon fibers.[17] This electrode has been used for the detection of mercury using DPASV, with a detection limit of 3.7 $\mu g\,L^{-1}$.

Catalytic polarography is used also for increasing the sensitivity of the measurement. The catalytic current is frequently much greater than the diffusion current. It may be used in different situations, such as catalytic cathodic stripping, catalytic ADSV, catalytic current polarography or catalytic hydrogen wave polarography.

Reviews of the environmental determination of inorganic species include fluoride,[18] organometallic compounds,[19] and total arsenic and arsenic compounds using EC methods.[20] Some inorganic methods included advanced electroanalytical techniques[21] and EC sensors.[22] The application of various electroanalytical techniques to environmental analysis has been reviewed.[23–25] The various techniques used include chronopotentiometric stripping,[26] voltammetry,[27] and polarography.[28] Many reviews on the environmental occurrence, properties, and determination of specific analytes have also been published. Voluminous reviews on environmentally significant metals have appeared in three volumes.[29–31] Other reviews have appeared on the electroanalytical determination of inorganic species in the environment,[32] and the chromatographic analysis of metals,[33] mercury and its compounds,[34] arsenic species,[35] organolead compounds,[36] cadmium,[37] and sulfur compounds.[38]

3 AIR ANALYSIS

The atmosphere is a protective blanket for the continuation of life on the earth; it is the source of carbon dioxide for plant photosynthesis and of oxygen for respiration. It provides the nitrogen that nitrogen-fixing bacteria and ammonia-manufacturing plants use to produce chemically bound nitrogen, an essential component of life molecules. As a basic part of the hydrological cycle the atmosphere transports water from the oceans to land. Unfortunately, the atmosphere has also been used as a dumping ground for many pollutant materials, ranging from sulfur dioxide to the refrigerant Freon which causes damage to vegetation and materials, shortens human life, and alters the characteristics of the atmosphere itself. The atmosphere absorbs most of the cosmic rays from outer space and protects organisms from their effect.

Protection of public health has long been an important force for the control of air pollution, beginning with the dramatic increases in daily deaths during the episodes of severe air pollution between 1930 and the early 1960s. Early measurements of air pollution in cities focused on two pollutants: sulfur dioxide and particulate matter. Present-day air quality control recognizes several

additional pollutants as potentially important to health, including oxidants and ozone, carbon monoxide, nitrogen oxides, lead, organics such as formaldehyde, and various aerosol constituents. Atmospheric aerosols are solid or liquid particles smaller than 100 μm in diameter. Pollutant particles in the 0.001–10-μm range are commonly suspended in the air near sources of pollution, such as the urban atmosphere, industrial plants, highways, and power plants. For the most part, aerosols consist of carbonaceous material, metal oxides and glasses, dissolved ionic species (electrolytes), and ionic solids. Some of the metals found predominantly as particulate matter in polluted atmospheres are known to be hazardous to human health. All of these except beryllium are so-called ''heavy metals''. Lead is the toxic metal of greatest concern in the urban atmosphere because it comes closest to being present at a toxic level; mercury ranks second. Others include beryllium, cadmium, chromium, vanadium, nickel, and arsenic.

3.1 Reviews and Monographs

A monograph covers several aspects of aerosol characterization.[39] The application of several microanalytical techniques for characterization of individual environmental particles has also been discussed.[40]

Stanley-Wood and Lines[41] edited a book on particle size analysis. Smoley[42] authored a text on methods of determining metals in environmental samples. Merian[30] published a review on the occurrence, analysis, and biological significance of metals in the environment. Other reviews dealt with electroanalytical methods for Zn, Cd, Se, and As,[23,24] procedures for determination of arsenic,[35,43] and determination of cadmium in biological and environmental samples.[31]

There are many reviews of analysis of atmospheric environmental samples ranging over seven major subcategories including sample collection, radon measurements, airborne microorganisms, site-specific assessments, methods of analysis, field instrumentation, and composition of aerosols.

The subject of collecting samples of air was reviewed with respect to particulate sampling technology,[44] consideration of sampling specifics for compliance to USOSHA (United States Occupational Safety and Health Administration) standards,[45] and sampling indoor atmospheres in workplace air monitoring.[46] Substantial interest has been shown in actual methods of analysis for indoor air monitoring of environmental carcinogens,[47] organic vapors,[48] and tracer gases for ventilation studies.[49] Noteworthy reviews from India[50] and from China[51] were used to describe modern instrumental methods of atmospheric pollution analysis for toxic compounds. Extensive coverage has been given to reviews of aerosols.[52–59] Especially relevant here is the review of chemical composition of aerosols[58] and trace elements in indoor/outdoor dust.[60]

Air monitoring is a global concern as shown by reviews of air pollution determination by gaseous and particulate sampling methods,[61] the state of environmental analytical chemistry and pollution climates in Europe,[62] and the monitoring of atmospheric pollutants from space.[63] Various combinations of filter media, extraction and analysis techniques used in the determination of inorganic species have been compared.[64]

3.2 Sample Collection

Sample collection is probably the most important step in trace analysis. Meaningful results will only be obtained if all risks of contamination are either minimized or eliminated. This, however is not an easy task.

Sampling of airborne matter was the topic of a review.[65] The use of diffusive samplers to assess community air pollution problems was also discussed.[66]

Kasahara et al.[67] described the optimum sampling conditions for analysis of atmospheric aerosols by particle-induced X-ray emission (PIXE) methods.

In order to obtain time resolution for aerosol composition, samplers have been developed that can deposit aerosol particles at different locations on the collection surface. These samplers are referred to as streakers. PIXE has sufficient sensitivity to be able to analyze a small quantity of deposited aerosol and PIXE multielement analysis has been applied to aerosol characterization in Milan,[68] Hungary,[69] and with cascade impactor sampling in Vienna.[70] In one work[71] air samples from Ankara were collected on glass-fiber filters and were analyzed by differential pulse polarography (DPP).

3.3 Techniques

DPP has been successively applied to the determination of copper, lead, cadmium, zinc and titanium in air particulates.[71] Air samples were collected on glass-fiber filters and acid digested prior to analysis. Using EDTA (ethylenediamine tetraacetic acid)–sodium acetate electrolyte at pH 2.0, it was possible to determine all five metals in a single polarographic scan.

Stripping voltammetry has been used for the determination of trace metals in particulate matter. DPASV was used to determine Cd, Pb, Cu and Zn, while the detection of Se required differential pulse cathodic stripping voltammetry (DPCSV).[72] An HMDE was used with both stripping methods. The only sample pretreatment required was filtration through a filter to remove the coarse particulate matter and adjustment of the filtrate to

pH 2.0 with HCl. Trace metals bound to the airborne dust particles were determined by subjecting the collected particles, together with the filter media, to low-temperature ashing. In this way, the metal ions contained in the organic layer on the surface of the dust particles were digested completely and were accessible to subsequent DPASV measurement.

Adsorptive CSV has been applied to determinations of Cu, Cd, Ni, and Co in atmospheric samples.[73] For determination of Pb and Cd, samples were collected on glass-fiber filters and were analyzed by DPASV.[74]

For the trace metal determination of Pb and Cd, stripping voltammetry and measurements based on ion selective electrodes were used.[75] A review of EC sensors for monitoring environmental pollutants[76] presents a survey of advances in EC sensor technology which have attempted to improve the performance of these devices. Three main areas of development have been addressed: advances in sensor design and measurement techniques, novel approaches to increasing electrode selectivity and the use of microminiaturization and microelectronics fabrication techniques. An improved voltammetric procedure has been developed[77] for determination of Zn, Cd, Pb, and Cu.

3.4 Sources

The existence of metals in the environment is a result of both natural and artificial processes. Lithogenic origins, e.g. ore deposits and volcanoes, were the predominant sources of metal before the Industrial Revolution. However, the present-day levels of heavy metals in the environment are almost exclusively from anthropogenic sources.[78] These metals are released into the environment either directly by wastewater discharges and run-off, or indirectly via the atmosphere adsorbed onto precipitates and dust. Owing to the ubiquity of heavy metals in the environment, they can become incorporated in food chains and drinking water, eventually being taken up by humans.

Among the most hazardous heavy metals are Cd, Pb and Hg. These three metals are toxic at all concentrations and have no biological function.[79] A second group of metals, containing As, Bi, In, Sb and Tl, have no biochemical requirement by man and are only tolerated at trace levels in biochemical systems. Finally, there is a third group of metals which are known to be biochemically essential. Among these essential elements are Cu, Zn, Co, Ni, V, Se, Cr and Fe. Even this last group of metals, though essential below their respective threshold levels, become progressively toxic above certain levels. Some of these metals, e.g. Ni, Cr, Cu and Se, are known to display carcinogenic effects[80] due to their interaction with nucleic acids.

In general, the toxicity of metals stems from the fact that they are biologically nondegradable and have a tendency to accumulate in vital organs, e.g. brain, liver or bones of humans, and by their accumulation become progressively more toxic.[81,82] However, the degree of toxicity exerted on a biological system depends on the chemical form or oxidation state of the particular heavy metal. For example, studies have revealed that ionic copper is far more toxic towards aquatic organisms than copper bound by organic ligands, and that the more stable the copper complex, the lower is its toxicity.[83,84] It is also known that As(III) is much more toxic than As(V).[85]

Perhaps the most insidious forms of heavy metals are their alkylated derivatives. Many of such compounds, e.g. alkylleads and alkyltins, have been found in environmental samples.[86] Apart from anthropogenic sources, some of these compounds are produced naturally by chemical[87,88] or biological processes.[89,90] Consequently, many heavy metals released into the environment can become even more toxic once transformed to their alkylated forms. In particular, the alkyl compounds of mercury and lead are especially toxic because they are lipid-soluble.[91] Generally, the toxicity of alkylated species tends to increase with the degree of alkyl substitution and carbon chain length of the alkyl group.[86]

3.4.1 Fixed Sources

In the category of fixed or stationary sources of airborne environmental pollutants, a broad range of types and kinds of source encompasses industrial to natural sources such as minerals, soils, waters, some plants and volcanic eruptions. Industrial and fugitive emissions, incinerators, stacks, residues and nuclear plants are among the fixed sources of airborne pollutants. Municipal incinerators are those in which the fuel or feedstock principally originates from domestic waste including paper, wood, kitchen scraps, and so forth. In contrast to the strong interest in organic compounds found in emissions from municipal incinerators, few reports have been found on the subject of metals, especially toxic metals, in the flue gases from such incinerators. The interest in metals in such effluent or flue gases is restricted to mercury, hexavalent chromium, arsenic, gallium, lead and selenium.[92–95] The trace metals in emissions from coal-fired incinerators can be determined by AAS, NAA, PIXE as well as by electroanalytical methods.

3.4.2 Mobile Sources

Mobile sources, particularly vehicles with internal combustion engines, have been examined since the 1970s for the chemical composition of their exhausts. Various mobile sources were targeted for chemical characterization of airborne pollutants and these included exhausts

from combustion-based engines with gasoline[96,97] and with diesel fuels,[98–100] jet craft engines,[101] and cigarettes.[102,103] Cigarettes can also be regarded as a type of mobile airborne pollutant source both outdoors and indoors. Tobacco smoke condensates were analyzed (as an indoor pollutant) for metals such as As, Cd, and Pb[101] at toxicologically relevant concentrations and for Se.[102]

3.4.3 Ambient Air

Analysis of air in urban and industrial settings for inorganics,[104,105] principally metals, has been accomplished using several techniques such as PIXE, X-ray fluorescence measurement, NAA and also electroanalytical techniques. The sampling method of choice for practical applications was high-volume filtering of air with trapping of particles which are later analyzed. A comparatively high level of interest was taken in metal determination in nonurban or natural locations. Examples are analyses of marine aerosols for Ti, Cr, Fe, and Ni,[106] and determination of transition metals in aerosols at the Brazilian Antarctic Station and at a biomass burning site in the Amazon basin.[107] Other examples include determinations of metal ions in the auroral lower E-region[108] and of atmospheric aluminum in the air at the French Riviera compared with determinations of Pb, Cd, Cu and Zn in Saharan dust.[109] Zn, Cd, Pb and Cu ions in ambient air can be determined by ASV.[77]

3.5 Accidents and Emergencies

Air monitoring during accidents or emergencies has received little attention, presumably due to the unplanned or unscheduled nature of an emergency and the speed at which events can overtake systematic sampling and analysis. The fallout and environmental poisoning from the Chernobyl nuclear facility figured prominently in terms of ^{235}U in soil particles as a source of dust,[110] and of ^{90}Sr in air filters in the late postaccident period using gamma spectroscopy.[111]

Vehicle fires in traffic tunnels, a special type of emergency, were considered in terms of sampling strategies with a mixture of passive and active collectors.[112]

3.6 Atmospheric Chemistry, Transport and Deposition

Wet deposition includes the collection of airborne pollutants in condensed rather than vapor form within rain drops, snow, or liquid aerosols (fog).

Davidson[113] discusses the mechanisms of wet and dry deposition on snow. The determination of heavy metals is complicated by adsorption on the collection funnel walls[114] and by contamination of Cd and Zn from metal surfaces.[115] Programs and analytical methods for rainwater monitoring have been developed.[116,117]

Ultratrace determinations of heavy metals can be carried out in rainwater[118] and in snow and ice.[119]

4 WATER ANALYSIS

Throughout history, the quality of drinking water has been a factor in determining human welfare. Currently, waterborne toxic chemicals pose the greatest threat to the safety of water supplies in industrialized nations. The quality of groundwater is subject to a number of chemical influences. There are many possible sources of chemical contamination of water. These include wastes from industrial chemical production, metal plating operations, and pesticide runoff from agricultural lands. Some specific pollutants include industrial chemicals such as chlorinated hydrocarbons, heavy metals (including cadmium, lead and mercury), saline water, bacteria and general municipal and industrial wastes. Since World War II there has been a tremendous growth in the manufacture and use of synthetic chemicals. Many of the chemicals used or manufactured by industry have contaminated water supplies. Water pollution should be a concern of every citizen. Understanding the sources, interactions, and effects of water pollutants is essential for controlling pollutants in an environmentally safe and economically acceptable manner.

Some of the heavy metals are significant water pollutants. Cadmium, copper, lead and mercury ions bind to cell membranes, hindering transport processes through the cell wall. Heavy metals may also precipitate phosphate-containing biochemicals or catalyze their decomposition. Some of the metalloids, arsenic, selenium, and antimony, are also known as significant pollutants. Manufacture of inorganic chemicals has the potential to contaminate water with trace elements. Among the industries regulated for potential trace element pollution of water are those producing sodium dichromate, chlor-alkali, hydrofluoric acid, aluminum fluoride, chrome pigments, copper sulfate, nickel sulfate, sodium bisulfate, sodium bisulfite, sodium hydrosulfate, titanium dioxide, hydrogen cyanide and various plating industries.

4.1 Reviews and Monographs

There are an increasing number of articles appearing in the literature describing the analysis of trace pollutants in environmental waters. Toxicity data for many hazardous compounds were compiled by the European Chemical Industry Ecology and Toxicology Centre selection criteria and the structure of this database and its uses have been described.[120] Soluble filters can be used for preconcentration of trace elements in water.[121] Contamination in the course of analysis of trace metals can

be minimized.[122] Other topics discussed are instrumentation and automation used in wastewater treatment,[123] pollutant monitoring and determination,[124] as well as analytical methods for the determination of trace components.[125] Analytical techniques, as applied to water analysis, including in situ voltammetric measurements, were described in a review containing 154 references.[126]

4.2 Sample Collection

Sample collection is a very important step in determination of traces of metals in water. The first problem is to choose the type of sampling bottle that will be used. When the often preferred high-pressure (low-density) polyethylene bottles are used,[127] they should be degreased and leached by an acid to remove trace metals present.[127] Alternatively, Teflon™ bottles may be used as this is the only material that does not release zinc into a water sample.[128] Surface water samples taken in the open ocean cannot be taken directly from a large ship because the ship is continually releasing copper, lead and zinc from its antifouling paints and cathodic protection devices.[129] In such situations, the only safe way to sample is to use a small rubber dinghy which is rowed to the sampling site. Sampling is undertaken at the bow of the boat whilst slowly rowing into the wind to prevent contamination of the sampled water by the boat.[130] Collection of surface samples from coastal waters, estuaries, rivers and lakes is usually undertaken from small boats using a telescopic bar with an attached sample bottle.[127]

For the collection of samples at depths greater than 10 m the aforementioned methods are unsuitable. The normal procedure involves clamping a sampler to a hydro wire and lowering the sampler to the required sampling depth. However, this method is prone to sample contamination by rust and grease from the hydro wire.[131] This procedure can introduce a large amount of trace metal contamination, especially lead. These contamination problems can be circumvented by using the all-Teflon™ Patterson CIT sampler.[132] This is one of the few sampling devices that can collect deep seawater samples without metal contamination.

4.2.1 Storage of Water Samples

Storage methods employed depend on the nature of the measurements which are to be undertaken. For speciation studies, water samples should not be acidified as this would change the nature of the speciation.[133] Likewise, freezing of water samples is not possible. As a water sample freezes, metal ions are concentrated in the unfrozen liquid in the center of the container, where they may undergo chemical and physical alteration.[129] Thus, samples for speciation studies should be filtered through a 0.45-μm membrane filter to remove any suspended material. The filtrate, containing the dissolved heavy metal fraction, should subsequently be stored at 4 °C, as this appears to be the safest method for both freshwater and seawater samples.[134,135] Storage of samples at higher temperatures may result in speciation changes. This has been observed for cobalt and manganese when water samples were stored at room temperature.[136] Changes induced by a pH drop due to bacterial respiration, in which oxygen was consumed and carbon dioxide was released, resulted in precipitation of cobalt and manganese carbonates.

If the total concentration of dissolved heavy metals is to be determined in a water sample, then the filtered sample is usually adjusted to pH 2.0 by adding a small amount ($1\,mL\,L^{-1}$) of concentrated hydrochloric acid or perchloric acid.[6] Acidification of the sample helps prevent loss of metal by adsorption on the walls of the storage container. However, in some instances, acidification may cause leaching of metals, especially lead and zinc from the container walls.[127] The best method of sample storage is freezing to −20 °C or at liquid nitrogen temperature.[137] This eliminates leaching of trace metals from the container and also losses by adsorption.

When filtered samples are stored at room temperature for less than a week, acidification is unnecessary.[138] In the case of long-term storage, filtered samples, whether acidified to pH 2.0 or unacidified, have to be deep frozen.[138]

4.3 Techniques

Polarographic and voltammetric methods are mostly used in water analysis. Among the polarographic methods the most sensitive direct method of analysis is DPP. It has been used to determine the total chromium content of a river water sample.[139] Before DP polarographic analysis, all chromium present in the water sample was converted to Cr(VI) by oxidation with H_2O_2 followed by determination in 1 M NaOH. The result was to be in good agreement with that obtained by flame AAS.

Direct determination of As(III) and Pb(II) in the run-off water from waste disposal grounds has also been accomplished by DPP.[140] The sample pretreatment simply consisted of addition of a supporting electrolyte (2 M NaCl and 2 M H_2SO_4) followed by deaeration and recording the DP polarogram at a DME. The complete analysis took 20 min to perform. A high sulfuric acid content was required to ensure sufficient separation of the responses of Pb(II) and As(III). If As(V) is present, it must first be reduced to As(III) using hydrazine, before it can be determined. With this method, determination limits of $40\,\mu g\,L^{-1}$ or less are attainable. Generally,

direct determinations are only possible when the metal concentrations are sufficiently high. For example, the detection limit for cadmium in water matrices is $10\,\mu g\,L^{-1}$ when DPP is used.[141]

When surface active organic traces are present, the detection limit is restricted to higher concentrations. For the determination of ultratrace levels of metals, stripping voltammetry is required. The voltammetric mode most frequently used for the determination of metals at such ultratrace levels is DPASV. This method has been successfully applied for the determination of trace metals in rainwater, seawater, freshwater, saline water and snow samples. Ni and Co can be simultaneously determined in rainwater samples.[8,9] ASV, CSV and adsorptive stripping voltammetric methods are also among the most frequently used methods.[142–150]

Many new procedures for trace analyses include determinations of iron,[142] aluminum,[143] zinc,[144] platinum,[145] molybdenum,[146] copper,[147] titanium,[148] nickel and cobalt,[149] and cadmium, lead and copper,[150] in seawater,[142,143,144–150] rainwater[148] and river water[149] by using some chelating agents.

A development in environmental analysis has been the construction of on-line sensors for trace metals based on stripping analysis.[151] An automatic voltammetric analyzer for the determination of four trace metals (Cd, Pb, Cu and Zn) in water has been developed.[152] This analyzer was employed for the intermittent on-line analysis of tapwater supplied by water works, although it can also be applied to the analysis of wastewater samples. With this device, samples are automatically taken from the mains water supply, acidified and analyzed by DPASV at an HMDE. Each sample solution is analyzed twice followed by two standard additions for calibration purposes. The detection limit is $0.1\,\mu g\,L^{-1}$.

Another form of automated system is a flow injection analysis (FIA) using ASV for detection.[153] In this type of system, small volumes of sample solution are injected into a carrier stream, which transports them towards the detector. Deposition occurs while the sample is flowing through the detector and the stripping step is effected after the sample volume has passed the detector. The most commonly used electrodes are the HMDE and MFE. Solid gold electrodes are used for monitoring metals with oxidation potentials more positive than that of mercury.[154] The stripping mode utilized depends on the analyte concentration. The linear mode is used to monitor in the $5–50\,\mu g\,L^{-1}$ concentration range. In this concentration range, short deposition periods (1–3 min) are employed so that the analysis takes 2–4 min, which means that carrying out 15–30 measurements per hour is possible. At lower concentrations the DP mode is used, but this considerably lengthens the time of each analysis, given the slow scan rate ($2–5\,mV\,s^{-1}$) in the stripping stage. An automated ASV flow system, based on a mercury-coated graphite tubular electrode, has been applied to the continuous monitoring of copper and zinc in marine environments.[155]

Another type of a flow-through detection system for the determination of trace metals involves combining high-performance liquid chromatography (HPLC) with EC detection. An example of such an approach is the selective determination of copper in tapwater using reversed-phase HPLC.[156] It involves complexation of copper with dithiocarbamate (DTC) ligands. The preferred method for formation of the $Cu(DTC)_2$ complex was inclusion of a DTC salt in the mobile phase and injection of the aqueous sample onto the column to form the complex in situ. Determination of the copper complex, using the oxidative process, was undertaken at a gold electrode. The DC response of the electrode was monitored at +0.6 V. The results by this HPLC/EC detection method were compared with those obtained by AAS, and excellent agreement has been found for results obtained by these two methods. The use of HPLC with EC detection has also been extended to multielement determinations. By forming diethyldithiocarbamate complexes of Cu, Ni, Cr and Co it was possible to determine simultaneously all four metals.[157] When pyrrolidine DTC was used as the complexing ligand, Cd, Cu, Co, Pb, Hg and Ni were determined.[158]

Other electroanalytical methods used in trace metal analysis include anodic stripping coulometry[159] and (chrono)potentiometric stripping analysis (PSA),[160,161] which is a relatively new analytical method, but resembles voltammetric stripping analysis by involving two discrete steps. In the first step, the analytes are deposited at or dissolved in a suitable electrode by application of an appropriate potential. In the second step, the preconcentrated analyte is stripped at a constant or zero current and change in the potential of the working electrode is recorded as a function of time. Both oxidative and reductive PSA methods have now been developed.[162] The APASA on-line analyzer by Tacussel is designed for control in water processing plants. It automatically performs on-line polarographic determination of trace amounts of Ni, Zn, Cd, Pb, Cu and Cr in primary waters of rivers, estuaries, seawater and so on. A built-in ultraviolet mineralization stage allows the system to treat waters loaded with organic material directly. The water is filtered before being pumped into the EC system for final sample preparation and analysis. Metrohm has recently marketed an automatic VA Trace Analyser that incorporates an autosampler.

4.4 Sources

The role of surface waters, rivers and lakes as pollutant transporters, as well as analyses of groundwater, drinking

water, seawater, municipal wastewater and industrial wastewater will be discussed in this section.

4.4.1 Surface Water, Rivers and Lakes

Interest in surface waters has increased, and this trend reflects an increased awareness of the importance of lakes and rivers in the transport of pollutants. Inorganic analytes were studied most often. Examples are the use of DPP in determination of the total chromium content of river water[139] or that of DPASV for a simultaneous determination of Tl(I) and $(CH_3)_2Tl(I)$ ionic species in buffered lake water samples with estimated detection limits of 3.2 ppb and 3.4 ppb, respectively.[163] Since the consequences of mixing organothallium compounds with mercury salts were unknown, the plating of a thin mercury film on a GCE was performed separately from the sample deposition step, rather than using in situ film deposition and compound accumulation. The other heavy metal ions such as Pb(II), Zn(II) and Cd(II) interfere in the determination of the thallium species, so that an addition of EDTA is necessary, which results in a slight shift of the peak potentials to more negative values and in a decrease in the peak current due to Tl(I) ion. One of the simplest procedures based on an ASV measurement has been developed for the speciation of copper in river water.[164] In this scheme samples of water were adjusted to pH 8.0 and the copper content measured by DPASV. Following this step the sample was acidified to pH 2.0 and the copper concentration was again determined. There was a tenfold increase in the measurable copper concentration on decreasing the pH of the water sample. The smaller stripping current at high pH was attributed to the nonlabile nature of the copper species present. The nickel and cobalt level of the surface water of the Mediterranean Sea was determined by adsorptive cathodic stripping square-wave voltammetry.[165] The concentration of the dissolved cobalt ions in the surface water was 0.168 nM but it decreased to 0.045 nM at a depth of about 300 m.

4.4.2 Groundwater

Two reviews[166,167] and a monograph covering the problem of groundwater contamination[168] have been published. Some other reports have described the monitoring program at Savannah River,[169] the quality control program at Hanford,[170] and an overview of statistical methods for groundwater monitoring.[171] Monitoring data from 500 waste disposal sites showed that volatile organic compounds are the most abundant class of organic contaminants in groundwater. Adsorptive catalytic stripping voltammetry was used for the determination of total chromium in groundwater with a detection limit of 1 ppt.[172] DPP and differential pulse catalytic adsorptive stripping voltammetry (DPCASV) techniques were used for the speciation analysis of organic and inorganic Se compounds in natural waters. The detection limit was $10\,\mu g\,L^{-1}$ in the DPP analysis and $0.04\,\mu g\,L^{-1}$ in the DPCASV analysis.[173]

4.4.3 Drinking Water

Contamination of drinking water and the natural environment by toxic metals is a serious problem. In the USA alone there are over 30 000 abandoned chemical dump sites, many of which are leaking various pollutants into sources of local waters. This provides a strong incentive for the development of continuous metal analyzers. Monitoring of drinking water is demanding since the permissible metal levels are very low, generally at parts per billion levels (US Environmental Protection Agency recommended standards are Pb 0.005 ppm and Cu 0.02 ppm). Using a portable (chrono) potentiometric stripping analyzer, Cu(II) at 0.1–5 ppm and Pb(II) at 1–50 ppb could be detected in a 0.4-mL drinking water sample.[174]

4.4.4 Seawater

Great interest has been shown in the analysis of seawater, as reflected by the development of new methods for the determination of inorganic species. Speciation studies of Cu and Ni use competitive ligand equilibration CSV, DPASV and graphite furnace AAS.[175] Mercury is determined by ASV with a GCE spin coated with a hexacosane derivative.[176] Arsenic is determined by CSV.[177]

A particularly important area of study is the quantification of dissolved metal species in both freshwater and seawater samples. Electroanalytical determination of the dissolved metal content in these samples requires a limited sample pretreatment, which minimizes error sources arising from contamination. The usual procedure involves sample filtration through a 0.45-μm filter to remove particulate matter present and acidification of the filtrate to pH 2.0 with concentrated HCl. Samples are usually left to stand for several hours to allow the completion of decomposition of most complexes present. The metal ions Cu, Cd and Pb are then simultaneously determined by DPASV at an MFE, with the possibility of determining these three metals to levels as low as $1\,ng\,L^{-1}$.[178] Because of the lower hydrogen overvoltage of the MFE, less negative plating potentials of −1.0 V are applied in media at pH 2.0. Hence, zinc which is reduced at more negative potentials has to be determined separately at pH 4.5.[6]

ASV is especially suited for the analysis of seawater and other saline waters because the high-salt matrix

provides excellent conductivity, whereas it is a source of interference in AAS.[179] DPASV was employed for the simultaneous determination of Zn, Cd, Pb, Cu, Bi(II) and Sb(III) in different matrices such as seawater,[180,181] suspended particulate matter and plankton[182] at the microgram per liter level. A routine speciation method was also proposed[181] in a pollution survey of coastal seawater. A combination of DPASV at an HMDE and linear ASV was used with collection at a rotating ring disc electrode (RRDE) for identification of soluble Zn, Cd, Pb, Cu, Bi and Sb trace metal species. A limit of detection between 0.005 and 0.25 mg L^{-1} was obtained for antimony determination. A simple method that allows Sb(III) and Sb(V) to be determined selectively is based on CSV after accumulation as amalgam from hydrochloric acid solutions.[183] The disadvantage of this simple method is the interference of bismuth and copper. It is possible to determine dissolved antimony as well as Sb(III) in freshwater and seawater by CSV[184] in the presence of catechol with a detection limit for Sb of 0.16 nmol L^{-1}. Alternatively it is possible to differentiate antimony oxidation states using ADSV[185] with chloranilic acid as a complexing agent. Spiked seawater samples were analyzed and detection limits of 210 ng L^{-1} for Sb(III) with 5 min preconcentration and 560 ng L^{-1} for Sb(V) with 10 min preconcentration were reported. Another study revealed that antimony could be accumulated in the presence of triphenylmethane dyes[186] in order to develop more sensitive and selective methods for antimony speciation. Automated constant-current stripping analysis for Sb(III) and Sb(V) with gold-fiber working electrodes enabled analyses of river and seawater reference samples at microgram per liter levels.[187] DPASV using an MFE has been employed to determine the levels of dissolved Cd, Pb and Cu in coastal and inland waters.[188] This study revealed that Cd and Pb levels were usually elevated along main shipping routes, while in the immediate vicinity of these routes the levels dropped abruptly to the values of the local coastal waters. It was also found that areas where the water is rich in suspended particulate matter tended to have low dissolved metal concentrations, typically 5 ng L^{-1} Cd, 18 ng L^{-1} Pb and 130 ng L^{-1} Cu. This is related to the high metal uptake by plankton organisms and/or metal trapping due to chemisorption at suspended inorganic and dead organic matter surfaces.

A recent survey used DPASV at an MFE to determine total and dissolved concentrations of Cd distributed in coastal and open ocean waters.[189] This survey found that coastal waters contained higher cadmium concentrations than open ocean waters. Generally, cadmium was preferentially present in the dissolved state, although in heavily polluted waters with large amounts of suspended particulate matter, the total levels were higher than the dissolved contents. Because Cd was present preferentially in the dissolved state, it was suggested that it can be used as a suitable tracer when studying the movement of pollution plumes in coastal waters.

Another toxic metal frequently found in the environment is mercury. A method has been developed for the determination of mercury in water samples, after acidification of the sample to pH 1.0, by DPASV at an activated gold electrode.[190] The detection limit of the method was 50 ng L^{-1}. This method was found to be suitable for the simultaneous determination of copper and mercury. Furthermore, for high levels of copper, the gold electrode is more suitable than the mercury electrode, because problems associated with the limited solubility of copper as an amalgam in mercury are eliminated. For the determination of copper in seawater, medium exchange after the plating stage was necessary. This was required because the response corresponding to chloride oxidation interferes, and the chloride formed would damage the surface of the gold electrode. For the determination of ultratrace levels, below 30 to 50 ng L^{-1} in seawater, a method based on DPASV in the subtractive mode at a twin gold electrode was recommended.[191] The separation of the two halves of the gold disc by an insulating epoxy resin allows application of different potentials to both halves. Operation of the subtractive mode ensures a low background current. There was good correlation between results obtained by this method and those by cold vapor AAS. However, as the voltammetric response is complicated, it was suggested that for routine ultratrace determinations of Hg the application of cold vapor AAS was more convenient.

The voltammetric determination of mercury in seawater has been improved by using an epoxy resin-cast gold disc electrode,[192] which allows extremely low concentrations of the toxic metal to be detected (2 ng L^{-1}). Seawater samples collected off the Swedish coast in both the Baltic and North Seas were analyzed by DPASV with a gold disc electrode, yielding mercury concentrations of 8.8 and 7.0 ng L^{-1}, respectively, with a precision of about ±20%.

Adsorptive CSV has been applied for the determination of many trace elements in seawater. A review including 51 references by Van Den Berg[193] discussed the applications of this technique to analysis of seawater. The preconcentration step in this method can consist of precipitation of an insoluble salt, or adsorption of a surface-active complex, prior to the reduction of the deposited material. The former method gives an irreversible reduction current, and is therefore not very sensitive to low concentrations of trace elements. The latter procedure results in the formation of a monolayer of complexed species at the surface of the mercury electrode, which is reduced reversibly and completely, yielding a high sensitivity. A variety of ligands have been used to

form adsorptive complexes useful for the determination of trace elements in seawater. Among them catechol was found suitable for the determination of Cu, Fe, Sb, Sn, V at pH levels of 6 to 8,[194–196] DMG was used for the determination of Co, Ni and Pd at pH 8–10,[197] nioxime for that of Co at pH 7.5,[198] quinolin-8-ol for Cd, Cu, Mo, Pb and U at pH 7–9,[199] diethylenetriaminepentaacetic acid (DTPA) for Cr,[200] mandelic acid (MA) for Ti at pH of 3.0[201] and tropolone for Mo and Sn at pH 2.0[202] in seawater samples. In these procedures the reduction of the metal ion from the adsorbed complex took place in the stripping phase. Platinum in seawater was determined by the catalytic current of hydrogen evolution in 0.5 M sulfuric acid medium.[203] The total tin in estuarine waters was determined by adsorptive CSV using tropolone[204] as the complexing ligand. A deposition potential of −0.8 V and a reoxidation potential of −0.4 V were used. Starting at the latter, the current–voltage was next recorded toward more negative potentials. Samples had to be ultraviolet photolyzed prior to analysis in order to remove interfering surface-active organic compounds. The limit of detection was 5 pM tin using a 10-min stirred deposition prior to the scan.

The total tin content in the shipyard water was determined by DPP.[205] A 100-mL sample solution was evaporated until 5–7 mL remained, then Sn(II) was oxidized in acidic solution with H_2O_2 to Sn(IV). In an EDTA solution at pH 5 the interference of lead could be eliminated and tin level of the water was found to be 1.2×10^{-6} M. For the simultaneous determination of Co and Ni in seawater, adsorptive cathodic stripping square-wave voltammetry has been applied.[165] Here the complexes of cobalt and nickel with DMG are analyzed by a voltammetric method based on adsorptive collection at an HMDE. This method had greater sensitivity than the DPCSV method and a faster scan rate. At pH 8.0, with an adsorption potential of −0.7 V and adsorption time of 60–180 s, depending on the concentration of cobalt and nickel in the sample, and a scan rate of 140 mV s^{-1}, detection limits for cobalt and nickel were obtained at approximately 8 pM and 0.05 pM, respectively.

The accuracy of the method for the determination of Bi(III) in seawater at its natural pH of 8.1[206] is similar to those previously reported for strongly acidified samples.[207] This method employs the square-wave mode for the anodic potential scanning at a bare glassy carbon rotating-disc electrode instead of at a mercury-plated one. In seawater, the GCE is highly selective for bismuth in the presence of the other commonly found metal ions; only Cu(II) reacts, but bismuth can be determined even in the presence of a 100-fold concentration of Cu(II) by selecting an appropriate accumulation potential. This method is sensitive to bismuth ions (including hydrated Bi(III), oxy- and hydroxy-species and various labile complexes) and not to its total concentration. The bismuth content in a sample collected near the coast of Heligoland was found to be 12 ± 2 ng L^{-1}. The newly developed PSA method has been used for the determination of Pb, Cd and Zn in seawater using an MFE.[208]

4.4.5 Municipal Wastewater

Inorganic species of concern included Pb and Cd and free available chlorine and total residual chlorine. Several commercial instruments designed for field analysis of toxic metals have been introduced. A portable, semi on-line voltammetric analyzer developed by Chemtronics uses a glassy carbon working electrode, operates on 0.1–5-mL batch samples and provides automatic deaeration and system calibration. Targeted at industrial effluent and domestic water supply analysis, measurable metals include Sb, As, Bi, Cd, Cu, Au, Pb, Hg, Tl, and Zn over the concentration range down to 10 ppb.

4.4.6 Industrial Wastewaters

Studies on the analyses or composition of wastewaters from industry or the other facilities such as governmental laboratories can be divided conveniently into categories by product or activity. The pulp and paper industries, miscellaneous manufacturing industries, chemical industries, and plating or tannery and dye manufacture industries are the main sources of inorganic species causing water pollution. As the contamination of the natural environment by toxic metals is important, all water runoffs have to be monitored for many of the major pollutants. This is providing a strong incentive for the development of continuous metal analyzers. The other major incentive is the need to monitor effluents from existing waste generators such as electroplating plants, manufacturers of electronics products and some mining operations. In these applications two types of analytical measurement are generally required. In the first, monitoring of the performance of waste treatment equipment is carried out to detect a malfunction, for example saturation of an ion-exchange column. The second type of measurement is an end-of-pipe analysis to ensure that the plant discharge meets compliance. In the former case, fairly simple metal monitors which can provide an alarm response in the range 5–20 ppm are adequate whereas the requirements for discharge monitoring are much more rigorous. These may be required to indicate compliance with the local discharge standards which in California, for example, are 0.1 ppm for Pb and 0.5 ppm for Cu. Trace metals in wastewater streams are also routinely determined for wastewater management. The concentrations of the metals can vary considerably as a function of time because of the level and type of activity at the generating source. Thus on-line, near real-time measurement capabilities are highly

desirable. An example is an on-line voltammetric analyzer for trace metals in wastewater.[209] By this device, Cr, Ni, Zn, Cd, Pb and Cu were determined automatically with detection limits close to 50 $\mu g\,L^{-1}$ to characterize a wastewater stream.

For automated trace metal analysis on-line ASV is now the technique of choice. In the last few years there have been some major advances in cell design, instrumentation and control systems. An improved version of the wall-jet cell using microprocessor control has been used in continuous flow ASV,[210] while a similar approach using a thin-layer cell and flow injection has also been described.[211] Eutech Cybernetics[210] has developed a hand-held metal monitor, also based on the ASV principle. A novel feature of this system is the use of disposable electrode strips on which the electrodes have been screen printed. Designed as a dedicated metal monitor for copper or lead, each unit has high and low operating ranges: 0.01–0.7 ppm and 0.5–3.0 ppm. The main application of the Electrascan[210] is also for monitoring of drinking water and of industrial effluents. The use of disposable electrodes, as was discussed earlier, has some significant advantages including elimination of the need for electrode preconditioning which is a tedious feature of most stripping voltammetric methods. It also avoids problems of cross-contamination of samples which may occur in continuous flow systems. A two-point calibration is first carried out using one of the two electrodes in the Electrascan sample chamber, using calibrating standards. Next, one drop of the activating solution is added to a 1.0-mL sample, then a 100-μL aliquot is introduced into the cell chamber. The analysis is completed and the result displayed in approximately 100 s. Improvement in the design and fabrication of the DME by Novotny and Heyrovski[212] has provided a much more robust, versatile sensor with wide ranging applications in environmental monitoring. Potentiometry was used to determine Hg(II) in chloralkali effluents.[213]

5 SOIL AND SEDIMENT ANALYSIS

Metals are released into the environment either directly by wastewater discharges and runoff, or indirectly via the atmosphere adsorbed to precipitates and dust. Because of the large quantities of heavy metals in the environment, they can become incorporated in food chains and drinking water, eventually being taken up by humans. The lead content of roadside soil has been found to vary between 100 and 3000 ppm, depending on the location and the volume of traffic. In contrast, the average lead content of the earth's crust is about 16 ppm. Marten and Hammond[214] investigated three types of soil samples and found that samples near a battery smelter, near a road, and from a greenhouse contained 680, 59, and 12 ppm of lead, respectively. Most of the lead in soil exists in sparingly soluble forms. When 2784 ppm of lead nitrate were added to soil, it was found that after three days the soluble lead content was only 17 ppm. It is to be expected that all ions in nature will accumulate as their less soluble compounds, such as oxides, carbonates, silicates and sulfates, the relative proportions of each depending on the nature of the soil and on the solubility.

Meat, seafood, milk, eggs and many other foods are sources of many toxic elements. The amount of these elements is influenced by the element content of the soil. The minerals move from the soil to the plants that grow on it, then to the animals that eat plants. The concentration of selenium in human milk samples obtained from vegetarian women was significantly greater (22 $ng\,mL^{-1}$) than from nonvegetarian women (16 $ng\,mL^{-1}$).[215] The milk samples obtained from a farm which was near to a road supporting heavy traffic contained 60 $\mu g\,L^{-1}$ lead, whereas 22 $\mu g\,L^{-1}$ of lead was normally expected.[216]

5.1 Reviews

Although considerable work on the determination of metals in soils and sediments has been reported, most studies represent small improvements in the application of existing methodology rather than novel analytical approaches. In a natural aquatic environment heavy metals will be deposited in the surface layers of sediments. Although sediments contain the greatest concentration of metals, only a few studies on the speciation of heavy metals in sediments have been undertaken.[217] Chemometric techniques can be used to investigate to what extent soil sampling presents a representative picture of soil as a whole.[218] Three methods of fluvial sediment sampling were studied to determine how valid and representative is the geochemical information obtained from each technique.[219] Bottom sediments and trap sediments showed similar results for metals. Methods for the determination of heavy metals in soils and sediments were reported in many studies. The chemical speciation and environmental mobility of heavy metals were reviewed by Sager.[220] An interlaboratory study where the extraction techniques were specified but the measurement method was optional showed a strongly differing degree of comparability for the various elements, extraction techniques, and measurement techniques used by the different laboratories.[221] In other studies, the use of an ultrasonic slurry sampler[222] and the recovery efficiencies of heavy metals during acid digestion by different acid mixtures and by block heater and microwave heating[223] were reported. Developments in the analysis of estuarine and coastal sediments for heavy metals have been

reviewed.[224] Methods have also been developed for specific metals in soils and sediments.[225,226]

5.2 Sampling

Problems associated with techniques and strategies of soil sampling[227] and with the collection and preparation of soil samples for the Federal Soil Survey Laboratory Program have been reviewed.[228] Various sampling schemes for soils have been described.[229,230] Different sampling designs are needed, depending upon whether the soil contamination is expected to be ''spread'' over the whole area or exists in localized ''hot spots''.[231] A decision-support system for the sampling of aquatic sediments in lakes was described by Wehrens et al. and was applied to a real environmental problem.[229] Lame and Defize showed that the fundamental sampling error for soils affects the analytical variance only when sample sizes are less than 10 g.[230] For larger samples, the variance is determined by the segregation error. A sampling board method for estimation of the segregation error was described. A two-way compositing strategy can be used to attribute detected contamination in composited samples directly to constituent samples without further analyses.[232]

Evaluation of various soil and sediment samplers has been reported.[233,234] The sediment shovel proved highly practical, but was limited because small particles tend to be lost when the shovel is lifted.[233] A cryogenic sediment sampler was less convenient to use, but allowed the collection of nearly undisturbed samples. Another device has been suggested for the automatic subsampling of soil, sediment and plant material for proficiency testing.[234] In another study, freeze-sampling collects representative sediment samples, whereas grab-sampling introduces a bias in the textural composition of the 120-mesh fraction, caused by washout and elutriation of the finer fractions.[235]

5.2.1 Sample Preparation

The risks of sample contamination using inappropriate materials, containers and tools, as well as possible analyte loss during sample handling, have been discussed.[236] The influence of grinding procedures was discussed in another study, which found that the availability of some analytes was significantly influenced by the degree of grinding in some soils.[237]

Significant interest still exists in microwave extraction methods.[238–241] In one investigation, the microwave extraction of cadmium in a soil reference material gave results comparable to those found after using conventional extraction procedures.[238]

In another study, microwave versus conventional dissolution was compared for soils, sludges, sediments, and oils.[239] By microwave digestion of dust samples with a nitric acid/hydrofluoric acid mixture, over 90% of Pb and Cd were recovered within 30 min.[240] Microwave digestion procedures for the analysis of metal-contaminated soils were also reviewed.[241]

Other extraction procedures for metals in soils and sediments were studied. For the determination of lead in roadside soil, HCl, HNO_3, $HClO_4$, aqua regia and EDTA were used for the digestion.[242] Among them the digestion with a mixture of aqua regia and EDTA yielded the highest level of lead content. However, 90% extraction was possible when EDTA was used as the extraction medium. The EDTA extraction procedure was evaluated in a collaborative study between six laboratories.[243] All laboratories produced some extreme outlying results, but most results were in good agreement once the outliers were removed. Acceptable accuracy and precision were obtained for metals in soil using an ultrasonic bath digestion procedure.[244] An interlaboratory comparison study was reported involving 160 accredited hazardous materials laboratories.[245] In this study, each laboratory performed a mineral acid digestion on five soils spiked with As, Cd, Mo, Se, and Tl.

5.3 Techniques Used for Soil and Sediment Analysis

Square-wave CSV was used for the determination of selenium in soils and sediments.[246] Speciation of As(III) and As(V) in soils was accomplished under different pH conditions by CSV.[247] Other developments in the determination of heavy metals in soil and sediments were reported.[248,249] The lead content of roadside soil in Ankara was determined using ASV.[242] The deposition potential was -0.800 V in the presence of EDTA and $HClO_4$ at pH 2.0 and the deposition time was 5 min. After a rest period of 45 s, a potential sweep of 60 mV s^{-1} was applied in a positive direction and the current was recorded. Under these conditions, 10^{-7} M lead could be determined with better than 95% accuracy. As expected the lead content of the soil increased with increasing traffic volume, changing from 193 to 710 ppm. A convenient and accurate analytical procedure has been developed for the simultaneous determination of Cd, Cu, Pb, and Zn in soils.[250] This method is based on DPASV measurement, at an HMDE, of the analytes resulting from the digested soil. Some samples were analyzed by DPASV and AAS to ascertain the applicability of the developed procedure. Good agreement was found between the results obtained with both methods. Selenium has been analyzed in soils and plants by DPCSV at the HMDE.[251] Oxygen-flask digestion of the sample was followed by treatment with HCl to reduce Se(VI) to Se(IV). The voltammogram was then recorded and Se concentration was determined by means of its second

stripping peak. Results obtained by DPCSV showed reasonable agreement with those of fluorimetric analysis for Se content above $0.1\,\mu g\,g^{-1}$, and the detection limit of $5\,ng\,g^{-1}$ was similar for both methods. A simple and sensitive adsorptive catalytic stripping (voltammetric and (chrono)potentiometric) procedure for determining trace amounts of chromium in soil samples was developed by Wang et al.[252] A preconcentration time of 1 min using a preconcentration potential of $-0.82\,V$ (versus Ag/AgCl) resulted in a detection limit of $1\,ng\,L^{-1}$. The identical stripping response for Cr(III) and Cr(VI) solutions makes the method applicable to the measurement of the total Cr. Groundwater and soil samples from contaminated nuclear energy sites were also analyzed.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
ADSV	Adsorptive Stripping Voltammetry
ASV	Anodic Stripping Voltammetry
CSV	Cathodic Stripping Voltammetry
DC	Direct Current
DME	Dropping Mercury Electrode
DMG	Dimethylglyoxime
DP	Differential Pulse
DPASV	Differential Pulse Anodic Stripping Voltammetry
DPCASV	Differential Pulse Catalytic Adsorptive Stripping Voltammetry
DPCSV	Differential Pulse Cathodic Stripping Voltammetry
DPP	Differential Pulse Polarography
DTC	Dithiocarbamate
DTPA	Diethylenetriaminepentaacetic Acid
EC	Electrochemical
EDTA	Ethylenediamine Tetraacetic Acid
FIA	Flow Injection Analysis
GCE	Glassy Carbon Electrode
HMDE	Hanging Mercury Drop Electrode
HPLC	High-performance Liquid Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
MA	Mandelic Acid
MFE	Mercury Film Electrode
NAA	Neutron Activation Analysis
PIXE	Particle-induced X-ray Emission
PSA	Potentiometric Stripping Analysis
RRDE	Rotating Ring Disc Electrode
TGA	Thioglycolic Acid
USOSHA	United States Occupational Safety and Health Administration

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Laser-induced Breakdown Spectroscopy, Elemental Analysis • Laser Mass Spectrometry in Trace Analysis • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Luminescence in Environmental Analysis • Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Neutron Activation in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses • Soil Sampling for the Characterization of Hazardous Waste Sites • Solid-phase Microextraction in Environmental Analysis • Supercritical Fluid Extraction of Inorganics in Environmental Analysis • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

REFERENCES

1. J.O. Nriagu, J.M. Pacyna, *Nature (London)*, **333**, 134 (1988).

2. M. Bond, *Modern Polarographic Methods in Analytical Chemistry*, Marcel Dekker, New York, 1980.
3. G.C. Barker, A.W. Gardner, *Z. Anal. Chem.*, **173**, 79 (1960).
4. S.G. Mairanovski, *Catalytic and Kinetic Waves in Polarography*, Plenum Press, New York, 1968.
5. T.R. Copeland, R.K. Skogerboe, *Anal. Chem.*, **46**, 1257A (1974).
6. H.W. Nurnberg, *Anal. Chim. Acta*, **164**, 1 (1984).
7. A. Pihlar, P. Valenta, H.W. Nurnberg, *Fresenius' J. Anal. Chem.*, **307**, 337 (1981).
8. L. Vos, Z. Komy, G. Reggers, E. Roekens, R. Van Grieken, *Anal. Chim. Acta*, **184**, 271 (1986).
9. Z. Komy, E. Roekens, R. Van Grieken, *Anal. Chim. Acta*, **204**, 179 (1988).
10. R. Inam, G. Somer, *Anal. Sci.*, **13**, 653 (1997).
11. T.M. Florence, *J. Electroanal. Chem.*, **168**, 207 (1984).
12. J. Wang, *Environ. Sci. Technol.*, **16**, 104A (1982).
13. T. Miva, L. Jin, A. Mizuike, *Anal. Chim. Acta*, **160**, 135 (1984).
14. E.S. Stewart, R.B. Smart, *Anal. Chem.*, **56**, 1131 (1984).
15. R.P. Baldwin, J.K. Christensen, L. Kryger, *Anal. Chem.*, **58**, 1790 (1986).
16. J. Gardea-Torresdey, D. Darnall, J. Wang, *Anal. Chem.*, **60**, 72 (1988).
17. G.J. Svoboda, J.P. Sottery, C.V. Anderson, *Anal. Chim. Acta*, **166**, 297 (1984).
18. C.Y. Wang, J. Xu, *Fluoride*, **26**(3), 197–202 (1993).
19. Mennie, P.J. Craig, *Met. Ions Biol. Syst.*, **29**, 37–77 (1993).
20. H. Greschonig, K.J. Irgolic, *J. Appl. Organomet. Chem.*, **6**(7), 565–577 (1992).
21. P.M. Bersier, J. Howell, C. Bruntlett, *Analyst*, **119**(2), 219–232 (1994).
22. B. Fleet, H. Gunasingham, *Talanta*, **39**(11), 1449–1457 (1992).
23. B. Lichiang, D. James, R. Magee, *J. Sci. Int. (Lahore)*, **4**(1), 25–37 (1992).
24. B. Lichiang, D. James, R. Magee, *J. Sci. Int. (Lahore)*, **4**(1), 25–37 (1992).
25. R. Kalvoda, *Electroanalysis*, **2**(5), 341–346 (1990).
26. P. Hansen, *Am. Lab.*, **23**(6), 52, 54–56, 58 (1991).
27. J.M. Fernandez, P.J. Hayes, M.R. Smyth, *Comp. Anal. Chem.*, **27**, 345–379 (1992).
28. G. Henze, *NATO ASI Ser., Ser. G. (Met. Speciation Environ.)*, **23**, 391–408 (1990).
29. E. Merian (ed.), *Metals and Their Compounds in the Environment, Occurrence, Analysis, and Biological Relevance*, VCH, Weinheim, 1990.
30. E. Merian (ed.), *Metals and their Compounds in the Environment, Occurrence, Analysis, and Biological Relevance*, VCH, Weinheim, 1991.
31. M. Stoeppler, *Tech. Instrum. Anal. Chem., (Hazard. Met. Environ.)*, **12**, 1–8 (1992).
32. G. Svehla, M.R. Smith, J.G. Vos (eds.), *Comprehensive Analytical Chemistry*, Volume XXVII, Analytical Voltammetry, Elsevier, Amsterdam, 345–370, 1992.
33. K. Robards, P.J. Worsfold, *Trends Anal. Chem.*, **10**(4), 121–127 (1991).
34. B.S. Sengar, A. Kumar, *Indian J. Environ. Prot.*, **11**(11), 843–852 (1991).
35. IUPAC Commission on Microchemical Techniques and Trace Analysis, *Pure Appl. Chem.*, **64**(4), 575–590 (1992).
36. D. Chakraborti, *Chem. Environ. Res.*, **1**(1), 3–12 (1992).
37. K. Robards, P. Worsfold, *Analyst (London)*, **116**(6), 549–568 (1991).
38. W. Wardenski, B. Zugmunt, *Anal. Chim. Acta*, **255**(1), 1–13 (1991).
39. T. Kouimtzia, C. Samara, *Airborne Particulate Matter*, Springer, Berlin, 1995.
40. W. Jambers, L. De Bock, R. Van Grieken, *Fresenius' J. Anal. Chem.*, **355**, 521–527 (1996).
41. N.G. Stanley-Wood, R.W. Lines (eds.), *Particle Size Analysis, The Royal Society of Chemistry*, Cambridge, UK, 1992.
42. C.K. Smoley, *Methods for the Determination of Metals in Environmental Samples*, Lewis Publishers, Boca Raton, FL, 1992.
43. K.J. Irgolic, *Tech. Instrum. Anal. Chem.*, **12**, 287–350 (1992).
44. R. Neissner, *Tech. Instrum. Anal. Chem.*, **13**, 3–22 (1993).
45. R. Cee, J.C. Ku, *Analyst*, **119**(1), 57–63 (1994).
46. T. Norseth, *Analyst*, **119**(1), 3–7 (1994).
47. M.R. Peterson, D.F. Naugle, M.A. Berry, *ARC Sci. Publ.*, **109**(12), 126–135 (1993).
48. K. Leichnitz, *IARC Sci. Publ.*, **109**(12), 211–220 (1993).
49. C.Y. Shaw, *Indoor Environ.*, **2**(5–6), 374–382 (1993).
50. B. Basu, D.K. Tuli, S.K. Jain, A.K. Bhatnagar, *Indian J. Environ. Prot.*, **12**(6), 416–430 (1992).
51. P. Lu, X. Zhang, L. Yang, G. Xu, Y. Zhang, *Anal. Sci.*, **10**(2), 241–245 (1994).
52. J.H. Vincent, *Analyst*, **119**(1), 13–18 (1994).
53. T. Schneider, *Analyst*, **119**(1), 103–107.
54. D. Mark, C.P. Lyons, S.L. Upton, D.J. Hall, *Aerosol Sci.*, **23**(1), S611–S614 (1992).
55. J.H. Vincent, *Analyst*, **119**(1), 19–25 (1994).
56. P.A. Baron, *Analyst*, **119**(1), 35–40 (1994).
57. R.P. Streicher, E.R. Kennedy, C.D. Lorberau, *Analyst*, **119**(1), 89–97 (1994).
58. R.J. Harrison, D.J.T. Smith, *J. Aerosol Sci.*, **23**(1), S853–S856 (1992).
59. J.P. Mitchell, *J. Aerosol Sci.*, **23**(1), S289–S292 (1992).
60. Y. Feng, R.S. Barratt, *Int. J.Environ. Health Res.*, **3**(1), 18–31 (1993).
61. D.L. Fox, *Anal. Chem.*, **63**, 292R–301R (1991).
62. J. Slanina, *Analusis*, **20**, M15–M19 (1992).
63. J. Fishman, *Environ. Sci. Technol.*, **25**, 612–621 (1991).

64. J. Mateu, R. Forteza, V. Cerda, M. Colom-Altes, *Water, Air, Soil Pollut.*, **84**, 61–79 (1995).
65. W. Hollaender, *Handb. Environ. Chem.*, **4**, 143–173 (1995).
66. R.H. Brown, *Pure Appl. Chem.*, **67**, 1423–1425 (1995).
67. M. Kasahara, K. Yoshida, K. Takahashi, *Nucl. Instrum. Methods Phys. Res., Sect. B*, **75**, 240–244 (1995).
68. Manini, M. Maugeri, A. Novo, *Ing. Ambientale*, **24**, 410–419 (1995).
69. Borbely-Kiss, E. Koltay, G. Szabo, E. Meszaros, A. Molnar, L. Bozo, *Int. J. PIXE*, **4**, 243–257 (1994).
70. H. Horvath, M. Kasahara, P. Pesava, *J. Aerosol Sci.*, **27**, 417–435 (1996).
71. G. Somer, G. Özyörük, M.E. Green, *Analyst*, **110**, 151–153 (1985).
72. V.D. Nguyen, P. Valenta, H.W. Nurnberg, *Sci. Total Environ.*, **12**, 151 (1979).
73. M. Nimmo, G. Fones, *Anal. Chim. Acta*, **291**, 321–328 (1994).
74. D.Q. Nam, F. Skacel, P. Buryan, *Sci. Total Environ.*, **144**, 87–92 (1994).
75. K. Florian, M. Galova, L. Koller, E. Krakovska, L. Lux, M. Matherny, H. Nicke, N. Pliesovska, *Acta Chim. Hung.*, **129**, 611–617 (1992).
76. B. Fleet, H. Gunasingham, *Talanta*, **39**, 1449–1457 (1992).
77. E. Cassassas, A.M. Perez-Vendrell, L. Poignou, *Int. J. Environ. Anal.Chem.*, **45**, 55–63 (1991).
78. H.W. Nurnberg, *Anal. Chim. Acta*, **164**, 1 (1984).
79. J. Berg, M. Zapella, *J. Ment. Defic. Res.*, **8**, 44 (1964).
80. M.A. Sirover, L.A. Loeb, *Science*, **194**, 1434 (1976).
81. H.W. Nurnberg, *Sci. Total Environ.*, **12**, 35 (1979).
82. H.W. Nurnberg, *Electrochim. Acta*, **22**, 935 (1977).
83. E. Steeman Nielsen, W. Wium-Andersen, *Marine Biol.*, **6**, 93 (1970).
84. G.W. Bryan, *Proc. R. Soc. London*, **177**, 389 (1971).
85. E.A. Crecelius, *Environ. Health Perspect.*, **19**, 147 (1977).
86. Y.K. Chau, P.S. Wong, *NBS Spec. Publ.(US)*, **618**, 65 (1981).
87. J.S. Thayer, G.J. Olson, F.E. Brinkman, *Environ. Sci. Technol.*, **18**, 726 (1984).
88. I. Ahmad, Y.K. Chau, P.S. Wong, A.J. Carthy, L. Taylor, *Nature*, **287**, 716 (1980).
89. W.P. Ridley, L.J. Dizikes, J.M. Wood, *Science*, **197**, 329 (1977).
90. S. Sensen, A. Jernelov, *Nature*, **223**, 759 (1969).
91. R.M. Sterritt, J.N. Lester, *Sci. Total Environ.*, **14**, 5 (1980).
92. US EPA, US Environmental Protection Agency, Official Air Quality Planning Stand and (Technical Report) EPA (1991) 1, EPA 450/4-92-013.
93. K. Larjava, T. Laitinen, T. Kiviranta, V. Siemens, D. Klockow, *Int. J. Environ. Anal. Chem.*, **52**(1–4), 65–73 (1993).
94. E. de'Munari, A. Mainardi, V. Pessina, B. Zani, *Boll. Chim. Igien, Parte Sci.*, **44**(S5), 305–314 (1993).
95. K. Larjava, T. Laitinen, T. Vahlman, S. Artmann, V. Siemens, J.A.C. Broekaert, D. Klockow, *Int. J. Environ. Anal. Chem.*, **49**(1–2), 73–85 (1992).
96. R. Westerholm, H. Li, J. Almen, *Chemosphere*, **27**(8), 1381–1384 (1993).
97. C. Weickhardt, U. Boesl, E.W. Schlag, *Anal. Chem.*, **66**(7), 1062–1069 (1994).
98. B.W. Wenclawiak, T.E. Jensen, J.F.O. Richert, *Fresenius' J. Anal. Chem.*, **346**(6–9), 808–812 (1993).
99. M. Loepfe, H. Burtcher, H.C. Siegmann, *Water, Air, Soil Pollut.*, **68**(1–2), 77–84 (1993).
100. S. Manabe, N. Kurihara, O. Wada, S. Izumikawa, K. Asakuno, M. Moita, *Environ. Pollut.*, **80**(3), 281–286 (1993).
101. V. Krivan, G. Schneider, H. Baumann, U. Reus, *Fresenius' J. Anal. Chem.*, **348**(3), 218–225 (1994).
102. S. Chakrabarty, R.K. Mishara, *Asian Environ.*, **14**(4), 44–49 (1992).
103. L.M. Hildemann, M.A. Mazurek, G.R. Cass, B.R.T. Simoneit, *Aerosol Sci. Technol.*, **20**(4) (1994).
104. M. Maugery, M. Valentini, A. Novo, *J. Aerosol. Sci.*, **23**(S1), S987–S990 (1992).
105. B. Finley, K. Fehling, M. Falerios, D. Paustenbach, *Appl. Occup. Environ. Hyg.*, **8**(3), 191–200 (1993).
106. J. Injuk, L. Breitenbach, R. Van Grieken, U. Waetjen, *Microchim. Acta*, **114–115**, 313–321 (1994).
107. P. Artaxo, M.L.C. Rabello, F. Watt, G. Grime, *Nucl. Instrum. Methods Phys. Res. Sect. B*, **75**, 521–525 (1993).
108. A. Steinweg, D. Krankowsky, P. Laemmerzahl, B. Anweiler, *J. Atmos. Terr. Phys.*, **54**(6), 703–714 (1992).
109. C. Migon, P. Courau, B. Gentili, *Int. J. Environ. Anal. Chem.*, **50**(3), 183–191 (1993).
110. Yu, A. Lyul, G.M. Kolesov, V.O. Cherkezyan, *Zh. Anal. Khim.*, **48**(10), 1683–1689 (1993).
111. E.I. Vapirev, A.V. Hristova, *J. Environ. Radioact.*, **20**(1), 23–34 (1993).
112. H. Wichmann, V. Zelinski, B. Scholz-Boettcher, W. Lorenz, M. Bahadir, *Chemosphere*, **26**(6), 1159–1166 (1993).
113. C.I. Davidson, *Phys., Chem., Earth Sci. Res. Rep.*, **8**, 29–51 (1989).
114. P. Winkler, M. Schulz, W. Dannecker, *Fresenius' J. Anal. Chem.*, **340**, 575–579 (1991).
115. E. Bujisman, Report, RIVM-228703013, Order No. PB 91-143149, NTIS, 1990.
116. R.B. See, T.C. Willoughby, M.H. Brooks, J.D. Gordon, Report, USGS/WRI-90-4029, Order No. PB 91-214759, NTIS, 1990.
117. F.Z. Grandi, L. Szpyrkowitz, *Fresenius' J. Anal. Chem.*, **341**, 625–630 (1991).
118. T. Jikelis, M.M. Kane, A. Rendell, T. Davies, M. Tranter, K.E. Jarvis, *Anal. Proc.*, **29**, 288–291 (1992).
119. C.F. Boutron, J.P. Candelon, U. Gorlach, *Analusis*, **20**, M24–M27 (1992).
120. J.F. Solbe, B. Buyle, W. Guhl, T. Hutchinson, R. Laenge, U. Mark, R. Munk, N. Scholz, *Sci. Total Environ.*, (Suppl., Part 1), 47–61, 1993.

121. K. Goto, S. Taguchi, *Anal.Sci.*, **9**(1), 1–7 (1993).
122. J.O. Nriagu, G. Lawson, H.K.T. Wong, J.M. Azcue, *J. Great Lakes Res.*, **19**(1), 175–182 (1993).
123. M.W. Sweeny, *Water Environ. Res.*, **65**(4), 374–377 (1993).
124. R. Goo, *Water Environ. Res.*, **66**(4), 298–301 (1994).
125. E.R. Blatchley, III, *Water Environ. Res.*, **65**(4), 353–360 (1993).
126. M.L. Tercier, J. Buffle, *Electroanalysis*, **5**(3), 187–300 (1993).
127. L. Mart, *Talanta*, **29**, 1035 (1982).
128. K.S. Subramanian, C.L. Chakrabarti, J.E. Sueiras, I.S. Maines, *Anal. Chem.*, **50**, 444 (1978).
129. A. Zirino, in *Marine Electrochemistry*, eds. M. Whitfield, D. Jagner, John Wiley & Sons, Chichester, 422–503, 1981.
130. L. Mart, *Fresenius' J. Anal. Chem.*, **299**, 97 (1979).
131. I. Ahmad, Y.K. Chau, P.T.S. Wong, A.J. Carthy, L. Taylor, *Nature*, **287**, 716 (1980).
132. B. Schaule, C.C. Patterson, in *Lead in the Marine Environment*, eds. M. Branica, Z. Konrad, Pergamon Press, Oxford, 31, 1980.
133. E.A. Crecelius, *Environ. Health Perspect.*, **19**, 147 (1977).
134. T.M. Florence, *Talanta*, **29**, 345 (1982).
135. T.M. Florence, *Anal. Chim. Acta*, **141**, 73 (1982).
136. P. Benes, E. Steinnes, *Water Res.*, **9**, 741 (1975).
137. H. Scheuermann, H. Hartkamp, *Fresenius' J. Anal. Chem.*, **315**, 430 (1983).
138. L. Mart, *Fresenius' J. Anal. Chem.*, **296**, 350 (1979).
139. W.F. Smyth, L. Goold, D. Dadgar, M.R. Jan, M.R. Smyth, *Int. Lab.*, **Sept.**, 43–50 (1983).
140. H.N. Hecker, *Fresenius' J. Anal. Chem.*, **261**, 29 (1972).
141. L. Mart, H.W. Nurnberg, P. Valenta, *Fresenius' J. Anal. Chem.*, **300**, 350 (1980).
142. C.M.G. Van Den Berg, Z.Q. Huang, *J. Electroanal. Chem.*, **177**, 269 (1984).
143. C.M.G. Van Den Berg, K. Murphy, J.P. Riley, *Anal. Chim. Acta*, **188**, 177 (1986).
144. C.M.G. Van Den Berg, *Talanta*, **31**, 1069 (1984).
145. C.M.G. Van Den Berg, G.S. Jacinto, *Anal. Chim. Acta*, **211**, 129 (1988).
146. C.M.G. Van Den Berg, *Anal. Chem.*, **57**, 1532 (1985).
147. C.M.G. Van Den Berg, *Anal. Chim. Acta*, **164**, 195 (1984).
148. J. Wang, J.S. Mahmoud, *J. Electroanal. Chem.*, **208**, 383 (1986).
149. J.R. Donat, K.W. Bruland, *Anal. Chem.*, **60**, 240 (1988).
150. C.M.G. Van Den Berg, *J. Electroanal. Chem.*, **215**, 111 (1986).
151. J. Wang, *Int. Lab.*, **Sept.**, 12–25 (1983).
152. P. Valenta, L. Sipos, I. Cramer, P. Krumpen, H. Rutzel, *Fresenius' J. Anal. Chem.*, **312**, 101 (1982).
153. J. Wang, H.D. Dewald, *Anal. Chim. Acta*, **162**, 189 (1984).
154. R.W. Andrews, D.C. Johnson, *Anal. Chem.*, **48**, 1056 (1976).
155. A. Zirino, S.H. Lieberman, C. Clavell, *Environ. Sci. Technol.*, **12**, 73 (1978).
156. A.M. Bond, G.G. Wallace, *Anal. Chem.*, **53**, 1209 (1981).
157. A.M. Bond, G.G. Wallace, *Anal. Chem.*, **54**, 1706 (1982).
158. A.M. Bond, G.G. Wallace, *J. Liq. Chromatogr.*, **6**, 1977 (1983).
159. R. Eggli, *Anal. Chim. Acta*, **91**, 129 (1977).
160. D. Jagner, A. Graneli, *Anal. Chim. Acta*, **83**, 19 (1976).
161. A. Hussam, J.F. Coetzee, *Anal. Chem.*, **57**, 581 (1985).
162. S. Jaya, T.P. Rao, *Rev. Anal. Chem.*, **6**, 343 (1982).
163. L.K. Hoeflich, R.J. Gale, M.L. Good, *Anal. Chem.*, **55**, 1591 (1983).
164. R.D. Guy, C.L. Chakrabarti, *Chem. Can.*, **28**, 26 (1976).
165. H. Zhang, R. Wollast, J.C. Vire, G.J. Patriarche, *Analyst*, **114**, 1597–1602 (1989).
166. A.S. Mayer, A.J. Rabideau, R.J. Mitchell, P.T. Imhoff, M.I. Lowry, C.T. Miller, *Water Environ. Res.*, **65**(4), 486–534 (1993).
167. M.J. Barcelona, A.J. Helfrich, 'Current Practice in Groundwater Vadose Zone Investigation', *ASTM Spec. Tech. Publ.*, **1118**, 3–23 (1992).
168. S. Lesage, R.E. Jackson (eds.), 'Groundwater Contamination and Analysis at Hazardous Waste Sites', *Environmental Science Pollution Control, Series 4*, Marcel Dekker, New York, 1992.
169. Westinghouse Savannah River Co., Exploration Resources Inc. Report ESH-EMS-900134, Order No. DE92011908, NTIS, 1991.
170. S.H. Hall, S.P. Juracich, Report PNL-7836, Order No. DE92004452, NTIS, 1991.
171. R.D. Gibbons, 'Groundwater Contamination and Analysis at Hazardous Waste Sites', *Environmental Science Pollution Control, Series 4*, Marcel Dekker, New York, 199–243, 1992.
172. J. Wang, J. Lu, *Analyst*, **117**(12), 1913–1917 (1992).
173. L. Campanella, T. Ferri, R. Morabito, A.M. Padetti, *Chim. Ind. Milan*, **69**, 90 (1987).
174. D. Jagner, E. Sahlin, B. Axelsson, R. Ratana-Ohpas, *Anal. Chim. Acta*, **278**(2), 237–242 (1993).
175. J.R. Donat, K.A. Lao, K.W. Bruland, *Anal. Chim. Acta*, **284**(3), 547–571 (1993).
176. I. Turyan, D. Mandler, *Nature*, **362**(6422), 703–704 (1993).
177. J. Zima, C.M.G. Van Den Berg, *Anal. Chim. Acta*, **289**(3), 291–298 (1994).
178. P. Valenta, L. Mart, H. Rutzel, *J. Electroanal. Chem.*, **82**, 327 (1977).
179. T.M. Florence, G.E. Batley, *Talanta*, **23**, 179 (1976).
180. G. Gillain, G. Duyckaert, A. Di Steche, *Anal. Chim. Acta*, **106**, 23 (1979).
181. G. Gillain, C. Brihaye, *Oceanologica Acta*, **8**, 231 (1984).
182. G. Gillain, *Talanta*, **29**, 651 (1989).
183. T. Frank, R. Neeb, *Fresenius' J. Anal. Chem.*, **327**, 670 (1987).
184. G. Capodaglio, C.M.G. Van Den Berg, G. Scarponi, *J. Electroanal. Chem.*, **235**, 275 (1987).

185. W. Wagner, S. Sander, G. Henze, *Fresenius' J. Anal. Chem.*, **354**, 11 (1996).
186. Kh. Brainina, A.B. Tchernyshova, *Talanta*, **21**, 287 (1974).
187. H. Huiliang, D. Jagner, L. Renman, *Anal. Chim. Acta*, **202**, 123 (1987).
188. L. Mart, H.W. Nurnberg, P. Valenta, M. Stoeppler, *Thalassia Jugosl.*, **14**, 171 (1978).
189. L. Mart, H.W. Nurnberg, H. Rutzel, *Fresenius' J. Anal. Chem.*, **317**, 201 (1984).
190. L. Sipos, J. Golimowski, P. Valenta, H.W. Nurnberg, *Fresenius' J. Anal. Chem.*, **298**, 1 (1979).
191. L. Sipos, P. Valenta, H.W. Nurnberg, M. Branica, *J. Electroanal. Chem.*, **77**, 263 (1977).
192. I. Gustavsson, *J. Electroanal. Chem.*, **214**, 31 (1986).
193. C.M.G. Van Den Berg, *Analyst*, **114**, 1527–1530 (1989).
194. C.M.G. Van Den Berg, Z.Q. Huang, *J. Electroanal. Chem.*, **177**, 269 (1984).
195. G. Capodaglio, C.M.G. Van Den Berg, G. Scarponi, *J. Electroanal. Chem.*, **235**, (1987).
196. C.M.G. Van Den Berg, Z.Q. Huang, *Anal. Chim. Acta*, **164**, 209 (1984).
197. B. Pihlar, P. Valenta, H.W. Nurnberg, *Fresenius' J. Anal. Chem.*, **307**, 337 (1981).
198. J.R. Donat, K.W. Bruland, *Anal. Chem.*, **60**, 240 (1988).
199. C.M.G. Van Den Berg, *J. Electroanal. Chem.*, **215**, 111 (1986).
200. J. Golimowski, P. Valenta, H.W. Nurnberg, *Fresenius' J. Anal. Chem.*, **322**, 315 (1985).
201. H. Li, C.M.G. Van Den Berg, *Anal. Chim. Acta*, **221**, 269 (1989).
202. J. Wang, J. Zadeij, *Talanta*, **34**, 909 (1987).
203. C.M.G. Van Den Berg, G.S. Jacinto, *Anal. Chim. Acta*, **211**, 129 (1988).
204. C.M.G. Van Den Berg, S.H. Khan, J.P. Riley, *Anal. Chim. Acta*, **222**, 43–53 (1989).
205. G. Somer, A. Arslantas, *Analyst*, **119**, 1257–1259 (1994).
206. S. Komorsky-Lovric, *Anal. Chim. Acta*, **204**, 161 (1988).
207. T.R. Gilbert, D.N. Hume, *Anal. Chim. Acta*, **65**, 451 (1973).
208. I. Drabark, P.P. Madsen, J. Sorensen, *Int. J. Environ. Anal.Chem.*, **15**, 153 (1983).
209. B.R. Clark, D.W. de Paoli, D.R. Mc Taggart, B.D. Patton, *Anal. Chim. Acta*, **215**, 13 (1988).
210. H. Gunasingham, K.P. Ang, C.C. Ngo, P.C. Thiak, B. Fleet, *J. Electroanal. Chem.*, **186**, 51 (1985).
211. J.A. Wise, W.R. Heineman, P.T. Kissinger, *Anal. Chim. Acta*, **172**, 1–12 (1985).
212. L. Novotny, M. Heyrovski, *Trends Anal. Chem.*, **6**, 176 (1987).
213. K. Loomba, G.S. Pandey, *Int. J. Environ. Anal. Chem.*, **50**, 15–18 (1993).
214. G.C. Marten, P.B. Hammond, *Agron. J.*, **58**, 553 (1966).
215. B. Debski, A.D. Finley, M.F. Piciano, B. Lonnerdal, J. Milner, *J. Nutr.*, **119**, 215–220 (1989).
216. R. Inam, G. Somer, *Euroanalysis*, **10**, 409 (1998).
217. B. Welte, N. Bles, A. Montiel, *Environ. Technol. Lett.*, **4**, 79 (1983).
218. J. Einax, B. Machelett, S. Geiss, K. Danzer, *Fresenius' J. Anal. Chem.*, **342**(4–5), 267–272 (1992).
219. J.P. Vernet, P.Y. Favarger, D. Span, C. Martin, *Trace Met. Environ.*, **1**, 397–404 (1991).
220. M. Sager, *Tech. Instrum. Anal. Chem. (Hazard. Met. Environ.)*, **12**, 133–175 (1992).
221. A. Desaules, P. Lischer, R. Dahinden, H.J. Bachmann, *Comm. Soil Sci. Plant Anal.*, **23**(3–4), 363–377 (1992).
222. A.H. Van Den Akker, H. Van Den Heuvel, *At. Spectrosc.*, **13**, 72–74 (1992).
223. C.K. Lo, Y.S. Fung, *Int. J. Environ. Anal. Chem.*, **46**, 277–291 (1992).
224. J.R. Senten, R.H. Charlier, *Int. J. Environ. Stud.*, **37**(1–2), 79–96 (1991).
225. L.M. Azzaria, A. Aftabi, *Water, Air, Soil Pollut.*, **56**, 203–217 (1991).
226. M. Hempel, H. Hintelmann, R.D. Wilken, *Analyst*, **117**, 669–672 (1992).
227. G.U. Fortunati, C. Banfi, M. Pasturenzi, *Fresenius' J. Anal. Chem.*, **348**(1–2), 86–100 (1994).
228. L.E. Brown, T.G. Reinsch, *ASTM Spec. Tech. Publ.*, **1162**, 32–40 (1993).
229. R. Wehrens, P. Van Hoof, L. Buydens, G. Kateman, M. Vossen, W.H. Mulder, T. Bakker, *Anal. Chim. Acta*, **271**, 11–24 (1992).
230. F.P.J. Lame, P.R. Defize, *Environ. Sci. Technol.*, **27**(10), 2035–2044 (1993).
231. A.B. Mc Bratney, G.M. Laslett, *Soil. Environ.*, **1**, 435–439 (1993).
232. J.R. Skalski, J.Q. Word, *Environ. Toxicol. Chem.*, **13**, 15–19 (1994).
233. E. Ristenpart, R. Gitzel, M. Uhl, *Water Sci. Technol.*, **25**, 63–69 (1992).
234. V.J.G. Houba, *Fresenius' J. Anal. Chem.*, **345**(2–4), 156 (1993).
235. M.C. Thoms, *J. Geochem. Explor.*, **51**, 131–141 (1994).
236. R. Rubio, A. Ure, *Int. J. Environ. Anal. Chem.*, **51**(1–4), 205–217 (1993).
237. V.J.G. Houba, W.J. Chardon, K. Roelse, *Commun. Soil Sci. Plant Anal.*, **24**(13–4), 1591–1602 (1993).
238. G.S.R. Krishnamurti, P.M. Huang, K.C.J. Van Rees, L.M. Kozak, H.P.W. Rostad, *Commun. Soil Sci. Plant Anal.*, **25**, 615–625 (1994).
239. H.M. Kingston, P.J. Walter, *Spectroscopy*, **7**, 20, 22, 24–7 (1992).
240. Y. Feng, R.S. Barratt, *Total Environ.*, **143**, 157–161 (1994).
241. C.M. Reynolds, in *Engineering Aspects Metal-Waste Management*, eds. I.K. Iskander, H.M. Selim, Lewis, Boca Raton, FL, 49–61, 1992.
242. G. Somer, H. Aydin, *Analyst*, **110**, 631–633 (1985).
243. A.R. Crosland, S.P. Mc Grath, P.W. Lane, *Int. J. Environ. Anal. Chem.*, **51**(1–4), 153–160 (1993).

244. J. Sanches, R. Garcia, E. Millan, *Analusis*, **22**, 222–225 (1994).
245. D.E. Kimbrough, J. Wakakuwa, *Analyst*, **119**, 383–388 (1994).
246. C.L. Rojas, S.B. de Maroto, P. Valenta, *Fresenius' J. Anal. Chem.*, **348**, 775 (1994).
247. D. Rurikova, A. Beno, *Chem. Pap.*, **46**, 23–27 (1992).
248. P. Del Castillho, I. Rix, *Int. J. Environ. Anal. Chem.*, **51**(1–4), 59–64 (1993).
249. A.M. Ure, P. Quevauviller, H. Muntau, B. Griepink, *Int. J. Environ. Anal. Chem.*, **51**(1–4), 135–151 (1993).
250. S.J. Reddy, P. Valenta, H.W. Nurnberg, *Fresenius' J. Anal. Chem.*, **313**, 390 (1982).
251. S. Forbest, G.P. Bound, T.S. West, *Talanta*, **26**, 473 (1979).
252. J. Wang, J.M. Lu, K. Olsen, *Analyst*, **117**, 1913 (1992).

Ion Chromatography in Environmental Analysis

Peter E. Jackson
Dionex® Corporation, Sunnyvale, USA

1 Introduction **1**
1.1 Historical Perspective 1
1.2 Definition and Scope 2
2 Alternative Analytical Techniques **2**
3 Sample Handling and Preparation **3**
3.1 Sample Storage and Preservation 3
3.2 Sample Dissolution 4
3.3 Sample Cleanup 5
4 Regulatory Methods of Analysis **6**
5 Water and Wastewater Analysis **7**
5.1 Drinking Water and Wastewater 7
5.2 Natural Waters 13
5.3 Brines 15
6 Soil, Sludge, and Solid Waste Analysis **17**
6.1 Soil and Soil Extracts 17
6.2 Sludge, Leachates, and Solid Waste 18
7 Quality Assurance **19**
Acknowledgment **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

Ion chromatography (IC) is a subset of liquid chromatography applied to the determination of ionic solutes, such as inorganic anions, cations, transition metals, and low-molecular-weight organic acids and bases. Although these solutes can be analyzed using a number of separation and detection modes, ion-exchange is the primary separation mode and suppressed conductivity is the primary method of detection in IC. Method detection limits (MDLs) for inorganic anions and cations are typically in the low parts per billion range and recoveries obtained for ions in spiked water samples are generally in the order of 80–110%. The linear calibration range extends from low parts per billion to mid parts per million concentrations for most applications. IC is well established as regulatory method for the analysis of inorganic anions in environmental samples as there are few alternative methods which can determine multiple anions in a single analysis. However, there are relatively few regulatory methods for cation analysis which use IC. Methods for cation analysis tend to be based upon spectroscopic techniques, although IC offers the advantage of providing information on metal speciation. The main interferences in IC are generally other ions which elute within the timeframe of the ion(s) of interest, and the determination of trace ions in the presence of high levels of other ionic species remains the most difficult of analyses with this technique.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Historical Perspective

The origins of modern IC were laid down by Hamish Small et al. at Dow Chemical in 1975, when they first described a novel ion-exchange chromatographic method for the separation and conductimetric detection of ionic species.[1] They employed a low-capacity, ion-exchange stationary phase for the separation of analyte ions, in conjunction with a second ion-exchange column and conductivity detector, which allowed for continuous monitoring of the eluted ions. The second column was called a 'stripper' column (later termed a 'suppressor') and served to reduce the background conductance of the eluent and enhance the detectability of the eluted ions. In the case of anion analysis, this was achieved by exchanging hydronium ions from the cation-exchange suppressor for cocations (e.g. sodium) in the eluent, prior to the measurement of conductance. This results in conversion of the eluent anion to its less-conductive weak-acid form, while enhancing the conductance of the analyte ion pair as a result of replacing the less-conductive cocation with the more conductive hydronium ion. The reverse situation can be described for cation analysis, where the suppressor is an anion-exchange column and the eluent is converted to its less-conductive weak-base form. The term 'IC' was subsequently introduced when this technology was licensed to the Dionex® Corporation for commercial development.

The introduction of IC provided a foundation for renewed interest in the determination of ionic solutes and prompted much investigation into the use of alternative separation and detection approaches for the liquid chromatographic analysis of inorganic compounds. In 1979, Fritz et al. showed that a suppressor was not essential to sensitive conductivity detection, provided that appropriate low-capacity stationary phases and low-conductance eluents were used.[2,3] In addition, the separation of inorganic ions using traditional C_{18} reversed-phase columns was also being explored at about the same time.[4] Since that time, a considerable variety of separation and detection methods have been

employed for the determination of ionic species, as discussed in the article **Ion Chromatography**. This, in turn, has significantly expanded the range of solutes and applications to which IC is now applied.

1.2 Definition and Scope

This diversity has led to the point where IC is typically defined by the range of solutes to which it is applied, rather than any specific combination of separation and detection modes.[5] IC can therefore be considered to encompass liquid chromatographic techniques which can be used for the determination of the following ionic solutes: inorganic anions; inorganic cations including alkali metal, alkaline earth, transition metal, and rare earth ions; low-molecular-weight (water soluble) carboxylic acids plus organic phosphonic and sulfonic acids, including detergents; carbohydrates; low-molecular-weight organic bases; and ionic metal complexes.[5–8] Ion-exchange remains the primary separation mode used in IC today, although other approaches used for the separation of inorganic species include ion interaction, ion exclusion, and chelation chromatography, in addition to reversed-phase separations of metal complexes. Advances in suppressor technology have improved the sensitivity and ease of use of suppressed conductivity detection, although nonsuppressed conductivity and indirect (or vacancy) detection methods are still employed as alternatives for universal detection. Direct detection methods have proven to be highly selective for ultraviolet (UV)-absorbing or electroactive species, whereas postcolumn derivatization followed by UV/VIS (ultraviolet/visible) absorption or fluorescence is an important detection approach for transition metals, lanthanides, and actinides. Additionally, the use of more advanced detection techniques for IC, such as mass spectrometry (MS) and inductively coupled plasma mass spectrometry (ICPMS), continue to be explored.[5]

The growth of IC was very rapid because it provided a reliable and accurate method for the simultaneous determination of many common inorganic ions. In the early stages of its development, IC was viewed as a tool for the determination of simple inorganic species, particularly in environmental samples. The vast majority of the early applications of IC were for the analysis of inorganic anions and cations in environmental samples, such as air-filter extracts, soil extracts, drinking water, and natural-water samples.[9,10] As the range of solutes that could be determined by IC continued to expand, so did the application areas in which the technique was applied. In addition to environmental applications, IC is now routinely used for the analysis of ionic compounds in diverse areas. These include the chemical and petrochemical industries, semiconductor and high-purity water applications, food and beverage applications, the clinical and pharmaceutical industries, and mining and metallurgical applications.

IC can now be considered a well-established, mature technique for the analysis of ionic species and many organizations, such as ASTM (American Society for Testing and Materials), AOAC (Association of Official Analytical Chemists), and USEPA (United States Environmental Protection Agency), have standard or regulatory methods of analysis based upon IC. Despite the diverse range of solutes and sample types currently analyzed by IC, environmental analysis continues to be the largest application area in terms of new instrument sales and the total number of samples analyzed. In terms of the solutes analyzed in environmental applications of IC, inorganic anions are by far the most important. The primary reason for this is the lack of alternative methods for anion analysis, which is not the case for cations, where many other instrumental techniques are available. Consequently, the simultaneous analysis of the common inorganic anions in drinking water and wastewater remains the most important routine application of IC.

2 ALTERNATIVE ANALYTICAL TECHNIQUES

IC is a well-established regulatory technique for the analysis of inorganic anions in environmental samples. Acceptance of IC for the analysis of anionic solutes was very rapid, primarily due to the lack of alternative methods that could determine multiple anions in a single analysis. A variety of methods have been employed for the analysis of inorganic anions: traditional spectroscopic techniques such as colorimetry; wet chemical methods such as gravimetric analysis, turbidimetry, and titrimetry; and electrochemical techniques such the use of an ion-selective electrode (ISE) and amperometric titrations.[11] Many of these methods suffer from interferences and limited sensitivity, they can be labor intensive and are often difficult to automate. The use of flow injection analysis (FIA) enables the automation of certain colorimetric and ISE methods of analysis, although still only for one analyte at a time. Multiple analytes can be determined by adding additional channels to an FIA system; however, this adds complexity and cost to the instrument.

During the early development of IC, many comparisons between wet chemistry methods and IC were performed in order to validate the latter technique. For instance, in a 1984 comparison, IC was found to be equivalent to conventional wet chemistry methods for the determination of common anions (such as chloride, nitrate, and sulfate) in

Table 1 Conventional methods for the analysis of inorganic anions in drinking water[12]

Analyte	Conventional analytical method	Method
Fluoride	ISE	USEPA 340.2
Chloride	Potentiometric titration	[b]APHA 407C
Nitrite	Automated (FIA) sulfanamide[a]	USEPA 354
Nitrate	Automated (FIA) cadmium reduction[a]	USEPA 365.1
Phosphate	Automated (FIA) ascorbic acid[a]	USEPA 353.2
Sulfate	Turbidimetric	USEPA 375.1

[a] FIA methods use spectrophotometry for quantification after appropriate color formation.
[b] American Public Health Association.

drinking-water samples.[12] Table 1 details the approved conventional methods used (at the time) for the analysis of the inorganic anions commonly found in drinking water. Considering that these six individual test procedures could be replaced by one 30 min chromatographic separation, it is not surprising that IC quickly became accepted by regulatory bodies worldwide for the analysis of anions in drinking water.

However, the situation regarding the analysis of cations in environmental samples is quite different to the case for anions. Many rapid and sensitive spectroscopic methods (such as AAS (atomic absorption spectroscopy), ICPAES (inductively coupled plasma atomic emission spectroscopy), and ICPMS) and electrochemical methods (such as polarography and anodic stripping voltammetry) are available for cation analysis. Many of these are multielement techniques and therefore duplicate one of the major attractions of chromatographic methods. Regulatory methods for cation (metal) analysis in environmental samples tend to be based primarily upon AAS and ICP (inductively coupled plasma) instrumentation.[11] However, IC does offer an advantage over spectroscopic techniques for cation analysis in the area of metal speciation. For instance, as IC is a separation-based technique, it can distinguish different metal oxidation states, such Cr(III) and Cr(VI) and also Fe(II) and Fe(III).[13,14] In addition, IC has also been used to determine stable metal complexes, such as metallocyanides and organic arsenic species.[5] Nevertheless, there are only a limited number of environmental regulatory methods for cations based upon IC methodology.

3 SAMPLE HANDLING AND PREPARATION

The primary concerns when collecting environmental samples for analysis using any measurement technique is that the sample collected is representative of the total sample matrix, and that no contamination occurs during the sampling process.[5] Also, appropriate storage and preservation of the sample is required, in order that the final sample analysis is representative of the analyte concentrations present when the sample was originally taken from the field. Appropriate procedures for the collection of representative samples are discussed in detail in the article **Sampling Considerations for Biomonitoring**; however, sample handling and preparation techniques specific to IC analysis are discussed below.

3.1 Sample Storage and Preservation

Water samples collected for analysis by IC ideally should be collected in plastic containers, such as polystyrene or polypropylene bottles, as glass bottles can contribute ionic contamination when performing trace analysis.[15] The bottles should be thoroughly rinsed with reagent-grade

Table 2 Sample preservation and holding times for anions and cations commonly determined by IC[11,16–19]

Analyte	Preservation	Holding time (days)
Acetate	Cool to 4 °C	2
Bromate[a]	Add 50 mg L^{-1} EDA[b]	28
Bromide	None required	28
Chlorate[a]	Add 50 mg L^{-1} EDA	28
Chloride	None required	28
Chlorite[a]	Add 50 mg L^{-1} EDA, cool to 4 °C	14
Chromate	Adjust pH to 9–9.5 with eluent[c]	1
Cyanide	Adjust pH to >12 with NaOH, cool to 4 °C	14
Fluoride	None required	28
Formate	Cool to 4 °C	2
Nitrate[d]	Cool to 4 °C	2
Nitrite[d]	Cool to 4 °C	2
o-Phosphate	Cool to 4 °C	2
Sulfate	None required	28
Ammonium	Filtration, cool to 4 °C	7
Calcium	Filtration	42
Magnesium	Filtration	42
Potassium	Filtration	42
Sodium	Filtration	42
Metals, e.g. Co, Ni, Zn	Acidify to pH < 2 with nitric acid, i.e. 1.5 mL conc. HNO_3 per liter sample	6 months

[a] Samples collected for oxyhalide analysis should be immediately sparged with an inert gas (e.g. nitrogen, argon, or helium) for 5 min to remove active gases such as chlorine dioxide or ozone. Samples for chlorite should be stored in amber containers.
[b] EDA = ethylenediamine.
[c] Eluent = 250 mM ammonium sulfate and 100 mM ammonium hydroxide.
[d] Holding times can be increased by adjusting to pH 12 with sodium hydroxide.

water before use. Sample preservation requirements and holding times for anions and cations typically determined by IC are listed in Table 2.

3.2 Sample Dissolution

The majority of the water samples collected for IC analysis require little or no sample pretreatment. Drinking-water samples, for instance, typically require no pretreatment other than filtration through a 0.45 μm filter to remove particulates. Higher ionic strength water samples, e.g. wastewater, often only require dilution (and filtration) to bring the analytes of interest into the working range of the method. In fact, this so-called dilute-and-shoot approach to sample preparation is one of the advantages of IC when it comes to the practical application of this technique.[20] However, solid samples, such as soils and sludge, are not directly amenable to IC analysis and require additional sample pretreatment.

3.2.1 Aqueous Extraction

The high solubility of ionic species in water means that such solutes can often be removed from solid samples prior to IC analysis simply by aqueous extraction of the finely divided sample. Generally, a weighed amount of the sample is mixed with a known volume of water, or other extracting solution, and homogenized using a blender or ultrasonic probe; alternatively, it is mixed with a magnetic stirrer or rotation flask.[17,21] The choice of extracting solution is dependent on both the sample matrix and the nature of the solute ions.[5] Water is preferred, in order to avoid introducing extraneous peaks into the final chromatogram. However, the following have all been used to extract ionic species from solid samples prior to IC analysis:[5] water combined with a miscible solvent such as methanol; solutions of dilute acid or base; dilute salt solutions, such as potassium chloride or phosphate buffers; or even the eluent solution used for the IC separation. Recoveries from the aqueous extraction of solid samples for low parts per million spikes of leachable solutes, such as fluoride, chloride, bromide, and nitrate, are typically of the order of 82–101%.[22]

3.2.2 Acid Digestion

Many solid samples, such as shale or rocks, are not amenable to simple aqueous extraction and it is necessary to digest the sample to obtain quantitative recoveries of ionic species. This is traditionally performed using concentrated acids (e.g. nitric acid) or their mixtures (e.g. nitric/hydrochloric acids) and is widely used in the preparation of environmental samples prior to analysis using spectroscopic techniques, such as AAS or ICPAES.[11,23] However, acid digestion is often inappropriate for IC because the excess of the acid coanion can lead to the appearance of a large, interfering peak in the final chromatogram and can also cause column overloading.[5] Hence, this approach is typically not employed for anion analysis, unless further sample pretreatment is being used. One important example of this approach is the determination of total cyanide by IC. Cyanide is strongly complexed to certain metals, e.g. iron, and strong acid digestion in the presence of a magnesium chloride catalyst is required to liberate free cyanide. The resulting hydrocyanic acid can be removed from the digestion matrix by reflux distillation and absorbed into a sodium hydroxide solution which can be analyzed for cyanide using IC with amperometric detection.[24] The same approach can also be used for sulfide and fluoride in solid samples.

In general, acid digestion is better suited to preparing samples which are to be analyzed for cations using IC. For instance, total nitrogen (as the ammonium ion) has been determined in environmental samples using IC by direct analysis of the Kjeldahl digest matrix.[25] In addition, transition metals and rare earth elements are frequently analyzed using IC with postcolumn reaction detection after acid digestion of the samples.[5]

3.2.3 Alkali Fusion

Fusion techniques can be an attractive alternative to acid digestion for the preparation of samples of geological materials. A sample is mixed with an alkaline flux and heated at high temperatures (800–1100 °C) until the flux becomes molten. The melt is then cooled and dissolved in a suitable digestion solution prior to analysis by IC. Typical flux materials include sodium peroxide, sodium carbonate, lithium tetraborate, and sodium hydroxide.[26] Care must again be taken to ensure that the final digest solution is compatible with the IC separation. This approach has been used for the determination of fluoride and chloride in geological materials after fusion with sodium carbonate and subsequent injection into an IC device using a carbonate/bicarbonate eluent.[5] Occasionally, this approach is the only means available for obtaining complex samples in solution. For instance, alkali fusion followed by acid digestion proved to be necessary for the complete dissolution of mineral sands when analyzing thorium and uranium by IC with postcolumn reaction detection.[26]

3.2.4 Combustion Methods

Sample combustion is another approach which is commonly used to prepare solid matrices for analysis by IC. This involves total combustion of the sample in oxygen, which converts nonmetallic elements to volatile gaseous

compounds. These gases can be collected into a suitable absorbing solution which can then be analyzed using IC. Combustion methods have been used for the determination of halides and total sulfur, nitrogen, and phosphorus in samples such as plant materials, silicate rocks, coal, and oil shales.[17] The Schoeniger flask is the simplest apparatus for sample combustion. However, the oxygen pressure is limited to atmospheric pressure, which limits the sample size and ultimately the sensitivity of the analysis. Larger samples can be accommodated in a Parr bomb where oxygen pressures as high as 40 atm are used to facilitate sample combustion.[5] The absorbing solution used for halides, which are converted to their corresponding acid gas (e.g. HCl) during combustion is typically dilute sodium hydroxide. A reducing agent, such as hydrazine sulfate, may be added to prevent the formation of oxyanions for bromide and iodide.[27] Alternatively, sulfur and phosphorus are typically quantitated as sulfate and phosphate, respectively, hence the absorbing solution typically contains an oxidizing agent, such as hydrogen peroxide.[28]

3.3 Sample Cleanup

Once a solution has been obtained it is typically necessary to perform some degree of sample pretreatment or cleanup prior to injection into the IC device. This pretreatment may be as simple as filtration, or may be a complicated time-consuming matrix-elimination step. The typical intent of sample cleanup is to achieve one or more of the following goals: (1) removal of particulates which could cause blockages or damage to the instrument; (2) reduce the overall sample loading on the column; (3) concentration or dilution of the target analytes; and (4) removal of matrix interferences.[5]

3.3.1 Filtration

As is the case with all liquid chromatographic methods, samples analyzed by IC should be free of particulates to avoid blockages or damage to connecting tubing, column end frits, and other hardware components. Samples are typically filtered through a 0.45 μm (or less) membrane-based filter. Disposable syringe filters are commercially available and their use greatly simplifies sample filtration. Also, certain types of autoinjectors will automatically filter the sample before injection into the IC instrument. However, the possibility of sample contamination from these devices can be a concern, particularly when performing trace analysis. Rinsing filters with 20 mL of deionized water prior to filtration of the sample has been shown to remove most inorganic contaminants.[29]

3.3.2 Matrix Elimination

Complex aqueous samples, such as wastewaters and solid leachates, often require further chemical modification (cleanup) of the sample in order to eliminate matrix interferences. Solid-phase extraction (SPE) cartridges represent the most convenient means of removing interferences prior to ion chromatographic analysis.[30] These commercially available disposable cartridges enable rapid sample pretreatment and require only small volumes of sample. SPE cartridges are available with many different chromatographic packing materials; including silica, alumina, C_{18}, anion-exchange resins (OH^- form), cation-exchange resins (H^+, Ag^+, Ba^+ forms), neutral polymer, amino, and activated carbon.[5] These cartridges can be employed in a number of different modes of operation with IC analyses, as discussed below.

- Hydrophobic SPE cartridges, e.g. C_{18} and neutral polymers, can be used to remove neutral organic compounds, while allowing inorganic ions to pass through unretained. Hydrophobic organic compounds do not typically interfere during an IC separation; however, they can be strongly retained on the stationary phase material which can lead to decreased column lifetimes. This approach is typically required when using IC for the analysis of ions in food and biological matrices, or wastewaters and soil leachates containing high levels of organics, e.g. humic acids.[17,21]
- Cation- and anion-exchange SPE cartridges in the H^+ and OH^- forms, respectively, can be used to adjust sample pH and reduce total ionic strength without adding a potentially interfering coanion (for acids) or cocation (for bases) to the sample.[30]
- Cation-exchange SPE cartridges in the H^+ form can be used to remove carbonate and cationic species, such as $iron^{III}$ and aluminum, which may precipitate under alkaline eluent conditions.[17] Cation-exchange cartridges in the Ag^+ and Ba^+ forms can be used to selectively remove halides and sulfate from samples by precipitating insoluble silver halides or barium sulfate, respectively.[30] The Ag^+ form cartridges are widely used to selectively remove chloride from environmental waters to allow trace analysis of anions, such as bromate in ozonated waters, which would otherwise be masked by the excess chloride.[31]

Recently, SPE disks have become available as an alternative to the cartridge configuration. These disks are available with many of the same packings as the cartridges, although their geometry allows the use of higher sample loading flow rates.[32] The use of dialysis across membranes offers another means of reducing sample interferences, and this approach has been used

to reduce sample acidity,[33] basicity,[34] and also chloride in brine samples prior to IC analysis.[35]

3.3.3 On-line Matrix Elimination

Many of the sample preparation techniques described above can also be performed using on-line instrumentation. This approach offers the benefits of greater precision and the process can often be automated, although the instrumentation is usually more complex. The most common application of on-line matrix elimination in IC is sample preconcentration. A measured volume of sample is first passed through an ion-exchange concentrator column and the solute ions are retained on this column, while the bulk sample solution (i.e. water) passes through to waste. The concentrator column is then switched to be in line with the eluent, which carries the solute ions through to the analytical column to be separated and detected in the usual manner.[5] This approach is frequently used for the determination of trace ions in high-purity waters, e.g. steam and boiler feed water for power station generators.[36]

Sample preconcentration can also be successfully applied to high-ionic-strength samples, particularly if the solutes of interest are strongly retained on the concentrator column. The strongly retained solutes can be trapped on the concentrator column while weakly retained solutes and water pass through to waste. This approach, which effectively combines sample preconcentration with matrix elimination, has been used for the determination of anions in aqueous vegetation extracts and low parts per billion levels of precious metal cyanide complexes in gold tailings solutions.[37,38] A variation on using ion-exchange columns to selectively concentrate ionic solutes prior to IC analysis is the use of chelating resins. These materials can be used to selectively complex transition metal and rare earth elements in the presence of high levels of alkali and alkaline earth metals. This approach has been employed for the on-line cleanup and concentration of low parts per billion levels of transition metals and lanthanides in complex geological matrices and seawater prior to analysis by IC or ICPMS.[39]

Automated matrix elimination can also be performed using heart-cut techniques. This involves loading the sample onto a short column, then carefully switching only the fraction of eluent containing the solute(s) of interest towards the separator column. This approach, which typically performs best when the sample matrix has a consistent composition, has been used for the determination of trace anions in samples containing high levels of chloride.[20]

4 REGULATORY METHODS OF ANALYSIS

IC has been approved by many standard or regulatory organizations in numerous countries for the analysis of

Table 3 Regulatory IC methods approved in the USA for environmental water and waste analysis

Method	Analytes	Matrices[a]
EPA Method 300.0 (A)	F, Cl, NO_2, Br, NO_3, PO_4, SO_4	rw, dw, sw, ww, gw, se
EPA Method 300.0 (B)	BrO_3, ClO_3, ClO_2	Raw water, dw
EPA Method 300.1 (A)	F, Cl, NO_2, Br, NO_3, PO_4, SO_4	rw, dw, sw, gw
EPA Method 300.1 (B)	BrO_3, Br, ClO_3, ClO_2	rw, dw, sw, gw
EPA SW-846 9056	F, Cl, Br, NO_3, PO_4, SO_4	Combustion extracts, all waters
ASTM D 4327-97	F, Cl, NO_2, Br, NO_3, PO_4, SO_4	dw, ww
Standard Methods[11] 4110	Cl, NO_2, Br, NO_3, PO_4, SO_4	rw, dw, ww
EPA Method 300.6	Cl, NO_3, PO_4, SO_4	Wet deposition, rain, snow, dew, sleet, hail
ASTM D 5085-90	Cl, NO_3, SO_4	Wet deposition, rain, snow, sleet, hail
EPA Method B-1011	NO_2, NO_3	rw, dw
EPA SW-846 9058[b]	ClO_4	rw, dw, gw
ASTM D 2036-97	CN	dw, sw, ww
ASTM D19.05.03.22[b]	Total $Fe(CN)_6^{3-}$ and $Fe(CN)_6^{4-}$	dw, gw, ww
EPA Method 218.6	Hexavalent chromium (CrO_4^{2-})	dw, gw, ww
EPA SW-846 7199	Hexavalent chromium (CrO_4^{2-})	dw, gw, ww
ASTM D 5257-93	Hexavalent chromium (CrO_4^{2-})	dw, sw, ww
ASTM D19.05.03.23[b]	Na, NH_4, K, Mg, Ca	rw, sw, gw, ww
EPA Method 300.7	Na, NH_4, K, Mg, Ca	Wet deposition, rain, snow, dew, sleet, hail
EPA Method 200.10[c]	Cd, Co, Cu, Pb, Ni, U, V	Brines, seawater, mw, ew
EPA Method 200.13[c]	Cd, Co, Cu, Pb, Ni	Brines, seawater, mw, ew

[a] Matrices: rw = reagent water; dw = drinking water; sw = surface water; ww = wastewater (mixed domestic and industrial); gw = ground water; se = solid extracts; mw = marine water; ew = estuarine water.
[b] Proposed method, in draft form only.
[c] IC used for cleanup and concentration prior to spectroscopic analysis.

both anions and cations in environmental samples. The majority of the approved methods are for the analysis of anionic solutes; however, some approved methods exist for cations. Although a complete listing of approved IC methods worldwide is beyond the scope of this article, a list of the most important regulatory IC methods used for environmental water and waste analysis in the USA is given in Table 3.[40]

Many different regulatory agencies promulgate what are essentially similar methods, as Table 3 illustrates. For instance, ASTM D 4327-97 uses the same methodology as USEPA Method 300.0(A); however, each agency has a unique method format and writing style. Also, differences exist between the methods in the area of quality control (QC). Thus QC is mandated in most USEPA methods whereas it is currently optional in ASTM methods. Different regulatory agencies exist even within the USEPA; hence Method 300.0 is applicable to the analysis of inorganic anions in drinking water and wastewater under direction of the Office of Groundwater and Drinking Water, whereas Method 9056 is applicable to the analysis of inorganic anions in all water types and combustion bomb extracts under direction of the Office of Solid Waste and Emergency Response.

In addition to the methods shown in Table 3 which use IC for environmental water and waste analysis, a considerable number of IC methods are employed for air analysis. Regulatory bodies, such as the National Institute of Occupational Safety and Health (NIOSH) and the Occupational Safety and Health Administration (OSHA), specify IC for the analysis of compounds in air and workplace atmospheres that form ionic species in solution (e.g. formic acid, sulfur dioxide, nitrous oxides, ammonia, and hexavalent chromium).

Although Table 3 details only methods approved in the USA, many industrial countries around the world have similar health and environmental standards; consequently, a considerable number of regulatory IC methods exist worldwide. For instance, German Methods DIN 38 405 (D 20) and DIN ISO (International Organization for Standardization) 10 304-1 are similar to USEPA Method 300.0, whereas DIN 38 405 (D 22) is applicable to the determination of chromate, iodide, sulfite, thiocyanate, and thiosulfate in water matrices. French Method AFNOR T90-042 is again similar to EPA Method 300.0, as is the Italian method UNICHIM 926 (1991). Several IC methods are approved in Japan for the analysis of anions in matrices, such as industrial waters (K0101), industrial wastewater (K0102), mine water, and wastewater (M0202).[41] Japan also has a method detailing general rules for IC analysis (K0127), whereas Standards Australia has a recommended practice for chemical analysis by IC (AS 3741-1990).[42]

5 WATER AND WASTEWATER ANALYSIS

5.1 Drinking Water and Wastewater

Water quality in the USA is legislated through the Safe Drinking Water Act (SDWA) and the Clean Water Act (CWA). The SDWA ensures the integrity and safety of US drinking water, whereas the goal of the CWA is to reduce the discharge of pollutants into US waters. The majority of the regulatory methods of analysis that use IC are validated for both drinking water and wastewater, hence these matrices will be considered together.

5.1.1 Common Inorganic Anions

The US National Primary Drinking Water Standards specifies a maximum contaminant level (MCL) for a number of common inorganic anions, including fluoride, nitrite, and nitrate. The MCLs are specified to minimize potential health effects arising from the ingestion of these anions in drinking water. For instance, high levels of fluoride causes skeletal and dental fluorosis, whereas nitrite and nitrate can cause methemoglobulinemia, which can be fatal to infants.[21] Consequently, the analysis of these anions in drinking waters is mandated, as are the analytical methods used for their quantification. Other common anions, such as chloride and sulfate, are considered secondary contaminants. The Secondary Drinking Water Standards are guidelines regarding taste, odor, color and certain aesthetic effects that are not federally enforced. However, they are recommended to all states as reasonable goals and many of the states adopt their own enforceable regulations governing these contaminants.[43]

IC has been approved for compliance monitoring of these common inorganic anions in drinking water in the USA since the mid-1980s, as described in EPA Method 300.0. This same method received interim approval for the analysis of inorganic anions in wastewater under the National Pollution Discharge Elimination System (NPDES) permits program in 1992. Method 300.0 specifies the use of a Dionex® AS4A anion-exchange column with an eluent of 1.7 mM sodium bicarbonate/1.8 mM sodium carbonate for the separation of common anions. An optional column may be substituted provided comparable resolution of peaks is obtained and that the QC requirements of the method can be met.[22] Conductivity detection is used for quantification after suppression of the eluent conductance with an anion micro-membrane suppressor (AMMS) or similar device.

Figure 1 shows a chromatogram of a standard containing low parts per million levels of common anions obtained using the conditions described in Method 300.0. All the anions are well-resolved within a total run time of less than 8 min. The application range and MDLs that can be achieved for each of the anions using

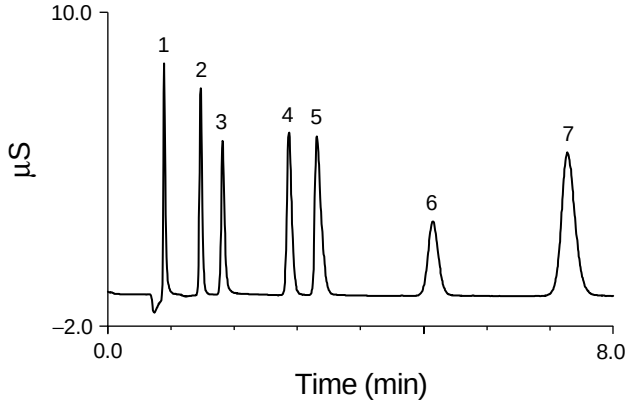

Figure 1 Separation of low parts per million anion standard using EPA Method 300.0. Conditions: column, Dionex® IonPac® AS4A-SC; eluent, 1.8 mM sodium carbonate/1.7 mM sodium bicarbonate; flow rate, 2.0 mL min^{-1}; detection, suppressed conductivity with an anion self-regenerating suppressor (ASRS) operated at 50 mA in recycle mode; injection volume, 25 µL; solutes, 1 = fluoride (2 mg L^{-1}), 2 = chloride (3 mg L^{-1}), 3 = nitrite (5 mg L^{-1}), 4 = bromide (10 mg L^{-1}), 5 = nitrate (10 mg L^{-1}), 6 = phosphate (15 mg L^{-1}), 7 = sulfate (15 mg L^{-1}).

Table 4 USEPA Method 300.0 application ranges and detection limits (DLs)[22]

Analyte	Application range (mg L^{-1})	MDL (mg L^{-1})
Fluoride	0.26–8.49	0.01
Chloride	0.78–26.0	0.02
Nitrite-N	0.36–12.0	0.004
Bromide	0.63–21.0	0.01
Nitrate-N	0.42–14.0	0.002
Orthophosphate-P	0.69–23.1	0.003
Sulfate	2.85–95.0	0.02

Method 300.0 are shown in Table 4. Similar methods, such as ASTM D 4327-97 and Standard Methods 4110, provide comparable performance. Advances in column and suppressor technology continues to improve the methodology for determination of these common anions. The IonPac® AS14 column provides complete resolution of fluoride and acetate and also improved resolution of fluoride from the void peak compared to the AS4A column. Figure 2 shows a chromatogram of a typical drinking-water sample obtained using an AS14 column with a 3.5 mM bicarbonate/1.0 mM carbonate eluent and suppressed conductivity detection.

EPA Method 300.0 is also validated for wastewater analysis, although such samples often require sample pretreatment before injection into the ion chromatograph. Dilution into the application range followed by filtration is often required, whereas pretreatment with SPE cartridges to remove hydrophobic organic material

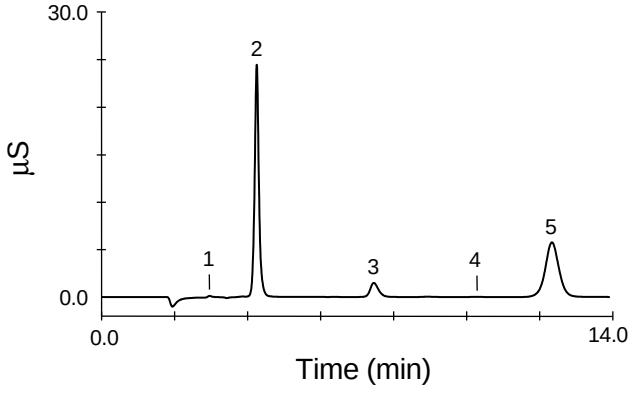

Figure 2 Determination of inorganic anions in drinking water. Conditions as for Figure 1, except: column, Dionex® IonPac® AS14; eluent, 3.5 mM sodium carbonate/1.0 mM sodium bicarbonate; injection volume, 50 µL; solutes, 1 = fluoride (0.03 mg L^{-1}), 2 = chloride (10.1 mg L^{-1}), 3 = nitrate (3.7 mg L^{-1}), 4 = phosphate (0.04 mg L^{-1}), 5 = sulfate (12.2 mg L^{-1}).

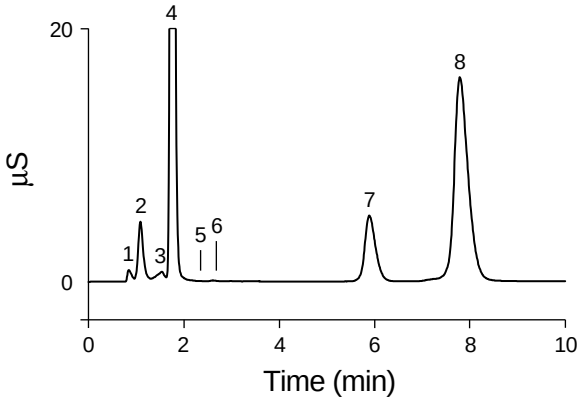

Figure 3 Determination of anions in domestic wastewater. Conditions as for Figure 1, except: injection volume, 50 µL; sample preparation, SPE cleanup with a Waters C_{18} Sep-Pak®; solutes, 1 = injection peak, 2 = acetate (4.2 mg L^{-1}), 3 = bicarbonate (not quantitated), 4 = chloride (49 mg L^{-1}), 5 = bromide (0.04 mg L^{-1}), 6 = nitrate (0.03 mg L^{-1}), 7 = phosphate (28 mg L^{-1}), 8 = sulfate (35 mg L^{-1}).

is recommended to prolong column lifetimes. Figure 3 shows a chromatogram of a typical domestic wastewater sample obtained using an AS4A column with a carbonate/bicarbonate eluent and suppressed conductivity detection. The performance of environmental methods, such as EPA Method 300.0, is typically validated through single- and multioperator precision and bias studies on spiked samples. Table 5 shows single-operator precision and bias data obtained using Method 300.0 for common anions spiked into reagent water, drinking water, and mixed domestic and industrial wastewaters. Acceptable precision and bias data were obtained for the determination of common anions in all three matrix types when using IC.

Table 5 USEPA Method 300.0 single-operator precision and bias[22]

Analyte	Matrices[a]	Added concentration ($mg\,L^{-1}$)	Mean[b] recovery (%)	Standard deviation ($mg\,L^{-1}$)
Fluoride	rw	2.0	91	0.05
	dw	1.0	92	0.06
	ww	1.0	87	0.07
Chloride	rw	20.0	96	0.35
	dw	20.0	108	1.19
	ww	20.0	101	5.2
Nitrite (N)	rw	10.0	91	0.14
	dw	10.0	121	0.25
	ww	5.0	91	0.50
Bromide	rw	5.0	99	0.08
	dw	5.0	105	0.10
	ww	5.0	105	0.34
Nitrate (N)	rw	10.0	103	0.21
	dw	10.0	104	0.27
	ww	10.0	101	0.82
Ortho-phosphate (P)	rw	10.0	99	0.17
	dw	10.0	99	0.26
	ww	10.0	106	0.85
Sulfate	rw	20.0	99	0.40
	dw	50.0	105	3.35
	ww	40.0	102	6.4

[a] Matrices: rw = reagent water; dw = drinking water; ww = wastewater (mixed domestic and industrial).
[b] Average of seven replicates.

5.1.2 Disinfection By-product Anions

USEPA Method 300.0 was revised in 1993 to include determination of the disinfection by-product (DBP) anions – bromate, chlorite, and chlorate. Bromate is a DBP produced from the ozonation of source water that contains naturally occurring bromide, whereas chlorite and chlorate are produced as a result of using chlorine dioxide as a disinfectant. These DBP anions pose significant health risks, even at low microgram per liter levels. Bromate has been judged as a potential carcinogen and the USEPA has estimated a potential cancer risk equivalent to 1 in 10^4 for a lifetime exposure to drinking water containing bromate at $5\,\mu g\,L^{-1}$.[44]

The occurrence of bromate and other DBPs in US drinking water has recently been documented by the USEPA through the comprehensive collection of data mandated by the information collection rule (ICR). The USEPA has proposed an MCL for bromate of $10\,\mu g\,L^{-1}$ and an MCL for chlorite of $1000\,\mu g\,L^{-1}$ in finished drinking water, which was promulgated in Stage I of the disinfectant and disinfection by-product (D/DBP) rule.[45] USEPA Method 300.0(B) specifies the use of a Dionex® AS9-SC anion-exchange column with a 1.7 mM sodium bicarbonate/1.8 mM sodium carbonate eluent and suppressed conductivity detection for the analysis of bromate, chlorite, and chlorate. However, Method 300.0(B), as originally published, cannot meet the quantitation requirements of the ICR and method modifications are required. These include the use of a weaker borate eluent to improve bromate and chloride resolution and the use of sample pretreatment to minimize chloride interference.[46]

In support of the publication of Stage I of the D/DBP rule, the USEPA recently developed Method 300.1 for the determination of inorganic and DBP anions. Method 300.1 specifies a Dionex® AS9-HC anion exchange column with a 9.0 mM sodium carbonate eluent and suppressed conductivity detection. The AS9-HC differs from the column specified in Method 300.0(B) in that it has higher capacity and improved separation of the key oxyhalide anions from potential interferences. Method 300.1(A) is applicable to common inorganic anions in drinking water, whereas 300.1(B) is applicable to the determination of DBP anions, and bromide, in drinking water. Methods 300.1(A) and (B) use different injection volumes to achieve different MDLs. Method 300.1(A) requires 10 µL whereas 300.1(B) requires 50 µL when using a 2 mm ID (internal diameter) column, or 50 µL and 200 µL injections, respectively, with a 4 mm ID column.[16] The relatively large injection volume is required for 300.1(B) in order to achieve the DLs necessary when analyzing DBP anions.

Figure 4 shows the separation of chlorite, bromate, and chlorate, in addition to the common inorganic anions using an AS9-HC column. The MDLs for 300.1(B) using a 50 µL injection and 2 mm ID column are 0.89, 1.44, 1.44, and $1.31\,\mu g\,L^{-1}$ for chlorite, bromate,

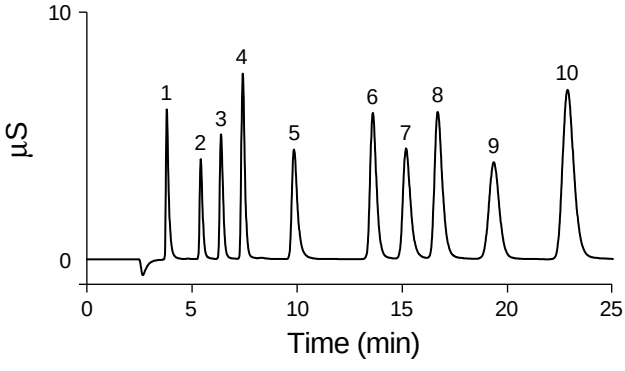

Figure 4 Separation of oxyhalides plus common inorganic anions using EPA Method 300.1. Conditions: column, Dionex® IonPac® AS9-HC (4 mm ID); eluent, 9.0 mM sodium carbonate; flow rate, $1.0\,mL\,min^{-1}$; detection, suppressed conductivity with an ASRS operated at 100 mA in external water mode; injection volume, 25 µL; solutes, 1 = fluoride ($3\,mg\,L^{-1}$), 2 = chlorite ($10\,mg\,L^{-1}$), 3 = bromate ($20\,mg\,L^{-1}$), 4 = chloride ($6\,mg\,L^{-1}$), 5 = nitrite ($15\,mg\,L^{-1}$), 6 = bromide ($25\,mg\,L^{-1}$), 7 = chlorate ($25\,mg\,L^{-1}$), 8 = nitrate ($25\,mg\,L^{-1}$), 9 = phosphate ($40\,mg\,L^{-1}$), 10 = sulfate ($30\,mg\,L^{-1}$).

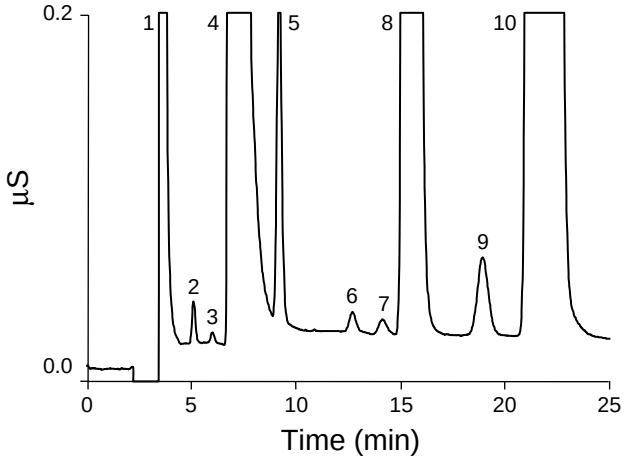

Figure 5 Determination of low-level oxyhalides and common anions in simulated drinking water. Conditions as for Figure 4, except: injection volume, 200 μL; solutes, 1 = fluoride (1 mg L^{-1}), 2 = chlorite (0.01 g L^{-1}), 3 = bromate (0.005 mg L^{-1}), 4 = chloride (50 mg L^{-1}), 5 = nitrite (0.1 mg L^{-1}), 6 = bromide (0.01 mg L^{-1}), 7 = chlorate (0.01 mg L^{-1}), 8 = nitrate (10 mg L^{-1}), 9 = phosphate (0.1 mg L^{-1}), 10 = sulfate (50 mg L^{-1}).

bromide, and chlorate respectively.[16] Figure 5 shows the chromatogram of a synthetic drinking water sample obtained using an AS9-HC column with a 9.0 mM carbonate eluent and suppressed conductivity detection. Despite using a 200 μL injection, no column overloading occurs and bromate can be determined at 5 μg L^{-1} in the presence of a 10 000-fold excess of chloride.[47] Figure 6 shows the application of Method 300.1(B) to the determination of DBP anions in drinking water from Sunnyvale, California. The water in this municipality is disinfected using hypochlorite, hence chlorate appears in the drinking water matrix, shown in Figure 6(a). The same drinking water spiked with 10 μg L^{-1} each of chlorite, bromate, bromide, and chlorate is shown in Figure 6(b), indicating that the DBP anions are clearly resolved from the common inorganic anions (and bromide) present in drinking water. Table 6 shows single-operator precision and bias data obtained using Method 300.1 for the DBP anions spiked into reagent water, drinking water, and high-ionic-strength drinking water.

5.1.3 Hexavalent Chromium

USEPA Methods 300.0 and 300.1, and equivalent methods in the USA and other countries, represent the most important and widely used applications of IC in environmental analysis. However, a number of other regulatory methods based on IC are applicable to drinking-water and wastewater analysis. Inorganic chromium is a primary drinking-water contaminant with an MCL of 0.1 mg L^{-1}. Hexavalent chromium is the most toxic form of the metal,

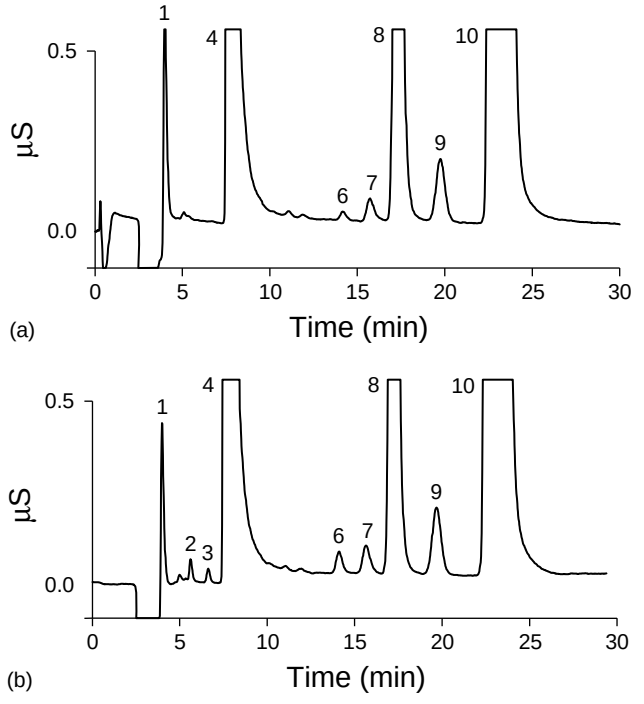

Figure 6 Determination of oxyhalides and common anions in Sunnyvale drinking water. Conditions as for Figure 4, except: sample, (a) drinking water and (b) drinking water spiked with 0.01 mg L^{-1} of oxyhalide anions and bromide; solutes, (a) 1 = fluoride (0.05 mg L^{-1}), 4 = chloride (19 mg L^{-1}), 6 = bromide (0.004 mg L^{-1}), 7 = chlorate (0.03 mg L^{-1}), 8 = nitrate (1.7 mg L^{-1}), 9 = phosphate (0.25 mg L^{-1}), 10 = sulfate (30 mg L^{-1}) and (b) 1 = fluoride (0.05 mg L^{-1}), 2 = chlorite (0.008 g L^{-1}), 3 = bromate (0.012 mg L^{-1}), 4 = chloride (19 mg L^{-1}), 6 = bromide (0.013 mg L^{-1}), 7 = chlorate (0.041 mg L^{-1}), 8 = nitrate (1.7 mg L^{-1}), 9 = phosphate (0.25 mg L^{-1}), 10 = sulfate (30 mg L^{-1}).

Table 6 USEPA Method 300.1 single-operator precision and bias[16]

Analyte	Matrices[a]	Added concentration (μg L^{-1})	Mean[b] recovery (%)	Standard deviation (μg L^{-1})
Chlorite	rw	100	96.2	0.95
	O3w	100	84.4	0.46
	hiw	100	102	2.19
Bromate	rw	5.00	101	0.45
	O3w	5.00	80.9	0.61
	hiw	5.00	97.5	0.95
Bromide	rw	20.0	104	0.80
	O3w	20.0	–[c]	3.67
	hiw	20.0	92.5	0.79
Chlorate	rw	100	98.3	0.80
	O3w	100	100	1.20
	hiw	100	86.1	1.47

[a] Matrices: rw = reagent water; O3w = ozonated drinking water; hiw = high-ionic-strength simulated drinking water.
[b] Average of nine replicates.
[c] Not calculated, as added amount was less than the unfortified amount.

in addition to being a suspected carcinogen. Hexavalent chromium can be separated as the chromate anion using a high capacity Dionex® AS7 anion-exchange column, as specified in USEPA Method 218.6.[18] In this case, detection is achieved using a UV/VIS detector after postcolumn reaction with diphenylcarbohydrazide, as this color-forming reagent provides a more sensitive and selective means for determining chromate than suppressed conductivity.

USEPA Method 218.6 and ASTM D 5257-93 are validated for the determination of hexavalent chromium in drinking water, groundwaters and industrial wastewaters. An MDL of $0.3\,\mu g\,L^{-1}$ for Cr(VI) in drinking and wastewater can be achieved using a 250 µL injection. Average recoveries in the order of 98–105% were obtained for $100\,\mu g\,L^{-1}$ Cr(VI) solutions spiked into reagent, drinking, ground, primary sewage waste and electroplating wastewaters.[16] Figure 7 shows a chromatogram of a spiked wastewater sample obtained using the conditions described in USEPA Method 218.6. No interfering peaks are observed when using this very specific detection approach for Cr(VI) analysis.

5.1.4 Cyanide

The highly toxic cyanide anion is a primary drinking-water contaminant which has an MCL of $0.2\,mg\,L^{-1}$. Sources of cyanide contamination in drinking water include effluents from the electroplating, steel, plastics and mining industries, in addition to certain fertilizers.[43]

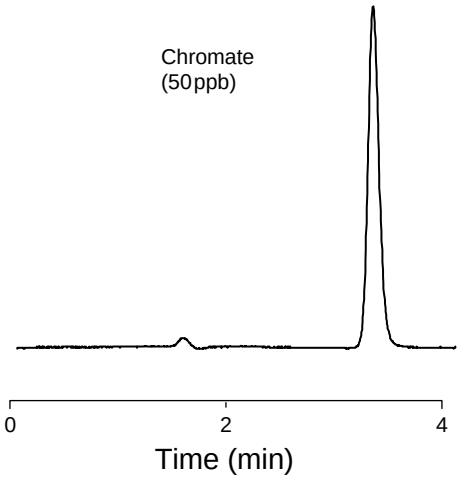

Figure 7 Determination of chromate in spiked wastewater using EPA Method 218.6. Conditions: column, Dionex® AS7; eluent, 250 mM ammonium sulfate/100 mM ammonium hydroxide; flow rate, $1.5\,mL\,min^{-1}$; detection, UV/VIS at 530 nm after postcolumn reaction with 2 mM diphenylcarbazide/10% methanol/1.0 N sulfuric acid delivered at $0.5\,mL\,min^{-1}$; injection volume, 100 µL; sample, filtered wastewater spiked with $50\,\mu g\,L^{-1}$ chromate.

Cyanide is classified according to its availability in the presence of complexing metals. Total cyanide refers to CN that can be released as hydrocyanic acid (HCN) from both the aqueous and particulate portions of a sample under total reflux distillation conditions and includes both free and complex cyanides. Free cyanide refers to CN that can be released as hydrocyanic acid from the aqueous portion of a sample by direct cyanide determination without reflux distillation. Cyanide amenable to chlorination refers to CN determined after chlorinating a portion of sample, and calculating the difference between total cyanide in the original and chlorinated samples. Weak acid dissociable cyanide refers to CN determined after distillation with a weak acid. This only releases CN bound in weak complexes, i.e. those with low stability constants, such as with Cu or Zn.[24]

In practice, the majority of CN determinations, particularly in wastewater samples, involve measurement of total CN, which is determined after reflux distillation of an alkaline sample in the presence of sulfuric acid and a magnesium chloride catalyst.[24] The released HCN is absorbed into a sodium hydroxide scrubber solution and the cyanide in this solution can be measured colorimetrically, by IC, FIA, titration, or ISE. An IC separation is recommended with electrochemical detection when sulfur, thiocyanate, or other sulfur-containing compounds are present in the sample, as H_2S codistills with HCN and can interfere with the FIA determination when using electrochemical detection.[24]

ASTM D 2036-97 is validated for the determination of total cyanide in drinking, ground, and surface waters, and both domestic and industrial wastes using IC, in addition to other analytical measurement techniques. Cyanide is separated on a Dionex® AS7 anion-exchange column using an eluent of 100 mM sodium hydroxide/500 mM sodium acetate/0.5% (V/V) ethylenediamine. The CN is then detected using an amperometric detector with a silver working electrode operated at −0.05 V.[24] This very sensitive detection approach provides an MDL of $2\,\mu g\,L^{-1}$ when using a 1.0 mL injection volume and can tolerate sulfur concentrations up to 100 times the cyanide level without degradation of method performance. Figure 8 shows a chromatogram of cyanide and sulfide in a spiked wastewater sample obtained using an AS7 column and amperometric detection. The applicable range of this method is from $10\,\mu g\,L^{-1}$ to $10\,mg\,L^{-1}$ cyanide; mean recoveries in the order of 85–98% were obtained for samples spiked with CN over the range $40–1000\,\mu g\,L^{-1}$.

5.1.5 Perchlorate

Ammonium perchlorate is a key ingredient in solid rocket propellants which has recently been found in groundwaters in regions of the USA where aerospace

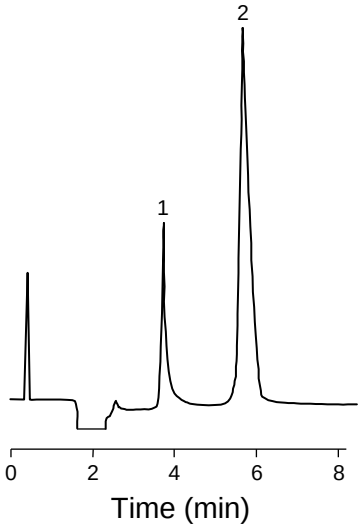

Figure 8 Determination of sulfide and cyanide in spiked wastewater using ASTM Method D2036-97. Conditions: column, Dionex® IonPac® AS7; eluent, 500 mM sodium acetate/100 mM sodium hydroxide; flow rate, 1.0 mL min^{-1}; detection, amperometry using a silver working electrode operated at −0.05 V versus Ag/AgCl reference; injection volume, 100 µL; sample, filtered spiked wastewater; solutes, 1 = sulfide (25 µg L^{-1}), 2 = cyanide (100 µg L^{-1}).

material, munitions, and fireworks have been developed, tested, or manufactured. Perchlorate has been found in ground and surface waters in California, Nevada, Utah, and West Virgina.[48] Perchlorate poses a human health risk and preliminary data from the USEPA reports that exposure to less than 4–18 µg L^{-1} as providing adequate health protection.[49] Perchlorate is listed on the USEPA contaminant candidate list as a research priority, although it is not currently regulated under the Federal SDWA.[43] Perchlorate contamination of public drinking-water wells is becoming a serious problem in some western States and the California Department of Health Services (CDHS) has adopted an action level for perchlorate in drinking water of 18 µg L^{-1}. To date, perchlorate has been detected in over 100 public drinking-water wells in California, with more than 20 wells being closed due to contamination.[49]

The CDHS developed an IC method based on the use of an hydrophilic IonPac® AS5 column, large loop injection, and suppressed conductivity detection to quantify perchlorate at low microgram per liter levels.[50] However, the use of an IonPac® AS11 column with an eluent of 100 mM hydroxide, 1000 µL injection, and suppressed conductivity detection provides an MDL for perchlorate of 0.3 µg L^{-1} without the need for an organic modifier in the mobile phase.[48] Figure 9 shows a chromatogram of perchlorate standard at 20 µg L^{-1}, whereas Figure 10 shows a drinking water sample spiked

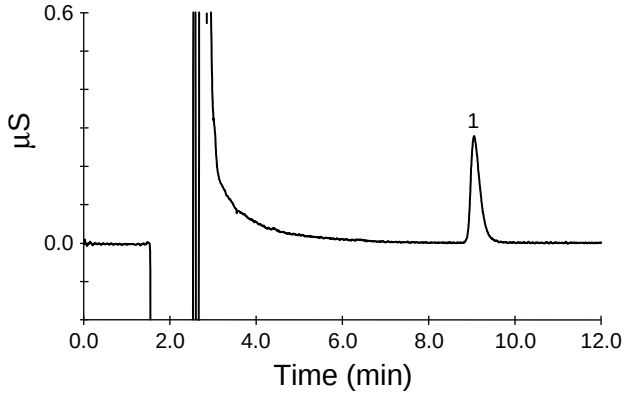

Figure 9 Chromatogram of 20 µg L^{-1} perchlorate standard. Conditions: column, Dionex® IonPac® AS11; eluent, 100 mM sodium hydroxide; flow rate, 1.0 mL min^{-1}; detection, suppressed conductivity with an ASRS-Ultra operated at 300 mA in external water mode; injection volume, 1000 µL; solute, perchlorate.

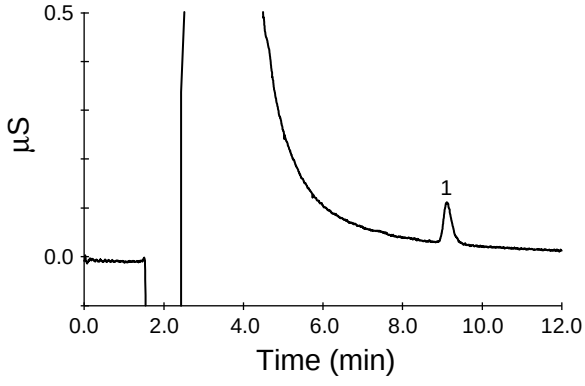

Figure 10 Analysis of perchlorate in spiked drinking water. Conditions as for Figure 9, except: sample, filtered Sunnyvale, CA, tapwater spiked with 6 µg L^{-1} perchlorate.

with 6.0 µg L^{-1} perchlorate. The applicable range of this method is from 2.0 to 100 µg L^{-1} perchlorate. The method is free of interferences from common anions and quantitative recoveries were obtained for low microgram per liter levels of perchlorate in spiked drinking and groundwater samples.[48]

5.1.6 Inorganic Cations and Ammonia

The preceding discussion shows that a number of regulatory methods based on IC are available for anion analysis; however, that is not case for cation analysis. The majority of the inorganic cations listed as primary drinking water contaminants are transition metals which are most commonly analyzed using spectroscopic methods, such as AAS, ICPAES, or ICPMS.[11] However, IC is a USEPA-approved method for the analysis of the ammonium cation, sodium, potassium, calcium, and magnesium

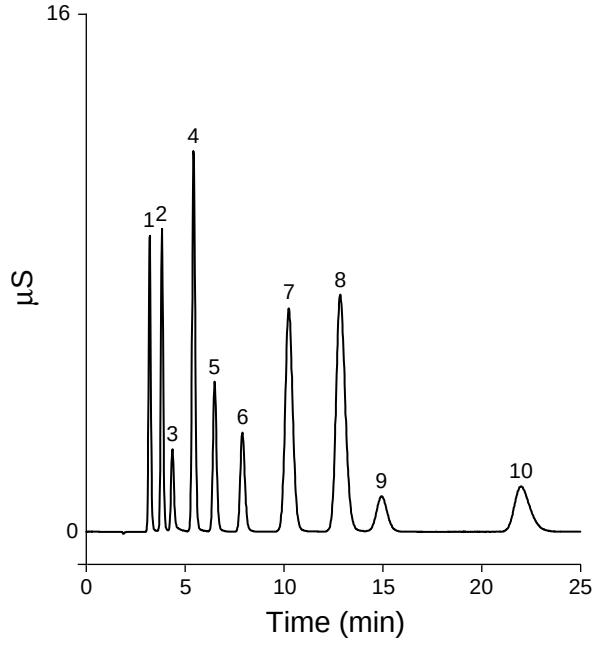

Figure 11 Separation of ammonia and alkali/alkaline earth cations. Conditions: column, Dionex® IonPac® CS12A; eluent, 18 mM methanesulfonic acid; flow rate, 1.0 mL min^{-1}; detection, suppressed conductivity with a cation self-regenerating suppressor (CSRS) operated at 100 mA in recycle mode; injection volume, 25 µL; solutes, 1 = lithium (1 mg L^{-1}), 2 = sodium (4 mg L^{-1}), 3 = ammonia (5 mg L^{-1}), 4 = potassium (10 mg L^{-1}), 5 = rubidium (10 mg L^{-1}), 6 = cesium (10 mg L^{-1}), 7 = magnesium (5 mg L^{-1}), 8 = calcium (10 mg L^{-1}), 9 = strontium (10 mg L^{-1}), 10 = barium (10 mg L^{-1}).

in rain water and wet precipitation, and provides a straightforward method for the simultaneous analysis of these cations.[19] Also, the ISO has recently published an ISO Method for the simultaneous determination of dissolved alkali and alkaline earth cations, ammonia, and manganese in water and wastewater using IC.[51] Figure 11 shows an example of a typical separation of alkali and alkaline earth cations and ammonia obtained using a Dionex® CS12A column with an eluent of methanesulfonic acid and suppressed conductivity detection.

IC is particularly beneficial for the simultaneous determination of cations plus ammonia in wastewaters containing amines, as these compounds can interfere with the conventional colorimetric or ISE methods used for ammonia analysis. Recently, an ASTM method based on IC has been proposed for the analysis of ammonia, sodium, potassium, calcium, and magnesium in reagent, drinking, surface, and groundwater and mixed industrial and domestic wastewater. This method uses a Dionex® CS15 column which has crown ether functionalities incorporated into the stationary phase to provide maximum resolution of ammonia and the

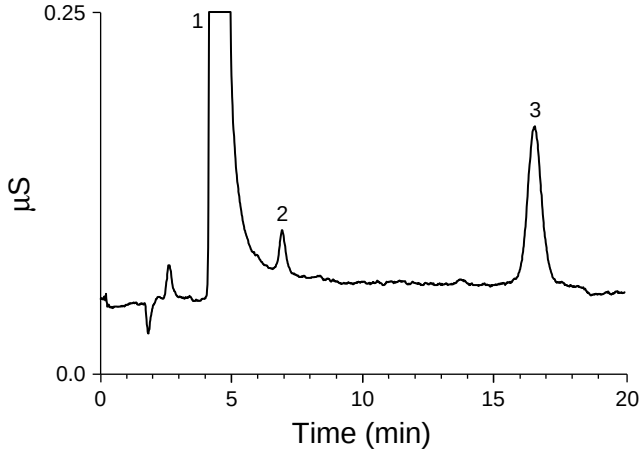

Figure 12 Determination of ammonia in industrial wastewater. Conditions: column, Dionex® IonPac® CS15; temperature, 40 °C; eluent, 10 mN sulfuric acid/10% methanol; flow rate, 1.2 mL min^{-1}; detection, suppressed conductivity with a CSRS operated at 100 mA in external water mode; injection volume, 25 µL; solutes, 1 = sodium (100 mg L^{-1}), 2 = ammonia (0.025 mg L^{-1}), 3 = calcium (not quantitated).

adjacent sodium peak. The use of this column with an eluent of 13 mN sulfuric acid and suppressed conductivity detection can achieve an MDL for ammonia of 1.0 µg L^{-1} using a 100 µL injection loop.[52] Figure 12 shows the application of this column for the determination of ammonia in industrial wastewater containing a large excess of sodium.

5.2 Natural Waters

In addition to being approved for a number of drinking-water and wastewater analyses, IC is also widely used for analysis of natural water samples. Many of the regulatory methods described in section 5.1 are also validated for natural waters, such as ground and surface waters. Natural waters encompass a wide variety of sample matrices, including: rain and acid rain; mineral spring waters and other groundwaters; surface waters (such as river, stream, lake, and pond waters); soil pore waters; runoff waters; snow, hail, and sleet; ice and ice cores; well and bore waters, etc. There are few regulations governing the analysis of such samples, hence a great diversity of IC methods are applied to a much wider range of analytes than in the highly regulated area of drinking water and wastewater analysis.

Although the key regulatory methods which use IC were described in the preceding pages, it is beyond the scope of this article to describe all of the applications of IC to natural-water analysis. Good overviews of the early use of IC for environmental applications are available,[9,10] and two very comprehensive listings of environmental analyses using IC were published in

1990.[5,53] In addition, proceedings of the International IC Symposium, published annually since 1991 by the *Journal of Chromatography*, contain excellent summaries of current developments and applications of IC for environmental and other analyses.

The analysis of rain water and acid rain is one of the more important applications of IC. The determination of ionic components in rain waters by IC is frequently used to estimate the effects of acidification on the natural and urban environments caused by acid rain.[54] The major ionic components of acid rain consist of the hydronium ion, sodium, ammonium, potassium, calcium, magnesium, and chloride, nitrate, and sulfate. Clearly, separate analyses using the separations presented in Figures 1 and 11 would provide the major components of an ion balance for an acid rain, once the sample pH had been measured. USEPA Methods 300.7 and 300.6 are based on similar, although somewhat dated, methods for the determination of these cations and anions, respectively, in acid rain and rain water.[19,55]

Many researchers have developed methods that enable the simultaneous determination of both anions and cations in rain waters and atmospheric aerosols, with perhaps the most comprehensive approach involving an automated quadruple IC system to perform isocratic anion, gradient anion (with and without sample preconcentration), and gradient cation separations using

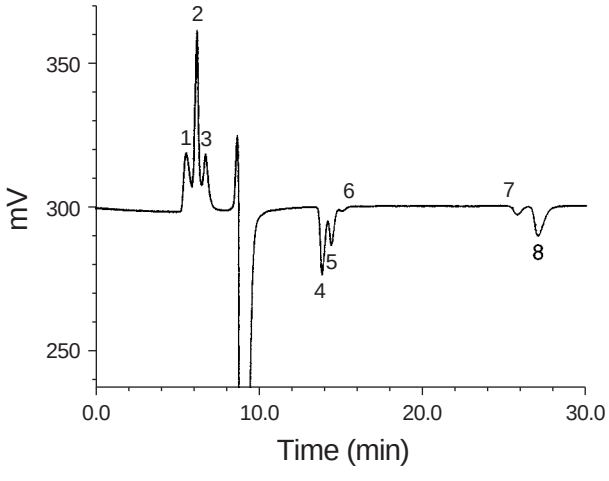

Figure 13 Simultaneous determination of anions and cations in acid rain. Conditions: column, Tosoh® TSKgel OA-PAK; eluent, 5 mM tartaric acid/7.5% methanol; flow rate, 1.2 mL min^{-1}; detection, nonsuppressed conductivity; injection volume, 100 μL; sample, acid rain from Nagoya, Japan; solutes, 1 = sulfate, 2 = chloride, 3 = nitrate, 4 = sodium, 5 = ammonia, 6 = potassium, 7 = magnesium, 8 = calcium. (Reprinted from K. Tanaka, K. Ohta, J.S. Fritz, S. Matsushita, A. Miyanaga, 'Simultaneous Ion-exclusion Chromatography Cation Exchange Chromatography with Conductometric Detection of Anions and Cations in Acid Rain Waters', *J. Chromatogr.*, **671**, 239–248, Copyright (1994) with permission from Elsevier Science.)

a single sample injection.[56] Tanaka et al. developed a less complex approach that enables the simultaneous determination of anions and cations in acid rain based upon a simultaneous ion-exclusion–cation-exchange separation and non-suppressed conductimetric detection.[54] Figure 13 shows an example of a chromatogram of anions and cations in rain water obtained using this approach.

Gradient elution is typically required to quantify all of the minor acid components of rain-water samples, as organic acid anions, such as formate, acetate, or methanesulfonate, are often present at low levels in rain-water samples.[57] Figure 14 shows an example of the resolving power that can be achieved with an hydroxide gradient using modern IC instrumentation, whereas Figure 15 shows the separation of inorganic anions and organic acids in rain water obtained using a borate gradient with a Dionex® AS11 column and suppressed conductivity detection. Gradient separations can be used to quantify a wide range of anionic solutes in complex water samples, such as hazardous-waste leachates. In addition to rain-water analysis, IC has also been used to analyze terrestrial waters and ice cores from pristine environments, such as Antarctica, in order to establish baseline levels of pollutants.[58,59]

The analysis of ground and surface waters is another common application of IC. The determination of inorganic solutes in waters from rivers, streams, lakes, and ponds is similar in complexity to the analysis of typical wastewater samples. Filtration followed by pretreatment with SPE cartridges to remove hydrophobic organic material is recommended when analyzing most ground and surface waters. Mineral waters are typical examples of natural groundwaters which are commonly analyzed by IC. These samples can significantly differ in both total mineralization and also chemical composition.[60] IC has been shown to be applicable to the determination of over 98.6% of the total cation composition and over 98.9% of the noncarbonate anion composition of mineral waters.[61] Similarly, IC is also frequently applied to the determination of inorganic ions in surface waters. Figure 16 shows the analysis of inorganic anions in lake water from Salt Lake in Utah obtained using an AS4A column and suppressed conductivity detection.

IC tends to be more commonly utilized for the determination of metal species in natural waters than in drinking-water and wastewater samples. The ability of IC to quantify different oxidation states, such as Fe(II) or Fe(III); and stable metal complexes, such as metallocyanides, or both inorganic and organic arsenic species is of great benefit in determining the fate, transport, and toxicity of metals in natural waters.[14,62] Figure 17 shows an example of the determination of low parts per billion levels of metal cations in

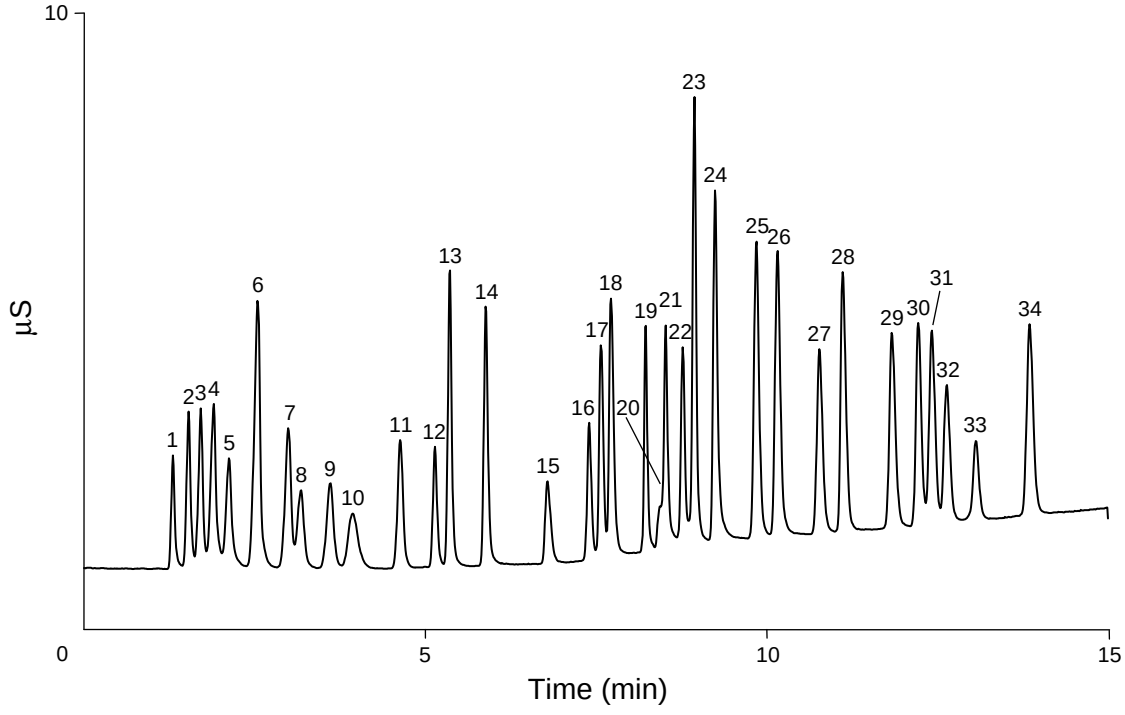

Figure 14 Hydroxide gradient separation of inorganic and organic acid anions. Conditions: column, Dionex® IonPac® AS11; eluent, deionized water/sodium hydroxide gradient; flow rate, 2.0 mL min^{-1}; detection, suppressed conductivity; solutes, 1 = isopropylethylphosphonate (5 mg L^{-1}), 2 = quinate (5 mg L^{-1}), 3 = fluoride (1 mg L^{-1}), 4 = acetate (5 mg L^{-1}), 5 = propionate (5 mg L^{-1}), 6 = formate (5 mg L^{-1}), 7 = methanesulfonate (5 mg L^{-1}), 8 = pyruvate (5 mg L^{-1}), 9 = chlorite (5 mg L^{-1}), 10 = valerate (5 mg L^{-1}), 11 = monochloroacetate (5 mg L^{-1}), 12 = bromate (5 mg L^{-1}), 13 = chloride (2 mg L^{-1}), 14 = nitrite (5 mg L^{-1}), 15 = trifluoroacetate (5 mg L^{-1}), 16 = bromide (3 mg L^{-1}), 17 = nitrate (3 mg L^{-1}), 18 = chlorate (3 mg L^{-1}), 19 = selenite (5 mg L^{-1}), 20 = carbonate (5 mg L^{-1}), 21 = malonate (5 mg L^{-1}), 22 = maleate (5 mg L^{-1}), 23 = sulfate (5 mg L^{-1}), 24 = oxalate (5 mg L^{-1}), 25 = ketomalonate (10 mg L^{-1}), 26 = tungstate (10 mg L^{-1}), 27 = phthalate (10 mg L^{-1}), 28 = phosphate (10 mg L^{-1}), 29 = chromate (10 mg L^{-1}), 30 = citrate (10 mg L^{-1}), 31 = tricarballylate (10 mg L^{-1}), 32 = isocitrate (10 mg L^{-1}), 33 = *cis*-aconitate (10 mg L^{-1}), 34 = *trans*-aconitate (10 mg L^{-1}).

river water obtained after preconcentrating 40 mL of sample.[63]

5.3 Brines

The analysis of brines by IC is complicated by the high ionic strength and excess sodium chloride in the sample. Nevertheless, IC is frequently used for the analysis of inorganic solutes in natural-water brines, which include seawater, subsurface brines, geothermal brines, and very high salinity groundwaters. The analysis of major components, e.g. chloride and sulfate, in brines is relatively straightforward, only requiring a dilution before direct injection into the IC device.[64] However, the analysis of minor components in brines typically requires careful selection of analytical conditions, and a wide variety of approaches have been used for the analysis of these minor components. The use of IC for the analysis of anions in high salt concentration environmental waters has recently been reviewed by Singh et al.[65] In addition to the use of conventional suppressed IC, other approaches include: using sodium chloride as the eluent; sample pretreatment with Ag-form membranes and SPE cartridges; the use of heart-cut matrix elimination or preconcentration combined with matrix elimination; and the use of more selective detection methods, such as UV/VIS absorption, amperometry, postcolumn derivatization, and even ICPMS.[30,32,58,65–72]

The use of sodium chloride as an eluent combined with low-wavelength UV detection allows the determination of UV-absorbing anions, such as nitrate, iodide, and molybdate, in samples containing up to 20 000 mg L^{-1} of chloride without loss of chromatographic performance.[66] A similar approach with an eluent containing chloride and an ion-pairing reagent has been used for the analysis of thiosulfate and polythionates in natural saline waters.[68] Sample pretreatment with Ag-form SPE cartidges or membranes to reduce chloride is a commonly used approach to allow determination of minor anionic components in brines. Chloride precipitates as AgCl, although other halides are also removed to a significant extent using this approach and poor recoveries can

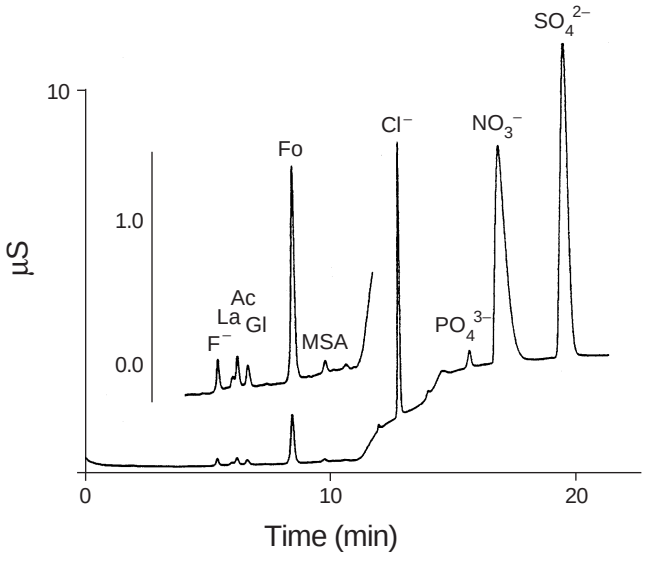

Figure 15 Determination of inorganic anions and organic acids in rain water using a borate gradient. Conditions: column, Dionex® IonPac® AS11; eluent, deionized water/sodium tetraborate gradient; flow rate, 1.0 mL min^{-1}; injection volume, 25 μL; detection, suppressed conductivity using an AMMS. (Reprinted from A.A. Ammann, T.B. Ruttimann, 'Simultaneous Determination of Small Organic and Inorganic Anions in Environmental Water Samples by Ion-exchange Chromatography', *J. Chromatogr.*, **706**, 259–269, Copyright (1995), with permission from Elsevier Science.)

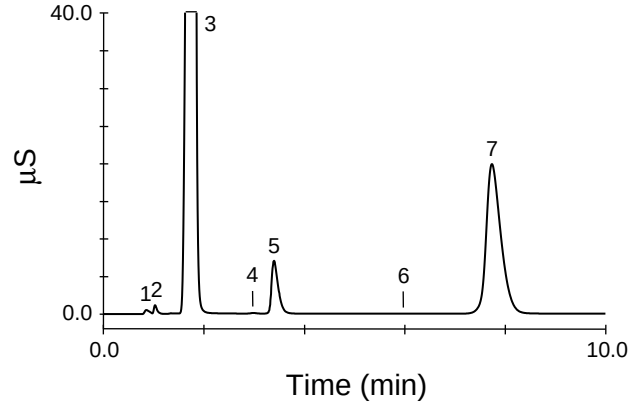

Figure 16 Determination of anions in lake water. Conditions as for Figure 3, except: sample, water from Salt Lake, Utah; solutes, 1 = injection peak, 2 = fluoride (0.28 mg L^{-1}), 3 = chloride (142 mg L^{-1}), 4 = bromide (0.18 mg L^{-1}), 5 = nitrate (11.2 mg L^{-1}), 6 = phosphate (0.28 mg L^{-1}), 7 = sulfate (44 mg L^{-1}).

be obtained for nitrite.[30,32] Selective detection can be applied to the determination of specific ions; for instance, iodide has been determined in brine using pulsed amperometry at a silver working electrode after separation on a Dionex® AS11 column with an eluent of 50 mM nitric acid. Iodide could be quantified at 16 μg L^{-1} in 30% NaCl

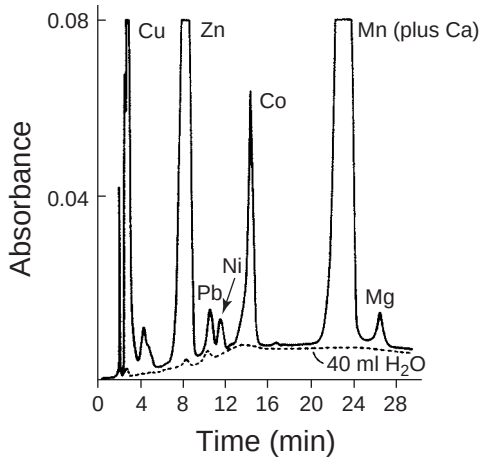

Figure 17 Determination of metal cations in river water obtained using sample preconcentration. Conditions: column, Nucleosil® C_{18}; eluent, 2 mM octanesulfonate/0.35–0.5 M tartrate gradient; detection, UV/VIS at 510 nm after postcolumn reaction with 4-(2-pyridylazo)-resorcinol (PAR); sample, 40 mL river water preconcentrated; solutes, copper (5 μg L^{-1}), zinc (47 μg L^{-1}), lead (0.9 μg L^{-1}), nickel (7.2 μg L^{-1}), cobalt (1.8 μg L^{-1}), manganese (38 μg L^{-1}), magnesium (430 μg L^{-1}). (Reprinted with permission from R.M. Cassidy, S. Elchuk, J.O. McHugh.[63] Copyright (1982) American Chemical Society.)

after a 10-fold dilution and calibration using standard addition.[69]

Automated matrix elimination techniques have also been used to determine anions and metals in seawater samples. Nitrite, bromide, nitrate, and sulfate have been determined in brine using a heart-cut and recycling system. A unresolved cut of the sample containing

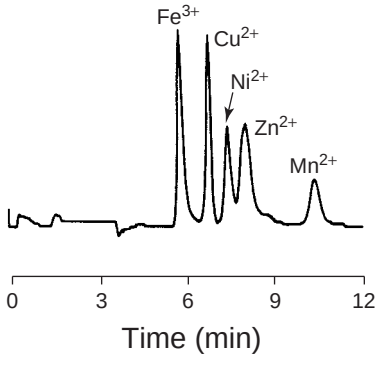

Figure 18 Determination of transition metals in seawater obtained using chelation concentration. Conditions: column, Dionex® IonPac® CS5; eluent, 35 mM pyridine 2,6-dicarboxylic acid; flow rate, 1.0 mL min^{-1}; detection, UV/VIS at 530 nm after postcolumn reaction with PAR; sample, 40 g seawater preconcentrated; solutes, iron (2.23 μg L^{-1}), copper (1.75 μg L^{-1}), nickel (1.34 μg L^{-1}), zinc (5.19 μg L^{-1}), manganese (1.85 μg L^{-1}). (Reprinted with permission from A. Siriraks, H.M. Kingston, J.M. Riviello.[71] Copyright (1990) American Chemical Society.)

chloride and nitrite is trapped in a sample loop and reinjected onto an anion-exchange column using four- and six-port switching valves. This method permits a DL of 0.5 mg L^{-1} for nitrite in spiked seawater samples when using IC with suppressed conductivity detection.[70] The use of a highly cross-linked, iminodiacetate functionalized macroporous resin has been shown to selectively retain transition metals and lanthanides in the presence of high levels of alkali and alkaline earth metals.[71] A Dionex® MetPac™ CC-1 chelating concentrator has been used to concentrate transition metals from seawater, after which the metals are eluted, via an intermediate cation exchange concentrator, to a CS5 analytical column with a pyridine 2,6-dicarboxylic acid eluent. The separated metals were detected using UV/VIS absorption after postcolumn derivatization with PAR. A chromatogram obtained using this approach for the determination of transition metals in seawater is shown in Figure 18; a similar approach has also been used for the analysis of uranium and thorium in seawater.[72]

6 SOIL, SLUDGE, AND SOLID WASTE ANALYSIS

6.1 Soil and Soil Extracts

The analysis of inorganic ions in soils was amongst the earliest applications of IC.[9,10] As is also the case with natural waters, the ability of IC to quantify different metal oxidation states and stable metal complexes in soil extracts is beneficial in determining the fate, transport and toxicity of metals in soil.[14,62] Also, the analysis of total nitrogen, phosphorus, and sulfur, and their corresponding oxide anions, e.g. nitrite, nitrate, phosphate, and sulfate, is of importance in assessing soil condition and fertility.

Ideally, solutions injected for IC analysis should be low in organic materials, strong acids, and soluble salts; hence, many of the traditional methods used for the extraction or digestion of soils are not compatible with IC.[73] However, modifications of traditional extraction methods, including the use of water or more dilute salt solutions, have now been developed to allow the application of IC to soil analysis. In fact, a wide variety of extraction and dissolution conditions have now been used for soil samples prior to IC analysis. Table 7 gives examples of species which have been determined in soils by IC, and typical extraction and dissolution conditions used for sample preparation. The extraction process generally involves forming a slurry of a dried soil sample with approximately 10 times its mass of extracting solution. The slurry is then mixed using a suitable mechanical device, such as a rotation apparatus, wrist shaker, or sonic bath, for a period anywhere from 10 min to 24 h depending upon the sample.[17] The slurry is then centrifuged, filtered, pretreated using SPE if necessary, and injected into the ion chromatograph. Figure 19 shows an example of the determination of anions in an aqueous extract of a fertilized soil sample.

Alkali and alkaline earth cations can simply be extracted from soils using a dilute mineral acid.[73] However, transition metals are typically bound to complexing agents in soils and require more severe extraction conditions or total digestion of the sample. Metals, such as Cd, Cu, Mn, or Zn, have been determined in soils after

Table 7 Examples of typical extraction and dissolution conditions used for soil samples prior to IC analysis[5,8,73–78]

Extraction/dissolution conditions	Analytes determined
Aqueous extraction	Cl, NO_2, Br, NO_3, PO_4, SO_4
Carbonate/bicarbonate buffer extraction	Cl, NO_2, Br, NO_3, PO_4, SO_4
0.01 M sodium hydroxide extraction	F, Cl, NO_2, NO_3, PO_4, SO_4
Aqueous extraction	Cl, NO_3, SeO_3, SO_4, SeO_4
Aqueous extraction	Cl, NO_3, WO_5, SO_4
10 mM potassium chloride extraction	NO_3, SO_4
0.15% calcium chloride extraction	NO_3, SO_4
10% potassium chloride extraction, dilution	NO_2, NO_3
Aqueous extraction	AsO_3, AsO_5
Fusion in carbonate, aqueous extraction	Br, I
16 mM phosphate extraction	SO_4
Combustion in carbonate, digestion in 1.0 M ammonium acetate, dilution	SO_4
Aqueous extraction	ClO_4
1.0 M ammonium acetate extraction, ashing, dissolution in 5 mM HCl	Na, K, Mg, Ca
40 mM nitric acid extraction	Li, Na, NH_4, K, Mg, Ca
Digestion in nitric, perchloric, and hydrofluoric acids, extraction using dithizone in chloroform, dissolution in nitric acid	Cd, Cu, Mn, Zn
0.36 M ammonium oxalate extraction	Fe(II), Fe(III)
0.01 M calcium carbonate extraction	Al, Al–F complexes

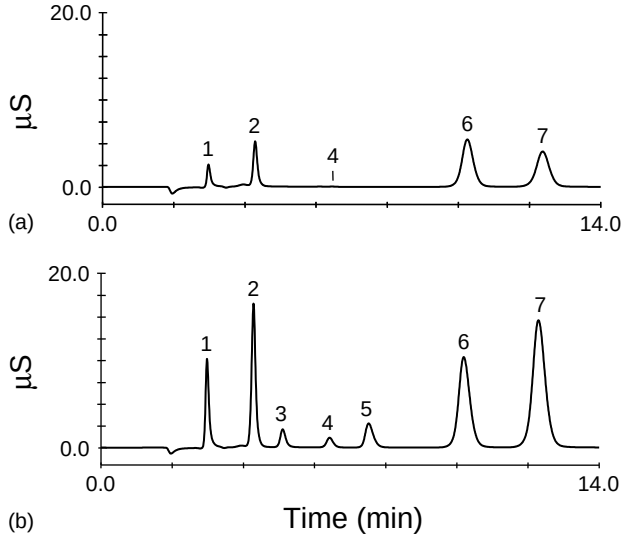

Figure 19 Determination of inorganic anions in an aqueous soil extract. Conditions as for Figure 2 except: sample, (a) aqueous (1:7) soil extract (b) aqueous (1:7) soil extract spiked with 2, 5, 2, 2, 5, 20, and 20 mg L^{-1} of solutes 1–7, respectively; solutes (a) 1 = fluoride (0.6 mg L^{-1}), 2 = chloride (2.6 mg L^{-1}), 4 = bromide (0.3 mg L^{-1}), 6 = phosphate (25.5 mg L^{-1}), 7 = sulfate (8.2 mg L^{-1}), (b) 1 = fluoride (2.7 mg L^{-1}), 2 = chloride (6.7 mg L^{-1}), 3 = nitrite (1.8 mg L^{-1}), 4 = bromide (2.1 mg L^{-1}), 5 = nitrate (4.6 mg L^{-1}), 6 = phosphate (47.6 mg L^{-1}), 7 = sulfate (27.1 mg L^{-1}).

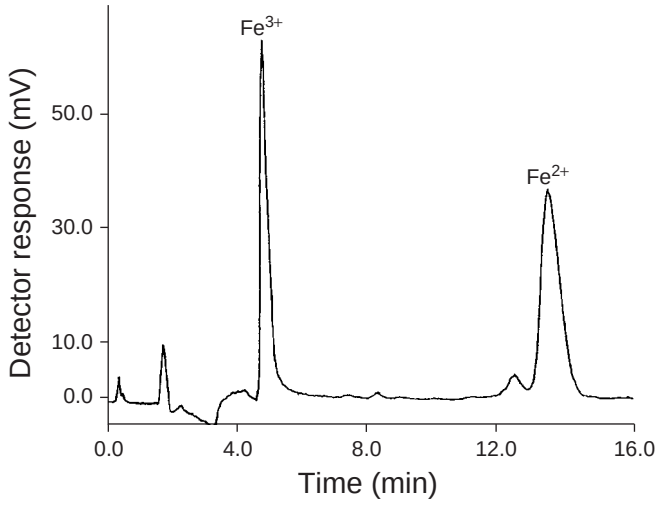

Figure 20 Determination of ferric and ferrous ions extracted from soil with 0.5 M HCl. Conditions as for Figure 18 except: eluent, 6 mM pyridine 2,6-dicarboxylic acid/50 mM acetic acid/50 mM sodium acetate; sample, 0.5 M HCl (1:10) soil extract; solutes, Fe(III) at 375 μM, Fe(II) at 239 μM. (Reprinted from M.Y. Ye, Y. Shen, C.C. West, W.G. Lyon,[77] 551–565 by courtesy of Marcel Dekker Inc.)

digestion in mixed nitric, perchloric, and hydrofluoric acids, followed by extraction using dithizone in chloroform, and destruction of the metal–dithizonate complex in nitric acid. The metals were chromatographed using a CS5 analytical column with a pyridine 2,6-dicarboxylic acid eluent and detected using UV/VIS absorption after postcolumn derivatization with PAR.[76] In some circumstances, metals can be released from soils using simple extraction procedures. The microbial reduction of Fe(III) to Fe(II) in soils and sediments is an important geochemical process and both hydrochloric acid (0.5 M) and ammonium oxalate (0.36 M) have been used to extract ferrous and ferric ions from these samples prior to IC analysis.[77] Figure 20 shows a chromatogram of Fe(III) and Fe(II) in a 0.5 M HCl extract of a soil sample.

6.2 Sludge, Leachates, and Solid Waste

The quantitation of inorganic ions in sludge, leachates and similar solid wastes by IC is similar in practice to the analysis of soil samples. Samples such as sludges, landfill, and solid (or hazardous) wastes are typically leached under aqueous conditions, then filtered and pretreated using SPE if necessary prior to injection. Although IC is most commonly applied to the analysis of anions in these sample types,[8,79,80] the determination of volatile fatty acids in landfill leachates is an important indication of the maturity of the waste deposit.[80,81] Low-molecular-mass organic acids can be separated either using an ion-exclusion separation or by gradient elution with an hydroxide eluent and suppressed conductivity detection. Figure 21 shows a chromatogram of volatile fatty acids in an immature landfill leachate, after 40 × dilution and acid treatment to remove carbonate, obtained using an ion-exclusion separation on a Dionex® HPICE/AS1 column.[80] Alternatively, gradient elution with an hydroxide eluent and suppressed conductivity detection can also be used for the determination of these volatile fatty acids by IC.[80,82]

The quantitation of metals in sludges and solid wastes also requires more severe extraction conditions or digestion of the sample, as is the case with soil samples. Total metals have been determined in sewage sludges by IC after dissolution in mixed nitric, perchloric, and hydrofluoric acids, followed by extraction using dithizone in chloroform, and destruction of the metal–dithizonate complex in nitric acid.[83] The metals were then chromatographed using a CS5 cation-exchange column and detected using UV/VIS absorption after postcolumn derivatization with PAR. Metals, such as Cr(III), Al(III), and Fe(II), have been analyzed in tannery sludges after extraction with sulfuric acid. The dried sludge samples were contacted for 24 h in concentrated sulfuric acid (1:20 ratio), then filtered, diluted, and passed through

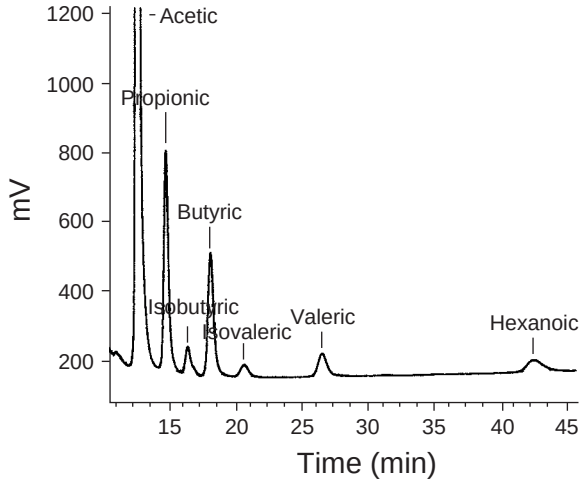

Figure 21 Determination of volatile fatty acids in immature landfill leachate. Conditions: column, Dionex® HPICE/AS1 (high-performance ion chromatography exclusion–anion separator 1); eluent, 1.0 mM octanesulfonic acid; flow rate, 2.0 mL min^{-1}; detection, suppressed conductivity with an AMMS/ICE (anion micro-membrane suppressor/ion chromatography exclusion) suppressor; injection volume, 50 μL; sample, aqueous (1 : 40) extraction of immature landfill leachate treated with octanesulfonic acid to remove carbonate. (Reprinted from D.A.C. Manning, A. Bewsher, 'Determination of Anions in Landfill Leachates by Ion Chromatography', *J. Chromatogr.*, **770**, 203–210, Copyright (1997) with permission from Elsevier Science.)

XAD-7 resin to remove organic matter.[84] The metals were derivatized with 8-hydroxyquinoline and the neutral complexes separated on a C_{18} reversed-phase column and detected using UV/VIS absorption after postcolumn reaction with PAR. Recoveries in the range of 64–80% were obtained for Cr(III), Al(III), and Fe(II) in the sludge samples.[84]

Sludges and solid-waste samples can also prepared for analysis by IC using combustion methods. USEPA Method 9056 is applicable to the determination of common inorganic anions in solid-waste combustion extracts.[23] The weighed sample is placed in a sample cup and combusted in a Parr bomb containing oxygen under pressure. Released acid gases are trapped in a dilute carbonate/bicarbonate solution which can be analyzed for anions using an anion-exchange column with a carbonate/bicarbonate eluent and suppressed conductivity detection. The method is applicable to the determination of anions at levels greater than 500 mg L^{-1} in solid wastes, virgin and used fuel oils, fuels, and related materials.[23]

7 QUALITY ASSURANCE

QC is an essential part of environmental analysis when it comes to generating reliable results using IC, or any analytical method. A great deal of literature is available in the area of quality management and the subject of QC in environmental analysis is covered in detail in the article **Quality Assurance in Environmental Analysis**; the specifics of quality assurance in environmental analysis have also been described elsewhere.[85] In addition, USEPA Methods contain detailed (and mandatory) QC sections specific to the method of analysis. For instance, USEPA Methods 300.0 and 300.1 provide detailed instructions on QC procedures to be implemented when analyzing inorganic anions and DBP anions in environmental samples using IC.[22,16] A example of the requirements of a typical QC section contained in an IC method intended for environmental analysis is detailed in Table 8.[86]

Table 8 Typical QC section of an IC method used for environmental analysis[86]

QC procedure	Intent of procedure
IDP	Analysis of seven replicates of IDP solution to demonstrate laboratory (or operator) proficiency using the test method
Initial calibration verification using CVS	Run CVS to check calibration standards and acceptable instrument performance
Run one CVS with each sample batch. A batch is typically defined as somewhere from 10 up to a maximum of 20 samples	Ongoing verification of previously established calibration curves; analyte concentrations to fall within acceptable limits (typically ±15% known value)
Run one reagent blank with each sample batch	Ongoing check for contamination introduced by the laboratory or method
Run one QCS with each sample batch	The analyte recoveries of the QCS should fall within control limits of $x \pm 3S$, where x is the mean recovery and S is the standard deviation
Run one matrix spike with each sample batch	Ongoing test of method recovery
Run one matrix duplicate with each sample batch	Ongoing test of method precision
Additional QC	Any laboratory may perform additional QC as desired or appropriate to their own internal quality program

CVS, calibration verification standard; IDP, initial demonstration of performance; QCS, quality control sample.

ACKNOWLEDGMENT

All chromatograms are courtesy of the Dionex® Corporation, unless otherwise stated. The author would like to thank the Dionex® Corporation for permission to publish the chromatograms used in this article.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AMMS	Anion Micro-membrane Suppressor
AMMS/ICE	Anion Micro-membrane Suppressor/ Ion Chromatography Exclusion
AOAC	Association of Official Analytical Chemists
ASRS	Anion Self-regenerating Suppressor
ASTM	American Society for Testing and Materials
CDHS	California Department of Health Services
CSRS	Cation Self-regenerating Suppressor
CVS	Calibration Verification Standard
CWA	Clean Water Act
DBP	Disinfection By-product
D/DBP	Disinfectant and Disinfection By-product
DL	Detection Limit
FIA	Flow Injection Analysis
HPICE/AS1	High-performance Ion Chromatography Exclusion–Anion Separator 1
IC	Ion Chromatography
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICR	Information Collection Rule
ID	Internal Diameter
IDP	Initial Demonstration of Performance
ISE	Ion-selective Electrode
ISO	International Organization for Standardization
MCL	Maximum Contaminant Level
MDL	Method Detection Limit
MS	Mass Spectrometry
NIOSH	National Institute of Occupational Safety and Health
NPDES	National Pollution Discharge Elimination System
OSHA	Occupational Safety and Health Administration
PAR	4-(2-Pyridylazo)-resorcinol
QC	Quality Control
QCS	Quality Control Sample
SDWA	Safe Drinking Water Act
SPE	Solid-phase Extraction
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of • Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring

Forensic Science **(Volume 5)**
Capillary Ion Electrophoresis in Forensic Science

Liquid Chromatography **(Volume 13)**
Ion Chromatography

REFERENCES

1. H. Small, T.S. Stevens, W.C. Bauman, 'Novel Ion Exchange Chromatographic Method Using Conductometric Detection', *Anal. Chem.*, **47**, 1801–1809 (1975).
2. D.T. Gjerde, J.S. Fritz, G. Schmuckler, 'Anion Chromatography with Low-conductivity Eluents', *J. Chromatogr.*, **186**, 509–519 (1979).
3. D.T. Gjerde, G. Schmuckler, J.S. Fritz, 'Anion Chromatography with Low-conductivity Eluents', *J. Chromatogr.*, **187**, 35–45 (1980).
4. I. Molnar, H. Knauer, D. Wilk, 'High-performance Liquid Chromatography of Ions', *J. Chromatogr.*, **201**, 225–240 (1980).
5. P.R. Haddad, P.E. Jackson, 'Ion Chromatography: Principles and Applications', in *J. Chromatogr. Library*, Elsevier, Amsterdam, Vol. 46, 1990.
6. C.A. Pohl, J.R. Stillian, P.E. Jackson, 'Factors Controlling Ion-exchange Selectivity in Suppressed Ion Chromatography', *J. Chromatogr.*, **789**, 29–41 (1997).

7. H. Small, *Ion Chromatography*, Plenum Press, New York, 1989.
8. J. Weiss, *Ion Chromatography*, 2nd edition, VCH, Weinheim, 1995.
9. E. Sawicki, J.D. Mulik, E. Wittgenstein (eds.), *Ion Chromatographic Analysis of Environmental Pollutants*, Ann Arbor Science Publications, Ann Arbor, MI, Vol. 1, 1978.
10. J.D. Mulik, E. Sawicki (eds.), *Ion Chromatographic Analysis of Environmental Pollutants*, Ann Arbor Science Publications, Ann Arbor, MI, Vol. 2, 1979.
11. A.E. Greenberg, L.S. Clesceri, A.D. Eaton (eds.), *Standard Methods for the Examination of Water and Wastewater*, 18th edition, American Public Health Association, Washington, DC, 1992.
12. A. Eaton, M. Carter, A. Fitchett, J. Oppenheimer, M. Bollinger, J. Sepikas, 'Comparability of Ion Chromatography and Conventional Methods for Drinking Water Analysis', in *Proc. AWWA Water Qual. Tech. Conf.*, 175–188, 1984.
13. S.S. Heberling, D. Campbell, S. Carson, 'Analysis of Chromium Plating Solutions and Wastewater by Ion Chromatography', *Plating Sur. Fin.*, November, 58–62 (1990).
14. S. Schnell, S. Ratering, K. Jansen, 'Simultaneous Determination of Iron(III), Iron(II), and Manganese(II) in Environmental Samples by Ion Chromatography', *Environ. Sci. Technol.*, **32**, 1530–1537 (1998).
15. ASTM D 5542-94, 'Standard Test Methods for Trace Anions in High Purity Water by Ion Chromatography', ASTM, Philadelphia, PA, 696–703, Vol. 11.01, 1995.
16. USEPA Method 300.1, 'The Determination of Inorganic Anions in Water by Ion Chromatography', USEPA, Cincinnati, OH, 1997.
17. C.A. Lucy, 'Practical Aspects of Ion Chromatographic Determinations', *LC/GC*, **14**, 406–415 (1996).
18. USEPA Method 218.6, 'Determination of Dissolved Hexavalent Chromium in Drinking Water, Groundwater and Industrial Wastewater Effluents by Ion Chromatography', USEPA, Cincinnati, OH, 1991.
19. USEPA Method 300.7, 'Dissolved Sodium, Ammonium, Potassium, Magnesium, and Calcium in Wet Deposition by Chemically Suppressed Ion Chromatography', USEPA, Cincinnati, OH, 1986.
20. S.R. Villasenor, 'Matrix Elimination in Ion Chromatography by Heart-cut Column Switching Techniques', *J. Chromatogr.*, **602**, 155–161 (1992).
21. P.E. Jackson, P.R. Haddad, S. Dilli, 'The Determination of Nitrate and Nitrite in Cured Meats Using High Performance Liquid Chromatography', *J. Chromatogr.*, **295**, 471–478 (1984).
22. USEPA Method 300.0, 'The Determination of Inorganic Anions in Water by Ion Chromatography', USEPA, Cincinnati, OH, 1993.
23. USEPA 'SW-846 Test Methods for Evaluating Solid Waste Physical/Chemical Methods', Update III, USEPA, Washington, DC, 1996.
24. ASTM D 2036 – 97, 'Standard Test Methods for Cyanide in Water', ASTM, Philadelphia, PA, 1997.
25. P.E. Jackson, J. Krol, A.L. Heckenberg, M. Mientjes, W. Staal, 'Determination of Total Nitrogen in Food, Environmental and Other Samples by Ion Chromatography after Kjeldahl Digestion', *J. Chromatogr.*, **546**, 405–410 (1991).
26. P.E. Jackson, J. Carnevale, H. Fuping, P.R. Haddad, 'Determination of Thorium and Uranium in Mineral Sands by Ion Chromatography', *J. Chromatogr.*, **671**, 181–191 (1994).
27. A.M. Quinn, K.W.M. Siu, G.J. Gardener, S.S. Berman, 'Determination of Heteroatoms in Organic Compounds by Ion Chromatography after Schoniger Flask Decomposition', *J. Chromatogr.*, **370**, 203–205 (1986).
28. J.R. Kreling, F. Block, G.T. Louthan, J. DeZwaan, 'Sulfur and Chlorine Organic Micro-analysis using Ion Chromatography', *Microchem. J.*, **34**, 158–165 (1986).
29. R. Bagchi, P.R. Haddad, 'Contamination Sources in the Clean-up of Samples for Inorganic Ion Analysis', *J. Chromatogr.*, **351**, 541–547 (1986).
30. I.K. Henderson, R. Saari-Nordhaus, J.M. Anderson, 'Sample Preparation for Ion Chromatography by Solid Phase Extraction', *J. Chromatogr.*, **546**, 61–71 (1991).
31. R.J. Joyce, H.S. Dhillon, 'Trace Level Determination of Bromate in Ozonated Drinking Water Using Ion Chromatography', *J. Chromatogr.*, **671**, 165–171 (1994).
32. R. Saari-Nordhaus, L.M. Nair, J.M. Anderson, 'Elimination of Matrix Interferences in Ion Chromatography by the Use of Solid Phase Extraction Disks', *J. Chromatogr.*, **671**, 159–163 (1994).
33. P.E. Jackson, W.R. Jones, 'Membrane-based Sample Preparation Device for the Pretreatment of Acidic Samples Prior to Cation Analysis by Ion Chromatography', *J. Chromatogr.*, **586**, 283–289 (1991).
34. P.R. Haddad, S. Laksana, 'On-line Analysis of Alkaline Samples with a Flow-through Electrodialysis Device Coupled to an Ion Chromatograph', *J. Chromatogr.*, **671**, 131–139 (1994).
35. P.E. Jackson, W.R. Jones, 'Hollow-fiber Membrane-based Sample Preparation Device for the Clean-up of Brine Samples Prior to Ion Chromatographic Analysis', *J. Chromatogr.*, **538**, 497–503 (1991).
36. D.F. Pensenstadler, M.A. Fulmer, 'Pure Steam Ahead', *Anal. Chem.*, **53**, 859A–868A (1981).
37. P.E. Jackson, P.R. Haddad, 'Studies on Sample Preconcentration in Ion Chromatography VII. Review of Methodology and Applications of Anion Preconcentration', *J. Chromatogr.*, **439**, 37–48 (1988).
38. P.R. Haddad, N.E. Rochester, 'Ion-interaction Reversed-phase Chromatographic Method for the Determination of Gold(I) Cyanide in Mine Process Liquors using Automated Sample Preconcentration', *J. Chromatogr.*, **439**, 23–36 (1988).

39. USEPA Method 200.10, 'Determination of Trace Elements in Marine Waters by On-line Chelation Preconcentration and Inductively Coupled Plasma/Mass Spectrometry', USEPA, Cincinnati, OH, 1992.
40. Dionex®, 'Official US Analytical Methods Specifying Ion Chromatography', Dionex® Corporation, Sunnyvale, CA, 1996.
41. Dionex®, 'Environmental Analysis Seminar', Dionex® Corporation, Sunnyvale, CA, 1995.
42. Australian Standard AS 3741-1990, 'Recommended Practice for Chemical Analysis by Ion Chromatography', Standards Australia, North Sydney, NSW, 1990.
43. USEPA, Fed. Reg., 62 FR 52193, 1997.
44. USEPA, Fed. Reg., 59 (145) FR 38709, 1994.
45. USEPA, Fed. Reg., 63 (241) FR 69389, 1998.
46. USEPA, 'DBP/ICR Analytical Methods Manual', USEPA, Cincinnati, OH, 1996.
47. L.K. Jackson, R.J. Joyce, M. Laikhtman, P.E. Jackson, 'Determination of Trace Level Bromate in Drinking Water by Direct Injection Ion Chromatography', *J. Chromatogr.*, **829**, 187–192 (1998).
48. K. Wirt, M. Laikhtman, J. Rohrer, P.E. Jackson, 'Low-level Perchlorate Analysis in Drinking Water and Groundwater by Ion Chromatography', *Am. Env. Lab.*, **10**(4), 1–5 (1998).
49. California Department of Health Services, 'Perchlorate in California Drinking Water', Update, October, 1998.
50. California Department of Health Services, 'Determination of Perchlorate by Ion Chromatography', California Department of Health Services, Berkeley, CA, 1997.
51. ISO 14911, 'Water quality – Determination of Dissolved Li^+, Na^+, NH_4^+, K^+, Mn^+, Ca^{2+}, Mg^{2+}, Sr^{2+} and Ba^{2+} using Ion Chromatography – Method for Water and Wastewater', ISO, Geneva, 1998.
52. ASTM D 19.05.03.23 (draft), 'Standard Test Method for Determination of Inorganic Cations and Ammonia in Water and Wastewater by Ion Chromatography', ASTM, Philadelphia, PA, 1999.
53. W.T. Frankenberger, H.C. Mehra, D.T. Gjerde, 'Environmental Applications of Ion Chromatography', *J. Chromatogr.*, **504**, 211–245 (1990).
54. K. Tanaka, K. Ohta, J.S. Fritz, S. Matsushita, A. Miyanaga, 'Simultaneous Ion-exclusion Chromatography Cation Exchange Chromatography with Conductometric Detection of Anions and Cations in Acid Rain Waters', *J. Chromatogr.*, **671**, 239–248 (1994).
55. USEPA Method 300.6, 'Chloride, Orthophosphate, Nitrate, and Sulfate in Wet Deposition by Chemically Suppressed Ion Chromatography', USEPA, Cincinnati, OH, 1986.
56. E. Dabek-Zlotorzynska, J.F. Dlouhy, 'Automated Simultaneous Determination of Anions and Cations in Atmospheric Aerosols by Ion Chromatography', *J. Chromatogr.*, **640**, 217–226 (1993).
57. A.A. Ammann, T.B. Ruttimann, 'Simultaneous Determination of Small Organic and Inorganic Anions in Environmental Water Samples by Ion-exchange Chromatography', *J. Chromatogr.*, **706**, 259–269 (1995).
58. K.A. Welch, W.B. Lyons, E. Graham, K. Neumann, J.M. Thomas, D. Mikesell, 'Determination of Major Elements in Terrestrial Waters from Antarctica by Ion Chromatography', *J. Chromatogr.*, **739**, 257–263 (1996).
59. J. Ivask, J. Pentchuk, 'Analysis of Ions in Polar Ice Core Samples by Use of Large Injection Volumes in Ion Chromatography', *J. Chromatogr.*, **770**, 125–127 (1997).
60. N. Gros, B. Gorenc, 'Simple Recognition of Similar Samples for the Ion Chromatographic Determination of the Main Cations', *J. Chromatogr.*, **789**, 323–327 (1997).
61. N. Gros, B. Gorenc, 'Performance of Ion Chromatography in the Determination of Anions and Cations in Various Natural Waters with Elevated Mineralization', *J. Chromatogr.*, **770**, 119–124 (1997).
62. J.M. Riviello, A. Siriraks, R.M. Manabe, R. Roehl, M. Alforque, 'Using IC to Enhance the Performance of Inductively Coupled Plasma/Optical Emission Spectrometry and Inductively Coupled Plasma/Mass Spectrometry', *LC/GC*, **9**, 704–712 (1991).
63. R.M. Cassidy, S. Elchuk, J.O. McHugh, 'Determination of Metals in Groundwaters by Trace Enrichment and Liquid Chromatography', *Anal. Chem.*, **54**, 727–731 (1982).
64. R. Kadnar, J. Rieder, 'Determination of Anions in Oilfield Waters by Ion Chromatography', *J. Chromatogr.*, **706**, 301–305 (1996).
65. R.P. Singh, N.M. Abbas, S.A. Smesko, 'Suppressed Ion Chromatographic Analysis of Anions in Environmental Waters Containing High Salt Concentrations', *J. Chromatogr.*, **733**, 73–91 (1996).
66. Marheni, P.R. Haddad, A.R. McTaggart, 'On-column Matrix Elimination of High Levels of Chloride and Sulfate in Non-suppressed Ion Chromatography', *J. Chromatogr.*, **546**, 221–228 (1991).
67. A.C.M. Brandao, W.W. Buchburger, E.C.V. Butler, P.A. Fagan, P.R. Haddad, 'Matrix-elimination Ion Chromatography with Post-column Reaction Detection for the Determination of Iodide in Saline Waters', *J. Chromatogr.*, **706**, 271–275 (1996).
68. S.I. Weir, E.C.V. Butler, P.R. Haddad, 'Ion Chromatography with UV Detection for the Determination of Thiosulfate and Polythionates in Saline Waters', *J. Chromatogr.*, **671**, 197–203 (1994).
69. Dionex® Application Update 122, 'Determination of Iodide in Brine', Dionex® Corporation, Sunnyvale, CA, 1998.
70. P.F. Subosa, K. Kihara, S. Rokushika, H. Hatano, T. Murayama, T. Kuboto, Y. Hanoaka, 'Ion Chromatography of Inorganic Anions in Brine Samples', *J. Chrom. Sci.*, **27**, 680–684 (1989).
71. A. Siriraks, H.M. Kingston, J.M. Riviello, 'Chelation Ion Chromatography as a Method for Trace Elemental

Analysis in Complex Environmental and Biological Samples', *Anal. Chem.*, **62**, 1185–1193 (1990).
72. M.P. Harrold, A. Siriraks, J. Riviello, 'High Performance Ion Chromatographic Separation of Uranium and Thorium in Natural Waters and Geological Materials', *J. Chromatogr.*, **602**, 119–125 (1992).
73. M.A. Tabatabai, N.T. Basta, 'Ion Chromatography' in *Soil Analysis: Modern Instrumental Techniques*, 2nd edition, ed. K.A. Smith, Marcel Dekker, New York, 1982.
74. M.A. Tabatabai, N.T. Basta, H.J. Pirela, 'Determination of Total Sulfur in Soils and Plant Materials by Ion Chromatography', *Commun. Soil. Sci. Plant Anal.*, **19**, 1701–1714 (1988).
75. S. Mou, 'Environmental Applications of Ion Chromatography in China', *J. Chromatogr.*, **546**, 317–324 (1991).
76. N.T. Basta, M.A. Tabatabai, 'Ion Chromatographic Determination of Total Metals in Soils', *Soil. Sci. Soc. Am. J.*, **54**, 1289–1297 (1990).
77. M.Y. Ye, Y. Shen, C.C. West, W.G. Lyon, 'Analysis of Ferric and Ferrous Ions in Soil Extracts by Ion Chromatography', *J. Liq. Chrom. Relat. Technol.*, **21**, 551–565 (1998).
78. I.R. Willett, 'Direct Determination of Aluminum and its Cationic Fluoro-complexes by Ion Chromatography', *Soil. Sci. Soc. Am. J.*, **53**, 1385–1391 (1989).
79. B. Gade, 'Ion Chromatographic Investigations of Leachates From a Hazardous Waste Landfill', *J. Chromatogr.*, **640**, 227–230 (1993).
80. D.A.C. Manning, A. Bewsher, 'Determination of Anions in Landfill Leachates by Ion Chromatography', *J. Chromatogr.*, **770**, 203–210 (1997).
81. K. Fischer, A. Chodura, J. Kotalik, D. Bieniek, A. Kettrup, 'Analysis of Aliphatic Carboxylic Acids and Amino Acids in Effluents of Landfills, Composting Plants and Fermentation Plants by Ion Exclusion and Ion Exchange Chromatography', *J. Chromatogr.*, **770**, 229–241 (1997).
82. R. Baziramakenga, R.R. Simard, 'Low Molecular Weight Aliphatic Acid Contents of Composted Manures', *J. Environ. Qual.*, **27**, 557–561 (1998).
83. N.T. Basta, M.A. Tabatabai, 'Determination of Total Metals in Sewage Sludges by Ion Chromatography', *J. Environ. Qual.*, **20**, 79–88 (1991).
84. A. Lopez, T. Rotunno, F. Palmisano, R. Passino, G. Tiravanti, P.G. Zambonin, 'Simultaneous Determination of Chromium(III), Aluminum(III), and Iron(II) in Tannery Sludge Acid Extracts by Reversed-phase High-performance Liquid Chromatography', *Environ. Sci. Technol.*, **25**, 1262–1266 (1991).
85. W.P. Cofino, 'Quality Assurance in Environmental Analysis', in *Environmental Analysis – Techniques, Applications and Quality Assurance*, ed. D. Barcelo, Elsevier, Amsterdam, 1993.
86. ASTM D 19.05.03.21 (draft), 'Standard Test Method for Bromate, Bromide, Chlorate, and Chlorite in Drinking Water by Chemically Suppressed Ion Chromatography', ASTM, Philadelphia, PA, 1998.

Ion-selective Electrodes in Environmental Analysis

Robert J. Forster and Tia E. Keyes*
Dublin City University, Dublin, Ireland

1 Introduction, Opportunities and Advantages 1
2 Principles 2
2.1 Background 2
2.2 Theoretical Considerations 2
2.3 Dynamic Range 3
3 Practical Considerations in Using Ion-selective Electrodes for Environmental Analysis 4
3.1 Theoretical Versus Practical Analytical Conditions 4
3.2 Activity Versus Concentration 4
3.3 Calibration 5
3.4 Reference Electrodes 5
3.5 Response Time 6
3.6 Quality Control in Ion-selective Electrode Measurements 6
4 Types of Ion-selective Electrodes 6
4.1 Solid-membrane Electrodes 6
4.2 Liquid Ion-exchange Electrode 7
4.3 Gas-sensing Membrane Electrodes 8
4.4 Solid-contact and Coated-wire Ion-selective Electrodes 9
4.5 Arrays 9
4.6 Ion-selective Field-effect Transistors 10
5 Ion-selective Electrodes in Environmental Analysis 11
5.1 Techniques: Flow Injection Analysis 11
5.2 Water-based Monitoring 13
5.3 Arrays in Water Quality Monitoring 16
5.4 Soil Monitoring 17
5.5 Gaseous Environmental Pollutants 18
6 Conclusions and Future Directions 20
Abbreviations and Acronyms 21
Related Articles 21
References 21

An ion-selective electrode (ISE) is an electroanalytical sensor that produces a potential whose magnitude is dictated by the activity of a small number of ions. The signal is generated by selective partitioning of a charged analyte across the sample–electrode interface. These sensors allow a wide range of environmental pollutants to be detected and quantified rapidly, sensitively and selectively.[1] The ideal ISE can be inserted into the sample and will display the result of the chemical analysis within a few seconds with high precision and selectivity. Often no sampling, dilution or reagent addition is required, and changes in the analyte concentration or activity can be displayed in real time. Beyond the classical glass pH electrode invented in the early 1900s, the field now includes electrodes with several selective membranes for both sensing analytes and eliminating interferences, as well as ion-selective field-effect transistors (ISFETs). These new sensing approaches have greatly extended the range of environmentally relevant molecules that can be measured, and sensors exist for environmental pollutants as diverse as atmospheric carbon dioxide and radionucleotides in highly acidic solutions.

1 INTRODUCTION, OPPORTUNITIES AND ADVANTAGES

The technological development of the past century has seen a dramatic rise in emissions of anthropogenic pollutants. As a result, environmental analysis has become an area of intense interest for scientists and the public alike. The field has evolved, with the growing concern for the welfare of the environment requiring that qualitative and quantitative information on the nature, concentration and temporal changes in pollutants be available. Environmental monitoring is often mandated by legislation that places limits on allowable pollutant levels in soil, air and water samples. Furthermore, the financial penalties of ''polluter pays'' legislation make it essential for industrialists to continuously monitor their environmental impact. The challenge for analytical scientists is to provide suitable sensors for analysis of environmental samples in a sensitive, reproducible, rapid and accurate manner. This information is then applied to risk analysis, environmental impact studies and as a guide for remedial action. As shown in Table 1, these demands have led to the development of many important analytical approaches. Sensors have a significant role to play since they can be used, often on-site, to rapidly identify the type and concentration of pollutants. The goal of the analysis can range from preliminary, qualitative, field-screening to exposure assessments, time-dependent studies, speciation, etc. that require quantitative, dynamic information across a wide range of concentrations. Organic, inorganic and biological materials must be catered for and the impact of a wide range of sample matrices and interferences on the performance of the

* Present Address: Dublin Institute of Technology, Dublin, Ireland.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Analytical methods for environmental monitoring of common pollutants

Chemical pollutant	Analytical technique
Carbon dioxide	IR spectroscopy, ISE
Carbon monoxide	IR spectroscopy
Nitrogen oxides (NO_x)	CL, ISE
Methane	IR spectroscopy, GC
Ozone	CL
Sulfur dioxide	Flame photometry, sulfonation rate by spectroscopy, ISE
Elemental	AAS, AES, XRF
Organics	UV/VIS spectroscopy, TLC with fluorescence spectroscopy, GC
F^-, Cl^-, NO_3^-, CN^-, NH_3, etc.	ISE
Metals	AAS, adsorptive stripping voltammetry, ISE

IR, infrared; CL, chemiluminescence; GC, gas chromatography; AAS, atomic absorption spectroscopy; AES, Auger electron spectroscopy; XRF, X-ray fluorescence; UV/VIS, ultraviolet/visible; TLC, thin-layer chromatography.

method assessed. Many of the pollutants detailed in Table 1 cause concern even when their concentrations are at the micromolar level. Thus, environmental analysis is often characterized by the need to measure trace levels, while demanding that the analytical technique has a high degree of selectivity, sensitivity, precision and accuracy.

Potentiometry fulfills these requirements for several species that pollute water, soil and air samples. The technique is distinguished by its versatility and simplicity, and the following list highlights some of the main advantages.

- **Portability:** the sensors and instrumentation are physically small, simple, and are often sufficiently robust to allow on-site monitoring of pollutants.
- **Sensitivity and dynamic range:** the logarithmic response of ISEs gives rise to a broad dynamic range in comparison with other analytical techniques such as spectroscopy. One reason why ISEs are especially attractive for environmental monitoring is that analytes can be determined over a wide range of concentrations using the same equipment with minimum changes in the analytical method.
- **Established technology:** the ISE is possibly the earliest example of a chemical sensor; preliminary reports of the pH electrode were made in 1906 by Cremer and subsequently in 1909 by Haber and Klemensiewicz who carried out systematic studies of the glass electrode's response.[2,3] Commercial production of the pH electrode commenced in the 1930s when Corning produced their $NaO-CaO-SiO_2$ glass electrodes. Research over the following decades improved the glass composition for pH analysis and also led to the production of other ion-selective glass electrodes. Other membrane systems have been developed in addition to glass, and ISEs are now firmly established as the most common chemical sensors.
- **Low reagent demand:** small volumes of solvents or reagents are required to perform an analysis.
- **Ease of use:** nonspecialists can be quickly trained to operate and diagnose the performance of the sensors and instrumentation. Moreover, the data can be easily processed to extract quantitative information.
- **Multianalyte capabilities:** multiple sensors can be incorporated within a single device, either to improve the reliability of the analytical information or to allow the concentration of several target analytes to be measured within a single sample.
- **Continuous measurement:** with careful design, ISEs and ISFETs can be fabricated to provide meaningful information on the second and millisecond timescales. Therefore, while their use as detectors in chromatography is not especially widespread, they can be used in flow injection analysis (FIA) to achieve high sample-throughput and automated screening of samples.

2 PRINCIPLES

2.1 Background

An ISE is a potentiometric sensor that responds to the activity of certain ions. Figure 1 shows a general scheme for the measurement process. In ISEs, a selective membrane separates the internal reference solution and the analyte. Irrespective of whether the membrane is glass, a poly(vinyl chloride) (PVC) film containing an ionophore or a single lanthanum fluoride crystal, its function is to selectively exchange ions at the solution–membrane boundary. Ion exchange results in the formation of a potential difference which is measured using a high-input impedance voltmeter. The high-input impedance means that the voltmeter draws essentially no current through the electrochemical cell and the cell potential is measured under conditions that are very close to equilibrium.

2.2 Theoretical Considerations

For an ideal ion-specific membrane, the magnitude of the potential developed is related to the activity of a single ion in solution and is given by the Nernst equation (Equation 1)

$$E = E_{\text{constant}} \pm \frac{RT}{zF} \ln[a_i] \qquad (1)$$

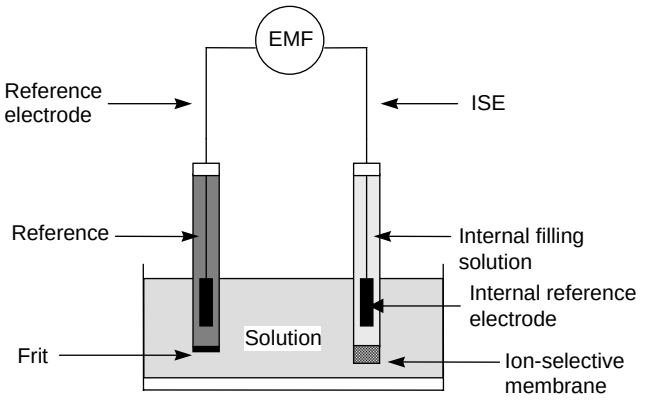

Figure 1 Typical ISE measuring system, including an ISE with a selective membrane separating an internal reference electrode from the sample solution and an external reference electrode against which the transmembrane ISE potential is measured.

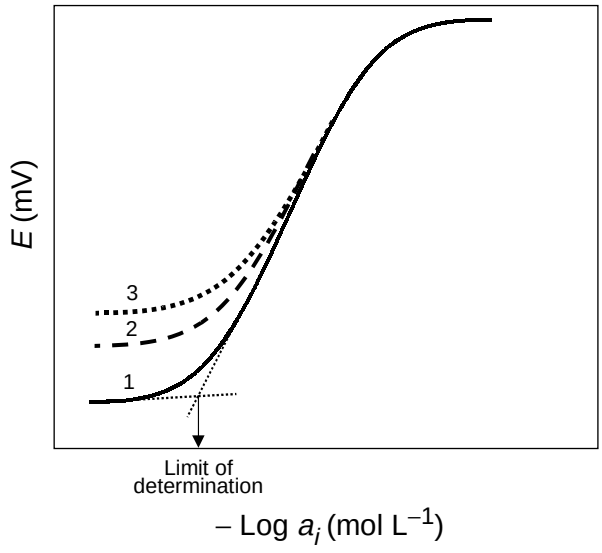

Figure 2 Typical ISE calibration curves. Curve 1 represents the situation found in the absence of interferences while curves 2 and 3 correspond to increasingly higher interference concentrations.

where E_{constant} is the cell constant, a_i is the activity of the ion in solution i, and z is the charge on the ion. For an ideal membrane, the slope would be 59.16 mV dec^{-1} for a univalent cation at 25 °C.

In reality, all electrodes are selective rather than specific and ions other than the primary ion contribute to the observed potential. These ions represent interferences to the measurement. Equation (2) describes the Nickolskii–Eisenmann equation that accounts for these interferences to the ISE response,

$$E = E_{\text{constant}} \pm \frac{RT}{z_j F} \ln \; a_{ik} + \sum K_{jkl}^{\text{pot}} a_{il}^{(z_k/z_l)} \qquad (2)$$

where a_{ik} is the activity of the primary ion in solution i, a_{il} is the activity of any interfering ion l, z_k and z_l are the charges on the primary and interfering ions, respectively, and K_{jkl}^{pot} is the selectivity coefficient of the electrode j against the interfering ion l. The empirical selectivity coefficient is unique to a particular cell and measurement conditions. The summation term of Equation (2) indicates how well the ISE can discriminate between the primary ion and interferences. The smaller the value of K_{ijl}^{pot} the more selective the membrane. An ideal membrane would have $K_{ijl}^{\text{pot}} = 0$ for all nonprimary ions and the Nickolskii–Eisenmann equation (Equation 2) would reduce to the Nernst equation (Equation 1). Typically a log scale is used to express the selectivity coefficients since ions that are dissimilar to the primary ion (principally through size and charge differences) will have selectivity coefficients close to zero, while similar ions may have K_{ijl}^{pot} close to unity. As illustrated in Figure 2, the concentration of interfering ions has a significant effect on both the linear calibration range and the limit of determination. Effects of this kind are especially important in environmental analysis, not only because low detection limits (DLs) are often essential, but because the absolute concentrations of interferences may be high.

To reduce the influence of interferences, one seeks to use ISEs that exhibit the smallest values of the selectivity coefficient and hence $K_{ijl}^{\text{pot}} a_j$. For many environmental samples, the background composition does not vary greatly from sample to sample. Thus, it is possible to use Equation (3) to estimate the maximum allowable value of the selectivity coefficient if the interferences are not to introduce error in the measured value of the primary ion activity, a_l, larger than P_{ij} percent

$$K_{ij}^{\text{pot}}(\max) = \frac{a_i(\min)}{[a_{ij}(\max)]^{n/z}} \times \frac{P_{ij}}{100} \qquad (3)$$

where $a_i(\min)$ is the lowest expected activity of the primary ion and $a_{ij}(\max)$ is the highest expected activity of the interferent. It may be possible to correct the responses of ISEs that fail to achieve the desired accuracy using numerical procedures.[4]

2.3 Dynamic Range

The dynamic or response range of an ISE defines the range of concentrations of the analyte ion to which the electrode will provide an analytically meaningful signal. Instrumental noise is intrinsic to any technique and often defines the lower limit of detection (LOD). This LOD may be described as the lowest concentration level of analyte that can be determined to be statistically different from the associated analyte blank. In practice, the LOD is the concentration of analyte which provides

a signal that is three times the standard deviation of the background noise. Typical ISEs for environmentally important analytes can be calibrated between approximately 10^{-5} and 10^{-1} M. The dilute end of this scale falls within the trace analysis range since 10^{-5} M corresponds to sub-parts per million (milligrams per kilogram) levels for simple anions and cations. It is sometimes possible to achieve even lower DLs by serial dilution of standards. However, use of ion buffers is an approach that can dramatically extend the calibration ranges of ISEs. For example, in the case of the fluoride sensor the typical lower calibration limit of 10^{-5} M can be pushed down to almost 10^{-10} M by using ion buffers containing Zr(IV), Th(IV) and La(III).[5]

Bakker has made significant contributions to improve the selectivity and sensitivity of ion-selective potentiometric and optical sensors.[6,7] In particular, Mathison and Bakker recently reported on the effect of transmembrane electrolyte diffusion on the DL of carrier-based potentiometric ISEs.[8] This work emphasizes the importance of considering the effects of both extraction and diffusion processes, e.g. the presence of lipophilic anions in the inner electrolyte solution increases the LOD owing to increased extraction into the membrane. Stirring the sample decreases the DL because it increases the mass transport rate from the membrane surface to the bulk sample. Bakker et al. highlighted the fact that many ionophores which achieve only micromolar LODs in ISEs can realize picomolar DLs in optically based approaches.[9] In membrane-based ISEs, poor DLs may arise because of primary ions leaching from the internal filling solution into the sample. Sokalski et al. used an internal electrolyte with a low activity of the primary ion and a high activity of an interfering ion.[10] This filling solution establishes a concentration gradient in the membrane causing a constant flux of the primary ion toward the inner compartment. In this way, the passage of the primary ion into the sample is prevented. Using this innovative approach, Sokalski et al. were able to push the conventional LOD for a Pb^{2+} electrode from 4×10^{-6} M to 5×10^{-12} M!

3 PRACTICAL CONSIDERATIONS IN USING ION-SELECTIVE ELECTRODES FOR ENVIRONMENTAL ANALYSIS

3.1 Theoretical Versus Practical Analytical Conditions

While the conditions under which Equation (1) is obeyed are well defined, they are not necessarily fulfilled for practical analysis, which may cause measurement errors. For example, the Nernst equation is based on thermodynamic principles and is strictly applicable only to ideally reversible systems. However, potentiometric measurements are usually carried out at relatively short timescales, typically tens of seconds, and the Nernst (Nickolskii–Eisenmann) equation may be incorrectly applied to systems that are electrochemically irreversible over long periods. Moreover, besides the desirable ion-exchange reaction of the primary ion, there are several other processes that can contribute to the observed potential change, including redox, adsorption and membrane dissolution processes. The Nernst equation is not strictly applicable under these circumstances since it is valid only when one well-defined process occurs at an electrode. These secondary processes limit the range over which there is a linear dependence between potential and log activity.

3.2 Activity Versus Concentration

As discussed, the Nernst equation underpins the operation of ISEs. Therefore, it is the activity of ions in solutions that is actually measured. Measuring activity rather than concentration is advantageous for speciation purposes and for the study of acid–base, complexation and precipitation equilibria. However, it represents a significant disadvantage when the objective is to measure the total content, i.e. both free and bound forms, of an analyte in a sample. As discussed later, where the total content is important, the analyte must be released from the matrix by chemical pretreatment or distillation.

Determining the true activity of an ion in solution places considerable demands on the analytical procedures employed and the manner in which the samples are handled. In particular, the following conditions must be met.

(1) Chemical pretreatments, e.g. reagent addition, evaporation, dilution or freeze-drying must be avoided.

(2) The relationship between the signal and the ion activity must be known, i.e. the response function of the ISE under identical operating conditions must be determined so as to calibrate the response.

(3) The ISE should be specific, i.e. broad-band selectivity is not adequate, since the Nernst equation applies only to electrochemical systems involving a single well-defined process. The sample cannot be chemically modified to mask interferences.

Therefore, in the context of environmental monitoring, samples should ideally be analyzed in situ or directly read at a later stage without the addition of further reagents. This condition is often difficult or impossible to achieve because analytes that are of environmental interest must be preconcentrated, treated or digested prior to analysis because of low concentrations or matrix effects.

3.3 Calibration

The selectivity coefficient should equal the thermodynamic exchange constant for the transfer of an ion from solution into the selective membrane of the ISE under equilibrium conditions. Therefore, K_{ij}^{pot} is sensitive to both the selective element within the membrane, e.g. the host within a host–guest complex, and to the composition of the membrane. For example, for polymer-membrane-based ISEs, the solvent, plasticizer, polymer, and their respective concentrations, will all influence the transmembrane transport processes and hence the selectivity coefficients. Thus, there are several possible difficulties in both the calibration of individual sensors and comparing calibration data between separate devices.

First, changing ionic strength or variable liquid junction potentials may mean that the standard cell potential E^{o} is not constant over the entire concentration range. This shift will cause nonlinear calibrations or non-Nernstian slopes to be observed. Moreover, an incorrect membrane composition or redox-active impurities may mean that the characteristic electrode slope is not given by RT/zF. Processes such as slow partitioning of the analyte across the membrane/solution boundary, slow diffusion within the membrane or adsorption/desorption cause many ISEs to be slow to reach equilibrium. Failure to achieve global equilibrium means that the apparent selectivity coefficients become time dependent. This ''kinetic selectivity''[11] has been exploited in high-throughput FIA to improve the overall analytical performance, since the short timescale response is often more sensitive to the primary ion concentration than to interferences.[12]

Second, the composition of the membrane may change significantly over time. For example, leaching of the ionophore can be a particular problem for liquid-state electrodes where the ionophores and plasticizer are typically only physically entrapped within the matrix and tend to dissolve into the samples. Also, partitioning of water and lipophilic sample components into the ISE element will decrease the ionophore concentration within the membrane. The most significant effect of reducing the ionophore loading is not to decrease the sensor's sensitivity, but to degrade its ability to discriminate between ions. In the most extreme situation, the sensor loses the selectivity achieved by the incorporated ionophore and the sensor response will ultimately follow the Hofmeister or lyotropic series, e.g. for anions $SO_4^{2-} > CF_3COO^- > Br^- > ClO_4^- \approx NO_3^-$, and for cations $Na^+ > Ca^{2+} \approx K^+ > NH_4^+ > Li^+$.

Third, the electrode geometry and measuring conditions must be considered. For example, macroscopic and microscopic ISEs constructed using identical materials and techniques can exhibit differences in the observed potential of 10 mV or more. These differences may arise because the membrane properties depend on the electrode's size. However, there is some evidence that, as found for amperometric microsensors,[13,14] changes in the mass transport mode with size may influence the rate of ion exchange at the membrane/solution interface. These effects may give rise to an electrokinetic potential causing the electrode slope observed at a microelectrode to decrease relative to that observed for a macroelectrode.

Fourth, the background composition of the solutions must be constant if measurements or devices are to be compared. Here, our focus is on the possibility that the composition of the solution at the membrane/solution interface differs from that of bulk. These differences may arise simply because of the ion-exchange reaction itself[15] or because solvent or other components partition into, or out of, the sensing element.

3.4 Reference Electrodes

As illustrated in Figure 1, a reference electrode is an integral part of ISE measurements since it is against this potential that the ISE signal is measured. If operated incorrectly, then the reference electrode can act as a source of significant error. Temperature is an important variable in many environmental monitoring situations, especially for in situ water quality monitoring since the temperature can change dramatically with depth. However, the reference electrode potential is temperature dependent, e.g. for the traditional calomel reference electrode the reference potential shifts by approximately $0.7\,\text{mV}\,\text{K}^{-1}$. In real terms, this shift translates into a few percent relative error on measurements performed a few degrees apart. Also, as described by the Nernst equation (Equation 1), the ISE potential depends on the temperature. Therefore, it is important that the analysis and calibration temperatures are closely matched if accurate results are to be obtained. An alternative is to use a reference electrode, such as the $Zn/ZnCl_2$ system, whose potential is very insensitive to temperature.

It is also highly desirable to make the electrochemical cell as symmetrical as possible to minimize the effect of changes in temperature on the observed potential. This objective means not only that the same internal and external reference electrodes are used, but that the composition of the inner filling solution of the ISE matches the sample composition as closely as possible. Thus, in contrast to most reference electrode filling solutions where 6 M or even saturated solutions are commonly used, ISEs are often filled with a 0.1 M solution of the primary ion. This lower concentration within the internal filling solution also helps prevent contamination of the samples by leakage, e.g. in the case of chloride determination.

3.5 Response Time

Under special conditions, ISEs can respond rapidly to abrupt changes in the concentration of their primary ion around the midpoint of their calibration range, achieving 90% of the total potential change within 100 ms. However, typical ISEs respond much more slowly with equilibrium responses only being obtained after seconds or even tens of seconds.

3.6 Quality Control in Ion-selective Electrode Measurements

Quality control is an important issue when using ISEs for environmental analysis, to ensure accurate and reproducible results. A typical quality control system involves periodic confirmation of the ISE calibration by determining the concentration of a known, independently prepared sample. For example, immediately after an ISE is calibrated, an initial calibration verification (ICV) standard should be analyzed. The ICV contains a known concentration of the primary ion close to the midpoint of the calibration range and is prepared independently of the calibration standards. For acceptable performance, the ICV recovery must be between 90% and 110% otherwise the source of error must be identified and corrected. This verification standard is typically analyzed after every ten samples. If the ISE calibration has changed, all samples analyzed since the last acceptable continuing calibration verification (CCV) standard, must be re-analyzed. After each ICV and CCV sample analyzed, a reagent blank is often analyzed. This blank is frequently high-purity water but, if reagents are being added to the samples, then a 1 : 1 dilution of the buffer is often used. The objective in measuring the blank is to identify possible contamination of the samples. Spiking the samples is another common approach to ensure that quality measurements are performed. The spike concentration must be ten times the DL and the volume added must be negligible (less than or equal to one-thousandth of the sample volume). Spike recovery must be 75–125%, otherwise samples must be analyzed by performing standard additions.

4 TYPES OF ION-SELECTIVE ELECTRODES

All ISEs rely on a membrane to separate the analyte and reference solutions. Therefore, ISEs may be conveniently classified according to the nature of this membrane material. While the membranes may absorb solvent, ideally they are as immiscible as possible with their analyte solution. Hydrophobic organic liquid ion-exchange electrolytes, solid polyelectrolyte ion exchangers and solid ion-conducting electrolytes are the most frequently employed electrode membranes types. The best membranes are ionically conducting, for example solid or liquid electrolytes that contain substances which ionize depending on the membrane dielectric constant and extent of solvent penetration.

4.1 Solid-membrane Electrodes

The success of the pH glass electrode is without parallel in sensors. Although somewhat mechanically fragile, it resists a wide variety of sample media and, with the exception of hydroxide, is largely free from interferences. Moreover, pH-sensitive glass electrodes form the basis of many successful sensors for environmentally important gases including carbon dioxide, sulfur dioxide and nitric oxide. Thus, glass membranes represent an important class of solid-membrane ISEs. As illustrated in Figure 3, these ISEs have a thin glass membrane fused to the end of a glass or plastic body. The main body of the electrode contains an internal reference electrode, typically Ag/AgCl, and is filled with a solution that is usually the chloride salt of the primary ion. The membranes are primarily made from lithia, aluminosilicate or multicomponent glasses whose selectivity pattern depends on the composition of the glass. As shown in Table 2, excellent discrimination against common cationic species is achieved. There are no anion-selective glass membrane materials and glass membranes are primarily used for sensing of monovalent cations.

As illustrated in Figure 4, homogeneous membrane electrodes contain insoluble crystalline materials either

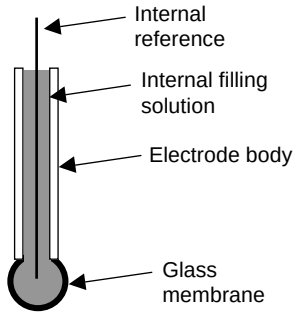

Figure 3 Schematic diagram of a typical glass electrode. The internal reference electrode is usually Ag/AgCl.

Table 2 Selectivity coefficients of some glass-membrane electrodes

Membrane	Log selectivity coefficients				
	H^+	Na^+	K^+	NH_4^+	Ag^+
NAS 11-18	10^3	1	10^{-3}	3×10^{-5}	400
NAS 27-04		0.1	1	0.3	

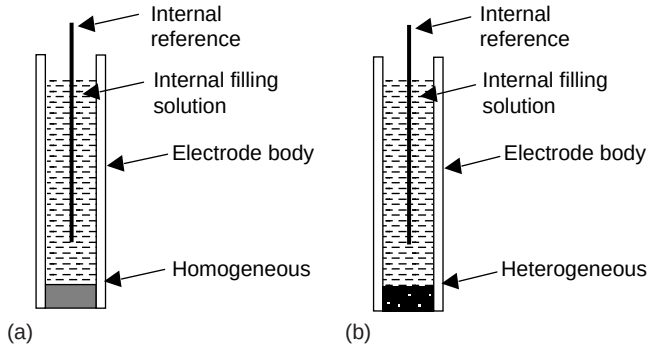

Figure 4 Diagram of (a) a solid homogeneous membrane electrode and (b) an electrode with a solid heterogeneous membrane.

as single crystals or as compressed powder pellets. An important example of an electrode of this type is the fluoride electrode. This electrode consists of a membrane of a single crystal of lanthanum fluoride doped with europium(II) fluoride to improve conductivity. The internal filling solution consists of 0.1 M NaF and NaCl, with an internal reference electrode of silver–silver chloride. This membrane only conducts fluoride but hydroxides may interfere with the measurement. Other homogenous membrane electrodes that have found significant use in environmental analysis include sensors for pollutants such as cadmium, cyanide and lead.

In contrast, heterogeneous membrane electrodes comprise an active crystalline material dispersed through another matrix, for example a polymer that improves the mechanical robustness of the sensor. However, plastics are electrical insulators and to achieve sufficient electrical conductivity the active ingredient must be present in high concentration. There are not many materials which can be easily prepared as single crystals and this approach to sensor design has extended the number of target analytes. Beyond dispersion in a polymeric matrix, pressed polycrystalline materials containing two or more inorganic salts, e.g. silver sulfide and iodide plus a binder to form an iodide selective ISE, are popular. Environmental analysis characteristically places heavy demands on the stability and durability of an electrode material, and, as shown in Table 3, many useful sensors have been developed using these approaches.

4.2 Liquid Ion-exchange Electrode

As illustrated in Figure 5, the traditional membrane of a liquid ion-exchange electrode consists of a homogeneous, organic, water-immiscible liquid absorbed into a millipore or nucleopore filter that incorporates mobile, ionic components. The mobile ionophores include compounds such as hydrophobic acids, bases or salts. Liquid membranes typically show Nernstian responses in the absence of interferences, usually for oppositely charged counterions to their own electrical charge. This design is more flexible than the solid-membrane materials considered in section 4.1 and ISEs sensitive to a much wider range of compounds can be fabricated using this arrangement, including K^+, Ca^{2+} and the environmentally important NO_3^- ions.[16] However, these ISEs have been largely superseded by other approaches, notably neutral carrier sensors, because of technical problems.

Table 3 Applications of solid-membrane ISEs to environmental analysis

Electrode type	DLs (M)	Interference	Environmental applications
Fluoride: Single crystal LaF_3	0.02 ppm	OH^-, Fe(III), Al(III)	Water quality, rain, air and vegetation. Drinking water fluoridation, minerals, soil, etching solutions containing H_2F_2
Chloride: AgCl, $AgCl/Ag_2S$ and Hg_2Cl_2/HgS	25 ppb	Br^-, I^-, CN^-, S^{2-}, Br^-, $S_2O_3^{2-}$, SCN^-	Water quality, seawater, treated sewage, wastewater, soil extracts
Bromide: $AgBr/Ag_2S$	50 ppb	OH^-, Cl^-, I^-, CN^-	Water quality
Cyanide: AgI/Ag_2S	5 ppb	S^{2-}, I^-	Industrial wastewater, galvanic baths, mineworks
Iodide: AgI and AgI/Ag_2S	12 ppb	OH^-, Cl^-, Br^-, CN^-	Water quality
Sulfide: Ag_2S	10 ppb	CN^-, I^-, CrO^{2-}, IO^{3-}	Industrial effluent, sewage, water quality, food and drinks
Cadmium: CdS/Ag_2S	0.1 ppm	Hg^{2+}, Ag^+, Cu^{2+}, Pb^{2+}, S^{2-}, Cl^-, I^-	Water quality, industrial effluents
Lead: PbS/Ag_2S	0.2 ppm	Hg^{2+}, Ag^+, Cu^{2+}, Fe^{3+}	Water and air quality, soil analysis, vehicular emissions
Copper: CuSe single crystal or CuS/Ag_2S	0.05 ppm	Hg^{2+}, Ag^+, Fe^{3+}	Industrial effluent, water quality

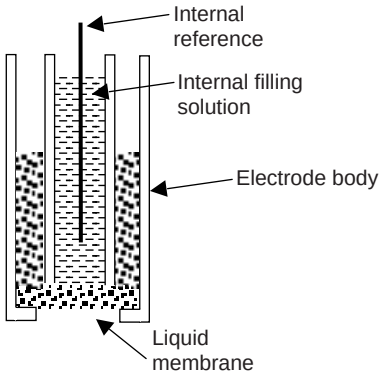

Figure 5 Diagram of a classical liquid-membrane electrode.

These difficulties include imperfect filling of the support membrane pores by the ionophoric liquid, leading to sample permeation into the internal filling solution that ''short circuits'' and fails to generate a measurable transmembrane potential. Also, the liquid membranes themselves tend to be somewhat unstable, becoming contaminated by lipophilic substances within the samples or mechanically dislodged from the support if mishandled.

Neutral-carrier ISEs are typically based on PVC in which a suitable electrically neutral ionophore, plasticizer and ion-exchanger are co-immobilized. The ionophores typically possess a polar cavity that contains electron-donating atoms or polar groups capable of interacting with the target ions. The stability of the complex, and hence the selectivity coefficient, is determined by a ''best-fit'' mechanism and the neutral carrier is more selective towards ions that best fit the molecular cavity. Larger ions will not be able to effectively penetrate the cavity, while smaller ions will interact with too few donor atoms to form stable complexes. The strength of binding must be carefully balanced, since while strong binding favors specificity, forming too strong a complex will lead to irreversible responses. The types of molecules used in neutral-carrier ISEs continues to expand, and include crown ethers,[17] calixarenes,[18,19] complex macromolecules, and antibiotics such as valinomycin[20] and nonactin.[21] Neutral-carrier ISEs are commonly used for ions such as Na^+, K^+, Li^+, NH_4^+, Ca^{2+}, Cs^+ and Ba^{2+}.

4.3 Gas-sensing Membrane Electrodes

Gas sensing is an extremely important aspect of environmental monitoring in areas ranging from smokestack emissions, notably for sulfur and nitrogen oxides, to carbon dioxide levels in ambient air. The earliest and most commercially successful gas-sensing probe is based on the Severinghaus electrode,[22] in which a glass pH-electrode is placed behind a gas-permeable membrane. Table 4 shows common gas-sensing electrodes that are important in environmental analysis. These electrodes differ from those described above in two important ways. First, they do not directly detect the target analyte. They are not true ISEs, but are complete electrochemical cells capable of indirectly determining the concentration of the target analytes using a conventional ISE that is internal to the cell. A popular approach to gas sensing is to analyze the by-products of the target chemical's reaction with the internal electrolyte, most commonly by monitoring pH changes in the cell. Second, because of the mode of operation, gas-sensing electrodes employ two membranes; the first allows the gas of interest to permeate into the electrochemical cell, and the second selectively generates the ISE potential.

Figure 6 shows a schematic of a typical gas-sensing probe. The probe comprises an outer gas-permeable membrane which prevents nongaseous interferences from accessing the inner cell. Once dissolved in the inner-cell filling solution, the analyte establishes a chemical equilibrium with the electrolyte and the products are then detected by the internal ISE. Gas-sensing probes are often highly selective since many interferences are removed by the gas-permeable outer membrane and the analyzing ISE is typically a highly selective pH electrode. The main disadvantages of such sensors are their poor long-term stability and their long response times, associated with the slow rate of gas diffusion across the membrane.

The second type of gas-sensing ISE is the air-gap electrode. In this arrangement, the internal filling solution surrounds the internal ISE's surface as a thin film, and an air bubble replaces the outer gas-permeable membrane.

Table 4 Environmentally important gases that can be determined using an ISE

Analyte	Internal filling solution	Equilibrium reaction	Internal ISE
CO_2	$NaHCO_3$	$CO_2 + H_2O \longrightarrow H^+ + HCO_3^-$	pH
NH_3	NH_4Cl	$NH_3 + H_2O \longrightarrow NH_4^+ + OH^-$	pH or NH_4^+
NO_x	$NaNO_2$	$2NO_2 + H_2O \longrightarrow 2H^+ + NO_3^- + NO_2^-$	pH
SO_2	$NaHSO_3$ or $K_2S_2O_3$	$SO_2 + H_2O \longrightarrow H^+ + HSO_3^-$	pH
H_2S	Citrate buffer	$H_2S \longrightarrow H^+ + HS^-$	S^{2-}
HF	1 M H^+	$HF \longrightarrow H^+ + F^-$	F^-

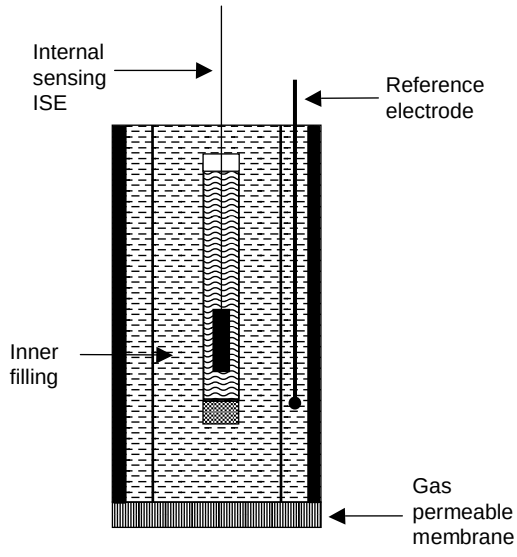

Figure 6 Schematic diagram of a classical Severinghaus-type gas sensor.

This arrangement offers some advantages over the Severinghaus-type electrode in terms of sensitivity and response time. However, the electrode surface must not be allowed to dip into the analyte solution. The electrode is operated by passing the analyte close to the electrode surface, either in the form of a gas stream or, for solution samples, by stirring the sample solution rapidly in the presence of the electrode. The air-gap cyanide sensor is an example of such an arrangement and has important applications in atmospheric analysis.

4.4 Solid-contact and Coated-wire Ion-selective Electrodes

Solid-state electrodes of the types illustrated in Figures 7 and 8 that do not contain free-flowing liquids represent an important and growing area of ISE research. Unlike conventional solution filled sensors, their response does not depend on orientation, they can be more easily

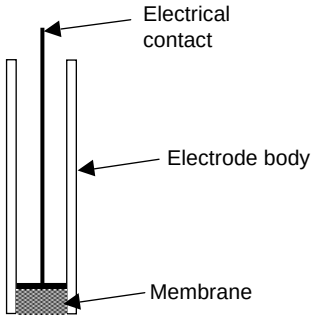

Figure 7 Solid-contact electrode without any internal filling solution.

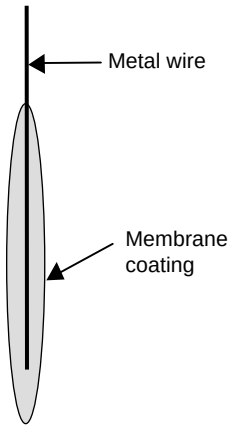

Figure 8 Coated-wire electrode.

miniaturized and they are mechanically flexible. As illustrated in Figure 8, coated-wire electrodes (CWEs) are formed by depositing a thin layer of material, often a polymer, onto the surface of an inert metal wire such as platinum or gold. CWEs have the advantages of rapid response, small size and suitability for temperature and noise compensation and for the manufacture of arrays. The difficulty with these sensors is that the potential that develops at the layer/wire interface can be unstable over long periods.

4.5 Arrays

Using a single sensor to determine the concentration of an analyte demands that it be extremely selective if its response is to be overwhelmingly dominated by a single ion. Combining several individual sensors within an array is attractive since it allows multicomponent analyses to be performed and provides the high-quality analytical information necessary to model the role of interferences. Accurate modeling of the array response may allow partially selective sensors to be used.[23] Moreover redundancy, i.e. having more than one sensor within the array that is capable of performing the analysis, leads to more reliable sensing systems. Beyond these issues, considerable attention is being paid to developing ''smart'' sensors with much improved abilities to process analytical information. A smart sensor exhibits the following characteristics:

- able to process information for automatic calibration or compensation of baseline drift
- can communicate with other sensors or devices
- adapts to changes in its environment, e.g. automatic temperature compensation
- performs self-diagnosis of performance.

With the exception of true liquid-membrane electrodes, all of the ISEs described can formally be integrated within a sensor array. However, those that employ fabrication procedures that can produce planar membranes, e.g. solid-state and neutral-carrier membranes, are the most amenable to producing arrays. Another important consideration in selecting ISEs for arrays is the stability of the calibration. Stability, rather than selectivity, is perhaps the single most important issue in sensor arrays. For example, a factorial design is preferred for calibration purposes requiring L^{f-1} solutions, where L is the number of levels (e.g. high and low concentrations) and f is the number of factors (e.g. four ions). Thus, to calibrate a four-electrode array using a two-level, four-factor design requires 32 solutions. Clearly, the sensor output must be stable during this calibration and subsequent measuring period if an accurate model is to be constructed. Highly selective sensors are not necessarily desirable since an artificially high concentration of the interferences must be used to sufficiently perturb the sensor response.

4.6 Ion-selective Field-effect Transistors

ISFETs[24–26] represent a hybrid of technologies from solid-state physics and electrochemistry. As shown in Figure 9, ISFETs employ conventional ion-selective membranes that either replace the gate of a metal-oxide field-effect transistor (MOSFET) or are placed in direct contact with the gate. The intense interest in these devices during the late 1970s and 1980s has decreased somewhat recently because of the problems in establishing a stable internal reference potential where the PVC membrane comes in contact with the field-effect transistor (FET) surface. Also, solid-state sensors tend to suffer from poor sensor-to-sensor reproducibility even within a single batch of devices, requiring that sensors be calibrated individually. However, better fabrication technologies, materials and visualization techniques promise to solve many of these problems.

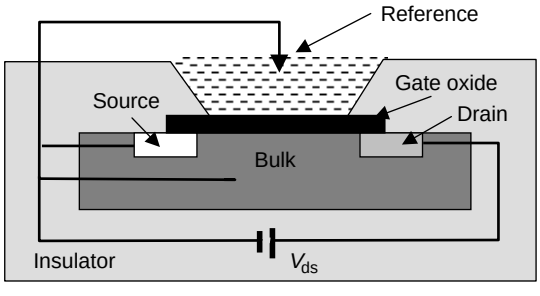

Figure 9 Typical ISFET.

The main advantages of ISFETs are their small size; the fact that they are compatible with planar silicon-processing technology, allowing them to be cheaply mass-produced; the ease with which they can be incorporated into arrays; the rapidity with which they respond to changes in analyte concentration; and their low impedance which makes them much less sensitive to electromagnetic interference. Depending on the materials used in their construction, light exposure may need to be controlled to obtain accurate readings.

Similar to a metal-oxide semiconductor FET, in an ISFET electric current flows from the source to the drain via a channel. The channel resistance depends on the electric field perpendicular to the direction of the current and on the potential difference over the gate oxide. Therefore, the source–drain current, i_{ds}, is influenced by the potential at the oxide/aqueous solution interface. At a fixed source–drain potential (V_{ds}), changes in the gate potential can be compensated by modulation of the gate–source potential, V_{gs}. This adjustment should be carried out in such a way that the changes in V_{gs} applied to the reference electrode are exactly opposite to the changes in the gate-oxide potential. This is automatically performed by the ISFET feedback amplifier that allows one to obtain constant source–drain current. In this case, the gate–source potential is determined by the surface potential at the insulator/electrolyte interface.

When SiO_2 is used as the insulator, the chemical nature of the interface oxide is reflected in the measured source–drain current. The surface of the gate oxide contains OH^- functionalities, which will be in electrochemical equilibrium with ions in the sample solutions (H_3O^+ and OH^-). The hydroxyl groups at the gate-oxide surface can be protonated and deprotonated and thus, when the gate oxide contacts an aqueous solution, a change of pH will change the SiO_2 surface potential. Typical pH sensitivities measured with SiO_2 ISFETs are 37–40 mV per pH unit rather than the 59 mV per pH unit predicted by the Nernst equation.[27] However, significant improvements on this performance have recently been achieved by using pH-ISFETs based on silicon nitride and tantalum oxide gates. In the Si_3N_4 system, silanol and amine sites make the assembly sensitive to the solution pH, while in the Ta_2O_5 ISFETs, Ta–O sites provide the pH sensitivity and close to ideal Nernstian behavior is observed.[28]

The selectivity and chemical sensitivity of the ISFET are completely controlled by the properties of the electrolyte/insulator interface. Therefore, by immobilizing a suitable membrane on the gate oxide, solid-state sensors for a range of environmentally important analytes can be fabricated.

5 ION-SELECTIVE ELECTRODES IN ENVIRONMENTAL ANALYSIS

The application of ISEs to environmental analysis can be divided into two different approaches. First, solution-phase analysis, e.g. water and effluent analysis. Second, gas-phase analysis, e.g. smokestack emissions and airborne pollutants. Many environmental monitoring applications require dynamic information about changing sample compositions within relatively short periods of time. One useful approach to achieving a high sample-throughput is FIA and we first review its mode of operation.

5.1 Techniques: Flow Injection Analysis

Although their responses are often too slow for use as end-column detectors in chromatography, the selectivity of ISEs can be exploited to achieve a high sample-throughput using FIA. Continuous methods in environmental analysis offer significant advantages over more traditional batch methods. Most importantly, it provides a means of observing temporal changes in environmentally important agents and may offer advance warning of impending pollution.

As described in the seminal work of Ruzicka,[29] in the flow injection method a carrier stream is continuously pumped around an analytical manifold and reproducible volumes of sample are injected into the stream at selected intervals.[30] After passage through reaction structures in the analytical assembly, the product or unreacted sample is passed to the detector, producing a signal whose magnitude depends directly on the concentration of the analyte. Usually the carrier contains a low concentration of the analyte ion in order to stabilize the ISE membrane potential. As illustrated in Figure 10, since the sample plug contains a higher concentration of the analyte, the analytical signal is an unsymmetrical peak whose shape reflects the effects of hydrodynamics and the kinetics of ion uptake and release.

As illustrated in Figure 11, the ISE can be integrated into the analytical assembly in one of four ways, namely by a flow-through, flow-past, tangential or wall-jet regime. In the flow-through and flow-past approaches, the electrode membrane replaces part of the tubing through which the analyte stream is travelling. In the tangential approach, the sample and membrane are at an acute angle of less than 90°, while in the wall-jet method the sample stream approaches the membrane at 90° and is then dispersed at the sides.

Operating ISEs in flowing streams offers a number of advantages over a batch-type approach. The Nernst diffusion layer that exists close to the membrane/solution interface is highly reproducible and stable, giving improved response times and very reproducible signals. Contamination problems arising from leaking reference electrodes are eliminated since the reference can be placed downstream of the detection point. The carrier solution can be easily modified so as to optimize the selective response of the ISE, e.g. controlling the ionic strength, masking interferences, etc. FIA offers a high sample-throughput, e.g. the flow rate is typically 0.5–1.0 ml min^{-1} and a 100 µl sample stays in contact with the membrane for only 6 s. Provided that the return to baseline is rapid, throughputs of 120 samples per hour or even higher may be achieved. Table 5 contains some examples where ISEs coupled to FIA have been exploited in environmental analysis.

FIA/ISE systems have been developed for use without a requirement for a preceding separation step. For example, copper has been determined in tap water

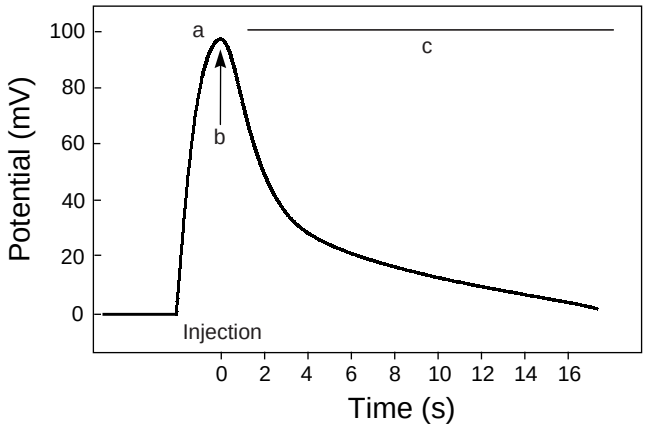

Figure 10 Typical FIA/ISE peak. The injection is made when the membrane is in equilibrium with the carrier solution. During the period labeled (a), the ISE potential increases rapidly owing to the boundary potential. A pseudo-steady-state situation is reached at the top of the peak (b). As the plug leaves the detector (c) and the membrane-carrier equilibrium is re-established, the potential returns to the baseline.

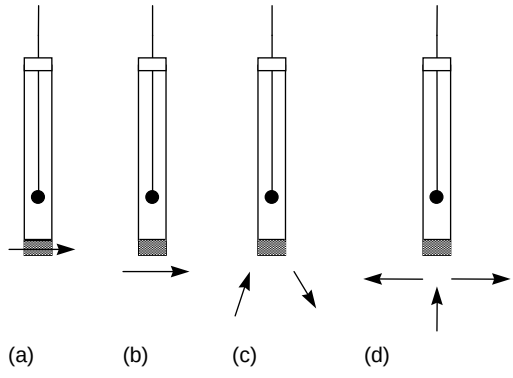

Figure 11 Different flow regimes used in FIA/ISE systems: (a) flow-through, (b) flow-past, (c) tangential and (d) wall-jet assemblies.

Table 5 FIA with ISE detection for environmental analysis

Analyte	Matrix	Electrode/membrane	Cell	Other features	Refs.
K^+	Soil	Valinomycin/di(iso-octyl)fatty acid ester/potassium tetra(4-chlorophenyl) borate/PVC	Tubular	Addition of the main ion to the carrier stream	31
Ca^{2+}	Tap water	(−)-(*R*,*R*)-*N*,*N* bis[11-(ethoxy-carbonyl)undecyl]-*N*,*N*′-4,5-tetramethyl-3,6-dioxaoctane-diamide/diethyl *N*,*N*′-[(4*R*,5*R*)-4,5-dimethyl-1,8-dioxo-3,6-dioxaoctamethylene]bis(12-methylaminododecanoate)(ETH 1001)			32
Ca^{2+}	Tap, river, and lake water	Calcium bis{bis-[4-(1,1,3,3-tetramethyl-butyl)phenyl]phosphate}/PVC	Tubular	Surfactants present	33
NO_3^-	Seawater	Commercial		In parallel with Na^+, K^+, Ca^{2+}	34
NO_3^-	Soil solution	Valinomycin/ETH 1001/tetradodecyl ammonium nitrate	Miniature	Multi-ion sensor cell	35
		Mobile carrier/PVC	Tubular	No inner reference solution	36
		Commercial	Tubular	Process control	37
CN^-	Industrial	CN^- sensor	Variable volume		38
S_2^{2-}	Natural and wastewater	AgS/Ag-based conductive epoxy glue	Tubular		39
	Sewage effluent	Commercial		Peroxide inhibition of interferences	40
	Effluents	Ag_2S	Tubular	Favorable comparison with commercial cascade electrode	41
F^-	Rainwater	Commercial		Limitation of linear response in FIA	42
Cl^-	Soil solution	Ag/AgCl	Miniature	Computer-controlled data acquisition	35
	Natural and wastewater	Coated solid-state Ag/AgCl	Tubular	High sample-throughput	43, 44
Br^-	Soil solution	Coated solid state sensor	Tubular	Primary ion added to carrier stream	45

using a detection system comprising two Cu^{2+} selective electrodes as indicator and reference.[46] In common with most FIA/ISE analyses, a relatively high concentration of electrolyte (0.1 M KNO_3 in this case) was added to the carrier stream to ensure a steady baseline. The results obtained compare well with those observed using a standard spectrophotometric approach. Moreover, as highlighted in Table 5, the primary ion is frequently added to the carrier stream, e.g. 5000 ppm cyanide was added to the carrier stream during the determination of cyanide in wastewater.[47]

ISEs have also been used as detectors in FIA following a separation step to ensure optimum sensor performance by efficiently eliminating interferences. Ion-exchange columns have also been incorporated with the twin objectives of improving sensitivity (pre-concentration step) and selectivity (separation step). For example, a porous glass column coated with 8-quinolinol has been used to preconcentrate copper.[48] Interferences from trace metal can be removed on-line using chelating ion-exchangers. This approach achieves DLs that are between 10^{-7} and 10^{-8} M.

Flow injection also lends itself to multianalyte determinations by incorporating as many ISEs as there are analytes into the detection manifold. The sensors are often placed in series. However, careful attention must be paid to achieving minimum dispersion and dead volume, i.e. high fluid velocities and well-defined sample plugs across each sensor should be employed. Multi-ion analyzers have been constructed to monitor potassium, calcium, nitrate and chloride in soil extracts, and to determine sodium, potassium, calcium, bicarbonate, nitrate and chloride in drinking water.[35]

While FIA certainly has the great advantages of high sample-throughput and highly reproducible sample handling, there are some difficulties in using ISEs as

detectors in FIA. First, as discussed in section 2.2 dealing with theoretical issues, the analyte often does not reach global equilibrium with the sensing element even for slow flow rates or stopped flow approaches. Since kinetic selectivity[49] often depends on the analyte concentration it causes plots of potential vs logarithm of analyte activity to become nonlinear. Therefore, slow transport or other kinetic difficulties can be a source of calibration errors because the rate at which equilibrium is achieved differs between standards and samples. Second, when the dilute electrolyte carrier solution moves along the ISE surface, excess charges outside the double layer are carried along by the flow, causing a streaming potential. The magnitude of the streaming potential depends on the ionic strength of the carrier, the solvent viscosity and the flow rate. It can be reduced by either stopping the flow to make the measurement or by using a carrier of high ionic strength. Another important technical consideration is contamination of the electrodes by components within the samples, e.g. highly lipophilic species in the case of neutral carrier ISEs. However, the short residence time of the sample compared with batch-type measurements usually means that this is a minor problem.

5.2 Water-based Monitoring

The late twentieth century has seen a dramatic increase in the consumption of water for domestic, agricultural and industrial purposes. As water is recycled, it is important to confirm that the water quality is acceptable and that it will not contaminate other stocks, e.g. reservoirs, storage tanks, rivers, lakes, etc. Historically, water analysis often involved simply checking the pH and biological oxygen demand, with the total salt concentration being estimated from the solution conductivity. However, this nonspecific information is becoming increasingly inadequate as limits are set for acceptable discharge levels for a wide variety of analytes ranging from heavy metals to anions. Potentiometry plays an important role in providing direct information about the concentrations of many important pollutants in potable and naturally occurring bodies of water.

5.2.1 Metallic Species

The ability to determine in situ the distribution of metal species using a simple, rapid technique is extremely valuable since the toxicity, transport and ultimate fate of the metallic species depends not only on the identity of the species in question but also on their oxidation state. Current environmental protection methods, e.g. those that have been evaluated and approved for use by the United States Environmental Protection Agency (USEPA),[50] focus heavily on the use of AAS for the determination of metals in the environment. However, AAS destroys the sample, requires a separate protocol for each metal, cannot differentiate valence states or species, is expensive and time-consuming and has difficulty in simultaneously determining the concentration of several metals. Therefore, there is growing interest in using ISEs for the environmental monitoring of metallic species.

5.2.1.1 Mercury Classical mercury ISEs exhibit linear calibration ranges in the 10^{-5}–10^{-7} M range. However, the unbound Hg^{2+} concentration is often in the 10^{-9} M range in natural environments. Therefore, highly sensitive sensors that are almost completely mercury specific, or whose responses can be corrected for the effects of interferences, are required. Epoxy-based ISEs have been developed that are capable of determining Hg^{2+} concentrations at the 10^{-14} M level, making them useful tools for the determination of both free and inorganic-bound mercury in natural waters.[51] However, Fe^{3+} represents a significant interference, not only because of poor selectivity, but also because Fe^{3+} can be present at high concentrations in environmental samples.

Host molecules capable of selectively binding heavy metals have been immobilized within PVC membranes to form polymeric ISEs. For example, calixarenes have been synthesized that incorporate soft donor atoms, such as nitrogen and sulfur, within the ion-binding cavity to create hosts for ISEs that are sensitive to Hg(II), Ag(I), Cu(II) and Pb(II).[52] A combination of the correct number and location of soft donor atoms makes the calixarenes selective for heavy metals rather than the alkali cations that are usually the primary ions for calixarene-based ISEs.

5.2.1.2 Lead The lead electrode is useful for monitoring aqueous concentrations of Pb^{2+}. Provided that the sample pH is between approximately 4 and 7, the linear range typically extends from 10^{-1} to 10^{-6} M with a DL of 10^{-7} M being observed. Under basic conditions, the LOD is restricted because of the formation of metal hydroxides. At lower pH values, the lead and silver sulfides used to form the sensing element become more soluble causing the LOD to increase. The ability of the ISE to discriminate against alkali and alkaline earth metals is good, but Cu^{2+} and Fe^{3+} represent significant interferences. High concentrations of chloride may also cause difficulties in accurately monitoring lead levels. In practical use, the surface of the electrode tends to become oxidized, and mechanical or electrochemical repolishing is required periodically. The lead electrode has also been quite widely used for the indirect determination of sulfate.

5.2.1.3 Copper Copper-sensitive ISEs can be formed by co-precipitation of copper and silver sulfides that are then pressed and sintered to form disks. Alternatively, the precipitate can be dispersed within an inert binder.

The resulting sensing element contains an intimate mixture of the two sulfides and a ternary compound jalapite ($Ag_{1.5}Cu_{0.5}S$). Mercury(II), silver(I) and chloride interfere significantly in the copper determinations. In common with the lead and cadmium electrodes, the response of the copper ISE does not always give the predicted Nernst slope of approximately 30 mV dec^{-1} when complexes are present. The DL for free metal concentrations is impressive, at 10^{-10} M, but is rarely required in environmental analysis where divalent metal ion concentrations are normally above 10^{-7} M.

5.2.1.4 Alkali Metals Neutral-carrier-membrane ISEs, e.g. hydrophobic esters of phosphoric acid or *N*,*N*,*N*′,*N*′-tetracyclohexyl-3-oxapentanediamide (ETH 129), or functionalized calixarenes, immobilized within plasticized PVC, are the most important calcium sensors. Calcium ISEs typically exhibit Nernstian responses to Ca^{2+} activity in the range 10^{-1}–10^{-7} M in metal ion buffers. The observed potentials are usually independent of pH between 5.5 and 9. The majority of applications lie in the clinical domain, e.g. calcium levels in renal and muscle cells, as well as neurons, and the only environmental application of note is the determination of water hardness. The direct potentiometric method can be used to rapidly measure water hardness. However, an indirect potentiometric method involving ethylenediamine tetraacetic acid (EDTA) titration is preferred. Potassium ISEs based on immobilized valinomycin are very successful and have been used for potassium determination in ores, slags and coolants in water-cooled power reactors, and for monitoring fertilizer run-off from agricultural land.

5.2.2 Anions

5.2.2.1 Chloride ISEs are approved by the USEPA for the determination of simple chloride ions in drinking waters, natural surface waters, groundwaters, domestic and industrial wastewaters and in soil extracts.[53] The DL of the method is 2.0 ppm and chloride concentrations between 0.4 and 1000 ppm can be measured, although concentrations less than 2.0 ppm can be overestimated by approximately 60%. Measurements in solutions that contain high levels of iodide, cyanide or sulfide should be avoided since these anions form silver salts that are less soluble than silver chloride, which forms the basis of the sensor. Erroneous readings may also be obtained if the sensor is exposed to ammonia, because these solutions tend to dissolve silver chloride. Sulfide, cyanide and ammonia can be removed as interferences by acidifying the solution to pH 4 using sulfuric acid, while addition of acidified potassium bromate oxidizes the iodide and bromide and removes them as interferences. Other interferences arise from metal ions, e.g. Fe^{3+} and Al^{3+}, that complex chloride preventing it from being measured. Under these circumstances, the sample is usually treated with EDTA, which forms strong complexes with the metal ions.

5.2.2.2 Cyanide Cyanide in aqueous samples is an environmentally important analyte that can be quantified using a potentiometric sensor.[54] Using this approach, the free cyanide and hydrocyanic acid concentration can be measured, or the total cyanide level determined if the sample is first distilled. The DL is 0.05 ppm and cyanide concentrations from 0.01 to 10 ppm can be determined. However, results less than 0.05 ppm may be up to 120% too high. The main interferences arise from transition metal ions that form strong complexes with cyanide which cannot be measured by the ISE. Distillation decomposes the metal–analyte complexes allowing the total cyanide concentration to be determined. The method can be extended to the analysis of soil samples after a suitable extraction step.

5.2.2.3 Fluoride The first significant anion-selective ISE was the fluoride-selective electrode reported by Frant and Ross[55] in 1966. This sensor is especially important as there are not many alternative methods available for fluoride monitoring. The selectivity and overall performance of this sensor are impressive, with hydroxyl representing the only significant interference. The determination of fluoride levels in potable water is important since a level of 1 ppm fluoride is recommended to promote dental health. The fluoride ISE offers excellent performance, with a typical DL of 0.5 ppm and a linear range between 0.025 and 500 ppm being obtained. However, using a linear calibration procedure may result in samples with fluoride concentrations less than 0.5 ppm being read approximately 160% too high. As is the case for chloride, polyvalent metal ions represent a significant difficulty if the objective is to determine the total, rather than free, fluoride concentration. This problem can be reduced by adding a buffer to adjust the total ionic strength which contains a strong metal chelator. The sample pH is a critical issue since the selectivity coefficient for hydroxide is approximately 0.1. Therefore, the pH is usually adjusted to between 5 and 5.5. Too acidic a medium should be avoided because at low pH values fluoride forms HF_2^- which cannot be detected by the ISE.

There is a long history of using ISEs for industrial hygiene monitoring, especially through the determination of fluoride in urine.[56] A National Institute of Occupational Safety and Health (NIOSH) method for fluoride has been established using this analytical approach.[57] In keeping with the demands of the workplace, the method is uncomplicated and involves the simple dilution of a urine sample with buffer, and direct measurement of the fluoride activity using the ISE. The addition of the pH buffer

minimizes the impact of H_2O_2 and Ca^{2+}. The reported range of the method is approximately 0.20–2000 ppm of fluoride.

5.2.2.4 Sulfide ISEs can be used to determine the total sulfide concentration in aqueous samples. The method is applicable only to samples distilled from a scrubber solution that contains oxygen scavengers such as salicylic or ascorbic acid, because of possible interference from mercury and silver. Also, distillation decomposes sulfide complexes to free sulfide dianions that can be determined by the ISE. The calibration range is between approximately 0.1 and 12 000 ppm. Results obtained for solutions containing less than 1 ppm sulfide may be approximately 90% too low. Silver, mercury, humic acid or thiosulfate do not interfere significantly with the analysis.

Underground corrosion, spoilage of petrol, paper pulp and foodstuffs can be caused by anaerobic sulfate-reducing bacteria. Therefore, the detection of these bacteria is an important economic objective. The sulfide ISE is not only capable of determining sulfide concentrations at the parts per billion level, it can also detect the initial growth of sulfate-reducing bacteria.[58] Further studies demonstrated that sulfate, sulfite, thiosulfate, pyrosulfite and dithionite, but not dithionate, are converted stoichiometrically to sulfide by bacteria such as *Desulfovibrio desulfuricans*, *D. gigas* and *D. vulgaris*.[59] Sulfide ISEs have also been used to monitor sulfide and thiols in petroleum process streams.[60]

5.2.2.5 Nitrate The USEPA has deleted the brucine–sulfanilamide method for determining the solubilized nitrate concentration (method 9200) and replaced it by an ISE method.[61] The original approach has been abandoned because it produced unreliable results. The ISE method gives a LOD of 2.0 ppm and nitrate concentrations from 0.2 to 1000 ppm can be measured. It is important to note that simple linear calibration models can introduce significant errors at both the high and low ends of this range. For example, results obtained for samples containing <2 ppm nitrate may be up to 400% too high, while results for solutions containing >400 ppm nitrate may underestimate the nitrate concentration by approximately 50%. The nitrate ISE is prone to a number of interferences that are typically masked by adding appropriate reagents. For example, cyanide, bisulfide, bicarbonate, carbonate and phosphate are removed as interferences by adjusting the pH to 4 with boric acid. The halides are removed as precipitates by adding silver sulfate. Nitrite can be removed by adding sulfamic acid. As discussed in section 3.2, adding reagents will change the ionic strength of the sample, which may in turn change the activity of the target ion thus altering the ISE potential. Therefore, one must be careful that samples and calibration standards are treated similarly, or that an ionic-strength-adjusting buffer is used. Also, some prior knowledge of the interference concentration is desirable if the interferences are to be successfully eliminated. One must also be careful about contaminating the samples with chloride, e.g. from a leaking saturated calomel reference electrode, and the use of a mercury sulfate reference is recommended.

While the nitrate-selective electrode discriminates well between nitrate and chloride, $K^{pot}_{NO_3,Cl^-} = 4 \times 10^{-3}$, it is in fact more sensitive to perchlorate than nitrate, $K^{pot}_{NO_3,ClO_4} = 550$. Therefore, the nitrate ISE suffers from a strong interference from perchlorate. However, perchlorate is rarely present in samples where nitrate has to be determined. Despite the small value observed for $K^{pot}_{NO_3,Cl^-}$, the response of the nitrate electrode will be compromised where the chloride concentration is high. This difficulty has lead to the use of ''ion-buffers'' for sample pretreatment. In the case of a high background chloride concentration, silver sulfate is added to form a precipitate. Other ''buffers'' include the addition of aluminum sulfate to complex anions of organic acids, and the use of boric and sulphamic acid to decompose nitrite. When using these pretreatments, it is necessary to adjust the pH to approximately 3 by adding sulfuric acid so that any organic acids present remain undissociated and the hydrogen carbonate level is controlled.

5.2.2.6 Nitrite Nitrite levels within aqueous environmental samples can also be determined using an ISE.[62] The DL of the method is 0.05 ppm with a calibration range that extends from 0.05 to 20 ppm. In common with many anion determinations by ISE, when measuring nitrite levels a buffer that adjusts the sample pH and masks interferences must be added. Table 6 summarizes the selectivity coefficients observed for major interferences of the method.

5.2.3 Miscellaneous Species

5.2.3.1 Ammonia The gas-sensing ammonia electrode, the ammonium-sensitive liquid-membrane electrode and the ammonium-sensitive glass electrode can be used for determining ammonia in water. The liquid-membrane electrode has been used to monitor ammonia levels in natural bodies of water and boiler water feeds. However, it suffers from significant interference from alkali metal cations and has a higher DL than the gas-sensing ISE. In general, the poor selectivity of these ISEs makes them suitable only for application in high-purity waters.

5.2.3.2 Organic Amines Beyond the ability of functionalized calixarenes to sequester alkali and heavy

Table 6 Nitrite electrode selectivity data

Interfering ion	Log K_{ij}^{pot}	10% error ratio (ppm)
Hydroxide	2.8	–
Fluoride	−3.1	170
Chloride	−3.1	320
Chlorate	−3.4	1600
Perchlorate	−3.1	830
Bromide	−3.0	570
Iodide	−1.2	15
Sulfate	−4.1	1100
Nitrate	−3.3	200
Phosphate	−4.0	9500
Polyphosphate	−4.4	3400
Bicarbonate	−3.3	870
Acetate	−3.2	720
Lactate	−4.9	Unlikely to interfere
Phtalate	−2.5	380
Ascorbate	−4.2	Unlikely to interfere
Salicyclate	−0.8	7.0

metals, they are also able to complex organic cations. The best responses are obtained for primary amines with no substituents adjacent to the amino group, e.g. 1-octylamine, dopamine and hexamine. It appears that the host–guest interaction involves hydrogen bonding between the carbonyl groups of the calix[6]arene and the protonated amino group of the organic amine. Locating the alkyl chain of the amine within the hydrophobic cavity of the calixarene further enhances the stability of the inclusion complex.

5.3 Arrays in Water Quality Monitoring

There are many ion-selective chemical sensors that can play useful roles in determining water quality. However, relatively few of these sensors can be used for the *direct* selective potentiometric measurements in naturally occurring bodies of water or waste effluents that contain large and widely varying concentrations of interferences. Indirect measurements can enhance the reliability and quality of the analytical information while extending the number of target analytes. However, sampling and sample pretreatment are often an integral part of these procedures. One of the most promising approaches to addressing the difficult issues of reliability, multiple analyte detection, minimum sample pretreatment, real-time monitoring, etc. is the use of sensor arrays of the type illustrated in Figure 12.[51]

A sensor array for the detection of Cu^{2+}, Pb^{2+}, Cd^{+}, Zn^{2+}, Fe^{3+}, H^{+}, and Cl^{-} has been developed.[63]

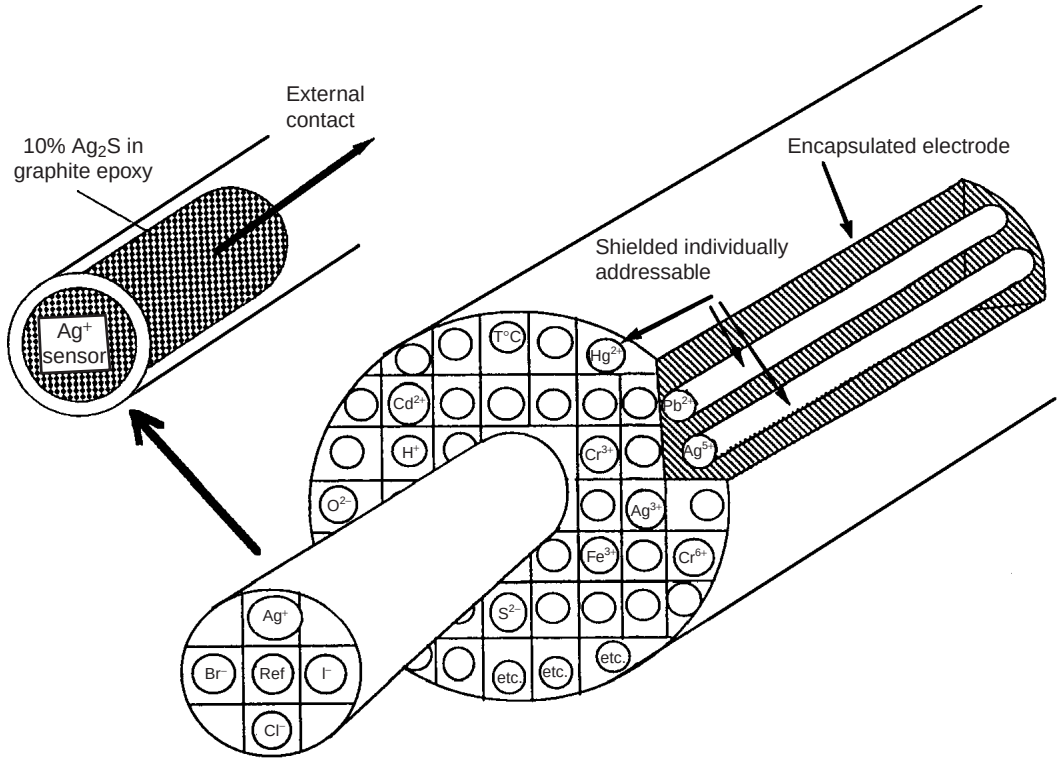

Figure 12 Schematic representation of a 50-electrode array. The expanded section shows that each electrode is independently addressable, allowing simultaneous detection of many environmentally relevant ions, correction for inter-electrode interferences, and subsequent speciation into major components. (Reproduced from Shatkin et al.[51] by permission of the American Chemical Society.)

Over thirty different sensor materials, mainly chalcogenide vitreous alloys, were considered. The vitreous materials exhibit good sensitivity towards their primary ion but respond significantly to the other analytes of interest. While such ''cross-selectivity'' would be detrimental to the performance of a single device, in a sensor array it enhances the quality of the analytical information available, allowing each sensor to provide a pseudo-independent determination of the concentration of each analyte. An important consideration when using ISEs for direct water-quality monitoring, is to consider the effects of changing temperature and pressure on the ISE response. In this regard, ISEs fabricated using incompressible solid-state materials offer improved performance and outputs that are much less sensitive to the external pressure. For example, as illustrated in Figure 13, the observed potential for a chloride-selective ISE based on a crystalline membrane of $AgCl-Ag_2S$ varies linearly with concentration over three decades of concentration even when operated with an external pressure of 10 bar.

5.4 Soil Monitoring

Metals, notably the heavy metals, can be extremely toxic and unlike hazardous organic materials cannot be broken down in the environment. Therefore, there is a strong need for simple sensors capable of providing analytical, time- and distance-dependent information about metal concentrations in soil for both protection and cleanup purposes. However, all soils contain trace levels of metal and their presence is not indicative of contamination. Moreover, while Table 7 provides information regarding the content of various metal and metal-like elements in soil,[64] local variations can be rather dramatic. For example, the average concentration of 0.3 ppm for Se can naturally reach 1200 ppm in seleniferous soils.

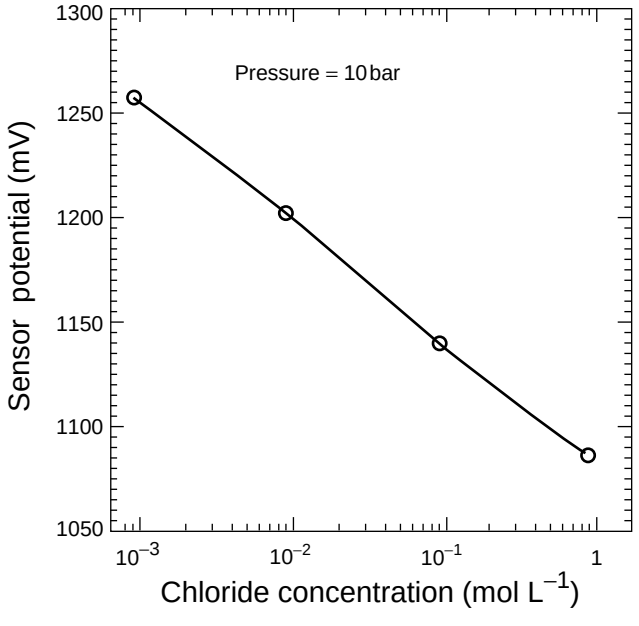

Figure 13 Chloride-selective sensor response under the pressure and in the typical concentrations found in natural and seawater. (Reproduced from Legin et al.[63] by permission of Elsevier Science.)

Metals within soils can exist either within the soil solution or bound within the organic and inorganic components. The metals may be free (uncomplexed), e.g. Cd^{2+}, Cu^{2+}, Hg^{2+}, in various complexes, e.g. $CdSO_4$, $ZnCl^+$, AsO_4^{2-}, or exist in their elemental forms. Therefore, not only is total metal concentration of interest, it is highly desirable to know bound vs free concentrations and valence states.

Soil analysis using an ISE is typically performed by either investigating samples of soil solution, i.e. collecting water runoff within the soil, or by extracting the soil into an aqueous medium. After suitable cleanup, e.g. to remove waste organic matter and humic acid, these samples can then be analyzed using the same ISE-based methods described for water samples. Several of the metals listed in Table 7, e.g. copper, cadmium, silver and lead occur at concentrations that can be directly read using an ISE. Others, e.g. mercury, require either specialized sensors capable of ultralow-level determinations, or more usually a preconcentration step.

Beyond metal analysis, nitrate determination in solids is an important application of ISEs. Concern about the intensive application of nitrogen-based fertilizers, and their potentially adverse effects on the environment, has driven the development of improved fertilizer management. Different soil-nitrate levels within a single field mean that application levels ought to be adjusted across the field to provide the optimum nutrient level for the particular crop. Use of conventional sampling and analysis techniques is not practical in this instance because

Table 7 Typical content of various metals in soils

Metal	Selected average for soils (ppm)	Common range for soils (ppm)
Al	71 000	10 000–300 000
Fe	38 000	7000–550 000
Mn	600	20–30 000
Cu	30	2–100
Cr	100	1–1000
Cd	0.06	0.01–0.70
Zn	50	10–300
As	5	1–5
Se	0.3	0.1–2
Ni	40	5–500
Ag	0.05	0.01–5
Pb	10	2–200
Hg	0.03	0.01–0.3

of the cost and time required for intensive sampling. Real-time nitrate sensors make it feasible to locally map the nitrate levels within farmland, thus maximizing yields while minimizing cost and environmental impact.

5.5 Gaseous Environmental Pollutants

Gaseous pollutants represent one of the most formidable threats to the environment because they possess a unique ability to travel considerable and often unpredictable distances. Furthermore, substances not commonly regarded as gases may also be released into the atmosphere as aerosols, including heavy metals such as lead. Such releases represent an immediate threat through inhalation. Moreover, they can be rapidly deposited, e.g. absorbed by soil, water or plant-life. Gaseous chemicals can be incorporated into rain, e.g. dissolution of NO_2 and SO_2 results in destructive acid rain. Table 8 shows the more important gaseous atmospheric pollutants that can be analyzed by ISEs. The majority of gaseous tropospheric pollutants are produced as a consequence of combustion and volcanic activities.

Air monitoring presents some unique problems to the analyst as a result of the gaseous state of the analyte. Continuous monitoring is often required, making spectroscopic approaches particularly attractive. Although some ISEs are capable of performing in situ determinations of analyte concentrations in gaseous samples, the majority of ISE-based determinations involve aqueous samples. Thus, ISE-based analysis of gases is typically a batch process in which the sample is adsorbed within a solid or liquid matrix and subsequently analyzed as an aqueous sample.

Table 8 Gaseous analytes and sources that are amenable to determination by ISE

Compound	Biogenic sources	Anthropogenic sources
CO_2		Oxidation most commonly from combustion of hydrocarbon materials
NO	Soil biochemistry, electric storms, natural fires	Combustion of oil, gas and coal
NO_2	Soil biochemistry, electric storms, natural fires	Combustion of oil, gas and coal
NH_3	Aerobic decomposition of biological material in the soil	Coal and fuel oil combustion, waste treatment, agriculture
SO_2	Volcanic activity, oceans and swamp emissions	Nonferrous metallurgical extraction, combustion of oil and coal
H_2S	Volcanic activity, anaerobic fermentation	Oil refining, farming, polymer and paper industry
HF	Volcanic activity	

Wet collection systems, usually involving buffered aqueous solutions, are employed to extract the analyte from the gaseous sample. Combinations of bubblers and diffusers provide efficient air-to-absorbant contact. This method offers certain advantages, e.g. since the samples are preconcentrated, very low DLs can be achieved in comparison to direct-reading methods. However, the ability to absorb a target analyte may be limited by the low solubility of the particular gas in aqueous media or by slow diffusion across the gas-permeable membrane, which leads to slow response times and poor selectivity.

5.5.1 Carbon Dioxide

This greenhouse gas is produced ubiquitously during combustion processes. Monitoring of CO_2 emissions is important for understanding its role as a greenhouse gas, for detecting rain-forest destruction, and for controlling biotechnological activities. Many very effective Severinghaus-type gas-permeable membrane electrodes are commercially available for measurement of CO_2, and ISEs are the method of choice for carbon dioxide monitoring.

These sensors can determine CO_2 levels either in solution or in humidified air. The probe uses glass and reference electrodes mounted behind a gas-permeable membrane such as silicone rubber. A thin film of aqueous sodium hydrogen carbonate solution separates the membrane and the surface of the glass electrode. Sample CO_2 gas traverses the permeable membrane until the partial pressures of CO_2 in the sample and in the internal filling solution are equal. CO_2 reacts with the $NaHCO_3$ internal filling solution to form carbonic acid. The glass electrode monitors the pH of the filling solution and the magnitude of the pH shift can be used to indirectly determine the CO_2 concentration. Buffering of samples and standards below pH 4.5 is often required to ensure conversion of carbonate and bicarbonate to CO_2. Because CO_2 diffuses rapidly out of acidic solution, it is important to minimize the surface area of the sample, to avoid vigorous stirring and to make measurements as soon as possible after acidification.[65]

The LOD is rather low, with a limit of 1×10^{-5} M being observed for most electrodes. Depending on the membrane employed, this method may be prone to interferences. Microporous membranes provide little selectivity and are particularly prone to permeation of water vapor, which may swamp the electrode and produce significant drift. The most important interferences are

volatile gaseous species such as SO_2 and NH_3. While silicone rubber offers good selectivity and a high diffusion coefficient for CO_2, it is prone to interference from the passage of organic acids which may change the pH of the detector.

5.5.2 Nitrogen Oxides

NO is liberated as a product of many combustion processes but is oxidized by ambient air into NO_2. This is an extremely toxic gas at high levels and an irritant at low levels, and is a primary component of photochemical smog. The NO_x level in air usually varies from 1 to over $100\,\mu g\,m^{-3}$. Therefore, it is an important target analyte in environmental monitoring. An ISE using a chalcogenide glass ($Se_{60}Ge_{28}Sb_{12}$) as the sensing element can selectively determine NO_2 without interference from nitric oxide, sulfur dioxide, carbon monoxide, methane and other gases. However, a number of other approaches to the potentiometric analysis of NO_x have been taken. The traditional nitrogen oxide electrode arrangement is analogous to the Severinghaus-type gas sensor described in section 4.3. This electrode responds to dissolved nitrite or nitrous acid. In air-quality monitoring applications, the gaseous NO_2 is collected in an appropriate collection medium where it is hydrolyzed to acidic nitrite which is then detected by the glass electrode. Unlike the CO_2 electrode, the inner filling solution is unbuffered, but maintains a high nitrite concentration. The dissolved NO and NO_2, which are in equilibrium with the HNO_2 in the acidified sample, diffuse across the electrode membrane and establish an equilibrium between NO_2^- in the sample and electrode thin film.[66]

As with all Severinghaus electrodes, interference arises from the nonselectivity of the membrane to other pH-altering species. To address the problem of ISE selectivity, Meyerhoff[67] replaced the glass pH electrode detector with a polymer-membrane electrode coupled to a membrane dialyzer injection loop. The NO_x was pumped through a Teflon® membrane and trapped in an oxidizing recipient solution where it was converted to nitrite, which was then detected by the polymer-membrane electrode. This system enhanced the selectivity of NO_x detection and was semiautomated for continuous flow. Modifying the gas-permeable membrane may also be productive, e.g. the selectivity and sensitivity of nitrite sensors is enhanced by incorporating cobalt(III)cobrinate into microporous PVC membranes.[68,69]

Nitrogen oxides can be chemisorbed on lead dioxide heated to 190 °C to form lead nitrate which, after extraction, can be measured using a nitrate ISE. Carbon dioxide, water vapor and nitrogen do not absorb and therefore do not interfere with the measurement. However, when the sulfur dioxide concentration is more than ten times the nitrogen oxide concentration, positive errors will be observed. Nitrogen oxides can also be selectively oxidized by passing them through a 2% hydrogen peroxide solution for 24 h at a rate of $2\,L\,min^{-1}$. Unreacted peroxide is then removed by adding manganese dioxide and the nitrate level is quantified using a nitrate ISE. Unlike the lead dioxide method, even a forty-fold excess of sulfur dioxide does not interfere with the analysis.

One of the great advantages of ISEs is that their small size and portability lends itself to in situ monitoring. A personal dosimeter that first oxidizes NO to NO_2 using sodium dichromate and sulfuric acid on a glass filter, followed by further oxidation using peroxide and determination of the resulting nitrate, has been developed.

5.5.3 Sulfur Dioxide

Sulfur dioxide is an important component of air pollution and its production by power-station smokestacks is routinely monitored. The major global contribution to SO_2 comes from natural volcanic activity, and the anthropogenic contributions are produced in the metallurgical extraction industry and from the burning of high-sulfur coal. When in the troposphere, this gas actively participates in physicochemical processes such as cloud and aerosol formation, making it the major contributor to acid rain. The atmospheric concentration of SO_2 varies very significantly and can reach $5\,mg\,m^{-3}$ in heavily polluted air.

Atmospheric monitoring of SO_2 is generally carried out using batch analysis, because preconcentration is required owing to the insensitivity of most analytical methods. Preconcentration is typically achieved by bubbling the gaseous sample through a buffered solution of tetrachloromercurate(II) whose pH is maintained at a value of 6.9 to ensure quantitative adsorption. The adsorption process results in the formation of $Hg(SO_3)_2^{2-}$. After chemical removal of interfering nitrite (by addition of amidosulfonic acid), the solution is acidified to release the sulfur dioxide gas. A number of more selective electrodes have been reported, for example Meyerhoff et al. have detected SO_2 after reduction by a sulfite-selective polymeric membrane electrode based on bis(diethyldithiocarbamato) mercury(II).[70–72]

5.5.4 Ammonia

Ammonia plays an active role in atmospheric pollution chemistry. It is thought to contribute substantially to atmospheric NO_x and, as the only basic trace gas constituent of the atmosphere, it plays an important role in the chemistry of the acidic gases NO_2 and SO_2. Concentrations of NH_4^+ and NH_3 in the unpolluted

atmosphere do not exceed a few micrograms per cubic meter. Therefore, sensors with low LODs or preconcentration are required.

The Severinghaus-type ammonia gas sensor is one of the more effective ISEs available for atmospheric monitoring. It responds very selectively to dissolved ammonia and, with the exception of volatile amines, is subject to very few interferences. Use of a PVC–nonactin composite membrane[73] is reported to improve the analytical performance with semiquantitative detection of ammonia gas in the range 260–0.4 ppb being achieved. In the more common batch technique, gaseous ammonia is collected in an absorbing medium either directly or after adsorption onto a solid support.[74]

5.5.5 Halides in Air

The solid-state fluoride-selective electrode is one of the earliest examples of the application of ISEs to atmospheric monitoring.[75,76] Sampled air is usually passed through a collection medium and monitored by either a solid fluoride ISE or a HF gas-sensing probe. Chlorine is most commonly measured using the chlorine gas-sensitive probe, where the sensing probe is either acid- or chloride-sensitive. Airborne halide-containing acids, e.g. hydrochloric and hydrobromic acids, which ionize to yield free halide ions, both pose health risks and cause corrosion problems. In common with the analysis of metal particles in air, the halide analyte is first trapped on an impregnated filter through which a known volume of air is drawn. Cellulose acetate is commonly used as a sorbent for both particulate and gaseous halides. The halide species is extracted or wet-ashed in alkaline solution, diluted appropriately and analyzed by ISE. By using a filter to preconcentrate the analyte, the poor LODs associated with these ISEs can be circumvented and trace levels can be measured successfully.

5.5.6 Heavy Metals in Air

Typically, to perform a potentiometric determination of heavy metals that may be present in air, particulate material is first collected using a filter with a sufficiently small pore size to trap the material. A given volume of air is drawn through the filter at a predetermined rate and then the filter is digested, typically in acid. Following suitable neutralization and dilution, the filter contents are analyzed. This approach has been used to determine the concentration of mercury in air.[77] While voltammetry is often the preferred electroanalytical technique because of its lower LODs and multianalyte capabilities, potentiometry avoids the high background signal of adsorptive stripping voltammetry due to intermetallic effects and double layer charging when attempting to determine mercury levels. The potentiometric analysis of mercury within gas samples taken from flue exhausts compares favorably with results obtained using cold-vapor AAS.[77]

6 CONCLUSIONS AND FUTURE DIRECTIONS

A significant range of environmentally important analytes can now be determined using simple and inexpensive ISEs. The range of accessible analytes will undoubtedly continue to grow as synthetic and materials chemists devise ever more robust and selective membranes. Recent developments in ISEs, including sensor arrays, miniaturized devices, portable instrumentation and advanced ISFETs, have not yet been widely exploited in environmental monitoring. Indeed, compared with other analytical methods such as spectroscopy and chromatography, electroanalytical methods in general (and ISEs in particular), have been somewhat under-utilized. In the past, ISEs have suffered from problems of mechanical ruggedness, interferences and long-term stability. However, these difficulties have largely been solved for many analytes allowing for more routine application.

Improving the recognition chemistry is pivotal to ISEs making further contributions in the area of environmental monitoring. Research will continue to refine the selectivity using existing approaches, e.g. the ability to tune ionophore selectivity by both the nature of the substituents and the conformation of the host will continue to increase the range of target analytes. These synthetic efforts will benefit from increasingly sophisticated molecular modeling that will reduce the synthetic cycle time and increase the likelihood that highly selective ionophores will be produced. New transduction mechanisms will be exploited, e.g. phase changes, and combined detection mechanisms, e.g. optical and electrochemical signals, will be developed. Successful ISE-based environmental monitoring must be interdisciplinary and demands that instrumentation specialists, surface scientists, mathematicians, theoretical chemists and materials chemists form strategic alliances. New materials, in particular those with attractive properties (e.g. easily processed, good physical and chemical stability) for applications ranging from encapsulating materials to solid-state reference electrodes will play important future roles.

''Intelligent sensing systems'' capable of recognizing and responding to changes in the concentration of environmental pollutants in a dynamic manner are likely to be an important future development. Smart sensors incorporate electronic logic, control and processing functions to offer enhanced measurement capabilities, information quality and functional performance. While these features are currently predominately found in

silicon-based transducers, they are appearing in ISEs, e.g. on-line temperature compensation.

From a fabrication perspective, the major research drive is to fabricate sensors using planar technologies. Therefore, there is intense interest in developing thin and thick films suitable for chemical sensing. Self-assembly, spontaneous adsorption, Langmuir–Blodgett and chemical vapor deposition (CVD) are all being used to produce films whose structure and chemical composition is carefully controlled and optimized for the detection of specific analytes. Miniaturization will be a recurring theme. In particular, potentiometric sensing in solids, nonpolar solvents, supercritical fluids and gases, will greatly diversify the range of environmental media in which ISEs can be successfully applied.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AES	Auger Electron Spectroscopy
CCV	Continuing Calibration Verification
CL	Chemiluminescence
CVD	Chemical Vapor Deposition
CWE	Coated-wire Electrodes
DL	Detection Limit
EDTA	Ethylenediamine Tetraacetic Acid
FET	Field-effect Transistor
FIA	Flow Injection Analysis
GC	Gas Chromatography
ICV	Initial Calibration Verification
IR	Infrared
ISE	Ion-selective Electrode
ISFET	Ion-selective Field-effect Transistor
LOD	Limit of Detection
MOSFET	Metal-oxide Field-effect Transistor
NIOSH	National Institute of Occupational Safety and Health
PVC	Poly(vinyl chloride)
TLC	Thin-layer Chromatography
USEPA	United States Environmental Protection Agency
UV/VIS	Ultraviolet/Visible
XRF	X-ray Fluorescence

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Electroanalysis and Biosensors in Clinical Chemistry • Electrolytes, Blood Gases, and Blood pH

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Flow-injection Techniques in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by Electrochemical Methods

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals

Liquid Chromatography **(Volume 13)**
Ion Chromatography

REFERENCES

1. J. Fuckskó, K. Tóth, E. Pungor, 'Application of Ion Selective Electrodes in Environmental Analysis. Determination of Acid and Fluoride Concentrations in Rain-water with a Flow Injection Analysis System' *Anal. Chim. Acta*, **194**, 163 (1987).
2. M. Cremer, 'Über die Ursache der Elektromotorischen Elektrolytketten', *Z. Biol.*, **47**, 562 (1906).
3. F. Haber, Z. Klemensiewicz, 'Electric Forces at Phase Boundaries', *Z. Phys. Chem.*, **67**, 385 (1909).
4. A. Lewenstam, A. Hulanicki, 'Selectivity Coefficients of Ion-sensing Electrodes', *Selective Electrode Rev.*, **12**, 161 (1990).
5. E.W. Baumann, 'Trace Fluoride Determination with Specific Ion Electrode', *Anal. Chim. Acta*, **54**, 189 (1971).
6. E. Bakker, 'Selectivity Comparison of Neutral Carrier-based Ion-selective Optical and Potentiometric Sensing Schemes', *Anal. Chim. Acta*, **350**, 329 (1997).
7. E. Bakker, 'Selectivity of Carrier-based Ion-selective Electrodes: Is the Problem Solved?', *Trends Anal. Chem.*, **16**, 252 (1997).
8. S. Mathison, E. Bakker, 'Effect of Transmembrane Electrolyte Diffusion on the Detection Limit of Carrier-based Potentiometric Ion Sensors', *Anal. Chem.*, **70**, 303 (1998).
9. E. Bakker, P. Buhlmann, E. Pretsch, 'Carrier-based Ion-selective Electrodes and Bulk Optodes. 1. General Characteristics', *Chem. Rev.*, **97**, 3083 (1997).

10. T. Sokalski, A. Ceresa, T. Zwicki, E. Pretsch, 'Large Improvement of the Lower Detection Limit of Ion-selective Polymer Membrane Electrodes', *J. Am. Chem. Soc.*, **119**, 11 347 (1998).
11. L. Ilcheva, K. Cammann, 'Flow Injection Determination of Chloride in Tap and Sewage Water Using Ion-selective Electrode Detection', *Fresenius' J. Anal. Chem.*, **322**, 323 (1985).
12. F.J. Sáez de Viteri, D. Diamond, 'Determination and Application of Ion-selective Electrode Model Pa Using Flow-injection and Simplex Optimization', *Analyst*, **119**, 749 (1994).
13. R.J. Forster, 'Miniaturized Chemical Sensors', in *Principles of Chemical and Biological Sensors*, ed. D. Diamond, J. Wiley & Sons, New York, 1998.
14. R.J. Forster, 'Microelectrodes – New Dimensions in Electrochemistry', *Chem. Soc. Rev.*, 289 (1994).
15. A. Hulanicki, A. Lewenstam, 'Model for Treatment of Selectivity Coefficients for Solid-state Ion-selective Electrodes', *Anal. Chem.*, **53**, 1401 (1981).
16. J. Paul, R. Carlson, 'Nitrate Detemination in Plant Extracts by Nitrate Electrode', *J. Agr. Food Chem.*, **16**, 766 (1968).
17. J.R. Allen, T. Cynkowski, J. Deasy, L.G. Bachas, 'Crown-ether Derivatives of Anthroquinones as Ionophores in Ion-selective Electrodes', *Electroanalysis*, **4**, 533 (1992).
18. R.J. Forster, A. Cadogan, M. Telting-Diaz, D. Diamond, S.J. Harris, M.A. McKervey, 'Calixarenes as Active Agents for Chemical Sensors', *Sensors and Actuators B*, **4**, 325 (1991).
19. K. Cunningham, G. Svehla, S.J. Harris, M.A. McKervey, 'Calix[4]arenes in the Partial Cone Conformation as Ionophores in Selver Ion-selective Electrodes', *Analyst*, **118**, 341 (1993).
20. S. Ozawa, P.C. Hauser, K. Seile, S.S.S. Tan, W.E. Morf, W. Simon, 'Ammonia-gas-selective Optical Sensors Based on Neutral Ionophores', *Anal. Chem.*, **63**, 640 (1991).
21. Y.M. Fraticelli, M.E. Meyerhoff, 'Automated Determination of Ammonia with a Potentiometric Gas Sensor and Flowing Internal Electrolyte', *Anal. Chem.*, **53**, 992 (1981).
22. J.W. Severinghaus, A.F. Bradley, 'Electrodes for Blood pO_2 and pCO_2 Determination', *J. App. Physiol.*, **13**, 515 (1958).
23. R.J. Forster, D. Diamond, 'Nonlinear Calibration of Ion-selective Electrode Arrays for Flow Injection Analysis', *Anal. Chem.*, **64**, 1172 (1992).
24. J. Janata, R.J. Huber, 'Ion Sensitive Field Effect Transistors', *ISE Rev.*, **1**, 31 (1979).
25. J. Janata, in *Solid State Chemical Sensors*, eds. J. Janata, R.J. Huber, Academic Press, New York, 1985.
26. A. Sibbald, 'Semiconductor Device Sensors', *J. Mol. Electron.*, **2**, 51 (1986).
27. A. van den Berg, P. Bergveld, D.N. Reinhoudt, E.J.R. Sudholter, 'Sensitivity Control of ISFETS by Chemical Surface Modification', *Sensors and Actuators*, **8**, 129 (1985).
28. C. Cane, I. Garcia, A. Merlos, 'Microtechnologies for pH ISFET Chemical Sensors', *J. Microelectron.*, **28**, 389 (1997).
29. J. Ruzicka, *Flow Injection Analysis*, J. Wiley & Sons, New York, 1981.
30. C.B. Ranger, 'Flow Injection Analysis, Principles, Techniques, Applications, Design', *Anal. Chem.*, **53**, 20A (1981).
31. H. Cui, S. Zhang, 'Preparation of PVC Tubular Flow-through Potassium Ion-selective Electrode and its Application in Flow Injection Analysis', *Fenxi-Huaxue*, **16**, 320 (1988).
32. H. Wada, T. Ozawa, G. Nakagawa, Y. Asano, S. Ito, 'Preparation and Examination of Calcium Ion-selective Electrodes for Flow Analysis', *Anal. Chim. Acta*, **211**, 213 (1988).
33. A.J. Frend, G.J. Moody, J.D.R. Thomas, B.J. Birch, 'Flow Injection Analysis with Tubular Membrane Ion Selective Electrodes in the Presence of Anionic Surfactants', *Analyst*, **108**, 1357 (1983).
34. K. Cammann, 'Flow Injection Analysis with Electrochemical Detection', *Fresenius' J. Anal. Chem.*, **329**, 691 (1988).
35. T.J. Cardwell, R.W. Cattrall, P.C. Hauser, I.C. Hamilton, 'A Multi-ion Sensor Cell and Data Acquisition System for Flow-injection Analysis', *Anal. Chim. Acta*, **214**, 359 (1988).
36. S. Alegret, J. Alonso, J. Bartroli, J.M. Paulis, J.L.F.C. Lima, A.A.S.C. Machado, 'Flow-through Tubular PVC Matrix Membrane Electrode Without Inner Reference Solution, for Flow Injection Analysis', *Anal. Chim. Acta*, **164**, 167 (1984).
37. J. Alonso, J. Bartroli, J.M. Del-Valle, M. Escalada, R. Barber, 'Sandwich Techniques in Flow Injection Analysis, Part I, Continuous Recalibration Techniques for Process Control', *Anal. Chim. Acta*, **199**, 191 (1987).
38. P. Petak, K. Stulik, 'Use of Ion-selective Electrodes in Industrial Flow Analyzers', *Anal. Chim. Acta*, **185**, 171 (1986).
39. J.L.F.C. Lima, S.M. Rocha, 'FIA Tubular Potentiometric Detectors Based on Homogeneous Crystalline Membranes–Their Use in the Determination of Chloride and Sulphide Ions in Water', *Int. J. Environ. Anal. Chem.*, **38**, 127 (1990).
40. M.G. Glaister, G.J. Moody, J.D.R. Thomas, 'Studies on Flow Injection Analysis with Sulfide Ion-selective Electrodes', *Analyst*, **110**, 113 (1985).
41. J.F. Van-Staden, 'Determination of Sulphide Using Flow Injection Analysis with a Coated Tubular Solid-state Silver Sulphide Ion-selective Electrode', *Analyst*, **113**, 885 (1988).
42. M. Trojanowicz, W. Matusewski, 'Limitation of Linear Response in Flow-injection Systems with Ion-selective Electrodes', *Anal. Chim Acta*, **138**, 71 (1982).

43. J.F. Van-Staden, 'Comparison of the Influence and Contribution of Response Times of Coated Open-tubular Solid-state Bromide-selective and Chloride-selective Electrodes on the Analytical Throughput (Dispersion) in Flow Injection Systems', *Analyst*, **117**, 51 (1992).
44. J.F. Van-Staden, 'Effect of Coated Open-tubular Inorganic-based Solid-state Ion Selective Electrodes on Dispersion in Flow-injection', *Analyst*, **115**, 581 (1990).
45. J.F. Van-Staden, 'Flow Injection Determination of Inorganic Bromide in Soils with a Coated Tubular Solid-state Bromide-selective Electrode', *Analyst*, **112**, 595 (1987).
46. A. Sakai, Y. Asano, 'Flow Injection Analysis for Chloride Ion Using a Flow-through Type Chloride Ion-selective Electrode Detector and Correlation with JIS Method', *Flow Injection Anal.*, **6**, 144 (1989).
47. E.A. Jones, D. Barnes, J.F. Van Staden, 'Determination of Cyanide by Flow-injection Analysis Using a Tubular Solid-state Ion-selective Electrode as a Detector', *Afr. J. Chem.*, **42**, 25 (1989).
48. L. Risinger, 'Preconcentration of Copper (II) on Immobilized 3-Quinolinol in a Flow-injection System with an Ion-selective Electrode Detector', *Anal. Chim. Acta*, **179**, 509 (1986).
49. R.J. Forster, F. Regan, D. Diamond, 'Modeling of Potentiometric Electrode Arrays for Multicomponent Analysis', *Anal. Chem.*, **63**, 876 (1991).
50. 'Methods for Chemical Analysis of Water and Wastes', EPA Environmental Monitoring and Support Laboratory, EPA-600/4-79-020, Protocols 245.1 and 245.2, Research Triangle Park, NC, 1983.
51. J.A. Shatkin, H.S. Brown, S. Licht, 'Composite Graphite Ion-selective Electrode Array Potentiometry for the Detection of Mercury and Other Relevant Ions in Aquatic Systems', *Anal. Chem.*, **67**, 1147 (1995).
52. K.M. O'Connor, D.W.M. Arrigan, G. Svehla, 'Calixarenes in Electroanalysis', *Electroanalysis*, **7**, 205 (1995).
53. 'Methods for Chemical Analysis of Water and Wastes', EPA Environmental Monitoring and Support Laboratory, Method 9212, Research Triangle Park, NC.
54. 'Methods for Chemical Analysis of Water and Wastes', EPA Environmental Monitoring and Support Laboratory, Method 9213, Research Triangle Park, NC.
55. M.S. Frant, J.W. Ross, 'Electrode for Sensing Fluoride Ion Activity in Solution', *Science*, **154**, 1553 (1966).
56. M.-W. Sun, 'Fluoride Ion Activity Electrode for Determination of Urinary Fluoride', *Am. Ind. Hyg. Assoc. J.*, **30**, 133 (1969).
57. *NIOSH Manual of Analytical Methods*, 2nd edition, ed. D.G. Taylor, Method No. P&CAM 114, DHEW/PHS, Atlanta, GA, 1977.
58. D.J. Crombie, G.J. Moody, J.D.R. Thomas, 'Application of a Sulphide Ion-selective Electrode for Monitoring the Growth of *Desulphovibrio desulphuricans*', *Lab. Pract.*, **29**, 259 (1980).
59. I.K. Al Hitti, G.J. Moody, J.D.R. Thomas, 'Determination of Sulphide Produced by *Desulfovibrio* Species of Sulphate-reducing Bacteria', *Analyst*, **108**, 43 (1983).
60. S. Bateson, G.J. Moody, J.D.R. Thomas, 'Automated Potentiometric Titrations of Sulphides and Thiols in Petroleum Systems', *Analyst*, **111**, 3 (1986).
61. 'Standard Method of Test for Nitrate in Water', Technical Bulletin 601, PN 221890-001, USEPA, Research Triangle Park, NC, 1994.
62. 'Methods for Chemical Analysis of Water and Wastes', EPA Environmental Monitoring and Support Laboratory, Method 9216, Research Triangle Park, NC.
63. A.V. Legin, E.A. Bychkov, B.L. Seleznev, Yu.G. Vlasov, 'Development and Analytical Evaluation of a Multisensor System for Water-quality Monitoring', *Sensors and Actuators B*, **26**, 377 (1995).
64. W.L. Lindsay, *Chemical Equilibria in Soils*, J. Wiley & Sons, New York, 1979.
65. E. Sacarano, P. Nagger, R. Belli, 'A High Precision Procedure for Carbon Dioxide Determinations with the Carbon Dioxide Sensing Probe', *Anal Lett.*, **16**, 723 (1983).
66. S.S.M. Hassan, F.S. Tadros, 'Performance Characteristics and Some Applications of the Nitrogen Oxide Sensor', *Anal. Chem.*, **57**, 162 (1985).
67. G.B. Martin, M.E. Meyerhoff, 'Tubular Debubbler for Segmented Continuous-flow Automated Analyzers', *Anal. Chim. Acta*, **186**, 71 (1986).
68. S.A. O'Reilly, S. Daunert, L.G. Bachas, 'Nitrogen-oxide Gas Sensor Based on a Nitrite-selective Electrode', *Anal. Chem.*, **63**, 1278 (1991).
69. U. Schaller, E. Bakker, U.E. Spichiger, E. Pretsch, 'Nitrite-selective Microelectrodes', *Talanta*, **41**, 1001 (1994).
70. M.E. Meyerhoff, D.M. Pranitis, H.S. Yim, N.A. Chaniotakis, S.B. Park, 'New Anion and Gas-selective Potentiometric Sensors', ACS Symposium Series 990, American Chemical Society, Washington, DC, 116, 1989.
71. D.M. Pranitis, M.E. Meyerhoff, 'Sulfite-sensitive Solvent/Polymeric-membrane Electrode Based on Bis(diethyldithiocarbamato)mercury(II)', *Anal. Chim. Acta*, **217**, 123 (1989).
72. I.H.A. Badr, M.E. Meyerhoff, S.S.M. Hassan, 'Novel Response Mechanism and Application of Sulfite Sensitive Polymeric Membrane-electrode Based on Dithiocarbamated Complexes of Mercury(II)', *Anal. Chim. Acta*, **310**, 211 (1995).
73. M.E. Meyerhoff, 'Polymer Membrane Electrode Based Potentiometric Ammonia Gas Sensor', *Anal. Chem.*, **52**, 1532 (1980).
74. N.A. Dimmock, G.B. Marshall, 'The Determination of Free Ammonia in Ambient Air with Diffusion/Denuder Tubes', *Anal. Chim. Acta*, **185**, 159 (1983).

75. L.A. Elfers, C.E. Decker, 'Determination of Fluoride in Air and Stack Gas Samples by Use of an Ion Specific Electrode', *Anal. Chem.*, **40**, 1658 (1968).
76. R.J. Thompson, T.B. McMullen, G.B. Morgan, 'Fluoride Concentrations in the Ambient Air', *J. Air Poll. Control Assoc.*, **21**, 484 (1971).
77. H. Huiliang, D. Jagner, L. Renman, 'Determination of Mercury in Air by Means of Computerised Flow Constant-current Stripping Analysis with a Gold Fibre Electrode', *Anal. Chim. Acta*, **201**, 269 (1987).

Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis

Steven F. Durrant
State University of Campinas, Campinas, Brazil

1 Introduction **1**
2 Laser Ablation Systems **2**
2.1 Laser Principles 2
2.2 Laser Ablation Inductively Coupled Plasma Mass Spectrometry Hardware 3
2.3 Laser–Solid Interactions 4
3 Analysis Procedures **5**
3.1 Sample Preparation 5
3.2 Optimization 5
3.3 Quantitation 6
4 Applications in Environmental Elemental Analysis **7**
4.1 Metals 9
4.2 Geological Materials 9
4.3 Biological Materials 12
4.4 Polymers 12
4.5 Other Matrices 12
5 Isotope Ratio Analysis **13**
6 Performance and Status **15**
7 Development Prospects **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

Soon after the invention of the laser by Maiman, laser pulses were used for excitation and ionization of sample material for inorganic analysis by optical emission spectrometry (OES) or mass spectrometry (MS). This direct approach is complicated by the complexity of the laser–sample interaction and the need to optimize simultaneously the sampling and excitation (or ionization) steps. Use of a laser for solids sampling with separate atomization and excitation in an inductively coupled plasma (ICP) retains the advantages of direct solids analysis, including spatially resolved concentration data, and adds those attributable to the ICP, including wide elemental coverage, a linear dynamic range of six or more orders of magnitude, and excellent detection limits (DLs). The resulting techniques, denominated laser ablation/inductively coupled plasma/optical emission spectrometry (LA/ICP/OES) and laser ablation/inductively coupled plasma/mass spectrometry (LA/ICP/MS), are based on recording of relative emission intensities or ion count rates, respectively. Both methods are treated here, but emphasis is given to LA/ICP/MS, a rapidly developing technique in use for less than 15 years, which in addition to offering DLs typically 10^3 times lower than those obtained by LA/ICP/OES, also offers isotope ratio analysis. Hardware, analytical procedures, performance, and prospects are addressed. The power and limitations of the methods are illustrated by environmental applications, including the analysis of metals, soils, rocks, and plastics.

1 INTRODUCTION

The ICP has been used for a quarter of a century as a source for atomic emission spectrometry (AES), in which detection of element-specific optical emissions is used to determine analyte concentrations in aqueous sample solutions. The advantages of inductively coupled plasma/atomic emission spectrometry (ICP/AES) or inductively coupled plasma/optical emission spectrometry (ICP/OES), include multielemental analysis, rapid sample throughput, low noise, low background, freedom from interferences, and DLs down to the sub ng mL^{-1} range.[1]

However, ever greater analytical performance required for new applications in biomedical, geological, and environmental science led, in the mid 1970s, to searches for yet more powerful analytical instrumentation. One idea for a novel analytical technique was to marry the advantages of the ICP as an ion source with the convenience and sensitivity of MS. The resulting technique, now known as inductively coupled plasma/mass spectrometry (ICP/MS), was developed by research groups in Britain, Canada, and the USA.[2] Advantages of the union of the ICP ion source and the quadrupole mass spectrometer include rapid multielemental analysis, high sample throughput (up to about 70 samples per day), typical DLs in the pg mL^{-1} range, calibration against simple aqueous standards, a linear dynamic range of seven or more orders of magnitude, relatively simple spectra, and an intrinsic isotope ratio analysis capability.[3]

Samples for ICP/MS analysis are introduced as an aerosol into the central channel of a torch consisting of several concentric quartz tubes. The plasma is maintained by radio-frequency power (typically about 1.3 kW) fed to a load coil near the mouth of the torch through which argon is fed. The resulting fireball is an almost perfect environment for the desolvation, volatilization, dissociation, and ionization of the sample.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Principally singly charged cations are produced, which are extracted from the plasma at atmospheric pressure by an ingeniously designed interface directly into the low-pressure analyzer of the spectrometer.

Both ICP/OES and ICP/MS, today used worldwide for elemental analysis, are dealt with in this article. More specifically, one form of sample introduction is addressed: laser ablation (LA). However, far greater emphasis is given to LA/ICP/MS since this method is younger, developing more rapidly, and offers more prospects for innovation than LA/ICP/OES. Henceforward, therefore, LA/ICP/MS is addressed but much of the commentary also applies to LA/ICP/OES.

Samples for analysis by either of the ICP methods are usually put into aqueous solution and dispersed into fine droplets. The droplets are traditionally produced by a pneumatic or an ultrasonic nebulizer, which discharges into a spray chamber. This method is extremely successful, allowing straightforward calibration against aqueous standards and the use of internal standards to correct for any drift in instrument sensitivity. However, the defects of nebulization include temporal variation in the particle size distribution; the adhesion of certain elements, such as iodine, to the inner walls of the glass spray chamber and their later release during analysis (so-called memory effects); and low efficiency, especially for pneumatic nebulization, where around 98% of the sample solution is lost to the drain. Moreover, sample dissolution often requires lengthy (half a day or more) digestions, frequently employing potentially hazardous reagents such as hydrofluoric acid. Analytically too, dissolution introduces the possibility of the loss of volatile elements, contamination from glassware, and the unavoidable introduction of constituent elements of the reagents used, which can, in ICP/MS, produce spectral interferences in the sample analysis. For example, when hydrochloric acid is used as the solvent the species $^{40}Ar^{35}Cl^{+}$ may be produced in the plasma and be preserved in the mass spectrometer, leading to unrepresentative intensities at a mass to charge ratio (m/z) of 75, where arsenic is determined.

Partly in response to such difficulties, many alternative sample introduction strategies previously used with ICP/OES have been coupled to ICP/MS over the last decade. Rapid sample heating is the basis of some of these methods, specifically of electrothermal vaporization (ETV), spark ablation (SA), and LA. In ETV the sample is placed on a graphite rod and subjected to a rapid temperature ramp, resulting in vaporization of the sample into a carrier gas stream (usually argon) which transports it to the plasma. Similarly, in SA surface material is removed by the action of a high voltage spark. Advantages of ETV and SA include the ability to work with microliter sample volumes and direct solids analysis, respectively.

One of the most versatile methods of sample introduction is the use of laser pulses to mobilize sample material directly from the solid and, as with ETV or SA, entrain it in a carrier gas flow for transport to the ICP. Advantages of LA/ICP/MS encompass:

- independent optimization of the ablation and spectrometer parameters;
- little or no sample preparation;
- high sample throughput;
- very rapid semiquantitative analysis;
- reduced spectral interferences;
- spatially resolved analysis.

In the following sections the main features of the lasers used, LA systems, and laser–solid interactions are outlined. Some alternative analysis procedures are described. Specific illustrative applications of the technique to elemental environmental analysis are then given. Isotope ratio analysis is addressed separately. An overview of the analytical performance and current status of the method follows. Finally, prospects for development are assessed.

2 LASER ABLATION SYSTEMS

2.1 Laser Principles

Following the introduction of the first coherent source of microwave radiation, the maser, in 1953, attention turned to the possibility of producing a coherent visible light source – the laser (an acronym for light amplification by stimulated emission of radiation). This was achieved in 1960 by Maiman[4] of the Hughes Research Laboratory using a ruby crystal as the active medium. A typical ruby

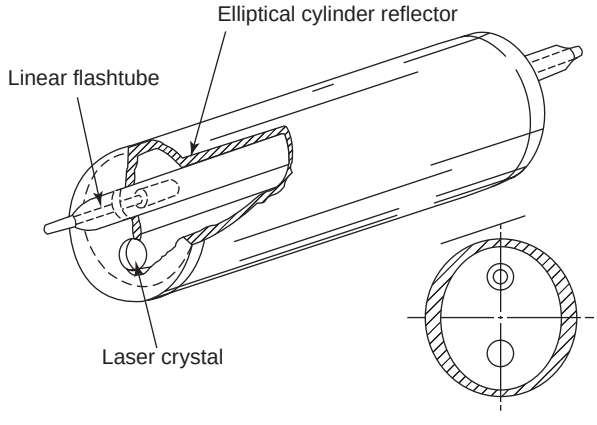

Figure 1 Ruby laser with a straight flashlamp in an elliptical cylinder configuration. (Reproduced by permission of John Wiley & Sons Inc., from B.A. Lengyel, *Lasers*, © Wiley-Interscience, p. 106 (1971).)

laser is shown in Figure 1. The elliptical reflector encloses a ruby rod and flash tube. The end surfaces of the ruby are highly polished and silvered, one end fully the other partially. The latter allows pulses of light to escape when lasing occurs. (Alternatively, mirrors external to the ruby may be used.)

Ruby consists of aluminum oxide, Al_2O_3, to which has been added a small proportion (about 0.05% by weight) of Cr_2O_3. The chromium, in the form Cr^{3+}, provides the energy levels between which stimulated emission is triggered. Chromium ions in the ground state are excited to an upper energy band by an intense flash of white light from the flash tube in a process known as optical pumping. A spontaneous non-radiative transition (phonon relaxation) to a lower metastable level then occurs. As a metastable state has a relatively long lifetime, its population can be increased at the expense of that of the ground state, producing a population inversion.

Spontaneous transitions emit photons which trigger further emissions. The photons traveling perpendicular to the ends of the ruby rod build up in the resonator and are released from the partially silvered ruby face. Once started, stimulated emission depopulates the upper lasing level much faster than the pumping rate of the flash tube. Consequently, in this free-running or normal (N) mode, the light output consists of many intense spikes. Pulse widths from about 0.1 to 1 ms are typical. However, it is possible to control the output of a laser by building up such a high population inversion through use of a light seal on the active medium that when it is opened a giant pulse is produced. The quality (Q) factor of a resonator is a measure of the net energy stored. Thus rapid modification of the Q factor, by use of a shutter for example, is the key to the method known as Q-switching. The pulses so produced, known as Q-switched or Q mode pulses, have typical durations in the range 0–1000 ns. Figure 2 shows schematically the principles of production and the form of free-running and Q-switched pulses.

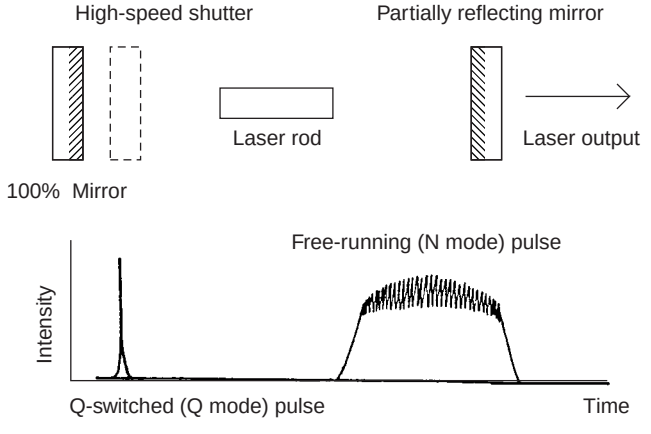

Figure 2 Schematic of a ruby laser with a high-speed shutter for Q-switching. The temporal forms of N and Q mode pulses are also shown.

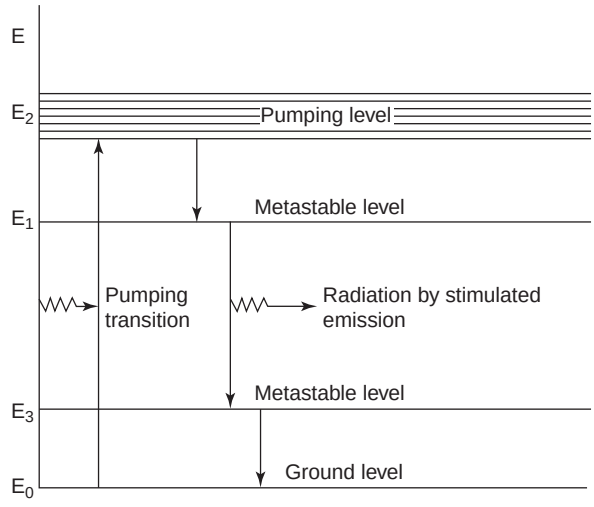

Figure 3 Energy-level diagram for Nd : YAG, indicating the optical pumping and laser transitions. (Reproduced by permission of Elsevier Science from Moenke-Blankenberg, 'Laser Micro Analysis', *Prog. Analyt. Spectrosc.*, **9**, 335–427 (1986).)

The ruby laser is of some interest since it was the first used to ablate solids for introduction to ICP/MS.[5] Subsequently, neodymium:yttrium aluminum garnet (Nd : YAG) lasers[6] have mainly been used. The energy level system of Nd : YAG is shown schematically in Figure 3. The lasing transition is between levels E_1 and E_3. As the population of E_3 is very low at room temperature, relatively little optical pumping power is required to produce a population inversion compared to a ruby system.

2.2 Laser Ablation Inductively Coupled Plasma Mass Spectrometry Hardware

A typical LA system[7] is shown in Figure 4. The laser may be mounted as shown or in the horizontal plane, the beam being reflected onto the sample through 90° by a

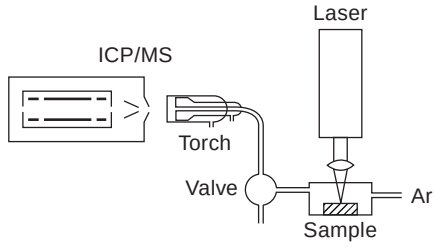

Figure 4 Laser, ablation cell, and associated gas lines. (Reproduced by permission of the Japan Society for Analytical Chemistry from Mochizuki et al.[7])

mirror. Samples are placed on a stage within a glass (or quartz) cell. The stage can be moved in the x, y, and z directions by a stepper-motor under computer control. The size and direction of the steps is controllable, and the laser is synchronized to fire only when the sample is at rest. A pre-selected number of laser shots are fired onto the sample surface, where rapid local heating causes the eruption of material in the form of molten droplets and vapor. Spectrometer data acquisition may be triggered by the first laser shot or after a pre-set delay. The sample surface may be viewed either by eye through a microscope equipped with a suitable protective filter or remotely via a video camera. Traditionally, visual and laser focusing are arranged to coincide.

A gas flow, usually of argon, is fed to the ablation cell via a precision flowmeter and entrains some of the ablated material, which is thus transported to the ICP. The cell is easily accessible for rapid sample exchange. However, as air must not be admitted to the ICP, a by-pass valve is installed beyond the cell exit line to provide an uninterrupted flow of argon and to allow purging of the cell to atmosphere.

2.3 Laser–Solid Interactions

Owing to the complexity of the process and the large number of physical variables to be considered, the influence of the incidence of a laser pulse on a solid surface cannot be analyzed as a single problem. We may say that a portion of the incident light is absorbed to a depth of a few microns. Laser photons are absorbed by electrons in the conduction band of the material and energy is given up by collisions with other electrons and lattice phonons. As the time between collisions is much less than the shortest laser pulse, the absorbed laser energy is effectively transferred instantaneously to the solid as heat.

The rapid transfer of heat even from N mode laser pulses is enough to melt, or boil, many matrices, and material erupts from the region of impact at high velocity, forming a crater. This complex process is shown schematically in Figure 5(a) and (b). To illustrate further, scanning electron micrographs[8] of a pit produced in nickel by a pulse from a ruby laser are shown in Figure 6(a) and (b). The crater's edge is very well defined and the pit is roughly conical. Shots of higher energy produce larger craters with a wide region of melted material near the crater mouth. Q mode pulses of similar energy produce only a shallow surface melting.

To simplify our discussion two regimes: low and high irradiances, corresponding to irradiances of $<10^8\,\mathrm{W\,cm^{-2}}$ and $>10^9\,\mathrm{W\,cm^{-2}}$ may be distinguished. Very crudely these correspond to N and Q mode pulses, respectively. In the former regime, most of the material is removed as liquid, i.e. in the form of droplets. In the high irradiance regime much more vapor is produced. The gas above the sample is ionized and a plasma is formed, which in turn absorbs some of the incoming radiation and partially shields the sample surface. The complexity of the interaction is illustrated by the formation of a plasma even with N mode pulses, and this may in fact occur. As material erupts from the sample in the form of a plume, complex interactions occur between it, the plasma, and the incoming laser light.

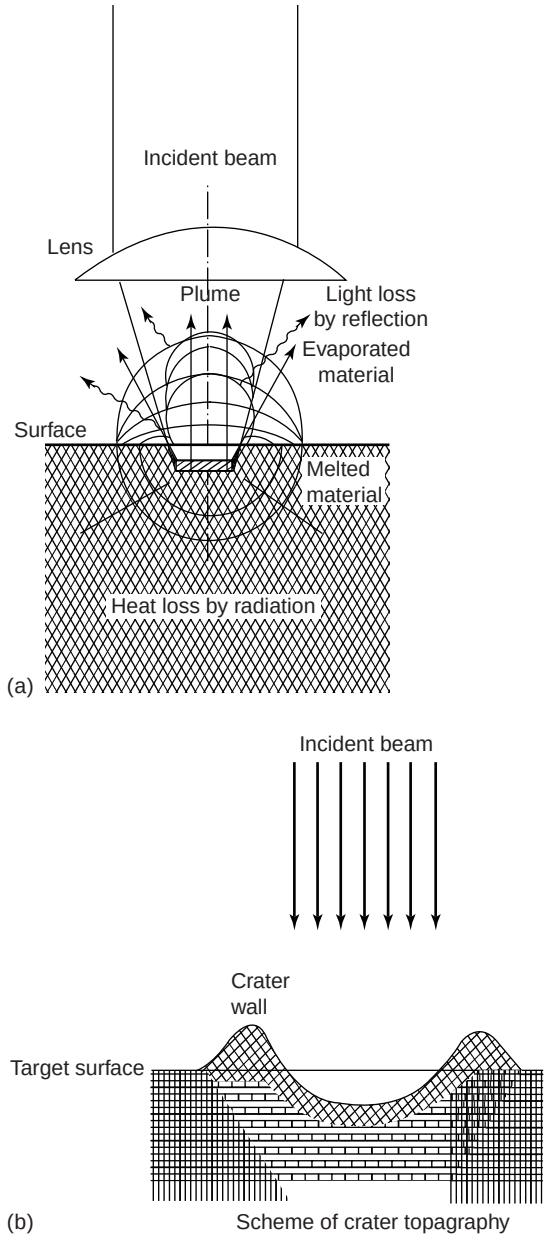

Figure 5 (a) Laser–solid interactions. (b) Cross-section of an ablated crater in a solid. (Reproduced by permission of Elsevier Science from K. Dittrich, R. Wennrich, *Prog. Anal. Spectrosc.*, **7**, 139–198 (1984).)

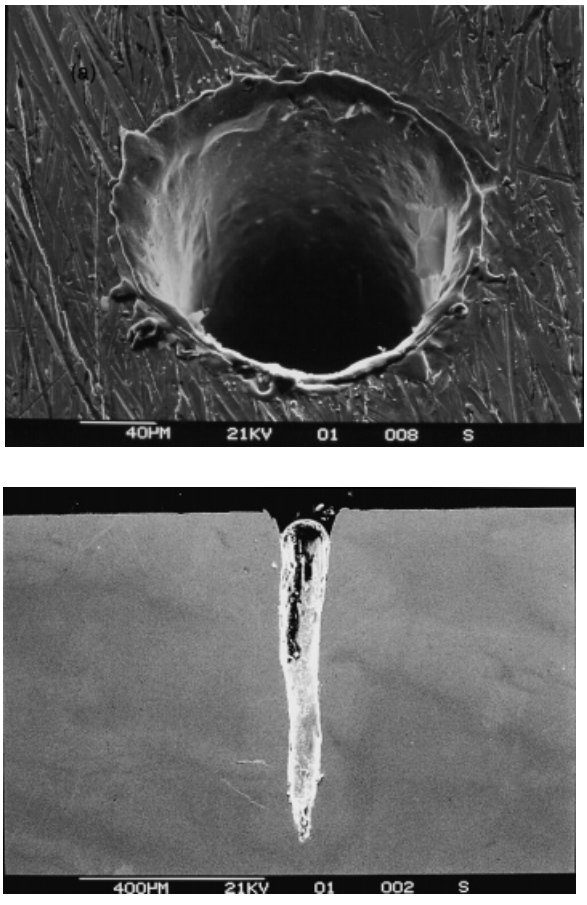

Figure 6 The effect of the impact of a 0.3-J N-mode shot from a ruby laser. (a) Crater produced in a nickel disc. (b) Cross-section of an ablated crater. Ablated mass ≅ 120 μg. (Reproduced from Durrant.[8])

Although it is always prudent to have an understanding of the underlying physics and chemistry of the ablation process as a foundation for the development of analytical laser techniques, the interactions are so complex and dependent on so many laser and sample parameters that specific experimentation is always needed. Readers interested in pursuing the subject matter of this section will find an excellent start with the review by Moenke-Blankenburg.[9]

3 ANALYSIS PROCEDURES

LA/ICP/MS can provide an elemental analysis of the bulk of a sample, or some spatially-resolved data. Small spot sizes are required for the latter but not for the former. For the present, bulk analysis is treated, spatially-resolved or laser microprobe (LM) analysis being dealt with later (section 4.2).

3.1 Sample Preparation

One of the great advantages of LA is simply that samples do not have to be put into solution. Examples of preparation include selection of a fragment that fits within the ablation cell; mounting a piece in epoxy, cutting to expose a fresh surface, polishing and cleaning in methanol; pressing a powder, with or without a binder, into a pellet; production of a glass fusion. Even such limited sample preparation can bring associated problems such as contamination produced by ball-milling; or insufficient mixing with a binder, producing a heterogeneous sample. Any use of a binder also inevitably dilutes the sample. Weak laser to sample-surface coupling, e.g. between pulses from a ruby laser and pressed milk powder, may sometimes be increased by mixing the sample with a binder that shows greater absorption at the laser wavelength being used.

3.2 Optimization

The newcomer to LA/ICP/MS is faced with a bewildering array of parameters to optimize for quantitative elemental analysis. These relate to the laser, the sample, and the spectrometer. The outcome of the laser–solid interactions depends upon, among others, the wavelength, energy, mode, focusing, and angle of incidence of the laser beam. Wavelengths of 694 nm [ruby], 1064 nm, 532 nm (frequency doubled), 266 nm (frequency quadrupled) [Nd : YAG], and 193 nm [F excimer] have been employed, with the historical trend towards shorter wavelengths, i.e. from infrared (IR) to ultraviolet (UV), for more precisely controlled LA. Generally, however, the light wavelength is determined by the laser hardware and is therefore fixed for any specific laboratory system. Greater laser energies generally produce greater ablated mass and therefore greater responses. Laser shot energies may be varied by the use of beam filters or masks.

As already noted, the ablation process is strongly influenced by the use of N or Q mode pulses. The latter vapor-producing pulses should theoretically produce less fractionation, i.e. the selective removal from the sample of one element relative to another. However, such pulses, being extremely intense and rapid, sometimes produce memory effects by vaporization or re-suspension of material previously deposited in the ablation cell and transport tubing. Moreover, Q mode pulses are associated with an acoustic shock which is transported to the ICP, decreasing the reproducibility of the analytical signal. Such deleterious effects may be partially offset by the use of multiple low-energy (mJ) pulses at high repetition rates, e.g. 10 Hz, and such schemes have been practicable since the introduction of Nd : YAG lasers.

Focusing of the laser beam is an obvious parameter of interest. Both surface- and off-surface focusing have

been employed: the former to increase ablated mass and spatial resolution; the latter to deliberately reduce these for particular applications.

The angle of incidence of the laser beam will influence the complex plasma shielding produced above the sample. Thus instead of normal incidence an angle between the beam and the sample surface of 45° is sometimes employed.[6]

Unfortunately, the ablation process depends not only on the laser parameters but also on the nature of the sample – its color, reflectivity, thermal conductivity, mineralization, and so on. Indeed, it is just such dependencies that form the Achilles' Heel of fully quantitative analysis by LA/ICP-spectrometries. For our purposes one example of the difficulties involved in quantitation will suffice. Figure 7 shows the size distribution of droplets produced by ablation of stainless steel. Only droplets of a few microns in diameter will be transported to the ICP. Therefore, a shift in the size distribution, for example between ablation of a sample and a standard, will directly influence sample transport and hence the count rate measured by the spectrometer, leading to errors in quantitation.

Finally, the spectrometer parameters should be optimized. To maximize elemental responses the cell flow and ion lens voltages may be systematically varied while monitoring the $^{12}C^+$ signal, which appears to be due to the presence of CO_2 in the Ar flow. This procedure is simple but truly optimal performance is ensured only by using realistic conditions, such as the continuous ablation of a solid of known composition. A subtle complication is that varying the cell flow alters scavenging as well as the optimal sampling depth (torch to sampling cone separation), so that the latter should also be systematically studied.

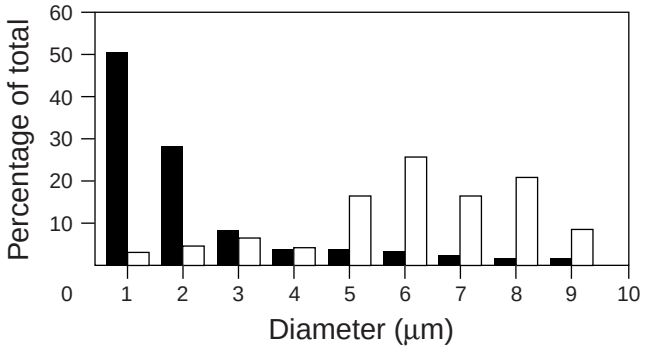

Figure 7 Particle size and mass distributions of material produced by ablation of stainless steel by lightly Q-switched pulses from a ruby laser. The black bars represent the number of particles and the open bars the relative mass. (Reproduced by permission of The Royal Society of Chemistry from S. Chenery, M. Thompson, K. Timmins, *Anal. Proc.*, **25**, 68–69 (1988).)

Owing to considerations such as these, extensive systematic studies are very time-consuming, though some have been reported.[8,10] The requirement for multielemental analysis is sufficient sensitivity for the determination of the proposed analytes and the best practicable reproducibility. Higher laser energies ablate more sample and so give higher count rates, but excess ablation leads to the deposition of excess material in the cell and exit line, often leading to memory effects. Also, excess ablation will first cause a saturation of the counts of individual matrix elements, which may not be seriously detrimental to the analysis. However, higher rates of sample ablation will cause plasma disequilibrium, disrupting the ionization process in the ICP and preventing quantitation. Thus a set of compromise conditions must be found.

3.3 Quantitation

We may note first that quadrupole spectrometers collect data at specific m/z values (peak hopping) or across a range or set of ranges of m/z (scanning). As continuous exposure to very high count rates may damage the detector, any permanently saturated peaks, such as that of $^{40}Ar^+$, are deliberately skipped.

Before specific calibration procedures are considered, brief mention is made of blanks and DLs. In solution nebulization/inductively coupled plasma/mass spectrometry (SN/ICP/MS) a blank spectrum obtained from an analysis of the solvent (typically 2% v/v HNO_3) is usually subtracted from any sample spectrum. The most rigorous procedure for LA/ICP/MS is to produce a blank spectrum by ablating an inert target such as poly(tetrafluoroethylene) (PTFE) since this exactly simulates the sample analysis. The DL, a measure of the minimum detectable concentration, is defined as 3SDs of the blank. Either this definition or, perhaps more realistically, the line equals background equivalent concentrations for a blank may be used in LA/ICP/MS.

A single laser pulse produces a transient response, typically of 20 to 30 s, depending on the cell flow rate. The peak height or area may be used for quantitative analysis, but this approach is generally inferior to the use of the more stable response generated by multiple pulses of low energy, which yields better analytical precisions. Typical responses from multiple laser pulses are shown in Figure 8. Multielemental analyses based on transient responses are also impeded because of the need for rapid scan rates from the quadrupole spectrometers.

Rapid survey analysis was always possible by LA/ICP/MS,[6] but to obtain precise and accurate elemental concentrations is the principal challenge to researchers in the field. A semiquantitative analysis requires a standard containing a few known concentrations. A response versus mass curve is then produced (using the sensitivities

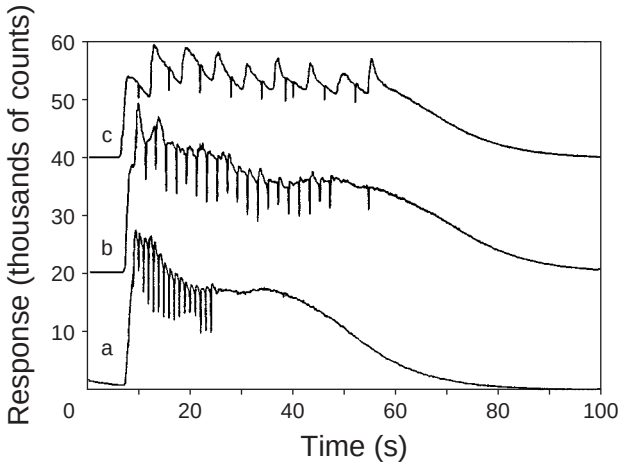

Figure 8 Response of ^{25}Mg obtained by LA/ICP/MS by ablation of a nickel disc using 0.3-J N-mode pulses from a ruby laser at (a) 1 Hz; (b) 0.5 Hz; (c) 0.17 Hz. To avoid overlapping, the responses are offset in the *Y* direction. (Reproduced from Durrant.[8])

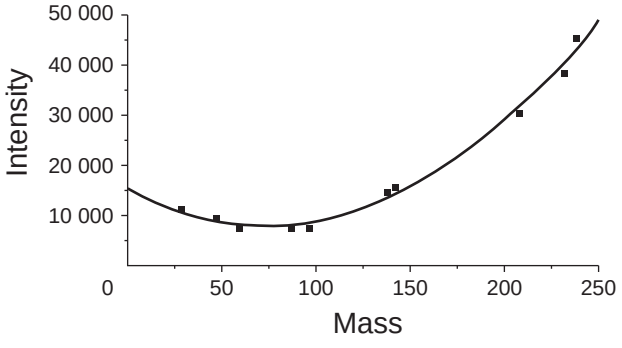

Figure 9 A response curve obtained by the ablation of NIST SRM 612 Glass using 80-mJ Q-mode pulses at 15 Hz. Elements measured: Si, Ti, Ni, Mo, Sr, La, Ce, Pb, Th, U. (Reproduced by permission of Springer-Verlag from J.L. Imbert, P. Telouk, *Mikrochim. Acta*, **110**, 151–160 (1993).)

from the known elements and correcting for isotopic abundance) as shown in Figure 9. This has the advantage of allowing determination of elements not certified in the standard, but usually at the expense of reduced accuracy.

Simple element-for-element calibration against the sensitivities of individual elements of known concentration in an external standard has been employed.[11] For better accuracy and precision the standard has to have a similar structure and composition, i.e. the sample and standard are matrix-matched. Alternatively, a full set of well-characterized external standards may be used to obtain plots of response versus concentration. All laser and spectrometer parameters are kept constant for analysis of the samples and standards. Obvious difficulties include the non-availability of suitable standards and the expense of extensive use of standard reference material for calibration. In addition, standards are often inadequately characterized: elements of interest have not been determined or only ''guidance'' concentrations are given, or elements are certified at extremely low concentrations which do not provide sufficient counts for useful calibrations. Artificial standards made, for example, by adding aqueous standard solutions to a suitable matrix, drying, mixing, and pressing into pellets are sometimes successfully employed.

Internal calibration using responses obtained from elements in the sample itself may be used, requiring an independent determination of the concentration of a few elements in the sample.

Other approaches include external calibration on sensitivities obtained from the nebulization of standard aqueous solutions,[12] though this involves the loss of the advantage of working with a dry plasma, namely a reduction in spectral interferences. To overcome just such complications, internal standards[13] have recently been introduced to the outlet of an ablation cell by ETV. Similarly, Günther et al.[14] retained the advantages of calibration against solution samples by using a dual desolvation/LA system for the determination of rare earth elements (REEs) in soil, bestonite, and synthetic polyethylene, obtaining precisions of <10%.

Internal standardization is widely used in SN/ICP/MS to improve analytical precisions. In LA/ICP/MS its use is indispensable since differences in laser energy, surface texture, and so on, ensure that the ablated sample mass is not constant even for sequential analyzes. Normalizing the responses to that of a known element, usually a minor isotope of a major constituent, corrects for this type of variation. Between replicate internal standardization is easily performed because it does not require knowledge of the concentration of the element used. The procedure may be applied between the standard and the sample if the concentrations of an element are known in both.

4 APPLICATIONS IN ENVIRONMENTAL ELEMENTAL ANALYSIS

The specialty steels used in the construction of petroleum and petrochemical plants must be within the specifications for elemental composition if they are not to pose a safety hazard. Integrated electronics are now of such small dimensions that α-particle-induced damage is a serious problem, necessitating the use of aluminum wiring with U 7440-61-1 and Th 7440-29-1 impurities of $<1\,ng\,g^{-1}$. Knowledge of the concentration and distribution of elements in peridotites and their constituent minerals is essential to understanding the chemical evolution and composition of the lithospheric mantle. Such are

Table 1 Comparison of certified values with LA/ICP/MS data. All values in %. Samples: diamond steel rubbings[a]

Element	CRM 455		CRM 456		CRM 458		CRM 460		Sample A[b]	
	Certified value	LA/ICP/MS	Certified value	LA/ICP/MS	Certified value	LA/ICP/MS	Certified value	LA/ICP/MS	Certified value	LA/ICP/MS
B	–	–	0.0015	0.0018 ± 0.0004	0.0061	0.0065 ± 0.0004	0.0028	0.0029 ± 0.0004	–	–
Ti	0.022	0.026 ± 0.003	–	0.0036 ± 0.0003	–	0.010 ± 0.002	–	0.0034 ± 0.0005	0.005	0.0052 ± 0.0002
V	–	0.0066 ± 0.0004	0.022	0.023 ± 0.0005	0.108	0.110 ± 0.002	0.06	0.056 ± 0.003	0.01	0.008 ± 0.0004
Cr	0.21	0.206 ± 0.002	–	0.019 ± 0.0008	–	0.029 ± 0.001	–	0.091 ± 0.0009	1.25	1.26 ± 0.03
Mn	0.4	0.40 ± 0.04	0.2	0.15 ± 0.01	0.49	0.434 ± 0.019	0.67	0.71 ± 0.05	0.8	0.73 ± 0.06
Co	–	0.0090 ± 0.0004	0.052	0.051 ± 0.0008	0.21	0.211 ± 0.003	0.014	0.0146 ± 0.0006	0.01	0.008 ± 0.0003
Ni	0.35	0.37 ± 0.03	–	–	–	–	–	–	1.6	1.56 ± 0.05
Cu	0.038	0.050 ± 0.004	–	0.030 ± 0.003	–	0.056 ± 0.002	–	0.024 ± 0.001	0.05	0.04 ± 0.002
Sn	0.085	0.084 ± 0.006	–	0.0030 ± 0.0002	–	0.005 ± 0.0007	–	0.0027 ± 0.0001	0.005	0.003 ± 0.0003

[a] Each sample was analyzed on five different sites: site 1 was used to establish the calibration graph; sites 2–5 were treated as unknown samples. ^{57}Fe was used as internal standard.
[b] Sample A: sample of known concentration.
(Reproduced by permission of the Royal Society of Chemistry from A. Raith, R.C. Hutton, I.D. Abell, J. Crighton, *J. Anal. At. Spectrom*, **10**, 591–594 (1995).)

the challenges that LA/ICP/MS must meet. Illustrative applications are given below.

4.1 Metals

Metals are a logical starting point for study of the efficacy of LA/ICP/MS since they may be considered ''liquid solutions'' of uniform composition unless spatial resolution to the size of metal grains is considered. An example of the quantitative analysis of steels based on sensitivities obtained from a suite of steel standards is given in Table 1. As there are many characterized steels, this approach is more practicable than for other types of matrix. The sample was ablated using a 5×5 raster of points, covering a sample area of about $9\,mm^2$. Each crater was ablated five times with Q mode shots from a Nd : YAG laser at a repetition frequency of 4 Hz. Five such measurements were made per sample. Accuracies and precisions were typically $<10\%$ and DLs in the range $1-100\,\mu g\,g^{-1}$.

Additional examples of metals analysis are documented in Table 2.

4.2 Geological Materials

The difficulty of summarizing a large and rapidly growing literature is especially acute for geological applications. Several major themes, however, may be discerned: elemental analysis; fingerprinting; microprobe studies. Isotope ratio analysis, another active area, is dealt with separately in section 5.

Summaries of a number of studies of the determination of major, minor, and trace elements in matrices such as soils, silicates and carbonates are given in Table 3. Early studies using ruby lasers for ablation included the analysis of basalt[7] and Chinese reference soils.[11] A spectrum reproduced from the former shows the responses obtained for the REEs, Th and U (Figure 10). The peak integrals ranged from 150 to 4800 counts for the analyte isotopic concentrations of 0.16 to $5.8\,\mu g\,g^{-1}$. In the second study[11] the elements Na (7440-23-5), Mg (7439-95-4), Al (7429-90-5), Ca (7440-70-2), Sc (7440-20-2), Ti (7440-32-6), V (7440-32-6), Cr (7440-47-3), Mn (7439-96-5), Fe (7439-89-6), Co (7440-48-4), As , Rb (7440-17-7), Sb (7440-36-0), Cs (7440-46-2), Ba (7440-39-3), La (7439-91-0), Ce (7440-45-1), Nd (7440-00-8), Sm (7440-19-9), Eu (7440-53-1), Dy (7429-91-6), Ho (7440-60-0), Yb (7440-64-4), Hf , Ta , W , Th and U were determined in seven soils. Precisions were typically in the range 2–10% RSD, but some were considerably poorer for elements present at trace concentrations. DLs were as low as $0.1\,\mu g\,g^{-1}$ for many elements.

Often the objective of an analysis is identification of the source of the material and this does not require full quantitation. ''Fingerprinting'' of gold,[31] steel from safes, and glasses,[32] for example, is possible. The method relies on the identification of patterns of elemental responses as illustrated in Figure 11 for three separate analyses over three months of steel from five safes; four groups of elements are compared. These so-called *ternary plots* represent the direct comparison of the relationship between three components in a system. For data to plot in any corner of a triangle, representing the ternary association of the three components, would indicate that the concentration of that component indicated at the corner was 100% relative to the other two components.

While samples with different provenance will be discriminated in ternary plots by a suitable choice of analytes, samples with the same provenance will not be discriminated under any variation of analytes. The plots of Figure 11, for example, show a tight packed relation for safes 3 and 5 in all but the last (lower right), the latter indicating a different provenance. This fingerprinting technique has already proved a practical success in forensic science.

The ability to provide spatially-resolved analysis is one of the anticipated advantages ofLA/ICP-spectrometry, and has been markedly improved over the last ten years. Ablation by single laser shots, multiple shots on the same site, individual sites in a row or in a plane in principle allow the analysis of individual grains, the production of

Table 2 Applications in metals analysis

Matrix	Analytes/Comments	Laser details	Ref.
NIST SRM stainless steels 444, 445, 446, 447, 448, 450	Ni,Cr/LA/ICP/OES	XeCl excimer, 308 nm, 50 mJ/40 Hz	15
British Chemical Standard nickel alloy 346, NIST nickel alloys 897, 898, and 899	B, Al, Ti, Cr, Co, Zr, Mo	Ruby 0.1-J N mode 1 Hz	16
Al–Si–Cu	Sc, Cr, As, Sb, Hf, Th, U	Nd : YAG 250 mJ/10 Hz	17
NIST SRM 1162 low alloy steel	V, Cr, Co, Cu, Mo	Nd : YAG/10 Hz	18

NIST, National Institute of Standards and Technology.

Table 3 Applications in geological analysis

Matrix	Analytes/Comments	Laser details	Ref.
SARM 5 (Pyrexonite, South African Bureau of Standards)	Al, K, Na, Si, Ba, Mn, Sr, Ti LA/ICP/OES	Nd : YAG, 2-J N mode Single pulse	19
Silicate rocks, limestones	Si, Al, Ca, Mg, Fe, Ti, Na, K Lithium tetraborate fusions IS: Li and B	Nd : YAG, 355 nm, 10 mJ Q mode 10 Hz	20
NIST fly ash 1633a, NIST basalt 688	Mg, Al, Sc, Ti, V, Cr, Mn, Fe, Co, Cs, Ba,	Nd : YAG 1060 nm 200 mJ Q mode 10 Hz	21
Syenite SY-2, SY-3, serpentine UB-N	REEs	Nd : YAG 1060 nm 200 mJ Q mode 10 Hz	22
BCS 393 limestone	Mg, Mn, Sr, Ba	Nd : YAG 1064 nm	23
GSJ JLS-1	Mg, Mn, Sr	125 mJ Q mode 15 Hz	
IGGE GSR-6 limestone	Mg, Mn, Sr, Ba, Pb		
BCS-CRM 368 dolomite	Mn, Sr, Pb		
GSJ JDO-1 dolomite	Mn, Sr		
BCS-CRM 389 magnesite	Mn		
GXR-1, 3, 4, 5 and 6	Totalquant, multielemental	Nd : YAG	24
NIST SRM 1645 river sediment		1064 nm 50 mJ	
NIST SRM 1632a bituminous coal		Q mode 7 Hz	
Zeolite 1 + 1 fusion lithium carbonate–lithium metaborate mixture	Al, Si, Ti, Fe	Nd : YAG Q mode 7 Hz	25
Glass	IS ^{42}Ca; RSD 5–10%	10 Hz	
Synthetic graphite standards Anhydrite	Na, Mg, Al, P, K, V, Cr, Fe, Co, Ni, Cu, Zn, As, Se, Cd, Sb, Hg, Pb Standardized against response of NIST 612 glass	Nd : YAG 1064 nm Q mode 4 Hz	26
Titanium-based ores ilmenite and rutile	Mg, Al, Si, Ca, V, Cr, Mn, Nb, Pb, Th, U		
Ferromanganese crusts	Pt, REE Calibration against briquetted powder pellets of USGS NOD-A-1 and NOD-P-1 Precisons 4–5% for Pt; <4% for REE	Nd : YAG 1064 nm N mode 2 Hz	27
Soils BCR 143 sewage sludge Ammended Soil; BCR 146 sewage Sludge of mainly industrial origin	Cu, Cd, Cr, Ni, Pb, Zn, Mn	Nd : YAG 1064 nm 210 mJ N mode 10 Hz	28
Diamond, chromites, garnets	Ca, Ti, V, Co, Cu, Zn, Ga, Ge, Se, Rb, Sr, Y, Zr, Nb, Mo, Sn, Sb, Te, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Er, Tm, Yb, Lu, Hf, W, Ti, Pb, Bi, Th, U	Nd : YAG Diamond Q mode 6 Hz Garnet N mode Double pulse per site Chromite single pulse	29
Apatite, monazite, chromite, olivine	IS: Ca, Ce, Si or Mn Spatial resolution 60–150 μm	Nd : YAG 1064 nm Q mode	30
Monazite	Sr, Y, Zr, Nb, Ba, La, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Lu, Th, U, Ti, V, Mn, Co, Ni, Cu, Zn, Ga, Ge, Zr, Nb, Hf		
Chromite	Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Y, Zr		
Olivines	Nb		
Sy-2, AGV-1	REEs		

RSD, relative standard deviation.

depth profiles, line profiles, or area analyses, respectively. Of major importance to the analytical performance of laser microprobe/inductively coupled plasma/mass spectrometry (LM/ICP/MS) is the spatial resolution. Some of the early studies in which improved laser optics allowed the ablation crater diameter to be reduced from 150 μm to about 25 μm are briefly reviewed elsewhere.[33]

Moissette et al.[34] used pulses at 266 nm from a Nd : YAG laser operated at 10 Hz to ablate synthetic fluid inclusions in halite, microwells containing aqueous solutions, and glass reference materials. In this field, elemental intensities are reported as ratios to Na but owing to difficulties in measuring this element were reported ratioed to Sr (7440-24-6). Elemental ratios for Li (7439-93-2), Mg, K (7440-09-7), Ca, Mn, Cu (7440-50-8), Zn (7440-66-6), Rb, Cs, Ba, Pb (7439-92-1), B (7440-42-8), Cl (7782-50-5) and Br (7726-95-6) in the synthetic fluid inclusions were linear over several orders of

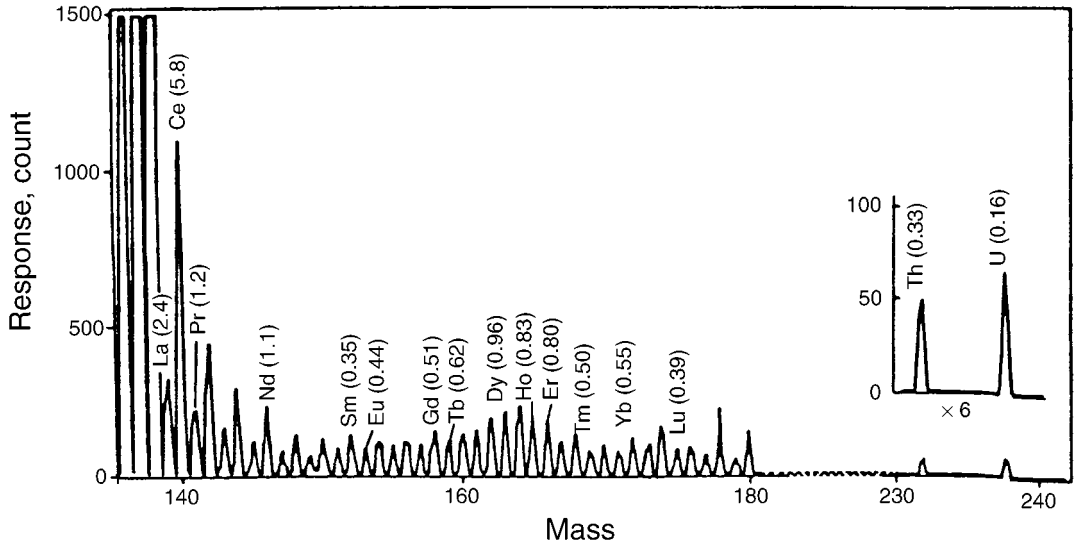

Figure 10 LA/ICP/MS spectrum of GSJ silicate standard sample JB-2. The concentrations of the isotopes determined are indicated in $\mu g\,g^{-1}$. The m/z region 181 to 230 was skipped. (Reproduced by permission of the Japan Society for Analytical Chemistry from Mochizuki et al.[7])

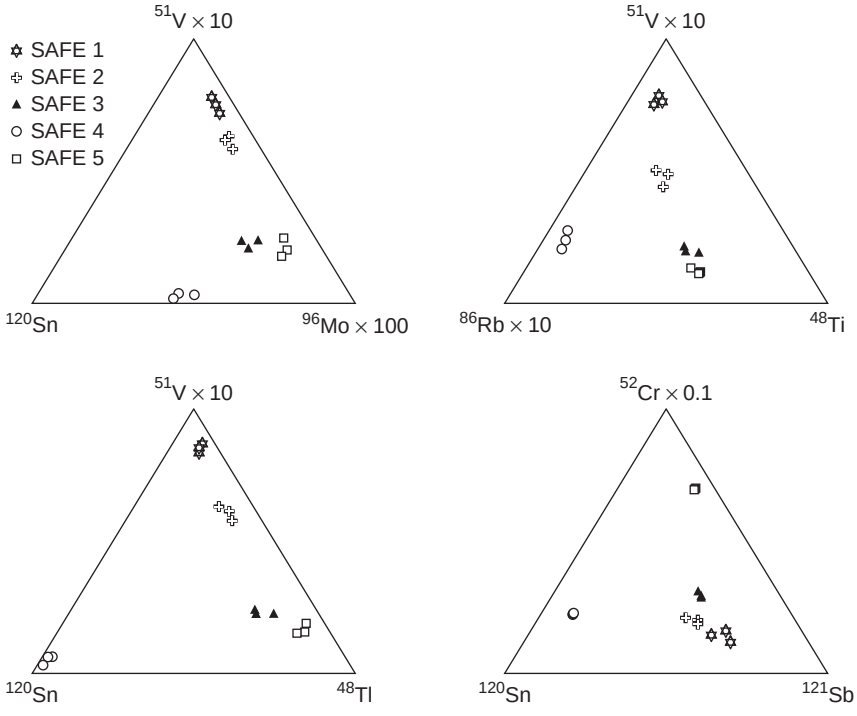

Figure 11 Ternary plots obtained by mapping relative elemental responses from LA/ICP/MS analysis of steel from 5 safes (three replicates). Four combinations of isotopes are plotted. See text for discussion. (Reproduced by permission of The Royal Society of Chemistry from R.J. Watling, B.F. Lynch, D. Herring, *J. Anal. At. Spectrom.*, **12**, 195–203 (1997).)

magnitude and in close agreement with those for the NIST SRM 611 glass and microwell solutions, irrespective of the inclusion size (20–100 μm) and depth in the sample (up to 80 μm). Thus microwells and glasses constitute suitable alternatives to synthetic fluid inclusions for the calibration and routine analysis of natural fluid inclusions. Similarly, concentrations for Sr in the range 90 to 153 $\mu g\,g^{-1}$ and Rb (98 to 100 $\mu g\,g^{-1}$) were determined in large (up to 1 mm diameter) fluid inclusions in halite[35] with precisions of 5–31%.

Further improvement in analytical performance is expected from the use of excimer lasers with wavelengths below 200 nm. The output from these devices gives extremely well-defined ablation pits and is also expected to produce less severe fractionation effects. Successful application of such instruments requires more sophisticated laser optics than used hitherto. Such a system, incorporating an argon fluoride 193 nm excimer laser, has been mounted, with a common UV/visual objective on a modified petrographic microscope with reflected and transmitted light illumination.[36] The improved beam quality allows the ablation of pits of only 4 μm diameter, giving DLs of $<10\,\mu g\,g^{-1}$ for selected elements between Rb and U.

4.3 Biological Materials

The author and colleagues reported the bulk multielemental analysis of botanical material[37] and foodstuffs.[38] To illustrate the sensitivity even of early ruby laser and quadrupole instruments, Figure 12 shows a spectrum obtained by ablating NIST 1577a bovine liver with 10 1.4-J N-mode shots. Mercury and lead were present at 4 and $135\,ng\,g^{-1}$, respectively. The isotopes of Hg (7439-97-6), Pb, and Bi (7440-69-9) are clearly visible.

LM analysis (variously abbreviated in the literature as LM, LAM, or LAMP) now offers more sophisticated possibilities. Outridge and Evans,[39] for example, outlined the difficulties encountered in analyzing teeth with ablation by pulses of wavelengths of 532 and 266 nm, finding that the analysis is wavelength, beam energy, and matrix dependent. More successful was a minor and trace element analysis across the growth bands of a shellfish *Arctica islandica*, with the resulting elemental profile being used as an indicator of heavy metal pollution through time. A spatial resolution of $<10\,\mu m$ was provided by the UV laser employed.[40]

The improved analytical performance of LM/ICP/MS over the last five years has apparently been little applied to biological analyses and is open for exploitation.

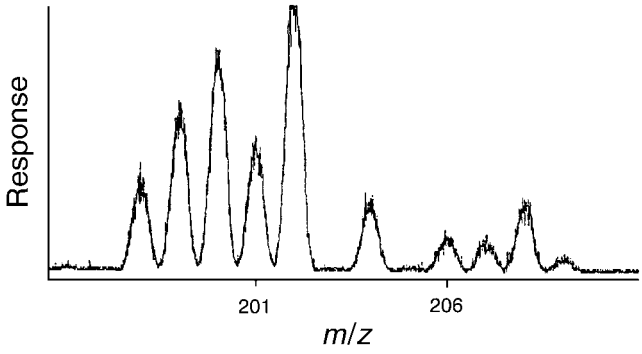

Figure 12 Spectrum obtained by LA/ICP/MS analysis of animal tissue. See text for details. The ^{201}Hg integral was 11 171 counts; the total certified Hg concentration was $4\,ng\,g^{-1}$. (Reproduced from Durrant.[8])

Table 4 summarizes key features of studies of biological materials.

4.4 Polymers

Analysis of plastics is illustrated by the early work of Marshall et al.[42] Successful analysis of polyester and polypropylene is demonstrated by the concentration data, reproduced in Table 5.

Improvement in analytical figures of merit can be expected with the use of the latest laser technology. This has been shown for the determination of Al, Ba, Ca, Cd (7440-43-9), Mg, Sb, Sn (7440-31-5) and Ti in poly(vinyl)chloride using LA/ICP/OES by Mermet's group,[43] where DLs obtained using a UV laser compare favorably with those obtained by SN/ICP/OES.

4.5 Other Matrices

Table 6 gives details of the analysis of diverse matrices by LA/ICP/OES and LA/ICP/MS. In nuclear applications where analysis of radioactive or toxic materials is common a complete LA/ICP/MS system may be sealed in a glove box. Laser sampling may readily be performed

Table 4 Applications in biological analysis

Matrix	Analytes/Comments	Laser details	Ref.
Chinese reference hair SINR 0920	Mg, Al, P, K, Ca, Cr, Mn, Ni, Cu, Zn, Sr, Ba, Hg, Pb	Ruby N 1 Hz 0.3 J (6 pulses)	38
NIST 8431a mixed diet	Na, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Mo	0.4 J (20 pulses)	
NIST 1549 Non-fat milk powder	Na, Al, P, Cl, K, Ca, Fe, Cu, Zn, Rb	1 J (10 pulses)	
Fin ray from a sturgeon	Ag, Ca, Hg Resolution 50 μm	Nd : YAG 70 mJ Q mode 1.5 Hz	18
Teeth from walruses and beluga Whales	Cd, Pb Resolution 200 μm	70 mJ Q mode 2 Hz	
Tree rings	Mg, Al, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Sr, Cd, Ba, Tl, Pb, Bi, U	Nd : YAG 200 mJ N mode 10 Hz	41

Table 5 Quantitative analysis of plastics by LA/ICP/MS

Element	Polyester		Polypropylene	
	LA/ICP/MS result/$\mu g\,g^{-1}$	Nominal value/$\mu g\,g^{-1}$	LA/ICP/MS result/$\mu g\,g^{-1}$	Nominal value/$\mu g\,g^{-1}$
Al	357	350	105	100
Si	720	770	600	750
P	93	105	60	45
Co	32	37	7	NA[a]
Zn	45	50	205	200
Sb	155	170	ND[b]	ND[b]

[a] NA = not available.
[b] ND = not detectable.
(Reproduced by permission of the Royal Society of Chemistry from Marshall et al.[42])

Table 6 Miscellaneous applications

Matrix	Analytes/Comments	Laser details	Ref.
Lithium and potassium chlorides	U, La, Cr, Nd, Y/ LA/ICP/OES DL (U) $25\,\mu g\,g^{-1}$	Nd : YAG 1064, 532, 355 nm Q mode	44
ZrN, TiN and ZrTiN ceramic films (3 μm thick) on steel	Zr, Ti, N/ LA/ICP/OES	Nd : YAG 355 nm Q mode 5 mJ	45
$ZrO_2-Y_2O_3$/NiCrAlY/steel	Depth profiling/LA/ICP/OES	Nd : YAG 266 nm	46
ZrO_2-CeO_2/NiCrAlY/steel	Zr, Y, Ce, Ni, Cr, Al, Fe	Q mode 5 mJ 10 Hz	
Concrete	Quadrupole and HR ^{99}Tc, ^{232}Th, ^{233}U, ^{237}Np DL $10\,ng\,g^{-1}$ (Quadrupole) $1\,ng\,g^{-1}$ or better (HR)	Nd : YAG 15 mJ 10–20 Hz	47
Uranium	Si, Cr, Fe, Co, Cu, Mo, Sn, Pb DLs $<0.1\,\mu g\,g^{-1}$	XeCl excimer 308 nm 100 mJ 5 Hz	48
Ceramic $La_{0.65}Sr_{0.3}MnO_3$ Yttria-stabilized ZrO_2	Li, B, Mg, Al, Si, P, K, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Sr, Nb, Mo, Ag, Cd, In, Ba, La, Ce, Ho, Re, Tl, Pb, Bi	Nd : YAG 266 nm Q mode 10/20 Hz	49
Airborne particulates	Mg, Si, Ca, Cr, Mn, Ni, Cu, Sr, Cd, Pb	Nd : YAG 1060 nm 200 mJ 10 Hz	50
Airborne particulates	As	Nd : YAG N-mode single shot 140 mJ	51
Airborne particulates	Na, Al, K, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Ag, Cd, Sb, Ce, W, Pb	Nd : YAG 1064 nm Q mode 90 mJ	52

far from the spectrometer since extending the laser cell to spectrometer tubing produces very little reduction in analyte responses.

5 ISOTOPE RATIO ANALYSIS

Isotope ratios may readily be determined by SN/ICP/MS. Factors of importance in such determinations[53] include

- sensitivity and counting statistics
- dead time
- resolution and abundance sensitivity and
- mass bias.

Thus count rates well above background are required, but these must not be so high that the detector is overwhelmed. However, to a degree, a dead-time correction can be applied at the onset of saturation, thus extending the working concentration range. Peaks nominally one mass unit apart must be clearly resolved, but resolution improves at the cost of sensitivity, so over-resolution is to be avoided. In practice tailing of one peak into an adjacent peak also occurs and sufficient abundance sensitivity, as it is known, to resolve a small peak from an adjacent intense one is necessary. In addition, the measured isotope ratio may deviate from the true value as a function of the difference in mass between the two isotopes measured. This effect, known as mass bias, may

be measured by comparison with a standard of known isotopic composition.

Owing to these considerations and to the ICP being a fairly noisy source, the best precisions achievable by SN/ICP/MS are typically 0.1% RSD. Unfortunately, the addition of signal fluctuations associated with sample introduction by LA leads to even poorer precisions, rending useful isotopic ratio measurements by LA/ICP/MS extremely difficult.

In 1987, Tye et al.[54] reported the measurement of the ratios $U^{234}:U^{235}$, $U^{234}:U^{236}$ and $U^{235}:U^{236}$ in U_3O_8 with percentage RSDs of 1.340, 0.774 and 1.020 respectively.

Subsequent improvement in laser optics allowed the determination of $^{207}Pb:^{206}Pb$ ratios and hence the calculation of ages of zircons with a grain size of $>60\,\mu m$ and a ^{207}Pb concentration of $>3\,\mu g\,g^{-1}$, to within 1% of the ages determined from thermal ionization mass spectrometry (TIMS) measurements.[55] Precisions for the average of the mean $^{207}Pb:^{206}Pb$ ratios from different pits in one grain were generally <1.5%. Although U : Pb ratios could not be accurately or precisely determined by LM/ICP/MS, its performance was otherwise comparable to the specialized technique usually used for such measurements, namely the sensitive high-resolution ion microprobe (SHRIMP).

Relatively poor precision is the stumbling block to the wider useful application of LA/ICP/MS to isotope ratio measurements but is overcome by the use of more sophisticated high-resolution spectrometers. In magnetic sector instruments the sampling interface is held at a high potential, e.g. +8 kV, and the quadrupole analyzer is replaced by an electrostatic and a magnetic analyzer, which transmit ions by kinetic energy and m/z, respectively. Following this double focusing the transmitted ions are detected by a Faraday cup (or cups). An up-to-date, simple and concise introduction to such spectrometers is available.[56]

Use of LA with a magnetic sector mass analyzer equipped with seven Faraday detectors[57] allows the determination of Pb isotope ratios with precisions comparable to those obtained by TIMS.[58] In this analysis the isotopic composition of lead within NIST SRM 610 glass was measured using Q mode pulses of 2 mJ from a Nd : YAG laser for sample ablation. Craters of $<40\,\mu m$ diameter were produced, corresponding to a total ablated mass of Pb of about 9×10^{-11} g per crater. Simultaneous measurement of ^{202}Hg, ^{203}Tl, ^{204}Hg and ^{204}Pb, ^{205}Tl, ^{206}Pb, ^{207}Pb, and ^{208}Pb isotopes was made. As the ratio $^{205}Tl : {}^{203}Tl$ had previously been shown to be related to the mass bias, this ratio was measured and used to correct the measured lead isotope ratios. The response of an isotope of mercury, ^{204}Hg, overlapped with that of ^{204}Pb but was corrected by measurement of the ^{202}Hg isotope. Table 7 shows the lead isotope ratios for repeat analyses, together with the standard error. The mean and standard deviation of each set is also given and compared with data obtained by TIMS and, where available, by secondary ionization mass spectrometry (SIMS). The important feature of the data is the precision, which is comparable to that obtained by the traditional methods of choice for such analyses.

Although not illustrated by the above analysis, high-resolution spectrometers are often also able to resolve out interfering polyatomic species. For example, a resolution of 2500, defined as the mass M divided by the mass width ΔM, resolves $^{56}Fe^+$ from $^{40}Ar^{16}O^+$, thereby allowing determination of iron at m/z 56.

Another innovative approach to isotope ratio determinations was reported by Houk et al.[59] using a double quadrupole instrument. The rationale for such an instrument is to eliminate flicker noise from the LA process by splitting the ion beam into two, each part being sent to its own quadrupole mass analyzer and detector. The sample was ablated with pulses from a Nd : YAG laser frequency doubled to operate at 532 nm and operating at

Table 7 Lead isotope ratio measurement of NIST SRM 610 glass by LA/ICP/MS

Date	$^{208}Pb:^{204}Pb$ ±ISE	$^{207}Pb:^{204}Pb$ ±ISE	$^{206}Pb:^{204}Pb$ ±ISE	$^{208}Pb:^{206}Pb$ ±ISE	$^{207}Pb:^{206}Pb$ ±ISE	No. of craters
07/05/92	36.942 ± 0.030	15.516 ± 0.011	17.061 ± 0.016	2.1653 ± 0.0009	0.9094 ± 0.0004	10
08/05/92	36.948 ± 0.038	15.498 ± 0.015	17.050 ± 0.017	2.1671 ± 0.0017	0.9090 ± 0.0004	9
11/05/92	36.941 ± 0.043	15.516 ± 0.017	17.048 ± 0.016	2.1669 ± 0.0011	0.9101 ± 0.0004	14
11/05/92	36.944 ± 0.017	15.508 ± 0.007	17.045 ± 0.007	2.1674 ± 0.0002	0.9098 ± 0.0001	12
12/05/92	36.985 ± 0.032	15.518 ± 0.010	17.060 ± 0.009	2.1679 ± 0.0009	0.9097 ± 0.0002	13
12/05/92	36.929 ± 0.015	15.499 ± 0.006	17.041 ± 0.007	2.1671 ± 0.0005	0.9095 ± 0.0001	12
Mean ± 2SD	36.948 ± 0.038	15.509 ± 0.018	17.051 ± 0.016	2.1670 ± 0.0018	0.9096 ± 0.0008	
TIMS ± 2SD[7]	36.989 ± 0.024	15.506 ± 0.010	17.049 ± 0.012	2.170 ± 0.002	0.9095 ± 0.001	
SIMS ± 2SD[7]	–	–	–	2.160 ± 0.006	0.9056 ± 0.003	

Reproduced from Walder et al., 'Lead Isotope Ratio Measurement of NIST 610 Glass by LA/ICP/MS', *Spectrochim. Acta*, **48B**(3), 397–402 (1993) with permission from Elsevier Science.

15 Hz and about 70 mJ per pulse. Count rates obtained simultaneously at m/z 52 and 53 from the ablation of NIST SRM 1263 showed an RSD of 3%, but for the ratio of the two signals the RSD improved to 0.54%. When five consecutive ratios were averaged, the average of five such averaged ratios was 0.24%. The latter was improved to 0.058% by increasing the dwell time (the time spent in accumulating data) from 0.5 to 2.0 s. Such precisions, of ratios of isotopes of one element, are only slightly above the counting (Poisson) statistics limit.

6 PERFORMANCE AND STATUS

As we have noted, the anticipated advantages of LA/ICP/MS and LA/ICP/OES include little sample preparation, rapid sample throughput, and excellent DLs.

Occasionally no sample preparation is needed and this represents an enormous saving in the workload of a skilled analyst. Frequently, however, some sample manipulation is required for maximum analytical performance. Throughput, i.e. the number of samples analyzed per working day, is excellent, being at worst equal to that of SN/ICP/MS. The DLs achieved are element and matrix specific but are now $<1\ \mu g\ g^{-1}$ for ablated crater diameters of only 20–40 μm.[60]

Another anticipated advantage of LA/ICP/MS with a quadrupole spectrometer is a significant reduction in spectral interferences owing to a reduction in the water loading of the ICP. This reduces water-derived polyatomic species present below mass 80 in SN/ICP/MS analyses as indicated by the data of Mochizuki et al.[7]

Table 8 Comparison of background intensities obtained by LA and solution introduction methods

Mass	Molecule	Analyte isotope (abundance, %)	Ratio of peak integrals (laser/solution[a])
28	$^{14}N_2^+$	Si(92.18)	0.67
29	$^{14}N_2H^+$	Si(4.71)	0.10
30	$^{14}N^{16}O^+$	Si(3.12)	0.16
31	$^{15}N^{16}O^+$, $^{14}N^{16}OH^+$	P(100)	0.22
32	$^{16}O_2^+$	S(95.02)	0.01
33	$^{16}O_2H^+$	S(0.75)	0.00_1
34	$^{16}O^{18}O^+$	S(4.22)	0.01
54	$^{40}Ar^{14}N^+$	Fe(5.90), Cr(2.38)	0.35
56	$^{40}Ar^{16}O^+$	Fe(91.52)	0.01
57	$^{40}Ar^{16}OH^+$	Fe(2.25)	0.11
76	$^{36}Ar^{40}Ar^+$	Ge(7.76), Se(9.12)	1.05
80	$^{40}Ar_2^+$	Se(49.96), Kr(2.27)	1.07

[a] Ion intensity for solution technique was obtained by nebulizing distilled water. (Reproduced by permission of the Japan Society for Analytical Chemistry from T. Mochizuki et al.[7])

presented in Table 8. (The Ar dimers are not water-related and levels are essentially the same in both SN/ICP/MS and LA/ICP/MS).

The linear dynamic range of ICP spectrometry is so familiar to users that it is perhaps overlooked. When pulse counting and current detection are both used, as they are in some ICP/MS instruments, for low and high count rates, respectively, the linear dynamic range extends over seven orders of magnitude or more.

Difficulties in obtaining fully quantitative concentration data have been apparent since Gray's original feasibility study.[5] Ingenious methods of calibration and internal standardization have been developed, and improved precisions are obtained by the use of radiation in the UV or far UV instead of the IR. Figures of merit are further improved by the use of magnetic sector instruments, which are now in much wider use than in the early years following their introduction in 1988.

One contribution to inaccuracy and imprecision, however, has not been fully overcome, namely chemical fractionation. Some studies of this effect have been undertaken. For example, elements such as Nb (7440-03-1), Ta, Zr (2440-67-7), and Hf show decreased fractionation trends with respect to Si in NIST SRM 610 glass during repeated LA, while elements such as Rb, Sr, Pb, Ba, and Ca show increasing trends. The effect, greater for IR than for UV/LA,[61] may be due in part to the formation of a complex plasma which both ablates the sample and shields it from incoming laser light. In a comparison of the ablation of glass by pulses of wavelength 1064 nm (IR), 532 nm (green) and 266 nm (UV), elemental fractionation was dependent on the melting point of the elemental oxide only for the first two wavelengths.[62]

Relatively little fractionation may occur with ablation by excimer lasers,[63] though the evidence is not definitive and more studies are needed. Indeed, an important observation by Outridge et al.[64] is that fractionation may occur not only in ablation but also in the transport of material to the ICP. Thus many of the previous studies of this problem have concerned only the product of these two effects and are therefore somewhat inadequate.

Finally, LA/ICP/OES deserves mention. Generally, this technique is less sensitive than ICP/MS and does not offer isotope ratio measurements. Against this should be set its greater simplicity of operation and maintenance and lower capital cost. Moreover, ICP/OES is also less prone to matrix effects than ICP/MS. In short, LA/ICP/OES is a very powerful technique and has received relatively little attention here only because of the *relative* strength of LA/ICP/MS. It should also be noted that most of the comments given concerning LA/ICP/MS also apply to LA/ICP/OES. For example, fractionation leads to a sample unrepresentative of the

original material and this influences either technique. In contrast, polyatomic interferences are a problem specific to ICP/MS.

7 DEVELOPMENT PROSPECTS

Means to improve analytical performance in LA/ICP/MS will continue to be developed. The wider use of mixed-gas plasmas, for example, would be advantageous in LA/ICP/MS to increase sensitivities and reduce relative oxide levels.[65] A new soft ablation technique[66] in which a slowly ramped increase in laser power produces a smoother response profile has given Pb : U age data in excellent agreement with those obtained by TIMS and SIMS. The improvement in performance, attributed to reduced fractionation effects, clearly merits further study.

Perhaps the most exciting foreseeable developments, however, lie in alternative spectrometers. We have already met double quadrupole instruments, but these are likely to remain a research tool in selected laboratories only. Other candidates include Fourier transform ion cyclotron resonance (FTICR), ion trap (IT), and time-of-flight (TOF) instruments. The possible use of such systems is outlined by Hieftje et al.,[67] who also indicate why inductively coupled plasma/time-of-flight/mass spectrometry (ICP/TOF/MS) may be the most promising among them. A prototype instrument has already been used in conjunction with LA.[68] Indeed, this technique has recently been applied to the analysis of metals.[69] Here we may note that the advantages of ICP/TOF/MS include high ion transport efficiency, high sensitivity, and extremely fast ($<60\,\mu s$) spectral collection across the whole mass range.

The reader is also directed to other exciting applications of LA/ICP/MS indicated by recent literature, namely biological speciation[70] and archeological studies.[71] As we have seen, successful applications of LA/ICP/MS and LA/ICP/OES abound in metallurgy, geology, materials science, pollution studies, nuclear science, and so on; both techniques deserve greater attention from botanists, biologists, and medical scientists.

Concerning the development of new instrumentation, alternative analytical procedures, and novel applications, the next decade promises to be even more exciting than the last.

ABBREVIATIONS AND ACRONYMS

AES	Atomic Emission Spectrometry
DL	Detection Limit
ETV	Electrothermal Vaporization
FTICR	Fourier Transform Ion Cyclotron Resonance
ICP	Inductively Coupled Plasma
ICP/AES	Inductively Coupled Plasma/Atomic Emission Spectrometry
ICP/MS	Inductively Coupled Plasma/Mass Spectrometry
ICP/OES	Inductively Coupled Plasma/Optical Emission Spectrometry
ICP/TOF/MS	Inductively Coupled Plasma/Time-of-Flight/Mass Spectrometry
IR	Infrared
IT	Ion Trap
LA	Laser Ablation
LA/ICP/MS	Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry
LA/ICP/OES	Laser Ablation/Inductively Coupled Plasma/Optical Emission Spectrometry
LM	Laser Microprobe
LM/ICP/MS	Laser Microprobe/Inductively Coupled Plasma/Mass Spectrometry
MS	Mass Spectrometry
Nd : YAG	Neodymium:Yttrium Aluminum Garnet
NIST	National Institute of Standards and Technology
OES	Optical Emission Spectrometry
PTFE	Poly(tetrafluoroethylene)
REE	Rare Earth Element
RSD	Relative Standard Deviation
SA	Spark Ablation
SHRIMP	Sensitive High-resolution Ion Microprobe
SIMS	Secondary Ionization Mass Spectrometry
SN/ICP/MS	Solution Nebulization/Inductively Coupled Plasma/Mass Spectrometry
TIMS	Thermal Ionization Mass Spectrometry
TOF	Time-of-flight
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Atomic Spectroscopy **(Volume 11)**
Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Ablation in Atomic Spectroscopy • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

REFERENCES

1. M. Thompson, J.N. Walsh (eds.), *Handbook of Inductively Coupled Plasma Spectroscopy*, Chapman & Hall, New York, 1989.
2. R.S. Houk, V.A. Fassel, G.D. Flesch, H. Svec, A.L. Gray, C.E. Taylor, 'Inductively Coupled Argon Plasma as an Ion Source for Mass Spectrometric Determination of Trace Elements', *Anal. Chem.*, **52**, 2283–2289 (1980).
3. A.L. Gray, 'The ICP as an Ion Source – Origins, Achievements and Prospects', *Spectrochim. Acta*, **40B**(10–12), 1525–1537 (1985).
4. T.H. Maiman, 'Stimulated Optical Radiation in Ruby', *Nature*, **187**, 493–494 (1960).
5. A.L. Gray, 'Solid Sample Introduction by Laser Ablation for Inductively Coupled Plasma Source Mass Spectrometry', *Analyst*, **110**, 551–556 (1985).
6. P. Arrowsmith, 'Laser Ablation of Solids for Elemental Analysis by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **59**, 1437–1444 (1987).
7. T. Mochizuki, A. Sakashita, H. Iwata, T. Kagaya, T. Shimamura, P. Blair, 'Laser Ablation for Direct Elemental Analysis of Solid Samples by Inductively Coupled Plasma Mass Spectrometry', *Anal. Sci.*, **4**, 403–409 (1988).
8. S.F. Durrant, 'The Application of Laser Ablation for Solid Sample Introduction to Inductively Coupled Plasma-source Mass Spectrometry', PhD Thesis, University of Surrey, Guildford, Surrey, 1989.
9. L. Moenke-Blankenburg, 'Laser Micro Analysis', *Prog. Anal. Spectrosc.*, **9**, 335–427 (1986).
10. I.D. Abell, 'Performance Benefits of Optimisation of Laser Ablation Sampling for ICP/MS', *Applications of Plasma Source Mass Spectrometry*, eds. G. Holland, A.N. Eaton, The Royal Society of Chemistry, Cambridge, 209–217, 1990.
11. S.F. Durrant, N.I. Ward, 'Rapid Multielemental Analysis of Chinese Reference Soils by Laser Ablation Inductively Coupled Plasma-source Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **345**, 512–517 (1993).
12. E.F. Cromwell, P. Arrowsmith, 'Semiquantitative Analysis with Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**(1), 131–138 (1995).
13. W. Doherty, P.M. Outridge, D.C. Grégoire, 'Technique for the Introduction of Dry Atomic Vapours for Improved Optimization and Diagnostic Studies of Laser Ablation Inductively Coupled Plasma Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1123–1126 (1996).
14. D. Günther, H. Cousin, B. Magyar, I. Leopold, 'Calibration on Dried Aerosols for Laser Ablation/Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 165–170 (1997).
15. M.E. Tremblay, B.W. Smith, M.B. Leong, J.D. Winefordner, 'Laser Ablation for the Introduction of Solid Metals into an Inductively Coupled Plasma', *Spectroscopy Lett.*, **20**(4), 311–318 (1987).
16. S.F. Durrant, 'Inorganic Analysis of Solids by Laser Ablation Inductively Coupled Plasma-source Mass Spectrometry', *Revista Brasileira de Aplicações de Vácuo*, **10**(1), 39–43 (1991).
17. P. van de Weijer, P.J.M.G. Vullings, W.L.M. Baeten, W.J.M. de Laat, 'Determination of Uranium and Thorium in Aluminum with Flow Injection and Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **6**(8), 609–614 (1991).
18. P. Richner, D. Evans, C. Wahrenberger, V. Dietrich, 'Application of Laser Ablation and Electrothermal Vaporization as Sample Introduction Techniques for ICP/MS', *Fresenius' J. Anal. Chem.*, **350**, 235–241 (1994).
19. S.A. Darke, S.E. Long, C.J. Pickford, 'Laser Ablation System for Solid Sample Introduction into the Inductively Coupled Plasma', *Anal. Proc.*, **26**, 159–160 (1989).
20. V. Kanicky, J.-M. Mermet, 'Selection of Internal Standards for the Determination of Major and Minor Trace Elements in Silicate Rocks and Limestones by Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry', *Appl. Spectrosc.*, **51**(3), 332–336 (1997).
21. A.A. van Heuzen, 'Analysis of Solids by Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LA/ICP/MS)-I. Matching with a Glass Matrix', *Spectrochim. Acta*, **46B**(14), 1803–1817 (1991).
22. A.A. van Heuzen, J.B.W. Morsink, 'Analysis of Solids by Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LA/ICP/MS)-II. Matching with a Pressed Pellet', *Spectrochim. Acta*, **46B**(14), 1819–1828 (1991).
23. W.T. Perkins, R. Fuge, N.J.G. Pearce, 'Quantitative Analysis of Trace Elements in Carbonates Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **6**, 445–449 (1991).
24. E.R. Denoyer, 'Analysis of Powdered Materials by Laser Sampling ICP/MS', in *Applications of Plasma Source Mass Spectrometry*, eds. G. Holland, A.N. Eaton, Royal Society of Chemistry, Cambridge, 199–208, 1991.
25. J. Franks, J. Marshall, I. Brown, L. Garden, 'Analysis of Inorganic Solids by Laser Ablation Inductively Coupled Plasma Spectrometry' *Anal. Proc.*, **29**, 23–24 (1992).
26. A. Raith, R.C. Hutton, 'Quantitation Methods Using Laser Ablation ICP/MS Part 1: Analysis of Powders', *Fresenius' J. Anal. Chem.*, **350**, 242–246 (1994).
27. C.-D. Garbe-Schönberg, G.M. McMurtry, 'In-situ Microanalysis of Platinum and Rare Earths in Ferromanganese

Crusts by Laser Ablation ICP/MS (LAICPMS)', *Fresenius' J. Anal. Chem.*, **350**, 264–271 (1994).
28. H. Scholze, H. Stephanowitz, E. Hoffmann, J. Skole, 'Element Analysis of Inhomogeneous Samples by Laser Ablation ICP Mass Spectrometry' *Fresenius' J. Anal. Chem.*, **350**, 247–252 (1994).
29. R.J. Watling, H.K. Herbert, I.S. Barrow, A.G. Thomas, 'Analysis of Diamonds and Indicator Minerals for Diamond by Laser Ablation/Inductively Coupled Plasma Mass Spectrometry', *Analyst*, **120**, 1357–1364 (1995).
30. Z. Chen, W. Doherty, D.C. Grégoire, 'Application of Laser Sampling Microprobe Inductively Coupled Plasma Mass Spectrometry to the In Situ Trace Element Analysis of Selected Geological Materials', *J. Anal. At. Spectrom.*, **12**, 653–659 (1997).
31. R.J. Watling, H.K. Herbert, D. Delev, I.D. Abell, 'Gold Fingerprinting by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta*, **49B**, 205–219 (1994).
32. R.J. Watling, B.F. Lynch, D. Herring, 'Use of Laser Ablation Inductively Coupled Plasma Mass Spectrometry for Fingerprinting Scene of Crime Evidence', *J. Anal. Atom. Spectrom.*, **12**, 195–203 (1997).
33. S.F. Durrant, 'Laser Ablation Inductively Coupled Plasma Mass Spectrometry', in *Surfaces, Vacuum, and their Applications*, eds. I. Hernández-Calderón, R. Asomoza, AIP Conference Proceedings, **378**, AIP Press, Woodbury, New York, 94–97, 1996.
34. A. Moissette, T.J. Shepherd, S.R. Chenery, 'Calibration Strategies for the Elemental Analysis of Individual Aqueous Fluid Inclusions by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 177–185 (1996).
35. A.H. Ghazi, T.E. McCandless, D.A. Vanko, J. Ruiz, 'New Quantitative Approach in Trace Elemental Analysis of Single Fluid Inclusions: Application of Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LAICPMS)', *J. Anal. At. Spectrom.*, **11**, 667–674 (1996).
36. D. Günther, R. Frischknecht, C.A. Heinrich, H.-J. Kalhert, 'Capabilities of an Argon Fluoride 193 nm Excimer Laser for Laser Ablation Inductively Coupled Plasma Mass Spectrometry Microanalysis of Geological Materials', *J. Anal. At. Spectrom.*, **12**, 939–944 (1997).
37. N.I. Ward, S.F. Durrant, A.L. Gray, 'Analysis of Biological Standard Reference Materials by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal At. Spectrom.*, **7**(7), 1139–1146 (1992).
38. S.F. Durrant, N.I. Ward, 'Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LAICPMS) for the Multielemental Analysis of Biological Materials: A Feasibility Study', *Food Chemistry*, **49**, 317–323 (1994).
39. P.M. Outridge, R.D. Evans, 'Effect of Laser Parameters and Tooth Type on the Ablation of Trace Metals from Mammalian Teeth', *J. Anal. Atom. Spectrom.*, **10**, 595–600 (1995).
40. A. Raith, W.T. Perkins, N.J.G. Pearce, T.E. Jeffries, 'Environmental Monitoring on Shellfish Using Laser Ablation ICP-MS', *Fresenius' J. Anal. Chem.*, **355**, 789–792 (1996).
41. E. Hoffmann, C. Lüdke, H. Scholze, H. Stephanowitz, 'Analytical Investigations of Tree Rings by Laser Ablation ICP-MS', *Fresenius' J. Anal. Chem.*, **350**, 253–259 (1994).
42. J. Marshall, J. Franks, I. Abell, C. Tye, 'Determination of Trace Elements in Solid Plastic Materials by Laser Ablation/Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **6**, 145–150 (1991).
43. M. Hemmerlin, J.M. Mermet, M. Bertucci, P. Zydowicz, 'Determination of Additives in PVC Material by UV Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry', *Spectrochim. Acta*, **42B**, 421–430 (1997).
44. P. Goodall, S.G. Johnson, 'Laser Ablation/Inductively Coupled Plasma Atomic Emission Spectrometry for the Determination of Lanthanides and Uranium in Fuel Reconditioning Materials: Problems, Solutions and Implications', *J. Anal. At. Spectrom.*, **11**(7), 469–478 (1996).
45. V. Kanicky, J. Musil, J.-M. Mermet, 'Determination of Zr and Ti in 3-μm-Thick ZrTiN Ceramic Coating Using Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry', *Appl. Spectrosc.*, **51**(7), 1037–1041 (1997).
46. V. Kanicky, I. Novotny, J. Musil, J.-M. Mermet, 'Depth Profiling of Thick Layers of Graded Metal-Zirconia Ceramic Coatings Using Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry', *Appl. Spectrosc.*, **51**(7), 1042–1046 (1997).
47. M. Gastel, J.S. Becker, G. Küppers, H.-J. Dietze, 'Determination of Long-lived Radionuclides in Concrete Matrix by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta*, **52B**, 2051–2059 (1997).
48. C. Leloup, P. Marty, D. Dall'Ava, M. Perdereau, 'Quantitative Analysis of Impurities in Uranium by Laser Ablation Inductively Coupled Plasma Mass Spectrometry: Improvements in the Experimental Setup', *J. Anal. At. Spectrom.*, **12**, 945–950 (1997).
49. J. Th. Westheide, J. Sabine Becker, R. Jäger, H.J. Dietze, J.A.C. Broekaert, 'Analysis of Ceramic Layers for Solid Oxide Fuel Cells by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 661–666 (1996).
50. C. Lüdke, E. Hoffmann, J. Skole, 'Comparative Studies on Metal Determination in Airborne Particulates by LA-ICP-MS and Furnace Atomization Non-Thermal Excitation Spectrometry', *Fresenius' J. Anal. Chem.*, **350**, 272–276 (1994).
51. C.F. Wang, S.-L. Jeng, F.-J. Shieh, 'Determination of Arsenic in Airborne Particulate Matter by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 61–67 (1997).

52. S. Tanaka, N. Yashushi, N. Sato, T. Fukasawa, S.J. Santosa, K. Yamanaka, T. Ootoshi, 'Rapid and Simultaneous Multi-element Analysis of Atmospheric Particulate Matter Using Inductively Coupled Plasma Mass Spectrometry with Laser Ablation Sample Introduction', *J. Anal. At. Spectrom.*, **13**, 135–140 (1998).
53. K.E. Jarvis, A.L. Gray, A.L. Houk, *Handbook of Inductively Coupled Plasma Mass Spectrometry*, Blackie, Glasgow, Chapter 11, 1992.
54. C. Tye, J. Gordon, P. Webb, 'Analysis of Uranium Using Laser Ablation ICP-MS', *International Laboratory*, **17**(10), 34–41 (1987).
55. R. Feng, N. Machado, J. Ludden, 'Lead Geochronology by Laser Probe/Inductively Coupled Plasma Mass Spectrometry (LP-ICP-MS)', *Geochim. Cosmochim. Acta*, **57**(14), 3479–3486 (1993).
56. L. Moens, N. Jakubowski, 'Double-focusing Mass Spectrometers in ICP-MS', *Anal. Chem.*, April 1, 251A–256A (1998).
57. A.J. Walder, P.A. Freedman, 'Isotopic Ratio Measurement Using a Double Focusing Magnetic Sector Mass Analyzer with an Inductively Coupled Plasma as an Ion Source', *J. Anal. At. Spectrom.*, **37**(3), 571–575 (1992).
58. A.J. Walder, I.D. Abell, I. Platzner, P.A. Freeman, 'Lead Isotope Ratio Measurement of NIST 610 Glass by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta*, **48B**(3), 397–402 (1993).
59. L.A. Allen, J.J. Leach, H.-M. Pang, R.S. Houk, 'Precise Measurement of Ion Ratios in Solid Samples Using Laser Ablation with a Twin Quadrupole Inductively Coupled Plasma Spectrometer', *J. Anal. At. Spectrom.*, **12**, 171–176 (1997).
60. T.E. Jeffries, W.T. Perkins, N.J.G. Pearce, 'Comparisons of Infrared and Ultraviolet Laser Probe Microanalysis Inductively Coupled Plasma Mass Spectrometry in Mineral Analysis', *Analyst*, **120**, 1365–1371 (1995).
61. T.E. Jeffries, N.J.G. Pearce, W.T. Perkins, A. Raith, 'Chemical Fractionation During Infrared and Ultraviolet Laser Ablation Inductively Coupled Plasma Mass Spectrometry – Implications for Mineral Microanalysis', *Analytical Communications*, **33**, 35–39 (1996).
62. D. Figg, M.S. Kahr, 'Elemental Fractionation of Glass Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **51**(8), 1185–1192 (1997).
63. F. Bea, P. Montero, A. Stroh, J. Baasner, 'Microanalysis of Minerals by an Excimer UV-LA-ICP-MS System', *Chem. Geol.*, **133**, 145–156 (1996).
64. P.M. Outridge, W. Doherty, D.C. Grégoire, 'Ablative and Transport Fractionation of Trace Elements During Laser Sampling of Glass and Copper', *Spectrochim. Acta*, **52B**, 2093–2102 (1997).
65. S.F. Durrant, 'Feasibility of Improvement in Analytical Performance in Laser Ablation Inductively Coupled Plasma/Mass Spectrometry (LA-ICP-MS) by Addition of Nitrogen to the Argon Plasma', *Fresenius' J. Anal. Chem.*, **349**, 768–771 (1994).
66. T. Hirata, 'Soft Ablation Technique for Laser Ablation/Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**(11), 1337–1342 (1997).
67. G.M. Hieftje, D.P. Myers, G. Li, P.P. Mahoney, T.W. Burgoyne, S.J. Ray, J.P. Guzowski, 'Toward the Next Generation of Atomic Mass Spectrometers', *J. Anal. At. Spectrom.*, **12**, 287–292 (1997).
68. P.P. Mahoney, G. Li, G.M. Hieftje, 'Laser Ablation/Inductively Coupled Plasma Mass Spectrometry with a Time-of-flight Mass Analyser', *J. Anal. At. Spectrom.*, **11**, 401–405 (1996).
69. L. Allen, S. Georgitis, D.P. Myers, K. Brushwyler, 'Trace Elemental Analysis of Metals by Laser Ablation Inductively Coupled Plasma Time-of-flight Mass Spectrometry', *Physica Status Solidi A, Applied Research*, **167**(2), 357–364 (1998).
70. J.L. Neilsen, A. Abildtrup, J. Christensen, P. Watson, A. Cox, C.W. McLeod, 'Laser Ablation Inductively Coupled Plasma/Mass Spectrometry in Combination with Gel Electrophoresis: A New Strategy for Speciation of Metal Binding Serum Proteins', *Spectrochim. Acta B*, **53**(2), 339–345 (1998).
71. W.J. Eastwood, N.J.G. Pearce, J.A. Westgate, W.T. Perkins, 'Recognition of Santorni (Minoan) Tephra in Lake Sediment from Golhisar Golu, Southwest Turkey by Laser Ablation ICP-MS', *J. Archeol. Sci.*, **25**(7), 677–687 (1998).

Liquid Chromatography/ Mass Spectrometry in Environmental Analysis

M. Careri
Università degli Studi di Parma, Parma, Italy

1 Introduction **1**
1.1 Mass Spectrometers: Characteristics and Recent Developments in Mass Analyzers 2
1.2 Chromatographic Mechanisms and Solvent Systems in Liquid Chromatography/Mass Spectrometry 3
1.3 Tandem Mass Spectrometry and Related Techniques 4
2 Environmental Applications **4**
2.1 Pesticides in Environmental Water Samples 4
2.2 Polycyclic Aromatic Hydrocarbons 10
2.3 Dyes 14
2.4 Surfactants and their Biodegradation Products in Wastewater, Surface Water and Drinking Water 17
2.5 Metals, Metal Complexes and Organometallic Compounds: the Problem of Speciation 20
3 Sample Treatment: Off-line and On-line Procedures **23**
Acknowledgments **25**
Abbreviations and Acronyms **25**
Related Articles **26**
References **27**

On-line liquid chromatography/mass spectrometry (LC/MS) has become a powerful technique in environmental analysis because of the separation capabilities of liquid chromatography (LC) and the detection and identification potential of mass spectrometry (MS), the characteristics of both methods making coupling very attractive. Among the interfacing systems used to couple LC with MS, those based on the atmospheric pressure ionization (API) offer advantages in terms of sensitivity and capability to analyze highly polar, ionic, heat sensitive and high-molecular-weight substances. Further, the on-line coupling of chromatographic separations with element-selective detection systems such as inductively coupled plasma mass spectrometry (ICP/MS) is an attractive technique for trace element speciation.

The main advantage of LC/MS lies in its ability to provide qualitative information (including molecular mass and fragmentation pattern) on the detected peaks. In addition, when all the ionization parameters are optimized, quantitative analyses can be carried out and low detection limits (at the ppb level) are obtained performing selective ion detection.

In environmental analysis, universal, selective and sensitive methods are needed to afford the assay of a wide variety of target and nontarget substances. Even though gas chromatography/mass spectrometry (GC/MS) is extensively applied to the analysis of organic contaminants in several matrices, LC/MS offers major advantages over GC/MS for analyzing polar, nonvolatile and thermolabile compounds. Ultraviolet (UV) or diode-array detector (DAD) are commonly used to perform detection in high-performance liquid chromatography (HPLC). However, UV-based detection methods for HPLC have several disadvantages deriving from nonspecific detection and lack of sensitivity. A more reliable identification is sometimes possible using DAD for certain compound classes, e.g. nitrophenols, in environmental samples; however, structural information is lacking for some classes of pollutants, among which are some pesticides, because their UV spectra are often almost identical. Using unspecific LC detectors, detection may be complicated by false-positive results; in such cases, MS has proved to be an extremely valuable technique for unequivocal identification of microcontaminants in a variety of environmental samples.

This overview discusses recent instrumental developments concerning mass analyzers of the mass spectrometers and on-line sample treatment procedures, with emphasis on the environmental field. Among the environmental applications of LC/MS, methods for the analysis of pesticides, polycyclic aromatic hydrocarbons (PAHs), dyes, surfactants, inorganic and organometallic compounds in environmental samples are discussed.

1 INTRODUCTION

High-performance liquid chromatography/mass spectrometry (HPLC/MS) realizes the combination of an analytical liquid-phase-based separation with mass spectrometric analysis for determination of nonvolatile, polar, thermolabile and ionic compounds.

On-line LC/MS has become a powerful technique in qualitative and quantitative analysis also in the environmental field because of the separation capabilities of LC and the detection and identification potential of MS, the characteristics of both methods making coupling very attractive.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The interfacing of LC with MS has been the object of fervent research for almost 30 years, since the first studies of Tal'roze et al. in 1969.[1] Different approaches have been devised to provide an effective combination of these two techniques including direct liquid introduction (DLI),[2] moving belt (MB),[3] particle beam (PB),[4] thermospray (TS),[5] continuous-flow fast atom bombardment (CFFAB)[6] and more recently electrospray (ES),[7] pneumatically-assisted electrospray (PAES) or ionspray (IS)[8] and heated nebulizer/atmospheric pressure chemical ionization (HN/APCI).[9] Finally, the on-line coupling of chromatographic separations with element-selective detection systems such as ICP/MS is an attractive technique for trace element speciation.[10]

LC/MS has progressed rapidly in the last years because it has benefited from the marked development of atmospheric pressure MS sources utilizing ES ionization; these ionization techniques are capable of producing ions from highly polar, ionic, heat sensitive and high molecular-weight substances. The main advantage of LC/MS is its ability to provide qualitative information (including molecular mass and fragmentation pattern) on the detected peaks. In addition, when all the ionization parameters are optimized, quantitative analyses can be carried out and low limits of detection (LOD) (at the ppb level) are obtained performing selective ion detection; selective ion monitoring (SIM) acquisition is a useful approach to obtain maximum sensitivity in the quantitation of target compounds.

In environmental analysis, universal, selective and sensitive methods are needed to afford the assay of a wide variety of target and nontarget substances. Even though GC/MS is still the most popular technique for the analysis of toxic organic micro-contaminants in several matrices, LC/MS offers major advantages over GC/MS for analyzing polar, nonvolatile and thermolabile compounds. Conventional detection in HPLC is usually achieved with UV or DAD. However, UV-based methods of HPLC have several disadvantages deriving from nonspecific detection and lack of sensitivity. The most serious drawbacks of UV detectors arise from matrix interferences particularly when working at low wavelengths, since several compounds lack a strong chromophore in the UV region for sensitive detection. When using DAD, a more reliable identification in environmental samples is sometimes possible for certain compound classes, e.g. nitrophenols. However, structural information is lacking for some classes of pollutants, for example, some pesticides, because their UV spectra are often almost identical. Therefore, UV detection and DAD are not well suited for universal application. If unspecific LC detectors are used, detection may be complicated by false-positive results; in such cases, MS has proved to be an extremely valuable technique for unequivocal identification of micro-contaminants in a variety of environmental samples.

1.1 Mass Spectrometers: Characteristics and Recent Developments in Mass Analyzers

Recent instrumental improvements to LC/MS systems are related to the use of mass analyzers instead of the quadrupole mass spectrometers.

Quadrupole instruments are the most common in LC/MS applications in various research areas and also in the environmental field; the relatively low costs, the ease of operation and fast data acquisition account for the widespread use of this mass analyzer. Other features of the quadrupole are the adequate dynamic range, the unit-mass resolution and the good quantitation capabilities; the restricted mass range (up to 4000 Da) is the main drawback, even though it is not a problem in environmental analysis.

Quadrupoles are readily combined with other types of analyzers, including other quadrupoles in arrangements for tandem mass spectrometry (MS/MS) experiments; triple quadrupole mass spectrometers are used in MS/MS for enhanced selectivity and improved detectability in trace analysis. This subject will be discussed in the section 1.3.

Commercial LC/MS instruments are also based on ion trap (IT) technology; interesting features of an ion trap mass spectrometer (ITMS) are the high sensitivity due to ion accumulation capabilities (100 times higher than by scanning quadrupole MS), the specificity because of the MS/MS options, the unlimited mass range and the scanning speed. On the other hand, IT systems generally suffers from space charge effects, ion–molecule reactions and from a limited dynamic range. Multiple-stage tandem mass spectrometry (MS^n) experiments which can be performed with this mass analyzer make the ITMS attractive particularly in structure elucidation rather than in quantitative target compound analysis. It should be noted that commercial IT-based LC/MS instruments are considerably less expensive than triple quadrupole systems and thus they could be ideally suited for target-type screening procedures in environmental applications.

Another mass analyzer which shows good perspectives in environmental LC/MS analysis is the time-of-flight (TOF) instrument. Time-of-flight mass spectrometry (TOFMS) can provide high sensitivity and a fast acquisition rate, combined with higher mass accuracy compared with scanning techniques such as quadrupole MS; the last feature is especially important for high-resolution capillary separation methods, such as capillary HPLC and capillary electrophoresis. Since it is a non-scanning instrument, mass resolution is not related

to sensitivity, so that high sensitivity can be achieved with a reasonably high resolution. TOFMS can be used in combination with pulsed ion sources, such as matrix assisted laser desorption/ionization (MALDI) making this system particularly useful in biological applications where high-molecular mass compounds are to be analyzed.

The magnetic sector mass spectrometer is used in environmental applications which require very high resolution, to aid structural elucidation and confirm molecular formulae of toxic micro-contaminants. Besides accurate mass measurements, this instrument can provide several benefits such as good sensitivity and selectivity and an excellent dynamic range. The high costs of the equipment, the complexity of operation are the main drawbacks of the magnetic sector mass spectrometer. Further, because of the high vacuum requirements, the combination of this mass analyzer with high-pressure LC/MS interfacing/ionization systems is not ideally suited; in fact, the enhancement in sensitivity due to improved ion transmission efficiency in the sector falls off because of losses in ion transmission in the passage from the high-pressure ion source to the high-vacuum analyzer.

High-resolution measurements are achievable with the Fourier transform ion cyclotron resonance (FTICR) instrument; the exact mass assignment capabilities, the excellent sensitivity and the feasibility of performing multiple-stage MS/MS are outweighed by the high prices and the severe vacuum restrictions of this sophisticated mass analyzer. For these reasons, it is rarely used in environmental applications and is unlikely to be employed widely in this field in the future.

1.2 Chromatographic Mechanisms and Solvent Systems in Liquid Chromatography/Mass Spectrometry

LC/MS coupling often puts restrictions on LC flow range or mobile phase composition. The eluent flow-rate plays a very important role in interfacing to MS. As for PB, the acceptable liquid flows vary in the $100-1000\,\mu L\,min^{-1}$ range according to the different nebulizer designs.[11] With the TS interface the total effluent from a conventional HPLC column (over $1\,mL\,min^{-1}$ of an aqueous mobile phase) can be introduced directly into the MS vacuum system. Interfaces such CFFAB and ES can accommodate only limited flow-rates (about $5\,\mu L\,min^{-1}$), whereas IS is capable of operating at $50-100\,\mu L\,min^{-1}$ thanks to the independent pneumatic nebulization; for this reason, in research to develop the various LC/MS interfacing systems, considerable attention has been paid to miniaturization of the HPLC columns to obtain the required linear velocity, without the need for conventional post-column splitting for flow-rate reduction. On the other hand, various manufacturers have proposed the introduction of higher flow-rate ES interfaces, such as turbo-ionspray™ (Perkin-Elmer Sciex trademark), ultrasonically assisted ES, thermally assisted ES and various high-flow options to be used in combination with pneumatically assisted ES; these instrumental developments with reference to currently commercially available LC/MS interfaces are extensively discussed in a recent review paper.[12] Among the API techniques, HN/APCI offers high solvent flow-rate capability (up to $2\,mL\,min^{-1}$); further, this interface/ionization system is compatible with a variety of mobile phase compositions, including pure aqueous solvents, making this technique highly versatile. The need for using volatile buffers with the other interfaces is a serious drawback, since chromatographic mechanisms, such as ion-exchange and ion-pair (IP) LC, which are useful for separating analytes which are also of environmental concern, cannot be universally applied. A possible solution to the requirements of volatile buffers and ion-pairing agents involves either replacement of involatile compounds by volatile equivalents or post-column suppression techniques.[13,14] In target-compound analysis, the phase-system switching (PSS) approach, based on heart cutting using valve-switching and coupled-column techniques, is successfully used to overcome some of the compatibility problems; following this on-line post-column liquid–solid extraction, a trapping column at the outlet of the analytical column is used to connect HPLC containing nonvolatile buffers and IP agents in the mobile phase to the mass spectrometer.[14–17] Vreeken[18] extensively discussed the application of post-column liquid–solid extraction in liquid chromatography/thermospray/mass spectrometry (LC/TS/MS) in a book devoted to applications of LC/MS in the environmental field.

With regard to the on-line coupling with ICP/MS, ion chromatography (IC) and ion-pairing chromatography are frequently used for inorganic and organometallic speciation. IC can separate both organic and inorganic charged species and, since it requires low buffer and organic solvent concentrations, unlike IP chromatography, is more amenable for coupling with ICP/MS. In general, liquid chromatography/inductively coupled plasma/mass spectrometry (LC/ICP/MS) based methods require the appropriate choice of chromatographic and plasma conditions to obtain efficient separations and to maintain good detector performance by preventing blocking of the interface or extinguishing of the plasma. Organic solvents used in LC eluents can lessen the performance of plasma MS because of the instability of the plasma to organic vapors and deposition of carbon on the sampling cone and torch. Chromatographic solvents used in normal-phase (NP) partition or adsorption present difficulty for effective plasma interfacing. Reversed-phase (RP) HPLC is most acceptable in LC/ICP/MS experiments; however, as the

percentage of organic eluent components increases, increase of radio-frequency power and the use of water-cooled spray chamber to enhance condensation of volatile solvents is needed. One approach can be microbore HPLC, with lower mobile phase flow-rates and a simpler interface. Mobile phases containing salts can cause short-term signal fluctuation, affect signal-to-noise (S/N) and give a more complex mass spectrum, particularly for $m/z < 80$. In addition, the deposition of salts in the nebulizer and on the sampling cone can cause blockage.[19,20]

1.3 Tandem Mass Spectrometry and Related Techniques

Interfacing systems such as TS, electrospray/ionspray (ES/IS) and HN/APCI yield soft ionization which can be complemented by recourse to liquid chromatography/tandem mass spectrometry (LC/MS/MS) as realized with the use of a triple-quadrupole system or quadrupole ITs; in the case of the API techniques, collision-induced dissociation (CID) in the interface itself (in-source CID) can be invoked to obtain structural information.

The most common MS/MS operating modes of tandem mass spectrometers are precursor ion scan, neutral-loss or product-ion scan and the selective reaction monitoring (SRM) mode,[21] the operational mode chosen being dependent on the analytical information required. Precursor ion, neutral-loss and product-ion scan modes are very selective and thus useful in screening, but they lack the sensitivity necessary for trace analysis. Instead, the use of SRM provides adequate selectivity and sensitivity for the determination of trace compounds in quantitative environmental assay; the choice of a limited number of precursor ion-product IPs to be monitored affords maximal selectivity and better S/N ratio. Both triple quadrupoles and IT analyzers can be used to carry out MS/MS experiments in product-ion mode and SRM mode; with the IT instruments, multiple-stage MS/MS can also be performed, as discussed above (section 1.1).

The selectivity of MS/MS reduces the need for complete chromatographic resolution of individual components; in fact, LC/MS/MS methods may facilitate the identification of substances in unresolved LC peaks, since a full-scan CID mass spectrum is characteristic of the structure of the parent compound. Fast LC/MS/MS analysis reduces the analytical run time thus providing high sample throughput. In addition, in some cases the use of MS/MS equipment allows direct analysis of even complex mixtures without chromatographic separation, making flow injection analysis/tandem mass spectrometry (FIA/MS/MS) helpful for time-saving.

The power of LC/MS/MS as a technique for the detection of analytes of environmental concern has been successfully demonstrated for quantitative analysis particularly in selected reaction monitoring mode.[22]

2 ENVIRONMENTAL APPLICATIONS

LC/MS has acquired a role of growing importance in environmental analysis, as is attested by the variety of applications appearing in recent years in this field and by the large number of reviews[23–32] and books[33–35] on the subject. In a review dealing with the occurrence, sample handling and chromatographic determination of pesticides in the aquatic environment, Barceló[23] has discussed the choice of gas chromatography (GC) or LC methods in combination with MS for the unequivocal characterization of pesticides and their breakdown products in environmental samples. In another paper, Barceló[24] has reviewed different coupled techniques such as GC/MS, LC/MS and MS/MS for the analysis of polar pesticides, surfactants, PAHs and polychlorinated biphenyls. Advantages and limitations of different LC/MS interfacing systems with reference to selected biological and environmental applications are discussed by the same author.[25] Careri et al.[26] have highlighted features and limitations of the various LC/MS approaches in an overview dealing with the application of this combined technique for the analysis of pollutants (pesticides, PAHs, inorganic and organometallic compounds, surfactants and dyes). An authoritative review of LC/MS which describes both the interfacing techniques and applications in environmental field is that published by Slobodník et al.:[27] the application-oriented section of the overview primarily deals with polar pesticides and related substances with emphasis on the on-line and off-line trace enrichment procedures. Chromatographic methods for the analysis of various classes of surfactants (nonionic surfactants, polar aromatic surfactants) in environmental samples have been extensively reviewed in recent papers.[28–30] An overview of current analytical methods including LC/MS for the analysis of sulfonated azo dyes in environmental samples has been published by Riu et al.[31] More recently, Di Corcia[32] has reviewed the state-of-the-art in the characterization of complex mixtures of isomers, oligomers and homologs of surfactants as well as their biointermediates in environmental matrices using LC/MS methodologies.

2.1 Pesticides in Environmental Water Samples

LC/MS is now indispensable in environmental organic analysis and surpasses GC/MS in the determination of polar pesticides; unlike GC, LC is not restricted by compound volatility or thermal lability, from which some pesticides suffer. In this respect, carbamates, triazines

Table 1 The EC priority pollutants list

Class	Compound class[a]
OCP	Organochloro pesticides
CP	Chlorophenols
PAH	Polyaromatic hydrocarbons
INORG	Inorganic – metals
OPP	Organophosphorus pesticides
VA	Volatile aromatics
AA	Amino aromatics
VHO	Volatile halogenated organics
COH	Halogenated hydroxyl compounds
CCOOH	Halogenated carboxyl compounds
CA	Chloroanilines
CNB	Chloronitrobenzenes
SVHO	Semivolatile halogenated organics
CNT	Chloronitrotoluenes
CT	Chlorotoluidines
TRIA	Triazines
PhAP	Phenoxy acid pesticides
ORGSn	Organotin compounds
BDs	Benzidines
CE	Chloroethers
A	Amines
PU	Phenylurea compounds
MISC	Miscellaneous

[a] Nomenclature used by the EC.

(TRIA), phenylureas and their corresponding anilines, and chlorinated phenoxy acids are groups of pesticides for which methods using LC are usually applied, making this technique appealing for multiresidue analysis.

The large number of different pesticides used within the European Community (EC) requires the development of rapid screening multiresidue methods for as many as possible of the most relevant pollutants. In addition, a European Union directive[36] demands methods that allow the verification of the concentration limit of $100\,ng\,L^{-1}$ for individual pesticide species in water intended for human consumption, which means that analytical methods should be able to detect pesticides at about one-tenth of this limit or lower, i.e. $10–20\,ng\,L^{-1}$. In Table 1 the classes of pesticides included in the EC priority pollutant list are reported.

In the US system of environmental monitoring, the maximum contaminant levels set by the Environmental Protection Agency (EPA) National Primary Drinking Water Regulations are in the $2–700\,\mu g\,L^{-1}$ range;[37] these guidelines demand limits of 40 and $200\,\mu g\,L^{-1}$ for carbofuran and oxamyl among carbamates in water, whereas limits of 50 and $70\,\mu g\,L^{-1}$ are established for fenoprop (2,4,5-TP) and dichlor phenoxyacetic acid (2,4-D) among phenoxy acid herbicides.

Because it is applicable to a wide range of compounds and is rugged, the TS interface has dominated LC/MS applications for pesticide analysis in environmental samples over the past 10 years. Thermospray mass spectrometry (TS/MS) is mainly used for target compound analysis rather than for the identification of unknown compounds in the environmental field; in fact, TS mass spectra often yield scarce fragmentation and thus lack structural information. To overcome this drawback, several approaches have been used, e.g. different ionization modes (the filament-on mode and the discharge ionization mode)[38] or the coupling of the TS interface with MS/MS.[39] The MS/MS approach is an elegant way of achieving relevant structural information, even though at the cost of a greater complexity of the technique; it is also highly useful for the characterization of unknown pesticide metabolites formed during degradation processes. Complementary information can also be gained exploiting both positive-ion (PI) and negative-ion (NI) operational modes.[40]

LC/MS with the PB interface is effective in the analysis of medium-polarity low-mass analytes of environmental concern. The combination of retention times, electron impact (EI) mass spectra and isotope distribution patterns for compounds containing naturally occurring isotopes provides excellent information for the unambiguous identification of both target and unknown compounds. The nonlinearity effects at low concentrations and the low sensitivity are the major drawbacks in the quantitative analysis performed by particle beam/mass spectrometry (PB/MS); better quantitative results in terms of linearity and detection limits can be achieved when operating in the presence of mobile phase additives or co-eluting internal standards.[11]

In Table 2 the most important classes of pesticides are briefly summarized, together with the LC/MS interfacing techniques used for their determination and the corresponding references.[17,41–53] To date LC/TS/MS and LC/PB/MS have been largely replaced by API techniques ES/IS and atmospheric pressure chemical ionization (APCI); the number of reports dealing with the application of electrospray/mass spectrometry (ES/MS) and HN/APCI/MS in pesticide analysis is rapidly growing. For this reason, this article will cover the most recent application examples in which the API techniques are used. Selected examples of multiresidue analyses of polar pesticides in environmental water samples (drinking water, ground- and surface waters) will be described in the following sections. Polar pesticides (phenylurea, sulfonylurea, phenoxy acid and triazine herbicides, carbamates, phenols, organophosphorus compounds) are considered, since they have become an important part of groundwater monitoring programmes, being the most likely to leach to groundwater.

Solid-phase extraction (SPE) for the preconcentration of the organic micropollutants is usually carried out; in some cases on-line monitoring procedures have been

Table 2 Selected applications of LC/MS techniques for the analysis of pesticides in environmental water samples

Class	Sample	Analytical technique	Ref.
Phenoxy acids	Surface water	LC/PSS/TS/MS; NI detection	17
	Surface water	SPE/LC/TS/MS; PI, NI detection	41
	Surface water	LC/PB/MS; EI; PI detection	42
	Surface water	SPE/LC/HFIS/MS; NI detection	43
	Surface water	LC/IS/MS; NI detection	44
	Drinking water		
	Surface water	SPE/LC/PB/MS; EI, PCI, NCI detection	45
		SPE/LC/APCI/MS; PI, NI detection	
	Groundwater	LC/IS/MS; LC/IS/MS/MS; NI detection	46
TRIA	Surface water	SPE/LC/TS/MS; PI, NI detection	41
	Surface water	LC/PB/MS; EI; PI detection	42
	Surface water	SPE/LC/PB/MS; EI, PCI, NCI detection	45
		SPE/LC/APCI/MS; PI, NI detection	
	Surface water	LC/TS/MS; PI detection	47
		LC/HFIS/MS; PI detection	
	Tap water	SPE/LC/IS/MS; SPE/LC/IS/MS/MS;	48
	Surface water	SPE/LC/APCI/MS; SPE/LC/APCI/MS/MS; PI, NI detection	
	Surface water	LC/APCI/MS/MS; PI detection	49
Phenylureas	Surface water	LC/TS/MS; PI detection	50
	Surface water	SPE/LC/TS/MS; PI, NI detection	41
	Surface water	LC/PB/MS; EI; PI detection	42
	Surface water	LC/TS/MS; PI detection	47
		LC/HFIS/MS; PI detection	
	Tap water	SPE/LC/IS/MS; SPE/LC/IS/MS/MS;	48
	Surface water	SPE/LC/APCI/MS; SPE/LC/APCI/MS/MS; PI, NI detection	
	Surface water	SPE/LC/APCI/MS/MS; PI detection	49
Carbamates	Surface water	SPE/LC/TS/MS; PI, NI detection	41
	Surface water	LC/PB/MS; EI; PI detection	42
	Tap water	SPE/LC/IS/MS; SPE/LC/IS/MS/MS;	48
	Surface water	SPE/LC/APCI/MS; SPE/LC/APCI/MS/MS; PI, NI detection	
	Tap water	LC/TS/MS; PI detection	51
	Surface water	LC/IS/MS; PI detection	
OPP	Surface water	SPE/LC/PB/MS; EI, PCI, NCI detection	45
		SPE/LC/APCI/MS; PI, NI detection	
	Tap water	SPE/LC/IS/MS; SPE/LC/IS/MS/MS;	48
	Surface water	SPE/LC/APCI/MS; SPE/LC/APCI/MS/MS; PI, NI detection	
	Groundwater	SPE/LC/APCI/MS; PI, NI detection	52
Phenols	Surface water	LC/PB/MS; EI, PCI, NCI detection	45
		LC/APCI/MS; PI, NI detection	
	Groundwater	LC/IS/MS; LC/IS/MS/MS; NI detection	46

LC/HFIS/MS, liquid chromatography/high flow ionspray/mass spectrometry; LC/APCI/MS, liquid chromatography/atmospheric pressure chemical ionization/mass spectrometry; LC/IS/MS, liquid chromatography/ionspray/mass spectrometry.

proposed. The combination of sample treatment by means of SPE with LC/MS in one set-up will be thoroughly discussed in section 3.

2.1.1 Triazines, Phenylureas, Anilides, Carbamates, Organophosphates, Nitrophenols

Column LC with IS or APCI interfaces using MS or MS/MS has been used for the analysis of 17 pesticides belonging to the cited classes in tap and surface waters.[48] Surface water samples were spiked with propazine (CAS 139-40-2) and metoxuron (CAS 19937-59-8) as internal standards. Sample aliquots were preconcentrated on PLRP-S (styrene–divinylbenzene copolymer) cartridges. Separations were performed on a C_{18} column, using acetonitrile–water as the mobile phase. The use of the on-line SPE/LC/API/MS/MS system made it possible to achieve detection limits at low nanograms per liter levels (Table 3). Negative ion acquisition mode was in most cases less sensitive, but organophosphorus compounds and nitrophenols gave adequate signals only in the NI mode. Detection limits obtained with APCI and IS interfaces are similar for the most compounds. However, the IS system appeared to be more prone

Table 3 Detection limits (S/N ratio 3) of 17 pesticides obtained from SPE/LC/PAES/MS and SPE/LC/APCI/MS in full-scan, SIM and MS/MS modes, using 100-mL samples[48]

No.	Compound[a]	SPE/LC/PAES/MS		SPE/LC/APCI/MS			Class
		Full-scan ($ng\,L^{-1}$)	SIM ($ng\,L^{-1}$)	Full-scan ($ng\,L^{-1}$)	SIM ($ng\,L^{-1}$)	MS/MS ($ng\,L^{-1}$)	
1	Oxamyl (23135-22-0)	500	3	500	3	n.m.	Carbamates
2	Dimethoate (60-51-5)	30	1	20	1	7	Organophosphates
3	Aldicarb (116-06-3)	100	2	25	2	n.m	Carbamates
4	Monuron (150-68-5)	40	2[b]	10	0.6	3	Phenylureas
5	Propoxur (114-26-1)	30	0.5	30	1.5	n.m.	Carbamates
6	Diuron (330-54-1)	300	2[b]	20	0.6	8	Phenylureas
7	Propazine (139-40-2)	8	0.1	7	0.4	1	TRIA
8	Terbutylazine	10	0.1	8	0.4	1	TRIA
9	Fenamiphos (22224-92-6)	10	0.2	10	0.8	1	Organophosphates
10	Alachlor (15972-60-8)	50	3	60	6.0	30	Anilides
11	Neburon (555-37-3)	40	2	20	1.5	4	Phenylureas
12	Fenthion (55-38-9)	n.d.	n.m.	3000	n.m.	n.m.	Organophosphates
13	Coumaphos (56-72-4)	500	20	350	15	75	Organophosphates
14	Fenchlorphos	n.d.	n.m.	n.d.	n.m.	n.m.	Organophosphates
15	Chlorpyriphos (2921-88-2)	n.d.	n.d.	2500	150	n.m.	Organophosphates
16	Trifluralin (1582-09-8)	n.d.	n.d.	3000	200	n.m.	Nitrophenols
17	Bromophos-ethyl	n.d.	n.m.	n.d.	n.m.	n.m.	Organophosphates

n.m., not monitored; n.d., not detected at highest concentration analyzed.
[a] CAS registry number enclosed in parenthesis, when available.
[b] SIM detection limits obtained in the NI mode. (Reproduced with permission of Elsevier Science from ref. 48 © 1996.)

to the formation of alkali metal adducts and the ion intensities were strongly dependent on the amount of sodium or potassium deriving from glassware and solvents. LC/APCI/MS in the PI mode appears to be the preferred method for the analysis of the classes of pesticides investigated, except for the organophosphorus compounds (Figure 1).

In MS/MS experiments, the excellent sensitivity of the SIM mode required for trace-level determination of pesticide residues can be combined with the capability in obtaining diagnostic product-ion spectra; Figure 2 shows the APCI and IS product-ion spectra obtained from protonated molecules of pesticides belonging to different classes. The acquisition of MS/MS spectra makes it possible to compile a library of product-ion spectra for the identification of the analytes in real samples.

2.1.2 Acidic Pesticides (Phenoxy Acids, Sulfonylureas, Phenols)

PAES interface has been used for the determination of 18 pesticides, including phenoxy acids, sulfonylureas, phenols, or their degradation products in groundwater samples by means of LC/MS and LC/MS/MS.[46]

To obtain satisfactory chromatographic results with these acidic analytes in terms of peak resolution, ion suppression with an acidic mobile phase is necessary. The chromatographic separation was obtained on a C_{18} narrow-bore column with a gradient elution of methanol–water–acetic acid. Porapak SPE cartridges were used for sample clean-up. Because of the acidic characteristics of the pesticides tested, mass analysis was performed under NI conditions in SIM mode; MS/MS experiments were performed in product-ion scanning mode or SRM mode. The highest sensitivity was achieved using MS and SIM detection. On the other hand, CID MS/MS experiments give a higher degree of confidence in the identification of the compounds by providing structural information.

In Table 4, in addition to the list of the analytes considered, comparison of the instrumental LOD obtained in the two MS detection modes is reported. For most compounds, LOD were three–four times lower using MS with respect to CID MS/MS. LC/IS/MS response was linear in the $5–500\,\mu g\,L^{-1}$ range for most analytes, except for four phenols [DNOC, dinoseb, bromoxynil and ioxynil] and bentazon. The method detection limits (MDLs) were determined on tap water samples spiked with the pesticides, using single MS in SIM mode. The results together with recovery values and precision are quoted in Table 5. The method was then applied to real samples of groundwater, which were first analyzed for the target herbicides using single MS in SIM mode. Positive findings were verified in a subsequent run using MS/MS in SRM mode by detecting diagnostic product ions. An example is reported in Figure 3; the concentration of phenoxy acids determined in the sample were $0.012\,\mu g\,L^{-1}$ of 2,4-D, $0.006\,\mu g\,L^{-1}$ of dichlorprop and $0.010\,\mu g\,L^{-1}$ of MCPA.

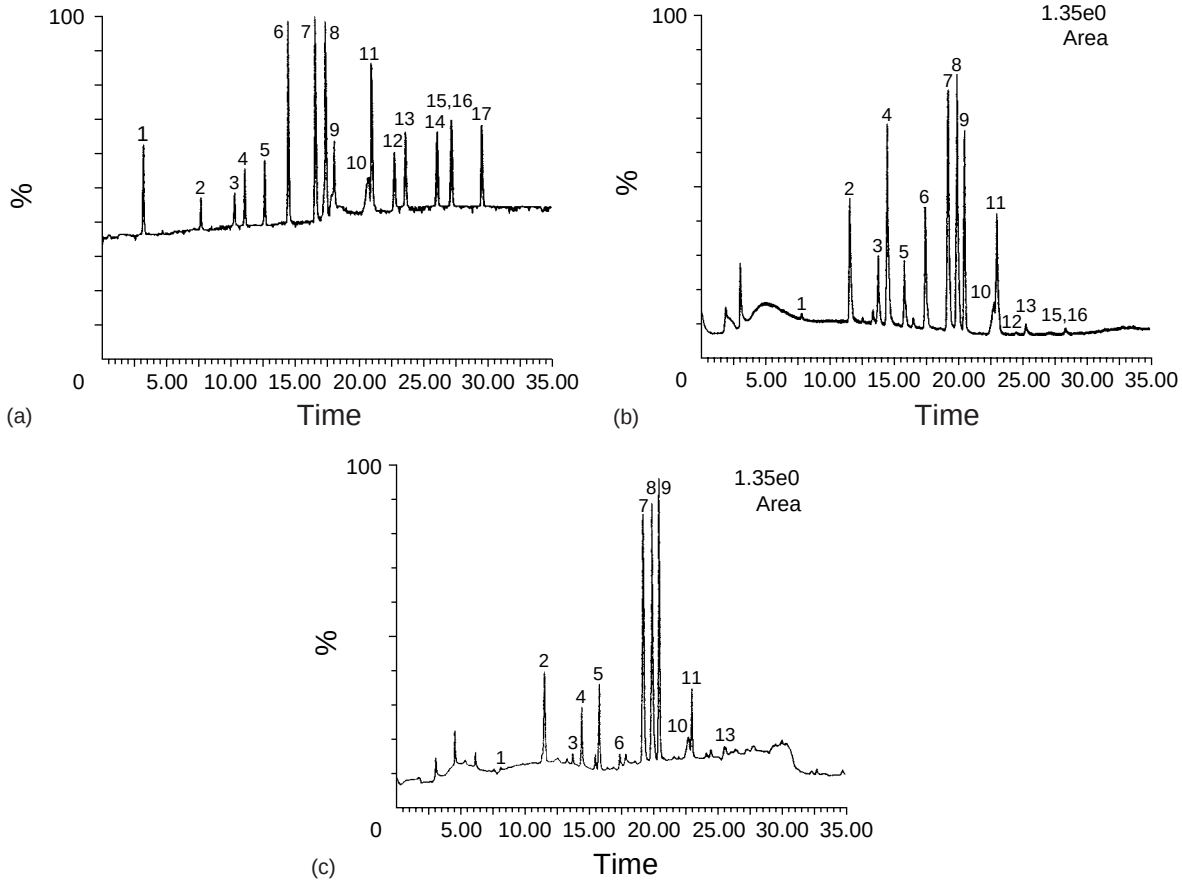

Figure 1 Chromatograms of the mixture of 17 pesticides obtained by (a) LC/UV, (b) on-line SPE/LC/APCI/MS and (c) SPE/LC/PAES/MS. Injected amount in (a) is 125 ng of each pesticide; the UV detector was kept at 215 nm. Enriched volume in (b) and (c) is 100 mL tap water; concentration of each analyte is 2.5 $\mu g\,L^{-1}$, MS is operated in full-scan mode, scan range m/z 150–420. Peaks: 1: oxamyl; 2: dimethoate; 3: aldicarb; 4: monuron; 5: propoxur; 6: diuron; 7: propazine; 8: terbutylazine; 9: fenamiphos; 10: alachlor; 11: neburon; 12: fenthion; 13: coumaphos; 14: fenchlorphos; 15: chlorpyriphos; 16: trifluralin; 17: bromophos-ethyl. (Reproduced by permission of Elsevier Science Publishers B.V. from J. Slobodník, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, *J. Chromatogr. A*, **741**, 59–74 © 1996.)

2.1.3 Carbamates

Ten carbamate pesticides of widely varying polarity were determined in waters and sediments using LC/IS/MS.[51] Concerning the interfacing conditions, the optimum cone extraction voltage with respect to analyte detectability and identification capabilities was between 15 and 35 V; this parameter is known to affect the sampling efficiencies in API sources.[54] In the mass spectra of most compounds the sodium adduct ion was the base peak. Chromatographic separations were performed on a reversed phase column using methanol–water mixtures as the eluent in gradient mode. The detection limits (S/N = 5–10) obtained with LC/IS/MS were in the 10–60 pg range (Table 6); linearity was checked in the 5–400 $\mu g\,L^{-1}$ range.

Concerning the applications, estuarine water samples were preconcentrated using liquid–solid extraction by passing 2 L of sample on C_{18}-bonded silica disks; the pesticides were recovered by two extractions with 10 mL of acetonitrile; after evaporation, the samples were dissolved in methanol and injected. Figure 4 shows the chromatogram of a surface water sample from the Ebro river and the mass spectrum of the peak eluting at 20 min, which allows the identification of carbofuran at low concentration levels, that is, at a concentration of about 1 $ng\,L^{-1}$ in the original water sample. Large-volume injections (500 µL) of tap water spiked at a level of 0.1 $\mu g\,L^{-1}$ of selected carbamates demonstrated the potential of LC/IS/MS with respect to trace-level detectability, making feasible the assay of these pollutants at levels below those of EU drinking water directives.[36]

2.1.4 Organophosphorus Pesticides

Twelve OPP have been determined in groundwater using LC/APCI/MS.[53] In Figure 5 the formulae of the

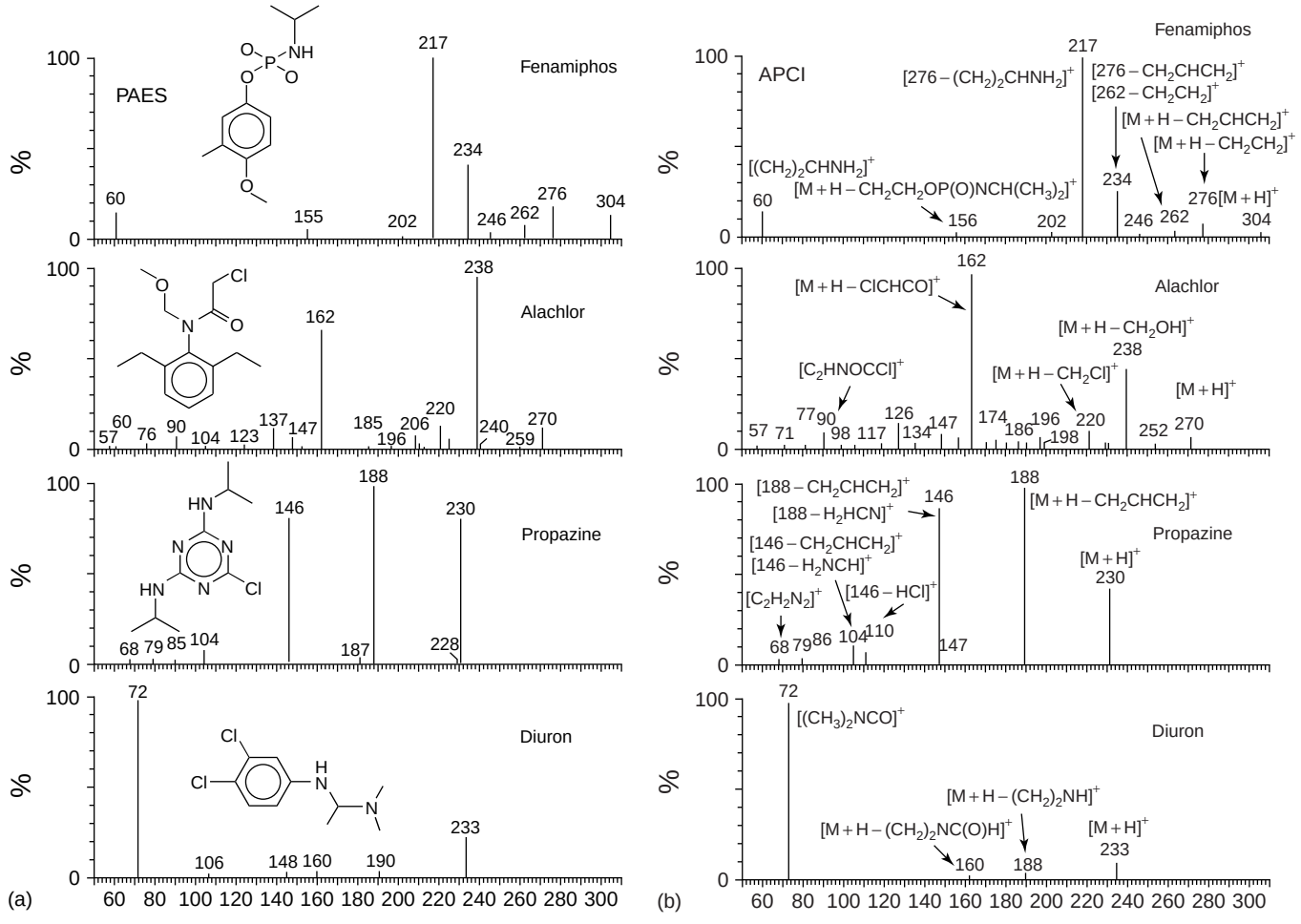

Figure 2 Product-ion spectra from protonated molecules of fenamiphos, alachlor, propazine and diuron in (a) PAES/MS/MS and (b) APCI/MS/MS. The APCI experiment was performed with injections of 125 ng of each compound onto a single short column. Temperature of the ion source, 150 °C; LC eluent flow-rate, 1 mL min^{-1}. The PAES/MS/MS spectra were obtained at a flow-rate of 25 µL min^{-1}; temperature of the ion source, 70 °C. (Reproduced by permission of Elsevier Science Publishers B.V. from J. Slobodník, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, *J. Chromatogr. A*, **741**, 59–74 © 1996.)

considered analytes are reported. An automated on-line liquid–solid extraction procedure was applied on RP (C_{18}) cartridges. Separation was achieved using a C_8 column and a gradient elution of methanol–water (1% acetic acid). Using APCI/MS with time-scheduled SIM in PI mode, the detection and quantitation limits reported in Table 7 were obtained. At the limit of quantitation (LOQ) concentration level, repeatability ($n = 5$) ranged from 17 to 25%. Operating in NI acquisition mode, best results were obtained for the parathion group compounds, whereas for the other compounds higher LOQs than those obtained acquiring the PI signals were achieved. Calibration graphs were linear in the range 0.0125–1 ng mL^{-1} for most of the analytes using SIM in PI or NI mode. Figure 6 shows the chromatograms for 100 mL of water samples spiked with 1000 ng L^{-1} of the OPP. The method was validated by analyzing certified samples of groundwater containing organophosphorus derivatives at levels between 10 and 100 ng L^{-1}.

In a recent study, the effects of the probe temperature and of extraction voltage on the fragmentation of the OPP were investigated with the aim of finding the best sensitivity and structural information for identification purposes;[54] compounds having the P=S bond were found to be more thermally stable with respect to those containing the P=O bond. In general, the optimum conditions as regards to flow-rate, probe temperature and extraction voltage were 1.2 mL min^{-1}, 400 °C and 20 V, respectively, even though for some compounds, e.g. trichlorfon (CAS 52-68-6), thermal degradation was observed. For these low thermolabile compounds ES/MS

Table 4 Instrument detection limits ($\mu g\,L^{-1}$ in injected solution, S/N ratio = 3[a]) of 18 pesticides or pesticide degradation products obtained by LC/ES using single MS and MS/MS modes[46]

Compound[b]	t_R (min)	Single MS			MS/MS		
		Full-scan	SIM	Ion	Full-scan	SRM	Ion (CE[c] (eV))
2,4-D (94-75-7)	20.9	1	0.5	219	10	3	161 (20)
2,4-Dichlorophenol	23.3	70	10	161	–[c]	–[d]	–[d]
Benazolin	17.0	15	2	242	37	5	170 (20)
Bentazon (25057-89-0)	17.2	0.7	0.1	239	18	4	132 (30)
Bromoxynil (1689-84-5)	20.5	0.7	0.1	276	33	15	79 (30)
Chlorsulfuron (64902-72-3)	19.0	4	0.4	356	19	7	139 (20)
Dicamba (1918-00-9)	14.5	50	9	219	68	8	175 (10)
Dichlorprop (120-36-5)	23.8	3	0.6	233	10	2	161 (20)
Dinoseb (88-857)	29.2	1	0.3	239	14	2	193 (30)
DNOC	20.7	0.8	0.1	197	71	23	122 (30)
Flamprop	23.5	4	0.3	320	19	3	121 (20)
Fluazifop	23.9	3	0.4	326	4	3	254 (20)
Ioxynil	22.3	0.4	0.06	370	100	21	127 (30)
MCPA (94-74-6)	22.0	8	1	199	16	4	141 (20)
Mecoprop (7085-19-0)	24.6	3	0.4	213	11	3	141 (20)
Metsulfuron-methyl (74223-64-6)	18.0	4	0.4	380	12	6	139 (20)
Thifensulfuron-methyl	17.5	5	0.5	386	23	6	139 (30)
Triasulfuton	17.2	5	0.5	400	23	3	139 (20)

MCPA, (4-chloro-2-methylphenoxy)acetic acid; DNOC, 4,6-dinitro-*o*-cresol.
[a] Measured on mass chromatogram of most abundant ion (single MS) or product ion (MS/MS).
[b] CAS registry number enclosed in parentheses, when available.
[c] CE = collision energy.
[d] No detectable product ion in MS/MS. (Reproduced with permission of Elsevier Science from ref. 46 © 1998.)

Table 5 Recoveries at a concentration level of $0.05\,\mu g\,L^{-1}$, MDLs and repeatability precision[46]

Compound	Recovery[a] (%)	MDLs ($\mu g\,L^{-1}$)	Precision (R.S.D.)[a] (%)
2,4-D	81	0.003	11
2,4-Dichlorophenol	47	0.026	17
Benazolin	44	0.008	19
Bentazon	61	0.002	9
Bromoxynil	57	0.003	10
Chlorsulfuron	35	0.004	19
Dicamba	33	0.032	15
Dichlorprop	89	0.003	11
Dinoseb	60	0.007	22
DNOC	51	0.003	14
Flamprop	87	0.004	11
Fluazifop	93	0.004	10
Ioxynil	55	0.005	13
MCPA	85	0.003	14
Mecoprop	86	0.003	10
Metsulfuron-methyl	28	0.004	28
Thifensulfuron-methyl	27	0.002	29
Triasulfuton	30	0.009	21

[a] Mean of duplicates from thirteen series of analyses. (Reproduced with permission of Elsevier Science from ref. 46 © 1998.)

is likely the most suitable technique, whereas the other polar OPP and the parathion group are more amenable for LC/APCI/MS analysis.

The examples described prove that LC/MS with IS or APCI interfaces provides instrumental detection limits, which make possible determinations of pesticides in environmental water samples at concentration levels below those of the EC Council directive 80/778/EEC ($0.1\,\mu g\,L^{-1}$ for individual pesticide).

2.2 Polycyclic Aromatic Hydrocarbons

The presence of PAHs in environmental samples is a legitimate cause for concern, because some PAHs are found to be highly carcinogenic.[55] Widespread in the environment, they originate from the incomplete combustion of organic material from both natural and anthropogenic sources;[55] PAHs have been observed in airborne particulates, coal soot, coal tar, particulate exhaust emissions of diesel- and gasoline-powered engines, marine sediments and water samples.

Owing to the carcinogenic properties of some of these substances, selective and sensitive analytical methods are required to assess their potential health risk and their impact in the environment. PAHs are priority pollutants on both the EC (Table 1) and US EPA lists. PAH determination is mostly limited to the analysis of 16 compounds (EC, EPA) (Table 8) in drinking, surface and wastewater by HPLC and HPLC/MS. A limit of $200\,ng\,L^{-1}$ of benzo[*a*]pyrene (CAS 50-32-8) in drinking water is set by the Office of Water of US

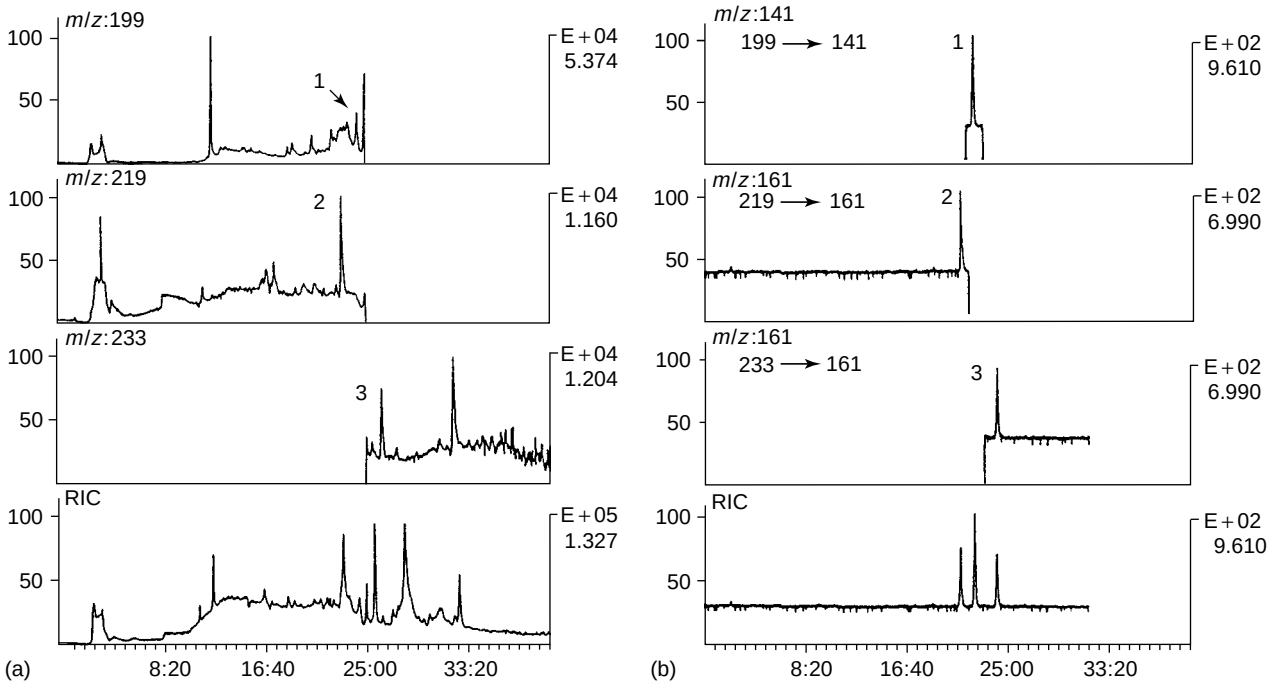

Figure 3 Typical chromatograms from an LC/ES/MS analysis of an authentic groundwater sample following pre-concentration by solid phase extraction. The sample was analyzed using single MS in SIM mode (a) and, subsequently, by MS/MS in SRM mode (b) for verification. Compound identification: 1: MCPA; 2: 2,4-D; 3: dichlorprop. (Reproduced by permission of Elsevier Science Publishers B.V. from B. Køppen, N.L. Spliid, *J. Chromatogr. A*, **803**, 157–168 © 1998.)

Table 6 LOD of carbamates in LC/TS/MS and LC/IS/MS, using time-scheduled SIM, at an extraction and a capillary voltage of 20 and 3500 V respectively[51]

Compound[a]	Type of ion	*m/z*	t_R (min)	LOD (pg)	
				TS	IS
Oxamyl	$[M + Na]^+$	242	4.1		15
(23135-22-0)	$[M + NH_3]^+$	237	5.9	2500	
Methomyl	$[M + Na]^+$	185	5.0		25
(16752-77-5)	$[M + NH_3]^+$	180	6.8	2500	
Carbendazim	$[M + H]^+$	192	13.2		30
(10605-21-7)	$[M + H]^+$	192	15.3	900	
Aldicarb	$[M + Na]^+$	213	13.2		15
(116-06-3)	$[M + NH_3]^+$	208	15.0	1000	
Propoxur	$[M + Na]^+$	232	19.0		15
(114-26-1)	$[M + NH_3]^+$	227	18.8	1200	
Carbofuran	$[M + Na]^+$	244	19.7		15
(1563-66-2)	$[M + NH_3]^+$	239	19.5	900	
Aminocarb	$[M + H]^+$	209	20.1		25
(2032-59-9)	$[M + H]^+$	209	18.8	500	
Pirimicarb	$[M + H]^+$	239	23.2		60
(23103-98-2)	$[M + Na]^+$	261	23.2		10
	$[M + H]^+$	239	22.8	500	
Promecarb	$[M + Na]^+$	230	28.5		5
(2631-37-0)	$[M + NH_3]^+$	225	28.9	600	
Barban	$[M + Na]^+$	280	29.8		8000
(101-27-9)	$[M + NH_3]^+$	275	30.1	900	

[a] CAS registry number enclosed in parenthesis, when available. LC was performed on a 12.5 cm × 3 mm i.d. column containing 3-μm C_8-deactivated silica, using a methanol–water gradient as LC eluent. (Reproduced with permission of Elsevier Science from ref. 51 © 1996.)

EPA.[37] The same tolerance level has been established in Directives of the Commission of the European Economic Communities.[36]

A number of SRMs have been issued by the US National Institute of Standards and Technology (NIST) to assist in validating the accuracy of PAH determinations. These SRMs range from a standard solution of 16 priority pollutant PAHs in acetonitrile (SRM 1647) to a complex mixture of PAHs from coal tar (SRM 1597), from air particulate material (SRM 1649) and from diesel particulate matter (SRM 1650).

The analysis and detection of PAHs in complex environmental samples is possible by a large number of different chromatographic methods; capillary GC combined with MS has been shown to be a very selective and sensitive analytical technique, even though its applicability is limited by the low volatility of high-molecular weight PAHs. Larger derivatives are more easily analyzed by LC methods using spectrophotometric or spectrofluorimetric detectors; on the other hand, the lower efficiency of LC demands the use of more selective detection methods, such as MS. A survey of the applications of LC/MS methods in the analysis of PAHs is shown in Table 9.[56] In the past MB and PB have been the interfacing systems most frequently used for the LC/MS assay of these low-polar analytes.[56–60,63] Generally, using the PB interface detection limits in the nanogram

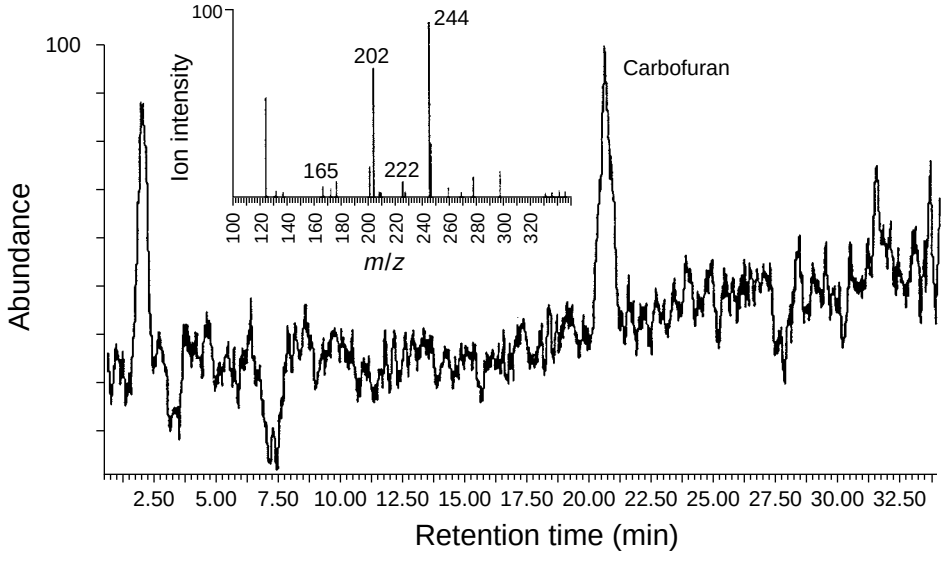

Figure 4 LC/IS/MS ion chromatogram of the ion at *m/z* 244 obtained after full-scan (*m/z* 50–300) acquisition and injecting 25 μL of a water sample (after 4000-fold preconcentration) from the Ebro river on to a 12.5 cm × 3 mm i.d. LC column. LC eluent flow-rate, 0.3 mL min^{-1}; extraction voltage, 60 V. The mass spectrum of the peak at a retention time of 20 min, shown as an inset, could be attributed to carbofuran ($[M + Na]^+$, *m/z* 244). (Reproduced by permission of Elsevier Science Publishers B.V. from M. Honing, J. Riu, D. Barceló, B.L.M. van Baar, U.A.Th. Brinkman, *J. Chromatogr. A*, **733**, 283–294 © 1996.)

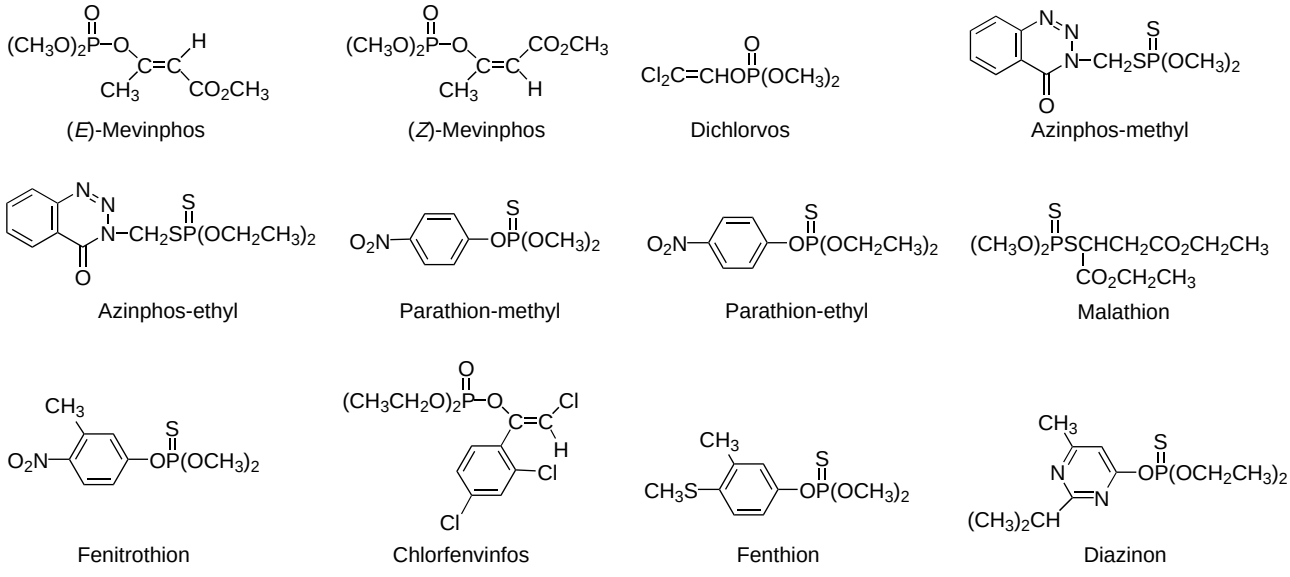

Figure 5 Chemical structures of the OPP studied. (Reproduced by permission of American Chemical Society from S. Lacorte, D. Barceló, *Anal. Chem.*, **68**, 2464–2470 © 1996.)

range were reported for PAHs containing four or more rings;[58,59] in particular, PB/MS detection affords to obtain detection limits of 0.2, 0.5, 1 and 2 ng for chrysene, benzo[*g*,*h*,*i*]perylene, dibenzo[*a*,*h*]anthracene and coronene (CAS 191-07-1) respectively.[59] In a more recent work, the HN/APCI interface for LC/MS has been demonstrated to be the most suitable technique for trace-level determination of PAHs;[60] in contrast to LC/PB/MS, the LC/HN/APCI/MS system yields a response linearity of at most three orders of magnitude and LOD in the picogram range for both low- and high-molecular-weight congeners. Positive ion APCI mass spectra of PAHs are characterized by both molecular ions and protonated molecules with no fragmentation; proton transfer from protonated water clusters leading to MH^+ results to prevail over electron transfer to species such as

Table 7 Calibration data obtained with LC/APCI/MS in time-scheduled SIM/PI mode for the studied pesticides (spiked from 0.0125 to 1 ng mL^{-1}) after on-line preconcentration of 100 mL of water[52]

Compound[a]	Linear range ($ng\,mL^{-1}$)	LOD[b] ($ng\,L^{-1}$)	LOQ[c] ($ng\,L^{-1}$)
(*E*)-Mevinphos (7786-34-7)	0.0125–1	2	10
(*Z*)-Mevinphos	0.0125–1	2	10
Dichlorvos (62-73-7)	0.0125–1	6	15
Azinphos methyl (86-50-0)	0.0250–1	12	30
Parathion-methyl	0.0125–1	9	22
Malathion (121-75-5)	0.0125–1	3	10
Fenitrothion (122-14-5)	0.0125–1	9	22
Azinphos-ethyl	0.0250–1	14	35
Parathion-ethyl (56-38-2)	0.0250–1	15	37
Fenthion (55-38-9)	0.0250–1	10	25
Chlorfenvinphos (470-90-6)	0.0125–1	1	5
Diazinon (333-41-5)	0.0125–1	2	10

[a] CAS registry number enclosed in parenthesis, when available.
[b] LOD calculated by using S/N ratio of 3–4 (the ratio between the peak intensity and the noise). LOD repeatability ($n = 5$) varied between 28 and 45%.
[c] LOQ obtained from LOD (2.5–5 times). LOQ repeatability ($n = 5$) varied between 17 and 25%. (Reproduced with permission of American Chemical Society from ref. 52 © 1996.)

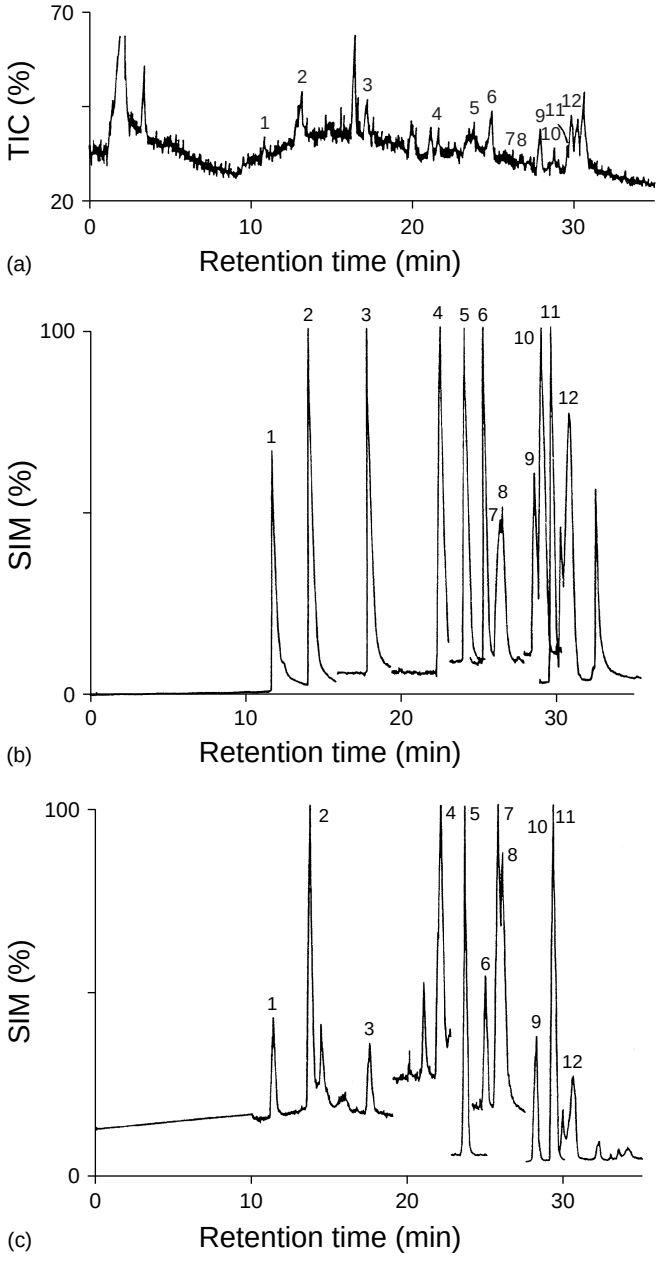

Figure 6 On-line liquid–solid extraction using OSP-2 autosampler, followed by LC/APCI/MS of 100 mL of groundwater samples spiked with 1000 ng L^{-1} of OPP under PI mode and full-scan conditions (a), time-scheduled SIM (b), and SIM using NI mode of operation (c) of 1: (*E*)-mevinphos; 2: (*Z*)-mevinphos; 3: dichlorvos; 4: azinphos-methyl; 5: parathion-methyl; 6: malathion; 7: fenitrothion; 8: azinphos-ethyl; 9: parathion-ethyl; 10: fenthion; 11: chlorfenvinphos; 12: diazinon. (Reproduced by permission of American Chemical Society from S. Lacorte, D. Barceló, *Anal. Chem.*, **68**, 2464–2470 © 1996.)

$N_2^{\bullet+}$ and $O_2^{\bullet+}$ with formation of M^+ as the size of the PAHs increases. The potential of the LC/HN/APCI/MS technique was demonstrated for a NIST certified coal tar reference material (SRM 1597), which is a natural, complex mixture of PAHs isolated from coal tar and certified for concentrations of 12 PAHs ranging from naphthalene to benzo[*g*,*h*,*i*]perylene. Internal standardization with perdeuterated PAH standards was used to perform the quantitative assay of the analytes using both the PB and APCI systems; in the first case, only seven out of 17 PAHs were detected, because of the poor transfer through the interface of the PAHs having molecular weights lower than 200 or greater than 380 Da. Conversely, using APCI interface all 17 of the target compounds were detected and accurately quantified; in general, satisfactory agreement with certified concentrations was observed when perdeuterated internal standards were used. The results of both PB and APCI measurements are summarized in Table 10.

PAHs, which are neutral low-polar compounds, cannot be detected under normal ES/MS conditions. The advantage of charge-transfer complexation to yield radical cations in solution for neutral substances for subsequent ES experiments has been demonstrated by Van Berkel et al.[61,62] However, ES/MS techniques do

Table 8 Compound names, molecular weights and CAS registry numbers of 16 priority pollutant PAHs

Compound	Molecular weight	CAS RN[a]
Naphthalene	128	91-20-3
Acenaphthylene	152	208-96-8
Acenaphthene	154	83-32-9
Fluorene	166	86-73-7
Phenanthrene	178	85-01-8
Anthracene	178	120-12-7
Fluoranthene	202	206-44-0
Pyrene	202	129-00-0
Benzo[*a*]anthracene	228	56-55-3
Chrysene	228	218-01-9
Benzo[*b*]fluoranthene	252	205-99-2
Benzo[*k*]fluoranthene	252	207-08-9
Benzo[*a*]pyrene	252	50-32-8
Dibenzo[*a*,*h*]anthracene	278	53-70-3
Benzo[*g*,*h*,*i*]perylene	276	191-24-2
Indeno[1,2,3-*cd*]pyrene	276	193-39-5

[a] CAS RN, Chemical Abstract Service registry number.

Table 9 Selected applications of LC/MS techniques for the analysis of PAHs in environmental samples

Sample	Analytical technique	Refs.
Coal tar	LC/MB/MS; PI detection	56
Sediment Water from the Exxon Valdez oil spill	LC/PB/MS; EI, PI detection	58
Standard solutions	LC/PB/MS; EI, PI detection	59
Coal tar SRM 1597	LC/PB/MS; EI, PI detection LC/APCI/MS; PI detection	60
Standard solutions	LC/ES/MS; PI detection	61, 62
Standard solutions	LC/PB/MS; EI, PCI, NCI detection	63

not provide adequate sensitivity for trace environmental analysis of PAHs.

2.3 Dyes

Dyes are widely used in industrial applications including textile dyes, printing inks, paint pigments, foodstuffs, plastics and cosmetics.[64] The textile industry is the largest consumer of these chemicals and accounts for two-thirds of the dyestuff market.[65] Dyes encompass structures that include (sulfonated) azo, anthraquinone, phthalocyanine, arylmethane and coumarin moieties. Azo dyes represent approximately 50% of the dyes available on the market today; they are classified in azo, diazo, triazo classes and found in various categories, e.g. acid, basic, disperse and pigments. The manufacturing and use of dyes creates wastes, which are discharged with the effluents into the receiving waters, thus making the determination of these chemicals in water samples an important task for environmental analysis. These products represent a potential human health risk because of their toxicity or carcinogenicity; in addition, for intermediates formed during the synthesis of dyes and degradation products of dyestuffs, such as naphthylamines, substituted phenylamines or BD analogs, mutagenic and carcinogenic activity is also suspected.[66]

Synthetic dyes are very polar or ionic and thermally labile substances, so that they are beyond the reach of capillary GC. Analytical methods for azo dyes are mostly based on HPLC with ultraviolet/visible (UV/VIS) detection; recently, the US EPA has approved several LC and LC/MS methods including Method 553 for BDs, which are carcinogenic compounds formed during synthesis of dyes.[67]

Various types of LC/MS set-up have been proposed for the analysis of dyes, with particular attention to azo dyes (Table 11). Scarce structural information and poor sensitivity were obtained using TS LC/MS under both PI and NI conditions;[68–70,73,75] in order to better characterize the dyes, the LC/TS/MS/MS approach proved to be useful.[70] Valuable MS data for the structural elucidation of dyes have been obtained using the LC/MS system with the PB interface;[73,74] again low sensitivity with detection limits in the 0.5–5 μg range were achieved.[74] A comparison of TS and PB mass spectra of an azo dye is shown in Figure 7. An application of capillary liquid chromatography (CLC) coupled with CFLSI/MS has been reported by Brumley et al.;[75] seven monosulfonated dyes with azo, diazo, triarylmethane and anthraquinone groups were considered among synthetic dyes. Using CFLSI/MS as a detection system for capillary LC, mass peak-profiling at high mass resolution was demonstrated as a potentially useful tool in the structure elucidation of unknowns (Figure 8).

The feasibility of API techniques for LC/MS and LC/MS/MS of azo dyes have been demonstrated by several authors.[71,72,74,76] A pioneering study on LC/MS analysis of sulfonated azo dyes is that of Bruins et al.,[71] who used HN/APCI and IS interfacing systems in NI mode. Compounds having acidic functional groups are most suited for negative ion mode LC/MS because of their acidic properties. Disulfonated azo dyes were found to be more difficult to analyze by MS than the monosulfonated derivatives; the NI APCI spectra are given in Tables 12 and 13. By comparing the two interface systems, sensitivity for sulfonated azo dyes obtained with IS was remarkably higher than by HN/APCI. By performing liquid-phase ionization by IS, sensitivity was 10 times

Table 10 Results of LC/MS analyses of a coal tar reference material (NIST SRM 1597) compared with certified and information values[a (60)]

No.	Compound	Certfd. concn.	PB interface			Heated pneumatic nebulizer		
			LC/MS concn.	RSD (%) ($n = 3$)	Reltv error (%)	LC/MS concn.	RSD (%) ($n = 3$)	Reltv. error (%)
1	Naphthalene ($C_{10}H_8$)	1160 ± 50				1149	2.9	−1.0
2	Acenaphthylene ($C_{12}H_8$)	250[b]				247	3.6	−1.1[b]
3	Acenaphthene ($C_{12}H_{10}$)					9.2	1.0	
4	Fluorene ($C_{13}H_{10}$)	140[b]				161	4.6	−15[b]
5	Phenanthrene ($C_{14}H_{10}$)	462 ± 3				461	3.2	−0.1
6	Anthracene ($C_{14}H_{10}$)	101 ± 2				107	5.4	6.1
7	Fluoranthene ($C_{16}H_{10}$)	322 ± 4	338	13	4.9	351	11	9.0
8	Pyrene ($C_{16}H_{10}$)	235 ± 2	238	11	1.2	241	5.0	2.6
9	Benzo[*a*]anthracene ($C_{18}H_{12}$)	98.6 ± 3.6	97.2	2.4	−1.4	105	4.6	6.8
10	Chrysene ($C_{18}H_{12}$)	71.7 ± 1.0	59.6	1.5	−17	70.4	7.5	−1.8
11	Benzo[*b*]fluoranthene ($C_{20}H_{12}$)	66[b]	64	2.1	−3.0[b]	63	7.9	−4.5[b]
12	Benzo[*k*]fluoranthene[c] ($C_{20}H_{12}$)	43[b]	41	5.5	−4.7[b]	41	8.0	−4.7[b]
13	Benzo[*a*]pyrene ($C_{20}H_{12}$)	95.8 ± 5.8	94.7	4.7	−1.2	84.8	1.1	−11
14	Benzo[*g*,*h*,*i*]perylene ($C_{22}H_{14}$)	53.7 ± 7.6				57.0	3.3	6.2
15	Indeno[1,2,3-*cd*] pyrene[d] ($C_{22}H_{12}$)	60.2 ± 4.4				60	25	−0.4
16	Dibenzo[*a*,*h*]anthracene ($C_{22}H_{14}$)	6.8[e]				7.4	6.6	8.8[e]
17	Coronene ($C_{24}H_{12}$)	11[b]				10.7	1.2	−2.7[b]

[a] Concentrations in $\mu g\, g^{-1}$. Internal standards were perdeuterated compounds (DPAC-1[7]) except where noted.
[b] Values not certified by NIST; information values reported by NIST.
[c] Perdeuteriobenzo[*b*]fluoranthene was used as internal standard.
[d] Perdeuteriobenzo[*g*,*h*,*i*]perylene was used as internal standard.
[e] Value not certified by NIST; information value obtained by HPLC by NIST. (Reproduced with permission of American Chemical Society from ref. 60 © 1995.)

Table 11 Selected applications of LC/MS techniques for the analysis of dyes in environmental samples

Class	Sample	Analytical technique	Ref.
Azo, diazo, anthraquinone	Wastewater Gasoline	LC/TS/MS; PI detection	68
Azo, methyne, arylmethane, anthraquinone, coumarin, xanthene	Dye wastes	LC/TS/MS; PI, NI detection	69
Disperse Azo	Wastewater	LC/TS/MS; PI detection LC/TS/MS/MS	70
Mono-, disulfonated azo	Lake water, Wastewater	LC/APCI/MS; LC/APCI/MS/MS; NI detection LC/IS/MS; LC/IS/MS/MS; NI detection	71
Sulfonated azo	Wastewater	LC/IS/MS/MS; NI detection	72
Azo	Dye wastes	LC/PB/MS; EI, NCI detection	73
Azo, diazo	Commercial dye samples	LC/PB/MS; EI detection LC/TS/MS; PI detection LC/ES/MS; NI detection	74
Azo, diazo, triarylmethane, anthraquinone	Water, Soil	CFLSI/MS; NI detection	75
Mono-, disulfonated azo	Spiked drinking water	LC/TS/MS; PI, NI detection LC/HFIS/MS; NI detection LC/APCI/MS; PI, NI detection	76

CFLSI/MS, continuous flow liquid secondary ion/mass spectrometry.

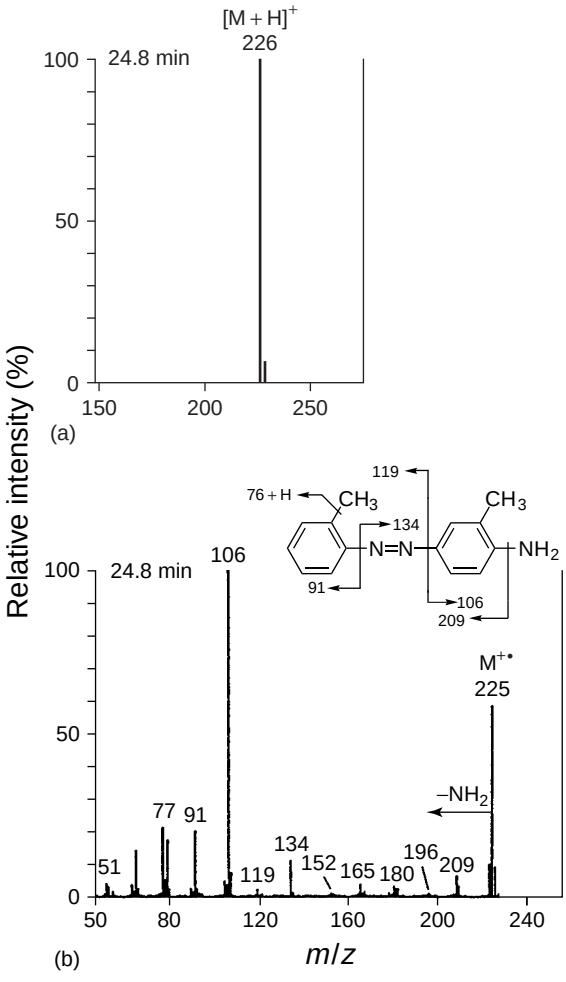

Figure 7 (a) HPLC/TSP and (b) HPLC/PB mass spectra of Solvent Yellow 3 (M_r 225). Amount injected, 37 µg; solvent methanol–water gradient. (Reproduced by permission of Elsevier Science Publishers B.V. from R. Straub, R.D. Voyksner, J.T. Keever, *J. Chromatogr.*, **627**, 173–186 © 1992.)

higher for mono- than for disulfonated derivatives and the detection of approximately 10 ng of a monosulfonated product was demonstrated in full-scan conditions. Gas-phase ionization by APCI was 20–50 times less sensitive for disulfonated azo dyes compared with monosulfonated ones. Using LC/IS/MS/MS, product-ion spectra were generated by CID from the parent ions $[M - H]^-$; the $SO_3^{\bullet -}$ ion at *m/z* 80 was found to be a diagnostic fragment in the MS/MS analysis of this class of dyes so that SRM of the fragmentation reactions $[M - H]^- \longrightarrow SO_3^{\bullet -}$ for each analyte could afford selective detection of the dyes, as illustrated in Figure 9.

The better suitability of IS with respect to APCI for this class of substances has been more recently confirmed by Ràfols and Barceló[76] (Tables 14 and 15). On the other hand, using APCI in PI mode these authors observed a much better sensitivity towards disulfonated azo dyes than that previously found,[71] likely owing to a more efficient nebulization in the APCI interface.

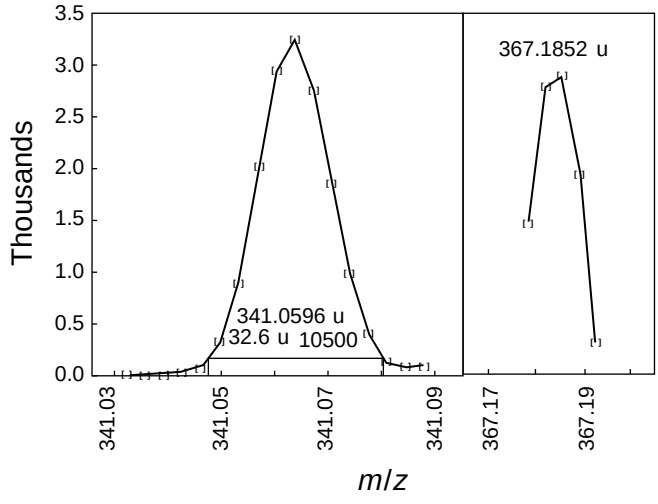

Figure 8 Mass peak-profiling at 10000 resolution on CLC/MS chromatographic peaks: *m/z* 341.0596 of orange 8 and *m/z* 367.1816 of glycerol matrix as the lock mass. (Reproduced by permission of Elsevier Science Publishers B.V. from W.C. Brumley, C.M. Brownrigg, A.H. Grange, *J. Chromatogr. A*, **680**, 635–644 © 1994.)

Table 12 Negative ion APCI spectra[a] of monosulfonated azo dyes[b(71)]

	$[M - H]^-$		Others	
	m/z	%	*m/z*	%
Acid Orange 7 (633-96-5)	327	100		
Acid Red 337 (12270-02-9)	410	100	238	30
			252	10
Acid Red 88 (2 isomers) (1658-56-6)	377	100	207	5
			222	10
Acid Red 151 (6406-56-0)	431	100	349	10
			277	20
Acid Black 52 (very impure mixture) (5610-64-0)	438	100	215	10
			201	70
			217	70
			218	60
			240	40
			283	20
			328	30
			410	20
Acid Yellow 151 (sulfonamide) (12715-61-6)	375	60	213	30
			230	50
			254	100
			314	20

[a] Mass range *m/z* 200–500; CAS RN enclosed in parenthesis.
[b] Introduced via LC column; M refers to free acid or free ligand of a metal complex. (Reproduced with permission of American Chemical Society from ref. 71 © 1987.)

Table 13 Negative ion APCI spectra[a] of disulfonated azo dyes[b(71)]

	$[M-H]^-$		Others	
	m/z	%	*m/z*	%
Acid Red 1 (3734-67-6)	464	40	422	40
			446	100
Acid Black 1 (1064-48-8)	571	80	214	15
			224	20
			242	10
			258	25
			400	10
			422	100
Acid Blue 113 (3351-05-1)	636		298	100
			311	10

[a] Mass range *m/z* 200–700; CAS RN enclosed in parenthesis.
[b] Introduced via LC column; M refers to the free acid. (Reproduced with permission of American Chemical Society from ref. 71 © 1987.)

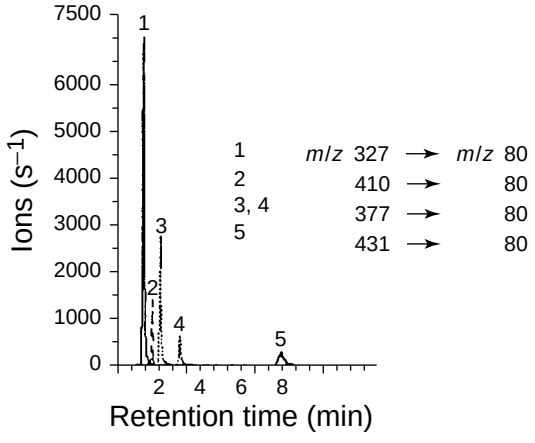

Figure 9 Ion spray LC/MS/MS of a mixture of five monosulfonated azo dyes (Acid Orange 7, Acid Red 337, 2 isomers of Acid Red 88, Acid Red 151) extracted from fortified Cayuga Lake water, selected reaction monitoring of $[M-H]^- \longrightarrow SO_3^{\bullet -}$: collision gas, argon; collision energy, $E = 120$ eV; sample size, approximately 20 ng of each component; liquid phase ionization by ion spray; 1 mm I.D. × 100 mm Hypersil 3-μm C_{18} column; flow rate, 40 μL min^{-1}, 30% acetonitrile, 70% 0.001 M ammonium acetate in water. (Reproduced by permission of American Chemical Society from A.P. Bruins, L.O.G. Weidolf, J.D. Henion, W.L. Budde, *Anal. Chem.*, **59**, 2647–2652 © 1987.)

2.4 Surfactants and their Biodegradation Products in Wastewater, Surface Water and Drinking Water

Surfactants are environmentally relevant because they are extensively used and occur at high levels of concentration in water.[77] These chemicals are used in large volumes both industrially and domestically in a broad variety of detergents and cleaning products. Surfactants have hydrophilic and hydrophobic centers both of which influence their physical and chemical properties; they may be classified as anionic surfactants, e.g. alkylbenzenesulfonates, cationic surfactants, e.g. quaternary ammonium salts, and nonionic surfactants such as those produced by reacting alkylphenols with ethylene oxide. General formulae and acronyms of the most widespreadly used anionic and nonionic surfactants which are discussed in

Table 14 Calibration data obtained with LC/ES/MS in time-scheduled SIM/NI mode for the dyes studied[76]

Compound	*m/z*	Linear range (ng)	LOD (ng)	LOQ (ng)
Acid Red 1	464	20–600	20	60
	232	60–600	60	150
Acid Red 13	228	3–900	3	11
Acid Red 14	228	3–800	3	11
Acid Red 73	511	12–800	12	50
Mordant Yellow 8	401	12–800	12	45
Acid Yellow 23	211	76–800	76	180
Direct Yellow 28	317	1.4–800	1	5
	318	2.6–800	3	10
Mordant Black 11	266	13–800	13	30
	294	26–800	26	80
Mordant Black 17	221	70–700	70	160
	249	70–700	70	160
Acid Blue 113	317	2–700	2	5
	318	2–700	2	5

Number of data points, $n = 7$, except for A. Red 13, A. Yellow 23 and M. Black 11, where $n = 5$. Standard errors for the intercept and the slope varied from 50–100 and from 3–10%, respectively. Mordant Red was not detected. (Reproduced with permission of Elsevier Science from ref. 76 © 1997.)

Table 15 Calibration data obtained with LC/APCI/MS in time-scheduled SIM/NI mode for the dyes studied[76]

Compound	*m/z*	Linear range (ng)	LOD (ng)	LOQ (ng)
Acid Red 13	236	140–2000	140	550
	298	70–2000	70	
	350	80–2000	80	240
Acid Red 14	236	100–3200	100	550
	298	70–3200	70	210
	350	70–3200	70	210
Acid Red 73	197	60–2100	60	200
	431	60–2000	60	200
	350	60–2000	60	200
Acid Yellow 23	172	50–2000	50	150
Direct Yellow 28	319	50–2200	50	150
Mordant Black 17	165	700–2000	700	700
	260	70–2000	70	210
Acid Blue 113	298	70–2300	70	210
	313	120–2300	120	360

Number of data points, $n = 7$, except for A. Red 13, A. Blue 113 and M. Black 11, where $n = 5$. Standard errors for the intercept and the slope varied from 100 to 200 and from 3 to 10% respectively. Acid Red 1, Mordant Yellow 8 and Mordant Black 11 were not detected at the concentrations studied. (Reproduced with permission of Elsevier Science from ref. 76 © 1997.)

Table 16 General formulae and acronyms of the most widely used anionic and nonionic surfactants

Common name (acronym)	Hydrophobic group	Hydrophilic group
Anionic surfactants		
Linear alkylbenzenesulfonates LAS	$C_6H_4-C_nH_{2n+1}$ $n = 10-14$	SO_3^-
Alkylethoxysulfates AES	C_nH_{2n+1} $n = 12-15$	$-(OCH_2CH_2)_n-OSO_3^-$ $n = 1-8$
Alkylsulfates AS	C_nH_{2n+1} $n = 12-15$	OSO_3^-
Nonionic surfactants		
Nonylphenol ethoxylates NPEOs	$C_9H_{19}-C_6H_4$	$-(OCH_2CH_2)_n-OH$ $n = 1-23$
Octylphenol ethoxylates OPEOs	$C_8H_{19}-C_6H_4$	$-(OCH_2CH_2)_n-OH$ $n = 1-23$
Aliphatic ethoxylate alcohols AEOs	C_nH_{2n+1} $n = 12-18$	$-(OCH_2CH_2)_n-OH$ $n = 1-23$

Table 17 Selected applications of LC/MS techniques for the analysis of surfactants in environmental water samples

Class	Sample	Analytical technique	Ref.
LAS PEG-type surfactants, PEG PPG-type surfactants	Wastewater	LC/TS/MS, PI detection LC/TS/MS/MS, PI detection	79
LAS APEO	Surface water	LC/APCI/MS, PI detection	80
AS, AES	Sewage Surface water	LC/IS/MS, NI detection	81
AS, AES, LAS OPEO, NPEO, PEG-type surfactants, PEG	Surface water	LC/ES/MS, PI, NI detection LC/APCI/MS, PI, NI detection FIA/IS/APCI/MS/MS	82
AEO	Surface water	LC/TS/MS, PI detection	86
AEO, NPEO	Raw and treated Sewage Surface water Drinking water	LC/IS/MS, PI detection	87

PEG, poly(ethylene glycol); PPG, polypropylenglycol; APEOs, alkylphenol ethoxylates.

this article are reported in Table 16. Table 17 summarizes selected applications of LC/MS-based methods for the analysis of surfactants in water samples.

2.4.1 Anionic Surfactants

Anionic surfactants have been extensively studied in aquatic environments; particular attention has been paid to LAS due to their biodegradability which makes them potentially hazardous with respect to surface and groundwater, affecting also drinking water quality. LAS (Table 16) are water-soluble materials used extensively in detergent formulations as surface-active agents.

Commercial LAS materials consist of a mixture of homologs and isomers. Individual LAS homologs and isomers are identified by the length of the hydrophobic alkyl substituent and by the position of attachment of the alkyl chain to the benzene ring.

Other anionic surfactants commonly used in consumer products are AES and AS. After use, the majority of sulfonate and sulfate surfactants is disposed to sewage treatment plants (STPs); during the sewage treatment process anionic surfactants are extensively removed as the result of biodegradation processes[78] and adsorption to solids.[79]

Growing attention is being paid to the analysis of the degradation products of commonly used surfactants; concerning LAS, their biotransformation leads to the formation of mono- and dicarboxylic sulfophenyl acids via the w-oxidation of the alkyl chain.[80] In support of biodegradation and ecotoxicological studies, reliable analytical methodologies are required to determine these

substances at trace levels in wastewater treatment plant influents and effluents, as well as river waters. Various analytical methods are available for the measurement of these chemicals, including simple colorimetric methods, GC and LC, as well as MS methodologies. Anionic surfactants are not sufficiently volatile to permit direct measurement by GC; as a result, they must be either derivatized to form a volatile compound or, in the case of LAS, desulfonated to form the respective linear alkylbenzenes. HPLC has been widely applied to the measurement of these surfactants both in NP and in RP in conjunction with UV and fluorimetric detection. IP RP LC, which has become the method of choice for the separation of sulfonates and their degradation products sulfophenyl carboxylates, is incompatible with MS detection (section 1.2). An elegant approach was that of coupling LC to IS/MS via a postsuppressor split with an IC system for the separation and detection of alkyl sulfates and sulfonates;[13] the function of the micromembrane suppressor was that of removing over 99.9% of the IP agents required for the chromatography from the mobile phase. LC/MS has recently been applied widely to the analysis of anionic surfactants and their degradation biointermediates in water (Table 17).[13,79–82] A comprehensive review on the use of TS interface for LC/MS and flow injection analysis/mass spectrometry (FIA/MS) analysis of these chemicals and the corresponding metabolites has been written by Schröeder.[83] Anionic surfactants have been detected either in PI mode as MNH_4^+ adduct ions in LC/TS/MS[79] or as MNa^+ adduct ions in LC/APCI/MS,[80] or in the NI mode as deprotonated molecules using TS/MS[79] or ES/MS detection systems.[13,81] LC/MS with ES ionization has been proved ideally suited for trace-level identification and quantification of anionic surfactants;[81] a quantitative LC/MS method for the determination of AES in environmental samples using IS interface has been validated for three aqueous samples, i.e. sewage influent, sewage effluent and river water. Detection of 36 components, down to ppb levels, was feasible in full-scan mode with a single injection, whereas multiple injections in SIM conditions allowed one to determine 50-fold lower concentrations, limits-of-detection being about $0.01\,\mu g\,L^{-1}$ for most of the analytes in river water.

2.4.2 Nonionic Surfactants

Nonionic surfactants represent the second most important class of surfactants after the anionic class. The three nonionic surfactants discussed here are those listed in Table 16. AEOs were introduced for replacing the more toxic OPEOs and NPEOs and are being used increasingly both in the household and in industry. Although APEOs and AEOs are not classified as dangerous substances, investigation on the removal efficiency of these organics by STPs and their ecotoxicological impact on aquatic life demands the development of appropriate analytical methods able to determine them in influents and effluents of STPs as well as in the natural aquatic environment. Chronic toxicity data of nonionic surfactants to aquatic life forms such as algae, fish and invertebrates range from 0.1 to $20\,mg\,L^{-1}$, depending on substance structures and experimental conditions. In addition, more toxic degradation products of nonionic surfactants can be produced as a consequence of the metabolic activities of aquatic microbial communities. The PEG family has been indicated as primary key intermediates of the aerobic biodegradation of AEOs.[78] The progressive hydrolytic shortening of the polyethoxy chain of NPEOs leading mainly to the formation of metabolites with one and two ethoxy units together with nonylphenol is the most favored pathway of APEO aerobic biodegradation.[84] Nonylphenol and octylphenol have been reported to display oestrogenic activity, being suspected of altering reproduction in fish.[85]

NP HPLC and RP HPLC have been used to separate and determine the homologs and oligomers of nonionic surfactants by the alkyl chain and ethoxylate lengths. If HPLC with fluorimetric detection provides the most simple and suitable technique for the analysis of these compounds in environmental samples, on the other hand LC/MS offers a unique possibility for obtaining information on the occurrence and distribution of homologs, oligomers and isomers at the ppb level. Also investigation on PEG-type compounds in the aquatic environment is made easier by the use of LC/MS methodologies; in fact, the high polarity of these metabolites combined with the lack of any chromophore group in their structures makes them not amenable to GC and LC determination. Nonionic surfactants and their biointermediates have been analyzed using TS,[79,86] and ES[87] interfaces for LC/MS. LC/TS/MS has been proven to exhibit adequate sensitivity and selectivity for AEO determination, allowing the quantitation of total and individual alcoholethoxylate species at the ppb level. In addition, the technique discriminates branched AEO from linear ethylene-based AEO, which may co-elute. The on-line coupling of solid phase extraction with LC/TS/MS used for the enrichment of AEOs from groundwater and wastewater allowed to quantify individual alcohol ethoxylates in the 60 ppt and 2.17 ppb range.[86] In another study,[79] the comparison between the results obtained by LC/TS/MS and those from flow injection analysis/thermospray/mass spectrometry (FIA/TS/MS) analyses of AEOs and their metabolites in wastewater samples showed that the chromatographic separation does not contribute significantly to the information obtainable by the direct injection of the extracts in flow injection analysis (FIA) mode.

In addition, since determination by LC/MS took 25–30 times longer than FIA/MS, this technique provided higher sample throughout.

Recently, the LC/MS technique with ES interface has been promising in terms of sensitivity and selectivity for the analysis of nonionic polyethoxylate surfactants, such as AEOs and NPEOs in various environmental aqueous samples (influent and effluent waters from STPs, river waters, drinking waters).[87] An extraction/enrichment procedure involving the application of large water volumes (up to four liters) on a graphitized carbon black cartridge allowed to differentially elute the AEOs and NPEOs and their biointermediates according to differences in their acid strengths. The chromatographic method developed could afford the elution of all the oligomers of NPEO and of the various AEO homologs into single peaks (Figure 10), in such a way that quantitative assay was simplified and detection levels improved. The resulting detection limits for each single homolog of AEO and for NPEOs in the influent, effluent, river water and drinking water samples investigated were estimated to be 0.6, 0.02, 0.002 and 0.0002 $\mu g\,L^{-1}$ respectively. The selectivity and sensitivity of the method is illustrated in the chromatograms obtained by analyzing 4-liter tap water samples (Figure 11).

Efforts to characterize the degradation intermediates of nonionic surfactants in environmental matrices have increased in recent years. Concerning the metabolites of AEOs, TS/MS and thermospray tandem mass spectrometry (TS/MS/MS) techniques have been successfully used for the analysis of PEGs in various aqueous samples;[83] mass trace analysis allowed detection of individual PEG homologs of the general empirical formula $HO-(CH_2-CH_2-O)_nH$ $[n = 5–13]$ at low concentrations, whereas these compounds were not observed in the reconstructed ion chromatogram (RIC). An ES/MS-based method[87] was shown not to be adequate for the detection of degradation products of NPEOs containing one or two ethoxy units; this result provided low accuracy in the quantitation of NPEOs in STP effluents, since the oligomer distribution of a biodegradated NPEO mixture is different from that of NPEOs in the calibrated standard solutions.

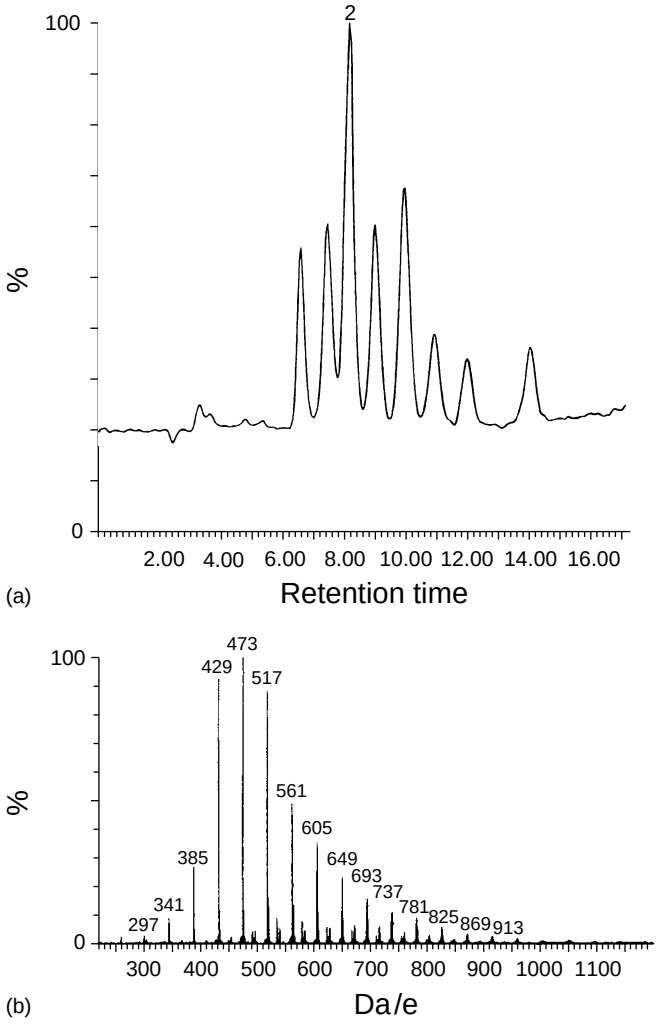

Figure 10 (a) LC/MS chromatogram obtained by injecting the composite working standard solution containing NPEOs and AEOs. The analytes in this synthetic mixture are, in order of elution: 1: NPEOs; 2: C-12 ethoxylates (EOs); 3: C-13 EOs; 4: C-14 EOs; 5: C-15 EOs; 6: C-16 EOs, 7: C-18 EOs. Internal standard, C-10 EO_6. (b) Mass spectra of the C-12 EO oligomers eluted as the single peak 2. As an example, the mass spectrum with $m/z = 473$ is for $[C\text{-}12\,EO_6 + Na]^+$. (Reproduced by permission of American Chemical Society from C. Crescenzi, A. Di Corcia, R. Samperi, A. Marcomini, *Anal. Chem.*, **67**, 1797–1804 © 1995.)

2.5 Metals, Metal Complexes and Organometallic Compounds: the Problem of Speciation

Nowadays, probably the most powerful technique for the analysis of metals and organometallic compounds at trace and sub-trace level is ICP/MS. The main features of this technique are the high element specificity, the capabilities of multielement detection, the feasibility of isotopic ratio determinations, the very low detection limits (usually in the low-ppb and ppt range) and the wide linearity range (in most cases up to 5 or 6 orders of magnitude). On the other hand, an inherent limitation of ICP/MS is that it provides only total elemental concentration information, thus denying the speciation and identification of the chemical and organometallic form of the analyte. This fact presents an important drawback in some problems of environmental concern, since the mobility of metals in the environment and their bioavailability depend on their chemical form (free metal, complexed metal, organometallic species). Consequently,

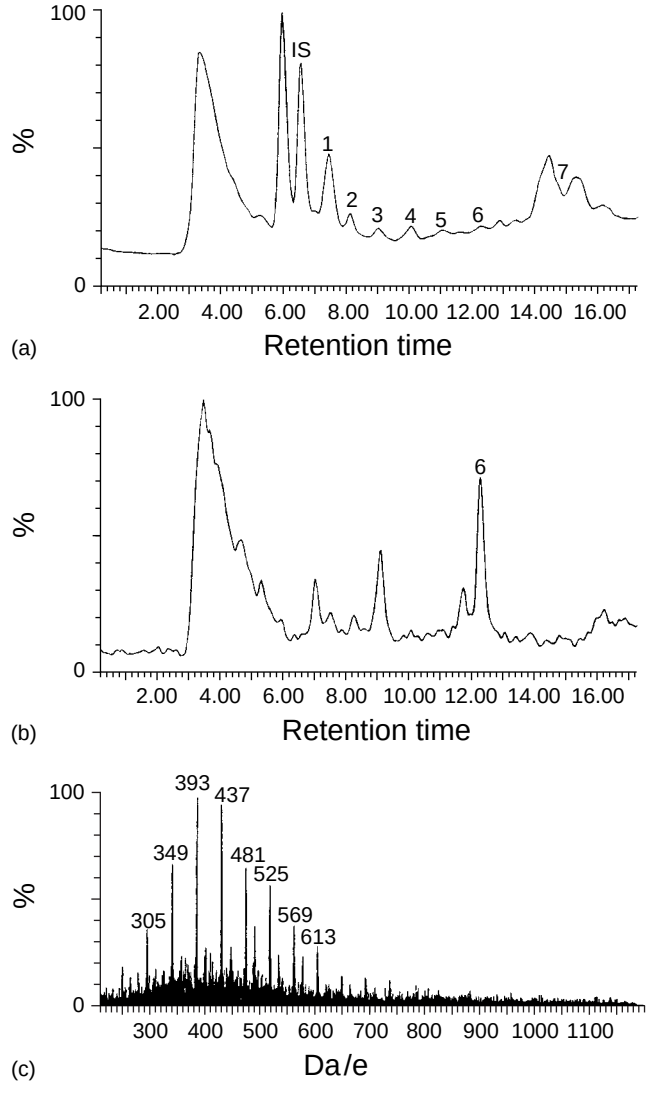

Figure 11 (a) LC/MS chromatogram obtained by injecting 3/100 of the final extract of a tap water sample (4 L). (b) Extracted chromatogram for *m/z* 441, 485, 529, and 573 relative to the most abundant $[M + Na]^+$ adducts of C-16 EO oligomers. Peak numbering is the same as in Figure 10. As measured by this method, the concentrations ($ng\,L^{-1}$) were as follows: NPEOs, 61; C-12 EOs, 10; C-13 EOs, 5.6; C-14 EOs, 5.1; C-15 EOs, 1.7; C-16 EOs, 1.2; and C-18 EOs, 0.5. (c) Mass spectra of the initial part of the chromatogram showing the presence of biointermediates of polyethoxylate surfactants, i.e. PEGs, in the water sample analyzed. As an example, *m/z* 393 is relative to a Na^+ adduct of octaethylene glycol. (Reproduced by permission of American Chemical Society from C. Crescenzi, A. Di Corcia, R. Samperi, A. Marcomini, *Anal. Chem.*, **67**, 1797–1804 © 1995.)

toxicity and environmental impact assessment cannot be adequately observed on the basis of the total concentration of an element.

On-line coupling of LC with ICP/MS permits the elemental detection capabilities of plasma MS to be added to the temporal information for the differentiation of species containing the same element, thus performing speciation analysis.[10] Another advantage of LC/ICP/MS is the possibility of reducing the spectral and non-spectral interferences, which in some cases can represent a severe limitation in the ICP/MS technique. The most commonly used chromatographic mechanisms are IC and IP chromatography; in some cases, the usual RP partition mechanism has been applied and size-exclusion chromatography (SEC) has been proposed for the separation of metals bound to large molecules. The problems due to the chromatographic requirements in LC/ICP/MS have been discussed in section 1.2.

Despite the potential applications, LC/ICP/MS cannot yet be considered a routine technique, owing to the complexity and high costs of the instrumentation and to experimental difficulties, particularly for quantitative analysis; nevertheless, interesting applications to environmental analytical problems have been reported in recent literature (Table 18).[88–94]

An isotope dilution mass spectrometric (IDMS) method has been proposed for the simultaneous determination of the complexes of 11 heavy metals (Ag, Cd, Cu, Mo, Ni, Pb, Tl, U, W, Zn, and Zr) with humic substances (HS).[91] The application of this technique involves the addition to the sample of a known amount of a spike enriched in one of the isotopes of the element under investigation; after mixing of the sample and the spike, the resulting isotope ratio is measured with the mass spectrometer. The use of IDMS for calibration requires that two isotopes of the element to be assayed are free from interference, unless it is possible to correct them. This calibration strategy reduces the influences of short-term fluctuations, long-term drifts and matrix effects on the accuracy of the determination, since isotope ratios instead of absolute responses are used as experimental measurements. Therefore it is particularly suitable for the LC/ICP/MS technique, for which long-term drifts are not negligible owing to the long analysis times. The samples considered were brown water, i.e. natural water characterized by a high content of humic acids, groundwater, sewage and seepage water and a sample of isolated fulvic acids. The total concentrations of the heavy metals and the concentrations of the metal complexes with HS were determined, obtaining detection limits in the 5–110 $ng\,L^{-1}$ range. Figure 12 shows the HPLC/ICP/IDMS chromatograms obtained using SEC and monitoring the signal of copper, zinc and molybdenum respectively, and the corresponding UV absorption profiles of a sewage water sample. The UV curve shows peaks corresponding to different fractions of HS over a broad range of molecular sizes. As it can be seen from the MS profiles, different metals are associated to single HS fractions, indicating specific

Table 18 Selected applications of LC/ICP/MS techniques for the analysis of metals, metal complexes and organometallic compounds in environmental water samples

Analyte	Sample	Separation/quantitation method	Ref.
Cu, Mo	River water	SEC/ICP/IDMS	90
Ag, Cd, Cu, Mo, Ni, Pb	Sewage water	SEC/ICP/IDMS	91
Tl, U, W, Zn, Zr	Seepage water from soil Groundwater		
As, Cr	Wastewater Drinking water	IC/ICP/MS	92
Cr, Cr complexes	Azo dyes (Acid Blue 158, Acid Blue 193)	IC/ICP/MS RP/ICP/MS	93
ORGSn	Water	RP/IP/ICP/MS	94
ORGSn	Sediment	RP/ICP/MS RP/APCI/MS	97
As	Certified reference materials (oyster, mussel)	IC/IS/MS/MS	96

interactions between each metal and HS of a particular molecular size.

Owing to the widespread presence of chromium and arsenic compounds in the environment via industrial processes, speciation of arsenic and chromium is of particular interest, since the toxicity of these elements greatly depends upon their chemical form. The inorganic arsenic species are highly toxic, whereas the organometallic compounds are less toxic or relatively nontoxic; as for chromium, Cr(VI) is known to be very toxic and carcinogenic, whereas Cr(III) is essential to mammals. IC with ICP/MS detection has been used for the simultaneous analysis of the toxic species Cr(VI), As(III) and As(V) in presence of Cr(III) and organoarsenic compounds.[92] Cr(III) and arsenic species did not interfere with the determination, the former being eliminated before analysis and the latter being separated from inorganic species of arsenic. The separation of Cr(VI) and the arsenic and organoarsenic species arsenobetaine (AB), monomethylarsonic acid (MMA) and dimethylarsonic acid (DMA) is shown in Figure 13. The detection limits (three times the standard deviation of the noise) were $0.5\,\mathrm{mg\,L^{-1}}$ for Cr(VI), As(III) and As(V). The method developed was applied to the analysis of wastewater and drinking water.

LC/ICP/MS has been used for the speciation of the chromium-based azo dyes Acid Blue 193 and Acid Blue 158 reported in Figure 14,[93] with the aim to establish whether the chromium-complex dyes considered are stable in the environment as used, or if they release the metal under real chemical, physical and microbiological conditions. Cr(III) and Cr(VI) were separated on a IC column, whereas the separation of the free and complexed forms of chromium in Acid Blue 193 samples was performed in RP LC. Figure 15 shows the chromatogram of this dye with ICP/MS detection; to avoid the interference of ArC^+ at m/z 52, the signal of chromium at m/z 53 was monitored. The second peak in the chromatogram has the same retention time as that of Cr(III) standard and therefore can be attributed to uncomplexed metal. The chromatogram of Acid Blue 193 obtained under the same conditions, but using UV/VIS detection shows two peaks corresponding to the first and third peaks of the ICP/MS chromatogram; these peaks can thus be assigned to species containing both the metal and the organic species. A third peak, which was not present in the ICP/MS profile, appears in the UV/VIS chromatogram; this indicates that the chromium metal was not incorporated in the dye.

ORGSn represent other analytes of interest from a toxicological point of view. As a consequence of the use of tributyltins (TBTs) and triphenyltins (TPhTs) as antifouling paints for fish nets and ship hulls, a considerable release of these compounds into the aquatic environment occurs. RP IP was the LC mechanism employed for the separation of tri-substituted tins from mono- and di-substituted forms; in particular, the method was developed for the determination of TBT and TPhT in water samples.[94] ICP/MS detection was carried out by monitoring the tin major isotope at m/z 120. Figure 16 shows the IP chromatogram of a triethyl-, triphenyl- and TBT mixture. Since the detection limit achieved with the system used ($200\,\mathrm{ng\,L^{-1}}$) was inadequate for applicative purposes, a preconcentration procedure using C_{18} cartridges was applied on spiked water samples. In such a way, near 100% recoveries were obtained, reaching detection limits of typically $2\,\mathrm{ng\,L^{-1}}$ for both TBT and TPhT.

In addition to HPLC/ICP/MS methods, analytical methodologies based on the use of ES/IS techniques have been proposed for organometal analysis.[95–97] The development of LC/API/MS methods originates from the lack of molecular information on the organometallic species,

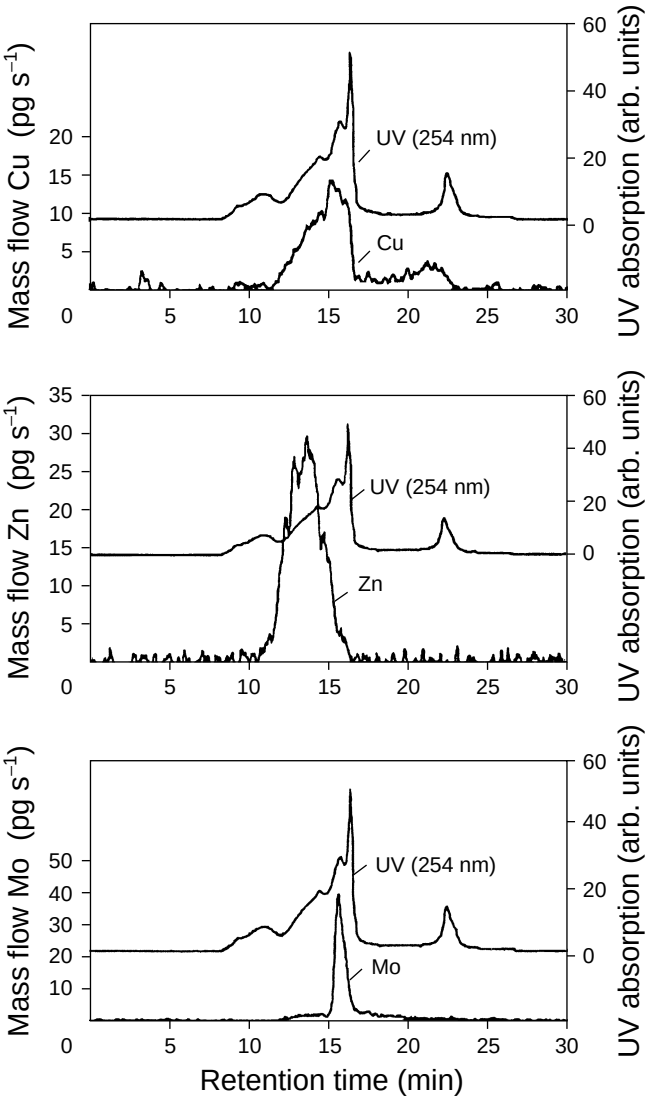

Figure 12 HPLC/ICP/IDMS chromatograms of copper, zinc and molybdenum, together with the UV absorption curve, of a sewage water sample using SEC separation (column: TSK Gel 3000 PW). (Reproduced by permission of Springer-Verlag from J. Vogl, K.G. Heumann, *Fresenius' J. Anal. Chem.*, **359**, 438–441 © 1997.)

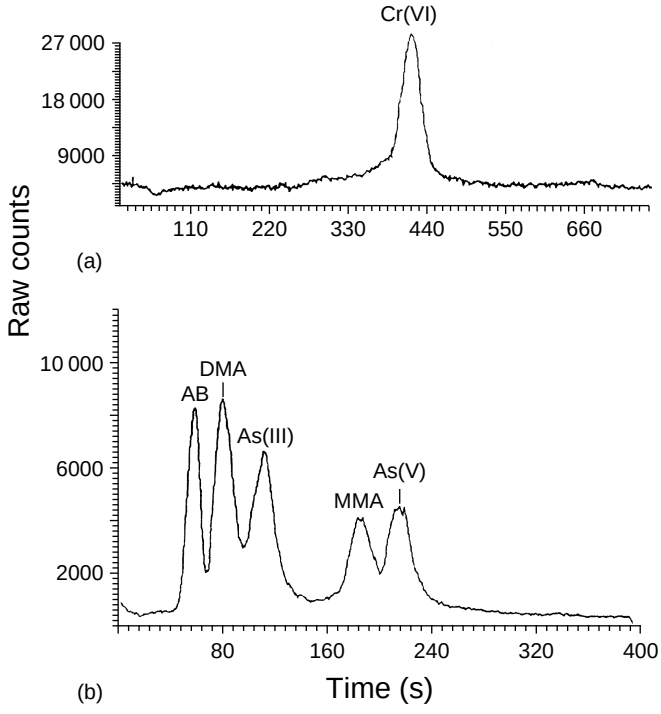

Figure 13 Chromatograms for the separation of (a) 10 μg L^{-1} Cr(VI) at *m/z* 52 and (b) 10 μg L^{-1} AB, DMA, As(III) and 6 μg L^{-1} MMA and As(V) at *m/z* 75 using isocratic conditions with 20 mM KNO_3 eluent at pH 9.8. (Reproduced by permission of Elsevier Science Publishers B.V. from M. Pantsar-Kallio, P.K.G. Manninen, *J. Chromatogr. A*, **779**, 139–146 © 1997.)

which is the main drawback of ICP/MS. A pioneering study on the use of IS/MS/MS in FIA mode for the characterization of organotin species is that of Siu et al.;[95] this technique proved to be extremely sensitive for TBT determination in environmental samples, the minimum detectable amount being 0.2 μg of Sn g^{-1} of sediment. On the other hand, FIA of the organometallic compounds has serious limitations when used for the assay of real sample extracts; using API techniques, the HPLC separation of the analytes is needed to achieve useful information for organometallic speciation in environmental samples. Methods based on the use of LC coupled with API/MS techniques have been applied to the organoarsenic speciation in marine certified reference materials[96] and recently to the organotin speciation in sediments.[97] In this work, a valuable chromatographic method using chromatographic conditions compatible with both APCI/MS and ICP/MS systems has been proposed for the determination of dibutyltin, TBT, diphenyltin and TPhT in sediment samples.[97] The two LC/MS techniques were shown to be useful in providing complementary information about the organotin species present in environmental samples, combining the molecular information of the intact organometallic compounds with the high selectivity and sensitivity of ICP/MS. Using LC/ICP/MS, detectability of the organotin species was approximately 200 pg mL^{-1}, whereas the LC/APCI/MS technique provided higher detection limits, being in the 50–100 ng mL^{-1} range.

3 SAMPLE TREATMENT: OFF-LINE AND ON-LINE PROCEDURES

Every instrumental method requires a chemical preconcentration if it does not fulfill the required limit

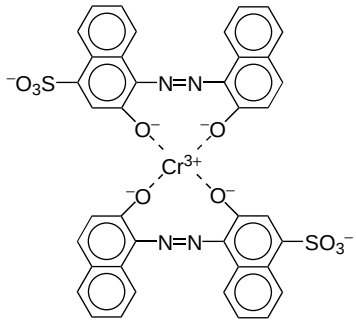

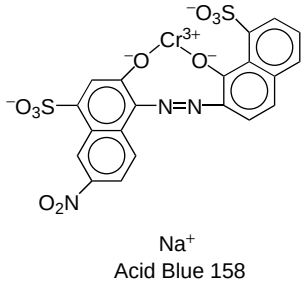

Figure 14 Structures of Acid Blue 193 and 158. (Reproduced by permission of Elsevier Science Publishers B.V. from G.K. Zoorob, J.A. Caruso, *J. Chromatogr. A*, **773**, 157–162 © 1997.)

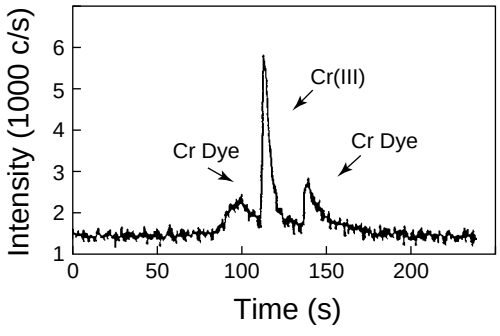

Figure 15 Chromatogram of 50 ng of Acid Blue 193 by RP LC. The mobile phase composition was water–methanol (80 : 20, v/v). The flow-rate was 1 mL min^{-1}. (Reproduced by permission of Elsevier Science Publishers B.V. from G.K. Zoorob, J.A. Caruso, *J. Chromatogr. A*, **773**, 157–162 © 1997.)

of detection. Recently, attention has been paid to the development and application of off-line and on-line preconcentration procedures in environmental analyses. With on-line methods, sample treatment generally involves SPE. The schematic apparatus involving SPE followed by LC/MS is shown in Figure 17. In such systems, the preconcentration of analytes from 50–250 mL water samples is obtained using a small cartridge packed with a

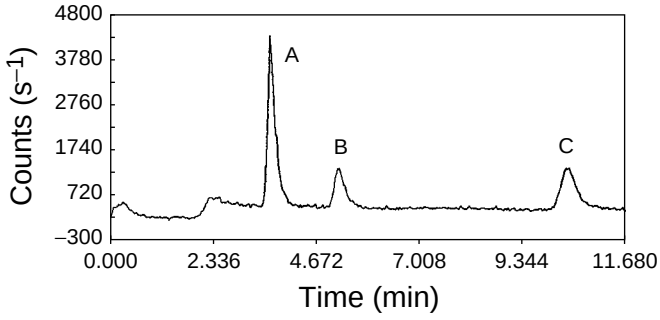

Figure 16 IP chromatogram of triethyltin (A), TPhT (B) and TBT (C) separated on the Kromasil-100 column using a mobile phase of 84% (v/v) methanol, 0.3% (v/v) triethylamine, 2.5% (v/v) acetic acid and 10 mg/100 mL oxalic acid; 100-µL injection of a mixture of 4.81 ng mL^{-1} triethyltin and 4.76 ng mL^{-1} of both triphenyl and TBT as their salts. (Reproduced by permission of Elsevier Science Publishers B.V. from B. Fairman, T. Catterick, B. Wheals, E. Polinina, *J. Chromatogr. A*, **758**, 85–92 © 1997.)

hydrophobic sorbent and performing subsequent gradient elution of the analytes extracted over an analytical column; detection is usually carried out with a UV DAD or MS. Another approach involves the use of single-column LC systems, in which analyte enrichment and chromatographic separation are performed on one short analytical column; the combination of this fast procedure with MS or MS/MS detection has been demonstrated to be quite promising.[48]

The on-line coupling of sampling treatment with LC/MS techniques represents an important advance in the development of efficient methods of extraction and analysis for organic pollutants from aqueous samples

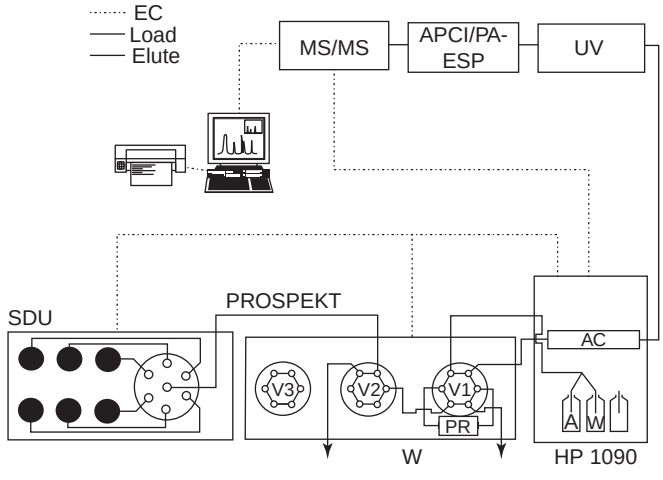

Figure 17 Schematic representation of the automated on-line SPE/LC/API/MS/MS system. (Reproduced by permission of Elsevier Science Publishers B.V. from J. Slobodník, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, *J. Chromatogr. A*, **741**, 59–74 © 1996.)

at trace levels. Besides sensitivity enhancement, advantages of the on-line approaches are sample throughput increase and automation capabilities, which make the overall LC/MS set-up routine for environmental analysis. Using adequate on-line trace-enrichment procedures, an improvement of analyte detectability of one or two orders of magnitude can be achieved and detection limits at nanogram per liter levels have been obtained for polar pesticides in water samples.[41,43,45,48–50,53] On-line monitoring systems referred to polar pesticides have been extensively discussed in the reviews of Slobodník et al.[27] and of Brouwer et al.[98]

Compared with on-line SPE, off-line SPE is more laborious and time-consuming; in addition, the problems which can be encountered in the application of this procedure are poor recoveries due to analyte losses during the solvent evaporation step. Conversely, because of the ''closed'' nature of an on-line set-up, contamination of the extract during treatment is reduced and losses of the target analytes do not occur. In addition, interferences present in the cartridges or the matrix have to be taken into account during the preconcentration of pollutants from water.

On-line preconcentration techniques have been also applied in LC/ICP/MS analysis of metals and organometallic compounds.[88] Even though ICP/MS is one of the most sensitive detection system for HPLC and generally it does not require trace enrichment, in speciation analysis there is a demand for ever lower detection limits. An off-line sample preconcentration approach has been recently reported for the determination of TBT and TPhT in water,[94] as discussed in section 2.5.

ACKNOWLEDGMENTS

The author gratefully thanks Professor Alessandro Mangia (University of Parma, Parma, Italy) for his valuable discussions and suggestions in the preparation of this manuscript and Dr. Lisa Elviri (University of Parma, Parma, Italy) for technical assistance.

ABBREVIATIONS AND ACRONYMS

A	Amines
AA	Amino Aromatics
AB	Arsenobetaine
AEO	Aliphatic Ethoxylate Alcohol
AES	Alkylethoxysulfates
APCI	Atmospheric Pressure Chemical Ionization
APEO	Alkylphenol Ethoxylate
API	Atmospheric Pressure Ionization
AS	Alkylsulfates
BD	Benzidine
CA	Chloroanilines
CAS RN	Chemical Abstract Service Registry Number
CCOOH	Halogenated Carboxyl Compounds
CE	Chloroethers
CFFAB	Continuous-flow Fast Atom Bombardment
CFLSI/MS	Continuous Flow Liquid Secondary Ion/Mass Spectrometry
CID	Collision-induced Dissociation
CLC	Capillary Liquid Chromatography
CNB	Chloronitrobenzenes
CNT	Chloronitrotoluenes
COH	Halogenated Hydroxyl Compounds
CP	Chlorophenols
CT	Chlorotoluidines
DAD	Diode-array Detector
DLI	Direct Liquid Introduction
DMA	Dimethylarsonic Acid
DNOC	4,6-Dinitro-*o*-cresol
EC	European Community
EI	Electron Impact
EO	Ethoxylate
EPA	Environmental Protection Agency
ES	Electrospray
ES/IS	Electrospray/Ionspray
ES/MS	Electrospray/Mass Spectrometry
FIA	Flow Injection Analysis
FIA/MS	Flow Injection Analysis/Mass Spectrometry
FIA/MS/MS	Flow Injection Analysis/Tandem Mass Spectrometry
FIA/TS/MS	Flow Injection Analysis/ Thermospray/Mass Spectrometry
FTICR	Fourier Transform Ion Cyclotron Resonance
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HN/APCI	Heated Nebulizer/Atmospheric Pressure Chemical Ionization
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
HS	Humic Substances
IC	Ion Chromatography
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
IDMS	Isotope Dilution Mass Spectrometric
INORG	Inorganic – Metals

IP	Ion-pair
IS	Ionspray
IT	Ion Trap
ITMS	Ion Trap Mass Spectrometer
LAS	Linear Alkylbenzenesulfonates
LC	Liquid Chromatography
LC/APCI/MS	Liquid Chromatography/Atmospheric Pressure Chemical Ionization/Mass Spectrometry
LC/HFIS/MS	Liquid Chromatography/High Flow Ionspray/Mass Spectrometry
LC/ICP/MS	Liquid Chromatography/Inductively Coupled Plasma/Mass Spectrometry
LC/IS/MS	Liquid Chromatography/ Ionspray/Mass Spectrometry
LC/MS	Liquid Chromatography/ Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LC/TS/MS	Liquid Chromatography/ Thermospray/Mass Spectrometry
LOD	Limits of Detection
LOQ	Limit of Quantitation
MALDI	Matrix Assisted Laser Desorption/Ionization
MB	Moving Belt
MCPA	(4-Chloro-2-methylphenoxy)acetic Acid
MDLs	Method Detection Limits
MISC	Miscellaneous
MMA	Monomethylarsonic Acid
MS	Mass Spectrometry
MS^n	Multiple-stage Tandem Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NI	Negative-ion
NIST	National Institute of Standards and Technology
NP	Normal-phase
NPEO	Nonylphenol Ethoxylate
OCP	Organochloro Pesticides
OPEO	Octylphenol Ethoxylate
OPP	Organophosphorus Pesticides
ORGSn	Organotin Compounds
PAES	Pneumatically-assisted Electrospray
PAH	Polycyclic Aromatic Hydrocarbon
PB	Particle Beam
PB/MS	Particle Beam/Mass Spectrometry
PEG	Poly(ethylene Glycol)
PhAP	Phenoxy Acid Pesticides
PI	Positive-ion
PPG	Polypropylenglycol
PSS	Phase-system Switching
PU	Phenylurea Compounds
RIC	Reconstructed Ion Chromatogram
RP	Reversed-phase
SEC	Size-exclusion Chromatography
SIM	Selective Ion Monitoring
S/N	Signal-to-noise
SPE	Solid-phase Extraction
SRM	Selective Reaction Monitoring
STP	Sewage Treatment Plant
SVHO	Semivolatile Halogenated Organics
TBT	Tributyltin
TOF	Time-of-flight
TOFMS	Time-of-flight Mass Spectrometry
TPhT	Triphenyltin
TRIA	Triazines
TS	Thermospray
TS/MS	Thermospray/Mass Spectrometry
TS/MS/MS	Thermospray Tandem Mass Spectrometry
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VA	Volatile Aromatics
VHO	Volatile Halogenated Organics
2,4-D	Dichlor Phenoxyacetic Acid
2,4,5-TP	Fenoprop

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Dyes, Environmental Analysis of • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Organometallic Compound Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. V.L. Tal'roze, V.E. Skurat, G.V. Karpov, 'Use of a Mass Spectrometer with a Capillary System for Admitting the Sample as a Liquid Chromatograph Detector', *J. Phys. Chem.*, **73**(2), 452–453 (1969).
2. E.D. Lee, J.D. Henion, 'Micro-liquid Chromatography/Mass Spectrometry with Direct Liquid Introduction', *J. Chromatogr. Sci.*, **23**(6), 253–264 (1985).
3. W.H. McFadden, H.L. Schwartz, S. Evans, 'Direct Analysis of Liquid Chromatographic Effluents', *J. Chromatogr.*, **122**, 389–396 (1976).
4. P.C. Winkler, D.D. Perkins, W.K. Williams, R.F. Browner, 'Performance of an Improved Monodisperse Aerosol Generation Interface for Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **60**(5), 489–493 (1988).
5. C.R. Blakley, J.J. Carmody, M.L. Vestal, 'A New Soft Ionization Technique for Mass Spectrometry of Complex Molecules', *J. Am. Chem. Soc.*, **102**(18), 5931–5933 (1980).
6. R.M. Caprioli, T. Fan, J.S. Cottrell, 'A Continuous-flow Sample Probe for Fast Atom Bombardment Mass Spectrometry', *Anal. Chem.*, **58**(14), 2949–2954 (1986).
7. C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, 'Electrospray Interface for Liquid Chromatographs and Mass Spectrometers', *Anal. Chem.*, **57**(3), 675–679 (1985).
8. A.P. Bruins, T.R. Covey, J.D. Henion, 'Ionspray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**(22), 2642–2646 (1987).
9. E.C. Horning, M.G. Horning, D.I. Carroll, I. Dzidic, R.N. Stillwell, 'New Picogram Detection System Based on a Mass Spectrometer with an External Ionization Source at Atmospheric Pressure', *Anal. Chem.*, **45**(6), 936–943 (1973).
10. J.J. Thompson, R.S. Houk, 'Inductively Coupled Plasma Mass Spectrometric Detection for Multielement Flow Injection Analysis and Elemental Speciation by Reversed-phase Liquid Chromatography', *Anal. Chem.*, **58**(12), 2541–2548 (1986).
11. C.S. Creaser, J.W. Stygall, 'Particle Beam Liquid Chromatography/Mass Spectrometry: Instrumentation and Applications', *Analyst*, **118**(12), 1467–1480 (1993).
12. W.M.A. Niessen, 'Advances in Instrumentation in Liquid Chromatography/Mass Spectrometry and Related Liquid-introduction Techniques', *J. Chromatogr. A*, **794**(1 + 2), 407–435 (1998).
13. J.J. Conboy, J.D. Henion, M.W. Martin, J.A. Zweigenbaum, 'Ion Chromatography/Mass Spectrometry for the Determination of Organic Ammonium and Sulfate Compounds', *Anal. Chem.*, **62**(8), 800–807 (1990).
14. R.J. Vreeken, W.D. van Dongen, R.T. Ghijsen, G.J. de Jong, H. Lingeman, U.A.Th. Brinkman, R.G.J. van Leuken, G.T.C. Kwakkenbos, R.S. Deelder, 'Coupling of Column Liquid Chromatography and Thermospray Mass Spectrometry via Phase-system Switching with an Ion-exchange Trapping Column', *Biol. Mass Spectrom.*, **21**(6), 305–314 (1992).
15. E.R. Verhey, H.J.E.M. Reeuwijk, W.M.A. Niessen, U.R. Tjaden, J. van der Greef, 'Phase-system Switching: A New Approach in Liquid Chromatography/Mass Spectrometry Coupling', *Biomed. Environ. Mass Spectrom.*, **16**(1–12), 393–397 (1988).
16. D. Barceló, G. Durand, R.J. Vreeken, G.J. de Jong, U.A.Th. Brinkman, 'Non Polar Solvents for Normal-phase Liquid Chromatography and Post Column Extraction in Thermospray Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **62**(15), 1696–1700 (1990).
17. R.J. Vreeken, R.T. Ghijsen, R.W. Frei, G.J. de Jong, U.A.Th. Brinkman, 'Coupling of Ion-pair Liquid Chromatography and Thermospray Mass Spectrometry via Phase-system Switching with a Polymeric Trapping Column', *J. Chromatogr.*, **654**(1), 65–77 (1993).
18. R.J. Vreeken, 'On-line Post-column Liquid–Solid and Liquid–Liquid Extractions in Environmental Liquid Chromatography–Thermospray/Mass Spectrometry', in *Applications of LC/MS in Environmental Chemistry*, ed. D. Barceló, Journal of Chromatography Library, Elsevier, Amsterdam, 447–480, Vol. 59, 1996.
19. N.P. Vela, L.K. Olson, J.A. Caruso, 'Elemental Speciation with Plasma Mass Spectrometry', *Anal. Chem.*, **65**(13), 585A–597A (1993).
20. P.C. Uden, 'Element-specific Chromatographic Detection by Atomic Absorption, Plasma Atomic Emission and Plasma Mass Spectrometry', *J. Chromatogr. A*, **703**(1 + 2), 393–416 (1995).
21. J.V. Johnson, R.A. Yost, 'Tandem Mass Spectrometry for Trace Analysis', *Anal. Chem.*, **57**(7), 758A–768A (1985).
22. W.M.A. Niessen, 'Developments in Interface Technology for Combined Liquid Chromatography, Capillary Electrophoresis and Supercritical Fluid Chromatography/Mass Spectrometry', in *Applications of LC/MS in Environmental Chemistry*, ed. D. Barceló, Journal of Chromatography Library, Elsevier, Amsterdam, 3–70, Vol. 59, 1996.
23. D. Barceló, 'Occurrence, Handling and Chromatographic Determination of Pesticides in the Aquatic Environment', *Analyst*, **116**(7), 681–689 (1991).
24. D. Barceló, 'Mass Spectrometry in Environmental Organic Analysis', *Anal. Chim. Acta*, **263**, 1–19 (1992).
25. J.F. Garcia, D. Barceló, 'An Overview of LC/MS Interfacing Systems with Selected Applications', *J. High Resolut. Chromatogr.*, **16**(11), 633–641 (1993).

26. M. Careri, P. Manini, M. Maspero, 'Liquid Chromatography–Mass Spectrometry in Environmental Analysis', *Ann. Chim.*, **84**(12), 475–508 (1994).
27. J. Slobodník, B.L.M. van Baar, U.A.Th. Brinkman, 'Column Liquid Chromatography–Mass Spectrometry: Selected Techniques in Environmental Applications for Polar Pesticides and Related Compounds', *J. Chromatogr. A*, **703**(1 + 2), 81–121 (1995).
28. A.T. Kiewiet, P. de Voogt, 'Chromatographic Tools for Analyzing and Tracking Non-ionic Surfactants in the Aquatic Environment', *J. Chromatogr. A*, **733**(1 + 2), 185–192 (1996).
29. T. Reemtsma, 'Methods of Analysis of Polar Aromatic Sulfonates from Aquatic Environments', *J. Chromatogr. A*, **733**(1 + 2), 473–489 (1996).
30. T. Cserháti, E. Forgács, 'Separation and Quantitative Determination of Non-ionic Surfactants Used as Pesticide Additives', *J. Chromatogr. A*, **774**(1 + 2), 265–279 (1997).
31. J. Riu, I. Schoensee, D. Barceló, C. Ráfols, 'Determination of Sulphonated Azo Dyes in Water and Wastewater', *Trends Anal. Chem.*, **16**(7), 405–419 (1997).
32. A. Di Corcia, 'Characterization of Surfactants and Their Biointermediates by Liquid Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **794**(1 + 2), 165–185 (1998).
33. M.A. Brown (ed.), 'Liquid Chromatography/Mass Spectrometry Applications' in *Agricultural, Pharmaceutical and Environmental Chemistry*, ACS Symposium Series, American Chemical Society, Washington, DC, 1–300, Vol. 420, 1990.
34. M.H. Lamoree, R.T. Ghijsen, U.A.Th. Brinkman, *Environmental Analysis: Techniques, Applications and Quality Assurance*, ed., D. Barceló, Elsevier, Amsterdam, 521, 1993.
35. W.M.A. Niessen, *Liquid Chromatography–Mass Spectrometry*, Second edition, Revised and Expanded, Chromatographic Science Series, Marcel Dekker, New York, Vol. 79, 1999.
36. EEC Council Directive 80/778/EEC, *Off. J. European Communities*, No. L229, August 30, 11, 1980.
37. US EPA Drinking Water Regulations, EPA-822-B-96-002, October, 1996.
38. T.R. Covey, E.D. Lee, A.P. Bruins, J.D. Henion, 'Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **58**(14), 1451A–1461A (1986).
39. K.S. Chiu, A. Van Langenhove, C. Tanaka, 'High-performance Liquid Chromatographic–Mass Spectrometric and High-performance Liquid Chromatographic–Tandem Mass Spectrometric Analysis of Carbamate Pesticides', *Biomed. Environ. Mass Spectrom.*, **18**(3), 200–206 (1989).
40. D. Barceló, G. Durand, R.J. Vreeken, G.J. de Jong, H. Lingeman, U.A.Th. Brinkman, 'Evaluation of Eluents in Thermospray Liquid Chromatography–Mass Spectrometry for Identification and Determination of Pesticides in Environmental Samples', *J. Chromatogr.*, **553**(1–2), 311–328 (1991).
41. S. Chiron, S. Dupas, P. Scribe, D. Barceló, 'Application of On-line Solid-phase Extraction Followed by Liquid Chromatography–Thermospray Mass Spectrometry to the Determination of Pesticides in Environmental Waters', *J. Chromatogr. A*, **665**(2), 295–305 (1994).
42. A. Cappiello, G. Famiglini, F. Bruner, 'Determination of Acidic and Basic/Neutral Pesticides in Water with a New Microliter Flow Rate LC/MS Particle Beam Interface', *Anal. Chem.*, **66**(9), 1416–1423 (1994).
43. S. Chiron, S. Papilloud, W. Haerdi, D. Barceló, 'Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography-high-flow Pneumatically Assisted Electrospray Mass Spectrometry for the Determination of Acidic Herbicides in Environmental Waters', *Anal. Chem.*, **67**(9), 1637–1643 (1995).
44. C. Crescenzi, A. Di Corcia, S. Marchese, R. Samperi, 'Determination of Acidic Pesticides in Water by a Benchtop Electrospray Liquid Chromatography Mass Spectrometer', *Anal. Chem.*, **67**(13), 1968–1975 (1995).
45. C. Aguilar, I. Ferrer, F. Borrull, R.M. Marcé, D. Barceló, 'Comparison of Automated On-line Solid-phase Extraction Followed by Liquid Chromatography–Mass Spectrometry with Atmospheric Pressure Chemical Ionization and Particle Beam Mass Spectrometry for the Determination of a Priority Group of Pesticides in Environmental Waters', *J. Chromatogr. A*, **794**(1 + 2), 147–163 (1998).
46. B. Køppen, N.L. Spliid, 'Determination of Acidic Herbicides Using Liquid Chromatography with Pneumatically Assisted Electrospray Ionization Mass Spectrometry and Tandem Mass Spectrometric Detection', *J. Chromatogr. A*, **803**(1 + 2), 157–168 (1998).
47. C. Molina, G. Durand, D. Barceló, 'Trace Determination of Herbicides in Estuarine Waters by Liquid Chromatography-high-flow Pneumatically Assisted Electrospray Mass Spectrometry', *J. Chromatogr. A*, **712**(1), 113–122 (1995).
48. J. Slobodník, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Trace-level Determination of Pesticide Residues using On-line Solid-phase Extraction-column Liquid Chromatography with Atmospheric Pressure Ionization Mass Spectrometric and Tandem Mass Spectrometric Detection', *J. Chromatogr. A*, **741**(1), 59–74 (1996).
49. A.C. Hogenboom, P. Speksnijder, R.J. Vreeken, W.M.A. Niessen, U.A.Th. Brinkman, 'Rapid Target Analysis of Microcontaminants in Water by On-line Single-short-column Liquid Chromatography Combined with Atmospheric Pressure Ionization Tandem Mass Spectrometry', *J. Chromatogr. A*, **777**(1), 81–90 (1997).
50. D. Volmer, K. Levsen, G. Wünsch, 'Thermospray Liquid Chromatographic–Mass Spectrometric Multi-residue Determination of 128 Polar Pesticides in Aqueous Environmental Samples', *J. Chromatogr. A*, **660**(1 + 2), 231–248 (1994).
51. M. Honing, J. Riu, D. Barceló, B.L.M. van Baar, U.A.Th. Brinkman, 'Determination of Ten Carbamate Pesticides

in Aquatic and Sediment Samples by Liquid Chromatography–Ionspray and Thermospray Mass Spectrometry', *J. Chromatogr. A*, **733**(1 + 2), 283–294 (1996).
52. S. Lacorte, D. Barceló, 'Determination of Part per Trillion Levels of Organophosphorus Pesticides in Groundwater by Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry Using Positive and Negative Ion Modes of Operation', *Anal. Chem.*, **68**(15), 2464–2470 (1996).
53. S. Lacorte, C. Molina, D. Barceló, 'Temperature and Extraction Voltage Effect on Fragmentation of Organophosphorus Pesticides in Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **795**(1), 13–26 (1998).
54. R.D. Voyksner, T. Pack, 'Investigation of Collisional-activation Decomposition Process and Spectra in the Transport Region of an Electrospray Single-quadrupole Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **5**(6), 263–268 (1991).
55. M.L. Lee, M.B. Novotny, K.D. Bartle, *Analytical Chemistry of Polycyclic Aromatic Compounds*, Academic Press, New York, 1981.
56. H. Perreault, L. Ramaley, P.G. Sim, F.M. Benoit, 'Use of a Moving-belt Interface for On-line High-performance Liquid Chromatography/Mass Spectrometry Characterization of Polycyclic Aromatic Compounds of Molecular Weight up to 580 Da in Environmental Samples', *Rapid Commun. Mass Spectrom.*, **5**(12), 604–610 (1991).
57. M.A. Brown, R.D. Stephens, I.S. Kim, 'Liquid Chromatography–Mass Spectrometry – A New Window for Environmental Analysis', *Trends Anal. Chem.*, **10**(6), 330–336 (1991).
58. D.R. Doerge, J. Clayton, P.P. Fu, D.A. Wolfe, 'Analysis of Polycyclic Aromatic Hydrocarbons Using Liquid Chromatography/Particle Beam Mass Spectrometry', *Biol. Mass Spectrom.*, **22**(11), 654–660 (1993).
59. R.P. Singh, I.D. Brindle, T.R.B. Jones, J.M. Miller, M. Chiba, 'Determination of Polycyclic Aromatic Hydrocarbons by Liquid Chromatography–Particle Beam–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**(11), 898–905 (1993).
60. J.F. Anacleto, L. Ramaley, F.M. Benoit, R.K. Boyd, M.A. Quilliam, 'Comparison of Liquid Chromatography/Mass Spectrometry Interfaces for the Analysis of Polycyclic Aromatic Compounds', *Anal. Chem.*, **67**(22), 4145–4154 (1995).
61. G.J. Van Berkel, S.A. McLuckey, G.L. Glish, 'Performing Ions in Solution via Charge-transfer Complexation for Analysis by Electrospray Ionization', *Anal. Chem.*, **63**(18), 2064–2068 (1991).
62. G.J. Van Berkel, S.A. McLuckey, G.L. Glish, 'Electrochemical Origin of Radical Cations Observed in Electrospray Ionization Mass Spectra', *Anal. Chem.*, **64**(14), 1586–1593 (1992).
63. L. Bonfanti, M. Careri, A. Mangia, P. Manini, M. Maspero, 'Simultaneous Identification of Different Classes of Hydrocarbons and Determination of Nitro-polycyclic Aromatic Hydrocarbons by means of Particle Beam Liquid Chromatography–Mass Spectrometry', *J. Chromatogr. A.*, **728**(1 + 2), 359–369 (1996).
64. M.S. Reisch, 'Better Times Ahead for US Dye Producers', *Chem. Eng. News*, **66**(29), 7–14 (1988).
65. S.V. Kulkarni, C.D. Blackwell, A.L. Blakard, C.W. Stackhouse, M.W. Alexander, 'Textile Dyes and Dyeing Equipment: Classification, Properties and Environmental Aspect', EPA-600/2-85/010, Research Triangle Park, NC, US Environmental Protection Agency, 1985.
66. M. Boeninger, 'The Carcinogenicity and Metabolism of Azo Dyes, Especially Those Derived from Benzidine', Publication No. 80-119, US Department of Health and Human Services, National Institute for Occupational Safety and Health, 1980.
67. EPA Report 600/R-92/129, 'Methods for the Determination of Organic Compounds in Drinking Water', Supplement II, US Environmental Protection Agency, Washington, DC, Method 553, Rev. 1.1, August, 1992.
68. R.D. Voyksner, 'Characterization of Dyes in Environmental Samples by Thermospray High-performance Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **57**(13), 2600–2605 (1985).
69. J.M. Ballard, L.D. Betowski, 'Thermospray Ionization and Tandem Mass Spectrometry of Dyes', *Org. Mass Spectrom.*, **21**(9), 575–588 (1986).
70. L.D. Betoswki, S.M. Pyle, J.M. Ballard, G.M. Shaul, 'Thermospray LC/MS/MS Analysis of Wastewater for Disperse Azo Dyes', *Biomed. Environ. Mass Spectrom.*, **14**(7), 343–354 (1987).
71. A.P. Bruins, L.O.G. Weidolf, J.D. Henion, W.L. Budde, 'Determination of Sulfonated Azo Dyes by Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**(22), 2647–2652 (1987).
72. P.O. Edlund, E.D. Lee, J.D. Henion, W.L. Budde, 'Determination of Sulphonated Azo Dyes in Municipal Waste Water by Ion Spray Liquid Chromatography/Tandem Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **18**(4), 233–240 (1989).
73. C. Pace, M. Roby, 'Characterization and Analysis of the Base/Neutral Fraction from Azo Dyes Waste Samples', EPA Report 600/X-91/147, EMSL, Las Vegas, NV 89193, USA, 1991.
74. R. Straub, R.D. Voyksner, J.T. Keever, 'Thermospray, Particle Beam and Electrospray Liquid Chromatography–Mass Spectrometry of Azo Dyes', *J. Chromatogr.*, **627**(2), 173–186 (1992).
75. W.C. Brumley, C.M. Brownrigg, A.H. Grange, 'Capillary Liquid Chromatography–Mass Spectrometry and Micellar Electrokinetic Chromatography as Complementary Techniques in Environmental Analysis', *J. Chromatogr. A*, **680**(2), 635–644 (1994).

76. C. Ràfols, D. Barceló, 'Determination of Mono- and Disulphonated Azo Dyes by Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *J. Chromatogr. A*, **777**(1), 177–192 (1997).
77. K.G. Malle, 'Foam Formation in Waters. Mechanism and Importance of Surfactants', *Git Fachz. Lab.*, **37**(5), 422–429 (1993).
78. R.D. Swisher, *Surfactant Biodegradation*, 2nd edition, Surfactant Science Series, Marcel Dekker, New York, 692, Vol. 18, 1987.
79. H.Fr. Schröeder, 'Surfactants: Non-biodegradable, Significant Pollutants in Sewage Treatment Plant Effluents. Separation, Identification and Quantification by Liquid Chromatography, Flow-injection Analysis–Mass Spectrometry and Tandem Mass Spectrometry', *J. Chromatogr. A*, **647**(2), 219–234 (1993).
80. S.D. Scullion, M.R. Clench, M. Cooke, A.E. Ashcroft, 'Determination of Surfactants in Surface Water by Solid-phase Extraction, Liquid Chromatography and Liquid Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **733**(1 + 2), 207–216 (1996).
81. D.D. Popenoe, S.J. Morris, III, P.S. Horn, K.T. Norwood, 'Determination of Alkyl Sulfates and Alkyl Ethoxysulfates in Wastewater Treatment Plant Influents and Effluents and in River Water Using Liquid Chromatography/Ion Spray Mass Spectrometry', *Anal. Chem.*, **66**(10), 1620–1629 (1994).
82. H.Fr. Schröeder, 'Analysis of Polar Organic Pollutants in the Elbe River by Flow Injection Analysis and High-performance Liquid Chromatography with Tandem Mass Spectrometry', *J. Chromatogr. A*, **777**(1), 127–139 (1997).
83. H.Fr. Schröeder, 'Separation, Identification and Quantification of Surfactants and Their Metabolites in Waste Water, Surface Water and Drinking Water by LC/TSP/MS, FIA/TSP/MS and MS/MS', in *Applications of LC/MS in Environmental Chemistry*, ed. D. Barceló, Journal of Chromatography Library, Elsevier, Amsterdam, 263–324, Vol. 59, 1996.
84. M. Ahel, W. Giger, M. Koch, 'Behavior of Alkylphenol Polyethoxylate Surfactants in the Aquatic Environment. I. Occurrence and Transformation in Sewage Treatment', *Water Res.*, **28**(5), 1131–1147 (1994).
85. J. Miyamoto, W. Klein, 'Natural and Anthropogenic Environmental Oestrogens: The Scientific Basis for Risk Assessment. Environmental Exposure, Species Differences and Risk Assessment', *Pure Appl. Chem.*, **70**(9), 1829–1845 (1998).
86. K.A. Evans, S.T. Dubey, L. Kravetz, I. Dzidic, J. Gumulka, R. Mueller, J.R. Stork, 'Quantitative Determination of Linear Primary Alcohol Ethoxylate Surfactants in Environmental Samples by Thermospray LC/MS', *Anal. Chem.*, **66**(5), 699–705 (1994).
87. C. Crescenzi, A. Di Corcia, R. Samperi, A. Marcomini, 'Determination of Nonionic Polyethoxylate Surfactants in Environmental Waters by Liquid Chromatography/Electrospray Mass Spectrometry', *Anal. Chem.*, **67**(11), 1797–1804 (1995).
88. E.M. Heithmar, T.A. Hinners, J.T. Rowan, J.M. Riviello, 'Minimization of Interferences in Inductively Coupled Plasma Mass Spectrometry using On-line Preconcentration', *Anal. Chem.*, **62**(8), 857–864 (1990).
89. J.W. McLaren, J.W.H. Cam, S.S. Berman, K. Akatsuka, M.A. Azeredo, 'On-line Method for the Analysis of Seawater for Trace Elements by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**(2), 279–286 (1993).
90. L. Rottmann, K.G. Heumann, 'Development of an On-line Isotope Dilution Technique with HPLC/ICP/MS for the Accurate Determination of Elemental Species', *Fresenius' J. Anal. Chem.*, **350**(4–5), 221–227 (1994).
91. J. Vogl, K.G. Heumann, 'Determination of Heavy Metal Complexes with Humic Substances by HPLC/ICP/MS Coupling Using On-line Isotope Dilution Technique', *Fresenius' J. Anal. Chem.*, **359**(4–5), 438–441 (1997).
92. M. Pantsar-Kallio, P.K.G. Manninen, 'Simultaneous Determination of Toxic Arsenic and Chromium Species in Water Samples by Ion Chromatography–Inductively Coupled Plasma Mass Spectrometry', *J. Chromatogr. A*, **779**(1 + 2), 139–146 (1997).
93. G.K. Zoorob, J.A. Caruso, 'Speciation of Chromium Dyes by High-performance Liquid Chromatography with Inductively Coupled Plasma Mass Spectrometric Detection', *J. Chromatogr. A*, **773**(1 + 2), 157–162 (1997).
94. B. Fairman, T. Catterick, B. Wheals, E. Polinina, 'Reversed-phase Ion-pair Chromatography with Inductively Coupled Plasma Mass Spectrometry Detection for the Determination of Organo-tin Compounds in Waters', *J. Chromatogr. A*, **758**(1), 89–92 (1997).
95. K.W.M. Siu, G.J. Gardner, S.S. Berman, 'Ionspray Mass Spectrometry/Mass Spectrometry: Quantitation of Tributyltin in a Sediment Reference Material for Trace Metals', *Anal. Chem.*, **61**(20), 2320–2322 (1989).
96. J.J. Corr, E.H. Larsen, 'Arsenic Speciation by Liquid Chromatography Coupled with Ionspray Tandem Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**(12), 1215–1224 (1996).
97. S. White, T. Catterick, B. Fairman, K. Webb, 'Speciation of Organo-tin Compounds Using Liquid Chromatography–Atmospheric Pressure Ionisation Mass Spectrometry and Liquid Chromatography–Inductively Coupled Plasma Mass Spectrometry as Complementary Techniques', *J. Chromatogr. A*, **794**(1 + 2), 211–218 (1998).
98. E.R. Brouwer, S. Kofman, U.A.Th. Brinkman, 'Selected Procedures for the Monitoring of Polar Pesticides and Related Microcontaminants in Aquatic Samples', *J. Chromatogr. A*, **703**(1 + 2), 167–190 (1995).

Luminescence in Environmental Analysis

Andres D. Campiglia
North Dakota State University, Fargo, USA

1 Introduction 1
2 Luminescence Techniques 2
2.1 Fluorimetry 2
2.2 Phosphorimetry 4
2.3 Shpol'skii Spectrometry 6
3 Field Monitoring in Environmental Analysis 7
4 New Trends in Luminescence Analysis of Environmental Samples 8
5 Final Remarks 11
Abbreviations and Acronyms 11
Related Articles 12
References 12

An overview of the applications of photoluminescence techniques for the analysis of polycyclic aromatic compounds (PACs) in environmental samples is presented. The basic experimental procedures of several techniques, such as room-temperature fluorimetry, low-temperature phosphorimetry (LTP), micelle-stabilized room-temperature phosphorimetry (MSRTP), sensitized room-temperature phosphorimetry (RTP), solid-surface room-temperature phosphorimetry (SSRTP), and Shpol'skii spectrometry at 77 K are provided. Analytical figures of merit, such as linear dynamic ranges (LDRs) of calibration curves, limits of detection, and relative standard deviations (RSDs), are compared. Basic spectrofluorimeters for direct laboratory analysis of organic pollutants and instrumental systems utilized for chromatographic detection are discussed. Emphasis to portable devices using laser excitation is given to illustrate the usefulness of photoluminescence techniques for field and remote analysis of contaminated sites. Specific examples of emerging technology developed in the author's laboratory are presented to illustrate recent advances in photoluminescence analysis of environmental samples.

1 INTRODUCTION

Environmental analysis of organic pollutants is mostly performed by chromatographic techniques. The United States Environmental Protection Agency (USEPA) 500 series for drinking water analysis is a representative example.[1] The USEPA 500 series is summarized in Table 1. Gas chromatography (GC) and high-performance liquid chromatography (HPLC) associated with a wide variety of detection techniques are the basis of USEPA methodology. Only three methods (547, 549/549.1, and 550/550.1) use fluorescence detection. The hydrophobic nature of organic pollutants leads to concentration levels below the limit of detection (LOD) of chromatographic methods. Purge-and-trap extraction (PTE), liquid–liquid extraction (LLE) and solid–liquid extraction (SLE) techniques provide preconcentration of organic pollutants. In addition to preconcentrating the sample, extraction techniques simplify matrix composition and facilitate analytical resolution in the chromatographic column.

Although chromatography provides comprehensive methodology for a wide variety of organic pollutants, the assays are tedious, laborious, time-consuming and expensive for routine analysis of numerous samples. Elution times of 30–60 min are typical and standards must be run periodically to verify retention times. If the concentrations of target species are found to lie outside the detector's response range, the sample must be diluted and the process repeated. On the other end of the concentration range, many samples are ''zeroes'', i.e. the concentrations are below the LOD. Problems arise when laboratory procedures are scaled up to handle thousands of samples under mass screening conditions.

Therefore, there is a continuous search for simple and rapid screening techniques that can tell beforehand whether environmental samples merit the full-blown chromatographic procedure. The main trust of luminescence techniques in environmental analysis is to provide screening methodology and field analysis of organic pollutants in a direct way. Field screening techniques lead to a shorter turnaround analysis time and improve the cost effectiveness for environmental monitoring and remediation. Although a small fraction of environmental pollutants are strong luminescence emitters, which provides inherent selectivity to fluorimetry and phosphorimetry, the limited applicability often can be overcome by using derivatization or indirect methods based on effects such as quenching, complexation, enzymatic catalysis, and immunochemical equilibria. For instance, an excellent review on the direct and indirect analysis of pesticides in environmental samples was reported recently.[2]

This article deals specifically with the luminescent techniques and instruments used to monitor and characterize PACs in environmental samples. PACs, which comprise

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 USEPA 500 series: methods for the analysis of organic pollutants in drinking water

Method	Organic pollutants	Method description
501	Trihalomethanes	GC/ECD or GC/PID/ELCD
502.1	Volatile halogenated organic compounds	Purge-and-trap GC
502.2	Volatile organic compounds	Purge-and-trap Capillary GC/PID/ELCD
503.1	Volatile aromatic and unsaturated organic compounds	Purge-and-trap GC
504/504.1	1,2-Dibromoethane; 1,2-dibromo-3-chloro-propane; 1,2,3-trichloropropane	Microextraction and GC
505	Organohalide pesticides and polychlorinated biphenyl products	Microextraction and GC
506	Phthalates and adipate esters	LLE or SLE and GC/PID
507	Nitrogen- and phosphorus-containing pesticides	LLE and GC/NPD
508	Chlorinated pesticides and aroclors	LLE and GC/ECD
508.1/508.A	Chlorinated pesticides, herbicides and organohalides	LSE and GC/ECD
509	Ethylene thiourea	GC/NPD
513	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin	LLE or LSE and GC with high-resolution MS
515.1	Chlorinated acids	GC/ECD
515.2	Chlorinated acids	LSE and GC/ECD
524.1/524.2	Volatile organic compounds	GC/MS
525.1	Organic compounds	LSE and capillary GC/MS
525.2	Semi-volatile organic compounds	LSE and GC/MS
531/531.1	*n*-Methylcarbamoyloximes and *n*-methylcarbamates	HPLC with postcolumn derivatization
547	Glyphosate	HPLC with postcolumn derivatization and fluorescence detection
548	Endothall	Derivatization, LSE, and GC/ECD
548.1	Endothall	GC/MS
549/549.1	Diquat and paraquat	LSE and HPLC with fluorescence detection
550	PAHs	LLE and HPLC with coupled ultraviolet and fluorescence detection
550.1	PAHs	LSE and HPLC with coupled ultraviolet and fluorescence detection
551	Chlorination disinfection by-products and chlorinated solvents	LLE and GC/ECD
552	Haloacetic acids	LLE, derivatization and GC/ECD
552.1	Haloacetic acids and dalapon	Ion-exchange LSE and GC/ECD
553	Benzidines and nitrogen-containing pesticides	LLE or LSE and HPLC with particle beam MS
554	Carbonyl compounds	Derivatization and HPLC
555	Chlorinated acids	HPLC with photodiode array ultraviolet detector

PID, photoionization detector; ELCD, electrolytic conductivity detector; GC, gas chromatography; GC/PID/ELCD, gas chromatography with photoionization and electrolytic conductivity detectors in series; LLE, liquid–liquid extraction; SLE, solid–liquid extraction; NPD, nitrogen–phosphorus detector; LSE, liquid–solid extraction; MS, mass spectrometry; HPLC, high-performance liquid chromatography; PAH, polycyclic aromatic hydrocarbon.

condensed multi-ring benzenoid compounds, are important environmental pollutants originating from a wide variety of natural and anthropogenic sources. Owing to the intrinsically strong fluorescence emission from the rigid π-electron system of PACs, luminescent techniques have been utilized extensively for PAC identification and quantification in environmental samples. The parent homocyclic species, which contain only carbon and hydrogen, are the PAHs. Sixteen PAHs are of extreme environmental concern due to their carcinogenic effect. Their molecular structures are presented in Figure 1. In addition to PAHs, there are thousands of PACs with substituent groups such as alkyl, amino, chloro, cyano, hydroxo, oxy or thio groups. In this review, the term PAC is used to include both heterocyclic and homocyclic compounds, whereas the term PAH is used to designate only homocyclic hydrocarbons.

2 LUMINESCENCE TECHNIQUES

2.1 Fluorimetry

Room-temperature fluorimetry offers the advantages of simplicity, relatively inexpensive instrumentation, calibration curves with LDR over three to four orders of magnitude, and LOD ranging from the picograms per milliliter to the nanograms per milliliter level.[3] In most cases, sample deoxygenation is not necessary and introducing the quartz sample cell into the compartment of a spectrofluorimeter easily performs fluorescence measurements.

A basic spectrofluorimeter for steady-state fluorescence measurements is equipped with a continuous wave (CW) excitation source, which is usually a xenon arc lamp operated at high pressure for continuous radiation between 200 and 800 nm. Excitation wavelengths

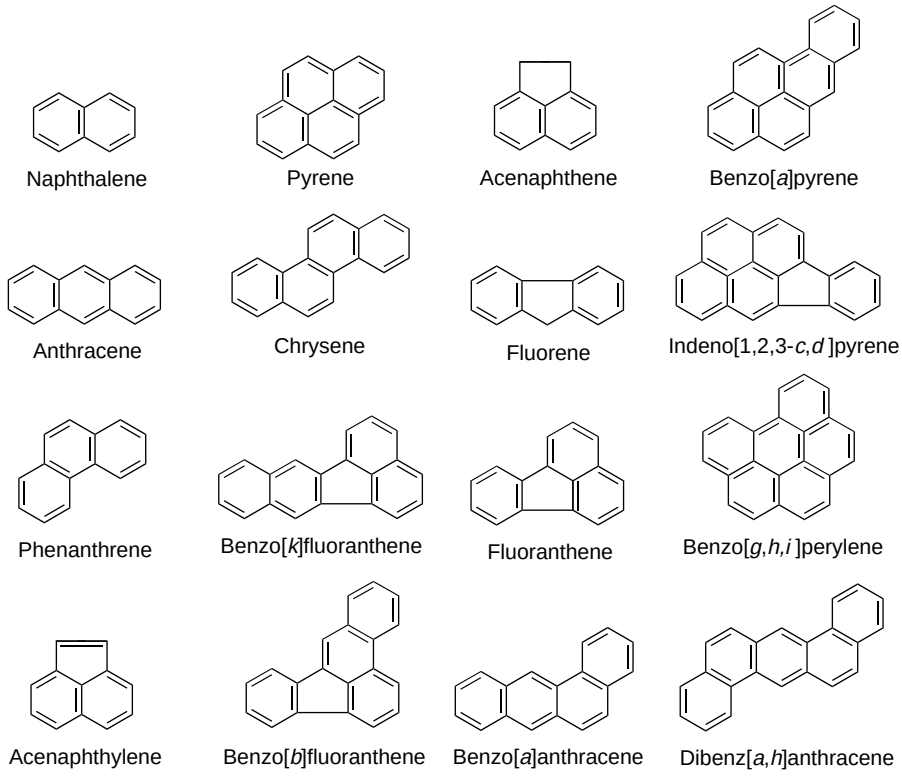

Figure 1 Chemical structures of the 16 PAHs listed on the USEPA priority pollutant list.

may be selected with bandpass filters or with an excitation monochromator. The detection system consists of: an emission monochromator, which selects the luminescence wavelength region of interest; the photodetector – usually a photomultiplier tube – which converts the luminescent radiant flux to an electrical signal; and the amplifier-readout system, which amplifies and processes the signal as required and displays it in a convenient fashion. A computer may be interfaced to the spectrofluorimeter for data acquisition and instrumental control.

A spectrofluorimeter with a dual sample–reference measurement capability may be used to compensate for instrumental distortions such as source intensity fluctuations and wavelength dependence, efficiency of the excitation wavelength selector and optical components, and wavelength dependence of detector response. Several accessories are available commercially that can be used to collect corrected excitation spectra. Improvements in signal-to-noise ratio can be obtained by using a photon-counting detection system.

Conventional fluorimetric methods – in which either the excitation or emission wavelength is set at its maximum position while the other is scanned – present limited selectivity. In the analysis of complex mixtures, the similarity and/or overlapping of broad fluorescence bands at room temperature usually interfere in the characterization process of single components. Thus, fixed-wavelength scans often do not provide either quantitative or even qualitative information for many real-world, multicomponent samples.

Several luminescence techniques have been developed to improve the selectivity of luminescence measurements. Warner et al.[4–6] have been pioneers on the use of excitation–emission matrices (EEM), lifetime and polarization measurements, and their simultaneous application to achieve a high degree of selectivity. Additionally, one can take advantage of miscellaneous methods such as quenching effects, surfactant-enhanced luminescence (micelles, etc.), chemical derivatization, and fluorescence detection in chromatographic separations such as thin-layer chromatography (TLC), paper chromatography, and HPLC.

A relatively simple approach for the simultaneous analysis of multicomponent mixtures is the synchronous excitation technique. Originally proposed by Lloyd for spectral fingerprinting of oil samples in forensic analysis, synchronous excitation was later developed by Vo-Dinh et al. for preliminary screening of PACs in environmental samples and for on-site luminescence analysis of PACs.[7–9] This approach consists of varying simultaneously both the excitation (λ_{exc}) and emission (λ_{em}) wavelengths while keeping a constant wavelength interval

$\Delta\lambda(\lambda_{em} - \lambda_{exc})$ between them. The advantages of synchronous excitation include simplification of emission spectra, reduction of the spectral range for a given analyte, and narrowing for spectral bandwidth. Synchronous excitation can be performed with any commercial spectrofluorimeter in which excitation and emission monochromators can be interlocked. Spectral acquisition is performed using narrow excitation and emission slits. Although the use of narrow slits limits the sensitivity of the technique, it avoids undesired scattering and stray light interference resulting from the small $\Delta\lambda$ value.

The use of laser excitation provides several advantages to fluorimetric measurements over commercial spectrofluorimeters. Because the irradiance of a laser is greater than that of a conventional source with typical excitation bandpass values, laser excitation increases the intensity of fluorescence emission. Depending on the behavior of background emission upon laser excitation, an improvement in LOD can be observed. Laser-induced fluorimetry (LIF) has been employed successfully for monitoring trace organic pollutants in a wide variety of environmental samples.[10]

Laser excitation is also advantageous for synchronous luminescence spectroscopy of PACs. Because of the small $\Delta\lambda$ value usually employed in synchronous excitation, the use of large excitation slits with commercial spectrofluorimeters may create scatter and stray light interference. The fluorescence intensity of synchronous spectra, therefore, is usually limited by the use of narrow spectrometer slits. The narrow line width of the laser beam allows one to use small $\Delta\lambda$ values and still achieve high irradiance excitation. LOD at the zeptomole (10^{-21} mol) level has been estimated for several PACs without previous separation from complex matrices.[11,12]

The strong fluorescence signals exhibited by PACs contribute to the popularity of HPLC detectors. The high sensitivity of fluorescence has been augmented by the commercial availability of programmable detectors. These detectors allow the excitation and emission wavelengths to be changed throughout the chromatographic run in order to target compounds during their retention time windows. Commercial spectrofluorimeters can now scan at rates of many hundreds of nanometers per second, which allows on-the-fly spectra collection of chromatographic peaks if a flow cell is used. This type of spectrofluorimeter has been employed to collect full PAH EEM in a variety of water samples.[13,14]

Fluorescence lifetime is another parameter that can be used to distinguish among compounds with overlapping excitation and emission spectra. Because fluorescence lifetime values range from picoseconds to microseconds, their timescale is much faster than the timescale of chromatography. Pulsed laser sources and fast detection systems allow temporal resolution of fluorescence decays during peak elution. McGown et al.[15] have been the most active researchers in this area. On-the-fly fluorescence detection has been applied to the ultratrace-level determination of PAHs in a variety of environmentally related samples.

2.2 Phosphorimetry

Phosphorimetric techniques include solution LTP, MSRTP, sensitized room-temperature phosphorimetry in liquid solution (SRTPL), and SSRTP. The interested reader is referred to comprehensive texts concerning modern phosphorimetric analysis.[16,17]

The first generation of commercial equipment for phosphorimetric measurements consisted of a spectrofluorimeter equipped with a phosphoroscope for time discrimination of short-lived phenomena. Recent spectrofluorimeters are equipped with a pulsed lamp and a gated detection system, which allow collection of phosphorescence lifetimes in the long microsecond to millisecond time domain.

Early work in phosphorimetry was carried out at liquid nitrogen temperature (77 K). By using cryogenic equipment, the solutions were frozen to a rigid matrix to minimize the collisions competing with phosphorescence for deactivation of the triplet state. Because of the long lifetime of the triplet state, careful sample deoxygenation is recommended to avoid quenching of phosphorescence intensity. Sample deoxygenation prior to phosphorescence measurement can be accomplished by purging pure nitrogen gas through the sample solution. The main advantages of LTP are low LOD (nanomolar ranges), calibration curves with wide LDR (10^4 to 10^6-fold change in analyte concentrations) and narrow phosphorescence bands that enhance selectivity.[18] However, very little LTP research has been done in recent years. The lack of widespread use is mainly attributed to three reasons: the high cost of cryogenic equipment; the time-consuming sample deoxygenation; and the poor reproducibility of measurements. Although a skilled technician can obtain a RSD lower than 5%, it is not unusual to obtain RSD values as high as 20%.

Cline-Love et al.[19,20] first applied MSRTP in aqueous solution as an analytical method. The method was employed to determine a wide variety of compounds, including ketones, aldehydes, alcohols, carboxylic acids, phenols, and amines. Organized media stabilization was made with the use of sodium dodecyl sulfate micelles in the presence of thallium(I) or silver(I) as a phosphorescence enhancer. Micellar solutions are colloidal aggregates or detergents that incorporate analyte molecules in their micelle cavity to protect the triplet state from collisional and vibrational deactivation. The aggregation of surfactants only occurs above a certain

concentration, known as the critical micelle concentration (CMC). This concentration depends upon the chemical structure of the surfactant and the nature of the solvent. If no data on CMC values are available in the literature, the CMC of a particular surfactant–solvent system can be determined by fluorescence titration of a probe compound,[19,20] capillary electrophoresis,[21–23] or microcalorimetric titration.[24] The use of microemulsions as organized media has also been proposed to enhance RTP emission of PAHs and pesticides in liquid solutions.[25,26] Cline-Love et al. reported on the use of different cyclodextrin-stabilized RTP methods using β-cyclodextrin.[27–29]

The main disadvantage of MSRTP is the necessity of thorough sample deoxygenation, which results in analysis times of approximately 1 h. Shorter analysis times can be obtained by using sodium sulfide as oxygen scavenger.[30] There are, however, considerable advantages in the use of the micelle method. In most cases, the LOD and precision of MSRTP are similar to those obtained by LTP. The RTP analysis of PACs is facilitated by the high solubility of these species in the micellar environment. Finally, the micelle approach offers a means to conduct phosphorescence analysis in flow systems and liquid chromatography.[16,17]

RTP detection in continuous flows and chromatographic systems can be performed also by means of SRTPL.[31,32] In this technique, the weakly phosphorescent or nonphosphorescent analyte (donor) transfers its triplet energy to a given acceptor. A suitable acceptor should have a triplet energy lower than that of the donor, a high quantum phosphorescence yield, and a lower molar absorption in the excitation region of the donor. In addition, the energy gap between the triplet states of the donor and the acceptor should be enough to prevent significant reverse mechanisms. For instance, biacetyl has shown to be an appropriate acceptor for the analysis of polyhalogenated naphthalenes, dibenzofurans and polychlorinated biphenyls. As a detection technique in flow injection analysis and HPLC, SRTPL obtained LOD in the nanomolar range. In addition to competitive LOD, several advantages justify the use of SRTPL. Because it is an emission technique, SRTPL is inherently more sensitive than ultraviolet absorption. The background due to scattering and fluorescence impurities is negligible at the emission wavelength of biacetyl (522 nm), and it is particularly selective because most sample constituents do not promote the sensitized phosphorescence signal of the acceptor. Further research, however, is necessary to develop new acceptor–donor systems in order to expand the analytical potential of SRTPL.

In the early 1970s, Schulman and Walling[33] reported the observation of strong RTP from a wide variety of ionic compounds adsorbed on solid substrates. Silica, alumina, paper, asbestos, and glass fiber were employed to restrict vibrational and collisional deactivation of the triplet state. Following these early reports, and certainly attracted by the simplicity of the sample procedure, a considerable number of studies on SSRTP have been reported.[16,17]

The classic SSRTP procedure includes two steps: co-spotting microliters of sample and phosphorescence enhancers on the solid substrate with a microsyringe; and drying the sample in an oven (or under an infrared lamp) to minimize water quenching of phosphorescence emission. As with any other analytical method, a satisfactory LOD is obtained with low background signals. The main contribution to background emission in SSRTP is given by the solid substrate. Many materials have been searched to find a substrate with low phosphorescence emission between 350 and 650 nm, which is the spectral region where most compounds phosphoresce. These include filter paper, silica gel, and various combinations of sodium acetate, poly(acrylic acid) and cyclodextrin with sodium halides. Among the materials investigated, filter paper has been shown to be a good, general SSRTP substrate. Its main advantages are low cost, simplicity of handling, and availability of a wide selection of papers, which enable one to select the best substrate for specific phosphors. On the other hand, filter paper presents broad, featureless bands of excitation spectra from 200 to 400 nm, and phosphorescence spectra from 375 to 625 nm. Several attempts have been made to decrease this signal, to some extent, using solvent extraction or chromatography, chemical or heat treatment, or signal background with the help of a computer-assisted spectrophosphorimeter.[16,17] However, the most significant reductions – which were above 90% of the original signal – were obtained with ultraviolet irradiation[34] or a combination of water Soxhlet extraction and ultraviolet irradiation.[35] Because the low background signal remains constant after treatment, batches of treated paper can be stored in a desiccator to be used in routine analysis.

Depending on the PAC and the phosphorescence enhancer, calibration curves with a 10^2–10^3-fold change in the LDR of the analyte concentration, absolute LOD in the nanogram to subnanogram level, and RSD from 5 to 15% can be obtained with a relatively simple procedure. In certain cases, the sensitivity of SSRTP has been improved with matrix-modifier substrates. Diethylene triaminepentaacetic acid, cyclodextrin, and sodium dodecyl sulfate are just three of the many compounds used to modify the surface of filter paper and enhance the enhancing efficiency of heavy atoms and/or PAC RTP emission.[36–38]

If the monitoring of targeted components in complex matrices is of interest, selective external heavy-atom perturbation (SEHAP),[39] second-derivative methods,[40]

and synchronous excitation[7] techniques can be employed to reach specific compound identification. SEHAP, which was first developed by Vo-Dinh and Hooyman,[39] is based on the appropriate choice of a heavy atom salt, which enhances the RTP emission of the target PAC and causes phosphorescence quenching of potential phosphorescent concomitants.

Although these alternatives improve the selectivity of SSRTP, separation prior to RTP detection is convenient when dealing with complex matrices. Paper chromatography and TLC have been proposed as separation techniques for SSRTP.[41,42] The possibility of complete analysis (separation and detection) on the same solid substrate, and the use of RTP spectral characteristics as additional parameters for compound identification, turn paper chromatography/SSRTP and thin-layer chromatography/solid-surface room-temperature phosphorimetry (TLC/SSRTP) into attractive screening tools.

SSRTP has been employed also as a detection technique for HPLC. The chromatographic fractions can be collected manually and spotted onto a RTP substrate,[16,17] or they can be analyzed with an SSRTP automated system that completely eliminates any manual steps.[43,44] In both cases it is possible to characterize single components in partially resolved mixtures, which simplifies the separation process in the chromatographic column. In addition to luminescent parameters such as excitation and emission wavelengths, excited-state lifetimes, synchronous excitation, and second-derivative methods, SEHAP – which is unique to SSRTP – can be used to enhance the selectivity toward a specific component. The combination of these parameters, associated with chromatographic retention times, is certainly valuable for qualitative purposes and for the simplification of chromatographic separation of complex samples.

Very few attempts have been made to improve the detection levels of PACs by laser-induced SSRTP.[45,46] When compared with the absolute LOD obtained with commercial spectrofluorimeters, laser-induced SSRTP provides an improvement of approximately 1000-fold. This improvement results from the high irradiance of the laser source and the small area of solid substrate probed by the laser beam. It is possible to measure nanoliter volumes of sample with good reproducibility. Within the LDR of the studied PAC, the RSD values varied from 7% to 9.5%. The RSD obtained with a commercial spectrofluorimeter[45] for the same analyte mass on the paper substrate varied from 30 to 40%. Even when small excitation slits were employed (10 nm), the irradiated area on the paper substrate excited the analyte signal and the background emission of the solid substrate. Smaller excitation slits (5 nm) resulted in unacceptable analyte-to-background signal ratios. By focusing the laser beam exclusively on the sample spot, it was possible to achieve selective excitation of the analyte, thus avoiding the excitation of background impurities. The low absolute LOD obtained by laser-induced SSRTP simplified soil sample preparation procedures. Because the evaporation of extracting solvent was not necessary to preconcentrate the sample, the direct deposition of nanoliter volumes of organic extract on the solid substrate was sufficient to determine PACs at the nanogram to picogram level.[46]

2.3 Shpol'skii Spectrometry

Shpol'skii spectrometry is part of a group of high-resolution luminescence techniques that include fluorescence line narrowing,[47] matrix isolation,[48] and supersonic jet spectrometry.[49] Because fluorescence line narrowing and matrix isolation can be considered complementary to Shpol'skii spectrometry, and supersonic jet spectrometry has received very little attention in current analytical literature, this review will give an overview of the application of Shpol'skii spectrometry to environmental analysis.

As mentioned previously, common room-temperature fluorescence and phosphorescence spectra of PAHs in solution show broad excitation and emission bands with typical full width at half-maximum (fwhm), values of the order of 300 cm^{-1}. When PAHs are incorporated in a crystalline or polycrystalline matrix (formed at 77 K or below, using appropriate organic solvents), their excitation and luminescent spectra (fluorescence and/or phosphorescence) exhibit fine structure with sharp line (quasi-line) widths (fwhm = 1–10 cm^{-1}). Under the Shpol'skii conditions, the influence of solution inhomogeneity on the luminescence spectra of PAHs is strongly reduced by isolating the guest molecule (analyte) in a host matrix (usually *n*-alkanes) under low (nitrogen or helium) temperature. As a result of the highly structured orientation of solute molecules in the solid solvent, vibrationally resolved excitation and emission spectra, favorable for fingerprinting identification, can be observed. Excellent reviews have discussed the fundamental and applied aspects of Shpol'skii spectrometry.[50,51] These include comprehensive tabulations listing the optimum solvent(s) for observing quasi-line spectra of PAHs, instrumentation, types of environmental samples, and procedures for sample cooling.

Early work in Shpol'skii spectrometry was carried out with excitation sources such as mercury-vapor discharge lamps or xenon lamps. Selection of excitation wavelength was performed with wide-band filters, interference filters, or monochromators. In recent years, the use of pump lasers and dye lasers for sample excitation has dominated the area of Shpol'skii spectrometry. Fassel et al.[52–55] examined extensively the utilization of a tunable dye laser with the Shpol'skii technique for the analyses of PAHs

at 15–20 K in solvent-refined coal and shale oil samples. The merger of quasi-linear absorption spectra with sharp-line laser excitation sources resulted in high selectivity. This technique, which was called laser-excited Shpol'skii spectrometry (LESS), has been employed for selective identification of nitrogen-, oxygen-, and sulfur-containing PACs[56] and isomeric PAHs.[57]

Despite the ability of Shpol'skii spectrometry and/or LESS to provide direct identification of PAHs and PACs in crude extracts of environmental samples, their analytical potentials are still scarcely employed in chemical analysis. From the point of view of quantitative analysis, there are some critical parameters such as the freezing rate, analyte concentration, final temperature, and the necessity of working with an appropriate Shpol'skii solvent that affect the signal intensity (and bandwidth) and, consequently, the analytical results. Other problems include the limited LDR of the calibration curves and the strong scatter observed. According to Fassel and D'Silva,[55] signal reproducibility can be solved by using a careful and reproducible freezing rate and a low analyte concentration (10^{-6} M or lower). The scatter is not a problem if the excitation wavelength is at least 10 nm lower than the emission wavelength, and the matrix compatibility is not an issue because *n*-heptane can be used with several PAHs with three- to six-ring structures. To compensate for the inner filter effect or enhancement effects, the standard addition and the internal standard method can be employed.

However, one of the main problems associated with the lack of analytical use of Shpol'skii spectrometry is the impractical sampling procedure. Two types of sampling systems are commonly used. For a temperature of 77 K, the sample is introduced into a glass or quartz tube followed by direct immersion into a Dewar flask with liquid nitrogen. The drawbacks are several, beginning with the somewhat fragile Dewar vessel. Small ice particles in the liquid nitrogen act as nucleation sites, which cause bubbling at irregular intervals and disruption of the optical path. Condensation on the outside of the Dewar flask similarly degrades the quality of the spectra. The excitation and emission light must pass through many airglass and other interfaces that scatter the exciting light, and the snowy Shpol'skii matrix is also strongly scattering. As a result of the irreproducibility from sample to sample, a RSD as high as 15–20% can be obtained. For temperatures below 15 K, cryostats or closed-cycle refrigerators are used, where the coolant can be liquid helium. However, helium closed-cycle refrigerators, Joule–Thomson miniature refrigerators, cryostats that cool the sample in cold vapor, contact cooling, etc. lack the simplicity of ''dunking'' the sample into liquid nitrogen.

3 FIELD MONITORING IN ENVIRONMENTAL ANALYSIS

The high coupling efficiency of lasers to fiber optics is useful in remote sensing applications. LIF with optical fibers has been used to monitor PACs in environmental samples.[58–62] The specificity of LIF sensors can be improved by coating selective receptors at the distal end of the fiber optic. For instance, a highly selective detector has been reported for pyrene and other PAHs that use β-cyclodextrin as a selective receptor.[58] Its high selectivity toward PAHs results from the hydrophobic nature of the β-cyclodextrin tip, which does not complex with other heterocyclic concomitants in contaminated water samples. Similarly to fluorescence measurements in solution, spectral overlapping of fluorescence species can be improved by temporal resolution.[59–62]

A remote fiber-optic sensor has been developed to detect PAHs on a real-time basis.[63] The device, which uses a pulsed nitrogen laser as excitation source and a gated charge-coupled device (CCD) detector at the focal plane of a spectrograph, is able to detect anthracene at nanogram per milliliter levels in 10 μs. Another instrumental system that can be used with standard cuvette sample holders or fiber-optic probes was developed recently to monitor PACs in environmental samples.[11] It is based on laser-induced synchronous luminescence and allows several laser dyes to be used in a single scan (multidye scan), which extends the wavelength excitation range. When the laser wavelength reaches the edge of the lasing region of one dye, the computer control program pauses for a few seconds to allow manual insertion of a new dye solution. The dye exchanging process is rapid and simple because no optical realignment is necessary. The wavelength regions scanned by different laser dyes in a multidye scan need not be continuous and different $\Delta\lambda$ values can be used. By using the fiber-optic approach, benzo[*a*]pyrene and perylene were detected at the parts-per-billion ($ng\,mL^{-1}$) level in a water sample.

A SSRTP fiber-optic probe has been developed recently for the analysis of PACs in contaminated waters.[64,65] A pulsed nitrogen laser and a dye laser were used as excitation sources. At the tip of a bifurcated fiber-optic bundle, the RTP probe, which consisted of a cavity between two cylindrical parts attached by concentric incised threads, provides the paper substrate compartment. A hole at the distal end of the probe allows the entry of water into the substrate cavity. By dipping the probe into the water sample, the paper substrate is imbibed with water, which is retained on the solid substrate. RTP measurement is performed after removing the probe from the water sample and drying the substrate under an infrared lamp. Several PAHs and nitrogen

heterocyclic compounds of environmental importance were determined at the parts-per-billion level.[64,65]

Although the SSRTP probe does not allow in situ analysis of water samples, fiber-optic sensing based on phosphorescence emission offers higher selectivity over fluorescence optosensing. The higher selectivity is based on the fact that phosphorescence is less common than fluorescence at room temperature. The interference from short-lived scattering and possible fluorescence signals can be avoided by using appropriate delay times after the excitation pulse. This is particularly useful in reducing or even eliminating the spectral interference of humic and fulvic acids, a problem usually encountered with fluorescence optosensing of PACs.[58–62] SEHAP can provide enhanced selectivity for the direct determination of a target compound in complex mixtures with several luminescent (fluorescent and/or phosphorescent) compounds. Other advantages include the use of low-cost optical fibers (phosphorescence occurs at emission wavelengths longer than 450 nm) and relatively-low-cost instrumentation for time-resolved measurements, because triplet-excited lifetimes are in the millisecond (or higher) time domain and, therefore, they are easier to resolve instrumentally than fluorescence lifetimes.

The RTP probe presents unique features for field analysis. Its design minimizes contact between the sample and quenching species such as oxygen and atmospheric humidity. By using appropriate heavy-atom salts, the excitation wavelengths of the compounds of interest can be shifted to overlap the excitation line of the laser source and avoid the change in excitation source. By spotting standard solutions on the paper substrate imbibed with the sample, on-site identification and determination of compounds can be performed. By changing the paper substrate in the sample compartment, the RTP probe can be used for an infinite number of measurements, which reduces analysis time and cost when compared to coated fiber-optic sensors.

4 NEW TRENDS IN LUMINESCENCE ANALYSIS OF ENVIRONMENTAL SAMPLES

SLE is nowadays the recommended method for the sample pretreatment of complex matrices, and the use of LLE has dropped dramatically following obligatory reduction of chlorinated solvent usage in analytical laboratories. In water analysis, SLE concentrates the organic pollutants by sorption onto an octadecyl (C_{18}) organic phase chemically bonded to silica particles. When compared to LLE, SLE reduces solvent consumption, has fewer steps, saves labor, provides better extracting efficiency, prevents emulsions, enables easy sample collection, and is more amenable to automation.[66]

SLE, however, presents certain disadvantages for chromatographic analysis. The mechanism of isolation on C_{18} membranes is van der Waals' forces (also called nonpolar, hydrophobic or reversed-phase interactions) and occasionally secondary interactions such as hydrogen bonding and dipole–dipole interactions. Nonreversed-phase interactions (hydrogen bonding and dipole–dipole interactions) result from unreacted (residual) silanol and siloxane groups on the surface of octadecyl silica phases. These interactions are stronger than van der Waals' forces and can cause poor analyte recovery. An additional consideration for efficient analyte elution is the removal of water from the C_{18} sorbent. In order for the elution solvent to wet effectively the bed of the sorbent, it is necessary to remove the water in the void volume, trapped water, and water sorbed to the silica. Air or a vacuum may be used to remove the majority of the water in the void volume. The water in the smallest pores and water that is sorbed to the silica surface will be tightly trapped or bound to the silica surface. The bound water may be removed by drying the sorbent under a vacuum (25 mmHg) for at least 15 min. Another approach is to use an eluting solvent that is miscible with water. For instance, a mixture of methylene chloride–ethyl acetate (1 : 1) gives recoveries of 85–95% for the 46 pollutants listed in the USEPA method 525.1.[1] Methylene chloride, on the other hand, gives low analyte recoveries (40–70%). These low recoveries most likely arise from the inability of methylene chloride to displace and penetrate the water saturating the internal pores of the silica phase. Finally, a source of potential contamination in SLE is the presence of impurities in the cartridge or disk (such as phthalate esters) that leach from the adsorbing material into ethyl acetate and methylene chloride and produce a variable background in the water sample.

A new approach for screening PACs in water samples is being developed in our laboratory.[67–69] The new analytical tool combines SLE with SSRTP. Because the analysis of PACs is performed on the adsorbing material, the drawbacks associated with SLE procedures are eliminated. The ability of SSRTP to measure nanogram to subnanogram levels of PACs on the extraction membrane allows one to use relatively small water sample volumes. This is advantageous because it simplifies the extraction procedure and reduces analysis time. Figure 2 shows the apparatus employed for 10-mL SLE of water samples. Conventional water SLE procedures, which rely on processing 1 L of water through the extraction membrane, require bulky glassware, a vacuum pump, an oven to dry the sample following water extraction, and relatively long analysis times (~1 h per sample).

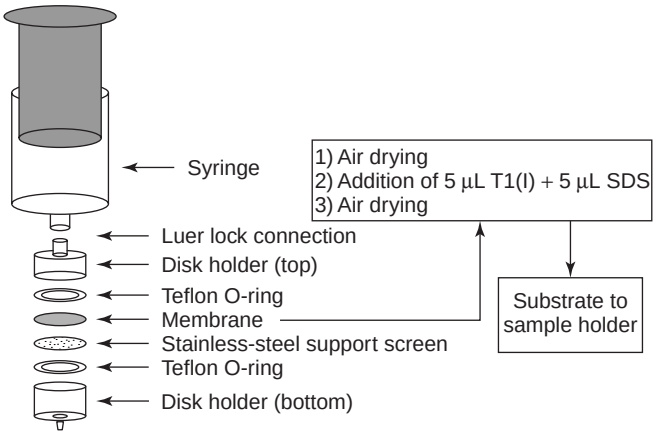

Figure 2 SLE apparatus and sample preparation procedure for RTP analysis. SDS, sodium dodecyl sulfate.

By using syringes and stainless-steel filter holders, the bulky glassware and vacuum pump are removed from the extraction procedure. The oven-drying step is substituted by air sample drying, which employs positive pressure to remove water excess from the extraction syringe. As a result, analysis times of 3–5 min per sample are obtained, which are extremely attractive for screening large number of samples. The LOD values for the 16 PAHs of the USEPA priority list (see Figure 1) were estimated with a commercial spectrofluorimeter at the picogram per milliliter level. LOD values at the parts-per-trillion level were also estimated for polychlorinated biphenyls and dibenzofurans. Selective identification of total PAHs, polychlorinated biphenyls and dibenzofurans was obtained by appropriate selection of heavy-atom salt and delay and gate times. Although further studies are necessary to prove solid–liquid extraction/room-temperature phosphorimetry (SLE/RTP) as a reliable analytical tool, the simplicity of the new experimental procedure, the analytical figures of merit and the selectivity of SLE/RTP show its analytical potential for initial screening of numerous water samples.

A similar approach is being developed by Hurtubise et al.[70,71] at the University of Wyoming (WI, USA). Whatman 1PS filter paper is used as the solid-phase microextraction medium for PAHs in water. This filter paper is based on the same cellulose substrate as Whatman no. 1, but it is impregnated with silicone oil in the presence of an organotin catalyst. It is believed that the tin remaining in the coating may act as a heavy atom, facilitating the acquisition of RTP spectra. The silicone provides a hydrophobic surface onto which PAHs can be extracted from water. The PAHs on the surface of the Whatman 1PS paper are detected at the nanogram per milliliter level by using fluorescence and/or phosphorescence measurements.

A fiber-optic probe was employed recently in our laboratory to test a new sampling procedure for photoluminescence measurements at 77 K. The probe consisted of one excitation fiber surrounded by six emission fibers. The launch and collection fibers were bundled with vacuum epoxy. Figure 3 shows a schematic diagram of the fiber-optic probe and measuring procedure. A 1-mL aliquot of sample solution was introduced into a 5-mL Pyrex test tube, and the test tube was then introduced into a container with liquid nitrogen. Figure 4 shows the instrumental system used with the fiber-optic probe. The intensifier in front of the CCD chip acts as a super-fast shutter, capable of operating on a nanosecond timescale. The parameters for data acquisition (gate delay, gate width, and the gate step) are entered in a computer, using

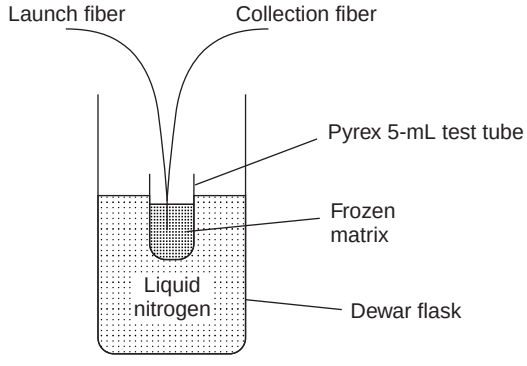

Figure 3 Fiber-optic probe for photoluminescence measurements in Shpol'skii matrices.

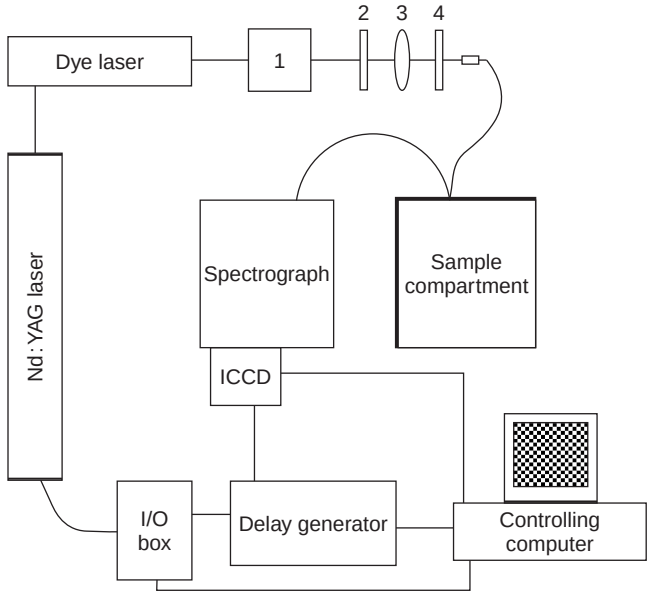

Figure 4 Instrumental system for laser-excited time-resolved EEM. 1 = doubling crystal; 2 = short-pass filter; 3 = focusing lens; 4 = density filter; I/O = input–output; ICCD = intensified charge-coupled device.

appropriate software. The software sends these intensifier set-up details via a general purpose interface bus (GPIB) interface to the pulse generator. Once triggered by the laser, the pulse generator uses this information to determine when the image intensifier in the detector head is gated on (gate delay), for how long it is gated on (gate width), and the duration of the steps by which the gate delay is progressively increased in the course of the sequence of acquisitions (gate step). The intensifier is gated on for the duration of the pulse, allowing the CCD to acquire data. When the intensifier is gated off, the acquired data are transferred from the detector head to the controller card (32-bit Intelligent Bus-Mastering PCI card) in the computer. With this instrument, it is possible to collect wavelength–time matrices in a few nanoseconds. The entire sample procedure – which includes sample freezing, spectra collection, and probe rinsing for the following measurement – takes less than 5 min, which is extremely attractive for routine screening. Because the fiber-optic probe eliminates scattering surfaces between sample excitation and fluorescence collection, the RSD for several PAHs was excellent, varying from 2% (pyrene) to 7% (phenanthrene). Table 2 shows the LDR and LOD obtained with the new system using optimum delay and gate times. The appropriate choice of delay and gate times for each PAH was based on the analyte and background fluorescence lifetimes. Compounds with long lifetimes, such as pyrene, allow the collection of strong PAH fluorescence emission at delay times where background contribution is negligible (approximately equal to instrumental noise). On the other hand, for a PAH such as chrysene, the complete time resolution of background emission is limited by the short fluorescence decay of the PAH. Therefore, the choice of a delay time for the best analyte-to-background signal ratio involves a compromise between the relative intensities of analyte and background signals.

In addition to the improvement in analyte-to-background signal ratios, time resolution enhances the selectivity of LESS toward the target PAH. Figure 5 shows the fluorescence spectrum of benzo[*a*]pyrene recorded from a standard solution in *n*-heptane. Figure 6 shows the low-temperature fluorescence spectrum of a small volume of oil sample spiked into *n*-heptane. By selecting the delay (10 ns) and gate (50 ns) times based on the fluorescence lifetime of benzo[*a*]pyrene (39.9 ns), it was possible to identify the target PAH in a direct way.

Several separation techniques have been coupled to Shpol'skii spectrometry for the analysis of PAHs in "real world" samples. In the specific case of water

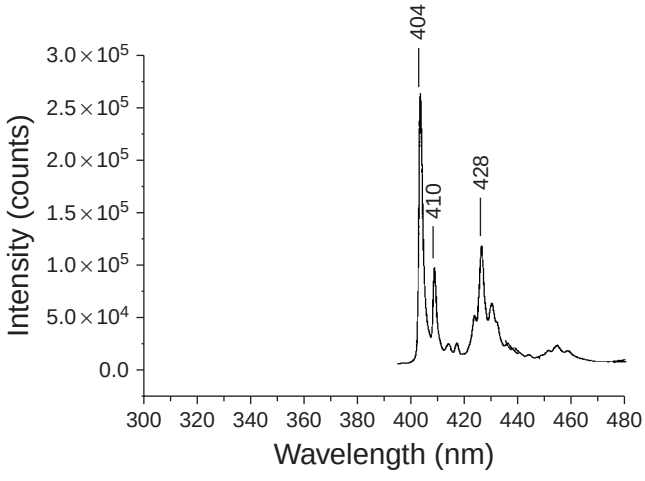

Figure 5 Low-temperature fluorescence spectrum of benzo[*a*]-pyrene in *n*-heptane.

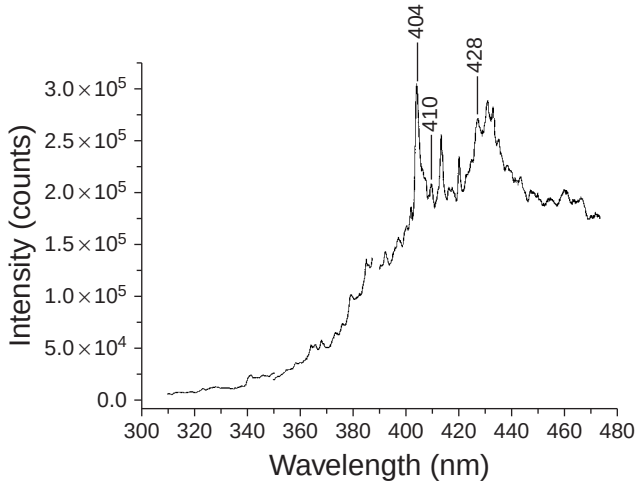

Figure 6 Low-temperature fluorescence spectrum of a contaminated oil sample spiked into *n*-heptane.

Table 2 LDRs and limits of detection obtained by LESS using a fiber-optic probe and a time-resolved EEM instrument

Analyte[a]	$\lambda_{exc}/\lambda_{em}$[b] (nm)	LDR[c]	LOD[d] (ng mL^{-1})
Pyrene	280/384	6×10^4	0.5
Phenanthrene	287/365	1.7×10^3	12
1,2 : 5,6-Dibenzanthracene	290/394	1.7×10^4	0.06
Chrysene	286/381	2.8×10^2	0.7
Benzo[*g*,*h*,*i*]perylene	290/421	6.7×10^3	0.2
Naphthalene	282/323	7.5×10^3	0.8

[a] Analyte solutions were prepared in *n*-heptane.
[b] Excitation and emission wavelengths. Emission wavelengths correspond to maximum emission wavelengths.
[c] LDR is calculated by dividing the upper linear concentration by the LOD.
[d] LOD is calculated based on the equation LOD = $3S_b/m$, where S_b is the standard deviation of 16 blank measurements and m is the slope of the calibration curve.

Table 3 Limits of detection estimated by SLE and LESS

Analyte[a]	$\lambda_{exc}/\lambda_{em}$[b] (nm)	LOD[c] (pg mL^{-1})
Pyrene	280/384	0.09
Benzo[*a*]pyrene	290/410	0.2
Chrysene	286/381	0.6
Fluoranthene	290/438	34

[a] Analyte solutions were prepared in *n*-heptane.
[b] Excitation and emission wavelengths. Emission wavelengths correspond to maximum emission wavelengths.
[c] LOD is calculated based on the equation LOD = $3S_b/m$, where S_b is the standard deviation of 16 blank measurements and m is the slope of the calibration curve.

sample preparation, LLE has been the approach used the most.[50] In addition to the previously discussed advantages of SLE over LLE, the combination of SLE with LESS presents a unique feature, which is obtained if the eluting and the Shpol'skii solvent are the same. A three-step sample procedure (extraction, elution, and fluorescence measurement) then is obtained, which provides a quick and simple method of analysis. Table 3 shows the LOD for several compounds obtained by SLE coupled with LESS. The preconcentration factors achieved by SLE, associated with the small volume of organic solvent used for PAH elution and fluorescence measurement, result in LOD at the femtogram per milliliter level. These values compare favorably with those obtained by room-temperature fluorimetry, where several factors such as lack of scatter, large excited sample volume, spectral bandwidth, and depth of excitation contribute to very low detection levels.[3]

5 FINAL REMARKS

The monitoring and characterization of PACs in complex environmental samples require a battery of analytical methods and instrumentation. Laboratory analysis of PACs is mostly performed by GC and HPLC, which can provide complete characterization of complex matrices after sample clean-up and preconcentration. Because chemical analysis is increasingly moving from the traditional laboratory setting toward in situ applications, analytical chemistry faces the challenge of developing the tools for on-line and field monitoring in real-time bases.

Photoluminescence is inherently a multiparameter phenomenon. The examples presented in this review demonstrate the main thrust of photoluminescence in environmental analysis, which is obviously related to the uniqueness of its multidimensional nature. The measurement of fluorescence and/or phosphorescence from a sample involves the simultaneous use of excitation and emission wavelengths and luminescence lifetimes. The appropriate combination of these parameters, sometimes associated with polarization measurements, chemical derivatization, and/or measurement of effects such as quenching, enzymatic catalysis, and immunochemical equilibrium, provides highly selective and sensitive techniques. To measure the various luminescent parameters and explore the full potential of photoluminescence techniques, the analyst should use a combination of commercial spectrofluorimeters and instrumental set-ups with laser excitation sources, fiber optics, monochromators and spectrographs, photodetectors, and imaging systems. The continuous development of luminescent techniques and instrumentation, associated with the parallel development of novel computer algorithms for reduction and interpretation of large data sets, provides a valuable number of spectrochemical tools for direct laboratory determination of PACs and remote sensing under field conditions.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
CMC	Critical Micelle Concentration
CW	Continuous Wave
EEM	Excitation–Emission Matrix
fwhm	Full Width at Half-maximum
GC	Gas Chromatography
GPIB	General Purpose Interface Bus
HPLC	High-performance Liquid Chromatography
LDR	Linear Dynamic Range
LESS	Laser-excited Shpol'skii Spectrometry
LIF	Laser-induced Fluorimetry
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
LTP	Low-temperature Phosphorimetry
MSRTP	Micelle-stabilized Room-temperature Phosphorimetry
PAC	Polycyclic Aromatic Compound
PAH	Polycyclic Aromatic Hydrocarbon
PTE	Purge-and-trap Extraction
RSD	Relative Standard Deviation
RTP	Room-temperature Phosphorimetry
SEHAP	Selective External Heavy-atom Perturbation
SLE	Solid–Liquid Extraction
SLE/RTP	Solid–Liquid Extraction/Room-temperature Phosphorimetry
SRTPL	Sensitized Room-temperature Phosphorimetry in Liquid Solution
SSRTP	Solid-surface Room-temperature Phosphorimetry
TLC	Thin-layer Chromatography

TLC/SSRTP	Thin-layer Chromatography/Solid-surface Room-temperature Phosphorimetry
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Solid-phase Microextraction in Environmental Analysis

REFERENCES

1. US Environmental Protection Agency, 'Methods for the Determination of Organic Compounds in Drinking Water', EPA 600/4-88/039, 1991.
2. A. Coly, J.J. Aaron, 'Fluorimetric Analysis of Pesticides: Methods, Recent Developments and Applications', *Talanta*, **46**(5), 815–843 (1998).
3. A. Jurgensen, E.L. Inman, Jr, J.D. Winefordner, 'Comprehensive Analytical Figures of Merit for Fluorimetry of Polynuclear Aromatic Hydrocarbons', *Anal. Chim. Acta*, **131**(1), 187–194 (1981).
4. C.-N. Ho, I.M. Warner, 'Multicomponent Mixture Analysis by Multidimensional Phosphorimetry', *Anal. Chem.*, **54**(14), 2486–2491 (1982).
5. C.-N. Ho, I.M. Warner, 'Multidimensional Phosphorimetry', *Trends Anal. Chem.*, **1**(7), 159–163 (1982).
6. I.M. Warner, G. Patonay, M.P. Thomas, 'Multidimensional Luminescence Measurements', *Anal. Chem.*, **57**(3), 463A–483A (1985).
7. T. Vo-Dinh, 'Multicomponent Analysis by Synchronous Luminescence Spectrometry', *Anal. Chem.*, **50**(3), 396–401 (1978).
8. T. Vo-Dinh, R.B. Gammage, 'Singlet–Triplet Energy Difference as a Parameter of Selectivity in Synchronous Phosphorimetry', *Anal. Chem.*, **50**(14), 2054–2058 (1978).
9. T. Vo-Dinh, 'Synchronous Luminescence Spectroscopy: Methodology and Applicability', *Appl. Spectrosc.*, **36**(5), 576–581 (1982).
10. S.A. Soper, I.M. Warner, L.B. McGown, 'Molecular Fluorescence, Phosphorescence, and Chemiluminescence Spectrometry', *Anal. Chem.*, **68**(12), 477R–494R (1998).
11. C.L. Stevenson, T. Vo-Dinh, 'Laser Excited Synchronous Luminescence Spectrometry', *Appl. Spectrosc.*, **47**(4), 430–435 (1993).
12. C.L. Stevenson, T. Vo-Dinh, 'Analysis of Polynuclear Aromatic Compounds using Laser-excited Synchronous Fluorescence', *Anal. Chim. Acta*, **303**(23), 247–253 (1995).
13. R. Ferrer, J. Guiteras, J.L. Beltran, 'Development of Fast-scanning Fluorescence Spectra as a Detection System for High-performance Liquid Chromatography. Determination of Polycyclic Aromatic Hydrocarbons in Water Samples', *J. Chromatogr. A*, **779**(12), 123–130 (1997).
14. J.L. Beltran, J. Guiteras, R. Ferrer, 'Three-way Multivariate Calibration Procedures Applied to High-performance Liquid Chromatography Coupled with Fast-scanning Fluorescence Spectrometry Detection. Determination of Polycyclic Aromatic Hydrocarbons in Water Samples', *Anal. Chem.*, **70**(9), 1949–1955 (1998).
15. L.B. McGown, F.V. Bright, 'Phase-resolved Fluorescence in Chemical Analysis', *CRC Crit. Rev. Anal. Chem.*, **18**(3), 245–298 (1987).
16. T. Vo-Dinh, *Room Temperature Phosphorimetry for Chemical Analysis*, John Wiley & Sons, New York, 1984.
17. R.J. Hurtubise, *Phosphorimetry: Theory, Instrumentation, and Applications*, VCH, New York, 1984.
18. R.J. Hurtubise, 'Phosphorimetry', *Anal. Chem.*, **55**(6), 669A–680A (1983).
19. M. Skrilec, L.J. Cline Love, 'Room Temperature Phosphorescence Characteristics of Substituted Arenes in Aqueous Thallium Lauryl Sulfate Micelles', *Anal. Chem.*, **52**(11), 1559–1564 (1980).
20. L.J. Cline Love, M. Skrilec, J.G. Habarta, 'Analysis by Micelle-stabilized Room Temperature Phosphorescence in Solution', *Anal. Chem.*, **52**(4), 754–759 (1980).
21. A. Cifuentes, J.L. Bernal, J.C. DiezMasa, 'Determination of Critical Micelle Concentration Values using Capillary Electrophoresis', *Anal. Chem.*, **69**(20), 4271–4274 (1997).
22. H. Nakamura, A. Sano, K. Matsuura, 'Determination of Critical Micelle Concentration of Anionic Surfactants by Capillary Electrophoresis using 2-Naphthalenemethanol as a Marker for Micelle Formation', *Anal. Sci.*, **14**(2), 379–383 (1998).
23. C.E. Lin, T.Z. Wang, T.C. Chiu, C.C. Hsueh, 'Determination of the Critical Micelle Concentration of Cationic Surfactants by Capillary Electrophoresis', *J. High Res. Chromatogr.*, **22**(5), 265–270 (1999).
24. P.R. Majhi, S.P. Moulik, 'Energetics of Micellization: Reassessment by High-sensitivity Titration Microcalorimeter', *Langmuir*, **14**(15), 3986–3990 (1998).
25. C. Cruces Blanco, A. Segura Carretero, A. Fernández Gutiérrez, 'Experimental Design Applied to a Room-temperature Phosphorimetric Method for the Determination of Acenaphthene in a Microemulsion', *Anal. Chim. Acta*, **318**(3), 357–363 (1996).
26. A. Segura Carretero, C. Cruces Blanco, A. Fernández Gutiérrez, 'Application of Variable-angle Synchronous Phosphorimetry in a Microemulsion Medium for the Simultaneous Determination of Three Polyaromatic Hydrocarbons', *Anal. Chim. Acta*, **329**(1/2), 165–172 (1996).

27. S. Scypinski, L.J. Cline Love, 'Room Temperature Phosphorescence of Polynuclear Aromatic Hydrocarbons in Cyclodextrins', *Anal. Chem.*, **56**(3), 322–327 (1984).
28. S. Scypinski, L.J. Cline Love, 'Cyclodextrin-induced Room Temperature Phosphorescence of Nitrogen Heterocycles and Bridged Biphenyls', *Anal. Chem.*, **56**(3), 331–336 (1984).
29. L.J. Cline Love, M.L. Grayeski, J. Noroski, R. Weinberger, 'Room-temperature Phosphorescence, Sensitized Phosphorescence and Fluorescence of Licit and Illicit Drugs Enhanced by Organized Media', *Anal. Chim. Acta*, **170**(1), 3–12 (1985).
30. M.E. Díaz García, A. Sanz-Medel, 'Facile Chemical Deoxygenation of Micellar Solutions for Room Temperature Phosphorescence', *Anal. Chem.*, **58**(7), 1436–1440 (1986).
31. J.J. Donkerbroek, J.J. Elzas, C. Gooijer, R.W. Frei, N.H. Velthorst, 'Some Aspects of Room-temperature Phosphorescence in Liquid Solutions', *Talanta*, **28**(10), 717–723 (1981).
32. J.J. Donkerbroek, J.J. Elzas, C. Gooijer, N.H. Velthorst, R.W. Frei, 'Sensitized Room Temperature Phosphorescence in Liquid Solutions with 1,4-Dibromonaphthalene and Biacetyl as Acceptors', *Anal. Chem.*, **54**(6), 891–895 (1982).
33. E.M. Schulman, C. Walling, 'Triplet-state Phosphorescence of Adsorbed Ionic Organic Molecules at Room Temperature', *J. Phys. Chem.*, **77**(3), 902–905 (1973).
34. D.L. McAleese, R.B. Dunlap, 'Reduction of Background Emission in Room-temperature Phosphorescence', *Anal. Chem.*, **56**(3), 600–601 (1984).
35. A.D. Campiglia, C.G. de Lima, 'Room-temperature Phosphorimetry of Carbaryl in Low-background Paper', *Anal. Chem.*, **59**(23), 2822–2827 (1987).
36. R.T. Parker, R.S. Freelander, E.M. Schulman, R.B. Dunlap, 'Room-temperature Phosphorescence of Selected Pteridines', *Anal. Chem.*, **51**(12), 1921–1926 (1979).
37. C.G. de Lima, M.M. Andino, J.D. Winefordner, 'Effects of Heavy Atom Containing Surfactants in the Room Temperature Phosphorescence of Carbaryl', *Anal. Chem.*, **58**(13), 2867–2869 (1986).
38. G. Ramis Ramos, M.C. Garcia Alvarez-Coque, A.M. O'Reilly, I.M. Khasawneh, J.D. Winefordner, 'Paper Substrate Room-temperature Phosphorimetry of Polyaromatic Hydrocarbons Enhanced by Surface-active Agents', *Anal. Chem.*, **60**(5), 416–420 (1988).
39. T. Vo-Dinh, J.R. Hooyman, 'Selective Heavy-atom Perturbation for Analysis of Complex Mixtures by Room-temperature Phosphorimetry', *Anal. Chem.*, **51**(12), 1915–1921 (1979).
40. T. Vo-Dinh, R.B. Gammage, 'The Applicability of the Second-derivative Method to Room-temperature Phosphorescence Analysis', *Anal. Chim. Acta*, **107**(1), 261–271 (1979).
41. C.D. Ford, R.J. Hurtubise, 'Room-temperature Phosphorescence of the Phthalic Acid Isomers, *p*-Aminobenzoic Acid, and Terephthalamide Adsorbed on Silica Gel', *Anal. Chem.*, **50**(4), 610–613 (1978).
42. C.G. de Lima, M. de Nicola, 'Analytical Application of the Room and Low Temperature (77 K) Phosphorescent Properties of some 1,8-Naphthyridine Derivatives', *Anal. Chem.*, **50**(12), 1658–1665 (1978).
43. A.D. Campiglia, L.M. Perry, J.D. Winefordner, 'Two-nebulizer Automatic System for Solid-surface Room-temperature Phosphorescence Analysis', *Appl. Spectrosc.*, **44**(4), 729–732 (1990).
44. A.D. Campiglia, A. Berthod, J.D. Winefordner, 'Solid-surface Room-temperature Phosphorescence Detection for High-performance Liquid Chromatography', *J. Chromatogr.*, **508**, 37–49 (1990).
45. A.D. Campiglia, D.M. Hueber, T. Vo-Dinh, 'Laser-induced Solid-surface Room-temperature Phosphorimetry of Polycyclic Aromatic Hydrocarbons', *Appl. Spectrosc.*, **50**(2), 252–256 (1996).
46. A.D. Campiglia, D.M. Hueber, T. Vo-Dinh, 'Analysis of Polycyclic Aromatic Compounds in Soil Samples using Laser-induced Phosphorimetry', *Polycyclic Aromat. Compd.*, **8**, 117–128 (1996).
47. R. Jankowiak, G.J. Small, 'Fluorescence Line Narrowing: a High-resolution Window on DNA and Protein Damage from Chemical Carcinogens', *Chem. Res. Toxicol.*, **4**(3), 256–269 (1991).
48. E.L. Wehry, G. Mamantov, 'Matrix Isolation Spectroscopy', *Anal. Chem.*, **51**(6), 643A–655A (1979).
49. J.M. Hayes, G.J. Small, 'Supersonic Jets, Rotational Cooling, and Analytical Chemistry', *Anal. Chem.*, **55**(4), 565A–574A (1983).
50. C.G. de Lima, 'The Shpol'skii Effect as an Analytical Tool', *CRC Crit. Rev. Anal. Chem.*, **16**(3), 177–221 (1986).
51. C. Gooijer, I. Kozin, N.H. Velthorst, 'Shpol'skii Spectrometry, a Distinct Method in Environmental Analysis', *Mikrochim. Acta*, **127**, 149–182 (1997).
52. Y. Yang, A.P. D'Silva, V.A. Fassel, M.K. Iles, 'Laser Excited Shpol'skii Spectrometry (LESS) for the Selective Excitation and Determination of Polynuclear Aromatic Hydrocarbons', *Proc. Soc. Photo-opt. Instrum. Eng.*, **286**, 126–129 (1981).
53. Y. Yang, A.P. D'Silva, V.A. Fassel, 'Deuterated Analogs as Internal Reference Compounds for the Direct Determination of Benzo[*a*]pyrene and Perylene in Liquid Fuels by Laser-excited Shpol'skii Spectrometry', *Anal. Chem.*, **53**(13), 2107–2109 (1981).
54. G.D. Renkes, S.N. Walters, C.S. Woo, M.K. Iles, A.P. D'Silva, V.A. Fassel, 'Determination of Aromatic Hydrocarbons in Particulate Samples by High-temperature Extraction and Shpol'skii Spectrometry', *Anal. Chem.*, **55**(14), 2229–2231 (1983).
55. A.P. D'Silva, V.A. Fassel, 'Laser-excited Shpol'skii Spectrometry', *Anal. Chem.*, **56**(8), 987A–1000A (1984).
56. A.E. ElSaiid, R. Walker, S. Weeks, A.P. D'Silva, V.A. Fassel, 'Site-selective Shpol'skii Spectrometry of Sulfur-, Oxygen-, and Nitrogen-containing Aromatic Compounds

in Complex Samples', *Appl. Spectrosc.*, **42**(5), 731–740 (1988).
57. I.S. Kozin, C. Gooijer, N.H. Velthorst, 'Direct Determination of Isomeric Polycyclic Aromatic Hydrocarbons in Environmental Samples by Conventional and Laser Excited Shpol'skii Spectrometry', *Int. J. Environ. Anal. Chem.*, **61**, 285–297 (1995).
58. J.P. Alarie, T. Vo-Dinh, 'A Fiber-optic Cyclodextrin-based Sensor', *Talanta*, **38**(5), 529–534 (1991).
59. W.A. Chudyk, M.M. Cabarra, J.E. Kenny, 'Remote Detection of Groundwater Contaminants using Far-ultraviolet Laser-induced Fluorescence', *Anal. Chem.*, **57**(7), 1237–1242 (1985).
60. S.M. Inmam, P. Thibado, G.A. Theriault, S.H. Lieberman, 'Development of a Pulsed-laser, Fiber-optic Based Fluorimeter: Determination of Fluorescence Decay Times of Polycyclic Aromatic Hydrocarbons', *Anal. Chim. Acta*, **239**(1), 45–51 (1990).
61. U. Panne, 'Laser Remote Sensing', *Trends Anal. Chem.*, **17**(8/9), 491–500 (1998).
62. R. Niessner, U. Panne, H. Schroder, 'Fiber-optic Sensor for the Determination of Polynuclear Aromatic Hydrocarbons with Time-resolved, Laser-induced Fluorescence', *Anal. Chim. Acta*, **255**(2), 231–243 (1991).
63. T. Vo-Dinh, J.P. Alarie, G.D. Griffin, R.S. Burlage, W. Hill, in *FY-97 Report for the ETPI Project on Autotropic Biofilms: Technical Development for Removing Contaminants from Industrial Wastewater*, Oak Ridge National Laboratory, Oak Ridge, TN, 1998.
64. A.D. Campiglia, J.P. Alarie, T. Vo-Dinh, 'Development of a Room Temperature Phosphorescence Fiber Optic Sensor', *Anal. Chem.*, **68**(9), 1599–1604 (1996).
65. A.D. Campiglia, T. Vo-Dinh, 'Fiber Optic Sensor for Laser-induced Room-temperature Phosphorescence Detection of Polycyclic Aromatic Compounds', *Talanta*, **43**(10), 1805–1814 (1996).
66. E.M. Thurman, M.S. Mills, 'Solid-phase Extraction Principles and Practice', in *Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications*, ed. J.D. Winefordner, John Wiley & Sons, New York, 54–62, Vol. 147, 1998.
67. E.D. Hagestuen, A.D. Campiglia, 'Phosphorimetric Detection of Polycyclic Aromatic Hydrocarbons on Solid-phase Extraction Membranes', *Appl. Spectrosc.*, **52**(8), 1096–1102 (1998).
68. E.D. Hagestuen, A.D. Campiglia, 'New Approach for Screening Polycyclic Aromatic Hydrocarbons in Water Samples Based on Solid-phase Extraction–Room-temperature Phosphorimetry', *Talanta*, **49**(3), 547–560 (1999).
69. A.F. Arruda, A.D. Campiglia, 'Determination of Trace Levels of Polychlorinated Biphenyls on Reversed-phase Octadecyl Bonded Silica Membranes', *Anal. Chim. Acta*, **386**(3), 271–280 (1999).
70. J. Chen, R.J. Hurtubise, 'Solid-phase Microextraction with Whatman 1PS Paper and Direct Room-temperature Solid-matrix Luminescence Analysis', *Talanta*, **45**(6), 1081–1087 (1998).
71. A.H. Ackerman, R.J. Hurtubise, 'Solid-matrix Fluorescence and Phosphorescence and Solid-phase Microextraction of Polycyclic Aromatic Hydrocarbons with Hydrophobic Paper', *Appl. Spectrosc.*, **53**(7), 770–775 (1999).

Mercury Analysis in Environmental Samples by Cold Vapor Techniques

Kim A. Anderson
Oregon State University, Corvallis, USA

1 Introduction **1**
2 Chemistry **2**
2.1 Properties of Mercury 2
2.2 Properties of Mercury Compounds 2
3 Distribution and Uses of Mercury **2**
3.1 Sources of Mercury in the Environment 2
3.2 Technical Uses 3
4 Sample Preparation **3**
4.1 Collection and Storage 3
4.2 Sample Pretreatment 4
4.3 Preparation of Environmental Samples 4
5 Analytical Methodology **5**
5.1 Cold Vapor Technique 5
5.2 Instrumental Developments 6
5.3 Amalgamation 7
5.4 Detection Limits 8
5.5 Interferences 8
5.6 Speciation 9
6 Environmental Regulations **9**
7 Quality Control and Troubleshooting **10**
7.1 Data Quality Objectives 10
7.2 Instrument Performance 10
7.3 Standard Reference Materials 10
8 Method Comparison **11**
8.1 Instrumental Technique Comparison 11
8.2 Other Techniques 11
Acknowledgments **11**
Abbreviations and Acronyms **11**
Related Articles **11**
References **12**

The accurate, precise, and automated determination of mercury in biological and environmental samples is complex, not only because of the various oxidation states of mercury (i.e. 0, I and II), but also because of biotic and abiotic methylation and the volatility of several forms of mercury(I and II). Mercury is unique among the metals because of its large vapor pressure at ambient conditions. Mercuric ions in solution can be reduced by tin(II), or sodium borohydride to produce volatile elemental mercury, referred to as the cold vapor process. The mercury is swept out of solution with a carrier gas (in the gas of sodium borohydride, the byproduct is hydrogen) into the cell (or torch) where the atomic absorption (or fluorescence, or emission) is measured. The cold vapor technique removes the mercury from the sample matrix, concentrates the mercury in the analytical sample into a small plug of carrier gas, and provides for a relatively long residence time. Detection limits vary depending on the instrument but are in the submicrogram to nanogram per liter range for atomic absorption spectrometry (AAS), atomic fluorescence spectrometry (AFS), and inductively coupled plasma atomic emission spectrometry (ICPAES).

1 INTRODUCTION

A naturally occurring element, mercury's unique properties have been valued for several centuries. The metal cycles in the environment as a result of natural and anthropogenic activities. Mercury in most environmental matrices occurs at low concentrations; therefore, sensitive analytical methods are necessary for its detection. Mercury is a known toxicant, and is known to bioaccumulate especially in the aquatic food web. In the 1950s and 1960s two major epidemics of methylmercury poisoning through fish consumption occurred in Minamata Bay and Niigata (both in Japan), where thousands of people were affected. Environmental monitoring of mercury is meaningful. Due to the complexity of mercury biogeochemical cycling, monitoring of mercury requires the analysis of many different matrix types, such as air, waters, soils, sediments, sludges, as well as range of biological matrices. Mercury determination is complicated by the multiple oxidation states of naturally occurring mercury but also because of biotic and abiotic methylation and the volatility of several forms of mercury.

Typically, samples are decomposed in the presence of acid(s) and oxidizing agents (which are necessary to prevent volatilization of mercury). The excess oxidizing agent is then prereduced under mild conditions. The samples are then ready for cold vapor analysis, in which a strong reducing agent, such as tin(II) or sodium borohydride, is added to reduce the mercury to the volatile elemental form. In the case of AAS, the mercury is swept from the system by an inert gas into the cell where the absorption is measured. Similar designs are used for AFS and for ICPAES. The cold vapor technique removes mercury from the liquid matrix, reducing interferences.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

In addition, the technique concentrates all the mercury from the entire sample into a small plug within the carrier gas. Sample sizes as large as 50 mL are routinely used; mercury from the entire 50 mL is purged out as volatile mercury, and aspirated into an AAS, AFS, or ICPAES instrument. This represents a huge increase in sample size as compared to typical aspiration sample delivery. This process is referred to as the cold vapor mercury analysis technique. The sensitivity is several orders of magnitude better than for conventional AAS, AFS, or ICPAES analysis.

2 CHEMISTRY

2.1 Properties of Mercury

The heavy metal mercury (Hg), atomic number 80 and atomic weight 200.59, is a silver-white heavy liquid at typical ambient temperatures and pressures (specific gravity 13.5 $kg\,L^{-1}$ CAS 7439-97-6). Mercury's melting point is −38.89 °C and boiling point (bp) is 357.25 °C under atmospheric pressure. Mercury is the only metal which is a liquid at room temperature. The reason for this is that mercury has a very high ionization energy (first ionization potential = 1007 $kJ\,mol^{-1}$)[1] which makes it difficult for electrons to participate in metallic bonding. Mercury metal has a high vapor pressure (0.16 Pa at 20 °C), and vaporizes readily under ambient conditions; a saturated atmosphere would contain approximately 15 $mg\,Hg\,m^{-3}$ at 20 °C.[2]

Elemental mercury dissolves numerous other metals (gold, silver, platinum, uranium, copper, lead, sodium, and potassium) to alloys, called amalgams. Mercury is unusual among metals because it tends to form covalent bonds rather than ionic bonds.

2.2 Properties of Mercury Compounds

Mercury can exist in three oxidation states: Hg^0 (metallic), Hg_2^{2+} (mercurous), and Hg^{2+} (mercuric), equivalently designated as Hg^0, Hg^I and Hg^{II}. The properties and chemical reactivity of mercury strongly depend on the oxidation state. Mercurous and mercuric mercury form numerous inorganic and organic chemical compounds; however, mercurous mercury is rarely stable under typical environmental conditions. Only a few mercurous compounds are known; mercury(I) is unique in that these compounds contain the mercurous ion $(Hg{-}Hg)^{2+}$, not Hg^+. However, halides of mercurous mercury form readily, and Hg_2Cl_2 is used extensively in chemical applications (often referred to by its trivial name, calomel).

Most mercury in the environment, including waters, soils, sediments and biota, occurs in the forms of inorganic mercuric salts and organomercurics. Organomercurics are defined by the presence of a covalent Hg−C bond. The mercury compounds most likely to be found in the environment (atmospheric mercury being the exception) include the following: mercuric salts, for example $HgCl_2$, $Hg(OH)_2$, and HgS, methylmercury compounds, for example methylmercuric chloride (CH_3HgCl) and methylmercuric hydroxide (CH_3HgOH); and, to a lesser extent, dimethylmercury and phenylmercury.[3] Most organomercurics are not readily soluble and do not react with weak acids or bases due to the low affinity of mercury for oxygen bonded to carbon. However, methylmercuric hydroxide (CH_3HgOH) is highly soluble due to the strong hydrogen bonding capability of the hydroxide group. The mercuric salts vary widely in solubility, $HgCl_2$, $Hg(NO_3)_2$, and $Hg(ClO_4)_2$ are very soluble in water, whereas HgS is insoluble ($K_{sp} \approx 10^{-54}$). Further information on the chemistry of mercury can be found in the excellent discussions by Mason et al.[4] and by Nriagu.[5]

3 DISTRIBUTION AND USES OF MERCURY

3.1 Sources of Mercury in the Environment

As a naturally occurring element, mercury cannot be created or destroyed; mercury is present throughout the environment in both environmental media and biota. Mercury is found at trace levels in many minerals, with continental rocks containing an average of 80 mg Hg per kilogram (i.e. ppb).[6] Soil mercury concentrations may vary dramatically depending on the local geology and industry; typical soils are reported to contain 0.01–0.5 ppm.[7] Mercury is mined as the bright red ore cinnabar (HgS) predominantly in Russia, Spain, Mexico, and Algeria. Mercury is normally present in plant tissues in the range 30–700 ppb.[8] Mercury concentrations in waters vary, typical levels are reported as ranging from 0.0001 to 2.8 $\mu g\,Hg\,L^{-1}$ in fresh water, and from 0.01 to 0.22 $\mu g\,Hg\,L^{-1}$ in seawater.[7] The mercury content in air varies dramatically depending on the source, and is influenced heavily by volcanic and anthropogenic activities. Mercury in clean air is reported at $<0.01\,ng\,m^{-3}$, whereas contaminated air values for mercury range from 0.009–2.8 $ng\,m^{-3}$ in Europe, 0.07–38 $ng\,m^{-3}$ in North America, 1.6 $ng\,m^{-3}$ in Japan, to 18–250 $ng\,m^{-3}$ in Hawaii.[7]

Mercury is emitted into the environment by both natural and anthropogenic (human) processes. Input into the atmosphere occurs by degassing of the earth's crust, emissions from volcanoes, soil erosion, as well as from anthropogenic sources. Most mercury in the atmosphere is elemental mercury vapor. Anthropogenic activities that have contributed to emissions include burning

mercury-containing fuels and materials, and industrial processes. Fossil-fuel coal and lignite contain mercury, often at levels ≥ 100 ppb.[6] The largest contributions to mercury emission from human activities include coal-burning electrical utilities, municipal waste combusters, commercial and industrial boilers, medical waste incinerators, and chlor-alkali plants.[3]

3.2 Technical Uses

Historically, mercury has been used extensively (and still is to a lesser degree) in the extraction of precious metals (for example, gold and silver) as amalgams. The major use of mercury is in electrolytic cells (as the cathode) for the production of NaOH and Cl_2. Metallic mercury is used as a liquid contact material for electrical switches, in vacuum technology in diffusion pumps, thermometers, barometers, tachometers and thermostats, and in mercury-vapor lamps. Mercury is widely used in batteries. The standard calomel (Hg_2Cl_2) electrode, is used as the reference electrode for measurements of potentials in analytical electrochemistry. Mercuric oxide (HgO) has been used in antifouling paint for ships.

Mercury has been used in medicines, cosmetics, and in dentistry as silver amalgams for tooth restorations. Mercury chloride has been used in eye drops for newborns, for syphilis prophylaxis. Although banned in many parts of the world, mercury iodide (3%) or mercury amidochloride (10%) are used as skin-lightening creams and soaps by dark-skinned people.[9]

Organic mercury compounds have been used as pesticides, particularly fungicides. Mercury compounds used in this capacity include aryl mercurics such as phenyl mercuric dimethyldithiocarbamate, which was used as a slimicide and mold-retardant in the paper/pulp industry, and alkyl mercurics such as ethylmercuric chloride (C_2H_5HgCl) which is used as a seed fungicide. Organic mercury compounds are also used for their germicidal properties, for example phenyl mercuric acetate.

4 SAMPLE PREPARATION

4.1 Collection and Storage

Collection and storage are a critical part of any analysis for metals in environmental samples. Questions about trace-metal preservation and storage have resurfaced periodically over the years. These concerns stem from the health risk associated with certain metals, as the requirements for detection limits continue to lower. Because the mercury analytical technique is sensitive, great care in sampling protocols should be used. All sampling, storage and manipulation devices should be mercury free. Because of the risk of mercury in some laboratory environments, all laboratory equipment and sample collection tools should be stored in a clean mercury-free environment. Typical glassware/plasticware cleaning protocols include routine laboratory washing, rinsing with clean (mercury free) water, an acidic rinse and further rinsing with laboratory-grade clean (mercury free) water. The acidic rinses may be $1+1$ nitric acid, $1+1$ hydrochloric acid or aqua regia (three parts concentrated hydrochloric acid and one part concentrated nitric acid). Rinse times reported are for 12–24 h at room temperature, whereas some reports recommend rinsing at 70 °C for 24 h.

The type of analysis and matrix will dictate the sample collection, storage, and preservation protocols. For determination of mercury in waters the specific matrix may be: dissolved, suspended, total, or acid-extractable mercury, and this will dictate the type of collection protocol. Dissolved mercury analysis requires prefiltering samples prior to preserving. Filtration is typically through a 0.45-μm membrane filter. Preservation of other aqueous samples is typically with nitric acid (HNO_3) to pH < 2.0. The quality of nitric acid or other preservation chemicals should match the quality and precision of the analysis.

Unpreserved or acid-preserved mercury water samples may be collected in quartz, Teflon®, linear polyethylene, or borosilicate glass to prevent loss of mercury through disproportionation and volatilization. Polypropylene containers were found to be unsatisfactory unless the preservation included concentrated hydrochloric acid, reportedly to form the soluble $HgCl_4^{2-}$ complex.[10] Mercury concentrations may increase in samples stored in plastic bottles in mercury-contaminated laboratories; therefore, borosilicate or quartz is recommended for precise work. Glassware used in analyses where mercury is used as a reagent should be avoided (for example total Kjeldahl nitrogen (TKN), chloride or chemical oxygen demand (COD) procedures). Dedicated glassware for mercury-only analysis is ideal.

After acidification, it is recommended that samples be stored at 4 °C to minimize evaporation. Under these conditions, mercury is reported to be stable for only 5 weeks.[11] For longer-term storage of mercury, other procedures have been proposed, such as 5.0% HNO_3 and 0.05% $K_2Cr_2O_7$ – for example, 2 mL L^{-1} of 20% (w/v) $K_2Cr_2O_7$ in $1+1$ HNO_3 has been recommended.[12] This preservation procedure for mercury has been reported to be stable for up to 6 months in glass or 30 days in polyethylene containers.[10]

Although soil, sediments, and plant tissue samples may be analyzed without drying, most work is reported on a dry-weight basis. Moisture may be driven off in a drying oven at 60 °C. No mercury losses have been observed

by using this drying step.[3] The dry sample may be pulverized and thoroughly mixed before the aliquot is weighed. Samples should be stored in a dry, mercury-free environment until analysis.

Clinical sample collection and preservation include urine collected in glass or plasticware and preserved with glacial acetic acid or 1.5% nitric acid.[9] Blood samples are collected as blood–EDTA (ethylenediaminetetraacetic acid). Recommended storage is at 4 °C until analysis; samples should be analyzed as soon as possible. For tissue samples immediate freezing is recommended.[13]

4.2 Sample Pretreatment

Several authors recommend preparing mercury standard working solutions with about 2 μg L^{-1} potassium dichromate ($K_2Cr_2O_7$) for enhanced stability.[14,15] However, others have noted at high concentrations, 0.34 mM, that a 5% loss of sensitivity was seen after 24 h.[16] This pretreatment has also been recommended for sample preservation as discussed above.

4.3 Preparation of Environmental Samples

Most matrices require digestion of some type to free mercury from the inorganic and organic forms contained in the sample. Preparation/digestion methods for mercury analysis is still an active area of research.[17,18] However, two sample types which may be analyzed directly include clinical aqueous samples, such as urine, and environmental aqueous samples where the DQO (data quality objective) is for dissolved mercury only. All other matrices require decomposition, typically by wet oxidation mineralization (section 5.6 discusses other digestions). The organic and inorganic matrix must be decomposed (digested) and all mercury present must be converted (oxidized) to mercury(II) prior to the cold vapor process. The oxidized mercury is then reduced to elemental metallic mercury vapor by one of the cold vapor techniques (section 5.1).

The digestion methods (decomposition) may be generalized in that each approach uses a strong acid or combination of acids, such as nitric acid, hydrochloric acid, perchloric acid and/or sulfuric acid, generally in combination with an oxidant such as hydrogen peroxide, potassium permanganate, potassium dichromate, potassium persulfate, or vanadium oxide. The function of the acid(s) is to decompose the inorganic and organic matter of the biological matrix. A host of processes/mechanisms are needed to represent the whole decomposition of solids and/or organic matter; details of these processes are not present here, but an excellent review is available.[17] Briefly, most wet oxidation processes that involve organic matter are oxidative-hydrolytic processes. The samples are then typically treated with an oxidant. The purpose of the oxidant is to oxidize any organically bound mercury together with any unoxidized organic matter. In this way mercury is ensured to be in the soluble (nonvolatile) free form of Hg^{II}. The use of the oxidant potassium permanganate is described in Equation (1):

$$MnO_4^- + (\text{inorganic or organic}) - Hg \longrightarrow MnO_2 + Hg^{2+} + CO_2\uparrow + \text{inorganic salts} \quad (1)$$

Potassium chromate may be used in a similar fashion – samples may be preserved with an oxidant such as chromate to ensure that mercury remains as Hg^{2+}. Equation (2) illustrates how chromium(IV) prevents the loss of mercury through volatilization, by preventing the formation of Hg^0.

$$2CrO_4^{2-} + 16H^+ + 3Hg \longrightarrow 2Cr^{3+} + 3Hg^{2+} + 8H_2O \quad (2)$$

Potassium permanganate oxidizes many of these compounds, but some studies have indicated that a number of organic mercurials, including phenyl mercuric acetate and methyl mercuric chloride, are only partially oxidized by this reagent.[3] However, other studies have reported good recoveries of phenyl mercuric acetate using potassium permanganate.[14] Potassium persulfate has been reported to give good recoveries when used as the oxidant with these compounds.[3]

The digestions are routinely done at modest temperatures below 95 °C, due to the possible loss of mercury at elevated temperatures; for example, the bp of dimethylmercury is 96 °C. The exception to this is if closed-vessel digestions are used, such as microwave closed-vessel techniques or other sealed vessels such as Teflon®- or quartz-lined bombs, where losses of mercury are minimized.

Excess oxidant (e.g. hydrogen peroxide, potassium permanganate, potassium dichromate, potassium persulfate, or vanadium oxide) must be removed prior to the cold vapor chemistry. Excess oxidant will interfere (consume) the reducing reagent(s) in the cold vapor chemistry process, potentially leaving insufficient reagent for the mercury chemistry. The addition of reagents such as oxalic acid,[14] or hydroxylamine hydrochloride[19] reduces excess oxidant such as permanganate without reducing the mercury. Oxalic acid is added to reach the final end-point, which is a colorless solution (i.e. Mn^{2+} is colorless, MnO_4^- is purple, and MnO_2 is a brown precipitate;) Equation (3):

$$MnO_4^- + H_2C_2O_4 + 2H^+ \longrightarrow Mn^{2+} + 2CO_2\uparrow + 2H_2O \quad (3)$$

Oxalic acid does not react with the mercury, leaving it as nonvolatile mercury(II). This mild reduction should be performed immediately prior to the cold vapor process.

A typical digestion procedure might be described as follows: take 1 g of biological material, add 3 mL of concentrated nitric acid to a 10 mL digestion test tube, heat to 95 °C for 12 h, cool and treat dropwise with a saturated solution of potassium permanganate until the solutions remains purple (MnO_4^-) or a brown precipitate persists (MnO_2), heat to 95 °C for 10 min. Add oxalic acid to remove excess oxidant. Dilute to final volume and analyze by CVAAS (cold vapor atomic absorption spectrometry).

4.3.1 Clinical

Most techniques for the determination of mercury in solid matrices require a multistep approach. First, the conversion of all bound mercury in the sample to mercury(II) by a wet oxidation procedure (mineralization) and, second, the reduction of mercury(II) to mercury(0) vapor for analysis (discussed in section 5.1). If mineralization is required, the volatile nature of mercury should be kept in mind. There are numerous approaches to the digestion of clinical samples, as discussed above. Research on the determination of mercury in clinical and biological matrices has been extremely active over the last decade.

Most recently, the analysis of liquid clinical samples such as urine and blood have been analyzed directly, without pretreatment.[20] Blood samples analyzed directly often require the addition of an antifoaming agent (e.g. octanol) to the cold vapor reagents.[20] Numerous publications discuss the determination of mercury in biological matrices using several reagent combinations and equipment. These include microwave digestions[21–24] and various combinations of acidic wet decomposition.[14,18,19,23] Solid biological matrices are digested with acid, often with an oxidant.[13]

Hair samples are often pretreated with an acetone washing prior to digestion. The acetone removes dust and oils from the hair matrix, but otherwise the samples may be treated like other biological solid matrices.

4.3.2 Agricultural

Methods applicable to biological tissues are often validated with agricultural plant matrices.[14,22,25] Agricultural materials have been digested by a host of techniques, such as nitric–sulfuric acids with vanadium oxide.[15,26–28] Potassium permanganate is a favored oxidant for biological and agricultural matrices (including soils).[13,14,29] Closed-vessel digestions are gaining popularity for digesting agricultural matrices (plant material, soils and biological material) as microwave digestion ovens become more common in analytical laboratories.[29]

4.2.3 Waters

Depending on the DQO, the water sample type may be dissolved, suspended, total, or total recoverable. For dissolved mercury the samples are filtered prior to preservation, with a 0.45 μm filter. Suspended mercury is the sample left on the filter. Total mercury is the complete digestion of the entire original sample, and total recoverable is a caveat that some digestion techniques have yet to demonstrate recoveries from all forms of mercury. Digestion of water matrices may proceed as outlined above; increased sensitivity may be achieved by using larger samples.

5 ANALYTICAL METHODOLOGY

5.1 Cold Vapor Technique

Mercury is the only element that exists in metallic form at ambient conditions and at the same time exhibits a considerable vapor pressure. These unique properties enable unique methods of detection to be exploited. In addition, the traditional techniques, such as flame AAS, AFS, or ICPAES, exhibit poor mercury sensitivity. As mercury has a high vapor pressure (0.16 Pa at 20 °C), mercury may be determined by AAS without the use of an atomizer. Mercury must be simply reduced to metallic mercury from its compounds and transferred as the vapor phase. This is accomplished by a chemical reduction reaction used to generate the gaseous mercury species.

There are two primary advantages of the cold vapor process. First, the mercury analyte is removed from the sample matrix, which reduces the potential for matrix interferences. Second, the detection limits are improved because all of the mercury sample is introduced into the atomizer (or nebulizer in the case of ICPAES) within a few seconds. Therefore, the density of mercury in the cell during data collection (absorption, fluorescence, or emission, depending on the detection technique) is greatly enhanced as compared to typical sample introduction.

The two reducing agents used exclusively for cold vapor analysis are tin(II) and sodium borohydride. Tin(II) chloride ($SnCl_2$) was used in most early applications of mercury cold vapor analysis; more recently sodium borohydride ($NaBH_4$; the correct nomenclature is sodium tetrahydridoborate) has gained some favor. Tin(II) chloride reacts with mercury as described by Equation (4):

$$Sn^{2+} + Hg^{2+} \longrightarrow Sn^{4+} + Hg^0 \qquad (4)$$

The tin chloride technique requires that the metallic mercury be transported by an inert gas stream (for example argon or nitrogen) bubbled (purged) through the solution to drive out the mercury to the absorption

cell. Later applications use tin sulfate (stannous sulfate) instead of tin(II) chloride.[29,30]

Sodium borohydride in an alkaline solution is becoming the preferred reagent; it requires no other reagents for the reduction. Sodium borohydride is a stronger reducing agent than tin(II) chloride. In addition, sodium borohydride produces dihydrogen as part of the reduction chemical reaction. The hydrogen can therefore transport the majority of the metallic mercury from the solution into the absorption cell. The sample digestate is acidic from the mineralization process (Equation 5):

$$BH_4^- + H^+ + 3H_2O \longrightarrow H_3BO_3 + 8H^\bullet + Hg^{2+} \longrightarrow Hg^0 + 4H_2 \quad (5)$$

Sodium borohydride, which is prepared in basic media (e.g. NaOH) for stabilization purposes, should be prepared daily. Because of the production of hydrogen, the borohydride technique is limited to open systems.

5.2 Instrumental Developments

The cold vapor method for mercury has been described for several instruments, including AAS (most favored), AFS, and ICPAES. The earliest report of using mercury vapor, to monitor mercury in air, was in 1939 by Woodson.[31] Later the use of Sn^{II} to produce mercury vapor was reported by Kimura and Miller,[32] and this approach was subsequently optimized by Poluektov et al.[33] One of the early descriptions of CVAAS for the analysis of mercury in environmental samples, soil and rock, was by Hatch and Ott in 1968.[34] Cold vapor generators of numerous special designs have subsequently been described in the literature.[35]

Overall, the CVAAS technique may be categorized as batch or flow systems. The AAS procedure is a physical method based on the absorption of radiation at 253.7 nm by mercury vapor. The mercury is reduced to the elemental state by the cold vapor process and aerated from solution. All of the mercury in the sample solution placed in the reaction vessel is chemically atomized and transported to the sample cell for measurement. The mercury vapor passes through a cell positioned in the light path of an atomic absorption spectrophotometer. Absorbance (peak height) is measured as a function of mercury concentration.

In the batch mode, specific volumes of sample and reductant are mixed at the start of the process, which culminates in the generation of a transient absorbance signal. A typical batch system for cold vapor generation is shown in Figure 1. The advantages of a batch system are that large volumes may be used, which increase sensitivity; sample volumes of 50 mL are routinely reported. The disadvantages include large reagent consumption, large dead volume, and the method is relatively slow and labor intensive.

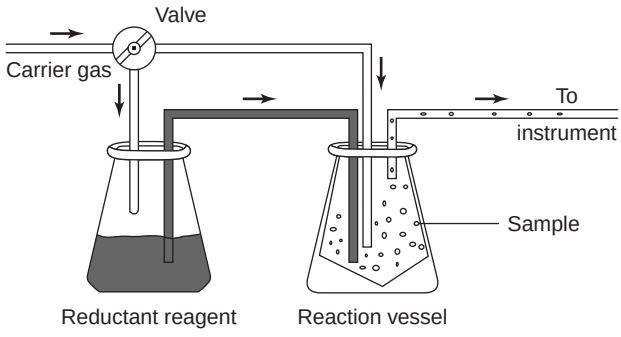

Figure 1 Batch system for mercury cold vapor generation.

In the flow injection system reagents are continuously pumped, the sample is injected into the carrier flow, and the flow then ultimately goes through a reaction (mixing) coil and a gas–liquid separator, thereby generating a transient absorption signal. In a continuous flow system the reagent and sample are continuously pumped, and the system flow proceeds as described above for flow injection. As the sample and reagents are continuously pumped through the system, this technique generates a continuous, flame-like absorbance signal. A typical continuous flow system for cold vapor generation is shown schematically in Figure 2. The advantages of a flow injection system are that smaller sample volumes are used, with consequent lower reagent consumption. The continuous and flow injection techniques are easily automated. Dead volume is reduced, which results in greater absolute sensitivities. Overall, the batch and flow techniques have similar absolute sensitivities, although batch systems use larger sample volumes to achieve the same sensitivity. Both systems have unique attributes and their suitability will depend on the application.

Any atomic absorption unit having an open sample presentation area capable of mounting an absorption cell is suitable. Instrument operating parameters are set according to the manufacturer. Instruments designed

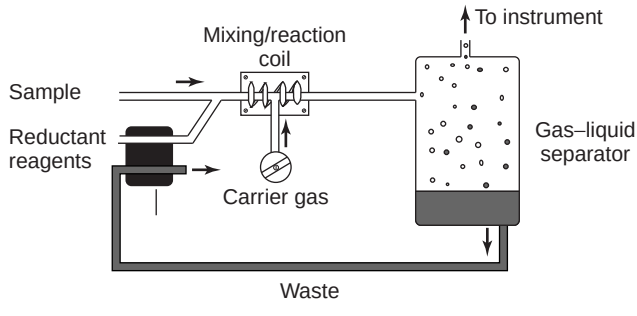

Figure 2 Flow system for mercury cold vapor generation.

specifically for the measurement of mercury using the cold vapor technique are commercially available and may be substituted for the atomic absorption spectrophotometer. Standard spectrophotometer cells 10 cm long and having quartz end windows may be used; alternatively, suitable cells may be constructed from plexiglass tubing.

Cold vapor for AFS works on the same principles as for AAS. The increase in sensitivity is a function of the selectivity of introducing mercury vapor into the instrument detection system, while eliminating most of the rest of the matrix constituents through the liquid waste. A CVAFS (cold vapor atomic fluorescence spectrometry) system consists of a vapor generator, mercury fluorescence detector, and measurement/computer system. The mercury vapor is removed in a gas–liquid separator by an inert gas (e.g. argon), the mercury is piped through a chimney past a light source and photomultiplier that are at right angles to each other. As the vapor absorbs the mercury light, it fluorescences at the same wavelength (typically 254.7 nm). The design allows for the efficient isolation of the required excitation and emission wavelength for sensitive quantitation of mercury. Numerous cold vapor techniques have proven satisfactory for this technique. Instrument operating parameters are set according to each particular manufacturer. Mercury sensitivities of 4–10 ng g^{-1} have been reported.[36]

Cold vapor may also be used in ICPAES applications, i.e. CVICPAES (cold vapor inductively coupled plasma atomic emission spectrometry). Mercury(II) is reduced by sodium borohydride in a continuous manifold. A standard pneumatic nebulizer effects the gas–liquid separation of mercury and concentrates the mercury into a small plug of gas, which is carried to the plasma. A cold vapor flow system for ICPAES is shown in Figure 3. This also provides longer residence time in the plasma, which further increases sensitivity.[14] The introduction of volatile mercury increases the sensitivity by over 1000-fold as compared to liquid sample introduction for ICPAES analysis. This increase is because the transport efficiency of the gaseous mercury to the plasma is higher. In addition, atomization and excitation of the gaseous mercury is very efficient because the process avoids the energy-consuming desolvation and vaporization phases associated with liquid samples. Matrix effects are also minimized as most of the other constituents in the samples are in the liquid fraction and are directed past the sample introduction system in the ICPAES and into the waste stream. Instrument operating parameters are set according to each particular manufacturer. Mercury is quantitated by ICPAES at 194.232 nm. The sensitivity is reported to be 0.2 μg L^{-1}. Gas–liquid separators for ICPAES are commonly used and are available from many manufacturers; they may be used with either tin(II) or sodium borohydride. These same gas–liquid separators are also used for hydride generation for ICPAES.

5.3 Amalgamation

As the need for lower and lower detection limits is required by the analytical community, further enhancement of the cold vapor technique has continued to evolve. Again, capitalizing on the unique properties of mercury affords yet another development in the enrichment of the cold vapor technique. Mercury readily forms amalgamations with precious metals. With this technique, mercury is reduced to Hg^0 vapor and transported by an inert gas (for example argon) in the typical manner used in the cold vapor technique. In this application, the mercury vapor is passed through an absorber,[37] for example a gold gauze, where it is retained (Figure 4). In this way, mercury may be collected from large sample volumes or over extended periods, or both, thereby permitting high preconcentration factors. The mercury is subsequently released by rapidly heating the absorber to about 600 °C and mercury is determined in the typical manner. Detection limits of 3 ng L^{-1} have been reported.[9]

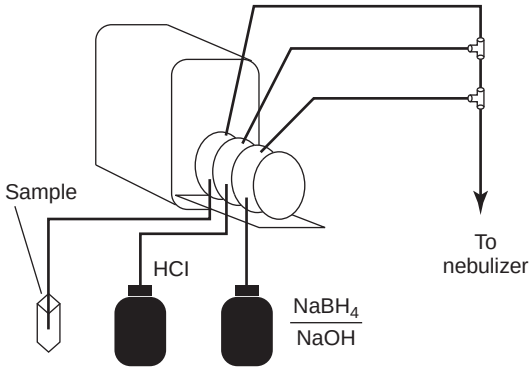

Figure 3 Continuous flow manifold for mercury cold vapor generation for ICPAES.

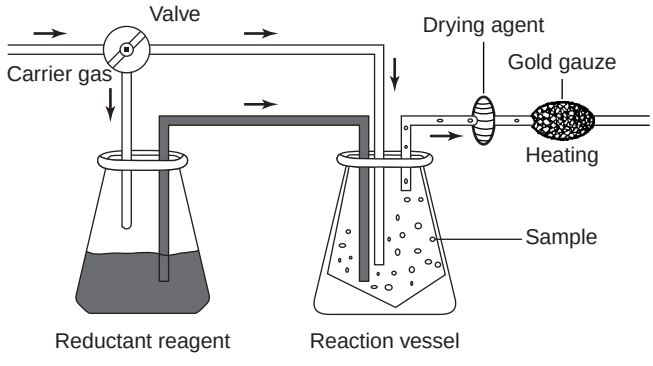

Figure 4 Batch mercury cold vapor generator with drying agent and gold amalgamation preconcentration.

Table 1 Mercury cold vapor detection limits

Cold vapor technique	Mercury sensitivities ($\mu g\,L^{-1}$)	Refs.
CVAAS	0.2	3, 9, 38
CVAFS	0.004–0.010	3, 36, 38
CVICPAES	0.2	14
Amalgamation CVAAS	0.003	9, 22, 38

5.4 Detection Limits

Detection limits depend in some measure on the matrix to be analyzed. The matrix may also dictate the type of analytical approach, together with the sensitivity requirements (Table 1).

5.5 Interferences

The most frequent source of error in the determination of mercury is typically the digestion which either does not break down the organomercury compounds or does not convert all of the mercury to the proper oxidation states. The other major source of error is mercury contamination from the laboratory environment. Typically, spectral and chemical interferences are of minor consequence in comparison. One of the primary advantages of the cold vapor process is that the analyte (mercury) is removed from the sample matrix, which reduces the potential for interferences. The matrix constituents are normally left in the reaction vessel. Therefore, due to the small number of constituents in the gas phase in the atomizer (nebulizer) spectral interferences are virtually excluded. Possible interference from water vapor was thought to interfere with the absorption at 253 nm in AAS analyses; this was later determined not to be a spectral interference.[38] The water interference was due to droplets of solution, which were carried by the gas stream and subsequent condensation of the vapor in the absorption cell, reducing signal intensity. The water interference can be eliminated by either using a suitable drying agent or by warming the AAS absorption cell.

Fortunately, chemical interferences are nearly excluded for the two most common techniques used in the mercury cold vapor generation, namely Sn^{II} and sodium borohydride. Metals that can be reduced form fine particulate material, which can retain mercury. Tin(II) chloride is not a strong reductant; therefore, it does not reduce many other metals which may be present in the matrix. Selenium and iodide are the only elements that have appreciable effects on the Sn^{II} reaction.[38] Sodium borohydride is a strong reductant; copper, and nickel, may form particulates which can retain mercury. The concentrations in most typical environmental matrices, which are required of these elements to interfere, seldom exceed the range of interference-free determination. Hazardous waste or other matrices with unique or high concentrations of these trace elements should be evaluated individually for potential interferences.

The hydride-forming elements may cause interference in the gas phase; Sn^{II} is a weak reductant and therefore represents a minor source of error. Sodium borohydride easily forms hydrides (with As, Se, etc.) and therefore can represent a large source of error, depending on the concentration of these hydride-forming elements.

Sodium borohydride also may reduce metals such as gold, silver, palladium, and platinum; mercury can form amalgams with these metals, thereby causing a suppression of the mercury signal.[38] It has been shown that increasing acid concentration in the sample solution reduces interference from these metals.[9] It has also been shown that the flow injection AAS system is less susceptible to interferences of this type. These types of interference are caused by the reduced form of the interferent, because the reduction takes some finite time; if flow conditions are optimized the mercury vapor reaches the absorption cell before significant amounts of interferent are precipitated and retain mercury (a kind of kinetic discrimination[38]).[9]

Nitric acid digestions used with sodium borohydride as the reducing agent in a flow injection AAS can cause poor reproducibility for mercury determination. It has been reported that the interference was due to inhibition of the mercury reduction by the volatile nitrogen oxides which were scavenging the reducing agent.[39] This interference can be overcome by purging the sample with an inert gas (e.g. argon).

Seawaters, brines and industrial effluents high in chlorides require additional permanganate. During the oxidation step, chlorides are converted to free chlorine, which will also absorb radiation at 253 nm.[3] This causes an enhancement of the signal. Care must be taken to ensure that free chlorine is absent before the mercury is reduced and swept into the cell. This may be accomplished by using an excess of hydroxylamine sulfate reagent. Alternatively, a sulfuric acid/dichromate digestion has been used successfully on high-chloride-content matrices.[38] Other reports ascribe the interference of chlorine on the mercury cold vapor technique to the re-oxidation of elemental mercury by chlorine inside the reduction–aeration system, causing suppression of mercury recovery.[40] Nonchemical removal of the chlorine interference has been accomplished by the use of an additional heating step after digestions, which allows evaporation of chlorine yet retains mercury in the sample solution.

Report on CVICPAES analyses indicate no chemical interference for sodium, calcium, potassium, magnesium, chromium, manganese, iron, cobalt, nickel, copper, aluminum or zinc at 0.4%.[14] Copper at or above 1000 $\mu g\,g^{-1}$

did interfere with the mercury emission signal,[14] copper was observed to be plating out on the tubing.

Possible interference from sulfide is eliminated by the addition of potassium permanganate. Concentrations as high as $20\,mg\,L^{-1}$ of sulfide as sodium sulfide do not interfere with the recovery of added inorganic mercury from distilled water.[3]

5.6 Speciation

Speciation of mercury is meaningful because of the toxicity of certain mercury compounds, and the importance of understanding the movement, bioaccumulation, and transformations of mercury. Methylation of mercury is the critical reaction in the entry of mercury into the food chain. Biotransformation of inorganic mercury species to methylated organic mercury species occurs in the sediment and water column. However, not all inorganic mercury is converted to methylmercury; reactions such as demethylation and volatilization also occur.[3] The rates of these reactions are not well understood. All forms of mercury accumulate to some degree in organisms, but methylmercury accumulates to a much greater extent. In addition, elimination of methylmercury by organisms (e.g. fish-eating birds, fish, and humans) is slow compared to inorganic forms of mercury. The toxicological risks of methyl mercury compounds are therefore more significant.

Speciation of mercury is a relevant analytical goal because, as yet, no single method permits unequivocal identification of all mercury species. More often, broad classes of mercury such as ''inorganic mercury'' and ''organic mercury'' are commonly reported. One of the early works on mercury speciation was based on the preparation of a biological sample using alkaline digestion,[41,42] often referred to as the Magos method. The method depends in part on the reaction rate kinetics for conversion of methylmercury by the reductant to be slow compared to the analytical process. The sample preparation is based on alkaline digestion with NaOH and cysteine. Briefly, about 1 g of sample is solubilized in 10 mL of 1:1:1 1% NaCl, 45% NaOH and 1% cysteine at 80 °C.[43] The cysteine is thought to complex with mercury and minimizes mercury volatility.[44] The high-pH sample is then reduced with the addition of tin(II) to form metallic mercury vapor as described in Equation (4); the sample is then analyzed in typical fashion by AAS. The mercury determined in this manner is termed inorganic mercury. Total mercury is analyzed by the same procedure, but a Sn–Cd reductant is used which reduces inorganic and organic mercury to metallic mercury vapor. Organic mercury is estimated as the difference between total and inorganic mercury. Current reports indicate that this technique overestimates the concentration of inorganic mercury in the presence of methylmercury.[43] Degradation of methylmercury during the formation of the analytical signal is thought to account for the overestimate.[43] Similar approaches, such as staging digestions, have been used with some success.

Most recent reports depend on hyphenated methods for speciation,[45,46] such as liquid chromatography (LC)/ICPMS, (inductively coupled plasma mass spectrometry)[47] electrothermal (ET)/ICPMS,[48] and gas chromatography (GC)/CVAFS.[49] Other speciation methods include physical/mechanical separation (separation based on molecular size), and potentiometric methods (separation based on differences in electrical potential).

6 ENVIRONMENTAL REGULATIONS

Concentrations of mercury in air and water are typically low, and pose little health risk. However, once mercury enters water, either directly or through air deposition, it can bioaccumulate in fish and animal tissue as methylmercury. Human exposure to mercury occurs primarily through eating contaminated fish. Exposure to high levels of mercury has been associated with serious neurological and developmental effects in humans.[3]

Table 2 lists some examples of environmental mercury standards, including some drinking water limits from various agencies and countries. Because of the diverse mechanisms by which mercury may be introduced into the environment, mercury regulations often span several agencies. The USEPA regulates mercury in pesticides and mercury releases into the environment (air, water, and land disposal), the FDA regulates mercury in

Table 2 Environmental regulatory standards for mercury[3,7]

Environmental matrix	Mercury limit ($\mu g\,L^{-1}$) or ($\mu g\,Kg^{-1}$)	Agency or country
Fish	1000	FDA (USA)
Ingesting aquatic organisms and water	0.144	USEPA
Air	0.01 mg m^3 eight hour limit	USOSHA
Drinking water	1.0	World Health Organization
Drinking water	2.0	USA
Drinking water	0.5	Japan
Drinking water	1.0	Chinese (proposed)
Drinking water	1.0	UK

FDA, Food and Drug Administration; USEPA, United States Environmental Protection Agency; USOSHA, United States Occupational Safety and Health Administration.

Table 3 Summary of USEPA methods for mercury analysis

Method number	Matrix	Digestion, instrument, sensitivity, comments
0060	stack emissions	Hg collection procedure, analysis by 7470 (CVAAS)
200.15	aqueous	ICPAES ultrasonic nebulizer, 3 μg L^{-1}
245.1	aqueous	CVAAS, batch system, 0.2 μg L^{-1}
245.2	aqueous	CVAAS, flow system, 0.2 μg L^{-1}
245.5	sediments, solids	CVAAS, batch system, 0.2 μg g^{-1}, aqua regia digestion
1620	all environmental	consolidation of methods 245.1, 245.2, and 245.5
1631	aqueous	CVAFS, oxidation, purge and trap, 0.5 ng L^{-1}
1669	all environmental	Hg sampling ambient water protocol
3051	soils, sludges, etc.	digestion procedure only, microwave–nitric acid
3052	organic material	digestion procedure only, microwave–nitric acid
6010B	predigested	ICPAES method only, detection limit ca. 17 μg L^{-1}
7470A	liquid waste, etc.	CVAAS, batch system, nitric/sulfuric acids, $KMnO_4$
7417A	solid waste, etc.	CVAAS, aqua regia, $KMnO_4$
7472	solids, aqueous	ASV, mercury(II) only, 0.1 μg L^{-1}

food, cosmetics and dental products, and the USOSHA regulates mercury in air exposure in the workplace.

The USEPA program for water, Contract Laboratory Program (CLP), and the Resource Conservation and Recovery Act (RCRA) accept results from several different digestions/instruments. A summary of USEPA methods for the determination of mercury is given in Table 3.

7 QUALITY CONTROL AND TROUBLESHOOTING

7.1 Data Quality Objectives

The DQOs are typically project defined. Spikes and check standards will typically have >95% recovery. Concentrations close to the detection limit will have inherently more variability. Typically, recoveries of SRMs (standard reference materials) are reported at >85% for total mercury. Speciation of mercury is currently not reported to be nearly as accurate, with some species of mercury recovered at <85%. Speciation of mercury is still an active research area.

7.2 Instrument Performance

A good, general preventive maintenance program will keep equipment in top performance longer and therefore is highly recommended. A review of troubleshooting is beyond the scope of this article and has been covered elsewhere.[10,13,28] Briefly, in the case of erroneously low results or no signal, cold vapor reagents should be checked to ensure they are mixing and have not deteriorated; other items to check are gas lines, contamination of blank, and matrix interferences. In the case of erroneously high results, check that blank correction was performed properly and that standards have not deteriorated, also check for memory effects and matrix interferences.

7.3 Standard Reference Materials

A minimum number of quality control (QC) samples should be analyzed as part of the QC program defined by the DQOs, as follows. Laboratory blanks prepared with the samples are a matrix blank, used to cancel interference or contamination from reagents or analytical manipulations and should be included in the analytical batch. Laboratory spikes should be analyzed for. An example would be to prepare by spiking a sample with 50 ng mL^{-1} of 1 mg L^{-1} mercury solution and calculating the percentage recovery; other concentrations may be more appropriate depending on the specific application. If the recovery falls within 85–115%, the analysis has passed QC; the recovery window is defined by the project DQOs. SRM should be included in the analytical batch. Depending on the matrix, the best possible similar SRM should be included. Examples are: NIST (National Institute of Standards and Technology), oyster tissue, NIST pine needles, NIST bovine liver, and NIST citrus leaves (for biological matrices); NIST Buffalo River sediment, National Research Council of Canada (NRC) estuarine sediment, NRC harbour sediment, BCR (European Commission, Belgium) river sediment, BCR light sandy soil, BCR sewage sludge amended soil, BCR sewage sludge of domestic origin (for soils, sludge, and sediment matrices). For aqueous environment samples NIST, NRC, BCR, or other vendor's trace-metal SRMs should be included. The SRM should have mercury levels within the concentration ranges that are expected in the samples to be analyzed. In-house SRM may be substituted if a sufficient confidence level can be demonstrated with the mercury concentration. The frequency of the QC samples depends on the project DQOs; 10% of each QC type is often used as a default frequency; SRMs for the speciation of mercury are currently inadequate/unavailable.

Table 4 Instrument sensitivities for mercury non-cold-vapor techniques

Technique	Mercury sensitivities ($\mu g\,L^{-1}$)	Refs.
Anodic stripping voltametry	0.5–1.0	9, 50
Flame AAS	2200	50, 51
ETAAS	0.5	50, 51
ICPAES	10–50	50, 51
NAA	3	50, 51
ICPMS	0.01–0.001	51, 52
AFS	0.2	50, 51

ETAAS, electrothermal atomic absorption spectrometry; NAA, neutron activation analysis.

8 METHOD COMPARISON

8.1 Instrumental Technique Comparison

Detection limits depend in some measure on the matrix to be analyzed. In addition, the matrix dictates the type of analytical technique approach, in addition to sensitivity requirements. Detection limits using the cold-vapor process for detection of mercury are given in Table 1. Instrument detection limits for non-cold-vapor techniques are given in Table 4.

8.2 Other Techniques

There are numerous approaches to mercury analysis and some techniques require fairly specialized equipment. Mercury in environmental matrices has been routinely determined by several techniques,[13] including CVAAS, ETAAS,[53] AFS,[54,55] CVICPAES,[14] GC/MS,[56] ICPMS,[52,57–59] and NAA.[60,61] Examples of some selected mercury methods specifically for water and environmental aqueous solutions include LC/ICP/MS (liquid chromatography/inductively coupled plasma/mass spectrometry),[62,63] X-ray fluorescence (XRF),[64] and spectrophotometric techniques.

ACKNOWLEDGMENTS

The author would like to thank Wes Stone and Brandon Isaacs for assistance in the creation of the illustrations.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AFS	Atomic Fluorescence Spectrometry
BCR	European Commission, Belgium
bp	Boiling Point
CLP	Contract Laboratory Program
COD	Chemical Oxygen Demand
CVAAS	Cold Vapor Atomic Absorption Spectrometry
CVAFS	Cold Vapor Atomic Fluorescence Spectrometry
CVICPAES	Cold Vapor Inductively Coupled Plasma Atomic Emission Spectrometry
DQO	Data Quality Objective
EDTA	Ethylenediaminetetraacetic acid
ET	Electrothermal
ETAAS	Electrothermal Atomic Absorption Spectrometry
FDA	Food and Drug Administration
GC	Gas Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LC	Liquid Chromatography
LC/ICP/MS	Liquid Chromatography/Inductively Coupled Plasma/Mass Spectrometry
NAA	Neutron Activation Analysis
NIST	National Institute of Standards and Technology
NRC	National Research Council of Canada
QC	Quality Control
RCRA	Resource Conservation and Recovery Act
SRM	Standard Reference Material
TKN	Total Kjeldahl Nitrogen
USEPA	United States Environmental Protection Agency
USOSHA	United States Occupational Safety and Health Administration
XRF	X-ray Fluorescence

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges)

• Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Soxhlet and Ultrasonic Extraction of Organics in Solids

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Atomic Spectroscopy for Forensic Applications

General Articles **(Volume 15)**
Traceability in Analytical Chemistry

REFERENCES

1. J.D. Lee, *Concise Inorganic Chemistry*, 4th edition, Chapman & Hall, NY, 1991.
2. R.C. Weast, M.J. Astle, *Handbook of Chemistry and Physics*, 64th edition, CRC Press, Boca Raton, Florida, 1983.
3. United States Environmental Protection Agency, *Mercury Study Report to Congress*, Volume III, Fate and Transport of Mercury in the Environment, Document #EPA-452/R-97-005, USEPA, 1997.
4. R.P. Mason, W.F. Fitzgerald, F.M.M. Morel, 'The Biogeochemical Cycling of Elemental Mercury: Anthropogenic Influences', *Geochim. Cosmochim. Acta*, **58**(15), 3191–3198 (1994).
5. J.O. Nriagu, *The Biogeochemistry of Mercury in the Environment*, Elsevier, NY, 1979.
6. S.E. Manahan, *Environmental Chemistry*, 5th edition, Lewis Publishers, Chelsea, MI, 1991.
7. I. Bodek, W.J. Lyman, W.F. Reehl, D.H. Rosenblatt, *Environmental Inorganic Chemistry*, Pergamon Press, NY, 1988.
8. A. Kabata-Pendias, H. Pendias, *Trace Elements in Soils and Plants*, revised edition, CRC Press, Boca Raton, FL, 1995.
9. H.G. Seiler, A. Sigel, H. Sigel, *Handbook on Metals in Clinical and Analytical Chemistry*, Marcel Dekker, NY, 1994.
10. R.K. Smith, *Handbook of Environmental Analysis*, Genium Publishing, NY, 1993.
11. C. Feldman, 'Preservation of Dilute Mercury Solutions', *Anal. Chem.*, **46**, 99–102 (1974).
12. A.E. Greenberg, L.S. Clesceri, A.D. Easton (eds.), *Standard Methods for the Examination of Waste and Wastewater*, 18th edition, ALPHA Publications, Washington, DC, 1992.
13. K.A. Anderson, *Analytical Techniques for Inorganic Contaminants in Veterinary Toxicology*, Agriculture and the Environment, AOAC International Publishing, Arlington, VA, 1999.
14. K.A. Anderson, B. Isaacs, M. Tracy, G. Moller, 'Cold Vapor Generation for Inductively Coupled Argon Plasma Atomic Emission Spectrometric Analysis. Part 3 Mercury', *J. AOAC Int.*, **77**(2), 473–480 (1994).
15. J.R. Knechtel, J.L. Frazer, 'Wet Digestion Method for the Determination of Mercury in Biological and Environmental Samples', *Anal.Chem.*, **51**, 315–317 (1979).
16. P. Dominski, D.E. Shrader, *Automated Cold Vapor Determination of Mercury: EPA Stannous Chloride Methodology*, Publication No. aa-51, Varian Associates Inc., CA, 1985.
17. Z. Sulcke, P. Povondra, *Methods of Decomposition in Inorganic Analysis*, CRC Press, Boca Raton, FL, 1989.
18. S.B. Adeloju, H.S. Dhindsa, R.K. Tandon, 'Evaluation of Some Wet Decomposition Methods for Mercury Determination in Biological and Environmental Materials by Cold Vapor Atomic Absorption Spectroscopy', *Anal. Chim. Acta*, **285**, 359–364 (1994).
19. J. Pineau, M. Piron, H.L. Boiteau, M.J. Etourneau, O. Guillard, 'Determination of Total Mercury in Human Hair Samples by Cold Vapor Atomic Absorption Spectrometry', *J. Anal. Toxicol.*, **14**, 235–238 (1990).
20. DFG, German Science Foundation, *Analyses of Hazardous Substances in Biological Materials*, eds. J. Angerer, K.H. Schaller, Weinheim, Vol. 2, 1985.
21. M. Gallignani, H. Bahsas, M.R. Brunetto, M. Burguera, J.L. Burguera, Y. PitidePena, 'A Time Based Flow Injection Cold Vapor Atomic Absorption Spectrometry System with On Line Microwave Sample Pretreatment for the Determination of Inorganic and Total Mercury in Urine', *Anal. Chim. Acta*, **369**, 57–67 (1998).
22. J.M. Ombaba, 'Total Mercury Determination in Biological and Environmental Standard Samples by Gold Amalgamation followed by Cold Vapor Atomic Absorption Spectrometry', *Microchem. J.*, **53**, 195–200 (1996).
23. U. Tinggi, G. Craven, 'Determination of Total Mercury in Biological Materials by Cold Vapor Atomic Absorption Spectrometry After Microwave Digestion', *Microchem. J.*, **54**, 168–173 (1996).
24. L. Aduna de Paz, A. Alegrai, R. Barbera, R. Farre, M.J. Lagarda, 'Determination of Mercury in Dry Fish Samples by Microwave Digestion and Flow Injection Analysis System Cold Vapor Atomic Absorption Spectrometry', *Food Chem.*, **58**(1–2), 169–172 (1997).
25. M. Navarro-Alarcon, M.C. Lopez-Martinez, M. Sanchez-Vinas, H. Lopez-Garcia de la Serrana, 'Determination of Mercury in Crops by Cold Vapor Atomic Absorption Spectrometry after Microwave Dissolution', *J. Agric. Food Chem.*, **38**, 2223–2225 (1991).
26. M. Malaiyandi, J.P. Barrette, 'Wet Digestion Methods for the Determination of Submicrogram Quantities of Mercury in Cereal Grains', *J. AOAC*, **55**, 951–959 (1972).
27. F.D. Deitz, J.L. Sell, D. Bristol, 'Rapid, Sensitive Method for Determination of Mercury in a Variety of Biological Samples', *J. AOAC*, **56**, 378–382 (1983).
28. Y.P. Kalra, *Handbook of Reference Methods for Plant Analysis*, CRC Press, Boca Raton, FL, 1998.

29. W. van Delft, G. Vos, 'Comparison of Digestion Procedures for the Determination of Mercury in Soils by Cold Vapor Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **209**, 147–156 (1988).
30. B.T. Sturman, 'Development of a Continuous Flow Hydride and Mercury Vapor Generation Accessory for Atomic Absorption Spectrophotometry', *Appl. Spectrosc.*, **39**, 48–56 (1985).
31. T.T. Woodson, 'A New Mercury Vapor Detector', *Rev. Sci. Instrum.*, **10**, 308–311 (1939).
32. Y. Kimura, V.L. Miller, 'Mercury Determination at the Microgram Level by Reduction–Aeration Method of Concentration', *Anal. Chim. Acta*, **27**, 325–331 (1962).
33. N.S. Poluektov, R.A. Vitkun, Y.V. Zelyukova, 'Determination of Milligram Amounts of Mercury by Atomic Absorption in the Gaseous Phase', *Zh. Analit. Khim.*, 937–942 (1964).
34. W.R. Hatch, W.L. Ott, 'Determination of Sub-microgram Quantities of Mercury by Atomic Absorption Spectrometry', *Anal. Chem.*, **40**(14), 2085–2087 (1968).
35. F.M. Bauza De Mirabo, A.Ch. Thomas, E. Rubi, R. Forteza, V. Cerda, 'Sequential Injection Analysis System for Determination of Mercury by Cold-vapor Atomic Absorption Spectroscopy', *Anal. Chim. Acta*, **355**, 203–210 (1997).
36. R.D. Jones, J. West-Thomas, C. Arfstrom, 'Closed-ampule Digestion Procedure for the Determination of Mercury in Soil and Tissue using Cold Vapor Atomic Fluorescence Spectrometry', *Bull. Environ. Contam. Toxicol.*, **59**, 29–34 (1997).
37. K.W. Jackson, H. Qiao, 'Atomic Absorption, Atomic Emission and Flame Emission Spectrometry', *Anal. Chem.*, **64**, 50R–66R (1992).
38. B. Welz, *Atomic Absorption Spectrometry*, 2nd edition, VCH, Germany, 1985.
39. I. Rokkjaer, B. Hoyer, N. Jensen, 'Interference by Volatile Nitrogen Oxides in the Determination of Mercury by Flow Injection Cold Vapor Atomic Absorption Spectrometry', *Talanta*, **40**(5), 729–735 (1993).
40. S. Landi, F. Fagioli, 'The Adaptation of the Dichromate Digestion Method for Total Mercury Determination by Cold Vapor Atomic Absorption Spectrometry to the Analysis of Soils, Sediments and Sludges', *Anal. Chim. Acta*, **298**, 363–374 (1994).
41. L. Magos, T.W. Clarkson, 'Atomic Absorption Determination of Total, Inorganic and Organic Mercury in Blood', *J. AOAC*, **55**, 966–971 (1972).
42. L. Magos, 'Mercury', in *Hazardous Metals in Human Toxicology*, ed. A. Vercruysse, Elsevier, Amsterdam, 1984.
43. B. Lind, E. Holmgren, L. Friberg, M. Vahter, 'Demethylation of Methylmercury During Inorganic Mercury Determination by the Magos Cold Vapor Atomic Absorption Technique', *Fresenius' J. Anal. Chem.*, **348**, 815–819 (1994).
44. R.S. Daniels, D.C. Wigfield, 'The Effect of Experimental Parameters on the Cold Vapor Mercury Atomic Absorption Determination: Speciation Analysis of Sulfhydryl-bound Mercury', *J. Anal. Toxic.*, **13**, 214–217 (1989).
45. K.W. Jackson, T.M. Mahood, 'Atomic Absorption, Atomic Emission and Flame Emission Spectrometry', *Anal. Chem.*, **66**, 252R–279R (1994).
46. F. Palmisano, P.G. Zambonin, N. Cardellicchio, 'Speciation and Simultaneous Determination of Mercury Species in Dolphin Liver by Liquid Chromatography with On-line Cold Vapor Atomic Absorption Spectrometry', *Fresenius' J. Anal. Chem.*, **346**, 648–652 (1993).
47. C.C. Wan, C.S. Chen, S.J. Jiang, 'Determination of Mercury Compounds in Water Samples by Liquid Chromatography Inductively Coupled Plasma Mass Spectrometry with an In-situ Nebulizer/Vapor Generator', *J. Anal. At. Spectrom.*, **12**, 683–688 (1997).
48. S.N. Wilie, D.C. Gregoire, R.E. Sturgeon, 'Determination of Inorganic and Total Mercury in Biological Tissues by Electrothermal Vaporization Inductively Coupled Mass Spectrometry', *Analyst*, **122**, 751–755 (1997).
49. L. Liang, N.S. Bloom, M. Horvat, 'Simultaneous Determination of Mercury Speciation in Biological Materials by GC/CVAFS after Ethylation and Room-temperature Pre-collection', *Clin. Chem.*, **40**, 602–606 (1994).
50. J.C. Van Loon, *Selected Methods of Trace Metal Analysis: Biological and Environmental Samples*, John Wiley & Sons, NY, 1985.
51. J.D. Ingle Jr, S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, NJ, 1988.
52. J.P. Schmit, M. Youla, Y. Gelinas, 'Multi-element Analysis of Biological Tissues by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **249**, 495–501 (1991).
53. K. Suetomi, H. Takahashi, T. Konishi, 'Fractional Determination of Ionizable and Stable Forms of Mercury in Animal Tissue Using Atomic Absorption Spectrometry', *Analyst*, **116**, 261–264 (1991).
54. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Determination of Mercury in Filtered Seawater by Flow Injection with On-line Oxidation and Atomic Fluorescence Spectrometric Detection', *J. Anal. At. Spectrom.*, **11**, 511–514 (1996).
55. G. Vermeir, C. Vandecasteel, R. Dams, 'Atomic Fluorescence Spectrometry Combined with Reduction Aeration for the Determination of Mercury in Biological Samples', *Anal. Chim. Acta*, **242**, 203–298 (1991).
56. S.K. Aggarwal, M. Kinter, D.A. Herold, 'Mercury Determination in Blood by Gas Chromatography–Mass Spectrometry', *Biol. Trace Elem. Res.*, **41**, 89–93 (1994).
57. C.J. Amariasiriwardena, B. Gercken, M.D. Argentine, R.M. Barnes, 'Semi-quantitative Analysis by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **5**, 457–462 (1990).
58. M.J. Campbell, G. Vermeir, R. Dams, 'Influence of Chemical Species on the Determination of Mercury in

a Biological Matrix (Cod Muscle) Using Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **7**, 617–620 (1992).

59. J. Yoshinaga, M. Morita, 'Determination of Mercury in Biological and Environmental Samples by Inductively Coupled Plasma Mass Spectrometry with the Isotope Dilution Technique', *J. Anal. At. Spectrom.*, **12**, 417–422 (1997).
60. A. Moauro, P.L. Carconi, 'A European Intercomparison of Vegetal Standard Reference Materials, Based on INAA and some Non-nuclear Spectrochemical Techniques', *J. Radioanal. Nucl. Chem.*, **151**(1), 149–157 (1991).
61. M. Bichler, 'Determination of Mercury in Mineral Samples by Employing a High Temperature Volatilization Technique and Activation Analysis', *J. Radioanal. Nucl. Chem.*, **154**(4), 255–263 (1991).
62. M.J. Bloxman, A. Gachanjz, P.J. Worsold, 'Determination of Mercury Species in Seawater by Liquid Chromatography with Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 145–148 (1996).
63. C.C. Wan, C.S. Chen, S.J. Jiang, 'Determination of Mercury in Compounds in Water Samples by Liquid Chromatography Inductively Coupled Plasma Mass Spectrometry with an In-situ Nebulizer Vapor Generator', *J. Anal. At. Spectrom.*, **12**, 683–688 (1997).
64. B. Holynska, B. Ostachowics, D. Wegrzynek, 'Simple Method of Determination of Copper, Mercury and Lead in Potable Water with Preliminary Pre-concentration by Total Reflection X-ray Fluorescence Spectrometry', *Spectrochim. Acta B*, **51**, 769–773 (1996).

Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis

Frank E. Smith
Laurentian University, Sudbury, Canada

Guohua Xiong
Jiujiang Teachers College, Jiujiang, PR of China

1 Introduction 1
2 Microwave-assisted Extraction 2
3 Focused Microwave-assisted Soxhlet Extraction 4
4 Microwave-assisted Headspace Analysis 4
5 Microwave-assisted Derivatization 4
6 Microwave-assisted Saponification and Microwave-assisted Decomposition 5
7 Conclusions 6
Acknowledgments 6
Abbreviations and Acronyms 6
Related Articles 7
References 7

This report describes innovative methods for sample pretreatment in organic environmental analysis that involve the use of microwave-assisted techniques. The methods described include: microwave-assisted extraction (MAE), focused microwave-assisted Soxhlet extraction (FMASE), microwave-assisted headspace analysis, microwave-assisted derivatization (MAD), microwave-assisted saponification (MAS) and microwave-assisted decomposition. A brief history of the development of each technique is provided. Reference is made to literature published up to the beginning of 1999.

1 INTRODUCTION

Modern analytical instruments can offer extremely low detection limits with high precision, resolution, reproducibility and speed of operation. But most of these methods require quite extensive sample pretreatment due to the great complexity of analytes and matrixes in environmental samples. This is usually the most time-consuming step in the whole analytical process. Therefore, the development of novel methods and techniques for sample pretreatment has attracted a great deal of attention from researchers. Since the mid-1980s a number of innovative pretreatment techniques have been introduced. These include: supercritical fluid extraction (SFE),[1] solid-phase extraction (SPE),[2] solid-phase microextraction (SPME),[3] the supported liquid membrane (SLM) technique for extraction,[4] accelerated solvent extraction (ASE)[5] and high-performance capillary electrophoresis (HPCE).[6] This report will outline other innovative methods for sample pretreatment in organic environmental analysis that involve the use of microwave-assisted techniques.

Interestingly enough, the first recorded uses of microwave ovens in chemical laboratories involved sample pretreatment in preparation for analysis. In 1974, Hesek and Wilson reported that a commercial microwave oven was employed to dry inorganic samples in process control,[7] then subsequently Abu-Samra et al. reported the use of microwave energy for the wet ashing of some biological samples prior to determination of metals content.[8] But the major breakthrough came with the development of microwave-assisted acidic pressure digestions[9,10] and microwave-assisted chemical synthesis[11,12] and the subsequent availability of cheap, microwave-transparent vessels able to withstand the rigorous conditions. Since then, this approach has been progressively established as a standard method for preparation of samples for elemental analysis.[13–15] Recently, microwave energy has been further employed to develop a four-stage sequential extraction method for metal fractionation in soil and sewage sludge samples.[16,17]

In the mid-1980s, Ganzler et al. pioneered the use of microwave energy for organic extractions in sample pretreatment.[18] They used microwave radiation from a conventional household microwave oven to enhance the extraction of various types of organic compounds from soil, seeds, foods and feeds. This MAE method gave higher recoveries and proved environmentally friendly in yielding considerable savings in time, energy and solvent compared with the conventional methods. Moreover, MAE had the advantage of being admirably suited for the isolation of thermally labile substances due to the dramatic savings in extraction time. The advantages of MAE have led to its widespread acceptance by practicing analysts in this field.

In addition, a series of new microwave-assisted techniques, such as MAS,[19] MAD[20] and microwave-assisted decomposition[21] have also been successfully developed in the last few years. These techniques can be separately employed for sample pretreatment in organic analysis or used in combination with MAE.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Details of the various microwave systems available and the theoretical basis of microwave heating are given elsewhere in this volume. Several useful reviews have appeared which put the various developments into perspective and demonstrate the wide applicability of microwave-assisted processes.[13,14,22–27]

2 MICROWAVE-ASSISTED EXTRACTION

Of all the microwave-assisted techniques for sample pretreatment for organic environmental analysis, MAE is the longest-established, best-studied and most widely-used. In performing MAE, a number of important parameters must be controlled, including sample size, solvent (extractant) type and volume, microwave irradiation time, magnetron power output and the temperature of the extraction system. The pressure in the sample vessel is also important when a closed-vessel system is used for MAE. All the parameters can be optimized by chemometrics and factorial design. Generally, the duration of an MAE cycle is less than 30 min, of which microwave irradiation time is about 10 min, and the quantity of solvent needed is less than 50% of that required for traditional Soxhlet extraction. The two principal microwave systems employed for MAE are the closed-vessel MAE system and the open focused microwave-assisted extraction (FMAE) system. For the FMAE both an open-vessel and a focused microwave unit are employed. There are some significant differences in experimental parameters and application areas between these two techniques.

MAE has been used in isolating a variety of analytes from a range of sample types. The increasing applications of this method are described in several review articles.[13,15,23,28–34] The original work by Ganzler et al. involved isolating various types of organic compounds and biologically active compounds from seeds, foods, feeds and rat feces.[18,35–37] This type of application of the process has been continued and expanded. In two patents held by Pare et al., the extraction of essential oils and other oils from biological materials such as plant and fish tissue,[38] and extraction of natural products from mint, garlic, etc.,[39] have been described. Prior to the issue of these patents, Craveiro et al.[40] also developed a method for the extraction of essential oils from fresh plant materials, which made use of MAE. In addition, the extraction of the essential oil components from rosemary leaves, including a study of the kinetics of the MAE process, has been reported by Chen and Spiro.[41] The extraction of antibodies from the surface of red blood cells was investigated by Torloni et al.[42] Banerjee et al.[43] have reported a microwave-based DNA extraction from paraffin-embedded tissue for polymerase chain reaction (PCR) amplification. Young[44] has used the technique for isolating the fungal metabolite ergosterol and total fatty acids. Leray et al.[45] reported an MAE method for lipid analysis in biological samples with an open-vessel microwave system. Akhiar and Croteau[46] described the MAE of salinomycin from finished layers ration. Carro et al.[47] studied the extraction of monoterphols from must samples, while Wang et al.[48] reported isolating Veltol® and Veltol-Plus® using MAE.

In addition to its applications in bioanalysis, food analysis and natural products extraction, MAE has also been used in isolating additives from polymers[34,49–51] and in the extraction of oligomers from poly(ethylene terephthalate) (PET).[52]

Probably the greatest number of reported applications of MAE involves sample preparation for organic environmental analysis. Examples include extracting polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), organochlorine pesticides (OCPs), organophosphorus pesticides (OPPs), phenolic compounds, chlorinated benzenes, petroleum hydrocarbons, triazines, imidazolinone herbicides, sulfonylurea herbicides, methylmercury and other organometallic compounds, from soils, sediments, sewage, air particles, plants, and tissue from marine mammals.[18,20,21,53–91] Polychlorinated dibenzo-*p*-dioxin (PCDD) and polychlorinated dibenzofuran (PCDF) were also extracted from solid samples such as fly ash, soil, wet sediment and sewage sludge.[58] The work of Lopez-Avila et al.[32,55,56,82,83] in evaluating the relative merits of MAE and the standard methods used by the United States Environmental Protection Agency (USEPA) for the extraction of the principal organic pollutants is worthy of note. Their research results demonstrate[55,56] that most of these compounds can be recovered in good yields from the soil/sediment samples that they studied using MAE. For example, they report recoveries of 80–120% for 79 of the 95 compounds listed in USEPA Method 8250,[92] 38 of the 45 OCPs listed in Method 8081,[93] and 34 of the 47 OPPs listed in Method 8141A.[94] For 15 compounds in a reference soil, the recoveries of 14 compounds by MAE were equal to or better than recoveries obtained by Soxhlet extraction (naphthalene being the exception). For selected OCPs, recoveries from spiked soil samples were at least 7% higher for MAE than for either Soxhlet or sonication extraction. Moreover, MAE also compared favourably with SFE.[83] In summary, it would appear that MAE can minimize pollution in the laboratory, improve target analyte recoveries and reduce sample preparation costs.

As well as being used for isolating organic analytes from diverse solid matrixes, MAE can also be applied to the

preparation of water samples. For example, microwave protocols for the extraction of PCBs and chlorinated benzenes from water samples have been developed by Onuska and Terry.[80,81] The use of aqueous extractants is especially suited to microwave-assisted techniques, and several new methods have been reported.[62–66] The MAE method has been used in combination with other extraction techniques such as SPE[85,86] and SPME.[48]

The methods which use MAE in combination with MAD, saponification, saponification–derivatization, decomposition or saponification–decomposition, will be discussed in the appropriate sections of this article.

A selection of cited applications of MAE for sample preparation in organic environmental analysis is given in Table 1. Although several nonpolar or weakly polar solvents are listed as extraction media, it is essential that some moisture or polar solvent be present in the system to enable the absorption of the microwave energy.

Table 1 MAE for sample preparation in organic environmental analysis

Sample	Analyte	Extractant (solvent)	Ref.
Soil	OPPs (bromophos, parathion)	Methanol	20
Sediment	OCPs	Isooctane–acetonitrile, 1 : 1	53
Sediment	Trialkyl and triaryl phosphates	Hexane/DCM	54
Soil, sediment	PAHs, base/neutral compounds, phenols, OCPs	Acetone–hexane, 1 : 1 (v/v)	55
Topsoil	187 compounds of interest to USEPA, 5 Aroclors (PCBs)	Acetone–hexane, 1 : 1 (v/v)	56
Soil	PAHs	Acetone, acetone/hexane	57
Solid samples (sediment/soil/fly ash, etc.)	PCDD, PCDF	Toluene	58
Marine sediment	OCPs	Acetone–hexane, 1 : 1 (v/v)	59
Marine sediment	PAHs	Acetone–hexane, 1 : 1 (v/v)	60
Ambient-air polyurethane foam	PAHs and their surrogates	Cyclohexane	61
Soil	Imidazolinone herbicides	0.1 M NH_4OAc/NH_4OH, pH 10.0	62
Plant tissue	Imidazolinone herbicides & their metabolites	H_2O	63
Soil	Atrazine and its degradates	0.35 M HCl	64
Soil	Atrazine	H_2O	65
Soil	Atrazine, simatrine, prometryne	H_2O	66
Soil, sediment, mussels	PCBs	Methanolic 1 M KOH solution	21
Marine mammals	Organochlorine compounds	Hexane	67
Soil	Phenol and methylphenol isomers	Acetone/hexane	68
Soil	Phenolic compounds	Hexane/acetic anhydride/pyridine In situ MAE/MAD	69
Sediment	Hydrocarbons, pesticides	Toluene–water, 10 : 1 (v/v)	70
Plants	PAHs	Hexane–acetone, 3 : 2 (v/v)	71
Soil	Phenolic compounds	Methanol–water, 4 : 1, with 2% TEA	72
Soil	Sulfonylurea herbicides	DCM–methanol, 9 : 1	73
Sediment	OCPs	THF	74
Sediment	PAHs	DCM	75
Soil, sewage	PAHs, PCBs	Hexane–acetone, 3 : 1	76
Air particles	PAHs	Acetone–hexane, 1 : 1 (v/v)	77
Soil, sewage	PCBs	Hexane–acetone, 3 : 1	78
Water, olives, sunflower, etc.	Edible oils, mineral oils, petroleum, gasoline, etc.	CCl_4	79
Water	PCBs	Isooctane (NaCl was added)	80
Water	Chlorinated benzenes	NaCl was added, helium used to purge the analytes into hexane	81
Sediment	Methylmercury	Toluene (6 M HCl was added)	88
Sediment	Methylmercury	2 M HNO_3	89
Mussels	Arsenic species	Methanol/water	90
Sediment	Organotin compounds	Methanolic 0.5 M ethanoic acid	20
Biomaterial	Organotin compounds	Acetic acid	91

DCM, dichloromethane.

3 FOCUSED MICROWAVE-ASSISTED SOXHLET EXTRACTION

One of the most recent applications of microwave energy for sample pretreatment in organic environmental analysis is FMASE, which has recently been described by Garcia-Ayuso et al.[95] The device itself consists of two main parts: the microwave irradiation zone, which includes a modified Prolabo Microdigest 301 (200 W maximum power) microwave source along with a modified Soxhlet extractor, and a distillation flask which is connected to a cooling unit. Two peristaltic pumps are used for solvent aspiration and siphoning in addition to the Soxhlet action. A special thermometer is used for monitoring the extraction temperature; three monitors control the microwave unit, pumps, and the thermometer. An electrical mantle is used as a heating source for the distillation flask.

The FMASE apparatus described above has been used to extract three different kinds of organic contaminants, namely alkanes, PAHs and herbicides, from clay soil. DCM was used as the extractant. For a 7-g sample, ultrapure water (1.0 mL or 1.5 mL) was added to the cartridge and the sample was allowed to soak for 10–15 min (water mainly acts to enhance the absorption of microwave energy). Then, the cartridge thimble was covered with a loose wad of glass wool and inserted into the extraction vessel, which was placed in the microwave irradiation zone. The extraction was performed as follows:

1. the extraction vessel was filled with 50 mL of fresh extractant delivered by pump 1;
2. the sample was subjected to microwave irradiation for 15 s at 50% of maximum power output;
3. the extraction vessel was drained by pump 2.

This extraction cycle was repeated a number of times depending on the target analytes. The duration of each cycle was 6 min, so a whole extraction program took 50–60 min.

The results of the study demonstrate that:

- FMASE provides efficiencies similar to those obtained by the conventional Soxhlet method with extraction times at least eight times shorter (50–60 min versus 8 h).
- FMASE can be used for almost all the applications of a conventional Soxhlet extraction under similar operating conditions.
- FMASE can be used for treating wet samples whereas the conventional Soxhlet method is usually suitable for dry samples only.

4 MICROWAVE-ASSISTED HEADSPACE ANALYSIS

The isolation and analysis of the volatile organic compounds (VOCs) in water samples may be carried out by transferring the volatile material in the headspace above the liquid surface to a gas chromatography (GC) device. Pare et al. have made use of microwave heating to improve the conventional headspace method.[96] The standard headspace sampler is still used to sample the vapor above the water and then to transfer the analytes into a GC device but the different method of heating gives very significant improvements in isolating efficiency.

The VOCs, namely benzene, methylbenzene, ethylbenzene, chlorobenzene, *o*-, *m*-, and *p*-xylenes, and 1,2-, 1,3-, and 1,4-dichlorobenzene, were isolated from both spiked and actual samples at parts per trillion (ppt) to parts per million (ppm) levels. Compared with the conventional headspace method, this new technique offers enhanced sensitivity as evidenced by a 35% increase in signal strength and a dramatic reduction in the time required for the establishment of equilibrium (from 30 min to 1 min).

5 MICROWAVE-ASSISTED DERIVATIZATION

Since the first studies by Smith et al.,[11,12] the use of microwave energy for the acceleration of organic reactions has increased enormously.[22,97] Recently, microwave-assisted esterification was used for the gas chromatography/mass spectrometry (GC/MS) determination of fatty acids in food oils.[98] The esterification of the fatty acids was carried out in methanolic 0.4 M KOH solution in a closed vessel with 30 s of microwave irradiation (600 W). Considerable savings of time and solvent were obtained with this method as compared with the conventional derivatization procedure.

Although derivatization reactions were among the first studied by the early researchers in this field,[99] the use of MAD for sample pretreatment in organic environmental analysis first appeared in a report by Lalere et al.[20] For the speciation analysis of organotin compounds in sediments by capillary GC with flame photometric detection (FPD), microwave-assisted ethylation of the organotin compounds was combined with MAE of the ethylated products. First, MAE of the target compounds from a sediment sample was performed with methanolic 0.5 M ethanoic acid as the extractant. Next, the separated supernatant solution was placed in a 50-mL reaction vessel and the derivatization reagent (sodium tetraethylborate, $NaBEt_4$), extractant (isooctane) and a certain volume of artificial seawater were added. Then the mixture was

exposed to a focused microwave field. The results showed that at volumes lower than 20 mL, quantitative recoveries were obtained with irradiation for 3 min at a microwave power of 70 W for all the compounds except monobutyltin (MBT), for which the recovery was close to 100% following a microwave irradiation for 10 min. The conventional process is also a multistage operation. The ethylation involves 30 min of magnetic stirring followed by extraction of the ethylation products. For the simultaneous extraction–derivatization reaction, the kinetics of the two processes were studied. Ethylation is a nucleophilic substitution, and the extraction is transfer of the derivatized analytes to the organic phase. It was demonstrated that the ethylation was apparently faster than the transfer process, and could be completed within 5 min if the reaction product was removed from the reaction medium or the phases were homogenized. Therefore, extraction does not affect derivatization in this case.

For the speciation analysis of organotin compounds in biomaterials, Pereiro et al.[102] developed an interesting microwave-assisted method by which sample decomposition, extraction, and derivatization of organotin compounds could be carried out simultaneously in a single reactor. Acetic acid was selected as the reaction medium, $NaBEt_4$ as the derivatization reagent, and nonane as the extractant. The reactions were conducted in a focused microwave field with 40 W power for 3 min to reach a temperature of 130 °C. It was found that ethyl derivatives of mono-, di-, and triorganotins were quantitatively detected in the supernatant organic phase. Furthermore, the new method was validated by analyzing a certified reference material (fish tissue, NIES11). The kinetics of simultaneous extraction–derivatization reaction were also investigated. Thus the hours-long multistep sample preparation procedure may be replaced by a one-step integrated method taking only 3 min.

Rodriguez et al.[103] also reported a speciation analysis of organotin compounds in marine biomaterials via microwave-assisted sample preparation using a closed-vessel microwave unit. Tetramethylammonium hydroxide was used as the reaction medium, and the pH of the reaction system was adjusted to 5. $NaBEt_4$ was used as the derivatization reagent, and isooctane as the extractant. This caused the simultaneous alkaline decomposition of the sample, the derivatization of the analytes (organotins) and extraction of the derivatization products. This study also included an investigation of the stability of organotin compounds on exposure to microwaves.

Llompart et al.[104] described an in situ MAE and derivatization for sample preparation to determine phenol and methylphenol isomers in soils. The sample, to which small quantities of acetic anhydride and pyridine had been added in order to catalyse the acetylation process, was immersed in hexane in a closed vessel. Microwave irradiation parameters and other conditions concerning the method were optimized using a factorial design approach. Compared with ultrasonic extraction, the microwave-assisted method gave superior recoveries (about two-fold) and greater precision, in addition to shorter extraction times. The analytes extracted in the hexane phase could be directly analyzed on a GC/MS system without preliminary clean-up or concentration steps.

6 MICROWAVE-ASSISTED SAPONIFICATION AND MICROWAVE-ASSISTED DECOMPOSITION

There have been many applications of MAS or hydrolysis. Microwave energy was used by Chiou and Wang[105] for peptide and protein hydrolysis without significant degradation of the analytes. Tatar et al.[106] also described a microwave hydrolysis system to determine the amino acid composition of protein. Arai and Kakishita[107] reported the manufacture of soap from olive oil and the saponification of $PhCO_2Me$ by microwave heating. Rapid acid hydrolyses of albumin[108] and benzodiazepines[109] were investigated by Morales-Rubio et al. Taketomi et al.[110] established a microwave-mediated saponification of galactosylceramide (cerobroside) and galactosylceramide I^3-sulfate (sulfide) to prepare their lyso-compounds. Moreover, Perez and Haswell[111] developed a flow injection analysis (FIA) method with online MAS of vitamin A to retinol; Young[44] reported a simultaneous MAE and saponification to isolate the fungal metabolite ergosterol from various samples and the fatty acids from spores for their determination. Recently, Pineiro-Avila et al.[19] established a MAS of animal grease for cholesterol determination.

MAS also has applications in organic environmental analysis. Tseng et al.[89] developed a method to detect inorganic mercury and methylmercury in biological samples by MAS coupled with hydride generation, cryofocusing and atomic absorption spectroscopy (AAS) determination. Saponification was performed in tetramethylammonium hydroxide or methanolic KOH solution with microwave irradiation of 40–60 W for only 2–4 min, which was much faster than previously reported methods. This new method was validated by the analysis of four biological certified reference materials, CRM 463, DORM-1, TORT-1 and Tuna fish 2. Recently, Xiong et al.[21] reported a MAS of mussels for PCB determination using a closed-vessel microwave system. A 15-g sample in 30 mL of methanolic KOH solution could be completely saponified in less than 2 min without affecting the PCB analytes. For the speciation of

butyltin compounds in biomaterials, Szpunar et al.[112] used saponification of the sample in tetramethylammonium hydroxide in a low-power microwave field. The biomaterial could be dissolved in less than 5 min, leaving the organotin moiety intact. In addition, as mentioned in section 5, MAS was also studied by Rodriguez et al.[103] for the speciation of butyltin compounds in marine biomaterials. There have been several reports of the use of similar methods for the speciation of other organotin compounds.[113–115]

It is important to distinguish between microwave-assisted decomposition and microwave-assisted digestion (wet ashing) for elemental analysis. In this article, microwave-assisted decomposition for organic environmental analysis refers to the selective decomposition of the sample matrix leaving the target analytes unchanged, whereas in microwave-assisted digestion (wet ashing) the whole sample is broken down.

When gas chromatography/electron capture detection (GC/ECD) is employed for determining PCBs in real-world samples, the co-existing OCPs, such as benzene hexachloride (BHC) and DDT (1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane) often interfere with the determination of the target analytes. Xiong et al.[21,116,117] have developed a microwave-assisted decomposition–extraction method for soil/sediment sample preparation and a MAS–decomposition–extraction method which addresses this problem. Methanolic 1 M KOH solution was selected as the reaction medium and the reaction conditions are optimized so that BHC, *o*,*p*-DDT, *p*,*p*-DDT and *p*,*p*-DDD (1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethane) are completely decomposed after microwave heating for 2 min. Although dieldrin, aldrin and *p*,*p*-DDE (1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene) are only partly degraded under these conditions, most of their decomposition products and dieldrin itself can be eliminated with an additional rinse using concentrated H_2SO_4.

In some publications,[20,102] microwave-assisted acidic decomposition of biomaterial samples for the speciation of organometallic compounds is reported. Generally, acetic acid is chosen because, compared with hydrochloric acid, it has a higher boiling point, lower acidity (and thus better compatibility with derivatization–extraction) and neutral behavior towards the carbon–metal bond.[102] According to these reports, microwave-assisted acidic decomposition, actually acidic hydrolysis, may be carried out rapidly and to completion.

7 CONCLUSIONS

For sample preparation in organic environmental analysis, a number of published reports have demonstrated that MAE possesses significant advantages over conventional methods such as Soxhlet and sonication extractions. These advantages include increased recoveries and great savings of time, energy and the quantity of organic solvents consumed. Compared with SFE, MAE generally also has a higher efficiency and wider field of application. Similarly, MAD, MAS, microwave-assisted decomposition, FMASE and microwave-assisted headspace analysis all show significant benefits over the earlier methods. Therefore, it can be predicted that in the near future, more and more laboratories will replace the conventional methods with these newly developed techniques, and they will become routine tools for sample preparation in organic environmental analysis.

However, there is still a need for further developments. In particular:

- microwave instruments, especially the closed vessel systems, need to be more adaptable to automation, making on-line preparation/determination for organic environmental analysis more readily achievable
- to minimize the use of organic solvents, further development of cheap, less toxic aqueous systems for microwave-assisted processes is required
- simultaneous application of several microwave-assisted techniques may further increase the selectivity for isolating target analytes
- the field of application for microwave-assisted techniques for sample preparation in organic environmental analysis may be greatly widened by using them in conjunction with other methods of sample preparation.

ACKNOWLEDGMENTS

Guohua Xiong would like to thank Professor Zhanxia Zhang for providing guidance while he was studying for his Ph.D. at Zhongshen (Sun Yet-Sen) University, Guangzhou PR of China. His research related to this article was supported by the National Natural Science Fund of China (No. 29967001).

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
ASE	Accelerated Solvent Extraction
BHC	Benzene Hexachloride
DCM	Dichloromethane
DDD	1,1-Dichloro-2,2-bis(*p*-chlorophenyl)-ethane

DDE	1,1-Dichloro-2,2-bis(*p*-chlorophenyl)-ethylene
DDT	1,1,1-Trichloro-2,2-bis(*p*-chlorophenyl)-ethane
FIA	Flow Injection Analysis
FMAE	Focused Microwave-assisted Extraction
FMASE	Focused Microwave-assisted Soxhlet Extraction
FPD	Flame Photometric Detection
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/MS	Gas Chromatography/Mass Spectrometry
HPCE	High-performance Capillary Electrophoresis
MAD	Microwave-assisted Derivatization
MAE	Microwave-assisted Extraction
MAS	Microwave-assisted Saponification
MBT	Monobutyltin
OCPs	Organochlorine Pesticides
OPPs	Organophosphorus Pesticides
PAHs	Polycyclic Aromatic Hydrocarbons
PCB	Polychlorinated Biphenyl
PCDD	Polychlorinated Dibenzo-*p*-dioxin
PCDF	Polychlorinated Dibenzofuran
PCR	Polymerase Chain Reaction
PET	Poly(ethylene terephthalate)
ppm	Parts Per Million
ppt	Parts Per Trillion
SFE	Supercritical Fluid Extraction
SLM	Supported Liquid Membrane
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
USEPA	United States Environmental Protection Agency
VOCs	Volatile Organic Compounds

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Detection and Screening of Chemicals Related to the Chemical Weapons Convention • Verification of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment • Flow-injection Techniques in Environmental Analysis • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Organic Analysis in Environmental Samples by Electrochemical Methods • Soxhlet and Ultrasonic Extraction of Organics in Solids • Waste Extraction Procedures

General Articles **(Volume 15)**
Microwave Techniques

REFERENCES

1. H. Engelhalt, A. Gross, 'Supercritical Fluid Extraction and Chromatography: Potential and Limitations', *Trends Anal. Chem.*, **10**, 64–70 (1991).
2. B.A. Bidlingmeyer, 'Guidelines for Proper Usage of Solid-phase Extraction', *LC Liq. Chromatogr. HPLC Mag.*, **2**, 578–580 (1984).
3. C.L. Arthur, J. Pawliszyn, 'Solid Phase Microextraction with Thermal Desorption Using Fused Silica Optical Fibers', *Anal. Chem.*, **62**, 2145–2148 (1990).
4. G. Audunsson, 'Aqueous/Aqueous Extraction by Means of a Liquid Membrane for Sample Cleanup and Preconcentration of Amines in a Flow System', *Anal. Chem.*, **58**, 2714–2723 (1986).
5. B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: A Technique for Sample Preparation', *Anal. Chem.*, **68**, 1033–1039 (1996).
6. R. Kuhn, F. Stocklin, F. Erni, 'Chiral Separations by Host–Guest Complexation with Cyclodextrin and Crown Ether in Capillary Zone Electrophoresis', *Chromatogr.*, **33**, 32–36 (1992).
7. J.A. Hesek, R.C. Wilson, 'Use of Microwave Oven in In-process Control', *Anal. Chem.*, **46**, 1160–1164 (1974).
8. A. Abu-Samra, J.S. Morris, S.R. Koirtyohann, 'Wet Ashing of Some Biological Samples in a Microwave Oven', *Anal. Chem.*, **47**, 1475–1477 (1975).
9. F.E. Smith, B. Cousins, J. Bozic, W. Flora, 'The Acid Dissolution of Sulfide Mineral Samples Under Pressure in Microwave Ovens', *Anal. Chim. Acta*, **177**, 243–245 (1985).
10. J. Huang, D. Goltz, F.E. Smith, 'A Microwave Dissolution Technique for the Determination of Arsenic in Soils', *Talanta*, **35**, 907–908 (1988).
11. R. Gedye, F.E. Smith, K. Westaway, H. Ali, L. Baldisera, L. Laberge, J. Roussell, *Tetrahedron Letters*, **27**, 279–282 (1986).

12. R.N. Gedye, F.E. Smith, K.W. Westaway, 'The Rapid Synthesis of Organic Compounds in Microwave Ovens', *Can. J. Chem.*, **66**, 17–26 (1988).
13. A. Zlotorzynski, 'The Application of Microwave Radiation to Analytical and Environmental Chemistry', *Crit. Rev. Anal. Chem.*, **25**, 43–76 (1995).
14. F.E. Smith, E.A. Arsenault, 'Microwave-assisted Sample Preparation in Analytical Chemistry', *Talanta*, **43**, 1207–1268 (1996).
15. H. Kingston, L.B. Jassie, *Introduction to Microwave Sample Preparation: Theory and Practice*, ACS, Washington DC, 1988.
16. E. Campos, E. Barahona, M. Lachica, M.D. Mingorance, 'A Study of the Analytical Parameters Important for the Sequential Extraction Producedure Using Microwave Heating for Pb, Zn and Cu in Calcareous Soils', *Anal. Chim. Acta*, **369**, 235–243 (1998).
17. B. Perez-Cid, I. Lavilla, C. Bendicho, 'Application of Microwave Extraction for Partitioning of Heavy Metals in Sewage Sludge', *Anal. Chim. Acta*, **378**, 201–210 (1999).
18. K. Ganzler, A. Salgo, K. Valko, 'Microwave Extraction. A Novel Sample Preparation Method for Chromatography', *J. Chromatogr.*, **371**, 299–306 (1986).
19. G. Pineiro-Avila, A. Salvador, M. de la Guardia, 'Microwave-assisted Saponification of Animal Greases for Cholesterol Determination', *Anal. Chim. Acta*, **371**, 297–303 (1998).
20. B. Lalere, J. Szpunar, H. Budzinski, P. Garrigues, O.F.X. Donard, 'Speciation Analysis for Organotin Compounds in Sediments by Capillary Gas Chromatography with Flame Photometric Detection After Microwave-assisted Acid Leaching', *Analyst*, **120**, 2665–2673 (1995).
21. G.-H. Xiong, X.-Q. He, M.-Q. Zhao, B.-Y. Tang, Z.-X. Zhang, 'Gas Chromatographic Determination of PCBs in Mussel Using Microwave-assisted Saponification Coupled with Alkaline Degradation for Sample Preparation', *The Proceedings of the Third International Symposium of World Chinese Scholars on Analytical Chemistry*, Hong Kong Baptist University, Hong Kong, 6-018, 312 (1998).
22. D.M.P. Mingos, D.R. Baghurst, 'Applications of Microwave Dielectric Heating Effects to Synthetic Problems in Chemistry', *Chem. Soc. Rev.*, **20**, 1–17 (1991).
23. J.R.J. Pare, J.M.R. Belanger, S.S. Stafford, 'Microwave-assisted Process (MAP): A New Tool for the Analytical Laboratory', *Trends Anal. Chem.*, **13**, 176–184 (1994).
24. M. Burguera, J.L. Burguera, 'Microwave Sample Pretreatment in Analytical Systems. A Review', *Quimica Analytica*, **15**, 112–122 (1996).
25. C.R. Strauss, R.W. Trainor, 'Developments in Microwave-assisted Organic Chemistry', *Aust. J. Chem.*, **48**, 1665–1692 (1995).
26. J.R. Dean, 'Microwave-enhanced Solvent Extraction', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, Inc., 2797–2821, 1998.
27. H.M. Kingston, S.J. Haswell (eds.), *Microwave-enhanced Chemistry – Fundamentals, Sample Preparation and Applications*, American Chemical Society, Washington DC, 1997.
28. B.W. Renoe, 'Microwave-assisted Extraction for Environmental Analyses', *Am. Environ. Lab.*, **7**, 26–27 (1995).
29. T.S. Reighard, S.V. Olesik, 'Bridging the Gap Between Supercritical Fluid Extraction and Liquid Extraction Techniques: Alternative Approaches to the Extraction of Solid and Liquid Environmental Matrices', *Crit. Rev. Anal. Chem.*, **26**, 61–99 (1996).
30. J. Szpunar, V.O. Schmitt, O.F.X. Donard, R. Lobinski, 'Low-power Focused Microwave Technology as a New Tool for Rapid Preparation of Solid Samples for Speciation Analysis', *Trends Anal. Chem.*, **15**, 181–187 (1996).
31. N.M. Kuz'min, I.V. Kubrakova, 'Microwave Sample Preparation', *J. Anal. Chem. (Transl. of Zh. Anal. Khim.)*, **51**, 38–41 (1996) (*Chem. Abstr.*, **124**, abs. no. 192439y (1996)).
32. V. Lopez-Avila, J. Benedicto, 'Microwave-assisted Extraction Combined with Gas Chromatography and Enzyme-linked Immunosorbent Assay', *Trends Anal. Chem.*, **15**, 334–341 (1996).
33. O. Zuloaga, N. Etxebarria, L.A. Fernandez, J.M. Madariaga, 'Comparison of Accelerated Solvent Extraction with Microwave-assisted Extraction and Soxhlet for the Extraction of Chlorinated Biphenyls in Soil Samples', *Trends Anal. Chem.*, **17**, 642–647 (1998).
34. H.J. Vandenburg, A.A. Clifford, K.D. Bartle, J. Carroll, I.D. Newton, L.M. Garden, J.R. Dean, C.T. Costley, 'Analytical Extraction of Additives from Polymers', *Analyst*, **122**, 101R–115R (1997).
35. K. Ganzler, J. Bati, K. Valko, 'A New Method for the Extraction and High-performance Liquid Chromatographic Determination of Vicine and Convicine in Faba beans', *Symp. Biol. Hung.*, **31**, 435–442 (1986).
36. K. Ganzler, A. Salgo, 'Microwave-extraction–A New Method Superseding Traditional Soxhlet Extraction', *Z. Lebensm.-Unters. Forsch.*, **184**, 274–276 (1987).
37. K. Ganzler, I. Szinai, A. Salgo, 'Effective Sample Preparation Method for Extracting Biologically Active Compounds from Different Matrices by a Microwave Technique', *J. Chromatogr.*, **520**, 257–262 (1990).
38. J.R.J. Pare, 'Microwave Extraction of Volatile Oils and Apparatus Therefore, Eur. Pat. Appl. EP 485, 668, 1992', *Chem. Abstr.*, **117**, 157431y (1992).
39. J.R.J. Pare, M. Sigouin, J. Lapointe, 'Microwave-assisted Natural Products Extraction', U.S. Patent 5,002,784, 1991.
40. A.A. Craveiro, F.J.A. Matos, J.W. Alencar, 'Microwave Oven Extraction of an Essential Oil', *Flavour Fragrance J.*, **4**, 43–44 (1989).
41. S.S. Chen, M. Spiro, 'Kinetics of Microwave Extraction of Rosemary Leaves in Hexane, Ethanol and a Hexane + Ethanol Mixture', *Flavour Fragrance J.*, **10**, 101–112 (1995).

42. A.S. Torloni, J.S. McCullough, M.E. Brecher, L.J. Tribble, M.G. Hill, 'Microwave Elution of Red Cell Antibodies', *28th International Microwave Power Institute Symposium*, 118, 1993.
43. S.K. Banerjee, W.F. Makdisi, A.P. Weston, S.M. Mitchell, D.R. Campbell, 'Microwave-based DNA Extraction from Paraffin-embedded Tissue for PCR Amplification', *Biotechniques*, **18**, 768–773 (1995).
44. J.C. Young, 'Microwave-assisted Extraction of the Fungal Metabolite Ergosterol and Total Fatty Acids', *J. Agric. Food. Chem.*, **43**, 2904–2910 (1995).
45. C. Leray, T. Grcic, G. Gutbier, M. Bnouham, 'Microwave Oven Extraction Procedure for Lipid Analysis in Biological Samples', *Analusis*, **23**, 65–67 (1995).
46. M.H. Akhtar, L.G. Croteau, 'Extraction of Salinomycin from Finished Layers Ration by Microwave Solvent Extraction Followed by Liquid Chromatography', *Analyst*, **121**, 803–806 (1996).
47. N. Carro, C.M. Garcia, R. Cela, 'Microwave-assisted Extraction of Monoterpenols in Must Samples', *Analyst*, **122**, 325–329 (1997).
48. Y. Wang, M. Bonilla, H.M. McNair, M. Khaled, 'Solid Phase Microextraction Associated with Microwave Assisted Extraction of Food Products', *J. High Resolut. Chromatogr.*, **20**, 213–216 (1997).
49. W. Freitag, O. John, 'Fast Separation of Stabilizers from Polylefins by Microwave Heating', *Angew. Makromol. Chem.*, **175**, 181–185 (1990).
50. R.C. Nielson, 'Extraction and Quantitation of Polyolefin Additives', *J. Liq. Chromatogr.*, **14**, 503–519 (1991).
51. H.J. Vandenburg, A.A. Clifford, K.D. Bartle, J. Carroll, I.D. Newton, 'Comparison of Pressurised Fluid Extraction and Microwave Assisted Extraction with Atmospheric Pressure Methods for Extraction of Additives from Polypropylene', *Analyst*, **124**, 397–400 (1999).
52. C.T. Costley, J.R. Dean, I. Newton, J. Carroll, L.M. Garden, 'Extraction of Oligomers from Poly(ethylene terephthalate) by Microwave-assisted Extraction', *Anal. Commun.*, **34**, 89–91 (1997).
53. F.I. Onuska, K.A. Terry, 'Extraction of Pesticides from Sediments Using a Microwave Technique', *Chromatogr.*, **36**, 191–194 (1993).
54. H. De Geus, B.N. Zegers, H. Lingenman, U.A.Th. Brinkman, 'Determination of Trialkyl and Triaryl Phosphates in Sediment Using Microwave Extraction and Packed-capillary Supercritical Fluid Chromatography', *Intern. J. Environ. Anal. Chem.*, **56**, 119–132 (1994).
55. V. Lopez-Avila, R. Young, W.F. Beckert, 'Microwave-assisted Extraction of Organic Compounds from Standard Reference Soils and Sediments', *Anal. Chem.*, **66**, 1097–1106 (1994).
56. V. Lopez-Avila, R. Young, J. Benedico, P. Ho, R. Kim, W.F. Beckert, 'Extraction of Organic Pollutants from Solid Samples Using Microwave Energy', *Anal. Chem.*, **67**, 2096–2102 (1995).
57. I.J. Barnabas, J.R. Dean, I.A. Fowlis, S.P. Owen, 'Extraction of Polycyclic Aromatic Hydrocarbons from Highly Contaminated Soils Using Microwave Energy', *Analyst*, **120**, 1897–1904 (1995).
58. M. Schlabach, A. Biseth, H. Gundersen, M. Oehme, 'Microwave Assisted Extraction of Solid Samples for PCDD/PCDF Analysis', *Organohalogen Compd.*, **23**, 105–108 (1995).
59. K.K. Chee, M.K. Wong, H.K. Lee, 'Determination of Organochlorine Pesticides in Water by Membranous Solid-phase Extraction, and in Sediment by Microwave-assisted Solvent Extraction with Gas Chromatography and Electron-capture and Mass Spectrometric Detection', *J. Chromatogr. A*, **736**, 211–218 (1996).
60. K.K. Chee, M.K. Wong, H.K. Lee, 'Optimization of Microwave-assisted Solvent Extraction of Polycyclic Aromatic Hydrocarbons in Marine Sediments Using a Microwave Extraction System with High-performance Liquid Chromatography-fluorescence Detection and Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **723**, 259–271 (1996).
61. R.C. Lao, Y.Y. Shu, J. Holmes, C. Chiu, 'Environmental Sample Cleaning and Extraction Procedures by Microwave-assisted Process (MAP) Technology', *Microchem. J.*, **53**, 99–108 (1996).
62. S.T. Stout, A.R. daCunha, D.G. Allardice, 'Microwave-assisted Extraction Coupled with Gas Chromatography/Electron Capture Negative Chemical Ionization Mass Spectrometry for the Simplified Determination of Imidazolinone Herbicides in Soil at the ppb Level', *Anal. Chem.*, **68**, 653–658 (1996).
63. S.J. Stout, A.R. daCunha, G.L. Picard, M.M. Safarpour, 'Microwave-assisted Extraction Coupled with Liquid Chromatography/Electrospray Ionization Mass Spectrometry for the Simplified Determination of Imidazolinone Herbicides and Their Metabolites in Plant Tissue', *J. Agric. Food Chem.*, **44**, 3548–3553 (1996).
64. T.R. Steinheimer, 'HPLC Determination of Atrazine and Principal Degradates in Agricultural Soils and Associated Surface and Ground Water', *J. Agric. Food Chem.*, **41**, 588–589 (1993).
65. G.-H. Xiong, J.-M. Liang, S.-C. Zou, Z.-X. Zhang, 'Microwave-assisted Extraction of Atrazine from Soil Followed by Rapid Detection Using Commercial ELISA Kit', *Anal. Chim. Acta*, **371**, 97–103 (1998).
66. G.-H. Xiong, B.-Y. Tang, X.-Q. He, M.-Q. Zhao, Z.-P. Zhang, Z.-X. Zhang, 'Comparison of Microwave-assisted Extraction of Triazines from Soils Using Water and Organic Solvents as the Extractants', *Talanta*, **48**, 333–339 (1999).
67. K. Hummert, W. Vetter, B. Luckas, 'Fast and Effective Sample Preparation for Determination of Organochlorine Compounds in Fatty Tissue of Marine Mammals Using Microwave Extraction', *Chromatogr.*, **42**, 300–304 (1996).

68. M.P. Llompart, R.A. Lorenzo, R. Cela, J.R.J. Pare, 'Optimization of a Microwave-assisted Extraction Method for Phenol and Methylphenol Isomers in Soil Samples Using a Central Composite Design', *Analyst*, **122**, 133–137 (1997).
69. M.P. Llompart, R.A. Lorenzo, R. Cela, K. Li, J.M.R. Belanger, J.R.J. Pare, 'Evaluation of Supercritical Fluid Extraction, Microwave-assisted Extraction and Sonication in the Determination of Some Phenolic Compounds from Various Soil Matrixes', *J. Chromatogr. A*, **774**, 243–251 (1997).
70. A. Pastor, E. Vazquez, R. Ciscar, M. de la Guardia, 'Efficiency of the Microwave-assisted Extraction of Hydrocarbons and Pesticides from Sediments', *Anal. Chim. Acta*, **344**, 241–249 (1997).
71. M. Tomaniova, J. Hajslova, J. Pavelka, Jr, V. Kocourek, K. Holadova, I. Klimova, 'Microwave-assisted Solvent Extraction – A New Method for Isolation of Polynuclear Aromatic Hydrocarbons from Plants', *J. Chromatogr. A*, **827**, 21–29 (1998).
72. M.C. Alonso, D. Puig, I. Silgoner, M. Grasserbauer, D. Barcelo, 'Determination of Priority Phenolic Compounds in Soil Samples by Various Extraction Methods Followed by Liquid Chromatography–Atmospheric Pressure Chemical Ionisation Mass Spectrometry', *J. Chromatogr. A*, **823**, 231–239 (1998).
73. N. Font, F. Hernandez, E.A. Hogendoorn, R.A. Baumann, P. van Zoonen, 'Microwave-assisted Solvent Extraction and Reversed-phase Liquid Chromatography–UV Detection for Screening Soils for Sulfonylurea', *J. Chromatogr. A*, **798**, 179–186 (1998).
74. I. Silgoner, R. Krska, E. Lombas, O. Gans, E. Rosenberg, M. Grasserbauer, 'Microwave-assisted Extraction of Organochlorine Pesticides from Sediments and its Application to Contaminated Sediment Samples', *Fresenius' J. Anal. Chem.*, **362**, 120–124 (1998).
75. M. Letellier, H. Budzinski, 'Influence of Sediment Grain Size on the Efficiency of Focused Microwave Extraction of Polycyclic Aromatic Hydrocarbons', *Analyst*, **124**, 5–14 (1999).
76. B. Enders, G. Schwedt, 'Microwave-assisted Extraction for the Analysis of PCB and PAH from Soil and Sewage Sludge', *GIT Fachz. Lab.*, **40**, 172, 174–176 (1996) (in German); (*Chem. Abstr.*, **124**, abs. no. 305998a (1996)).
77. K.K. Chee, M.K. Wong, H.K. Lee, 'Microwave-assisted Solvent Extraction of Air Particulates for the Determination of PAHs', *Environ. Monit. Assess.*, **44**, 391–403 (1997).
78. B. Enders, G. Schwedt, 'Supercritical Fluid Extraction, Microwave-assisted Extraction, and Soxhlet Extraction for the Analysis of PCBs in Soil and Sewage Sludge', *J. Prakt. Chem./Chem.-Ztg.*, **339**, 250–255 (1997) (in German); *Chem. Abstr.*, **127**, abs. no. 8625u (1997).
79. Y. Daghbouche, S. Garrigues, M. Delaguardia, 'Flow Injection–Fourier Transform Infrared Spectrometric Determination of Oil and Greases: Preliminary Microwave-assisted Extraction Studies', *Analyst*, **121**, 1031–1036 (1996).
80. F.I. Onuska, K.A. Terry, 'Microwave Extraction in Analytical Chemistry of Pollutants: Polychlorinated Biphenyls', *J. High Resolut. Chromatogr.*, **18**, 417–421 (1995).
81. F.I. Onuska, K.A. Terry, 'Microwave Extraction in Analytical Chemistry of Pollutants: Chlorinated Benzenes', *J. Microcol. Sep.*, **7**, 319–326 (1995).
82. V. Lopez-Avila, J. Benedico, R. Young, C. Charan, W.F. Beckert, 'Determination of PCBs in Soils/Sediments by Microwave-assisted Extraction and GC/ECD or ELISA', *Environ. Sci. Technol.*, **29**, 2709–2712 (1995).
83. V. Lopez-Avila, R. Young, N. Teplitsky, 'Microwave-assisted Extraction as an Alternative to Soxhlet, Sonication, and Supercritical Fluid Extraction', *J. AOAC Int.*, **79**, 142–156 (1996).
84. G.-H. Xiong, J.-M. Liang, S.-C. Zhou, X.-Q. He, Z.-X. Zhang, 'Microwave-assisted Extraction of PAHs from Soils', *Chem. J. Chinese Univ.*, **19**, 1560–1565 (1998) (in Chinese).
85. K.K. Chee, M.K. Wong, H.K. Lee, 'Optimization of the Microwave-assisted Solvent Extraction of Petroleum Hydrocarbons in Marine Sediment and Soil', *Environ. Monit. Assess.*, **44**, 587–603 (1997).
86. K.K. Chee, M.K. Wong, H.K. Lee, 'Membrane Solid-phase Extraction (SPE) with Closed-vessel Microwave Elution for the Determination of Phenolic Compounds in Aqueous Matrixes', *Mikrochim. Acta*, **126**, 97–104 (1997).
87. J.R. Fish, R. Revesz, 'Microwave Solvent Extraction of Chlorinated Pesticides from Soil', *LC-GC*, **14**, 203–232, 234 (1996).
88. M.J. Vazquez, A.M. Carro, R.A. Lorenzo, R. Cela, 'Optimization of Methylmercury Microwave-assisted Extraction from Aquatic Sediments', *Anal. Chem.*, **69**, 221–225 (1997).
89. C.M. Tseng, A. de Diego, F.M. Martin, O.F.X. Donard, 'Rapid and Quantitative Microwave-assisted Recovery of Methylmercury from Stand Reference Sediments', *J. Anal. At. Spectrom.*, **12**, 629–635 (1997).
90. T. Dagnac, A. Padro, R. Rubio, G. Rauret, 'Optimisation of the Extraction of Arsenic Species from Mussels with Low Power Focused Microwaves by Applying a Doehlert Design', *Anal. Chim. Acta*, **364**, 19–30 (1998).
91. O. Donard, B. Lalere, F. Martin, R. Lobinski, 'Microwave-assisted Leaching of Organotin Compounds from Sediments for Speciation Analysis', *Anal. Chem.*, **67**, 4250–4254 (1995).
92. US Environmental Protection Agency, *Test Methods for Evaluating Solid Waste*, 3rd edition, SW-846, Method 8250 (Revision 1. November 1992), US Environmental Protection Agency, Washington DC, 1986.
93. US Environmental Protection Agency, *Test Methods for Evaluating Solid Waste*, 3rd edition, SW-846, Method 8081A (Revision 0. November 1992), US Environmental Protection Agency, Washington DC, 1986.

94. US Environmental Protection Agency, *Test Methods for Evaluating Solid Waste*, 3rd edition, SW-846, Method 8141A (Revision 1. November 1992), US Environmental Protection Agency, Washington DC, 1986.
95. L.E. Garcia-Ayuso, M. Sanchez, A.F. de Alba, M.D. Luque de Castro, 'Focused Microwave-assisted Soxhlet: An Advantageous Tool for Sample Extraction', *Anal. Chem.*, **70**, 2426–2431 (1998).
96. J.R.J. Pare, J.M.R. Belanger, K. Li, S.S. Stafford, 'Microwave-assisted Process (MAP): Application to the Headspace Analysis of VOCs in Water', *J. Microcol. Sep.*, **7**, 37–40 (1995).
97. M. Khan, J.P. Williams, *Lipids*, **28**, 953 (1993).
98. G.-K. Li, X.-Q. He, Z.-X. Zhang, 'Determination of Fatty Acids in Food Oils by GC/MS After Microwave-assisted Derivatization', *Anal. Lab. (in Chinese)*, (in press).
99. R. Gedye, K. Westaway, F.E. Smith, 'Microwave Ovens in the Laboratory', *Educ. in Chem.*, **25**, 55–56 (1988).
100. F. Martin, O.F.X. Donard, 'Interference Mechanisms and Reduction During the Speciation of Organotin Compounds by Hydride Generation, Cryoseparation and Detection by Atomic Absorption Spectrometry', *Anal. At. Spectrom.*, **9**, 1143–1151 (1994).
101. M. Ceulemans, C. Witte, R. Lobinski, F.C. Adams, 'Simplified Sample Preparation for GC Speciation Analysis of Organotin in Marine Biomaterials', *Appl. Organomet. Chem.*, **8**, 451–461 (1994).
102. I.R. Pereiro, V.O. Schmitt, J. Szpunar, O.F.X. Donard, R. Lobinski, 'Speciation Analysis for Organotin Compounds in Biomaterials After Integrated Dissolution, Extraction, and Derivatization in a Focused Microwave Field', *Anal. Chem.*, **68**, 4135–4140 (1996).
103. I. Rodriguez, M. Santamarina, M.H. Bollain, M.C. Mejuto, R. Cela, 'Speciation of Organotin Compounds in Marine Biomaterials After Basic Leaching in a Non-focused Microwave Extractor Equipped with Pressured Vessels', *J. Chromatogr. A*, **774**, 379–384 (1997).
104. M.P. Llompart, R.A. Lorenzo, R. Cela, J.R.J. Pare, J.M.R. Belanger, K. Li, 'Phenol and Methylphenol Isomers Determination in Soils by In situ Microwave-assisted Extraction and Derivatization', *J. Chromatogr. A*, **757**, 153–164 (1997).
105. S.-H. Chiou, K.-T. Wang, 'A Rapid and Novel Means of Protein Hydrolysis by Microwave Irradiation Using Teflon–Pyrex Tubes', in *Current Research of Protein Chemistry: Technique, Structure, and Function (Papers of 3rd Annual Symposium of Protein Society 1989)*, ed. J.J. Villafranca, Academic Press, San Diego, California, 139–149, 1990.
106. E. Tatar, M. Khalifa, G. Zaray, I. Molnar-Perl, 'Comparison of the Utility of Capillary Zone Electrophoresis and High-performance Liquid Chromatography in Peptide Mapping and Separation', *J. Chromatogr. A*, **672**, 109–115 (1994).
107. T. Arai, N. Kakishita, 'A Chemical Experiment by Microwave Heating. I. Hydrolysis and Synthesis of Esters', *Kagaku to Kyoiku*, **39**, 564–565 (1991) (*Chem. Abstr.*, **116**, abs. no. 20387y, 1992) (in Japanese).
108. A. Morales-Rubio, J. Cerezo, A. Salvador, M. de la Guardia, 'Rapid Acid Hydrolysis of Albumin in a Microwave Oven', *Microchem. J.*, **47**, 270–277 (1993).
109. A. Morales-Rubio, J.V. Julian-Ortiz, A. Salvador, M. de la Guardia, 'Hydrolysis of Bensodiazepines in a Microwave Oven and Ultraviolet Derivative Analysis of Their Benzophenones', *Microchem. J.*, **49**, 12–19 (1994).
110. T. Taketomi, A. Hara, K. Uemura, H. Kurahashi, E. Sugiyama, 'A Microwave-mediated Saponification of Galactosylceramide and Galactosylceramid I^3-sulfate and Identification of Their Lyso-compounds by Delayed Extraction Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry', *Biochem. Biophys. Res. Commun.*, **224**, 462–467 (1996).
111. E.L. Perez, S.J. Haswell, 'Development of an FIA Method with Online Microwave-assisted Saponification of Vitamin A to Retinol and Direct In-line Spectrophometric Determination', *Anal. Proc.*, **32**, 85–89 (1995).
112. J. Szpunar, V.O. Schmitt, R. Lobinski, J.-L. Monod, 'Rapid Speciation of Butyltin Compounds in Sediments and Biomaterials by Capillary Gas Chromatography-microwave-induced Plasma Atomic Emission Spectrometry After Microwave-assisted Leaching/Digestion', *J. Anal. At. Spectrom.*, **11**, 193–199 (1996).
113. V.O. Schmitt, J. Szpunar, O.F.X. Donard, R. Lobinski, 'Microwave Accelerated Sample Preparation for Speciation Analysis of Organotin Compounds in Biomaterials', *Can. J. Anal. Sci. Spectrosc.*, **42**, 41–49 (1997).
114. J. Szpunar, M. Ceulemans, V.O. Schmitt, F.C. Adams, R. Lobinski, 'Microwave-accelerated Speciation Analysis for Butyltin Compounds in Sediments and Biomaterials by Large Volume Injection Capillary Gas Chromatography Quartz Furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **332**, 225–232 (1996).
115. S. Shawky, H. Emons, H.W. Durbeck, 'Speciation of Organotin Compounds in Fish Samples', *Anal. Commun.*, **33**, 107–110 (1996).
116. Z.-X. Zhang, G.-H. Xiong, G.-K. Li, X.-Q. He, 'Environmental Sample Pretreatment with Microwave-assisted Techniques', *Chem. J. Chinese Univ.*, **20**, 542–557 (1999).
117. G.-H. Xiong, X.-Q. He, Z.-X. Zhang, 'Microwave-assisted Extraction or Saponification Combine with Microwave-assisted Decomposition Applied in Pretreatment of Soil or Mussel Samples for the Analysis of Polychlorinated Biphenyls', *Anal. Chim. Acta*, (in press).

Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis

John R. Dean
University of Northumbria at Newcastle, Newcastle upon Tyne, UK

1 Introduction to Environmental Organic Sample Preparation 1
2 Principles of Microwave Heating for Organic Solvents 2
3 Apparatus for Microwave-enhanced Solvent Extraction 2
4 Applications of Microwave-enhanced Solvent Extraction in Environmental Analysis 3
4.1 Extraction from Solid Matrices 3
4.2 Extraction from Aqueous Matrices 8
5 Conclusions 9
Abbreviations and Acronyms 9
Related Articles 10
Further Reading 10
References 10

Microwave-enhanced solvent extraction is a relatively new approach for the extraction of organic pollutants from solid environmental matrices. This approach uses microwave technology to heat organic solvent in contact with the sample in either a sealed (pressurized) or open (atmospheric) vessel. The major advantage of this approach over conventional extraction procedures, such as Soxhlet extraction, is the speed of the process. As well as the principles of microwave heating, the article also describes the instrumentation required to effect the extraction process. Finally, selected applications of microwave-enhanced solvent extraction are described. Applications are divided into extraction from solid matrices and extraction from aqueous matrices. The specific analytes covered include the extraction of polycyclic aromatic hydrocarbons (PAHs), hydrocarbons, polychlorinated biphenyls (PCBs), pesticides and phenols from solid matrices; direct extraction of chlorinated benzenes from aqueous matrices; microwave elution of a range of analytes from solid-phase extraction (SPE) media and, finally, volatile compounds, e.g. benzene, toluene, ethylbenzene and xylenes (BTEX) from aqueous samples. Finally, some recommended operating conditions for the extraction of analytes from solid matrices using pressurized microwave-assisted extraction (MAE) are provided.

1 INTRODUCTION TO ENVIRONMENTAL ORGANIC SAMPLE PREPARATION

Analysis of organic pollutants in the environment is necessary because of past or present industrial activity, accidental spillage and deliberate (unlicensed) disposal, all of which can have ecotoxicological effects on living organisms. In order to assess the potential damage, remediate known sites or simply to have an early warning of potential problems requires environmental measurement of the levels of organic pollutants. Organic pollutants are located in all compartments of the environment. Thus, typically we refer to pollution in terms of natural waters (e.g. lakes, rivers, groundwater, potable water), solid or semisolid matrices (e.g. soil, sludge, vegetation) and the atmosphere (e.g. air). In this article we are concerned principally with solid-type matrices, although some attention is given to aqueous sample preparation. A recently published book discusses the different types of sample preparation approaches for both liquid and solid samples of environmental origin.[1]

Preparation of solid samples for organic analysis can traditionally be subdivided into two classes, with heating and without heating. In the latter case, we refer to the use of a shake flask, while in the former Soxhlet extraction is used. Shake-flask extraction involves placing a solid sample into a container together with an organic solvent and agitating, either manually or by the use of a laboratory shaker, for a given period of time. Subsequently the organic solvent containing the extract is removed via filtration, a separating funnel or decanting. Soxhlet extraction involves refluxing warm organic solvent through the sample repeatedly for several hours. This is achieved by heating organic solvent contained in a round-bottomed flask, on an isomantle. The vaporized solvent is then condensed, via a water-cooled condenser, and falls into a thimble containing sample. This in turn returns to the round-bottomed flask. The whole process is repeated frequently until the preset extraction time is reached. As the extracted analyte will normally have a higher boiling point than the solvent it is preferentially retained in the flask and fresh solvent recirculates. This ensures that only fresh solvent is used to extract the analyte from the sample in the thimble. A disadvantage of this approach is that the organic solvent is below its boiling point when it passes through the sample contained in the thimble. In practice this is not

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

necessarily a problem as Soxhlet extraction is normally done over long time periods i.e. 6, 12, 18, or 24 h.

More recently, alternative approaches have become available. In addition to the shake-flask approach it is possible to agitate the sample using a sonic bath or sonic probe. Modifications of Soxhlet extraction have appeared which offer some degree of solvent reduction and automation, e.g. Soxtec. Perhaps the biggest changes have occurred in terms of the instrumentation available for extraction of pollutants from solids. The 1980s saw the arrival of commercial apparatus for supercritical-fluid extraction (SFE), while the early 1990s heralded the commercial availability of MAE closely followed in 1995 by accelerated solvent extraction (ASE).

This article is concerned with the use of microwave technology for the extraction of organic pollutants from environmental (solid and liquid) matrices. While the first papers on the subject were published in 1986[2] using a domestic microwave oven, this article only considers the work published in the 1990s when commercial systems became available.

2 PRINCIPLES OF MICROWAVE HEATING FOR ORGANIC SOLVENTS

Microwaves are high-frequency electromagnetic radiation. The most commonly used frequency of operation of microwaves is 2.45 GHz. The choice of solvent for MAE is essential. The solvent must be able to absorb microwave radiation and pass it on in the form of heat to other molecules in the system. The ability to pass on this energy is measured in terms of the dissipation factor ($\tan\delta$), see Equation (1):

$$\tan\delta = \frac{\varepsilon''}{\varepsilon'} \qquad (1)$$

where, ε'' is the dielectric loss (a measure of the efficiency of conversion of microwave energy into heat energy) and ε' is the dielectric constant (a measure of the polarizability of a molecule in an electric field).

This polarizability is achieved by the reorientation of permanent dipoles by the applied electric field. This means that under microwave conditions, a polarized molecule will rotate to align itself with the electric field at a rate of approximately 10^9 times per second. This results in very rapid heating. An estimate of the ability of the microwave oven to couple to any molecule can be obtained by considering its ε' values (Table 1). It is not suprising to find therefore that polar solvents, such as water, acetone and methanol all readily absorb microwaves and hence are heated up when subjected to microwave radiation, while nonpolar solvents, such as hexane and toluene, do not heat up when they are subjected to microwave irradiation. In addition, since the extraction process typically takes place in a closed vessel, the solvent chosen can be heated well above its normal boiling point (Table 1). This will reduce the time required for the extraction process.

Table 1 Solvent properties for MAE

Solvent	Dielectric constant[a]	Boiling point (°C)[b]	Closed-vessel temperature (°C) at 175 psig[b,c]
Hexane	1.89	68.7	NH
DCM[d]		39.8	140
Acetone	20.7	56.2	164
Methanol	32.63	64.7	151
Acetonitrile	37.5	81.6	194
Acetone : hexane (1 : 1 v/v)	–	52	156

NH – no heating in microwave.
[a] See *CRC Handbook of Physics and Chemistry*.[3]
[b] Adapted from Renoe.[4] (1 psi = 6894.76 Pa.)
[c] Pounds per square inch gauge.
[d] Dichloromethane.

3 APPARATUS FOR MICROWAVE-ENHANCED SOLVENT EXTRACTION

Two types of microwave heating systems are commercially available: an open-vessel (atmospheric) system and a closed-vessel (pressurized) system. In the open-style system, individual sample vessels are heated sequentially. A typical commercial system is the Soxwave from Prolabo Ltd., France. A schematic diagram of the atmospheric-MAE system is shown in Figure 1. This system operates at percentage power increments from 0–100%, corresponding to a maximum of 300 W. These power increments can be operated in stages and for various time intervals (up to 9 h). The sample and solvent are introduced into

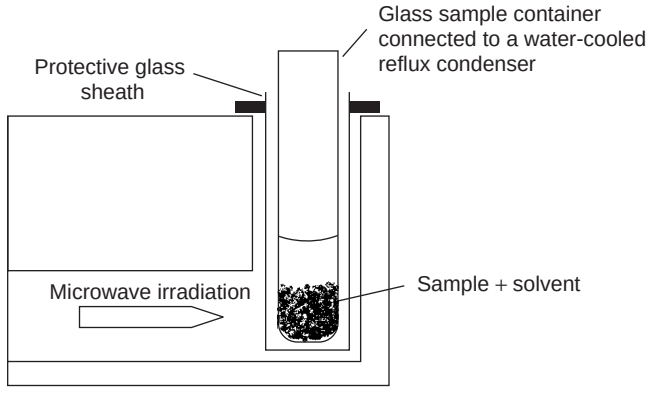

Figure 1 Schematic diagram of an atmospheric MAE system.

a glass container which is fitted with either an air or a water condenser to prevent loss of volatiles and solvent. The sample container is located within a protective glass sheath. The organic solvent is then heated, by means of microwave energy, and refluxed through the sample.

A common commercial closed system (pressurized-MAE) is the MES-1000 microwave solvent extraction system (CEM Corp., USA). A schematic diagram of the pressurized-MAE system is shown in Figure 2. (Note: nine extraction vessels only are shown.) This system allows up to 12 extraction vessels to be irradiated simultaneously, in 1% increments, up to 950 W of microwave energy at 100% power. This system has in-built safety features, such as an audible alarm for the detection of an unexpected release of flammable and toxic organic solvent, a Teflon-lined microwave cavity, and an exhaust fan which continues to operate in the event of a solvent escape. Microwave conditions can be monitored via in situ pressure and temperature sensors (located within a single extraction vessel only). Pressure is measured using a water manometer that allows readings of up to 200 psi (1 psi = 6894.76 Pa) to be made. The temperature probe, a fiber optic with a phosphor sensor, allows extraction temperatures to be selected from 20–200 °C in 1 °C increments. Samples are placed into Teflon perfluoroalkoxy (PFA) vessels (100 mL capacity) which are themselves located within a polyetherimide protective casing (bodies and caps). Each extraction vessel contains a rupture membrane that is designed to fail at elevated pressure (200 psi). Each extraction vessel (up to 12) is located in a carousel which rotates through 180° during microwave operation. In the center of the carousel is an expansion chamber which is connected to each extraction vessel, and which acts to contain escaping vapors in the event of rupture membrane failure.

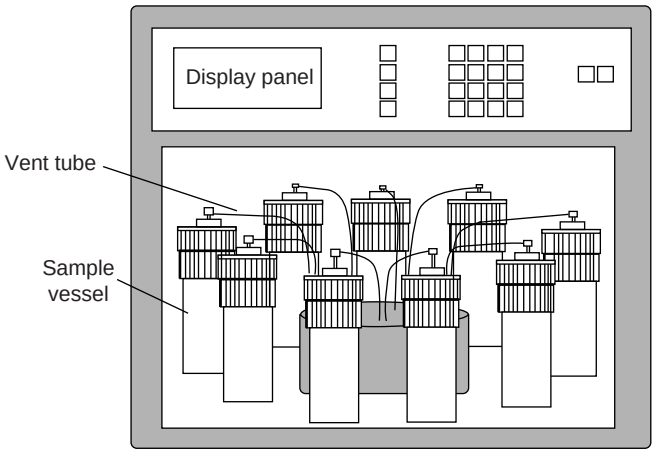

Figure 2 Schematic diagram of a pressurized MAE system.

4 APPLICATIONS OF MICROWAVE-ENHANCED SOLVENT EXTRACTION IN ENVIRONMENTAL ANALYSIS

4.1 Extraction from Solid Matrices

4.1.1 Atmospheric Microwave-enhanced Solvent Extraction

Akhtar and Croteau[5] reported the extraction of sodium salinomycin (CAS 55721-31-8, referred to as salinomycin antibiotic) from finished chicken feed. It is added to the feed at a level of 60 mg kg^{-1}. Salinomycin is used to prevent coccidiosis in broiler chickens. A conventional microwave oven (800 W) was used for the extraction followed by postcolumn derivatization high-performance liquid chromatography (HPLC). The feed (0.6 g) was placed in a polyethylene container and extracted with ethanol (15 mL) and propan-2-ol (2 mL) for 8 s. After irradiation, the organic layer was decanted off and the microwave process repeated twice more. All extracts (three times) were combined, centrifuged and evaporated to dryness. The residue was reconstituted in 10 mL of mobile phase. The results were compared with a conventional approach based on liquid–solid extraction and vortex mixing for 3 min. The same solvent system and volumes were used. The sample extract was treated the same as for the microwave extraction. The results in the finished feed were as follows: 34, 60, 103 and 160 mg kg^{-1} by MAE and 27, 45, 59 and 112 mg kg^{-1} using liquid–solid extraction. The expected values, based on calculation, were 30, 60, 90 and 150 mg kg^{-1}. The authors concluded that MAE shows great potential for the monitoring of veterinary drug and pesticide residues in animal feeds and edible tissues and eggs.

Focused MAE (Maxidigest 350, Prolabo, France) was applied to the extraction of PAHs from reference materials (sediment, soil, air particulates) by Letellier et al.[6] Initial studies were performed on marine sediment (National Institute of Science and Technology (NIST) standard reference material (SRM) 1941a) where it was found that small quantities of organic solvent (30 mL DCM) at short extraction times (10 min) and 30 W microwave power provided good recoveries (average recovery 87% for 13 PAHs) with standard deviations comparable to those obtained using Soxhlet extraction (2 × 250 mL DCM for 48 h). This approach was then applied to two other reference samples: soil (Community Bureau of Reference (BCR), certified reference material (CRM) 524) and air particulate/urban dust (NIST SRM 1649a). The results, compared to certificate values, were in good agreement (average recovery 97% with a relative standard deviation (RSD) of 6.4%, $n = 18$). The procedure was also applied to two sediment samples (industrial harbour marine sediment and yatching

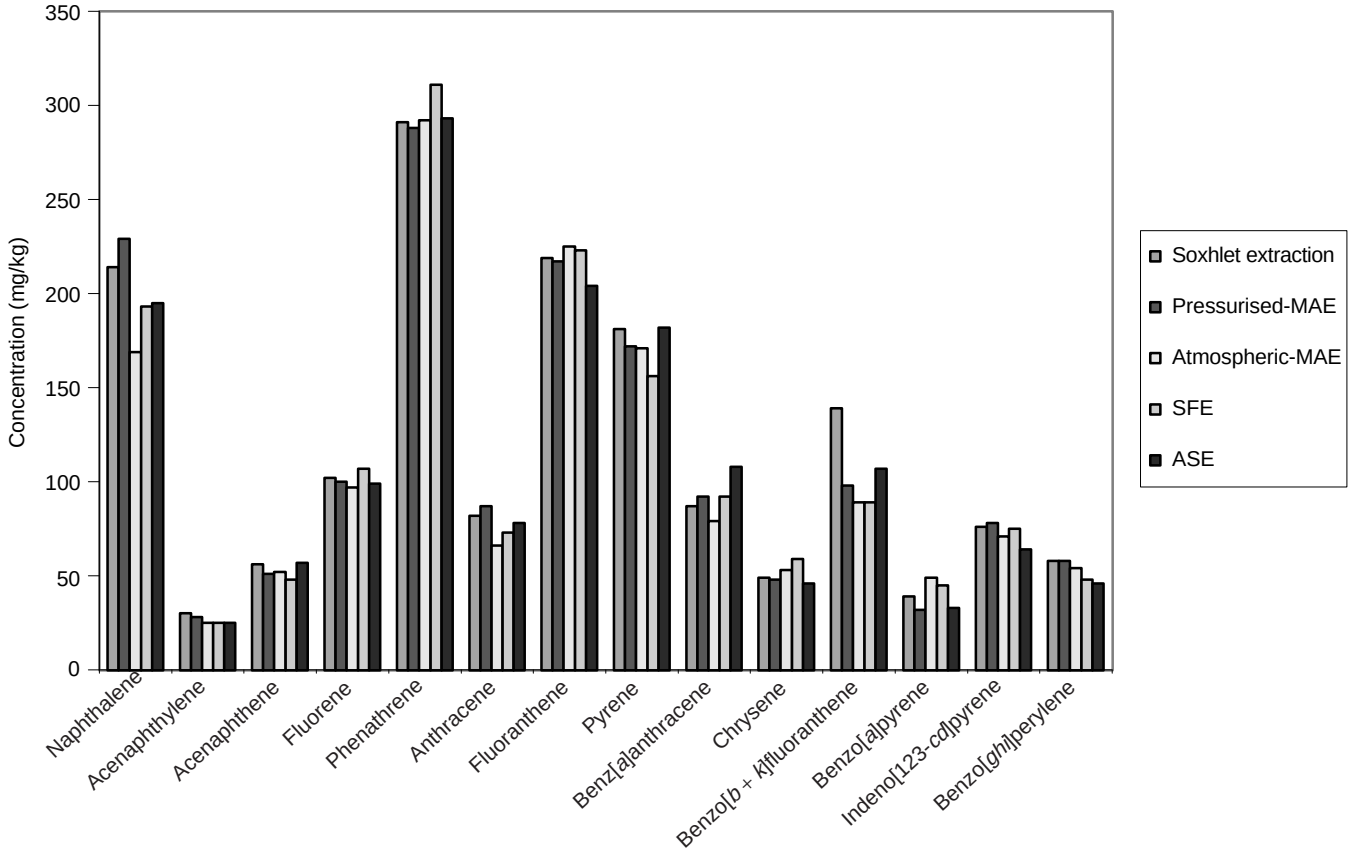

Figure 3 Extraction of PAHs from contaminated soil. (Adapted from Saim et al.[7])

harbour sediment). The results, compared to Soxhlet extraction, produced slightly lower recoveries (average recovery 90% with an RSD of 7.9%, $n = 24$). The benefits of this open MAE approach are the speed of analysis, simplicity of operation and safety in use. It is noted however, that the group did not include the most volatile PAHs in their remit, omitting two-ring aromatics, i.e. naphthalene (CAS 91-20-3).

Saim et al.[7] compared the merits of Soxhlet extraction, pressurized-MAE, SFE and ASE with atmospheric-MAE for the extraction of PAHs from native contaminated soils. Atmospheric-MAE was done using a Soxwave-100 system. A 2-g soil sample was irradiated with 70 mL DCM for 20 min at 99% power. The extract was filtered and reduced in volume to 5 mL prior to analysis. The results (Figure 3), for 14 PAHs, show that comparable recoveries are obtainable by all extraction methods. The sum of the 14 PAHs by each extraction technique were as follows: 1623, 1578, 1492, 1544 and 1537 mg kg^{-1} by Soxhlet extraction, pressurized-MAE, atmospheric-MAE, SFE and ASE, respectively. Finally, the individual PAH recoveries were statistically compared with Soxhlet extraction. It was found, at the 95% confidence interval, that the majority of PAH determinations were not significant for MAE (atmospheric and pressurized). The authors concluded however, that the method choice, based on the data presented, still appears to be Soxhlet extraction.

4.1.2 Pressurized Microwave-enhanced Solvent Extraction

Pressurized-MAE is by far the most common approach that has been utilized for the extraction of organic pollutants from solid environmental matrices. Accordingly this section has been subdivided with respect to the analyte of interest.

4.1.2.1 Polycyclic Aromatic Hydrocarbons PAHs are the most commonly studied organic pollutant by MAE. As part of an on-going evaluation of new sample preparation techniques by the United States Environmental Protection Agency (USEPA), especially those that minimize waste solvents, MAE has been evaluated.[8] Using a 1 : 1 hexane : acetone solvent mixture, the influence of extraction temperature (80, 115 and 145 °C) and extraction time (5, 10 and 20 min) on the recovery of PAHs from CRMs was compared with room-temperature extraction using the same solvent and contact time. It

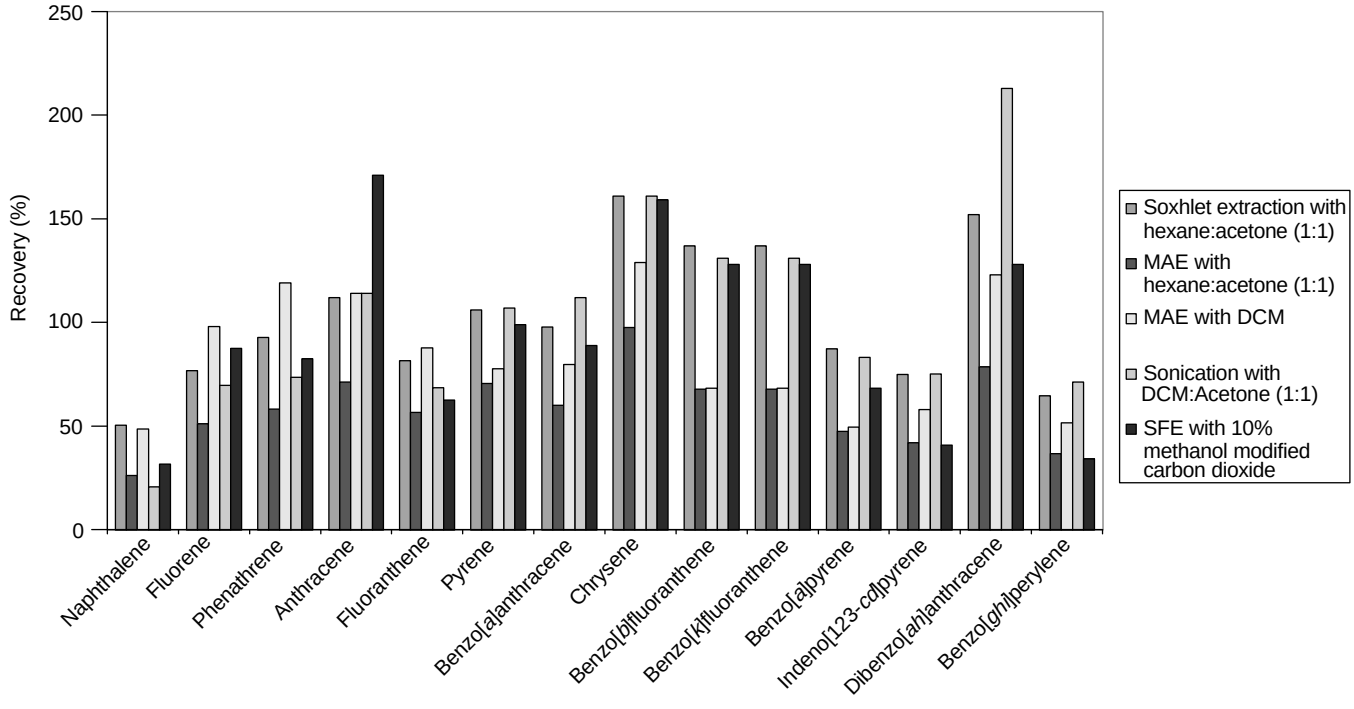

Figure 4 Extraction of PAHs from sediment (NIST SRM 1941a). (Adapted from Lopez-Avila et al.[10])

was found that the average recoveries by MAE for the 17 PAHs studied ranged from 70–75%. The authors concluded that MAE offers reduced extraction time (typically 10 min for extraction and 40 min for cooling, centrifugation and extract concentration) and reduced solvent use (30 mL versus 300 mL for Soxhlet/Soxtec). In addition, up to 12 samples can be extracted simultaneously. This work was extended to incorporate a spiked soil sample with and without ageing.[9] It was noted that recoveries from aged samples decreased, compared to freshly spiked samples. The same group[10] extended this work to compare MAE with Soxhlet, sonication and SFE. Again freshly spiked soil samples were evaluated along with two CRMs. A whole range of compounds were considered (95 compounds in total). It was concluded that MAE gave superior or equivalent results to the other techniques considered. Of the compounds studied, the number of recoveries >80% was 51 by MAE, 50 by Soxhlet, 63 by sonication, and 37 by SFE. As an example, the recoveries of 14 PAHs from a certified sediment sample (SRM 1941a) by the different extraction techniques are shown in Figure 4.

Similarly, Dean et al. have evaluated the effectiveness of MAE for the extraction of PAHs from industrially contaminated soil.[11–13] Initial work considered the robustness of MAE.[11] It was found that acetone was the solvent of choice compared to hexane : acetone (1 : 1) and DCM only. An experimental design approach was used to investigate the dependence of operating variables (temperature, extraction time and solvent volume). These variables were considered between strict operating limits, i.e. temperature, 40–120 °C; extraction time, 5–20 min and solvent volume, 30–50 mL. It was concluded that MAE was independent of the variables considered (and their limits) for the extraction of PAHs from native contaminated soil. The PAH recoveries from the soil by MAE (20 min extraction plus 30 min cooling time using 40 mL acetone) were compared with Soxhlet (extraction for 6 h using DCM), SFE (1 h extraction using 10% methanol-modified CO_2) and ASE (12 min extraction using 25 mL acetone : DCM 1 : 1).[12,13] It was found that MAE compared favorably with SFE and ASE. All techniques gave superior results compared to Soxhlet extraction. The sum of 16 individual PAHs was 422.9, 297.4, 458.0 and 421.0 mg kg^{-1} by MAE, Soxhlet, SFE and ASE, respectively. Solvent choice was considered to be an important variable for extraction of native contaminants.

Certified marine sediments and soil have been extracted by MAE and analyzed for PAHs.[14] The sample (1–5 g) was irradiated with microwave radiation in the presence of 30 mL of hexane : acetone (1 : 1), for 10 min at 115 °C. Recoveries of 14 PAHs from certified soil (standard reference soil (SRS)-103–100) ranged from 67% with an RSD of 22.1% for dibenzo[*ah*]anthracene (CAS 53-70-3) to 111% with an RSD of 10.2% for benzo[*k*]fluoranthene (CAS 207-08-09).

The certified values ranged from 14.2–1924.9 ppm. Similarly, recoveries of 16 PAHs from marine sediments

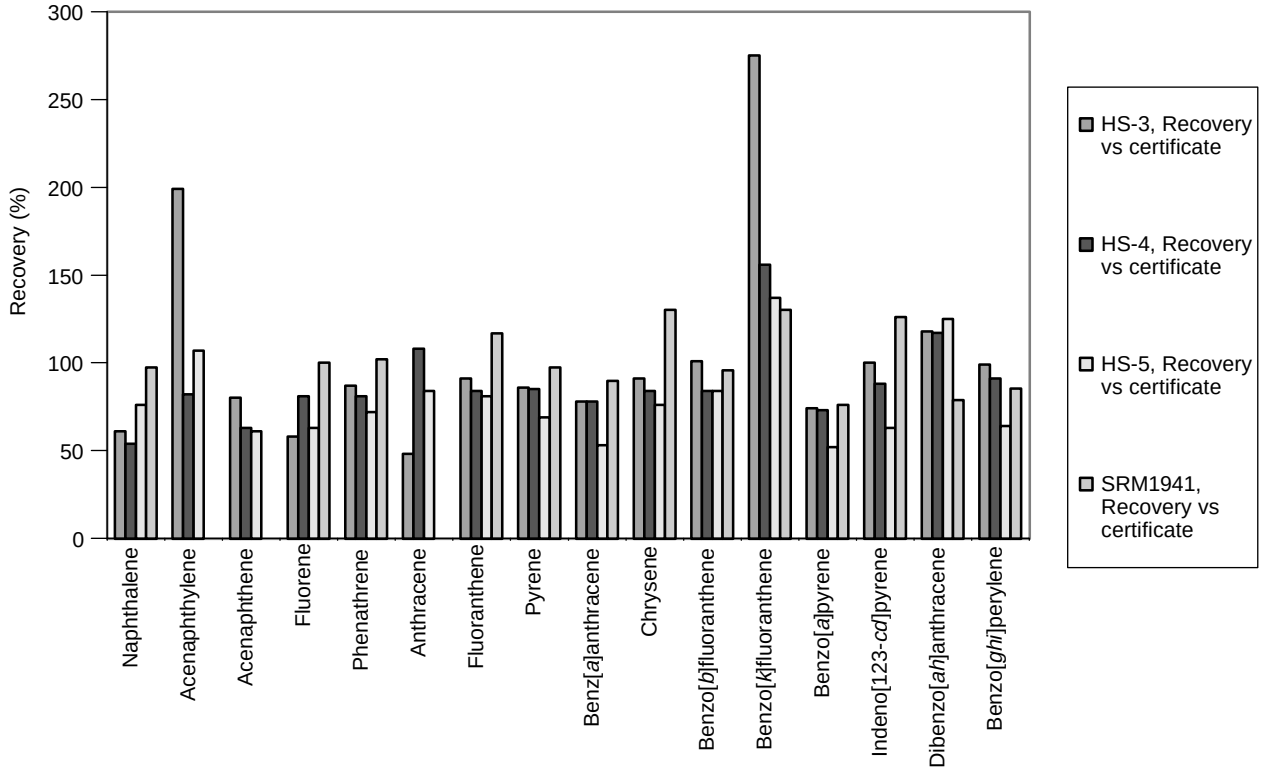

Figure 5 MAE of PAHs from CRMs. (Adapted from Li et al.[14])

(HS-3, HS-4 and HS-5) ranged from 61% (naphthalene) to 275% (benzo[*k*]fluoranthene) for HS-3; 54% (naphthalene) to 156% (benzo[*k*]fluoranthene) for HS-4; and, 52% (benzo[*a*]pyrene, CAS 50-32-8) to 137% (benzo[*k*]fluoranthene) for HS-5. Average recoveries, based on 16 individual PAHs, for HS-3, HS-4 and HS-5 were 103, 88 and 79%, respectively. In addition, a further marine sediment was also analyzed (SRM 1941) for 15 PAHs. Recoveries ranged from 72.4% (perylene) to 130.2% (benz[*k*]fluoranthene) with an average recovery, based on 15 individual PAHs, of 99%. The results for the extraction of PAHs from certified reference samples are summarized in Figure 5.

An evaluation of extraction solvent for the MAE of PAHs from sediment samples has been reported.[15] Three solvent systems were evaluated: acetone : hexane (1 : 1), hexane : water and toluene : water. The acetone : hexane mixture was included in this study because it has been by far the most studied extraction solvent system for MAE. For the other two solvent systems, water was added at 10% (v/v) in all cases. MAE was done by placing the sample (2–10 g) and solvent mixture in a hermetically sealed polytetrafluoroethylene (PTFE) reactor and then irradiating at 660 W power for 6 min. After cooling, the sample container was opened and 8 g of sodium sulfate added to dry the organic phase. All extracts were then evaporated to dryness and dissolved in 2 mL hexane. Copper wire was then added to remove sulfur and the solution was finally passed through a microcolumn of activated Florisil and eluted with 30 mL of DCM : hexane (70 : 30 v/v). Finally, the extract was evaporated to dryness and dissolved in 200 µL solvent. Final extracts were quantified using gas chromatography (GC) with either flame ionization detection (FID) or mass selective detection (MSD). The recoveries, compared to Soxtec extraction, for a marine sediment were in agreement (average recoveries were 101, 99 and 102% using acetone : hexane, hexane : water and toluene : water, respectively). The automated Soxtec extraction was based on 2 g sample, 60 mL hexane : acetone (1 : 1 v/v) for 1 h sample immersion followed by 1 h of heating under reflux.

4.1.2.2 Hydrocarbons A method was developed for the quantitative extraction of gasoline- and diesel-range petroleum hydrocarbons from soil and marine sediment samples.[16] Microwave operating parameters were evaluated using a two-level orthogonal array design. The operating parameters varied were as follows: solvent (DCM and acetone), temperature (80 and 115 °C) and extraction time (5 and 15 min). It was found that the optimum conditions were: 30 mL of acetone at a temperature of 115 °C and heated for 5–15 min. The effects of

wetting the sample on the recoveries of 14 hydrocarbons were investigated using spiked soil and marine sediment. Under dry conditions recoveries ranged from 60.0 to 88.3% with an RSD of 3.5–10.1% for marine sediment and from 60.4 to 89.6% with an RSD of 3.7–13.5% for soil. Under wet conditions (10% w/w) the recoveries improved; 80.5–100% with an RSD of 2.0–7.9% for marine sediment and 80.2–100.4% with an RSD of 2.3–7.4% for soil. The method was then applied to a range of local (Singapore) sediment and soil samples. It was concluded that the level of hydrocarbon contamination was low in the areas sampled.

MAE of linear hydrocarbons (C_{16}–C_{32}) from a lyophilized sediment sample using a mixture of toluene and water (90 : 10% v/v) is proposed.[15] For further details of the experimental procedure, see section 4.1.2.1 on PAHs. The authors reported that, provided the microwave power was maintained above 66%, quantitative recoveries could be obtained irrespective of the extraction time (3–12 min). Similarly, maintaining a sample : toluene ratio greater than 1 : 2 provided quantitative recoveries. Extraction yields from the native sediment samples were assessed by comparison with Soxhlet extraction.

4.1.2.3 Polychlorinated Biphenyls PCBs have been extracted from a range of CRMs (marine sediment and soils) using MAE.[9] The following proven conditions were used: solvent, 30 mL of hexane : acetone (1 : 1 v/v); temperature 115 °C; and extraction time 10 min. The recoveries compared with certificate values were as follows: Arochlor 1254 from marine sediment HS-1 and HS-2, 93.2 and 76.7%, respectively; Arochlor 1260 from Environmental Research Associates (ERA) soil (lot number 9801), 89.9%; and Arochlor 1248 from Superfund site soils, 102, 157, 86.8 and 75.3%. Values for the Superfund site soils were based on those obtained by an independent laboratory using Soxhlet extraction and gas chromatography/electron capture detection (GC/ECD). The same certified reference samples (HS-1, HS-2 and the ERA soil) were analyzed by Lopez-Avila et al.[17] to compare the method of detection (GC/ECD or ELISA (enzyme-linked immunosorbent assay) of the PCBs after MAE. It was found that the results closest to the certificate values were obtained using GC/ECD. However, it was concluded that the ELISA approach was more rapid. A batch of 10 samples could be prepared and analyzed in approximately 1 h.

4.1.2.4 Pesticides Sample ageing, either artificially or by natural weathering, is an important factor to consider when evaluating any new extraction technique. However, it is unfortunate to note that most published methods for MAE are based on freshly spiked samples. Onuska and Terry[18] aged a slurry spiked air-dried sediment for at least one month prior to MAE extraction of organochlorine pesticides (OCP). It was found that the optimum extraction conditions consisted of an extraction solvent of isooctane : acetonitrile (1 : 1 v/v), a mimimum water content of the sample of 15% and an extraction time of >3 min. The results for the extraction of 15 OCPs from a spiked sediment sample (at spiking levels between 50 and 250 $\mu g\,kg^{-1}$) were good, the minimum recovery was 74% (3% RSD) for *p*,*p*′-DDE (1,1′-(2,2-dichloroethylenylidene)-bis(4-chlorobenzene) or 2,2-bis(4-chlorophenyl)-1,1-dichloroethylene) (CAS 72-55-9) while the best recovery was 95.3% (3.9% RSD) for methoxychlor (CAS 72-43-5) ($n = 5$).

The nature of the MAE process is such that heating can only occur if the solvent has a permanent dipole moment. As an alternative to the conventional approach, Hummert et al.[19] utilized a microwave transformer (Weflon®, an inert material based on carbon-containing teflon) to extract OCPs from seal blubber using a nonpolar solvent, e.g. hexane. It was experimentally determined that the best approach for MAE was to operate with seven extraction cycles (30 s duration at full power). Average recoveries for the seven organochlorine compounds from seal blubber were 96.9 ± 0.5%.

Extraction of a fungicide, hexaconazole, which had been weathered over periods of time up to one year was reported from two characterized soils.[20] The soils, which had been collected at various times after application of the fungicide, were air dried for 48 h and sieved through a 2–3.5 mm sieve prior to storage at −18 °C. The levels of hexaconazole in the two soil types were assessed using Soxhlet extraction, MAE, SFE and ASE. Extraction with microwave energy was done by irradiating a 5-g sample with 30 mL of acetone at 115 °C for 15 min. The soil extract was then precleaned using C_{18} SPE media and after elution with methanol the volume reduced to 1 mL. The results for the sandy clay soil (organic matter, 1.5%) were comparable by all extraction techniques (Figure 6). For example, the 52-week aged sample gave recoveries of 0.08, 0.072, 0.073 and 0.070 $mg\,kg^{-1}$ by Soxhlet, SFE, MAE and ASE, respectively. However, this was not

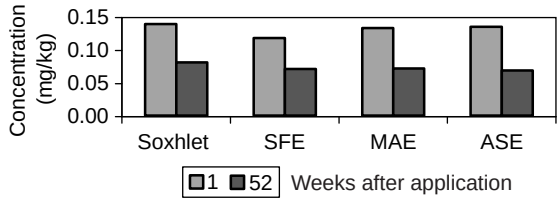

Figure 6 Extraction of hexaconazole from aged soils: sandy clay soil (organic matter, 1.5%). (Adapted from Frost et al.[20])

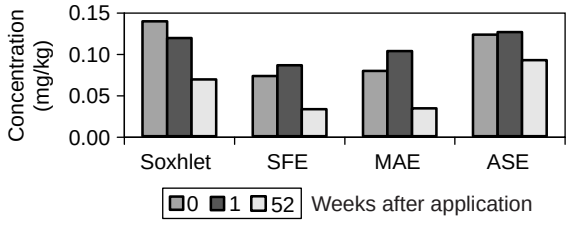

Figure 7 Extraction of hexaconazole from aged soils: sandy loam soil (organic matter, 5.7%). (Adapted from Frost et al.[20])

the case for the sandy loam soil (organic matter, 5.7%) where recoveries for the aged sample (52 weeks) were 0.07, 0.034, 0.035 and 0.093 by Soxhlet, SFE, MAE and ASE, respectively (Figure 7). It was concluded that ASE was unaffected by analyte–matrix interactions, providing good recoveries for all sample types.

4.1.2.5 Phenols The continued interest by USEPA in new sample preparation techniques has focused on a wide range of compounds including phenols.[21] In this work, 16 phenolic compounds were extracted (30 mL acetone : hexane (1 : 1 v/v), at a temperature of 115 °C for 10 min) from freshly spiked clay soil, topsoil, sand, organic compost and topsoil with 5% humic acid. The recoveries of the phenolic compounds, spiked at the $50\,\mu g\,g^{-1}$ level, were determined. It was found that the majority of the samples analyzed gave recoveries above 70%. It was however noted, that 2,4-dinitrophenol (CAS 51-28-5) and 4,6-dinitro-2-methylphenol (CAS 534-52-1) may have degraded or become irreversibly adsorbed to the soil-type matrices. The method was then applied to the extraction of six phenols from a certified soil (ERA soil, lot number 330). The results were compared with Soxhlet and sonication extraction. The average recoveries for the six phenols ranged between 71%, 52% and 57% for MAE, Soxhlet and sonication, respectively. The most precise results were also obtained by MAE.

Llompart et al.[22] have utilized a central composite design for the MAE of phenol (CAS 108-95-2) and methylphenol isomers from spiked soils that have been aged for 25 days to simulate weathering processes. The following MAE operating variables were optimized: temperature (70–130 °C); solvent type (acetone–hexane mixtures); and solvent volume (15–50 mL). The extraction time was maintained at 10 min and the amount of sample was fixed at 5 g. At the 95% confidence interval, temperature was significant for all four phenols (phenol, *o*-cresol (CAS 95-48-7), *m*-cresol (CAS 108-39-4) and *p*-cresol (CAS 106-44-5)), while the acetone : hexane ratio and solvent volume were found to be significant in some cases only. Quantitative recoveries from the spiked and aged soil samples, were obtained using 10 mL of acetone : hexane (80 : 20 v/v) at 130 °C. Typical recoveries ranged from 94.5–104.4% with an RSD of 6.1–9.9%. The same group[23] developed a simultaneous MAE-derivatization procedure for phenol and methylphenol isomers. Samples (aged spiked and real soils) were irradiated in a closed vessel while immersed in hexane that had been previously doped with acetic anhydride and pyridine. The procedure was optimized using a factorial design. The developed method was applied to a soil sample from a coke plant known to be highly contaminated. The results obtained by this approach were as follows: $1.5\,\mu g\,g^{-1}$ for *o*- and *p*-cresol, $4.4\,\mu g\,g^{-1}$ for *m*-cresol and $7\,\mu g\,g^{-1}$ for phenol; results obtained using a sonic probe were approximately half the values reported.

4.2 Extraction from Aqueous Matrices

4.2.1 Direct Extraction

A novel approach for the microwave oven has been its use to extract chlorinated benzenes directly from water samples.[24] An aqueous sample was placed in a screw-capped container located in the microwave oven. Through the cap of the container were inserted a helium purge line and an exhaust line. The exhaust line led to a trap which was cooled in an ice bath. This dynamic extraction procedure was completed as follows: aqueous sample (1 L) containing 20 g of salt was purged with helium for 20 min whilst operating the microwave oven at full power for 7 min. The extract was collected in the cooled trap. Excess water was removed by passing the extract through a microcolumn of anhydrous sodium sulfate; the column was eluted with 2×1 mL of hexane and concentrated under a stream of nitrogen to 2 mL. Further cleanup was necessary using a Florisil microcolumn prior to analysis. This approach was compared with the traditional approach of liquid–liquid extraction and found to compare favorably at the nanogram per liter level. Average recoveries for 10 chlorinated benzenes spiked into water at the $1\,ng\,L^{-1}$ level were 63% by a microwave-assisted process (MAP) and 69% by liquid–liquid extraction, while at the $10\,ng\,L^{-1}$ level the average recoveries increased to 93% by MAP and 98% by liquid–liquid extraction. The same group[25] applied the approach to the extraction of PCBs from 500 mL of water. Recoveries of the spiked congeners, at two levels 85.5 and $175\,pg\,L^{-1}$, in organic-free water were in the range 68–85%. The technique was then applied to Hamilton Harbour water and the results were compared with liquid–liquid extraction. Average recoveries ($n = 7$) of PCB 209 by MAP were 90.9% compared with 80.6% by liquid–liquid extraction at a concentration of approximately $20\,pg\,L^{-1}$.

4.2.2 Solid-phase Extraction/Microwave Solvent Elution

MAE has been combined with SPE for the extraction of pollutants from aqueous samples.[26] The procedure is as follows: analytes of interest are retained on a C_{18} membrane disk (Empore), the disk is then rolled up and transferred into the closed PTFE-lined vessels of the MAE. Elution is achieved by placing organic solvent into the extraction vessel and irradiating the disk. The approach has been applied to a range of compounds, including OCPs, PCBs, PAHs, phthalate esters, organophosphorus pesticides (OPPs), fungicides, herbicides and insecticides, in aqueous solution.[26] The operating variables have been optimized. The variables considered were: elution solvent (acetone and DCM); temperature (80, 100 and 120 °C); and extraction time (1, 3, 5 and 10 mins) at 50% power. The optimum conditions for microwave elution were acetone at a temperature of 100 °C for 7 min. Excellent results compared to liquid–liquid extraction, were obtained for PAHs and phthalates (spiked at 2 μg L^{-1}), OCPs (spiked at levels between 0.1–0.2 μg L^{-1}), PCBs (spiked at 0.5 μg L^{-1}) and OPPs, fungicides, insecticides and herbicides (spiked at levels between 1–2 μg L^{-1}) in either reagent water or seawater.

4.2.3 Gas-phase Microwave-assisted Extraction

The liberation of volatile organic compounds (VOCs) from water samples by microwave energy has been described by Pare et al.[27] The aqueous sample is heated via a microwave oven, vaporizing VOCs into the headspace of the sample. By utilizing a conventional headspace sampler the VOCs are introduced directly into GC with FID. This approach was compared with a conventional 30 min static headspace sampling apparatus and the results obtained were favorable. It was noted, however, that the microwave approach gave higher detector responses, with better precision and in a shorter timescale than the conventional approach. The approach was subsequently extended to a range of volatile compounds including simple aromatics, i.e. benzene (CAS 71-43-2), toluene (CAS 108-88-3), ethylbenzene (CAS 100-41-4), *o*-, *m*- and *p*-xylene (CAS 1330-20-7), propylbenzene (CAS 103-65-1) and butylbenzene (CAS 104-51-8), as well as alkanes, i.e. C_8, C_{10}, C_{12} and C_{13}.[28] In each case, a sensitivity enhancement was noted for the microwave approach (average enhancement factor was 4.4).

5 CONCLUSIONS

The use and application of MAE continues to expand. However, as the reader will be aware after reading this article, the diversity of operating conditions and solvent choice suggested by others, merely adds to the difficulty in accepting MAE in the laboratory. To assist the reader in this choice Table 2 contains guidelines for MAE method development. As pressurized-MAE is the most common approach, details for this mode of operation only are provided. The parameters shown in Table 2 are based on the use of a microwave system capable of delivering a minimum power of 900 W. Ideally, the microwave system should be equipped with the following safety features: a fan and appropriate ducting to allow ventilation of the cavity in the event of an organic vapor release; a solvent sensor that automatically shuts off the microwave source in the event of an organic solvent leakage (minimizing the risk of fire) and extraction vessels that will automatically vent into a solvent collection system at a preset pressure (to minimize the risk of explosion).

Table 2 Recommendations for pressurized-MAE

Variable	Operating condition
Temperature	>115 °C but <145 °C.
Pressure	<200 psi NOTE: Extraction vessels should have a safety feature, e.g. rupture membrane, that prevents this pressure being exceeded.
Microwave power	100%
Extraction time (time at parameter)	>5 min but <20 min. NOTE: The longer time is recommended if 12 vessels are to be extracted simultaneously.
Extraction solvent volume	30–45 mL per 2–5 g of sample
Extraction solvent	Hexane–acetone (1 : 1 v/v) has been the most commonly used. Other solvents also appear to be satisfactory, e.g. acetone.

ABBREVIATIONS AND ACRONYMS

ASE	Accelerated Solvent Extraction
BCR	Community Bureau of Reference
BTEX	Benzene, Toluene, Ethylbenzene and Xylenes
CRM	Certified Reference Material
DCM	Dichloromethane
ELISA	Enzyme-linked Immunosorbent Assay
ERA	Environmental Research Associates
FID	Flame Ionization Detection
GC	Gas Chromatography

GC/ECD	Gas Chromatography/Electron Capture Detection
HPLC	High-performance Liquid Chromatography
MAE	Microwave-assisted Extraction
MAP	Microwave-assisted Process
MSD	Mass Selective Detection
NIST	National Institute of Science and Technology
OCP	Organochlorine Pesticide
OPP	Organophosphorus Pesticide
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PFA	Perfluoroalkoxy
p,*p*′-DDE	(1,1′-(2,2-dichloroethylenylidene)-bis(4-chlorobenzene) or 2,2-bis(4-chlorophenyl)-1,1-dichloroethylene)
psig	Pounds Per Square Inch Gauge
PTFE	Polytetrafluoroethylene
RSD	Relative Standard Deviation
SFE	Supercritical-fluid Extraction
SPE	Solid-phase Extraction
SRM	Standard Reference Material
SRS	Standard Reference Soil
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 4)**
Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Gas Chromatography **(Volume 12)**
Sample Preparation for Gas Chromatography

General Articles **(Volume 15)**
Microwave Techniques • Traceability in Analytical Chemistry

FURTHER READING

E.D. Ramsey (ed.), *Analytical Supercritical Fluid Extraction Techniques*, Kluwer Academic, London, 1998.
A.J. Handley (ed.), *Extraction Methods in Organic Analysis*, Sheffield Academic Press, Sheffield, 1999.
E.M. Thurman, M.S. Mills, *Solid-phase Extraction: Principles and Practice*, Wiley-Interscience, New York, 1998.
J. Pawliszyn, *Solid-phase Microextraction: Theory and Practice*, Wiley-VCH, New York, 1997.
J. Pawliszyn (ed.), *Applications of Solid-phase Microextraction*, Royal Society of Chemistry, Cambridge, 1999.

REFERENCES

1. J.R. Dean, *Extraction Methods for Environmental Analysis*, John Wiley & Sons, Chichester, 1998.
2. K. Ganzler, A. Salgo, K. Valko, 'Microwave Extraction. A Novel Sample Preparation Method for Chromatography', *J. Chromatogr.*, **371**, 299–306 (1986).
3. *CRC Handbook of Physics and Chemistry*, 76th edition, ed. D.R. Lide, Boca Raton, FL, 1995–96.
4. B.W. Renoe, 'Microwave Assisted Extraction', *Am. Laboratory*, August, 34–40 (1994).
5. M.H. Akhtar, L.G. Croteau, 'Extraction of Salinomycin from Finished Layers Ration by Microwave Solvent Extraction followed by Liquid Chromatography', *Analyst*, **121**, 803–806 (1996).
6. M. Letellier, H. Budzinski, P. Garrigues, S. Wise, 'Focused Microwave-assisted Extraction of Polycyclic Aromatic Hydrocarbons in Open Cell from Reference Materials (Sediment, Soil, Air Particulates)', *Spectroscopy*, **13**, 71–80 (1996/1997).
7. N. Saim, J.R. Dean, Md. P. Abdullah, Z. Zakaria, 'Extraction of Polycyclic Aromatic Hydrocarbons from Contaminated Soil using Soxhlet Extraction, Pressurized and Atmospheric Microwave-assisted Extraction, Supercritical Fluid Extraction and Accelerated Solvent Extraction', *J. Chromatogr.*, **791**, 361–366 (1997).
8. V. Lopez-Avila, R. Young, W.F. Beckert, 'Microwave-assisted Extraction of Organic Compounds from Standard Reference Soils and Sediments', *Anal. Chem.*, **66**(7), 1097–1106 (1994).
9. V. Lopez-Avila, R. Young, J. Benedicto, P. Ho, R. Kim, W.F. Beckert, 'Extraction of Organic Pollutants from Solid Samples using Microwave Energy', *Anal. Chem.*, **67**(13), 2096–2102 (1995).
10. V. Lopez-Avila, R. Young, N. Teplitsky, 'Microwave-assisted Extraction as an Alternative to Soxhlet, Sonication and Supercritical fluid Extraction', *J. AOAC Int.*, **79**(1), 142–156 (1996).
11. I.J. Barnabas, J.R. Dean, I.A. Fowlis, S.P. Owen, 'Extraction of Polycyclic Aromatic Hydrocarbons from Highly Contaminated Soils using Microwave Energy', *Analyst*, **120**, 1897–1904 (1995).

12. J.R. Dean, I.J. Barnabas, I.A. Fowlis, 'Extraction of Polycyclic Aromatic Hydrocarbons from Highly Contaminated Soils: A Comparison between Soxhlet, Microwave and Supercritical Fluid Extraction', *Anal. Proc.*, **32**, 305–308 (1995).
13. J.R. Dean, 'Accelerated Solvent Extraction of Polycyclic Aromatic Hydrocarbons from Contaminated Soil', *Anal. Commun.*, **33**, 191–192 (1996).
14. K. Li, J.M.R. Belanger, M.P. Llompart, R.D. Turpin, R. Singhvi, J.R.J. Pare, 'Evaluation of Rapid Solid Sample Extraction using the Microwave-assisted Process (MAP) under Closed-vessel Conditions', *Spectroscopy*, **13**, 1–13 (1996/1997).
15. A. Pastor, E. Vazquez, R. Ciscar, M. de la Guardia, 'Efficiency of the Microwave-assisted Extraction of Hydrocarbons and Pesticides from Sediments', *Anal. Chim. Acta*, **344**, 241–249 (1997).
16. K.K. Chee, M.K. Wong, H.K. Lee, 'Optimization of the Microwave-assisted Solvent Extraction of Petroleum Hydrocarbons in Marine Sediment and Soil', *Environ. Monitoring and Assessment*, **44**, 587–603 (1997).
17. V. Lopez-Avila, J. Benedicto, C. Chatan, R. Young, W.F. Beckert, 'Determination of PCBs in Soils/Sediments by Microwave-assisted Extraction and GC/ECD or ELISA', *Environ. Sci. Technol.*, **29**, 2709–2712 (1995).
18. F.I. Onuska, K.A. Terry, 'Extraction of Pesticides from Sediments using a Microwave Technique', *Chromatographia*, **36**, 191–194 (1993).
19. K. Hummert, W. Vetter, B. Luckas, 'Fast and Effective Sample Preparation for Determination of Organochlorine Compounds in Fatty Tissue of Marine Mammals using Microwave Extraction', *Chromatographia*, **42**, 300–304 (1996).
20. S.P. Frost, J.R. Dean, K.P. Evans, K. Harradine, C. Cary, M.H.I. Comber, 'Extraction of Hexaconazole from Weathered Soils: A Comparison between Soxhlet Extraction, Microwave-assisted Extraction, Supercritical Fluid Extraction and Accelerated Solvent Extraction', *Analyst*, **122**, 895–898 (1997).
21. V. Lopez-Avila, R. Young, R. Kim, W.F. Beckert, 'Accelerated Extraction of Organic Pollutants using Microwave Energy', *J. Chromatogr. Sci.*, **33**, 481–484 (1995).
22. M.P. Llompart, R.A. Lorenzo, R. Cela, J.R. Jocelyn Pare, 'Optimization of a Microwave-assisted Extraction Method for Phenol and Methylphenol Isomers in Soil Samples using a Central Composite Design', *Analyst*, **122**, 133–137 (1997).
23. M.P. Llompart, R.A. Lorenzo, R. Cela, J.R. Jocelyn Pare, J.M.R. Belanger, K. Li, 'Phenol and Methylphenol Isomers Determination in Soils by in-situ Microwave-assisted Extraction and Derivatisation', *J. Chromatogr.*, **757**, 153–164 (1997).
24. F.I. Onuska, K.A. Terry, 'Microwave Extraction in Analytical Chemistry of Pollutants: Chlorinated Benzenes', *J. Microcolumn Sep.*, **7**(4), 319–326 (1995).
25. F.I. Onuska, K.A. Terry, 'Microwave Extraction in Analytical Chemistry of Pollutants: Polychlorinated Biphenyls', *J. High Resolut. Chromatogr.*, **18**, 417–421 (1995).
26. K.K. Chee, M.K. Wong, H.K. Lee, 'Microwave-assisted Solvent Elution Technique for the Extraction of Organic Pollutants in Water', *Anal. Chim. Acta*, **330**, 217–227 (1996).
27. J.R.J. Pare, J.M.R. Belanger, K. Li, S.S. Stafford, 'Microwave-assisted Process (MAP): Application to the Headspace Analysis of VOCs in Water', *J. Microcolumn Sep.*, **7**, 37–40 (1995).
28. J.R.J. Pare, J.M.R. Belanger, D.E. Thornton, K. Li, M.P. Llompart, M. Fingas, S.A. Blenkinsopp, 'Gas-phase Extraction using the Microwave-assisted Process (MAP) for the Analysis of Dissolved Volatile Organic Compounds (VOCs) in Water', *Spectroscopy*, **13**, 23–32 (1996/1997).

Neutron Activation in Environmental Analysis

Keith Randle
University of Birmingham, Birmingham, UK

1 Introduction 1
2 Scope of Neutron Activation Analysis 2
3 Analysis of Biological Materials 3
3.1 Introduction 3
3.2 Sampling 4
3.3 Standard Reference Materials 4
3.4 Irradiations 4
3.5 Radiochemical Separations 5
3.6 Applications 6
4 Analysis of Water 7
4.1 Introduction 7
4.2 Sampling 8
4.3 Preconcentration Procedures 8
4.4 Postirradiation Procedures: Radiochemical Neutron Activation Analysis 10
4.5 Biomonitoring and Hydrological Studies Using Neutron Activation Analysis 11
5 Analysis of Air Particulates 11
5.1 Introduction 11
5.2 Particulate Matter: Sources, Composition and Inhalation 11
5.3 Sampling 12
5.4 Measurements 13
5.5 Applications 14
6 Analysis of Soil, Sediments and Coal 15
6.1 Soil 15
6.2 Sediments 15
6.3 Coal 16
7 Role of Neutron Activation Analysis in Speciation Studies 17
7.1 Introduction 17
7.2 Speciation of Elements in Water 18
7.3 Speciation of Elements in the Atmosphere 18
Abbreviations and Acronyms 18
Related Articles 19
References 19

After a brief historical survey of the role of neutron activation analysis (NAA) in environmental analysis there follows a section on the main types of NAA and their application to environmental problems and a table of ideal sensitivities is given. The rest of the article deals with NAA as it is applied to the main divisions of the environment. Each of these sections includes a short summary of the sampling requirements suitable for NAA. Section 3 deals with its application to biological samples and includes a brief discussion on standard reference materials (SRMs) in NAA. Applications to various types of biological material are discussed, ending with a survey of biomonitoring using NAA. The next section (the longest) is the analysis of water by NAA. This is mainly based on the preconcentration and postirradiation procedures currently used for the major types of water (seawater, surface water and underground water). A table of element concentrations in seawater and surface water illustrates the low limits of detection required to determine many trace elements by instrumental neutron activation analysis (INAA). Applications are discussed in the main body of the text. There then follows a section on the use of NAA in air particulate analysis for trace elements. The importance of PM10s (particles with aerodynamic diameter less than 10 μm) to human health and the reasons for concentrating on the analysis of these particles by NAA is discussed. Having discussed the main areas of NAA used there follows an application section. Large-scale biomonitoring procedures are outlined, particularly those involving plant materials. The final section on major environmental divisions covers the use of NAA in analysis of soils and sediments together with the analysis of coal. This latter topic also includes on-line analysis of coal using isotope sources. Finally, there is a discussion of speciation and the function of NAA in environmental speciation, particularly atmospheric and aqueous speciation analysis.

1 INTRODUCTION

NAA is usually said to have begun with the work of Hevesy and Levi[1,2] on the analysis of the lanthanides, although it was only in their second paper that they described the determination of an element (Eu) by NAA. Having only weak neutron sources meant that the method remained essentially a curiosity. Virtually no further developments of the technique occurred until after the end of the Second World War and the availability of far higher neutron intensities from newly constructed nuclear reactors. These allowed analyses of samples at a sensitivity comparable to, or even better than, most other analytical techniques then available. This fact, combined with its inherent multielement capability, capacity to

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

analyze a large number of samples, ease of sample preparation and freedom from contamination following irradiation meant that it had many distinct advantages over comparable techniques such as atomic absorption spectroscopy (AAS), flame emission spectroscopy (FES), mass spectrometry (MS) and polarography. Despite some disadvantages, such as the availability of a large expensive reactor facility and its relative insensitivity for some elements of importance in toxicology and pollution it rapidly became the choice for multielement analysis.

Even by 1968, however, Meinke commented that NAA was not a panacea for all analytical problems and that it was best considered as only one of a number of analytical techniques which can complement each other. Thus, the best way of viewing the solution to an analytical problem is to regard the various techniques available as complementary and not competitive. No one technique (not even ICPMS (Inductively Coupled Plasma Mass Spectrometry)) is ''best''. Centers that provide a variety of analytical techniques are probably the ones that will provide the most accurate and wide-ranging analyses, particularly of environmental samples. On the other hand, the development of ICPMS systems with the provision for direct analysis of solids (using electrothermal evaporation, etc.) might see this technique assuming a dominant position.

Since the late 1970s a number of significant improvements have occurred, such as the routine use of epithermal neutron activation analysis (ENAA) and cyclic NAA, which have overcome some of the aforementioned problems for some important elements. NAA, however, is now regarded as a ''mature'' technique, i.e. one which has reached a plateau as far as new developments in the techniques expected in the future.

Several books and reviews on NAA have been published detailing the technique and its applications but few deal exclusively, or even substantially, with environmental applications. Probably the best of these is the two-volume book edited by Alfassi[3] which covers the techniques and methodology in volume I and environmental applications in volume II. An older but still very useful book which deals more with radioanalytical NAA in environmental analysis is that by Das et al.[4] as well as a similar book by Tolgyessy and Kyrs.[5] There is a review article by Eisenbud[6] of the whole field and one by Ehmann and Vance[7] covers advances in NAA to 1989, including environmental analysis. Other good review articles have appeared on the role of NAA in air particulate analysis.[8,9] Other areas of environmental research have not been so well reviewed. Starting in 1986, Ehmann et al. periodically published literature surveys covering all fields of NAA, including applications in environmental analysis, until 1994.[10–14] Yates published a similar survey in 1997.[15]

2 SCOPE OF NEUTRON ACTIVATION ANALYSIS

The actual technique and instrumentation used in NAA has been discussed in articles **Chemical Analysis by Nuclear Methods: Introduction**, **Instrumental Neutron Activation Analysis** and **Nuclear Detection Methods and Instrumentation**. NAA is usually subdivided into (1) INAA, which is nondestructive and is purely instrumental and (2) radiochemical neutron activation analysis (RNAA), which involves dissolution of the sample and some more or less complex chemical separations, usually postirradiation. RNAA produces the ultimate in sensitivity limits for the method and is essential for determining certain elements such as P and Tl in any matrix, or particular elements in specific matrices, owing to inherent interference in the γ-ray spectrum. The penalty to be paid for this improvement in sensitivity and precision is that of complexity and time for analysis. INAA is a much simpler technique and geared to processing quite large numbers of samples in a relatively automated way while still retaining good accuracy and precision. It is, therefore, the method usually chosen, if possible. Probably, still the most widely used method of NAA is the comparator method, using several single element standards or a single, multielement standard. The development of single element comparator methods (especially the k_0 method), has reduced analysis time and improved accuracy.

INAA can be further subdivided into thermal (and epithermal) NAA and fast neutron activation analysis (FNAA). Thermal NAA takes advantage of the high intensity of neutrons available from the thermalization of fission neutrons in a nuclear reactor and the large thermal neutron cross-sections for most isotopes. This makes the technique suitable for trace element determinations which is its major application in chemical analysis. INAA usually refers to its use with reactor thermal neutrons. Sensitivities for element determinations by INAA are matrix dependent but some calculated values for interference-free sensitivities are given in Table 1.

An additional factor in environmental research is the fact that the matrix elements C, H, N and O produce no significant activity and so virtually no interference problems. Other matrix elements such as Mg, Al, Si and Ca give rise to fairly short-lived activities that may only interfere with a few short-lived trace element radioisotopes.

FNAA usually refers to irradiations with neutrons with the specific energy of 14.7 MeV. FNAA is not usually regarded as a trace element technique owing to the much lower fluxes generally available and the generally much smaller reaction cross-sections. The major exception to this rule is for the determination of oxygen

Table 1 Calculated best INAA detection sensitivities for some environmentally important elements with γ-emitting nuclides, in the absence of any interfering activities[a]

Detection sensitivity[b] (μg)	Element (nuclide used, γ-ray energy used)[c] (keV)
$1-3 \times 10^{-7}$	Eu (^{152m}Eu, 122), Dy (^{165}Dy, 95), In (^{116m}In, 417)
$4-9 \times 10^{-7}$	Ho (^{166}Ho, 81)
$1-3 \times 10^{-6}$	Mn (^{56}Mn, 847), Au (^{198}Au, 412), Sm (^{153}Sm, 103)
$4-9 \times 10^{-6}$	Lu (^{176m}Lu, 88), Rh (^{104m}Rh, 51), Re (^{188}Re, 155), Ir (^{194}Ir, 328)
$1-3 \times 10^{-5}$	U (^{239}U, 75), Co (^{60m}Co, 59), Hg (^{197}Hg, 69-78), Cs (^{134m}Cs, 127), Er (^{171}Er, 308), As (^{76}As, 559), Ga (^{72}Ga, 834), I(^{128}I, 443), La (^{140}La, 487), W (^{187}W, 72), Cu (^{64}Cu, 511)
$4-9 \times 10^{-5}$	Pd (^{109}Pd, 88), Br (^{82}Br, 554), Yb (^{177}Yb, 150), V (^{52}V, 1434), Na (^{24}Na, 1368), Th (^{233}Th, 87), Ru (^{105}Ru, 724), Sb (^{122}Sb, 564)
$1-3 \times 10^{-4}$	Ba (^{139}Ba, 166), Sr (^{87m}Sr, 389), Nd (^{149}Nd, 211), Te (^{131}Te, 160), Sc (^{46}Sc, 889), Ge (^{75}Ge, 265), Ta (^{182}Ta, 68)
$4-9 \times 10^{-4}$	Hf (^{180m}Hf, 215), Pt (^{197}Pt, 78), Tb (^{160}Tb, 299), Se (^{77m}Se, 162), Gd (^{159}Gd, 364), Cd (^{111m}Cd, 245), Tm (^{170}Tm, 84), Cl (^{38}Cl, 1642)
$1-3 \times 10^{-3}$	Ag (^{110}Ag, 658), Sn (^{123m}Sn, 160), Zn (^{69m}Zn, 439), Ce (^{143}Ce, 293), Al (^{28}Al, 1779), Os (^{191}Os, 129), Mo (^{101}Mo, 192)
$4-9 \times 10^{-3}$	Pr (^{142}Pr, 1576), Ti (^{51}Ti, 320), K (^{42}K, 1525), Y (^{90m}Y, 202), Rb (^{86m}Rb, 556), Ni (^{65}Ni, 367), Cr (^{51}Cr, 320)
$1-3 \times 10^{-2}$	Mg (^{27}Mg, 844)
$4-9 \times 10^{-2}$	Zr (^{97}Zr, 747), Nb (^{94m}Nb, 871)
0.1–0.3	Ca (^{49}Ca, 3083), F(^{20}F, 1633)
1–3	Fe (^{59}Fe, 1099)
4–9	Si (^{31}Si, 1266)
10–30	Pb (^{207m}Pb, 570), S (^{37}S, 3102)

[a] Data from V.P. Guinn, J. Hoste, in *Elemental Analysis of Biological Materials*, IAEA Technical Report Series No. 197, IAEA, Vienna, Chapter 7, 1980.
[b] Sensitivities based on the following parameters: $\phi_{th} = 10^{17}\,nm^{-2}\,s^{-1}$, irradiation time = 5 h max., decay time = 0, count time = 100 min max., 40 cm^3 Ge (Li) detector with nuclide 2 cm distance.
[c] Elements are listed in a given group in decreasing order of sensitivity.

where FNAA is probably the most sensitive technique available.

ENAA is a useful extension of INAA in that it enhances the activation of a number of trace elements relative to the major matrix elements. Here, epithermal is taken to be neutrons with energies from the Cd cut-off of 0.55 eV up to approximately 1 MeV. In general, the activation cross-sections of the matrix elements of environmental samples are inversely proportional to the neutron energy (1/v law). The trace elements also follow this general trend but many of them have large activation cross-sections at specific energies in the epithermal energy region. Thus, in irradiations with mainly epithermal neutrons the relevant trace elements are well activated and the resultant γ-ray peaks are usually more distinct and so more precisely analyzed than when using thermal NAA, even though the peaks may actually contain less counts. ENAA is particularly useful for biological and seawater samples which contain high concentrations of Na and Cl, which follow the 1/v law.

Finally, a number of more specialized techniques can be used to enhance the sensitivity of INAA. One of these is cyclic instrumental neutron activation analysis (CINAA), which is used to improve the measurement of some very short-lived isotopes (e.g. Pb, which can be activated to ^{207m}Pb with 0.8 s half life) by repeated irradiation and counting. CINAA has also been used to improve the determination of Cu, F, Pb and Se amongst others, in biological samples.

A special form of INAA results from the γ-rays emitted immediately following the capture reaction. These are known as prompt gamma rays and, hence, the method is called prompt gamma neutron activation analysis (PGNAA). Here, irradiation and counting have to be undertaken simultaneously but in principle, all elements are determinable by PGNAA since all elements capture neutrons but detection limits (DLs) vary from ~0.1 ppm up to tens of ppm depending on the element and the available neutron flux. Chau et al.[16] have discussed the wider application of PGNAA to environmental and biological materials.

3 ANALYSIS OF BIOLOGICAL MATERIALS

3.1 Introduction

The determination of the minor and trace elements in biological materials can have a number of objectives. One

of these is to estimate the normal or baseline concentrations of elements so that any significant deviations from these levels can perhaps be related to specific changes or diseases in the organism or tissue. A second objective is to determine those elements deemed essential or toxic to an organism and at what levels they function in these roles. Thus, in humans some 18 elements (Ca, Cl, Co, Cr, Cu, F, Fe, I, K, Mg, Mn, Mo, Na, Ni, P, S, Se and Zn) are now thought to be essential for the correct metabolism of an individual, ranging from a daily requirement of 2–5 g for K down to 3 μg for Co. It is well known, however, that almost any substance taken in excess will ultimately become harmful, i.e. toxic. In other words, there is no essential qualitative difference between essential and toxic elements. A third objective of trace element determinations is to ascertain if the concentration of an element(s) in a particular tissue or component can be an indicator of a specific disease. An example is the elevated concentration of Cu in the livers of people suffering from Wilson's disease.

3.2 Sampling

In environmental monitoring the quality of the data is related to three principal factors: (1) the design and performance of a representative sampling program, (2) the satisfactory storage of the sample to prevent changes to the analyte and (3) the use of suitable analytical methods which yield the requisite accuracy and precision.

Sampling is a problem for all types of material but sampling of biological materials is particularly difficult. This is in part due to the very complexity of the components of biological samples but also in part to the often limited amounts available and to potential changes to the constituents following sampling, e.g. coagulation of milk and blood or general bacterial attack. Thus, sampling may be the key factor in the quality of the data for some biological analyses.

3.2.1 Methods of Sampling and Sample Preparation

An important consideration in the collection of vegetation for any type of analysis is that the sample variation within a collection area probably greatly exceeds all other analytical variations. This necessitates careful consideration being given to the collection strategy and will often require proper statistical design for sample collection if meaningful results are to be obtained. Such considerations are beyond the scope of this article but are treated by, for example, Barnett[17] and Webster.[18] Vegetation in this context could include grasses, vegetables of all sorts, leaves of trees or needles from conifers and fresh water and marine plants, particularly seaweed.

Sampling tissues, organs, and so on is a specialized job and normally the analyst will be supplied with the necessary material from a suitable laboratory. Urine and faecal samples may be required from both humans and animals, and suitable arrangements for their collection are required.

3.3 Standard Reference Materials

In order to verify the quality of the data from INAA, the availability of SRMs with certifiable (trace) element concentrations is important. They may also be used as standards in some circumstances in the comparator method. Some 200 SRMs are now currently available, covering virtually all the matrices met within the analysis of biological materials. The International Atomic Energy Agency (IAEA) in Vienna has issued a comprehensive list of these SRMs and their suppliers.[19] Such materials conform to the basic requirements of certifiable element concentrations, homogeneity and long-term stability. Byrne[20] has given a useful survey of the role of SRMs in environmental monitoring.

3.4 Irradiations

Biological samples are normally irradiated in acid-cleaned polythene vials for irradiations up to a day or so. For much longer irradiations, quartz ampoules are normally used. These must be cleaned in boiling aqua regia followed by thorough rinsing in doubly distilled water prior to irradiation. A blank irradiation, with no sample present is usually necessary owing to the presence of trace impurities such as K, Na, Sb and Sc in the silica. A suitable standard(s) is used in the comparator method, although Byrne and Dermelj[21] have described a method based on the use of an endogenous internal standard of already known concentration for the analysis of biological samples which eliminates the need to know the sample mass and greatly reduces errors due to geometrical differences, and in RNAA it eliminates the necessity of determining chemical yields.

Owing to the generally low concentrations of nonmatrix elements in biological samples coupled with the fact that the matrix elements (C, H, N, O and P) do not produce γ-emitting nuclides, biological samples are often irradiated for long periods to enhance the activities of the analytes. The problem of bremsstrahlung from some of the induced β-emitters (e.g. ^{32}P) is reduced by placing a plastic absorber between the samples and the detector. Using a flux of approximately $10^{17}\,nm^{-2}\,s^{-1}$ the sensitivity for many elements is inferior to that of many other analytical procedures and is frequently slower. Even if fluxes greater than $10^{17}\,nm^{-2}\,s^{-1}$ were available, they could rarely be used for biological samples owing to the radiation damage that they would cause.

In INAA of biological materials, the normal procedure is to use a short irradiation of about 30 min to determine short-lived nuclides such as ^{27}Al, ^{27}Mg, ^{47}Ca, ^{52}V, ^{56}Mn, ^{80}Br, ^{86m}Rb, ^{101}Mo and ^{128}I, together with ^{24}Na, ^{42}K and ^{82}Br. Very short-lived isotopes such as ^{38m}Cl(710 ms), ^{79m}Br(4.8 s), ^{20}F(11.1 s) and ^{77m}Se(18 s) can also be determined using a pneumatic transfer system in the technique of short-time NAA. Here, CINAA may be effectively used to increase the sensitivity of the determination. Short-time NAA allows the determination of some elements, e.g. Pb and Se, that would otherwise require a postirradiation chemical separation. These short irradiations are then followed by a long irradiation of 8–48 h duration for the determination of ^{51}Cr, ^{59}Fe, ^{60}Co, ^{65}Zn, ^{76}As, ^{86}Rb, ^{121}Sb, ^{137}Cs, ^{133}Ba, lanthanides, ^{182}Ta, ^{233}Th and ^{233}Pa(U). INAA is very useful for samples that are difficult to obtain, such as biopsy samples. Because of the essentially nondestructive nature of the technique, the samples are available for repeated determinations or analyses by other methods.

The high activities of ^{24}Na, ^{38}Cl, ^{56}Mn and ^{82}Br induced in most biological samples mean that many trace elements can only be determined after pre- or postirradiation chemical separation to remove these interfering nuclides or alternatively, by employing ENAA since the interfering nuclides follow the 1/v law. ENAA is particularly useful for the determination of Ag, As, Br, Cd, Cs, I, Mo, Rb, Sb, Se, Sr and Zn. Silicon can be determined by ENAA in biological materials but the corrections for the interfering ^{31}P(n,α)^{28}Al and ^{27}Al(n,γ)^{28}Al reactions are so large for fruit, legumes, meat and milk as to make the method inapplicable. Both Cd and B compounds have been used as thermal neutron absorbers, although there can be thermal problems when using boron compounds (BN and B_4C_3). FNAA is also used as a complement to INAA and ENAA to extend the range of elements determined in biological materials, e.g. N, P, Si and Ca. It has been used particularly for the determination of nitrogen which, in turn, has been used as an estimate of the protein content of samples. The reaction used is ^{14}N(n,2n)^{13}N, giving a positron-emitting nuclide with a half-life of 9.97 min.

Some elements in biological samples are also determined by PGNAA. It is particularly useful for boron determinations using the 478-keV γ-ray from ^{11}B. The nondestructive nature of the method makes it useful for studying the uptake of boron by plants from the soil.[22]

3.5 Radiochemical Separations

Although for reasons discussed in section 1, instrumental forms of NAA are usually preferred because of the simplicity and ability to process large numbers of samples in a relatively short time and to retain the sample intact, nevertheless, for certain applications there is no escaping dissolution of the sample and chemical processing. Generally, postirradiation separations are favored because of the lack of contamination problems but occasionally preirradiation separations must be used. In either case, the procedures for the dissolution of materials may result in the loss of more volatile trace elements such as Hg, I, Sb, Se, and when dry ashing is used, even elements such as Ce, Co, Mn, Pa, Ru and Zn can be partly volatilized, even at low ashing temperatures.

3.5.1 Preirradiation Separations

These are usually performed when there is the possibility of rapid sample deterioration or in order to reduce the amount of activity to be handled after the irradiation. In either case, either the analyte is separated or the matrix elements removed. When using preirradiation separations, because there is the possibility of chemical contamination, a blank determination is necessary. The removal of sodium prior to irradiation may be required owing to its short half-life (15 h) and the intense activity that may be induced in some samples. This can be achieved by dissolving the sample, stirring in powdered hydrated antimony pentoxide (HAP) and then filtering to remove the HAP which will contain most of the sodium.

3.5.2 Postirradiation Separations

In order to improve the sensitivity of the determination of a number of elements by removal of spectral interferences, or to enable a general reduction in the background activity and elimination of any blank value, a postirradiation separation is used. An important aspect of modern RNAA is the development of standardized and automated separation systems which can handle relatively large numbers of samples and are less prone to radiation hazards, although these tend to be laboratory specific. It has been found to be more effective to have a number of separate separation procedures for a number of small sets of elements rather than a single complicated procedure for all the elements, usually at the expense of high and constant yield.[23]

The majority of postirradiation separations use either ion-exchange chromatography or solvent extraction but precipitation is also used. Here the addition of carriers is essential for all the analyte elements as well as major interfering elements. Platinum has been determined by electrodeposition from acid solution using a cell with a niobium cathode. A DL of about 1 ppb has been achieved with this technique.

A more recent development has been the introduction of new ion-exchange materials, particularly inorganic exchangers. An interesting example is analysis for Li which can be determined in biological materials by

using the following reactions: $^{6}Li(n,\alpha)T$, $^{16}O(t,n)^{18}F$. The sample solution is placed on a ZrO_2 column and the ^{18}F absorbed. Interfering nuclides are eluted off with pH 1 to ~3 solution and the column then counted to determine the ^{18}F activity.[24] These developments have been reviewed by Braun et al.[25]

One of the more rapid methods for the separation of trace elements following dissolution of the sample is by adsorption on activated carbon (AC).[26] The solution is simply filtered through a layer of AC on a membrane filter. The number and type of elements that are collected in a single filtration can be varied by the addition of certain reagents to the solution. Thus, filtration at low pH without the addition of any reagents will result in the adsorption of ^{64}Cu, ^{110m}Ag, ^{197}Hg, ^{198}Au and $^{233}Pa(Th)$. If the filtrate from this is then treated with ammonium pyrrolidine dithiocarbamate (APDC) and oxine and filtered again through fresh carbon, this results in the adsorption of ^{64}Cu, ^{99}Mo, ^{122}Sb, ^{187}W and ^{197}Hg. This can be extended to the point where some 15–25 radionuclides can be collected on different carbon fractions which can then be counted to determine the relevant nuclides.

3.6 Applications

3.6.1 Tissues, Blood and Bone

Almost all human organs and tissues have been subjected to analysis by INAA. Many of these studies have simply determined the levels of all the detectable elements. Others have been used to assess levels connected with specific diseases. Some international comparison studies have revealed the difficulties in obtaining reliable data for some elements in some tissues. Values for As, Cd, Hg and Pb obtained by a number of analytical techniques including NAA in bone, lung and brain samples are rather poor.

A number of studies have focused on the analysis of the brain by INAA with attempts in particular to try to relate instances of decreased concentrations of specific elements with Alzheimer's disease.[27] It is well known that high concentrations of Al in the brain are associated with the disease but other trace elements may play a role.

Blood has been widely analyzed by INAA, again with much of the emphasis on the concentration range of trace elements found in blood or blood components.[28] Some efforts have been made to correlate various diseases with blood composition and there is evidence for some specific changes of blood composition for some occupational groups. Also, certain elements (e.g. La and Tb) tend to increase in concentration with age.

Much of the work on bone has been directed primarily at determining the fluoride content of the bone, due to the potential role of fluoride in the development of bone diseases. This has also led to investigations of changes of other elements in bone such as Al, Ca, Mg, Na and P from various diseases or forms of treatment. Al, Mg and Na are all best determined by INAA but FNAA is the more sensitive method for Ca, F and P. If lower energy neutrons (<10 MeV) are used, the most sensitive $^{19}F(n,\alpha)^{16}N$ reaction is effectively employed.[29] An interesting application of bone analysis is the correlation of the N/F ratio in fossil bones with the age of the bone, F increasing with age as it replaces the P in apatite.[30]

3.6.2 Biomonitoring

An increasing trend in environmental analysis is that of biomonitoring. Here, some living system (usually a plant) which incorporates and concentrates a particular pollutant is sampled and used to estimate the total amount of the pollutant over a wide area or to estimate the effect of pollutants on the local ecosystem. In this way the biomonitor records environmental changes and can be used to follow the effect of changed pollutant conditions. Alternatively, from determining the concentration of a particular pollutant in specific tissues or body fluids in human beings, biomonitoring can be used to estimate better the total dose absorbed and to assess the risk of environmental or occupational exposure to the pollutant.

Since the late 1960s there have been numerous studies of the trace element composition of human hair (and to a lesser extent, nails) with the aim of estimating patterns of composition for normal populations in a given region or geographical area and then using these ''normal'' patterns to reveal significant variations which may indicate exposure to high levels of a pollutant or possible correlations with certain diseases. The IAEA has coordinated international programs of nuclear-based research on hair analysis.[31,32] The later IAEA program[32] was directed towards collection of hair only from males of a certain age who died by accident or were subject to sudden unexpected death.

The data from the first IAEA program provided data on 40 elements. It showed that use of geometric means and standard geometric deviations gave a much better characterization of the distribution of trace elements in most cases than the equivalent arithmetic ones. Also, statistical variations of element concentrations could be quite large without being ''abnormal''. Studies have also revealed geographical differences in some elements, e.g. Se higher in Asian than European populations.

More frequently, however, the interest is in how the concentration of elements in hair reflects internal contamination or burden of an element(s). An example is that of peasants who consumed grain contaminated with a mercury fungicide.[31] Many of these ingested toxic

quantities and variations along the length of hair recorded the history of the poisoning.

3.6.3 Plants

The analysis of trace elements in plants has proved a fruitful field for both RNAA and INAA. Again, much of the work has concentrated on determining the range of trace elements found both in whole plants and in different plant structures such as seeds. Essentially the same 15–20 elements found in biological materials are also determined in plants by INAA.[33] RNAA then yields a further 10–12 elements, among them Ag, Au, Cd, Cu, Hg, Mo, Sr and Zn. FNAA and ENAA also extend the range of elements determined, including nitrogen which is useful for determining the protein content of plants.

4 ANALYSIS OF WATER

4.1 Introduction

The analysis of water for trace elements and, in particular, trace metals is one of the most important aspects of environmental monitoring. This results from the role of water in all aspects of life on our planet, ranging from the purity of drinking water to the pollution levels in the seas and oceans. At the same time the seas provide a valuable source of some heavy precious metals. In addition, the data obtained can be used to elucidate transport processes in the aquatic environment.

Water can be broadly segregated into sea, surface (lakes, rivers, drinking, etc.), ground and waste. Apart from the latter, levels of most elements in water are at the trace or ultratrace level (see Table 2). The exceptions to this general rule are the concentrations of Na, Cl and Br and one or two other elements, particularly in seawater. Levels of certain elements in wastewater can also be high but vary according to the source of this water.

In one respect, water presents the perfect matrix for determinations of trace elements by NAA since it is essentially not activated. On the other hand, the low concentrations of most elements would require the use of large sample volumes and long irradiations with high neutron fluxes for measurements by INAA. These requirements present several problems: first, there is a

Table 2 Concentration of elements in seawater and fresh surface waters determinable by INAA

Element	Seawater[a] concentration (ppm)	Fresh surface waters[b] mean concentration (range) (ppm)	Element	Seawater concentration (ppm)	Fresh surface waters mean concentration (range) (ppm)
Cl	1.95×10^4	7.8(4.9–12.1)	U	3.2×10^{-3}	$\sim 1.0 \times 10^{-3}$
Na	1.077×10^4	6.3(2.9–11)	V	2.5×10^{-3}	$< 1.0 \times 10^{-3}$
Mg	1.29×10^3	4.1(1.5–5.6)	Al	2.0×10^{-3}	–
S	9.05×10^2	3.7	Fe	2.0×10^{-3}	$0.67(1 \times 10^{-2} - 1.4)$
Ca	4.12×10^2	15	Ni	1.7×10^{-3}	$1.0 \times 10^{-2}(0–0.3)$
K	3.80×10^2	2.3	Ti	1×10^{-3}	$\sim 7 \times 10^{-3}(5 \times 10^{-5} - 5 \times 10^{-2})$
Br	67	~0.02	Zn	5×10^{-4}	$\sim 1.0 \times 10^{-2}(0–0.32)$
Sr	8	$\sim 0.09(3.0 \times 10^{-3} - 0.8)$	Cs	4×10^{-4}	$(5 \times 10^{-5} - 2 \times 10^{-4})$
Si	2	6.1(1.8–10.8)	Cr	3×10^{-4}	–
F	1.3	$\sim 9.0 \times 10^{-2}$	Sb	2.4×10^{-4}	–
Rb	0.12	$\sim 1.2 \times 10^{-3}(3 \times 10^{-4} - 2 \times 10^{-3})$	Mn	2×10^{-4}	$> 1 \times 10^{-2}$
I	6×10^{-2}	–	Se	2×10^{-4}	–
Ba	2×10^{-2}	$\sim 5 \times 10^{-2}(1.5 \times 10^{-2} - 0.5)$	Cd, W	1×10^{-4}	–
Mo	1×10^{-2}	–	Cu	1×10^{-4}	$\sim 8 \times 10^{-3}(0–6 \times 10^{-2})$
As	3.7×10^{-3}	$4.0 \times 10^{-4}(0–0.1)$	Ge	5×10^{-5}	–
Hg	3×10^{-5}	–	Ce	1×10^{-6}	–
Zr	3×10^{-5}	$2.6 \times 10^{-3}(5 \times 10^{-5} - 2.3 \times 10^{-2})$	Er, Yb	8×10^{-7}	
Nb, Sn, Tl, Th	1×10^{-5}	–	Gd	7×10^{-7}	–
Hf	7×10^{-6}	–	Pr, Sc	6×10^{-7}	–
Au, Re	4×10^{-6}	–	Pb	5×10^{-7}	$\leq 1 \times 10^{-2}$
Co	3×10^{-6}	$\sim 9 \times 10^{-3}(0–1.9 \times 10^{-2})$	Ho, Lu, Tm	2×10^{-7}	–
La, Nd	3×10^{-6}	–	In, Tb, Te	1×10^{-7}	–
Ag	2×10^{-6}	$\sim 9 \times 10^{-5}(1 \times 10^{-5} - 3.5 \times 10^{-3})$	Sm	5×10^{-8}	–
Ta, Ga	2×10^{-6}		Eu	1×10^{-8}	–
Y	1.3×10^{-6}	–			

[a] Data from *Seawater: Its Composition, Properties and Behaviour*, Open University Course Team, Pergamon Press, Oxford, 1989.[34]
[b] Data from D.A. Livingstone, 'Chemical Composition of Rivers and Lakes', *Data of Geochemistry*, 6th edition, US Geol. Survey Prof. Paper 440-G, Chapter G, G1–G61, 1963.[35]

severe limitation on sample size in most reactors and second, long or intense neutron irradiation of water results in the radiolysis of the water leading to problems from the subsequent gas generation. Shorter irradiation times result in unacceptably long counting times. A final factor militating against INAA of many water samples is the high activities of ^{24}Na, ^{38}Cl and ^{82}Br induced in these elements. Although these are all relatively short-lived activities, it makes handling the sample following irradiation difficult and precludes the determination of the short-lived trace elements in many cases.

Consequently, water as such is rarely irradiated (although methods have been developed to allow irradiation of large volumes of water, up to 250 cm^3 for 20 h) and the more usual approach is to concentrate the elements prior to irradiation in a solid form (preconcentration) or, more rarely, following the irradiation (postconcentration).

4.2 Sampling

Environmental sampling of liquids will usually involve liquids flowing in open systems (rivers, canals, effluent streams, etc.) or liquids in large open bodies (seas, lakes, ponds, etc.). The major difficulty in sampling flowing liquids for this type of system is that it may vary in two dimensions, i.e. both with depth and along its length. The sampling will probably have to take into account changes in flow rate, both seasonal and on a shorter timescale.

Samples are collected in wide-mouth bottles or small buckets, probably of plastic construction. If samples are required at various depths then special containers are used which can be filled at a specific depth and then closed before removal to the surface.

For water in large open bodies it will again probably be necessary to obtain samples from various parts of the body and at various depths. The samples can be taken with similar devices discussed above. Special devices have also been constructed for sampling relatively shallow lakes and ponds on a regular basis to obtain samples quickly from several fixed depths.

4.3 Preconcentration Procedures

4.3.1 Evaporation and Freeze Drying

These are both simple methods, can handle large volumes of water and are free of potential chemical contamination. They do, however, suffer from the time required for their completion and the fact that no reduction in the Br, Cl or Na content occurs. Freeze drying is preferred as even evaporation at relatively low temperatures can result in the loss of some volatile elements and possible contamination of the open system. The resultant solid that is produced can then be subject to INAA with the above limitations due to the presence of Br, Cl and Na.

4.3.2 Chemical Preconcentration

The aim here is to concentrate the trace analytes whilst separating them from potential interfering elements at the same time. In general, the most useful methods are those that concentrate a large number of trace elements, although procedures have been developed for the concentration of a small number of elements that are of particular concern or troublesome to determine by alternative concentration techniques. The most appropriate method will be one which is rapid and can actually be carried out at the sampling site, thus eliminating problems due to adsorption or desorption during storage of the sample. Examples of such methods are filtration, adsorption of the analyte during passage of the water through a column of a suitable solid, solvent extraction or the collection of a precipitate. Direct precipitation is usually not possible owing to the very low levels of most of the elements. The answer is to use coprecipitation whereby a suitable bulk chemical is precipitated and this carries the trace elements. In particular, organic coprecipitants are suitable as C, H, O and N do not form radioisotopes except for small amounts of very short-lived ^{16}N and ^{19}O. Alternatively, the trace elements may be scavenged from solution by a suitable solid, e.g. MgO and AC. A method requiring more time and involving greater complexity is ion exchange, mainly of complexed species. Rarely, electrodeposition is used for some specific elements (e.g. Hg). The enrichment factors for these techniques, however, are usually much smaller than those obtained by coprecipitation methods.

The major drawback of preconcentration is the potential contamination of the final product by the chemicals used. Hence, the emphasis is on the purity of all the chemicals used in a procedure. A second disadvantage is the necessity to determine a blank value for the elements determined, in other words, carry out the entire procedure with very pure water to determine the contamination from the procedure itself. Finally, the chemical yield for the analytes must be determined or it must be established that a given procedure produces an essentially quantitative yield. Despite these drawbacks preconcentration methods are the most widely used in the analysis of water samples by NAA.

4.3.2.1 Coprecipitation Coprecipitation may be conveniently divided into methods based on either organic or inorganic precipitants. Some more recent organic precipitants have also incorporated metals that produce only very short-lived or long-lived γ-emitting nuclides or only β-emitters to increase yields, e.g. Pb and Bi in conjunction with the precipitant pyrrolidine dithiocarbamate (PDC).

Organic Precipitates. A number of organic compounds, mainly chelates, have been used in preconcentration by coprecipitation. Thus, oxine (8-hydroxyquinoline) is used for the determination of Cd, Cu and Mn in seawater and, in the form of reversed-phase extraction, chromatography using silica-immobilized oxine has been developed for the simultaneous determination of Co, Cd, Cu, Hg, V and Zn in rain and surface waters.[36]

Dithiocarbamates, in the form of the Pb or Bi salts of PDC carry many elements when coprecipitated with them, which is carried out at pH < 6 for the best yields for most elements. APDC in the presence of Bi^{3+} carrier carries Cd, Co, Cu, Mo, V and Zn quantitatively from lake and river-water samples[37] and arsenic as As(III) has been determined in surface waters by precipitation of the dibenzyldithiocarbamate complex followed by filtration and then INAA of the solid. Dithiocarbamic acids possess an S−H group which is readily oxidized in aqueous solution to insoluble disulfides which form the precipitate. Some other organic molecules with oxidizable S−H groups have been examined as potential coprecipitants,[38] in particular quinolyl-8-thiol and its derivatives which give rise to bis(8-quinolyl) disulfides. Waters with high fulvic or humic acid concentrations result in poor or no coprecipitation of trace metals and so these organic contaminants in the water have to be destroyed prior to disulfide coprecipitation.

PAN (1-(2-Pyridylazo-2-naphthol)) when added as a dilute solution in ethanol to the water sample results in the PAN crystallizing out, whereupon the solid is filtered and collected for irradiation. In this way, excellent recoveries of Cr(III), Cu, the lanthanides, Mn(II), Ni and Zn can be obtained together with lower recoveries of Ag, Cd, Co, Fe, Hf, Hg, In, Sc, Sn, Th and U in surface and seawater.[39] Under these conditions, preconcentration factors $> 2 \times 10^5$ are possible. Combinations of these precipitants with each other or other organic complexes are also used to improve the yield or separation of various elements. Examples are bismuth salts of mixed oxine and thionalide (2-mercapto-*N*-2-naphthylacetamide) used in combination with Pb or Bi salts of (PDC) for the determination of As, Cd, Co, Cr, Cu, Fe, In, La, Mn, Mo, Ni, Sb, Sc, Se, U and Zn in tap water and seawater. A mixture consisting of APDC, Cupferron and 1-(2-thiazolylazo)-2-naphthol (TAN) can be precipitated from drinking water and seawater samples at pH 6 for the determination of Cd, Co, Cu, Hg, Mn, Th, U, V and Zn.[40] Finally, oxine is also used to determine Al, Cu, Mo, V, U and Zn in underground, lake and seawater by converting these metals to oxine chelates, extracting these with molten *o*-phenylphenol and irradiating the subsequently cooled solid.

Inorganic Precipitates. A problem with some potential precipitates is that the matrix metal, e.g. Fe, is activated requiring its removal either prior to irradiation or prior to counting. $Fe(OH)_3$ has been used to collect a number of trace inorganics from water. Arsenic in seawater can be determined by coprecipitation with $Fe(OH)_3$, with the precipitate redissolved in acid and the solution irradiated.[41] The As (^{76}As) is isolated on a column of Al_2O_3 for counting. The same precipitant can be used for the determination of several lanthanides in river and seawater following a group separation after irradiation. A combination of $Fe(OH)_3$, $Al(OH)_3$ and PbS has been used to determine a number of elements in river water, each coprecipitant removing a specific set of elements.[42] PbS coprecipitates Ag, Al, Ba, Br, Cl, Cu, I, In, K, La, Mg, Mn, Na, Sb, Sn, Sr, U, V and Zn from various potable mineral waters. Lead phosphate has also been used to coprecipitate Ag, Cd, Cr(III), Cu, Mn(II), Th(IV), U(VI) and Zr(IV) from surface and seawater.[43] Bismuth sulfide, Bi_2S_3, has been used to coprecipitate a limited number of elements, e.g. Pd in seawater. Short irradiations produce no activity from Pb or Bi and the beta activity from the sulfur and phosphorus is not a problem.

4.3.2.2 Scavenging This term as used here refers to the process whereby a finely divided solid such as carbon or MgO is added to the water sample with the aim of adsorbing trace elements present. The process may, in fact, involve physicochemical processes other than adsorption, such as ion exchange. In general, scavenging satisfies the requirements of a fast simple means of preconcentration.

By far the most widely used method of scavenging is that based on AC.[26] In its simplest form, this simply requires the water sample to be stirred with a suitable mass of AC (or for the water to be poured through a column of AC), followed by filtration to collect the AC and then irradiation of the dried solid. Ultrapure carbon results in virtually no activity since the carbon itself is not activated and only trace impurities produce γ-emitters. The levels of these trace impurities vary according to the source of the carbon and this requires a blank measurement, so that the sample data can be corrected where necessary. Preparation and irradiation of a blank are necessary for all scavenging techniques.

In practice, it is found that adsorption isotherms for almost all elements are increased by chelation of the element so that the element is adsorbed as the chelated species.[26] Commonly used chelates are APDC (Cd, Co, Cr, Cu, Mn, Mo, V and W) and L-ascorbic acid (Se, U). Bismuthiol-II (3-phenyl-5-mercapto-1,3,4-thiadiazole-2(3*H*)-thion potassium salt) has also been used to chelate Se. For Hg no chelation is used but the

process involves reduction of Hg with stannous chloride ($SnCl_2$) followed by its volatilization and subsequent collection on AC.

Magnesium oxide, MgO, was originally applied to the determination of trace elements in surface water.[44] In this process light MgO is stirred with the water for 30 min and then allowed to stand for ~8 h. This produces a stable mixture of MgO and $Mg(OH)_2$ of pH 10, which then adsorbs trace elements from the water via an ion-exchange mechanism. The filtered and dried solid can be subject to INAA since, again, Mg does not produce any long-lived activity. In this way, Ag, As, Au, Cd, Ce, Co, Cr, Fe, La, Mo, Ni, Sb, Sn, U and Zn are all absorbed. Washing the irradiated MgO with water removes interfering ^{24}Na, ^{42}K and ^{82}Br. It is also found that the lower oxidation states of some elements are much better absorbed than the higher ones [Cr(III) > Cr(VI), Sb(III) > Sb(V), Se(IV) > Se(VI)]. This provides for the possibility of separating the oxidation states of these three elements. When MgO scavenging is used for seawater, adsorption from acidified seawater is necessary in order to destroy the chloro complexes of some metals such as Hg and Ni and allow their absorption as the cationic species.

4.3.2.3 Ion Exchange The method is used not only as a preconcentration technique, but also as a means of removing many interfering elements from natural waters. A great advantage of ion exchange over other preconcentration techniques is that virtually unlimited volumes of (nonsaline) water can be used in the exchange of trace elements. After washing the column, the resin can be removed, dried and subjected to INAA. For conventional acidic cationic or basic anionic exchangers, seawater is obviously a particular problem owing to the presence of large amounts of Br, Cl and Na in particular and they are not normally used. In general, anionic exchangers are more widely used than cationic ones for the preconcentration of trace metals from surface waters, and metals determined in this way are Ag, As, Au, Cd, Co, Cu, Hg, Mn, Th, V and Zn.

In the case of seawater, the choice is usually between a chelating resin or an inorganic exchanger. Chelating resins readily take up those metals that form strong chelates with the functional group. This excludes Na to a great extent so that large volumes of seawater can be used. Such resins take up Al, Ba, Ca, Cr, Eu, Fe, La, Mn, Mo, Ni, Sc, Sn, Ti and V in addition to those mentioned with anion exchangers. Thus, carboxymethylated polyethyleniminepolymethylenephenylene isocyanate chelating resin can be used for the preconcentration of Cd, Pb and Zn from seawater or Cd, Co, Cu, Hg and Zn from surface water by shaking the water with the resin beads. Enrichment factors of ~400 are obtained with this material.[45] Conventional polystyrene powder containing conformationally flexible aminocarboxylic acid groups have high sorption for most elements at pH 3–7 except the alkali metals. Direct INAA on the sorbent allows determination of Au, Ba, Cd, Co, Cr, Cu, Eu, Fe, Hg, In, La, Mn, Sc, Sr and V.

The most commonly used inorganic exchanger is HAP. The material is prepared, crushed and sieved and then packed into a column as with an exchange resin. The key feature of HAP is its ability to absorb Na strongly. Thus by passing seawater through such a column, the interfering Na is removed. Other chemically modified exchangers have been used for preconcentration in special circumstances.

4.3.2.4 Solvent Extraction This method like coprecipitation or scavenging has the advantage of rapidity and simplicity. Usually, the trace metals are extracted as chelates and the method simply requires the water sample to be shaken with the chelate in an immiscible organic solvent (usually CCl_4 or $CHCl_3$). The metal reacts with the chelate and the metal chelate is then extracted into the organic phase. APDC and sodium diethyldithiocarbamate (NaDDC) are the most widely used chelating agents. Extractions based on both these agents are pH-dependent so that the elements extracted can be controlled by controlling the pH of the water sample. Following the extraction, the organic solution may be subjected to INAA or sometimes a back extraction is performed so that the extracted metals are back extracted into an aqueous phase for irradiation.

4.4 Postirradiation Procedures: Radiochemical Neutron Activation Analysis

It was argued at the beginning of this article that a major advantage of NAA over other analytical techniques was the possibility of chemical separation following irradiation (RNAA) by which contamination problems were eliminated. Unfortunately, with water samples, RNAA is not so advantageous. This is partly due to the limited volume of water and neutron dose that can be used in the irradiation, usually resulting in poor DLs even after chemical separation unless inordinately long counting times are used. Additionally, the presence of Br, Cl and Na, particularly for seawater samples, makes sample handling immediately after irradiation very difficult and also creates problems for the determination of short-lived elements. Thus RNAA is much less used for water analysis than for solid environmental samples.

The majority of postirradiation separations are based on either ion exchange or solvent extraction. A method that is commonly used is to pass the irradiated water through a column of HAP to remove (mainly) ^{24}Na, which is often the most active nuclide present a few hours

after irradiation. This one-step procedure may then allow many of the other elements to be determined directly. Other inorganic exchangers have been investigated for selective removal of specific nuclides, with variable success. Conventional anion exchange resins are the most commonly used exchangers where chromatographic or group separations are required. For solvent extraction, extraction of chelates (PDC, in particular) is used with extraction into a number of solvents (CCl_4, $CHCl_3$, methyl isobutyl ketone (MIBK)).

Coprecipitation is occasionally used, as in the determination of As, Mo, Sb and Se, coprecipitated with Bi_2S_3. Substoichiometric isotope dilution analysis has also occasionally been used (determination of Cr). Finally, iodine is sometimes determined in irradiated water by isotope exchange.[46] Here, irradiated water is shaken with a solution of iodine in CCl_4 when some 95% of the active iodine present in the water can be exchanged and then determined by counting an aliquot of CCl_4.

4.5 Biomonitoring and Hydrological Studies Using Neutron Activation Analysis

Monitoring the marine environment is one of the most difficult but most essential tasks in environmental studies. One method of monitoring is to analyze plant or animal tissues which are known to incorporate or concentrate pollutant elements. These materials are more amenable to INAA than the water itself but reflect the marine conditions. An example of this is the brown algae which inhabit most of the world's coastal seas and show an affinity for a number of pollutants, such as Ag, As and Sr, and levels of many other trace elements, reflecting the waters they inhabit. They have been shown to be useful as monitors of trace elements in coastal waters.

Other marine plants are also useful in this context such as certain phanerogams.[47] Some 43 trace elements have been determined in this plant by INAA, and comparisons between those growing in polluted and unpolluted areas reveal that the plant can be used to assess levels of pollution provided sampling is careful. In addition, these plants concentrate elements known for their biotoxicity from seawater and so can be used as indicators of this type of pollution.

The levels of As in drinking water have also been assessed by analysis of toenail clippings using NAA.[48] It was found that a 10-fold increase in the concentration of As in the water led to about a twofold increase in toenail As. Thus, toenail As levels could be used to assess As levels in water of a particular region.

NAA has begun to play a part in hydrological studies in recent years. In the past, radioactive tracers were introduced into aqueous systems to study the behavior of large bodies of water, necessitating the use of large quantities of long-lived nuclides. Fears over potential contamination of the environment have led to the use of activable tracers instead. Here, an ordinary stable element but one not present in the water, is introduced. At a later time, the water is sampled, treated in a manner previously discussed and then activated. Measurements of induced activity indicate dispersal and other properties of the element and, hence, information on the water system itself.[49]

5 ANALYSIS OF AIR PARTICULATES

5.1 Introduction

The analysis of the air constitutes the second major area of environmental analysis, along with that of water. The importance of air arises from its fundamental role in the existence of all living things. The air we breathe is essentially a mixture of gases with suspended solid particles and liquid droplets which are referred to as aerosols. NAA of air is essentially confined to the analysis of the suspended particulate matter. Although this may seem a very restricted role for NAA, the position of particulate matter in human health is currently regarded as one of the most serious environmental problems. There is particular concern regarding particles with aerodynamic diameters less than 10 μm, the so-called PM10. The IAEA began a global coordinated research program (CRP) in this area in 1992 and a report of the work has been issued.[50] A review of this whole area and methods of determining compliance with current United States air quality standards for suspended particles has also appeared.[51] An earlier review of the role of activation techniques in the analysis of aerosols also emphasized the importance of these particles.[8]

5.2 Particulate Matter: Sources, Composition and Inhalation

The particulate matter found in air covers a wide range of sizes from about 100 μm down to about 0.01 μm diameter; however, the most important size range with respect to human health is that fraction $< 10\,\mu m$ diameter. This range is further subdivided into those particles between 0.08 μm and $\sim 2\,\mu m$, which constitute the fine particle size fraction and those between $\sim 2\,\mu m$ and 10 μm which constitute the coarse particle size fraction. The majority of fine particles are of anthropogenic origin whereas the majority of coarse particles are of geological origin. Many schemes for sampling the air allow separation of these two fractions. The main sources of PM10 are shown in Table 3. The actual fraction of each of these components present in PM10 at a given site will be

Table 3 Relative abundance of elements, inorganic chemicals and carbon in PM10s from different emission sources

Source of emission	Dominant particle size	Relative abundances of elements, inorganic chemicals and carbon in percent mass			
		<0.1%	0.1–1%	1–10%	>10%
Paved road dust	Coarse (2.5–10 μm)	Cr, Sr, Pb, Zr	SO_4^{2-}, Na^+, K^+, P, S, Cl, Mn, Zn, Ba, Ti	(EC), Al, K, Ca, Fe	OC, Si
Construction	Coarse	Cr, Mn, Zn, Sr, Ba	SO_4^{2-}, K^+, S, Ti	OC, Al, K, Ca, Fe	Si
Agricultural soil	Coarse	NO_3^-, NH_4^+, Cr, Zn, Sr	SO_4^{2-}, K^+, Na^+, S	OC, Al, K, Ca, Fe	Si
Natural soil	Coarse	Cr, Mn, Sr, Zn, Ba	Cl^-, Na^+, EC, P, S, Cl, Ti	OC, Al, Mg, K, Ca, Fe	Si
Lake bed	Coarse	Mn, Sr, Ba	K^+, Ti	OC, SO_4^{2-}, Na^+, Al	Si
Motor vehicle	Fine (0–2.5 μm)	Cr, Ni, Y, Sr, Ba	Si, Cl, Al, P, Ca, Mn, Fe, Zn, Br, Pb	Cl^-, NO_3^-, SO_4^{2-}, NH_4^+, S	OC, EC
Vegetative Burning	Fine	Ca, Mn, Fe, Zn, Br, Rb, Pb	NO_3^-, SO_4^{2-}, NH_4^+, Na^+, S	Cl^-, K^+, Cl, K	OC, EC
Residual oil combustion	Fine	K^+, OC, Cl, Ti, Cr, Co, Ga, Se	NH_4^+, Na^+, Zn, Fe, Si	V, OC, EC, Ni	S, SO_4^{2-}
Incinerator	Fine	V, Mn, Cu, Ag, Sn	K^+, Al, Ti, Zn, Hg	NO_3^-, Na^+, EC, Si, S, Ca, Fe, Br, La, Pb	SO_4^{2-}, NH_4^+, OC, Cl
Coal-fired boiler	Fine	Cl, Cr, Mn, Ga, As Se, Br, Rb, Zr	NH_4^+, P, K, Ti, V, Ni, Zn, Sr, Ba, Pb	SO_4^{2-}, OC, EC, Al, S, Ca, Fe	Si
Oil-fired power plant	Fine	V, Ni, Se, As, Br, Ba	Al, Si, P, K, Zn	NH_4^+, OC, EC, Na, Ca, Pb	S, SO_4^{2-}
Marine	Fine and coarse	Ti, V, Ni, Sr, Zr, Pd, Ag, Sn, Sb, Pb	Al, Si, K, Ca, Fe, Cu, Zn, Ba, La	NO_3^-, SO_4^{2-}, OC, EC	Cl^-, Na^+, Na, Cl

EC, elemental carbon; OC, organic carbon.
Reproduced by permission of Air & Waste Management Association from Chow.[51]

strongly dependent on the source and it is possible to classify and identify sources of PM10 from the relative values of these fractions.

The reason for the interest in PM10 is that these represent the inhalable fraction of particulates since most particles larger than 10 μm are removed in the mouth or nose and so do not enter further into the body. Up to 60% of particles < 10 μm diameter may be deposited in the lungs. The health implications lie in the fact that these particles are contaminated by a wide range of elements and compounds, including toxic and carcinogenic ones. In addition, Si particles may result in lung damage even at low levels.

5.3 Sampling

The majority of environmental air sampling involves aerosol collection and the most widely used methods of collection are by filtration and impingement. Filtration is simple and provides sufficient material for subsequent NAA. A basic filtration unit consists of a filter holder, filter, an air mover (pump) and a flow-rate meter. Uniform flow of air across the filter is desirable since measurements of element concentrations are nearly always made directly on the filter. Accurate measurement of flow rate is also important in order to determine the total volume of air from which the sample was collected.

A wide variety of filter media are available but the two types most routinely used for environmental sampling are fibrous filters (cellulose paper or glass fiber mainly) and membrane filters (gel-type or nuclepore). Cellulose papers are relatively cheap and robust but suffer from variability in collection efficiency. Nevertheless, for the rapid collection of sufficient material for an analysis they are quite adequate. Glass fiber filters have more uniform collection efficiencies and tend to collect more material at the surface of the filter but may give rise to high blank values for some elements.

Membrane filters are quite different in construction from fibrous filters but the most significant difference to the fibrous type is that the pore size is both controllable and variable and they can be used for the collection of particles down to 0.1 μm. Their collection efficiency is better than for fibrous filters and for a pore size of 0.6 μm they collect essentially all particles. For the collection of relatively large amounts of material (several mg), however, the fibrous type is the usual choice.

Impingement for the collection of aerosols relies on a stream of air being forced through a nozzle at the back of

which is some form of obstruction. Particles carried in the airstream then collide forcibly with the obstruction and tend to collect on it. The collection efficiency is usually increased by attaching a ''sticky'' film to the surface of the obstruction such as a film coated with petroleum jelly. The main use of impingers is the ability to collect fractions of different particle size so that the range of particle sizes associated with the majority of the airborne elements can be determined. The major disadvantage of impingement devices, however, is the relatively small mass of material that can be collected. The cascading impactor is perhaps the most useful design of impingement equipment. It consists of a series of stages, each of which collects successively smaller particle fractions.

Kist et al.[52] have described a sampling system for the small-size particles (<0.1 μm) and vapor-gas phases involving the use of very thin polyethylene films placed directly in the atmosphere. Various configurations were investigated and data indicated the potential for trapping submicron and vapor-gas aerosol phases.

5.4 Measurements

In the case of particulate matter collected alone, e.g. from an impaction collector, the material can be packed directly into cleaned polyethylene irradiation vials or compressed into a pellet first. For filters, the filters plus deposit may be folded and then packed into the irradiation vials. Even better is to compress the filter into a pellet which gives improved geometry for both irradiation and counting. Some filters are brittle and may crumble whilst being handled with possible loss of some of the deposit. Flexibility can sometimes be improved by exposure to acetone. For long irradiations, aluminum or silica may replace polyethylene which may become brittle after exposure to a high neutron fluence. Standards are usually prepared by pipetting accurate volumes of solution containing known concentrations of the elements under investigation onto filter discs and then after drying, treating them the same as the samples. A blank correction is always required when any type of filter disc is used since these always contain measurable levels of some trace elements, requiring a correction to the analyzed sample data. For the usual comparator method of INAA, suitable standards are required (either a single multielement standard, frequently a suitable SRM, or a series of filter standards containing only a single element).

Most of the trace elements in air particulate samples can be determined by INAA. This is particularly true for aerosols collected on organic filters. As discussed earlier, a major fraction of any aerosol will be organic matter, carbon, sulfate, nitrate or ammonium, none of which produce significant gamma activity. The inorganic fraction, mainly consisting of the oxides of Ca, Fe, Si and Mg, plus P and Cl again produce low or short-lived activities. Sodium and bromine can be a problem in some marine aerosols and this may require a delay in counting until the ^{24}Na activity has decayed to an acceptable level, usually a few days after irradiation. Such a delay time precludes the analysis of any short-lived trace elements by this technique. In general, two irradiations are carried out as for most other types of environmental sample. A short irradiation (1–5 min duration depending on the available neutron flux) uses a rabbit system for the short-lived nuclides of Al, Ca, Cl, Cr, Cu, Mg, Na and V. This may also include CINAA for the determination of the very short-lived nuclides (<10 s) ^{38m}Cl, ^{77m}Se, ^{185m}W and ^{207m}Pb. These are then followed by a long irradiation (2–8 h, again depending on the flux) and several counting periods at different decay times for the remaining long-lived elements. In this way some 33 elements (Al, As, Ba, Br, Ca, Ce, Cl, Co, Cr, Cs, Cu, Eu, Fe, Hf, Hg, K, La, Mg, Mn, Mo, Na, Pb, Rb, Sb, Sc, Se, Sm, Sn, Sr, V, W, Zn and Zr) can be determined with DLs ranging from 4×10^{-4} ng m^{-3} for Eu to 30 ng m^{-3} for Ca and Cl, based on filtering 1000 m^3 of air and irradiations with a flux of $\sim 2 \times 10^{17}$ nm^{-2} s^{-1}. A more complete set of DLs is given by Kolthoff et al.[53]

These data are based on the normal multielement comparator technique. Some data have also been published based on the single comparator method which reduces the analysis time of the standards considerably. It has also been shown that the k_0 method gives reliable results for suitable air particulate reference materials and has been used in some air pollution studies. A few more elements (Ag, Au, F, In, Ta, Th and U) are better determined by ENAA. Also, Landsberger and Wu[54] showed that ENAA in conjunction with a Compton suppression γ-ray detection system gave improved DLs for As, Cd, I, Mo, Sb, Si, and U, with a particularly low DL for iodine. Finally, we have already mentioned that F is probably best determined by the ^{19}F(n,α)^{16}N reaction using 6 MeV neutrons from an accelerator in conjunction with a rabbit system.[29]

The use of low-energy photon spectrometry (LEPS) i.e. photons in the range 50–200 keV has also been investigated for use with both INAA and ENAA. The technique, however, does not markedly add to the list of elements that can be determined using INAA beyond the lanthanides Lu, Sm, Tm and Yb and an advantage for the determination of Cu. In combination with ENAA, it improves the sensitivity for the determination of Br, Co, In, Sb and V.

FNAA plays a minor role in air particulate analysis. In general, it is confined to the determination of the matrix elements (Al, Ca, Cl, Fe, Mg, Mn, Si and Ti). In a few laboratories where high fluxes of 14 MeV neutrons are available ($\sim 5 \times 10^{12}$ n s^{-1}) it is possible to determine

a number of trace elements although, again, the range compared with INAA and ENAA is not extended and the small sample weight militates against low DLs.

RNAA also plays little part in air particulate analysis. This is partly due to the fact that monitoring air pollution requires the analysis of large numbers of samples fairly rapidly but also due to the problems of sample size and dissolution.

Accuracy and precision are problems in any type of analysis. For INAA, the problems associated with the technique are well known and in most circumstances lead to a precision of $\sim\pm 1-2\%$. For air particulate analysis, the largest error probably arises from the inhomogeneity of the sample; however, as long as the entire sample is subjected to analysis, this should not detract from the normal precision of the technique. Kist et al.[52] revealed differences in the concentration of elements such as Au, Br, Cl, Hg and W determined by their passive absorption method compared with samples obtained by the standard filtering methods. This was ascribed to the fact that these elements exist in a thin particulate or a vapor-gas phase which are only sampled with poor efficiency by the use of standard filters.

5.5 Applications

5.5.1 Direct Methods of Analysis

A large number of studies of air pollution and related studies have certainly been undertaken since about 1985. The majority of these apply to the monitoring of the air in specific urban areas with some data from adjoining rural areas to provide a comparison. Alternatively, a number of studies have also been concerned with the analysis of the emissions in the neighborhood of incinerators, smelters and power stations since these can be the sources of large quantities of particulate (some of it PM10) and other toxic components, both inorganic and organic. In the earlier phases of these studies most of the effort was confined to the developed world. A number of such studies are detailed by Alfassi[3] in Chapter 13 of Volume 2 of his book and in an earlier review by Alian and Sansoni.[55] However, attention has shifted to areas such as southeast Asia since their rapid economic rise has been accompanied by that of severe air pollution. Studies have shown that of the 15 cities with the highest total particulate matter in their air, 12 of these are in Asia. This has stimulated an IAEA CRP in the Asia and Pacific region. Determinations of large numbers of trace elements in air particulates on filters by a combination of methods, including INAA, can be used as internal checks on the quality of data. Such consistent data can then be reliably used for receptor model analyses.

Fly ash from power stations is a particular problem, being composed of the inorganic fraction of the fuel. Studies based on INAA and other techniques have shown that some toxic and radioactive species (e.g. As, Mo, Ni, Pb, Th, V and Zn) are concentrated significantly in this fly ash relative to their initial values in the fuel.[56] This process is exacerbated by adsorption on the fly ash of these components in the flue gas.

Studies on the quality of air in the workplace environment have become of increasing importance. A number of industries (e.g. metal plating, metallurgical processing, etc.) lead to the release of toxic metals and other pollutants into the local atmosphere. Equally, levels of pollutants in indoor public areas due to, say, tobacco smoking have become of concern. Analysis of filter samples of indoor air polluted by tobacco smoke using an ENAA technique revealed elevated levels of As, Cd and Sb, with consequential potential health risks beyond those normally associated with passive smoking.[57]

5.5.2 Indirect Methods of Analysis

Monitoring the ambient air by the methods discussed above has limitations, on both a geographical and a temporal scale. These are due to the necessity of setting up monitoring stations in suitable areas, which are expensive in themselves, and then the necessity for analyzing the samples collected. Such monitoring stations are not really suitable for covering wide geographical areas, particularly if remote areas with difficult accessibility are involved. Also, studies over a long time period (several years) can involve the analysis of a large number of samples, even from one station.

In these circumstances, another approach is taken, that of indirect monitoring. The majority of such monitoring usually involves sampling various types of plant materials which can retain the elements deposited on them from the atmosphere and thus act as monitors of the atmosphere in which they are growing. Such materials are referred to as biomonitors. Other, nonorganic materials such as snow have also been used as accumulators of atmospherically deposited trace elements.

Lichens and mosses have almost invariably been used in air pollution studies since these plants absorb elements directly from the air because they lack root systems. They also concentrate and retain the elements adsorbed. This is in contrast to vascular plants which take elements directly from the air and, more importantly, from the soils in which their roots grow. Hence, variations in element concentration cannot normally be directly correlated with changes in atmospheric concentration. An exception to this is the use of evergreen oak leaves to monitor vehicle pollution using both INAA and photon activation analysis (PAA) (for Pb).[58] These leaves were shown to be very good monitors for air pollution studies, particularly for Pb and a few other elements (Ce, Co, Eu, Hf, Sb and Zn),

which were shown not to reach the leaves through the tree roots. There was good correlation between Pb on the leaves and the level in the air.

Generally, however, lichens and mosses have been used as biomonitors. One of the most comprehensive studies using mosses was that by Steinnes et al.[59] Here a specific moss was collected from over 500 sites throughout Norway and the moss samples subjected to INAA, with some 22 elements determined. A further four elements were determined by flame AAS. The data from these studies allowed different sources of pollution to be identified varying from long-range atmospheric transport of pollutants from other parts of Europe to local sources within Norway. There was also an indication of atmospheric transport from the marine environment. Similar studies by Sloof and Wolterbeck based on lichens collected in the Netherlands were used to create geographical concentration patterns and to estimate the sources of pollution.[60] Analysis of lichens and mosses by INAA has also been used to investigate pollution around industrial plants such as smelters.

A more unusual example of indirect monitoring is that based on the analysis of snow.[61] The method is based on the fact that snow scavenges particulates from the air whilst larger particles sink into the snow and are immobilized in it. Samples of melted snow from the Montreal area of Canada were analyzed by INAA allowing enrichment factors to be determined and thus allocated to natural or anthropogenic sources.

6 ANALYSIS OF SOIL, SEDIMENTS AND COAL

6.1 Soil

Considering the importance of soil for the production of food, there has not been a vast number of analyses by either INAA or RNAA. For INAA it is usual to perform a short and a long irradiation, as in most areas of environmental analysis, and possibly two counting periods following the long irradiation. This results in the determination of up to 24 elements (Al, Ba, Br, Ce, Co, Cr, Cs, Dy, Eu, Fe, Hf, La, Lu, Mn, Na, Rb, Sb, Sc, Sm, Tb, Th, V, Yb and Zn) in most soils and a further 16 (As, Au, Ca, Cl, Ho, I, K, Mg, Nd, Sr, Ta, Ti, Tm, U, W and Zr) in many soils. This number is augmented by nine more elements (Ag, Cd, Gd, In, Ir, Mo, Ni, Pd and Sn) by use of ENAA although, again, a few of the elements determined by INAA are better determined by ENAA. Further elements such as Hg and Se can be determined by RNAA but this technique has not been widely used for soils.

A number of soil and sediment SRMs are available to act as multielement standards or as reference materials, with some having quite high concentrations of several trace elements. These are often suitable as standards in the analysis of soils and sediments polluted by heavy metals.

Soils have also been analyzed in attempts to study the mobility of heavy trace metals through them. This is of particular concern where sewage sludge, often containing high concentrations of heavy metals has been used as fertilizer, a growing practice in many countries. The potential mobility of these metals in soils is of concern and has been investigated using both INAA and ICPS (inductively coupled plasma spectroscopy). INAA is probably a better method owing to the presence of some metals that are located generally in minerals resistant to acid digestion and hence give low readings by the ICPS technique. The data indicate that, not surprisingly, different metals have different mobilities but that significant amounts of Cd, Cu, Hg and Zn remain in the topsoil up to 15 years after application of the sludge.[62]

Regional surveys of surface soils have also been carried out using INAA as well as other analytical techniques in order to construct accurate geochemical databases for environmental, agricultural and related purposes. Obviously, such surveys require the analysis of large numbers of samples for which INAA is admirably suited.[63] Similarly, baseline element concentrations have been determined for soils and some associated vegetation in a given region with the aim of investigating landscape patterns of the elements. Additionally, soils have been used as pollution monitors in regions where they are subject to potentially high concentrations of toxic elements.

A number of studies using INAA have also investigated the uptake and concentration of heavy metals by plants growing in normal or contaminated soils. This is of concern where plant material for human consumption is grown. An example was the concern for the levels of Cd in cabbages grown domestically in the southwest of England when the levels of Cd in the soil were found to be several times the local background level.

6.2 Sediments

By sediments we mean the material deposited at the bottoms of lakes, rivers, estuaries and seas. The techniques and elements that can be determined by INAA are very similar to those discussed for soils. In sediment analysis, however, a much greater emphasis is placed on determining element concentrations in sediment cores. Core samples are often obtained using a simple plastic tube, 2–5 cm diameter and up to a meter in length. The tube

is inserted into the sediment to obtain the core which can later be extruded from the tube. This core is then usually cut into sections (often 5 or 10 cm in length but shorter lengths for greater detail). Core sections, after drying (freeze drying preferred) and pulverizing can then be packaged for NAA. INAA, ENAA and FNAA have all been used to analyze sediment cores.

The core is essentially a historical record of the sediment with the age of the sediment increasing down the core; each cut segment then represents a specific time zone. Analyzing the segments reveals changes in the concentration of the elements with time and may reflect particular events in the history of the region. The total length of time represented by a single core obviously depends on the sedimentation rate, which itself may vary with time. A plot of change in concentration with core depth is known as a concentration–depth profile. In some cases the concentration–depth profile may reveal the sedimentation rate if it is assumed to be constant. Sodium concentration–depth profiles may reflect the change of sediment origin from lacustrine through brackish (possibly estuarine) and possibly even to marine for sediments found close to the sea. Cores taken from the sea bed adjacent to the outflow of a major river such as the Amazon or the Mississippi can distinguish the river sediments from those of marine origin and so can be used to investigate the factors controlling the discharge of river sediments into the ocean.

A number of countries (e.g. Canada, South Africa, UK) have also used INAA to analyze large numbers of samples obtained from lake and river sediments from throughout the country concerned again to provide regional geochemical databases.[63] These can help in the geochemical exploration for commercially viable mineral deposits and potentially useful regions can then be investigated in greater detail by geological investigation.

The method of activable tracers has also been used to investigate accretion, reworking and erosion of sediments. Suitable tracers are certain rare earth elements (Dy and Sm) and in one case a Cs-saturated phlogopite (a form of mica with exchangeable K) where the Cs replaced the K of the phlogopite.[64] In all cases the tracer is applied to the sediment. After a given period of time (ranging from days to years), the sediment is sampled and after activating the tracer, its concentration can be determined and used to investigate changes in the sediment.

Finally, studies of the transport and ultimate fate of pollutants can be studied by examining sediments from lakes or rivers. Analyses of such sediments is particularly useful for following the transport and ultimate fate of fertilizers washed out of fields and into rivers and lakes. Fertilizers of phosphate origin tend to have relatively high concentrations of the rare earth elements and this will result in increases of these elements in the lake or river sediments. INAA is a very sensitive method for the determination of most of the rare earths usually making it the preferred method in this type of study. Toxic elements such as As have been determined by RNAA in high concentrations in lake sediments as a result of the use of arsenite pesticides sprayed on to adjacent agricultural areas.

6.3 Coal

NAA has been widely used in coal analysis, much of it directed towards on-line analysis of bulk coal samples prior to their use in power or conversion (to liquid fuel or gas) plants. The other role of NAA in coal science is in the investigation of trace element concentrations and patterns.

Trace and minor element analysis of coal usually employs a short and long irradiation strategy, allowing for the determination of about 38 elements (Al, Ag, As, Ba, Cr, Ca, Ce, Cl, Co, Cr, Cs, Dy, Eu, Fe, Hf, I, K, La, Lu, Mg, Mn, Na, Nd, Rb, S, Sb, Sc, Se, Sm, Ta, Tb, Ti, Tl, U, V, W, Yb and Zn). In addition, B is sometimes determined by PGNAA. Coal SRMs are available to act as both multielement standards in INAA or to provide checks on the accuracy of analyses using other standards. Such analyses can serve a number of purposes.

One of these involves exploration and mining of coal where the trace element composition has been used to try to ''fingerprint'' coal seams. Most of these attempts, however, have not proved commercially viable owing to the large variations of trace element patterns within a given seam. More hopeful is the fingerprinting of certain formations that enclose a given seam.

Of more importance, however, is the use of trace element data to investigate the fate of volatile or more or less toxic elements when coal is burnt.[65] Trace elements in coals might be expected to follow the crustal abundance of such elements. This is found to be the case for many of the elements although some such as As, Br, Cd, I, Mo, Pb, S, Sb and Se are significantly enriched. Levels of S vary widely in coals and, because of its crucial role in the production of acid rain, levels have to be carefully monitored.

The combustion of coal releases significant amounts of trace elements either in the form of gaseous discharge (Br, Cl, I, S and Se) or as fly ash, particulate matter of varying dimensions, to the atmosphere. These releases are then studied as part of air pollution monitoring, as discussed in section 5 of this article.

The rest of the trace elements not released to the atmosphere are retained in the ash that remains at the end of combustion. It is found that most nonvolatile elements are considerably concentrated in the ash. The total ash formed depends on the initial inorganic (mainly

silicates) content of the coal. The disposal of this ash remains a major environmental problem since the total world annual output is hundreds of millions of tons. Most of this ash is disposed of by landfill or minefill. An important problem is the leaching of toxic elements from this material by groundwater and leaching studies based on INAA have been undertaken by Jervis et al.[66]

On-line analysis of coal has become increasingly important and activation techniques have predominated because they are nondestructive, can be automated and are capable of producing rapid results. Also, radioactivity induced in the coal by neutron irradiation is both short-lived and negligible. Coal analysis in this context is used to determine the ash (from the concentrations of Al, Fe and Si), sulfur, nitrogen (important for its role in affecting catalysis in coal conversion) and chlorine (high chlorine coals cause boiler corrosion) contents of coal along with its calorific value. The latter is not determined directly but is based on the carbon content of the coal.

At present, the favored approach to on-line coal analysis is based on PGNAA.[67] A fast neutron source (either isotope- or accelerator-based) is used and the neutrons thermalized within the coal to allow neutron capture necessary for PGNAA. For on-line situations radioisotope sources offer many advantages over accelerators, being physically small, rugged, not requiring any external power or maintenance and having long useful lifetimes. The sources most generally used are ^{252}Cf and ^{241}Am/Be.

Unfortunately, the measured intensity of any capture γ-ray peak is not directly proportional to the element concentration as it also depends on the thermal neutron flux in the sample, which is itself dependent on the concentrations of all the elements in the sample and, in particular, small concentrations of B or Cl. On the other hand, the ratio of one capture γ-ray peak area to that of another is directly proportional to the ratio of the elemental concentrations. Therefore, if one capture γ-ray-producing element can be absolutely determined by a separate technique, a quantitative assay of all the other capture γ-ray producing elements can be made. Fortunately, this is possible in the case of hydrogen whose absolute determination can be made based on the measurement of the epithermal neutron flux escaping from the coal. This then allows for the determination of Al, C, Cl, Fe, N, S and Si by the peak ratioing technique.

Two problems still remain with on-line analysis based only on PGNAA. The first is that although C can be determined by detection of the prompt γ-rays, the method is severely limited by the small (n,γ) cross-section, so that the precision of the determination is poor. The second and more important problem is that PGNAA does not allow for the determination of O because its capture cross-section is extremely small and detection of subsequent γ-rays is almost impossible. Thus determination of *O* is only possible by difference, which involves the determination of the concentration of all the other major elements and the difference from 100% is assumed to be oxygen. This is the method that must be used in all other techniques for coal analysis.

Both problems are solved by the use of inelastic scattering reactions to produce γ-ray peaks. Here a ^{241}Am/Be source must be used because of its ability to excite more noncapture reactions than ^{252}Cf. Thus, C is determined from the 4.43 MeV γ-rays resulting from the ^{12}C(n,n$'\gamma$) inelastic reaction. Oxygen is determined indirectly using the ratio of C to O peak areas produced in the inelastic scattering reactions on the two elements (6.13 MeV γ-ray from ^{16}O(n,n$'\gamma$)). Since carbon can be determined either by PGNAA or absolutely using the inelastic scattering peak (to an accuracy of $\pm$10%), this gives a method for the determination of oxygen. The disadvantage is that large expensive NaI(Tl) (thallium-activated sodium iodide detector) crystals must be used, which are also susceptible to vibration and thermal shock making them unsuitable for on-line purposes.

7 ROLE OF NEUTRON ACTIVATION ANALYSIS IN SPECIATION STUDIES

7.1 Introduction

An important aspect of the analysis of trace elements in environmental samples is that of speciation. Speciation analysis can be defined as the concentration of an element in each of the physicochemical forms present in the system and which together represent the total concentration of that element. For example, natural waters usually contain some particulate or colloidal material so that a particular element may be present in solution or bound to the particles or colloids. A similar situation exists for other environmental samples such as air, where we have already seen that particulate matter is present, and in biological systems. The presence of different species (maybe in different oxidation states) of an element will affect a number of environmental parameters, such as its mobility, transport, bioavailability and toxicity.

NAA by itself cannot distinguish between the various species in which the element may be found; however, in combination with other analytical techniques it may be used very effectively in speciation analysis. The most important of these other analytical techniques are various separation methods which bring about the isolation of the respective fractions or forms which can then each be analyzed by the available NAA techniques.

7.2 Speciation of Elements in Water

There is a gradation in size between simple molecules present in solution and macroscopic particles rather than abrupt changes. It is important, therefore, that the method adopted for size fractionation does not itself disturb the original distribution. Size fractionation methods based on both filtration and ultrafiltration are commonly applied. For ultrafiltration, where the aim is to isolate the low-molecular-weight (low-MW) species, dialysis, the use of hollow-fiber cartridges or a series of ultrafilter membranes are usually the preferred methods. Hollow-fiber cartridges have the advantage that the species do not absorb on the fibers to any great extent. Cartridges with MW cut-off of 10^3, 10^4 or 10^5 are simply placed in the water and the species of mass below the particular MW cut-off are filtered through. Similarly, a series of filters and ultrafilters may be used to separate both larger, suspended particles and then successive MW fractions down to as low as MW 500. This ultrafiltrate can then be collected directly in polythene irradiation vials along with the the solid on the filter for INAA. Alternatively, a polythene container is used to collect the filtrates if RNAA is necessary. Analysis by INAA can give information for up to about 40 elements in many cases Thus, it is possible to show the major forms in which these elements exist in water, e.g. alkali metals occur mainly as free cations, Al, Au, Sb, U and V mainly as hydroxy complexes whilst other elements such as Co, Cr, the lanthanides, Ni and Zn tend to form soluble organic and inorganic complexes.[68] In addition, the size distribution pattern of a large number of elements can be readily investigated.

Water containing fulvic and humic acids (complex organic material) would be expected to give rise to organic complexes of several elements. One way to investigate such complexation would be to decompose the organic matter prior to ultrafiltration so that the complexed elements are released back into solution. The destruction method must ideally be capable of operating in situ and a candidate is γ-irradiation.[68] The elements are again determined by INAA following γ-irradiation of the water and ultrafiltration. In this way it can be shown that the dissolved species of Co, Cr, Ho, Lu, Ni, Sc, Tm and Yb predominately exist as soluble organic complexes, whereas Ag, Ce, Eu, Fe, La and Sm are chiefly associated with organic colloids.

For elements such as As, Cr and Sb, where their toxicity and biological activity depend on their oxidation state it is important to differentiate between species in different oxidation states. In order to do this some form of chemical separation is essential. We have already referred to the possible separation of As(III)/As(V) and Sb(III)/Sb(V) on MgO (section 4.3). The oxidation states of Sb can also be separately determined by INAA by use of the fact that only Sb(III) is extracted into a chloroform solution of APDC. Similarly, As(V) may be determined by its complexation with molybdate followed by coprecipitation with tetraphenylphosphonium chloride, with any As(III) present first removed with dibenzyldithiocarbamate.[69] Finally, Cr(III) and Cr(VI) can be separated based on the variation of the coprecipitation yields with $Pb(PDC)_2$ as a function of pH, with Cr(VI) coprecipitated at pH 4.0 but Cr(III) only at pH 9.[70] INAA of the precipitate thus yields the concentration of either species directly. Similarly, iodide and iodate speciation in water can be determined following preconcentration on bismuth thionalide.[71] Following irradiation of the precipitate, kinetic differences are exploited in the isotopic exchange of iodide and iodate with iodine in chloroform solution, since the isotopic exchange of iodide occurs almost instantaneously but that for iodate is very slow.

7.3 Speciation of Elements in the Atmosphere

One method that has been used is based on the analysis of rain water.[72] The rain water is filtered through a membrane to collect the solid particulate matter whilst the soluble trace elements are collected in the filtrate. The filtrate is dried to form a solid and both this and the solid collected on the filter are irradiated and elements determined by INAA. In this way, a crude estimation can be made of the trace elements present in the particulate matter and those present in the vapor-gas phase. Based on such experiments it was concluded that Au, Br, Cl, Cr, Cs, Fe, K, La, Mn, Se and Sm are found in the particulate fraction whereas As, Hg, Sb and Zn occur mainly in the vapor-gas phase.

A method for separating particulate, inorganic and organic forms of iodine in the air has also been described.[73] The particulate matter was collected in the usual way by filtration through a membrane filter. Inorganic iodine was collected by passing the air through two filters impregnated with lithium hydroxide and organic iodine in two beds of chemically treated activated charcoal. The filters or activated charcoal were then irradiated and ^{128}I counted. Results showed that almost 90% of the atmospheric iodine was in gaseous form in which the organic form dominated.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AC	Activated Carbon
APDC	Ammonium Pyrrolidine Dithiocarbamate
CINAA	Cyclic Instrumental Neutron Activation Analysis

CRP	Coordinated Research Program
DL	Detection Limit
EC	Elemental Carbon
ENAA	Epithermal Neutron Activation Analysis
FES	Flame Emission Spectroscopy
FNAA	Fast Neutron Activation Analysis
HAP	Hydrated Antimony Pentoxide
IAEA	International Atomic Energy Agency
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICPS	Inductively Coupled Plasma Spectroscopy
INAA	Instrumental Neutron Activation Analysis
LEPS	Low-energy Photon Spectrometry
MIBK	Methyl Isobutyl Ketone
MS	Mass Spectrometry
MW	Molecular Weight
NAA	Neutron Activation Analysis
NaDDC	Sodium Diethyldithiocarbamate
OC	Organic Carbon
PAA	Photon Activation Analysis
PAN	1-(2-Pyridylazo-2-naphthol)
PDC	Pyrrolidine Dithiocarbamate
PGNAA	Prompt Gamma Neutron Activation Analysis
PM10	Particles with Aerodynamic Diameter Less Than 10 μm
RNAA	Radiochemical Neutron Activation Analysis
SRM	Standard Reference Material
TAN	1-(2-thiazolylazo)-2-naphthol

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Atomic Fluorescence in Environmental Analysis • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Actinides and other Alpha-emitters, Determination of • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation • Speciation of Radionuclides in the Environment

REFERENCES

1. G. Hevesy, H. Levi, 'Action of Neutrons on the Rare Earth Elements', *Kgl. Danske Videnskab. Selskab. Mathfys. Medd.*, **14**(5), 34 (1936).
2. G.V. Hevesy, H. Levi, 'Artificial Activity of Hafnium and Some Other Elements', *Kgl. Danske Videnskab. Selskab. Math-fys. Medd.*, **15**(11), 18 (1938).
3. Z.B. Alfassi (ed.), *Activation Analysis*, CRC Press, Boca Raton, FL, Volumes I and II, 1989.
4. H.A. Das, A. Faanhof, H.A. van der Sloot, 'Environmental Radioanalysis', in *Studies in Environmental Science 22*, Elsevier, Oxford, 1983.
5. J. Tolgyessy, M. Kyrs, *Radioanalytical Chemistry*, Ellis Horwood, Chichester, Volumes I and II, 1989.
6. M. Eisenbud, 'Environmental Radiochemistry', in *Encyclopedia of Physical Science and Technology: 1990 Yearbook*, ed., R.A. Meyers, Academic Press, San Diego, 1990.
7. W.D. Ehmann, D.E. Vance, 'Advances in Neutron Activation Analysis', *CRC Crit. Rev. Anal. Chem.*, **20**, 405–443 (1989).
8. R. Dams, 'Nuclear Activation Techniques for the Determination of Trace Elements in Atmospheric Aerosols, Particulates and Sludge Samples', *Pure Appl. Chem.*, **64**(7), 991–1014 (1992).

9. S.J. Landsberger, 'Analytical Methodologies for Instrumental Neutron Activation Analysis of Airborne Particulate Matter', *J. Trace Microprobe Tech.*, **10**(1), 1–41 (1992).
10. W.D. Ehmann, S.W. Yates, 'Nuclear and Radiochemical Analysis', *Anal. Chem.*, **58**, 49R–65R (1986).
11. W.D. Ehmann, S.W. Yates, 'Nuclear and Radiochemical Analysis', *Anal. Chem.*, **60**, 42R–62R (1988).
12. W.D. Ehmann, J.D. Robertson, S.W. Yates, 'Nuclear and Radiochemical Analysis', *Anal. Chem.*, **62**, 50R–70R (1990).
13. W.D. Ehmann, J.D. Robertson, S.W. Yates, 'Nuclear and Radiochemical Analysis', *Anal. Chem.*, **64**, 1R–22R (1992).
14. W.D. Ehmann, J.D. Robertson, S.W. Yates, 'Nuclear and Radiochemical Analysis', *Anal. Chem.*, **66**, 229R–251R (1994).
15. S.W. Yates, 'Reviews in Radioanalytical and Nuclear Chemistry', *J. Radioanal. Nucl. Chem.*, **215**(2), 311 (1997).
16. L.N. Chau, N.T. Hiep, V.T. Ha, N.C. Hai, 'Use of PGAA (Prompt γ-ray Neutron Activation Analysis) at the Dulat Reactor', *J. Radioanal. Nucl. Chem.*, **165**(6), 351–362 (1992).
17. V. Barnett, *Elements of Sampling Theory*, Hodder and Stoughton, London, 1974.
18. R. Webster, *Quantitative and Numerical Methods in Soil Classification and Survey*, Clarendon Press, Oxford, 1977.
19. E. Cortes Toro, R.M. Parr, S.A. Clements, 'Biological and Environmental Reference Materials for Trace Elements, Nuclides and Organic Microcontaminants – A Survey', IAEA/RL/128 (Rev. 1), IAEA, Vienna, 1990.
20. A.R. Byrne, 'Some Considerations Regarding Reference Materials and Their Role in Environmental Monitoring', *Analyst*, **117**, 251–258 (1992).
21. A.R. Byrne, M. Dermelj, 'An Endogenous Standard, Radioisotopic Method in Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **223**(1), 55–60 (1997).
22. R.N. Sah, P.H. Brown, 'Techniques for Boron Determination and Their Application to the Analysis of Plant and Soil Samples', *Plant Soil*, **193**, 15–33 (1997).
23. J.J.M. De Goeij, J.R.W. Woittiez, 'Position of RNAA in Trace Element Analysis', *J. Radioanal. Nucl. Chem.*, **168**(2), 429–437 (1993).
24. J.Y. Yang, C.L. Tseng, J.M. Lo, M.H. Yang, 'Determination of Traces of Lithium in Biological, Environmental and Metal Samples by Neutron Activation Analysis', *Fresnius' J. Anal. Chem.*, **321**, 141–145 (1985).
25. T. Braun, P. Bull, J. Fardy, I. Haiduc, F. Macasek, W.J. McDowell, Z. Misak, J.D. Navratil, T. Sato, 'Some Developments for Radioanalytical Separations', *J. Radioanal. Nucl. Chem.*, **84**(2), 461–472 (1984).
26. H.A. van der Sloot, 'Neutron Activation Analysis of Trace Elements in Water Samples after Preconcentration on Activated Carbon', ECN-1, Netherlands Energy Research Foundation, Petten, Netherlands, 1976.
27. W.D. Ehmann, W.R. Markesbery, T.I.M. Hossain, M. Alauddin, D.T. Goodin, 'Trace Elements in Human Brain Tissue by INAA', *J. Radioanal. Chem.*, **70**(1–2), 57–65 (1982).
28. H. Nakahara, Y. Nagame, Y. Yoshizawa, H. Oda, S. Gotoh, Y. Murakami, 'Trace Element Analysis of Human Blood Serum by Neutron Activation Analysis', *J. Radioanal. Chem.*, **54**(1–2), 183–190 (1979).
29. K. Randle, 'Determination of Fluorine in Geological Samples Using Accelerator Derived Neutrons', *J. Radioanal. Nucl. Chem.*, **90**(2), 355–361 (1985).
30. P. Eisenbarth, P. Hille, 'Nondestructive Method for Age Determination of Fossil Bone', *J. Radioanal. Chem.*, **40**, 203–211 (1977).
31. Y.S. Ryabukhin, 'Nuclear-based Methods for the Analysis of Trace Element Pollutants in Human Hair', *J. Radioanal. Nucl. Chem.*, **60**(1), 7–30 (1980).
32. E. Cortes Toro, J.J.M. De Goeij, J. Bacso, Yuan-Di Cheng, L. Kinova, J. Matsubara, S. Niese, T. Sato, G.R. Wesenberg, Y. Muramatsu, J.M. Parr, 'The Significance of Hair Mineral Analysis as a Means for Assessing Internal Body Burdens of Environmental Pollutants: Results from an IAEA Coordinated Research Programme, *J. Radioanal. Nucl. Chem.*, **167**(2), 413–421 (1993).
33. W.A. Haller, R. Filby, L.A. Rancitelli, J.A. Cooper, 'The Instrumental Determination of Fifteen Elements in Plant Tissue by Neutron Activation Analysis', in *Modern Trends in Activation Analysis*, ed., J.R. DeVoe, US NBS Spec. Publ. 312, Vol. 1, US National Bureau of Standards, Washington, DC, 177–183, 1968.
34. Open University Course Team, 'Seawater: Its Composition, Properties and Behaviour, Pergamon Press, Oxford, 1989.
35. D.A. Livingstone, 'Chemical Composition of Rivers and Lakes', in *Data of Geochemistry*, 6th edition, US Geol. Survey Prof. Paper 440-G, Chapter G, 1963.
36. R.R. Rao, D.G. Goski, A. Chatt, 'Determination of Trace Elements in Acid-rain by Reversed-phase Extraction Chromatography and Neutron Activation', *J. Radioanal. Nucl. Chem.*, **161**(1), 88–99 (1992).
37. P.I. Beazley, R.R. Rao, A. Chatt, 'Preconcentration Neutron Activation Analysis of Trace Elements in Surface Waters by Coprecipitation with Pyrrolidinedithiocarbamate in the Presence of Bi(III)', *J. Radioanal. Nucl. Chem.*, **179**(2), 267–276 (1994).
38. M. Vircavs, 'Organic Disulphides – Collectors for Trace Element Preconcentration', *Rev. Anal. Chem.*, **14**, 167–203 (1995).
39. H. Bem, D.E. Ryan, 'Determination of Seven Trace Elements in Natural Waters by Neutron Activation Analysis After Preconcentration with 1-(2-pyridylazo)-2-naphthol', *Anal. Chim. Acta*, **166**, 189–197 (1984).
40. R.R. Rao, A. Chatt, 'Preconcentration Neutron Activation Analysis of Trace Elements in Seawater by Coprecipitation with 1-(2-thiazolylazo)-2-naphthol,

Pyrrolidinedithiocarbamate and *N*-nitroso-phenylhydroxylamine', *J. Radioanal. Nucl. Chem.*, **168**(2), 439–448 (1993).
41. G.D. Wals, H.A. Das, H.A. van der Sloot, 'The Determination of Inorganic Arsenic in Water by Thermal Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **57**(1), 215–222 (1980).
42. S. Nagatsuka, Y. Tanizaki, 'The Determination of Trace Elements in Tonegawa River Water by Neutron Activation Analysis', *Radioisotopes*, **25**(7), 375–379 (1976).
43. J. Holzbecher, D.E. Ryan, 'Determination of Trace Metals by Neutron Activation after Coprecipitation with Lead Phosphate', *J. Radioanal. Chem.*, **74**(1), 25–30 (1982).
44. S.J. Yeh, J.C. Wei, C.N. Ke, C.Y. Yang, 'Determination of Trace Elements in Surface Water by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **179**(2), 221–229 (1994).
45. P. Rama Devi, T. Gangaiah, G.R.K. Naidu, 'Determination of Trace Metals in Water by Neutron Activation Analysis after Preconcentration on Poly (acrylamidoxime) Resin', *Anal. Chim. Acta*, **249**(2), 533–537 (1991).
46. W. Susetyo, F. Lahagu, Sigit, H.A. Das, 'Evaluation of the Determination of Iodine in Drinking Well Water by Neutron Activation Analysis and Separation by Isotope Exchange', *J. Radioanal. Nucl. Chem.*, **164**(6), 373–380 (1992).
47. A. Grauby, H. Augier, R. Lion, O. Charlent, 'Neutron Activation Analysis of Elemental Composition in a Marine Phanerogam, Posidonia-Oceanica (L) Delile – A Biological Indicator of Pollution', *Environ. Exptl. Botany*, **31**(3), 255–265 (1991).
48. M.R. Karagas, J.S. Morris, J.E. Weiss, V. Spate, C. Baskett, E.R. Greenberg, 'Toenail Samples as an Indicator of Drinking Water Arsenic Exposure', *Cancer Epidemiology Biomarkers Prevention*, **5**(10), 849–852 (1996).
49. S.P. Stanescu, S. Spiridon, 'An Improved Method for Analysis of In-EDTA, I- and Br- Used as Activable Tracers', *J. Radioanal. Nucl. Chem.*, **163**(2), 301–305 (1992).
50. R. Zeisler, N. Haselberger, M. Hakarewicz, R. Ogris, R.M. Parr, S.F. Stone, O. Valkovic, V. Valkovic, E. Wehrstein, 'Nuclear Techniques Applied to Air Particulate Matter Studies', *J. Radioanal. Nucl. Chem.*, **217**(1), 5–10 (1997).
51. J.C. Chow, 'Measurement Methods to Determine Compliance with Ambient Air Quality Standards for Suspended Particles', *J. Air Waste Management Assoc.*, **45**, 320–382 (1995).
52. An.A. Kist, A.A. Kist, L.I. Zhuk, 'Passive Sorption Sampling Method with Neutron Activation Analysis Seems to Offer Perspective in Atmospheric Trace Element Investigation', *J. Radioanal. Nucl. Chem.*, **167**(2), 227–235 (1993).
53. I.M. Kolthoff, P.J. Elving, V. Krivan (eds.), 'Nuclear Activation and Radioisotope Methods of Analysis', *Treatise on Analytical Chemistry*, 2nd edition, John Wiley & Sons, New York, Part I, Vol. 14, Section K, 1986.
54. S. Landsberger, D. Wu, 'Improvement of Analytical Sensitivities for the Determination of Antimony, Arsenic, Cadmium, Indium, Iodine, Molybdenum, Silicon and Uranium in Airborne Particulate Matter by Epithermal Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **167**(2), 219–222 (1993).
55. A. Alian, B. Sansoni, 'A Review on Activation Analysis of Air Particulate Matter', *J. Radioanal. Nucl. Chem.*, **89**(1), 191–275 (1985).
56. H. Rausch, I.L. Sziklai, J. Borossay, K. Torkos, T. Rikker, E Zemplen-Papp, 'Distribution of Toxic and Radiation Components in Air Particulates', *Sci. Total Environ.*, **173/174**, 283–291 (1995).
57. S. Landsberger, D. Wu, 'The Impact of Heavy Metals from Environmental Tobacco-smoke Indoor Air-quality as Determined by Compton Suppression Neutron Activation Analysis', *Sci. Total Environ.*, **173**(1–6), 323–337 (1995).
58. G. Capannesi, A. Cecchi, C. Ciavola, A.F. Sedda, 'Feasibility of Oak Leaves as Monitor for Airborne Pollution', *J. Radioanal. Nucl. Chem.*, **167**(2), 309–320 (1993).
59. E. Steinnes, J.P. Rambaek, J.E. Hanssen, 'Large Scale Multielement Survey of Atmospheric Deposition using Naturally Growing Moss as a Biomonitor', *Chemosphere*, **25**, 735–752 (1992).
60. J.E. Sloof, H.Th. Wolterbeck, 'Patterns in Trace Elements in Lichens', *Water Air Soil Pollut.*, **57–58**, 785–795 (1991).
61. L. Zikovsky, M. Badillo, 'An Indirect Study of Air Pollution by Neutron Activation Analysis of Snow', *J. Radioanal. Nucl. Chem.*, **114**(1), 147–153 (1987).
62. M.B. McBride, B.K. Richards, T. Steenhuis, J.J. Russo, S. Sauve, 'Mobility and Solubility of Toxic Metals and Nutrients in Soil Fifteen Years after Sludge Application', *Soil Sci.*, **62**(7), 487–500 (1997).
63. P.W.B. Friske, E.H.W. Hornbrook, 'Canada National Geochemical Reconnaissance Program', *Trans. Inst. Mining Metallurgy Sect. B- Appl. Earth Sci.*, **100**, B47–B56 (1991).
64. J.R. Krezoski, M. Oladipo, L. Adjarova, C. Shick, T. Tisue, 'Cesium-labelled Phlogopite as a Tracer for In-situ Studies of Sediment Burial, Reworking, and Resuspension Rates', *Int. J. Environ. Anal. Chem.*, **59**(2–4), 213–223 (1995).
65. W.S. Lyon, 'Trace Element Measurements at the Coal-fired Steam Plant', CRC Press, Boca Raton, FL, 1977.
66. R.E. Jervis, S.S. Krishnan, M.M. Ko, L.D. Vela, T.G. Pringle, A.C. Chan, L. Xing, 'Biological Incinerator Emissions of Toxic Inorganics, Their Residues and Their Availability', *Analyst*, **120**, 651–657 (1995).
67. M.R. Wormald, C.G. Clayton, 'In-situ Analysis of Coal by Measurement of Neutron Induced Prompt γ-rays', *Int. J. Appl. Radiat. Isot.*, **34**(1), 71–82 (1983).

68. Y. Tanizaki, T. Maeno, M. Nakamura, N. Yamazaki, 'Use of an Ultrafiltration and Gamma-ray Irradiation Technique for Studying the Speciation of Trace Elements in River Waters', *J. Environ. Sci. Health*, **A31**(4), 913–940 (1996).
69. J.T. Vanelteren, N.G. Haselager, H.A. Das, C.L. Deligny, J. Agterdenbos, 'Determination of Arsenate in Aqueous Samples by Precipitation of the Arsenic(V) Molybdate Complex with Tetraphenylphosphonium Chloride and Neutron Activation Analysis or Hydride Generation Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **252**(1–2), 89–95 (1991).
70. C.R. Lan, C.L. Tseng, M.H. Yang, Z.B. Alfassi, '2-Step Coprecipitation Method for Differentiating Chromium Species in Water Followed by Determination of Chromium by Neutron Activation Analysis', *Analyst*, **116**(1), 35–38 (1991).
71. G.J. Evans, K.A. Hammad, 'Radioanalytical Studies of Iodine Behaviour in the Environment', *J. Radioanal. Nucl. Chem.*, **192**(2), 239–247 (1995).
72. A.A. Kist, 'Investigation of Element Speciation in Atmosphere', *Biol. Trace Element Res.*, **43–45**, 259–266 (1994).
73. S. Yoshida, Y. Muramatsu, 'Determination of Organic, Inorganic and Particulate Iodine in the Coastal Atmosphere of Japan', *J. Radioanal. Nucl. Chem.*, **196**(2), 295–302 (1995).

Nitroaromatics, Environmental Analysis of

Torsten C. Schmidt
Swiss Federal Institute for Environmental Science and Technology (EAWAG), Dübendorf, Switzerland

Klaus Steinbach and Ulf Bütehorn
Philipps-University, Marburg, Germany

1 Introduction **1**
2 Official Methods of Analysis **2**
3 Sample Preparation **5**
3.1 Water 5
3.2 Soil and Particulates 6
4 Methods Without Analyte Separation and In Situ Techniques **7**
4.1 Chemical Methods 7
4.2 Electrochemical Methods 8
4.3 Immunoassays and Sensors 9
5 Separation Techniques **10**
5.1 Gas Chromatography 10
5.2 High-pressure Liquid Chromatography 12
5.3 Capillary Electrophoresis and Related Techniques 14
5.4 Thin-layer Chromatography 15
6 Outlook **15**
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **17**
References **17**

Nitroaromatic compounds (NACs) are important anthropogenic chemicals, which may enter all areas of the environment in significant amounts. Once there, they may be transformed, mainly via redox reactions. Most NACs and their transformation products are rather toxic, both in humans and ecosystems, and they are especially known for their carcinogenic potential. Thus, input and detrimental effects of NACs make environmental surveillance necessary, which in turn requires selective, sensitive, unbiased and robust methods suitable for the analysis of complex mixtures. These requirements are mainly fulfilled by the combination of an enrichment step with chromatographic separation and use of an appropriate detector. Official methods today use liquid–liquid extraction (LLE) or solid-phase extraction (SPE), followed by high-performance liquid chromatography (HPLC) with ultraviolet (UV) detection, or gas chromatography (GC) with electron capture detection (ECD). These methods are widely established and have been thoroughly tested in interlaboratory comparisons. However, new methods for extraction, separation and detection are emerging, which are superior in terms of the required characteristics, and thus may replace the existing methods in the future. Both standard methods and new developments for the environmental analysis of NACs are discussed in this article. Regarding the references given, priority was given to the most recent developments, and thus about 70% of the cited literature is less than five years old.

1 INTRODUCTION

For decades, NACs have been among the most widely used anthropogenic chemicals. Nitration of aromatic compounds is one of the key industrial processes, since the nitro group offers a variety of further functionalizations. Thus NACs are used as solvents, pesticides, explosives and fragrances and are intermediates in the production of pharmaceuticals, dyes, polymers, whitening agents and photochemicals.[1] Seventy NACs are currently listed as OECD (Organisation for Economic Co-operation and Development) high-production-volume chemicals, with a production of more than 1000 t per year in at least one member country. Of the parent compound nitrobenzene alone, 740 000 t were produced in the United States in 1994, mainly as a source for aniline. Emissions of NACs can occur during the production, the use and the disposal of the compounds. The Toxics Release Inventory of the United States Environmental Protection Agency (USEPA) estimates that the total emission of nitrobenzene to the environment in the United States was 320 t in 1997. In 53 of the USEPA Superfund Sites NACs are contaminants of concern. In addition to direct emissions, it has become obvious in recent years that NACs may also be formed in the atmosphere due to the reaction of nitrogen oxides with aromatics. This reaction is important both for polycyclic aromatic hydrocarbons (PAHs) to yield nitro-PAHs[2] and for benzenes to yield mainly nitrated phenols.[3]

There are hardly any naturally occurring NACs. One exception is aristolochic acid (Figure 1), which was used as a drug ingredient before its toxic effects were known. The molecular formulae of some important NACs are given in Figure 1. It can be seen that many compounds have more than one functional group governing their

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

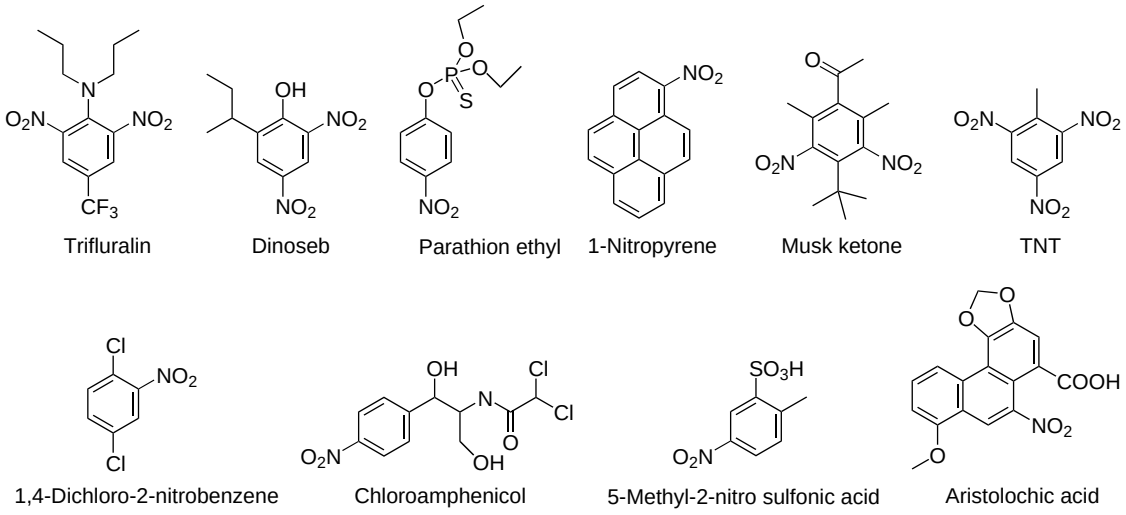

Figure 1 Molecular formulae of selected NAC.

behavior, therefore some of the compounds may also be covered in other articles.

Table 1 gives the CAS no. and some important properties of nitrobenzene, 2,4-dinitrotoluene and the compounds shown in Figure 1. As might be expected from the diversity of the compounds, their physicochemical properties differ widely. In general, NACs are electron-acceptor compounds and the acidity of aromatic acids is much increased because of the electron-withdrawing nitro group. The introduction of a nitro group does not significantly alter the water solubility of nonionogenic compounds. Many NACs are known to be highly toxic and are proven or suspected carcinogens.[5,6] In contrast, the easily water-soluble nitrosulfonic acids do not show significant toxicity. However, their mobility and persistence in water make them possible drinking-water contaminants. A very common reaction in soil and water is the abiotic or biotic reduction of the nitro group to an amino group via three two-electron transfer steps. The corresponding anilines, which may be of even more environmental and toxicological concern, thus have to be considered in many cases. An extensive treatment on biodegradation of NACs can be found in the book by Spain.[7] The complexity of wastewater generated during the manufacture of trinitrotoluene was demonstrated by Spanggord et al.[8] Rosenblatt et al.[9] showed the multitude of products from phototransformations of NACs, including for the first time nitrobenzoic acids resulting from photooxidation of the methyl group. Recently, these and other ionic compounds were found in groundwater[10,11] and soil[12] samples, which suggests other mechanisms in effect besides photooxidation. In summary, most NACs require monitoring due to their high production, their toxicity and their behavior in the environment.

2 OFFICIAL METHODS OF ANALYSIS

To the best of our knowledge, NACs are not included in legislatory regulations as such. However, there might be regulations resulting from the use of specific NACs, e.g. as pesticides. These are not discussed further here since they are covered in other articles. Furthermore, the USEPA has issued health advisory levels for several explosive-related NACs, which are given by Tomkins in **Explosives Analysis in the Environment**. Despite the lack of general regulations, maximum NAC emissions from factories and contents in water and soil at former industrial sites might have been specified. For example, at the former ammunition plant in Stadtallendorf, Germany, remediation aims of 20–80 mg NAC per kg soil, depending on the proposed use of the area, were constituted by regional authorities.

Official methods of analysis for NACs are available for the determination of explosives, dinitroaniline pesticides and for a number of NACs in a general analytical scheme for semivolatile compounds. Table 2 briefly summarizes these methods.

As shown in Table 2 the official USEPA method 8030 has also been adopted for water analysis. Sample preparation involves a salting-out assisted solvent extraction with acetonitrile, as developed by Leggett et al.[13] In the new draft method 8095 this procedure is replaced by SPE which needs considerably less organic solvent. The official method requires primary analysis on a C_{18} column with confirmation on a second CN column since a single-wavelength UV detector is used.

Both the German DIN method 38407-17 and the USEPA draft method for explosives utilize GC for separation. The USEPA method includes even very

Table 1 Physicochemical data of selected NACs

Substance	CAS no.	Use	Molecular weight ($g\,mol^{-1}$)	Density ($kg\,L^{-1}$)	Melting point (°C)	Vapor pressure (Pa)	Henry constant c_{air}/c_{water}	Water solubility ($g\,L^{-1}$)	Log K_{ow}	Production	Toxicity data
2,4-DNT	121-14-2	Important industrial chemical, e.g. intermediate in polymer production	182.14	1.3208	70	2.9×10^{-2}	1.84×10^{-4}	2.7×10^{-1}	1.98	HPV	r: C s: N, R, L, B, K
5-Nitro-2-methyl sulfonic acid	121-03-9	Intermediate in the production of whitening agents	217.20	n.a.	134	n.a.	n.a.	660	−0.8[a]	HPV	None identified
1,4-Dichloro-2-nitrobenzene	89-61-2	Intermediate in dye production	192.00	1.439	55	1.37×10^{-4}	n.a.		3.09	HPV	n.a.
Musk ketone	81-14-1	Fragrance	294.31	1.28	135	n.a.	n.a.	<1	4.31[a]	HPV	n.a.
Trifluralin	1582-09-8	Herbicide	335.28	1.294	48.5	1.46×10^{-2}	6.2×10^{-4}	3.2×10^{-4}	5.34		n.a.
Dinoseb	88-85-7	Herbicide	240.22	1.29	45	1.3×10^{-4}	2.4×10^{-7}	5.2×10^{-2}	2.29	HPV	r: D, R s: C, N, Resp
Parathion	56-38-2	Insecticide	291.26	1.26	6	8.9×10^{-4}	9.7×10^{-6}	2.4×10^{-2}	3.83	HPV	s: N, D, R, Resp, E, B, S
2,4,6-TNT	118-96-7	Explosive	227.15	1.654	81	0.85×10^{-3}	1.1×10^{-6}	0.11	1.6	HPV	s: N, Resp, L, B, S
1-Nitropyrene	5522-43-0	Combustion product from fuels	247.26	n.a.	155	n.a.	n.a.	<1	4.68[a]		r: C
Nitrobenzene	98-95-3	Aniline production, solvent	123.11	1.196	6	41	9.5×10^{-4}	1.9	1.83	HPV	r: C s: N, R, Resp, L, B, K
Chloroamphenicol	56-75-7	Antibacterial drug	323.13	n.a.	149	n.a.	n.a.	2.5	1.14		r: C s: N, D, Resp, L, B, S

n.a., not available; CAS, chemical abstracts service; DNT, dinitrotoluene; HPV, high production volume chemical; 2,4,6-TNT, 2,4,6-trinitrotoluene.
Toxicity data abbreviations: r, recognized; s, suspected; C, carcinogenic; N, neurotoxic; R, reproductive toxicity; D, developmental toxicity; Resp, respiratory toxicity; E, endocrine toxicity; L, liver toxicity; B, blood toxicity; K, kidney toxicity; S, skin toxicity. All toxicity data were obtained from Scorecard (http://www.scorecard.org).
[a] Estimated with fragment constant method as described in Lyman et al.[4]

Table 2 Official methods for the analysis of NACs

Method	Analytes	Matrix	Sample preparation	Separation and detection
USEPA SW 846/8030 (http://www.epa.gov/SW-846/8330.pdf)	14 explosives-related compounds, including 12 NACs	Soil (also adapted for water)	Ultrasonic extraction with acetonitrile (soil); salting-out extraction with acetonitrile (water)	HPLC/UV
USEPA draft 8095	17 explosives-related compounds, including 13 NACs	Soil/water	Acetonitrile extract (soil); SPE with Sep-Pak™ C18 according to USEPA SW 846/3535 (water)	GC/ECD
DIN 38407-17	12 explosives-related NAC	Water	SPE with RP18 or LLE with toluene	GC/ECD or GC/PND
USEPA 627	Dinitroaniline pesticides	Wastewater	LLE with dichloromethane	GC/ECD
USEPA SW 846/8270 (http://www.epa.gov/SW-846/8270c.pdf)	Organic compounds, including at least 21 NACs	Soil/water	According to USEPA 3500 methods	GC/MS
USEPA TO-4	Organic compounds, including parathion and trifluralin	Air	Sampling of 288–403 m^3 air on filters	GC/PND or GC/ECD

DIN, Deutsches Institut für Normung; GC/ECD, gas chromatography/electron capture detection; GC/MS, gas chromatography/mass spectrometry; PND, phosphorus–nitrogen detector.

thermolabile nitramines and is therefore carried out on a short, wide-bore column (6 m × 0.53 mm) of low polarity (DB™-1 or DB™-5 or equivalents) for primary analysis and a mid-polar column (RTX™-200 or RTX™-225 or equivalents) for confirmatory analysis. The DIN method also requires a confirmatory analysis on a column of different polarity, but the recommended columns have an internal diameter of 0.2–0.4 mm and a length of 20–60 m. Suggested stationary phases range in polarity from DB™-1 to DB™-1701 or equivalents.

The standardized methods can help the practitioner in the fast establishment of a new method and are often required by regulatory bodies. The main advantage of these methods is the assurance of data quality by thorough testing and interlaboratory comparisons.[14] However, it should be emphasized that their use also poses some problems. Usually, the standardization requires several years before being officially adopted. Considering the rapid developments in the area of instrumental analysis they therefore often demand the use of rather outdated methods or equipment. As an example, USEPA method 8030 for the analysis of nitroaromatic explosives in soil and water is discussed here. The processing of water samples requires the use of salting-out extraction with 175 mL acetonitrile per sample. Some 97% of the acetonitrile remains in the aqueous phase and has to be disposed of with the sample. Furthermore, the extraction is very time-consuming, requiring several extraction/back-extraction steps. Modern SPE methods are able to provide at least comparable extraction efficiencies with much less solvent and time consumption. For soil samples an 18 h acetonitrile extraction in an ultrasonic bath is used. Here it has already been shown that faster methods such as accelerated solvent extraction (ASE) or supercritical fluid extraction (SFE), which are rather easy to automate and consume little solvent, are applicable. Nevertheless, none of these methods are included in the new draft method of the USEPA. In the determination step a primary column with an internal diameter of 4.6 mm is recommended. Again, it has been known for a long time that 4- and even 3-mm columns can be used with almost all HPLC equipment, providing even better separation in many cases, with much less consumption of the mobile phase. Loadability of the smaller columns is lower but in general this is not a severe problem in environmental analysis. Detection is done with a single-wavelength UV detector at 254 nm. To confirm positive results therefore requires a second column of a different selectivity. However, in most HPLC laboratories nowadays a diode-array detector (DAD) is in use which allows the on-line confirmation of peak identity with the comparison of UV spectra. This of course requires much less time than a second run for the same sample. Because of the drawbacks of the official method, several method adjustments were recently published to improve its performance.[15,16]

The USEPA method was selected as an example for outdated procedures because it is still in use worldwide,

but, as already mentioned here, the general problem exists for all standard methods.

3 SAMPLE PREPARATION

3.1 Water

3.1.1 Sampling and Storage

Water samples should be taken in precleaned brown-glass bottles. Bottles and stoppers should be rinsed several times with the sample on-site. Since some of the analytes are volatile the bottles should be filled completely, without leaving a headspace. Extraction of water samples should be done as soon as possible after sampling. If storage cannot be avoided the samples should be stored at temperatures below 4 °C. Even then losses are possible with prolonged storage times. The loss is mainly due to microbial degradation, hence the stability is enhanced if growth of microorganisms is inhibited. For that purpose, sodium azide and mercury(II) chloride are mainly used, but the use of mercury salts is not recommended due to problems with the disposal of the samples after processing. A 1-mL portion of a saturated solution of sodium azide in water is sufficient to depress microbial activity in a 1-L sample.

3.1.2 Extraction

A brief overview of important characteristics of some extraction techniques for aqueous samples is given in Table 3. More detailed information can be found in the following sections and the cited literature.

Liquid–Liquid Extraction Several solvents are suitable for LLE of NACs. Most often, toluene (as in DIN 38407-17) and dichloromethane (as in USEPA method 625) are used.[17] In general, the choice between the two is partly determined by the detection method. For GC/ECD, toluene has to be used, while for the GC/MS and HPLC methods, dichloromethane is preferred (for HPLC this is because solvent exchange is easier with the more volatile dichloromethane). In the DIN method the extraction is done with two portions of 20 mL toluene (with 10 min of shaking in a separatory funnel) for a water sample of 1 L after adjusting the sample pH to between 6 and 8. In contrast, the dichloromethane extraction is done after salting-out of the sample with 100 g sodium chloride (for 1 L water). It is then extracted three times with 50 mL dichloromethane. In both cases, the extracts are dried over anhydrous sodium sulfate, filtered and reduced in volume under reduced pressure. Alternatively, the approach of Leggett et al.[13] (as used in the adaptation of USEPA method 8030 for water) might be used. They extracted with acetonitrile and achieved a phase separation between water and acetonitrile with the addition of high amounts of salt (see section 2). A general disadvantage of all LLE procedures is the use of rather high amounts of organic solvents, especially when more polar degradation products have to be included. An attractive further development in LLE is the use of a micro-extraction technique as described by Welsch and Block,[18] who found satisfactory recoveries (>70%) for nitroaromatic explosives and reduction products with the use of only 4.5 mL methyl *tert*-butyl ether in a multiple extraction scheme. Limit of detection (LOD) values given were in the low parts per trillion (ppt) range for di- and trinitroaromatics and in the high ppt range for mononitroaromatics.

Solid-phase Extraction SPE has increasingly replaced LLE in the analysis of NACs in water over the last decade. This is due to the reduction in organic solvent use, the high recoveries obtained in many cases, the ease of sample clean-up and the possibility of coupling SPE on-line with chromatographic separation. SPE procedures are included in both the DIN method and the new USEPA draft method for NACs in water. In both cases a C_{18} phase is used for that purpose. As an example, the DIN method is described in more detail here. A 2-g sample of the solid phase (for 1 L sample volume) is conditioned with 15 mL of methanol and washed with 15 mL of water.

Table 3 Characteristics of extraction procedures for aqueous samples

Method	Typical sample volume (L)	Typical extraction time (h)	Solvent consumption (mL)	Cost of equipment (apparatus/consumables)	Automation
LLE	1	1–2	40–200	++/o	−
Cartridge SPE	0.1–1	0.5–2	3–10	+/o	+
Disk SPE	1	0.3–1	3–10	o/−	o
SPME	0.01	0.2–1	0	+/+ (off-line) −/+ (on-line)	− (HPLC) ++ (GC)

SPME, solid-phase microextraction.

The water sample is salted out with 100 g of sodium chloride per liter of sample and the pH is adjusted to between 6 and 8. The extraction is performed with a flow of about 16 mL min^{-1}. Afterwards, residual water is blown off with nitrogen and the analytes eluted with 6 mL of ethyl acetate. If necessary, to achieve lower detection limits, the eluent is reduced in volume to 1 mL. C_{18} phases were described most often in experimental work carried out at the end of the 1980s and the beginning of the 1990s. However, their use in the analysis of NACs is already declining because of the commercial advent of polymeric phases, which offer higher capacities, a better retention of more polar compounds and enhanced pH stability, which is important for the analysis of degradation products. Styrene–divinylbenzene copolymers, which are available from most manufacturers of SPE materials, are mainly used for the extraction of NACs.[19] The extraction can easily be automated, as has been shown for explosives,[19,20] as well as for nitroaromatic pesticides.[21] Solid phases in disk geometry allow higher flows than packed cartridges and have less susceptibility to clogging, which are distinct advantages, especially for the extraction of large volumes of aqueous samples. In a recent monitoring study for selected pesticides (including NACs) in groundwater, a flow of 30 mL min^{-1} was used without breakthrough of analytes or pressure increase due to clogging.[22] Disk-type solid phases are already employed in the general USEPA method for performing SPE (SW 846/3535), and laboratories using cartridge types have to prove comparable performance.

An interesting new approach for the selective enrichment of NACs was described by Bucheli et al.[23] It is based on the interaction between electron-deficient aromatic rings and electron-rich clay minerals, which leads to the formation of electron donor–acceptor (EDA) complexes. These are strongest for polynitroaromatics, while for mononitroaromatics they are less efficient than the previously mentioned polymeric phases. The method has been proved to be suitable for routine analysis and on-line SPE.

Solid-phase Microextraction SPME is a relatively new, solvent-free extraction method that so far has rarely been described for the enrichment of NACs from water. However, there are reports on the extraction of nitrophenols,[24] dinitroaniline pesticides,[25] and several explosives, including trinitrotoluene and DNT, in seawater.[26] These reports have shown the potential of SPME, especially in conjunction with GC/MS. In general, SPME is still limited to GC separations in routine analysis and it requires the use of internal standards (preferably isotopically labeled ones) for accurate quantitation, because it is a nonequilibrium method.

3.2 Soil and Particulates

3.2.1 Sampling and Storage

A thorough treatment of the representativeness of soil sampling for explosives analysis has been carried out by Tomkins, **Explosives Analysis in the Environment**.[12] Briefly, a well-designed sampling plan is required to ensure the quality of spatially resolved information on contamination levels. Even inside a sample it is often difficult or impossible to obtain a homogeneous distribution of the analytes. Thus it is also important to make sure that a representative subsample is used for further analysis. Soil samples can be refrigerated without significant loss of analytes for several weeks but should be kept in the dark to prevent photodegradation.

For the sampling of particulates, high-volume air samplers are most frequently employed, which suck about 2000 m^3 air through a filter. The choice of an appropriate filter type is crucial in proper sampling.[27]

3.2.2 Extraction

There is a wide range of methods available for the extraction of soil samples. The classical approach is either extraction at elevated temperatures in a Soxhlet apparatus or mechanical shaking of the solid with an appropriate solvent. More recently, these methods have been partly replaced by extraction in an ultrasonic bath. These three methods and the use of a homogenizer–sonicator were compared by Jenkins and Grant[28] for the analysis of explosives in soil. They found extraction in an ultrasonic bath to be superior to the other methods. Lately, SFE and ASE have attracted considerable attention, and for some NACs there are applications for their use in extractions of solids. A brief overview of the important characteristics of some extraction techniques for solids is given in Table 4. More detailed information can be found in the following sections and the cited literature.

Soxhlet Extraction Even today, Soxhlet extraction is the standard method in many application areas, for example for the extraction of diesel emission particulates on filter paper.[27] Several solvents are commonly used, such as diethylether, dichloromethane, cyclohexane, and toluene. Recoveries are high for many compounds, but the temperature must be carefully controlled because many nitroaromatics are thermolabile. The distinct disadvantages of the Soxhlet method are its very high solvent consumption, usually 100–500 mL per sample, and the long extraction times needed (up to 48 h is common). The time of each cycle varies between 15 and 30 min.

Ultrasonic Treatment Ultrasonic treatment for the extraction of nitro-aromatic explosives was thoroughly

Table 4 Characteristics of extraction procedures for soil and particulates samples

Method	Typical sample mass (g)	Typical extraction time (h)	Solvent consumption (mL)	Cost of equipment (apparatus/consumables)	Automation
Soxhlet	5–20	2–48	100–500	o/−	o
Ultrasonic bath	5–10	1–18	5–50	o/+	−
ASE	5	0.5–4	10–50	−/o	+
SFE	5	0.5–1	1–2 (modifier)	−−/−	+

investigated by Jenkins et al.,[28,29] and shown to give reproducible results in an interlaboratory comparison.[30] As a result of these studies, the use of an ultrasonic bath in soil extraction became mandatory in USEPA method 8330 for explosives. In this method a 2 g subsample is extracted with 10 mL aceto-nitrile for 18 h. The fine particles are then flocculated with a calcium chloride solution (concentration = 5 g L^{-1}). After 15 min settling time the sample is filtered through a 0.5 μm filter. An internal standard should be added to the extraction solvent. In this method, acetonitrile is chosen as the extractant because of its better solvation of the nitramines, while, for the NACs, methanol gave comparable results regarding extraction efficiency. For other NACs, ethyl acetate, ethanol and dichloromethane have also been used. A general advantage of the ultrasonication technique is the possibility of processing several samples at the same time. A disadvantage is that it is not well suited for automation.

Accelerated Solvent Extraction In the past few years accelerated extraction from solids at elevated pressures and temperatures has been employed in many fields. Surprisingly there was only one report in the open literature on the use of this method for NAC extraction.[31] The authors compared six different extraction techniques for nitroaromatic explosives in real soil samples from a former ammunition plant. They found ASE with methanol to be comparable to standard methods in terms of extraction efficiency and reproducibility, but superior in terms of solvent consumption and extraction time. Saito and Yamaguchi[27] expect ASE to replace both Soxhlet and ultrasonic extraction for the analysis of particulates in the future.

Supercritical Fluid Extraction In contrast to the more recently introduced ASE, SFE employs mainly CO_2 as the extracting medium. CO_2 is chosen because of its favorable critical parameters (i.e. critical pressure 74 bar, critical temperature 32 °C) which help to keep it in the supercritical state. Lopez-Avila and Beckert prepared a thorough overview on the use of SFE in environmental analysis, including the extraction of 36 NACs from sand and a reference soil.[32] In the last few years the influence of different modifiers on the extraction efficiency has been intensively investigated.[33–35] For trinitrotoluene and trinitrobenzene Wujcik and Seiber[34] found the following conditions to be optimal: 5% methanol, 150 °C, 517 bar (7500 psi), 5 min static and 15 min dynamic extraction. Deuster et al.[35] emphasized that direct addition of the liquid modifier to the sample was superior to the use of premixed CO_2/modifier. Toluene and acetonitrile were comparable in enhancing the extraction efficiency of pure CO_2 for nine NACs, whereas methanol was less effective. The optimized conditions were: 2 mL modifier, 90 °C, 400 bar, 15 min static and 15 min dynamic extraction. SFE was also found to be suitable for the extraction of parathion and 4-nitrophenol from soil[36] and it is assumed that SFE should be similarly suited to the analysis of nitro-PAHs.[27]

4 METHODS WITHOUT ANALYTE SEPARATION AND IN SITU TECHNIQUES

4.1 Chemical Methods

Numerous chemical reactions of NACs leading to colored products are described in the literature. Such reactions are the basis for spot tests, spraying reagents for thin-layer chromatography (TLC) and titration procedures. Several spot tests on NACs are described by Jungreis and Ben-Dor.[37] The principles of the tests are: transformation of a polynitroaromatic with strong base to a highly colored Jackson–Meisenheimer anion or quinoidal form (Janovsky reaction); derivatization with phenylacetonitrile in an alkaline dimethylformamide (DMF) medium to a blue- or yellow-colored addition product; formation of a charge-transfer complex with *N*,*N*-diethylaniline.

Of these, only the Janovsky reaction has so far been employed in environmental analysis. A procedure reported by Medary[38] uses methanol as the extraction solvent for soil samples. The absorbance of the purple color produced by addition of sodium hydroxide to the soil extracts is measured at 516 nm and found to be directly proportional to the concentration of trinitrotoluene. The estimated detection limit of the method is 4–8 ppm.

The same reaction was utilized for a field-test kit for trinitrotoluene allowing semiquantitative measurements.[39] To achieve this, small glass columns are filled

with a polystyrene-based anionic exchanger with a thin layer of a solid base (NaOH, CaO or MgO) on top. In the field, 5–50 mL of water or a methanolic soil extract is injected with a disposable syringe. The colored Jackson–Meisenheimer anions are formed in contact with the base and extracted with the anion exchanger. The length of the colored zone in the column differs depending on the amount of trinitrotoluene present. As might have been expected, there are a number of interferences, for example due to other organic or inorganic anions in varying concentrations.

The adaptation of the Janovsky reaction to an automated sequential injection system was recently reported for the rapid quantitative determination of TNT in soil.[40] Acetone extracts of the soil samples were injected with a reagent solution of potassium hydroxide and sodium sulfite into a carrier stream (acetone/water 88/12). After a reaction time of 30 s the mixture was dispensed by the carrier through a flow cell of a spectrophotometer and the absorbance was measured at 470 nm. It was shown that the method is selective for TNT and had only little or no response to mono- and dinitroaromatic compounds. The detection limit was 0.5 $\mu g\,mL^{-1}$ trinitrotoluene in solution.

Another possibility for a spot test is the facile reduction to the corresponding aromatic amines, which can subsequently be derivatized to form highly colored azo dyes.[41] If the dyes are extracted with *n*-butanol a semi-quantitative colorimetric or photometric measurement is possible.

The general advantage of such simple chemical methods is their quick and easy use in the field, which allows a sound selection of samples for further analysis. However, due to their low specificity and sensitivity, their use will be limited to investigations on sites where rather high concentrations of NACs are found.

4.2 Electrochemical Methods

Determination of NACs by electrochemical methods utilizes the reduction of the nitro group. Depending on the pH of a solution, the reduction leads to the corresponding hydroxylamino compounds or anilines, which require 4 e^- or 6 e^-, respectively, for each nitro group. Thus polynitroaromatics in particular can be very sensitively detected, a technique which is also used in electrochemical detectors in HPLC (see section 5.2). With differential pulse polarography (DPP) and adsorptive accumulation methods, detection limits in the low $\mu g\,L^{-1}$ range are achievable. Voltammetric methods have been intensively investigated for the determination of NACs, but there are few applications in environmental analysis. Recently, it was shown that with DPP tri-, di- and mononitrotoluenes can be distinguished due to their different reduction potentials.[42] After minimizing the differences in reduction potentials of positional isomers, this method was used to establish a sum parameter for rapid determinations in process water from a treatment plant for contaminated soil. The principle of the method is shown in Figure 2. Firstly, the concentration of TNT is calculated from H_{SI} by comparison with a TNT standard H_{PI}. The contribution of this TNT concentration to SII is calculated and subtracted from the overall peak height to yield H_{SII}, which gives the concentration of DNTs. The same procedure is repeated for SIII and the mononitrotoluene concentration is calculated from H_{SIII}. Method validation with a standard HPLC method showed good agreement of the results.

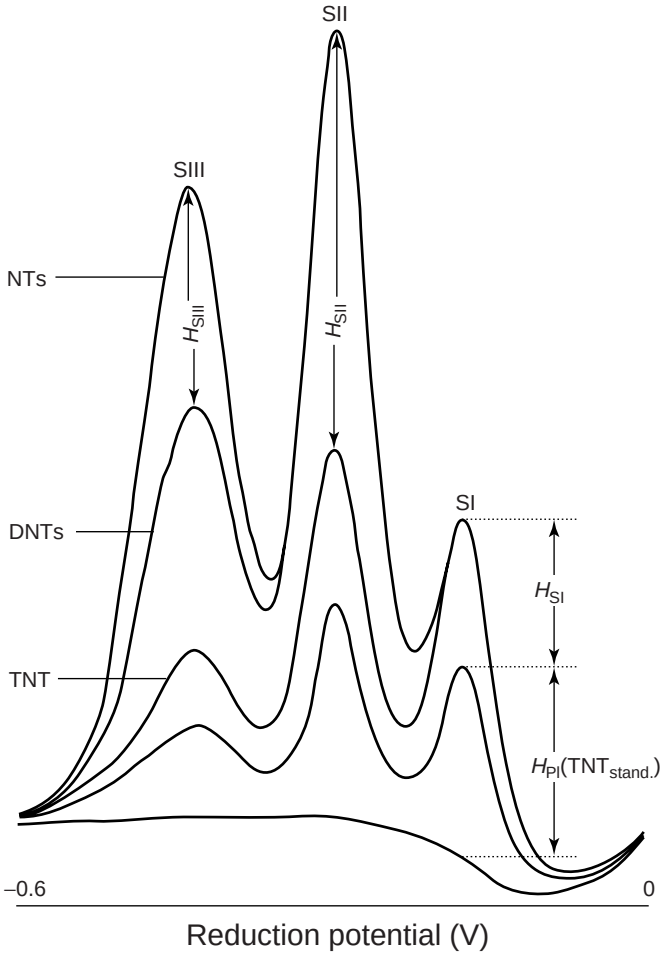

Figure 2 Voltammetric detection of nitrotoluenes. Reduction at different potentials allows quantification of tri-, di- and mononitrotoluenes. SI to SIII: sum peaks for tri-; di- and tri-; mono-, di- and trinitrotoluenes, respectively. H_{SI} to H_{SIII}: peak height for tri-, di- and mononitrotoluenes, respectively. (Adapted from Weibels[42] by permission of Görich & Weiershäuset, Marburg.)

4.3 Immunoassays and Sensors

Immunochemical methods have been successfully used in biochemical and biomedical analysis for many years. In the last decade they have received increasing acceptance in environmental analysis. Enzyme-linked immunosorbent assays (ELISAs) are generally faster and less expensive than conventional analytical methods and require only simple equipment. They are highly specific and can be used without or only minimal sample clean-up and are therefore especially suited for field screening tests.

The procedure for a heterogeneous ELISA involves the following steps. First the surface of the test device, in most cases a microtiter plate coated with an analyte antibody, is incubated with the analyte solution and an analyte–enzyme conjugate. In a competitive reaction both can bind to the antibodies on the surface. After a specific reaction time, the system is washed to remove unbound molecules. In the next step the enzyme substrate is added, which is converted to a colored reaction product by the enzyme of the bound analyte–enzyme conjugate. The intensity of the color is inversely proportional to the concentration of the analyte and can be determined photometrically.

Several immunological assays have been developed for the determination of the explosive TNT. A strategy for the validation of commercially available immunoassay test kits is described by Krämer.[43] Eck et al.[44] reported the development of assays for 1,3-dinitroaromatic compounds. The limits of detection for 2,4-DNT, 2,6-DNT, 1,3,5-trinitrobenzene and TNT were $1\,ng\,mL^{-1}$. No cross-reactivity with mononitroaromatic compounds was observed. 4-Nitrophenol and monosubstituted 4-nitrophenols can be determined with an ELISA at levels as low as 0.2 ppb (parts per billion) in water samples without extraction and clean-up.[45]

Keuchel et al.[46] described the development of a sensitive and selective immunoassay and its application for screening of TNT in groundwater. The method has a detection range from 40 ppt to 10 ppm for TNT in water, while di- and mononitroaromatic compounds show cross-reactivities below 2%.

A rapid immunofiltration assay was developed especially for on-site testing, allowing the determination of TNT in water and soil samples within an assay time of about 6 min. The test shows a quantitative color response to concentrations ranging from 1 to $30\,\mu g\,L^{-1}$ TNT in water and $50–1000\,\mu g\,kg^{-1}$ TNT in soil.[47]

Dosch et al.[48] used a homogeneous apoenzyme reactivation immunoassay system (ARIS) for the detection of TNT with a LOD of $5\,\mu g\,L^{-1}$. The system is less sensitive than heterogeneous ELISA methods, but faster and more simple, because it does not require any washing step.

NACs and their biodegradation products can be bound covalently to the organic matter of the soil. These bound residues are nonextractable with organic solvents and are therefore not detectable with common analytical methods. For the quantitation of such residues, immunoassays can be extremely useful. A first example in the literature is the use of a sandwich immunoassay to detect bound trinitrophenyl residues in polluted soil samples after exhaustive Soxhlet extraction.[49] The humic acid–analyte complex (bound NAC residue) is immobilized on the surface of the microtiter plate via nonspecific interactions with a basic protein. The bound trinitrophenyl fraction is then detected by a TNT-selective antibody. In the final step the bound TNT–antibody couples to an enzyme-labeled antibody. The color of the enzyme reaction is directly proportional to the amount of bound analyte.

Most immunoassays have been designed as highly specific methods with only minimal cross-reactivity. In screening tests, however, a high cross-reactivity is useful to detect the analyte, its metabolites and structurally related compounds. Broadening the antibody specificity offers the possibility of developing a group-specific assay for the analysis of 2,4,6-TNT, its precursors and metabolites in soils.[50]

Quantitation of 1-nitropyrene, which is suspected to have a high mutagenic and carcinogenic potential, in diesel soot and urban aerosols with an ELISA was investigated by Fröschl et al.[51] The cross-reactivities with several other NACs possibly present in air particulate matter were determined.

Numerous sensor systems for the detection of NACs, especially of trinitrotoluene, have been reported in the literature. Buttner et al.[52] describe the development of an amperometric sensor for the selective in situ detection and quantification of TNT and other nitrated explosives in soils. Thermal decomposition of the contaminated soil samples by a platinum wire heater (900 °C) produces volatile nitrogen–oxygen compounds which were detected with a sensitive amperometric gas sensor. The system has a detection limit of $0.5\,mg\,kg^{-1}$ for trinitrotoluene in soil.

An electrochemical sensor with rapid square-wave voltammetric scanning was used by Wang et al.[53] for the continuous monitoring of TNT in natural waters with low ppb detection limits. Fast detection of nitro- and aminoaromatic compounds with an electrochemical sensor system based on cyclic voltammetry was reported by Krausa et al.[54]

Several immunosensors have been developed for the detection of TNT on the basis of a competitive immunoassay (Shriver-Lake et al.,[105] Bart et al.[106]). A comparison of the detection limits, the assay time and the run-to-run stability of different sensor system is given by Narang et al.[55]

5 SEPARATION TECHNIQUES

The standard work on chromatographic methods for the separation of nitroaromatic and other explosives was published by Yinon and Zitrin in 1993.[56] Because of their experience, mass spectrometric methods and forensic analysis are emphasized in the book, whereas there is only a brief chapter on environmental analysis.

5.1 Gas Chromatography

Despite the limited thermostability of many NACs, separations with GC are possible without derivatization for all nonionic compounds. GC in fact is very popular for the analysis of all kinds of NACs. The reasons are the high separation efficiency compared with liquid chromatography, the range of detectors, and the ease of ''hyphenation'', especially with mass spectrometry (MS). Furthermore, the presence of the nitro group in analytes does not require the use of special columns, thus nonpolar polydimethylsiloxanes are widely used. However, especially in explosives analysis, short wide-bore columns are often used to prevent thermodegradation on the column. Thus, for example, the USEPA draft method employs a low-polarity $6\,m \times 0.53\,mm$ column for primary analysis. Besides, polynitroaromatics may be thermally degraded in the injector. Therefore the use of a filled liner is not recommended and the liner has to be replaced regularly. No single injection technique was found to be superior for all NACs. Methods in the literature comprise split, splitless, on-column and programmed temperature vaporizer (PTV) injection, depending on solvent type, sensitivity required and availability of instrumentation. If further functional groups are present, they will often determine the choice of column and injector. Yinon recommended the use of a cooled, temperature-programmable injector for the analysis of explosives, including DNTs and TNT, in water samples.[57] Large-volume injection of water into a PTV was used by Müller et al.[58] for the investigation of NACs. After evaporation of water and adsorption of the analytes from the gas phase, they are thermally desorbed in the PTV. An inherent disadvantage of this method is the very low injection rate due to the high enthalpy of vaporization of water, leading to prolonged analysis times.

A short overview of the most important GC detectors for NAC analysis is given in Table 5.

As for all organic compounds, FID may be used as the detection technique. But since a variety of more specific and sensitive detectors for NACs are available, it is seldom used today. In contrast, a modified instrument called a PND, or thermoionic detector (TID), which has an additional glass bead containing an alkali metal, is especially sensitive towards nitrogen- and phosphorus-containing compounds and is thus suited for the analysis of NACs. It is especially useful in the investigation of parathion insecticides, since these contain both elements,[59] but has also been used for the detection of nitro-PAHs in dichloromethane extracts of airborne particulates.[60]

Table 5 Comparison of detector types in GC

Detector	Instrument detection limits (pg)	Selectivity	Notes
FID	10–100	−−	Wide linear range
PND	1–100	o	Allows simultaneous detection of anilines, sensitivity increases with number of nitrogens
ECD	0.1–10	o	Linear range limited, sensitivity increases with number of nitro groups
TEA	10–100	++	Specialized detector for nitro and nitroso compounds, sensitivity increases with number of nitro groups
MS, EI, SIM	1–10	+	Sensitive detection requires SIM mode
MS, NCI, SIM	0.1–1	++	Sensitive detection requires SIM mode, quantification difficult, sensitivity increases with number of nitro groups

FID, flame ionization detection; TEA, thermal energy analysis; EI, electron impact; SIM, selected ion monitoring; NCI, negative chemical ionization.

From the number of publications, ECD seems to be the most popular detection technique for NACs besides MS. This is due to the ease of use and low cost of the detector, which provides adequate selectivity in many cases. The rather high electron affinity of NACs allows a sensitive detection. In general, the sensitivity increases with the number of nitro groups in the compound. A general disadvantage observed by many authors is the limited linear range of the detector. Examples of the use of ECD from the recent literature are for the determination of nitroaromatic pesticides in groundwater[22] and the analysis of nitroaromatic explosives in water samples.[61] The use of ECD is demanded in DIN method 38407-17 and USEPA draft method 8095. The DIN method alternatively allows the use of a PND, and a splitting of the column effluent to provide simultaneous detection is recommended.

The highest selectivity towards nitro (and nitroso) groups is shown by TEA. Since it is not a generally known

GC detection technique, its principle is described here briefly. After separation of the compounds the column effluent is pyrolyzed at 600–1000 °C. The decomposition leads to the formation of NO_2, which is catalytically reduced to NO. After passing a cold trap to remove interfering pyrolysis products, it is oxidized with ozone to yield electronically excited NO_2*. When relaxing to the ground state, light of wavelength 600 nm is emitted, which is detected by a photomultiplier. The use of TEA for the analysis of explosives was thoroughly discussed by Yinon and Zitrin.[56] Although TEA exhibits some unique features it is not widely used in environmental analysis. Probably this results from its high selectivity, which limits its use to nitro and nitroso compounds, and thus it is only available in specialized laboratories.

Maybe the most important detection technique in GC today, because of its general applicability along with its high selectivity, sensitivity and additional spectral information for compound identification or confirmation, is MS. With the standard EI ionization at 70 eV, radical cations are formed (M^+) which subsequently lose an oxygen radical $[M - 16]^+$ or, after rearrangement to the corresponding nitrite ester, nitrogen oxide $[M - 30]^+$. Loss of nitrogen dioxide is also favored, especially for nitrotoluenes, since a stable tropylium cation $[M - 46]^+$ can be formed. Nitrotoluenes are also a classical example of the ''*ortho*-effect'', which often allows the distinction of *ortho*-isomers from *meta*- and *para*-isomers. In *o*-nitrotoluene, interaction between a methyl proton and an oxygen of the nitro group leads to the loss of a hydroxyl radical $[M - 17]^+$. The fragmentation patterns of 46 NACs in EI ionization are summarized by Feltes et al.[62] The advantages of EI are the structural information obtained from fragmentation which, for example, usually allows the easy determination of the number of nitro groups present in a molecule, and the ability to use large spectrum libraries.

With chemical ionization, detection in negative ion mode is often used, since the formation of a radical anion is favored by the high electron affinity of NACs, especially in the case of polynitroaromatics. NCI allows more sensitive detection: a factor of 20 in single ion monitoring was reported for 2,6-DNT,[62] and more than two orders of magnitude for 1-nitropyrene and 3-nitrofluoranthene.[27] However, it yields much less fragmentation and is seldom used for quantitative analysis due to reproducibility problems.

In the authors' laboratory the following method is used in routine analysis of water samples with GC/MS, though the enrichment step with dichloromethane will be replaced by a SPE method in the near future. As internal standards, the deuterated analogs 2-nitrotoluene-*d*7 and 2,4-DNT-*d*3 are spiked into each sample prior to extraction. The 500 mL sample of water is extracted three times, each with 40 mL dichloromethane, and the combined extracts are concentrated in a rotary evaporator at 40 °C to about 1 mL and placed into 1.4 mL autosampler vials. If necessary for lower detection limits, the extract is further reduced in volume under a gentle stream of nitrogen. A 1 µL sample of the extract is injected on-column. A separation column with medium polarity is especially suitable since it allows a better separation of the DNT isomers than a nonpolar column. A typical separation is shown in Figure 3. For this separation DB™-1701 (or equivalent), 30 m × 0.25 mm, and 0.25 µm film thickness was used. The GC temperature was held at 40 °C for 1 min, then raised at 10 °C min^{-1} to 230 °C. The carrier gas was helium at 1.03 bar (15 psi) column head pressure, and detection done with electron ionization in SIM mode. The use of deuterated standards in such analyses is highly recommended to allow compensation of analyte losses during sample preparation and of changes in ionization efficiencies. For this purpose, some deuterated NACs are commercially available (e.g. from Promochem, Wesel, Germany).

Even higher selectivity and (due to the enhanced signal-to-noise ratio) sensitivity is possible with the use of MS^n methods, mostly employing a triple–quadrupole instrument (allowing MS/MS, tandem MS or MS^2) or an ion-trap detector, which is in principle able to provide even higher dimensions in MS detection. However, practical reasons limit the use mostly to MS/MS/MS (or MS^3). Recent examples of the use of tandem MS in the analysis of NACs describe the analysis of nitro-PAHs in air[63] and soil.[64] In the former application the best MS conditions were found to be selection of the molecule ion in the first stage of the MS instrument and subsequent reaction monitoring with detection of NO_2^- at m/z 46.

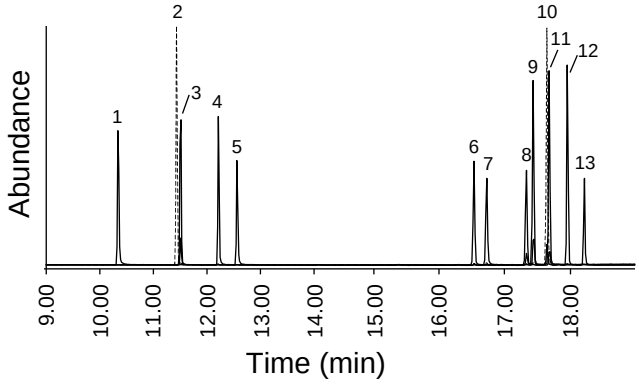

Figure 3 GC/MS chromatogram (EI, SIM) of a NAC mixture. Peak assignment: 1, nitrobenzene; 2, 2-nitrotoluene-*d*7; 3, 2-nitrotoluene; 4, 3-nitrotoluene; 5, 4-nitrotoluene; 6, 2,6-DNT; 7, 1,3-dinitrobenzene; 8, 2,5-DNT; 9, 2,3-DNT; 10, 2,4-DNT-*d*3; 11, 2,4-DNT; 12, 3,5-DNT; 13, 3,4-DNT.

A new, very interesting approach for the determination of NACs in soil samples was introduced by Williams and Pappas.[65] They used ultrahigh-speed GC on 0.5 m capillary columns and subsequent detection with a surface acoustic wave detector after thermodesorption of the analytes from soil for separation and detection of trinitrotoluene and DNT isomers in less than 2 min. The method was shown to be useful in the field for the approximate quantitation of those analytes, thus allowing a substantiated selection of samples for further investigations.

5.2 High-pressure Liquid Chromatography

HPLC is often the separation method of choice for the analysis of NACs. This applies especially for environmental analysis, since in many cases partly reduced metabolites, some of which are not amenable to GC analysis, have to be determined at the same time. Organic acids such as nitrotoluenesulfonic acids or nitrobenzoic acids also cannot be separated by GC without prior derivatization, and are mainly analyzed by HPLC. Another advantage of HPLC is the easier on-line coupling, with an enrichment step in comparison to GC, thus facilitating the set-up of totally automated analysis systems. A disadvantage of HPLC is the incompability of common solvents used in LLE with aqueous eluents in reversed-phase HPLC. In these cases a blow-down of the extractant and reconstitution in a suitable solvent is necessary, which might lead to the loss of more volatile analytes. Recent reviews were published on the HPLC analysis of explosives[66] and nitro-PAHs.[67]

Separation of NACs is usually performed with reversed phases and aqueous eluents. Methanol is the most widely used organic modifier, followed by acetonitrile. Depending on the analytes to be separated, isocratic elution may suffice or a gradient elution may be necessary. Some authors describe the isocratic elution with a ternary eluent,[68] although this is seldom used in routine analysis. C_{18} and C_8 stationary phases are mainly used. USEPA method 8030 demands the use of a CN column for confirmatory analysis since the elution order is completely different in comparison with the C_{18} primary column. The separation method used in the authors' laboratory is described here in more detail. It is superior to many of the published methods in terms of isomer separation and allows the analysis of at least 16 NACs in one chromatographic run. A typical chromatogram of a reference mixture is shown in Figure 4. For the separation a Nucleosil® 120 3-C_{18}, 250 × 3 mm column (Macherey-Nagel, Dueren, Germany) was used with the following gradient: methanol–water 35/65 (v/v), in 12 min to 45% water, isocratic for 28 min, in 10 min to 0% water (for column flushing), then re-equilibration for 15 min with the starting eluent. Flow was 0.35 mL min^{-1}, and detection was done with a DAD.

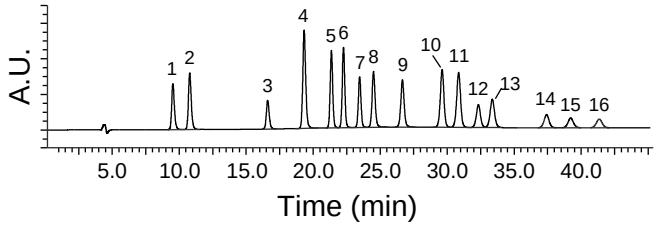

Figure 4 HPLC/DAD chromatogram of a NAC mixture, detection wavelength 230 nm. Peak assignment: 1, 2,6-diamino-4-nitrotoluene; 2, 2,4-diamino-6-nitrotoluene; 3, trinitrobenzylalcohol; 4, 1,3,5-trinitrobenzene; 5, 2-methyl-3-nitroaniline; 6, 4-methyl-3-nitroaniline; 7, 2-methyl-5-nitroaniline; 8, 1,3-dinitrobenzene; 9, trinitrotoluene; 10, 4-amino-2,6-DNT; 11, 2-amino-4,6-DNT; 12, 2,6-DNT; 13, 2,4-DNT; 14, 2-nitrotoluene; 15, 4-nitrotoluene; 16, 3-nitrotoluene.

A selective analysis of NACs is possible with the use of electron donor–acceptor complexes (EDA complexes) formation,[23,69] which allows separation according to the electron deficiency of the aromatic system. The stationary phase consists of clay conglomerates[23] or silica-based materials which are derivatized to yield electron-rich aromatic systems as end groups (naphthalene, *N*-methyl indole, anisole).[69]

The separation of neutral and acidic NACs in one chromatographic run was described by Caton and Griest[70] with a mixed-mode C_{18} anion exchanger column. A similar result can be achieved using acidified eluents[10] or ion-pairing chromatography,[71] since the retention of neutral NACs is hardly affected by pH or the presence of an ion-pairing reagent.

If more powerful separations are needed, the use of two-dimensional chromatography may drastically increase the peak capacity and separation power of the chromatographic system. To maximize these parameters the use of an orthogonal separation mechanism is ideal. Examples in the analysis of NACs are the selective analysis of nitro-PAHs in diesel exhaust particulates[72] and the analysis of water samples for nitroaromatic explosives.[73] In the latter case a safrol-modified silica provided the necessary selectivity difference from the primary C_{18} column.

A short overview of the most important HPLC detectors for NAC analysis is given in Table 6.

All NACs exhibit the more or less shifted benzene UV absorption bands. Therefore, a UV detector can be generally used, and is indeed by far the most widely applied detector.[74] Instead of using a confirmatory column, nowadays the use of a DAD is generally accepted, which allows compound verification via the on-line measurement of the compound's UV spectrum. The sensitivity of photometric detection is not very good,

Table 6 Comparison of detector types in HPLC

Detector	Instrument detection limits (pg)	Selectivity	Notes
UV/DAD	500–10 000	–	
Electrochemical	10–1000	o	Interferences by other reducible species, sensitivity increases with number of nitro groups
Fluorescence	10–500	+	Requires transformation, usually to aromatic amines, eventually subsequent derivatization
Chemiluminescence[a]	0.1–10	+	Requires transformation, usually to aromatic amines, and subsequent derivatization
MS, API, SIM (negative mode)	1–1000	++	Sensitivity increases with number of nitro groups
NMR	1 000 000	o	Only method for isomer identification

[a] Based on bis(trichlorophenyl)oxalate/hydrogen peroxide reaction with nitro PAHs.[67]
API, atmospheric pressure ionization; NMR, nuclear magnetic resonance.

therefore an enrichment step is always necessary for trace analysis.

Direct fluorescence detection of NACs is not possible because fluorescence is effectively quenched by the nitro group. To utilize such a detector therefore requires reduction to the corresponding amino compounds, some of which show intense native fluorescence (especially the amino-PAHs), but which can also easily be converted with standard derivatization reagents, e.g. fluorescamine, to yield highly fluorescing pyrrolinones.[75] The reduction can either be accomplished pre-column (mainly off-line) or post-column (mainly on-line) with a variety of reductants, including iron,[75] zinc[76] and hydrosulfide.[77] A disadvantage of the pre-column approach is the more demanding separation of the corresponding anilines in comparison with NACs, which mainly result from interactions with residual silanol groups. An especially interesting approach for post-column reduction of nitro-PAHs was recently described by Murayama and Dasgupta,[78] who used a coulometric detector for on-line reduction before subsequent fluorescent detection. By switching the coulometric detector on and off and subtracting both chromatograms they achieved an extraordinary selectivity. Another means to increase selectivity is a post-column reaction scheme which allows chemiluminescence detection.[76,77]

Electrochemical detection of NACs can principally be used in the reductive mode to yield the corresponding hydroxylamines or anilines (see section 4.2). In the review by Lloyd[66] on HPLC of explosives, different techniques for the reductive determination are thoroughly discussed, including hanging mercury drop electrodes and thin-layer electrodes. Cvacka et al.[67] give some examples of the use of amperometric detection in the reductive mode (potential in the range −500 to −650 mV) for environmental analysis of nitro-PAHs. However, a severe drawback of these methods is the interfering oxygen reduction, which leads to high background currents. Therefore a rigorous degassing of eluent and sample and the sole use of stainless-steel tubing is necessary. Still, in many routine applications oxygen interference cannot be avoided sufficiently. Thus the use of electrochemical detection in environmental analysis is more popular for NACs possessing other, oxidizable groups, such as aminonitro compounds or nitrophenols[79,80] which allow operation in the oxidative mode. The electrochemical detector shows a better selectivity (and often sensitivity) than a single-wavelength UV detector. However, use of an electrochemical detector requires much more operational efforts than does a UV detector. Besides, there are still problems with baseline stability in gradient elution, although the developments of the last few years are promising in that regard. An interesting approach was described by Krull and LaCourse,[81] who used a post-column photochemical reactor to cleave nitrite from the analytes, which was then detected in the oxidative mode.

In the past few years, high-performance liquid chromatography/mass spectrometry (HPLC/MS) has gained much attention for NAC analysis, as probably in all fields of HPLC analysis. The MS detection, particularly MS^n, allows in many cases unequivocal identification of analytes, even in difficult matrixes. Many commercialized interfaces have been used for the environmental analysis of NACs: thermospray,[82,83] particle beam,[84] electrospray ionization (ESI),[82,85] and atmospheric pressure chemical ionization (APCI).[82] From the authors' experience it is expected that ESI and APCI will become the standard interfaces over the next few years. NACs are mostly detected in the negative ion mode since the formation of radical anions

is favored by the electron-withdrawing nitro group. The sensitivity increases with the number of nitro groups and also for organic anions. Astratov et al.[82] compared the mass spectra of 26 NACs obtained with a thermospray, an APCI and an ESI interface, all operated in negative ion mode with methanol, water and formic acid as the mobile phase. In general, they observed the M^- or $[M-H]$ peak as the base peak with all interfaces. In a few cases, however, the use of APCI and ESI yielded clusters with methanol or formate as the base peak. The fragmentation increased from ESI over APCI to thermospray, reflecting the different thermal stresses prior to or during ionization. Astratov et al. emphasized that there is still a lot of variation in the mass spectra obtained, depending on the use of a specific instrument or interface. They did not provide data on the applied cone (also called fragmentation or capillary exit) voltage. The choice of the cone voltage is crucial for the fragmentation behavior of all NACs but in particular for acidic compounds such as nitrobenzoic acids and nitrotoluenesulfonic acids,[71] regardless of the interface type. With increasing cone voltage a higher degree of fragmentation is observed. The same behavior was found for polycarboxylates[86] and aromatic sulfonates.[87] Thus the cone voltage always has to be set at the same value to obtain reproducible mass spectra. On the other hand the possibility of in-source induced fragmentation offers the possibility of enhanced identification power even in single-quadrupole systems which are not capable of MS^n experiments. The ionization behavior of NACs (and other organic chemicals) in modern liquid chromatography coupled interfaces is not at all understood yet and requires a lot of further work.

Another hyphenated method which is increasingly being investigated is high-performance liquid chromatography/nuclear magnetic resonance (HPLC/NMR). Regarding the use for NACs, Preiss et al. did a number of initiative studies, mainly applied to the analysis of contaminated groundwater at ammunition waste sites.[88,89] This method is unique in its power to reveal the chemical structure of compounds, as even the distinction of positional isomers is possible. Therefore it is well suited for the most complex problems of structural analysis. However, despite advances in the last years, its lack of sensitivity is a major drawback in environmental analysis. Besides, it is very costly both in acquisition and operation and needs an experienced supervisor. Thus it is not expected to become a routine method in the analysis of NACs.

5.3 Capillary Electrophoresis and Related Techniques

During the last few years capillary electrophoresis (CE) has gained more and more importance in environmental analysis.[90] Among the capillary electroseparative techniques, CE, capillary electrochromatography (CEC) and especially micellar electrokinetic chromatography (MEKC) are of interest for the analysis of NACs. Instrumentation for these methods is similar: the main component is a fused silica capillary of a length between 20 and 100 cm and an internal diameter smaller than 100 μm. The capillary, which is filled with the separation electrolyte, is immersed at both ends into vials also filled with electrolyte. After the sample is injected in the capillary, a high voltage (up to 35 kV) is applied between two electrodes immersed into the vials. The temperature of the capillary has to be controlled and the hydrodynamic flow has to be minimized. Detection is usually carried out directly on a segment of the capillary (mostly by photometric or fluorimetric detection).

Whereas in CE only charged solutes can be analyzed because its separation principle is based on different electrophoretic mobilities of analytes, both CEC and MEKC make use of the distribution between two phases, enabling the separation of electrically neutral analytes.

CEC can be regarded as a hybrid technique between HPLC and CE. Using electro-osmotic flow, a mobile phase is driven through a stationary phase which is packed in the capillary. Bailey and Yan[91] used CEC to separate a series of 14 nitroaromatic and nitramine explosive compounds, achieving baseline resolution for all of the compounds in less than 7 min.

In MEKC, an ionic surfactant (usually sodium dodecyl sulfate, SDS) is added to a CE solution, forming micelles. The migration velocity of these micelles differs from that of the surrounding aqueous phase. The micelles are able to incorporate solutes and act as a pseudostationary phase. MEKC does not only extend the applicability of CE to neutral solutes, it also offers improvements in the separation of very complex mixtures. This is due to the numerous possibilities that are available for the manipulation of separation selectivity.

In comparison to classical chromatographic methods like GC and HPLC, capillary electroseparative techniques possess some unique advantages. High separation efficiencies (up to 300 000 plates per capillary) can be achieved easily under routine conditions, even for highly polar or thermal unstable substances, which are in general difficult to analyze. Sample volumes are very small, analysis time is short, and consumption of reagents and costs of columns are low. Moreover, minimal sample clean-up is necessary due to the exchange of the separation electrolyte in CE and MEKC after each run. Irreversible sorption of matrix constituents to the stationary phase cannot occur. Direct sample injection is often possible.

The analysis of trinitrotoluene and its transformation products in a variety of matrixes has been a popular subject of studies employing MEKC.[92,93] Mussenbrock and Kleiböhmer[94] optimized the separation

of 24 compounds of explosives via MEKC in less than 12 min, including all USEPA priority substances. Oehrle[95] used MEKC to analyze wastewater samples (called ''pink water'') from ammunition plants.

One major disadvantage of capillary electroseparative techniques is the relatively low LOD. Although featuring impressive mass sensitivity comparable to average GC systems, CE with its very small injection volumes of only a few nanoliters is usually not sensitive enough for environmental trace analysis when using photometric detection.

To enhance concentration sensitivity, three basic approaches are possible: increasing detection performance; off-line sample preconcentration; and on-line sample concentration.

The optical pathlength for detection can be increased using special detection cells such as z-shaped cells, multireflection cells, rectangular cells and bubble cells. The use of these cells may decrease separation efficiency, however.

For the analysis of NACs with their well-defined redox behavior, electrochemical detection can be useful. Hilmi et al.[96] presented an off-column amperometric cell for the determination of explosives by MEKC. In comparison to photometric detection, a ten-fold improvement of detection limit was achieved for most of the explosives tested in this study. The results were in good agreement with those obtained by the liquid chromatographic method recommended by the USEPA.

Thermal lens detection can also be used as alternative detection technique. Employing MEKC for the separation of nitrophenol pesticides, Seidel et al.[97] attained 1–2 orders of magnitude lower detection limits with thermal lens detection than with conventional photometric detection. Another promising effort to enhance sensitivity as well as selectivity is coupling of CE with MS.[98]

Using off-line sample preconcentration, CE can be employed for the analysis of nitroaromatics in water at ppb levels. Most SPE techniques can be transferred to CE. Rodriguez et al.[99] used cross-linked polystyrene for off-line concentration of nitro- and chlorophenols from water before separation with CE. Kaniansky et al.[100] used preparative capillary isotachophoresis (ITP) as a sample pretreatment technique for the analysis of nitrophenols. The following capillary zone electrophoresis (CZE) separation resulted in limits of detection of 2–8 ppb.

Furthermore, capillary electroseparative techniques offer the possibility of on-line concentration techniques that are not applicable in HPLC. One of them is sample stacking, which was initially developed for CZE. A large sample volume is injected. Differences in the effective migration mobilities of the solutes, caused by differences in the electric conductivity of bordering zones, result in focusing of the analytes in a narrow zone. This mechanism is also suitable for the on-line concentration of charged solutes in MEKC. With some modifications it can also be used for the concentration of electrically neutral species in MEKC, as the solutes exhibit an effective migration mobility during the stacking process, due to interaction with the pseudostationary phase.

In summary, capillary electroseparative techniques possess a high potential in this field. In some cases, they already offer a real alternative to GC and HPLC.

5.4 Thin-layer Chromatography

The use of TLC for the environmental analysis of NACs offers some significant advantages. Costs are low and analysis time is short, making this method ideal for screening purposes. Thermal degradation of analytes is insignificant. Moreover, TLC often does not require any sample clean-up. When analyzing soil samples in some cases it is necessary to use size-exclusion chromatography (SEC) to remove large amounts of humic acids.[101]

Due to the relatively poor separation efficiency of TLC, the method is not sufficient for certain identification of analytes. Two-dimensional TLC can improve separation significantly, employing binary as well as ternary mixtures of solvents.[41,102]

Automated multiple development high-performance TLC with a commercially available instrument was also successfully applied for the analysis of nitroaromatics and other explosives.[101] The development of the plate was done employing a stepwise gradient (25 steps) of the solvent, leading to a run time of about 4.5 h. Detection was performed with UV scanning. Detection limits ranged from 5 to 20 ng.

TLC employing different stationary phases was also used for the analysis of nitro-PAHs in environmental samples.[103]

Kessel and Hauck[104] developed a TLC method for qualitative screening and quantitative analysis of trinitrotoluene and its main degradation products in water samples after enrichment using SPE.

6 OUTLOOK

In general, the trend in environmental analysis of NACs is determined by advances in extraction, separation and detection methods. As an example it is to be expected that SFE and ASE, due to their ease of use, good extraction efficiencies and automatization possibilities, will become routine methods for the investigation of soil and particulate samples for NACs of all kinds. Separation methods will move to miniaturized systems. With the advent of more and more sensitive detectors,

the use of electrophoretic techniques will increase since they allow the most efficient separations of the whole range of NACs, from rather nonpolar compounds like nitro-PAHs (e.g. with CEC) to nitrosubstituted organic acids such as nitrosulfonic acids. The latter are increasingly recognized as pollutants of concern since their polarity allows fast movement in aqueous systems and penetration in drinking water reservoirs. The importance of oxidized metabolites of nitroaromatic explosives is also underestimated so far, which shows that despite numerous studies we still do not have a complete picture of the transformation processes for NACs in soil and water systems. Besides, nitration processes in the atmosphere are still not well understood. Atmospheric chemistry studies are extremely demanding in terms of analytical methods, and it can be expected that nitration will prove to be of more general importance for the formation of NACs than is known today. Regarding detection, the fast developments in MS are especially promising, both in terms of selectivity (enhancing the reliability of data) and sensitivity. For all ionization methods except EI there is still the need for a better understanding of the underlying mechanisms, thus allowing a better interpretation and prediction of mass spectra. The reproducibility of measurements has to be improved, since quantitation power is still limited. NMR methods may complement MS detection for identification of unknowns if they become easier to handle and cheaper to obtain. However, ECD in GC and a DAD in HPLC will remain the workhorses for NAC analysis in routine laboratories, at least for a number of years.

More sophisticated analytical methods allow the detection of lower and lower quantities of NACs, even in complex environmental matrixes. But to derive reliable information on environmental systems still requires more efforts in quality assurance of ultra-trace analyses. One way of providing unbiased data is the use of in situ methods, which will become more important, especially for the surveillance of point sources to aqueous systems. They already allow fast measurements with high temporal and spatial resolution without the problems involved with sampling and storage. Nowadays detection limits and stability of the systems are often not satisfying, but it is expected that both will improve over the next few years.

ACKNOWLEDGMENTS

The authors would like to thank Kerstin Heck for technical assistance and Dr Matthias Weibels for permission to use an amended version of Figure 2.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
ASE	Accelerated Solvent Extraction
CAS	Chemical Abstracts Service
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CZE	Capillary Zone Electrophoresis
DAD	Diode-array Detector
DIN	Deutsches Institut für Normung
DMF	Dimethylformamide
DNT	Dinitrotoluene
DPP	Differential Pulse Polarography
ECD	Electron Capture Detection
EDA	Electron Donor–Acceptor
EI	Electron Impact
ELISA	Enzyme-linked Immunosorbent Assay
ESI	Electrospray Ionization
FID	Flame Ionization Detection
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
HPLC/NMR	High-performance Liquid Chromatography/Nuclear Magnetic Resonance
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
MEKC	Micellar Electrokinetic Chromatography
MS	Mass Spectrometry
NAC	Nitroaromatic Compound
NCI	Negative Chemical Ionization
NMR	Nuclear Magnetic Resonance
OECD	Organisation for Economic Co-operation and Development
PAH	Polycyclic Aromatic Hydrocarbon
PND	Phosphorus–Nitrogen Detector
ppb	parts per billion
ppt	parts per trillion
PTV	Programmed Temperature Vaporizer
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TEA	Thermal Energy Analysis

TLC	Thin-layer Chromatography
TNT	Trinitrotoluene
USEPA	United States Environmental Protection Agency
UV	Ultraviolet

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Laser- and Optical-based Techniques for the Detection of Explosives

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites • Soxhlet and Ultrasonic Extraction of Organics in Solids

REFERENCES

1. G. Booth, 'Nitro Compounds, Aromatic', in *Ullmann's Encyclopedia of Industrial Chemistry*, 5th edition, VCH, Weinheim, 411–456, Vol. A17, 1991.
2. R. Atkinson, J. Arey, 'Atmospheric Chemistry of Gas-phase Polycyclic Aromatic Hydrocarbons: Formation of Atmospheric Mutagens', *Environ. Health Perspect. Suppl.*, **102**(Suppl. 4), 117–126 (1994).
3. J. Lüttke, V. Scheer, K. Levsen, G. Wünsch, J.N. Cape, K.J. Hargreaves, R.L. Storeton-West, K. Acker, W. Wieprecht, B. Jones, 'Occurrence and Formation of Nitrated Phenols In and Out of Cloud', *Atmos. Environ.*, **31**, 2637–2648 (1997).
4. W.J. Lyman, W.F. Reehl, D.H. Rosenblatt, *Handbook of Chemical Property Estimation Methods*, American Chemical Society, Washington, 1990.
5. D.E. Rickert, *Toxicity of Nitroaromatic Compounds*, Hemisphere Publishing, Washington, 1985.
6. P.C. Howard, S.S. Hecht, F.A. Beland, *Nitroarenes, Occurrence, Metabolism, and Biological Impact*, Plenum Press, New York, 1990.
7. J.C. Spain (ed.), *Biodegradation of Nitroaromatic Compounds*, Plenum Press, New York, 1995.
8. R.J. Spanggord, B.W. Gibson, R.G. Keck, D.W. Thomas, J.J. Barkley, Jr, 'Effluent Analysis of Wastewater Generated in the Manufacture of 2,4,6-Trinitrotoluene. 1. Characterization Study', *Environ. Sci. Technol.*, **16**, 229–232 (1982).
9. D.H. Rosenblatt, E.P. Burrows, W.R. Mitchell, D.L. Parmer, 'Organic Explosives and Related Compounds', in *The Handbook of Environmental Chemistry*, ed. O. Hutzinger, Springer, Heidelberg, Vol. 3G, 1991.
10. M. Godejohann, A. Preiss, K. Levsen, G. Wünsch, 'Determination of Nitrophenols, Nitrobenzoic Acids and Polar Explosives by HPLC–Diode Array Detection in Groundwater Samples of Former Ammunition Plants', *Chromatographia*, **43**, 612–618 (1996).
11. T.C. Schmidt, K. Steinbach, E.v. Löw, G. Stork, 'Highly Polar Metabolites of Nitroaromatic Compounds in Ammunition Wastewater', *Chemosphere*, **37**, 1079–1091 (1998).
12. D. Bruns-Nagel, T.C. Schmidt, O. Drzyzga, E.v. Löw, K. Steinbach, 'Identification of Oxidized TNT Metabolites in Soil Samples of a Former Ammunition Plant', *ESPR – Environ. Sci. & Pollut. Res.*, **6**, 7–10 (1999).
13. D.C. Leggett, T.F. Jenkins, P.H. Miyares, 'Salting-out Solvent Extraction for Preconcentration of Neutral Polar Organic Solutes from Water', *Anal. Chem.*, **62**, 1355–1356 (1990).
14. C.F. Bauer, C.L. Grant, T.F. Jenkins, 'Interlaboratory Evaluation of High-performance Liquid Chromatographic Determination of Nitroorganics in Munition Plant Wastewater', *Anal. Chem.*, **58**, 176–182 (1986).
15. C.A. Weisberg, M.L. Ellickson, 'Practical Modifications to USEPA Method 8330 for the Analysis of Explosives by HPLC', *Am. Lab.*, **30**, 32N, 32P–32Q, 32S–32V (1998).
16. J. Ding, K. Xie, E. Liu, Y. Zhang, Y.-P. Hsia, Z.-P. Luo, 'Study of Surrogates for SW846 Method 8330', *J. Chromatogr. Sci.*, **37**, 65–70 (1999).
17. K. Levsen, P. Mußmann, E. Berger-Preiß, A. Preiß, D. Volmer, G. Wünsch, 'Analysis of Nitroaromatics and Nitramines in Ammunition Wastewater and in Aqueous Samples from Former Ammunition Plants and Other Military Sites', *Acta Hydrochim. Hydrobiol.*, **21**, 153–166 (1993).
18. T. Welsch, H. Block, 'Separation and Enrichment of Traces of Explosives and their By-products from Water by Multiple Micro Liquid Extraction for their Determination by Capillary Gas Chromatography', *Fresenius' J. Anal. Chem.*, **357**, 904–908 (1997).

19. J. Kruppa, A. Preiß, K. Levsen, H.P. Kabus, 'Off-line and On-line Extraction of Explosives and Related Compounds from Aqueous Samples Using Solid Sorbents', *Acta Hydrochim. Hydrobiol.*, **24**, 226–231 (1996).
20. S.D. Harvey, T.R.W. Clauss, 'Rapid On-line Chromatographic Determination of Trace-level Munitions in Aqueous Samples', *J. Chromatogr., A*, **753**, 81–89 (1996).
21. T.C.R. Santos, L.C. Rocha, D. Barcelo, 'Multiresidue Analysis of Pesticides in Water from Rice Cultures by On-line Solid Phase Extraction Followed by LC–DAD', *Int. J. Environ. Anal. Chem.*, **70**, 19–28 (1998).
22. S.A. Senseman, T.L. Lavy, T.C. Daniel, 'Monitoring Groundwater for Pesticides at Selected Mixing/Loading Sites in Arkansas', *Environ. Sci. Technol.*, **31**, 283–288 (1997).
23. T.D. Bucheli, S.R. Müller, P. Reichmuth, S.B. Haderlein, R.P. Schwarzenbach, 'Spherical Clay Conglomerates: A Novel Stationary Phase for Solid-phase Extraction and ''Reversed-phase'' Liquid Chromatography', *Anal. Chem.*, **71**, 2171–2178 (1999).
24. B. Schäfer, W. Engewald, 'Enrichment of Nitrophenols from Water by Means of Solid-phase Microextraction', *Fresenius' J. Anal. Chem.*, **352**, 535–536 (1995).
25. R. Eisert, K. Levsen, 'Determination of Organophosphorus, Triazine and 2,6-Dinitroaniline Pesticides in Aqueous Samples via Solid-phase Microextraction (SPME) and Gas Chromatography with Nitrogen Phosphorus Detection', *Fresenius' J. Anal. Chem.*, **351**, 355–362 (1995).
26. S.A. Barshick, W.H. Griest, 'Trace Analysis of Explosives in Seawater Using Solid-phase Microextraction and Gas Chromatography Ion Trap Mass Spectrometry', *Anal. Chem.*, **70**, 3015–3020 (1998).
27. N. Saito, K. Yamaguchi, *Chromatographic Analysis of Nitrobenzene and Related Compounds*, Chromatographic Science Series 77, Marcel Dekker, New York, 169–212, 1998.
28. T.F. Jenkins, C.L. Grant, 'Comparison of Extraction Techniques for Munitions Residues in Soil', *Anal. Chem.*, **59**, 1326–1331 (1987).
29. T.F. Jenkins, M.E. Walsh, P.W. Schumacher, P.H. Miyares, C.F. Bauer, C.L. Grant, 'Liquid Chromatographic Method for Determination of Extractable Nitroaromatic and Nitramine Residues in Soil', *J. – Assoc. Off. Anal. Chem.*, **72**, 890–899 (1989).
30. C.F. Bauer, S.M. Koza, T.F. Jenkins, 'Liquid Chromatographic Method for Determination of Explosives Residues in Soil: Collaborative Study', *J. – Assoc. Off. Anal. Chem.*, **73**, 541–552 (1990).
31. E. Alter, G. Donnevert, C. Sabel, 'Analysis of Explosives. A Comparison of Methods', *Umweltwiss. Schadst.-Forsch.*, **10**, 66–74 (1998).
32. V. Lopez-Avila, W.F. Beckert, *Supercritical Fluid Extraction in Environmental Analysis*, ACS Symposium Series 488 (Supercritical Fluid Technology), American Chemical Society, Washington, 179–205, 1992.
33. T.S. Reighard, S.V. Olesik, 'Comparison of the Extraction of Phenolic and Nitroaromatic Pollutants Using Supercritical and Enhanced-fluidity Liquid Methanol CO_2', *J. Chromatogr., A*, **737**, 233–242 (1996).
34. C.E. Wujcik, J.N. Seiber, 'Supercritical Fluid Extraction of 2,4,6-Trinitrotoluene and 1,3,5-Trinitrobenzene from Soil', *J. Environ. Sci. Health, Part A*, **31**, 1361–1377 (1996).
35. R. Deuster, N. Lubahn, C. Friedrich, W. Kleibohmer, 'Supercritical CO_2 Assisted Liquid Extraction of Nitroaromatic and Polycyclic Aromatic Compounds in Soil', *J. Chromatogr., A*, **785**, 227–238 (1997).
36. M.J. Wong, Q.X. Li, B.D. Hammock, J.N. Seiber, 'Method for the Analysis of 4-Nitrophenol and Parathion in Soil Using Supercritical Fluid Extraction and Immunoassay', *J. Agric. Food Chem.*, **39**, 1802–1807 (1991).
37. E. Jungreis, L. Ben-Dor, 'Organic Spot Test Analysis', in *Wilson and Wilson's Comprehensive Analytical Chemistry*, ed. G. Svehla, Elsevier, Amsterdam, Vol. X, 1980.
38. R.T. Medary, 'Inexpensive, Rapid Field Screening Test for 2,4,6-Trinitrotoluene in Soil', *Anal. Chim. Acta*, **258**, 341–346 (1992).
39. T.F. Jenkins, M.E. Walsh, 'Development of Field Screening Methods for TNT, 2,4-DNT and RDX in Soil', *Talanta*, **39**, 419–428 (1992).
40. R.T. Echols, M.M. Christensen, R.M. Krisko, J.H. Aldstadt, 'Selective Determination of TNT in Soil Extracts by Sequential Injection Spectrophotometry', *Anal. Chem.*, **71**, 2739–2744 (1999).
41. R. Haas, G. Stork, 'Konzept zur Untersuchung von Ruestungsaltlasten. 1. Untersuchung ehemaliger TNT-Fabriken und Fuellstationen', *Fresenius' J. Anal. Chem.*, **335**, 839–846 (1989).
42. M. Weibels, *Entwicklung und Optimierung eines Differential-Puls-Polarographischen Verfahrens zur Bestimmung eines Summenaparameters für Nitro-Verbindungen*, Görich & Weiershäuser, Marburg, 1995.
43. P.M. Krämer, 'A Strategy to Validate Immunoassay Test Kits for TNT and PAHs as a Field Screening Method for Contaminated Sites in Germany', *Anal. Chim. Acta*, **376**, 3–11 (1998).
44. D.L. Eck, M.J. Kurth, C. Macmillan, 'Trinitrotoluene and Other Nitroaromatic Compounds', in *Immunochemical Methods for Environmental Analysis*, American Chemical Society, Washington, 80–94, 1990.
45. Q.X. Li, M.S. Zhao, S.J. Gee, M.J. Kurth, J.N. Seiber, B.D. Hammock, 'Development of Enzyme-linked Immunosorbent Assay for 4-Nitrophenol and Substituted 4-Nitrophenols', *J. Agric. Food Chem.*, **39**, 1685–1692 (1991).
46. C. Keuchel, L. Weil, R. Niessner, 'Einsatz eines Enzymimmunoassays als Screening-verfahren für die Bestimmung von Trinitrotoluol im Grundwasser', *Vom Wasser*, **81**, 7–15 (1993).
47. C. Keuchel, R. Niessner, 'Rapid Field Screening Test for Determination of 2,4,6-Trinitrotoluene in Water and

Soil with Immunofiltration', *Fresenius' J. Anal. Chem.*, **350**, 538–543 (1994).
48. M. Dosch, M.G. Weller, A.F. Bückmann, R. Niessner, 'Homogeneous Immunoassay for the Detection of Trinitrotoluene (TNT) based on Reactivation of Apoglucose Oxidase Using a Novel FAD–Trinitrotoluene Conjugate', *Fresenius' J. Anal. Chem.*, **361**, 174–178 (1998).
49. P. Pfortner, M.G. Weller, R. Niessner, 'Detection of Bound Nitroaromatic Residues in Soil by Immunoassay', *Fresenius' J. Anal. Chem.*, **360**, 781–783 (1998).
50. P. Jülicher, E. Mussenbrock, R. Renneberg, K. Camman, 'Broadening the Antibody Specificity by Hapten Design for an Enzyme-linked Immunoassay as an Improved Screening Method for the Determination of Nitroaromatic Residues in Soils', *Anal. Chim. Acta*, **315**, 279–287 (1995).
51. B. Fröschl, D. Knopp, R. Niessner, 'Possibilities and Limitations of an Enzyme-linked Immunosorbent Assay for the Detection of 1-Nitropyrene in Air Particulate Matter', *Fresenius' J. Anal. Chem.*, **360**, 689–692 (1998).
52. W.J. Buttner, M. Findlay, W. Vickers, W.M. Davis, E.R. Cespedes, S. Cooper, J.W. Adams, 'In Situ Detection of Trinitrotoluene and Other Nitrated Explosives in Soils', *Anal. Chim. Acta*, **341**, 63–71 (1997).
53. J. Wang, R.K. Bhada, J. Lu, D. MacDonald, 'Remote Electrochemical Sensor for Monitoring TNT in Natural Waters', *Anal. Chim. Acta*, **361**, 85–91 (1998).
54. M. Krausa, J. Doll, K. Schorb, W. Boke, G. Hambitzer, 'Fast Electrochemical Detection of Nitro- and Aminoaromates in Soil and Liquids', *Propellants, Explos., Pyrotech.*, **22**, 156–159 (1997).
55. U. Narang, P.R. Gauger, F.S. Ligler, 'Capillary-based Displacement Flow Immunosensor', *Anal. Chem.*, **69**, 1961–1964 (1997).
56. J. Yinon, S. Zitrin, *Modern Methods and Applications in Analysis of Explosives*, John Wiley & Sons, Chichester, 1993.
57. J. Yinon, 'Trace Analysis of Explosives in Water by Gas Chromatography Mass Spectrometry with a Temperature-programmed Injector', *J. Chromatogr., A*, **742**, 205–209 (1996).
58. S. Müller, J. Efer, W. Engewald, 'Gas Chromatographic Water Analysis by Direct Injection of Large Sample Volumes in an Adsorbent-packed PTV Injector', *Chromatographia*, **38**, 694–700 (1994).
59. S.B. Lartiges, P. Garrigues, 'Gas Chromatographic Analysis of Organophosphorus and Organonitrogen Pesticides with Different Detectors', *Analusis*, **23**, 418–421 (1995).
60. L. Warzecha, 'Separation and Analysis of Nitroarenes from Airborne Particulate Organic Matter by HPLC and Capillary GC–MS Methods', *J. High Resolut. Chromatogr.*, **19**, 639–642 (1996).
61. M.E. Walsh, T. Ranney, 'Determination of Nitroaromatic, Nitramine, and Nitrate Ester Explosives in Water Using Solid-phase Extraction and Gas Chromatography Electron Capture Detection: Comparison with High-performance Liquid Chromatography', *J. Chromatogr. Sci.*, **36**, 406–416 (1998).
62. J. Feltes, K. Levsen, D. Volmer, M. Spiekermann, 'Gas Chromatographic and Mass Spectrometric Determination of Nitroaromatics in Water', *J. Chromatogr.*, **518**, 21–40 (1990).
63. M. Vincenti, C. Minero, E. Pelizzetti, M. Fontana, R. De Maria, 'Sub-parts-per-billion Determination of Nitrosubstituted Polynuclear Aromatic Hydrocarbons in Airborne Particulate Matter and Soil by Electron Capture–Tandem Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **7**, 1255–1265 (1996).
64. M. Niederer, 'Determination of Polycyclic Aromatic Hydrocarbons and Substitutes (Nitro-, Oxy-PAHs) in Urban Soil and Airborne Particulate by GC–MS and NCI–MS/MS', *Environ. Sci. Pollut Res.*, **5**, 209–216 (1998).
65. D. Williams, G. Pappas, 'Ultra-rapid Soil Screening with the Use of GC-SAW Technology for TNT/DNT and Related Nitroaromatic Contaminants', *Field Anal. Chem. Technol.*, **2**, 299–308 (1998).
66. J.B.F. Lloyd, 'HPLC of Explosive Materials', in *Advances in Chromatography*, eds. J.C. Giddings, E. Yrushka, P.R. Brown, Marcel Dekker, New York, Vol. 32, 1992.
67. J. Cvacka, J. Barek, A. Fogg, J.C. Moreira, J. Zima, 'High-performance Liquid Chromatography of Nitrated Polycyclic Aromatic Hydrocarbons', *Analyst*, **123**, 9R–18R (1998).
68. T.F. Jenkins, D.C. Leggett, C.L. Grant, C.F. Bauer, 'Reversed-phase High-performance Liquid Chromatographic Determination of Nitroorganics in Munition Wastewater', *Anal. Chem.*, **58**, 170–175 (1986).
69. M. Markuszewski, J.D. Krass, T. Hippe, B. Jastorff, R. Kaliszan, 'Separation of Nitroaromatics and their Transformation Products in Soil Around Ammunition Plants: New High Performance Liquid Chromatographic Charge Transfer Stationary Phases', *Chemosphere*, **37**, 559–575 (1998).
70. J.E. Caton, W.H. Griest, 'Determination of Explosives and Some Metabolites of TNT in Biological and Environmental-samples by Liquid-chromatography on a Mixed-mode C-18-Anion Column', *J. Liq. Chromatogr.*, **19**, 661–677 (1996).
71. T.C. Schmidt, K. Steinbach, U. Buetehorn, K. Heck, U. Volkwein, G. Stork, 'Synthesis of Reference Substances for the Determination of Highly Polar Metabolites of Nitroaromatic Compounds', *Chemosphere*, **38**, 3119–3130 (1999).
72. E. Veigl, W. Posch, W. Lindner, P. Tritthart, 'Selective and Sensitive Analysis of 1-Nitropyrene in Diesel Exhaust Particulate Extract by Multidimensional HPLC', *Chromatographia*, **38**, 199–206 (1994).
73. A.P. Koehne, U. Dornberger, T. Welsch, 'Two-dimensional High Performance Liquid Chromatography for

the Separation of Complex Mixtures of Explosives and their By-products', *Chromatographia*, **48**, 9–16 (1998).
74. L. Huber, R. Schuster, 'Environmental Applications of Diode Array Detection in HPLC', *Chromatogr. Sci. Ser.*, **62**, 325–350 (1993).
75. D. Djozan, M.A. FarajZadeh, 'Liquid Chromatographic Determination of Nitroaromatics in Water Following Solid Phase Extraction, Pre-column Reduction and Derivatization', *Chromatographia*, **43**, 25–30 (1996).
76. K. Hayakawa, N. Terai, P.G. Dinning, K. Akutsu, Y. Iwamoto, R. Etoh, T. Murahashi, 'An On-line Reduction HPLC/Chemiluminescence Detection System for Nitropolycyclic Aromatic Hydrocarbons and Metabolites', *Biomed. Chromatogr.*, **10**, 346–350 (1996).
77. K. Hayakawa, T. Murahashi, M. Butoh, M. Miyazaki, 'Determination of 1,3-Dinitropyrene, 1,6-Dinitropyrene, and 1,8-Dinitropyrene and 1-Nitropyrene in Urban Air by High-performance Liquid-chromatography Using Chemiluminescence Detection', *Environ. Sci. Technol.*, **29**, 928–932 (1995).
78. M. Murayama, P.K. Dasgupta, 'Liquid Chromatographic Determination of Nitro-substituted Polynuclear Aromatic Hydrocarbons by Sequential Electrochemical and Fluorescence Detection', *Anal. Chem.*, **68**, 122–1232 (1996).
79. K. Spiegel, T. Welsch, 'Monitoring Degradation Processes of Explosives by HPLC Analysis with UV and Amperometric Detection', *Fresenius' J. Anal. Chem.*, **357**, 333–337 (1997).
80. U. Lewin, J. Efer, W. Engewald, 'High-performance Liquid Chromatographic Analysis with Electrochemical Detection for Residues of Explosives in Water Samples Around a Former Ammunition Plant', *J. Chromatogr., A.*, **730**, 161–167 (1996).
81. S. Krull, W.R. LaCourse, 'Post-column Photochemical Derivatizations for Improved Detection and Identification in HPLC', in *Reaction Detection in Liquid Chromatography*, ed. I.S. Krull, Marcel Dekker, New York, 1986.
82. M. Astratov, A. Preiss, K. Levsen, G. Wünsch, 'Identification of Pollutants in Ammunition Hazardous Waste Sites by Thermospray HPLC/MS', *Int. J. Mass Spectrom. Ion Processes*, **167**, 481–502 (1997).
83. P.M. Gates, E.T. Furlong, T.F. Dorsey, M.R. Burkhardt, 'Determination of Nitroaromatic Explosives and their Degradation Products in Unsaturated-zone Water Samples by High-performance Liquid Chromatography with Photodiode-array, Mass Spectrometric, and Tandem Mass Spectrometric Detection', *Trends Anal. Chem.*, **15**, 319–325 (1996).
84. L. Bonfanti, M. Careri, A. Mangia, P. Manini, M. Maspero, 'Simultaneous Identification of Different Classes of Hydrocarbons and Determination of Nitro-polycyclic Aromatic Hydrocarbons by Means of Particle Beam Liquid Chromatography Mass Spectrometry', *J. Chromatogr., A*, **728**, 359–369 (1996).
85. B. Casetta, F. Garofolo, 'Characterization of Explosives by Liquid-chromatography Mass-spectrometry and Liquid-chromatography Tandem Mass-spectrometry Using Electrospray-ionization and Parent-ion Scanning Techniques', *Org. Mass Spectrom.*, **29**, 517–525 (1994).
86. H. Frauendorf, R. Herzschuh, 'Application of High-performance Liquid Chromatography/Electrospray Mass Spectrometry for Identification of Carboxylic Acids Containing Several Carboxyl Groups from Aqueous Solutions', *Eur. Mass Spectrom.*, **4**, 269–278 (1998).
87. T. Storm, Th. Reemtsma, M. Jekel, 'Use of Volatile Amines as Ion Pairing Agents for the High-performance Liquid-chromatographic–Tandem Mass Spectrometric Determination of Aromatic Sulfonates in Industrial Wastewater', *J. Chromatogr., A*, **854**, 175–185 (1999).
88. M. Godejohann, A. Preiss, C. Mügge, G. Wünsch, 'Application of On-line HPLC–H-1 NMR to Environmental Samples: Analysis of Groundwater Near Former Ammunition Plants', *Anal. Chem.*, **69**, 3832–3837 (1997).
89. M. Godejohann, M. Astratov, A. Preiss, K. Levsen, C. Mugge, 'Application of Continuous Flow HPLC Proton Nuclear Magnetic Resonance Spectroscopy and HPLC Thermospray Mass Spectroscopy for the Structural Elucidation of Phototransformation Products of 2,4,6-Trinitrotoluene', *Anal. Chem.*, **70**, 4104–4110 (1998).
90. L. Song, Z. Xu, J. Kang, J. Cheng, 'Analysis of Environmental Pollutants by Capillary Electrophoresis with Emphasis on Micellar Electrokinetic Chromatography', *J. Chromatogr., A*, **780**, 297–328 (1997).
91. C. Bailey, C. Yan, 'Separation of Explosives Using Capillary Electrochromatography', *Anal. Chem.*, **70**, 3275–3279 (1998).
92. L. Pucik, J. Hughes, 'Capillary Electrophoretic Separation of TNT and its Transformation Products', *J. Capillary Electrophor.*, **3**, 209–213 (1996).
93. T.M. Chow, J. Liu, M. Piwoni, N.R. Adrian, 'Analysis of the Transformation of Nitroaromatic Compounds in Wastewater by Bacteria Using Micellar Electrokinetic Capillary Chromatography', *J. Capillary Electrophor.*, **4**, 189–194 (1997).
94. E. Mussenbrock, W. Kleiböhmer, 'Separation Strategies for the Determination of Residues of Explosives in Soils Using Micellar Electrokinetic Capillary Chromatography', *J. Microcol. Sep.*, **7**, 107–116 (1995).
95. S.A. Oehrle, 'Analysis of Nitramine and Nitroaromatic Explosives in Pink Water by Capillary Electrophoresis', *Electrophoresis*, **18**, 300–302 (1997).
96. A. Hilmi, J.H.T. Luong, A.L. Nguyen, 'Development of Electrokinetic Capillary Electrophoresis Equipped with Amperometric Detection for Analysis of Explosive Compounds', *Anal. Chem.*, **71**, 873–878 (1999).

97. B.S. Seidel, W. Faubel, H.J. Ache, 'Thermal Lens Detection System for Capillary Electrophoresis', *J. Biomed. Optics*, **2**(3), 326–331 (1997).
98. W.C. Brumley, W. Winnik, 'Capillary Electrophoresis/Mass Spectrometry Applied to Environmental Analysis', *J. Chrom. Libr.*, **59**, 481–527 (1996).
99. I. Rodriguez, M.I. Turnes, M.H. Bollain, M.C. Mejuto, R. Cela, 'Determination of Phenolic Pollutants in Drinking Water by Capillary Electrophoresis in the Sample Stacking Mode', *J. Chromatogr., A*, **778**, 279–288 (1997).
100. D. Kaniansky, E. Kremova, V. Madajova, M. Masar, 'Capillary Zone Electrophoresis of Nitrophenols with Off-line Isotachophoretic Sample Pretreatment', *Electrophoresis*, **18**, 260–267 (1997).
101. C. Steuckart, E. Berger-Preiss, K. Levsen, 'Determination of Explosives and their Biodegradation Products in Contaminated Soil and Water from Former Ammunition Plants by Automated Mutiple Development High-performance Thin-layer Chromatography', *Anal. Chem.*, **66**, 2570–2577 (1994).
102. L. Bagnato, G. Grasso, 'Two-dimensional Thin Layer Chromatography for the Separation and Identification of Nitro-derivatives in Explosives', *J. Chromatogr.*, **357**, 440–444 (1986).
103. K. Trypien, B. Janoszka, D. Bodzek, 'Application of Planar Chromatography to the Analysis of Polycyclic Aromatic Hydrocarbons and their Derivatives in Environmental Samples', *J. Chromatogr.*, **774**, 111–120 (1997).
104. S. Kessel, H. Hauck, 'Qualitative and Quantitative Determination of 2,4,6-TNT, Octogen, Aminonitro- and Nitrocompounds in Ammunition Wastes with Modified TLC and HPTLC Precoated Layers', *Chromatographia*, **43**, 401–404 (1996).
105. L.C. Shriver-Lake, B.L. Donner, F.S. Ligler, 'On-site Detection of TNT with a Portable Fiber Optic Sensor', *Environ. Sci. Technol.*, **31**, 837–841 (1997).
106. L.L. Bart, J.C. Judd, K.E. Hoffman, A.M. Wilkins, A.W. Kunsterbeck, 'Application of a Portable Immunosensor to Detect the Explosives TNT and RDX in Groundwater Samples', *Environ. Sci. Technol.*, **31**, 1501–1511 (1997).

Nuclear Magnetic Resonance for Environmental Monitoring

Janet S. MacFall
Elon College, USA

Anthony A. Ribeiro
Duke University, Durham, USA

1 Introduction **2**
1.1 History and Background 2
1.2 Limitations 4
1.3 Potential Applications 5
2 Environmental Applications **5**
2.1 Magnetic Resonance Imaging of Soil 5
2.2 Field Application of Nuclear Magnetic Resonance Spectroscopy for Soil Study 8
2.3 Aluminum in Surface Waters 10
2.4 Carbon in Soil – Nuclear Magnetic Resonance of Humic Materials 11
2.5 Pesticide Monitoring In Vivo and In Soil 12
Acknowledgments **12**
Abbreviations and Acronyms **12**
Related Articles **13**
References **13**

Nuclear magnetic resonance (NMR) technology is based on discoveries made in 1946, which have provided the basis for many applications in chemical identification, characterization, and imaging. The earliest use was for NMR spectroscopic studies of the molecular distribution of hydrogen, the nucleus detected by NMR with the greatest sensitivity. The use of NMR imaging was first explored in early 1970s, using fruit and vegetable specimens which provided good models for engineering. These early experiments clearly demonstrated the potential utility of NMR for nondestructive visualization of the internal components of intact specimens. Nuclei which can be detected and studied by NMR include those with unpaired electrons, such as ^{1}H, ^{2}H, ^{31}P, ^{14}N, ^{15}N, ^{17}O, ^{19}F, ^{27}Al, ^{23}Na and ^{39}K. Contemporary NMR units used for either imaging or spectroscopy have three major components: a large magnet, a radiofrequency (rf) coil, and a computer-driven console to control the pulse sequence and collect data. Samples are placed into the center of the magnet, allowing nuclei to become aligned either parallel or antiparallel to the magnetic field while precessing at a defined frequency which is proportional to the magnetic field strength. The nuclear precession is then realigned to a temporary magnetic field which is perpendicular to the primary magnetic field and is induced by the rf coil. Finally the rf coil is turned off, allowing spins to realign with the primary magnetic field and give off energy which is detected as an alternating current in the process. In NMR spectroscopic experiments, the molecular environment of each nucleus induces slight alterations in the magnetic field, which are detected as slight changes in the frequency of precession. This information is used to create spectra showing details of the molecular environment of the nuclei being detected, and information describing the structure of the molecule. With imaging experiments, electromagnetic coils are placed in the bore of the magnet, to create linearly defined alterations in the magnetic field. As with spectroscopic experiments, the specimen is placed in the center of the stationary magnet. Nuclei are realigned by pulses from an rf coil, and the energy given off is detected as an alternating current. With imaging, changes in the frequency of precession provide information describing the spatial location of the specimen within the bore, and allow images of the specimen volume to be acquired without disruption to the specimen. Use of electromagnets which create a linear magnetic field allows the precise position of the specimen and its morphology to be determined. Contrast within the images is determined both by the content of detected nuclei (usually ^{1}H within water), and by the interactions between nuclei and their surroundings. The advantage of analysis by NMR spectroscopy and imaging is that the samples can be studied without disruption to the specimen. Chemical analysis can even be done in vivo or in an undisturbed setting. Similarly, imaging can be done on live, intact specimens, the only requirement being that they remain immobile for some period of time. The limitation of these techniques is that only specific nuclei can be detected, with most imaging experiments primarily limited to water distribution and binding. Specimens must be sufficiently small to fit within the bore of the stationary magnet. The magnet bore size may range from a few millimeters for some spectroscopic experiments to up 100 cm with some imaging units. Additionally, the technique is relatively insensitive compared with other analytical techniques such as immunoassays or fluorescent assays. Nevertheless, NMR spectroscopy and imaging provide a unique and powerful set of tools for environmental assessment. Applications include examination of humic substances in soil, soil chemistry, chemistry of surface waters, aluminum chemistry in situ, pesticide fate when introduced into the environment, water uptake and translocation by plants, and structural studies of living specimens such as plant roots within a soil volume. New applications are constantly

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

appearing as magnet technology improves, new pulse sequences are developed, and coil design progresses.

1 INTRODUCTION

1.1 History and Background

NMR technology is based on discoveries made in 1946 by two groups of researchers, headed by Felix Bloch and Edward Purcell respectively. Its earliest use was in NMR spectroscopic studies of the molecular distribution of hydrogen, the nucleus detected by NMR with the greatest sensitivity. Today, similar spectroscopic experiments provide detailed information on molecular structure, and are used extensively in chemical analysis of many types of molecules. Many texts are available describing applications of NMR in determining molecular structure.[1,2]

The use of NMR for imaging was first explored in the early 1970s, using fruit and vegetable specimens which provided good models for engineering.[3] These early experiments clearly demonstrated the potential utility of NMR for nondestructive visualization of the internal components of intact specimens.

1.1.1 Introductory Theory

Nuclei which can potentially be studied with NMR are those with unpaired electrons. These include ^{1}H, ^{2}H, ^{31}P, ^{14}N, ^{15}N, ^{17}O, ^{19}F, ^{27}Al, ^{23}Na, and ^{39}K. Of these, ^{1}H can be detected with the greatest sensitivity. The high natural abundance of this nucleus in the water within both soil and biological tissue (such as plant roots), combined with its high detectability, makes this the most commonly used nucleus for NMR imaging and spectroscopic studies.[2]

There are three major components to NMR instruments: a large magnet, an rf coil, and a computer-driven console to control the pulse sequence and to collect and process received signals. Magnets used in contemporary spectroscopic and imaging units are usually static, superconducting magnets. They create a magnetic field which is parallel with the center bore of the magnet.

Another fundamental piece of hardware required for NMR imaging experiments is the rf coil which is placed either around or adjacent to the specimen. This is tuned to a frequency specific to the field strength of the magnet and the nucleus to be detected. It is also used to receive the alternating current generated by manipulating the nuclear spins within the bore of the magnet.

In practice, the sample is placed inside or adjacent to an rf coil. The specimen and the rf coil are then positioned in the center of the bore of the magnet, where the spins align themselves either parallel or antiparallel to the primary magnetic field. The distribution of alignments depends on the magnetic field strength and the identity of the nucleus. Since the detection of NMR signals depends on relatively small differences between the number of nuclei in parallel or antiparallel alignments, NMR is a rather insensitive technique compared with many others such as chemical or radioisotopic assays.

Magnetic resonance imaging (MRI) units are also composed of several computers linked by a communications network. The computers control all aspects of image acquisition and processing. The pulse sequences used to generate the images are stored in the main computers, and these control gradient generation, as well as controlling the rf coils needed for flipping the nuclei to be detected into defined angles of orientation and detecting the alternating current generated with each pulse. A substantial amount of memory is necessary to store the software required to run the machine, and to acquire and process images. Generally a minimum of 1 GB is needed for the scanner itself. Additional computers for off-line image reconstruction, processing and analysis are also desirable to maximize the time the scanner can be used for image acquisition.

The final component is the library of software necessary to run the gradients, power the rf coil, detect and amplify the alternating current received from the sample, and process the detected signal into an image using multiple Fourier transformations. Within the medical community new pulse sequences are continually being developed, either for specific diagnostic applications or to decrease the time necessary to acquire an image. Many of these have shown promise for enhancing capabilities for environmental monitoring.

NMR experiments can be viewed from the classical perspective, where the nuclear magnetism is derived from the spinning of the nucleus. The precessing motion of the nuclear magnetic moment within the externally applied magnetic field (the superconducting magnet) is described by the Larmor equation, a relationship which is fundamental to all NMR experiments. This relationship can be described as shown in Equation (1)

$$w = \gamma B_0 \tag{1}$$

where w is the frequency which with the nucleus can be considered to precess, γ is the gyromagnetic ratio for the nucleus being considered and is unique for each nucleus, and B_0 is the strength of the magnetic field.

Following placement in the magnet, the rf coil is pulsed in a prescribed manner at the frequency of precession of the nucleus of interest. This frequency is defined by the Larmor equation, and is directly proportional to the field strength. For example, at 2 T ^{1}H nuclei precess at 85 MHz, while at 7 T the frequency of precession is

300 MHz. The rf pulse creates a magnetic field, designated B_1, which is perpendicular to the static magnetic field created by the superconducting magnet. This causes the net magnetization of the spins to ''tip sideways'', with the degree of tipping depending on the strength and duration of the applied rf pulse. When the rf radiation is turned off, the spins realign with the static magnetic field. During realignment, energy is given off, which is detected by the rf coil as an alternating current. These detected signals are then amplified, filtered, and sent to the computer for digitization, recording, and processing. The process of exciting the nuclei by application of an external rf pulse, realignment, and signal collection is repeated many times for acquisition of an image as well as for most spectra.

For NMR spectroscopic experiments, slight variations in the frequency of precession of individual nuclei are created by interactions with the local environment, creating slight localized changes in the magnetic field, for example patterns of binding or paramagnetic interactions. Spectroscopic experiments, therefore, often provide detailed information describing the molecular environment of the nuclei of interest. There is no information regarding spatial localization of the nuclei in traditional NMR spectroscopy.

MRI is similarly based on the Larmor equation. It capitalizes on the frequency of precession being dependent on the specimen position within the magnet bore. Electromagnetic gradients within the bore of the superconducting magnet create linear, defined increases and decreases in the static magnet field, depending on the direction and distance down the bore from the magnetic field center. If we again consider the Larmor equation, the addition of magnetic field gradients gives the relationship shown in Equation (2)

$$w = \gamma(B_0 + rG_r) \tag{2}$$

wherein r is the position along the axis of the gradient G_r. At the isocenter, or the center position of the gradients, the magnetic field is equal to B_0. The gradient strength increases or decreases linearly from the center position.

1.1.2 Fundamental Parameters

There are three fundamental additional parameters inherent in NMR studies: spin–lattice relaxation time (T_1), spin–spin relaxation time (T_2), and diffusion. In addition, the number of detectable spins plays an important role in the detected signal intensity. Spin–lattice relaxation is a measure of time required for the spin population to return to equilibrium following excitation by the rf pulse. The spin system transfers its energy to its surroundings, termed the lattice, as a result of interactions with other local, fluctuating magnetic fields. The spins gain magnetization in the longitudinal, static orientation, with reorientation defined as an exponential gain. Spin–lattice relaxation time is described by the exponential function in Equation (3)

$$M_x(t) = M_0(1 - e^{-t/T_1}) \tag{3}$$

where $M_x(t)$ is the magnetization in the x plane at time t and M_0 is the initial magnetization. Spin–spin relaxation is an exponential decay process, describing loss of magnetization in the transverse plane B_1 following the rf pulse. Generally, both T_1 and T_2 decrease as water becomes more tightly bound to surrounding substrates and less mobile.

Using imaging experiments, early workers demonstrated that plant roots could be visualized in a variety of fields.[4] The presence of paramagnetic ions frequently found in soils such as Fe^{2+}, Fe^{3+}, Mn^{2+}, and Cu^{2+} shortens relaxation times, with the dominant effect being on T_1 at low concentrations, and increasing in effect on T_2 with concentration. When there are sufficiently high concentrations, especially of particulate iron, the relaxation times may be shortened to such an extent that either the image becomes distorted or there is a localized loss of signal in the area near the paramagnetic or ferromagnetic ions. Relaxation rates can therefore give information not only on water binding properties, but also provide indirect information about the chemical environment of the molecules of interest. For example, the measured T_1 for water in sand can range from 565 ms to 1465 ms, although all sand samples contained 25% w/w water and were of about the same particle size. The differences in relaxation time were due primarily to varied iron contents.[5]

In NMR images, contrast between different regions of a volume being imaged, between a root and the surrounding soil for example, depends entirely upon the inherent properties of the specimen and upon the choice of protocol used for image acquisition.[5,6] Image or spectroscopic acquisition approaches have been developed which enhance differences in T_1, T_2, water distribution, diffusion, and/or flow. Most often, it is a combination of these factors which allows the investigator to distinguish specific features of the specimen under study.

1.1.3 Resolution

NMR spectroscopic signals from molecules in solution have a characteristic lineshape, determined by the nuclear magnetic properties, resonance frequency, and spin–spin relaxation times. The natural linewidth at half-height ($v_{1/2}$) can be described by Equation (4)

$$\frac{1}{T_2} = \pi \Delta v_{1/2} \tag{4}$$

In general, signals or linewidths broaden as molecules become increasingly immobilized and T_2 decreases.

Therefore, spectra of molecules in metabolites and in extracts frequently have narrower linewidths than spectra of molecules in living tissue or in macromolecules (such as polyphosphates in fungal vacuoles or membrane phospholipids). Generally, as lines broaden they become less detectable or appear as broad lines underlying the signal of other molecules. Magnetic field homogeneity is a critical feature in NMR studies as resonance frequency is determined by magnetic field strength. If the magnetic field is not homogeneous in a spectroscopic study, significant line-broadening can occur, reducing the detectability of nuclei of interest and causing overlap of adjacent peaks. Nuclear magnetic imaging experiments also rely on a homogeneous magnetic field created by the static superconducting magnet. Without this static homogeneous magnetic field, magnetic field gradients created by the electromagnets may not provide the defined, linear magnetic field upon which good spatial resolution is dependent.[2]

Resolution in MRI refers to the digital resolution and is calculated by the field of view size divided by the number of pixels used to create the image. For example, an image of an apple might have a field of view of 100 mm, divided by 256 pixels, thus giving a resolution of 0.39 mm, the measurement on each side of each pixel. Images are acquired from three-dimensional (3-D) specimens, creating a two-dimensional (2-D) image from something which has volume. Typically, in a single slice plant imaging experiment, a slice might be 1 mm thick. Therefore, in an MRI experiment, the term voxel is used to denote the 3-D properties of the specimen. In the apple example, therefore, a voxel would have dimensions of 0.39 mm $\times$ 0.39 mm $\times$ 1 mm, or a volume of 0.152 mm^3.

Resolution considerations also usually dictate the choice of imaging equipment. Generally large bore scanners, as would be used in clinical applications, provide resolutions of 0.9 mm or more with a slice thickness of several millimeters. While these resolutions are appropriate for large specimens, generally environmentally relevant samples such as fine roots of plants are of a much smaller diameter and would likely not be detected under such imaging conditions. In contrast, if one is interested in the water distribution throughout a large specimen such as a mature taproot, such equipment might be most appropriate.

An additional consideration with respect to resolution is the process of diffusion, whereby nuclei (i.e. water for most imaging experiments) physically move between voxels during the period of image acquisition. This movement causes loss of signal as the water moves from voxel to voxel, changing precessional frequency as it experiences slight changes in magnetic field strength. This sets the theoretical limit of in-plane resolution to about 1 μm, based on diffusion losses in signal detection.[7] With recent developments in rf coils, the practical resolution in certain cases is approaching this theoretical limit, but generally resolution this fine also dictates a need for small specimen sizes.

1.2 Limitations

There are several limitations to the use of MRI in the study of environmental samples. These include the following: the subject for study must generally be smaller than 30 cm in diameter (to fit into the bore of a high-resolution imaging magnet); the specimen must be removed from the study site and brought to the lab for study since it must be placed within the magnet bore; and acquisition of a high-resolution spectrum may take several hours. New pulse sequences (the protocols used for acquiring images and spectra) are continually under development for medical applications, and many focus on shortening the time required to complete a study, which may reduce the time limitation. Many of these new pulse sequences show promise for plant, root, and soil studies. Several previously published reviews summarize applications for MRI in the study of plant physiological processes.[5,8]

One of the primary disadvantages of NMR is the inherent lack of sensitivity, which can be expressed as a signal-to-noise ratio (S/N) between spectral lines and the background signal (in spectroscopic experiments) or between the signal detected in each pixel and the background signal (in imaging experiments). The S/N is dependent on a wide range of factors including:

(1) relative sensitivity of the specific nucleus;
(2) the volume of the specimen or sample contributing to each signal;
(3) magnetic field strength;
(4) the design and construction of the rf coil;
(5) the number of times each pulse is repeated and data collected;
(6) line-width of the detected signal;
(7) relaxation times and their relationship to the chosen pulse sequence and data acquisition parameters;
(8) concentration of the specific detectable nuclei (i.e. specific isotopes) present in the sample or specimen.[2]

Generally the minimum concentration required for a specific molecule under study to be detected is 0.2 mM, although experiments can be designed for detection of molecules both above and below this concentration.

There is one fundamental difference between NMR spectroscopic and imaging experiments when determining sensitivity. If we consider a specimen with a volume of 5 mL, when a spectroscopic experiment is performed the signal is acquired from the entire 5 mL volume and is used to produce the resultant spectra. In imaging, however,

spatial encoding digitally separates the specimen into many voxels, each of which is probed individually and contributes its own signal. Many 3-D image sets are comprised of 128 ''slices'', with each slice arranged in 256×256 pixels. The entire volume is thus divided into 838 8608 voxels. In imaging, therefore, the 5 mL volume is divided into 838 8608 voxels, each with a volume of 0.0006 μL, which is many orders of magnitude less than the volume providing the signal in a spectroscopic experiment. As the signal detected is proportional to the number of nuclei, the signal detected from a single voxel is far less than the signal detected without spatial encoding. For this reason, although many other nuclei can be probed with NMR spectroscopy, MRI is generally restricted to nuclei such as ^{1}H, which are most readily detected and are highly abundant.

1.3 Potential Applications

Based on its versatility and ability to analyze relatively unprocessed samples, there are numerous applications for NMR techniques in environmental analysis. NMR spectroscopy is often used for the chemical analysis of environmental samples, either directly showing the chemical structure of compounds or through isotopic labelling techniques using NMR-detectable nuclei to show changes in the chemical structure of compounds of interest. Materials studied cover a wide range, including lignin from plants, the fate of pesticides in the environment, phosphorus metabolism in plants, carbon chemistry both in vivo and during decomposition in the environment (root decomposition in soil for example), and speciation of aluminum in biological material as well as in solution (surface waters, for example). Field applications have included the construction of a mobile field unit for the measurement of soil moisture (as described in section 2.2), as well as the study of water content and potential in soil samples. Several texts describe the range of NMR spectroscopic applications to environmental analysis.[9–12]

NMR imaging provides spatial information but lacks the sensitivity required for chemical analysis for most detectable nuclei. This technique has found wide acceptance as a diagnostic for human health, and use for environmental monitoring has been strengthened by developments within the medical community. Applications include the imaging of soil, providing some of the first views of soil properties and biology in undisturbed samples. MRI provides an alternative technique to the traditional study of soil which depends on destructive sampling (i.e. removal of plants from the soil volume or disruption of a sample soil core). The strength of the technique is in its ability to nondestructively show root and water distribution within a soil column, either within a core of field soil or a potting container. MRI has also proven extremely valuable in the study of intact, living biological specimens, including animal models for toxicological studies, plant structure and anatomy, water transport and distribution through plant roots, stems, and leaves, and rooting architecture and function.

2 ENVIRONMENTAL APPLICATIONS

2.1 Magnetic Resonance Imaging of Soil

The study of plants in soil has historically presented significant challenges to researchers interested in investigating undisturbed or intact systems. The inherently opaque nature of soil makes direct study of roots or of water distribution within soil difficult without using techniques for remote sensing, such as mini-rhizotrons, X-ray diffraction, computed tomography, γ-ray attenuation, or MRI.[13,14] The researcher is physically separated from the specimens, such as plant roots or water distributed throughout a volume of soil, and disruption of the soil volume destroys the integrity and geometry of the system.

Soil can present many challenges to MRI and spectroscopic experiments, owing to the inherent heterogeneity of soils. Early in MRI research, it was demonstrated that MRI techniques have the potential for nondestructively observing the spatial distribution of water and large roots in soil.[4] In an extensive examination of the MRI properties of soil, plant roots were imaged when potted into 30 different agricultural and 8 artificial soils. Image quality and the ability to detect roots suffered significantly when the ferromagnetic particle content was greater than 4%. The presence of paramagnetic or ferromagnetic ions and particles within a soil volume creates localized alterations of the magnetic field within the soil volume, distorting the signal received and creating significant artifacts in the subsequent image. Below 4%, there was still significant variation in the suitability of a soil as a root imaging substrate.[15] This is especially important if the objective of the experiment is to make quantitative measurements of root volume, root area, water content, water distribution/flow/diffusion within the soil volume, or other similar measurements.

2.1.1 Experimental Design Considerations

There are several choices to be made when selecting an MRI unit, with a range of sizes and magnetic field strengths determining sensitivity and possible image resolution. The units commonly used for medical applications are constructed of a superconducting magnet with a field strength ranging from 0.5 T to 3 T (1 T or tesla = 10 000 gauss). The size, weight, cost, and technical

difficulties increase with magnetic field strength, however, making high-field magnets with bores large enough to image intact biological specimens or human patients generally impractical with current technology.

In contrast, MRI units that are primarily designed for research generally have magnets with a field strength from 2 T to 9 T. The size of the opening in the magnet for specimen position generally decreases with field strength, ranging from 30 cm to less than 1 cm, depending on the magnet design. For small-volume samples in which the highest resolution is desired, a small-bore, high-field magnet is usually most appropriate. If the specimen is large and the highest resolution may not be necessary, for example in the imaging of logs or large soil volumes, larger bore, lower field strength magnets are used.

2.1.2 Water Content and Distribution

Owing to its nondestructive and noninvasive attributes, water potential, content, and movement can be determined through NMR imaging and spectroscopic techniques. MRI has been demonstrated to give information on the pore structure and wettability of natural sandstone, providing visualization of wetting heterogeneities.[16] The T_1 relaxation time of water in natural sandstone was shown to increase significantly with silanization, changing the wettability of the surface from moderately hydrophilic to hydrophobic. Samples were infiltrated with water, then imaged. Subsequently, there have been other studies using MRI techniques to describe water distribution and flow through porous media such as sandstone, providing information describing the complicated flow and diffusion patterns.[17,18]

Similarly, the binding properties of water by fine sand as reflected by the relationship between relaxation time and water content have been studied. T_1 values increased from 472 ms when containing 5% water (w/w) to 1265 ms at 25% water. The increase in T_1 was linear between 5% and 20% with a slight reduction in slope between 20% and 25%. The water potential was measured as between 0.01 and 0 MPa, indicating lack of tight water binding by the sand grains.[5,19]

In addition to the study of soil/water interactions and the water content of soil, MRI provides the ability to study plant/soil interactions at a resolution not previously possible. In addition to spin–lattice relaxation time changes with water content, the detected signal intensity of water in sand increased linearly with water content with some pulse sequences complementing study by more traditional techniques such as neutron activation probes or soil moisture blocks. Studies of both the water content and distribution within the soil/sand matrix surrounding roots and visualization of the undisturbed root system can be achieved. The linear relationship between signal intensity from moist sand and water content allows direct measurement of the water content of sand by MRI when a standardized reference tube filled with copper sulfate solution is placed in the image field of view. The development of varied pulse sequences and the ability to control the environmental conditions within the magnet bore provide the possibility for study of patterns of water distribution and binding within sand (and the influence of roots of transpiring plants) as well as the architecture of roots in an undisturbed soil volume.

Experiments have been conducted to measure quantitative uptake of water from sand by plant roots. In these experiments, pine seedlings were planted into fine sand, watered to field capacity, and imaged repeatedly during a period of transpirational water uptake. Plants remained undisturbed within the bore of the magnet and continued to transpire under illumination from a fiber optic lamp placed into the magnet bore. Over time, regions of low detectable signal were observed to form around the roots which were shown to be regions from which water was withdrawn from the sand by the plants. The ''dark'' regions surrounding the roots were first observed around the top of the taproot and the fine roots. The margins were seen to expand over time, showing expansion of the root-associated water depletion zones as the plants continued to withdraw water from the sand through transpirational uptake.[5,19,20]

Water uptake by specific roots can be determined by MRI. Specific regions defined as containing pine seedling roots were imaged, and the areas of the water depletion zone surrounding each root were measured. The amount of water extracted from sand was measured by determining the volume of the water depletion zone by tracing the margins visualized in the images and determining the water content of the sand both within and outside the depletion zones. Comparative measures of water extraction for fine roots, woody lateral roots and the taproot of loblolly and red pine seedlings were calculated. It was found that the fine roots were very efficient at water extraction based on root weight and surface area. What was perhaps more surprising, however, was the formation of a water depletion zone surrounding the developing taproots.

In these experiments it is necessary to first calibrate a standard curve for detected signal intensity, and to include a reference material within each field of view. It is useful to include a glass tube filled with a mixture of water and $CuSO_4$ solution. The signal intensity of the reference tube and of specific regions of sand are measured from the acquired images. The ratio of these two signal intensities is then fitted to a predetermined regression equation, and the water content of the sand can then be estimated.

This method has provided some of the first views of water extraction by plant roots, and clear evidence that

the woody taproot of pines contributes to water uptake. There are some limitations to the method, however. The first is that image acquisition takes approximately 1 h, using a multislice technique wherein images of nonadjacent slices can be simultaneously imaged. For very rapidly transpiring plants, however, this may not be sufficient time for acquisition of high-quality images. A second limitation is that acquisition of a true 3-D volumetric image of the entire soil column and root system requires substantially more time – up to several hours in some cases. Again, if the subject of study is rapidly transpiring, the volume of the water depletion zone may change substantially during the period of image acquisition. The strength and power of the technique, however, generally outweights its limitations in that it provides definitive information showing the volume and location of the water depletion zone, its spatial relationship to the root system in situ, and water extracted by individual identified roots. A single container can also be repeatedly studied, showing changes in water distribution through time.

2.1.3 Root Distribution and Physiology

Another application is in the direct study of root geometry, architecture, growth, and turnover within the soil column. The nondestructive nature of MRI and NMR spectroscopy allows repeated study of a specific root system. When a root becomes nonfunctional with respect to water uptake, it becomes extremely difficult to distinguish it from the surrounding soil matrix, presumably because it has equilibrated in water content and binding with the soil around it. Images with contrasts primarily weighted by differences in T_1 show substantial differences with respect to signals detected between roots and water in the surrounding sand. Thus, root elongation, senescense, and replacement can be seen. This allows functional pine roots to be readily distinguished from nonfunctional roots, a critical distinction with regard to plant health, within an undisturbed soil volume (Figures 1 and 2).

The development of MRI protocols for the acquisition of truly 3-D image sets allows digital extraction of the root system from the soil column, without physical disturbance of the root system in soil. The digitally extracted root system can be measured for volume and surface area using software packages such as VoxelView (BioImage, Ames, Iowa). If appropriate markers are placed in the soil column, or are identified on the root system, image sets can even be registered with one another so more subtle changes can be studied by placing one digital root system alongside or within another. A root can be repeatedly imaged at multiple points in time, allowing quantitative analysis of growth, death, and replacement.

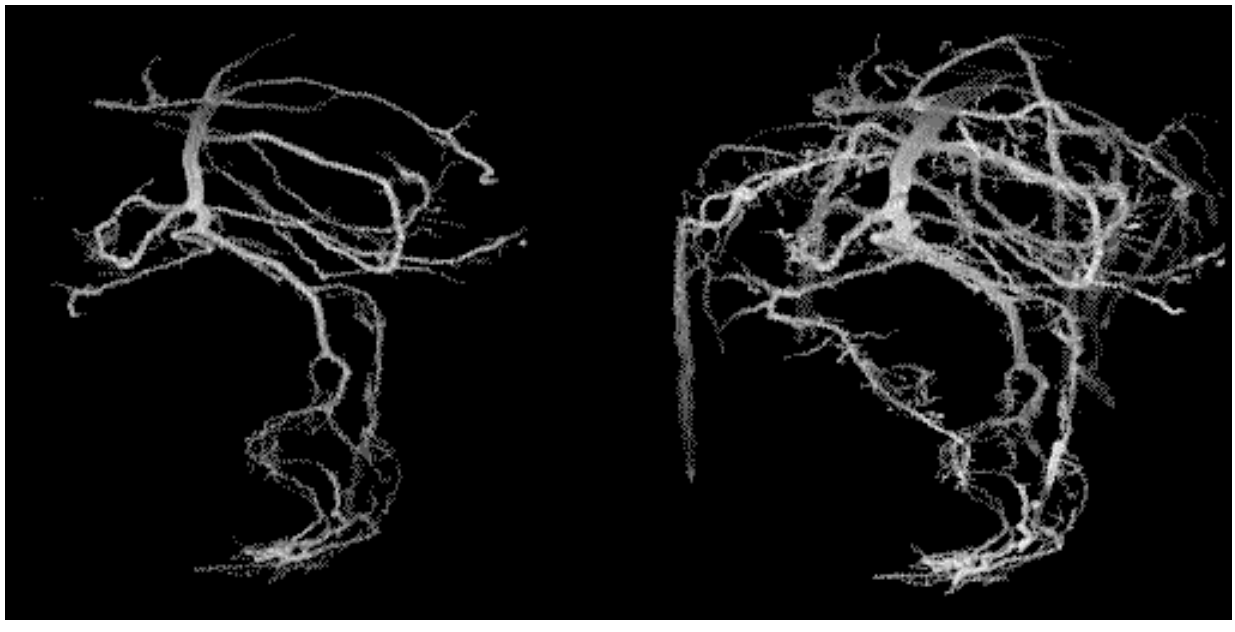

(a) (b)

Figure 1 3-D image acquired by MRI of a pine seedling root system. The plant was grown in a sand-filled container and watered routinely to field capacity. The image set was acquired with a spin–echo pulse sequence. The image set has been volume rendered to make the sand and water surrounding the root system transparent, allowing the intact root system to be viewed without removal from the container. Both images are of the same seedling: (a) seedling at approximately 6 weeks of age; (b) seedling at approximately 10 weeks of age.

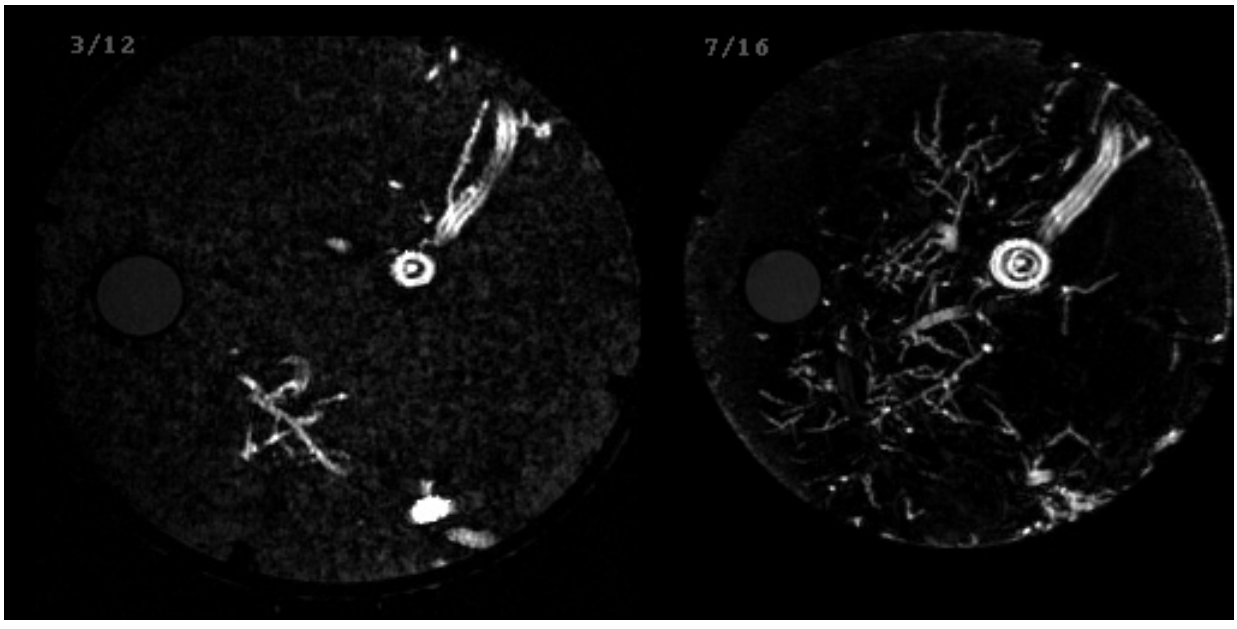

Figure 2 Single slice image through a sand-filled container planted with a pine seedling. The view shows a cross-section through the taproot and bunches of fine roots within the soil volume. Slice thickness was approximately 1 mm. Seedling on left was approximately 6 weeks old. Seedling on right was approximately 20 weeks old.

Again, however, one of the present limitations of the technique is the time required for acquisition of a high-quality 3-D image set using the more common image acquisition approaches. A time requirement of 3–7 h is not uncommon for a single 3-D image, depending on the study requirements. As use of the equipment can be extremely costly, and plant study applications are often competing with studies which generate substantial funding (drug trials, for example, or clinical diagnoses from paying clients), this time requirement may become a serious limitation.

2.1.4 Future for Magnetic Resonance Imaging Study of Soil Systems

The technology is constantly changing, with new imaging protocols in development. In clinical research, it is to the diagnostician's benefit to decrease the time required for a clinical diagnosis so more patients can have access to the machine in a given period of time. These interests have caused the recent development of many new pulse sequences with a substantially shortened time requirement. While many are not applicable to plant root studies because of the S/N reduction (the smaller voxel size in most plant studies dictates less signal detected for each voxel) compared with medical applications, several hold promise for bringing MRI from the realm of casual interest to becoming a true research tool in the study of plant roots and below-ground processes. Unfortunately, the number of plant researchers who have access to such equipment is limited at present, but with time hopefully this will change.

2.2 Field Application of Nuclear Magnetic Resonance Spectroscopy for Soil Study

One unique application of NMR in environmental monitoring is the use of a mobile, low-magnetic-field-strength unit used in forestry and field settings to measure soil moisture. An NMR unit was designed to be pulled at relatively slow speeds behind a tractor, allowing relatively rapid measurement of surface soil water content over large areas. Field tests have been made at ground speeds from 0–17 km h^{-1}, with a correlation to water content measurements made gravimetrically estimated to be $r = 0.98$.[21]

As the emphasis in environmental monitoring expands to the examination of landscape-scale questions, development of methods to acquire environmentally relevant information at these scales becomes increasingly critical. In support of these studies, remote-sensing techniques (aerial photography and satellite imagery) have emerged as critical tools in the examination of environmental issues. Interpretation of these information sets, however, requires careful ground validation and supplemental testing of the land surface. The mobile NMR spectrometer was designed to characterize soil water content over large areas, to calibrate and verify remote sensor output. Accurate ground-truth information to verify interpretation of measurements made through remote sensing is critical to

accurate interpretation of remote sensor output. Water content, therefore, was determined in the field by measuring the strength of the ^{1}H NMR signal from water in the soil.

2.2.1 *Experimental Design Considerations and Limitations*

Construction of a unit designed for soil water measurements in a field setting requires a magnet to generate the static magnetic field required for resonance, an rf coil with tuning capacitor, and electronics to control the pulse sequence and data acquisition and processing. In addition, the unit must be sufficiently rugged to withstand movement across uneven terrain and exposure to the out-of-doors. In this design, an electromagnet rather than a superconducting magnet was used. The magnet and rf coil with tuning capacitor were designed to be mounted on the back of a tractor, with the electronics located in a weatherproof, air-conditioned steel enclosure mounted just behind the tractor cab. The magnet was approximately 60 cm long with a pole separation of 23 cm. It operated at a relatively low field strength of 700 G. The rf coil operated at a frequency of 3.0 MHz. Electrical power was supplied by a generator mounted on the front of the tractor. Electronics included rf pulse generators, amplifiers, signal processors, and the magnet power supply. The electrical requirements for magnet operation are determined by depth of measurement. Approximately 300 W of electrical power is required for a measurement depth of 51 mm, but approximately 500 W is required with the sensor in the flat configuration. The magnet was fabricated from 512 turns of 0.16-mm thick Al ribbon, bonded to a steel core and encased in a nonmagnetic stainless-steel housing for environmental protection.

Special consideration was needed in the design of the magnet and rf coil. In most laboratory applications, the sample is surrounded by the rf coil, and the assembly is placed in the center of a superconducting magnet. In this arrangement, the sample (the soil surface) was external to both the rf coil and the electromagnet. The size and location of the measurement region are determined by the placement and magnitude of the static magnetic field and the rf field. The poles of the electromagnet are positioned ahead of the measurement region, below the rf coil. As the unit moves across the landscape, nuclei become aligned in the static magnetic field prior to reorientation by the rf field pulse. In this arrangement, the time for the nuclei to align with the static magnetic field is determined by the magnet extension in front of the rf coils, and the transport speed.

In addition to the design constraints, the unit must be constructed for the soil to contact or nearly contact the surface of the unit. The sensor unit can be designed with two different configurations, depending on the application. In the flat arrangement, the sensor rides directly on the soil surface. This provides a stationary arrangement, most suited to flat terrain. In the extended configuration, extended magnetic pole pieces can be positioned into two 10-cm deep furrows. The poles can be adjusted to different soil depths, and NMR measurements of soil water content have been shown to be highly correlated with gravimetrically measured soil water contents at 38, 51, and 64 mm below the soil surface.

The unit was field tested by examining the soil water content of three clay samples (montmorillonite, kaolinite, and illite) and sandy loams with varied levels of organic matter. The soil water contents were measured in the laboratory using standard gravimetric techniques (oven drying and weighing), for comparison with NMR determined measurements.

2.2.2 *Software Considerations*

A spin–echo pulse sequence was used to measure soil water content, with the amplitude of the spin–echo a measure of the water content. In this sequence, spins are initially aligned parallel or antiparallel with the static magnetic field created by the electromagnet. The rf coil is pulsed, causing spins to realign to a 90° orientation relative to the static field. Spins are pulsed a second time, creating a 180° flip of the spins. The rf coil is then turned off, and the spins are allowed to realign with the primary, static field. Use of the second 180° pulse creates the formation of a signal "spin–echo", detectable by the rf coil. The rf pulse sequence is completed in 40 μs, and it is repeated 10 times per second, making the pulse sequence repetition time 40 ms. Accumulated signals are acquired for 25.6 s, then averaged.

Spin–lattice and spin–spin relaxation times were determined for water in loamy fine sand (400 ms/40 ms and 200 ms/1 ms respectively) and silty clay (200 ms/5 ms and 130 ms/700 ms respectively). Relaxation times are a reflection of the degree of binding and interaction between water and the surrounding matrix, (soil particles in this case). For the signal detected to be proportional to the number of spins in the sample in NMR spectroscopic experiments (water content in soil in this case) the repetition times between rf pulses should be equal to at least four times T_1. Under most conditions, this constraint is met, so the detected signal in the array is proportional to the water content in soil with this choice of pulse sequence and application.

It should be noted, however, that both T_1 and T_2 values decrease with increasing field strength of the stationary magnet. In addition, the relaxation time for water in other soil types must be determined, since the binding potential of soil and the presence of paramagnetic or

ferromagnetic ions in the soil will greatly affect both spin–spin and spin–lattice relaxation times. Therefore, relaxation times should be determined for each apparatus and field application, and appropriate pulse sequence parameters established for water content determinations.

2.3 Aluminum in Surface Waters

Some forms of Al(III) have been demonstrated to exhibit toxic effects in aquatic and terrestrial ecosystems, affecting fish, plants, mammals, and other organisms, making aluminum study in environmental samples extremely important. Development of methods for the detection and measurement of aluminum is crucial to our understanding of Al(III) chemistry and toxicity in natural waters, soil solutions, and environmental samples. Traditional colorimetric assays, by their very nature, alter Al(III) solution chemistry, potentially biasing measurements. Chemical analysis by complex formation gives specific information on the complexed aluminum species, but may not provide information on other Al(III) forms present in the sample. Magnetic resonance spectroscopic techniques have been developed to measure unaltered Al(III) in aqueous solution, whether from soil, surface waters, environmental samples, or extracts, providing an important tool for the study of this complex, environmental toxicant.

^{27}Al NMR spectroscopy (spin 5/2, 100% natural abundance) allows study of the chemical characteristics of the sample without alteration of the inherent chemistry. In addition, as with NMR imaging, the analysis is nondestructive, allowing repeated examination of the sample as needs dictate.

2.3.1 Experimental Design Considerations and Limitations

While Al exists as the most abundant element in the Earth's crust, it usually occurs at dilute levels (in the order of 10^{-5}–10^{-6} M[9–12]) in areas of environmental interest such as lakes, rivers, streams, biological samples (fish, plants, etc.), and soil extracts. This is at least one order of magnitude below the sensitivity of most commercial NMR probes. In addition, most vendor-supplied NMR probes and sample tubes designed for multinuclear NMR contain appreciable amounts of Al, resulting in high background signals that mask the desired signals from natural samples.[22]

Techniques to circumvent interference from the probe background include:

(1) study of solutions with concentrations $> 10^{-3}$ M;
(2) filtering the broad background signal by increasing the preacquisition delay from 150–200 μs to few microseconds;
(3) use of a spin–echo pulse sequence to edit out the background; and
(4) use of difference spectra to digitally subtract the probe background.

Since samples of biological and environmental relevance generally have levels of Al in solution well below 10^{-3} M, studies at high concentrations are often impractical. Concentration of samples to improve sensitivity is also likely to alter the solution chemistry of Al, making this a less-than-desirable choice. Filtering and spin–echo methods can sometimes be effective, but fail when the background and resonance-of-interest have similar linewidths, and may potentially decrease the sensitivity of detection. In particular, broad resonance from Al(III) complexes with biologically important ligands (e.g. organic acids, phosphate, and hydroxy ligands) could go undetected. In difference spectroscopy, the low levels of detectable Al in natural samples make subtraction uncertain, resulting in incomplete background subtraction and ''baseline roll''.

To address these challenges and to produce a probe for NMR spectroscopy of Al in solution, a novel probe was designed and constructed to reduce probe background and maximize sensitivity for the detection of ^{27}Al in solution. The successful construction of such probes provides new tools for the study of Al(III) speciation in dilute, natural samples collected from the environment.[23,24]

2.3.2 Coil Design for Environmentally Relevant Measurements

A novel probe was designed and constructed for analysis of environmentally relevant concentrations of Al both in solution and in vivo. To minimize the detectable Al background signal from the probe itself, materials for construction were selected based on prior analysis by neutron activation analysis for Al content. Three primary components were constructed:

(1) a probe body constructed of polycarbonate replacing the aluminum components of the housing in commercial probes;
(2) a solenoid rf coil designed for increased sensitivity to ^{27}Al from the sample, tuned by inductive coupling of three separate copper rings rather than by a variable capacitor (which are usually constructed of either ceramic or sapphire, both of which contain detectable aluminum); and
(3) a sample holder specifically milled from Teflon® to fit into the rf coil support.

Two separate chambers were constructed, one for a reference solution (8×10^{-4} M $Al(OH)_4$, pH 13) and

one for the sample material, either in solution or as biological material. The magnet was a 7-T 89-mm wide-bore vertical magnet. The solenoid coil was assembled around a 3.6-cm-o.d. cylindrical Teflon® insert.

To construct the coil, a 3-cm × 11-cm sheet of 2-mil (50-μm) thick microwave substrate was etched to remove about two-thirds of the copper from one side of the strip, while leaving all the copper intact on the other side. The strip was soldered to form a coil sliding freely on the coil support. Two additional copper loops, each 1 mm wide, were placed on the coil form on either side of the main coil. One of these was attached to the coaxial cable to serve as an inductively coupled pickup coil. Tuning and matching were accomplished by adjusting the distances between the three coils on the coil support.

To additionally reduce pickup of aluminum signals from the instrument hardware, unshielded coaxial cables were used for connecting electronic components in the NMR probe, as aluminum braid is frequently used as rf shielding. To reduce background detection, the entire probe was wrapped in copper foil.

2.3.3 Pulse Sequence Design

The coil was tuned to 78.199 MHz at 22 °C, and operated without lock. The internal reference material was used to normalize the spectrum. Acquisitions generally included a 40-μs pulse, with a 348-ms recycle delay, a 51-μs readout delay, 8192 points, 11 764-Hz sweepwidth, and 10-Hz line broadening. The number of transients collected varied between samples, depending on aluminum concentration.[2]

2.3.4 Sensitivity Enhancement

The study of Al in natural samples, without alteration of the chemistry present at the time of study or sampling, requires sensitivity as well as specificity. NMR spectroscopy under specific conditions can provide the sensitivity needed for these kinds of studies. Several considerations need to be made to enhance detection. The first is that not all Al ligands are readily detectable. As a quadrupolar nucleus, ^{27}Al has an inherently broader line and shorter T_2 times than dipolar nuclei such as ^{1}H. The less symmetrical the Al-containing molecule, the broader the line and the less detectable the ligand. This property in itself provides information describing the molecular geometry, but also presents a challenge to detection. Therefore, molecules which are most likely to be detected at environmentally relevant concentrations are those which are symmetrical, such as $Al(OH)_4^-$ or $AlCl_3$.

A second consideration is that ^{27}Al gives a substantially less detectable signal than ^{1}H nuclei, a property linked to both the quadrupolar nature of the nucleus as well as the gyromagnetic ratios. Having more sample available for study is an asset, so use of a wide-bore magnet is extremely valuable in the NMR study of ^{27}Al. In previously reported studies, approximately 11 mL of solution was within the sample volume detectable by the solenoid rf coil. Construction of larger diameter sample containers, to be used at higher field strengths (also increasing detectable signal), would improve the detectability of ^{27}Al as a potential candidate for environmental NMR study.

2.4 Carbon in Soil – Nuclear Magnetic Resonance of Humic Materials

Humic substances are complex mixtures of organic acids found in soil sediments and in natural waters of the geosphere, and are derived mainly from the weathering and decomposition of plants. Humic material includes macromolecules that have ill-defined structures and are lacking in repeating units. This molecular heterogeneity and complexity has left open to debate many fundamental questions about the structural chemistry of humic substances. NMR spectroscopy has provided useful insights that broaden our understanding of the origins of humic substances, their influence in land and water management, and the fate of pollutants and contaminants.

Humic substances have been extracted and/or methylated and studied in deuterated solvents since the earliest continuous-wave ^{1}H and ^{13}C NMR studies of humic materials.[9] The first good Fourier transform ^{13}C NMR spectra showing detailed separation of aromatic and aliphatic resonances of dissolved humic and fulvic acids appeared in 1976.[9] Direct examination of humic substances in the solid state avoids structural changes due to solubilization, and the first solid-state cross-polarization magic angle spinning spectra of humic material appeared in 1980.[25–27] This is of particular benefit for NMR studies of the insoluble humin fraction. Complete references to early NMR work are found in reviews published in 1983[28] and in two books published in 1987.[10,29] Recent reviews have focused on new ^{15}N and ^{13}C NMR enrichment techniques to trace the reactivity of humic substances with regard to degradation and interaction with contaminants and pollutants.[30,31] The International Humic Substances Society has catalogued standard and reference fulvic and humic acids by ^{13}C and ^{1}H NMR spectroscopy.[32]

Traditionally, humic substances have been separated into humic acids (base-soluble fractions), fulvic acids (fractions soluble under all pH conditions) and humins (insoluble fractions). Precipitation of the humic acid fraction at acid conditions could hydrolyze phosphate esters. Thus, many researchers today prefer to use

the total extract of the geo-organic material instead of individual fractions, particularly when preparing samples for ^{31}P NMR. Removal of paramagnetic ions, e.g. reduction of Fe^{3+} and Mn^{2+} with dithionite[33] can lead to improved NMR spectra of humic substances, particularly for quantitative data[9,31,32] in both the solution and solid state. Removal of mineral matter with HF also improves the NMR spectra of organic matter in sediments and soils.[34]

^{13}C NMR spectra of humic substances show overlapping signals in four distinct chemical shift ranges: aliphatic carbon signals (0–50 ppm) from alkyl side chains in amino acids, lipids, and plant polymers; *O*-alkyl carbon signals (50–110 ppm) from plant carbohydrate and degradation products; aromatic carbons (100–160 ppm); and carbonyl carbon signals (170–200 ppm) from carboxylic acid and carboxamide groups.[29,35] ^{1}H NMR in the liquid state with poorly resolved envelopes in the alkyl (0–2.5 ppm), *O*-alkyl (3–4 ppm), and aromatic (6–8 ppm) region is by comparison much less widely used.[9,10] ^{31}P NMR of soils and sediments allow observation of the inorganic phosphate pool (P_i) and the organic phosphate pool (P_o) such as phosphate esters, pyrophosphates and phosphonates with their direct carbon–phosphorus bond.[36,37] ^{15}N NMR of soil humic acids incubated with ^{15}N-enriched nitrate showed signals from secondary amides, most likely from protein material.[38]

A brief survey of NMR applications to humic substances reveals recurring themes of comparisons of humic materials from differing geochemical sources, studies of the decomposition processes in humification, studies of binding to metal ions, studies of interactions with pollutants and contaminants, and studies of formation of adducts.[39] Thus, ^{13}C NMR reveals dissolved humic substances from seawater to be enriched in total carbon but depleted in saturated carbon relative to humic substances in river water. ^{31}P NMR reveals persistence of phosphate esters and phosphonates at various depths in marine sediments from pelagic, lower shelf, and continental shelf depositional environments.[40] Solid-state ^{13}C NMR yields the insight that vegetational background (peat bog, spruce, pine, and oak forests) has a significant influence on the chemical composition of the humus in the soil[41] and that carbohydrates decrease in relative concentration while paraffinic substances persist in the decomposition processes of mangrove leaves in a tropical estuary.[42] The solid-state ^{13}C NMR spectra of humic acid, fulvic acid, and humin fractions sum to give a composite spectrum virtually identical to that of whole soil, indicating that conventional humic material fractionation does not change the carbon content of different functional groups in soil.[43] Effects of land use are seen in ^{31}P NMR results, revealing a larger inorganic phosphate pool in grass and arable soils compared with spruce and deciduous forest soils.[44] Solid-state ^{13}C and ^{15}N NMR of ^{13}C- and ^{15}N-labeled decomposing ryegrass reveals the *O*-alkyl-C signal (probably from carbohydrates) and the ^{15}N signal (probably from ^{15}N-amides) decreased, while the alkyl- and methyl-C content of soil increased over time.[45,46]

2.5 Pesticide Monitoring In Vivo and In Soil

Dissolved geo-organic matter in lakes and rivers can increase the solubility of pesticides, fertilizers, pollutants, and contaminants and threaten the quality of water supplies. Insights into the interactions of these pollutants with humic materials are needed to understand the migration of these pollutants in the water table. Chemical binding of pollutants to humic substances might decrease the risk of further transport. Irreversible binding, e.g. formation of stable chemical adducts, might be a natural detoxification process. Thus, while much early work centered on assessing the stereochemistry of pesticides themselves and some metabolites,[12] current approaches focus on the combination of labeling and NMR strategies to investigate specific molecular interactions. Anderson has reviewed ^{1}H and ^{19}F solution NMR of pesticide intermolecular interactions,[47] discussing solution complexes of the herbicide atrazine, and the pesticide 1,1,1-trichloro-2,2-bis[*p*-chlorophenyl]-ethane (DDT), DDT interactions with serum albumin and with phospholipid membranes, and fluoroaromatic–humic acid molecular interactions. Bortiatynski et al.[48] have presented ^{13}C labeling in combination with T_1 relaxation time measurements to study the binding of chlorophenols to humic acid. A covalent complex forms in the presence of an oxidoreductive enzyme. Herbert and Bertsch have used ^{2}D and ^{19}F NMR to study the interaction of nonionic organic contaminants with a Lakeland soil.[49]

ACKNOWLEDGMENTS

The Duke NMR Center is partially supported by NIH NCI P30-CA-14236. NMR instrumentation in the Duke NMR Center was funded by the NSF, the NIH, the NC Biotechnology Center and Duke University. The Duke Center for In Vivo Microscopy is supported by funds from the USDA, the NSF, and the NIH.

ABBREVIATIONS AND ACRONYMS

DDT	1,1,1-Trichloro-2,2-bis[*p*-chlorophenyl]-ethane
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance

rf	radiofrequency
S/N	Signal-to-noise Ratio
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Imaging, Functional • Multinuclear Magnetic Resonance Spectroscopic Imaging

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Neutron Activation in Environmental Analysis • Organic Analysis in Environmental Samples by Electrochemical Methods • Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Soil Instrumental Methods • Soxhlet and Ultrasonic Extraction of Organics in Solids

Food **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials • Water Determination in Food

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Nuclear Magnetic Resonance of Geological Materials and Glasses • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction

REFERENCES

1. P.T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Clarendon Press, Oxford, 1991.
2. D.G. Gadian, *Nuclear Magnetic Resonance and its Applications to Living Systems*, 2nd edition, Oxford University Press, New York, 1995.
3. P.C. Lauterbur, 'Image Formation by Induced Local Interactions: Examples Employing NMR', *Nature (London)*, **242**, 191–195 (1973).
4. K.M. Omasa, M. Onoe, H. Yamada, 'NMR Imaging for Measuring Root System and Soil Water Content', *Environ. Control Biol.*, **23**, 99–102 (1985).
5. J.S. MacFall, H. Van As, 'Magnetic Resonance Imaging of Plants', in *Nuclear Magnetic Resonance in Plant Biology*, ed. Y. Shachar-Hill, American Society of Plant Physiology, Rockville, MD, 33–70, 1996.
6. P.T. Callaghan, C.J. Clark, L.C. Forde, 'Use of Static and Dynamic NMR Microscopy to Investigate the Origins of Contrast in Images of Biological Tissues', *Biophys. Chem.*, **50**, 225–235 (1994).
7. C.B. Ahn, Z.H.Cho, 'A Generalized Formulation of Diffusion Effects in μm Resolution Nuclear Magnetic Resonance Imaging', *Med. Phys.*, **16**, 22–28 (1989).
8. T.J. Schaafsma, H. Van As, W.D. Palstra, J.E.M. Snaar, P.A. de Jager, 'Quantitative Measurement and Imaging of Transport Processes in Plants and Porous Media by ^{1}H NMR', *Magn. Reson. Imaging*, **10**, 827–836 (1992).
9. M.A. Nanny, R.A. Minear, J.A. Leenheer, *Nuclear Magnetic Resonance Spectroscopy in Environmental Chemistry*, Oxford University Press, New York, 1997.

10. M.A. Wilson, 'NMR Techniques and Applications in Geochemistry and Soil Chemistry', 1st edition, Pergamon Press, New York, 353, 1987.
11. Y. Shachar-Hill, P.F. Pfeffer, *Nuclear Magnetic Resonance in Plant Biology*, American Society of Plant Physiology, Rockville, MD, 260, 1996.
12. J.D. McKinney, N.D. Wilson, L.H. Keith, A.L. Alford, in *Mass Spectrometry and NMR Spectroscopy in Pesticide Chemistry*, ed. R.A.B. Haque, F.J. Biros, Plenum Press, New York, 136–160, 1974.
13. L.A.G. Aylemore, 'The Use of Computer Assisted Tomography in the Study of Water Movement Around Plant Roots', *Adv. Agronomy*, **49**, 1–54 (1993).
14. S.H. Anderson, J. Hopmans (eds.), *Tomography of Soil–Water–Root Processes*, Special Publication No. 36, Soil Science Society of America, 1994.
15. H.H. Rogers, P.A. Bottomley, 'In Situ Nuclear Magnetic Resonance Imaging of Roots: Influence of Soil Type, Ferromagnetic Particle Content, and Soil Water', *Agronomy J.*, **79**, 957–965 (1987).
16. G. Guillot, C. Chardiare-Riviere, S. Bobroff, A. Leroux, L. Cuiec, J.C. Roussel, 'Characterization of Wetting Heterogeneities in Sandstone Rocks by MRI', in *Workshop on MRI in the Applied Science*, ed. G.A. Johnson, Society of Magnetic Resonance in Medicine, Berkeley, CA, 1992.
17. M. Shattuck, B. Behringer, G.A. Johnson, J. Georgiadis, 'Onset and Stability of Convection in Porous Media: Visualization by Magnetic Resonance Imaging', *Phys. Rev. Lett.*, **25**, 1934 (1995).
18. S.W. Sinton, A.W. Chow, 'NMR Flow Imaging of Fluids and Solid Suspensions in Poiseuille Flow', *J. Rheol.*, **35**, 735–772 (1991).
19. J.S. MacFall, G.A. Johnson, P.J. Kramer, 'Observation of a Water Depletion Region Surrounding Loblolly Pine Roots by Magnetic Resonance Imaging', *Proc. Natl. Acad. Sci. USA*, **897**, 1203–1207 (1990).
20. J.S. MacFall, G.A. Johnson, P.J. Kramer, 'Comparative Water Uptake by Roots of Different Ages in Seedlings of Loblolly Pine (*Pinus taeda* L.)', *New Phytol.*, **119**, 551–560 (1991).
21. R.F. Paetzold, A.D.L. Santos, G.A. Matzkanin, 'Pulse Nuclear Magnetic Resonance Instrument for Soil Water Content Measurement: Sensor Configurations', *Soil Sci. Soc. Am. J.*, **51**, 287–290 (1987).
22. P.M. Bertsch, 'Aqueous Polynuclear Aluminum Species', in *The Environmental Chemistry of Aluminum*, ed. G. Sposito, CRC Press, New York, 87–115, 1989.
23. B.C. Faust, W.L. Labiosa, K.H. Dai, J.S. McFall, B.A. Browne, D.D. Richter, 'Speciation of Aqueous Mononuclear Al(III)-Hydroxo and Other Al(III) Complexes at Concentrations of Geochemical Relevance by Aluminum-27 Nuclear Magnetic Resonance Spectroscopy', *Geochim. Cosmochim. Acta*, **59**, 2651–2661 (1995).
24. J.S. MacFall, A.A Ribeiro, G.P. Cofer, K.G. Dai, B.C. Faust, D.D. Richter, 'Design and Use of Background-reduced 27Al NMR Probes for the Study of Dilute Samples from the Environment', *Appl. Spectrosc.*, **49**, 156–162 (1995).
25. P.G. Hatcher, R. Rowan, M. Mattingly, '^{1}H and ^{13}C NMR of Marine Humic Acids', *Org. Geochem.*, **2**, 77–85 (1980).
26. P.G. Hatcher, D.L. Vanderhart, W.L. Earl, 'Use of Solid-state ^{13}C NMR in Structural Studies of Humic Acids and Humin from Holocene Sediments', *Org. Geochem.*, **2**, 87–92 (1980).
27. P.G. Hatcher, I.A. Breger, G.E. Maciel, N.E. Szeverenyi, 'Geochemistry of Humin', in *Humic Substances in Soil, Sediment and Water: Geochemistry, Isolation and Characterization*, ed. G.R. Aiken, J. Wiley & Sons, New York, 275–302, 1985.
28. R.F. Christman, E.T. Gjessing, *Aquatic and Terrestrial Humic Materials*, Ann Arbor Press, Ann Arbor, MI, 1983.
29. R.L. Wershaw, M.A. Mikita, *NMR of Humic Substances and Coal-techniques, Problems and Solutions*, Lewis Publishers, Chelsea, MI, 1987.
30. M.A. Nanny, '31-P FT of Concentrated Lake Water Samples', in *Nuclear Magnetic Resonance in Environmental Chemistry*, eds. M.A. Nanny, R.A. Minear, J.A. Leenheer, Oxford University Press, New York, 221–246, 1997.
31. J.M. Bortiatynski, P.G. Matcher, H. Knicker, 'NMR Techniques (C, N, and H) in Studies of Humic Substances', in *Humic and Fulvic Acids: Isolation, Structure and Environmental Role*, eds. J.S. Gaffney, N.A. Marley, S.B. Clark, ACS Symposium Series No. 651, American Chemical Society, Washington, DC, 57–77, 1996.
32. K.A. Thorn, D.W. Folan, P. McCarthy, 'Characterization of the International Humic Substances Society Standard and Reference Fulvic and Humic Acids by Solid-state Carbon-13 and Hydrogen-1 Nuclear Magnetic Resonance Spectrometry', US Geological Survey Water Investigations Report 89–4196, Denver, CO, 1989.
33. M. Oades, A.M. Vassallo, A.G. Waters, M.A. Wilson, 'Characterization of Organic Matter in Particle Size and Density Fractions from a Red Brown Earth by Solid-state ^{13}C NMR', *Aust. J. Soil Res.*, **25**, 71–82 (1987).
34. M.W. Schmidt, H. Knicker, P.G. Hatcher, I. Kogel-Knabner, 'Improvement of ^{13}C and ^{15}N CPMAS NMR Spectra of Bulk Soils, Particle Size Fraction and Organic Material by Treatment with 10% Hydrofluoric Acid', *Eur. J. Soil Sci.*, **48**, 319–328 (1997).
35. F.J.G. Vila, H. Lentz, H.D. Ludemann, 'Fourier Transform ^{13}C Nuclear Magnetic Resonance Spectra of Natural Humic Substances', *Biochem. Biophys. Res. Commun.*, **72**, 1063–1070 (1976).
36. R.H. Newman, K.R. Tate, 'Phosphorus Characterization by ^{31}P Phosphorus Isotope Nuclear Magnetic Resonance', *Commun. Soil Sci. Plant Anal.*, **11**, 835–842 (1980).
37. R.H. Newman, K.R. Tate, P.F. Baron, M.A. Wilson, 'Towards a Direct, Non-destructive Method of Characterizing Soil Humic Substances Using ^{13}C Carbon Isotope NMR', *J. Soil Sci.*, **31**, 623–631 (1980).

38. L. Benzing-Purdie, J.A. Ripmeester, C.M. Preston, 'Elucidation of the Nitrogen Forms in Melanoidins and Humic Acid by Nitrogen-15 Cross Polarization–Magic Angle Spinning Nuclear Magnetic Resonance Spectroscopy', *J. Agric. Food Chem.*, **31**, 913–915 (1983).
39. J.L. Hedges, P.G. Hatcher, J.R. Ertel, K.J. Meyers-Schulte, 'A Comparison of Dissolved Humic Substances from Seawater with Amazon River Counterparts by ^{13}C-NMR Spectroscopy', *Geochim. Cosmochim. Acta*, **56**, 1753–1757 (1992).
40. E.D. Ingall, P.A. Schroeder, R.A. Berner, 'The Nature of Organic Phosphorus in Marine Sediments: New Insights from ^{31}P NMR', *Geochim. Cosmochim. Acta*, **54**, 2617–2620 (1990).
41. M. Krosshavn, T.E. Southon, E. Steinnes, 'The Influence of Vegetational Origin and Degree of Humification of Organic Acids on Their Chemical Composition Determined by Solid-state ^{13}C NMR', *J. Soil Sci.*, **43**, 485–493 (1992).
42. R. Benner, P.G. Hatcher, J.L. Hedges, 'Early Diagenesis of Mangrove Leaves in a Tropical Estuary: Bulk Chemical Characterization Using Solid-state ^{13}C NMR and Elemental Analysis', *Geochim. Cosmochim. Acta*, **54**, 2003–2013 (1990).
43. M. Krosshavn, I. Kogel-Knabner, T.E. Southon, E. Steinnes, 'The Influence of Humus Fractionation on the Chemical Composition of Soil Organic Matter Studied by Solid-state ^{13}C NMR', *J. Soil Sci.*, **43**, 473–483 (1992).
44. G. Guggenberger, B.T. Christensen, G. Rubek, W. Zech, 'Land-use and Fertilization Effects on P Forms in Two European Soils: Resin Extraction and ^{31}P NMR Analysis', *Eur. J. Soil Sci.*, **47**, 605–614 (1996).
45. D.W. Hopkins, J.A. Chudek, E.A. Webster, D. Barraclough, 'Following the Decomposition of Ryegrass Labelled with ^{13}C and ^{15}N in Soil by Solid-state Nuclear Magnetic Resonance Spectroscopy', *Eur. J. Soil Sci.*, **48**, 623–631 (1997).
46. H. Knicker, R. Frund, H. Ludemann, 'Characterization of Nitrogen in Plant Composts and Native Humic Material by Natural-abundance ^{15}N CPMAS and Solution NMR Spectra', in *Nuclear Magnetic Resonance Spectroscopy in Environmental Chemistry*, eds. M.A. Nanny, R.A. Minear, J.A. Leenheer, Oxford University Press, New York, 272–294, 1997.
47. S.J. Anderson, 'Proton and ^{19}F NMR Spectroscopy of Pesticide Intermolecular Interactions', in *Nuclear Magnetic Resonance Spectroscopy in Environmental Chemistry*, eds. M.A. Nanny, R.A. Minear, J.A. Leenheer, Oxford University Press, New York, 51–72, 1997.
48. J.M. Bortiatynski, P.G. Hatcher, R.D. Minard, 'The Development of ^{13}C Labeling and ^{13}C NMR Spectroscopy Techniques to Study the Interaction of Pollutants with Humic Substances', in *Nuclear Magnetic Resonance Spectroscopy in Environmental Chemistry*, eds. M.A. Nanny, R.A. Minear, J.A. Leenheer, Oxford University Press, New York, 26–50, 1997.
49. B.E. Herbert, P.M. Bertsch, 'A ^{19}F and ^{2}H NMR Spectroscopic Investigation of the Interaction Between Nonionic Contaminants and Dissolved Humic Material', in *Nuclear Magnetic Resonance Spectroscopy in Environmental Chemistry*, eds. M.A. Nanny, R.A. Minear, J.A. Leenheer, Oxford University Press, New York, 73–90, 1997.

Optical Emission Inductively Coupled Plasma in Environmental Analysis

Mark E. Tatro
Spectra Spectroscopy & Chromatography Specialists, Inc., Warwick, NY, USA

Dulasiri Amarasiriwardena
School of Natural Science, Hampshire College, Amherst, MA, USA

1 Introduction **1**
2 Inductively Coupled Plasma Designs **2**
3 Front-end Improvements in ICP-AES **3**
3.1 Solution Introduction via Nebulizer 3
3.2 Hydride Generation (HG) 3
3.3 Chemical Speciation with ICP-AES 4
3.4 Flow Injection Analysis (FIA) in ICP-AES 4
4 Environmental Sample Preparation for Inductively Coupled Plasma/Atomic Emission Spectroscopy Analysis **4**
4.1 Sources of Information 4
4.2 Hot Plate Digestions 5
4.3 Microwave Digestions 6
5 Environmental Inductively Coupled Plasma/Atomic Emission Spectroscopy Methods of Analysis **6**
5.1 Sources of Methods 6
5.2 Recommended Inductively Coupled Plasma-atomic Emission Spectroscopy Emission Lines 6
5.3 Compensating for Interferences 7
5.4 Quality Control Protocols 7
Acknowledgments **9**
Abbreviations and Acronyms **9**
Related Articles **10**
References **10**

The ability of the inductively coupled plasma-atomic (or optical) emission spectroscopy (ICP-AES (or OES))) system to perform multielement trace metal analysis of environmental samples provided commercial laboratories with the needed incentive to enter into the business of efficient trace metal analysis. Previously, trace metals were typically analyzed using colorimetric techniques, which were both cumbersome and subject to interferences, or flame atomic absorption techniques, which, although almost interference free, were labor intensive owing to their one-element-at-a-time analytical mode. Even the furnace atomic absorption technique, for years the standard bearer of low-level trace metal analysis, is giving way to axial and radial viewed ICP-AES techniques. Now, ICP-AES has become an affordable and well-established multielement analytical method.

Taking the USEPA (United States Environmental Protection Agency) SW-846 solid waste analysis manual as an example, this article reviews the methods used for the preparation of samples and the ICP-AES analysis of the prepared samples for trace metals in environmental matrices. This review includes recent developments in front-end improvements in ICP-AES and a detailed overview of the quality control (QC) requirements for environmental ICP-AES analysis.

1 INTRODUCTION

Philosophers tell us that when we look back at our life's journey, the events that we viewed as a series of random unconnected acts when we were living them became instead a path that has cohesion and rationality. Such has been the history of ICP-AES for environmental analysis. These authors remember the pre-ICP (inductively coupled plasma) days when to perform multielement analysis flame atomic absorption spectrometers were relied upon, which required the use of a different hollow cathode lamp and in some cases a different burner head for the analysis of each element. In the ICP-AES approach, the energy to ICP torch is induced by radio frequency (RF) energy through the three-ring, water-cooled load coil surrounding the three concentric quartz tube torches. Plasma is initiated by introducing a highly energetic electron surge by a Tesla discharge or piezoelectric solid-state transducer. The resulting ions and electrons interact with the oscillating magnetic field produced by the load coil, promoting further ionization of argon gas atoms via collisional interactions, thereby creating sustainable plasma rich with Ar^0, Ar^+, and free electrons. The electron temperature in the plasma can reach 8000–10 000 K. Thus, the ICP facilitates sufficient energy to the aerosolized sample stream to vaporize ($MX_{(g)} \rightarrow M_{(g)} + X_{(g)}$) and to atomize analyte elements. In the hot zone of the plasma, electrons in analyte atoms are promoted to higher energy levels (excitation), and, at appropriate plasma conditions, some of these atoms are ionized to form ions ($M^+_{(g)}$). Excited analyte atoms later

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

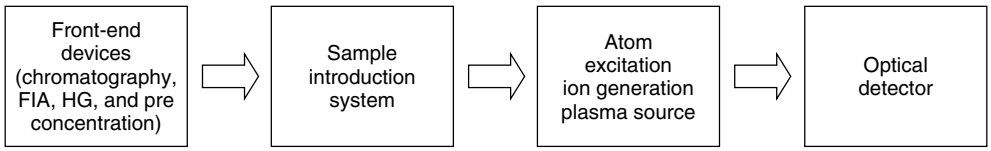

Figure 1 Block diagram for inductively coupled plasma atomic emission spectroscopy (ICP-AES) system.

reach the ground state, emitting characteristic energy at appropriate wavelengths. These wavelengths are unique to each analyte atom, and they can be used to identify unknown elements, while intensities of these emission signals are directly related to the concentration of analyte atoms in the test sample. Argon-based ICP is particularly suitable for exciting a variety of metal atoms and thus facilitates simultaneous multielement capability (Figure 1). By means of an appropriate suite of elemental standards, the instrument can be calibrated. By employing the linear calibration functions obtained for each analyte, one can measure the concentration of these elements in the sample.[1,2,3]

Environmental applications were the driving force in the development of ICP instrumentation. In 1975, under contract to the USEPA, Winge et al.[4] of Ames laboratory reported on the development of ICP-AES for the simultaneous determination of trace element pollutants in water. The USEPA established its first official ICP-AES method for wastewater in 1982 as Method 200.7.[5]

In the 1980s, the decade of ICP, the USEPA approved ICP-AES for the following programs: National Pollutant Discharge Elimination System (NPDES); Safe Drinking Water Act (SDWA); Contract Laboratory Program (CLP); and Resource Conservation and Recovery Act (RCRA). An overview of the USEPA-approved methods for ICP/AES can be found in the *Code of Federal Regulations* (40 CFR, Parts 136 and 141) and in the *Guide to Environmental Analytical Methods*.[6]

This article expands on the USEPA programs listed above, their effect on the growth of ICP-AES for environmental analysis, and recent ''front-end'' developments of several sample introduction methods for ICP-AES.

2 INDUCTIVELY COUPLED PLASMA DESIGNS

ICP-AES gained a strong foothold in environmental laboratories in the 1980s and 1990s due, in part, to the advantages over its predecessor flame atomic absorption spectroscopy listed here:

- multielement analysis capability
- large dynamic linear range
- reduction of matrix interferences
- good long-term stability: <1% (relative standard deviation) RSD with good instruments[2]
- improved detection limits for refractory elements and
- enhancement of productivity.

The early ICP-AES designs of the 1980s that were approved for environmental analysis were of two types: simultaneous and sequential spectrometers with radial viewed torches.[7] The use of the radial viewed plasma results in a shortened path length, which puts the detection limit above the detection limits required by USEPA for elements such as arsenic, lead, selenium, and thallium. For these elements, the USEPA usually required the use of a heated graphite atomizer (HGA).[8]

Later, ICP-AES designs of the 1990s approved for environmental analyses were axial viewed simultaneous and sequential spectrometers.[9] The use of the axial view of the plasma results in a longer path length that greatly improved detection limits and nearly eliminated the need for the use of the HGA to meet the USEPA-detection requirements.[10] With the advent of inductively coupled plasma mass spectrometry (ICP-MS) and by taking the benefit of its high sensitivity and low elemental detection limits, Method 200.8 was first introduced by USEPA in 1990 for the analysis of water.[11]

Recent commercial ICP-AES instruments are not bulky and do not require expensive laboratory space to house them; and they are often bench top versions. In addition, the hardware is more robust, the instrument operational software is user friendly, and the unit price of decent ICP-AES instruments are now more affordable for commercial and research laboratories. During the past 10 years, the major analytical advances in ICP-AES analytical technology have been somewhat slow except for the introduction of solid-state charge transfer detectors and the use of microsample introduction systems including micronebulizer. Solid-state charge transfer detectors (CTD) such as charged-coupled devices (CCDs) and charged-injection devices (CIDs) are common nowadays in commercial instruments. These detectors offer simultaneous multielement capability and have lowered the detection limits obtained by ICP-AES, but further work should be done to minimize the analytical limits imposed by the analytical response at the UV range and issues related to the dynamic range and limited pixilation.[12] These solid-state charge transfer

detectors (CTD) were developed in parallel with the recent development in ICP-mass spectrometry (MS)[2,13] and reviewed by Mermet.[13] Plasma robustness can be evaluated by measuring the signal intensity of Mg(I) (at 285.213 nm) and Mg(II) (at 280.270 nm).[14] Finally, ICP-AES spectrometers that could be programmed for both radial and axial view became the state of the art where the less sensitive radial view was preferred for the higher concentration elements such as aluminum, calcium, iron, magnesium, potassium, and sodium in environmental samples, while the more sensitive axial view was preferred for the lower concentration elements such as the transition series.[15]

3 FRONT-END IMPROVEMENTS IN ICP-AES

The efficient introduction of liquids that offer maximum analytical information is always a challenge for popular pneumatic nebulizer-based ICP-AES techniques. As a result, a series of so-called front-end improvements have taken place during the past two decades (see Figure 1). Some of these improvements were developed in parallel with similar developments in ICP-MS, and some were borrowed from well-established atomic absorption spectrophotometry (AAS). These alternatives include the improvement of nebulizers that can deliver liquids efficiently, introduction of analytes as a gas-phase (i.e. hydride generation (HG) and electrothermal vapor introduction) that will eliminate the introduction of difficult sample matrices, flow injection analysis (FIA) that can introduce small sample volumes that increase the analytical throughput via online analysis, and the separation of elemental species (i.e. As(III) and As(V)) before introduction to the ICP discharge. The former approach has led to the new and exciting field of *chemical speciation* that offered vital analytical information of environmental and clinical interest. For example, As(III) is more toxic than As(V),[16] and the ratio of these species' presence in a given environmental medium is of great importance for an environmental chemist. Resolution of these chemical species requires powerful separation techniques, for example, high performance liquid chromatography (HPLC), gas chromatography (GC), capillary electrophoresis (CE), or size-based separation methods like field-flow fractionation (FFF). These chromatographic or separation methods offer unique chemical and physical separation capabilities before introduction of the separated chemical species to the ICP-based elemental detector. The analytical developments in the chemical speciation area have been noteworthy with ICP-AES, and these advances are extended particularly with ICP-MS. This is partly because the mass spectrometer (MS) is a more sensitive detector than the optical emission detector (OES or aka AES). Some of these so-called front-end improvements are discussed and critiqued at length in the recent literature by Linge,[12] Mermet,[13] Becker,[17] Hansen, and Miró.[18]

3.1 Solution Introduction via Nebulizer

Although the customary method of sample introduction is a pneumatic nebulizer ($\sim$1 mL min^{-1}) that converts the sample liquid into appropriate-sized aerosol droplets that can be introduced into the plasma source for excitation, this approach has been limited by poor sample-/analyte-transport efficiency.[17] The alternative to this technique is to introduce small sample volumes using a microflow nebulizer, and a spray-chamber configuration that provides improved sample transport to the plasma and reduced sample consumption.[2,17,19] A direct-injection, high-efficiency nebulizer, which is widely known as *DIHEN*, can reduce the sample consumption rate down to 30 μL min^{-1}, and it was successfully applied to minimize sample matrix effects in both axially and radially viewed ICP-AES.[20] An attempt has been also made to incorporate a sample introduction system to the ICP excitation source using a microflow nebulizer that is easily snapped to the base of the ICP torch assembly. This torch-integrated sample introduction configuration (*aka* TISIS) has minimized memory effects, and it has been found that matrix effects and nonspectroscopic interferences are minimal.[21] The elemental sensitivity of ICP-AES and MS had enhanced two to five times by the use of ultrasonic nebulizer (USN), in which the sample is nebulized with the aid of a piezoelectric transducer driven by a high frequency generator.[2] Clearly, these advancements are a welcome addition to atomic spectroscopists and can be advantageous in analytical applications where the sample size is limited and improved detection limits are needed. These devices are now available commercially through several manufactures that specialize in front-end ICP spectrometry accessories.

3.2 Hydride Generation (HG)

The premise of the HG method is to produce a gas-phase elemental hydride by a reduction reaction,[22] and to introduce the resulting gaseous hydride (MH_n) directly to the atom excitation/ionization source (i.e. ICP, flame, and furnace). These reducing reactions are rapid, and are usually achieved by the reaction of the sample with an alkaline solution of sodium tetraborate ($NaBH_4$) in an acidic medium (the reducing agent that generates nascent hydrogen). The ensuing hydride ($M^{n+} + (n+2)\,H^+ \rightarrow MH_n + H_2$) formed is flushed with a stream of an inert striping gas like argon. HG reactions

can be achieved on line or by flow-injection modes; therefore, it is convenient for automation. The HG method has been successfully applied for metalloid-based hydrides (AsH_3, SbH_3, and BiH_3) and other elemental hydrides, for example, Ge, Pb, Se, Te, and Sn.[23,24] More recently, the scope of this approach has extended to transition metals (Zn) and noble metals such as Ru and Rh.[25,26] The HG method has enhanced analyte-transport efficiency (from 1 to 100%) and has eliminated most of the matrix-based interferences. As a result, the analytical sensitivity of ICP-AES has improved and lower detection limits have been achieved. Furthermore, the sample is introduced as a gas phase, and the input power to the argon plasma is minimized in contrast with liquid-based sample introduction. Selective reduction and subsequent measurement of inorganic arsenic species (As(III), As(V)), monomethylarsonic acid (MMA), and dimethylarsinic acid (DMA) using continuous HG mode has been demonstrated with ICP-AES.[23,27]

3.3 Chemical Speciation with ICP-AES

Although ICP-AES is not the most sensitive detector for elemental speciation, it has been used effectively in several instances where the sensitivity of ICP-AES is sufficient enough to achieve the detection capability of interested chemical species. The element-based chemical speciation field is dominated by ICP-MS as a sensitive detector for many speciation applications. To accomplish chemical speciation, often chromatography-based separation techniques such as HPLC or GC are required. Once separated, the ions (e.g. different oxidation states of Sb(III) and Sb(V)) or metal-bearing chemical species MMA or diethyl mercury – $(C_2H_5)_2Hg$) can be directly transported into the ICP-AES nebulizer as a sample. The resulting chromatogram (i.e. emission intensity at a selected wavelength vs time) indicates the retention time of the chemical species–associated metals as a peak. The peak area or height can provide quantitative information about these species if an appropriate calibration is made prior to sample analysis. The mobile phase flow rates of HPLC-based separations ($1\,mL\,min^{-1}$) are suitable for interfacing the chromatograph eluent with the ICP-AES detector, thus providing the capability of automation of metal speciation. HPLC-ICP-AES techniques have been successfully applied for selenoamino acids speciation[28] and the study of cobalt–cysteine interactions,[29] while GC-MS-ICP-AES is reported by Kato et al., (1992) using an axial view of ICP-AES.[30] Size-based separation of metals associated with air particulates in PM_{10} fractions were determined by flow field-flow fractionation (flow-FFF) coupled with ICP-AES.[31] The authors used sedimentation (Sd) FFF-ICP-AES and applied USN to overcome the poor sensitivity of ICP-AES and small sample mass and dilutions resulting from the Sd-FFF separation process. Later, Saeseaw et al. used FFF-ICP-AES to distinguish Ca^{2+}- and Zn^{2+}-induced aggregation of β-lactoglobulin aggregates.[32] Application of hyphenated ion chromatography (IC),[33] GC,[34] and CE separation methods to ICP-AES and MS detectors are discussed in the literature.[33,12]

3.4 Flow Injection Analysis (FIA) in ICP-AES

A small, discrete sample is injected to the flowing carrier solution where the dispersion of the sample into the carrier is controlled by the convection current owing to the carrier stream flow and diffusion owing to the concentration difference between the sample and the carrier stream. The introduced sample can be subjected to on-line reagents to carry out predetermined chemistries, or separation or preconcentration of analytes, to prepare for subsequent detection analytes of interest downstream. In this particular case, the detector is a sensitive multielement detector like ICP-AES or ICP-MS. The resulting FIA curve (or ''fiagram'') is a profile of the intensity signal against time. The width of this profile is influenced by dispersion of sample resulting from convection and diffusion forces between the sample and carrier liquids. The recent advancements in FIA analysis are reviewed by Hansen and Miro.[18] FIA has been used effectively to incorporate analysis of hydride-forming elements such as As, Sb, Bi, Se, Te, and Ge,[18] for isolation and preconcentration of aluminum species,[35] for on-line standard addition and dilution of trace metals in plant materials,[36] for speciation of Cr(III) and Cr(VI) species,[37] and for on-line multielement analyses of alkaline salts by ICP-AES.[38] The FIA method facilitates rapid throughput of samples and requires a small sample size, therefore it is an attractive method in terms of time and cost of analysis. FIA may suffer from degraded sensitivity. Nevertheless, FIA methods are as good as most traditional analytical methods as far as accuracy and precision are concerned and the selectivity is often better.[2]

4 ENVIRONMENTAL SAMPLE PREPARATION FOR INDUCTIVELY COUPLED PLASMA/ATOMIC EMISSION SPECTROSCOPY ANALYSIS

4.1 Sources of Information

Sample preparation methods for ICP analysis of trace metals can be found in the following sources:

- The USEPA's Office of Solid Waste prepares and publishes the SW-846 manuals for the analysis of

groundwater, extracts, industrial wastes, soils, sludges, sediments, and solid wastes as dictated by the RCRA.[39]

- The American Public Health Association, the American Water Works Association, and the Water Environment Federation jointly prepare and publish the *Standard Methods* manual for the analysis of drinking water and wastewater.[40]
- The USEPA's Office of Research and Development prepares and publishes its *Methods for the Determination of Metals in Environmental Samples* for the analysis of drinking water, surface water, saline, and industrial and domestic wastes.[41]
- The USEPA published its *Methods for the Determination of Metals in Environmental Samples* Supplement I that updates sample preparation methods.[42]
- The USEPA's Superfund program prepares and publishes its *CLP* manual for the analysis of wastewater and solid waste.[43]

The following method requirements are taken from the SW-846 manuals to provide the reader with an example of what sample preparation for ICP analysis entails. A compilation of approved methods of ICP sample preparation is presented in Table 1.

The sample introduction for ICP environmental applications requires the samples to be in a liquid form so that they can be aspirated through a nebulizer and that suspended solids be removed to prevent clogging of the nebulizer. Therefore, solid samples such as soils and sludges must first be acid digested to extract the metals of interest from the solids, and the resulting digestate must be filtered, centrifuged, or allowed to sit to remove the suspended solids prior to analysis. Drinking water with turbidity greater than 1 (nephelometric turbidity unit) NTU must also be acid digested followed by the removal of suspended solids prior to analysis.

4.2 Hot Plate Digestions

The acid digestion of waters and solids, employing nitric and hydrochloric acids, can be effected by either hot plate (or hot block) methods or microwave methods. The hot plate methods for the preparation of waters can be quite confusing as they are subcategorized as being either ''total'' or ''total recoverable'' – or referred to by some as ''hard'' and ''soft'' respectively – where the ''total'' digestion is more rigorous than the ''total recoverable'' method. An example of a hot plate ''total'' digestion

Table 1 Digestion methods for environmental ICP/AES analysis

Method	Matrix	Reagents	Digestion type	References
3005A	Ground water/surface water	HNO_3/HCl	Total recoverable	16
3010A	Waste water (also applicable to extraction procedure (EP) and mobility-procedure extract)	HNO_3/HCl	Total	16
3015A	Water, mobility-procedure extracts, and wastes that contain suspended solids	HNO_3	Microwave assisted in Teflon vessels	39
3031	Oils and petroleum sludges	Various (HNO_3), H_2SO_4, HCl, $KMnO_4$	Total	39
3040A	Oils	Dissolved and diluted in organic solvent	Dissolution	39
3050B	Soil	HNO_3/H_2O_2/HCl	Total recoverable	39
3050B	Soil	HNO_3/H_2O_2/HCl	Microwave-assisted	39
3051A	Soil, sludges, sediments	HNO_3/HCl	Microwave-assisted	39
3052	Silicates, oil contaminated soils, organic ash, biological tissue	HNO_3/H_2O_2/HCl/HF	Microwave-assisted Total (use fluoro carbon digestion vessel)	39
3030E	Water	HNO_3	Total	40
3030F[a]	Water	HNO_3/HCl	Total	40
3030F[b]	Water	HNO_3/HCl	Total recoverable	40
4.1.3	Water	HNO_3/HCl	Total	41
4.1.4	Water	HNO_3/HCl	Total recoverable	41
200.2	Water/soil	HNO_3/HCl	Total recoverable	42
ILMO4.0	Water	HNO_3/HCl	Total recoverable	43
ILMO4.0	Soil	HNO_3/H_2O_2/HCl	Total recoverable	43
200.7	Water/solids, and bio solids	HNO_3/HCl	Total recoverable	44
200.15	Water	HNO_3/HCl	Total recoverable	45

[a] Total HNO_3/HCl procedure.
[b] Recoverable HNO_3/HCl procedure.

for wastewater is the SW-846 method 3010A and an example of a hot plate ''total recoverable'' method for ground water, surface water, and wastewater is the SW-846 method 3005A. An example of a hot plate digestion for solids, including sludge and soils, is SW-846 3050B.

4.3 Microwave Digestions

Microwave digestion methods using either closed-vessel or open-vessel microwave digestion systems are approved by the USEPA. Examples of closed-vessel microwave methods for wastewaters and solids are the SW-846 methods 3015A and 3051A respectively. An example of an open-vessel microwave method is the SW-846 method 3050B. Finally, the SW-846 method 3052, a closed-vessel microwave method, which employs hydrofluoric acid for the preparation of siliceous and organic-based matrices as well as biological tissue, oils, sediment, soil, and sludges, is the only truly ''total'' digestion approved by the USEPA in which all suspended solids are solubilized.

5 ENVIRONMENTAL INDUCTIVELY COUPLED PLASMA/ATOMIC EMISSION SPECTROSCOPY METHODS OF ANALYSIS

5.1 Sources of Methods

The most used ICP methods of environmental compliant analysis are the USEPA's Office of Research and Development method 200.7 revision 4.4[44] and the USEPA's Office of Solid Waste SW-846 method 6010B.[16] Other environmental ICP methods of analysis that are similar to method 6010B include the standard method 3120B,[40] the CLP,[43] and the air force center for engineering and the environment (AFCEE) at http://www.afcee.af.mil/.

The USEPA also approves the use of its method 200.15 for drinking water analysis.[45] This method employs an USN on the front end of an ICP-AES system to provide better detection limits required for drinking water analysis. It is important to note that the USEPA has not yet approved ICP-AES, with or without ultrasonic nebulization, for the analysis of Pb, Sb, Se, and Tl in drinking water. Effective from January 2006, drinking water supply operators have to comply with the new public health goal (PHG) for arsenic in drinking water promulgated by USEPA in 2002. The new maximum contaminant level (MCL) for As in drinking water is 10 $\mu g\ L^{-1}$.[46] This MCL is routinely difficult to achieve by ICP-AES and consequently the EPA has withdrawn the ICP-AES methods 200.7 and the standard method 3120B for the determination of arsenic in drinking water. Now, one has to use either method 200.9 graphite furnaces (GF)-AAS or method 200.8, which is based on ICP-MS.[46]

The SW-846 method 6010B is used here to provide an example of a typical environmental ICP method. For a comparison of various environmental ICP methods, see Wagner.[6]

5.2 Recommended Inductively Coupled Plasma-atomic Emission Spectroscopy Emission Lines

Table 2, taken from method 6010B, documents the suggested ICP emission line of choice and the estimated instrument detection limit (IDL) for radial viewed plasma. Some programs such as the CLP publish contract-required detection limits (CRDLs) that must be met by the instrument of choice (see Table 3). The CRDLs for Pb, Se, and Tl required by the CLP's Statement of Work (SOW) Inorganic Laboratory Manual Operations (ILMO) 4.0 are 3, 5, and 10 respectively. A radial viewed ICP does not meet these required detection limits. However, an axial viewed ICP meets these

Table 2 Recommended wavelengths and estimated instrumental detection limits (IDLs) for radial view ICP-AES[43]

Element	Wavelength (nm)	Estimated IDL ($\mu g\ L^{-1}$)
Aluminum	308.215	30
Antimony	206.833	21
Arsenic	193.696	35
Barium	455.403	0.87
Beryllium	313.042	0.18
Boron	249.678 × 2	3.8
Cadmium	226.502	2.3
Calcium	317.933	6.7
Chromium	267.716	4.7
Cobalt	228.616	4.7
Copper	324.754	3.6
Iron	259.940	4.1
Lead	220.353	28
Lithium	670.784	2.8
Magnesium	279.079	20
Manganese	257.610	0.93
Mercury	194.227 × 2	17
Molybdenum	202.030	5.3
Nickel	231.604 × 2	10
Phosphorus	213.618	51
Potassium	766.491	– [a]
Selenium	196.026	50
Silicon	251.611	17
Silver	328.068	4.7
Sodium	588.995	19
Strontium	407.771	0.28
Thallium	190.864	27
Tin	189.980 × 2	17
Titanium	334.941	5.0
Vanadium	292.402	5.0
Zinc	213.856 × 2	1.2

Reproduced from Ref. 43.
[a] Detection limit is subject to plasma conditions.

Table 3 The CLP inorganic target analyte list[20]

Analyte	CRDL ($\mu g\,L^{-1}$)	Analyte	CRDL ($\mu g\,L^{-1}$)
Aluminum	200	Magnesium	5000
Antimony	60	Manganese	15
Arsenic	10	Mercury	0.2
Barium	200	Nickel	40
Beryllium	5	Potassium	5000
Cadmium	5	Selenium	5
Calcium	5000	Silver	10
Chromium	10	Sodium	5000
Cobalt	50	Thallium	10
Copper	25	Vanadium	50
Iron	100	Zinc	10
Lead	3		

Reproduced from Ref. 20.

required detection limits and, therefore, has become the instrument of choice for environmental analysis.

5.3 Compensating for Interferences

ICP emission suffers from spectral interferences, and the analyst must document that these interferences have been eliminated. The two most common interferences experienced in environmental ICP analysis are (i) the baseline shift caused mainly by high concentrations of Ca and Mg in samples and due to stray light and (ii) direct spectral overlap due to an emission line from an element in the sample that falls at or near the analyte wavelength and causes a false analyte concentration. Figure 2 shows the effect of both types of interference on the Sb 206.833-nm emission line for an axial view ICP. The calibration standard of 1000-ppb Sb (——) and the calibration blank (- - - -) are superimposed on the graphic that also displays the emission signal for an interference check standard (ICS[A]) with major elements only (-·-·-·-·), which contains 500-ppm Al, Ca, and Mg, and 200-ppm Fe as well as the emission signal for a standard that contains 10-ppm Cr, Mn, and V (· · ·).

If left unchecked, the ICS [A] would give a false positive reading of approximately 800-ppb Sb. To prevent this, a background-correction point is programmed slightly to the left of the Sb calibration standard peak (L). During analysis, the emission at the background-correction point wavelength is subtracted from the emission at the 206.833-nm peak wavelength, which effectively reduces the reading of the ICS[A] to less than detectable.

Referring again to Figure 2, the background-correction point does not reduce the reading from the 10-ppm Cr, Mn, V standard since there is a false peak for Sb arising from one of the three elements in this standard (in this case, the interference is due to Cr). To correct for this direct spectral overlap, an inter element correction (IEC) is required and must be documented.

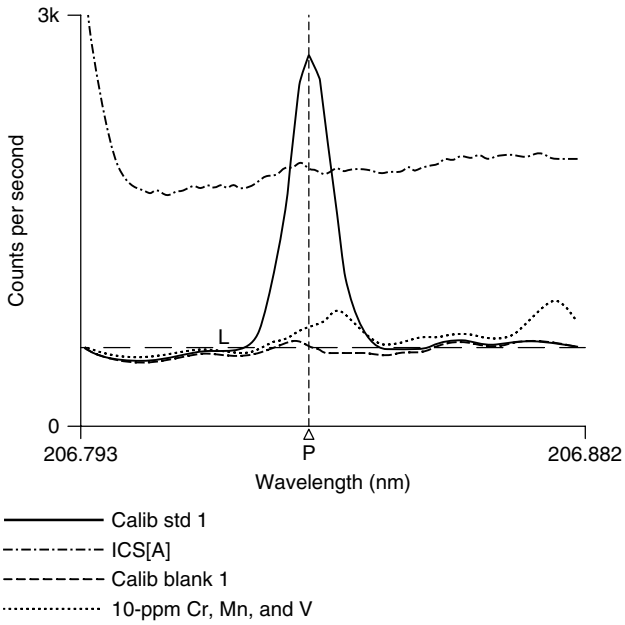

Figure 2 ICP/AES spectral overlays of the antimony 206.833-nm emission line (——) and a blank (- - - -) in the presence of an ICS[A] (-·-·-·-·) and a mixed standard of chromium, manganese, and vanadium (· · · ·). The overlays are used to set the background-correction position and to verify if inter element correction factors (IECs) are required.

Other interferences normally encountered in environmental ICP analysis are physical or matrix interferences caused by differences in viscosity or the amount of dissolved solids in samples compared to standards. To reduce matrix interferences, analysts are required to prepare all standards and QC samples in the same acid matrix that exists for the digested samples. As a further control, analysts should use an internal standard such as Sc or Y to compensate for physical interferences.

5.4 Quality Control Protocols

After optimizing the instrument according to the manufacturer's specifications, the analyst is required to analyze a number of QC samples throughout the run where there are decisions to be made on the basis of a window of acceptance for each QC sample analyzed.

The concept of the sample batch is used to describe the number of samples (typically a maximum of 20 samples) that are digested prior to ICP analysis and that are accompanied by QC samples that are also digested. The failure of the digested QC samples (see below) to meet acceptance criteria requires that the samples be redigested.

The complete list of the QC samples as per SW-846 method 6010B is given in the following sections.

5.4.1 Initial Calibration

For SW-846 method 6010B, calibration of the ICP-AES requires one calibration standard within the linear range and one calibration blank. The correlation coefficient of linear calibration curve should be at least 0.995 or better.

5.4.2 Initial Calibration Verification Standard

This standard is prepared from a source different from the source used for the calibration of the ICP. The concentrations of the elements should be at a level that is 50% of the concentrations in the calibration standard. The initial calibration verification (ICV) standard is analyzed after initial calibration of the ICP to verify the calibration curve and must fall within an error of ±10% of the true value. The same analytical wavelength used for calibration of the instrument must be used for sample analyses. Any analyte(s) that fails must be reanalyzed.

5.4.3 Continuing Calibration Verification Standard

This is the same as the ICV but the analysis is after every 10 samples and at the end of the run to monitor the drift. If the continuing calibration verification (CCV) standard falls outside the error window of ±10%, then the 10 samples preceding the failed CCV must be reanalyzed for the failed analyte(s). The concentration of this standard should be closer to the midrange of the calibration curve.

5.4.4 Initial Calibration Blank

This contains the matrix acids and is analyzed after the ICV to monitor contamination and memory effects. If the (initial calibration blank) ICB is greater than three times the IDL, the instrument must be recalibrated for the failed analyte(s).

5.4.5 Continuing Calibration Blank

This is the same as the ICB but the analysis is after each CCV. If the continuing calibration blank (CCB) is greater than three times the IDL, then the 10 previous samples must be reanalyzed for the failed analyte(s). However, if any of the 10 previous samples are greater than 10 times the failed CCB, then these sample(s) do not have to be reanalyzed.

5.4.6 Method Blank

This blank is carried through the entire sample preparation process. Although method 6010B does not provide a definitive QC window, laboratories usually assign the same criteria for the method blank (MB) as for the CCB. In other words, if the MB is greater than three times the IDL, then the samples must be redigested for the failed analyte(s) except for those samples that are more than 10 times the failed MB.

5.4.7 Check Standard

A known standard solution is prepared by an external laboratory.

5.4.8 Interference Check Standard

This standard contains 500-ppm Al, Ca, Mg, 200-ppm Fe, and either 500 or 1000 ppb of all the analytes. The normal nomenclature used for this standard refers to the solution containing the major elements only (Al, Ca, Fe, and Mg) as ICS[A]; the solution containing the 500 or 1000-ppb analytes only as ICS[B]; and the combination of the two solutions as ICS[AB]. The ICS[AB] is used to verify that the background-correction points and the IEC values are programmed correctly. Any analyte(s) that falls outside the error window of ±20% must be reanalyzed.

5.4.9 Matrix Spike/Matrix Spike Duplicate

One duplicate sample is spiked with a known concentration of analytes and is carried through the entire sample preparation process. A separate MS/MSD (matrix spike/matrix spike duplicate) must be prepared for waters, soils, and extracts. The relative percent difference (RPD) of the matrix spike duplicate (MSD) must fall within ±20% error and the spike recovery must be within ±25% error.

5.4.10 Additional Quality Control

In the event that the matrix spike fails to fall within the allowed 75–125% recovery limits, the analyst must perform (i) a dilution test and (ii) a postdigestion spike addition.

1. Dilution test – a 1:5 dilution of the sample is prepared only if the analyte concentration is greater than 10 times the IDL after dilution. The diluted concentration must agree with the undiluted concentration within ±10%.
2. Postdigestion spike addition – a spike containing the analytes at concentration between 10 and 100 times the IDL is added to the digested sample. The spike recovery must be within 75 – 125%.

5.4.10 Preanalysis Quality Control

Prior to any analysis, the following data must be gathered for each ICP-AES emission line:

Table 4 SW-846 method 6010B ICP/AES run log

Sample	QC error limits
Calibration blank	
Calibration standard	
ICV standard	±10%
ICV blank	<3 × IDL
ICS[A]	±20%
MB	<3 × IDL
Sample 1	
Sample 1 + MS	±25%
Sample 1 + MSD	±20% RPD
Sample 2	
Sample 3	
Sample 4	
Sample 5	
Sample 6	
CCV standard	±10%
CCV blank	<3 × IDL
Sample 7	
Sample 8	
Sample 9	
Sample 10	
CCV standard	±10%
CCV blank	<3 × IDL

- linear dynamic range;
- method detection limits;
- IDLs (instrument detection limits, the concentration equal to an analytical signal resulting from three times the blank signal = $3^{*}\sigma_{b}$, measure 7 [n] replicates);
- IEC factor determined by aspirating a solution with major concomitant ions such as Ca, Mg, Al, and Fe; and
- background-correction points.

A typical run log for ICP-AES analysis that follows the SW-846 method 6010B is shown in Table 4.

ACKNOWLEDGMENTS

The author would like to thank Oliver Fordham, Jr. of the USEPA Office of Solid Waste and Emergency Response for his input on the historical aspects of the USEPA methods of ICP-AES analysis, and Kathy Zinn of Progress Environmental Laboratory and Dr. Ramon Barnes of University Research Institute for Analytical Chemistry, Amherst, MA for their review and critique of the various parts of this manuscript.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrophotometry
AES	Atomic Emission Spectroscopy
AFCEE	Air Force Center for Environmental Excellence
CCB	Continuing Calibration Blank
CCDs	Charged-coupled Devices
CCV	Continuing Calibration Verification
CE	Capillary Electrophoresis
CFR	Code of Federal Regulations
CIDs	Charged-Injection Devices
CLP	Contract Laboratory Program
CRDL	Contract-required Detection Limit
CTD	Charge Transfer Detectors
DIHEN	Direct-injection High-efficiency Nebulizer
DMA	Dimethylarsinic Acid
EP	Extraction Procedure
FFF/ICP	Flow Field-flow Fractionation/Inductively Coupled Plasma
FIA	Flow Injection Analysis
Flow-FFF	Flow-field Flow Fractionation
GC	Gas Chromatography
GC-MS	Gas Chromatography/Mass Spectrometry
GF	Graphite Furnaces
HG	Hydride Generation
HGA	Heated Graphite Atomizer
HPLC	High Performance Liquid Chromatography
HPLC-ICP	High Performance Liquid Chromatography Inductively Coupled Plasma
IC	Ion Chromatography
ICB	Initial Calibration Blank
ICP/AES	Inductively Coupled Plasma/Atomic Emission Spectroscopy
ICP	Inductively Coupled Plasma
ICP-MS	Inductively Coupled Plasma/Mass Spectrometry
ICS	Interference Check Standard
ICV	Initial Calibration Verification
IDL	Instrument Detection Limit
IEC	Interelement Correction Factor
ILMO	Inorganic Laboratory Manual Operations
MB	Method Blank
MCL	Maximum Contaminant Level
MMA	Monomethylarsonic Acid
MS/MSD	Matrix Spike/Matrix Spike Duplicate
MS	Mass Spectrometer
NPDES	National Pollutant Discharge Elimination System
NTU	Normal Turbidity Unit
PHG	Public Health Goal
QC	Quality Control
RCRA	Resource Conservation and Recovery Act
RF	Radio Frequency
RPD	Relative Percent Difference
RSD	Relative Standard Deviation

SDWA	Safe Drinking Water Act
SOW	Statement of Work
TISIS	Torch-integrated Sample Introduction System
USEPA	United States Environmental Protection Agency
USN	Ultra Sonic Nebulizer

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Industrial Waste Dumps, Sampling and Analysis • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Inductively Coupled Plasma/Optical Emission Spectrometry

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration

REFERENCES

1. A. Montaser, D.W. Gollightly (eds.), *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*, 2nd edition, VCH Publishers, New York, NY, 1992.
2. A. Montaser (ed.), *Inductively Coupled Plasma Mass Spectrometry*, 1st edition, Wiley-VCH Publishers, New York, 1998.
3. D. Beauchemin, 'A Comparison of ICP Atomic Spectrometry', *Spectroscopy (Eugene)*, **7**(7), 12–18 (1992).
4. R.K. Winge, V.A. Fassel, R.N. Kniseley, E. Kalb, W.J. Haas Jr, 'Determination of Trace Elements in Soft, Hard, and Saline Waters by the ICP Multi-element Atomic Emission Spectroscopic Technique', *Spectrochim. Acta, Part B*, **32**, 327 (1977).
5. USEPA-ICP Users Group, 'Inductively Coupled Plasma Atomic Emission Spectrometric Method for Trace Element Analysis of Water and Wastes – Method 200.7, Rev. 1.0, printed 1979 and published in 1982 in eds. T.D. Martin, J.F. Kopp, *Methods for the Determination of Metals in Environmental Samples – Supplement 1*, USEPA Office of Research and Development, EPA/600/R-94/111, 1994.
6. R.E. Wagner (ed.), *Guide to Environmental Analytical Methods*, 4th edition, Genium Publishing Corporation, Schenectady, NY, 1998.
7. G.A. Meyer, 'A Review of Sequential and Simultaneous ICP Spectrometers for Trace Metal Analysis', *Anal. Chem.*, **59**, A A 1345–1354 (1987).
8. V.A. Letourneau, B.M. Joshi, L.C. Butler, 'Comparison of Furnace AA Spectral Interferences between a Continuum and Zeeman Background Correction Spectrometer', *At. Spectrosc.*, **8**(5), 149 (1987).
9. D.E. Dobb, J. Heitschmidt, T.M. Hall, 'A Comparison of Radial and Axial-viewed ICP/AES Instruments for the Determination of As, Ag, Cd, Pb, Se and Tl in Environmental Samples', *The Pittsburgh Conference*, Chicago, IL, March 5, Paper 238P, 1996.
10. M.E. Tatro, L.C. Camp, R.K. Christenberry, 'Analysis of As, Pb, Se and Tl in Solid Waste using a Simultaneous Trace Analyzer ICP', *Spectroscopy*, **9**(7), 36 (1994).
11. J.T. Creed, C.A. Brockhoff, T.D. Martin, Method 200.8, 'Determination of Trace Elements in Waters and Wastes by Inductively Coupled Plasma – Mass Spectrometry', Revision 5.4 (1994), on line: http://web1.er.usgs.gov/nemi/method_pdf/4665.pdf, accessed December, 2007.
12. K.L. Linge, 'Trace Element Determination by ICP-AES and ICP-MS: Development and Applications Reported During 2004–2005', *Geostand. Geoanal. Res.*, **30**(3), 157–174 (2006).
13. J.M. Mermet, 'Is it Still Possible, Necessary, and Beneficial to Perform Research in ICP-Atomic Emission Spectrometry?' *J. Anal. At. Spectrom.*, **20**, 11–16 (2005).
14. J. Dennaud, A. Howes, E. Poussel, J.-M. Mermet, 'Study of Ionic-to-Atomic Line Intensity Ratios for Two Axial Viewing-Based Inductively Coupled Plasma Atomic Emission Spectrometers', *Spectrochim. Acta, Part B*, **56**, 101–112 (2001).
15. M. Duffy, R. Thomas, 'Benefits of a Dual-view ICP/OES for the Determination of B, P and S in Low Alloy Steels', *At. Spectrosc.*, **17**(3), 128 (1996).
16. C. Baird, 'Toxic Heavy Metals', *Environmental Chemistry*, 2nd Edition, W.H. Freeman and Co, New York, Chapter 7, 381–418, 1999.
17. J.S. Becker, 'Trace and Ultra Trace Analysis in Liquids by Atomic Spectrometry', *Trends Anal. Chem.*, **24**(3), 243–254 (2003).
18. E.H. Hansen, M. Miro, 'How Flow-injection Analysis (FIA) Over the Past 25 Years has Changed Our Way

of Performing Chemical Analyses?' *Trends Anal. Chem.*, **26**(1), 18–26 (2006).

19. J.S. Becker, R.S. Soman, K.L. Sutton, J.A. Caruso, H.-J. Dietze, 'Determination of Long-lived Radionuclides by Inductively Coupled Plasma Quadrupole Mass Spectrometry Using Different Nebulizers', *J. Anal. At. Spectrom.*, **14**, 933–937 (1999).
20. S-A.E. O'Brien, J.R. Chirinos, K. Jorabchi, K. Kahen, M.E. Cree, A. Montaser, 'Investigation of the Direct Injection High Efficiency Nebulizer for Axially and Radially Viewed Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **18**, 910–916 (2003).
21. J.L. Todolí, J.M. Mermet, 'Study of Matrix Effects Using an Adjustable Chamber Volume in a Torch-Integrated Sample Introduction System (TISIS) in ICP-AES', *J. Anal. At. Spectrom.*, **17**, 913–924 (2002).
22. W. Holak, 'Gas Sampling Techniques for Determination of Arsenic by Atomic Absorption Spectrometry', *Anal. Chem.*, **41**, 1712–1713 (1969).
23. P. Pohl, 'Hydride Generation-Recent Advances in Atomic Emission Spectrometry', *Trends Anal. Chem.*, **23**(2), 87–100 (2004).
24. A.D. Campbell (ed.), 'A Critical Survey of Hydride Generation Techniques in Atomic Spectrometry - Technical Report', *Pure & Appl. Chem.*, **64**(2), 227–244 (1992).
25. P. Smichowski, S. Farías, S.P. Arisnabarreta, 'Chemical Vapour Generation of Transition Metal Volatile Species for Analytical Purposes: Determination of Zn by Inductively Coupled Plasma-Optical Emission Spectrometry', *Analyst*, **128**, 779–785 (2003).
26. P. Phol, W. Zyrnicki, 'Study of the Reaction of Ir, Os, Rh and Ru ions with $NaBH_4$ in the Acid Medium by the Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **17**, 746–749 (2002).
27. R.K. Anderson, M. Thompson, E. Culbard, 'Selective Reduction of Arsenic Species by Continuous Hydride Generation.' Part I. Reaction Media', *Analyst*, **111**, 1143–1152 (1986).
28. K. Abbas-Ghaleb, N. Gilon, G. Crétier, J.M. Mermet, 'Pre concentration of Selenium Compounds on Porous Graphitic Carbon Column in View of HPLC-ICP-AES Speciation Analysis', *Anal. Bioanal. Chem.*, **377**, 1026–1031 (2003).
29. C. Bresson, C. Colin, F. Chartier, C. Moulin, 'Cobalt Speciation Study in the Cobalt-cysteine System by Electrospray Ionization Mass Spectrometry and Anion-exchange Chromatography Inductively Coupled Plasma Emission Spectrometry', *Appl. Spectrosc.*, **59**(5), 696–705 (2005).
30. T. Kato, T. Uehiro, A. Yasuhara, M. Morita, 'Determination of Methylmercury Species by Capillary Column Gas Chromatography with Axially Viewed Inductively Coupled Plasma Atomic Emission Spectrometric Detection', *J. Anal. At. Spectrom.*, **7**, 15–18 (1992).
31. U. Kumtabtim, J. Shiowatana, A. Siripinyanond, 'Sedimentation Field-Flow Fractionation-Inductively Coupled Plasma Optical Emission Spectrometry: Size-based elemental Speciation of Air Particulates', *J. of At. Anal. Spectrom.*, **20**, 1185–1190 (2005).
32. S. Saeseaw, J. Shiowatana, A. Siripinyanond, 'Observation of Salt-induced β-Lactoglobulin Aggregation Using Sedimentation Field-Flow Fractionation', *Anal. Bioanal. Chem.*, **386**(6), 1681–1688 (2006).
33. A. Seubert, 'On-line Coupling of Ion Chromatography with ICP-AES and ICP-MS', *Trends Anal. Chem.*, **20**(6–7), 274–287 (2001).
34. J.C.A. Wuilloud, R.G. Wuilloud, A.P. Venderheide, J.A. Caruso, 'Gas Chromatography /Plasma Spectrometry-an Important Analytical Tool for Elemental Speciation Studies', *Spectrochim. Acta, Part B*, **59**, 755–792 (2004).
35. B. Fairman, A. Sanz-Medel, 'Flow Injection-Mini-Column Technique with ICP-AES Detection for the Isolation and Preconcentration of the Fast Reactive Aluminum Fraction in Waters', *Fresenius J. Anal. Chem.*, **355**(7–8), 757–762 (1996).
36. B.F. dos Reis, M.F. Giné, F.J. Krug, H.B. Filho, 'Multipurpose Flow Injection System. Part 1. Programmable Dilutions and Standard Additions for Plant Digests Analysis by Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **7**, 865–868 (1992).
37. A.G. Cox, I.G. Cook, C.W. McLeod, 'Rapid Sequential Determination of Chromium (III)-Chromium (VI) by Flow Injection Analysis-Inductively Coupled Plasma Atomic-Emission Spectrometry', *Analyst*, **110**, 331–333 (1985).
38. Y. Israel, A.P. Krushevska, H. Foner, L.J. Martines, R.M. Barnes, 'Flow Configurations for interfacing Automatic Ion-Exchange Devices with Sequential Inductively Coupled Plasma Atomic Emission Spectrometry: Application to Trace Metal Determination in Pure Alkali Metal Salts', *J. Anal. At. Spectrom.*, **8**, 467–474 (1993).
39. *Test Methods for Evaluating Solid Waste: Physical/Chemical Methods*, Government Printing Office, Washington, DC, Vol. 1, December 1996 (also see the US-EPA web site: SW-846 online – http://www.epa.gov/sw-846/html for current updates, accessed November, 2007).
40. *Standard Methods for the Examination of Water and Wastewater*, 20th edition, part 3000, American Public Health Association, Washington, DC, 1998.
41. EPA/600/4-79-020, 'Methods for Chemical Analysis of Water and Wastes, USEPA', Office of Research and Development, Washington, DC, 1983.

42. EPA/600/R-94/111, 'Method 200.2 Sample Preparation Procedure for Spectrochemical Determination of Total Recoverable Elements in Water and Wastes by Inductively Coupled Plasma/Atomic Emission Spectrometry, revision 4.4 in Methods for the Determination of Metals in Environmental Samples – Supplement I', USEPA Office of Research and Development, May 1994.

43. EPA-540/R-95/121, 'USEPA Contract Laboratory Program Statement of Work for Inorganic Analysis, Document No. ILM04.0', US Environmental Protection Agency, Office of Solid Waste and Emergency Response, Government Printing Office, Exhibit D, Sec. III, D5, Washington, DC, 1995.

44. EPA/600/R-94/111, 'Method 200.7 Determination of Metals and Trace Elements in Water and Wastes by Inductively Coupled Plasma/Atomic Emission Spectrometry, revision 4.4 in Methods for the Determination of Metals in Environmental Samples-Supplement I', USEPA Office of Research and Development, May 1994.

45. EPA/600/R-94/111, 'Method 200.15 Determination of Metals and Trace Elements in Water by Ultrasonic Nebulization Inductively Coupled Plasma/Atomic Emission Spectrometry, revision 1.2 in Methods for the Determination of Metals in Environmental Samples – Supplement I', USEPA Office of Research and Development, May 1994.

46. B. Spence, K. Harper, J. Wills, F. Keenan, 'ICP Source Spectrometry Techniques in Regulated Water Analysis, Applications of ICP and ICP-MS Techniques for Today's Spectroscopists', Supplement to Spectroscopy, *Spectroscopy*, Special Issue, (October 2007), 8–23 (2007).

Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of

Sigrid Peldszus
University of Waterloo, Ontario, Canada

1 Introduction **1**
2 Background **2**
2.1 Air and Atmospheric Precipitation 2
2.2 Groundwater 3
2.3 Drinking Water 3
2.4 Wastewater 4
2.5 Landfill Leachates 4
3 Ion Chromatography **4**
3.1 Ion Exchange Chromatography Versus Ion Exclusion Chromatography 4
4 Sampling and Preservation **5**
4.1 Air 5
4.2 Rain 6
4.3 Ice 6
4.4 Groundwater 6
4.5 Drinking Water 6
4.6 Preservation 6
5 Sample Preparation and Measurement **7**
5.1 Air 7
5.2 Rain and Snow 7
5.3 Ice 7
5.4 Groundwater and Soil Pore Water 10
5.5 Drinking Water 10
5.6 Wastewater 11
6 Other Complementary Techniques **11**
Abbreviations and Acronyms **12**
Related Articles **12**
References **12**

The analysis of environmental samples typically focuses on aliphatic short-chain organic acids. The origin of these organic acids and the motivation for their analyses are unique for each type of matrix. Environmental sample matrices are very diverse and can range from air and atmospheric precipitation such as fog, rain or snow to wastewater, drinking water or landfill leachates and soil pore water. This diversity in matrix leads to substantial differences in organic acid concentrations and different organic acid distributions in the various types of environmental samples. Hence, the analytical techniques employed have to be adapted to the particular matrix under investigation.

*Ion chromatography is one of the most commonly used techniques for organic acid analysis. Competing ion chromatographic methodologies are anion (ion)-exchange chromatography (IC) and ion exclusion chromatography (ICE). The separation principles underlying each of these techniques are very different and may be used to their advantage. In IC, anions are separated by anion exchange processes between the cationic exchange groups on the resin and the eluent. Hence, inorganic anions and carboxylate anions may be analyzed at the same time (e.g. rain). Limitations of this technique set in when the inorganic anion concentration is much larger than the carboxylate concentration. As a consequence, inorganic anion peaks may mask or co-elute with the analyte of interest. However, techniques have been developed to handle this problem, for example in drinking-water analysis. The separation mechanism in ICE is much more complex than in IC involving Donnan exclusion, adsorption and steric exclusion. ICE is generally used for the separation of weak organic acids since strong organic acids elute in the system peak in the front of the chromatogram. Hence, large concentrations of strong inorganic anions, for example sulfate or chloride, will not interfere with the analysis of weak organic acids. However, organic acids with relatively low p*Ka values may be difficult to separate from the system peak. In addition, weak inorganic acids such as carbonate have the potential to interfere with the analysis of certain organic acids if present in large enough concentrations.*

Other techniques than ion chromatography have been used for organic acid analysis. Direct injections of samples into a gas chromatograph are very common for measuring volatile fatty acids (VFAs) in wastewater. More complex gas chromatography (GC) methods, which include concentration and derivatization steps, have been employed to analyze for specific types of organic acids such as dicarboxylic acids in rain. Capillary electrophoresis (CE) is a newer technique, which has not been used widely in routine analysis. However, speciality applications such as the determination of organic acids in a single raindrop show the potential of this type of methodology.

1 INTRODUCTION

Organic acids are hydrocarbons, which are characterized by their carboxylate function. Their hydrocarbon structure can vary considerably resulting in a wide range of organic acids with different physical and to a certain

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Structures and p*K*a values of organic acids

	Formula	Structure	pKa_1	pKa_2	Formula weight
Monocarboxylic acids					
Formic acid	CH_2O_2	HCOOH	3.75[a]	–	46.02
Acetic acid	$C_2H_4O_2$	CH_3-COOH	4.75[a]	–	60.05
Propionic acid	$C_3H_6O_2$	CH_3-CH_2-COOH	4.87[a]	–	74.08
Butyric acid	$C_4H_8O_2$	$CH_3-(CH_2)_2-COOH$	4.82[a]	–	88.10
Valeric acid	$C_5H_{10}O_2$	$CH_3-(CH_2)_3-COOH$	4.86[a]	–	102.13
Hydroxyacids					
Glycolic acid[d]	$C_2H_4O_3$	$H_2COH-COOH$	3.83[b]	–	76.05
Lactic acid[e]	$C_3H_6O_3$	$CH_3-HCOH-COOH$	3.89[a]	–	90.08
Ketoacids					
Glyoxylic acid[f]	$C_2H_2O_3$	CHO–COOH	3.34[c]	–	74.04
Pyruvic acid[g]	$C_3H_4O_3$	$CH_3-CO-COOH$	2.49[c]	–	88.06
Dicarboxylic acids					
Oxalic acid[h]	$C_2H_2O_4$	HOOC–COOH	1.23[b]	4.19[b]	90.04
Malonic acid[i]	$C_3H_4O_4$	$HOOC-CH_2-COOH$	2.83[b]	5.69[b]	104.06
Ketomalonic acid[j]	$C_3H_2O_5$	HOOC–CO–COOH			118.05
Oxalacetic acid[k]	$C_4H_4O_4$	$HOOC-CO-CH_2-COOH$	2.22[a]	3.89[a]	132.07

[a] Ohta et al,[1]
[b] Weast,[2]
[c] Budavari,[3]
[d] Hydroxyacetic acid,
[e] 2-hydroxypropanoic acid,
[f] oxoacetic acid = formylformic acid,
[g] 2-oxopropanoic acid = α-ketopropionic acid = acetylformic acid = pyroracemic acid,
[h] ethanedioic acid,
[i] propanedioic acid = methanedicarboxylic acid,
[j] mesoxalic acids = oxopropanedioic acid = oxomalonic acid,
[k] oxobutanedioic acid = oxosuccinic acid = ketosuccinic acid.

extent chemical properties. The analysis of environmental samples typically focuses on aliphatic short-chain organic acids. The most common ones are listed together with their structure and p*K*a values in Table 1. The origin of these organic acids in environmental samples and the motivation for their analyses varies from matrix to matrix as outlined in the background section. Table 2 gives an overview of the different types of environmental sample. A brief general discussion of the ion chromatographic techniques commonly utilized and their advantages and disadvantages with respect to the organic acid analysis follows.

2 BACKGROUND

2.1 Air and Atmospheric Precipitation

Organic acids in rain were first detected in the 1970s.[4] They also have been found to be present in air and all other forms of wet atmospheric precipitation such as fog, snow and ice. Possible sources of emission are reported to be of anthropogenic as well as biogenic origin. However, little information is available about biogenic sources.[5] Direct sources of anthropogenic emission include vehicle exhaust fumes[6,7] and combustion of coal or wood.[8–10] Directly emitted hydrocarbons can be viewed as indirect anthropogenic sources of organic acids in the atmosphere. These hydrocarbons presumably act as organic precursors, which are oxidized through complex photochemical reactions into organic acids. The ongoing discussion in the literature indicates that it is difficult to elucidate the interactions and the detailed reaction mechanisms of the radicals involved, especially when taking into account the role of the gaseous phase, the water phase and adsorption processes onto particles.[10–12]

Formic and acetic acid are usually present in much higher concentrations in all of these matrices than any of the other organic acids detected, including dicarboxylic and ketoacids. The total amount of organic acids can be quite significant. It is reported that in North American rain up to one-third of the total free acidity is caused by organic acids.[13,14] However, it is postulated that organic acids do not contribute significantly to the long-term acidification of the environment since they are easily biodegradable.[13] Research is focusing on gaining more insight into reaction mechanisms as well as cataloging possible emission sources of organic acids.

Table 2 Types of environmental sample

Matrix	Type of organic acid	Expected organic acid concentration	Other matrix characteristics
Air	Mainly formic and acetic acid	Very low ppbv = low $\mu g\,L^{-1}$ volume	SO_2, NO_x, CO, O_3 at low to medium ppbv Particles present too
Rain, snow	Formic and acetic acid predominate Other mono- and dicarboxylic acids, keto- and hydroxy acids	Low to high $\mu g\,L^{-1}$ depending on organic acid species and location	Inorganic cations and anions at $\mu g\,L^{-1}$ especially SO_4^{2-}, Cl^- and NO_3^- Other organic compounds such as hydrocarbons from fuel emissions or terpenes from natural biogenic sources
Ice	Acetic and formic acid most abundant, some glycolic and oxalic acid	Low $ng\,g^{-1}$	Inorganic anions at low $ng\,g^{-1}$
Groundwater, soil pore water	Monocarboxylic acids especially acetic acid	Low to high $\mu g\,L^{-1}$ depending on location and other factors	Vary depending on geology, contamination, location etc.
Drinking water (incl. raw and treated water)	Ozonation by-products: Predominantly acetic, formic and oxalic acid Other mono- and dicarboxylic acids, hydroxy and ketoacids	Low to medium $\mu g\,L^{-1}$	Inorganic anions and cations at low to medium $mg\,L^{-1}$ Other organic compounds: typical TOC 2–7 $mg\,C\,L^{-1}$
Wastewater	Predominantly aliphatic, monocarboxylic acids (C_1–C_7 = VFA) especially in anaerobic digesters	High $mg\,L^{-1}$	Medium to high $mg\,L^{-1}$ of inorganic cations and anions High concentrations of organic compounds
Landfill leachates	Monocarboxylic acids, keto- and hydroxyacids	Low to medium $mg\,L^{-1}$ depending on type of acid and age of landfill	Medium to high $mg\,L^{-1}$ of inorganic cations and anions High concentrations of organic compounds

TOC, total organic carbon.

2.2 Groundwater

Organic acids can be found close to seeps of naturally occurring hydrocarbons into groundwater, or in the proximity of groundwater sites contaminated with petroleum hydrocarbons. Biodegradation of these hydrocarbons under the anaerobic conditions of the groundwater aquifer leads to a variety of metabolic intermediates, especially organic acids.[15,16] Although it is expected that the organic acid intermediates are structurally related to the original hydrocarbons, high concentrations of short-chain aliphatic organic acids are most commonly observed, even when aromatic hydrocarbons are the original source. The localized concentration of formate, acetate and isobutyrate combined may be as high as $9000\,\mu mol\,L^{-1}$,[17] but varies considerably depending on the rate of organic acid production and consumption. Hence, factors such as the proximity to the hydrocarbon source, hydrocarbon loading, presence of microbial inhibitors, availability of nutrients and availability of electron acceptors affect the organic acid concentration in groundwater.

The main effects of high organic acid concentrations in groundwater are mineral dissolution, which leads to changes in soil structure, and complexation of metals such as Fe or Al, which may subsequently be mobilized into the aquifer.[17] Hence, it is important to take the organic acid production and its consequences into account when observing and projecting the transport of hydrocarbon contamination in groundwater[17] and when estimating changes in the porosity of oil reservoirs due to organic acid formation in oilfield water.[18,19]

2.3 Drinking Water

Ozone is frequently applied during drinking-water treatment, mainly for disinfection purposes but also to break down taste and odor compounds. Other reasons for its utilization include color removal and pretreatment.[20] In Europe, ozone has been established in many drinking-water treatment plants for quite some time, whereas in North America, it has only gained interest since the late 1980s. In North America, ozone is often applied with the intention of reducing chlorinated disinfection by-products further by partially substituting chlorine with ozone as the disinfectant.

However, organic compounds which are present naturally, especially in surface water, are oxidized during ozonation, and organic acids and aldehydes are formed as well as a range of other products. Various short-chain aliphatic organic acids (for example formate, acetate and oxalate) have been identified and quantified in partially treated water as a result of the utilization of ozone during drinking-water treatment.[21–26]

Concentrations of aliphatic organic acids are highest in the effluent of ozone contactors, where they may reach concentrations of $250\,\mu g\,L^{-1}$ for formate or $195\,\mu g\,L^{-1}$ for acetate.[25] Subsequent steps in the drinking-water treatment process reduce these concentrations to low microgram per liter levels or remove them completely.[25] The finished drinking-water is then transported to the customer through a drinking-water distribution system, where any remaining organic acids may contribute to bacterial regrowth, since they are a good nutrient source for microorganisms. Current research conducted in cooperation with the drinking-water industry continues to assess the formation, the removal and the impact of ozonation by-products such as organic acids.

2.4 Wastewater

Organic acids are produced in high milligram per liter concentrations during wastewater treatment in anaerobic digesters. These acids are called VFAs and are defined as ''water soluble fatty acids which can be distilled at atmospheric pressure''.[27] VFAs comprise monocarboxylic aliphatic organic acids with chain lengths of up to seven C-atoms.

Wastewater usually has a high organic carbon (OC) content, which needs to be reduced before the water is discharged into the receiving water body. This is often accomplished by a two-step biological degradation in anaerobic digesters.[28,29] First, acidogenic bacteria convert the OC into VFAs with acetate being the major product, followed by metabolization of these VFAs by methanogenic bacteria into methane. These two processes have to be in balance to ensure successful treatment. The difficulty lies in the fact that methanogenic bacteria metabolize at a slower rate than the acidogenic bacteria. In addition, methanogenic bacteria are usually more susceptible to sudden changes in their environment than the acidogenic bacteria.[28–30] Hence, it is possible that VFAs may not be consumed at the same rate as they are produced, with the consequence that VFA concentrations increase. This in itself may disturb the balance between these processes even further.

Daily monitoring of VFAs is utilized as a sensitive measure of the status of the OC biodegradation. Changes in VFA concentrations indicate disturbances of the biodegradation at an early stage, therefore, making VFA measurement a valuable operational tool in wastewater treatment.[28,29]

2.5 Landfill Leachates

The anaerobic degradation of organic waste in landfills passes through several stages.[31] During the acidogenic phase organic acids, mainly VFAs, are formed in high concentrations and consequently contribute significantly to the OC concentration in landfill leachates. One problem associated with VFA production is the possible mobilization of heavy metals out of the landfill. This is especially of concern when older landfill sites are involved. Older landfill sites often have no liner or only incomplete seals towards the groundwater aquifer and heavy metals may actually be transported into the groundwater.

By monitoring VFA concentrations in landfill leachates, the degradation stage of the landfill itself can be determined. Once VFA concentrations are rising, precautions can be taken to prevent leaking of mobilized heavy metals into the groundwater.

3 ION CHROMATOGRAPHY

3.1 Ion Exchange Chromatography Versus Ion Exclusion Chromatography

Ion chromatographic techniques commonly used for the determination of organic acids are IC and ICE which differ significantly in their separation mechanisms.

In IC, separation is primarily accomplished through partitioning of the analyte ions between the mobile phase and the ion-exchange groups bound to the surface of the column resin. Secondary mechanisms may involve adsorption processes. IC columns for the separation of anions are comprised of a copolymer core, usually divinylbenzene crosslinked with polystyrene, to which modified quaternary ammonium groups are bound on the surface. With the development of membrane suppressors, conductivity detectors have found widespread use in IC. They allow for very sensitive detection of ions by suppressing the background conductivity of the mobile phase, which usually consists of hydroxide, borate, bicarbonate or carbonate. Other detectors utilized in IC include amperometric and UV (ultraviolet)-detectors. IC is normally applied to the analysis of strong inorganic ions. However, compounds such as organic acids, which dissociate at the high pH of the mobile phase into ions, may also be analyzed by IC. Depending on the matrix, inorganic anions and carboxylate anions may be determined at the same time, although separation becomes difficult when the concentration of inorganic anions is much higher than

the concentration of organic acids. This may even result in co-elution and hence masking of organic acid peaks, rendering the particular method unsuitable for organic acid analysis.

Separation in ICE is a much more complex process than in IC and involves Donnan exclusion, adsorption and steric exclusion. ICE columns have high capacities and are usually composed of totally sulfonated divinylbenzene/polystyrene copolymers, which carry strong cation exchange groups. Traditionally UV detection predominates, although conductivity detection is becoming more common. Eluents applied in ICE are aqueous solutions of acids. Dilute mineral acids are mainly in use as eluents in combination with UV detection, whereas strong organic acids such as methanesulfonic or octanesulfonic acid are in use in combination with suppressed conductivity detection. Owing to its specific separation mechanism, ICE is mainly applied to the analysis of weak organic acids in complex matrices. Anions of strong acids, for example sulfate or chloride, elute with the system peak at the front of the chromatogram, thus making this technique suitable for the determination of weak organic acids in samples with high concentrations of strong inorganic anions. Nevertheless, weak inorganic acids, for example carbonate, elute between the organic acids and may, if present in sufficiently high concentrations, interfere with the analysis of the organic acids. In addition, organic acids with relatively low p*K*a values may be difficult to analyze with ICE, since they elute early, close to the system peak.

More detailed descriptions of theoretical background, instrumentation and applications in IC and ICE can be found in books by Weiss[32] and by Haddad and Jackson.[33]

4 SAMPLING AND PRESERVATION

Sampling itself is an important part of the overall analytical process. The ideal sampling technique is adapted to the matrix being sampled, prevents contamination and results in an unaltered sample which reflects the average content of the analyte at the time when the sample was taken. To fulfill these criteria different procedures are applied for different types of matrices. A summary of different sampling techniques is given in Table 3.

When dealing with very low concentrations of organic acids, contamination with the almost ubiquitously present organic acids has to be avoided. It was found that organic acids are present on the human skin and wearing gloves is highly advisable in order to minimize contamination from this source.[53]

4.1 Air

Air samples are usually taken by pumping ambient air for a defined time interval, first, through a Teflon® prefilter to remove particles, and then through a device which retains the organic acids. Devices and processes used for this purpose include filters, scrubbers, denuders, condensation or cryogenic trapping.[34,35,49–51,54] Solutions used in scrubbers may be water or alkaline solutions.[34,35] Filters are also impregnated with alkaline solutions[11,36,51] and denuders are coated with either NaOH or KOH.[37,49] Contamination of these sampling devices with the ubiquitous organic acids is easy. Precautions include thorough cleaning of the sampling devices accompanied by regular measurements of blanks. Another problem is the possible formation of artifacts during long sampling periods in the presence of alkali and/or water which makes

Table 3 Summary of sampling techniques

Matrix	Technique used	Comments	Refs.
Air	(a) Scrubber (b) Denuder: NaOH or KOH coated (c) Filters impregnated with KOH or Na_2CO_3 (d) Cryogenic trapping	• Sampling time varies from 30 min to several hours • DLs depend on sampling time • Contamination of water phase possible • In alkaline solutions formation of acetate from PAN possible	(a) 34, 35 (b) 49, 37 (c) 11, 51, 36 (d) 50, 51
Rain	(a) Setup varies from bottle + funnel to fully automated wet only collection systems (b) Collection of individual raindrops possible	(a) Collection either by rain event or in bulk (b) Analysis only by micro HPLC or CZE	(a) 4–6, 13, 14, 38–46 (b) 47, 52
Ice	Drill cores	Contamination of outer core unavoidable. Remove outer layers prior to melting	48
Groundwater	Well with screen Utilize pump or bailer	Pump until pH and conductivity are stable, then take sample	15, 16
Drinking water	Use sampling ports or bailer	Flush sampling line until stagnant water is removed	–

DL, detection limit; PAN, peroxyacetylnitrate; HPLC, high-performance liquid chromatography; CZE, capillary zone electrophoresis.

it difficult to distinguish between organic acids trapped from the air and those formed while sampling.[34,50]

Recoveries and DLs in particular vary considerably between these different sampling techniques. DLs can be as low as 10 pptv for formic and acetic acid when utilizing the more established scrubber (recoveries 55–98%)[35,50] or the relatively new cryogenic trapping technique (recoveries 100%).[50] When utilizing filters or denuders, DLs are considerably higher (20–450 pptv), though recoveries between 84 and 99% ensure reliable data.[50]

4.2 Rain

Rain collectors, which are typically set up high above ground, for example on rooftops, consist of a collection container and a funnel. Materials utilized are glass, polyethylene, Teflon® and stainless steel. The design of rain collectors varies from a simple setup like a wide neck bottle with a funnel, which is operated manually, to sophisticated automatic wet-only collection systems.[4–6,13,14,38–46] The rain collectors must be scrupulously cleaned before each sampling in order to avoid contamination.

Generally, rain collectors are kept shut until the rain event starts so that dry deposition is excluded from sampling.[38] Rainwater is sampled either by rain event or, less frequently, in bulk. When collecting in bulk, rain collectors are replaced after a defined time period. Although this makes operation relatively easy, samples represent only an average of all of the rain events during the past time interval. When sampling by rain event, collectors are exchanged after a single rain event resulting in an average sample of this specific event.

It is also possible to collect individual raindrops, classify them by size and analyze them for organic acids as well as for inorganic cations and anions.[47,52] This is accomplished with a so-called 'Guttalgor'. Raindrops entering this collection system during a brief opening period freeze immediately when they come into contact with liquid nitrogen and sink to the bottom of the vessel. The frozen raindrops may then be grouped by size by using sieves with different mesh ranging from 0.1–1.0 mm. The analysis of these individual raindrops with their extremely small volume is accomplished by either CE or micro HPLC.[47,52]

4.3 Ice

Ice is sampled by drilling cores. These cores are divided into segments which are then analyzed individually. By relating each segment to its date of origin, information can be gained about the composition of the atmosphere in the past.[10] The organic acid concentration in ice from remote areas is usually quite low and care has to be taken to avoid contamination. It was shown that it is not possible to avoid contaminating the outer parts of ice cores considerably with organic acids and ions during the sampling process. These outer layers have to be removed and only the inner parts of the ice core should be used for analysis.[48] Even extended exposure of the melted ice to the laboratory atmosphere may result in contamination with organic acids.[48]

4.4 Groundwater

Groundwater samples are taken from wells by using either a bailer or a pump. To get a representative sample substantial volumes of groundwater have to be removed until pH and conductivity are stable. Only then should the actual sample be taken.[16] Contamination should be avoided by working with clean equipment and storing samples in clean glass containers.

4.5 Drinking Water

Most drinking-water treatment plants have special sampling ports at the different stages of their treatment process. Before taking samples, the lines leading to these ports need to be flushed thoroughly until the water at the sampling port has the same composition as the water at this particular stage of the treatment process. In plants which do not have these sampling ports, bailers have to be used. During sampling, skin contact should be avoided since organic acids are present in significant concentrations on the human skin.[53] Wearing gloves was found to be an appropriate measure to minimize contamination from this source.

4.6 Preservation

Preservation is closely related to the sampling procedure and the measurement technique applied. Ideally it should ensure stable concentrations of the analytes during sample storage while being compatible with the measurement technique utilized later.

The use of IC or ICE techniques requires the sample to be in a liquid aqueous phase prior to injection. If samples are not already in this form, analytes are transferred into the aqueous phase (for example extraction of filters from air samples with deionized water). Samples are usually stored at 4 °C. However, several authors found this to be insufficient to prevent organic acid losses due to biological degradation.[13,39,53,55] As a consequence, preservatives, usually chloroform or mercuric chloride, are added to aqueous samples. When using IC or ICE for sample analysis, adding a small volume of neat chloroform is the preferred preservation technique, since this does not interfere with the measurement.[13,39,42,53,55] In contrast, when using mercuric chloride as a preservative, the mercury cation has to be removed prior to injection (e.g. by

passing the sample through a cation exchanger) in order to prevent poisoning of the IC or ICE column.[24] Hence, mercuric chloride as a preservative is more prevalent in the context of GC measurements.[15,16,43–46] Another, although less common preservative for environmental samples is benzalkonium chloride, which was recently utilized for stabilizing drinking-water samples.[26]

Freezing samples is another option for sample preservation. Ice cores are usually kept frozen and melted just before their measurement.[10,48] Filters used for sampling air may be stored at $-4\,^{\circ}C$ or $-20\,^{\circ}C$ until they are extracted with deionized water in preparation for their measurement.[11,34,51] Sometimes aqueous sample solutions are also stored at subzero temperatures.[56]

5 SAMPLE PREPARATION AND MEASUREMENT

Sample preparation and measurement are intertwined but differ from matrix to matrix. A summary of methods discussed herein, sorted by matrix can be found in Table 4.

5.1 Air

Concentrations of organic acids in air are quite low (very low parts per billion by volume concentrations). They are significantly lower than concentrations of trace gases such as NO_x or CO. During the sampling procedure and the subsequent sample preparation organic acids are enriched, resulting in aqueous solutions which contain organic acids in low micromol per liter concentrations.

If alkaline filters or denuders are used as the main trapping device, sample preparation consists in extraction with either deionized water or eluent followed by preservation.[34,49] If using a scrubber, condensate process or cryogenic trapping as the main sampling technique,[12,34,50] aqueous solutions are the result and no further sample preparation is required other than immediate preservation. Prefilters, which are used to remove particles prior to air sampling, may be extracted as well. First they are wetted with methanol and then extracted with either deionized water or eluent.[34] The organic acids extracted from these prefilters have been adsorbed onto particles, which were trapped with these prefilters.

All these extracts contain inorganic anions in addition to the organic acids. Fortunately, inorganic anions are present in relatively low concentrations and thus do not interfere with organic acid separation. Organic acid analysis focuses mainly on formate and acetate utilizing both IC and ICE methods, with IC being applied more frequently.[12,34,49–51] Standard columns and conditions are employed as described briefly in Table 4. Additional organic acids such as pyruvate or oxalate may also be analyzed.[12,50,61] If utilizing newer columns, which have been developed for the determination of organic acids, an even wider range of organic acids may be analyzed.[41]

5.2 Rain and Snow

Sample preparation for snow samples consists merely of melting, filtration if necessary and preservation. Organic acid content in rain and melted snow range from low to high micrograms per liter concentrations depending on location, precipitation event and type of acid.[4–6,13,14,39,40,42–47] The more predominant inorganic anions in rain (such as sulfate, sulfite, nitrate and chloride) usually display a somewhat higher concentration.[50] However, in these concentrations they do not cause any major difficulties for the separation of organic acids from inorganic anions. Again, formate and acetate are the major acids of concern and samples may be analyzed by the same standard IC and ICE methods utilized for the analysis of organic acid in air-sample extracts (Table 4). Better yet, some newer IC methods allow for the determination of additional acids such as oxalate as well as for the codetermination of various common inorganic anions (F^-, Cl^-, NO_2^-, NO_3^-, Br^-, SO_3^{2-}, SO_4^{2-} and PO_4^{3-}).[40,41] Hence, these newer IC methods may be preferred over ICE methods, which by nature are only capable of determining weak organic acids.

5.3 Ice

When drilling ice cores it is generally unavoidable to contaminate the outer ice core layers. Hence, these layers have to be removed prior to melting the ice and preserving the resulting solution.[48]

Ice from remote areas usually contains very low concentrations of organic acids and inorganic anions (i.e. low nanograms per gram concentrations). In order to measure these low concentrations organic acids have to be enriched. Utilizing a concentrator column on-line with an analytical column achieves the required low DLs while keeping the sample preparation at a minimum. Concentrator columns suitable for this purpose are small high-capacity anion-exchange columns, which retain the carboxylate anions quantitatively upon loading, if appropriate conditions have been chosen. ICE columns cannot be used as concentrators since water itself acts as an eluent with the consequence that organic acids are not retained quantitatively.

In order to analyze organic acids in ice samples, anion-exchange concentrators have been combined successfully either with an anion-exchange analytical column or with an ICE analytical column.[48,57] Organic acids determined with these setups are either propionate and butyrate, or

Table 4 Method overview organized by matrix

	Type of column[a]	Eluent	Detector[b]	Compounds	DLs	Comments	Refs.
Air							
IC	AG 9, AS 9	$Na_2B_4O_7$ 1 M	Cond. A	Acetic, formic acid	0.1 ppbv[c]	Filters stored at subzero temperatures	51
IC	AS 11	NaOH gradient	Cond.	Acetic, formic, pyruvic, oxalic acid	[c]	Simultaneous detection of inorganic anions possible (Cl^-, NO_2^-, SO_4^{2-}) Sampling with cryogenic trapping	50
IC	AS 4	$NaHCO_3$ 0.4 mM	Cond. A	Formic, acetic, pyruvic acid	0.02–0.1 µM (IC limits)	DLs for 60 min sampling interval: 5–20 pptv Matrix: mist	12
IC	AG 4, AS 4	$Na_2B_4O_7$ or $Na_2CO_3/NaHCO_3$	Cond. A or B	Formic, acetic acid	[c]	Reference gives comparison of different sampling techniques	34
ICE	AS 1	HCl 1 mM	Cond.	Formic, acetic acid	[c]	Reference gives comparison of different sampling techniques	34
ICE	AS 1	Octane sulfonic acid 1 mM + 2% isopropylalcohol	Cond. D	Formic, acetic acid	[c]	Sampling with KOH coated denuder	49
Rain							
IC	AG 4, AS 4A	$Na_2B_4O_7$ 1.5 mM	Cond. A	Formic, acetic, pyruvic acid	0.2 µM	Focus on formic and acetic acid	42
IC	AG 9, AS 9	$Na_2CO_3/NaHCO_3$ gradient	Cond. A	Formic, acetic, oxalic acid	20–100 µg L^{-1}	Simultaneous detection of inorganic anions (F^-, Cl^-, NO_2^-, NO_3^-, Br^-, SO_3^{2-}, SO_4^{2-} and PO_4^{3-}) Matrix: rain, snow	40
IC	ATC-1 (a) AG 11, AS 11 (b) AG 10, AS 10	(a) Borate gradient (b) Borate 7 mM isocratic	Cond. A or B	2- and 3-hydroxy-butyric, lactic, acetic, glycolic, propionic, formic, butyric, pyruvic, valeric, oxalic acid	0.5-1 µM	Simultaneous detection of inorganic anions (F^-, Cl^-, NO_2^-, CO_3^{2-}, PO_4^{3-}, Br^-, SO_3^{2-}, SO_4^{2-}) Other matrices: lake sediment pore water, fog	41
ICE	Separator: ICE 30580	0.002 N HCl; pump rate 10%	Cond., suppressor column: ICE 30960	Formic, acetic, citric, lactic, glycolic, propionic, butyric and valeric acid	na; but <0.6 mg L^{-1}	Sample preservation with $CHCl_3$	39
Ice							
IC/ICE	IC concentrator column and HPX-87H, Biorad	MSE acid at pH 2.7	UV at 200 nm	Formic, acetic, propionic and butyric acid	5.6–9.4 µg L^{-1}	Concentrator column conditioned with MSE at pH 9 Injection volume 10 ml	57

IC/IC	2 ATC-1 in series; TAC1[d] (a) AG 5, AS 5 (b) CH10, Pax 500	(a) + (b) different NaOH gradients	Cond. A	Formic, acetic, oxalic, glycolic acid	0.2–0.6 ng g^{-1}	Simultaneous detection of inorganic anions (F^-, Cl^-, NO_2^-, NO_3^- and SO_4^{2-}) Injection volume 5 mL	57
Groundwater							
ICE	ICE-AS 1	1 mM HCl isocratic	Cond.	Acetic, propionic, isobutyric and *n*-butyric acid	0.5 µM	–	58
Drinking Water							
IC	ATC-1 AG 11, AS 11	NaOH gradient	Cond. A	Formic, acetic, oxalic acid; propionic, pyruvic, glyoxylic, ketomalonic, dichloroacetic acid	2–3 µg L^{-1}; MRL 15 µg L^{-1}	Removal of Hg^{2+} and other cations through H^+-cartridge essential Daily maintenance of suppressor with 0.5 N H_2SO_4 imperative	24, 26
IC	ATC-1, AG 10, AS 10	NaOH gradient	Cond. C	Formic, acetic, butyric, β-hydroxybutyric, glycolic, pyruvic, α-ketobutyric, oxalic acid	1–9 µg L^{-1}	Oxalate determination in matrices with high ionic strength requires switching technique	53, 55, 59
IC	AG 11[d], AG 11, AS 11	NaOH gradient	Cond.	Formic, acetic, glycolic, oxalic acid	20–40 µg L^{-1}; concentr. 0.1–0.5 µg L^{-1}	Sample matrix: Ozonated model water. No testing on "real" drinking water samples described	60
Landfill Leachate							
ICE	(a) Polyspher, OA-HY (Merck) (b) HPICE, AS-6 (Dionex)	(a) H_2SO_4; 5 mM and 50 mM (b) PFBA 0.4 mM and 1.6 mM	(a) UV at 210 nm (b) Cond. D	Formic to *n*-valeric acid, pyruvic, glyoxylic, glycolic, lactic, glyceric, succinic acid	na; analyzed from 50 to 50 000 µM	Sample stored frozen	56

na = not available. MRL, minimum reporting level; PFBA, perfluorobutyric acid; MSE, methanesulfonic acid.

[a] Columns: Columns are manufactured by Dionex if not mentioned otherwise. AG refers to guard columns, AS refers to analytical columns, ATC-1 is the anion trap column used to purify eluents.

[b] Detection: Cond. is the conductivity detection, no further details available, Cond. A is the chemical suppressed conductivity detector using H_2SO_4 as regenerant, Cond. B is the suppressed conductivity detection in auto suppression mode, Cond. C is the suppressed conductivity detection in external water mode, Cond. D is the chemical suppressed conductivity detector using tetrabutylammonium hydroxide as regenerant.

[c] DLs vary with sampling time for air samples.

[d] Used as concentrator column.

glycolate and oxalate, in addition to formate and acetate. The newer method using the IC/IC combination[48] achieves lower DLs than the IC/ICE method. Moreover, the IC/IC method has the added advantage over the IC/ICE method[57] in that inorganic anions (F^-, Cl^-, NO_2^-, NO_3^-, SO_4^{2-}) can be determined within the same chromatogram as the organic acids.

Concentrator column use is limited by the fact that the analytes have to be retained quantitatively. If anions with strong affinities to the anion-exchange concentrator resin such as inorganic anions are present in higher concentrations, carboxylate anions may be forced off the concentrator column during the loading process, thus leading to irreproducible injections and, hence, irreproducible results. However, inorganic anion concentrations in ice samples are usually low enough so that overloading of the concentrator is very unlikely. Nevertheless, the possibility of column overloading should always be taken into consideration when working with concentrators.

5.4 Groundwater and Soil Pore Water

When dealing with samples from contaminated groundwater sites, organic acid content and matrix can vary significantly depending on factors such as type of contamination, proximity to the contamination and geology. Therefore, organic acid content can fluctuate from low to very high microgram per liter concentrations.[15,16] Inorganic anion concentrations as well as type and concentration of contaminants also differ widely from site to site. As a consequence, analytical techniques may have to be adapted to the particular sample matrix.

Another aspect of dealing with contaminated groundwater sites is the possible presence of organic solvents and/or metals in these samples. Columns should be compatible with these specific organic solvents, and metals should be removed from the sample prior to injection by passage through H^+ cartridges in order to prevent column poisoning. Nevertheless, filtration is often the only sample preparation step employed in the IC and ICE methods described below.

An ICE method with standard conditions has been successfully applied to organic acid analysis of samples from a groundwater site, which has been contaminated with petroleum hydrocarbons.[58] Similar ICE methods have been used in the analysis of extracts derived from air samples, and therefore, may also be utilized for groundwater analysis (Table 4). One advantage of ICE methods is the fact that anions derived from strong acids such as mineral acids elute with the system peak at the front of the chromatogram. Hence, inorganic anions such as chloride or sulfate will not interfere with the separation of weak organic acids, even if present in higher concentrations.

Anion-exchange chromatography has been used for organic acid determination in soil pore water[41] and may also be used for organic acid quantification in a contaminated groundwater matrix, although inorganic anions might interfere with the separation of specific organic acids (for example sulfate/oxalate), especially if they are present at much higher concentrations than the organic acids. Similar problems are encountered during drinking-water analysis, which is why strategies applied in drinking-water analysis may also be utilized for groundwater samples containing high inorganic anion concentrations.

5.5 Drinking Water

Organic acids in drinking-water including partially treated water are only present at low microgram per liter concentrations, if at all. The exceptions are effluents from ozone contactors, an accepted drinking-water treatment practice, where organic acid concentrations may reach medium microgram per liter levels.[22,24–26] In contrast to other matrices, concentrations of inorganic anions (Cl^-, SO_4^{2-}, CO_3^{2-} ...) in drinking water are much higher compared to organic acids, ranging from low to medium milligram per liter concentrations.

Although different approaches have been taken, it is surprising that no ICE methods for organic acid analysis in drinking water are reported in the current literature. Problems anticipated when using ICE include interferences from weak inorganic acids such as carbonate and separation problems between the early eluting system peak and organic acids with low p*K*a values such as oxalate or pyruvate.

Anion-exchange chromatography has been used to determine organic acids in drinking water, although very low concentrations of carboxylate anions have to be measured in the presence of large, possibly interfering, inorganic anion concentrations. Slightly different approaches have been taken in the few IC methods dealing with this analytical problem. The focus is usually on formate, acetate and oxalate, although a few other organic acids may be determined as well.

Injection of large sample volumes (for example 760 μL) is one way of ensuring low DLs.[53,55] Owing to the so-called ''relaunch effect''[33] analytes are focused into a relatively small band at the beginning of the column resulting in a chromatogram with narrow peaks, even when injecting larger sample volumes. Subsequent separation is accomplished by a gradient developed specifically for this purpose on a standard anion-exchange column. However, very high concentrations of sulfate can obscure the oxalate peak to the extent that its quantification is impossible. For this type of sample, a two-phase switching technique may be used.[55,59] First, samples

are injected and separated in the manner described previously. At the time window when oxalate is known to elute, the column effluent is redirected onto a concentrator column where the oxalate is trapped. Subsequent reinjection of the concentrator contents onto the same analytical column and application of a different specific gradient makes the quantification of oxalate possible.

In another IC method Hg^{2+} cations, which were initially added to the samples for preservation purposes, have to be removed prior to injection in order to avoid column poisoning.[24] This is accomplished by putting a H^+ cartridge, which is exchanged daily, in line just before the injection valve.[24,26] Separation takes place on a specific organic acid column with subsequent suppressed conductivity detection. Utilization of the H^+ cartridge is also essential for achieving satisfactory chromatograms, since otherwise organic acids eluting before chloride display broad, plateau-type shapes instead of narrow peaks. Although the injection volume is only 50 µL, sufficiently low DLs are achieved.[24,26]

A different approach is to remove chloride and sulfate before IC measurement through pretreatment with barium and silver cartridges.[60] This should be followed by passing the sample through a H^+ cartridge, so that Ag^+ cations, which may have leached into the sample from the previous Ag^+ cartridge, are trapped thus preventing poisoning of the analytical column with Ag^+ cations. A very similar pretreatment procedure is used for low-level bromate analysis in drinking water.[62,63] The treated sample may be injected directly into the ion-exchange chromatograph, where organic acids are separated on a specific organic acid column followed by suppressed conductivity detection. DLs are 20–40 µg L^{-1} for an injection volume of 25 µL.

Using the same approach these limits can be lowered substantially by introducing a concentrator column.[60] The pretreated sample is passed through a strong anion-exchange concentrator column, where the carboxylate anions and other anions are retained. The trapped anions are then eluted from the concentrator onto the separation column in the opposite direction to the loading process. Separation and detection are performed in the same manner as for the direct injection. However, different matrices can contain widely varying concentrations of inorganic anions. Although the majority of the inorganic anions are removed during pretreatment, in the case of very high concentrations this pretreatment may be insufficient. Inorganic anions that remain in the sample may lead to overloading of the concentrator column. Consequently, the less strongly retained carboxylate anions are partially eluted from the concentrator column during sample loading thus leading to irreproducible results. It is therefore important to prevent any breakthrough of the carboxylate anions on the concentrator column when using this method.

5.6 Wastewater

Wastewater, especially from anaerobic digesters, displays a high content (milligrams per liter concentrations) of VFAs, which comprise short-chain aliphatic organic acids. The wastewater matrix is quite complex with very high concentrations of organic material and inorganic compounds.[28,29] Landfill leachates have a very similar composition and may therefore be analyzed by similar methods.[31]

VFAs are most often analyzed by GC methods in these matrices.[30,64,65] Only recently has ICE been applied successfully to the analysis of these organic acids in landfill leachate.[56] Samples are usually stored at −18 °C directly after sampling. Sample preparation consists of centrifugation, filtration and removal of humin-like substances by adsorption onto polyvinylpyrrolidine cartridges. Two different ICE methods have been developed allowing for the determination of organic acids in the samples so prepared. These ICE methods analyze for a wider range of organic acids than the GC methods, hence giving a more complete picture of the organic acid composition in these matrices.

6 OTHER COMPLEMENTARY TECHNIQUES

A variety of other techniques beside ion chromatography are used to determine organic acids in various matrices. Only a few methods, which are either of specific interest or very commonly used, will be mentioned.

One of the most frequently used techniques is GC. Determination of VFAs, usually in wastewater but also in other matrices, is often accomplished by direct aqueous injection of samples after minimal sample preparation onto specialty GC columns.[30,64–66] Typically, only VFAs with a chain length from C_2 to C_5 are analyzed by this method. If the gas chromatograph is maintained properly these direct GC injections deliver fast results, which are utilized as operational parameters in wastewater treatment. Nevertheless, direct GC injection methods are only applicable to samples with high organic acid concentrations (milligrams per liter), since they lack the sensitivity necessary to measure low organic acid concentrations without further extensive sample preparation. In addition, acids such as formic acid or oxalic acid can not be measured by direct GC injection.

Other GC methods focus on the analysis of specific types of organic acid. These methods utilize more

complex sample preparations resulting in methods with high sensitivities and, hence, low DLs.[21–23,54,43–46] They usually consist of enrichment, derivatization and GC measurement. In fact, even aqueous derivatization may be applied which is the case when determining ketoacids in drinking water.[21–23] Organic acids determined by this type of GC method include monocarboxylic acids, ketoacids and dicarboxylic acids, often in rain or air matrices.

CE has gained considerable interest in environmental analysis. Advantages of this technique include very fast analysis times, high resolution and the necessity of using only small sample volumes. However, matrix-dependent changes in the migration times of compounds make peak identification difficult even when utilizing internal standards. More detailed information about CE can be found in various books.[67–69] CZE is the CE technique best suited for the analysis of small ions. One example of applying CZE to the analysis of environmental samples is the quantification of organic acids in single raindrops.[47,52]

ABBREVIATIONS AND ACRONYMS

CE	Capillary Electrophoresis
CZE	Capillary Zone Electrophoresis
DL	Detection Limit
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
IC	Ion-exchange Chromatography
ICE	Ion Exclusion Chromatography
MRL	Minimum Reporting Level
MSE	Methanesulfonic Acid
OC	Organic Carbon
PAN	Peroxyacetylnitrate
PFBA	Perfluorobutyric acid
TOC	Total Organic Carbon
UV	Ultraviolet
VFA	Volatile Fatty Acid

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis • Sampling Considerations for Biomonitoring

Forensic Science **(Volume 5)**
Capillary Ion Electrophoresis in Forensic Science

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Ion Chromatography

REFERENCES

1. K. Ohta, K. Tanaka, P.R. Haddad, 'Ion-exclusion Chromatography of Aliphatic Carboxylic Acids on an Unmodified Silica Gel Column', *J. Chromatogr. A*, **739**, 359–365 (1996).
2. R.C. Weast, (ed.), *CRC Handbook of Chemistry and Physics*, 57th edition, CRC Press, Cleveland, OH, 1976–1977.
3. S. Budavari, (ed.), *Merck Index*, 11th edition, Merck, Rathway, NJ, 1989.
4. J.N. Galloway, G.E. Likens, E.S. Edgerton, 'Acid Precipitation in the Northeastern United States: pH and Acidity', *Science*, **194**, 722–723 (1976).
5. W.C. Keene, J.N. Galloway, 'Considerations Regarding Sources for Formic and Acetic Acids in the Troposphere' *J.Geophys. Res.*, **91**, 14 466–14 474 (1986).
6. K. Kawamura, I.R. Kaplan, 'Biogenic and Anthropogenic Organic Compounds in Rain and Snow Samples Collected in Southern California', *Atmos. Environ.*, **20**(1), 115–124 (1986).
7. K. Kawamura, I.R. Kaplan, 'Motor Exhaust Emissions as a Primary Source for Dicarboxylic Acids in Los Angeles Ambient Air', *Environ. Sci. Technol.*, **21**, 105–110 (1987).
8. R.W. Talbot, K.M. Becher, R.C. Harris, W.R. Cofer III, 'Atmospheric Geochemistry of Formic and Acetic Acids at a Mid-latitude Temperate Site', *J. Geophys. Res.*, **93**, 1638–1652 (1988).
9. G. Helas, H. Bingemer, M.O. Andreae, 'Organic Acids Over Equatorial Africa: Results From DECAFE 88', *J. Geophys. Res.*, **97**, 6187–6193 (1992).
10. M. Legrand, M. De Angelis, 'Origins and Variations of Light Carboxylic Acids in Polar Precipitation', *J. Geophys. Res.*, **100**(D1), 1445–1462 (1995).
11. C.G. Nolte, P.A. Solomon, T. Fall, L.G. Salmon, G.R. Cass, 'Seasonal and Spatial Characteristics of Formic and Acetic Acids Concentrations in the Southern California Atmosphere', *Environ. Sci. Technol.*, **31**(9), 2547–2553 (1997).
12. R.W. Talbot, B.W. Mosher, B.G. Heikes, D.J. Jacob, J.W. Munger, B.C. Daube, W.C. Keene, J.R. Maben, R.S. Artz, 'Carboxylic Acids in the Rural Continental Atmosphere Over the Eastern United States During the Shenandoah Cloud and Photochemistry Experiment', *J. Geophys. Res.*, **100**(D5), 9335–9343 (1995).
13. W.C. Keene, J.N. Galloway, 'Organic Acidity in Precipitation of North America', *Atmos. Environ.*, **18**(11), 2491–2497 (1984).

14. H. Sakugawa, I.R. Kaplan, L.S. Shepard, 'Measurements of H_2O_2, Aldehydes and Organic Acids in Los Angeles Rainwater: Their Sources and Deposition Rates', *Atmos. Environ.*, **27B**(2), 203–219 (1993).
15. I.M. Cozzarelli, J.S. Herman, M.J. Baedecker, 'Fate of Microbial Metabolites of Hydrocarbons in a Coastal Plain Aquifer: The Role of Electron Acceptors', *Environ. Sci. Technol.*, **29**(2), 458–469 (1995).
16. I.M. Cozzarelli, M.J. Baedecker, R.P. Eganhouse, D.F. Goerlitz, 'The Geochemical Evolution of Low-molecular-weight Organic Acids Derived from the Degradation of Petroleum Contaminants in Groundwater', *Geochim. Cosmochim. Acta*, **58**(2), 863–877 (1994).
17. P.B. McMahon, D.A. Vroblesky, P.M. Bradley, F.H. Chapelle, C.D. Gullett, 'Evidence for Enhanced Mineral Dissolution in Organic Acid-rich Shallow Ground Water', *Ground Water*, **33**(2), 207–216 (1995).
18. R.K. Stoessel, E.D. Pittman, 'Secondary Porosity Revisited: The Chemistry of Feldspar Dissolution by Carboxylic Acids and Anions', *Am. Assoc. Petrol. Geol.*, **74**, 1795–1805 (1990).
19. R.C. Surdam, D.B. MacGowan, 'Oilfield Waters and Sandstone Diagenesis', *Appl. Geochem.*, **2**, 613–619 (1987).
20. J.M. Montgomery, *Water Treatment: Principle and Design*, John Wiley & Sons, New York, 1985.
21. Y. Xie, D.A. Reckhow, 'A New Class of Ozonation-byproducts: The Ketoacids', Proceedings of the American Water Works Association Annual Conference, Vancouver, Canada, 251–265, 1992.
22. Y. Xie, D.A. Reckhow, 'Research Note: Formation of Ketoacids in Ozonated Drinking Water', *Ozone Sci. Eng.*, **14**, 269–275 (1992).
23. R.M. Le Lacheur, L.B. Sonnenberg, P.C. Singer, R.F. Christman, M.J. Charles, 'Identification of Carbonyl Compounds in Environmental Samples', *Environ. Sci. Technol.*, **27**, 2745–2753 (1993).
24. C.-Y. Kuo, H.-C. Wang, S.W. Krasner, M.K. Davis, 'Ion Chromatographic (IC) Determination of 3 Short-chain Carboxylic Acids in Ozonated Drinking Water', *ACS Symposium Series*, **649**, 350–365 (1996).
25. G.A. Gagnon, S.D.J. Booth, S. Peldszus, D. Mutti, F. Smith, P.M. Huck, 'Carboxylic Acids: Formation and Removal in Full-Scale Plants', *J. Am. Water Works Assoc.*, **89**(8), 88–97 (1997).
26. C.-Y. Kuo, 'Improved Application of Ion Chromatographic Determination of Carboxylic Acids in Ozonated Drinking Water', *J. Chromatogr. A*, **804**, 265–272 (1998).
27. L.S. Clesceri, A.E. Greenberg, A.D. Eaton, (eds.), 'Method 5560 Volatile Organic Acids', *Standard Methods for the Examination of Water and Wastewater*, 5th edition, American Public Health Association, Water Environment Federation, American Water Works Association, 48–50, Vol. 5, 1995.
28. C.N. Sawyer, P.L. McCarty, G.F. Parkin, *Chemistry for Environmental Engineering*, McGraw Hill, New York, 1994.
29. Metcalf, Eddy, *Wastewater Engineering – Treatment/Disposal/Reuse*, 3rd edition, Revised by G. Tschobanoglous, F.L. Burton, McGraw Hill, New York, 1991.
30. B.K. Ahring, M. Sandberg, I. Angelidaki, 'Volatile Fatty Acids as Indicators of Process Imbalance in Anaerobic Digesters', *Appl. Microbiol. Biotechnol.*, **43**, 559–565 (1995).
31. E.A. McBean, F.A. Rovers, G.F. Farquhar, *Solid Waste Landfill Engineering and Design*, Prentice Hall PTR, New Jersey, 1995.
32. J. Weiss, *Ion Chromatography*, 2nd edition, VCH, Weinheim, 1995.
33. P.R. Haddad, P.E. Jackson, *Ion Chromatography: Principles and Applications*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 46, 1990.
34. W.C. Keene, R.W. Talbot, M.O. Andreae, K. Beecher, H. Berresheim, M. Castro, J.C. Farmer, J.N. Galloway, M.R. Hoffmann, S.-M. Li, J.R. Maben, J.W. Munger, R.B. Norton, A.A.P. Pszenny, H. Puxbaum, H. Westberg, W. Winiwarter, 'An Intercomparison of Measurement Systems for Vapors and Particulate Phase Concentrations of Formic and Acetic Acids', *J. Geophys. Res.*, **94**(D5), 6457–6471 (1989).
35. W.R. Cofer III, G.C. Collins, R.W. Talbot, 'Improved Aqueous Scrubber for Collection of Soluble Atmospheric Trace Gases', *Environ. Sci. Technol.*, **19**, 557–560 (1985).
36. D. Grosjean, 'Organic Acids in Southern California Air: Ambient Concentrations, Mobile Source Emissions, In Situ Formation and Removal Processes', *Environ. Sci. Technol.*, **23**, 1637–1648 (1989).
37. W. Winiwarter, H. Puxbaum, S. Fuzzi, M.C. Facchini, G. Orsi, N. Beltz, K. Enderle, W. Jaeschke, 'Organic Acid Gas and Liquid-phase Measurements in Po-Valley Fall – Winter Conditions in the Presence of Fog', *Tellus*, **40B**, 348–357 (1988).
38. J.N. Galloway, G.E. Likens, W.C. Keene, J.M. Miller, 'The Composition of Precipitation in Remote Areas of the World', *J. Geophys. Res.*, **87**(11), 8771–8786 (1982).
39. W.C. Keene, J.N. Galloway, J.D. Holden, 'Measurement of Weak Organic Acidity in Precipitation From Remote Areas in the World', *J. Geophys. Res.*, **88**(9), 5122–5130 (1983).
40. P. Hoffmann, V.K. Karandashev, T. Sinner, H.M. Ortner, 'Chemical Analysis of Rain and Snow Samples from Chernoglovka/Russia by IC/TXRF and ICP/MS', *Fresenius' J. Anal. Chem.*, **357**, 1142–1148 (1997).
41. A. Amman, T.B. Rüttimann, 'Simultaneous Determination of Small Organic and Inorganic Anions in Environmental Water Samples by Ion-exchange Chromatography', *J. Chromatogr. A*, **706**, 259–269 (1995).
42. J.A. Morales, H.L. de Medina, M.G. de Nava, H. Velasquez, M. Santana, 'Determination of Organic Acids

by Ion Chromatography in Rain Water in the State of Zulia, Venezuela', *J. Chromatogr. A*, **671**, 193–196 (1994).
43. K. Kawamura, 'Identification of C_2-C_{10} w-Oxocarboxylic Acids, Pyruvic Acid, and C_2-C_3 α-Dicarbonyls in Wet Precipitation and Aerosol Samples by Capillary GC and GC/MS', *Anal. Chem.*, **65**, 3505–3511 (1993).
44. K. Kawamura, I.R. Kaplan, 'Capillary Gas Chromatography of Volatile Organic Acids in Rain and Fog Samples', *Anal. Chem.*, **56**, 1616–1620 (1984).
45. K. Kawamura, S. Steinberg, I.R. Kaplan, 'Capillary GC Determination of Short-chain Dicarboxylic Acids in Rain, Fog and Mist', *Int. J. Environ. Anal. Chem.*, **19**, 175–188 (1985).
46. S. Steinberg, K. Kawamura, I.R. Kaplan, 'The Determination of α-Ketoacids and Oxalic Acid in Rain, Fog and Mist by HPLC', *Int. J. Environ. Anal. Chem.*, **19**, 251–260 (1985).
47. K. Bächmannn, I. Haag, T. Prokop, A. Röder, P. Wagner, 'Chromatographic Methods for the Analysis of Size-classified and Individual Rain-drops', *J. Chromatogr.*, **643**, 181–188 (1993).
48. M. Legrand, M. De Angelis, F. Maupetit, 'Field Investigation of Major and Minor Ions Along Summit (Central Greenland) Ice Cores by Ion Chromatography', *J. Chromatogr.*, **640**, 251–258 (1993).
49. J.E. Lawrence, P. Koutrakis, 'Measurement of Atmospheric Formic and Acetic Acids: Methods Evaluation and Results from Field Studies', *Environ. Sci. Technol.*, **28**(5), 957–964 (1994).
50. U. Hofmann, D. Weller, C. Ammann, E. Jork, J. Kesselmeier, 'Cryogenic Trapping of Atmospheric Organic Acids under Laboratory and Field Conditions', *Atmos. Environ.*, **31**(9), 1275–1284 (1997).
51. K. Granby, C.S. Christensen, C. Lohse, 'Urban and Semirural Observations of Carboxylic Acids and Carbonyls', *Atmos. Environ.*, **31**(10), 1403–1415 (1997).
52. A. Röder, K. Bächmannn, 'Simultaneous Determination of Organic and Inorganic Anions in the Sub-µmol/l Range in Rain Water by Capillary Zone Electrophoresis', *J. Chromatogr.A*, **689**, 305–311 (1995).
53. S. Peldszus, P.M. Huck, S.A. Andrews, 'Determination of Carboxylic Acids in Drinking Water at Low µg/l Concentrations: Method Development and Application', Proceedings of American Water Works Association Water Quality Technology Conference, WQTC, Boston, USA, 1996.
54. K. Kawamura, H. Kasukabe, 'Source and Reaction Pathways of Dicarboxylic Acids, Ketoacids and Dicarbonyls in Arctic Aerosols: One Year of Observations', *Atmos. Environ.*, **30**(10/11), 1709–1722 (1996).
55. S. Peldszus, S.A. Andrews, P.M. Huck, 'Determination of Short-chain Aliphatic, Oxo- and Hydroxy-acids in Drinking Water at Low Microgram per Liter Concentrations', *J. Chromatogr. A*, **723**, 27–34 (1996).
56. K. Fischer, A. Chodura, J. Kotalik, D. Bienik, A. Kettrup, 'Analysis of Aliphatic Carboxylic Acids and Amino Acids in Effluents of Landfills, Composting Plants and Fermentation Plants by Ion-exclusion and Ion-exchange Chromatography', *J. Chromatogr.*, **770**, 229–241 (1997).
57. P.R. Haddad, P.E. Jackson, 'Studies on Sample Preconcentration in Ion Chromatography. VIII. Preconcentration of Carboxylic Acids Prior to Ion-exclusion Separation', *J. Chromatogr.*, **447**, 155–163 (1988).
58. P.M. Bradley, F.H. Chapelle, D.A. Vroblesky, 'Does Lead Affect Microbial Metabolism in Aquifer Sediments Under Different Terminal Electron Acceptors', *Geomicrobiol. J.*, **11**, 85–94 (1993).
59. S. Peldszus, P.M. Huck, S.A. Andrews, 'Quantitative Determination of Oxalate and Other Organic Acids in Drinking Water at Low µg/l Concentrations', *J. Chromatogr. A*, **793**, 198–203 (1998).
60. H.S. Weinberg, W.H. Glaze, 'A Unified Approach to the Analysis of Polar Organic By-products of Oxidation in Aqueous Matrices', *Water Res.*, **31**(7), 1555–1572 (1997).
61. S.R. Souza, M.F.M. Tavares, L.R.F. de Carvalho, 'Systematic Approach to the Separation of Mono- and Hydroxycarboxylic Acids in Environmental Samples by Ion Chromatography and Capillary Electrophoresis', *J. Chromatogr. A*, **796**, 335–346 (1998).
62. Dionex Application Note 101, 'Trace Level Determination of Bromate in Ozonated Drinking Water Using Ion Chromatography', Dionex Corp., Sunnyvale, CA, 1995.
63. R.W. Slingsby, C.A. Pohl, 'Approaches to Sample Preparation for Ion Chromatography, Sulfate Precipitation on Barium-form Ion Exchangers', *J. Chromatogr. A*, **739**, 49–55 (1996).
64. F.J. Duisterwinkel, B.G. Wolthers, W. van der Slik, J. Dankert, 'Determination of Volatile Fatty Acids by Gas Chromatography on a Capillary and Megabore Column', *Clin. Chim. Acta*, **156**, 207–214 (1986).
65. A.M. Eilersen, M. Henze, L. Kloft, 'Effect of Volatile Fatty Acids and Trimethylamine on Nitrification in Activated Sludge', *Wat. Res.*, **28**(6), 1329–1336 (1994).
66. C.A. Hordijk, I. Burgers, G.J.A. Phylipsen, T.E. Cappenberg, 'Trace Determination of Lower Volatile Fatty Acids in Sediments by GC with Chemically Bonded FFAP Columns', *J. Chromatogr.*, **511**, 317–323 (1990).
67. D.R. Baker, 'Capillary Electrophoresis', John Wiley & Sons, New York, 1995.
68. S.F.Y. Li, *Capillary Electrophoresis: Principles, Practice and Applications*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 52, 1992.
69. J.P. Landers (ed.), *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, FL, 1994.

Organic Analysis in Environmental Samples by Capillary Electrophoresis

Carmen García-Ruiz, Antonio L. Crego, and Maria L. Marina
Universidad de Alcalá, Madrid, Spain

1 Introduction **1**
2 Separation Modes for Environmental Analysis **2**
2.1 Capillary Zone Electrophoresis 2
2.2 Micellar Electrokinetic Chromatography 5
2.3 Other Kinds of Electrokinetic Chromatography 13
2.4 Capillary Electrochromatography 14
3 Analysis of Pollutants in Environmental Samples **14**
3.1 Sensitivity Enhancement Procedures 15
3.2 Analysis of Pollutants in Environmental Samples 22
4 Conclusions **24**
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **25**
References **26**

Capillary electrophoresis (CE) is an instrumental technique that employs narrow-bore (10–100 μm ID, internal diameter) capillaries under the influence of a high-gradient electric field to perform high-efficiency separations. Different working modes of CE have been described. Capillary zone electrophoresis (CZE) and micellar electrokinetic chromatography (MEKC) have become the most popular modes of CE in environmental applications. Other kinds of electrokinetic chromatography (EKC) and other modes of CE like capillary electrochromatography (CEC) have also been used.

CE techniques have been considered as a good alternative for analyzing pollutants because of their high performance and versatility related to selectivity and range of application. In addition, CE is a powerful tool in the field of chiral separations and it is considered one of the so-called clean analytical techniques. Although the main drawback of CE techniques in environmental analysis is their insufficient sensitivity for detecting trace-level pollutants, some procedures have been developed in order to overcome this problem.

1 INTRODUCTION

Chromatographic techniques are important tools for carrying out analysis of environmental pollutants. Gas chromatography (GC) and high-performance liquid chromatography (HPLC) are the most common techniques used routinely and in complementary form in this type of analysis. Nevertheless, another instrumental technique has been developed as an interesting alternative to HPLC and GC techniques, that is, CE. The utilization of capillary tubes of ID generally ranging from 10 to 100 μm to achieve electrophoretic separations has enabled high efficiencies and an enormous resolution power to be obtained.[1,2]

Nowadays, the use of CE is one of the major trends in analytical chemistry and the number of publications has increased exponentially.[3] The limitations that GC techniques present for separating nonvolatile compounds (polar, acidic or basic, and thermally labile compounds) and the poor performance that sometimes characterizes HPLC techniques causing loss in resolution and sensitivity, have led to consideration of CE techniques as a good alternative for analyzing pollutants unsuitable for GC and affected by the poor efficiency of HPLC. In fact, they combine the high performance of GC techniques with the versatility of HPLC techniques with respect to selectivity and range of application.[4]

The importance of chirality in environmental control (enantiomers may have a different toxicity or one of them may not be toxic)[5] has given rise to the search for new and effective means of resolving enantiomeric mixtures.[6] The extent to which various chemicals are being applied in agriculture (many pesticides are produced as racemic mixtures and often only one enantiomer is biologically active) will soon require such potential ecological hazards to be monitored enantioselectively[7] and CE can be considered as a powerful separation tool in the field of chiral separations.[8] Aside from all other considerations mentioned here which show the great potential of CE, it can also be considered to be one of the so-called *clean* analytical techniques. This is due to the small volumes of mobile phase (electrolyte solution) and sample required to perform separations by this technique. This property confers great environmental interest on CE.

Despite the impressive potential for CE to achieve analytical separations, this technique has a major drawback: CE is not sensitive enough to detect trace-level environmental pollutants. Actually, government regulations require determination of lower and lower levels of pollutants in environmental samples so that the concentration

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

sensitivity of commercial CE instruments is far from ideal. In order to overcome this problem, some procedures to enhance CE sensitivity have been developed and are described below.

2 SEPARATION MODES FOR ENVIRONMENTAL ANALYSIS

Environmental pollutants are organic compounds from ionizable compounds, moderately hydrophobic compounds to highly hydrophobic compounds. Correspondingly, several working modes are needed, and some parameters must be taken into account such as electrolytic solution composition, capillary dimensions, capillary temperature, applied voltage and mode and time of injection. In order to optimize some of the above-mentioned parameters, some mathematical designs have been described.[9]

CZE and MEKC have become the most popular modes of CE in environmental applications.[10,11] Thus, their fundamentals will be described prior to outlining their potential for environmental analyses. Additionally, although less used in the environmental field, other kinds of EKC and other modes of CE like CEC, will be described.[11]

2.1 Capillary Zone Electrophoresis

2.1.1 Fundamentals

The most practical and simple working mode in CE is CZE. Separation by CZE is carried out in a capillary filled with a continuous background electrolyte (buffer).[12–14] The direction and migration velocity of analytes are determined by two phenomena, electroosmosis and electrophoresis.[15] Electroosmosis originates from ionized silanol groups on the inner wall of the fused-silica capillary which attract positively charged buffer ions to form an electrical double layer. The buffer ions in the mobile region of the double layer will migrate toward the cathode (negative pole) as soon as the voltage has been switched on and induce electroosmotic flow (EOF) of the entire liquid in the capillary. On the other hand, analytes are separated on the basis of their different electrophoretic migration velocities, which are related to their charge densities (mainly based on differences in solute size and charge at a given pH). Generally, the EOF will be higher than the electrophoretic migration velocity of most anionic analytes (which tend to migrate toward the positive pole). Consequently, both cations and anions will migrate in the same direction and can be separated in the same run, eluting initially the positively charged substances, the neutral ones (not separated), and finally the negatively charged solutes. This simplicity has contributed to the popularity of CZE.

CZE can only be used for the separation of ionic or ionizable compounds with low molecular weight and sufficient water solubility. These types of analyte are normally not very amenable to GC requiring, as a minimum, a sample derivatization which presents long sample preparation times and incomplete recoveries for many analyte derivatives. Therefore, the major competition for CZE techniques comes from the HPLC field. However, the higher efficiency of CZE and its different selectivity makes CZE complementary to HPLC. When comparing similar applications, CZE often requires a shorter analysis time and less effort in method development, owing to its huge separation power and flexibility in selectivity tuning. In addition, applications on CZE instruments are much more cost-effective than on HPLC, requiring inexpensive columns and a few milliliters of solutions.

The most important experimental parameter in CZE separations is the pH, except when completely ionized ions are separated. For this reason, some mathematical models have been developed for theoretical prediction of resolution as a function of the pH of the buffer.[16] Other less important parameters are choice of the buffer and its concentration and applied voltage. Unfortunately, many environmental pollutants are related isomers or have very close chemical structures, and therefore CZE often cannot provide sufficient resolution. To improve the separation, CZE has been used with additives such as organic solvents,[17] neutral cyclodextrins (CDs)[18] or neutral polymers.[19,20] Organic solvents are used as EOF and electrophoretic mobility modifiers to enable the separation of isomers or compounds with closed structures. The use of CDs in CZE has been recognized as a new kind of CZE named cyclodextrin-modified capillary zone electrophoresis (CDCZE). CDs are toroidal, with an axial void–apolar cavity and an outer hydrophilic surface, which form CD-inclusion complexes with certain molecules, based on the concordance of the solute molecular size with the cavity diameter of CD in addition to the hydrophobicity and chemistry.[21] Neutral CDs, despite migrating with the EOF, can form inclusion complexes with molecular compounds which have specific structural arrangements, modifying their electrophoretic mobilities and allowing the separation of pairs of enantiomers or improving multicomponent separation. On the other hand, neutral polymers interact with some analytes decreasing their electrophoretic mobilities and improving the separation.

Recently, cations like cetyltrimethylammonium bromide (CTAB), hexadimetrine bromide (HDB), and other alkylammonium salts have been used to produce a reversal of the EOF and to reduce analysis time of the anionic analytes which migrate in the same direction

as EOF. This principle is known as coelectroosmotic CE or reversed-polarity CE.[22–24]

2.1.2 *Applications to the Analysis of Compounds of Environmental Interest*

According to the principles described before, CZE can only be employed to analyze ionic or ionizable pollutants such as pesticides, plant growth regulators, phenols, amines, aromatic sulfonates, alkylsulfates, organometallic compounds and dyes. Table 1 summarizes the analytical conditions used to separate these compounds. These separations have for the most part been monitored by using UV (ultraviolet) absorbance detectors generally used in CE.

Pesticides (e.g. herbicides, fungicides, insecticides) include some compounds that persist in soils or aquatic environments and can be toxic to humans. Ionic and ionizable pesticides can be separated by CZE at the appropriate pH using acetate or phosphate buffers at pH between 3 and 8. In many cases, these compounds have been separated using only a buffer without additives. Examples are paraquat and diquat,[25] sulfonylureas,[26]

Table 1 Compounds of environmental interest analyzed by CZE

Compounds	Separation media	Detection	Application	Ref.
Pesticides	*Buffer*: phosphate (pH 3.5)	UV–254 nm	Paraquat and diquat	25
	Buffer: acetate (pH 5) *Additive*: acetonitrile	MS	Sulfonylureas	26
	Buffer: acetate (pH 4.8)	UV–200 nm	Phenoxy acids	27
	Buffer: borate (pH 6)	Indirect UV 200 nm	Alkylphosphonic acids	28
	Buffer: phosphate (pH 6) *Additive*: acetonitrile	UV–220 nm	Triazines	29
	Buffer: acetate (pH 4.8) *Additives*: heptakis(2,6-di-*o*-methyl)-β-CD and γ-CD	UV–200 nm		27
	Buffer: acetate (pH 4.45) *Additive*: TM-β-CD	UV–230 nm	Enantiomeric separation of optically active phenoxy acids	30
	Buffer: phosphate/borate (pH 5) *Additives*: β-CD, TM-β-CD alone or mixed	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 420$ nm		31
Plant growth regulators	*Buffer*: phosphate/borate (pH 7.5/8.1) *Additives*: α-CD, β-CD and γ-CD	Diode-array 196–210 nm	Three indole, three chlorophenoxyacetic and two naphthaleneacetic acids	32 33
Phenols	*Buffer*: phosphate (pH 6.9)	UV–210 nm	Chlorophenols	34
	Buffer: phosphate/borate (pH 7.7)	UV–215 nm	Dichlorophenol isomers	35
	Buffer: phosphate/borate (pH 9.8)	UV–210 nm	USEPA priority phenols	34
	Buffer: phosphate (pH 10.5)		Substituted phenols	36
	Buffer: phosphate/borate (pH 11) *Additives*: HDB and 2-propanol	UV–254 nm	Substituted phenols	22 23
Amines	*Buffer*: TRIS/acetate (pH 4.8)	UV–200 nm	Phenylenediamine isomers	37
	Buffer: acetate (pH 3.75)	UV–214 nm	Chloroaniline isomers	38
	Buffer: phosphate/citrate (pH 2) *Additives*: NaCl, methanol	Diode-array 190–263 nm	12 heterocyclic amines	11
Aromatic sulfonates	*Buffer*: borate (pH 8.3)	UV–214 nm	Eight aromatic sulfonates	39
	Buffer: borate (pH 9) *Additive*: acetonitrile		C_2–C_{12} linear alkylbenzene sulfonates	40
Alkylsulfates	*Buffer*: *p*-toluensulfonate *Additives*: methanol/ acetonitrile	Indirect UV 214 nm	C_9–C_{12} alkylsulfates	41

Table 1 (*continued*)

Compounds	Separation media	Detection	Application	Ref.
Organo-metallics	*Buffer*: pyridine (pH 2.5) *Additive*: CTAB	Indirect UV 254 nm	Alkyltin compounds	42
	Buffer: tartaric acid (pH 2.6) *Additives*: methanol, BTMA	Indirect UV 210 nm	Alkyltin compounds	43
	Buffer: borate (pH 8.24)	Diode-array 200–400 nm	Methylmercury	44
Dyes	*Buffer*: borate (pH 7.5) *Additive*: β-CD	UV–245 nm	Dyestuffs	45
	Buffer: TES/imidazole (pH 6.8) *Additives*: PEG, β-CD	UV–254 nm	Food colorants	46
	Buffer: BTP with HCl (pH 6.5) *Additives*: PEG and PVP	UV–214 nm	Nine synthetic organic dyes	20
	Buffer: citrate (pH 3.25) *Additives*: acetonitrile and CTAB	Diode-array 400–610 nm	Six reactive dyes	24

MS, mass spectrometry; LIF, laser induced fluorescence; TRIS, tris(hydroxymethyl)aminomethane; BTMA, benzyltrimethylammonium; TES, *N*-tris(hydroxymethyl)methyl-2-aminoethane sulfonic acid; PEG, polyethylene glycol; BTP, bis-tris-propane; PVP, polyvinylpyrrolidone; USEPA, United States Environmental Protection Agency.

phenoxy acids[27] and alkylphosphonic acids.[28] In other cases, the CZE buffer contains additives such as organic solvents or CDs. Thus, prometon and prometryne (triazine herbicides) have been separated with a neutral buffer containing acetonitrile.[29] On the other hand, it is important to note that the addition of CDs to the separation buffer allows enantiomeric separation of several phenoxy acid herbicides.[27,30,31] Figure 1 shows the chiral separation of three optically active phenoxy acid herbicides in their six enantiomers by CZE using an acid buffer with tri-*o*-methyl-β-cyclodextrin (TM-β-CD) as additive.[30] They were separated from another nonchiral phenoxy acid at microgram per milliliter levels in an analysis time close to 12 min.

Plant growth regulators have been easily separated using a basic buffer with a cocktail of α-, β- and γ-CDs in order to enhance the selectivity.[32,33]

Phenols are important environmental pollutants owing to their high toxicity even at low concentrations. They are weak acids which can ionize in aqueous solutions and then can be analyzed easily by CZE using a buffer without additives. The pH chosen for the buffer depends on the types of compound, using phosphate or phosphate/borate buffers for neutral or basic pH values, respectively. Thus, the analysis of several chlorophenols requires a neutral pH,[34,35] although the eleven priority pollutant phenols listed by the USEPA (including chlorophenols, alkylphenols and nitrophenols)[34] and eight substituted phenols (different methyl-, dimethyl- and trimethylphenols)[36] have been separated using basic conditions. The 11 USEPA priority phenols can be completely resolved in less than 15 min by CZE (see Figure 2).[34] In this case, CZE provided better results than MEKC, because the analysis time was noticeably shorter compared with the 45 min obtained with MEKC,[47] or the 25 min typically required by HPLC.[48] Optimum conditions included a small buffer concentration and a basic pH, under which all phenols except one were ionized. Detection was performed with an on-column UV detector and good linearities ($r \geq 0.999$) were obtained for concentrations up to at least 50 ppm, with detection limits less than 1 ppm. On the other hand, the rapid analysis of substituted phenols has been achieved by using a buffer with a high pH value and HDB with propanol as additives.[22,23] In this method, the principle of coelectroosmotic CE, described above, was used. The application of CE techniques to the analysis of phenols has been reviewed.[49]

Amines such as heterocyclic amines are carcinogenic and mutagenic compounds that are found in cooked foods. Others like aromatic amines can contaminate water and soils. Based on their positive charge in acidic solution, several phenylenediamine isomers[37] and chloroaniline isomers[38] have been separated by CZE easily using only an acid buffer. On the other hand, the separation of 12 heterocyclic amines was optimized by orthogonal array design and the optimum conditions included the use of an acid buffer with NaCl and methanol as additives.[11]

Aromatic sulfonates are compounds separated by CZE using only basic buffers.[39] Alkylbenzene sulfonates are anionic surfactants widely used in detergent formulations, which have been separated by CZE using basic buffers with acetonitrile as additive.[40] The results show that CZE is more selective than MEKC for these compounds and the resulting analysis time is at least three times shorter compared with HPLC.[50]

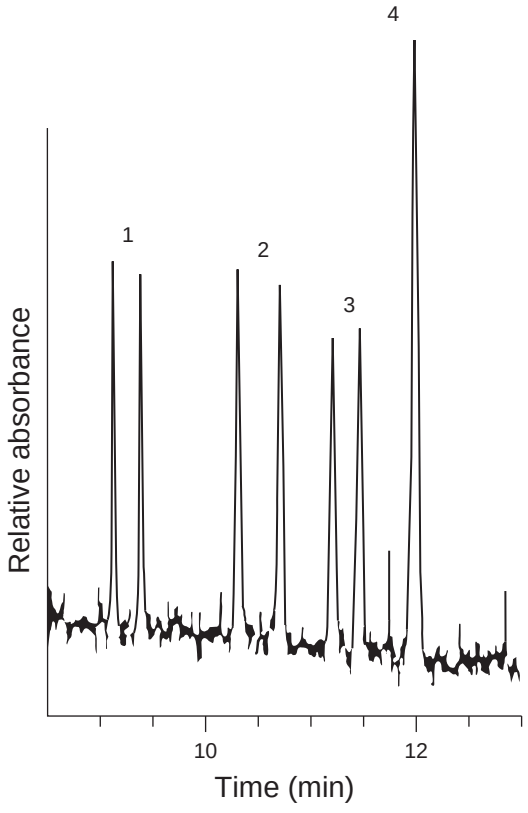

Figure 1 Chiral separation of three optically active phenoxy acid herbicides and another nonchiral phenoxy acid by CDCZE. Conditions: separation media, 0.05 M acetate buffer (pH = 4.45) with 0.025 M TM-β-CD; capillary tube, 75 μm ID × 50 cm (to the detector); hydrodynamic injection, 5 s; applied voltage (current), 20 kV (41 μA); UV-detection at 230 nm; temperature, 30 °C. Peaks (analyte concentration, 1 μg mL^{-1}): (1) 2-(2,4,5-trichlorophenoxy)propionic acid, (2) 2-(2-methyl-4-chlorophenoxy)propionic acid, (3) 2-(2,4-dichlorophenoxy)propionic acid and (4) 2,4-dichlorophenoxyacetic acid. (Reproduced by permission of Elsevier Science Publishers. Copyright (1994), from A.W. Garrison, P. Schmitt, A. Kettrup, *J. Chromatogr. A*, **688**, 317–327.[30])

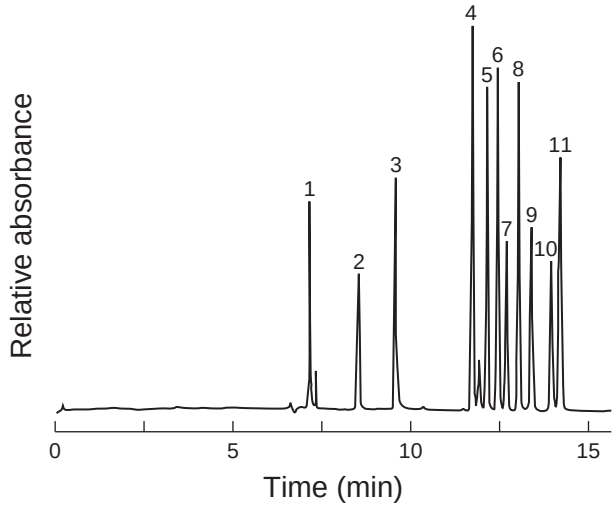

Figure 2 Electropherogram corresponding to a mixture of 11 USEPA priority phenols separated by CZE. Conditions: separation media, 0.010 M phosphate–borate buffer (pH 9.8); capillary tube, 75 μm ID × 100 cm (65 cm to the detector); injection by vacuum, 10 s; applied voltage (current), 22.5 kV (53 μA); UV-detection at 210 nm. Peaks (analyte concentration, 25 μg mL^{-1}): (1) 2,4-dimethylphenol, (2) phenol, (3) 4-chloro-3-methylphenol, (4) pentachlorophenol, (5) 2,4,6-trichlorophenol, (6) 2,4-dichlorophenol, (7) 2-methyl-4,6-dinitrophenol, (8) 2-chlorophenol, (9) 2,4-dinitrophenol, (10) 4-nitrophenol and (11) 2-nitrophenol. (Reproduced by permission of Elsevier Science Publishers. Copyright (1995) from G. Li, C. Locke, *J. Chromatogr.*, **669**, 93–102.[34])

Separation of C_9–C_{12} sodium *alkylsulfates* was followed using indirect UV detection because there is no chromophore in the analyte molecules.[41]

Organometallic compounds such as organotin compounds have been analyzed by CZE employing an acid buffer and pyridine[42] or BTMA[43] to enable indirect UV detection. Methylmercury has been separated from an excess of cysteine at basic pH, using different reference materials.[44]

Dyes are colorants employed in the food and textile industries which frequently contaminate the aquatic environment. Some of them are anionic compounds suitable for analysis by CZE. Thus, seven food additive dyestuffs[45] and eleven permitted food colorants[46] have been separated using neutral buffers with CDs. Other synthetic dyes used as colorants for textiles have been separated using a neutral buffer with PEG and PVP as additives.[20] Reactive dyes have also been separated with an acid buffer and acetonitrile as additive.[24]

2.2 Micellar Electrokinetic Chromatography

2.2.1 Fundamentals

Even though CE was initially limited to the separation of charged compounds, the introduction of a micellar system in the separation buffer in 1984,[51] which resulted in the so-called MEKC, enabled CE to be applied to the separation of noncharged compounds.[52–56] In this technique, ionic surfactants are added to the CZE buffer at concentrations exceeding the critical micelle concentration (CMC) to form micelles which act as a pseudostationary phase.[57] The micelles are spherical aggregates whose hydrophobic groups are oriented toward the center of the micelle, with polar or charged groups along the sphere's surface. In this situation, neutral analytes are distributed between the buffer phase (which migrates with the velocity of the EOF) and the micellar phase (which moves at a velocity slower than EOF), resulting in separation. Because the micellar phase moves toward the detector, an elution window is created that

is bordered by a column void time (t_0, mobility of the EOF) and a micelle migration time (t_{MC}, mobility of the micelle). All the analytes must elute between these two limits, t_0 and t_{MC}, depending upon their partition between aqueous and micellar phases.[58] On the other hand, the migration behavior of ionizable solutes is more complicated because their electrophoretic mobilities have a great influence on the migration behavior and selectivity.[47]

Unfortunately, highly hydrophobic compounds tend completely to associate with the micelles and are not resolved because they coelute near t_{MC}. In order to enhance the usefulness of this technique, organic solvents (methanol, acetonitrile, tetrahydrofuran, or acetone),[59–61] and urea[62] were added to the buffer. In this case, the separation selectivity in MEKC can be controlled through the nature and concentration of the buffer and surfactant in the electrolyte solution and the nature and concentration of organic additives.[63] As an alternative to the addition of organic solvents (aggregation numbers and CMCs are affected by the addition of small amounts of organic modifier, and micelles are generally not stable in buffer with a high organic solvent content),[64,65] neutral CDs[66–71] are extensively used as modifiers in MEKC, giving rise to cyclodextrin-modified micellar electrokinetic chromatography (CDMEKC). In CDMEKC, neutral CDs will not interact with the micelle owing to its hydrophilic outside surface, but a hydrophobic solute is partitioned between the micelle (which migrates with the micellar velocity) and CD (which migrates with the EOF). Accordingly, the differential partition of the solute between CD-monomers and the micelles allows a separation.

The most widely used surfactants in MEKC have been those which are anionic in nature, especially sodium dodecyl sulfate (SDS) alone or with the additives previously described. Other possibilities, such as mixed micelles between SDS and neutral surfactants (Brij 35 (polyoxyethylenelauryl ether)[72] or PEG 400 monolaurate[73,74] or SDS homologs (sodium tetradecyl[52] or decyl[75] sulfates)) were tested to alter resolution and selectivity, but narrower retention windows or irreproducible times were obtained.

Bile salts, especially sodium cholate (SC) and sodium deoxycholate (SDC), have been considered promising pseudostationary phases in MEKC. They are important biological surfactants, based on a hydroxy-substituted steroid backbone, which form helical micelles with the hydrophobic portions of the monomer facing the aqueous solution while the hydrophilic portions turn inward. Thus, in some respects, bile salt micelles can be envisioned as ''inverted micelles''.[76,77] Their structure and aggregation behavior allow chiral recognition[78–80] and reduction of retention factor values with respect to those obtained with SDS.[81,82] In addition, bile salt micelles appear to tolerate high concentrations of organic solvents without drastic loss of efficiency or dramatic increase in analysis time.[65] These factors facilitate the analysis of hydrophobic compounds more easily than with SDS.

On the other hand, in situ charged micelles[83–86] are essentially anionic complexes formed between alkyl- or steroidal-glycoside surfactants and borate anions. They possess many unique and attractive features:

1. The surface charge density and the elution window increase with an increase in pH (upper 8) or borate concentration (25–400 mM).
2. Owing to the more balanced hydrophilic–lipophilic character than alkyl ionic surfactants, they are useful for separating hydrophobic compounds.
3. Some in situ charged micelles have chiral selectivity.[87]

Finally, although employed less in the analysis of environmental pollutants, other compounds have been used as micellar phases in MEKC:

1. different micelles such as alkyltrimethylammonium,[88] especially CTAB, disodium 5,12-bis-(dodecyloxymethyl)-4,7,10,13-(tetraosa)-1,16-hexadecanedisulfonate (DBTD)[89] and butyl acrylate–butyl methacrylate–methacrylic acid copolymers sodium salts (BBMA);[90]
2. oligomers with micelle-like structure such as sodium 10-undecylenate (SUA)[91] and sodium 10-undecylsulfate (SUS);[92]
3. macromolecules that are considered to mimic micelles such as dendrimers[93–95] and resorcarenes.[96]

MEKC systems compared with CZE are more complicated because of the additional experimental parameters that must be optimized[97] and because their electrolytic solutions are less stable than CZE buffer solutions owing to the effect of temperature on the distribution equilibria involved. In addition, the theoretical plate number in MEKC is not usually as high as could be achieved by CZE owing to the mass transfer resistance introduced by solute partitioning between the two phases. Consequently, sensitivity in MEKC is expected to be lower than in CZE.[98] Despite these drawbacks, MEKC has extended the enormous power of CZE to the separation of uncharged solutes and, moreover, it has demonstrated additional selectivity. This offers a good alternative for pollutants that are not amenable to CZE because they are electrically neutral solutes.

2.2.2 *Applications to the Analysis of Compounds of Environmental Interest*

MEKC is a mode of CE suitable to analyze neutral environmental pollutants such as mycotoxins, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), polychlorodibenzo-*p*-dioxins (PCDDs), or phthalates, and ionizable or ionic contaminants such as pesticides, phenols, dyes, amines, nitroaromatic compounds, organometallic compounds and sulfonates. The experimental conditions used to separate these compounds are grouped in Table 2. MEKC has been performed with several detection systems, such as UV absorbance, LIF, amperometric detection, and MS.

Mycotoxins are produced by fungi and they constitute a common class of environmental pollutants.[150] Conventional methods for mycotoxins analysis are thin-layer chromatography and HPLC;[151] however, these methods are very time-consuming and labor-intensive. These compounds, which cannot be analyzed by CZE, have been successfully resolved using MEKC. Three aflatoxins (mycotoxins that occur in a variety of foodstuffs intended for both livestock and human consumption) were separated by MEKC using SDC as surfactant in about 15 min.[65] The separation of 10 mycotoxins was optimized (fully resolved chromatograms in only 15 min) using two different MEKC systems, SDS with acetonitrile

Table 2 Compounds of environmental interest analyzed by MEKC

Compounds	Separation media	Detection	Application	Refs.
Mycotoxins	*Buffer*: phosphate/borate *Micelle*: SDC *Additive*: acetonitrile	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 450$ nm	Three aflatoxins	65
	Buffer: phosphate/borate (pH 9.3) *Micelle*: SDS or SDC *Additives*: acetonitrile or acetonitrile and γ-CD	UV–254 nm	Acidic and neutral mycotoxins (including four aflatoxins)	99
PAHs	*Buffer*: CHES (pH 10) *Micelle*: SDS *Additive*: propanol	UV–214 nm	Eight USEPA priority PAHs and 15 aromatic compounds	63
	Buffer: borate (pH 9) *Micelle*: SDS *Additives*: urea and γ-CD	UV–210/220/230/270 nm	The 16 USEPA priority PAHs and other homologs	66 70 100–102
	Buffer: borate (pH 9) *Micelle*: SDS *Additives*: urea, γ-CD and DMSO	UV–220/245 nm		103
	Buffer: borate (pH 9) *Micelle*: SDS *Additives*: urea and β-CD	UV–220 nm	Four USEPA priority PAHs and five aromatic compounds	67
	Buffer: borate (pH 7) *Micelle*: SDS *Additive*: γ-CD	UV–254 nm	Seven USEPA priority PAHs	69
	Buffer: phosphate/borate *Micelle*: SDS *Additives*: urea, γ-CD, (2-propanol)	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 400$ nm	Seven methyl-substituted benzo[*a*]pyrene isomers	71
	Buffer: borate (pH 9) *Micelle*: SDS *Additives*: urea, γ-CD and β-CD	UV–230 nm	10 hydroxylated homologs	104
	Buffer: phosphate/borate (pH 9) *Micelle*: SC or SDC *Additive*: methanol		Three USEPA priority PAHs	65
		LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 400$ nm	Four binaphthyl enantiomers	81
	Buffer: TRIS/phosphate *Micelle*: SDC *Additive*: DMF		Six USEPA priority PAHs	105

Table 2 (*continued*)

Compounds	Separation media	Detection	Application	Refs.
PAHs	*Buffer*: borate (pH 9.5) *Micelle*: CTAB *Additive*: acetonitrile	Diode-array 190–350 nm	Seven USEPA priority PAHs	106
	Buffer: borate (pH 10) *Micelle*: MEGA 10	UV–254 nm	Two USEPA priority PAHs and aromatic compounds	83
	Buffer: borate (pH 8.2) *Micelle*: SUA *Additive*: acetonitrile	UV–275 nm	11 USEPA priority PAHs	91
	Buffer: tetraborate *Micelle*: SUS *Additives*: acetonitrile or methanol	UV–214 nm	Six USEPA priority PAHs and six homologs	92
	Buffer: borate (pH 10.1) *Micelle*: SBD(X)-C_{12} *Additive*: methanol	UV–254 nm	Three USEPA priority PAHs and three homologs	95
	Micelle: resorcarene (pH 13.2) *Additives*: urea and acetonitrile	UV–260 nm	12 USEPA priority PAHs	96
	Buffer: borate (pH 9) *Micelle*: poly(D-SUV) *Additive*: γ-CD	UV–280 nm	Enantiomeric separation of naphthyl derivatives	107
PCDDs	*Buffer*: borate (pH 8/9) *Micelle*: SDS *Additives*: urea and γ-CD	UV–220 nm	TCDD isomers	66 108
PCBs	*Buffer*: borate (pH 8) *Micelle*: SDS *Additive*: urea and γ-CD	UV–220 nm	11 PCB congeners	66
	Buffer: CHES (pH 10) *Micelle*: SDS *Additives*: urea and γ-CD	UV–240 nm	14 PCB congeners	109
	Buffer: CHES (pH 10) *Micelle*: CTAB		Ten PCB congeners	110
	Buffer: CHES (pH 10) *Micelle*: SDS *Additives*: urea, γ-CD or γ-CD and β-CD	UV–235 nm	Enantiomeric separation of multicomponent mixtures of eight or nine chiral PCBs	111 112
	Buffer: CHES (pH 10) *Micelle*: SC *Additive*: urea		Enantiomeric separation of three PCBs	113
Phthalate esters	*Buffer*: phosphate/borate (pH 6/9) *Micelle*: SDS *Additive*: methanol	UV–210 nm	Phthalate esters (including priority USEPA phthalate esters)	114 115
Pesticides	*Buffer*: phosphate/borate (pH 5) *Micelle*: MEGA 10 *Additive*: TM-β-CD	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 420$ nm	Phenoxy acids	33
	Buffer: borate (pH 10) *Micelle*: MEGA 8			83
	Buffer: borate (pH 10) *Micelle*: OS or OM or OG or MEGA 9	UV–240 nm	Urea herbicides	84 85
	Buffer: borate (pH 7) *Micelle*: MEGA 10 *Additives*: PNA and acetonitrile	Indirect fluorimetry $\lambda_{ex} = 365$ nm $\lambda_{em} = 400$ nm	Lindane and herbisan	86

(*continued overleaf*)

Table 2 (*continued*)

Compounds	Separation media	Detection	Application	Refs.
Pesticides	*Buffer*: Borate (pH 10) *Micelle*: Deoxy Big CHAP		Enantiomeric separation of silvex herbicides	87
	Buffer: phosphate (pH 7) *Micelle*: TTAC and OTAC or DTAC	UV – 240 nm	Phenylureas	88
	Buffer: phosphate *Micelle*: SDS or Brij 35 *Additive*: methanol	Diode-array 200 – 300 nm	Chlorinated phenoxy acids	116
	Buffer: phosphate/borate *Micelle*: SDS *Additive*: urea	UV – 254 nm	Urea herbicides	
	Buffer: borate (pH 8) *Micelle*: SDS *Additive*: methanol	UV – 214 nm	Triazines and other herbicides	117 118
	Buffer: phosphate (pH 7) *Micelle*: SDS *Additives*: urea and butanol or methanol	LIF $\lambda_{ex} = 488$ nm $\lambda_{em} = 520$ nm	Phenoxy acids	119
	Buffer: phosphate (pH 7) *Micelle*: SDS *Additives*: butanol or methanol	UV – 205 nm	Phenylureas and phenoxy acids	120
	Buffer: phosphate/borate (pH 8.7) *Micelle*: SDS *Additive*: TBA	UV – 214 nm	Imidazole and methylimidazoles	121
	Buffer: borate (pH 9) *Micelle*: SDS *Additives*: urea, γ-CD, 2-methyl-2-propanol	UV – 254 nm	Enantiomeric separation of diniconazole and uniconazole	122 123
	Buffer: borate (pH 9) *Micelle*: SDS *Additives*: γ-CD and acetonitrile	UV – 200 nm	Achiral and chiral separation of organochloride herbicides	124
	Buffer: phosphate (pH 7) *Micelle*: OG or NG	UV – 240 nm	Enantiomeric separation of phenoxy acids	125
	Buffer: phosphate (pH 9) *Micelle*: OG	UV – 210 nm	Triazines and another herbicide	126
	Buffer: phosphate *Micelle*: OM	LIF $\lambda_{ex} = 329/341$ nm $\lambda_{em} = 380$ nm	Phenoxy acids	127
Phenols	*Buffer*: borate/phosphate (pH 7)	UV – 270 nm	14 substituted phenols	51
	Micelle: SDS	UV – 220 nm	Chlorophenols	53 128
	Buffer: borate/phosphate (pH 7)	UV – 215 nm	Six isomers of dichlorophenols	129 130
	Micelle: SDS	UV – 254 nm	USEPA priority phenols	47 131
	Buffer: borate (pH 8.3) *Micelle*: SC *Additive*: acetone	UV – 214 nm	Several USEPA priority phenols	61
	Buffer: borate (pH 9) *Micelle*: SDS *Additive*: HP-β-CD	UV – 220 nm	Homologs of alkylphenols and isomers of 4-nonyl-phenols	132

Table 2 (*continued*)

Compounds	Separation media	Detection	Application	Refs.
Phenols	*Buffer*: quinine sulfate (pH 6.8) *Micelle*: SDS *Additives*: methanol	Indirect fluorimetry	Methylphenol isomers and two phenol derivatives	133
	Buffer: borate/phosphate (pH 7) *Micelle*: SDS	Amperometric (0.8 V)	Diphenol isomers	134
Dyes	*Buffer*: borate (pH 9) *Micelle*: SDS *Additive*: methanol	UV–210/254 nm	Acidic and reactive dyes	135 136 137
	Buffer: phosphate/borate (pH 8) *Micelle*: SDS	Diode-array 200 nm	Xanthene and azo dyes	138
	Buffer: borate (pH 8.3) *Micelle*: SC *Additive*: acetone	UV–214 nm	Monosulfonated dyes	139
Amines	*Buffer*: phosphate *Micelle*: SDS	MS	Aromatic amines	140
	Buffer: phosphate/borate (pH 6.2) *Micelle*: SDS *Additive*: γ-CD	UV–210 nm	Aniline derivatives	141
	Buffer: phosphate/carbonate (pH 11) *Micelle*: SDS *Additive*: acetonitrile	UV–214 nm	Aromatic amines	142
Nitroaromatics	*Buffer*: phosphate/borate (pH 7) *Micelle*: SDS	UV–254 nm	Eight nitroaromatics	143
	Buffer: borate (pH 8.1) *Micelle*: SDS *Additives*: urea or methanol	UV–220/230/254 nm	Nitrotoluenes, nitroaromatic and nitramines (USEPA priority substances)	144 145 146
Aromatic sulfonates	*Buffer*: borate (pH 9) *Micelle*: SDS *Additive*: acetonitrile	UV–214 nm	Alkylbenzene sulfonates	40
	Buffer: borate (pH 9.5) *Micelle*: Brij 35	Diode-array 220 nm		147
Organo-metallics	*Buffer*: phosphate/borate (pH 6) *Micelle*: SDS *Additive*: β-CD	UV–210 nm	Organolead and organoselenium compounds	148
	Buffer: phosphate/borate (pH 9) *Micelle*: SDC *Additive*: acetone	UV–232 nm	Enantiomeric separations of tri aza aromatic ligand compounds of iron(II)	149

CHES, 2-(*N*-cyclohexylamino) ethanesulfonic acid; DMSO, dimethyl sulfoxide; DMF, *N*,*N*-dimethylformamide; MEGA 10, decanoyl-*N*-methylglucamide; SBD, starburst dendrimers; MEGA 8, octanoyl-*N*-methylglucamide; MEGA 9, nonanoyl-*N*-methylglucamide; Poly(D-SUV), poly(sodium *N*-undecylenyl-D-valinate); OS, *n*-octanoylsucrose; OM, octyl-β-D-maltopyranoside; OG, octyl-β-D-glucopyranoside; PNA, *N*-phenyl-1-naphthylamine; Deoxy Big CHAP, *N*,*N*-bis-(3-D-gluconamidopropyl)-deoxycholamide; TTAC, tetradecyltrimethylammonium chloride; OTAC, octyltrimethylammonium chloride; DTAC, dodecyltrimethylammonium chloride; TBA, tetrabutylammonium hydrogensulfate; NG, *n*-nonyl-β-D-glucopyranoside; HP-β-CD, hydroxypropyl-β-cyclodextrin.

and γ-CD, or SDC with acetonitrile. Comparing the retention characteristics of sample and standard components under different MEKC conditions increased the reliability of peak identification.[99]

The separation of highly hydrophobic compounds and closely related compounds (PAHs, PCBs and PCDDs) is generally difficult by electrophoretic techniques, as explained before. In addition, it is often difficult to dissolve sufficient analyte in the aqueous micellar solution to enable detection by standard on-column UV detection. In an attempt to address these problems, several research groups have experimented with different separation procedures, which have been devoted to the separation of PAHs as shown in Table 2. PAHs which contain

at least two aromatic rings in their basic structure are of great environmental concern owing to their carcinogenicity. Multicomponent separations of PAHs have been carried out by MEKC using basic buffers with different surfactants to form the micellar phase such as SDS,[63,66,67,69–71,100–104] SC,[85] SDC[65,81,105] or CTAB.[106] The addition of organic modifiers to an electrophoretic medium is particularly advantageous for the analysis of hydrophobic analytes because they reduce the analytes' relative affinity for the micellar phase and increase solubility. CDs (CDMEKC) have been also employed to enhance the separation of highly hydrophobic compounds such as PAHs. The separation of a mixture of 20 PAHs by CDMEKC is illustrated in Figure 3. Finally, other authors have reported the separation of several PAHs using different pseudostationary phases:

1. in situ charged micelles such as MEGA 10;[83]
2. mimetic micelles such as oligomers (SUA[91] and SUS[92]), dendrimers (SBD(X)-C_{12})[95] or resorcarenes;[96]
3. combined polymerized chiral micelle with CD.[107]

PCBs and PCDDs are persistent environmental pollutants toxic to humans and animals. The separation of these compounds by MEKC is difficult owing to the low selectivity of the system in separating highly hydrophobic compounds. The addition of CDs and urea (increasing the solubility of hydrophobic compounds and CDs in aqueous media[63]) to the separation buffer in MEKC has enabled the separation of PCBs and PCDDs. CDMEKC using SDS with γ-CD and urea in the separation buffer has enabled successful separation of 11 trichlorobiphenyl isomers,[66] three pairs of trichlorodibenzo-*p*-dioxins,[66] different PCDD isomer pairs,[108] and 14 PCB congeners.[109] CTAB has been used as surfactant in CDMEKC to separate 10 PCB congeners.[110] In addition, CDMEKC with SDS has been employed to achieve the enantiomeric separation of 12 chiral PCBs each one in its two enantiomers using γ-CD[111] alone or mixed with β-CD[112] in the buffer. Multicomponent mixtures of nine[111] or eight[112] chiral PCBs were separated. Figure 4 shows the separation of the 18 enantiomers of a mixture of nine chiral PCBs, containing one octachlorinated biphenyl, in an analysis time close to 35 min.[111] Bile salts are alternate micellar systems for the separation of chiral compounds by MEKC and their use to separate three chiral PCBs, each one in its two enantiomers in less than 30 min has been described.[113]

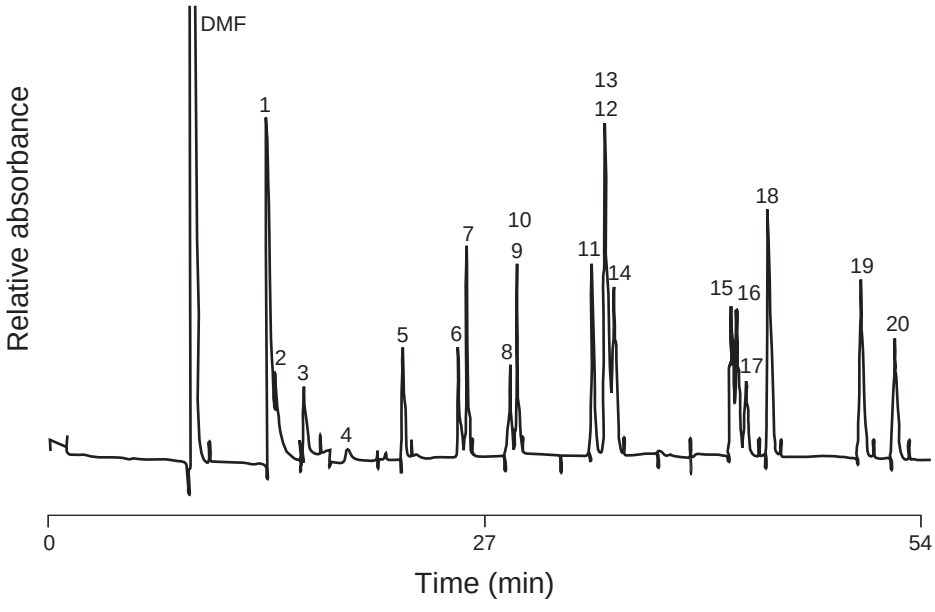

Figure 3 Electropherogram corresponding to the separation of 20 PAHs by CDMEKC. Conditions: separation media, 0.14 M borate buffer (pH = 9.0) with 2.5 M urea, 0.10 M SDS, 0.042 M β-CD and 0.026 M γ-CD; capillary tube, 50 μm ID × 65 cm (50 cm to the detector); injection by pressure, 0.02 min at 20 mbar; applied voltage, 15 kV; UV–detection at 230 nm; temperature, 35 °C. Peaks (analyte concentration 100 μg mL^{-1}): (1) acenaphthene, (2) acenaphthylene, (3) naphthalene, (4) benzo(*g*,*h*,*i*)perylene, (5) fluorene, (6) phenanthrene, (7) pyrene, (8) chrysene, (9) perylene, (10) anthracene, (11) benzo(*e*)pyrene, (12) benzo(*a*)pyrene, (13) benzo(*a*)anthracene, (14) fluoranthene, (15) dibenzo(*a*,*h*)anthracene, (16) benzo(*k*)fluoranthene, (17) triphenylene, (18) benzo(*j*)fluoranthene, (19) benzo(*b*)fluoranthene, and (20) indeno(1,2,3-*cd*)pyrene. (Reproduced by permission of Elsevier Science Publishers. Copyright (1997) from B. Jiménez, D.G. Patterson, J. Grainger, Z. Liu, M.J. González, M.L. Marina, *J. Chromatogr. A*, **792**, 411–418.[103])

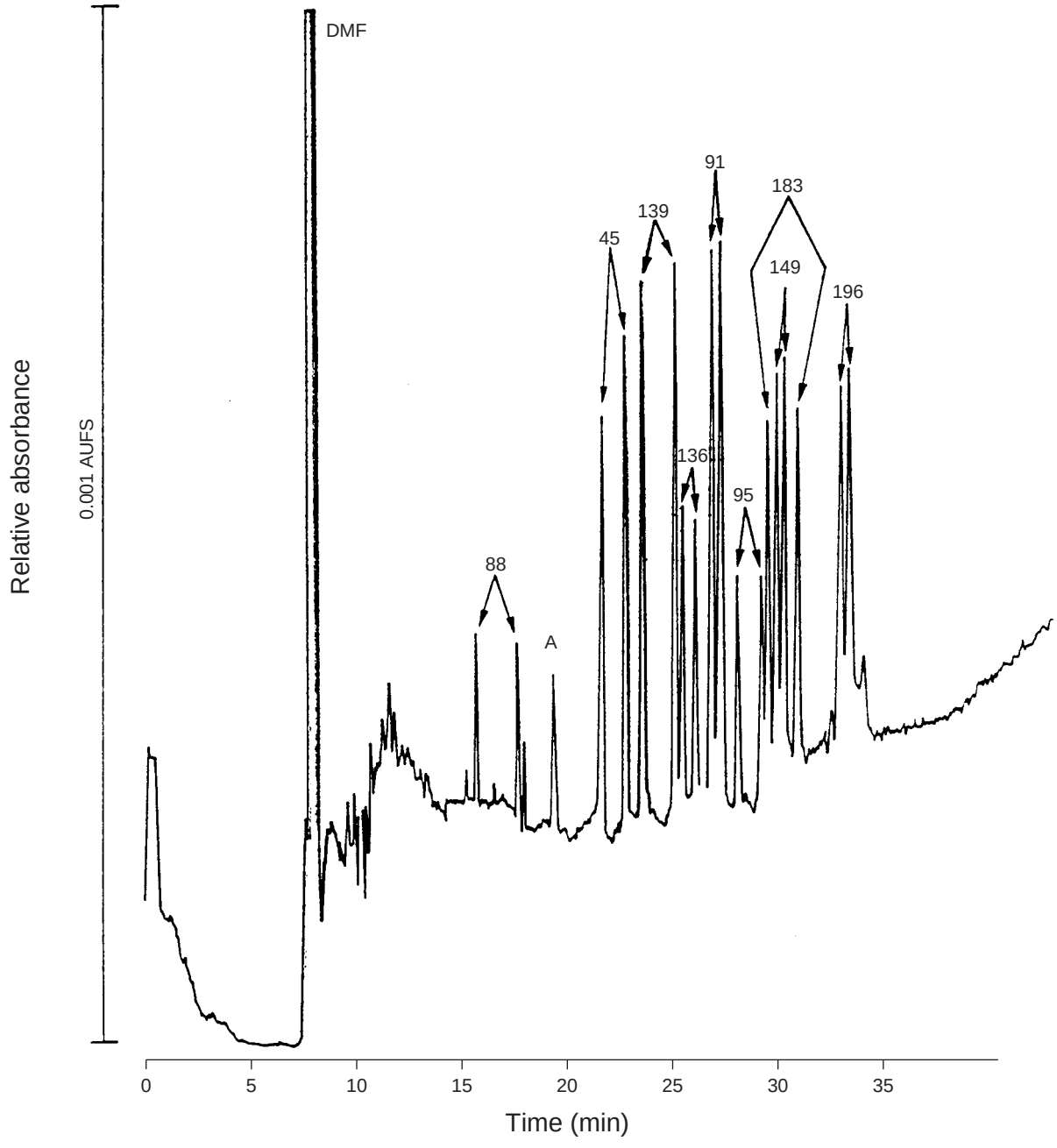

Figure 4 Electropherogram corresponding to the separation of a mixture of nine chiral PCBs by CDMEKC. Conditions: separation media, 0.10 M CHES buffer (pH = 10.0) with 2 M urea, 0.11 M SDS and 0.05 M γ-CD, capillary tube, 50 μm ID × 65 cm (50 cm to the detector), injection by pressure, 0.02 min at 20 mbar, applied voltage (current), 15 kV (56 μA), UV-detection at 235 nm, temperature, 45 °C. Peaks: DMF, each pair of enantiomers is identified by IUPAC number (according to Ballschmiter nomenclature; K. Ballschmitter, M. Zell, *Fresenius Z. Anal. Chem.*, **302**, 20–31 (1980)[152]); and (A) unknown peak. (Reproduced by permission of Chromatographia Vieweg-Publishing from M.L. Marina, I. Benito, J.C. Díez-Masa, M.J. González, *Chromatographia*, **42**, 269–272 (1996).[111])

Phthalates are widely used as plasticizers in industry, and can migrate from the plastics into the environment, so they are common contaminants in aquatic environments.[153] USEPA method 606 specifies analysis by GC with an electron capture detector, but appropriate pretreatment is necessary. The separation of six and ten phthalate esters, including some priority pollutants, was achieved by MEKC using SDS alone[114] or with methanol as additive[115] respectively. The use of SUA as a pseudostationary phase, in the presence of relatively high

amounts of organic modifier (acetonitrile), allows the separation and analysis of hydrophobic analytes such as phthalates.[91] The major disadvantage of this new phase lies in its laborious preparation (>2 days).

Pesticides have been widely studied by MEKC as shown in Table 2, since although they can be also separated by CZE, in some cases the separation can be enhanced by MEKC. Thus, multicomponent separation of different types of herbicides such as phenoxy acids, triazines, and phenylureas have been carried out by MEKC using mainly SDS[116–121,144] as anionic surfactant and other in situ charged micelles such as MEGA 8, MEGA 9 or MEGA 10.[33,83–86] Also, seven phenylurea herbicides have been successfully separated using cationic micelles such as TTAC with OTAC or DTAC.[88] On the other hand, enantiomeric separation of diniconazole and uniconazole,[122,123] and chiral separation of organochloride herbicides[124] have been performed by MEKC using SDS micelles with CDs as additives. Optical isomers of silvex herbicides have been separated using in situ charged Deoxy Big CHAP micelles.[87] Other chiral separations of phenoxy acid herbicides have been accomplished using neutral chiral surfactants such as OG, OM or NG.[125–127]

MEKC can also provide enhanced resolution of phenols in comparison with CZE. Thus, MEKC using SDS has been employed to separate substituted phenols,[51] chlorophenols,[53,128–130] and USEPA priority phenols.[47,131] A bile salt like SC has also been used as surfactant for a rapid separation of some priority phenols.[61] On the other hand, the addition of HP-β-CD in separation buffer with SDS enables the separation of 4-nonylphenol isomers and other alkylphenol homologs.[132] However, the separation mechanism of ionizable phenols by MEKC is complicated by the effect of the combined electrophoretic and chromatographic migration on separation. This explains the different elution orders shown in MEKC versus CZE. Although UV detection has been generally employed to determine phenols, a mixture of phenols has been analyzed by indirect fluorimetry[133] or amperometric detection.[134]

Finally, different types of compounds such as *dyes*,[135–139] *amines*,[140–142] *nitroaromatics*,[143–146] *aromatic sulfonates*[147] and *organometallics*[148,149] have been separated by MEKC using SDS micelles with or without additives such as organic solvents (acetonitrile, methanol, acetone), urea or CDs (see Table 2). The use of a bile salt like SDC has permitted the enantiomeric separation of tri aza aromatic ligand compounds of iron(II).[149] On the other hand, MEKC and CZE appear to be complementary in the analysis of *alkylbenzene sulfonates*.[40] MEKC with SDS is less selective than CZE for the separation of linear alkylbenzene sulfonates, although it appears to be much more selective than CZE for the separation of alkylbenzene sulfonate isomers.[40] However, MEKC with a nonionic surfactant such as Brij 35 instead of SDS improved the separation of these compounds with respect to CZE.[147]

2.3 Other Kinds of Electrokinetic Chromatography

One of the working modes most used in CE is EKC, which includes several techniques with the common feature that separations are based on the partitioning of the analytes between two phases which have different velocities relative to each other. One of the phases is a buffer solution (which migrates with the velocity of the EOF) and the other is a compound distributed homogeneously in the buffer solution, which is strongly electrophoretically retarded and acts as a pseudostationary phase. Micelles are the most used pseudostationary phase (MEKC is one kind of EKC), but other pseudostationary phases such as CD derivatives which have ionic groups, microemulsions, cationic or anionic surfactant monomers, or other alkyl–ionic compounds, which cannot be absolutely classed as micelles, have been successfully employed in EKC for the analysis of environmental compounds resulting in different modes of EKC.

Cyclodextrin electrokinetic chromatography (CDEKC) is a mode of EKC that can employ charged CD or mixtures of neutral and charged CD as the pseudostationary phase. In the first case, the distribution process between the analytes and the CD allows the separation by a host–guest interaction. As an example, cresol isomers have been successfully separated by using a negatively charged CD, carboxymethyl-β-cyclodextrin (CM-β-CD).[154] Using neutral and charged CD mixtures the separation was performed on the basis of differential partitioning of analytes between the two CDs moving relatively, neutral CD (which migrates with the EOF) and charged CD (which migrates with a given electrophoretic velocity and acts as a pseudostationary phase). Mixtures of methyl-β-cyclodextrin (M-β-CD) and sulfobutyl ether-β-cyclodextrin (SB-β-CD) have been used to provide efficient separation of PAH isomers.[155] The separation of these isomers using mixtures of α-CD or β-CD with CM-β-CD was also successful.[156] Thus, the 16 PAHs listed by USEPA as priority pollutants have been separated with the system formed by M-β-CD and SB-β-CD[157,158] in less than 20 min.

In microemulsion electrokinetic chromatography (MEEKC), the pseudostationary phase is a microemulsion, that is, an aggregate with a similar structure to that of a micelle, except that it has a core of a minute droplet of oil. Microemulsions are prepared by mixing oil (hydrocarbon), water, surfactant, and co-surfactant (medium alkyl-chain alcohol) in the appropriate proportions. This technique has been studied in comparison

with MEKC using phenol derivatives[159] and phenylurea and chlorsulfuron herbicides as model mixtures (see Figure 5).[73] Results have shown that with MEEKC it is very easy to extend the retention window through an increase in the surfactant concentration, and moreover MEEKC provides different selectivities from MEKC.

Finally, in hydrophobic interaction electrokinetic chromatography (HIEKC), the pseudostationary phase can be formed by cationic or anionic surfactant monomers, or other alkyl–ionic compounds in hydro-organic solutions with high concentrations of organic solvent (≥40%). Under these conditions, micelles are not formed because of the high concentration of organic solvent. The separation mechanism is attributed to the hydrophobic interaction between a monomolecular pseudostationary phase and neutral analytes. A series of alkyl aryl ketone homologs[160] has been successfully separated, in less than 6 min, by HIEKC using phosphate buffer at pH 7.0 with SDS or at pH 2.8 with CTAB as anionic or cationic surfactant, respectively, in 50% acetonitrile. Several mixtures of PAH homologs have been well resolved using a basic buffer (about pH 9) in 40–50% acetonitrile and with tetrahexylammonium perchlorate (five homologs),[161] or tetraheptylammonium bromide (16 homologs)[162] or sodium dioctylsulfosuccinate (23 homologs).[163]

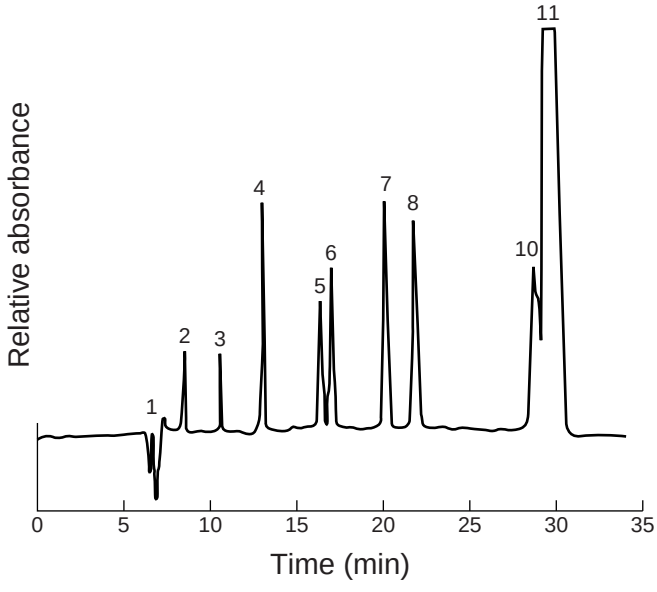

Figure 5 Separation of seven herbicides by MEEKC. Conditions: separation media, 0.0124 M dihydrogenophosphate/0.0038 M borate buffer (pH = 7.0) with 0.070 M *n*-octane, 0.800 M *n*-butanol and 0.050 M SDS; capillary tube, 50 μm ID × 75 cm (55 cm to the detector); hydrodynamic injection, 10 cm × 5–20 s; applied voltage 24 kV; UV-detection at 254 nm; temperature, 30 °C. Peaks: (1) DMF, (2) chlorsulfuron, (3) fenuron, (4) monuron, (5) fluometuron, (6) chloroturon, (7) dinuron, (8) linuron, (10) impurity in *o*-diphenylbenzene and (11) *o*-diphenylbenzene. (Reproduced by permission of Elsevier Science Publishers. Copyright (1995) from L.G. Song, Q.Y. Ou, W.L. Yu, G.Z. Li, *J. Chromatogr. A*, **699**, 371–382.[73])

2.4 Capillary Electrochromatography

The combination of CZE and HPLC, called CEC, is one of the newest working modes in CE.[164–166] In CEC, the separation of uncharged analytes is based on partitioning between a mobile phase (buffer which migrates with the velocity of the EOF) and a stationary phase (HPLC material which is static at all times). The stationary phase can be either coating the inner wall of a capillary (open tubular column, OTC) or packing the inner diameter of the capillary (packed capillary column, PCC). The advantage of using EOF[164] as the driving force is an increased column efficiency because of the flat-flow profile, and moreover, in the case of PCCs the ability to use smaller particles than those used in HPLC (particle diameters as small as 1.5 μm) because EOF is not limited by the column back pressure that results from the passage of a mobile phase through a packed bed, the flow being limited in this case only by the voltage applied. The separation of charged analytes is also possible in CEC, but in this case, the analyte migration velocity will be determined by the partition between the two phases described previously and its electrophoretic mobility. Although two kinds of column can be used (OTC or PCC), most applications have employed PCCs with an ID of less than 100 μm, and these have been commercialized, while OTCs are still in the developmental stage. Generally, PAHs[167–170] have been used as standards to evaluate OTCs. Some compounds such as PAHs,[171–173] benzene derivatives,[174] nitrotoluenes and biphenyls[175] have been utilized to evaluate PCCs. Figure 6 shows an electrochromatogram of 16 USEPA priority PAHs with a PCC of 75 μm ID and 3 μm octadecylsilica (C_{18}) particles using a buffer with a high percentage of organic solvent.[176] Attempts have been made to separate the same mixture of PAHs in HIEKC[162] and MEKC,[66] but with poorer results. A linear gradient elution in CEC was used for the separation of 16 PAHs showing a better performance than in isocratic elution.[177]

It should be emphasized that CEC is still not widely applied owing to the difficulty in manufacturing and operating the PCCs.

3 ANALYSIS OF POLLUTANTS IN ENVIRONMENTAL SAMPLES

Most environmental applications in CE are restricted to demonstrations of high-resolution separations with standards of specific pollutants and only a few refer to the determination of pollutants in environmental samples.

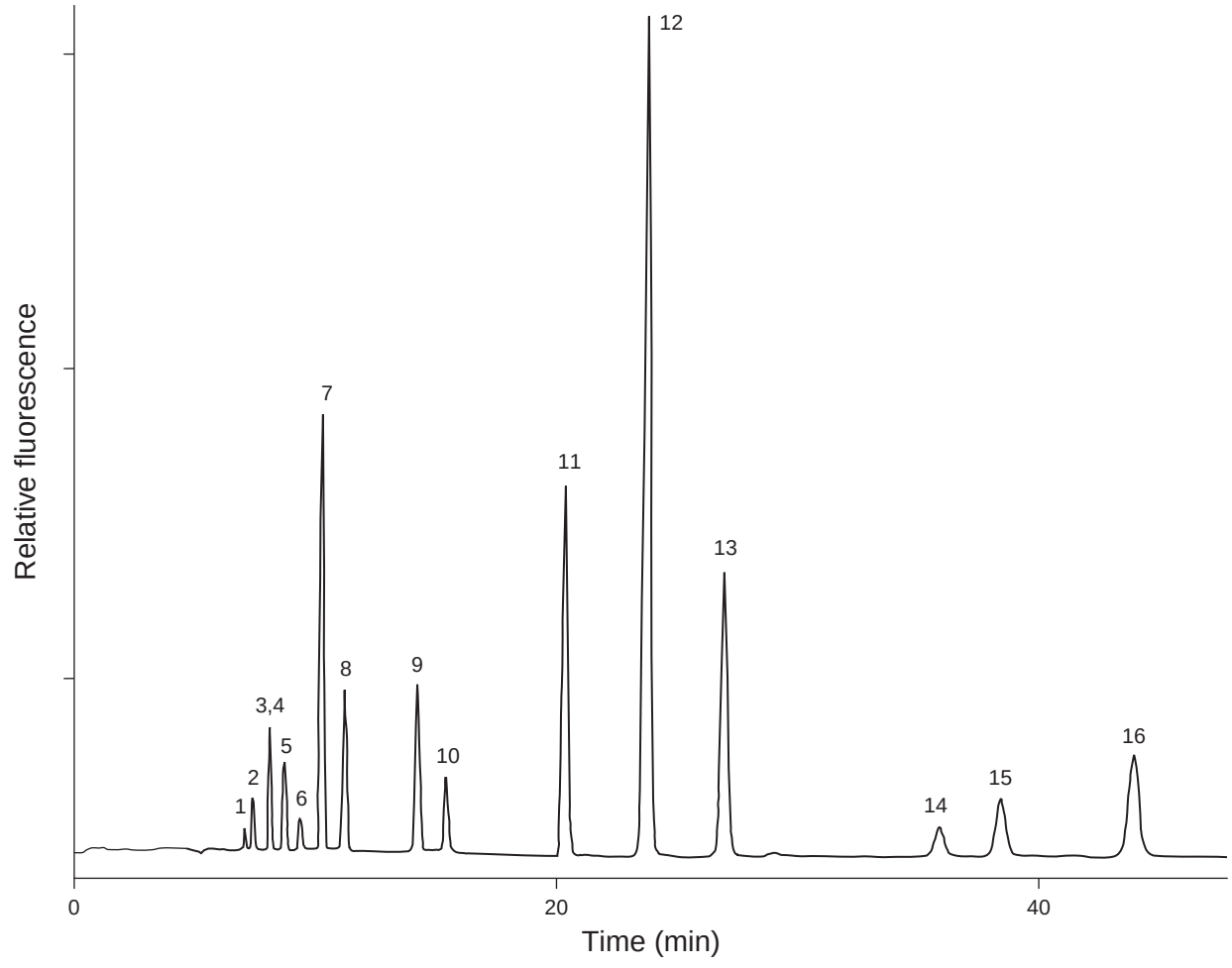

Figure 6 Electrochromatogram corresponding to a mixture of 16 USEPA priority PAHs. Conditions: separation media, 0.004 M borate buffer with 80% acetonitrile, capillary tube, 75 μm ID × 43 cm (33 cm packed length), electrokinetic injection, 5 s × 5 kV, applied voltage, 15 kV, LIF-detection ($\lambda_{exc} = 257$ nm, $\lambda_{em} = 400$ nm). Peaks (analyte concentration 10^{-6}–10^{-8} M): (1) naphthalene, (2) acenaphthylene, (3) acenaphthene, (4) fluorene, (5) phenanthrene, (6) anthracene, (7) fluoranthene, (8) pyrene, (9) benzo(*a*)anthracene, (10) chrysene, (11) benzo(*b*)fluoranthene, (12) benzo(*k*)fluoranthene, (13) benzo(*a*)pyrene, (14) dibenzo-(*a*,*h*)anthracene, (15) benzo(*g*,*h*,*i*)perylene and (16) indeno(1,2,3-*cd*)pyrene. (Reproduced by permission of American Chemical Society. Copyright (1995) from C. Yan, R. Dadoo, H. Zhao, R.M. Zare, *Anal. Chem.*, **67**, 2026–2029.[176])

There are two main reasons for this. First, CE commercial instruments have been introduced only relatively recently and are not yet in widespread use in environmental analytical laboratories. Second, the sensitivity obtained for the analysis of pollutants by CE has not been adequate for trace analysis. Unfortunately, the injection volumes that can be tolerated in CE without decreasing separation performance, are typically a few nanoliters only. Therefore, although commercial CE instruments equipped with UV absorbance detectors are able to detect quantities as low as several hundred femtograms,[6] very poor sensitivity is obtained as expressed in sample concentration terms (several hundred parts per billion or above 10^{-6} M), which are generally at least three orders of magnitude above levels of pollutants in environmental samples (low parts per billion to parts per trillion level). Because the detection ability of commercial CE instruments is far from ideal and they are not sensitive enough to determine compounds in environmental samples, improvements needed to be made to detection systems or preconcentration procedures.

3.1 Sensitivity Enhancement Procedures

Several approaches have been developed to improve the limit of detection (LOD) for CE instruments,[178] such as extending the optical path length of the capillaries, using highly sensitive detection systems, and employing several strategies for the preconcentration of samples, namely, solid-phase extraction (SPE), field-amplified sample stacking (FASS), or isotachophoresis (ITP).

3.1.1 Improved Capillary Geometry and Detector Cell Design

Unfortunately, the detection system that can be used in CE, without decreasing the separation performance, is an on-column detection, which implies that the optical path length for detection is limited by the ID of the capillary column. Some authors have extended the optical path length by using Z-shaped flow cells[179] which improve the LOD by a factor of 20–60, or bubble cells[84,180] which allow a multiplying of sensitivity usually by a factor of 3–5.

3.1.2 Alternative Detection Systems

Fluorescence detection can provide LODs several orders lower than UV absorbance. In applications of environmental interest the excitation of molecules is done with different lasers in order to achieve very high sensitivities of LIF. Thus, the laser most used in LIF for pollutants analysis is the He–Cd laser ($\lambda_{ex} = 325\,nm$). It is employed (see Table 2) for the determination of hydrophobic compounds with native fluorescence such as aflatoxins[65] or PAHs[65,71,81,95,105,157,181–183] in which LODs are 10^{-9}–10^{-11} M. However, most environmental pollutants are not intrinsically fluorescent, like some phenoxy acids herbicides[33,127] which are analyzed by LIF after a derivatization with a fluorophore (e.g. 7-aminonaphthalene-1,3-disulfonic acid (ANDSA)) achieving a LOD around a few parts per billion. The Ar ion laser ($\lambda_{ex} = 257\,nm$) has been employed to analyze 16 USEPA priority PAHs.[127,157] A semiconductor laser[105,184] has also been employed for detection of PAHs obtaining worse sensitivity (LOD 10^{-8} M) than with a He–Cd laser. Finally, another alternative for nonfluorescent analytes is the use of indirect LIF detection. This detection is the result of the decreased signal from a fluorescing reagent present in the running buffer, which is caused by the displacement of the reagent with nonfluorescing analytes. Indirect LIF detection was employed in the separation of six aromatic compounds,[185] two neutral herbicides,[86] 11 USEPA priority phenols[186] and a mixture of phenols (phenol, 1-chlorophenol and cresol isomers).[133] LODs for these compounds were in the 10^{-6}–10^{-7} M range.

MS is one of the most powerful detection methods for obtaining structural information on separated analytes. Recently, capillary electrophoresis/mass spectrometry (CE/MS) has been recognized as a promising technique for analysis of mixtures of pollutants. Thus, sulfonylurea herbicides have been analyzed by CZE using ion spray MS.[26] Moreover, aromatic amines have been analyzed by MEKC using an electrospray-chemical ionization interface.[140] However, some problems still exist in CE/MS. This detector has relatively poor concentration sensitivity. The use of nonvolatile buffers is generally avoided (composition of buffer must be chosen previously). On the other hand, the coupling of MEKC with MS is difficult owing to the presence of high concentrations of micelles in the running buffer, which can cause low ionization efficiency. Nevertheless, current research is rapidly changing this situation by developing new CE/MS interfaces, and novel pseudostationary phases.

Amperometric detection is advantageous because LOD should not increase with a decrease in the ID of the capillary column as occurs with UV detection. However, this type of detection has been scarcely used in CE, particularly in the analysis of environmental pollutants, owing to its difficult manipulation and to the limited number of pollutants with electroactive character. Actually, amperometric detection has been used for the analysis of phenols[134,187] and chlorophenols[188] reaching a LOD close to 10^{-8} M.

3.1.3 Preconcentration by Solid-phase Extraction

Liquid–liquid extraction (LLE) and SPE are the most common methods for liquid sample enrichment. LLE is a successful method but with recognized disadvantages: it is laborious, time consuming and subject to problems such as the emulsification of the samples during extraction and the disposal of large volumes of toxic solvents. Thus, trace enrichment of environmental samples in CE analysis is preferably performed via SPE (see Table 3), providing concentration factors of 10–10^6, and being especially effective for aqueous samples. In off-line SPE, large-volume samples are passed through a small cartridge packed with sorbent. Analytes are trapped by the sorbent and then eluted with a small volume of a suitable solvent. An alternative technique based on the SPE principle but with a piece of sorbent membrane disk to replace the SPE cartridge has been reported.[189] Because the sorbent disk has a larger contacting area than a SPE cartridge, larger flow rates are permitted, and therefore the extraction for large-volume (1 L or more) samples can be performed in much shorter times than with a SPE cartridge. Nevertheless, with the off-line preconcentration technique, the sample is not introduced directly into the capillary, requiring major handling of the samples and analysis time. For this reason, an on-line coupled SPE/CZE has been described[29,190] using a reversed-phase coated capillary column to perform SPE preconcentration before CZE separation. However, owing to the small amount of reversed-phase coating on the inner wall of the capillary, the breakthrough volumes and the capacity, and hence the enrichment factors, are rather limited (10–35-fold). This is why on-line SPE is used with a short plug of

Table 3 Experimental conditions for the analysis of pollutants in environmental samples by CE

Compounds	CE mode	Sample preconcentration	Separation media	Capillary length[a] × ID (mm)	Injection	Detection	Application	Ref.
Aflatoxins	MEKC	Off-line SPE	*Buffer*: 50 mM phosphate/6 mM borate *Micelle*: 50 mM SDC *Additive*: 10% acetonitrile	350 × 0.025	Electromigration	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 450$ nm	Determination of aflatoxins in Indiana corn meal extract spiked with 16 ppm	181
Amines	CZE	Off-line SPE	*Buffer*: 50 mM phosphate (pH 2.35) *Additive*: 1,3-diaminopropane	600 × 0.050	–	UV–280 nm	Determination of anilines in groundwater and soil extracts from industrial plant	192
Dyes	MEKC	Off-line SPE	*Buffer*: 50 mM borate (pH 8.35) *Micelle*: 100 mM SC *Additive*: 10% acetone	500 × 0.050	–	UV–214 nm	Determination of dyes in soil extracts spiked with 3 ppm	138
Pesticides	CZE	FASS	*Buffer*: 50 mM acetate (pH 4.8) *Additive*: 4 mM α-CD	760 × 0.050	–	UV–200 nm	Chiral separation of herbicides in a Rhine river water sample spiked with 5 ppb of phenoxy acid	193
		ITP	*Leading buffer*: 10 mM citric acid (pH 6.08), 0.2 mg mL^{-1} MHEC *Terminating buffer*: 9.8 mM citric acid (pH 6.08) 0.15 mg mL^{-1} PEG	ITP column: 80 × 0.085 CZE column: 250 × 0.025	–	UV–310 nm	Determination of paraquat and diquat in water samples	194
		Off-line SPE	*Buffer*: 40 mM borate (pH 8.5)	500 × 0.075	–	UV–214 nm	Monitoring of herbicides in water sources	195

Pesticides	CZE	–	*Buffer*: 100 mM borate (pH 9.6) *Additive*: 10% methanol	720 × 0.050	Hydrodynamic	UV–240 nm	Determination of glyphosate in human serum treated (acid precipitation and centrifugation)	196
		Off-line SPE	*Buffer*: 40 mM borate (pH 9.0)	500 × 0.075	Hydrodynamic	UV–214 nm	Determination of sulfonylureas in tap water	197
		–	*Buffer*: acetate (pH 4.0) *Additive*: NaCl	500 × 0.050	Electromigration	UV–205 nm	Determination of paraquat, diquat and dibenzoquat in crop water	198
	CD/CZE	Off-line SPE	*Buffer*: phosphate (pH 5.6) *Additives*: 4 mM α-CD, 1 mM β-CD	400 × 0.050	Hydrodynamic	UV–200 nm	Separation of chlorophenoxy acids in lake water spiked with 20 ppb. Resolution of two pairs of enantiomers	199
	MEKC	Off-line SPE	*Buffer*: 60 mM borate (pH 9.2) *Micelle*: 50 mM SDS	500 × 0.075	Hydrodynamic	UV–214 nm	Determination of triazines spiked in bottled water	200
		Off-line SPE	*Buffer*: 20 mM borate (pH 9.0) *Micelle*: 50 mM SDS	450 × 0.050	Hydrodynamic	UV–205 nm	Determination of cycloheximide (fungicide) in microbial extracts	201
		Off-line SPE	*Buffer*: 30 mM borate (pH 7.0) *Micelle*: 80 mM SDS *Additives:* 14% methanol, 20% isopropanol	500 × 0.075	–	UV–214 nm	Determination of sulfonylureas in soil extract spiked with 20 ppb	202

(*continued overleaf*)

Table 3 (*continued*)

Compounds	CE mode	Sample preconcentration	Separation media	Capillary length[a] × ID (mm)	Injection	Detection	Application	Ref.
Nitro-aromatics	MEKC	–	*Buffer*: 15 mM borate (pH 8.75) *Micelle*: 25 mM SDS	510 × 0.075	Hydrodynamic	UV – 230 nm	Separation of explosives extracted from contaminated soils	203
Organo-metallics	MEKC	Off-line SPE	*Buffer*: 30 mM phosphate/15 mM borate (pH 7.65) *Micelle*: 50 mM SDS	500 × 0.050	–	UV – 200 nm	Determination of organolead and organotin species in spiked drainage and tap water samples	204
	CDMEKC	LLE	*Buffer*: 25 mM phosphate/borate (pH 6.0) *Micelle*: 50 mM SDS *Additive*: 5 mM β-CD	440 × 0.050	–	UV – 210 nm	Determination of organoselenium and organolead compounds in water samples	148
PAHs	MEKC	–	*Buffer*: 50 mM borate (pH 8.35) *Micelle*: 100 mM SC *Additive*: 10% acetonitrile	500 × 0.050	–	UV – 214 nm	Determination of PAHs in an extract of creosote-contaminated soil	61
		–	*Buffer*: 8.5 mM borate (pH 9.9) *Micelle*: 8 mM SDS *Additive*: 50% acetonitrile	280 × 0.050	Hydrodynamic	Diode-array 190 – 350 nm	Quantitation of PAHs in a soil extract contaminated with a mixture of standard PAHs and a machine oil extract	106
	CDEKC	–	*Buffer*: 50 mM borate (pH 9.0) *Additives*: 25 mM SB-β-CD and 20 mM β-CD	500 × 0.050	Hydrodynamic	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 370$ nm	Determination of the 16 USEPA priority PAHs in a contaminated soil extract	157

Phenols	CZE	–	*Buffer*: 40 mM borate (pH 10)	560 × 0.050	Hydrodynamic	UV–230 nm	Determination of pentachloro-phenol in 100 mL of tap water sample spiked with 0.5 ppb	180
		–	*Buffer*: 15 mM borate (pH 9.9) with 1 mM fluorescein	450 × 0.020	Electromigration	Indirect fluorimetry	Determination of the 11 USEPA priority phenols from industrial wastewater	186
		–	*Buffer*: 20 mM CHES (pH 10.1)	625 × 0.050	Electromigration	Amperometric (+1.5 V)	Determination of the 11 USEPA priority phenols from industrial wastewater	187
		LLE	*Buffer*: 15 mM borate/45 mM phosphate (pH 8)	350 × 0.025	Electromigration	Amperometric (+1.4 V)	Determination of chlorophenols from industrial wastewater	188
Sulfonates	MEKC	Off-line SPE	*Buffer*: 50 mM borate *Micelle*: 100 mM SDS	700 × 0.075	–	UV–230 nm	Determination of naphthalene sulfonates in spiked river water	205
	CZE	Off-line SPE	*Buffer*: borate (pH 8.75)	800 × 0.075	–	LIF $\lambda_{ex} = 325$ nm $\lambda_{em} = 370$ nm	Identification and quantitation of two trisulfonates and amino-trisulfonates in Elbe river samples	205

[a] Capillary length to the detection. MHEC, methylhydroxyethylcellulose.

a reversed-phase material (particles) which is packed at the beginning of the separation capillary.[191] These capillaries have a somewhat higher capacity and will show higher breakthrough volumes and enrichments (up to 100-fold). However, the main disadvantage is that these capillaries are only available from a limited number of suppliers, and are much more expensive (at least 30 times) than the simple fused-silica capillaries commonly used in CE. In addition, an environmental sample can almost never be directly introduced because the system could suffer plugging and severe contamination. Therefore, the sample must be cleaned or extracted and should not contain a high percentage of organic solvent.

3.1.4 On-column Preconcentration by Field-amplified Sample Stacking

FASS occurs when the injected sample has a lower ionic strength than the running buffer. Then, an analyte ion dissolved in the sample matrix will experience locally an increased field strength and will migrate with a higher velocity. When the analyte reaches the boundary between the sample matrix zone and the running buffer, it will slow down again and stack into a zone much shorter than the original sample zone, thus the analyte has been preconcentrated or focused on-column. Cationic compounds are focused at the front of the sample zone and anionic compounds at the back of the sample zone.[193]

In the simplest form of FASS, a plug of sample (dissolved in diluted buffer or in water) with a very short length compared to the total length of the capillary, is introduced hydrodynamically into the capillary. In this case, LODs can be improved by a factor of 10. However, when attempting to inject a larger volume of sample, a laminar flow occurs and destacks the stacked analyte ion.[206] Fortunately, Burgi and Chien[207] provided a simple method for removing the relatively large sample matrix zone during or after the enrichment step, but before the CZE separation started. This can be accomplished by pumping out the matrix zone, using the EOF. For anionic analytes, this involves field polarity reversion after the capillary is filled with the sample, but for cationic analytes, this demands a reversed EOF. Therefore, the enrichment factor attainable by field amplification can also benefit from the EOF and LODs can then be improved by a factor of 200. The same technique can also be performed by electrokinetic injection with LODs improved by a factor of 1000. However the reproducibility is poor.

FASS can be also used in MEKC. For ionic analytes, the procedure is the same as that for CZE. However, for neutral analytes the procedure involves dissolving them in a micellar solution with a low concentration just above the CMC. Then, the micelles will be stacked up under applied voltage and will cause stacking of neutral analytes.

3.1.5 On-line Preconcentration by Isotachophoresis

ITP is a working mode in CE which is performed in a discontinuous buffer system.[208] The analyte ions of an injected sample migrate between a leading buffer having the highest mobility and a terminating buffer having the lowest mobility. When voltage is applied, the analyte ions will be focused and migrate as consecutive zones, according to their mobility, between the leading and terminating buffers.[209] From the point of view of trace analysis, on-line coupling of ITP and CZE is very attractive[193,210] because it reduces the LOD by a factor of 10^3–10^4. However, this technique has seldom been employed in environmental analysis (see Table 3) because it requires complicated instrumentation.

ITP can be also performed before CZE separation as an ITP-like state realized only for a limited period of time (transient-ITP preconcentration step).[193,210] Transient-ITP is possible when the sample contains an excess of a matrix ion with higher electrophoretic mobility than analytes, and at the same time, the CZE electrolyte has a lower mobility than analytes. In this case, the matrix ion can act as a leading buffer and the CZE electrolyte as a terminating buffer. This technique can be used in commercial CE instruments by carefully selecting the CZE electrolyte to form an ITP buffer system and it could reduce the LOD by a factor as high as 200 when analytes migrate just behind the matrix ion. However, as in ITP, there are some disadvantages:

1. It might be difficult occasionally to find a suitable combination of leading and terminating buffers.
2. The migration times of analytes in unknown samples may be hard to predict.
3. Irreproducibility of EOF will give bad enrichment performance.

Finally, the enrichment procedures described above (FASS and ITP) can be problematic in some environmental samples where matrix components may also be concentrated; significant background matrix may degrade separation and mask the analytes of interest. Another drawback is that concentration enhancement can be reduced, eliminated or even reversed when the sample has a relatively high electrolyte content, because the sample zone may broaden, decreasing the separation efficiency and then the concentration sensitivity may decrease significantly. To resolve these problems, some authors employ a sample preconcentration technique like SPE coupled with FASS[143] or with ITP/CZE[194] for separating some pollutants. Such combination allows

the minimum detectable concentration of analytes to be decreased to low parts per billion or even parts per trillion levels without using sophisticated and expensive detection systems.

3.2 Analysis of Pollutants in Environmental Samples

CE has been employed to analyze pollutants in environmental samples and environmental matrices spiked with standards. Table 3 shows experimental conditions employed to analyze these compounds arranged alphabetically.

MEKC has been employed to determine hydrophobic compounds such as *aflatoxins*. High-speed separation of four aflatoxins in less than 30 s was optimized[181] and a bile salt such as SDC was used with acetonitrile as additive. This separation coupled with an on-column LIF detector allows LODs in the range 0.05–0.9 fmol for underivatized aflatoxins. This method was applied to the analysis of a spiked (16 ppm) corn meal extract obtained by SPE.[181] *Amines* from groundwater and soil samples preconcentrated by off-line SPE have been easily analyzed by CZE based on their positive charge in acidic solution.[192]

Dyes and *pesticides* can be separated by both CZE and MEKC. In the case of seven *monosulfonated dyes*[138] from spiked water and soil samples, MEKC was selected for its better selectivity. On the other hand, the potential and drawbacks of CE for the monitoring of pesticides water pollution have been discussed.[195] CZE has been employed primarily[193–199] (seven of ten works in Table 3), although MEKC has also been used.[200–202] UV detection was used in all cases (Table 3) and different sample treatments (off-line SPE, FASS and ITP) were employed. Figure 7 shows the electropherogram corresponding to a Rhine river water sample spiked with 5 ppb of four herbicides, two of them chiral, by using CZE and a derivative of β-CD as chiral selector in the separation buffer (sample volume 20 mL).[193] In Figure 8, the versatility and efficiency of MEKC in the separation of 11 herbicides from distilled water are shown. In the same electropherogram, ionic herbicides and various nonionized herbicides are separated in a short analysis time.[195]

Other pollutants such as *nitroaromatic* and *organometallic* compounds have been determined by MEKC in heavily contaminated soil samples and in water samples, respectively. Because the concentrations of nitroaromatic compounds were relatively high (soil from military zones), the analysis was accomplished by MEKC with UV detection without preconcentration.[203] The possibility of determining organolead and organotin compounds in other environmental samples such as drainage and tap waters has also been demonstrated.[204] Figure 9 presents

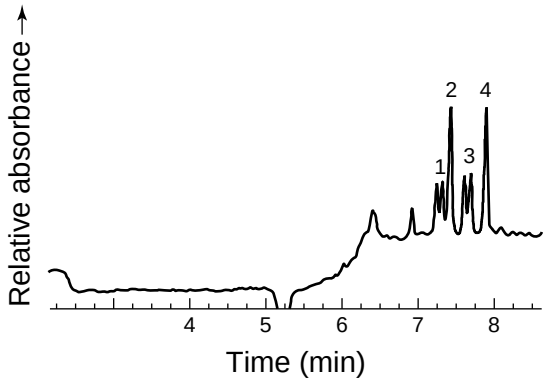

Figure 7 Electropherogram corresponding to a Rhine river water sample (sample volume 20 mL) spiked with 5 ppb of four herbicides (two chiral) by CDCZE and using preconcentration by off-line SPE (C_{18} disks) coupled with FASS. Peaks: (1) 2-(2-methyl-4-chlorophenoxy) propionic acid, (2) 2-methyl-4-chlorophenoxyacetic acid, (3) 2-(2,4-dichlorophenoxy) propionic acid, and (4) 2,4-dichlorophenoxyacetic acid. Conditions: separation media, 0.03 M acetate buffer (pH = 4.8) with 20 g L^{-1} heptakis-(2,6-di-*o*-methyl)-β-CD, capillary tube, 50 µm ID × 76 cm, applied voltage, 30 kV, UV-detection at 200 nm, temperature, 30 °C. (Reproduced by permission of Elsevier Science Publishers. Copyright (1993) from M.W.F. Nielen, *Trends Anal. Chem.*, **12**, 345–356.[193])

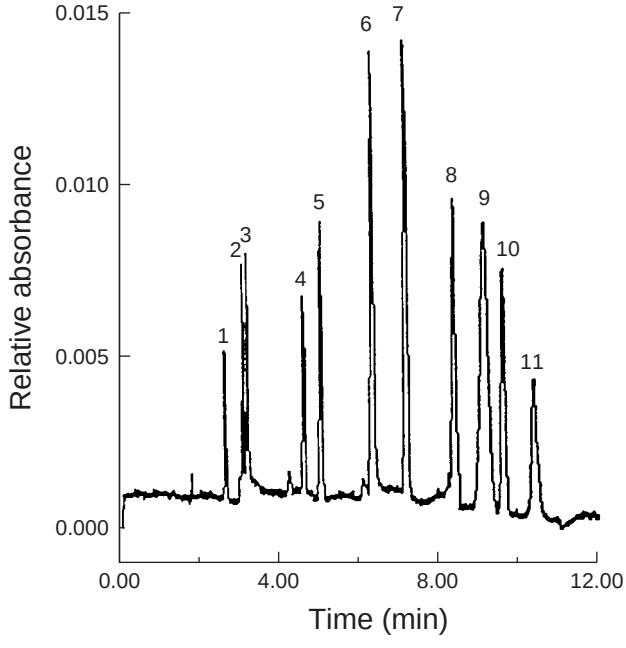

Figure 8 Separation of 11 herbicides from distilled water by MEKC. Peaks (analyte concentration, 1–2 mg L^{-1}): (1) tribenuron, (2) chlorsulfuron, (3) metsulfuron, (4) paraquat, (5) simazine, (6) atrazine, (7) linuron, (8) terbuthylazine, (9) alachlor, (10) metolachlor, (11) trifluralin. Conditions: separation media, 0.030 M borate buffer (pH 8.0) with 0.030 M SDS; capillary tube, 75 µm ID × 50 cm, applied voltage, 25 kV, UV-detection at 214 nm, temperature, 30 °C. (Reproduced by permission of Elsevier Science Publishers. Copyright (1996) from G. Dinelli, A. Vicari, P. Catizone, *J. Chromatogr. A*, **733**, 337–347.[195])

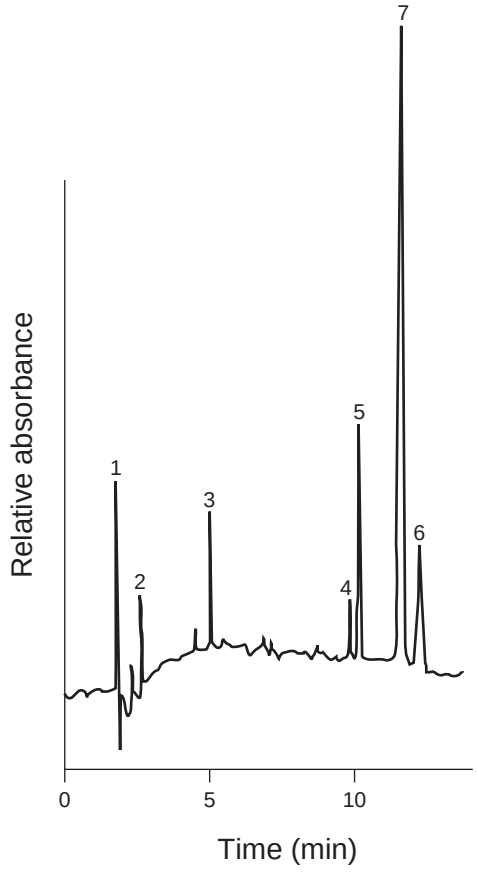

Figure 9 Electropherogram of four organometallic compounds in the extract from 1 L rain water by MEKC. Peaks: (1) methanol, (2) trimethyltin (49.8 ng mL^{-1}), (3) trimethyllead (18.4 ng mL^{-1}), (4) unknown, (5) triethyllead (50.0 ng mL^{-1}), (6) tributyltin (28.8 ng mL^{-1}), (7) NaDDC (sodium diethyldithiocarbamate) + unknown. Conditions: separation media, 0.030 M dihydrogenphosphate/0.015 M borate buffer (pH 7.65) with 0.050 M SDS, capillary tube, 50 µm ID × 60 cm (50 cm to the detector), hydrodynamic injection, 10 cm × 10–20 s, applied voltage, 20 kV, UV-detection at 200 nm. (Reproduced by permission of Marcel Dekker Inc from K. Li, S.F.Y. Li, H.K. Lee, *J. Liq. Chromatogr.*, **18**, 1325–1347 (1995).[204])

the separation by MEKC of the most common organolead and organotin compounds (trimethyltin, trimethyllead, triethyllead and tributyltin) in the extracts from 1 L of rain water samples after disk SPE extraction. In addition, CDMEKC was also applied to separate two organolead and two organoselenium compounds from water samples using LLE for sample enrichment.[148]

PAHs have been analyzed in contaminated soil samples by different kinds of EKC and detection systems. Thus, several PAHs were determined in an extract of creosote-contaminated soil by MEKC with a bile salt and UV detection at 214 nm.[61] Quantitation of seven PAHs in soil samples contaminated with a mixture of standard PAHs and machine oil was carried out by MEKC with SDS and diode-array detection[106] and 16 USEPA priority PAHs contained in an extract obtained by supercritical fluid extraction (SFE) of a contaminated soil sample were determined by CDEKC using a mixture of charged and neutral CDs and LIF detection ($\lambda_{ex} = 325$ nm).[157]

CZE has been used to determine several *chlorophenols* in different water samples.[180,188] The fact that phenols respond to a sensitive detection method such as electrochemical detection has allowed determination of seven chlorophenol isomers and three neutral phenols in industrial wastewater by CZE with on-column electrochemical detection (see Figure 10).[188] LODs in the picomole range were achieved by thermostating the separation capillary and no interference from the impurities present in industrial wastewater samples were observed, using only a simple LLE with chloroform–diethyl ether. Therefore, the use of an electrochemical detector provides excellent sensitivity and selectivity without derivatization. Thus, the separation of 11 USEPA priority phenols by CZE with on-column amperometric detection has been successfully applied to the analysis of industrial wastewaters.[187] In order to minimize the large electrophoretic currents, which seriously interfere with amperometric detection, CHES was used as the running buffer because of its

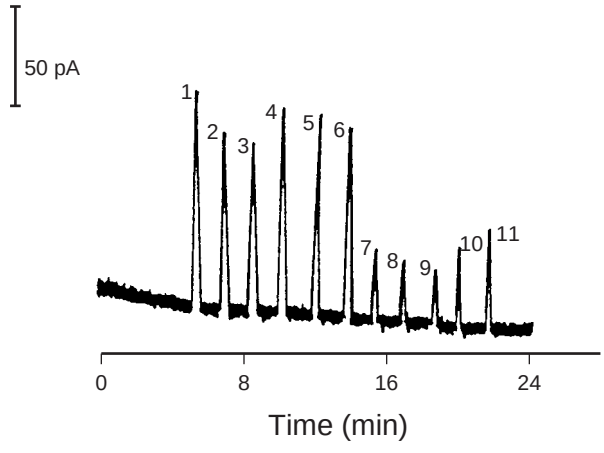

Figure 10 Electropherogram of an industrial wastewater sample obtained by CZE with electrochemical detection. Peaks: (1) phenol, (2) 2-chlorophenol (concentration, 50 ng mL^{-1}), (3) 4-chlorophenol, (4) 2,4-dichlorophenol, (5) 2,6-dichlorophenol, (6) *o*-phenylphenol, (7) catechol, (8) 2,4,6-trichlorophenol, (9) 2,3,4,6-tetrachlorophenol, (10) 4,5,6-trichloroguiacol, and (11) pentachlorophenol. Conditions: separation media, 0.045 M dihydrogenphosphate/0.015 M borate buffer (pH 8.0), capillary tube, 25 µm ID × 65 cm (30 cm to the detector), electrokinetic injection, 20 kV × 30 s, applied voltage, 20 kV; amperometric detection using a carbon fiber at +1.4 V versus SCE (standard calomel electrode); temperature, 20–25 °C. (Reproduced by permission of Elsevier Science Publishers. Copyright (1990) from C.D. Gaitonde, P. Pathak, *J. Chromatogr.*, **514**, 389–393.[188])

zwitterionic nature. The analysis of similar samples was achieved by CZE with indirect laser-induced fluorescence detection.[186] This detection technique involves the addition of a noninteracting and fluorescing ion to the running buffer to create a constant fluorescence background and when a charged analyte is present, it displaces the fluorescing ion of the same charge owing to local charge neutrality, resulting in a decreased background signal even though the analyte does not absorb or fluoresce.

Finally, other pollutants separated by CE were *sulfonates*, which were analyzed by MEKC with UV detection in spiked river samples.[205] or by CZE with LIF detection in unspiked river samples.[205]

In general, Table 3 shows that the analysis of pollutants in environmental samples can be resolved by two means: (i) sample preconcentration with off-line SPE (principally with a C_{18} disk), or (ii) use of detectors with enough sensitivity (LIF or amperometric detection). In both cases, it is possible to obtain a LOD of about 10^{-9} M which is adequate for the determination of pollutants in several environmental matrices or samples.

4 CONCLUSIONS

CE is a newly developed separation technique, which has demonstrated its enormous capability and potential for separating ionic and neutral solutes owing to its simplicity and high efficiency. The first option for analyzing ionic or ionizable compounds by CE is CZE with or without additives to increase separation selectivity. When this selectivity is poor or the compounds to be analyzed are not ionizable, separation by CE should be achieved by means of MEKC with or without additives to improve separation resolution. Finally, other kinds of EKC or CEC can be used to enhance the separation or decrease the analysis time.

CE is very suitable for those analytes that are not amenable to GC (polar and acid or basic pollutants, e.g. modern pesticides in aqueous samples) or when existing HPLC methods do not offer sufficient separation power. Many efficient and rapid CE separations of standards, including compounds of environmental interest, have been described. Despite the excellent prospects for CE, the analysis of pollutants in environmental samples where these compounds are at trace level (governmental regulations require determination of lower and lower levels of pollutants in environmental samples) requires the use of enrichment procedures or highly sensitivity detectors. Off-line extraction, preconcentration techniques and special injection methods are being developed. A simple clean-up, followed by a large-volume injection and a high-efficiency separation in a CE system, should be adequate to fulfill the requirements of real-life environmental applications. Several research groups are developing their own highly sensitive detectors such as LIF, which has become popular mainly because of its capability to provide extremely high sensitivity (10^{-12} M). Unfortunately, this type of detector is not generally applicable in environmental analysis. Many interesting pollutants cannot be derivatized using any of the few fluorescence labels which match the particular laser lines. Other sensitive detection systems, such as laser-induced resonance energy transfer,[211] chemiluminescence[212] and inductively coupled plasma MS[213–215] are being developed. With respect to MS which is one of the most powerful detection systems, current research is devoted to developing new CE/MS interfaces in order to increase the possibility of CE/MS being able to solve problems in environmental analysis.

ACKNOWLEDGMENTS

The authors thank the Comunidad Autónoma de Madrid (Spain) for project 07M/0049/98. They also thank C. Marina for linguistic assistance.

ABBREVIATIONS AND ACRONYMS

ANDSA	7-Aminonaphthalene-1,3-disulfonic Acid
BBMA	Butyl Acrylate–Butyl Methacrylate–Methacrylic Acid
Brij 35	Polyoxyethylenelauryl Ether
BTMA	Benzyltrimethylammonium
BTP	Bis-tris-propane
CD	Cyclodextrin
CDCZE	Cyclodextrin-modified Capillary Zone Electrophoresis
CDEKC	Cyclodextrin Electrokinetic Chromatography
CDMEKC	Cyclodextrin-modified Micellar Electrokinetic Chromatography
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CHES	2-(*N*-Cyclohexylamino) Ethanesulfonic Acid
CM-β-CD	Carboxymethyl-β-cyclodextrin
CMC	Critical Micelle Concentration

CTAB	Cetyltrimethylammonium Bromide
CZE	Capillary Zone Electrophoresis
DBTD	Disodium 5,12-Bis(dodecyl-oxymethyl)-4,7,10,13-(tetraosa)-1,16-hexadecanedisulfonate
Deoxy Big CHAP	*N*,*N*-Bis-(3-D-gluconamido-propyl)-deoxycholamide
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl Sulfoxide
DTAC	Dodecyltrimethylammonium Chloride
EKC	Electrokinetic Chromatography
EOF	Electroosmotic Flow
FASS	Field-amplified Sample Stacking
GC	Gas Chromatography
HDB	Hexadimetrine Bromide
HIEKC	Hydrophobic Interaction Electrokinetic Chromatography
HP-β-CD	Hydroxypropyl-β-cyclodextrin
HPLC	High-performance Liquid Chromatography
ID	Internal Diameter
ITP	Isotachophoresis
LIF	Laser Induced Fluorescence
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
M-β-CD	Methyl-β-cyclodextrin
MEEKC	Microemulsion Electrokinetic Chromatography
MEGA 8	Octanoyl-*N*-methylglucamide
MEGA 9	Nonanoyl-*N*-methylglucamide
MEGA 10	Decanoyl-*N*-methylglucamide
MEKC	Micellar Electrokinetic Chromatography
MHEC	Methylhydroxyethylcellulose
MS	Mass Spectrometry
NaDDC	Sodium Diethyldithiocarbamate
NG	*n*-Nonyl-β-D-glucopyranoside
OG	Octyl-β-D-glucopyranoside
OM	Octyl-β-D-maltopyranoside
OS	*n*-Octanoylsucrose
OTAC	Octyltrimethylammonium Chloride
OTC	Open Tubular Column
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PCC	Packed Capillary Column
PCDD	Polychlorodibenzo-*p*-dioxin
PEG	Polyethylene Glycol
PNA	*N*-Phenyl-1-naphthylamine
Poly(D-SUV)	Poly(sodium *N*-undecylenyl-D-valinate)
PVP	Polyvinylpyrrolidone
SB-β-CD	Sulfobutyl Ether-β-cyclodextrin
SBD	Starburst Dendrimers
SC	Sodium Cholate
SCE	Standard Calomel Electrode
SDC	Sodium Deoxycholate
SDS	Sodium Dodecyl Sulfate
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction
SUA	Sodium 10-Undecylenate
SUS	Sodium 10-Undecylsulfate
TBA	Tetrabutylammonium Hydrogensulfate
TES	*N*-Tris(hydroxymethyl)methyl-2-aminoethane Sulfonic Acid
TM-β-CD	Tri-*o*-methyl-β-cyclodextrin
TRIS	Tris(hydroxymethyl)-aminomethane
TTAC	Tetradecyltrimethylammonium Chloride
USEPA	United States Environmental Protection Agency
UV	Ultraviolet

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis

Environment: Water and Waste cont'd **(Volume 4)**
Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Waste Extraction Procedures

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Herbicide Residues in Biota, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis • Sampling and Sample Preparation in Process Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Micellar Electrokinetic Chromatography

REFERENCES

1. S.F.Y. Li, *Capillary Electrophoresis. Principles, Practice, and Applications*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 23, 1992.
2. N.A. Guzman, *Capillary Electrophoresis Technology*, Chromatographic Science Series, Marcel Dekker, New York, Vol. 24, 1993.
3. D.R. Baker, *Capillary Electrophoresis*, J. Wiley & Sons, New York, 1995.
4. H.J. Issaq, 'Comparison of Chiral Separations in Capillary Zone Electrophoresis with Other Methods', *Instrum. Sci. Technol.*, **22**, 119–149 (1994).
5. M.W.F. Nielen, 'The Potential of Capillary Electrophoresis in Environmental Analysis', in *Environmental Analysis: Techniques, Applications and Quality Assurance*, ed. D. Barceló, Elsevier, Amsterdam, 607–633, 1993.
6. J. Snopek, I. Jelínek, E. Smolková-Keulemansová, 'Chiral Separation by Analytical Electromigration Methods', *J. Chromatogr. A*, **609**, 1–17 (1992).
7. M. Novotny, H. Soini, M. Stefansson, 'Chiral Separation Through Capillary Electromigration Methods', *Anal. Chem.*, **66**, 646A–655A (1994).
8. M.L. Marina, A.L. Crego, 'Capillary Electrophoresis: A Good Alternative for the Separation of Chiral Compounds of Environmental Interest', *J. Liq. Chromatogr. Related Technol.*, **20**, 1337–1365 (1997).
9. J. Wu, M.K. Wong, H.K. Lee, C.N. Ong, 'Orthogonal Array Design for Optimizing the Capillary Zone Electrophoretic Analysis of Heterocyclic Amines', *J. Chromatogr. Sci.*, **34**, 139–145 (1996).
10. Z. El Rassi, 'Capillary Electrophoresis of Pesticides', *Electrophoresis*, **18**, 2465–2481 (1997).
11. L. Song, Z. Xu, J. Kang, J. Cheng, 'Analysis of Environmental Pollutants by Capillary Electrophoresis with Emphasis on Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **780**, 297–328 (1997).
12. J.W. Jorgenson, K.D. Luckas, 'Capillary Zone Electrophoresis', *Science*, **222**, 266–272 (1983).
13. J.W. Jorgenson, 'Capillary Zone Electrophoresis in Open-tubular Capillaries', *Trends Anal. Chem.*, **3**, 51–54 (1984).
14. K. Altria, C. Simpson, 'Measurement of Electroendosmotic Flows in High-voltage Capillary Zone Electrophoresis', *Anal. Proc.*, **23**, 453–454 (1986).
15. J.H. Knox, 'Terminology and Nomenclature in Capillary Electroseparation Systems', *J. Chromatogr. A*, **680**, 3–13 (1994).
16. W. Friendl, E. Kenndler, 'Resolution as a Function of the pH of the Buffer Based on the Analyte Charge Number for Multivalent Ions in Capillary Zone Electrophoresis Without Electroosmotic Flow: Theoretical Prediction and Experimental Evaluation', *Anal. Chem.*, **65**, 2003–2009 (1993).
17. S. Fujiwara, S. Honda, 'Effect of Addition of Organic Solvent on the Separation of Positional Isomers in High-voltage Capillary Zone Electrophoresis', *Anal. Chem.*, **59**, 487–490 (1987).
18. A. Guttman, A. Paulus, A.S. Cohen, N. Grinberg, B.L. Karger, 'Use of Complexing Agents for Selective Separation in High Performance Capillary Electrophoresis. Chiral Resolution via Cyclodextrins Incorporated Within Polyacrylamide Gel Columns', *J. Chromatogr.*, **448**, 41–53 (1988).
19. Y. Esaka, Y. Yamaguchi, K. Kano, M. Goto, H. Haraguchi, J. Takahashi, 'Separation of Hydrogen-bonding Donors in Capillary Electrophoresis using Polyesters as Matrix', *Anal. Chem.*, **66**, 2441–2445 (1994).
20. P. Blatny, C.H. Fisher, A. Rizzi, E. Kenndler, 'Linear Polymers Applied as Pseudo-phases in Capillary Zone Electrophoresis of Azo Compounds used

as Textile Dyes', *J. Chromatogr. A*, **717**, 157–166 (1995).
21. D. Sybilska, E. Smolkova-Keulemansova, 'Applications of Inclusion Compounds in Chromatography', in *Inclusion Compounds*, eds. J.L. Atwood, J.E.D. Davies, D.D. Mac Nicol, Academic Press, London, Vol. 3, 174–243, 1984.
22. S.M. Masselter, A.J. Zemann, 'Influence of Organic Solvents in Coelectroosmotic Capillary Electrophoresis of Phenols', *Anal. Chem.*, **67**, 1047–1053 (1995).
23. S.M. Masselter, A.J. Zemann, 'Influence of Buffer Electrolyte pH on the Migration Behavior of Phenolic Compounds in Coelectroosmotic Capillary Electrophoresis', *J. Chromatogr. A*, **693**, 359–365 (1995).
24. D.A. Oxspring, W.F. Smyth, R. Marchant, 'Comparison of Reversed-polarity Capillary Electrophoresis and Adsorptive Stripping Voltammetry for the Detection and Determination of Reactive Textile Dyes', *Analyst*, **120**, 1995–2000 (1995).
25. J. Cai, Z. El Rassi, 'Capillary Zone Electrophoresis of Two Cationic Herbicides, Paraquat and Diquat', *J. Liq. Chromatogr.*, **15**, 1193–1200 (1992).
26. F. García, J. Henion, 'Fast Capillary Electrophoresis Ion Spray Mass Spectrometric Determination of Sulfonylureas', *J. Chromatogr.*, **606**, 237–247 (1992).
27. M.W.F. Nielen, 'Enantio-separation of Phenoxy Acid Herbicides using Cyclodextrin Modified Capillary Zone Electrophoresis', *J. Chromatogr.*, **637**, 81–90 (1993).
28. G.A. Pianetti, M. Taverna, A. Baillet, G. Mahuzier, D. Baylocq-Ferrier, 'Determination of Alkylphosphonic Acids by Capillary Zone Electrophoresis using Indirect UV detection', *J. Chromatogr.*, **630**, 371–377 (1993).
29. J. Cai, Z. El Rassi, 'On-line Preconcentration of Triazine Herbicides with Tandem Octadecyl Capillaries–Capillary Zone Electrophoresis', *J. Liq. Chromatogr.*, **15**, 1179–1192 (1992).
30. A.W. Garrison, P. Schmitt, A. Kettrup, 'Separation of Phenoxy Acid Herbicides and their Enantiomers by High-performance Capillary Electrophoresis', *J. Chromatogr. A*, **688**, 317–327 (1994).
31. Y. Mechref, Z. El Rassi, 'Capillary Electrophoresis of Herbicides: Precolumn Derivatization of Chiral and Achiral Phenoxy Acid Herbicides with a Fluorescent Tag for Cyclodextrin Modified Micellar Electrokinetic Chromatography', *Anal. Chem.*, **68**, 1771–1777 (1996).
32. S.K. Yeo, C.P. Ong, S.K.Y. Li, 'Optimization of HP-CE of Plant Growth Regulators Using the Overlapping Resolution Mapping Scheme', *Anal. Chem.*, **63**, 2222–2225 (1991).
33. S.K. Yeo, H.K. Lee, S.F.Y. Li, 'Separation of Plant Growth Regulators by Capillary Electrophoresis', *J. Chromatogr.*, **594**, 335–340 (1992).
34. G. Li, C. Locke, 'Separation of the Eleven Priority Pollutant Phenols by Capillary Zone Electrophoresis', *J. Chromatogr.*, **669**, 93–102 (1995).
35. C. Lin, W. Lin, W. Chiou, 'Migration Behavior and Optimization of Selectivity of Dichlorophenols in Capillary Zone Electrophoresis', *J. Chromatogr. A*, **705**, 325–333 (1995).
36. S.C. Smith, M.G. Khaledi, 'Prediction of the Migration Behavior of Organic Acids in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **65**, 193–198 (1993).
37. M.W.F. Nielen, 'Separation of Phenylenediamine Isomers by Capillary Zone Electrophoresis', *J. Chromatogr.*, **625**, 387–391 (1992).
38. J.C. Jacquier, C. Rony, P.L. Desbene, 'Computer Assisted pH Optimization for the Separation of Geometric Isomers in Capillary Zone Electrophoresis', *J. Chromatogr. A*, **652**, 337–345 (1993).
39. W.C. Brumley, 'Qualitative Analysis of Environmental Samples for Aromatic Sulfonic Acids by High-performance Capillary Electrophoresis', *J. Chromatogr.*, **603**, 267–272 (1992).
40. P.L. Desbene, C. Rony, B. Desmazieres, J.C. Jacquier, 'Analysis of Alkylaromatic Sulfonates by HP-CE', *J. Chromatogr.*, **608**, 375–383 (1992).
41. H. Salimi-Moosavi, R.M. Cassidy, 'Application of Nonaqueous Capillary Electrophoresis to the Separation of Long-chain Surfactants', *Anal. Chem.*, **68**, 293–299 (1996).
42. F. Han, J.L. Fasching, P.R. Brown, 'Speciation of Organotin Compounds by Capillary Electrophoresis Using Indirect Ultraviolet Detection', *J. Chromatogr. B*, **669**, 103–112 (1995).
43. E. Pobozy, B. Glód, J. Kaniewska, M. Trojanowicz, 'Determination of Triorganotin Compounds by Ion Chromatography and Capillary Electrophoresis with Preconcentration Using Solid-phase Extraction', *J. Chromatogr. A*, **718**, 329–338 (1995).
44. A.M. Carro-Diaz, R.A. Lorenzo-Ferreira, R. Cela-Torrijos, 'Capillary Electrophoresis of Methylmercury with Injection by Sample Stacking', *J. Chromatogr. A*, **730**, 345–351 (1996).
45. S. Razee, A. Tamura, T. Masujima, 'Improvement in the Determination of Food Additive Dyestuffs by Capillary Electrophoresis using β-Cyclodextrin', *J. Chromatogr. A*, **715**, 179–188 (1995).
46. M. Masár, D. Kaniansky, V. Madajová, 'Separation of Synthetic Food Colorants by Capillary Zone Electrophoresis in a Hydrodynamically Closed Separation Compartment', *J. Chromatogr. A*, **724**, 327–336 (1996).
47. C.P. Ong, C.L. Ng, N.C. Chong, H.K. Lee, S.F.Y. Li, 'Analysis of Priority Substituted Phenols by Micellar Electrokinetic Chromatography', *Environ. Monit. Assess.*, **19**, 93–103 (1991).
48. H.K. Lee, S.F.Y. Li, Y.H. Tay, 'High-performance Liquid Chromatography of Eleven Priority Substituted Phenols by Isocratic Elution', *J. Chromatogr.*, **438**, 429–432 (1988).
49. A.L. Crego, M.L. Marina, 'Capillary Zone Electrophoresis Versus Micellar Electrokinetic Chromatography in

the Separation of Phenols of Environmental Interest', *J. Liq. Chromatogr. Related Technol.*, **20**, 1–20 (1996).
50. A.D. Corcia, M. Marchetti, R. Samperi, A. Marcomino, 'Liquid Chromatographic Determination of Linear Alkylbenzenesulfonates', *Anal. Chem.*, **63**, 1179–1182 (1991).
51. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open-tubular Capillaries', *Anal. Chem.*, **56**, 111–113 (1984).
52. S. Terabe, K. Otsuka, T. Ando, 'Electrokinetic Chromatography with Micellar Solution and Open-tubular Capillaries', *Anal. Chem.*, **57**, 834–841 (1985).
53. K. Otsuka, S. Terabe, T. Ando, 'Electrokinetic Chromatography with Micellar Solutions. Retention and Separation of Chlorinated Phenols', *J. Chromatogr.*, **348**, 39–47 (1985).
54. S. Terabe, K. Otsuka, T. Ando, 'Band Broadening in Electrokinetic Chromatography with Micellar Solutions and Open-tubular Columns', *Anal. Chem.*, **61**, 251–260 (1989).
55. J.P. Foley, 'Optimization of Micellar Electrokinetic Chromatography', *Anal. Chem.*, **62**, 1302–1308 (1990).
56. J. Vindevogel, P. Sandra, *Introduction to Micellar Electrokinetic Chromatography*, Hüthig, Heidelberg, 1992.
57. S. Terabe, 'Electrokinetic Chromatography: an Interface between Electrophoresis and Chromatography', *Trends Anal. Chem.*, **8**, 129–134 (1989).
58. S. Yang, M.G. Khaledi, 'Chemical Selectivity in Micellar Electrokinetic Chromatography: Characterization of Solute–Micelle Interactions for Classification of Surfactants', *Anal. Chem.*, **67**, 499–510 (1995).
59. A.T. Balchunas, M.J. Sepaniak, 'Extension of Elution Range in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **59**, 1466–1470 (1987).
60. J. Gorse, A.T. Balchunas, D.F. Swaile, M.J. Sepaniak, 'Effects of Organic Mobile Phase Modifiers in Micellar Electrokinetic Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**, 554–559 (1988).
61. W.C. Brumley, W.J. Jones, 'Comparison of Micellar Electrokinetic Chromatography with Capillary Gas Chromatography in the Separation of Phenols, Anilines and Polycyclic Aromatic Hydrocarbons. Potential Field-screening Applications of Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **680**, 163–173 (1994).
62. S. Terabe, Y. Ishihama, H. Nishi, T. Fukuyama, K. Otsuka, 'Effect of Urea Addition in Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **545**, 359–368 (1991).
63. M.A. García, J.C. Díez-Masa, M.L. Marina, 'Determination of Solute–Micelle Association Constants for a Group of Benzene Derivatives and Polycyclic Aromatic Hydrocarbons with Sodium Dodecyl Sulfate by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **732**, 345–359 (1996).
64. A.T. Balchunas, M.J. Sepaniak, 'Gradient Elution on Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **60**, 617–621 (1988).
65. R.O. Cole, M.J. Sepaniak, W.L. Hinze, J. Gorse, K. Oldiges, 'Bile Salts Surfactants in Micellar Electrokinetic Chromatography. Application to Hydrophobic Molecule Separations', *J. Chromatogr.*, **557**, 113–123 (1991).
66. S. Terabe, Y. Miyashita, O. Shibata, E.R. Barnhart, L.R. Alexander, D.G. Patterson, B.L. Karger, K. Hosoya, N.J. Tanaka, 'Separation of Highly Hydrophobic Compounds by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **516**, 23–31 (1990).
67. H. Nishi, M. Matsuo, 'Separation of Corticosteroids and Aromatic Hydrocarbons by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Liq. Chromatogr.*, **14**, 973–986 (1991).
68. H. Nishi, T. Fukuyama, S. Terabe, 'Chiral Separation by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **553**, 503–516 (1991).
69. Y.F. Yik, C.P. Ong, S.B. Khoo, H.K. Lee, S.F.Y. Li, 'Separation of Polycyclic Aromatic Hydrocarbons by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **589**, 333–338 (1992).
70. S. Terabe, Y. Miyashita, Y. Ishihama, O. Shibata, 'Cyclodextrin-modified Micellar Electrokinetic Chromatography: Separation of Hydrophobic and Enantiomeric Compounds', *J. Chromatogr.*, **636**, 47–55 (1993).
71. C.L. Copper, M.J. Sepaniak, 'Cyclodextrin-modified Micellar Electrokinetic Chromatography Separations of Benzopyrene Isomers: Correlation with Computationally Derived Host–Guest Energies', *Anal. Chem.*, **66**, 147–154 (1994).
72. H.T. Rasmussen, L.K. Goebel, H.N. McNair, 'Cyclodextrin-modified Micellar Electrokinetic Chromatography Employing Sodium Alkyl Sulfates and Brij 35', *J. Chromatogr.*, **517**, 549–559 (1990).
73. L.G. Song, Q.Y. Ou, W.L. Yu, G.Z. Li, 'Separation of Six Phenylureas and Chlorsulfuron Standards by Micellar Electrokinetic Chromatography and Microemulsion Electrokinetic Chromatography', *J. Chromatogr. A*, **699**, 371–382 (1995).
74. L. Song, S. Zhang, Q. Ou, W. Yu, 'Electrokinetic Chromatography of Diterpenoid Alkaloids from *Aconitum Sinomontanum* Nakai', *J. Microcolumn Sep.*, **7**, 123–126 (1995).
75. D.E. Burton, M.J. Sepaniak, M.P. Maskarinec, 'Evaluation of the Use of Various Surfactants in Micellar Electrokinetic Chromatography', *J. Chromatogr. Sci.*, **25**, 514–518 (1987).
76. D. Attwood, A.T. Florence, *Surfactant Systems*, Chapman & Hall, London, 185–196, 1983.
77. A.R. Campanelli, S.C. De Sanctis, E. Chiessi, M. D'Alagni, E. Giglio, L. Scaramuzza, 'Sodium Glyco- and

Taurodeoxycholate: Possible Helical Models for Conjugated Bile Salt Micelles', *J. Phys. Chem.*, **93**, 1536–1542 (1989).
78. S. Terabe, M. Shibata, Y. Miyashita, 'Chiral Separation by Micellar Electrokinetic Chromatography with Bile Salt Micelles', *J. Chromatogr.*, **480**, 403–411 (1989).
79. H. Nishi, T. Fukuyama, M. Matsuo, S. Terabe, 'Chiral Separation of Optical Isomeric Drugs Using Micellar Electrokinetic Chromatography and Bile Salts', *J. Microcolumn Sep.*, **1**, 234–241 (1989).
80. H. Nishi, T. Fukuyama, M. Matsuo, S. Terabe, 'Chiral Separation of Diltiazem, Trimetiphinol and Related Compounds by Micellar Electrokinetic Chromatography with Bile Salts', *J. Chromatogr.*, **515**, 233–243 (1990).
81. R.O. Cole, M.J. Sepaniak, W.L. Hinze, 'Optimization of Binaphthyl Enantiomer Separations by Capillary Zone Electrophoresis Using Mobile Phase Containing Bile Salts and Organic Solvent', *J. High Resolut. Chromatogr.*, **13**, 579–582 (1990).
82. H. Nishi, T. Fukuyama, M. Matsuo, S. Terabe, 'Separation and Determination of the Ingredients of a Cold Medicine by Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **498**, 313–323 (1990).
83. J.T. Smith, W. Nashabeth, Z. El Rassi, 'Micellar Electrokinetic Chromatography with in situ Charged Micelles. 1 Evaluation of *N*-D-gluco-*N*-methylalkanamide Surfactants as Anionic Borate Complexes', *Anal. Chem.*, **66**, 1119–1133 (1994).
84. J.T. Smith, Z. El Rassi, 'Micellar Electrokinetic Chromatography with in situ Charged Micelles: Evaluation and Comparison of Octylmaltoside and Octylsucrose Surfactants as Anionic Borate Complexes in the Separation of Herbicides', *J. Microcolumn Sep.*, **6**, 127–138 (1994).
85. J.T. Smith, Z. El Rassi, 'Micellar Electrokinetic Chromatography with in situ Charged Micelles. IV. Influence of the Nature of the Alkylglycoside Surfactant', *J. Chromatogr. A*, **685**, 131–143 (1994).
86. J.T. Smith, Z. El Rassi, 'Micellar Electrokinetic Chromatography with in situ Charged Micelles: 5. Evaluation of Background Fluorophores for Indirect Fluorescence Detection', *J. Cap. Elec.*, **1**, 136–143 (1994).
87. Y. Mechref, Z. El Rassi, 'Micellar Electrokinetic Chromatography with in situ Charged Micelles. VI. Evaluation of Novel Chiral Micelles Consisting of Steroidal–Glycoside Surfactant–Borate Complexes', *J. Chromatogr. A*, **724**, 285–296 (1996).
88. D. Crosby, Z. El Rassi, 'MEKC with Cationic Surfactants', *J. Liq. Chromatogr.*, **16**, 2161–2187 (1993).
89. M. Tanaka, T. Ishida, T. Araki, A. Masuyama, Y. Nakatsuji, M. Okahara, S. Terabe, 'Double Chain Surfactant as a New and Useful Micelle-forming Reagent for MEKC', *J. Chromatogr.*, **648**, 469–473 (1993).
90. H. Ozaki, S. Terabe, A. Ichihara, 'Micellar Electrokinetic Chromatography Using High-molecular Surfactants. Use of Butyl Acrylate-Butyl Methacrylate-Methacrylic Acid Copolymers Sodium Salts as Pseudo-stationary Phases', *J. Chromatogr. A*, **680**, 117–123 (1994).
91. C.P. Palmer, M.Y. Khaledi, H.M. McNair, 'A Monomolecular Pseudo-stationary Phase for Micellar Electrokinetic Capillary Chromatography', *J. High Resolut. Chromatogr.*, **15**, 756–762 (1992).
92. C.P. Palmer, S. Terabe, 'A Novel Sulfate Polymer as a Pseudo-stationary Phase for Micellar Electrokinetic Chromatography', *J. Microcolumn Sep.*, **8**, 115–121 (1996).
93. N. Tanaka, T. Tanigawa, K. Hosoya, K. Kimata, T. Araki, S. Terabe, 'Starburst Dendrimers as Carriers in Electrokinetic Chromatography', *Chem. Lett.*, 959–962 (1992).
94. P.G.H.M. Muijselaar, H.A. Claessens, C.A. Cramers, J.F.G.A. Jansen, E.W. Meijer, E.M.M. de Brabander van den Berg, S. Van der Wal, 'Dendrimers as Pseudo-stationary Phases in Electrokinetic Chromatography', *J. High Resolut. Chromatogr.*, **18**, 121–123 (1995).
95. N. Tanaka, T. Fukutome, T. Tanigawa, K. Hosoya, K. Kimata, T. Araki, K.K. Unger, 'Structural Selectivity Provided by Starburst Dendrimers as Pseudo-stationary Phases in Electrokinetic Chromatography', *J. Chromatogr. A*, **699**, 331–341 (1995).
96. K. Bachmann, A. Bazzanella, Y. Haag, K.Y. Han, R. Arnecke, V. Bohmer, W. Vogt, 'Resorcarenes as Pseudo-stationary Phases with Selectivity for Electrokinetic Chromatography', *Anal. Chem.*, **67**, 1722–1726 (1995).
97. M.G. Khaledi, S.C. Smith, J.K. Strasters, 'Micellar Electrokinetic Chromatography of Acidic Solutes: Migration Behavior and Optimization Strategies', *Anal. Chem.*, **63**, 1820–1830 (1991).
98. G.M. Janini, H.J. Issaq, 'Micellar Electrokinetic Chromatography: Basic Considerations and Current Trends', *J. Liq. Chromatogr.*, **15**, 927–960 (1992).
99. R.D. Holland, M.J. Sepaniak, 'Qualitative Analysis of Mycotoxins Using Micellar Electrokinetic Chromatography', *Anal. Chem.*, **65**, 1140–1146 (1993).
100. Z. Liu, P. Sam, S.R. Sirimanne, P.C. McClure, J. Grainger, D.G. Patterson, 'Field-amplified Sample Stacking in Micellar Electrokinetic Chromatography for On-column Sample Concentration of Neutral Molecules', *J. Chromatogr. A*, **673**, 125–132 (1994).
101. K. Jinno, Y. Sawada, 'Relationships Between Capacity Factors and Hydrophobicity of Polycyclic Aromatic Hydrocarbons in Cyclodextrin Modified Micellar Electrokinetic Chromatography Using Surface Treated Capillaries', *J. Liq. Chromatogr.*, **18**, 3719–3727 (1995).
102. K. Jinno, Y. Sawada, 'Relationships Between Capacity Factors and Hydrophobicity of Polycyclic Aromatic Hydrocarbons in Cyclodextrin Modified Micellar Electrokinetic Chromatography', *J. Cap. Elec.*, **1**, 106–111 (1994).
103. B. Jiménez, D.G. Patterson, J. Grainger, Z. Liu, M.J. González, M.L. Marina, 'Enhancement of the Separation

Selectivity of a Group of Polycyclic Aromatic Hydrocarbons Using Mixed Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **792**, 411–418 (1997).
104. U. Krismann, W. Kleiböhmer, 'Separation of Hydroxilated Polycyclic Aromatic Hydrocarbons by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **774**, 193–201 (1997).
105. T. Kaneta, T. Yamashita, T. Imasaka, 'Separation of Polycyclic Aromatic Hydrocarbons by Micellar Electrokinetic Chromatography with Laser Fluorescence Detection', *Anal. Chim. Acta*, **299**, 371–375 (1995).
106. O. Brüggemann, R. Freitag, 'Determination of Polycyclic Aromatic Hydrocarbons in Soil Samples by Micellar Electrokinetic Chromatography with Photodiode-array Detection', *J. Chromatogr. A*, **717**, 309–324 (1995).
107. J. Wang, J.M. Warner, 'Combined Polymerized Chiral Micelle and γ-Cyclodextrin for Chiral Separation in Capillary Electrophoresis', *J. Chromatogr. A*, **711**, 297–304 (1995).
108. J. Grainger, P.C. McClure, Z. Liu, B. Botero, S. Sirimanne, D.G. Patterson Jr, M. Sewer, C. Gillyard, K. Kimata, K. Hosoya, T. Araki, N. Tanaka, S. Terabe, 'Isomer Identification of Chlorinated Dibenzo-*p*-dioxins by Orthogonal Spectroscopic and Chromatographic Techniques', *Chemosphere*, **32**, 13–23 (1996).
109. I. Benito, J.M. Saz, M.L. Marina, J. Jiménez-Barbero, M.J. González, J.C. Díez-Masa, 'Micellar Electrokinetic Capillary Chromatographic Separation of Polychlorinated Biphenyl Congeners', *J. Chromatogr. A*, **778**, 77–85 (1997).
110. J. Pérez, A.L. Crego, M.L. Marina, J.M. Saz, 'Study of the Retention Behavior of Highly Hydrophobic Compounds in Micellar Electrokinetic Chromatography Using Hexadecyltrimethylammonium Bromide Micelles and γ-Cyclodextrin in the Separation Buffer. Some Mechanistic Considerations', unpublished results.
111. M.L. Marina, I. Benito, J.C. Díez-Masa, M.J. González, 'Chiral Separation of Polychlorinated Biphenyls by Micellar Electrokinetic Chromatography with γ-Cyclodextrin as Modifier in the Separation Buffer', *Chromatographia*, **42**, 269–272 (1996).
112. M.L. Marina, I. Benito, J.C. Díez-Masa, M.J. González, 'Separation of Chiral Polychlorinated Biphenyls by Micellar Electrokinetic Chromatography Using β- and γ-Cyclodextrin Mixtures in the Separation Buffer', *J. Chromatogr. A*, **752**, 265–270 (1996).
113. A.L. Crego, M.J. González, M.L. Marina, 'Chiral Separation of Polychlorinated Biphenyls by Micellar Electrokinetic Chromatography with Sodium Cholate', *Electrophoresis*, **19**, 2113–2118 (1998).
114. C.P. Ong, C.L. Ng, H.K. Lee, S.F.Y. Li, 'Separation of Phthalates by Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **542**, 473–481 (1991).
115. S. Takeda, S. Wakida, M. Yamane, A. Kawahara, K. Higashi, 'Migration Behavior of Phthalate Esters in Micellar Electrokinetic Chromatography with or without Added Methanol', *Anal. Chem.*, **65**, 2489–2492 (1993).
116. Q. Wu, H.A. Claessens, C.A. Cramers, 'The Separation of Herbicides by Micellar Electrokinetic Chromatography', *Chromatographia*, **34**, 25–30 (1992).
117. H. Süsse, H. Müller, 'Application of Micellar Electrokinetic Chromatography to the Analysis of Pesticides', *Fresenius' J. Anal. Chem.*, **352**, 470–473 (1995).
118. H. Süsse, H. Müller, 'Pesticide Analysis by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **730**, 337–343 (1996).
119. M. Jung, W.C. Brumley, 'Trace Analysis of Fluorescein-derivatized Phenoxy Acid Herbicides by Micellar Electrokinetic Chromatography with Laser-induced Fluorescence Detection', *J. Chromatogr. A*, **717**, 299–308 (1995).
120. A. Farran, S. Ruiz, 'Effect of Aliphatic Alcohols as Mobile Phase Modifiers on Separation of Phenylurea and Phenoxyalkyl Acid Herbicides by Micellar Electrokinetic Chromatography', *Anal. Chim. Acta*, **317**, 181–188 (1995).
121. C.P. Ong, C.L. Ng, H.K. Lee, S.F.K. Li, 'Separation of Imidazole and its Derivatives by Capillary Electrophoresis', *J. Chromatogr. A*, **686**, 319–324 (1994).
122. R. Furuta, T. Doi, 'Enantiomeric Separation of Diniconazole and Uniconazole by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **676**, 431–436 (1994).
123. R. Furuta, T. Doi, 'Chiral Separation of Diniconazole, Uniconazole and Structurally Related Compounds by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *Electrophoresis*, **15**, 1322–1325 (1994).
124. P. Schmitt, A.W. Garrison, D. Freitag, A. Kettrup, 'Application of Cyclodextrin-modified Micellar Electrokinetic Chromatography to the Separation of Selected Neutral Pesticides and their Enantiomers', *J. Chromatogr. A*, **792**, 419–429 (1997).
125. Y. Mechref, Z. El Rassi, 'Capillary Electrophoresis of Herbicides. II. Evaluation of Alkylglucoside Chiral Surfactants in the Enantiomeric Separation of Phenoxy Acid Herbicides', *J. Chromatogr. A*, **757**, 263–273 (1997).
126. J. Cai, Z. El Rassi, 'Micellar Electrokinetic Chromatography of Neutral Solutes with Micelles of Adjustable Surface Charge Density', *J. Chromatogr.*, **568**, 31–45 (1992).
127. Y. Mechref, Z. El Rassi, 'Capillary Electrophoresis of Herbicides: IV. Evaluation of Octylmaltopyranoside Chiral Surfactant in the Enantiomeric Separation of Fluorescently Labeled Phenoxy Acid Herbicides and their Laser-induced Fluorescence Detection', *Electrophoresis*, **18**, 220–226 (1997).
128. K. Otsuka, S. Terabe, T. Ando, 'Quantitation and Reproducibility in Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **396**, 350–354 (1987).
129. C. Lin, W. Lin, 'Migration Behavior of Dichlorophenols for Replicate Separations without Replenishment

of Buffer Electrolyte in Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **732**, 361–367 (1996).
130. C. Lin, W. Lin, W. Chiou, 'Migration Behavior and Selectivity of Dichlorophenols in Micellar Electrokinetic Chromatography. Influence of Micelle Concentration and Buffer pH', *J. Chromatogr. A*, **722**, 333–343 (1996).
131. C.P. Ong, C.L. Ng, N.C. Chong, H.K. Lee, S.F.Y. Li, 'Retention of Eleven Priority Phenols Using Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **516**, 263–270 (1990).
132. Y. He, H.K. Lee, 'Separation of Structural Homologs of Alkylphenols and Isomers of 4-nonylphenol by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **749**, 227–236 (1996).
133. L.N. Amankwa, W.W.G. Kuhr, 'Indirect Fluorescence Detection in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **63**, 1733–1737 (1991).
134. S. Hu, Z. Wang, P. Li, J. Cheng, 'Amperometric Detection in Capillary Electrophoresis with an Etched Joint', *Anal. Chem.*, **69**, 264–267 (1997).
135. G.M. McLaughlin, A. Weston, K.D. Hauffe, 'Capillary Electrophoresis Methods Development and Sensitivity Enhancement Strategies for the Separation of Industrial and Environmental Chemicals', *J. Chromatogr. A*, **744**, 123–134 (1996).
136. S.M. Burkinshaw, D. Hinks, D.M. Lewis, 'Capillary Zone Electrophoresis in the Analysis of Dyes and other Compounds Employed in the Dye-manufacturing and Dye-using Industries', *J. Chromatogr.*, **640**, 413–417 (1993).
137. K.P. Evans, G.L. Beaumont, 'Role of Capillary Electrophoresis in Special Chemical Research', *J. Chromatogr.*, **636**, 153–169 (1993).
138. W.C. Brumley, C.M. Brownrigg, A.H. Grange, 'Capillary Liquid Chromatography–Mass Spectrometry and Micellar Electrokinetic Chromatography as Complementary Techniques in Environmental Analysis', *J. Chromatogr. A*, **680**, 635–644 (1994).
139. S. Suzuki, M. Shiao, M. Aizawa, H. Nakazawa, K. Sasa, H. Sasagawa, 'Determination of Synthetic Food Dyes by Capillary Electrophoresis', *J. Chromatogr. A*, **680**, 541–547 (1994).
140. Y. Takada, M. Sakairi, H. Koizumi, 'On-line Combination of Micellar Electrokinetic Chromatography and Mass Spectrometry Using an Electrospray–Chemical Ionization Interface', *Rapid Commun. Mass Spectrom.*, **9**, 488–490 (1995).
141. S. Takeda, S. Wakida, M. Yamane, A. Kawahara, K. Higashi, 'Separation of Aniline Derivatives by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **653**, 109–114 (1993).
142. C. Quang, J.K. Strasters, M.G. Khaledi, 'Computer-assisted Modeling, Prediction, and Multifactor Optimization in Micellar Electrokinetic Chromatography of Ionizable Compounds', *Anal. Chem.*, **66**, 1646–1653 (1994).
143. Y.F. Yik, H.K. Lee, S.F.Y. Li, 'Separation of Nitroaromatics by Micellar Electrokinetic Chromatography', *J. High Resolut. Chromatogr.*, **15**, 198–200 (1992).
144. U. Pyell, U. Bütehorn, 'Optimization of Resolution in Micellar Electrokinetic Chromatography via Computer-aided Simultaneous Variation of Concentrations of Sodium Dodecyl Sulfate and Urea as Modifier', *J. Chromatogr. A*, **716**, 3–10 (1995).
145. D.M. Northop, D.E. Martire, W. Mac Crehan, 'Separation and Identification of Organic Gunshot and Explosive Constituents by Micellar Electrokinetic Chromatography', *Anal. Chem.*, **63**, 1038–1042 (1991).
146. E. Mussenbrock, W. Kleiboher, 'Separation for the Determination of Residues of Explosives in Soils Using Micellar Electrokinetic Capillary Chromatography', *J. Microcolumn Sep.*, **7**, 107–116 (1995).
147. M.J. Cugat, F. Borrull, M. Calull, 'Comparative Study of Capillary Zone Electrophoresis and Micellar Electrokinetic Chromatography for the Separation of Twelve Aromatic Sulfonate Compounds', *Chromatographia*, **46**, 204–208 (1997).
148. C.L. Ng, H.K. Lee, S.F.Y. Li, 'Determination of Organolead and Organoselenium Compounds by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **652**, 547–553 (1993).
149. M.M. See, S. Elshihabi, J.A. Burke, M.M. Bushey, 'Resolution Effects of Organic Additives on the Micellar Electrokinetic Chromatography Enantiomeric Separations of Tri Aza Aromatic Ligand Compounds of iron(II)', *J. Microcolumn Sep.*, **7**, 199–206 (1995).
150. R.D. Zare, G.J. Diebold, 'Laser Fluorimetry: Subpicogram Detection of Aflatoxins Using High-performace Liquid Chromatography', *Science*, **196**, 1439–1441 (1977).
151. K. Helrich, *Official Methods of Analysis*, 15th edition, AOAC, Arlington, VA, Chapter 49, 1990.
152. K. Ballschmitter, M. Zell, 'Analysis of Polychlorinated Biphenyls (PCBs) by Capillary Gas Chromatography', *Fresnius' J. Anal. Chem.*, **302**, 20–31 (1980).
153. N.L. Wolfe, L.A. Burns, W.C. Steen, 'Use of Linear Free Energy Relationships and an Evaluated Model to Assess the Fate and Transport of Phthalate Esters in the Aquatic Environment', *Chemosphere*, **9**, 393–402 (1980).
154. S. Terabe, H. Ozaki, K. Otsuka, T. Ando, 'Electrokinetic Chromatography with 2-*o*-carboxymethyl-β-cyclodextrin as a Moving Pseudo-stationary Phase', *J. Chromatogr.*, **332**, 211–217 (1985).
155. O.H.J. Szolar, R.S. Brown, J.H.T. Luong, 'Separation of Polycyclic Aromatic Hydrocarbons by Capillary Electrophoresis with Laser-induced Fluorescence Detection Using Mixtures of Neutral and Anionic β-cyclodextrins', *Anal. Chem.*, **67**, 3004–3010 (1995).
156. M.J. Sepaniak, C.L. Copper, K.W. Whitaker, V.C. Anigbogu, 'Evaluation of a Dual-Cyclodextrin Phase Variant of Micellar Electrokinetic Chromatography for

Separations of Nonionizable Solutes', *Anal. Chem.*, **67**, 2037–2041 (1995).
157. R.S. Brown, J.H.T. Luong, O.H.J. Szolar, A. Halasz, J. Hawari, 'Cyclodextrin-modified Capillary Zone Electrophoresis: Determination of Polycyclic Aromatic Hydrocarbons in Contaminated Soils', *Anal. Chem.*, **68**, 287–292 (1996).
158. A.L. Nguyen, J.H.T. Luong, 'Separation and Determination of Polycyclic Aromatic Hydrocarbons by Solid-phase Microextraction/Cyclodextrin-modified Capillary Electrophoresis', *Anal. Chem.*, **69**, 1726–1731 (1997).
159. S. Terabe, N. Matsubara, Y. Ishihama, Y. Okada, 'Microemulsion Electrokinetic Chromatography: Comparison with Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **608**, 23–29 (1992).
160. E.S. Ahuja, J.P. Foley, 'Separation of very Hydrophobic Compounds by Hydrophobic Interaction Electrokinetic Chromatography', *J. Chromatogr. A*, **680**, 73–83 (1994).
161. Y. Walbroehl, J.W. Jorgenson, 'Capillary Zone Electrophoresis of Neutral Organic Molecules by Solvophobic Association with Tetraalkylammonium Ion', *Anal. Chem.*, **58**, 479–481 (1986).
162. Y.C. Shi, J.S. Fritz, 'Capillary Zone Electrophoresis of Neutral Organic Molecules in Organic Aqueous Solution', *J. High Resolut. Chromatogr.*, **17**, 713–718 (1994).
163. Y. Shi, J.S. Fritz, 'High Performance Capillary Zone Electrophoresis of Nonionic Compounds Using a Novel Anionic Surfactant Additive', *Anal. Chem.*, **67**, 3023–3027 (1995).
164. A.L. Crego, A. González, M.L. Marina, 'Electrochromatography', *Crit. Rev. Anal. Chem.*, **26**, 261–304 (1996).
165. L.A. Colon, Y. Guo, A. Fermier, 'Capillary Electrochromatography', *Anal. Chem.*, **69**, 461A–467A (1997).
166. M.M. Robson, M.G. Cikalo, P. Myers, M.R. Euerby, K.D. Bartle, 'Capillary Electrochromatography: a review', *J. Microcolumn Sep.*, **9**, 409–419 (1997).
167. G.J. Bruin, P.P.H. Tock, J.C. Kraak, H. Poppe, 'Electrically Driven Open-tubular Column Liquid Chromatography', *J. Chromatogr.*, **517**, 557–572 (1990).
168. Y. Guo, L.A. Colón, 'Modification of the Inner Capillary Surface by the Sol–Gel Method: Application to Open-tubular Electrochromatography', *J. Microcolumn Sep.*, **7**, 485–491 (1995).
169. Y. Guo, L.A. Colón, 'A Stationary Phase for Open-tubular Column Liquid Chromatography and Electrochromatography Using Sol–Gel Technology', *Anal. Chem.*, **67**, 2511–2516 (1995).
170. Y. Guo, L.A. Colón, 'Open-tubular Liquid Chromatography Using a Sol–Gel Derived Stationary Phase', *Chromatographia*, **43**, 9–10 (1996).
171. J.H. Knox, I.H. Grant, 'Electrochromatography in Packed Tubes Using 1.5 to 50 μm Silica Gels and Octadecylsilica Bonded Silica Gels', *Chromatographia*, **32**, 317–328 (1991).
172. P.B. Wright, A.S. Lister, J.G. Dorsey, 'Behavior and Use of Nonaqueous Media without Supporting Electrolyte in Capillary Electrophoresis and Capillary Electrochromatography', *Anal. Chem.*, **69**, 3251–3259 (1997).
173. M.M. Dittmann, G.P. Rozing, 'Capillary Electrochromatography: Investigation of the Influence of Mobile Phase and Stationary Phase Properties on Electroosmotic Velocity, Retention and Selectivity', *J. Microcolumn Sep.*, **9**, 399–408 (1997).
174. H. Yamamoto, J. Baumann, F. Erni, 'Electrokinetic Reversed-phase Chromatography with Packed Capillary Columns', *J. Chromatogr.*, **593**, 313–319 (1992).
175. H. Rebscher, U. Pyell, 'A Method for the Experimental Determination of Contributions to Band Broadening to Electrochromatography with Packed Capillaries', *Chromatographia*, **38**, 737–743 (1994).
176. C. Yan, R. Dadoo, H. Zhao, R.N. Zare, 'Capillary Electrochromatography: Analysis of Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **67**, 2026–2029 (1995).
177. C. Yan, R. Dadoo, R.N. Zare, D.J. Rakestraw, D.S. Anex, 'Gradient Elution in Capillary Electrochromatography', *Anal. Chem.*, **68**, 2726–2730 (1996).
178. M. Albin, P.D. Grossman, S.P. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 486A–497A (1993).
179. A. Mainka, K. Bächmann, 'UV-detection of Derivatized Carbonyl Compounds in Rain Samples in Capillary Electrophoresis Using Sample Stacking and a Z-shaped Flow Cell', *J. Chromatogr. A*, **767**, 241–247 (1997).
180. M.I. Turnes, M.C. Mejuto, R. Cela, 'Determination of Pentachlorophenol in Water Samples by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **733**, 395–404 (1996).
181. R.O. Cole, R.D. Holland, M.J. Sepaniak, 'Factors Influencing Performance in the Rapid Separation of Aflatoxins by Micellar Electrokinetic Chromatography', *Talanta*, **39**, 1139–1147 (1992).
182. T. Imasaka, K. Nishitani, N. Ishibashi, 'Determination of Anthracene Derivatives by Cyclodextrin-modified Micellar Electrokinetic Chromatography Combined with He–Cd Laser Fluorimetry', *Anal. Chim. Acta*, **256**, 3–7 (1992).
183. S.M. Nie, R. Dadoo, R.N. Zare, 'Ultrasensitive Fluorescence Detection of Polycyclic Aromatic Hydrocarbons in Capillary Electrophoresis', *Anal. Chem.*, **65**, 3571–3575 (1993).
184. T. Imasaka, K. Nishitani, N. Ishibashi, 'Cyclodextrin-modified Micellar Electrokinetic Chromatography Combined with Semiconductor Laser Fluorimetry', *Analyst*, **116**, 1407–1409 (1991).
185. T. Kaneta, T. Imasaka, 'Indirect Detection of Aromatic Hydrocarbons by Semiconductor Laser Fluorimetry in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **67**, 829–834 (1995).
186. Y. Chao, C. Whang, 'Capillary Zone Electrophoresis of Eleven Priority Phenols with Indirect Fluorescence Detection', *J. Chromatogr. A*, **663**, 229–237 (1994).

187. I.C. Chen, C.W. Whang, 'Capillary Zone Electrophoresis of Eleven Priority Phenols with Amperometric Detection', *J. Chin. Chem. Soc.*, **41**, 419–424 (1994).
188. C.D. Gaitonde, P. Pathak, 'Capillary Zone Electrophoresis Separation of Chlorophenols in Industrial Waste Water with On-column Electrochemical Detection', *J. Chromatogr.*, **514**, 389–393 (1990).
189. D. Barceló, G. Durand, V. Boucot, M. Nielen, 'Use of Extraction Disks for Trace Enrichment of Various Pesticides from River Water and Simulated Seawater Samples Followed by Liquid Chromatography–Rapid-scanning UV–Visible and Thermospray-Mass Spectrometry Detection', *Environ. Sci. Technol.*, **27**, 271–277 (1993).
190. M.E. Swartz, M. Merion, 'On-line Sample Preconcentration on a Packed-inlet Capillary for Improving the Sensitivity of Capillary Electrophoretic Analysis of Pharmaceuticals', *J. Chromatogr.*, **632**, 209–213 (1993).
191. Waters AccuSep® C/PRP Concentrator Capillarie, Waters, Milford, MA, 1992.
192. A. Cavallaro, V. Piangerelli, F. Nerini, S. Cavalli, C. Reschiotto, 'Selective Determination of Aromatic Amines in Water Samples by Capillary Zone Electrophoresis and Solid-phase Extraction', *J. Chromatogr. A*, **709**, 361–366 (1995).
193. M.W.F. Nielen, 'Trace Enrichment of Environmental Samples in Capillary Zone Electrophoresis', *Trends Anal. Chem.*, **12**, 345–356 (1993).
194. D. Kaniansky, F. Iványi, F.I. Onuska, 'On-line Isotachophoretic Sample Pretreatment in Ultratrace Determination of Paraquat and Diquat in Water by Capillary Zone Electrophoresis', *Anal. Chem.*, **66**, 1817–1824 (1994).
195. G. Dinelli, A. Vicari, P. Catizone, 'Monitoring of Herbicide Pollution in Water by Capillary Electrophoresis', *J. Chromatogr. A*, **733**, 337–347 (1996).
196. M. Tomita, T. Okuyama, Y. Nigo, B. Uno, S. Kawai, 'Determination of Glyphosate and its Metabolite, (Aminomethyl)phosphonic Acid, in Serum Using Capillary Electrophoresis', *J. Chromatogr.*, **571**, 324–330 (1991).
197. G. Dinelli, A. Vicari, P. Catizone, 'Use of Capillary Electrophoresis for Detection of Metsulfuron and Chlorsulfuron in Tap Water', *J. Agric. Food Chem.*, **41**, 742–746 (1993).
198. M.C. Carneiro, L. Puignou, M.T. Galceran, 'Comparison of Capillary Electrophoresis and Reversed-phase Ion-pair High-performance Liquid Chromatography for the Determination of Paraquat, Diquat and Dibenzoquat', *J. Chromatogr. A*, **669**, 217–224 (1994).
199. Y.Z. Hsieh, H.Y. Huang, 'Analysis of Chlorophenoxy Acid Herbicides by Cyclodextrin-modified Capillary Zone Electrophoresis', *J. Chromatogr. A*, **745**, 217–223 (1996).
200. R. Carabias Martínez, E. Rodríguez Gonzalo, A.I. Muñoz Domínguez, J. Domínguez Alvarez, J. Hernández Méndez, 'Determination of Triazine Herbicides in Water by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **733**, 349–360 (1996).
201. G. Lai, J.P. Pachlatko, 'Identification of Cycloheximide in Microbial Extracts with Micellar Electrokinetic Chromatography', *J. High Resolut. Chromatogr.*, **17**, 565–566 (1994).
202. G. Dinelli, A. Vicari, V. Brandolini, 'Detection and Quantitation of Sulfonylurea Herbicides in Soil at the ppb Level by Capillary Electrophoresis', *J. Chromatogr. A*, **700**, 201–207 (1995).
203. W. Kleiböhmer, K. Cammann, J. Robert, E. Mussenbrock, 'Determination of Explosives Residues in Soils by Micellar Electrokinetic Chromatography and High-performance Liquid Chromatography', *J. Chromatogr.*, **638**, 349–356 (1993).
204. K. Li, S.F.Y. Li, H.K. Lee, 'Determination of Organolead and Organotin Compounds in Water Samples by Micellar Electrokinetic Chromatography', *J. Liq. Chromatogr.*, **18**, 1325–1347 (1995).
205. S.J. Kok, E.H.M. Koster, C. Gooijer, N.H. Velthorst, U.A. Th. Brinkman, 'Separation of Twenty-one Naphthalene Sulfonates by Means of Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **19**, 99–104 (1996).
206. D.S. Burgi, R.P. Chien, 'Optimization in Sample Stacking for High-performance Capillary Electrophoresis', *Anal. Chem.*, **63**, 2042–2047 (1991).
207. D.S. Burgi, R.P. Chien, 'On-column Sample Concentration Using Field Amplification in Capillary Zone Electrophoresis', *Anal. Chem.*, **64**, 489A–496A (1992).
208. F.M. Everaerts, J.L. Beckers, P.E.M. Verheggen, *Isotachophoresis–Theory, Instrumentation and Applications*, Elsevier, Amsterdam, 1976.
209. D. Kaniansky, J. Marak, 'On-line Coupling of Capillary Isotachophoresis with Capillary Zone Electrophoresis', *J. Chromatogr.*, **498**, 191–204 (1990).
210. M. Mazereeuw, U.R. Tjaden, N.J. Reinhoud, 'Simple Capillary Isotachophoresis–Zone Electrophoresis – Current Practice and Prospects', *J. Chromatogr. Sci.*, **33**, 686–697 (1995).
211. M.G. Bissell, A.O. Okorodudu, J.R. Petersen, A.A. Mohammad, 'Laser Induced Energy Transfer: a Novel Approach Towards Achieving High Sensitivity in Capillary Electrophoresis. Part B: Applications for Environmental Testing', *J. Chromatogr.*, **767**, 217–221 (1997).
212. B. Huang, J. Li, L. Zhang, J. Cheng, 'On-line Chemiluminescence Detection for Capillary Ion Analysis', *Anal. Chem.*, **68**, 2366–2369 (1996).
213. Q. Lu, S.M. Bird, R.M. Barnes, 'Interface for Capillary Electrophoresis and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**, 2949–2956 (1995).

214. Y. Liu, V. López-Ávila, J.J. Zhu, D.R. Wiederin, W.F. Beckert, 'Capillary Electrophoresis Coupled On-line with Inductively Coupled Plasma Mass Spectrometry for Elemental Speciation', *Anal. Chem.*, **67**, 2020–2025 (1995).

215. J.W. Olesik, J.A. Kinzer, S.V. Olesik, 'Capillary Electrophoresis Inductively Coupled Plasma Spectrometry for Rapid Elemental Speciation', *Anal. Chem.*, **67**, 1–12 (1995).

Organic Analysis in Environmental Samples by Electrochemical Methods

Ignacio Naranjo-Rodríguez and José Luis Hidalgo-Hidalgo de Cisneros
University of Cádiz, Spain

1 Introduction 1
2 Sample Preparation 3
3 Environmental Applications 3
4 Environmental Regulations and Methods 6
5 Method Development 7
5.1 Nitrated Organic Explosives Components 7
5.2 Aromatic Amines 9
5.3 Other Nitrogen-containing Compounds 9
5.4 Aromatic Hydrocarbons 10
5.5 Polychlorinated Biphenyls 10
5.6 Phenols 13
5.7 Pesticides 17
5.8 Surfactants 20
5.9 Organometallics 21
5.10 Miscellaneous Compounds 22
6 Comparison with Other Techniques 22
Abbreviations and Acronyms 24
Related Articles 24
References 25

Concern for the environment is not limited to ecological aspects, but encompasses matter such as the origin and the effects of pollutant discharges, the evaluation of environmental programs, quality control and quality assurance related to analytical precedures, and the analysis of the content of environmental matrices – both inorganic and organic – in order to control and evaluate environmental pollution. Thus, it is right to say that analytical chemistry has an important role to play in environmental analysis.

A large number of analytical methods are used in environmental analysis. Some are widely used in environmental programs, and are reference/standard/official methods for the analysis of a great variety of compounds.

However, there is no single universal procedure that can be applied to the analysis of any chemical species in any environmental matrix, and for any in situ/remote measurement.

Electrochemical (EC) methods are widely used to investigate environmentally significant compounds, both organic and inorganic. EC research covers all important branches of environmental analysis, such as in situ/remote measurements, flow systems, combined applications with a separation technique, (bio)sensors for selective analysis, and monitoring and control procedures.

This article discusses the application of these EC methods to the study of organic compounds occurring in environmental water and waste samples.

1 INTRODUCTION

Modern EC methods are applicable to a very large number of organic compounds of great importance, not only in environmental monitoring but in many other fields.[1]

Most applications of chemical analysis to environmental protection involve trace determinations; consequently, highly sensitive methods that offer good selectivity, precision, and accuracy are needed. Other desirable characteristics are easy sample pretreatment, quick analysis, easy automatization of the method, the availability of a portable device for field measurements, low cost of instrumentation and analysis.[2]

Assignments given to environmental analysis include: the quantitative and qualitative control of pollutant emissions in the environment; the elucidation of the path of the emitted pollutant from its source to its interaction with humans; the monitoring/study of the conversion and transformation of pollutants as well as their natural reactions; and the discovery and elucidation of processes for the prevention of the emissions or processes for their neutralization. Unfortunately, it is impossible to monitor the content, distribution and fate of all chemical substances present in the environment, whether man-made or naturally occurring.

Polarography was one of the first techniques for monitoring some anthropogenic pollutants at the trace level in real environmental samples, in particular for heavy metals. Nowadays, EC methods have to compete with other methods – the whole analytical procedure, from problem definition to data evaluation, must be taken into account when choosing the right determination method. Economic aspects are also very important, and include the cost of instrumentation, the qualification and number of operators, sample requirements (such as the sampling method, necessity and manner of storage), and the possibilities and difficulties of the process automatization;[3] of increasing importance for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

environmental monitoring is quality assurance and quality control.

A large number of important organic environmental compounds are electrochemically active, either directly or indirectly. In most cases, the sensitivity is adequate for organic environmental analysis; the use of stripping techniques, especially adsorptive stripping analysis for organic compounds, increases this sensitivity. Selectivity is limited by several factors, such as the number of compounds that are active in a same range, the overlapping of signals, and surface alterations of the electrodes. Using chemical reactions it is sometimes possible to achieve experimental situations to solve most problems without using special separation techniques. However, in many organic analyses the variety of compounds is much larger, and similar compounds are only badly or not distinguished at all by their EC properties. For these more complex problems, the use of EC sensors after high-performance liquid chromatography (HPLC) or other separation technique offers an attractive solution. Although EC methods have a moderate equipment cost, the whole analytical process must be priced for a particular sample. Experienced staff are also necessary for the accurate analysis of complex environmental samples when using modern methods and instrumentation for EC trace analysis.

In voltammetry, a continuous changing potential is applied and the current that flows is continuously recorded; the term 'polarography' is reserved for voltammetric methods in which an electrode with a renewed surface is used. The most-used techniques for analytical purposes are cyclic, differential pulse, and square wave voltammetry. Potentiometry involves the measurement of the potential of a working electrode (WE) as a constant current is passed through an EC cell; sensors based on MOSFET (metal oxide semiconductor field effect transistor) semiconductors is a developing field in potentiometric analysis. Coulometry is based on the deposition (or conversion to another oxidation state) of an analyte on an electrode at a constant potential.

One of the main applications of electroanalysis is in trace determination of organic pollutants of interest; the needs of environmental analysis include the use of flow systems, mainly on-line flow systems coupled to separation methods with electrochemical detection (ED) for the analysis of complex environmental samples.

Chromatography is the most important and popular separation technique, but a sensitive detector is needed for the separated compounds. For organic analysis, the most common EC measuring technique in chromatography/flowing liquids is amperometry, i.e. the measurement at a constant electrode potential. Amperometric detectors were introduced by Kissinger et al. in the early 1970s for oxidative ED of catecholamines at carbon electrodes. Phenomena at an interface between a solution and an electrode can be exploited for potentiometric, coulometric, and voltammetric detectors. Potentiometric detectors are based on the measurement of the electromotive force under zero current conditions. With voltammetric detectors, the current is measured as a function of an applied potential. Amperometric, pulsed amperometric, and scanning voltammetric detectors are all classified as voltammetric detectors. In coulometric detectors an electrode reaction efficiency of practically 100% is obtained.[4]

Separation techniques can also be used for preconcentration and enhancement of sensitivity, and thus lower detection limits can be obtained. Although stripping techniques involve a preconcentration step, the separation is usually a secondary effect as the aim of stripping is to improve the sensitivity of the determination.[5]

The most important EC methods applied to environmental analysis are based on the measurement of signals due to phenomena at an interface between an electrode and a solution. However, there are some problems associated to this superficial interaction. The sensitivity of electrode interfaces in stripping methods, with respect to the adsorption of interferences or surface alterations, is the main source of problems for their application to real environmental samples. This is particularly true in the case of incomplete destruction of organic matrix components. Selectivity is limited in voltammetry because of overlapping voltammetric signals. One of the most critical aspects of the routine analysis of complex environmental samples is the stability and reproducibility of the WE surface. The use of renewable mercury electrodes avoids this problem, but its toxicity is an increasing matter of concern; solid electrodes with stable and reproducible surfaces are a very interesting field where very important progress is being made.

The application of EC (bio)sensors is a possibility, due to several characteristics, such as intrinsic specifity, fast analysis, and minimal requeriments for sample treatment. However, widespread application of EC (bio)sensors for environmental purposes is remains limited by the low stability of (bio)sensors and their high detection limits.[6] EC (bio)sensors can be useful for field monitoring, to provide enough information for the routine testing and screening of samples. In this way they can be used to establish an unmanned measuring system for the continuous monitoring of the environment at selected sites, or used as a component of a portable instrument for incorporation in flow systems.[7]

In situ measurement offers several advantages. The operations connected with the collection, transport and storage of samples are eliminated, and so the cost of the global process is lower, and the delay and

error incurred in such operations can be avoided. The measurements are carried out in real time and the pollutant is analyzed in its own environment. The results are obtained quickly, and consequently the problem can be solved promptly. Remote EC sensors for monitoring organic pollutants are a very interesting field, and advances in EC technology now enable the use of modified electrodes, ultramicroelectrodes, the design of highly selective recognition layers, sensor arrays, and development in the areas of microfabrication, flow detectors, and compact, low-powered, user-friendly instrumentation.[8]

2 SAMPLE PREPARATION

The use of analytical procedures applied to environmental analysis involves a succession of steps, including: sampling and storage of environmental samples; the collection of treatments that must be carried out on the sample for quantification; and final determination and presentation of results. Several operations can be considered of general application to most procedures. After sampling, samples are filtered, usually through a 0.45 μm filter, and stored in the dark at 2–4 °C. After treatment, and when a chromatographic system is used, evaporation of the solvent prior to chromatographic analysis is carried out, in order to reduce the sample to an appropiate volume for injection. Most analytical papers dealing with environmental applications include an exact procedure for the sample treatment.

In environmental analysis, the analyzed species are seldom accessible directly to the detector. Treatment procedures include digestion, extraction and separation steps, which can add considerable uncertainly to the overall results. Matrix influences also have to be evaluated. The inestability of the analyte in the sample must be considered. Due to the high ionic content of some samples it is advisable to determine several species directly, as in seawater. Sometimes it is necessary to form a derivative which can be determined electrochemically.

Usually, UV (ultraviolet) digestion by sulphuric acid/hydrogen peroxide or by microwave digestion is used. In the analysis of organic pollutants, when water samples are too complex and/or the analytes very diluted, preliminary preparation prior to the analysis is needed. Traces of organic pollutants must be extracted from aqueous media and concentrated, and other components which can also be co-extracted and co-concentrated, and which may interfere in the analysis, must be removed from the matrix. The most usual methods for extraction and concentration of organic pollutants are liquid–liquid extraction (LLE) and solid-phase extraction (SPE).

The LLE technique has been used for many years due to its simplicity and good results. If a nonpolar solvent is chosen as the extractant, the addition of a supporting electrolyte is necessary to ensure a solution of sufficient conductivity before EC determination. The popularity of SPE has increased because of the reduced use of organic solvents. Micro-LLE is always an attractive solution. In the field of SPE, new sorbents are being developed in order to improve the selectivity in this preliminary step, to enable subsequent steps of removing interferents to be eliminated or simplified.[9]

Even though separation techniques are somehow always undesirable, because they will always increase the amount of labor and cost of the analysis, in many cases their application is necessary to reduce interference and/or to increase the sensitivity of the method by preconcentrating the analyte.[5] Environmental analysis includes the use of flow systems coupled to separation methods, such as chromatography, with ED for the analysis of complex environmental samples. As chromatography is only a way to separate compounds from each other, is necessary to use a detector which is sensitive to the separated compound. For organic analysis, the most common EC measuring technique in chromatography is amperometry, though some others detectors, such as coulometric, potentiometric, pulsed amperometric and scanning voltammetric detectors are also used.

The combination of capillary electrophoresis and electrochemistry is a developing field within EC research, and has been used for the analysis of catecholamines, carbohydrates, amino acids, chlorophenols and explosive compounds.[10] ED appears to be promising for analysis of compounds with well-defined redox behavior. Capillary electrophoresis has a high resolution, is suitable for ionic/polar and apolar compounds, but has a low loading capacity, in the range of few nanoliters; this last characteristic is not a disadvantage for very sensitive techniques. Table 1 gives several examples of the preparation/treatment/extraction procedures used with environmental samples.

3 ENVIRONMENTAL APPLICATIONS

The number of papers related to EC methods applied to environmental analysis is increasing. Different matrices, compounds and specific strategies of analysis have been reviewed to provide a general view of the applications of these methods. Analytical techniques for the determination of chemical species in natural water, drinking water and wastewater have been sumarized, including EC methods.[39]

In general, the sensitivity required in the determination is not a restriction when EC methods are

Table 1 Examples of preparation/treatment/extraction procedures used with environmental samples prior to EC analysis

Analyte	Sample	Previous treatment	Extraction/treatment for measurement	Ref.
Organic explosives (Organic nitro compounds)	groundwater	stored in the dark at 2–4 °C; exposition to daylight for differents periods of time for monitoring degradation processes	direct injection into HPLC system	11
	well/surface water	(a) stored in the dark at 4 °C; (b) extracted inmediately	(a) LLE with methylene chloride; (b) SPE	12
	groundwater	addition of sodium azide; stored in the dark at 4 °C	LLE with dichloromethane pH 9, retention of the neutral/basic fraction; LLE with dichloromethane pH 2, retention of the acidic fraction	13
Aromatic amines	drinking water	without filtration	on-line preconcentration/ extraction system by two precolumns	14
	groundwater	refrigerated at 5 °C; filtered through a 0.22 μm filter disk	LC with precolumn sample preconcentration	15
Aromatic hydrocarbons	tap water, motor oil		tetrahydrofuran was used to dilute water samples or to extract oil samples Ce(IV) is added to oxidize aromatic hydrocarbons to quinones	16
	ground, surface, river, rain, drinking water	addition of 2-propanol; filtered through a 0.45 μm filter	SPE; evaporation of the solvent and the residue diluted with methanol–water for HPLC analysis	17
PCBs (TCB)	wastewater	filtration and dissolution of the insoluble material in methanol		18
(Aroclors®)	industrial water effluent	drying agent	LLE with a water-immiscible solvent, mechanical shaking for 20 min	19
Phenols	seawater	filtered, adjusted to pH 2 and stored at 4 °C	SPE and elution with acetonitrile	20
	seawater	filtered through a 0.45 μm filter, adjusted to pH 2 and frozen at −20 °C	SPE	21
	river, tap water	filtered through a 0.45 μm filter, adjusted to pH 2 and stored at 4 °C	solid-phase disk extraction	22
	river water		on-line SPE	23
(Total phenols)	industrial wastewater	stored at 4 °C	SPE	24
(Catecol, phenol)	wastewater	filtered through a membrane filter, adjusted to pH 5	SPE	25
(Phenol, *o*-cresol)	tap water, wastewater		comparision between drect injection, LLE and on/off line SPE	26
Pesticides (Amitrole)	drinking and groundwater	water samples (5 mL) concentrated to 0.5 mL under a nitrogen stream; filtered through a 0.5 μm membrane filter		27
(Phenylurea herbicides)	tap, mineral and spring water	stored in the dark at 2–4 °C	SPE by Sep-Pak cartridges (C_{18}); elution with methanol	28

Table 1 (*continued*)

Analyte	Sample	Previous treatment	Extraction/Treatment for measurement	Ref.
(Phenylurea and chlorotriazine herbicides)	river, estuarine and coastal seawater	(a) on-line analysis: sample filtered through a glass microfiber (b) off-line analysis: sample filtered through a glass microfiber, porosity 0.7 μm	(a) precolumn packed with styren–divinylbenzene copolymer; elution by an acetonitrile–water gradient via a LC pump (b) SPE (C_{18}); elution with acetonitrile; evaporation to dryness and dissolved in acetonitrile–water (20 : 80)	29
Surfactants (Poly(ethylene glycols))	river water		LLE with ethyl acetate and chloroform	30
(Nonionic surfactants)	river water		LLE with ethyl acetate	31
	surface water and wastewater		reaction with molybdophosphoric acid for complexation	32
(Linear alkylbenzene sulfonates)	tap water and sewage	complete evaporation to dryness of the sample; dissolution with methanol and evaporation of the solvent	nitration with fuming nitric acid for 15 min at 30 °C; evaporation of nitric acid and dissolution of the residue with Britton–Robinson buffer solution of pH 12	33
Organometallics (Tributyltin)	seawater	preconcentration by a cation exchanger; elution with HCl–methanol	LLE with dichloromethane	34
(Methyl-, ethyl-, phenylmercury)	river water	addition of sodium pyrrolidenedithiocarbamate for complexation reaction with the mercury compounds	SPE and elution with methanol; evaporation of the solvent to dryness; reconstitution of the residue prior to injection into the chromatographic system	35
(Alkyllead compounds)	seawater	stored in the dark at 4 °C; addition of sodium diethylthiocarbamate	double extraction with 20 + 10 mL hexane; re-extraction for carbamate complexes to an acid aqueous solution	36
Miscellaneous (9,10-Anthraquinone)	natural water, industrial samples		Continuous steam distillation/extraction device	37
(Phthalate esters)	wastewater		LLE with hexane (two or three extractions are required); alkaline hydrolysis of the esters	38

LC, liquid chromatography; PCBs, polychlorinated biphenyls; TCB, trichlorobiphenyl.

applied to the analysis of organic pollutants. Detection limits obtained with these methods compare well with those obtained by other methods; several papers describe comparative studies between ED and other detectors.

Problems associated with possible interferences are, sometimes, particular when EC methods are used. It is known that selectivity of some EC applications is poor. Many organic pollutants have similar molecular structures. Therefore, the use of an unspecific electrode without a previous separation step results in the overlapping of the redox responses of similar compounds when their mixtures are analyzed. It is difficult to find a specific reaction or specific sustrate electrode for each organic compound.

Most EC processes used in environmental analysis take place on an electrode surface, so reproducibility and stability of the active surfaces is an important consideration. Unfortunately, these characteristics are not easy to achieve. As a result of the problems associated with the use of mercury electrodes, most research deals with new and more sensitive combinations of EC detectors and separation techniques, specific (bio)sensors for organic pollutants, and the stabilization of electrode surfaces.

Trace organic contaminants in surface water have been considered.[40] The methods described in literature or used in laboratories often derive from those used for drinking water, with additional treatments. There is a general trend to develop LC methods, due to their advantages over gas chromatography (GC) for a large

number of environmental pollutants. EC methods for the determination of anilines and chloroanilines are described, with detection limits below $0.1\,\mu g\,L^{-1}$.

As most applications use a separation step combined with ED, there are reviews that discuss the advantages and limitations of combining separation technique and voltammetry,[5] as well as reviews that describe the EC detectors used for LC and capillary electrophoresis.[4] Different types of detectors are considered, and much effort is directed at overcoming the difficulties of using combined techniques to achieve selective determinations.

Biosensors offer many possibilities for the analysis of organic pollutants. Despite the problems associated with biosensors, their attributes, such as selectivity, in situ/remote measurements, and minimal requirements for sample analysis are very important, and research in this field is increasing. Although amperometric and potentiometric transducers are used,[41] an electrical transducer and a potentiostat can also be used for detection. Most reports refer to amperometric biosensors, because they offer higher sensitivity and selectivity. Although potentiometric devices are not used as much, they still make an important contribution. For example, an enzyme (cholinesterase) potentiometric biosensor can be used for preliminary testing of the total pollution of domestic and industrial discharges.[42]

The most used biosensors are tyrosinase based and peroxidase based. Environmental organic pollutants are either enzyme substrates or enzyme inhibitor, for both types of biosensors.[43] Whereas thioureas, mercaptoethanol, hydroxylamine, methylisothiocyanate, and cyanide are detected with a peroxidase biosensor, carbamates are detected with a tyrosinase biosensor. The use of cholinesterase-based biosensors has also been described.[44] These biosensors are applied to the analysis of organophosphate and carbamate pesticides; acettylcholinesterase and butyrylcholinesterase are the enzymes used for the analysis of pesticides in real environmental samples. Potentiometric or amperometric transducers can be used for the analysis of these pesticides. Most potentiometric enzimatic biosensors are based on the immobilization of an appropriate esterase and measurement of the pH change which is produced; a glass electrode or a pH-sensitive ion-sensitive field effect transistor (ISFET) is usually used as the internal detector for the pH change.[45] Different amperometric transducers are employed. They use the reduction of oxygen, or the oxidation of hydrogen peroxide or thiocholine. In general, amperometric detection reaches lower detection limits than potentiometric detection.[46]

EC immunoassays are based on modifications of enzyme immunoassays (EIAs) and the enzyme activities are determined potentiometrically or amperometrically. Some EC immunosensors have been described, including one for the analysis of PCBs using a conducting polymer-based immunosensor.[47]

Among specific strategies, a review on developments and applications of screen-printed EC sensors and their applications to environmental analysis can be found.[48] Several analytes have been studied in environmental samples, and the research focuses on obtaining portable devices and the ability of performing environmental analysis away from a laboratory. Amperometric biosensors have been applied to hydrazines, hydroquinona, phenols and pesticides. Two interesting alternatives are the use of remote sensors and submersible probes for organic analyltes, by appropriate modification of the surface of the electrode and deposition of chemical recognition layers.[8]

Phenolic compounds belong to a very important group of organic pollutants. The EC literature on these compounds is wide, and the methods developed for their determination include EC detectors coupled to separation techniques, amperometric (bio)sensors, and remote measurements. Methodologies for the determination of phenolic compounds in some water types have been reviewed including ED, amperometric biosensors, and emerging techniques.[49] Applications of EC methods to the determination of other groups of organic pollutants in environment have been also described, such as organoarsenic and organotin compounds,[50,51] surfactants,[52,53] ethylenediaminetetraacetic acid (EDTA)/diethylenetriaminepentaacetic acid (DTPA).[54]

4 ENVIRONMENTAL REGULATIONS AND METHODS

In spite of the proven applicability of EC methods for organic analysis, there are few examples of EC methods being used as reference and/or official methods for the organic analysis of environmental samples.

Most current official analytical methods are based on LLE followed by GC using different detectors, such as electron capture and mass spectrometry (MS). However, there is a general trend to change these procedures to SPE followed by HPLC.[49]

The methods for organic chemical analysis of municipal and industrial wastewater, used by the United States Environmental Protection Agency (USEPA), include HPLC with an ED procedure for the determination of benzidines at a glassy carbon electrode, or equivalent. An LLE procedure is used in the sequence chloroform–acid–chloroform; the final chloroform extract is exchanged to methanol while it is being concentrated using a rotary evaporator for introduction into the HPLC system. Detection limits of $0.08–0.13\,\mu g\,L^{-1}$ are reported.[55]

Polarography has been proposed as an official German standard method for nitrilotriacetic acid determination in water, wastewater, and sludge.[1] The German ESB (environmental specimen banking) programme includes the regular determination of a wide variety of organic and inorganic species; electroanalytical methods and other analytical techniques are used to carry out these determinations.[3]

European Community Directive 76/464/ECC lists organic compounds to be eliminated from discharges. The drinking water standard methods are described in Directive 79/869/ECC. These methods are chemical group-specific analytical parameters, and nowdays are insufficient for the analysis of pollutants in water samples of different nature and origin. For example, nonionic surfactants are determined from potentiometric measurement after precipitation with bismuth. Many of the newly developed methods have been derived from those used for drinking water. So, a selective preconcentration technique has been described for the analysis of anilines and chloroanilines by cation exchangers coupled to a oxidative ED; these highly toxic substances are included in Directive 76/464/ECC, and their determination is possible thanks to the selectivity of the procedure. River water samples are analized, and detection limits below $0.1\,\mu g\,L^{-1}$ are reported.[40]

5 METHOD DEVELOPMENT

5.1 Nitrated Organic Explosives Components

Nitrated organic compounds have been widely used as munitions components; they have been produced in large quantities, and have to be regulated by environmental agencies.

There are several reasons for the great interest in the determination of organic explosives. First, their inherent toxicity; moreover, some of the biotransformation products, mainly aminoaromatic compounds, are known or suspected cancerogens or mutagens. Second, the localization of these explosives; surface and groundwater are generally contaminated by explosives in the vicinity of former ammunition plants, and decommissioning of military bases for other uses must be accompanied by verification that all pollution has been removed. Table 2 shows the application of EC methods to the analysis of nitrated organic explosives in environmental samples.

An HPLC method with UV and ED is described for monitoring the degradation of organic explosives. Amino and phenolic compounds are determined by oxidative ED (+1.1 V) at a glassy carbon electrode, whereas UV detection is used for nitro compounds. The samples (groundwater taken near a former ammunition plant) is injected directly in the HPLC system. To achieve the degradation of the explosives, samples are exposed to daylight for different periods of time.[11] Figure 1 shows the effect of light exposure on water samples for differents periods of time.

The use of hanging mercury drop electrode (HMDE) and the mercury film electrode (MFE) have been described for the reductive ED of organic explosives.[56,12] Well and surface water samples are collected, and two different extraction techniques are used, LLE and SPE. The extract is injected in an HPLC system with reductive ED (−1.0 V) at a MFE.

A glassy carbon WE in an HPLC system with ED allows the application of both reductive and oxidative detection.

Table 2 Environmental analysis of nitrated organic explosives by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
Organic explosives (Organic nitro compounds)	groundwater	–	monitoring degradation processes, HPLC with amperometric detection of amino derivatives, glassy carbon WE at +1.0 V	11
	well/surface water	–	HPLC with reductive amperometric detection at −1.0 V, gold–mercury thin-film WE	12
	groundwater	[a] 4–25 $\mu g\,L^{-1}$	HPLC with reductive/oxidative detection, 0.6–1.2 V and −(0.6–1.2) V, glassy carbon WE	13
	groundwater		electrokinetic capillary electrophoresis with amperometric detection at −0.7 V	
		(a) 70–110 $\mu g\,L^{-1}$	(a) gold WE	
		(b) 140–300 $\mu g\,L^{-1}$	(b) silver-on-gold WE	10
(Trinitrotoluene)	river/drinking water		EC sensor for remote continuous monitoring of 2,4-trinitrotoluene	
	groundwater		Square-wave voltammetry at a carbon fiber electrode	57

[a] Value for oxidative mode; for reductive mode, 5–10 times higher detection limits are obtained.

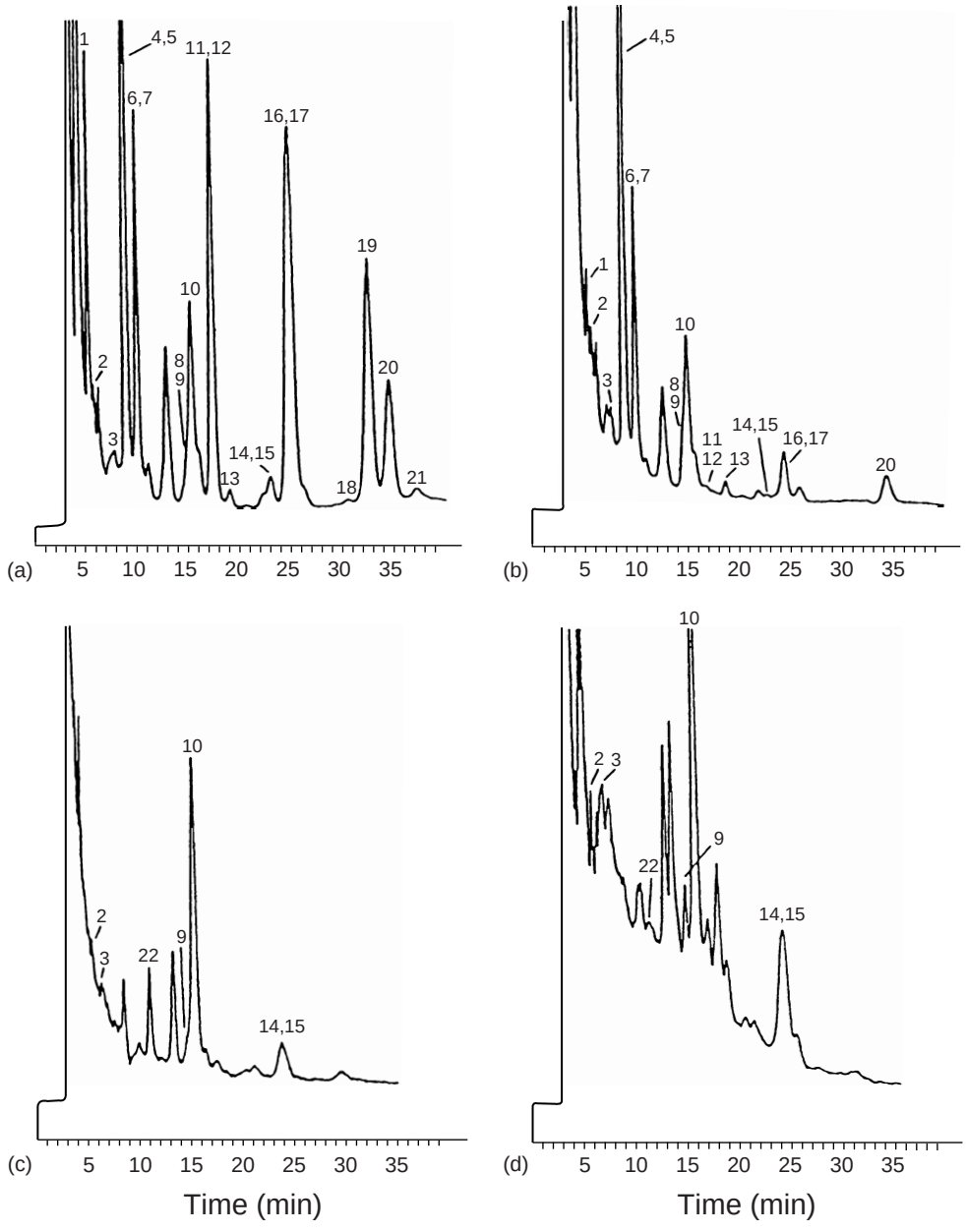

Figure 1 Representative chromatograms of directly injected water samples: (a) UV detection, sample not exposed; (b) UV detection, sample exposed for 80 days; (c) ED, sample not exposed; (d) ED, sample exposed for 80 days. (Reproduced by permission of Springer-Verlag from ref. 11.)

Nitroaromatics, nitramines, aminoaromatics and nitrophenols were determined in groundwater samples from the surroundings of a former ammunition plant, and the results compared with those obtained by UV detection. A two-step LLE procedure is used: dichloromethane at pH 9 to obtain the neutral/basic fraction, and at pH 2 for the acidic fraction.[13]

Electrokinetic capillary electrophoresis with amperometric detection has been used for the determination of organic explosives.[10] ED is carried out by a silver-on-gold electrode at −0.7 V. A groundwater sample obtained in the vicinity of the remediation site is used as sample. Detection limits in the range 70–110 $\mu g\,L^{-1}$ were found for all explosives, lower in all cases than with UV detection. Excellent agreement was found between the results obtained by capillary electrophoresis with amperometric detection and the standard USEPA method.

A remote EC sensor reported by Wang et al.[57] allows the continous monitoring of the explosive 2,4,6-trinitrotoluene. Square wave voltammetry as the EC

technique, and a submersible carbon-fiber as the WE are used. The electrode is connected to a 50 ft shielded cable. The response is very selective due to the low detection potential for the explosive. Low mg L^{-1} detection limits are obtained for untreated river water, groundwater and drinking water spiked with 2,4,6-trinitrotoluene.[57]

5.2 Aromatic Amines

Amines are present in the environment at the trace level. They enter the environment from a variety of sources, such as industrial discharges, from pesticide degradation, or from bacterial conversion of azo compounds into the precursor amines.

Aromatic amines are included in the EPA list of priority pollutants. The EEC directives give a concentration limit of 1 μg L^{-1} for all aromatic amines in drinking water, and 0.5 μg L^{-1} for the concentration of individual compounds. Table 3 shows the application of EC methods to the analysis of aromatic amines in environmental samples.

LC with ED has been described for the analysis of aromatic amines as these compounds are easily oxidized. Different aqueous environmental samples have been analyzed using LC or HPLC,[15,58] and a potential in the range 0.4–1.0 V is usually required together with previous sample preconcentration.

Cyclic voltammetry and hydrodynamic voltammetry can be used to study the voltammetric characteristics of these compounds. Aromatic diamines have a significantly lower oxidation potential than the monoamines, so the applied potential can be used for the individual analysis of aromatic mono- or di-amines.[59]

The UV detection limits at 235 nm and for ED at 0.9 V are always below the regulatory limits. Although UV detection can be used alone, ED provides complementary information for identification of the analytes.[14]

The use of DNA recognition layers has enabled the development of a DNA-modified electrode for detection of toxic aromatic amines.[60] It can be used as a screening tool for these compounds.

5.3 Other Nitrogen-containing Compounds

An amperometric sensor for flow injection analysis (FIA) of hydrazines has been developed; the electrode is a Nafion–ruthenium(III) modified glassy carbon electrode for the study of electroxidation of hydrazine compounds.[61] EC oxidation of hydrazine catalyzed by 4-hydroxy-2,2,6,6-tetramethyl-piperminyloxy at a glassy carbon electrode has also been applied to waste and river water.[62]

For hydrazines, a DNA-coated carbon paste electrode has been used as a sensitive biosensor for the detection of hydrazine compounds, by means of potentiometric stripping analysis. DNA immobilization is carried out at 1.7 V; for measurement, a stripping current of 8 μA is used.[63] Sensors for field monitoring studies are always an important trend in environmental analysis. A remote EC sensor for field monitoring has been described for hydrazine, based on a catalytic film of 3,4-dihydroxybenzaldehyde. The electrode is coupled to a shielded cable for remote operation.[64]

Capillary electrophoresis with ED has been used for the analysis of hydroxylamine. An ultramicro palladium particle modified carbon fiber electrode was employed and applied to the analysis of river and wastewater.[65]

Polarography of nitrilotriacetic acid complexed with Bi(III) is widely applied to verify the pollution of waters.

Table 3 Environmental analysis of aromatic amines and hydrazines by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
Aniline derivatives	drinking water	10–20 μg L^{-1}	LC with amperometric detection	14
Aromatic amines	river, groundwater	0.1–60 × 10^{-7} M	potentiometric stripping analysis at a DNA-modified carbon paste electrode	60
	groundwater	25 ng L^{-1}	LC with amperometric detection at a carbon paste or glassy carbon electrode	15
	tap water		HPLC with amperometric detection	59
	groundwater	<0.1 μg L^{-1}	HPLC with ED	58
Hydrazynes	river, groundwater	<2 mg L^{-1}	potentiometric stripping analysis at a DNA-modified carbon paste electrode	63
	river, lake, groundwater		electrocatalytically modified electrode based on a film of 3,4-dihydroxybenzaldehyde for remote monitoring	64
	river, wastewater	1 × 10^{-6} M	voltammetry determination based on catalytic reaction at a glassy carbon electrode	62
	drinking, river water	1 × 10^{-7} M	amperometric sensor based on a Nafion–ruthenium(III) modified glassy carbon electrode	61

Table 4 Environmental analysis of hydroxylamines, nitrilotriacetic acid and aromatic hydrocarbons by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
Hydroxylamine	river, wastewater	5×10^{-8} M	capillary electrophoresis with ED at a paladium particle modified carbon fiber microdisk array electrode	65
Nitrilotriacetic acid	wastewater	$2\,\mu g\,L^{-1}$	DPP complexed to bismuth(III)	66
	natural, wastewater		single-sweep polarography of the Bi(III)–NTA[a] complex	67
Aromatic hydrocarbons	tap water, motor oil	$5–50\,\mu g\,L^{-1}$	LC with reductive ED at a glassy carbon electrode	16
	ground, surface, river, rain, drinking water	$0.26–0.71\,\mu g\,L^{-1}$	HPLC with amperometric detection at a glassy carbon electrode	17

[a] NTA, nitrilotriacetic acid.
DPP, differential pulse polarography.

A correct determination of nitrilotriacetic acid in the presence of Fe(III) and Al(III) has been described,[66] and applications to natural and wastewaters have been carried out.[67] Tables 3 and 4 show applications of EC methods to the analysis of hydrazines, hydroxylamine, and nitrilotriacetic acid.

5.4 Aromatic Hydrocarbons

Polycyclic aromaric hydrocarbons (PAHs) are products of an incomplete combustion of organic materials, and can be found in some environmental matrices. Table 4 shows applications of ED methods to the analysis of aromatic hydrocarbons in environmental samples.

Either LC or HPLC is used for the analysis of PAHs; the analysis can be carried out by previous oxidation of the analytes with Ce(IV), and the final product (quinones) detected by reductive LC/ED at -0.4 V. The procedure is applied to tap water and motor oil.[16]

The direct analysis of PAHs is carried out by oxidative HPLC/ED at $+1.35$ V. An SPE procedure is used for the preceding extraction/preconcentration step.[17]

5.5 Polychlorinated Biphenyls

PCBs have been produced in many countries under different trade names, e.g. Aroclor®, Delor®, Sovtol®, or Kanerol®. The best known is Aroclor®, which is a mixture of chlorinated biphenyls characterized and numbered by the total percentage of chlorine in the mixture. The PCBs are toxic, the lethal oral dose (LD_{50}) is $2–10\,g\,kg^{-1}$ in mammals, and produce cancer. Because of their toxicity, the use of PCBs has been subject to some governmental regulations in many countries.

The determination of PCBs is of great environmental concern, because their release from industrial processes, combined with their insolubility in water and their resistence to biological degradation, has resulted in widespread contamination of water and soil. For several years PCBs have been recognized as ubiquitous environmental pollutants and these molecules are still presents in the environment, even though the production of PCBs has been banned. The conventional methods for quantifying PCBs are GC/MS or GC with electron capture detection.[68] Table 5 shows applications of EC methods to the analysis of PCBs in environmental samples. Aromatic halogen-containing organic compounds can be reduced according to:[69]

$$RX + 2e^- + H^+ = RH + X^-$$

with reduction potentials being markedly affected by the nature and number of substituents in the aromatic ring. Although EC methods have not been widely applied for quantifying PCBs, some examples can be found, such as the following (all cited in[69]): Kemula and Kreminska (1960) combined column chromatography and polarography to separate various isomers of dichlorodiphenyltrichloroethane (DDT) quantitatively; Colas et al. (1964) used silica gel to separate a mixture of two isomers of endosulfan, followed by polarography in a water–acetone mixture; Supin and Budnikov (1973) used the properties of α-benzene hexachloride at a slow DME to its determination at levels down to $60\,\mu g\,L^{-1}$ in pure solution; and Farwell et al. (1975) reported the use of interrupted-sweep voltammetric analysis to identify polychlorinated insecticides and other polychlorinated aromatics. However, this method requires at least $9\,\mu g$ of relatively pure compound to obtain positive identification. Therefore it is not very helpful for the analysis of pesticide residues in the environment.

In 1984 Lam and Kopanica[18] carried out an important EC application for the environmental analysis of PCBs. They determined TCB by adsorptive stripping voltammetry at a HMDE in the concentration range $0.004–1\,mg\,L^{-1}$. Preconcentration is achieved by adsorption at a potential of -0.40 V, and desorption at -1.10 V. This method was applied to the analysis of waste and natural waters with a relative standard deviation (RSD)

Table 5 Environmental analysis of PCBs and phenols by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
PCBs	waste mineral oils	12 mg g^{-1}	dechlorination by sodium and potentiometric measurement	70
(TCB)	waste and natural water	0.0015 mg L^{-1}	adsorptive stripping voltammetry at a HMDE: deposition potential −0.4 V, potential range from −0.4 to 1.3 V	18
(Aroclors®)	industrial effluent water	0.39–3.3 µg L^{-1}	EC immunosensor, specific binding between PCB and anti-PCB antibody-immobilized conducting polymer matrix	19
Phenols (2-Nitrophenol)	seawater	0.03 mg L^{-1}	FIA coupled to differential pulse voltammetry at a bentonite-modified carbon paste electrode	88
(2-Methyl-4,6-dinitrophenol)	seawater	3.2 µg L^{-1}	differential pulse voltammetry at a hidepower-modified carbon paste electrode	86
(Total phenols)	industrial wastewater		LC or FIA system with two carbon electrode in series: coulometric and amperometric detectors	24
Phenols	surface water		peroxidase modified carbon paste/solid graphite electrode at −0.05 V (versus saturated calomd electrode)	93
	seawater	0.01–1.9 µg L^{-1}	LC with ED using a dual electrode analytical cell equipped with a coulometric detector and an amperometric electrode	20
	seawater	[a]0.9–6.1 µg L^{-1} [b]5–45 ng L^{-1}	LC with ED at a glassy carbon electrode at 1.25 V	21
	river, tap water	0.01–1 µg L^{-1}	LC with ED at a glassy carbon electrode at 1.1 V	22
	tap, wastewater	5×10^{-8} M	adsorptive stripping voltammetry at 1.8 V using a carbon paste electrode with solid paraffin as a binder	95
	industrial wastewater		amperometric biosensor based on polyphenol oxidase	101
	wastewater	0.01–0.12 mg L^{-1}	Capillary electrophoresis with ED	97
	municipal waste		HPLC with ED	98
(Catecol, phenol)	surface water	0.25 µg L^{-1}	tyrosinase-modified solid graphite electrode	92
		4 µg L^{-1}	tyrosinase containing apoxy/graphite electrode	92
	wastewater	2–3 nM	tyrosinase-modified solid graphite electrode	25
	river, wastewater		Biosensor: catecol is determined by the reduction of its oxidation product on glassy carbon electrode after conversion to quinone in an on-line coconut-based reactor by FIA	99
	river water	6 nM	dual enzyme electrode chromatographic detection; carbon paste tyrosinase amperometric detector using short-chain hydrocarbons as pasting liquids	100
(Priority phenolic pollutants)	effluent water	0.03–2 µg L^{-1}	LC with ED at a glassy carbon electrode	89
	tap, drinking water	<35 ng L^{-1}	HPLC with amperometric detection	96
	river water	>0.4 ng L^{-1}	HPLC with coulometric detection using a dual electrode analytical cell equipped with two glassy carbon electrodes	23
(Chlorinated phenols)	natural water	25 ng L^{-1}	HPLC with ED using a supported liquid membrane extraction system	102
(Phenol, *p*-cresol)	river, groundwater		tyrosinase modified carbon paste electrode coupled to a 50 ft shielded cable for field monitoring	94
(Phenol, *o*-cresol)	tap, wastewater		LC with amperometric detection at a glassy carbon electrode	26
(Phenol, alkyl-, hydroxy-, and chlorophenols)	waste, river water	2.75–950 µg L^{-1}	HPLC with EC dual detector	90

[a] Direct injection.
[b] After SPE.

of below 5%. The cyclic voltammogram of TCB yielded a sharp cathodic peak, about four times higher than the anodic peak. TCB was accumulated on the surface of the HMDE in the more positive potential range and later it was stripped off at more negative potentials.

The current measured in the voltammetric determination of TCB depends on the pH of the supporting electrolite. The optimun pH is in the range 6–7, where the TCB peak is well separated from hydrogen evolution. Because of the insolubility of TCB in water, voltammetric determination must be done in a buffer solution containing methanol. The dependence of the peak current on the TCB concentration was linear in the range 1×10^{-8}–1×10^{-6} mol L^{-1} TCB and the detection limit was found to be 1.5 μg L^{-1}.

Polarographically active metal ions such as Cu^{2+}, Pb^{2+}, Cd^{2+}, or Zn^{2+} would not interfere if their concentration in the sample solution was up to 10 time greater than that of TCB; otherwise, masking with EDTA was necessary. However, surface-active compounds and organic compounds exhibiting adsorption properties similar to those of TCB may interfere; so, DDT can be determined under practically the same conditions as TCB. Thus it interferes with the TCB determination. The procedure was applied to different kinds of water and the results agreed very well with those obtained by the spectrophotometric measurements (Table 6).

The screening of PCBs in used mineral oils is carried out by a fully automated robotic method based on the dechlorination by sodium and a potentiometric measurement. The robot performs the weigh-in of the sample, adds the reagents, develops an LLE and obtains the concentration using a chloride-selective electrode. The detection limit is below the regulatory limit established for a contaminated waste oil.[70]

A current trend dealing with the EC analysis of PCBs is the development of EC immunosensors.[47] Del Carlo and Mascini[71] have developed an EIA with amperometric flow-injection analysis. To do that, they used a horseradish peroxidase (HRP) as a label, and its application to the determination of PCBs. They employed a competitive assay, where conjugated PCB–protein (gelatin) was immobilized to the solid phase (microtitre assay plate) and the competition was carried out with PCB standard using a limiting amount of anti-PCB IgG. The extent of the competition was evaluated using a secondary, HRP labeled, IgG. The PCB range was 0.1–50 μg mL^{-1}. Another development is a disposable screen-printed electrode for the immunochemical detection of PCBs.[72] Three different EC approaches were compared (amperometry, chronoamperometry and differential pulse voltammetry) for the detection of α-naphthol and a DPP procedure was optimized to evaluate the detection of low levels of alkaline phosphatase, an enzyme which catalyzes the hydrolysis of α-naphthyl phosphate to α-naphthol. Finally, a competitive EIA for PCBs, made of alkaline phosphatase, was performed using the optimized ED of α-naphthol. The procedure was not applied to real water samples.

Sadik and Bender[19] have developed a direct EC immunosensor for the determination of PCBs in water. The assay is based on the measurement of the current due to the specific binding between PCB and anti-PCB antibody-immobilized conducting polymer matrix. The immunosensor showed a high selectivity for PCBs in the presence of a potential interference such as chlorinated anisole, benzenes, and phenols. Electrode antibody membranes were prepared using three main formats (Figure 2).

- *EC microassembly*. The PCB immunosensors were prepared by constant potential (0.7 V) EC deposition from a 0.2 M pyrrole monomer solution onto platinum or silver substrates. An antibody raised against PCBs served as the counterion in the polymerization process. Antibody layers were further adsorbed onto the pyrrol layer by dipping the polypyrrole (PPY)/$H_2PO_4^-$ electrodes into a solution containing PCB antisera, which was made up in phosphate buffer saline (PBS) buffer (pH 7.4).
- *Physical entrappment*. EC polymerization of the PPY/antibody membrane was carried out. The films were grown using a charge of 160 mC, and the thickness of the antibody layers was controlled by means of the charge which passed during the polymerization.
- *Passive adsorption*. A silver or platinum substrate (0.3 cm^2) was preconditioned and used as the base electrode. The antibody layers were also adsorbed onto the substrates by dipping the electrodes into a solution containing PCB antisera, which was made up in PBS buffer (pH 7.4). The electrodes were inmersed

Table 6 Determination of TCB in waste and natural waters. (Reproduced by permission of Elsevier Science from ref. 18.)

Sample no.	TCB found (mg L^{-1})		
	Spectr. method	Proposed method[a]	RSD (%)
1	0.07	0.065 ± 0.001[b]	1.6
2	1.04	1.025 ± 0.04[c]	4.2
3	11.20	10.75 ± 0.19[d]	1.8
4	68.30	67.8 ± 3.05[d]	4.5
5	167.08	160.2 ± 5.3[d]	3.3

[a] Mean and standard deviation for four measurements.
[b] Drinking water.
[c] Creek water.
[d] Wastewater.

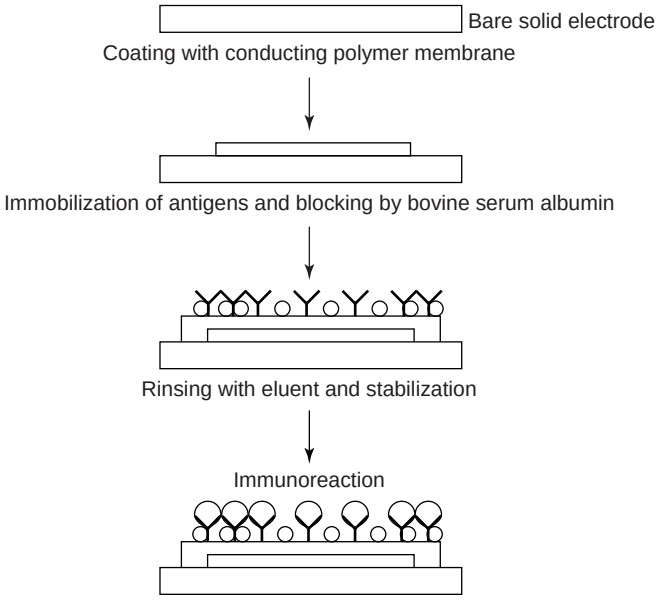

Figure 2 Diagrams of the procedures for preparation and measurement with an immunosensor. (Reproduced by permission of the American Chemical Society from ref. 19.)

in PBS at pH 7.4 after preparation for removal the adsorbed proteins and then thoroughly rinsed with PBS at pH 7.4, followed by air-drying. The dried PPY/antibody electrodes were stored covered at 4 °C.

Antibody membrane electrodes are characterized by using cyclic voltammetry, chronoamperometry, and enzyme-linked immunosorbent assay (ELISA) protocols. The immunoreaction detection technique employed is referred to as a pulse-accelerated immunoassay technique (PAIT). The analytical signal was generated by applying a pulsed waveform between +0.60 and −0.60 V with a pulse frequency of 120 and 480 ms. The oscillating potential reversibly drove the antigen-antibody binding process. The current arising from this process was monitored in real time.

The sensor can be used for analyte detection in a static or flow injection (FI) mode by applying pulsed potentials between the sensor surface (or WE) and the reference electrode. The current obtained can be directly related to the rate of EC reactions occurring at the working (or sensing) electrode, which may be related partly to the concentration of the analyte in solution. Linear calibration curves of the peak heights (nA) versus concentration ($\mu g\,L^{-1}$) were obtained for various Aroclors®. The immunosensor exhibited a high selectivity for PCBs in the presence of potential interferences such as the chlorinated anisole, the benzenes, and the phenols. The highest cross-reactivity measured for chlorinated phenolic compounds relative to Aroclor® 1248 was less than 3%. Recoveries of spiked Aroclors® 1242 and 1254 from industrial effluent water, rolling mill, and seafood plant pretreated water ranged from 103 to 106%. The detection method may be applied to the continuous monitoring of effluent such as waste streams and groundwater.

5.6 Phenols

Phenolic compounds are released into the environment in a variety of ways, because they are used in many industrial processes. These include the manufacture of industrial products, such as plastics, pharmaceuticals, dyes and pesticides; agriculture also uses insecticides containing toxic nitrophenols.[73–76]

The environmental concern of phenol pollution results from the high toxicity of many phenol derivatives to living organisms.[77] Many phenols have a well established reputation as toxic pollutants due to their toxicity and persistency in the environment. According to European Community legislation, the maximum admissible concentration of phenols in water for human consumption is $0.5\,\mu g\,L^{-1}$ for the total content and $0.1\,\mu g\,L^{-1}$ for individual compounds, excluding those natural phenols which do not react with chlorine; in bathing water the maximum admissible value is $5\,\mu g\,L^{-1}$.[78,79] The analysis of phenols is very important in environmental programs due to their presence in natural and wastewaters. The USEPA lists eleven phenols as priority pollutants.[80]

Different methods have been developed for the analysis of phenols. Traditional methods recommended for their determination are based on the measurement of an index of phenols using spectrophotometric methods based on 4-aminoantipyrine (4-AAP) or 2-nitrophenol;[81,82] the result is expressed as total phenols, as individual phenols are not measured. These procedures have also been adapted to FIA, involving in-line concentration by solvent and sorbent extraction.[73]

GC allows the speciation of the various substituted phenols after LLE. The USEPA Method 604 for phenols determination involves the serial extraction of an acidified sample with dichloromethane.[83] The extract is dried and exchanged to 2-propanol, typically using Kuderna–Danish glassware, after that the phenols are analyzed by GC using a flame-ionisation detector. Alternatively, USEPA Method 625 for extractable base/neutrals and acids can be used for phenols determination.[84] The major limitation of these GC methods is the extraction procedure, where losses can occur during the steps of extraction, concentration, and solvent exchange. The extraction is also very laborious and intensive, so it requires specialized glassware as well as the use of chlorinated solvents. Table 5 shows applications of EC

methods to the analysis of phenols in environmental samples.

EC methods have been applied to the determination of phenols, with or without a separation step. Differential pulse voltammetry[85,86] and FI differential pulse voltammetry[87,88] were used for the determination of several nitrophenols in spiked seawater and/or spiked river water employing modified carbon paste electrodes. The reduction waves of nitrophenols were used for quantification. Furthermore the interferent effects of several anionic species, cationic species, and other phenols were studied. The major interferences on individual nitrophenols were caused by some cationic species and other nitrophenols; chlorophenols, generally, did not interfere. The use of an adequate potential to accumulate certain nitrophenols on the WE can be used to avoid the effect of some interferents.[86] The modified carbon paste electrodes can be also used for oxidations and the determination of several phenols in real water samples. However its sensitivity and selectivity could be improved if a preconcentration and/or separation step is applied.

A FI method using a dual EC detector was described for the analysis of total phenols in water and wastewater samples.[24] The method relies on an in-line oxidative EC removal or an elimination of interferences at an upstream large surface area (coulometric) electrode combined with the oxidative detection of all phenols at an amperometric electrode. The method was compared with the standard colorimetric procedure based on 4-AAP. The results clearly indicate the increased accuracy of the FI method for the analysis of phenols in comparison with the data derived from the colorimetric procedure.

LC with an EC detector is widely used for the analysis of phenols. As with GC, LC also allows the speciation of various substituted phenols. The main advantage when using LC are that there is no requeriment for precolumn derivatization and the need for sample pretreatment is minimal. So, LC and HPLC with ED are being widely applied to the measurement of phenols in environmental samples. Paterson et al.[89] have developed a liquid chromatographic method for the determination of priority pollutant phenols at low microgram per liter levels in environmental waters. The phenols were separated using a Nova-Pak Phenyl column and optimized continuous (concave) gradient elution conditions, and subsequently they were detected using an EC detector equipped with a glassy carbon electrode, operated at +1.15 V (Figure 3). This approach resulted in minimal sample interferences and practical quantitation limits in the low microgram per liter range. The advantage this method offers over the established GC and GC/MS phenol methods is that no sample extraction is required, hence avoiding the losses which occur during the steps of

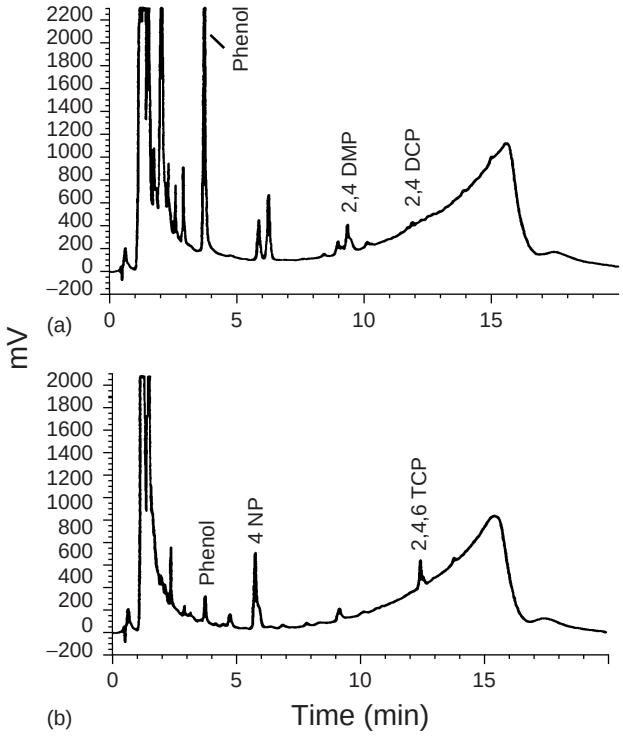

Figure 3 Example LC/ED chromatograms obtained for (a) saline oil refinery effluent (2,4 DCP = 2,4-dichlorophenol; 2,4 DMP = 2,4-dimethylphenol) and (b) paper mill effluent samples (4 NP = 4-nitrophenol; 2,4,6 TCP = 2,4,6-trichlorophenol). (Reproduced by permission of Elsevier Science from ref. 89.)

extraction, concentration and solvent exchange required for the GC-based analyses. The method was evaluated following the determination of the 11 priority pollutant phenols in five 'blind' environmental water matrices. These samples were supplied (in duplicate) as part of a round-robin testing program involving nine laboratories. The samples, a saline oil refinery effluent, a spiked refinery effluent, a paper mill effluent, a spiked paper mill effluent and a spiked reagent grade water, were analyzed using either GC or LC methodology by the nine laboratories. The LC/ED sample chromatograms were virtually free of sample interferences. The LC/ED results and recoveries agreed with the results reached by the nine laboratories.

The HPLC technique also offers the possibility of separating high-polar and high-boiling phenols. The differences in their chromatographic behavior require gradient techniques to achieve good resolution in a short analysis time. ED can increase the sensitivity and selectivity for phenols, in comparison with other detection techniques, especially when an EC dual detector is used. Thus, Hagen et al. propose an isocratic HPLC with programmed flow-rate-variation and ED on a glassy carbon dual electrode in parallel mode for the determination of phenol, alkylphenols, hydroxyphenols,

and chlorophenols.[90] The dual mode is used to increase selectivity. Whereas the WE, WE1, is held at +1.1 V, the other electrode, WE2, is held at +0.8 V. The potential of +1.1 V is necessary to reach the diffusion-controlled region of oxidation for all phenols. At a potential of +0.8 V only *p*-cresols (no *o*- and *m*-) are oxidizable. Therefore, an EC separation of the chromatograpically nonresolved cresols is possible. The methods were successfully applied to the direct determination of phenols in samples of wastewater of a coal industry, and in samples of river water.

Table 7 shows the obtained results. The qualitative classifications were performed according to capacity factors (k') and peak-current ratios (I_{WE1}/I_{WE2}). Calibration curves served as the basis for quantification.

For the lowest phenols concentrations a preconcentration step is usually carried out. LC/electrochemistry was employed for the analysis of phenol traces in the range from 100 parts per trillion to 500 parts per million.[26] Various sample preparation schemes were examined. The method was applied to wastewater from an oil refinery, municipal drinking water, and shale oil viscous materials. Direct injection was satisfactory for wastewater samples with phenol concentrations greater than 20 ppb. Municipal drinking water was analyzed for trace of phenols and chlorophenols in the 0.1–10 ppb range using trace enrichment. Regarding viscous shale oil materials, simple LLE into a base was enough to guarantee the measurement selectivity when coupled to LC/EC. Dual electrode detection provided additional information about peak identity.

Table 7 Determination of phenols in waste and river waters. (Reproduced by permission of Springer-Verlag from ref. 90.)

Pollutant	Wastewater concentration ($mg\,L^{-1}$)	River water concentration $\mu g\,L^{-1}$ sample			
		1	2	3	4
Hydroquinone	2.69	–	–	–	–
Resocinol	9.64	14	368	212	230
Catechol	0.72	–	–	–	–
Phenol	5.39	61	–	41	31
Cresol	4.37	197	13	–	83
p-Cresol	1.46	92	–	–	57

– not detected.

Most reports about the analysis of phenols use SPE as the extraction technique. Ruana et al. have described a method for the determination of trace amounts of phenols in drinking and river water.[91] They employ solid–liquid extraction of samples with laboratory-made microcolumns and LC with UV and ED to separate and identify the 11 EPA priority pollutant phenols. The EC detector was connected in series to the outpout of the UV detector and worked in the amperometric mode. The EC detector has a pretreatment function that allows the cleaning of the WE surface electrochemically. This function can occur before or in the middle of the analysis, avoiding premature fouling of the electrode. The detection limits obtained with prior concentration of the samples were 40–600 $ng\,L^{-1}$, depending on the analyzed phenol.

The efficiency and the reproducibility of different solid-phase adsorbents such as C_{18} cartridges, C_{18} and poly(styrene–divinylbenzene) (PS-DVB) membrane extraction disks were evaluated for extractions of phenols from water at microgram per liter concentration levels. LC with coulometric detection was used for the analysis of the organic extracts.[20] The limit of detection (LOD) for phenol and various chlorophenols in seawaters were estimated by the three different SPE protocols (Table 8). For C_{18} cartridges, LODs higher than those for SPE disks were obtained. The low breakthrough volume of phenol (20 mL) for C_{18} cartridges prevented the determination of this compound at concentrations lower than 0.25 $\mu g\,L^{-1}$.

Table 8 LOD ($mg\,L^{-1}$) of phenols in seawater using LC/ED after enrichment with different SPE cartridges and disks. (Reproduced by permission of Elsevier Science from ref. 20.)

Compound	SPE technique				
	Cartridge C_{18}			Disk C_{18} 250 mL seawater	Disk PS-DVB 250 mL seawater
	20 mL seawater	100 mL seawater	250 mL seawater		
Phenol	0.25	–[a]	–[a]	–[b]	0.1
o-Chlorophenol	0.33	0.08	–[a]	–[a]	0.01
2,4,6-Trichlorophenol	0.37	0.09	0.03	0.03	0.03
2,4-Dimethylphenol	0.46	0.07	0.03	0.01	0.01
4-Chloro-3-methylphenol	1.64	0.31	0.12	0.04	0.04
2,4-Dichlorophenol	1.52	0.27	0.13	0.05	0.04
Pentachlorophenol	1.90	0.33	0.12	0.04	0.04

[a] Sample volume higher than breakthrough volume.
[b] An interferene prevented quantification.

For C_{18} and PS-DVB membrane extraction disks better LODs (between 0.01 and 0.04 $\mu g\,L^{-1}$) than for C_{18} cartridges were obtained because larger water volumes could be preconcentrated and, moreover, in a shorter period of time. The PS-DVB disks allowed the analysis of all phenols studied at the legislated level of 0.1 $\mu g\,L^{-1}$.

The PS-DVB membrane extraction disks were used as sorbents for the on-line solid phase extraction of 13 phenols (nitro and chlorophenols) in river and tap water by LC/ED.[22] High water volumes, up to 250 mL, can be preconcentrated without loss of phenols (recoveries between 80% and 100%) except for the more polar ones. The LODs were studied in HPLC-grade water, tap water, and river water after preconcentrating differents volumes of each one spiked by a low level concentration of phenols. In natural waters, detection limits were always higher than in HPLC-grade water. This fact was due to the matrix effects of natural water which especially affected to the compounds eluted with relatively short retention times; these limits were between 0.01 and 0.1 $\mu g\,L^{-1}$ in tap water and between 0.1 and 1.0 $\mu g\,L^{-1}$ in river water. The method was applied to two river water samples (both crossing industrial zones) employing the standard addition method for quantitative analysis.

A high-sensitivity method for the determination of polar phenolic compounds in water samples was developed. Water samples were preconcentrated using on-line SPE with a styrene–divinylbenzene sorbent. They were analysed by LC/dual coulimetric detection, with the first electrode set at low potential for sample clean-up and the second electrode being used for analytical purposes.[23] The method was applied to spiked groundwaters. Detection limits at part per trillion level were attained.

A reversed-phase HPLC procedure with gradient elution and ED, coupled with a preconcentration step by SPE using a derivated PS-DVB, is described for the determination of various phenolic compounds.[21] The successful method was used to determine phenols in seawater samples collected from a contaminated coastal area by wastewaters from an iron works, a steel factory, and a refinery; although the matrix was complex and contaminated by aliphatic and polycyclic aromatic hydrocarbons, the determination of phenolic compounds was simple and not affected by interferences.

A present trend for the analysis of phenols in environmental samples is the use of biosensors. The catalytic properties of several enzymes have been exploited for the construction of biosensors with narrow or broad selectivity for phenolic compounds. The sensors can be applied to the determination of phenols in real surface waters. Analyte-specific biosensors (type I) can be constructed by choosing a specific enzyme that catalyzes the conversion of a specific analyte structure, leading to a narrow substrate selectivity. A group-specific biosensor (type II), leading to a broader substrate selectivity, can be constructed by choosing an enzyme that catalyzes a whole group of related analytes of similar structure or several analyte-specific enzymes that are co-immobilized in the same biosensor. Both types of biosensors have been developed on the basis of phenol oxidases (tyrosinases, laccases) and peroxidases, using different electrode material, flow systems, and sample pretreatment techniques.

Although oxidases and peroxidases have different enzymatic mechanisms, their action in amperometric biosensors for detection of phenolic compounds can be illustrated with a similar reaction sequence at an enzyme-modified electrode (Figure 4).

Amperometric electrodes for the determination of phenolic compounds are essentially two types:

- Surface-modified solid electrodes, where the enzyme is physically adsorbed, covalently immobilized or membrane retained.
- Bulk-modified composite electrodes, where the enzyme is incorporated in the bulk of the electrode material.

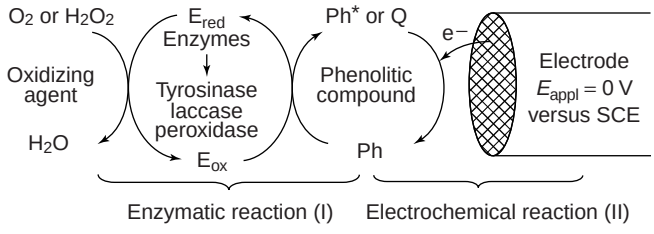

Figure 4 Operational mechanism of enzyme-modified electrodes for the determination of phenolic compounds. E_{red} and E_{ox} are the reduced and oxidized forms of the enzyme, respectively; Ph, Ph*, and Q are the phenolic molecule, its phenoxyradical and its quinone form, respectively. (Reproduced by permission of Elsevier Science from ref. 49.)

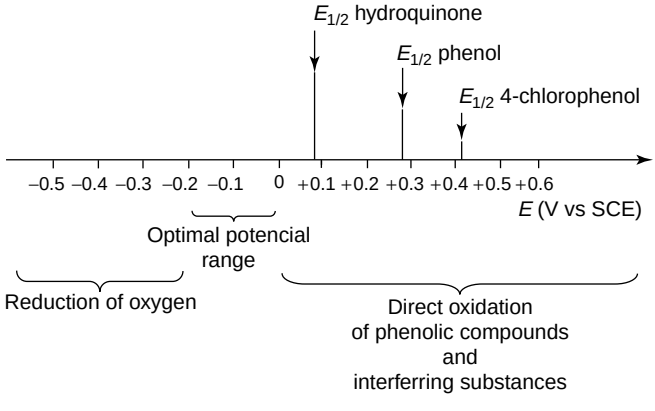

Figure 5 Optimal potential range for the operation of amperometric biosensors. (Reproduced by permission of ref. 92.)

The optimal potential range for the development of amperometric phenolic biosensors lies between −0.2 and 0 V (Figure 5). In this region the background current switches signs, leading to low noise and background currents, and thereby enabling the detection of low analyte concentrations. Outside this region several interference reactions can occur; special care should be taken with environmental water samples that contain different amounts of electrochemically active matrix compounds, such as humic substances.

For use with real water samples, biosensors are generally incorporated into fully integrated analytical flow systems, whose choice is often determined by the concentration of phenolic compounds in the sample and the complexity of the sample matrix. At levels of $1\,\mu g\,L^{-1}$ and below, direct measurements with any type of biosensor are practically impossible because of the lack of sensitivity (except for the best enzyme substrates).[92] A sample-enrichment step is required to extract the phenolic compounds from larger sample volumes, which is not usually a limiting factor for environmental applications. Solute extraction can be performed off-line on a membrane extraction disk or a disposable extraction cartridge, where the enriched sample is injected into a flow system. On-line systems are designed to eliminate tedious manual sample-handling procedures and contamination. SPE is an efficient sample handling technique that allows phenols to be monitored in highly complex samples at higher level, such as industrial wastewaters and effluents.

Biosensor applications to the determination of phenols in real samples are not still very abundant. An amperometric biosensor based on the phenol oxidase tyrosinase immobilized on solid graphite electrodes as a selective and sensitive detector in column liquid chromatography (CLC), with SPE as sample handling step, was used for the determination of phenolic compounds in spiked wastewater samples obtained from a pulp-producing plant in the south of Sweden.[25] Peroxidase-modified electrodes as amperometric detectors in FI systems were employed to determine phenols in humic-containing surface water from the Ebro river in Spain.[92,93] A remote EC biosensor for field monitoring of phenolic compounds[94] was proposed. The sensor is based on a tyrosinase enzyme electrode connected to a voltammograph via an enviromentally sealed connector and a 50 ft shielded cable. The submersible probe offers highly selective measurements of micromolar concentrations of phenolic substrates in untreated river and groundwater samples.

5.7 Pesticides

Pesticides are materials created to destroy or control any pest. The term 'pesticides' may also refer to substances used as plant-growth regulators, defoliants, agents for preventing the premature fall of fruit, and chemicals applied to crops either before or after the harvest to inhibit deterioration during storage or transportation. The term pesticide includes the diverse range of substances variously classified as insecticides, fungicides, molluscicides, bactericides, rodenticides, fumigants, and herbicides.

Pesticides are highly toxic compounds. The health hazards arising from their use include bone marrow disorders, carcinogenicity, infertility, cytogenic effects, immunological and respiratory problems, and some of them (organophosphates) are powerful inhibitors of enzymes involved in nerve functions.[103] The organochlorine insecticides (such as DDT, aldrin, and lindane) used in the past have been progressively replaced with organophosphates and carbamates, which have low environmental persistence though they are still highly toxic compounds. For these reasons, the 1989 European Community Water Act states that the maximum admissible concentration of all pesticides in drinking water should be below $0.5\,\mu g\,L^{-1}$ and the individual pesticide concentration below $0.1\,\mu g\,L^{-1}$.

A number of methods can be used for the determination of the pesticide content in environmental samples.[104] Most of the early methods used were based on spectrophotometric procedures or the determination of total phosphorus. These methods were slowly superseded by GC which is undoubtedly the most common and, when coupled to FTIR or MS, has the ability to fingerprint pesticides. However, because of the thermal lability and low volatility of some pesticides, cold methods such as LC and EC methods are increasingly used for the determination of these compounds. Table 9 shows the application of EC methods to the analysis of pesticides in environmental samples.

The direct determination of pesticides in water without separation/preconcentration techniques has rarely been carried out; the EC reduction of 1,2-dibromoethane (EDB) in aqueous solution has been studied with direct current, normal pulse and reverse pulse polarography, voltammetry, and coulometry.[105] The reduction mechanism involves a slow two-electron transfer with a fast displacement of bromide leading to the formation of ethene. The EDB can be determined directly from the limiting reduction current or indirectly from the limiting current for the anodic oxidation of mercury in the presence of the reduction product, bromide.

Adsorptive stripping differential pulse voltammetry was employed for the determination of the organophosphorus pesticide azinphos-methyl (Gution).[106] Accumulation is achieved by the adsorption of the compound on an HMDE. The method was applied to the determination of Guthion in river water. The samples were collected

Table 9 Environmental analysis of pesticides by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
Pesticides (Azinphos-methyl, Guthion)	river water	$0.2\,\mu g\,L^{-1}$	adsorptive stripping voltammetry at a SMDE,[a] accumulation potential −5 mV, potential range from −5 mV to −1.1 V	106
(Paraquat)	river water	$0.1\,mg\,L^{-1}$	CSV at a carbon paste electrode modified with Amberlite® XAD-2 resin	107
	well, river and lake water	$1-18 \times 10^{-7}\,M$	selective potentiometric sensor based on octamethylcyclotetrasiloxane; use as detector for FIA	112
(Ioxinil and 2-methyl-3-nitroaniline)	drinking water	0.1^{b}, $0.3^{c}\,mg\,L^{-1}$	voltammetry at a C_{18}-modified carbon paste electrode	108
(Triazine herbicide, Metamitron)	natural water	$0.07\,mg\,L^{-1}$	adsorptive stripping with differential pulse or square wave voltammetry at a carbon paste electrode modified with silica, preconcentration potential −0.8 V or open circuit potential conditions	109
	tap water	$<1\,mg\,L^{-1}$	thick film electrode printed with monoclonal antibodies as biorecgnition element	118
(Carbamate pesticides)	river water	$2.6-22\,\mu g\,L^{-1}$	microarray amperometric detection using a Kel-F®-graphite (Kelgraph) electrode at 1.1 V	110
(Phenylurea herbicides)	tap, mineral and spring water	$2.1-4.9 \times 10^{-4}\,\mu g\,L^{-1}$	HPLC coupled to four coulometric array cell modules, each containing four electrochemical detector cells; 0 V at electrode 1, with increments of 80 mV at each subsequent electrode until a value of 1.2 V at electrode 16	28
(Amitrole)	drinking and groundwater	$0.05\,\mu g\,L^{-1}$	HPLC/coulometric array electrode detector, equipped with 16 porous graphite WEs	111
	drinking and groundwater	$0.1\,\mu g\,L^{-1}$	HPLC/amperometric detection with a glassy carbon WE at 1.1 V	27
(Organophosphate pesticide, Malathion)	drinking water	$0.1\,mg\,L^{-1}$	biosensor, inhibition of acetylcholinesterase activity; enzyme membrane placed on the surface of an ISFET with Al_2O_3 as the pH-sensitive layer	115
	river water	$4-7 \times 10^{-1}\,M$	screen printed amperometric biosensor in a FI system	119
(Atrazine)	drinking water	$<0.1\,\mu L^{-1}$	electrochemiluminescence FI immunoassay	120
	natural water	$13-38\,mg\,L^{-1}$	differential pulse voltammetry at a SMDE	121
(Propazine)	drinking water	$2.8 \times 10^{-7}\,M$	DPP in a micellar and emulsified media	122
(Nitralin herbicide)	groundwater	$8.7 \times 10^{-11}\,M$	adsorptive stripping differential pulse voltammetry at a mercury electrode	123
(Chlorpyrifos)	wastewater		cathodic adsorptive stripping voltammetry at a HMDE	124
	underground water, wastewater	$8.7 \times 10^{-7}\,M$	DPP	125
(2,4-Dichorophenoxyacetic acid)	tap water	$<0.01\,mg\,L^{-1}$	screen-printed amperometric electrode and monoclonal antibodies	126
	drinking water		amperometric FI immunoanalysis	127
(Pyridafenthion)	river water	$34\,mg\,L^{-1}$	adsorptive stripping voltammetry at a HMDE	128

[a] SMDE, static mercury drop electrode.
[b] Detection limit for loxinil.
[c] Detection limit for 2-methyl-3-nitroaniline.
CSV, cathodic stripping voltammetry. Kel-F®, registered 7 M of 3 M Co.

from a river near where a factory that produces the insecticide is located. As no residues of Guthion were found, the procedure was applied to spiked river water. The determination limit was 1.5 ng mL^{-1}.

CSV was used for the determination of the herbicide paraquat (1,1′-dimethyl-4,4′-bipyridylium ion) on a carbon paste electrode modified with Amberlite® XAD-2 resin.[107] The procedure exhibits good linearity for concentrations of paraquat lower than 1.08 mg mL^{-1} with a detection limit of 0.10 g mL^{-1}. The method was applied to river water samples, but the presence of paraquat was not detected. To establish possible matrix effects, samples of river water were spiked with paraquat then analyzed with the proposed method, obtaining recoveries of 104% with a precision of 6%.

A voltammetric procedure for the determination of the pesticide Ioxynil (4-hydroxy-3,5-diiodobenzonitrile) by adsorptive stripping using a C_{18} carbon paste-modified electrode was carried out and applied to determination of the compound in drinking water;[108] however, the sensitivity was not enough to control concentrations of pesticides according to the national regulations. Also an anodic voltammetric assay using a silica modified carbon paste electrode for the herbicide Metamitron (4-amino-3-methyl-6-phenyl-1,2,4-triazin-5-(4*H*)-one) was proposed to determine the compound in natural water[109] with a detection limit of 75 ng L^{-1}.

LC with ED is one of the most-used techniques for the determination of pesticides and has been applied to the determination of carbamate pesticides spiked into a river water matrix using a series of Kelgraf microarray electrodes.[110] The Kelgraf electrode effectively discriminates between oxidation reactions limited by the rate of mass transport and reactions (including solvent oxidation) limited by the rate of electron transfer or other surface processes, affording improved detection limits at high applied potentials. Subnanogram detection limits are obtained with minimal sample clean-up or pretreatment. The HPLC technique with coulometric electrode array detection[111] and with amperometric detection[27] was also employed for determination of amitrole (3-amino-1,2,4-triazole) in spiked drinking and groundwater. A sensitive and simple method for the simultaneous evaluation of phenol and 26 substituted phenols and phenylurea herbicides was developed using HPLC with multielectrode ED.[28] After extraction from the samples on solid-phase cartridges, the compounds were separated on a reversed phase column by using a combined gradient of organic modifier and counterion. The identification of the compounds was based on the retention time comparison with authentic standards. The detection limits for the 27 compounds studied was found to be much lower (5×10^{-4} g L^{-1} for the least-sensitive compound, the herbicide Linurom) than that indicated by the European Community. The method was applied to the determination of three phenylurea herbicides and four phenols in three different samples of water, namely tap water from an aqueduct, water from a spring at 1600 m above sea level, and commercial mineral water.

The monitoring of trace levels of phenylurea and chlorotriazine herbicides and their metabolites was performed on river, estuarine and coastal seawater by off-line (*n*-octadecyl silica, C_{18}) and on-line (styrene–divinylbenzene copolymer-based) preconcentration techniques using reversed phase chromatography and UV and ED.[29] It was shown that the off-line and on-line technologies fit monitoring of trace-level phenylurea and triazine analysis (detection limits 10–50 ng L^{-1}) in fresh water as well as seawater and in natural water with high dissolved organic carbon content.

Other flow systems, such as FIA, have been combined with ED for the environmental analysis of pesticides. Thus an ion selective electrode (ISE) has been developed for the herbicide paraquat based on a membrane of octamethylcyclotetrasiloxane and highly lipophilic PVC-based components (i.e. plasticizer and anionic sites), fixed on the surface of platinum wire and subsequently used as a flow-through potentiometric detector. Analysis of 0.01 mM paraquat dichloride in the background of well, river and lake waters when performed in the FIA mode gave excellent recoveries.[112] A technique for the rapid and sensitive FI monitoring and analysis of mixtures of the triazine herbicides simazine, atrazine, and propazine using filter-supported bilayer lipid membranes was also described.[113]

EC biosensors and immunosensors are being developed for the environmental analysis of pesticides. In general, the pesticides are not the substrates of enzymes but acts as specific inhibitors of their activity. Consequently their detection is not generally based on the enzymatic catalysis of the pesticide but on the inhibiting capacity of a specific enzyme activity.[48] The biosensors used for determination of pesticides are of two types, potentiometric and amperometric.

Cholinesterases, inhibited by organophosphorus and carbamate pesticides, are practically the only enzymes used in pesticide potentiometric biosensors. The prevalence of the use of biosensors for organophosphorus and carbamate biosensors may be related to the fact that in 1994 they accounted for more than 57% of the total world insecticide market. Moreover, activity of cholinesterase enzymes can be easily monitored by EC methods. The two types of potentiometric transducer mainly employed are the ion-selective electrode (the glass pH electrode),[114] and the ISFET – a semiconductor that monitors pH change.[115,116] Other transducers, such as

Table 10 LOD of some carbamate and organophosphorus insecticides: comparison of potentiometric and amperometric transduction. (Reproduced by permission of ref. 46.)

Insecticide	LOD (ppb)	
	Potentiometric pH	Amperometric thiocholine oxidation
Aldicarb	1140	1.9
Carbaryl	1000	19
Carbofuran	6	0.02
Dichlorvos	300	22

light-addressable potentiometric sensor (LAPS), are less used.

With amperometric biosensors, choline oxidase is employed in addition to the esterases. This allows the indirect monitoring of choline by measurement of the consumption of oxygen or the production of hydrogen peroxide.[117] Furthermore, the direct anodic measurement of choline or thiocholine using a biosensor either with or without a suitable mediator of the electron transfer, can be used.[45]

The sensitivity reached with cholinesterase sensors depends on the type of cholinesterase and the source. Table 10 shows detection limits reached for amperometric and potentiometric sensors applied to the determination of several carbamate and organophosphorus insecticides; lower detection limits are always obtained for amperometric detection.

EIA technologies are a rapidly growing tool for pesticide measurement, and several EIA test kits are commercially available. These methods are indirect techniques, and detection of molecular interactions needs the use of a label such as an enzyme, fluorophore, or radioactive isotope. Using these kits it is possible to detect 0.02–2 $\mu g\,L^{-1}$ in water, depending on the pesticide and the antibodies involved.[46] The disadvantages of these commercial kits include slow equilibration time and irreversible binding which prevents re-use or continuous application.

The wide application of disposable strip tests in clinical analysis and environmental analysis for field control suggests that EC biosensor and immunosensors could also find a wide acceptance for the determination and monitoring of pesticides. Such sensors could serve as an alternative to chromatographic methods which, although more versatile in providing information about numerous species from one run, are much more expensive, require well-trained personnel, and cannot be used readily in the field. However, there are practical obstacles to be overcome before EC biosensors and immunosensors will make a significant impact in environmental applications.[45,7] In particular, it is necessary to improve the transduction mechanisms and to combine such improvements with the choice of biological transduction sequences and materials, and EC science.

5.8 Surfactants

Surfactants are an important source of pollution due to their extensive industrial and domestic use. Reviews on the determination of ionic and nonionic surfactants are provided, with application to surface water and

Table 11 Environmental analysis of surfactants and organometallics by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
Surfactants (Nonionic surfactants)	surface water	6–8 $mg\,L^{-1}$	indirect tensammetric method	31
	surface water and wastewater	–	polarographic determination as complexes with molybdophosphoric acid	32
(Poly(ethylene glycols))	river water		indirect tensammetric method	30
(Linear alkylbenzene sulfonates)	sewage and tap water		DPP of the nitro derivatives	33
(Ionic surfactants)	river/lake water and seawater		ISFET sensor	52
Organometallics (Tributyltin)	water		differential pulse anodic stripping voltammetry at a binuclear ruthenium complex coated glassy carbon electrode at −0.7 V	133
(Tributyltin)	seawater		alternating current polarography	34
(Methyl-, ethyl-, phenylmercury)	river water	(a) 8–20 $ng\,L^{-1}$ (b) 30–140 $ng\,L^{-1}$	HPLC with: (a) coulometric detection with a glassy carbon WE at +0.90 V; (b) amperometric detection with a carbon paste or glassy carbon WE at +1.15 V	35
(alkyllead compounds)	seawater	0.5 $ng\,L^{-1}$	differential pulse anodic stripping voltammetry at a hanging mercury drop WE	135

sewage.[129,130] Particular attention is given to indirect tensammetric techniques useful in the analysis of non-ionic surfactants. EC techniques applied for the determination of surfactants have been reviewed.[53] Table 11 shows applications of EC methods to the analysis of surfactants in environmental samples.

Potentiometric electrodes selective for ionic surfactants have been developed. These electrodes are ISEs with a liquid membrane containing a complex of an ionic surfactant that is insoluble in water and behaves as a liquid ion exchanger. Several ISFET devices have been developed to detect different ionic surfactants in environmental water samples, such as river, lake, and seawater.[52]

A membrane electrode based on poly(ethylene oxide) for determination by potentiometric titration of nonionic surfactants has been described as an alternative method for determination in wastewater.[131]

Primary biodegradation of nonionic surfactants can be tested with an indirect tensammetric procedure.[132] These compounds are also determined in river water by the same technique; the method for nonionic surfactants adsorbed on particles of river water combined the bismuth-active substances method with the indirect tensammetric method. Detection limits in the range $3–8\,mg\,L^{-1}$ are reported.[31]

Metabolic products from the biodegradation of surfactants are also the object of analysis. Poly(ethylene glycols) is one of the major products of the biodegradation of nonionic surfactants and their determination is of great interest for control of the aquatic environment. A process of extraction must be carried out in order to separate poly(ethylene glycols) and nonionic surfactants. Applications to river water are reported.[30]

Polarography has been applied to the determination of nonionic surfactants and linear alkylbenzene sulfonates. Nonionic surfactants are transformed into complexes with molybdophosphoric acid prior to their polarographic determination; applications to surface water and wastewater are reported.[32] Linear alkylbenzene sulfonates are anionic surfactants commonly used in the detergent industry; for these compounds, a previous nitration step with fuming nitric acid is proposed. Applications to sewage are described by DPP.[33]

5.9 Organometallics

A variety of organoarsenic compounds are found in the environment as the result of industrial activity. EC determinations of organic arsenic compounds offer the possibility of differentiating between different oxidation states of arsenic compounds. New EC developments include sensors for the determination of gaseous arsines and the use of capillary electrophoresis for a prior separation step.[50]

Organotin compounds have been extensively used as wood preservatives, stabilizers for PVC plastics, fungicides and antifouling paint additives. Organotin compounds have different degrees of toxicity depending on the nature and number of the alkyl groups bonded to the tin atom. Alkyltin compounds are more toxic than the aryl compounds, and the trisubstituted derivatives more toxic than the di- or tetrasubstituted derivatives. Analytical methods for organotins in the environment can be found in a review.[51] Most analytical methods applied to the analysis of environmental samples involve a derivatization step to transform organotin compounds into volatile hydrophobic analytes and their analysis by GC. However, methods based on LC or supercritical fluid chromatography are being developed to avoid some problems related to this procedure.[51] For example, a differential pulse anodic stripping voltammetry method at a binuclear ruthenium complex coated glassy carbon electrode has been described for organotins.[133] Alternating current polarography is used for the determination of tributyltin in seawater; samples were passed through a cation exchanger, and the eluate extracted with dichloromethane.[34]

Environmental toxicity of mercury compounds depends on their different chemical form and oxidation states. Symmetric organomercurials are inactive, whereas alkyl and aryl mercury have a high toxic activity; methylmercury leads to extremely toxic effects on biota and humans. Thus, it is important to know both the concentration of mercury compounds and their chemical forms and oxidation states. River water samples have been analyzed by reversed-phase HPLC with amperometric and coulometric detection, and the contents of methyl-, ethyl-, and phenylmercury have been determined. The procedure includes the complexation reaction of mercury species with pyrrolidinedithiocarbamate, and SPE of the complexes, prior to their introduction into the chromatographic system. A carbon paste or glassy carbon electrode is used for amperometric detection, and a glassy carbon electrode for coulometric detection. Limits of detection lower than $0.2\,mg\,L^{-1}$ were reached in all cases; the best results are always obtained with coulometric detection.[35] An EC enzyme probe has been also developed for the determination of methylmercury and ethylmercury.[134]

Organolead compounds appear in the environment due to their use as an additive in gasoline. These compounds are emitted from motor vehicles, and the emissions include tetraalkyllead compounds and ionic alkylleads. Ionic alkyllead compounds are determined in natural waters by differential pulse anodic stripping voltammetry by means of the reextraction of the compounds from

carbamate complexes to an acid aqueous solution.[36,135] Table 11 shows the applications of EC methods to the analysis of organometallics in environmental samples.

5.10 Miscellaneous Compounds

Ethylenethiourea is a environmental metabolite of some fungicides, high contents of which cause thyroid enlargement and cancer. The low volatility of ethylenethiourea needs derivatization for GC analysis, and the low UV absorbance maximum (232 nm) makes LC with UV detection susceptible to the interferences found in typical environmental samples. Pulsed amperometric detection (PAD) is used for the analysis of thiourea derivatives. The potential–time waveform is a +0.6 V/+1.2 V/−0.6 V/+0.6 V sequence.[136]

9,10-Anthraquinone is of great interest in connection with production of dyestuffs and pharmaceuticals. The determination is carried out by differential pulse voltammetry. The origin of the water sample dictates the procedures used for sample treatment.[37]

A procedure for the determination of total phthalate esters as phthalic acid in environmental samples is based on differential pulse voltammetry. The phthalate esters are used as plasticizers and, due to the increasing use of plastics, these esters have become widely distributed and are important organic pollutants. The method has been successfully applied to determine total phthalate esters in wastewater with a previous extraction procedure.[38] Applications of EC methods to the analysis of miscellaneous compounds in environmental samples described above are shown in Table 12.

6 COMPARISON WITH OTHER TECHNIQUES

Numerous analytical techniques can be employed for the determination of organic pollutants in environmental water and wastewater samples. Among them EC ones play an important role.

In general terms it can be said that the determination of the smallest amounts of substances is the domain of radiochemical methods, especially of activation analysis and tracer techniques, which can attain detection limits, in optimal cases, of the order of 10^{-21} g. Among the physicochemical analytical methods, MS, MS/GC or MS/LC, UV fluorescence, emission spectral analyses reach detection limits from the former amount to around 10^{-12} g.

In the range from about 10^{-6}–10^{-10} g, or in the concentration range 10^{-6}–10^{-10} M, a number of physicochemical methods can be used, starting with spectrophotometric, atomic-absorption spectroscopy and fluorescence, through kinetic and catalytic methods, to EC methods. With EC stripping detection limits up 10^{-12} M can be attained.

The most frequent comparison carried out in the environmental analysis of water samples and involving electroanalytical techniques, is between UV detection and the ED using chromatographic techniques. Thus, comparative studies between both types of detection for the determination by LC of priority phenols in river waters[141] and in drinking waters and river waters[91] have been carried out. In order to compare the performance of LC/UV and LC/EC, the EC detector was connected in series to the outpout of the UV detector in amperometric mode.

ED of phenolics requires applied potentials of around 1 V, which can cause other matrix components to be oxidized, thus increasing the background current. Apart from that, the use of high applied potentials can cause electropolymerization of phenolics at the electrode surface, with a decline in signal response with time. A high amount of humic substances in water samples can also cause a reduction in the signal response. Nevertheless, the electrode surface can generally be cleaned by applying a

Table 12 Environmental analysis of miscellaneous compounds by EC methods

Analyte	Sample	Detection limit	Comments about the method	Ref.
Miscellaneous compounds	groundwater	5 μg L^{-1}	PAD at a gold electrode	136
(Ethylenethiourea)	natural water	2 mg L^{-1}	DPP	37
(9,10-Anthraquinone)	industrial products			
(Motor oils)	polluted water		adsorptive stripping voltammetry at a HMDE	137
(Phthalate esters)	wastewater	5×10^{-7} M	DPP	38
(Peroxides)	river, groundwater		peroxidase-based biosensor; submersible bioprobe for continuous monitoring	138
	wastewater		polarographic determination	139
	drinking water		reversed-phase HPLC with amperometric detection	140

cycle of alternating negative and positive potentials, or by the use of the PAD.

However, not all the solvents used in LC can be used with an EC detector, as some can attack certain internal parts of the detector. Working with gradients in ED causes both the conductivity of the mobile phase and the potential of the solid state reference electrode to be slightly changed during elution of the analytes. This makes the real applied potential for every analyte dependent on their retention times. However, by recording the calibration graph always under the same experimental conditions the quantification is not affected.

By comparing the chromatographic profiles obtained from the two detectors when processing 50–100 mL of river water sample spiked with phenolics at the $1\,\mu g\,L^{-1}$ level, it can be seen that interferences make the identification and quantification of analytes when using UV detection difficult. This is especially true for the humic substances which can be found in the river water, which affect the beginning of the chromatogram, thus hampering the monitoring of low-retention-time analytes such as phenol. A cleaner profile was obtained when working with an EC detector, which allows identification and quantification of all the target compounds. Nevertheless for very high contents of humic substances the electrode response may be altered and it is found appropriate to diminish the sample volume.

In Table 13 the detection limits obtained in groundwater with both types of detectors are shown. It can be observed that the sensitivity of ED is around 100 times better than for UV detection, for all the phenols except the nitrophenols, because the working potential used (+1 V) was not optimum for determining nitrophenols. It can be concluded that the use of both type of electrodes in LC can be the most appropriate procedure for phenols determination at the required environmental levels.

A FI method using a dual EC detector for the analysis of total phenols in water and wastewater samples has been proposed as an alternative to the standard colorimetric procedure.[24] It is verified that whereas the 4-AAP colorimetric method for the determination of total phenols is unreliable when the water samples contain appreciable levels of certain substituted phenols, the FI/EC procedure significantly improves the accuracy of the determination of total phenols when there are substituted phenols.

The comparison of UV and ED in HPLC for the analysis of some organophosphorus pesticides[142] and some nitrogen-containing pesticides[143] has also been performed. In general, ED is more sensitive and, particularly for the nitrogen-containing compounds studied, the EC method was found to be 20–240 times more sensitive than the UV method. Furthermore, the reproducibility of the EC method is also better than for UV.

ED in HPLC is also more sensitive than UV detection for the analysis of residues of explosives in water samples.[13] Moreover, ED shows other advantages, as it allows the selective detection of nitrophenols in the presence of some nitramines and some nitrobenzoic acids, and it shows several chromatographic peaks undetectable with the UV detector (Figure 6).

Table 13 Detection limits ($\mu g\,L^{-1}$) in groundwater using on-line procedures with UV and ED. (Reproduced by permission of Elsevier Science from ref. 141)

Analyte	Sample volume 10 mL		Sample volume 100 mL	
	UV (280 nm)	EC (1 V)	UV (280 nm)	EC (1 V)
Phenol	10	0.02	4	4
4-Methylphenol	1.5	0.01	0.7	0.005
2,4-Dimethylphenol	0.8	0.03	0.05	<0.001
2-Nitrophenol	1.2	2	0.05	0.7
4-Nitrophenol	0.8[a]	3	0.1[a]	1
2,4-Nitrophenol	0.5	3	0.01	1.5
4-Chloro-3-methylphenol	2	0.01	0.1	0.001
2-Chlorophenol	1.5	0.05	0.4	0.001
3-Chlorophenol	1.7	0.05	0.4	0.001
4-Chlorophenol	1.5	0.05	0.5	0.001
2,4-Dichlorophenol	2	0.03	0.2	0.001
2,4,6-Trichlorophenol	2	0.03	0.1	0.001
2,3,5-Trichlorophenol	2	0.03	0.1	0.001
2,3,4-Trichlorophenol	2	0.03	0.1	0.001
3,4,5-Trichlorophenol	2	0.05	0.1	0.001
Pentachlorophenol	1[a]	0.03	0.01[a]	0.001

[a] 310 nm.

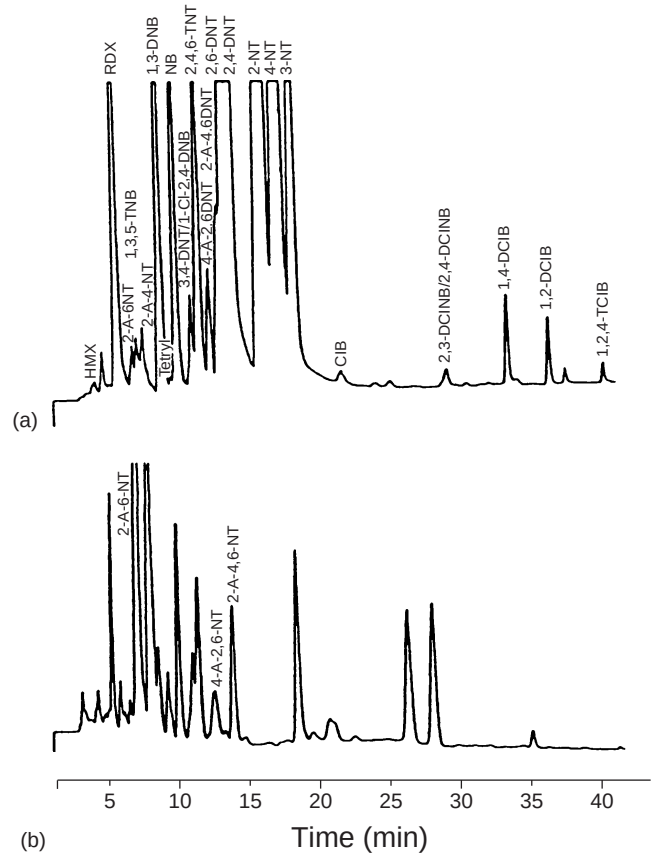

Figure 6 Chromatograms of the neutral basic/fraction of a groundwater from Elsnig: (a) UV detection (254 nm); (b) ED (+1.2 V). (Reproduced by permission of Elsevier Science from ref. 13.)

An off-column amperometric cell was also used for monitoring nitroaromatic explosives separated by electrokinetic capillary electrophoresis.[10] The amperometric detector measured several explosives in extracts prepared from contaminated soils and in groundwater samples successfully. Such a detection system offered a 10-fold detection limit improvement over UV measurement for most of the explosives tested.

ABBREVIATIONS AND ACRONYMS

CLC	Column Liquid Chromatography
CSV	Cathodic Stripping Voltammetry
DDT	Dichlorodiphenyltrichloroethane
DPP	Differential Pulse Polarography
DTPA	Diethylenetriaminepentaacetic Acid
EC	Electrochemical
ED	Electrochemical Detection
EDB	1,2-Dibromoethane
EDTA	Ethylenediaminetetraacetic Acid
EIA	Enzyme Immunoassay
ELISA	Enzyme-linked Immunosorbent Assay
ESB	Environmental Specimen Banking
FI	Flow Injection
FIA	Flow Injection Analysis
GC	Gas Chromatography
HMDE	Hanging Mercury Drop Electrode
HPLC	High-performance Liquid Chromatography
HRP	Horseradish Peroxidase
ISE	Ion Selective Electrode
ISFET	Ion-sensitive Field Effect Transistor
LAPS	Light-addressable Potentiometric Sensor
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
MFE	Mercury Film Electrode
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MS	Mass Spectrometry
PAD	Pulsed Amperometric Detection
PAHs	Polycyclic Aromaric Hydrocarbons
PAIT	Pulse-accelerated Immunoassay Technique
PBS	Phosphate Buffer Saline
PCB	Polychlorinated Biphenyl
PPY	Polypyrrole
PS-DVB	Poly(Styrene–Divinylbenzene)
RSD	Relative Standard Deviation
SMDE	Static Mercury Drop Electrode
SPE	Solid-phase Extraction
TCB	Trichlorobiphenyl
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
WE	Working Electrode
4-AAP	4-Aminoantipyrine

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Flow-injection Techniques in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Nitroaromatics, Environmental Analysis of • Organometallic Compound Analysis in Environmental Samples • Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis

in Environmental Samples • Solid-phase Microextraction in Environmental Analysis

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • Organophosphorus Pesticides in Water and Food, Analysis of

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Pulse Voltammetry

REFERENCES

1. A. Ivaska, A. Lewenstam, R. Sara (ed.), *Contemporary Electroanalytical Chemistry*, Plenum Press, New York, 1990.
2. R. Kalvoda (ed.), *Electroanalytical Methods in Chemical and Environmental Analysis*, Plenum Press, New York, 1987.
3. H. Emons, P. Ostapczuk, 'Electroanalysis for the Purpose of Environmental Monitoring and Specimen Banking: Is there a Future?', *Analyst*, **121**(12), 1917–1921 (1996).
4. W. Buchberger, 'Electrochemical Detectors – Tailormade Techniques for Liquid Chromatography and Capillary Electrophoresis', *Fresenius' J. Anal. Chem.*, **354**(7–8), 797–802 (1996).
5. A. Möeller, F. Scholz, 'Adventages and Limitations of Combining Separation Techniques with Voltammetry', *Fresenius' J. Anal. Chem.*, **356**(3–4), 160–168 (1996).
6. J. Svitel, S. Miertus, 'Development of Tyrosinase-based Biosensor and its Applications for Monitoring of Biomediation of Phenol and Phenolic Compounds', *Environ. Sci. Technol.*, **32**(6), 828–832 (1998).
7. J. Parellada, A. Narváez, M.A. López, E. Domínguez, J.J. Fernández, V. Paulov, I. Katakis, 'Amperometric immunosensors and enzyme electrodes for environmental applications', *Anal. Chim. Acta*, **362**(1), 47–57 (1998).
8. J. Wang, 'Remote Electrochemical Sensors for Monitoring Inorganic and Organic Pollutants', *Trends Anal. Chem.*, **16**(2), 84–88 (1997).
9. Ph. Quevauviller, W. Cofino, L. Cortez, 'Use of Reference Materials in Accreditation Systems for Environmental Laboratories', *Trends Anal. Chem.*, **17**(4), 241–248 (1998).
10. A. Hilmi, J.H.T. Luong, A.-L. Nguyen, 'Development of Electrokinetic Capillary Electrophoresis Equipped with Amperometric Detection for Analysis of Explosive Compounds', *Anal. Chem.*, **71**(4), 873–878 (1999).
11. K. Spiegel, T. Welsch, 'Monitoring Degradation Processes of Explosives by HPLC Analysis with UV and Amperometric Detection', *Fresenius' J. Anal. Chem.*, **357**(3), 333–337 (1997).
12. M.P. Maskarinec, D.L. Manning, R.W. Harvey, W.H. Griest, B.A. Tomkins, 'Determination of Munitions Components in Water by Resin Adsorption and High-performance Liquid Chromatography-electrochemical Detection', *J. Chromatogr.*, **302**, 51–63 (1984).
13. U. Lewin, J. Efer, W. Engewald, 'High-performance Liquid Chromatographic Analysis with Electrochemical Detection for Residues of Explosives in Water Samples Around a Former Ammunition Plant', *J. Chromatogr. A*, **730**(1+2), 161–167 (1996).
14. M.-C. Hennion, P. Subra, V. Coquart, R. Rosset, 'Determination of Polar Aniline Derivatives in Aqueous Environmental Samples Using On-line Liquid Chromatographic Preconcentration Techniques', *Fresenius' J. Anal. Chem.*, **339**(7), 488–493 (1991).
15. J.R. Rice, P.T. Kissinger, 'Liquid Chromatography with Precolumn Sample Preconcentration and Electrochemical Detection: Determination of Aromatic Amines in Environmental Samples', *Environ. Sci. Technol.*, **16**(5), 263–268 (1982).
16. J.R. Mazzeo, I.S. Krull, P.T. Kissinger, 'Use of Ce(IV) Oxidizing Agent for the Derivatization of Polycyclic Aromatic Hydrocarbons for Liquid Chromatography-electrochemical Detection', *J. Chromatogr.*, **550**(1+2), 585–594 (1991).
17. H.-P. Nirmaier, E. Fisher, A. Meyer, G. Henze, 'Determination of Polycyclic Aromatic Hydrocarbons in Water Samples Using High-performance Liquid Chromatography with Amperometric Detection', *J. Chromatogr. A*, **730**(1+2), 169–175 (1996).
18. N.K. Lam, M. Kopanica, 'Determination of Trichlorobiphenyl by Adsorptive Stripping Voltammetry', *Anal. Chim. Acta*, **161**, 315–324 (1984).
19. S. Bender, O.A. Sadik, 'Direct Electrochemical Immunosensor for Polychlorinated Biphenyls', *Environ. Sci. Technol.*, **32**(6), 788–797 (1998).
20. M.T. Galcerán, O. Jáuregui, 'Determination of Phenols in Seawater by Liquid Chromatography with Electrochemical Detection after Enrichment by Using Solid-phase Extraction Cartridges and Disks', *Anal. Chim. Acta*, **304**(1), 75–84 (1995).
21. N. Cardellicchio, S. Cavalli, V. Piangerelli, S. Giandomenico, P. Ragone, 'Determination of Phenols in Environmental Samples by Liquid Chromatography-electrochemistry', *Fresenius' J. Anal. Chem.*, **358**(6), 749–754 (1997).
22. O. Jáuregui, M.T. Galcerán, 'Determination of Phenols in Water by On-line Solid-phase Disk Extraction and Liquid Chromatography with Electrochemical Detection', *Anal. Chim. Acta*, **340**(1–3), 191–199 (1997).
23. D. Puig, D. Barceló, 'Determination of Polar Priority Phenols ar Parts per Trillion Levels in Water Using On-line Liquid-solid Extraction Followed by Liquid

Chromatography with Coulometric Detection', *J. Chromatogr. A*, **778**(1+2), 313–319 (1997).
24. M.J. Christophersen, T.J. Cardwell, 'Determination of Total Phenols in Waters and Wastewatyers Using Flow Injection with Electrochemical Detection; an Alternative to the Standard Colorimetric Procedure', *Anal. Chim. Acta*, **323**(1–3), 39–46 (1996).
25. F. Ortega, E. Domínguez, E. Burestedt, J. Emnéus, L. Gorton, G. Marko-Varga, 'Phenol Oxidase-based Biosensors as Selective Detection Units in Column Liquid Chromatography for the Determination of Phenolic Compounds', *J. Chromatogr. A*, **675**(1+2), 65–78 (1994).
26. R.E. Shoup, G.S. Mayer, 'Determination of Environmental Phenols by Liquid Chromatography/Electrochemistry', *Anal. Chem.*, **54**(7), 1164–1169 (1982).
27. A. Pachinger, E. Eisner, H. Begutter, H. Klus, 'A Simple Method for the Determination of Amitrole in Drinking and Groundwater', *Fresenius' J. Anal. Chem.*, **342**(4/5), 413–415 (1992).
28. G. Achilli, G.P. Cellerino, G.M. d'Eril, S. Bird, 'Simultaneous Determination of 27 Phenols and Herbicides in Water by High Performance Liquid Chromatography with Multi-electrode Electrochemical Detection', *J. Chromatogr. A*, **697**(1+2), 357–362 (1995).
29. S. Dupas, P. Scribe, J.F. Dubernet, 'On-line and Off-line Solid-liquid Extraction and Liquid Chromatographic Analysis at Trace Levels for Monitoring of Herbicides and Their Degradation Products in River and Fluvio-Estuarine Freshwater-Seawater Interfaces', *J. Chromatogr. A*, **737**(1), 117–126 (1996).
30. Z. Lukaszewski, 'Determination of Poly(ethylene glycol)s in Environmental Samples by the Indirect Tensammetric Method', *Analyst*, **121**(12), 1897–1901 (1996).
31. B. Wyrwas, A. Szymanski, Z. Lukaszewski, 'Determination of Nonionic Surfactants Adsorbed on Articles of Surface Water by an Indirect Tensammetric Method Combined with the BiAS Separation Scheme', *Anal. Chim. Acta*, **331**(1–2), 131–139 (1996).
32. I. Zjawiony, Z. Klima, 'Indirect Polarographic Determination of Trace Amounts of Nonionic Surfactants', *Chem. Anal.*, **36**(4), 741–747 (1991).
33. J.P. Hart, W.F. Smith, B.J. Birch, 'Polarographic Determination of Some Linear Alkylbenzene Sulphonates', *Analyst*, **104**, 853–859 (1979).
34. M. Ochsenkuhn-Petropulu, K.-M. Ochsenkuhn, G. Parissakis, A.N. Giannakis, H.D. Smith, 'Investigation of the Distribution of Tributyltin in the Seawater from the Port of Piraeus, Greece', *Can. J. Appl. Spectrosc.*, **40**(3), 66–71 (1995).
35. M.P. da Silva, J.R. Procopio, L. Hernández, 'Reversed-phase High-performance Liquid Chromatography of Pyrrolidinedithiocarbamate Complexes of Mercuric Species Using Amperometric and Coulometric Detection', *J. Chromatogr. A*, **761**(1+2), 139–146 (1997).
36. N. Mikac, M. Branica, Y. Wang, R.M. Harrison, 'Organolead Compounds in Mussels (*Mytilus galloprovincialis*) from the Eastern Adriatic Coast', *Environ. Sci. Technol.*, **30**(2), 499–508 (1996).
37. J.O. Bronstad, K.H. Schroder, H.O. Friestad, 'Determinations of Trace Amounts of 9,10-Anthraquinone in Aqueous Systems by Differential Pulse Polarography', *Anal. Chim. Acta*, **119**, 243–249 (1980).
38. K. Tanaka, M. Takeshita, 'Determination of Total Phthalate Esters in Wastewaters by Differential Pulse Polarography', *Anal. Chim. Acta*, **166**, 153–161 (1984).
39. A.M. Dietrich, J.N. Jensen, W.F. da Costa, 'Measurement of Pollutants. Chemical Species', *Water Environ. Res.*, **68**(4), 391–406 (1996).
40. M.-C. Hennion, V. Pichon, D. Barceló, 'Surface Water Analysis (Trace-organic Contaminants) and EC Regulations', *Trends Anal. Chem.*, **13**(9), 361–372 (1994).
41. A.N. Reshetilov, 'Biosensors Models Based on Potentiometric and Amperometric Transducers for Medicine, Biotechnology and Environmental Monitoring', *Prikl. Biokhim. Mikrobiol.*, **32**(1), 78–93 (1996).
42. G.A. Evtugyn, E.P. Rizaeva, E.E. Stoikova, V.Z. Latipova, H.C. Budnikov, 'The Application of Cholinesterase Potentiometric Biosensor for Preliminary Screening of the Toxicity of Wastewaters', *Electroanalysis*, **9**(14), 1124–1128 (1997).
43. E.I. Iwuoha, M.R. Smith, 'Organic-phase and Inhibition-base Amperometric Biosensors for the Determination of Environmental Pollutants', *NATO ASI Ser. 2*, **38**, 289–304 (1997).
44. P. Skladal, 'Biosensors Based on Cholinesterase for Detection of Pesticides', *Food Technol. Biotechnol.*, **34**(1), 43–49 (1996).
45. M. Trojanowicz, M.L. Hitchman, 'Determination of pesticides Using electrochemical biosensors', *Trends Anal. Chem.*, **15**(1), 38–45 (1996).
46. J.L. Marty, D. Garcia, R. Rouillon, 'Biosensors: Potential in pesticide detection', *Trends Anal. Chem.*, **14**(7), 329–333 (1995).
47. O.A. Sadik, J.M. Van Emon, 'A Status Report on Electroanalytical Techniques for Immunological Detection', *ACS Symp. Ser.*, **646**, 127–147 (1996).
48. J.P. Hart, S.A. Wring, 'Recent Developments in the Design and Applications of Screen-printed Electrochemical Sensors for Biomedical, Environmental and Industrial Analysis', *Trends Anal. Chem.*, **16**(2), 89–103 (1997).
49. D. Puig, D. Barceló, 'Determination of Phenolic Compounds in Water and Wastewater', *Trends Anal. Chem.*, **15**(8), 362–375 (1996).
50. H. Greschonig, 'Electrochemical Behavior and Electroanalytical Methods for the Determination of Arsenic Compounds', *Sci. Pap. Univ. Pardubice, Ser. A, Volume date 1997*, **3**, 293–305 (1998).

51. K.M. Attar, 'Analytical Methods for Speciation of Organotins in the Environment', *Appl. Organomet. Chem.*, **10**(5), 317–337 (1996).
52. L. Campanella, C. Colapicchioni, L. Aiello, M. Tomassetti, 'Solid State Sensors (ISFETs) for Environmental Analysis', *Sens. Microsyst.*, **1**, 157–162 (1996).
53. M. Gerlache, J.M. Kauffmann, G. Quarin, J.C. Vire, G.A. Bryant, J.M. Talbot, 'Electrochemical Analysis of Surfactants: An Overview', *Talanta*, **43**(4), 507–519 (1996).
54. M. Sillanpaa, M.-L. Sihvonen, 'Analysis of EDTA and DTPA', *Talanta*, **44**(8), 1487–1497 (1997).
55. US Environmental Protection Agency, *EPA method 605*, Part VIII, 40 CFR, Part 136, US EPA, Washington, DC, 1978.
56. J.B.F. Lloyd, 'High Performance Liquid Chromatography of Organic Explosives with Electrochemical Detection at a Pendant Mercury Drop Electrode', *J. Chromatogr.*, **257**, 227–236 (1983).
57. J. Wang, R.K. Bhada, J. Lu, D. MacDonald, 'Remote Electrochemical Sensor for Monitoring TNT in Natural Waters', *Anal. Chim. Acta*, **361**(1–2), 85–91 (1998).
58. L. Wennrich, U. Lewin, J. Efer, W. Engelwald, 'Analysis of Diamino- and Aminonitroaromatics in Water Using Solid Phase Extraction and HPLC', *Vom Wasser*, **86**, 341–352 (1996).
59. V. Concialini, G. Chiavari, P. Vitali, 'Electrochemical Detection in High-performance Liquid Chromatography Analysis of Aromatic Amines', *J. Chromatogr.*, **258**, 244–251 (1983).
60. J. Wang, G. Rivas, D. Luo, X. Cai, F.S. Valera, N. Dontha, 'DNA Biosensor for the Detection of Hydrazines', *Anal. Chem.*, **68**(13), 2215–2254 (1996).
61. I. Casella, M.R. Guascito, A.M. Salvi, E. Desimini, 'Catalytic Oxidation and Flow Detection of Hydrazine Compounds at a Nafion/Ruthenium(III) Chemically Modified Electrode', *Anal. Chim. Acta*, **354**(1–3), 333–341 (1997).
62. H. Xia, H.L. Li, D.Q. Yang, 'Voltammetric determination of Hydrazine based on Catalytic Reaction in the Presence of 4-Hydroxy-2,2,6,6,-Tetramethylpiperidinyloxy (TEMPOL) Radical', *Electroanalysis*, **9**(18), 1429–1431 (1997).
63. J. Wang, M. Chicharro, G. Rivas, D. Luo, X. Cai, N. Dontha, P.A.M. Farias, H. Shiraishi, 'DNA-modified Electrode for the Detection of Aromatic Amines', *Anal. Chem.*, **68**(24), 4365–4369 (1996).
64. J. Wang, Q. Chen, G. Cepria, 'Electrocatalytic Modified Electrode for Remote Monitoring of Hydrazines', *Talanta*, **43**(8), 1387–1391 (1996).
65. T. You, M. Wu, E. Wang, 'Determination of Hydroxylamine by Capillary Electrophoresis-electrochemical Detection with a Palladium Particle Modified Carbon Fiber Microdisk Array Electrode', *Anal. Lett.*, **30**(5), 1025–1036 (1997).
66. F. Guerrieri, G. Bucci, 'Differential Pulse Polarographic Determination of Nitrilotriacetic Acid in Natural Waters and Wastewater Effluents', *Anal. Chim. Acta*, **167**, 393–397 (1985).
67. J. Jou, D. Tian, X. Shi, 'Determination of the Content of NTA in Natural Water and Wastewater by Single Sweep Polarography', *Nanjing Huagong Daxue Xuebao*, **20**(1), 76–79 (1998). (CA 129 : 206892).
68. M. Erickson, *Analytical Chemistry of PCBs*, CRC Press, Boca Raton, FL, 1992.
69. W. Franklin Smith, '*Voltammetric Determination of Molecules of Biological Significance*', John Wiley and Sons, Chichester, 1992.
70. A. Velasco-Arjona, M.D. Luque de Castro, A. Izquierdo, 'Fully Automated Robotic Method for the Screening of Polychlorinated Byphenyls in used Mineral Oils', *Analyst*, **123**(3), 509–512 (1998).
71. M. Del Carlo, M. Mascini, 'Enzyme Immunoassay with Amperometric Flow-injection Analysis Using Horseradish Peroxidase as a Label. Application to the Determination of Polychlorinated Biphenyls', *Anal. Chim. Acta*, **336**(1–3), 167–174 (1996).
72. M. Del Carlo, I. Lionti, M. Taccini, A. Cagnini, M. Mascini, 'Disposable Screen-printed Electrodes for the Immunochemical Detection of Polychlorinated Biphenyls', *Anal. Chim. Acta*, **342**(2–3), 189–197 (1997).
73. W. Frenzel, S. Kekler, 'Spectrophotometric Determination of Total Phenolics by Solvente Extraction Optosensing Using Flow Injection Methodology', *Anal. Chim. Acta*, **310**(3), 437 (1995).
74. J. Wang, Y. Lin, L. Chen, 'Organic-phase Biosensors for Monitoring Phenol and Hydrogen Peroxide in Pharmaceutical Antibacterial Products', *Analyst*, **118**(3), 277–280 (1993).
75. J.W. Moore, S. Ramamoorthy, *Organic Chemical in Natural Waters: Applied Monitoring and Impact Assessment*, Springer Verlag, New York, 1984.
76. B. Schultz, 'Determination of Priority Pollutant Nitrophenols in Water by High-performance Liquid Chromatography', *J. Chromatogr.*, **269**, 208–212 (1983).
77. A. Buikema, M. McGinnes, J. Cairns, 'Phenols in Aquatic Exosystems: A Selected Review of Recent Literature', *Marine Environ. Res. 2*, Applied Science Publishers, London, 1979.
78. *Drinking Water Directive 80/778/EEC*, Commission of the European Communities, Brussels, 1980.
79. *Directive 76/160/EEC*, Commission of the European Communities, Brussels, 1975.
80. United States Environmental Protection Agency, *Toxic Substances Control Act (TSCA)*, USEPA, Washington, DC, 1979.
81. US Public Health Association, *Standard Methods for the Examination of Water and Wastewater*, 16th edition, US PHA, Washington DC, 1985.
82. W. Fresenius, K.E. Quentin, W. Schneider (eds.), *Water Analysis*, Springer, Berlin, 1988.

83. US Environmental Protection Agency, *EPA Method 604-Phenols*, 40 CFR Part 136, 43290; Federal Register 49, No. 209, US EPA, Washington, DC, 1984.
84. US Environmental Protection Agency, *EPA Method 625-Base/neutrals and acids*, 40 CFR Part 136, 43385; Federal Register 49, No. 209, US EPA, Washington, DC, 1984.
85. I. Naranjo Rodríguez, M. Barea Zamora, J.M. Barberá Salvador, J.A. Muñoz Leyva, M.P. Hernández-Artiga, J.L. Hidalgo Hidalgo de Cisneros, 'Voltammetric Determination of 2-Nitrophenol at a Bentonite-modified Carbon Paste Electrode', *Mikrochim. Acta*, **126**(1–2), 87–92 (1997).
86. M.M. Cordero-Rando, I. Naranjo-Rodríguez, J.L. Hidalgo-Hidalgo de Cisneros, 'Voltammetric Study of 2-Methyl-4,6-dinitrophenol at a Modified Carbon Paste Electrode', *Anal. Chim. Acta*, **370**(2–3), 231–238 (1998).
87. I. Naranjo-Rodríguez, J.A. Muñoz-Leyva, J.L. Hidalgo-Hidalgo de Cisneros, 'Use of a Bentonite-modified Carbon Paste Electrode for the Determination of Some Phenols in a Flow System by Differential-pulse Voltammetry', *Analyst*, **122**(6), 601–604 (1997).
88. I. Naranjo-Rodríguez, J.A. Muñoz-Leyva, J.L. Hidalgo-Hidalgo de Cisneros, 'Use of a Carbon Paste Modified Electrode for the Determination of 2-Nitrophenol in a Flow System by Differential Pulse Voltammetry', *Anal. Chim. Acta*, **344**(3), 167–173 (1997).
89. B. Paterson, C.E. Cowie, P.E. Jackson, 'Determination of Phenols in Environmental Waters Using Liquid Chromatography with Electrochemical Detection', *J. Chromatogr. A*, **731**(1+2), 95–102 (1996).
90. A. Hagen, J. Mattusch, G. Werner, 'Flow-rate Variated HPLC-/EC-determination of Phenols', *Fresenius' J. Anal. Chem.*, **339**(1), 26–29 (1991).
91. J. Ruana, I. Urbe, F. Borrull, 'Determination of Phenols at the ng/l Level in Drinking and River Waters by Liquid Chromatography with UV and Electrochemical Detection', *J. Chromatogr. A*, **655**(2), 217–226 (1993).
92. G. Marko-Varga, J. Emnéus, L. Gorton, T. Ruzgas, 'Development of Enzyme-based Amperometric Sensors for the Determination of Phenolic Compounds', *Trends Anal. Chem.*, **14**(7), 319–328 (1995).
93. T. Ruzgas, J. Emnéus, L. Gorton, G. Marko-Varga, 'The Development of a Peroxidase Biosensor for Monitoring Phenol and Related Aromatic Compounds', *Anal. Chim. Acta*, **311**(3), 245–253 (1995).
94. J. Wang, Q. Chen, 'Remote Electrochemical Biosensor for Field Monitoring of Phenolic Compounds', *Anal. Chim. Acta*, **312**(1), 39–44 (1995).
95. H. Wang, A. Zhang, H. Cui, D. Liu, R. Liu, 'Adsorptive Stripping Voltammetric Determination of Phenol at an Electrochemically Preteated Carbon-paste Electrode with Solid Paraffin as a Binder', *Microchem. J.*, **59**(3), 448–456 (1998).
96. N. Masque, E. Pocurull, R.M. Marce, F. Borrull, 'Determination of Eleven Priority EPA Phenolics at ng L^{-1} levels by On-line Solid-phase Extraction and Liquid Chromatography with UV and Electrochemical Detection', *Chromatographia*, **47**(3+4), 176–182 (1998).
97. G. Carrieri, S. Cavalli, P.P. Ragone, N. Cardellicchio, 'Characterization of Phenolic Compounds in Landfill Leachate of Urban Solid Waste By HPLC with an Electrochemical Detector', *Ann. Chim.*, **88**(9–10), 619–627 (1998).
98. Y. Fang, F. Gong, X. Liu, X. Fang, J. Ye, 'Determination of Priority Phenols in Wastewater by Capillary Electrophoresis with Electrochemical Detection', *Fenxi Huaxue*, **26**(6), 694–697 (1998). (CA 129:85637).
99. A.W.O. Lima, V.B. Nascimento, J.J. Pedrotti, L. Angnes, 'Coconut-based Plant Tissue Reactor for Biosensing of Cathecol in Flow Injection Analysis', *Anal. Chim. Acta*, **354**(1–3), 325–331 (1997).
100. J. Wang, F. Lu, S.A. Kane, Y.K. Choi, M.R. Smith, K. Rogers, 'Hydrocarbon Pasting Liquids for Improved Tyrosinase-based Carbon-paste Phenol Biosensors', *Electroanalysis*, **9**(14), 1102–1106 (1997).
101. I. da Cruz Vieira, O. Fatibello-Filho, 'Amperometric Biosensor for the Determination of Phenols Using a Crude Extract of Sweet Potato', *Anal. Lett.*, **30**(5), 895–907 (1997).
102. M. Knutsson, L. Mathiasson, J.A. Joensson, 'Supported Liquid Membrane Work-up in Combination with Liquid Chromatography and Electrochemical Detection for the Determination of Chlorinated Phenols in Natural Water Samples', *Chromatographia*, **42**(3/4), 165–170 (1996).
103. World Health Organization, 'Public Health Impact of Pesticides Used in Agriculture', WHO, Geneva, 1990.
104. J. Sherma (ed.), *Analytical Methods for Pesticides and Plant Growth Regulators*, Academic Press, San Diego, CA, 1989.
105. R. Tokoro, R. Blewicz, J. Osteryoung, 'Polarographic Reduction and Determination of 1,2-Dibromoethane in Aqueous Solutions', *Anal. Chem.*, **58**(9), 1964–1969 (1986).
106. P. Hernández, E. Lorenzo, L. Hernández, 'Determination of Azinphos-Methyl by Adsorptive Stripping Voltammetry', *Anal. Chim. Acta*, **238**(2), 383–387 (1990).
107. E. Alvarez, M.-T. Sevilla, J.M. Pinilla, L. Hernández, 'Cathodic Stripping Voltammetry of Paraquat on a Carbon Paste Electrode Modified with Amberlite® XAD-2 Resin', *Anal. Chim. Acta*, **260**(1), 19–23 (1992).
108. C. Faller, A. Meyer, G. Henze, 'Voltammetric Determination of Ioxynil and 2-Methyl-3-Nitroaniline Using C_{18} Modified Carbon Paste Electrodes', *Fresenius' J. Anal. Chem.*, **356**(3–4), 279–283 (1996).
109. A. Arranz, M.F. Villalba, S. Fdez de Betoño, J.M. Moreda, J.F. Arranz, 'Anodic Voltammetric Assay of the Herbicide Metamitron on a Carbon Paste Electrode', *Fresenius' J. Anal. Chem.*, **357**(6), 768–772 (1997).
110. J.L. Anderson, K.K. Whiten, J.D. Brewster, T.S. Ou, W.K. Nonidez, 'Microarray Electrochemical Flow

Detector at High Applied Potentials and Liquid Chromatography with Electrochemical Detection of Carbamate Pesticides in River Water', *Anal. Chem.*, **57**(7), 1366–1373 (1985).
111. I. Bernwieser, G. Sontag, 'HPLC with Electrode Array Detection: Determination of Amitrole in Drinking and Groundwater', *Fresenius' J. Anal. Chem.*, **347**(12), 499–501 (1993).
112. B. Saad, M.M. Ariffin, M.I. Saleh, 'Paraquat Sensors Containing Membrane Components of High Lipophilicities', *Anal. Chim. Acta*, **338**(1–2), 89–96 (1997).
113. D. Nikolelis, C. Siontorou, 'Flow Injection Monitoring and Analysis of Mixtures of Simazine, Atrazine, and Propazine Using Filter-supported Bilayer Lipid Membranes', *Electroanalysis*, **8**(10), 907–912 (1996).
114. K. Stein, G. Schwedt, 'Comparision of Immobilization Methods for the Development of an Acetylcholinesterase Biosensor', *Anal. Chim. Acta*, **272**(1), 73–81 (1993).
115. C. Ristori, C. Del Carlo, M. Martini, A. Barbaro, A. Ancarani, 'Potentiometric Detection of Pesticides in Water Samples', *Anal. Chim. Acta*, **325**(3), 151–160 (1996).
116. A.M. Nyansi Hendji, N. Jafrezic-Renault, C. Martelet, P. Clechet, 'Sensitive Detection of Pesticides Using a Differential ISFET-based System with Immobilized Cholinesterases', *Anal. Chim. Acta*, **281**(1), 3–11 (1993).
117. G. Palleschi, M. Bernabei, C. Cremisini, M. Mascini, 'Determination of Organophosphorus Insecticides with a Choline Electrochemical Biosensor', *Sens. Actuators, B*, **7**(1–3), 513–517 (1992).
118. A.J. Baumner, R.D. Schmid, 'Development of a New Immunosensor for Pesticide Detection: A Disposable System with Liposome-enhancement and Amperometric Detection', *Biosens. Bioelectron.*, **13**(5), 519–529 (1998).
119. J.J. Rippeth, T.D. Gibson, J.P. Hart, I.C. Hartley, G. Nelson, 'Flow-injection Detector Incorporating a Screen-printed Disposable Amperometric Biosensor for Monitoring Organophosphate Pesticides', *Analyst*, **122**(11), 1425–1429 (1997).
120. R. Wilson, M.H. Barker, D.J. Schiffrin, R. Abuknesha, 'Electrochemiluminescence Flow Injection Immunoassay for Atrazine', *Biosens. Bioelectron.*, **12**(4), 277–286 (1997).
121. C.M.P. Vaz, S. Crestana, S.A.S. Machado, L.H. Mazo, L.A. Avaca, 'Electroanalytical Determination of the Herbicide Atrazine in Natural Waters', *Int. J. Environ. Anal. Chem.*, **62**(2), 65–76 (1996).
122. M.A. Mateos, R. Galvez, F.J. Manuel de Villena, J.M. Pingarron, 'Determination of Propazine by Differential Pulse Polarography in Micellar and Emulsified Media', *Mikrochim. Acta*, **120**(1–4), 339–349 (1995).
123. A. Arranz, S. Fdez De Betono, J.M. Moreda, J.F. Arranz, 'Study of the Electro-adsorptive Behavior of the Herbicide Nitralin by means of Voltammetric Techniques', *Talanta*, **45**(2), 417–424 (1997).
124. M.S. El-Shahawi, M.M. Kamal, 'Determination of the Pesticide Chlorpyrifos by Cathodic Adsorptive Stripping Voltammetry', *Fresenius' J. Anal. Chem.*, **362**(3), 344–347 (1998).
125. A.S.R. Al-Meqbali, M.S. El-Shahawi, M.M. Kamal, 'Determination of the Pesticide Chlorpyrifos by Cathodic Adsorptive Stripping Voltammetry', *Electoanalysis*, **10**(11), 784–786 (1998).
126. T. Kalab, P. Skladal, 'Disposable Multichannel Immunosensors for 2,4-Dichlorophenoxyacetic Acid Using Acetylcholinesterase as an Enzyme Label', *Electroanalysis*, **9**(4), 293–297 (1997).
127. D. Trau, T. Theuerl, M. Wilmer, M. Meusel, F. Spener, 'Development of an Amperometric Flow Injection Immunoanalysis System for the Determination of the Herbicide 2,4-Dichlorophenoxyacetic Acid in Water', *Biosens. Bioelectron.*, **12**(6), 499–519 (1997).
128. M.C. Sampedro, Z. Gomez de Balugera, A. Coicolea, R.J. Barrio, 'Determination of the Insecticide Pyridafenthion in River Water, Soils and Wine by Adsorptive Stripping Voltammetry', *Food Addit. Contam.*, **15**(7), 793–800 (1998).
129. Z. Lukaszewski, 'Determination of Surfactants in Surface Water. Critical Review', *Ann. Univ. Mariae Curie-Sklodowska, Sect. AA: Chem., Volume Date 1995*, **50–51**, 1–22 (1996).
130. E. González-Mazo, A. Gómez-Parra, 'Monotoring Anionic Surfactants (LAS) and their Intermediate Degradation Products in the Marine Environmen', *Trends Anal. Chem.*, **15**(8), 375–380 (1996).
131. R. Schulz, 'Potentiometric Titration of Nonionic Surfactants with Surfactant-sensitive Electrodes', *SOFW J.*, **122**(14), 1022–1028 (1996).
132. W. Haman, P. Kikolski, E. Dluska-Smolik, 'Analytical Methods Used for the Determination of Primary and Ultimate Biodegradation of Anionic and Nonionic Surfactants', *Pollena: Tluszcze, Sroddki Piorace, Kosmet.*, **40**(9), 369–374 (1996).
133. Y. Hoshino, 'Detection of Organotin Compound at the Trace Level By Voltammetric Method with Binuclear Ruthenium Complex Coated Electrode. Quantitative Analysis of Tributyltin Compounds', *Nagasaki Daigaku Kyoikugakubu Shizen Kagaku Kenkyu Hokoku*, **58**, 15–22 (1998). (CA 128:330511).
134. A. Amine, C. Cremisini, G. Palleschi, 'Determination of Mercury(II), Methylmercury and Ethylmercury in the ng/ml Range with an Electrochemical Enzyme Glucose Probe', *Mikrochim. Acta*, **121**(1–4), 183–190 (1995).
135. N. Mikac, M. Branica, 'Determination Of Ionic Alkyllead Compounds in Natural Waters by Differential Pulse Anodic Stripping Voltammetry', *Anal. Chim. Acta*, **264**(2), 249–258 (1992).
136. D.R. Doerge, A.B.K. Yee, 'Liquid Chromatographic Determination of Ethylenethiourea Using Pulsed

Amperometric Detection', *J. Chromatogr.*, **586**(1), 158–160 (1991).
137. M. Kovac, R. Kalvoda, L. Novotny, A. Berka, 'Adsorptive Stripping Tensammetry of Motor Oils in Aqueous Solutions', *Electroanalysis*, **5**(2), 171–176 (1993).
138. J. Wang, G. Cepria, Q. Chen, 'Submersible Bioprobe for Continous Monitoring of Peroxide Species', *Electroanalysis*, **8**(2), 124–127 (1996).
139. V.N. Klochkova, T.N. Praeobrazhenskaya, A.A. Gaifullin, 'Polarographic Determination of Ethylbenzene Hydroperoxide in Aqueous Solution', *J. Anal. Chem.*, **53**(1), 44–47 (1998).
140. O. Evans, 'High-performance Liquid Chromatography of Selected Organic Peroxides with Oxidative Amperometric Detection', *Analyst*, **123**(4), 705–710 (1998).
141. D. Puig, D. Barceló, 'Comparative Study of On-line Solid Phase Extraction Followed by UV and Electrochemical Detection in Liquid Chromatography for the Determination of Priority phenols in River Water Samples', *Anal. Chim. Acta*, **311**(1), 63–69 (1995).
142. G.J. Clark, R.R. Goodin, J.W. Smiley, 'Comparasion of UV and Reductive Amperometric Detection of Ethyl and Methyl Parathion in Green Vegetables and Surface Water Using High Performance Liquid Chromatography', *Anal. Chem.*, **57**(12), 2223–2228 (1985).
143. S. Yao, A. Meyer, G. Henze, 'Comparasion of amperometric and UV-spectrophotometric Monitoring in the HPLC Analysis of Pesticides', *Fresenius' J. Anal. Chem.*, **339**(4), 207–211 (1991).

Organometallic Compound Analysis in Environmental Samples

Les Ebdon and Andy Fisher
University of Plymouth, Devon, UK

1 Introduction **1**
2 Sample Collection and Storage **2**
3 Sample Preparation **3**
3.1 Aim of Sample Preparation 3
3.2 Extraction of Species from Liquid Samples 4
3.3 Extraction of Species from Solid Samples 4
3.4 Derivatization 5
4 Sample Analysis **7**
4.1 Coupled Techniques 7
4.2 Other Methods of Speciation 13
5 Method Validation **15**
5.1 Mass Balance Data and Spiking/Recovery Experiments 15
5.2 Certified Reference Materials 16
5.3 Other Methods of Validation 17
6 Conclusions **17**
Abbreviations and Acronyms **18**
Related Articles **18**
References **19**

Many organometallic species exist in the environment and these may range from being completely nontoxic, e.g. the arsenic compound arsenobetaine, to exceptionally toxic, e.g. methyl mercury. Compounds of intermediate toxicity, such as the arsenic species dimethylarsinic acid and monomethylarsonic acid also exist. Determining the chemical form or speciation of samples is vital in environmental impact assessment as it affects bioavailability and mobility in the environment. Speciation analysis usually comprises a number of analytical steps. Sampling and sample storage are critical and must be performed in such a way that ensures a representative sample whilst preventing changes in speciation or losses of some of the more labile species. Sample preparation normally requires the extraction of the species of interest from the sample materials. This may be achieved in a number of ways that are discussed in detail in the relevant section. Sample cleanup and derivatization methods that may be required if gas chromatographic methods of analysis are to be used are also discussed. The actual sample analysis stage is discussed in detail and this concentrates mainly upon different chromatographic techniques coupled with assorted atomic spectrometric methods of detection. This is because these so-called ''coupled'' techniques are used for the majority of speciation analyses. Other nonchromatographic methods of sample introduction to atomic spectrometric detection, such as hydride generation (HG), are also discussed together with techniques including more conventional detectors for the chromatographic techniques (such as ultraviolet (UV)/visible, fluorescence, flame ionization detectors (FIDs), electrochemical detectors, etc.), electrochemical methods, nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy. In addition to obtaining quantitative data, techniques such as NMR and IR spectroscopy may be used to obtain structural information on individual species.

Method validation is as important for speciation analysis as it is for any other analysis. This may be achieved using mass balance data, the analysis of certified reference materials (CRMs) or using completely independent analysis techniques. The text gives an overview of all stages of the analysis, highlighting potential pitfalls and areas of uncertainty.

1 INTRODUCTION

Since the late 1980s, there has been increasing interest in determining not only ''total'' concentrations of analytes, but also the concentrations of individual forms of the element. The determination of the different forms of elements is known as speciation analysis. The interest in this area stems from increased knowledge of the chemistry of the environment and the way in which anthropogenic inputs can affect it. Several different elements can form a number of species. These forms can be of different oxidation states but many have an organic moiety associated with them. These organometallic compounds usually have both different biogeochemical cycles and different toxicity when compared with the inorganic forms and hence they are of interest to food scientists, environmental scientists and ecologists. In general, it can be said for some elements, e.g. lead and mercury, that the organometallic forms of elements are more toxic than the inorganic forms. Obvious exceptions to this general rule are arsenic and selenium. Inorganic arsenic is a well-known poison, but many of the organic forms, e.g. arsenobetaine and arsenocholine, are generally regarded as being nontoxic. This is because these molecules are extremely stable and resist enzymatic or oxidative attack even by reagents such as concentrated nitric acid. The

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

potentially dangerous arsenic is therefore effectively immobilized. Therefore if a substantial amount of seafood is consumed, the arsenobetaine that will be present, often at high concentration, will not be attacked by the enzymes or stomach acid of the body and will be excreted in the normal way without any adverse effects.

Organic forms of many elements are regarded as being more toxic than inorganic forms. This is because the organic forms are more readily absorbed through the gut lining and because they pass through the blood–brain barrier. The latter is especially true for methyl mercury. This compound attacks the blood–brain barrier and disrupts the metabolism of the nervous system. In addition, it can cross the placenta more easily and it is this property that led to the Minamata poisoning case in Japan. In this incident a factory released mercury into the sea; this mercury underwent biomethylation, thereby contaminating local fish stocks. Ingestion of the fish by expectant women led to 22 infants being born with serious brain damage whilst the mothers showed only marginal symptoms of mercury poisoning.

There are a number of natural phenomena that effect the speciation of elements. Many organic compounds are destroyed by the action of light (photolysis), e.g. organotin compounds and arsenobetaine. There are a number of bacteria capable of biomethylation and there are other bacteria that demethylate metals. Elements whose speciation is affected by these bacteria include mercury, tin and arsenic. It is also known that some moulds are capable of methylating inorganic arsenic. Methylation is also hypothesized as the method by which vertebrates detoxify inorganic arsenic.

Anthropogenic inputs to the environment such as tributyl tin (TBT) from antifouling paints or triphenyl tin from fungicides are known to cause deleterious effects to shellfish (e.g. imposex and shell thickening). In the environment, these species will slowly be detoxified by stepwise dealkylation, forming dibutyl tin, followed by monobutyl tin and finally inorganic tin. The same process also occurs for the phenyl tins, which are dearylated in a stepwise manner. Once in the environment, these tin species may also be methylated by bacteria. Hence, a plethora of different species may be formed.

From this brief overview it is obvious that a very large number of different species for many elements exist in the environment. To determine the overall toxicity of a sample, or to gauge the extent to which a breakdown process has occurred, it is necessary to perform speciation analysis. There have been a very large number of methods employed, of which the relative advantages, disadvantages and pitfalls of each are discussed in later sections. A recent publication has summarized these, but has not gone into specific problems.[1] Other books that have summarized speciation include one by Quevauviller.[2] This book has several chapters on the speciation of individual elements and also a chapter devoted to speciation analysis of environmental samples. In addition, possible sources of error and the preparation of CRMs to validate analytical methods are also discussed. The latter subject was the main focus of a previous book.[3] An overview of analytical methods used for elemental speciation has also been made.[4] Numerous other reviews have concentrated on the speciation analysis of certain elements. Examples include reviews on arsenic,[5] lead,[6] tin,[7] mercury[8] and selenium.[9,10]

2 SAMPLE COLLECTION AND STORAGE

As for normal analytical chemistry, the sampling and storage strategies are of utmost importance. If performed incorrectly, all subsequent steps in the analysis are pointless. A book produced by the Royal Society of Chemistry gives invaluable information on sample collection and storage.[11] It is important that the correct sample is taken. For instance, if a sediment is to be analyzed, it should be taken whilst ensuring that weed, slimes, twigs and leaves are not sampled simultaneously. Care should be taken to store an anoxic sediment under nitrogen and to avoid undue temperature increases. Another good overview on sampling strategies and sample storage has also been produced.[12]

Different sampling methodologies are required for different sample types. Sample homogeneity is easily ensured for gaseous and liquid samples, but solids may cause problems. Contamination may be a problem for all three sample types, but can be especially bad for liquid or gaseous samples. For water samples, the type of collection vessel used will depend on the depth from which the water is to be sampled. Clean plastic bottles may be used to collect samples from shallow water or close to the surface. If sampling from deeper water is required, GoFLO™ samplers or other specialized sample collection devices may be necessary. The material from which the container used to collect the sample is made will also depend on the analyte. Mercury vapor is known to pass through low-density plastics and therefore collection and storage in a plastic bottle should be avoided. Similarly, some organotin compounds are used as plasticizers and therefore sample collection and storage in plastic containers may lead to contamination. Under circumstances such as these, it may be better to collect and store samples in glass bottles. It has become popular to collect the sample by isolating the species of interest on a microcolumn of adsorption media. Once the sample has been collected, the microcolumns can be returned to the laboratory and inserted into a flow injection

(FI) system for elution/quantitative determination. This approach has been used to collect mercury and lead species in seawater that is pumped through a microcolumn containing dithiocarbamate groups. It was found however that the storage stability of the samples was poor, indicating that elution and analysis should be performed as soon after collection as possible. Methods such as these limit the possibility of analyte loss or contamination but, as was found in this example, care must be taken to ensure stability of the species and hence obtain accurate results. Care must also be taken to ensure that all the species of interest are collected quantitatively, without changes in oxidation state or ligands, on the adsorption media. If one species is collected in preference to another, then the overall speciation picture may be distorted.

Sampling gaseous matrices can also be performed in several different ways. Commercially available bags or evacuated containers may be used. Alternatively, some workers have isolated the species of interest from the bulk matrix. Examples include the use of several solid adsorbents for sampling mercury species in effluent gases from chimneys of waste incinerators, and cryogenic trapping, which has also been used to isolate the analytes from the atmosphere. Papers that have reported the cryogenic trap method include one by Pecheyran et al. who trapped a number of volatile species of analytes such as Pb, Sn, As, P and Se.[13] The compounds were collected on a small glass wool column held at $-175\,^{\circ}C$. Another approach to the collection of mercury species has been to trap it on silver or gold traps.[14] Aerosols have been analyzed by collection of the particulate material on filter papers.[15] In this particular application, Cr^{III} and Cr^{VI} were speciated. A general overview of the speciation of chemical forms of metals and metalloids in the air has been published by Pecheyran.[16]

Sampling of solid matrices can be problematic. Sample inhomogeneity often requires large amounts of material to be collected. Sampling methods depend on the material of interest, but for soils and sediments, corers have found a great deal of usage. If specialized corers are not available, then make-shift ones made ''in-house'' using a rigid plastic pipe may be used. The use of corers allows depth profiles to be obtained. Once collected the core may be removed from the pipe and split into the respective horizons. For other solid matrices, e.g. plant tissue, collection may be as simple as picking the leaves or fruits or removal of stems. For root or tuber analysis, soil must be removed (usually by washing) to prevent sample contamination. Other biological environmental samples, such as fish or animal tissue, may be collected by the use of nets or, in the case of shellfish, by harvesting them from rocks. In the case of fauna, it is often necessary to collect several animals of approximately the same size and hence the same age.

Sample preservation is not as straightforward as in ''total'' metal determinations. Under normal circumstances, water samples are acidified to prevent adsorption of metals to glass containers. The addition of acid may alter the speciation for many analytes, e.g. change the oxidation state or perhaps demethylate some species. Similarly, potassium dichromate is often added to samples to stabilize mercury. This may also lead to species oxidation. The effect of the preserving agent on the species of interest should always be determined using pure standards prior to sample collection. If the speciation is found to be unaltered, it is permissible to use the preservation agent.

Preservation of solid matrices is relatively simple. Often, the sample must be dried although care must be taken to ensure that this does not lead to a change in speciation or loss of volatile species. Freeze drying is often preferred to oven drying, as the risk of species loss is less. Drying of the material prevents the growth of mould on many biological samples (which may change the speciation). Some laboratories may irradiate the samples to prevent the growth of mould. This should again be viewed with suspicion as irradiating samples may also lead to species conversion (e.g. dealkylation of organotin species). It is worth emphasizing again that the effects of any process performed on, or reagent added to, the sample should first be checked using pure standards of the species of interest. Only then is it possible to deduce whether or not it is a valid step in the analytical process.

Sample storage can also lead to analyte losses, contamination or speciation change. As described in the introduction, most organotin and some arsenic species are light sensitive. It is therefore important that they be kept in the dark. This may be achieved by storage in darkened glass bottles or by wrapping clear glass containers with aluminum foil. Any sample degradation may be slowed by freezing the sample, or by storing it below room temperature (e.g. at $4\,^{\circ}C$ in a refrigerator). It must be noted that freeze-dried samples will slowly absorb water and therefore if they are analyzed at a later date, a further drying stage (on a separate subsample to that being analyzed) must be performed so that an accurate moisture content correction can be applied. As the moisture content increases, the likelihood of bacterial decomposition increases and hence refrigeration becomes even more important. Storage of ''anoxic'' sediments may be particularly problematic.

3 SAMPLE PREPARATION

3.1 Aim of Sample Preparation

The aim of sample preparation is to obtain the species of interest in a form and at a concentration in which

it may most easily be determined using the techniques at the disposal of the analyst. Sample preparation techniques may consist of a number of different analytical steps, all of which have some sources of error or uncertainty associated with them. An overview of the pitfalls in speciation analysis has been produced by Rauret and Rubio.[17] The amount of sample preparation required depends on several factors including the matrix type, the species of interest and the method of analysis. Gaseous samples normally require least sample preparation because often they may be introduced directly to an instrumental technique such as gas chromatography (GC). Liquid samples may require preconcentration so that the concentration of the species is increased to a level where determination is facilitated. This may be achieved in a number of ways, including solvent extraction and enrichment using solid-phase extraction media. Another sample preparation step that may be necessary is filtration to remove unwanted solids, e.g. twigs or suspended particulates. This may also pose problems, because some species, e.g. TBT, are known to adsorb very strongly to particulate material. If the analyst filters the sample, the majority of TBT will be removed and hence erroneously low estimations of its concentration will be made. Solid matrices require more attention. The species must be extracted from the solid material in a quantitative (or at least representative) manner, whilst ensuring that no change to the speciation has occurred. This is often the most difficult part of the analysis, with numerous different methodologies having been developed.

Once isolated from the matrix, the species may sometimes be determined directly, but depending on the method of analysis further sample manipulation may be required. GC requires the analytes to be volatile. Therefore if the species are relatively involatile, they must be converted into a form that can be vaporized more easily. This requires a derivatization step.

The extraction, derivatization and cleanup steps will be described in more detail below. A brief overview of extraction procedures for speciation analysis of environmental samples has been made by Morabito.[18] This paper briefly discusses the most common extraction techniques for Pb, Sn, Hg, Cr, As, Se and Sb in liquid and solid samples.

3.2 Extraction of Species from Liquid Samples

This process is one of the easier steps in speciation analysis. Simple liquid–liquid extractions have proved popular because they afford large preconcentration factors (in excess of 100), whilst separating the species of interest from the bulk of the matrix (and hence potential interferences). Other species require derivatization followed by extraction into a suitable solvent prior to analysis.

Solid-phase extraction is becoming far more popular, as it is readily automated. An overview of solid-phase extraction applied to metal speciation has been published by Urasa et al.[19] There are a variety of techniques that may be used, including assorted resin-based media, extraction discs, polyurethane foam and, more innovatively, bacteria and fibers such as sulfydryl cotton. After retention of the species of interest on the solid medium, they may be removed by a variety of eluents, normally an organic solvent, such as hexane, toluene or, if the species have any polarity, methanol.

Often, species of interest may be extracted from the matrix by a direct derivatization process (e.g. Grignard derivatization) that leaves the analytes in a gaseous form. If this occurs, the analytes may be swept to detection by a flow of inert gas.

3.3 Extraction of Species from Solid Samples

The extraction of organometallic species from solid matrices is fraught with problems. Normal wet chemical procedures for sample decomposition, e.g. fusion, acid digestion with concentrated nitric/perchloric and hydrofluoric acids and dry ashing, cannot be used as they will lead to species alteration or complete speciation breakdown. Much more gentle methods of extraction are therefore required. It is often difficult to obtain 100% extraction efficiency for all species. The analyst is then confronted with a problem. Is the portion not extracted because of incomplete extraction of one or all of the species present, or, are there one or more species present that are not extracted at all? This problem can be impossible to answer when the analyst is uncertain what species are present in the first place, which unfortunately is likely to be the case for the vast majority of samples. Spiking experiments are of limited use because in many instances standards of all the relevant species are unavailable. Examples where standards are unavailable include many of the arsenosugars (arsenoribofuranosides) and many of the environmentally occurring species of antimony. The subject of quality control and valid analytical measurement of speciation analysis will be discussed further in section 5.

Before species extraction can be attempted, some preliminary sample preparation is normally required. For ''total'' metal determination, the sample is usually ground into a fine powder, but occasionally this may not be a good idea for speciation analysis as discrimination may occur against species that associate with larger particle sizes or, in the case of anoxic sediments, oxidation may occur destroying some species. This is especially a problem for samples such as soils and sediments where some species may be associated with the larger silica particles that resist grinding, whereas others may be associated with

humic substances that are more easily ground. This is not a problem for organotin speciation because it has been shown that the species are strongly bound to the fine-grained sediment fractions (<63 μm).

There are a number of extraction methods reported in the literature. New methods of extraction should first be tested on pure standards of the species of interest to ensure that no alteration to the speciation occurs. Once it has been established that the extraction protocol does not alter the speciation, further experiments to elucidate extraction efficiency from real samples may be performed. Overviews of different sample extraction methodologies have been produced, including one by Morabito[18] and another by Wells.[20]

Extraction using an organic solvent is still proving popular. There are a number of approaches to this, depending on the analytes of interest. Arsenic species have been extracted from biological tissues using a mixture of methanol–water, with the process being accelerated by the use of an ultrasonic bath. Soxhlet extraction may also be used, but care must be taken to ensure that losses of very volatile species, e.g. organomercury compounds, do not occur. Often, a reagent such as tropolone, is used to improve extraction efficiency. This reagent also improves the extraction efficiency of many unwanted matrix constituents that may affect the chromatographic separation of the analyte species. These concomitants may have to be removed in a cleanup step prior to the analysis being performed.

Alkaline extraction using tetramethylammonium hydroxide has also been achieved. This reagent is, however, very aggressive and hence is efficient at extracting many species, but may lead to a change in speciation. Another gentle extraction method for use with biological sample preparation is that of enzymolysis. The enzymes used will depend on the sample type. For proteinaceous material, e.g. fish/shellfish or meat tissues, a protease or trypsin has been found to yield good recoveries of arsenic species. For carbohydrate-based material, e.g. plant tissue (cellulose), trypsin was found to have a poor extraction efficiency, but the enzyme cellulase has been found to have an efficiency of over 85%. A detailed protocol for an enzyme digestion is shown in Figure 1. The enzyme used here is trypsin, but the procedure is almost identical for a cellulase digestion except that an optimal pH of 5 is required. In addition to arsenic, tin species have also been determined after sample decomposition using enzymes.

Supercritical fluid extraction is a method that is growing in popularity. Specialized equipment is required, but the technique yields good recoveries. Several analytes have been extracted including organotin and organomercury compounds.

Vacuum distillation is a technique that has been used to extract organomercury species. This technique has faced some criticism by workers who claim that artificially high results are obtained because the process leads to the methylation of inorganic mercury. This underlines the problems associated with speciation analysis. Many of the materials that have been certified for methylmercury have used the vacuum distillation process for matrix removal. The validity of these materials is therefore doubtful.

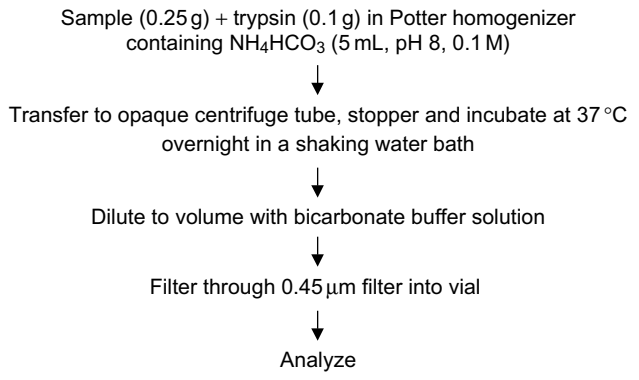

Figure 1 Enzymolysis procedure for extracting organometallic species from proteinaceous biological materials.

Table 1 Approaches used to extract organotin compounds from mussel tissue

Tropolone + benzene + HBr/ethanol/ascorbic acid	Extraction accelerated with ultrasonication
Extraction with 1 M HCl	Focused MW accelerated extraction
Homogenization with tetramethylammonium hydroxide	Alkaline extraction extracts butyl tins, but may cause rearrangements with phenyl tin compounds
NaOH and methanol	Butyl tin compounds extracted
Dichloromethane/ether + tropolone	Butyl and phenyl tins extracted
Acetic acid + back-extraction into toluene	Butyl and phenyl tins extracted, but method takes more than one day
Hexane + HCl	Butyl tins extracted
Protease + lipase	Enzymatic digestion extracts butyl and phenyl tins

MW, microwave.

Table 1 shows some of the approaches used to extract organotin compounds from mussel tissue. It is not a comprehensive review, but does give an indication of some of the methods performed.

3.4 Derivatization

As described previously, analysis techniques that incorporate a separation by GC may require a derivatization step to ensure that the analyte is in a form volatile enough to be

eluted from the column. This step is especially necessary for analytes such as many organotin, organolead and organomercury compounds. Many of these compounds are ionic in nature and are therefore relatively refractory. They must therefore be reacted with a reagent to form volatile, completely organic molecules. For example, tetramethyl lead is nonionic and is therefore readily determined by GC. Other species such as Me_3Pb^+, Et_3Pb^+, Me_2EtPb^+, Me_2Pb^{2+} and so on require derivatization. One of the most common methods of chemical derivatization is the use of a Grignard reagent. These all have the basic structure: organic moiety–Mg–halogen, where the organic moiety may be an alkyl or aryl group. This organic moiety may be used to replace the anion (often chloride) on the analyte species, hence converting them to a volatile form. Care must be taken when selecting the Grignard reagent, to prevent confusion of the results. For example, if EtMgBr was used for derivatization, then the lead species Me_2EtPb^+ and Me_2Pb^{2+}, which are both found in environmental samples, would yield the same species once derivatized, i.e. Me_2Et_2Pb. To prevent such confusion, Grignard reagents containing propyl or butyl groups are used most often.

Many mercury compounds may also be derivatized before separation using GC. The majority of laboratories determining $MeHg^+$ use an ethylation derivatization, which can be introduced via a Grignard reagent, but is more commonly based on the use of tetraethylborate. A review on the use of sodium tetraethylborate as a derivatization agent has been published.[21] Methylmercury must normally be separated from the bulk matrix before ethylation can be attempted, to prevent severe interference effects exerted by sulfur-containing compounds. Such compounds interfere with the efficiency of ethylation and may therefore cause erroneously low and imprecise data to be obtained. Extraction of the material using methanolic potassium hydroxide has been reported as being a suitable pretreatment method to achieve this.

Ionic organotin compounds must also be derivatized prior to GC separation. This may be done in two basic ways, i.e. in situ hybridization/ethylation using sodium tetrahydroborate or sodium tetraethylborate, or by an extraction followed by the use of a Grignard reagent. Assorted Grignard reagents have been used for organotin speciation. These include methyl, ethyl, propyl, pentyl and hexyl magnesium chlorides/bromides, although the ethyl and pentyl forms are most commonly used. It is reported that the alkylation is quantitative and that the derivatives are stable once formed.

The problem with some derivatization protocols is that they may not be reproducible or quantitative. This may arise because of partial destruction of the derivatizing agent or for a number of other reasons, such as impurities in the derivatization agent. If the derivatization process is not reproducible, the data obtained could be inaccurate and/or imprecise. Despite the potential problems, many laboratories experienced in speciation analysis have developed successful and reliable procedures. A brief overview of some of the applications of derivatization is given in Table 2.

Postcolumn derivatization may be necessary for some high-performance liquid chromatography (HPLC), FI or capillary electrophoresis (CE) applications. This is especially the case when either fluorescence or UV/visible absorption is used as the method of detection. Few organometallic species fluoresce or absorb UV/visible light naturally and hence after chromatographic separation they must be reacted with an agent that does. Examples of such reagents are nitrilotriacetic acid (this absorbs UV/visible light) and 8-hydroxyquinoline-5-sulfonic acid (this fluoresces).

Cleanup of the sample may take several forms. Enzymolysis of solid samples does not produce a completely solid-free matrix. Instead, it leaves a suspension that has most analyte species in the liquid phase. It is common for suspensions such as these to be passed through

Table 2 Some applications of derivatization

Analyte	Separation/preconcentration	Derivatization	Matrix
Organolead	Diethyldithiocarbamate extraction	Propylation	Surface waters
Organolead	Diethyldithiocarbamate extraction	Butylation	Rainwater
Organolead	Diethyldithiocarbamate extraction	Butylation	Soil
Organolead	Diethyldithiocarbamate extraction	Propylation	Aerosol
Organotin	Methanol/HCl	$NaBH_4$	Oyster
Organotin	HCl (2 M)	$NaBH_4$	Mussel
Organotin	Enzymolysis	$NaBEt_4$	Mussel
Organotin	Ethanol/KOH	Propyl magnesium chloride	Fish
Phenyl tins	Hexane	Methyl magnesium chloride	Potatoes
Organotins	Tropolone/HCl	Ethyl magnesium chloride	Sediment
Organotins	HCl/HBr/tropolone	Pentyl magnesium bromide	Sediment
Methyl mercury		$NaBEt_4$	Fish
Methyl mercury		$LiBEt_3H$	

a 0.45-μm polypropylene filter. This method can prevent blocking of HPLC columns, but care must be taken to ensure that some species are not lost. In some cases, it may be best if filtration does not occur and that a disposable guard column be fitted prior to the analytical columns. Occasionally, after a liquid–liquid solvent extraction has been performed, a back-extraction may be necessary either to preconcentrate the analyte further, or to separate the analyte from other extracted matrix constituents. Small quantities of activated alumina or silica or silica gel have also been used to achieve this. Small columns of fluorisil have also been used to clean up extracts containing organotin compounds. As always, it is necessary to ensure that none of the species of interest are lost using the cleanup procedure.

4 SAMPLE ANALYSIS

There are a number of instrumental methods used for speciation analysis. Of these, the most common is a coupling between a separation method (chromatography) and a selective method of detection (atomic spectrometry or mass spectrometry (MS)). There are however a number of other methods in common use. The following sections will describe the methods in more detail, highlighting the relative advantages and disadvantages of each and illustrating them with selected applications.

4.1 Coupled Techniques

As described previously, this is the most common method of affecting speciation analysis. There are a number of chromatographic methods of separation of analytes and an even larger number of atomic or MS methods of detection. Methods of chromatography include HPLC, GC and CE. Occasionally ion-exchange media may be packed in microcolumns and these are then used on-line with a suitable detector. Such techniques are often termed FI procedures. Atomic spectrometric methods of detection include flame atomic absorption spectrometry (FAAS), electrothermal atomic absorption spectrometry (ETAAS), inductively coupled plasma atomic emission spectrometry (ICPAES), inductively coupled plasma mass spectrometry (ICPMS), microwave-induced plasma atomic emission spectrometry (MIPAES) and atomic fluorescence spectrometry (AFS). The choice of which techniques to couple together will be governed by several factors including:

- the nature of the analyte
- the ease of coupling
- the sensitivity required
- the availability of the instrumentation.

In theory, most of the separation techniques can be coupled with any of the detectors, although there are practical limitations. For all of the methods, the sample is introduced as a discrete aliquot whose volume can range from 50–100 μL for HPLC down to 1 nL for CE. The individual analyte species will be detected as a transient signal. The distance between the chromatograph and the spectrometer should be kept to a minimum to avoid peak broadening as this will lead to decreased precision at lower concentrations and hence poorer limit of detection (LOD).

Harrison and Rapsomanikis have edited a book on the subject of environmental analysis using chromatography coupled with atomic spectroscopy.[22] Although elderly, the book gives a good overview of many of the techniques available and describes numerous early applications. A paper that gives an overview of chromatography coupled with atomic spectrometry and assesses the performance of such techniques has been produced by Quevauviller.[23] Several reviews have been made of the coupling between chromatography and atomic spectrometric detectors. These include ones by Hill et al.,[24] two by Ebdon et al.[25,26] and a more recent one by Ellis and Roberts.[27] In addition, a comprehensive review of chromatography and CE coupled with ICPMS has also been produced.[28] The following sections will describe the coupling of each type of chromatography with atomic spectrometry in more detail.

4.1.1 Liquid Chromatography Coupled with Atomic Spectrometry

There are several types of resin used in HPLC, including anion exchange, cation exchange, reversed phase (which is often used in conjunction with an ion pairing reagent to separate the species of interest) and chelation exchange. In general, it is the ion-exchange resins that have found most use as they are less affected by matrix components in real samples. The eluent (mobile phase) used will be governed by the nature of the analyte and by the type of resin used. The eluent may be inorganic in nature, e.g. it may contain salts dissolved in water, or it may be based on an organic solvent such as methanol or acetonitrile.

HPLC may be coupled with FAAS by simply connecting the outlet of the HPLC column to the uptake tube of the nebulizer. The flow rate of HPLC is typically 1 mL min^{-1}, which is substantially less than the natural uptake rate of the nebulizer (5–8 mL min^{-1}) caused by the Venturi effect. If the incompatible flow rates are not compensated for, the signal will be very noisy, hence making the observation of a transient signal more difficult. The problem of incompatible flow rates may be overcome easily by placing a piece of rubber with a slit in it in the junction between the HPLC and the nebulizer uptake

tube. The slit in the rubber tubing allows the ingress of air and this ''air bleed'' compensates for the differences in the flow rates. Applying this small modification to the interface leads to improved peak shapes and improved LOD because of the improved noise characteristics.

As described earlier, the eluent may be inorganic or organic in nature. There are potential problems with both types of mobile phase. Those that contain substantial amounts of dissolved solids may lead to salting up (blockage) of the nebulizer or, less likely, the slot on the burner head. If this occurs, the instrument may have to be shut down whilst the nebulizer is unblocked using a very thin piece of wire. Organic solvents often have the advantage of increasing sensitivity, because the reduced surface tension of these solvents leads to improved nebulization efficiency and hence increased sample transport to the flame. However, once in the flame the organic solvent combusts and carbon deposits may start to clog the slot on the burner head. Such deposits will cause gaps to appear in the flame, hence reducing the path length and thereby reducing sensitivity. The deposits may often be removed (often without recourse to flame extinction) by gently rubbing the slot with a noncombustible item such as a spatula.

The coupling between HPLC and FAAS is the simplest of the coupled speciation techniques because it is so straightforward and the flame is so tolerant of all the mobile phases. The largest drawback is the lack of sensitivity. For the majority of applications, the technique is incapable of detecting species at the concentrations at which they occur in the environment. Modifications to the system can be made, however, to improve the sensitivity. Optimization of the flame chemistry is the most obvious way of achieving this. For instance, when compared with an air–acetylene flame, a hydrogen flame exhibits a substantially decreased background signal in the UV region of the spectrum. This flame should therefore be the first choice for determining tin at the 224.6-nm line. Specialized nebulizers such as the hydraulic high-pressure nebulizer increases the sample transport efficiency and hence improvements in sensitivity of ten-fold may be achieved. This nebulizer was developed especially for use with sample introduction via an HPLC pump.

Another method to increase the sensitivity is the use of a slotted tube atom trap (STAT). This is a quartz tube placed on the burner head so that the light beam from the hollow cathode lamp passes longitudinally through it. The sample is introduced via the slot in the tube and, once present, the analyte species will be atomized by the flame in the normal way. Instead of passing out of the light beam at high speed as they are in the absence of the STAT, they are contained within the tube for a longer period, hence increasing their residence time in the light beam. The analytical sensitivity is therefore improved.

Despite the overall poor sensitivity of the technique, several applications of organometallic speciation using HPLC coupled with FAAS have been published. HPLC has also been coupled with ETAAS (also known as graphite furnace atomic absorption spectrometry). In ETAAS the sample is introduced to a graphite tube which then undergoes a heating program in which it is raised to increasingly high temperatures first to dry, then ash the sample to remove as many of the matrix constituents as possible, thereby decreasing interference effects, and finally to atomize the analyte. It is a technique that yields detection limits that are typically a factor of between 10 and 100 times better than FAAS. Under normal circumstances, it is a batch technique, i.e. a discrete aliquot of sample is introduced, the temperature program is run and the result obtained and then the process is repeated for the next sample. It is this batch nature that has led to only a few applications of speciation. When HPLC is coupled with ETAAS, the eluent, which is a continuous flow of liquid, passes into the graphite tube via a transfer line. To obtain atomization of the eluting species, the normal temperature program is discarded in favor of maintaining the tube at a constantly high temperature (typically 1400–3000 °C, depending on the analyte and on the furnace design). Many HPLC separations can take in excess of 15 min to complete and running the graphite tube at such an elevated temperature for such a long time rapidly leads to tube oxidation and destruction. The high price of graphite tubes can therefore make the on-line coupling of HPLC with ETAAS prohibitive for many laboratories. Despite this, there have been several applications that have used its coupling. Some of these have been cited in the review by Ebdon et al.[26]

Speciation analysis has also been achieved by HPLC separation of analytes followed by detection of the species off-line by ETAAS. This is achieved by collecting discrete aliquots of the eluent (perhaps 30 s worth) and then analyzing the aliquots using ETAAS in its normal mode of operation. Analysis of the eluent fractions enables an analyte concentration against elution time graph to be plotted. Although this method is very sensitive, it does suffer from some drawbacks. With so many aliquots being collected during a single chromatographic run, it is easy to mislabel an aliquot and hence confuse the issue by inadvertently ''changing'' the retention time of one of the species. Another problem arises when more than one species elutes close together. A continuous method of analysis may distinguish between two closely eluting species, however these may ''co-elute'' when aliquots covering 30 s are collected. Despite these drawbacks, the method has been used for speciation analysis by many workers.

The coupling between HPLC and inductively coupled plasma (ICP) instrumentation can be relatively straightforward. The uptake rate of an ICP nebulizer is typically $1\,mL\,min^{-1}$ which is comparable to the flow rate of most HPLC systems. The coupling is not without its problems though, and careful optimization of operating conditions and the interface is necessary to ensure optimal separation and sensitivity. Glass concentric pneumatic nebulizers, such as the Meinhard, which are frequently used for both ICPAES and ICPMS analyses, are easily blocked by dissolved or suspended solid materials. A dissolved solid content above 0.2% m/v is likely to lead to nebulizer blockage and therefore the use of these nebulizers may prove to be inappropriate for many HPLC separations (especially those that use an inorganic mobile phase containing dissolved salts). Fortunately, there are a large number of other nebulizer types that can tolerate dissolved solids more readily. Examples include the de Galan, the Ebdon and assorted other V-groove and frit nebulizers. These should be used in preference to the Meinhard when mobile phases containing dissolved solids are to be aspirated.

The next part of the sample introduction system is the spray chamber. There are a variety of different types available commercially, including the Scott double pass, miniature versions of the Scott double pass, single pass and cyclone. Each type has its associated advantages and disadvantages. The majority of spray chambers are made of glass, although inert plastic ones are also available for some designs. The Scott double pass has a large internal surface area and regions of dead volume. A large internal surface area that is made of glass will give ample opportunity for analytes such as lead (and its species) to adsorb to the walls. This may lead to some peak broadening which may affect the resolution, the precision and LOD. Similarly, regions of dead volume (i.e. regions where there is a relatively stagnant flow of transport gas and sample) also lead to memory effects. These problems are less acute in reduced-volume spray chambers, although these may be less capable of damping out noise arising from the pump and be less efficient at partitioning the droplets produced by the nebulizer according to their size. This means that larger droplets may occasionally enter the plasma, which will lead to poorer noise characteristics. Single pass and cyclone spray chambers have a smaller internal surface area when compared with the Scott double pass and also have fewer areas of dead volume.

The sensitivity, resolution and noise characteristics of the analysis depend critically on the nebulizer/spray chamber assembly. It is worth spending time in optimizing the combination as drastic improvements in performance may be obtained.

A number of specialized nebulizers have been developed that operate at reduced sample flow rates (at the microliter per minute level) and therefore these may aspirate the sample directly into the plasma torch, thereby bypassing the need for a spray chamber whilst improving sample transport rate to close to 100%. Many of these specialist nebulizers therefore have improved LOD when compared with conventional pneumatic ones. Examples of some of these nebulizers include the direct injection nebulizer (DIN), the oscillating capillary nebulizer (OCN), the electrospray and the thermospray. A more detailed list of the uses of such nebulizers may be found in the annual Atomic Spectrometry Update reviews that may be found in the June edition of the *Journal of Analytical Atomic Spectrometry* (for ICPAES) or the October edition of the same journal (for ICPMS).

Other problems may be encountered during speciation analysis using HPLC coupled with ICP instrumentation. As well as possible blockage of the nebulizer, mobile phases containing dissolved solids may also block the injector tubes of some plasma torches. This again may be overcome by optimizing the injector type used. Many torches are partially demountable, i.e. the injector may be replaced by another of different internal diameter, different design or made from a different material. Dissolved solids may also block the sampling cone orifice of an ICPMS instrument. Blockage of either the injector or the sampling cone orifice will lead to excessive signal drift until, when blockage is complete, no signal is obtained. The other problem associated with coupling HPLC with plasma instrumentation is the use of organic solvents as mobile phase. These solvents tend to have a higher vapor pressure than water and hence will have higher transport efficiency to the plasma. This will often lead to plasma quenching and possibly extinction. The magnitude of the problem will depend not only on the type of solvent used, but also on the individual instrument used. Some instruments cope more readily with the introduction of organic solvents than others, because it depends on the speed and quality of the instrument's matching network, as well as the frequency of the generator (40.68-MHz generators cope more readily than those fixed at 27.1 MHz). Another problem associated with the introduction of organic solvents when using ICPMS instrumentation is that the solvent decomposes thermally, producing large amounts of carbon deposits. These block the sampling orifice and also clog the ion lens stack. Blockage of the sampling orifice leads to signal drift and will eventually require the plasma to be turned off, so that it may be cleaned. If the lens stack becomes dirty, complete instrumental shutdown is required whilst it is cleaned. This can take several hours, which is obviously undesirable in a busy laboratory.

Because so many problems can arise through the introduction of organic solvents, in the majority of cases it is best to reduce the amount of solvent reaching the plasma. This can be achieved in a number of ways. The simplest way of doing it is to chill the spray chamber. Most spray chambers are surrounded by a jacket through which a cooling liquid is circulated. Normally this is water at a temperature of approximately 4 °C, but when organic solvents are being aspirated, it is better to add antifreeze to the cooling water. In this way the temperature may be reduced to −5 °C or below, depending upon the power of the chiller/circulator. The cold temperature reduces the vapor pressure of the solvent, causing less of it to be transported to the plasma whilst still allowing the analyte ions to be transported. This method is a standard procedure in many laboratories where speciation analysis is performed.

An alternative method is to use a specialized desolvation device. Some instrument manufacturers have marketed such devices, but many workers develop their own instrumentation. An example of such a device could use a mixture of IR lamps (to vaporize as much eluent as possible) followed by a permselective membrane (so that vapor present preferentially passes through the membrane and is removed by a counterflow of inert gas) followed by a region chilled by Peltier coolers (which condenses out remaining organic solvent at −30 °C). The condensed organic solvent could then be pumped to waste. This device can be placed between the nebulizer and the plasma torch. They are found to have a desolvation efficiency of approximately 90%, whilst not affecting the analytical sensitivity, and may be used for either ICPMS or ICPAES instrumentation.

Any solvent that does reach the plasma may be prevented from clogging the ICPMS sample cone orifice or lens stack by introducing a flow of oxygen into the nebulizer gas flow. Introduction may be either via a gas blender (a device capable of introducing between 0 and 100% of a gas into the flow of another gas) and a mass flow controller, or by an alternative gas inlet that many of the more modern instruments possess. Oxygen at 4% v/v in argon is sufficient to ''burn off'' carbon produced by 85–100% methanol (even without a desolvation device). It should be noted that excess oxygen leads to oxidative attack of the sampling cone, which for standard nickel cones can be drastic. Excess oxygen can reduce the lifetime of a nickel cone from months to seconds. The use of platinum-tipped cones that are more resistant to oxidative attack are therefore recommended when oxygen is introduced.

Despite this list of drawbacks associated with coupling HPLC and ICP instrumentation, many workers have successfully made the coupling. This is because the problems are not insurmountable and once the instrumentation has been optimized, sensitive and reliable analyses can be performed. A comprehensive listing of the applications is inappropriate, but mention will be made of some reviews of the subject. As well as the reviews described previously[24,26] that supply the reader with an indication of the range of analytes, matrices and methods used, a review that focuses on the coupling between HPLC using microbore columns and atomic spectrometry has also been published.[29]

Liquid chromatography has also been coupled with microwave-induced plasmas (MIPs) to affect speciation analysis. MIPs can be operated economically using helium rather than the argon that is used for ICPs. Helium is a far more ionizing gas than argon (first ionization potential of 24.59 eV compared with 15.76 eV) and it can therefore be used to excite analytes such as the halogens, sulfur, nitrogen and phosphorus that are, at best, only slightly excited in an ICP. Traditionally, the problem associated with MIPs is that they are easily extinguished by liquid entering them. In early research, even a few microliters of solvent injected into a GC column had to be vented before the analyte reached the plasma. It was thought that coupling with liquid chromatography would be impossible, but by increasing the power of the plasma to several hundred watts, liquid has been introduced. This is a potential area of speciation analysis for the future, especially as many laboratories now have access to desolvation systems. In absolute terms, femtogram levels of metals and nonmetals could be determined.

Liquid chromatography has now been coupled with AFS detection, both with and without an additional MW-assisted reduction and hydride generation (HG) stage. It is worth mentioning FI techniques here, because many of the points described for HPLC are equally applicable for these methods. Many of the FI techniques used for speciation have been reviewed in a publication that contains 58 references.[30] FI techniques use microcolumns containing ion-exchange resins or chelating resins, fibers such as sulfydryl cotton or other exchange media. Since the volume of the columns is so small, analyte separation may be achieved by using discrete volumes of eluent. With only 50–100 μL of eluent entering the plasma, the problems encountered with organic solvents are diminished. However, the use of oxygen to remove the solvent becomes complicated because of the lack of solvent entering the plasma continuously. When an inorganic eluent is used, the problem with injector and cone blockage is reduced substantially.

4.1.2 Gas Chromatography Coupled with Atomic Spectrometry

GC has been coupled less often with atomic spectrometry than HPLC. This has in part been due to the extra

sample preparation steps required for a GC analysis (i.e. derivatization reactions), but also because of the ready availability of GC/MS systems. The chief complexity of GC/atomic spectrometry is that the analytes must pass from an elevated-temperature GC column to the atom cell. For this to occur, a heated transfer line is required to prevent the analytes condensing on cool spots and hence becoming lost analytically. The transfer line should not be too hot otherwise thermal decomposition of the analyte species may occur and this may also lead to condensation of less volatile products. The construction of heated transfer lines has been discussed in the literature but the complexity depends critically on the analytes of interest. For volatile analytes, a transfer line consisting of a metal tube surrounded by a nichrome winding will suffice, but for very refractory analytes, e.g. various porphyrins which require the temperature to stay above 350–400 °C, specialist interfaces are required. Injector and column temperatures must also be optimized to prevent condensation or species breakdown.

Coupling with FAAS is one of the easier of the speciation techniques. The heated transfer line may be placed between the chromatograph and a STAT or an ordinary quartz tube with a hole in it (so that the transfer line may introduce the eluent) placed on the burner head of the FAAS instrument. The technique is relatively straightforward but suffers from a lack of sensitivity. Despite this, there have been several applications published. These include many of those cited in the review by Ebdon et al.[25] Coupling between GC and ICP instrumentation has been achieved by several workers, but at present can hardly be regarded as routine. Plasmas tend to be enclosed to prevent eye damage to operators. The transfer line must therefore often have to pass through torch box walls or doors to reach the plasma. This means that holes must occasionally be machined in the ICP instrumentation and there is a reluctance to do this.

The transfer line must usually pass as far up the injector tube of the torch as possible to prevent condensation of the analyte, without risking it picking up radiofrequency (RF) power. This is especially likely when the transfer line contains metallic components. If the line does pick up RF it may act as an aerial and transmit the power back to the gas chromatograph or possibly the operator, which could cause damage to either. Coupling of GC with ICPMS instrumentation can lead to absolute LOD substantially below a picogram. Several workers have published reports using GC/ICPAES and GC/ICPMS.

GC has also been coupled with MIPAES. A commercial instrument has been produced, and these instruments have been used by several workers to achieve speciation analysis. An example is the determination of organotin compounds. A review of the environmental applications of GC/MIPAES has been made that focuses on sample preparation and the element-specific detection of both nonmetallic and metallic pollutants.[31]

4.1.3 Capillary Electrophoresis Coupled with Atomic Spectroscopy

Capillary zone electrophoresis (CZE) is a relatively new separation technique. It has a very low flow rate (nanoliters per minute) and at present this is not readily compatible with the uptake rate of most atomic spectrometric detectors. The extremely small sample volume used (nanoliter range) means that as much sample should reach the atom cell as possible and that the actual concentration of the analyte must be quite high (typically at the microgram per milliliter range). DINs that have a transport efficiency of close to 100% are ideal for this type of small volume sample introduction, but even these have a natural uptake rate at the microliter per minute level. Specialist interfaces must therefore be constructed to ensure that the Venturi effect exerted by the nebulizer does not draw the sample through the column at an accelerated rate, thereby ruining the separation of the species. This has been reported in a few research papers, but as yet it cannot be regarded as a routine technique. A status report detailing the requirements and potential of CE coupled with plasma spectrometry has been published.[32] Recently, there have been a number of interface designs published.[33–35]

4.1.4 Vapor Generation Coupled with Atomic Spectrometry

4.1.4.1 Introduction There are a number of metallic and metalloid elements that form vapors, e.g. volatile hydrides and mercury vapor, at room temperature. These vapors may then be transported in a flow of inert gas to a detector at an efficiency of close to 100%. The high sample transport rate makes the process of vapor generation very popular for these analytes because of the improved LOD. There are problems associated with the technique however, the most important of which being that not all species of an analyte form a gaseous hydride. Examples of such species include the organoarsenic compounds arsenobetaine and arsenocholine and the organoselenium compounds selenomethionine, selenocysteine and even inorganic Se^{VI}. For species such as these to be determined, specialized sample pretreatment is necessary. This will be discussed later. Another important problem is that caused by transition metals. The presence of these elements in the sample interferes with the hydride-forming process and therefore erroneously low hydride formation yields may result, causing an underestimate of the true concentration

of the analytes. This problem can be alleviated by the addition of masking agents such as L-cystine or L-cysteine.

Several different methods of detection are available for HG techniques. These include atomic absorption spectrometry (AAS), atomic emission spectrometry (AES), ICPAES, ICPMS and AFS. The technique of AFS has the advantages of being very sensitive and having a linear range that covers several orders of magnitude. Commercial instruments have been marketed that have a vapor generation unit coupled with an element-selective AFS detector. Although relatively cheap, they are often the first choice of instrumentation for analytes such as mercury, that can be troublesome by other methods. The coupling between the vapor generation system and the detector is relatively straightforward, because the sample is transported as a gas. It should be noted however that some plasma-based instruments may be disturbed by excessive amounts of hydrogen that may be a product of the hydride-forming reaction. Similarly, the presence of large volumes of hydrogen going into the flame of an AFS instrument may also lead to flame perturbation, dilution and decreased sensitivity. The chemistry of the hydride-forming process will therefore have to be optimized to prevent flame or plasma perturbation and decreased sensitivity. It is therefore essential to consider the system as a whole rather than optimizing the hydride-forming process and the detection process independently.

For normal HG analysis the sample is mixed with a reductant in an acidic medium and the analytes form gaseous hydrides that may be swept from a gas–liquid separator by an inert gas. Several reductants may be used, but sodium tetrahydroborate is the most common for many analytes although tin(II) chloride is often used to reduce mercury ions to elemental mercury vapor. A good overview of the use of sodium tetrahydroborate (also known as sodium borohydride) for elemental speciation has been produced by Howard.[36]

4.1.4.2 Cryogenic Trapping Methods Some simple organometallic forms can be separated from others by cryogenic trapping. In this process, the hydride of the analytes should be formed in the normal way, but instead of transportation directly to the atom cell they may be trapped in a U tube packed with a silanized chromatographic support or glass beads. The U tube should be immersed in liquid nitrogen, hence freezing the gaseous hydrides. The U tube may then be removed from the liquid nitrogen and allowed to rise to room temperature (or possibly above room temperature if a heating device is used). If this is performed whilst flushing an inert gas through the system, the individual hydrides revert to gaseous form at their own respective boiling points and are transported to the detector. In this way a pseudochromatographic separation may be achieved. Detection may be by atomic emission, atomic absorption or ICPMS. Often the atom cell is a quartz T-piece placed on the burner head of an atomic absorption instrument and heated either by the flame or electrothermally. The T-piece operates in a similar manner to the STAT, i.e. it enhances sensitivity by increasing the residence time of the analyte in the light beam. A masking agent (L-cysteine) and a cryogenic trap have been used to determine inorganic arsenic, monomethylarsonate and dimethylarsinate. The cysteine has the added bonus of giving uniform sensitivity for all the species determined. A schematic diagram of the overall apparatus for HG used in conjunction with cryogenic trapping is shown in Figure 2.

4.1.4.3 Chromatography Coupled with Hydride Generation Techniques Cryogenic trapping experiments have been used in association with chromatographic separation of the species prior to the HG step. Both GC and, more commonly, HPLC have been coupled with HG. A chromatographic separation is of more use than simple cryogenic trapping because it allows the determination of more analyte species, but cryogenic trapping can allow concentration of the analyte.

For compounds such as those described earlier that do not form a hydride, several on-line techniques have

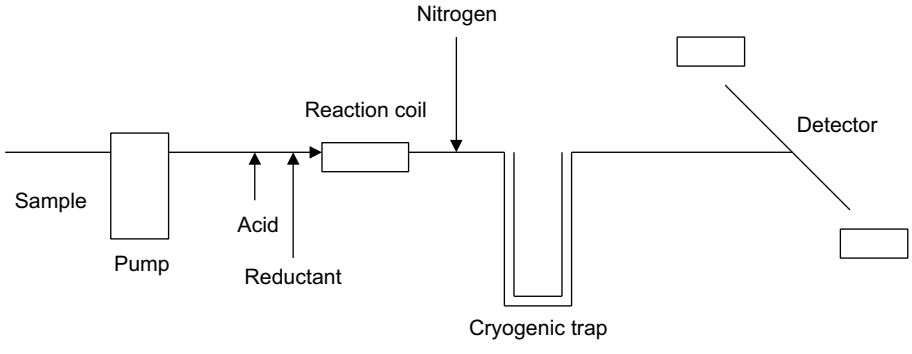

Figure 2 Schematic diagram of the instrumentation required for cryogenic trapping of hydrides.

Table 3 List of some applications utilizing on-line sample pretreatment

Element	Matrix	Method	Comments
As	Sediment	HPLC/UV/HG/AAS	Photooxidation in presence of 2% m/v $K_2S_2O_8$
As	Fish	HPLC/MW/HG/AAS	Thermal oxidation in MW using 5% m/v $K_2S_2O_8$ and 5% m/v NaOH
As	Marine tissues	HPLC/MW/HG/AAS	Thermal oxidation in MW with 5% m/v $K_2S_2O_8$ and 3.4% m/v NaOH followed by on-line reduction with 2.5% L-cysteine
Se	Water	HPLC/MW/HG/AAS	On-line oxidation using 3% m/v $K_2S_2O_8$ and 2% m/v NaOH
Se	Urine	HPLC/MW/HG/AAS	On-line digestion using BrO_3^-/Br^-

been developed that transform the species in a known and reproducible way. If these methods are used, then all the species present in a sample may be separated chromatographically and then transformed, perhaps to a single species that forms a hydride. The components are then detected as a series of transients in a time-resolved way. Several on-line methodologies are available. The use of a strong oxidizing agent such as potassium persulfate is known to decompose compounds such as arsenobetaine. Other methods include photooxidation and the use of heating. Many reported methods have used one or more of these techniques to transform the species into a reducible form. The use of MW ovens on-line has therefore found use in this area as the MWs heat the eluent very rapidly thereby assisting in both the thermal and chemical decomposition of the species. Numerous examples exist in the literature. Some of these on-line methodologies are presented in Table 3. A thermal bath may also be used, but this acts far slower than an MW oven and is therefore less efficient for on-line techniques. In general the efficiency with which sample decomposition occurs decreases in the order MW digestion > UV photolysis > thermal bath. Addition of a reagent such as persulfate will accelerate the decomposition processes further.

The majority of applications involving speciation via chromatography coupled with HG and atomic spectrometric detection have probably involved the determination of As, although other analytes such as Sn and Se have also been determined effectively.

The use of such techniques with detection by AAS yields LODs of approximately 0.5–10 $\mu g\,L^{-1}$ (which for an injection volume of 0.1 mL leads to an absolute LOD of 0.05–1 ng), depending on the species of interest. For some analytes, the use of AFS or ICPMS detectors will lead to improvements in the LOD. Some research papers have now been published linking HPLC/UV/ICPAES. This application may be used to photolyze arsenobetaine and arsenocholine prior to arsine generation and detection. The coupling between chromatography, HG and assorted detectors is likely to be exploited more often in the future for speciation analysis in an ongoing attempt to improve versatility and sensitivity. A schematic diagram of the instrumentation used is shown in Figure 3.

4.2 Other Methods of Speciation

The coupling of a separation technique (chromatography) with an element-selective (or specific) detector (such as atomic spectroscopy) is a common method of speciation analysis. There are however several other techniques

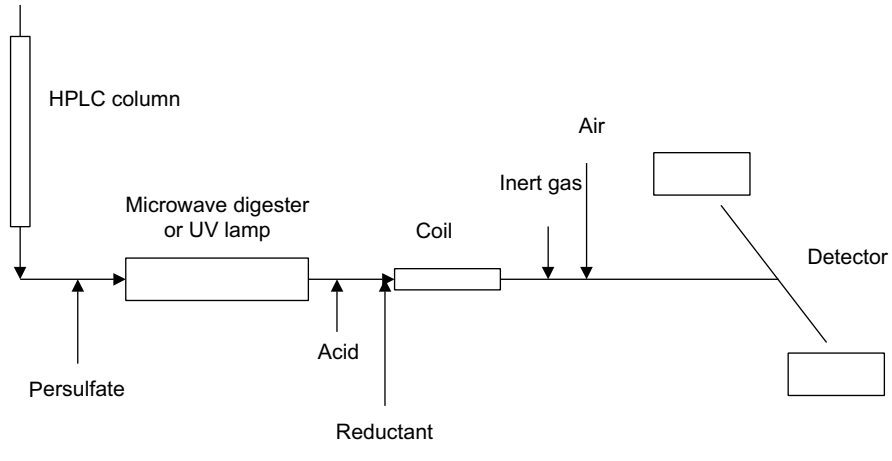

Figure 3 Schematic diagram of the instrumentation required for the determination of organometallic compounds by HPLC/HG/AAS.

that have been used, although some of them suffer from poor sensitivity or interferences. Many still use chromatography to separate the species but then use conventional detectors. Examples include GC separation followed by MS (GC/MS). This mature technique has been used, for example, for the speciation of mercury in fish and water samples, trimethyllead in rainwater, arsenic speciation in an aqueous environment and the determination of an assortment of analytes (tin, antimony and bismuth compounds) in landfill and fermentation gases. Other detectors that have been used in conjunction with GC for speciation analysis include the flame photometric detector (FPD), which has found special use for tin speciation, the FID that has been used for several analytes, including arsenic compounds, and the electron capture detector (ECD). Each has its own specific advantages and problems. The FID is quite sensitive, but will give a response to any compound that contains carbon. Therefore, for many sample types, the chromatograms may have a plethora of peaks, only a few of which relate to the species of interest. Other detectors such as the FPD, which is also very sensitive, may suffer from interference effects. Only a single wavelength isolation device is used, which allows a broad band of radiation to be detected. Therefore any components in the sample that emit light at a wavelength similar to the analyte will also be detected. There is often a need for matrix elimination when this technology is being used. This is normally possible with the aid of derivatization reactions, but some sample types may still be problematic. One such type occurs when organotin compounds are to be determined in a sample that contains numerous volatile sulfur compounds. The sulfur will interfere strongly with the detection of the tin. When used for tin speciation, the FPD can yield a LOD of $0.02\,ng\,mL^{-1}$. The technique can therefore be regarded as being extremely sensitive (more so than coupled HPLC/ICPMS). The latter technique however would not suffer the interference from sulfur. The ECD is very sensitive for certain elements, e.g. the halogens, but not for most metals, and therefore its application to environmental analysis is limited.

Similarly, with HPLC separations, the traditional methods of detection, i.e. UV/visible absorption and fluorescence, may be used. The UV/visible method has poor sensitivity. Fluorimetric detection has also been reported. This method yields high sensitivity, but often requires postcolumn derivatization. The method has been used to speciate aluminum complexes and organotins in waters. As with all fluorescence-based detection techniques, the method is very sensitive and has a long linear range.

CE has also been used in conjunction with an array of different detectors. These include UV detection for assorted arsenic compounds and for the determination of methylmercury in fish and crab. Amperometric detection has also been used for mercury species. This latter application enabled a LOD of $0.2\,ng\,mL^{-1}$ for Hg^{2+} and $3\,ng\,mL^{-1}$ for methylmercury, which for a 40-nL injection volume relates to an absolute LOD of 8 fg and 120 fg, respectively. The LOD by this technique is therefore extremely impressive, but it does suffer from the drawback of only being able to detect cationic mercury species that are electrochemically active. It was reported that in the analysis of a sediment at least one mercury species was unsuitable for detection. Another novel coupling has been CE with particle-induced X-ray emission (PIXE). An overview of speciation studies using CE has also been published.[37] Many applications require postchromatographic derivatization to transform the species into a state where they may be detected more easily. The current trend for speciation analyses is to couple several different techniques together in an attempt to improve applicability and sensitivity. CE has been coupled with HG prior to detection of the analytes (arsenic species) by ICPAES.

Assorted mass spectrometric methods other than GC/MS have also been used for speciation analysis. These include several versions of liquid chromatography/MS, such as liquid chromatography/electrospray/MS/MS (tandem mass spectrometry) and ionspray-MS. As the instrumentation required becomes more readily available and reliable, these approaches can be expected to grow in popularity.

Electrochemical techniques have also been used for speciation analysis. These methods are inexpensive and are also relatively portable. They are therefore commonly found aboard ships and hence are popular with oceanographers and marine chemists. Often these techniques are used to differentiate between oxidation states, but may occasionally be used to analyze for organometallic compounds or compounds that are associated with an organic moiety, e.g. ligands. An overview comparing electroanalytical techniques with atomic spectrometric techniques has been made by Bersier et al.[38] Several other papers of interest in this area may be found in the literature. Examples include the dissolved phase speciation of zinc in the Humber Estuary,[39] cadmium speciation and complexation by natural organic ligands in fresh water[40] and the soil solution speciation of lead.[41] In this latter paper, ''total'' lead was determined by ETAAS, and the labile lead determined by differential pulse anodic stripping voltammetry (DPASV). The labile Pb values were used to calculate the free Pb^{2+} activity based on the assumption that organolead complexes are not DPASV labile. The techniques may be problematic. For instance, the determination of organolead compounds may be interfered with by the presence of inorganic lead. This must be removed prior to analysis by the use of a resin such as chelex, by ethylenediamine tetraacetic acid or by coprecipitation

with barium sulfate. The sample preparation for these techniques therefore tends to be complex. Differential pulse polarography and differential pulse cathodic stripping voltammetry have been used to determine selenium species. It has been found that the presence of some transition metals (e.g. Cr^{VI}, Cu^{II}, Mo^{VI}, Ni^{II}, Zn^{II} and V^{V}) caused interferences and that these too had to be removed by the use of chelex prior to analysis.

The other drawback with electrochemical techniques such as anodic stripping voltammetry is that they are limited to only certain elements, i.e. only certain elements are electroactive. These metals include copper, lead, cadmium, zinc thallium, antimony, tin, bismuth, chromium cobalt, manganese, nickel and mercury. Fortunately, these are the elements of most environmental interest, but these limitations could be prohibitive for other sample types.

Neutron activation analysis has also been used as an offline detector for HPLC separations of different species. This technique can hardly be described as routine though.

NMR has been used to elucidate the structure of many arsenosugars found naturally in seaweed and algae. The poor sensitivity of NMR (even Fourier transform instruments) means that the analytes of interest have to be present at an appreciable concentration. In one paper, the authors started with several kilograms of material so that they could extract sufficient analyte to be determined.[42] To make the spectra more simple to interpret, purifications and separations of the species were performed using gel permeation chromatography or ion-exchange chromatography prior to the ^{1}H- and ^{13}C-NMR analysis. The determination of arsenolipids in seaweed has also been performed in a similar manner.[43] Other analytes have also had their structure elucidated using NMR spectroscopy. Examples include the speciation of copper in the bovine rumen environment[44] and the reaction of aquatic humic substances with aluminum.[45] The latter paper is typical of an NMR-based speciation analysis, because aluminum is the most commonly determined analyte by the technique.

A multimethod approach to studying humic compounds and metal speciation has been described by Petronio et al.[46] These workers used both NMR and IR spectroscopy as well as elemental analysis and thermogravimetry to study metal–ligand structures. In another paper[47] the Cr^{III} binding abilities of humic acids were determined using NMR and IR spectroscopy. It was found that the bonding was between the Cr and a carboxylate group.

5 METHOD VALIDATION

As with all types of analytical chemistry, methods should be validated to ensure that the data being produced are both accurate and reproducible. Ideally, all species should be isolated from the matrix with 100% efficiency and with no change to the speciation. This is rarely the case. Often extraction efficiency is less than 100% and the analyst must decide what course of action to take. Many species are lipophilic (i.e. they dissolve in the fat layer of biological materials such as fish). These species will be extracted very poorly by a polar solvent system such as methanol–water which has been used by several workers to extract arsenic species. On such occasions, the analyst must try to determine whether the species that have been extracted are the only ones present in the sample and that there was a small amount of them dissolved in the fat; or whether there may be one or more species present only in the fat that were not extracted. Several other problems exist for speciation analysis. Often, standards of all the relevant species are unavailable commercially. This can therefore cause problems with the identification of some species and will definitely cause problems when quantification is attempted. The absence of standards will also make spiking/recovery experiments more difficult. Examples of species that do not have commercially available standards include many of the arsenosugars and many antimony species. Ideally, CRMs that contain all the relevant species in the matrix of interest would be available. Unfortunately this is rarely the case. Despite these problems, method validation can be achieved under certain circumstances. Several papers have been produced that describe quality control techniques for speciation analysis. These include one that concentrated on hyphenated techniques.[48]

5.1 Mass Balance Data and Spiking/Recovery Experiments

Mass balance data can be obtained if the sum of the analytes extracted are compared with the ''total'' amount of analyte. Therefore, if a water sample is to be analyzed for tin species, the water could be aspirated into a spectrometer (e.g. ICPMS) directly and the ''total'' concentration of tin determined. If a separate subsample then undergoes speciation analysis using perhaps HPLC/ICPMS, then the sum of the species should be the same as the ''total'' concentration. Similarly, if the material of interest is a solid, ''total'' concentrations of analyte may be determined by a strong acid digestion, and the speciation data may be obtained after an extraction technique is performed on a separate subsample. If the sum of the species equals the ''total'' concentration, then it is usually assumed that extraction is 100% efficient. This method does not account for any speciation changes during the extraction process. A summary of how mass balance data is obtained is shown in Figure 4. An example could be arsenic speciation in fish

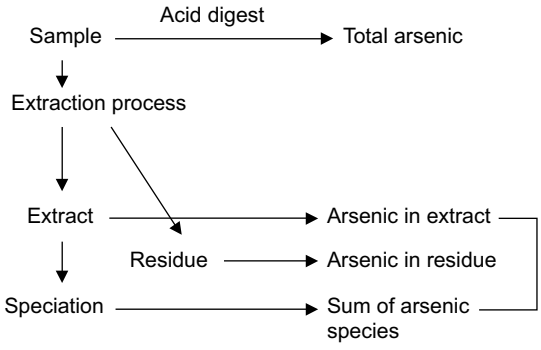

Figure 4 Process for obtaining mass balance data.

tissue using an extraction method with methanol–water and a ''total'' digest using nitric acid and hydrogen peroxide. If the ''total'' concentration was found to be $20\,\mu g\,g^{-1}$ and the sum of the species was found to be $17.5\,\mu g\,g^{-1}$ (as arsenic, not the species), then the overall extraction efficiency is given by Equation (1):

$$\frac{17.5}{20} \times 100 = 87.5\% \quad (1)$$

The recovery of the arsenic species may then be corrected for the losses observed during extraction (i.e. the concentrations of all observed species multiplied by the recovery factor, in this case 100/87.5). As described previously, this method is not infallible because there may be one or more species present that have 0% extraction efficiency whilst the others have 100% efficiency. By multiplying the results by the overall extraction efficiency, incorrect answers may be obtained.

Spiking/recovery experiments are another common way of method validation. It should be noted that the method of spiking is similar to that used during normal analysis, i.e. a small volume of a concentrated standard is spiked into a subsample and allowed to equilibrate (''soak in'') for at least 24 h. The now dry sample can then be extracted in the normal way. Unspiked subsamples should also be extracted in the same manner. It is also worth noting that the spike should be approximately the same concentration as that already present in the sample (or at least within a factor of 2–5). If the spike concentration is far in excess of this, unreliable data will be obtained. Similarly, the spike should be allowed to equilibrate for at least 24 h to ensure that it behaves at least partially like the analyte species already present in the sample (i.e. it has time to become adsorbed or chemisorbed to the matrix). Although this situation is not ideal, because it can take weeks or months for the same process to occur naturally in the environment, it is the only practical way of doing it. The problem does not arise with water samples because simple mixing is sufficient to ensure equilibration.

This method of validation is useful when standards of the species to be determined are available. It allows the recovery of individual species to be determined, so the recovery factor of each can be used to estimate the true concentration of that species in the sample.

If spiking/recovery experiments are used in conjunction with mass balance data, then reasonable estimates of the speciation can be obtained.

5.2 Certified Reference Materials

Speciation analysis is a subject that has only relatively recently become hugely popular. It is for this reason that the number of CRMs available with any speciation data are limited. There are an infinite number of environmental samples occurring naturally and several elements that form organometallic species. It is an impossible task to speciate all relevant analytes in all possible sample matrices and therefore manufacturers of CRMs have concentrated on only a few of most interest. A list of some of the CRMs available from different manufacturers is given in Table 4. In addition to these, there are also materials that, although not certified, do have consensus values agreed by several laboratories. The dogfish reference muscle DORM-1 available from the National Research Council Canada is one such material. Its total arsenic value is $17.7\,\mu g\,g^{-1}$ and its arsenobetaine concentration (as arsenic) is generally regarded to be in the range $15.5–16.5\,\mu g\,g^{-1}$.

As described previously, ideally a certified material with speciation data of every matrix would be available. This however is not the case. Even those samples that are ostensibly the same type may have only limited use. Several certified sediments are available with data for organotin compounds, but the actual matrix of the

Table 4 List of CRMs available with speciation data

NRCC PACS-1 harbor sediment	Organotins
NIES 11 fish tissue	Organotins
BCR 462 coastal sediment	Organotins
BCR 424 harbor sediment	TBT
BCR 477 mussel tissue	Organotins
NRCC TORT-1 lobster hepatopancreas	Hg and MeHg
BCR 463 tuna	Hg and MeHg
BCR 464 tuna	Hg and MeHg
BCR 584 estuarine sediment	Hg and MeHg
BCR 580 sediment	Hg and MeHg
BCR 545 welding dust	Cr^{III} and Cr^{VI}
BCR 544 freeze-dried solution	Cr^{III} and Cr^{VI}
NIES scallop muscle	As species
NIES algae	As species
BCR 627 tuna fish	As species
BCR solutions 602 and 603 (reference materials)	Se^{IV} and Se^{VI}
BCR 605 urban dust	Organoleads

sediments may differ widely. Some have a sandy matrix, whereas others may be very rich in humic acids and other organic materials. A given extraction method may have a higher efficiency for some matrices than others. Therefore, if sediments are to be analyzed, it may not be sufficient to analyze any certified sediment, as misleading data may result. Instead, it is necessary to pick a certified sediment that resembles the matrix of the sample sediments as closely as possible.

The difficulty in certifying samples for speciation data has been highlighted by unsuccessful certification campaigns. For example, The European Community attempted to produce a simulated rainwater sample containing organolead species. It was found that the lead species were unstable and that losses occurred. Another example was for the organotin speciation of a highly polluted sediment. It was found that data were obtained with an unacceptably high uncertainty and that the laboratories that relied on a derivatization step produced the least reliable results. Despite these setbacks, the production of certified materials for speciation analysis continues to grow. Quality procedures necessary to produce CRMs for speciation analysis have recently been reviewed in an excellent book by Quevauviller.[2]

5.3 Other Methods of Validation

The analysis of an appropriate CRM is the best procedure for method validation. Unfortunately these are not universally available for every speciation problem. The use of two completely different analytical methods is the next best approach but again often only one analytical method, or extraction procedure, is possible practically. Alternatives such as standard additions or spiking, as discussed above, have to be used.

The use of isotope dilution analysis (IDA) can also assist in obtaining accurate results. Although a full description of isotope dilution is beyond the scope of this text it may readily be found elsewhere.[11] Basically, the added isotope acts as an intrinsic internal standard that can overcome problems associated with incomplete extraction, but cannot correct for changes in speciation. It can however be a useful method of determining when changes to speciation occur. Several papers have been published that have used IDA for speciation purposes. In addition, two book chapters have also been published.[3] Where the isotopically enriched spike is added in the form of the species to be determined, additional value is obtained.

Although not strictly a method validation technique, it is worth mentioning the reporting of the data. For normal analysis, it is sufficient to give a ''total'' concentration of an analyte. For speciation analysis though, it is necessary to describe what actually has been measured, i.e. the species or the element. For instance, if the total arsenic concentration is $20\,mg\,kg^{-1}$, it is necessary to describe the concentration of the species in terms of arsenic rather than arsenobetaine and so on. This will prevent confusion. If the analyte species' concentration is reported, it is necessary to say so in the report.

6 CONCLUSIONS

Speciation analysis is a still relatively young area of analytical chemistry, even though it gives far more information on the overall toxicity of a sample and may be used to elucidate the reactions performed by naturally occurring phenomena. A list of the elements that have been speciated in environmental samples is shown in Table 5. A diagram of the LOD achievable using different techniques is indicated in Figure 5. The LODs have been quoted as absolute amounts for chromatographic based techniques, but it is worth remembering that the absolute LOD is dependent on the discrete volume introduced. Techniques such as CZE, which have a very impressive absolute LOD but only require nanoliter injection volumes, therefore require the analyte species to be relatively concentrated (e.g. at the microgram per milliliter level). There are several problems associated with the procedures, with methods that work for one sample not necessarily working for another. The development of robust and universal methods is therefore still a great challenge. Method validation is still an area under expansion as

Table 5 List of elements that have been speciated in environmental samples

Element	Species
As	Arsenobetaine, dimethylarsinic acid, monomethylarsonic acid, As^{III}, As^{V}, arsenosugars, arsenocholine, arsenolipids
Se	Se^{IV}, Se^{VI}, selenocystine, selenomethionine
Sn	Mono-, di- and TBT, mono-, di- and triphenyl tin, methylated tin
Pb	Alkyl lead compounds, lead ligands
Hg	Methyl- and dimethylmercury, ethyl- and diethylmercury, higher alkyl mercury compounds, mixed alkyl mercury compounds
Si	Siloxanes
Ge	Inorganic and methylated species
Assorted transition metals e.g. Cu, Cd, Zn, Fe	Metal–ligand complexes
Al	Humic and fulvic complexes
Sb	Sb^{III}, Sb^{V}, Sb-complexes

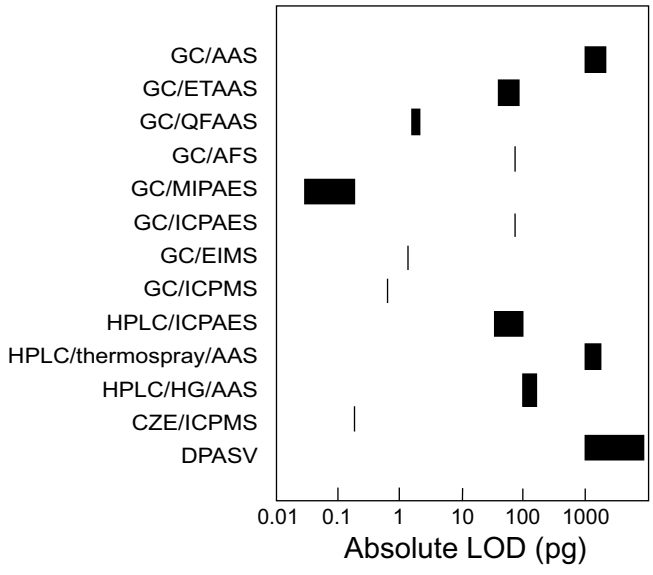

Figure 5 Absolute LOD for different speciation techniques. QFAAS = quartz furnace atomic absorption spectrometry.

the requirement for certified materials far outstrips their supply. As yet, the number of CRMs available is very limited, with some analytes, e.g. antimony compounds, not being certified in any matrix.

It is envisaged that as analysts adapt their methodologies to other sample types, whilst looking for more analytes at a lower concentration, the use of hyphenated techniques will continue to grow. The use of multiply hyphenated techniques is already appearing and this is likely to increase further to improve the LOD and enable more analytes to be determined. Similarly, more efficient methods of sample extraction and preconcentration will also be developed.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
AFS	Atomic Fluorescence Spectrometry
CE	Capillary Electrophoresis
CRM	Certified Reference Material
CZE	Capillary Zone Electrophoresis
DIN	Direct Injection Nebulizer
DPASV	Differential Pulse Anodic Stripping Voltammetry
ECD	Electron Capture Detector
ETAAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Flame Atomic Absorption Spectrometry
FI	Flow Injection
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
GC	Gas Chromatography
HG	Hydride Generation
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IDA	Isotope Dilution Analysis
IR	Infrared
LOD	Limit of Detection
MIP	Microwave-induced Plasma
MIPAES	Microwave-induced Plasma Atomic Emission Spectrometry
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MW	Microwave
NMR	Nuclear Magnetic Resonance
OCN	Oscillating Capillary Nebulizer
PIXE	Particle-induced X-ray Emission
RF	Radiofrequency
STAT	Slotted Tube Atom Trap
TBT	Tributyl Tin
UV	Ultraviolet

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Biological Samples in Environmental Analysis: Preparation and Cleanup • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by

Electrochemical Methods • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Inorganics in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Inductively Coupled Plasma/Optical Emission Spectrometry • Microwave-induced Plasma Systems in Atomic Spectroscopy

Gas Chromatography **(Volume 12)**
Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Supercritical Fluid Chromatography

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. S.J. Hill, 'Speciation of Trace Metals in the Environment', *Chem. Soc. Rev.*, **26**, 291–298 (1997).
2. P. Quevauviller, *Method Performance Studies for Speciation Analysis*, Royal Society of Chemistry, Cambridge, 1998.
3. P. Quevauviller, E.A. Maier, B. Griepink (eds.), *Quality Assurance for Environmental Analysis*, Elsevier, Oxford, 1995.
4. J.C. Van Loon, R.R. Barefoot, 'Overview of Analytical Methods for Elemental Speciation', *Analyst*, **117**(3), 563–570 (1992).
5. M. Burguera, J.L. Burguera, 'Analytical Methodology for Speciation of Arsenic in Environmental and Biological Samples', *Talanta*, **44**(9), 1581–1604 (1997).
6. D.C. Baxter, W. Frech, 'Speciation of Lead in Environmental and Biological Samples–Technical Report', *Pure Appl. Chem.*, **67**(4), 615–648 (1995).
7. K.M. Attar, 'Analytical Methods for Speciation of Organotins in the Environment', *Appl. Organomet. Chem.*, **10**, 317–337 (1996).
8. R. Puk, J.H. Weber, 'Critical Review of Analytical Methods for Determination of Inorganic Mercury and Methylmercury Compounds', *Appl. Organomet. Chem.*, **8**, 293–302 (1994).
9. R.M. Olivas, O.F.X. Donard, C. Camara, P. Quevauviller, 'Analytical Techniques Applied to the Speciation of Selenium in Environmental Matrices', *Anal. Chim. Acta*, **286**, 357–370 (1994).
10. X. Dauchy, M. Potingautier, A. Astruc, M. Astruc, 'Analytical Methods for the Speciation of Selenium Compounds–A Review', *Fresenius' J. Anal. Chem.*, **348**(12), 792–805 (1994).
11. E. Prichard, G.M. MacKay, J. Points (eds.), *Trace Analysis: a Structured Approach to Obtaining Reliable Results*, Royal Society of Chemistry, Cambridge, 1996.
12. R. Rubio, A. Ure, 'Approaches to Sampling and Sample Pretreatment for Metal Speciation in Soils and Sediments', *Int. J. Environ. Anal. Chem.*, **51**, 205–217 (1993).
13. C. Pecheyran, C.R. Quetel, F.M.M. Lecuyer, O.F.X. Donard, 'Simultaneous Determination of Volatile Metal (Pb, Sn, In) and Nonmetal Species (Se, P, As) in Different Atmospheres by Cryofocusing and Detection by ICPMS', *Anal. Chem.*, **70**, 2639–2645 (1998).
14. R. Ebinghaus, S.G. Jennings, W.H. Schroeder, T. Berg, T. Donaghy, J. Guentzel, C. Kenny, H.H. Kock, K. Kvietkus, W. Landing, T. Muhleck, J. Munthe, E.M. Prestbo, D. Schneeberger, F. Slemr, J. Sommar, A. Urba, D. Wallschlager, Z. Xiao, 'International Field Comparison Measurements of Atmospheric Mercury Species at Mace Head, Ireland', *Atmos. Environ.*, **33**, 3063–3373 (1999).
15. R. Nusko, K.G. Heumann, 'Cr^{III}/Cr^{VI} Speciation in Aerosol Particles by Extractive Separation and Thermal Ionization Isotope Dilution Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **357**, 1050–1055 (1997).
16. C. Pecheyran, 'Speciation of Chemical Forms of Metals and Metalloids in the Air: Issues and Perspectives', *Analusis*, **25**, M41–M44 (1997).
17. G. Rauret, R. Rubio, 'Sources of Error in Speciation Analysis', *Quim. Anal.*, **16**, 119–130 (1997).
18. R. Morabito, 'Extraction Techniques in Speciation Analysis of Environmental Samples', *Fresenius' J. Anal. Chem.*, **351**, 378–385 (1995).

19. I.T. Urasa, S.F. Macha, W. El Maaty, 'Application of Solid-phase Extraction in Metal Speciation', *J. Chromatogr. Sci.*, **35**, 519–524 (1997).
20. D.E. Wells, 'Extraction and Preconcentration of Organometallic Species from Environmental Samples', *Mikrochim. Acta*, **109**, 13–21 (1992).
21. S. Rapsomanikis, 'Derivatization by Ethylation with Sodium Tetraethylborate for Speciation of Metals and Organometallics in Environmental Samples – a Review', *Analyst*, **119**, 1429–1439 (1994).
22. R. Harrison, S. Rapsomanikis (eds.), *Environmental Analysis Using Chromatography Interfaced with Atomic Spectroscopy*, Ellis Horwood, Chichester, 1988.
23. P. Quevauviller, 'Atomic Spectrometry Hyphenated to Chromatography for Elemental Speciation: Performance Assessment Within the Standards, Measurements and Testing Programme (Community Bureau of Reference) of The European Union', *J. Anal. At. Spectrom.*, **11**, 1225–1231 (1996).
24. S.J. Hill, M.J. Bloxham, P.J. Worsfold, 'Chromatography Coupled with Inductively Coupled Plasma Atomic Emission Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**, 499–515 (1993).
25. L. Ebdon, S. Hill, R.W. Ward, 'Directly Coupled Chromatography–Atomic Spectroscopy, Part 1. Directly Coupled Gas Chromatography–Atomic Spectroscopy: a Review', *Analyst*, **111**, 1113–1138 (1986).
26. L. Ebdon, S. Hill, R.W. Ward, 'Directly Coupled Chromatography–Atomic Spectroscopy, Part 2. Directly Coupled Liquid Chromatography–Atomic Spectroscopy: a Review', *Analyst*, **112**, 1–16 (1987).
27. L.A. Ellis, D.J. Roberts, 'Chromatographic and Hyphenated Methods for Elemental Speciation Analysis in Environmental Media', *J. Chromatogr., A*, **774**, 3–19 (1997).
28. K. Sutton, R.M.C. Sutton, J.A. Caruso, 'Inductively Coupled Plasma Mass Spectrometric Detection for Chromatography and Capillary Electrophoresis', *J. Chromatogr., A*, **789**, 85–126 (1997).
29. H. Garraud, A. Woller, P. Fodor, O.F.X. Donard, 'Trace Elemental Speciation by HPLC Using Microbore Columns Hyphenated to Atomic Spectrometry: a Review', *Analusis*, **25**, 25–31 (1997).
30. L. Campanella, K. Pyrzynska, M. Trojanowicz, 'Chemical Speciation by Flow Injection Analysis – a Review', *Talanta*, **43**, 825–838 (1996).
31. S. Pedersen Bjergaard, T. Greibrokk, 'Environmental Applications of Capillary Gas Chromatography Coupled with Atomic Emission Detection – a Review', *HRC-J. High Res. Chromatogr.*, **19**, 597–607 (1996).
32. R.M. Barnes, 'Capillary Electrophoresis and Inductively Coupled Plasma Spectrometry: Status Report', *Fresenius' J. Anal. Chem.*, **361**, 246–251 (1998).
33. I. Karen, J.A. Caruso, 'Interfacing Electrophoresis with Inductively Coupled Plasma Mass Spectrometry', *LC/GC Mag. Sep. Sci.*, **17**, 36 (1999).
34. A. Tangen, W. Lund, B. Josefsson, H. Borg, 'Interface for the Coupling of Capillary Electrophoresis and Inductively Coupled Plasma Mass Spectrometry', *J. Chromatogr. A*, **826**, 87–94 (1998).
35. K.A. Taylor, B.L. Sharp, D.J. Lewis, H.M. Crews, 'Design and Characteristics of a Microconcentric Nebulizer Interface for Capillary Electrophoresis Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1095–1100 (1998).
36. A.G. Howard, '(Boro)Hydride Techniques in Trace Element Speciation – Invited Lecture', *J. Anal. At. Spectrom.*, **12**, 267–272 (1997).
37. E. Dabek Zlotorzynska, E.P.C. Lai, A.R. Timerbaev, 'Capillary Electrophoresis: the State of the Art in Metal Speciation Studies', *Anal. Chim. Acta*, **359**, 1–26 (1998).
38. P.M. Bersier, J. Howell, C. Bruntlett, 'Advanced Electroanalytical Techniques Versus Atomic Absorption Spectrometry, Inductively Coupled Plasma–Atomic Emission Spectrometry and Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis', *Analyst*, **119**, 219–232 (1994).
39. M. Gardner, 'Dissolved Phase Speciation of Zinc in the Humber Estuary', *Chemosphere*, **38**, 2117–2124 (1999).
40. H.B. Xue, L. Sigg, 'Cadmium Speciation and Complexation by Natural Organic Ligands in Fresh Water', *Anal. Chim. Acta*, **363**, 249–259 (1998).
41. S. Sauve, M. McBride, W. Hendershot, 'Soil Solution Speciation of Lead (II): Effects of Organic Matter and pH', *Soil Sci. Soc. Am. J.*, **62**, 618–621 (1998).
42. J. Edmonds, M. Morita, Y. Shibata, 'Isolation and Identification of Arsenic Containing Ribofuranosides and Inorganic Arsenic from Japanese Edible Seaweed *Hizikia Fusiforme*', *J. Chem. Soc. Perkin Trans.*, **1**, 577–580 (1987).
43. M. Morita, Y. Shibata, 'Isolation and Identification of Arseno-lipid from a Brown Alga *Undaria Pinnatifida* (Wakame)', *Chemosphere*, **17**, 1147–1152 (1988).
44. R.S. Reid, M.A. Attaelmannan, 'NMR Studies of Copper Speciation in the Bovine Rumen', *J. Inorg. Biochem.*, **69**, 59–65 (1998).
45. R.F. Howe, X.Q. Lu, J. Hook, W.D. Johnson, 'Reaction of Aquatic Humic Substances with Aluminum: a Al-27 NMR Study', *Mar. Freshw. Res.*, **48**, 377–383 (1997).
46. B.M. Petronio, B. Cosma, A. Mazzucotelli, P. Rivaro, 'A Multimethod Approach to Study Humic Compounds and Metal Speciation in Marine Sediment Samples', *Int. J. Anal. Chem.*, **54**, 45–56 (1993).
47. M. Fukushima, K. Nakayasu, S. Tanaka, H. Nakamura, 'Chromium(III) Binding Abilities of Humic Acids', *Anal. Chim. Acta*, **317**, 195–206 (1995).
48. P. Quevauviller, 'Improvement of Quality Control of Speciation Analysis Using Hyphenated Techniques – a Decade of Progress Within the European Community', *J. Chromatogr., A*, **750**, 25–33 (1996).

Pervaporation, Analytical

M.D. Luque de Castro and L. Gámiz-Gracia
University of Córdoba, Córdoba, Spain

1 Introduction 1
2 Principles of Pervaporation 2
3 The Analytical Pervaporator 3
3.1 Body of the Analytical Pervaporator 3
3.2 Membranes 4
4 The Auxiliary Manifold 4
5 Types of Analytical Pervaporation 5
6 Analytical and Physical Parameters of Pervaporation 5
6.1 Sensitivity 5
6.2 Selectivity 8
6.3 Precision 8
6.4 Physical Parameters of Analytical Pervaporation: Kinetics of the Mass-transfer Process 9
7 Scope of Analytical Continuous Pervaporation 9
7.1 Types of Samples 11
7.2 Derivatization Reactions 11
7.3 Individual Determinations 11
7.4 Multideterminations 11
7.5 Pervaporation as an Alternative to Headspace Methods 13
7.6 Speciation Studies 13
7.7 Integrated Pervaporation/Detection 14
7.8 Types of Detectors 15
7.9 Pervaporation as an Interface Between Evolving Systems and Monitoring 15
8 Prospects for Pervaporation 16
Abbreviations and Acronyms 16
Related Articles 16
References 16

The use of pervaporation in analytical chemistry is a relatively new development. Pervaporation is a separation technique for the removal of volatile analytes or their volatile derivatives from the sample matrix. It can be defined as the integration of evaporation and gas diffusion in a single module. The volatile substances present in a heated donor phase evaporate through a porous membrane and the vapor condenses on the surface of a cool acceptor stream on the other side of the membrane. The temperature difference across the membrane is the driving force for the separation. An important characteristic of analytical pervaporation is the presence of a constant-volume air gap between the sample in the donor chamber and the membrane, which hinders any contact between them, thus avoiding clogging and/or deterioration of the membrane.

Pervaporation can also be used for sample pretreatment, e.g. for solid samples where leaching and derivatization of the analytes are done simultaneously. In this article, the applications of analytical pervaporation that have been developed so far are discussed. Special attention is paid to its use as an alternative to gas diffusion, to liquid samples containing solids in suspension and/or corrosive, to solid samples in which a single analyte must be determined, to multideterminations, and to speciation studies. The integration of this separation technique with detection and the coupling of a pervaporator to a gas chromatograph (GC), which constitutes a valid alternative to headspace sampling, is also commented upon.

1 INTRODUCTION

Recent developments in automatic continuous analyzers have led to the use of nonchromatographic continuous separation techniques for the determination of one or more analytes, the separation of components into groups, and interference removal. The different principles on which current nonchromatographic continuous separation techniques rely overcome a variety of analytical problems associated with preliminary operations, which are normally complicated, slow, variable, difficult to control, labor-intensive, occasionally hazardous to both laboratory personnel and the environment, and the sources of major errors. These techniques (which can be applied in various areas, including environmental pollution, clinical and pharmaceutical chemistry, and food and industrial analysis), have emerged as very promising alternatives to chromatographic techniques on account of such assets as reduced human participation, easy on-line coupling with many instruments, diminished errors and hazards, increased sample throughput, and sparse sample and reagent consumption. Nonchromatographic continuous separation techniques can improve the sensitivity of a method via preconcentration, and its selectivity through avoidance of matrix effects or particular interferents. Coupling two equal or different continuous separation techniques can be a useful way of dealing with complex samples by exploiting a synergistic effect.[1]

Membrane-based nonchromatographic continuous-separation techniques are especially useful for increasing

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

selectivity. In some cases, they also permit analyte preconcentration, thereby resulting in improved sensitivity. Continuous gas diffusion, whether in isolation or coupled (in series or parallel) with other separation techniques such as dialysis and ion exchange, makes an excellent choice for achieving high selectivity and sensitivity. However, these techniques suffer from two serious drawbacks, namely potential clogging of membrane pores by suspended particles or components of a high molecular weight occasionally present in the sample plug, and deterioration of the membrane through contact with the sample. Both shortcomings can be overcome by using pervaporation (so-called in order to emphasize that the analyte or its reaction product, known as the ''permeate'', undergoes a phase change from liquid to vapor before reaching the membrane). Pervaporation has long been employed in industry, in competition with other traditional separation processes (e.g. distillation, extraction, adsorption); by contrast, it has scarcely been used for analytical purposes. On the laboratory scale, analytical pervaporation can be defined as the integration of two different analytical separation principles (evaporation and gas diffusion) in a single micromodule. Major simplification and miniaturization of the experimental set-up are thus obtained, and human participation is also reduced so as to increase analytical quality and productivity. Pervaporation is in principle a separation technique for the removal of volatile analytes or their volatile derivatives from the sample matrix but can also be used for sample pretreatment, e.g. for solid samples where leaching and derivatization of the analytes are done simultaneously.

2 PRINCIPLES OF PERVAPORATION

Industrial pervaporation[2–4] selectively separates a liquid feed mixture by partial vaporization through a nonporous polymeric membrane. The separation is not based on relative volatilities, as in the case of distillation or evaporation, but rather on the relative rates of permeation through the membrane. The mixture is circulated in contact with the membrane and the permeating component diffuses through it. A vacuum (vacuum pervaporation) or sweeping gas (sweeping-gas pervaporation) is applied to the membrane on the permeate side and the permeating component desorbs from the membrane as a vapor, which can be collected or released, as required. The driving force for separation is the chemical potential on both sides of the membrane. The permeating component is transported through the membrane because its partial pressure on the permeate side is lower than that in the saturated vapor.[5]

Separation from mixtures is obtained because the membrane transports one component more readily than the other, even if the driving forces are equal. The effectiveness of pervaporation is measured by two parameters, namely flux, which determines the rate of permeation, and selectivity, which measures the membrane's separation efficiency (controlled by the intrinsic properties of the polymers used for the membrane). The coupling of fluxes affecting the permeability of a component present in a mixture can be divided into two parts, namely a thermodynamic part, expressed as solubility, and a kinetic part, expressed as diffusivity. In the thermodynamic part, the concentration change of one component in the membrane due to the presence of another is caused by mutual interactions between the permeates in the membrane, in addition to interactions between the individual components and the membrane material. On the other hand, kinetic coupling is due to the dependence of the concentration on the diffusion coefficients of the permeates in polymers.[2]

For a binary mixture of components A and B, the flux can be expressed for the entire permeate (J_T = total flux) or for each component (J_A, J_B = flux of components A and B, respectively), having dimensions of mass/(area × time). The flux can be calculated provided that the mass of the permeating component, the membrane area, and the time of measurement are known. Also it can be defined according to Equation (1):

$$J_i = \frac{-L_i \Delta\mu_i}{l} \tag{1}$$

where L_i is the phenomenological diffusion coefficient, $\Delta\mu_i$ the chemical potential driving force across the membrane, and l the membrane thickness.

The selectivity is given by the ratio of the mass fractions of components A and B for the permeate and feed. In the case of selective permeation of the component A, Equation (2) applies:

$$\alpha_B^A = \frac{(y_A/y_B)}{(x_A/x_B)} \tag{2}$$

where x_A and x_B are the mass fraction of components A and B, respectively, in the feed, and y_A and y_B those in the permeate. A value greater than unity indicates selective permeation of A over B, whereas a value lower than unity indicates selective permeation of B over A. The selectivity is dimensionless and occasionally described as an enrichment factor β, which is the ratio of a component concentration in the permeate to its concentration in the feed.[5]

The process is favored by increased temperatures, which result in a vapor pressure difference that increases the permeability of the substances through the membrane by allowing the diffusion of permeated molecules and decreasing interactions between permeates. The membrane morphology (rubbery/glassy polymers) also influences the efficiency of the separation technique.

Even though pervaporation appeared at the beginning of the 20th century, research activity in this field for industrial applications increased in the 1970s owing to the energy crisis and has been focused on three types of separation: the dehydration of an aqueous–organic binary mixture by using water-selective hydrophilic membranes (e.g. dehydration of ethanol produced by fermentation in the sugar refinery of Bethéniville in France), the permeation of organic(s) from an aqueous–organic mixture using hydrophobic membranes (concerning solvent recovery, hazardous waste treatment, water purification, and beverage processing) and the permeation of a particular organic from an anhydrous mixture (e.g. aromatics–paraffin, branched hydrocarbons–paraffin, olefins–paraffin, and isomeric mixtures), which is not yet viable with currently available membranes. The advantages of pervaporation over conventional industrial separation processing can be summarized as follows: energy savings, easy separation of azeotropic mixtures, easy dehydration of multicomponent mixtures, no contamination of the product with entrained compounds, no environmental pollution caused by treatment of entrained compounds, easy operation, space savings, and easy installation on-site as the pervaporator is skid-mounted.[2]

Analytical pervaporation is a process in which volatile substances in a heated donor liquid phase evaporate through a porous hydrophobic membrane. The vapor condenses on the surface of a cool acceptor stream on the other side of the membrane. Surface-tension forces withhold the liquids from the pores and prevent direct contact between them. A temperature difference, which results in a vapor pressure difference across the membrane, provides a strong driving force for the separation, which also occurs in the absence of temperature gradient. Evaporation will occur at the solution surface if the vapor pressure is greater than that at the condensate surface. One important feature of the pervaporation module used for analytical purposes is the air gap between the donor stream and the hydrophobic membrane, which avoids any contact between them and diminishes the problems associated with fouling of the membrane.

The mechanism of transport by pervaporation can be described by the solution-diffusion model,[6] which comprises the following steps:

- evaporation of the analyte into the air gap;
- sorption into the membrane on the downstream side;
- diffusion of the sorbed component through the polymer matrix;
- desorption into a liquid or gas phase on the upstream side.

The last three steps are also included in industrial pervaporation.

Several applications of laboratory pervaporation have been developed for the determination of analytes or their reaction products in both liquid and solid samples, integrated or not with detection.[7–26] Since pervaporation is a very mild process that can be operated at the required temperature and needs no high pressure, cross-flow velocity, or additional chemical, it is particularly well adapted for integration with fermentation processes.

3 THE ANALYTICAL PERVAPORATOR

A laboratory-scale pervaporator is a very simple unit that can be designed and built in the laboratory according to requirements.[6,7,27–30] Two different parts must be considered in it: the body of the separation module and the membrane.

3.1 Body of the Analytical Pervaporator

The body of the pervaporation module consists of the following parts (Figure 1): a lower, donor chamber through which the feed stream, containing the analyte or its volatile product, is circulated (a); spacers of varying thickness, from 2 to 10 mm, which can be placed below and/or above the membrane support in order to increase the volumes of the corresponding chambers (b); a thin membrane support (ca. 1 mm thickness) (c); an upper, acceptor chamber with inlet and outlet orifices through which the acceptor stream is circulated and in which the gaseous analyte is collected (d); and fluid-circulating thermostats (e) which increase the difference

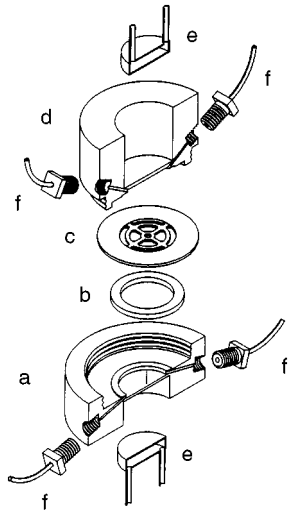

Figure 1 Parts of a conventional pervaporation module. (a) Donor/sample chamber; (b) spacer; (c) membrane support; (d) acceptor chamber; (e) fluid-circulating thermostats; (f) connectors.

in temperature between the two chambers, so improving the pervaporation process.

The whole module can be made of methacrylate, which permits continuous checking during experiments, although a Perspex pervaporation unit has also been proposed recently.[30] The membrane support is made of polytetrafluoroethylene (PTFE). The upper and lower chambers (and the membrane support) are kept together by screwing them. The temperature is adjusted by circulating a fluid with the appropriate temperature,[29] by immersing the module in a thermostated waterbath,[6,7,8–15,19,20,21–23,25,26] by inserting an electronic temperature controller,[28] or by focused microwaves.[18]

This kind of unit has been used in combination with a flow-injection (FI) manifold for the determination of various analytes in liquid samples; the separation is followed by detection and the donor chamber is usually heated by a conventional water bath. Alterations of either the auxiliary dynamic manifold or the pervaporator itself are required when (i) the pervaporation process is assisted by focused microwaves, (ii) the separation step assists in the continuous monitoring of an evolving system, (iii) solid samples are used with no pretreatment, and (iv) pervaporation is integrated with detection.

(i) The module used for microwave-assisted pervaporation[18] is similar to that shown in Figure 1, but with smaller dimensions (external diameter less than 35 mm) so that it can be accommodated inside the microwave vessel. Since no metal parts are admitted in the microwave chamber, the cell (screws included) is constructed entirely in PTFE/methacrylate.

(ii) When the pervaporation unit is used for the continuous monitoring of a fermentation process or the determination of analytes in samples containing suspended particles, the module in Figure 1 must be altered by increasing the diameter of both channels (inlet and outlet) of the donor chamber, which could otherwise be clogged.

(iii) For solid samples, a septum is placed at the inlet of the lower chamber through which the reagents are injected with a syringe. The outlet of the same chamber is shut with a screw so as to avoid leakage and the consequent analyte losses.

(iv) When separation and detection of the volatile species occur simultaneously, a hole is drilled at the center of the upper chamber top of the pervaporation cell in order to accommodate the sensor[10,11] by means of suitable adapters. This module can be used for liquid[10,11] and solid samples;[10] the flow of the acceptor stream is stopped during measurement in order to permit the accumulation of the analyte released from the matrix during the process and hence increase the sensitivity of the method.

3.2 Membranes

In general, permeability in glassy polymers (e.g. cellulose) is lower than in rubbery polymers (e.g. poly(dimethylsiloxane) PDMS) and selectivity is primarily determined by the molecular dimensions of the permeating species.[31] It has also been demonstrated that both the pore size of the membrane and its thickness strongly affect the pervaporation process and its selectivity.[30] The polymers used as membranes for the pervaporation process are similar to those employed for gas separation and have a ''dense'' nonporous macroscopic structure; the difference between the two techniques lies in the transport mechanism and arises mainly from a large difference in affinity between the permeating molecules and the polymeric membrane.

The membranes used for analytical pervaporation are hydrophobic membranes of the types usually employed in ultrafiltration and gas-diffusion processes. In practice, PTFE is the most frequently used membrane material,[6–16,18–20,27,28] followed by hydrophobic poly(vinylidene fluoride) (PVDF).[6–8,19] Ultrafiltration membranes[6,8,9,15,16,18–20,23] are very thin, which, in combination with the large surface area of both the donor and acceptor chambers, leads to their easy bending. This results in a change in the flux of the permeating component through an altered membrane area, and hence in a change in the efficiency of the process; as a consequence, the membranes must be changed fairly often. Gas-diffusion[11–14] membranes are not so easily bent because of their thickness, so the same membrane can be used over longer periods. The pore size of the membrane (or its directly proportional thickness) influences the efficiency of the process and hence the sensitivity of the method; as a result, it affords application to samples with variable analyte concentrations.[13,14]

4 THE AUXILIARY MANIFOLD

A typical flow injection/pervaporation configuration is shown in Figure 2(a). The sample is either injected into the carrier stream or directly aspirated and mixed (or not) with the reagent (R1), in passing through the mixing coil, RC1. The resulting stream reaches the lower part of the pervaporation cell, which is thermostated at the required temperature. The volatile compound evaporates, diffuses through the hydrophobic membrane, and is collected in an appropriate solution with which it may or not react.

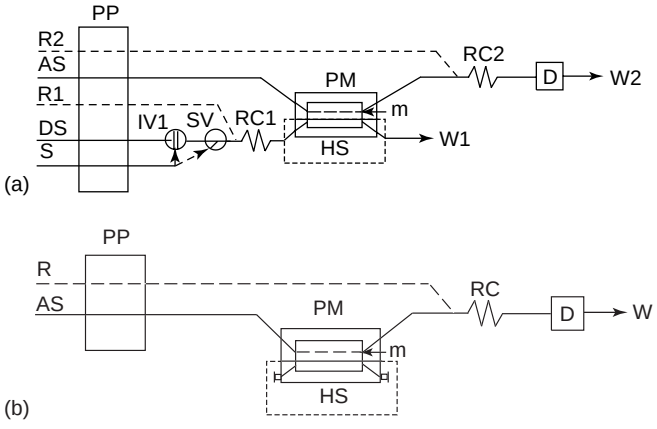

Figure 2 Dynamic configuration–pervaporation approaches (a) for liquid samples and (b) for solid samples. S = sample; DS = donor solution; R = reagent; AS = acceptor solution; PP = peristaltic pump; IV = injection valve; SV = selection valve; RC = reaction coil; PM = pervaporation module; HS = heating source; D = detector; W = waste system; m = membrane. The dashed lines refer to potential derivatization reactions and the dotted line to continuous sampling. (Reprinted from D. Bryce, A. Izquierdo, M.D. Luque de Castro, 'Continuous Microwave Assisted Pervaporation/Atomic Fluorescence Detection: An Approach for Speciation in Solid Samples', *Anal. Chim. Acta*, **324**, 69–75.[18] Copyright (1996), with permission from Elsevier Science.)

The outlet of the acceptor chamber can be either led to the detector or merged with another reagent stream to give the finally detected product.

5 TYPES OF ANALYTICAL PERVAPORATION

The qualifiers ''continuous'' and ''discontinuous'' for pervaporation are used to refer to different aspects of the process. In fact, analytical pervaporation is a continuous technique because, while the sample is in the separation module, mass transfer between the phases is continuous until equilibrium is reached. ''Continuous'' also refers to the way the sample is introduced into the dynamic manifold for transfer to the pervaporator. When the samples to be treated are liquid, the overall manifold to be used is one such as that shown in Figure 2(a). The sample can be continuously aspirated and mixed with reagent(s) if required (continuous sample introduction). The ''discontinuous'' mode of sample introduction is performed by injection of a liquid sample, either with an injection valve of the manifold, followed by sample transport to the pervaporator, or by means of a syringe furnished with a hypodermic needle, directly into the lower/donor chamber of the separation module when no dynamic manifold is used. In any case, the sample reaches the lower chamber of the separation module and the volatile analyte (or its reaction product) evaporates, diffuses across the membrane, and is accepted in the upper chamber by a stream or static solution that drives it continuously or intermittently, respectively, to the detector (except in the cases where separation and detection are integrated). The pervaporation unit can also be used with solid samples in a discontinuous way. In this case, the sample is weighed in the donor chamber; then, the module is tightly closed and the reagents, if required, are injected through septa. The manifold used in this case is shown in Figure 2(b). A heating source at the appropriate temperature, acting on the donor chamber, starts the process. In this case, the separation module is employed as a leaching and reaction vessel with or without integrating the detection step. Sample manipulation and losses of analyte are hence limited.

The flow of donor (sample) and acceptor solutions (lower and upper chamber, respectively) can be continuous or stopped at any time during the separation step in order to increase the efficiency of the process.

6 ANALYTICAL AND PHYSICAL PARAMETERS OF PERVAPORATION

6.1 Sensitivity

The sensitivity of a method involving pervaporation basically depends on the efficiency of the separation step. In order to fit the signal provided by the analyte to the linear portion of the calibration curve of a given method, the efficiency of the mass transfer through the membrane can be manipulated thus avoiding either a dilution or a preconcentration step.

6.1.1 Efficiency of Pervaporation

The efficiency of a given pervaporation process must be evaluated in order to act on the overall system so as either to boost or to decrease the mass transfer to the acceptor chamber if required. This evaluation can be done either in relative terms by comparing the signal provided under different working conditions by the analyte (or its reaction product), previously collected in the upper chamber and driven to the detector, or in absolute terms by comparing the signal obtained under the working conditions with that corresponding to 100% mass transfer.

The relative procedure is used in optimization experiments in order to maximize sensitivity, which depends not only on mass transfer, but also on the best conditions of the derivatization reactions (prior and/or subsequent to pervaporation) and on dispersion along the continuous system, among others. No especial alteration of the

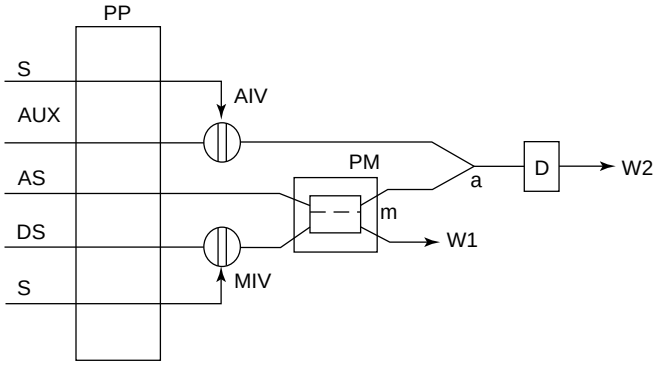

Figure 3 Simplified scheme of a hydrodynamic manifold for evaluation of the efficiency of a pervaporation process. Note that when a derivatization reaction is required before or after analyte separation, additional channel(s) for the reagent solution(s) must be included in the donor or acceptor subsystem, respectively. AIV and MIV = auxiliary and main injection valve, respectively; PM = pervaporation module; m = membrane; a = merging point; D = detector; W = waste; AUX = auxiliary channel containing acceptor solution; S = sample; AS = acceptor solution; DS = donor solution. (Reproduced by permission of John Wiley and Sons.)

manifold is needed in this case, other than those resulting from the optimization process.

The absolute procedure requires an auxiliary channel in the upper subsystem, through which the acceptor solution is also passed, furnished with an auxiliary injection valve (AIV in Figure 3) of the same characteristics as the main injection valve (MIV) that is used to inject the sample for analyte separation and monitoring. Two sequential sample injections (the sequence is uninfluential) from both valves are needed to evaluate the efficiency of the mass-transfer process. First, a volume of sample containing a given analyte concentration is injected via MIV; the analyte evaporates to the air gap, diffuses through the membrane, is absorbed by the acceptor stream, and is driven to the detector. The signal produced is proportional to the amount of analyte transferred from the donor to the acceptor stream. Second, an identical volume of the same sample is then injected through AIV and driven to the detector. A signal corresponding to the whole amount of injected analyte is obtained. The comparison of the two signals provides the amount of analyte transferred in the separation process relative to that in the sample. Note that the dispersion undergone by the monitored species in the two channels is different, so the two must be compared in terms of the areas of the transient signals (representative of the total amount of monitored species which has passed by the detection point) rather than their heights, which are representative of the concentration.[19]

6.1.2 Adjusting the Pervaporation Efficiency

Most often, improved efficiency is the aim of optimizing a method in order both to lower the determination limit and to obtain a steeper slope for the calibration curve. When the concentration of the target analyte in the sample varies widely, a real or pseudo dilution step is mandatory in order to fit the concentration of the analyte in some samples to the linear range of the calibration curve.

6.1.2.1 Increasing the Pervaporation Efficiency Some ways of increasing mass transfer across the membrane are obvious and can be inferred from the nature of the separation process; others, however, are sophisticated or ingenious approaches. An increase in temperature in the donor solution increases the vapor pressure of the analyte in the donor chamber and boosts mass transfer across the membrane. The usual way of heating this chamber is by plunging it into a water bath; however, when pervaporation is assisted by microwaves focused on the donor chamber, higher temperatures can be reached with closer control.[18] An increase in the sample volume also increases the amount of volatile species in the donor chamber in the stopped-flow mode, or prolongs the time during which the sample passes through the chamber in the continuous dynamic mode, which results in increased amounts of volatiles crossing the membrane. A continuous sample introduction mode is the best approach to increase sensitivity.[6] The best way of increasing sensitivity by avoiding dispersion when the sample is scarce and the continuous sample introduction mode cannot be implemented is using a combined discontinuous/continuous manifold (similar to that used for solid samples, Figure 2b), suppressing the dynamic lower subsystem and introducing the sample into the donor chamber by injection. The pore size of the membrane can also facilitate passage through it, thus improving efficiency. Magnetic stirring facilitates removal of gases from the donor solution, thus accelerating attainment of the separation equilibrium. Ultrasound has a similar but greater effect than magnetic stirring; however, it favors leakage and hence losses of the gas phase. The presence of chemically inert beads of a suitable size in the donor chamber of a dynamic system boosts the transfer of volatile organic compounds (VOCs) to the air gap by delayed passage of the solution through the tortuous way formed by the beads.[6]

One other way of enhancing the efficiency of a pervaporation process is by increasing the pressure in the donor chamber. The additional pressure must not be imposed by a stream of external, inert gas as this creates an overpressure under the donor solution which totally or partially hinders evaporation. The best way of creating overpressure in this chamber and thus improving evaporation is by lengthening the waste line.

This overpressure both forces mass transfer through the membrane and reduces the air gap over the sample, thus decreasing the amount of evaporate analyte necessary for establishment of the equilibrium in the gas phase which results in a faster mass transfer through the membrane. Thus, the efficiency obtained by using a waste tube of 280 cm rather than one of 27 cm (both of the same inner diameter) is 14 times higher.[9]

Halting the flow as a means of improving the pervaporation efficiency can be done in one or both chambers of the separation module. Halting the flow in the donor chamber only increases the efficiency when the sample is injected since equilibrium or a near-equilibrium state is obtained in this way. In the continuous sample introduction mode, stopping the flow impoverishes the sample solution in volatiles and leads to a poorer equilibrium state in the evaporation step. The most effective halting procedure is that involving the acceptor stream as this allows equilibrium or the closest to equilibrium situation to be reached. There are two ways of implementing this step. The first is by stopping the propulsion unit for this stream. This mode makes mandatory the use of separate propulsion units for the lower and upper streams (usually the units can propel four or more channels at once) and disturbs the detector baseline through the start-and-stop sequence. Both shortcomings are avoided by the second mode, which uses an injection valve in whose loop the upper chamber of the separation module is accommodated (Figure 4a). In the filling position, the acceptor solution in the loop remains static while the solution stream passes through the bypass to the detector at the established flow rate. After the preconcentration time has elapsed, the valve is switched to the load position and the stream passed through the loop to drive its contents to the detector. In this way, a situation as close to the mass-transfer equilibrium as required can be achieved without disturbing the detector baseline.

A more efficient way of favoring mass transfer by displacing the equilibrium between the phases involved is by continuous removal of the transferred species from the upper chamber, followed by concentration, either in a minicolumn prior to the detection point or in the flow cell itself. (a) A minicolumn packed with an ion-exchange sorbent can be placed between the pervaporation module and the detector, either in the transport tubing or, better, in the loop of an injection valve (Figure 4b). The latter approach allows the retention and elution steps to take place in opposite directions in the column, thus avoiding a gradual increase in compactness that, in addition, increases pressure in the dynamic system.[20] (b) Retention of the pervaporated species in the detection zone allows the integration of these two steps and hence monitoring the kinetics of the separation process since the volatile species reaches the detector at the same rate as it crosses the membrane (Figure 4c). Obviously, this approach can only be implemented when the detector used is nondestructive. Both approaches allow fresh portions of the acceptor solution to be in contact with the membrane in such a way that equilibrium of the separation process is never reached.

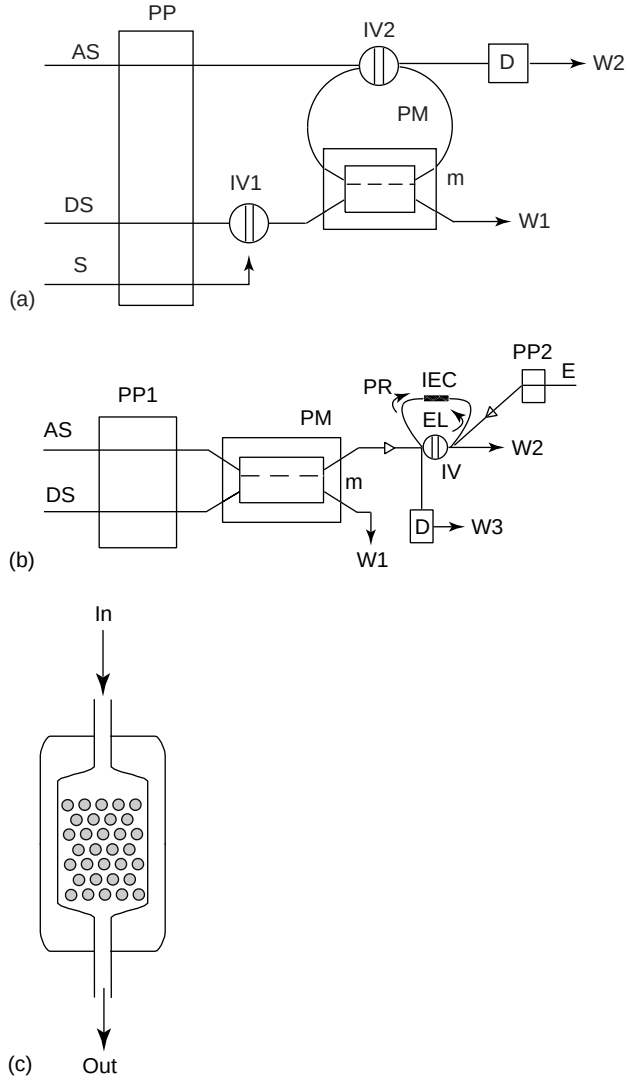

Figure 4 (a) Halting of the flow in the acceptor chamber without disrupting the overall dynamic system. (b) In-line retention of transferred volatile species and elution in the opposite direction. (c) Packed flow cell for integration of reaction and detection (monitoring of the pervaporation kinetics). E = eluent; IEC = ion-exchange column; PR = preconcentration; EL = elution. See previous figures for other abbreviations. (Reproduced by permission of John Wiley and Sons.)

6.1.2.2 Decreasing the Pervaporation Efficiency

When the concentrations of the target analytes in the sample are higher than the upper limit of the linear range

of the calibration curve, a dilution or pseudodilution step is mandatory in order to fit the unknown concentration to this portion of the calibration curve, and hence increase the precision of measurements. In addition to the usual prior dilution of the sample or the weighing of a smaller amount, various strategies can be used to avoid the errors involved in the dilution step and in the weighing of small amounts of a solid. Such strategies include the following: (a) use of a lower pervaporation temperature; (b) changing the chemical conditions to a less favorable situation if a derivatization reaction is involved, in order to reduce the yield of the volatile or monitored species (reaction prior or subsequent to pervaporation, respectively); (c) expanding the air gap between the donor liquid level and the membrane by intercalating as many spacers as required); (d) use of a thicker membrane; (e) use of a smaller loop for the injection valve if the sample is liquid; and (f) use of a higher flow rate of the donor and/or acceptor stream.

6.2 Selectivity

The fact that the pervaporation separation process can only be applied to the determination of volatile analytes could be considered as a shortcoming of the technique but, at the same time, this increases the selectivity of the analyte of interest, and allows its analysis in ''dirty'' samples, such as biological or environmental samples,[32] without any previous treatment, using the pervaporation as a cleanup step. If the analyte of interest is not volatile, but it can be transformed into a volatile compound by means of a selective (bio)chemical reaction, it can also be analyzed by pervaporation. The formation of the volatile analyte is done by the addition of the appropriate reagent to the sample, either continuously or by direct injection into the matrix. In any case, a derivatization reaction is often required for the formation of the species to be finally detected. The possibility of employing the derivatization reaction prior to or after the separation step broadens the application field of the technique to a variety of analytes and, at the same time, permits the coupling of the unit to several types of detectors. Pervaporation can also be applied in speciation analysis, selectively analyzing the different forms of an analyte of interest (see section 7.6).

In the case of multideterminations and when the target volatile analytes have different boiling points, they can be selectively separated by subjecting the unit to different temperatures. In this case, a heating system allowing reproducible control of time and temperature is mandatory.

Discrimination between polar and nonpolar volatile analytes can also be achieved by using microwave irradiation of the lower chamber, so as to favor the evaporation of the polar species. The evaporation of the nonpolar compounds can then be done by conventional heating.

Selectivity can also be increased by changing the membrane used in the separation process. In general, permeability in glassy polymers is lower than in rubbery polymers and selectivity is primarily determined by the molecular dimensions of the permeating species.[31] In recent work, three PTFE membranes with pore sizes of 0.5, 2.0, and 5.0 μm and thicknesses of 0.16, 0.18, and 0.20 mm, respectively, were compared with respect to permeability to hydrogen cyanide. The membrane with a 2.0 μm pore size and 0.18 mm thickness gave the highest permeability. The 5.0 μm pore size was supposed to give the best permeability, but the larger pore size of the membrane only allows better diffusion of larger molecular gases, and not of the small molecular gases as HCN. In such cases other considerations, such as the thickness and the pore density of the membrane, should be taken into account. The thickness is inversely proportional to the permeability whereas the pore density is proportional to it. It is expected that a larger pore size will require a lower pore density to ensure acceptable mechanical stability of the membrane. This means that the overall area of the pores through which gas diffusion takes place could even decrease with increasing pore size and hence give lower permeability.[30]

6.3 Precision

The donor chamber of the unit depicted in Figure 1 was found to give rise to increasingly less reproducible signals with successive injections owing to nonhomogeneous, irreproducible dispersion. In order to circumvent this shortcoming, magnetic stirring and packing the chamber with glass beads were tried. Improved precision, expressed as relative standard deviation (RSD), was obtained in both cases, especially with the latter method (RSD = 11.29% without glass beads or magnetic stirrer, 0.79% with glass beads, and 1.87% with a magnetic stirrer).[6]

Repeatability can also be improved by aspirating the detector waste at a flow rate equal to that of the acceptor stream or, if a chemical reaction with an additional reagent is involved, to the sum of the individual flow rates.[7]

When the concentrations of the analytes to be determined are relatively low, a low precision is obtained, which can be attributed to the fact that equilibrium conditions in the gas phase in the donor chamber are not readily reached. This problem can be overcome by previously saturating the air gap with the volatile species, i.e. by injecting the required volume of a standard solution containing an appropriate volatile concentration and driving it to the lower part of the separation module.

The flows of both the donor and the acceptor stream are then stopped for a certain interval and, once resumed, the experiment can be continued with fairly good results (the precision thus obtained in the determination of sulfide was 5.46% as RSD).[8]

The use of long waste tubes in the lower part of the pervaporation cell can also improve the precision of a method since the pressure in the module is thus increased and small variations of pressure caused by both internal and external noise yield a smaller overall effect. In this way, for injection of the sample, the RSD decreased from 5.46 to 2.86%; for continuous sample introduction, it decreased from 7.03 to 1.61%. Halting the flows of both the acceptor and donor streams resulted in a repeatability of 2.51% at the normal waste tube length (27 cm), and 1.21% for a longer length.[9]

The use of gas-diffusion membranes instead of ultrafiltration membranes can also improve the precision of the method, since these are not easily curved, a fact that could change the flux of the permeate owing to both an irreproducible membrane area and an increased volume of the upper chamber.[12,13]

6.4 Physical Parameters of Analytical Pervaporation: Kinetics of the Mass-transfer Process

Monitoring of the mass-transfer kinetics is of paramount importance in optimization studies in order both to elucidate the factors that influence the process and to establish the appropriate pervaporation time in order to achieve a high sample frequency with minimal detriment to the sensitivity of the method. There are two ways of monitoring the kinetics of the pervaporation process: (a) by integrating pervaporation and detection and (b) by integrating a retention step with detection, after pervaporation.

Continuous monitoring of the mass transfer across the membrane can be accomplished by placing the active surface of a sensor opposite the membrane, as shown in Figure 5(a). In this way, a rising signal is obtained that reflects the enrichment of the static acceptor stream with the pervaporated species. The signal evolves to a plateau when the partitioning equilibrium of the analyte between the donor liquid–air gap–acceptor solution is reached. Electroanalytical probes (potentiometric, voltammetric, or amperometric) and fiber-optic-assisted probes have been used in this context. Optical fibers can also be used for monitoring color changes or luminescence emission.

A flow cell packed with a suitable support (ion exchanger, sorption material, etc.) can be accommodated in a nondestructive detector for continuous retention of the pervaporated species or its reaction product (Figure 5b). In this way, the content of the upper chamber is continuously removed and driven to the detector, and the fresh acceptor stream that reaches the chamber displaces the mass-transfer equilibrium, thus increasing the efficiency of the pervaporation process. A rising recording is obtained, the slope of which evolves to a plateau but never becomes flat since the concentration of volatile species in the lower chamber becomes increasingly lower, but is never zero. After the measurement period, the duration of which depends on the quantitativeness to be achieved, has elapsed, a switching valve (SV) located prior to the detector allows the acceptor stream to be replaced with an eluting solution that removes the retained species and drives them to waste. On switching the valve again, the flow of the acceptor solution through the detector is restarted, the baseline re-established and the system made ready for a fresh sample.

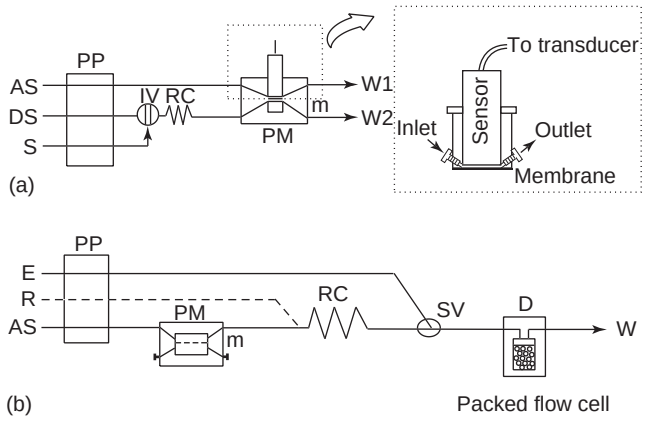

Figure 5 Monitoring of the pervaporation kinetics: (a) by integrating separation and detection; (b) by coupling pervaporation with a retention step integrated with detection. See previous figures for abbreviations. The dashed lines refer to potential derivatization reactions and the dotted line to continuous sampling.

7 SCOPE OF ANALYTICAL CONTINUOUS PERVAPORATION

Table 1 summarizes the applications of analytical continuous pervaporation reported so far. Applications can be classified according with the nature of the sample or whether a derivatization reaction is involved in the overall process. When a reaction is involved, it can take place prior to the separation step in order to convert the target analyte into a volatile compound, or subsequent to the mass-transfer step (in the upper subsystem) in order to obtain a product measurable by the detector. The target reaction can be of chemical or biochemical nature. One other possible classification is according to whether a single or multiple parameters are determined.

Table 1 Applications of pervaporation

Analyte	Type of sample	Derivatization reaction	Detection	Special features	Refs.
Ammonia	Water	Chemical	Conventional photometry	Mass-transfer studies	19
Mercury compounds	Solid (CRM)		Atomic fluorimetry	Microwave-assisted pervaporation, speciation	18
Ethanol	Wine	Chemical	Conventional photometry	Optimization of separation module	6
Ethanol	Wine	Biochemical	Conventional fluorimetry	Immobilized enzyme	21
Ethanol	Wine	Biochemical	Conventional photometry	Immobilized enzyme	35
Ethanol	Bakers' yeast culture, beer	Biochemical	Voltammetry, fluorimetry	Electronically controlled temperature, immobilized enzyme	28
Ethanol	Yeast extract	Biochemical	Fluorimetry	Immobilized enzyme	27
S^{2-}	Kraft liquors	Chemical	Conventional photometry	Improvement of precision and sensitivity	8, 9
Diacetyl	Yeast extract	Biochemical	Fluorimetry	Immobilized enzyme	27
Acetaldehyde	Fermentation media	Biochemical	Conventional photometry	Immobilized enzyme, continuous monitoring	13
F^-	Fertilizers, ceramic industry wastewater	Chemical	Conventional potentiometry	Stopped-flow preconcentration	7
Ammonia/urea	Serum, urine	Biochemical	Conventional potentiometry	Sequential determination, immobilized enzymes	16
Ammonia/urea	Soil	Biochemical	Integrated potentiometry	Sequential determination, speciation	17
VOCs	Soil		ECD	Alternative to headspace	12
Trimethylamine	Fish	Chemical	Conventional photometry	Pretreatment integrated with pervaporation	22
Diacetyl	Beer	Chemical	Conventional photometry	Stopped-flow preconcentration	23
Acetaldehyde	Fruit juices, yoghurt, bread	Chemical	Conventional photometry	Food samples	14
F^-	Tree leaves, fertilizers, ceramic industry wastewater	Chemical	Integrated potentiometry	Hybrid continuous/ discontinuous manifold	10
Mercury	CRM		Atomic fluorimetry	Alternative to headspace, speciation	20
Oxalate	Urine, serum	Biochemical	Conventional potentiometry	Hybrid continuous/discontinuous manifold, immobilized enzyme	15
COD/inorganic carbon	Bleaching liquors, domestic sewage, river water	Chemical	Conventional photometry/integrated potentiometry	Simultaneous determination of COD and inorganic carbon	11
SO_2	Wine	Chemical	Conventional photometry	Speciation of free and bound SO_2	25
Acidity	Wine	Chemical	Conventional photometry	Sequential determination of total and volatile acidity	26
Cyanide	Samples from the mining industry	Chemical	Amperometric	Determination of cyanide in the presence of sulfide, interference study	30
Pectinesterase activity	Fruit and vegetables	Biochemical	Conventional fluorimetry	Through methanol monitoring	36

The availability of multidetermination can be based on the separation process (different temperatures for analytes with different vapor pressures), on a reaction prior or subsequent to pervaporation, on a discriminating separation technique, or on the characteristics of the detector. Whether the separation step is integrated with detection also gives rise to two different types of methods. Finally, other applications of pervaporation such as speciation studies and continuous monitoring of evolving systems, as alternatives to headspace methods, etc. are also commented on below.

7.1 Types of Samples

Two broad categories can be established according to the state of aggregation of the target sample (solid or liquid). As was noted in section 5, both the manifold used and the way of introduction of the sample are different, depending on the type of sample (Figures 2a and b). The most significant difference between the treatment of liquid and solid samples is that with the former, the overall process from sample injection to delivery of the results can be fully automated. In contrast, some human intervention is required in sample weighing and fastening of the pervaporation module for solid samples. Partial automation can be achieved in this case as the rest of the process (derivatization (if required), pervaporation, transport to the detector, monitoring and data acquisition and processing, and delivery of the results) can be performed in an automated way.

7.2 Derivatization Reactions

When a derivatization reaction is developed in any of the subsystems (in the lower one to form a volatile product or in the upper one to adapt the monitored product to the detector characteristics), additional channels must be connected to the main lines in order to ensure optimal development of the target reaction. Solid reagents, the more usual situation in enzyme-catalyzed reactions, are packed in a tube (solid reactor), placed between the injection valve and the separation module when the analyte is nonvolatile in order to obtain a volatile product amenable to separation; however, the solid material containing the biocatalyst can also be placed in the donor chamber for two chief purposes: (a) to increase the yield of the biocatalyzed reaction (as the volatile product is removed from the solution into the air gap and then into the acceptor stream, the reaction equilibrium is gradually displaced to quantitativeness) and (b) to avoid dispersion of the sample in the carrier solution when the analyte concentration is low and the sample scarce (inadequate for implementing the continuous sample-introduction mode). When the purpose of the biocatalyzed reaction is to obtain a product amenable to detection, the reactor, packed with the support–enzyme conjugate, is placed between the acceptor chamber and the detector. The enzyme has also been immobilized in the upper chamber of the pervaporation module, thus integrating separation and enzymatic reaction with immobilized biocatalyst. This method has been applied to the monitoring of ethanol production during wine fermentation processes, providing longer use of the immobilized biocatalyst.[21] Figure 6 illustrates all these possibilities.

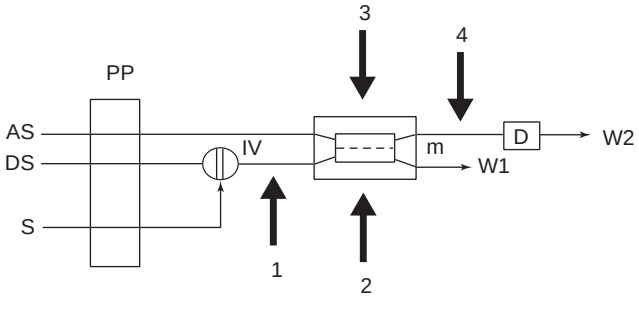

Figure 6 Possible locations of a support–enzyme conjugate in an FI/pervaporation assembly. (1), (2) Formation of a volatile product; (3), (4) formation of a measurable product. In (1) and (4), the conjugate is packed and constitutes an enzyme reactor. In (2), the conjugate is in the bottom of the lower chamber. In (3), the conjugate is retained in the wall of the upper chamber, facing the membrane. See previous figures for abbreviations. (Reproduced by permission of John Wiley and Sons.)

7.3 Individual Determinations

The simplest use of an FI pervaporation approach is for the separation/determination of a single analyte; thus, most of the methods proposed so far involve individual determinations. The determination of acetaldehyde in semisolid and solid food samples is an example of a single analyte determination,[14] in which important sample preparation steps, such as filtration, removal of existing dyestuffs and turbidity from Carrez solution, or centrifugation are avoided. Another example is the determination of trimethylamine (an objective parameter for fish quality evaluation, correlating well with organoleptic estimation).[22] The method is based on pervaporation of the analyte and monitoring the color change of bromothymol blue caused by the basic character of the amine, obtaining results in good agreement with those provided by the official method.

7.4 Multideterminations

Multideterminations can also rely on the manipulation of parts or parameters affecting the overall approach. One of the common ways of determining two analytes is by sequential formation of volatile compounds. Thus,

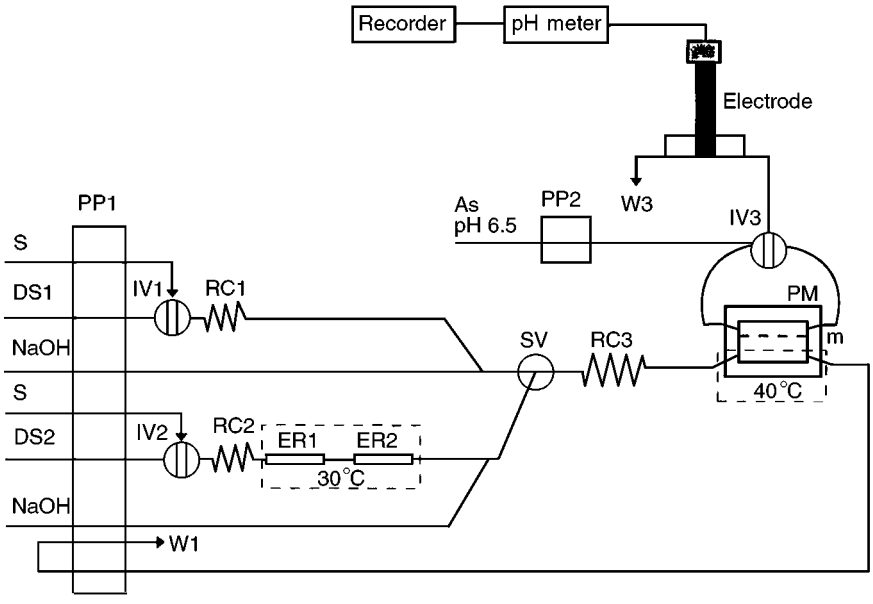

Figure 7 Manifold for the sequential determination of urea and ammonia based on enzymatic reactions, and separation of the target analytes by pervaporation, followed by potentiometric detection. RC = open reactor; ER = enzymatic reactor. See previous figures for other abbreviations. (Reproduced by permission of John Wiley and Sons.)

the determination of ammonium and urea in serum and urine is accomplished in a manifold such as that shown in Figure 7. Ammonium is determined by injecting the sample through valve IV1 into a basic stream that converts the ammonium into ammonia, which is removed from the donor stream and driven to the potentiometric sensor by the acceptor carrier. For the determination of urea, the sample is injected via IV2, passes through a glutamate dehydrogenase reactor (ER1) to remove the endogenous ammonium and then through a urease reactor (ER2) that converts urea into ammonia and carbon dioxide. The latter is transformed into nonvolatile HCO_3^-, while the NH_3 is pervaporated and detected.[16] The same analytes were determined in a sequential speciation analysis of nitrogen in soil. Ammonium is first converted into volatile ammonia, which is removed by pervaporation and detected by an ammonium-ion-selective electrode located on the top of the separation unit (thus, separation and detection are integrated). Urea is converted to ammonia by reaction with dissolved urease and detected in the same way.[17]

A reaction subsequent to pervaporation allows two volatile analytes to be discriminated. Such is the case with CO_2 and SO_2, which are removed by pervaporation under similar working conditions; the contribution of both to the pH change of the acceptor stream is monitored potentiometrically and then the well-known *p*-rosaniline–formaldehyde reaction with SO_2 yields a colored product that is measured photometrically.[24] The concentration of CO_2 is calculated by difference between the results provided by the two detectors.

Change of temperature during the pervaporation process is another way of accomplishing multideterminations. The best way of producing fast, reproducible changes in the temperature of the lower chamber of a pervaporator is by equipping the module with a focused microwave device. In this way, speciation of mercury in a solid matrix (certified reference material (CRM)) is carried out with subsequent monitoring of the target compounds by atomic fluorescence spectroscopy (AFS).[18]

Multidetermination capabilities can be provided by a high discrimination separation technique implemented after the pervaporator. The separation technique can be selected according to both the number of analytes to be determined and their similarities. A GC connected to a pervaporator furnished with a Tenax minicolumn as interface is used for preconcentration of VOCs before their separation and electron capture detection (ECD).[12]

Two serially arranged detectors also permit the determination of two analytes, as noted before for CO_2 and SO_2 (potentiometric–spectrophotometric assembly).[24] A method for the simultaneous determination of chemical oxygen demand (COD) and inorganic carbon by spectrophotometric and potentiometric detection, respectively, has been also proposed, the detectors being arranged in parallel. The carbon dioxide is removed by pervaporation and accepted in the upper part of the separation unit by a buffer stream, the pH change of which is measured by a pH electrode situated on the top of the unit. The donor stream is, at the same time, led to a spectrophotometer for the measurement of COD. The

inorganic carbon is calculated as the difference between the total concentration of CO_2 monitored potentiometrically and the COD monitored photometrically.[11]

7.5 Pervaporation as an Alternative to Headspace Methods

One of the most promising uses of pervaporation is as an alternative to headspace methods. Two methods of this type have so far been proposed for mercury speciation and for VOCs in solid samples; both testify to the advantage of pervaporation over headspace methodology. Both use the overall assembly shown in Figure 8, by which the analytical process is developed in four steps. (1) *Pervaporation of the analytes*. About 0.5 g of sample is placed in the donor chamber of the pervaporation cell. The cell is then closed, connected to the system and placed either in a water bath or in the vessel of a focused microwave device, depending on the polar or nonpolar nature of the target analytes, which evaporate into the gas layer above the sample and then diffuse through the membrane to the argon acceptor stream. (2) *Preconcentration/desorption*. A minicolumn of Tenax, positioned in the loop of an injection valve that is placed in an ice bath, is used to preconcentrate the volatile species that pass through the membrane of the pervaporation unit. In order to desorb the retained species and inject them onto the chromatographic column, the flow of Ar from the pervaporation unit is diverted so that the valve loop is shut off with no gas flow passing through it. The Tenax minicolumn is removed from the ice bath and placed in a muffle oven at a suitable temperature for nonpolar analytes or in a microwave device for polar analytes, following which the valve is switched to the ''inject'' position to allow a He flow to pass through the column and flush the desorbed species from the loop of the injection valve through the connector between the injection valve and the chromatograph, and onto the chromatographic column. (3) *Chromatographic separation*. The He stream containing the desorbed compounds enters the injector in a continuous manner via a hypodermic needle fixed to the injector inlet. This is then mixed with an additional stream that flows through the normal carrier gas inlet into the injector. The chromatograph oven is programmed as required and the species are separated. (4) *Detection of the separated species*. As the individual mercury species emerge from the column, they are passed through a pyrolysis unit set at 800 °C in order to break the compounds down to elemental mercury. This stream is then mixed with a make-up gas on its way to the detector. A second stream of the same gas (Ar) is used as sheath gas, which increases the reproducibility as a result of the He/Ar/Hg flow to be maintained in the beam of the atomic fluorescence detector. VOCs require no pyrolysis unit, so the separated analytes are driven directly to the ECD unit. Hydrophilic membranes could not be used in these studies, as they permitted the passage of water-vapor through them, which, as a result, was led to the GC column. Therefore, PTFE hydrophobic membranes were used, as they provided better sensitivity and higher thermal stability than other hydrophobic membranes, such as the PVDF type.[12]

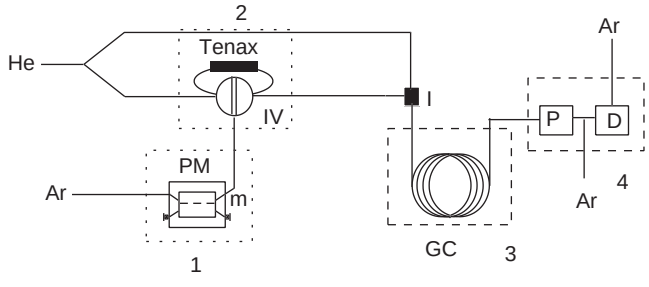

Figure 8 Speciation of mercury in solid samples using pervaporation/GC/atomic fluorescence detection. Tenax = Tenax minicolumn; I = injector; GC = gas chromatograph; P = pyrolysis unit; He = helium; Ar = argon. Sections marked 1, 2, 3, and 4 are used in the pervaporation, retention/desorption, chromatographic separation, and detection steps, respectively. See previous figures for other abbreviations. (Reprinted with permission from I. Papaefstathiou, M.D. Luque de Castro, 'Integrated Pervaporation/Detection: Continuous and Discontinuous Approaches for the Treatment/Determination of Fluoride in Liquid and Solid Samples', *Anal. Chem.*, **67**, 3916–3921.[10] Copyright (1995) American Chemical Society.)

The advantages of pervaporation over headspace methodology are as follows: (a) the thin air-gap layer above the sample uses very small amounts of analytes to establish equilibrium with the solid or liquid phase and mass transfer across the membrane; (b) continuous removal of the volatilized analytes across the membrane from the air gap displaces the mass-transfer equilibrium and increases the efficiency of the separation process; (c) continuous removal of the pervaporated analytes to the preconcentration column allows fresh portions of acceptor gas to come continuously into contact with the diffused species, thus displacing the mass-transfer equilibrium; (d) the separation process can readily be automated for laboratory use; and (e) there is no need for a water-vapor condenser, as in the case of the ''purge and trap'' mode, or a hydrophobic sorbent, since no water crosses through the hydrophobic membrane.

7.6 Speciation Studies

When a pervaporation module is used as a fundamental device in speciation analysis, this goal can be achieved in the following ways.[33]

First, (bio)chemical treatment can be applied as a previous discriminating reaction. Hence, it can favor the

formation of a volatile compound from one of the forms of the analyte of interest, thus permitting the simultaneous determination of volatile and nonvolatile compounds by monitoring both the acceptor and the donor streams, as in the method proposed for the determination of COD and inorganic carbon by spectrophotometric and integrated potentiometric detection, previously commented upon.[11] A similar method has also been applied to the determination of total and free SO_2 in wine. The sulfur dioxide formed in an acidic stream is removed from the liquid sample and collected in an acceptor solution containing *p*-rosaniline and formaldehyde, the colored compound formed being monitored photometrically.[25]

(Bio)chemical treatment can also be applied in samples containing both a volatile and non-volatile form of the analyte, promoting the conversion of the nonvolatile form into the volatile form. This is the case in the sequential determination of ammonia and urea in biological fluids, making use of immobilized enzymes and a pervaporator coupled to an FI manifold,[16] commented upon in a previous section.

In other cases, the analytes to be determined are both involved in (bio)chemical reactions in order to yield volatile compounds. One example is the speciation of mercury in solid samples making use of microwave-assisted pervaporation/atomic fluorescence detection in a continuous-flow system. A focused microwave device has proved highly efficient when the analytes are sufficiently polar to interact with microwave energy [e.g. $Hg(NO_3)$ and PhHgOAc]. As can be seen in Figure 9, the experimental set-up required is very simple. In this manifold, the sample is weighed in the donor chamber of the pervaporation cell and the cell is tightly closed and connected to the loop of the injection valve, which acts as a SV. This allows the carrier stream (water or argon) to pass either through the acceptor (upper) chamber of the pervaporation unit and then into the detector, or via a bypass to the detector while the flow in the pervaporation cell is stopped. Although the flow

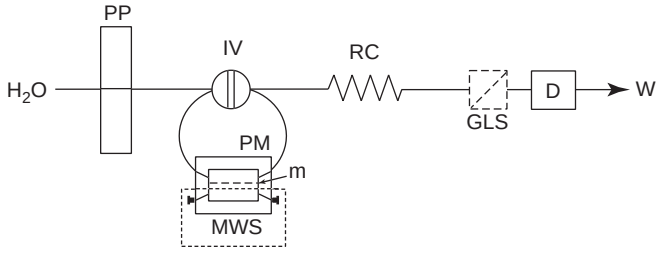

Figure 9 Manifold for the pervaporation of Hg(II). GLS = gas–liquid separator; MWS = microwave system. See previous figures for other abbreviations. (Reprinted from M.D. Luque de Castro, I. Papaefstathiou, 'Analytical Pervaporation: a New Separation Technique', *Trends Anal. Chem.*, **17**, 41–49.[37] Copyright (1997) with permission from Elsevier Science.)

through the pervaporator is stopped, the baseline does not change because the carrier stream proceeds to the gas–liquid separator via the bypass loop at the same rate as it circulates through the pervaporation cell. Once the sample is in the donor chamber, the reagents are injected via one of the inlets of the pervaporation cell. Septa similar to those used in GCs allow the syringe needle to penetrate the donor chamber, inject the reagents and be removed with no loss of either sample or reagents. One minute after the reagent is injected, the microwave device is switched on at a preset power for a given time, after which the cell is allowed to cool for 1 min before the SV is activated to divert the carrier through the acceptor chamber of the cell, collecting one of the mercury compounds to the gas–liquid separator if the acceptor carrier is a liquid, or directly to the atomic fluorescence detector if the carrier is gaseous. When the detector baseline is restored, the valve is switched again to isolate the pervaporator and the microwave digestor is switched on again for another preset time and at a suitable power setting for the other mercury compound. After the pertinent peak has been obtained, the valve is switched to isolate the cell and the system is thus made ready for the next sample.[18]

Second, the temperature of the donor chamber can be changed according to the vapor pressure of the target compounds. Exact control of temperature is then required, as was commented upon previously.

Third, the pervaporator can be connected to high-discrimination separation equipment such as a GC, as described in the previous section.

7.7 Integrated Pervaporation/Detection

The best way of monitoring the kinetics of pervaporation is by inserting a probe-type sensor in the acceptor chamber, with its active surface faced to the membrane. A number of methods based on this principle have been developed. The determination of fluoride in liquid (ceramic industry wastewater, dissolved fertilizers) and solid samples such as tree leaves by formation of a volatile product with hexamethyldisiloxane using manifolds such as that shown in Figure 10 is an excellent example of this approach.[10] The most outstanding advantages derived from integrating pervaporation and detection are as follows: (1) the response time is shortened relative to the conventional location of the sensor after the separation module since transport of the target species from the separation module to the detector is avoided; (2) the kinetics of mass transfer across the membrane can be monitored, which affords a better understanding of the separation step and easier, in-depth optimization; and (3) the set-up required for the determination of a volatile analyte or reaction product in solid samples can be miniaturized by integrating the leaching, derivatization, separation, and detection steps.

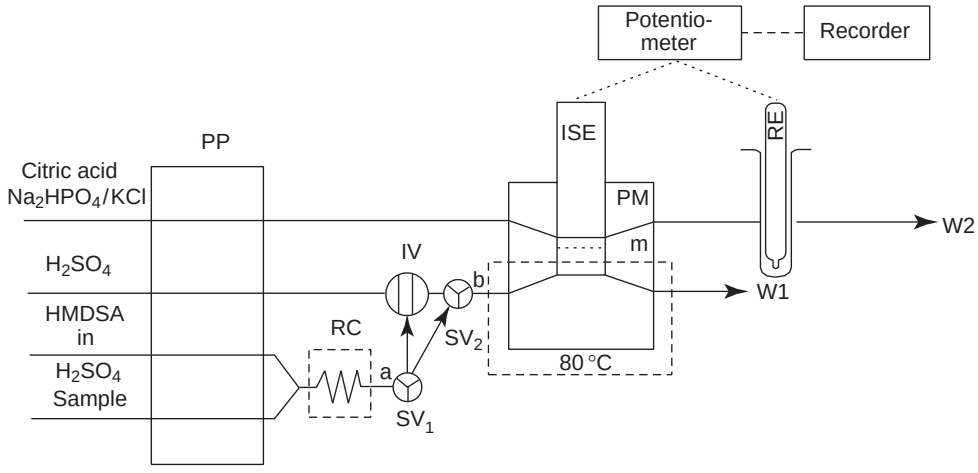

Figure 10 FI/pervaporation/detection arrangement for the determination of fluoride in liquid samples. ISE = ion-selective electrode. The dashed lines indicate thermostated zones. See previous figures for other abbreviations. (Reprinted with permission from D.W. Bryce, A. Izquierdo, M.D. Luque de Castro, 'Pervaporation as an Alternative to Headspace', *Anal. Chem.*, **69**, 844–847.[20] Copyright (1997) American Chemical Society.)

7.8 Types of Detectors

Any type of detector, whether molecular or atomic, can be coupled to a pervaporator through an appropriate interface such as a transport tube, a microcolumn packed with adsorptive or ion-exchange material, or a gas–liquid separator. The acceptor carrier stream can be either liquid or gaseous depending on the characteristics of the detector. The detectors often used in conjunction with pervaporation are of the spectroscopic [atomic (fluorimetric) or molecular (photometric, fluorimetric)], electroanalytical (potentiometric, voltammetric, or amperometric), electron-capture and flame ionization types. The low selectivity of some detection techniques such as photometry, ion-selective electrode-based potentiometry, and flame ionization, is countered by that of the pervaporation step, which endows the overall analytical process with the selectivity required to process highly complex matrices.

7.9 Pervaporation as an Interface Between Evolving Systems and Monitoring

The fact that the presence of solid material, suspended particles, macromolecules, etc. does not disturb the pervaporation process, and hence the passage of the volatile compounds into the acceptor stream is independent of the state of aggregation, is of great interest. Fermentation processes, environmental systems, and, in general, any evolving process that must be continuously or intermittently monitored are potential candidates for implementation of pervaporation-assisted monitoring. The presence of colloidal particles, macromolecules, etc. in fermenters calls for the use of filters of different nature and pore size in order to ensure cleanliness of the sample prior to its insertion into the analyzer. This step dramatically complicates the overall set-up and can even hinder full automation or make it expensive. In addition, a separation module such as a dialyzer or gas diffuser may be necessary for subsequent separation of the target analyte from the rest of the matrix. The use of a pervaporator as an interface between fermenters and monitoring devices (which can be extended to other evolving systems) has proved a suitable means for avoiding the cleanup step (provided the analytes are volatile or can be converted into volatile compounds, which is very often the case in fermentation processes). The only requirement for using a pervaporator as an interface between the evolving system and the monitoring device is that the transport tube and the inlet/outlet of the separation module should be of appropriate diameters, which will be a function of those of the particles in the stream. Stirring of the donor chamber avoids sedimentation in the bottom, which can be facilitated by aspirating the donor waste. When the system evolves slowly and continuous monitoring is unnecessary, a modified injection valve (also of an appropriate loop diameter and sample path) can be a more effective choice. The injection mode is known to improve continuous monitoring in that the baseline is restored after each transient signal.[34] As a result, any changes in the baseline due to detector problems (e.g. fouling of the flow cell), are immediately detected as such; on the other hand, such changes during continuous monitoring are taken to be changes in the evolving system. This fact, which does not pertain to pervaporation, but to continuous-flow systems, must be taken into account in choosing continuous or discontinuous monitoring.

8 PROSPECTS FOR PERVAPORATION

Pervaporation is a useful tool in the analytical laboratory, and is characterized by simplicity, ease of automation, and the possibility of miniaturization. The potential of pervaporation for the pretreatment of liquid or solid samples can be inferred from both the intrinsic features of the technique and existing methods using a pervaporator, two subject matters of this article. The most interesting applications of analytical pervaporation developed so far are as an alternative to gas diffusion, its use with liquid samples containing solids in suspension or of a corrosive nature, its use with solid samples for the determination of a single analyte or multideterminations, integrated pervaporation–detection, and coupling of a pervaporator to a GC, as an alternative to headspace sampling.[37]

Consolidation of this technique as an effective analytical tool relies on two previous, sequential developments: (a) the availability of sufficient research experience on the subject to compile a cook-book of methods where potential users can search for solutions for specific problems and (b) the commercialization of precise, inexpensive pervaporation modules meeting the requirements of a variety of samples and analytes. The development of the methods and the theoretical studies on pervaporation should permit the consolidation of the technique as an effective analytical tool in routine analysis that can be applied in various fields, including environmental, clinical, food, and industrial chemistry.

ABBREVIATIONS AND ACRONYMS

AFS	Atomic Fluorescence Spectroscopy
AIV	Auxiliary Injection Valve
COD	Chemical Oxygen Demand
CRM	Certified Reference Material
ECD	Electron Capture Detection
FI	Flow-injection
GC	Gas Chromatograph
MIV	Main Injection Valve
PDMS	Poly(dimethylsiloxane)
PTFE	Polytetrafluoroethylene
PVDF	Hydrophobic Poly(vinylidene fluoride)
RSD	Relative Standard Deviation
SV	Switching Valve
VOCs	Volatile Organic Compounds

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Flow-injection Techniques in Environmental Analysis

Food **(Volume 5)**
Sample Preparation, Headspace Techniques

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis

Atomic Spectroscopy **(Volume 11)**
Flow Injection Analysis Techniques in Atomic Spectroscopy

Gas Chromatography **(Volume 12)**
Sample Preparation for Gas Chromatography

REFERENCES

1. M. Valcárcel, M.D. Luque de Castro, *Nonchromatographic Continuous Separation Techniques*, Royal Society of Chemistry, Cambridge, 1991.
2. R.Y.M. Huang (ed.), *Pervaporation Membrane Separation Processes*, Elsevier, Amsterdam, 1991.
3. W. Nell, *Pervaporation*, Tec & Doc Lavoisier, Cachan, 1998.
4. M.D. Luque de Castro, I. Papaefstathiou, *Encyclopedia of Environmental Analysis and Remediation*, John Wiley & Sons, Chichester, 3462–3476, 1998.
5. P.J. Hickey, C.S. Slater, 'The Selective Recovery of Alcohols from Fermentation Broths by Pervaporation', *Sep. Purif. Methods*, **19**(1), 93–115 (1990).
6. I.L. Mattos, M.D. Luque de Castro, M. Valcárcel, 'Pervaporation: an Integrated Evaporation/Gas-diffusion Approach to Analytical Continuous Separation Techniques', *Talanta*, **42**(5), 755–763 (1995).
7. I. Papaefstathiou, M.T. Tena, M.D. Luque de Castro, 'On-line Pervaporation Separation Process for the Potentiometric Determination of Fluoride in ''Dirty'' Samples', *Anal. Chim. Acta*, **308**, 246–252 (1995).
8. I. Papaefstathiou, M.D. Luque de Castro, M. Valcárcel, 'Flow Injection/Pervaporation Coupling for the Determination of Sulphide in Kraft Liquors', *Fresenius' J. Anal. Chem.*, **354**, 442–446 (1996).
9. I. Papaefstathiou, M.D. Luque de Castro, 'Approaches for Improving the Precision and Sensitivity of Continuous Pervaporation Processes', *Anal. Lett.*, **28**(11), 2063–2076 (1995).
10. I. Papaefstathiou, M.D. Luque de Castro, 'Integrated Pervaporation/Detection: Continuous and Discontinuous Approaches for Treatment/Determination of Fluoride in Liquid and Solid Samples', *Anal. Chem.*, **67**(21), 3916–3921 (1995).
11. I. Papaefstathiou, M.D. Luque de Castro, 'Simultaneous Determination of Chemical Oxygen Demand/Inorganic Carbon by Flow Injection–Pervaporation', *Int. J. Environ. Anal. Chem.*, **66**, 107–117 (1997).

12. I. Papaefstathiou, M.D. Luque de Castro, 'Hyphenated Pervaporation–Solid-phase Preconcentration–Gas Chromatography for the Determination of Volatile Organic Compounds in Solid Samples', *J. Chromatogr.*, **779**, 352–359 (1997).
13. I. Papaefstathiou, U. Bilitewski, M.D. Luque de Castro, 'Pervaporation: an Interface between Fermentors and Monitoring', *Anal. Chim. Acta*, **330**, 265–272 (1996).
14. I. Papaefstathiou, U. Bilitewski, M.D. Luque de Castro, 'Determination of Acetaldehyde in Liquid, Solid and Semi-solid Food after Pervaporation–Derivatization', *Fresenius' J. Anal. Chem.*, **357**, 1168–1173 (1997).
15. P. Cañizares, M.D. Luque de Castro, 'Enzymatic Interference-free Assay for Oxalate in Urine', *Fresenius' J. Anal. Chem.*, **357**, 777–781 (1997).
16. P. Cañizares, M.D. Luque de Castro, 'Determination of Urea and Ammonia in Biological Fluids Based on the Use of Enzymes–Pervaporation–Potentiometry', *Quími. Anal.*, **15**, 239–244 (1996).
17. I. Papaefstathiou, M.D. Luque de Castro, 'Nitrogen Speciation Analysis in Solid Samples by Integrated Pervaporation and Detection', *Anal. Chim. Acta*, **354**, 135–142 (1997).
18. D. Bryce, A. Izquierdo, M.D. Luque de Castro, 'Continuous Microwave Assisted Pervaporation/Atomic Fluorescence Detection: an Approach for Speciation in Solid Samples', *Anal. Chim. Acta*, **324**, 69–75 (1996).
19. I.L. Mattos, M.D. Luque de Castro, 'Study of Mass-transfer Efficiency in Pervaporation Processes', *Anal. Chim. Acta*, **298**, 159–165 (1994).
20. D.W. Bryce, A. Izquierdo, M.D. Luque de Castro, 'Pervaporation as an Alternative to Headspace', *Anal. Chem.*, **69**(5), 844–847 (1997).
21. F. Delgado-Reyes, I. Papaefstathiou, J.M. Fernández Romero, M.D. Luque de Castro, 'Monitoring Ethanol Production During Wine Fermentation Processes by a Pervaporation–Enzymic Derivatisation Approach', *Analyst*, **123**, 2367–2372 (1998).
22. J.A. Garcìa-Garrido, M.D. Luque de Castro, 'Determination of Trimethylamine in Fish by Pervaporation and Photometric Detection', *Analyst*, **122**, 663–666 (1997).
23. J.M. Izquierdo-Ferrero, J.M. Fernández-Romero, M.D. Luque de Castro, 'On-line Flow Injection–Pervaporation of Beer Samples for the Determination of Diacetyl', *Analyst*, **122**, 119–122 (1997).
24. P. Linares, M.D. Luque de Castro, M. Valcárcel, 'Simultaneous Determination of Carbon Dioxide and Sulfur Dioxide in Wine by Gas-diffusion/Flow Injection Analysis', *Anal. Chim. Acta*, **225**, 443–448 (1989).
25. E. Mataix, M.D. Luque de Castro, 'Determination of Total and Free Sulfur Dioxide in Wine by Pervaporation–Flow Injection', *Analyst*, **123**, 1547–1549 (1998).
26. E. Mataix, M.D. Luque de Castro, 'Sequential Determination of Total and Volatile Acidity in Wines Based on a Flow Injection–Pervaporation Approach', *Anal. Chim. Acta*, **381**, 23–28 (1999).
27. U. Prinzing, I. Ogbomo, C. Lehn, H.L. Schmidt, 'Fermentation Control with Biosensors in Flow Injection Systems: Problems and Progress', *Sens. Actuators B*, **1**, 542–545 (1990).
28. I. Ogbomo, A. Steffl, W. Schumann, U. Prinzing, H.L. Schmidt, 'On-line Determination of Ethanol in Bioprocesses Based on Sample Extraction by Continuous Pervaporation', *J. Biotechnol.*, **31**, 317–325 (1993).
29. C. Cámara, R. Muñoz, personal communication.
30. H. Sulistyarti, T.J. Cardwell, M.D. Luque de Castro, S.D. Kolev, 'On-line Determination of Cyanide in the Presence of Sulfide by Flow Injection with Pervaporation', *Anal. Chim. Acta*, **390**, 133–139 (1999).
31. C.M. Bell, F.J. Gerner, H. Strathmann, 'Selection of Polymers for Pervaporation Membranes', *J. Membr. Sci.*, **36**, 315–329 (1988).
32. M.D. Luque de Castro, I. Papaefstathiou, 'Pervaporation: a Technique with a Promising Future in Environmental Analysis', *Rev. Anal. Chem.*, **17**, 191–205 (1998).
33. M.D. Luque de Castro, I. Papaefstathiou, 'Pervaporation: a Useful Tool for Speciation Analysis', *Spectrochimica Acta, Part B*, **53**, 311–319 (1998).
34. M.D. Luque de Castro, M. Valcárcel, 'Calibration in Process Monitoring by Using Unsegmented Continuous-flow Systems', *J. Autom. Chem.*, **11**(6), 260–265 (1989).
35. V. Fernández-Pérez, I. Papaefstathiou, M.D. Luque de Castro, Unpublished work.
36. F. Delgado-Reyes, J.M. Fernández-Romero, M.D. Luque de Castro, Unpublished work.
37. M.D. Luque de Castro, I. Papaefstathiou, 'Analytical Pervaporation: a New Separation Technique', *Trends Anal. Chem.*, **17**(1), 41–49 (1997).

Phenols Analysis in Environmental Samples

Monika Moeder
Centre for Environmental Research Leipzig/Halle, Germany

1 Introduction **1**
1.1 Occurrence of Phenolic Compounds and their Penetration of the Environment 1
1.2 Properties and Toxicological Importance of Selected Phenols 2
2 Regulations **3**
2.1 General 3
2.2 Protection of Water Resources by National Programs 4
2.3 Regulations Regarding the Determination of Phenols in Solid Material 5
2.4 Recommendations for Air Monitoring 6
2.5 Current Analytical Methodologies 6
3 Determination of Phenols as a Group Parameter **7**
4 Determination of Phenols **8**
4.1 Preparation of Aqueous Samples 8
4.2 Preparation of Solid and Gaseous Samples 12
5 Instrumental Techniques **15**
5.1 Gas Chromatography 15
5.2 High-performance Liquid Chromatography 15
5.3 Capillary Zone Electrophoresis 16
6 New Developments **17**
6.1 Sensors 17
6.2 Immunoassay 18
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Phenolic compounds are widely distributed in our ecosystems. In addition to naturally occurring phenols, the majority of toxic phenolic substances are released into the environment by industry and agriculture. As the manufacture and use of phenols requires qualitative and quantitative control, a wide variety of methods have been developed to determine phenolic compounds. In addition to coarse screening to rapidly obtain a general overview, many highly specialized techniques exist that are geared to individual analysis problems. The range of available methods extends from steam distillation and classic liquid–liquid extraction (LLE) as sample preparation combined with photometric or ultraviolet (UV) detection, to more sophisticated techniques such as liquid–solid extraction (LSE) including solid-phase extraction (SPE) and solid-phase microextraction (SPME) coupled to gas chromatography/mass spectrometry (GC/MS) and high-performance liquid chromatography/mass spectrometry (HPLC/MS). This article describes the most suitable methods, taking account of special factors such as substance class, chemical properties, concentration range and accompanying matrices. Some new trends in phenol determination are also discussed in comparison with the established methods. The basic factors governing the selection of suitable analysis methods are also covered.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The class of hydroxylated aromatic compounds known as the phenols is widely distributed in the natural environment. A great variety of phenolic compounds are also released into the environment as a result of human activity. In particular, chlorinated, long-chain alkylated, nitrated and sulfonated phenols contribute to the dangerous pollution of surface water and groundwater.

Owing to their toxicological effects and their partial persistence in both the environment and in biological systems, contamination by phenolics must be subjected to environmental monitoring. As a result of their wide range of physicochemical and biological properties, phenolic substances are ubiquitously distributed. Being naturally occurring organic substances, efficient biological degradation pathways exist for phenols.[1] For example, phenol oxidases convert phenols to corresponding chinones, which act as radical traps. Furthermore, phenols are typical metabolites of oxidative transformation processes of aromatics. Their high reactivity enables the easy formation of, for example, nitro-, halogeno-, and sulfonic phenols, and also supports rapid degradation reactions.

1.1 Occurrence of Phenolic Compounds and their Penetration of the Environment

Phenols are widespread in nature as constituents of plant dyestuffs, tanning agents, lignin, flavonoids, etheric oils, substances used as fragrance additives and flavorings, steroids, alkaloids, and natural antibiotics. Phenol is also

produced by normal human metabolism from tyrosines and applied pharmaceuticals via areneoxides.

The bactericidal properties of phenol make it a useful substance in natural defence mechanisms for biological systems, resulting in its application as a disinfectant. In 1865, a 5% phenol solution (also known as carbol or carbolic acid) was introduced as a disinfectant in surgery, and for a long time remained virtually the only antiseptic available. However, these properties are no longer of such interest, and its usage as a disinfectant has declined. Nevertheless, the use of phenol in cosmetics is still permitted when the relevant restrictions are observed.[2]

Nowadays, phenol is much more attractive as an educt for technical products such as phenol-based resins, ε-caprolactam, bisphenol A, adipic acid, alkylated phenols, aniline, chlorophenols, plasticizers and antioxidants.

Phenol used to be produced from coal tar, a ton of which contains around 250 g of phenol as a decomposition product of fossilized oxygen-containing plant constituents. Hence, much of today's environmental pollution attributed to phenolic compounds stems from recent manufacturing processes involving lignite and derived products. Every year thousands of tons of phenol alone (5174 kt in 1985)[3] are produced by splitting cumolhydroperoxide.

Environmental problems caused by phenolic compounds are attributable to their long-term, large-scale uncontrolled release, as well as their persistence and, above all, their toxicity (especially halogenated phenols).

The toxicities of individual phenols vary depending on the type, number and position of the substituents. They can be estimated via their lipophilicity, which can be determined using reversed-phase chromatography.

In human exposure studies, phenol itself is rapidly absorbed into the body via ingestion, skin absorption or inhalation, and rapidly excreted.[4]

Exposure to phenol vapor may cause severe irritation, neurosis, affect the central nervous system, and damage the liver and kidneys. However, there is a lack of exposure data on chronic effects in humans. As evidence of the carcinogenicity of phenol in humans is inadequate, it is not classifiable in this regard.[5]

Even a few types of halogenosubstituted phenols occur naturally, as the isolation of bromophenol from marine organisms has demonstrated. However, the majority of halogenophenols are of anthropogenic origin, such as from wood processing and preservation plants, kraft pulp mills, pesticide manufacturing, and formulation plants. A typical representative is pentachlorophenol (PCP), which is used as an insecticide and a fungicide to protect timber from fungal rot and wood-boring insects, and which is also applied as a biocide in industrial water systems. The by-products of technical grade PCP, such as polychlorodibenzo-*p*-dioxins and hexachlorobenzene, intensify the environmental pollution problems many times over.

The discovery of these highly toxic by-products was yet another reason for the use of PCP being phased out. Other reasons include the toxicological and ecological effects of PCP.

1.2 Properties and Toxicological Importance of Selected Phenols

PCP is rapidly absorbed through the gastrointestinal tract following ingestion. It is known to be extremely toxic to humans because it affects the circulatory system and is accompanied by heart failure. Accumulation can occur in the liver, kidneys, plasma protein, brain, spleen, and fat.[6] It is eliminated and excreted unaltered or in conjugated form via the urine. The United States Environmental Protection Agency (USEPA) estimates that the consumption of $0.03\ \mathrm{mg\,kg^{-1}}$ per day (reference dose) of PCP over a lifetime is unlikely to result in the occurrence of chronic effects.[7]

PCP may be highly toxic to plants[8] and to many species of fish; for example, $LD_{50} = 52\ \mu\mathrm{g\,L^{-1}}$ for rainbow trout. The high bioaccumulation factors caused by the lipophilicity of PCP lead to extremely high levels in several species of fish, invertebrates and algae that are significantly higher (up to 10 000 times) than the concentration in the surrounding waters.[6]

In soil, PCP is moderately persistent with a reported field half-life of 45 days.[9] The sorption of PCP is strongly dependent on the pH conditions of the surrounding medium and increases with higher proportions of organic matter. Under neutral or alkaline conditions, PCP is poorly adsorbed and is mobile in many soils.[9]

In the aquatic environment, PCP is mainly bound to sediments and suspended particles,[10] often hampering analytical determination.

Some phenols are used for munitions manufacturing, handling and disposal, causing the uncontrolled input of (initially) nitrophenols into the soil. During their transformation into the corresponding aminophenols, the toxicity changes (mostly carcinogenically) and, under unfavorable conditions, the contaminants may reach the groundwater.

Another area of research is the environmental behavior of nitrated and aminophenols, which is characterized by various negative effects on organisms. For instance, 2,4-dinitrophenol is used in the manufacture of dyes, wood preservatives, insecticides,[11] explosives, herbicides, photographic developers, picric acid and picramic acid. It is poisonous to humans by ingestion, inhalation and skin contact. It is also an experimental teratogen and mutagen, has phytotoxic effects, and harbors the danger of cumulative effects.[12] Nitrated phenols and their metabolites

(partially amino-substituted) enter the environment via soil leaching. The risk of groundwater contamination by nitrophenols is especially high in the vicinity of munitions manufacturing and dumps.[13]

Phenolic compounds such as bisphenol A (80-05-7), long-chain alkylphenol ethoxylates, and derived phenols are known to be endocrine-disrupting xenobiotics (Scheme 1).

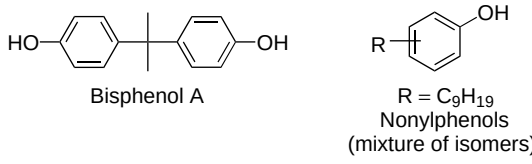

Scheme 1 Chemical structures of endocrine-disrupting phenolic compounds bisphenol A and isomeric nonylphenols.

Bisphenol A is more relevant as a food contaminant because it is used as an additive in polycarbonate plastics and epoxy resins, both of which are used in food and drink packaging, as well as in lacquers to coat the inside of food cans and water supply pipes. Bisphenol A exhibits hormonal effects at a concentration level as low as $2-5\,\mu g\,L^{-1}$.[14] Although human exposure is more probable via the food chain, uptake from drinking water cannot be ruled out, especially if water supply devices are coated by plastics. Some greater environmental human health risks may arise from alkylphenols present in surface, ground- and drinking water.

Long-chain alkylated phenols, such as nonyl- and octylphenols, are produced and released into the environment, whereas the degradation of alkylphenol ethoxylates and tris(nonylphenol phosphites) used predominantly as industrial detergents or surfactants and as antioxidants in plastics results in the release of nonylphenols. Investigations into the environmental behavior of these industrial products indicates only the partial degradation of these compounds during their passage through sewage treatment plants. The resulting metabolites (including the free phenolics) enter the aquatic environment and are persistent in rivers and adjoining groundwater, soil and sediments.[15] High bioconcentration factors have led to the contamination of fish and birds between 10 and several thousand times higher than in the surrounding environment. Diverse nonylphenol concentrations (10 and $1500\,\mu g\,L^{-1}$) in river water have been published.[16] Drinking water analysis in the USA has indicated a concentration of about $1\,\mu g\,L^{-1}$ total nonylphenol-derived compounds.[17] Environmental interest in this class of compounds arises from its direct toxicity to organisms and the endocrine-disrupting effects determined for the nonylphenols. Biotests demonstrated them to be 10 times more toxic to fish and daphnia than the original alkylphenol ethoxylates.

Although the estrogenic (estrogen-mimicking) effects of the alkylphenols were first described in 1938, it was only in 1991 that the effects of nonylphenols on cultured human breast cells finally led to concern over health risks. Tissue concentration levels of just 0.1–1 μmol octyl- and nonylphenol affect fetal steroid synthesis and may cause masculinization,[18–20] according to experiments with adult male rainbow trout. After exposure to $30\,\mu g\,L^{-1}$ octyl- or nonylphenol, the fish began to produce the female egg yolk protein vitellogenin. The exposure concentration used in these experiments corresponded to the amounts found in river water polluted by industrial effluents.

When the environmental concentration of alkylphenols, their high bioaccumulation factors, and their estrogenic effect levels are taken into account, it seems likely that the current amounts of alkylphenols may affect the endocrine system of some organisms. The extent to which these results can be transferred to humans is not yet clear because these effects have only been the target of scientific interest over the past few years. The data available concerning occurrence, environmental accumulation and human exposure are still few (compared for example with polyaromatic hydrocarbons), a situation exacerbated by the lack of official regulations and current analytical standard protocols for the environmental monitoring of alkylphenols. The initial consequences of the studies have been the partial or complete ban of the use of alkylphenols as cleaning agents, although they are still used in paints, pesticide formulations, and lubricants.[21] The phase-out of mainly nonylphenols as planned by the Paris Commission is to be carried out in stages and should be complete by the end of 2004.

2 REGULATIONS

2.1 General

As population growth, urbanization and industrialization continue, the demand for water will increase further, with at least a partial depletion resulting from the nonreversible conversion of pure water to wastewater. Directives, regulations and laws for water management are therefore needed to preserve the quality of our limited water resources.

These regulations depend on the classification of waters. Although guidelines for surface water control contain priority lists of dangerous substances to be monitored and eliminated from surface waters, they do not set any maximum concentration limits. Directives concerning drinking water and surface water used for drinking purposes that establish water quality by recommended or maximum levels are more restrictive.

Since the late 1970s, increasingly stringent regulations have been established worldwide governing the production, transport, application and discharge of toxic substances, as well as the equipment at industrial and wastewater treatment plants. These regulations and the more intensive control of surface water quality have brought about a general reduction in pollution.

To preserve and maintain this encouraging development, governments are endeavoring to follow national and international trends concerning the production, usage and discharge into the environment of hazardous substances, and to revise and adapt existing regulations to actual environmental situations.

For example, in 1996 the Safe Drinking Water Act Amendment was introduced in the USA. Its purpose is to bring about substantial changes to the US national drinking water program, and to provide greater protection and public information. Several regulatory actions are included such as the Disinfectant/Disinfection By-products Rule and the Interim Enhanced Surface Water Treatment Rule, both of which address the discharge and environmental distribution of pathogens and disinfectants including chlorinated phenols. Streamlining of the environmental and health protection program is proposed in the near future, to improve its effectiveness under reduced technical and administrative conditions.

Phenols and selected alkylated, nitrated, and chlorinated phenols are included in various national lists of toxic chemicals[22] as well as in programs that control dangerous substances discharged into the environment (e.g. USEPA in the USA, Municipal/Industrial Strategy for Abatement (MISA) Program in Canada, German National Standards Organization (DIN) in Germany, and the European Community (EC) directives).

2.2 Protection of Water Resources by National Programs

The current USEPA regulations are based on one of the major environmental laws of the USA, the 1977 Clean Water Act, which regulates the discharge of pollutants into water. Certain phenolics were included on the USEPA list of priority pollutants, chiefly due to their toxic properties and environmental persistence.[23,24]

In every country, specific priority lists and state guidelines exist depending on the particular industrial or agricultural substructures. For example, the Canadian Ministry of Environment has serialized the analytical monitoring of nine industrial sectors discharging dangerous material into the environment. Within the framework of the MISA, the guidelines force the control of a total of 104 volatile and semivolatile organics, including the extractable phenols (MISA group 20). The attribute "extractable" indicates an LLE procedure for sample preparation.

Slightly different limits for relevant substances are often encountered when different national guidelines are compared. Within the European framework, the input of environmentally relevant substances into aquatic systems is an essential part of EC Directive 76/464/EEC. This guideline lists dangerous and persistent compounds that have to be controlled, and provides quality parameters for drinking water, surface water used for drinking purposes and secondary waters (bathing, sea and river water). The more detailed directive 75/440 lays down the maximum permissible content of total phenols (determined by steam distillation) in surface water used for drinking water production. This group-specific total concentration of phenolics should not exceed 1 or $10\,\mu g\,L^{-1}$, depending on the water treatment techniques used. The regulation does not stipulate the determination of individual phenolics. However, judging by today's criteria, this requirement seems to be an inadequate description of the environmental situation. The latest findings concerning the specific toxicological properties of isomeric phenolics and semivolatile phenolics, such as nonylphenols or bisphenol A, demonstrate that specific trace analysis for some individual compounds is needed.

By contrast, the World Health Organization (WHO) gives quite exact recommendations regarding, for example, the presence of PCP in drinking water, where it should not exceed a concentration of $9\,\mu g\,L^{-1}$.[25]

The priority pollution list of The Netherlands, updated in 1994, also laid down recommended and maximum permissible values for phenolic compounds (mostly as sum parameters). This provides a useful tool for the risk assessment of soil and groundwater contamination by toxic chemicals, although its general significance as a point of reference is waning with the introduction and increasing acceptance of EC regulations. A comprehensive review of the EC regulations concerning surface water control is given by Hennion et al.[26] At present, although globally valid regulations such as the International Standardization Organization (ISO) norms are available, they mostly concern analytical procedures such as the determination of monovalent phenols for water quality monitoring (ISO-8165-1)[27] by gas chromatography (GC). There are no general global regulatory restrictions limiting the concentration levels of the discharge of phenols into the environment. All we have are recommendations based on scientifically sound toxicological research results.

It is highly likely that most of the phenols enter the environment via waterways in the form of industrial or municipal wastewater. Hence, primary interest must be devoted to the occurrence of phenolic substances in aqueous systems. The standard methods currently used

were therefore developed to determine phenols in various waters.

For example, USEPA method 604[28] has been laid down for municipal and wastewater control, whereas USEPA-625[23] addresses phenols such as polycyclic aromatics in wastewater. More comprehensively, the USEPA series 8000 deals with methodologies for identifying and quantifying priority organic pollution contained in the USEPA lists (Appendix VIII and IX) in groundwater, wastewater, and solid material deposited or discharged. In addition to the USEPA method 604, the list of phenols to be determined by USEPA 8040[29] also includes five isomeric trichlorophenols, two more tetrachlorophenols, 2,6-dichlorophenol, all cresols, and dinoseb (2-*sec*-butyl-4,6-dinitrophenol). The USEPA method 8041[24] includes additional phenolic substances and pesticides in line with the latest research findings regarding the toxicity and environmental behavior of phenol derivatives.

Parallel to this, the USEPA Contract Laboratory Program (CLP) features analytical methods to determine compounds of the target compound list (TCL) in groundwater, sediments and soil from refuse dumps. The protocols developed by the Environmental Monitoring Systems Laboratory in Las Vegas are based on the Superfund Amendments Reauthorization Act (February 1988 statement of work). The target analyte list of semivolatiles contains six selected phenols to be determined (Table 1).

Reference compounds for method validation and calibration are available from various manufacturers as individual components or as mixtures prepared in accordance with the corresponding target lists and official determination methods.

2.3 Regulations Regarding the Determination of Phenols in Solid Material

Methods describing the determination of phenolic compounds in soils, sediments and sludge are often derived from or combined with phenol determination procedures used for wastewater. The main difference is to be observed in the sample extraction procedures where

Table 1 Priority phenols that have to be monitored by different national environmental programs

Target analyte	CAS No.	USEPA methods[a]					Canada[a] GW MISA Group 20	EC[a] GW Directive No. 75/440
		WW 604 LOD ($\mu g\,L^{-1}$)	WW 625 LOD ($\mu g\,L^{-1}$)	WW, S 8040 LOD ($\mu g\,L^{-1}$)	GW CLP	A TO-8		
Phenol	108-95-2	0.14	1.5	2.2	+	+	+	+
2-Methylphenol	95-48-7			+		+	+	+
3-Methylphenol	108-39-4			+		+		+
4-Methylphenol	106-44-5			+		+	+	+
2,4-Dimethylphenol	105-67-9	0.32	2.7	0.63	+		+	
2-Chlorophenol	95-57-8	0.31	3.3	0.58	+		+	
2,4-Dichlorophenol	120-83-2	0.39	2.7	0.68	+		+	
2,6-Dichlorophenol	87-65-0			+			+	
4-Chloro-2-methylphenol	1570-64-5			1.8				
4-Chloro-3-methylphenol	59-50-7	+	3.0	+	+		+	+
2,3,4-Trichlorophenol	15950-66-0	+		+			+	+
2,3,5-Trichlorophenol	933-78-8			+			+	+
2,3,6-Trichlorophenol	933-75-5			+				+
2,4,5-Trichlorophenol	95-95-4			+			+	+
2,4,6-Trichlorophenol	88-06-2	0.64		0.58	+		+	+
3,4,5-Trichlorophenol	609-19-8			+				
2,3,4,5-Tetrachlorophenol	4901-51-3			+			+	
2,3,4,6-Tetrachlorophenol	56818-02-1			+			+	
2,3,5,6-Tetrachlorophenol	935-95-5			+			+	
PCP	87-86-5	7.4		0.59	+		+	+
2-Nitrophenol	88-75-5	0.45	3.6	0.77	+			
4-Nitrophenol	100-02-7	2.8		0.7	+		+	
2,4-Dinitrophenol	51-28-5	13.0		+	+		+	
2-Methyl-4,6-dinitrophenol	534-52-1	16.0		+	+		+	
2-*sec*-Butyl-4,6-dinitrophenol	88-85-7			+				
N-Nitrosodi-*n*-propylamine	621-64-7				+			
Hexachlorocyclopentadiene	77-47-4				+			

[a] WW = wastewater; GW = groundwater; S = soil, sediment, sludge; A = air; + = monitoring required.
LOD, limit of detection.

LSE methods such as Soxhlet (USEPA 3540 and 3545[30]) and sonication (USEPA 3550) are preferred for standardization. The subsequent steps of clean-up, analyte concentration, and analysis are comparable to those applied for wastewater. The problems that may appear in the analysis of suspended phenolic compounds are discussed later. Approved methods that provide conditions for phenol determination from wastewater as well as from soil and sludge include USEPA 8120[29,31] and USEPA 8040,[32] which use the sample extraction techniques mentioned above. The achievable LOD closely depends on the sample matrix present and the method of sample preparation used. The derived practical quantification limits are geared towards the LOD values of the reference method corrected by a theoretical factor that characterizes matrix effects. For example, all LOD values listed in Table 1 (column USEPA 8040) have to be multiplied by a factor of 10 for the wastewater matrix or 10 000 for high-level soil and sludge extracted by sonication.[33] Nevertheless, the corrected values are not always attainable if the matrix is too complex.

2.4 Recommendations for Air Monitoring

The control of air pollution by phenolic compounds is mostly limited to the working environment and indoor air monitoring, along with emissions monitoring of the volatile phenols that may, for example, be discharged during the manufacture of pesticides and phenol resins or released from furniture or wood treated with preservatives.

Methods for air monitoring (including workplace and indoor air measurements of dangerous compounds) are defined in NIOSH (National Institute for Occupational Safety and Health) and OSHA (Occupational Safety and Health Administration) protocols.

The phenols and cresols present in indoor air are determined using NIOSH 2001/OSHA 32 methods. The pollution of ambient air is monitored using the USEPA method for toxic compounds TO-8,[33] which samples the analytes by impinger traps (72 L at a rate of 100 mL min^{-1} for 12 h) and analyzes the phenol- and cresol-containing extracts by high-performance liquid chromatography (HPLC)/UV or electrochemical detection.

The USEPA method TO-17 determines phenol from smaller volumes (1–4 L, 16–67 mL min^{-1}) using a shorter sampling time (1 h) and thermal desorption/GC/MS for determination.[34] The multisorbent tubes are packed with, for instance, Carbopack™ graphitized carbon black and Carboxen™-1000 type carbon molecular sieve. Other sorbent packages have been proposed for VOC (volatile organic compound) enrichment, each combined with optimum gas flows and temperatures for sorbent conditioning, air sampling, and for thermal desorption/GC/MS.

Phenylphenol, used as a germicide and fungicide on foods, has to be detected in accordance with the USEPA TO-10A method,[35,36] which uses a polyurethane filter cartridge for sampling and HPLC for the analysis of this special air pollutant.

The USEPA standard methods are often used as global reference protocols and have been widely accepted in many countries for environmental analysis, allowing the global exchange of data between laboratories.

2.5 Current Analytical Methodologies

The analytical procedures of the USEPA methods developed for phenolic determination include LLE from the acidified water sample. Nearly all methods use GC coupled with various detectors such as flame ionization detection (FID), electron capture detection (ECD) and mass spectrometry (MS) for analysis.

The rapid development of analytical instrumentation has overtaken some standard techniques, such as the use of packed columns for GC separation.

However, owing to new regulatory strategies, such as those preferring environmentally friendly methodologies that reduce solvent consumption and do not use toxic chemicals, government regulation agencies have become quite flexible about accepting and introducing new analytical techniques (e.g. SPE or micro-LLE instead of LLE).

The need to use state-of-the-art methods for monitoring dangerous organic pollutants such as phenolics, halogenated compounds, polyaromatic hydrocarbons, pesticides, amines, phthalates or surfactants in aquatic systems is now recognized.

Most standard procedures have been derived from methods used for drinking water. This means that the detection limits described were obtained without any clean-up, because of the lack of a matrix. If groundwater with low amounts of associated organic matter has to be analyzed, these standard methods may be easily used. However, care must be taken if these methods are to be transferred and adapted to matrix-loaded surface or wastewaters. In such cases, some preliminary sample preparation (clean-up[37]) is needed over and above the extraction, concentration and chromatographic separation (GC or HPLC) procedures usually applied.

If a good method of optimization succeeds, i.e. high enrichment factors are attained for the target analytes, the matrix effects are eliminated by clean-up. If the chosen analysis technique is selective enough, the detection limits may be of the same order in both drinking water and surface water.

Finally, two examples of official methodologies in common usage for water monitoring are shown in Figure 1.

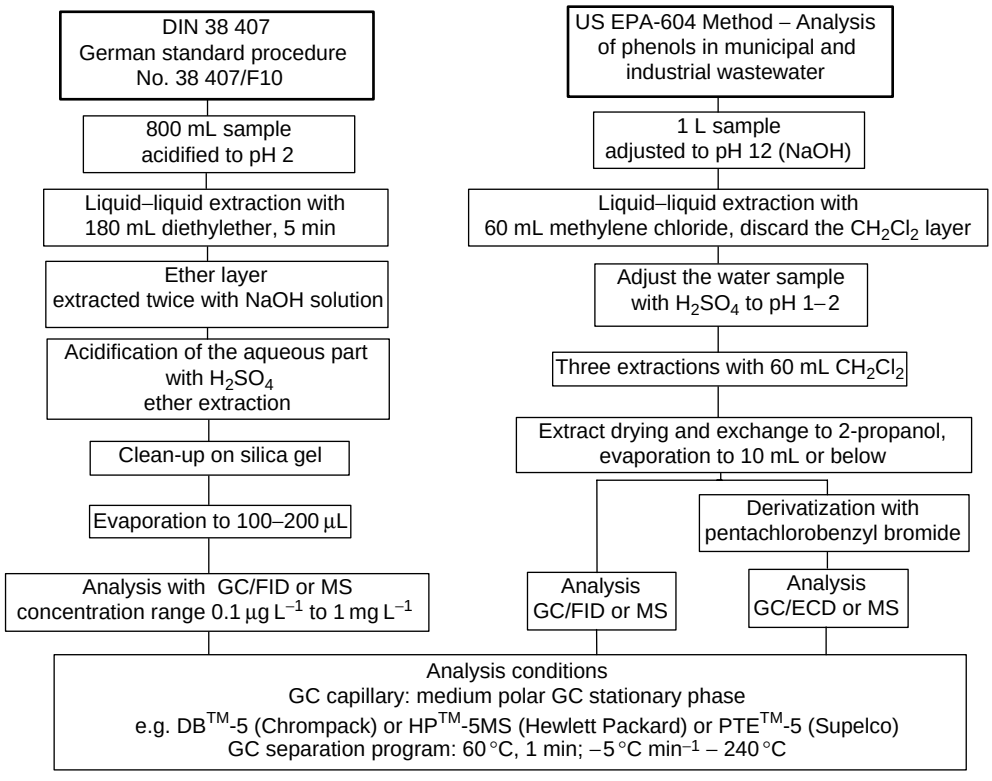

Figure 1 Examples of two currently used official methodologies for determining phenolic compounds in water.

The LLE technique has been proposed in many official methods and is still used because of its ease of use, even though it no longer corresponds to the state of the art or the latest requirements to replace methods using large amounts of toxic solvents. The SPE and microextraction techniques should be preferred in future developments.

As analytically produced data are used as a basis for global conclusions, risk assessment and ultimately to lay down regulatory limitations for environmental pollution, the need for analytical quality control is becoming increasingly evident.

Analytical quality assurance is an additional criterion whose significance is growing in the assessment of environmental data.[38] Only the critical validation of each individual step in sampling, sample transport, storage and preparation, analytical determination and data processing can guarantee the reliable interpretation of data.

3 DETERMINATION OF PHENOLS AS A GROUP PARAMETER

A detailed analysis of individual phenols is not always required for on-site screening and alarm monitoring. Instead, it is the ability to detect a sudden increase in the concentration of pollutants as soon as possible that is needed. The ''phenol index'' is a group-selective parameter that indicates the content of antipyrine dyestuff-forming compounds by photometric detection. However, although phenols are the main target substances, interference may arise from an assortment of different compounds. In addition, the phenol index is not well designed for the determination of para-substituted phenols, thus this method can lead to underestimated results.[39,40]

Based on the draft version of international norm ISO/DIS 6439, the DIN has drawn up the method DIN 38 409 H 16[41] to determine the phenol index in water, wastewater, and sludge.

Depending on the water quality, various procedures exist for determining the phenol index. Direct photometric determination using phenol–antipyrine dyestuffs is recommended for investigating drinking water or surface water with average pollution-containing concentrations of about 10–150 µg L^{-1} (related to phenol[41]). The products absorb at 460 nm and result from the oxidizing coupling of phenols with 4-aminoantipyrine (4-amino-2,3-dimethyl-1-phenyl-3-pyrazoline-5-one) under alkaline conditions and in the presence of potassium peroxodisulfate. The absorption intensity of the dyestuff extracted with trichloromethane is used as a measure of quantity (standard method variant DIN 38 409 H 16-1[41]).

Some aromatic amino compounds are also able to form antipyrine products that are photometrically sensitive in

the same range as phenols. Thus interference may occur and this method is unsuitable for examining water samples containing cyanide or amino compounds.

Alternatively, a slightly different version of the sum parameter method may be applied, which excludes most of the amino compounds prior to analysis. This is achieved by adjusting the sample solution to pH 0.5, forcing the conversion of the amino compounds into salt forms. Prior to the coupling reaction, the phenols are then separated from the water by water vapor distillation (DIN 38 409 H 16-2).[41] As the interfering components are now absent, distillation can be performed using just a small amount of acidic solution (pH 4).

At higher concentration levels ($0.1–10\,mg\,L^{-1}$ and above), the sample must be diluted and extraction of the resulting dyestuff is generally unnecessary. Photometric detection is performed directly on the water sample (DIN 38 409 H 16-3[41]). The procedure is very sensitive to oxidizing compounds (chlorine), reducing substances (SO_2 or NO_2^-), colored components, and cyanides. In any event, the procedure requires several chemicals and time-consuming steps.

To minimize the requirements for solvents and reagents, the entire procedure has been scaled down, and made suitable for on-site application. The cuvette test has been developed from the above-mentioned phenol index method and is divided into two categories. Phenols can be determined in concentration ranges of $0.05–5\,mg\,L^{-1}$ in water, wastewater, soil and waste (LCK® 345[42]), and at higher levels ($5–200\,mg\,L^{-1}$) from wastewater, exhaust gases and condensates such as those released from the manufacture of, for example, coal tars or asphaltenes (LCK® 346[42]).

The photometric determination of phenols as a group is still being developed. In particular, the improved preconcentration of phenols from aqueous solutions is achieved by using a two-phase system such as poly(ethylene glycol)–ammonium sulfate in water.[43]

As UV spectroscopy enables chlorophenols in the milligram per liter range to be rapidly screened, combination with the mathematical interpretation of the spectra using for instance multivariate calibration techniques provides good prediction for analyzing a group of selected phenols.[44] The rapid results obtainable highlight the usefulness of sum parameter methods for early-warning purposes, and for the permanent monitoring of dynamic systems such as industrial discharge or river water control with high sample throughput.

4 DETERMINATION OF PHENOLS

In addition to rapid screening, mostly at higher concentration ranges, trace analyses of individual phenols are sufficient to provide an appropriate stock of data for risk assessment and environmental protection. The determination of a large variety of compounds in the low microgram per liter range requires high analytical performance and sophisticated sample preparation techniques. The development of an analytical strategy depends primarily on the nature of the accompanying matrix and on the analyte concentration.

A fundamental review of the determination of phenols in water samples has recently been published.[45] The current methods are described as LLE and LSE, off-line and on-line, combined with GC and liquid chromatography (LC). Use of LSE with subsequent thin-layer chromatography (TLC) can provide an easy, rapid, and inexpensive route to obtaining an overview for screening purposes.

4.1 Preparation of Aqueous Samples

A concise review of the main steps and techniques for specific phenol determination is contained in *Water Analysis*,[46] This booklet features practical hints, analysis flow diagrams and operating conditions, which provide a convenient introduction to the methodologies currently used.

In general, for aqueous sample preparation, LLE, LSE, SPE, and SPME are preferred in combination with GC/FID, ECD, MS and the option of prior derivatization. HPLC with UV detection does not require any additional derivatization, and is increasingly being used for phenol determination from water samples.

4.1.1 Liquid–Liquid Extraction

The classic variant of sample preparation is LLE under acidic conditions. The implementation of this method in major standard procedures accounts for its widespread use in environmental laboratories, although more recent methods are now available with a number of advantages such as simplicity, good reproducibility, and sensitivity.

USEPA standard methods 604 and 625 involve the methylene chloride extraction of water samples after acidification to pH 2. Subsequent GC analysis reaches detection limits in and slightly below the microgram per liter range (Table 1).

The analytes in surface water and wastewater are often accompanied by organic material, such as humic and fulvic acids. These matrix components may notably affect analytical performance because of their competing extraction and accumulation under acidic conditions. Therefore, in LLE the presence of interfering material probably requires additional sample clean-up.

Another important aspect concerning matrix influence is division of the pollutants between the suspended matter and the water. One argument in favor of LLE is that

contaminants are extracted from both the free water and the suspended matter, whereas LSE mainly analyzes contaminants dissolved in the water.

As large amounts of mostly toxic and flammable solvents are required for extraction, LLE is not an environmentally friendly or economic method. The formation of emulsions during extraction, additional clean-up and solvent evaporation steps can cause some reproducibility problems or even the loss of analytes. These features make automation difficult. Hence, there is a general trend to substitute current LLE procedures by LSE protocols.

4.1.2 Liquid–Solid Extraction and Solid-phase Extraction

The SPE method corresponds to the current trend towards automation and establishing environmentally friendly techniques in laboratories.

It can be used off-line, or on-line linked up to a chromatographic device. The separate operation of the extraction step from analysis has some advantages regarding the versatility of application. In method development, for varying extraction problems (matrix, concentration), and for on-site sampling, an instrument-independent SPE step is favored. Furthermore, the eluates obtained can be extensively treated as often as required and analyzed using different methods. Automation is possible using preparation units that sequentially extract the samples and clean them up for automatic injections.

Nevertheless, there is a general trend towards on-line combinations because fully automatic systems allow high sample throughput without any loss in analytical performance. Generally, reproducibility and detection limits are improved because the closed system minimizes the loss of analytes and limits sample manipulation problems. Unfortunately, the restricted size of the precolumns may cause problems with low-breakthrough analytes such as phenol.[47]

Off-line techniques have the disadvantage of lower sensitivities owing to the injection of an aliquot and analyte loss during the evaporation steps. On-line coupling of SPE to the chromatographic system avoids these problems. Moreover, the sample volume can be reduced because the entire sample is transferred and analyzed.

LSE combines all the techniques that use the analyte partition between a solid and a corresponding liquid phase in order to separate and accumulate the target compounds from an aqueous sample. The separation mechanisms are based on absorption as well as adsorption phenomena without achieving partition equilibria at any points of extraction. Several liquid chromatographic methods also belong to this category and may be conveniently combined with SPE as a prior sample enrichment and clean-up procedure.

Numerous sorbent materials are commercially available for various purposes. Suitable surface modifications adapted to special extraction problems may solve most of the sample enrichment problems in environmental analysis. The bonded silica used until recently is being increasingly substituted by polymeric sorbents with higher capacity and efficiency regarding the extraction of polar compounds, such as the phenols.[47]

Prefabricated sorbent material packed as columns, cartridges and disks are supplied by various manufacturers. The larger the active sorbent surface, the more effective is the extraction of polar compounds. Consequently, highly porous supports and strongly cross-linked coatings have gained acceptance. Breakthrough volumes, recoveries and detection limits related to sample volume and the analyzer used are important parameters for characterizing SPE performance. Lichrolut EN™ (available from Merck), a very suitable sorbent for the extraction of polar compounds, allows the trace-level detection of the phenols listed in Table 2 from ground- and river water with recoveries of 80% and higher.[45] Several surface and coating modifications have been examined to increase the capacity of sorbents and to improve the analytes' extraction and desorption behavior. Beneficial modifications have been achieved by acetylation, hydroxymethylation, or sulfonation,[48] providing improved recoveries for several compounds such as phenol and catechol with low breakthrough volumes on commercially produced SPE material.

One material that predominantly adsorbs analytes is based on graphitized carbon. Other than for phenol itself, the extraction of phenolics from water is basically possible at a recovery of about 90%.[49,50] The inherent difficulties of analyte desorption are due to the excessive or even irreversible retardation of the adsorbed molecules, such as 2,4-dimethylphenol. Backflash elution may improve the desorption but is difficult to automate. Some improvements in mechanical stability and the handling of carbon-derived sorbents have been achieved by immobilizing porous graphite carbon on a silica surface. This modification showed good results for the concentration of aminophenols and of the highly polar catechol.[49]

Heating or microwave excitation may also support the yield and the selectivity of the desorption step.[51]

The development of new sorbent materials continues for improved selectivity (see also section 6.2), capacity, stability, and feasibility for field application.

Very different sorbent breakthrough volumes as exhibited by the phenols may cause problems in simultaneous determination. For example, high sorption capacities for PCP contrast with low adsorption capability for phenol, monochlorophenols, and nitrophenols. Although the new Empore™ disk containing a polystyrene–divinylbenzene copolymer can compensate for insufficient recoveries of

Table 2 Detection limits of phenolic compounds in groundwater and river water when working with on-line and off-line LSE procedures using Lichrolut EN™ (sample volumes 100 mL and 1000 mL for on-line and off-line, respectively)

Analyte	Detection limits ($\mu g\,L^{-1}$)			
	On-line		Off-line	
	Groundwater	River water	Groundwater	River water
Phenol	0.70	3.50	0.80	3.00
Catechol	2.00	9.00	0.50	6.00
4-Methylphenol	0.40	2.00	0.25	1.70
2,4-Dimethylphenol	0.03	0.50	0.10	1.00
2-Nitrophenol	0.05	0.80	n.d.[a]	n.d.[a]
4-Nitrophenol	0.01	0.08	0.05	0.25
2,4-Dinitrophenol	0.09	0.30	0.10	0.80
4-Chloro-3-methylphenol	0.05	1.00	0.50	1.30
2-Chlorophenol	0.20	1.20	0.10	1.00
3-Chlorophenol	0.20	1.20	0.10	0.90
4-Chlorophenol	0.20	1.40	0.25	1.40
2,4-Dichlorophenol	0.10	0.80	0.10	0.50
2,4,6-Trichlorophenol	0.08	0.30	0.10	0.60
2,3,5-Trichlorophenol	0.07	0.30	0.10	0.20
2,3,4-Trichlorophenol	0.08	0.30	0.10	0.10
PCP[b]	<0.01	0.02	0.10	0.20

[a] n.d. = not detected.
[b] At 310 nm.
(Reproduced with permission of Elsevier Science from: D. Puig, D. Barceló, 'Determination of Phenolic Compounds in Water and Wastewater', *Trends Anal. Chem.*, **15**(8), 362–375 (1996).)

Table 3 Commonly used derivatization reagents for phenols

Reagent	CAS No.	Transferred group	Analysis by
Trimethylsilylation			
Bis(trimethylsilyl)acetamide	10416-59-8	$(CH_3)_3Si$	GC/FID, MS
Bis(trimethylsilyl)trifluoroacetamide	25561-30-2	$(CH_3)_3Si$	GC/FID, MS
Trimethylchlorosilane	75-77-4	$(CH_3)_3Si$	GC/FID, MS
1,1,1,3,3,3-Hexamethyldisilazane	999-97-3	$(CH_3)_3Si$	GC/FID, MS
N-(Trimethylsilyl)diethylamine	996-50-9	$(CH_3)_3Si$	GC/FID, MS
N-(Trimethylsilyl)imidazol	18156-74-6	$(CH_3)_3Si$	GC/FID, MS
Acylation			
Heptafluorobutyric anhydride	336-59-4	C_3F_7CO	GC/ECD, MS, FID
N-[Methyl-bis(trifluoroacetamide)]	685-27-8	CF_3CO	GC/ECD, MS, FID
Trifluoroacetic anhydride	407-25-0	CF_3CO	GC/ECD, MS, FID
Acetic anhydride	108-24-7	CH_3CO	GC/FID
Alkylation			
Pentafluorobenzyl bromide	1765-40-8	$C_6F_5CH_2$	GC/ECD, MS, FID

Reproduced with permission of John Wiley & Sons, Inc. from Blau and Halket.[52]

the chlorophenols, phenol still remains a problem with just 20% recovery.

4.1.3 Derivatization

As described in section 2.5, derivatization is particularly required in phenols determination when using GC separation techniques. Due to the high polarity and low volatility of some phenols, a derivatization step improves thermal stability, gas chromatographic behavior of the analytes, and most notably the selectivity and sensitivity of detection. Commonly used derivatization reagents are listed in Table 3.[52]

Typically, three basic reactions are recommended for the derivatization of the hydroxy group of phenols. O-Acylation, silylation and alkylation are possible, with the

selection of the most suitable reagent depending on the analysis method applied. Acylation converts compounds with active hydrogens, such as phenols or alcohols, into esters through the action of carboxylic acids, anhydrides, or carboxylic acid halogenides. Alkylation replaces the active hydrogen by an aliphatic or aliphatic–aromatic group and silylation forms trimethylsilylether.[53–55] The GC retention times of the derivatives are influenced in various ways. Acylation decreases retention due to the increased volatility of the products. In contrast, trimethylsilylation mostly delays the retention times.

The highest detection sensitivity is obtained in combination with fluorine-containing reagents and GC/ECD analysis. As MS detection and even tandem mass spectrometry (MS/MS) techniques are available, a derivatization reagent should be chosen that produces characteristic fragment ions to improve identification. For example, the introduction of a trimethylsilyl (TMS) group results in quite abundant ions representing the complete derivate group at 73 amu ($(CH_3)_3Si]^+$) and methyl abstraction from the molecular ion ($M-CH_3]^+$). The molecular ions of the TMS derivatives often appear weak. More predominant in the mass spectrum are the molecular ions of the trifluoroacetyl (TFA) derivatives. Furthermore, every spectrum of such a phenol derivative is marked by the characteristic loss of the TFA group ($M-CF_3CO]^+$). This fragmentation can be used by MS/MS techniques for the highly selective detection of phenolic compounds even without preseparation of the total mixture. The rapid screening of lignite-derived wastewaters and asphaltene products is also possible.[56]

The lack of specificity of most reagents should be mentioned as a disadvantage of derivatization because reactions with other compounds bearing acidic hydrogens, such as amines, mercaptans, alcohols, and carboxylic acids, cannot be prevented without prior clean-up. Furthermore, undesirable by-products such as acids are formed during reactions with anhydrides and acylhalogenides. Such aggressive by-products have to be removed prior to GC analysis, in order to avoid damaging the GC separation phase. In this respect perfluoroacylimidazole is preferred because no interfering by-products occur.

Similarly, in the HPLC/UV analysis of phenols, sensitivity can be improved by postcolumn photochemical derivatization reactions. The UV-initiated decomposition of dansyl derivatives of phenolic compounds in a methanol–water mixture leads to the formation of highly fluorescent dansyl hydroxides or dansyl methoxides. Detection limits below the nanogram per liter level can be achieved by dansyl derivatization and direct fluorescence detection.[57] Fluorescence can also be used to detect phenols at trace levels without derivatization.[58]

4.1.4 Liquid–Solid Extraction and Solid-phase Microextraction

Another emerging approach combines sample preparation and analysis. The SPME method is a solvent-free LSE technique when combined with GC. Combination with GC, HPLC, or capillary electrophoresis (CE) allows the analysis of different pollutant classes from the air, water and soil–water slurries at trace-level concentrations. A comprehensive review of the theoretical and practical aspects of SPME is given by Pawliszyn.[59]

The SPME technique is not easy to classify as merely a sample preparation or a sample introduction method that transfers analyte molecules into an analytical system, because SPME combines both aspects. The main advantage is the absence of any solvent within the extraction step, which complies with the current trend towards the development of environmentally friendly and economic analytical methodologies. Furthermore, particularly high sensitivities may be achieved for substances with high octanol–water coefficients ($\log K_{ow}$).[59] First described by Pawliszyn and coworkers,[60–64] SPME is a simple, rapid and inexpensive method which may be conveniently introduced into laboratory protocols. It is now commercially available as an automated technique coupled on-line to a gas chromatograph (Varian auto-SPME/GC). A fused silica fiber coated with a polymer film is fixed inside a syringe (Figure 2).

For analysis, the fiber is exposed to the sample and the analytes are absorbed in the stationary phase until their partition equilibrium is reached. Additional sample filtration prior to extraction is not needed. After extraction, the fiber is then removed from the sample and the analytes trapped are thermally desorbed in the injector of a gas chromatograph. Care and attention has to be paid to septum leaks which can arise when the SPME needle penetrates the GC septum, but numerous reports have proved the high reproducibility and reliability of SPME/GC if professionally applied.

Different fiber coating materials with specific selectivity and extraction mechanisms (absorptive, adsorbtive) and appropriate SPME fiber assemblies are commercially available. In principle, SPME coating materials are very similar to the polymers used in the stationary phase of GC. Of the diverse range of coatings examined for phenol determination, polyacrylate (PA) has proved to be the most suitable.

The SPME technique was first used to analyze phenols by Buchholz and Pawliszyn.[65,66] The detection limits for the phenolics included in the USEPA 604 method were in the nanogram per liter range using GC/FID and GC/MS. The optimized protocol used a saturated sodium chloride solution at pH 2 to increase SPME sensitivity. The optimum sampling time depends on the sample matrix, namely on the diffusion coefficients of the

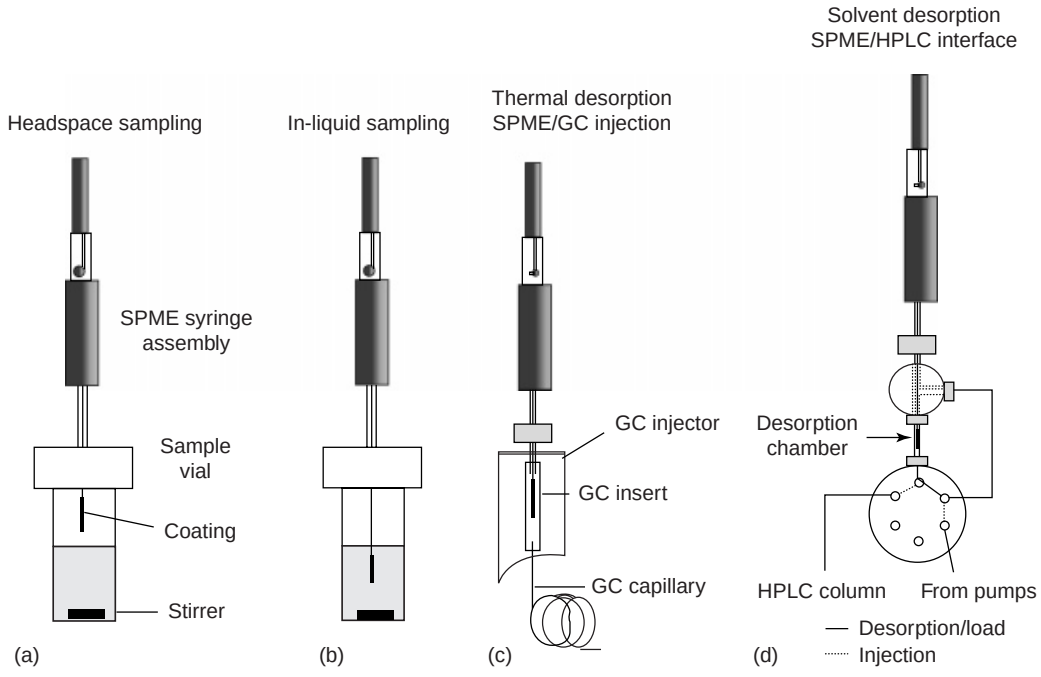

Figure 2 Syringe assembly for SPME and individual sampling steps. 1. The SPME fiber is exposed to headspace (a) or in liquid (b) while suitable agitation of the sample (stirring, sonication, fiber vibration) takes place. 2. After the extraction period, the fiber is rejected and removed from the sample. 3. Thermal (c) or solvent-assisted (d) desorption of the sample into the GC (c) or HPLC (d) instrument.

analytes under the respective conditions. As they can vary between 30 min and 1 h,[67] optimization prior to a series of analyses is recommended.

For comparison, the GC/MS analysis of a phenol mixture was performed on samples prepared by directly injecting an extract and by SPME of the aqueous sample (Figure 3).

Although derivatization before and/or during SPME sampling can enhance the sensitivity and selectivity of analysis, the incorporation of this additional step typically complicates the entire process, and should only be considered when necessary. Quantification requires some additional effort to ensure reproducible results.[59]

Higher amounts of organic matter can influence the extraction process, recoveries and detection limits achieved[67] by delay of the analyte diffusion or by partial blocking of the surface coating. In these cases, reliable quantification requires standard addition to compensate for complex matrix effects. Standard addition involves the multiple, phased addition of defined amounts of suitable reference compounds and subsequent analysis. The resulting relationship between analytical values detected and the analyte amounts added gives a calibration curve whose regression leads to the initial concentration of the target analyte in the original sample. This additional step is only needed in the presence of interfering matrix material; otherwise, for low-matrix-polluted waters, such as tap water or groundwater, quantification using an internal, usually isotope-labeled, standard is recommended.

In contrast to SPE, the SPME techniques have significant advantages of time and lower solvent consumption per analysis (Figure 4). The SPME technique also has potential use with mobile and miniaturized GC devices for on-site measurements.

4.2 Preparation of Solid and Gaseous Samples

Typical procedures for sample preparation of solid material involve LSE with an organic solvent followed by both clean-up and concentration stages. This methodology is time-consuming and costly in the amount of toxic solvents required. Furthermore, aggressive shaking and/or temperature conditions are required to release the analytes adsorbed on the soil matrix. The conventionally used method to extract organic pollutants from soil is Soxhlet extraction, which consists of a simple distillation process repeated a number of times. In total, extraction times vary between 4 and 24 h depending on the soil material and the analytes to be extracted. Alternatively, the soil slurry can be sonicated but with the risk of a lack of sensitivity owing to possible analyte degradations.

Thermally and pressure-assisted extractions, generally described as enhanced solvent extraction (ESE), are increasingly being applied and accepted by official regulations. Examples are, supercritical fluid extraction

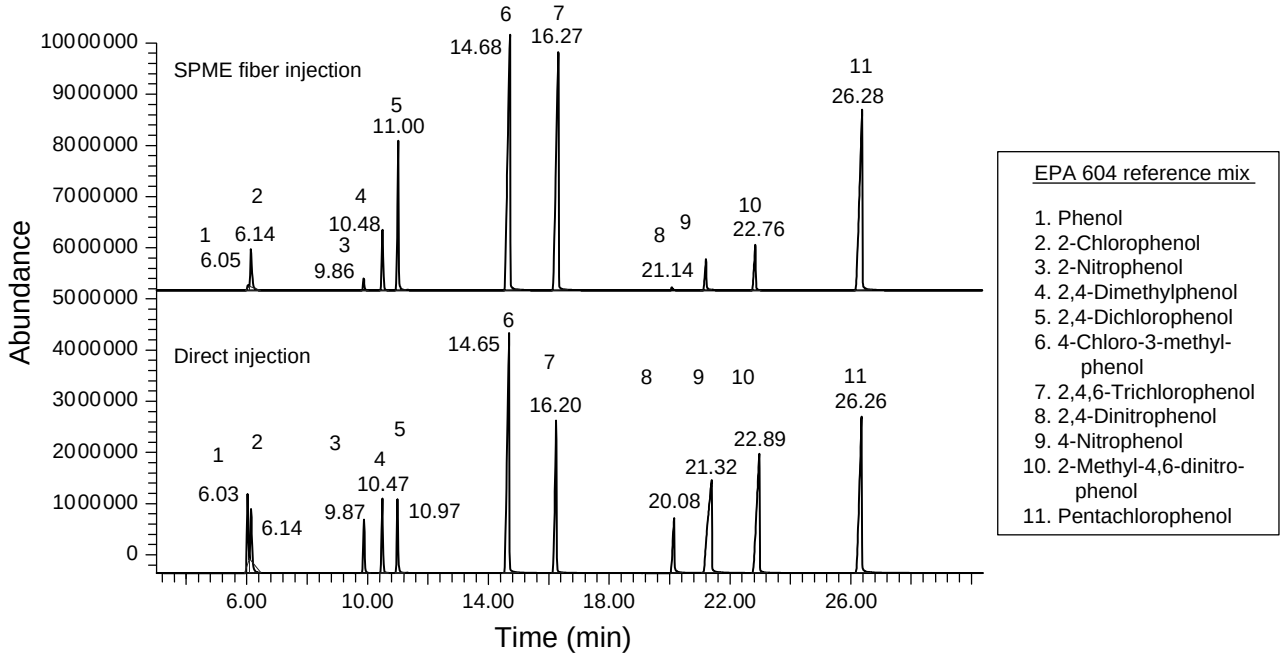

Figure 3 Lower trace shows the GC/MS chromatogram of the USEPA 604 phenols directly injected from a mixture of defined concentration. The upper trace is of the same mixture sampled and analyzed by SPME using a PA fiber coating.

(SFE) and accelerated solvent extraction (ASE™). Microwave-assisted sample extraction has also been introduced into sample preparation protocols.

4.2.1 Subcritical and Supercritical Fluid Extraction

Since the rapid developments of supercritical fluid chromatography in the 1980s, SFE has been considered as a sample preparation technique in environmental analysis owing to its efficiency and time-saving extraction capabilities. Summaries and method reviews have been given by Hedrick et al.,[68] Barnabas et al.[69] and Bøwadt and Hawthorne.[70]

In practice, more than 90% of all analytical SFE applications use carbon dioxide (CO_2) as the supercritical fluid due to its low critical pressure of 74 bar and critical temperature of 32 °C. Carbon dioxide is also nontoxic, is not flammable or explosive, is chemically inert, and is easy to remove after extraction; however it is not sufficiently polar to extract phenols. At supercritical conditions carbon dioxide has a comparable polarity to liquid pentane, and is therefore an excellent fluid for the extraction of lipophilic compounds. A few percent of an organic solvent such as methanol are often added as modifier to enhance the solubility of more polar compounds. Alternatively, the polarity of the analytes can be reduced by derivatization or complexation using chelating agents in order to improve the SFE efficiency. For example, chlorinated phenols from sediments were derivatized in situ using acetic anhydride in the presence of triethylamine. The subsequent SFE was carried out with CO_2 at 80 °C and 370 bar using 10 min of static extraction and 5 min of dynamic extraction with a flow rate of 2 mL min^{-1}. The recoveries obtained varied between 84% and 100%.[71] Apart from several advantages for the extraction of lipophilic compounds from solid matter, SFE has not won wide recognition in phenol determination because it is not easy to use for routine analyses. Moreover, commercial apparatus is expensive and not very widespread.

Extraction methods using organic solvents or even water at moderate pressure and temperature have been used. Under specific conditions, water is an ideal solvent for a wide range of pollutants of different polarity. The intrinsic polarity of water supports the preferred extraction of polar phenols. By changing the temperature at moderate pressure (subcritical conditions), a wide range of water polarities can be achieved,[72] which allows the sequential extraction of different polar compounds from soil; however, supercritical water may be corrosive to stainless steel vessels, thus the operator should take care.[72]

For GC analysis the analytes have to be transferred from water into an organic solvent (e.g. trichloromethylene). After release of the water extract from the extraction cell, it bubbles through an organic solvent and the analytes partition between both phases. This step may lower the extraction efficiency of polar compounds owing to their high water solubility. Additional

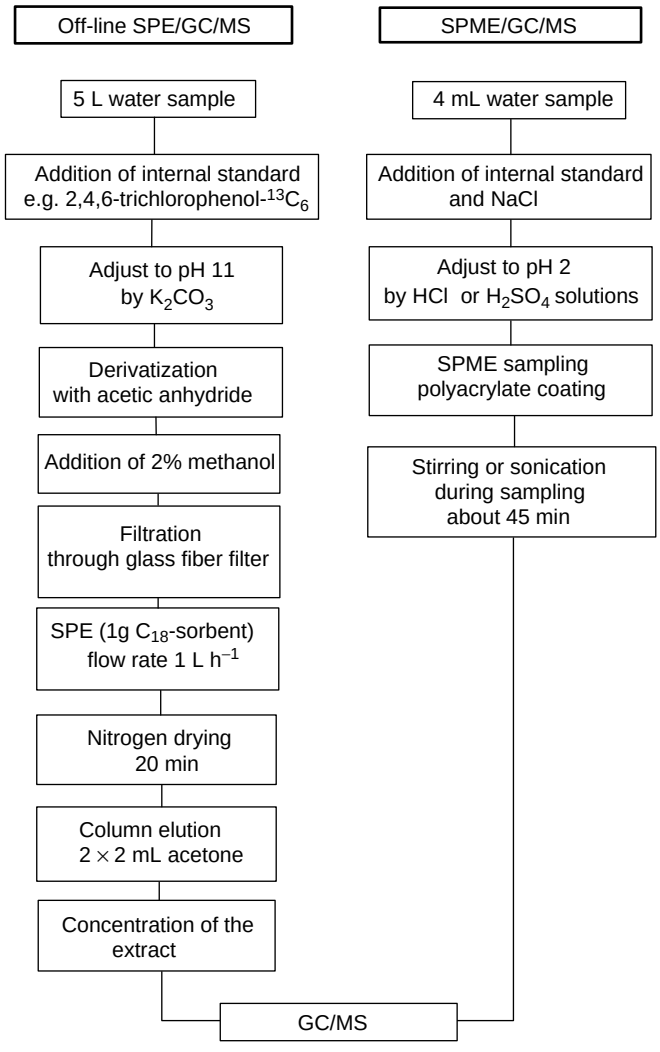

Figure 4 Analysis flow charts for off-line SPE/GC/MS and SPME/GC/MS.

LLE steps, also assisted by sonication, SPE or SPME, can enhance the recoveries. Chlorinated phenols can be efficiently extracted by subcritical water, as the recoveries of 96–100% prove.[72]

4.2.2 Accelerated Solvent Extraction

The ASE™ protocols, used for the extraction of basic, neutral and acidic environmentally relevant compounds from solid samples, meet the requirements of official methods such as USEPA 3545.[32] Commonly used solvents are applied to the extraction from solid material under increased pressure (10 MPa) and temperature (100 °C) conditions.[73] The extraction conditions change the properties of the solvents and improve the wetting of solid surfaces. The release of analytes from the solid matrix is possible in shorter times (usually 12 min) and is managed with a small volume of solvent, about 15 mL per 10 g of soil. The fully automated process is easy to operate and gives recoveries that are comparable with Soxhlet extraction.

For example, the extraction of nonylphenols from soil and solid material can be performed by Soxhlet extraction, steam distillation/solvent extraction or more economically by ASE™ and SFE. Optimal yields of nonylphenols were obtained by ASE™ using pure methanol and a fairly low pressure of about 15 MPa.[74]

Following hot water extraction or an aqueous ASE™ of the contaminated solid materials, the incidental water extracts are analyzed by SPME/GC/MS.[75]

Optimum ASE™ conditions using water as solvent for the extraction of soil spiked with USEPA 604 phenols are 10 MPa at 125 °C. The extraction cycle of 10 min is repeated three times. The addition of 5% (v/v) acetonitrile as solvent modifier can improve phenol recovery slightly. The detection limits obtained after SPME/GC/MS analysis range from 1.1 μg kg^{-1} for 2-chlorophenol to 6.7 μg kg^{-1} for 4-chloro-3-methylphenol. The SPME sampling was performed with a PA fiber coating at conditions currently used for SPME of phenols. The ASE™ methodology using only pure water gave slightly higher detection limits, but within the lower microgram per kilogram range.

Optimum conditions obtained for the extraction of spiked soils are usually less efficient when real samples are extracted. Thus extraction efficiencies are overestimated, especially in cases of strongly different soil matrices (such as by varying the content of organic matter).[76–82]

4.2.3 Microwave-assisted Extraction

The number of procedures using microwave-assisted extraction (MAE) of organic compounds from environmental samples is increasing. Agitation by microwave energy causes molecular motion by migration of ions and rotation of dipoles without changing of molecular structures. The theory and main applications of microwave-assisted processes (MAP™) are well described by Dupeyron et al.[83] and Paré et al.[84]

Only dipole-containing compounds are affected by microwaves. Therefore, with nonpolar solvents such as hexane or toluene, polar additives such as acetone are required. The main advantages of microwave use are the reduction of solvent, shorter extraction times, and the greater number of samples that can be simultaneously extracted. In an experimental set-up[85] MAE and the most currently used Soxhlet extraction were compared using a phenol-spiked clay soil from an industrial downfall. The recoveries for most of the phenols studied were comparable. 2-Methylphenol and 2-nitrophenol gave better results with MAE. Generally, low recoveries for alkylated phenols are reported, probably attributable to intensive analyte–soil interactions.

In every case, the extracts contain, among the phenols, a great variety of matrix compounds and a clean-up using LC on alumina, silica gel or aluminosilicates is required.[86,87] Gel permeation chromatography or preparative HPLC can improve and automate this cleaning process.

As the organic solvents have been replaced by water, the resulting aqueous extracts can be immediately sampled and analyzed by SPME/GC/MS.

4.2.4 Air Monitoring

Compared to soil and wastewater, air samples are easier to handle. The decisive and most important step is the enrichment of phenols under defined conditions. Passive and active (or dynamic) sampling are the two main ways for collecting phenols from gaseous samples. Canisters, sampling bags and bulbs commonly used in air sampling are not suited for polar and surface-active analytes (like phenols). Despite passivation of the canister and bag materials a nonreversible adsorption of phenol cannot be prevented completely. Sorbent tubes provide a low-cost and versatile alternative to containers.[88] The active air sampling onto sorbent tubes needs only a convenient pump assembly and a thermodesorption interface for subsequent GC analysis. Tubes prepacked with sorbents are commercially available, such as Tenax™, Chromosorb™, Carbopack™, or Amberlite XAD™ resins. They can be used individually or in combinations of two or three.[35] After desorption they are ready for reuse, but after a longer storage period the sorbents should be cleaned thermally. The sorbents used are stable enough for about 100 cycles before they have to be replaced. The most sensitive parameter that has to be considered in method development is the breakthrough volume of the analytes of interest, which depends on the amount and quality of sorbents used in the tube. Also important is the gas flow used for conditioning, sampling and desorption, the influence of humidity during sampling, and operating temperatures, especially those of the desorption step. The desorbed analytes are collected on a cooled focusing trap prior to GC analysis. After desorption the trap is heated to transfer the analytes to the GC capillary for analysis. The USEPA air monitoring methods (TO-8 and TO-17) can be used to select the appropriate procedures for the determination of volatile phenols in gaseous samples.

5 INSTRUMENTAL TECHNIQUES

5.1 Gas Chromatography

The analytical approaches most frequently used in conjunction with LLE and LSE are GC and LC. In particular, GC is combined with FID, ECD, or MS. Each combination is accepted by official USEPA methods. Of the detectors, MS is the most sensitive and selective, including the identification of unknown components. However, FID and especially ECD (which is highly selective and sensitive for polyhalogenated compounds) are more affordable and are therefore widely used in the routine determination of priority phenolics.

For the determination of endocrine-disrupting phenols (e.g. nonylphenols) and their precursors, the alcohol ethoxylates, a number of method combinations are available.[89] Off-line SPE is the preferred enrichment procedure from water samples using Amberlite XAD™-2 and XAD™-4. Excellent recoveries of 90–100% have been obtained using four different solvents for SPE elution.[90] GC/MS and HPLC/UV are the most favored analytical methods for determination. The most easy, solvent- and time-saving procedure with which to determine nonylphenols in aqueous samples is SPME/GC/MS using a PA or poly(dimethylsiloxane)/divinylbenzene (PDMS/DVB)-coated fiber. Detection limits for the nonylphenol isomers are between 1 and 10 $\mu g\,L^{-1}$.

5.2 High-performance Liquid Chromatography

The most commonly used procedure for the extraction and concentration of phenolics from aqueous samples is LSE, and on-line coupling to HPLC is becoming increasingly popular. This configuration does not require any prederivatization steps, especially for phenol determination.

Octadecyl (C_{18}) and octyl (C_8) HPLC columns are currently used to determine phenols. Although most phenols may be detected by UV at 280 nm, the detection of nitrophenols and PCP is usually more sensitive at 310 nm.[45,91] Full spectral information is obtainable using a diode array detector, which may assist in the identification of unknowns or the confirmation of assigned analytes by spectra libraries usually contained in the software packages of HPLC instruments. Table 2 compares the on-line and off-line SPE/HPLC/UV analysis of selected phenols, showing the clear advantages of on-line coupling.[45] Otherwise, the detection limits depend strongly on the detector used. Electrochemical detection instead of UV can improve sensitivity by a factor between 10 and 100.[92,93]

The high detection sensitivity results in a possible reduction of sample volume which complies well with on-line SPE/LC and minimizes the input of interfering material. Unfortunately, oxidation-susceptible matrix components can interfere with the background current, resulting in low reproducibility of the detection.

If a high oxidizing material content is present then cleaning of the detector cell is frequently required

to restore the initial response, so that electrochemical detectors are not necessarily advisable for automated wastewater analysis. The use of biosensors as LC detectors can overcome some of these problems because of their high analyte specifity, but unfortunately there is still a lack of sensitivity.

Coulometric detection has not found wide acceptance in spite of the low detection limits attainable. The increase in sensitivity as compared to the usual electrochemical detector is caused by nearly complete oxidation of the analyte inside a porous electrode. The more sophisticated and therefore more expensive variant is an array detector, which further improves the detection limits (0.03 and 0.38 ng L^{-1}) and extends the analyte response.

Special equipment is required to couple MS with LC. Today several LC/MS interfaces are available. Nearly all the currently used interfaces have been examined for phenol determination but only a few give full responses for the entire list of priority phenols. For example, thermospray is not suited to the ionization of phenol and some methylated phenols.[94] The technique is based on ion formation by the addition of volatile buffers and thermally assisted release of the solvent. Under these conditions most phenols produce negative ions by deprotonation, excepting the few acidic phenols.

Particle beam interface is not successful in phenol analysis because of its low sensitivity.[95]

The LC/MS interfaces, generally classified as atmospheric pressure ionization (API) interfaces, are the most promising in phenols analysis. The interface with the highest sensitivity for phenols is atmospheric pressure chemical ionization (APCI).[96,97] Ion formation occurs by acid/base reactions already in the liquid phase of the LC eluent and additionally in the gas phase of the spray produced by a corona discharge electrode. A suitable buffer ensures hydrogen abstraction from the phenolic hydroxyl group, and produces negative pseudo-molecular ions ($M-H]^-$). The coupling of APCI/MS with on-line SPE using Lichrolut EN™ as sorbent in selected ion monitoring (SIM) mode gives detection limits in the nanogram per liter range. Even in full scan operation the detection limit ranges from 0.5 μg L^{-1} for 4-nitrophenol to 20 μg L^{-1} for 2,4-dimethylphenol.[97]

The same trend as for thermospray ionization is observed, in that phenol, methyl-, dimethyl- and also the monochlorophenols show lower sensitivities than the more acidic phenols, such as the nitro-substituted analogs. However, electrospray or ion spray interfaces can be used that provide more ion abundance, particularly for the phenols that are problematic to APCI. In this approach the post-column addition of volatile basic buffers such as triethylamine or KOH solution is required to generate phenolate ions.[98] In API method development, very careful optimization of both separation and ionization techniques is necessary to obtain maximum sensitivity at optimum reproducibility. For example, phenols require acidic conditions for LC separation but a basic pH for API ionization. In addition to these additional operational difficulties matrix components such as humic matter can negatively interfere with the method response and system stability. The newest development of an orthogonally configured spray needle can delay contamination and blocking of the sample cone and skimmer assembly by solid matter. An important advantage of MS detection is its eminent suitability for the screening of unknowns, especially when the mass spectrometric information can be improved by collision-induced dissociation (CID). For this purpose an additional potential is set to either the cone voltage or to an octapole field (instrument specific). Both variants result in an increase in fragmentation of the normally predominating pseudomolecular ion. Unfortunately, interpretation of the resulting more ion-rich spectrum needs considerable experience and cannot be confirmed by commercially available spectra libraries. This is a general problem in API applications because the majority of differences are instrument dependent. Finally, more expensive instruments with MS/MS functionality can be used for structure elucidation.

Ongoing developments in HPLC (e.g. new microcolumns and capillary columns), in MS (e.g. the coupling of API with MS/MS capable ion traps), or fast-scanning time-of-flight instruments will further increase the attraction of LC/API/MS in environmental trace analysis.

5.3 Capillary Zone Electrophoresis

In addition to phenol determination by ion chromatography,[99] electrophoretic separation is of interest in the analysis of ionic and polar molecules, including the phenols. The main advantages are high chromatographic resolution, and sample enrichment in the capillary by isotachophoresis[100] which may compensate the low loading capacity of CE. Furthermore, the normally disturbing humic and fulvic acids are separated from the analytes by their strongly different migration behavior.

The most common electrolytes are sodium borate or phosphate buffer,[101] which have to be replaced by volatile buffer systems if electrospray ionization (ESI)/MS detection is used; however, UV detection is generally favored. The combination of off-line SPE and capillary zone electrophoresis (CZE)/UV gives detection limits of phenols at low microgram per liter levels.[101] As a result of the basic electrolyte conditions and the high voltage applied, undesirable reactions such as hydrolysis or polymerization of phenols can occur. This can be overcome by micellar electrokinetic chromatography using sodium dodecyl sulfate (SDS) as an additive.[100]

6 NEW DEVELOPMENTS

6.1 Sensors

Most biosensors have been developed based on phenol-oxidizing enzymes coupled to an amperometric detection system. A typical example is shown in Figure 5. Tyrosinase (EC 1.14.18.1), laccase (EC 1.10.3.2) and various peroxidases (horseradish peroxidase[102]) are used as receptor compounds.[103] For example, the tetrameric enzyme tyrosinase which contains 1 mol active copper(I) per monomer catalyses the ortho hydroxylation of phenol to catechol and additionally the dehydrogenation of catechol to orthoquinone.[104] The oxidation of para-substituted phenols is usually catalyzed. The oxygen consumption of an enzymatic cycles can then be used as sensor for phenols oxidized within the process. Alternatively, the electroactive quinoid products can be detected at a cathodically potentiostated carbon electrode. An amplification of the enzymatic electrochemical signal is obtained when the orthoquinone is again reduced via catalytically active conducting polymer or redox mediators. Ferrocyanide, tetracyanochinodimethane, or 5-methylphenazonium cations usually act as an electron shuttle between the cathode and the orthoquinoid products of the enzyme reaction.[105]

Sensor developments are focused on improvements in sensitivity, enhanced resistance against changing medium conditions (pH, temperature), the influence of inorganic and organic matrix components, improved long-term stability, and robustness for field application.

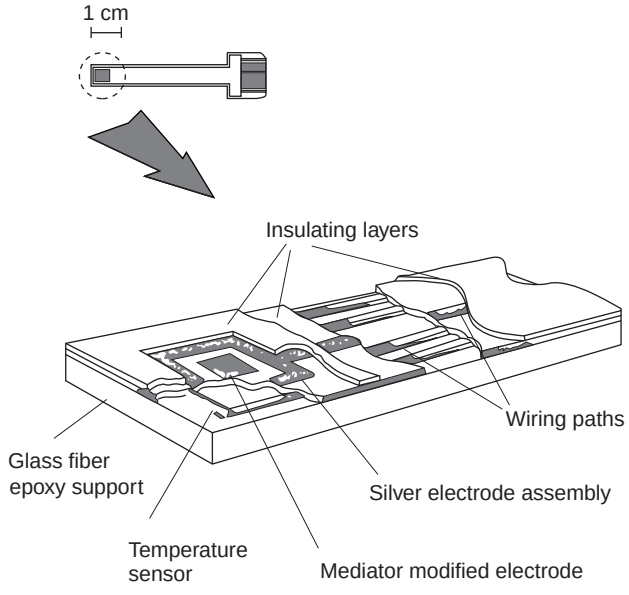

Figure 5 Design of a tyrosinase-based phenol sensor.[111] (Reprinted with permission from H. Kotte, B. Gründig, K.-D. Vorlop, B. Strelitz, U. Stottmeister, 'Methylphenazonium-Modified Enzyme Sensor Based on Polymer Thick Films for Subnanomolar Detection of Phenols', *Anal. Chem.*, **34**, 65–70, Copyright (1995) American Chemical Society.)

Sensor stability depends strongly on the kind of enzyme immobilization on the transducer surface. Different possibilities have been examined to fix the enzyme, such as behind a dialysis membrane,[106] mixing in graphite powder and hexadecanol,[107] carbon paste,[108] epoxy/graphite,[109] in a composite of carbon paste/polyamide,[110] or tyrosinase immobilized by entrapment within a polycarbamoylsulfonate polymere.[111] The latter variant is marked by high sensitivity for 2,4-dimethylphenol (635% related to phenol as 100%), 4-chlorophenol (225%), and 4-methylphenol (185%). In total this special biosensor responds to eight phenols, including phenol, catechol, 2,4-dimethylphenol, and para-substituted phenols. The detection of meta- and ortho-substituted phenols is not possible due to the specificity of the catalytic process, which forms intermediate quinoid structures. The optimum pH of sensor operation is 7.0, within a narrow pH range of 6.5–7.2 where 90% of the maximum response is retained. The influence of temperature is restricted to the range 10–30 °C; below and above this range the enzyme activity decreases and a loss of steady-state stability occurs, respectively.

Long-term stability of 100 days is guaranteed when the sensor is stored at 5 °C in buffer solution or dry at −18 °C. Very short analysis times of 6 min including rinsing of the sensor and adjustment to the initial baseline, detection limits in the low microgram per liter range, and cost-effective production of biosensors – especially on the basis of screen-printing technology using polymer pastes – make this method attractive for field applications.

Tyrosinase is often implemented in phenol biosensors and co-immobilized with laccase. The range of detectable phenols is extended by additional responses from hydroquinone and 2-amino-4-chlorophenol.[112,113] For the analysis of vapors containing phenols an enzyme/gas-diffusion electrode with tyrosinase has been investigated that allowed the determination of phenol in the ppm range.[114]

Biosensors using cells of *Pseudomonas putida* have been designed for the determination of phenol, benzoic acid and their monochlorinated derivatives from water and hexane solutions.[115] This approach is based on oxygen consumption in relation to the analyte oxidation and, therefore, on the cells immobilized on an oxygen electrode. The detection limit obtained for phenol was 0.1 μM and for chlorophenols 8 μM. Different cell induction procedures influenced the sensor selectivity, thus the selective production of various sensor types should be possible in the future. The intrinsic specificity and rapid operation of biosensors make them potentially interesting for fast screening in field analysis and, in combination with SPE, for selective detection coupled to LC.[108,116]

A review covering the development of sensor arrays and their applications is given by Carey.[117]

6.2 Immunoassay

The immunoassay methods combine a rapid and inexpensive analysis with high analyte specifity at low expenditure in sample preparation.[118] They are promising for on-site measurements to select samples for more detailed analysis and for early-warning purposes. In the USA, the immunoassay approach has been implemented in many environmental pollution studies, together with the use of conventional techniques such as confirmation by GC/MS.[94]

Although considerable effort has been devoted to the development of immunoassays only a few are available, such as for PCP. The assay features high selectivity and low response to matrix influences.[119] Only for two tetrachlorophenols (2,4,5,6- and 2,3,4,6-) was significant cross-reactivity observed (54.1% and 15.1%, respectively), and a detection limit for PCP of 60 ng L^{-1} emphasizes suitability as a specific trace method. In comparative studies the immunoassay values were confirmed by accepted methods (GC/MS and HPLC). Further developments tend to assay applications for toxins including phenols in biological materials. 4-Nitro- and 2,4-dinitrophenol in urine and water samples could be detected with limits of 7 μg L^{-1}.[120] As this immunoassay is for LC detection, an 8–10-fold higher sensitivity is achieved than with UV detection. The adjustment of an immunoassay for HPLC detection is difficult to achieve because of the incompatibility of antibodies and organic solvents used for HPLC separation.

Immunoassays are best suited for the detection of selected phenols after ASE™ with water or SFE if a water-based collection solvent is used. Progress in biotechnological research and molecular design is likely to extend the development of immunoassays to a wider range of applications.

ACKNOWLEDGMENTS

The author is grateful to Dr P. Popp and S. Schrader (Centre for Environmental Research Leipzig-Halle) for critical revision and assistance in preparing this manuscript.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
ASE	Accelerated Solvent Extraction
CE	Capillary Electrophoresis
CID	Collision-induced Dissociation
CLP	Contract Laboratory Program
CZE	Capillary Zone Electrophoresis
DIN	German National Standards Organization
EC	European Community
ECD	Electron Capture Detection
ESE	Enhanced Solvent Extraction
ESI	Electrospray Ionization
FID	Flame Ionization Detection
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
ISO	International Standardization Organization
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
LSE	Liquid–Solid Extraction
MAE	Microwave-assisted Extraction
MAP	Microwave-assisted Processes
MISA	Municipal/Industrial Strategy for Abatement
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NIOSH	National Institute for Occupational Safety and Health
OSHA	Occupational Safety and Health Administration
PA	Polyacrylate
PCP	Pentachlorophenol
PDMS/DVB	Poly(dimethylsiloxane)/Divinyl-benzene
SDS	Sodium Dodecyl Sulfate
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TCL	Target Compound List
TFA	Trifluoroacetyl
TLC	Thin-layer Chromatography
TMS	Trimethylsilyl
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
VOC	Volatile Organic Compound
WHO	World Health Organization

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Immunoassay Techniques in Environmental Analyses • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Neutron Activation in Environmental Analysis • Nitroaromatics, Environmental Analysis of • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

REFERENCES

1. O. Lampire, I. Mila, M. Raminosoa, V. Michon, Ch. DuPenhoat, N. Faucheur, O. Laprevote, A. Scalbert, 'Polyphenols Isolated from the Bark of *Castanea sativa* Mill', *Phytochemistry*, **49**(2), 623–631 (1998).
2. Commission of the European Communities, 15. Guideline 92/86/EEC, 'Adjustment of Legislations of the European Communities on Cosmetics to the Technical Advance', ABI, EC No. L325, pp. 18, Brussels, 1992.
3. J. Falbe, M. Regitz (eds.), *Römpp Lexikon, Chemie*, Georg Thieme Verlag, Stuttgart, 3248, Vol. 4, 1998.
4. B. Baranowska-Dutkiewicz, 'Skin Absorption of Phenol from Aqueous Solutions in Men', *Int. Arch. Occup. Environ. Health*, **49**, 99–104 (1981).
5. International Agency for Research on Cancer (IARC), *Some Organic Solvents, Resin Monomers and Related Compounds, Pigments and Occupational Exposures in Paint Manufacture and Painting*, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No 47, IARC, Lyon, France, 263–287, 1989.
6. US Agency for Toxic Substance and Disease Registry, *Toxicological Profile for Penta-chlorophenol*, Draft Report, Atlanta, GA, 1992.
7. US Environmental Protection Agency, *Health Effects Assessment for Pentachlorophenol*, Report EPA/540/1–86/043, Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Office of Research and Development, Cincinnati, OH, 1988.
8. US Environmental Protection Agency, *Guidance for the Registration of Pesticide Products Containing Chlorobenzilate as the Active Ingredient*, Washington, DC, 6–7, 1983.
9. P.W.M. Augustijn-Beckers, A.G. Hornsby, R.D. Wauchope, 'SCS/ARS/CES Pesticide Properties Database for Environmental Decisionmaking II. Additional Compounds', *Rev. Environ. Contam. Toxicol.*, **137**, 6–16 (1994).
10. P.H. Howard (ed.), *Handbook of Environmental Fate and Exposure Data for Organic Chemicals. Pesticides*, Lewis Publishers, Chelsea, MI, 6–13, 1991.
11. T.A. Gasiewicz, 'Nitro Compounds and Related Phenolic Pesticides', in *Handbook of Pesticide Toxicology*, eds. J.W. Hayes, Jr, E.R. Laws, Jr, Academic Press, New York, 6–70, 1991.
12. L. Bretherick (ed.), *Hazards in the Chemical Laboratory*, 4th edition, The Royal Society of Chemistry, London, 315, 1986.
13. P. Mußmann, R. Eisert, K. Levsen, G. Wünsch, 'Determination of Nitrophenols, Diaminotoluenes, and Chloroaromatics in Ammunition Wastewater', *Acta Hydrochim. Hydrobiol.*, **23**, 13–19 (1995).
14. A.V. Krishnan, P. Starhis, S.F. Permuth, L. Tokes, D. Feldman, 'Bisphenol-A: an Estrogenic Substance is Released from Polycarbonate Flasks During Autoclaving', *Endocrinology*, **132**, 2279–2286 (1993).
15. M. Ahel, C. Schaffner, W. Giger, 'Behaviour of Alkylatphenol Polyethoxylate Surfactants in the Aquatic Environment. III. Occurrence and Elimination of their Persistent Metabolitea During Infiltration of River Water to Groundwater', *Water Res.*, **30**, 37–46 (1996).
16. M. Ahel, W. Giger, C. Schaffner, 'Behaviour of Alkylatphenol Polyethoxylate Surfactants in the Aquatic Environment. II. Occurence and Transformation in Rivers', *Water Res.*, **28**, 1143–1152 (1994).
17. L.B. Clark, R.T. Rosen, T.G. Hartman, J.B. Louis, I.H. Suffet, R.L. Lippincott, J.D. Rosen, 'Determination of Alkylphenol Ethoxylates and their Acetic Derivatives in Drinking Water by Particle Beam Liquid Chromatography/Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **147**, 167–180 (1992).

18. R. White, S. Jobling, S.A. Hoare, J.P. Sumpter, M.G. Parker, 'Environmental Persistent Alkylphenolic Compounds are Estrogenic', *Endocrinology*, **135**, 175–182 (1994).
19. A.M. Soto, C. Sonnenschein, K.L. Chung, M.F. Fernandez, N. Olea, F.O. Serrano, 'The E-SCREEN Assay as a Tool to Identify Estrogens: an Update on Estrogenic Environmental Pollutants', *Environ. Health Persp.*, **103**, 113–122 (1995).
20. S. Jobling, D. Sheahan, J.A. Osborne, P. Matthiessen, J.P. Sumpter, 'Inhibition of Testicular Growth in Rainbow-trout Exposed to Estrogenic Alkylphenolic Chemicals', *Environ. Toxicol. Chem.*, **15**, 194–202 (1996).
21. Communication from the Commission to the Council and the European Parliament, 'Community Strategy for Endocrine Disrupters – a range of substances suspected of interfering with the harmone systems of humans and wildlife', Ed. Commission of the European Communities, COM 706, p. 5, Brussels, 1999.
22. US Environmental Protection Agency, *Emergency Planning and Community Right-to-know*, EPA, Office of Pollution Prevention and Toxics, Washington, section 313, 1996.
23. US Environmental Protection Agency, *Method 625 – Base/neutrals and Acids in Federal Register*, EPA, Washington, DC, Part VIII, 40 CFR Part 136, 153–174, 1984.
24. US Environmental Protection Agency, *EPA Method 8041, Phenols by Gas Chromatography, Capillary Column Technique, EPA*, Washington, DC, 1–28, 1995.
25. H.H. Dieter (ed.), *New Drinking Water Guide Values of Chemical Substances*, WHO, Bundesgesundheitsamt, Bundesgesundheitsblatt, 36/5, 1993 Tab. 3: Pesticides, 1993.
26. M.C. Hennion, V. Pichon, D. Barceló, 'Surface Water Analysis (Trace-organic Contaminants) and EC Regulations', *Trends Anal. Chem.*, **13**(9), 361–372 (1994).
27. International Standardization Organization, *Water Quality Determination of Selected Monovalent Phenols. Gas Chromatographic Methods after Enrichment by Extraction*, ISO-8165-1, release 01.92, ISO, Beuth Verlag, Berlin, 1992.
28. US Environmental Protection Agency, *Methods for Organic Chemical Analysis of Municipal and Industrial Wastewater, Method 604 – Phenols in Federal Register*, part VIII, 40 CFR Part 136, EPA, Washington, DC, 58–66, 1984.
29. V. Lopez-Avila, N.S. Dodhiwala, J. Milanes, W.F. Beckert, 'Evaluation of EPA Method 8120 for Determination of Chlorinated Hydrocarbons in Environmental Samples', *J. Assoc. Off. Anal. Chem.*, **72**, 593–602 (1989).
30. US Environmental Protection Agency, *US EPA Method 600/4-81-055, Interim Methods for the Sampling and Analysis of Priority Pollutants in Sediments and Fish Tissue*, EPA, Washington, DC, section 3.1.3, 1988.
31. F. Tschochner, W. Pilz-Mittenburg, T. Benz, H. Brunner, W. Jäger, H. Hagemaier, 'Determination of Chlorophenols in Aqueous, Solid and Gas Samples by GC/ECD and GC/MS', *Z. Wasser-Abwasser-Forsch.*, **22**, 267–271 (1989).
32. US Environmental Protection Agency, *Method 8040 – Phenols*, EPA, Washington, DC, 1986.
33. US Environmental Protection Agency, Centre for Environmental Research Information, Office of Research and Development, *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air, Compendium Method TO-8*, 2nd edition, EPA, Washington, DC, 1988.
34. US Environmental Protection Agency, *Compendium Method TO-17, Determination of Volatile Organic Compounds in Ambient Air Using Active Sampling onto Sorbent Tubes*, EPA, Cincinnati, 1997.
35. US Environmental Protection Agency, *Compendium Method TO-10A, Determination of Organochlorine Pesticides in Ambient Air*, EPA, Cincinnati, 1997.
36. H. Kontsas, Ch. Rosenberg, P. Jäppinen, M.-L. Riekkola, 'Glass Fiber Prefilter-XAD-2 Sampling and Gas Chromatographic Determination of Airborne Chlorophenols', *J. Chromatogr.*, **636**, 255–261 (1993).
37. B.S.K. Buisson, P.W.W. Kirk, J.N. Lester, 'Determination of Chlorinated Phenols in Water, Wastewater, and Wastewater Sludge by Capillary GC/ECD', *J. Chromatogr. Sci.*, **22**, 339–342 (1984).
38. Association Frangeais de Normalisation, *Recueil de Normes Francaises, Qualité de L'eau-environment*, section 90–109, AFNOR, Paris, 510, 1994.
39. S. Dupeyron, M. Astrue, M. Marbach, 'Comparison of the Phenol Index, an Automated SPE/HPLC and a Destillation/HPLC Method', *Analusis*, **23**, 474–476 (1995).
40. S.S. Prasad, 'Trends in Quality Assurance in the Characterization of Chemical Comtaminants in the Environment', *Trends Anal. Chem.*, **13**(4), 157–162 (1994).
41. DIN 38 409 H 16, *German Standard Methods for the Examination of Water, Wastewater and Sludge, General Measures of Effects and Substances (Group H), Determination of the Phenol Index (H 16)*, VCH Verlagsgesellschaft, Weinheim, 2–17, Vol. II, 1984.
42. *Handbook of Photometrical Operation Analysis*, 'Phenol Index', LCK® 345 and LCK® 346, Application note 93/10, Dr. Bruno Lange Ltd., Berlin, Vol. 10, 1993.
43. Y.I. Korenman, T.A. Kuchmenko, S.A. Karavaev, 'Extraction-spectrophotometric Determination of Phenol Using a Reaction with 4-Aminoantipyrine in the Two-phase Poly(ethyleneglycol)–Ammonium Sulfate–Water System', *J. Anal. Chem.*, **53**(3), 258–263 (1998).
44. J.C.E.G. da Silva, M.C.P.O.M. Laquipai, 'Method for Rapid Screening of Chlorophenols Using a Reduced Calibration Set of UV Spectra and Multivariate Calibration Techniques', *Anal. Lett.*, **31**(14), 2549–2563 (1999).

45. D. Puig, D. Barceló, 'Determination of Phenolic Compounds in Water and Wastewater', *Trends Anal. Chem.*, **15**(8), 362–375 (1996).
46. R. Schuster, 'Phenols', in *Water Analysis. Organic Micropollutants*, eds. R. Soniassy, P. Sandra, C. Schlett, Hewlett-Packard, Waldbronn, 141–162, 1994.
47. D. Puig, D. Barceló, 'Comparison of Different Sorbent Materials for On-line Liquid–Solid Extraction followed by Liquid Chromatographic Determination of Priority Phenolic Compounds in Environmental Waters', *J. Chromatogr. A*, **733**, 371–381 (1986).
48. J.S. Fritz, P.J. Dumont, L.W. Smitd, 'Methods and Materials for Solid-phase Extraction', *J. Chromatogr. A*, **691**, 133–140 (1995).
49. V. Coquart, M.-C. Hennion, 'Trace-level Determination of Polar Phenolic Compounds in Aqueous Samples by High-performance Liquid Chromatography and On-line Preconcentration on Porous Graphitic Carbon', *J. Chromatogr.*, **600**, 195–201 (1992).
50. A. Di Corcia, A. Bellioni, M.D. Madbouly, S. Marchese, 'Trace Determination of Phenols in Natural Waters. Extraction by a new Graphitized Carbon Black Cartridge followed by Liquid Chromatography and Reanalysis after Phenol Derivatization', *J. Chromatogr. A*, **733**, 383–393 (1996).
51. E. Pocorull, R.M. Marcé, F. Borull, M.L. Serna, L. Toribio, J.L. Bertnal, 'Determination of Phenolic Compounds in Natural Waters by Liquid Chromatography with Ultraviolet and Electrochemical Detection after On-line Trace Enrichment', *J. Chromatogr. A*, **738**, 1–9 (1996).
52. K. Blau, J.M. Halket (eds.), *Handbook of Derivatives for Chromatography*, John Wiley & Sons, New York, 1993.
53. J.A. Perry, C.A. Feit, 'Derivatization Techniques in Gas–Liquid Chromatography', in *GLC and HPLC Determination of Therapeutic Agents*, eds. K. Tsuji, W. Morowich, Dekker, New York, Part I, 1978.
54. C.F. Poole, A. Zlatkis, 'Derivatization Techniques for the Electron Capture Detector', *Anal. Chem.*, **52**, 1002A (1980).
55. J. Cheung, R.J. Wells, 'Analysis of Phenolic Compounds in Effluent by Solid-phase Extraction and Gas Chromatography Mass Spectrometry with Direct On-column Benzylation: Sensitive Negative ion Chemical Ionization Gas Chromatography Mass Spectrometry Detection of Phenyl Benzyl Ethers', *J. Chromatogr. A*, **771**(1–2), 203–211 (1997).
56. M. Möder, D. Zimmer, J. Stach, R. Herzschuh, 'GC/MS and MS/MS Investigations of Derivatized Hydroxyaromatic Compounds from Coal-derived Liquids', *Fuel*, **68**, 1422–1428 (1989).
57. G. Lamprecht, J.F.K. Huber, 'Ultra-trace Analysis of Phenols in Water Using High-performance Liquid Chromatography with On-line Reaction Detection', *J. Chromatogr. A*, **667**, 47–54 (1994).
58. S. Dupeyron, M. Astrue, M. Marbach, 'Automated Solid-phase Extraction for Routine Determination of Phenol and Chlorophenols at Trace Levels in Water by High-performance Liquid Chromatography', *Analusis*, **23**, 470–473 (1995).
59. J. Pawliszyn, *Solid Phase Microextraction – Theory and Practice*, Wiley-VCH, Weinheim, 1997.
60. C.L. Arthur, J. Pawliszyn, 'Solid-phase Microextraction with Thermal Desorption Using Fused Silica Optical Fibers', *Anal. Chem.*, **62**, 2145–2148 (1990).
61. D. Louch, S. Motlagh, J. Pawliszyn, 'Dynamics of Organic Compound Extraction from Water Using Liquid-coated Fused Silica Fibers', *Anal. Chem.*, **64**, 1187–1199 (1992).
62. C.L. Arthur, D.W. Potter, K.D. Buchholz, S. Motlagh, J. Pawliszyn, 'Solid-phase Microextraction for the Direct Analysis of Water: Theory and Practice', *LC-GC*, **10**, 656–661 (1992).
63. Z. Zhang, M.J. Yang, J. Pawliszyn, 'Solid-phase Microextraction Integrates Sampling, Extraction, Concentration, and Sample Introduction into a Single Step', *Anal. Chem.*, **66**, 844A (1994).
64. A.A. Boyd-Boland, M. Chai, Y.Z. Luo, Z. Zhang, M.J. Yang, J. Pawliszyn, T. Gorecki, 'New Solvent-free Sample Preparation Techniques based on Fiber and Polymer Technologies', *Environ. Sci. Technol.*, **28**, 569A–574A (1994).
65. K.D. Buchholz, J. Pawliszyn, 'Determination of Phenols by Solid-phase Microextraction and Gas Chromatographic Analysis', *Environ. Sci. Technol.*, **27**, 2844–2848 (1993).
66. K.D. Buchholz, J. Pawliszyn, 'Optimization of Solid-phase Microextraction Conditions for Determination of Phenols', *Anal. Chem.*, **66**, 160–167 (1994).
67. M. Möder, S. Schrader, U. Franck, P. Popp, 'Determination of Phenolic Compounds in Wastewater by Solid-phase Microextraction', *Fresenius' J. Anal. Chem.*, **357**, 326–332 (1997).
68. J.L. Hedrick, L.J. Mulcahey, L.T. Taylor, 'Supercritical Fluid Extraction – a Fundamental Review', *Mikrochim. Acta*, **108**, 115–132 (1992).
69. I.J. Barnabas, J.R. Dean, S.P. Owen, 'Supercritical Fluid Extraction of Analytes from Environmental Samples', *Analyst*, **19**, 2381–2394 (1994).
70. S. Bøwadt, S.B. Hawthorne, 'Supercritical Fluid Extraction in Environmental Analysis – a Review', *J. Chromatogr. A*, **703**, 549–571 (1995).
71. H.B. Lee, T.E. Peart, R.L. Hong-You, 'Determination of Phenolics from Sediments of Pulp Mill Origin by *in situ* Supercritical Carbon Dioxide Extraction and Derivatization', *J. Chromatogr.*, **636**, 263–270 (1993).
72. S.B. Hawthorne, Y. Yang, D.J. Miller, 'Extraction of Organic Pollutants from Environmental Solids with Sub- and Supercritical Water', *Anal. Chem.*, **66**, 2912–2920 (1994).

73. Dionex Corporation, Application Note 317, 'Extraction of Basic, Neutral and Acidic Environmental Relevant Compounds by ASE™ (accelerated solvent extraction)', Sunnyvale, CA, 1995.
74. A. Kreisselmeier, H.W. Dürbeck, 'Determination of Alkylphenols, Alkylphenolethoxylates and Linear Alkylbenzenesulfonates in Sediments by Accelerated Solvent Extraction and Supercritical Fluid Extraction', *J. Chromatogr. A*, **775**, 187–196 (1997).
75. L. Wennrich, P. Popp, M. Möder, 'Determination of Chlorophenols in Soil Using Accelerated Solvent Extraction Combined with Solid-phase Microextraction', *Anal. Chem.*,**72**, 546–551 (2000).
76. S.B. Hawthorne, J.J. Langenfeld, D.J. Miller, M.D. Burford, 'Comparison of Supercritical $CHClF_2$, N_2O, and CO_2 for the Extraction of Polychlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **64**(14), 1614–1622 (1992).
77. S.B. Hawthorne, D.J. Miller, M.D. Burford, J.J. Langenfeld, S. Eckert-Tilotta, P.K. Louie, 'Factors Controlling Quantitative Supercritical Fluid Extraction of Environmental Samples', *J. Chromatogr.*, **642**, 301–317 (1993).
78. J.J. Langenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Effects of Temperature and Pressure on Supercritical Fluid Extraction Efficiencies of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls', *Anal. Chem.*, **65**, 338–344 (1993).
79. S. Bøwadt, B. Johansson, 'Analysis of PCBs in Sulfur-containing Sediments by Off-line Supercritical Fluid Extraction and HRGC-ECD', *Anal. Chem.*, **66**, 667–673 (1994).
80. S. Bøwadt, B. Johansson, S. Wunderli, M. Zennegg, L.F. de Alencastro, D. Grandjean, 'Independent Comparison of Soxhlet and Supercritical Fluid Extraction for the Determination of PCBs in an Industrial Soil', *Anal. Chem.*, **67**, 2424–2430 (1995).
81. I.J. Barnabas, J.R. Dean, W.R. Tomlinson, S.P. Owen, 'Experimental Design Approach for the Extraction of Polycyclic Aromatic Hydrocarbons from Soil Using Supercritical Carbon Dioxide', *Anal. Chem.*, **67**, 2064–2069 (1995).
82. V. Seidel, W. Lindner, 'Evaluation of a Supercritical Fluid Extraction Method for Hexachlorobenzene from Artificially Spiked and Naturally Contaminated Oil Seeds and Soil Samples', *Int. J. Environ. Anal. Chem.*, **59**, 1–13 (1995).
83. S. Dupeyron, P.M. Dudermel, D. Couturier, 'Focused Microwave Assisted Extraction (FMAE) of Polynuclear Aromatic Hydrocarbons from Contaminated Soil: Role of Acetone and Water Content Impact on Microwave Efficiency', *Analusis*, **25**, 286–292 (1997).
84. J.R.J. Paré, J.M.R. Bélanger, S.S. Stafford, 'Microwave-assisted Process (MAP™): a New Tool for the Analytical Laboratory', *Trends Anal. Chem.*, **13**(4), 179–184 (1994).
85. A. Egizabal, O. Zuloaga, N. Etxebarria, L.A. Fernández, J.M. Madariaga, 'Comparison of Microwave-assisted Extraction and Soxhlet Extraction for Phenols in Soil Samples Using Experimental Design', *Analyst*, **123**, 1679–1684 (1998).
86. J.S. Warner, M.C. Landes, L.E. Slivon, 'Development of a Solvent Extraction Method for Determining Semivolatile Organic Compounds in Solid Wastes', in *Hazardous and Industrial Solid Waste Testing: Second Symposium*, eds. R.A. Conway, W.P. Gulledge, American Society for Testing and Materials, Philadelphia, PA, 203–213, 1983.
87. M. Valcárcel, E. Ballestero, M. Gallego, 'Continuous Liquid–Liquid Extraction and Derivatization Module Coupled On-line with Gas Chromatographic Detection', *Trends Anal. Chem.*, **13**(2), 68–73 (1994).
88. E. Woolfenden, 'Monitoring VOCs in Air Using Sorbent Tubes followed by Thermal Desorption–Capillary GC Analysis: Summary of Data and Practical Guidelines', *J. Air Waste Manage. Assoc.*, **47**, 20–36 (1997).
89. B. Thiele, K. Günther, M.J. Schwuger, 'Alkylphenol Ethoxylates: Trace Analysis and Environmental Behaviour', *Chem. Rev.*, **97**(8), 3247–3272 (1997).
90. P. Jones, G. Nickless, 'Characterization of Non-ionic Detergents of the Polyethoxylated type from Water Systems II, *J. Chromatogr.*, **156**, 99–110 (1978).
91. D. Puig, D. Barceló, 'Off-line and On-line Solid-phase Extraction Followed by Liquid Chromatography for the Determination of Priority Phenols in Natural Waters', *Chromatographia*, **40**, 435–441 (1995).
92. M.T. Galcerán, O. Jáuregui, 'Determination of Phenols in Sea Water by Liquid Chromatography with Electrochemical Detection after Enrichment by Using Solid Phase Extraction Cartridges and Disks', *Anal. Chim. Acta*, **304**, 75–80 (1995).
93. D. Puig, D. Barceló, 'Comparative Study of On-line Solid Phase Extraction followed by UV and Electrochemical Detection in Liquid Chromatography for the Determination of Priority Phenols in River Water Samples', *Anal. Chim. Acta*, **311**, 63–69 (1995).
94. R.J. Vreeken, U.A.Th. Brinkman, G.J. Jong, D. Barceló, 'Chloroacetonitrile as Eluent Additive in Thermospray Liquid Chromatography/Negative Ion Mass Spectrometry for the Characterization of Organic Pollutants', *Biomed. Environ. Mass Spectrom.*, **19**, 481–492 (1990).
95. A. Cappiello, G. Famiglini, P. Palma, A. Berloni, F. Bruner, 'New Approach for the Analysis of Acidic Pesticides in Water by LC/MS with a Particle Beam Interface', *Environ. Sci. Technol.*, **29**, 2295–2303 (1995).
96. D. Puig, I. Silgoner, M. Grasserbauer, D. Barceló, 'Comparision of Three Different Liquid Chromatography-mass Spectrometry Interfacing Techniques for the Determination of Priority Phenolic Compounds in Water', *J. Mass. Spectrom.*, **31**, 1297–1308 (1996).

97. M.C. Alonso, D. Puig, I. Silgoner, M. Grasserbauer, D. Barceló, 'Determination of Priority Phenolic Compounds in Soil Samples by Various Extraction Methods Followed by Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **823**(1–2), 231–239 (1998).
98. J.M.E. Quirke, C.L. Adams, G.J. Van Berkel, 'Chemical Derivatization for Electrospray Ionization Mass Spectrometry. 1. Alkyl Halides, Alcohols, Phenols, Thiols and Amines', *Anal. Chem.*, **66**, 1302 (1994).
99. C. Cresenzi, A. Di Corcia, S. Marchese, R. Samperi, 'Determination of Acidic Pesticides in Water by a Benchtop Electrospray Liquid Chromatography Mass Spectrometer', *Anal. Chem.*, **67**, 1968–1975 (1995).
100. P. Praus, V. Dombek, 'Utilization of Polyethylene Glycol for the Separation of Chlorophenols by Capillary Isotachophoresis', *Anal. Chim. Acta*, **283**, 917–922 (1993).
101. M. Aguilar, A. Farran, V. Marti, 'Analysis of Phenoxyalkyl Acid Herbicides and Chlorophenols by Capillary Zone Electrophoresis', *Sci. Total Environ.*, **132**, 133–139 (1993).
102. J. Wang, Y. Lin, 'On-line Organic-phase Enzyme Detector', *Anal. Chim. Acta*, **271**, 53–56 (1993).
103. G. Marko-Varga, J. Emnéus, L. Gorton, T. Ruzgas, 'Development of Enzyme-based Amperometric Sensors for the Determination of Phenolic Compounds', *Trends Anal. Chem.*, **14**, 319–327 (1995).
104. H.W. Duckworth, J.E. Coleman, 'Physicochemical and Kinetic Properties of Mushroom Tyrosinase', *J. Biol. Chem.*, **245**, 1613–1625 (1970).
105. B. Gründig, B. Strehlitz, C. Krabisch, H. Thielemann, H. Kotte, M. Gomoll, H. Kopinke, J. Pitzler, in *Biosensors: Fundamentals, Technologies and Applications*, eds. R.D. Schmidt, F. Scheller, GBF-Monographs, No. 17, VCH, Weinheim, 275–285, 1992.
106. J. Kulys, R.D. Schmid, 'A Sensitive Enzyme Electrode for Phenol Monitoring', *Anal. Lett.*, **23**, 589–597 (1990).
107. J. Svitel, S. Miertus, 'Development of Tyrosinase Based Biosensor and its Application for Monitoring of Bioremediation of Phenol and Phenolic Compounds', *Environ. Sci. Technol.*, **32**(6), 828–832 (1998).
108. F. Ortega, E. Domìnguez, E. Burestedt, J. Emnéus, L. Gorton, G. Marko-Varga, 'Phenol Oxidase-based Biosensors as Selective Detection Units in Column Liquid Chromatography for the Determination of Phenolic Compounds', *J. Chromatogr. A*, **675**, 65–68 (1994).
109. J. Wang, K. Varughese, 'Polishable and Robust Biological Electrode Surfaces', *Anal. Chem.*, **62**, 318–324 (1990).
110. Y.D. Zou, J.Y. Mo, 'The 2.5th Order Differential Voltammetric Determination of Phenol with a Composite Carbon Paste Polyamide Electrode', *Anal. Chim. Acta*, **353**(1), 71–78 (1997).
111. H. Kotte, B. Gründig, K.-D. Vorlop, B. Strelitz, U. Stottmeister, 'Methylphenazonium-modified Enzyme Sensor Based on Polymer Thick Films for Subnanomolar Detection of Phenols', *Anal. Chem.*, **34**, 65–70 (1995).
112. J. Wang, L. Fang, D. Lopez, 'Amperometric Biosensor for Phenols Based on a Tyrosinase–Graphite–Epoxy Biocomposite', *Analyst*, **119**, 455–459 (1994).
113. M. Lutz, E. Burestedt, J. Emnéus, H. Lidén, S. Gobhadi, L. Gorton, G. Marko-Varga, 'Effects of Different Additives on a Tyrosinase Based Carbon Paste', *Anal. Chim. Acta*, **305**, 8–14 (1995).
114. A. Kaisheva, I. Illiev, S. Christoph, R. Kazareva, 'Electrochemical Gas Biosensor for Phenol', *Sens. Actuators, B*, **44**(1–3), 571–577 (1997).
115. T.T. Bachmann, U. Bilitewski, R.D. Schmid, 'A Microbial Sensor Based on *Pseudomonas putida* for Phenol, Benzoic Acid and their Monochlorinated Derivatives which can be Used in Water and *n*-Hexane', *Anal. Lett.*, **31**(14), 2361–2373 (1998).
116. E. Burestedt, J. Emnéus, L. Gorton, G. Marko-Varga, E. Domìnguez, F. Ortega, A. Narváez, H. Irth, M. Lutz, D. Puig, D. Barceló, 'Optimization and Validation of an Automated Solid Phase Extraction Technique Coupled On-line to Enzyme-based Biosensor Detection for the Determination of Phenolic Compounds in Surface Water Samples', *Chromatographia*, **41**, 207–215 (1995).
117. W.P. Carey, 'Multivariate Sensor Arrays as Industrial and Environmental Monitoring Systems', *Trends Anal. Chem.*, **13**(5), 210–217 (1994).
118. A. Brecht, R. Abuknesha, 'Multi-analyte Immunoassays Application to Environmental Analysis', *Trends Anal. Chem.*, **14**, 361–371 (1995).
119. C.S. Hottenstein, S.W. Jourdan, M.C. Hayes, F.M. Rubio, D.P. Herzog, T.S. Lawruk, 'Determination of Pentachlorophenol in Water and Soil by a Magnetic Particle-based Enzyme Immunoassay', *Environ. Sci. Technol.*, **29**, 2754–2759 (1995).
120. Q.X. Li, M.S. Zhao, S.G. Gee, M. Kurth, J. Seiber, B.D. Hammock, 'Development of Enzyme-linked Immunosorbent Assays for 4-Nitrophenol and Substituted 4-Nitrophenols', *J. Agric. Food Chem.*, **39**, 1685–1691 (1991).

Polychlorinated Biphenyls Analysis in Environmental Samples

Bert van Bavel, Mats Tysklind, and Gunilla Lindström
Umeå University, Umeå, Sweden

1 Introduction 1
2 A Short History of Polychlorinated Biphenyl Analyses in Environmental Samples 1
3 State-of-the-art Analytical Methodologies 2
3.1 Detection Methods 5
3.2 Sample Enrichment and Clean-up 7
3.3 Extraction 11
4 Recent Developments in Analytical Methodology 12
4.1 Supercritical Fluid Extraction Coupled with Liquid Chromatography 12
4.2 Gas Chromatography Coupled with Time-of-flight Mass Spectrometry 13
5 Quality Assurance and Quality Control by International Comparison 15
Abbreviations and Acronyms 16
Related Articles 16
References 17

The analysis of polychlorinated biphenyls (PCBs) in a wide variety of environmental samples is discussed, starting with the more traditional methodology and moving on to more recent developments that are expected to continue in the twenty-first century. Several detection techniques, including gas chromatography (GC) coupled with electron capture detection (ECD) and mass spectrometry (MS), are described in the first part of this article. The use of MS as a detection technique is discussed in detail, from both a qualitative and a quantitative point of view. To a large extent the detection technique used determines the sample pretreatment and clean-up, and in later paragraphs column clean-up on adsorbents is discussed, including fat removal, alumina column chromatography and planar separation on carbon. A new methodology for the analysis of PCBs in environmental samples is also described. Both speed of analysis using gas chromatography/time-of-flight mass spectrometry (GC/TOFMS) and environmentally friendly analysis using supercritical CO_2 are discussed in detail. In the last section several aspects of quality assurance/quality control (QA/QC) are discussed using an international intercalibration study as an example.

1 INTRODUCTION

Production of PCBs, which are characterized by containing several of the 209 possible chlorination products of biphenyl, occurred mainly from 1930 to the late 1970s. An overall world production until 1980 of 2×10^9 kg has been estimated.[1] As early as 1936 workplace threshold limits were set due to toxic effects observed in connection with occupational exposure. But the widespread environmental contamination and its effects were not fully accounted for until the late 1960s, when the global occurrence of PCBs in all biological material was shown to be a fact. In the 1960s chemists using the newly introduced GC techniques to analyze chlorinated pesticides present in the lipid fraction of biological materials found interfering peaks in the chromatograms which soon were identified as being caused by bio-magnifying PCB residues.[2,3] These findings eventually led to restrictions in the 1970s, in both the use and production of PCBs in most countries. Measures were also taken to minimize the generation and distribution of PCBs from secondary sources. Today, some thirty years and tens of thousands of PCB analyses later, our analytical techniques are tuned to detect each and every individual PCB congener down to parts per trillion (ppt) levels. And due to extensive efforts in combating the PCB problem since the 1980s, declining PCB trends are reported.

2 A SHORT HISTORY OF POLYCHLORINATED BIPHENYL ANALYSES IN ENVIRONMENTAL SAMPLES

The Swedish researcher Sören Jensen[2,4,5] discovered PCBs in the environment using a packed column gas chromatograph with an electron capture detector. Traditionally, therefore, GC with ECD is used to analyze PCBs. Packed columns were originally used until capillary columns were introduced at the start of the 1980s. Major environmental monitoring studies were done using packed-column gas chromatography/electron capture detection (GC/ECD). Quantification was achieved by comparison with Aroclor™ mixtures,[6] total areas analysis or quantification of individual congeners.[7] While

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

changing from packed-column GC/ECD to capillary high-resolution gas chromatography (HRGC), problems involving quantification arise. Results acquired by the different methods are not directly comparable because of several basic differences. The main problem with using packed columns is the poor resolution of the PCBs. PCBs are represented as a poorly resolved series of peaks and reported as a total area or related to a commercial PCB mixture. The introduction of capillary gas columns improved the resolution and made it possible to produce congener-specific data. The next problem was that it was difficult to compare the old packed-column data with the newer high-resolution column data. The older methods used Aroclor™ mixtures as the external standards, based on specific peaks[8] or total area.[9] Although these methods can be applied with capillary gas chromatographic data, the errors in the determination become larger and more obvious due to differences in PCB pattern between the different samples. Thus it is difficult to compare levels of PCBs measured and quantified with different methods. Although largely dependent on the sample matrix, PCBs measured as the total area overestimate the ''real'' levels due to interferences (chlorinated pesticides or other halogenated compounds). The methods based on specific peak quantification generally result in higher levels[10] for the packed column analysis although in some cases good agreement[11,12] is observed. While capillary or HRGC columns offer a more specific and in most cases a more sensitive analysis, ECD was often used as the detection method. This technique specifically detects halogenated compounds and although interferences of nonhalogenated compounds are eliminated, it is still possible that halogenated compounds eluting at the same time are interfering. A large number of chlorinated pesticides and brominated flame retardants are known to elute within the same time-window as PCBs. Often the HRGC resolution will be enough to separate these compounds from the PCBs, but depending on the levels the interferences might cause erratic, higher, values for some of the PCB congeners. The quest for more specificity and in some cases more sensitivity has resulted in the use of MS as the detection technique instead of ECD. The specificity is enhanced by monitoring the molecular ions of the PCBs in combination with their unique molecular chlorine clusters. Gas chromatography/mass spectrometry (GC/MS) in combination with the use of ^{13}C-labeled internal standards has become the preferred detection method using ion-trap or quadrupole MS. High-resolution mass spectrometry (HRMS) using magnetic sector instruments is nowadays often used for specific applications (planar PCBs) to further enhance specificity and detection levels.

3 STATE-OF-THE-ART ANALYTICAL METHODOLOGIES

With respect to the analysis of PCBs it is important to note that there are 209 different PCB compounds, with different properties and environmental behaviors. The basic chemical structure is given in Figure 1 and chlorine substitution at the different positions is indicated by a number on the first ring (1–5) and/or the second ring (1′–5′) as illustrated. To facilitate the reporting of the different PCB congeners, Ballschmiter and Zell organized the 209 PCBs into numerical order and assigned IUPAC (International Union of Pure and Applied Chemistry) numbers from 1 to 209.[13] Throughout this article these so-called Ballschmiter numbers are used for identification of the different PCBs. One example of the different properties of the 209 PCBs is the lower bioaccumulation potential of the less chlorinated PCBs as compared with the more chlorinated PCBs. More subtle substitution differences, like chlorine substitution in the *meta* or *para* position, will also influence bioaccumulation.[14] Even more important with respect to the analysis of PCBs in environmental samples are the differences in toxicity of the different congeners. PCBs that can adopt a planar or coplanar configuration show dioxin-like toxicity and are known to inhibit 7-ethoxyresorufin-*O*-deethylase (EROD) activity.[15] It might therefore be important to specifically analyze these isomers as well as the most abundant PCBs or total PCB levels. A listing of the dioxin-like PCBs according to the World Health Organization (WHO)[16] is given in Table 1. This table shows that PCBs with no or only one chlorine in the *ortho* position are assigned a so-called TEF value in relation to the toxicity of 2,3,7,8-tetradibenzo-*p*-dioxin. Toxicological considerations are not the only reason for congener-specific PCB analysis; the large differences in PCB patterns in environmental samples are another reason. Because of these large differences in the PCB patterns it is not very useful to report a sum parameter, the sum of only seven isomers, or to relate the levels to a commercial PCB oil mixture (such as Aroclor™). The differences in PCB patterns are illustrated in Figure 2, where the PCB patterns of both human and whale samples are shown. The whale sample contains nearly

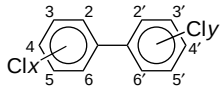

Figure 1 Chemical structure of the PCBs and designation of the Cl-substituted sites on the two aromatic rings. The prime (′) indicates the first or the second aromatic ring. The *ortho* position corresponds with chlorine substitution at 2, 2′, 6 or 6′, the *meta* position with chlorine substitution at 3, 3′, 5 or 5′, and the *para* position with chlorine substitution at 4 or 4′.

Table 1 WHO toxic equivalency factors (TEFs) for humans, mammals, fish and bird[16]

Congener	Humans/ mammals	Fish	Birds
2,3,7,8-TCDD	1	1	1
Non-o-PCBs			
3,4,4′,5-Tetra-PCB (#81)	0.0001	0.0005	0.1
3,3′,4,4′-Tetra-PCB (#77)	0.0001	0.0001	0.05
3,3′,4,4′,5-Penta-PCB (#126)	0.1	0.005	0.1
3,3′,4,4′,5,5′-Hexa-PCB (#169)	0.01	0.00005	0.001
Mono-o-PCBs			
2,3,3′,4,4′-Penta-PCB (#105)	0.0001	<0.000005	0.0001
2,3,4,4′,5-Penta-PCB (#114)	0.0005	<0.000005	0.0001
2,3′,4,4′,5-Penta-PCB (#118)	0.0001	<0.000005	0.00001
2′,3,4,4′,5-Penta-PCB (#123)	0.0001	<0.000005	0.00001
2,3,3′,4,4′,5-Hexa-PCB (#156)	0.0005	<0.000005	0.0001
2,3,3′,4,4′,5′-Hexa-PCB (#157)	0.0005	<0.000005	0.0001
2,3′,4,4′,5,5′-Hexa-PCB (#167)	0.00001	<0.000005	0.00001
2,3,3′,4,4′,5,5′-Hepta-PCB (#189)	0.0001	<0.000005	0.00001

TCDD, tetrachlorodibenzo-*p*-dioxin.

every PCB present in commercial PCB oils. In the human sample several PCBs present in a commercial mixture are metabolized. Although within species and within samples from the same origin sum parameters might be useful, this is certainly not the case when different sample types are to be compared. The emphasis is thus on congener-specific PCB analysis.

Nowadays most analytical methods use capillary GC columns to achieve gas chromatographic separation of the PCBs. Generally a nonpolar DB-5™ column (5% diphenyl dimethyl siloxane) is used for analysis with different column diameters (0.15–0.50 mm), film thicknesses (0.1–0.5 μm) and lengths (10–60 m). Different columns have been used and the retention order and relative retention times of the 209 PCBs on a variety of columns are now known.[17,18] In some cases it is necessary to use another column to resolve co-elution or interference. One should be aware that the common use of the DB-5™ column in analysis has resulted in a generalization of the results. For example PCB #138 is reported in numerous reports, but it has been shown that this isomer co-elutes with PCB #163 (10–30% in environmental samples).[19] In many reports PCB #138 is measured on a DB-5™-like column and thus actually reports the sum of PCB #138 and PCB #163. Several other co-elutions occur, and in some cases it is thus necessary to use two different GC columns.

ECD or MS is generally used for detection after GC separation. Depending on the type of samples and the questions to be answered the analytical method can be adapted. Although sample clean-up and extraction are important, the most important factor in choosing the analytical method is the detection. For many purposes ECD will be sufficient, but more and more MS detection is being favored. ECD is specific to halogen-containing molecules, and thus shows good sensitivity for the analysis of PCBs. It is also easily coupled to capillary columns and is fast enough to measure the PCBs eluting from the GC column even when fast GC and short columns are used. The disadvantage is that an ECD detector is unable to separate halogen compounds eluting at the same time as the target compound. For most applications ECD might still be good enough with the use of internal or external standard methods. Note that if an internal standard is added before sample extraction and clean-up using ECD, it must be a compound containing a halogen that is not present in the sample, and in some way resembles the analytical behavior of PCBs. With regard to the type of samples such a compound might be hard to find and often compromises have to be made.

Recently MS has gained in popularity due to the development of easy-to-use, benchtop instruments, most prominently low-resolution quadrupole and ion-trap instruments. With a low-resolution mass spectrometer it is possible to monitor specific masses, in most cases the molecular ion, to detect the target compounds. One specific property of chlorine, the existence of the natural isotopes ^{35}Cl and ^{37}Cl in a ratio of 3 : 1, can be used to identify or confirm the presence of any compound containing chlorine. An example of the typical chlorine cluster is given in Figure 3. A compound containing chlorine atoms shows a typical mass cluster and for PCBs this results in very specific mass spectra with peaks at the molecular mass (m) and two mass units higher ($m + 2$). Theoretically the ratio $(m)/(m + 2)$ can be calculated and thus used to identify PCBs or as an extra QA/QC criterion. Another advantage of the use of MS is that ^{13}C-labeled internal standards can be utilized during the sample extraction and clean-up. A ^{13}C-labeled PCB is a synthesized stable isotope of the PCB, which exhibits the same chemical behavior as the ''native'' PCB. This implies that the ^{13}C-labeled internal standards behave chemically in exactly the same way as the ''native'' PCBs. The only difference is in the mass, which is 12 units higher. The mass spectrometer can detect this difference in mass. This technique, which is called isotope dilution, allows the analytical chemist to use miscellaneous extraction

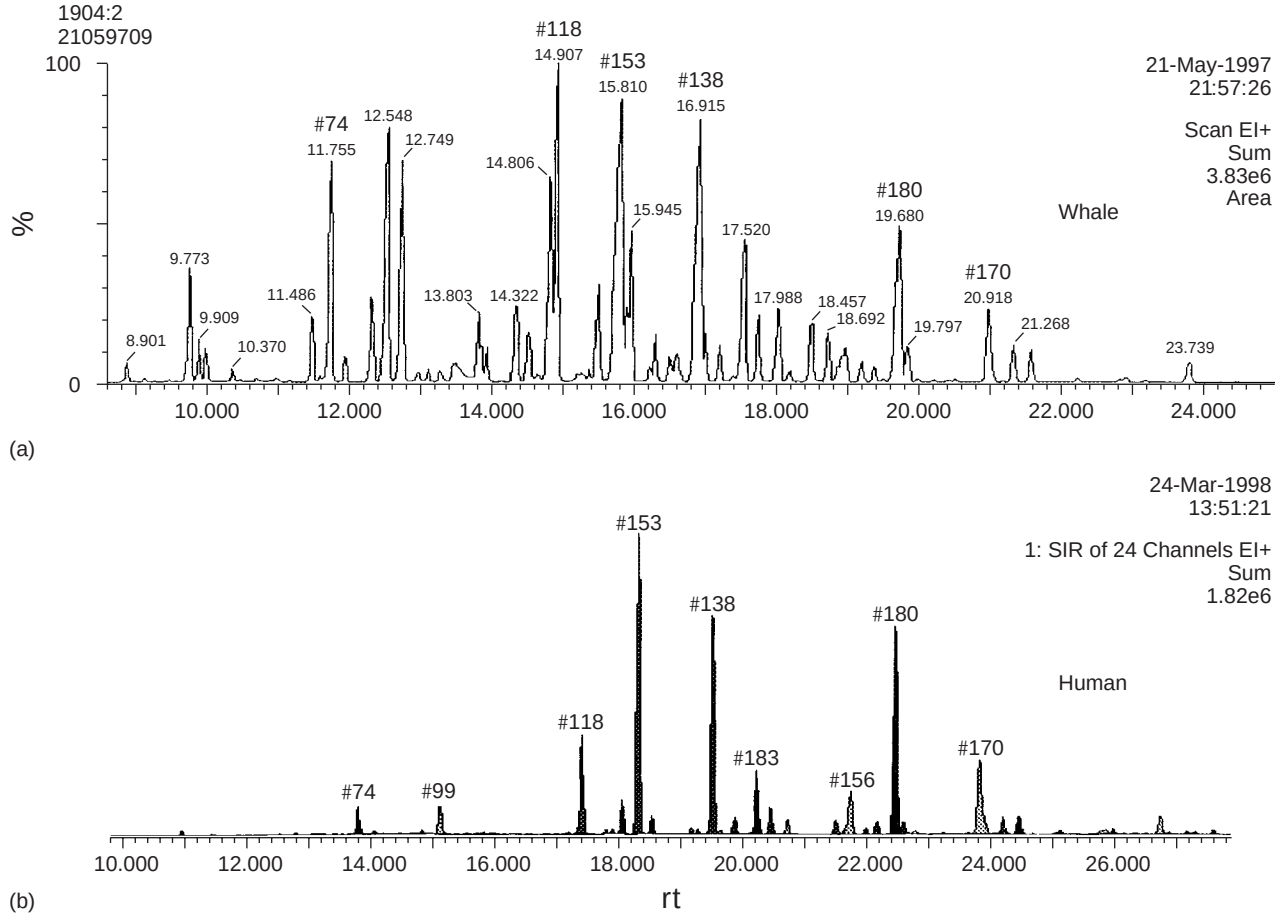

Figure 2 PCBs in (a) a whale and (b) a human sample. The most abundant congeners are indicated by their IUPAC number. This example clearly illustrates the differences in PCB patterns between the two samples. The whale sample contains nearly all of the congeners present in commercial PCB mixtures whereas the human sample only contains a number of specific congeners. Comparing noncongener-specific analyses of these samples is rather useless.

and clean-up techniques for qualitative analysis. These advantages over ECD as the detection technique make GC/MS the most prominent technique from a QA/QC point of view. In specific cases HRMS is used as the detection method. HRMS enhances the specificity by being more mass-specific, operating at a mass resolution of over 10 000 compared with a resolution of 100–1000 for a low-resolution instrument. Use of HRMS reduces the change due to interference at the more exact PCB masses monitored. More important though is the improvement in signal-to-noise ratio for high-resolution instruments, resulting in lower detection limits. HRMS is mostly used for the detection of the three planar PCBs (PCB #77, #126, #169). These PCBs are normally present in samples at 10–100 times lower concentrations than the normal ''bulk'' PCBs but are of toxicological importance as is evident from Table 1.

In conclusion it can be stated that there is no real universal method of analyzing PCBs. The preferred method depends on many factors, including sample type, levels and most of all the objective of the study. Questions on the accuracy of the results, expected levels, or objective of the study are preferably discussed before analysis. By doing this the analytical chemist can choose the best available technique for the purpose of the study. Because the extraction and clean-up techniques largely depend on the analytical detection technique that will be used in the end, the detection methods will be discussed in detail first, followed by the sample enrichment, clean-up and extraction techniques in later sections. For the interested reader many reviews and publications on the analytical chemistry of PCBs can be found in the international literature, but by far the most comprehensive publication is *Analytical Chemistry of Polychlorinated Biphenyls* by Erickson.[20] In addition, several standardized analytical procedures are described in more detail in this book.[20] Common United States Environmental Protection Agency (USEPA) procedures for PCB analyses include Methods 505, 508, 525, 608, 625, 680, 1618, 8080, 8081, 8250, 8270 and 8275.

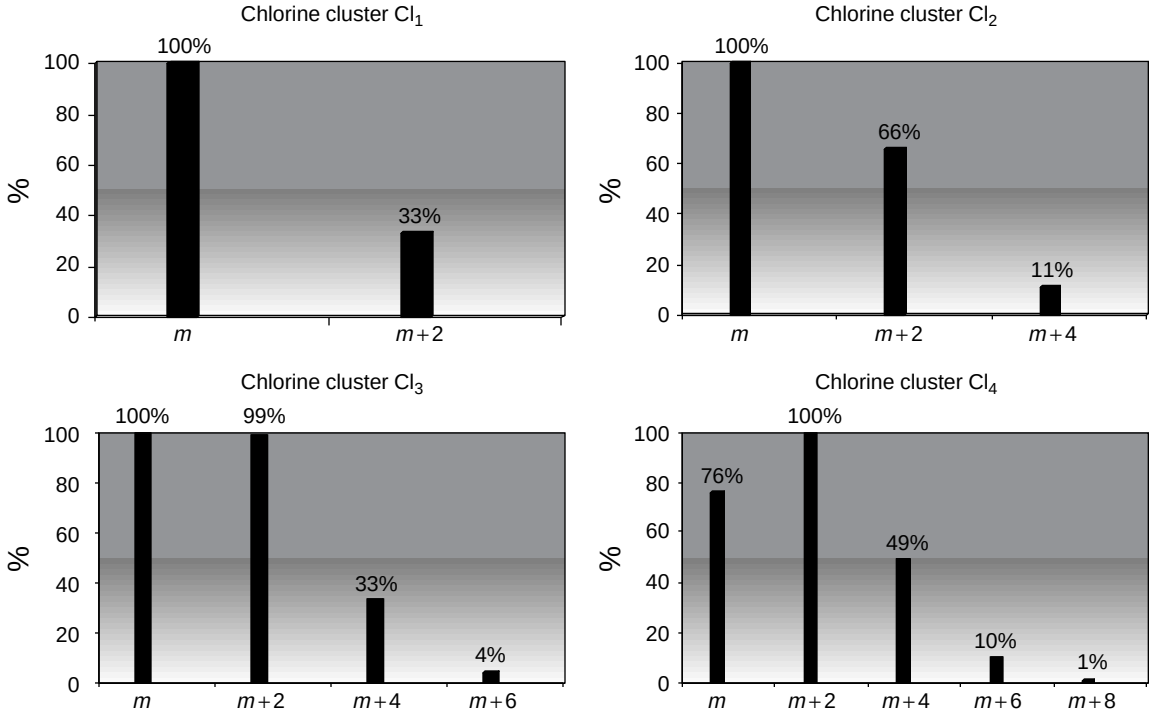

Figure 3 Theoretical chlorine clusters of compounds containing Cl_1, Cl_2, Cl_3 or Cl_4. The natural abundance of ^{35}Cl and ^{37}Cl results in a molecular ion of molecular mass m, $m+2$, $m+4$, $m+6$ and $m+8$ for mono through tetra PCBs as shown.

3.1 Detection Methods

3.1.1 Gas Chromatography with Electron Capture Detection

GC/ECD is probably still the most used analytical technique for environmental samples because of its selectivity and high sensitivity for PCBs. Originally developed by Lovelock and Lipsky[21] in the 1960s, ECD quickly became the most important analytical tool for the analysis of chlorinated environmental pollutants. The electron capture detector consists of a radioactive β^- particle emitter (^{63}Ni), emitting towards an anode over an applied voltage (10–50 V), thus resulting in a current. When specific molecules with electron-adsorbing groups (halogens, nitrogroups or conjugated double bonds) pass by, these electrons are absorbed causing the current to drop, and this drop in current is recorded as the detector response. Modern electron capture detectors use pulsing signals to enhance the linear range. The linear range is around 10^4 and ECD devices are able to detect 0.1 pg Cl s^{-1}. In Table 2, GC/ECD detection is compared with several other detection techniques used for the detection of PCBs. Although NCIMS in some cases produces a slightly better detection limit, the detection limit of GC/ECD is nearly as good. Compared with GC/NCIMS, ECD is much more stable and exhibits a larger linear range. GC/NCIMS is not normally used in routine analysis, which might be because of the difficulties in achieving stable ion source conditions during analysis while a reactant gas ion is flowing into the ion source. GC/ECD devices are nearly as sensitive as the much larger and thus more expensive HRMS instruments. These instruments however are much more selective because they use MS instead of the electron-absorbing capability of halogens. However from an instrumental point of view GC/ECD is easy to use and readily available in many configurations.

Until the development of capillary GC columns at the beginning of the 1980s, packed columns were used to separate the PCBs from each other and from other electron-absorbing compounds. In most cases these were other chlorinated environmental pollutants displaying the same properties as the PCBs, including the well-known pesticide dichlorodiphenyltrichloroethane (DDT) and its metabolites chlordane, hexachlorobenzene, aldrin and dieldrin. In some early publications using packed GC columns these compounds almost certainly interfered and resulted in an over-estimation of the real levels. Nowadays with the use of capillary columns the separation capacity has improved enormously but fractionation prior to analysis is often still necessary. It is therefore important to know which analytical detection method will be used before designing a sample clean-up procedure. Use of a less specific detection technique will require a more

Table 2 Detection limits for the different detection techniques, adapted from Erickson.[20] Depending on the injection volume and the state of the instrument, amounts of less than a femtogram can be detected using NCIMS. More routine techniques such as GC/HREIMS and GC/ECD are able to detect in the low femtogram level

Detector	Limit of detection(g)											
		mg			ng			pg			fg	
	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}	10^{-13}	10^{-14}	10^{-15}	10^{-16}
GC/GC/NCIMS	→	→	→	→	→	→	→	→	→	→	→	→
GC/NCIMS	→	→	→	→	→	→	→	→	→	→	→	
GC/NCIMS/NCIMS	→	→	→	→	→	→	→	→	→	→		
GC/HREIMS	→	→	→	→	→	→	→	→	→			
GC/ECD	→	→	→	→	→	→	→	→	→			
GC/EIMS	→	→	→	→	→	→	→	→				
GC/IR	→	→	→	→	→	→	→	→				
GC/FID	→	→	→									

NCIMS, negative chemical ionization mass spectrometry; HREIMS, high-resolution electron impact mass spectrometry; EIMS, electron impact mass spectrometry; IR, infrared; FID, flame ionization detection; GC/NCIMS, gas chromatography/negative chemical ionization mass spectrometry.

rigorous sample clean-up and fractionation. It might also be necessary to run different fractions obtained during sample enrichment and clean-up separately, which might be avoidable using a different detection technique (e.g. GC/MS). On the other hand the objective of the project is important; from a toxicological point of view the analysis of the three planar PCBs might need extensive fractionation while using GC/ECD, while if analyzing a PCB spill the sum of several PCBs might be sufficient and only minimal clean-up and fractionation is required.

3.1.2 Gas Chromatography Coupled to Mass Spectrometry

The preferred detection technique from an analytical point of view is GC/MS. Using GC/MS several additional QA/QC features can be added to the analytical method. Several MS instruments are available and benchtop instruments such as the ion trap or quadrupole are nowadays widely used in PCB analysis and are expected to replace GC/ECD more and more. In special cases high-resolution mass spectrometers are used, especially when the three planar PCBs (#77, #126 and #169) are to be analyzed at low levels or in the same fraction as toxicologically similar dioxins.

The mass spectrometer can be used in two ways:

1. to monitor all the masses in a certain mass range (full scan)
2. to monitor only specific masses in so-called selective ion monitoring (SIM) or selective ion recording (SIR).

Ionization can take place by means of electron impact (EI) or negative chemical ionization (NCI), a process similar to electron capture in the ECD device. Although NCI is somewhat more sensitive, as shown in Table 2, EI is by far the preferred method. This is probably because it is difficult to establish stable conditions in the ion source while the reactant gas is pumped into this source using NCI. Another reason might be that the relative response factors (RRFs) of the PCB homologs, that is PCBs with the same chlorination level, are not constant. This implies that more internal and quantification standards are needed. Accurate results have however been reported using GC/NCIMS.[22]

The bombardment of the PCB molecules in the ion source during EI ionization results in very specific fragmentation of the PCBs. EI ionization is generally performed at 70 eV. Typical full-scan mass spectra are acquired over a mass range of 100–600 mass units at a scan speed of 1 scan s^{-1}. Full-scan mass spectra from penta-, hexa-, hepta- and octachlorinated biphenyls are given in Figures 4–7. These mass spectra serve as a kind of fingerprint for the PCBs and in this way PCBs can be identified in environmental samples. Looking more closely at the mass spectra, a number of features can be used for quantitative analysis. First, from the molecular ion the two most abundant masses can be used to monitor the PCBs. One mass can be used to quantify the PCBs, and the other to confirm that it really is an undisturbed PCB peak by calculating the ratio $(m)/(m+2)$ and comparing this with the theoretical chlorine ratio. It can also be seen that two major fragments appear because of the loss of Cl and the loss of Cl_2 as illustrated in Figure 6. The loss of Cl does not present any problems but the loss of Cl_2 results in a mass fragment with the same mass as the PCB with two fewer chlorine atoms. The heptachlorinated biphenyls produce an interfering mass at the molecular masses of the penta-PCBs and the hexa-PCBs produce

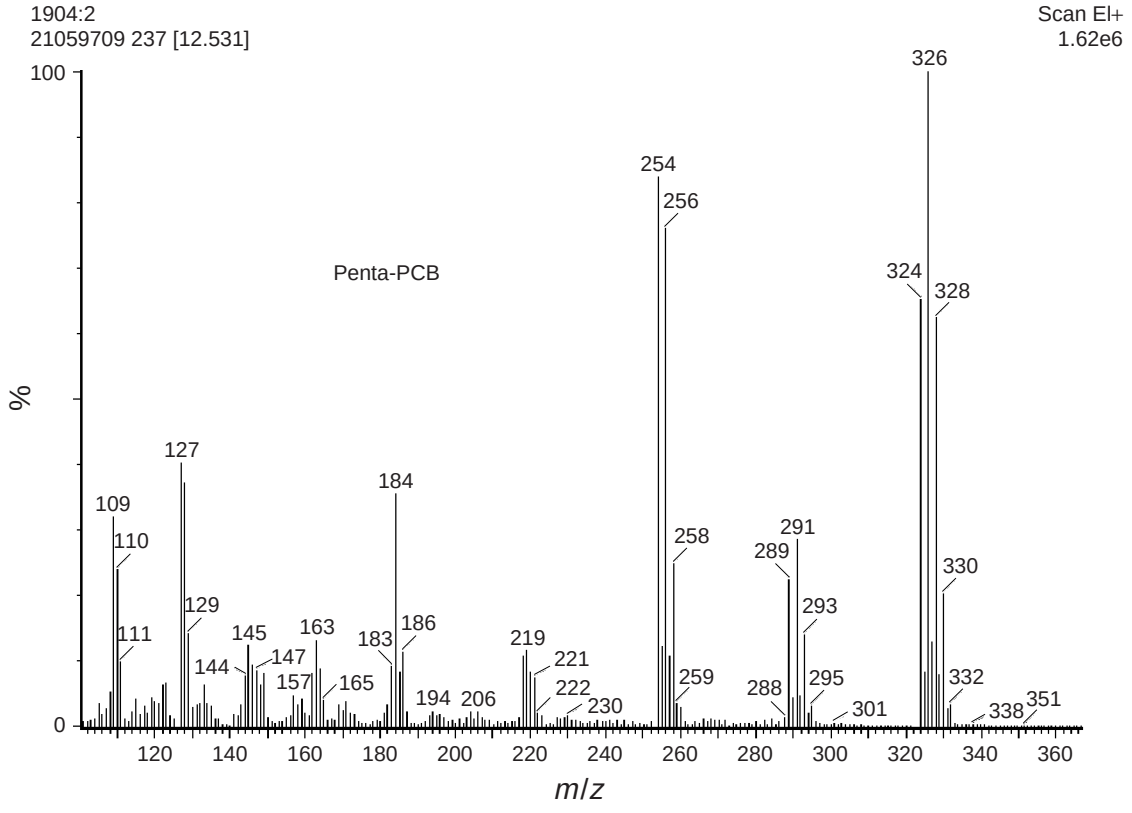

Figure 4 Full-scan spectrum of a penta-PCB in a whale sample. The spectrum was acquired at 70 eV over a mass range of 100–600 amu at a scan speed of 1 scan s^{-1}.

a mass fragment at the same mass as the tetra-PCBs. Normally this is not a problem because these PCBs will be separated on the GC column.

One of the biggest advantages of GC/MS is that chemically identical ^{13}C-labeled stable isotopes can be used as internal and recovery standards. This is illustrated in Figure 8, where a SIR GC/MS run of a human sample is shown. The top channel represents the mass (371.88) for the ^{13}C-labeled internal standards added to the sample before sample extraction and clean-up and the recovery standard added just before GC injection. The lower two channels (359.84 and 361.64) are the two most abundant molecular ions of the PCBs actually present in the sample. The most abundant mass (359.84) is used for quantification; the other mass is used to compare its ratio with the theoretical ratio of $m + 2$ and $m + 4$ for a hexa-PCB (0.82). The three internal standards are used for quantification against an external standard containing at last one PCB at each chlorination level. The recovery standard is used to calculate the percentage of the internal standard lost during sample extraction, fractionation and clean-up.

The example above illustrates the capability of a low-resolution GC/MS system, in particular a quadrupole instrument. An ion-trap instrument performs in a similar way but can also be used for tandem mass spectrometry (MS/MS), to monitor specific fragments. Both instruments perform at mass resolution, which means that mass 360 can be separated from mass 361. High-resolution magnetic sector mass spectrometers operating at mass resolutions of 10 000 or more can differentiate between 359.96 and 360.00. This reduces the risk of interferences and significantly improves the signal-to-noise ratio. HRMS devices are thus somewhat more sensitive than the low-resolution instruments. For routine analysis of environmental samples it is not necessary to use HRMS. For several special applications, for example the analysis of the three planar PCBs (#77, #126, #169), high-resolution gas chromatography/high-resolution mass spectrometry (HRGC/HRMS) is widely used. The reasons for this are the relatively low levels of these PCBs in environmental samples, in addition to the fact that in several clean-up procedures these PCBs end up in the dioxin fraction and are easily run together with the dioxins in the same GC/MS run.

3.2 Sample Enrichment and Clean-up

The choice of the sample clean-up and enrichment procedure is dependent on many elements. The sample levels and matrixes determine the clean-up as well

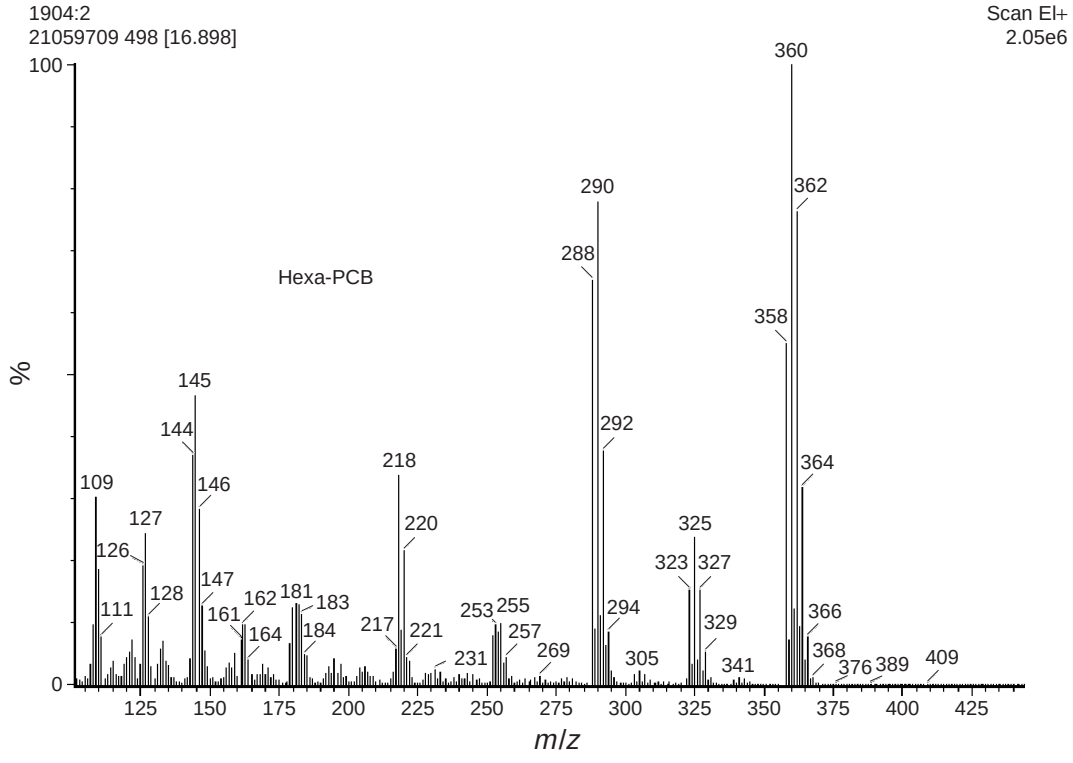

Figure 5 Full-scan spectrum of a hexa-PCB in a whale sample. The spectrum was acquired at 70 eV over a mass range of 100–600 amu at a scan speed of 1 scan s^{-1}.

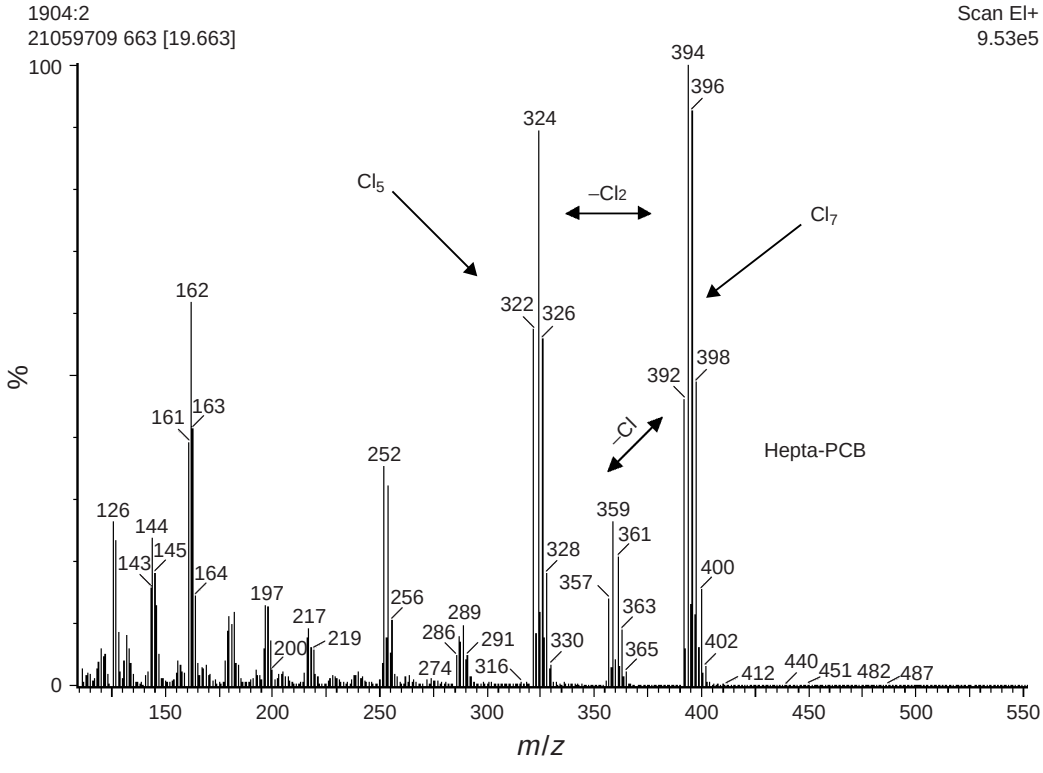

Figure 6 Full-scan spectrum of a hepta-PCB in a whale sample. The spectrum was acquired at 70 eV over a mass range of 100–600 amu at a scan speed of 1 scan s^{-1}. The typical molecular Cl_7 mass cluster, the loss of Cl to form a fragment containing Cl_6, and the typical loss of two chlorines resulting in a more abundant Cl_5 cluster are all shown.

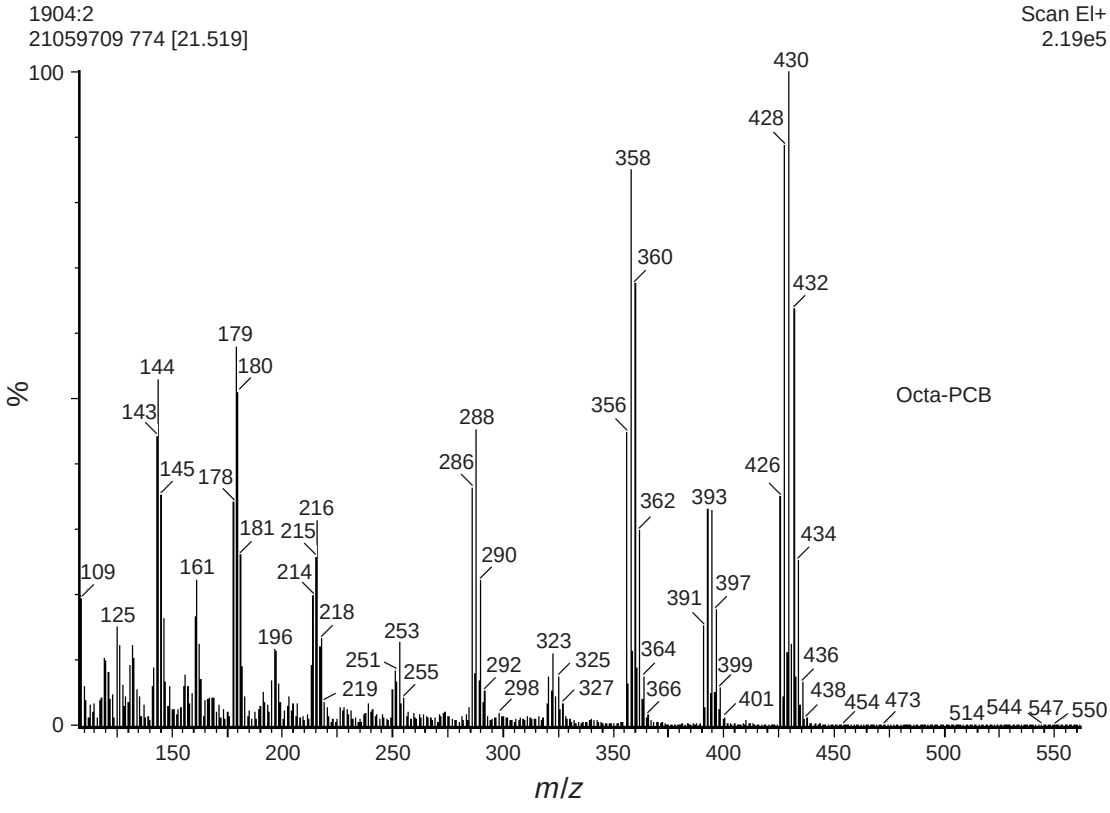

Figure 7 Full-scan spectrum of an octa-PCB in a whale sample. The spectrum was acquired at 70 eV over a mass range of 100–600 amu at a scan speed of 1 scan s^{-1}.

as the detection technique used. The more selective the detection technique, the less laborious the sample clean-up that has to be performed on the sample extract. Generally, when using GC-based detection techniques, sample clean-up is necessary. The clean-up can be divided into three steps: an acid/base clean-up including a fat removal step; an open-column chromatographic step; and, for the detection of mono *ortho* or planar PCBs, a planar separation step on carbon is essential.

3.2.1 Acid/Base Clean-up/Fat Removal

Because PCBs are such stable compounds, rigorous clean-up schemes can be applied without risk of destroying the target compounds. The sample extract can thus be digested by a strong acid[23] or base[24] with relatively small analyte losses. Back-extraction with an organic solvent (hexane, methylene chloride) partitions the PCBs back to the organic phase while the fats are destroyed and other more polar interferences partition to the acid or base phase. Acid or base digestion is mainly used for large sample sizes or heavily contaminated samples. A more sophisticated approach involves the use of acid/base-modified silica. Activated silica is ''loaded'' with, for example, sulfuric acid or sodium/potassium hydroxide. Again, depending on the sample size, silica can be modified by 22% or 44% of acid (wt%). The use of combined acid/base/activated silica open-column chromatography in one single column is a commonly used clean-up procedure using hexane or a combination of hexane and methylene chloride as the elution solvent.

Other options for the removal of large amounts of fat or other large compounds are gel permeation chromatography (GPC) or dialysis using semipermeable membrane devices (SPMDs). Size-exclusion chromatography (SEC) or GPC is based on the stronger retention of smaller compounds as compared with macromolecules (fats, humic acids) in the 600–1500 amu range. In GPC the stationary phase is a hydrophobic gel (Styragel™, Bio-Beads™, Sephadex™) into which the smaller molecules can penetrate while the larger macromolecules are not retained. The large macromolecules therefore elute first, followed by the smaller molecules. By using model compounds or commercial PCB oil the GPC column can be calibrated using ultraviolet (UV) detection. GPC is widely used for the separation of PCBs and fat in biological samples and is well described in the literature.[25–27]

SPMDs are based on the diffusion of the target compounds through a semipermeable membrane with a typical cut-off size of 10 Å (500 amu). A polyethylene tube

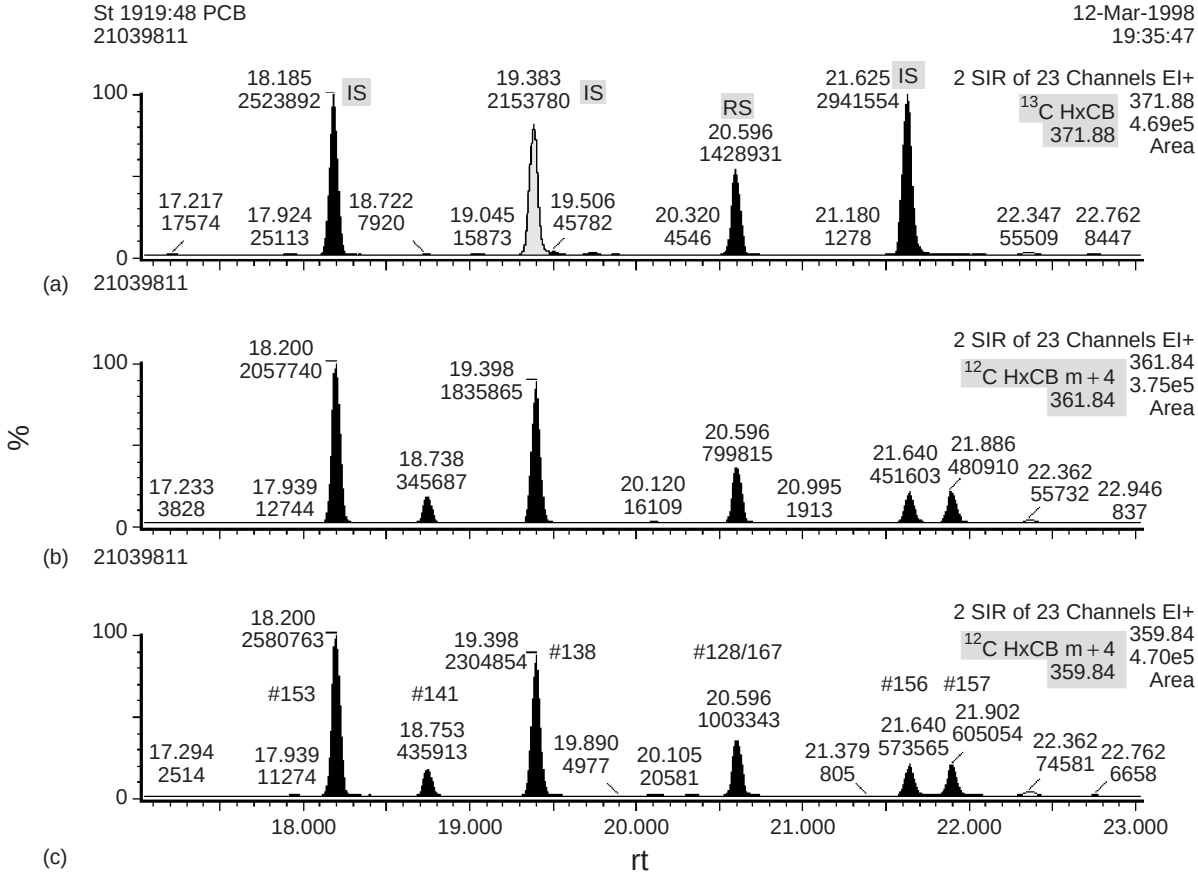

Figure 8 SIR chromtogram of hexa-PCBs in a human sample. The (a) channel represents the ^{13}C-labeled internal standard and recovery standard at mass 371.88, (b) represents the channel with the $m+2$ mass of the molecular chlorine cluster at mass 361.64, and the (c) channel shows the most abundant $m+4$ mass of the molecular chlorine cluster at 359.84. The most abundant hexa-PCBs in a human sample are marked by their Ballschmiter numbers (#153, #141, #138, #128/167, #156, #157).

is filled with the samples, sealed and placed in a column filled with pentane. Lipid removals are in the 90–99% range using two to three volumes of cyclopentane collected after 24 h of dialysis. The recoveries of the PCBs and pesticides in the combined dialysis fractions range from 88% to 98%.[28–30] SPMDs have recently been used[31,32] as sampling devices for PCBs in water, reflecting the bioaccumulating fraction of hydrophobic chemicals, using trioline within the polyethylene tube.

3.2.2 Open-column Chromatography

Open-column adsorption chromatography is the most common clean-up technique and is based on the differences in polarity, solubility and consequently partitioning between the stationary phase and the elution solvent. The affinity of PCBs and interfering compounds depends on the surface activity of the adsorbent and the solvent strength of the elution solvent. By using different adsorbent materials (silica, Florisil™, alumina) which can be activated at different temperatures, clean-up procedures can be fine-tuned for specific applications. Open-column chromatography is used in many configurations and column sizes, with a wide variety of organic solvents (either alone or in combination). Before using a clean-up column, the column has to be calibrated. The elution volume in which the PCBs elute has to be determined, preferably with the real sample matrix present. Adsorbents used a relatively long time after activation, or even those from different batches, can become less active and thus should be calibrated again. Azulene, an aromatic dye, has been used to facilitate easy calibration of several adsorbents used for the clean-up of PCBs.[33] Another precaution that should be taken when doing trace analysis is to check the adsorbent for PCB contamination. Several silicas have been reported to contain trace amounts of PCBs,[34] and thus careful washing of the adsorbent with organic solvents is needed.

One of the most commonly used absorbents in clean-up procedures for PCBs is Florisil™, a synthetic magnesium silicate. Normally Florisil™ is designed by the manufacturer to be activated at 650 °C and should be

stored at a temperature of 130 °C before use. Florisil™ is used both fully activated and 2–5% deactivated with water. Traditionally the elution solvent is an ethanol–diethylether (2:98) mixture, but the use of several other mixtures has been reported, including ether–petroleum and ethyl ether–hexane. More recently hexane–methylene chloride combinations or hexane alone have gained more popularity. Note that the different combinations and thus solvent strengths can change the elution patterns of PCBs and other compounds considerably.

As well as acid or base modified, granular silica can also be used activated or deactivated by applying small amounts of water. The hydroxyl groups of silica bind to the compounds applied and thus a separation on polarity is achieved. The PCBs interact with silica with basically the same strength and only minor separation of the PCBs due to size can be achieved on silica. In several procedures the ''rough'' silica clean-up is used before moving on to a more specific Florisil™, alumina or carbon column. *n*-Hexane is usually used as the elution solvent, sometimes in combination with methylene chloride.

Alumina (Al_2O_3) or AlO_x is used in a wide variety from activity grade I (0% water) to V (15% water), also known as ''super'' AlO_x. Alumina has gained its popularity as an adsorbent in ''dioxin'' analysis because it can be used to separate the more polar dioxins and furans from the less polar PCBs. When applied to an alumina column the PCBs are less well retained and typically elute with hexane or hexane with a small percentage of methylene chloride (2–5%), while the dioxins and furans need a stronger solvent such as methylene chloride or a 50% mixture of methylene chloride and hexane. In some special cases structural separation based on planarity of the PCBs has been achieved, but not as effectively as on a carbon column.

3.2.3 Carbon Column

Initially carbon (Fisher coconut charcoal) was used to separate PCBs from the DDT group of pesticides.[35] The ability of active carbon (Darco-C™) to separate on the number of chlorine atoms in the *ortho* position was then discovered.[36] The prospect of using active carbon (PX-21™) was later on fully exploited for the separation of the ''bulk'', *o*-substituted PCBs and pesticides from the planar PCBs (#77, #126, #169) and dioxins.[37] Both dioxins and non-*o*-PCBs can adapt a planar configuration. This results in a strong retention of these compounds on carbon and even strong solvents like methylene chloride cannot elute these planar compounds from a carbon column. After applying the sample to the active carbon column, nonplanar compounds can be eluted from the column with a mixture of hexane–methylene chloride. The dioxins and planar PCBs have to be eluted with an aromatic solvent (toluene). In some cases only back-elution of the column will result in the elution of the planar compounds in reasonable amounts of aromatic solvent. In the literature[38] more than 34 different protocols are mentioned, using 10 different carrier materials. The majority of the methods use PX-21™, or its successor AX-21™, mixed with different carrier materials. Other commonly used carbon materials include Wako active carbon and Carbopack C™. Problems with the reproducibility of the different active carbons have been observed and recently high-performance liquid chromatography (HPLC) columns have been introduced to obtain better results. Good separation between the planar and nonplanar PCBs was achieved on porous carbon gel (PCG) HPLC columns.[39] Along with the establishment of TEFs for the mono-*o*-substituted PCBs (see Table 1), the interest in separation by the number of *ortho* chlorine atoms was revived. The separation on the classic PX-21™ was reported[40] and in addition the isolation of the mono-*o*-substituted PCBs was successfully done using PX-21™ dispersed on an octodecylsilane HPLC column.[41] This separation was also reported upon a PCG column.[42] Typically three fractions are obtained from the HPLC: a fraction with the pesticides and ''bulk'' PCBs, a fraction with the mono-*o*-PCBs and a fraction with the planar PCBs, dioxins and furans. In Figure 9 a typical clean-up protocol for a combined PCB and dioxin analysis is shown as an illustration of a fine-tuned clean-up procedure.

3.3 Extraction

The octanol–water partition coefficient (K_{ow}) of the PCBs varies between 4.3 for the mono-substituted PCBs and 8.26 for the deca-substituted PCBs.[43] This indicates a partition coefficient between the organic and the water phase of 10^4–10^8. This implies relatively easy extraction of PCBs from the water phase. Liquid–liquid extraction by shaking the water sample with an organic

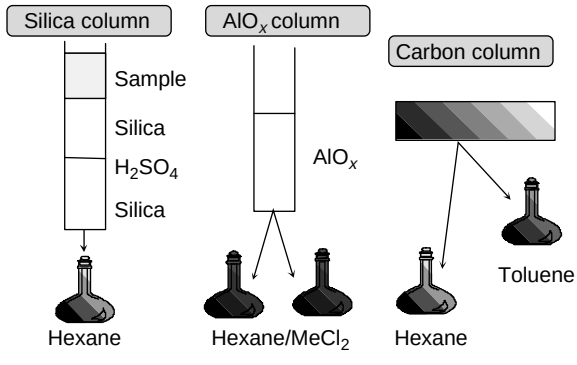

Figure 9 An illustration of a typical clean-up protocol for a combined dioxin and PCB analysis. Several columns are fine-tuned and combined to achieve several fractions free from interferences.

solvent in a separating funnel is generally enough to transfer PCBs into the organic phase. The most common procedure is a three-phase extraction with methylene chloride or a methylene–hexane mixture resulting in PCB recoveries of over 99%. Continuous liquid–liquid extraction techniques can also be used. More recently solid-phase extraction (SPE) has been developed for large water samples, which are normally difficult to extract by traditional liquid–liquid extraction. In SPE the water is pumped through an adsorbent (XAD-2™, polyurethane foam, C_{18}-silica) leaving the hydrophobic PCBs behind on the absorbent. The PCBs are dissolved from the absorbent by an organic solvent. Care should be taken when the water contains particulate or organic matter.

For solid samples Soxhlet extraction is the preferred method although other solvent extraction methods are known, e.g. centrifugation after mixing with an acetone–methylene chloride–hexane mixture, column extraction with organic solvents after mixing with sodium sulfate, or sonication with an organic solvent. During Soxhlet extraction a solvent is refluxed through a cooler where it condenses and drips back into a thimble containing the sample. This process continues until the thimble is filled to the top and liquid siphons into the round-bottom flask containing the extraction solvent in a heater. These cycles are continued during the extraction period (12–48 h) resulting in a very efficient extraction.

Recently even more efficient extraction has been achieved using accelerated solvent extraction (ASE™). This technique is based on more efficient extraction at higher temperatures. Basically the organic solvent is pressurized, making it possible to perform extractions above the boiling point of the extraction solvent. Extraction cycles are performed at temperatures between 100 °C and 200 °C, resulting in short extraction times (10–20 min) and considerable reductions in solvent use. PCBs are found in the fat fraction of biological samples and extraction of the lipids is normally sufficient. Column extraction of the sample homogenized with a drying agent (sodium sulfate) is one of the most commonly used extraction techniques. Soxhlet extraction can also be used on biological sample homogenates. The lipid fraction is most easily extracted from body fluids and other liquid biological samples (for example milk) by liquid–liquid extraction.

4 RECENT DEVELOPMENTS IN ANALYTICAL METHODOLOGY

The analytical chemistry of PCBs and other environmental pollutants has changed enormously over the last 30 years and this development is expected to continue into the next century. In this section some recent developments are described which are still at a developmental stage but are expected to have an impact on the analysis of persistent organic pollutants (POPs), including PCBs, in the near future. One major issue that will be improved is the speed of analysis, as both sample extraction and sample clean-up are expected to become faster and more efficient. In general more information is demanded more quickly and more accurately. This will lead to a combination of congener-specific PCB analysis, and for example pesticide or brominated flame retardant analysis in one extraction. In addition a strong demand to reduce the usage of (chlorinated) organic solvents is foreseen. The same demand for speed of analysis will be focused on the gas chromatographic detection techniques and fast GC is expected to be incorporated in the routine analysis of a wide variety of POPs. In sections 4.1 and 4.2 two promising approaches are discussed in more detail as examples of recent developments in the analytical chemistry of PCBs.

4.1 Supercritical Fluid Extraction Coupled with Liquid Chromatography

4.1.1 Introduction to Supercritical Fluid Extraction

Solvents and gases under high pressures at elevated temperatures are becoming more and more important as extraction media in environmental analysis. The use of CO_2 at supercritical conditions, with or without modifiers, has increased enormously for both biological and solid samples.[44–46] In addition, extraction by organic solvents at elevated pressure and temperature has also recently been introduced as ASE™ or enhanced solvent extraction. Several organic solvents exhibit much better extraction efficiencies under pressure at higher temperatures.

Which technique to use depends on the sample matrix and the compounds to be analyzed. An important consideration is how specific the extraction should be. ASE™ or enhanced solvent extraction is in most applications rather unspecific and extracts not only the target compounds but also many interfering compounds. Extracts obtained in this way often need further sample enrichment or sample clean-up. Supercritical fluid extraction (SFE) has the advantage that it can be made very specific by varying the temperature, pressure and the addition of different modifier concentrations. To find the optimum SFE extraction conditions for different sample matrixes and target compounds, however, demands somewhat more developed methods. When these optimum conditions are established the advantages of using SFE are obvious and minimal clean-up and sample enrichment is needed. So, not only is the use of organic solvents during extraction

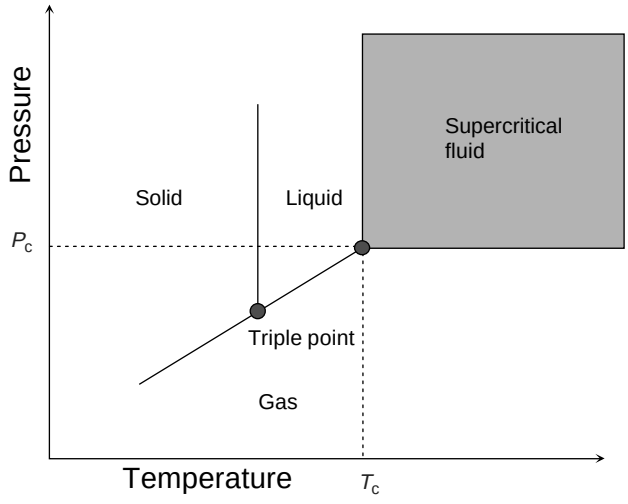

Figure 10 Temperature/pressure phase diagram, indicating solid, liquid, gas and supercritical phases.

reduced, but the extensive use of organic solvents during sample clean-up is also reduced to a minimum.

4.1.2 The Theory of Supercritical Fluid Extraction

A gas or a liquid becomes a supercritical fluid above the critical point in its temperature/pressure phase diagram (Figure 10). The critical temperature (T_c) and the critical pressure (P_c) define this point. In the supercritical state a substance is neither a gas nor a liquid. The properties of a supercritical fluid are therefore intermediate between those of the gas and the liquid phases. The density is comparable with that of a liquid whereas the viscosity is equivalent to that of a gas. As a consequence, the diffusion rate is somewhere in-between those of the liquid and the gas phases. Supercritical fluids can be compressed to almost the same density as a liquid, and thus show liquid-like solvation properties. With respect to extraction, a supercritical fluid has a higher diffusion rate than a liquid, while still exhibiting the solvation capacity of a liquid. Consequently the mass transfer kinetics are rapid, and by controlling the pressure and temperature the solvent strength can be adjusted. Optimization by changing the pressure and temperature, to achieve selective extraction, is thus possible.

4.1.3 Supercritical Fluid Extraction Instrumentation

Substantial improvements have been made in recent years in SFE instrumentation, going from home-made instruments to sophisticated commercial instruments with the convenience of autosamplers and computer operation. The instrumentation required is basically rather simple (Figure 11). A gas cylinder where liquid CO_2 is withdrawn from the bottom is connected to a pump where the liquid is

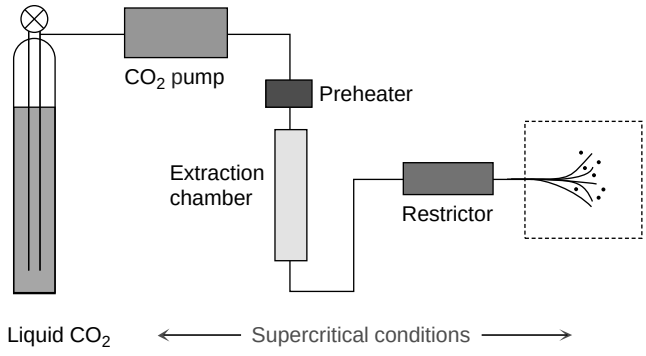

Figure 11 A block diagram of an SFE system.

compressed to the desired pressure. The gas is preheated to the extraction temperature in an oven in which the extraction vessel is placed. A restrictor placed after the extraction chamber controls the CO_2 flow and a variable pressure regulation system controls the pressure in the extraction vessel. Collection of the PCBs after expansion of the supercritical CO_2 can be done in a liquid or on a solid phase.

4.1.4 The Use of Supercritical Fluid Extraction in Polychlorinated Biphenyl Analysis

SFE is very suitable for the extraction of PCBs in a wide variety of samples. By optimizing the extraction conditions cleaner extracts are obtained which need less further sample clean-up. There is also the possibility of combining sample extraction and clean-up by utilizing the solid-phase trap for supercritical fluid extraction/liquid chromatography (SFE/LC). Excellent results have been achieved with direct injection of concentrated SFE extracts containing PCBs without further sample clean-up. The extraction of PCBs from soils, sediments, sewage sludge and absorbents has been successfully reported by a number of research groups.[47–50] Methods using SFE performed well in certification exercises for soil.[51] Successful SFE of PCBs from biological samples has also been reported upon in the literature.[52–55] The most encouraging approach is the combined POPs analysis using SFE/LC[56] or supercritical fluid extraction/gas chromatography/mass spectrometry (SFE/GC/MS).[57,58] In both procedures a large number of compounds were reported besides the PCBs, including dioxins, furans, DDT, chlordane and brominated flame retardants.

4.2 Gas Chromatography Coupled with Time-of-flight Mass Spectrometry

The sample preparation procedure, including both extraction and clean-up, has become more efficient due to the use of new extraction techniques.[46] By combining

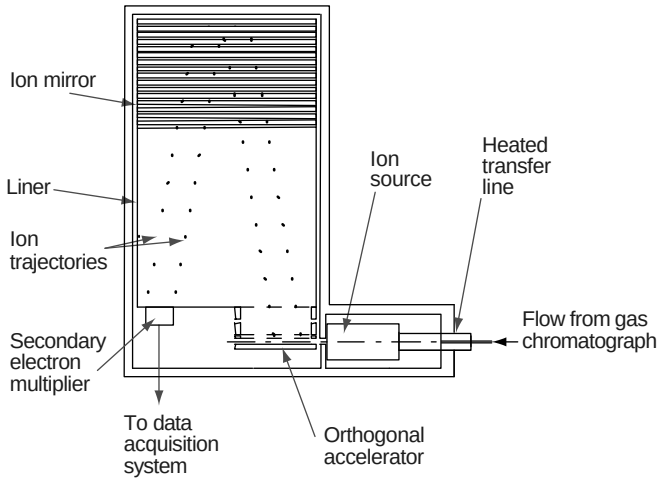

Figure 12 General schematic diagram of orthogonal extraction TOFMS device coupled to GC column.

extraction and clean-up using SFE/LC, the time required to produce an extract ready for GC/MS injection has been considerably reduced, from several days to only 20 min.[59] When achieving such fast clean-up and extraction, suddenly GC/MS analysis becomes the bottleneck in the analytical process. The introduction of fast GC is a step in the right (''fast'') direction, but there are problems getting conventional mass spectrometers to achieve the very fast scanning times needed. Time-of-flight mass spectrometry (TOFMS) devices do not experience these kinds of problems and are able to scan more than 50 times s^{-1}, more than enough for fast GC applications. A schematic diagram of a time-of-flight mass spectrometer is given in Figure 12. In a TOFMS instrument operated in the EI mode, the ions are accelerated into the flight tube at a frequency of 36 kHz. After reflection from the ion mirror the ions can be measured at a rate of up to 100 scans min^{-1}. Figure 13 shows the reconstructed GC/MS runs of both the TOFMS and the quadrupole analysis of a human adipose tissue sample. The summation of mass channels 360, 394 and 318, corresponding to the masses for hexa-PCB, hepta-PCB and 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene respectively, is shown here. Figure 13 clearly illustrates the difference in analysis time between the GC/TOFMS run and the SIM quadrupole run. It takes less than 7 min to elute all compounds of interest and still achieve enough resolution to identify and quantify all PCBs and pesticides of interest in a human sample from the GC/TOFMS run with the fast GC. It is possible to quantify more than 30 different PCB congeners in a human sample above the detection limit. Combining SFE/LC as the sample extraction and pretreatment technique with GC/TOFMS analysis reduces the analysis time for screening of environmental pollutants from several days to less than 1 h. The fast analysis with a GC/TOFMS instrument, utilizing a ''fast'' GC column, still gives enough resolution of 30 PCBs because of the fast scanning times of a TOFMS instrument.

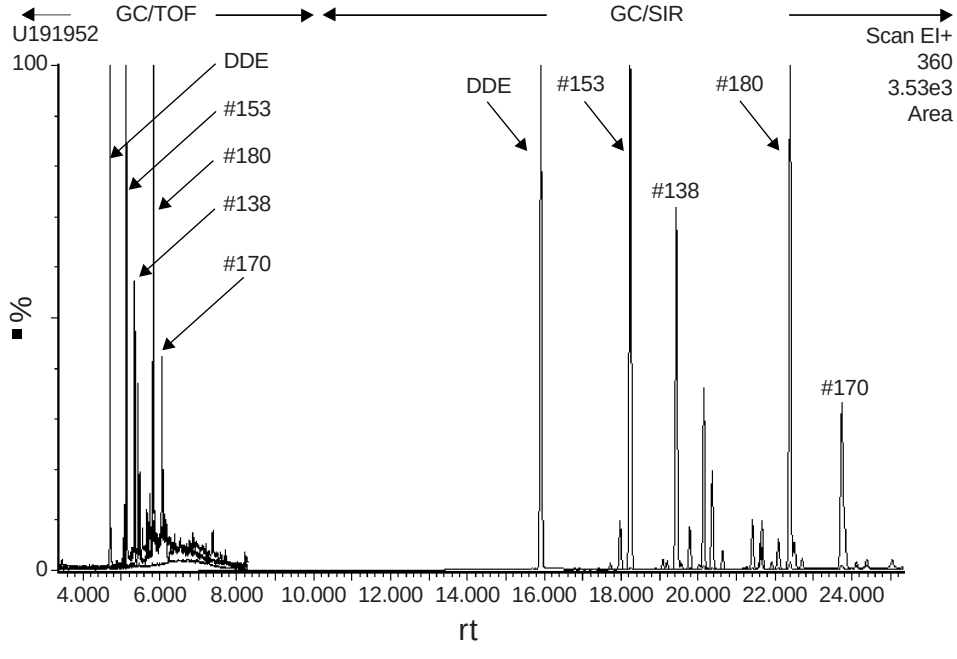

Figure 13 A reconstructed chromatogram of a human adipose tissue extract showing both the GC/TOF run and the GC/SIR run on the same timescale. Masses 360, 394 and 318 are shown. DDE, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene.

5 QUALITY ASSURANCE AND QUALITY CONTROL BY INTERNATIONAL COMPARISON

To illustrate the current analytical methodologies used in leading laboratories in Europe, Japan and the USA for analyzing PCBs in biological material, the reported methodologies from two recent interlaboratory studies are described in this section and Table 3. The summary is based on 14 laboratories within the IUPAC fish oil study and 19 laboratories in the WHO/EURO milk study. There is some overlap between the 33 participating laboratories, i.e. some of the laboratories took part in both studies. The main steps in the PCB analyses are summarized in Table 3: (a) *extraction* of the fat including the analytes, (b) quantitative *determination* of the amount *of fat*, (c) *clean-up*, i.e. isolation of the analytes from the fat and from other interfering components, (d) *detection* and (e) *quantification* of the analytes.

Table 3 Methodology for PCB analysis used in the IUPAC fish oil study and the WHO/EURO milk study

Analysis stage	No. of labs used
Extraction[a]	
Not reported	2
None; dilution	7
Soxhlet	4
Liquid extraction	16
Column extraction	2
SFE	2
Fat determination[b]	
Total; gravimetric	17
Aliquot; gravimetric	2
Pre-extracted (IUPAC)	14
Clean-up	
Fat reduction	
SPE	1
SFE	2
GPC	6
Acid/base[c]	24
Fractionation	
HPLC	1
Open column[d]	28
SFE, SPE and other	4
Detection	
GC/ECD[e]	
Single column	5
Double columns	11
GC/MS[f]	
High-resolution (>1000)	8
Low-resolution (<1000)	9
Quantification[g]	
External standards	3
Internal standards	
Single	11
Several	19
Some native standards used: ^{12}C-PCB #22, 29, 72, 66, 104, 112, 155, 185, 196, 199 and 207	
Isotope-labeled standards: ^{13}C-PCB #28, 52, 101, 105, 118, 138, 153, 180, 194, 202	

[a] Solvents used: *i*-octane (dilution), toluene (Soxhlet), hexane, diethyl ether, ethanol, acetone, cyclohexane, pentane (liquid or column) and CO_2 (SFE).
[b] In the IUPAC study the sample was pre-extracted and the total weight measured was used as fat weight.
[c] H_2SO_4 treatment most common.
[d] Column materials used: modified silica, alumina, Florisil™ and carbon.
[e] GC columns used (all capillary 30–60 m): Ultra 1™, Ultra 2™, DB-5™, SPB-5™, RTX-5™, DB-DIOX™, DB-17™, SPB-17™, DB-210™, SIL19-CB™, CP-SIL5™, CP-SIL8CB™, SPB-20™ and SPB-Octyl.
[f] MS technique: SIR, EI.
[g] In both studies referred to here individual concentrations of PCBs were to be reported.

Of the 14 laboratories in the IUPAC study 8 were found to have a standard deviation (S_r) of within-laboratory repeatability of 4–17% for individual congeners, while the rest were within 45%. The standard deviation of between-laboratory reproducibility (S_L) was between 15% and 48% for nine of the labs, four of the labs were off by up to 80% and one was >100% off. The following PCB congeners were most accurately determined in the fish oils: 28, 52, 66, 74, 101, 105, 114, 118, 128, 153, 156, 167, 170, 180 and 183. Other congeners were not included when calculating the standard deviations.

In reporting total PCB levels (also denoted sum of PCBs) in a biological sample several approaches can be used and will eventually result in greater deviations in the results generated in different laboratories than when reporting concentrations of single congeners. The comparison, for instance in trend studies, of total levels of PCBs reported in biological materials today compared with reports from the 1960s to the 1980s, based on quantification with Aroclor™ and other technical mixtures as standards, is a problem. Also the use of different quantification techniques and lack of documentation of more precise calculations make comparisons very difficult.

From the interlaboratory studies mentioned one can conclude that the *state-of-the-art methodology* today for PCB determination, as far as this is defined by *most-used methodology*, is one that uses a ''very traditional'' liquid extraction, followed by ''traditional'' open-column chromatography clean-up, and finally a ''less traditional'' detection and quantification step. (''Traditional'' is used to indicate the progression in methodology since the early 1950s.) Further, in comparing results between different laboratories the reported PCB levels in the same sample can be assumed to have a standard deviation of less than 50%.

ABBREVIATIONS AND ACRONYMS

ASE™	Accelerated Solvent Extraction
DDT	Dichlorodiphenyltrichloroethane
ECD	Electron Capture Detection
EI	Electron Impact
EROD	7-Ethoxyresorufin-*O*-deethylase
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GC/NCIMS	Gas Chromatography/Negative Chemical Ionization Mass Spectrometry
GC/TOFMS	Gas Chromatography/Time-of-flight Mass Spectrometry
GPC	Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
HRGC	High-resolution Gas Chromatography
HRGC/HRMS	High-resolution Gas Chromatography/ High-resolution Mass Spectrometry
HRMS	High-resolution Mass Spectrometry
IUPAC	International Union of Pure and Applied Chemistry
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NCI	Negative Chemical Ionization
NCIMS	Negative Chemical Ionization Mass Spectrometry
PCB	Polychlorinated Biphenyl
PCG	Porous Carbon Gel
POP	Persistent Organic Pollutant
ppt	Parts per Trillion
QA/QC	Quality Assurance/Quality Control
RRF	Relative Response Factor
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SFE/GC/MS	Supercritical Fluid Extraction/Gas Chromatography/Mass Spectrometry
SFE/LC	Supercritical Fluid Extraction/Liquid Chromatography
SIM	Selective Ion Monitoring
SIR	Selective Ion Recording
SPE	Solid-phase Extraction
SPMD	Semipermeable Membrane Device
TEF	Toxic Equivalency Factor
TOFMS	Time-of-flight Mass Spectrometry
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
WHO	World Health Organization

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Supercritical Fluid Chromatography in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Waste Extraction Procedures

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Lipid Analyses in Food • Liquid Chromatography in Food Analysis • Sample Preparation Analytical Techniques for Food

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Pharmaceuticals and Drugs **(Volume 8)**
Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers • Supercritical Fluid Chromatography of Polymers

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Biopolymer Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Isotope Ratio Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

REFERENCES

1. K. Ballschmiter, M. Zell, 'Analysis of Polychlorinated Biphenyls (PCB) by Glass Capillary Gas Chromatography', *Fresenius' Z. Anal. Chem.*, **302**, 20–31 (1980).
2. S. Jensen, 'Report of a New Chemical Hazard', *New. Sci.*, **32**, 612 (1966).
3. C. Bernes, 'Persistent Organic Pollutants. A Swedish View of an International Problem', *Swedish Environmental Protection Agency Monitor*, **16**, 9–23 (1998).
4. S. Jensen, A.G. Johnels, M. Olson, G. Oterlind, 'DDT and PCBs in Marine Animals from Swedish Waters', *Nature*, **224**, 247–250 (1969).
5. S. Jensen, 'The PCB Story', *Ambio*, **1**(4), 123–131 (1972).
6. R.G. Webb, A.C. McCall, 'Quantitative PCB Standards for Electron Capture Gas Chromatography', *J. Assoc. Off. Anal. Chem.*, **55**(4), 746–752 (1972).

7. L.D. Sawyer, 'Quantification of Polychlorinated Biphenyl Residues by Electron Capture Gas–Liquid Chromatography: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **61**(2), 282–291 (1978).
8. R.G. Webb, A.C. McCall, 'Identities of Polychlorinated Biphenyl Isomers in Aroclors™', *J. Assoc. Off. Anal. Chem.*, **55**, 746–752 (1972).
9. G.B. Collins, D.C. Holmes, F.J. Jackson, 'The Estimation of Polychlorobiphenyls', *J. Chromatogr.*, **71**, 443–449 (1972).
10. R. Turtle, R.J. Norstrom, B.G. Collins, 'Comparison of PCB Quantification Methods: Reanalysis of Archived Specimens of Herring Gull Eggs from the Great Lake', *Chemosphere*, **22**, 201–213 (1991).
11. M.S. Wolf, J. Thornton, A. Fischbein, R. Lilis, I.J. Selikoff, 'Disposition of Polychlorinated Biphenyl Congeners in Occupational Exposed Persons', *Toxicol. Appl. Pharmacol.*, **62**, 294–306 (1982).
12. V.W. Burse, D.F. Groce, M.P. Korver, P.C. McClure, J.F. Donahue, L.L. Needham, 'Determination of Selected Organochlorine Pesticides and Polychlorinated Biphenyls in Human Serum', *J. Anal. Toxicol.*, **14**, 137–141 (1990).
13. K. Ballschmiter, M. Zell, 'Analysis of Polychlorinated Biphenyls (PCBs) by Glass Capillary Gas Chromatography. Composition of Technical Aroclor™- and Clophen-PCB Mixtures', *Fresenius' Z. Anal. Chem.*, **302**, 20–31 (1980).
14. B. van Bavel, P. Andersson, H. Wingfors, P.-A. Bergqvist, L. Norrgren, C. Rappe, M. Tysklind, 'Multivariate Modelling of PCB Bioaccumulation in Three-spined Stickleback (*Gasterosteus aculeatus*)', *Environ. Toxicol. Chem.*, **15**, 299–307 (1996).
15. M. Tysklind, A.T.C. Bosveld, P. Andersson, E. Verhallen, T. Sinnige, W. Seinen, C. Rappe, M. van den Berg, 'Inhibition of Ethoxyresorufin-*O*-deethylase (EROD) Activity in Mixtures of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin and Polychlorinated Biphenyls', *Environ. Sci. Pollut. Res.*, **2**(4), 211–216 (1995).
16. M. van den Berg, L. Birnbaum, A.T.C. Bosveld, B. Brunström, P. Cook, M. Feeley, J.P. Geisy, A. Hanberg, R. Hasegawa, S.W. Kennedy, T. Kubiak, J.C. Larsen, F.X.R. van Leeuwen, A.K.D. Liem, C. Nolt, R.E. Peterson, L. Poellinger, S. Safe, D. Schrenk, D. Tllitt, M. Tysklind, M. Younes, F. Waern, T. Zacharewski, 'Toxic Equivalency Factors (TEFs) for PCBs, PCDDs, PCDFs for Humans and Wildlife', *Environ. Health Perspect.*, **106**(12), 775–791 (1998).
17. M.D. Mullin, C.M. Pochini, S. McCrindle, M. Romkes, S.H. Safe, L.M. Safe, 'High-resolution PCB Analysis and Chromatographic Properties of All 209 PCB Congeners', *Environ. Sci. Technol.*, **18**, 468–476 (1984).
18. G.M. Frame, 'A Comprehensive, Multicenter Study of PCB Elution on Capillary GC columns', Proceedings Pittcon '95, New Orleans, 5–10 March, paper 717, 1995.
19. B. Larsen, J. Riego, 'Interference from 2,3,5,6,3′,4′-Hexachlorobiphenyl (CB 163) in the Determination of 2,3,4-2′,4,5-Hexachlorobiphenyl (CB 138) in Environmental and Technical Samples', *Int. J. Environ. Anal. Chem.*, **40**, 59–68 (1990).
20. M.D. Erickson, *Analytical Chemistry of PCBs*, 2nd edition, CRC Lewis Publishers, Boca Raton, 1997.
21. J.D. Lovelick, S.R. Lipsky, 'Electron Affinity Spectroscopy – a New Method for the Identification of Functional Groups in Chemical Compounds Separated by Gas Chromatography', *J. Am. Chem. Soc.*, **82**, 431 (1960).
22. V.S. Ong, R.A. Hites, 'Determination of Pesticides and Polychlorinated Biphenyls in Water: a Low Solvent Method', *Environ. Sci. Technol.*, **29**, 1259–1266 (1995).
23. P.G. Murphy, 'Sulphuric Acid for the Cleanup of Animal Tissues for the Analysis of Acid Stable Chlorinated Hydrocarbon Residues', *J. Assoc. Off. Anal. Chem.*, **55**, 1360–1362 (1972).
24. S. Tanabe, N. Kannan, T. Wakimoto, R. Tatsukawa, 'Method for the Determination of Three Toxic Non-*ortho* Chlorine Substituted Coplanar PCBs in Environmental Samples at ppt Levels', *Int. J. Environ. Anal. Chem.*, **29**, 199–213 (1987).
25. D.L. Stalling, R.C. Tindle, J.L. Johnsson, 'Cleanup of Pesticides and Polychlorinated Biphenyl Residues in Fish Extracts by Gel Permeation Chromatography', *J. Assoc. Off. Anal. Chem.*, **55**, 32–38 (1972).
26. M.A. Ribick, G.R. Dubay, J.D. Petty, D.L. Stalling, C.J. Schmitt, 'Toxaphene Residues in Fish: Identification, Quantification, and Confirmation at Part per Billion Levels', *Environ. Sci. Technol.*, **16**, 310–318 (1982).
27. D.W. Kuehl, E.N. Leonard, 'Isolation of Xenobiotic Chemicals from Tissue Samples by Gel Permeation Chromatography', *Anal. Chem.*, **50**, 182–185 (1978).
28. J.N. Huckins, N.W. Tubingen, J.A. Lebo, R.W. Gale, T.W. Schwartz, 'Polymeric Film Dialysis in Organic Solvent Media for Cleanup of Organic Contaminants', *J. Assoc. Off. Anal. Chem.*, **73**, 290–293 (1990).
29. J. Meadows, D. Tillitt, J. Huckins, D. Schroeder, 'Large-scale Dialysis of Sample Lipids Using a Semipermeable Membrane Device', *Chemosphere*, **26**, 1993–2006 (1993).
30. B. Stranberg, P.-A. Bergqvist, C. Rappe, 'Dialysis with Semipermeable Membranes as an Efficient Lipid Removal Method in the Analysis of Bioaccumulative Chemicals', *Anal. Chem.*, **70**, 526–533 (1998).
31. J.N. Huckins, G.K. Manuweera, J.D. Petty, D. Mackay, J.A. Lebo, 'Lipid-containing Semipermeable Membrane Devices for Monitoring Organic Contaminants in Water', *Environ. Sci. Technol.*, **27**, 2489–2496 (1993).
32. P.-A. Bergqvist, B. Strandberg, R. Ekelund, C. Rappe, Å. Granmo, 'Temporal Monitoring of Organochlorine Compounds in Seawater by Semipermeable Membranes following a Flooding Episode in Western Europe', *Environ. Sci. Technol.*, **32**, 3887–3892 (1998).

33. H.G. Nowicki, 'Application of Azulene as Visual Aid to Monitor Column Chromatographic Fractionation of Samples for Pesticide and Polychlorinated Biphenyl Determination', *J. Assoc. Off. Anal. Chem.*, **64**, 16–18 (1981).
34. Å. Bergman, L. Reutergårdh, M. Åhlman, 'Determination of Polychlorinated Biphenyls in Commercial Silica Gel by Capillary Gas Chromatography', *J. Chromatogr.*, **291**, 392–397 (1984).
35. O.W. Berg, P.L. Diosady, G.A.V. Rees, 'Column Chromatographic Separation of Polychlorinated from Chlorinated Hydrocarbon Pesticides, and their Subsequent Gas Chromatographic Quantitation in Terms of Derivatives', *Bull. Environ. Contam. Toxicol.*, **7**, 338–347 (1972).
36. S. Jensen, G. Sundström, 'Structures and Levels of most Chlorobiphenyls in Two Technical PCB Products and in Human Adipose Tissue', *Ambio*, **3**, 70–76 (1974).
37. D.L. Stalling, J.N. Huckins, J.D. Petty, J.L. Johnson, H.O. Sanders, 'An Expanded Approach to Study and Measurement of PCBs and Selected Planar Halogenated Aromatic Environmental Pollutants', *Ann. NY Acad. Sci.*, **320**, 48–59 (1979).
38. C.S. Creasar, F. Krokos, J.R. Startin, 'Analytical Methods for the Determination of Non-*ortho* Substituted Chlorobiphenyls: a Review', *Chemosphere*, **25**, 1981–2008 (1992).
39. C.S. Creasar, A. Al-Haddad, 'Fractionation of Polychlorinated Biphenyls, Polychlorinated Dibenzo-*p*-dioxins, and Polychlorinated Dibenzofurans on Porous Graphite Carbon', *Anal. Chem.*, **61**, 1300–1302 (1989).
40. L.M. Smith, T.R. Schwartz, K. Feltz, T.J. Kubiak, 'Determination and Occurrence of AHH-active Polychlorinated Biphenyls, 2,3,7,8-Tetrachloro-*p*-dioxin and 2,3,7,8-Tetrachlorodibenzofuran in Lake Michigan Sediment and Biota', *Chemosphere*, **21**, 1063–1085 (1990).
41. K. Lundgren, P.-A. Bergqvist, P. Haglund, C. Rappe, 'Separation of Polychlorinated Aromatic Pollutants on a HPLC Carbon Column', *Organohal. Comp.*, **8**, 103–106 (1992).
42. Y. Zebühr, C. Näf, C. Brandh, D. Broman, R. Ishaq, H. Pettersen, 'An Automated HPLC Separation Method with 2 Coupled Columns for the Analysis of PCDD/Fs, PCBs and PACs', *Chemosphere*, **27**, 1211–1219 (1993).
43. W.Y. Shiu, D. Mackay, 'A Critical Review of Aqueous Solubilities, Vapor Pressures, Henry's Law Constants, and Octanol–Water Partition Coefficients of the Polychlorinated Biphenyls', *J. Phys. Chem. Ref. Data*, **15**, 911–929 (1986).
44. S.B. Hawthorne, J.D. Miller, 'Extraction and Recovery of Organic Pollutants from Environmental Solids and Tenax-GC using Supercritical CO_2', *J. Chromatogr. Sci.*, **24**, 258–264 (1986).
45. S.B. Hawthorne, D.J. Miller, 'Direct Coupled Supercritical Extraction Gas Chromatographic Analysis of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls from Environmental Samples', *J. Chromatogr.*, **403**, 63–76 (1987).
46. S.B. Hawthorne, 'Analytical-scale Supercritical Fluid Extraction', *Anal. Chem.*, **62**, 633A–642A (1990).
47. M.H. Liu, S. Kapila, A.F. Yanders, T.E. Clevenger, A.A. Elseewi, 'Role of Entrainers in Supercritical Fluid Extraction of Chlorinated Aromatic Compounds from Soil', *Chemosphere*, **23**, 1085–1095 (1991).
48. S.B. Hawthorne, J.J. Langenfield, D.J. Miller, M.D. Burford, 'Comparison of Supercritical $CHClF_3$, N_2O and CO_2 for the Extraction of Polychlorinated Biphenyls and Polychlorinated Aromatic Hydrocarbons', *Anal. Chem.*, **64**, 1614–1622 (1992).
49. J.J. Langenfield, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Effects of Temperature and Pressure on Supercritical Fluid Extraction Efficiencies of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls', *Anal. Chem.*, **65**, 338–344 (1993).
50. E.G. van der Velde, W. de Haan, A.K.D. Liem, 'Supercritical Fluid Extraction of Polychlorinated Biphenyls and Pesticides from Soil. Comparison with other Extraction Methods', *J. Chromatogr.*, **626**, 135–143 (1992).
51. S. Bøwadt, B. Johansson, S. Wunderli, M. Zennegg, L.F. de Alencastro, D. Granjean, 'Independent Comparison of Soxhlet and Supercritical Fluid Extraction for the Determination of PCBs in an Industrial Soil', *Anal. Chem.*, **67**, 2424–2430 (1995).
52. M.V. Djordjevic, D. Hoffman, J. Fan, B. Prokopczyk, M.L. Citron, S.D. Stellman, 'Assessment of Chlorinated Pesticides and Polychlorinated Biphenyls in Adipose Breast Tissue using a Supercritical Fluid Extraction Method', *Carcinogenesis*, **15**, 2581–2585 (1994).
53. K.S. Nam, S. Kapila, A.F. Yanders, R.K. Puri, 'Supercritical Extraction and Cleanup for the Determination in Biological Samples', *Chemosphere*, **20**, 873–880 (1990).
54. S. Bøwadt, B. Johansson, P. Fruekilde, M. Hanson, D. Zille, B. Larsen, J. de Boer, 'Supercritical Fluid Extraction of Polychlorinated Biphenyls from Lyophilised Fish Tissue', *J. Chromatogr., A*, **675**, 189–204 (1994).
55. R.C. Hale, M.O. Gaylor, 'Determination of PCBs in Fish Tissue Using Supercritical Fluid Extraction', *Environ. Sci. Technol.*, **29**, 1043–1047 (1995).
56. B. van Bavel, M. Järemo, L. Karlsson, G. Lindström, 'Development of a Solid Phase Carbon Trap for Simultaneous Determination of PCDDs, PCDFs and PCBs in Environmental Samples Using SFE-LC', *Anal. Chem.*, **68**, 1278–1283 (1996).
57. H.R. Johanson, G. Becher, T. Greibokk, 'Determination of PCBs in Biological Samples Using On-line SFE-GC', *Fresenius' J. Anal. Chem.*, **344**, 486–491 (1992).
58. H.R. Johanson, G. Becher, T. Greibokk, 'Determination of Planar PCBs by Combining On-line SFE-HPLC and ECD or GC/MS', *Anal. Chem.*, **66**, 4068–4073 (1994).
59. B. van Bavel, B.P. Dahl, L. Karlsson, L. Hardell, C. Rappe, G. Lindström, 'Supercritical Fluid Extraction of PCBs from Human Adipose Tissue for HRGC/LRMS Analysis', *Chemosphere*, **30**, 1229–1236 (1995).

Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Michal Fisher and Israel Schechter
Israel Institute of Technology, Haifa, Israel

This chapter reviews conventional and novel methods of analysis of polynuclear aromatic hydrocarbons in environmental samples. Sampling and pretreatment techniques of air, soil and water are discussed, as well as various clean-up and extraction methods. Most of the relevant polycyclic aromatic hydrocarbons (PAHs) separation techniques are reviewed. The principles and performance of the common PAH detection systems are presented. An introduction to novel analytical techniques, which will probably be introduced in the near future, is also provided.

1 Introduction **1**
1.1 Definition of Polycyclic Aromatic Hydrocarbons 1
1.2 Formation and Emission of Polycyclic Aromatic Hydrocarbons 1
1.3 Health, Biological and Environmental Hazards of Polycyclic Aromatic Hydrocarbons 2
1.4 Environmental Occurrence of Polycyclic Aromatic Hydrocarbons 2
2 Sampling, Pretreatment and Extraction Techniques for Polycyclic Aromatic Hydrocarbons **3**
2.1 Environmental Sampling of Polycyclic Aromatic Hydrocarbons 3
2.2 Extraction Methods 4
2.3 Clean-up Techniques 6
3 Separation and Analysis of Polycyclic Aromatic Hydrocarbons **6**
3.1 Gas and Liquid Chromatographic Techniques 6
3.2 Optical Methods 9
3.3 Photoelectric Aerosol Sensor 14
3.4 Mass Spectrometry 15
3.5 Immunological Methods for the Detection and Quantification of Polycyclic Aromatic Hydrocarbons 17
4 Recent Advances in the Analysis of Environmentally Occurring Polycyclic Aromatic Hydrocarbons **19**
4.1 Laser-induced Multiphoton Ionization/Fast Conductivity 19
4.2 Multiphoton Ionization–Time-resolved Ion-induced Charges 20
4.3 Imaging Methods 20
5 Concluding Remarks **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **23**

1 INTRODUCTION

The human environment contains numerous anthropogenic substances. Most of these compounds are ignored, because there is insufficient technology at present to analyze them in environmental samples.[1] Many of the compounds that can be detected are considered harmful. An important class of such substances is PAHs. Since many of these compounds are known carcinogenic or mutagenic agents, their environmental concentrations are restricted by legislation in many countries. Therefore, development of new analytical methods for fast and reliable detection of PAHs is of considerable environmental importance.

1.1 Definition of Polycyclic Aromatic Hydrocarbons

PAHs comprise a class of condensed multimembered benzenoid-ring compounds. The parent homocyclic PAHs contain only carbon and hydrogen atoms. Polycyclic aromatic compounds (PACs) is a broader definition, which also includes thousands of substituted PAHs that contain various substitute groups such as alkyl, amino, hydroxy, oxy, thio, chloro, or cyano groups. In addition, another group of PACs includes compounds that contain hererocyclic substituents. This review will focus mainly on the environmental analysis of PAHs and their derivatives.

1.2 Formation and Emission of Polycyclic Aromatic Hydrocarbons

PAHs are generally formed during incomplete combustion, pyrosynthesis or pyrolysis of organic matter containing carbon and hydrogen, for example, coal, tar, wood, gasoline, and diesel fuel. One of the common mechanisms of PAH formation is pyrosynthesis. Hydrocarbons with low molecular mass, including methane, may serve as precursors to PAHs. At temperatures of ~500–800 °C, carbon–hydrogen and carbon–carbon bonds are broken and form free radicals. The free radicals, C_1, C_2, and higher radicals undergo dehydrogenation

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and combine rapidly, forming aromatic ring structures. Another mechanism of PAH formation is pyrolysis, the ''cracking'' of organic compounds, such as higher alkanes, forming radicals. These radicals combine and form condensed aromatic molecules. The final products of this procedure are those which are thermodynamically preferred.[2]

1.3 Health, Biological and Environmental Hazards of Polycyclic Aromatic Hydrocarbons

Many of the PAH and PAC compounds are known carcinogens and/or mutagens. Among the first PAHs that were found to be carcinogenic in laboratory animals were dibenz[*a*,*h*]anthracene and benzo[*a*]pyrene.[3,4] The carcinogenicity of PAHs depends upon activation by a class of enzymes, cytochromes P-450. When oxygen is added to PAHs by these enzymes, the product is an epoxide, which is a potent electrophile. The epoxide, generated inside the cell, can diffuse into the nucleus and react strongly with DNA, resulting in gene altering.[5]

Some microorganisms have the ability to metabolize certain PAHs, e.g. benzo[*a*]pyrene, forming a mixture of metabolites. Some of these metabolites, epoxides for instance, are carcinogenic in higher organisms, but some derivatives are detoxified derivatives.[6]

1.4 Environmental Occurrence of Polycyclic Aromatic Hydrocarbons

Since there are many processes of incomplete combustion and pyrolysis that favor production of PAHs, these pollutant compounds are abundantly encountered in the atmosphere, soil, and water. The occurrence of PAHs can be divided into two major categories. One involves natural processes such as forest burning, and the other category involves human-made procedures, e.g. aluminum production, wood burning, wood stove smoke, coal steam power plants, oil and natural gas combustion, and motor vehicle exhaust. Coal tars and petroleum residues, such as road and roofing asphalt, also contain and emit high levels of PAHs. The levels of PAHs in the atmosphere may often reach $20\,\mu g\,m^{-3}$. These high concentrations are most likely encountered in polluted urban atmospheres and in areas of natural fires, such as forest fires. Coal furnace stack gas may contain over $1000\,\mu g\,m^{-3}$ of PAHs and cigarette smoke contains almost $100\,\mu g\,m^{-3}$ of PAHs. Owing to the low vapor pressure of many of the PAH compounds, they are most likely found in aerosols or condensed onto aerosol particles. PAHs are often adsorbed on soot particles.[7] An important source of PAHs in urban environments is the diesel exhaust. The diesel engine is more economical than the gasoline engine, therefore it has become popular. The exhaust emission contains

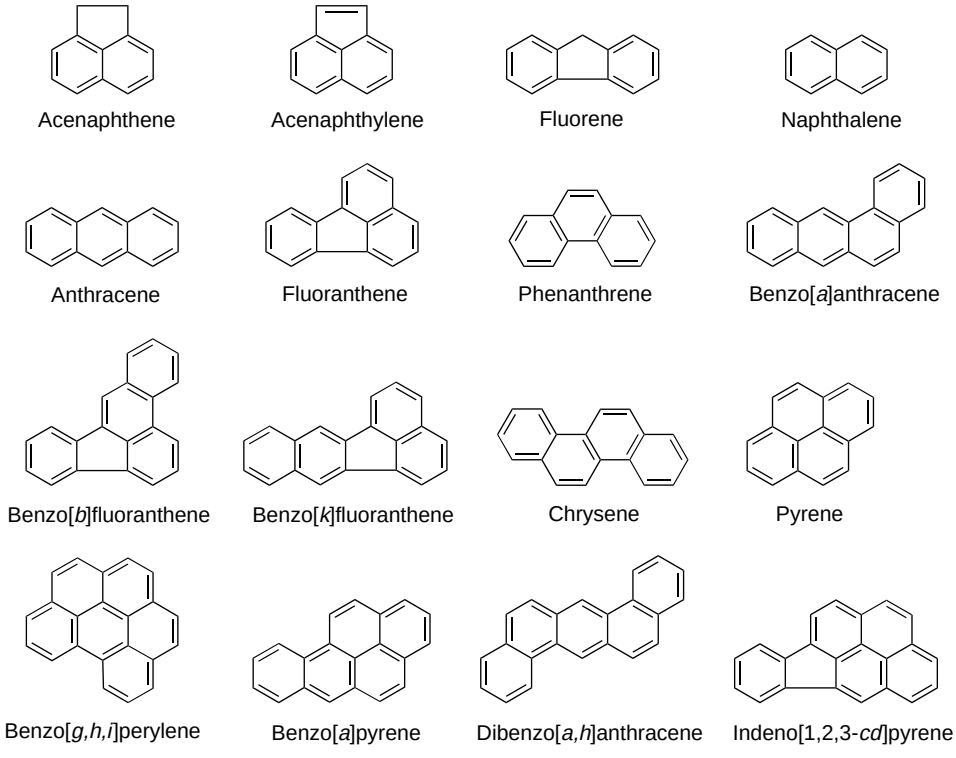

Figure 1 The 16 PAHs included in USEPA regulations.

soot and condensed organic compounds. Among these compounds are many PAHs, e.g. pyrene, fluoranthene, benz[*a*]anthracene and benz[*a*]pyrene.[8] As a result of industrial activities, the PAH levels range from 5 ppb (parts per billion) in an undeveloped area in Alaska to 1.8×10^6 ppb for an oil-refining zone in the UK.[6]

The occurrence of PAHs in water systems depends on many factors. PAHs can enter oceans by various routes, including oil spills, industrial waste and atmospheric fallout.[9] The concentration of individual dissolved PAH varies between sub-ppt ($pg\,g^{-1}$) for groundwater, to the order of ppm ($\mu g\,g^{-1}$) for highly contaminated water.[10]

PAHs in soil are found at various depths. A considerable accumulation of low molecular weight PAHs, partly of natural origin, was detected in the subsurface and in deep soils. High-molecular-weight PAHs, mostly of anthropological origin, were detected in the soil surface. These findings were repeated for a few types of soils in different places. The surface PAH contamination is probably a result of aerosol fallout.[11] It was also reported that the major fraction of PAH contamination in soil and groundwater is anthropogenic rather than natural.[12]

Although several hundreds of PAHs exist, most of the studies deal with only a limited number, namely the 16 PAHs included in the US Environmental Protection Agency (USEPA) priority pollutants list (Figure 1).

2 SAMPLING, PRETREATMENT AND EXTRACTION TECHNIQUES FOR POLYCYCLIC AROMATIC HYDROCARBONS

Conventional environmental sampling is based on the collection of the sample, followed by sample preparation, pretreatment and extraction, and finally analysis. These procedures have to be followed as a part of most standard analytical methods. However, recently developed analytical methods, which will be briefly discussed in the last section of this review, allow for direct analysis of environmental samples. It is hoped that these new techniques will avoid the time-consuming procedures of pretreatment or extraction.

2.1 Environmental Sampling of Polycyclic Aromatic Hydrocarbons

The goal of collecting environmental samples is to obtain a small fraction that reflects the properties of the evaluated population. The confidence in data quality generally depends on the number of samples and on the location and timing of sampling.[7]

2.1.1 Air Sampling

Air is comprised mainly of gas molecules. However, air also carries many organic compounds of environmental interest in aerosols. These compounds are usually found at low concentrations and may be distributed between two phases, solid–gas or liquid–gas. Low analyte concentration requires high sample volume, therefore a proper air sampling may become the bottleneck of airborne PAH analysis.

2.1.1.1 Sampling of Vapor Phase Polycyclic Aromatic Hydrocarbons Steel canister or Teflon® bags are commonly used for collecting a measured volume of air. The air is passed through a solid sorbent that traps the organic compounds. The use of solid cyclodextrins (CDs) for extraction of vapor-phase PAHs from air is a well-established method. CDs are cyclic oligosaccharides which possess a hydrophilic exterior and a less polar interior cavity. This property enables extraction of PAHs by formation of inclusion complexes.[13] Other types of solid sorbents are organic polymeric sorbents (e.g. Tenax), inorganic sorbents (e.g. silica gel and alumina), and carbon. Another technique for sampling volatile or unstable species is cryo-sampling, where the sample is condensed in a cryogenic trap. In this method, the total volume of sampled air can be calculated from the volume of the collected water.[7]

Another sampling method is passive monitoring, where the analyte is collected from air by diffusion and introduced into the monitoring device. This technique can not provide reliable information concerning the analyte concentration.

2.1.1.2 Sampling of Solid-phase Polycyclic Aromatic Hydrocarbons As mentioned previously, because of the low vapor pressure of many PAH compounds, they are most likely found in aerosols or condensed onto aerosol particles.[14] A proper analytical sampling of aerosols, which maintains the original chemical composition and does not introduce artifacts due to geometrical or morphological considerations, is a most challenging task.

Airborne particulate matter is usually sampled using filtration techniques. Porous membrane filters (e.g. cellulose esters and Teflon®), and fibrous filters (e.g. cellulose fibers and glass fibers) are commonly used for PAH sampling. During the filtration process, particles depart from the air stream, collide with the filter and attach to it. Several mechanisms are responsible for deposition of aerosols on filter-based samplers, including interception, inertial impaction, diffusion, gravitational settling, and electrostatic attraction.[15]

The above sampling methods are well established; nevertheless, they are not free of artifacts. Among these pitfalls are changes in particle size distribution due

to electrostatic deposition or vaporization and artifacts caused by various interactions. The development of on-line and in situ sampling, followed by direct analysis, has the potential to avoid artifacts in airborne PAH detection.[16]

Another recently reported sampling technique, suitable for both PAH vapor and airborne particulates, is based on a renewable water microdroplet.[17] In this method, a water microdroplet comes in contact with ambient air and quasi-equilibrium conditions are achieved. The contaminated droplet can be analyzed for PAH compounds, and a fresh one is readily produced for the next sampling cycle.

Meteorological effects such as wind velocity, wind direction, temperature and atmospheric pressure should also be considered while sampling.[7] Such information may affect the sampling validation on one hand, and may provide useful insight to the analytical results, on the other.

2.1.2 Water Sampling

Water sampling for PAH analysis depends on the type of water under consideration, for instance, surface waters (e.g. rivers, lakes), groundwater, or wastewater. Many of these water types require particular sampling procedures and treatment. Another difficulty in water sampling is the heterogeneity of water, both spatially and temporally. Organic materials containing PAHs can either float on the water surface, sink to the bottom or be suspended in the flow, depending on their weight and density.[7] Many additional problems need to be considered for accurate PAH sampling.

The relevant concentration of dissolved individual PAH compounds varies between sub-parts per trillion ($pg\,g^{-1}$) for groundwater, to the order of parts per million ($\mu g\,g^{-1}$) for highly contaminated water. The low PAH concentration samples require application of preconcentration techniques prior to analysis. Furthermore, analysis of low PAH concentrations can lead to serious errors due to analyte losses or contamination in the sampling process. PAHs can be deposited on any surface with which they come in contact during sampling and extraction.[10] It was found that PAH solutions at the ppb concentration level exhibited 80% PAH adsorption loss after 4 h of stirring in a glass container.[18] Other PAH sampling losses are related to chemical or biological degradation in the vessels. Therefore, water samples must be held in the dark, at subzero temperature in glass containers, pretreated for removing any organic traces. The best way of avoiding some of these sampling problems is the direct sampling into the extraction vessel, followed by extraction steps, without any delay.[10]

Solid adsorbents such as Tenax and Amberlite resin are convenient for sampling large volumes of water. However, this procedure leads to the loss of the volatile compounds. Liquid extraction of PAHs from water systems can also be performed using an organic solvent that could easily be separated from the water, e.g. methylene chloride.[10]

2.1.3 Soil Sampling

The heterogeneity of soil along with the fact that PAHs concentration in soil may depend on sampling depths and may possess various distributions, make soil sampling very complicated.[7] No simple sampling procedure is prescribed, but, it is usually recommended that soil samples are preserved by freeze-drying, followed by graining and sieving.

2.2 Extraction Methods

One of the most significant difficulties in analysis of environmental samples is related to the complexity of the matrix. Environmental samples are usually composed of numerous analytes and interfering matrix components. Extractions are performed in order to separate the analyte from the interference.

2.2.1 Soxhlet Extraction

Conventional extraction of PAHs involves different organic solvents, e.g. hexane, acetone, methanol, acetonitrile, toluene. An efficient method of extraction of PAHs using organic solvents is the Soxhlet extraction, in which the solvent is vaporized in a distillation flask, rises through an exterior tube to a condenser and then liquefies by cooling. The liquid drops onto a porous vessel containing the sample. The solution of the solvent and the dissolved matter is drained back into the distillation flask. This circulation is usually carried out overnight, and finally, the solution containing the extracted matter is collected. Automatic devices for speeding up the extraction are available. Soxhlet extraction of PAHs usually serves as a reference for new extraction techniques.

2.2.2 Alkaline Saponification

Alkaline saponification is an effective extraction method based on hydrolysis of esters, using KOH/methanol mixture. This method is especially useful for PAH extraction from contaminated soil samples. One of the major problems in PAH extraction from soil is related to the fact that PAH molecules might be strongly bounded to the soil matrix (humic polymer).[19] Therefore, conventional organic solvent extraction is not entirely effective, and a more rigid extraction procedure is necessary in these cases. Application of saponification as an extraction tool, in addition to conventional

organic solvent extraction, succeeded in recovering a significant higher amount of PAH compounds from soil, comparing to the amount obtained by organic solvent extraction alone.[19–21] These findings were attributed to cleavage of ester bonds, which form parts of the humic macromolecules. This process results in releasing PAH molecules trapped in the humic polymer.[19]

2.2.3 Sonication

Sonication is another traditional extraction technique, based on cold liquid–solid extraction. Originally, mechanical shaking was applied and, later, better results were accomplished by ultrasonic treatment. Nevertheless, the recoveries of PAHs extracted by sonication are lower than those obtained by other techniques, e.g. Soxhlet extraction.[22]

Recently, several other extraction techniques have been developed for the determination of PAHs in environmental matrices such as solids and sediments. These techniques are supposed to improve recoveries of PAHs and to reduce both the extraction time and the amount of solvent used. Among these techniques are supercritical fluid extraction (SFE), accelerated solvent extraction (ASE), and microwave-assisted extraction (MAE), which will briefly be discussed in the following sections.

2.2.4 Supercritical Fluid Extraction

SFE is a highly promising alternative to conventional Soxhlet extraction of environmental samples. Heating a gas above its critical temperature or compressing a liquid above its critical pressure produces supercritical fluids. Under these conditions, the molar volume is the same whether the original form was a liquid or a gas. Due to its physical properties, i.e. low critical pressure and temperature, CO_2 has become a major solvent for SFE. The properties of supercritical fluids provide some advantages for analytical extractions. Supercritical fluids may have solvating powers similar to liquid organic solvents, but with higher diffusivity, lower viscosity, and lower surface tension. The solvating power can be adjusted by changing the pressure or temperature, or by adding modifiers, e.g. methanol, methylene chloride and toluene to the supercritical fluid. Further advantages of using supercritical fluids for environmental sample pretreatment are the low cost and the facts that they are contaminant free. Moreover, their safe disposal is less expensive than the cost of organic solvents.[23]

SFE has been used to extract PAHs from solid environmental samples, including complex matrices and analytes. Attempts were made to investigate the effects of various factors affecting PAH recoveries in SFE, such as matrix characteristics, analytes, extraction fluid, modifiers, temperature and pressure.[22,24–30] It was found that the PAH recoveries from various matrices differ significantly due to the interactions between the matrix, the analytes and the supercritical fluid. The effects of matrix characteristics on the recoveries of five PAHs (anthracene, pyrene, chrysene, benzo[*a*]pyrene, and perylene) were recently investigated. The PAHs were adsorbed on two series of adsorbent matrices, i.e. polystyrene divinylbenzene copolymer (XAD resin) and silica. It was found that the recoveries of the PAHs were relatively higher if they were extracted from a matrix of a smaller surface area and larger pore size. Considering the analyte–matrix interactions, for a given matrix, it was reported that the PAH recoveries decreased with the increasing molecular weight. In addition, the analyte desorption improved when the temperature was increased.[26]

The effectiveness of SFE of different PAHs from natural matrix standard reference materials (SRMs), e.g. water sediments, diesel particles, and urban dust, were also studied. The results revealed the limitation of SFE. The recoveries of specific compounds extracted from these samples were matrix dependent. The lowest recoveries of high-molecular-weight PAHs were obtained from diesel particles.[25] In these cases the recovery figures were much lower (<20%) than the Soxhlet results.

Optimization of SFE of PAHs from real soil samples were also studied and compared to Soxhlet extraction. It was established that low-molecular-weight PAHs (in the range of 128–178 Da) can be extracted with pure CO_2, while high molecular weight PAHs (202–278 Da) can only be extracted using modifiers added to CO_2. A higher temperature was found to improve recoveries of all PAHs under USEPA control. The addition of small amount of water also improved extractability of these PAHs because of the swelling effect of the matrix.[22] SFE was accepted as an official method by the USEPA.[31]

2.2.5 Accelerated Solvent Extraction

ASE uses a conventional solvent up to 200 °C as an extraction medium. At high temperatures the solvating ability of certain solvents increases. As a result, PAH recoveries also increase due to changes in diffusion coefficients. In this method, extraction is performed under pressure (0.3–20 kPa), in order to keep the solvent in the liquid phase. Under these conditions, extraction kinetics and desorption of the analyte from the matrix are significantly enhanced. As far as extraction of high molecular weight PAHs from diesel materials was concerned, ASE was proved as an efficient method, compared to both Soxhlet extraction and SFE.[22,25,32] The USEPA also accepted the ASE as an official extraction method for PAHs.[33]

2.2.6 Microwave-assisted Extraction

Microwave energy is used as a heat source that is capable of heating the whole sample simultaneously without heating the vessel. In addition, in a sealed apparatus, the solvent may reach a higher boiling point. The combination of the higher boiling point and the rapid heating increases extraction efficiency and reduces the overall extraction time.[34] Microwave energy was used for extraction of PAHs from contaminated soil and sediment samples. The results indicate that in many cases MAE is an achievable alternative to the traditional Soxhlet and sonication techniques.[34,35]

2.3 Clean-up Techniques

The extraction techniques (section 2.2) are optimized for providing a high PAH recovery. Therefore the extract solution may contain a significant amount of side materials (co-extracts), such as other organic compounds. Clean-up methods for PAHs are applied to extract solutions in order to eliminate sample interference. The most commonly used clean-up techniques for PAHs are briefly discussed in the following sections.

2.3.1 Solid-phase Extraction Clean-up

Liquid adsorption chromatography for PAH clean-up is performed in normal phase using alumina (USEPA method 3611),[36] Florisil® (USEPA method 3620),[37] and silica gel (USEPA method 3630).[38] These columns may be prepared in the laboratory or solid-phase extraction (SPE) cartridges can be used.[39] SPE using a Florisil® column and dichloromethane was utilized for clean-up of PAHs in sewage sludge and soils.[40] SPE using silica column and methanol–dichloromethane (2 : 3) was also applied for clean-up of PAHs in sediments.[41]

2.3.2 Gel Permeation Chromatography Clean-up

Gel permeation chromatography (GPC), also called size-exclusion chromatography (SEC), utilizes porous particles in order to separate molecules by their different size. This method is capable of separating high molecular-weight materials from the sample. The GPC method was approved by the USEPA (method 3640)[42] and was recommended for PAH clean-up.

2.3.3 Acid–Base Partitioning Clean-up

The acid–base partitioning clean-up method is based on liquid–liquid partitioning for separating acid analytes, e.g. organic acids and phenols, from basic or neutral analytes such as PAH compounds, using a proper pH adjustment. The solvent extract (from a prior extraction procedure) is shaken with strongly basic water. The acid analytes partition into the aqueous layer, whereas the basic and natural compounds stay in the organic solvent. The latter fraction, containing the PAH compounds, is concentrated and is then ready for further clean-up or analysis. This method was also approved by the USEPA as a clean-up method (method 3640).[43]

3 SEPARATION AND ANALYSIS OF POLYCYCLIC AROMATIC HYDROCARBONS

3.1 Gas and Liquid Chromatographic Techniques

PAHs are routinely analyzed by chromatographic procedures. Two major types of chromatography are commonly used, namely, gas chromatography (GC) and high-performance liquid chromatography (HPLC). GC is most useful for smaller PAH molecules, because they are generally volatile. HPLC is suitable for high molecular weight or nonvolatile PAHs. Generally, GC is used for PAHs of up to 24 carbons, whereas HPLC can be used for PAHs as small as naphthalene. The upper limit for HPLC is the condensed PAHs with 38 carbons (13 rings).[44] Actually, the basic requirement for HPLC analysis is that the sample is soluble in the mobile phase, and many PAHs fulfill this simple requirement.

3.1.1 Gas Chromatography

The high resolution and short analysis times of capillary GC have made it the most popular analytical technique for PAHs. Capillary columns, with resolutions of several hundred thousand, can separate extremely complex samples such as tars and oils. The high resolution is especially needed in environmental PAH analysis due to the extreme complexity of most samples. These samples may contain many types of PAH compounds and, sometimes, also several isomers.[44]

3.1.1.1 The Stationary Phase The stationary phases in capillary GC was initially statically coated onto the column walls. These phases were thermally unstable at high temperatures. This limited their application in separation of low-molecular-weight PAHs. This limitation was overcome by the development of cross-linked polymeric or chemically bonded GC phases. These stationary phases are thermally stable (up to more than 100 °C), more reproducible, and more uniformly coated. In many cases, the GC separation of PAHs is very difficult because chemical affinities and boiling points of many PAH isomers are very similar. Sometimes the three-dimensional (3-D) structure of some isomers is slightly different and GC separation is possible due to stationary phases capable of

''molecular shape'' differentiation.[44] Liquid–crystalline phases have been shown to provide such capabilities.[45,46]

A significant advancement in this field was achieved by coupling liquid–crystalline to a stable polysiloxane polymer, which made the synthesis of bound liquid–crystalline possible.[47] Liquid–crystal phases can separate the PAHs by their molecular shapes, since their separation is based on permeation of the PAH molecules between the layered phase. Correlations between various PAH topological and geometrical parameters have been studied and reported.[48,49]

3.1.1.2 Gas Chromatography Detectors for Analysis of Polycyclic Aromatic Hydrocarbons The mass spectrometer (MS) is a most widely used GC detector for PAH analysis. MS is used to determine molecular weights, and also some structural information on the analyzed PAHs. The fragmentation of PAH compounds may provide valuable information about PAH isomers present in the sample.

Fourier transform infrared (FTIR) spectrometers can also be coupled to GC for PAH detection and provide a high sensitivity. FTIR was found to be a powerful tool for the determination of various compounds of environmental interest. For example, this detection technique was successful in determining fluoren-9-ones in diesel particulates, while nitrogen-containing species were erroneously suggested by the GC/MS technique.[50]

Although GC detection methods based on absorbance or fluorescence are generally sensitive and somewhat selective, they are not popular for PAH analysis. The reason is the technical difficulty of maintaining these compounds in the gas phase during detection on the one hand and the lack of gas phase reference spectra on the other.

Atomic emission spectroscopic detectors are sometimes applied for monitoring the atomic emission of individual elements such as nitrogen, sulfur and oxygen in PAH derivatives or in PACs. For instance, this method was applied for the quantification of polycyclic aromatic sulfur heterocycles (PASHs) in several SRMs from the National Institute of Standards and Technology (NIST), namely, SRM 1597, coal tar; 1580, shale oil; and 1582, crude oil. In this case it was shown that due to the co-elution problem of some PAHs (e.g. phenanthrene) along with other PASHs (e.g. dibenzothiophene and naphthothiophene), the NIST concentration figures were too high.[51] The flame ionization detector (FID), which is the most common GC detector, is not of great use for PAH detection, because of its universality.[44]

3.1.2 High-performance Liquid Chromatography

Although GC resolution is much better than that of HPLC, the larger nonvolatile or thermally unstable PAHs cannot be analyzed using GC techniques. These compounds can be separated only by HPLC. The smaller (nonpolar) PAHs are soluble in the weak normal-phase solvents such as *n*-hexane or in the weak reversed-phase solvents such as methanol and acetonitrile. PAHs that are less soluble in these solvents can generally be dissolved in stronger solvents such as dichloromethane, ethyl acetate, or tetrahydrofuran (THF). Mixing of solvents and solvent gradients are often carried out in order to achieve better separation.[44] The variety of mobile phase solvents, which can be applied in HPLC separations, makes this method suitable for environmental samples containing a large number of different PAH molecules. Other HPLC parameters, such as temperature gradient and variations in the stationary phase, also improves PAH separation.

Generally, two HPLC techniques are applied: the normal phase and the reverse phase. Normal-phase separation is based on hydrogen bonding and acid–base interactions between the silica (stationary phase) and the sample molecules. Another type of normal-phase HPLC applies alumina and derivatized silica (e.g. amino-, cyano-, phenyl-silica) as stationary phase.

Separation is based on polarizability and induction effects between the stationary phase and the sample molecules. The less polar PAHs elute before the more polar PAHs (nitrogen-containing PACs, for instance). This technique can be applied for separating PAHs by their aromatic ring number, or more accurately, by their π-electron number. The stronger the interaction between the analyte's π-electrons and the polar stationary phase molecules is, the longer the retention times are. Therefore, the planarity of unsubstituted PAH compounds directly affect their retention times. Planar PAH isomers are retained longer than the non-planar ones.[52]

Another factor affecting retention time of planar PAHs is their geometrical characteristic: the isomer with a longer molecular axis (i.e. one with a higher length to breadth ratio (L/B)), has a longer retention time.[53] The reason is, again, the interaction extent of the analyte molecules and the polar-bonded phase.

Substituent groups also affect retention times: alkyl chains reduce the retention of PAHs relative to unsubstituted PAHs, probably due to a shielding effect, which reduces the interaction between the π-electrons and the polar-bonded phase. Cycloalkane substituents, on the other hand, increase retention times of PAHs, presumably due to a polarity increase.[44]

Separations of PAHs are also carried out by the reverse-phase chromatographic technique. Here, the separation utilizes a nonpolar bonded phase and a polar solvent. The separation of PAHs can be performed using monomeric or polymeric bonded phases. Monomeric

phases are composed of bonded alkyl-substituted siloxane, while polymeric phases are based on bonded alkyl-substituted polysiloxane. Monomeric-phase based separations differentiate the number of carbon atoms in the analyte molecules. The polymeric phases, however, are usually very structured and organized, and therefore, separate PAHs by their 3-D shapes. These properties result in an efficient isomeric separation.

The L/B ratio and planarity of PAHs affect the retention times as well. PAHs with larger L/B ratio retain longer both in monomeric and polymeric bonded phases.[44] Octadecylsilyl-bonded silica gel (C_{18} column) reverse-phase chromatography is commonly used for PAH separation.[54–57]

Shape selective retention of PAHs by applying other types of stationary phases, were reported. Metalloprotoporphyrins covalently bonded to silica were examined as stationary phases for reverse-phase HPLC separation of PAHs. The results showed very good shape selectivity for planar PAHs over nonplanar ones (Figure 2). The retention mechanism is believed to be primarily due to face-to-face π-interactions.[58]

The retention behavior of PAHs on acridin derivative (dodecylacridinium) stationary phase of reverse-phase HPLC was also examined. It was reported that nonplanar

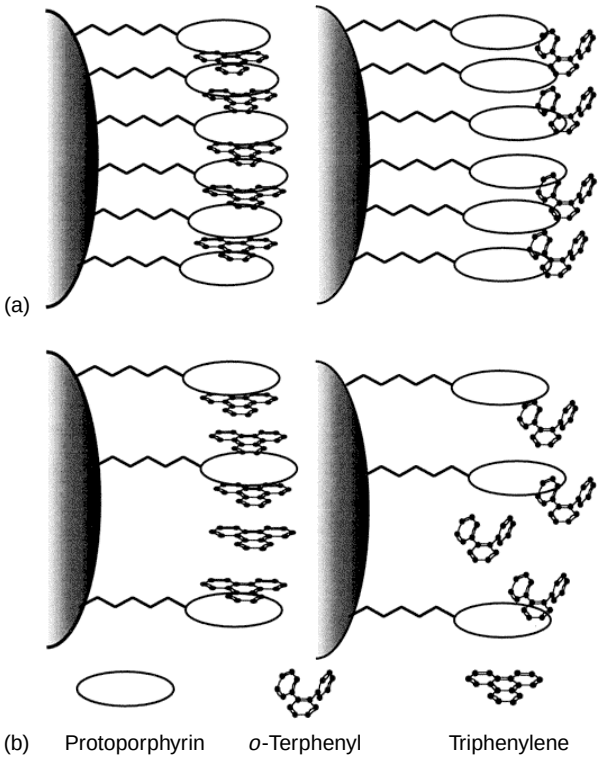

Figure 2 Schematic diagram of a slot type interaction mechanism proposed for PAHs. (Reproduced from Chen, Meyerhaff,[58] by permission of the American Chemical Society.)

PAHs exhibited smaller retention times relative to the figures obtained using octadecylsilyl-bonded silica gel. These results indicate that interactions between PAHs and the acridinyl ring dominate the PAH retention. In addition, the preferential retention of planar PAHs, rather than nonplanar ones, is attributed to the stronger interaction of the former with the acridinyl ring of the stationary phase.[59]

Recent investigations succeeded in shortening HPLC analysis times and obtaining better information.[55–57] Rapid and highly resolved separation (700 000 theoretical plates per m) of the 16 PAHs under USEPA control, was achieved by applying capillary electrochromatography. In this case, 28 kV was applied to a column packed with 1.5-μm octadecylsilica particles, using a mobile phase with low conductivity (acetonitrile mixed with sodium tetraborate). A sample containing all 16 PAHs was separated isocratically in less than 10 min.[57]

A mathematical model for fast determination of the breakthrough volumes of PAHs was also developed. This procedure was applied for calculating the breakthrough volumes of 15 PAHs in a C_{18} preconcentration column, for aqueous solutions containing 10% acetonitrile as organic modifier. This method is particularly useful for the less polar PAHs that have high breakthrough volumes (e.g. benzo[*a*]pyrene and benz[*a*]anthracene).[55]

The HPLC detectors applied for PAH analyses are required to differentiate PAHs from other compounds and to differentiate PAHs and PAH isomers as well. Therefore, the most commonly used detectors for HPLC analyses of PAHs are UV absorbance and fluorescence detectors. The energies corresponding to the electronic transitions in PAH molecules are determined by their π-electrons, thus, the electronic spectra (UV absorbance and fluorescence) of PAHs are usually more featured than other organic species. The details of the electronic transitions are governed by the size and shape of the PAH compounds.[60]

During the last few years the PAH separation by the HPLC technique has been greatly improved by multichannel detectors such as diode arrays[61] and charge-coupled devices (CCDs).[62] When these detectors are applied, the spectrally resolved transmitted or emitted light is simultaneously detected, providing a full spectrum of the analyte at any time. The full spectrum of PAHs, in addition to the retention time, is a valuable tool for PAH analysis.

Fast scanning fluorescence spectrometry detection for HPLC may also provide full spectral information of PAHs. Although spectral data are not collected simultaneously, in most cases the scanning time is adequate for unperturbed results. The combination of this method with 3-D multivariate calibration, based on retention time, emission wavelengths, and fluorescence

intensity, has provided successful analysis of PAHs in water samples.[56]

Several fluorescence based detection systems are coupled to HPLC for PAH analysis. Owing to reduced background, the sensitivity of fluorescence detection is very high, on the order of ppb. The sensitivity can be improved by up to two orders of magnitude when a laser is used as the excitation source.[63] An ultraviolet laser-induced fluorescence (UV/LIF) detection system was applied for PAH analysis by HPLC and improved sensitivity and selectivity were reported.[57]

Time-resolved laser-induced fluorescence, where fluorescence lifetimes are measured in addition to the fluorescence spectra, was successfully applied to identification of PAH isomers following HPLC separation.[64] The identification principle takes into account the fact that isomers of a similar structure and spectra may differ in their fluorescence lifetimes. Nevertheless, since fluorescence lifetimes depend on the solvent used, appropriate standards must be available for identification of unknown peaks.[44]

As previously mentioned, mass spectrometry provides a powerful detection tool for HPLC analysis. Indeed, HPLC/MS has been applied to a variety of PACs, providing informative fragmentation patterns. Although this technique is well established for routine laboratory operation, it still suffers from some inherent problems concerning the connection between the HPLC exit and the vacuum system of the MS.[44] Therefore, HPLC/MS is currently not suitable for fast environmental surveys.

3.1.3 Other Techniques

Since PAHs constitute a large class of uncharged compounds, various separation techniques have been developed by inducing selective interactions between PAHs and charged additives. The most common technique is the micellar electrokinetic chromatography, in which separation is based on the partitioning of solutes into charged micelles, containing hydrophobic cores.[65] This method was applied to PAH separation using both cationic and anionic micelles.[66] Although successful separations were reported, it was found that the binding of PAHs to the micelle core was so strong, that selectivity could barely be achieved. Moreover, this method suffers from the instability of the micelles, which may be encountered in some experimental conditions.

Capillary electrophoresis (CE) is a separation technique of high efficiency, which is based on the differences in the abilities of charged compounds to migrate in an electric field. CE was successfully applied for separation of PAHs, by using charged CD derivatives, without the presence of micelles and based solely on the formation of inclusion complexes.[67,68] The use of nonaqueous media for electrokinetic chromatography of PAHs is a promising alternative to aqueous solutions. This is due to the better solubility of PAHs in organic solvents. Moreover, certain organic solvents allow for application of higher electric field intensities and higher ionic strengths. Recently, separation of PAHs by nonaqueous CE was successfully performed using charge transfer (CT) complexation with planar organic cations (2,4,6-triphenylpyrylium ion and tropylium ion). It was reported that 2,4,6-triphenylpyrylium separates PAHs primarily according to their size, whereas for tropylium ion, separation is based on both PAH size and shape.[69]

3.2 Optical Methods

3.2.1 Ultraviolet/Visible Absorption Spectroscopy

Ultraviolet/visible (UV/VIS) absorption spectroscopy is a simple and useful tool for PAH identification purposes. The spectral information can be enhanced and improved by application of low-temperature UV/VIS absorption spectroscopy. One of the most useful applications of this method utilizes the Shpol'skii effect in an appropriate solvent type at very low temperature (4 K).

Derivative spectroscopy is another method for obtaining reach-structured spectral information, which is useful for PAH identification purposes. Although the use of derivative spectroscopy decreases the signal-to-noise ratio, it may improve the detection of spectral bands in a broad background, or of shoulders in a broad main band. The combination of a low-temperature technique and derivative spectroscopy yields quite characteristic spectra that can be used for identification of PAHs in complex mixtures such as environmental samples.[70]

Absorption spectroscopy, as a whole, suffers of low sensitivity. Therefore probably the most commonly used application of UV absorption spectroscopy in PAH analysis is for HPLC detection systems (as described in section 3.1.2).

3.2.2 Fluorescence Spectrometry

Luminescence spectroscopic techniques are widely recognized as excellent analytical tools, especially due to their high inherent sensitivity (zero-background measurements). Many PAH compounds possess strong fluorescence; therefore, several analytical techniques based on this property were developed.[71]

Unfortunately, the fluorescence spectra of different PAHs present in an environmental sample often overlap. Therefore, identification and quantification of PAHs are not possible under these conditions. However, one can still perform chromatographic separation of a PAH mixture, prior to the fluorimetric analysis of the individual compounds.

Various methods have been developed in order to reduce the spectral overlap and to improve the quality of the information obtained by fluorescence analysis of environmental samples. Among these methods are synchronous fluorescence spectrometry (SFS), full excitation–emission fluorescence spectrometry (EEM), Shpol'skii spectroscopy (SS), laser-induced fluorimetry (LIF), time-resolved laser-induced fluorimetry (TRLIF), and multivariate calibration methods combined with fluorimetry. The applications of these techniques to environmental PAH analysis are briefly discussed in the following.

3.2.2.1 Synchronous Fluorescence Spectrometry In conventional fluorescence spectroscopy, the emission spectrum is monitored by scanning the emission wavelength (λ_{em}), while the excitation wavelength (λ_{ex}) is fixed. The excitation spectrum is obtained by scanning λ_{ex} at a fixed λ_{em}. However, in SFS, λ_{ex} and λ_{em} are simultaneously varied while the wavelength interval $\Delta(\lambda_{em} - \lambda_{ex})$ is kept constant. Theoretically, when a small wavelength interval is selected, synchronous fluorescence excitation analysis provides only one spectral peak for each PAH compound in the analyzed sample. Therefore, the synchronous fluorescence technique is often regarded as spectroscopic chromatography.[70]

Several environmental applications of this powerful technique were suggested. Of special interest is the determination of several PAHs in spiked natural water samples, at concentration levels of a few ng ml. This rather demanding task was successfully accomplished by combining synchronous fluorescence and multivariate calibration methods.[71]

3.2.2.2 Full Excitation–Emission Fluorescence Spectrometry The main advantage of full EEM is that the two-dimensional (2-D) information obtained can be used for resolution of PAH mixtures. Usually, chemometric techniques are required for a proper data analysis, and quite often only separation of PAH families (rather than individual compounds) is possible.

Better results are obtained for simple PAH mixtures. The application of EEM for identification and quantification of PAHs in simple mixtures was accomplished in several laboratories.[72–76] In more complicated cases, such as analysis of samples containing 10 different PAHs, a combination of EEM and multivariate calibration methods was required for obtaining successful results.[77]

3.2.2.3 Shpol'skii Spectroscopy It is well known that broad-banded spectra are observed from compounds in condensed liquid and solid phases. This is a result of several effects, one of which is the inhomogeneous line broadening, i.e. each analyte molecule experiences a different influence from its surrounding solvent cage. This causes a slightly different Jablonski diagram for each analyte molecule. The overall spectrum is described as a Gaussian distribution of narrow lines with a total bandwidth of typically several hundreds per cm for each transition. The Shpol'skii effect is the spectral line narrowing. The reduction of the inhomogeneous line broadening can be accomplished by using an appropriate organic solvent at very low temperatures. The Shpols'kii effect is based on the geometric relationship between the organic solvent, i.e. the host molecules, and the PAH molecules, as shown in Figure 3.

It is believed that the narrow banded spectra are obtained by isolated molecules, trapped in the matrix during the cooling procedure and substituting one or more of the solvent molecules in the polycrystalline lattice.[78]

Ariese[79] demonstrated the power of SS by comparing the emission spectra of benzo[*k*]fluoranthene at room temperature (Figure 4a), spectrum and at 26 K (Figure 4b), spectrum. It is evident that the latter spectrum provides detailed information about the vibration levels of the electronic ground state.

The highly characteristic Shpol'skii spectra are suitable for qualitative and quantitative purposes in analytical and environmental chemistry. Actually many investigations showed that both conventional and laser excited SS can be successfully applied to the determination of PAHs in environmental samples.[80–83]

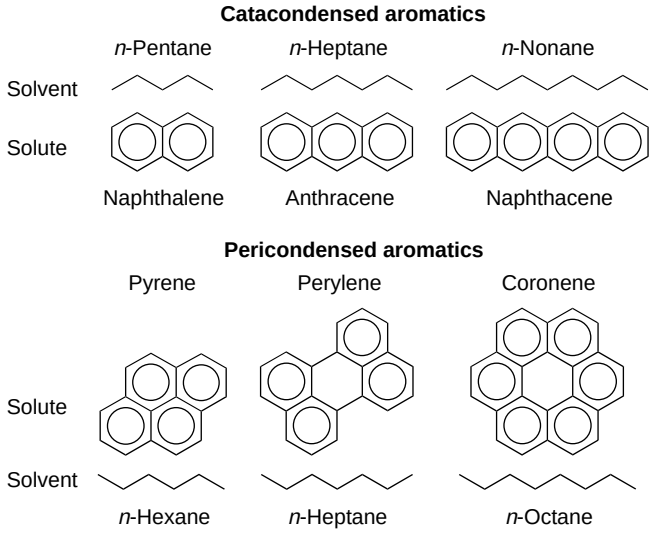

Figure 3 Geometric relationship of *n*-alkane and PAH molecules, which attempt to rationalize the Shpols'kii effect. (Reproduced from D'Silva, Fassel,[78] by permission of the American Chemical Society.)

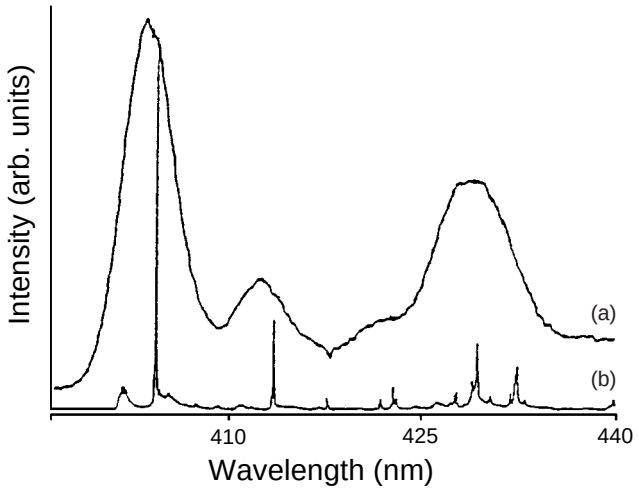

Figure 4 Emission spectra of benzo[*k*]fluoranthene in *n*-octane, using xenon lamp excitation at 308 nm. (a) Room temperature spectrum, concentration 10^{-4} M. (b) Shpol'skii spectrum at 26 K, concentration 10^{-6} M. Spectra were recorded using the same experimental setting; intensities are on the same scale. (Reproduced from Gooijer et al.,[83] by permission of Elsevier Science.)

3.2.3 Laser-induced Fluorimetry and Time-resolved Laser-induced Fluorimetry

Laser-induced fluorimetry (LIF) has been widely recognized as a standard technique for monitoring molecular species at very low concentrations. In this method a tuneable source of laser light of narrow spectral bandwidth is brought into resonance with an electronic transition to an exited level. The measured fluorescence signal originates from the spontaneous radiation decay. Among the most important advantages of the LIF technique in PAH analysis are its high sensitivity and its power in probing various environments, even hostile ones such as flames. Additional analytical advantages are attributed to the possibility of applying pulsed lasers to on-line time resolved measurements.[84] Optical fibers and various waveguides have largely contributed to LIF applications where remote sensing measurements are concerned. LIF techniques have already been coupled to many detection systems and analytical instruments such as mass spectroscopy, chromatography, plasma ionization and laser induced breakdown spectroscopy.[85–93] (These methods and their applications to environmental PAH analysis are briefly discussed here.)

An important factor that directly affects all on-line environmental analyses by spectroscopic techniques, is related to the homogeneity of the sample. Actually, all environmental samples are not homogeneous, although this fact is commonly ignored. For instance, environmental water is not homogeneous due to persistent hydrosols or microparticulates. Nevertheless, proper filtration and sample pretreatment are not possible in on-line applications; therefore analytical techniques that can be applied under such ambient conditions are of considerable importance. The effects of suspended materials on laser induced fluorescence of PAH molecules and its analytical performance, were recently investigated.[94] The particle's overall mass, light scattering and absorbing coefficients and particle size distributions were shown to effect fluorescence intensity of PAH in aqueous solutions. However, it was pointed out that a proper model can take into account all these effects and allow for reasonable PAH analysis.

On-line LIF and TRLIF were also evaluated for analysis of PAH contaminated aerosols.[16] The experimental set-up used for this purpose (shown schematically in Figure 5) consisted of a fast nitrogen laser for fluorescence excitation, a monochromator for spectral detection and an oscilloscope for temporal resolution. Remote analysis was carried out via quartz optical fibers. Detection limits of 100 ng pyrene per m^3 were achieved by using this set-up.

3.2.4 Phosphorescence Spectrometry

Several PAH molecules emit phosphorescence radiation and this characteristic can be utilized for analytical purposes. It is well known that the inherent phosphorescence quantum yield of PAHs can be enhanced by external heavy-atom perturbers. This method was originally developed for improving sensitivity and selectivity of low-temperature phosphorescence. However, phosphorescence enhancement by heavy atoms (e.g. thallium(I) acetate, lead(II) acetate) is also successfully used for room-temperature phosphorescence spectrometry (RTP).[70] Unlike the low-temperature phosphorescence techniques, sample preparation for the RTP measurements is very simple and does not require cryogenic equipment. Therefore, this method is readily applied to PAH analysis.

The most simple and common RTP method is solid-surface room-temperature phosphorescence spectrometry (SSRTP), which has become a well-established method for analysis of environmental samples containing PAH compounds. A solid support is often required for SSRTP analysis. Among the solid supports used are silica gel, sodium acetate, and cellulose. Cellulose, and particularly filter paper, is the most convenient, inexpensive, and common substance for SSRTP.

It was found that the sensitivity of the SSRTP method could be considerably improved by applying a laser for excitation.[95–98] Application of laser-induced SSRTP (nitrogen laser–pumped dye laser), with thallium(I) acetate as a phosphorescence enhancer, provided picogram limits of detection for several PAH compounds (phenanthrene, pyrene, benzo[*g*,*h*,*i*]perylene, chrysene, coronene, and 1,2-benzofluorene).[95]

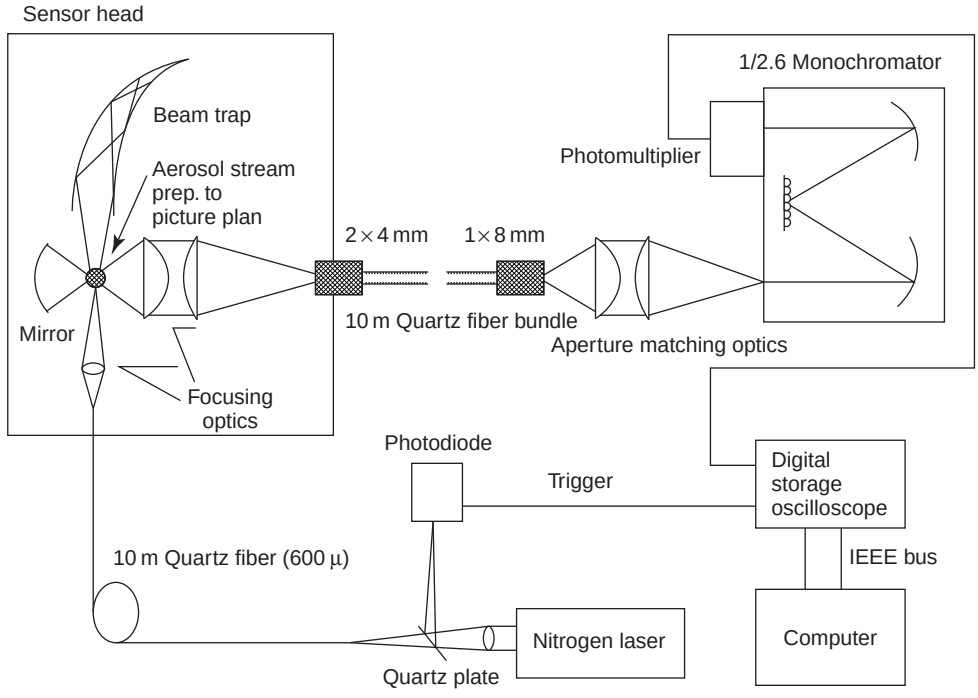

Figure 5 Experimental set-up for on-line aerosol fluorescence detection of particle-bound PAHs. (Reproduced from Niessner,[16] by permission of Elsevier Science.)

The coupling of laser-induced RTP to fiber optic sensors may potentially be useful for field applications where fast determination of PAHs is needed.[96,97] Limits of detection at the order of picograms were obtained for several PAHs, using a SSRTP fiber-optic sensor. Figure 6 shows typical emission spectra obtained by a laser-induced RTP sensor.

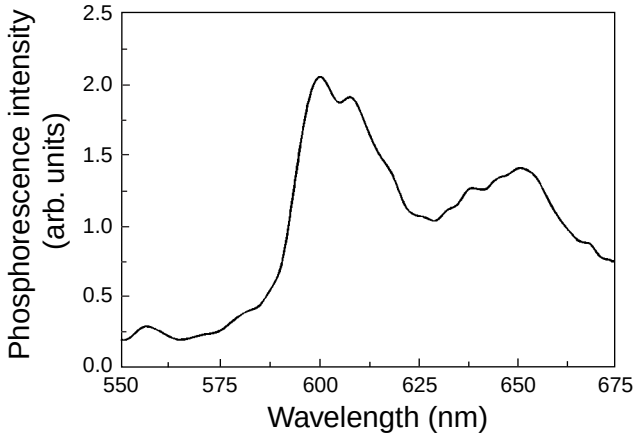

Figure 6 Laser induced RTP emission spectrum of a 10^{-6} M pyrene solution obtained with a fiber-optic sensor. Phosphorescence emission was enhanced by 0.1 M TlOAc. Excitation wavelength was 337 nm. (Reproduced from Campiglia et al.,[97] by permission of the American Chemical Society.)

3.2.5 Infrared Spectrometry

Conventional dispersive infrared (IR) spectrometry is not a preferred method for PAH analysis. The disadvantages of this technique for PAH analysis are the lack of unique and significant structures that exhibit IR spectra, the relatively weak bands, and the fact that the IR band strength is not proportional to the concentration of the analyte. Moreover, as far as environmental samples are concerned, they usually contain numerous PAHs which all exhibit the typical C–H stretching in the region 3200–3000 cm^{-1} and the out of plane bending vibrations at 900–675 cm^{-1}.[99] Other IR absorption bands may be contributed by heteroatoms in the ring structure or by substituent groups on the ring.[100] Thus, overlapping of spectral features due to various PAH compounds in a sample is very likely.

The development of FTIR spectroscopy along with the application of matrix isolation (MI) techniques has made IR spectroscopy more useful for PAH analysis. In particular, FTIR can follow the different stages of chromatographic separation and provide successful detection of PAH compounds.[100]

In the MI procedure the compounds of interest are isolated in a rare-gas, IR transparent, matrix at low temperature (~10 K). Since interactions between the studied compound and the matrix are very weak, the molecules are trapped in a quasi-gas-phase environment. Rotations of the PAH molecules are reduced at the

low temperature, and the IR bands are narrowed, thus reducing the possibility of overlapping.[101]

The positions of the absorption bands of matrix-isolated PAH molecules typically undergo only small shifts in the 0–15 cm^{-1} range, relative to their gas-phase values.[102,103] IR band positions of matrix-isolated PAHs were reported for a large number of PAHs.[104]

Several successful applications of MI FTIR spectroscopy in PAH analysis have been reported recently. Matrix-isolated techniques have been applied to measure the mid-IR spectra of several PAHs (naphthalene, anthracene, phenanthrene, 1,2-benzoanthracene, chrysene, pyrene, tetracene, and triphenylene).[101] The results showed a good agreement with theoretical calculations for the majority of IR-active vibrations of the examined PAHs. In addition, the analysis of 12 nitro-PAHs was accomplished recently.[104]

The use of FTIR spectrometry to detect PAH compounds separated by GC analysis has been suggested and shown to provide complementary information to mass spectrometric data. Naturally, mass spectrometry cannot provide sufficient information for distinguishing between PAH isomers, whereas their IR spectra are significantly different. The GC/FTIR technique is carried out by passing the GC effluent through a gold-coated light pipe coupled to IR-transparent windows, while interferograms are continuously obtained. The interferograms of the various components undergo Fourier transformation, resulting in analysis of the PAH mixture.[105]

Another GC/FTIR detection procedure is based on the MI technique, in which the effluent compound is trapped in a frozen matrix of an IR transparent gas on a rotating gold-plated disk. Interferograms are obtained after the GC run is completed. The detection limit for this method for one compound, pyrene, was 710 ng.[106]

It is generally concluded that the applications of IR spectroscopy to PAH analysis provide complementary information that may not be obtained using other methods. Nevertheless, the IR-related techniques provide relatively poor sensitivity and selectivity in analysis of PAH compounds. Therefore, these techniques are recommended when the unique IR information (e.g. separation of isomers) is required.

3.2.6 Raman Spectroscopy

Raman spectroscopy is the measurement of the frequencies and the intensity of the light inelastically scattered from a molecule. The range of frequencies shifted from the incident light (Stokes and anti-Stokes lines) comprises the Raman spectrum of a molecule. The Raman shifts correspond to vibrational transitions in the scattering molecule.

Even though Raman scattering frequencies provide information about the vibrational transitions of the analyte molecule, as the IR absorption dose, the information obtained by these two methods is not identical. Frequencies permitted in the IR may be forbidden in Raman, and vice versa.

Most commonly, the Raman excitation is performed using a UV/VIS light source, hence the Raman bands are also observed around this region. Detection in this spectral range is easy, because many detecting devices are available, e.g. photomultiplier tubes, diode arrays, and CCDs. The ideal light source for Raman spectroscopy is a laser.[107] The most frequently used lasers in Raman spectroscopy are the Ar or Kr-ion types. Sometimes these lasers pump CW dye lasers, providing tuneable radiation. High repetition rate excimer laser-pumped pulsed-dye lasers are also applied. When a UV/VIS excitation source is used for PAH analysis, a broad fluorescence band is also expected. The fluorescence interferes with the Raman scattering; thus, the former has to be subtracted from the spectra, if possible. This fluorescence interference can be avoided in the analysis of most PAHs when an NIR laser is applied for excitation (e.g. CW neodymium : yttrium aluminum garnet (Nd : YAG) laser, irradiating at 1064 nm). The main drawbacks of using this spectral range are the loss of sensitivity due to the λ^{-4} dependence of the scattering event and the decreasing quantum efficiency of the CCD detectors.[108]

The major disadvantage of Raman spectroscopy is its low quantum yield: only 1 out of 10^7 photons of the incident light exhibits the Raman effect. This results in very weak signals. Therefore, spectral analysis based on Raman scattering requires high light throughput spectrophotometers, such as the Fourier transform type. Raman spectra free of fluorescence interference are commonly obtained by a FTIR system, which combines a Fourier transform interferometer and a Nd : YAG laser.[109] Near-IR Fourier transform Raman spectra of several PAHs have been measured. These spectra exhibited distinct features, which enabled identification and analysis of multicomponent mixtures of PAHs.[110]

Resonance excitation may raise the scattering power of PAHs by several orders of magnitude, allowing detection of low PAH levels. The resonance Raman intensities strongly depend on the excitation wavelength. Thus, by using a tuneable laser source one can selectively excite individual PAH compounds in a mixture. Resonance Raman spectroscopy using pulsed UV lasers is considered the most selective application of Raman spectroscopy in PAH analysis.[107] Naphthalene, anthracene, phenanthrene, pyrene, and various substituted derivatives were examined by pulsed UV laser and exhibited resonance Raman bands.[111] The UV resonance Raman enhancement enabled trace analysis of these species down to the ppb level.

This technique has also been applied as a detection system for HPLC. A mixture of five PAHs (naphthalene, fluorene, anthracene, pyrene, and chrysene), each compound at a concentration of 10^{-3} M, was separated by HPLC. The UV resonance Raman spectra (λ_{ex} 230 nm) of each compound supplied enough information for their identification.[112]

Recently, the Raman technique has been applied to remote environmental sensing. Dielectric stack interference filters were incorporated within an UV optical fiber Raman probe head. This set-up enabled the removal of background silica Raman signals and Rayleigh scattering. Using a 244 nm UV optical fiber probe, resonance Raman spectra from 2 ppm (Figure 7a), 200 ppb (Figure 7b), and 20 ppb (Figure 7c) of pyrene in water were collected. The 1632 cm^{-1} pyrene band is clearly observed in all spectra. In addition, equimolar aqueous solutions of pyrene and a visible fluorophore (R6G) were recorded (Figure 7d). This spectrum demonstrates the selectivity and sensitivity of this method. The R6G bands (Figure 7e) are not seen in the Raman spectra of pyrene and R6G. This technique was applied for in situ, on line monitoring of low PAH concentrations.[113]

Since Raman spectroscopy is based on scattering of light, non-transparent samples can be examined. This is an advantage over other spectroscopic methods, which fail in analysis of solid environmental samples.

3.3 Photoelectric Aerosol Sensor

A sensor is usually a device that transduces a physical or chemical parameter into an electrical or optical device. Sensing techniques are expected to provide both high sensitivity and in situ, real-time information.[114] Photoelectric aerosol sensing was studied using artificially PAH-coated carbon and sodium chloride aerosols. Aerosols were irradiated at 185 nm and photoelectric signals were measured. The signals obtained were linearly correlated to the amounts of each single PAH bound to the particle.[115,116] It was found that the total photoelectric signal of internally mixed aerosols is described as a simple addition of the individual contributions of the adsorbed PAHs.[117] The photoelectric aerosol sensor (PAS) was also applied to detection of particle-bound PAHs in diesel exhaust.[16] In this case, the sensing system included a dilution unit and a measuring device. The system allowed dilution of exhaust aerosols in clean air at ratios of up to 1 : 32. Within the measuring device the neutral aerosols were separated from the charged particles and then introduced into the illumination section, were PAH mass-related particle photocharging took place. (The irradiation wavelength was 185 nm.) Under these conditions, only surface-enriched molecules possessing an extended π-electron system (e.g. PAH molecules) exhibit high photoelectric quantum yield. The PAS

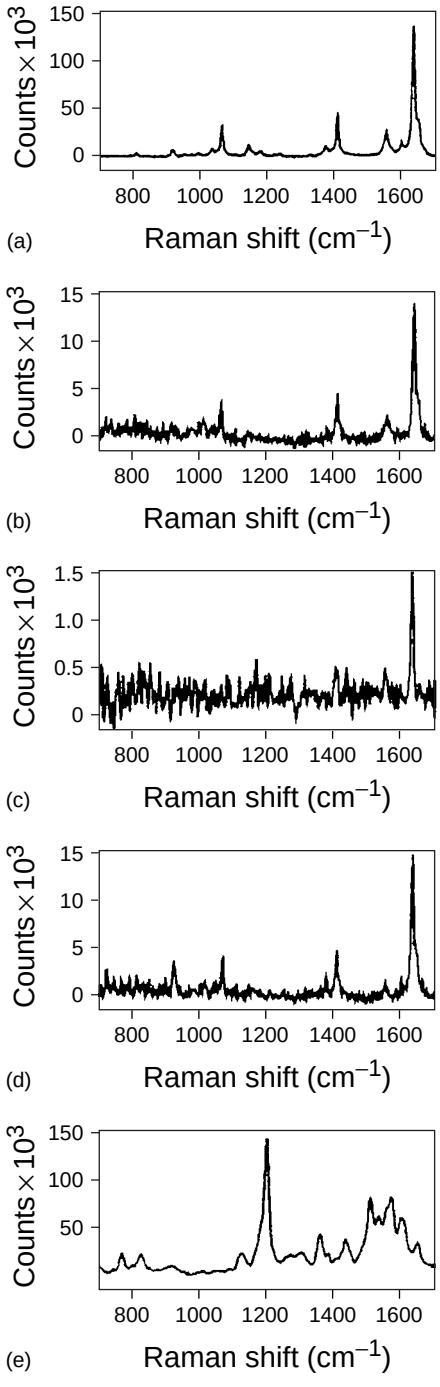

Figure 7 UV Raman spectrum from: (a) 10 μM (2 ppm); (b) 1 μM (200 ppb); and (c) 100 nM (20 ppb) pyrene in water; (d) 1 μM pyrene in a 1 μM aqueous solution of R6G; and (e) 100 μM R6G. Spectra were obtained by using a UV optical fiber probe dipped directly into the analyte solutions. Excitation conditions: (a–d) 244 nm, 3 mW, 200 s accumulation time; (e) 244 nm, 7.5 mW, 300 s. (Reproduced from Munro et al.,[113] by permission of the Society for Applied Spectroscopy.)

provided a continuous sensitivity ($<$ng m^{-3}) of four and more ring-membered PAHs, but the photoelectric signal represented the sum of individual signals of all adsorbed PAHs.

3.4 Mass Spectrometry

Mass spectrometry is a well established method for PAH analysis. Various techniques for ion generation, separation and detection are used in MSs. In general, MSs rely on differences in the mass-to-charge ratio (m/z) of ionized PAH molecules for separation, identification, and quantification. The basic operating procedures of analytical mass spectrometry include: creating gas phase ions, separating the ions in space or time according to their (m/z) ratio, and finally identification and quantification. Numerous mass spectrometric analytical methods are known, but this review will focus on two recent techniques, which are of special relevance to environmental detection of PAH compounds: laser desorption/laser ionization (two-step laser) time-of-flight (L2TOF) mass spectrometry and multiphoton ionization (MPI) time-of-flight mass spectrometry.

3.4.1 Laser Desorption/Laser Ionization Time-of-flight Mass Spectrometry

L2TOF mass spectrometry is used for direct analysis of PAH compounds. In the first step an IR laser pulse desorbs the analyte molecules from the sample surface, or ablates neutral molecules from a solid sample or from a polymer membrane containing the analyte. No fragmentation or ionization occurs at that step. In the second step, a pulse from a tuneable ultraviolet laser is used for resonance-enhanced two-photon ionization of the desorbed material. This procedure results in the formation of molecular ions. This mechanism is suitable for PAH molecules which are known to absorb the UV photons efficiently. Nonresonant two-photon ionization may also occur under these conditions, but with several orders of magnitude lower probability. Fragmentation is avoided by tuning the laser intensity and thus preventing absorption of another photon. Finally, ions are extracted by an electrical field and separated by their m/z ratio using time-of-flight mass spectrometry.[118]

A laser desorption/laser ionization mass spectrum of a PVC membrane with eight PAH compounds, is shown in Figure 8. Only the intact parent ions are produced and detected.

The L2TOF mass spectra of 17 PAH compounds were recorded, when excitation was performed in the range 238–310 nm. It was proven that efficient ionization depends upon the existence of an intense absorption band. In addition, a long lifetime of the S_1 state is essential for efficient resonance enhanced photoionization.[118]

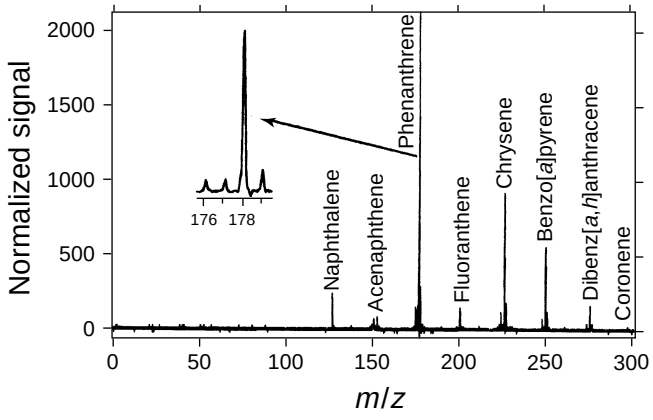

Figure 8 Typical L2MS spectrum. PVC membrane with eight PAHs. About 100 fmol of each PAH is ablated by a single laser shot. Ionization at 250 nm. Mass resolution: routinely about 1200. No low-mass fragments are observed. (Reproduced from Haefliger and Zenobi,[118] by permission of the American Chemical Society.)

Environmental samples contain numerous PAHs. Among them are volatile PAH compounds, with vapor pressures between 10 Pa for naphthalene, to 7×10^{-4} Pa for pyrene. Mass spectrometry of solid environmental samples suffers from losses of the volatile material, e.g. volatile PAHs, due to the high vacuum conditions required. A possible solution to this problem is lowering the temperature of the sample, thus reducing the analyte's vapor pressure. Another approach is forming a non-volatile PAH complex which dissociates during desorption or ionization. Complexation of PAHs with CDs, which has been mentioned in section 2.1, is also a useful technique for this purpose.[13] The formation of solid-phase CT complexes of PAH–picric acid may also reduce PAH vapor pressure and thus reduce the loss of volatile PAHs such as anthracene during L2TOF/MS.[119]

The spatial resolution capability of the laser desorption, combined with the high sensitivity of L2TOF/MS, enables the detection of microscopic PAH samples.[120–126] Spatially resolved L2TOF/MS was applied to the detection and analysis of PAHs in atmospheric particles.[126–129] Recent studies demonstrated that single-shot analysis of PAHs could be performed with micrometer-sized samples.[127] PAHs in a single micrometer-sized diesel particle (from NIST SRM sample of diesel particulate matter, SRM 1640) were detected by L2TOF/MS. The limit of detection was estimated as lower than 20 μg g^{-1} diesel soot. In addition, other PAHs, which were not certified by NIST, were also detected.[130]

Two-step laser time-of-flight mass spectrometry is a selective and sensitive technique, which utilizes soft ionization conditions for providing parent-dominated ion spectra. This method requires very little sample preparation. Furthermore, measurements can be performed

within a few minutes. In addition, spatial resolution can be obtained by using a focused laser beam for desorption (resolution of 1–2 μm).[130] These characteristics make L2TOF/MS an efficient technique for analyzing complex environmental samples. However, this method suffers of two main problems: it cannot distinguish between isomers, and quantification is difficult.[118] Although the second problem has been addressed recently,[131] the difficulty still remains, especially for complex environmental samples.

3.4.2 Multiphoton Ionization Time-of-flight Mass Spectrometry

When a molecule interacts with electromagnetic radiation it can absorb energy and become ionized if the photon energy is larger than its ionization energy. However, under certain conditions, a molecule can also be ionized when the photon energy is lower than its ionization energy. This nonlinear process is called multiphoton ionization/time-of-flight mass spectrometry (MPI/TOFMS) and it involves simultaneous absorption of several photons. The mechanism of MPI requires intensive radiation ($\sim 10^{30}$ photons $cm^{-2}\,s^{-1}$) and short pulses, such as those readily obtained by lasers.

There are two classes of MPI processes: nonresonant MPI and resonance-enhanced multiphoton ionization (REMPI). The former process can be described as absorption of n photons through short-lived ($\iota \sim 10^{-15}$ s) virtual states (Figure 9a).

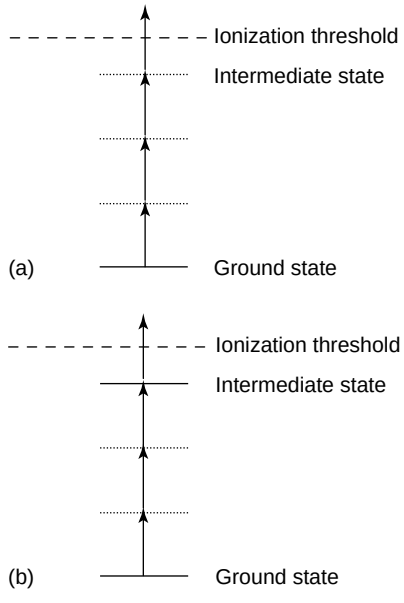

Figure 9 MPI: (a) direct n-photon ionization; (b) n-photon ionization through m-photon resonant transition to an intermediate state.

In REMPI the energy of m photons ($m < n$) approaches the energy of a real intermediate state (Figure 9b). If the m-photon transition is allowed, the cross-section for n-photon ionization increases significantly.[132]

In the case of naphthalene, for example, the first excited singlet state lies at an energy of 4.15 eV relative to the ground state. This energy gap corresponds to a wavelength of 299 nm. A laser that operates at this wavelength can reach the first PAH excited state with one photon. Another photon of this energy will ionize the molecule because naphthalene's first ionization potential is 8.12 eV. In general, PAHs are particularly suitable for MPI because their absorption of light in the region 250–350 nm is very efficient.[133]

Techniques based on MPI can be roughly divided into two main categories: MPI conductivity measurements, in which the photocurrent (electrons emitted in the MPI process) is detected; and MPI/MS measurements, in which the mass of the molecular ions is detected. The former will be discussed in section 4.1. The most common technique that detects the mass of ions produced by MPI processes is the MPI/TOFMS.

MPI/TOFMS is based on the differences in flight-time of equally accelerated PAH ions. Following the MPI step, an electric field accelerates all ions into a field-free drift region, where they obtain a kinetic energy of qV (q is the ion charge and V is the applied voltage). Since the kinetic energy of the ion is $0.5\,mv^2$, lighter ions reach higher velocities and therefore reach the detector after a shorter period of time.[131]

Several applications of MPI or REMPI[134] in MS of airborne particulate matter have been reported.[135–139] On-line flame-sampling resonance-enhanced multiphoton ionization time-of-flight mass spectrometry (REMPI/TOFMS) has been performed in order to determine the concentrations of naphthalene, fluorene, and anthracene in a pure methane + oxygen–argon (1 : 5) diffusion flame.[138] Detection limits as low as ppb were reported using a one-color (301.6 nm) or two-color (361.1 nm, excitation beam and 308 nm, ionization beam) REMPI on jet-cooled samples extracted from the flame. Figure 10 shows a REMPI/TOFMS mass spectrum of a jet-cooled sample from a heavily sooting flame.[138]

Recently, a mobile laser MS (REMPI/TOFMS) was applied to on-line measurements of PAHs in a waste incineration pilot plant.[140] The mobile device was adapted to harsh conditions of heat, dust and vibrations. PAHs were ionized selectively from the gas matrix by using UV laser irradiation (Nd : YAG at 266 nm or excimer KrF at 248 nm). Detection limits of naphthalene, for instance, were in the order of 10 ppt.

Unfortunately, despite relatively high material selectivity and spectral resolution offered by various MS techniques, they are limited, at times, by the quite severe

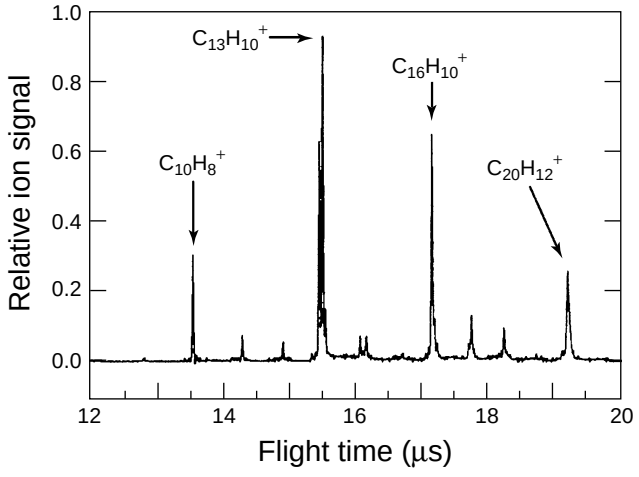

Figure 10 One-color laser ionization time-of-flight-mass spectrum (301.6 nm, 100 shots averaged) of a jet-cooled sample from the luminous zone of a heavily sooting $CH_4/O_2/Ar$ diffusion flame. Peaks in the spectrum do not reflect the relative abundance of each mass in the sample. (Reproduced from Gittins et al.,[138] by permission of the American Chemical Society.)

vacuum conditions needed for most of the accurate time-of-flight measurement schemes.

3.5 Immunological Methods for the Detection and Quantification of Polycyclic Aromatic Hydrocarbons

Immunological methods are based on the high affinity reaction of an antibody–antigen complex. These types of reactions, which are characterized by specificity and extremely high sensitivity (femtomole levels) are used for PAH analysis. The two major immunoassays are radioimmunoassay (RIA)[141] and enzyme-linked immunosorbent assay (ELISA).[142]

Antibodies (immunoglobulins) are proteins produced by an animal in response to the presence of a foreign substance, called an antigen. Low molecular weight organic compounds (MW < 5000), such as PAHs, are not immunogenic by themselves. These compounds, called haptens, must be coupled to a carrier protein (e.g. BSA, bovine serum albumin) in order to induce an immunogenic response.[143] Either polyclonal or monoclonal antibodies can be produced. Polyclonal antibodies are produced by immunizing an animal and collecting its serum, which contains the antibodies. These antibodies are called polyclonal because they are produced by many different antibody-producing cells, and therefore they differ to some extent in their affinity for the antigen. In contrast to polyclonal antibodies, monoclonal antibodies are identical because they are produced by a population of identical cells (a clone). Each of these cells is descended from a single hybridoma cell formed by fusing an antibody-producing cell with a tumor cell, which is characterized by its capacity for unlimited proliferation.[144]

Almost all antibody–antigen or antibody–hapten reactions can be modified for use as competition assays. Such modifications have been developed for measuring cross-reactivity among different haptens or antigens for the same antibody preparation, or quantitatively measuring the concentration of the same hapten or antigen in unknown samples.[142]

The similar shapes, hydrophobicity, lack of hydrogen-bonding atoms, and other characteristics of PAHs contribute to the difficulties in the preparation of antibodies, which are highly selective for individual PAH compounds.[145] Therefore, cross-reactivity is expected, especially when analyzing environmental samples. Selectivity can be improved by separating the compounds (e.g. by chromatography) prior to quantification by immunoassay.[143]

3.5.1 Radioimmunoassay

In RIA the antigen or hapten is radiolabeled and serves as a tracer, while the nonradiolabeled antigen or hapten serves as an inhibitor. RIA involves competition between labeled and unlabeled antigen (or hapten) for binding to antibodies. The basic principle of RIA is that the likelihood of a molecule, labeled or unlabeled antigen, of binding the antibody is proportional to its concentration in the original mixture. Therefore, an isotope-dilution standard curve can be constructed. This is carried out by addition of ever-increasing known amounts of unlabeled antigen (inhibitor), while the concentration of labeled antigen (tracer) is kept constant. Then, the bound and free antigens (or haptens need to be separated). Separation can be accomplished by precipitating the antibody–antigen complex (e.g. by applying a second antibody). Finally, concentrations are determined from the radioactivity readings. After construction of the standard curves, an unknown antigen can be quantified by measuring the inhibition of the tracer binding.[142,143] Although RIA is not a common analytical technique, several applications to the analysis of biological PAHs have been proposed during the last few years.[146–148]

3.5.2 Enzyme-linked Immunosorbent Assay

The ELISA is the most commonly used immunoassay technique for identification and quantification of PAHs. Figure 11 illustrates the general steps of the ELISA procedure.

In the first step the hapten (PAH molecule)–carrier protein conjugates are adsorbed onto the surface of the reaction vessel (a 96-microwell plate). A solution of the examined antibodies and antigen standards (or

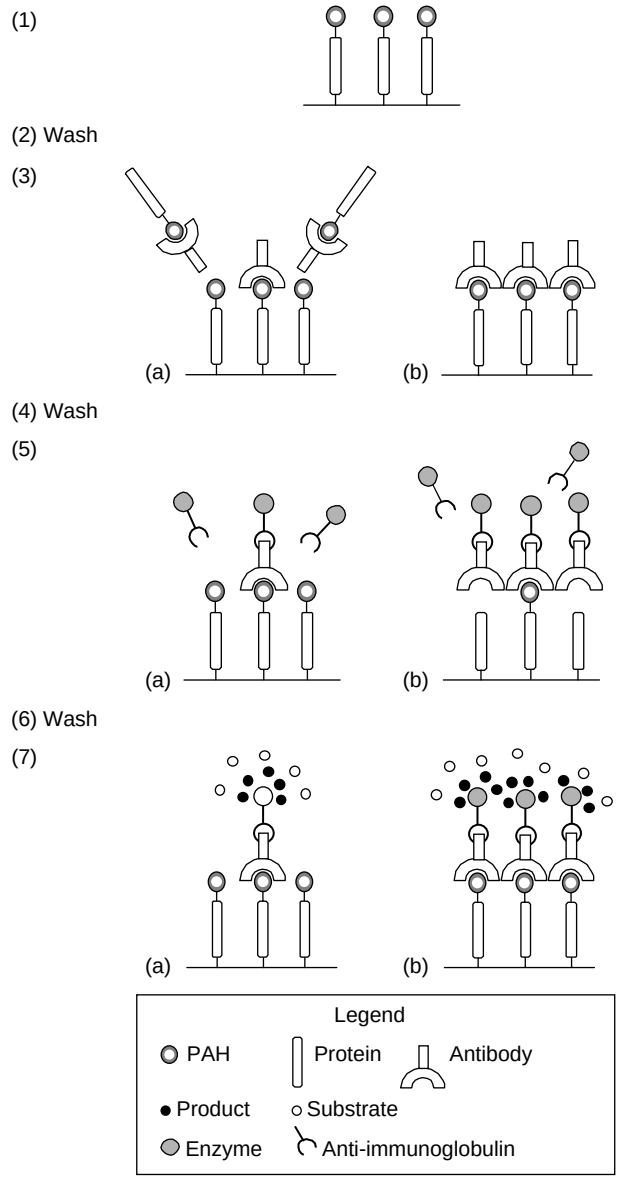

Figure 11 Competitive ELISA: (1) Conjugates of PAH molecule-carrier protein are adsorbed onto the well; (2) wash; (3) incubation with specific antibodies in the presence (a) or absence (b) of standard or sample conjugates of PAH molecule–carrier protein; (4) wash; (5) incubation with anti-immunoglobulin enzyme; (6) wash; (7) incubation with enzyme substrate and measurement of product.

hapten–protein complexes) is then added. The binding of the antibodies to the immobilized haptens is competitively reduced due to the antigen addition. Wells are thoroughly washed so that only antibodies bound to immobilized haptens remain in the vessel. A second reagent is then added. This reagent is a conjugate of anti-immunoglobulin (an antibody originally produced from another animal, which binds the constant region of all antibodies) and an enzyme. Finally, after washing again, a colorless substrate is added and color develops as the enzymatic reaction progresses. The enzymatic product concentration is inversely proportional to the concentration of the added standard antigen. Quantification is performed by spectrophotometric measurements of the color intensity.

ELISA methods are generally much more sensitive than RIA, and therefore more popular for PAH analysis. However, as previously mentioned, the production of antibodies, which are highly selective for a specific PAH compound, is difficult. This is due to the similar characteristics of most PAH compounds. Cross-reactivity is expected, especially when environmental samples are considered.

ELISA techniques have also been applied to PAH analysis in air particulate matter, which has been analyzed by using a specific enzyme immunoassay for [benzo[*a*]pyrene].[149,150] In this case, however, the results were not compared with a conventional method. Recently, ELISA was applied to the analysis of 1-nitropyrene on diesel exhaust particles.[150,151] This compound is one of the numerous products of incomplete combustion. It was chosen as a target analyte because of its high mutagenic and carcinogenic properties and because of its common appearance mainly on exhaust particles, emitted from diesel engines. It was shown that the detection of 1-nitropyrene, using the ELISA method, was applicable for the analysis of a large number of samples, in a short time. However, there were always substances present in the atmospheric samples, which could lead to overestimation of the target analyte due to their cross-activity in the ELISA.[151] Direct application of the ELISA method resulted in a 6-fold overestimation. However, after a sample clean up, results were lowered to ~1.6-fold overestimation, which was attributed to the cross-reaction of 2-nitropyrene and 2-nitrofluoranthene.[150] Therefore, it is concluded that absolute quantification of a single compound by the ELISA method is limited. Quite often, other substances present in the sample can cause an over-estimation of the target analyte due to their ELISA cross-reactivity.[150,151]

A new approach of hapten screening was recently presented. The purpose of this research was to step towards developing sensitive ELISA for multianalyte PAH samples. Several new PAH haptens were applied for a systematic study of cross-reactivity with monoclonal antibodies. PAH haptens with different lengths of carboxylic acid chains, placed at various positions, were synthesized, and coupled with proteins. Several PAHs were used, e.g. naphthalene, fluorene, anthracene, pyrene, benzo[*a*]pyrene. The goal was to maximize the difference in PAH cross-reactivity in the ELISA technique. As a result, the benzo[*a*]pyrene detection limit

obtained by ELISA was ca. 150-fold lower than that obtained by GC.[145]

Compact kits for on-site PAH detection and screening, applying the ELISA technology, were developed and are now commercially available.[152,153] Such a test kit for screening PAHs in soils has been validated by the USEPA (method number 4035).[154]

3.5.3 Immunoaffinity Extraction and Chromatography

Selective isolation, extraction and clean-up of PAHs from complex environmental samples (soil, water and waste) can be accomplished using antigen–antibody interactions.[155–157] These methods are based on extraction of PAHs from samples onto an immunosorbent. The benefits of immunosorbent extraction techniques are the simplicity, the reduced amount of hazardous organic solvents and the relatively low cost.

Coupling of immunosorbent SPE with liquid chromatography analysis was successful in the determination and quantification of PAHs in environmental samples, down to the ppb level.[155,156] The unavoidable cross-reactivity of specific antibodies (anti-pyrene for instance) with a number of PAHs, can be advantageous for extraction and clean up applications. Thus, antibodies that possess an extended cross-reactivity are required for preparation of class-specific immunosorbents.

Immunoaffinity chromatography (IAC) utilizes the interactions between an immobilized antibody and the analyte (antigen or hapten) in order to selectively extract the analyte from a complex sample matrix (e.g. environmental samples).[158,159] Immunoaffinity columns are packed with support material, e.g. silica, which contains immobilized antibodies. The latter are either covalently bonded or physically adsorbed to the support material.

Unfortunately, both of these immobilization techniques suffer from some disadvantages: the physical adsorption of antibodies may result in unstable columns, while the covalent binding may impose changes in antibody conformation. A partial solution for these problems has been suggested. It is based on physical trapping of the antibodies in the pores of a silicate glass prepared by sol–gel techniques.[158,159] The mild conditions under which the column is prepared do not destroy the molecular recognition.

Several PAH antibodies have been applied in IAC. In particular, anti-pyrene[159] and anti-1-nitro-pyrene[158] have been studied in detail. It is commonly concluded that the main drawback of the IAC technique is that the selectivity of these type of columns is limited due to the cross-reactivity of the antibodies with other PAH compounds.

4 RECENT ADVANCES IN THE ANALYSIS OF ENVIRONMENTALLY OCCURRING POLYCYCLIC AROMATIC HYDROCARBONS

In this section several new analytical techniques which have considerable potential in future environmental PAH analysis are discussed. The feasibility of these techniques has already been demonstrated, but, they are not yet available commercially. Further research and development is required for final validation and general acceptance of these methods.

4.1 Laser-induced Multiphoton Ionization/Fast Conductivity

The principle of laser-induced MPI has been described in section 3.4.2. It was mentioned that MPI-based techniques are roughly divided into two main categories: MPI/MS (previously described) and MPI conductivity measurements, in which the photocurrent (electrons emitted in the MPI process) is detected. The latter is described in this section.

A photoionization spectrum, namely the photocurrent vs the laser excitation wavelength, λ_{exc}, can be obtained using the conductivity MPI technique. These spectra possess peaks in the vicinity of the resonant wavelengths. In resonant MPI conductivity measurements, changing λ_{exc} results in photocurrent signals that correspond to excitation of intermediate resonance. The MPI spectrum is, therefore, a fingerprint of the analyte molecule.[132]

A block diagram of a typical multiphoton ionization fast-conductivity (MPI/FC) facility is shown in Figure 12.

Laser two-photon MPI conductivity measurement is a selective and sensitive method for the detection of photoabsorbing molecules such as various PAH compounds. This technique has been applied in both polar

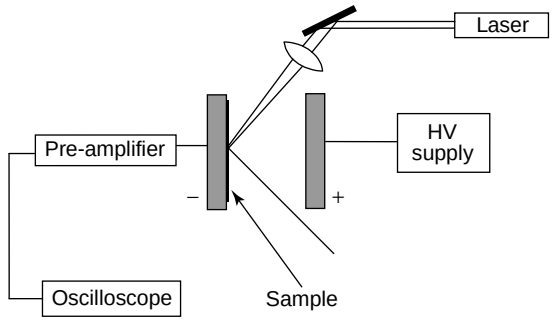

Figure 12 A block-diagram of an MPI/FC facility. Laser irradiation is used for MPI of PAH traces, positioned on a conducting substrate. The high voltage electrode enables spatial charge separation and transport. Readouts are then recorded and analyzed using a preamplifier.

and nonpolar solvents,[160–163] on metal surfaces,[164,165] on a water surface,[166–171] and on an interface region in a two-phase system.[172]

MPI conductivity measurements are of considerable environmental value, due to their potential for operating under ambient conditions. Therefore, this technique allows for in situ, on line measurements of PAHs in environmental samples without pretreatment and extraction procedures. Quite obviously, measurements of MPI photocurrents require the substrate to be conductive. However, most environmental matrixes are insulators. Therefore, for many years it was believed that this technique could not be applied to environmental PAH analysis. A recent study showed that MPI measurement, under certain conditions, can be performed on insulators, provided that a certain humidity is present. The effects due to humidity were quantified and it was found that the natural humidity of environmental samples, due to their exposure to ambient air, is adequate for allowing MPI analysis. For example, natural soil moisture is sufficient for trace analysis of organic contamination, e.g. PAHs, by fast-conductance MPI.[173,174] Clearly, this finding opened up the possibility of applying similar experimental schemes to any porous substrate of interest, e.g. glass/paper filters, plant leaves (open field vegetation), and certain food products. Preliminary field tests were also conducted and reported. In these applications the data acquisition facilities were of a reasonably low cost, as well as being user-friendly and simple to operate.[175]

Recently, the MPI/FC technique has been applied to the detection of PAH-polluted aerosols, i.e. motor car exhaust emissions and cigarette smoke. Aerosols were sampled on line by means of renewable water microdroplets. As shown in Figure 13, on-line PAH analysis is feasible. Detection limits as low as 1 pg were obtained for pyrene-contaminated renewable microdroplets.[17]

Clearly, the liquid droplet technique offers an easily engineered renewable sampling of gaseous PAH pollutants and PAH contaminated aerosols. Some poorly dissolving airborne PAH compounds could readily reach the aqueous solubility limit and precipitate on the surface of the sampling water microdroplet. However, the same MPI/FC technique can be applied for measuring low aqueous solubility of PAHs and characterization of their adsorption, without any sample preparation.[176]

4.2 Multiphoton Ionization–Time-resolved Ion-induced Charges

A new, simple, detection system, based on time resolution of ion-induced charges in MPI of PAH compounds in air has been suggested.[177,178] The system is suitable for on-line air monitoring and has been tested for the detection of simple PAH molecules.

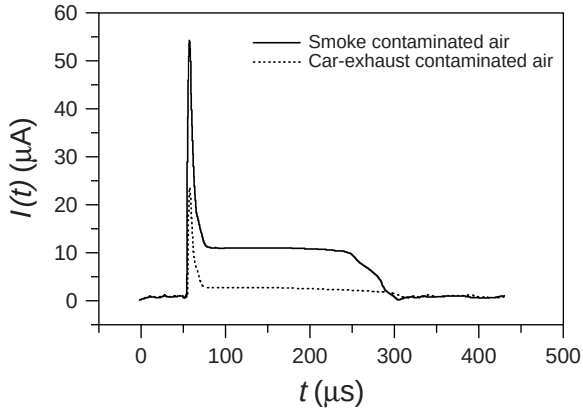

Figure 13 Photoionization current, $I = I(t)$, obtained by the MPI/FC technique, applied to a renewable water droplet. The droplet was contaminated with motor exhaust gases and cigarette smoke. Note that the area under the respective $I(t)$ curves corresponds to the detected photocharge, Q, which, in turn, is proportional to the amount of photoabsorbing organic traces. Excitation was carried out by a pulsed nitrogen laser at 337.1 nm. (Reproduced from Gridin et al.,[17] by permission of the American Chemical Society.)

Two-photon quasi-resonant ionization is carried out by a few mW cm^{-2} of a KrF excimer laser radiation at 248 nm. Under these conditions, none of the regular atmospheric gas is ionized. The quasi-resonant ionization of PAHs at 248 nm provides the required selectivity.

This technique provides a certain level of mass analysis, using the following set-up. The examined air is introduced into a long cylindrical chamber, and its pressure is reduced until the mean free path reaches an order of 1 cm. A parallel laser beam passes into this chamber, resulting in MPI of the PAHs. Two long electrodes are placed parallel to the laser beam. The MPI charges are separated and slightly accelerated towards the electrodes. Quantification of the PAHs is based on the number of charges, while mass detection is based on the time-resolved electrode voltage change during the ions' flight. The latter is a result of the mirror-charges induced by the PAH ions, and it was shown that their temporal profile is a clear function of the masses. This new detection system is characterized by its simplicity, large ionization volume, and high ion collection efficiency.

A set of computer algorithms was required in order to analyze the data and to ascertain the ionic mass. Detection limits in the order of ng m^{-3} in the detection chamber were reported, which corresponds to ambient air concentrations in the sub-ppm range.

4.3 Imaging Methods

Chemical imaging, which combines spectral and spatial resolution, is a relatively new technique in analytical

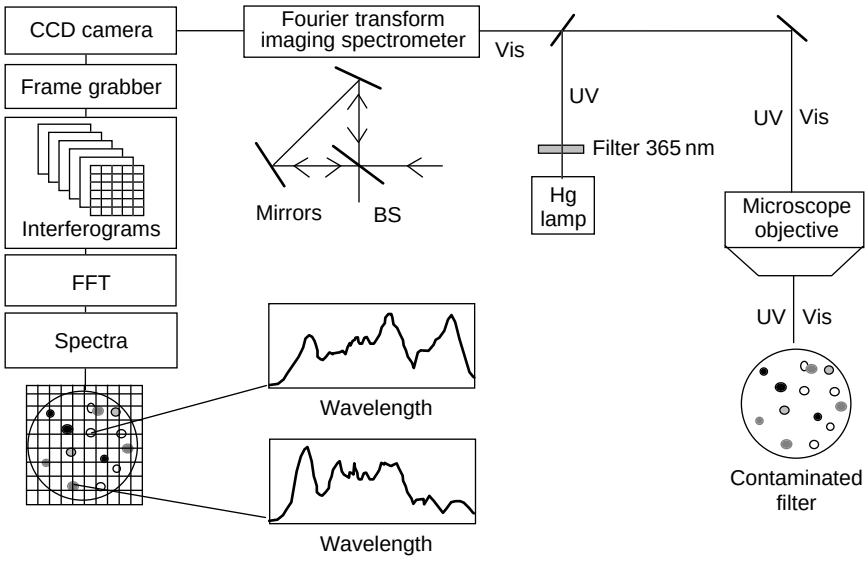

Figure 14 Schematic experimental set-up for microscopic Fourier transform spectral imaging. (Reproduced from Fisher et al.,[187] by permission of the American Chemical Society.)

chemistry. This promising technology, which is still under investigation, may be applied successfully to environmental analysis of PAH compounds.

Several imaging methods have been investigated recently. Most of these methods applied imaging Raman analysis,[179–183] but fluorescence imaging has been developed as well.[184–186] Old-type fluorescence imaging techniques were based on irradiation of the sample with UV light, and observation of the imaging fluorescence through optical filters. This method provided very limited spectral information, which was not sufficient for the identification of compounds or the resolution of mixtures.

The above problem, as well as optical throughput problems, were recently solved by Fourier transform imaging spectrometry. Both spatial and spectral resolution were obtained when a Fourier transform spectrometer was coupled to a microscope and a CCD camera. This imaging device provides a full fluorescence spectrum at each pixel. The imaging fluorescence provides both chemical and morphological information, which is useful in environmental analysis.

The imaging fluorescence technique has been applied to PAH particulate materials.[187,188] The experimental set-up consisted of a fluorescence microscope, a Fourier transform imaging spectrometer and a CCD detector, as shown schematically in Figure 14. This set-up was applied to the analysis of PAH aerosols on glass fiber filters.

Inspection of PAH contaminated filters through a fluorescence microscope provides 3-D maps, as shown in Figure 15(a). In this example, a glass filter contaminated with a mixture of PAH aerosols was inspected (the image area was $213 \times 232\,\mu m$). Here, the wavelength-integrated image is shown. The peaks represent the emitting contaminating aerosols, while the contour map at the bottom indicates their precise location and size. This figure presents the integrated intensities, but the imaging Fourier transform spectrometer provides a full spectrum at each pixel. Therefore, a proper classification can be carried out, resolving the chemical identity of each peak. The spectral classification procedure results in a 2-D mapping of the filter contamination (Figure 15b). This figure corresponds to the same filter area as Figure 15(a). The dark and light gray pixels represent contamination by perylene and coronene respectively.

In addition to the combined chemical and morphological information, this technique provides excellent sensitivity. The absolute detection limits (for aerosols collected on a glass fiber filter) were estimated to be as low as 0.25 pg, corresponding to an air concentration of $0.5\,pg\,m^{-3}$, when integrated in 1 min.[187] These results correspond to almost on-line particulate PAH analysis.

The above fluorescence imaging technique was also applied to the analysis of PAHs compounds sorbed on sand particles.[188] The absolute sensitivity in this case was in the pg range, which corresponds to global sand contamination in the ppm range.

A somewhat different imaging technique, based on an acousto-optic tuneable filter (ATOF) combined with a CCD detector, was also reported. The ATOF is a solid-state electronic device, which replaces the grating of a traditional spectrometer. Its principle of operation

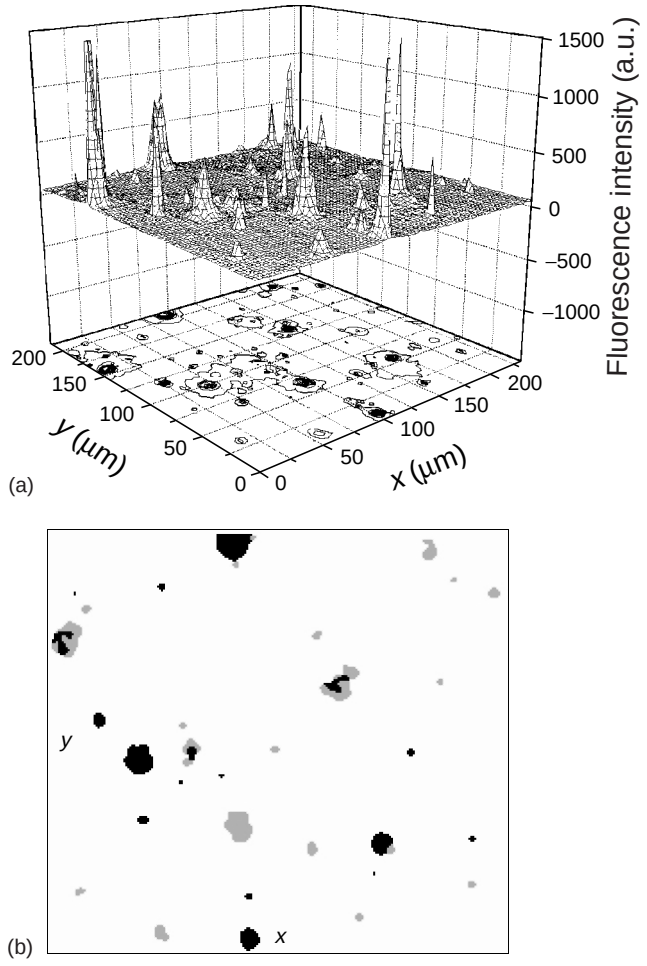

Figure 15 (a) 2-D wavelength integrated fluorescence imaging map, obtained from a PAH contaminated filter. Fluorescence intensities of emitting aerosols are shown, together with their contour map. (b) 2-D classification map obtained from the contaminated filter shown in (a). The dark and light gray pixels represent perylene and coronene aerosols, respectively. (Reproduced from Fisher et al.,[187] by permission of the American Chemical Society.)

is simple. When an acoustic wave propagates in an anisotropic crystal it produces a periodic modulation of the refractive index. The result is a grating that diffracts portions of an incident light beam.[189] When the ATOF is placed in front of a CCD detector it enables both spectral and spatial resolution. The advantages of the ATOF device for imaging purposes are the lack of moving parts and the relatively rapid tuneability over a broad spectral range. An imaging spectrometer for solid-surface room-temperature phosphorimetry, based on ATOF, was recently reported as a tool for PAH analysis.[98] Nevertheless, the major drawback of ATOF devices for PAH analysis is their relatively low optical throughput. This problem becomes crucial when PAH traces are concerned.

5 CONCLUDING REMARKS

Although numerous analytical techniques are available, environmental PAH analysis is still considered a challenging task. Simple PAH mixtures in homogeneous matrixes can now be handled by a variety of methods, but analysis of a complicated environmental sample requires time-consuming investigation of the particular case in hand. Consequently, not all qualified analytical laboratories are capable of handling complicated PAH analysis. In most cases, this is a task for specialists. Nevertheless, several emerging technologies are now under investigation in many laboratories (some of them were briefly mentioned here). A promising direction is the development of PAH sensors, based on recent advances in material science (chemical sensors) and in optoelectronics (physical sensors). It is hoped that these methods will provide a better solution to the problem of on-line and in situ detection of PAHs in the environment.

ABBREVIATIONS AND ACRONYMS

ASE	Accelerated Solvent Extraction
ATOF	Acousto-optic Tuneable Filter
CCD	Charge-coupled Device
CD	Cyclodextrin
CE	Capillary Electrophoresis
CT	Charge Transfer
EEM	Excitation–Emission Fluorescence Spectrometry
ELISA	Enzyme-linked Immunosorbent Assay
FID	Flame Ionization Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
IAC	Immunoaffinity Chromatography
IR	Infrared
LIF	Laser-induced Fluorimetry
L2TOF	Laser Desorption/Laser Ionization (two-step laser) Time-of-flight
MAE	Microwave-assisted Extraction
MI	Matrix Isolation
MPI	Multiphoton Ionization
MPI/FC	Multiphoton Ionization Fast-conductivity
MPI/TOFMS	Multiphoton Ionization/Time-of-flight Mass Spectrometry
MS	Mass Spectrometer
Nd : YAG	Neodymium : Yttrium Aluminum Garnet

NIST	National Institute of Standards and Technology
PAC	Polycyclic Aromatic Compound
PAH	Polycyclic Aromatic Hydrocarbon
PAS	Photoelectric Aerosol Sensor
PASH	Polycyclic Aromatic Sulfur Heterocycle
ppb	parts per billion
REMPI	Resonance-enhanced Multiphoton Ionization
REMPI/TOFMS	Resonance-enhanced Multiphoton Ionization Time-of-flight Mass Spectrometry
RIA	Radioimmunoassay
RTP	Room-temperature Phosphorescence Spectrometry
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SFS	Synchronous Fluorescence Spectrometry
SPE	Solid-phase Extraction
SRM	Standard Reference Material
SS	Shpol'skii Spectroscopy
SSRTP	Solid-surface Room-temperature Phosphorescence Spectrometry
THF	Tetrahydrofuran
TRLIF	Time-resolved Laser-induced Fluorimetry
USEPA	United States Environmental Protection Agency
UV/LIF	Ultraviolet Laser-induced Fluorescence
UV/VIS	Ultraviolet/Visible
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses • Industrial Waste Dumps, Sampling and Analysis • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Luminescence in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Waste Extraction Procedures

REFERENCES

1. J.G. Grasselly, 'Analytical Chemistry – Feeding the Environmental Revolution?', *Anal. Chem.*, **64**, 677A (1992).
2. B.J. Finlayson-Pitts, J.N. Pitts, Jr, *Atmospheric Chemistry: Fundamental and Experimental Techniques*, John Wiley & Sons, NY, 870 – 958, 1986.
3. E.L. Kennaway, I. Hieger, *Br. Med. J., II*, 1044 (1930).
4. I.Y. Wang, R.E. Rasmussen, T.T. Crocker, 'Isolation and Characterization of a DNA-binding Metabolite of Benzo[*a*]pyrene from Hamster Liver Microsomal Incubation System', *Biochem. Biophys. Res. Commun.*, **49**(4), 1142 – 1149.
5. T.G. Spiro, W.M. Stigliani (eds.), *Chemistry of the Environment*, Prentice Hall, Upper Saddle River, NJ, 1996.
6. E. Galli, 'The Role of Microorganisms in the Environmental Decontamination', in *Contaminants in the Environment, A Multidisciplinary Assessment of Risks to Man and Other Organisms*, eds. A. Renzoni, N. Mattei, L. Lari, M.C. Fossi, Lewis Publishers, CRC Press, Boca Raton, FL, 235 – 246, 1994.
7. L.H. Keith, 'Environmental Sampling', in *Environmental Sampling and Analysis: A Practical Guide*, Lewis Publishers, CRC Press, Boca Raton, Florida, 3 – 57, 1991.
8. M.L. Yu, R.A. Hites, 'Identification of Organic Compounds on Diesel Engine Soot', *Anal. Chem.*, **53**(7), 951 – 954 (1981).
9. T. Vo-Dinh, 'Significance of Chemical Analysis of Polycyclic Aromatic Compounds and Related Biological Systems', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 1 – 25, 1988.
10. D.J. Futoma, S.R. Smith, T.E. Smith, J. Tanaka, *Polycyclic Aromatic Hydrocarbons in Water Systems*, CRC Press, Boca Raton, FL, 1981.
11. A.N. Gennadiyev, Yu.I. Pikovsky, I.S. Kozin, 'Polycyclic Aromatic Hydrocarbons in Soils. Environmental and Behavioral Issues', *Polycyclic Aromatic Compounds*, **9**, 169 – 176 (1996).

12. D.A. Dzombak, R.G. Luthy, 'Estimating Adsorption of Polycyclic Aromatic Hydrocarbons on Soils', *Soil Science*, **137**(5), 292–308 (1984).
13. M.T. Butterfield, R.A. Agbaria, I.M. Warner, 'Extraction of Volatile PAHs from Air by Use of Solid Cyclodextrin', *Anal. Chem.*, **68**(7), 1187–1190 (1996).
14. D.A. Lane, 'Polycyclic Aromatic Hydrocarbons: Atmospheric Physics and Chemistry', in *Organic Chemistry of the Atmosphere*, eds. L.D. Hansen, D.J. Eatough, CRC Press, Boca Raton, FL, 155–198, 1991.
15. W.C. Hinds, 'Filtration', in *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*, John Wiley, New York, 164–210, 1982.
16. R. Niessner, 'Sampling Techniques for Air Pollutants', in *Environmental Analysis Techniques, Applications and Quality Assurance*, ed. D. Barcelo, Elsevier Science Publishers, Amsterdam, 3–18, 1993.
17. V.V. Gridin, I. Litani-Barzilai, M. Kadosh, I. Schechter, 'A Renewable Liquid Droplet Method for On-line Pollution Analysis by Multi-photon Ionization', *Anal. Chem.*, **69**(11), 2098–2102 (1997).
18. H.S. Hertz, W.E. May, S.A. Wise, S.N. Chesler, 'Trace Organic Analysis', *Anal. Chem.*, **50**(4), 429A (1978).
19. A. Eschenbach, M. Kästner, R. Bierl, G. Schaefer, B. Mahro, 'Evaluation of a New, Effective Method to Extract Polycyclic Aromatic Hydrocarbons from Soil Samples', *Chemosphere*, **28**(4), 683–692 (1994).
20. A. Eschenbach, R. Wienberg, B. Mahro, 'Fate and Stability of Nonextractable Residues of [14C]PAH in Contaminated Soils Under Environmental Stress Conditions', *Environ. Sci. Technol.*, **32**(17), 2585–2590 (1998).
21. R. Martens, 'Concentrations and Microbial Mineralization of Four to Six Ring Polycyclic Aromatic Hydrocarbons in Composted Municipal Waste', *Chemosphere*, **11**(8), 761–770 (1982).
22. J.D. Berset, M. Ejem, R. Holzer, P. Lischer, 'Comparison of Different Drying, Extraction and Detection Techniques for the Determination of Priority Polycyclic Aromatic Hydrocarbons in Background Contaminated Soil Samples', *Anal. Chim. Acta*, **383**(3), 263–275 (1999).
23. B.M. Tissue, Science Hypermedia, Inc., 1400 Gladewood Drive, Blacksburg, VA, http://www.scimedia.com, 1997.
24. V. Janda, K.D. Bartle, J. Clifford, 'Comparison of Different Drying, Extraction and Detection Techniques for the Determination of Priority Polycyclic Aromatic Hydrocarbons in Background Contaminated Soil Samples', *J. Chromatogr.*, **642**(1–2), 283–299 (1993).
25. B.A. Benner, Jr, 'Summarizing the Effectiveness of Supercritical Fluid Extraction of Polycyclic Aromatic Hydrocarbons from Natural Matrix Environmental Samples', *Anal. Chem.*, **70**(21), 4594–4601 (1998).
26. Y.Z. Hsieh, A.C. Chang, 'Effects of Analyte–Matrix Interactions on Supercritical Fluid Extraction Efficiencies of Polycyclic Aromatic Hydrocarbons', *Anal. Chim. Acta*, **342**(1), 41–50 (1997).
27. J.J. Langenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Kinetic Study of Supercritical Fluid Extraction of Organic Contaminants from Heterogeneous Environmental Samples with Carbon Dioxide and Elevated Temperatures', *Anal. Chem.*, **67**(10), 1727–1736 (1995).
28. M. Barzegar, A. Daneshfar, M. Ashraf-Khorassari, 'Effect of Parameters on Supercritical Fluid Extraction of Nitro Polynuclear Aromatic Hydrocarbons from Sand', *Anal. Chim. Acta*, **349**(1–3), 245–252 (1997).
29. S.B. Lagenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Effects of Temperature and Pressure on Supercritical Fluid Extraction Efficiencies of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls', *Anal. Chem.*, **65**(4), 338–344 (1993).
30. J. Dachs, I. Fernandez, J.M. Beyona, 'Combined Experimental Design and Information Theory for the Optimization of Supercritical Fluid Extraction of Organic Priority Pollutants from Sediment', *Anal. Chim. Acta*, **351**(1–3), 377–385 (1997).
31. US EPA Method 3561, Revision 0, January 1995, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
32. M.M. Schantz, J.J. Nichols, S.A. Wise, 'Evaluation of Pressurized Fluid Extraction for the Extraction of Environmental Matrix Reference Materials', *Anal. Chem.*, **69**(20), 4210–4219 (1997).
33. US EPA, Method sw-846, 3545, 1995, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
34. I.J. Barnabas, J.R. Dean, I.A. Fowlis, S.P. Owen, 'Extraction of Polycyclic Aromatic Hydrocarbons from Highly Contaminated Soils Using Microwave Energy', *Analyst*, **120**(7), 1897–1904 (1995).
35. V. Lopez-Avila, R. Young, 'Microwave-assisted Extraction of Organic Compounds from Standard Reference Soils and Sediments', *Anal. Chem.*, **66**(7), 1097–1106 (1994).
36. US EPA Method 3611, Revision 2, December 1996, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
37. US EPA Method 3620, Revision 2, December 1996, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
38. US EPA Method 3630, Revision 3, December 1996, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
39. D.E. Wells, 'Extraction, Clean-up and Recoveries of Persistent Trace Organic Contaminants from Sediment and Biota Samples', in *Environmental Analysis Techniques, Applications and Quality Assurance*, ed. D. Barcelo, Elsevier Science Publishers, Amsterdam, 80–105, 1993.
40. S.R. Wild, K.S. Waterhouse, S.P. McGrath, K.C. Jones, 'Organic Contaminants in an Agricultural Soil with a Known History of Sewage Sludge Amendments: Polynuclear Aromatic Hydrocarbons', *Environ. Sci. Technol.*, **24**(11), 1706–1711 (1990).
41. G. Morel, O. Samhan, P. Literathy, H. Al-Hashash, L. Moulin, T. Saeed, K. Al-Matrouk, M. Martin-Bouyer, A. Saber, L. Paturel, J. Jarosz, M. Vial, E. Combet,

C. Fachinger, 'Evaluation of Chromatographic and Spectroscopic Methods for the Analysis for Petroleum-derived Compounds in the Environment', *Fresenius' J. Anal. Chem.*, **339**(10), 699–715 (1991).
42. US EPA Method 3640, Revision 1, September 1994, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
43. US EPA Method 3650, Revision 2, December 1996, http://www.epa.gov/epaoswer/hazwaste/test/main.htm
44. J.C. Fetzer, 'Gas- and Liquid-chromatographic Techniques', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 59–101, 1989.
45. K.E. Markides, M. Nishioka, B.J. Tarbet, J.S. Bradshaw, M.L. Lee, 'Smectic Biphenylcarboxylate Ester Liquid Crystalline Polysiloxane Stationary Phase for Capillary Gas Chromatography', *Anal. Chem.*, **57**(7), 1296–1299 (1985).
46. W.-S. Lee, G.-P. Chang-Chien, 'All-hydrocarbon Liquid Crystalline Polysiloxane Polymer as Stationary Phase in Gas Chromatography Capillary Column for Separation of Isomeric Compounds of Polynuclear Aromatic Hydrocarbons', *Anal. Chem.*, **70**(19), 4094–4099 (1998).
47. M.L. Lee, R.C. Kong, C.L. Woolley, J.S. Bradshaw, 'Fused Silica Capillary Column Technology for Gas Chromatography', *J. Chromatogr. Sci.*, **22**(4), 136–142 (1984).
48. M. Nishioka, B.A. Jones, B.J. Tarbet, J.S. Bradshaw, M.L. Lee, 'Gas Chromatographic Retention Behavior of Polycyclic Aromatic Compounds on Smectic Liquid-crystalline Polysiloxane Stationary Phases', *J. Chromatogr.*, **375**(1), 79–91 (1986).
49. R. Kaliszan, H. Lamparczyk, 'A Relationship Between the Connectivity Indices and Retention Indices of Polycyclic Aromatic Hydrocarbons', *J. Chromatogr. Sci.*, **16**(6), 246–248 (1978).
50. M.D. Erichson, D.L. Newton, E.D. Pellizzari, K.B. Tomer, 'Identification of Alkyl-9-fluorenones in Diesel Exhaust Particulate', *J. Chromatogr. Sci.*, **17**(8), 449–454 (1979).
51. B. Schmid, J.T. Andersson, 'Critical Examination of the Quantification of Aromatic Compounds in Three Standard Reference Materials', *Anal. Chem.*, **69**(17), 3476–3481 (1997).
52. J.C. Fetzer, W.E. Biggs, 'The High-performance Liquid Chromatography of Peropyrene-type Polycyclic Aromatic Hydrocarbons', *J. Chromatogr.*, **295**(1), 161–169 (1984).
53. S.A. Wise, W.J. Bonnett, F.R. Guenther, W.E. May, 'A Relationship Between Reversed-phase C18 Liquid Chromatographic Retention and the Shape of Polycyclic Aromatic Hydrocarbons', *J. Chromatogr. Sci.*, **19**(9), 457–454 (1981).
54. F. Sun, D. Littlejohn, M.D. Gibson, 'Ultrasonication Extraction and Solid Phase Extraction Clean-up for Determination of US EPA 16 Priority Pollutant Polycyclic Aromatic Hydrocarbons in Soils by Reversed-phase Liquid Chromatography with Ultraviolet Absorption Detection', *Anal. Chim. Acta*, **364**(1–3), 1–11 (1998).
55. R. Ferrer, J.L. Beltran, B.J. Guiteras, 'Mathematical Procedure for the Determination of the Breakthrough Volumes of Polycyclic Aromatic Hydrocarbons (PAHs)', *Anal. Chim. Acta*, **346**(2), 253–258 (1997).
56. J.L. Beltran, J. Guiteras, R. Ferrer, 'Three-way Multivariate Calibration Procedures Applied to High-performance Liquid Chromatography Coupled with Fast-scanning Fluorescence Spectrometry Detection. Determination of Polycyclic Aromatic Hydrocarbons in Water Samples', *Anal. Chem.*, **70**(9), 1949–1955 (1998).
57. R. Dadoo, R.N. Zaro, C. Yan, D.S. Anex, 'Advances in Capillary Electrochromatography: Rapid and High-efficiency Separations of PAHs', *Anal. Chem.*, **70**(22), 4787–4792 (1998).
58. S. Chen, M. Meyerhoff, 'Three-way Multivariate Calibration Procedures Applied to High-performance Liquid Chromatography Coupled with Fast-scanning Fluorescence Spectrometry Detection. Determination of Polycyclic Aromatic Hydrocarbons in Water Samples', *Anal. Chem.*, **70**(9), 2523–2529 (1998).
59. J. Chu, T. Takeuchi, T. Miwa, 'Retention Behavior of Polycyclic Aromatic Hydrocarbons in Microcolumn Liquid Chromatography with Acridine Derivative as Stationary Phase', *Anal. Chim. Acta*, **370**(2–3), 125–130 (1998).
60. E. Clar, *The Aromatic Sextet*, Academic Press, New York, 1972.
61. J.C. Fetzer, W.E. Biggs, 'Use of Full-spectrum Absorbance and Emission Detectors in Environmental Analyses', *J. Chromatogr.*, **642**(1–2), 319–327 (1993).
62. J. Johansson, D.T. Witte, M. Larsson, S. Nilson, 'Real-time Fluorescence Imaging of Isotachophoretic Preconcentration for Capillary Electrophoresis', *Anal. Chem.*, **68**, 2766–2770 (1996).
63. H.J. Richardson, *Modern Fluorescence Spectroscopy*, ed. E.L. Wehry, Plenum Press, New York, Vol. 3, 1981.
64. D.J. Desilets, J.T. Coburn, D.A. Lantrip, P.T. Kissinger, F.E. Lytle, 'On-the-fly Determination of Fluorescence Lifetimes from Two-point Decay Measurements', *Anal. Chem.*, **58**(6), 1123–1128 (1986).
65. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open-tubular Capillaries', *Anal. Chem.*, **56**(1), 111–113 (1984).
66. D.E. Burton, M.J. Sepaniak, M.P.J. Maskarinec, 'Evaluation of the Use of Various Surfactants in Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. Sci.*, **25**(11), 514–518 (1987).
67. O.H.J. Szolar, R.S. Brown, J.H.T. Luong, 'Separation of PAHs by Capillary Electrophoresis with Laser-induced Fluorescence Detection Using Mixtures of Neutral

and Anionic β-Cyclodextrins', *Anal. Chem.*, **67**(17), 3004–3010 (1995).
68. R.S. Brown, J.H.T. Luong, O.H.J. Szolar, A. Halasz, J. Hawari, 'Cyclodextrin-modified Capillary Electrophoresis: Determination of Polycyclic Aromatic Hydrocarbons in Contaminated Soils', *Anal. Chem.*, **68**(2), 287–292 (1996).
69. J.L. Miller, M.G. Khaledi, D. Shea, 'Separation of Polycyclic Aromatic Hydrocarbons by Nonaqueous Capillary Electrophoresis Using Charge-transfer Complexation with Planar Organic Cations', *Anal. Chem.*, **69**(6), 1223–1229 (1997).
70. M. Zander, 'Ultraviolet Absorption and Luminescence Spectrometry: An Overview of Recent Developments', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 171–200, 1989.
71. F. Ariese, C. Gooijer, N.H. Velthorst, 'Application of Fluorescence Spectroscopic Techniques in the Determination of PAHs and PAH Metabolites', in *Environmental Analysis Techniques, Applications and Quality Assurance*, ed. D. Barcelo, Elsevier Science Publishers, Amsterdam, 449–479, 1993.
72. J. Guiteras, J.L. Beltran, R. Ferrer, 'Quantitative Multicomponent Analysis of Polycyclic Aromatic Hydrocarbons in Water Samples', *Anal. Chim. Acta*, **361**(3), 233–240 (1998).
73. I.M. Warner, G.D. Christian, E.R. Davidson, 'Analysis of Multicomponent Fluorescence Data', *Anal. Chem.*, **49**(4), 564–573 (1977).
74. C.-N. Ho, G.D. Christian, E.R. Davidson, 'Application of the Method of Rank Annihilation to Quantitative Analyses of Multicomponent Fluorescence Data From the Video Fluorometer', *Anal. Chem.*, **50**(8), 1108–1113 (1978).
75. C.-N. Ho, G.D. Christian, E.R. Davidson, 'Application of the Method of Rank Annihilation to Fluorescent Multicomponent Mixtures of Polynuclear Aromatic Hydrocarbons', *Anal. Chem.*, **52**(7), 1071–1079 (1980).
76. K.S. Booksh, A.R. Muroski, M.L. Myrick, 'Single-measurement Excitation/Emission Matrix Spectrofluorometer for Determination of Hydrocarbons in Ocean Water. 2. Calibration and Quantitation of Naphthalene and Styrene', *Anal. Chem.*, **68**(20), 3539–3544 (1996).
77. J.L. Beltran, R. Ferrer, J. Guiteras, 'Multivariate Calibration of Polycyclic Aromatic Hydrocarbon Mixtures from Excitation–Emission Fluorescence Spectra', *Anal. Chim. Acta*, **373**(2–3), 311–319 (1998).
78. A. D'Silva, V.A. Fassel, 'Laser-excited Shpol'skii Spectrometry', *Anal. Chem.*, **56**(8), 985A–1000A (1984).
79. F. Ariese, *Shpols'kii Spectroscopy and Synchronous Fluorescence Spectroscopy*, Ph.D. Thesis, Free University, Amsterdam, 1993.
80. F. Ariese, S.J. Kok, M. Verkaik, G.Ph. Hoornweg, C. Gooijer, N.H. Velthorst, J.W. Hofstraat, 'Chemical Derivatization and Shpol'skii Spectrofluorometric Determination of Benzo[*a*]pyrene Metabolites in Fish Bile', *Anal. Chem.*, **65**(8), 1100–1106 (1993).
81. F. Ariese, C. Gooijer, N.H. Velthorst, J.W. Hofstraat, 'Shpol'skii Spectrofluorometric Determination of Polycyclic Aromatic Hydrocarbons in Biota', *Anal. Chim. Acta*, **232**(2), 245–251 (1990).
82. I.S. Kozin, C. Gooijer, N.H. Velthorst, J. Harmsen, R. Wieggers, 'Direct Determination of Isomeric Polycyclic Aromatic Hydrocarbons in Environmental Samples by Conventional and Laser Excited Shpol'skii Spectroscopy', *Intern. J. Environ. Anal. Chem.*, **61**(4), 285–297 (1995).
83. C. Gooijer, F. Ariese, J.W. Hofstraat, N.H. Velthorst, 'Shpol'skii Spectroscopy, a Cryogenic High-resolution Molecular Fluorescence Technique with a Distinct Potential in Analytical Chemistry', *Trends Anal. Chem.*, **13**(2), 53–61 (1994).
84. J. Pfab, 'Laser-induced Fluorescence Spectroscopy', in *Applied Laser Spectroscopy: Techniques, Instrumentation, and Applications*, ed. D.L. Andrews, VCH Publishers, Inc., NY, 111–184, 1992.
85. G.R. Haugen, F.E. Lytle, 'Quantitation of Fluorophores in Solution by Pulsed Laser Excitation and Time-filtered Detection', *Anal. Chem.*, **53**(11), 1554–1559 (1981).
86. K.D. Hughes, F.E. Lytle, M.D. Huber, 'Identification of Immobilized Bacteria by Aminopeptidase Profiling', *Anal. Chem.*, **61**(15), 1656–1660 (1989).
87. N.T. Azimi, M.D. Huber, J.E. Whitaker, R.P. Haugland, F.E. Lytle, 'Multiple Reagent [for] Aminopeptidase Profiling of Bacteria', *Appl. Spectrosc.*, **44**(3), 400–403 (1990).
88. J.D. Fasset, J.C. Travis, 'Analytical Applications of Resonance Ionization Mass Spectrometry (RIMS)', *Spectrochim. Acta, Part B*, **43B**(22), 1409–1422 (1988).
89. H.-B. Kim, M. Hayashi, K. Nakatani, N. Kitamura, K. Sasaki, J.-I. Hotta, H. Mashuhara, 'In Situ Measurements of Ion-exchange Processes in Single Polymer Particles: Laser Trapping Microspectroscopy and Confocal Fluorescence Microspectroscopy', *Anal. Chem.*, **68**(3), 409–414 (1996).
90. G.J. Diebold, R.N. Zare, 'Laser Fluorimetry: Subpicogram Detection of Aflatoxins Using High-pressure Liquid Chromatography', *Science*, **196**(4297), 1439–1441 (1977).
91. G.J. Diebold, N. Karny, R.N. Zare, L.M. Seitz, 'Laser Fluorometric Determination of Aflatoxin B1 in Corn', *J. Assoc. Off. Anal. Chem.*, **62**(3), 564–569 (1979).
92. W.L. Wiese, 'Spectroscopic Reference Data', in *Inductively Coupled Plasma in Analytical Atomic Spectroscopy*, eds. A. Montaser, D.W. Golightly, VCH Publishers, Inc., New York, 451–471, 1987.
93. D.A. Cremers, J.E. Barefield, A.C. Koskelo, 'Remote Elemental Analysis by Laser-induced Breakdown Spectroscopy Using a Fiber-optic Cable', *Appl. Spectrosc*, **49**(6), 857–860 (1995).

94. B. Horowitz, V.V. Gridin, V. Bulatov, I. Schechter, 'Laser-induced Fluorescence of Perylene in a Microparticle Suspension Environment', *Anal. Chem.*, **70**(15), 3191–3197 (1998).
95. A.D. Campiglia, D.M. Hueber, T. Vo-Dinh, 'Laser-induced Solid-surface Room-temperature Phosphorimetry of Polycyclic Aromatic Hydrocarbons', *Appl. Spectrosc.*, **50**(2), 252–256 (1996).
96. A.D. Campiglia, T. Vo-Dinh, 'Fiber Optic Sensor for Laser-induced Room-temperature Phosphorescence Detection of Polycyclic Aromatic Compounds', *Talanta*, **43**(10), 1805–1814 (1996).
97. A.D. Campiglia, J.P. Alarie, T. Vo-Dinh, 'Development of a Room-temperature Phosphorescence Fiber-optic Sensor', *Anal. Chem.*, **68**(9), 1599–1604 (1996).
98. A.D. Campiglia, F. Moreau, D.M. Hueber, T. Vo-Dinh, 'Phosphorescence Imaging System Using an Acousto-optic Filter-based Charge Coupled Device', *Anal. Chim. Acta*, **351**(1–3), 229–239 (1997).
99. M. Silverstein, G.C. Bassler, *Spectroscopic Identification of Organic Compounds*, Wiley, New York, 1967.
100. P. Stout, G. Mamantov, 'Recent Advances Infrared Analysis of Polycyclic Aromatic Compounds', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 411–432, 1989.
101. D.M. Hudgins, S.A. Sandford, 'Infrared Spectroscopy of Matrix Isolated Polycyclic Aromatic Hydrocarbons. 1. PAHs Containing Two to Four Rings', *J. Phys. Chem. A*, **102**(2), 329–343 (1998).
102. C. Joblin, L. d'Hendecourt, A. Leger, D. Defourneau, 'Infrared Spectroscopy of Gas-phase PAH Molecules I. Role of the Physical Environment', *Astron. Astrophys*, **281**(3), 923–936 (1994).
103. C. Joblin, P. Boissel, A. Leger, L. d'Hendecourt, D. Defourneau, 'Infrared Spectroscopy of Gas-phase PAH Molecules. II. Role of the Temperature', *Astron. Astrophys*, **299**(3), 835–846 (1995).
104. P.L. Stout, G. Mamantov, 'Matrix Isolation FTIR Studies of Nitro Polycyclic Aromatic Hydrocarbons', *Appl. Spectrosc.*, **41**(6), 1048–1052 (1987).
105. K.S. Chiu, K. Biemann, K. Krishnan, S.L. Hill, 'Structural Characterization of Polycyclic Aromatic Compounds by Combined Gas Chromatography/Mass Spectrometry and Gas Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chem.*, **56**(9), 1610–1615 (1984).
106. D.M. Hembree, A.A. Garrison, R.A. Crocombe, R.A. Yokley, E.L. Wehry, G. Mamantov, 'Matrix Isolation Fourier Transform Infrared Spectrometric Detection in the Open Tubular Column Gas Chromatography of Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **53**(12), 1783–1788 (1981).
107. D.L. Gerrard, H.J. Bowley, 'Raman Spectroscopy', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 433–450, 1989.
108. F. Adar, R. Geiger, J. Noonan, 'Raman Spectroscopy for Process/Quality Control', *Appl. Spectrosc. Rev.*, **32**(1–2), 45–101 (1997).
109. D.B. Chase, 'Fourier Transform Raman Spectroscopy', *J. Am. Chem. Soc.*, **108**(24), 7485–7488 (1986).
110. B. Schrader, 'Raman Spectroscopy of Mineral Oil Products Part I: NIR/FT-Raman Spectra of Polycyclic Aromatic Hydrocarbons', *Appl. Spectrosc.*, **45**(8), 1230–1232 (1991).
111. S.A. Asher, 'Ultraviolet Resonance Raman Spectrometry for Detection and Speciation of Trace Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **56**(4), 720–724 (1984).
112. C.M. Jones, T.A. Naim, M. Ludwig, J. Murtaugh, P.L. Flaugh, J.M. Dudik, C.R. Johnson, S.A. Asher, 'Analytical Applications of Ultraviolet Resonance Raman Spectroscopy', *Trends Anal. Chem.*, **4**(3), 75–80 (1985).
113. C.H. Munro, V. Pajcini, S.A. Asher, 'Dielectric Stack Filters for Ex Situ and In Situ UV Optical-fiber Probe Raman Spectroscopic Measurements', *Appl. Spectrosc.*, **51**(11), 1722–1729 (1997).
114. R. Niessner, 'The Chemical Response of the Photoelectric Aerosol Sensor (PAS) to Different Aerosol Systems', *J. Aerosol. Sci.*, **17**(4), 705–714 (1986).
115. R. Niessner, P. Wilbring, 'Ultrafine Particles as Trace Catchers for Polycyclic Aromatic Hydrocarbons: The Photoelectric Aerosol Sensor as a Tool for In Situ Sorption and Desorption Studies', *Anal. Chem.*, **61**(7), 708–714 (1989).
116. R. Niessner, W. Robers, P. Wilbring, 'Laboratory Experiments on the Determination of Polycyclic Aromatic Hydrocarbon Coverage of Submicrometer Particles by Laser-induced Aerosol Photoemission', *Anal. Chem.*, **61**(4), 320–325 (1989).
117. R. Niessner, B. Hemmerich, P. Wilbring, 'Aerosol Photoemission for Quantification of Polycyclic Aromatic Hydrocarbons in Simple Mixtures Adsorbed on Carbonaceous and Sodium Chloride Aerosols', *Anal. Chem.*, **62**(19), 2071–2074 (1990).
118. O.P. Haefliger, R. Zenobi, 'Laser Mass Spectrometric Analysis of Polycyclic Aromatic Hydrocarbons with Wide Wavelength Range Laser Multiphoton Ionization Spectroscopy', *Anal. Chem.*, **70**(13), 2660–2665 (1998).
119. S.M. Hankin, P. John, G.P. Smith, 'Effect of Carbon Loading on the Extraction Properties of C-18 Bonded Silica Used for Solid-phase Extraction of Acidic and Basic Analytes', *Anal. Chem.*, **69**(15), 2972–2975 (1997).
120. P. Voumard, Q. Zhan, R. Zenobi, 'A New Instrument for Spatially Resolved Laser Desorption/Laser Multiphoton Ionization Mass Spectrometry', *Rev. Sci. Instrum.*, **64**(8), 2215–2220 (1993).

121. M.S. De Vries, D.J. Elloway, H.R. Wendt, H.E. Hunziker, 'Photoionization Mass Spectrometer with a Microscope Laser Desorption Source', *Rev. Sci. Instrum.*, **63**(6), 3321–3325 (1992).
122. A. Hachimi, G. Krier, E. Poitevin, M.C. Schweigert, S. Peter, J.F. Muller, 'Characterization of Nitro-PAH Adsorbed on Environmental Micro Particles: A Comparison of LAMMA, Fourier Transform Mass Spectrometry Laser Microprobes and Gas Chromatography–Tandem Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **62**(3), 219–230 (1996).
123. M.J. Dale, A.C. Jones, S.J.T. Pollard, P.R.R. Langridge-Smith, 'Direct Determination of Polycyclic Aromatic Hydrocarbons in Environmental Matrixes Using Laser Desorption Laser Photoionization Time-of-flight Mass Spectrometry', *Analyst*, **119**(4), 571–578 (1994).
124. M.J. Dale, O.H.J. Downs, K.F. Costello, S.J. Write, P.R.R. Langridge-Smith, J.N. Cape, 'Direct Analysis of Polycyclic Aromatic Hydrocarbons in Cloud-water Aerosol Filtrates Using Laser Desorption Mass Spectrometry', *Environ. Pollut.*, **89**(2), 123–129 (1995).
125. M.J. Dale, A.C. Jones, S.J.T. Pollard, P.R.R. Langridge-Smith, A.G. Rowley, 'Application of Two-step Laser Mass Spectrometry to the Analysis of Polynuclear Aromatic Hydrocarbons in Contaminated Soils', *Environ. Sci. Technol.*, **27**(8), 1693–1695 (1993).
126. R. Zenobi, 'Advances in Surface Analysis and Mass Spectrometry Using Laser Desorption Methods', *Chimia*, **48**(3), 64–71 (1994).
127. R. Zenobi, 'In Situ Analysis of Surfaces and Mixtures by Laser Desorption Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **145**(1/2), 51–77 (1995).
128. Q. Zane, P. Voumard, R. Zenobi, 'Application of Two-step Laser Mass Spectrometry to the Chemical Analysis of Aerosol Particle Surfaces', *Rapid Comm. Mass Spectrom.*, **9**(2), 119–127 (1995).
129. D.Z. Bezabeh, A.D. Jones, L.L. Ashbaugh, P.B. Kelly, 'Screening of Aerosol Filter Samples for PAHs and Nitro-PAHs by Laser Desorption Ionization TOF Mass Spectrometry', *Aerosol Sci. Technol.*, **30**(3), 288–299 (1999).
130. S.M. Hankin, P. John, 'Laser Time-of-flight Mass Analysis of PAHs on Single Diesel Particulates', *Anal. Chem.*, **71**(6), 1100–1104 (1999).
131. O.P. Haefliger, R. Zenobi, 'New Sample Preparation for Quantitative Laser Desorption Mass Spectrometry and Optical Spectroscopy', *Rev. Sci. Instrum.*, **69**(4), 1828–1832 (1998).
132. K. Siomos, 'Laser Multiphoton Ionization Spectroscopy of Polyaromatic Molecules', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 263–325, 1989.
133. R.A. Hites, 'Mass Spectrometry of Polycyclic Aromatic Compounds', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 219–261, 1989.
134. M.N.R. Ashfold, 'Multiphoton Absorption Spectroscopy', *Analytical Proc.*, **28**(12), 416–418 (1991).
135. M.J. Dale, A.C. Jones, S.J.T. Pollard, P.R.R. Langridge-Smith, 'Direct Determination of Polycyclic Aromatic Hydrocarbons in Environmental Matrixes Using Laser Desorption Laser Photoionization Time-of-flight Mass Spectrometry', *Analyst*, **119**(4), 571–578 (1994).
136. X. Fei, G. Wei, K.K. Murray, 'Aerosol MALDI with a Reflectron Time-of-flight Mass Spectrometer', *Anal. Chem.*, **68**(7), 1143–1147 (1996).
137. S.S. Alimpiev, M.E. Belov, V.V. Mlinsky, S.S. Nikiforov, 'Laser Mass Spectrometry of a Frozen Water Matrix in Environmental Analysis', *Analyst*, **119**(4), 579–582 (1994).
138. C.M. Gittins, M.J. Castaldi, S.M. Senkan, E.A. Rohlfing, 'Real-time Quantitative Analysis of Combustion-generated Polycyclic Aromatic Hydrocarbons by Resonance-enhanced Multiphoton Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **69**(3), 286–293 (1997).
139. C. Weickhardt, U. Boesl, E.W. Schlag, 'Laser Mass Spectrometry for Time-resolved Multicomponent Analysis of Exhaust Gas', *Anal. Chem.*, **66**(7), 1062–1069 (1994).
140. H.J. Heger, R. Zimmermann, R. Dorfner, M. Beckmann, H. Grieble, A. Kettrup, U. Boesl, 'On-line Emission Analysis of Polycyclic Aromatic Hydrocarbons Down to pptv Concentration Levels in the Flue Gas of an Incineration Pilot Plant with a Mobile Resonance-enhanced Multiphoton Ionization Time-of-flight Mass Spectrometer', *Anal. Chem.*, **71**(1), 46–57 (1999).
141. D. Parratt, H. McKenzie, K.H. Nielsen, S.J. Cobb, *Radioimmunoassay of Antibodies*, Wiley, New York, 1982.
142. E.T. Maggio, *Enzyme-immunoassay*, CRC Press, Boca Raton, FL, 1983.
143. R.M. Santella, M. Stefanidis, 'Immunological Methods for the Detection and Quantification of Exposure to Aromatic Hydrocarbons', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley, Toronto, Canada, 391–410, 1989.
144. L. Stryer, 'Exploring Proteins', in *Biochemistry*, ed. W.H. Freeman, New York, 62–63, 1988.
145. K. Li, R. Chen, B. Zhao, M. Liu, A.E. Karu, V.A. Roberts, Q.X. Li, 'Monoclonal Antibody-based ELISAs for Part-per-billion Determination of Polycyclic Aromatic Hydrocarbons: Effects of Haptens and Formats on Sensitivity and Specificity', *Anal. Chem.*, **71**(2), 302–309 (1999).
146. B.V. Herikstad, S. Overeboe, A. Haugan, I. Hagen, 'Determination of Polycyclic Aromatic Hydrocarbons in Urine from Coke-oven Workers with a Radioimmunoassay', *Carcinogenesis*, **14**(2), 307–309 (1993).
147. E.W. Weiler, 'Use of Immunoassay in Plant Science. Part II. Radioimmunoassays for the Differential and Direct Analysis of Free and Conjugated Abscissic Acid in Plant Extracts', *Planta*, **148**(3), 262–272 (1980).

148. N.Y. Kado, E.T. Wei, 'Radioimmunoassay for Benzo[*a*]-pyrene', *J. Natl. Cancer Inst.*, **61**(1), 221–225 (1978).
149. A. Roda, A. Pistillo, A. Jus, C. Armanino, M. Baraldini, 'Analysis of Air Particulate Benzo[*a*]pyrene by a Specific Enzyme Immunoassay: Correlation with Chemical and Atmospheric Parameters', *Anal. Chim. Acta*, **298**(1), 53–64 (1994).
150. J. Zühlke, D. Knopp, R. Niessner, 'Determination of 1-Nitropyrene with Enzyme-linked Immunosorbent Assay Versus High-performance Column Switching Technique', *J. Chromatogr. A.*, **807**, 209–217 (1998).
151. B. Fröschl, D. Knopp, R. Niessner, 'Possibilities and Limitations in Enzyme-linked Immunosorbent Assay for the Detection of 1-Nitropyrene in Air Particulate Matter', *Fresenius' J. Anal. Chem.*, **360**, 689–692 (1998).
152. D. Barcelo, A. Oubina, J.S. Salau, S. Perez, 'Determination of PAHs in River Water Samples by ELISA', *Anal. Chim. Acta*, **376**(1), 49–53 (1998).
153. P.M. Krämer, 'A Strategy to Validate Immunoassay Test Kits for TNT and PAHs as a Field Screening Method for Contaminated Sites in Germany', *Anal. Chim. Acta*, **376**, 3–11 (1998).
154. US EPA Method 4035, Revision 0, December 1996, http://www.epa.gov/epaoswer/hazwaste/test/main.htm.
155. S. Perez, I. Ferrer, M.-C. Hennion, D. Barcelo, 'Isolation of Priority Polycyclic Aromatic Hydrocarbons from Natural Sediments and Sludge Reference Materials by an Anti-fluorene Immunosorbent Followed by Liquid Chromatography and Diode Array Detection', *Anal. Chem.*, **70**(23), 4996–5001 (1998).
156. V. Pichon, M. Bouzige, M.-C. Hennion, 'New Trends in Environmental Trace-analysis of Organic Pollutants: Class-selective Immunoextraction and Clean-up in One Step Using Immunosorbents', *Anal. Chim. Acta*, **376**(1), 21–35 (1998).
157. S. Kipp, H. Peyrer, W. Kleibohmer, 'Coupling Superheated Water Extraction with Enzyme Immunoassay for an Efficient and Fast PAH Screening in Soil', *Talanta*, **46**(3), 385–393 (1998).
158. J. Zühlke, D. Knopp, R. Niessner, 'Sol–gel Glass as a New Support Matrix in Immunoaffinity Chromatography', *Fresenius' J. Anal. Chem.*, **352**(7–8), 654–659 (1995).
159. M. Cichna, D. Knopp, R. Niessner, 'Immunoaffinity Chromatography of Polycyclic Aromatic Hydrocarbons in Columns Prepared by the Sol–gel Method', *Anal. Chim. Acta*, **339**(3), 241–250 (1997).
160. T. Ogawa, K. Kise, T. Yasuda, H. Kawazumi, S. Yamada, 'Trace Determination of Benzene and Aromatic Molecules in Hexane by Laser Two-photon Ionization', *Anal. Chem.*, **64**(11), 1217–1220 (1992).
161. E. Voigtman, J.D. Winefordner, 'Two-photon Photoionization Detection of Polycyclic Aromatic Hydrocarbons and Drugs in a Windowless Flow Cell', *Anal. Chem.*, **54**(11), 1834–1839 (1982).
162. T. Yasuda, S. Sato, H. Kawazumi, T. Ogawa, 'Laser Two-photon Ionization Spectrometry of Perylene and Naphthacene in Hexane', *Anal. Chem.*, **59**(22), 2719–2721 (1987).
163. T. Ogawa, S. Sato, M. Tachibana, K. Ideta, T. Inoue, K. Nakashima, 'Dependence of the Laser Two-photon Ionization Signal of Anthracene on the Electron Mobility and the Excess Energy in Non-polar Solvents', *Anal. Chim. Acta.*, **299**(3), 355–359 (1995).
164. T. Ogawa, T. Yasuda, H. Kawazumi, 'Laser Two-photon Ionization of 4,4‴-bis[2-butyloctyl)oxy]-p-quaterphenyl on a Platinum Surface in Ambient Air: Highly Sensitive Detection of Surface Molecules', *Anal. Sci.*, **8**(1), 81–82 (1992).
165. T. Ogawa, T. Yasuda, H. Kawazumi, 'Laser Two-photon Ionization Detection of Aromatic Molecules on a Metal Surface in Ambient Air', *Anal. Chem.*, **64**(21), 2615–2617 (1992).
166. H. Chen, T. Inoue, T. Ogawa, 'Solvents Effects in Determination of Anthracene on Solutions Surfaces by Laser Two-photon Ionization Technique', *Anal. Chem.*, **66**, 4150 (1994).
167. K. Masuda, T. Inoue, T. Yasuda, K. Nakashima, T. Ogawa, 'Highly Sensitive Detection of Pyrene by Laser Multi-photon Ionization on the Surface of Water', *Anal. Sci.*, **9**(2), 297–298 (1993).
168. T. Inoue, K. Masuda, K. Nakashima, T. Ogawa, 'Highly Sensitive Detection of Aromatic Molecules by Laser Two-photon Ionization on the Surface of Water in Ambient Air', *Anal. Chem.*, **66**(7), 1012–1014 (1994).
169. T. Ogawa, H. Chen, K. Masuda, T. Inoue, 'Highly Sensitive Determination of Aromatic Molecules on the Water Surface by Laser Two-photon Ionization at 355 nm', *Anal. Sci.*, **10**(1), 219–221 (1994).
170. T. Ogawa, H. Chen, T. Inoue, K. Nakashima, 'Laser Two-photon Ionization Spectrometry and Photoionization Threshold of Perylene on the Surface of Water', *Chem. Phys. Lett.*, **229**(3), 328–332 (1994).
171. T. Ogawa, H. Chen, T. Inoue, K. Nakashima, 'Laser Two-photon Ionization Spectrometry of Perylene, Naphthacene and Methylanthracene on Water Surface', *Anal. Sci.*, **10**(6), 849–852 (1994).
172. Y.Q. Li, T. Inoue, T. Ogawa, 'Highly Sensitive Detection of Anthracene at the Interface Region of a Hydrocarbon–Water Two-phase System Using Laser Two-photon Ionization', *Anal. Sci.*, **12**(5), 691–693 (1996).
173. V.V. Gridin, A. Korol, V. Bulatov, I. Schechter, 'Laser Two-photon Ionization of Pyrene on Contaminated Soils', *Anal. Chem.*, **68**(19), 3359–3363 (1996).
174. V.V. Gridin, A. Korol, V. Bulatov, I. Schechter, 'Particulate Material Analysis by a Laser Ionization Fast Conductivity Method', *Anal. Chem.*, **69**(3), 478–484 (1997).
175. V. Bulatov, V.V. Gridin, F. Polyak, I. Schechter, 'Application of Pulsed Laser Methods to In Situ Probing

of Highway Originated Pollutants', *Anal. Chim. Acta*, **343**(1–2), 93–99 (1997).

176. V.V. Gridin, I. Litani-Barzilai, M. Kadosh, I. Schechter, 'Determination of Aqueous Solubility and Surface Adsorption of Polycyclic Aromatic Hydrocarbons by Laser Multiphoton Ionization', *Anal. Chem.*, **70**(13), 2685–2692 (1998).
177. I. Schechter, H. Schröder, K.L. Kompa, 'New Sensor for Ecological Determination of Benzene in Ambient Air', *Anal. Chem.*, **65**(14), 1928–1931 (1993).
178. I. Schechter, H. Schröder, K.L. Kompa, 'Quantitative Laser Mass Analysis by Time Resolution of the Ion-induced Voltage in Multiphoton Ionization Processes', *Anal. Chem.*, **64**(22), 2787–2796 (1992).
179. C.J.H. Brenan, I.W. Hunter, J.M. Brenan, 'Noninvasive Confocal Raman Imaging of Immiscible Liquids in a Porous Medium', *Anal. Chem.*, **69**(1), 45–50 (1997).
180. H.R. Morris, C.C. Hoit, P. Miller, P.J. Treado, 'Liquid Crystal Tunable Filter Raman Chemical Imaging', *Appl. Spectrosc.*, **50**(6), 805–811 (1996).
181. L. Markwort, B. Kip, E. Da Silva, B. Roussel, 'Raman Imaging of Heterogeneous Polymers: A Comparison of Global Versus Point Illumination', *Appl. Spectrosc.*, **49**(10), 1411–1430 (1995).
182. M.D. Schaeberle, V.F. Kalasinsky, J.L. Luke, E.N. Lewis, I.W. Levin, P.J. Treado, 'Raman Chemical Imaging: Histopathology of Inclusions in Human Breast Tissue', *Anal. Chem.*, **68**(11), 1829–1833 (1996).
183. J. Zhao, R.L. McCreery, 'Multichannel Fourier Transform Raman Spectroscopy: Combining the Advantages of CCDs with Interferometry', *Appl. Spectrosc.*, **50**, 1209–1214 (1996).
184. J.E. Vitt, R.C. Engstrom, 'Imaging of Oxygen Evolution and Oxide Formation Using Quinine Fluorescence', *Anal. Chem.*, **69**(6), 1070–1076 (1997).
185. A.A. Panova, P. Pantano, D.R. Walt, 'In Situ Fluorescence Imaging of Localized Corrosion with a pH-sensitive Imaging Fiber', *Anal. Chem.*, **69**(8), 1635–1641 (1997).
186. S. Nomura, M. Nakao, T. Nakanishi, S. Takamatsu, K. Tomita, 'Real-time Imaging of Microscopic pH Distribution with a Two-dimensional pH-imaging Apparatus', *Anal. Chem.*, **69**(5), 977–981 (1997).
187. M. Fisher, V. Bulatov, S. Hassom, I. Schechter, 'Fast Aerosol Analysis by Fourier Transform Imaging Fluorescence Microscopy', *Anal. Chem.*, **70**(11), 2409–2414 (1998).
188. C. Sluszny, V. Bulatov, I. Schechter, 'Classification and Quantification of Polycyclic Aromatic Hydrocarbons on Quartz Sand Particles by Direct Fourier Transform Imaging Fluorescence', *Anal. Chim. Acta*, **367**(1–3), 1–10 (1998).
189. C.D. Tran, 'Principles and Analytical Applications of Acousto-optic Tunable Filters, an Overview', *Talanta*, **45**(2), 237–248 (1997).

Proton-induced X-ray Emission in Environmental Analysis

Grazia Ghermandi
Modena University, Modena, Italy

1 Introduction **2**
2 The Particle-induced X-ray Emission Technique Performed by Proton Beams **2**
2.1 X-ray Production by Means of Accelerated Protons 2
2.2 Experimental Set-up 4
2.3 Measurements of Thin and Thick Targets 6
2.4 Intrinsic Sensitivity of the Technique 7
3 The Use of Proton-induced X-ray Emission in Environmental Study **9**
3.1 Preparation of Targets from Environmental Samples 9
3.2 Coupling Target Features and Particle-induced X-ray Emission Measurement Conditions 14
3.3 Matrix Effects and Interferences 14
3.4 Calibration and Measurement Errors 15
3.5 Detection Limits and Ranges 17
4 Interpretation of Proton-induced X-ray Emission Spectra of Environmental Samples **17**
5 Method Development for Environmental Purposes **18**
5.1 Proton Microbeam Application 18
6 Comparison with Other Spectrometric Techniques **20**
6.1 Methods that use Particle Beams 20
6.2 Other Technologies 21
6.3 Combined use of Particle-induced X-ray Emission with Other Techniques 22
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **23**

Particle-induced X-ray emission (PIXE) is an analytical method based upon X-ray spectrometry. In the vast majority of applications, a proton beam is used to eject inner-shell electrons from atoms in a solid specimen (target). When the resulting vacancies are filled by outer-shell electrons, characteristic X-rays whose energies identify the particular atom are emitted.

Proton PIXE is performed by a small particle accelerator (the most widely used is the Van de Graaff machine at a 2–3 MV voltage, more recently also small tandem accelerators), that provides a beam of protons. The beam emerging from the accelerator, stabilized and directed by electrostatic and magnetic steering elements in a high vacuum line, enters the specimen chamber through suitable collimators. The X-ray detection system, housed in the chamber, is generally a Si(Li) device, that combines the advantage of high efficiency in the X-ray energy region of interest (usually 2–20 keV) with a good energy resolution.

A typical PIXE spectrum consists of characteristic X-ray peaks superimposed on a background due to bremsstrahlung radiations and nuclear reactions induced by the beam. The area of each peak is related directly to the concentration of the corresponding element in the specimen. Various software codes have been developed to deconvolute spectra and to accurately calculate peak areas.

PIXE intrinsic detection limit (DL) is not very much below 1 ppm (mg kg^{-1}) in a given matrix. It offers its maximum sensitivity in two atomic number (Z) regions: $20 < Z < 35$ and $75 < Z < 85$. Measurement errors are in the order of 10%, depending mainly on the target preparation procedure and on the slight variability of the proton flux.

The PIXE technique allows fast (few minutes), non-destructive, highly sensitive simultaneous determinations of a wide group of elements ($12 \leq Z \leq 85$), without large variations in sensitivity among different elements. The total element concentration is measured by PIXE, independent of different oxidation states. Adequate sample preparation procedures may be required in environmental PIXE applications since solid targets have to be exposed to the particle beam. Thin targets are preferable, in order to minimize matrix effects. For thin samples and for light-element matrices, PIXE has superior sensitivity compared to X-ray fluorescence (XRF). In heavy-element matrices XRF is preferable. DLs for inductively coupled plasma atomic emission spectroscopy (ICPAES) may be lower than for PIXE (depending on the sample dilution factor), but the topic of interferences is much more complex for ICPAES than for PIXE.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Highly sensitive analytical techniques and adequate procedures are of primary importance in order to perform reliable concentration measurements of trace inorganics in the environment.

The study of environmental processes requires the examination of different media, such as air, water and sediment. Among the analytical techniques that employ accelerators in the study of the environment, PIXE has proven a useful tool for the study of trace element distribution in ecosystems, both in terms of pollution measurement monitoring as well as for geochemical and geophysical studies. The major features of this technique are multi-element analysis, high sensitivity and excellent DLs across a wide range of atomic numbers. An additional advantage is the speed of analysis for a wide variety of samples. Given its main capabilities, PIXE also provides a preliminary insight into the behavior of the elements in the samples while showing the relative abundances of the elements present.

However, it must be stressed that adequate sample-preparation procedures and suitable measurement conditions are required in order to make full use of PIXE capabilities. Since the concentrations of trace elements in the examined media range from parts per million (milligrams per kilogram) to fractions of a part per trillion (nanogram per kilogram), a preconcentration treatment is generally needed since the PIXE intrinsic DL is not very much below 1 ppm in a given matrix. In addition, PIXE requires solid specimens, so that liquid samples must undergo some pretreatment before exposure to the beam, and both liquid and aerosol samples must be deposited on suitable backings. The optimum sensitivity of PIXE is obtained in the analysis of trace elements in a matrix of light elements and with thin samples of low mass. It follows that some effort is needed to prepare targets with these features using different natural specimens.

Samples of any size and shape can be irradiated using the so-called external-beam technique, where the particle beam (protons in the vast majority of applications) is taken out in the ambient air through a thin window at the end of the vacuum line of the accelerator. Obviously the sensitivity is impaired by the air layer crossed by the beam, therefore this version of PIXE is not suitable for accurate environmental analysis. On the other hand, environmental studies can benefit from the development of another variation of proton PIXE, called proton microprobe, or microparticle-induced X-ray emission (μ-PIXE), that in principle is analogous to the electron microprobe or scanning electron microscope–energy dispersive X-ray analyzer (SEM/EDAX), where a proton accelerator replaces the electron gun. The proton beam can be scanned across the surface of the specimen, giving an elemental mapping of the tested area. The μ-PIXE DLs are better by a factor of about 100 than the electron microprobe DLs, and it is the real advantage of μ-PIXE compared to SEM/EDAX, that is surely the best technique to obtain images of the specimen at the micron-size scale.

2 THE PARTICLE-INDUCED X-RAY EMISSION TECHNIQUE PERFORMED BY PROTON BEAMS

2.1 X-ray Production by Means of Accelerated Protons

PIXE is based on spectrometry of the characteristic X-rays emitted by target atoms whose innermost orbitals have been ionized by accelerated particles (Coulomb interaction). When the resulting vacancies are filled by outer-shell electrons, characteristic X-rays whose energies identify the particular atom are emitted (mainly K and L X-rays, M in a smaller amount). The vacancy production for protons at the energy used may be calculated by means of different theoretical approximations [binary encounter approximation (BEA), semi classical approximation (SCA), plane wave Born approximation (PWBA)[1] and energy (loss during collision) coulomb (field deflection) perturbation (of atomic) stationary states (relativistic) effects (ECPSSR)[2]] which all give similar results for the cross-sections of producing vacancies (i.e. shell or sub-shell ionization cross-sections). Significant K, L-shell ionization cross-sections for protons predicted by ECPSSR are shown in Figure 1. Analytical

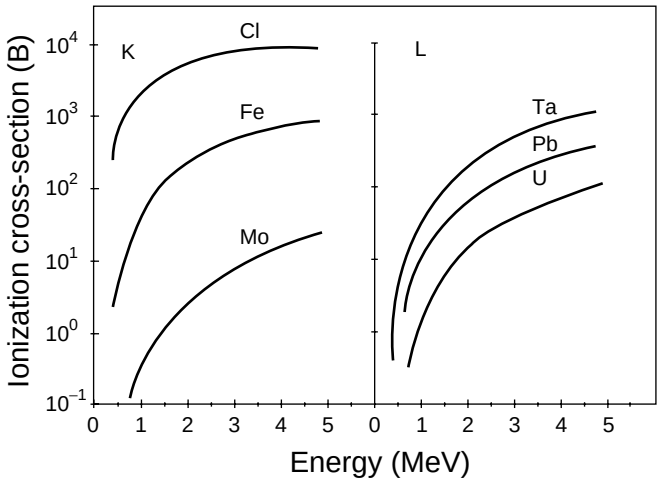

Figure 1 Significant K, L-shell ionization cross-sections for protons (barns) predicted by ECPSSR as a function of proton energy. (Reproduced by permission of Wiley & Sons from S.A.E. Johansson, J.L. Campbell, *P.I.X.E A Novel Technique for Elemental Analysis*, John Wiley & Sons, 1988.)

formulae for cross-section calculation based on experimental data[3] and on measured K- and L-shell ionization cross-sections are available in the literature. These clearly show that ionization cross-sections increase with increasing proton energy and decreasing target atomic number (with slight variations between neighboring elements).

The probability of an X-ray emission when a vacancy is filled, instead of an Auger electron, is known as the fluorescence yield[4] for the shell or subshell in question (Figure 2). The cross-section for emission of an X-ray is the product of the ionization cross-section and the fluorescence yield for a given shell or subshell. The filling of vacancies by means of transitions from outer shells, produces X-ray lines of different intensities (relative emission rates), theoretically calculated[5] and experimentally measured by several authors. The K and L or L and M lines from the same element may be simultaneously present in a PIXE spectrum. The number of emitted X-ray photons for a given atomic transition is proportional to its production cross-section for protons at the particular energy, to the relative intensity of the transition and, obviously, to the flux of protons.

The PIXE spectrum resulting from an irradiated specimen (Figure 3) is the sum of different contributions:

- the characteristic X-ray peaks (mainly K lines for medium and L lines for heavy elements);
- the continuum bremsstrahlung (Figure 4) of beam particles projectile bremsstrahlung (PB) and the secondary electron bremsstrahlung (SEB) ejected from the specimen atoms by beam particles or emitted as Auger electrons, which lose energy while decelerating into the target matrix;

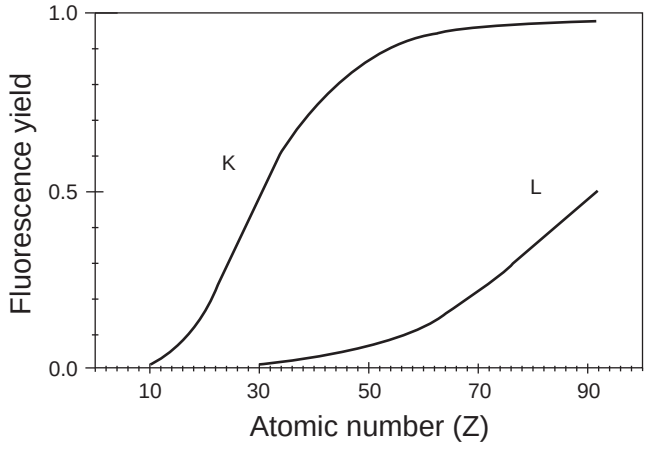

Figure 2 The fluorescence yield for the K and L shells as a function of atomic number *Z*. (Reproduced by permission of Wiley & Sons from S.A.E. Johansson, J.L. Campbell, *P.I.X.E A Novel Technique for Elemental Analysis*, John Wiley & Sons, 1988.)

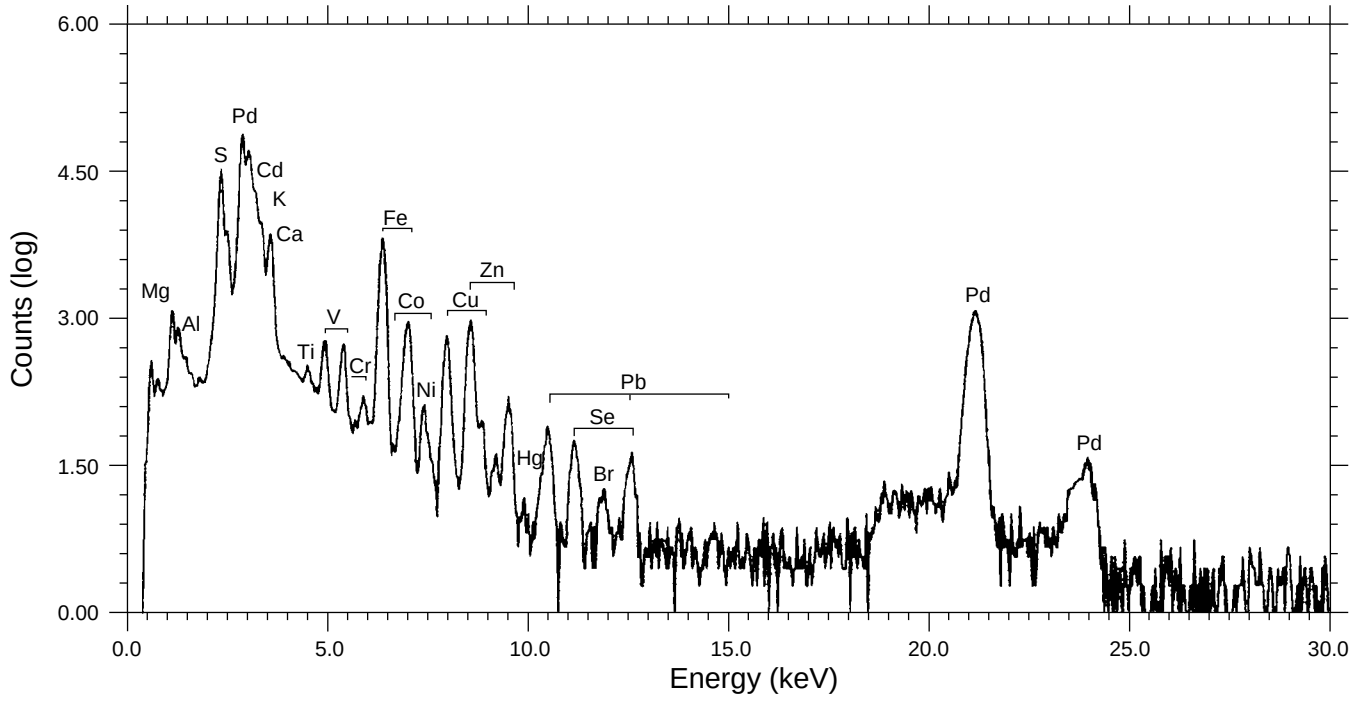

Figure 3 A proton PIXE spectrum of a 15 mL water sample with element concentrations between 100 (Fe) and 2.5 (Ni) ppb ($\mu g\,kg^{-1}$). Pd (as internal standard) and S are introduced by the target preparation procedure described in section 3.1.1. The proton energy was 1.8 MeV and the accumulated charge 10 μC.

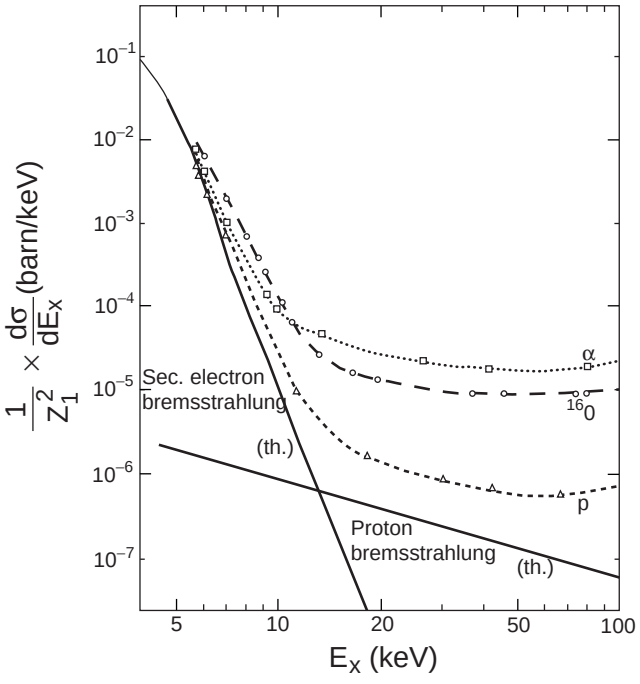

Figure 4 Continuous background from PB and SEB for three different projectiles, 3 MeV each per nucleon, incident on a thin carbon target. Data for helium and oxygen are normalized by the inverse of the square of the projectile nuclear charge Z_1 (three equal-velocity projectiles). (Reproduced by permission of Wiley & Sons from S.A.E. Johansson, J.L. Campbell, *P.I.X.E A Novel Technique for Elemental Analysis*, John Wiley & Sons, 1988.)

- the γ-rays from nuclear reactions, γ background (GB) induced by the beam.

The intensity of the proton bremsstrahlung is much less than that induced by an electron beam, because PB is proportional to the square of the charged particle deceleration (i.e. the square of the Coulombic force for particle mass unit acting on the projectiles). Given the higher proton mass, the intensity of their bremsstrahlung is negligible in comparison with the electron one. Even if of low intensity, the proton bremsstrahlung cannot disappear, because the theoretical PB cross-section contains the term (Equation 1):

$$\left(\frac{Z_p}{A_p} - \frac{Z_t}{A_t}\right)^2 \tag{1}$$

where $Z_{p,t}$ and $A_{p,t}$ are atomic number and atomic mass for projectiles and target. For low-atomic-number targets Z_t/A_t is approximately 0.5: it follows that the PB should have zero intensity for heavy projectile ions, but not for protons. The particle bremsstrahlung is responsible for the high-energy background in a PIXE spectrum, and decreases with increased beam particle energy.[6]

The SEB is the principal contribution to the low-energy continuum X-ray background in PIXE X-ray spectra. It results from the slowing down in the target of electrons ejected from atoms through inelastic collisions. The maximum kinetic energy T_m that can be transferred in a head-on collision from a projectile to a free stationary electron of mass m is given by Equation (2):

$$T_m = \left(\frac{4m}{M_p}\right) E_p \tag{2}$$

where M_p, E_p are projectile mass and energy. It follows that SEB is a continuum spectrum which can be very intense below energy T_m, and which falls off very rapidly above this photon energy.[6] Given the dependence of T_m on E_p, the SEB effect increases with rising particle energy.

SEB has an unisotropic distribution in angle; its maximum intensity is at (somewhat less than) 90° backward angle, while diminishing asymmetrically on both sides of this maximum.[7] Depending on the slope of the curve, an angle of 135° is generally preferred for X-ray detector placement.

γ-Radiation from nuclear reactions (resulting from the high cross-sections for γ-ray-producing nuclear reactions of a few of the lighter elements with $Z < 20$) interact by Compton scattering in the Si(Li) detector. This results in an essentially flat background in the 1–30 keV region of the PIXE spectrum, worsening the DLs for elements heavier than iron. This effect is dominant for $Z > 30$ when proton energies of 3 and 5 MeV are used. The yield of γ background emissions is related to nuclear resonances excitation by given protons slowing down in the target matrix: depending on the target composition, it is improved in thick targets, even if light elements (N, F, P, Na...) are present also at trace levels.[7] The intensity of GB in thin targets is not easily predictable; however, it may be very low.

Because of the dependence of the related cross-sections on the incident particle energy, in a PIXE measurement GB increases with rising proton energy; at proton energies lower than 2–3 MeV, cross-sections for such nuclear reactions decrease rapidly, at higher energies they progressively increase. Thus, significant γ-radiation background is created by thick specimens bombarded by heavy ions.

2.2 Experimental Set-up

The great majority of accelerators employed for analysis are single-ended Van de Graaff machines, employing protons as projectiles for PIXE (in a few cases helium or heavier ions are also used), and which accelerate through a voltage of typically 2–3 MV.[8] These accelerators may

provide currents up to tens of microamperes, but for PIXE the current is limited to tens of nanoamperes onto the target, also minimizing sample damage by heating and high detector dead times. There is also increased use of small machines, the tandem accelerators, that allow the use of half the voltage required by conventional accelerators to provide particles of the same energy, with considerable reduction in size and price.

The beam emerging from the accelerator is at first stabilized by passing through a slit into a deflector magnet, then driven axially in a beam line, evacuated typically to 10^{-6} Torr, by means of electrostatic equipment and steering magnets, to the irradiation chamber.[6] Before entering the chamber, the beam is diffused through a thin metal foil and then collimated in order to produce a spot on the specimens of well-defined size and with rectangular distribution of proton flux across the beam.

Various kinds of specimens may be analyzed by PIXE, both thin samples, for which the loss of particle beam energy passing the sample may be neglected, or thick samples, that completely stop the beam. Among the devices installed in the irradiation chamber, a Faraday cup is placed beyond the sample (with respect to the incident beam), which measures the transmitted particle current intensity, in order to evaluate the charge collected on the target. Obviously it is not possible to perform this measure for thick samples: in this case the current has to be monitored in other ways, either by intercepting the beam before it enters the chamber or by means of current integrators connected to the (conducting) specimen.

The X-rays emitted from the sample are detected (in most PIXE set-ups) by means of a high-resolution, high-efficiency Si(Li) detector. These detectors have an energy resolution of about 150 eV (at 5.9 keV), that permits resolution of the K lines of neighboring elements such as the transition metals. The Si(Li) efficiency varies with X-ray energy (it is close to 100% in the energy range 5 to 15 keV, and higher than 80% between 2 and 20 keV), because it is limited at low energies by X-ray absorption (in various windows and layers) and at high energies by penetration through the crystal; experimental efficiency determination is usually carried out using calibrated X-ray sources.

The resulting PIXE spectrum collected in a multichannel analyzer may be immediately passed to an on-line computer to be examined by suitable software codes.

The PIXE set-up used[9] for environmental samples at the National Institute of Nuclear Physics (INFN) Laboratory of Legnaro (Padova, Italy) is shown in Figure 5. The incident particles are protons accelerated to 1.8 MeV by a Van de Graaff accelerator. To obtain a homogeneous circular beam spot at the target, the beam is diffused through a Ni foil, of thickness 450 $\mu g\,cm^{-2}$, and subsequent carbon collimators pick off the central part of the beam, approximately 1.5 cm diameter. The Si(Li) detector (resolution: 150 eV at 5.9 KeV, solid angle between the target and the sensitive detector area: 0.0198 sr) is placed at 135° from the beam and in front

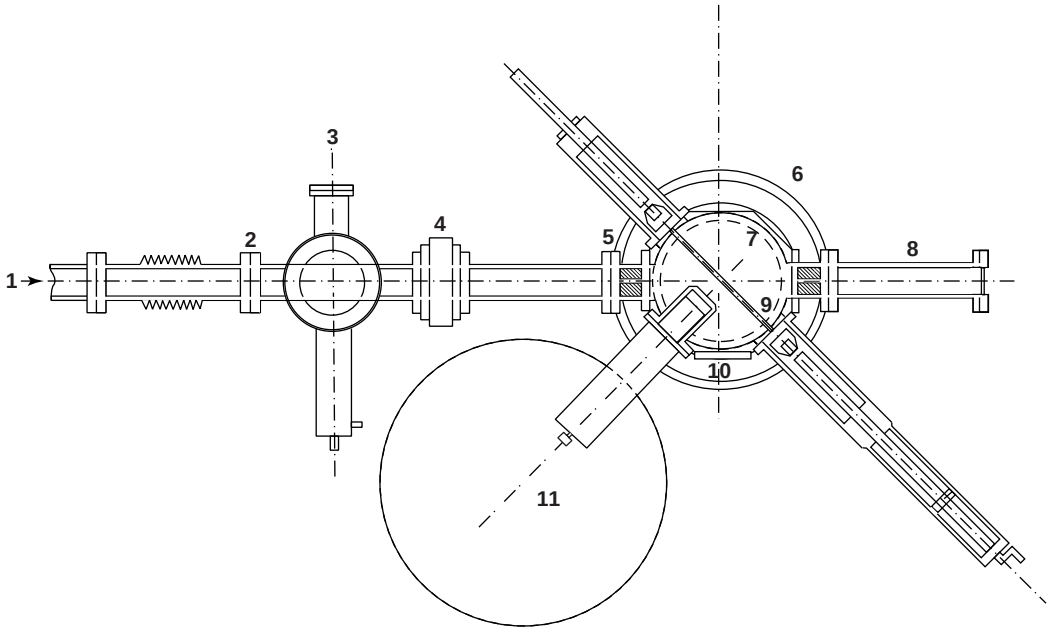

Figure 5 The PIXE set-up for environmental analysis at INFN Laboratory of Legnaro (Padova, Italy): 1, Particle beam; 2, Ni diffuser; 3, quartz; 4, valve; 5, carbon collimator; 6, irradiation chamber; 7, carbon covering; 8, Faraday cup; 9, sample holder; 10, window; 11 Si(Li) detector and electronic devices.

of the target (the detector axis is normal to the target surface). Under these conditions X-ray attenuation in the target is limited. Mylar™ absorbers of various thickness or a ''funny'' filter (an absorber with a central hole) are placed in front of the detector.

Typically a 150–300 μm Mylar™ absorber is used in sea and fresh water samples analyses, in which major light elements may be exceptionally abundant, while so-called ''funny'' filters are preferable for aerosol specimens. The aerosol specimens yield X-ray intense peaks of low-Z elements superimposed on an intense low-energy background, that would both be strongly attenuated by Mylar™ absorbers. By piercing a central hole in the absorbing filter foil, a percentage of the X-rays within the detector's solid angle (related to the hole area) is not attenuated.[7] This ''funny'' filter balances the intensity between the low- and high-energy regions of the spectrum. The use of ''critical'' absorbers is generally not required in these analyses. For pore water and mainly for sediment samples in which the relative abundance of a measured element can be particularly high, an absorber that selectively attenuates (i.e. a critical absorber) the element's most intense lines may be used. Such an absorber attenuates the X-rays that lie just above its absorption edge and transmits a greater fraction of the photons at higher (and at some lower) energies.

2.3 Measurements of Thin and Thick Targets

The energy loss of a proton beam and the X-ray attenuation are considered to be negligible in a so-called ''infinitely thin'' target. This definition obviously depends on the incident proton energies E_p, that is assumed constant within the target material, and on the nature and thickness of the exposed sample matrix, which would determine the extent, if any, of the proton energy degradation in the target. A thin environmental specimen is generally prepared by deposition of a small amount of material (such as powder, solid residue or precipitate from liquids, ultrathin slice, etc.), on a low-Z backing foil that should be as thin as possible. A conducting backing would be preferable, in order to avoid an improvement of the background by periodic discharging (to the nearest conductor) of the backing. On the other hand, the use of conducting (higher Z) backing improves PB intensity; in addition, the energies of the characteristic X-rays emitted from the backing material are sufficiently high to interfere with the peaks of the elements to be analyzed. For these reasons low-Z insulating material backings are generally used, and several solutions to their charging up are available:[6] the specimen may be coated by a conducting carbon film, or sprayed with an electron current sufficient to maintain neutrality (the risk of these procedures is the introduction of spurious trace elements).

For example, for a proton beam of 2–4 MeV energy, and with thin carbon or polymer backings, a thin target is considered to be one with a thickness of the order of less than 1 mg cm^{-2}.

The number N of characteristic X-ray photons emitted (i.e. the yield of X-rays) as a result of a transition occurring in a particular element, of atomic number Z and atomic mass A, due to the passage of N_p protons in a thin target is given by Equation (3):

$$N(Z) = N_p M(Z) \sigma_z(E_p) w_z I_z \varepsilon_z \left(\frac{\Omega}{4\pi} \quad t_z \frac{N_{av}}{A} \right. \qquad (3)$$

where $\sigma_z(E_p)$ is the ionization cross-section of the given shell/subshell for proton energy E_p; w_z and I_z, the fluorescence yield and the relative intensity for the particular transition; ε_z, the absolute detector efficiency at the considered X-ray energy; t_z, the transmission factor through any absorber interposed between the target and the detector at the considered X-ray energy; N_{av}, Avogadro's number; and Ω, the fractional solid angle subtended by the detector. $M(Z)$ is the mass of a thin, uniform elemental film per unit beam area, in the case of a beam normal to the target surface. $M(Z)$ is given by the product of target areal density (ρ_z) and target thickness, τ. Equation (3) allows absolute analysis: $M(Z)$ may be calculated if all the other terms were determined, with an accuracy of about 10%[6] for K shells, with more uncertainty for L shells. In spite of it, and even if a reliable data base and physical quantities are available, an experimental calibration by means of measurements of standard samples is frequently performed. It is the case for targets from environmental specimens that reference samples of similar matrices are more suitable for use to calibrate the set-up, as will be discussed in section 3.4.

The target may be thick enough to entirely stop the beam; the range of 2 MeV protons in a carbon foil is about 8 mg cm^{-2},[1] while the range is less than 20 mg cm^{-2} for 3 MeV protons in organic material. This is the case for specimens that cannot be reduced to a thin sample; archaeological specimens or artefacts apart, among environmental samples the mineralogical ones, that should be kept in their original state, are generally thick enough to stop the beam. Other examples are pelletized samples, mainly biological or geological. In all these cases the beam proton energy, E_p, is reduced to zero within the target, and an integration of the ionization cross-sections and of the photon transmission from energy $E = E_p$ down to zero is needed. This variant of PIXE is sometimes referred to as thick target particle-induced X-ray emission (TTPIXE) (Thick Target PIXE). The X-ray yield from a given element (Z) uniformly distributed within a thick specimen is given

by Equation (4):

$$N(Z) = N_p C_z w_z I_z \varepsilon_z \left(\frac{\Omega}{4\pi} \quad t_z \frac{N_{av}}{A} \quad \int_{E_p}^{0} \frac{\sigma_z(E) T_z(E)}{S(E)} \quad dE \right) \tag{4}$$

where C_z is the concentration of the element Z of interest. $T_z(E)$ describes the transmission of X-rays in the thick specimen, and includes the mass absorption coefficient for photons of the Zth element in the specimen matrix, obtained by summing over the matrix constituents, which should be known. $S(E)$ is the stopping power[7] of the sample matrix, obtainable from a combination of the component stopping powers. For a more detailed description, see Johansson and Campbell[6] and Johansson, Campbell and Malmqvist.[7] The other quantities have been already defined.

The PIXE capability for trace and ultratrace measurements represents a great advantage in environmental analysis: this feature may be impaired when high-intensity peaks from very abundant elements and when considerable matrix background are present in the spectrum. Thick targets frequently fall in this category. Their background emission may be further improved by charging insulating specimen materials. It follows that, even if largely used, TTPIXE finds its more relevant application mainly in biology, geology or archaeology, as previously noticed, than in the environmental field.

Specimens may be prepared, following the thin target protocol, having thicknesses up to a few milligrams per square centimeter, involving an appreciable proton energy loss within the sample. The X-ray yield in these intermediate thickness targets may be obtained from the same approach used for thick samples, with the difference being that the proton energy integration limits becomes the entrant E_p energy and the final E_f energy of the proton beam, as it emerges from the target. A knowledge of the target thickness and of its matrix composition is required in order to evaluate the matrix mass absorption coefficient for characteristic X-rays and the stopping power in the specimen. A simple scheme for the matrix corrections for these intermediate thickness specimen, that has been proposed and developed,[10] is presented in section 3.3. Simple correction factors allow one to relate the yield for intermediate thickness targets to thin target ones, without any previous knowledge of the composition of the specimen.

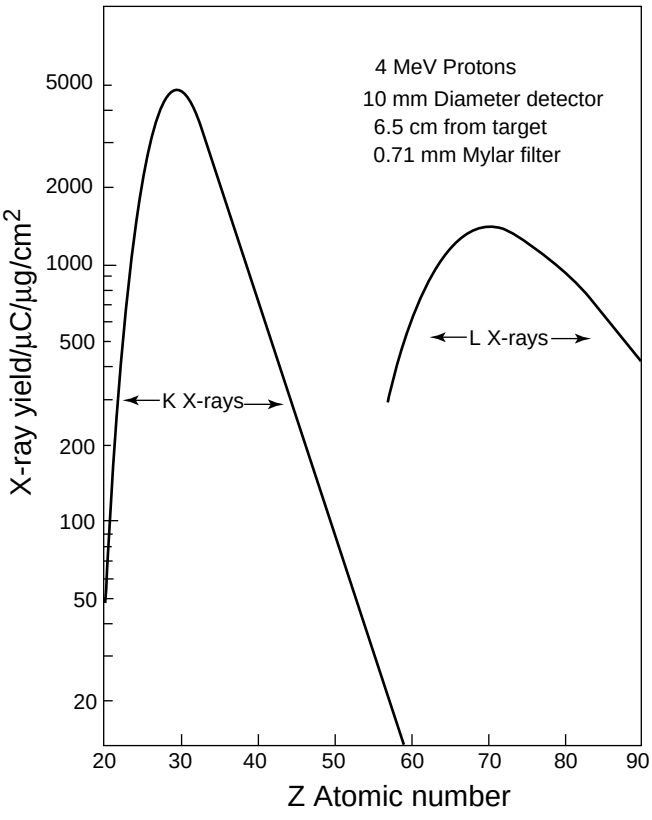

Figure 6 The sensitivity curves (i.e. the yield behavior) measured on thin specimens by means of standard Micromatter targets. (Reproduced by permission of Wiley & Sons from S.A.E. Johansson, J.L. Campbell, *P.I.X.E A Novel Technique for Elemental Analysis*, John Wiley & Sons, 1988.)

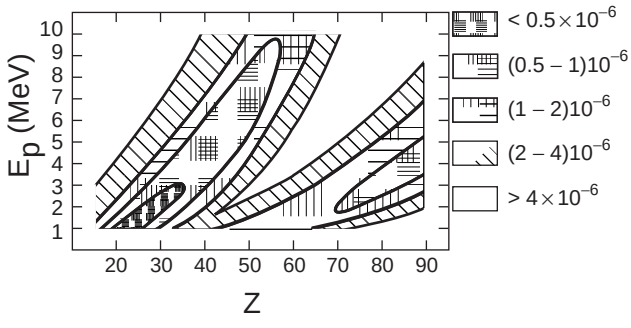

Figure 7 The LOD as a function of atomic number and bombarding energy for a thin organic specimen on a thin backing foil (0.1 mg cm^{-2}) in terms of concentration. See text for measurement parameters. (Reproduced by permission of Elsevier from S.A.E. Johansson, T.B. Johansson, *Nucl. Instrum. Methods Phys. Res.*, **137**, 473–516 (1976).)

2.4 Intrinsic Sensitivity of the Technique

PIXE analytical sensitivities are approx. 1 ppm in the exposed target material. This means that the technique is able to detect trace elements at picogram level in a suitable matrix. Figure 6 presents sensitivity curves (i.e. the yield behaviors) measured on thin specimens by means of standard targets. The region of highest yield corresponds to the limit of detection (LOD) minima. Figure 7[11] shows the limits of detection for a thin organic specimen on a thin backing foil (0.1 mg cm^{-2}) in terms of concentration, with the following experimental parameters:

detector resolution 165 eV, solid angle $0.003 \times 4\pi$, collected charge 10 μC. The highest sensitivities are obtained for $20 < Z < 40$ and $Z > 75$ and with low proton energies (1–3 MeV). Instrumental parameters and measurement features may be optimized in order to improve limits of detection. The sensitivity of the PIXE analysis is determined by the signal-to-noise ratio (S/N)[12] of the detected X-ray emissions. It follows that sensitivity is improved by increasing this ratio. The pioneer work of T.A. Cahill,[13] proposes an evaluation of the LOD variation as a function of detector, target, ion beam and measurement conditions.

As a function of the detector parameters, the LOD variation due to:

Solid angle (Ω), is $(\Omega)^{-1/2}$
Energy resolution (ΔE), is $(\Delta E)^{1/2}$
Count rate (CRM), is $(\mathrm{CRM})^{-1/2}$

As a function of target parameters, the LOD variation due to:

Target thickness for given surface density, (τ), is $(\tau)^{-1/2}$
Characteristic ionization cross-section (σ_z), is $(\sigma_z)^{-1}$
Backing cross-section (σ_{Bz}), is $(\sigma_{Bz})^{1/2}$

As a function of ion beam parameters, the LOD variation due to:

Ion Charge (Z_i), is $(Z_i)^{-1/2}$
Ion Integrated Current (Q), is $(Q)^{-1/2}$
The optimum ion velocity is about $0.1\,c$ ($c = 3 \times 10^8\ \mathrm{m\,s^{-1}}$)

Finally, the LOD variation due to measurement time (t_m) is $(t_m)^{-1/2}$. Once the main beam features are optimized, it clearly follows that the LOD (i.e. sensitivity) scales as Equation (5):

$$(\Delta E)^{1/2}(\Omega Q \tau)^{-1/2} \tag{5}$$

The estimation of the LOD, in the generally accepted formulation, is related to the standard definition of statistical significance of the signal (number of counts in a peak N) with respect to the noise (background counts N_b in the same spectral region). The definition (Equation 6), corrected for situations in which a high enough number of counts is obtained (>10), is

$$N = 3\sqrt{N_b} \tag{6}$$

Equation (6) gives a confidence level of 99.86% for normally distributed measurements. The energy interval within which N_b is determined is generally assumed to be equal to the peak fwhm (full width at half-maximum).

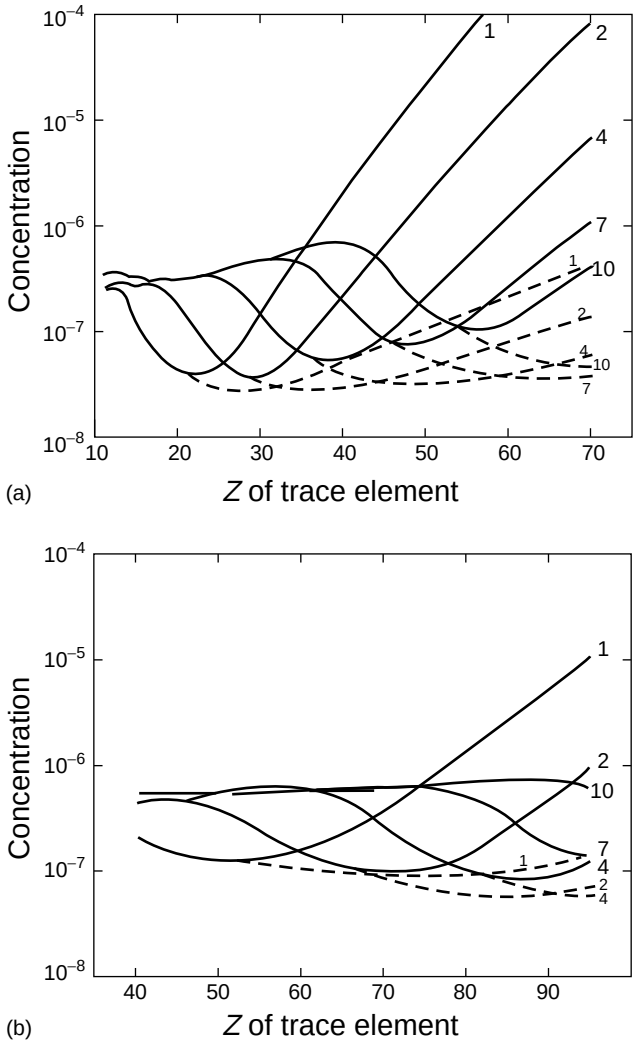

Figure 8 (a) K line and (b) L line LOD (evaluated from Equation (6) over one fwhm of the corresponding peaks) in a thin (1 mg cm^{-2}) carbon target, for 100 μC of total proton charge, at various projectile energies (indicated on each LOD curve). The dashed lines indicate the SEB contribution alone. Detector resolution 150 eV at 5.9 keV, solid angle 0.038 sr. (Reproduced by permission of Plenum Press, from F. Folkmann in: *Ion Beam Surface Layer Analysis*, eds. O. Meyer, G. Linker, O. Kappeler, Plenum Press, New York, 695–718, 1976.)

Even though Equation (6) is widely used for LOD calculations, other similar expressions are given in the literature. If the background counts N_b are determined as an average from several blank sample spectra, its variance is V_b. The detectable minimum may be defined[14] for that particular value of counts in a peak N^* such that $(N^* - N_b)$ equals its standard deviation. With this assumption, the LOD is given by $(N_b + V_b)^{1/2}$. N^* must be ≥ 10 pulses. Such an expression is suitable for environmental applications, where the measurement of several blank samples is strongly recommended.

Representative curves of DLs are given in Figure 8, chosen from among the numerous examples given in literature.

3 THE USE OF PROTON-INDUCED X-RAY EMISSION IN ENVIRONMENTAL STUDY

3.1 Preparation of Targets from Environmental Samples

The fundamental importance of adequate sampling procedures in every environmental study must be underlined. Briefly, the sample containers and equipment employed have to be made of suitable material (plastics such as high-density polyethylene (HDPE) and low-density polyethylene (LDPE), and fluorinated ethylene propylene (FEP, Teflon) are preferable). They should be cleaned following a standard procedure used for laboratory ware, then immediately sealed in plastic bags to avoid any possibility of contamination prior to sampling. The same attention needs to be devoted to the equipment used in the sampling procedure. This applies to all items of equipment: to water samplers, to filtration system (used to separate suspended and dissolved fractions in water samples by means of filter membranes), as well as the drill used for sediment and ice-core collection or any other tools used in aerosol sampling. Clean room suits should be worn during sample collection and handling. Several suggestions are given in the literature about sampling treatment at collection and about sample conservation in order to preserve the original sample features.

It must be stressed that in trace and ultratrace element analysis all laboratory facilities and procedures must be designed to eliminate any possible contamination.[15] If a so-called ''clean room'' is not available, good results may been obtained by performing all the target preparation inside a specially equipped laboratory.[12] The laboratory's main features should be: ordinary dimensions (4×4 m), floor covered with a polyvinyl chloride (PVC) sheet, walls and ceiling painted with resistant epoxy paint, access through a double door to ensure perfect closure from the outside. The air inside should be purified by means of class 100 laminar flow clean benches which are operated continuously. The commercially available laminar flow benches forces the air into the work area and in turn expels the air through conventional high-efficiency particulate air (HEPA) filters. Other benches (custom made), more adequately equipped, force air flow into the work area through more than one filter. A satisfactory system includes two conventional HEPA pre-filters and a third special HEPA filter which is 99.999% efficient at 0.3 μm. In addition the air is passed through two activated charcoal cells before being expelled into the lab. To reduce any metal contamination, this bench has been custom made out of natural polypropylene (PP) and polycarbonate (PC). PIXE target preparations are performed on this clean bench.

Throughout the lab operations high-purity water (18 MΩ cm resistivity) is used. This is produced using a Millipore Super Q ion-exchange resin unit. High-purity acids from the US National Institute of Standards and Technology (NIST) as well as high-purity salts and standard solutions from Merck (Suprapur) and Johnson Matthey GmbH, are also employed for target preparations. The assorted sample containers, vessels and lab tools that are used are made of conventional LDPE and FEP. They are either custom-made or supplied by Nalgene. Borosilicate glass (Duran) can be used for higher concentrations. Immersing these containers in a succession of heated acid baths of increasing purity for several weeks ensures cleanliness. The preliminary cleaning procedure can be performed using a Mielabor glass washer and an ultrasonic bath that operates with ultrapure water. Finally, all the personnel in the laboratory wear clean room suits and use LDPE gloves.

Extremely accurate blanks are prepared in order to allow correction for any spurious additions to the samples during both the ''in-field'' sampling operations and the target preparations. The blank specimens prepared together with the environmental samples are measured under the same experimental conditions. The average blank contribution is subtracted from the concentration measured in the samples, while variance due to the blank fluctuation can be accounted for in the error calculation.[16]

3.1.1 Water Samples

The number of studies dealing with the PIXE determination of trace elements suspended and/or dissolved in water is far smaller than those of aerosols. In great part this has been due to the greater difficulties in preparing targets, even if quantitative evaluations are aided by the use of the internal standard method. The matrix of natural waters may pose problems in the pre-analysis treatment of samples which, especially in the case of ocean water, contain several elements of interest at very low concentrations. Techniques of preconcentration are therefore necessary. In addition, as previously mentioned, solid targets have to be prepared for PIXE analysis, therefore aqueous samples require pretreatment. The preconcentration procedure has to be chosen taking into account the different types of samples: in fresh and seawater the concentrations of the so-called major lighter elements are generally higher than the heavy elements, but in ground and pore water the concentration ratios may be completely different. Rain water, melted snow and ice

behave (generally) as dilute solutions compared to the other natural waters.

In spite of the advantages of using thin sample backings, the preconcentration of natural water may also be performed on thick backings, since these are easier to handle. The sample preparation may be performed either by spraying large sample volumes (up to 1000 mL) or by direct evaporation of smaller samples (10–30 mL) onto thick carbon backings (of the order of 0.3 g cm^{-2}). The spraying procedure requires quite large sample volumes, resulting in better DLs of the order of 0.1 ppm of the original specimen. Sample evaporation at a suitably low temperature is certainly preferable because it allows a very high preconcentration without introducing any chemical reagent into the sample, which enables lower DLs. However, problems arise if the sample contains some nonvolatile elements in exceptional amounts, as, for example, is often the case with major cations in surface waters. The presence of a superabundant element on the target always hinders a PIXE measurement as well as affecting the sample preparation because visible crystallization can take place on the backing. Such surface inhomogeneities result in complex matrix effects which are difficult to evaluate. In addition, even if the water deposit is mainly on the target backing surface, a minor part can penetrate into the backing involving target inhomogeneities difficult to evaluate. The sensitivity is impaired by an approximate estimation of matrix effects, with DLs of the order of 0.01 ppm of the original sample.

Methods that use water spraying and water vaporization to prepare thin targets are described in the literature. The water sample[17] to be analyzed is induced to evaporate when placed in a container separated from a vacuum chamber by a thin filter membrane. At complete evaporation the resulting deposit on the membrane may be measured by PIXE, with LOD ranges from 0.1 to 50 ppb (μg kg^{-1}). In another set-up,[18] the water is sprayed into an air system where it forms a fine aerosol, collected by means of an impactor on thin polystyrene film (that will be the target for PIXE). The resulting LOD ranges from 0.1 to 1 ppb. Thin target preparation by means of specimen evaporation has been performed by Laj et al.[19,20] The procedure employs nonboiling evaporation (lower than 60–70 °C) on liquid samples ranging from 30 to 100 mL. Once preconcentrated to about 100–200 μL, 10 μL of Pd or Y (1000 ppm) solution can be added as the internal standard. The sample is then spotted onto a thin backing and evaporated until dryness. Experiments performed on such samples[21] have shown that losses by evaporation are very limited for nonvolatile chemical species. The recovery efficiency of heavy volatile elements is certainly impaired.

In order to obtain a background as low as possible we initially adopted laboratory-made polyvinylacetate (PVA) backings.[22] However its low resistance to attack by acid solutions forced us to select thicker but more resistant backing materials. Another drawback of the PVA backings was the need to remove impurities from them (typically 10 ng cm^{-2} of Cu, Ni, Zn and a tenth of ng cm^{-2} of Fe and Br), that is quite difficult to achieve without using acids. Acid pre-cleaned PC foils were adopted (the Bayer AG type 1 mg cm^{-2}, specification which matches Nuclepore filters). The surface density of samples prepared following this procedure from ice and firn specimens,[20] which had been previously melted at laboratory temperatures, resulted typically in 0.3 mg cm^{-2} thicknesses (backing excluded), which is at the limits of the thin targets. In Table 1 are reported mean blanks, LOD and analytical uncertainties from samples excited with a 1.8 MeV proton beam, a collected charge of 15 μC, and with a 50 μm Mylar™ ''funny'' filter (hole diameter: 7.06% of the total area).[20]

Table 1 Water samples evaporated onto thin PC backings

Element	Mean blank (ng cm^{-2})	LOD (ng cm^{-2})	Precision (%)
Al	12	10	10
Si	50	10	10
S	20	5	10
K	7	2	5
Ca	9	2	5
Fe	4	2	10

Mean blanks, LODs and analytical uncertainties (i.e. percent standard deviation of measurement reproducibility) measured with 1.8 MeV proton beam, collected charge 15 μC, and with a 50 μm Mylar™ 'funny' filter. (Reproduced by permission of American Geophysical Union from P. Laj, G. Ghermandi, R. Cecchi, V. Maggi, C. Riontino, Sungmin Hong, J.P. Candelone, C. Boutron, *J. Geoph. Res.*, **102**, C12, Nov. 30, 26 615–26 623 (1997).)

Given the aforementioned problems arising from the high concentration of major elements in natural water, it is necessary to employ selective preconcentration. This also reduces the intensity of background radiation. For this purpose the use of suitable chemical preconcentration procedures and the resulting thin targets are preferred.

There are many methods of preconcentrating aqueous samples and many of them can be used in conjunction with PIXE and XRF.[23] Many treatments use chelating reagents to extract trace metals. The complexes resulting can be extracted into liquid organic phases (several compounds are used for this purpose,[24] a suitable one being ammonium pyrrolidine dithiocarbamate (APDC), as will be discussed in what follows) or absorbed on activated carbon;[25] subsequently, the activated carbon can be removed by filtration and then pelletized.[26] In another application,[27] metal chelates (adding 5 mg of APDC to 500 mL of sample) are adsorbed onto 5 mg of activated carbon, then separated from the water by decanting and

centrifugation. The carbon follows acid digestion (boiling with 35 μL of 8N nitric acid) and finally is spotted onto a thin plastic (PC or polystyrene) backing foil.

The complexes can also be coprecipitated together with a collector element on thin membrane filters. A methodology that involves minimal sample handling is based on the development of a procedure[28] which performs extraction of (trace) elements as carbamates. The technique is based on the high stability of the carbamates of several elements and on their low solubility in water which is mainly determined by pH: the solubility of the carbamate complexes decreases as the pH of the solution increases. When pH values reach a suitable range, from 10% sulfuric acid up to 9, a large number of carbamate complexes precipitate and can be collected by filtration on a thin backing using a coprecipitant agent.

This procedure is suitable for natural waters in which the trace elements are generally already in their more stable state, namely the one complexed by the carbamate chelant. Acidification of the sample at the point of collection or during the sample pretreatment (for pore water or for any aqueous fraction filtered in the field) predisposes most metals to appear in their reduced form as ions. The acidification also breaks chemical bonds of the majority of organic compounds. Obviously this procedure is not recommended for aqueous liquid waste samples which generally require prior treatment to convert the elements into a form readily suitable for analysis. In such cases, other complexing agents like ethylenediaminetetraacetic acid (EDTA) may be discharged in significant concentrations from farms and industrial plants and could result in possible reduction of carbamate extraction efficiency. The use of chelating reagent and coprecipitant on the other hand creates an additional ''chemical background'' in the sample with characteristic emissions at low-energy. This partially reduces the advantages gained from using thin targets as well as affecting the detection of light elements (as is the case with snow and ice samples).

The largest number of elements, 41, precipitate at pH 9 with sodium diethyldithiocarbamate (NaDDTC) as chelant mainly as carbamates but also as carbamates and hydroxides or as hydroxides alone, as a result of the coprecipitation effect.[28] This represents an advantage when very dilute solutions (under parts per billion (micrograms per kilogram) level) are treated, as is the case of open-ocean water. In addition the carbamate hydroxide precipitate acts as a 'scavenger' for constituents in solution. However it becomes a problem with fresh water, because, even if NaDDTC does not form chelates with alkaline earth metals, at pH 9 some major cations, like Ca, precipitate following the equilibria of their other salts. This phenomenon strongly affects the features (thickness, uniformity...) of the resulting target. For this reason, even if satisfactory results can be obtained with precipitation at pH 9 with NaDDTC,[14,29] modified target preparation procedures were performed at pH 4 and which made use of the synergistic effect of two chelating agents, NaDDTC and APDC, which also does not chelate the alkaline earths.

Palladium is used as a coprecipitant agent for quantitative carbamate recovery. It can also acts as the internal PIXE standard and is favored for this role owing to the high stability of its carbamate complexes and for its rarity in most environmental specimens. In addition, its K and L lines do not interfere with the detection of the other elements of interest. In the case of Cd and Ag (whose emissions interfere with Pd lines), Cu is used as the internal standard.

Co-precipitation is carried out as follows: 100 μg of Pd (from a 1000-ppm Pd solution in 20% HCl) is added to the sample. Ammonia is added to keep the pH at a value of 4. The carbamate elements are precipitated by adding 1 mL of a freshly prepared NaDDTC APDC (1% q/vol concentration of each) solution which has been previously purified by shaking it with an equal volume of 1,1,2-trichloro-2,2,1-trifluoroethane (Freon). The sample is allowed to stand for a variable period of time, depending on the treated volume (from 30 min to 12 h) and is then filtered through a Nuclepore PC membrane to provide a sample collection area of 0.79 cm^2. The specifications of the Nuclepore filters are: 0.4 μm pore size, 10 μm thickness, 1 mg cm^{-2} surface density and 10^8 pore cm^{-2}. The choice of the 0.4 μm as pore size (among the various sizes available) is based both upon filtration efficiency (which has been evaluated to be 99% when using a double filtration technique) and the time needed for sample filtration. Moreover if unfiltered natural water is processed the 0.4 μm pore size ensures good collection of suspended particles (assumed to be greater than 0.45 μm). The collected precipitate forms an apparently homogeneous layer on the filter. Every possible inhomogeneity can be integrated by performing PIXE measurements with a uniformly diffused beam (through a Ni 450 μg cm^{-2} foil in the set-up described in Figure 5) collimated to produce a beam size larger than the filtration area on the target.

This procedure has been widely tested and calibrated (see section 3.4) for Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Se, Hg, Pb, Bi, Mo, Y, Tl, Sb, Ag and Cd. The recovery efficiency as evaluated[29] and reported in Table 2, remains generally higher than 80% for Fe, Co, Ni, Cu, Zn, Se, Hg, Pb, Ag, Cd, higher than 50% for Ti, V, Cr, Mo, Sb, but lower than 25% for Y and Tl.

3.1.2 Sediment Samples

Sediment and soil solid samples may be analyzed as such, or reduced to slices (if possible), powders or solutions.

Table 2 Recovery efficiency (in percent) for several elements precipitated as carbamates at different pH

Element	Recovery (pH 4) (%)	Recovery (pH 5.5) (%)	Recovery (pH 9) (%)
Ti	64.2	74.5	81.3
V	70.7	77.1	
Cr	55.2	77.9	82.8
Mn			69.7
Fe	90.2	90.3	90.6
Co	91.5	99.1	94.4
Ni	89.0	94.0	93.7
Cu	95.3	98.0	97.5
Zn	81.3	84.0	88.4
Se	87.7	80.0	
Y	23.9		62.5
Mo	56.1		
Ag			86.7
Cd			91.3
Sb	66.7		
Hg	96.5	94.5	96.6
Tl	8.4		
Pb	87.0	85.2	94.7
Bi	87.8	80.1	97.4

Reproduced (partly modified) by permission of Elsevier from R. Cecchi, G. Ghermandi, R. Zonta, *Nucl. Instrum. Methods Phys. Res. B*, **49**, 283–287 (1990).

Since most samples are not inherently homogeneous, problems may arise in measuring them in the ''as-received'' state to achieve a representative bulk analysis. The specimens might be desiccated and then pulverized into fine grains: sieves made of silk bolting clothes or nylon which have low metal content are suitable for sieving ground-soil materials. The addition of an internal standard to the sample is then also feasible with a powdered and sieved sample material.

The powdered samples may undergo either a nondestructive or a destructive preparation. In the first case,[30] a target from soil powdered to micromillimetre fineness is prepared by scattering a few milligrams of soil on a 4 μm Mylar™ film and using a drop of 0.1 mL of polyvinyl acetate, diluted with acetone, as adhesive. The film of soil results in a homogeneous and adequately thin specimen to meet the thin target thickness limits. In PIXE analysis with 3.5 MeV protons, an Al absorber 50 μm thickness was used to decrease the low-energy background. The most widely used nondestructive methods involve pellet preparation from grained specimens: obviously then, thick targets are prepared. As an example, a soil sample mixed with high-purity graphite into 10 : 3 soil–graphite ratio was compressed into a disc shape with 0.5 mm thickness and 10 mm diameter. Mylar™ absorbers have to be used in front of the detector in order to reduce the pile-up effect of low-energy X-rays due to the high concentrations of light elements prevalent in soils. In another study,[31] a dried and accurately ground and sieved sediment sample (about 5 g) was pressed in a 5 ton press, and pellets of 1 cm diameter and approximately 2 mm thickness were obtained. The elemental concentrations measured ranged from a few ppm to a few percent when the measurement sensitivity was of the order of the PIXE intrinsic LOD.

In the destructive methods, the powdered specimens are solubilized using commonly accepted procedures (not especially performed for PIXE), employing reagents that should be ultrapure grade. The extraction of 10 g of soil may be performed with 50% HCl and 50% HNO_3, then diluted to 100 mL. For PIXE analysis, an aliquot of 0.4 mL of the extract can be dried onto a 4 μm Mylar™ foil.[32] Another procedure, especially suited to extract the non-lattice-held fraction from sediment samples[33] involves the digestion of 5 g of desiccated and ground sediment with 100 mL of 1N solution HCl and agitation for 12 h. The resulting liquid (sediment extract) specimens are preconcentrated as NaDDTC and APDC carbamate on Nuclepore filters, using the procedure previously described for waters. The results are comparable with those of atomic absorption spectroscopy (AAS) analysis. It has to be underlined that the number of studies dealing with PIXE determinations of sediment composition is very small. The Proceedings of the 7th PIXE International Conference contain only one paper on this topic. In fact, even if satisfactorily applied, PIXE cannot completely demonstrate its capabilities in this field. The matrix background, the low-energy X-ray tails and the pile-up pulses worsen the spectrum on each side of the most intense characteristic emissions, and these interferences may result in large inaccuracies in the measurements.[12]

3.1.3 Atmospheric Aerosol Samples

Atmospheric aerosols may be considered to be any liquid matter or solid particles suspended in the atmosphere, and which represent a small fraction, generally below 0.1 part per million by mass, of the total atmosphere.

The study of atmospheric aerosols requires the analysis both of composition of wet and dry fractions and also of particle-size measurements. Concerning the latter, several studies have been performed and protocols optimized on aerosol sampling techniques, that have generally found that particle collection and size-fractionation by impaction, is the most satisfactory technique. Investigations of aerosol composition present considerable difficulties for standard chemical methods since it is very difficult to collect enough mass in each particle size range for analysis, and, also, the concentrations are very low in the liquid fraction.

Among the methods for aerosol analysis based on the use of accelerated particles, the most frequently used is PIXE which has been shown to be a very suitable tool

for the direct analysis of bulk aerosol specimens collected on filters. This is surely the environmental field in which PIXE has found its largest successful application. That results from PIXE's sensitivity, and also on the simplicity of preparing aerosol targets. In fact, the membranes used in air samplers may be very suitable for immediate PIXE analysis, without further handling: once suitable thin aerosol backings have been selected, and provided an adequate air sampler is used, few problems can arise in preparing aerosol targets for PIXE.

The commercially available air samplers, built to conform to the protocols certified by governments (such as the PM10 specification) in order to check air pollution, can collect a large area aerosol deposit (typically 47 mm diameter) on the backing. Given that a PIXE set-up does not provide so large a beam, the PIXE analysis is obviously limited to small fractions of the collected specimen. Even if more than one target may be obtained from the same specimen, its surface inhomogeneities are not negligible; on the contrary, the use of defocused proton beams larger than the deposit would allow the integration of the whole target surface. It follows that specially designed air samplers can be adapted to produce aerosol deposits of size suitable for PIXE analysis.

The most widely used air samplers are mentioned below; a more detailed review of these devices may be found in Johansson, Campbell and Malmqvist.[7] These samplers collect aerosol particles separated into two or more size classes: the aerosol deposit may be obtained by filtration or by inertial impaction.

The dichotomous aerosol samplers yield two size fractions: among them, the ''Gent'' SFU (stacked filter unit)[34] (with a suitable inlet that provide a 50% cut-off diameter d_p of 10 μm) separates aerosol by differential filtration into a coarse ($2.5 < d_p < 10$ μm) particle fraction and a fine particle ($d_p < 2.5$ μm) collection onto 47 mm diameter membranes; respectively, 8 μm median pore size and 0.4 μm pore size Nuclepore filters are used in most applications. The unit's air flow rate is typically 17 L min^{-1}. The interagency monitoring of protected visual environment (IMPROVE) dichotomous samplers includes a cyclone preseparator that eliminates the coarse particles.[7]

Among the cascade impactors, the unit most commonly used for PIXE is the model I-1 PIXE International Cascade Impactor (PCI).[35] This is a single-orifice impactor with seven plus two (option) impaction stages and a back-up filter stage. The collection surfaces of the various stages are mounted on 25 mm diameter PC rings, and the loading and unloading of these substrate rings in each stage of the PCI is fairly straightforward. The aerosol particles are collected in each impaction stage as a single deposit, which can be easily enveloped by a narrow proton beam (8 mm diameter). The particle size cut-point of the last impaction stage is for 0.25-μm particle equivalent aerodynamic diameter (EAD). Those of extra stages are 0.12 and 0.06 μm EAD. The PCI sampler's air flow rate is 1 L min^{-1}, which is somewhat lower than other impactors.

A rotating version of the micro-orifice uniform deposit impactor (MOUDI)[36] has stage cut-off limits down to about 0.05 μm EAD and operates at 25 to 30 L min^{-1} flow rates. The particles are collected on each stage over a large ring area, so that, as in the case of SFU, only a small portion (<10% of deposit) can be covered by the proton beam during PIXE measurements, thereby decreasing any advantage of the higher flow rate. The Davis, rotating-drum, universal size cut, monitoring (DRUM) sampler has eight slowly rotating drums giving air particle-size fractions of cuts from 0.088 to 9.6 μm EAD. Its vaseline-coated Mylar™ substrate moves at 2 mm per 8 hours. It can operate up to 28 days without attention. Its flow rate is 1 L min^{-1}. The DRUM[37] aerosol sample specimen that results from prolonged aerosol sampling is a strip that, for immediate PIXE analysis and best time resolution, should be mounted on a movable frame which can be advanced, on command, two millimeters per step for proton irradiation of each increment.

The small deposit area impactor[38] (SDI) has been designed to collect size-fractionated samples for subsequent PIXE analysis. The SDI is a 12-stage, multinozzle device, with a deposit area with diameter less than 8 mm and accepts the same 25 mm diameter substrate rings as the PCI. The air flow rate is 11 L min^{-1}. The experimental aerodynamic cut-off diameters for the 12-stage SDI are given in Table 3.

Table 3 Characteristics of jet dimensions, operating pressure and calculated and experimental aerodynamic cut diameter (d_{50} values) for the 12 stage SDI

Stage	Number of jets	Jet diameter (mm)	p_s (hPa)	d_{50} (calc.) (μm)	d_{50} (exp.) (μm)
12	1	7.80	1013	8.55	8.50
11	3	3.55	1013	4.46	4.08
10	8	1.80	1012	2.69	2.68
9	14	1.10	1010	1.68	1.66
8	14	0.85	1004	1.07	1.06
7	10	0.80	990	0.809	0.796
6	17	0.55	968	0.577	0.591
5	32	0.35	925	0.364	0.343
4	27	0.30	811	0.223	0.231
3	22	0.30	574	0.145	0.153
2	29	0.30	305	0.085	0.086
1	53	0.30	135	0.044	0.045

Data for $p = 1013$ hPa, $t = 23$ °C. p_s is the absolute pressure downstream of the stage. (Reproduced (partly modified) by permission of Elsevier from W. Maenhaut, R. Hillamo, T. Makela, J.-L. Jaffrezo, M.H. Bergin, C.I. Davidson, *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 482–487 (1996).)

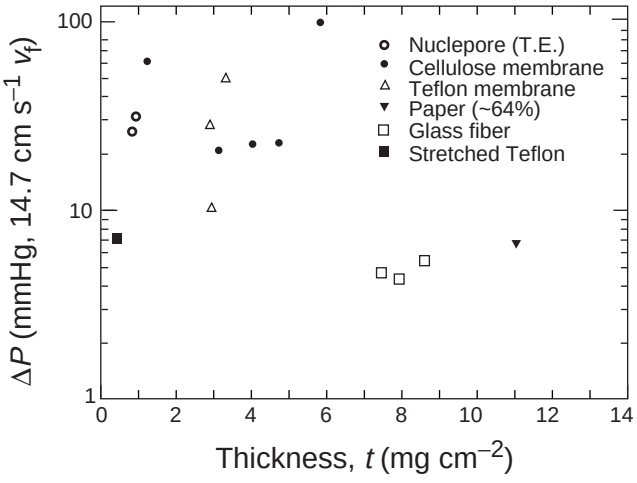

Figure 9 The pressure variation (Δp) for a fixed flux velocity (v_f) versus target thickness (t) for commonly used filter media. (Reproduced by permission of Wiley & Sons from S.A.E. Johansson, J.L. Campbell, K.L. Malmqvist, *Particle-induced X-ray Emission (P.I.X.E)*, John Wiley & Sons, Inc. 1995.)

Various types of thin film can be used as aerosol collection substrates, whereas the samplers that operate by filtration employ various porous filter media. The parameters of commonly used filters are listed in Figure 9.[7] PCI and SDI employ a 1.5 mm Kimfol PC film as collection substrate. In order to reduce particle bounce-off effects, the film is coated with vaseline (especially the larger diameter stages) or paraffin (smaller diameter stages), or apiezon L. The targets from all SDI impaction stages, when irradiated by means of 2.4 MeV proton beams (beam current typically 150 nA, total collected charge 100 μC, funny filter absorber) allow the detection of 22 elements ranging from Na to Pb.

Even if aerosol samples are generally considered infinitely thin, this assumption has to be validated. In fact the inhomogeneous accumulation of particles on the backing is possible, at least in the larger diameter stages, especially with a small deposit area. Given the small quantity of collected material, the target thickness is not significantly improved, but the target handling becomes more critical, with possible loss of particulate from the deposit surface. It may impair the determination of the mass collected, and the subsequent quantitative analysis.

3.2 Coupling Target Features and Particle-induced X-ray Emission Measurement Conditions

The sensitivity of PIXE analysis is determined (among other factors) by the S/N of the detected X-ray emission. The PIXE sensitivity is improved by increasing this ratio. In terms of the mass and the energy of accelerated particles, the sensitivity is increased by higher X-ray production cross-sections and low background but both the X-ray production cross-sections and the γ and SEB background increase with increased particle energy, while the PB decreases. Heavier ion beams do not produce significant PB but the resulting GB is more than ten times greater than for protons. The use of a 1–3-MeV beam of protons limits the γ-radiation while the corresponding X-ray production cross-section values are acceptable. SEB radiation may be also reduced taking into account its angular dependence. The background effects can also be minimized by using low Z and thin targets (of the order of 1 mg cm^{-2}).

Concerning measurements of the thin targets, it must be stressed that, even if the number of counts in a peak (6), and the related estimation of the LOD, is improved by increasing Q, which may be achieved by operating at higher beam currents, limitations are usually set either by the extent of target damage or by the maximum count-rate that can be handled. Target damage can be decreased, but not avoided, by increasing the target irradiation area and uniformly diffusing the beam to cover this area. Damage can also be decreased by increasing Ω so that, for a fixed CRM, the beam current is minimized. Since the CRM is proportional to the beam current, the peak signal (and subsequently the LOD) is improved when the maximum CRM capabilities of the system are increased. The increasing of the CRM, given the acquisition time for high-resolution detector systems, is limited by increased dead time (and possible pulse pile-up events), relative to the number of emitted X-rays (i.e. to the target composition). The time required for the PIXE analysis of a target from a typical environmental sample (with the set-up of Figure 5) generally ranges from 10 to 30 min. A 10–30-μC integrated beam charge is used, with 10–40-nA current on the target. The majority of the emissions resulting from the abundant light elements and/or the chemicals used for preconcentration fall in the lower energy part of the X-ray spectrum, and can be attenuated by use of suitable Mylar™ absorbers. If the count-rate does not significantly exceed 1000 counts per second, the detector system operates with dead time less than 5% and with negligible pile-up effect.

The other means to increase Q is by increasing the irradiation time, but that results in more expensive measurements.

3.3 Matrix Effects and Interferences

When using non-infinitely-thin targets, the evaluation of matrix effects is necessary. Correction factors for X-ray attenuation, for proton energy degradation and for secondary X-ray production (by primary photon absorption) in the samples have been evaluated[12] separately. The yield of X-rays for thin targets may

be assumed to be approximately equal to the product of the yield for intermediate thickness targets and these correction factors.[10,6] Calculations have been performed for targets with surface densities ranging from 0.5 to 2.5 mg cm^{-2} prepared from laboratory solutions containing all the calibration elements, from Ti to Bi, in variable concentrations (parts per million to parts per billion), and on environmental samples (precipitated for targets as carbamates at pH 4 with palladium and deposited on Nuclepore filters).[12] Surface density here refers to the precipitate deposit alone; the backing thickness is excluded. The calculations have also taken into account any accumulation of the precipitate in the Nuclepore filter holes. The effects of secondary X-ray production[39] was found to be negligible. The ratio between secondary and primary radiation was of the order of 10^{-4} in the case of the sample composition tested, and thus may be neglected.

The matrix correction for each Z element [MC(Z)] is then limited to the effects of other two phenomena, and is evaluated[10] as product of a proton energy slowing down correction and an X-ray attenuation correction, separately, as Equation (7):

$$\mathrm{MC}(Z) = \frac{\sigma(E_\mathrm{p})}{\sigma(E_\mathrm{m})} \left(\frac{\mu}{\rho} \quad \frac{\rho d/\cos\zeta}{1 - \exp(-\mu/\rho\, \rho d/\cos\zeta)} \right. \qquad (7)$$

where $\sigma(E_\mathrm{p})$ is the cross-section at the particle incident energy E_p, $\sigma(E_\mathrm{m})$ is the cross-section at the particle energy E_m that is assumed as ''average'' energy of X-ray production within the target, evaluated empirically as: $E_\mathrm{m} = E_\mathrm{p} - \Delta E/2$. The energy loss, ΔE, within the target layer (as presented to the beam) is calculated from the specimen stopping power, by means of published equations,[40] assuming a linear behavior of energy loss as a function of target thickness, for these thin samples.

The quantity μ/ρ is the attenuation coefficient for the characteristic X-rays in the matrix calculated from Thiesen,[41] ρd is the surface density (g cm^{-2}) of a target of thickness d and density ρ, ζ is the angle of X-ray emission with respect to the target axis (i.e. $\cos 90^\circ = 1$ using the geometry of Figure 5).

In order to calculate the stopping power and attenuation coefficient it is necessary to first establish the matrix composition. When unknown samples are treated, this calculation is performed as follows: the medium–high-Z constituents are measured by PIXE without matrix correction. The main low-Z components arising from the carbamate chelants can be determined indirectly knowing the bonding ratio with chelated atoms. If chemical treatments were not performed, all the target components would be at first evaluated by PIXE without matrix correction. Using an internal standard S (Pd in the experiment), the correction factor applied is the ratio MC(Z)/MC(S) between matrix correction for the element Z and the internal standard S, thus considerably reducing the weight of the correction factor.[42]

The multielemental PIXE spectrum having a large number of X-ray emissions involves unavoidable line interferences owing to the limited resolving power of the Si(Li) detector. This problem is adequately solved fortunately by suitable computer codes that include a complete database of all X-ray production phenomena and also optimized experimental values. The interference resolution in PIXE spectra is not, in general, a very critical matter. The more frequently encountered interferences are between the K_α lines of an element Z and the K_β lines of element $(Z-1)$ in the transition element region. For higher-Z elements, the energy differences are large enough to permit the detectors to separate interfering lines. For $Z \geq 35$ interferences occur between L-lines from heavy elements and K-lines from lighter ones. For Z around 80, interferences can also occur between M-lines from these elements and K from the lightest detected elements.

Given that the relative K-line intensity from one element is constant, the use of different exciting particles cannot reduce the secondary emission (i.e. the interfering line effects are unaltered). In general no more than 20 elements are simultaneously present in a PIXE spectrum, and this limits the interference problems.

The quantitative treatment of line interferences will be briefly discussed in section 3.4.

3.4 Calibration and Measurement Errors

The calibration of PIXE equipment consists of the determination of the X-ray yields for each element of interest: its accuracy depends on a precise knowledge of the basic physical quantities involved and of the experimental parameters. In fact Equation (3) lends itself to absolute analysis; nevertheless many analysts prefer to perform calibration measurements using external standard samples or with the internal standard method. In the case of thin samples, the sensitivity may be determined by measuring the K_α or L_α X-ray yields (as a function of Z) per microcoulomb of proton charge per unit areal density, employing thin, uniform standard targets, and a particle beam normal to the target surface (Equation 3). These external standard specimens are commercially available, the most widely used being from the MicroMatter Company. They consist of very thin, certified, layers (20–50 μg cm^{-2}) of various elements evaporated onto Mylar™ or Nuclepore membrane. The accuracy of the estimation of the sensitivity curve with a large number of points has been found[6] to be better than 10%. Loss of volatile elements (Hg, As, Br) may occur during proton bombardment, but was found to be negligible[11] (apart from Br) with beams of tens of nanoamperes or less.

Several laboratories of applied nuclear physics have prepared their own thin external standards by means of element evaporation or similar techniques. The reliabilities of these standard specimens are in many cases of the highest level. Calibration measurements[12,20] have been performed with satisfactory results using standard thin targets prepared at the INFN Laboratory at Legnaro, Padova (Italy). These were prepared by vaporization of elements onto PC backings. The calibrated elements are: Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Co, Cu, Ni and Zn. An alternative approach consists of the use of standard solutions pipetted in micro-drops and left to dry onto a thin backing (Mylar™, Nuclepore...). The deposit has to be completely covered by the proton beam, whose cross-sectional area must be also measured. There are advantages to choosing the elements introduced in the target and to deposit them on the same backing used for environmental samples. Problems may arise if dried solutions form a nonuniform deposit with clumps of crystals on the backing. Given that the standard specimens must be as similar as possible to the features of the environmental samples, it follows that this procedure is preferable for aerosol analysis.

The use of internal standards is surely to be preferred when the environmental specimen must be chemically treated to obtain a suitable PIXE target which is the case for water samples. In fact after addition the yield $N(S)$ of the standard element S will be subjected to all the same losses as the sample followed during the target preparation procedure. The yield $N(Z)$ of any other elements can then be experimentally determined with respect to recovery of the internal standard.

From Equation (3) it follows that when the internal standard is added, the determination of the ratio $N(Z)/N(S)$ in a given target does not require the knowledge of some physical quantities: N_p, Ω. The mass ratio $M(Z)/M(S)$ may be considered equal to the ratio $C(Z)R(Z)/C(S)R(S)$, where C is the concentration of element Z and the standard in the original water samples and R the percentage recovery of the internal standard. The statistical processing[9] of several calibration measurements made under the same experimental conditions leads to the evaluation of a parameter $K(Z, S)$ that relates the ratio between number of X-ray counts $N_C(Z)$ in the peak of an element Z and the number of X-ray counts $N_C(S)$ in the standard peak to the corresponding concentration ratio $C(Z)/C(S)$ (Equation 8)

$$\frac{C(Z)}{C(S)} = \frac{K(Z, S)N_C(Z)}{N_C(S)} \tag{8}$$

where it is obviously assumed that a linear behavior of yields exists within the concentration range involved. The parameter $K(Z, S)$ represents the ratio between the other terms of Equation (3), with recovery efficiency included, for both the standard and the element to be determined.

The calibration of the procedure for target preparations from water samples by means of element precipitation as carbamates (section 3.1.1), has been calibrated with mono- and multielemental solutions (for both fresh and artificial seawater) and with reference materials for a wide group of elements (Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Se, Hg, Pb, Bi, Mo, Y, Tl, Sb, Ag, Cd), in the parts per trillion (nanogram per kilogram)–parts per million concentration range, after preconcentration of variable sample volumes ranging from 500 to 15 mL.[9,14,29] Satisfactory replicate results have been verified, with measurement errors generally lower than 10%. Problems arise for Mn, whose precipitation as a carbamate is quantitative at pH greater than 6, and for Ti, Cr and Y, which are extractable mainly as hydroxides. The reproducibility of the PIXE measurement of a given target under the same experimental conditions is quite high: the uncertainty, e_s, coming from the slight variability of the set-up parameters (such as the proton flux) may be experimentally estimated (generally $e_s < 5\%$).

The evaluation of another part of the measurement errors in PIXE analysis can be directly performed by the computer codes that examine the spectrum. In fact, each peak fitting system estimates an uncertainty that has to be accounted for in error evaluation: the formulation of this fit error, e_f, depends on the particular spectrum deconvolution procedure used by the code.

For the specimens (such as aerosols) that do not require a target preparation procedure, the PIXE measurement error may result from the contribution of e_f, of e_s and of the variance V_b of a suitable number of blank samples (backings) measured under the same experimental conditions. When a target preparation procedure is needed, the measurement, ε, error expression may be expressed as Equation (9):

$$\varepsilon = V_b + e_f + bC(Z) \tag{9}$$

where b is the average percentage error of the measured concentration, $C(Z)$, evaluated from calibration measurements at the same concentration range of $C(Z)$, and includes e_s. The values of b and of experimental calibration parameters (such as $K(Z, S)$ involved with the internal standard method, Equation (8)) may be determined[9] by a system of nonlinear equations which can be solved by a quick converging iterative process. One equation can be obtained by imposing the condition that the relation shown in Equation (10):

$$\sum_{i=1}^{n} = \left(\frac{\Delta C(Z)_i}{\varepsilon_i}\right)^2 \tag{10}$$

be a minimum (best fit), where $\Delta C(Z)_i$ is the difference between theoretical and measured concentrations of the ith sample, i ranging from 1 to the number n of the calibration samples analyzed. The quantity ε_i is the error (Equation 9) for the same ith sample. A second equation is obtained by imposing the condition that Equation (10) at minimum be equal to the degrees of freedom $(n - 1)$.

3.5 Detection Limits and Ranges

The PIXE intrinsic sensitivity, the LOD calculation and its improvement by coupling target and measurement features, has been previously detailed. The LOD should be quoted with information of integrated beam charge, experimental geometry, energy resolution and background interval. The experimentally evaluated LODs from water and aerosol specimens are here reported. In the case of water samples preconcentrated by elemental precipitation as chelates, the procedure suggested by[27] gives the DLs for marine water and blank listed in Table 4. Unfortunately, information about the measurement parameters are not reported in this work. The DLs calculated for water samples precipitated as carbamates[12] and obtained using 15 μC of 1.8 MeV protons for an averaged blank spectrum and for the Nuclepore backing alone, using K X-rays, are reported in Figure 10. The LOD for the Nuclepore samples (calculated using the external standard calibration method) are comparable with the expected values in a 1 mg cm^{-2} carbon target,[43] for the same integrated charge. The LOD values for the blank are generally close to the values obtained for Nuclepore, even if they are affected by the slightly increased target thickness and by the behavior of each element as a carbamate. The LOD for heavy elements (based on their L X-rays), for an average blank, are 56 (Sb), 8 (Hg), 111 (Tl), 9 (Pb), 10 (Bi) ng cm^{-2}.

Table 4 Data for seawater and blanks

Element	Conc. (ng L^{-1})	Recovery (%)	LOD (seawater) (ng L^{-1})	LOD (blanks)	
				(ng cm^{-2})	(ng L^{-1})
Cr	30	40	2.5	–	–
Mn	–	–	–	1.9	6.4
Fe	9000	70	1.4	2.8	9.5
Co	42	100	3.2	–	–
Ni	420	100	0.8	0.3	0.8
Cu	500	100	1.2	0.7	1.9
Zn	710	65	1.8	0.3	0.8
Pb	160	75	7.0	–	–
Ag	30	100	6.4	–	–
U	430	100	5.3	–	–

Concentrations, percent recoveries and LODs for some elements in seawater and in blanks prepared following the same procedure used for seawater samples. Only five elements could be detected in blank samples, and four of them were detected in the seawater sample. The relatively large concentration of Fe increases the background and hence affects the detection of neighboring elements (Mn, when close to the DL, and mainly Co). (Reproduced by permission of Elsevier from E.M. Johansson, S.A.E. Johansson, *Nucl. Instrum. Methods Phys. Res. B*, **3**, 154–157 (1984).)

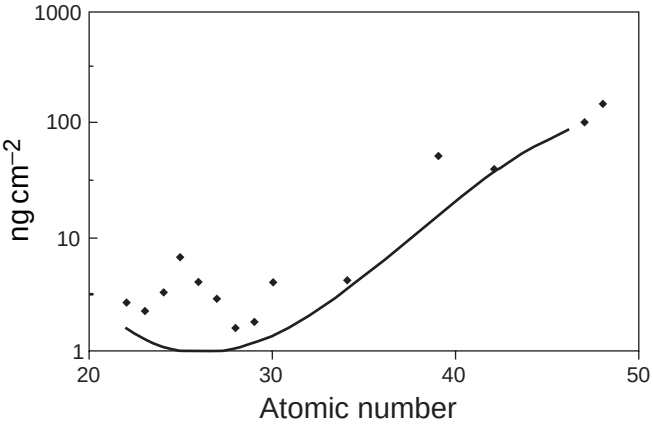

Figure 10 DLs in ng cm^{-2} for the calibrated K X-rays in the averaged spectrum from several blanks obtained by carbamate precipitation (dots), and in a PC Nuclepore filter 1 mg cm^{-2} (line), for 15 μC of protons at 1.8 MeV, detector resolution 150 eV at 5.9 KeV, solid angle 0.0198 sr. (Reproduced by permission of Elsevier from G. Ghermandi, R. Cecchi, P. Laj, *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 63–70 (1996).)

The minimum detectable limit for a 2 ng cm^{-3} PIXE analysis of some air samplers[7] are: SFU 0.6 ng m^{-3}, IMPROVE 0.14 ng m^{-3}, DRUM 1.1 ng m^{-3}, MOUDI 3.3 ng m^{-3}.

The DLs as a function of Z for PIXE analysis of PCI and SDI air samplers are given in Figure 11 (2.4 MeV proton beam, 300 nA cm^{-2}, corresponding to 90 μC charge for 0.5 cm^2 beam size).

These results confirm that the LOD variation with Z follows substantially the rapid decrease in cross-sections mitigated by the fluorescence yield and by the background shape, and is slightly affected by the recovery efficiency of the target preparation procedure (if applicable).

4 INTERPRETATION OF PROTON-INDUCED X-RAY EMISSION SPECTRA OF ENVIRONMENTAL SAMPLES

The analysis of the PIXE spectrum is performed by computer codes that deconvolute the spectrum in order to calculate peak areas. Briefly, for proton-induced X-ray emission, these codes[44] (the most widely use are AXIL, HEX, NANOLEX, PIXAN, SESAMIX and their further improvement, and more recently GUPIX) apply linear or nonlinear (or both) least squares fit (LSF) fitting procedure. The background (to

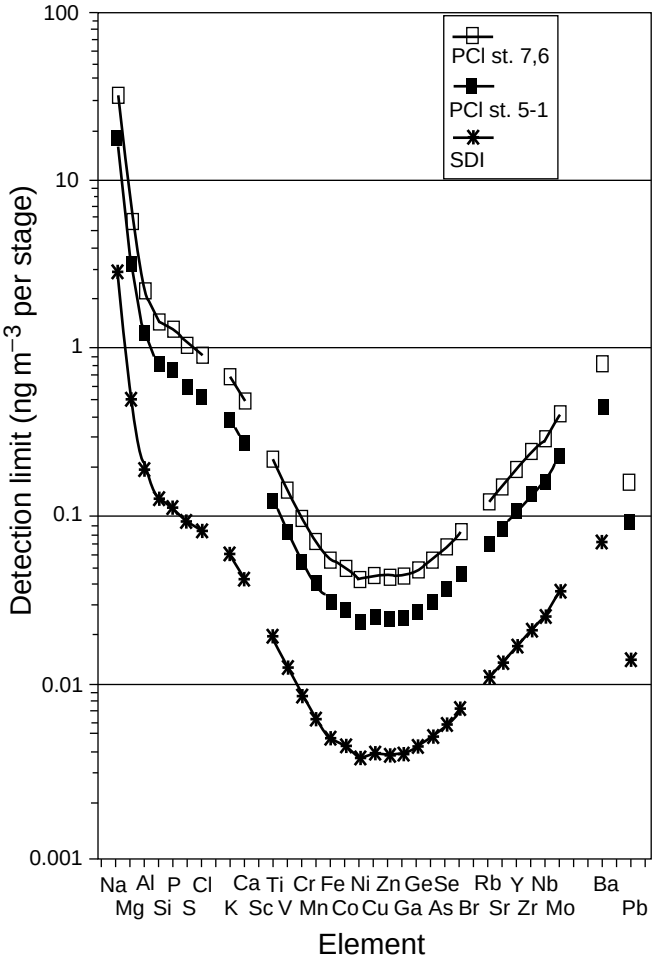

Figure 11 The DLs for K X-ray emissions (evaluated from Equation (6), neglecting the contribution from blank variability) as a function of Z for PIXE analysis of PCI and SDI air samplers, for 2.4 MeV proton beam and 300 nA cm^{-2}, corresponding to 90 μC charge for 0.5 cm^2 beam size. (Reproduced by permission of Elsevier from W. Maenhaut, R. Hillamo, T. Makela, J.-L. Jaffrezo, M.H. Bergin, C.I. Davidson, *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 482–487 (1996).)

be subtracted) is mainly modeled by an absorption-modified exponential polynomial, taking into account the composition and thickness of the target backing (if present). Some codes use digital filters to remove background. The peak shapes are gaussian in some cases plus low-energy tails and steps, and generally pile-up and escape peaks are also accounted for in the spectrum model. All these codes perform thin target analysis, while some of them (and improvements are in progress) can treat intermediate and thick targets (by applying suitable matrix effect calculations). GUPIX also examines multilayer specimens. Taking into account X-ray transmission through the whole detection equipment, the area of each resolved peak in the spectrum is finally directly related to the concentration of the corresponding element by means of the X-ray yield. The commercial availability of continuously evolving computer codes such as GUPIX[45] represents a great advantage for PIXE workers. Based on a large database and a deep knowledge of the physics of X-ray productions by accelerated protons, this code may satisfactorily examine PIXE spectra from each set-up whose technical and experimental parameters have been accurately determined. The peak intensity errors (for $26 \leq Z \leq 83$) through fitting of various spectra are estimated to within <10% for peak areas less than 500 counts and <5% for larger peak areas.

A different approach has been applied to trace element analysis in thin targets which are prepared from chemically pretreated liquid samples. The SPECTRE program[16] is based on experimentally evaluated parameters and calibrations and operates with linear LSF. The background shape is taken as the averaged spectrum from several blank targets measured in the same PIXE region, thus including the chemical treatment background. The peak shape is gaussian. Escape lines and first-order pile-up effects are taken into account. Relative intensities of X-rays and transmission through the whole set-up have been experimentally evaluated by measuring reference targets. The peak identification is performed (without previous instructions or assumptions) among all the elements for which the recovery of the target preparation procedure is significant. The X-ray yields to be used are obtained from calibration measurements of water samples that followed the same chemical treatment. The concentrations in the original water sample are then calculated. Even if suitable for a restricted number of applications, this code provides reliable determinations of trace and ultratrace elements, with fit error generally <5% for peak areas of about 100 counts or greater.

5 METHOD DEVELOPMENT FOR ENVIRONMENTAL PURPOSES

The possibility of coupling PIXE analysis with other techniques in the same irradiation chamber will be discussed in section 6.3. The high-energy proton microprobe outlined here represents a considerable increase in proton PIXE power and flexibility in the environmental field.

5.1 Proton Microbeam Application

In typical microbeam equipment,[6] a beam from the accelerator passes through an analyzing magnet which stabilizes the particle energy, and is then defined by a collimator, with an aperture whose diameter is in the range 10–100 μm. This aperture is placed at a

focusing point where the beam diameter is a minimum: the divergent beam is then further focused by a lens system (usually a magnetic quadrupole). The aperture (more commonly called crossed microslits) is a critical component in a microbeam line, because of scattering of the particles by the edge of the slits. It can give rise to a halo around the central beam that diminishes the resolution. At the end of the beam line, in the vacuum irradiation chamber, scanning over the specimen surface is accomplished by moving either the specimen or the beam. A microscope allows the optical viewing of the sample. A PIXE X-ray detector and the electronic devices for spectral analysis are the same as are used for diffused-beam PIXE. The layout of the microbeam irradiation chamber at INFN Laboratory of Legnaro (Padova, Italy)[46] is depicted in Figure 12.

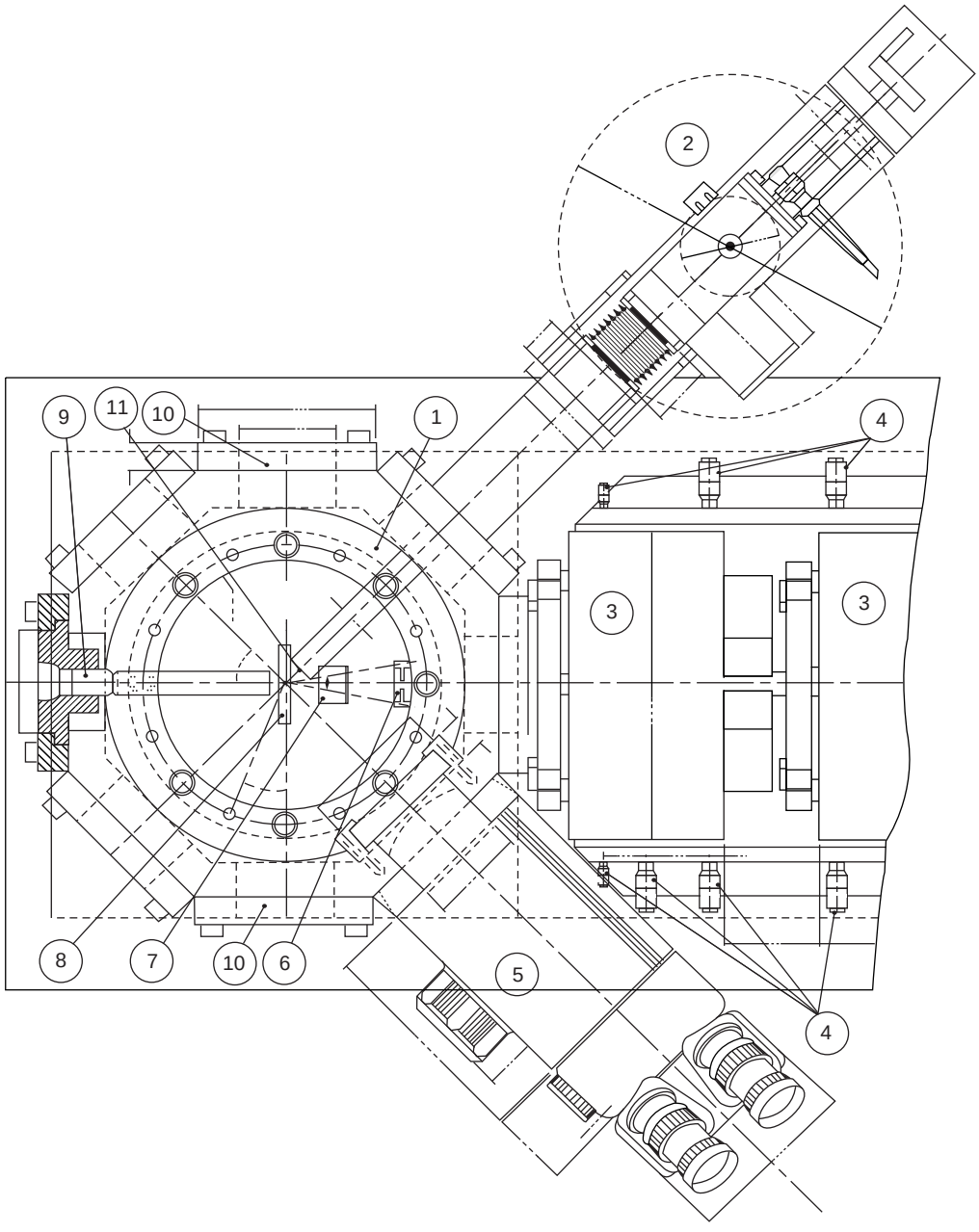

Figure 12 The irradiation chamber of the proton microprobe at INFN Laboratory of Legnaro (Padova, Italy). 1, Reaction chamber; 2, PIXE Si(Li) detector; 3, magnet quadrupole lens; 4, micrometers for lens reg.; 5, stereo microscope; 6, Rutherford backscattering spectrometry (RBS) annular detector (internal diameter = 4 mm); 7, microscope mirror center; 8, target holder; 9, Faraday cup; 10, sample changing port; 11, detector filter holder. (Reproduced by permission of Elsevier from P. Boccaccio, D. Bollini, D. Ceccato, G.P. Egeni, P. Rossi, V. Rudello, M. Viviani, *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 94–98 (1996).)

Most applications of microbeam PIXE are for mapping spatial elemental concentrations. A specimen is moved in small steps (positional reproducibility better than 1 μm) in two orthogonal directions perpendicular to the fixed beam or by linear scanning. A major problem with sequential scanning is specimen heating. Alternatively, the beam may be deflected in two orthogonal directions onto the surface of a fixed specimen. This method is mechanically simpler and reduces sample heating since the scanning is fast. A PIXE spectrum is recorded at each position.

The spatial resolution of the best PIXE microbeams is about 1 μm, but it is not comparable with electron microprobe resolution, which is better than 0.1 μm for surface analyses. In targets thicker than a few tens of micrometers (which includes the most samples) electron scattering broadens the beam, and the resolution of the electron microprobe is considerably impaired: given that the penetration of electron beam is about 10 μm, this is also the effective lateral resolution. The proton microbeam resolution is maintained (1–2 μm) for tens of micrometers of penetration depth, making this technique preferable for specimens thicker than a few micrometers.

The sensitivity of proton microbeam (proton μ-PIXE) may be scaled from proton-induced X-ray emission sensitivity, taking into account that the LOD is inversely proportional to the square root of the integrated charge (Equation 5). In proton μ-PIXE the current in the target (typically hundreds of picoamperes) is one or two orders of magnitude less than in the diffused beam technique. In addition, assuming the same total X-ray measuring time, in the scanning mode, the total collected charge is divided among all the scanned positions. If for a PIXE measurement generally a few minutes are required, proton μ-PIXE requires longer measuring times (some hours to obtain a spatial map). With the same assumptions of section 2.4 for operating parameters, the proton μ-PIXE DLs that result are of the order of 10 ppm of the exposed target, which are better than the LOD for an electron microprobe analyzer, which is about 0.1%. Trace elements are seen only with the proton μ-PIXE: the small size of the irradiated area (few micrometers diameter) allows detection of absolute quantities of 10^{-15}–10^{-16} g. In addition, given the energy of the proton beam (some megaelectron volts) compared to the electron microprobe (typically a few of tens of kiloelectron volts), and the high intensity of the electron bremsstrahlung emission, the detection of elements heavier than (about) $Z = 30$ by means of their K-lines is possible only with proton μ-PIXE.

The use of a proton microbeam to map the elements in a specimen involves the problem of maintaining, in the target, the spatial distribution of the original sample. It may be particularly difficult in biological applications, where the structure even of single cells is examined. In geological applications thin slices of minerals are readily available. In the environmental field, proton μ-PIXE may be a very powerful tool in the investigation of major and trace elements in aerosol particles. The microbeam map of a portion of an aerosol deposit on a thin backing allows the determination of the elements' associations and relative abundances, providing an immediate characterization of the studied aerosol. The image of a single grain of at least some micrometers size can be studied; however, the resolution is probably better with the SEM/EDAX technique, which is preferable for any morphological study.

6 COMPARISON WITH OTHER SPECTROMETRIC TECHNIQUES

Comparison of various features of proton-induced X-ray emission with those of other techniques are included in this paper, in order to support an understanding of its capabilities and limits. Here are reviewed briefly the other major spectrometric techniques and methodologies which are complementary to proton PIXE.

6.1 Methods that use Particle Beams

Deuteron beams are little used in PIXE (and never in environmental analysis), mainly because of the emission of neutrons from nuclear reactions.

α-particles, which do not produce PB in specimens with matrix light elements having $Z/A = 0.5$, may be used in place of protons with very good results, especially when high sensitivity is not required. α-particles of 10–20 MeV (which are often used) provide a high GB contribution to the background, thus limiting the sensitivity. In addition, the large energy loss of α-particles causes appreciable target damage by heating. The beam intensity must then be limited, which offsets the advantage of the higher α X-ray-production cross-sections.[7,11] A study[47] of high interest indicates that for thin organic specimens 5 MeV helium ions provide a significant improvement of LOD compared to 1–3 MeV proton beam.

Heavy ion beams of few hundreds of kiloelectron volt energies have been successfully used in surface analysis given their large X-ray-production cross-sections. The small number of studies performed with heavy ion beams (^{7}Li, ^{12}C, ^{16}O, ^{28}Si, ^{80}Se) at energies of the order of tens of megaelectron volts do not indicate any particular advantage for trace element analysis compared to the use of proton beams. A major problem with heavy ion PIXE is the GB emission: in Figure 4 data are reported for helium and oxygen projectiles (in comparison to protons at equal particle velocities). Given that only the protons generate PB while all three ions generate SEB, the

intense backgrounds (more than ten times greater than for protons) of the helium and oxygen ions at higher energy depend on γ-radiation from nuclear reactions in the specimen. Other problems are the low-energy continuum background from the filling of molecular orbital vacancies and the associated increased spectrum complexity.[7,11]

Electron microprobes were compared with proton μ-PIXE in section 5.1.

6.2 Other Technologies

XRF[48,7] induced by X-ray tubes has some common features with proton PIXE: it allows multielemental, fast, nondestructive analyses of solid specimens. XRF is not an alternative to PIXE in the determination of trace elements in thin samples, such as environmental ones. The sample characteristics cannot be neglected in sensitivity evaluations; proton PIXE LODs for thin, light matrix targets are in the 0.1–1 ppm range (for transition elements and heavy metals up to lead), while for XRF the corresponding LODs are 1–10 ppm, and the difference is more pronounced in absolute terms (a few picograms can be detected by PIXE, and hundred nanograms for XRF). The situation is quite different for thick specimens, such as mineralogical specimens, soil or sediment pellets, where proton PIXE and XRF are comparable, the latter being preferable for heavy-element matrices.

A significant development of XRF is the total reflection X-ray fluorescence (TRXRF) technique. This technique reduces the main source of spectrum background (scattering of primary radiation from the specimen support) by using an incident angle less than the critical one, so that total reflection results with very little scattering. This methods allows absolute DLs of a few picograms, but requires that a thin (<10 μm) sample layer is spread out over the entire quartz plate surface. In diluted natural water (rain, melted snow), sub-parts per billion concentrations may be detected. Groundwater or marine water with high levels of dissolved salts require a selective preconcentration treatment. In solid samples, the relative LODs are the same for XRF as for PIXE.

The use of electron synchrotron radiation as a primary source for XRF (SRXRF) is very useful, because its radiation beam is very intense, with high brightness, naturally collimated and almost totally polarized in the plane of the electron beam, thus providing (with suitable measurement geometry) a reduced scatter from a specimen into the detector. This results in a low background and improves its LOD (0.1–10 ppm in biological reference samples[49]). Given that the SRXRF beams are collimated by apertures typically 30 μm^2, the technique has several applications similar to μ-PIXE (such as mapping). SRXRF microbeams (spot dimension around 10 μm and possible improvement to 1 μm) are under development with the expectation to reach LODs in the 10–100 ppb range.

While an advantage of conventional XRF with respect to PIXE is surely the low cost and availability of experimental equipment, SRXRF has the same main limitations as PIXE but to a much greater extent: small number of existing SRXRF facilities and higher costs.

Neutron activation analysis (NAA) is inherently a multielemental technique like PIXE and XRF; its best LODs are of the order of a few parts per billion and, up to parts per trillion level by employing postirradiation chemical separation. Its sensitivity varies irregularly as a function of Z. Longer times are needed for NAA analyses; the time for activation of the samples has to be taken into account as well as required decay intervals and time for the spectrometric analysis.

Several methods apply both absorption and emission atomic spectrometry.[48,7] These methods require that the samples be in liquid form, to be atomized into the instrument. Now, in atomic emission spectroscopy the most common light source used for the atomic excitation of the specimen is the inductively coupled plasma (ICPAES). AAS uses a flame for sample vaporization. All these optical methods do not have the intrinsic multielemental capability of PIXE, even if some improvements can allow the simultaneous detection of several element. The LODs are of the order of 1–30 ppb for a liquid specimen (in absolute term a few picograms). Depending on the dilution ratio of the original material, these LODs may correspond to sample concentrations ranging from 0.1 to 30 ppm. The interferences are also more complex owing to the larger number of emitted lines and severe matrix interferences.

The requirement for liquid specimens to be injected into the instruments led to the common use of these techniques in environmental studies of natural waters, snow and ice. Trace element determination at the parts per trillion level has been performed using AAS by several authors in sea[50] and fresh water and in polar snow[51] and ice samples. The analysis of solid material involves complex preparation, with possible contamination or changes of the sample, which is destroyed during the analysis. A variant, electrothermal atomization atomic absorption spectroscopy (ETAAAS) deals with solid specimens, providing better LODs than ICPAES, but matrix effects are very appreciable, and, additionally, it requires longer analysis times.

Coupling ICP excitation with mass spectroscopy (ICPMS) reduces the problem of spectral interferences, while the matrix effects become more serious with increased concentrations of major elements, that can interfere with the detection of trace elements. Nevertheless this technique allows the detection of almost all the elements, and has no high variation in sensitivity from

one element to the next. In addition, the cost of high-resolution ICPMS instruments reduces the availability of this equipment.

Finally, neither NAA nor optical spectroscopy methods offer a microbeam capability.

This short review shows that, in environmental applications, other methods can compete with PIXE in water and sediment analysis, while for air pollution particulates, the PIXE technique is surely the most suitable in terms of analytical capability, sensitivity and economy.

6.3 Combined use of Particle-induced X-ray Emission with Other Techniques

Some complementary techniques may be used simultaneously with proton PIXE, with a suitable set-up and appropriate set of detectors placed in the irradiation chamber. Particle elastic scattering analysis (PESA) is based on the dependence of forward scattered particles on the Z of the scattering target atom and on the angle of diffusion. A multielemental analysis may be performed by measuring the energy spectrum of particles scattered at fixed angles, at least for light elements $Z \leq 20$,[8] hydrogen included, by proton–proton scattering. PESA entails more stringent restrictions about sample thickness. It is currently used with proton PIXE[7] to perform hydrogen measurements in air particulate filters, and absolute accuracy of 3% has been achieved with sample thickness below $1\,\mathrm{mg\,cm^{-2}}$, which is the surface density range of almost all the aerosol filters.

Particle-induced γ-ray emission (PIGE) is based on the detection of particle-induced γ-rays. Given the light elements ($Z < 20$) responsible for these emissions (see section 2.1), PIGE may be used in conjunction with proton PIXE for the detection of light elements of interest such as fluorine. Its use for the light elements measurable also by proton PIXE has to be carefully assessed, given the better PIXE sensitivity and the quite high cost of PIGE.[7,8]

RBS is based on the elastic backscattering of incoming ions by nuclei in the sample. RBS allows the determination of the mass of the scattering nuclei and the depth of the nuclei in the target. RBS is used in combination with PIXE[7] for matrix determination, light element determination and depth profiling. The RBS technique is particularly useful for thin specimens (such as aerosol filters) to detect light elements of especial interest (carbon, nitrogen, oxygen). Its detector has to be placed as close as possible to 180°, near the beam axis at a backward angle.

Finally, scanning transmission ion microscopy (STIM)[52] may be performed on a light ion beam that has passed through a sufficiently thin target, by means of a detector placed behind the sample at 0° to the beam. The energy loss caused by ion–electron interaction provides information about density, thickness and structure of the sample. If a scanning microprobe is utilized, STIM provides transmission images. STIM at 0° is highly efficient with low ion currents, which enables spatial resolution of the order of 50 nm,[7] but this condition is not compatible with simultaneous use of RBS and PIXE. Satisfactory results with RBS, PIXE and STIM in biological and aerosol particle studies have been obtained placing the detector at 15°–25° to the beam axis, even if this use of STIM impairs its precision and provides the identification only of the main structures of the target.

ACKNOWLEDGMENTS

The author is very grateful to Rodolfo Cecchi for the essential scientific support and the suggestions in the preparation of this work.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
APDC	Ammonium Pyrrolidine Dithiocarbamate
BEA	Binary Encounter Approximation
CRM	Count Rate
DL	Detection Limit
DRUM	Davis, Rotating-drum, Universal Size Cut, Monitoring
EAD	Equivalent Aerodynamic Diameter
ECPSSR	Energy Coulomb Perturbation Stationary States Effects
EDTA	Ethylenediaminetetraacetic Acid
ETAAAS	Electrothermal Atomization Atomic Absorption Spectroscopy
FEP	Fluorinated Ethylene Propylene
fwhm	Full Width at Half-maximum
GB	γ Background
HDPE	High-density Polyethylene
HEPA	High-efficiency Particulate Air
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
IMPROVE	Interagency Monitoring of Protected Visual Environment
INFN	National Institute of Nuclear Physics
LDPE	Low-density Polyethylene
LOD	Limit of Detection
LSF	Least Squares Fit
MOUDI	Micro-orifice Uniform Deposit Impactor
NAA	Neutron Activation Analysis
NaDDTC	Sodium Diethyldithiocarbamate

NIST	National Institute of Standards and Technology
PB	Projectile Bremsstrahlung
PC	Polycarbonate
PCI	International Cascade Impactor
PESA	Particle Elastic Scattering Analysis
PIGE	Particle-induced γ-ray Emission
PIXE	Particle-induced X-ray Emission
PP	Polypropylene
PVA	Polyvinylacetate
PVC	Polyvinyl Chloride
PWBA	Plane Wave Born Approximation
RBS	Rutherford Backscattering Spectrometry
SCA	Semi Classical Approximation
SDI	Small Deposit Area Impactor
SEB	Secondary Electron Bremsstrahlung
SEM/EDAX	Scanning Electron Microscope–Energy Dispersive X-ray Analyzer
SFU	Stacked Filter Unit
S/N	Signal-to-Noise Ratio
SRXRF	Synchrotron Radiation X-ray Fluorescence
STIM	Scanning Transmission Ion Microscopy
TRXRF	Total Reflection X-ray Fluorescence
TTPIXE	Thick Target Particle-induced X-ray Emission
XRF	X-ray Fluorescence
μ-PIXE	Microparticle-induced X-ray Emission

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Neutron Activation in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Surfaces **(Volume 10)**
Scanning Electron Microscopy in Analysis of Surfaces

Nuclear Methods **(Volume 14)**
Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission) • Rutherford Backscattering Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Sample Preparation for X-ray Fluorescence Analysis

REFERENCES

1. F. Folkmann, 'Ion Induced X-rays, General Description', in *Material Characterization Using Ion Beams*, eds. A. Cachard, J.P. Thomas, Plenum Publishing, New York, 239, 1977.
2. W. Brandt, G. Lapicki, 'Energy-loss Effect in Inner-shell Coulomb Ionization by Heavy Charged Particles', *Phys. Rev. A*, **23**, 1717–1729 (1981).
3. H. Paul, 'An Analytical Cross-section Formula for K X-ray Production by Protons', *Nucl. Instrum. Methods Phys. Res. B*, **3**, 5–10 (1984).
4. W. Bambynek, B. Crasemann, R.W. Fink, H.U. Freund, H. Mark, C.D. Swift, R.E. Price, P. Venugopala Rao, 'X-ray Fluorescence Yields, Auger, and Coster-Kronig Transition Probabilities', *Rev. Mod. Phys.*, **44**, 716–816 (1972).
5. J.H. Scofield, 'Exchange Correction of K X-ray Emission Rates', *Phys. Rev. A*, **9**(3), 1041–1049 (1974).
6. S.A.E. Johansson, J.L. Campbell, *P.I.X.E A Novel Technique for Elemental Analysis*, John Wiley & Sons, New York, 347, 1988.
7. S.A.E. Johansson, J.L. Campbell, K.L. Malmqvist, *Particle-induced X-ray Emission (P.I.X.E)*, John Wiley & Sons, New York, 1995.
8. R. Cecchi, G. Ghermandi, 'The Use of Accelerators in Environmental Study', in *Physics in Environmental and Biomedical Research*, eds. S. Onori, E. Tabet, World Scientific Publishing, River Edge, NJ, 63–74, 1986.
9. G. Aprilesi, R. Cecchi, G. Ghermandi, G. Magnoni, R. Santangelo, 'Calibration and Errors in the Detection of Heavy Metals in Fresh and Seawater by PIXE in the ppb-ppm Range', *Nucl. Instrum. Methods Phys. Res. B*, **231**(B3), 1–3, 172–176 (1984).
10. P. Alougopiannis, G. Robaye, I. Roelandts, G. Weber, J.M. Delbrouck-Habaru, J.P. Quisefit, 'Use of Rutherford Scattering on a Secondary Carbon Target for Correcting Intermediate Thickness Sample PIXE Measurements',

Nucl. Instrum. Methods Phys. Res. B, **14**, 297–303 (1986).
11. S.A.E. Johansson, T.B. Johansson, 'Analytical Application of Particle Induced X-ray Emission', *Nucl. Instrum. Methods Phys. Res. B*, **137**, 473–516 (1976).
12. G. Ghermandi, R. Cecchi, P. Laj, 'Procedures of Target Preparation to Improve PIXE Efficiency in Environmental Research', *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 63–70 (1996).
13. T.A. Cahill, 'Ion-excited X-ray Analysis of Environmental Samples', in *New Uses of Ion Accelerators*, ed. Ziegler, Plenum Press, New York, 1–71, 1975.
14. R. Cecchi, G. Ghermandi, G. Calvelli, 'Ultratrace PIXE Analysis in Water with Low Selective Metal Preconcentration at Various pH', *Nucl. Instrum. Methods Phys. Res. B*, **22**, 460–464 (1987).
15. C.F. Boutron, 'A Clean Laboratory for Ultralow Concentration Heavy Metal Analysis', *Fresenius' J. Anal. Chem.*, **337**, 482–491 (1990).
16. R. Cecchi, M.G. Ferrari, G. Ghermandi, M. Rossi, 'SPECTRE: a Program for Ultra Trace PIXE Analysis of Thin Targets', *Nucl. Instrum. Methods Phys. Res. B*, **49**, 146–151 (1990).
17. P.C. Simms, F.A. Rickey, EPA Report 60011-78-058, 1978.
18. H.C. Hansson, E.M. Johansson, A.K. Ekholm, 'A Nonselective Preconcentration Technique for Water Analysis by PIXE', *Nucl. Instrum. Methods Phys. Res. B*, **3**, 158–162 (1984).
19. P. Laj, G. Ghermandi, R. Cecchi, D. Ceccato, 'Coupling PIXE and SEM-EDAX for the Characterization of Atmospheric Aerosols in Ice Cores', *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 252–257 (1996).
20. P. Laj, G. Ghermandi, R. Cecchi, V. Maggi, C. Riontino, Sungmin Hong, J.P. Candelone, C. Boutron, 'Distribution of Ca, Fe, K and S Between Soluble and Insoluble Material in the GRIP Ice Core', *J. Geoph. Res.*, **102**, C12, Nov. 30, 26 615–26 623 (1997).
21. U. Gorlach, C.F. Boutron, 'Heavy Metal Concentration in Surface Snow from Central Greenland', in *Heavy Metals in the Environment*, ed. J.P. Vernet, Page Bros, Norwich, 24–27, 1989.
22. C. Boni, A. Caridi, E. Cereda, G.M. Braga Marcazzan, 'A PIXE-PIGE Set-up for the Analysis of Thin Samples', *Nucl. Instrum. Methods Phys. Res. B*, **47**, 133–142 (1990).
23. A.T. Ellis, D.E. Leyden, W. Wegscheider, B.B. Jablonski, W.B. Bodnar, 'Preconcentration Methods for the Determination of Trace Elements in Water by X-ray Fluorescence Spectrometry', *Anal. Chim. Acta*, **142**, 73–87 (1982).
24. L.G. Danielsson, B. Magnusson, S. Westerlund, K. Zhang, 'Trace Metal Determinations in Estuarine Waters by Electrothermal Atomic Absorption Spectrometry after Extraction in Dithiocarbamate Complexes into Freon', *Anal. Chim. Acta*, **144**, 183–188 (1982).
25. B.M. Vanderborght, R.E. Van Grieken, 'Enrichment of Trace Metals in Water by Adsorption on Activated Carbon', *Anal. Chem.*, **49**(2), 311–316 (1977).
26. E.M. Johansson, K.R. Akselsson, 'A Chelating Agent–Activated Carbon–PIXE Procedure for Sub-ppb Analysis of Trace Elements in Water', *Nucl. Instrum. Methods Phys. Res. B*, **181**, 221–226 (1981).
27. E.M. Johansson, S.A.E. Johansson, 'PIXE Analysis of Water at the Parts per Trillion Level', *Nucl. Instrum. Methods Phys. Res. B*, **3**, 154–157 (1984).
28. C.L. Luke, 'Determination of Trace Elements in Inorganic and Organic Materials by X-ray Fluorescence Spectroscopy', *Anal. Chim. Acta*, **41**, 237–250 (1986).
29. R. Cecchi, G. Ghermandi, R. Zonta, 'PIXE Analysis to Study Metal Diffusion and Sedimentation Phenomena in Venice Lagoon', *Nucl. Instrum. Methods Phys. Res. B*, **49**, 283–287 (1990).
30. V.R. Navarrete, G. Izawa, T. Shiokowa, M. Kamiya, S. Morita, 'The Quantitative Analysis of Bowen's Kale by PIXE Using the Internal Standard', *Radiochem. Radioanal. Lett.*, **36**(2–3), 151–153 (1978).
31. O. Valkovic, I. Bogdanovic', 'PIXE and XRF Analysis of Marine Sediments', *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 488–492 (1996).
32. V. Valkovic, 'Sample Preparation Techniques in Trace Elements Analysis by X-Ray Emission Spectroscopy', IAEA–TECDOC-300, Vienna, 1983.
33. S. Bernardi, R. Cecchi, F. Costa, G. Ghermandi, S. Vazzoler, R. Zonta, 'A Preliminary Investigation on the Distribution of Heavy Metals in Surface Sediments of the Cona Tidal Marsh (Venice Lagoon)', *Nuovo Cimento*, **11C**(5–6), 667–678 (1988).
34. P. Hopke, P.K.Y. Xie, T. Raunemaa, S. Biegalski, S. Landsberger, W. Maenhaut, P. Artaxo, D. Cohen, 'Characterization of the Gent Stacked Filter Unit PM_{10} Sampler Technical note', *Aerosol Sci. Technol.*, **27**, 726–735 (1997).
35. Instruction manual for International Cascade Impactor (PCI) PIXE International Corporation, Tallahassee, FL.
36. Instruction manual for Model 100 Micro-Orifice Uniform Deposit Impactor (MOUDI), MSP Corporation, Minneapolis, MN, 1989.
37. T.A. Chaill, P.J. Feeney, R.A. Elderd, 'Size-time Composition of Aerosols Using the DRUM Sampler', *Nucl. Instrum. Methods Phys. Res. B*, **22**, 344–348 (1987).
38. W. Maenhaut, R. Hillamo, T. Makela, J.-L. Jaffrezo, M.H. Bergin, C.I. Davidson, 'A New Cascade Impactor for Aerosol Sampling with Subsequent PIXE Analysis', *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 482–487 (1996).
39. W. Reuter, A. Lurio, F.C. Cardone, J.F. Ziegler, 'Quantitative Analysis of Complex Targets by Proton-induced X-rays', *J. Appl. Phys.*, **46**(7), 3194–3202 (1975).
40. H.H. Andersen, J.F. Ziegler, *Hydrogen Stopping-power and Ranges in all Elements*, Pergamon Press, New York, 1975.
41. R. Theisen, D. Vallath, *Tables of X-ray Mass Coefficient*, Verlag stahleisen M.B.H., Dusseldorf, 1976.

42. W. Maenhaut, L. De Reu, H.A. Van Rinsvelt, J. Cafmeyer, P. Van Espen, 'Particle-induced X-ray Emission (PIXE) Analysis of Biological Materials: Precision, Accuracy and Application to Cancer Tissue', *Nucl. Instrum. Methods Phys. Res. B*, **168**, 557–562 (1980).
43. F. Folkmann, in: *Ion Beam Surface Layer Analysis*, eds. O. Meyer, G. Linker, O. Kappeler, Plenum Press, New York, 695–718, 1976.
44. U. Watjen, 'Currently Used Computer Programmes for PIXE Analysis', *Nucl. Instrum. Methods Phys. Res. B*, **22**, 29–33 (1987).
45. J.A. Maxwell, W.J. Teesdale, J.L. Campbell, 'The Guelph PIXE Software Package II', *Nucl. Instrum. Methods Phys. Res. B*, **95**, 407–421 (1995).
46. P. Boccaccio, D. Bollini, D. Ceccato, G.P. Egeni, P. Rossi, V. Rudello, M. Viviani, 'The LNL Proton Microprobe: Original Technical Solutions and New Developments', *Nucl. Instrum. Methods Phys. Res. B*, **109/110**, 94–98 (1996).
47. S.A.E. Johansson, 'Optimization of the Sensitivity in PIXE analysis', *Int. J. PIXE*, **2**, 33–46 (1992).
48. W. Maenhaut, 'Recent Advances in Nuclear and Atomic Spectrometric Techniques for Trace Element Analysis. A New Look at the Position of PIXE', *Nucl. Instrum. Methods Phys. Res. B*, **49**, 518–532 (1990).
49. A.J.J. Bos, R.D. Vis, H. Verheul, M. Prins, S.T. Davies, D.K. Bowen, J. Makjanic, V. Valkovic, 'Experimental Comparison of Synchrotron Radiation with Other Modes of Excitation of X-rays for Trace Element Analysis', *Nucl. Instrum. Methods Phys. Res. B*, **3**, 232–240 (1987).
50. F. Baffi, R. Frache, A. Dadone, B. Cosma, 'The Determination of Heavy Metals in the Ligurian Sea. III Distribution of Copper, Iron and Nickel in the 0–200 Metre Layer in Open Sea', *Chem. Ecol.*, **1**, 233–244 (1983).
51. C.F. Boutron, E.W. Wolff, 'Heavy Metal and Sulfur Emissions to the Atmosphere from Human Activities in Antarctica', *Atmos. Environ.*, **23-8**, 1669–1675 (1989).
52. G. Bench, A. Saint, G.J.F. Legge, M. Cholewa, 'Applications of Energy Loss Contrast STIM', *Nucl. Instrum. Methods Phys. Res. B*, **77**, 175–183 (1993).

Quality Assurance in Environmental Analysis

Malcolm J.R. Clark
Ministry of Environment, Lands and Parks, Victoria, Canada

1 Introduction **2**
1.1 Introduction to Quality Assurance 2
1.2 Economics of Quality Assurance 2
1.3 Quality Control and Quality Assessment 3
1.4 Reasons for Concern 3
1.5 Scientific versus Administrative Studies 3

2 Environmental Analysis **4**
2.1 Good Scientific Practice 4
2.2 International Standardization of Methodologies 5

3 Quality System for the Environmental Analysis Process **6**
3.1 Quality Assurance for Program Design 7
3.2 Quality Assurance for the Study Plan 8
3.3 Quality Assurance for Protocols 8
3.4 Quality Assurance for Presampling Preparation 8
3.5 Quality Assurance for Training and Liaison 8
3.6 Quality Assurance for Sample Collection 8
3.7 Quality Assurance for Sample Handling 9
3.8 Quality Assurance for Laboratory Analyses 10
3.9 Quality Assurance for Data Transmission 11
3.10 Quality Assurance for Data Validation 13
3.11 Quality Assurance for Data Approval 14
3.12 Quality Assurance for Data Provision 14
3.13 Quality Assurance for Statistical Analyses 15
3.14 Quality Assurance for Reporting 16

4 Error Control **16**
4.1 Random Errors 16
4.2 Nonrandom Errors 17
4.3 Gross Errors 20

5 Detection Limits and Limits of Quantitation **20**
5.1 The International Union of Pure and Applied Chemistry Detection Limit Model 20
5.2 Calculation of the Limit of Detection 21
5.3 Different Types of Limits of Detection 22

6 Associated Topics **22**
6.1 Quality Assurance Requirements of Legal Samples 22
6.2 Certified Reference Materials 23
6.3 Interlaboratory Proficiency Testing 23
6.4 Quality Assurance Concerns for Field Measurements 24
6.5 Quality Assurance for Automated Continuous Monitoring 24
6.6 Performance-based Methods 25
6.7 Calculation of Confidence Intervals 25
6.8 Control Charts 26
6.9 Laboratory Automation 27

Acknowledgments **28**
Disclaimer **28**
List of Symbols **28**
Abbreviations and Acronyms **28**
Related Articles **28**
References **29**

Vigorous and thorough programs of quality assurance (QA) are vital to ensure that environmental analysis studies yield results which are trustworthy, scientifically credible, and of known quality commensurate with their intended use. Mistakes in any step of the environmental analysis process can result in a substantial increase in random and nonrandom errors. Poor design of an environmental analysis program or failure to adhere to good scientific practices (GSP) for every step of the environmental analysis process can result in compromised or even meaningless results. Therefore, a holistic approach must be taken to ensure adequate QA is implemented for each and every step of the environmental analysis process, from initial study design through final information reporting. However, there is a substantial economic cost to

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

incorporate QA on such a thorough and comprehensive basis. Therefore, QA efforts are unlikely to succeed unless management is committed to the value of these efforts. Increased recognition of the importance of QA, plus broadened international adoption of harmonized standard QA methodologies, has substantially improved the reliability of environmental analyses. QA ensures that environmental monitoring results are compatible with project goals, are comparable between different agencies, and maintain a high degree of scientific credibility. The key elements of QA programs include comprehensive planning, defined data quality objectives (DQOs), thorough training of personnel, standard operating procedures, detailed documentation, timely resolution of problems, regular reporting, routine independent audits plus regular challenges of study elements.

1 INTRODUCTION

1.1 Introduction to Quality Assurance

Resource management decisions underpin our environmental, economic, and industrial prosperity. It is important that monitoring results upon which such decisions are based be trustworthy. Therefore, every environmental study should include a QA component to provide confidence that the study results are credible. The key elements of the QA component of environmental analysis studies include

- comprehensive planning;
- documentation of study objectives;
- documentation of study methodologies;
- thorough training of personel;
- comprehensive record keeping;
- timely identification and resolution of problems;
- regular reporting of results;
- routine challenge of study elements; and
- regular independent audits to ascertain whether study elements are in control.

There are a number of somewhat different approaches to QA. While most people would agree with the concept that environmental measurements should be reliable, there often is debate as to how best to achieve this goal at an affordable cost. Nonetheless, there is a broad consensus that QA is a critical consideration for every element or activity of the environmental analysis process.[1,2]

There are different reasons why samples are collected. Samples may be taken to determine compliance with government regulations (legal sampling). In other instances, samples might be collected to determine the presence or absence of some particular contaminant and also the extent of its dispersal in the environment (environmental monitoring). Sometimes samples are collected to determine change in environmental conditions or change in levels of contamination (trend monitoring). The specific purpose underlying environmental analysis programs often dictates the choices as to sampling methodology, sampling sites, sample collection strategy (random, systematic or judgmental), sample frequency, types of container, and intensity of QA effort. Therefore, it is important that the specific purpose of environmental analysis programs be defined before any routine process of sample collection and sample analysis is established. DQOs must be defined in the initial project planning stage. Field personnel, laboratory analysts, data users, statisticians, and systems analysts should be jointly involved from the study onset, since each has a critical role towards achieving the collection of results which are both useful and credible.

It should be emphasized that QA for the field sampling aspects of environmental analyses and for the laboratory analysis aspects are two distinctly separate but interrelated activities. The first responsibility of field personnel is to ensure that samples collected truly are representative of the ecosystem being sampled. The first responsibility of laboratory analysts is to ensure that the analysis results accurately describe each sample received. Sampling errors and analytical errors often occur independently of each other. High analytical precision in a laboratory will not compensate for errors made during sample collection. Careful collection of representative samples in the field will not compensate for poor precision in the laboratory.

1.2 Economics of Quality Assurance

From a ''total quality management'' point of view, quality cost is usually described as a framework for analyzing an organization's quality policy. On the one hand there is the cost of ensuring (prevention cost) and of assuring (appraisal cost) that a product meets a certain quality standard. Balanced against these costs are the costs when this standard of quality is not achieved (failure cost).

The QA component of an environmental analysis study seeks to increase the likelihood that the investment cost of collecting and analyzing samples returns data both credible and useful for the intended purpose. QA also seeks to offer some guarantee or assurance that monitoring data collected by two or more agencies will be comparable. Certainly the main benefit to environmental analysis studies derived from QA is enhanced credibility of results to colleagues, peers, and clients. The main disadvantage is a significant increase in overall costs, often by 10–20% or more. Most of this additional cost is the

direct result of the costs associated with control samples, audit samples, standard reference materials (SRMs), and equipment preventive maintenance. Nonetheless, the cost of adding QA is much less than the cost of redoing a study, and very much less than costs associated with erroneous resource management decisions made upon dubious or flawed information. Both the costs and benefits of QA have been tabulated by Taylor,[3] as compiled in Table 1. Although Taylor's tabulation of costs and benefits were only for the laboratory analysis portion of environmental analysis studies, nonetheless these costs and benefits equally apply to the entire environmental analysis process.

Table 1 Costs and benefits of QA. (Reprinted with permission from J.K. Taylor, *Quality Assurance of Chemical Measurements*, CRC Press, Boca Raton, FL, 1987. Copyright CRC Press, Boca Raton, FL.)

Direct costs
- Test materials
- Standards
- Quality assurance equipment – test instruments
- Analysis of QC/quality assurance samples
  - Time of personnel
  - Time of supervision
- QA official
- Committee work
- Round-robin costs
- Travel/attendance at meetings

Indirect costs
- Training
- Extra cost for quality people
- Extra quality equipment
- Extra quality supplies
- Relaxed work schedules

Benefits
- More efficient outputs
- Fewer replicates for same reliability
- Fewer do-overs
- Greater confidence of:
  - Staff
  - Laboratory
  - Customers

Intangible benefits
- Promote external image
- Improve internal image
- Promote client confidence
- Add credibility to results
- Prevent hasty decisions
- Minimize indecision
- Eliminate unnecessary redundancies
- Promote continuity of effort
- Provide for retention of vital records
- Set forth goals and objectives
- Provide guidance to staff
- Provide basis for training

1.3 Quality Control and Quality Assessment

QA consists of two separate but complementary aspects, quality control (QC) and quality assessment. For this reason, QA sometimes is referred to as quality assessment plus quality control (QA/QC). QC includes all those activities having the purpose of controlling the quality of produced information to ensure it is adequate for the purpose intended. Quality assessment includes all those activities undertaken to confirm that the QC activities were performed effectively. It is useful to consider that QC refers to those aspects of QA which are fully under an investigator's control (such as instrument calibration), whereas quality assessment refers to those aspects of QA which are not under an investigator's control (such as independent audits). To eliminate the confusion resulting from the similar terms ''quality assurance'' and ''quality assessment'' and their identical acronyms, some organizations prefer the terms ''internal QC'' (synonymous with ''QC''), and either ''external quality assessment'' or ''quality surveillance'' (both synonymous with ''quality assessment''). In this article, the acronym QA is restricted to the term ''quality assurance'' except when it occurs within the double acronym QA/QC. It should be noted that some authors use the term ''QA'' both as defined above and additionally as synonymous with ''quality assessment'' (there is an example of this in Table 1).

1.4 Reasons for Concern

Experienced scientists [e.g. Mesley et al.,[4] Rose and Smith[5]] continue to warn that a significant proportion of environmental data sets are not fit for the purpose intended. Unreliable data sets continue to be accumulated in the literature and on electronic databases. Ultratrace analyses are particularly vulnerable to serious error. King[6] states that ''This is both a costly and a dangerous problem which could get worse as the demands of society increase''. Additionally, many present-day environmental data sets will be of interest in future decades and centuries. The credibility of present-day data sets to future scientists requires that trustworthy data be unambiguously distinguishable from untrustworthy data. Also, since equipment and methodologies are changing at a rapid pace, future scientists must be able to ascertain whether or not various different methods employed over time are comparable to each other. Both reliability and comparability assessments will require convenient (i.e. electronic) access to permanent archives of complete chronological records including methodology details, plus results from comprehensive programs of QA.

1.5 Scientific versus Administrative Studies

An unfortunate fact of the real world is that studies do occur without QA. Reports from studies with QA and

reports from studies without QA can look very similar in appearance, particularly to nonscientists. Clark[7] refers to studies having comprehensive QA as ''scientific studies'' and to studies without QA (or with inadequate QA) as ''administrative studies''. Clark[7] notes his experience that scientists and administrators have very different points of view as to the relative value of scientific versus administrative studies, as indicated in Table 2. Administrators often consider QA to add unnecessary project costs, whereas scientists view QA as essential to keep project results scientifically credible. It is important that both scientists and administrators recognize these differences in viewpoint. It cannot be emphasized too strongly that studies that lack QA also lack genuine credibility, and put at risk the very environmental resources the studies are meant to protect. Management commitment to QA is essential for QA/QC efforts to succeed.

Often it is marketplace arguments, not scientific merit, which bring administrators to endorse QA. Environmental remediation costs, environmental liability insurance fees, lawyers fees for court appearances, regulator fees and fines, impacts of adverse public relations – all are expensive costs. QA is the best strategy to avoid or minimize such costs. There is increasing recognition that QA genuinely can contribute to the ''bottom line'' by reducing overall operating costs and vulnerability to regulatory fines. It can be argued that international recognition of this fact underpins International Organization for Standardization (ISO) 9000 and ISO 14000, both of which are discussed later in this article.

Table 2 Comparison of a scientific study with an administrative study. (Source: Clark.[7])

	Scientific study	Administrative study
(a) From a scientist's point of view		
Time frame	Years	Months
Level of examination	Detailed	Cursory
Expertise	PhD	Student
Cost	Costly	Inexpensive
Data	Sufficient	Insufficient
Data credibility	High	Low
Recommendations	Credible	Not trustworthy
Publication in scientific journals	Yes	No
Overall credibility	High	None
(b) From an administrator's point of view		
Speed	Slow	Fast
Details	Too many	Few
Expertise	PhD	Student
Cost	Costly	Inexpensive
Data	Too many	Sufficient
Data credibility	Good	OK
Recommendations	Fuzzy	Precise
Report publication	Too many pages	OK
Understandability	Low	High
Overall credibility	Adequate	Good

2 ENVIRONMENTAL ANALYSIS

''Environmental analysis'' is the overall process of sampling or monitoring environmental systems so as to derive understanding of environmental quality and ecosystem operating mechanisms. Sample collection and laboratory analysis each forms only part of the overall process of environmental analysis, as described by Porter et al.[8] (Figure 1) and by Clark and Whitfield[2] (Figure 2). Environmental analysis itself is a key component of the resource management (sometimes called ''environmental management'') process. In order to have credible results produced, every component of the environmental analysis process must be documented to have a program of QA in place and successfully implemented. It should be noted that some scientists employ the term ''environmental analysis'' to refer solely to laboratory operations concerning the analysis of environmental samples.

2.1 Good Scientific Practice

It is easy to challenge the credibility of environmental studies. Environmental samples are collected under imprecisely controlled conditions, but the environments from which these samples are collected are both heterogeneous and dynamic. Also, samples might be unintentionally modified at any of the steps that make up the environmental analysis process. To ensure environmental data are scientifically credible, environmental scientists must document in detail their adherence to ''GSP''. GSP (sometimes called ''accepted scientific practices'') are those operating procedures that have been accepted by scientists of authority to be the minimum requirements to promote quality and integrity of scientific findings and results. GSP include detailed documentation of objectives, of experimental design, of methods and procedures employed, of hypotheses tested, of unusual events, of deviations from regular practices, of results from ongoing evaluations of bias and precision, and of results from regular independent audits. Together these records provide a complete chronological record to support the validity of the environmental analysis results.

The literature includes papers describing GSP for environmental analysis studies in general.[1,2,9] The literature also contains numerous papers describing GSP for each of the elements of environmental analysis studies, particularly program design,[5,10–13] sample collection and handling,[14–16] laboratory analyses,[17,18] data reporting,[8,19] data assessment,[13,20] and statistical data

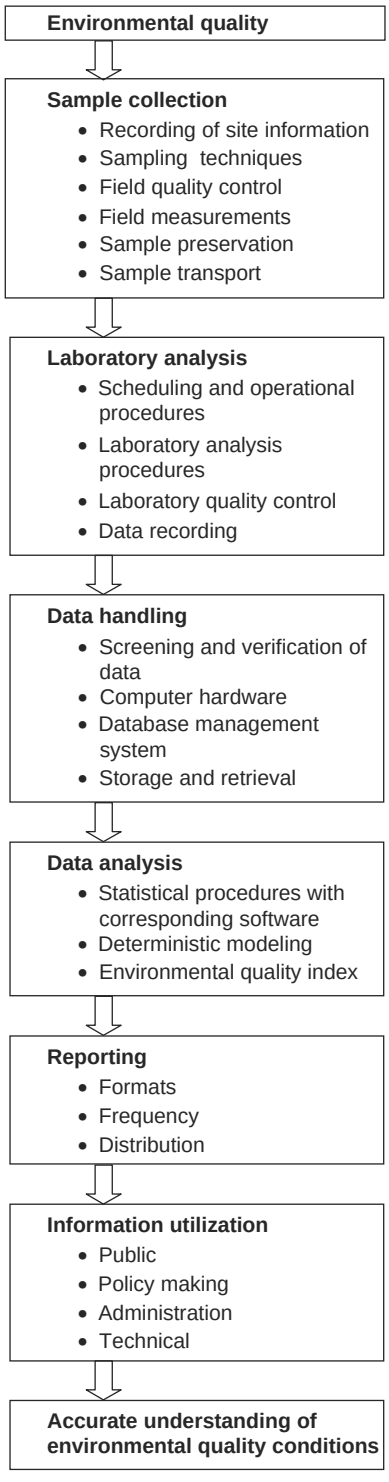

Figure 1 Flow of information in an environmental analysis system. Note that sampling and laboratory analysis form two parts of a series of processes. (Reprinted with permission of the American Chemical Society from P.S. Porter, R.C. Ward, H.F. Bell, 'The Detection Limit', *Environ. Sci. Technol.*, **22**(8), 856–861 (1988). Copyright 1988, American Chemical Society) (modified slightly from the original figure, which pertained to water quality systems).

evaluation[11,21] (this is just a small selection of citations from many hundreds of papers). Synonymous terms for ''GSP'' regarding subelements of the total assay process include ''good monitoring practice'', ''good measurement practice'', ''good sampling practice'', and ''good laboratory practice'' (GLP). The flowchart in Figure 3 shows the integral relationship between GLP and QA. By replacing the term ''laboratory'' in the phrase ''good laboratory practice'' by an appropriate term, this schematic could apply to other elements of the environmental analysis process.

Environmental analysis studies benefit greatly by organizing QA on a very formal basis. Allocation of responsibilities should be clearly delineated, and reports should bear signatures. Each major study element should be described in a ''Quality Assurance Plan'', should be detailed in a ''Quality Assurance Manual'', and should be the responsibility of a ''Quality Assurance Officer''. It is widely recognized that independent audits by impartial expert auditors offer a guarantee that GSP are maintained. An even stronger guarantee results when such independent audits are performed on behalf of an authoritative certifying organization, and methods are vetted against published standard protocols (e.g. ISO/IEC Guide 25) as part of a formal accreditation process. Basically accreditation is a formal recognition of the technical competence of an organization to carry out specific defined methodologies and to produce credible results.

2.2 International Standardization of Methodologies

There has been increasing recognition of the need for international adoption of harmonized standard methodologies to ensure data collected by different agencies are comparable. This is particularly important with regard to data sets collected to characterize transboundary and global environmental problems. ISO has been taking a lead role towards such international standardization, and has published a number of standard methodologies. In 1987 the ISO released ISO 9000, its first core series of standards for QA. The ISO 9000 series of standards addresses quality management and QA essentially with regard to any development, production, or service business, including analytical laboratories. Albu[22] summarizes the implementation of ISO 9000 standards from the point of view of a company which had recently attained ISO accreditation. In 1996 ISO released the ISO 14000 series of standards, that specifically address environmental quality management, as summarized by Cichowicz.[23] To date some 80 countries have adopted the ISO 9000 series as national standards. However, global implementation of these standards still remains to be accomplished. A major revision to the ISO 9000 standards is scheduled for release in 2000 or soon thereafter.

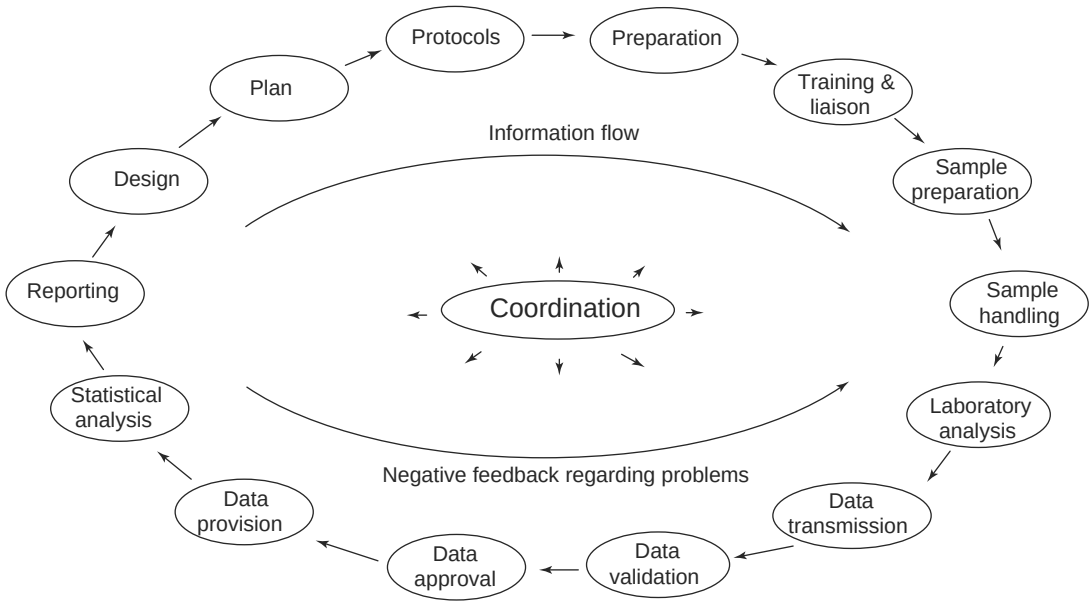

Figure 2 Fourteen-element iterative cycle of the environmental analysis process. (Reproduced with permission of the American Water Resources Association from M.J.R. Clark, P.H. Whitfield, 'A Practical Model Integrating Quality Assurance into Environmental Monitoring', *Water Resour. Bull.*, **29**(1), 119–130 (1993).)

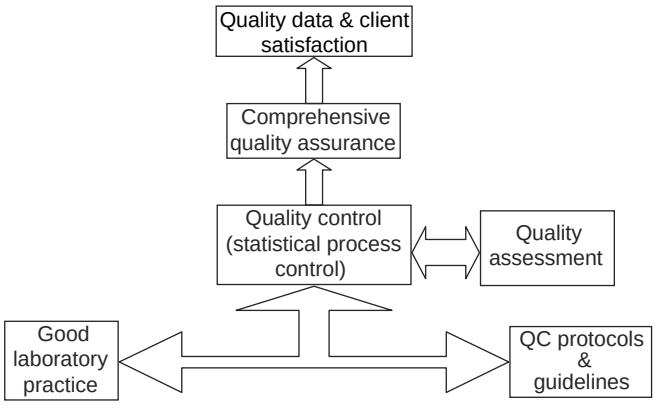

Figure 3 GLP as a component of QA. (Reproduced with permission of the National Laboratory for Environmental Testing from H. Agemian, *Quality Assurance in the National Water Quality Laboratory*, 3rd edition, Canada Centre for Inland Waters, Burlington, ON, 1989.)

Harmonized technical protocols are being produced jointly by the ISO and other international scientific bodies such as the International Union of Pure and Applied Chemistry (IUPAC) and the Association of Official Analytical Chemists (AOAC) International. Horwitz[24] has summarized the history to 1992 of the IUPAC/ISO/AOAC Harmonization Program. The ISO standards have had a major impact because of the associated rigorous certification process. Such certification is now firmly in place for analytical laboratories. Harmonized international standards and certification for other elements of the environmental analysis process have progressed more slowly. Sample collection is particularly lagging behind to this regard, and as a result often is viewed as the Achilles' heel of the environmental analysis process.

With specific regard to laboratory accreditation, it should be pointed out that there is a common misunderstanding that registration to ISO 9000 automatically assures the validity of specific test results to a level equivalent to accreditation to ISO/IEC Guide 25. Both this guide and also EN 45001 (a laboratory guide well known in Europe) do assure validity of test data to an established technical competence. However, ISO 9000 does not as thoroughly address technical credibility and competence. (It is expected that the proposed new International Standard ISO/IEC 17025 will soon replace both ISO/IEC Guide 25 and EN 45001.)

3 QUALITY SYSTEM FOR THE ENVIRONMENTAL ANALYSIS PROCESS

Quality planning is absolutely essential if an environmental analysis program is to produce credible results. Therefore, senior management must ensure that quality planning is understood, implemented, and maintained at all levels of an organization. For environmental analysis programs, it is strongly preferred that organizations formally conduct their operations in accord with appropriate international standard protocols. Even where no such

format arrangements are in place, it is mandatory that GSP be employed on a rigorous basis. Each organization involved in the environmental analysis process must establish a ''documented quality system'' (DQS) for each and every function where controls are crucial to ensure the quality of the data. This DQS must be approved and endorsed by senior management, and must also be communicated to all staff. Typically there are three types of DQS documents:

- the ''QA Manual'', which states the organization's policies and commitments to quality, with emphasis upon commitments to national or international standards;
- the ''QA Plan'', which details how the organization will meet and achieve the commitments of the QA Plan;
- standard operating protocols (SOPs), which are task-specific instructions for correct activity performance, and for quality criteria to be achieved.

(Note that some organizations use ''QA Manual'' with the above definition for ''QA Plan'' and vice versa; other organizations use one term to cover both; other organizations use quite different terms. Also, some organizations prefer to employ SOP as the acronym for standard operating procedures.) Every person participating in the environmental analysis process has responsibility to maintain a high standard of QA. Nonetheless, large organizations usually find it beneficial to assign one person primary responsibility to ensure that QA actually is maintained within an organization. This person usually is referred to as the ''QA Officer.''

Environmental analysis programs can be divided into a number of discrete interlinked elements. Clark and Whitfield[2] have suggested a 14-element model of such a linked multielement environmental analysis process. A schematic of this model is presented in Figure 2. This model divides the overall monitoring process into an ordered cyclical series of linked elements. For large environmental programs, each element (sometimes several adjacent elements) will be performed by different organizations, and therefore will have its own separate QA Manual, QA Plan and QA Officer. It is important that good communications be maintained between persons responsible for the various elements, so that effective and timely feedback can occur whenever problems occur. Large programs require central coordination of activities to ensure effective communication and overall-project management. Each element of the environmental analysis process is subject to changes and enhancements over time, reflecting improvements in instrumentation and methodology. The reliability of the information often is determined by the weakest link of the interlinked elements. In view of the rapid pace of technological change, tracking and characterizing the impacts of these changes on the collected data warrants a substantial QA effort. QA forms an integral part of each element of the environmental analysis process, as summarized in the following discussions.

3.1 Quality Assurance for Program Design

The 'design element' of environmental analysis programs defines the study boundaries, the specific study objectives, and the data collection strategies to meet those objectives. Every environmental analysis program should have specifically designed goals (in the QA Manual) and also effective strategies to meet those goals (in the QA Plan). Nonetheless, many environmental programs continue to be poorly designed and implemented, to the extent that the results compromise the intended objectives. Laane and ten Brink[12] and also Maher et al.[13] note that a poor standard of design and implementation of environmental surveys has resulted in massive collections of data which are ''data rich but ... information poor''. One strategy to ensure environmental analysis programs do produce data fit for their intended purpose, has been to specify that DQOs must be established at the design stage of the environmental analysis process. DQOs are derived from the intended use of the data, and serve as criteria for data acceptability, as is described by Keith.[25] Two examples DQOs are (a) that the analytical variance for each parameter should not exceed 4% of the observed local variance and (b) that the sum of the analytical variance plus sampling variance should not exceed 20% of the observed local variance. Both of these DQO rules have been employed in genuine circumstances, as described by Ramsey et al.[26] and Clark et al.[27] Such methods are cost effective, particularly when they permit one to weed out inappropriate methods which fail to produce useable data. The disadvantage of DQOs and of statistically valid sampling designs in general is that some environmental analysis programs may not be economically realistic, and the study manager must grapple with the choice of finding additional funds, designing a reduced-scale program, developing new methodologies, or collecting data that are of questionable merit.

The primary QA goal for the design element of environmental analysis is to ensure that specific study objectives are defined, and appropriate data collection strategies are implemented. Comprehensive consideration of the statistical aspects of program design is a priority. Fortunately, there is a sizable literature on this topic, e.g. Rose and Smith,[5] Hurlbert,[10] Ward and Loftis,[11] and Maher et al.[13] Comprehensive consideration of the chemical aspects of program design is equally a priority. Quantitative understanding of the chemistry of ecosystems

involves both thermodynamics and kinetics. Equilibrium approaches are useful in understanding the inorganic composition and reactions in simple aqueous ecosystems. However, as detailed by Brezonik[28] in an excellent textbook, kinetic approaches are essential when it comes to understanding of the composition and reactions within environments where nonequilibrium conditions persist. An advantage of the cyclic model of environmental analysis is that study objectives, study area boundaries, DQOs, and sampling strategies can be optimized as new information comes available each cycle.

3.2 Quality Assurance for the Study Plan

The ''study plan'' is a DQS document which details the responsibilities, schedule, and budget of every participant on an environmental analysis project. This plan defines the tasks and steps necessary to attain each goal or objective specified in the design. The QA aspect of the study plan should ensure that all participants are familiar with the plan, that each participant has clearly defined responsibilities, and that there is effective communication between study participants. Any deviations from budgets and schedules should be identified and rectified on a timely basis. Evaluation of historical data (for either the study ecosystems or for similar ecosystems), pilot studies, and mock projects using surrogate data all assist in early identification of conflicts between study participants with regard to objectives, responsibilities, schedules, or methods. These practice studies also serve to identify training requirements, and equipment or methodology deficiencies.

3.3 Quality Assurance for Protocols

The ''study protocols'' (variously called ''standard operating protocols'', ''standard operating procedures'', ''SOPs'', or ''work instructions'') are formal written procedures for all methods employed during the course of an environmental analysis study. Included in the SOPs are details of records to be maintained, QA activities, and remedial responses to be employed in response to various types of method failure. The SOPs should specify the reporting and evaluation of QA results on a regular and timely basis, and should also specify acceptance criteria for specific QA programs. The QA aspect of study protocols must ensure that these protocols are kept up-to-date and are readily available to all members of the study team, that they do genuinely reflect actual practice, and that a copy of the protocols is safely maintained in some permanent archive. In spite of the best of intentions to use only the most trustworthy instruments and most up-to-date methods, sometimes this is not possible. Instrument failure, accident or illness, or other unexpected events may necessitate the use of nonstandard practice. Therefore, provision must be made within the protocols for handling contingencies and deviations from normal practices.

3.4 Quality Assurance for Presampling Preparation

''Presampling preparation'' includes those activities that take place prior to traveling to the sampling locations. These activities include preparation of reagents and sample bottles, maintenance and calibration of equipment, and shipping of materials to remote locations. The QA for presampling preparation include ensuring that maintenance and calibration activities are performed correctly, on schedule, according to SOP specifications, and also that appropriate maintenance and calibration records are kept. It is particularly important that reagents, reagent water, and sample bottles be checked for contamination. If material is being shipped to remote locations, late or incomplete shipments can result in significant program interruption. Checklists are particularly useful at this stage of the environmental analysis process. Table 3 is based upon a checklist published by Keith.[1]

3.5 Quality Assurance for Training and Liaison

''Training and liaison'' covers both training programs plus communication between headquarters staff and field personnel. Staff training should be ongoing, and particularly should be emphasized where changes in personnel are frequent. Associated QA includes verification that training records are maintained and up-to-date, plus verification that personnel assigned specific tasks have the necessary education, training, and experience. Safety training is essential for everyone, even the most senior staff.

For those environmental analysis programs that have separate headquarters staff and field staff, effective two-way communication is important. Information regarding problems, unusual events, and deviations from standard practices must be communicated rapidly between study participants, in order to avoid misunderstandings and continued use of flawed methods, or inappropriate use of flawed data. Associated QA includes verification that communication records are complete and up-to-date, and verification that personnel are familiar with protocols for handling contingencies and deviations from normal practices.

3.6 Quality Assurance for Sample Collection

The ''sample collection'' element of the environmental analysis process includes those activities which take place immediately preceding and during the collection of samples. These activities include on-site maintenance and calibration of equipment, site selection, sample collection, on-site sample handling, in situ measurements, photography and all related observations, measurements,

Table 3 Example presampling preparation checklist. (Reprinted with permission from L.H. Keith, *Environmental Sampling and Analysis: a Practical Guide*, CRC Press, Boca Raton, FL, 1991. Copyright CRC Press, Boca Raton, FL.)

What observations at sampling sites are to be recorded?
Has information concerning DQOs, analytical methods, LODs. etc., been included?
Have instructions for modifying protocols in case of problems been specified?
Has a list of all sampling equipment been prepared?
• Does it include all sampling devices?
• Does it include all sampling containers?
• Are the container compositions consistent with analytes?
• Are the container sizes consistent with the amount of samples needed?
• Does it include all preservation materials/chemicals?
• Does it include materials for cleaning the equipment?
• Does it include labels, tape, waterproof pens, and packaging materials?
• Does it include chain of custody forms and sample seals?
• Does it include chemical protective clothing or other safety equipment?
Are there instructions for cleaning equipment before and after sampling?
• Are instructions for equipment calibration and/or use included?
• Are instructions for cleaning or handling sample containers included?
Have instructions for completing sample labels been included?
Have instructions for preserving each type of sample been included?
• Do they include maximum holding times of samples?
Have instructions for chain of custody procedures been included?
Have safety plans been included?

LOD, limit of detection.

and records. Photographs and detailed field records are invaluable during data interpretation. For the environmental samples to be representative of the ecosystem, spatial and temporal gradients should be sampled adequately to characterize the gradients. Sharp gradients are best avoided if practical, since they require very intensive monitoring effort to be characterized. A selection of papers from the large literature on this topic includes Kramer,[14] Stroomberg et al.,[15] Wagner,[16] and Barcelona.[29] Heterogeneity at a site is usually explored by collection of multiple samples.

QA aspects of the sample collection should ensure that SOPs are followed, details as to site location and condition are recorded, and details as to sample collection and sample characteristics also are recorded. Proactive QA should include specific protocols to confirm that samples are representative of the ecosystems being studied, that samples have not been contaminated through the sampling process or otherwise modified, and that each sample has been properly and uniquely identified. Since ecosystems are both heterogeneous and dynamic, any collected samples also will be both heterogeneous and dynamic. This fact has major consequences for the QA aspects of any environmental analysis program. Highly dynamic variables must be either preserved or measured in situ or on-site. Those ecosystems that are highly heterogeneous over time or space require intensive sampling effort in order to be adequately characterized. It is often worthwhile to employ on-site measurements of a few variables in order to target effective sample collection from such heterogeneous ecosystems (these samples subsequently being analyzed under laboratory conditions for the entire suite of relevant contaminants). With regard to collection of samples for trace and ultratrace analytes, contamination is difficult to avoid. Therefore, scrupulous care and intensive QA are required, such as that suggested by Stewart and Anderson[30] for metals. Proactive QA should include intensive use of blank samples, replicate samples and other QA/QC samples. Table 4, based on Cavanagh et al.,[31] tabulates common QA/QC sample types along with brief descriptions as to the purpose of each type. Note that replicate samples should be collected as closely located in space and time as possible. Preferably a multiple sampling device should be used to collect simultaneous, immediately adjacent replicate samples. GSP call for some proportion of QA samples to be submitted on a blind basis. GSP also call for independent audits of sampling. Additionally, where two study teams are carrying out environmental analysis studies having overlapping or adjacent study boundaries, simultaneous independent sampling at a common site by the two sampling teams is encouraged.

3.7 Quality Assurance for Sample Handling

The ''sample handling'' element of environmental analysis includes all activities which occur between collection

Table 4 Summary of QA/QC sample types. (Reproduced with permission of the Ministry of Environment, Lands and Parks, British Columbia, Canada, from N. Cavanagh, R.N. Nordin, L.W. Pommen, L.G. Swain, *Guidelines for Interpreting Water Quality Data*, Ministry of Environment, Lands and Parks, Victoria, 1998.)

Sample type		Measures
• Laboratory replicates	⟶	Analytical precision
• Field replicates	⟶	Sampling + environmental + analytical precision
• Certified reference samples	⟶	Analytical accuracy (bias)
• Certified reference replicates	⟶	Analytical accuracy (bias) and precision
• Spiked samples	⟶	Analytical accuracy (bias)
• Field blank	⟶	Contamination (bias and imprecision) introduced during sample handing in the field and laboratory
• Trip blank	⟶	Contamination (bias and imprecision) introduced by the container, by preservative, and/or during transportation
• Equipment blank	⟶	Contamination (bias and imprecision) introduced through improper cleaning techniques
• Filtration blank	⟶	Contamination (bias and imprecision) introduced from the filtration apparatus and inadequate cleaning of apparatus
• Laboratory blank	⟶	Contamination (bias and imprecision) introduced during laboratory analysis

of a sample and receipt of that sample at a scientific institution or laboratory responsible for cataloguing, warehousing, or analysis of the sample. Included in the sample handling activities are sample preservation, sample packaging, and sample transportation. The QA aspect of sample handling includes those activities to assure sample integrity. Proactive QA should include protocols to establish that unstable samples are stabilized and handled appropriately, and also to assure that samples are not contaminated through the sample handling activities. GSP call for the stability of analyte to be confirmed for each type of sample being studied. Note that some analytes will degrade even if deep-frozen. Rate-determining factors for sample degradation usually include the rates of hydrolysis, of photolysis, and of oxidation.

Poor technique with regard to sample handling will produce unrepresentative samples and contribute to uncertainty of data. Contamination of sample may come from use of improperly cleaned containers, dirty container caps, dirty or inappropriate sampling equipment or filtration equipment, contaminated preservatives, cross-contamination between samples, exposure of samples to air, and loosely capped containers. GSP call for strict attention to appropriate SOPs both to avoid contamination, plus intensive in-house QC studies to identify contamination should it occur.

3.8 Quality Assurance for Laboratory Analyses

''Laboratory analysis'' activities encompass all activities from receipt of samples at the laboratory doors to issuance of analysis results either in a signed laboratory report or in electronic form. Every laboratory activity must have an associated QA activity. Comprehensive specifications for all laboratory QA/QC must be detailed in the DQS documents, i.e. the Quality Manual, Quality Procedures and Work Instructions or SOPs. The laboratory must implement internal quality audits to verify the effectiveness of the quality system. Internal quality audits should be carried out by personnel (e.g. the QA Officer) independent of those having direct responsibility for the activity being audited. The results of internal quality audits must form an integral part of the input to management review activities.

The primary objectives of QA/QC are to ensure that unstable samples are either stabilized or analyzed expeditiously, to ensure that heterogeneous samples are either homogenized or properly subsampled, to ensure that samples are not contaminated through the subsampling and analytical processes, to ensure that the results of analyses are both consistent and comparable to those from other laboratories, and to ensure that GLP are followed. Both sample degradation and sample contamination are controlled by strict adherence to published SOPs, accompanied by regular in-house studies designed to identify such problems. These in-house studies typically include both routine analyses of sample blanks and reference materials within each batch of analytical test samples, plus replicate analyses of a regular proportion of samples. Such a QC program is often referred to as ''batch-specific QC'', or as ''batch QC''. The choice of samples for reanalysis should be randomized, unless the analyses are automated, in which case the replication may be a fixed interval (e.g. every 10th test).

Every laboratory should have general benchmark limits or DQOs that must be met for each variable. These DQOs pertain to the key operating characteristics of analytical methods, including:

- trueness (or bias)
- precision (repeatability and reproducibility)
- detection limit
- sensitivity
- selectivity
- measurement range.

(Readers are cautioned that terms such as ''accuracy'', ''reproducibility'', and ''repeatability'' have been used with very different meanings by different schools of analytical chemistry.) These characteristics sometimes are referred to as ''figures of merit'', since they form the basis for evaluating whether or not particular methods are appropriate for some specific purpose. As one example of a DQO, Richardson and Morrison[32] reported that their laboratory has an expectance that ''over a 5-year period, precision and trueness will not exceed $\pm 10\%$ (2σ) for a given element when that element's concentration is more than 10 times the detection limit'' of the method in question. Solsky[33] noted that selective exclusion of data to achieve such DQO goals is fraudulent. Laboratory results are proven to be comparable to those provided by other laboratories via analyses of certified reference materials (CRMs) and through participation on interlaboratory proficiency studies. GSP for laboratories is usually referred to as GLP. GLP is usually verified by independent audit, often associated with an ISO-based certification process. The flowchart in Figure 3, from Agemian,[17] shows the relationship between GLP and QA.

The laboratory that performed the sample analyses must carry out final inspection and validation of the test results in accordance with the Quality Plan, or as documented in the Quality Manual, in order to complete the evidence of conformation of the test results to laboratory-specified DQOs. No test results should be dispatched until all activities specified in the documented procedures have been satisfactorily completed and the associated QC data are available and authorized. Before release of data, statistical evaluation must indicate conformance of test data with regard to all essential elements of laboratory DQOs. Written reports must identify (usually by signature) the inspection authority (usually the QA Officer or Laboratory Manager) responsible for the release of the results.

3.9 Quality Assurance for Data Transmission

Once a laboratory has completed the chemical analysis of a sample, the results are transmitted to a client, either by paper report or by electronic transfer to a computer database. ''Data transmission'' includes all those operations and activities of the laboratory reporting process and of the client data entry process. The associated QA protocols ensure that a permanent record is kept of algorithms and procedures used to check, flag, calculate, validate, censor, or otherwise manipulate the results. The QA protocols also ensure that every sample is correctly identified, that all requisitioned tests had results reported, that analyzing laboratory and analyzing methods are unambiguously identified, that any changes and corrections to the data are formally documented, and that all results are backed up at some secure location to avoid data loss on event of fire or earthquake. Data residing on computer databases should be regularly audited to confirm data integrity and data authenticity. Worthington et al.[34] have published a summary of audit requirements for environmental data systems. Their paper includes both a checklist and recommended plans for corrective action.

Data assessment is an important subelement of the data transmission process. Each result is individually assessed against the DQOs. Values which meet these DQOs are accepted, values which are marginally outside the DQOs are accepted with some appropriate qualifier or flag, and values which are severely fail the DQOs are rejected. Figure 4 presents a guidance flowchart for data assessment. After the data assessment process, data usually are flagged with single-character qualifier codes to indicate the data quality. Since currently there is no consistency among various agencies in the use of such qualifier codes, reports including such codes should clearly define any qualifier codes employed. Several example sets of data quality codes are given in Table 5.

GSP call for analytical results to be reported uncensored and with unambiguous estimate of analytical error. Nonetheless, many laboratories prefer to report data which are both left-censored and without any accompanying statement of analytical precision. Fyles et al.[37] have published a detailed overview of this problem. They reported that a recent survey of environmental laboratories in Canada, Europe, and the Pacific Northwest United States found that both digit truncation (i.e. rounding or simple truncation of numerical values) and distributional truncation (i.e. replacement of a numerical value by text such as ''nd'' for ''not detected'') is nearly universally practised. Fyles et al.[37] also demonstrated how both types of truncation genuinely do result in loss of information. Horwitz[38] noted that both random and nonrandom components of laboratory uncertainty should be reported, and explained why chemists are loathe to do this because of the very large magnitude of uncertainty for trace and ultratrace analyses. Byrialsen et al.[39] suggested that the uncertainty for a measurement cannot be determined with a degree of scientific objectivity equivalent to that for the measurement itself. The arguments in the literature both for and against left-censoring have been summarized by Clark and Whitfield.[19] Porter et al.[8] called for reporting of measurement precision, and also listed other papers making similar recommendations. Even when laboratories do report analytical precision,

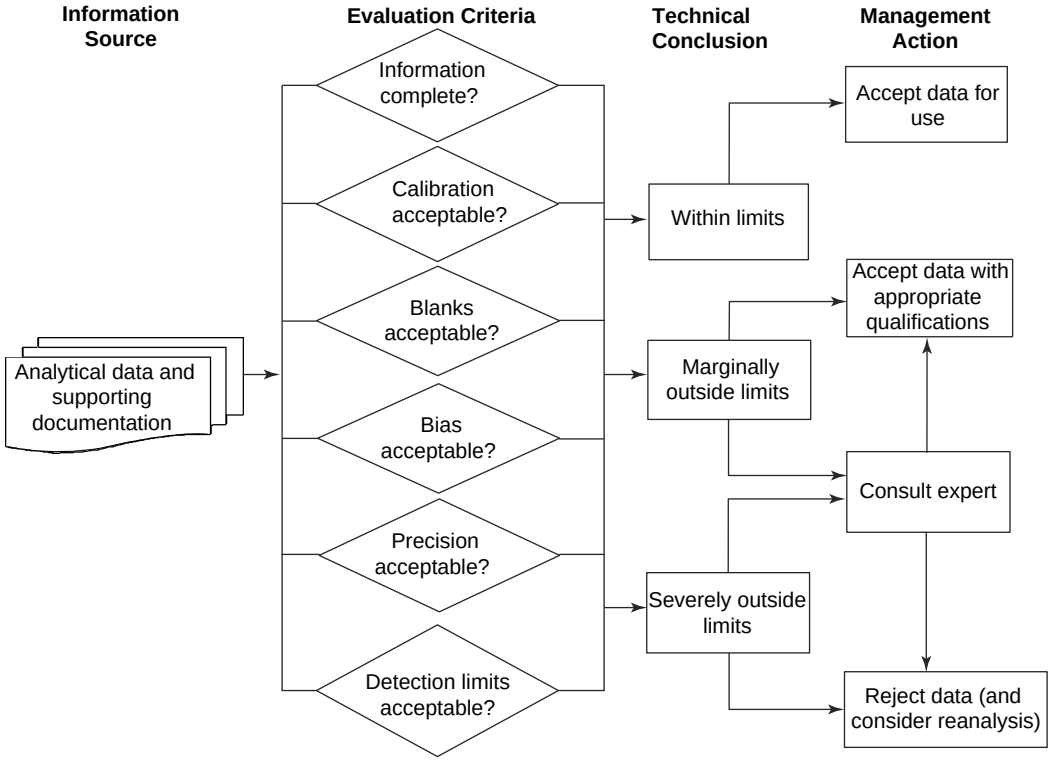

Figure 4 Guidance flowchart for assessment of data quality. (Reproduced with permission of the US Environmental Protection Agency, Region 10, Seattle, WA, from PTI Environmental Services, 'A Project Manager's Guide to Requesting and Evaluating Chemical Analyses', EPA Contract No. 68D80085, 1991.)

various algorithms are employed. Therefore, it is important that laboratories should unambiguously define the precision algorithm used. Currently the 95% CI appears to be gaining support to become the standard algorithm to be employed for expression of precision. (Personally the author favors the 95% nonparametric CI, but that is not a popular choice).

Each method of chemical analysis has a range of concentrations for which it can produce reliable results for a specific chemical matrix. By IUPAC definition, this is known as the ''region of quantitation'' and is bracketed by the ''limit of quantitation'' (LOQ) and the ''upper limit of quantitation'' (ULOQ), as described in Figure 5. A continuing problem with regard to data transmission is how to cope with results that fall below the LOQ. Many laboratories define a detection limit or ''LOD'' and report data above the LOD as if quantitative. Both LOD and LOQ are based upon σ, and thus are directly related to reliable estimation of measurement uncertainty. The term ''method detection limit'' (MDL) is synonymous with LOD. For results below the LOD some laboratories transform the values to a text entry such as ''nd'' (for ''not detected''). Other laboratories revise the observed result to the LOD and set an appropriate data qualifier flag. Still other laboratories report down to the lowest increment reading on the analytical instrument and flag values either ''T'' below the ''criterion of detection'' (see Table 5 for definition), or ''W'' below the lowest increment reading. Many laboratories prefer to transform or to flag results below the LOD in order to emphasize their lack of confidence in the numeric results. However, as summarized by Clark and Whitfield[19] and Fyles et al.,[37] such ''left-censoring'' does result in information loss and inconvenience. When censored data must be discarded as unusable, the effort expended to collect and analyze those samples has been wasted. For situations where censored data are usable, there still is some wasted effort, since many additional samples are required to reach statistically meaningful decisions. (Moreover, where analytical chemistry fails to provide estimates of the concentrations of target elements or compounds, risk assessors will turn to more dubious sources of information such as economic production or disappearance data, with worst-case scenarios.) GSP call for all analyses to yield results that primarily fall within the region of quantitation. Therefore, an analytical method preferably should only be employed if the estimated LOQ falls well below the lowest expected experimental value or regulatory criterion. Where inappropriate methods are employed, GSP call for the results to be unambiguously identified as suspect.

Table 5 Examples of data quality codes employed by different organizations. (Note that the definitions have been paraphrased from the original literature for consistency and brevity.)

(a) Developed for US Department of Energy (Source: Pink et al.[35])

B	Indicates warning of potential high bias since the analyte was found in both sample and associated analytical blank
E	Warning of possible low bias since concentration exceeded upper limit of analytical range
D	Indicates extra dilution step prior to analysis
J	Estimated value
T	Value reported is below criterion of detection, i.e. value does not differ significantly from zero but exceeds smallest increment of analytical instrument
U	Analyte was analyzed but not detected
*	QC samples associated with the result failed the control limits set in the analytical protocol.
W	Value is less than the smallest increment of analytical instrument

(b) US Environmental Protection Agency (Source: PTI Environmental Services[20])

D	Dilution step prior to analysis
E	Value estimated because of interference
J	Analyte positively identified but numerical value reported uncertain
M	Duplicate injection precision was not met
N	Predigestion matrix recovery beyond control limits
R	Data rejected and should not be used for any purpose
S	Value was determined by method of standard additions and associated value is as reliable as unqualified results
U	Analyte not present above detection concentration
UJ	Both U and J codes apply
W	Postdigestion spike recovery beyond control limits for graphite furnace atomic absorption spectroscopy
*	Duplicate analysis beyond control limits
+	Value was determined by method of standard additions and correlation coefficient is <0.995

(c) Puget Sound Estuary Program, USA (Source: PTI Environmental Services[20])

C	Analyte is combined with unresolved substances
E	Estimated value
G	Analyte is present at concentrations greater than value reported
K	Analyte is present at concentrations below detection limit reported
L	Analyte is present at concentrations less than value reported
M	Value reported is a mean of several analyses
Q	Result is questionable
R	Rejected value
T	Analyte present but detected below the detection limit reported
U	Analyte undetected for detection limit reported
X	For isotope dilution, recovery was below 10%
Z	Reported result was blank corrected

(d) British Columbia Ministry of Environment Lands and Parks, Canada (Source: British Columbia Ministry of Environment, Lands and Parks[36])

A	Result passes all validation tests
B	Result is missing for associated CRM information, or is lacking associated CI and detection limit information, or the result has been either truncated or left-censored.
C	The batch-specific QC data are either missing or more than 4 months old, or replicate values within the batch-specific QC vary by more than 25%.
D	Preservation or collection media codes were omitted
E	The sample was held too long before being analyzed
F	Sample identifier information is missing or the result and CI are not numerical values
L	Legal sample

CI, confidence interval.

3.10 Quality Assurance for Data Validation

The ''data validation'' element of environmental analysis includes validation checks of accumulated data sets to confirm that DQOs have been achieved. It also includes validation checks of field and laboratory QA/QC results associated with those accumulated data sets. Data validation confirms that data sets are accurate, complete, and defensible for the intended use. The DQS component of data validation should specify what methods and

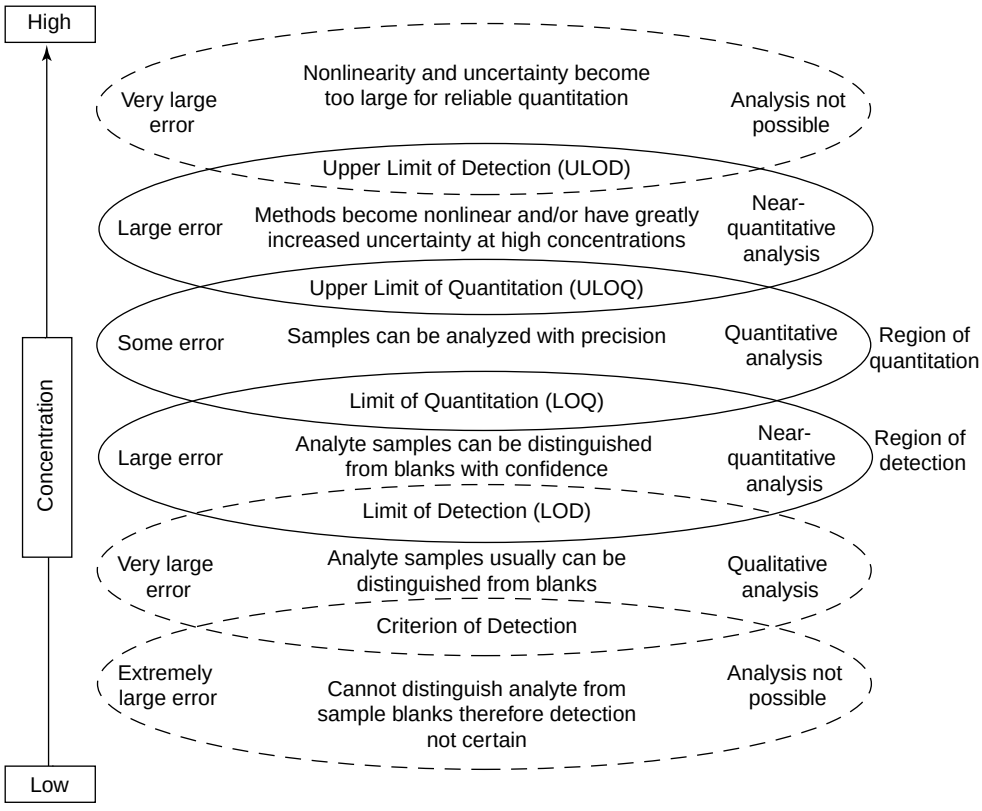

Figure 5 Schematic presenting IUPAC-style detection limit definitions. (Reproduced with permission of the American Water Resources Association from M.J.R. Clark, P.H. Whitfield, 'Conflicting Perspectives About Detection Limits and About the Censoring of Environmental Data', *Water Resour. Bull.*, **30**(6), 1063–1079 (1994).)

checks are to be employed to check the data, what DQOs employed, and what remedial actions to be taken when data fail validation. The characteristics of the data sets evaluated usually should include precision, trueness, representativeness, completeness and interlaboratory comparability. It is preferable that all data results should be fully validated. Where all results are not fully validated, it is important that the degree of validation be clearly indicated. Currently no standard set of validation codes has been established. Three example sets of validation codes are given in Table 6. Note that (a) data inspection and validation by a laboratory's QC process are required before results are released, (b) data assessment of individual results is required before data are added to a computer database, and (c) data validation of data sets (multiple results) is required prior to use and interpretation of the data sets. It is particularly important to note that these are three independent processes. The fact that one of them was performed is never an excuse for either of the other two to be omitted.

3.11 Quality Assurance for Data Approval

''Data approval'' refers to the process by which an agency gives formal acceptance to data, prior to their release to other organizations or to the public. Often the terms ''validated'' and ''preliminary'' are used to distinguish between data that have been officially approved for release and that are believed to be credible, and data that have not been officially released and that may or may not be credible. Senior management (i.e. having executive responsibility) must ensure that the Quality Plan for the overall environmental analysis process documents procedures for the inspection, validation, and approval of test data in order to ensure that the specified requirements in the DQOs are met. Records of validation must be kept. Where data fail to meet DQOs, the procedures for control of nonconforming data shall apply. This might include rejection of results, or acceptance with a warning flag. If preliminary data are released to other parties or to the public, the definition of the term ''preliminary'' and the implications for data quality and data interpretations must be clearly defined for the benefit of the data recipients.

3.12 Quality Assurance for Data Provision

The ''data provision'' element covers the release and distribution of environmental data sets, whether to the public, to university scientists, to commercial business, or

Table 6 Examples of data validation codes employed by different organizations. (Note that the definitions have been paraphrased from the original literature for consistency and brevity.)

(a) Developed for US Department of Energy (Source: Pink et al.[35])	
Quality level 1	The quality of the data set is established and meets the objectives of the study
Quality level 2	The quality of the data does not consistently meet the objectives of the study but the results are usable; additional information may be required to confirm the results
Quality level 3	High probability of false positives and false negatives; data should not be used to draw conclusions as to presence of environmental problems
(b) US Environmental Protection Agency (Source: PTI Environmental Services[20])	
Level 1	100% of the data including laboratory QC data are independently validated using the DQOs established for the project
Level 2	20% of sample data and 100% of laboratory QC samples are validated
Level 3	Limited review based upon summary results
Level 4	No validation is performed beyond that routinely performed by the analytical laboratory
(c) Prepared for the Institute of Ocean Sciences, Sidney, Canada (Source: Fyles et al.[40])	
Rating 0	Data are found to have errors and major discrepancies
Rating 1	Data are not internally consistent; patterns or trends within the data probably are not real
Rating 2	Insufficient data provided to assess the reliability of the data set
Rating 3	Data are consistent internally but are not standardized. Data may be used to evaluate internal patterns and trends but may not be used for comparisons to other data sets
Rating 4	Data are both internally consistent and standardized. Data set may be compared with other data sets of this rating

to government agencies. The QA associated with data provision should ensure that the results are perceived by all parties to be complete, accurate, believable, timely, and understandable. There is increasing recognition of the necessity for comprehensive and efficient mechanisms to report both the data results of environmental analyses, plus also the judgmental information as to the quality and usability of data. Pink et al.[35] reported the development of an approach which includes assessing, documenting, and reporting what is known about the data quality and consequently the usability of data. Their approach integrates field data, analytical results, QC data, and historical data. As the world moves more and more to electronic (''high-tech'') paperless exchange of information, QA must include provision for the security of the integrity of data sets.

3.13 Quality Assurance for Statistical Analyses

''Statistical analyses'' of environmental analysis data range from simple hypothesis testing methods to sophisticated time-series analyses and forecasting techniques. Before statistical evaluations commence, targeted data sets should be examined for reliability caution flags, lack of completeness, anomalous values, left-censoring, data truncation, nonhomogeneous variance, pseudoreplication, autocorrelation, serial correlation, and nonnormal distribution of random error – all these problems being commonplace for environmental data sets. Long-term environmental analysis data sets must also be examined for artifacts introduced by changes in frequency or timing of sample collection, and by changes in sampling methods or in analytical methods. In addition, the sampling uncertainty, the analytical uncertainty, and the total assay uncertainty must all meet the study DQOs and be significantly smaller than the local environmental uncertainty. Ideally both the statistical techniques to be employed and also the responses to each of the aforementioned problems should be predefined in a series of study DQOs developed (with the assistance of a qualified statistician) prior to initiation of the project. Clark and Whitfield[19] and Fyles et al.[37] have summarized the literature regarding statistical methods to cope with the commonly encountered problem of left-censoring.

GSP call for correct statistical methods to be effectively applied. However, Millard[41] reported that there is widespread misuse of statistics in environmental monitoring applications. Millard attributed this problem to a general failure of environmental agencies to hire professional statisticians. (Millard additionally expressed concern that the widespread availability of statistical packages for personal computers and the growing trend towards expert systems seem likely to exacerbate this problem.) For hypothesis testing, GSP call for α to be always reported, and also specifically call for β to be reported whenever a statistical analysis does not reject a null hypothesis (α is the probability of type I error, the rejection of a true null hypothesis; β is the probability of type II error, the failing to reject a false null hypothesis). Peterman[42] reported in 1990 that some 98% of the papers he had surveyed in fisheries and aquatic sciences had failed to report β whenever a null hypothesis had not been rejected. Peterman also reported that 52% of papers

he surveyed had incorrectly drawn conclusions for those situations as if the null hypothesis had been proven to be true. To ensure that appropriate statistical procedures are employed to cope with the statistical aspects of environmental analysis, independent audit by a professional statistician is strongly advised.

3.14 Quality Assurance for Reporting

Decision makers prefer reports which are concise, which make explicit recommendations, and which clearly identify important variables. Also, as noted by Cullen,[43] it is important to appreciate that conflicting advice from otherwise credible sources tends to be disregarded. While it is important that scientists understand the fine details comprising an environmental analysis, many of the people who encounter the published reports will be nonscientists. Therefore, summary-type explanations should be emphasized. Conversely, scientists reading the reports will want the detail, preferably in machine-readable form. Short interpretive reports accompanied by separate technical appendices can be both effective and cost efficient.

QA for the reporting element of environmental analysis has five main objectives:

1. to ensure that the overall environmental analysis program from start to finish is credible;
2. to ensure that the reporting presentation is relevant to known environmental problems and pending resources management decisions;
3. to ensure that the findings are issued on a timely basis;
4. to ensure that the interpretive reports are understandable to nonscientists; and
5. to ensure that total assay results and associated QA/QC results are available to other scientists and secured on a permanent archive.

4 ERROR CONTROL

All scientific measurements are subject to error. The goal of ''good measurement practice'' is to minimize errors sufficiently that they do not adversely affect resource management decisions. Samples collected from some particular population in any particular ecosystem must be so chosen, so collected, so preserved, so analyzed, so reported, and so interpreted such that the results do truly and accurately represent the entire population being characterized. Wagner[16] discussed this topic in some detail, and emphasized that ''The main problem for QA in environmental sampling is how to guarantee biological, ecological and geographical representativeness of the samples''.

As indicated in Figure 6, there are many sources from which errors may arise during the environmental analysis process. The nomenclature of errors usually refers either to the general source of the errors or to the character of the errors. Where discrete samples are not taken, usually the errors are simply referred to as ''measurement errors''. If discrete samples are collected and subsequently taken elsewhere for analysis, ''sampling errors'' are those errors which arise from the instant of sampling until the instant the laboratory takes receipt of the sample. ''Analysis errors'' are those errors which arise after the instant a sample arrives at the laboratory. ''Total assay error'' is the sum of sampling error and analysis error, as discussed by Clark et al.[27] With regard to the character of errors, there are three general categories of error. These are ''random errors'', ''nonrandom errors'', and ''gross errors''. Wagner[16] has published a useful summary of error characteristics, quality, and magnitude, although using a slightly different nomenclature to that employed here. Wagner's summary is reproduced in Table 7. It should be noted that the term ''uncertainty'' is often employed as synonymous to ''error''.

4.1 Random Errors

Even the most carefully performed measurements are subject to small irregular random errors. When the random error results from numerous sources of small errors, it is reasonable to assume (because of the central limit theorem) that the distribution of random error is normal. The assumption that errors are primarily random in nature forms the essential foundation of modern quantitative statistics. This assumption is also fundamental to the techniques described by Ramsey et al.[26] and Clark et al.[27] for determining the relative proportions of analytical error, sampling error, and local environmental heterogeneity by variable within environmental analysis data sets. Those data sets influenced primarily by random error (i.e. minimally influenced by nonrandom error) are usually described as being ''in control'' or ''in statistical control''.

It is important to appreciate that an analytical method may not yield usable results even when that analytical procedure is fully under control with no gross errors, no observable nonrandom error, and minimal random error. For example, an environmental analysis DQO for some specific purpose might set a precision limit at $\pm 10\%$ for all analytes. However, according to Cantillo and Lauenstein,[44] laboratories typically set limits of precision DQOs for trace organic control substances at $\pm 30\%$ on average for all analytes and $\pm 35\%$ for

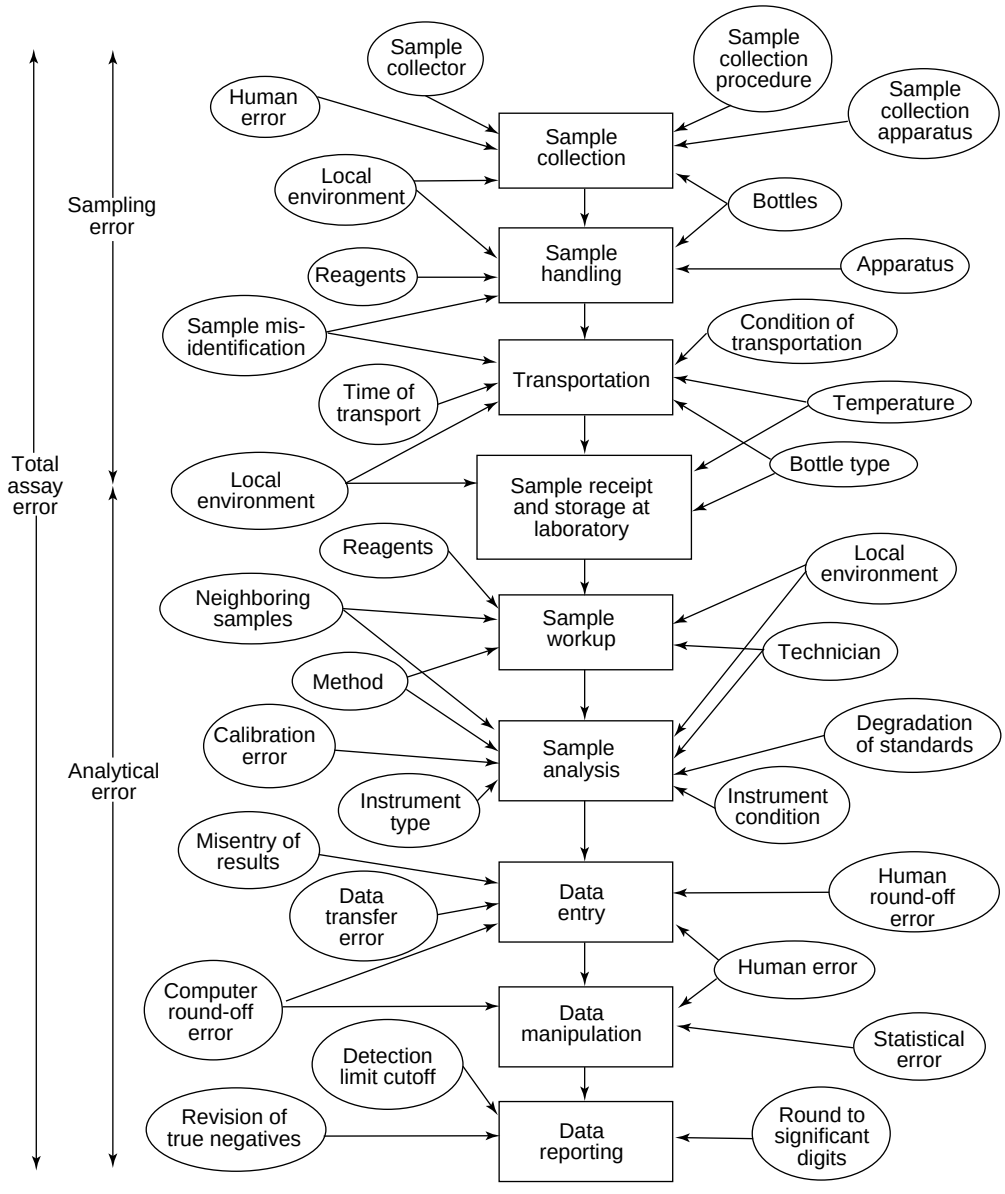

Figure 6 Multiple sources of total assay error. For discrete samples, ''sampling errors'' are those errors which arise from the instant of sampling until arrival of the sample at a laboratory. ''Analysis errors'' are those errors which arise after the instant a sample arrives at a laboratory. (Reproduced with permission of the American Water Resources Association from M.J.R. Clark, P.H. Whitfield, 'Conflicting Perspectives About Detection Limits and About the Censoring of Environmental Data', *Water Resour. Bull.*, **30**(6), 1063–1079 (1994).)

individual analytes. These data would be unsuitable for the intended purpose even though fully in control. This problem of finding suitable analytical methods becomes more difficult for ultratrace analyses. From retrospective investigation of a number of interlaboratory proficiency studies, Horwitz et al.[45] have determined that the interlaboratory precision expressed as coefficient of variation (CV) is an inverse function of concentration expressed on a mass/mass basis, as illustrated in Figure 7. In other words, the lower the concentration of analyte in a given chemical matrix, the greater is the random uncertainty of analytical results.

4.2 Nonrandom Errors

Nonrandom errors are not random and usually are not normally distributed. As a class they confound statistical analysis. Therefore, an important role of QA is to recognize and to remedy any occurrence of nonrandom error. There are five important types of nonrandom

Table 7 Characteristics, quality and magnitude of errors categorized by different elements of environmental analysis monitoring. (Note that the nomenclature used in this table is slightly different from that employed in the main article and also that the table was modified slightly from the original for clarity. Reprinted from G. Wagner, 'Basic Approaches and Methods for Quality Assurance and Quality Control in Sample Collection and Storage for Environmental Monitoring', *Sci. Total Environ.*, **176**, 63–71 (1995), Copyright 1995, with permission from Elsevier Science.)

Procedure and step	Main source and characteristics of possible errors	Quality of errors[a]	Quantitative estimation (worst case)[b]	Risk of serious errors
Planning				
• Definition of the area	Spatial variability, heterogeneity (hot-spot effects)	Systematic + random	Factor 2/2 or more	High
• Selection of specimens	Ecological or physiological variability	Systematic	Factor 5/5	Moderate
• Stratification	Biological, physiological or spatial variability	Systematic	Factor 3/3	Moderate
• Sampling method	Statistical and physiological representativeness, skewed distribution, contamination or loss	Random and/or systematic	Factor 2/2 or more	High, partly controllable
• Number of samples	Statistical representativeness	Random	±50%	High
• Sample mass	Statistical representativeness	Random	±few %	Normally low
• Timing	Temporal variability, trends	Random or systematic	±30% and more	High
Sampling				
• Weather conditions	Irreproducible deposition, leaching or matrix effects	Systematic	±50%	Very high
• Packaging	Contamination of loss by tools and container material	Systematic	±100% or more	Controllable
• Sample conservation during the sampling	Losses by metabolism, volatilization or translocation (most serious for air filters, water and animal tissues)	Systematic or absolute	−100%/+few %	Moderate
Transportation	Contamination or loss by volatilization (most serious for soil and water samples)	Systematic–absolute	±100% or more	High
Storage				
• Short-term, long-term	Contamination or loss, metabolism, alteration of binding form or weight basis, speciation, solubility	Systematic–absolute	±100% or more	High–very high
Sample preparation				
• Cleaning, washing	Contamination or loss by leaching	Systematic	±100% or more	High
• Drying	Loss, contamination	Systematic	±100% or more	Moderate
• Homogenization	Contamination, disregard of skewed distribution	Systematic	±100% or more	High
• Subsampling, aliquotation	Particle and analyte distribution and heterogeneity, statistical representativeness	Random	±50%	Moderate
Sample pretreatment				
• Digestion, extraction, separation, or concentration	Contamination by reagents or container material, loss by adsorption, precipitation or coprecipitation	Systematic	±100% or more	Controllable
• Matrix modification	Contamination by reagents	Systematic	±100% or more	Controllable

Table 7 (*continued*)

Procedure and step	Main source and characteristics of possible errors	Quality of errors[a]	Quantitative estimation (worst case)[b]	Risk of serious errors
Analysis				
• Injection	Inaccurate or badly adjusted tools	Random or systematic	Few %	Moderate
• Calibration	Physical interferences	Random or systematic	Some %	Moderate
• Detection	For example, spectral interferences	Systematic	±100% or more	Low
• Peak identification	Automatic evaluation	Qualitative	Qualitative	Low
• Quantitation	Baseline shift	Random or systematic	Few ‰ or %	Very low
Data evaluation				
• Averaging	Disregard of asymmetric distribution	Systematic	Up to 50%	Moderate
• CI	Disregard of asymmetric distribution	Systematic	Up to 50%	Moderate
• Trend detection	Disregard of natural variability and fluctuations	Qualitative	Qualitative	High

[a] "Absolute error" is the sum of absolute deviations from a target value divided by the number of measurements.
[b] The term "factor 2/2" indicates that the true result is between 1/2 and 2 times the reported value.

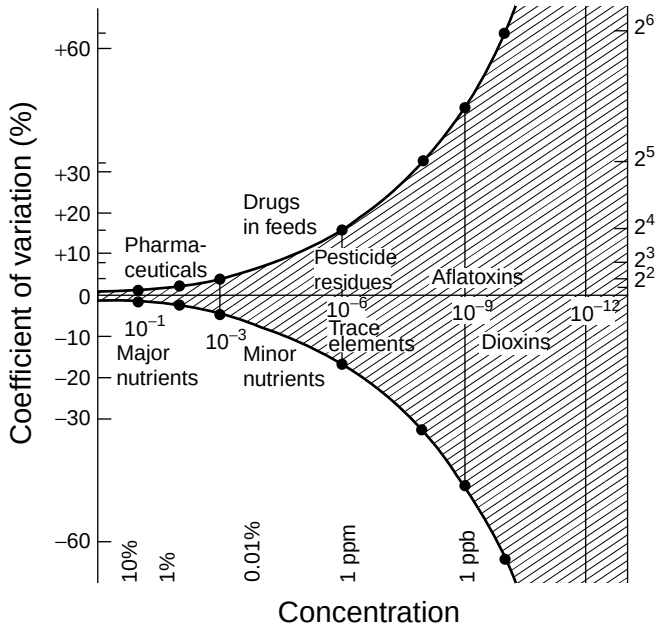

Figure 7 Interlaboratory CV as a function of mass-to-mass concentration. (Reprinted with permission of AOAC International from W. Horwitz, L.R. Kamps, K.W. Boyer, 'Quality Assurance in the Analysis of Foods for Trace Constituents', *J. Assoc. Off. Anal. Chem.*, **63**(6), 1344–1354 (1980). Copyright 1980, by AOAC International.)

error commonly encountered for environmental analysis. The first type is very similar to random error in that the overall error is the sum of small errors from many different sources. However, one or more of the nonrandom sources of error are of sufficient magnitude so that the overall error distribution is not normal. This situation is best identified by testing error distributions for normality. Appropriate distribution-free (sometimes called nonparametric) statistics often can cope with this situation. (Random error may appear to be non-normal if transformed or collected in inappropriate units; transformation will return the data distribution to normal but cannot rectify nonnormal error. Distribution-free statistics is preferable to transformation if there is doubt as to whether the data distribution is truly or apparently non-normal.) The second type of nonrandom error is referred to as "systematic error" because it shows some reproducible pattern. This type of random error is usually recognized through use of CRMs, from results of interlaboratory proficiency studies, or from comparisons of independent methods of analysis. Where the systematic error is highly reproducible, it may be possible to correct it by either calculation or instrumentation. This is the basis of blank correction. A third type of nonrandom error occurs as infrequent transient events where the error is of similar magnitude to the measurement being taken. This type of error is recognized through the use of statistical outlier tests. A fourth type of nonrandom error ("persistent nonrandom error") both occurs frequently in a data set and is of similar magnitude to the measurements themselves. Like systematic error, this type of random error is usually recognized through use either of CRMs or of independent methods of analysis. Data sets containing persistent nonrandom error are not in statistical control and are best discarded, with more appropriate assay methods being used thereafter. Data sets having a majority of results below the LOQ are particularly vulnerable to persistent nonrandom error. This type of error also occurs for data

sets based on methods that are susceptible to interference or that are inadequately selective to the analyte of interest. A fifth type of nonrandom error has periods of random error intermixed with periods of persistent nonrandom error. This often occurs for situations where one or more series of measurements above the LOQ are intermixed with series of measurements below the LOQ. It may be possible to segregate the trustworthy from the nontrustworthy results in such situations.

In classical stoichiometric chemistry (concentrations $>0.1\%$), systematic error is viewed as the predominating type of nonrandom error, and also is viewed as being generally correctable. Much of the literature on stoichiometric chemistry treats the term ''systematic error'' as synonymous with ''nonrandom error''. Analytical error was viewed as being composed of a random error component plus a systematic component or bias, and the bias could be estimated and subtracted. It must be emphasized that the simple model of errors employed for classical stoichiometric chemistry is inappropriate for modern trace analysis (concentrations $< 0.1\%$). Although the five types of nonrandom error offer a convenient classification model for trace analyses, often it is not possible to classify the source of discrepancies in environmental analysis results. Often different sources of error are confounded with each other.

A question that often arises during discussion of errors is whether or not statistical outliers should be removed from data sets. GSP permit the removal of statistical outliers only if anomalous experimental conditions can be documented. Removal of statistical outliers without proof of anomaly is controversial, but Grant et al.[46] argued that the removal of single aberrant values can be justified. Multiple outliers are frequently encountered in environmental analysis, and often there is considerable incentive to remove multiple outliers without proof of anomaly, since leaving outliers in data sets distorts the latter, and tracing their origin is a never-ending, expensive task. Nonetheless, it is preferable to leave unexplained outliers within data sets, and to address this problem both with a reasonable (i.e. affordable) effort to trace their cause, and through the use of statistical procedures that are robust to the presence of outliers. In this regard, it should be noted that outliers have a considerable impact on statistical parameters such as mean and standard deviation, but little impact upon statistical parameters such as the median or the inner quartile range (range from 25th to 75th percentiles).

Samples that truly do not contain any analyte of interest but which are reported to contain the analyte are called ''false positives''. Samples that do truly contain the analyte of interest but are reported to have no analyte present are called ''false negatives''. It should be appreciated that the detection limit model always permits some false positive and false negative values. However, the proportion of false positives and false negatives should be that permitted by chance alone. If nonrandom error predominates over random error (as it well might near the detection limit, or for sample matrices which differ substantially from the matrix of the calibration samples), then the occurrence of false positives and false negatives is not controlled. For such a situation it is even possible for a majority of observed results to be false positives or false negatives. Interpretation of such untrustworthy data sets can yield misleading conclusions.

4.3 Gross Errors

Gross errors result from accidents, equipment malfunction, or human blunders. These errors are of large magnitude, and will greatly distort data interpretations if not identified. An amusing element of gross outliers is the great variety of names assigned to this type of error by various authors, e.g. blunders, catastrophic errors, flagrant outliers, fliers, gross errors, lurking outliers, obvious blunders, outlandish results, outrageous outliers, and rogue values. Fortunately, the large size of these errors often makes them easy to spot. These errors particularly stand out from routine results on graphic plots. They may also be identified as statistical outliers by appropriate statistical tests. The most confusing type of gross error stems from simple misidentification of samples. It must be emphasized that even the most careful program is vulnerable to gross errors, and one should never give in to the temptation to relax QA programs to economize. Unidentified gross outliers can so greatly distort data as to render them meaningless.

5 DETECTION LIMITS AND LIMITS OF QUANTITATION

5.1 The International Union of Pure and Applied Chemistry Detection Limit Model

GSP call for all environmental samples to be analyzed by analytical methods that have a definite relationship (often but not always linear) between observed and true values, and that also have analytical error that is minimal, homogeneous, and random. Most analytical procedures specify a concentration range for which such a definite relationship has been established, within allowances for analytical error. For measurements made outside this concentration range, the relationship between ''true'' and observed concentrations of analyte can become nonlinear. Both random and nonrandom analytical errors can grow rapidly larger with increasing distance from the linear range. However, with the growing importance of trace and ultratrace analyses of environmentally significant

contaminants, a number of analyses are now performed at concentrations substantially below the lower extreme of the quantitative range. The literature contains numerous procedures, definitions, and algorithms for determining analytical boundaries (detection limits and quantitation limits) to distinguish between data that are meaningful, somewhat meaningful, and not meaningful. A 1990 review by Oresic and Grdinic[47] on this topic included over 500 citations.

As discussed by Long and Winefordner,[48] IUPAC has issued a standard set of definitions regarding detection limits. Figure 5 presents a schematic representation of a detection limit model that conforms to definitions endorsed by IUPAC. For this model, the most reliable quantitative data fall in the ''region of quantitation''. The lower boundary of this region is called the ''LOQ''. Below the LOQ boundary, there is a range of concentrations called the ''region of detection'' for which there is a high probability that samples genuinely containing the analyte of interest can be distinguished from samples (called ''sample blanks'') of similar matrix which genuinely do not contain the analyte of interest. The lower boundary of the region of detection is called the ''LOD'' (one often hears this boundary referred to as the ''detection limit''). Below the LOD, the probability of distinguishing between samples which genuinely contain analyte and sample blanks falls rapidly with decreasing concentration. That concentration below which test samples and sample blanks cannot be distinguished is called the ''criterion of detection''. There are a number of flaws in the detection limit concept, as discussed by Kurtz et al.[49] and Clark and Whitfield.[19] Not the least of these is the fact that a number of laboratories do not precisely follow the IUPAC approach, even though they may use similar nomenclature. The detection limit concept as correctly applied offers a useful tool towards choosing analytical methods appropriate to the objectives for programs of environmental analysis. The Kurtz et al.[49] paper includes advice from a panel of scientists on how data at or near the detection limit should be produced and reported. Their paper includes four real-world examples, and particularly advises that reported results at or near the detection limit should have statistical support.

The detection limit concept was developed for the specific intention of comparing analytical instruments and methodologies. Estimated LODs and LOQs are only meaningful for interference-free conditions. Clark and Whitfield[19] advised that results below the LOQ should not be assumed to be in statistical control. GSP call for such assumptions to be proven. This is particularly important when test samples have greatly different chemical composition (matrix) than the calibration samples. To avoid problems with regard to nonrandom error (including false positives and false negatives), all experimental analyses should fall well above the LOQ. Many protocols define the LOQ as some simple multiple (usually a multiple of 3 or 10) times the LOD. Winefordner and Rutledge[50] even suggested that a method should only be considered for application if the calculated detection limit is at least 100 times below the lowest experimental values expected.

5.2 Calculation of the Limit of Detection

Above the LOD, there is a statistical probability that samples that genuinely contain a specific analyte can be distinguished from samples that do not contain that analyte. When reaching such a statistical decision, there are two ''errors of decision'' which must be considered. For a type I decision error, one would have decided that there was analyte present in a sample when in fact that sample contained no analyte. For a type II decision error, one would have decided that there was no analyte present in a sample, when in fact the analyte was present. Depending upon the use of the results, sometimes one decision error is much more important than the other, and only the single important decision error is statistically tested. For other purposes both decision errors are equally important and both decision errors are tested.

If one assumes that analytical error is random, normal in distribution, and identical for test samples and blanks at the LOD concentration, then fairly straightforward decision rules can be established to ascertain the LOD. For example, if one repeatedly analyses a very low concentration of analyte in a specific sample matrix, and if one is willing to accept a 5% chance of a type I decision error and uncontrolled type II decision error, then the LOD will be 1.65 times the standard deviation of these results for a large number (>30) of analyses (Equation 1):

$$\mathrm{LOD} = 1.65s \tag{1}$$

(This is because for a normal distribution 95% of the area falls below the mean plus 1.65σ and 5% falls above that limit. The population standard deviation, σ, is represented by the sample standard deviation, s.) For a small number of analyses the LOD should be based on the t-distribution (Equation 2):

$$\mathrm{LOD} = t_{0.95}s \tag{2}$$

If one wished neither type I decision error nor type II decision error to exceed 5%, then for the same circumstances, the LOD would be estimated from Equation (3) for a large number of analyses or from Equation (4) for a small number of analyses:

$$\mathrm{LOD} = 3.29s \tag{3}$$

$$\mathrm{LOD} = 2t_{0.95}s \tag{4}$$

A large number of papers have been published regarding the derivation and application of the algorithms to calculate LODs. Adams[51] has published a practical summary of the topic.

5.3 Different Types of Limits of Detection

There are a number of particulars that substantially impact the final estimate of LODs and must therefore be detailed in SOPs and publications. First, is the LOD for an instrument exclusive of sample preparation and pretreatment? Or is the LOD for the entire method including all digestions, extractions, and other preparation steps? Or does the LOD include some but not all of the sample preparation and pretreatment steps? Second, what are the analyte concentrations and characteristics of the test matrix used to determine the LOD? The four usual preferences are deionized water, deionized water spiked with low concentrations of the target analyte, test samples having very low concentrations of target analyte, or test samples spiked with low concentrations of analyte. Third, are both type I and type II errors to be equivalently controlled, is one to be more greatly controlled than the other, or is only one to be controlled? Fourth, over what time frame were the test samples analyzed – all in one day would give a rather overoptimistic estimate of LOD. The two most common strategies are either analysis of multiple sets of replicate aliquots over several weeks or analysis of a large number of duplicate samples over many months. Finally, were the analyses performed by more than one instrument, more than one technician, more than one laboratory? All this information has substantial impact on the results, and therefore must be specified in the laboratory SOPs and permanent archive records.

As noted previously, false positives can occur by random chance alone for a concentration near or below the detection limit. Also, nonrandom error is usually difficult to control at these low concentrations. This usually increases the number of false positives. The problem of false positives is of particular concern to agencies dealing with potent carcinogens or toxicants, where even a single false positive might cause undue public alarm. For these situations, some agencies and laboratories arbitrarily adjust the detection limit upwards to what is often called a ''practical detection limit'' (PDL). A similar strategy is sometimes taken to resolve difficulties with merging data from different laboratories having different LODs. The PDL strategy does substantially increase information loss through left-censoring. For merged data of different LODs, the worse information loss paradoxically is for the best (most sensitive) methods.

6 ASSOCIATED TOPICS

6.1 Quality Assurance Requirements of Legal Samples

Some samples are collected with the intent that the results will be used in a court of law, whether to enforce government regulations or for the purpose of other litigation. Such ''legal samples'' require rigorous adherence to protocols additional to those required for routine environmental monitoring. Both persons who collect the samples and persons who analyze the samples must be able to document appropriate training for duties performed with regard to the samples, and must be able to document and testify that the appropriate technical protocols were rigorously followed. These persons must also be able to testify as to results of the QA/QC programs associated with the legal samples. They also must be able to prove that neither sample containers nor reagents employed could have contaminated the samples.

Both persons who collected the samples and those who analyzed the samples must be able to establish, without doubt, that the samples are correctly identified and could not possibly have been tampered with, contaminated, or substituted. Secure possession of the samples must be documented from the moment of sample collection (and earlier with regard to reagents, collection bottles and sampling equipment) until results are reported in court. This secure possession must be maintained on a continuous basis as samples are handled by various parties, and is referred to as ''chain-of-custody''. Every person who comes into possession of a legal sample should be able to identify the sample and be prepared to testify where he or she received it, when, and from whom. As each person assumes possession of the sample, he or she should mark the date, time, and his or her initials on the sample or to an attached tag. Each person receiving a legal sample should record separately, in a bound notebook kept for this purpose, the sample identification information, date and time of sample receipt, and from whom the sample was received. Similar information should be recorded in the notebook when possession of the sample is passed to another person. Note that some statutes make provision for taking sufficient quantities of samples so that they may be split between litigants. A number of useful guides have been published as to the special steps and records required for legal samples. Environment Canada,[52,53] for example, has published useful check list guides towards maintaining chain-of-custody both in the field and during shipping of samples, and also has produced an associated CD-ROM that summarizes the respective roles of samplers, laboratory analysts, legal staff, and expert witnesses. Frank[54] published a very readable summary of the types of records courts expect laboratories to

maintain. Frank also discussed the types of problems that can result when laboratories do not maintain good record-keeping practices. Horwitz[38] noted that assay uncertainty pertains to legal requirements expressed by the terms ''preponderance of evidence'' and ''beyond a reasonable doubt''.

One caution worth noting is that if a record is introduced into evidence, then that record may not be available for other purposes for some years. Also, if a record consists of several pages extracted from a larger record, it might be possible that the other party in court might be entitled to demand the entire document. Therefore, if one is collecting evidence for several different court cases, it is preferable to maintain separate records. Also, extraneous material should be excluded. Finally, it should be kept in mind that the other side will be examining photographic evidence closely for deviations from accepted GSP.

6.2 Certified Reference Materials

CRMs are well characterized materials that have their analyte content guaranteed to be a best estimate by some institution of authority. The certificate accompanying a CRM will describe the purpose of the material, the sample matrix, and the conditions of certification. The certificate may also include other details such as recommended procedures for drying and storage of the CRM. The development and use of CRMs is included in ISO Guides 33 and 35. The development, preparation and utilization of reference materials according to the ISO guidelines is addressed in publications by Mesley et al.[4] and Kane.[55] In the USA, CRMs issued by the National Institute of Standards and Technology (NIST) are referred to as ''SRMs''. Some CRMs are available in limited numbers. GSP call for any CRMs available in limited supply to be used on an occasional basis only, for specific short-term evaluations, so as to ration the supply. They should not be consumed in routine long-term QC programs. On the other hand, some CRMs are so readily available that they can be used for instrument calibration, as described by Watters and Koch,[56] or other routine QC procedures.

Despite efforts to improve the quality of environmental measurements and assays, reports continue to be published of major discrepancies between results of samples analyzed at different laboratories. As a result, it has now become the usual practice for analytical laboratories to challenge their analytical methods against CRMs so as to determine method bias. This permits laboratories to establish that their methodologies are comparable to those for other laboratories in a manner traceable to the primary institution that guarantees the CRM. Many different reference materials are now available for use by analytical laboratories. Available sample matrices include animal tissues, plant tissues, coal, rock, sediments (freshwater and marine), soil, and water (freshwater and marine). However, the production of reference materials having chemical matrices similar to those of test samples poses a number of difficult problems, as discussed by Griepink et al.[57] A number of publications describe the history, preparation, availability, and appropriate use of CRMs, including papers by Mesley et al.,[4] Cantillo and Lauenstein,[44] Watters and Koch,[56] and Griepink et al.[57]

6.3 Interlaboratory Proficiency Testing

Interlaboratory proficiency testing (also referred to as ''round robins'' or ''collaborative trials'') are collaborative studies where a number of different laboratories (preferably six or more) analyze identical aliquots of one or more stabilized and homogenized large samples. There are two distinctly different types of laboratory proficiency test studies. The first type is intended to determine whether or not a new test method is reliable. These studies usually are of short duration and often are carried out by a small number of participating laboratories, all following identical procedures. The second type of laboratory proficiency evaluation is intended to measure the performance of individual laboratories in comparison with the results of a large number of laboratories. These studies usually consist of a long-term sequence of short-term studies, several per year. Statistics are calculated primarily for each individual study but also for the sequence of studies. Although some interlaboratory studies require all laboratories to follow the same method, usually laboratories are free to use methods of their preference. Laboratory proficiency testing often forms an important part of the laboratory accreditation evaluation. Protocols establishing the minimum requirements for the design, conduct, and interpretation of laboratory proficiency studies have been internationally accepted, based on the recommendations by consensus at the IUPAC Workshop on the Harmonization of Collaborative Analytical Studies held in May 1987 in Geneva, Switzerland.[58]

Four important features of laboratory proficiency studies are currently the subject of debate. First, should the results for individual laboratories be kept confidential by the study organizers? Second, should statistical outliers be removed prior to calculation of method bias and precision? Third, should samples be sent to laboratories on a nonblind basis? Fourth, should laboratories be considered proficient in all methods of determining a particular analyte if they only use their best methods for proficiency evaluations? (To elaborate the last question, if a laboratory has three methods in routine use to analyze some analyte, and one method is substantially

more sensitive and selective than the other two methods, should that laboratory be certified as being proficient in all three methods if only the best procedure was evaluated in a proficiency study?) Most laboratory proficiency studies have adopted all four of the aforementioned practices.

Horwitz[24,45] has reviewed the data from a number of laboratory proficiency testing studies and for a variety of sample matrices and analytical techniques. He discovered that the interlaboratory (i.e. among-laboratory) precision expressed as CV (i.e. relative standard deviation in %) is generally less than twice the values predicted by Equation (5):

$$\mathrm{CV} = 2^{(1-0.5\log C)} \approx 2C^{-0.1505} \tag{5}$$

where C is the analyte concentration expressed as a mass/mass decimal fraction. Horwitz[24] reported that this equation (presented in Figure 7) is sufficiently robust for use in checking that interlaboratory studies are themselves in control.

6.4 Quality Assurance Concerns for Field Measurements

Field measurements include both measurements made in situ and also measurements or analyses made of collected discrete samples, provided that those measurements are made in the close vicinity to a sample-collection site and soon after sample collection. There are two general categories of field measurements, namely approximate measurements and precise measurements. Approximate measurements are not quantitative and require less operator training and less QC effort. Although approximate measurements are not reliable, they do offer an inexpensive means of ascertaining major characteristics at a site. Approximate measurements are particularly useful in targeting the collection of representative samples in polluted or high-gradient environments. The major flaw to approximate measurements is that they are not reliable and can cause confusion if intermixed with reliable measurements. Therefore, an important aspect of field measurement QA is to ensure that reliable and nonreliable data are unambiguously identified and segregated.

Reliable field measurements require the same detailed QA verification that is required for laboratory measurements. All field instruments should be calibrated under laboratory conditions on a regular basis, and many also require in-field calibration. (It should be emphasized that thermometers and barometers do require calibration.) Even greater effort is required to maintain field instruments in calibration than for laboratory instruments, because of the much greater stress (e.g. vibrations, temperature extremes) encountered by field instruments during transportation and use. Therefore, intensive calibration and calibration reverification are mandatory. Ample time must be given for instruments to warm up prior to calibration and instruments must be protected from temperature extremes, particularly where electronics might be influenced by temperature. Also, many sensors need to reach steady-state conditions with the environment prior to calibration and measurement. The key elements to QA for field measurement include thorough operator training, detailed method manuals, regular laboratory-based calibration and maintenance, intensive on-site calibration checks, audit checks against second instruments and against the results of discrete samples, regular on-site QA audits by a QA Officer or independent party, and detailed records of maintenance, calibration, and use. There are numerous publications regarding SOPs for field measurements. For example, Environment Canada[52] has published SOPs that apply across Canada with regard to field inspection of fresh waters. Similarly, Wells et al.[59] have published the SOPs followed by the US Geological Survey for field measurements for streams in Texas.

Field measurements must follow the same rigid QA protocols as laboratory analyses, including data management and documentation. QC will depend heavily upon blanks, replicates and reference standards. It must be emphasized that field protocols, SOPs, and associated records must be maintained in weather-tolerant binders, and must be readily accessible in case of accidents or emergencies.

6.5 Quality Assurance for Automated Continuous Monitoring

Measurements which are automated, continuous, and in situ have many advantages over the more traditional collection of grab samples. Most importantly, continuous monitoring offers real-time analysis of analytes. This permits transient events to be examined in detail. In addition, continuous monitoring curtails difficulties associated with collecting representative discrete samples and also with sample preservation during transport and storage. This is particularly important with regard to substances that are not readily stabilized. The primary elements of continuous monitors are sensors, controllers, and data loggers. Sensors are devices which produce signals continuously in response either to physical parameters or to concentrations of chemical analytes. Controllers accept the sensor signals, perform various signal conversion and instrument control functions, and pass signal data to the data logger for temporary storage until the information is passed to a central database via a data verification and sometimes a data reduction process. A great advantage of continuous monitors is the large amount of information produced, which is extremely valuable for QA purposes. There is a huge literature

regarding the development of numerous types of sensors, both electronic and fiber-optic based. However, many potential sensors have problems of inadequate sensitivity, inadequate stability, inadequate sensor response time, or inadequate lifetime. Because of these problems, at present only a limited number of sensor types are routinely utilized for long-term continuous unattended monitoring. Camara et al.[60] have published a detailed summary of this topic, with primary focus upon fiber-optical sensors.

The key elements for continuous monitoring QA generally are identical with those for other field measurements: thorough operator training, detailed method manuals, regular laboratory-based calibration and maintenance, on-site calibration checks, regular on-site QA audits, audit checks against second instruments and against the results of discrete samples, and detailed maintenance, calibration, and measurement records. Whitfield and Wade[61] have reported from their experiences with automated monitoring in surface waters that standard QA protocols established for sampling and analysis of discrete samples are not adequate for continuous monitors. They suggest that three areas where standard protocols need to be established specifically for continuous monitoring are the problem of field verification, the matter of time controls, and the relationship between sensor accuracy (bias) and precision to frequency of instrument servicing.

6.6 Performance-based Methods

A number of jurisdictions, particularly in the USA and Canada, are adopting ''performance-based methods'' (PBMs), particularly with regard to sampling and analytical methodologies. For PBMs, a regulatory agency publishes generic methods, with certain key steps specified as mandatory. (For simple target compounds having readily available reference materials, there usually are few mandatory steps. However, for aggregate substances such as petroleum hydrocarbon mixtures, and for substances lacking reference materials, there usually are relatively more mandatory steps.) The regulatory agency also publishes DQOs (such as specific data on recovery, reproducibility, calibration validity, and MDLs) which are to be achieved. The focus of a PBM is less on the method per se, and more upon the DQOs associated with the method. Other than the mandatory steps, samplers and analysts are free to use equipment and methodologies of their choice provided that the specified performance criteria are unambiguously proven to have been achieved. The advantage claimed for regulatory agencies is that fewer method SOPs require vetting, that published method manuals do not need to be revised as frequently, and that much of the responsibility and costs of verifying methods is moved from the regulatory agencies to the sampling agencies and analytical laboratories. The advantage claimed for samplers and analysts is that they can readily change methods and equipment in accord with changes in circumstances and advances in technology. The major drawback to PBMs is the substantial initial or front-end transition cost.

Implementation of PBMs in the USA was summarized by Stevenson.[62] Stevenson noted that ''The PBM paradigm is that each member of the characterization and remediation team is responsible for his or her [own] work. ... It is the job of those involved in the planning, logistics, and storage to have data that demonstrates the entire process complies with the DQO''. The overall PBM process is described by the flow-chart in Figure 8, as it is being implemented in Canada.

6.7 Calculation of Confidence Intervals

If the population distribution for a series of measurements is normal and without nonrandom error, then the probability that the population mean will fall within an interval extending either side of the sample mean can be related to the area under the Gaussian curve within that interval. If one has a sample mean and one knows the population standard deviation, σ, then the CI would be given by Equation (6):

$$\text{CI} = \text{mean} \pm \frac{Z\sigma}{\sqrt{n}} \tag{6}$$

where Z = probability factor which depends on the level of confidence desired, σ = population standard deviation, n = number of measurements used to calculate the mean, and mean = sample mean. Usually one does not know σ but must estimate the sample standard deviation, s, from a set of measurements. There are two situations: either s is determined from the same set of measurements as the sample mean, or s is determined from another set measurements. For both situations the CI is calculated from Equation (7):

$$\text{CI} = \text{mean} \pm \frac{ts}{\sqrt{n}} \tag{7}$$

The t value (sometimes called Student's t) is always larger than Z, and corrects for the fact that the standard deviation is not known with confidence for small data sets. The value of t depends both upon the level of confidence desired and upon the degrees of freedom of the data set used to estimate the standard deviation. The mean must always be calculated from actual measurements, but the population standard deviation may be estimated either from the actual measurements or from measurements of another set from the same population. For environmental studies the mean is often

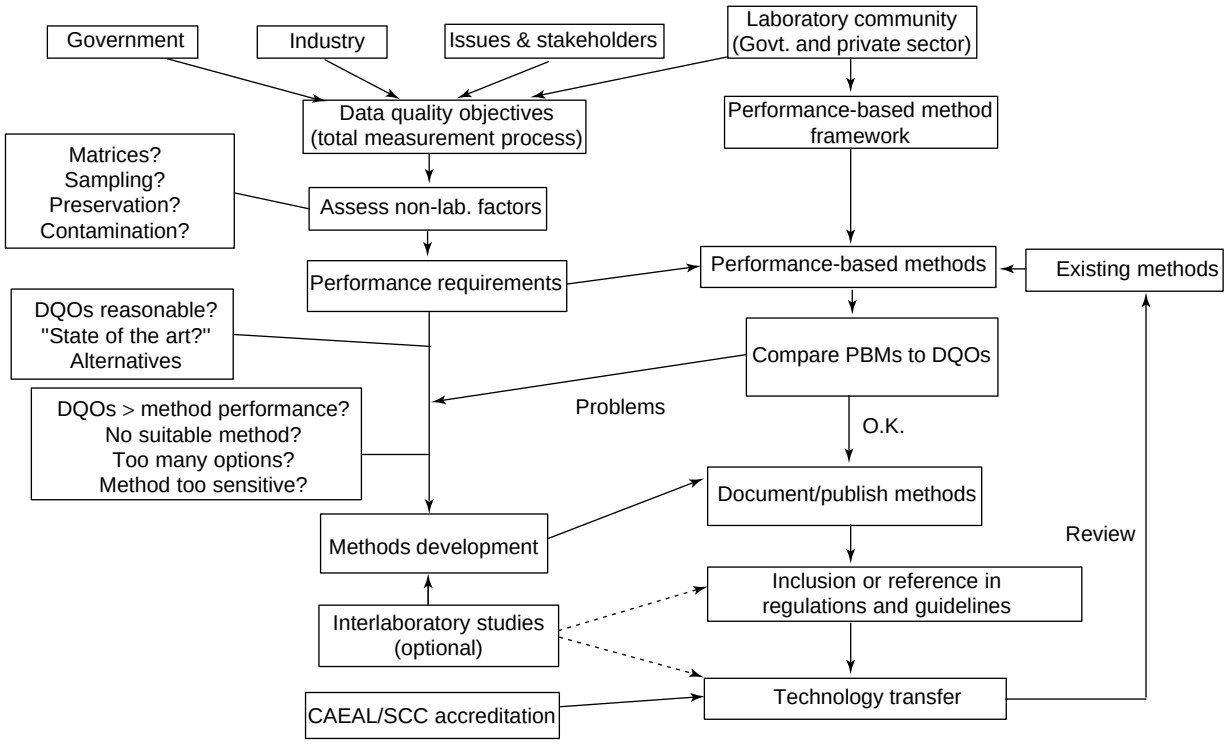

Figure 8 Schematic of the PBM process as currently being implemented in British Columbia, Canada. (Reproduced with permission of Environment Canada, Pacific Environmental Science Centre (PESC), North Vancouver, Canada, from their standalone drawing.)

estimated from only a few measurements, sometimes even from a single measurement. CIs based on a small number of measurements would be fairly wide, and infinitely wide for a single measurement. However, Taylor[3] suggested that for analytical systems proven by control charts to be in statistical control, the values for s and for n may be determined from the appropriate long-term QC results. The advantage of employing the long-term QC results to estimate the standard deviation is that often very large numbers of such data are available from the laboratory QC program, thus yielding small s and large n, and hence narrow CIs. (It is important to note that both Z and t are two-tailed for CI calculations but are single-tailed for LOD calculations.)

6.8 Control Charts

Control charts provide an understandable means to demonstrate that laboratory analyses are under statistical control, or conversely to diagnose analytical problems for systems which are not in control. Taylor[3] has written an excellent chapter on this topic. Of the several different types of control charts employed by laboratories, the most frequently employed is the Shewhart chart. For these charts, aliquots of some reference material are repeatedly analyzed over time, and the chronological sequence of results are plotted. An example plot is shown in Figure 9.

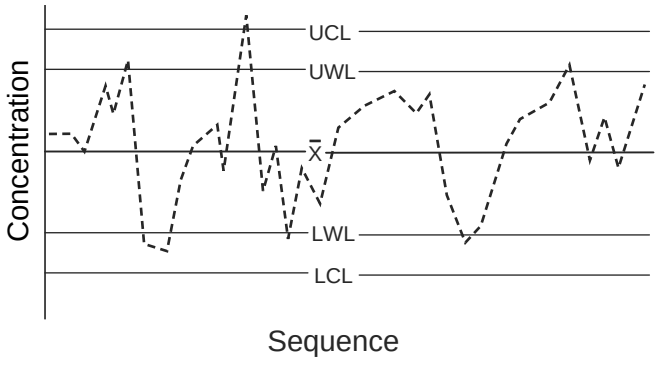

Figure 9 Example of the Shewhart control chart. LCL = lower control limit; LWL = lower warning limit; X = mean value; UWL = upper warning limit; UCL = upper control limit. (Reprinted with permission from J.K. Taylor, *Quality Assurance of Chemical Measurements*, CRC Press, Boca Raton, FL, 1987. Copyright CRC Press, Boca Raton, FL.)

The central line on the chart corresponds either to the experimentally determined mean for the reference material, or to the certified ''true'' amount for a CRM. There are several variations to the Shewhart chart. The most common variation has both control and warning limits established on either side of the center line, these limits established according to Equations (8) and (9) or Equations (10) and (11). One or more measurements beyond

the control limits signals out-of-control conditions and indicates that the analytical method requires remedial attention. Values beyond the warning limit do not require remedial attention unless there is a string of two or more such values adjacent in sequence. There are additional rules to indicate out-of-control conditions, as summarized by Taylor.[3]

Some laboratories choose to plot individual determinations upon their control charts, and other laboratories prefer to plot the mean of several determinations. It is important to distinguish between these two types of Shewhart charts since they employ different algorithms for the control and warning limits, as noted below:

(a) Plots of individual determinations:

$$\text{control limits} = \pm 3s \tag{8}$$

$$\text{warning limits} = +2s \tag{9}$$

(b) Plots of means of n-multiple determinations:

$$\text{control limits} = \pm 3s/\sqrt{n} \tag{10}$$

$$\text{warning limits} = \pm 2s/\sqrt{n} \tag{11}$$

Another style of control chart is the Cusum chart. For repeated analyses of a reference material, the cumulative sum (hence the name Cusum) of the deviation for each determination from the target mean plus deviations for previous determinations is calculated and plotted sequentially. Each point X_n on the chart is determined according to Equation (12):

$$\text{Cusum}(X_n) = \sum_{i=1}^{n} (X_i - \text{mean}_{\text{target}}) \tag{12}$$

These points are plotted either chronologically or in sequence of analysis. Cusum charts tend to be more sensitive than Shewhart charts to some types of problems. However, they are more difficult to comprehend and to program, since the action limits both are not parallel to the axes and also change with addition of each new data point. For this reason the Shewhart charts are more popular. (The author prefers to employ Shewhart and Cusum charts in tandem, but this is an unusual approach.) Garland et al.[63] compared Shewhart, Cusum and other approaches with regard to predictive power of analytical fault detection. They found the Cusum technique to be the most effective. Garfield[18] similarly reports Cumsum charts to provide a faster warning of process deviation from control.

A third style of control chart is the range control chart. The distinctive feature of these charts is that they utilize results from actual test samples, whereas Shewhart and Cusum charts utilize results from analysis of reference materials. Range control charts are particularly useful in circumstances where reference materials can be challenged as inappropriate. These charts are employed to monitor measurement precision. However, the deficiency to range control charts is that they do not respond to systematic shifts in measurement response. Range control charts require that a constant number of replicate analyses (usually duplicate analyses) be performed on a routine selection of test samples. The average values for the range, the warning limits, and the control limits are established from statistical analysis of the initial eight or more [Taylor[3] prefers 15 or more] replicate sets. Taylor[3] has summarized this method in some detail, and also has summarized several additional types of control charts not included in this article.

6.9 Laboratory Automation

Most laboratories now employ "laboratory information management systems" (LIMS) to automate data management. Jobs performed by LIMS include the preparation of work lists, sample tracking, instrument calibration, verification of QC samples, and delivery of results and invoices to clients. Increasingly sophisticated statistical techniques and graphic presentations are becoming available to both bench analysts and to laboratory QA Officers. Much of the literature regarding LIMS, e.g. Peak et al.,[64] focuses upon their successful employment to review QC results. This past decade has seen considerable improvements in both computer hardware and software utilized for LIMS, and an increasingly larger proportion of routine analysis is being automated. Reports supplied to clients can now include QA verification information and custom formats as specified by the clients. Additional benefits of laboratory automation include reduction in transcription errors, automatic verification against site-specific historical results, production of archive information, and more efficient billing. Benefits to clients include on-line access to preliminary results and tracking of cost estimates for work in progress. Benefits to laboratories include facilitation of audit trails, facilitation of system validation, timely access to original test results and to associated QA/QC data, plus convenient access to methodology information.

Drawbacks to laboratory automation include difficulties in intermarrying hardware and software from different vendors and of different vintage. This is a continuing problem because of the present rapid pace of technological change with regard to computers, instrumentation, and associated software. A second drawback is the increasing workload for computer experts to keep these systems operational as they become increasingly more sophisticated. The following factors are important considerations when establishing a new LIMS system:

- convenient installation and set-up
- convenient upgrade and expansion

- flexible configuration
- documentation complete and regularly updated
- long-term commitment by computer, instrument and software vendors
- timely support by vendor technical staff
- system easy to learn and use
- system security
- flexible retrievals to reflect specific requirements by different clients
- internet and intranet access.

QA for LIMS includes comprehensive records as to hardware and software, timely identification and resolution of problems, routine challenge of LIMS elements, and independent audits by external LIMS experts. A laboratory's QA manual should include well documented policies and procedures regarding data management via LIMS: data records to be maintained, handling of absent and nonconforming results (including client notification), data reporting formats, special rules for low-level results and for calculated results, data validation, test report content, procedures test report revision, test report authorization, transmission of results to clients, transmission of revisions to clients, access to records, confidentiality of records, and disposal of records. The main focus of QA with regard to LIMS must always be protection of the results. Is there mandatory tracking of all revisions to results? Is there adequate protection against power failure and against introduction of computer viruses? Is there off-site backup in case of fire or earthquake?

ACKNOWLEDGMENTS

Thanks are expressed to H. Agemian, H. Alkema, G. Kan, H. Quon, E. Tradewell, P. Whitfield for suggesting topics to be included in this article.

DISCLAIMER

The opinions expressed in this article are those of the author and are not the policy of the author's employer, the BC Ministry of Environment, Lands and Parks.

LIST OF SYMBOLS

σ	Population standard deviation
α	Probability of type I error, the rejection of a true null hypothesis (e.g. concluding an analyte is present when in fact it is absent)
β	Probability of type II error, the failing to reject a false null hypothesis (e.g. concluding an analyte is absent when it is in fact present)
C	Analyte concentration expressed as a mass/mass decimal fraction
n	Number of measurements used to calculate the mean
s	Sample standard deviation
t	Student's t, the value of which is determined by both the level of confidence specified and the degrees of freedom upon which the standard deviation is based
Z	Probability factor, the value of which depends on the level of confidence specified

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
CI	Confidence Interval
CRM	Certified Reference Material
CV	Coefficient of Variation
DQO	Data Quality Objective
DQS	Documented Quality System
GLP	Good Laboratory Practice
GSP	Good Scientific Practices
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
LIMS	Laboratory Information Management System
LOD	Limit of Detection
LOQ	Limit of Quantitation
MDL	Method Detection Limit
NIST	National Institute of Standards and Technology
PBM	Performance-based Method
PDL	Practical Detection Limit
QA	Quality Assurance
QA/QC	Quality Assessment Plus Quality Control
QC	Quality Control
SOP	Standard Operating Protocol
SRM	Standard Reference Material
ULOQ	Upper Limit of Quantitation

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants

Chemometrics **(Volume 11)**
Chemometrics

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. L.H. Keith, *Environmental Sampling and Analysis: A Practical Guide*, CRC Press, Boca Raton, FL, 1991.
2. M.J.R. Clark, P.H. Whitfield, 'A Practical Model Integrating Quality Assurance into Environmental Monitoring', *Water Resour. Bull.*, **29**(1), 119–130 (1993).
3. J.K. Taylor, *Quality Assurance of Chemical Measurements*, Lewis Publishers, Chelsea, MI, 1987.
4. R.J. Mesley, W.D. Pocklington, R.F. Walker, 'Analytical Quality Assurance. A Review', *Analyst*, **116**, 975–990 (1991).
5. K.A. Rose, E.P. Smith, 'Experimental Design: the Neglected Aspect of Environmental Monitoring', *Environ. Manage.*, **16**(6), 691–700 (1992).
6. B. King, 'Data Quality in the 1990s: Targets and Approaches', *Anal. Proc.*, **29**, 184–186 (1992).
7. M.J.R. Clark, 'Characteristics of a Successful Water Quality Monitoring Network: the Canada–British Columbia Water Quality Monitoring Agreement Experience', Paper presented to Assessment of Water Quality Monitoring Networks – Design and Redesign Project, Vancouver, Canada, February 5–8, 1996, NATO International Scientific Exchange Programme Linkage Grant No. 950779, 1996.
8. P.S. Porter, R.C. Ward, H.F. Bell, 'The Detection Limit', *Environ. Sci. Technol.*, **22**(8), 856–861 (1988).
9. S.S. Prasad, 'Trends in Quality Assurance in the Characterization of Chemical Contaminants in the Environment', *Trends Anal. Chem.*, **13**(4), 157–164 (1994).
10. H. Hurlbert, 'Pseudoreplication and the Design of Ecological Field Experiments', *Ecol. Monogra.*, **54**(2), 187–211 (1984).
11. R.C. Ward, J.C. Loftis, 'Establishing Statistical Design Criteria for Water Quality Monitoring Systems: Review and Synthesis', *Water Resour. Bull.*, **22**(5), 759–767 (1986).
12. R.W.P.M. Laane, Ir.B.J.E. ten Brink, 'Data-rich, Information-poor: The Modern Monitoring Syndrome?', *Land Water Int.*, **68**, 12–16 (1990).
13. W.A. Maher, P.W. Cullen, R.H. Norris, 'Framework for Designing Sampling Programs', *Environ. Monit. Assess.*, **30**, 139–162 (1994).
14. K.J.M. Kramer, 'Inorganic Contaminants in the Water Column: Sampling and Sampling Strategy', *Int. J. Environ. Anal. Chem.*, **57**, 179–188 (1994).
15. G.J. Stroomberg, I.L. Freriks, F. Smedes, W.P. Cofino, 'Quality Assurance and Quality Control of Surface Water Sampling', in *Quality Assurance in Environmental Monitoring. Sampling and Sample Pretreatment*, ed. P. Quevauviller, VCH, Weinheim, 51–90, 1995.
16. G. Wagner, 'Basic Approaches and Methods for Quality Assurance and Quality Control in Sample Collection and Storage for Environmental Monitoring', *Sci. Total Environ.*, **176**, 63–71 (1995).
17. H. Agemian, *Quality Assurance in the National Water Quality Laboratory*, 3rd edition, Canada Centre for Inland Waters, Burlington, ON, 1989.
18. F. Garfield, *Quality Assurance Principles for Analytical Laboratories*, Association of Official Analytical Chemists, Arlington, VA, 1991.
19. M.J.R. Clark, P.H. Whitfield, 'Conflicting Perspectives About Detection Limits and About the Censoring of Environmental Data', *Water Resour. Bull.*, **30**(6), 1063–1079 (1994).
20. PTI Environmental Services, 'A Project Manager's Guide to Requesting and Evaluating Chemical Analyses', EPA Contract No. 68D80085, Prepared for the US Environmental Protection Agency, Region 10, Office of Coastal Waters, Seattle, WA, 1991.
21. C.N. Hewitt (ed.), *Methods of Environmental Data Analysis*, Elsevier Applied Science, London, 1992.
22. M.L. Albu, 'ISO 9000 and the Analytical Laboratory', *Am. Lab.*, September, 52–55 (1996).
23. J.A. Cichowicz, 'ISO 14000: New Opportunities for Expanding P2', *Pollut. Prev. Rev.*, Spring, 1–11 (1996).
24. W. Horwitz, 'History of the IUPAC/ISO/AOAC Harmonization Program', *J. AOAC Int.*, **75**(2), 368–371 (1992).
25. L.H. Keith, 'Cost-effective Systematic Sampling', *Environ. Test. Anal.*, September, 18–23 (1996).
26. M.H. Ramsey, M. Thompson, M. Hale, 'Objective Evaluation of Precision Requirements for Geochemical Analysis Using Robust Analysis of Variance', *J. Geochem. Explor.*, **44**, 23–36 (1992).
27. M.J.R. Clark, M.C.A. Laidlaw, S.C. Ryneveld, M.I. Ward, 'Estimating Sampling Variance and Local Environmental Heterogeneity for Both Known and Estimated Analytical Variance', *Chemosphere*, **32**(6), 1133–1151 (1996).
28. P.L. Brezonik, *Chemical Kinetics and Process Dynamics in Aqueous Systems*, CRC Press, Boca Raton, FL, 1994.
29. M.J. Barcelona, 'Overview of the Sampling Process', in *Principles of Environmental Monitoring*, American Chemical Society, Washington, DC, 3–23, 1988.
30. R.E. Stewart, J. Anderson, 'Metals Monitoring', *Water Environ. Technol.*, August, 65–70 (1998).
31. N. Cavanagh, R.N. Nordin, L.W. Pommen, L.G. Swain, *Guidelines for Interpreting Water Quality Data*, Ministry of Environment, Lands and Parks, Victoria, 1998.

32. J.M. Richardson, J.L. Morrison, 'Method Validation in the Ontario Geoscience Laboratories', *Analyst*, **120**, 1519–1526 (1995).
33. J.F. Solsky, 'Questionable Practices in the Laboratory', presented at the 14th Annual Waste Testing and Quality Assurance Symposium, Washington, DC, 1998.
34. J.C. Worthington, S.A. Coll, TechLaw, Inc., 'Auditing Environmental Data Systems', in *ASQC Quality Congress Transactions, 46th Annual Quality Congress, May 18–20, 1992, Nashville, TN*, American Society for Quality Control, Milwaukee, WI, 692–705, 1992.
35. P.N. Pink, J.S. Jessup, D.K. Knight, 'Reporting Environmental Data of Known Quality', in *Waste Testing and Quality Assurance: Third Volume*, ed. D. Friedman, ASTM STP 1075, American Society for Testing and Materials, Philadelphia, PA, 339–352, 1991.
36. 'BC Environmental Monitoring System: Using EMS', Release 1.3, British Columbia Ministry of Environment, Lands and Parks, Victoria, 1997.
37. T.M. Fyles, B.A. King, M. Pawluk, P.R. West, 'Implications of Data Truncation in Environmental Decision-making', Report prepared by the University of Victoria for the BC Ministry of Environment, Lands, and Parks, Victoria, 1997.
38. W. Horwitz, 'Uncertainty – A Chemist's View', *J. AOAC Int.*, **81**(4), 785–794 (1998).
39. K. Byrialsen, J. Kristiansen, J.M. Christensen, 'Trends in Quality Assurance of Metal Determination in Clinical Chemistry', *Analyst*, **123**, 7–12 (1998).
40. T.M. Fyles, B.A. King, M.P. Pawluk, T. Sowden, P.R. West, 'The Development of Protocols for the Appraisal of Survey and Experimental Events in a Contamination Information System', in *EnviroAnalysis: Practical Methods for Environmental Analysis. Proceedings of the EnviroAnalysis Conference, May 13–16, 1996, Ottawa, Ontario, Canada*, eds. R.E. Clement, R.E. Burk, Polyscience Publications, Morin Heights, PQ, 1996.
41. S.P. Millard, 'Environmental Monitoring, Statistics, and the Law: Room for Improvement', *Am. Stat.*, **41**(4), 249–253 (1987).
42. R.M. Peterman, 'Statistical Power Analysis Can Improve Fisheries Research and Management', *Can. J. Fish. Aquat. Sci.*, **47**, 2–15 (1990).
43. P. Cullen, 'The Turbulent Boundary Between Water Science and Water Management', *Freshwater Biol.*, **24**, 201–209 (1990).
44. A.Y. Cantillo, G.G. Lauenstein, 'Use of Reference Materials in Coastal Monitoring Quality Assurance', *Fresenius' J. Anal. Chem.*, **252**, 152–156 (1995).
45. W. Horwitz, L.R. Kamps, K.W. Boyer, 'Quality Assurance in the Analysis of Foods for Trace Constituents', *J. Assoc. Off. Anal. Chem.*, **63**(6), 1344–1354 (1980).
46. C.L. Grant, A.D. Hewitt, T.F. Jenkins, 'Experimental Comparison of EPA and USATHAMA Detection and Quantitation Capacity Estimators', *Am. Lab.*, February, 15–33 (1991).
47. L.S. Oresic, V. Grdinic, 'Kaiser's 3-Sigma Criterion: a Review of the Limit of Detection', *Acta Pharm. Jugosl.*, **40**, 21–61 (1990).
48. G.L. Long, J.D. Winefordner, 'Limit of Detection: a Closer Look at the IUPAC Definition', *Anal. Chem.*, **55**(7), 712A–724A (1983).
49. D.A. Kurtz, J.K. Taylor, L. Sturdivan, W.B. Crummett, C.R. Midkiff, Jr, R.L. Watters, Jr, L.J. Wood, W.W. Hanneman, W. Horwitz, 'Real-world Limitations to Detection', *ACS Symp. Ser.*, **361**, 288–316 (1988).
50. J.D. Winefordner, M. Rutledge, 'Comparison of Calculated Detection Limits in Molecular Absorption, Molecular Luminescence, Raman, Molecular Ionization, and Photothermal Spectrometry', *Appl. Spectrosc.*, **39**(3), 377–391 (1985).
51. M.J. Adams, 'Errors and Detection Limits', in *Methods of Environmental Data Analysis*, ed. C.N. Hewitt, Elsevier Science, New York, Chapter 5, 181–212, 1992.
52. Environment Canada, *The Inspector's Field Sampling Manual*, Environmental Protection Service, Ottawa, Canada, 1995.
53. Environment Canada, 'Legal Sampling Training and Reference', CD-ROM, Pacific Environment Science Centre, North Vancouver, 1998.
54. R.S. Frank, 'Records – Why Keep Them?', in *Optimizing Chemical Laboratory Performance Through the Application of Quality Assurance Principles*, eds. F.M. Garfield, N. Palmer, G. Schwartzman, Association of Official Analytical Chemists, Arlington, VA, 129–154, 1980.
55. J.S. Kane, 'Reference Samples for Use in Analytical Geochemistry: Their Availability, Preparation, and Appropriate Use', *J. Geochem. Explor.*, **44**, 37–63 (1992).
56. R.L. Watters, Jr, W.F. Koch, 'The Use of Standard Reference Materials in Spectrometric and Electroanalytical Quality Assurance Programs', *Microchem. J.*, **45**, 282–290 (1992).
57. B. Griepink, P. Quevauviller, E.A. Maier, S. Vandendriessche, 'BCR: a Service to Quality Assurance in Analytical Chemistry – Some Experiences and Achievements with Regard to Reference Material Preparation', *Fresenius' J. Anal. Chem.*, **346**, 530–535 (1993).
58. International Union of Pure and Applied Chemistry, 'Protocol for the Design, Conduct and Interpretation of Collaborative Studies', *Pure Appl. Chem.*, **60**, 855–864 (1988).
59. F.C. Wells, W.J. Gibbons, M.E. Dorsey, 'Guidelines for Collection and Field Analysis of Water-quality Samples from Streams in Texas', Open File Report 90–127, US Geological Survey, Austin, TX, 1990.
60. C. Camara, C. Perez-Conde, M.C. Moreno-Bondi, C. Rivas, 'Fiber Optical Sensors Applied to Field Measurements', in *Quality Assurance for Environmental Analysis*, eds. Ph. Quevauviller, E.A. Maier, B. Griepink, Elsevier, Amsterdam, 165–194, 1995.
61. P.H. Whitfield, N.W. Wade, 'Quality Assurance Techniques for Electronic Data Acquisition', *Water Resour. Bull.*, **29**(2), 301–308 (1993).

62. R. Stevenson, 'Twin Stars Shine: Environmental Immunoassays and Performance-based Methods', *Sep. Sci.*, October, 4–7 (1995).
63. S.W. Garland, B. Lees, J.C. Stevenson, 'DXA Longitudinal Quality Control: a Comparison of Inbuilt Quality Assurance, Visual Inspection, Multi-rule Shewhart Charts and Cusum Analysis', *Osteoporosis Int.*, **7**, 231–237 (1997).
64. S.R. Peak, P.E. Duerksen, K.W. Wong, E. Szeto, 'Early Warning Report: Automated Checking of QC Data', in *Waste Testing and Quality Assurance: Third Volume*, ed. D. Friedman, ASTM STP 1075, American Society for Testing and Materials, Philadelphia, PA, 362–376, 1991.

Sample Preparation for Elemental Analysis of Biological Samples in the Environment

Kunnath S. Subramanian
Health Canada, Ottawa, Canada

1 Introduction **1**
2 Flame Atomic Absorption Spectrometry **2**
2.1 Introduction 2
2.2 Dilution 3
2.3 Protein Precipitation (Deproteinization) 3
2.4 Acid Extraction 3
2.5 Chelation/Solvent Extraction 3
2.6 Destruction of Organic Matter 4
2.7 Hydride-forming Elements 5
2.8 Cold Vapor Atomic Absorption Spectrometry 5
2.9 Tissue Solubilization 5
2.10 Solid Sampling 5
2.11 Methods for Various Elements 5
2.12 Methods for Urine, Hair, Nail, Feces, Milk, Tooth and Bone 7
3 Graphite Furnace Atomic Absorption Spectrometry **9**
3.1 Introduction 9
3.2 Direct Methods 9
3.3 Dilution 9
3.4 Deproteinization 10
3.5 Matrix (Chemical) Modification 10
3.6 Dissolution of Tissues 10
3.7 Complete Oxidation of Tissues and Fluids 10
3.8 Solid Sampling 11
4 Inductively Coupled Plasma Atomic Emission Spectrometry **11**
4.1 Direct Methods 11
4.2 Deproteinization 12
4.3 Oxidative Decomposition 12
4.4 Solid Sampling 13
5 Inductively Coupled Plasma Mass Spectrometry **13**
5.1 Introduction 13
5.2 Dilution 14
5.3 Oxidative Decomposition 14
5.4 Solid Sampling 15
5.5 Speciation 16
5.6 Other Studies 16
6 Anodic Stripping Voltammetry **16**
6.1 Introduction 16
6.2 Direct Methods 17
6.3 Acid Decomposition 17
6.4 Dry Ashing 18
6.5 Other Voltammetric Techniques 18
7 X-ray Spectrometry **18**
7.1 Introduction 18
7.2 Biological Fluids 19
7.3 Tissues 19
8 Neutron Activation Analysis **20**
Abbreviations and Acronyms **21**
Related Articles **21**
References **22**

This article focuses on biological sample preparation methods which are unique to each of the commonly used instrumental techniques used in trace element analysis. The biological samples covered are mainly of human and animal origin. The preparation methods considered span the entire gamut and include direct solid or liquid sample introduction involving dilution or matrix modification; dry ashing; wet oxidation including microwave digestion and high-pressure ashing; deproteinization; and tissue solubilization. The instrumental techniques covered are flame atomic absorption spectrometry (FAAS), graphite furnace atomic absorption spectrometry (GFAAS), inductively coupled plasma atomic emission spectrometry (ICPAES), inductively coupled plasma mass spectrometry (ICPMS), X-ray fluorescence (XRF) spectrometry, neutron activation analysis (NAA) and anodic stripping voltammetry (ASV). The choice of a given sample preparation method would be governed in general by the type of biological matrix, sample size and the type of instrumental technique used. The advantages and disadvantages of the various sample preparation methods have been emphasized for each of the instrumental techniques. Also, an attempt has been made to point out the optimum sample preparation method(s) suitable for the particular biological matrix and instrumental technique.

1 INTRODUCTION

The sample preparation method to be used for a given analysis will be governed by the nature and concentration of the analyte, the nature (solid or liquid) and type of biological matrix, the sample size (amount or volume available) and the type of instrumental technique

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

used. Generally, solids will require some form of digestion/dissolution, although direct analysis may be feasible with many of the spectroscopic techniques. Liquids may be amenable to direct analysis, but matrix interferences in many cases may necessitate some form of sample pretreatment. This article will focus on sample preparation methods which are unique to each instrumental technique. The instrumental techniques covered are FAAS, GFAAS, ICPAES, ICPMS, XRF spectrometry, NAA and ASV, since they are most commonly used in biological trace element analysis. These preparation methods, which are strictly confined to samples of human and animal origin, have been discussed in detail elsewhere recently with literature citations covering the period to 1993.[1] Therefore, the present treatment will be brief, and literature citations will be primarily restricted to the period since 1993. For sample preparation methods dealing with botanicals and botanically derived products, the reader may refer to the book by Rossbach et al.[2]

2 FLAME ATOMIC ABSORPTION SPECTROMETRY

2.1 Introduction

The sample preparation step is a major concern in atomic absorption spectrometry (AAS) as it basically governs the analytical performance characteristics, namely detection limit, accuracy, precision, speed, throughput and cost. Because of the selective nature of AAS, classical preparatory work, i.e. complete decomposition of organic matter followed by separation and preconcentration of the analyte(s) from the matrix and other concomitants, is not mandatory.

The choice of FAAS or GFAAS depends on the concentration of the element sought and the analytical sensitivity. Table 1[1] presents a guide to the relative sensitivities of these two techniques. Generally, elements coded A can be measured using FAAS with relative ease, those coded B and C are progressively more difficult and those coded D are impossible with FAAS. The relative lack of sensitivity of FAAS restricts its use to those coded A, namely Na, K, Ca, Mg, Li, Au, Cu and Zn. However, the practical advantages of FAAS such as speed, simplicity, precision, cost and ease of operation are ideal and necessary for the many situations where large throughput is required for the diagnosis and management of a patient's clinical condition. Thus, many practical measurements in clinical laboratories, hospitals, occupational health, toxicology and nutrition involve FAAS.

For details on FAAS sample preparation methods, the reader may consult a recent book,[1] *Atomic Spectroscopy* (a Perkin-Elmer publication), which publishes twice yearly a bibliography with the papers being listed in alphabetical order of authors, and the *Journal of Analytical Atomic Spectrometry*, which publishes AAS updates.

The major FAAS sample preparation methods are dilution, deproteinization, acid extraction, chelation/solvent

Table 1 A guide to the choice of FAAS and GFAAS for the determination of elements in biological materials, especially blood and serum[a,b]

Essential			Therapeutic			Nonessential		
Element	FAAS	GFAAS	Element	FAAS	GFAAS	Element	FAAS	GFAAS
Ca	A	A	Au	A	A	Pb	B	A
Mg	A	A	Li	A	A	Cd	C	A
Cu	A	A	Al	C	B	Ag	C	A
Zn	A	A	Bi	C	B	Tl	C	B
Fe	B	A	Ga	D	B	As	D	B
Se	C	A	Pt	D	C	Be	D	B
Mn	D	B				Hg	D	C
Cr	D	C				Ni	D	C
Co	D	D				Sb	D	C
Mo	D	D				Te	D	D
V	D	D						

[a] Adapted from Iyengar et al.[1]

[b] A, easily determinable by both FAAS and GFAAS (requires <1 mL of sample by FAAS and <10 μL by GFAAS); B and C, progressively more difficult (B requires 1–10 mL by FAAS and 10–50 μL by GFAAS while C requires 10–100 mL by FAAS and 50–1000 μL by GFAAS); D, extremely difficult by FAAS and difficult by GFAAS (>100 mL by FAAS and >1000 μL by GFAAS). The volumes required are for the generation of a signal of 0.05 absorbance unit from either a dilution/concentration to 1 mL for FAAS or a direct injection with a practical upper limit of 50 μL for GFAAS. The volumes are based on the mid-range of reference levels for the essential, the lower end of the target range for the therapeutic and the upper limit of acceptable levels (in the case of 'nonexposed' individuals) for the nonessential elements.

extraction, destruction of organic matter through dry ashing or wet digestion, tissue solubilization, atom trapping/preconcentration to improve sensitivity and mercury and hydride generation techniques. Each of these methods is briefly described below.

2.2 Dilution

Dilution methods are mainly used for biological fluids, especially serum and urine. The main function of diluents is to overcome the viscosity effect of the matrix (e.g. protein content of blood) on the sample uptake rate. Simple aqueous or acidic (0.1 M HCl) dilutions that can be used to overcome this effect are 50-, 12- and 5-fold for whole blood, serum and urine, respectively. The acid concentration used in dilutions should be as low as possible. Solutions with acid concentrations >5% should be nebulized using a nebulizer with a Pt–Ir capillary. With relatively low temperature flames, the addition of La(III), Sr(II) or ethylenediaminetetraacetic acid (EDTA) is necessary to overcome the suppressive effect of phosphate on Ca absorption and the influence of ionization on Na and K. For example, Ca, Mg, Na and K can be determined by a 50-fold dilution of serum with 0.25% $SrCl_2$.

In direct and dilution procedures, the viscosity and solids content of biological fluids must be such as to permit nebulization without giving rise to problems associated with variation in aspiration rate, burner blockage or nebulizer 'salting-out'. The long-path air–acetylene burner head will accommodate 2–3% solids at 40 psi oxidant pressure, and up to 10% solids by progressively reducing the oxidant pressure. In many cases, however, the design of the nebulizer, premix chamber and burner may not permit complete removal of the matrix effect caused by the protein in serum, especially at dilution factors <10, and compensating factors may have to be introduced. These include the addition of butanol, acetone or Triton X-100 to the diluent, deproteinization and comparison with aqueous standards, positive pressure sampling and matching the viscosity of standards to that of the plasma by the addition of glycerol.

Considerable variation in the protein content of plasma occurs among normal subjects and very wide differences occur between normal and pathological specimens. In most chronic diseases, the albumin content falls substantially, and to obtain accurate results, separate standards of appropriate concentrations have to be prepared. Hence the technique of loading standards with a protein-like material to match the sample viscosity cannot be considered suitable for clinical work, and simple dilution of the sample is recommended. Since significant alterations in aspiration rate can occur even when plasma is diluted 10-fold, a minimum dilution of 12-fold is recommended. Reduced signal strength may require scale expansion leading to an unstable baseline, but this problem can be overcome by employing a strictly timed sequence of measurements. Matrix effects may be pronounced, requiring matrix-matched calibration. Fortunately, the matrix properties of blood and serum are constant, and calibration solutions can be prepared in pooled spiked blood or a few previously analyzed specimens can be used as standards. Flow injection analysis (FIA) in conjunction with FAAS can used to facilitate automatic sample handling such as on-line dilution and buffer additions, reduction in sample consumption and a minimization of interferences arising from sample viscosity.[1]

2.3 Protein Precipitation (Deproteinization)

Protein precipitation involves removal of protein from serum and blood by precipitation with a reagent such as trichloroacetic acid (TCA) and FAAS analysis of the supernatant fluid. TCA, employed as a 5–10% solution, is the most popular reagent. Elements determined by this method include Cu, Fe and Zn. However, the TCA precipitation method for Cu and Zn gives higher values than the dilution method owing either to volume exclusion errors or to some nonspecific interferences. There seems little value in using the TCA method in view of the established accuracy of the dilution procedures.

2.4 Acid Extraction

Zinc could be quantitatively extracted from homogenized diet and feces samples with 1 M HCl. This simple yet efficient method can be extended to other metals in diets and feces. Quantitative release of Mn from liver homogenates with 10% Triton X-100 could be obtained heating for 1 h in 2 M HCl at 60 °C. Extraction of bovine liver with 1% nitric acid gave quantitative recoveries of Cd, Cu, Mn and Zn.

2.5 Chelation/Solvent Extraction

When the analyte concentration falls below the detection limit or is low enough to make a precise direct determination impossible, preconcentration by solvent extraction is used in conjunction with FAAS to concentrate the analyte (by reducing the solvent/aqueous phase volume ratio) or to remove matrix interferences, or both. Also, organic solvents generally produce an enhanced signal in the flame. The most popular chelating agents are sodium diethyldithiocarbamate (NaDDC) and ammonium pyrrolidinedithiocarbamate (APDC) because they complex with many metals. Also, a concentration factor as high as 100 is possible with APDC, and the analyte can be easily separated from the biological matrix which

may cause difficulties with nebulization and atomization. The most widely used solvent is methyl isobutyl ketone (MIBK) because of its extraction and nebulization efficiencies, combustibility and high enhancement factor. If required, selective extraction can be performed by adjustment of the sample pH and use of more specific chelating agent/solvent systems. In this way, a particular metal which may cause subsequent interference in the flame analysis can be removed or the analyte under consideration can be isolated from the matrix. Delves et al.[3] used this approach to extract sequentially 11 metals from 1 mL of blood.

The solvent extraction approach should be used only when the sensitivity of direct procedures is inadequate, severe matrix effects are to be avoided or certain specific chemical species or binding state of the analyte are being studied. This is because it is tedious and prone to risk of contamination. Nowadays, solvent extraction methods in FAAS methods are used mainly for the concentration of low levels of metals from urine and tissue specimens after destruction of organic matter. Extractions are most efficiently performed on ashed samples in which the analyte is converted into an ionic form, but some elements, such as Cd, Ni, Pb and Tl, have been extracted from blood hemolyzed from Triton X-100 or from the supernatants of TCA-deproteinized blood or serum. Ashing prior to extraction is certainly required for As, Au, Hg and Se since they are strongly bound to the organic matrix.

Recent preconcentration studies using FIA for on-line solid sorbent extraction and subsequent FAAS detection seem very promising.[4] The combined advantages of automation, speed, microsample requirement, preconcentration, separation of analyte from matrix concomitants and sensitivity enhancement by the organic solvents permit several elements to be determined by FAAS which were previously accessible only by GFAAS, and which thus seem to be ideally suited for clinical application. The on-line extraction of Cu and Cd in urine as the diethyldithiocarbamate (DDTC) complexes on reversed-phase C_{18} immobilized on silica gel and elution into the FAAS system using ethanol, the on-line ion-exchange preconcentration of Al in dialysis fluids and the determination of Pb in acid-digested blood and liver samples by on-line coprecipitation with the hexamethyleneammonium hexamethylenedithiocarbamate (HMAHMDTC)–Fe(II) complex are typical examples.[4] The coprecipitation procedure has also been used for the determination of Cd, Co and Ni in blood, urine and animal and plant tissues. The procedure is unique in that the precipitate was collected on the walls of a knotted reactor made of 150 cm of Microline tubing of 0.5 mm i.d. The precipitate was dissolved in a stream of MIBK which was fed directly into the flame. The preconcentration factor, signal enhancement factor, coprecipitation time and sampling frequency were 20, 66, 30 s and $90\,h^{-1}$, respectively. On-line preconcentration has become even more attractive since it could be combined with on-line sample digestion, as has been demonstrated in the case of Bi, Hg, Pb and Sn in urine.[4]

2.6 Destruction of Organic Matter

The introduction of tissue specimens into an FAAS system necessitates conversion of organic matter to a relatively simple inorganic form in solution. Wet digestion and dry ashing are the two main techniques employed in tissue destruction. Wet digestion methods include decomposition by acids either alone or in mixtures carried out in *open* vessels in tubes on a hot-plate or aluminum heating block, and in *closed* vessels at increased pressure via a bomb or high-pressure asher (HPA) or microwave heating. Microwave digestion is an attractive method, especially for small samples. The simplest dry ashing technique for the preparation of tissue specimens is to ash the sample in a conventional muffle furnace at $<600\,^{\circ}C$, usually with some type of ashing aid. This approach often leads to contamination and loss of analytes, and thus has tended to fall from favor. Acceleration of the reaction can be accomplished for small samples by heating at $<150\,^{\circ}C$ in an atmosphere of radio frequency-generated oxygen. This low-temperature ashing (LTA) minimizes the loss of volatile elements.

In general, complete oxidation of organic matter is not necessary (but see below) for FAAS work as long as the analyte is quantitatively brought into an appropriate acidic solution, and the blank level can be reduced. FAAS tolerates any acid up to a concentration of <5%. The majority of the procedures involve HNO_3, but HCl and $HCl + HNO_3$ are also effective. Complete digestion is essential for those elements which are most prone to volatile loss, namely As, Ge, Hg, Sb, Se, Sn and Te. In these cases, wet digestion should be performed under carefully controlled conditions. The completeness and speed of these digestion procedures can be enhanced by using pressurized vessels; also, they minimize blanks and loss of volatile elements.

The concentration of many elements (e.g. Co, Cr) in tissue samples is greater than that in body fluids, and it is possible to determine them by FAAS. Many digestion procedures are available, but the $HNO_3 + HClO_4$ procedure described by Taylor is satisfactory for elements such as Na, K, Ca and Mg.[1]

Blanks can be minimized by using small sample sizes, pure reagents and labware materials, thorough cleaning, shorter digestion times, using reflux conditions or closed vessels, low temperature, controlled heating and clean air conditions.

2.7 Hydride-forming Elements

Conventional FAAS offers poor sensitivity and is prone to interferences for the hydride-forming elements such as As, Se and Sb. However, measurements via the hydride-generation reaction circumvent these problems and provide a useful method for the determination of these elements at trace and ultratrace levels.[4] The formation of the hydride requires that the element in the sample should be present in ionic form. Hence the destruction of organic matter to provide a simple inorganic matrix is necessary prior to hydride generation. The majority of published work prefers acid digestion using HNO_3, H_2SO_4 and $HClO_4$ to accomplish this. Loss of analyte during digestion is more of a potential problem and recovery studies indicate that it is crucial to ensure that the digestion vessel has a sufficiently long neck (Kjeldahl tube or tall boiling tube) to act as condenser and to prevent excessive heating, or to use a closed pressure digestion vessel. The preparation of specimens for the formation of hydrides should ensure that the digestion is complete, nitric acid is removed and loss by volatilization does not occur. The $HNO_3-H_2SO_4-HClO_4$ procedure is adequate for all types of biological samples. In the case of As, Sb, Se and Te, the oxidative attack must be followed by controlled reduction. As and Sb must be reduced to the trivalent state and Se and Te to the tetravalent state[4] prior to generation of the hydride. An alkaline solution of $NaBH_4$ is the preferred reductant, which reacts with the acidified (usually HCl) sample solution to generate the analyte hydride. The concentration of $NaBH_4$ used has ranged from 0.1 to 10%. Interelement interference is a problem, but the interfering elements can be effectively removed using chelating ion-exchange resins.[4]

2.8 Cold Vapor Atomic Absorption Spectrometry

Mercury is invariably determined by cold vapor AAS because of its high sensitivity and selectivity.[1] Acid digestion with $H_2SO_4 + KMnO_4$ or $HNO_3 + HClO_4$ is the sample preparation technique commonly used for urine, and in systems protected by reflux condensers for blood and tissues prior to cold vapor determination. Tin(II) chloride ($SnCl_2$) is the preferred reductant because it is too weak a reagent to cause any precipitation of transition metals and thus completely eliminates any interference from these metals.[4] In current procedures, L-cysteine is added to stabilize the sample solutions and prevent the premature loss of mercury during the addition of the reductant. Magos developed a direct method without prior digestion involving mixed $CdCl_2-SnCl_2$ for total Hg and $SnCl_2$ alone for inorganic Hg in blood, urine, tissues, foodstuffs and feces.[4] Determination of Hg at the lowest level (down to $3\,ng\,L^{-1}$) requires preconcentration and separation. The collection of Hg as an amalgam on a gold or silver gauze adsorber is the most common approach.[4] Collection from large sample volumes and/or over extended periods of time permits high concentration factors. The Hg is released for determination by electrically heating the amalgam at 600 °C.

2.9 Tissue Solubilization

Tissue solubilization with tetramethylammonium hydroxide (TMAH) is simple, permits large throughputs, gives a clear, homogeneous solution, offers two- to fourfold improved sensitivity due to the organic solvent effect and avoids hazards intrinsic to certain wet digestion procedures. TMAH is commercially available in aqueous, alcohol, 1-propanol, toluene and other solvents such as Lumatom, Soluene 100 and Soluene 350. Typically, 0.1–1 g of dry sample is solubilized with 1–5 mL of TMAH in 8–48 h at ambient temperature or in 2–8 h at 50–70 °C. Occasional agitation is useful. Homogenates are further diluted with water, alcohol, MIBK, etc. as required, just prior to analysis, and then fed directly into the flame for Cd, Cu, Fe, Mn, Pb and Zn determination in liver, brain, placenta, etc.[1] Some workers have found that addition of APDC (25% alcoholic solution of TMAH in 2% APDC) minimizes losses of metals. Standard addition calibration is required. Some of the disadvantages of TMAH solubilization include limited time stability of the homogenates compared with the acid digestates and prolonged digestion times.

2.10 Solid Sampling

Some work exists on the use of solid sampling techniques for the FAAS analysis of tissue samples.[1] Large samples (9 g of liver or beef steak) were processed in a Brinkman Polytron stainless-steel homogenizer to yield low- to submicrometer diameter aqueous homogenate particles (~60% solids). A 500-μL aliquot of the diluted (13% solids) homogenate was injected into a Babington clog-free, high-solids nebulizer. Cu, Mn and Zn were determined by this approach in an air–acetylene flame using the method of standard additions.

2.11 Methods for Various Elements

It is important to emphasize that interferences associated with viscosity in the case of ionizing elements are influenced by nebulizer and burner designs. Not every method listed below will, therefore, transfer directly to all atomic absorption (AA) instruments. Analysts are advised to test for these interferences and ways of removing them. The methods given here refer to specimens and analytes routinely examined. For the not-so-common

preparation methods, and for details on individual elements, the literature should be consulted.[1,5–7] The element groupings listed below are based on the Delves classification.[8]

2.11.1 *Electrolyte Elements Sodium, Potassium, Calcium and Magnesium*

Table 2 presents the volumes of blood, serum or urine required for the FAAS determination of the physiological levels of these elements. The relatively high physiological concentrations and the good FAAS sensitivities of Na, K, Ca and Mg allow sufficiently high dilutions of the biofluids to overcome all viscosity and surface tension interferences. Typical serum dilutions are 1 + 249 for K and 1 + 99 for Na with the addition of Li or Cs as ionization suppressant, and 1 + 49 for Ca and Mg with the addition of a releasing agent such as La (usually 0.01 M La in 0.05 M HCl) to eliminate matrix interferences from phosphate. The determination of these metals in urine can be done using even higher aqueous dilutions than those for serum specimens. Note that flame emission is the preferred method for Na and K, although FAAS is equally good. This is mainly because many commercially available flame emission instruments specifically designed for the simultaneous measurement of Na and K are available.

Table 2 FAAS procedures for some elements in blood and urine[a]

Metal[b]	Procedure	Refs.
Na[c] (S, P)	250-fold aqueous dilution	8–10
Na (U)	50-fold aqueous dilution	9, 10
K[d] (S, P)	50-fold aqueous dilution	9, 10
K (U)	500-fold aqueous dilution	9, 10
Ca, Mg[e] (S, P, U)	50-fold dilution with 0.01% (or 0.01 M) La in 1% (or 0.05 M HCl)	8–10
Cu, Zn (S, P)	5-fold aqueous dilution or 10-fold dilution with 6% butanol	8, 9
Cu, Zn[f] (U)	Direct aspiration	9
Zn (U)	5-fold dilution with 6% butanol	8
Fe[g] (S)	2-fold dilution with protein precipitation solution	9, 10
Fe[h] (U)	10-fold aqueous dilution	9
Au (S)	2-fold dilution with water or 5% Triton X-100	8, 9
Au (U)	Direct	9
Li[i] (S)	10-fold dilution with a diluent composed of 140 mM Na and 5 mM K or 25-fold dilution with water	8, 9

[a] Adapted from Iyengar et al.[1]
[b] S = serum; P = plasma; U = urine.
[c] The use of Li as internal standard affords good analytical control.
[d] Hemolysis or failure to separate plasma from cells within about 4 h causes release of K from erythrocytes and the results will be erroneously high. Unreliable results can be obtained for K with serum, and plasma is recommended for this determination.
[e] Addition of La overcomes phosphate interference. EDTA or Sr(II) may also be used. A dilution factor >50 may be required for urine from individuals who are hyperexcretors of Ca and Mg.
[f] Wilson's disease patients will have high Cu excretion requiring appropriate dilution of urine.
[g] Precipitation solution is composed of 10% TCA, 3% thioglycolic acid and 16.6% HCl. Sample preparation procedure for total iron binding capacity (TIBC) is given in Taylor.[9]
[h] Procedure given is for Fe in urine following chelation therapy.
[i] Serum should be used for Li measurement since lithium heparin is a widely used anticoagulant, and plasma may be grossly contaminated with Li.

2.11.2 *Essential Elements Zinc, Copper, Iron, Selenium and Manganese*

Table 2[1,8–10] shows the reference ranges in serum, blood or urine and the volumes required for a given analysis. Table 2 shows that FAAS methods, which require dilution as the only sample preparation approach, is limited to Zn in serum and urine, and Fe and Cu in serum. Zn is best determined using a modified Meret–Henkin method involving a 10-fold dilution of 100 µL of serum with 6% butanol to overcome all viscosity effects from serum proteins. Delves[8] determined Zn in whole blood by a 1 + 9 dilution with 0.05% Triton X-100, in urine by a 1 + 5 dilution with 6% butanol and in cerebrospinal fluid (CSF) by a 1 + 1 dilution with 6% butanol. The distribution of Zn among the proteins separated from serum has been measured by a combined ion-exchange/FAAS method.

Delves[8] adapted the Zn method for determining Cu by diluting 0.2 mL of serum with 1.8 mL of butanol at a 25× scale expansion. The scale expansion was necessary because at the 10-fold dilution used, the absorbance was only 0.013. If smaller dilutions are used, viscosity effects from the serum proteins must be taken into account. One approach would be to calibrate using viscosity-adjusted standards prepared in glycerol (e.g. 5% for fivefold dilution of serum, 11.6% for fourfold dilution and 22% for twofold dilution). In the case of urine, direct analysis (4 mL + 0.2 mL H_2SO_4) is feasible because of the high physiological concentration (0.2–0.8 µm) of Cu, but NaDDC/MIBK or APDC/MIBK extraction provides greater accuracy.

Fe in serum can be determined by a fivefold dilution with water followed by pulsed nebulization/FAAS. However, the presence of hemoglobin in serum can lead to erroneously high results because the erythrocyte Fe concentration is 1000-fold higher than the serum Fe level. Therefore, protein precipitation methods are advisable. One method involves addition of 1 mL of a solution of 0.1% ascorbic acid in 0.5 M HCl to 2 mL of serum, dilution

of the supernatant with 0.5 mL of acetone to overcome viscosity interference and FAAS analysis. The use of TCA in place of ascorbic acid was not satisfactory because it could not remove the interference from NaCl at the 248.3-nm line. Urine Fe measurement requires solvent extraction and preconcentration. The high background at the 196-nm line and the poor sensitivity preclude the use of conventional FAAS for measuring Se in biological materials. Hydride generation techniques have resulted in sensitive methods.[9]

The determination of Mn in serum, whole blood and urine by FAAS requires considerable improvement in detection limit through preconcentration. Therefore, scrupulous control of contamination would be essential. Delves determined Mn in 2-mL samples of whole blood after HNO_3–$HClO_4$–H_2SO_4 oxidation and extraction at pH 8.5 with 8-hydroxyquinoline (oxine)–MIBK.[8] A fourfold concentration allowed measurements to be made down to 50 nM. Van Ormer and Purdy analyzed 100-mL volumes of urine after acid digestion and extraction with cupferron at pH 7 with 30 mL of MIBK.[1] They could detect down to 1.8 nM. Many other chelating–extraction schemes have been employed.[8] Analyses of metals such as Co, Cr, Mo and V require some form of pretreatment to provide absorbance signals of sufficient magnitude to allow precise measurement of reference ranges. Since FAAS analyses of these metals in biofluids require preconcentration from large volumes, alternative techniques such as GFAAS are more appropriate.

Digestion and dissolution, solubilization or acid extraction are the possible approaches for diets, feces and tissues.[1] In the case of electrolyte elements, a simple ashing procedure prior to dilution will suffice. An attractive procedure for the determination of Zn in diets and feces without sample oxidation involves acidification of the samples with 6 mL of 1 M HCl for 24 h, extraction of the fatty components with chloroform and centrifugation of the solution after the addition of methanol to reduce surface tension. Tissues such as liver and kidney which are rich in elements such as Cu, Zn, Fe and Mn are also amenable to FAAS analyses.

2.11.3 Therapeutic Elements Gold and Lithium

The FAAS of Au in the serum of arthritic patients receiving chrysotherapy can be carried out easily by 1 + 2 aqueous dilution because of the high concentrations encountered. However, solvent extraction methods are advisable owing to matrix interference problems.[1] Li can be determined by a 25-fold aqueous dilution of serum because of its high therapeutic range (0.8–1.5 mM). No matrix problems are encountered at this dilution.

2.11.4 Nonessential Elements Lead, Cadmium, Nickel, Arsenic and Mercury

The poor sensitivity of FAAS for Pb precludes its direct determination in blood. FAAS method requires preconcentration, usually by solvent extraction.[1] Lead may be released from blood proteins by TCA precipitation, haemolysis with Triton X-100 or complete acid digestion prior to extraction. Acid digestion as pretreatment for solvent extraction with APDC/MIBK has been used for the FAAS analysis of Pb in tissues, diets and feces. In the case of urine, Pb may be directly extracted at pH 2.8 using APDC/MIBK. The large volumes (5 mL) of blood required for these solvent extraction/FAAS methods will be unacceptable for pediatric work. Pediatric samples may be analyzed by discrete sampling FAAS using the Delves cup method. The procedure involves oxidation of 6 µL of blood with 20 µL of 20% H_2O_2 in a nickel cup prior to insertion into an air–acetylene flame. The method could also be used to measure Pb in lung, liver and kidney samples.[1]

The direct determination of Cd in blood and urine is not possible because of the poor FAAS sensitivity. The APDC/MIBK extraction method has been used for measuring Cd in tissues, diets, feces and urine after acid digestion or after TMAH solubilization. However, unacceptably large volumes of blood (at least 10 mL) would be required to measure the very low concentration of Cd ($<0.2\,\mu g\,L^{-1}$) observed in unexposed humans. Discrete sampling FAAS with the Delves cup method may be used to measure Cd in 15 µL of blood down to $0.3\,\mu g\,L^{-1}$. Solvent extraction and preconcentration are necessary to measure the low levels of Ni found in body fluids and tissues. Sunderman et al. have determined Ni by APDC/MIBK extraction of acid digested serum, urine and tissues.[1]

2.12 Methods for Urine, Hair, Nail, Feces, Milk, Tooth and Bone

The above discussion mainly dealt with blood components and soft tissues. Given below are sample preparation methods applicable to some of the other matrices.

A urine sample should be vigorously shaken before aliquoting since some elements, such as As, Hg and Pb, tend to adsorb on container walls or coprecipitate with the particulate matter (Cd, Cr, Cu, Hg, Ni, Pb) settled on storage. The urine matrix is so variable that it can give rise to serious and unpredictable matrix effects. Therefore, few elements can be determined by direct FAAS involving only dilution. Also, untreated or inadequately treated urine tends to create foaming problems. Release of the analyte from its chemical bond with the urine matrix or destruction of the organic matter can be attained by complete ashing or by a simple minor

pretreatment of urine with HNO_3, TCA or H_2O_2. The latter approach is recommended for its speed, simplicity and lower blank; it is often practised prior to extraction and other preconcentration steps, and in direct assays by hydride generation or cold vapor AAS. Wet digestion and dry ashing methods can also be used.[1]

One of the preparatory steps in hair analysis is washing.[11] Ideally, any reliable washing procedure should remove exogenous dirt, adherent air particulates, dead skin and adsorbed vapor (e.g. Hg) without significantly leaching the endogenously incorporated trace elements. A number of washing solutions have been applied, such as water, dilute acid (1% HCl or HNO_3), dilute EDTA, detergent solutions, organic solvents, and also different washing conditions, such as soaking, stirring, ultrasonification and heating of wash solution to 100 °C. Mild washing with water or detergent solutions closely resembles the usual in situ washing of hair, but has resulted in the partial leaching of elements from hair (e.g. Cu, Fe, Rb, Zn, Ca, Cl, Na, K, Mg). Organic solvents effectively remove grease and fat with little extraction of the trace elements. Ether, acetone and ethanol are most often used, but have been criticized as unnatural and ineffective. Strong washing solutions such as EDTA and acids are not recommended as they extract both the trace elements and inorganic matrix components of hair. Even aqueous solutions of detergents with pH < 6–7 have been found to leach trace elements. For the sake of standardization, many analysts follow the International Atomic Energy Agency (IAEA)-recommended procedure of washing, which consists of five successive washes, namely acetone–water–water–water–acetone, each for 10 min in 25 mL of solvent.[1] However, this has also been criticized. In summary, there is no consensus on the washing procedure for hair. The main difficulty arises in cases of major exogenous contamination, especially with the determination of ubiquitous elements such as Al, Cr, Fe, Mn, Ni, Pb, Si and V.

Homogeneous pools of hair can be prepared by cutting hair into 2–3-mm pieces or by grinding at liquid nitrogen temperature.[10] Hair can be brought into solution by acid treatment or TMAH solubilization. Generally, 10–50 mg of the sample are digested with <1 mL of HNO_3 in small test-tubes which can be screw-capped or pressurized. The resulting HNO_3 digests are diluted and injected directly into the flame for determination of metals such as Cu, Fe and Zn. In the case of volatile elements such as As, Se and Hg, complete digestion may be required; this can be accomplished using HNO_3–H_2O_2 or HNO_3–$HClO_4$ prior to hydride generation or cold vapor AAS. In the case of TMAH solubilization, usually 50-mg specimens are incubated at 50 °C with 0.1 mL of water and 1 mL of base (Soluene, Lumatom, etc.) for 3–24 h, and the contents are subsequently diluted with alcohol or toluene. Dry ashing of hair at 450 °C, followed by dissolution of the ash in HNO_3, is also practised in the determination of non-volatile elements in larger samples.

Nail[1] specimens are, in general, found to be more severely contaminated with metals than hair, and hence require more thorough cleaning. The surface layer is scraped off using a quartz or titanium knife, a platinum-coated razor blade or Teflon®-coated forceps. This is followed by cleaning in an aqueous nonionic detergent or an organic solvent, or both. Application of heating or ultrasound during washing is also practised. Nail specimens are easily digested with HNO_3, HNO_3–$HClO_4$ or HNO_3–H_2SO_4. For most nonvolatile elements except noble metals, TMAH solubilization is also useful. Complete digestion will be required for volatile elements.

In the case of feces analysis,[1] thorough homogenization is required because of the highly variable composition. Feces can be homogenized by adding an equal amount of water followed by shaking for a few hours. Alternately, homogenization can be effected by the brittle fracture technique, which involves grinding or shattering at liquid nitrogen temperature. Smaller samples can be ground after lyophilization. The homogenates should be stored deep-frozen or lyophilized or refrigerated if analysis is to be deferred. Feces are a difficult matrix because of their high content of organic matter, especially fat, and the inorganic salts, especially those of Na, K, Ca and Mg. In order to destroy the fat content completely, digestion with a mixture containing $HClO_4$ or H_2O_2 is desirable. Typically, 1–2 g of feces is digested with 10–15 mL of HNO_3 with controlled heating followed by oxidation with $HClO_4$ or H_2O_2. Several other wet ashing procedures have been described by Tsalev.[11] Acid-extraction procedures involving incomplete digestion of organic matter have also been described. Dry ashing of feces at 450 °C has been applied, but insoluble residues of oxides, phosphates and silicates tend to persist, giving rise to problems during FAAS analysis. Also, volatile elements such as As, Cd, Cu, Hg, Pb, Se and Zn are not quantitatively recovered.

In general, milk[1] is a difficult matrix owing to its high and variable fat content (approximately 4% maximum) and inorganic salt content, especially of Ca, K and P. Fat is not completely decomposed by most conventional wet digestion procedures. Hence pressurized decomposition in polytetrafluoroethylene (PTFE)-lined bombs with HNO_3 or $HNO_3 + V_2O_5$ is advisable. Combustion in oxygen is an effective technique for volatile elements such as As, Hg, Sb, Se, Sn and Te. Other approaches include dilution of the sample with an aqueous detergent such as 0.1% Meriten (nonyl phenyl polyglycol ether), 0.1% saponin or 0.05% Triton X-100 for Cu, Fe and Zn, protein precipitation with TCA acid for Cu, Fe and Zn and acid extraction with 1–2 M HCl for Mn and Pb or H_2SO_4

for Cd, Cu and Fe.[1] Dry ashing at 450–500 °C is often practised owing to its simplicity, thorough fat oxidation and applicability to many elements. Depending on the individual element to be quantitated, the ash may require further treatment with acids. Tsalev[11] has detailed the procedures available for the various individual elements. Preconcentration and separation from the matrix are often required for Cd and Pb, and especially for Ag, Bi, Co, Sn and V.

In the case of teeth,[1] the sample is scraped from adherent soft tissue and tartar using polyethylene (PE) or PTFE instruments, then washed in high-purity water with a stiff plastic brush and finally rinsed with high-purity water. In determining elements such as Cd and Pb, the quantitative separation of dentine and enamel is essential since these elements are higher in dentine by one to two orders of magnitude. One of the most popular techniques for the separation of enamel and dentine involves powdering the tooth by crushing and grinding, mixing the powder with a heavy fluid such as bromoform, ultrasonification and then centrifugation. Another density separation may be required to remove the carious tissue. Pressurized digestion with HNO_3 in PTFE-lined bombs is the most popular digestion method for dissolving small specimens of tooth enamel. Whole tooth and bone, which contain more organic matter and fat, are digested with combinations of acids: $HNO_3 + HClO_4$, $HNO_3 + HCl + H_2O_2$, etc. Rapid microwave digestion of teeth with a mixture of $HNO_3 + HClO_4 + H_2SO_4$ has also been described. Only Fe, Sr and Zn, and perhaps Mn and Pb, can be determined directly in the bone/tooth digests by FAAS. Because of matrix effects in these direct techniques, matrix-matched calibration should be used. Determination of Cd, Cu and Pb is best effected by chelation–extraction of the digested bone or tooth. Coprecipitation methods have also been described for Ba and Sr.[11]

3 GRAPHITE FURNACE ATOMIC ABSORPTION SPECTROMETRY

3.1 Introduction

Sample preparation methods for GFAAS have been comprehensively reviewed by several authors.[1] Discussions on individual elements are also contained in these reviews and in books.[5–7,11] Therefore, this article will be a summarized version of these reviews, which the reader may consult for detailed information.

The available sample preparation methods can, in general, be classified as follows: (i) direct injection; (ii) use of diluent solutions to minimize matrix and molecular absorption interferences; (iii) partial decomposition techniques in which metals are extracted from proteins with acids; (iv) dissolution of tissue samples without complete oxidation; and (v) complete oxidation procedures such as dry ashing, wet digestion at ambient and elevated pressures and LTA with reactive gases at low pressures. In addition, solid and slurry sampling methods have been used in the case of tissue samples. Each of these methods is briefly discussed below.

3.2 Direct Methods

Direct methods seem ideal for laboratories engaged in the routine analysis of a large number of samples because they are simple and rapid, involve little sample pretreatment, reduce the risks of contamination and do not require considerable operator skill. Earlier publications advocated direct injection of microliter volumes of biological fluids such as blood or serum into the furnace. However, the results obtained were not accurate because of excessive background absorption, carbonaceous residue buildup within the graphite tube and frothing, foaming and splattering of the samples during the drying stage. Hence some form of sample pretreatment was required to obtain reproducible and reliable results.

3.3 Dilution

Biological fluids have been diluted with water, Triton X-100, acids (especially nitric acid) or ammonia. However, aqueous dilution of whole blood is often associated with time-dependent interferences. A precipitate of red cell membrane appears within 1–2 min and reaches a maximum within 15 min. This is of particular importance when large numbers of samples may be prepared hours in advance of the analysis. Dilution of serum, plasma or urine seldom presents serious problems. Dilution factors as high as 50 or 100 can be used for metals such as Cu, Fe and Zn because of their high levels in serum and excellent GFAAS sensitivity. At such high dilutions, matrix effects are insignificant and these three elements can be easily determined in serum with a 50-fold dilution with water. Oxidizing acids such as HNO_3 or $HClO_4$ at the 1–10% level have been used as diluents for biological fluids to help reduce molecular absorption interferences and as reagents to modify the effect of the matrix upon analyte sensitivity. However, dilution of blood samples with these acids is generally unsuccessful because the protein/red cell precipitation results in a turbid mixture which is more difficult to pipet than the original solution. An often overlooked reagent is dilute ammonia solution. A 1 + 19 dilution of whole blood, using 1% (v/v) ammonia solution, provides complete cell lysis within minutes of mixing, and the diluted samples are stable for up to 48 h when stored at 4 °C.

Triton X-100, a nonionic surfactant, has found widespread use as a diluent because it causes cell lysis, provides a clear homogeneous solution from whole blood, minimizes frothing of whole blood, reduces the sample/graphite interfacial tension and improves the contact between sample and furnace walls. However, this reagent can diffuse into the graphite, leading to variable rates of atomization. The use of a pyrolytically coated graphite tube in conjunction with a L'vov platform can reduce but not eliminate this problem. The concentration of Triton X-100 used has varied from 0.05 to 2%. However, the final concentration injected into the atomizer should be preferably less than 0.25% (v/v). Higher levels of Triton X-100 can cause buildup of a carbonaceous crust within the graphite tube and can degrade the furnace performance.

3.4 Deproteinization

Treatment of diluted whole blood with 3–5% nitric acid leads to complete deproteinization and quantitative release of metals. Centrifugation of the deproteinized samples yields clear supernatant fractions that contain the metals, and which can be easily dispensed into the atomizer using autosamplers. The method is simple, rapid and contamination-free since all operations can be done in the same sampling tube. Other deproteinizing agents used include TCA and $HClO_4$ (10% final concentration) in acetone. The use of $HClO_4$ in acetone is potentially hazardous and is not recommended.

3.5 Matrix (Chemical) Modification

A few simple chemical reagents added directly to biological fluids or to tissue digests can minimize matrix effects upon analyte sensitivity either by modifying the volatility of the analyte or of the matrix. The modifier generally allows the formation of a more thermally stable analyte/modifier compound which allows the matrix to be separated from the analyte and removed from the atom cell prior to atomization. Many reagents have been used as modifiers in biological analysis.[1] The various ammonium phosphates, either alone or in combination with magnesium nitrate, have been used for the determination of Cd and Pb in whole blood and urine. These permit pyrolysis temperatures of approximately 650 and 950 °C to be used for Cd and Pb, respectively. Other matrix modifiers used include Ag, Cu, Ni, Pt, Pd and V. Since Cu and Ni are often determined in biological materials, their use as matrix modifiers is not advisable because of possible furnace contamination. Welz,[4] based on his group's studies and the studies of Ni et al., advocated the use of Pd(II) either alone or in conjunction with magnesium nitrate as a 'universal' modifier applicable to as many elements as possible.

3.6 Dissolution of Tissues

Clear homogeneous solutions were obtained when 100–300-mg samples of brain, liver, muscle and placenta were heated at 60 °C with 1–3 mL of a 25% alcoholic solution of TMAH containing 2% APDC, and the cooled solutions were diluted with toluene. The solutions were analyzed for Cu, Fe, Mn, Pb and Zn. Although 26 different types of tissues have been dissolved using only a 25% alcoholic TMAH solution followed by water dilution, the standard solutions of metals formed precipitates with the basic TMAH in the absence of APDC. Therefore, the use of TMAH in conjunction with APDC is recommended.

3.7 Complete Oxidation of Tissues and Fluids

The procedures mentioned in sections 3.1–3.5 are applicable only to biological fluids, and not to tissues. In the case of tissues, it is necessary to oxidize them completely using either hot mineral acids (wet ashing) or by prolonged heating at elevated temperatures (dry ashing) prior to bringing them into solution suitable for analysis. The book by Tsalev[11] may be consulted for the various combinations of acids used in wet digestion. Particular mention may be made of vapor-phase oxidation and pressure digestion. The sample is placed inside a reaction vessel containing nitric acid and fitted with a reflux condenser. The sample attack, which takes 3–4 h, is from the vapor phase only, ensuring that no acid accumulates in the sample and the contamination is minimized. The digestion of tissues can be accomplished rapidly by heating the tissue samples to 110–170 °C with oxidizing acids in sealed pressure vessels. The process can be made especially faster using microwave heating, which has become very popular in recent years.

Dry ashing of samples in muffle furnaces is rarely used because of potential contamination, analyte loss and possible matrix effects because of the high levels of the dry ashing aids (e.g. magnesium salts) used. LTA using gas mixtures such as ozone or carbon tetrafluoride in oxygen seems very effective for the decomposition of biological tissues and blood, and merits further study.

All the sample preparation methods described above may be used with solvent extraction and other separation techniques to concentrate the analyte and separate it from the matrix.[1] However, the potential for contamination with preconcentration is great with a sensitive technique such as GFAAS. Nevertheless, two preconcentration approaches are worthy of mention. The on-line FIA coprecipitation system mentioned earlier in connection with FAAS has been adapted for the GFAAS determination of Cd and Ni in digested whole blood. The small sample volumes of 0.12–0.15 mL (1.2–1.5 mL of digest) required for the preconcentration make it ideal for sample-limited applications such as pediatric blood

and serum analysis. In the second approach, the gaseous hydrides generated in a continuous or flow injection system are introduced via a quartz capillary into a graphite tube atomizer coated with a palladium–iridium modifier where the hydrides are decomposed and the metalloid atomized.[1] This approach not only gives high sensitivity because of the preconcentration, but also eliminates any kinetic interferences in the generation stage by separating it from the atomization step.

3.8 Solid Sampling

Solid sampling is an attractive approach for the GFAAS analysis of biological tissues, especially hair, nail, bone, teeth, etc. Direct introduction of solid tissues into the graphite furnace has several advantages,[12] including shorter sample preparation time compared with dry ashing or wet digestion, minimal sample contamination, reduced loss, especially of volatile elements, and increased sensitivity. The enhanced sensitivity is particularly useful when only small amounts of sample are available for analysis. Herber[13] has summarized the feasibility of the solid sampling approach for the determination of a number of elements in biological materials. Readers may also consult the proceedings of four European colloquia on solid sampling.[14] Solid sampling has been applied to the determination of Al, As, Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Se and Zn in biological tissues, especially liver and muscle.[12,13] Soft tissues such as liver, muscle and kidney need only freeze-drying and subsequent milling. Hair may be crushed by adding liquid nitrogen and stirring with a quartz rod. Hard tissues such as bone and teeth can be treated with liquid nitrogen followed by milling or pulverizing. Suitable milling devices can be made from agate, carbide or titanium. Calibration may be performed against natural or synthetic solid standards, against standard addition or against aqueous standards. Herber[15] has given a list of the available Bureau Communautaire de Réference (BCR) biological solid standards. Although solid sampling is attractive, there are major problems. Thus, it is difficult to obtain a homogeneous, representative sample in microgram quantities since samples have to be introduced into the furnace in microgram amounts in order to minimize inorganic matrix effects.[14] The microweighing of samples is difficult and tedious, and the transportation of these microgram quantities requires special care even when the sample is weighed directly on a boat, platform or cuvette that can be directly inserted into the furnace.[12]

One possible solution to the problem is to use slurries of the biological tissues.[12,14] Homogenization of the slurry can be achieved in situ by automated ultrasonic agitation[16] and the homogeneous slurry can be immediately pipetted into the graphite furnace using conventional liquid-handling devices such as autosamplers or pipets. According to Miller-Ihli,[16] reliable results can be obtained using as little as 1 mg of finely ground ($<500\,\mu m$) sample suspended in 1 mL of dilute acid (e.g. 5% HNO_3) or water which contains 0.005% Triton X-100 to aid particle dispersion. Nevertheless, slurry sampling of biological materials is still in its infancy, and more studies will be required before its applicability is firmly established. Slurry sampling seems best suited for powders such as soils, and biological tissues are not powders unless they are freeze-dried and pulverized, in which case many of the advantages of solid sampling may be lost. For example, the reagents used in slurry preparation may lead to contamination. Also, the small mass used in slurry sampling entails a more stringent need for homogeneity than in the case of solid sampling.

4 INDUCTIVELY COUPLED PLASMA ATOMIC EMISSION SPECTROMETRY

4.1 Direct Methods

Direct methods with or without dilution are applicable to the ICPAES and ICPMS analyses of biological fluids provided the analyte concentrations are at least 10 times higher than the detection limit. This is the case for the ICPAES determination of Na, K, Ca, Mg, Ba, Cu, Fe and Zn in serum and urine.[1] The samples are usually diluted with high-purity water or acids by a factor of 5–10 to minimize clogging of the nebulizer by the lipid matter, and the resulting solution is injected into the plasma. Al, Cu and Zn have been determined by dilution of serum with 1.3 M KCl. The use of KCl serves to enhance the Al signal and minimize matrix effects. Sample sizes of >1 mL of both the fluids are usually recommended. It is also possible to minimize both the sample size and dilution factor using an on-line FIA system for the simultaneous determination of Ca, Fe, K, Li, Mg, Na and Zn in 20 µL of a 1 : 1 aqueous diluted serum. Dilution of serum with Triton X-100 results in fouling of the nebulizer and torch by protein deposits, and therefore is not recommended.

An interesting direct analysis approach involves the electrothermal vaporization (ETV) of microliter volumes of biological fluids in a graphite furnace, cup or rod, or tungsten loop, and the subsequent transportation of the generated aerosol vapor into the inductively coupled plasma (ICP) in a carrier-gas stream. The ETV/ICPAES method requires only 5–10 µL of samples, offers good precision and minimizes matrix effects.[17] The method has been used for the analysis of whole blood, serum, urine and ceruloplasmin. Also, some tissue samples have been dissolved in TMAH, and microliter volumes of the solution have been analyzed by the ETV/ICPAES method.

4.2 Deproteinization

The simultaneous screening of Na, K, Ca, Mg, Fe, Cu and Zn in serum has been achieved by vortex mixing 1-mL sample aliquots with a reagent composed of TCA and HCl (deproteinizing agents), hydroxylamine sulfate (for Fe reduction) and yttrium (internal standard), centrifuging the mixture and aspirating the clear supernatant into the ICPAES system.[1] A nitric acid deproteinization method has been used for the ICPAES determination of high levels of Cd ($>30\,ng\,mL^{-1}$) and Pb ($700\,ng\,mL^{-1}$) in whole blood.[1] As in the case of AAS, some workers have dissolved tissue samples in 5–10% TMAH, and aspirated the resulting solution into the ICP.

4.3 Oxidative Decomposition

Oxidative decomposition by either wet or dry ashing is the most widely used sample preparation method for the ICPAES analysis of biological materials. Generally, the digestion for multielement analyses would require complete destruction of the organic matrix and decomposition of the inorganic concomitants. Any of the wet or dry decomposition methods used in AAS analyses can be applied. The reagents most commonly used are nitric acid–perchloric acid or aqua regia. The nitric acid–perchloric acid method will destroy most organic matrices and decompose the inorganic constituents. Aqua regia digestion is also effective unless there is concern over volatile analytes such as arsenic.

The sample preparation studies of Dalquist and Knoll and Ward et al. have been the most comprehensive.[1] They evaluated the suitability of dry ashing and wet ashing, with either nitric acid–perchloric acid or nitric acid–hydrogen peroxide, for the ICPAES determination of 18 elements in serum, bovine liver and muscle tissues. The nitric acid–perchloric acid method had a slight edge over the other two methods. Although nitric acid–perchloric acid requires the use of a perchloric acid fume hood, quantitative recoveries were obtained for all the elements investigated, including potassium. Templeton[18] has summarized selected sample preparation methods that he considers would lead to reliable analysis of biological fluids and tissues by ICPAES. The information is presented in Table 3. The nitric acid–peroxide digestion procedure proved to be tedious because of the need to add the peroxide dropwise to the cool solution for maximum oxidizing power. The high dilution factors required to give a solution that can be nebulized over several hours of operation without building up fats in the nebulization system is also a distinct limitation for the determination of trace elements.

In recent years, digestion involving mineral acids, especially HNO_3 or HNO_3–$HClO_4$ in closed vessels under pressure, and heated thermally or with microwave energy, has become the preferred method.[19] PTFE vessels may not lead to complete decomposition of samples because of temperature and pressure limitations, but quartz vessels which can withstand temperatures up to 300 °C at a pressure of 80 bar permit complete destruction of organic matter even without the use of $HClO_4$. Closed-vessel digestions minimize external

Table 3 Selected biological sample preparation procedures for ICPAES analysis[a]

Sample type	Elements determined	Preparation procedure
Whole blood, hair	Al, As, Cd, Cr, Cu, Fe, Ni, Pb, Se, Zn	HNO_3–H_2O_2 digestion; hair washed prior to digestion
Serum	Al, Cu, Zn	Dilution with 1.3 M KCl
Serum, urine	Al, Au, Li, Pt	(i) Direct analysis with ETV (ii) HNO_3–$HClO_4$ bomb ashing
Serum	Ca, Fe, K, Li, Mg, Na, Zn	Flow injection with aqueous or detergent (Brij) dilution
Urinary stones	Al, Cd, Co, Cr, Cu, Fe, Li, Mn, Mo, Ni, Pb, Sr, Ti, Zn	Sample ground; digested with aqua regia–H_2O_2
Liver, kidney	Ag, Ba, Be, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, P, Sr, Th, Ti, Zn	HNO_3–$HClO_4$ digestion
Bovine liver, several rat organs	Ag, Al, As, Ba, Ca, Cd, Co, Cr, Cu, Fe, Hg, K, Li, Mg, Mn, Mo, Na, P, Pb, S, Sb, Se, Si, Sr, Ti, Tl, V, Zn	Several wet ashing methods
Bone	(A) Al, As, Ba, Ca, Fe, Hg, Mg, Mn, Na, P, Se, Sr, Zn; (B) Cd, Co, Cu, Mo, Ni, Pb, Ti, V	Powdered bone wet ashed with HNO_3–H_2O_2; group B elements determined after concentration on poly(dithiocarbamate) resin

[a] Adapted from Iyengar et al.[1]

contamination and loss of volatile elements, permit rapid sample decomposition (e.g. 20 min for microwave vs 3 h for conventional hot-plate) and give low analytical blanks. These systems have become increasingly automated and reliable, and are commercially available.

Dry ashing is relatively simple and ideally suited for batch-type procedures. Also, a smaller dilution factor could be employed to improve the sensitivity of some of the trace elements since the solid loading of the dissolved ash is small. However, volatile elements such as arsenic and selenium may be lost if care is not exercized in controlling the ashing temperature. In addition, low recoveries have often been obtained for Al, Cr and Fe.

The ICPAES determination of trace elements (e.g. Cd, Co, Cr, Mn, Ni, Pb, Sn and V) in biological samples requires some form of preconcentration/separation because of inadequate detection limits. It is seldom possible to dissolve the residue resulting from wet or dry ashing without diluting the original analyte concentration. Any preconcentration/separation method chosen should preserve the multielement capabilities of the ICP, should provide a high concentration factor and should minimize matrix effects. These criteria seem to be satisfied mainly by chelating ion-exchange resins, and it is no surprise that the majority of preconcentration/separation procedures are based on these resins. In addition, automation of the preconcentration/separation procedure using flow injection techniques in conjunction with ICPAES measurements has been shown to improve efficiency, sensitivity and selectivity.[20] Recently, on-line preconcentration/separation systems have become available as accessories for spectrometric techniques.

Horváth et al.[21] have comprehensively reviewed the preconcentration and separation techniques for ICPAES and ICPMS. Some selected examples are presented here. Thirteen elements were separated and concentrated from perchloric acid–nitric acid- or perchloric–nitric–sulfuric acid-digested solutions of bovine liver and oyster tissue samples using the Chelex-100 chelating resin in the pH range 4.7–5.5. Only Cd, Cu, Mo, Ni and Zn could be consistently recovered quantitatively in the Chelexed fraction by elution with dilute nitric acid. The As, Sb and Se in an aliquot of the digestate were concentrated, separated and quantitatively recovered as the hydrides, which were introduced into the plasma. Horváth et al.[21] used the poly(dithiocarbamate) chelating resin to preconcentrate and determine simultaneously 10 or more trace elements in urine by ICPAES. They obtained a concentration factor of 125 and were able to reduce the potential interference of alkali and alkaline earth metal ions since the resin showed little affinity for these major elements. Elements such as As, Bi, Sb, Se, Sn and Te were introduced into the plasma as the hydrides. They were also able to speciate Cr(III) and Cr(VI), Se(IV) and Se(VI) and Te(IV) and Te(VI) by carefully adjusting the pH prior to resin complexation. The same approach was extended to the analysis of bone. In the case of serum or bovine liver, the sample was pressure digested with nitric acid–perchloric acid, the digested residue was diluted with water, the pH was appropriately adjusted (e.g. pH 4.5 for Cd, Co, Mo and V and pH 8 for Cu, Fe and Zn), the solution was passed through the poly(dithiocarbamate) resin column, the resin was decomposed using nitric acid–hydrogen peroxide and the digested resin solution was used for the ICPAES analysis. Schramel et al.[22] reported the reliable simultaneous determination of Cu, Mn, Ni, V and Zn in certified reference samples of mussel, bovine liver, milk powder and human hair by an on-line preconcentration of the nitric acid pressure-digested solution at pH 6.5 on an oxine–cellulose microcolumn coupled to the ICP.

4.4 Solid Sampling

Powdered samples of bovine liver have been analyzed by the ETV/ICPAES approach.[1] The sample was micropulverized using the cold brittle fracture technique, 0.5 mg was weighed into a graphite microboat, which was placed on the carbon rod atomizer, the sample was atomized and the aerosol vapor was transported into the ICP for the simultaneous determination of Ca, Cd, Cu, Fe, K, Mg, Na, Pb and Zn. The results agreed with the certified values within ±15%. The analysis of single strands of hair using a direct sample introduction device for Cu and Zn has also been described.[1] The ETV/ICPAES technique is of recent origin, and increased interest in its use for biological analysis is expected. Rapid ETV/ICPAES procedures with as little sample preparation as possible would be useful in hospital and clinical laboratories.

5 INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY

5.1 Introduction

The biological applications of ICPMS have been reviewed.[23] Few specimens can be analyzed directly. Direct analysis causes problems such as blockage of the pneumatic nebulizer, the central quartz tube of the torch and the sampling orifice due to the high protein content of biological fluids. The problem is especially severe when the dissolved solids content is >0.2%. The high solids content can also cause spectral and matrix interferences. Anions and metals in Groups IA and IIA of the periodic table are the major interferents, and are present

in abundance in biological materials. Therefore, some form of sample preparation mainly to convert the biological specimen into an appropriate solution is required. Depending on analysis requirements and specific interferences encountered, separation and preconcentration steps would also be required during sample preparation. The sample preparation methods used in ICPMS are presented below.

5.2 Dilution

The dilution method has been applied to the analysis of whole blood, serum, urine and milk powder. Dilution is possible in the case of ICPMS because the majority of elements have detection limits below 1 $\mu g\,L^{-1}$. Delves et al. diluted whole blood 25-fold with a diluent composed of ammonia, $(NH_4)_2H_2EDTA$, $NH_4H_2PO_4$, Bi and Triton X-100 for the measurement of total Pb concentration and stable Pb isotopic ratios at blood Pb levels $>100\,\mu g\,L^{-1}$.[1] The ammonia, EDTA and phosphate solutions impart a temporal stability of at least 48 h to the diluted blood solution by maintaining the pH at 8.5 to prevent any precipitation. The use of Bi as the internal standard overcomes the diurnal variations in sensitivity. Addition of Triton X-100 prevents blockage of the nebulizer, improves the wash-out of the sample after the analysis and improves the nebulizer efficiency.

Al, Ba, Bi, Br, Cd, Co, Cs, Cu, Fe, Li, Mg, Mn, Mo, Ni, Pb, Rb, Sn, Sr and Zn have been determined in human serum by a 5- or 10-fold dilution of the sample either with 0.14 M ultrapure nitric acid or with ultrapure water.[23] Internal standards such as Be, In or Tl were used to correct for matrix effects. Uranium in human plasma was determined by a threefold dilution of the sample with a diluent consisting of 1% nitric acid and 25 $mg\,L^{-1}$ Eu (internal standard).[1]

Attempts have been made to determine directly the elemental composition of urine either by using a Hilderbrand grid nebulizer, which permits aspiration of concentrated salt solutions or using on-line FIA together with direct injection nebulization.[1] A 3- or 10-fold dilution of the urine sample with water or 1% or 0.14–0.28 M nitric acid has also been used.[22]

Direct, simultaneous determination of 35 elements in CSF by direct introduction of the sample into the ICPMS system has been reported.[1]

The simple direct and dilution methods outlined above have been found to be inadequate for the measurement of elements between 40 and 80 u, especially elements such as As, Cr, Se and V because of spectral interference caused by the formation of polyatomic species such as $^{16}O^{35}Cl^+$, $H^{16}O^{35}Cl^+$ and $^{40}Ar^{35}Cl^+$ by interaction with the NaCl present in serum and urine. Minimization of these interferences requires separation of the analytes from the matrix. Thus, chromatographic methods such as gel filtration, ion chromatography and anion exchange have been used to remove the chloride interference prior to ICPMS determination of total elemental levels or isotopic composition of aqueous dilutions of serum and urine samples.[1] Arsenic has been preconcentrated and separated as the hydride from a 100-fold diluted urine.[1]

5.3 Oxidative Decomposition

The most widely used approach involves wet ashing techniques followed by dissolution and dilution of the residue to a specific volume prior to analysis. Nitric acid is preferred since it produces the fewest polyatomic interferences, in contrast with $HClO_4$ and H_2SO_4, which introduce polyatomic ions such as ClO^+ and SO^+. Although open digestion (e.g. hot-plate) has been used, the current trend is to use microwave digestion since this permits complete decomposition of the sample with nitric acid alone in the majority of cases (see discussion in section 4). The digested residue is dissolved in water or dilute nitric acid, but the final acid concentration should be kept below 10% to minimize corrosion of the metal sampler and skimmer cones. Vanhoe[23] determined 11 elements in various biological materials such as milk powder, wheat flour, rice flour, bovine liver and pig kidney after subjecting 0.3–0.4 g of sample to a two-step microwave digestion with nitric acid. The two-step program was necessary to avoid damage to the decomposition vessels. The digested residue was suitably diluted with ultrapure water prior to analysis. Others have used a single-step hot-plate, pressure bomb or microwave digestion for the multielement analysis of the above samples and also cod liver, shellfish tissue, lobster hepatopancreas, dogfish liver and muscle tissue, human autopsy specimens, rat serum and liver and human whole blood and serum.[1] In general, 100–400-mg tissue samples and 0.5–1-mL blood or serum samples were used in these digestion procedures. Internal standards were used in all cases.

Friel et al. used a nitric acid–microwave digestion followed by an open-beaker digestion (hot-plate) with nitric acid–hydrogen peroxide for bovine liver, oyster tissue and animal muscle.[1] The open-beaker digestion was found to be necessary for complete decomposition of the organic matrix and for reduction of the chlorine content which form polyatomic species that interfere with the determination of As, Cr and V. Also, they found a severe loss of sensitivity when imperfectly digested samples were analyzed.

Plantz et al. found it necessary to digest urine samples in nitric acid–perchloric acid, separate the analytes by complexation with bis(carboxymethyl)dithiocarbamate and adsorb the complexes on a poly(styrene–divinylbenzene)

Table 4 Selected biological sample preparation procedures for ICPMS analysis[a]

Sample type	Preparation method/ internal standard	Elements determined
Whole blood	Dilution, matrix modifier/Bi	Pb
Serum	Dilution with 0.14 M HNO_3/In	Co, Cs, Cu, Fe, Mo, Rb, Sr, Zn
Serum	Dilution with 0.14 M HNO_3/Be	Li
Serum	Dilution with 1% HNO_3/Rh	B, Cu, Fe, Li, Mg, Rb, Sr, Zn
Plasma protein	Dilution with water/In	Al, Cr, Cu, Fe, Mg, Mn, Ni, Se, Zn
Urine	Dilution with 1% HNO_3/Rh	Ni
Urine	Digestion with HNO_3-HClO_4/In	As,[b] Cd, Co, Hg, Pb, Tl
Bovine liver	Digestion with HNO_3-HClO_4/In	Co, Cu, Fe, Mn, Rb, Sr, Zn
Bovine liver	Microwave digestion with HNO_3	Ca, Cd, Cu, Fe, Mg, Mn, Mo, Pb, Rb, Sr, Zn
Tissues	Microwave with $HNO_3-H_2O_2-HClO_4$/Re	Ba, Bi, Th, Tl, U
Tissues	Microwave with HNO_3/In	Al, Cd, Co, Cu, Fe, Mg, Mn, Mo, Rb, Sr, Zn

[a] Adapted from Iyengar et al.[1]
[b] Arsenic was determined after precipitation of chloride with AgI.

resin column for the reliable measurement of Bi, Co, Cr, Cu, Hg, Mo, Pt and V.[1] The complexation and adsorption separated the trace elements from the easily ionized Na, K and Ca so that matrix effects were insignificant. They could not, however, preconcentrate and separate some of the important metals such as Cd, Pb and Zn under these conditions. Minor changes in the experimental conditions or the complexing agent itself might permit the preconcentration and separation of these metals. Buckley et al.[24] separated and preconcentrated Se as the hydride prior to its determination in several biological materials. Table 4 summarizes selected digestion procedures used in the ICPMS analysis of biological fluids and tissues.[18]

In addition to chemical means, the separation of analytes from the matrix has been accomplished instrumentally, e.g. by ETV. Sample introduction by means of ETV significantly reduces polyatomic interference since the samples are introduced in the absence of any accompanying solvent. An example is the determination of Fe. All the Fe isotopes, $^{54-57}Fe$, suffer severe polyatomic interferences from $^{40}Ar^{14}N$, $^{40}Ar^{16}O$ and $^{40}Ar^{16}OH$. These interferences were easily overcome in the determination of Fe in serum by ETV/ICPMS.[1] Branch et al.[25] successfully overcame the chloride interference ($^{40}Ar^{35}Cl^+$, $ArCl^+$ and ClO^+ species) in the determination of As in urine by diluting it 10-fold with 2% HNO_3 and by adding as little as 0.032 L min^{-1} (2%) nitrogen to the Ar nebulizer gas.

5.4 Solid Sampling

There has been considerable interest in recent years in reducing sample preparation times by the use of solid sampling techniques. These approaches are especially useful when the sample is difficult, hazardous or tedious to digest and when contamination errors may arise during conventional dry and wet ashing procedures. One of the most promising approaches to solid sampling is via nebulization of aqueous suspensions of fine powders, termed slurry atomization. The use of a clog-free, high-solids Babington-type nebulizer combined with the high temperature and long residence time of the ICP make it a suitable technique for the slurry analysis of solid samples in which the analytes are present at trace levels. The most critical parameter in slurry atomization is particle size. In general, particles with a diameter larger than 8 μm will not pass through the sample introduction system, reach the plasma and be atomized. The particle size can be reduced to <5 μm by wet grinding 0.25–1-g samples with ZrO beads in a PE bottle on a laboratory shaker for 1–4 h (depending on the sample), and this can result in slurries giving an atomization efficiency equivalent to that of an aqueous solution of the same analyte concentration. Dried milk powder is naturally suited to slurry analysis, as the product is spray-dried during production and has a particle size of <5 μm. Milk powder samples (0.25–1 g) were suspended in a 1% Triton X-100 solution to which Mg was added as an internal standard. The slurried suspension was continuously stirred and introduced via a peristaltic pump and a high-solids nebulizer. Excellent results were obtained for Al and Pb.[1] In the case of Pb, the stable isotope ratio was also measured in the slurry using isotope dilution analysis ICPMS technique. Disadvantages with slurry atomization involve sample contamination and sample heterogeneity.

Other solid sample introduction techniques proposed for ICPMS include ETV and laser ablation (LA). The use of LA in ICPMS has been reviewed.[26] Briefly, a highly intense pulsed Nd:YAG laser beam is focused on to the solid sample placed in an enclosed sampling cell. The resulting sample vapor is swept into the plasma using argon. Laser sampling can be used for semiquantitative screening and for obtaining isotopic information. Deposition of vapor on the transfer tube and sampling cell can lead to memory effects.

Application of this technique to biological materials is yet to appear in the literature. The development of procedures for calibration, the availability of appropriate

(corresponding to the high sensitivity of the LA/ICPMS technique), well characterized reference materials and a fundamental understanding of the processes involved in the laser sampling step will undoubtedly promote application of LA/ICPMS to the analysis of biological samples.

5.5 Speciation

Preparation procedures for trace element speciation in biological materials generally involve 'digestion' of the organic matrix without changing the species followed by separation and preconcentration of the species prior to determination. The 'digestion' method usually involves extraction with organic solvents, and the separation/preconcentration methods are invariably based on chromatographic techniques. The potential of these methods when coupled to ICPMS as a detection system has been highlighted.[23,27] At present, the most commonly used chromatographic technique is high-performance liquid chromatography (HPLC), possibly because of the simplicity of the interface, the ability to vary the stationary and mobile phases easily to enhance separation and the feasibility of performing the analysis in solution at ambient temperature. Two forms of HPLC have been used, namely ion pairing (both anion and cation pairing) and ion exchange (commonly called ion chromatography). Byrdy and Caruso[27] have outlined some of the problems associated with the HPLC method for ICPMS detection, and have suggested possible approaches to minimize these problems. The major problems associated with sample preparation include (i) clogging of the torch tip and orifices when the total dissolved salt concentration in the mobile phase is $>0.2\%$, which can be minimized by decreasing the buffer concentration, using mixed gas plasmas and using a nitric acid wash between runs, and (ii) high concentrations of organics in the mobile phase may cause carbon deposition on the sampler and skimmer, which will cause plasma instability. This can be alleviated using mixed gas plasmas and cooling the spray chamber.

A few typical examples reviewed by Iyengar et al.[1] are given below. The arsenic species from dogfish muscle were extracted using a methanol–chloroform method, and were then separated into As(III), As(V), monomethylarsenic acid (MMA), dimethylarsenic acid (DMA) and arsenobetaine by ion pairing and ion-exchange HPLC, and were determined on-line by ICPMS. Similar approaches have been used for the speciation of arsenic in urine. Bushee determined methylmercury in albacore tuna by ion chromatography/ICPMS while Beauchemin et al. determined organomercury in dogfish and lobster reference materials after extraction with toluene as chloride and back-extraction into an aqueous medium of cysteine acetate. Tributyl- and dibutyltin have been separated in the lobster hepatopancreas reference material (LUTS-1, National Research Council of Canada) using either a strong cation-exchange resin column or reversed-phase HPLC. Other examples include the detection of gold drug metabolites in human blood serum using HPLC/ICPMS, and cadmium speciation in pig kidney using enzyme digestion and size-exclusion chromatography coupled directly to ICPMS. Another attractive approach which is yet to be explored is supercritical fluid chromatography (SFC) coupled to ICPMS. Use of SFC would permit the analysis of thermally labile, nonvolatile, high-molecular-weight compounds.[27]

5.6 Other Studies

Other studies with ICPMS include isotope ratio measurements in exposure assessment studies and isotope tracer analysis in biokinetics investigations, where the use of radiotracers is undesirable or unavailable.[1] Delves and Campbell used the dilution method described earlier to determine the $^{208}Pb/^{206}Pb$ isotope ratio in blood in the presence of Pt ($^{194}Pt/^{195}Pt$ ratio) as an internal standard to correct for drift. Janghorbani et al. have extensively dealt with the sample size, sample preparation and separation aspects in connection with the ICPMS measurement of stable isotope ratios in mineral metabolism. The samples include serum, whole blood, urine and stool. Serfass et al. determined the zinc isotope ratio by lyophilizing human plasma and feces samples, dry ashing them at at 480 °C, dissolving the ash in 6 M HCl and adjusting the pH to 2.5–3.0. Other applications include studies on Se pools, dietary Cu and Zn absorption, Fe incorporation into erythrocytes and clearance of ingested Ni.[23]

6 ANODIC STRIPPING VOLTAMMETRY

6.1 Introduction

Wang,[28] and Ostapczuk and Froning[29] have reviewed the application of voltammetry to biological analysis including sample preparation. Among electrochemical techniques, ASV, including differential-pulse anodic stripping voltammetry (DPASV), has been the method of choice for the analysis of biological specimens because of its low cost and excellent sensitivity for the frequently determined metals Cd, Cu, Pb and Zn. Therefore, this section will focus on sample preparation methods used in ASV analysis of biological materials.

The reliability of the data generated in the analysis of biological samples using ASV is more dependent on

how well the sample has been decomposed than with atomic spectroscopic techniques. In general, the metal ions to be determined should be released from their binding in the organic matrix and should be presented in the solution state for voltammetric analysis. Incomplete sample destruction results in two major types of matrix interferences. In one type, the matrix concomitants other than the analyte can be electrolytically reduced or oxidized in the applied voltage range giving rise to electrode reactions and hence high background currents (also called faradaic residual current), which can cause problems in the evaluation of the analyte signal.[30] The second type of interference comes from surface-active matrix components which adsorb on the working electrode, inhibit the electrode process (i.e. hinder access of the metal ions to the working electrode) and distort and suppress the voltammetric signal of the metals. Furthermore, the catalysis of hydrogen evolution by the adsorbed organic matter containing nitrogen tends to mask the Cd peak and renders its determination impossible. Needless to say, the limiting factor in ASV analysis of biological materials is the sample preparation step because it is not easy to devise a sample treatment method that would selectively release all the bound trace elements from the sample matrix. In general, solid sampling and direct analysis involving minimal sample pretreatment are out of the question. Some typical examples are presented below.

6.2 Direct Methods

One of the simplest and fastest preparation methods involves the use of Metexchange reagent for the ASV determination of trace elements. The reagent, which is composed of 1.07% $CrCl_3 \cdot 6H_2O$, 1.43% calcium acetate and 0.0028% Hg(II), rapidly decomplexes the trace metals contained in the sample via an ion-exchange substitution reaction. The reagent has been extensively used in the ASV determination of Pb in whole blood, urine and milk. The main limitations of the method are as follows: (i) the large dilution factor of 30 (0.1 mL blood + 2.9 mL reagent) which prevents reliable measurement of Pb levels below $100\,\mu g\,L^{-1}$; (ii) its high cost; (iii) incomplete removal of organic matter, which usually fouls the electrode by adsorption; and (iv) its ineffectiveness in the presence of competing ligands such as EDTA, which is used as an anticoagulant. Hence the use of Metexchange requires heparin-stabilized blood. Other simple approaches involve the use of low-energy ultraviolet (UV) or high-energy γ-irradiation to decompose the organic matrix and release the bound metals of blood samples. Direct determination of Pb in urine after simple dilution with water leads to erroneous results owing to adsorption of the urine matrix constituents on the electrode surface, complexation of Pb by the surface-active ligands and water-soluble proteins frequently found in urine and the slow oxidation of the mercury film on the electrode by the urine matrix. Bond and Reust[1] directly determined Pb in acidified (0.01 M HCl) urine by adsorptive removal of the organic matter using a Sep-Pak C_{18} cartridge, an apolar chromatographic medium, followed by filtration through a 0.45-μm filter. The filtration has also been performed in situ at the electrode surface using a permselective polymeric coating such as cellulose acetate or Nafion. Acid hydrolysis of the urine sample at pH < 0.5 followed by filtration of the solution over a small column of macroreticular resin has also been attempted.[29] Stauber and Florence collected sweat on a 0.45-μm Millipore membrane filter, leached the filter with a supporting electrolyte of $NaNO_3$–HNO_3 and determined Cd, Cu, Pb and Zn in the leachate by ASV. Also attractive is the use of fused silica to remove interferences by urinary organics.

6.3 Acid Decomposition

As in the case of spectroscopic techniques, acid digestion is the preparation method of choice in ASV.[1] The digestion method chosen should completely destroy the biological matrix and release the metals to be determined; it should not result in loss or contamination of the metal; the digestate (the residue remaining after mineralization) should be dissolved completely in the diluent used; and any excess of the reagents remaining from the digestion should not interfere with the measurement step.[30] Wet digestion procedures with nitric–perchloric–sulfuric acids or digestion with HNO_3 and subsequent $HClO_4$ treatment have proved to be the most effective for the ASV determination of Cd, Cu, Pb, Tl and Zn in blood, urine, fish and crustacean samples. The same digestion procedure was also effective for As and Sb in urine and blood. In the case of As, the ashed materials were subjected to a reductillation procedure (reduction followed by distillation) to reduce As(V) to As(III) and to separate arsenic from the sample matrix.

Tooth samples have been decomposed with HNO_3–$HClO_4$ and the residue was taken up in 0.25% HNO_3 for the measurement of Cd, Cu, Pb and Zn. Some studies have used only nitric acid–sulfuric acid, and avoided the use of perchloric acid because of the risk of explosion when organic matter is digested with this acid. However, repeated additions of nitric acid were required to decompose the organic matter completely. Urine, liver and muscle samples were digested in this way for the determination of Cd, Cu, Pb, and Zn. Co, Ni and Se were also measured in the same digestate, but by using

different voltammetric techniques as outlined later. As in the case of the Metexchange reagent, digestion methods also require high dilutions (up to 50-fold), which do not permit the measurement of metals (e.g. Cd) present at micrograms per liter levels in blood. Pressure digestion with nitric acid alone in a bomb or HPA in quartz vessels at around 300 °C for 3–4 h was found to be sufficient for the ASV analysis of hair, teeth, liver, mussels, herring gull eggs, earthworm, feces and urine for Bi, Cd, Pb, Tl and Zn. Ultrapure acids should be used in order to avoid the high blanks resulting from the attack of the glass and quartz laboratory ware during the digestion process.

6.4 Dry Ashing

Adeloju et al. evaluated three dry ashing procedures for the simultaneous ASV determination of Cd and Pb in bovine liver, and veal tissue samples.[1] The methods considered were (i) direct, (ii) with nitric acid as ashing aid and (iii) with sulfuric acid ashing aid. Methods (i) and (ii) gave less than 90% recovery. Cd and Pb were quantitatively recovered only by method (iii) because the conversion of the metals to their sulfate might have prevented the risk of loss by volatilization. Also, it is advisable to use HCl for the dissolution of the ash since nitric acid undergoes a redox reaction with the mercury drop to give nitrite ion, which interferes with the electrode process. Finally, the dry ashing of biological materials usually gives rise to a carbonaceous residue even when ashing aids are used. This residue will not be dissolved by the HCl used in dissolving the ash. The residue can create problems with ASV analysis. Therefore, it is important to dilute the the sample sufficiently after dissolution of the ash to allow adequate separation of the carbon from the solution by settling. Valenta et al. found LTA in a microwave-excited oxygen plasma to be very effective for the complete destruction of organic matter. They used LTA method for the ASV determination of Cd, Cu and Pb in whole blood and fish.[1]

6.5 Other Voltammetric Techniques

Other voltammetric methods used include cathodic stripping voltammetry for Se in $HNO_3-H_2SO_4-HClO_4$-digested whole blood and serum, adsorptive stripping voltammetry for Ni in human nails, Co and Ni in blood and Pt in human whole blood and potentiometric stripping voltammetry (PSV) for Hg in urine and Cd and Pb in whole blood and urine.[1,2] In all cases, the samples were wet ashed in a bomb or HPA, except in the case of Pt where the sample was dry ashed followed by the adsorptive–catalytic reduction of the Pt–formazone complex.

7 X-RAY SPECTROMETRY

7.1 Introduction

Energy dispersive X-ray fluorescence (EDXRF) and particle-induced X-ray emission (PIXE) are the most common X-ray techniques used for elemental screening of biological materials. Both EDXRF and PIXE can usually detect elements at the micrograms per gram (parts per million, ppm) level or a few tenths of 1 ppm. Samples are prepared for EDXRF and PIXE analysis either as thin or as thick targets. A thin film is defined as one in which the sample mass/unit area is $<0.1\,g\,cm^{-2}$ and a thick target is one in which the mass of material is sufficient to stop the the projectile beam.[31] For example, the range of a 2.5-MeV proton in a biological matrix is 100 μm, corresponding to $\sim 10\,mg\,cm^{-2}$ for material of unit density. Although a thick target presents the maximum amount of sample to the beam, thereby improving sensitivity, calibration is more difficult than with thin samples because of matrix effects. Therefore, thin targets are invariably preferred.

In general, PIXE is two to five times more sensitive than EDXRF and is more suited to the analysis of small sample sizes. However, direct ultratrace analysis is not feasible with either technique because of poor sensitivity, and a preconcentration step is essential to bring the trace metals in biological specimens to detectable levels. Preconcentration is generally effected by air drying, ashing (oven or LTA) or freeze-drying. Since the preconcentrated material is not self-supporting, especially with thin targets, the preconcentration is usually effected on suitable 'backings'. In general, the backing must consist of low-Z material, should provide minimal contamination and its thickness should be ≤ 4 μm to minimize the X-ray background and thus to improve sensitivity. The mechanical strength of the backing should be such as to withstand irradiation, handling and sample preparation procedures. Many materials have been used as backings, including thin foils of carbon, collodion, Formvar, Kapton, Mylar, different filters (Millipore, Nuclepore) and others.[1] In addition, the sample is often doped with internal standards such as Ag, Sc, Ti, V, Y or a combination of them to facilitate calibration and to minimize errors due to instrumental fluctuations, sample preparation variations and geometry variations in sample density. The simultaneous use of two or even three internal standards provides a valuable control of the analysis, and, if the internal standards differ significantly in atomic number, such as Sc and Y, assures accuracy of the matrix correction applied. The endogenous concentration of the selected internal standards should be low in the samples. Räisänen et al. have summarized some of the problems encountered in the preparation of thin-film targets.[32]

7.2 Biological Fluids

The sample preparation procedures used for the EDXRF multielement (usually Br, Ca, Cu, Fe, K, Rb, Se and Zn) analysis of plasma and serum have been reviewed.[1,33] Preparation information on samples other than serum and plasma can be found in a review by Williams.[31] Sample preparation in the case of blood usually involves addition of 50 μL of a 1000 mg L^{-1} solution of at least two internal standards to 1 mL of serum, plasma or whole blood, vortexing, pipetting 250 μL of the vortexed mixture on to the center of a 4-μm-thick film of Formvar (caution should be exercized in using Formvar as occasional charging problems have been observed), Mylar or polypropylene suspended on a Teflon® ring that fits into the sample changer of the XRF unit, and drying overnight in a laminar-flow clean-air hood at ambient temperature (air drying) or by placing the sample in an evaporation chamber under vacuum for 1 h. The target prepared in this way ranges in thickness from 5 to 10 mg cm^{-2}, and provides an elemental concentration enhancement factor of 11.3 on average. The concentration factor can be further enhanced by ashing the sample in an oven at 400 °C or in a current of microwave-induced oxygen plasma (LTA), but loss of volatile elements such as Br and Se is a distinct possibility under these conditions. In the freeze-drying approach, the serum or serum albumin sample is lyophilized for 72 h, the resulting powder is homogenized, 0.1 g of the powder is doped with an aqueous solution of internal standard (e.g. 1 mL of 200 mg L^{-1} Ag) and 10 μL of the doped solution are pipetted on to a Mylar film mounted on a PE target frame and vacuum dried as above. In some cases, the dried material has been pelletized for the elimination of errors resulting from sample heterogenity. Some workers recommend a 50% dilution of the sample to minimize background continuum. For example, Rastgar et al. and Maier et al. diluted samples of serum and lavage fluid (collected from the human bronchial alveoli) 50% with an aqueous solution containing two internal standards (Y and V) and air-dried a 20–50-μL aliquot as above for EDXRF analysis, and Sargentini-Mayer et al. determined Pt in plasma and filtered plasma by diluting the sample 50% with 0.5% poly(ethylene glycol) to ensure a better mechanical cohesion of the dried sample film.[1]

Selective enrichment procedures have been used for single-element determinations. For example, the pretreatment steps used for the measurement of Se in serum and whole blood and in IAEA fish soluble certified reference material (CRM) A-6/1975 involved wet digestion followed by coprecipitation with Te, which also served as an internal standard.[1,33] Pd, Ge and Cr have also served as internal standards. For the EDXRF determination of Cu, Pb and Zn in urine, Agarwal et al. prepared a composite solution of 4 mL of urine, 0.5 mL of 3.4 mg L^{-1} yttrium (internal standard) and 0.5 mL of a suitable standard solution of the elements of interest, preconcentrated the elements by adding the composite to 0.5 g of Chelex-100 resin (sodium form) and mixing for nearly 2 h on a rotating mixer at pH 7, filtered off the solution, dried the resin and placed a portion in a sample holder for analysis.[1]

Räisänen et al.[32] found that when the sample size is small (<30 μL), as in the case of CSF, the usual methods of preparing thin-film targets were not successful for PIXE analysis. They prepared thin, self-supporting CSF films by pipetting 20–40 μL (2–4 drops) of sample into 2-mm holes made in a 125-μm thick Kapton foil and drying at room temperature after the addition of each drop. The holes were made with an Mo spike, as Mo was not measured and did not produce any overlapping peaks. Yttrium (14 mg L^{-1}) was added to the sample as an internal standard and EDTA (5 μg L^{-1} of sample) was added to carry and homogenize the ionic components. Reliable data were obtained for Br, Ca, Cu, Fe, K and Zn.

7.3 Tissues

The techniques and problems associated with the preparation of thin biological tissues for XRF and PIXE analysis have been reviewed.[1] Preparation methods used include sliced tissue sections supported on ultrathin polystyrene films, or between Formvar foils, evaporation of homogenates on to suitable backings as in the case of biological fluids, direct irradiation of ashed specimens and self-supporting tissue sections. Elements frequently determined in tissues are the same as in the case of biofluids. Wallin et al. determined Co, Cr, Cu, Fe, Mn, Ni, Pb and Zn in proteins associated with tubulin and microtubules prepared from bovine brain by lyophilizing the sample containing added internal standards of Y and Sc, pressing 8–25 mg of the powder into pellets and mounting them on platic foils for EDXRF analysis.[1]

Maenhaut et al. prepared human autopsy kidney samples and bovine liver [National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) 1577] for PIXE analysis as follows. The samples were freeze-dried, pulverized and homogenized. A 0.1–0.2-g amount of the homogenate was then doped with an aqueous solution of Ag as an internal standard and prepared for PIXE analysis on a Mylar backing as described for serum above. The concentration factor and hence the sensitivity can be enhanced by dry ashing the homogenate prior to doping; the same group prepared bovine liver (NIST SRM 1577a), animal muscle (BCR CRM 63) and milk powder (IAEA CRM H4) by

microwave digestion of the sample (0.1–0.2 g) with nitric acid (1 mL/0.1 g) containing 9 mg L^{-1} Y as internal standard, pipetting 10 μL of the digestate on to a 1.5-μm-thick polycarbonate film mounted on a 25-mm diameter PE target frame and drying the target in a vacuum desiccator.[1] Gunther et al.,[34] on the other hand, digested the bovine liver sample with nitric acid using a hot-plate, dissolved the residue in 5% HNO_3, spiked a 450-μL aliquot with 50 μL of 25 mg L^{-1} Ga as internal standard, pipetted 20 μL of this mixture on to a quartz-glass carrier, evaporated to dryness using an infrared (IR) lamp and analyzed the residue by total reflection XRF. The digestion procedure, however, resulted in loss of Br with recoveries of only 8–26%. For the EDXRF analysis of Pb in human teeth, 100–200-μg samples were dissolved in 50 μL of 50% HNO_3 containing 5 mg L^{-1} Y and 5 μL of this solution were deposited on a 4-μm thick polypropylene film and dried at room temperature.

Bloch et al. measured the Pb level of children's teeth in situ by EDXRF with a pencil beam of γ-rays from a ^{57}Co source. PIXE analysis of deciduous molar enamel for Pb was carried out by powdering the dried teeth in a mechanical mill, separating the enamel from dentine by a flotation method and compressing the enamel powder into pellets 5 mm in diameter and 0.5–2 mm thick. The crown, root and jaw bone located around and underneath each tooth of a 97-year-old female cadaver were similarly analyzed for Br, Cu, Fe, Mo, Pb, Sr and Zn by powdering and pelletizing each sample and attaching the pellets either with 20 μL of a 15% ethanolic solution of polyvinylpyrrolidone or Krazy glue to Kapon films.[1] Williams has summarized the application of PIXE to the trace element analysis of hair.[31]

Renan et al. preferred the use of 'thick' liver samples for PIXE analysis.[1] The use of 'thick' samples simplifies preparation techniques, minimizes contamination, enhances sensitivity, reduces target heating and minimizes selective volatilization of labile analytes such as Br. However, literature studies with thick samples are not extensive, possibly because of matrix effects. They also prepared murine liver sample for PIXE analysis by homogenizing, lyophilizing at −45 °C for 8 h, powdering, compressing the powder into a disk (13 mm diameter, 2 mm thickness) and mounting it on a Perspex holder for irradiation.

8 NEUTRON ACTIVATION ANALYSIS

According to Heydorn,[35] NAA is becoming less and less popular because the geographical separation between an NAA laboratory (with its associated nuclear reactor) and a hospital prevents the routine use of this technique in a clinical setting. Apart from this problem of inaccessibility, the instrumentation is very expensive and requires highly skilled professionals to carry out the analysis. Also, its sensitivity is poor for some of the key elements such as Pb. Hence NAA cannot be considered as a routine technique. It may serve as a reference method because of its insensitivity to contamination by stable elements after irradiation, minimal handling required prior to irradiation and the absence of a blank value. However, it is important to ensure that no contamination of the sample occurs before irradiation. Reference laboratories such as NIST are well versed in NAA sample preparation methods, and the IAEA has dealt with the practical aspects of NAA in a technical document.[36] A brief outline mostly based on this document is given below.

The high sensitivity of NAA for many elements and the necessity to use optimally the space available within the irradiation container call for the use of small sample sizes in the range 50–500 mg. Clearly, sample homogeneity is critical. Samples in dry form are preferred for irradiation in order to avoid potential radiolysis problems and the attendant precautions. However, liquids and solids containing moisture can be irradiated when irradiations last only a few seconds and where irradiation ports and cooling facilities are available.

In NAA, samples have to be irradiated in a container. The most commonly used container materials are PE (both conventional PE and linear PE) and quartz. When the flux needs to be modified, the container can be lined with graphite. PE is suitable for most purposes. It is inexpensive and available in high purity, can be fabricated in various forms and sizes, can be efficiently cleaned with HNO_3 and can be easily sealed. For most applications, it is sufficient to use a snap-on lid. Capsules cut from PE microcentrifuge tubes will hold 100 mg of sample powder and can be conveniently bundled into a rabbit. PE screw-capped bottles are useful for large samples or for containing numerous samples.

The major disadvantage of PE is that volatile elements such as Hg and halogens diffuse through it during long irradiations, and in these situations samples have to be irradiated in quartz containers. High-purity quartz should be used since mineral quartz is activated strongly, and it is difficult to transfer biological materials to a pure capsule for counting. Graphite containers are preferred when a pneumatic transfer system (e.g. cyclic determinations) is used for irradiation in a very high neutron flux. Graphite is obtainable in high purity, but is very expensive. Therefore, it is used only in special cases.

All major elements in the human body except P can be directly determined in vivo using prompt γ neutron activation analysis (PGNAA). This information allows the calculation of the body composition with

the exception of trace elements and P, and for this purpose PGNAA has been accepted as a reference method. In in vitro biological trace element analysis, direct determination [referred to as instrumental neutron activation analysis (INAA)] is seldom possible because of interferences. The interfering activities have to be eliminated by radiochemical separation [radiochemical neutron activation analysis (RNAA)]. This involves addition of a suitable amount of inactive carrier of the analyte to the sample to control and minimize losses, irradiation, decomposition of the irradiated sample using nitric, perchloric or sulfuric acid, and radiochemical separation. Complete isolation of the analyte is not required, and several group separation schemes have been developed. Particularly simple separations may be effected by adsorption on activated carbon or by elimination of only ^{24}Na from the sample.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
AAS	Atomic Absorption Spectrometry
APDC	Ammonium Pyrrolidine-dithiocarbamate
ASV	Anodic Stripping Voltammetry
BCR	Bureau Communautaire de Réference
CRM	Certified Reference Material
CSF	Cerebrospinal Fluid
DDTC	Diethyldithiocarbamate
DMA	Dimethylarsenic Acid
DPASV	Differential-pulse Anodic Stripping Voltammetry
EDTA	Ethylenediaminetetraacetic Acid
EDXRF	Energy Dispersive X-ray Fluorescence
ETV	Electrothermal Vaporization
FAAS	Flame Atomic Absorption Spectrometry
FIA	Flow Injection Analysis
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HMAHMDTC	Hexamethyleneammonium Hexamethylenedithiocarbamate
HPA	High-pressure Asher
HPLC	High-performance Liquid Chromatography
IAEA	International Atomic Energy Agency
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
INAA	Instrumental Neutron Activation Analysis
IR	Infrared
LA	Laser Ablation
LTA	Low-temperature Ashing
MIBK	Methyl Isobutyl Ketone
MMA	Monomethylarsenic Acid
NAA	Neutron Activation Analysis
NaDDC	Sodium Diethyldithiocarbamate
NIST	National Institute of Standards and Technology
PE	Polyethylene
PGNAA	Prompt γ Neutron Activation Analysis
PIXE	Particle-induced X-ray Emission
PSV	Potentiometric Stripping Voltammetry
PTFE	Polytetrafluoroethylene
RNAA	Radiochemical Neutron Activation Analysis
SFC	Supercritical Fluid Chromatography
SRM	Standard Reference Material
TCA	Trichloroacetic Acid
TIBC	Total Iron Binding Capacity
TMAH	Tetramethylammonium Hydroxide
UV	Ultraviolet
XRF	X-ray Fluorescence

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Electroanalysis and Biosensors in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Neutron Activation in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Proton-induced X-ray Emission

in Environmental Analysis • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications • X-ray Fluorescence in Forensic Science

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure • Spectroscopic Techniques in Industrial Hygiene

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis • Sampling and Sample Preparation in Process Analysis

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Pulse Voltammetry

Liquid Chromatography **(Volume 13)**
Supercritical Fluid Chromatography

Nuclear Methods **(Volume 14)**
Instrumental Neutron Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Sample Preparation for X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. G.V. Iyengar, K.S. Subramanian, J.R.W. Woittiez, *Element Analysis of Biological Samples*, CRC Press, Boca Raton, FL, 135–173, 1998.
2. M. Rossbach, J.D. Schladot, P. Ostapczuk (eds.), *Specimen Banking: Environmental Monitoring and Modern Analytical Approaches*, Springer, Berlin, 1992.
3. H.T. Delves, G. Shepherd, P. Vinter, 'Determination of 11 Metals in Small Samples of Blood by Sequential Solvent Extraction and AAS', *Analyst*, **96**, 260–273 (1971).
4. B. Welz, 'Atomic Absorption Spectrometry', in *Handbook on Metals in Clinical and Analytical Chemistry*, eds. H.G. Seiler, A. Sigel, H. Sigel, Marcel Dekker, New York, 85–106, 1994.
5. H.A. McKenzie, L.E. Smythe (eds.), *Quantitative Trace Analysis of Biological Materials*, Elsevier, Amsterdam, 1988.
6. H.G. Seiler, A. Sigel, H. Sigel (eds.), *Handbook on Metals in Clinical and Analytical Chemistry*, Marcel Dekker, New York, 1994.
7. R.F.M. Herber, M. Stoeppler (eds.), *Trace Element Analysis in Biological Specimens*, Elsevier, Amsterdam, 1994.
8. H.T. Delves, 'The Analysis of Biological and Clinical Materials', *Prog. Anal. At. Spectrosc.*, **4**, 1–48 (1981).
9. A. Taylor, 'Clinical Applications of Flame Techniques', in *Atomic Absorption Spectrometry*, ed. J.T. Haswell, Elsevier, New York, 359–379, 1991.
10. B.E. Walker, 'Clinical Applications of Flame Techniques', in *Atomic Absorption Spectrometry*, ed. J.E. Cantle, Elsevier, New York, 321–339, 1982.
11. D.L. Tsalev, *Atomic Absorption Spectrometry in Occupational and Environmental Health Practice*, CRC Press, Boca Raton, FL, Vol. II, 1984.
12. N.J. Miller-Ihli, 'Solids Analysis by GFAAS', *Anal. Chem.*, **64**, 964A–968A (1992).
13. R.F.M. Herber, 'Solid Sampling Analysis', in *Handbook on Metals in Clinical and Analytical Chemistry*, eds. H.G. Seiler, A. Sigel, H. Sigel, Marcel Dekker, New York, 195–199, 1994.
14. W. Slavin, 'Graphite Furnace AAS', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 68–85, 1994.
15. R.F.M. Herber, 'Solid Sampling Atomic Absorption Spectrometry and Matrix Composition of Organic Reference Materials', *Pure Appl. Chem.*, **63**, 1213–1220 (1991).
16. N.J. Miller-Ihli, 'Advances in Ultrasonic Slurry Graphite Furnace Atomic Absorption Spectrometry', *Fresenius' J. Anal. Chem.*, **345**, 482–489 (1993).

17. G.A. Meyer, P.N. Keliher, 'An Overview of Analysis by Inductively Coupled Plasma–Atomic Emission Spectrometry', in *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*, 2nd edition, eds. A. Montaser, D.W. Golightly, VCH, New York, 473–516, 1992.
18. D.M. Templeton, 'Inductively Coupled Plasma–Atomic Emission Spectrometry (ICPAES) and Inductively Coupled Plasma–Mass Spectrometry (ICPMS)', in *Handbook on Metals in Clinical and Analytical Chemistry*, eds. H.G. Seiler, A. Sigel, H. Sigel, Marcel Dekker, New York, 168–178, 1994.
19. M. Thompson, R.M. Barnes, 'Analytical Performance of Inductively Coupled Plasma–Atomic Emission Spectrometry', in *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*, 2nd edition, eds. A. Montaser, D.W. Golightly, VCH, New York, 290–291, 1992.
20. A. Sanz-Medel, 'Inductively Coupled Plasma–Atomic Emission Spectrometry: Analytical Assessment of the Technique at the Beginning of the 90s', *Microchim. Acta II*, 265–275 (1993).
21. Z. Horváth, A. Lásztity, R.M. Barnes, 'Preconcentration and Separation Techniques for Inductively Coupled Plasma Atomic and Mass Spectrochemical Analyses', *Spectrochim. Acta Rev.*, **14**, 45–78 (1991).
22. P. Schramel, L.-G. Xu, G. Knapp, M. Michaelis, 'Multielement Analysis in Biological Samples by On-line Preconcentration on 8-Hydroxyquinoline–Cellulose Microcolumn Coupled to Simultaneous ICPAES', *Fresenius' J. Anal. Chem.*, **345**, 600–606 (1993).
23. H. Vanhoe, 'A Review of the Capabilities of ICPMS for Trace Element Analysis in Body Fluids and Tissues', *J. Trace Elem. Electrolytes Health Dis.*, **7**, 131–139 (1993).
24. W.T. Buckley, J.J. Budac, D.V. Godfrey, K.M. Koenig, 'Determination of Selenium by Inductively Coupled Plasma Mass Spectrometry Utilizing a New Hydride Generation Sample Introduction System', *Anal. Chem.*, **64**, 724–729 (1992).
25. S. Branch, L. Ebdon, M. Ford, M. Foulkes, P. O'Neill, 'Determination of Arsenic in Samples with High Chloride Content by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **6**, 151–154 (1991).
26. E.R. Denoyer, K.J. Fredeen, J.W. Hager, 'Laser Solid Sampling for Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **63**, 445A–457A (1991).
27. F.A. Byrdy, J.A. Caruso, 'Elemental Analysis of Environmental Samples Using Chromatography Coupled with Plasma Mass Spectrometry', *Environ. Sci. Technol.*, **28**, 528A–534A (1994).
28. J. Wang, 'Voltammetry', in *Trace Element Analysis in Biological Specimens*, eds. R.F.M. Herber, M. Stoeppler, Elsevier, Amsterdam, 140–145, 1994.
29. P. Ostapczuk, M. Froning, 'Advanced Electrochemical Techniques for the Determination of Heavy Metals in Specimen Bank Materials', in *Specimen Banking: Environmental Monitoring and Modern Analytical Approaches*, eds. M. Rossbach, J.D. Schladot, P. Ostapczuk, Springer, Berlin, 153–165, 1992.
30. H. Seiler, 'Voltammetry', in *Handbook on Metals in Clinical and Analytical Chemistry*, eds. H.G. Seiler, A. Sigel, H. Sigel, Marcel Dekker, New York, 132–133, 1994.
31. E.T. Williams, 'Application of Particle-induced X-ray Emission to Research in Biology and Medicine', *Biol. Trace Elem. Res.*, **12**, 19–31 (1987).
32. J. Räisänen, M. Hyvönen-Dabek, R. Lapatto, J.T. Dabek, 'Preparation of Thin Self-supporting Samples of Liquids for External Beam PIXE Analysis: Application to Human Cerebrospinal Fluid', *Appl. Radiat. Isot.*, **38**, 573–574 (1987).
33. H. Robberecht, H. Deelstra, R. Van Grieken, 'Determination of Selenium in Blood Components by X-ray Emission Spectrometry', *Biol. Trace Elem. Res.*, **25**, 149–185 (1990).
34. K. Gunther, A. von Bohlen, G. Paprott, R. Klockenkamper, 'Multielement Analysis of Biological Materials by Total-reflection X-ray Fluorescence and Inductively Coupled Plasma Mass Spectrometry in the Semiquant Mode', *Fresenius' J. Anal. Chem.*, **342**, 444–448 (1992).
35. K. Heydorn, 'Neutron Activation Analysis', in *Handbook of Metals in Clinical and Analytical Chemistry*, eds. H.G. Seiler, A. Sigel, H. Sigel, Marcel Dekker, New York, 184–191, 1994.
36. International Atomic Energy Agency, 'Practical Aspects of Operating a Neutron Activation Analysis Laboratory', Technical Document IAEA-TECDOC-564, International Atomic Energy Agency, Vienna, 1990.

Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges)

David Eugene Kimbrough
Castaic Lake Water Agency, Santa Clarita, USA

1 Introduction 1
2 Matrix 2
3 Physical Preparation 2
4 Acid Digestions for Elements 2
5 Solvent Extractions for Nonvolatile Organic Compounds 6
6 Conclusions and Future Trends 8
Abbreviations and Acronyms 9
Related Articles 9
References 9

The purpose of the analysis of environmental contaminants, such as pesticides, organic solvents, and toxic elements, in solids is to provide the basis for certain environmental decisions. Thus, the needs of the end user of the analytical procedures are the guide to understanding sample preparation. Two common goals for end users are regulatory compliance and data comparability, which is why it is very common to use standardized analytical procedures. There are two steps in the analysis of solids, the separation of analytes of interest from the rest of the sample and then the instrumental analysis of the analytes. The first step is the topic of this article. More than any other medium, solids are subject to severe matrix effects. The sample matrix is the portion of the sample not of interest except to the extent it affects the analysis of the portion that is of interest. Matrix effects can include reactions that prevent the separation of the analytes from the sample matrix or the inclusion of nontarget materials that interfere with the instrumental analysis. Other issues with sample matrix include heterogeneity and moisture content. A heterogeneous sample can give unrepresentative results and the moisture content can change over time and lead to variable results. These issues are often dealt with by drying and homogenization. The two most common types of analyte that require sample preparation are elements (e.g. copper and lead) and nonvolatile organic compound (NOC) (e.g. pesticides and polychlorinated biphenyls (PCBs)). Acid digestion is the most common sample preparation for elemental analysis. The acid must be aggressive enough to break the bonds between the analyte and the matrix and it must yield a form of the element soluble in the acid. NOCs are generally extracted from the sample matrix by organic solvents. The extraction solvent must be more soluble to the analyte than the sample matrix and the solvent must be able to penetrate into the sample matrix in order to solubilize the analyte. Thus the correct selection of acids and extraction solvents is the key to successful extraction.

1 INTRODUCTION

There are generally two parts to chemical analysis, sample preparation and instrumental analysis. There is one overriding consideration in selecting a sample preparation procedure: the data quality objectives of the final data user. Samples are analyzed in order to answer questions that will impact human health and the environment and different levels of accuracy and precision are needed to answer different questions.

Data users must consider the issue of data comparability. If the analytical results are to be compared to other sets of results, there is the possibility that the results may produce significant differences from those generated previously or expected. One possible explanation is that a different analytical method was used. For this reason researchers and especially regulatory agencies have traditionally used standardized (a.k.a. prescriptive or ''cook-book'') methods to eliminate this possibility. These may be used by either standard-setting organizations (e.g. American Public Health Association (APHA), American Water Works Association (AWWA), Water Environment Federation (WEF), American Society for Testing and Materials (ASTM)) or regulatory agencies (United States Environmental Protection Agency (USEPA), United States Occupational Safety and Health Administration (USOSHA)). In a great many cases the use of certain sample preparation procedures is a statutory, regulatory, or policy requirement.

However, because new analytes of environmental concern are being added constantly, and new or radically improved analytical instruments and techniques are being introduced, the ability of regulatory agencies to keep pace with these changes and incorporate the most recent procedures is severely limited. Likewise, standard-setting organizations are generally quite conservative about the introduction of new techniques and do not incorporate the latest techniques very quickly. As a result standardized methods tend to be behind the state of the art and nowhere near the cutting edge. Thus the data user must also consider the strictly scientific needs, and ask whether a given procedure can analyze the analytes of interest in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the matrix of interest over the range of concentrations of interest at the appropriate cost. For solids (soils, sludge, sediments, solid wastes, etc.) by far the most commonly used set of standardized methods is SW-846 prepared by the USEPA Office of Solid Waste for work associated with the Resource Conservation and Recovery Act (RCRA) and the Comprehensive Emergency Response Compensation, and Liability Act (CERCLA).[1] While this document was only intended to be a guide and not a required set of methods, in many situations it has become so. As a result, in many situations these methods are used for the sake of data comparability or simple conservatism.

2 MATRIX

The term ''sample matrix'' is one that is often used but rarely discussed.[2] The simplest explanation is that the sample matrix is the portion of the sample that is not of interest, it is all of the nontarget elements and compounds. It is also usually (but not always, especially with industrial wastes) the vast majority of the mass of the sample. The sample matrix often ''interferes'' with the extraction or digestion of target analytes and so special steps are often needed to remove or mitigate these interferences. Similar matrices should have similar effects on the same analysis. In many standard methods matrices are often defined such as ''soil'', ''solids'', or ''others''. It is not possible to predefine matrices in this fashion.

A matrix must actually be defined by the effect it has on an analysis. For example, in drinking water, the amount of calcium in the water has no effect on the analysis of volatile organic compounds (VOCs) so a very hard water and a very soft water are both the same matrix for VOC analysis. In contrast, calcium has a strong effect on the analysis of barium by flame atomic absorption spectroscopy (FAAS), and thus a very hard drinking water would be a different matrix from a very soft water. Thus two waters are the same matrix for one analysis and two different matrices for an other. In a similar fashion, how a sample matrix interacts with a sample preparation procedure is determined by the effect the matrix has on the final results, not on its physical state or appearance. Two tars might be physically very similar but one that is high in sulfur will be more difficult to analyze for barium than one that is low in sulfur.

There are many ways to group analytes of environmental interest. However, from a sample preparation point of view there are really only two types of analytes. Volatile analytes, those that will readily leave the solid phase of the sample and enter the gaseous phase of the sample container (generally the head space), the atmosphere, or some other equally useless space, generally have no sample preparation at all as any manipulation of the sample will result in loss of the analyte. Non-volatile analytes are those, such as metals and pesticides, which do not leave the solid phase for the air.

3 PHYSICAL PREPARATION

Solid samples can contain a wide ranges of moisture and must often be dried prior to analysis. This is done for a variety of reasons: one is so that comparisons between samples are more realistic and not skewed by differences in moisture content. Drying can also make other sample preparation procedures easier, such as milling and sieving. Finally, some analytical procedures are sensitive to moisture content and so by removing moisture the samples are more similar to the calibration standards. Some samples such as soils, sediments, and solids can be dried simply by placing them in drying ovens at 105 °C for 24 h.

Milling solid samples is often performed in order to achieve a more uniform particle size. Some analytical techniques are very sensitive to particle-size effects and milling minimizes this problem; this applies to X-ray fluorescence and neutron activation in particular. Particle-size reduction also increases the surface to volume ratio which is important for effective extraction procedures (see below). Milling also make sieving easier. Sieving is used to remove portions of solid samples that are extraneous, such as rocks, wood, and debris. It also complements milling by removing larger particles that need further milling. Milling and sieving also have the effect of homogenizing the sample, especially with samples with a heterogeneous distribution of analytes.

4 ACID DIGESTIONS FOR ELEMENTS

These procedures remove the elements of interest from the sample matrix, transfer them to liquid matrix, and then the liquid extract is analyzed by instruments and compared to other liquid standards. The most common approach is the use of hot acids to destroy chemical and physical bonds between the elements of interest and the matrix and to convert those elements into water-soluble forms, often leaving the bulk of nontarget elements, especially silicaceous materials, as solids that can be filtered out.[3,4]

In selecting an acid digestion procedure, two important factors need to be considered: is the acid (or acid mixture) aggressive enough to break the bonds between the analyte and the matrix, and will the element be soluble in the acid? For example, there are a number of acid digestion

procedures that use hot nitric acid. If copper oxide were bound to the surface of a carbonaceous particle, when exposed to hot nitric acid a significant portion of the carbon on the surface of the particle will be mineralized to CO_2, thus freeing the copper atoms. The copper will be converted to copper nitrate which is quite soluble in nitric acid and in water. In contrast, consider antimony trioxide in an oil. Nitric acid is not strong enough to break up the oil's molecular bonds. Further, the oil will float on top of the acid so that the nitric acid would only interact with the surface of the oil. Likewise, if the sample is composed of a small number of very large units, i.e. it is very ''lumpy'', then the acids can only attack the surface and will not be able to penetrate the sample matrix. It is here that milling and sieving are very important.

A wide variety of oxidants can be used: nitric acid, hydrochloric acid, sulfuric acid, perchloric acid, hydrofluoric acid, potassium permanganate, and hydrogen peroxide to name the more common. The most common techniques are designated in SW-846 as methods 3050, 3051, and 3031.[1] In addition, the USEPA has a parallel set of methods for its contract laboratory program (CLP). In this set it has used an acid digestion procedure included in the inductively coupled plasma mass spectrometer (ICPMS) procedure 6020 (CLP).[3] 3050 is in fact three different methods for non-oily samples using a hotplate as the heating source. One version is a simple nitric acid digestion which is nearly identical to the ASTM method 9.3.4[3] using 2 g of sample digested in concentrated nitric acid. A second version is a sequential digestion of 2 g of sample in nitric acid and then hydrochloric acid. In the third method 2 g of sample is digested in a 4 : 1 mixture of hydrochloric to nitric acid (similar to aqua regia), followed by a hot hydrochloric acid rinse, and then the filtered residue is digested again in hot hydrochloric acid and filtered again (all three of these filtrates are combined). Method 3051 is a simple nitric acid digestion using a microwave heating technique. Method 6020 (CLP) uses dilute acid (2 + 3) nitric acid and then dilute (1 + 4) hydrochloric acid heating on a hot plate. There are many other methods but they are not very different from these and these are the most commonly used.

In examining these procedures, it is useful to look at a limited number of elements and see how they interact. In one published study[3] four of these methods were compared on a variety of pure compounds, soils, and solid waste samples by inductively coupled plasma atomic emission spectroscopy (ICPAES). The linear range of a procedure is the range of quantities where there is a linear correlation between the actual quantity in the sample and the measured quantity. Each analytical procedure has upper and lower limits beyond which accurate and reproducible results cannot be obtained.

As can be seen on Table 1, where 1 μg (method 6020) or 2 μg (the others) of pure compounds were digested, not only do different elements have different linear ranges for different methods, the different salts of the same element can behave quite differently. Table 2 shows the results of the same methods on real-world soil and solid-waste samples. The same patterns are observed. The

Table 1 Linear range of four acid digestions

Element	Version of 3050			
	Plain HNO_3	Sequential $HCl + HNO_3$	Aqua regia	6020 CLP
Ag ($AgNO_3$)	NL	50–150	50–700	NL
Ba ($Ba(OH)_2$)	NL	50–700	50–2500	NL
Cr (elemental)	L	L	L	200–250 000
Mo (MoO_3)	50–500	50–60 000	L	200–60 000
Ni (elemental)	L	L	L	200–100 000
Pb (elemental)	L	50–200 000	50–50 000	200–10 000
Pb ($Pb(NO_3)_2$)	L	50–40 000	50–120 000	200–40 000
Pb (Pb_2O)	L	50–250 000	50–120 000	200–100 000
Pb (PbO)	L	50–60 000	50–120 000	200–20 000
Pb (PbO_2)	L	50–60 000	50–250 000	200–20 000
Sb (elemental)	NL	5000–20 000	50–50 000	3000–20 000
Sb (K(SbO)tartrate)	NL	5000–L	L	3000–L
Se (elemental)	L	L	L	50–20 000
TI (Tl_2SO_4)	3000–10 000	L	L	L
V (NH_4VO_3)	50–1000	50–250 000	L	200–250 000

All results in mg kg^{-1}. As_2O_3, $BeSO_4$, Cd, Co, $CoCl_2 \cdot 6H_2O$, CrO_3, $K_2Cr_2O_7$, Cu, $CuSO_4$, $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, $Ni(NO_3)_2 \cdot 6H_2O$, H_2SeO_3, and Zn were all analyzed and found to be linear for all four methods.
L = 50–1 000 000 mg/kg.
NL = Nonlinear.

Table 2 A comparison of four acid digestion procedures

Sample	Element	Version of 3050			
		Plain HNO_3	Sequential $HCl + HNO_3$	Aqua regia	6020 CLP
A	V	19 000	31 000	31 000	29 000
B	Ni	230	270	310	210
	V	19	28 000	32 000	28 000
C	Ba	220	340	510	<200
	Cu	930	1200	1100	1500
	Pb	880 000	340 000	110 000	140 000
	Zn	88	140	93	<200
D	As	120	280	400	<200
	Ba	97	210	310	<200
	Cr	290	300	260	<200
	Cu	670	1300	1200	1600
	Pb	430 000	280 000	140 000	150 000
	Sb	<50	1200	1300	510
	Zn	940	1200	1100	620
E	Pb	380 000	130 000	140 000	150 000
	Sb	<50	2600	5300	2400
F	As	<50	330	330	<200
	Cu	970	800	730	1300
	Pb	800 000	220 000	110 000	90 000
	Sb	690	23 000	97 000	13 000
G	Ag	<50	<50	230	<200
	Cd	1600	1600	2000	1800
	Cu	4400	5000	6600	5500
	Ni	390	290	370	390
	Pb	230 000	280 000	110 000	99 000
	Zn	1200	1400	1500	1600
H	Co	91	610	460	690
	Cr	9100	6000	9000	8000
	Cu	650	470	680	670
	Mo	3200	2300	3200	2400
	Ni	16 000	13 000	16 000	13 000
	Pb	280	190	350	450
I	Ag	290	<50	10 000	<200
	Ba	140	260	250	<200
	Cr	680	1000	500	340
	Cu	18 000	17 000	17 000	12 000
	Ni	610	600	550	350
	Pb	3800	3100	5300	2600
	Zn	4100	4100	3900	2700
J	Ag	23 000	<50	11 000	<200
	Ba	190	170	530	<200
	Cd	350	310	280	230
	Cu	8900	5100	3200	3000
	Pb	20 000	19 000	18 000	14 000
	Zn	51 000	44 000	42 000	42 000
K	Ag	<50	87	140	<200
	Ba	890	1300	1300	320
	Cr	960	750	1300	520
	Cu	2700	1900	3700	2400
	Mo	<50	81	230	<200
	Ni	490	410	760	430
	Pb	21 000	21 000	19 000	18 000
	Zn	5500	5300	4900	5300

All results in $mg\,kg^{-1}$.

Method 6020 (CLP) consistently got lower results for almost all elements and compounds in all matrices. This is no doubt because the digestion procedure is not aggressive using diluted acids in sequence and it was not effective at breaking apart the sample matrix. The plain nitric acid consistently solubilized more lead in all matrices than any of the other methods. This is because nitric acid reacts with lead to form lead nitrate with is extremely soluble in water and nitric acid. In contrast, this method solubilized none of the antimony in any of the matrices. Antimony reacts with nitric acid to form Sb_2O_5, which is very insoluble in either nitric acid or water.[5,6] Likewise silver reacts with nitric acid to form silver nitrate which is very soluble in water but has very limited solubility in nitric acid.

The two versions of 3050 that used a mixture of nitric acid and hydrochloric acid solubilized less lead than the plain nitric version because the hydrochloric acid reacts with the lead to form lead chloride which is not as soluble in water and hydrochloric acid than lead nitrate is in water and nitric acid. On the other hand while Sb_2O_5 is also produced by these two versions Sb_2O_5 reacts with hydrochloric acid to form Sb_2O_5Cl which is very soluble in hydrochloric acid. Similarly, silver reacts with the hydrochloric acid to form AgCl. The amount of both of these chlorides that can be solubilized is determined by the amount of chloride ion in solution. Thus the more hydrochloric acid is used, the more of these elements is solubilized. This can be seen in looking at the methods from least to most hydrochloric acid. The greater aggressiveness of the aqua regia procedure can be seen in it's efficiency in attacking the elemental pure compounds (Table 1) which are more resistant than the other molecular forms. When hydrochloric and nitric acids are added together a number of other reactive compounds are formed, including NOCl and Cl_2, giving additional aggressive qualities. These are by no means the only acid digestion procedures available but they are illustrative of the factors to consider in selecting an acid digestion procedure.

The aqua regia version of Method 3050 has one advantage over the other versions. It requires that the residue in the filter paper be re-digested. Table 3 shows how much analyte (total mass in milligrams) can be left in residue in the filter paper or in precipitate at the bottom of the filtrate. Barium, silver, and lead are the most conspicuous elements in the secondary filtrate (for reasons discussed above) and precipitate but other elements can be found there as well. The other elements are likely co-precipitated, becoming entrained in the precipitate of other elements.

There are a number of methods for heating the acids in these procedures. The most commonly used method is a

Table 3 Distribution of elements in aqua regia digestion

Sample	Element	Soluble primary filtrate	Insoluble primary filtrate	Soluble secondary filtrate
A	V	61 000		510
B	Ni	610		<100
	V	64 000		790
C	Ba	743		280
	Cu	1700		410
	Pb	57 000		166 000
	Zn	180		<100
D	As	800		<100
	Ba	370		260
	Cr	520		<100
	Cu	2100		390
	Pb	220 000		53 000
	Sb	2600		130
	Zn	2200		<100
E	Pb	231 000		53 000
	Sb	11 000		130
F	As	670		<100
	Cu	1450		<100
	Pb	149 000	6200	63 000
	Sb	180 000	6200	10 000
G	Ag	460		<100
	Cd	4000		<100
	Cu	12 500		690
	Ni	750		<100
	Pb	170 000		230 000
	Zn	3100		<100
H	Co	920		<100
	Cr	17 000	350	350
	Cu	1400		<100
	Mo	6000	290	140
	Ni	29 000	1300	810
	Pb	760		<100
I	Ag	12 000		8500
	Ba	400		<100
	Cr	1500		<100
	Cu	33 000		240
	Ni	1100		<100
	Pb	11 000		640
	Zn	7900		1500
J	Ag	18 000		3500
	Ba	420		<100
	Cu	7300		250
	Ph	36 000		640
	Zn	82 000		1500
K	Ag	280		<100
	Ba	1500		1000
	Cr	2500		<100
	Cu	7300		7300
	Mo	460		<100
	NI	820		<100
	Pb	38 000		820
	Zn	9600		210

All results in mg.

hotplate (SW-846 Methods 3031 and 3050 both use this) although water baths, steam baths, heating blocks, and microwave ovens are also used. Overheating can cause a variety of problems. If the acids are heated too rapidly, the reactions can occur too fast and overflow can result, particularly in samples when large quantities of organic materials are present. This is particularly true for mixed acid procedures, as the mixing of acids is usually exothermic. The heating source can impact the aggressiveness of a procedure if pressurized containers can be used. In this situation, the amount of energy available to break molecular bonds is greatly increased. There are important safety concerns that accompany the use of pressurized digestion containers that must be followed. When pressurization is not an issue then the heating source is really a matter of convenience as the chemistry and solubilities are the same for all heating sources.

5 SOLVENT EXTRACTIONS FOR NONVOLATILE ORGANIC COMPOUNDS

Samples to be analyzed for VOCs must not be manipulated in any way as this will result in the volatilization of the analytes of interest. The instrumental procedures generally analyze the sample directly without any preparation. The only precautions are that the samples should be refrigerated at 4 °C or lower.[7]

NOCs on the other hand must be extracted from the sample matrix in order to analyzed. The most common approach is to extract the analytes of interest into some liquid solvent.[8] The problem of matching extraction solvents to sample matrix and analyte is similar to that of the acid digestion procedures. The analyte must be more soluble in the extraction solvent than in the sample matrix and the solvent must be able to penetrate into the sample matrix in order to solubilize the analyte. It is very common to concentrate the analytes after extraction to increase the ability of the instrument to measure the analytes. This can be combined with cleanup procedures to remove interfering compounds or cleanup can occur separately.

It is very common to lump different sample matrices together into pre-defined groups as ''soils'' and ''solids''. However the effect of the matrix on solvent extractions depends on the polarity of the matrix, the solvent, and the analyte. Using soils as an example, they contain an inorganic portion made up of very polar oxides of silicon, aluminum, iron, etc. and an organic portion made up of decomposed plant material referred to as humic substances. Humic substances can range from very polar to very nonpolar, even in the sample soil particle. A soil can consist of pockets of very low polarity in a very polar particles or the exact opposite depending on the ratio of organic to inorganic materials and the nature of the humic substances. It would be extremely naive to assume that two such different matrices would have the same impact on the extraction of a given compound simply because they are both ''soils''. These principles are true for other sample types.

The polarity of the analyte of interest also impacts the extraction.[9] Very nonpolar compounds such as PCBs must be extracted with nonpolar solvents such as hexane, just as very polar compounds like acidic herbicides must extracted with polar solvents. However, using PCB-contaminated soils as an example, the PCBs, being nonpolar, are likely to be found in the least polar portions of the soil particles. If the soil particle is mainly polar then a very nonpolar solvent such as hexane will have a hard time penetrating the polar portions of the soil to reach the portions containing the PCBs. Combining more polar and less polar solvents can overcome this problem, such as combining hexane and propanone (acetone).

These solvents rarely extract only the analyte of interest; they often extract other organic materials, for example nonpolar humic substances in hexane. Sometimes these materials can interfere with the analytical process either by physically obstructing some part of the instrumentation or by being mistaken by a detector as an analyte.[10] Humic substances can build on the surface of an electron-capture detector (ECD) forming a plaque that causes severe drift. Humic substances passing an ECD in a gas chromatograph can absorb electrons and can be confused with halide-bearing compounds. Other substances can be extracted and interfere with other tests. For this reason, whenever possible, cleanup procedures are used.

USEPA's SW-846[1] lists a variety of methods for the extraction of different organic compounds from solids of various sorts. As with the acid digestions, there may be legal requirements for the use of these methods or they may be used simply for data comparability or to remain consistent with previous practice. It is not possible to review all possible sample extraction procedures or even those in SW-846 but it is helpful to examine a particular method and see all of the impacts of the above-mentioned variables.

For the analysis of PCBs in solids there are two accepted methods. Method 3540 uses 10 g of sample mixed with 10 g of anhydrous sodium sulfate. The mixture is placed in an extraction thimble (or glasswool) and the thimble is placed in a Soxhlet extractor. A round-bottomed flask, containing 300 mL of the extraction solvents and boiling chips, is attached to the Soxhlet extractor. The flask is heated just to boiling for 16–24 h. The solvent vapor rises from the flask into the condenser. The condensed solvent falls from the condenser into the extraction thimble. The sample and thimble are eventually soaked in the solvent. When a fixed volume of solvent has been collected in the

extractor, it is siphoned back into the boiling flask and the process begins again.

Method 3550 uses 30 g of sample placed in a 400 mL closed beaker. A 100 mL volume of extraction solvent is added and the mixture is sonicated for 3 min with the output control knob set at 10, with the mode switch on ''Pulse'' and the percent duty cycle set at 50%. After the sonication is complete, the solvent is either filtered or decanted. This step is repeated two or more times, after which the three extracts are combined. The solvent may be heated or left at ambient temperature. The recommended extraction solvents for both methods are dichloromethane (methylene chloride), 1 + 1 hexane–propanone (acetone) and 9 + 1 toluene–methanol, although this last mixture is rarely used. In practice, four other extraction solvent systems have been used, 1 + 1 dichloromethane–propanone, 1 + 1 dichloromethane–hexane, straight hexane or trimethylpentane (isooctane). In order to use an ECD, all of the dichloromethane must be removed as a solvent and exchanged with a more suitable solvent, usually hexane. The extract is concentrated to about 10 mL using a Kuderna–Danish concentrator then 50 mL of hexane are added. The solution is heated until all of the dichloromethane is boiled off and the remaining hexane is brought to volume with more hexane.

There is only one approved cleanup procedure for PCB in hexane extract. However many other USEPA approved methods for other analytes are often used for PCBs.

- Method 3610: This method is intended for the removal of interfering phthalate esters and nitrosoamines. The exchanged hexane extract is concentrated to 2 mL and added to the top of a preconditioned alumina column. The PCBs are then eluted from the column.
- Method 3620: The exchanged hexane is concentrated to 2 mL and added to a Florisil column. The PCBs are eluted from the column using 6% diethyl ether in hexane. The solvents are reconcentrated and exchanged back to hexane.
- Method 3630: This method is intended for polycyclic aromatic hydrocarbons (PAHs) or derivatized phenols. The extract is exchanged to cyclohexane and placed on the top of a column preconditioned with pentane. PAHs are eluted using dichloromethane–pentane (2 + 3) and then reconcentrated and exchanged to hexane.
- Method 3640: The extract must be dissolved in dichloromethane. It is passed through a gel permeation column and collected in a flask, reconcentrated and exchanged to hexane. The retention time for PCBs is determined by using methoxychlor as a surrogate. It elutes after the phthalates and before perylene and elemental sulfur. There is a 50% loss in the cleanup process for which the analyst is expected to make a correction.
- Method 3665: A 1–2-mL volume of the exchanged hexane is mixed with 5 mL of 1 + 1 sulfuric acid in a tightly sealed container. The phases are allowed to separate and the hexane is transferred into a clean 10 mL vial. The sulfuric acid is washed again with hexane and collected in the same 10 mL vial. Finally, the hexane is concentrated back to its original volume.

Table 4 shows the results of the interlaboratory study where 129 accredited hazardous materials laboratories analyzed five soils spiked with Aroclor 1260 (11–14). As can be seen, the Soxhlet extraction procedure is more efficient at extracting PCBs form the test soils. This is because the solvent was heated, the solvent that is presented to the sample is never saturated, and the extraction period is a great deal longer. The differences are most pronounced at higher concentrations where the extraction solvent has the greatest challenge. Table 5 shows how different solvents performed in the same study. The two nonpolar solvents (hexane and trimethylpentane) extracted significantly less PCBs than the more polar solvents. Given the high solubility of PCBs in hexane it must be concluded that these solvents were not able to penetrate the more polar soil matrix.

Table 6 presents the results for different cleanup procedures. There is little difference between the results from the laboratories which did not clean up, those that used the Florisil cleanup procedure, and those that use the sulfuric acid for the three higher concentration samples. However, for the sample spiked with 0.1 mg kg^{-1} of PCB there is a very significant difference. Humic materials are extracted from the soils and are creating a positive interference of little more than 0.1 mg kg^{-1}. At the higher

Table 4 Comparison of extraction equipment

	F	G	H	I
Spiked	100.00	10.00	1.00	0.100
	Overall results ($n = 129$)			
Mean	69.90	7.01	0.76	0.168
SD	32.10	3.05	0.36	0.236
<RL	0	1	3	34
	Soxhlet procedure ($n = 36$)			
Mean	81.90	7.90	0.87	0.241
SD	36.30	3.01	0.37	0.241
<RL	0	0	0	7
	Sonication procedure ($n = 91$)			
Mean	65.20	6.66	0.72	0.139
SD	30.80	3.03	0.35	0.231
<RL	0	1	3	26

All results are in mg kg^{-1}.
<RL = Less than reporting limit.
SD = Standard deviation.

Table 5 Comparison of extraction solvents

Sample	Spiked	Mean	SD	No.	<RL
		All solvents			
F	100	69.90	32.90	129	0
G	10	7.01	3.05	128	1
H	1.00	0.764	0.357	125	4
I	0.10	0.168	0.235	95	34
		Dichloromethane			
F	100	71.80	39.30	40	0
G	10	7.04	3.24	40	0
H	1.00	0.789	0.376	40	0
I	0.10	0.148	0.214	30	10
		Dichloromethane–Propanone			
F	100	69.70	31.10	33	0
G	10	7.76	3.30	33	0
H	1.00	0.84	0.31	32	1
I	0.10	0.18	0.34	27	6
		Propanone–Hexane			
F	100	78.80	31.70	32	0
G	10	7.33	2.62	31	1
H	1.00	0.752	0.304	31	1
I	0.10	0.094	0.078	23	9
		Hexane			
F	100	49.60	28.30	17	0
G	10	4.65	2.37	17	0
H	1.00	0.53	0.56	15	2
I	0.10	0.16	0.24	10	7
		Dichloromethane–Hexane			
F	100	80.70	17.70	4	0
G	10	7.89	1.20	4	0
H	1.00	0.797	0.130	4	0
I	0.10	0.105	0.014	4	0
		Trimethylpentane			
F	100	64.50	4.20	3	0
G	10	6.60	2.40	3	0
H	1.00	0.707	0.012	3	0
I	0.10	0.105		1	2

All results are in mg kg^{-1}.
<RL = Less than reporting limit.
SD = Standard deviation.

Table 6 Comparison of cleanup procedures

	Sample F	Sample G	Sample H	Sample I
Spiked	100	10	1	0.1
		Overall results ($n = 129$)		
Mean	69.90	7.01	0.764	0.168
SD	32.10	3.05	0.357	0.235
<RL		1	4	34
		No cleanup procedure ($n = 43$)		
Mean	69.10	6.75	0.732	0.202
SD	27.31	2.47	0.350	0.222
<RL			1	18
		Florisil only ($n = 20$)		
Mean	63.9	5.9	0.6	0.1
SD	31.3	2.5	0.3	0.1
<RL	0	1	1	7
		Sulfuric acid only ($n = 44$)		
Mean	75.90	7.40	0.833	0.137
SD	39.00	3.54	0.357	0.127
<RL	0	0	2	7
		Alumina column only ($n = 4$)		
Mean	30.90	3.75	0.391	0.053
SD	17.60	1.63	0.194	0.017
<RL			1	
		Mercury only ($n = 3$)		
Mean	75.70	102	1.09	0.120
SD	11.60	5.04	0.515	0.049
<RL				
		Gel permeation only ($n = 3$)		
Mean	79.60	7.93	0.857	0.14
SD	39.50	1.50	0.140	0.014
<RL			1	

All results are in mg kg^{-1}.
<RL = Less than reporting limit.

concentrations this is a trivial amount but when the amount of PCBs is 0.1 mg kg^{-1}, the effect is to double the measured amount of PCBs. Thus, depending on the data quality objectives of the end user, the amount of analyte, and the amount of interferents, the need for cleanup procedures will vary. However, there is no way of knowing how much interferent or analyte is present until the analysis is performed so until a sample is well characterized, the use of the appropriate cleanup procedure is a very prudent step.

As is also quite evident, the use of the wrong cleanup procedure is as bad as not using any cleanup procedure, if not worse. The laboratories that used alumina column cleanup consistently got lower recoveries at all concentrations than any of the other methods. At the critical concentration of 0.1 mg kg^{-1}, neither the gel permeation nor the sulfuric acid cleanup procedures performed as well on average as did the Florisil column procedure.

6 CONCLUSIONS AND FUTURE TRENDS

As should be evident, the sample preparation procedures are both an absolutely critical component of analytical analysis and one that can be very complex. The strengths and limitations of prescriptive standardized methods are quite evident here. These methods can be very effective when they are well designed. However, subtle differences in the chemistry of acid cocktails, extraction solvents, or cleanup columns can cause serious errors in analysis and it can be difficult to exploit these subtleties when trying to adhere to prescriptive methods. For this reason many regulators and analysts have proposed ''method streamlining'' and ''performance based methods systems''.[2] Method streamlining allows laboratories under regulatory constraints to use standardized methods

to modify these methods provided they demonstrate equal or greater efficiency. Performance-based methods systems would do away with any reference to standardized methods and simply allow the analyst to demonstrate that the data quality objectives of the data user are being met. How far these two initiatives will proceed has not yet been determined.

ABBREVIATIONS AND ACRONYMS

APHA	American Public Health Association
ASTM	American Society for Testing and Materials
AWWA	American Water Works Association
CERCLA	Comprehensive Emergency Response Compensation, and Liability Act
CLP	Contract Laboratory Program
ECD	Electron-capture Detector
FAAS	Flame Atomic Absorption Spectroscopy
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometer
NOC	Nonvolatile Organic Compound
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
RCRA	Resource Conservation and Recovery Act
USEPA	United States Environmental Protection Agency
USOSHA	United States Occupational Safety and Health Administration
VOC	Volatile Organic Compound
WEF	Water Environment Federation

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Soil Instrumental Methods • Soil Sampling for the Characterization of Hazardous Waste Sites

Food **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

REFERENCES

1. 'Test Methods for Evaluating Solid Wastes', 3rd edition, 3rd update, Office of Solid Waste and Emergency Response, US Environmental Protection Agency, Washington, DC, 1998.
2. D.E. Kimbrough, R. Spinner, 'Performance Based Methods for Regulatory Analytical Environmental Chemistry', *Am. Environ. Lab.*, November (1994).
3. D.E. Kimbrough, J.R. Wakakuwa, 'A Report of the Linear Ranges of Several Acid Digestion Procedures', *Proc. Sixth Ann. USEPA Symp. Solid Waste Testing and Quality Assurance*, US Government Printing Office, Washington, DC, II-214, July 1990. Re-printed in *Environ. Sci. Technol.*, **26**, 173–178, 1992.
4. Z. Sulcek, P. Povondra, *Methods of Decomposition in Inorganic Analysis*, CRC Press, Boca Raton, FL, 1989.
5. G.G. Hawley (ed.) *Condensed Chemical Dictionary*, 10th edition, Van Nostrand Reinhold Co., New York, 79, 1981.
6. *Kirk–Othmer Encyclopedia of Chemical Engineering*, 3rd edition, Wiley-Interscience, New York, 464–469, Vol. 3, 1978.
7. Method 8280, 'Test Methods for Evaluating Solid Wastes', 3rd edition, 3rd update, Office of Solid Waste and Emergency Response, US Environmental Protection Agency, Washington, DC, 1998.
8. A.L. Alford-Stevens, 'Analyzing PCBs', *Environ. Sci. Technol.*, **20**, 1194–1199 (1986).
9. D.E. Kimbrough, R. Chin, J. Wakakuwa, 'Widespread and Systematic Errors in the Analysis of Soils for Polychlorinated Biphenyls: Part 2. Comparison of Extraction Systems', *The Analyst*, **119**, 1283–1292 (1994).
10. D.E. Kimbrough, R. Chin, J. Wakakuwa, 'Widespread and Systematic Errors in the Analysis of Soils for Polychlorinated Biphenyls: Part 3. Gas Chromatography', *The Analyst*, **119**, 1293–1301 (1994).
11. D.E. Kimbrough, R. Chin, J. Wakakuwa, 'Widespread and Systematic Errors in the Analysis of Soils for Polychlorinated Biphenyls: Part 1. A Review of Interlaboratory studies', *The Analyst*, **119**, 1277–1283 (1994).
12. D.E. Kimbrough, R. Chin, J. Wakakuwa, 'Industry-wide Performance in a Pilot Performance Evaluation Sample Program for Hazardous Materials Laboratories. 2. Precision and Accuracy of Polychlorinated Biphenyls', *Environ. Sci. Technol.*, **26**, 2101–2104 (1992).

Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices

Sergio R. Marín
Chilean Commission of Nuclear Energy, Santiago, Chile

1 Introduction 1
2 Sampling 2
2.1 Water Sample Collection 3
2.2 Water Sample Handling 3
2.3 Solid Waste Sample Collection 3
2.4 Laboratory Sample Reception and Storage 5
2.5 Sample Reduction and Homogeneity 5
3 Laboratory Conditions for Sample Preparation 6
3.1 Laboratory Environment 6
3.2 Apparatus and Equipment 7
3.3 Reagent Quality 7
3.4 Container and Glassware Quality 10
3.5 Glassware Cleaning 10
3.6 Clean Analytical Laboratory 10
3.7 Safety 11
4 Sample Preparation for Water Analysis 11
4.1 Sample Preservation and Storage 12
4.2 Sample Preparation 12
4.3 Concentration Procedure 13
4.4 Seawater Preparation 14
4.5 Interference 14
5 Sample Preparation for Solid Waste Analysis 14
5.1 Open Wet Digestion 15
5.2 Dry Ashing 15
5.3 Wet Ashing Techniques 16
5.4 Pressurized Digestion 16
5.5 Microwave Digestion 16
5.6 Sequential Extraction 16
5.7 Sources of Error 17
6 Environmental Regulation and Methods 17
7 Quality Control 18
7.1 Instrument Calibration 19
7.2 The Blank 19
7.3 Use of Reference Materials 19
8 Development of the Method 20
8.1 Writing Up of the Method 20
Acknowledgments 20
Abbreviations and Acronyms 21
Related Articles 21
References 21

The reliability of results obtained from chemical analysis of environmental samples is related to the sample collection and preparation techniques. A succession of stages is required to obtain a result that is both accurate and representative of the initial material. The task begins by taking the appropriate precautions during collection of the liquid and solid samples, and during their later shipment to the laboratory. The environmental conditions in the laboratory where the sample will be prepared are important, as is the quality of the apparatus, containers, glassware, and reagents that are usually used. The preparation of solid waste samples requires the selection of appropriate digestion procedures, apparatus that is of high enough quality to resist the aggressiveness of some inorganic acids, and convenient systems of security for the procedures that are carried out. The information produced by the laboratory should have a high degree of reliability. For this reason, the procedures used usually permit the results obtained to be guaranteed. The more important procedures include the use of certified reference materials (CRM) and participation in interlaboratory programs. Some organizations, for example the United States Environmental Protection Agency (USEPA), have developed and distributed procedures for environmental sample preparation, which are used by many laboratories as normalized methods of reference.

1 INTRODUCTION

The sample preparation techniques and the sample collection processes are closely related to the accuracy and representativeness of the final results for the material being studied. This is because the values obtained in the analysis of this material depend fundamentally on the sample characteristics and the chemical procedures used. Sample collection is the first and most fundamental of all the stages within the analysis process for environmental materials, since the utility of the final results will depend on the representativeness of this initial collection. Very careful handling is essential during the sample collection process to avoid significant alterations in composition. Sample integrity must be ensured from collection up to production of the final report. Nevertheless, complete

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

stability of all the sample components is impossible, especially in domestic and industrial residual waters, since in the best case the conservation techniques can only delay the chemical and biological changes that inevitably take place after collection. On the other hand, sample containers can introduce errors in the measurement of some metals, as a consequence of container wall leaching and/or concentration decreases due to adsorption. For this reason, the samples collected in the field should be kept in borosilicate glass containers or in plastic containers, such as polyethylene, polypropylene or polytetrafluoroethylene (PTFE). Unfortunately, glass containers can be broken during transport or handling and the PTFE containers are very expensive thus polyethylene and polypropylene containers are preferred.

Special care should be taken regarding the working conditions in the laboratory, although a clean and well-arranged workplace and a supply of reagents and solutions of the appropriate quality are the basic requirements for good work, particularly with soil samples, sediment and polluted waters. Conditions in the laboratory should be more stringent when analysis of pure waters or precipitation waters is required, and extreme care should be taken with environmental air quality, material cleaning and reagent quality. Clean benches that use high-efficiency particulate air (HEPA) filters are advisable for handling this type of sample. Fume hoods should be used when the samples require an energetic process of decomposition by means of inorganic acids, and stainless steel or polyvinyl chloride (PVC) hoods when using perchloric acid. The glassware should be mainly made of borosilicate and should be produced according to the Deutsches Institut für Normung (DIN) and/or the International Standardization Organization (ISO) standards. The reagent quality should be at least of an analytical degree according to specifications of the American Chemical Society (ACS). The water quality should conform to the categories of the American Society for Testing and Materials (ASTM).

Pure water collected in the field does not usually require an initial treatment, but if sensitive methods are used for analysis, an evaporation concentration may be required in some cases. The decomposition of solid samples by open digestion procedures is frequently tiresome, and analytical errors are possible, since some elements could be lost or suffer contamination due to the material and reagents used. However, these procedures have the advantage of destroying the matrix, thus creating safe homogeneity and also eliminating some of the interferences present in the original sample. The choice of an appropriate procedure is difficult, because if the original sample is diluted to avoid high concentrations of dissolved solids, many elements cannot be determined.

It is also necessary to keep in mind that the information obtained in a chemical analysis laboratory usually plays a key part in social or economic decisions. For this reason, the accuracy and the precision of the results obtained should be such that they generate a high degree of reliability and recognition. Therefore the frequency of instrument calibration, the maintenance of equipment, the glassware quality and the reagent quality should be among the most important considerations of the laboratory. At present, one of the most widely used procedures to assure the quality of the results is the use of CRM that have a matrix very similar to the samples being analyzed. The National Institute of Standards and Technology (NIST) and the Bureau Communautaire de Réference (BCR) are the main suppliers of these materials.

2 SAMPLING

The sampling stage is considered one of the most important in the analysis of samples, since the representativeness of their results will depend upon the way they were collected.[1] The objective of the sample collection is to obtain a material portion whose volume is such that it can be easily transported and analysed in the laboratory, and will accurately represent the material from which it was drawn.[2] The sample obtained during the collection phase should not deteriorate or become contaminated before arriving at the laboratory. Any of the components of interest that are present in the sample are at parts per million concentration levels and even lower. It is possible that these may alter completely or partially if the collection is faulty, or if the person who collects the sample doesn't take the necessary precautions for conservation. The collection should be carried out carefully with the purpose of guaranteeing an analytical result that will faithfully represent the real composition.

The details of sample collection vary according to the local conditions, so recommendations that have universal application cannot be specified. Several systems have been proposed in the literature. Startin and Lansdown[3] present a system for the collection and treatment of wastewater from highways and paved areas. Robinson et al.[4] describe operational experiences on the collection, treatment and distribution of contaminated solutions that are generated in landfills due to the rotting of enormous quantities of waste. Blommaert et al.[5] present a study about the uncertainties that can hinder the representativeness of the sampling and/or correct interpretation of the data obtained in a complex waste processing installation. Lohani and Ko[6] present a general approach to determine the sample size and the number of samples in solid waste sampling. The

Department of Energy (DOE) and the USEPA consider the screening methods to be of great benefit in quickly and efficiently studying large numbers of highly polluted places that present a serious risk to inhabitants. Wise and Guerin[7] show the utility of the screening methods as a technique for direct sampling. They discuss the advantages of a direct sampling method in which an analyte is introduced directly from the sample into a mass spectrometer. Olsen et al.[8] propose stripping analysis as a field method to give quick, economic and reliable results in the determination of Cr, Cd, Zn, Cu and Pb in sediments.

2.1 Water Sample Collection

Water sample collection varies depending on the collection source. For domestic water systems, the sample should run down through the pipe to ensure its representativeness. When a water sample collected from a river is analyzed it is necessary to keep in mind that the concentrations of the elements under study will vary according to the depth, the stream speed, the distance from the bank and the width of the river. When sample collection is carried out in deep bodies of water it should be remembered that at depths larger than 5 m there is a tendency towards thermal and chemical stratification of the water. This is why the site and the position of the sampling should be registered exactly to obtain representative samples during the sampling period. This is not a problem for sample collection in waters with a depth less than 5 m, since the waters mix by the action of the wind. In lakes and swamps, besides the stratification problem, variations due to natural causes such as the quantity of rainfall, drainage and wind are also present.

In seawater samples the main problem is to prevent sample deterioration between collection and analysis. This is carried out in two steps. First, by filtering the sample through 0.2 μm or 0.45 μm membrane filters for quick separation of the suspended particulate material, and later by acidification of the sample to a pH around 1 to avoid adsorption by the container walls.[9,10] For certain components, the place where the sample is drawn from is very important;[11] excessively turbulent areas should be avoided to prevent the loss of volatile substances. In general, the samples are collected under the surface, from calm areas, keeping in mind the tests or analyses that will be carried out or the utility that the final results will have.

2.2 Water Sample Handling

Contamination of the sampling apparatus can produce serious errors during collection and storage of the samples, due to inappropriate cleaning of the apparatus. Several types of pollutants are caused by the material from which the sampling apparatus is built. For example, PVC can contribute Zn, Fe, Sb, and Cu; cellulose filters Cu, Ni, Zn, P, K, etc. and plastic tips Cu, Fe, Zn, Cd, Ni. All this contamination can be minimized by washing the material five times or more with distilled water, or allowing it to soak for two days in 5 N HCL and then rinsing it with ultrapure water. The acidity of the sample, the time that it is in contact with the container, and the presence of complex agents and dissolved active carbon all determine the adsorption of metals on the walls of the container. Table 1 summarizes the requirements for sample collection and handling. The blank used during the analysis is essential to assure the results, since this analysis can be used to identify unexpected pollution associated with the purity of the distilled water, inappropriate cleaning of the containers or contamination through air absorbed by the samples during the sampling process.

2.3 Solid Waste Sample Collection

One of the most important parameters for solid samples is how representative they are of the place being studied. Many factors affect this representativeness, but the most important is the degree of preliminary information about the sampling site. Many geological maps exist, to different scales, which can serve as a starting point for comprehensive knowledge of the sampling site. The selection of a good sampling model, appropriate instruments and safe storage will result in a good program of sample collection. If the analysis is for environmental reasons, horizontal samplings are recommended since they are more representative of the surface contamination. With sediment samples, the main problem in trace analysis is to obtain a representative sample. Under normal conditions, the best place to draw sediment samples from a lake is the deepest part of the lake, where the turbulence of the water column is at a minimum. However, for the purposes of trace analysis to assess pollution, the sample should be collected from places near to the discharge points or the coastal areas, which are often those that show greater turbulence. Sediments that have been deposited at the bottom of the lake consist of particles of different size, different hydrodynamic properties, and different chemical composition. These particles originate from algae, fecal pellets, and precipitated compounds whose composition has exceeded the solubility product, etc. The first problem for soil sampling is the soil itself, due to the heterogeneous distribution of particle sizes, chemical composition, etc. which makes it impossible to standardize on sampling and pretreatment procedures. A second aspect to consider is the users' expectations of soil research programs.

Table 1 Summary of requirements for sample collection and manipulation. (Reproduced from *Standard Methods for the Examination of Water and Wastewater*, 19th edition, 1/22, 1995, by permission of the American Public Health Association.)

Determination	Container[a]	Minimum sample size (mL)	Sample type[b]	Preservation[c]	Maximum storage (recommended/ regulatory[d])
Acidity	P, G(B)	100	g	Refrigerate	24 h/14 days
Alkalinity	P, G	200	g	Refrigerate	24 h/14 days
BOD	P, G	1000	g	Refrigerate	6 h/48 h
Boron	P	100	g, c	None required	28 days/6 months
Bromide	P, G	100	g, c	None required	28 days/28 days
Carbon, organic, total	G	100	g, c	Analyze immediately; or refrigerate and add H_3PO_4 or H_2SO_4 to pH < 2	7 days/28 days
Carbon dioxide	P, G	100	g	Analyze immediately	stat/NS
COD	P, G	100	g, c	Analyze as soon as possible, or add H_2SO_4 to pH < 2; refrigerate	7 days/28 days
Chloride	P, G	50	g, c	None required	28 days
Chlorine, residual	P, G	500	g	Analyze immediately	0.5 h/stat
Chlorine dioxide	P, G	500	g	Analyze immediately	0.5 h/NS
Chlorophyll	P, G	500	g, c	30 days in dark	30 days/NS
Color	P, G	500	g, c	Refrigerate	48 h/48 h
Conductivity	P, G	500	g, c	Refrigerate	28 days/28 days
Cyanide: Total	P, G	500	g, c	Add NaOH to pH > 12, refrigerate in dark[d]	24 h/14 days; 24 h if sulfide present
Amenable to chlorination	P, G	500	g, c	Add 100 mg $Na_2S_2O_3L^{-1}$	stat/14 days; 24 h if sulfide present
Fluoride	P	300	g, c	None required	28 days/28 days
Hardness	P, G	100	g, c	Add HNO_3 to pH < 2	6 months/6 months
Iodine	P, G	500	g, c	Analyze immediately	0.5 h/NS
Metals, general	P(A), G(A)	500	g	For dissolved metals filter immediately, add HNO_3 to pH < 2	6 months/6 months
Chromium(VI)	P(A), G(A)	300	g	Refrigerate	24 h/24 h
Copper by colorimetry					
Mercury	P(A), G(A)	500	g, c	Add HNO_3 to pH < 2, 4 °C, refrigerate	28 days/28 days
Nitrogen:					
Ammonia	P, G	500	g, c	Analyze as soon as possible or, add H_2SO_4 to pH < 2, refrigerate	7 days/28 days
Nitrate	P, G	100	g, c	Analyze as soon as possible or refrigerate	48 h/48 h (28 days for chlorinated samples)
Nitrate + nitrite	P, G	200	g, c	Add H_2SO_4 to pH < 2 refrigerate	none/28 days
Nitrite	P, G	100	g, c	Analyze as soon as possible or refrigerate	none/48 h
Organic, Kjeldahl	P, G	500	g, c	Refrigerate; add H_2SO_4 to pH < 2	7 days/28 days
Odor	G	500	g	Analyze as soon as possible; refrigerate	6 h/NS
Oil and grease	G, wide-mouth calibrated	1000	g, c	Add HCl to pH < 2, refrigerate	28 days/28 days
Organic compounds:					
Pesticides	G(S), TFE-lined cap	1000	g, c	Refrigerate; add 1000 mg ascorbic acid L^{-1} if residual chlorine present	7 days/7 days until extraction; 40 days after extraction
Phenols	P, G	500	g, c	Refrigerate, add H_2SO_4 to pH < 2	128 days
Purgeables by purge-and-trap	G, TFE-lined cap	2 × 40	g	Refrigerate; add HCl to pH < 2; add 1000 mg ascorbic acid L^{-1} if residual chlorine present	7 days/14 days

Table 1 (*continued*)

Determination	Container[a]	Minimum sample size (mL)	Sample type[b]	Preservation[c]	Maximum storage (recommended/ regulatory[d])
Oxygen, dissolved:	G, BOD bottle	300	g		
Electrode				Analyze immediately	0.5 h/stat
Winkler				Titration may be delayed after acidification	8 h/8 h
Ozone	G	1000	g	Analyze immediately	0.5 h/NS
pH	P, G	50	g	Analyze immediately	2 h/stat
Phosphate	G(A)	100	g	For dissolved phosphate filter immediately; refrigerate	48 h/NS
Salinity	G, wax seal	240	g	Analyze immediately or use wax seal	6 months/NS
Silica	P	200	g, c	Refrigerate, do not freeze	28 days/28 days
Sludge digester gas	G, gas bottle	–	g	–	NS
Solids	P, G	200	g, c	Refrigerate	7 days/2–7 days; see cited reference
Sulfate	P, G	100	g, c	Refrigerate	28 days/28 days
Sulfide	P, G	100	g, c	Refrigerate; add 4 drops of 2 N zinc acetate per 100 mL; add NaOH to pH > 9	28 days/7 days
Taste	G	500	g	Analyze as soon as possible; refrigerate	24 h/NS
Temperature	P, G	–	g	Analyze immediately	stat/stat
Turbidity	P, G	100	g, c	Analyze same day; store in dark up to 24 h, refrigerate	24 h/48 h

For determinations not listed, use glass or plastic containers; preferably refrigerate during storage and analyze as soon as possible.
[a] P = plastic (polyethylene or equivalent); G = glass; G(A) or P(A) = rinsed with 1 + 1 HNO_3; G(B) = glass, borosilicate; G(S) = glass, rinsed with organic solvents or baked.
[b] g = grab; c = composite.
[c] Refrigerate = storage at 4 °C in the dark.
[d] Environmental Protection Agency, Rules and Regulations. 40 CFR Parts 100–149, July 1, 1992. See this citation for possible differences regarding container and preservation requirements. NS = not stated in cited reference; stat = no storage allowed, analyze immediately.
BOD, biochemical oxygen demand; COD, chemical oxygen demand; TFE, tetrafluoroethylene.

2.4 Laboratory Sample Reception and Storage

Before analysis, the water samples are generally stored in one or more bottles or bags, which are made of glass, polypropylene, polyethylene and/or PTFE, each of these materials offering advantages and disadvantages. Once the sample has been taken, all the efforts will be directed towards minimizing the handling. The samples can change physically or chemically; this is why it is necessary to preserve them from deterioration. The preservation is important, since samples that don't require pretreatment should be kept at 4 °C as soon as they have been drawn and filtered. If the period of storage is longer than 2 or 3 days, it is convenient to maintain them at lower temperatures (down to −20 °C) in plastic containers,[12,13] as this reduces the activity of the bacteria. Samples for metal analysis should be acidified with nitric acid to a pH lower than 2. Those samples that will be analyzed by dissolved metals should be filtered before acidification. After preservation, they should be kept at 4 °C. As a general rule, 1.5 mL^{-1} of ultrapure HNO_3 per liter is used, as this is sufficient to give an appropriate acidity to the samples. All apparatus that will be in contact with the samples should first be cleaned and washed rigorously with 6 N HCl, and then with an excess of ultrapure water. The membranes used for filtering should be washed in 5 N HCl and then washed thoroughly with twice-distilled water. Great care should be taken to avoid sample contamination; twice-distilled or ultrapure water should be used to dilute and wash the material, and the acids and all other reagents used must be of high purity. All samples should be labelled at the site where they were collected; self-adhesive labels should be used to avoid false identifications. The labels should have at least the following information: number of the sample, name of the person who carried out the collection, date, time and place. A field record book should also be kept, containing all data regarding the sample collection process.

2.5 Sample Reduction and Homogeneity

Only a fraction of the field sample is used, and since this fraction undergoes the whole handling process, it is

necessary to keep in mind the possible contamination that could take place during the procedures and storage. Sediment and soil preparation involve drying and sieving, and in special cases size reduction by milling. The 63 μm fraction is suited for most chemical analyses, but it is not sufficiently abundant in many soils and sediments, where only few grams can be obtained. In these cases the particle size will be defined according to the purposes of the study. Furthermore, it is sometimes necessary to perform detailed examinations of the influence of the sieved and milled particle size in the samples under study. Soils and sediments should be kept in humidity-resistant containers. Drying temperatures should not be higher than 65 °C, since volatile elements present in the sample can be lost at higher temperatures. When drying, it is necessary to avoid the sample becoming a dry mud, since this makes the disintegration more difficult. After drying, the sample can be disintegrated and sieved. The object of the disintegration is not to mill the material but to break down the pieces to obtain a natural grain size distribution. The sample should preferably be sieved through a silk or synthetic fiber sieve, since stainless steel sieves, besides being more expensive, have usually been welded during their production. The sieves should be cleaned, preferably with compressed air.

3 LABORATORY CONDITIONS FOR SAMPLE PREPARATION

This section presents the basic conditions that a laboratory should meet to carry out work related to sample preparation, prior to analysis. Element concentrations in environmental materials can range from percentage levels down to a few parts per billion. Considering the very lowest concentration levels that will be analyzed and the special elements (e.g. lead, cadmium, arsenic, mercury, selenium, chromium, etc.), a clean area of the laboratory should be chosen for performing this activity.[14] Interesting work was carried out by Kanatharana et al.,[15] who established a clean laboratory from a traditional chemistry laboratory. In the last decade, in order to carry out environmental studies in Antarctic snow and ice,[16–19] the need to determine different heavy metals at the lowest concentration levels (down to parts per billion) has been increased. Boutron[20] describes a laboratory for the reliable determination of several heavy metals at picogram per gram concentration levels in the snow and ice of Antarctica and Greenland. This laboratory was developed in collaboration with Patterson's group of the Technological Institute of California.

Particular care should also be taken regarding the quality of the reagents and materials that will be used, since these are important sources of contamination. Likewise, it is necessary to keep in mind the quality of the apparatus and instruments used in the sample decomposition processes. This is also the case with hot plates, fume hoods, heating blocks, etc., since during the sample decomposition process, copious fumes and highly aggressive vapors are usually generated with most of the inorganic acids. This situation could cause damage to the instruments, as well as being a major source of contamination. Special care should be taken when cleaning the materials and containers used in the sample preparation processes, especially with substances that are strongly attached, such as the grease used to lubricate the taper joints.[21–23] The use of mixtures containing chromic acid should be avoided as much as possible, because of the risk when using this acid.[24] Attention should be paid to selecting the quality of the cleaning substances, since some pollutants that enter this way could significantly change the final results.

The Standard Method[25] gives particular consideration to security in the laboratory, since the characteristics of the work mean that personnel are potentially exposed to inhalation of different types of gases, as well as injuries such as burns and, to a smaller degree, poisoning by ingestion. Good training on the risks and precautions that must be taken in laboratory work is a necessary policy. Detailed and practical information appears in the manual by Bernabei.[26]

3.1 Laboratory Environment

One of the important factors to consider inside the laboratory is control of the air that enters during the analysis of trace and ultratrace elements in environmental materials, since this could be the cause of possible contamination. This problem has been increased by the increase in atmospheric particles due to industrialization. Pollutants that circulate in the atmosphere in the form of powder, fog or smoke can enter the laboratory through the ventilation systems, which is why the composition of the laboratory air becomes similar to that of the surrounding atmosphere. Usually, there are solid and liquid particles in the air derived from natural sources or coming from industrial processes. To achieve a clean atmosphere in the laboratory, in the first place it is recommended that all air inlets, such as the air-conditioning systems, and the heating and ventilation systems, are filtered efficiently. It is advisable not to smoke in the laboratory where the analyses are carried out. Painting, welding or other types of activity that would alter the normal atmospheric conditions should also be avoided. When performing special work, it is convenient to select and maintain an area of the laboratory under essentially clean conditions, in which

the walls and floor should be painted with epoxy-type paint and the amount of people entering should also be reduced. All unnecessary bookcases and pieces of furniture should be eliminated, since these can be sources of dust accumulation. All the windows should be tightly closed and sealed to prevent the entry of dust.

However, it is necessary to keep in mind that unless the work is carried out in a highly polluted atmosphere, the contamination that comes from the air is usually not the most critical in the analysis process. In general, the most sensitive situations are related to contamination from the reagents and containers, especially when they are subject to rigorous digestion processes. For this reason, before making big investments in building special clean rooms for the preparation of samples, it is necessary to carry out a study that determines if the contamination problems really come from a polluted atmosphere. Often good classification, suitable cleaning and appropriate selection of the quality of the reagents and glassware are enough.

3.2 Apparatus and Equipment

The first activity in the preparation of samples for elemental analysis is to measure a quantity of material. In the case of solid samples, this is carried out using an analytical balance. Consequently, it is necessary to take special care with the calibration of this instrument, since errors at this initial stage are difficult to detect and can invalidate the final results. It is usually advisable that this calibration be carried out daily, using several types of analytical certificate weights. The later stages in the preparation of solid samples involve material decomposition processes, which are usually carried out by means of aggressive attacks with inorganic acids.

Usually the apparatus used in the sample decomposition processes, such as hot plates, digestion systems and fume hoods for vapour extraction, is exposed to an atmosphere containing inorganic acid vapours. For this reason, the most important factors to consider at the sample digestion stage are the quality and the efficient operation of the gas extraction systems inside the laboratory. Highly efficient fume hoods which incorporate electronic detection systems are available. They also warn the operator when the extraction speed slows to dangerous levels. When $HClO_4$ is used in the sample decomposition process it is necessary to use hoods specifically designed to expel the dangerous vapours of this acid from the laboratory. To protect the hood from the action of $HClO_4$ and other corrosive reagents, the entire base of the interior is built of PVC. Hoods where the interior is built from a single piece of stainless steel 316, integrating the work surface and the waste discharge, are also used, but they are not suitable for digestions involving hydrochloric and nitric acid mixtures.

The hot plates used for the sample digestion should ideally have a cover manufactured from ceramic material, or alternatively from aluminum. Aluminum covers offer a more uniform temperature on the surface, but are less resistant to aggressive atmospheres. Some hot plates are specially designed for remote operation in hoods or other corrosive atmospheres, and the controls can be placed up to 1.0 m from the hot plate, for security and convenience. The best plates are made of steel covered with ceramic, and a cover that is especially resistant to corrosive vapours protects the electric cord that goes from the hot plate to the control. To a smaller degree the oven muffles are also affected by some of the corrosive substances that are produced from the sample decomposition processes. The risk of contamination of the samples is the main worry during the use of these items. In general the corrosiveness of the acids used in the digestion processes oxidize the component materials of the apparatus, a situation that can be a source of contamination. On the other hand, the movement of the air inside the hoods during the digestion process can also be a source of contamination.

For solid samples the preparation stage prior to decomposition habitually involves disintegration, sifting, and occasionally milling. Special care should be taken when selecting the types of material to use in these processes, because of the difficulty in evaluating the contamination introduced by the blank determination, as it is impossible to keep the blanks constant during these operations. It is necessary to carry out periodic and systematic maintenance of the apparatus used in the processes, since it affects the quality of the results obtained. However, the greatest source of error is associated with the sample gathering and chemical preparation processes.

3.3 Reagent Quality

The reagents should be at least of analytical quality according to the specifications of the ACS. For work where the determination of trace and ultratrace elements is required, it is necessary to use high purity reagents. The quality of the reagents used for sample preparation is important, since they could be sources of contamination. If the concentrations of the pollutants are very low, this could have an important effect on the final determination of some elements, due to the quantity of acid that is usually used in the digestion processes.

Different suppliers have produced a great variety of these reagents, with guarantee certificates that indicate that they can be used to analyze different elements at required levels, without contamination problems. Figure 1 shows the certificate of analysis for Suprapur® nitric acid and Figure 2 shows the corresponding certificate for hydrochloric acid, both from Merck. Ultrapure

CERTIFICATE OF ANALYSIS

MERCK

441 Nitric Acid 65% Suprapur®
Lot No.: B878741 Assay (acidimetric): 65.5%

	Guarantee values ($\mu g\ g^{-1}$ = ppm)	Actual values ($\mu g\ g^{-1}$ = ppm)
Chloride (Cl)	max. 0.05	<0.05
Phosphate (PO_4)	max. 0.01	<0.005
Sulfate (SO_4)	max. 0.2	<0.1
Ag (Silver)	max. 0.001	<0.001
Al (Aluminum)	max. 0.005	0.002
As (Arsenic)	max. 0.001	<0.001
Au (Gold)	max. 0.005	<0.001
Ba (Barium)	max. 0.005	<0.001
Be (Beryllium)	max. 0.001	<0.001
Bi (Bismuth)	max. 0.005	<0.001
Ca (Calcium)	max. 0.03	0.013
Cd (Cadmium)	max. 0.001	<0.001
Co (Cobalt)	max. 0.001	<0.001
Cr (Chromium)	max. 0.002	<0.001
Cu (Copper)	max. 0.001	<0.001
Fe (Iron)	max. 0.01	<0.004
Ga (Gallium)	max 0.005	<0.001
Ge (Germanium)	max 0.005	<0.001
Hg (Mercury)	max. 0.002	<0.001
In (Indium)	max. 0.002	<0.001
K (Potassium)	max. 0.01	<0.002
Li (Lithium)	max. 0.001	<0.001
Mg (Magnesium)	max. 0.01	0.006
Mn (Manganese)	max. 0.001	<0.001
Mo (Molybdenum)	max. 0.001	<0.001
Na (Sodium)	max. 0.01	<0.004
Ni (Nickel)	max. 0.01	<0.003
Pb (Lead)	max. 0.001	<0.001
Sn (Tin)	max 0.005	<0.001
Sr (Strontium)	max. 0.001	<0.001
Ti (Titanium)	max. 0.001	<0.001
Tl (Thallium)	max. 0.001	<0.001
V (Vanadium)	max. 0.001	<0.001
Zn (Zinc)	max. 0.005	<0.001
Zr (Zirconium)	max. 0.001	<0.001
Residue after ignition (as sulfate)	max. 2.0	<1.0

Actual analysis values are subject to unavoidable systematic variations in this concentration range
Date of release: 16. 07. 1996
Maximum shelf time: 60 months

Blum
PQ-Lab 1.3

Figure 1 Certificate of analysis for Suprapur® nitric acid. (Reproduced by permission of Merck.)

reagents are expensive, and cost several times as much as their analytic equivalents. Therefore many laboratories have opted to produce the inorganic acids that they need by means of a sub-boiling system. The laboratories usually use this apparatus to obtain HNO_3 and HCl, since they are the most frequently used acids. These acids are also advantageous because of their low boiling points, unlike sulfuric and perchloric acid. Perchloric acids have the added difficulty that they are dangerous.

Water is the solvent most used in the laboratory and is therefore most important. At present, many procedures exist for the purification of water used in the laboratory; including distillation, deionization, membrane filtration and reverse osmosis. Depending on the quality of the feed water and the volume of pure water required, it may

CERTIFICATE OF ANALYSIS

MERCK

318 Hydrochloric Acid 30% Suprapur®

Lot No.: B899318 Assay (acidimetric): 31.6%

	Guarantee values ($\mu g\ g^{-1}$ = ppm)	Actual values ($\mu g\ g^{-1}$ = ppm)
Colour (Hazen)	max. 10.0	<10.0
Bromide (Br)	max. 50.0	<2.0
Free chlorine (Cl)	max. 0.5	<0.2
Phosphate (PO_4)	max. 0.01	<0.01
Sulfate (SO_4)	max. 0.5	<0.2
Sulfite (SO_3)	max. 1.0	<0.5
Ag (Silver)	max. 0.001	<0.001
Al (Aluminum)	max. 0.01	<0.001
As (Arsenic)	max. 0.005	<0.001
Au (Gold)	max. 0.005	<0.001
Ba (Barium)	max. 0.005	<0.001
Be (Beryllium)	max. 0.001	<0.001
Bi (Bismuth)	max. 0.005	<0.001
Ca (Calcium)	max. 0.03	0.003
Cd (Cadmium)	max. 0.001	<0.001
Co (Cobalt)	max. 0.001	<0.001
Cr (Chromium)	max. 0.001	<0.001
Cu (Copper)	max. 0.001	<0.001
Fe (Iron)	max. 0.01	<0.003
Ga (Gallium)	max 0.005	<0.001
Hg (Mercury)	max. 0.005	<0.001
In (Indium)	max. 0.002	<0.001
K (Potassium)	max. 0.01	<0.001
Li (Lithium)	max. 0.001	<0.001
Mg (Magnesium)	max. 0.005	0.001
Mn (Manganese)	max. 0.001	<0.001
Mo (Molybdenum)	max. 0.001	<0.001
NH_4 (Ammonium)	max. 1.0	<0.5
Na (Sodium)	max. 0.02	<0.001
Ni (Nickel)	max. 0.005	<0.001
Pb (Lead)	max. 0.001	<0.001
Pt (Platinum)	max. 0.01	<0.001
Sn (Tin)	max 0.001	<0.001
Sr (Strontium)	max. 0.001	<0.001
Ti (Titanium)	max. 0.001	<0.001
Ti (Thallium)	max. 0.001	<0.001
V (Vanadium)	max. 0.001	<0.001
Zn (Zinc)	max. 0.002	<0.001
Zr (Zirconium)	max. 0.001	<0.001
Residue after ignition (as sulfate)	max. 2.0	<1.0

Actual analysis values are subject to unavoidable systematic variations in this concentration range

Date of release: 08. 10. 1996

Maximum shelf time: 60 months

Blum

PQ-Lab 1.3

Figure 2 Certificate of analysis for Suprapur® hydrochloric acid. (Reproduced by permission of Merck.)

be necessary to combine two of these systems. The ASTM establishes three categories regarding the quality of pure water for use in the laboratory, as shown in Table 2. Type I water is used for the preparation of solutions when trace and ultratrace elements are to be analyzed with maximum accuracy and precision. This water is obtained using a mixed resin, which eliminates dissolved inorganic compounds until a resistivity of 18 MΩ cm is reached. It is also passed through active carbon cartridges to eliminate dissolved organic matter. Type II water is recommended

Table 2 Classification of water quality to be used in the laboratory. (Reproduced from *Standard Methods for the Examination of Water and Wastewater*, 19th edition, 1/25, 1995, by permission of the American Public Health Association.)

Quality parameter	Type I	Type II	Type III
Bacteria (CFU mL^{-1})	<10[a]	<1000	NS
pH	NS	NS	5–8
Resistivity (MΩ cm at 25 °C)	>10	>1	0.1
Conductivity (MΩ cm^{-1} at 25 °C)	<0.1	<1	10
SiO_2 (mg L^{-1})	<0.05	<0.1	<1
Particulate matter[b]	0.22 μm filter	NS	NS
Organic contaminants[b]	Activated carbon[c]	NS	NS

[a] Preferably bacteria-free.
[b] Process specification; not measured by end user.
[c] Pretreatment, and possibly post-treatment for certain uses.
NS = not specified.

for most laboratory tests where the analysis of elements is not required to trace levels. It is prepared by distillation, and is recommended for use when it is important that no organic impurities are present. Finally, type III water is satisfactory for qualitative analysis. It can also be used for general glassware cleaning. It is obtained by distillation, deionization in mixed resin or inverse osmosis.

3.4 Container and Glassware Quality

Sample collection in the field is usually carried out using polypropylene or polyethylene containers, which should be of high quality and not coloured, since the presence of pigments means that they contain elements that could introduce some pollutant into the sample. Borosilicate glass should be used in the laboratory, and should meet international standards such as DIN/ISO 3585. For special work, quartz glass should be used. Borosilicate glass has the advantage of resisting acids, salt solutions, organic substances and also halogens such as chlorine and bromine. It also has a relatively good resistance to alkaline solutions. Only fluorohydric acid, concentrated phosphoric acid and strongly alkaline solutions attack the surface of the glass at high temperatures. Glass should not be exposed to abrupt changes of temperature, for example it should not be removed from a hot oven and placed on a laboratory surface that is cold or wet. This is especially important in glass apparatus with thick walls, such as filtering flasks or desiccators. Also glass apparatus should not be subjected to abrupt changes of pressure. Glass items such as Erlenmeyer flasks should not be subjected to pressure or vacuum.

3.5 Glassware Cleaning

Once field samples are received in the laboratory, and before beginning the analysis, the exterior of the containers should be washed with nitric acid, and then rinsed with high purity water to avoid possible contamination. A series of products have been developed for cleaning glassware or plastic, ceramic or metal containers used in the laboratory. These products have the advantage of being free from chlorine, and of being environmentally friendly, since all the components are biodegradable. Some varieties are also free of phosphate, which means that they can be used to clean material that will later be used to determine this element at trace levels. The products are quite reliable, since they clean plastics, glasses, and metals carefully, with analytic purity, and without leaving residues.

To take the best care of laboratory apparatus, it should be cleaned immediately after use, at low temperatures and for a short time. This is the only way to prevent dirt sticking and to avoid damage to the glass due to adherence of chemical substances. Laboratory glass should never be cleaned with abrasive agents since they can damage the surface. In the immersion bath process, the laboratory glassware is placed in the cleaning liquid for the period of time recommended by the detergent manufacturers.[27,28] It is then rinsed with flowing water and finally with distilled water. It is not advisable to expose the material to the cleaning solution for a long time, or to increase the temperature, except in the case of persistent dirt. Long periods of contact with the detergent at temperatures of 70 °C should be avoided, since some alkaline agents in the detergent can destroy the graduations. Cleaning in a dishwasher is a more careful treatment than cleaning in the immersion bath, since the glassware is in contact with the cleaning liquid only for short washing cycles. Boutron[20] recommends extremely rigorous procedures for cleaning plastic, PTFE and glass materials when the determination of metals is required at picogram per gram levels. The initial cleaning solution should contain up to 25% high purity nitric acid in high purity water.

3.6 Clean Analytical Laboratory

The use of clean laboratories with air filtered through HEPA filters is obligatory if reliable measures of heavy metals and other elements to trace and ultratrace concentration levels are required. Classification of clean atmospheres is based on the maximum number of particles between 0.5 and 5.0 mm diameter in the volume of air that exists in a certain space. Table 3 shows the different classes of filtered atmospheres. It has been proven that the most efficient HEPA filters are the dry type, since they have a greater capacity for the capture of small particles. The laminar flow hoods that are used

Table 3 Different classes of filtered atmospheres

Class[a]	Particle concentration per ft^3	Particle diameter (μm)
100	100	0.5 μm and larger
	0	5.0 μm and larger
10 000	10 000	0.5 μm and larger
	65	5.0 μm and larger
100 000	100 000	0.5 μm and larger
	700	5.0 μm and larger

[a] The standard requires laminar flow equipment to attain this level of cleanliness. Since dust particles smaller than 0.5 μm do not introduce substantial errors, 0.5 μm has been adopted as the criterion of measurement.

in clean places for preliminary sample manipulation are designed to protect the operator, the product and the laboratory environment. These cabinets are not recommended for work in the presence of volatile toxic reagents. Two vertical laminar flow clean benches were used by Boutron[20] for the most critical contamination problems, and for operations that generate vapors, such as the evaporation of acid solutions.

3.7 Safety

Laboratory activities related to the handling of chemical products and apparatus can be dangerous to employee's health. It is important that employees are aware of the risks that are present in a laboratory and the prevention mechanisms. The main safety features that should be available in a laboratory are fire extinguishers, showers, eye washes, storage facilities, gas extraction hoods, and warning signs. Clothes are the most important means of protection, and for this reason should be made from material that is resistant to possible acid or alkali splashes. Gloves should be resistant to hot inorganic acid vapours, and safety glasses should be made of shatter-proof glass. It is also sensible to have breathing apparatus for special situations.

The chemicals usually used in the laboratory are quite safe if they are employed carefully, while respecting the warnings on the product labels. Special care should be taken when perchloric acid is used, since this reacts violently or explosively when it comes into contact with organic matter. Hoods that have been used with perchloric acid should not be used with organic reagents. It is best to have a fume hood exclusively for work with perchloric acid. Risks can be reduced by taking the following precautions.

- Do not add $HClO_4$ to a hot solution that contains organic matter.
- If samples contain organic matter always test first with HNO_3 before the addition of $HClO_4$.
- Avoid repeated evaporation to vapours when working in hoods. If necessary, use special hoods for $HClO_4$.
- Never evaporate to dryness samples that are digested with $HClO_4$.

CAUTION don't use perchloric acid for the digestion of membrane filters.

The analysis of waters involves many risks, so it is advisable that during the sample manipulation processes the analysts use plastic gloves, preferably disposable ones. When working with very polluted samples, as residues or wastewater, pipette-fillers should be used to avoid ingestion accidents, preferably pipette fillers with disposable tips. Untreated waters can contain water-borne microorganisms, so it is necessary, before washing the glassware, to decontaminate all the used pipettes, by placing them in a jar containing a disinfectant solution. To avoid contamination by contact, it is important to maintain appropriate personal hygiene, which means frequently disinfecting the hands and the work surface.

Electrical installations, connections and apparatus should be installed according to the correct standards. Incorrect use of this equipment can lead to serious hazards, such as fires, explosions and electrical discharges. Protection or security devices should be prepared for transmission belts, pulleys, chain transmitters, axes and other types of mechanical transmission apparatus. The vacuum pump, mixers, mills and crushers also require this type of protection.

4 SAMPLE PREPARATION FOR WATER ANALYSIS

The procedures used in the preparation of water samples will depend on the purpose of the study, the sample type (natural, waste, sea, etc.) and the analysis technique that will be used. For this reason, it is necessary to decide if the elements to be determined correspond to dissolved metals (those that pass through a 0.45 mm membrane filter are only a fraction of the sample without acidification), suspended metals (those that are retained in the 0.45 mm filter membrane), total metals (elements that are in solution after intense digestion of the whole sample), or metals extracted with acid (elements that are present after the treatment of a sample without filtering with hot dilute mineral acid).

Conservation of the sample is practically impossible, especially if it is a sample of residual domestic or industrial water. In the best case the conservation techniques are only able to slow down the chemical and biological changes that inevitably develop with time after the sample is taken. Some types of samples are more affected

than others by the conservation, and certain elements will be lost. Some authors outline the advantages of carrying out preconcentration of the sample in the field. Okumura et al.[29] propose a procedure involving iron(II) preconcentration and spectrophotometric analysis of the elution, which is applicable to seawater and fresh waters.

Most of the elements that should be analyzed in an environmental water sample are usually present at very low concentration levels, so it is often necessary to use a preconcentration treatment, before determination of the elements.[30] Zou et al.[31] studied a preconcentration system involving separation of Cr(III) on-line and coprecipitation with lanthanum hydroxide. Luo Fangruo and Hou Xiandeng,[32] oxidized thallium(I) with a bromide solution and then preconcentrated it in a column of polyurethane plastic foam. Chikuma and Aoki[33] preconcentrated lead using a chelate-forming resin. Other authors recommend extraction systems to avoid matrix effects. Starostin and Witkiewicz[12] present a review of the literature regarding extraction procedures. Adachi et al.[34] proposed a method in which the vanadium is first separated from the water sample, by extraction with xylene. Paull et al.[35] studied trace elements in seawater, using a high-performance chelating ion chromatographic system, which incorporated a simple chelating-dyestuff-treated column for preconcentration and combined separation of the elements. Because of the interferences that can produce polyatomic ions in the determination of trace elements by inductively coupled plasma mass spectrometry (ICPMS), Bloxham et al.[36] proposed eliminating matrix effects in the analysis of seawater by chelating the analytes on an iminodiacetate-based resin. Chapple and Byrne[37] describe a method in which the seawater matrix interferences are eliminated by a combined process of in situ separation of the analyte and electrothermal vaporization before ICPMS. Special treatments are considered to carry out speciation of some elements. Tanzer and Heumann[38] proposed a methodology for the selective determination of different chemical species of selenium in environmental water samples. Horvat et al.[37] compared the distillation process with other traditional processes in the determination of methyl mercury.

4.1 Sample Preservation and Storage

The sampling methods and sample preparation procedures depend on the type of determination and the sample source (surface water, underground water, drinking water, etc.). About 250–1000 mL of water is necessary for a simple determination, although larger volumes would be required in some cases. After collection, the samples should be filtered to separate the dissolved metals, and acidified with concentrated nitric acid (HNO_3) to pH < 2, to prevent adsorption of the analytes on the container walls. For short-term conservation, $1.5\,mL^{-1}$ concentrated HNO_3 per liter of sample, is usually enough. For samples with high buffer capacity it is necessary to increase the quantity of acid (by up to 5 mL for some alkaline samples), or to use high-purity acid. The acidified sample should preferably be conserved at a temperature of 4 °C to avoid changes in volume caused by evaporation. Under these conditions, samples with metal concentrations of several milligrams per liter remain stable for up to six months. If the levels of metals are of the order of micrograms per liter, the samples should be analyzed as soon as possible after collection. To conserve the sample when the determination of mercury is required, it is advisable to add 2 mL of solution of 20% $K_2Cr_2O_7$ (prepared in HNO_3 1 + 1) per liter of sample. If the samples are to be stored for a long period without previous treatment, it is recommended that they are kept at temperatures of −20 °C, in plastic containers.[12] Once defrosted it will be necessary to filter them with 0.45 μm filters and acidify the fraction that will be used for metals analysis.

4.2 Sample Preparation

In general when the samples are colourless, transparent, of low turbidity, odorless and single-phase, the total metals are analyzed directly. Samples that contain particles or organic matter generally require pretreatment before analysis of the total metals.[40] To analyze dissolved metals, it is necessary to filter the sample, acidify the filtrate and then analyze it directly. To determine suspended metals, the sample is filtered, and the solution produced from digestion of the material that covers the filter is analyzed. To determine metals extracted by acid, the sample is not filtered, but is acidified with HNO_3, and treated with hot HCl. Because the digestion of some samples is unavoidable, it is necessary to work under strictly controlled conditions to obtain representative and reproducible results.

To reduce interference from organic matter and solubilize the associated metal, an appropriate digestion procedure should be used. It is preferable to use a less complex digestion method that allows complete and consistent recovery, but is still compatible with the analytical method and with the metal that is analyzed. Nitric acid decomposes most of the samples in an appropriate form, and nitrate is also an acceptable matrix for most of the chemical analysis techniques.[41] Some samples will require the addition of perchloric, hydrochloric or sulphuric acid for complete digestion. These acids can interfere in the analysis of some metals and they all provide a more problematic matrix for electrochemical analysis. Prior to the preparation it is necessary to confirm the recovery of the metal for each

digestion and analytical procedure used. As a general rule, HNO_3 is only used in the case of clean samples or materials that are oxidized easily. Digestion with $HNO_3-H_2SO_4$ or with HNO_3-HCl is easily adapted for oxidizable organic matter; in the case of organic matter or mineral that is difficult to oxidize it is necessary to use HNO_3-HClO_4 or HNO_3-HF. In the case of lead determination in the presence of high quantities of sulfate (for example, determination of Pb in fly ash of a thermal power plant) it is necessary to dissolve the precipitate of $PbSO_4$. When $HClO_4$ is used it should be noted that contact of this acid with organic matter can cause violent explosions (see section 3.7). If large quantities of organic matter are present, dry combustion is useful. Balint et al.[42] carried out digestion of wastewater samples prior to the determination of total mercury by spectrofluorimetry. A solution of potassium permanganate was added to a fraction of the filtered sample to cause preoxidation to ionic mercury, and later on the sample was treated with a mixture of nitric and sulphuric acid and kept at room temperature.

4.3 Concentration Procedure

When the analysis techniques don't have the appropriate sensitivity, it is necessary to use concentration procedures. These processes are inconvenient in that the sample is exposed to possible contamination and also, if the water samples don't come from very pure sources, some elements can be concentrated at levels that could produce interference in the analytical determination. The evaporation should preferably be carried out using some system that minimizes possible interferences. Good results can be obtained if the sample is heated inside a chamber with injection of filtered air,[14] as is shown in Figure 3. To avoid contamination due to atmospheric pollutants during the determination of elements at ultratrace levels, Görlach and Boutron[18] carried out the sample evaporation procedure inside a vertical laminar flow hood, as shown in Figure 4.

The digestion procedures are usually used to concentrate the sample for analysis. Therefore these processes usually require that the sample volume taken for analysis be reduced almost to dryness. Thus in the dissolution stage a much smaller sample volume is used. But these processes have the drawback, on most occasions, of causing

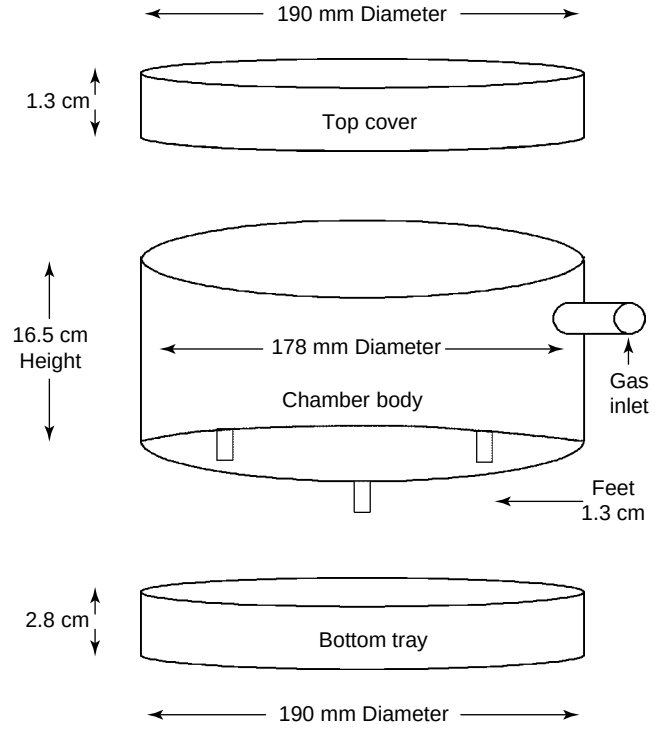

Figure 3 Environmental evaporation chamber for predigestion and evaporation. (Reproduced from *J. Anal. At. Spectrom.*, **11**, 751–757 (1996) by permission of The Royal Society of Chemistry.)

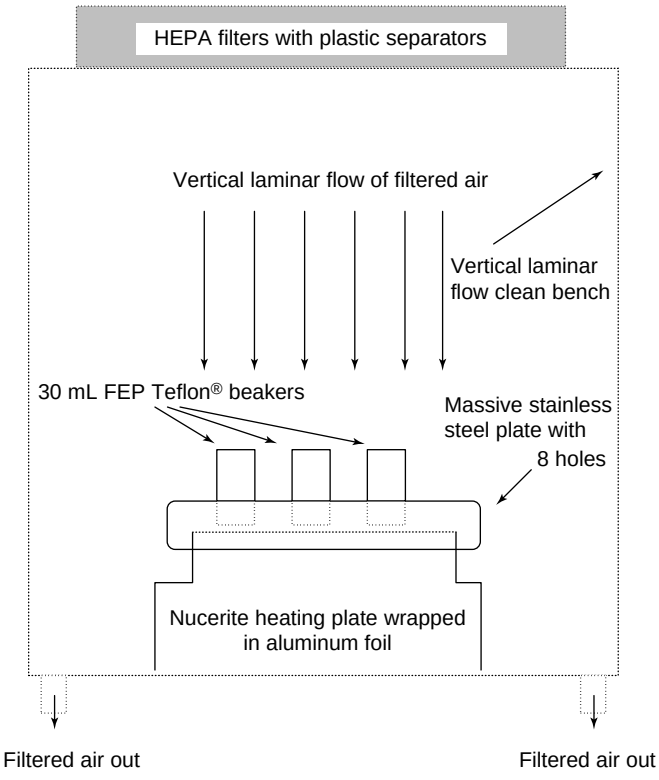

Figure 4 Vertical laminar flow clean bench used for preconcentration of water to extremely low concentrations. FEP, fore edges painted. (Reprinted from U. Görlach, C.F. Boutron, 'Preconcentration of Lead, Cadmium, Copper, and Zinc in Water at the pg g^{-1} Level Non-boiling Evaporation', *Anal. Chim. Acta*, **236**, 391–390. Copyright 1990, with permission from Elsevier Science.)

a high concentration of dissolved salts, which are concentrated as a consequence of the digestion process.

4.4 Seawater Preparation

The analysis of seawater is currently one of the most important challenges in environmental analytical chemistry, due to the very low concentration levels required to analyze the toxic elements that may be present. Usually preparation of the sample before analysis is very dependent on the technique to be used. Only the standard procedures during sample collection (i.e. immediate filtration to below 0.45 mm in the field, acidification to pH 1.5 with high purity nitric acid, and conservation at 4 °C) are common to all techniques. Voltammetry doesn't usually involve treatment of the samples, since the saline environment is very advantageous for analysis,[10] but generally the sample is irradiated with ultraviolet light to eliminate organic matter.[43] The direct determination of metals by atomic absorption spectrometry (AAS) with a graphite furnace is practically impossible. The sample is usually preconcentrated before determination of the elements,[44] or matrix modifiers are used.[45] Recently ICPMS has become one of the most frequently used techniques in the determination of seawater. Some authors analyze this type of water without pretreatment,[46] but it is more common to carry out some form of pretreatment, either solvent extraction[47] or column separation.[48]

4.5 Interference

During the preliminary collection or sample preparation treatments, great care must be taken not to introduce metals into the samples. Contact with rubber, metal-based paints, cigarette smoke, and paper napkins, as well as all metallic products, including stainless steel, galvanized metal and brass, should be avoided. Conventional fume hoods can contribute significantly to the contamination of the sample, mainly during digestion with acids in open containers. In the case of tips for micropipettes, care should be taken regarding their origin and quality, since they can be polluted with copper, iron, zinc and cadmium. If this is the case, they should be submerged in 2 N HNO_3 or 2 N HCl for several days, and later rinsed with an excess of deionizing water.

It is essential to check the purity of the acids and the quality of the reagents used for conservation, extraction and digestion. If excessive concentrations of metal are detected, the acids must be purified by distillation or ultrapure acids should be used. Tests with blanks should be carried out at all digestion and filtration stages, and the necessary corrections should be applied to the results.

5 SAMPLE PREPARATION FOR SOLID WASTE ANALYSIS

When selecting a good method of chemical decomposition for elemental analysis, the capacity of the method to completely digest all the elements of interest, without loss of any volatile species, should be borne in mind. Most of the analytical techniques used for quantitative inorganic elemental analysis require that the sample is in aqueous solution, since better homogeneity is achieved in this way. But the process of putting a material into solution is often the most critical part in the analysis, since there are many sources of potential errors, for example partial digestion of the analytes present, or some type of contamination from the containers or chemical products used. Some analyses don't require the complete digestion of the sample, for example the determination of recoverable elements from sediments.[49,50] However, on most occasions the total concentration of the element in solution needs to be determined. Therefore systems of open digestion have traditionally been used. The use of aqua regia and mixtures of acids such as $HClO_4$, HNO_3, HCl and HF are considered the most efficient methods for total elemental digestion.[51] But due to the high risk associated with using $HClO_4$ in contact with organic matter, some authors have studied the use of H_2O_2 in its place.[52,53] Decomposition with mixtures of $HClO_4$, HNO_3 and HCl has proved to be very effective for the total digestion of selenium,[54] selenium and arsenic,[55] and selenium and tellurium.[13] Other authors have carried out the digestion of arsenic, selenium and antimony using less energetic procedures, digesting the samples with concentrated HCl,[56,57] and digesting mercury with HNO_3 and then with HCl.[58] Horvat et al.[59] show the advantage of distillation, rather than digestion with hydrochloric acid and alkalis, for the preparation of samples for the determination of methyl mercury. Sample preparation techniques by means of fusion are considered to have little applicability, mainly because the process is slow, and there is a risk of contamination. Smeller[60] did a comparative study of digestion with hydrochloric acid and fusion with lithium metaborate, and considers it more efficient to use the fusion process.

Pressure digestion systems have been used with some success in the preparation of soils and sediments for the elemental determination of trace metals.[61,62] They have also been used in the preparation of these materials, before total determination of critical elements such as mercury,[63] platinum,[64] lead[65] and cadmium.[66] Takamatsu[67] proposed the use of a high-pressure double bomb for the quantitative recovery of total S from sediments. Zhang and Shan[68] compared the effectiveness of a pressure digestion system and a microwave

system in the decomposition of soils for the determination of rare earths by ICPMS. In the last few years the use of microwave digestion systems has offered the advantages of rapid preparation of samples and reduction of contamination risks.[69,70] This technique, which is the same as pressure systems with bombs, has been used in the digestion of marine sediments,[71] and also in soils and sediments in agriculture,[72] for the elemental determination of trace metals. Good results for the determination of total mercury were obtained by Ombaba,[73] Murphy et al.[74] and Kumar and Meeravali[75] when applying this digestion technique to environmental materials. Morrow et al.[76] used this digestion procedure for the preparation of sediment samples, before the simultaneous determination of Sb, As, Bi, Ge, Se and Te by inductively coupled plasma atomic emission spectrometry (ICPAES). Matusiewicz et al.[77] used it for the preparation of marine sediments and later determination of cadmium.

The determination of the total concentration of the elements present in sediments doesn't give enough information about the mobility of metals. For this reason it becomes necessary to analyze the metals associated with a particular sediment phase. Sequential extraction offers useful information on how the metals are bound to sediments. Many of the outlines of sequential extractions are based on a five-stage protocol, as suggested by Tessier et al.[78] Thomas et al.[79] used a three-stage sequential extraction procedure on two river sediments, following the protocols proposed by the BCR. Whalley and Grant[80] recommend acidifying the extracts after separation, since they detected significant differences in the extracts from calcium carbonate extraction, potassium feldspar, ferrihydrite and humic acid in samples analyzed before and after acidification. However, sequential extraction methods have some detractors. Several authors have criticized the wide use that these procedures have had, because reabsorption effects have been reported during the extraction,[81] and because of lack of geochemical specificity.[82]

5.1 Open Wet Digestion

The acid digestion procedures have advantages over the fusion methods, since the fusion reagents cannot be obtained in such pure forms as the acids, and also an excess of acid in the samples can often be eliminated by evaporation. The total digestion methods are based on oxidation of the matrix by nitric, perchloric and sulphuric acid mixtures. This digestion technique is recommended for many environmental samples, but the danger involved in the use of perchloric acid is one of the biggest disadvantages. Another disadvantage of these digestion methods is the high consumption of chemicals and the possibility of loss or contamination of certain elements. The loss of elements such as B, As, Be and Sb in the form of volatile compounds is a limitation in this type of acid digestion when hydrofluoric acid is used. When selecting an appropriate procedure there is a temptation to dilute the original sample as little as possible, so that the maximum number of trace elements can be determined. Unfortunately this can mean that the concentration of dissolved solids reaches values of 1–2%, which can cause physical problems with the injection of samples in some spectroscopic techniques. Fortunately, several techniques now exist that allow the direct determination of many trace elements in environmental materials, even when the digestion procedure involves a dilution of a hundred times or more.

5.2 Dry Ashing

Dry ashing is one of the most common techniques used for destroying organic or biological matter. In this procedure, the sample is put in a crucible (made of quartz, porcelain or platinum), and heated inside an oven muffle until all the organic material is burnt. Finally the small inorganic residue is used in the analysis. Atmospheric oxygen acts as an oxidizing agent. The normal result of the ashing process is an inorganic residue of metallic oxides, sulfates, silicates and so on. However, as the purpose is to determine the concentration of the elements, an appropriate digestion process is necessary to put the elements into solution. It is then possible to consider a soft digestion with inorganic acid or a vigorous attack with mixtures of acids. An alternative to these ashing processes involves passing a low pressure flow of oxygen (approximately 150–700 Pa) through a high frequency electric field to produce excited oxygen. This oxygen flows over the sample inside a quartz tube, oxidizing the sample at temperatures less than 200 °C. The use of low temperatures has the advantage that losses due to volatilization decrease significantly, as do losses due to reactions with the container. One of the problems with this procedure is that the oxygen acts mainly on the surface of the sample. This makes periodic agitation necessary, which causes the process to stop, thus making it quite slow and tedious. Microwave muffle furnaces offer big advantages in the use of this technique, since they reduce the ashing times to a few minutes. They also improve the environmental conditions in the laboratory, since they don't produce odors and their internal capacities are also quite large. The combination of high heating by the microwaves and the concentration of air (and therefore O_2) that flows through the porous walls of the muffle makes this equipment an excellent alternative to the ashing processes.

5.3 Wet Ashing Techniques

Wet ashing processes involve heating the sample (containing organic matter) in the presence of an acid mineral oxidizer or mixtures of acids. When the acids are sufficiently strong oxidizers, and the sample is heated vigorously for a while, it is possible to completely oxidize the sample, leaving several of the elements present in the acid solution for analysis.

Wet ashing using concentrated acids (H_2SO_4, HNO_3 or $HClO_4$), or a mixture of them, is a common and important method for the decomposition of samples containing organic matter. This procedure is particularly useful for the determination of trace metals in a great variety of samples, because many of the elements are transformed into nonvolatile inorganic cations, which remain in the acid solution. Wet ashing can be used for nitrogen, phosphorous and sulfur determination, among the nonmetallic elements, but some elements are totally or partially lost through volatilization; these include mercury, arsenic, selenium, boron, and antimony.

It should be noted that $HClO_4$ is highly dangerous when in contact with organic matter, since explosions can occur (see section 3.7). For this reason it is recommended that the organic matter is first oxidized with HNO_3 concentrate. Finally, leave the mixture at room temperature and add $HClO_4$ (preferably mixed with HNO_3) carefully and reheat until complete oxidation. In this form the use of this acid is safer.

5.4 Pressurized Digestion

The advantages of pressure decomposition techniques for the digestion of environmental samples are very important when the quantitative analysis of trace elements is required. Pressure digestion procedures are preferred to traditional methods in most of the laboratories that determine elements in environmental samples. Good reproducibility, very low risk of contamination and no loss of analytes are the main advantages of this system.

One of the most frequently used methods of decomposition over the last few years has been wet chemical decomposition with nitric acid in Teflon® pressure bombs. This digestion has the advantage that many substances which are inert to attack by nitric acid at low temperatures can be attacked successfully at high temperatures while avoiding loss of volatile elements. However, decomposition temperatures that can be reached by this method are limited by the decomposition temperature of the PTFE container. The container volume also limits decomposition in these techniques, since only small quantities of samples can be digested. In some cases the PTFE containers can absorb elements from the solution, and also release elements into the solution, and as a consequence loss or contamination can happen. There is a pressure digestion system that involves the use of quartz containers, thus avoiding damage due to mechanical problems that can take place in PTFE containers when working at high temperatures.

5.5 Microwave Digestion

The microwave digestion system has proven to be a very useful tool, since this technique has the advantage that because the sample can be prepared quickly, the risks of contamination are reduced. The digestion times decrease, because of the energy generated as a consequence of the action of the microwaves on the acid. Contrary to traditional heating, where the heat is introduced from the outside by convection and radiation, heat from microwaves is generated in the core of the solution, and spreads from the inside towards the outside. This system causes rapid heating of the solution, which can reach higher temperatures than the boiling point of the acid. The activity of the microwaves also causes mechanical agitation, facilitating the disintegration of the sample matrix. Digestions at high temperature and pressure provide complete mineralization of the material. Digestion in closed systems ensures complete recovery of the elements being studied. Some difficulties occur in the digestion of materials containing silicates,[69] and for this reason it is desirable that they are treated with HF for better recovery.

Microwave digestion of organic compounds presents problems due to rapid oxidation, which can generate exothermic reactions that are difficult to control. Microwave units now exist in which digestion of the organic matter can be started with a 250 W ''unpulsed'' program for a few minutes, thus beginning a soft oxidation of the sample and preventing an uncontrollable exothermic reaction. Another microwave digestion procedure consists of an open system in which the heating is generated in the center of the solution, as in a traditional microwave system. This system can be used in digestions with low volatility and high boiling point acids (H_2SO_4, $HClO_4$, and H_3PO_4).

5.6 Sequential Extraction

Sequential extraction systems are used to determine the different forms of metals in sediments. A sediment fraction of particle size less than 63 μm is subjected to three extraction stages, according to the protocol set down by the BCR.[79] In the first stage a solution of acetic acid is added to the sediment and it is leached with agitation for a period of 16 h. In a second stage, a solution of hydroxylammonium chloride is added to the washed residue, and later the same extraction procedure is reapplied. In the final stage the residue is subjected

to digestion with hydrogen peroxide, and the sample is almost taken to dryness. The residue is extracted with ammonium acetate, using a similar method to the first two stages.

5.7 Sources of Error

In general, during the sample analysis process, the main source of error is the digestion stage, since this is when the sample is usually subjected to rigorous transformations. When the concentrations of the analyte to be determined are more than $100\,\mu g\,g^{-1}$, contamination in routine operations doesn't greatly influence the analytical results. But these contaminations become more significant when trace element analysis is required, since the use of small quantities of reagents during the decomposition process can have a great effect on the final trace element concentration.

Fusion procedures are generally more efficient than acid digestions, but fusion reagents can increase the concentrations of the dissolved solids to inconveniently high levels. Also, these reagents can introduce important quantities of impurities, increasing the blank levels (and thus the limits of detection) for the trace elements. The loss of some elements due to volatilization can be severe, due to the high temperatures of the calcination processes. However, the problems of loss of volatile elements affect non-metallic elements more than metals. The elements most prone to loss by volatilization are the halogens, S, Se, P, As, Sb, Ge, Tl, and Hg. Other causes of error can be reaction of the crucible during calcination, as well as absorption of some analytes on the crucible walls. These problems can be largely avoided by using platinum crucibles. To reduce negative effects during the calcination procedure, the sample should be mixed with some material that helps oxidation, prevents loss due to volatilization or prevents reactions with the walls of the crucible. The better known materials for this purpose are mixtures containing sulfuric acid, which form nonvolatile sulfates, thus reducing losses due to volatilization. Nitric acid and nitrates of alkaline metals act as oxidizing agents, forming metallic oxides. Hydroxides and carbonates form nonvolatile compounds with nonmetallics, preventing the loss of some elements.

6 ENVIRONMENTAL REGULATION AND METHODS

The monitoring systems require environmental measurements to determine the water quality and the type of waste effluent. The USEPA carries out continuous investigations of new technologies to improve the analytical methods for chemical characterization. Good sensitivity, low cost, reliability, robustness, good precision and accuracy are the main aspects that a new technology should have for the chemical characterization of samples coming from a particular geographic area. For this reason the USEPA has developed and distributed analytical methods to measure the chemical and physical parameters that affect the quality of the environment, including pollutants that can potentially have adverse effects on health. The USEPA has carried out collaborative studies with other laboratories to validate the analytical methods.[83]

The analytical methods published by the USEPA are written as stand-alone documents, so that each method can be removed from the manual, photocopied, inserted in another portfolio, and used without loss of information. The revisions of these methods are usually available in similar formats to facilitate substitution of existing methods as new techniques are developed. This flexibility avoids duplication of material. The volume added to the manual is also small in comparison with the whole. An important aspect of the methods is the consistent use of terminology, which is especially useful in the quality control section. The meanings of the terms are carefully selected, and the definition list is small and consequently easy to understand and to use. The names of the authors of the methods are provided to help users obtain direct support when required. The methods don't seek to be USEPA regulations, or obligatory for monitoring programs or specific studies. The quality control sections are uniform and they contain the minimum requirements to develop a reliable monitoring program. It is recommended that different practices are used for quality control, such as routine analysis of reagent blanks, analysis of fortified reagent blanks and sample solutions with standard additions, and analysis of quality control samples.

looseness=1 Some authors have used these methods to determine mercury in groundwater and drinking water samples[84] and in drinking water, wastewater, and sediments.[85] In both studies USEPA method 245.1A was used for digestion of the samples. Pyle and Nocerino[86] carried out a combined study with the USEPA to compare AAS, ICPAES, potentiometric stripping analysis (PSA) and X-ray fluorescence (XRF) spectrometry, in the determination of cadmium and lead in soil. The samples were digested using USEPA method 3050. Skip Kingston[87] showed the advantages of using microwave preparation systems using USEPA method 3052, which is considered to be the most flexible and the most used inside the United States. The publication also presents a comparative table of several methods developed by different organizations.

Table 4 summarizes the different methods used in the monitoring programs.

Table 4 Summary of USEPA sample preparation methods

USEPA digestion procedure	Method number	Analytes	Matrixes	Comments
Unfiltered or filtered sample is heated with diluted HCl and HNO_3	3005	Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Tl, V, Zn	Groundwater, surface water	Total recoverable and dissolved metal by FLAAS, ICPAES, ICPMS
Vigorous digestion with HNO_3 and dilution with HCl	3010	Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Tl, V, Zn	Waste sample	Total recoverable metal by FLAAS, ICPAES, ICPMS
Digestion in Teflon® vessel with HNO_3 and heating in microwave unit	3015	Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Tl, V, Zn	Water, wastewater	Total recoverable metal by FLAAS, GFAAS, ICPAES, ICPMS
Vigorous digestion with HNO_3 and dilution with HNO_3	3020	Ag, Al, As, Ba, Be, Cd, Cr, Co, Cu, Mn, Ni, Pb, Sb, Tl, Zn	Waste sample	Total recoverable metals by GFAAS or ICPMS
Vigorous digestion with HNO_3, H_2SO_4, HCl and $KMnO_4$	3031	Ag, As, Ba, Be, Cd, Co, Cr, Cu, Mo, Ni, Pb, Sb, Se, Tl, V, Zn	Waste oils, oil sludge and other viscous petroleum	Determination by FLAAS, GFAAS, ICPAES
Vigorous digestion with HNO_3 and H_2O_2, and dilution with HNO_3 or HCl	3050	Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Tl, V, Zn	Soil, sludge, solid waste	Total recoverable metals by FLAAS and ICPAES, or GFAAS and ICPMS
Digestion in Teflon® vessel with HNO_3 and heating in microwave unit	3051	Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Tl, V, Zn	Sludge, soil, sediment and oils	Total recoverable metals by FLAAS, GFAAS, ICPAES or ICPMS
Sample decomposition with HNO_3 and HF in a microwave unit	3052	Al, Ag, As, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Tl, V, Zn	Fly ash, oil, oil-contaminated soil, sediment, sludge, soil	Total analysis by FLAAS, GFAAS, ICPAES, ICPMS
Aqueous sample Add HNO_3 and HCl and evaporate to small volume Solid sample Digestion with HNO_3 and HCl (30 min)	200.2	Al, Ag, As, B, Ba, Be, Ca, Cd, Co, Cu, Cr, Fe, Hg, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, Sb, Se, Si, Sr, Th, Ti, Tl, U, V, Zn	Groundwaters, surface waters, drinking waters, wastewater, sediment, sludge and solid waste	Total recoverable elements with exception of silica in sediment, sludge and solid waste. Analysis by GFAAS, ICPAES or ICPMS

FLAAS, flame atomic absorption spectrometry; GFAAS, graphite furnace atomic absorption spectrometry.

7 QUALITY CONTROL

Information obtained in a chemical analysis laboratory serves as a basis for scientific, technological, administrative and economic decisions, such that the results are recognized and accepted and provide a firm foundation for making the correct decision. At the moment, information regarding the quality of the environment, foods, pharmaceutical and industrial chemicals, etc. is in demand in many countries. Accuracy, which is the degree of agreement between the result and the true value, and representativeness, which is the degree of agreement between the collected sample and the definition of the problem, are the two properties that unequivocally define the quality of the results. The factors that affect the representativeness depend on the sampling, an activity that in general is unrelated to the work of the laboratory. However, accuracy, which should have an appropriate degree of precision, represents the quality of the analytical work.[88,89] Work that is carried out in the laboratory can be divided into three stages. The first one corresponds to the chemical treatment, and involves the preliminary preparation of the sample, digestion, dilution, separation procedures, etc. The second involves measuring the analytical signs, that is the use of an instrument that generates this information. The last corresponds to the acquisition and treatment of data. The quality of the results, which is the most important aspect, depends on

each one of these stages, and the effect that each phase has on the final result cannot be minimized. Nevertheless, the first stage, because of the variety of operations involved, is traditionally the one that introduces the greatest degree of uncertainty in the final result.[2]

One of the main factors that should be kept in mind during the chemical preparation stage involves the pollutants that can be introduced by the containers (especially in the digestion process), the chemists, the circulation of air inside the fume hoods, etc.

Other factors are related to the behaviour of the sample during this process (loss of elements by evaporation, co-precipitation, digestion of refractory elements, etc.). The first of these factors is controlled by the use of reagent blanks and the second using CRM. As with the reagent blanks, the CRM should undergo the whole process with the sample. A great variety of these certified materials exist, in different matrix types, which can be used to check the behavior of the procedures used. Among the materials most used by different authors are those related to seawaters,[35,46] river waters,[90] and soil and sediment.[11,62] The main suppliers of these materials are: NIST, BCR, Laboratory of the Government Chemist (LGC), Laboratoire National d'Essais (LNE) and the National Research Council of Canada (NRCC).

Many laboratories have produced this type of material.

7.1 Instrument Calibration

The calibration of instruments and apparatus should be a regular activity in all laboratories that work under a quality program. The frequency of calibration will depend on the nature of the instrument, its robustness, amount of use, and environment. It also depends on the demands of the analytical methodology used. The calibration programs should be designed to fit in with the periods of use of the instruments and the periods of time during which they are not in service. Instruments that are used infrequently or those that are particularly sensitive should be standardized every time that they are used. Others will only need periodic calibration, unless there are special circumstances.

Because of the complexity of modern instruments, the design of maintenance programs has placed demands on the laboratories. Laboratory personnel now have difficulty repairing instruments, and normally rely on the technical departments of the makers or their representatives, who offer maintenance contracts that incorporate calibration and validation of the instruments within their quality guarantees.

7.2 The Blank

To obtain a good analytical result, it is advisable that a blank undergoes the whole chemical analysis procedure, along with the samples.[91] This blank should include all the reagents that are used in the preparation of the samples, since many metals are present in the acids and the glassware in significant quantities that can influence the results. If the measurements carried out with the blank give higher concentrations than the detection limit of the method, preparation of the sample should be repeated using purer reagents and/or cleaner glassware. For the analysis of metals it is necessary to check that the glassware and other materials are especially clean. Experience has shown that a blank made with the same acids and subjected to the same digestion procedure as the sample can correct for impurities present in the acids, in the reagent solution or in the glassware.

7.3 Use of Reference Materials

Good accuracy is one of the fundamental premises in all chemical analysis, but the problem is knowing with reliability the true value. The use of CRM is one of the mechanisms most used by the laboratories to determine if the results obtained are near to the true values. These materials should satisfy the requirements of homogeneity, accuracy, traceability and stability, as well as such requirements as similarity with the sample matrix and precision. But frequently it is impossible to fulfill all of these requirements. Homogeneity is essential in these materials, to ensure that there are no differences from the certified value either within a single flask or between different flasks. The same situation applies with accuracy and traceability, since these materials are used to verify the analytical results or instrument calibrations. In the case of stability, in spite of being a basic requirement it is not always possible to guarantee stability in all sample types. Also in some cases these materials are sensitive to changes in temperature, intense light, humidity, etc. The value that should be most similar to that of the sample for analysis is the matrix of the reference material, a situation that can be fulfilled quite easily in the case of metals and minerals, but is more complicated in the case of environmental samples, where the analysis of pollutants is required. The certified values of a CRM should include the degree of precision with which it was obtained, which must be as large as possible. CRM are high-cost products, which limits their frequency of use inside a laboratory. For this reason it is advisable to prepare some type of control material inside the laboratory, which must have similar characteristics to the samples to be analyzed. These could be prepared from the samples after some of them had been analyzed several times and the two most similar methods chosen. These materials should be obtained systematically for each batch of samples to be analyzed.

8 DEVELOPMENT OF THE METHOD

The results generated by a chemical analysis laboratory, as mentioned in the previous section, should serve as a basis for scientific, technological or administrative decisions. Thus it is essential in the development of an analysis method that the final results have a high degree of reliability. This requires that a strict control is maintained of all the stages that the sample is subjected to during the analytical process. The development of an analysis method should fit within the laboratory quality system, so that all the information that is generated has the quality level that has previously been agreed upon. It is necessary to make a distinction between the quality of the work that is carried out and the quality of the results that are generated. This last meaning is most often used to define quality in analytical laboratories.

To generate reliable results it is necessary to use a CRM reference that ensures that the result obtained is traceable. Traceability[92] implies a complete sequence of comparisons, so that any stage that has a direct or indirect relationship with a measurement that will be used to generate an analytical result can be subtracted. When development of a method of chemical analysis is planned, it should cover the field or sample type to which it will be applied. If the most important issue is the quality of the result, this can imply a slower analytical response. On the other hand, if the important issue is the speed of the results, this can lead to deterioration in accuracy and precision. It should not be forgotten that the quality of the results has two essential components: accuracy and representativeness. This is because, however good the accuracy and precision of a result, if it is not representative of the original material the result won't have any value. Only the correct application of a good method can guarantee the quality of the results, which shows the need to maintain the quality control and evaluation systems.

8.1 Writing Up of the Method

To write up an analytical method it is necessary to follow a strict sequence, to avoid causing confusion to other chemists. Jargon should not be used, and if abbreviations are used then they should be defined. The writing should be clear and unambiguous. It is necessary to indicate the critical stages in the analysis and the consequences if they are not followed. The sentences should be short and clearly edited. Different formats have been used to write up methods, but they all, up to a point, cover the same sections. The sections, and their contents, are as follows.

- Method number: a number that uniquely identifies each method will facilitate traceability.
- Title: the title should be brief, and should contain the name of the analyte and the matrix.
- Scope and application: this indicates the matrix type to which the method will be applied, and the technical or instrumental procedures that could be used to determine the elements.
- Summary of method: a very brief summary of the procedure is useful.
- Definitions: acronyms or abbreviations that are used to identify solutions or processes should be defined.
- Interference: elements that could enter the sample solution, as well as materials or apparatus that could potentially cause contamination, are indicated.
- Apparatus and equipment: glassware, plastics, etc. that will be used are listed, plus the precautions that it is necessary to take with some of them are pointed out.
- Cautions and security: the risks involved in the manipulation of certain samples and the use of certain apparatus and reagents are pointed out. The precautions that it is necessary to take are indicated.
- Reagents and materials: the quality of reagents, prepared solutions, water and reference materials that will be used during the analysis process are indicated in some detail.
- Sample preservation and storage: sample conservation systems are indicated, and details of how samples will be stored until analysis are given.
- Calibration and standardization: details of which instruments require calibration, and how the calibration will be carried out, are recorded in this section.
- Quality control: the quality control program that will be applied to the samples is indicated.
- Procedure: the procedure used to obtain a solution suitable for analysis by some instrumental procedure is indicated in detail.
- References: bibliographical accounts of the experimental method, and also details of the validation studies, are indicated.

ACKNOWLEDGMENTS

The author would like to thank Cecilia Navarrete, librarian of the Nuclear Centre Lo Aguirre, for her valuable cooperation in the location of important bibliographic materials. He is also very grateful to Oscar Arlegui, of the Laboratory of Chemical Analysis of the Nuclear Centre Lo Aguirre, for his invaluable contributions to section 4.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
ACS	American Chemical Society
ASTM	American Society for Testing and Materials
BCR	Bureau Communautaire de Réference
BOD	Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
CRM	Certified Reference Materials
DIN	Deutsches Institut für Normung
DOE	Department of Energy
FEP	Fore Edges Painted
FLAAS	Flame Atomic Absorption Spectrometry
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HEPA	High-efficiency Particulate Air
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ISO	International Standardization Organization
LGC	Laboratory of the Government Chemist
LNE	Laboratoire National d'Essais
NIST	National Institute of Standards and Technology
NRCC	National Research Council of Canada
PSA	Potentiometric Stripping Analysis
PTFE	Polytetrafluoroethylene
PVC	Polyvinyl Chloride
TFE	Tetrafluoroethylene
USEPA	United States Environmental Protection Agency
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in

REFERENCES

1. *Standard Methods for the Examination of Water and Wastewater*, 19th edition, American Public Health Association, American Water Works Association, Water Pollution Control Federation, Washington, 1/18–1/24, 1995.
2. N.M. Kuz'min, 'Sample Preparation in Environmental Analysis', *J. Anal. Chem.*, **51**(2), 186–193 (1994).
3. J. Startin, R.V. Lansdown, 'Drainage from Highways and Other Paved Areas: Methods of Collection, Disposal and Treatment', *J. Inst. Water Environ. Manage.*, **8**(5), 518–526 (1994).
4. H.D. Robinson, M.J. Barr, S.D. Last, 'Leachate Collection, Treatment and Disposal', *J. Inst. Water Environ. Manage.*, **6**(3), 321–332 (1992).
5. W. Blommaert, P. Ferrari, J.P. Culot, C. Urtgen, 'Problems Concerning Representative Sampling for Possible Contamination in Complex Waste Processing Facility', *J. Radioanal. Nucl. Chem.*, **197**(2), 253–269 (1995).
6. N.B. Lohani, S.M. Ko, 'Optimal Sampling of Domestic Waste', *J. Environ. Eng.*, **114**(6), 1479–1483 (1988).
7. M.B. Wise, M.R. Guerin, 'Direct Sampling MS for Environmental Screening', *Anal. Chem.*, **69**(1), 26A–32A (1997).
8. K.B. Olsen, J. Wang, R. Setladlj, J. Lu, 'Field Screening of Chromium, Zinc, Copper, and Lead in Sediments by Stripping Analysis', *Environ. Sci. Technol.*, **28**(12), 2074–2079 (1994).
9. G. Scarponi, G. Capodaglio, P. Cescon, 'Anodic Stripping Voltametric Determination of the Contamination of Seawater Samples by Cadmium, Lead, and Copper During Filtration and Storage', *Anal. Chim. Acta*, **135**, 263–276 (1982).
10. C. Locatelli, T. Garai, F. Fagioli, 'Critical Comparison of Alternating Current and Differential Pulse Voltametry in the Determination of Heavy Metals in Seawater', *Mikrochim. Acta*, **107**(1–2), 41–48 (1992).
11. C. Capodaglio, G. Scarponi, G. Toscano, C. Barbante, P. Cescon, 'Speciation of Trace Metals in Seawater by Anodic Stripping Voltametry: Critical Analytical Steps', *Fresenius' J. Anal. Chem.*, **351**(4–5), 386–392 (1995).
12. L. Starostin, Z. Witkiewicz, 'Environmental Water Sample Preparation for Chemical Analysis', *Chem. Anal.*, **39**(3), 263–279 (1994).
13. A. Dulivo, 'Determination of Selenium and Tellurium in Environmental Samples', *Analyst*, **122**(12), 117R–144R (1997).

14. B.C. Rhoades, 'Clean Laboratory Chemistry for the Microwave-assisted Digestion of Botanical Samples', *J. Anal. At. Spectrom.*, **11**(9), 751–757 (1996).
15. P. Kanatharana, W. Sirinawin, R.R. Ohpas, 'Establishing a Clean Room from an Ordinary Chemistry Laboratory', *J. Environ. Sci. Health*, **28**(10), 2315–2321 (1993).
16. E.D. Suttie, E.W. Wolff, 'Preconcentration Method for Electrothermal Atomic Absorption Spectrometric Analysis for Heavy Metals in Antarctic Snow at Sub ng Kg^{-1} Levels', *Anal. Chim. Acta*, **258**(2), 229–236 (1992).
17. J.P. Candelone, S. Hong, C.F. Boutron, 'An Improved Method for Decontaminating Polar Snow for Ice Cores for Heavy Metal Analysis', *Anal. Chim. Acta*, **299**(1), 9–16 (1994).
18. U. Görlach, C.F. Boutron, 'Preconcentration of Lead, Cadmium, Copper and Zinc in Water at the pg g^{-1} Level Non-boiling Evaporation', *Anal. Chim. Acta*, **236**(2), 391–398 (1990).
19. M.A. Bolshov, C.F. Boutron, F.M. Ducroz, 'Direct Ultratrace Determination of Cadmium in Antarctic and Greenland Snow and Ice by Laser Atomic Fluorescence', *Anal. Chim. Acta*, **251**(1–2), 169–175 (1991).
20. C.F. Boutron, 'A Clean Laboratory for Ultralow Concentration Heavy Metal Analysis', *Fresenius' J. Anal. Chem.*, **337**(5), 482–491 (1990).
21. T.L. Lowry, 'Removing Silicone Grease from Glassware', *J. Chem. Educ.*, **74**(7), 841–841 (1997).
22. T.J. Reilly, 'Removing Silicone Greases from Round Bottom Flasks', *J. Chem. Educ.*, **73**(4), 381 (1996).
23. N. Narayanan, 'A Simple and Effective Method to Remove Grease from Glassware', *J. Chem. Educ.*, **69**(10), 849 (1992).
24. P.L. Manske, T.M. Stimpfel, 'Safety in the Chemical Laboratory', *J. Chem. Educ.*, **67**(11), A280–A282 (1990).
25. *Standard Methods for the Examination of Water and Wastewater*, 19th edition, American Public Health Association, American Water Works Association, Water Pollution Control Federation, Washington, 1/27–1/44, 1995.
26. *Security: Manual for the Laboratory*, ed. D. Bernabei, Merck, Darmstadt, Germany, 1994.
27. S.R. Marìn, S.G. Olave, O.E. Andonie, O.G. Arlegui, 'Comparison of Analytical Techniques for the Determination of Trace Elements in Rain Water', *Int. J. Environ. Anal. Chem.*, **52**, 127–136 (1993).
28. S. Cornejo, S. Marìn, S. Olave, C. Urrutia, 'Determination of Trace Elements in Aerosol Samples Collected on Polycarbonate Filters by Atomic Absorption Spectrometry', *Int. J. Environ. Anal. Chem.*, **60**, 377–387 (1995).
29. M. Okumura, Y. Seike, K. Fujinaga, K. Hirao, 'In Situ Preconcentration Method for Iron(II) in Environmental Water Sample Using Solid Phase Extraction Followed by Spectrophotometric Determination', *Anal. Sci.*, **13**(2), 231–235 (1997).
30. A. van Green, E. Boyle, 'Automated Preconcentration of Trace Metals from Seawater and Freshwater', *Anal. Chem.*, **62**(15), 1705–1709 (1990).
31. H. Zou, S. Xu, Z. Fang, 'Determination of Chromium in Environmental Samples by Flame AAS with Flow Injection On-line Coprecipitation', *At. Spectrosc.*, **17**(3), 112–118 (1996).
32. L. Fangruo, H. Xiandeng, 'Determination of Sub-ppb Levels of Thallium in Natural Water by STPF AASS after Preconcentration Using a Polyurethane Plastic Foam Column', *At. Spectrosc.*, **15**(5), 216–217 (1994).
33. M. Chikuma, H. Aoki, 'Preconcentration of Lead(II) in Environmental Waters Using a Chelating Resin and Subsequent Determination as a Resin–Water Suspension by Hydride Generation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **8**(3), 415–417 (1993).
34. A. Adachi, K. Ogawa, Y. Tsushi, N. Nagao, T. Kobayashi, 'Determination of Vanadium in Environmental Sample by Atomic Absorption Spectrometry', *Water Res.*, **31**(5), 1247–1250 (1997).
35. B. Paull, M. Foulkes, P. Jones, 'High-performance Chelation Ion Chromatographic Determination of Trace Metals in Coastal Seawater Using Dye-impregnated Resins', *Analyst*, **119**(5), 937–941 (1994).
36. M.J. Bloxham, S.J. Hill, P.J. Worsfold, 'Determination of Trace Metals in Sea-water and the On-line Removal of Matrix Interferences by Injection with Inductively Coupled Plasma Mass Spectrometric Detection', *J. Anal. At. Spectrom.*, **9**(9), 935–938 (1994).
37. G. Chapple, J.P. Byrne, 'Direct Determination of Trace Metals in Seawater Using Elecrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**(8), 549–553 (1996).
38. D. Tanzer, K.G. Heumann, 'Determination of Dissolved Selenium Species in Environmental Water Samples Using Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **63**(18), 1984–1989 (1991).
39. M. Horvat, L. Liang, N.S. Bloom, 'Comparison of Distillation with Other Current Isolation Methods for the Determination of Methyl Mercury Compounds', *Anal. Chim. Acta*, **282**(1), 153–168 (1993).
40. *Standard Methods for the Examination of Water and Wastewater*, 19th edition, American Public Health Association, American Water Works Association, Water Pollution Control Federation, Washington, 3/2–3/8, 1995.
41. F. Wang, S. Li, S. Liu, Y. Zhang, Z. Liu, 'Application of Differential Pulse Anodic Stripping Voltammetry with Stopped-flow System for the Simultaneous Determination of Pb, Cu, Sb, and Bi, in Environmental Samples', *Anal. Lett.*, **27**(9), 1779–1787 (1994).
42. L. Balint, I. Vedrina-Dragojevic, B. Sebecic, J. Momirovic-Culjat, M. Horvativ, 'Spectrofluorimetric Method for Determination of the Total Mercury Content in Environmental Samples – Waste Waters', *Mikrochim. Acta*, **127**(1–2), 61–65 (1997).

43. J. Pérez-Peña, J.J. Hernández-Brito, J.A. Herrera-Melián, C. Collados-Sánchez, C.M.G. Van den Berg, 'High-performance Adsorptive Cathodic Stripping Voltammetry of Nickel and Cobalt in Seawater', *Electroanalysis*, **6**, 1069–1076 (1994).
44. Z.S. Liu, S.D. Huang, 'Automatic On-line Preconcentration System for Graphite Furnace Atomic Absorption Spectrometry for the Determination of Trace Metals in Seawater', *Anal. Chim. Acta*, **281**(1), 185–190 (1993).
45. S. Sachsenberg, T. Klenke, W.E. Krumbein, H.J. Schellnhuber, E. Zeeck, 'Direct Graphite Furnace Atomic Absorption Spectrometric Determination of Metals in Seawater: Application of Palladium Modifiers and a Fractal Approach to their Analytical Support', *Anal. Chim. Acta*, **279**(2), 241–251 (1993).
46. I. Rodushkin, T. Ruth, 'Determination of Trace Metals in Estuarine and Seawater Reference Materials by High Resolution Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**(10), 1181–1185 (1997).
47. G.J. Batterham, N.C. Munksgaard, D.L. Parry, 'Determination of Trace Metals in Seawater by Inductively Coupled Plasma Mass Spectrometry After Off-line Dithiocarbamate Solvent Extraction', *J. Anal. At. Spectrom.*, **12**(11), 1277–1280 (1997).
48. S.N. Willie, Y. Lida, J.W. MacLaren, 'Determination of Cu, Ni, Zn, Mn, Co, Pb, Cd, and V in Seawater Using Flow Injection ICP–MS', *At. Spectrosc.*, **19**(3), 67–72 (1998).
49. T.W. May, R.H. Wiedmeyer, W.G. Brumbaugh, C.J. Schmitt, 'The Determination of Metals in Sediment Pore Waters and in 1 N HCl-extracted Sediments by ICP–MS', *At. Spectrosc.*, **18**(5), 133–139 (1997).
50. T.D. Martin, J.T. Creed, S.E. Long, 'Method 200.2 Sample Preparation Procedure for Spectrochemical Determination of Total Recoverable Elements', *ICP Information Newsletter*, **18**(3), 141–144 (1992).
51. U.R. Kackstaetter, G. Heinrichs, 'Validity of Lowcost Laboratory Geochemistry for Environmental Applications', *Water, Air, Soil Pollut.*, **95**(1/4), 119–131 (1997).
52. B. Kovacs, Z. Gyori, P.A. Daniel, 'A Study of Plant Sample Preparation and Inductively Coupled Plasma Emission Spectrometry Parameters', *Commun. Soil Sci. Plant Anal.*, **27**(5–8), 1177–1198 (1996).
53. P. Masson, A. Gomez, A. Tremel, 'Suitability of Open System Digestions for the Determination of Thallium in Environmental Samples', *Commun. Soil Sci. Plant Anal.*, **27**(1–2), 109–118 (1996).
54. P.M. Haygarth, A.P. Rowland, S. Sturup, K. Jones, 'Comparison of Instrumental Methods for the Determination of Total Selenium in Environmental Samples', *Analyst*, **118**(10), 1303–1308 (1993).
55. C.C.Y. Chan, R.S. Sadana, 'Determination of Arsenic and Selenium in Environmental Samples by Flow-injection Hydride Generation Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **270**(1), 231–238 (1992).
56. K.A. Anderson, B. Isaacs, 'Determination of Antimony in Environmental Samples by Hydride Generation–Inductively Coupled Plasma Spectrometry', *J. AOAC Int.*, **77**(6), 1562–1568 (1994).
57. K.A. Anderson, B. Isaacs, 'Simultaneous Determination of Arsenic, Selenium, and Antimony in Environmental Samples by Hydride Generation for Inductively Coupled Plasma Atomic Spectrometry', *J. AOAC Int.*, **78**(4), 1055–1060 (1995).
58. J. Yoshinaga, M. Morita, 'Determination of Mercury in Biological and Environmental Samples by Inductively Coupled Plasma Mass Spectrometry with the Isotope Dilution Technique', *J. Anal. At. Spectrom.*, **12**(4), 417–420 (1997).
59. M. Horvat, N.S. Bloom, L. Liang, 'Comparison of Distillation with Other Current Isolation Methods for the Determination of Methyl Mercury Compounds', *Anal. Chim. Acta*, **281**(1), 135–152 (1993).
60. J.M. Smeller, 'Comparison of Sample Preparation Methods for the Spectrophotometric Determination of Phosphorus in Soil and Coal Fly Ash', *Analyst*, **120**(1), 207–210 (1995).
61. J.W. McLaren, D. Beauchemin, S.S. Berman, 'Application of Dilution Inductively Coupled Plasma Mass Spectrometry to the Analysis of Marina Sediment', *Anal. Chem.*, **59**(4), 610–613 (1987).
62. J.W. McLaren, D. Beauchemin, S.S. Berman, 'Determination of Trace Metals in Marine Sediments by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **2**(4), 277–281 (1987).
63. M. Moskalova, M. Zemberyova, 'Determination of Total Mercury Content in Environmental Samples', *Chem. Pap.*, **51**(5), 273–275 (1997).
64. E. Beinrohr, M.L. Lee, P. Tschopel, G. Tolg, 'Determination of Platinum in Biotic and Environmental Samples by Graphite Furnace Atomic Absorption Spectrometry after its Electrodeposition into a Graphite Tube Packed with Reticulated Vitreous Carbon', *Fresenius' J. Anal. Chem.*, **346**(6–9), 689–692 (1993).
65. X.P. Yan, F. Adams, 'Flow Injection On-line Sorption Separation and Preconcentration with a Knotted Reactor for Electrothermal Atomic Absorption Spectrometric Determination of Lead in Biological and Environmental Samples', *J. Anal. At. Spectrom.*, **12**(4), 459–464 (1997).
66. R. Ma, W. Van Mol, F. Adams, 'Selective Flow-injection Sorbent Extraction for Determination of Cadmium, Copper and Lead in Biological and Environmental Samples by Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **293**(3), 251–260 (1994).
67. T. Takamatsu, 'A Convenient Determination Method of Total Sulfur in Environmental and Biological Samples', *Int. J. Environ. Anal. Chem.*, **62**(4), 303–310 (1996).
68. S. Zhang, X.-Q. Shan, 'The Determination of Rare Earth Elements in Soil by Inductively Coupled Plasma Mass Spectrometry', *At. Spectrosc.*, **18**(5), 140–144 (1997).

69. P. Quevauviller, J.L. Imbert, M. Ollé, 'Evaluation of the Use of Microwave Oven System for the Digestion of Environmental Samples', *Mikrochim. Acta*, **112**(1–4), 147–154 (1993).
70. N.M. Kuz'min, I.V. Kubrakova, 'Microwave Sample Preparation', *J. Anal. Chem.*, **51**(1), 38–41 (1996).
71. J.W. McLaren, D. Beauchemin, S.S. Berman, 'Analysis of the Marina Sediment Reference Materials PACS-1 by Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta, Part B*, **43**(4–5), 413–420 (1988).
72. R. Moral, J. Navarro-Pedreño, I. Gómez, J. Mataix, 'Quantitative Analysis of Organic Wastes: Effects of Sample Preparation in the Determination of Metals', *Commun. Soil Sci. Plant Anal.*, **27**(3–4), 753–761 (1996).
73. J.M. Ombaba, 'Total Mercury Determination in Biological and Environmental Standard Sample by Gold Amalgamation Followed by Cold Vapor Atomic Absorption Spectrometry', *Microchem. J.*, **53**(2), 195–200 (1996).
74. J. Murphy, P. Jones, S.J. Hill, 'Determination of Total Mercury in Environmental and Biological Samples by Flow Injection Cold Vapor Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **51**(14), 1867–1873 (1996).
75. S.J. Kumar, N.N. Meeravali, 'In-situ Trapping of Mercury Vapors on Au, Pd–Au Alloy or Pt–Rh Alloy in the Graphite Furnace for the Determination of Hg in Environmental Samples After Microwave Digestion', *At. Spectrosc.*, **18**(5), 166–168 (1997).
76. A. Morrow, G. Wiltshire, A. Hursthouse, 'An Improved Method for the Simultaneous Determination of Sb, As, Bi, Ge, Se, and Te by Hydride Generation ICP–AES: Application to Environmental Samples', *At. Spectrosc.*, **18**(1), 23–28 (1997).
77. H. Matusiewicz, M. Kopras, R.E. Sturgeon, 'Determination of Cadmium in Environmental Samples by Hydride Generation with In Situ Concentration and Atomic Absorption Detection', *Analyst*, **122**(4), 331–336 (1997).
78. A. Tessier, P.G.C. Campbell, M. Bisson, 'Sequential Extraction Procedure for the Speciation of Particulate Trace Metals', *Anal. Chem.*, **51**(7), 844–851 (1979).
79. R.P. Thomas, A.M. Ure, C.M. Davidson, D. Littlejohn, G. Rauret, R. Rubio, J.F. López-Sáncchez, 'Three-stage Sequential Extraction Procedure for the Determination of Metals in River Sediments', *Anal. Chim. Acta*, **286**(3), 423–429 (1994).
80. C. Whalley, A. Grant, 'Assessment of the Phase Selectivity of the European Community Bureau of Reference (BCR) Sequential Extraction Procedure for Metals in Sediment', *Anal. Chim. Acta*, **291**(3), 287–295 (1994).
81. E. Tipping, N.B. Hetherington, J. Hilton, D.W. Thompson, E. Bowles, J. Hamilton-Taylor, 'Artifacts in the Use of Selective Chemical Extraction to Determine Distributions of Metals Between Oxides of Manganese and Iron', *Anal. Chem.*, **57**(9), 1944–1946 (1985).
82. C. Kheboian, C.F. Bauer, 'Accuracy of Selective Extraction Procedure for Metal Speciation in Model Aquatic Sediments', *Anal. Chem.*, **59**(10), 1417–1423 (1987).
83. J.E. Longbottom, T.D. Martin, K.W. Edgell, S.E. Long, M.R. Plantz, B.E. Warden, 'Determination of Trace Elements in Water by Inductively Coupled Plasma–Mass Spectrometry: Collaborative Study', *J. AOAC Int.*, **77**(4), 1004–1023 (1993).
84. S. Mcintosh, J. Baasner, Z. Grosser, C. Hanna, 'Improving the Determination of Mercury in Environmental Samples Using a Dedicated Flow-injection Mercury System', *At. Spectrosc.*, **15**(4), 161–163 (1994).
85. C.P. Hanna, S.A. Mcintosh, 'Determination of Total Hg in Environmental Samples with Online Microwave Digestion Coupled to a Flow-injection Mercury System (FIMS)', *At. Spectrosc.*, **16**(3), 106–114 (1995).
86. S.M. Pyle, J.M. Nocerino, 'Comparison of AAS, ICP–AES, PSA, and XRF in Determining Lead and Cadmium in Soil', *Environ. Sci. Technol.*, **30**(1), 204–213 (1996).
87. H.M. Skip Kingston, 'Standardization of Sample Preparation for Trace Element Determination Through Microwave-enhanced Chemistry', *At. Spectrosc.*, **19**(2), 27–30 (1998).
88. M. Valcárcel, A. Rios, 'The Hierarchy and Relationships of Analytical Properties', *Anal. Chem.*, **65**, 781A–787A (1993).
89. A. Ríos, M. Valcárcel, 'Representativeness of Analytical Results', *Analyst*, **119**(1), 109–112 (1994).
90. D. Beuchemin, J.W. McLaren, A.P. Mykytiuk, S.S. Berman, 'Determination of Trace Metals in a River Water Reference Material by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **59**(5), 778–783 (1987).
91. J.K. Taylor, *Quality Assurance of Chemical Measurements*, Lewis Publishers, Chelsea, MI, USA, 123–128, 1987.
92. M. Valcárcel, A. Rios, 'Traceability in Analytical Chemistry', *Analyst*, **120**(8), 2292–2297 (1995).

Sampling Considerations for Biomonitoring

Tim Koeckritz, Olaf Wappelhorst, and Bernd Markert
International Graduate School Zittau, Germany

1 Introduction **1**
2 Demands on Sampling **2**
2.1 Axiom 1: Chemical Composition 2
3 Factors Involved in Heavy Metal Uptake by Organisms **2**
3.1 Abiotic Factors 2
3.2 Biotic Factors 3
4 Conduct of Sampling **9**
4.1 Choice of Organisms 9
4.2 Axiom 2: Distribution of Organisms 10
4.3 Axiom 3: Economic Aspects of Information Content 12
4.4 Influence of Size of the Organisms on Number of Samples 13
5 How to Carry Out a Biomonitoring Procedure **13**
5.1 On-site Sampling 13
5.2 Conclusions 15
Related Articles **15**
References **15**

This article looks at various factors involved in the first step in biomonitoring, namely the sampling process with special reference to heavy metal levels in living organisms and their interaction with terrestrial ecosystems. Because samples taken in the field naturally form the basis for the entire biomonitoring process, it is vitally important that the margin of error in this step is kept as low as absolutely possible. An important initial step is the decision about which organisms to use and where to gather them. It is important to consider factors such as chemical composition, relationship to the ecosystem involved and the number of individuals available, in addition to the abiotic and biotic factors which influence the amount of heavy metal the organism takes in from its environment, or its bioavailability (heavy metal accumulation is the practical example for biomonitoring used in this article).

How the samples are collected is as important as what is collected, and here various topics are covered, from choosing the most suitable organism(s), to deciding how many samples to take in order for them to be representative of the entire population at that site, to economic considerations. Although exact sampling methods vary according to the organisms involved and there are very few publications in this field, some basic guidelines are given to follow in any case, as well as references to the existing published guides.

1 INTRODUCTION

The damage caused to ecosystems by pollutants can be ascertained by analyzing substrate samples or by determining pollutant concentrations in organisms. An important advantage of using living organisms as opposed to other environmental compartments such as soil, water or air is that an organism exists in a particular ecosystem because of its specific requirements for survival and may therefore be considered representative of this ecosystem. In the context of monitoring programs, the pollutant concentrations found in organisms make it possible to draw conclusions about pollution at a particular location or in whole ecosystems.[1]

Since airborne and waterborne pollutants are highly subject to dilution by their environment, their occurrence is only temporary and therefore difficult to determine. But if an organism that lives in the ecosystem permanently is used, we may assume that determining its pollutant content will reveal an integrated overall burden at a particular location. Unlike inanimate material, however, living organisms constantly interact with their environment, with the result that their uptake of pollutants is not constant but influenced and sometimes dominated by abiotic and biotic factors. This means that the information on ecosystems yielded by the pollutant concentrations in organisms is at best limited unless the natural, ecological fluctuations are known.[2]

The interaction between an organism and its environment does not only result in fluctuations in pollutant levels over the course of time, it also affects the geographic distribution of the organisms, which may differ from the substrate-bound distribution of the pollutant at the site.

Before drawing conclusions about the overall burden on the ecosystem from the pollutant concentration found in an organism, it is therefore necessary to determine the variability of the organism's pollutant uptake in terms of time and space. As a first step in practical analytical work, sampling is of crucial importance since errors that occur at this stage continue throughout the process of data collection and cannot be corrected retrospectively.[2–10] Whether the information yielded by an environmental sample is enough to be representative of the pollution of an organism or population, a site or a whole ecosystem depends on how sampling is planned and conducted and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

its adjustment to the interactions of the organism or whole populations to be investigated.

2 DEMANDS ON SAMPLING

Since environmental monitoring can never be carried out continuously and thus all observations have to be expressed within a specified framework, any particular process can only be investigated by taking samples. In order to ensure that the monitoring method nevertheless makes the ecosystem as transparent as possible, the sample chosen must contain the greatest amount of information possible which, in turn, has to be applicable to and thus representative of the system as a whole. Like the information content, the degree to which a sample is representative is unknown and cannot be determined directly. But whereas the information content is not ascertained until the process of analysis and evaluation has begun, how representative a sample is, is defined at the beginning, during the planning of the experiment and during sampling.

The following axioms have to be followed to ensure that the samples taken are representative:[11]

- Axiom 1: The sample taken from the system should have exactly the same chemical composition as the original sample.
- Axiom 2: The probability of being removed from a total population must be the same for each individual.
- Axiom 3: The larger the degree of dispersion among individuals and the larger the number of individuals, the greater the amount of time and effort needed for sampling.

These three axioms for sampling may be regarded as a guideline for dealing with individual problems during the sampling procedure. They can be applied to a large number of questions that arise during monitoring programs.

2.1 Axiom 1: Chemical Composition

The demand made on sampling by Axiom 1 means that a certain number of samples must represent the average pollutant concentration of the overall system. But ''overall systems'' do not really exist in the ecological sense; they are in fact made up of an arbitrary number of individual systems that may be regarded as an overall system by scientific convention. When determining pollutants in the overall system it is therefore usual to choose individual systems that constitute a large part of the overall system in order to describe this. For example, to describe the pollutant load on a terrestrial ecosystem the sampling and analysis is usually carried out with soil samples. A great advantage of choosing soil to describe a terrestrial ecosystem is that the composition of soil can be depicted in the form of a model, and that the number and weight of the individual samples to be taken can be calculated.[12]

However, the degree to which a sample is representative of a set of data or an individual system is not an indication of the relevance of its information to the ecosystem as a whole. For example, to describe heavy metal levels in earthworms, further soil parameters such as the pH or cation exchange capacity may be included as well as the total concentrations in the soil samples.[13] This means that for certain frames of reference an analysis of the pollutants found in samples of organisms is more suitable for describing conditions in an ecosystem, since we may assume that the system ''organism'' includes and integrates the individual systems ''soil'' and ''soil chemistry''. The purpose of determining pollutants in organisms within an ecosystem is not to acquire information that has the same value as that yielded by a substrate sample. The intention is rather to acquire information on the integrating effect of the organisms in order to include this in a description of the ecosystem.

The demands on sampling that result from this are discussed below, using heavy metals as an example of environmental pollutants.

3 FACTORS INVOLVED IN HEAVY METAL UPTAKE BY ORGANISMS

3.1 Abiotic Factors

The total of all organisms found in or on a site may be regarded as a biocoenosis of one or several ecosystems. The properties of an ecosystem as a basis for the life of the organisms are governed by abiotic factors such as light, temperature and humidity, and also by the characteristics of the substrate concerned (water, soil, air). Depending on the exposure of the organisms the pollutants may be deposited, and thus collected for analysis, directly on their surfaces. This simplest case of exposure to a pollutant has led to the development of numerous monitoring methods for determining airborne pollution; most of them make use of plants.[9,14–26] As sessile organisms they function as a sink for airborne pollutants because of their large surface area. But these pollutants do not remain on the surface; they may be washed off again at irregular intervals by precipitation.

The effect of water on the distribution of polar substances is a fundamental challenge to the determination of pollutants in ecosystems, since water dilutes heavy metal concentrations in general and is largely responsible for their transportation. Even if the quantity of heavy

metals entering an aquatic and a terrestrial ecosystem is the same per unit of surface area, their distribution is greatly dependent on the distribution of the water in the system and the proportion of water contained in the predominant solid matrix (soil, sediment), since the heavy metals are subject to matrix-dependent adsorption and desorption mechanisms. In the simplest case, adsorption to a solid can be expressed by a Langmuir adsorption isotherm (Figure 1). The model assumes that a binding site in the matrix is occupied by a molecule or ion of the pollutant which is thus removed from the aqueous phase of the system (e.g. river water or water in the soil).

If an element has been bound by the substrate it is only available for uptake by an organism if the affinity of the pollutant for the organism is greater than that for the surrounding substrate, or if the affinity of the pollutant for the substrate decreases. In the Langmuir adsorption isotherm this area is marked by the plateau phase of the function in which the adsorption capacity of the substrate decreases and the pollutant remains in the solution.

The adsorption capacity of a matrix for polar pollutants such as heavy metal ions depends on several factors. The most important of the abiotic factors are the structural characteristics of the matrix, the pH of the solution, its redox potential and the presence of other cations. In areas with marine influence the high concentration of sodium and magnesium reduces the uptake of zinc and cadmium.[27] A fall in the pH of the soil does not only permit increased replacement of heavy metals by protons. The buffer systems of the soil are also broken down, and the heavy metals associated with them are released.[28] The same applies to a falling redox potential when there is a lack of oxygen in the soil due to waterlogging. In this case oxidic compounds are reduced and the associated heavy metals released.[29] In the sapropel of a lake or river bed, on the other hand, a positive redox potential can contribute to the oxidation of reduced sulfidic heavy metal compounds and thus to their mobilization.[30] It is important for both forms of mobilization that events such as rain, varying currents or temperature-related water circulation at the ecosystem level influence the oxygen content of the system and thus the mobility of heavy metals. They must therefore be taken into account during sampling as well. Table 1 shows some of the abiotic factors that may influence the mobility of heavy metals in ecosystems.

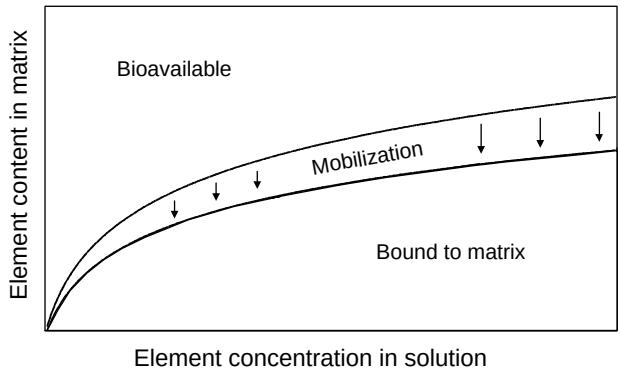

Figure 1 Theory of bioavailability of pollutants. The lower the adsorption capacity of the matrix for a particular pollutant, the greater is the bioavailable amount of the pollutant in solution.

Table 1 A selection of abiotic factors affecting the mobility of heavy metals in ecosystems that have to be taken into account during sampling, depending on the frame or reference

Particle size of the substrate	The smaller the particles, the larger is the adsorbing surface area
Organic matrix	In the organic matrix heavy metals are bound to functional groups, e.g. carboxyl groups
Clay content	Heavy metals are adsorbed by aluminosilicates and incorporated in the lattice structures over the course of time
pH	As the ion concentration increases, heavy metals are eluted from the matrix structures by exchange and dissolving processes
Redox potential	Manganese oxide compounds, especially, are reduced as the redox potential decreases, and the associated heavy metals are dissolved
Cationic effect	Elements compete for the binding sites at a surface

3.2 Biotic Factors

Biotic factors are those factors by which heavy metal uptake by an organism can be actively modified. They are numerous and occur in various forms, which makes it much more difficult to incorporate them in sampling. Their effect is partly due to the fact that the abiotic changes already described can be altered by the organism. Moreover, they include all the ecological and physiological parameters that characterize the interaction of an organism with its environment. Among the latter there are also some that show no connection with the primary uptake of heavy metals at first glance but constitute sources of error during sampling or subsequent analysis.

3.2.1 Determining the Monitor Species

Roth-Holzapfel[31] described the accumulation factors as the quotient of the organism and substrate concentrations of different metals in saprophytic species of a substrate. Depending on the metal and the species concerned (e.g. Diplopoda, Lumbricidae, Enchytraeidae, Collembola) the concentrations in the organisms differed in respect

to both the essential elements magnesium or potassium and heavy metals such as lead and cobalt.

But differences in the heavy metal concentrations found in the organisms in an ecosystem do not only occur between different taxonomic groups. They are also to be found between representatives of one family or genus and are attributable to the organisms' requirements for survival. Soil organisms such as earthworms construct extensive tunnel systems and take the soil into their bodies as they do so. This activity results in bioturbation of the soil, intensified contact with the polluted soil surface and an influence of the worms' digestive system on the mobility of the heavy metals in the soil.[32] Through the uptake and digestion of polluted food, pollutants are released from the food and taken in by the worms. Although different species of worms may exist side by side in the same ecosystem, their pollutant concentration also differs because of their various preferences in respect of food. Anecic earthworms feed on the leaf litter on the surface and the roots of plants. Their tunnels may extend several meters down into the soil. But endogeic earthworms feed on the detritus of the organic upper soil layers, which they rarely leave.[33] Although both types of worms are collected in the same way during sampling and come from the same soil, there may be significant differences in respect of their accumulated pollutant concentrations because of their different ways of life.[34] This means that a strict distinction between species must be made when determining pollutant concentrations in organisms.

3.2.2 Bioaccumulation and Biomagnification

Worm predators, such as hedgehogs, foxes and moles, are subject to the same laws concerning the uptake of pollutants through food, namely that contact with the pollutant through the food chain plays an important role in addition to a pollutant concentration in the surrounding environmental compartment. On the basis of the heavy metal levels found in vertebrates, Ma[35] found that insectivores have significantly higher heavy metal concentrations than herbivores in the same ecosystem. For herbivorous species of wild animals, Steinnes et al.[36] demonstrated that the cadmium levels found in the kidneys are higher when the animals have eaten herbs and mushrooms than after the intake of grasses.

The bioaccumulation and biomagnification of pollutants is highly significant for representative sampling, since the accumulation paths can usually be ascertained but remain unknown in the organism to be analyzed; this is because the preceding choice of food depends not only on the organism itself but also on the food available. So if conclusions are to be drawn about the burden on an ecosystem from the concentrations found in organisms, Axiom 1 is violated more and more as bioaccumulation increases. This violation decreases from the euryoecious species without a preference for certain foods to stenoecious species with food preferences, for in the latter case there is a greater probability of finding a connection between pollutant concentrations and both the food and the availability of food.

3.2.3 Effects of Biomass on the Determination of Pollutant Concentrations

3.2.3.1 Analytical Requirements Concerning the Composition of Biomass The level of analysis is an important dimension in the determination of pollutants because it sets the chemical and physical limits within which an investigation can be carried out. This means that it also has an influence on sampling and the choice of the organism to be sampled.

Increasing use has been made of inductively coupled plasma atomic emission spectrometry and inductively coupled plasma mass spectrometry to determine heavy metal concentrations in environmental samples.[23,36–38] In multielement analysis these techniques have the advantage of permitting a large sample throughput as well as low-detection levels and a very wide linear measuring range.

A specific problem of environmental analysis is that in many cases it is only possible to use a part of the original sample for instrumental analysis. But according to Axiom 1, this part must have the same chemical composition as the original sample. This can be achieved by homogenizing the original sample. If the available measuring technique is sensitive enough it is possible to analyze small quantities of sample material as they are, without previous homogenization. But if the elements to be analyzed only occur in very small amounts in the samples, the detection limit of the equipment makes it impossible to analyze these elements in small quantities of sample material without greatly increasing the risk of analytical error. Homogenization lessens the probability of such error while avoiding the necessity of testing every part of the sample organism separately.

3.2.3.2 Compartmentation To meet the statistical requirements of sampling it is usual to choose organisms that are both available in large numbers and characteristic of a particular ecosystem. These are often lower plants (lichens, mosses etc.) or lower animals (such as earthworms, snails or wood-lice). Because of the small mass of some organisms, individuals can only be analyzed for elements with relatively high concentrations; otherwise mixed samples have to be made up from several organisms in order to reach the analytical detection limit.

Table 2 Heavy metal concentrations in different compartments of selected organisms

Organism	Pollutant	Organ	Content ($\mu g\,g^{-1}$)	Biomass (mg)	Ref.
Brassica juncea (Cruciferae)	Lead	Shoot	258	165	
		Root	2675	76	39
Polytrichum formosum (Bryophyta)	Lead	Leaf	36.9	Nd	
		Shoot	36.2	Nd	
		Sporogonium	10.8	Nd	41
Oniscus asellus (Isopoda)	Copper	Hepatopancreas	1144	Nd	
		Remaining tissue	583	Nd	40
Aporrectodea longa (Oligochaeta)	Lead	Gut wall	32	Nd	
		Waste nodules	89	Nd	
		Whole worms	5.9	Nd	42
Marisa cornuartietis (male) (Gastropoda)	Tributyltin	Gonad	14.2	548	
		Kidney	169	22	
		Penis	7.3	100	
		Remaining tissue	7.6	1330	43

Nd = No data.

One of the problems involved in determining element concentrations from total digestions is that the information yielded by the pollutant levels in the individual organs of the organisms is lost. Pollutants are not distributed equally throughout organisms; they may be compartmentalized in different organs (Table 2). If, for example, a storage organ polluted with heavy metals accounts for a small proportion of the overall weight of the organism, a reduction in the size of this organ may also reduce the heavy metal concentration of the organism as a whole, although the effect on the weight of the organism is negligible.

With such compartmentation in mind, it is important to ensure that the distribution of the pollutants in the organism is close to the ideal of uniformity if a representative analysis is to be made from the total digestion of whole organisms. Otherwise ecophysiological fluctuations that cause accumulation or depletion of an element in a compartment may influence the overall concentration of the sample.

3.2.3.3 Influence of Homogeneity on Pollutant Concentrations The ecophysiological fluctuations in sample organisms include changes in the weight, size and structure of the organisms. In the case of spruce needles, Weissflog et al.[44] found that pollutant levels varied according to the age and surface area of the needles. The concentrations of polyaromatic hydrocarbons and lead fell significantly as the size of the needles increased. Such information could not have been acquired if needles of different age and size had been homogenized, and the logical conclusion is that only needles of the same size or age must be used if the analysis is to be carried out with mixed samples instead of individual needles. Otherwise the biomass with lower pollutant concentrations may have a diluting effect on the total sample.

The age and size distribution in organism samples is very important, in contrast to the substrates of an ecosystem, for there are relative differences between the two in respect of length of exposure to the pollutant. Whereas a substrate is constantly present at a particular site, the maximum presence of an organism at the same site is the length of its life. This means that the length of exposure of the organism or its compartments to the pollutant is many times shorter. Moreover, the size of an organism is not independent on age as is the case with a substrate; within this relationship it is determined by rapid growth processes.

Using the moss species *Polytrichum formosum* and *Brachythecium rutabulum*, Lötschert et al.[45] and Engelke[41] showed that the cadmium, lead and zinc concentrations in the plants increased with the age of the tissues. Koeckritz[46] investigated the copper concentrations in the earthworm species *Aporrectodea caliginosa* at three sites during one summer and although there was found to be a positive correlation between the levels in the soil and those in the worms, the concentrations in the worms varied greatly over the five-month observation period. At one site a maximum concentration in the earthworms resulted from the fact that the worms fell into a state of drought rigor after a long hot period of intensive sunshine without rain. The event was accompanied by considerable weight loss in the worms. A comparison of the concentrations in the animals as a function of their weight showed that the copper levels decreased exponentially as the weight of the worms increased (Figure 2).

As in the example of the spruce needles, the size distribution of a mixed sample of earthworms is an

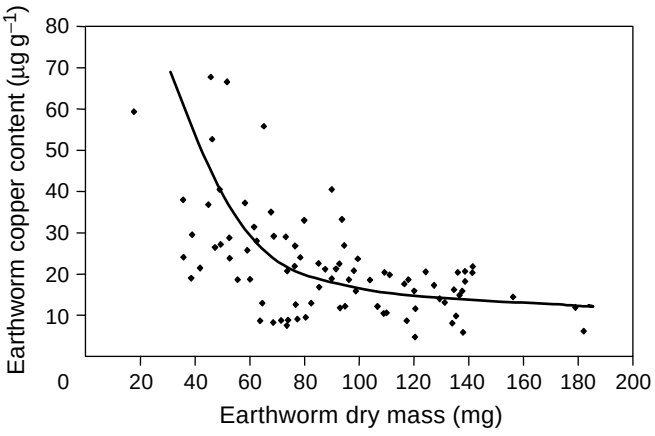

Figure 2 Fall in the copper content of *A. caliginosa* as the dry mass of the worms increased.

important condition for determining average pollutant concentrations over a period of time. The fall in pollutant levels as the size of the needles increases can be explained by the slow growth rate of spruce needles. Although the pollutant-adsorbing surface area becomes larger as the needles grow, it only increases to the second power in relation to the volume and weight of the needles; this results in a dilution of the pollutants by the needles' own biomass. In *A. caliginosa* this causality cannot readily be achieved. Heavy metal content was found to depend on the weight of the worms in the case of copper but not with lead, cadmium or zinc. Other biotic factors that had the same effect on the result must therefore have played a role.

So the requirement that only organisms of the same species should be used in one set of samples is not enough. It is also necessary to specify that only organisms of the same species, size, age and sex must be sampled or used in the analysis. The list can be extended indefinitely, as long as the requirements contribute to the homogeneity of the sample. Only empirical work can show how far each factor influences the pollutant content of an organism.

3.2.3.4 Surface of the Organisms The accumulation of heavy metals in a monitor organism depends not only on their presence but on the degree to which the pollutants are bioavailable as well. In order to be bioavailable, however, the pollutant naturally first has to reach the monitor organism and come into contact with its surface. In the earthworm example we have already discussed how such contact is intensified by the uptake of soil material into the digestive system, since this active intake increases the surface area. In the case of plants the surface area is enlarged by the assimilatory and respiratory system of the roots and leaves.

Plants come into contact with airborne pollutants mainly through the leaves; such contact can be described as dry and wet deposition. The amount of pollutant coming into contact with a surface depends on its concentration and also on the extent to which it is bound to airborne particles, on the size, nucleation, coagulation, condensation and sedimentation properties of these particles and the affinity of the leaf surface for them. Metals and semimetals are usually emitted in a solid form, bound to dust, whereas organic pollutants are mainly emitted as gases. Fly ash and other dusts with a particle diameter of more than 10 µm settle in the vicinity of the emitter, whereas gaseous substances and particles with a diameter of less than 10 µm remain suspended for long periods.[47] They are carried by the wind and deposited on the substrates of the ecosystems according to the direction of the prevailing winds and the topography of the region.

Whether or not an element passes into an organism depends to a large extent on its solubility in water, and this in turn depends on its chemical form, the surface of the particles and ambient conditions. According to Rohbock[48] the amounts of water-soluble heavy metals in rainwater can differ considerably, and the effect of the water has an important influence on uptake in addition to the element concentrations deposited on the surface of an organism.

But precipitation does not only deposit elements on the relevant surfaces. Ions already stored and particles adhering to the surfaces may be washed off again, depending on the nature of the precipitation. This means that the type of precipitation has an effect on the uptake processes[49] and that a brief, heavy rainfall results in less pollutant uptake than gentle rain over a longer period. Rühling and Tyler[50] demonstrated a connection between the amount of precipitation and lead concentration in *Hypnum cupressiforme*. In tests on poplar leaves, Wagner[51] showed that the levels of cadmium and zinc in older leaves were similar to those in younger leaves, whereas the lead concentration of the older leaves was twice that of the younger leaves. This phenomenon was explained by the fact that cadmium and zinc were more easily washed off and out of the leaves by heavy rain after the first third of the vegetation period.

This is a further example of how the age of an organism may affect the result of monitoring. In the above examples the age of the organism had an indirect effect on pollutant concentrations because the biomass varied over the course of time. But as the age of the organism increases, so does the likelihood of its being influenced by abiotic or biotic events capable of significantly altering heavy metal concentrations.

The surface of organisms is rarely constant, and in extreme cases, such as the ecdysis of insects, crustaceans and reptiles, it may be renewed completely in the course of time. The surface of plants varies too. According

to Hanway and Weber[52] and Ashley[53] the biomass distribution of the annual soybean changes greatly during a vegetation period. After the plant has flowered the biomass of all the parts increases exponentially, but the leaves and leaf-stalks, that is, the parts with the largest surface area and the greatest adsorption of pollutants, are cast off successively and renewed.

But the surface of plants does not only vary with time as the plant grows; it is also influenced by abiotic factors at the site. Packham and Willis[54] showed with *Oxalis acetosella* that the ratio of the biomass of the leaves to that of the roots increased as insolation decreased. In other words, the plant adjusted to a lack of light by increasing its surface area. An increase in the biomass of the roots in response to a lack of water in the soil may be regarded as a similar form of adjustment.[55]

The thallophytic mosses have a surface that is much more dependent on water than the cormophytes described above, and in spite of their rather simple structure the surfaces of mosses of different species are not directly comparable even if the same amount of water is available. In moss species the capacity to take up cations varies with the morphology of the biomonitor, which may differ greatly according to conditions at the site.[56] To give two examples, *Pleurozium schreberi* has small leaves arranged close to the stems that store the water from precipitation, whereas the leaves of *Polytrichum formosum* stand out from the stems. Moisture from dew and mist settles on these leaves and is then taken in by the plant.

A high surface-to-volume ratio such as that found in mosses not only leads to an increase in the parameters that encourage the accumulation of pollutants; those parameters that promote the variability of pollutant concentrations also acquire an influence.

If the surface area of an organism is very large in relation to its volume or weight, particulate heavy metals adhering to the surface may account for a large proportion of the overall content. It may therefore be advisable to wash off these heavy metals before analyzing the sample.[57,58]

3.2.3.5 Seasonal Changes Organisms growing in unpolluted areas contain a specific basic concentration of all the elements that occur naturally. These originate from natural element cycling – mainly through the atmospheric transportation of soil particles – from the marine environment and from volcanic activity. As a general rule it must be taken into account that the influence of all the biotic and abiotic factors mentioned increases as deposition decreases.[18,19,59,60] The influence of such factors on element concentrations is shown by the fact that the levels in the organisms do not correlate properly with those in the substrates, or that the concentrations in the organisms are subject to fluctuations, sometimes repeated at regular intervals, instead of being stable.

The determination of pollutants in organisms is often used in assessing pollutant loads on ecosystems. Fluctuations in content according to the pollutant and organism concerned must therefore be taken into account in an evaluation since natural fluctuations in the organism levels may overlay inputs from artificial sources, especially when concentrations in the environment are low. Important features of biological fluctuations are the stability of an ecosystem in terms of space, and above all in terms of time, that is expressed in regular events such as the seasons, the tides or the phases of the moon, and the adjustment of the organisms to these rhythms.

The effects of natural fluctuations on a monitoring result can be illustrated by the example of element concentrations in the moss *P. formosum*. In Figure 3 the magnesium, zinc and cadmium levels in the moss at one site are plotted over a period of two and a half years. In order to show the harmony of the fluctuations among themselves the amplitudes have not been adjusted to the scale, that is, only the mean element concentrations correspond to the axis of ordinates. It is plain that the fluctuations in element levels depend not only on the season but also on the mean element concentrations in the moss too, for when the mean cadmium concentrations fall, the amplitude of the element fluctuations is also reduced.

The fall in heavy metal levels in the mosses in spring can be explained by the diluting effect of the plants' biomass production at this time of the year.[22,61] But without an adequate knowledge of the time resolution of the element concentrations in the moss, any individual measurement would result in misinterpretations of the pollutant load

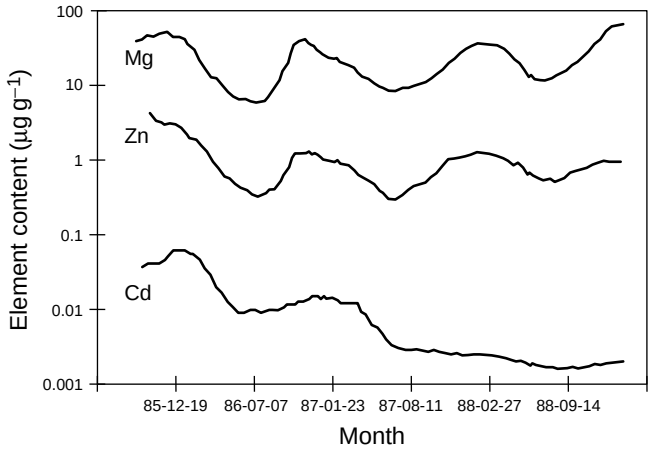

Figure 3 Seasonal fluctuations in metal concentrations in *Polytrichum formosum*. Unlike the mean pollutant loads, the amplitudes have not been adjusted to scale.

on the ecosystem, for the monitor may show a high level of pollution even if the level in the ecosystem is low.

The effect of fluctuation on a monitoring result does not only depend on the element concentration in the environmental compartment and the organism; it also depends on the organism's ability to accumulate a particular element. If an increase in the element concentration in an organism is to be attributed to artificial input, the increase in content must differ significantly from that caused by natural fluctuations. Magnesium and zinc are essential elements for organisms, but the levels found in the example of the moss *P. formosum* differ 100-fold. We may assume that for these elements special uptake and storage mechanisms exist that are responsible for the high overall level in the organism but are also subject to fluctuations. Because of the natural element concentrations it is not possible to show small inputs from anthropogenic sources within these fluctuations. Nonessential elements such as cadmium are also subject to periodic fluctuations, but because of the low levels in the organism additional inputs are more easily shown provided that they differ from the mean concentrations.

It is essential to find the reason for rhythmic fluctuations in element concentrations in order to compare different monitoring methods in which the same monitor organism has been used. The element concentrations shown for *P. formosum* follow a rhythm determined by the seasons. Although the months of the relevant years have been entered on the *x*-axis, this is only a nominal scale for the site concerned. In order to compare different sites at different times the time entries have to be made on an ordinal scale with a timing ecological component. On the basis of phenological observations in 50 gardens of identical genetic composition the progress of spring was recorded over a period of 30 years.[62] Whereas the speed of its progress from Portugal to Norway is about 60 km per day close to the Atlantic, its continental speed is only 35 km per day on the line from Greece to Finland. A speed reduction of three to four days per 100 km is also to be expected for every 100 m of altitude.

Since we may assume that the mean period of insolation and mean temperatures and precipitation determine the growth time of the moss, the fluctuations in pollutant levels may depend on these factors too. This means that the time of sampling in relation to climatic conditions may have a lasting influence on a monitoring result although the pollutant burdens on the ecosystem are the same (Figure 4). Scales such as mean temperatures, mean insolation and phenological observations must therefore be checked for their suitability as ordinal scales.

In the above example of the moss *P. formosum*, heavy metal concentrations were studied over a period of one year. The maximum concentrations were observed between December and March. These were followed by a fall in the metal levels from April to May, and until October the values were stable and low. In November the metal levels rose again by up to 100%. Because the metal concentrations fell by about 100% within 100 days in the spring, and in view of the phenological speed with which the spring advances in central Europe, we may assume that the error will be about 2% if the samples are taken at the same time but 100 km apart in a north–south direction. This error may be considered negligible for regional comparisons, but it acquires importance as soon as biomonitoring methods are used at the international level. We also have to remember that the rule about the continuation of errors applies here too, so that even in regional comparisons the overall error increases if the sampling is carried out over a period of several days or perhaps weeks. To avoid this error and ensure that the metal concentrations in moss from sites far apart are comparable, it is therefore advisable to choose the summer months of July and August. These two months are within the period of minimum metal concentrations and allow about 30 days' margin for variation before and after.

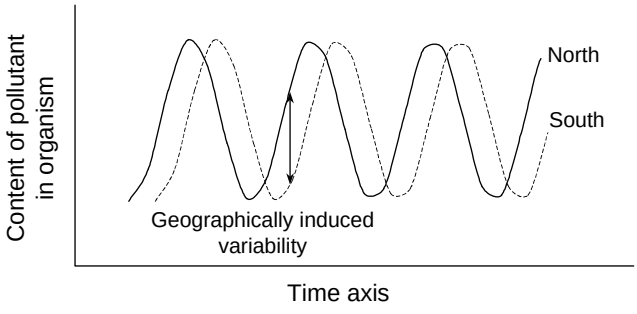

Figure 4 Theoretical influence of periodic element fluctuations on the monitoring result. If the sampling sites are far apart, the delayed phenological development of organisms from south to north during the year may result in different element concentrations in the monitor organism although the pollutant levels in the ecosystem, and the sampling dates, are the same.

3.2.3.6 Loss of Information through Sampling The above examples have shown that the heavy metal concentrations in organisms may depend on numerous factors. Thus it is only possible to determine the pollutant load on an ecosystem from the levels found in organisms if the parameters influencing the accumulation factor are known. But since these biotic and abiotic parameters affect all the organisms in an ecosystem and the accumulation factors of the organisms differ nevertheless, the sum of all the parameters cannot be universally correct for all the organisms in an ecosystem. Thus, one objective of biomonitoring must be to ascertain the main accumulation parameters for each species and

to quantify them. To do this it is necessary to view the monitor organism as a whole and take all the known interactions into account during sampling.

Figure 5 shows the main interactions that have been shown to affect the element concentrations in moss.

When attempting to view an organism as a whole within an ecosystem we also have to take into account those factors that do not have a direct effect on the organism. These may include the density of the leaf canopy over a moss population that affects the amount of sunlight or rain reaching the plants.

The abiotic factors that may influence the mobility of heavy metals have already been listed in Table 1. Table 3 shows the biotic factors of heavy metal uptake that have to be taken into account in biomonitoring procedures.

The list shows that the biotic factors are parameters that cannot be ascertained directly during sampling. In order to determine them it is necessary to carry out laboratory or long-term tests in individual cases, or investigations that may exceed the available laboratory capacity.

As far as possible, it is important that the sampling procedure keeps open the option of carrying out such tests at a later date. In other words, it must not affect either the composition of the organisms themselves or their composition relative to each other within the ecosystem. It is important to use sampling equipment that minimizes the risk of contamination from outside sources. The samples must be taken in such a way that the result of the analysis is capable of representing the undisturbed condition of the original sample. The organism must be removed from the population carefully and stored in an equally careful manner.[21] A mistake commonly made at this stage is mixing the samples from one site during the sampling procedure. When plant samples are cut, for example, cell, phloem and xylem sap may exude and dissolve and distribute pollutants. This effect is aggravated by mechanical action on the surfaces through the use of containers that are too small, and the material can then only be used as a mixed sample even if it was intended to analyze individual parts of the organism.

Further practical instructions for sampling are given by Ernst,[3] MacNaeidhe,[6] Markert,[8] and Wolterbeek.[63]

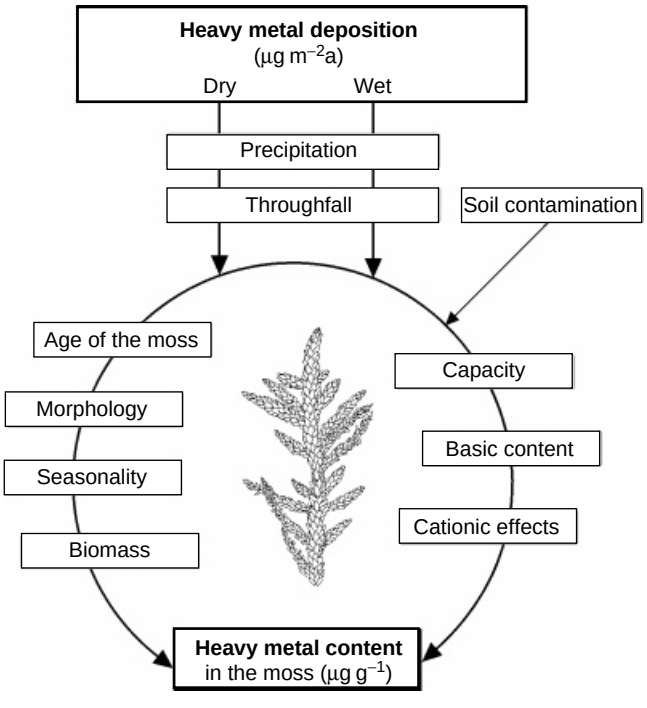

Figure 5 Factors affecting heavy metal uptake by mosses.

Table 3 Selection of biotic factors affecting heavy metal uptake

Biotic factor	Result
Formation of assimilatory and respiratory surfaces	Enlargement of the diffusion surface and increased heavy metal uptake
Bioturbation by mobile organisms	
Release of inorganic and organic acids, some of them complexing, by plants and microorganisms	
Soil respiration	
Digestive activity of vertebrates	Lowering of the pH and mobilization of heavy metals
Release of enzymes	Enzymatic dissolution of organic matrices containing heavy metals
Binding of heavy metals by phytochelatins or metallothionins	Transformation and compartmentation
Bioaccumulation, biomagnification	Increased influence of transfer factors
Biological fluctuations	Changes in element concentrations although the burden on the ecosystem remains constant
Physiology	Excretion, secretion, storage

4 CONDUCT OF SAMPLING

4.1 Choice of Organisms

Since monitoring procedures can show both the situation with respect to pollution and the accumulation variables, the organisms must be chosen to suit the frame of reference before sampling is carried out. We have already discussed what requirements an organism must meet in order to be representative of the burden on an ecosystem.

But the burden on an organism can also be used to investigate the influence of the accumulation variables on pollutant accumulation. This may be necessary if, for example, the effect of an anthropogenic measure on the behavior of pollutants in the ecosystem is to be investigated in an environmental impact study, even if the measure itself does not constitute an additional burden. An example of this would be the renaturation of an ecosystem after some form of intensive use, in which a former drainage system is abandoned to promote the development of a wetland. As the soil becomes wetter its redox potential may be reduced, and organic acids may be released as a result of anaerobic respiration by the soil organisms. This can mobilize heavy metals and thus make them bioavailable.

The bioconcentration factor mentioned above can be used as a parameter for measuring the availability of a pollutant for organisms. As the quotient of the pollutant contained in the organism and the substrate it describes an organism's ability to take the substance in.

The bioconcentration factor may increase although the pollutant concentrations in the substrate are constant if an organism takes in a larger amount of a pollutant. The cause of an increased intake may be assumed, initially, to be the sum of all the abiotic and biotic accumulation variables. The object of the investigation is to weight the factors. As in conventional monitoring methods it is necessary to analyze the pollutant concentrations in the substrate and the food in order to determine the bioaccumulation factor or biomagnification factor that describes accumulation through the food. But by determining the accumulation variables (pH, temperature etc.) at the same time, it is also possible to describe the main components of pollutant accumulation, whereas these remain unknown or a matter for speculation with ordinary monitoring methods. When introducing a new monitor organism it is therefore advisable to determine the bioconcentration factor, as this makes it possible to quantify the limits of a monitoring method.

According to Baker,[64] organisms can be classified according to the manner in which they take up pollutants (Figure 6). The nature of the pollutant intake determines whether the concentrations in an organism are capable of representing the burden on an environmental compartment.

The most suitable organism for all frames of reference is a monitor whose pollutant content depends directly on the level in the environmental compartment. The bioavailability of a pollutant for the monitor organism is influenced mainly by abiotic factors and is overlaid only slightly, if at all, by ecophysiological processes in the organism. In spite of linear pollutant accumulation the bioconcentration factor may be less than, equal to or greater than one; any differences are only in respect

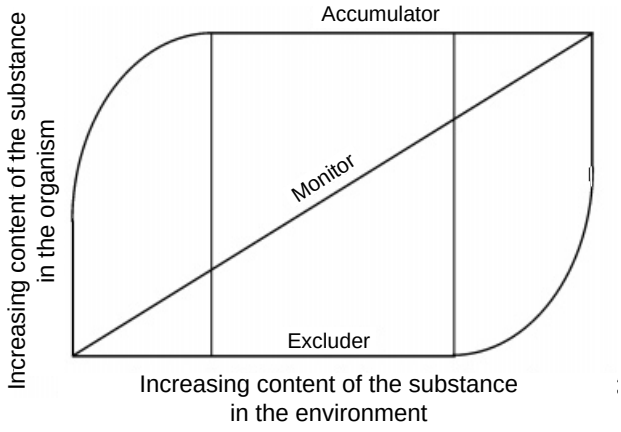

Figure 6 Different types of pollutant uptake by organisms.[64]

of the evaluation of the pollutant concentrations and the working range. If the bioconcentration factor is greater than one, even quite small fluctuations in the compartment levels are reflected in the organism levels, as the concentrations are duplicated. If the level in the organism is lower than that of the environmental compartment, only major changes in the pollutant concentration can be detected with the organism. But if the bioconcentration factor of such an organism becomes greater than one we may assume that the bioavailability of the pollutant concerned for the organism has increased as a result of abiotic factors.

In an organism that is ideal for biomonitoring there should be a linear relationship between the concentration of an investigated substance in the environment and the concentration in the organism across a very steep pollutant gradient. But with low environmental concentrations, especially, pollutant uptake is not linear and resembles that of an accumulator or an excluder. The accumulator builds up pollutants exponentially at low environmental concentrations, whereas the excluder only takes in small quantities of the pollutant initially as it is able to regulate uptake physiologically. Not until a tolerance threshold is exceeded does the concentration in the organism increase noticeably.

Table 4 contains a summary of the biotic variables that may influence a pollutant concentration in an organism and their suitability for the purposes of monitoring.

4.2 Axiom 2: Distribution of Organisms

In order to draw conclusions about the concentration of a pollutant in an ecosystem from that in an organism it is necessary to observe statistical conventions which always state that the number of organisms to be taken as samples must be representative of the chosen site. This requirement is not easy to fulfill, for the total population to be investigated and sampled is always unknown and

Table 4 Influence of biotic accumulation variables on the information value of pollutant concentrations in organisms

Ratio of the dry weight of the organism (O) to the weight required for analysis (E)	**O ≤ E**	**O ≥ E**	**O ≫ E**
Suitable for:	Preparation of mixed samples of medium size and composition	Excellent for digestion of the whole organism	Analysis conducted from organs
Necessary form of distribution of the pollutant in the organism	May be disregarded	Uniform distribution	Uniform distribution in the organ
Type of pollutant uptake	**Accumulator**	**Monitor**	**Excluder**
According to Baker (1981), suitable for showing:	Bioavailability	Burden on the ecosystem	Threshold values
Variation of pollutant concentrations due to abiotic and biotic influences	**High**		**Slight**
Organism suitable for:	Indicating accumulation through bioavailability		Representative determination of pollutants as a biomonitor
Variation of pollutant concentrations due to abiotic and biotic fluctuations	**High**		**Slight**
Suitability of the organism as a monitor	Unsuitable if ecophysiological conditions cannot be determined		Highly suitable, since changes in pollutant concentrations in the ecosystem can be determined directly
Suitability of the organism as an accumulation indicator	Suitable for determining biotic accumulation variables		Suitable for determining abiotic accumulation variables
Comparability of results from different regions	Poor		Excellent

is also variable. The distribution of individuals within a population, of populations within an ecosystem and of ecosystems themselves is not random. It is determined by climatic, chemicophysical and ecological factors. Interactions of the organisms with other forms of life and their own species may lead to changes in population density and distribution over the course of time.

In the following we shall assume that sampling serves the purpose of comparing pollutant concentrations at sites that are different and independent of each other, that is, the sampling procedure is not intended to ascertain the actual distribution at the site.

Since the real distribution of the organisms at each sampling site is unknown, it is necessary to use a sampling technique that is independent of the existing distribution. A method that has been used successfully for sampling the moss *P. formosum* is shown in Figure 7 after Markert.[11] Independently of the true distribution of the organisms, a random distribution of the sampling points over the site is generated by dividing the area up according to a grid consisting of four zones.

No organisms are taken from the inner zone of about $40\,m^2$. This is the area in which the physicochemical parameters that may be assumed to apply to the whole location are determined. These parameters include wind speed, temperature, precipitation and insolation, but not the pollution of the soil or its pH.

The second zone serves to separate the first from the third, in which the actual sampling is done. It should not be less than 10–20 m wide to ensure that the sampling has no direct effect on measuring work in the inner zone.

The actual sampling is carried out in the third zone, which is divided up in the form of a grid consisting of at least a hundred squares. The selection of the squares to be sampled is random and independent of the topography and so on. The squares are approached in a clockwise

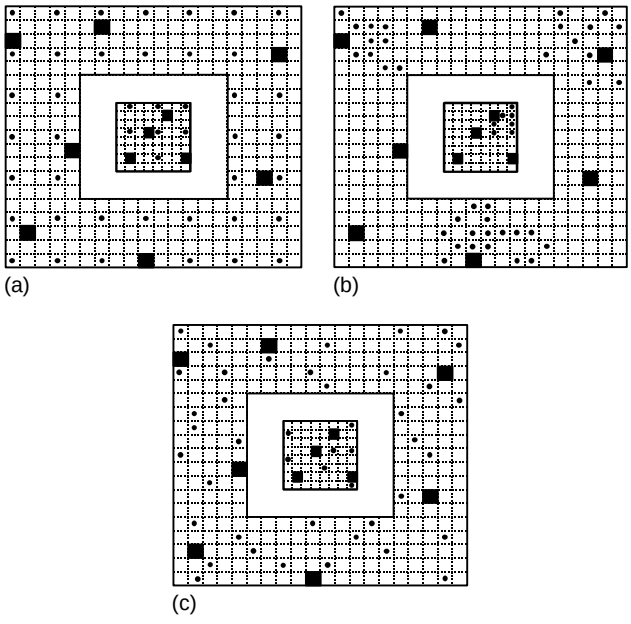

Figure 7 Conduct of random sampling that is independent of real distribution. (a) Uniform: uniform distribution results when organisms avoid each other and thus create territories. In the case of plants, uniform distribution may occur if the same resources are used for growth or if the plants inhibit each others growth. (b) Contagious: aggregations may form if, for example, a food resource is used by several organisms or the presence of one individual leads to that of others as in the case of flock formation. (c) Random: random distribution results if the probability of occupying a point within an area is the same for each organism and the presence of one organism does not prevent that of another.

direction to ensure undisturbed sampling. Should a square contain no organisms they may be sought in a clockwise direction around the point. During this simple random sampling procedure each square is only sampled once. It is therefore termed sampling without replacement, which is more precise than sampling with replacement.[65]

Since the analytically determined mean of the pollutant concentrations in the organism population at one site is intended to enable a comparison between different sites, the number of squares sampled has an important influence on the power of the method to differentiate. According to Eberhardt,[66] the number of samples required to carry out statistical procedures increases exponentially as the ratios of the means to be compared decrease. This means that it is not possible to state an exact number of squares to be sampled in this context, since the number depends on the frame of reference, the desired resolution of the method and the funds available.

4.3 Axiom 3: Economic Aspects of Information Content

In biomonitoring projects to determine the pollution of ecosystems, the target for the ratio of the means is often set too low for the number of samples taken. Because of the factors described above that influence the uptake of pollutants by organisms there is a great deal of variability in pollutant concentrations within a population sampled at a particular time. Standard deviations of 50–100% are the rule rather than the exception. A further difficulty is that a natural fluctuation in the pollutant concentrations may find its way into the result as well. Although both of these phenomena can be quantified and included in the result, the effort and expenditure required raises a question about the value of the information yielded by monitoring.

Since the pollutant concentrations and environmental conditions in an ecosystem are constantly changing, the results of monitoring only have the temporary character of a time window[2] and thus a limited information content. From the economic point of view the effort required to acquire the information must not exceed the value of the information. So if the ratio of the means to be compared is set very low (to achieve a high level of differentiation), the researcher is obliged to collect a large number of samples within a very short time. The time factor cannot be manipulated by the researcher, since it is determined by the natural fluctuations and the mobility of the pollutant in the system. The only variables that can be influenced are the number of sites to be sampled and the number of samples per site. Thus an important factor in practical, representative and comparative sampling to determine environmental pollutants in biological matrices is the personnel required to carry out the work.

Figure 8 shows the information content of sampling as a function of the work carried out. This general description

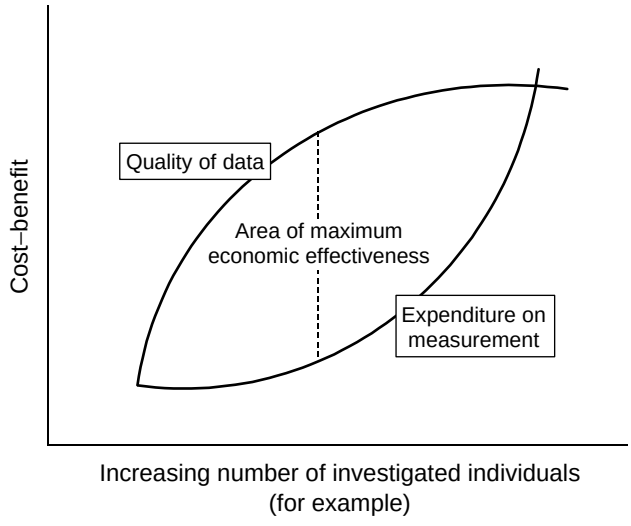

Figure 8 Increase in the information content of sampling as a function of the effort and expenditure required. (Adapted from Begon et al., 1991.[39])

of the relationship between work/cost and the resulting benefit is applicable to all the steps in biomonitoring and must be taken into account at the planning stage.

Although the method of random sampling described above is independent of the actual distribution at the site, an analysis of the individuals from one site reveals that their pollutant concentrations rarely show a normal distribution. In addition to binominal and Poisson distribution the most common forms found in environmental samples include negative binominal distribution, which requires a larger number of samples.[67] But if the necessary number of samples was not collected at the given time this does not mean that the sample is useless; it does, however, mean that the comparability of different locations has to be adjusted to the distributions found. As we have already said, the real distribution within the population is unknown at the time of sampling and cannot be determined until the results are evaluated. By this stage the time window for sampling has in all probability closed, so that additional sampling to adjust the number of random samples to the existing distribution is not permissible.

4.4 Influence of Size of the Organisms on Number of Samples

To show the distribution of a pollutant within a population it is necessary to carry out the pollutant analysis at the level of the individual, and this gives rise to the requirements described above in respect to the minimum size of the organism and the homogeneity of the biomass to be analyzed. In order to determine the manner in which pollutants are distributed statistically in a population the preparation of mixed samples must be avoided, and the biomass of an organism must therefore be sufficient to reach the analytical detection limit. If this is not the case, and the preparation of mixed samples is unavoidable, the number of random samples increases by the factor that is required in order to obtain a representative result for individuals of one size, age or other phenotypical characteristics.

5 HOW TO CARRY OUT A BIOMONITORING PROCEDURE

The aim of biomonitoring is to draw conclusions about the impact of heavy metal accumulation in organisms. The precondition is that all variables described above that might influence the heavy metal content in organisms are either minimized or known. As the methodical and measuring work involved entails significant cost (both financial and time-related), methodical conventions must be stipulated, aiming to guarantee a high level of ''ecological homogeneity'' of the examined material. This is even more true if the monitored organisms or their accumulation behavior is unknown.

In order to deduce the impact of heavy metals on ecosystems and exposition paths from the heavy metal content in organisms, it is imperative to ask if the corresponding organism represents the exposition path appropriately, thereby keeping in mind that many organisms of one and the same ecosystem can be exposed to various paths of heavy metal accumulation. Even though moss, for example, can appear with many plant communities in one ecosystem, its way of accumulating heavy metal differs, due to its thallophytic ecology and physiology, from the one shown by cormophytes in the same system. As a consequence, the simple measuring of heavy metal content in moss can only provide evidence about heavy metal exposition into the ecosystem via the air, whereas the impact of heavy metals on the soil can only be evaluated by means of other representative species.

In order to compare different monitoring procedures, it is necessary to select representative microorganisms and to take representative samples. So far, only very few practical guides (e.g. Markert,[11] Wappelhorst,[38] Rühling,[68] Tataruch[69] and Herpin, Markert, Siewers and Lieth[70]) have been published. This is because of the variety of exposition paths, types of ecosystem, biotic and abiotic mechanisms of accumulation, and the diverse demands of any biomonotoring process. Does it aim to determine ecophysiological accumulation variables or background impact, or to determine a source of emission systematically? As no generalization is possible, we will sum up only those factors that are relevant for a monitoring, which must, however, be adjusted to the corresponding kinds of investigation and organisms (Figure 9). In selecting an organism to be monitored, sampling sites, the samples themselves, and methods of analysis and assessment of data, the researcher does not face the challenge of collecting as much data as possible on the basis of a corresponding quantity of sample material. The crucial task is rather to minimize the heterogeneity arising from the many ways ecological factors can be combined, that is, to minimize it methodically in such a way that the findings can allow a distinct statement about the actual aim of the investigation to be made, which here is to determine the heavy metal content.

5.1 On-site Sampling

In order to assess representative data, it is imperative that the material contained in the sampled organisms is not damaged during the sampling process itself. As instrumental analysis cannot be applied on the site, the samples must be very carefully collected, transported, stored and processed. To avoid contamination plastic gloves must be worn, and the samples must be taken with

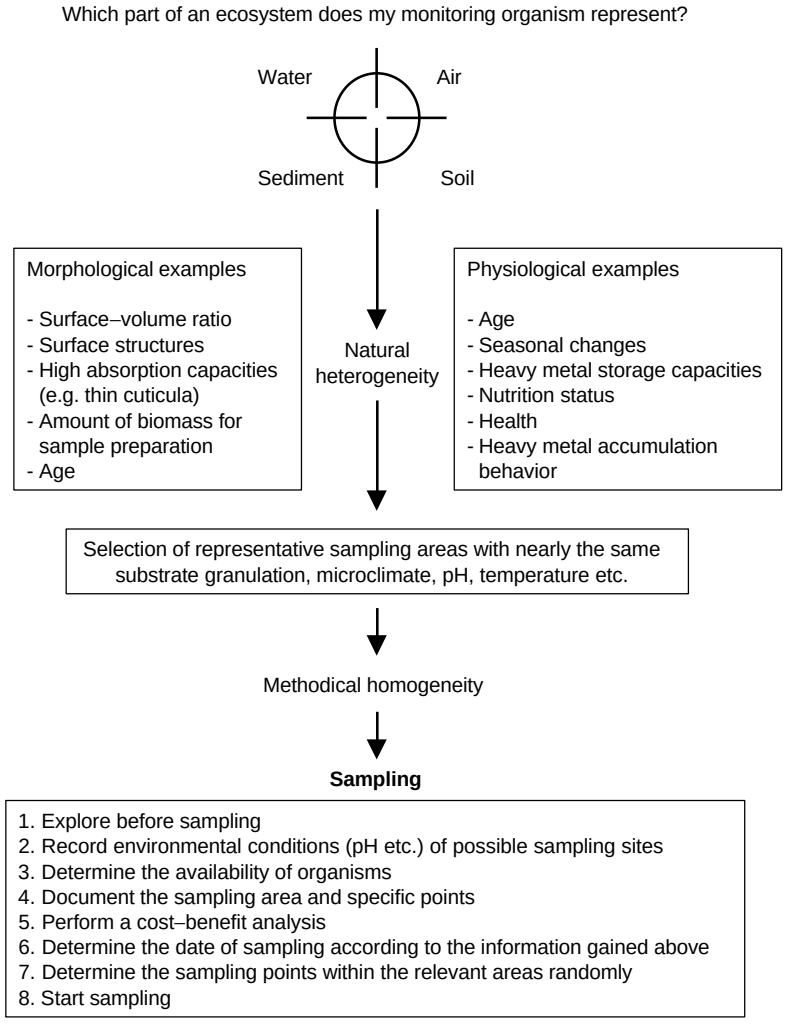

Figure 9 Ecological targeting.

metal-free tools (e.g. ceramic scissors, plastic tweezers) and be stored and transported in plastic containers. It is important to ensure secure transport and storage until measurement can be carried out.

It is also imperative for the researcher not to adulterate the sampling site itself. Especially when monitoring is carried out at one site over a long period of time, and when samples are taken consecutively, researchers must take care that the site disposes of a sufficient quantity of monitored organisms in order to avoid an abundant decrease which might provide locational advantage to other individuals of the same type or of another kind of organism. This could lead to increased growth of the remaining monitored organisms, and thus influence their capacity to adsorb heavy metals.

Assuming that the site is ecologically stable, it is likewise imperative to avoid deterioration of the microclimate which can be triggered by sampling. This includes, for instance, breaking twigs and branches, but also treading on the sampling area. Fixed paths should be used, in particular when samples are to be taken at consecutive dates, because, especially with plants, such damage will not necessarily be evident during the following season. However, a lasting change in the heavy metal content of such organisms can occur.

In order not to adulterate the sampled organism itself, the damage inflicted upon any one organism during sampling must be reduced to a minimum. This means that parts of the plant should not be removed by tearing or breaking, but by using scissors. Organisms are damaged once they are put under stress during sampling. An example is the collecting of earthworms using formaldehyde, electricity or mustard extracts, which aim to drive the earthworms out of the soil and to the surface. Earthworms react to these stress factors by producing coelom liquid; in turn, coelomocytes, which are rich in heavy metals, emanate from the body of the earthworms, leading to an adulteration of data. Sampling

should therefore be carried out by hand, even though this might take longer.[46]

The above described autolysis of organisms can cause the same errors in the assessment of sampling as does the human-caused damage to organisms. In both cases, liquids occurring naturally in the body can be emitted and distributed throughout the entire sample. This unwelcome homogenization of samples on-site must be minimized by not squeezing the samples in the sampling container, and by freeing them from foreign substances such as soil or leaves before placing them in the corresponding containers. This is important because body liquids can cause the heavy metals contained in the sample to mix with those of the foreign substances. It is likewise imperative to put the sampled organisms into cold storage on-site, but not to freeze them. Cold storage slows down the autolysis of the sampled material, avoiding stress factors by slowing the metabolism of invertebrates, which cannot always be killed on-site as they must empty their bowels for several hours before being processed for analysis.

5.2 Conclusions

The representative sampling of organisms of one ecosystem does not only demand high scientific competence of the staff dealing with the processing and the analysis of the samples, but also of the sampling staff, because none of the mistakes made during the sampling activities can be undone afterwards. Apart from the duty of care demanded of scientific staff, other factors are important. Thus the sampling activity must not only consider the statistical aspects of a representative group of organisms, but also account for ecological and economic factors. Only a thorough study of one organism can provide the guarantee that the investigated organism reflects the desired exposition path at the time of sampling. However, even a complete investigation does not entirely allow for a consideration of all ecological parameters concerning one particular organism. This means that even during the planning of the sampling, its execution must be defined in such a way that its information value does not suffer and that the broader view does not disappear behind the variety of ecological possibilities considered. This is only possible by balancing the heterogeneity of samples taken from one area with homogenizing sampling from similar sites, which show similar microclimatic conditions, and so on.

Although it is seldom mentioned in this context, the economic aspects of sampling should not be neglected. In contrast to sample processing and heavy metal analysis, sampling is, within the biomonitoring activity, the least cost-intensive element. However, as sampling is the first link in a long chain of analytical work, and because errors committed at that point will only become evident after evaluation of the results, ensuring quality at this stage should outweigh other cost considerations.

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Soil Sampling for the Characterization of Hazardous Waste Sites

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in

REFERENCES

1. B. Markert, J. Oehlmann, M. Roth, 'General Aspects of Heavy Metal Monitoring by Plants and Animals', in *Environmental Biomonitoring–Exposure Assessment and Specimen Banking*, eds. K.S. Subramanian, G.V. Iyengar, ACS Symposium Series 654, American Chemical Society, Washington DC, 18–29, 1997.
2. B. Markert, T. Koeckritz, O. Wappelhorst, 'Anforderungen an die Probenahme bei der Ermittlung von Schadstoffen in Pflanzenmatrizes', in *Qualitätssicherung bei Abfällen und Altlasten–Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 53–60, 1998.
3. W.H.O. Ernst, 'Sampling of Plant Material for Chemical Analysis', *Sci. Total Environ.*, **176**, 15–24 (1995).
4. E.I. Hamilton, 'State of the Art of Trace Element Determinations in Plant Matrices–an Overview', *Sci. Total Environ.*, **176**, 3–14 (1995).
5. H. Lieth, B. Markert (eds.), *Element Concentration Cadasters in Ecosystems, Methods of Assessment and Evaluation*, VCH mbH, Weinheim, 1990.
6. F. MacNaeidhe, 'Procedures and Precautions used in Sampling Techniques and Analysis of Trace Elements in Plant Matrices', *Sci. Total Environ.*, **176**, 25–31 (1995).
7. B. Markert (ed.), *Plants as Biomonitors–Indicators for Heavy Metals in the Terrestrial Environment*, VCH mbH, Weinheim, 1993.
8. B. Markert, 'Quality Assurance of Plant Sampling and Storage', in *Quality Assurance in Environmental Monitoring–Sampling and Sample Pretreatment*, ed. Ph. Quevauviller, VCH mbH, Weinheim, 215–254, 1995.
9. B. Markert, O. Wappelhorst, V. Weckert, U. Herpin, U. Siewers, K. Friese, G. Breulmann, 'The Use of Bioindicators for Monitoring the Heavy Metal Status of the Environment', *J. Radioanal. Nucl. Chem.*, **240**(2), 425–429 (1999).
10. G. Wagner, 'Basic Approaches and Methods for Quality Assurance and Quality Control in Sample Collection

and Storage for Environmental Monitoring', *Sci. Total Environ.*, **176**, 63–80 (1995).
11. B. Markert, *Instrumental Element and Multielement Analysis of Plant Samples–Methods and Applications*, John Wiley & Sons, Chichester, 1996.
12. W. Rasemann, 'Qualitätssicherung von Stoffsystemen durch Probenahme und Datenanalyse', in *Qualitätssicherung bei Abfällen und Altlasten–Probenahme und Datenanalyse*, ed. W. Rasemann, Trans Tech Publications, Clausthal-Zellerfeld, 9–21, 1998.
13. W.C. Ma, 'The Influence of Soil Properties and Worm-related Factors on the Concentration of Heavy Metals in Earthworms', *Pedobiologia*, **24**, 109–119 (1982).
14. Å. Rühling, G. Tyler, 'An Ecological Approach to the Lead Problem', *Bot. Notiser*, **121**, 321–342 (1968).
15. T.V. Callagan, N.J. Collins, C.H. Callaghan, 'Photosynthesis, Growth and Reproduction of *Hylocomium splendens* and *Polytrichum commune* in Swedish Lapland', *Oikos*, **31**, 73–88 (1978).
16. W. Lötschert, R. Wandter, 'Schwermetallakkumulation in *Sphagnetum magellanici* aus Hochmooren der Bundesrepublik Deutschland', *Ber. Dtsch. Botan. Ges.*, **95**, 341–351 (1982).
17. E. Steinnes, 'Monitoring of Trace Element Distribution by Means of Mosses', *Fresenius' J. Anal. Chem.*, **317**, 87–97 (1984).
18. Å. Rühling, L. Rasmussen, K. Pilegaard, A. Mäkinen, E. Steinnes, 'Survey of Atmospheric Heavy Metal Deposition–Monitored by Moss Analyses', *NORD*, 21 (1987).
19. H.B. Ross, 'On the Use of Mosses (*Hylocomium splendens* and *Pleurozium schreberi*) for Estimating Atmospheric Trace Metal Deposition', *Water, Air Soil Pollut.*, **50**, 63–76 (1990).
20. K. Grodzinska, G. Szarek, B. Godzik, S. Braniewski, E. Chrzanowska, 'Air Pollution Mapping in Poland by Heavy Metal Concentration in Moss, Proceedings Polish-American Workshop: Climate and Atmospheric Deposition Monitoring Studies in Forest Ecosystems', Nieborow, 6–9 October, 1993.
21. G. Wagner, M. Altmeyer, R. Klein, M. Paulus, J. Sprengart, 'Richtlinien der Umweltprobenbank des Bundes zur Probenahme und Probenbearbeitung–Fichte (*Picea abies*) und Kiefer (*Pinus sylvestris*)', Federal Ministry of the Environment, Nature Conversation and Nuclear Safety, Bonn, 1993.
22. B. Markert, V. Weckert, 'Time- and Site-integrated Long-term Biomonitoring of Chemicals by Means of Mosses', *Toxicol. Environ. Chem.*, **40**, 43–56 (1993).
23. B. Markert, V. Weckert, 'Higher Lead Concentrations in the Environment of Former West Germany after the Fall of the Berlin Wall', *Sci. Total Environ.*, **158**, 93–96 (1994).
24. G. Schüürmann, K.D. Wenzel, L. Weissflog, K. Wienhold, E. Müller, 'Ökotoxikologische Charakterisierung der Schwermetall Immissionsmuster', *UWSF – Z. Umweltchem. Ökotox.*, **6**, 265–270 (1994).
25. M. Paulus, R. Klein, M. Zimmer, J. Jacob, M. Rossbach, 'Die Rolle der Biometrischen Probencharakterisierung in der Umweltanalytik am Beispiel der Fichte (*Picea abies*)', *UWSF – Z. Umweltchem. Ökotox.*, **7**(4), 236–244 (1995).
26. U. Herpin, J. Berlekamp, B. Markert, B. Wolterbeck, K. Grodzinska, U. Siewers, H. Lieth, V. Weckert, 'The Distribution of Heavy Metals in a Transsect of the Three States the Netherlands, Germany and Poland', *Sci. Total Environ.*, **187**, 185–198 (1996).
27. E. Gjengedal, E. Steinnes, 'Uptake of Metal Ions in Moss from Artifical Precipitation', *Environ. Monitoring Assess.*, **14**, 77–87 (1990).
28. U. Herms, G. Brümmer, 'Löslichkeit von Schwermetallen in Siedlungsabfällen und Böden in Abhängigkeit vom pH-Wert, Redoxbedingungen und Stoffbestand', *Mitt. Deutsch. Bodenkdl. Ges.*, **27**, 23–43 (1978).
29. J. Thöming, W. Calmano, 'Applicability of Single and Sequential Extractions for Assessing the Potential Mobility of Heavy Metals in Contaminated Soils', *Acta Hydrochim. Hydrobiol.*, **26**(6), 338–343 (1998).
30. D.M. Di Toro, J.D. Mahony, D.J. Hansen, W.J. Berry, 'A Model of the Oxidation of Iron and Cadmium Sulfide in Sediments', *Environ. Toxicol. Chem.*, **15**, 12, 2168–2186 (1996).
31. M. Roth-Holzapfel, 'Multi-element Analysis of Invertebrate Animals in a Forest Ecosystem (*Picea abies* L.)', in *Element Concentration Cadasters in Ecosystems, Methods of Assessment and Evaluation*, eds. H. Lieth, B. Markert, VCH mbH, Weinheim, 281–295, 1990.
32. U. Rink, 'Experimente mit Freiland Containern zur Wirkung von Kompensationskalkung auf den Abbau bleibelasteter Waldstreu und auf Regenwürmer', *Mitt. Deutsch. Bodenkdl. Ges.*, **69**, 135–138 (1993).
33. V. Nuutinen, 'Earthworm Community Response to Tillage and Residue Management on Different Soil Types in South Finland', *Soil Tillage Res.*, **23**, 221–239 (1992).
34. J. Terhivuo, E. Pankakoski, H. Hyvärinen, I. Koivisto, 'Pb Uptake by Ecologically Dissimilar Earthworm (*Lumbricidae*) Species Near a Lead Smelter in South Finland', *Environ. Pollut.*, **85**, 87–96 (1994).
35. W.C. Ma, 'Methodological Principles of Using Small Mammals for Ecological Hazard Assessment of Chemical Soil Pollution, with Examples on Cadmium and Lead', in *Ecotoxicology of Soil Organisms*, eds. M.H. Donker, H. Eijsacker, F. Heimach, Lewis Publishers, Boca Raton, 267–276, 1994.
36. E. Steinnes, O. Johansen, O. Røyset, M. Ødegard, 'Comparison of Different Multielement Techniques for Analysis of Mosses used as Biomonitors', *Environ. Monitoring Assess.*, **25**, 87–97 (1993).
37. T. Berg, O. Røyset, E. Steinnes, 'Moss (*Hylocomium splendens*) Used as a Biomonitor of Atmospheric Trace Element Deposition: Estimation of Uptake Efficiencies', *Atmos. Environ.*, **29**(3), 353–360 (1995).
38. O. Wappelhorst, 'Charakterisierung atmosphärischer Depositionen in der Euroregion Neiße durch ein

terrestrisches Biomonitoring', Dissertation, International Graduate School Zittau, 1999.
39. M. Begon, J.L. Harper, C.R. Townsend, '*Ökologie. Individuen, Populationen und Lebensgemeinschaften*, Birkhäuser Verlag, Basel, 613, 1991.
40. M. Alikhan, V. Storch, 'Copper and Nickel Uptake and Accumulation, and their Effects on Redox and Electrical Potentials of Hepatopancreas Cells of *Oniscus asellus* Linneaus (Porcellionidae, Isopoda)', *Can. J. Zool.*, **68**, 651–655 (1990).
41. R. Engelke, 'Schwermetallgehalte in Laubmoosen Des Hamburger Stadtgebietes und Untersuchungen zur Sensibilität bei experimenteller Belastung', Dissertation, University of Hamburg, 1984.
42. C. Andersen, J. Laursen, 'Distribution of Heavy Metals in *Lumbricus terrestris*, *Aporrectodea longa* and *A. rosea* Measured by Atomic Absorption and X-ray Fluorescence Spectrometry', *Pedobiologia*, **24**, 347–356 (1982).
43. U. Schulte-Oehlmann, *Fortpflanzungsstörungen bei Süß- und Brackwasserschnecken, Einfluß der Umweltchemikalie Tributylzinn*, Wissenschaft und Technik Verlag, Berlin, 1997.
44. L. Weissflog, K. Wienhold, K.D. Wenzel, G. Schüürmann, 'Ökologische Situation der Region Leipzig-halle. I. Immissionsmuster luftgetragener Schwermetalle und Bioelemente', *UWSF – Z. Umweltchem. Ökotox.*, **6**(2), 75–80 (1994).
45. W. Lötschert, R. Wandter, H. Hiller, 'Schwermetallanreicherung bei Boden Moosen in Immissionsgebieten', *Ber. Dtsch. Botan. Ges.*, **88**(3), 419–431 (1975).
46. T. Koeckritz, 'Bioakkumulation von Schwermetallen in Bodenorganismen am Beispiel von *Aporrectodea caliginosa* (Oligochaeta, Lumbricidae) an unterschiedlich belasteten Standorten in Schleswig-Holstein', Degree Dissertation, Christian Albrechts University in Kiel, 1996.
47. H.G. Dässler (ed.), *'Einfluß von Luftverunreinigungen auf die Vegetation*, Gustav Fischer Verlag, Jena, 1991.
48. E. Rohbock, 'Atmospheric Removal of Airborne Metals by Wet and Dry Deposition', in *Deposition of Atmospheric Pollutants*, eds. H.W. Georgii, J. Pankrath, Reidel, Dordrecht, 159–171, 1982.
49. S.S. Groet, 'Regional and Local Variations in Heavy Metal Concentrations in Bryophytes in the North-eastern United States', *Oikos*, **27**, 445–456 (1976).
50. Å. Rühling, G. Tyler, 'Ecology of Heavy Metals–a Regional and Historical Study', *Bot. Notiser*, **122**, 248–259 (1969).
51. G. Wagner, 'Entwicklung einer Methode zur Großräumigen Überwachung mittels standardisierter Pappelblätterproben von Pyramidenpappeln (*Populus nigra* Italica) am Beispiel von Blei, Cadmium und Zink', in *Beiträge zur Umweltprobenbank*, Nr. 5, eds. M. Stoeppler, H.W. Dürbeck, Jül. Spez., 412–424, 1987.
52. J.H. Hanway, C.R. Weber, 'Dry Matter Accumulation in Eight Soybean (*Glycine Max* (L.) MERRILL) Varieties', *Agron. J.*, **63**, 227–230 (1971).
53. D.A. Ashley, 'Soybean', in *Crop–Water Relations*, eds. I.D. Teare, M.M. Peet, John Wiley & Sons, New York, 389–422, 1983.
54. J.R. Packham, A.J. Willis, 'The Effects of Shading on *Oxalis acetosella*', *J. Ecol.*, **65**, 619–642 (1977).
55. J.L. Harper, J. Ogden, 'The Reproductive Stragtegy of Higher Plants. I. The Concept of Strategy with Special Reference to *Senecio vulgaris* L.', *J. Ecol.*, **58**, 681–698 (1970).
56. U. Herpin, 'Moose als Bioindikatoren von Schwermetalleinträgen–Möglichkeiten und Grenzen für flächenhafte und zeitabhängige Aussagen', Dissertation, University of Osnabrück, 1997.
57. W.H.O. Ernst, 'Ökologische Risiken verursacht durch emittierte Feinstäube aus Verbrennungsmotoren. Folgerungen für die Analytische Chemie', *Chem. Rundschau*, **31**, 32–35 (1978).
58. J.R. Porter, 'Evaluation of Washing Procedures for Pollution Analysis of *Ailanthus altissima* Leaves', *Environ. Pollut., Ser. B*, **12**, 195–202 (1986).
59. T. Berg, E. Steinnes, 'Use of Mosses (*Hylocomium splendens* and *Pleurozium schreberi*) as Biomonitors of Heavy Metal Deposition: From Relative to Absolute Deposition Values', *Environ. Pollut. 98*, **1**, 61–71 (1997).
60. E. Steinnes, 'A Critical Evaluation of the Use of Naturally Growing Moss to Monitor the Deposition of Atmospheric Metals', *Sci. Total Environ.*, **160/161**, 243–249 (1995).
61. B. Markert, V. Weckert, 'Fluctuations of Element Concentrations during the Growing Season of *Polytrichum formosum* (Hedw.)', *Water, Air Soil Pollut.*, **43**, 177–189 (1989).
62. C. Polte-Rudolf, 'Frühlingseinzug in Europa', *DWD, Phän. J.*, **5**(3) (1999).
63. H.Th. Wolterbeek, P. Kuik, T.G. Verburg, U. Herpin, B. Markert, L. Thöni, 'Moss Interspecies Comparisons in Trace Element Concentrations', *Environ. Monitoring Assess.*, **35**, 263–286 (1995).
64. A.J.M. Baker, 'Accumulators and Excluders–Strategies in the Response of Plants to Heavy Metal', *J. Plant Nutr.*, **3**, 643–654 (1981).
65. G. Caughley, 'Sampling in Aerial Survey', *J. Wildlife Manag.*, **41**, 605–615 (1977).
66. L.L. Eberhardt, 'Transect Methods for Population Studies', *J. Wildlife Manag.*, **42**, 1–31 (1978).
67. C.J. Krebs, *'Ecological Methodology'*, Harper Collins, New York, 1989.
68. A. Rühling, 'Atmospheric Heavy Metal Deposition in Europe–Estimation Based on Moss Analysis, *NORD*, 9, (1994).
69. F. Tataruch, 'Die Belastung freilebender Wildtiere mit Umweltschadstoffen', in *Kind und Umwelt, Band 4, Verlag Kind und Umwelt*, Zürich, 111–139, 1996.
70. U. Herpin, B. Markert, U. Siewers, H. Lieth, 'Monitoring der Schwermetallbelastung in der Bundesrepublik Deutschland mit Hilfe von Moosanalysen', Forschungsbericht, 108 02 087, UBA-FB 94–125, Berlin, 1995.

Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses

Nancy J. Miller-Ihli and Scott A. Baker
Beltsville Human Nutrition Research Center, Beltsville, USA

1 Introduction 1
2 Evolution of Slurry Methods 2
3 Optimization of Graphite Furnace Atomic Absorption Spectrometry Methods for Slurry Analyses 3
3.1 Sample Grinding 3
3.2 Optimizing Slurry Preparations 4
3.3 Slurry Preparation 6
3.4 Slurry Mixing and Sampling 7
3.5 Operating Conditions 7
3.6 Number of Determinations 8
4 Environmental Applications 9
4.1 Characteristic Masses, Detection Limits and Ranges 9
4.2 Analytical Performance – Benefits and Limitations 11
4.3 Comparison of Sample Types 12
5 Comparison with Other Direct Analysis Methods 13
6 Future Expectations 15
Disclaimer 15
Abbreviations and Acronyms 15
Related Articles 16
References 16

Slurry sampling graphite furnace atomic absorption spectrometry (GFAAS) is a powerful tool for the determination of trace elements in environmental samples owing to its high sensitivity, instrumental simplicity, and relatively low cost. In this technique, discrete aliquots of the slurry or suspension are deposited in a graphite tube and then electrically heated to produce atomic vapors. The absorption of the atomic vapor is measured, typically with a hollow cathode lamp (HCL) or electrodeless discharge lamp (EDL), to determine the amount of analyte present in the sample. To ensure that a representative portion of the slurry is injected into the furnace, the slurry must be stabilized or mixed immediately prior to analysis. Slurry sampling GFAAS detection limits (DLs) are typically in the range of 10^{-2} to 1 part per million (pg absolute DLs). Precision (0.5–5%) and accuracy are also good, provided that proper guidelines for slurry preparation and instrument operation are followed, and samples are sufficiently homogeneous. An additional benefit of the technique is that matrix interferences are typically small; in fact, aqueous calibration standards are routinely used in the analysis of slurry samples. A limitation of slurry sampling GFAAS is that it is primarily a single-element technique, although continuum source and sequential atomic absorption spectrometry (AAS) systems have been developed. Furthermore, the linear range is typically less than two orders of magnitude unless alternative wavelengths, or some other means of reducing sensitivity (e.g. minigas flow during atomization), are used. In addition, some environmental samples have to be ground if the sample is not in the proper form or if the analyte is not homogeneously distributed. Care must be taken during the grinding process to avoid contamination of the sample or fractional distribution of the analyte.

1 INTRODUCTION

A growing awareness of the effects of trace elements on human health has fostered considerable interest in the analysis of a wide variety of environmental samples.[1] In many instances, the analyses are directed towards identifying and quantifying elemental pollutants at high levels; however, the establishment of baseline levels is also important since this information can be used to reconstruct the past distribution of toxic pollutants on both local and global scales. The source and extent of contamination can then be more easily determined. Based on these considerations, analytical techniques used in environmental analysis must be capable of measuring elemental concentrations over a wide range of concentrations (parts per million to parts per trillion levels) in samples from all phases of the environment (air, ocean waters, soils, etc.). Accurate and reliable methods, which are cost-effective, easily implemented, and, perhaps most importantly, applicable to real world samples are required. Obviously, no single analytical technique is ideally suited for such a diverse range of samples; however, GFAAS is one of several techniques that have enjoyed widespread application in the field of environmental analysis.[1] This is because the technique provides a relatively simple, inexpensive, and highly sensitive means of analysis. In addition, it offers some versatility since samples can be introduced to the graphite furnace in either solution or solid form.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Traditionally, samples have been introduced into the graphite furnace as solutions. This is particularly limiting since a large percentage of environmental samples are originally solids and require considerable processing to dissolve them prior to analysis. Over the past decade, solid sampling techniques for trace element determinations in environmental samples have increased in popularity owing to a number of potential advantages. The benefits include: reduced sample preparation time, since digestion procedures required for many environmental samples can be quite lengthy and complex; decreased analyte loss through volatilization prior to analysis or because of retention by an insoluble fraction of the sample matrix; reduced risk of sample contamination because less sample handling and fewer reagents are required; elimination of hazards associated with the use of acids and other corrosive reagents; and the potential for improved sensitivity since the sample is not diluted. Unlike nebulization techniques, GFAAS does not suffer significantly from particle size effects because it offers longer residence times and correspondingly higher atomization efficiencies. In addition, direct solids analysis by GFAAS provides unique information to the analyst with regard to the distribution of analytes in microamounts of sample and is particularly useful when only small quantities of sample are available for analysis.

The two principal methods for analyzing solids by GFAAS are to introduce the solid material directly into the furnace or to prepare a slurry or suspension, then inject the sample by using conventional liquid sampling methodology. Applications and the relative merits of both direct solid sampling and slurry sampling can be found in reviews by Bendicho and de Loos-Vollebregt[2] and Miller-Ihli.[3] For samples already in powder form (e.g. soils and sediments) or samples that can be efficiently ground, slurry sampling offers a number of advantages over direct solid sampling. Slurry sample introduction is straightforward and can be automated using commercial liquid autosamplers. This is advantageous for routine determinations and is a more reproducible means of sample introduction than weighing and transferring the sample, as is done with direct sample introduction. Slurry samples may also be ''diluted'' when the analyte concentration is too high to facilitate analysis in the linear range; however, consideration must be given to ensure that enough particles are still being sampled. With direct solids analysis, the sample can, in some cases, be diluted with graphite powder; however, this increases the risk of contamination. Both methods may take advantage of other methods for reducing sensitivity, such as choosing a less sensitive absorption line. This does, however, require additional optimization. Slurry sampling also allows for the addition of matrix modifiers directly to the slurry or suspension and calibration is typically easier, since aqueous standards can often be used. This is not the case for direct solid sampling, where matrix-matched calibration standards are routinely required to achieve accurate results.

The development and optimization of slurry sampling GFAAS methods for environmental samples requires consideration of a number of key issues. These include sample preparation protocols (e.g. grinding of samples and mixing of slurries), representative sampling of the slurry, matrix modification, optimization of furnace conditions, wavelength selection, and calibration. With careful consideration of these variables, slurry sampling GFAAS provides a unique and powerful tool for the analysis of environmental samples.

2 EVOLUTION OF SLURRY METHODS

Slurry sampling GFAAS has matured significantly since the first reports by Brady et al.[4,5] Obviously, one of the major concerns with this methodology is the necessity of introducing a representative sample aliquot into the furnace. This can only occur if the slurry is efficiently stabilized or homogenized during the time needed for sample introduction. Stabilization has been attempted with thixotropic agents, such as viscalex[6] and glycerol;[7] however, workers reported problems with reproducible pipetting because of material adherence to the exterior of the autosampler capillary and incomplete delivery of the sample with these high viscosity mediums. For this reason, sample homogenization using some form of physical agitation is preferred. Magnetic stir bar mixing,[8,9] vortex mixing,[10,11] gas bubbling,[12,13] and ultrasonic agitation[14,15] have all been employed for slurry homogenization.

Slurry analyses utilizing magnetic stir bar mixing and manual pipetting are often reported in the literature. Although several workers have successfully used this methodology, it is limited because manual pipetting typically provides poorer precision than can be achieved with autosamplers. In addition, modern analytical laboratories require automated analyses for increased throughput. A system employing miniature magnetic stir bar mixing combined with an autosampler was developed for the analysis of food slurries.[16] A potential problem with using magnetic stir bar mixing, however, is that some samples (e.g. soils) contain particles that adhere to the stir bar because of their magnetic properties.[17]

Vortex mixing is effective in agitating slurry samples before analysis, but sample aliquots must be removed quickly and deposited into the furnace to avoid sedimentation of larger and/or denser particles that may occur. If a significant portion of the analyte is associated with these

particles, slurry GFAAS results will be systematically low.[7] Another major disadvantage of vortex mixing for slurry analysis is that it is not compatible with autosampler operation and is therefore of limited utility for routine determinations.

Homogenization of slurries using Ar gas bubbling has been described in the literature.[12,13] A benefit of this methodology is that slurry mixing can be performed directly in an autosampler cup, thus allowing for automated analysis. The major disadvantage of gas bubbling is that the effectiveness is largely dependent on the particle size and density. Bendicho and de Loos-Vollebregt reported poor homogenization when using gas bubbling for glass slurries in the 82–341 μm range.[12] Similarly, Lopez-Garcia et al. found it necessary to modify the suspension medium to minimize sedimentation of particles in the determination of Pb in a marine sediment sample.[13]

Ultrasonic agitation with a compact probe is one of the most effective means of slurry homogenization.[14,18] Ultrasound can effectively break up particle clusters, disperse solids, and increase wettability, leading to improved slurry stability. In addition, the ultrasonic action can increase the extraction of analyte into the aqueous phase of the slurry, which typically results in improved precision. Importantly, ultrasonic slurry mixing can be performed directly in an autosampler cup, allowing for completely automated operation.[14] Of the mixing methods, only ultrasonic agitation has been employed in a commercial autosampler accessory (USS-100, Perkin–Elmer). The widespread acceptance of ultrasonic mixing for slurry preparation is evidenced by an international collaborative study in which 14 of the 16 collaborators used ultrasonic devices.[19]

Representative sampling is obviously crucial to the success of slurry GFAAS in solids analysis; however, a number of additional issues must be considered when introducing solid material into the furnace. In comparison to solutions, solids pose a number of unique challenges in GFAAS analysis. Historically, these challenges have included high background levels associated with the solid matrix, difficulty in calibration, difficulty with refractory elements, build-up of residual material in the furnace, sample-dependent peak shapes, and particle size dependence. Undoubtedly, one of the most significant developments in the evolution of slurry GFAAS for the analysis of solids was the introduction of the stabilized temperature platform furnace concept (STPF).[20,21] The STPF approach employs a platform furnace so that the analyte is atomized in a more isothermal environment. STPF also provides for matrix modification for analyte stabilization, rapid furnace heating, pyrolytically coated tubes to reduce chemical interferences, good background correction (such as the Zeeman effect), no gas flow during atomization, and integrated absorbance measurements for quantitation. Employing these conditions significantly improves the quality of slurry GFAAS analyses, allowing the technique to be routinely used for the determination of trace elements in solids. This was demonstrated in an international collaborative study where slurry GFAAS was used to analyze a variety of materials. It was concluded that the technique could be used for routine quantitative analysis, as long as the slurry analysis conditions were optimized and STPF conditions were employed.[19,22]

3 OPTIMIZATION OF GRAPHITE FURNACE ATOMIC ABSORPTION SPECTROMETRY METHODS FOR SLURRY ANALYSES

3.1 Sample Grinding

Slurries are typically prepared from finely ground, powdered material. Unfortunately, not all samples received are in a form suitable for slurry preparation. As a result, grinding of the sample is necessary prior to analysis. There are a variety of grinding methods available, depending on the material hardness and required particle size. Although several workers have reported the need to work with relatively small particles for slurry GFAAS, accurate analyses have been reported using particles up to 300–400 μm.[15,23] A general grinding method that is suitable for a variety of environmental materials is the use of a vibratory ball mill, such as the Spex Mixer/Mill (Spex, Metuchen, NJ, USA) or the Retsch Spectro Mill (K. Retsch Co., GmbH, Germany). These devices are available in an assortment of materials, such as stainless steel, agate, zirconia, tungsten carbide, boron carbide, and silicon nitride. To avoid contaminating the sample with the analytes of interest, it is extremely important to choose the proper grinding material.[24] Stainless steel mills should be avoided if at all possible. The use of ceramics offers much better abrasion resistance and reduced likelihood of contamination. Common contaminants, and their relative levels, introduced from various grinding materials can be found in the Spex Handbook.[24] Grinding times are dependent on the sample type, but typically are less than 30 min. For soils, Dobrowolski reported that 15 min of grinding (tungsten carbide ball mill) was sufficient to produce samples with 65–70% of the particles below 20 μm.[25]

Wet grinding methods have also been used for reducing particle sizes in sediment[26] and botanical samples.[27] Grinding was done in this laboratory[27] by adding beads (high purity zirconia, 2 mm in diameter) to the sample, which was in a polyethylene bottle, in a 10 : 1 weight ratio

(beads : sample). Approximately 10–15 mL of water was added to the bottles and the mixture was shaken for 2 h. Significant levels of contamination were observed for Fe, Cr, and Al with this grinding method.

A method for grinding of biological and botanical samples, using Teflon® beads and polyethylene bottles, has been described.[15] Samples (1–5 g) and Teflon® beads (12.7 mm in diameter) were placed in 125-mL polyethylene bottles with 15 mL of water (or 5% HNO_3). The mixture was shaken on a wrist action shaker for 20–60 min and then filtered through a coarse mesh (800 + μm) polyethylene screen to remove the Teflon® beads. Blanks for a range of analytes (Cu, Mn, Fe, Cr, Co, Pb, V, and Mo) were evaluated and indicated no significant contamination (<0.2 ppb). The data indicated that a 40-min grinding time was sufficient to produce samples with 85–90% of the particles less than 125 μm in diameter.[15] The procedure was evaluated for slurry GFAAS analysis of several in-house control food samples and a National Institute of Standards and Technology (NIST) reference material (RM 8431 Diet) and the results compared favorably with the certified reference concentrations.

A common error when analyzing solid materials is to sieve them to provide a uniform fraction for analysis. While this might appear to be advantageous, it should be avoided if the goal of the analysis is to provide an accurate representation of the original solid. For example, if the analyte is associated with a particular particle size or type, the results will be biased if any of the sample is discarded.

3.2 Optimizing Slurry Preparations

Several factors must be considered when preparing slurries for GFAAS analysis. An obvious consideration for slurry preparation is the amount of material that will be injected into the furnace for analysis. This can be calculated using Equation (1):

$$M_F = \frac{M_S}{V_S} \times V_F \quad (1)$$

where M_F is the mass injected in milligrams, M_S is the mass of sample used to prepare the slurry in milligrams, V_S is the volume of liquid (diluent) used to prepare the slurry in milliliters, and V_F is the volume of the slurry injected into the furnace in milliliters. For a slurry prepared using 10 mg of solid sample in 1 mL of diluent, a typical 20-μL injection will introduce 200 μg of solid into the furnace. With some knowledge of the expected analyte concentration, one can optimize the slurry preparation, in accordance with conventional GFAAS protocol, to ensure that the concentration falls within the optimum range for calibration on the working curve. In addition, quantitative determinations should be performed at levels that are at least five times greater than the DL. The more material used to prepare the slurry for a fixed volume of diluent and/or the greater the extraction of analyte into the liquid phase of the slurry, the more representative the determined concentration will be to the concentration in the original sample. If analyte concentrations are too high to work in the optimum calibration range, the slurry can be diluted, smaller injections can be used, or GFAAS conditions can be modified to reduce the sensitivity (e.g. use alternative wavelength). The effects of slurry dilution and smaller injection volumes will be discussed shortly, whereas GFAAS wavelength selection and other means of reducing sensitivity will be discussed later in the article.

Although the concentration of analyte in the slurry is important, other interrelated factors of interest in optimizing slurry preparations include particle size, material density and homogeneity, diluent viscosity and density, maximum slurry concentration, and the distribution of analyte (solid to liquid) in the slurry. Holcombe and Majidi have characterized errors associated with slurry sampling, considering the volume of sample, the number of particles in the sample volume, and variations in the mass of individual particles (volumetric errors); as well as sedimentation errors, which are dependent on the particle size, density of the sample and diluent, and viscosity of the diluent.[28,29] These workers concluded that volumetric errors can be minimized by working with small particles, concentrated slurries, and narrow particle size distributions; whereas sedimentation errors can be controlled by optimizing the sampling time window for a given matrix. The homogeneity of the material is a major consideration for slurry preparation. If the analyte is not homogeneously distributed in the solid material, grinding to a very small particle size (<10 μm) and increasing the number of particles in the sample will increase the likelihood that the measured concentration will be representative of the original material.

Calculating the number of particles in the injection volume is useful to ensure that sampling uncertainty is not a major source of imprecision. Assuming that the uncertainty associated with pipetting of the slurry is ±1 particle, and assuming that all particles contain representative amounts of the analyte, then a sample volume containing 50 particles would have a sampling uncertainty of less than 2%. The number of particles in the sample volume can be calculated based on the material density and particle size as given in Equation (2):

$$N_P = \frac{M_S}{\rho \times V_P} \quad (2)$$

where N_P is the number of particles in the injection volume, M_S is the mass of sample used to prepare the slurry in grams, ρ is the density of the material in grams per cubic centimeter, and V_P is the volume of the particle

Table 1 Slurry preparation guidelines

Minimum mass (mg)[a]	Particle size (diameter, μm)	Density ($g\,cm^{-3}$)
<0.05	25	0.5
0.08	50	
0.65	100	
10.2	250	
81.8	500	
<0.05	25	1.0
0.16	50	
1.30	100	
20.4	250	
163	500	
<0.05	25	1.5
0.24	50	
1.95	100	
30.6	250	
245	500	
<0.05	25	2.0
0.32	50	
2.60	100	
40.8	250	
327	500	
<0.05	25	2.5
0.40	50	
3.25	100	
51.0	250	
408	500	

[a] Minimum mass required to prepare a 1 mL slurry to produce 50 particles in a 20 μL injection.

Table 2 Densities of some common materials

Material	Density ($g\,cm^{-3}$)
Clay	1.8–2.6
Bone	1.7–2.0
Glass	2.4–2.8
Coal (anthracite)	1.4–1.8
Coal (bituminous)	1.2–1.5
Talc	2.7–2.8
Estuarine Sediment (SRM 1646)	2.47
Coal Fly Ash (SRM 1633a)	2.2
Peach Leaves (SRM 1547)	1.4
Oyster Tissue (SRM 1566a)	0.28

in cubic centimeters. Table 1 shows the minimum mass required to prepare a 1-mL slurry which will contain 50 particles in a 20-μL injection volume as a function of the material density and particle size. In all cases, the particles were assumed to be spherical. It is clear that the particle size has the largest influence on the required mass, since the volume has a cubic dependence on the particle radius. Densities for several common materials are provided in Table 2. To experimentally determine densities for powdered samples, a simple method has been described.[15] Small autosampler cups (0.5 mL) were filled with material and the mass recorded. Then, the dead volume was determined by adding surfactant to fill the cup and recording the volume added. The mass of sample per corrected volume (known cup volume − dead volume) was then used to calculate the density of the solid.

Although working with concentrated slurries is beneficial in reducing volumetric errors and problems related to heterogeneity, there is a practical limit to the amount of material that can be pipetted. The volume of solid to volume of liquid in a slurry preparation (volume to volume ratio) can be so high that the autosampler cannot pipette such a viscous slurry. This ratio can be calculated from Equation (3):

$$\frac{V_{\text{Solid}}}{V_{\text{Liquid}}} = \frac{M_S/\rho}{V_S} \tag{3}$$

where M_S is the mass of sample used to prepare the slurry in grams, ρ is the material density in grams per cubic centimeter, and V_S is the volume of liquid used to prepare the slurry in cubic centimeters. A volume/volume ratio of 0.25 or less is typically acceptable for reliable and efficient pipetting.[15,30] For a material with a density of $1\,g\,cm^{-3}$, this corresponds to a slurry prepared with as much as 250 mg diluted to 1 mL. For materials of lower density, the maximum sample mass decreases linearly with density for a fixed diluent volume.

It is important to have some knowledge of the analyte distribution in the liquid and solid phases of the slurry to derive the maximum benefit from the slurry GFAAS technique. Particle size and the type of diluent influence the amount of analyte extracted into the liquid phase. In addition, different amounts of an analyte can be extracted depending on whether the analyte is occluded in the particle or if it is simply adsorbed on the surface of the particle. The use of acids in the diluent generally increases the analyte extraction efficiency. Nitric acid is most commonly used in the diluent; however, some workers[12,13,31–33] have used hydrofluoric acid to increase the extraction efficiency for silicate-rich matrices, such as soils and sediments. Based on the authors' experience, HNO_3 concentrations between 2.5% and 5.0% result in approximately equal extraction efficiencies and have proven useful for a variety of sample types. The use of a more concentrated HNO_3 solution does not significantly increase the extraction efficiency and leads to rapid degradation of the graphite tube. Extraction efficiencies are measured by preparing the slurries and determining the analyte content, then separating the solid and liquid phases by ultracentrifugation and measuring the analyte content of the liquid phase.[24,25,34–35] Benefits of extraction are that it can lead to precision that approaches that obtainable with conventional liquid sample introduction and can also lead to analyses that are more representative of the original solid. It is important

to mention, however, that significant analyte extraction is not a requirement for accurate and precise analysis with slurry GFAAS.

The representative sample mass concept is useful in determining how much of the original sample is represented by a slurry analysis.[23] The two extremes of the concept are where no analyte is extracted and where 100% extraction occurs. In the case of no extraction, the representative sample mass is only the amount of solid injected into the furnace. For a 2 mg of solid in 1 mL of slurry preparation and a 20-μL injection volume, this corresponds to 0.04 mg. If 100% extraction occurs, then the analysis of any volume of the sample is representative of the entire 2 mg of sample used to make the slurry. Most often, the amount of analyte extracted lies somewhere in between the two extremes. If it is assumed that the analyte is distributed homogeneously in the solid and that the percentage extracted into the liquid phase is representative of the sample mass dissolved, then the sample mass represented by slurry analyses can be calculated using Equation (4):

$$M_a = V_F \times \frac{M_s}{V_S} \times (1 - f_x) + f_x \times M_s \qquad (4)$$

where M_a is the representative sample mass in milligrams, V_F is the volume injected into the furnace in milliliters, M_s is the mass of sample used to prepare the slurry in milligrams, V_S is the volume of the slurry preparation in milliliters, and f_x is the fraction of analyte in solution.[30] For the 2-mg slurry prepared in 1 mL, 5% analyte extraction would result in $M_a = 0.14$ mg, 20% analyte extraction would result in $M_a = 0.43$ mg, and 50% analyte extraction would result in $M_a = 1.02$ mg. Representative sample mass values for three apparently equal mass to volume slurry preparations are given in Table 3. Provided that analyte extraction occurs to some extent, the benefit of using larger masses of sample to increase the amount of sample represented by the analysis is clearly evident. If extraction does not occur, no benefit is gained by increasing the sample mass. Improvements in measurement precision with increasing M_a values have been reported for various sample types.[23,35] The representative sample mass concept also clearly illustrates the problem with discussing slurry preparations using terminology such as '0.2%', which can be found in the literature. It should be noted that M_a is only an indicator of the extent of sample dissolution and resultant analyte extraction. Because calculations of M_a are based on the assumption that the analyte is homogeneously distributed and the percentage of analyte extracted is representative of the amount of solid dissolved, it is not an absolute quantity. For example, if the analyte is adsorbed on the surface of the solid, the extraction efficiency could far exceed the actual amount of material dissolved.[23] A safe assumption is that the quantity of material represented by the analysis is not larger than the calculated M_a value.

Optimization of the slurry preparation obviously requires considerable planning to ensure that accurate and precise measurements are obtained. Each of the above considerations should be thoroughly examined. A sufficient quantity of material should be weighed out to prepare a homogeneous slurry, the analyte concentration should not exceed the optimum calibration range, a representative number of particles ($\geq$50) should be injected into the furnace, and the slurry should not be too viscous to pipette. Analyte extraction should also be examined since this can lead to improved precision and a better representation of the original sample.

3.3 Slurry Preparation

Slurries should be prepared in accordance with the criteria discussed above. Weighing should be performed on an electronic microbalance. Static electricity can be problematic when weighing small particles, so an ionization source such as the Staticmaster (NRD Inc., Grand Island, NY) is recommended. Samples (typically 1–50 mg) can be weighed directly into autosampler cups to avoid problems that can result from the transfer of small quantities of sample. Autosampler cups should be acid cleaned to minimize trace element contamination and be made of Teflon®. Polypropylene cups might also be acceptable depending on the analytes of interest; however, polystyrene cups should be avoided since significant levels of Fe, Cu, Cr, Al, and Mg have been

Table 3 Representative sample mass (mg) assuming a 20 μL injection volume. Apparently equal mass to volume slurry preparation

Amount extracted (%)	Slurry preparations		
	2 mg per 1 mL	20 mg per 10 mL	200 mg per 100 mL
0	0.04	0.04	0.04
5	0.14	1.04	10.0
20	0.43	4.03	40.0
50	1.02	10.0	100.0
100	2.00	20.0	200.0

observed after ultrasonic mixing.[27] Although a variety of diluents have been used (e.g. HF for silicate-rich matrices), the authors have found a mixture containing 5% (v/v) sub-boiling distilled HNO_3 and 0.005% (v/v) Triton X-100® surfactant to be useful for a wide variety of sample types. Nitric acid facilitates the extraction of analyte into the liquid phase of the slurry; whereas Triton X-100® (Rohm and Haas registered trademark for octylphenoxypolyethoxyethanol, Sigma Chemicals, St. Louis, MO, USA) serves as a wetting agent and assists in particle dispersion. If particle coagulation is not problematic and/or analyte extraction is not desirable (e.g. in homogeneity characterization), distilled deionized water (18 MΩ) can be used as the diluent. In some cases, it might be desirable to prepare larger mass and/or volume slurry preparations. For these larger preparations, sample and diluent can be placed in polyethylene test tubes, avoiding the need for microweighing. Care must be taken with this approach, however, to ensure that a representative subsample is transferred to the autosampler cup for analysis. Vortex mixing has been shown to be effective in homogenizing the slurry while an aliquot is removed.[10]

3.4 Slurry Mixing and Sampling

As previously discussed, several methods have been used for agitation of the slurry mixture. One of the most reliable means of slurry mixing is with an ultrasonic device, such as the automated Perkin–Elmer USS-100, combined with an autosampler (e.g. Perkin–Elmer AS-60). Handheld ultrasonic devices such as the Kontes Micro-ultrasonic cell disrupter (Vineland, NJ) or the Sonics and Materials Model VC-40 (Danbury, CT) can also be used. The power output of these devices should be adjusted so that the slurry is vigorously mixed with a rolling action, but not so high that splattering occurs. The power output of these devices is quite variable and tuning is somewhat subjective; therefore, no exact condition can be provided. In general, the 40-watt mode of operation is used and the power is adjusted to 60–80%. Mixing times of 15–25 s are generally adequate for all materials.[15] If the analyte is easily extracted into the liquid phase, even shorter mixing times can be used. One might expect that shorter mixing times would be required for low-density materials but that is not the case. Because particles with densities less than 1 mg cm^{-3} tend to float, ultrasound is necessary to wet the particles and keep them dispersed in the slurry. The use of a surfactant, such as Triton X-100®, is beneficial for the same reasons. Since the slurry technique can tolerate particle sizes up to several hundred microns, it is advisable that the diameter of the autosampler sampling orifice be checked to ensure that it is sufficiently large. In the authors' laboratory, the standard AS-60 tubing (approximately 300 μm in diameter) was replaced with Teflon® tubing which was 810 μm in diameter.

Although a particular slurry might be sufficiently homogenized during ultrasonication, it is important to consider the influence of particle sedimentation on the analytical results since a finite time exists between the end of the mixing procedure and the pipetting of sample. Potential errors associated with sedimentation can be predicted based on Stokes' Law.[28] The settling velocity for sample particles can be calculated from Equation (5):

$$v = 2gr^2 \frac{\rho_1 - \rho_2}{9\eta} \tag{5}$$

where v is the settling velocity in centimeters per second, g is the acceleration due to gravity (981 cm s^{-2}), r is the particle radius in centimeters, ρ_1 is the density of the solid in grams per cubic centimeter, and ρ_2 is the density of the suspending medium in grams per cubic centimeter, and η is the viscosity of the medium in grams per centimeter second.[35] It is clear from this expression that the sedimentation rate is reduced for smaller particles, when the particle and medium densities are closely matched, and when a high viscosity medium is used. As was previously mentioned, a major problem associated with using high-viscosity mediums, such as glycerol or viscalex, is the inability to pipette them efficiently. Fortunately, it is not necessary to match densities or use high-viscosity mediums, since the sampling depth can be varied in order to remove a representative portion of the slurry. For the 5% HNO_3 + 0.005% Triton X-100® diluent described above, which has a density of 1.02 g cm^{-3} and a viscosity of 0.01002 g cm^{-1} s^{-1}, it was demonstrated that a sampling depth of 7.2 mm or greater was required to remove a representative sample of a high-density ($\rho = 2.6$ g cm^{-3}) glass slurry.[35] Sampling depths of 2.5 mm or less precluded the sampling and measurement of particles 50 μm or greater (78% analyte recovery); whereas a depth of 7 mm did not allow for the sampling and measurement of particles 80 μm or larger (92% analyte recovery). Sampling too deep into the autosampler cup also proved problematic because of the build-up of particles in the conical-shaped bottom of the autosampler cup. With knowledge of sample particle sizes and density, the required sampling depth can be easily calculated to ensure representative sampling of the slurry. For high-density materials, it might be advisable to grind the sample to reduce the particle size and subsequently reduce the settling velocity.

3.5 Operating Conditions

Successful slurry determinations depend on proper selection of GFAAS operating conditions. Typically, operating conditions similar to those used with conventional acid

digests can be employed. The use of STPF conditions as described by Slavin et al.[20,21] is particularly useful because it facilitates the analysis of slurry samples using aqueous calibration standards. This greatly enhances the slurry GFAAS methodology, making it a suitable alternative to conventional acid digests.

As an example, typical conditions for the determination of Cu in a sediment sample are given in Table 4. Wavelength selection is dependent on the expected analyte concentration in the slurry preparation. For high analyte concentrations, less-sensitive nonresonance wavelengths should be used to extend the working range, particularly where there are concerns about the number of particles being injected into the furnace for analysis. Alternatively, the use of a minigas flow (typically 30–300 mL min^{-1}) during atomization has been successfully used to reduce analyte sensitivity in some cases (see for example Lopez-Garcia[33] and Belarra[36]); however, this is generally not recommended since it can lead to non-isothermal atomization conditions.[19,22] A more detailed discussion on minigas flows and other means of sensitivity reduction will be provided later.

Other essential aspects of the STPF concept are platform atomization, fast heating, good background correction, the use of integrated absorbance measurements for quantitation, and, if necessary, the use of a matrix modifier. Platform atomization should be used for all but the most refractory elements since more isothermal atomization conditions are achieved. Because of the large amount of solid injected into the furnace and subsequently large background signals, good background correction such as Zeeman or Smith–Hieftje is essential. The use of integrated absorbance measurements for quantitation eliminates problems associated with matrix dependent peak shapes. This is important if aqueous standards are to be used for calibration of slurried materials.

Table 4 Analytical conditions for the determination of Cu in sediments

Wavelength	Std. concentrations (ng mL^{-1})	m_o^a (pg)
Cu 324.7 nm resonance line	10–50	7.8
Cu 327.4 nm 2 × less sensitive	25–100	13
Cu 216.5 nm 5 × less sensitive	50–500	43
GFAAS heating program		
Temperature (°C)	Ramp time (s)	Hold time (s)
200	40	40
500	20	40
20	1	10
2300	0 (maximum ramp)	8 (no gas flow)
2700	1	5

[a] m_o is the characteristic mass (mass of analyte producing an integrated absorbance signal of 0.0044 s) 20 µL injection volume.

Matrix modification can be useful in stabilizing analytes during the charring step and in eliminating or reducing matrix effects. The use of a traditional matrix modifier (e.g. palladium or magnesium nitrate) is straightforward with slurry GFAAS, since the liquid modifier can be added directly to the slurry preparation. The amount of modifier will typically be larger than those used with liquids because of the larger masses of material in the furnace. Much of the work with chemical modification for slurry GFAAS has been carried out in order to stabilize highly volatile analytes such as Cd,[34,37,38] Pb,[8,16,19,37–39] Se,[32,40] and Hg.[9,26] For samples with a high organic content, air/oxygen ashing can be used to remove carbonaceous material prior to atomization.[23,41,42] This not only prevents residue from building up inside the furnace, but can also reduce background levels and minimize matrix effects. The addition of hydrogen peroxide to the slurry has also been used for the removal of carbonaceous material in the furnace.[37,43] For samples with a high silica content, the use of HF in the slurry diluent has been reported.[9,12,31–33] Some of the reported benefits of using HF are increased analyte extraction, the removal of silicon (as SiF_4) during the heating cycle which minimizes damage to the graphite tube and allows larger masses to be injected, and reduced matrix suppression effects because of the partial digestion of the matrix.[9]

The need for a matrix modifier is determined from systematic optimization of the drying, charring, and atomization conditions. It is important that these conditions are optimized for each sample type, since the matrix can significantly affect the volatilization characteristics of the analyte. In some instances, fast-temperature programs, in which the conventional drying and ashing steps are replaced by a single modified drying stage, have been used successfully.[9,32,33,43–45] The major benefit of fast-temperature methodology is a significant reduction in analysis time. It has been used for the determination of As, Pb, Tl, and Se in coal and coal fly ash;[45] Pb,[33,44] Cd,[33] Se,[32] and Hg,[9] and Tl[33] in soil; and Pb and Cd in vegetables.[43] Although chemical modifiers are often unnecessary since a conventional ashing stage is not used and the risk of analyte loss is small, they may be useful in delaying atomization of the analyte in the furnace until more isothermal conditions are achieved.[33]

3.6 Number of Determinations

A problem with solid sampling is the possibility of not analyzing a representative portion of the bulk sample.

Table 5 Recommended procedure for ultrasonic slurry GFAAS determinations

1. Grind the sample, if necessary, to produce a powder
   - particle sizes up to 300–500 μm may be acceptable
   - grinding methods: mortar grinding and Teflon® beads and polyethylene bottles (biological and botanical samples); grinding, pulverizing, or vibratory ball mills
   - minimize contamination risks (no stainless steel)
   - do not sieve samples
2. Optimize slurry preparations
   - calculate minimum mass required for analysis based on particle size and density (≥50 particles) and select appropriate analytical wavelength
   - factors of interest include homogeneity of solid, distribution of analyte in solid, sample density, particle size, diluent viscosity and density, and analyte partitioning in the slurry (calculate representative sample mass)
   - if analyte is heterogeneously distributed in the solid, strive for very small particles (<10 μm)
   - calculate the (volume of solid/volume of diluent) to ensure the ratio is ≤0.25
3. Prepare slurries for analysis
   - weigh samples using electronic microbalance (avoid static with ion source)
   - 1–50 mg of ground material may be weighed directly in a Teflon® autosampler cup
   - Add 1–1.5 mL of diluent (5% sub-boiling distilled HNO_3 containing 0.005% Triton X-100®)
   - slurries may be prepared using larger masses and/or volumes if conditions require
4. Ultrasonic slurry mixing and sampling
   - power output to ultrasonic probe should be adjusted to provide good mixing (typically 60–80%)
   - mix time of 15–25 s
   - consider particle sedimentation and adjust sampling depth accordingly
5. GFAAS conditions
   - wavelength selection based on analyte concentration in slurry
   - less sensitive, nonresonance lines may be useful for high analyte concentrations
   - GFAAS conditions should be systematically optimized (char and atomization temperatures)
   - STPF conditions should be employed
   - oxygen ashing may be useful for removal of carbonaceous material, HF may be useful with high-silica-content matrices
   - quantitation is accomplished using aqueous standards and peak area measurements
   - the use of a matrix modifier and char step may not be necessary
6. Number of determinations
   - typically five readings of each of five slurry preparations are adequate when analyzing an unknown sample
   - data should be reviewed to determine if measured concentrations suggest a dependence on sample weight or sample heterogeneity

Without detailed knowledge of the homogeneity of the material it is difficult to know how many subsamples should be weighed out and prepared as slurries to provide a representative measure of the mean analyte concentration in the original sample.[27] In addition, the influence of analyte extraction into the aqueous phase of the slurry significantly affects the mass of material represented by the slurry measurement. For this reason, it is important to consider what information is needed from the analysis. If the objective is to gain information on the homogeneity of the material at the microgram level, then analyte extraction should be minimized (no acid in diluent) so that the analysis is representative only of the amount of solid injected into the furnace. Homogeneity characterization levels (microgram to milligram) are established by the representative sample mass.

For high-accuracy measurements of bulk analyte concentrations, an analysis of variance (ANOVA) can be performed to determine the optimum number of slurry preparations and replicate analyses.[15] Based on practical experience, it is typically found that starting with 3–5 replicate readings of 3–5 slurry preparations is sufficient to determine if additional readings are necessary and if the slurry GFAAS technique is suitable for analyzing the sample of interest. In addition, it is advisable that the mass of material used to prepare the slurry be varied to determine if the measured concentrations suggest a dependence on sample mass or indicate sample heterogeneity. If heterogeneity is evident, it might be useful to grind the sample to reduce particle sizes and/or use a larger sample mass. The mass of material used to prepare the slurry must take into account each of those factors (e.g. particle size, density, diluent volume, etc.) previously discussed.

The recommended procedure for slurry GFAAS determinations is outlined in Table 5.

4 ENVIRONMENTAL APPLICATIONS

Slurry sampling GFAAS is well suited for the analysis of a range of environmental samples. Numerous applications of the technique for the determination of trace elements in environmental matrices can be found in past reviews[2,3] or in annual *Atomic Spectroscopy Updates* featured in the *Journal of Analytical Atomic Spectrometry*.[46,47] In this section, several of the most recent applications of slurry GFAAS for environmental analyses will be described and the performance of the technique will be discussed.

4.1 Characteristic Masses, Detection Limits and Ranges

The concept of characteristic mass (m_o) in AAS provides a measure of the sensitivity of the determination for a

Table 6 Characteristic masses and DLs for slurry GFAAS analyses of several environmental materials

Analyte	Matrix	Slurry preparation	Injection volume	m_o(pg)[a]	DL (ng g^{-1})[b]	Ref.
Pb	Sediment and plant material	3–10 mg per 1 mL	10 μL	18	7–13 (0.7–1.3 pg)	39
Sn				15–30	9 (0.9 pg)	
Pb	Soil and sediment	2.5–250 mg per 50 mL	10 μL	10		33
Cd		10–400 mg per 50 mL		0.5		
Tl		100–600 mg per 10 mL		17		
Pb	Soil	45 mg per 50 mL	10–20 μL	12.5		44
Pb	Plant material	100–300 mg per 21 mL	20 μL		15 (4.3 pg)	37
Cd					5 (1.4 pg)	
Se	Soil	100 mg per 1 mL	10 μL	25	30 (30 pg)	32
Cr	Soil	1.5 mg per 1 mL	20 μL	3.6		24
Co	Soil and sediment	3–5 mg per 2.5 mL	20 μL	7.5		48
Ni				14.1		
Ti	Plant material	2500 mg per 25 mL	20 μL	74	30 (75 pg)	49
As	Soil and sediment	4 mg per 1 mL	20 μL	23	1000 (80 pg)	31
Sb		100 mg per 1 mL		14	30 (60 pg)	
Hg	Food and sediment	50–100 mg per 1.1 mL	10–20 μL		100 (180 pg)	50
Hg	Marine sediment	1000 mg per 50 mL + additional 2-fold dilution	20 μL	124	70 (14 pg)	25
Hg	Soil and sediment	125 mg per 1 mL	10 μL	61	100 (125 pg)	9

[a] Characteristic mass (mass of analyte producing an integrated absorbance signal of 0.0044 s).
[b] Calculated absolute DLs are given in parentheses.

particular analyte. It is defined as the mass of analyte that produces an integrated absorbance signal of 0.0044 s. The DL, on the other hand, is defined as the smallest amount of analyte that can be detected with some stated level of confidence. It is generally defined as the concentration (or mass) of analyte that produces a signal that is three times greater than the standard deviation of replicate blank measurements. Since the DL considers the limiting noise source associated with the measurement of a particular analyte, it is useful in comparing different analytical techniques for determining one analyte or for determining several analytes by one technique.

Recently reported characteristic mass values and DLs for the determination of analytes in a variety of environmental samples by slurry GFAAS are given in Table 6. Characteristic masses and DLs are dependent on several factors including the design of the furnace, atomization conditions, absorption wavelength, background correction method, and the use of matrix modifiers. Additionally, both the characteristic mass and the concentration DLs are dependent on the amount of material used to prepare the slurry and the amount of material injected into the furnace. They do not, however, necessarily increase linearly with the mass of sample injected into the furnace. Reporting of the absolute (or mass) DL is often useful for comparison purposes since this takes into account the amount of sample used for the analysis. Absolute DLs have been calculated and included in Table 6. Characteristic mass values and absolute DLs are in the 1–100 pg range for most elements.

As previously mentioned, a limitation of GFAAS is the limited analytical range. For a given wavelength, the analytical range is generally only two or three orders of magnitude above the DL. For this reason, it is often necessary to 'dilute' the slurried samples or modify the GFAAS analytical conditions. The feasibility of diluting the slurry depends on the homogeneity of the material, the amount extracted into the liquid, and the number of particles that will be injected into the furnace. A preferable alternative is to use less-sensitive, nonresonance absorption wavelengths to extend the analytical range. Although this method has been widely used with excellent success, not all elements exhibit nonresonance lines with acceptable analytical characteristics.[36] In these cases, the use of a minigas flow (30–300 mL min^{-1}) during the atomization step might be useful in reducing sensitivity.[25,32,36,48] Mierzwa et al. reported good agreement between the certified and measured Ni concentration (~300 μg g^{-1}) in a soil reference material using both a weaker spectral line and a minigas flow (40 mL min^{-1}) to reduce sensitivity.[48] Similarly, Lopez-Garcia et al. reported that the use of a 100 mL min^{-1} minigas flow did not significantly affect results for the determination of Pb in soil and sediment samples.[32] Although the use of a minigas flow has only a slight cooling effect and thermal balance in the tube is not

Table 7 Selected applications of slurry GFAAS in environmental analysis

Sample	Analyte	Comments	Refs.
Plant material	Cd, Pb, Ni, Co	Ultrasonic mixing; Pd modifier necessary for Cd determinations; calibration with aqueous standards (Ni and Co) or standard additions (Cd and Pb)	34
Coal and coal fly ash	Ga	Magnetic stirring; Ni as matrix modifier; aqueous calibration	51
Marine sediments	Hg	Magnetic stirring; air ashing, matrix modification (Pd), and standard additions for quantitation	25
Soil, sediment, and biological materials	Mn, Zn, Cu, Cr, Pb	Ultrasonic mixing; oxygen ashing; aqueous calibration; influence of slurry preparation on accuracy discussed	30
Sediment and soil	As, Cd, Cu, Cr, Ni, Pb	Ultrasonic mixing; aqueous calibration except for Cr (matrix-matched calibration)	52
Sediment and soil	Se	Manual shaking; samples suspended in conc. HF to eliminate silicon during furnace heating; fast-heating methodology; aqueous calibration with nickel nitrate modifier	32
Sediments, peach leaves, liver tissue, coal fly ash, and glass	Pb, Cu, Cr	International collaborative study evaluating state of the art for slurry GFAAS; discusses problems relating to inaccurate results	19, 22
Soil and sediment	Pb, Cd, Tl	Magnetic stirring; use of HF in diluent; fast-heating methodology; aqueous (Cd and Pb) and standard additions (Tl) calibration	33
Soil	Ni, Cr	Ultrasonic mixing; aqueous (Cr) and standard additions (Ni) calibration; recommend using less sensitive lines rather than miniflow to reduce sensitivity	24
Sediment and leaves	Pb, Sn	Ultrasonic mixing; comparison of several chemical modifiers; aqueous calibration (Pb) and standard additions (Sn); comparison of fast-ramp and conventional furnace heating	39
Soil and sediment	Hg	Magnetic stirring and gas bubbling; aqueous calibration; use of HF in diluent; fast-heating methodology; use of silver nitrate and potassium permanganate as modifiers	9
Soil and sediment	Co, Ni	Ultrasonic mixing; used nonresonance wavelengths and miniflow to reduce sensitivity; aqueous calibration	48
Fish	Se	Ultrasonic mixing; Pd as modifier; aqueous calibration	40
Plant material	Co, Ni, Cu	Ultrasonic mixing; aqueous calibration	53
Plant material	Cd, Pb	Ultrasonic mixing; standard additions calibration; hydrogen peroxide added to aid in removal of carbonaceous matrix and for stabilization of analytes	37

significantly affected, the reliability of this approach must be confirmed for each matrix and analyte studied.[19,22]

4.2 Analytical Performance – Benefits and Limitations

Table 7 lists several applications of slurry GFAAS for the analysis of environmental materials and includes some comments on mixing methods, matrix modifiers, and calibration procedures used. Most workers have developed methodology that allowed for aqueous calibration. However, in some cases, standard additions were necessary to achieve accurate results.[25,26,33,34,37,39] The need for standard additions calibration was likely due to incomplete removal of the sample matrix or the presence of interfering atomic or molecular species. Accuracy of slurry GFAAS measurements is also dependent on the homogeneity of the sample, particle sizes, amount of sample used in the slurry preparation, and the extent of analyte extraction. For this reason, it is advisable to make multiple preparations using different masses of sample to determine the minimum mass required to achieve consistent measured concentrations. As previously discussed, reducing particle sizes by grinding might also be beneficial for improving the homogeneity of the sample. Precision is similarly affected by particle size, homogeneity of the sample, mass of slurry preparation, and extent of analyte extraction. When analyzing homogeneous samples under optimum analytical conditions, the slurry sampling GFAAS technique can be expected to provide accuracies of ±5% and measurement relative standard deviation (RSD) values of 5% or less.

In addition to the ability to obtain highly accurate and precise measurements with slurry sampling GFAAS, the technique offers several additional benefits. These include the often mentioned benefits associated with solids analysis techniques, such as reduced contamination, decreased sample preparation times, and reduced likelihood of analyte loss during sample preparation or because of retention by an insoluble fraction. Other benefits of the GFAAS technique are the simplicity and relatively low cost of the instrumentation as well as the ability to characterize solid materials on the μg to mg level. These attributes make slurry sampling GFAAS a powerful technique for the determination of trace analytes in environmental samples. However, the technique does have some limitations.

One of the major limitations of GFAAS is that it is primarily a single-element technique; however, multielement systems based on continuum sources (simultaneous) or multiple HCLs (sequential) have been developed. Sequential systems are limited by relatively slow duty cycles, making them less than ideal for use with transient GFAAS signals. Although not commercially available at present, continuum sources combined with fast multichannel detectors might prove useful in the future for multielement slurry GFAAS analyses. Using a prototype simultaneous multielement atomic absorption spectrometer with a continuous source (SIMAAC) with photomultiplier tubes as detectors, it has been reported that up to seven elements could be simultaneously determined with an average accuracy of $100 \pm 12\%$ and an average precision of $<10\%$ RSD.[7] Obviously, compromise conditions must be used when measuring several elements at once. Char temperatures should be low enough to avoid loss of the most volatile analyte, whereas atomization temperatures should be high enough to ensure that the most refractory element is atomized.

Conventional slurry GFAAS analyses require careful, systematic optimization of instrumental and sampling parameters for each analyte and matrix of interest. Obviously, this can be a time-consuming process and careful consideration should be given to what type of information is desired from the analytical measurement. The small sample masses injected into the furnace that are beneficial for homogeneity characterization might lead to unacceptably poor precision when bulk concentration of the analyte is the desired objective. Sample heterogeneity can be particularly problematic when dealing with complex environmental samples. The limited analytical range associated with AAS measurements can also prove problematic when analyzing samples with very different analyte concentrations. Although alternate wavelengths or minigas flows can be used to reduce sensitivity, these require additional optimization.

4.3 Comparison of Sample Types

As evidenced in Table 7, slurry sampling GFAAS has been used for the determination of a variety of elements in numerous environmental samples. In this section, some of the issues related to specific sample matrices will be discussed. Particle size plays an important role in slurry GFAAS analyses. Several workers have reported the need to work with relatively small particles ($<50\,\mu m$) for accurate and precise analyses; however, it has been demonstrated in the authors' laboratory that good results could be obtained for samples with particle sizes up to $500\,\mu m$ in diameter.[23] The composition and homogeneity of the matrix, as well as the percentage of analyte extracted, largely dictates what range of particle sizes can be used for the analysis. For samples that are easily ashed, such as biological materials, larger particle sizes can be tolerated. For samples that are more refractory in nature, such as geological materials, it might be necessary to use small particles for the analysis to ensure complete vaporization of the particle and/or complete release of the analyte. Even for a highly volatile element such as Hg, it has been reported that particle sizes of less than $20\,\mu m$ were necessary to ensure the complete release of Hg from marine sediment samples.[26] Another key point in determining what particle sizes can be used for analysis is the stability of the slurried sample. In section 3.4, the effect of particle sedimentation on the accuracy of slurry GFAAS measurements was discussed. For higher-density materials (see Table 2), it is imperative that the sampling depth be great enough to ensure that a representative aliquot of the sample is removed for injection into the furnace.[35] Because a finite time can exist between the end of the mixing step and pipetting of the sample (e.g. with automated ultrasonic mixing systems), and autosampler cups are of limited depth, it might be necessary to reduce particle sizes for high-density materials in order to minimize the rate of sedimentation and allow for representative slurry sampling.

Because of the relatively large amount of solid material being vaporized in the furnace, matrix effects can significantly affect the quality of slurry GFAAS measurements. The presence of organic material, not completely removed during the charring step, affects the volatility of some elements, increases background levels, and leads to the build-up of carbonaceous material in the furnace. These problems are most significant for biological matrices, such as plants and foods. However, appreciable levels of organic material can be found in many types of soil as well. To alleviate these problems, matrix modifiers can be used to stabilize volatile elements during furnace heating and good background correction methods (e.g. Zeeman or Smith–Hieftje) should be employed. An effective method

for removing carbonaceous material and thus avoiding its deleterious effects has been the use of an oxygen or air ashing step in the furnace program.[23,41,42] Hydrogen peroxide added to the slurry diluent has also been reported to be useful for this purpose.[37,43]

Samples high in silicate have been shown to seriously affect analyte signals in slurry GFAAS. In work with glass samples, Bendicho and de Loos-Vollebregt found that after 50–100 firings of the furnace absorbance signals decreased and precision became worse.[12] In addition, rapid deterioration of the graphite tube has been reported when analyzing high-silicate matrices due to the formation of silicon carbide at high temperatures.[33] In order to avoid problems related to graphite tube degradation and matrix suppression effects, silicon can be removed by addition of HF to the slurry diluent.[9,31–33] Silicon is volatilized as SiF_4 during the furnace heating cycle. In addition, HF increases analyte extraction and leads to a decrease in particle sizes because of partial dissolution of the matrix.[33] Fluoride compounds, such as NaF[31] and NH_4F,[54] have also been studied for the removal of silicon during furnace heating, but did not significantly decrease the amount of silicon prior to atomization. Because of the highly corrosive and toxic nature of HF, extreme caution should be exercised when using this acid. In addition, ultrasonic mixing should not be used to mix slurries containing HF because of damage caused to the titanium probe. For this reason, other methods for slurry agitation must be employed or alternatively, HF could be pipetted separately into the furnace (just as a matrix modifier might be added). A problem reported with the use of HF is the violent reaction of this acid with fine silica particles, which resulted in spattering and loss of analyte.[54] To avoid this problem, a low temperature (60 °C) reaction step lasting 3 min was used prior to drying and charring of the sample. Several other elements (B, As, Sb, and Ge) form volatile fluorides and may be lost along with silicon. However, their volatilities are highly dependent on oxidation states, which are controlled by a combination of acids used and sample composition.[55] Workers have reported difficulties in the determination of As in soils and sediments using HF due to analyte loss during the heating cycle.[31] The problem was alleviated by neutralizing the slurry with sodium hydrogen carbonate prior to injection of the sample into the furnace.

5 COMPARISON WITH OTHER DIRECT ANALYSIS METHODS

Numerous spectroscopic methods have been utilized for the elemental analysis of environmental samples; however, no single technique is capable of addressing all analytical needs. For this reason, it is essential that the analyst understand the benefits and limitations of specific techniques in order to choose a suitable method for a particular analytical measurement. In this section, several spectroscopic methods will be discussed and the performance of these methods will be compared with slurry GFAAS for the analysis of environmental samples.

The benefits associated with direct solids analysis have already been discussed. Briefly, these include reduced analysis times because samples can be analyzed with little to no sample preparation, reduced risk of contamination from reagents, reduced likelihood of analyte loss during dissolution procedures or because of retention by an insoluble portion of the sample, potential for improved sensitivity since the sample is not diluted, and the selective analysis of microamounts of solid. A wide variety of methods can be used to analyze solid environmental samples, including X-ray fluorescence (XRF) and total-reflection X-ray fluorescence (TXRF), electrothermal vaporization inductively coupled plasma atomic emission spectrometry (ETV/ICP/AES), and electrothermal vaporization inductively coupled plasma mass spectrometry (ETV/ICP/MS) laser ablation inductively coupled plasma atomic emission spectrometry (LA/ICP/AES), and laser ablation inductively coupled plasma mass spectrometry (LA/ICP/MS) laser induced breakdown spectrometry (LIBS), direct current arc atomic emission spectrometry (DC Arc/AES), glow discharge mass spectrometry (GDMS), both laser ablation/laser excited atomic fluorescence (LA/LEAFS) and electrothermal atomization/laser excited atomic fluorescence (ETA/LEAFS), electron microprobe analysis (EMP) and X-ray microprobe analysis (XMP), and neutron activation analysis (NAA) and instrumental neutron activation analysis (INAA) techniques. In this section, some of the more commonly used techniques, or techniques with considerable potential for direct analysis of solid environmental samples, will be discussed. In particular, these are XRF and TXRF, ETV/ICP/AES, ETV/ICP/MS, LA/ICP/AES, LA/ICP/MS, and LIBS. Typical analytical characteristics of these techniques are summarized in Table 8.

XRF has been widely used in the field of environmental analysis. It is a nondestructive, multielement technique capable of simultaneous detection of elements with atomic numbers $11 < Z < 92$. It is particularly useful for contamination monitoring, and portable devices capable of in situ measurements have been developed for this purpose.[56] The XRF technique, however, suffers from relatively poor DLs, making it unsuitable for trace analysis. In addition, matrix effects are often problematic unless very thin samples or residues are being analyzed. Significant improvements in terms of both absolute DLs and the occurrence of matrix effects can be achieved

Table 8 Comparison of techniques for direct solids analysis

	DL ($\mu g\,g^{-1}$)[a]	Matrix effects	Selectivity	Precision	Accuracy
GFAAS	0.01–1	Small	Good	0.5–5%	Good
XRF	0.1–100	Moderate	Good	2–5%	Good
TXRF	0.1–100	Small	Good	2–5%	Good
ETV/ICP/AES	0.01–10	Moderate	Moderate	0.5–5%	Moderate
ETV/ICP/MS	0.001–1	Moderate	Good[c]	1–5%	Good
LIBS	0.1–1	Moderate[b]	Moderate	2–10%	Moderate
LA/ICP/MS	0.001–1	Moderate[b]	Good[c]	2–10%	Moderate

[a] DL in the solid.
[b] Requires assumption that laser produces representative sampling; laser techniques, in general, require special consideration for quantitation and are more prone to matrix effects.
[c] Isotopic selectivity.

using TXRF. This technique is suitable for very small sample amounts (<100 μg) with thicknesses ranging from 10 nm–10 μm depending on the matrix.[57] Samples must be placed on polished carriers to ensure the absence of matrix effects. Calibration is typically performed by spiking the sample with an internal standard. DLs are in the picogram range ($\mu g\,g^{-1}$ in solid) and the technique is capable of high in-depth resolution (nanometers), making it well suited for depth profiling. Because of the limitations with respect to sample size and thickness, however, TXRF is suitable only for selected environmental applications.[58,59] When combined with dissolution and subsequent deposition of the sample on a polished carrier, the full advantages of the technique can be exploited.

ETV/ICP/AES and ETV/ICP/MS are becoming increasingly popular for the direct analysis of solid samples. Issues related to preparation of the sample and representative sampling of the solid are identical to those previously discussed for GFAAS. In ETV techniques, the graphite furnace is used to vaporize the sample before it is transported into the ICP source. Separation of the vaporization step and analyte excitation (for AES) or ionization (for MS) can lead to improved analytical performance since each process can be independently optimized. Because they are multielement techniques, with high sensitivity and wide linear dynamic ranges (5–7 orders of magnitude), ETV/ICP/AES and ETV/ICP/MS are powerful methods for the analysis of environmental samples. DLs for ETV/ICP/AES are generally poorer than those obtained with GFAAS; however, ETV/ICP/MS DLs are, in general, comparable or better (sub picogram levels, $ng\,g^{-1}$ in solid) than GFAAS. ICP/AES offers simultaneous multielement analysis, whereas ICP/MS relies on scanning of the quadrupole mass analyzer and provides sequential multielement analysis. This means that with ICP/MS, only about five elements can typically be measured during an analysis because of the transient nature of the ETV signal. In addition to providing DLs that are one hundred or more times better than ICP/AES, ICP/MS spectra are much simpler and spectroscopic interferences are less problematic than with ICP/AES. This is particularly advantageous when analyzing complex environmental matrices. Another important benefit of mass spectrometric detection is that it allows the determination of isotope ratios.[60,61]

The suitability of ETV/ICP/MS for very sensitive and accurate analysis of a variety of environmental samples is evidenced by a large number of applications of the technique.[62–69] Most workers have reported the need to use the method of additions to obtain satisfactory results with this technique, most likely because of differences in analyte transport between aqueous standards and solid samples or because of matrix effects in the ICP/MS. It has been demonstrated that the use of a physical carrier (e.g. Pd) or removal of part of the solid matrix (e.g. organics by oxygen ashing) can be useful in minimizing differences in analyte transport efficiency, thereby allowing the use of aqueous standards for quantitation.[62,63] Signal suppression due to space charge effects can also be problematic, especially when large amounts of solid are introduced into the ICP/MS. Removal of matrix components prior to vaporization can be used to reduce space charge effects. A very powerful quantification method available with ICP/MS techniques is isotope dilution. Since another isotope of the same element represents the ideal internal standard for that element, isotope dilution should provide accurate results even in the presence of matrix effects, instrument instability, and fluctuations (or loss) of analyte in vaporization and transport processes. Successful application of isotope dilution to ETV/ICP/MS measurements has been reported for the determination of Cd and Se in biological and environmental samples[68] and Cu, Cd, and Pb in sediments.[69]

Laser sampling techniques, such as LIBS and LA/ICP/MS are becoming increasingly popular for direct solids analysis, since little to no sample preparation is required. Samples can be analyzed in either compact or

powdered form. The modest spatial resolution (~10 μm lateral and ~1 μm in-depth) achieved with laser sampling can also provide the analyst with information on the distribution of elements in a sample. One of the limitations of laser sampling is that laser–material interactions are complex, which results in an unknown quantity of ablated material with a composition that may or may not be representative of the bulk material.[70] For this reason, internal standardization and matrix-matched standards are typically required to produce accurate results. The use of ultraviolet lasers, rather than infrared, has been shown to be beneficial because of improved spatial resolution, minimization of selective vaporization, and enhanced ablation efficiencies.[70,71] LIBS provides rapid multielement analysis with DLs for most analytes in the $\mu g\,g^{-1}$ range. In addition to being a useful laboratory technique, compact LIBS systems have been developed for on-site analysis.[72,73] LA/ICP/MS offers improved DLs compared to LIBS (100 times or more) and greater freedom from matrix interferences because the sampling and excitation (ionization) steps are separated and can be independently optimized. In addition, spectral interferences are less problematic with mass spectrometric detection, as previously discussed for ICP/AES and ICP/MS. Although calibration remains an impediment to the widespread applicability of LA/ICP/MS, alternatives to matrix-matched standardization, such as the use of fused-glass beads[74] or dual-sample introduction systems,[75–77] have been described for calibration purposes and the future of the technique appears to be very promising.

No single technique is capable of addressing the wide range of analytical needs. In choosing the most appropriate method for an analysis, careful consideration should be given to a number of issues, including the type of information required from the analysis (e.g. bulk or microanalysis), detection capabilities, number of elements to be determined, cost, and sample throughput. As a general rule, the simplest method that meets analysis requirements should be used. In many cases, considerable benefits can be gained by using one or more complementary techniques for analysis.

6 FUTURE EXPECTATIONS

The relative maturity of slurry sampling GFAAS for routine analysis of environmental materials has been demonstrated in an international collaborative study.[19,22] Since slurry sampling is ideally suited to the analysis of samples that are already in powdered form, it will continue to be widely used for the analysis of soils and sediments. Many other solid materials are well suited for slurry analysis, especially botanicals and food samples, since they can be easily ground with minimal risk of contamination. The use of ultrasonic mixing and autosampler pipetting has made sample introduction routine and reliable; in fact, the major source of imprecision in slurry GFAAS measurements is sample heterogeneity. For this reason, the technique is quite useful for homogeneity characterization of samples at the microgram to milligram level.

Relative DLs obtainable with slurry GFAAS are excellent, with only ICP/MS techniques providing better values. Although multielement ETV/ICP/MS and LA/ICP/MS appear to be quite promising, these techniques are not at the stage where they can compete with GFAAS for routine analyses, especially when multielement data are not required. Matrix effects are more problematic with ICP/MS and the instrumentation is more complex and expensive. It is likely that each of the techniques will continue to prosper and that choice of one technique over the other will be decided based on the required sensitivity, number of elements to be determined, and cost. The development of commercial, continuum source AAS instruments may greatly improve the applicability of slurry GFAAS for multielement, environmental analyses.

DISCLAIMER

Mention of trademark or proprietary products does not constitute a guarantee or warranty of the product by the US Department of Agriculture and does not imply their approval to the exclusion of other products that may also be suitable.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
ANOVA	Analysis of Variance
DC Arc/AES	Direct Current Arc Atomic Emission Spectrometry
DL	Detection Limit
EDL	Electrodeless Discharge Lamp
EMP	Electron Microprobe Analysis
ETA/LEAFS	Electrothermal Atomization/Laser Excited Atomic Fluorescence
ETV	Electrothermal Vaporization
ETV/ICP/AES	Electrothermal Vaporization Inductively Coupled Plasma Atomic Emission Spectrometry
ETV/ICP/MS	Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry

GDMS	Glow Discharge Mass Spectrometry
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HCL	Hollow Cathode Lamp
ICP	Inductively Coupled Plasma
INAA	Instrumental Neutron Activation Analysis
LA/ICP/AES	Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry
LA/ICP/MS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
LA/LEAFS	Laser Ablation/Laser Excited Atomic Fluorescence
LIBS	Laser Induced Breakdown Spectrometry
MS	Mass Spectrometry
NAA	Neutron Activation Analysis
NIST	National Institute of Standards and Technology
RSD	Relative Standard Deviation
SIMAAC	Simultaneous Multielement Atomic Absorption Spectrometer with a Continuous Source
STPF	Stabilized Temperature Platform Furnace
TXRF	Total-reflection X-ray Fluorescence
XMP	X-ray Microprobe Analysis
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Graphite Furnace Atomic Absorption Spectrometry

REFERENCES

1. R.E. Sturgeon, 'Future of Atomic Spectrometry for Environmental Analysis', *J. Anal. At. Spectrom.*, **13**, 351–361 (1998).
2. C. Bendicho, M.T.C. de Loos-Vollebregt, 'Solid Sampling in Electrothermal Atomic Absorption Spectrometry Using Commercial Atomizers', *J. Anal. At. Spectrom.*, **6**, 353–374 (1991).
3. N.J. Miller-Ihli, 'Solids Analysis by GFAAS', *Anal. Chem.*, **64**, 964A–968A (1992).
4. D.V. Brady, J.G. Montalvo, J. Jung, R.A. Curran, 'Direct Determination of Lead in Plant Leaves Via Graphite Furnace Atomic Absorption', *At. Absorpt. Newsl.*, **13**, 118–119 (1974).
5. D.V. Brady, J.G. Montalvo, G. Glowacki, A. Pisciotta, 'Direct Determination of Zinc in Sea-Bottom Sediments by Carbon Tube Atomic Absorption Spectrometry', *Anal. Chim. Acta.*, **70**, 448–452 (1974).
6. L. Ebdon, H.G.M. Parry, 'Direct Atomic Spectrometric Analysis by Slurry Atomisation. Part 2. Elimination of Interferences in the Determination of Arsenic in Whole Coal by Electrothermal Atomisation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **2**, 131–134 (1987).
7. N.J. Miller-Ihli, 'Slurry Sample Preparation for Simultaneous Multi-element Graphite Furnace Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **3**, 73–81 (1988).
8. R. Karwowska, K.W. Jackson, 'Atomisation Characteristics of Lead Determined in Alumina Matrices by Slurry Electrothermal Atomisation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **2**, 125–129 (1987).
9. I. Lopez-Garcia, M. Sanchez-Merlos, M. Hernandez-Cordoba, 'Determination of Mercury in Soils and Sediments by Graphite Furnace Atomic Absorption Spectrometry with Slurry Sampling', *Spectrochim. Acta B*, **52**, 2085–2092 (1997).
10. M. S. Epstein, G. R. Carnrick, W. Slavin, N.J. Miller-Ihli, 'Automated Slurry Sample Introduction for Analysis of a River Sediment by Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chem.*, **61**, 1414–1419 (1989).
11. E.D. Byrd, D.J. Butcher, 'Determination of Trace Elements in Biological Standard Reference Materials by Graphite Furnace Atomic Absorption Spectrometry with Solid and Slurry Sampling', *Spectrosc. Lett.*, **26**, 1613–1624 (1993).
12. C. Bendicho, M.T.C. de Loos-Vollebregt, 'The Influence of Pyrolysis and Matrix Modifiers for Analysis of Glass Materials by GFAAS using Slurry Sample Introduction', *Spectrochim. Acta B*, **45**, 679–693 (1990).
13. I. Lopez-Garcia, M. Sanchez-Merlos, M. Hernandez-Cordoba, 'Slurry Sampling Device for Use in Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **12**, 777–779 (1997).
14. N.J. Miller-Ihli, 'Automated Ultrasonic Mixing Accessory for Slurry Sampling into a Graphite Furnace Atomic Absorption Spectrometer', *J. Anal. At. Spectrom.*, **4**, 295–297 (1989).
15. N.J. Miller-Ihli, 'Advances in Ultrasonic Slurry Graphite Furnace Atomic Absorption Spectrometry', *Fresenius' J. Anal. Chem.*, **345**, 482–489 (1993).
16. S. Lynch, D.J. Littlejohn, 'Palladium as a Chemical Modifier for the Determination of Lead in Food Slurries by Electrothermal Atomisation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **4**, 157–161 (1989).

17. M.W. Hinds, K.W. Jackson, 'Determination of Lead in Soil by Vortex Mixing Slurry–Graphite Furnace Atomic Absorption Spectrometry', *At. Spectrosc.*, **12**, 109–110 (1991).
18. G.R. Carnrick, G. Daley, A. Fotinopoulos, 'Design and Use of a New Automated Ultrasonic Slurry Sampler for Graphite Furnace Atomic Absorption', *At. Spectrosc.*, **10**, 170–174 (1989).
19. N.J. Miller-Ihli, 'Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry: A Preliminary Examination of Results from an International Collaborative Study', *Spectrochim. Acta B*, **50**, 477–488 (1995).
20. W. Slavin, D.C. Manning, G.R. Carnrick, 'The Stabilized Temperature Platform Furnace', *At. Spectrosc.*, **2**, 137–145 (1981).
21. W. Slavin, *Graphite Furnace AAS: A Source Book*, The Perkin–Elmer Corporation, Norwalk, CT, 1984.
22. N.J. Miller-Ihli, 'Slurry Sampling Electrothermal Atomic Absorption Spectrometry: Results From the Second Phase of an International Collaborative Study', *J. Anal. At. Spectrom.*, **12**, 205–212 (1997).
23. N.J. Miller-Ihli, 'Influence of Slurry Preparation on the Accuracy of Ultrasonic Slurry Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **9**, 1129–1134 (1994).
24. *Spex CertiPrep Handbook of Sample Preparation and Handling*, Spex CertiPrep, Inc., Metuchen, NJ, 1997.
25. R. Dobrowolski, 'Determination of Ni and Cr in Soils by Slurry Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta B*, **51**, 221–227 (1996).
26. P. Bermejo-Barrera, J. Moreda-Pineiro, A. Moreda-Pineiro, A. Bermejo-Barrera, 'Palladium as a Chemical Modifier for the Determination of Mercury in Marine Sediment Slurries by Electrothermal Atomization Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **296**, 181–193 (1994).
27. N.J. Miller-Ihli, 'Slurry Sampling for Graphite Furnace Atomic Absorption Spectrometry', *Fresenius' J. Anal. Chem.*, **337**, 271–274 (1990).
28. V. Majidi, J.A. Holcombe, 'Error Analysis for Sampling of Slurries: Sedimentation Errors', *Spectrochim. Acta B*, **45**, 753–761 (1990).
29. J.A. Holcombe, V. Majidi, 'Error Analysis for Sampling of Slurries: Volumetric Errors', *J. Anal. At. Spectrom.*, **4**, 423–425 (1989).
30. N.J. Miller-Ihli, 'A Systematic Approach to Ultrasonic Slurry GFAAS', *At. Spectrosc.*, **13**, 1–6 (1992).
31. I. Lopez-Garcia, M. Sanchez-Merlos, M. Hernandez-Cordoba, 'Arsenic and Antimony Determination in Soils and Sediments by Graphite Furnace Atomic Absorption Spectrometry with Slurry Sampling', *Spectrochim. Acta B*, **52**, 437–443 (1997).
32. I. Lopez-Garcia, M. Sanchez-Merlos, M. Hernandez-Cordoba, 'Rapid Determination of Selenium in Soils and Sediments Using Slurry Sampling Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1003–1006 (1996).
33. I. Lopez-Garcia, M. Sanchez-Merlos, M. Hernandez-Cordoba, 'Slurry Sampling for the Determination of Lead, Cadmium, and Thallium in Soils and Sediments by Electrothermal Atomic Absorption Spectrometry with Fast-Heating Programs', *Anal. Chim. Acta*, **328**, 19–25 (1996).
34. R. Dobrowolski, J. Mierzwa, 'Determination of Trace Elements in Plant Materials by Slurry Sampling Graphite Furnace AAS – Some Analytical Problems', *Fresenius' J. Anal. Chem.*, **346**, 1058–1061 (1993).
35. N.J. Miller-Ihli, 'Ultrasonic Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry: Analytical Considerations for Analysis of a High Density Glass Material', *Spectrochim. Acta B*, **52**, 431–436 (1997).
36. M.A. Belarra, M. Resano, J.R. Castillo, 'Expanding the Working Concentration Range in Graphite Furnace Atomic Absorption Spectrometry. Effect of Non-Stop Flow on the Determination of Tin', *Spectrochim. Acta B*, **52**, 1223–1227 (1997).
37. I. Karadjova, M. Karadjov, 'ETAAS Determination of Cd and Pb in Plants', *Fresenius' J. Anal. Chem.*, **360**, 246–251 (1998).
38. F. Dolinsek, J. Stupar, V. Vrscaj, 'Direct Determination of Cadmium and Lead in Geological and Plant Materials by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**, 653–660 (1991).
39. V.I. Slaveykova, M. Hoenig, 'Electrothermal Atomic Absorption Spectrometric Determination of Lead and Tin in Slurries. Optimization Study', *Analyst*, **122**, 337–343 (1997).
40. G.S.B. Januzzi, F.J.Krug, M.A.Z. Arruda, 'Application of the Slurry Technique to the Determination of Selenium in Fish Samples by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **12**, 375–378 (1997).
41. L. Ebdon, A.S. Fisher, H.G.M. Parry, A.A. Brown, 'Direct Atomic Spectrometric Analysis by Slurry Atomisation. Part 10. Use of an Air-Ashing Stage in Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **5**, 321–324 (1990).
42. S.C. Stephen, D. Littlejohn, J.M. Ottaway, 'Evaluation of a Slurry Technique for the Determination of Lead in Spinach by Electrothermal Atomic Absorption Spectrometry', *Analyst*, **110**, 1147–1151 (1985).
43. P. Viñas, N. Campillo, I. Lopez-Garcia, M. Hernandez-Cordoba, 'Slurry Atomization of Vegetables for the Electrothermal Atomic Absorption Spectrometric Analysis of Lead and Cadmium', *Food Chem.*, **50**, 317–321 (1994).
44. M.W. Hinds, K.E. Latimer, K.W. Jackson, 'Determination of Lead in Soil by Slurry–Electrothermal Atomic Absorption Spectrometry With a Fast Temperature Programme', *J. Anal. At. Spectrom.*, **6**, 473–476 (1991).
45. D. Bradshaw, W. Slavin, 'Rapid Slurry Analysis of Solid Coal and Coal Fly Ash Samples', *Spectrochim. Acta B*, **44**, 1245–1256 (1989).

46. J.R. Dean, L.M. Garden, J. Armstrong, M.S. Cresser, M. Cave, P. Watkins, 'Atomic Spectroscopy Update – Environmental Analysis', *J. Anal. At. Spectrom.*, **12**, 19R–87R (1997).
47. J.R. Dean, O. Butler, A. Fisher, L.M. Garden, M.S. Cresser, P. Watkins, M. Cave, 'Atomic Spectroscopy Update – Environmental Analysis', *J. Anal. At. Spectrom.*, **13**, 1R–56R (1998).
48. J. Mierzwa, Y-C. Chang, M.-H. Yang, 'Determination of Co and Ni in Soils and River Sediments by Electrothermal Atomic Absorption Spectrometry with Slurry Sampling', *Anal. Chim. Acta*, **355**, 277–282 (1997).
49. I. Lopez-Garcia, P. Viñas, M. Hernandez-Cordoba, 'Slurry Procedure for the Determination of Titanium in Plant Materials Using Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **7**, 529–532 (1992).
50. I. Karadjova, P. Mandjukov, S. Tsakovsky, V. Simeonov, J.A. Stratis, G.A. Zachariadis, 'Determination of Mercury by Electrothermal Atomic Absorption Spectrometry Using Different Chemical Modifiers or a Slurry Technique', *J. Anal. At. Spectrom.*, **10**, 1065–1068 (1995).
51. S.X-quan, W. Wen, W. Bei, 'Determination of Gallium in Coal and Coal Fly Ash by Electrothermal Atomic Absorption Spectrometry Using Slurry Sampling and Nickel Chemical Modification', *J. Anal. At. Spectrom.*, **7**, 761–764 (1992).
52. W. Klemm, G. Bombach, 'Trace Element Determination in Contaminated Sediments and Soils by Ultrasonic Slurry Sampling and Zeeman Graphite Furnace Atomic Absorption Spectrometry', *Fresenius' J. Anal. Chem.*, **353**, 12–15 (1995).
53. D.T. Takuwa, G. Sawula, R. Wibetoe, W. Lund, 'Determination of Cobalt, Nickel, and Copper in Flowers, Leaves, Stem and Roots of Plants Using Ultrasonic Slurry Sampling Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **12**, 849–854 (1997).
54. R.St.C. McIntyre, D.C. Gregoire, C.L. Chakrabarti, 'The Direct Determination of Radium in Geological Materials by Slurry Sampling Electrothermal Vaporization-Inductively Coupled Plasma/Mass Spectrometry', *Spectroscopy*, **13**, 18–24 (1998).
55. K.E. Jarvis, A.L. Gray, R.S. Houk, *Handbook of Inductively Coupled Plasma Mass Spectrometry*, Chapman and Hall, New York, 1992.
56. A. Argyraki, M.H. Ramsey, P.J. Potts, 'Evaluation of Portable X-ray Fluorescence Instrumentation for In-Situ Measurements of Lead on Contaminated Land', *Analyst*, **122**, 743–749 (1997).
57. G. Tolg, R. Klockenkamper, 'The Role of Total-Reflection X-ray Fluorescence in Atomic Spectroscopy', *Spectrochim. Acta B*, **48**, 111–127 (1993).
58. J. Injuk, R. Van Grieken, R. Klockenkamper, A. von Bohlen, P. Kump, 'Performance and Characteristics of Two Total-Reflection X-ray Fluorescence and a Particle Induced X-ray Emission Setup for Aerosol Analysis', *Spectrochim. Acta B*, **52**, 977–984 (1997).
59. T. Stahlschmidt, M. Schulz, W. Dannecker, 'Application of Total-Reflection X-ray Fluorescence for the Determination of Lead, Calcium, and Zinc in Size-Fractionated Marine Aerosols', *Spectrochim. Acta B*, **52**, 995–1001 (1997).
60. D.C. Gregoire, B.M. Acheson, R.P. Taylor, 'Measurement of Lithium Isotope Ratios by Inductively Coupled Plasma Mass Spectrometry: Application to Geological Materials', *J. Anal. At. Spectrom.*, **11**, 765–772 (1996).
61. L.A. Allen, J.J. Leach, H.-M. Pang, R.S. Houk, 'Precise Measurement of Ion Ratios in Solid Samples Using Laser Ablation With a Twin Quadrupole Inductively Coupled Plasma Mass Spectrometer', *J. Anal. At. Spectrom.*, **12**, 171–176 (1997).
62. R.W. Fonseca, N.J. Miller-Ihli, 'Analyte Transport Studies of Aqueous Solutions and Slurry Samples Using Electrothermal Vaporization ICP/MS', *Appl. Spectrosc.*, **49**, 1403–1410 (1995).
63. R.W. Fonseca, N.J. Miller-Ihli, 'Influence of Sample Matrix Components on the Selection of Calibration Strategies in Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta B*, **51**, 1591–1599 (1996).
64. M.-J. Liaw, S.-J. Jiang, Y.-C. Li, 'Determination of Mercury in Fish Samples by Slurry Sampling Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta B*, **52**, 779–785 (1997).
65. D.C. Gregoire, N.J. Miller-Ihli, R.E. Sturgeon, 'Direct Analysis of Solids by Ultrasonic Slurry Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **9**, 605–609 (1994).
66. L. Moens, P. Verrept, S. Boonen, F. Vanhaecke, R. Dams, 'Solid Sampling Electrothermal Vaporization Inductively Coupled Plasma Atomic Emission Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta B*, **50**, 463–475 (1995).
67. Y-C. Li, S-J. Jiang, 'Determination of Cu, Zn, Cd, and Pb in Fish Samples by Slurry Sampling Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **359**, 205–212 (1998).
68. F. Vanhaecke, S. Boonen, L. Moens, R. Dams, 'Isotope Dilution as a Calibration Method for Solid Sampling Sampling Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 125–130 (1997).
69. M.-J. Liaw, S.-J. Jiang, 'Determination of Copper, Cadmium, and Lead in Sediment Samples by Slurry Sampling Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 555–560 (1996).
70. R.E. Russo, 'Laser Ablation', *Appl. Spectrosc.*, **49**, 14A–28A (1995).
71. C. Geertsen, A. Briand, F. Chartier, J.-L. Lacour, P. Mauchien, S. Sjostrom, J.-M. Mermet, 'Comparison Between

Infrared and Ultraviolet Laser Ablation at Atmospheric Pressure – Implications for Solid Sampling Inductively Coupled Plasma Spectrometry', *J. Anal. At. Spectrom.*, **9**, 17–22 (1994).

72. B.C. Castle, A.K. Knight, K. Visser, B.W. Smith, J.D. Winefordner, 'Battery Powered Laser-induced Plasma Spectrometer for Elemental Determinations', *J. Anal. At. Spectrom.*, **13**, 589–595 (1998).
73. K.Y. Yamamoto, D.A. Cremers, M.J. Ferris, L.E. Foster, 'Detection of Metals in the Environment Using a Portable Laser-induced Breakdown Spectroscopy Instrument', *Appl. Spectrosc.*, **50**, 222–233 (1996).
74. F.E. Lichte, 'Determination of Elemental Content of Rocks by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**, 2479–2485 (1995).
75. E.F. Cromwell, P. Arrowsmith, 'Semiquantitative Analysis with Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**, 131–138 (1995).
76. S.M. Chenery, J.M. Cook, 'Determination of Rare Earth Elements in Single Mineral Grains by Laser Ablation Microprobe/Inductively Coupled Plasma Mass Spectrometry – Preliminary Study', *J. Anal. At. Spectrom.*, **8**, 299–303 (1993).
77. D. Günther, H. Cousin, B. Magyar, I. Leopold, 'Calibration Studies on Dried Aerosols for Laser Ablation/Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 165–170 (1997).

Soil Instrumental Methods

M.R. Carter
Agriculture & Agri-Food Canada, Charlottetown, Canada

D. Curtin
New Zealand Institute of Crop & Food Research, Christchurch, New Zealand

E.G. Gregorich
Agriculture & Agri-Food Canada, Ottawa, Canada

Instrumental methods play an important role in the analysis of soil material. Soil is composed of mineral, organic, water, and air components. Analytical methods are required that characterize the fitness or quality of soils to perform various functions, such as providing a medium for plant growth, recycling waste products, and regulating and storing water, energy, and nutrients. The objective of this article is to describe the basic concept and approach to laboratory analysis of soil, and to outline the application of techniques used in the analysis of soil quality properties. Emphasis is placed on soil sampling and sample preparation, followed by a description of the possible instrumental methods available for the analysis of the chemical, biochemical, and physical properties of soil.

1 Introduction **1**
1.1 Soil Chemical Properties 2
1.2 Soil Biochemical Properties 2
1.3 Soil Physical Properties 2
2 Soil Quality **2**
3 The Soil Sample **3**
4 Instrumental Methods for Soil Chemical Analysis **4**
4.1 Mineralogical Analysis 4
4.2 Total Elemental Analysis 5
4.3 Organic Matter 6
4.4 pH and pE 6
4.5 Cation Exchange Capacity 7
4.6 Anion Adsorption 8
4.7 Analysis of Soil Solution and Aqueous Extracts of Soil 9
4.8 Analysis of Soil Air 10
4.9 Analysis of Soil Nutrients 10
5 Instrumental Methods for Soil Biochemical Analysis **12**
5.1 Isotopic Methods for Organic Matter 12
5.2 Analysis of Organic Matter Fractions 14
5.3 Microbial Biomass 15
5.4 Analyzing Microbial Processes 16
6 Instrumental Methods for Soil Physical Analysis **17**
6.1 Soil Water Analysis 17
6.2 Characterizing Porosity and Aeration 18
6.3 Macrostructure and Density 19
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

1 INTRODUCTION

Soils serve as an interface between aquatic, atmospheric, and terrestrial ecosystems. In this capacity, the soil provides important functions such as supporting plant and animal life, providing a sink and source of biogenic energy and nutrients, partitioning the flows of energy and water, regulating the solubility of toxic elements, providing a sink and source for greenhouse gases (e.g. CO_2, N_2O, CH_4) and mitigating their accumulation in the atmosphere, and acting as an environmental buffer. Soil forms at a relatively slow rate, so should be considered a nonrenewable natural resource.

Soil is composed of interacting mineral, organic, water and air components and has strong biological features. Knowledge of the chemical, physical, and biological properties of soil can only be fully obtained by recourse to analytical methods. The need for soil analytical methods is underlined, by both the scientific community and the general public, in the ongoing concern with soil and environmental quality protection, and the sustainability of the biosphere. Analysis of soil material in the laboratory can provide information on soil properties,[1] but it should be realized that some of these properties may not be always equivalent to the same property measured in situ, as the removal of the material will also remove the influence of adjacent soil and the ambient environment.

Soil is used for various purposes, such as a medium for plant growth, for recycling numerous waste products, and as a construction material. Thus, methods are required that can characterize soil material for a specific use or purpose. The wide range of disciplines that have an interest in soil analysis (e.g. agriculture, forestry, earth sciences, civil engineering, and environmental science) underlines the need for the continued development of analytical and testing methods. This article describes the basic concept

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and approach to laboratory analysis of soil material, and provides a select description of instrumental analysis of the chemical, physical, and biological properties of soil, with an emphasis on instrumental methods of analysis for soil quality attributes.

1.1 Soil Chemical Properties

Soil can be viewed as a three-phase system of solid, liquid, and gaseous components in an ever-shifting equilibrium. Chemical analysis can involve characterization of all three phases. The solid phase consists of various soil minerals and organo-mineral complexes. Generally, most analysis is concerned with specific constituents of the solid and liquid phases that have importance for a certain soil function (e.g. soil as a medium for plant growth). The interface between the solid (generally clay sized, $<2\,\mu m$, secondary phyllosilicate minerals) and liquid phases is usually a net-negatively-charged particle surface surrounded by a diffuse layer of hydrated counterions merging into the soil solution or liquid phase. In many cases, the chemical composition of both the diffuse layer and the soil solution is of interest. Analysis of the soil air or atmosphere is often conducted to determine soil aeration and the presence of physiologically active gases such as ethylene, toxic gases, and radiatively active gases. For plant growth studies, the availability of soil nutrients and chemical elements that impact on plant productivity are a major concern for ''soil test'' determinations.

1.2 Soil Biochemical Properties

Soil is a dynamic medium for many organisms that transform and cycle organic materials (i.e. mainly plant residues) and chemical elements. These chemical processes result in the formation of organic compounds that are important for the functioning of the soil. Many of these biochemical compounds and their related properties can be measured and characterized.

Although soil is a three-phase system of solid, liquid, and gaseous components, it also contains living organisms or a biotic component. Soil biochemical properties are related to the activities of these organisms and the biochemical substances they produce. The importance of microbes to soil ecology is reflected in the numerous analytical techniques available in the area of soil microbiology.[1–3]

1.3 Soil Physical Properties

Soil physics deals mainly with the properties of the soil solid phase, especially the framework it supplies for the storage of water and the movement of water and air. Physical properties of soil have many aspects that range from semipermanent (e.g. clay content, retention of soil water) to dynamic (e.g. porosity, structure). The latter tend to change under the influence of climate and management, and tend to be soil and site specific. Generally, the methodology for soil water content and porosity is based on clear or accepted physical principles. In contrast, soil structure measurements are very method-dependent and subjective. However, though this complex situation often confounds the development of quantitative methodology, it has produced a relatively small number of reference methods that utilize a common approach to soil structure assessment.[4,5]

2 SOIL QUALITY

Interest in evaluating the quality of soil has evolved around the various functions that soils perform in ecosystems.[6] Placing a value upon soil with regard to a specific function, purpose, or use leads to the concept of *soil quality*. However, in contrast to water and air, where the function can be directly related to human and animal consumption, the function placed upon soil is often diverse and usually not directly linked or involved with human health. Table 1 provides a list of common soil functions.

Soil quality can only be assessed by measuring properties. However, analysis of soil properties alone cannot provide a measure of soil quality unless the properties evaluated are calibrated or related against a designated role or function of the soil. Thus, implicit in any definition of soil quality is an understanding of the stated function of the soil, or what the soil does. Soil quality can be simply defined as ''the fitness of a soil for a specific use''. However, to accommodate the concepts of ''sustainability'' and ''ecosystem'', and to stress the need to protect the environment, definitions have been expanded and developed, as indicated by the following quotation:

> Soil quality is the capacity of a soil to function, within ecosystem and land-use boundaries, to sustain biological productivity, maintain environmental quality, and promote plant, animal and human health.[6]

Soil quality has two parts: an intrinsic soil part which covers a soil's *inherent* or natural capacity for a specific function (e.g. plant growth), and a *dynamic* part influenced by the soil user or how the soil is managed. This has important implications for the approach to soil analysis. Inherent soil quality properties, such as mineralogy and particle size distribution, are mainly viewed as almost static and usually show little change over time. In contrast, dynamic soil quality properties, such as organic matter, pH, and microbial parameters, are responsive to

Table 1 Function components of a soil with regard to crop production

Function component	Function characteristics/processes
Medium of plant growth	Suitable medium for seed germination and root growth Absence of adverse chemical conditions (acidity, salinity, sodicity) Supply balance of nutrients Suitable medium for microbes (nutrient cycling, decomposition) Promote root growth and development
Regulate water	Receive, store, and release moisture for plant use Adequate water retention to buffer and reduce effects of drought Adequate infiltration and storage capacity to reduce runoff
Regulate gases	Accept, hold, and release gases Permeable for adequate air movement and exchange with atmosphere
Regulate energy	Store and release (recycle) energy-rich organic matter
Buffer or filter	Accept, hold, and release nutrients Sequester energy compounds and/or biotoxic elements Detoxify substances harmful to plants

Adapted from Gregorich and Carter.[6]

soil management and manipulation, and are thus subject to change over relatively short time periods.

A useful framework to evaluate soil quality is based on the following sequence: functions, processes, attributes or properties, attribute indicators, and methodology.[6] Soil quality is evaluated on the basis of the function in question. Functions deal with ''what the soil does'', or ''what the soil is asked to do''. Each function can be characterized by specific soil processes that support the function which is being imposed upon the soil. Soil quality attributes can be defined as measurable soil properties that influence the capacity of the soil to perform a specific function. Generally, attributes describe a critical soil property involved with the process. In many cases the specific attribute or property may be difficult to measure directly, so an indicator can be used to serve as an indirect, practical measure of the attribute. Soil analysis is concerned with characterizing soil properties and/or indicators. It is generally acknowledged that indicators have the following characteristics: sensitivity to variations in soil management (but not over-sensitivity), relatively low sampling error, easily measurable with readily available methodology.

3 THE SOIL SAMPLE

Laboratory analysis of soil is dependent on obtaining a suitable soil sample. However, soil properties can exhibit both temporal and spatial variability, and also vary with depth. Thus, information about soil properties usually involves a sequence of steps and procedures as follows:

- description of soil sampling site;
- selection of optimum sampling time, procedure, and number of soil samples required;
- storage and preparation of soil samples;
- selection of suitable analytical method;
- calibration of the method and quality control;
- statistical analysis;
- interpretation of analytical results.

In many cases some information about site characteristics is needed as soils can vary in relation to their position within both a soil profile and the landscape.[1] Even where soils appear to be uniform, the various sources of soil variation demand that an adequate sampling scheme be adopted so that valid inferences can be made about the soil population from the sample analyzed. In most cases little is known about the soil location or site, so random or probability sampling schemes need to be utilized. To accommodate different situations, various soil sampling schemes and methods have been developed for both disturbed and undisturbed soils.[1,7] Some common sampling procedures are:

- *simple random sampling* – sample points are not influenced by other points;
- *systematic or grid sampling* – sample points are at regular distances from each other;
- *composite sampling* – sample points are grouped and the samples within each group are combined to give well mixed, but fewer, composite samples;
- *spatial structure sampling* – sample points are at regular distances to estimate relationships between distance and soil variation.[6]

The above sampling procedures allow some measure of soil variation to be calculated for the measured variables (x) in the number (n) of soil samples collected. The most basic statistics are the mean ($\bar{x}$), the variance (s^2),

and the standard deviation (s). They are calculated as shown in Equations (1) and (2).

$$\bar{x} = \sum \frac{x}{n} \quad (1)$$

$$s^2 = \sum \frac{(x - \bar{x})^2}{n - 1} \quad (2)$$

Dividing s by $\sqrt{n}$ gives the standard error of the mean. To assess the relative degree of variability among the measured variables, the coefficient of variation (CV) can be determined. The CV can vary among soil properties from a relative low value (<10%) for soil bulk density and water-filled pore space, to a medium value (10% < CV < 50%) for soil particle size distribution, and a high value (>50%) for soil water and air permeability. The calculation for CV is shown in Equation (3).

$$\text{CV} = \frac{s}{\bar{x}} \quad (100) \quad (3)$$

In addition to adequate soil sampling procedures, attention must also be given to soil handling and preparation prior to analysis. These quality control considerations are a prerequisite to minimize error in all phases of soil analysis, especially those errors caused by both soil population and measurement variability. Soil handling concerns involve such things as the protocol for drying field-moist samples and the degree of grinding and screening that soil samples require prior to analysis.[1,7,8] In addition, storage of soil samples before analysis requires attention to the state of the soil (e.g. dry or moist), the temperature, and the type of container used.[1,2] Special precautions and conditions apply for processing soil samples for biological analyses.[2] Quality control is also an important aspect of the actual laboratory analysis procedure. This includes the use of quality control procedures such as calibration of measuring instruments, use of both certified (obtained from international organizations) and internal reference samples where possible (i.e. for total concentrations of elements), and statistical quality control (i.e. routine evaluation of $\bar{x}$ and s of the analytical values over time).

4 INSTRUMENTAL METHODS FOR SOIL CHEMICAL ANALYSIS

Soil chemical quality is important for many soil functions (Table 1). Some specific processes involved are:

- storage of organic matter
- acceptance, retention, and release of nutrients
- regulating the solubility of elements.

The primary function related to soil chemical quality for agronomic purposes is to provide nutrients for crop growth. Chemical processes and reactions also underpin other functions, such as the ability of soil to act as a filter for potential pollutants and thus reduce the risk of contamination of rivers, lakes, and aquifers. As a result of the use of pesticides and the application of sewage sludges and other wastes to land, an ability to detoxify pesticides and immobilize heavy metals is important. Chemical properties that affect these functions are mineralogy, organic matter content, cation exchange capacity (CEC), anion adsorption capacity, pH, pE, soluble salt concentration (salinity), exchangeable sodium percentage (ESP), and nutrient-supplying power.

Storage of nutrients can be viewed in terms of the quantity–intensity concept where *quantity* represents the total amount of nutrient held in the various supply reservoirs (in organic matter or adsorbed to clay particles) and *intensity* is the concentration in the soil solution at any one time. The amount and type of clay minerals in a soil determine its ability to adsorb and retain nutrient cations, and also to buffer or regulate the release of plant nutrients. Organic matter contains a store of nutrients and energy-rich compounds that, upon decomposition by microbes, provide a source of plant-available N, S, and P. In addition, the functional groups of the humic component of soil organic matter can retain cations and remove pesticides from the soil solution. The CEC of the soil reflects the contributions of both clay minerals and organic matter. The pH and pE of the soil can have a controlling influence on the solubility of nutrients, mineral dissolution, and ion adsorption. Changes in pH have major impacts on the soil as a medium for plant growth. Acid soils may contain toxic levels of soluble Al and Mn, and suffer from deficiencies of certain nutrients (e.g. Ca and P). In alkaline soils deficiencies of trace elements, especially Fe and Zn, are common due to their low solubility in the upper pH range. An excess of soluble salts, as found in saline soils, can limit the growth of plants owing to decreased osmotic potential (making it more difficult for plant roots to absorb water) and nutrient imbalances or toxicities. In some cases, trace elements can accumulate to phytotoxic concentrations, an example being the excessive boron levels sometimes found in arid areas where leaching is limited.[9]

4.1 Mineralogical Analysis

Soil mineral composition is a function of the mineralogy of the parent material from which the soil is formed, and the soil weathering processes. The coarser fractions, i.e. gravel (>2 mm), sand (50 μm–2 mm), and silt (2–50 μm), consist mainly of primary minerals such as quartz, feldspars, and micas inherited from the parent material

(e.g. igneous and metamorphic rocks). The clay fraction (<2 μm) is largely composed of secondary minerals, formed by chemical reactions during soil formation or inherited from sedimentary rocks. These minerals are mainly aluminosilicates belonging to the phyllosilicate (or layer silicate) group of minerals. Important members are smectites, vermiculites, and kaolinites.[10] Clay, which is the most surface-active mineral fraction in the soil, has the ability to adsorb and hold nutrients, water, and other chemicals including pesticides. The dominant influence of clay on soil chemical behavior means that most mineralogical investigations focus primarily on that size fraction.

The aggregated nature of soil components represents a situation where crystalline minerals of varying size are associated with noncrystalline material, as well as other components such as organic matter and salts. Thus, before mineralogical analysis, the soil (sieved, <2 mm) must be pretreated to enhance separation or dispersion of aggregates. Organic matter and coatings of iron oxide and carbonate that bind particles together are normally removed using selective chemical treatments prior to mineralogical analysis. Soluble salts may also need to be removed as they could cause clay particles to coagulate or flocculate, thereby preventing dispersion. Chemical pretreatments are designed to have minimal effects on constituents other than those to be eliminated. Hydrogen peroxide is generally used to remove organic matter while the most widely used method of removing iron oxide coatings involves chemical reduction of iron to the soluble ferrous form.[7] Carbonates may be destroyed by treatment with dilute acid solution. Once the sample has been ''cleaned up'' by these pretreatments, it is dispersed in water using a suitable mixing device such as a shaker or ultrasonic vibrator.

Separation into particle-size fractions is achieved by a combination of sieving (to collect the >50 μm fraction, i.e. sand) and sedimentation (to isolate clay from silt). Sedimentation procedures are based on the relationship that exists between settling velocity and particle diameter, as defined by Stokes' law.[7] From this law it may be calculated that 2-μm clay will settle to a depth of 10 cm in 8 h at 20 °C. Clay can be removed by decanting the suspension that remains above the 10 cm depth after an 8 h settling period. The residue is resuspended and decanted repeatedly to recover those particles that were not previously at the top of the suspension at the beginning of the sedimentation period. In mineralogical investigations, it is common to subdivide clay into coarse (0.2–2 μm) and fine (<0.2 μm) fractions because of the tendency for certain minerals to occur within specific size limits. Subdivision of clay is achieved by sedimentation of particles under centrifugal acceleration.[7]

X-ray diffraction (XRD) analysis is considered to be the most powerful technique available for the identification of mineral species in soil clays. Criteria for X-ray identification of the common clay minerals have been described in many texts.[7,10] Other instrumental techniques that are used in specific situations include thermal analysis (TA) (differential TA and thermogravimetric analysis), infrared (IR) spectroscopy, and magnetic methods.[1,7] Quantitative estimation of the proportions of the various minerals in soil clays usually requires the use of a combination of analytical techniques including instrumental methods and determination of the elemental composition of the material. Sand-sized mineral components can be separated by specific gravity or magnetic susceptibility and identified on the basis of their crystallographic, optical, physical, and chemical properties.

Some soil clays contain a significant amount of noncrystalline or amorphous aluminosilicate material. As such materials are transparent to X-rays, alternative techniques are needed for their determination. Noncrystalline material may be measured by boiling clay in 0.5 M NaOH for 2.5 min and determining dissolved Al and Si by atomic absorption spectroscopy (AAS).[7] A clay/NaOH ratio (1 : 100) is used to avoid saturating the NaOH solution with Al or Si.

Weathering of soil may result in the accumulation of oxides of Fe and Al which frequently occur as surface coatings on other soil constituents. Pedogenic oxides (also called ''free'' oxides) are important because they exhibit a strong affinity for certain anions, particularly phosphate. Free iron is determined by reducing to the soluble ferrous (Fe^{2+}) form using sodium dithionite ($Na_2S_2O_4$) as reducing agent and measuring soluble Fe by AAS.[7] This procedure also extracts Al associated with iron oxides.

4.2 Total Elemental Analysis

Methods for analysis of total elemental composition of soils and soil components were developed from procedures used for analysis of rocks and minerals.[7] Elemental analysis of soils requires their decomposition into soluble forms by means of acid digestion or fusion with various fluxes. The two basic procedures for bringing elements into solution are as follows.

- fusion in Na_2CO_3 followed by dissolution of the fused sample in hydrochloric acid (HCl);
- digestion in hydrofluoric acid (HF) alone or in HF in combination with sulfuric (H_2SO_4) and perchloric ($HClO_4$) acid. The decomposition of silicate minerals by HF occurs by reaction of HF with Si to form gaseous SiF_4 which is lost by volatilization if digestion is carried out in an open crucible. One method of

preventing volatile loss of SiF_4 gas is by use of a specially designed Teflon® decomposition vessel fitted with a sealed cover.[3]

The acid extracts and digests can be analyzed for a range of elements (e.g. Si, Al, Ca, Mg, K, Na, Fe, Mn, Ti, and other elements) using AAS and flame emission spectroscopy (FES), as appropriate.

The recommended method for determination of total P requires conversion of mineral and organic forms of P to soluble orthophosphate by fusion in Na_2CO_3, followed by dissolution of the melt in H_2SO_4 solution. Phosphate in solution is then measured colorimetrically by the molybdate blue method.[5]

Soil sulfur can exist in both inorganic (including several oxidation states) and organic forms, with the latter predominating in surface soil horizons. Methodology for determination of total S generally involves converting all of the S to a single form using a variety of wet digestion (WD) and dry combustion (DC) techniques. Alkaline oxidation in sodium hypobromite (NaOBr) converts all S to sulfate, which is subsequently reduced to sulfide (H_2S) using a mixture of hydriodic acid, hypophosphorus acid and formic acid followed by colorimetric determination of H_2S as methylene blue.[3] Alternatively, the sulfate-S can be quantified by ion chromatography (IC).[1]

4.3 Organic Matter

Analysis of soil organic matter is common in agronomic and environmental studies. Organic matter (which contains 50–58% C) is usually estimated from the concentration of soil organic C and ranges from about 0.5% to 10% in temperate region mineral soils.[3] It is an important store of macro-elements that are crucial in plant nutrition (i.e. N, P, and S). N, P, and S are usually present in relatively low concentrations in soil organic matter (<5% for N, <2% for P and S). In soil science, the term ''organic'' is applied to elements that are associated with organic matter. For example, organic C refers to C in soil organic matter while inorganic C is found mainly in carbonate minerals such as calcite ($CaCO_3$) and dolomite ($CaCO_3 \cdot MgCO_3$). Methods used to determine the total amounts of organic elements usually involve WD or DC techniques to liberate the element or convert it to an inorganic form.[1,3] Generally, the following basic approaches can be used to determine organic C:

- WD using acid dichromate followed by back titration of the remaining dichromate using ferrous ammonium sulfate (Walkley–Black method, see Carter[1]);
- WD as above but with collection and determination of the evolved CO_2;
- DC in a furnace and determination of evolved CO_2 using an IR analyzer.

Modifications of these methods can include a separate treatment to measure any C present in carbonates (e.g. by neutralization of carbonate with acid and back-titration of the excess acid) which is subtracted from total C to give organic C. Alternatively, carbonates may be destroyed by treatment with dilute acid before C determination. Instrumental methods featuring automated versions of DC methods are now routinely used to determine total C and organic C in many laboratories.

Total soil N is also measured by WD (e.g. Kjeldahl method, see Carter[1]) and DC (e.g. Dumas method, see Carter[1]). In the Kjeldahl procedure, soil samples are digested with sulfuric acid in the presence of a catalyst and a salt to control the boiling temperature of the acid. Organic N is converted to NH_4 by the boiling H_2SO_4/salt solution. After addition of a strong alkali, NH_3 is steam-distilled into boric acid solution and titrated with standardized acid. Alternatively NH_4–N can be quantified by colorimetry using an autoanalyzer.[3] The DC method utilizes high temperatures and a hot copper catalyst to convert organic N to N_2, which may be quantified using a thermal conductivity (TC) detector.[3]

Part of the total soil N (up to 10% in topsoils and 30% in subsoils) resides as nonexchangeable ammonium (fixed N that cannot be replaced or extracted with a neutral K salt solution – see section 4.5), which is found primarily in the interlayer positions of certain (i.e. 2 : 1 type) clay minerals. This fixed N can be quantified by using KOBr to remove organic N followed by a HF/HCl mixture (5 : 1) to decompose the clay minerals containing the fixed N.[3] The released NH_4 is then measured using distillation or an autoanalyzer, as described previously.

Organic P is generally determined indirectly as the difference between total P and inorganic P. Total P may be determined by Na_2CO_3 fusion as described previously while inorganic P can be measured by sequentially extracting soil with HCl and NaOH.[3] Another popular method is to ignite the soil sample at 550 °C over a period of 1 to 2 h to convert organic P to inorganic P. The difference in amounts of inorganic P between an ignited sample and a separate un-ignited sample represents the amount of organic P in the soil.

As in the case of organic P, organic S is estimated as the difference between total S (determined as outlined previously) and inorganic S. Inorganic S in normally aerated soils may be assumed to be present entirely in the sulfate form, which can be extracted using a dilute phosphate solution (e.g. distilled water containing 16 mmol P L^{-1} as KH_2PO_4 or $Ca(H_2PO_4)$) and measured either using IC or colorimetrically following its reduction to sulfide.[3]

4.4 pH and pE

Soil pH exerts an important influence on growth of plants and on the activities of microbes and other soil

organisms such as earthworms. Processes of mineral dissolution generally exhibit a strong dependence on pH. In addition, the number of functional groups active in the retention of cation and anions is pH-dependent. Acid soils can be deficient in basic nutrients (e.g. Ca, P) while metals such as Al and Mn may be present in sufficiently high concentrations to be toxic to plants. Alkaline soils are often associated with deficiencies of trace elements, particularly Fe and Zn. Thus, pH provides an important indication of the soil chemical environment, with values in the 5.5 to 7.5 range being best for the growth of most plants.

Soil pH is measured using a glass, H^+-sensing electrode paired with a reference electrode attached to a suitable voltmeter (pH meter). Soil pH is determined by inserting the electrodes into a slurry or suspension obtained by mixing a sample with distilled water or a dilute electrolyte solution (e.g. 0.01 M $CaCl_2$). Factors which can influence the pH value are the soil-to-solution ratio (1 : 1 or 1 : 2 soil-to-solution ratios are commonly used) and the salt or electrolyte content (pH can be a half unit lower when measured in 0.01 M $CaCl_2$ compared with water).

Reduction–oxidation or ''redox'' reactions influence the solubility and speciation of elements that can exist in multiple oxidation states, especially Fe, Mn, N, and S. The redox status of a soil reflects the availability of electrons that may participate in reduction–oxidation reactions. The pE is defined as the negative logarithm of the free electron activity.[11] Soil pE is controlled by physical conditions (water content and porosity) and biological activity. Values are usually in the −6 to 13 range. The most common method of determining electron activity in soils is to measure the potential difference between a platinum indicator electrode and a reference electrode (e.g. calomel or Ag/AgCl electrode) with a voltmeter (pH meter).[12]

4.5 Cation Exchange Capacity

Fine mineral particles, especially clay, and organic matter carry a negative electrostatic charge on their surfaces. Such charge sites attract and hold cations which are referred to as *exchangeable cations* because they are easily displaced from the particle surface by other cations. In the case of clay particles, negative charge may originate through a process called isomorphous substitution whereby a cation of higher valence is replaced in the crystal structure by one of lower valence (e.g. Al^{3+} substitutes for Si^{4+}). This type of charge is independent of pH and is often referred to as *permanent negative charge*. Negative-charge sites also arise from hydrolysis reactions at broken bonds at the edges of clay particles and through dissociation of acidic functional groups (e.g. carboxylic groups) in organic matter. The number of such sites increases as pH increases. The total amount of cations balancing a soil's negative charge represents its CEC. Table 2 gives some typical CEC values of soil colloids.

Table 2 Typical CEC values of soil colloids[a]

Colloid	CEC (mmol (+) kg^{-1})
Illite	100–400
Vermiculite	1000–1500
Smectite	700–1200
Chlorite	100–300
Kaolinite	20–100
Allophane	300–1500[b]
Fe and Al oxide	<10
Humus (organic matter)	1000–3000[b]

[a] Adapted from McLaren and Cameron.[13]
[b] Value varies greatly with pH.

In some soils, part of the permanent charge may be neutralized by positively charged polymers of hydroxy-Al or Fe oxide. Soil acidification can cause deprotonation of these Al and Fe polymers and also of hydroxy-Al and -Fe associated with organic functional groups, resulting in a positive charge and potential for anion exchange as pH declines. In the vast majority of soils, anion exchange capacity (AEC) is small compared with CEC and AEC is not a routinely measured soil characteristic.

Soil CEC is not an independent or single-value property. The value depends on pH, as discussed previously, and, in weathered soils high in Fe and Al oxides, it can also be influenced by the electrolyte concentration of the soil solution. Methods to measure CEC may involve displacement of exchangeable cations using a salt solution (e.g. 1 M ammonium acetate) followed by measurement and summation of the displaced cations, Ca^{2+}, Mg^{2+}, K^+, Na^+ (summation method). An alternative method involves replacement of native exchangeable cations by treatment with a solution containing an index cation (e.g. Na^+, NH_4^+, or Ba^{2+}), which is then displaced and measured (displacement method). Depending on the purpose of the measurement, one may choose to determine CEC at a standard pH (commonly pH 7 or 8) or at the pH of the soil.

The following sequence is used for determination of exchangeable cations and CEC by the ammonium acetate (NH_4OAc) method at pH 7.

- A soil sample (5 g; <2 mm) is added to a centrifuge tube followed by 40 mL of 1 M NH_4OAc (adjusted to pH 7). The tube and contents are shaken, centrifuged, and the supernatant solution filtered into a 250-mL flask. A fresh aliquot of the NH_4OAc is added to the tube, followed by shaking, centrifugation, and filtration of the supernatant into the 250-mL flask.

This cycle may be repeated to ensure complete saturation of exchange sites with NH_4 (3 or 4 washes with NH_4OAc should be sufficient). In nonsaline and noncalcareous soils, the NH_4OAc extract is retained for determination of Ca, Mg, K, and Na by AAS and FES.

- The residual NH_4OAc is removed from the soil by washing several times with alcohol (ethanol or isopropanol), then the soil is washed with four 40-mL aliquots of 1 M KCl into a 250-mL flask. After bringing the KCl extract up to volume with distilled water, it is analyzed for NH_4 using an autoanalyzer. Exchangeable cations and CEC are usually expressed in units of mmol (+) kg^{-1} (millimoles of charge per kilogram of soil).

The nature of the exchangeable cation population reflects the chemical and mineralogical makeup of the soil. The exchangeable cations Ca^{2+}, Mg^{2+}, K^+, Na^+ are called *exchangeable bases* and they normally occur in the order listed. Together Ca^{2+} and Mg^{2+} account for over 90% of exchangeable bases in most soils. Exchangeable K varies with parent material and management, particularly K fertilizer use history. Only trace amounts of Na are found in soils in humid regions, but in salt-affected arid-zone soils, Na commonly occupies 10% or more of exchange sites. The presence of this latter level of exchangeable Na can cause clay particles to disperse, leading to dis-aggregation and structural deterioration. The ESP, i.e. sodium as a percentage of the total CEC, is used to diagnose the potential for structural problems in arid and semiarid regions.

Acid conditions lead to the release of Al from aluminosilicate minerals and Al^{3+} can become an important exchangeable cation when soil pH is below about 5.5. Exchangeable Al^{3+} and H^+ represent the *acidic* component of the CEC. In strongly acid soils, a separate measurement to determine exchangeable Al^{3+} and H^+ is often carried out. This procedure involves use of an unbuffered salt solution (usually 1 M KCl) to extract acidic cations that are present at the pH of the salt–soil mixture (approximately the pH of the soil). The total amount of Al^{3+} plus H^+ in the KCl extract can be determined by titration with NaOH, and Al can be measured by AAS, allowing H^+ to be estimated by difference (generally the amount of H^+ is small compared with Al^{3+}). In acid agricultural soils, exchangeable acidity is often used to assess lime requirements.

4.6 Anion Adsorption

Two mechanisms of anion adsorption in soils are generally recognized:

(1) simple electrostatic attachment to positively charged sites (nitrate and chloride are adsorbed through this mechanism);

(2) ligand exchange whereby certain anions (e.g. phosphate) displace $-OH$ and $-OH_2$ ligands of Fe and Al oxides and hydroxides.

These differences in adsorption mechanisms translate to major contrasts in behavior among the principal anions found in soils. Except for those that are highly weathered and acidic, soils have negligible amounts of positive charge and do not adsorb nitrate, which is easily leached as a consequence. In contrast most soils, but especially those high in Fe and Al oxide, can adsorb substantial quantities of phosphate and its concentration in the soil solution is typically maintained at less than about 5 μmol P L^{-1}. Sulfate is adsorbed electrostatically but it may also participate in ligand exchange to a limited extent. In general, soils with pH > 6 adsorb little sulfate. A simple index of sulfate adsorption can be obtained by shaking a 5-g soil sample with 25 mL of 0.01 M $CaCl_2$ containing 1.6 mmol SO_4–S L^{-1}.[14] The amount adsorbed is calculated from the decrease in solution sulfate concentration upon contact with the soil.

Characterization of a soil's ability to adsorb phosphate is often done by shaking samples of the soil with solutions containing a range of P concentrations and calculating the amount adsorbed as a function of the soluble P concentration. A plot of the amount adsorbed against the concentration in solution is called an isotherm (Figure 1). The isotherm may be used to estimate the soil's adsorption capacity and its affinity for phosphate. The latter measure gives an indication of the ease with which adsorbed P is released into the soil solution. Because phosphate is adsorbed mainly by oxidic components, the content of

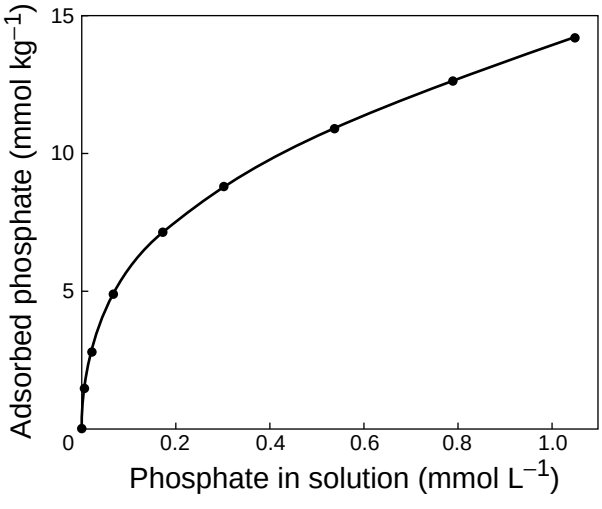

Figure 1 An example of a soil phosphate adsorption isotherm.

"free" Fe and Al can be used as an indirect gauge of the adsorption capacity.

4.7 Analysis of Soil Solution and Aqueous Extracts of Soil

The chemistry of the soil solution is of interest for disciplines such as plant nutrition and environmental science. The chemical composition of the solution phase of soils can vary depending on both climate and soil type (Table 3). Many chemical and biological processes are controlled by the composition of the soil solution. Methods used to isolate the soil solution include column displacement (e.g. using liquids such as saturated $CaSO_4$ solution or ethanol) or immiscible displacement in which a dense organic liquid such as carbon tetrachloride (specific gravity 1.6) displaces the soil solution under centrifugal force.[18,19] The analytical sequence for column displacement is as follows.

- Place portions of about 30 g, moist, sieved (<6 mm) soil in a plastic or glass column (5-cm diameter; 50-cm long, with a screen or mesh attached to the bottom of the column to retain the soil) and gently pack until the soil is within a few centimeters of the column top. Place about 1 cm of acid-washed sand on top of the soil. Add the displacement liquid (e.g. saturated $CaSO_4$ solution containing 4 g of KSCN per liter as a tracer) to the top of the column and maintain a head (~1 cm) of solution above the sand during the displacement process. The first 2–3 mL of leachate from the column can be discarded. Subsequently, leachate may be collected in small (~5 mL) aliquots and tested for the presence of SCN^- by adding a drop of leachate to a solution of $FeCl_3$. The procedure is continued until SCN^- is detected (generally several hours are required to complete the procedure).
- The displaced solution is filtered to remove particulate material and analyzed for the solutes of interest using AAS, FES, or IC, as appropriate. From the analytical data, activities of different chemical species in solution may be calculated using a suitable computer program.

Table 3 Chemical composition of solution phase of soils representing temperate, tropical (weathered soil), and semiarid, saline environments[a]

	Temperate	Tropical	Semiarid (saline)
pH	6.3	4.2	9.0
EC ($dS\,m^{-1}$)	0.5	0.2	12.0
Ca ($mmol\,L^{-1}$)	0.82	0.01	0.13
Mg ($mmol\,L^{-1}$)	0.37	0.06	4.5
K ($mmol\,L^{-1}$)	1.03	0.10	6.4
Na ($mmol\,L^{-1}$)	1.72	0.27	160.9
NH_4 ($mmol\,L^{-1}$)	0.39	0.67	nd[b]
Cl ($mmol\,L^{-1}$)	0.31	1.04	25.1
SO_4 ($mmol\,L^{-1}$)	0.44	0.10	60.4
NO_3 ($mmol\,L^{-1}$)	2.79	0.05	nd[b]
HCO_3 ($mmol\,L^{-1}$)	0.12	0.10	41.7

[a] Adapted from Edmeades et al.,[15] Gillman and Bell,[16] and Kohut and Dudas.[17] Values for phosphorus in soil solution are very low and generally not reported.
[b] nd = Not determined.
EC, electrical conductivity.

Relatively few column displacements can be performed by a single operator in an 8 h day. The centrifuge method is more suited to routine use but, since small soil samples (10–20 g) are used, the volume collected may be insufficient for complete characterization of soil solution. Inductively coupled plasma atomic emission spectrometry (ICPAES) can be an especially useful analytical technique where solution volume is limited. ICPAES features a very high temperature (7000–8000 K) excitation source that greatly reduces molecular interferences compared with the AAS technique. In addition to its relative freedom from chemical interferences, ICPAES has a wide analytical range and low detection limits. The sensitivity of the technique allows for the complete analysis of a small quantity of solution.[20,21]

Indirect estimates of soil solution composition can be obtained using soil–water extracts.[1,3] Such procedures include extracts at a fixed soil/water ratio (e.g. 1 : 2) and saturated paste extracts, where the amount of water added depends on the moisture-holding capacity of the soil. The saturated paste method is a popular way of measuring soluble salts (e.g. in saline soils). The procedure is as follows.

- Place 100–200 g of air-dry soil into a plastic container and add distilled water with stirring until it is saturated. Allow the mixture to stand for several hours to permit the soil to absorb the water. Check to see that the soil is saturated (the soil paste should glisten as it reflects light and flow slightly when the container is tipped) and adjust the moisture level if necessary.
- Transfer the saturated soil paste to a Buchner funnel fitted with a retentive filter paper, apply vacuum, and collect the filtrate. The EC of the saturated paste extract provides a measure of total dissolved salts. The EC value can be used in conjunction with functions describing the relationship between crop yield and EC to evaluate the salinity hazard posed by the soil. An EC of 4 decisiemens per meter is commonly used to separate saline from nonsaline soils. From individual cation concentrations (in $mmol\,L^{-1}$) in the

paste extract, the sodium adsorption ratio (SAR), an indicator of the amount of exchangeable Na in the soil, may be calculated as shown in Equation (4).

$$\mathrm{SAR} = \frac{\mathrm{Na}}{(\mathrm{Ca} + \mathrm{Mg})^{0.5}} \quad (4)$$

Soil solution samples may be obtained in the field by vacuum extraction. The most common vacuum sampler is a porous ceramic cup. The cup is inserted to the desired soil depth, taking care to achieve good soil–sampler contact. Sample collection is accomplished by applying a suction of 50–85 kPa to the sampler using a motor-driven or hand vacuum pump. To ensure fresh samples representative of the solution in the soil pores adjacent to the ceramic cup, the initial sample is usually discarded and the second used for analysis. The time required to collect a solution sample varies with the suction applied and the hydraulic conductivity and water content of the soil. For soils near field capacity with good conductivity, enough sample for most analytical needs can often be collected in about an hour. One of the major applications of in situ solution samplers is in studies of nitrate leaching.

Other in situ or remote instruments are capable of measuring the soil solution composition, e.g. total salt concentration. In situ instruments are used extensively in agricultural research where continuous monitoring of soil solution is often needed. These instruments include porous matrix sensors, four-electrode soil EC sensors, electromagnetic-induction soil EC sensors and time-domain reflectometry (TDR).[7] The first method measures the EC of soil water directly, whereas the last three methods measure the EC of bulk soil which depends on the soil water content and salt concentration.

4.8 Analysis of Soil Air

The composition of soil air reflects the various internal chemical and biological processes, and it is regulated by the soil moisture level and the rate of exchange between soil and the atmosphere. In addition to the major gases in the atmosphere (N_2, O_2, and CO_2), soil air may contain an array of other gases such as methane (CH_4), ethylene (C_2H_4), nitrous oxide (N_2O), nitric oxide (NO), ammonia (NH_3), and various sulfur gases. Generally, the major impediment to satisfactory analysis of soil air is the sampling process. In the field, various types of soil wells can be used to allow samples of the soil air to be withdrawn by syringe[7] or chambers can be placed over the soil to trap emitted gases.[22] In the case of disturbed samples, incubation flasks or flow-through columns can be utilized in the laboratory. Gas chromatography (GC)[1,3,23] utilizing an ultrasonic detector is the most versatile method for the separation, identification, and determination of individual gases, though the more sensitive electron capture detector may be needed to measure N_2O, a low concentration gas. Carbon dioxide and O_2 are usually measured by GC using a TC detector.[1] Determination of gases dissolved in the soil solution is not possible using GC, but it can be accomplished using electrochemical (e.g. platinum microelectrodes to measure diffusion of O_2) and amperometric (e.g. O_2 electrode) sensors.

In the field, air wells can be used that allow the soil atmosphere to be withdrawn by vacuum after a time of equilibrium,[24] or chambers can be inserted into the soil to measure the efflux of CO_2 from the soil. In static closed-chamber systems, alkali solutions (e.g. NaOH, KOH) are used to passively absorb the emitted CO_2, typically over a 24-h period. In dynamic closed-chamber systems the air circulates from the chamber headspace to an IR gas analyzer and returns to the chamber (Figure 2). The chamber is coupled to a portable CO_2 analyzer in a closed circuit and a small fan is used to mix the air within the closed chamber during the measurement. The portable CO_2 analyzer usually contains software to calculate CO_2 flux rates and has the capacity to store the data in memory which allows repeated measurements to be made. Measurements using these systems require placement of the chamber on the soil surface for a short time, typically 1–2 min; thus there is little impact of the chamber on the soil environment.

Nitrous oxide (N_2O) emissions can be measured by sequential soil air samples drawn from soil wells, or by taking a series of air samples with evacuated vacutainers at 10-min intervals over a 30-min period, and quantifying the N_2O in the laboratory with GC.

4.9 Analysis of Soil Nutrients

Utilization of soil as a medium to grow crops creates a need for measurement of nutrient *availability*. In contrast to total chemical analysis, the objective in *soil test* analysis is to characterize nutrient element availability to plants and assess the presence of potentially toxic elements. The soil test methodology involves selecting an appropriate chemical extraction technique which has been calibrated to crop response, for instance the level of extracted nutrient is related to crop growth and tissue nutrient concentration. This generally involves the use of mild chemical extractants that approximate the nutrient-absorbing capability of plant roots. For example, the soil test may provide a measure of nutrients in the soil solution plus those present on exchange sites. Often, measurements of ''plant-available'' nutrients are soil-type and climate specific and are directed mainly to agronomic situations where a relatively high level of plant nutrients is required. In reality there is not a well-defined form

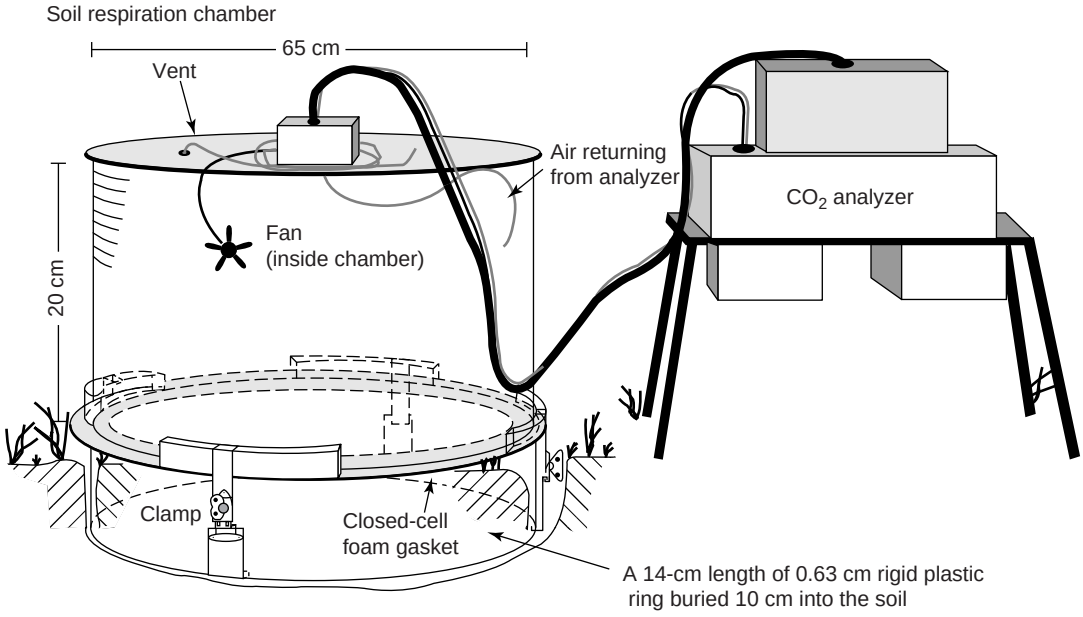

Figure 2 Closed chamber system for use in the field using a portable CO_2 analyzer.[24] (Reproduced by permission of Agricultural Institute of Canada.)

or pool of plant-available nutrient in the soil, rather a dynamic equilibrium exists among solution-exchangeable precipitated forms. Under natural ecosystems, where plant-available nutrients can be low, the rate of internal cycling may be more important than maintenance of a specific nutrient level.

A wide range of soil test extractants is available, particularly for water-soluble elements. Measurement of exchangeable nutrients (e.g. K, NH_4) requires utilization of another cation at sufficiently high concentration that it will replace most of the target nutrient. Ammonium acetate (1 M NH_4OAc) is commonly used to test for soluble and readily exchangeable Ca, Mg, and K. Some K may also be made available to plants through release of the element from within the structure of K-bearing minerals such as micas. This so-called non-exchangeable K fraction is estimated by boiling in strong acid such as 1 M HNO_3. For N, where plants utilize only the mineral forms NO_3 and NH_4, both KCl and K_2SO_4 are common extractants.[1]

Extraction with sodium bicarbonate ($NaHCO_3$) is used to test for available phosphate.[1,3] Phosphorus resides in a variety of inorganic and organic forms in soil, and sequential extraction schemes have been devised to fractionate soil P into various components.[25] Where detailed information is needed on a soil's P chemistry, the following sequence of extractants may be used: anion-exchange resin for soluble and readily-available P, 0.5 M $NaHCO_3$ for weakly-adsorbed P, NaOH for strongly adsorbed P, dilute HCl for P in calcium phosphate compounds, and concentrated HCl for residual P.[25]

In the case of S, soluble sulfate can be extracted using a weak $CaCl_2$ solution, while sulfate adsorbed to soil surfaces is removed using a dilute phosphate solution.[3]

Many elements that are only required in small amounts for plant growth may be toxic at higher concentrations. Also, other elements such as heavy metals, some of which are essential elements for plant growth, can at certain concentrations adversely affect animal and human health. For these elements selective extractants are required.[1,3] For example, micronutrient cations (Fe, Mn, Zn, Cu) and heavy metals such as Cd, Ni, and Pb that are naturally present in the soil or added as pollutants can be extracted using chelating agents, e.g. diethylenetriaminepentaacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA). Micronutrient anions (e.g. B, Mo, and Se) can be removed with various extractants such as hot water or NH_4HCO_3 + DTPA. Aluminum is a nonessential element that may be present in the soil solution at toxic concentrations under acid conditions. There is no specific soil test to assess Al toxicity, rather soil pH and the exchangeable Al level may be used as indirect indicators of a potential Al toxicity problem.

Multielement soil extractants, such as the following Mehlich III reagent, extract several "available" plant nutrients simultaneously:

- the extractant is composed of 0.2 M CH_3COOH, 0.25 M NH_4NO_3, 0.015 M NH_4F, 0.013 M HNO_3, and 0.001 M EDTA;
- phosphate is extracted by CH_3COOH and NH_4F;
- K, Ca, and Mg are extracted by NH_4NO_3 and HNO_3;

- the micronutrients Fe, Zn, Mn, and Cu are extracted by the chelating agent EDTA.

Nutrients in these extracts can be determined by colorimetric methods, AAS or ICPAES, as appropriate. The ICPAES technique offers advantages over other methods because matrix interferences are eliminated or minimized by use of high-temperature plasma, allowing multielement (e.g. K, Ca, Mg, Fe, Mn, Zn, Cu) determinations to be performed simultaneously with little or no sample preparation.[20,21]

5 INSTRUMENTAL METHODS FOR SOIL BIOCHEMICAL ANALYSIS

Soil biochemical quality is important for several soil functions relating to plant growth and the regulation of soil water and energy flows (Table 1). Specific processes involved are:

- supply balance of plant nutrients
- provide suitable medium for microbes and fauna
- influence soil water storage and movement
- store and release (recycle) energy-rich organic matter.

The main attributes or soil properties used to characterize the above processes are organic matter and various fractions of organic matter (e.g. microbial biomass, macroorganic matter, light fraction), soil respiration, N mineralization, enzymes, and fauna. Figure 3 illustrates the composition of soil organic matter.

Organic matter is the prime energy source for soil organisms and as such is a central component of soil biological processes. Changes in the quantity of total soil organic matter are especially reflected by changes in the labile light fraction and macroorganic matter. The microbial biomass is both a sink and a source for nutrients and is the main agent for organic matter recycling. The soil fauna play a key role in the comminution and redistribution of organic matter in soil and thus have a profound influence on decomposition processes. Soil respiration, N mineralization, and enzymes are measures of microbial activity.

5.1 Isotopic Methods for Organic Matter

The redistribution of added substrate and residues among organic-matter fractions, particle-size or density

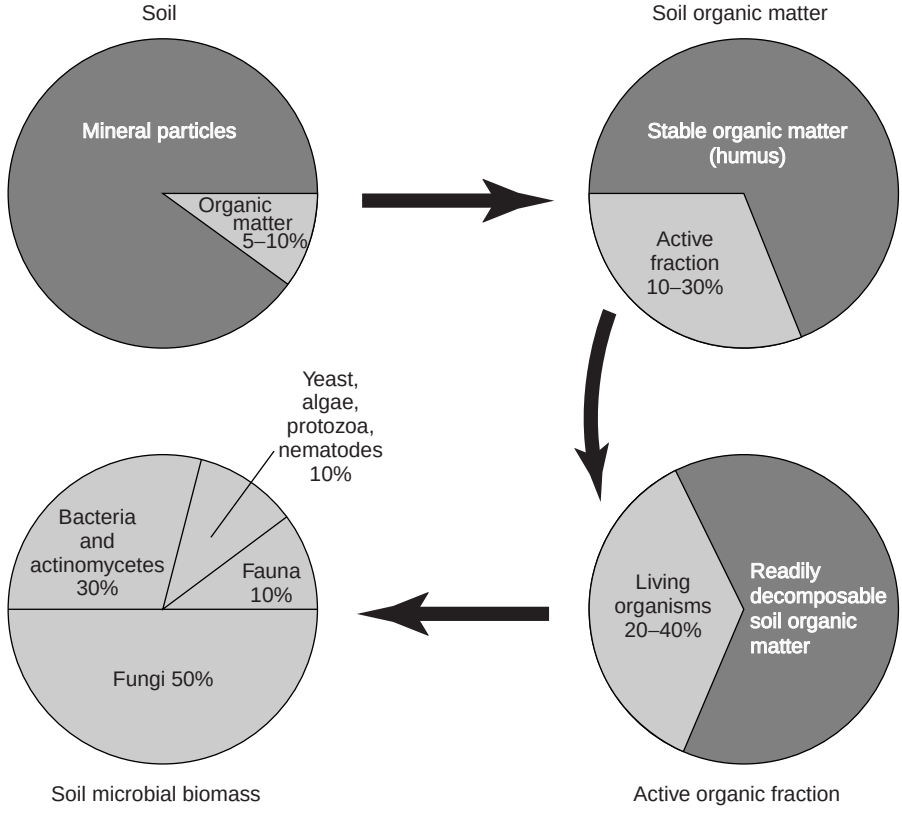

Figure 3 Composition of soil organic matter, the active fraction, and the soil microbial biomass.[6] (Reprinted from *Soil Quality for Crop Protection and Ecosystem Health*, Copyright 1997, 52, with permission from Elsevier Science.)

fractions, and aggregate-size fractions in soil is traceable with isotopes.[4] Various radioisotopic techniques have been widely used since the 1960s. Recently, however, instrumentation has become more sophisticated and readily available, allowing widespread use of stable isotopes in soil organic matter studies.The benefit of using isotopic methods is that they allow precise quantification of changes in organic matter during decomposition. Generally, ^{14}C is used to study short-term dynamics of organic matter, and the natural abundance of ^{13}C and ^{15}N is used to characterize the long-term status.

For finer levels of resolution, enrichment of a substrate with labelled C or N is needed to distinguish the added organic matter from soil organic matter, because the amount of C or N added is usually small relative to the native C or N present in soil. Substrates labelled with ^{14}C or ^{15}N that have been used in this way range from simple compounds, such as glucose and nitrate, to complex compounds, such as cellulose and proteins, to plant leaves and roots. This approach has been used to study rapid mineralization processes (usually less than five years) and the stabilization of decomposition products over the long term. Use of a ^{13}C- or ^{15}N-labelled substrate offers the potential to measure precise quantitative changes as well as to monitor the changes, as seen by ^{13}C or ^{15}N-nuclear magnetic resonance (NMR), in chemical structure associated with the decomposition and stabilization of the added material.

Difficulties previously encountered in dating soil organic matter in samples with relatively small masses, for ^{14}C dating and bomb ^{14}C, has been greatly reduced by using accelerator mass spectrometry (AMS). This technique reduces the sample size necessary for ^{14}C measurements by a factor of 2000 over conventional counting methods.[26] This refinement has allowed the determination of high-precision dates on as little as several tenths of a gram of sample, such as that of aggregate or size fractions separated from whole soils. Radiocarbon dating of soils is based on the decay of ^{14}C in plant material which had a $^{14}C/^{12}C$ ratio similar to that of atmospheric CO_2 at the time of its death and incorporation into the soil. The degree to which the $^{14}C/^{12}C$ ratio of soil organic matter differs from that of the plant material from which it is derived reflects the ''mean residence time'' of C in soils. Dates for whole soils and soil fractions are expressed as mean residence times, because the annual input of new plant material into soils contaminates the soil with modern C. For the purposes of assessing soil organic matter dynamics, ^{14}C dating is done on samples that have not been enriched with bomb ^{14}C, as described below. These soils are usually archived samples that were taken prebomb (before the 1960s) and stored in glass jars, or were obtained from under buildings. Various methods are available to calculate, report, and interpret ^{14}C.[26]

Detonation of thermonuclear devices in the 1950s and 1960s resulted in the enrichment of the atmosphere with ^{14}C. This spike input of bomb ^{14}C to the stratosphere subsequently entered the terrestrial ecosystem through plants and has then recycled through animals, microorganisms, and soil organic matter.[26] Measurement of the rate at which this pulse of ^{14}C moves through different fractions provides a unique opportunity to bridge the gap between the rapid turnover rates measured in labelling studies and the slow turnover rates measured in radiocarbon dating studies. Bomb ^{14}C is determined by comparing ^{14}C concentrations in soil sampled before 1960 (prebomb) with those in ^{14}C-enriched modern soils. The modern soil sample should be of the same soil type and should be collected near the location of the prebomb samples.

Natural ^{13}C and ^{15}N abundance is another isotopic method used to characterize organic matter.[4] Plants with a C4 photosynthetic pathway have higher $^{13}C/^{12}C$ ratios ($\delta^{13}C$ of -12 to -14%) than C3 plants ($\delta^{13}C$ -26 to -28%). The natural abundance of ^{13}C and ^{15}N is expressed relative to an international standard. The δ values are calculated from the measured mass isotope ratios of the sample and standard gases using Equation (5), where R is the mass 45/44 ratio of the sample or standard gas:

$$\delta^{13}C(\%) = \frac{R\ \text{sample} - R\ \text{standard}}{R\ \text{standard}} \times 1000 \qquad (5)$$

The use of natural ^{13}C abundance as a tracer is based on the facts that the ^{13}C content of soil organic matter corresponds closely to the ^{13}C content of the plant material from which it is derived, and that there is negligible fractionation of ^{13}C during the decomposition of plant material in soil. Thus the introduction of C4 plants to a soil previously developed under C3 vegetation (or vice versa) results in that soil containing two isotopically different sources of C and provides a means of partitioning soil organic matter according to origin. This approach has been used to determine the storage of soil organic matter and turnover of plant residue C in particle-size fractions in a wide range of soils.

Nitrogen enters soils either directly through nitrogen fixation and atmospheric deposition of nitrous oxides, or indirectly through the decomposition of plant and animal residues. Atmospheric N_2 has a ^{15}N abundance of 0.366% and is generally used as a standard The $\delta^{15}N$ values in living plants and residues, which vary from -9% to $+8.0\%$, increase quickly (to 10% or more) with increasing microbial transformation of the residues in litter, because of the microbial discrimination of ^{15}N during the catabolic processes occurring in soil. Thus

successive microbial transformations of N-containing material cause a progressive increase in ^{15}N of the soil organic matter. Changes in the $\delta^{15}N$ values allow information to be obtained on N_2 fixation in plants and N transformations (e.g. mineralization, immobilization, nitrification) in soil.[2]

5.2 Analysis of Organic Matter Fractions

Organic matter C is composed of 10–25% carbohydrates, 20% amino sugars and amino acids, and 10–20% fatty acids or lipids, with the remainder as aromatic humic compounds. The latter are a complex mixture of organic compounds.[27] Some organic matter fractions (e.g. particulate or macroorganic matter, and light fractions) are transitory intermediates between plant litter and stable organic matter and exist in soil as fragments of recently added plant material and nonhumified material (Figure 4).

These fractions can be separated from the soil using physical methods (i.e. sieving and density separation) prior to chemical analysis,[1] as follows.

- *Macroorganic matter*: an air-dried <2 mm soil sample is dispersed using sodium hexametaphosphate and the soil suspension is passed through a 50 μm sieve. The sand plus organic matter on the sieve is dried and analyzed for organic C.
- *Light fraction*: an air-dried <2 mm soil sample is mixed with a heavy liquid (usually between 1.5 and 2.0 g cm^{-3} at a soil/heavy liquid ratio of 1 : 2) such as sodium iodide or sodium metatungstate. After shaking, the solution is allowed to stand for ~48 h and the floating material is removed, washed, dried, and analyzed for organic elements. Alternatively the mixture of soil and heavy liquid may be shaken and centrifuged, and the organic matter floating on the dense solution recovered and analyzed.

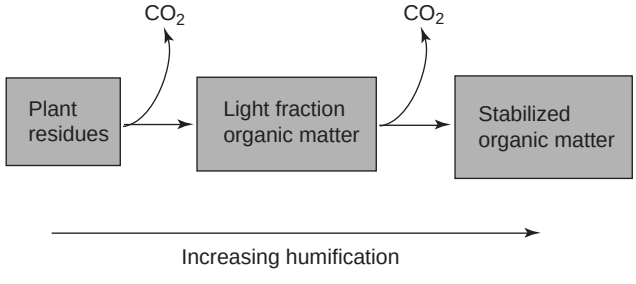

Figure 4 Light fraction organic matter as a transitory intermediate between plant litter and stable organic matter.[4] (Reprinted with permission from M.R. Carter, B.A. Stewart (eds.), *Structure and Organic Matter Storage in Agricultural Soils*, Copyright Lewis Publishers an Imprint of CRC Press, Boca Raton, FL, 1996.)

Soil carbohydrates are generally extracted by acid hydrolysis (e.g. concentrated H_2SO_4) to release saccharide monomers and the hydrolysate is analyzed for individual monosaccharides by liquid chromatography (LC) and GC, or for total saccharides by colorimetric methods.[1,8] Owing to the great variability in soil carbohydrate complexity and resistance to hydrolysis (e.g. cellulose), and the need to extract carbohydrates that are implicated in soil aggregation processes, dilute acid (e.g. 0.5 M H_2SO_4) is often used as the extractant. Nitrogen compounds (e.g. amino sugars or acids) are also extracted by acid hydrolysis (e.g. using concentrated HCl) and the neutralized hydrolysate is analyzed for the organic form of N using the Kjeldahl steam distillation procedure or by colorimetric methods.[1,3] Soil lipids are extracted using organic solvents (e.g. chloroform) which separate the lipids from their association with carbohydrates and proteins, and the extract is analyzed for lipid composition by chromatographic fractionation.[3] In many cases, an acid pretreatment is needed to separate lipids bound by clay minerals or iron and aluminum salts.

There are several spectroscopic techniques for chemical characterization of soil organic matter: NMR spectroscopy is a nondestructive technique to study organic C because the organic matter is not extracted from the soil for analysis. The results therefore provide a chemical view of the organic material as it exists in soil. A range of NMR techniques has been used to study organic materials in soils on different nuclei (^{13}C, ^{15}N, and ^{31}P) in whole soil or soil fractions in solutions and solids. Cross-polarization/magic angle spinning (CP/MAS) methods are often used to enhance signal intensity and obtain well-resolved spectra. Solid-state ^{13}C-NMR has been used to characterize organic C in different chemical structures that are differentiated on the basis of chemical shift (i.e. units of parts per million of applied magnetic field). Because soil organic matter comprises a heterogeneous mixture of different types of organic C, the spectra for most mineral soils with <50 g C kg^{-1} soil are usually divided into four regions of chemical shift: alkyl C (0–45 ppm), *O*-alkyl (45–110 ppm), aromatic C (110–165 ppm) and carbonyl C (165–220 ppm).[28] These C compounds denote the most dominant form of organic C present within each region. The integral of the signal intensity across a chemical region is usually expressed as a proportion of the total integrated area of the spectrum and the NMR results should not be considered quantitative unless detailed NMR experiments are performed.[28]

Other spectroscopic techniques that have been used to characterize the chemistry of soil organic matter involve mass spectrometry. In pyrolysis/mass spectrometry (py/MS) the sample is pyrolyzed (i.e. heated rapidly

at high temperature to release organic material) directly in the vacuum of a mass spectrometer.[27] The components can be identified using several ionization methods, including field, chemical, and laser ionization as well as fast atom bombardment. The spectra result in a pattern of pyrolysis products with signals at specific values which correspond to particular types of organic molecules. A fundamental assumption underlying these techniques is that little fragmentation occurs during the sample pyrolysis.

The development of techniques to separate humic compounds was driven partly by the need to isolate soil organic matter in forms that are both high in C and relatively soluble. The traditional approach to analyze soil humic compounds is by extraction with alkali (NaOH) followed by acidification (HCl) of the extract to obtain the humic, fulvic, and humin fractions. The three fractions can be analyzed by various chemical methods to obtain their overall chemical composition. However, since humic compounds are not discrete substances, but rather are heterogeneous and complex in nature and cannot be defined in specific chemical terms, information on their composition alone is of limited value. Most interest is directed to characterizing the major functional groups (i.e. carboxyl, alcohol, phenolic hydroxyl, and carbonyl) that provide information on the structural composition of humic compounds and how they function in soil and water ecosystems. It is imperative that a standard methodology be used in separating humic compounds because small changes in the type and concentration of the solute during the extraction significantly alter the partition of C into these compounds. Several instrumental methods are available to study the structure of humic substances, such as ^{13}C NMR spectroscopy, Fourier transform infrared (FTIR) spectroscopy, electron spin resonance (ESR) spectroscopy, py/MS, and supercritical gas extraction (SGE).[27]

5.3 Microbial Biomass

The magnitude of the soil microbial biomass is of interest, especially for ecological studies concerned with the biomass as a sink and source of nutrients. Light microscopic methods have been developed to estimate bacteria and fungal biomass by using cell length and width dimensions, and a dry weight/cellular volume conversion factor.[2] However, such methodology is time-consuming. Several techniques have utilized the change in the soil respiration rate, as a result of exposing the soil to a fumigant or by adding a C substrate (e.g. glucose), to estimate microbial biomass.[2,3] In the former approach, a soil sample is fumigated under specific conditions with a fumigant (e.g. $CHCl_3$ vapor) to kill the living microbes. The resulting increased decomposition, in comparison to an unfumigated soil, as measured by the release of CO_2 over a specific period of incubation (i.e. 10 days) is directly proportional to the size or amount of C in the biomass. The release or flush of CO_2 after fumigation is the result of germinating microbial spores utilizing the C substrate provided by microbial cells killed by the fumigant. Fumigation techniques are suitable for estimating the C, N, P, and S content of the soil microbial biomass.[1–3] Replacing this incubation step by directly extracting the soil immediately after fumigation provides a relatively rapid chemical method to estimate the size of the biomass. The following list gives the analytical sequence for determining microbial biomass C using the $CHCl_3$ fumigation–extraction method.

- *Soil sampling and preparation*: soil cores are obtained according to an acceptable sampling scheme to provide a measure of soil variability. The field-moist samples can be stored for up to a week at 4 °C or stored overnight at 15 °C. Soil is passed through a 4–6-mm sieve prior to analysis to homogenize the sample. Three soil subsamples (e.g. 15–50 g) are placed in glass beakers one of which is immediately extracted with K_2SO_4 (see final step).
- *Fumigation treatment*: under a fumehood, the two remaining soil subsamples are placed in a desiccator, along with a beaker containing purified or heptachlor epoxide stabilized (i.e. ethanol-free) $CHCl_3$ (50 mL) and boiling chips. A vacuum is applied until the $CHCl_3$ boils vigorously, then the sealed desiccator is maintained under vacuum in the dark at 25 °C for 24 h.
- *Extraction of microbial biomass C*: the vacuum in the desiccator is carefully released, the $CHCl_3$ removed (and disposed of as a hazardous waste), the beakers containing the soil samples removed, and 0.5 M K_2SO_4 added to each beaker (to give 1 : 2 to 1 : 5 ratios of soil to extract). After shaking for 1 h, the soil suspension is filtered and the filtrate analyzed for organic C on a total organic C analyzer.

After adjusting the soil weight to an oven-dry basis (using Equation 6) and correcting for the total volume of the solution in the extracted soil (Ext. mL = soil wet wt. – soil dry wt. + mL of K_2SO_4 added), the total weight of extractable C in fumigated (CF) (i.e. average for the two samples) and in unfumigated (CUF) soil, and the microbial biomass C are calculated using Equations (7) and (8)

$$\%H_2O\ (\text{mass}) = \frac{(\text{mass of wet soil, g}) - (\text{mass of dry soil, g})}{\text{mass of dry soil, g}} \times 100 \quad (6)$$

$$\text{CF or CUF } (\mu g\, g^{-1} \text{ soil}) = \text{extractable C } (\mu g\, mL^{-1}) \times \frac{\text{ext. mL}}{\text{g dry soil}} \quad (7)$$

$$\text{Microbial biomass C } (\mu g\, g^{-1} \text{ soil}) = \frac{CF - CUF}{0.25} \quad (8)$$

where 0.25 is the efficiency of extraction (k_{ec}) of microbial biomass C based on in situ calibration techniques.[1]

5.4 Analyzing Microbial Processes

Microbial activity can influence the mineralization and transformation of many soil elements, such as C and N, and influence the composition of the soil air (e.g. CO_2, N_2O, NO, O_2). Most methods devised to characterize microbial processes in soil are concerned with estimating C and N transformations. However, methods are also available to assess other microbially mediated transformations such as the oxidation and reduction of S, Fe, and Mn.[2] Carbon mineralization involves the biological oxidation of organic C, where CO_2 is released to the soil atmosphere from metabolizing soil microbes and fauna. Soil respiration is defined as the sum of the soil's metabolic activities that produce CO_2.[2] In order for the method to give an accurate measure of the activities in soil it is important that the conditions of the incubation (i.e. temperature and soil moisture) be carefully controlled under standardized conditions. Laboratory methods to measure soil respiration generally involve use of incubation chambers, where CO_2 evolved from a soil sample is trapped in an alkali solution as carbonate, which can then be measured by titrimetric, gravimetric, or conductimetric analysis. For analysis of CO_2 in gaseous samples, GC or IR gas analysis can be used. In the static method, titration of alkali solutions exposed to incubated soils determines the quantity of NaOH not reacted with CO_2.[1] The dissolved CO_2 is precipitated as $BaCO_3$ and phenolphthalein is used as the end-point indicator. The steps in the static method are as follows.

- A soil sample (W, g dry soil) in a beaker is placed in an incubation jar, along with a beaker containing a 1 M alkali solution (e.g. KOH, NaOH), and the jar is sealed. A water blank and a NaOH blank are also incubated separately.
- After a period of time the NaOH beaker is removed from the jar, and an excess of 1.5 M $BaCl_2$ with a few drops of 1% phenolphthalein indicator solution are added to the beaker. The solution is then titrated with 1 M HCl (S). Both the water and NaOH blank (B) are also titrated. Equation (9) is used to calculate mg CO_2–C g^{-1} soil for the incubation time.

$$\text{mg } CO_2\text{–C g}^{-1} \text{ soil} = \frac{\left(\begin{array}{l}\text{mL HCl used in } B \\ \quad - \text{mL HCl used in } S\end{array}\right) \times 6}{W} \quad (9)$$

The molarity of both the alkali solution used to trap the CO_2 and that of the acid used for titration can be adjusted according to the rate of soil respiration in order to accommodate titrating smaller volumes of acid.

Nitrogen mineralization involves the conversion of the various organic N components in soil to inorganic N. Methods to characterize this phenomenon usually employ incubation techniques similar to those used for C mineralization. The importance of N for plant growth has created a need to estimate the N mineralization potential in soil. This potential is determined by incubating a soil sample under standard conditions and periodically leaching the sample (e.g. using 0.01 M $CaCl_2$) and measuring the NO_3–N and NH_4–N in the leachate over time, until the cumulative N mineralized approaches a plateau.[1,2] Both the amount of cumulated N and the duration of incubation are used to calculate the N mineralization potential.

Biological denitrification is the stepwise process where NO_3 is reduced to N_2O and N_2. These gases can be quantified using a continuous-flow column approach or gas collection chambers (see section 4.8), although this methodology is mainly used for characterizing the soil–atmosphere exchange of gases rather than the denitrification rate.[2] Denitrification can also be estimated in the laboratory on soil cores obtained from the field, using the acetylene block technique. The acetylene blocks the last step in the denitrification process (i.e. $N_2O \longrightarrow N_2$). The accumulated nitrous oxide (N_2O), measured by GC, provides a measure of the soil denitrification rate.[1,2] In the presence of trace amounts of O_2, however, the accumulation of N_2O can be impeded as acetylene reacts with NO (a precursor to N_2O) reducing the production of N_2O and consequently causing a potential underestimate of denitrification. The specific sequence for the acetylene block technique is as follows.

- Soil cores are weighed (W, g) and placed in a glass jar sealed with a serum stopper. A 10 mL sample of gas is removed from the sealed jar and replaced with 10 mL of high-purity acetylene.
- After time to allow for acetylene diffusion through the soil core, a gas sample is removed from the sealed jar and the N_2O concentration measured using GC. In most cases, the amount of dissolved N_2O in the soil water is also calculated to give a measure of the total N_2O produced.

Microbial processes in soil are mediated by enzymes that serve as biological catalysts. They control organic matter decomposition and their activities reflect conditions in the ambient environment (i.e. soil moisture and other environmental parameters). Generally, laboratory methods used to measure enzyme activities in soil involve the incubation of a soil sample with a specific substrate, under standard conditions, followed by determination of the rate at which the substrate is converted to a measurable product.[2,3]

6 INSTRUMENTAL METHODS FOR SOIL PHYSICAL ANALYSIS

Soil physical quality is important for many soil functions such as the medium for plant growth, and regulation of water and air[29] (Table 1). Specific processes involved are:

- provision of a suitable medium for root growth and development
- acceptance, retention, and release of water
- permeability for air and water movement.

The main attributes or soil properties used to characterize the above processes are clay and organic matter content, hydraulic conductivity, pore size distribution, macropore space and organization, oxygen diffusion rate (ODR), bulk density, soil strength, and aggregate stability.[29] Many of these properties reflect the ability of soil to store and transmit water, solutes, and gases. Inherent soil properties, such as clay content, influence the magnitude of the soil water holding capacity. This is also augmented by the organic matter content. Dynamic soil properties, such as soil density and porosity, control the soil storage capacity for water and air. Pore size distribution is considered to be a good indicator of the soil structural condition and has proven useful for predicting water infiltration rates, water availability to plants, soil water storage capacity, and soil aeration status.[29] Macroporosity, air permeability, and/or ODR provide a measure of soil macropore continuity and organization. Soil strength and aggregate stability reflect a soil's ability to resist compaction and other stresses that can lead to loss of structure.[29]

Instrumentation, based on analytical chemical methods, is available for the analysis of soil water, porosity and aeration, and macrostructure and bulk density.

6.1 Soil Water Analysis

The water content of soil is one of the most commonly measured parameters in soil analysis.[30] A simple approach to estimate total water content is to determine the loss of mass after oven-drying a soil sample at 105 °C to a constant mass. The analysis can be expressed as the ratio of the mass (or the volume, i.e. assuming the bulk density of H_2O to be equal to one) of water and dry soil as shown in Equations (6) and (10).

$$\%H_2O\ (\text{volume}) = [\%H_2O\ (\text{mass})] \times \text{soil bulk density} \quad (10)$$

In many cases the ratio of the volume of water and dry soil is used, especially if the soil is a swelling soil. Soil water content can range from near zero to a variable level associated with the volume of the soil pore space. In addition to oven-drying procedures to calculate the mass of soil water, various instrumental techniques are available.[1,7] TDR is based on the electrical properties of the water molecule, specifically the dielectric constant of water which is much higher than that of soil solids or air.[31] Thus, the dielectric constant of soil is related to the water content. The TDR method is useful for both laboratory and field use. For mainly field studies, or with large bulk soil samples, the neutron thermalization (NT) method and the γ-ray attenuation (GRA) method can be used. NT uses high-energy fast-moving neutrons, which are scattered and slowed to the level of thermal energy by collisions with nuclei of similar mass and size, such as the H nuclei. As most H in soil is in water, the intensity of thermal neutrons provides a measure of soil water content. GRA uses γ-rays which are passed through the soil and absorbed by soil particles and water. Thus, under conditions where soil solids are maintained at constant density the degree of GRA or reduction in intensity is related to soil water content.

Water in soil is held in a certain physiochemical state that can be characterized in terms of free energy per unit mass or *potential*.[1,7] The potential (Φ) is related to the sum of the various forces in soil that hold on to water (e.g. matric, osmotic, gravitational). Matric potential arises from the attraction of the soil solids for water, while the osmotic potential relates to the influence of dissolved solutes on water activity. For methodology purposes, water potential relates to the amount of work expressed in terms of pressure (i.e. megapascals) required to remove water from the soil system. Matric potential can be measured using resistance blocks, tensiometers, and water desorption curves. Resistance blocks contain electrodes embedded in a matrix of porous material such as gypsum, fiberglass, or nylon. After equilibration in soil, the EC of the block can be related to the soil matric potential. Resistance blocks are generally better suited to dry soil conditions and have the disadvantage of a slow response time to changes in soil water. Tensiometers are composed of a porous ceramic cup attached to a tube with an airtight cap and a side arm containing a device

to measure the pressure of the water in the tube. The porous cup allows water to move freely between the soil and tube, and thus the hydrostatic pressure within the tube can be related to the soil matric potential. Use of tensiometers is generally restricted to wet soil conditions.

The thermal properties of soil are related to soil water content. On this basis, the TC of soil has been used as an indirect measurement for soil matric potential. TC sensors have been developed that use the thermal properties of a standard porous medium (e.g. ceramic porous media).[30] The sensor, placed in a soil, will at equilibrium give TC measurements that are related to the soil matric potential.

6.2 Characterizing Porosity and Aeration

Soil porosity parameters reflect the character of soil as a three-phase system of solids, liquids, and gases. Porosity measurements are useful indicators of the soil physical condition, especially the distribution of the total pore space into different categories of relatively large diameter pores that influence the movement of water and air.

Total soil porosity (St) is a measure of the volume percentage pore space and can be estimated from soil dry bulk density (Db) and soil particle density (Dp) as shown in Equation (11).

$$\mathrm{St} = 1 - \left(\frac{\mathrm{Db}}{\mathrm{Dp}} \right. \tag{11}$$

The Db parameter can be considered as the ratio of the mass of oven-dried soil solids to the bulk volume of the solids plus pore space, while Dp is often assumed to approximate the Dp of quartz (i.e. $2.65\,\mathrm{Mg\,m^{-3}}$). Greater accuracy for Dp can be obtained by the use of a pycnometer.[7] In moist soils there is usually a need to further characterize the total pore space into air-filled porosity and water-filled porosity, and to obtain a measure of pore size distribution.[1] The air-filled porosity (Fa) is the fraction of the soil bulk volume occupied by air at a certain moisture content, or preferably the soil matric potential.

The water-filled porosity parameter (Fw), often called ''relative saturation'', expresses the volume of water in soil relative to St. It can be simply calculated from St and Fa as follows: Fw = St − Fa. Generally, a Fw over 65–70% indicates anaerobic conditions.

These parameters of soil porosity are usually ''static'' as they provide little information about the movement of air or water in the soil. In many cases, both the size and continuity of pores is of interest. The latter characteristic is the main influence on air and water movement in soil. Micromorphological methods can be used to directly determine the shape and continuity of pores > 100 μm.[1,32] Soil cores can be sectioned and visible soil pores segregated or classified (e.g. macrostructure of pores > 200 μm) on the basis of size and shape to provide a semiquantitative analysis. Noninvasive instrumental methods, such as computed tomography (CT), have been developed to obtain images of the distribution of pore space.[33] Several types of CT techniques are available based on X- or γ-ray transmission, NMR, and ultrasound.

A sensitive method of assessing pore continuity involves use of a gas at low pressures. In contrast, use of water to assess pore continuity may significantly change the soil pore structure over time. Air permeability can be measured in the laboratory, using soil cores, by the constant pressure gradient method.[1,7,32] In this method an air flux, obtained from a pressurized air bottle and regulated by a pressure gauge, is passed through a soil core of a certain length (L, cm) and cross-sectional area (A, $\mathrm{cm^2}$) at a constant low pressure (e.g. 2.5 cm water) as measured with a water manometer, and the volume of air (V, $\mathrm{cm^3\,s^{-1}}$) passing through the core is measured with a flow meter. A low air pressure is used to prevent turbulent flow in the soil core which may change the pore size distribution as the air flows through the soil. The air permeability coefficient (K, $\mathrm{cm^2}$), termed the intrinsic permeability to air, is calculated as shown in Equation (12).

$$K = (V)\left(\frac{L}{A}\right.\quad(7.4 \times 10^{-8}) \tag{12}$$

The constant 7.4×10^{-8} represents the air viscosity at 22 °C and a constant air pressure of 2.5 cm water (0.25 kPa).

A useful method of assessing soil aeration involves the use of a platinum microelectrode to measure the ODR through the liquid phase of the soil.[1,7] The microelectrode (cathode) is connected by an electrical circuit to a reference Ag–AgCl electrode (anode) which has a KCl-saturated agar gel salt bridge. The procedure is given in the following list.

- The platinum microelectrode (A, surface area, $\mathrm{m^2}$) is inserted into the soil or soil core along with the reference electrode which is connected to the soil via the salt bridge.
- A voltage is then applied (usually −0.65 V or less) between the electrodes, and the steady-state current (i, μA) is measured. The latter is proportional to the flux of O_2 at the surface of the platinum microelectrode.
- The ODR ($\mathrm{\mu g\,m^{-2}\,s^{-1}}$) is then calculated as shown in Equation (13).

$$\mathrm{ODR} = \left(\frac{i}{A}\right.\quad(8.3 \times 10^{-5}) \tag{13}$$

The constant 8.3×10^{-5} equals the molecular weight of O_2 (32) divided by the product of the number of electrons (4) required to reduce one molecule of O_2 and the Faraday constant (96 500 coulombs per mole of O_2). The microelectrodes should be periodically cleaned to remove soil colloidal material or salts that could influence the potential and subsequently the output current. The method is conducive to sequential measurement by multiple microelectrodes, which can be retained in the soil at different depths. The platinum electrode, as a combined unit with the Ag–AgCl anode and KCl electrolyte, can also be used as an electrode to measure directly O_2 concentration in both the gaseous and solution phases of the soil.

6.3 Macrostructure and Density

Soil structure has been variously defined but in the broadest sense can be described as the spatial arrangement of both soil particles and voids. In this regard, soil structure has significant effects on air–water relationships in soil.[4,6,32] The structure of soil secondary particles can be placed into three broad categories as follows:

- single grained (sometimes referred to as 'structureless'), completely unattached particles as in sands;
- massive, tightly packed particles (e.g. compacted clay soil);
- aggregated, where primary particles associate to form secondary particles or aggregates.

Soil aggregates formed under natural conditions (called peds), without the aid of soil tillage or other forms of soil disturbance, are characterized by their shape, arrangement and size. However, the shape and size of aggregates in cultivated soils can be greatly modified by agricultural practices. Thus, the emphasis is placed not only on the size and arrangement of aggregates, but also on the stability of the aggregate arrangement.

The formation of soil aggregates is dependent upon both abiotic and biotic factors, the former being mainly related to soil clay content and the capacity for natural (e.g. alternating shrink–swell, freeze–thaw, wet–dry) structure-forming processes. The aggregation process is not totally random but occurs with some degree of order, in that for many soil types where biotic factors are the main aggregating agents small soil particles are apparently stabilized in large particles.[4,34] Although soil specific, this ordering or arrangement of particles and aggregates can be viewed conceptually as a hierarchy consisting of three main orders:[4]

- clay microstructures (<2 μm diameter)
- microaggregates (2–250 μm)
- macroaggregates (>250 μm).

In soils (e.g. Oxisols) where inorganic compounds (Fe and Al oxides) are the main aggregating agents this hierarchical order may not be evident. In clay microstructures, clay–organic matter complexes are stabilized by humic acids and inorganic ions (e.g. Ca). Microaggregates are stabilized directly by microbial materials such as polysaccharides, hyphal fragments, and bacterial cells or colonies.[4] In comparison, the formation of macroaggregates and their temporary stabilization is the result of a combination of mechanisms related to plant roots and the activity of soil fungi and fauna. Figure 5 provides a schematic diagram of the structure of micro- and macroaggregates.

Generally, the study of soil structure is usually not related to soil structure itself (i.e. the spatial heterogeneity of the different components or properties of soil) but rather to the "functionality" of soil structure, or how an aspect of structure affects specific soil processes. A useful approach to describe soil structure in agricultural soils, which accommodates both aggregates and pores, is to characterize structure on the following basis:[34]

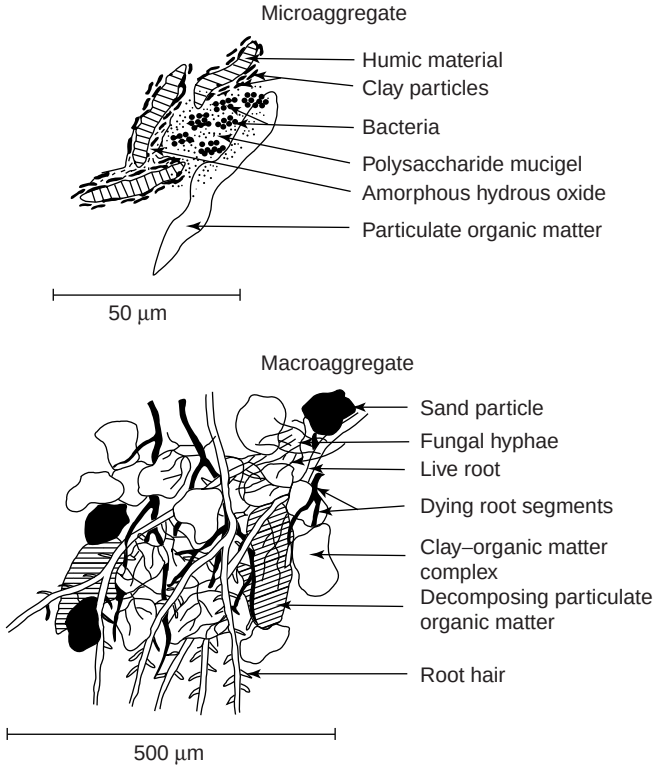

Figure 5 Diagram of the structure and components of soil micro- and macroaggregates.[4] (Reprinted with permission from M.R. Carter, B.A. Stewart (eds.), *Structure and Organic Matter Storage in Agricultural Soils*, Copyright Lewis Publishers an Imprint of CRC Press, Boca Raton, FL, 1996.)

- structural form – arrangement and size of the pore space;
- structural stability – ability to retain the distribution and size of its aggregates after exposure to various stresses (e.g. external forces of impact, shear, abrasion, and slaking);
- structural resiliency – ability of a soil to recover its pore space arrangement after the removal of a specific stress (e.g. compaction).

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AEC	Anion Exchange Capacity
AMS	Accelerator Mass Spectrometry
CEC	Cation Exchange Capacity
CP/MAS	Cross-polarization/Magic Angle Spinning
CT	Computed Tomography
CV	Coefficient of Variation
DC	Dry Combustion
DTPA	Diethylenetriaminepentaacetic Acid
EC	Electrical Conductivity
EDTA	Ethylenediaminetetraacetic Acid
ESP	Exchangeable Sodium Percentage
ESR	Electron Spin Resonance
FES	Flame Emission Spectroscopy
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GRA	γ-ray Attenuation
IC	Ion Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
IR	Infrared
LC	Liquid Chromatography
NMR	Nuclear Magnetic Resonance
NT	Neutron Thermalization
ODR	Oxygen Diffusion Rate
py/MS	Pyrolysis/Mass Spectrometry
SGE	Supercritical Gas Extraction
TA	Thermal Analysis
TC	Thermal Conductivity
TDR	Time-domain Reflectometry
WD	Wet Digestion
XRD	X-ray Diffraction

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Soil Sampling for the Characterization of Hazardous Waste Sites

Pesticides **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

REFERENCES

1. M.R. Carter (ed.), *Soil Sampling and Methods of Analysis*, Lewis Publishers/CRC Press, Boca Raton, FL, 1993.
2. R.W. Weaver, J.S. Angle, P.S. Bottomley (eds.), *Methods of Soil Analysis, Part 2, Microbiological and Biochemical Properties*, 1st edition, American Society of Agronomy, Madison, 1994.
3. A.L. Page (ed.), *Methods of Soil Analysis, Part 2, Chemical and Microbiological Properties*, 2nd edition, American Society of Agronomy, Madison, WI, 1982.
4. M.R. Carter, B.A. Stewart (eds.), *Structure and Organic Matter Storage in Agricultural Soils*, Lewis Publishers/CRC Press, Boca Raton, FL, 1996.
5. K.A. Smith, C.E. Mullins (eds.), *Soil Analysis: Physical Methods*, Marcel Dekker, New York, 1991.
6. E.G. Gregorich, M.R. Carter (eds.), *Soil Quality for Crop Production and Ecosystem Health*, Elsevier Science, Amsterdam, 1997.
7. A. Klute (ed.), *Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods*, 2nd edition, American Society of Agronomy, Madison, WI, 1986.
8. P.R. Hesse, *A Textbook of Soil Chemical Analysis*, John Murray, London, 1971.
9. U.C. Gupta, Y.W. Jame, C.A. Campbell, A.J. Leyshon, W. Nicholaichuk, 'Boron Toxicity and Deficiency: A Review', *Can. J. Soil Sci.*, **65**, 381–409 (1985).
10. J.B. Dixon, S.B. Weed (eds.), *Minerals in Soil Environments*, 2nd edition, Soil Science Society of America, Madison, WI, 1989.
11. D. Heil, G. Sposito, 'Chemical Attributes and Processes Affecting Soil Quality', *Dev. Soil Sci.*, **25**, 59–79 (1997).
12. R.J. Bartlett, B.R. James, 'Redox Chemistry of Soils', *Adv. Agron.*, **50**, 152–205 (1993).
13. R.G. McLaren, K.C. Cameron, *Soil Science: Sustainable Production and Environmental Protection*, Oxford University Press, Auckland, 1996.
14. D. Curtin, J.K. Syers, 'Extractability and Adsorption of Sulfate in Soils', *J. Soil Sci.*, **41**, 305–312 (1990).
15. D.C. Edmeades, D.M. Wheeler, O.E. Clinton, 'The Chemical Composition and Ionic Strength of Soil Solutions from New Zealand Topsoils', *Australian J. Soil Res.*, **23**, 151–165 (1985).
16. G.P. Gillman, L.C. Bell, 'Soil Solution Studies on Weathered Soils from Tropical North Queensland', *Australian J. Soil Res.*, **16**, 67–77 (1978).

17. C.K. Kohut, M.J. Dudas, 'Comparison of Immiscibly Displaced Soil Solutions and Saturated Paste Extracts from Saline Soils', *Canadian J. Soil Sci.*, **74**, 409–419 (1994).
18. F. Adams, C. Burmester, N.V. Hue, F.L. Long, 'A Comparison of Column-displacement and Centrifuge Methods for Obtaining Soil Solutions', *Soil Sci. Soc. Am. J.*, **44**, 733–735 (1980).
19. A. Mubarak, R.A. Olsen, 'Immiscible Displacement of the Soil Solution by Centrifugation', *Soil Sci. Soc. Am. J.*, **40**, 329–331 (1976).
20. A.J. Jeffrey, D.J. Lyons, 'An Inductively Coupled Plasma Atomic-emission Spectrometric Method for Determination of Several Elements in Soil Solutions Proximal to Banded Superphosphate Fertilizer', *Commun. Soil Sci. Plant Anal.*, **19**, 47–58 (1988).
21. P.N. Soltanpour, G.W. Johnson, S.M. Workman, J.B. Jones Jr, R.O. Miller, 'Advances in ICP Emission and ICP Mass Spectrometry', *Adv. Agron.*, **64**, 27–113 (1998).
22. G.L. Hutchinson, G.P. Livingston, 'Use of Chamber Systems to Measure Trace Gas Fluxes', in *Agricultural Ecosystem Effects on Trace Gases and Global Climate*, American Society of Agronomy Special Publication 55, eds. L.A. Harper, A.R. Mosier, J.M. Duxbury, D.E. Rolston, American Society of Agronomy, Madison, WI, 63–77, 1993.
23. K.A. Smith (ed.), *Soil Analysis: Modern Instrumental Techniques*, Marcel Dekker, New York, 1991.
24. P. Rochette, B. Ellert, E.G. Gregorich, R.L. Desjardins, R. Lessard, E. Pattey, B.G. Johnson, 'Description of a Dynamic Closed Chamber for Measuring Soil Respiration and its Comparison with Other Techniques', *Can. J. Soil Sci.*, **77**, 195–203 (1997).
25. M.J. Hedley, J.W. Stewart, 'Changes in Inorganic and Organic Phosphorus Fractions Induced by Cultivation Practices and Laboratory Incubations', *Soil Sci. Soc. Am. J.*, **46**, 970–976 (1982).
26. D.C. Coleman, B. Fry (eds.), *Carbon Isotope Techniques*, Academic Press, New York, 1991.
27. P. MacCarthy, C.E. Clapp, R.L. Malcolm, P.R. Bloom (eds.), *Humic Substances in Soil and Crop Sciences: Selected Readings*, American Society of Agronomy, Madison, 1990.
28. J.A. Baldock, P.N. Nelson, 'Soil Organic Matter', in *Handbook of Soil Science*, ed. M.E. Sumner, CRC Press, Boca Raton, FL, 1999.
29. M.R. Carter, 'Soil Quality Analytical Methods', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, J. Wiley & Sons, London, 1998.
30. G.C. Topp, W.D. Reynolds, R.E. Green (eds.), *Advances in Measurement of Soil Physical Properties: Bringing Theory into Practice*, Soil Science Society of America, Special Publication, No. 30, Madison, WI.
31. G.C. Topp, W.D. Reynolds, 'Time Domain Reflectometry: A Seminal Technique for Measuring Mass and Energy in Soil', *Soil & Tillage Res.*, **47**, 125–132 (1998).
32. W. Burke, D. Gabriels, J. Bouma (eds.), *Soil Structure Assessment*, A.A. Balkema, Rotterdam, 1986.
33. S. Crestana, C. Manoel Pedro Vaz, 'Non-invasive Instrumentation Opportunities for Characterizing Soil Porous Systems', *Soil & Tillage Res.*, **47**, 19–26 (1998).
34. B.D. Kay, 'Rates of Change of Soil Structure under Different Cropping Systems', *Adv. Soil Sci.*, **12**, 1–52 (1990).

Soil Sampling for the Characterization of Hazardous Waste Sites

Brian A. Schumacher, J. Jeffrey van Ee, and Evan J. Englund
United States Environmental Protection Agency, Las Vegas, USA

1 Introduction **1**
2 Sampling Planning **2**
2.1 Terminology 2
2.2 Data Quality Objectives 4
3 Quality Assurance/Quality Control **4**
4 Sampling Network Design **6**
4.1 Basic Spatial Sampling Strategies 7
4.2 Estimating Variances 8
4.3 Spatial Dependence 9
4.4 Sampling for Local Estimation 10
5 Documentation **11**
6 Site Characterization **12**
6.1 Site Observations 12
6.2 Site Mapping 13
7 Sample Collection **14**
7.1 Particulate Sampling Theory 14
7.2 Sources of Variation and Sampling Error 14
7.3 Application of Particulate Sampling Theory to Soil Sampling 18
7.4 Selection of Sampling Equipment 19
7.5 Soil Sampling Tools 21
8 Sample Handling and Preparation **23**
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **24**
References **25**

The planning for and collecting of samples at hazardous waste sites is nearly always the first step in any program that involves the analysis of real-world soil samples. Knowledge of numerous different, yet related, topics, such as data quality objectives (DQO), sampling network design, quality assurance (QA)/quality control (QC), documentation, site characterization, sample collection, and sample handling and preparation, is essential to ensure the success of any program in obtaining accurate (non-biased), precise, and representative soil samples. It is often quoted that during sample collection, 80% or more, for stable contaminants (e.g. heavy metals, polychlorinated biphenyls, polyaromatic hydrocarbons), ranging up to 99.9%, for less stable contaminants (e.g. volatile organic compounds (VOCs)), of the total measurement error associated with a sample occurs. Hence the goal of the program planner is to develop the most efficient sampling design that collects the desired samples while minimizing sample errors. If all of the aforementioned topics are not taken into account during the sample planning and collecting phases of the program, then no matter how good the analytical data may be, the data may lead to incorrect decisions with respect to the extent, distribution, and need for remediation of contaminants at a site.

1 INTRODUCTION

The sampling of soil, or any heterogeneous medium, requires an understanding of the spatial and temporal scales of interest to the decision makers. Depending on how the medium is sampled and analyzed and the data are processed, almost any ''valid'' contaminant concentration can be obtained. In the absence of other data, data from the sampling and analysis of soil (or hazardous waste) is frequently assumed to represent the actual contaminant concentration in the vicinity of the sampling point when emphasis is placed on characterizing the distribution of a contaminant in the soil. When additional data are obtained in the vicinity of that sampling point, differences in contaminant concentrations almost always occur. The dilemma, then, is to try to determine the true concentration of the contaminant within a given area or volume of soil using the most efficient sampling design and minimizing the errors that occur during sample collection.

This article will provide a basic understanding of the processes and factors involved in the sampling of soils at a hazardous waste site. These processes and factors can be divided into several general categories, which include sample planning, documentation, QA, QC, sampling network design, site characterization, sample collection, and sample handling and preparation concerns. Basic guidance on dealing with factors influencing accuracy, precision, representativeness, and comparability of data and how to develop concise statements of the sampling effort objectives will be emphasized. Interpretation of the collected data will be discussed as a means for not only answering the basic objectives of the sampling effort, but also determining whether the quality of the data is sufficient to meet the needs of the user. Information on how to answer some of the most frequently asked questions, such as, ''How many samples are needed and where

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

should they be taken?'', will be provided. Regrettably, no simple answers exist. Once the planning phase is completed and the soil sampler goes to the field, discussion of site features that influence the distribution and ultimate management of the contaminants at the site and how to correctly collect the soil sample will be presented. Unfortunately, no checklists are capable of addressing all of the important issues and factors that influence the sampling and analysis of soils and heterogeneous models. Site characterization and soil sampling are multimedia, multidimensional, multidisciplinary efforts that require vigorous communications between all of the principal parties involved. A genuine grasp of the basic concepts and critical definitions is extremely important to perform efficient and effective soil sampling for characterizing hazardous waste sites. The American Society for Testing and Materials (ASTM) is currently preparing a practical manual describing all of the different aspects of soil sampling at hazardous waste sites that are discussed in this article.[1] The ASTM manual provides ''real-world'' examples (including site data) and covers all aspects of soil sampling from planning through implementation and assessment making the manual a valuable companion document for this article.

2 SAMPLING PLANNING

Typically, at the beginning of a site investigation, historical information is sought to allow hypotheses and conceptual models of the site to be developed. All potential pathways of contaminant movement and exposure routes to living organisms should be identified and addressed. For example, if the contamination is in the soil, one pathway of concern may be the migration of the contaminant to the groundwater. The project planners should use the limited historical information about the site to determine what additional data are needed and where to collect it to address the critical questions of how much risk the site poses to humans and the environment.

Data are collected from a variety of sources and from a variety of locations to develop a ''conceptual model'' of the site. There may be no one complete conceptual model that exists for a given site and other models may need to be developed as the site investigation progresses and new information becomes available. A variety of mathematical models, and sometimes physical models, may be employed to estimate data between collected data points and to predict data at other times, at critical receptors, and in other media.

Far too often, efficient collection of samples and data is difficult because critical definitions and well-defined problem statements are often lacking at the beginning of an investigation, even though the conceptual models may have been well defined. Usually, the controlling factors regarding the amount, quality, and type of data collected are time and money and not whether the results will adequately represent the site, fill the data gaps, or allow for a defensible decision to be made. The traditional thinking is that the more data that can be collected, the better the final conceptual model and decision-making process become.

Data that are collected in a site investigation are often viewed in a spatial context. The increasing popularity of computer-based geographical information systems (GIS) permits the rapid display of both two- (2-D) and three-dimensional (3-D) data. While data from a site investigation are frequently analyzed using classical statistics, geostatistics is increasingly being used. None of these tools, however, can properly be used to answer critical questions about a site unless the critical questions have been properly highlighted and defined. Frequently, the generally accepted critical question that drives an investigation is whether an action level for one, or a variety of contaminants, is being exceeded. This question, by itself, is meaningless unless the ''support''[2] or spatial volume that accompanies the action level is prescribed. Sampling heterogeneous materials, such as soil, can lead to a wide range of valid contaminant concentrations from which comparisons with an action level may be made depending upon the volumes of material that are being represented. For example, the concentration of a contaminant in a 5-g sample may exceed the action level whereas the concentration of contaminant in a soil core sample from which the 5-g subsample was obtained may not exceed the action level.

2.1 Terminology

A number of spatial units need to be considered and defined early in the investigation of a hazardous waste site in order to permit the efficient collection of data and the answering of critical questions (Figure 1). All of these units are used in the development of the conceptual model for a site.

The first unit to consider is the ''distribution unit''. The distribution unit is the volume of soil that has been contaminated. A large site may consist of a number of distribution units (i.e. contaminated areas). Different processes may have been involved in the distribution unit for distributing the contaminant. The contaminant may have been ''uniformly'' distributed as one might expect in a settling pond. Conversely, a contaminant may be distributed irregularly if leaking drums were stored in the contaminated area. Knowing in advance how the contaminant may have been distributed at a site should influence the decision-making process for determining how and where samples must be collected to answer

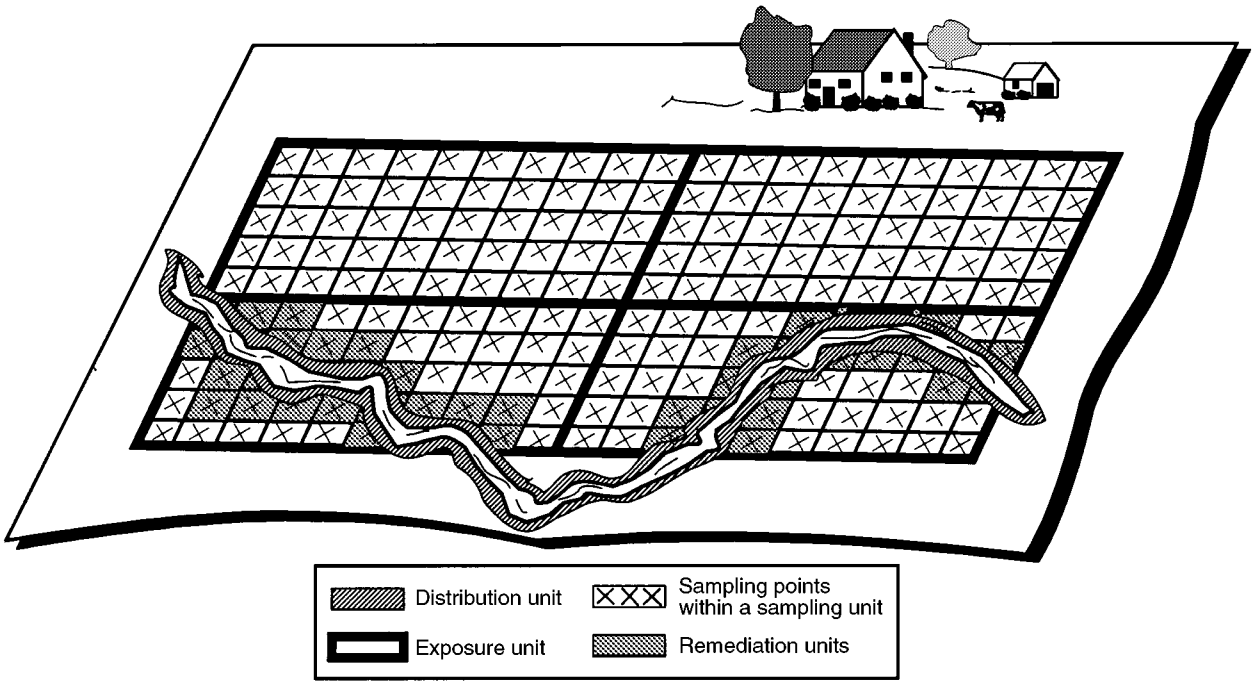

Figure 1 A site may be broken up into a variety of units to allow for critical questions to be addressed in the DQO process.

critical questions about the site. The sampling procedures and number of samples needed to characterize an area where leaking drums were stored can be expected to be different than for the bottom of a settling pond.

The second unit to consider before an investigation progresses significantly is the ''exposure unit''. The exposure unit is the volume of material, in this case soil, for which an action level or threshold concentration applies. Unfortunately, action levels are typically listed without the accompanying exposure unit. Action levels may be listed for humans for residential or industrial exposures, but usually the spatial area for which a person may be exposed to the action level (for some period of time and experience little risk) is seldomly defined precisely. Exposure units need to be defined as precisely as possible before extensive data collection begins if the exposure unit is going to serve as the foundation for the decision-making process. It may be necessary to spend some time to examine the basis and data for the action level to determine the type of exposure, underlying assumptions, and area for which actual exposure and risk of contact with the contaminant is of concern. For example, ''Does the 'residential exposure unit' assume 0.6 ha (1/4 acre) lots for which an individual may be exposed over a lifetime, on average, to unacceptable risks from the contaminant?''. Alternatively, ''are hot spots within the exposure unit of greater concern because the exposure and dose to children are expected to be higher?''. These are potential questions that must be considered when defining the exposure unit. The establishment of the exposure unit, together with the action level for a contaminant, is important in the development of the conceptual model and the associated decision-making process in a site investigation.

The third unit, the ''remediation unit'' is related to the ''exposure unit'' through the decision-making process and conceptual models in a site investigation. If the measured and estimated contamination within an exposure unit exceeds an action level, then presumably remediation needs to occur. A remediation unit may then be defined as the smallest practical volume of soil at which a cost-effective remediation can occur. The issue of the smallest practical volume is important since not only is it defined by the contaminant concentrations present but it is also operationally defined by the sampling and remediation equipment to be used during site restoration. For example, if sampling were done with tablespoons, and estimates of exposure units show that remediation is required, backhoes, rather than tablespoons, would more likely be used to remediate the area. Although samples were collected with a small support and decisions were based on those data, decisions would likely be made and based on a larger support. If only a small fraction of the exposure unit were contaminated (i.e. a hot spot), would it be necessary to remediate the entire volume of the exposure unit, or can smaller areas be remediated? It may be possible to make decisions that are less costly on the basis of remediation units, e.g. backhoe volumes, rather than entire exposure units.

Addressing the volume of the ''remediation unit'' at the beginning stages of a site investigation allows for development of decision rules (to be discussed) that can guide the decision-making process and the subsequent selection of sampling procedures, analytical methods, and concentration estimation methods.

The problem of ''hot spot'' identification and the required, optimal sampling network becomes easier to address if the remediation units have been defined at an early stage. If ''hot spots'' were identified through the use of tablespoon samples at a particular grid spacing, it would not be practical to define precisely the boundaries of the hot spot at a relatively small spatial scale if a backhoe were used to remove contaminated areas. High-resolution sampling for hot spots may not be cost-effective if the area to be effectively and economically remediated is on a larger scale. Definition of the ''remediation unit'' prior to the collection of a significant number of samples from the field permits the costs in obtaining data at various spatial densities (to address a decision rule) to be balanced against the costs in remediating the area with no further information or data being required.

2.2 Data Quality Objectives

All three previously defined units, the distribution, exposure, and remediation units, may be used in the DQO process to achieve an optimization of the site characterization process. The DQO process fosters communications and the development of critical decision rules for the study of soils and hazardous waste sites (Figure 2). A stepwise, iterative process, involving major participants in the decision-making process for a site, is used to focus attention and resources on the critical questions or hypotheses that need to be tested. A variety of questions may be answered in an investigation but there are usually just a few contaminants and questions that drive the decision-making process for a site investigation. An example of a decision rule would be, ''If the mean concentration of arsenic exceeds the action level within the exposure unit by 10%, then the unit should be remediated by progressively removing those remediation units that contribute to the exceedance''.

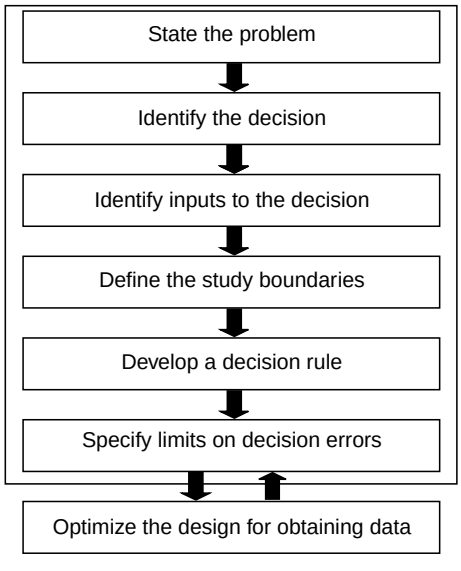

Figure 2 The DQO process.

The DQO process has been used in the past to establish objectives for precision, accuracy, representativeness, completeness, and comparability for the reported data. The major focus of QA/QC efforts has been on the analytical phases of the sample collection and measurement process. Current DQO guidance from the United States Environmental Protection Agency (USEPA) attempts to focus more efforts towards examining the sensitivity of the decision-making process to all key inputs, rather than just those related to sample analysis. In other words, ''What are the effects on the basic decision-making process (i.e. the decision rule) from imprecise, inaccurate, nonrepresentative, incomplete, and incomparable data from a variety of sources?''. This change in emphasis in the DQO process recognizes that there are numerous, and often significant, sources of variability outside the analytical process. There is variability in the process that is used to estimate contaminant concentrations and other physical parameters between measured points. There are errors introduced in the sample collection, subsampling, sample preparation, and sample transportation steps. All of these steps can introduce errors that are significantly greater than those encountered in the analytical process. A well-executed DQO process considers, in parallel, the tolerance for error in the basic decision-making process as well as the tolerance for error in the sample collection and measurement processes.

Further information on the DQO process may be found in the *Guidance for the DQO Process*[3] and *The DQO Process for Superfund: Interim Final Guidance*[4] documents. A related document, *Guidance for Data Quality Assessment*,[5] describes statistical methods for evaluating data to determine whether the DQOs have been met.

3 QUALITY ASSURANCE/QUALITY CONTROL

The overall goal of the QA/QC program in any investigation is to assess and ensure that the variability in the data is attributable, as much as possible, to the media and environment that is being measured and not to the measurement (including sampling) process. A properly designed and implemented QA/QC program can

achieve that goal. In the field, QA/QC samples may be placed in the sample collection and preparation processing stream at several different locations. Depending upon their placement, a variety of different errors can be assessed. However, the determination of the number, type, and placement of QA/QC samples within the sample stream is as problematic for a site investigation as the determination of the number, type, and location of routine samples that must be collected. All too often the number and type of samples to be used in both situations are determined by the projected budget for the site investigation and by estimated analytical costs and not by a judicious consideration of the acceptable error rates in making basic decisions (i.e. false positives and false negatives).

A rationale has been developed for how to determine the number, type, and placement of QA/QC samples.[6] The decision-making process for determining how extensive a QA/QC program needs to be for a particular investigation is not simple. A stepwise process needs to be developed in which data from the QA/QC effort are progressively examined and compared with the sample data and to DQOs. Basic concepts need to be understood and applied to ensure that the data collected meet the needs of the decision maker.

Data that are collected as part of an environmental investigation are going to exhibit variability. The expectation is that variability associated with the measurement process is sufficiently low that variability in the data can be attributed to the measured parameters or population. QA/QC programs are implemented to assess and control measurement variability, but these efforts are frequently insufficient or unable to assess or control all of the sources of bias and variability throughout the measurement process. Often, the QA/QC programs are targeted towards the analytical phase of the measurement process, but considerable bias and variability may be present in other areas.

As reported in the ''Rationale'' document,[6] Equations (1) and (2) are valid:

$$\sigma_t^2 = \sigma_m^2 + \sigma_p^2 \quad (1)$$

where σ_t^2 = total variability, σ_m^2 = measurement variability, and σ_p^2 = population variability, and

$$\sigma_m^2 = \sigma_s^2 + \sigma_h^2 + \sigma_{ss}^2 + \sigma_a^2 + \sigma_b^2 \quad (2)$$

where σ_s^2 = sampling variability (standard deviation (SD)), σ_h^2 = handling, transportation and preparation variability, σ_{ss}^2 = preparation variability (subsampling variability), σ_a^2 = laboratory analytical variability, and σ_b^2 = between batch variability (note: it is assumed that the data are normally distributed or that a normalizing data transformation has been performed).

Bias can change over time and become intertwined with imprecision – both of these contribute to variability in the measurement process. A variety of QA/QC samples may be used in a rigorous manner to try to assess the magnitude and source of errors in the sampling of soils at hazardous waste sites. Some of the errors may never be assessed because the true distribution and concentration of contaminants in the environment can never be truly measured; they can only be estimated. The best approach is to try to minimize bias and imprecision at the outset by carefully considering and selecting the sample collection, sample preparation, subsampling, and analytical techniques that are apt to produce ''representative'' data with low measurement error.

Some of the most common QA/QC samples used during the collection of soils include field evaluation samples, field blanks, field duplicates, and preparation splits. Field evaluation samples (e.g. performance evaluation materials or site-specific soil QA materials) are soil samples of known concentration that are subjected to the same manipulations as routine samples. The field evaluation samples should be introduced in the field at the earliest stage possible. These samples can then be used to estimate total measurement error and, if they are used in duplicate, precision estimates can be made.

Field blanks provide a measure of various cross-contamination sources, decontamination efficiency, and other potential sources of error introduced from sources other than the sample. Three common types of blanks used in the field include the field rinsate blank, preparation rinsate blanks, and trip blanks. Rinsate blanks (i.e. equipment blanks) are used to assess the efficiency of the decontamination process. These blanks are obtained by pouring deionized/distilled water over the sampling equipment (for the field rinsate blank) or preparation equipment (for the preparation rinsate blank) after it has been decontaminated and then collecting the rinsate for analysis. Trip blanks are generally used during the collection of VOCs. A trip blank consists of the sample containers filled with deionized/distilled water, solvent (e.g. methanol), or evacuated vials used during headspace analyses. For some sampling efforts (depending on the DQOs), a portion of the trip blanks will be opened in the field (excluding evacuated vials) to assess if there is any contamination that may have occurred during sampling, preparation, or shipment to the analytical laboratory. The remainder of the trip blanks will not be opened in the field and can be used to assess contamination due to incomplete sealing of volatile organic analysis (VOA) vials and sampling jars or improperly cleaned glassware.

The field duplicate sample is an additional sample taken near a routine field sample and may used to determine total within-batch measurement variability although differences in the two samples may also be attributable to short-range spatial variability in the soil. If the data from field duplicates are significantly different from the corresponding routine sample, then there is good cause to question the representativeness of the routine sample data and to employ data from other QA/QC samples to determine whether the problem is with short-range spatial variability. If the problem seems to be associated with short-range variability, then a larger support, or compositing of the samples on a small, spatial scale, may be all that is necessary to obtain more representative samples.

Preparation split samples are soil samples that are collected after the soil has been properly homogenized and, thus, all portions of the sample have theoretically equal concentrations. The preparation split samples are used to estimate the sum of subsampling, analytical, and data-handling error variations (i.e. total within-batch error minus the error that occurs during sampling, handling, transportation, and preparation/homogenization).

4 SAMPLING NETWORK DESIGN

A successful sampling design accomplishes the objectives of the investigation at minimal cost. The DQO process is one approach that provides a sound framework for sampling network design. The DQO process accomplishes three critical tasks. First, the DQO process develops a clear, quantitative statement of the decision that must be made. This is generally in the form of an if–then–else decision rule: if the sample measurements exhibit characteristic x, then take action a; else, take action b. Second, by specifying tolerable error limits on the decision, the DQO process allows for the fact that the true concentration in volume v is never precisely known. Finally, the design-performance diagram can be used to evaluate the expected performance of alternative sampling design options and to select from among the acceptable designs the one with the lowest associated cost.

In the DQO approach, error limits are established across the entire range of possible contaminant concentrations, and can be plotted against true concentration to form a ''design-performance'' diagram (Figure 3). The acceptable range of performance is the area including

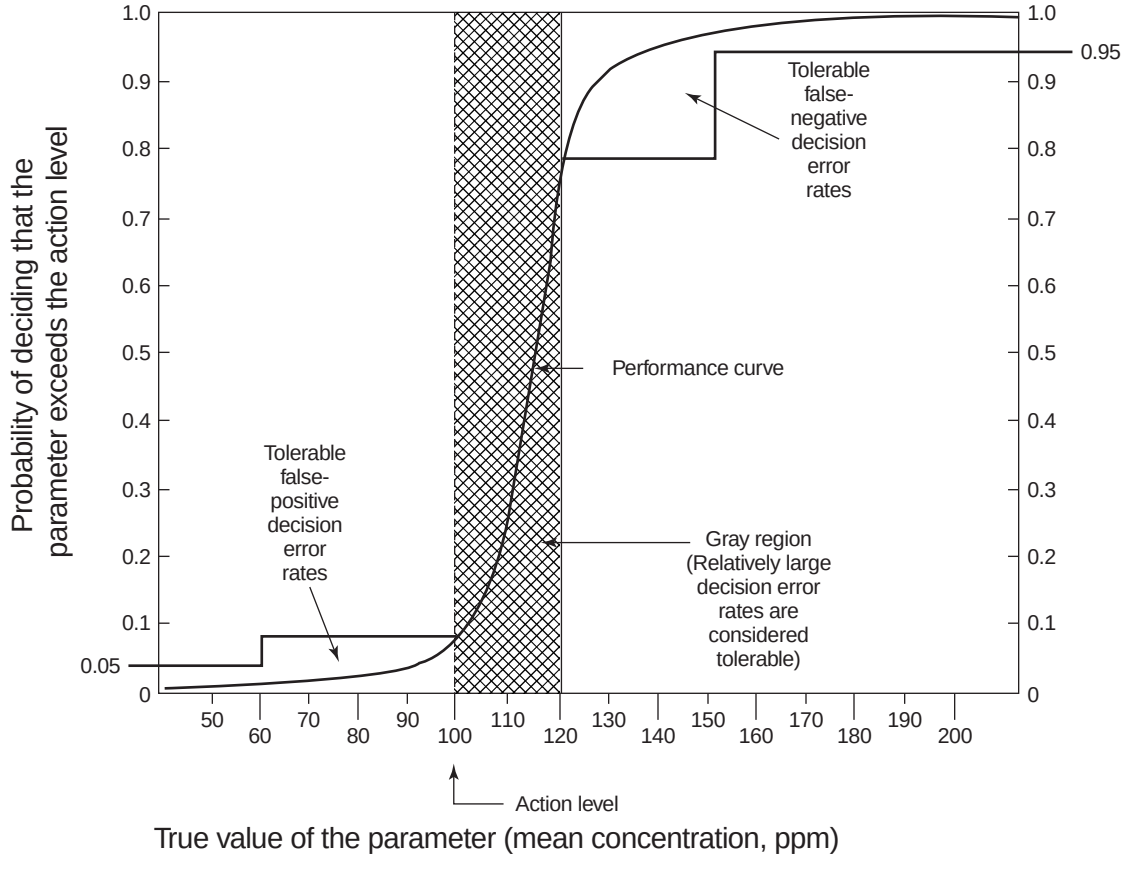

Figure 3 A typical design-performance curve used to evaluate the expected performance of alternative sampling design options and to select from among the acceptable designs the one with the lowest associated cost.

the gray region and the tolerable false-positive and false-negative regions. Each sampling scheme has an associated performance curve that shows the probability of making a correct or incorrect decision at any given true concentration. The performance curve (also known as a power curve) shown as the solid line in Figure 3 represents a sampling scheme that has acceptable performance because it falls within the specified limits. This particular curve also represents a minimum cost design because it meets but does not exceed the specifications. A sampling scheme with fewer samples or less accurate measurement would have higher error rates and be unacceptable, while a scheme with more samples or higher quality measurements would be acceptable, but unnecessarily expensive.

It is not easy to develop design-performance criteria. Determining acceptable decision error rates is a subjective process that is difficult and uncomfortable for many people to perform. There is a tendency to initially set the error rates so low that they lead to unacceptably high sampling costs, in which case the process must be repeated to obtain a more reasonable sampling design with lower overall associated costs.

4.1 Basic Spatial Sampling Strategies

To illustrate some of the basic spatial sampling strategies, two common decision scenarios in soil remediation will be examined. Additionally, alternative sampling strategies under the same decision scenarios will be discussed.

The first decision scenario involves the estimation of the mean concentration of contaminant in a specified volume of soil. Assume that a single distribution unit, or stratum, has been identified, such as the sediment in an abandoned waste lagoon, and that the decision maker will have to remediate the sediments if the mean concentration exceeds a specified threshold. From a design-performance diagram it has been determined that a sampling design will be acceptable if it the standard error (*se*) of the mean is ≤ 4, or equivalently, a variance (V_e) ≤ 16 (note: $se^2 = V_e$). Further, assume that the measurement error variance (V_m) for collecting and analyzing a soil core is 10 and the population variance (V_p) is 90 (i.e. 10% of the total variance is due to measurement error including sample collection, preparation and analysis while 90% is due to natural variability at the site). The total variance (V_t), including measurement error, is given by Equation (3):

$$V_t = V_p + V_m = 100 \tag{3}$$

For this scenario, a soil core costs \$25 to collect in the field and \$300 to analyze in the laboratory. The question that needs to be answered is, ''What is the best sampling design?''.

The classical approach to estimating a population mean is by the arithmetic mean of a random sample. Random sampling is always a valid method, even though it may not always be the most efficient for spatial sampling. In the random sampling approach, it is simple to calculate the standard error of the mean and evaluate the costs of alternate sampling schemes.

For a random sample of N measurements, the variance of the error of the mean (V_e) is simply equal to V_t/N. The design process now involves nothing more than finding the smallest value of N that will meet our example design requirement, namely, $V_e \leq 16$. In this case, $N = 7$, making $V_e = 100/7 = 14.3$, which is <16. The final cost of the random sampling design is $7 \times \$325 = \$2275$.

The bulk of the cost in the case of this random sampling design is due to laboratory analysis, so let us next consider a less expensive alternative. Assume that an alternative laboratory method is available at a cost per analysis of \$100, but that this method is less precise, with $V_m = 40$ instead of 10. If the initial study objectives were to obtain high-quality measurements of individual soil cores, this approach would not be useful. By taking four separate measurements of a single core sample, V_m could be reduced to $V_m/4 = 40/4 = 10$ and equal the precision/quality of a single measurement by the original method. However, the cost of the replicated analyses would be $4 \times \$100 + \$25 = \$425$ per sample, which is more expensive than with the original method selected. However, the original decision rule has nothing to do with measuring individual cores, only with estimating the mean of the sediments in the waste lagoon. For this purpose, $V_t = V_p + V_m = 90 + 40 = 130$. To meet the required objective, nine measurements would be needed so that $V_e = 130/9 = 14.4 < 16$. The cost of this design alternative becomes $9 \times \$125 = \$1125$, a substantial financial improvement from the \$2275 cost of the initial random sampling design with the more precise (but more expensive) analytical method.

Suppose that with the initial measurement method, where $V_m = 10$, the population variance is much higher than the measurement variance. In such cases, composite sampling can be a cost-effective alternative approach. In composite sampling, individual core samples are physically combined to make a single unique sample from which one or more measurements of the sample can be made. The overall estimation error is $V_e = V_p/N_c + V_m/N_a$, where N_c is the number of cores or increments that make up a composite sample and N_a is the number of analyses performed on a single composite sample. There will also be an additional error component and an added cost introduced by the process of homogenizing and subsampling the composite, but for this example it is assumed that this added error and cost

are negligible. The problems associated with subsampling a composite sample are far from trivial and will be discussed in detail in section 7.1.

The composite sampling design process becomes one of evaluating combinations of N_c and N_a and selecting the lowest-cost combination that meets the initial design requirements. In this case, the best result is obtained when $N_c = 15$ and $N_a = 1$ (i.e. a single analysis of a 15-core composite sample). The error associated with the alternative approach would be $V_e = 90/15 + 10/1 = 6 + 10 = 16$, and the associated cost is $(N_a \times \$300) + (N_c \times \$25) = \$675$. This achieves the original objectives at a lower cost than with individual samples.

From these examples, it has been shown that an improvement on the original design can be achieved by using a lower-cost analytical method or by compositing. Can the project planners do even better by combining both methods and analyzing composite samples by the cheaper method? Interestingly, in this particular example, the answer turns out to be no. With a single composite sample, measurement error (V_m) is again a significant factor, and the cost of reducing V_m through multiple analyses becomes cost prohibitive. The best option to meet the original goal of $V_e \leq 16$ is with $N_c = 15$ and $N_a = 4$ (where $V_e = V_p/N_c + V_m/N_a = 90/15 + 40/4 = 6 + 10 = 16$), for a total cost of \$775 [equals $(15 \times \$25) + (4 \times \$100)$]. Although this approach costs slightly more than the single, more precise analysis (\$400 vs \$300), the four replicate analyses on the composite sample provide at least some measure of QA in terms of precision (laboratory analytical repeatability).

The example above was intended to illustrate the type of economic analysis of sampling alternatives that is part of the DQO process, not to provide a blanket endorsement of composite sampling or cheap but imprecise analytical methods. With different costs, variances, and objectives the conclusions could be completely different.

4.2 Estimating Variances

While the analysis described above is relatively simple, the results are only strictly true if *all* of the assumptions are correct. The major assumptions were:

- the variance of the population distribution and the measurement error distribution for each measurement alternative are known;
- the true costs of sample collection and measurement are known; and
- all measurements are independent of each other.

In practice, these assumptions are at best only approximately true, so it is worth looking at how they affect the analysis of sampling alternatives. The relative magnitude of the variance of the population versus the variances of the alternative measurement methods and the relative costs of the alternatives are all important in determining the overall sampling strategy. Generally, the relative costs of measurement methods can be estimated reasonably accurately. The more difficult problem is estimating the error variances. The true variance of a contaminant in soil at a site or within a stratum will always remain unknown. The best way to estimate it is through the analysis of sample data, but this poses a dilemma. It is more difficult to estimate variances precisely than it is to estimate means. It is likely that by the time one has collected enough samples to obtain a satisfactory estimate of variance, there will be far more data than are needed to estimate the mean efficiently. The situation becomes worse if composite sampling or another alternative measurement method were evaluated since an estimate of the measurement variance(s), in addition to the total variance, is needed. Analytical variances are often well known, but measurement variance includes many sampling and subsampling error components that are site-specific. Estimating measurement variance from sample data requires the analyses of multiple field replicates. Estimating these variances can easily cost many times more than simply estimating the mean.

The alternative to estimating variances from data is to estimate them through educated guesswork (i.e. best professional judgement). Usually, at least a few measurements will be available from previous investigations which can in turn be used to estimate the measurement variance. Analogies can be made with data sets from other similar sites. Measurement variance can likewise be estimated from QA data collected at similar sites, or it can be estimated by combining the analytical variance from laboratory studies with estimates of preparation and subsampling variances from particulate sampling theory (as described later).

The question of which approach to take to estimate the variances should be addressed as part of the DQO process. In addition to determining the decision to be made and how accurately it must be made, it is also necessary to know how accurately the decision maker needs to know how accurate the decision is. Unfortunately, there is no standard formula for determining these requirements. In the example above, if the single-composite, single-analysis option is chosen, there will be no way to evaluate whether or not the desired precision is achieved. The driving factor will be the magnitude of the consequences of decision errors. Suppose in the above example that the variances were preliminary estimates based on prior investigations of similar sites. Suppose, also, that the decision involved is relatively minor, for example, whether to leave slightly contaminated soil in place or remove it to a municipal land fill at a cost of \$5000–\$10 000. In this situation, the

analysis described above might well be sufficient, and one of the compositing schemes discussed would be chosen. However, if the decision were between disposal in the municipal landfill versus in a hazardous waste facility at a typical cost of $100 000, it is reasonable to spend more money on sampling in order to be more confident about the quality of the decision. Again, there is no unique simple strategy for such cases; nonetheless, some form of sequential sampling may be appropriate.

In sequential sampling, an initial set of data is collected and evaluated. The decision to be made is now threefold: yes, remediation is required; no, the soil is ''clean'' and can be left in situ; or collect more data to help define the situation better. The initial data are used to estimate the mean, variance, and standard error, and if the standard error is low enough, the decision is made. If the standard error is too high, the variance is used to estimate how many additional samples are needed in the second sampling phase. The process is repeated until a confident yes or no decision can be made. The number of samples collected in the initial set is determined by economics and logistics. If measurements are being made in the field, it may be practical to re-evaluate the data after each measurement. This approach minimizes the total number of samples required. However, if the samples must be sent to a laboratory for measurement, there are time delays and mobilization costs each time a sampling crew returns to the field. This may make it advantageous to make a conservative initial estimate of variance and attempt to over-sample the initial phase to avoid further sampling.

Sequential sampling can be used with either individual or composite samples. Unfortunately, economic analysis of the alternatives is very difficult, especially when the population is not normally distributed. Real-world distributions of contaminated soil measurements are often highly skewed and contain non-detects (i.e. data below the instrument or method detection limit (DL)). In such cases, the only practical method for comparing the cost-effectiveness of alternative sequential designs is through elaborate computer simulations of the sampling process. Composite samples may well have an advantage in that the composite samples will be more normally distributed than the original population. The occurrence of a more normally distributed population makes it more likely that normality-based tests, such as Student's *t*-test, can be used with a small number of measurements.

4.3 Spatial Dependence

In the examples above, it was assumed the data were independent. However, the fate and transport of contaminant in soil are determined by physical and chemical processes that do not operate at random. This is also true of natural soil parameters, such as clay content and porosity, that can influence the spatial distribution of contaminant. In general, it is expected that soil measurements taken a few centimeters apart would be more similar in concentration than when the measurements are taken kilometers apart. Quantifying such spatial dependence and how it affects sampling and estimation is the subject matter of geostatistics. The reader is referred to Isaaks and Srivastava[7] for a practical introduction to geostatistics. Additionally, Webster and Oliver[8] provide an excellent discussion of the specifics involved in spatial dependence during soil sampling applications. Only a brief discussion of the basic concepts and a few rules of thumb for sampling and estimation will be provided in the following text.

Spatial dependence can be quantified in the form of a semi-variogram (Figure 4), which shows the variance of measurements as a function of the spatial vector (distance and direction) separating the points. More precisely, the semi-variogram shows the expected value of $[z(x) - z(x+h)]^2/2$, where z is the analyte concentration at points x and $x+h$, and sampling location $x+h$ is separated from sampling location x by the vector h. The semi-variogram incorporates all of the components of the total variance. The y-intercept at $h = 0$ includes all of the independent measurement variance components, such as field and laboratory preparation, subsampling, and analytical error (AE). The semi-variogram at very small h values (i.e. very small distances from the original sampling location) would reflect the variance of collocated samples. The shape of the semi-variogram curve defines the spatial dependence structure. The mean value of the semi-variogram over all possible sample pairs at all possible separation vectors is simply the total population variance. The semi-variogram provides a very useful model of both

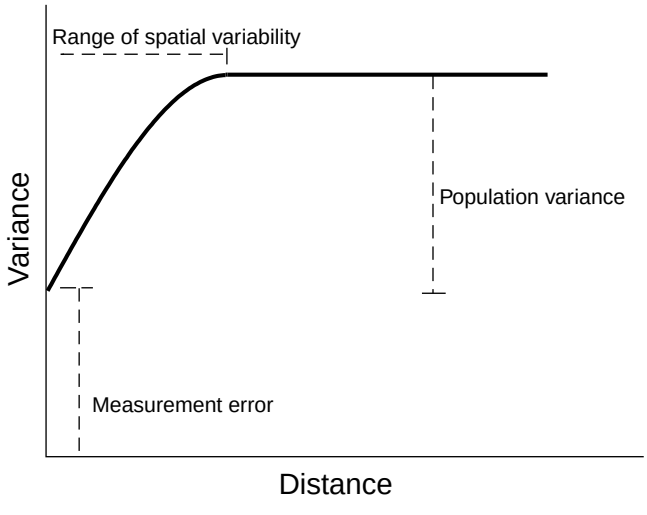

Figure 4 A typical semi-variogram showing the variance of measurements as a function of the spatial vector (distance and direction) separating the data points.

independent and dependent variance components, but like any of the components, creating a semi-variogram entirely from data can be very expensive.

When data are spatially dependent, the arithmetic mean is no longer the best (minimum error variance) estimate of the population mean. Instead, a weighted mean should be used. In geostatistics, a weighting method called kriging is frequently used, in which the semi-variogram is used to compute sample weights that minimize the estimation variance. When two or more samples are clustered very closely together (i.e. the sampling distance between two points is small), whether by chance or by design, they are effectively providing duplicate information. This duplication is indicated by low semi-variogram values among neighboring samples and smaller weights are assigned to such samples. Conversely, samples with no near neighbors receive higher weights. By down-weighting clustered data that are duplicative, kriging provides better estimates than simple averaging.

The most efficient sampling design, when spatial dependence exists, is one that minimizes the duplication of information or, equivalently, maximizes the distance from each sample to its nearest neighbor. The solution to this problem is simple: use a systematic sampling scheme and sample at regular grid nodes. Webster and Oliver[8] have shown that grid sampling can be significantly better than random sampling in some cases. They also reported that the arithmetic mean of a regular grid is essentially equivalent to the kriged mean for the same data set. This suggests that regular grids are also optimal for composite sampling where the physical mixing of the soil cores is equivalent to arithmetic averaging. Triangular grids are the most efficient (in two dimensions) but square grids are very nearly as good and are generally more convenient in the field. Since it is usually safe to assume the existence of some form of spatial dependence in soil contamination, regular grid sampling is recommended, preferably with a random origin and orientation.

4.4 Sampling for Local Estimation

The purpose of local estimation is usually to classify a site into higher-concentration areas that need some form of remedial action and lower-concentration areas that need none. When it is assumed that the site contains high- and low-concentration areas, it is the same as saying that there is spatial dependence (i.e. similar values tend to be grouped together). This spatial dependence becomes a significant factor in sampling network design. Similar to global estimation, sampling on regular grids is more efficient for local estimation.

Local estimation and decision making can become very complex. It is critical to establish a very clear decision rule at the beginning that must include the scale of the decision. The simplest approach to local estimation is to subdivide the site into an array of discrete units. These discrete units may be as large as exposure units or as small as remediation units. Generally, high remediation costs lead to the consideration of smaller decision units in an attempt to be more selective in which units must undergo the expense of being remediated. Each unit will be classified based on an estimate of its mean concentration. Subdividing a site into smaller remediation units means that many decisions will be made instead of just one and that the consequences of error for any given decision become much smaller. The DQO process should reflect this by establishing broader error tolerances for smaller units.

Composite sampling remains an attractive option for local estimation because compositing increases the area represented by each sample and reduces the sample variance. Compositing is particularly effective when individual core concentrations vary greatly over distances that are small compared with the scale of decision, and when this short-range variability is high relative to measurement error. Generally, the larger the area over which a composite is taken, the greater is the reduction in sample variability; however, the composite area should not be larger than the decision unit.

Local estimation differs significantly from global estimation because the local unit being estimated is a part of the larger population. As a result of spatial dependence, each unit is correlated to some extent with its neighboring units. It becomes possible to estimate the concentration of a unit by interpolation from nearby data, even when there are no samples in the unit itself. In this case the design problem is no longer one of adequately sampling each unit, but rather one of sampling the population area in order to adequately estimate each unit.

There are two basic approaches to optimizing sampling designs for local estimation. Both require an estimate of the semi-variogram. The first approach is based on the fact that when estimates of remediation units are made by kriging, estimates of the corresponding standard errors are also produced. The basic assumptions involved in kriging result in the standard error estimates being determined by only the semi-variogram and the sample locations, not by the data values, so it becomes possible to calculate kriging standard errors for hypothetical sampling network designs. The performance of alternate designs can be compared with the established decision error limits. The approach is detailed by McBratney et al.[9] and McBratney and Webster.[10] The major problem with this approach is that kriging estimates of kriging standard errors are sensitive to errors in estimating the semi-variogram and to errors in the basic assumptions.

A promising alternative approach to the problem of developing complex designs for local estimation was

described by Englund and Heravi.[11,12] This approach involves first creating a detailed computer model of the soil contamination distribution that is consistent with both the semi-variogram and with any available data using a geostatistical technique called ''conditional simulation''. The resulting model contains a very dense grid of data points (typically, 10 000 or more points.). A few of these grid points will contain actual measured data values while the remainder of the grid points will contain simulated measurements that are both reasonable and realistic in detail. This approach is less sensitive to errors in estimating the semi-variogram because the model is forced to fit the actual data values. After the detailed model has been constructed, it is possible to simulate the entire process of sampling, estimation, and decision making for any particular sampling design and to compare the cost-effectiveness of alternative designs. The basic procedure for evaluating a sampling design by conditional simulation is as follows:

1. Overlay a grid of decision units over the modeled area. Each unit will contain many simulated sample values.
2. Calculate the mean concentration of all the simulated values in each unit. These are the true concentrations of the units. Comparing these means to a specified action level determines what the correct remediation decision for each unit should be.
3. Select a sampling design scheme, such as 50 samples at randomly selected locations over the model area.
4. Determine a set of sample locations and draw the simulated sample concentration values from the model.
5. Estimate the mean concentration of each decision unit, apply the decision rule, and determine whether the decision is correct by comparison with the true concentration from step 2.
6. Repeat steps 4 and 5 at least 100 times with different sample sets generated by the same sampling scheme and keep track of all decisions obtained.
7. On a design-performance diagram, plot the proportion of correct decisions versus the concentration for each of the decision units. If all of the points fall within the design-performance limits, the sampling scheme is deemed acceptable.

The conditional simulation procedure is complex and clearly best suited to large, expensive remediation projects. Its primary advantages include the flexibility to evaluate alternate spatial designs, such as regular grids and composites, to tabulate detailed performance statistics, such as costs of sampling and remediation, and to quantify the amount of contaminant that would remain unremediated following the selected sampling design.

This simulation approach can also be used to evaluate the cost-effectiveness of the overall sampling/remediation strategy. In a design-performance diagram like that in Figure 3, the right-hand boundary of the ''acceptable performance'' zone limits the probability that highly contaminated soil will remain unremediated, while the other boundary limits the probability that relative uncontaminated soil will be unnecessarily remediated. The former is of primary interest to the regulator because it is protects against risk to humans and the environment. The latter boundary, however, is primarily an economic choice on the part of the party responsible for remediation. The trade-off is between sampling and remediation costs. In practice, the actual decision level will occur at estimated concentrations near the center of the gray region. Moving the left-hand boundary further to the left lowers the effective decision level and increases the amount of remediation to compensate for the greater uncertainty due to less sampling. Moving it to the right raises the effective decision level and increases the amount of sampling required. Simulation can be used to evaluate these trade-offs and find the most cost-effective solution that meets the regulatory requirements.

5 DOCUMENTATION

Accurate documentation is essential for the success of a field sampling program. Documentation should occur in all phases of a soil sampling program, including planning, sample collection, and laboratory analysis. Three documents usually required are the field sampling plan (FSP), quality assurance project plan (QAPP), and health and safety plan (HASP). The FSP provides guidance for all fieldwork by defining in detail the sampling and data-gathering methods to be used on a project.[13] Topics that should be discussed in the FSP are the site background, sampling objectives, sample location and frequency by matrix (including QA/QC samples), sample designation, sampling equipment and procedures with standard operating procedures (SOPs), if available, and sample handling and analysis. The QAPP describes the policy, organization, functional activities, and QA/QC procedures necessary to achieve the DQOs defined for the project. The detail to be included in QAPPs depends upon the category of investigation being undertaken. For example, investigations that have compliance or litigational implications require discussion of the following 14 points: project description; project organization and responsibilities; QA objectives; site selection and sampling procedures; sample custody; calibration procedures

and frequency; analytical procedures; data reduction, validation, and reporting; internal QC checks; performance and system audits; preventative maintenance; calculation of data quality indicators; corrective action; and QC reports to management.[14] If both a FSP and QAPP are required for the same project, to avoid duplicative efforts for sections that contain the same information, it is recommended simply to reference the overlapping section in the one document from the other document. The HASP is prepared as a support document to protect the health and safety of the soil sampler and to meet United States Occupational Safety and Health Administration (OSHA) guidelines and regulations.[13] Topics to be discussed in a HASP include identification of key health and safety personnel (e.g. site health and safety officer); hazard risk analysis by task and operation; employee training requirements; personal protective equipment (PPE); medical surveillance requirements; frequency and types of air monitoring; site control measures, decontamination procedures; SOPs for the site; emergency response procedures; and confined space entry should sampling be performed in a pit or hole.

During sampling and analysis, it is only through proper documentation that a soil sample can be linked with a sampling location (including depth) and the date and time of collection. Use of detailed field and laboratory logbooks to document routine information (e.g. sampling date, weather conditions, sample numbers, QA/QC sample use, air monitoring data), problems encountered during sampling (e.g. inability to penetrate to desired depth), sampling location changes, site characterization information (to be discussed), soil characteristics, production of the site map (to be discussed), or analytical procedural changes. The use of standardized forms greatly enhances field documentation by ensuring the collection of complete information required at the site. Other required documentation associated with sample identification and shipment include sample labels, chain-of-custody forms, custody seals, and shipping airbills.[15]

6 SITE CHARACTERIZATION

Whenever a soil sampler, sampling team, or management team approaches a new site that needs to be evaluated, one of the initial steps performed should be site characterization (i.e. site description). A good site characterization program consists of two major steps, namely site observation and subsequent site mapping. Site characterization may be limited to a reconnaissance survey or involve a full-scale (or definitive) survey. A reconnaissance survey is performed to confirm historical data through the collection of a few selected soil samples for analysis or to identify site features that may indicate contaminant sources, pathways, affected populations, and potential monitoring/sampling locations. A full-scale survey is undertaken to obtain soil samples for analysis with their results being used in the final decision-making process.[16]

For both survey types, a basic suite of characteristics should be noted and recorded during the site visit. This suite includes information and observations on climate and weather, slope, surface erosion and erodibility, surface runoff potential, vegetation, and the presence of macro- and mesofauna. Although these parameters are readily observed, they should be recorded because they may provide explanations for observed trends in the data and possible considerations for long-term planning and remediation efforts at the site.

In addition to the site characteristics, numerous soil characteristics should be examined and recorded in the field since they provide information on potential subsurface transport pathways, variability of soil properties, location of contaminants, and identify features that may influence remedial operations. More complete descriptions and information on site characteristics may be found in Cameron[17] and Boulding.[18]

6.1 Site Observations

Climate and weather are site characteristics that often influence the collection of soil samples. Climate, the average condition of the weather over a period of years, influences the time of year that sampling and site remediation efforts can occur.[17] Clearly, soil sampling during the winter months in the northeastern United States may be restricted owing to cold weather and frozen ground conditions, yet in contrast, winter sampling in the southwestern United States may be preferable to avoid the extreme high temperatures commonly encountered during the summer months. Weather conditions, the state of the atmosphere during the field investigation, influence the specific time of sampling. Specific conditions to note include temperature, precipitation, wind speed and direction, and humidity. Although often noted, but not recorded, these conditions generally influence the health and safety of the sampler and have little direct influence on soil sample collection with the exception of precipitation. For example, when the wind is blowing, samplers should always position themselves upwind to avoid inhalation of dust and other particles that may contain the contaminants of concern. During times of high winds, sampling should be avoided to prevent the spread of the contaminants sorbed on fugitive dust particles leaving the waste site.

It is important to note the slope of the soil surface because it influences the rate and amount of runoff and erosion of water, soil, and associated contaminants.[17,19]

Slope features to be noted include gradient, length, shape, and topographic position. Simple terms such as flat, moderate, or steep can be used to describe the gradient. If greater accuracy is required, slope gradients may be recorded as a slope percentage (the change in elevation per horizontal distance between two points) or a slope angle (measured in degrees from horizontal). Slope length, usually measured in meters or feet, influences the behavior of water and the potential for erosion. In general, longer slopes have greater runoff and erosion potential than shorter slopes.[19] Slope shape (e.g. convex, concave, or flat) and topographic position (e.g. summit, shoulder, sideslope, or floodplain) influence the movement of water and soil on the surface and in the subsurface. For example, steep convex shoulder positions are more likely to experience erosion than flat or concave floodplains in which soil deposition is likely to occur.[20]

Site characterization of surface erosion is important to assess soil loss or deposition in the past and to assess future erosion potential. The determination of surface erosion is easily done visually by comparison of observed changes in soil texture and color between surface and subsurface horizons.[17] If subsurface colors or textures are noted at or near the surface, then that area has undergone and is prone to erosion. The presence of rills, gullies, or other erosional features also indicate that erosion is occurring at the site and mark potential pathways for movement of contaminated soils within and from the site. Valuable information on surface erosion and erodibility of soils at a site can be found in county soil surveys published by the United States Department of Agriculture (USDA) National Resource Conservation Service (NRCS), formerly the USDA Soil Conservation Service.

Surface runoff potential is used to evaluate the potential for transport of contaminants at the soil surface to surface streams or other water bodies.[17] The runoff potential is controlled by slope and saturated hydraulic conductivity. Saturated hydraulic conductivity may be defined as the rate of movement of water through a soil or as a measure of the ability of a soil to transmit water under saturated conditions.[21] Runoff is directly proportional to slope (i.e. as slope increases, surface runoff potential also increases) and inversely proportional to the saturated hydraulic conductivity (i.e. as the hydraulic conductivity increases, the runoff potential decreases). Similarly to surface runoff, valuable information on soil hydraulic conductivities and surface runoff may be found in soil surveys published by the USDA NRCS.

Vegetation serves as an indicator of site history, erosion potential, and contaminant location. During site characterization, the sampler should observe the nature (e.g. hardwood forest), kind (e.g. maple trees with mixed understory), extent (e.g. dense, scattered, sparse, or bare), and distribution of the site vegetation. The presence of stunted vegetation, luxuriant plant growth in comparison with surrounding areas, discolored leaves, and ''burn'' spots can all indicate the toxic effects of contaminants in the soil. In areas of heavy metal contamination, sampling of vegetation along with soil sampling may be desirable to assess potential metals exposure through bioaccumulation.[17]

Macro- and mesofauna, like vegetation, can provide an indication of the presence of contaminants. Additionally, during their routine daily activities, fauna create potential pathways to enhance vertical movement and leaching of contaminants to depth within a soil, mix the soil, and are capable of moving large quantities of subsurface materials to the surface. Macrofauna are those animals that can be measured in centimeters (or inches) such as burrowing rodents, earthworms, and insects.[17] In contrast, mesofauna are the smaller animals including other insects, arthropods, nematodes, and smaller worms. Simple terms, such as many, common, few, and none, can be quickly recorded to indicate the presence of indigenous fauna. A lack of any visible fauna may indicate an area of high contaminant concentrations.

6.2 Site Mapping

The second step in the site characterization process is the production of a site map. Site mapping generally occurs in two phases: initial and final mapping. The site map is important because it allows for planning sampling locations and allows for the subsequent plotting of the contaminant spatial distribution. The initial site map is a crude sketch of the sampling site. Features that should be drawn on the site map include both permanent and ''temporary'' objects. Permanent objects to be mapped include playgrounds, buildings, roads, railroads, boulders, and bodies of water (e.g. streams or rivers). Temporary objects include trees, fences, power poles, and telephone poles. Drainage ways (e.g. gullies or depressions) are sometimes plotted since they represent potential pathways of contaminant migration. Other features that may be included on the initial site map are bare spots, percentage vegetative cover, and unusual surface features such as soil discolorations and surface staining.

Once the first draft sketch of the site has been completed, the approximate locations of the sampling points can be plotted based on the sampling design. More exact locations of the sampling points can be made using surveying equipment or global positioning system (GPS) units depending upon the needs of the program. It will be from these marked points that the sampler(s) will collect the soil samples. If changes in sampling location need to be made owing to unmapped surface features

or unforeseen subsurface obstacles (e.g. large tree roots, buried rocks), the sampler should indicate the type of obstacle encountered, move approximately 30 cm (12 in) in any direction away from the obstacle, and try to re-collect the sample. This process may have to be repeated several times in different directions from the initial point or at greater distances until a satisfactory location is found. The new sampling point should be marked on the site map to indicate the sampling location change.

After sampling has been completed, the final location of each sampling point needs to be accurately determined. Exact locations of the final sampling points can be made through a variety of different techniques ranging from measuring the distance and direction from a fixed reference point (or points) with tape measures, through the use of field survey equipment and GPS units. Each sampling point should be reconfirmed as to its approximate spatial location in the field and on the site map. A more formal site map may then be produced using this collected information in conjunction with, or separate from, the analytical results.

7 SAMPLE COLLECTION

After the site has been adequately mapped and all important site observations have been noted, the next step is to collect the soil samples. It is often quoted that it is during sample collection that a majority (80% or more) of the total measurement error occurs and yet it is during this phase of a program that the least attention to detail is paid by the decision makers, project managers, chemists, etc. The following text provides guidance on how to collect a soil sample properly, taking into account all the potential pathways for error to enter the system, provides means to minimize these errors, provides information on common concerns during the selection of sampling equipment, and discusses the advantages and disadvantages of commonly available sampling tools.

7.1 Particulate Sampling Theory

A particulate sampling theory was developed by Pierre M. Gy for the mining industry to provide a more accurate and precise means for identifying heterogeneous ore grades. The theory is based upon the relationship that exists between the variability of the material, the particle sizes within the material, the distribution of the component of interest (e.g. pollutant), and the size of the sample collected. Particulate sampling theory is based on sampling ''correctness'', which in turn is based on a property intrinsic to the material itself and the equipment used to extract the sample.[22] Gy defines a correct sample as ''A sample in which all particles in a randomly chosen sampling unit have the same probability of being selected for inclusion in the sample''. The importance of this theory to environmental soil sampling is that the theory identifies various sources of error that can influence the final determined contaminant concentration and presents the means for controlling these errors. The following text is a very brief synopsis of Gy's particulate sampling theory as it pertains to soil sampling. A more complete and in-depth discussion of all aspects of the sampling theory may be found in Pitard.[22]

Two models are described in Gy's particulate sampling theory: a continuous model which deals with variables through time and space and a discrete model which addresses sampling populations of fragments. Since soil sampling is more closely related to sampling populations of discrete fragments, only those errors associated with the discrete model will be discussed. The errors associated with the discrete model include the fundamental error (FE), grouping and segregation, short-range heterogeneity, increment delimitation, increment extraction, preparation, and AEs.

7.2 Sources of Variation and Sampling Error

One of the key concepts of Gy's particulate sampling theory and its determination of the FE is that the relationship between the maximum particle size in a collected sample and the weight of the sample to be collected must be clearly defined. By following the equations (to be presented), the sampler can select a particle size range that will reduce the relative variance in the final results to an acceptable level (as defined in the DQO process) and ensure that enough soil is collected to present an unbiased sample to the laboratory for analysis.

The FE is associated with the natural variability inherent (i.e. the heterogeneity that is inherent to the composition of each fragment or particles making up the lot) in the composition of every particle making up the ''lot'' to be sampled.[22] A lot is defined by Gy as the batch, volume, or sampling unit from which increments and samples are collected. Lots may range in size from a bottle of soil at the laboratory, to a dump truck load of soil, through the entire hazardous waste site under investigation. The FE is the only error that cannot be eliminated; however, it can be reduced by comminution of the maximum particle size or by taking larger sample sizes in the lot. Mathematically, the FE is defined as in Equation (4):

$$s^2_{\mathrm{FE}} = \left(\frac{1}{M_{\mathrm{S}}} - \frac{1}{M_{\mathrm{L}}}\right) IH_{\mathrm{L}} \qquad (4)$$

where s^2_{FE} = relative variance of the FE, M_{S} = mass of the sample (g), M_{L} = mass of the lot (g), and IH_{L} = constant factor of constitution heterogeneity. When the mass of

the lot, M_L, is large in relation to the sample weight to be collected, M_S, the equation can be simplified to Equation (5):

$$s^2_{FE} = \frac{IH_L}{M_S} \tag{5}$$

IH_L can be estimated using Equation (6):

$$IH_L = fgcld^3 \tag{6}$$

where f = shape factor (dimensionless), g = particle-size (or granulometric) factor (dimensionless), c = mineralogical factor ($g\,cm^{-3}$), l = liberation factor (dimensionless), and d = diameter of the largest particle (cm).

The shape factor (f) addresses the error due to all fragments not being perfect cubes and thus not fitting perfectly through the square holes in sieves used to screen the soil.[22] The shape factor is influenced by the number of particles passing through two consecutive upper and lower sieves, average particle diameter, average particle density, and the mass of the fraction to be analyzed. Mathematically it is expressed by Equation (7):

$$f = \frac{M}{p} d^3 \lambda \tag{7}$$

where M = mass of fraction collected on the lower sieve (g), p = number of individual particles, d = average diameter of particles passing through the two sieves (cm), and λ = average density of collected fraction ($g\,cm^{-3}$).

For all practical purposes, a quick examination of the sample under a microscope is sufficient to determine the shape factor based on visual observation and comparison with the following approximate shape factor values. The shape factor used in calculating the FE can be approximated by using values of (a) 0.5 or slightly less for most minerals and soils, (b) 0.1 for flat minerals such as micas, and (c) ≥ 1 for acicular minerals, such as asbestos and tourmalines. Shape factors as large as 10 may be used in the presence of long, thin, needle-shaped minerals.

The particle size distribution factor (g) addresses the error due to all fragments not being exactly the same size and thus not being the coarsest fragments in the sample.[22] Mathematically, it can be expressed by Equation (8):

$$g = \sum \frac{M_{L\alpha}}{M_L} \tag{8}$$

where $M_{L\alpha}$ = mass of the size fraction (α) of concern (g) and M_L = mass of the lot (g).

Two general assumptions for g that can made in calculating the FE are that for most soils, g is approximately equal to 0.25 and that if the soil has been screened to a particular size fraction, then g can be approximated by a value of 0.55.

The mineralogical factor (c), also known as the chemical or mineralogical composition factor, accounts for the maximum heterogeneity within the lot when the mineral (or constituent of interest, e.g. a pollutant) is completely liberated from the other material (gangue). It results from differing particle densities and associated differing concentrations.[22] While the full mathematical equation accounts for the average concentration of the lot and the varying densities of the fraction of concern and the gangue or rest of the material, for most soils Equations (9) and (10) can be used:

$$c = \frac{\lambda_M}{a_L} \quad \text{if } a_L < 0.1 \tag{9}$$

$$c = (1 - a_L)\lambda_g \quad \text{if } a_L > 0.9 \tag{10}$$

where λ_M = density of pure mineral or constituent of concern ($g\,cm^{-3}$), a_L = critical content (proportion of the constituent of interest in the lot) of the lot (dimensionless), and λ_g = density of gangue or background materials ($g\,cm^{-3}$). A particle density of 2.6–2.65 $g\,cm^{-3}$ is often used to represent the density of the soil in these equations. The critical content (a_L) of the lot is usually an approximation based on historical data about the proportion (or concentration) of the contaminant of concern at the site.

The liberation factor (l) is a correction factor taking into account that c is a measure of the maximum possible heterogeneity where the contamination occurs as completely separated (liberated) discrete particles within the soil matrix.[22] Hence l, whose values lie between 0 and 1, adjusts for heterogeneity of the sample or lot. Mathematically, it is expressed by Equation (11):

$$l = \frac{a_{max} - a_L}{1 - a_L} \tag{11}$$

where a_{max} = maximum proportion of the pollutant associated with the largest contaminated particles in the lot and a_L = average proportion of the pollutant in the lot.

The liberation factor can be estimated depending upon the degree of visual heterogeneity within the sample using values of 0.8, 0.4, 0.2, 0.1, and 0.05 for matrices that appear to be very heterogeneous, heterogeneous, average, homogeneous, and very homogeneous, respectively. At most hazardous waste sites where soil contamination is the result of a spill, l values range between 0.05 and 0.2. In contrast, where discrete contaminant particles have been spread out at a site (e.g. ground Pb plates from used automotive batteries), then a liberation factor of between 0.8 and 1 may be appropriate.

The final and perhaps most influential factor used in the determination of the FE is the diameter (d) of the opening of a square mesh retaining no more than 5% of

the sample.[22] The largest particles strongly influence the FE because they have the greatest inherent heterogeneity among particles.

By knowing the factors that control the FE, the sampler can minimize, but never eliminate, the FE. However, the FE is not the only source of error or bias that must be considered during the collection of a soil sample. The sampler must be careful to collect a ''correct'' sample and thus avoid introducing the other biases into the final results. These biases are the result of the grouping and segregation, increment delimitation, increment extraction, preparation, and AEs. Fortunately, these errors can be minimized or eliminated through the careful collection of the soil sample.

The grouping and segregation error (GE) results from the distributional heterogeneity, the heterogeneity that is inherent to the manner in which separate and distinct particles (or aggregates) are scattered or spread out within the lot to be sampled.[22] The grouping factor reflects the number of increments (i.e. the group of particles extracted from the lot in a single operation of a sampling device) making up the sample compared to the number of fragments (i.e. particles) making that sample. The grouping factor approaches zero as the number of increments approaches the number of fragments. The segregation factor accounts for the segregation of different types of particles and the natural range between the minimum and maximum distributional heterogeneity within a sample. The segregation factor ranges between zero and one. The major factor affecting GE is gravity. GE can occur due to gravitational separation that results from differences in:

- particle density;
- size;
- shape (e.g. round particles tend to move or roll more easily in the sample or pile than flat particles);
- magnetic properties (e.g. magnetite particles will preferentially tend to stick to themselves or steel sampling utensils);
- electrostatic charge (e.g. charges on the sides of plastic bottles or bags can cause the adhesion of charged clay particles);
- moisture, which tends to clump soil particles together;
- air turbulence during sample splitting or homogenization, which can cause loss of the fines;
- vibration (e.g. during transport from the site, on a laboratory cart, or from a vibrospatula used to obtain subsamples) in which heavy particles will slowly settle to the bottom of the sample container.

To minimize the grouping error, the collection of as many small increments as is practically possible is recommended. This process of collecting numerous small increments increases the probability of collecting all types of ''groups'' for inclusion in the sample, and thus reduces the grouping error. To minimize the segregation error, homogenization of the lot before sampling is recommended, taking care that the homogenization process itself is free from the factors that increase GE.

The short-range heterogeneity error (CE_1) is a random, discontinuous error that is influenced by FE and GE ($CE_1 = FE + GE$) and thus can only be minimized.[22] CE_1 is what most soil samplers are referring to when they discuss the ''natural variability'' of a soil through time or space and what geostatisticians call the ''nugget effect''. Since CE_1 is a random and discontinuous error, it can only be minimized by performing ''correct'' sampling and, thus, minimizing the factors that influence FE and GE.

Sampling is a selection process. The selection process can either be probabilistic (i.e. sampling with a random selection component) or nonprobabilistic. Nonprobabilistic sampling (e.g. grab sampling) can never be strictly ''correct''. Probabilistic sampling can be ''correct'' or ''incorrect''. Probabilistic sampling is correct when all of the constituents of the lot to be sampled have an equal probability of being sampled and when all constituents that are not of the lot have no probability of being selected. The error associated with the selection/sampling process is known as the materialization error (ME) and is the combination of the two relative variances, the delimitation error (DE) and the extraction error (EE); that is, $ME = DE + EE$. The ME must be minimized to approach the notion of correct sampling. [Note: although the preparation error (PE) (to be discussed) is technically part of the ME, it is usually treated separately.]

The increment DE can be a major source of sampling bias and results from incorrectly defining the boundaries or limits of the volume of material to be extracted and physically collected from the sample.[22] In order to define the proper shape, the ''dimensionality'' of the lot, which ranges from 0 to 3, must be defined. A zero-dimensional lot exists when the whole lot is used for analysis and, thus, no DE exists, or when the order (through time or space) of the sampling has been lost or is irrelevant (e.g. a processed laboratory sample). Similarly to a zero-dimensional lot, a one-dimensional (1-D) lot may be defined as a thin, continuous, elongated pile or stream, except that the order of the sample is important. For zero- or 1-D lots, a correct sample is obtained by collecting all of the soil between two *parallel* planes cut across the entire sample. An example of a 1-D lot may be a core sample from which a slice is cut for the analytical subsample.

A 2-D lot is correctly sampled by a cylinder with a constant cross-section and consists of a mass with an upper and lower boundary. The third dimension,

although existing, for a 2-D lot should be insignificant when compared with the other two dimensions. Typical examples of 2-D lots in a soil is a horizon (or a designated sampling layer) in which the thickness of the horizon is negligible compared with its horizontal distribution or a core sample collected through the entire soil horizon.

A 3-D lot is a sample in which all three dimensions are important and significant. Examples of 3-D lots include large waste piles, truck loads of soil, and an entire waste site. 3-D lots are nearly impossible to sample correctly since the proper shape of the sample is a sphere. 3-D samples are typically reduced to 2-D samples for correct subsampling purposes. The 1-D and 2-D lots are most frequently encountered during soil sampling. The 2-D lot is collected in the field whereas the 1-D lot is sampled during sample preparation and analysis.

The other important aspect of DE is that the sampling device must be able physically to collect all particles (or portions of particles) that lie within the edges of the sampling device (Figure 5). If a particle or fraction of particle lies outside the sampling device, then it must be excluded from the collected sample. However, collection of only those portions of particles that lie within the edges of the sampling device is neither practical nor possible since the sampling device cannot cut soil particles. To account for the cutting error associated with the sampling device as it delimits the sample, the increment EE has been defined. To extract a sample properly, the ''rule of the center of gravity'' must be respected.[22] The rule

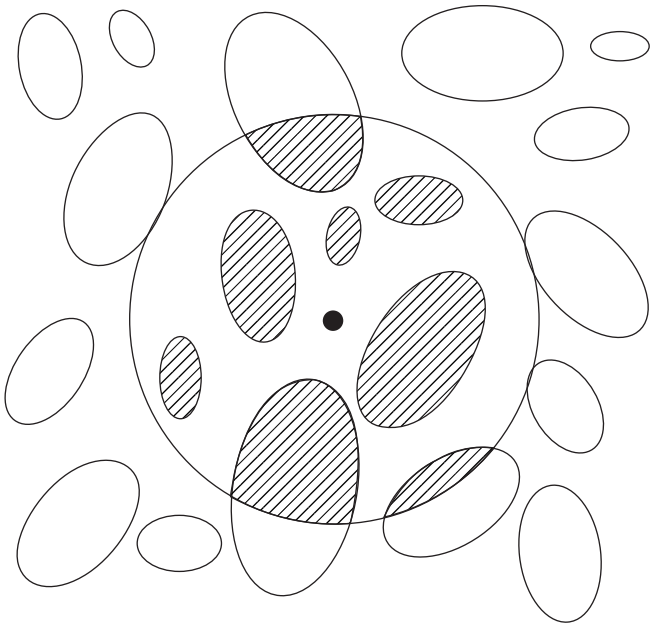

Figure 5 Illustration of proper increment delimitation. (Reprinted with permission from F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition. Copyright CRC Press, Boca Raton, FL. © 1993.)

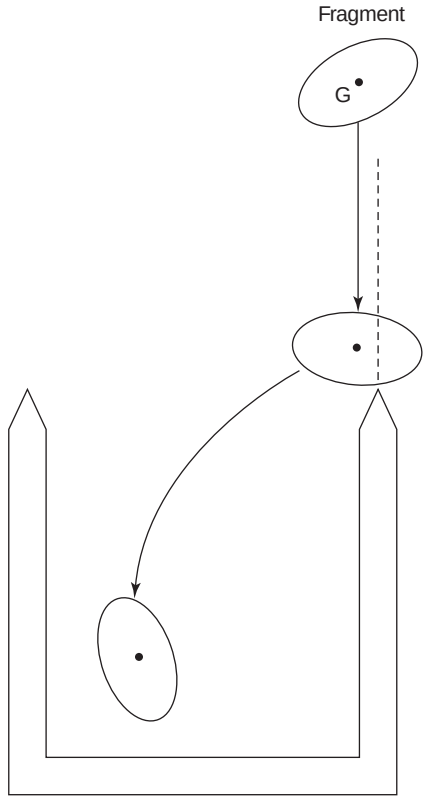

Figure 6 Illustration of the ''rule of the center of gravity''. G = the center of gravity of the fragment or soil particle. (Reprinted with permission from F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition. Copyright CRC Press, Boca Raton, FL. © 1993.)

of the center of gravity simply states that if the center of gravity for a particle lies within the sampling device boundaries, then the fragment must be included with high probability in the sample (Figure 6). Conversely, if the particle's center of gravity lies outside the sampling device boundaries, yet part of the particle lies within the sampling device's edges, then the particle must be excluded with high probability from the collected sample (Figure 7). Factors that can affect the center of gravity rule during soil sampling are as follows:

- The straightness of the cutting edge. The tips of the cutting edge should be superposable to each other by simple lateral translation (i.e. both edges of equal length).
- The shape of the cutting edge. The cutting edge should form an angle between the outer wall of the sampler, and the line parallel to the cutting edge should be less than 45° or equal to 90° (a flat cutting edge). If the cutting edge has two beveled edges from the tip, then a right-angle should be formed at the cutter tip between the two sides.

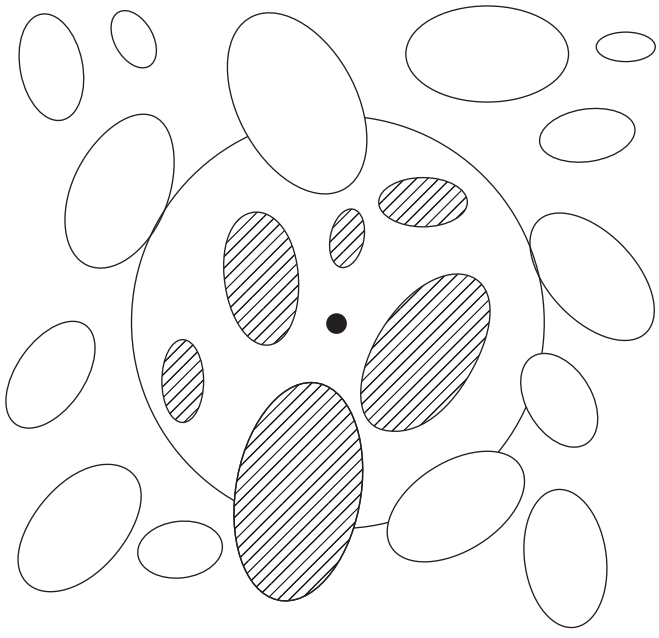

Figure 7 Illustration of proper increment extraction. (Reprinted with permission from F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition. Copyright CRC Press, Boca Raton, FL. © 1993.)

- The width of the cutting mouth. For $d \geq 3\,\text{mm}$, the width of the cutting mouth should be at least $3 \times d$. If $d < 3\,\text{mm}$, then the width should be at least 10 mm in diameter.

One possible means to minimize the EE is to select a larger diameter coring device.

The last two errors, the PE and AE, are the result of physical sample manipulation or from noise in the system.[22] The major sources of PE include contamination, alteration of chemical form (e.g. volatilization or redox reactions), human error, and loss of the sample via misplacement or spillage. PE may be unintentional and occur during sample grinding, sieving, and storage, or through other mistakes such as mislabelling or improper sample handling. Occasionally, intentional error (e.g. fraud or sabotage) may be encountered such that pollutants of concern are intentionally lost during preparation or sampling. AE is the error of which most analysts are cognizant. Unfortunately, it is in reducing, controlling, and quantifying AE that most QA/QC time, effort, and budget are expended. As presented throughout the discussion of particulate sampling theory, without giving proper attention to the actual soil sample collection, the sample received at the analytical laboratory may have already been so severely biased that controlling the AE may hardly be worth the time and expense.

7.3 Application of Particulate Sampling Theory to Soil Sampling

One of the keys to Gy's theory is that it provides a means to determine the quantity of soil to collect and submit a representative sample with an acceptable and set relative variance of the FE to the analytical laboratory based on the maximum particle diameter. Additionally, if a large sample is submitted, it provides a comminution (i.e. particle-size reduction) protocol to reduce the sample size yet maintain the FE below the acceptable relative variance as defined in the DQO process.[23] Of course, this assumes that the sample is collected correctly using all possible means to control and minimize the GE, DE, EE, PE, and AE, so that these errors are negligible compared with FE. Two approaches used to determine the required sample mass are to calculate the mass by solving for M_S in the FE equation or to create a sampling nomograph (Figure 8). Following the calculational pathway, for example, based on the assumptions of an allowable relative standard deviation (RSD) for FE of 15% (or a relative variance of FE of 0.0225), $M_L \gg M_S$, $d = 2\,\text{mm}$, $f = 0.5$, $g = 0.25$, $l = 0.2$, $\lambda = 2.65\,\text{g cm}^{-3}$, and contaminant concentrations (a_L) = 50, 100, 250, 500, and 1000 ppm, then the required sample weights are approximately 471, 236, 94, 47, and 24 g, respectively. By changing the allowable RSD of FE to 5% and following the same assumptions, the required sample weights would be 4240, 2120, 848, 424, and 212 g, respectively.

A sampling nomograph also allows the sampler to control sampling and sample processing such that no step exceeds the DQO established for allowable FE. The nomograph provides a visual interpretation of steps that may be necessary to reduce particle and sample size in addition to providing the required sample mass for a given maximum particle size. The construction of a nomograph simply involves solving and plotting the equation for the FE. The FE was defined according to Equation (12):

$$s_{FE}^2 = \left(\frac{1}{M_S} - \frac{1}{M_L}\right) fgcld^3 \qquad (12)$$

This equation can be rearranged to identify a sampling constant, C, by assuming constant values for f, g, c, and l and assuming that M_L is much larger (at least 10-fold) than M_S.[22] The equation then becomes Equation (13):

$$s_{FE}^2 = \frac{Cd^3}{M_S} \qquad (13)$$

where $C = fgcl$.

In order to put all the pertinent information on a single nomograph, a logarithmic coordinate system is used and thus the equation becomes Equation (14):

$$\log s_{FE}^2 = \log C + 3\log d - \log M_S \qquad (14)$$

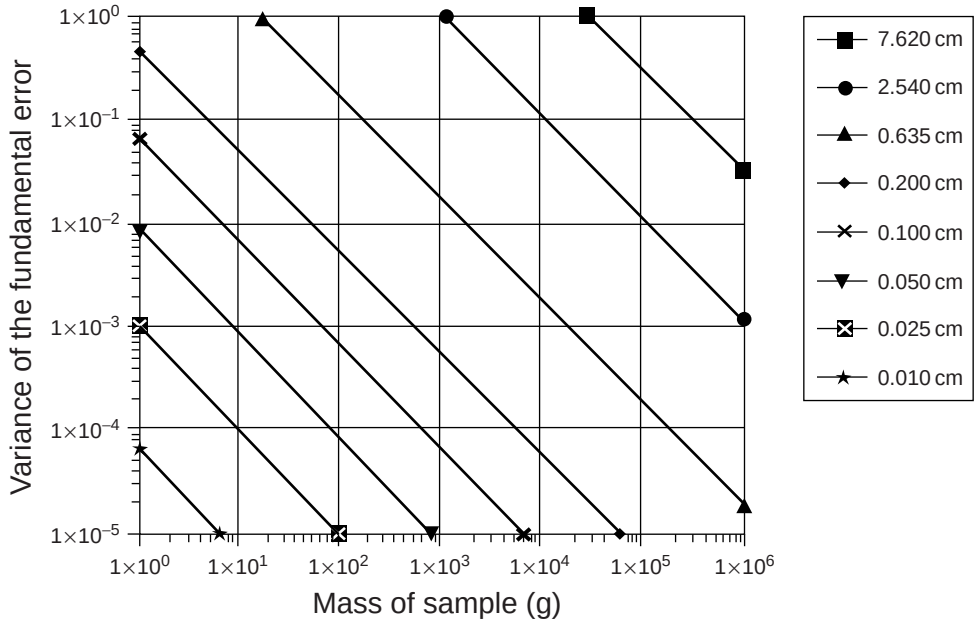

Figure 8 Typical sampling nomograph for soils where $g = 0.25$, $f = 0.5$, $l = 0.2$, $\lambda_M = 2.65\,g\,cm^{-3}$, and $a_L = 1000\,\mu g\,g^{-1}$ (ppm).

For a given particle size d and a sampling constant C, the value of $\log s^2_{FE}$ is directly proportional to $-\log M_S$.[22] The derivative of $\log s^2_{FE}$ with respect to M_S equals -1. Then, the line representing $\log s^2_{FE}$ as a function of $-\log M_S$ for a given particle size d and a sampling constant C has a slope of -1 on the nomograph. Therefore, to plot the nomograph for a given d with a constant C, $\log s^2_{FE}$ is plotted on the y-axis, $\log M_S$ is plotted on the x-axis, and $\log d$ is plotted as a family of parallel lines with a slope of -1 (Figure 8).

To identify a minimum required sample size given the largest particle diameter and acceptable relative variance limit, the sampler can follow the diagonal particle size line to the point where it intersects the horizontal variance line. At the point of intersection, the value on the x-axis equals the minimum sample size required to collect a representative sample.

Alternatively, if a large sample is collected whose mass must be reduced to a smaller size (e.g. to the size required for an extraction procedure), the nomograph provides information on how to reduce the sample mass without exceeding the acceptable DQO relative variance level. To identify the steps necessary (if any) to reduce the sample mass properly, the sampler needs to know the acceptable relative variance limit in addition to the starting and final sample weights. Following the same process for determining the required sample size, if the intersection of the particle size and variance lines results in a sample size smaller than the final required weight, then no particle size reduction is necessary, and the sample can be collected bearing in mind to minimize the other sampling errors. If not, then some form of particle size reduction (i.e. comminution) must be taken. The reduction in particle size can be performed by grinding, correct splitting, or by screening the sample to remove larger particle sizes.[23] The removal of the coarsest particle sizes should be based on the assumption that the largest particles do not contribute significantly to the contaminant concentration. To select an appropriate particle reduction scheme, simply follow the maximum particle diameter line from its intersection with the starting weight line until it intersects the acceptable variance line. At this point, a vertical line should be drawn until it intersects a smaller particle diameter line that falls below the variance line for the final required sample size. Once an appropriate smaller particle diameter line has been identified, then the whole sample needs to be reduced to have the coarsest particles equal to or smaller than that selected diameter. Several steps may be required to reduce the maximum particle sizes properly but the key is to not go above the acceptable relative variance limit.

7.4 Selection of Sampling Equipment

The selection of sampling equipment should be performed with care since the sampling equipment must be able to collect a representative and correct sample and because the sampling equipment comes in direct contact with the soil. Criteria that should be considered during the selection of the appropriate sampling tool are chemical and physical compatibility, matrix effects, volume capacity, physical requirements for equipment use, ease of

operation, time requirement, decontamination and reuse potential, cost, and soil type to be sampled.[16,24,25] Further, the appropriate sampling device(s) selected must be correct with respect to delimitation and EEs previously discussed. Each of these factors will be discussed in greater detail in the following text.

Since the soil comes into direct contact with the sampling tool, the sampling tool must be compatible both chemically and physically with the soil and the contaminants present in the soil.[24] Chemical compatibility is a concern primarily when organic contaminants are present. The sampling tool should not be a source of additional sample contamination and should be free from (or at least resistant to) chemical degradation that may occur due to interactions with the samples. For example, plastic or acetate sleeves commonly used as liners in coring equipment should be avoided when sampling for organics because they can ''bleed'' phthalates and other organic compounds into the sample as well as degrade owing to the ''solvent'' effect of some organics that may be present in the contaminated soil. Conversely, certain organic contaminants may partition into the plastic resulting in lower contaminant concentrations and cross-contamination if the plastic sampler is used to obtain another sample. Physical compatibility is a consideration in terms of the physical strength of the sampler to resist deformation when collecting the soil and for the possibility of sample contamination due to physical abrasion of the sampler, with the separated particles becoming part of the analytical sample. For example, the use of stainless-steel sampling equipment for sampling chromium-contaminated soils should be avoided since chromium is a major elemental component in stainless steel.

The effect of the sampling device on the soil matrix is an important concern because the design of the device influences the representativeness of the collected sample.[24] As previously discussed, if the sampling device excludes a given particle-size fraction (Gy's increment EE) or improperly collects the sample by influencing which part of the whole sample is collected (Gy's increment delineation error), then major sources of error enter into the sampling and analytical stream. Further, sample disturbance is a primary consideration when sampling soil for VOCs. An intact core is highly preferred (versus a disaggregated sample obtained from devices such as augers) during collection of VOC-contaminated soils since the minimization of atmospheric exposure of the sample and consequent VOC loss from the sample is imperative.

Most sampling devices provide adequate sample volume.[24] However, sampling device volumes should be compared with the volume necessary for all required analyses and, QA purposes, and provide sufficient excess sample for archiving and reanalyses purposes. If the device does not provide adequate volume, the following options should be considered: collecting multiple increments/samples in very close proximity to the original location, using a similar device with increased capacity, using an alternate device with increased capacity, or modifying the existing device.

The physical requirements, ease of operation, and time required for use of the sampling equipment all relate to the transport and operation of the selected sampling device.[24] Physical requirements to be considered are the device's size and weight since the sampler may have to carry the device to the sampling location and manually collect the sample. If the device is power driven, then the power source and ancillary equipment (e.g. drill rigs, trucks) are a concern in selecting these devices. Ease of operation concerns involve the training of personnel to use the sampling device effectively and properly. Fortunately, most manually operated sampling equipment is relatively simple to use, and representative samples can be collected as long as proper care is taken during sample collection. A factor influencing both the ease of operation and time requirement for sampling is the labor requirement. Most power-driven sampling tools require multiple-person teams to effectively and safely use the equipment but have the advantage of shorter sample collection times, especially when depths exceed 1 m (39 in). In contrast, manual sampling tools, while allowing a single sampler to collect surface samples quickly and efficiently, require much longer times to collect samples at depth than power-driven tools.

Decontamination, reuse potential, and cost are concerns when sampling equipment may be used for multiple sampling programs.[24] Three aspects of decontamination, and thus the device's reuse potential, to consider are the procedure's ease, success (i.e. ability to eliminate or minimize the potential for sample cross-contamination), and time involved. Ideally, the decontamination procedure should be simple, quick, and successful. Most soil sampling equipment can be effectively decontaminated with a simple soap and water wash followed by water rinse. However, where certain organic compounds are of concern, a solvent rinse may also be necessary. [Note: caution should be taken during the selection and use of solvents to rinse sampling equipment to avoid cross-contamination of the samples, health and safety issues, and spillage such that the solvents become part of the problem at the site.] If successful decontamination procedures are not possible owing to the soil matrix or contaminant (e.g. soil contaminated with heavy oils or tars), the use of disposable sampling tools should be considered to avoid sample cross-contamination. Cost considerations should be contemplated and compared with the life expectancy of the equipment and use expectancy (i.e. will the equipment be used for multiple projects within the organization or should it be leased for a short one-time project).

Several soil types make the routine collection of soil samples difficult and thus require specialized (or modified) sampling tools. These soil types include stony soils, noncohesive soils, and saturated (or nearly saturated) soils.[16,24] Sampling difficulties are encountered in stony soils owing to the blocking or plugging of the device's opening, thus preventing the collection of the sample. Large stones may also prevent the penetration of the sampling device to any significant depth. Noncohesive soils, such as loose sands, require special sampling equipment that retains the soil within the sampler and thus prevents the soil from flowing out of the open end of the sampling device. In contrast, saturated soils (e.g. mucks and muds) are generally difficult to collect and remove from the sampling tool, especially if the soils have high clay contents. Under these circumstances, specialized sampling tools are required that will allow for the simple removal of the soil from the sampling device.

7.5 Soil Sampling Tools

Soil sampling tools can be divided into two main categories based on the depth to which the samples can be collected. Surface samplers are generally designed to

Table 1 Guide to sampling equipment selection[a]

Type of sampler	Typical length/i.d. (in)[b,c]	Relative sample size[c]	Liners	Core samples	Suitable soil types			Manual or power	Labor[c]
					Cohesive[c]	Moisture[c]	Stony		
(A) Surface samplers									
1. Spoons, trowels, and shovels	Var	Var	No	No	All	All	Yes	Manual	Single
2. Soil punches and sampling tubes	6–10/1–3	Small	Yes	Yes	Coh	All	No	Manual	Single
3. Probes and tube samplers	6–24/1	Small	No	Yes	Coh	All	No	Manual	Single
4. Soil corers	2–12/1–3	Small	Yes	Yes	Coh	All	No	Manual	Single
5. Split core/barrel samplers	6–12/1–2	Small	Yes	Yes	Coh	Int	No	Manual	Single
6. Screw augers	6–12/1–2	Small	No	No	Coh	Int	No	Manual	Single
7. Soil recovery augers	8–12/2–3	Large	Yes	Yes	Coh	Int	No	Manual	Single
(B) Subsurface samplers									
1. Augers									
a. Power augers	36–60/2–6	Large	No	No	All	Int	No	Power	Multi
b. Solid-stem flight augers	36–60/2–6	Large	No	No	All	All	Yes	Power	Multi
c. Hollow-stem augers	36–60/2–4	Large	No	Yes	All	All	Yes	Power	Multi
d. Regular bucket augers	6[d]/2–4	Large	No	No	Coh	Int	No	Manual	Single
e. Sand bucket augers	6[d]/2–4	Large	No	No	Non	Int	No	Manual	Single
f. Mud bucket augers	6[d]/2–4	Large	No	No	Coh	Wet	No	Manual	Single
f. Dutch augers	6[d]/3	Large	No	No	Coh	Wet	No	Manual	Single
e. Screw augers	12[d]/1–2	Small	No	No	Coh	Int	No	Manual	Single
g. Posthole auger	6[d]/2–8	Large	No	No	Coh	Wet	No	Manual	Single
2. Tube-type samplers									
a. Shelby/thin-wall tube sampler	30/2–4	Large	No	Yes	Coh	Int	No	Both[e]	Multi
b. Subsoil probes	36–48/1	Small	No	Yes	Coh	All	No	Manual	Single
c. Split spoon/tube samplers	12–60/2–4	Large	Yes	Yes	Coh	Int	No	Both[e]	Multi
d. Continuous sampler	48–60/2–5	Large	Yes	Yes	Coh	All	No	Power	Multi
e. Veihmeyer sampling tubes	48–72/1	Large	No	Yes	Coh	Int	Yes	Manual	Single
f. Zero contamination sampler	12–24/1	Small	Yes	Yes	Coh	Wet-Int	No	Both[e]	Multi
g. Ring-lined barrel sampler	24–60/2–4	Large	Yes	Yes	All	Wet-Int	Yes	Both[e]	Multi
(C) Miscellaneous samplers									
1. Triers	6–60/1–3	Var	No	No	Coh	Int	No	Manual	Single
2. Thiefs	6–72/1	Var	No	No	Non	Dry	No	Manual	Single
3. Backhoes	Var	Large	No	No	All	All	Yes	Power	Multi

[a] This table does not purport to be exhaustive. Only the general names of the sampler types are presented. Varying names for the same sampler or same type of sampler with slight modifications are commonly identified in the literature and sales catalogs.
[b] Approximate lengths and internal diameters (i.d.), in inches.
[c] Var = variable depending upon length of sampler or size of hole dug; Coh = cohesive; Non = noncohesive; Int = intermediate wetness (i.e. not dry nor wet); Multi = two or more persons required to operate effectively.
[d] Manual auger lengths given are just for portion of the sampler where the actual sample is held. Most bucket augers have handles and extensions that allow for greater sampling depths with multiple passes.
[e] Manually operated versions of these samplers can be operated by one person.

collect the sample within the upper 30 cm (12 in) with a single pass (i.e. single insertion into the soil) of the sampling device. Subsurface devices are generally designed to sample to a depth of 1.5 m (60 in) in a single pass but can usually collect deeper samples with multiple passes. Table 1 presents commonly available sampling tools and information on the selection criteria previously discussed. Project-specific parameters, such as compatibility, matrix effects, decontamination requirements, and cost will need to be determined on a per-project basis.

While spoons, scoops, trowels, and shovels (or backhoes) are perhaps the most readily available and widely used sampling tools for the collection of surface samples, they also tend to introduce the greatest sampling error. The most common cause of the sampling error is the improper increment delimitation in which part of the sample is unintentionally excluded from the sample to be analyzed.[22] For example, when using a shovel, rarely is the hole dug such that the side walls are parallel to each other and perpendicular to the surface. The natural tendency is to dig a hole with side walls at an angle to the surface (Figure 9c). If the sample collected from the soil is removed from the hole, the lower portion of the sample is under-represented since less sample is collected at that depth than at the wider surface mouth. Under these circumstances, if the contaminant has been leached to a depth within a given sampling increment, a lower contaminant concentration may be obtained due to sample ''dilution'' with the cleaner overlying material. Conversely, if the contaminant is present at or near the surface, then the reported concentrations may overestimate the actual concentration due to inclusion of a greater proportion of surface versus subsurface soil. To overcome this problem, it is recommended to dig a hole to the required depth and collect the sample from the walls of the hole being careful to collect a uniform sample thickness from the entire length of the sampling stratum. Alternatively, care must be taken to dig the hole properly to include equal proportions of all fractions within the sampling increment (Figures 9a and b).

The same general principle is the basis for the error introduced when using spoons, scoops, and trowels. Most of these samplers have a rounded bottom which, when used to collect soil, preferentially collects less of the lower portion of the sample than the surface and thus biases the results (Figure 10a). Additionally, since none of these sampling utensils have side walls capable of holding the entire sample within the sampler, as the sample is removed and transported to the sample container, portions of the sample may be lost by ''falling'' off the sampler (Figure 11a). In some cases, both the upper and lower portion of the sample may be lost with the greatest sampled portion being collected from the middle of the sampling stratum. To collect a sample properly with these types of tools would require equipment modification to have parallel side walls and a flat bottom capable of holding the entire sampling increment (Figures 10b and 11b).

Tube-type samplers, including probes, punches, corers, tube samplers, barrel samplers, continuous samplers, and zero contamination samplers, tend to have a distinct advantage over other forms of samplers since an intact soil core is collected. The cylindrical shape of the intact core is the correct shape for a 2-D lot and, thus, reduces

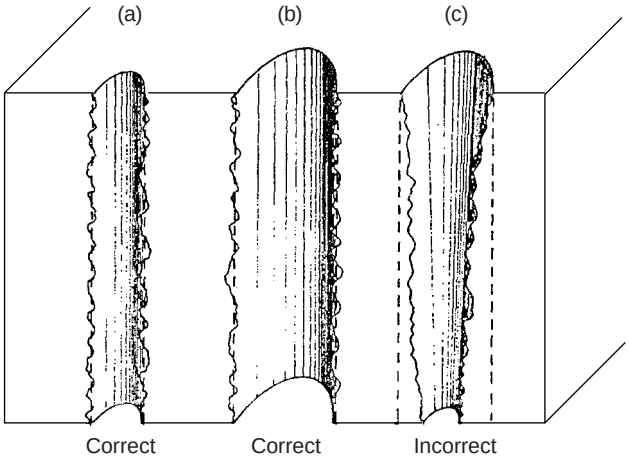

Figure 9 Illustration of correct and incorrect sample delimitation using a coring device (a) and shovels (b) and (c). (Reprinted with permission from F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition. Copyright CRC Press, Boca Raton, FL. © 1993.)

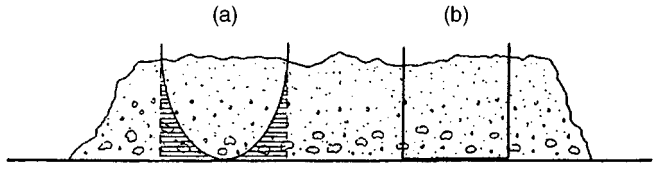

Figure 10 Incorrect increment delimitation introduced by a scoop with round bottom. (Reprinted with permission from F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition. Copyright CRC Press, Boca Raton, FL. © 1993.)

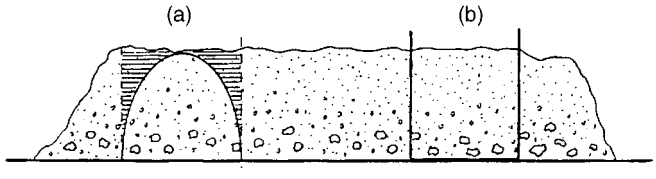

Figure 11 Incorrect increment delimitation introduced by scoop with no side walls to prevent material from falling. (Reprinted with permission from F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition. Copyright CRC Press, Boca Raton, FL. © 1993.)

the increment DE as long as care is taken to collect the entire layer or depth defined as the sampling increment (Figures 9a and b).[22]

When more sample is collected than required for the sampling increment [e.g. a 25-cm (10-in) core is collected and subsampled to collect the top 10 cm (4 in)], care must be taken to delineate the proper increment by cutting the new boundary parallel to the surface of the sample at the desired depth. For the thinner diameter tube-type samplers (<2.5 cm i.d.; 1 in), the increment EE may be a concern owing to exclusion of larger particles at the mouth of the sampling tube.

Augers, whether used to collect surface or subsurface samples, with the exception of the hollow-stem auger, collect samples that are disaggregated by the cutting bit or within the auger threads. This cutting action destroys natural soil structure and makes it difficult to delineate accurately soil horizon boundaries or sampling depths. This disaggregating property of augers also makes them unsuitable for the collection of soils contaminated with VOCs since the sample disturbance allows for the rapid loss of VOCs to the atmosphere. Increment DE may also occur with augers since the lowest portion of the sample (i.e. that portion where threads do not overlap and have an upper and lower thread to hold the sample) cannot be collected.

Thief and triers are two types of sampling tools that were not designed for the collection of soils but can be used when necessary. Thiefs are usually pushed into the sample and rotated to open ''windows'' and allow the sample to flow into the central chamber/barrel to be collected.[24] This type of filling action leads to the concerns about incorrect sample delimitation since (a) the sample below the windows, although truly a part of the sample, cannot be collected, (b) only those particles that flow easily are collected, (c) particles larger than the window diameter are excluded, thus biasing the collected sample to the finer particle sizes, and (d) the proper cylindrical shape for a 2-D lot is not guaranteed. Triers, on the other hand, while nearly capable of collecting the proper sample cylinder, do not have side walls to hold the sample once collected. This lack of side walls can lead to sample loss during sample removal and transfer operations, resulting in potentially biased analytical results.

8 SAMPLE HANDLING AND PREPARATION

Once the sample has been collected, the next steps that occur prior to preparing subsamples for analysis are sample handling and preparation. Various concerns and steps are involved which are briefly presented here to remind the sampler of their final responsibilities. These steps include sample compositing, homogenization, preservation, and storage.

Sample compositing is the process of combining several distinct subsamples to create a single sample for analysis.[26] The combined subsamples are then homogenized to make up the final sample submitted for analysis. The advantages of composite sampling are that it reduces the cost of analysis at a waste site and provides an estimate of the mean concentration. If the sampler follows the principles of Gy's particulate sampling theory, then sample compositing after the collection of numerous small increments is an excellent way to reduce GE. The disadvantages of sample compositing are a loss of information about variation within the sampling area,[27] loss of sensitivity because of sample dilution,[28] and the time required to homogenize the samples prior to laboratory subsampling. In situations where rapid, field-portable methods are used (e.g. field portable X-ray fluorescence spectrometry), compositing may not be necessary since individual samples can be analyzed in real time to provide a more definitive answer concerning the distribution of a contaminant at the waste site.

Numerous techniques have been used to homogenize soils prior to the removal of the analytical subsample.[29,30] Some of these techniques include bottle shaking, stirring, sheet mixing or rolling, tumbling, mechanical mixing (e.g. cement mixers or V-blenders), riffling, cone-and-quartering, and sectorial splitting. Homogenization is performed to diminish GE within the sample or lot. When performing sample homogenization, perhaps the most frequently asked question is ''when has homogeneity been reached?''. Although there is no patented answer to this question, most investigators will ''homogenize'' their samples for a fixed time, a fixed number of passes, or until visual homogeneity is reached without ever testing the degree of sample homogeneity. Investigators should always test their homogenization technique prior to full-scale usage especially since some ''homogenizers'' will actually lead to resegregation of the sample.

In one experiment to determine the effectiveness (degree of homogeneity) and the efficiency (time consumption) of riffling versus cone-and-quartering as homogenization techniques, it was found that using a closed-bin riffle splitter (which also reduced visual fine particle loss) was more consistently effective and efficient than cone-and-quartering, open-bin riffle splitting, or random sampling, when the sample was passed through the riffle splitter five times.[30] In this experiment, homogenization was performed on bulk soils (approximately 5 kg each with textures ranging from sand through clay) by either pouring the soil collected in both bins back through

the riffle splitter, or by forming a new cone as the individual quarters are removed during cone-and-quartering. While homogenization can help overcome grouping and segregation problems that occur during shipment and handling, properly performed incremental sampling can also produce the same benefits without the added time and expense.[23]

Sample preservation for most soil samples is simply to cool the sample to 4 °C or colder.[31] For soils contaminated with VOCs, ''preservatives'' such as methanol or other biocides are commonly used, in conjunction with cooling, to help reduce biological degradation and volatilization of the VOCs. Sample storage in either plastic or glass containers is generally acceptable for soils contaminated with inorganics. In contrast, organically contaminated soils should be stored in glass containers (with Teflon™-lined caps for VOC-containing soils). It should be noted that some compounds are light-sensitive, and such samples should be stored in the dark to avoid any photochemical reactions.

ACKNOWLEDGMENTS

The work presented here was based on an original chapter prepared by the authors for the *Encyclopedia of Environmental Analysis and Remediation*.[32] The US Environmental Protection Agency (EPA), through its Office of Research and Development, funded and performed the research described. The manuscript was subjected to the EPA's peer and administrative review and was approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABBREVIATIONS AND ACRONYMS

AE	Analytical Error
ASTM	American Society for Testing and Materials
CE_1	Short-range Heterogeneity Error
DE	Delimitation Error
DL	Detection Limit
DQO	Data Quality Objective
EE	Extraction Error
FE	Fundamental Error
FSP	Field Sampling Plan
GE	Grouping and Segregation Error
GIS	Geographical Information Systems
GPS	Global Positioning System
HASP	Health and Safety Plan
ME	Materialization Error
NRCS	National Resource Conservation Service
OSHA	United States Occupational Safety and Health Administration
PE	Preparation Error
PPE	Personal Protective Equipment
QA	Quality Assurance
QAPP	Quality Assurance Project Plan
QC	Quality Control
RSD	Relative Standard Deviation
SD	Standard Deviation
SOP	Standard Operating Procedure
USDA	United States Department of Agriculture
USEPA	United States Environmental Protection Agency
VOA	Volatile Organic Analysis
VOC	Volatile Organic Compound
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Soil Instrumental Methods

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction

Pesticides **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Steel and Related Materials **(Volume 10)**
Iron Ore, Sample Preparation and Analysis of • Metal Analysis, Sampling and Sample Preparation in

X-ray Spectrometry **(Volume 15)**
Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. ASTM, *Draft Manual on Practical Illustrations of the Planning, Implementation and Assessment of Sampling for Common Waste Management Activities*, American Society for Testing and Materials, W. Conshohocken, PA, 1999.
2. T.H. Starks, 'Determination of Support in Soil Sampling', *Math. Geol.*, **18**(6), 529–537 (1986).
3. USEPA, 'Guidance for the Data Quality Objective Process', EPA QA/G4, Quality Assurance Management Staff, Washington, DC, 1994.
4. USEPA, 'The Data Quality Objectives Process for Superfund: Interim Final Guidance', EPA/540/R-93/071, Quality Assurance Management Staff, Washington, DC, 1993.
5. USEPA, 'Guidance for Data Quality Assessment – Practical Methods for Data Analysis', EPA QA/G-9, EPA/600/R-96/084, Quality Assurance Management Staff, Washington, DC, 1996.
6. USEPA, 'A Rationale for the Assessment of Errors in the Sampling of Soils', EPA/600/X-89/203, Environmental Monitoring Systems Laboratory, Las Vegas, NV, 1989.
7. E.H. Isaaks, R.M. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.
8. R. Webster, M.A. Oliver, *Statistical Methods in Soil and Land Resource Survey*, Oxford University Press, New York, 1990.
9. A.B. McBratney, R. Webster, T.M. Burgess, 'The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables. I: Theory and Method', *Comput. Geosci.*, **7**(4), 331–334 (1981).
10. A.B. McBratney, R. Webster, 'The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables. II: Program and Examples', *Comput. Geosci.*, **7**(4), 335–365 (1981).
11. E.J. Englund, N. Heravi, 'Phased Sampling for Soil Remediation', *Environ. Ecol. Stat.*, **1**, 247–263 (1994).
12. E.J. Englund, N. Heravi, 'Conditional Simulation: Practical Application for Sampling Design Optimization', in *Geostatistics Troia '92*, ed. A. Soares, Kluwer, Dordrecht, 1993.
13. USEPA, 'Guidance for Conducting Remedial Investigations and Feasibility Studies under CERCLA', EPA/540/G-89/004, Environmental Protection Agency, Washington, DC, 1989.
14. G.F. Simes, 'Preparing Perfect Project Plans', EPA/600/8-89/087, Environmental Protection Agency, Cincinnati, OH, 1989.
15. M.E. Byrnes, *Field Sampling Methods for Remedial Investigations*, Lewis Publishers, Boca Raton, FL, 1994.
16. D.S. Barth, B.J. Mason, T.H. Starks, K.W. Brown, 'Soil Sampling Quality Assurance User's Guide', 2nd edition, EPA 600/8-89/046, Environmental Protection Agency, Las Vegas, NV, 1989.
17. R.E. Cameron, 'Guide to Site and Soil Description for Hazardous Waste Site Characterization, Vol. 1 – Metals', EPA/600/4-91/029, Environmental Protection Agency, Las Vegas, NV, 1991.
18. J.R. Boulding, 'Description and Sampling of Contaminated Soils – A Field Pocket Guide', EPA/625/12-91/002, Environmental Protection Agency, Cincinnati, OH, 1991.
19. Soil Survey Staff, *Soil Survey Manual*, US Department of Agriculture Handbook No. 18, US Department of Agriculture, Washington, DC, 1951.
20. L.M. Thompson, F.R. Troeh, *Soils and Soil Fertility*, McGraw-Hill, New York, 1978.
21. A. Klute, C. Dirksen, 'Hydraulic Conductivity and Diffusivity: Laboratory Methods', in *Methods of Soil Analysis, Part I – Physical and Mineralogical Methods*, 2nd edition, ed. A. Klute, American Society of Agronomy, Madison, WI, 1986.
22. F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, 2nd edition, CRC Press, Boca Raton, FL, 1993.
23. B.J. Mason, 'Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies', EPA/600/R-92/128, Environmental Protection Agency, Las Vegas, NV, 1992.
24. ASTM, *Draft Standard Guide for Selection of Sampling Equipment for Waste and Contaminated Media Data Collection Activities*, American Society for Testing and Materials, W. Conshohocken, PA, 1995.
25. T.E. Lewis, A.B. Crockett, R.L. Siegrist, K. Zarrabi, 'Soil Sampling and Analysis for Volatile Organic Compounds', EPA/540/4-91/001, Environmental Protection Agency, Las Vegas, NV, 1991.
26. ASTM, *Draft Standard Guide for Composite Sampling and Field Subsampling for Environmental Waste Management Activities*, American Society for Testing and Materials, W. Conshohocken, PA, 1996.
27. J. Crépin, R.L. Johnson, 'Soil Sampling for Environmental Assessment', in *Soil Sampling and Methods of Analysis*, ed. M.R. Carter, Lewis Publishers, Boca Raton, FL, 1993.
28. L.H. Keith, *Environmental Sampling and Analysis – A Practical Guide*, Lewis Publishers, Boca Raton, FL, 1991.
29. R.E. Cameron, G.B. Blank, D.R. Gensel, 'Sampling and Handling of Desert Soils', NASA Technical Report 32-908, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 1966.
30. B.A. Schumacher, K.C. Shines, J.V. Burton, M.L. Papp, 'Comparison of Three Methods for Soil Homogenization', *Soil Sci. Soc. Am. J.*, **54**, 1187–1190 (1990).
31. USEPA, 'Test Methods for Evaluating Solid Waste, SW-846', 3rd edition, Office of Solid Waste and Emergency Response, Washington, DC, 1986.
32. B.A. Schumacher, J.J. van Ee, E.J. Englund, 'Soil Sampling at Hazardous Waste Sites', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 1998.

Solid-phase Microextraction in Environmental Analysis

Janusz Pawliszyn
University of Waterloo, Waterloo, Ontario, Canada

1 Introduction **1**

2 Theoretical Aspects of Solid-phase Microextraction Optimization and Calibration **4**
2.1 Thermodynamics 4
2.2 Kinetics 6

3 Method Development **8**
3.1 Selection of Fiber Coating 8
3.2 Selection of the Derivatization Reagent 9
3.3 Selection of the Extraction Mode 10
3.4 Selection of the Agitation Technique 10
3.5 Selection of Separation and/or Detection Technique 11
3.6 Optimization of Desorption Conditions 11
3.7 Optimization of Sample Volume 13
3.8 Determination of the Extraction Time Profile in a Pure Matrix 14
3.9 Determination of the Extraction Time 15
3.10 Calculation of the Distribution Constant 15
3.11 Optimization of Extraction Conditions 17
3.12 Determination of the Linear Dynamic Range of the Method for a Pure Matrix at Optimum Extraction Conditions 18
3.13 Selection of the Calibration Method 18
3.14 Optimization of Extraction Conditions for Heterogeneous Samples 18
3.15 Verification of the Equilibration Time, Sensitivity and Linear Dynamic Range for Real Sample Matrices 19
3.16 Determination of Method Precision 20
3.17 Determination of Method Detection Limit 21
3.18 Validation of the Method 21
3.19 Automation of the Method 22

4 Applications of Solid-phase Microextraction **22**
4.1 Gaseous Matrices 22
4.2 Liquid Matrices 24
4.3 Solid Matrices 25
4.4 Environmental Applications 26
4.5 Food and Pharmaceuticals 28
4.6 Clinical and Forensic Analysis 29
4.7 Future Analytical Developments 30
4.8 Physicochemical Applications 30

Acknowledgments **31**

Abbreviations and Acronyms **31**

Related Articles **32**

References **32**

Solid-phase microextraction (SPME) uses a small volume of sorbent dispersed typically on a surface of small fibers to isolate and concentrate analytes from a sample matrix. After contact with the sample, analytes are absorbed or adsorbed by the fiber phase (depending on the nature of the coating) until an equilibrium is reached in the system. The amount of an analyte extracted by the coating at equilibrium is determined by the magnitude of the partition coefficient of the analyte between the sample matrix and the coating material. After the extraction step, the fibers are transferred, with the help of the syringe-like handling device, to an analytical instrument, for separation and quantitation of target analytes. This technique is able to integrate sampling, extraction and sample introduction in a simple way, facilitating on-site monitoring. The additional advantages include elimination of solvents from the sample preparation step and convenient introduction of extracted components into the analytical instrument. Applications of this technique to date include environmental, industrial hygiene, process monitoring, clinical, forensic, food and flavor, fragrance and drugs, in laboratory and on-site analysis.

1 INTRODUCTION

SPME was developed to address the need for fast, solvent-free and field-compatible sample preparation technologies. The information presented below is a summary of the comprehensive discussion of the topic covered in a published book.[1]

SPME was introduced as a solvent-free sample preparation technique in 1990. The basic principle of this approach is to use a small amount of the extracting phase, usually less than one microliter. The sample volume can be very large, when the investigated system, for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

example air in room or lake water, is sampled directly. The extracting phase can be either a high-molecular-weight polymeric liquid, similar in nature to stationary phases in chromatography, or it can be a solid sorbent, typically of a high porosity to increase the surface area available for adsorption.

To date the most practical geometric configuration of SPME utilizes a small fused-silica fiber, usually coated with a polymeric phase. The fiber is mounted for protection in a syringe-like device. The analytes are absorbed or adsorbed by the fiber phase (depending on the nature of the coating) until an equilibrium is reached in the system. The amount of an analyte extracted by the coating at equilibrium is determined by the magnitude of the partition coefficient (distribution ratio) of the analyte between the sample matrix and the coating material.

In SPME, analytes typically are not extracted quantitatively from the matrix. However, equilibrium methods are more selective because they take full advantage of the differences in extracting-phase/matrix distribution constants to separate target analytes from interferences. Exhaustive extraction can be achieved in SPME when the distribution constants are large enough. This can be accomplished for most compounds by the application of internally cooled fiber.[2] In exhaustive extraction, selectivity is sacrificed to obtain quantitative transfer of target analytes into the extracting phase. One advantage of this approach is that, in principle, it does not require calibration, since all the analytes of interest are transferred to the extracting phase. On the other hand, the equilibrium approach usually requires calibration when dealing with complex matrices. This is accomplished by using surrogates or standard addition to quantify the analytes and compensate for matrix-to-matrix variations and their effect on distribution constants.

Since equilibrium rather than exhaustive extraction occurs in microextraction methods, SPME is ideal for field monitoring. It is unnecessary to measure the volume of the extracted sample and therefore the SPME device can be exposed directly to the investigated system for quantitation of target analytes. In addition, extracted analytes are introduced to the analytical instrument by simply placing the fiber in the desorption unit (Figure 1b and c). This convenient, solvent-free process facilitates sharp injection bands and rapid separations.[3] These features of SPME result in integration of the first steps in the analytical process: sampling, sample preparation and introduction of the extracted mixture to an analytical instrument.

The equilibrium nature of the technique also facilitates speciation in natural systems since the presence of a minute fiber, which removes small amounts of target analytes, is not likely to disturb the system. Because of the small size, coated fibers can be used to extract analytes from very small samples. For example, SPME has been used to probe for substances emitted by a single flower bulb during its lifespan.

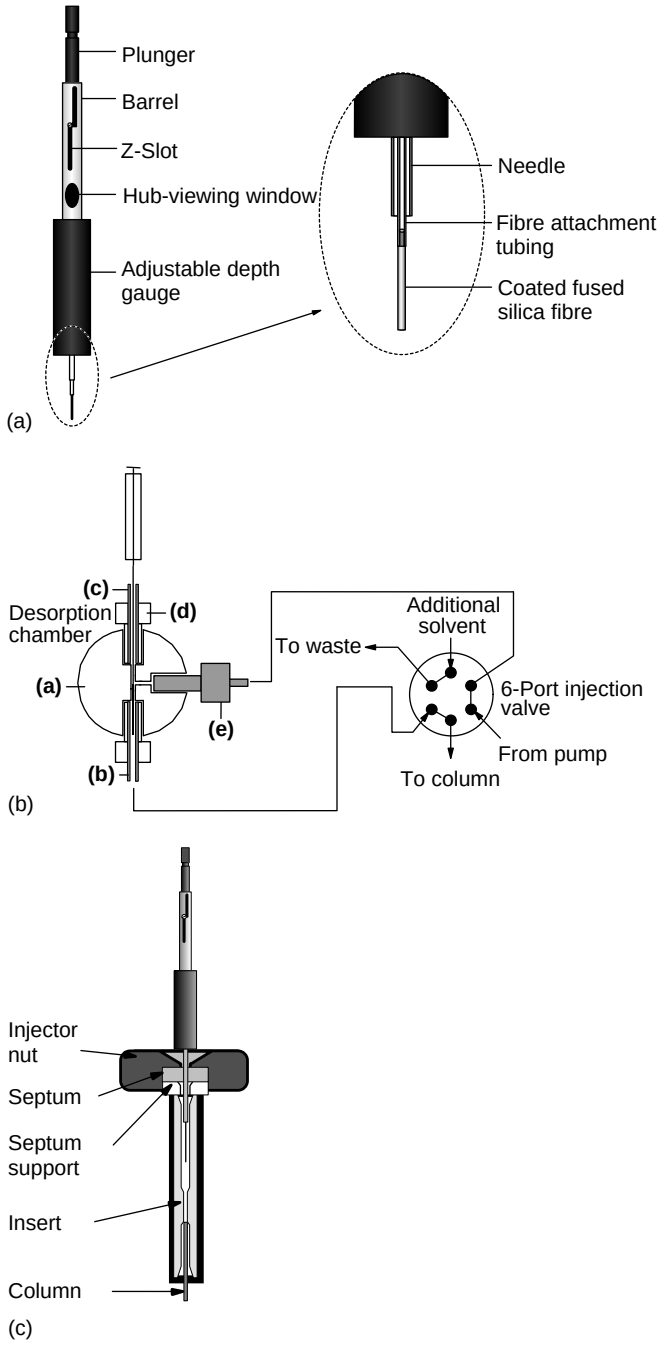

Figure 1 (a) Design of the commercial SPME device. (b) SPME/HPLC interface: (**a**) stainless steel (SS) 1/16″ tee, (**b**) 1/16″ SS tubing, (**c**) 1/16″ polyetheretherketone (PEEK) tubing (0.02″ ID), (**d**) two-piece finger-tight PEEK union, (**e**) PEEK tubing (0.005″ ID) with a one-piece PEEK union. (c) SPME/GC (solid-phase microextraction/gas chromatography) interface.

Figure 1(a) illustrates the commercial SPME device, manufactured by Supelco, Inc. (Bellefonte, PA). The

fiber, glued into a piece of stainless steel tubing, is mounted in a special holder. The holder is equipped with an adjustable depth gauge, which makes it possible to control repeatably how far the needle of the device is allowed to penetrate the sample container (if any) or the injector. This is important, as the fiber can be easily broken when it hits an obstacle. The movement of the plunger is limited by a small screw moving in the z-shaped slot of the device. For protection during storage or septum piercing, the fiber is withdrawn into the needle of the device, with the screw in the uppermost position. During extraction or desorption, the fiber is exposed by depressing the plunger, which can be locked in the lowered (middle) position by turning it clockwise (the position depicted in Figure 1a). The plunger is moved to its lowermost position only for replacement of the fiber assembly. Each type of fiber has a hub of a different color. The hub-viewing window enables a quick check of the type of fiber mounted in the device.

If the sample is placed in a vial, the septum of the vial is first pierced with the needle (with the fiber in the retracted position), and the plunger is lowered, which exposes the fiber to the sample. The analytes are allowed to partition into the coating for a predetermined time, and the fiber is then retracted back into the needle. The device is next transferred to the analytical instrument of choice. When gas chromatography (GC) is used for analyte separation and quantitation, the fiber is inserted into a hot injector, where thermal desorption of the trapped analytes takes place (Figure 1c). The process can be automated by using an appropriately modified syringe autosampler. For high-performance liquid chromatography (HPLC) applications, a simple interface mounted in place of the injection loop can be used to re-extract analytes into the desorption solvent (Figure 1b).

A SPME device is capable of both spot and time-averaged sampling. As described above, for spot sampling, the fiber is exposed to a sample matrix until the partitioning equilibrium is reached between the sample matrix and the coating material. In the time-averaged approach, on the other hand, the fiber remains in the needle during the exposure of the SPME device to the sample. The coating works as a trap for analytes that diffuse into the needle, resulting in an integral of concentration over time measurement.

SPME sampling can be performed in three basic modes: direct extraction, headspace extraction, and extraction with membrane protection. Figure 2 illustrates the differences between these modes. In direct extraction mode (Figure 2a), the coated fiber is inserted into the sample and the analytes are transported directly from the sample matrix to the extracting phase. To facilitate rapid extraction, some level of agitation is required to transport the analytes from the bulk of the sample to the vicinity of the fiber. For gaseous samples, the natural flow of air (e.g. convection) is frequently sufficient to facilitate rapid equilibration for volatile analytes, but for aqueous matrices, more efficient agitation techniques, such as fast sample flow, rapid fiber or vial movement, stirring or sonication are required to reduce the effect of the ''depletion zone'' produced close to the fiber as a result of slow diffusional analyte transport through the stationary layer of liquid surrounding the fiber.

In the headspace mode (Figure 2b), the analytes are extracted from the gas phase equilibrated with the sample. The primary reason for this modification is to protect the fiber from adverse effects caused by non-volatile, high-molecular-weight substances present in the sample matrix (e.g. humic acids or proteins). The headspace mode also allows matrix modifications, including pH adjustment, without affecting the fiber. In a system consisting of a liquid sample and its headspace, the amount of an analyte extracted by the fiber coating does not depend on the location of the fiber, in the liquid phase or in the gas phase, therefore the sensitivity of headspace sampling is the same as the sensitivity of direct sampling as long as the volumes of the two phases are the same in both sampling modes. Even when no headspace is used in direct extraction, a significant sensitivity difference between direct and headspace sampling can occur only for very volatile analytes. However, the choice of sampling mode

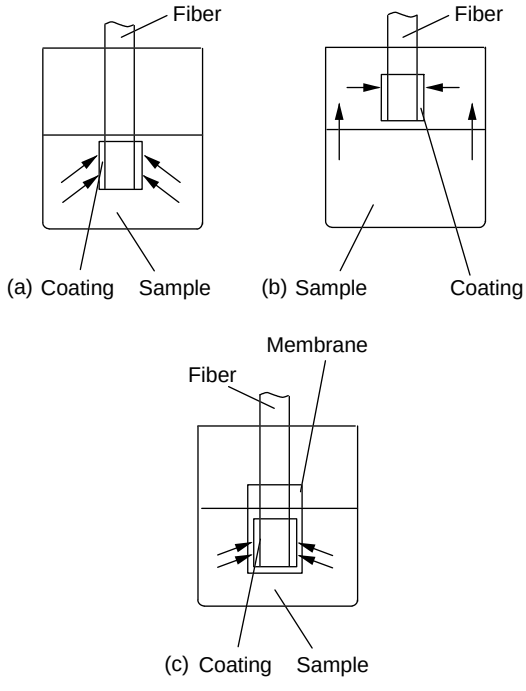

Figure 2 Modes of SPME operation: direct extraction (a), headspace extraction (b) and membrane-protected SPME (c). (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

has a very significant impact on the extraction kinetics. When the fiber is in the headspace, the analytes are removed from the headspace first, followed by indirect extraction from the matrix. Therefore, volatile analytes are extracted faster than semivolatiles. Temperature has a significant effect on the kinetics of the process, since it determines the vapor pressure of analytes. In general, the equilibration times for volatile compounds are shorter for headspace SPME extraction than for direct extraction under similar agitation conditions, for the following three reasons:

1. a substantial portion of the analytes is present in the headspace prior to the beginning of the extraction process;
2. there is typically a large interface between sample matrix and headspace; and
3. the diffusion coefficients in the gas phase are typically higher by four orders of magnitude than in liquids.

The concentration of semivolatile compounds in the gaseous phase at room temperature is small, so headspace extraction rates for those compounds are substantially lower. They can be improved by using very efficient agitation or by increasing the extraction temperature.

In the third mode (SPME extraction with membrane protection, Figure 2c), the fiber is separated from the sample with a selective membrane, which lets the analytes through while blocking the interferences. The main purpose of the membrane barrier is to protect the fiber against adverse effects caused by high-molecular-weight compounds when very dirty samples are analyzed. While extraction from headspace serves the same purpose, membrane protection enables the analysis of less volatile compounds. The extraction process is substantially slower than direct extraction because the analytes need to diffuse through the membrane before they can reach the coating. Use of thin membranes and increase in extraction temperature result in shorter extraction times.

2 THEORETICAL ASPECTS OF SOLID-PHASE MICROEXTRACTION OPTIMIZATION AND CALIBRATION

2.1 Thermodynamics

SPME is a multiphase equilibration process. Frequently, the extraction system is complex, as in a sample consisting of an aqueous phase with suspended solid particles having various adsorption interactions with analytes, plus a gaseous headspace. In some cases specific factors have to be considered, such as analyte losses by biodegradation or adsorption on the walls of the sampling vessel. In the discussion below we will only consider three phases: the fiber coating, the gas phase or headspace and a homogeneous matrix such as pure water or air. During extraction, analytes migrate between all three phases until equilibrium is reached.

The mass of an analyte extracted by the liquid polymeric coating is related to the overall equilibrium of the analyte in the three-phase system. Since the total mass of an analyte should remain constant during the extraction, we have Equation (1):

$$C_0 V_s = C_f^\infty V_f + C_h^\infty V_h + C_s^\infty V_s \tag{1}$$

where C_0 is the initial concentration of the analyte in the matrix; C_f^∞, C_h^∞, and C_s^∞ are the equilibrium concentrations of the analyte in the coating, the headspace, and the matrix, respectively; V_f, V_h, and V_s are the volumes of the coating, the headspace, and the matrix, respectively. If we define the coating/gas distribution constant as $K_{fh} = C_f^\infty / C_h^\infty$, and the gas/sample matrix distribution constant as $K_{hs} = C_h^\infty / C_s^\infty$, the mass of the analyte absorbed by the coating, $n = C_f^\infty V_f$, can be expressed as Equation (2):

$$n = \frac{K_{fh} K_{hs} V_f C_0 V_s}{K_{fh} K_{hs} V_f + K_{hs} V_h + V_s} \tag{2}$$

Also (Equation 3):

$$K_{fs} = K_{fh} K_{hs} = K_{fg} K_{gs} \tag{3}$$

since the fiber/headspace distribution constant, K_{fh} can be approximated by the fiber/gas distribution constant K_{fg}, and the headspace/sample distribution constant, K_{hs}, by the gas/sample distribution constant, K_{gs}, if the effect of moisture in the gaseous headspace can be neglected. Thus, Equation (2) can be rewritten as Equation (4):

$$n = \frac{K_{fs} V_f C_0 V_s}{K_{fs} V_f + K_{hs} V_h + V_s} \tag{4}$$

Equation (4) states, as expected from the equilibrium conditions, that the amount of analyte extracted is independent of the location of the fiber in the system. It may be placed in the headspace or directly in the sample as long as the volumes of the fiber coating, headspace and sample are kept constant. There are three terms in the denominator of Equation (4) which give measures of the analyte capacity of each of the three phases: fiber ($K_{fs} V_f$), headspace ($K_{hs} V_h$), and the sample itself (V_s). If we assume that the vial containing the sample is completely filled (no headspace), the term $K_{hs} V_h$ in the denominator, which is related to the capacity ($C_h^\infty V_h$) of the headspace, can be eliminated, resulting in Equation (5):

$$n = \frac{K_{fs} V_f C_0 V_s}{K_{fs} V_f + V_s} \tag{5}$$

Equation (5) describes the mass absorbed by the polymeric coating after equilibrium has been reached in the system. In most of determinations, K_{fs} is relatively small compared to the phase ratio of sample matrix to coating volume ($V_f \ll V_s$). In that situation the capacity of the sample is much larger compared to capacity of the fiber resulting in a very simple relationship (Equation 6):

$$n = K_{fs} V_f C_0 \tag{6}$$

Equation (6) emphasizes the field sampling capability of the SPME technique. It is not necessary to sample a well-defined volume of the matrix since the amount of analyte extracted is independent of V_s as long as $K_{fs} V_f \ll V_s$. The SPME device can be placed directly in contact with the investigated system to allow quantitation.

Strictly speaking, the above discussion (Equations 4–6) is limited to the partitioning equilibrium involving liquid polymeric phases such as PDMS. The method of analysis for solid sorbent coatings is analogous for low analyte concentration, since the total surface area available for adsorption is proportional to the coating volume if we assume constant porosity of the sorbent. For high analyte concentration the saturation of the surface can occur resulting in nonlinear isotherms. Similarly high concentration of a competitive interference compound can displace the target analyte from the surface of the sorbent. The simplest way to consider these high-concentration effects is to replace the volume of the fiber coating, V_f in the above equations as a measure of the total fiber surface area, by a fraction of the original coating volume corresponding to a free surface area available for adsorption.

2.1.1 *Prediction of Distribution Constants*

In many cases, the distribution constants present in Equations (4–6) which determine the sensitivity of SPME extraction can be estimated from physicochemical data and chromatographic parameters. This approach eliminates the need for calibration. For example, distribution constants between a fiber coating and gaseous matrix (e.g. air) can be estimated using isothermal GC retention times on a column with a stationary phase identical to the fiber coating material.[4] This is possible because the partitioning process in GC is analogous to the partitioning process in SPME, and there is a well-defined relationship between the distribution constant and the retention time. The nature of the gaseous phase does not affect the distribution constant, unless the components of the gas, such as moisture, swell the polymer, thus changing its properties. A most useful method for determining coating-to-gas distribution constants uses the linear temperature programmed retention index (LTPRI) system, which indexes compounds' retention times relative to the retention times of *n*-alkanes. This system is applicable to retention times for temperature programmed gas–liquid chromatography. The logarithm of the coating-to-air distribution constants of *n*-alkanes can be expressed as a linear function of their LTPRI values. For PDMS, this relationship is $\log K_{fg} = 0.00415 \times \text{LTPRI} - 0.188$.[5] Thus, the LTPRI system permits interpolation of the K_{fg} values from the plot of $\log K_{fg}$ versus retention index. The LTPRI values for many compounds are available in the literature, hence this method allows estimation of K_{fg} values without experimentation. If the LTPRI value for a compound is not available from published sources, it can be determined from a GC run. Note that the GC column used to determine LTPRI should be coated with the same material as the fiber coating.

Estimation of the coating/water distribution constant can be performed using Equation (3).[6] The appropriate coating/gas distribution constant can be found by applying techniques discussed above, and the gas/water distribution constant (Henry's constant) can be obtained from physicochemical tables or can be estimated by the structural unit contribution method.[7]

Some correlations can be used to anticipate trends in SPME coating/water distribution constants for analytes. For example, a number of investigators have reported the correlation between octanol/water distribution constant K_{ow} and K_{fw}. This is expected, since K_{ow} is a very general measure of the affinity of compounds for the organic phase. It should be remembered, however, that the trends are valid only for compounds within homologous series, such as aliphatic hydrocarbons, aromatic hydrocarbons or phenols; they should not be used to make comparisons between different classes of compounds, because of different analyte activity coefficients in the polymer.

2.1.2 *Effect of Extraction Parameters*

Thermodynamic theory predicts the effects of modifying certain extraction conditions on partitioning and indicates parameters to control for reproducibility. The theory can be used to optimize the extraction conditions with a minimum number of experiments and to correct for variations in extraction conditions, without the need to repeat calibration tests under the new conditions. For example, SPME analysis of outdoor air may be done at ambient temperatures that can vary significantly. The relationship that predicts the effect of temperature on the amount of analyte extracted allows calibration without the need for extensive experimentation[8] Extraction conditions that affect K_{fs} include temperature, salting, pH and organic solvent content in water. A brief discussion about the use of extraction parameters in SPME method optimization can be found in the next section.

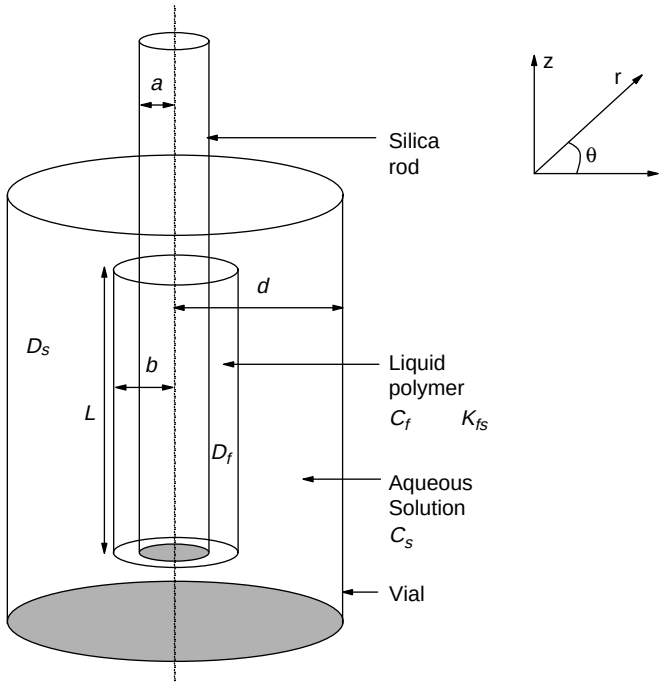

Figure 3 Graphic representation of the SPME/sample system configuration, with dimensions and parameters labeled as follows: a, fiber coating inner radius; b, fiber coating outer radius; L, fiber coating length; d, vial inner radius; C_f, analyte concentration in the fiber coating; D_f, analyte diffusion coefficient in the fiber coating; C_s, analyte concentration in the sample; D_s, analyte diffusion coefficient in the sample; K_{fs}, analyte distribution coefficient between fiber coating and sample; $K_{fs} = C_f/C_s$.

2.2 Kinetics

The kinetic theory is very useful in optimization of the extraction conditions by identifying ''bottlenecks'' of SPME and indicates strategies to increase extraction speed. In the discussion below we will limit our consideration to direct extraction (Figure 3).

2.2.1 Perfect Agitation

Let us first consider the case where the liquid or gaseous sample is perfectly agitated. In other words, the sample phase moves very rapidly with respect to the fiber, so that all the analytes present in the sample have access to the fiber coating. In this case, the equilibration time, defined as the time required to extract 95% of the equilibrium amount (Figure 4) of an analyte from the sample, corresponds to Equation (7):

$$t_e = t_{95\%} = \frac{2(b-a)^2}{D_f} \tag{7}$$

Using Equation (7) one can estimate the shortest equilibration time possible for the practical system by substituting appropriate data for the diffusion coefficient of an analyte in the coating (D_f) and the fiber coating thickness ($b - a$). For example, the equilibration time for the extraction of benzene from a perfectly stirred aqueous solution with a 100 μm PDMS film is expected to be about 20 s. Equilibration times close to those predicted for perfectly agitated samples have been obtained experimentally for extraction of analytes from air samples (because of high diffusion coefficients in gas) or when very high sonication power was used to facilitate mass transfer in aqueous samples. However, in practice, there is always a layer of unstirred water around the fiber. A higher stirring rate will result in a thinner water layer around the fiber.

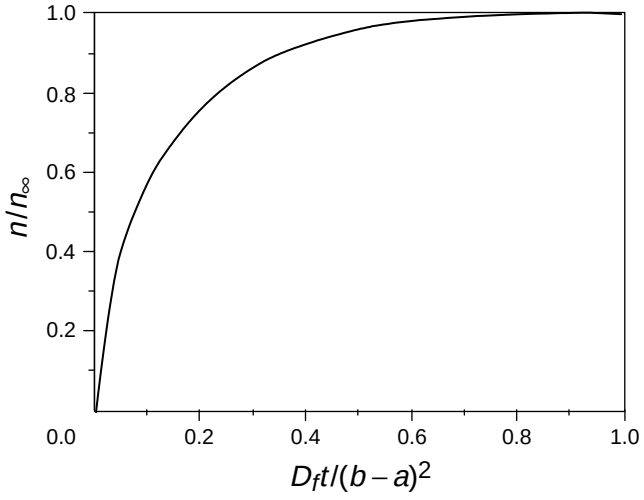

Figure 4 Mass absorbed versus time from perfectly agitated solution of infinite volume.

2.2.2 Practical Agitation

Independent of the agitation level, fluid contacting a fiber's surface is always stationary, and as the distance from the fiber surface increases, the fluid movement gradually increases until it corresponds to bulk flow in the sample. To model mass transport, the gradation in fluid motion and convection of molecules in the space surrounding the fiber surface can be simplified by a zone of a defined thickness in which no convection occurs, and perfect agitation in the bulk of the fluid everywhere else. This static layer zone is called Prandtl boundary layer (see Figure 5).[9] Its thickness, δ, is determined by the agitation conditions and the viscosity of the fluid.

The equilibration time can be estimated for practical cases from Equation (8):

$$t_e = t_{95\%} = 3\frac{\delta K_{fs}(b-a)}{D_s} \tag{8}$$

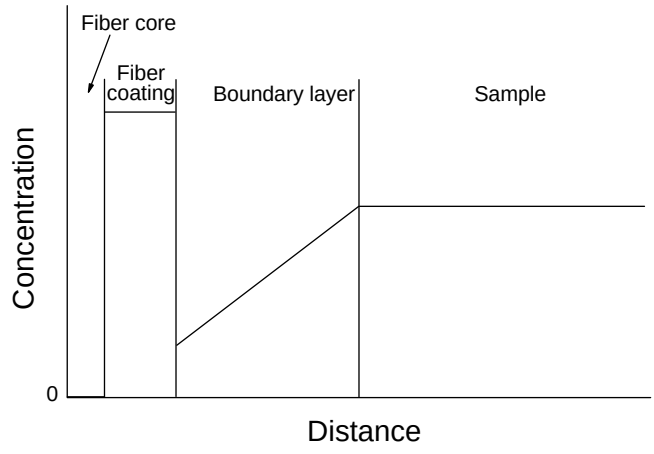

Figure 5 Boundary layer model configuration showing the different regions considered and the assumed concentration versus radius profile for the case when the boundary layer determines the extraction rate.

where $(b - a)$ is the fiber coating's thickness, D_s is the analyte's diffusion coefficient in the sample fluid and K_{fs} is the analyte's distribution constant between the fiber and the sample. Equation (8) can be used to predict equilibration times when the extraction rate is controlled by the diffusion in the boundary layer. In other words, the extraction time calculated by using Equation (8) must be longer than the corresponding time predicted by Equation (7).

2.2.3 Time-average Sampling Using in Tube Solid Phase Microextraction

In addition to the analyte concentration measurement at a well-defined place in space and time, obtained by using the approaches discussed above, an integrating sampling is possible with the SPME device. This is particularly important in field measurements when changes of analyte concentration over time, and place-to-place variations, must often be taken into account.

When the extracting phase is not exposed directly to the sample, but is contained in the protective tubing (needle) without any flow of the sample through it, the extraction occurs by the process of diffusion through the static gas phase present in the needle. The integrating system can consist of an externally coated fiber withdrawn into the needle. This simple geometric arrangement represents a very convenient method which is able to generate a response proportional to the integral of the analyte concentration over time and/or space (when the needle is moved through the space). The only mechanism of analyte transport to the extracting phase is diffusion through the gaseous phase contained in the tubing. During this process, a linear concentration profile is established in the tubing between the small needle opening, characterized by surface area A, and the distance Z between the needle opening and the position of the extracting phase. The amount of analyte extracted, dn, during time interval, dt, can be calculated by considering Fick's first law of diffusion (Equation 9):[4]

$$\mathrm{d}n = AD_g \frac{\mathrm{d}C}{\mathrm{d}Z}\,\mathrm{d}t = AD_g \frac{\Delta C(t)}{Z}\,\mathrm{d}t \tag{9}$$

where $\Delta C(t)/Z$ is a value of the gradient established in the needle between the needle opening and the position of the extracting phase, Z; $\Delta C(t) = C(t) - C_Z$, where $C(t)$ is a time dependent concentration of analyte in the sample in the vicinity of the needle opening, and C_Z the concentration of the analyte in the gas phase in the vicinity of the coating. If C_Z is close to zero for a high coating/gas distribution constant capacity, then: $\Delta C(t) = C(t)$. The concentration of analyte at the coating position in the needle, C_Z, will increase with integration time, but it will be kept low compared to the sample concentration because of the presence of the sorbing coating. Therefore the accumulated amount over time can be calculated by Equation (10):

$$n = D_g \frac{A}{Z} \quad C(t)\,\mathrm{d}t \tag{10}$$

As expected, the extracted amount of analyte is proportional to the integral of a sample concentration over time, the diffusion coefficient of analytes in the gaseous phase, D_g, the area of the needle opening, A, and inversely proportional to the distance of the coating position in respect of the needle opening, Z. It should be emphasized that this application of the SPME device is not an equilibrium measurement. Equation (10) is valid only in a situation where the amount of analyte extracted onto the sorbent is a small fraction (below relative standard deviation (RSD) of the measurement, typically 5%) of the equilibrium amount in respect to the lowest concentration in the sample. To extend integration times, the coating can be placed deeper into the needle (larger Z), the opening of the needle can be reduced by placing an additional orifice (smaller A), or a high-capacity sorbent can be used. The first two solutions will result in a low measurement sensitivity. An increase of sorbent capacity presents a more attractive opportunity. It can be achieved by increasing either the volume of the coating, or its affinity towards the analyte. An increase of the coating volume will require an increase in the size of the device. The optimum approach to increased integration time, is to use sorbents characterized by large coating/gas distribution constants, or derivatization reagent in the sorbent.

3 METHOD DEVELOPMENT

Developing a new SPME method in most cases requires the following steps:

- Selection of fiber coating
- Selection of the derivatization reagent, if required
- Selection of the extraction mode
- Selection of the agitation method
- Selection of separation and/or detection technique
- Optimization of desorption conditions
- Optimization of sample volume
- Determination of the extraction time profile in a pure matrix
- Determination of extraction time
- Calculation of the distribution constant
- Optimization of extraction conditions (pH, salt, temperature)
- Determination of the linear range for pure matrix at optimum extraction conditions
- Selection of the calibration method
- Optimization of the extraction conditions for heterogeneous samples
- Verification of the equilibration time, sensitivity and linear dynamic range for complex samples
- Determination of method precision
- Determination of method detection limit (DL)
- Validation of the method
- Automation of the method

In most cases, not all the steps have to be performed. Knowledge gained from previous experiments, as well as from literature, can often be applied to the problem at hand. Most SPME methods developed to date are used in combination with GC separation and an appropriate detection method. Use in conjunction with HPLC and other techniques should be also considered. Following is the discussion of the particular steps in method development, with main focus on aqueous samples.

3.1 Selection of Fiber Coating

The following coatings are currently (1997) commercially available from Supelco, Inc. (Bellefonte, PA): PDMS of three thicknesses: 7, 30 and 100 μm; 85 μm PA; 65 μm poly(dimethylsiloxane)/divinylbenzene (PDMS/DVB); 60 μm PDMS/DVB optimized for HPLC; 65 μm poly(ethylene glycol)/divinylbenzene (Carbowax™/DVB); 65 μm poly(ethylene glycol)/DVB template resin (Carbowax™/TR); and 80 μm PDMS/Carboxen. The PDMS and the PA coatings, still the most widely used, extract analytes by absorption. PDMS is a very viscous liquid, while PA is a low-density solid polymer at room temperature, which allows analytes to diffuse into the coating, but the diffusion coefficients are lower than for PDMS. The remaining coatings are porous polymers, in which DVB or Carboxen porous microspheres are immobilized onto the fiber by using either Carbowax™ or PDMS as the glue to hold them together. These coatings extract analytes by adsorption rather than absorption. The pores in the template DVB polymer are uniform, resulting in less adsorption discrimination as a function of analyte molecular weight.[10]

Figure 6 presents general guidelines for choosing a coating for a given application. Typically, the chemical nature of the target analyte determines the type of coating used. A simple general rule, ''like dissolves like'', applies very well for the liquid coatings. Porous polymer coatings are mostly nonselective. Selection of the coating is based primarily on the polarity and volatility characteristics of the analyte. For example, according to Figure 6, a 100 μm PDMS coating should be used for compounds of high to medium volatility, and low to medium polarity. In general, PDMS is the most useful coating and should be used whenever possible. It is very rugged and able to withstand high injector temperatures, up to about 300 °C. PDMS is a nonpolar phase, thus it extracts nonpolar analytes very well. However, it can also be applied successfully to more polar compounds, particularly after optimizing extraction conditions. An additional advantage of this phase is the possibility of estimating the distribution constants

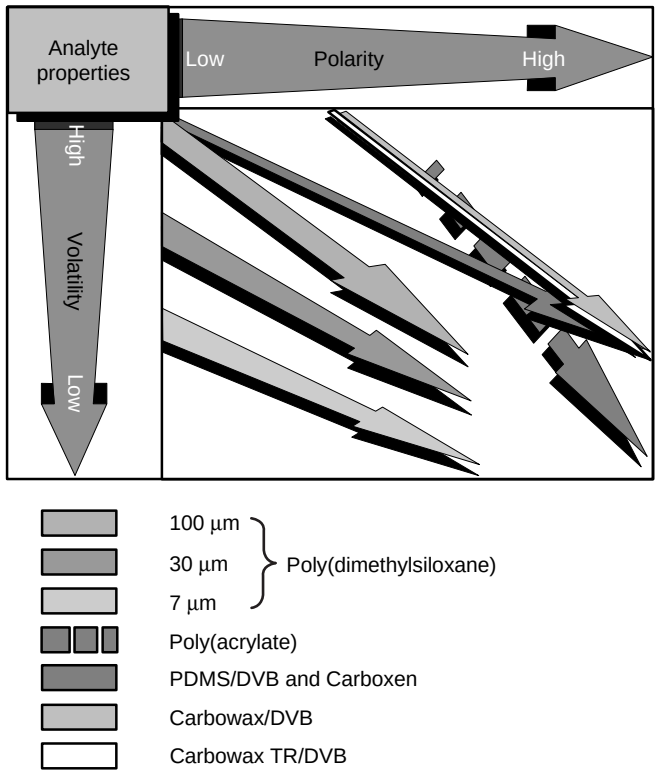

Figure 6 Coating selection guide.

for organic compounds from retention parameters on PDMS-coated GC columns (see section 2).

Both the coating thickness and the distribution constant determine the sensitivity of the method and the extraction time. Thick coatings offer increased sensitivity, but require much longer equilibration times. As a general rule, to speed up the sampling process, the thinnest coating offering the sensitivity required should be used. For 100 μm PDMS fibers, the equilibration time in direct aqueous extraction with magnetic stirring is less than 1 hr for compounds characterized by distribution constants, K, lower than ~10 000. Thinner PDMS coatings should be considered for compounds of larger K. For more polar compounds, such as phenols,[11] the PA coating is more suitable.

Porous polymer coatings have complementary properties compared to PDMS and PA. They are more suitable for volatile species, the respective distribution ratios typically being larger than for PDMS. Changing the 'glue' from PDMS to Carbowax™ results in different selectivity towards polar compounds, such as ketones and alcohols. The adsorption times are typically shorter for gaseous samples compared to 100 μm PDMS, since the analytes do not need to diffuse through the liquid polymeric phase.

The TR coating is designed to reduce molecular-weight discrimination for homologs differing in chain length. With regular DVB coatings characterized by nonuniform pore size distribution, the amount of analyte extracted from a solution of a given concentration may vary as a function of the size of a molecule. The TR has uniform pore diameters, therefore it exhibits uniform extraction efficiency.

When a group of analytes of different characteristics (e.g. pesticides) is to be determined by SPME, primary consideration should be given to the analytes that are the most difficult to extract. A coating should be chosen that enables the determination of all the analytes with enough sensitivity. When none of the commercially available coatings meets the requirements, custom coatings can sometimes be prepared. For example, Figure 7 shows the separation of alcohols extracted from unleaded gasoline using a fiber coated with commercially available Nafion™ perfluorinated resin.[12] The SPME device also facilitates microsolvent extraction. For example, porous coatings can be used to support organic solvent by placing the fiber into the solvent prior to extraction.[13] Solvent located in the pores extracts analytes from the matrix, when the fiber is placed in contact with a sample.

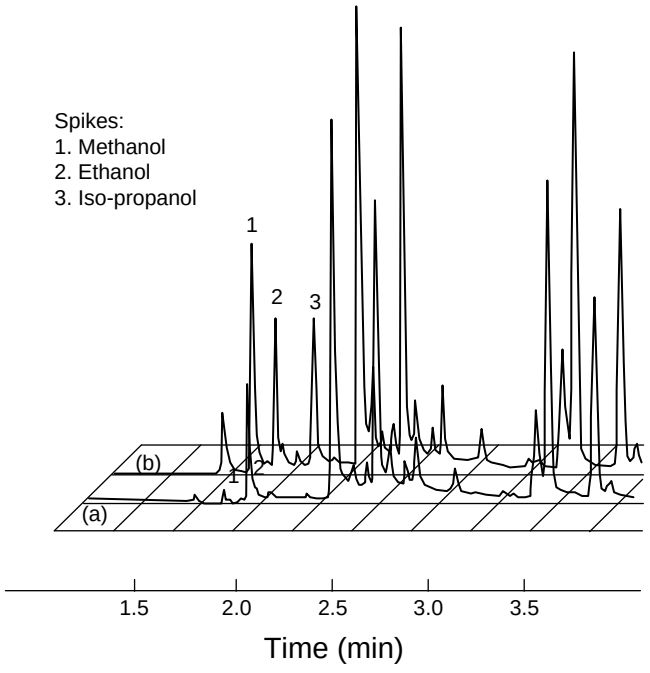

Figure 7 Separation of alcohols extracted from unleaded gasoline using a fiber coated with Nafion™ perfluorinated resin: (a) chromatogram of unleaded gasoline, (b) chromatogram of unleaded gasoline spiked with 333 ppm each of methanol, ethanol and iso-propanol. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

3.2 Selection of the Derivatization Reagent

Derivatization performed before and/or during extraction can enhance sensitivity and selectivity of both extraction and detection, as well as enabling SPME determination of analytes normally not amenable to analysis by this method, while post-extraction methods can only improve chromatographic behavior and detection. Incorporation of the derivatization step complicates the SPME procedure and therefore should only be considered when necessary. Selective reactions producing specific analogs result in less interference during quantitation. This approach can be used for analyte determination in complex matrices. Sensitivity enhancement can be achieved when the derivatizing reagent contains moieties that enhance detection. For example, the sensitivity of SPME determination of carboxylic acids in water can be greatly improved by using 1-(pentafluorophenyl)diazoethane as the derivatizing reagent.[14] This compound converts carboxylic acids directly in the aqueous matrix to pentafluorophenyl ester derivatives which can be detected by electron capture detection (ECD) with very good sensitivity. Method selectivity is also greatly improved in this way.

In many cases, derivatization reagents designed specifically for liquid extraction procedures can be used in SPME. For example, sodium tetraethylborate was used to enable the analysis of metal ions, including mercury and lead, in water, as well as to perform partial speciation of organometallic compounds by SPME/GC.[15,16]

The most interesting implementation of the derivatization procedure involves the use of doped fibers.

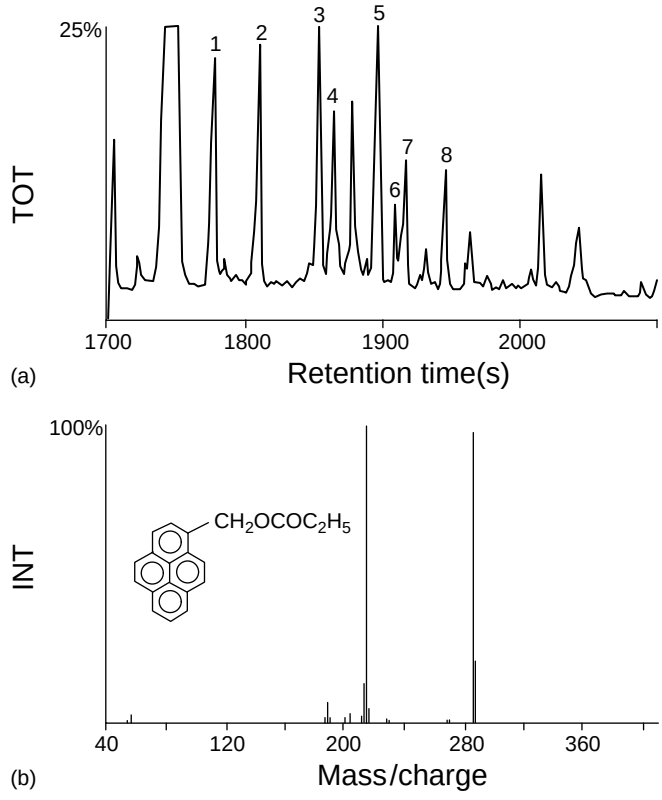

Figure 8 (a) Reconstructed gas chromatography/mass spectrometry (GC/MS) chromatogram illustrating the presence of short-chain fatty acids in a real sewage sample (SPME extraction with PA-coated fiber). Peak assignment: 1, acetic; 2, propionic; 3, isobutyric; 4, butyric; 5, pivalic; 6, isovaleric; 7, valeric; and 8, hexanoic acids. The peaks shown correspond to pyrenylmethyl esters of these acids; (b) An example of the mass fragmentogram and the structure of propionic acid/1-pyrenyldiazomethane (PDAM) ester. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

Figure 8 presents a chromatogram obtained by exposing a PA-coated fiber doped with PDAM to the headspace of a sewage sample.[17] All volatile carboxylic acids are detected. Without derivatization, SPME is not capable of extracting sufficient amounts of these acids because no appropriate selective coating exists as of yet. Such an approach results in high sensitivity and is compatible with field analysis requirements, since the derivatizing reagent and the derivatization products are nonvolatile. The doped fibers not only facilitate spot sampling, but can also measure long-term exposure to a given contaminant, since analytes are continuously accumulated onto the fiber until all the reagent is consumed.

3.3 Selection of the Extraction Mode

Extraction mode selection is based on the sample matrix composition, analyte volatility, and its affinity to the matrix. For very dirty samples the headspace or fiber protection mode should be selected. For clean matrices, both direct and headspace sampling can be used. The latter is applicable for analytes of medium to high volatility. Headspace extraction is always preferential for volatile analytes because the equilibration times are shorter in this mode compared to direct extraction. Extraction conditions for many compounds, including polar and ionic ones, can be improved by matrix modifications, as described later. Application of headspace SPME can be extended to semivolatile compounds and analytes strongly bound to the matrix by increasing the extraction temperature, also discussed below. Fiber protection should be used only for very dirty samples in cases where neither of the first two modes can be applied. General sampling mode guidelines are presented in Table 1.

3.4 Selection of the Agitation Technique

Equilibration times in gaseous samples are short and frequently limited only by the rate of diffusion of the analytes in the coating. A similar situation occurs when analytes characterized by large air/water distribution constants are determined in water by the headspace technique. When the aqueous and gaseous phases are at equilibrium prior to the beginning of the sampling process, most of the analytes are in the headspace. As a result, the extraction times are short even when no agitation is used. However, for aqueous samples, agitation is required in most cases to facilitate mass transport between the bulk of the aqueous sample and the fiber. Table 2 summarizes the properties of several agitation methods which have been tested with SPME.

Care must be taken when using magnetic stirring to ensure that the rotational speed of the stirring bar is constant and that the base plate does not change its temperature during stirring. This usually implies the use of high-quality digital stirrers. Alternately, with cheaper stirrers, the base plate should be thermally insulated from the vial containing the sample to eliminate variations in sample temperature during extraction. Magnetic stirring is efficient when fast rotational speeds are applied. Figure 9 illustrates the dependence of the equilibration

Table 1 Sampling mode selection criteria

Sampling mode	Analyte properties	Matrices
Direct	Medium to low volatility	Gaseous; liquid (preferably simple)
Headspace	High to medium volatility	Liquid (including complex); solid
Membrane protection	Low volatility	Complex samples

Table 2 Agitation methods in SPME

Method	Advantages	Disadvantages
Static (no agitation)	Simple, performs well for gaseous samples	Limited to volatile analytes and headspace SPME
Magnetic stirring	Common equipment, good performance	Require stirring bar in the vial
Intrusive stirring	Very good performance	Sealing the sample vial difficult
Vortex/moving vial	Good performance, no need for a stirring bar in the vial	Stress on needle and fiber
Fiber movement	Good performance, no need for a stirring bar in the vial	Stress on needle and fiber, limited to small sample volumes
Flow-through	Good agitation at rapid flows	Potential for cross-contamination, requires constant flow
Sonication	Very short extraction times	Noise, heating of the sample

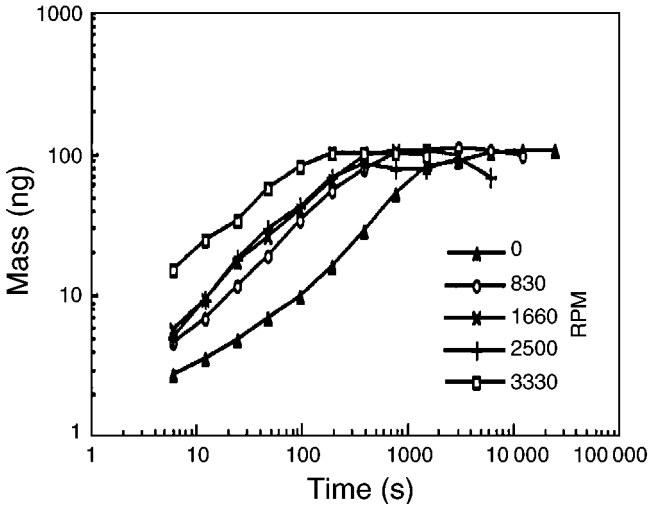

Figure 9 Extraction time profiles obtained for direct extraction with magnetic stirring of 100-ppb toluene solution using 56-μm PDMS coating.

time for toluene on the rotational speed. The equilibration time progressively decreases with stirring speed increase, and at 3330 rpm is only about 200 s. Intrusive stirring can improve the agitation further, but it requires a direct connection between the stirrer and the motor, which is difficult to seal.

Needle vibration technique uses an external motor and a cam to generate a vibrating motion of the fiber and the vial. This technique has been implemented by Varian in the SPME autosampler. In the Vortex technique the vial is moved rapidly in a circular motion. Both techniques generally provide good agitation, resulting in equilibration times similar to those obtained for magnetic stirring. However, for the needle vibration technique, good performance is limited to small vials and direct extraction mode. Both techniques can be conveniently applied to process a large number of samples since the sample vials do not require any manipulations, such as introduction of stirring bars.

Flow-through techniques are very useful in continuous monitoring applications and can also be automated. However, some additional flow-metering devices may be required to ensure reproducible mass-transfer conditions. The most efficient agitation method evaluated to date for SPME applications is the direct probe sonication, which can provide very short extraction times, frequently approaching the theoretical limits calculated for perfectly agitated samples. However, this technique has substantial drawbacks associated with large amounts of energy introduced into the system, which cause the sample to heat up and in some cases can lead to decomposition of the analyte. Most of these disadvantages are eliminated when sonication is applied in the flow-through mode.

3.5 Selection of Separation and/or Detection Technique

So far, most SPME applications have been developed for GC, but other separation techniques, including HPLC, capillary electrophoresis (CE) and supercritical fluid chromatography (SFC), can also be used in conjunction with this technique. The complexity of the extraction mixture determines the proper quantitation device. Regular chromatographic and CE detectors can normally be used for all but the most complex samples, for which mass spectrometry (MS) should be applied. As selective coatings become available, the direct coupling to tandem mass spectrometry (MS/MS) and inductively coupled plasma mass spectrometry (ICPMS) becomes practical.

3.6 Optimization of Desorption Conditions

Standard GC injectors, such as popular split-splitless types, are equipped with large volume inserts to accommodate the vapors of the solvent introduced during liquid injections. As a result, the linear flow rates of the carrier gas in those injectors are very low in splitless mode, and the transfer of the volatilized analytes onto the front of the GC column is very slow. No solvent is introduced

during SPME injection, therefore the large insert volume is unnecessary. Opening the split line during SPME injection is not practical, since it results in reduced sensitivity. Efficient desorption and rapid transfer of the analytes from the injector to the column require high linear flow rates of the carrier gas around the coating. This can be accomplished by reducing the internal diameter of the injector insert, matching it as closely as possible to the outside diameter of the coated fiber. Narrow-bore inserts for SPME are commercially available from Supelco for a range of GC instruments.

Other GC injectors, including programmed temperature vaporization (PTV) and septum-equipped programmable injector (SPI) (Varian), can be used with good results for SPME analysis. The use of on-column injectors is limited to those equipped with independent heating and capable of accommodating large diameter columns. Table 3 summarizes the characteristics of GC injectors suitable for SPME analysis.

Desorption time is determined by the temperature of the injector and the linear flow rate of the carrier gas around the fiber. Theoretical desorption times are very short since the diffusion coefficients of analytes in the coating increase and the gas/coating distribution constants rapidly decrease with temperature increase. Table 4 illustrates the decrease in the gas/coating distribution constant for three analytes when the temperature is increased from ambient to 250 °C. which is a typical injector temperature.

In practice, desorption temperature is determined by the thermal stability of the coating. It is advisable to use high desorption temperatures in order to achieve fast desorption. However, application of excessive temperatures adversely affects the coating's longevity and results in bleeding of the polymer, rendering the separation and quantitation difficult.

To achieve good separation efficiency, it is crucial that the injection band width is as small as possible. The few seconds needed to transfer the analytes onto the front of the column when narrow-bore inserts are used might be still too long to produce narrow bands for very volatile analytes. It is necessary, therefore, to refocus the band after the desorption is finished. A thick film column or cryofocusing can be used for this purpose. It is always advisable to consider the application of a short length of deactivated narrow-bore fused-silica tubing as a retention gap. This will eliminate the broadening effect created by the temperature gradient existing in the column close to the injector. Except for the most volatile analytes (gases), the combination of a thick film column and the retention gap is usually sufficient to refocus the injection band, and cryofocusing does not have to be applied.

Parameters which control the desorption process in the HPLC interface are analogous to those in GC applications. In addition to temperature and flow rate, the composition of the mobile phase also affects the process. In many cases, it is possible to use the mobile phase without any modifications, as in the analysis of polynuclear aromatic hydrocarbons (PAHs).[18] In some instances, addition of an appropriate solvent to the interface will assist desorption. The linear flow rate of the mobile phase should be maximized by choosing a small internal diameter tubing for the desorption chamber. Temperature also plays an important role in accelerating the desorption process as illustrated in Table 5. At ambient desorption temperatures, a significant carryover of Triton X-100 was observed, which was corrected when the temperature was raised to 135 °C.[19]

Table 3 GC injectors and their compatibility with SPME

Split/splitless	PTV	SPI
Can be used for SPME in splitless mode, low volume insert required	Can be used for SPME, low volume insert required	Low internal volume, best suitable for SPME

Table 4 Effect of temperature on gas/coating distribution constant, K_{fg}

Compound	K_{fg}	
	25 °C	250 °C
Benzene	300	0.4
Xylene	3000	1.4
Undecane	26 000	13

Table 5 Effect of desorption chamber temperature on carryover of Triton X-100 using Carbowax™/DVB fiber

Number of units in the ethoxylate chain	% Carryover	
	21 °C	135 °C
4	15	0
5	20	2
6	30	2
7	25	3
8	26	1
9	26	2
10	27	2
11	27	3
12	26	2
13	28	3
14	26	0

3.7 Optimization of Sample Volume

The volume of the sample should be based on the estimated distribution constant K_{fs}. The distribution constant can be estimated by using literature values for the target analyte or a related compound with the coating selected. For the PDMS coating, it can also be directly estimated from the chromatographic parameters.[4–6] In all other cases, K_{fs} has to be determined experimentally by equilibrating the sample with the fiber and determining the amount of analyte extracted by the coating. Care must be taken to avoid analyte losses via adsorption, evaporation, microbial degradation, etc., when very long extraction times are required to reach the equilibrium.

The sensitivity of the SPME method is proportional to the mass of the analyte, n, extracted from the sample and, for the direct extraction method, is given by Equation (11):

$$n = \frac{K_{fh} V_f V_s C_0}{V_s + K_{fh} V_f} \tag{11}$$

As the sample volume V_s increases, so does the amount of analyte extracted until the volume of the sample becomes significantly larger than the product of the distribution constant and volume of the coating (fiber capacity $K_{fs} \ll V_s$). At this point, the sensitivity of the method does not increase with further increase in volume. In practice, the limiting volume can be calculated from the dependence in Equation (12):

$$V_s = \frac{100 K_{fs} V_f}{E} \tag{12}$$

where E is the magnitude of the relative measurement error (%). For example, for a 5% error, it can be estimated that $V_s = 20 K_{fs} V_f$. This means that for K_{fs} values of up to about 200 and a 100 µm coating, a 2 mL vial is sufficient to give maximum sensitivity, while a 40 mL vial can be used for K_{fs} values smaller than 4000, etc. Using sample volumes larger than the limiting value does not only maximize the sensitivity, but also results in better precision since variations of the sample volume do not affect the results. Figure 10 illustrates this relationship graphically. The n_I corresponds to the amount of analyte extracted from an infinite sample volume. Even for a moderate distribution constant of 1000, characteristic for many volatiles, for the 100 µm PDMS coating (see Figure 10a) there is a substantial difference in the amount extracted when varying the sample volume within a few milliliters range. Increasing the volume from 1 mL to 10 mL results in sensitivity improvement by over 50%. The change is more dramatic for semivolatile analytes. The variation is less significant for thinner coatings (Figure 10b and c), but the sample volume should still be considered when analytes characterized by large

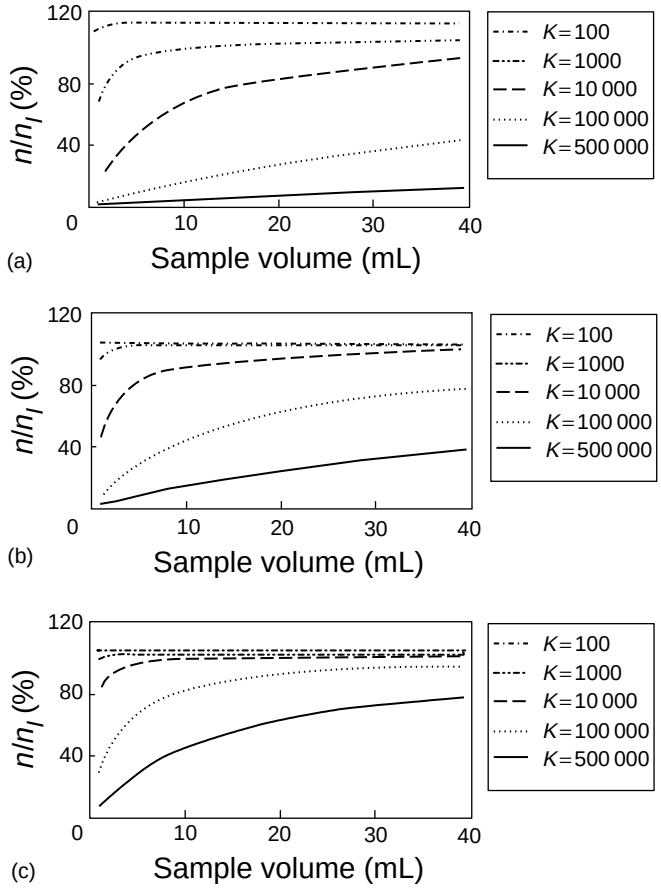

Figure 10 Dependence of n/n_I on V_s for (a) 100-µm fiber ($V_f = 0.65$ µL), (b) 30-µm fiber ($V_f = 0.14$ µL), and (c) 7-µm fiber ($V_f = 0.028$ µL) for direct SPME from small volume samples (two-phase system).

distribution constants, such as PAHs, are analyzed. In practice, the sample size is determined by the available vials. The discussion above should assist in the choice of the most appropriate vial size.

In headspace SPME, the analytes partition to the gaseous phase as well as to the coating. Very volatile compounds preferentially accumulate in the headspace resulting in a very substantial loss in sensitivity when the headspace volume is large. Figure 11 presents the dependence of the amount of analyte extracted by a 100 µm fiber on the headspace-to-sample volume ratio for three vial sizes: 4, 15 and 40 mL. The following values of distribution constants have been used: $K_{fs} = 400$, 2500 and 1000, and $K_{hs} = 0.15$, 0.7 and 1.24, corresponding to chloroform, 1,1,1-trichloroethane and carbon tetrachloride, respectively. In Figure 11, n_I is the amount of analyte extracted from an infinite sample volume. The n/n_I ratio is always the largest for chloroform, which has the lowest K_{hs} value, even though its K_{fs} value is also the lowest. This is because only a small fraction of the analyte is present in the sample

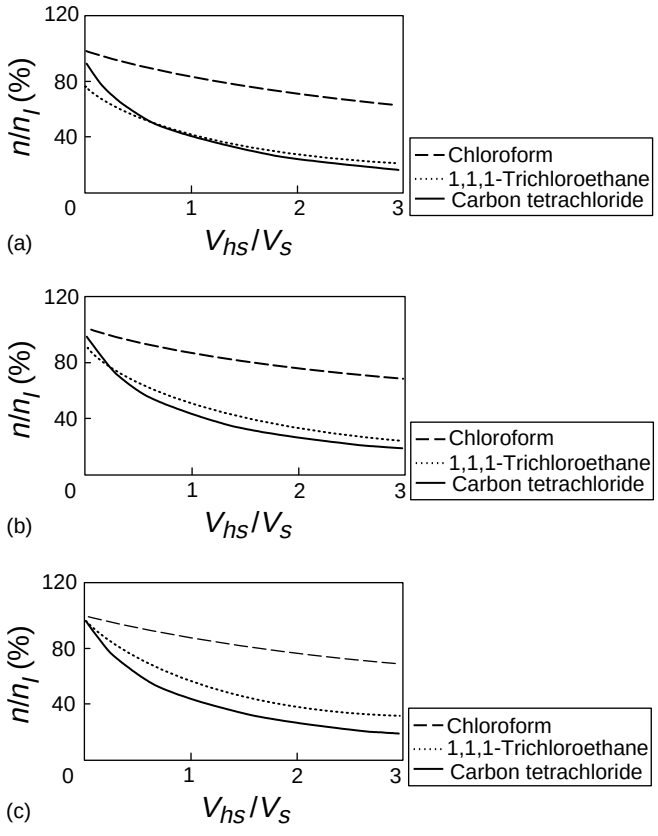

Figure 11 The effect of headspace volume on the amount of analyte extracted in a system of a constant volume of (a) 4, (b) 15 and (c) 40 mL by a 100-μm PDMS fiber for chloroform, 1,1,1-trichloroethane and carbon tetrachloride.

headspace, therefore the equilibrium concentration of the analyte in the sample remains relatively high. For the two other curves and low headspace volumes, n/n_I is higher for carbon tetrachloride than for 1,1,1-trichloroethane, which is consistent with the values of their distribution constants K_{fs}. However, as carbon tetrachloride has the largest K_{hs}, n/n_I drops faster for this compound than for 1,1,1-trichloroethane, as a result of which the respective two lines cross each other. The biggest relative increase in n/n_I when moving from small to large vials (~44%) is observed for 1,1,1-trichloroethane, characterized by the highest K_{fs} value. For compounds with lower K_{fs} values the relative increase in the amount extracted is much lower (~16 and ~12% for chloroform and carbon tetrachloride, respectively). It is therefore the combination of K_{fs} and K_{hs} for a given compound that determines the magnitude of the effect of sample volume on the amount of analyte extracted by the fiber in three-phase systems involving headspace.[20]

Headspace volume can have a significant effect on equilibration times (extraction kinetics). If the amount of the analyte extracted by the fiber at equilibrium is negligibly small compared to the amount present in the headspace equilibrated with the sample prior to the extraction, the analyte is extracted almost exclusively from the gaseous phase, and the process is much faster compared to the case when significant amounts of the analyte have to transported from the liquid sample to the fiber via the headspace. Assuming that this situation occurs when 95% of the analyte is extracted exclusively from the headspace, the criterion that must be fulfilled can be described by Equation (13). The assumption is reasonable, as a 5% difference usually falls within the limits of experimental error for trace SPME/GC analysis:

$$\frac{K_{fh}V_f}{K_{hs}V_h} < 0.05 \qquad (13)$$

The above criterion means that the capacity of the headspace ($K_{hs}V_h$) needs to be at least 20 times larger than the capacity of the fiber ($K_{fs}V_f$) to achieve rapid extraction. For a given sample volume V_s, this can be achieved by using a large enough headspace volume, or by increasing K_{hs}. The latter can be accomplished by increasing the temperature or by salting the analyte out of the liquid phase. When the criterion of Equation (13) is fulfilled, equilibration can take as little as a few minutes, and is almost independent of the agitation conditions (provided the analyte is equilibrated between the liquid phase and its headspace before the extraction begins). It should be emphasized, however, that the increase in headspace capacity causes a significant loss in sensitivity. In any case, it should be remembered that changing the vial size during the optimization process can affect not only the sensitivity, but also the equilibration time if the headspace/sample volume ratio is different in the new vial.

3.8 Determination of the Extraction Time Profile in a Pure Matrix

Extraction time profiles for the analyte should be determined first in a pure matrix, such as dry air, pure water or sand, to obtain basic understanding of the kinetics of analyte transfer from the matrix to the fiber under given mass-transfer conditions. The profile is determined by extracting samples of identical composition for progressively longer periods of time and plotting the resultant area counts vs. time. The mass of the analyte extracted can be quantified by a proper calibration procedure, e.g. syringe injection of a standard solution of the analyte in an appropriate solvent. In GC, this type of calibration is reliable only when non-vaporizing injectors are used (PTV, SPI, heated on-column).

3.9 Determination of the Extraction Time

An optimal approach to SPME analysis is to allow the analyte to reach equilibrium between the sample and the fiber coating. The equilibration time is defined as the time after which the amount of analyte extracted remains constant and corresponds within the limits of experimental error to the amount extracted after infinite time. Care should be taken when determining the equilibration times, since in some cases a substantial reduction of the slope of the curve might be wrongly taken as the point at which equilibrium is reached. Such a phenomenon often occurs in headspace SPME determinations of aqueous samples, where a rapid rise of the equilibration curve corresponding to extraction from the gaseous phase only is followed by a very slow increase related to analyte transfer from water through the headspace to the fiber. Figure 12 illustrates an example of such a curve for phenanthrene. The amount extracted increases rapidly in the first five minutes, but the equilibrium is not reached until much later. Determination of the amount extracted at equilibrium allows the calculation of the distribution constants.

When equilibration times are excessively long, shorter extraction times can be used. However, in such cases the extraction time has to be strictly controlled to assure good precision. Figure 13 presents equilibration time profiles for *p*,*p*′-1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (DDT) (a) and Dichlorvos (b). The equilibration time for DDT is about two hours. At equilibrium, small variations in the extraction time do not affect the amount of the analyte extracted by the fiber. On the other hand, at the steep part of the curve, even small variations in the extraction time may result in significant variations of the amount extracted. The relative error is the larger, the shorter is the extraction time. For a 10 min extraction (Figure 13a), the relative error caused by varying the extraction time by ±1 min is as high as 20%. It drops to 5% after 50 min extraction since more analyte accumulates onto the fiber. The timing is even more critical for short exposure times when the equilibration curve rises rapidly. Figure 13(b) shows an example when a 1 min deviation from 10 min sampling time results in 50% relative error of determination.

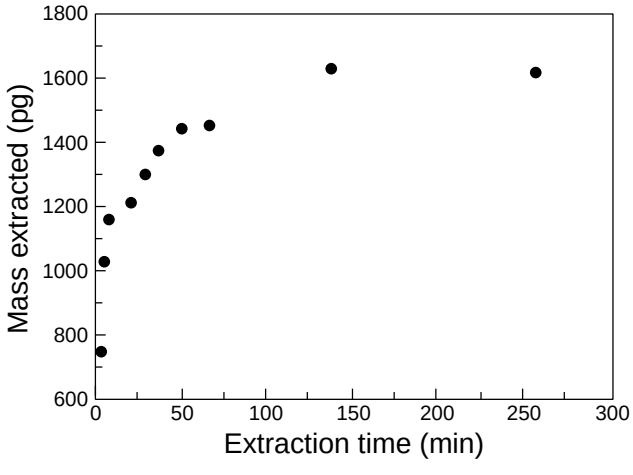

Figure 12 Extraction time profile corresponding to extraction of phenanthrene-d10 from water at 50 °C under magnetic stirring. Fiber coating: 7 μm PDMS.

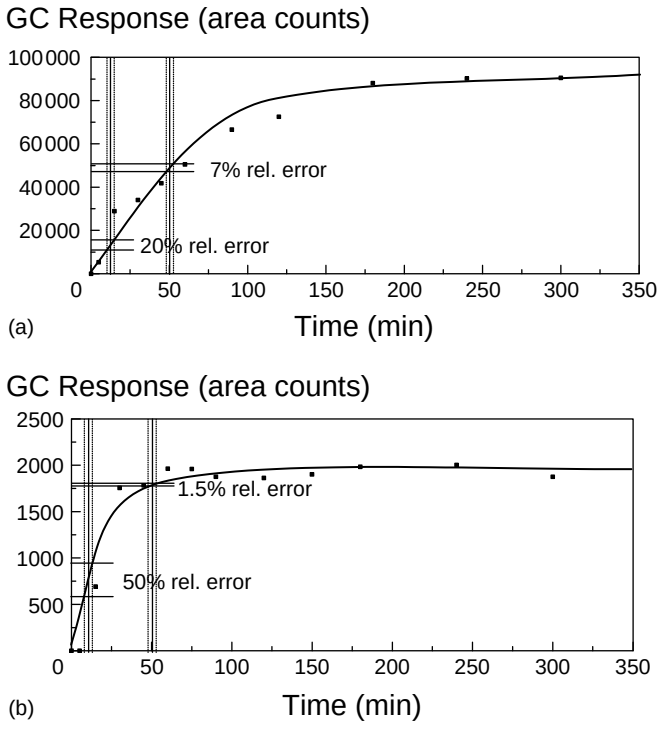

Figure 13 Selection of the extraction time based on extraction time profiles. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

Autosamplers can measure the time very precisely and the precision of analyte determination can be very good, even when equilibrium is not reached in the system. However, this requires the mass-transfer conditions and the temperature to remain constant during all experiments.

3.10 Calculation of the Distribution Constant

The distribution constant of the target analyte defines the sensitivity of the method. It is not necessary to calculate the fiber coating/sample matrix distribution constant, K_{fs}, when the calibration is based on isotopically labeled standards or standard addition, or when identical matrix and headspace volumes are used for the standard and for the sample with external calibration. However, it is always advisable to determine K_{fs} since this value gives more

Table 6 Ranges of K values obtained when the amount of the analyte extracted by the fiber, n, determined experimentally falls within ±5% (relative) of the true value, for two different fiber coating thickness and two sample volumes

100 μm			
2 mL sample		35 mL sample	
K	Range of K	K	Range of K
100	105–95	100	105–95
1000	1067–935	1000	1051–949
10 000	12 537–8172	10 000	10 598–9413
100 000	∞–36 190	100 000	115 748–86 928
1 000 000	∞–55 072	1 000 000	4 700 000–492 593
10 000 000	∞–58 104	10 000 000	∞–923 611
100 000 000	∞–58 426	100 000 000	∞–1 012 177
7 μm			
2 mL sample		35 mL sample	
K	Range of K	K	Range of K
100	105–95	100	105–95
1000	1051–949	1000	1050–950
10 000	10 574–9434	10 000	10 504–9496
100 000	112 903–88 785	100 000	105 422–94 622
1 000 000	3 500 000–558 824	1 000 000	1 093 750–913 462
10 000 000	∞–1 187 500	10 000 000	17 500 000–6 785 714
100 000 000	∞–1 338 028	100 000 000	∞–19 000 000

information about the experiment and aids optimization. K_{fs} can be used for calculation of the headspace volume, sample volume and coating thickness required to reach the desired sensitivity.

The distribution constant for the direct extraction mode can be calculated from the following dependence (Equation 14) obtained from Equation (11):

$$K_{fh} = \frac{nV_s}{V_f(C_0V_s - n)} \tag{14}$$

It is important to consider the sample volume when calculating the value of K_{fs}. Table 6 presents the ranges of K values that are obtained when the amount of the analyte extracted by the fiber, n, determined experimentally falls within ±5% (relative) of the true value, for two sample volumes (2 and 35 mL) and two fiber coating thicknesses (100 and 7 μm). A ±5% error can be assumed typical for trace analysis by SPME/GC. To correctly estimate the distribution constants over 1000 with 100 μm thick PDMS fiber, volumes of the sample larger than 2 mL need to be used. For very large K_{fs} values (>1 000 000), typical for high-molecular-weight hydrocarbons in water or air, no reliable estimates of K_{fs} can be obtained even with 35 mL sample when the 100 μm fiber is used, therefore only the 7 μm fiber should be used for this purpose.

If the distribution constant is low and the total volume of the sample is high, a good estimate of K_{fs} can be obtained from the relationship, Equation (15):

$$K_{fs} = \frac{n}{C_0V_f} \tag{15}$$

However, extreme caution is advised when using this equation, since it assumes that after the extraction, the analyte concentration in the sample does not change significantly. If this equation is used by mistake when the K_{fs} value is large or the sample volume is too small, the result of the calculation does not correspond to K_{fs}, but rather to the sample/coating phase volume ratio (V_s/V_f).

To determine the coating/matrix distribution constant when headspace is present in the vial, knowledge of the matrix/headspace distribution constant is required (Equation 16):

$$K_{fs} = \frac{n(K_{hs}V_h + V_s)}{V_f(C_0V_s - n)} \tag{16}$$

K_{fs} can be obtained for aqueous samples from tables of Henry's constants (K_H) (Equation 17):

$$K_{hs} = \frac{K_H}{RT} \tag{17}$$

For analytes with an unknown Henry's constant, Equation (18) can be used:

$$K_{fs} = \frac{nK_{fh}V_s}{K_{fh}V_fV_sC_0 - nK_{fh}V_f - nV_h} \tag{18}$$

Equation (18) is obtained by substituting $K_{hs} = K_{fs}/K_{fh}$ into Equation (16) and rearranging it. The fiber/gas distribution constant can be found by extracting target compounds from air mixtures by using a simple bulb experiment.[21] In addition, K_{fg} can be obtained directly from a chromatographic run when the stationary phase is made of the same material as the fiber coating. For compounds which have low vapor pressures, the amount of analyte present in the headspace is very small, especially if the volume of the headspace is kept to a minimum. Therefore, in this situation, the headspace volume can be neglected altogether and Equation (14), which corresponds to direct extraction, can be applied.

3.11 Optimization of Extraction Conditions

Extraction temperature increase causes an increase in the extraction rate, but simultaneously a decrease in the distribution constant. In general, if the extraction rate is of major concern, the highest temperature which still provides satisfactory sensitivity should be used. Figure 14 demonstrates the effect of temperature on extraction of methamphetamine from water. At room temperature equilibration takes several hours. When the temperature is raised to 60 °C, the equilibration time is less than 20 min for both compounds. When the temperature is further increased to 73 °C, the equilibration time is reduced to a few minutes (Figure 14a). Very short equilibration times indicate that the majority of the analytes extracted by the fiber originate from the headspace. This conclusion is further supported by the stirring data obtained for 60 °C (Figure 14b). At this temperature, the amount of analytes extracted is almost independent of the agitation conditions. The amount extracted is lower by over an order of magnitude compared to room temperature, but still provides sufficient sensitivity to quantify these drugs at the parts per billion level, which is adequate for screening applications.

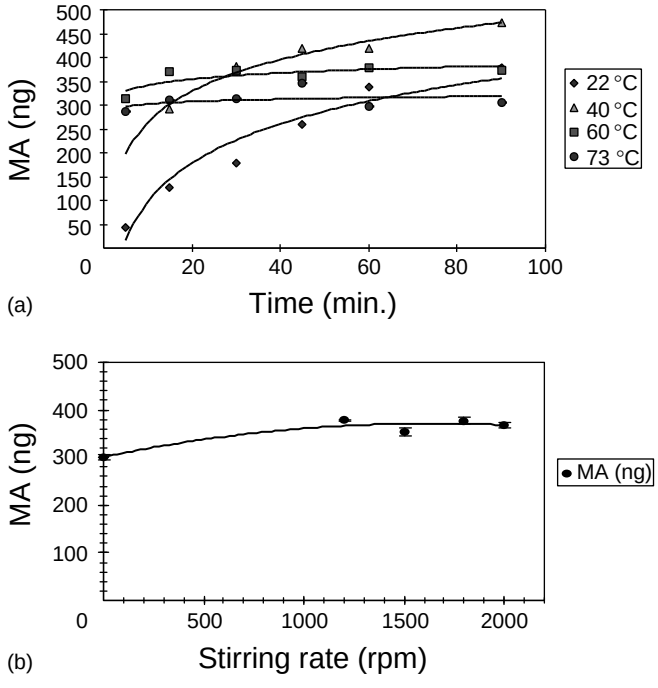

Figure 14 (a) Temperature dependence of the absorption time profile obtained for methamphetamine (MA): 22 °C, 40 °C, 60 °C and 73 °C. (b) Amount of sample adsorbed as a function of the stirring rate at 60 °C. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

Adjustment of the pH of the sample can improve the sensitivity of the method for basic and acidic analytes. This is related to the fact that unless special coatings are used, SPME can extract only neutral (nonionic) species from water. By properly adjusting the pH, weak acids and bases can be converted to their neutral forms, in which they can be extracted by the SPME fiber. To make sure that at least 99% of the acidic compound is in the neutral form, the pH should be at least two units lower than the pK_a of the analyte. For the basic analytes, the pH must be larger than pK_b by two units. Figure 15 illustrates the increase in the amount of amphetamine and methamphetamine extracted as the pH is increased.[22] In practice, it may be difficult to use pH adjustment with direct extraction, since the coating might be damaged by solutions of very high or very low pH values. Headspace SPME is therefore a natural choice for pH-adjusted matrices. The sample should be buffered to ensure good reproducibility when basic or acidic compounds are present in it.

Addition of any ionic salt to aqueous samples generally increases the fiber/matrix distribution constants of neutral organic molecules. However, when the analytes are in the dissociated form, a decrease in the amount extracted is observed. This is related to the fact that the activity coefficients of ionic species in water decrease with the increase in the ionic strength of the solution. It is important, therefore, to first convert the analytes to neutral

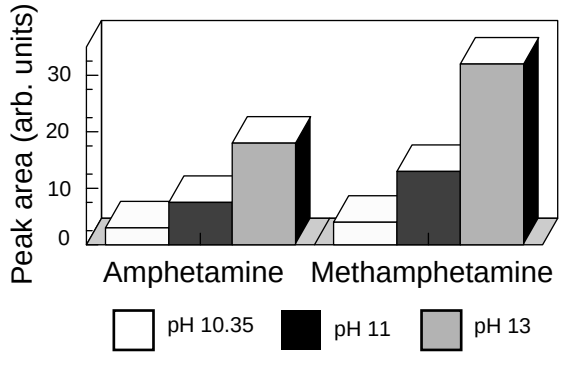

Figure 15 Effect of pH on amphetamine and methamphetamine extraction.

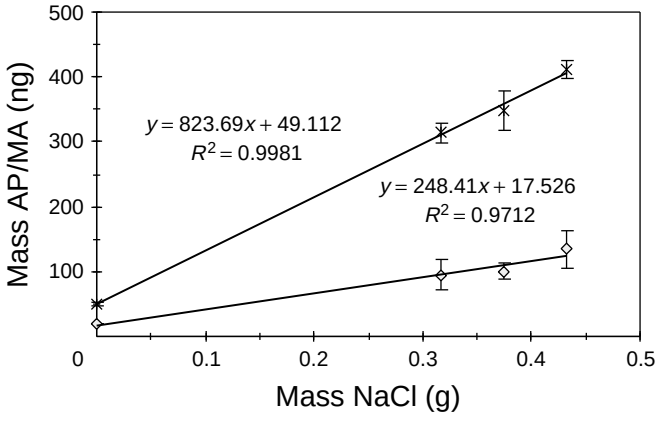

Figure 16 Effect of salt concentration on amphetamine (◇) and methamphetamine (×) extraction. 0.5 N KOH at 22 °C. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

forms. Figure 16 illustrates the increase in the amount of amphetamines extracted from a solution when any ionic salt is added after the pH is lowered by KOH addition.

3.12 Determination of the Linear Dynamic Range of the Method for a Pure Matrix at Optimum Extraction Conditions

Modification of the extraction conditions affects both the sensitivity and the equilibration time. It is advisable, therefore, to check the previously determined extraction time before proceeding to the determination of the linear dynamic range. This step is required if substantial changes of the sensitivity occur during the optimization process.

SPME coatings include polymeric liquids such as PDMS, which by definition have a very broad linear range. For solid sorbents, such as Carbowax™/DVB or PDMS/DVB, the linear range is narrower because of a limited number of sorption sites on the surface, but it still can span over several orders of magnitude for typical analytes in pure matrices. In some rare cases when the analyte has extremely high affinity towards the surface (e.g. basic proteins adsorption on poly(acrylic acid)[23] or bioaffinity coatings), saturation can occur at low analyte concentrations. In such cases, the linear range can be expanded by shortening the extraction time. In the majority of practical applications developed to date, it is the dynamic range of the detector that limits the linear response of SPME methods.

3.13 Selection of the Calibration Method

Standard calibration procedures can be used with SPME. The fiber blank should be first checked to ensure that neither the fiber nor the instrument cause interferences with the determination. The fiber should be conditioned prior to the first use by desorption in a GC injector or a specially designed conditioning device. This process ensures that the fiber coating itself does not introduce interferences. Fiber conditioning may have to be repeated after analysis of samples containing significant amounts of high-molecular-weight compounds, since such compounds may require longer desorption times than the analytes of interest.

When the matrix is simple (e.g. air or groundwater), the distribution constants are very similar to those for pure matrix. It has been shown, for example, that typical moisture levels in ambient air, as well as the presence of salt and/or alcohol in water in concentrations lower than 1%, usually do not change the K values beyond the 5% RSD typical for SPME determinations. In many such instances, calibration might not be necessary since the appropriate distribution constants which define the external calibration curve are available in the literature or can be calculated from chromatographic retention parameters. External calibration can also be used successfully when the matrix is more complex, but well defined (e.g. process streams of relatively constant composition). Of course calibration standards have to be prepared in such cases in the matrix of the same composition rather than in the pure medium (e.g. water).

A special calibration procedure, such as isotopic dilution or standard addition, should be used for more complex samples. In these methods, it is assumed that the target analytes behave similarly to spikes during the extraction. This is usually a valid assumption when analyzing homogeneous samples. However, it might not be true when heterogeneous samples are analyzed, unless the native analytes are completely released from the matrix under the conditions applied. Moreover, whenever any of these methods is used, an inherent assumption is made that the response is linear in the concentration range between the original analyte concentration and the spiked concentration. While this is usually true for fibers extracting the analytes by absorption (PDMS, PA) and detectors with wide linear range, for example, a flame ionization detector (FID), problems may arise when porous polymer fibers (PDMS/DVB, Carbowax™/DVB) are used, or when the detector applied has a narrow linear dynamic range. It is important, therefore, to check the linearity of the response using standard solutions before applying standard addition or isotopic dilution for calibration. To improve the accuracy and precision, multipoint standard addition should be used whenever practical.

3.14 Optimization of Extraction Conditions for Heterogeneous Samples

Optimizing the process that releases native analytes from the matrix components (e.g. solid particles) is more

difficult since the analyte/matrix association is usually poorly understood. Whenever a new type of complex matrix is considered, a study must be conducted to find the optimum extraction conditions which give the fastest and most complete release and extraction of native analytes. Typically, an empirical approach is taken and several parameters are varied, such as temperature and type of additives. Change of extracting phase type, used in optimization of many solvent extraction methods, is of limited value in SPME since the fiber coating does not interact directly with solids present in the matrix. Heating can release analytes from a solid matrix, but sometimes heating is not adequate because analytes are bound too strongly or the matrix or analytes are unstable at higher temperatures. Additives (matrix modifiers) which enhance the extraction of analytes are an alternative to heating. The additive selected should enhance the release of analytes from the matrix but not interfere with the partitioning of analytes into the fiber coating or with the analysis of the extracted compounds. Successful modifiers for SPME include nonvolatile acids, salts and water.[2,24] Better understanding of the analyte–matrix interaction would allow a more rational choice of appropriate additives. In-depth investigation of sample matrices should be considered.

Frequently, the conditions optimal for releasing the analytes from the matrix are different from the conditions that are optimal for partitioning them into the coating. In most cases, high temperature is used to remove native analytes from the matrix. Higher temperatures, however, reduce the distribution constant and thus the sensitivity of the method. To eliminate this disadvantage, a cooled-fiber approach can be applied to condense analytes onto the fiber.[2]

Optimization of extraction parameters is considered complete when the recoveries of analytes present in native samples match those for spikes. This verification is performed based on extraction results obtained for certified standard reference materials (SRMs), or by a comparison with standard extraction methods. Once the method is verified, quantitation is performed by comparing extracted amounts of native and spiked analytes. Table 7 illustrates the comparison of results of PAH analysis by SPME with certified reference values obtained for urban air particulate (NIST 1649).[25] The method consists of static hot water extraction at 250 °C, followed by a cool-down period to increase the amount of analytes absorbed by the PDMS coating.

Table 7 Quantitation of PAHs from urban air particulate (NIST 1649) by static high temperature water extraction

Compound	Cert. conc. $\mu g\,g^{-1}$ (%RSD)	Estimated concentration as % of certified concentration (% RSD)		
		100 mg 270 °C	50 mg 270 °C	50 mg 300 °C
Fluoranthene	7.1 (7)	137 (6)	149 (8)	128 (2)
Pyrene	7.2 (7)	113 (9)	121 (7)	105 (7)
Benzo[*a*]pyrene	2.9 (17)	49 (14)	72 (10)	70 (4)

3.15 Verification of the Equilibration Time, Sensitivity and Linear Dynamic Range for Real Sample Matrices

The sample matrix can affect not only the distribution constant, but also the equilibration time. When the matrix is heterogeneous (e.g. an aqueous sample with solid particulate matter), the kinetics of analyte release might determine the overall extraction rate. To operate at optimum sensitivity, extraction time needs to be adjusted accordingly.

The amount of analyte extracted by the fiber changes substantially when the coating is swollen owing to absorption of a large amount of an interfering major matrix component. To produce swelling, the amount of interference which needs to be absorbed must be substantial compared to the mass of the coating (more than 1%), which translates to about 10 μg for a 100-μm PDMS fiber. In trace analysis of homogeneous aqueous samples this would rarely be the case, since it requires that the interfering compound has good solubility in both the aqueous matrix and the coating. However, the phenomenon might occur if the interference exists as a separate phase, such as an oily suspension or dispersed hydrophobic humic material.[26] In such cases, SPME is of limited value as a quantitation tool, since both the distribution constant and the volume of the coating can change during extraction. The effect of high-molecular-weight compounds on the results of SPME analysis of aqueous samples can be eliminated by using microporous hollow fiber membranes to isolate the coating from direct contact with the sample.

Apart from the extraction time, the linear dynamic range and method sensitivity should also be examined for real matrices. Typically, no differences between model solutions and real samples occur when fibers extracting the analytes by absorption (PDMS, PA) are used. However, when porous polymer fibers are used, competitive adsorption of interfering matrix components with large distribution ratios can reduce the uptake of analytes of interest and cause nonlinearity of the calibration curves. The adverse effect of competitive sorption can be minimized by extracting the analytes under nonequilibrium conditions. For example, short extraction can be performed with a well-agitated sample, or static extraction can be used. In both cases, the uptake of volatile analytes with lower distribution ratios is only

slightly affected, while that of less volatile compounds with large distribution ratios is significantly reduced.

The sample matrix can affect not only the extraction process, but also the separation. It should be verified that no matrix components co-elute with the analytes of interest, interfering with proper quantitation.

3.16 Determination of Method Precision

The most important factors affecting precision in SPME are listed below:

- agitation conditions
- sampling time (if nonequilibrium conditions are used)
- temperature
- sample volume
- headspace volume
- vial shape
- condition of the fiber coating (cracks, adsorption of high-molecular-weight species)
- geometry of the fiber (thickness and length of the coating)
- sample matrix components (salt, organic material, humidity, etc.)
- time between extraction and analysis
- analyte losses (adsorption on the walls, permeation through Teflon®, absorption by septa)
- geometry of the injector
- fiber positioning during injection
- condition of the injector (pieces of septa)
- stability of the detector response
- moisture in the needle

The majority of these factors and their effect on method precision have already been discussed. A brief explanation of the remaining parameters follows.

Reproducibility of the sample and headspace volumes can impact on the precision of SPME methods. The volume of the sample has an impact on the amount extracted when only a few mL sample is used (see the discussion above). Therefore, small samples, such as 1 mL, need to be measured very carefully. For large samples, small variations in the sample volume do not have a substantial effect on precision of the data. Obviously, the volume of all the samples and standards must be the same to obtain valid results.

Headspace volume can be an important factor determining precision of the results in three-phase systems. It is relatively easy to measure sample volume accurately. However, vials are not manufactured to have exactly the same total volume. Wall thicknesses and bottom shapes may differ from vial to vial. Also, the shape of the septum in a closed vial can vary from concave to convex. All these factors will affect the total volume of the system, and therefore the headspace volume, which is the difference between the total volume and the sample volume. Such differences (especially related to septum shape) are usually more pronounced in small vials, therefore worse precision can usually be expected when they are used. In addition to keeping the sample and standard volumes constant throughout all the experiments, vials of the same volume should be used for all samples and standards, so that the headspace volume is constant from vial to vial. This becomes a bigger challenge for small vials, when a small absolute variation in headspace volume could translate to a large relative difference.

The condition of the fiber is very important. The presence of high-molecular-weight species, such as proteins and humic matter, adsorbed on the surface of the fiber, causes a change in sorption characteristics. In some cases, the substances adsorbed can be removed by washing the fiber in appropriate solvents. When the coating is badly cracked, a small amount of the matrix itself kept in the cracks by capillary forces might be transferred into the analytical instrument and adversely affect its operation.

The volume of the coating determines the amount of analytes extracted. Fibers which begin to lose the coating phase should be replaced immediately. The length of the fiber should be monitored to ensure equivalency when changing the fiber. If the length of the fiber (thus the volume of the coating) is different, an appropriate correction factor should be incorporated.

Time between preparation of standards or sample collection and extraction should be minimized. After placing aqueous samples or standards in a vial, analytes begin to interact with the walls of the container. Very hydrophobic substances are frequently strongly adsorbed on the surfaces and are not available for extraction by the coating. This results in lower than expected extracted amounts. One way to prevent this problem is to use vials with more inert walls, for example silanized glass or Teflon®. For volatile analytes, it is crucial that new Teflon®-lined septa are always used. Once the Teflon® lining is pierced, the analytes can come into direct contact with the septum material (typically PDMS), which usually results in losses through absorption. Finally, it should be remembered that many volatile compounds can permeate through Teflon® at a relatively high rate. This phenomenon can also contribute to analyte losses when the samples or standards are stored for prolonged periods of time.

Time between extraction and analysis should also be minimized. When this time is longer that a few seconds, the needle of the fiber assembly should be sealed with a septum and preferentially cooled down to reduce losses and interferences. After the fiber is withdrawn from the sample, it begins to equilibrate with ambient air. This process is slow when the fiber is withdrawn into the

needle. Therefore, the few seconds which are normally required to transfer the fiber to the injection port of a GC do not result in any substantial losses of the analytes. However, when the fiber is to be stored for some time before the analysis, sealing of the needle is necessary. In addition, this protection prevents contamination of the fiber with interferences or even target analytes which might be present in ambient air. Cooling of the SPME device might be necessary to extend the stability of the volatile analytes in the coating.[11]

3.17 Determination of Method Detection Limit

Several methods are described in the literature to determine the method DL. The most widely accepted definition is based on estimating the DL using low concentration spikes and calculating the standard deviation of the determination. The DL is then defined as 3 times the standard deviation obtained for the measurement value not higher than 10 times the method DL.

3.18 Validation of the Method

Validation of the method might include comparison of quantitation results with certified values obtained for SRMs which have similar matrix and target analytes. Another approach is to validate the method against officially accepted techniques for the analysis of target samples and analytes. Table 8 summarizes the results of multilevel validation of SPME against the standard purge and trap technique for the analysis of BTEX (benzene, toluene, ethylbenzene, *m*-xylene, *o*-xylene, *p*-xylene) compounds in water. The regression line has a slope close to one, with a very small intercept value and linear correlation coefficients better than 0.99 for all species. The very good agreement between both methods indicates suitability of headspace SPME for the analysis of volatile organic compounds (VOCs) in aqueous samples.[27]

Table 8 Regression line parameters for SPME versus purge and trap for clean water analysis

Compound	Regression parameters – purge and trap on *x*-axis		
	Slope	*y*-Intercept	Correlation
Benzene	1.07 ± 0.087	-5.9 ± 6.6	0.9951
Toluene	1.06 ± 0.060	-4.6 ± 4.6	0.9976
Ethylbenzene	1.06 ± 0.046	-3.8 ± 3.7	0.9986
m/*p*-Xylene	1.05 ± 0.073	-8.4 ± 11	0.9964
o-Xylene	1.07 ± 0.060	-4.0 ± 4.6	0.9977

Finally, interlaboratory studies are frequently performed to validate a method. Table 9 summarizes the results obtained in a round robin test on pesticide analysis by SPME, which involved several laboratories in Europe and North America.[28] In general, the results are characterized by good repeatability, which proves that SPME is a valid method for the determination of pesticides at trace levels. As expected, the interlaboratory and reproducibility standard deviations are higher since they include differences between laboratories. However, they are still satisfactory. The results indicate also that SPME is an accurate method. In all cases, the confidence intervals of the gross average and the true value overlap, which indicates that any differences between the two respective values are due to random factors. Interestingly, for 10 out of 12 compounds, the values of the gross average are slightly lower than the true values. This might be due in part to losses of analytes through adsorption (as described above) in cases

Table 9 Statistical characteristics of the results obtained in the round robin test on pesticide analysis by SPME

Compound	s_r	s_L	s_R	r	R	G.A.	C.I.	T.V.
Dichlorvos	2.06	5.04	5.44	5.83	15.4	27.3	27 ± 5.8	25 ± 1.35
EPTC	0.56	1.56	1.66	1.57	4.7	9.9	10 ± 1.6	10 ± 0.54
Ethoprofos	0.82	4.79	4.86	2.32	13.74	15.5	16 ± 2.3	17 ± 0.92
Trifluralin	0.27	0.57	0.63	0.76	1.79	1.6	1.6 ± 0.76	2 ± 0.11
Simazine	2.34	3.45	4.17	6.61	11.79	23.6	24 ± 6.6	25 ± 1.35
Propazine	1.21	2.04	2.37	3.42	6.71	9.5	10 ± 3.4	10 ± 0.54
Diazinon	0.63	2.13	2.22	1.79	6.29	8.2	8 ± 1.8	10 ± 0.54
M.chlorpyriphos	0.12	0.32	0.34	0.35	0.97	1.6	1.6 ± 0.35	2 ± 0.11
Heptachlor	2.03	2.89	3.53	5.75	10	8.9	9 ± 5.8	10 ± 0.54
Aldrin	0.54	0.73	0.91	1.53	2.58	2.0	2 ± 1.5	2 ± 0.11
Metolachlor	0.73	2.83	2.92	2.07	8.28	15.7	16 ± 2.1	17 ± 0.92
Endrin	0.87	3	3.13	2.47	8.85	8.8	9 ± 2.5	10 ± 0.54

All values expressed in $\mu g\,L^{-1}$; s_r – repeatability standard deviation, s_L – interlaboratory standard deviation, s_R – reproducibility standard deviation, r – repeatability, R – reproducibility, G.A. – gross average, C.I. – confidence interval of the gross average, T.V. – confidence interval of the "true" value.

where the aqueous pesticide solutions were not prepared directly before the analysis, as required by the test protocol.

3.19 Automation of the Method

SPME is a very powerful investigative tool, but it can also be a technique of choice in many application for processing a large number of samples. To accomplish this task would require automation of the methods developed. As automated SPME devices with more advanced features and capabilities become available, automation of the methods developed becomes easier. A currently available SPME autosampler from Varian enables direct sampling with agitation of the sample by fiber vibration, and static headspace sampling. In some cases, custom modifications to the commercially available systems can facilitate operation of the method closer to optimum conditions.

4 APPLICATIONS OF SOLID-PHASE MICROEXTRACTION

Most of the methods developed so far are applicable for clean matrix samples and only a few optimization parameters have been fully investigated. Many other potential applications are still waiting to be explored. The results obtained are very promising. A basic understanding is emerging of the basic matrix types: gaseous, liquid and solid and is summarized below. General approaches to facilitate successful extraction of these systems are emphasized. The object of this section is not to demonstrate a comprehensive review, but to give examples of successful SPME applications.

4.1 Gaseous Matrices

All SPME methods developed to date for gas analysis are directed towards air analysis. However, the approaches described are also suitable for the analysis of other gas mixtures.

Model studies on extraction from air have been performed by two methods: static and (the more realistic) dynamic. In the static method, the target compounds are introduced to a glass bulb equipped with a septum. After the analytes vaporize completely, the SPME fiber is introduced to the bulb through a septum. In practical ambient air measurements, the system is not static, since convection is always present. It is more appropriate, therefore, to use dynamic gas chambers for the modeling studies.[5,8] Equilibration times for extraction of trace contaminants from moving air are short, as predicted from the theory. Figure 17 shows the equilibration time profiles obtained for a 20 L chamber operated at 200 mL min^{-1}. The equilibration times are between 30 and 100 s, corresponding to pentane and undecane, respectively. They are close to theoretical values for perfect agitation

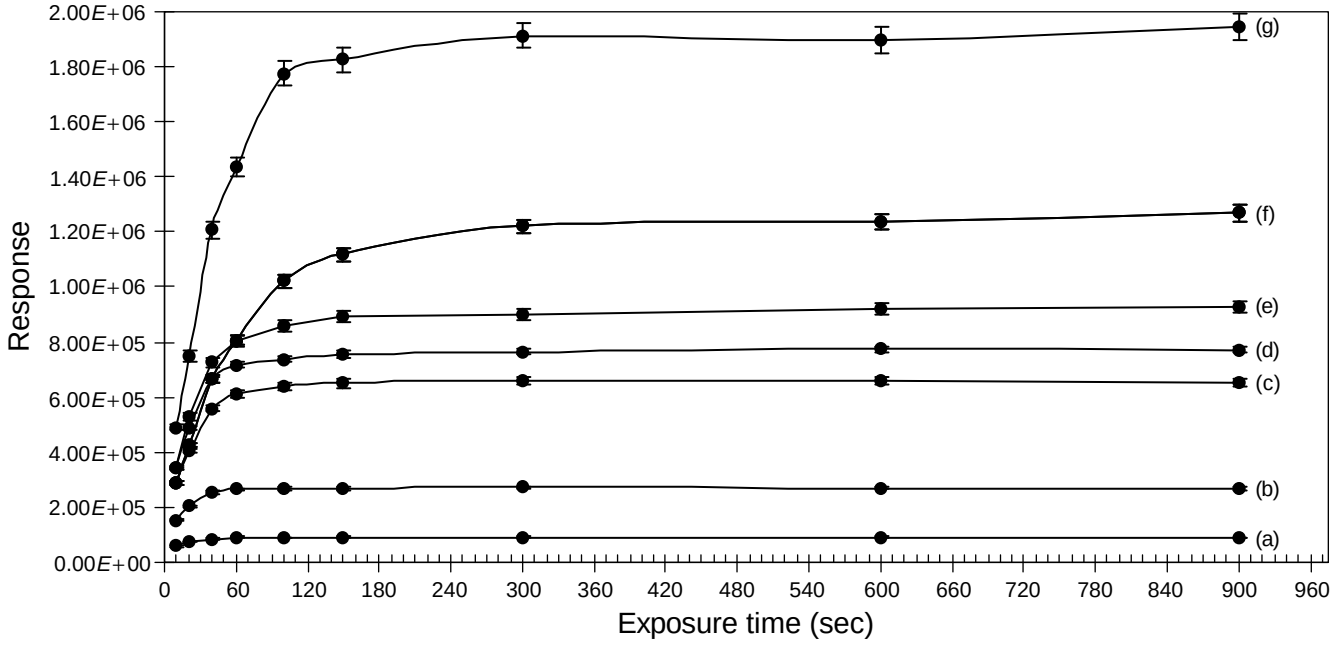

Figure 17 Representative extraction time profiles obtained for a standard gas mixture containing (a) benzene, (b) toluene, (c) ethylbenzene, (d) *p*-xylene, (e) *o*-xylene, (f) α-pinene and (g) 1,3,5-trimethylbenzene. PDMS-coated fiber.

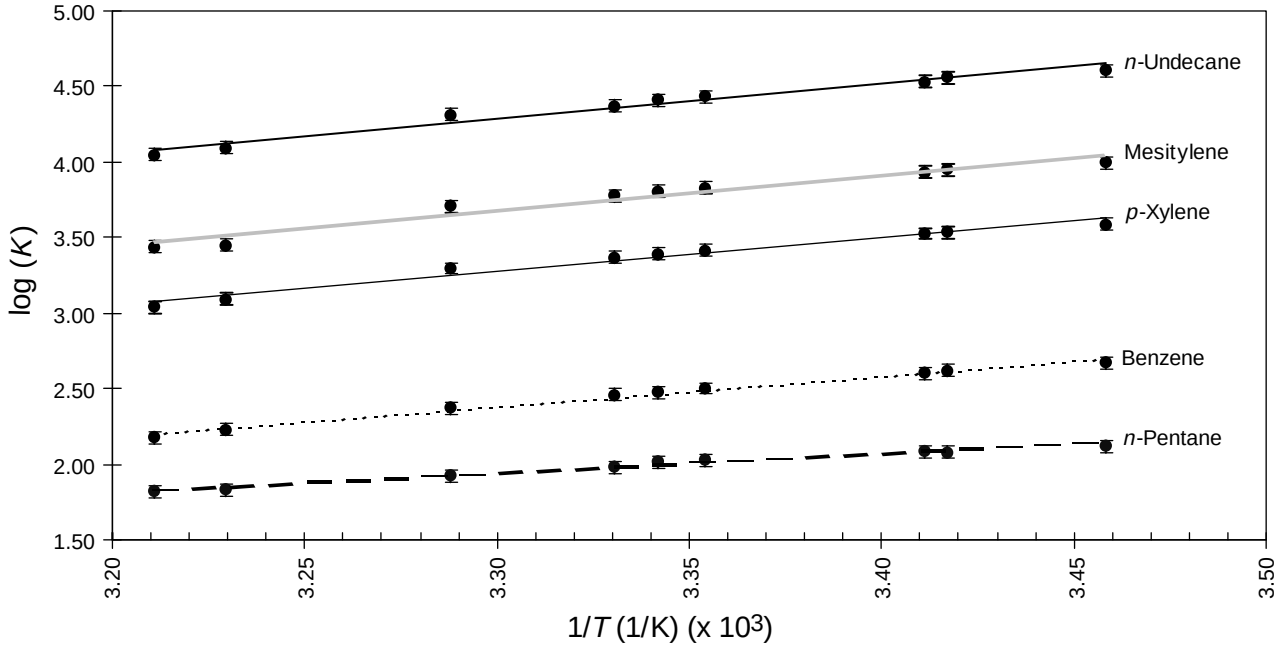

Figure 18 Representative plots for pentane, benzene, *p*-xylene, mesitylene and *n*-undecane demonstrating the linear relationship between $\log K_{fg}$ and $1/T$.

conditions. Increased air flow decreases the equilibration times for less volatile analytes.

PDMS coating is generally not affected by major air components. A small decrease in the response for organics was observed only for samples of relative humidity close to 100%. The primary experimental parameter which controls the response is temperature. However, the effect of temperature on the distribution constant can be easily predicted (Equation 19), since $\log K_{fg}$ is linearly related to $1/T$ and the heat of vaporization of the pure solute, ΔH^{V}:[5]

$$\log_{10} K_{fg} = \frac{\Delta H^{V}}{2.303RT} + \left(\log_{10} \frac{RT}{\gamma_i p^*} - \frac{\Delta H^{V}}{2.303RT^*} \right. \qquad (19)$$

where p^* is the analyte vapor pressure at a known temperature T^* for a pure solute, and γ_i is the activity coefficient of the solute in the coating. In other words, K_{fg} can be calculated for given extraction conditions using the temperature measured and the heat of vaporization for the target compound. Figure 18 illustrates the above relationship graphically for a range of compounds varying in volatility. The linear relationship is clearly illustrated.

In addition, as has been described in section 2, the heat of vaporization and activity coefficient are related to the retention time of a compound on the GC column using the same coating material as the SPME fiber. Thus, the appropriate distribution constants can be determined directly from the retention indices of the target analytes. For PDMS coating, the following relationship between the gas/coating distribution constant and the LTPRI has been found (Equation 20):

$$\log_{10} K_{fg} = 0.00415 \times \text{LTPRI} - 0.188 \qquad (20)$$

The differences between the values of distribution constants calculated from the retention indices and determined directly were generally found to be insignificant. This approach can be extended to other coatings as long as columns with appropriate coatings are available. In addition, using this approach it is possible to determine the distribution constants of unidentified compounds, therefore it is possible to use SPME for the determination of parameters like total petroleum hydrocarbons (TPH) in air. A comparison of SPME/PDMS with the standard charcoal tube technique obtained for an air sample drawn near a gas station, resulted in 262 and 247 $\mu g\,L^{-1}$, correspondingly.[5] The values show a very good agreement indicating that the approach described above can be used to quantify analytes in gaseous samples without the need for identification. It should be pointed out, however, that this approach can be used only with detectors whose response is more or less independent of the nature of the analyte (e.g. FID for the analysis of hydrocarbons).

SPME has superior sensitivity for short term monitoring compared to traditional devices, which are limited by the gas throughput. The SPME fibers can also be used at high temperatures. For example, Figure 19 illustrates the identification of several polycyclic aromatic hydrocarbons (PAHs) directly in diesel exhaust. A typical distribution of these pollutants is illustrated. The advantages of SPME

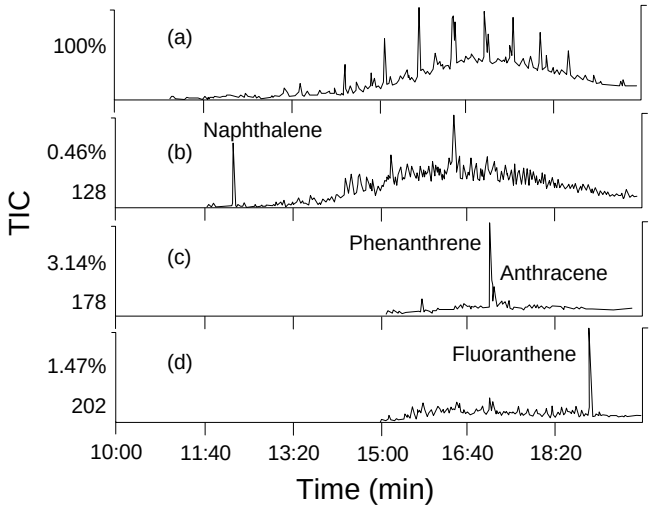

Figure 19 GC/MS traces corresponding to polycyclic aromatic hydrocarbons present in diesel exhaust sampled by SPME/PDMS; (a) total ion chromatogram; selected ion traces at: (b) $m/z = 128$, (c) 178, (d) 202.

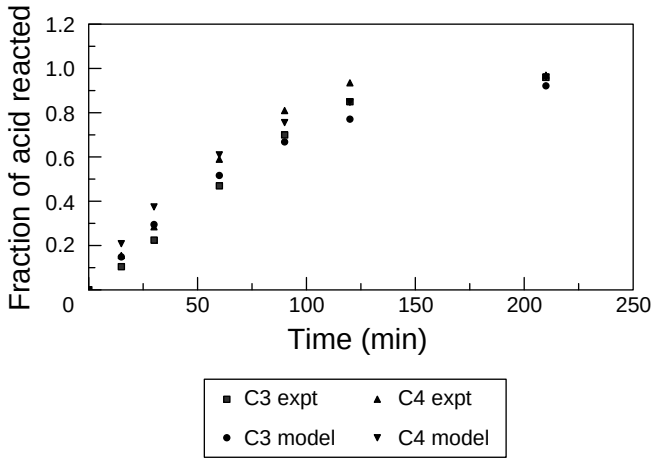

Figure 20 Comparison of experimental data versus theoretical model predictions for the reaction of gaseous propionic and butyric acids with PDAM in the PA-coated fiber.

for long-term sampling are not as clear at the present time. Typically, the fiber reaches equilibrium with the air components in the first few minutes and then does not accumulate any more analytes, independently of the exposure time. However, the sensitivity of the SPME technique can be improved by using thicker, more selective coatings or by cooling the fiber to increase the PDMS/gas distribution constant. Another approach is to incorporate the derivatization reagent in the coating to allow ''trapping'' of the analyte in the coating. All these modifications enhance the sensitivity of SPME in air monitoring.

SPME can also be used as an integrated sampling device. Theoretical principles of this approach have been described in section 2. The simplest way to accomplish this task is to retract the fiber into the needle. An alternative approach is to use a chemical reaction in the coating. Figure 20 shows the extraction of low-molecular-weight carboxylic acids from a small volume of air using a fiber doped with PDAM.[14] As expected, a linear dependence between the amount of acid extracted and time is obtained until about 60% of the acid is extracted from the sample. In practice, however, when the fiber is exposed to ambient air, analytes accumulate linearly with concentration for as long as the fiber is exposed or the derivatization reagent is substantially consumed. This approach is particularly useful for integrated sampling of very low analyte concentrations in air. Another interesting application of SPME coupled to an on-fiber derivatization step was demonstrated for analysis of aldehydes in air. In this case *o*-(2,3,4,5,6-pentafluorobenzyl)hydroxylamine was used as derivatization reagent and the products, oximes were analyzed by GC.[29]

Frequently, after the extraction, the fibers with analytes in the coating cannot be analyzed immediately. This is particularly true in field analysis, since an appropriate portable instrument might not be available. In such situations, the fibers have to be stored in a way assuring that analyte losses are minimal. This can be simply accomplished by retracting the fiber into the needle and sealing its open end with a piece of a septum. Cooling the needle provides additional protection against losses. For PDMS fibers, even with such precautions, losses of the most volatile analytes might occur when the fiber is stored for more than a few hours. When properly protected (retracted fiber, needle sealed with a septum, low temperature), fibers coated with PDMS/DVB can be safely stored for much longer periods of time.

4.2 Liquid Matrices

All methods for liquid matrices reported in the literature, up to the end of 1996, had been developed for aqueous samples. However, the technique is not limited to water analysis. When appropriate coatings are used, target analytes can be preconcentrated on SPME fibers from other types of matrices. For example, Nafion™ coating can be used to extract polar compounds from fuels.[12]

The fiber coating/water distribution constant can be calculated from Equation (21):

$$K_{fw} = K_{fg}K_{aw} \tag{21}$$

where K_{fg} can be determined from chromatographic data as discussed above, and K_{aw} is the air/water distribution constant for a given analyte which can be found in the tables of Henry's constant values. For example, the equation for a PDMS coating and aqueous matrix can be

calculated from Equation (22):

$$K_{fw} = 0.00415 \times \text{LTPRI} - 0.188 + \log_{10} K_{aw} \quad (22)$$

This equation is obtained after substitution of the expression for K_{fg} given by Equation (21) into Equation (20). The K_{fw} values found using this equation agree very well with experimental data[6] considering that typically the errors in Henry's constant determination are larger than 10%. In addition, the Henry's constants are similar for closely related compounds, as illustrated in Figure 21. As could be predicted from Equation (22), the slopes of the lines are similar, but their intercepts vary because of the differences between the average values of Henry's constant for the different classes of compounds. This relationship enables quantitation of unknown analytes in water provided that the class to which a compound belongs is known.

Several reports indicated the existence of a linear relationship between the coating/water distribution constant and octanol–water distribution constant, K_{ow}.[30] Considering the discussion above, such relationships are expected to exist only for individual classes of compounds, such as aromatic hydrocarbons. Because the activity coefficients of various classes of analytes in octanol are expected to be different from those in PDMS or other fiber coatings, the relationship between the two extracting phases will vary with the change of chemical properties of the analytes. It is possible, however, to predict the general trends in K_{fw} within a group of related compounds by using the corresponding values of K_{ow}.

The above discussion pertains to pure water as the sample matrix. The presence of other components in water might affect the distribution constants of the analytes. Also, liquid chromatographic experience gives some clues about the trends in the distribution constant change with modification of the matrix. For example,

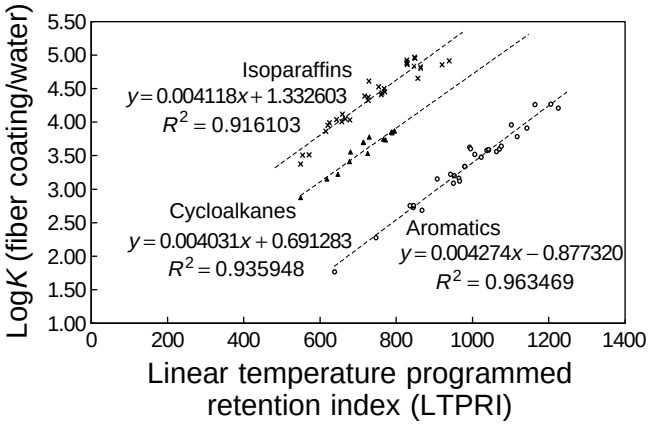

Figure 21 Log K_{fw} PDMS as a function of LTPRI for isoparaffins, substituted benzenes and cycloalkanes at 25 °C.

addition of any ionic salt would generally result in an increase in the distribution constant for neutral organics, but the change is expected to be noticeable only if the concentration of the salt exceeds 1%. Also, the presence of water-miscible polar organic solvents may change the properties of the matrix by reducing its polarity. In addition, swelling of the polymer with the solvent might occur for polar coatings, resulting in a change of the coating volume and possibly also of K_{fs}. However, the change is not expected to be substantial when the concentration of the solvent is below 1%.[31] When samples contain higher amounts of salt and/or dissolved organics, but they are well defined so that pure matrix can be prepared, external calibration using standards in the matrix may still be appropriate. Otherwise, standard addition (preferentially of isotopically labeled analytes) should be used to compensate for variations in the matrix composition.

A major analytical challenge is always associated with the analysis of samples containing solids, such as sludge. Several approaches can be implemented with SPME. Sometimes modification of the extraction conditions, such as temperature, pH, salt and other additives, facilitates the displacement of analytes into the aqueous phase or the headspace, resulting in distribution constants similar to those for pure water. In certain cases, direct extraction is not possible because of a very dirty matrix or pH conditions which damage the fiber (above pH 12 for the PDMS coating). In such situations, the headspace mode is more suitable for many applications. Even semivolatiles can be analyzed by this method as long as the extraction temperature is sufficiently high, and good agitation conditions are provided. Application of SPME with membrane protection can assist in the extraction of nonvolatile species in the presence of high-molecular-weight interferences which are able to alter the properties of the coating (e.g. proteins or humic material).[26]

4.3 Solid Matrices

Accurate quantitation of target analytes in solids represents a very significant challenge to the analytical community. Although SPME cannot be used directly to extract analytes from solids, several approaches can be taken to facilitate simple sample preparation. For volatiles, the typical approach is to perform headspace analysis. To release the analytes from the matrix quantitatively, the temperature needs to be increased. This facilitates extraction and improves the kinetics of the process, but it also decreases the distribution constant. Typically, a maximum is observed for the relationship between the amount of analyte extracted by the fiber and temperature.[2] Initially, the loss of sensitivity due to decreased distribution ratio is more than compensated

for by the increased concentration of the analyte in the headspace. At even higher temperatures, the loss of sensitivity becomes predominant and the amount of analyte extracted by the fiber decreases. The inherent loss of sensitivity associated with the decrease of the distribution constant at high temperatures can be compensated for by cooling the fiber. However, this approach has not yet been commercialized. Alternatively, matrix modifier can be added to the solid sample. Water has been found to be very effective in displacing the analytes from the solid surfaces for many types of samples, especially at elevated temperatures.

Another successful approach to SPME analysis of solids involves the use of water or a polar organic solvent, such as methanol, for the extraction of the analytes from the solid matrix prior to SPME analysis. When a polar solvent is used, the extract is spiked into pure water. Low-temperature water extraction followed by SPME was found to be a very useful approach for polar compounds, such as herbicides.[32] Application of methanol with water spiking, on the other hand, has been found to be useful for the analysis of volatile hydrocarbons.[33] An interesting modification to the above procedure involves volatilizing the extract followed by fiber extraction from the gaseous phase.[34] Quantitation of analytes in the gas phase is easier, as discussed above.

For less volatile analytes, the methanol approach described above can give good results. In addition, high-pressure hot water extraction is a very suitable solvent-free alternative.[25] Both static and dynamic hot water extraction have their advantages. The static method is very simple and inexpensive since it does not use high-pressure pumps. The dynamic approach, on the other hand, provides an extract which is much cleaner than the original matrix. However, it might be possible to extract many semivolatile target analytes from very dirty matrices with the high-temperature static system, when using the headspace mode of SPME.

4.4 Environmental Applications

Much research has been performed on samples of environmental origin, including air, water and sludge, as well as soil. A majority of applications have been developed for aqueous matrices. The results obtained for priority pollutants in water are very encouraging, indicating that in most cases the performance of SPME can meet United States Environmental Protection Agency (USEPA) method requirements. Table 10 summarizes the limits of detection for SPME of volatiles and semivolatiles, including phenols. In each case, the SPME method can be optimized to meet the requirements of the regulatory agency. The following is a brief discussion of the most important accomplishments obtained for several groups of analytes.

Table 10 Summary of DLs for SPME analysis of various classes of organic compounds

Analytes	LOD ($pg\,mL^{-1}$)	EPA specified method DLs ($pg\,mL^{-1}$)
BTEX	1–15	30–90
Polychlorinated HC solvents	1–100	10–100
PAHs	1.2–20	40
PCBs	3	60–100
Phenols	10–800	1500–42 000

LOD, limit of detection; PCBs, polychlorinated biphenyls.

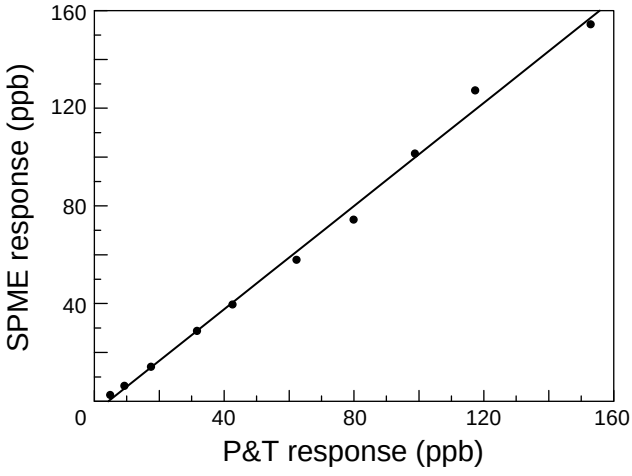

Figure 22 Regression plot for toluene. Each point on the graph represents a comparison between headspace SPME and purge and trap technique for the same water sample.

BTEX compounds have been the most investigated group of organic compounds of environmental interest. The sensitivity of the SPME/PDMS method for these compounds is very high, allowing mid to low part per trillion determinations in water. BTEX compounds are volatile, so the headspace SPME mode can be used successfully in many applications. Validation research has been conducted to compare the performance of this method with more established alternatives. Figure 22 illustrates an example of a multilevel validation for toluene. The correlation line, indicates the equivalence between the purge and trap technique and SPME.[27] The validation work has been extended to sludge and soils. Considering the complexity of the matrices, the correlation between the purge and trap and headspace SPME techniques was good. Validation work with other VOCs, such as chlorinated hydrocarbons, produced similar conclusions.[35]

The measure of TPH is frequently used to estimate the contamination level of a sample. For air samples, this value can be calculated directly from chromatographic

data without the need for identification of all individual components as emphasized in the section devoted to gaseous matrices. TPH can then be reported, together with more specific information related to distribution and type of contaminants present in the sample, without the need for performing calibration. Determination of TPH of water samples is possible if the compounds belonging to different classes are separated and identified by a chromatographic technique combined with a quantitation method such as MS. For solids, this value can be obtained after the analytes are transferred into the gaseous or aqueous phase during the extraction process, as described in the section devoted to solid matrices.

Polar volatile solvents like alcohols, ketones and aldehydes are difficult to quantify at trace levels in aqueous matrices. Extraction with nonpolar solvents leads to poor recoveries. More polar solvents cannot be used since they are miscible with water. SPME with a polar polymeric coating has the potential to provide the necessary alternative. Low parts per billion DLs are possible even with FID when PDMS/DVB coating and 35% salt concentration are used.[5]

Nonpolar semivolatiles, such as PAHs and PCBs are characterized by very large K_{fs} values, which results in very high sensitivities of SPME determination, frequently reaching low parts per trillion levels.[36,37] Figure 23 illustrates major components detected and identified in a pulp and paper-mill discharge by solid-phase microextraction/gas chromatography/ion trap mass spectrometry (SPME/GC/ITMS). The major advantage of SPME compared to alternative techniques is its field portability. Frequently, it is very difficult to collect and preserve a representative sample since nonpolar semivolatiles have a tendency to adsorb on solid surfaces of the sampling vessel or the particulate matter. By sampling directly in the field, these phenomena can be eliminated to a large extent.

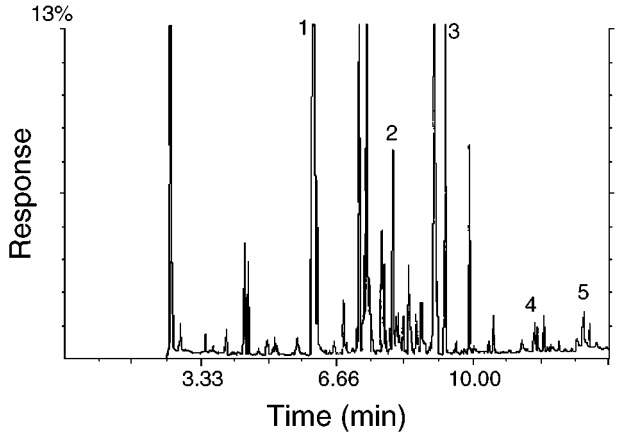

Figure 23 Total ion chromatogram showing organics detected and identified in a pulp and paper-mill discharge wastewater sample extracted by headspace SPME at 25 °C for 30 min: 1, naphthalene; 2, 1,1′-biphenyl; 3, dibenzofuran; 4, anthracene; 5, dibutylphthalate.

Polar semivolatiles in natural matrices represent a very significant analytical challenge. Adjustment of pH and salt addition help to reach the required DLs. For example, the LOD obtained for phenols with direct extraction from an acidified solution with a pH stable PA coating are lower than values specified by USEPA. Derivatization with acetic anhydride allows further improvement in sensitivity and chromatographic performance. The derivatization approach also enhances the extraction performance for carboxylic acids and amines, particularly for lower-molecular-weight species.[1]

Direct extraction of very polar heteroaromatic compounds frequently present in creosote-contaminated groundwater with a PA coating combined with sensitive ion trap MS detection allows low to sub parts per billion DLs for these very water-soluble compounds.[38]

Pesticides are an analytically difficult group of compounds because of their wide range of chemical structures and properties. Some of them are classified as nonpolar (organochlorine pesticides) and others as very polar semivolatiles (herbicides). Despite this diversity, SPME has been very successfully applied for their determination in aqueous matrices by several research groups.[39–43] Table 11 summarizes the SPME DLs obtained for several groups of pesticides. They are lower compared to USEPA method LODs. It is even possible to develop single extraction conditions for rapid screening of 60 target pesticides combining three classes: organochlorine, organophosphorus and organonitrogen.[46]

Application of sodium tetraethylborate as an in situ derivatization reagent enables SPME analysis of some metal ions and organometallic species in aqueous samples. This approach was first used for the determination of methylmercury in water and biota,[44] followed by the analysis of lead ion and organolead compounds.[45] In this method, the analytes are first derivatized to their fully ethylated forms before they are extracted by the PDMS fiber. The derivatization products are volatile, therefore the headspace extraction mode of SPME can be applied, facilitating rapid analysis of very complex matrices. The reaction time (about 15 min) typically determines the extraction time. Stirring must be very efficient during derivatization, since the reagent decomposes rapidly after adjustment of the pH. The affinity of neutral forms of organometallic compounds towards PDMS coating is very high, resulting in very low DLs, approaching sub parts per trillion levels for the determination of tetraethyllead in water by SPME/GC/ITMS. This method can be applied to partial speciation of various forms of these metals present in a sample, first by extracting native neutral organometallic species followed by the determination of the ionic forms (after derivatization).

Table 11 DLs for SPME coupled with various detectors

Pesticide classes	Target analyte subgroups	ECD (ng L^{-1})	NPD (ng L^{-1})	MS (ng L^{-1})	EPA specified method DLs (ng L^{-1})
Nitrogen-containing herbicides[a]	Thiocarbamates		20–60	0.05–.8	100–200
	Triazines		40–6000	0.4–3	100–800
	Nitroanilines		10–30	0.02–0.4	200
	Substituted uracils		200–400	0.1–1	2500–4500
	Substituted amide		800	15	500
	Acetanilide		200	0.01	700
	Diphenyl ether		300	6	NA
	Triazole		30	0.01	NA
Organochlorine pesticides[b]	HCH's	0.9–9*		0.01–0.04	10–25
	Hexachloro-cyclodienes	0.06–1.6		0.02–0.6	2.5–50
	Diphenyl aliphatics	0.05–4.7		0.06–4.5	10–75
Phosphorus-containing pesticides[a]	Phosphate		500	6	100–2500
	Phosphorothiolate		130	2	NA
	Phosphonothiate		15	8	9–200
	Pyrophosphate		16	1	6
	Phosphorodithioates		9–320	0.4–9	9–1500
	Phosphorothioates		11–280	0.7–100	4–2000

* Analyzed with 30% w/w salt added; NA – not available.
[a] Analyzed by SPME with 85 μm PA (poly(acrylate)) fiber.
[b] Analyzed by SPME with 100 μm poly(dimethylsiloxane) (PDMS) fiber.
NPD, nitrogen–phosphorus detection.

Application of HPLC to separation of species extracted by SPME opens new possibilities in the determination of nonvolatile and thermally unstable environmental pollutants. Analysis of several alkylphenol ethoxylate nonionic surfactants in water by SPME combined with HPLC has been demonstrated.[46] The DLs obtained were at low parts per billion level for individual ethoxamers with ultraviolet (UV) detection, which indicates that trace level determination in natural matrices should be possible by this method.

4.5 Food and Pharmaceuticals

A unique feature of SPME in the analysis of food and drugs is its ability, in some cases, to extract the analytes from the packaged product without opening the package. For example, flavors present in wine can be checked before sale or purchase by introducing the fiber into the headspace of the wine through the cork of the bottle. Since only a small amount of flavor is extracted, the composition of the product does not change, and the extracting phase does not contaminate the product since the polymeric phase is nonvolatile and insoluble. A similar process can be applied to on-line product monitoring to assure the best possible quality.

Food and pharmaceutical products are frequently contaminated by volatile organic solvents, which are used in manufacturing and processing of these goods. Considering the frequent complexity of their matrices, the analytical procedures can be quite involved. Static headspace is frequently used for this purpose, but SPME has been found to be a good alternative for the analysis of both contaminated pharmaceuticals[47] and foods[48] because of its superior sensitivity, especially for less volatile compounds. Samples can be analyzed by dissolving the product in water prior to determination, or the analytes can be extracted directly from the headspace above the product. Typically, standard addition is used for quantitation. The DLs are similar to those obtained for environmental matrices.

Flavors constitute another very interesting topic for investigation. SPME has been applied for the analysis of a wide range of food products, spices, oils, beverages[49,50] and even black and white truffle aroma.[51] Initial investigations indicate that good quantitation of volatile analytes in the matrix can be obtained after the solid components of the sample matrix are separated from the aqueous phase by centrifugation. Combining SPME extraction with GC separation using chiral stationary phase, provides information not only about the composition of the products, but also their natural purity.[52] Figure 24 shows the separation of peppermint oil (a) and peppermint-flavored candy (b) using a β-cyclodextrin column.

Another interesting application of SPME for the analysis of foods is to quantify caffeine in soft drinks.[53] The most convenient quantitation method for this determination is based on isotopically labeled standards which are readily available. In addition to caffeine, flavor components can also be detected in the same chromatographic run.

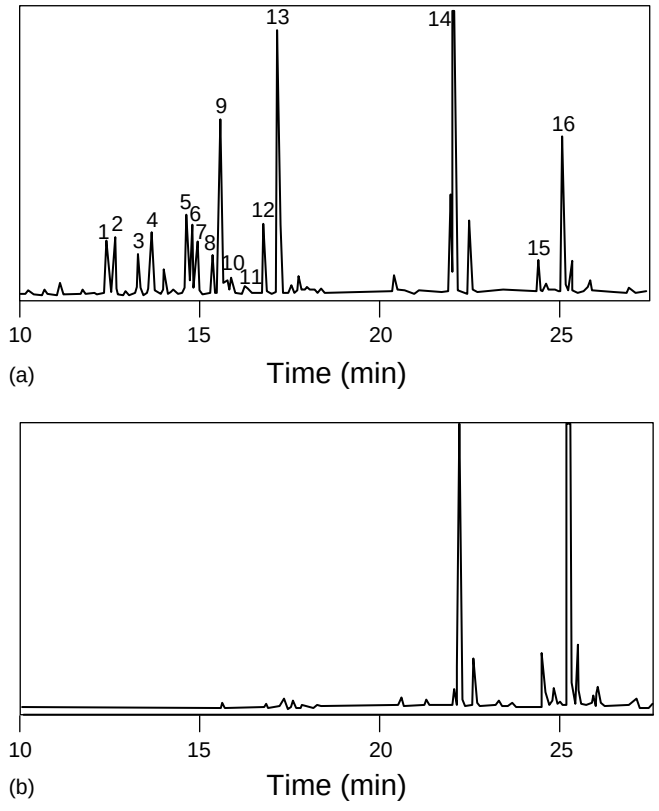

Figure 24 Headspace SPME extraction and chiral GC analysis of (a) peppermint oil and (b) peppermint-flavored candy using a β-cyclodextrin column. Oven temperature program: 40 °C for 2 min, then 40–220 °C at 4 °C min^{-1}. Peaks: 1, (−)-α-pinene; 2, (+)-α-pinene; 3, β-myrcene; 4, (+)-sabinene; 5, (−)-camphene; 6, (−)-β-pinene; 7, (+)-β-pinene; 8, α-terpinene; 9, 3-carene; 10, (−)-limonene; 11, (+)-limonene; 12, (±)-β-phellandrene; 13, γ-terpinene; 14, cineole; 15, menthone; 16, (+)-menthol; 17, (−)-menthol; 18, isomenthol; 19, (±)-menthyl acetone.

The number of research reports on semivolatiles in foods is more limited, but the scope of potential applications is as broad as in the environmental area. One of the interesting applications is determination of herbicides in wines.[32]

4.6 Clinical and Forensic Analysis

The major advantage of SPME for clinical and forensic applications is its portability, which might eventually be translated into on-site sampling followed by immediate analysis by portable instruments. This approach would allow better monitoring of a patient's condition during treatment or therapy, and better preservation of a crime scene since the objects would not need to be removed for the analysis to be performed.

Monitoring of human health can be achieved by non-invasive SPME methods. For example, breath contains the headspace of blood. Therefore, characterization of volatile components present inside the human body can be accomplished by adapting the SPME device to breath analysis. Figure 25 presents an example of breath analysis.[54] The DLs are at a low nanomol per litre level, which provides sufficient sensitivity for practical application of this approach to monitoring alcohol and acetone levels in blood. SPME can also be used for rapid analysis of body fluids. Figure 26 illustrates the detection of several aromatic amines in the milk of a mother who is a smoker.

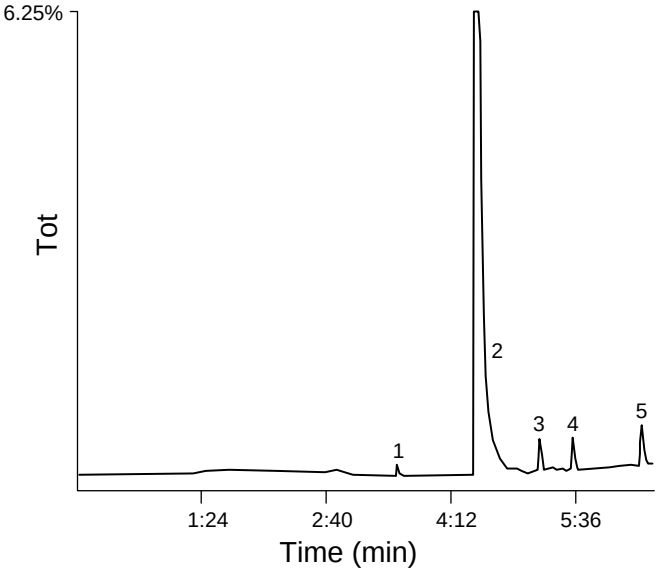

Figure 25 SPME/GC/ITMS analysis of a breath sample. The sample was taken after alcohol consumption. Fiber coating: Carbowax™/DVB. Peak assignment: 1, acetaldehyde; 2, ethanol; 3, acetone; 4, isoprene; 5, carbon disulfide.

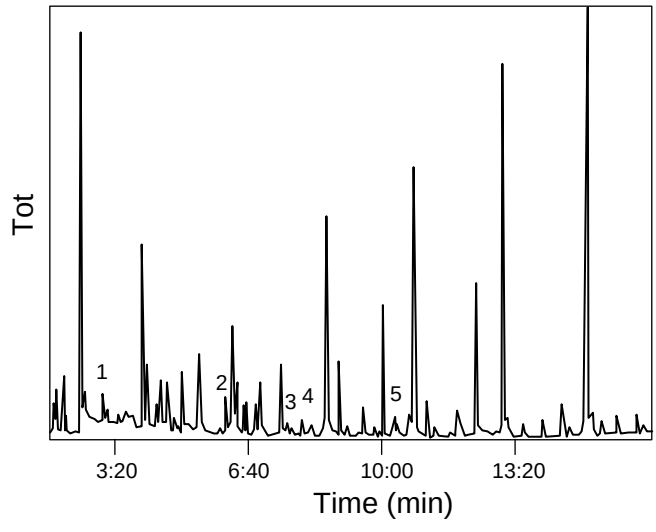

Figure 26 SPME/GC/ITMS analysis of human milk from a smoker. 1, aniline; 2, methylaniline; 3, dimethylaniline; 4, trimethylaniline; 5, nicotine.

Analysis for alcohol and drugs in body fluids is frequently performed in clinical and forensic laboratories. SPME is suitable for monitoring alcohol levels in both urine and blood. Drugs in body fluids are more difficult to analyze, since typically they are semivolatile substances. However, a number of published preliminary results indicate that headspace SPME analysis at an elevated temperature, or even direct extraction after dialysis of urine or blood serum, is possible. Most of the work to date was done with amphetamines,[55,56] but analysis of other drugs, including valproic acid,[57] cocaine, tricyclic antidepressants,[58] nicotine and local anesthetics have been also reported.

In forensic analysis, headspace SPME has been applied for the examination of suspected arson cases.[59,60] Figure 27 shows the total and selected ion chromatograms corresponding to a real arson sample and a gasoline standard, indicating the possibility of use of an accelerant.

4.7 Future Analytical Developments

SPME can be used as sampling and sample introduction technique for fast GC. Figure 28 demonstrates that using such an approach, volatile compounds in water can be analyzed in less than three minutes,[3] approximately ten times faster compared to the currently used purge and trap systems.

SPME is very well suited for field analysis. Field sampling and analysis reduce the costs and eliminate many possible sources of error associated with handling and storing of the samples. In addition, field analysis enables faster and better characterization of the problems at hand since analytical information is available immediately for evaluation and decision. It has been demonstrated that SPME can be succesfully used for fast GC analysis in the field with dedicated instrumentation.[61] Also, time-average sampling as described in section 2 would facilitate wider adoption of SPME in the on-site field monitoring.

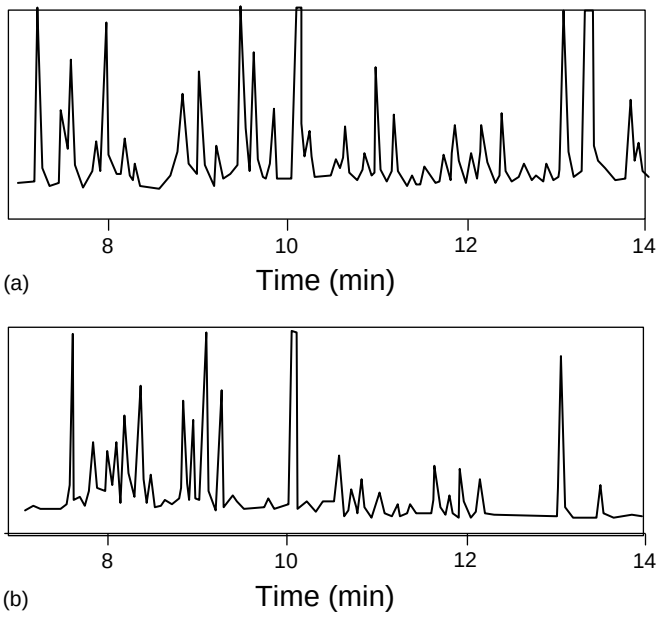

Figure 27 GC/ITMS chromatograms of headspace SPME of real fire debris sample: (a)/(b) selected ion trace for $m/z = 120$ for the sample and pure gasoline.

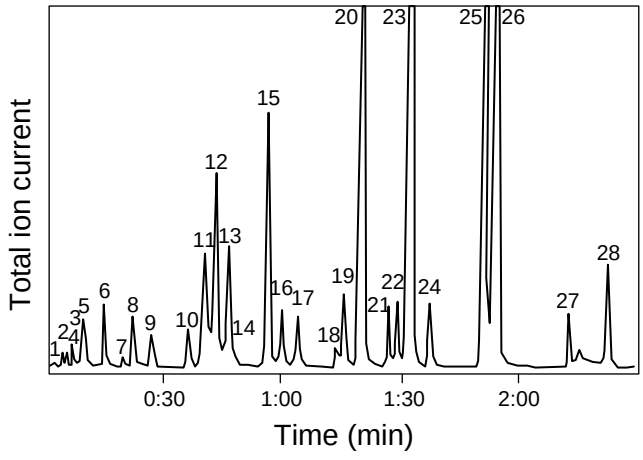

Figure 28 Separation of purgeables A, B and C on a VOCOL column. Conditions: 0 °C–30°/min–70°; 2.1 atm, dedicated injector, capacitor voltage 24 V, MS detector, mass range 45–250. 1, chloromethane; 2, vinyl chloride; 3, bromomethane; 4, chloroethane; 5, trichlorofluoromethane; 6, 1,1-dichloroethene; 7, dichloromethane; 8, 1,2-dichloroethene; 9, 1,1-dichloroethane; 10, trichloromethane; 11, 1,1,1-trichloroethane; 12, tetrachloromethane; 13, benzene; 14, 1,2-dichloroethane; 15, trichloroethene; 16, 1,2-dichloropropane; 17, bromodichloromethane; 18, 2-chloroethyl vinyl ether; 19, *cis*-1,3-dichloropropene; 20, toluene; 21, *trans*-1,3-dichloropropene; 22, 1,1,2-trichloroethane; 23, tetrachloroethylene; 24, dibromochloromethane; 25, chlorobenzene; 26, ethylbenzene; 27, tribromomethane; 28, 1,1,2,2-tetrachloroethane. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

Ion-exchange coating materials and crown ether ligands have been used to selectively extract inorganic ions from aqueous matrices, allowing not only quantitation, but also speciation.[13,62] Distribution of chemicals in complex samples can be characterized using SPME, as discussed in section 4.8, because of its equilibrium principle. Polyacrylic acid coating has been developed for the analysis of basic proteins.[23] The small size of the fiber will allow speciation of minute sample volumes, for example down to single cell dimensions.

4.8 Physicochemical Applications

SPME can be applied not only for extraction purposes, but also to perform measurements which better characterize the extraction system. The measurements can

include studying the properties of the fiber coating and investigation of multiphase equilibria in the matrix. In addition, since detailed mathematical treatment of the process is available, SPME can be applied to study the parameters involved in the extraction, including diffusion coefficients in both the coating and the extracted phase, as well as various distribution constants.

For understanding solute–solvent interactions at the molecular level and the thermodynamic processes involved in forming the solution, the study of infinite-dilution activity coefficients of probe solutes in a polymer phase is a fundamental approach. SPME can be a useful tool for determining these coefficients.[4] The stationary phases of interest can be coated on fibers made of suitable materials (fused silica, stainless steel, etc.). The SPME device with a selected coating can be used to extract a group of probe compounds, which are then separated on a standard commercially available column and quantified by GC/MS.

The SPME measurements of infinite-dilution activity coefficients of probe solutes in a selected stationary phase consist of two steps. The first step is to measure the infinite-dilution partition coefficients of the probe solutes between the stationary phase and the gas phase. Then, from these partition coefficients, the corresponding activity coefficients are determined.

Investigation of analyte distributions in natural systems is critical for many disciplines of science including environmental chemistry. Information gained during these studies is also very important to analytical chemists, since it facilitates the development of optimum sampling protocols and can lead to more rational optimization of extraction parameters. One approach to study multiphase equilibria involves separation of the individual phases after equilibrium has been reached and analysis of the individual phases to determine the concentrations of target compounds. For example, SPME was used to study the concentration of organic substances from octanol-saturated water.[63] However, separation of the phases is not always necessary. SPME has been applied to measure the distribution of chemicals between the components of a system consisting of water and dissolved humic organic matter (HOM).[64] Concentrations of the free forms of the target compounds were determined in pure water and water containing HOM. Based on a comparison of results obtained for the two samples, the partitioning of the analytes between water and HOM was determined. Good agreement was found between SPME and standard techniques, which indicates that it is well suited for such investigations.

Partitioning kinetics is another important issue for analytical chemists, since it determines the extraction rates. Interaction of nonpolar organic analytes with dissolved polymers is fast because physisorption does not require high activation energies. SPME can be used to investigate the kinetics of the equilibration process of organic pollutants between water and dissolved HOM. In such study the sample containing native analytes is spiked with isotopically labeled analogs. Rapid extraction from a stirred HOM–water sample can monitor concentration changes of target analytes with respect to time. Figure 29 illustrates the change of the concentration of labeled compound in water versus native analyte as a function of time. The curve indicates rapid equilibration of the species between water and HOM.

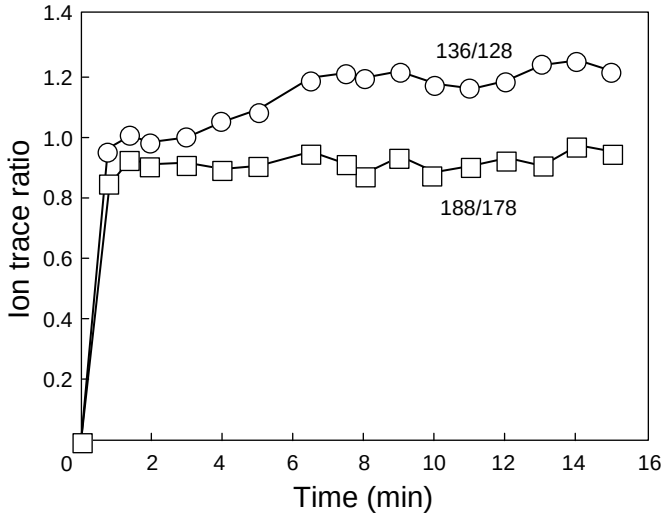

Figure 29 Kinetics of the sorption of naphthalene and anthracene onto Dissolved Organic Matter (DOM) (100 ppm) from water measured by SPME. *Y*-axis represent ratio of GC signals for deuterated spike and nondeuterated compound. Matrix: diluted wastewater with native PAHs. (Reproduced by permission from *Solid Phase Microextraction: Theory and Practice* Pawliszyn © 1997 Wiley-VCH.)

ACKNOWLEDGMENTS

The SPME work in my laboratory was supported by grants from the Natural Sciences and Engineering Research Council of Canada, Supelco, Varian, US Environmental Protection Agency Office of Exploratory Research, Dow Chemical, Imperial Oil, SRI Instruments, Chrompack, the Waterloo Centre for Groundwater Research, and the Ontario Ministry of Environment and Energy.

ABBREVIATIONS AND ACRONYMS

BTEX	Benzene, Toluene, Ethylbenzene, *m*-Xylene, *o*-Xylene, *p*-Xylene
CE	Capillary Electrophoresis
DDT	1,1,1-Trichloro-2,2-bis(*p*-chlorophenyl)ethane

DL	Detection Limit
DVB	Divinylbenzene
ECD	Electron Capture Detection
FID	Flame Ionization Detector
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HOM	Humic Organic Matter
HPLC	High-performance Liquid Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LOD	Limit of Detection
LTPRI	Linear Temperature Programmed Retention Index
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NPD	Nitrogen–Phosphorus Detection
PA	Poly(acrylate)
PAH	Polynuclear Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PDAM	1-Pyrenyldiazomethane
PDMS	Poly(dimethylsiloxane)
PDMS/DVB	Poly(dimethylsiloxane)/Divinylbenzene
PEEK	Polyetheretherketone
PTV	Programmed Temperature Vaporization
RSD	Relative Standard Deviation
SFC	Supercritical Fluid Chromatography
SPI	Septum-equipped Programmable Injector
SPME	Solid-phase Microextraction
SPME/GC	Solid-phase Microextraction/Gas Chromatography
SPME/GC/ITMS	Solid-phase Microextraction/Gas Chromatography/Ion Trap Mass Spectrometry
SRM	Standard Reference Material
TPH	Total Petroleum Hydrocarbons
TR	Template Resin
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
VOC	Volatile Organic Compound

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Soxhlet and Ultrasonic Extraction of Organics in Solids

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Food **(Volume 5)**
Sample Preparation for Food Analysis, General • Sample Preparation, Headspace Techniques

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction

Pharmaceuticals and Drugs **(Volume 8)**
Robotics and Laboratory Automation in Pharmaceuticals Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

REFERENCES

1. J. Pawliszyn, *Solid Phase Microextraction. Theory and Practice*, Wiley-VCH, New York, 1997.
2. Z. Zhang, J. Pawliszyn, 'Quantitative Extraction Using an Internally Cooled SPME Device', *Anal. Chem.*, **67**, 34–43 (1995).
3. T. Gorecki, J. Pawliszyn, 'Sample Introduction Approaches for SPME', *Anal. Chem.*, **67**, 3265–3274 (1995).
4. Z. Zhang, J. Pawliszyn, 'Studying Activity Coefficients of Probe Solutes in Selected Liquid Polymer Coatings', *J. Phys. Chem.*, **100**, 17648–17654 (1996).
5. P. Martos, A. Saraullo, J. Pawliszyn, 'Estimation of Air/Coating Distribution Coefficients for Solid Phase Microextraction Using Retention Indexes from Linear Temperature-programmed Capillary Gas Chromatography', *Anal. Chem.*, **69**, 402–408 (1997).
6. P. Martos, A. Saraullo, J. Pawliszyn, 'Water Analysis by SPME Based on Physical Chemical Properties of the Coating', *Anal. Chem.*, **69**, 1992–1998 (1997).
7. R. Schwarzenbach, P. Gschwend, D. Imboden, *Environmental Organic Chemistry*, John Wiley & Sons, New York, 109–123, 1993.

8. P. Martos, J. Pawliszyn, 'Calibration of SPME for Air Analysis Based on Physical Chemical Properties of the Coating', *Anal. Chem.*, **69**, 206–215 (1997).
9. A.D. Young, *Boundary Layers*, BSP Professional Books, Oxford, 1989.
10. V. Mani, (Supelco), private communication.
11. K. Buchholz, J. Pawliszyn, 'Determination of Phenols by SPME and GC Analysis', *Environ. Sci. Technol.*, **27**, 2844–2848 (1993).
12. T. Górecki, J. Pawliszyn, 'Strategies for the Analysis of Polar Solvents in Liquid Matrices', *Anal. Chem.*, **70**, 19–27 (1998).
13. C. Jia, Y. Luo, J. Pawliszyn, 'SPME combined with HPLC for Determination of Metal Ions Using Crown Ether as Selective Extracting Reagent', *J. Microcolumn. Sep.*, **10**, 167–173 (1998).
14. L. Pan, J. Pawliszyn, 'Derivatization/SPME – New Approach to Polar Analytes', *Anal. Chem.*, **69**, 196–205 (1997).
15. Y. Cai, J. Bayona, 'Determination of Methylmercury in Fish and River Water Samples Using In Situ Sodium Tetraethylborate Derivatization Followed by SPME and GC-MS', *J. Chromatogr.*, **696**, 113 (1995).
16. J. Poerschmann, F-D. Kopinke, J. Pawliszyn, 'SPME to Study the Sorbtion of Organotin Compounds onto Particulate and Dissolved Humic Organic Matter', *Environ. Sci. Technol.*, **31**, 3629–3636 (1997).
17. L. Pan, M. Adams, J. Pawliszyn, 'Determination of Fatty Acids Using SPME', *Anal. Chem.*, **67**, 4396–4403 (1995).
18. J. Chen, J. Pawliszyn, 'SPME Coupled to HPLC', *Anal. Chem.*, **67**, 2530–2533 (1995).
19. H. Daimon, J. Pawliszyn, 'Effect of Heating the Interface on Chromatographic Performance of SPME/HPLC', *Anal. Comm.*, **33**, 365–367 (1996).
20. T. Górecki, J. Pawliszyn, 'The Effect of Sample Volume on Quantitative Analysis by SPME', *Analyst*, **122**, 1079–1086 (1997).
21. M. Chai, J. Pawliszyn, 'Analysis of Environmental Air Samples by SPME/GC/ITMS', *Environ. Sci. Technol.*, **29**, 693–701 (1995).
22. L. Wang, 'Determination of Amphetamine and Methamphetamine in Urine and Blood Plasma by Headspace Solid Phase Microextraction', MSc Thesis, University of Waterloo, Waterloo, Canada, 1996.
23. J-L. Liao, C-M. Zeng, S. Hjerten, J. Pawliszyn, 'SPME of Biopolymers: Adsorption of Basic Proteins onto a Fiber Coated with Polyacrylic Acid', *J. Microcol. Sep.*, **8**, 1–4 (1996).
24. K. Buchholz, J. Pawliszyn, 'Optimization of SPME Conditions for Determination of Phenols', *Anal. Chem.*, **66**, 160–167 (1994).
25. H. Daimon, J. Pawliszyn, 'High Temperature Water Extraction Combined with SPME', *Anal. Com.*, **33**, 421–424 (1996).
26. Z. Zhang, J. Poerschmann, J. Pawliszyn, 'Direct SPME of Complex Aqueous Samples with Hollow Fiber Membrane Extraction', *Anal. Com.*, **33**, 129–131 (1996).
27. B. MacGillivray, P. Fowlie, C. Sagara, J. Pawliszyn, 'Headspace SPME vs. Purge and Trap for the Determination of Substituted Benzene Compounds in Water', *J. Chromatogr. Sci.*, **32**, 317–322 (1994).
28. T. Górecki, J. Pawliszyn, 'Pesticides by SPME. Results of a Round Robin Test', *Analyst*, **121**, 1381–1386 (1996).
29. P. Martos, J. Pawliszyn, 'Sampling and Determination of Formaldehyde Using SPME with On-fiber Derivatization', *Anal. Chem.*, **70**, 2311–2320 (1998).
30. J. Dean, W. Tomlinson, V. Makovskaya, R. Cumming, M. Hetheridge, M. Comber, 'SPME as a Method for Estimating the Octanol–Water Partition Coefficient', *Anal. Chem.*, **68**, 130–133 (1996).
31. C. Arthur, L. Killam, K. Buchholz, J. Pawliszyn, 'Automation and Optimization of SPME', *Anal. Chem.*, **64**, 1960–1966 (1992).
32. A. Boyd-Boland, J. Pawliszyn, 'SPME of Nitrogen-containing Herbicides', *J. Chromatogr.*, **704**, 163–172 (1995).
33. B. MacGillivray, 'Analysis of Substituted Benzenes in Environmental Samples by Headspace Solid Phase Microextraction', MSc Thesis, University of Waterloo, Waterloo, Canada, 1996.
34. A. Saraullo, 'Determination of Petroleum Hydrocarbons in the Environment by SPME', MSc Thesis, University of Waterloo, Waterloo, Canada, 1996.
35. T. Nilsson, F. Pelusio, L. Montanarella, B. Larsen, S. Facchetti, J. Madsen, 'An Evaluation of SPME for Analysis of Volatile Organic Compounds in Drinking Water', *HRC*, **18**, 617 (1995).
36. D. Potter, J. Pawliszyn, 'Rapid Determination of Polyaromatic Hydrocarbons and Polychlorinated Biphenyls in Water Using SPME', *Environ. Sci. Technol.*, **28**, 298–305 (1994).
37. J. Langenfeld, S. Hawthorne, D. Miller, 'Quantitative Analysis of Fuel Related Hydrocarbons in Surface Water and Wastewater Samples by SPME', *Anal. Chem.*, **68**, 144 (1996).
38. S. Johanssen, J. Pawliszyn, 'Trace Analysis of Hetero Aromatic Compounds (NSO) in Water and Polluted Groundwater by SPME', *HRC*, **19**, 627–632 (1996).
39. R. Eisert, K. Levsen, 'SPME Coupled to GC: A New Method for the Analysis of Organics in Water', *J. Chromatogr. A.*, **733**, 143 (1996).
40. I. Barnabas, J. Dean, I. Fowlis, S. Owen, *J. Chromatogr.*, **705**, 305 (1995).
41. K. Graham, L. Sarna, G. Webster, J. Graynor, H. Ng, 'SPME of the Herbicide Metolachlor in Runoff and Tile-drainage Water Samples', *J. Chromatogr.*, **725**, 129 (1996).
42. A. Boyd-Boland, S. Magdic, J. Pawliszyn, 'Simultaneous Determination of 60 Pesticides in Water Using SPME and GC/MS', *Analyst*, **121**, 929–937 (1996).

43. P. Popp, S. Mothes, L. Brueggemann, 'Determination of Pesticides in Water by SPME/GC', *Vom Wasser*, **85**, 229 (1995).
44. Y. Cai, J. Bayona, 'Determination of Methylmercury in Fish and River Water Samples Using In-situ Sodium Tetraethylborate Derivatization followed by SPME/GC/MS', *J. Chromatogr. A*, **696**, 113 (1995).
45. T. Górecki, J. Pawliszyn, 'Determination of Tetraethyllead and Inorganic Lead in Water by SPME/GC', *Anal. Chem.*, **68**, 3008–3014 (1996).
46. A. Boyd-Boland, J. Pawliszyn, 'SPME Coupled with HPLC for the Determination of Alkylphenol Ethoxylate Surfactants in Water', *Anal. Chem.*, **68**, 1521–1529 (1996).
47. S. Scypinski, A-M. Smith, L. Nelson, S. Shaw, private communication.
48. D. Page, G. Lacroix, 'Application of SPME to the Headspace GC Analysis of Halogenated Volatiles in Selected Foods', *J. Chromatogr.*, **648**, 199 (1993).
49. X. Yang, T. Peppard, 'SPME for Flavour Analysis', *J. Agric. Food Chem.*, **42**, 1925 (1994).
50. A. Steffen, J. Pawliszyn, 'Analysis of Flavour Volatiles Using Headspace SPME', *J. Agric. Food Chem.*, **44**, 2187–2193 (1996).
51. F. Pelusio, T. Nilsson, L. Montanarella, R. Tilio, B. Larsen, S. Facchetti, J. Madsen, *J. Agric. Food Chem.*, **43**, 2138 (1995).
52. V. Mani, C. Woolley, *LC-GC*, **13**, 734 (1995).
53. S. Hawthorne, D. Miller, C. Arthur, J. Pawliszyn, 'Solventless Determination of Caffeine in Beverages Using SPME with Fused Silica Fibres', *J. Chromatogr.*, **603**, 185–191 (1992).
54. C. Grote, J. Pawliszyn, 'SPME for the Analysis of Human Breath', *Anal. Chem.*, **69**, 587–596 (1997).
55. M. Yashiki, Y. Iwasaki, K. Hara, T. Kojima, 'Detection of Amphetamines in Urine Using Headspace SPME and Chemical Ionization Selected Ion Monitoring', *Forensic Sci. Int.*, **76**, 169 (1995).
56. L. Wang, 'Determination of Amphetamine and Methamphetamine in Urine and Blood Plasma by Headspace Solid Phase Microextraction', MSc Thesis, University of Waterloo, Waterloo, Canada, 1996.
57. M. Krogh, K. Johansen, F. Tonnesen, K. Rasmussen, 'SPME for the Determination of the Free Concentration of Valproic Acid in Human Plasma by Capillary GC', *J. Chromatogr.*, **673**, 299 (1995).
58. T. Kumazawa, X-P. Lee, M-C. Tsai, H. Seno, A. Ishii, K. Sato, 'Detection of Cocaine in Human Urine by SPME and Capillary GC with NPD', *Jpn. J. Forensic Toxicol.*, **13**, 25 (1996).
59. K.G. Furton, J. Bruna, 'A Simple, Inexpensive, Rapid, Sensitive and Solventless Technique for the Analysis of Accelerants in Fire Debris Based on SPME', *HRC*, **18**, 625 (1995).
60. A. Steffen, J. Pawliszyn, 'Determination of Liquid Accelerants in Arson Suspected Fire Debris Using Headspace SPME', *Anal. Com.*, **33**, 129–131 (1996).
61. T. Górecki, J. Pawliszyn, 'Field Portable SPME/Fast GC System for Trace Analysis', *Field Anal. Chem. Technol.*, **1**, 277–284 (1997).
62. E. Otu, J. Pawliszyn, 'Solid Phase Microextraction of Metal Ions', *Mikrochim. Acta*, **112**, 41–46 (1993).
63. P. Popp, A. Paschke, U. Schroter, G. Oppermann, 'Application of SPME/GC/FID to the Determination of Benzene and Halogenated Benzenes in Pure and Octanol Saturated Water', *Chemia Analityczna (Warsaw)*, **40**, 897 (1995).
64. J. Poerschmann, Z. Zhang, F. Kopinke, J. Pawliszyn, 'SPME for Determining the Distribution of Chemicals in Aqueous Matrices', *Anal. Chem.*, **69**, 597–600 (1997).

Soxhlet and Ultrasonic Extraction of Organics in Solids

G.R. Barrie Webster
Websar Laboratories, Inc., Ste. Anne, MB, Canada

1 Introduction **1**
2 Extraction Methods **2**
2.1 Soxhlet Extraction 2
2.2 Ultrasonic Extraction 3
3 Method Development and Comparison with Other Extraction Methods **7**
3.1 Benchmark Methods for Comparison 7
3.2 Comparison Issues 7
3.3 Extraction Variables 7
3.4 Making Use of the Data 8
4 Quality Control and Troubleshooting **8**
4.1 Data Quality Objectives 8
4.2 Appropriate Surrogate Standards and Concentrations 9
4.3 Matrix Spiking to Determine Extraction Efficiency 10
4.4 Coextractives and Contaminants 10
5 Environmental Applications **10**
5.1 Applicable Matrices for Extraction by these Methods 10
5.2 Interface with Toxicology 10
5.3 Confounding Phenomena to be Considered 10
6 Regulatory Issues **11**
6.1 Value of Total Residue Data 11
6.2 Relationship between Solvent-extractable Residue and Toxicology 11
6.3 New Improved Soxhlet Extraction Methods 11
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**
Note Added in Proof **13**

Soxhlet extraction is the removal and recovery of organic analytes from a permeable solid matrix by means of a solvent which is continually evaporated from a still-pot and condensed in such a manner that it falls into and permeates through the matrix which itself is held in a permeable container in a siphonable chamber. Soxhlet extraction has been used for over 120 years (since 1879) and is commonly used as a benchmark for total extractable organic residues.

Ultrasonic extraction is the removal and recovery of organic analytes from a permeable solid matrix by means of a solvent which is energized by sound energy at frequencies in excess of those audible to the human ear. Energy may be introduced into the sample by means of an ultrasonic probe which is inserted into the sample or an ultrasonic bath into which the sample plus solvent is immersed. The ultrasonic energy may be strong enough to disrupt and pulverize the matrix, thus increasing the extractability of the analyte.

The operation of the Soxhlet extractor is intuitively easy to grasp. Extraction is viewed as being complete or close to complete.

Ultrasonic extraction is convenient and straightforward. Ultrasonic baths provide relatively reliable sources of extraction energy.

Soxhlet extraction is solvent intensive and requires laboratory facilities in which flammable and toxic solvents can be used safely. The glassware required is expensive and vulnerable to breakage. Cooling of the condenser of the Soxhlet apparatus requires a constant supply of cooling water. Residues are concentrated in the still-pot and are thus vulnerable to thermal degradation. Only total residue can be extracted by Soxhlet extraction; little or no speciation is possible.

Ultrasonic extraction must be performed under conditions which protect the hearing of laboratory personnel. The ultrasonic generator may generate heat which may affect the extraction process. Ultrasonic probes become pitted when used for the extraction of abrasive matrices such as soil, thus changing the energy generated and affecting the extraction efficiency of the procedure. Ultrasonic baths provide less intense extraction energy. Extraction of total residue is the aim of ultrasonic extraction; little or no speciation is possible.

1 INTRODUCTION

Extraction techniques can have as their objective the extraction of all of the analyte, no matter what its environment in the sample, or the selective extraction of residues from defined environments within the sample matrix. Both Soxhlet extraction and ultrasonic extraction have as their goal the complete extraction of specified analytes or groups of analytes from solvent-accessible solid matrices. The methods are intended to be exhaustive, and performance that does not meet these expectations is usually viewed as being unsuccessful. In samples containing medium to high concentrations ($>20\,mg\,kg^{-1}$) of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the analyte or analyte group, virtually all residue will be readily extracted; it is unlikely to be adsorbed in such a way that extraction is prevented. In relating low residue levels ($<20\,mg\,kg^{-1}$) to toxicological interpretations, it is important to remember that all residues are not created equal. Those having the most significance toxicologically speaking are most often the freely available residues, the unassociated ones. Residues which are complexed, adsorbed, sequestered, or incorporated into a matrix either physically or covalently are less available or unavailable biologically, but are residues nonetheless. The degree to which the total residue is extracted from particular matrices will depend a great deal on the extraction conditions and the technique employed. Thus Soxhlet extraction may be carried out in conjunction with hydrolytic conditions to free residues conjugated with biological molecules or soil humic material. Ultrasonic extraction may need to be carried out with sufficient energy to assist the physical break-up of the matrix to free entrapped residues.

Soxhlet extraction is the removal and recovery of organic analytes from a permeable solid matrix by means of a solvent which is continually evaporated from a still-pot and condensed in such a manner that it falls into and permeates through the matrix which itself is held in a permeable container in a siphonable chamber. In Soxhlet extraction, the solvent is continuously purified and recycled and the extracted material is concentrated in the still-pot. Soxhlet extraction has been used for over 120 years and is commonly used as a benchmark technique for total extractable organic residues.

Ultrasonic extraction is the removal and recovery of organic analytes from a permeable solid matrix by means of a solvent that is energized by sound energy at frequencies in excess of those audible to the human ear. Energy may be introduced into the sample by means of an ultrasonic probe that is inserted into the sample or an ultrasonic bath into which the sample plus solvent is immersed. Officially recognized methods specify the former; however, the latter method has been used effectively, e.g. for herbicide residue extraction from soils.[1] The ultrasonic energy may be strong enough to disrupt and pulverize the matrix, thus increasing the extractability of the analyte. In ultrasonic extraction, the solvent remains with the sample and is heated only if sufficient energy is supplied to overcome heat dissipation at ambient temperature.

2 EXTRACTION METHODS

2.1 Soxhlet Extraction

Soxhlet extraction makes use of a three-piece apparatus (normally constructed of borosilicate glass) and

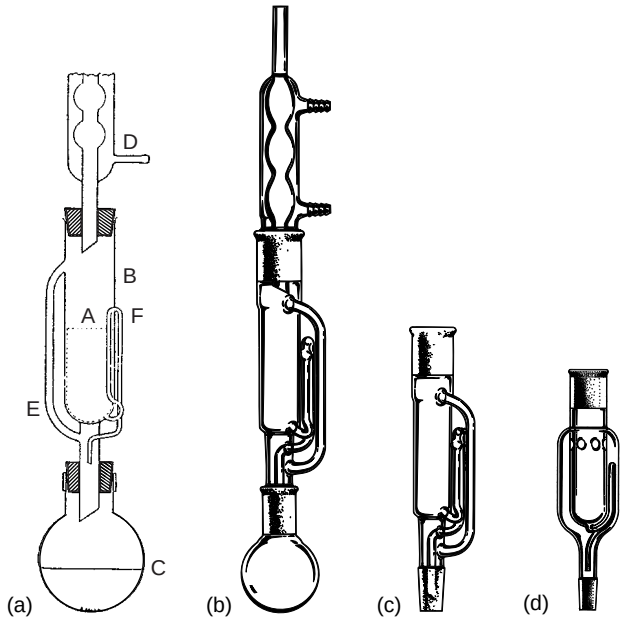

Figure 1 Soxhlet extraction apparatus: (a) the old style with the cork or rubber stopper connectors (ref. 2 with permission), A = extraction thimble showing the top of the thimble below the top of the siphon tube, B = sample-holding siphonable chamber, C = still-pot showing the optimum solvent level, D = condenser unit, E = solvent vapor by-pass tube, F = extract solution siphon tube; (b) the modern style Soxhlet extraction apparatus with ground glass joints (ref. 3 with permission); (c) modern sample-holding siphonable chamber with ground glass joints; (d) sample-holding siphonable chamber in which there is no solvent vapor by-pass tube.

a heat source. The still-pot, an appropriately sized round-bottomed flask, has a close-fitting joint to allow connection with the sample-holding siphonable chamber with solvent vapor by-pass tube, topped with the condenser unit (Figure 1). The heat source is usually an electric heating mantle sized to fit the still-pot. Early versions of this apparatus used drilled rubber or cork stoppers to join the sections; however, since the 1950s, Soxhlet extractors have been available with precision ground-glass joints to avoid escape of solvent vapors and to avoid contamination of the extract with material which is extracted inadvertently from the stopper matrix. Soxhlet apparatuses come in a variety of sizes from micro versions with 25-mL still-pots to gigantic pilot-plant scale multiliter-sized models. Custom construction is possible; however, the most common sizes are available from glassware suppliers and have 100- or 125-, 250-, 500-, or 1000-mL still-pots with the rest of the apparatus proportioned to fit; micro sized Soxhlet extractors are also now readily available and are suitable for the 1–5 g samples increasingly used to enable solvent volumes per sample to be reduced. Samples to be extracted are usually contained in paper or glass thimbles that fit inside

the sample-holding chamber, although in special circumstances thimbles can be made of other inert materials (e.g. Teflon® or stainless steel) as long as they have been suitably precleaned and pose no risk of contamination of the sample or sample extract. Thimbles are commonly available from the same supplier as the apparatus itself.

Some Soxhlet procedures make use of flat-bottomed still-pots such as Erlenmeyer flasks with hotplate heating sources (e.g. AOAC (Association of Official Analytical Chemists) Method 920.172, see Table 2).

Samples are placed in the thimbles and the tops are protected with paper or glass-wool covers to prevent splashing as the condensed solvent falls from the condenser. To ensure compatibility with the extraction solvent, the sample is often predried, pulverized, or mixed with a filter-aid such as sand or Celite® to facilitate solvent permeation. It may also be mixed with anhydrous sodium sulfate (Na_2SO_4). The sample in the thimble must not exceed the height of the solvent siphon (Figure 1).

Soxhlet extraction is described in the USEPA (United States Environmental Protection Agency) Method 3540C[4] as 'a procedure for extracting nonvolatile and semivolatile organic compounds from solids such as soils, sludges, and wastes' and one that 'ensures intimate contact of the sample matrix with the extraction solvent'. The method is 'applicable to the isolation and concentration of water-insoluble and slightly water soluble organics' and extracted analytes are then identified and quantified usually by means of a chromatographic technique.

In order to render a solid matrix permeable to the extracting solvent, one of two approaches may be followed:

- The matrix is either air-dried to remove water, or it is mixed with the drying agent, anhydrous sodium sulfate, to sequester water prior to the placing of the matrix in the extraction thimble; the extraction solvent is immiscible with water.
- The matrix is extracted with the water in place but the extracting solvent is prepared to render it compatible with the water; usually a pair of miscible solvents is used, one of them being water miscible and the other not.

The first approach allows a water-immiscible solvent to be used without fear of in situ water preventing contact with residues sorbed or otherwise associated with the matrix (e.g. soil). In view of the complex behavior of soil matrices in the presence and absence of water, it is often preferable to leave the water in place and use the second approach. The drying action of the first approach may intensify sorption of the analyte(s) to the matrix and sometimes render the residues unextractable. If it is important to determine the moisture content of the sample, a portion of it should be taken for this determination before the sample is dried or mixed with the drying agent (see section 2.2.2).

Method 3540C specifies the use of a 500-mL still-pot; however, the analyst would be well advised to modify the method to allow the use of smaller scale extraction providing that it is still compatible with the limits of detection and quantification desired, and that the sample being extracted is still representative of the matrix being analyzed. The larger sample sizes are sometimes used to compensate for sample inhomogeneity.

Extraction solvents used for Soxhlet extraction need to be of sufficient purity that the extract is able to be concentrated for subsequent steps without the impurities interfering with the analysis. For process samples, reagent grade solvent may be sufficient; however, environmental matrices will require the use of high-purity distilled-in-glass grade solvent. Typical solvents used are given in Table 1.

Soxhlet extraction dates from the 1870s; however, many approved standard methods of analysis incorporate this extraction technique including a significant number of the AOAC Official Methods[5] (Table 2). This table is extensive and the extraction methods are somewhat out of date in that many use solvents no longer permitted for use for occupational toxicological reasons, such as benzene and chloroform. However, they remain official standard methods of the AOAC in which Soxhlet extraction plays an important role.

2.2 Ultrasonic Extraction

The ultrasonic procedure promotes intimate contact of matrix and extraction solvent through energetic mixing using intense ultrasound as the energy source.

Ultrasonic extraction can be conducted by one of two procedures. Either an ultrasonic probe is used to introduce ultrasonic energy into the mixture being

Table 1 Typical solvents used in Soxhlet extraction

Analytes	Matrix	Solvent
Organics such as pesticide residues and industrial chemicals	soil/sediment	acetone–hexane (1:1) or dichloromethane–acetone (1:1)
Organics such as pesticide residues and industrial chemicals	other environmental matrices	dichloromethane or toluene–methanol (10:1)

extracted, or the mixture in a glass container is immersed in an ultrasonic water bath. Although the energy introduced using the probe technique is of greater local intensity, some authors believe that the bath technique is more likely to provide consistent extraction results. Extraction of abrasive matrices such as soil or sediment using a probe can lead to erosion of the probe tip, resulting in loss of the ability of the probe to deliver sonic energy efficiently to the system being sonicated.

Ultrasonic extraction is regarded by the USEPA Method 3550 as being not as rigorous as other extraction methods for soils and solids. It is therefore of utmost importance that, if the method is used, the method and the manufacturer's directions be followed carefully to achieve good extraction efficiency. Successful use of this extraction technique demands that the device has minimum power of 300 W and is equipped with the appropriate disrupter horn, tuned, and in good working condition. The horn tip must be inspected regularly for wear, as the shape of the horn tip will influence the effectiveness of the technique. The sample must be a free flowing powder, achieved for example by thorough mixing with anhydrous sodium sulfate prior to the addition of the solvent. The horns used for medium/high analyte concentrations are not interchangeable with those used for low concentrations.

Table 2 Official AOAC methods that make use of Soxhlet extraction

Analytes and matrices	Solvent	Official method
Food		
Polysorbate 60 in shortening, oils, dressings (on Hyflo Super-Cel™)	$CHCl_3$ – abs EtOH (93 : 7)	AOAC 974.11
Urine stains on foods and containers	acetone	AOAC 973.64
Sorbitol in foods	CH_3OH (anhydrous)	AOAC 973.28
Light filth in cracked wheat and flours	$CHCl_3$	AOAC 971.31
Aflatoxins in cocoa beans	*n*-hexane	AOAC 971.23
Sterols in macaroni products	$CHCl_3$	AOAC 969.14
Fat in cacao products	petroleum ether (30–60 °C)	AOAC 963.15
Piperonyl butoxide residues	CH_3Cl	AOAC 960.43
DDT pesticides residues	Et_2O or benzene	AOAC 960.41
Parathion pesticide residues	benzene	AOAC 960.42
Diethylstilbestrol in feeds	EtOH in CH_3Cl (7% v/v) (*sic*)	AOAC 956.10
Lactose in bread (with Filter-Cel™)	EtOH – water (126 : 61)	AOAC 952.05
Fat (crude) in nuts and nut products	Et_2O	AOAC 948.22
Oil in cereal adjuncts: petroleum ether extraction method	petroleum ether (35–60 °C)	AOAC 945.16
Saccharin in foods (baked goods)	petroleum ether	AOAC 941.10
Caffeine in tea	EtOH	AOAC 925.16
Caffeine in roasted coffee	EtOH	AOAC 920.94
Ether extractables in prepared mustard	Et_2O (anhydrous)	AOAC 920.172
Formulation analysis		
Chlordane in granular pesticide formulations	acetone	AOAC 973.16
Propoxur technical and pesticide formulations	CH_3CN	AOAC 984.12
Urea and methyleneureas (water-soluble) in fertilizers	acetone	AOAC 983.01
Parathion in pesticide formulations	Et_2O	AOAC 978.08
Aldicarb in pesticide formulations	CH_2Cl_2	AOAC 974.04
Benfluralin or trifluralin in pesticide formulations	$CHCl_3$, acetone	AOAC 973.13, AOAC 973.14
Chloral – dimethyl (DCPA) in pesticide formulations	benzene	AOAC 970.05
Heptachlor in pesticide formulations	pentane	AOAC 968.04, AOAC 962.07
Chlordane (technical) and pesticide formulations	pentane, benzene	AOAC 965.14, AOAC 962.05
Chlorine (total) in organohalogen pesticide formulations	acetone	AOAC 961.04
β-Naphthol in color additives	isopropyl ether	AOAC 950.70
Amines (nonvolatile unsulfonated) in color additives	petroleum ether	AOAC 950.69

Table 2 (*continued*)

Analytes and matrices	Solvent	Official method
Extracts of color additives	isopropyl ether, petroleum ether	AOAC 950.67
γ-BHC (lindane) in pesticide formulations	Et_2O	AOAC 949.05
α-Naphthylthiourea in rodenticide	acetone	AOAC 948.03
Pyrethrin in pesticide formulations	petroleum ether (30–60 °C)	AOAC 936.05
Santonin in *Santonica* (levant worm seed)	benzene	AOAC 932.27
Natural product analysis		
α-Tocopherol acetate in supplemental vitamin E concentrates	*n*-hexane–cyclohexane (1 : 1)	AOAC 989.09
α-Tocopherol and α-Tocopherol acetate in foods and feeds	EtOH abs, petroleum ether (35–60 °C)	AOAC 971.30, AOAC 948.26
Griseofulvin in feeds	CH_3Cl	AOAC 966.30
Volatile oil in spices	Et_2O	AOAC 962.17
Reserpine–rescinnamine group alkaloids in *Rauwolfia serpina* drugs	EtOH	AOAC 956.09
Lignin in plants	EtOH–benzene (1 : 2), (32 : 68)	AOAC 949.04, AOAC 932.01
Toxicology related analysis		
N-Nitrosamines in baby bottle rubber nipples	CH_2Cl_2	AOAC 987.05
N-Nitrosodibutylamine in latex infant pacifiers	CH_2Cl_2	AOAC 986.01
Hexachlorobenzene and mirex in adipose tissue	hexane	AOAC 980.22
Sporicidal activity of disinfectants	$CHCl_3$	AOAC 966.04
General use methods		
Soluble matter	$CHCl_3$	AOAC 950.81

The tip of the ultrasonic horn is positioned just below the surface of the extraction solvent, yet clear of the solid matrix material. Three extractions are performed with the apparatus in the specified pulse mode. Very active mixing of the slurry so produced should be observed during the procedure.

Recognized official methods using the ultrasonic procedure have been published by the AOAC (Table 3) and the USEPA (Table 4) with adjustments for the extraction of medium/high concentrations of analyte(s) and low concentrations in Table 5.

2.2.1 United States Environmental Protection Agency Method 3550 – Ultrasonic Extraction

This method is designed specifically for the extraction of nonvolatile and semivolatile organics from solids (soils, sludge, and wastes).[6] Common solvents suitable for a variety of analytes are given in Table 4. Adjustments for medium-high or low concentrations of analytes are given

Table 3 Ultrasonic extraction method approved by the AOAC

Matrix and analyte(s)	Solvent	Official method
Munition residues in soil	CH_3CN	AOAC 991.09

Table 4 The USEPA method 3550: ultrasonic extraction methods for nonvolatile and semivolatile organic analytes in soils, sludges, and wastes

Analyte(s)	Solvent
Semivolatile organics	acetone–DCM[a] (1 : 1 v/v) or acetone–hexane (1 : 1 v/v)
Organochlorine pesticides	acetone–hexane (1 : 1 v/v) or acetone–DCM[a] (1 : 1 v/v)
PCBs	acetone–hexane (1 : 1 v/v), acetone–DCM[a] (1 : 1 v/v) or hexane
OPs[b]	extensive validation required on real-world samples before use of any solvent system
Various analytes	other solvent systems provided that adequate performance can be demonstrated for the analyte in the sample matrix

[a] DCM = dichloromethane.
[b] The EPA has not validated this method for organophosphorus compounds (OPs) from solid matrices. There is also concern that the ultrasonic energy may be sufficient to cause breakdown of some OPs.

Table 5 Adjustment of ultrasonic extraction methods for medium to high (above 20 mg kg^{-1}) and low (below 20 mg kg^{-1}) concentrations of analytes

Parameter	Medium to high concentrations	Low concentrations
Sample size required	small	large
Relative speed of method	fast	slow
Rigor of extraction	lesser	greater
Preparation	2-g sample mixed with anhydrous Na_2SO_4[a] to form a free-flowing powder	30-g sample mixed with anhydrous Na_2SO_4 to form a free-flowing powder
Procedure	extract once with a 3 mm (1/8 in) tapered microtip attached to a 12.5 mm horn; remove aliquot of extract for analysis	extract three times with a 19 mm horn and separate under vacuum filtration or centrifugation

[a] Granular, anhydrous, sodium sulfate is dried and cleaned by heating at 400 °C for 4 h in a shallow tray, or by precleaning (e.g. by Soxhlet extraction) with dichloromethane. A method blank should be analyzed to demonstrate that the sodium sulfate is clean.

in Table 5. Sample handling is described in Table 6. The original method may be consulted for completeness.

2.2.1.1 Samples Containing Low Concentrations For samples expected to contain low concentrations of organics (<20 mg kg^{-1}), the extraction procedure should be performed as quickly as possible to avoid losses of the more volatile extractables. Weigh approximately 30 g of sample to the nearest 0.1 g into a 400-mL beaker and add anhydrous sodium sulfate if necessary to yield a free-flowing matrix. Add 1.0 mL of the surrogate standard to all samples (samples, spiked samples, QC samples) and blanks. For appropriate surrogate compounds and concentrations, see section 4.2.

Add 100 mL extraction solvent to the sample in the beaker and place the bottom of the disrupter horn about 1.25 cm below the surface of the solvent, but above the solid on the bottom of the beaker. (Ensure that the ultrasonic system and horn have been properly tuned according to the manufacturer's instructions.) Extract ultrasonically for 3 min with the output control set at full power, the mode switch on pulse, and the percentage duty cycle knob at 50% (energy on half of the time and off half of the time). Do not use the microprobe tip for samples containing <20 mg kg^{-1} of the analyte(s).

Table 6 Sample handling depending on the nature of the matrix

Matrix	Handling
Sediment/soil	decant any water (sediment sample); mix sample thoroughly; discard sticks, rocks, leaves
Waste	separate phases before extraction; ultrasonic extraction is for solids only
Dry waste amenable to grinding	grind or otherwise subdivide sample to pass a 1 mm screen or a 1 mm hole to yield at least 10 g of ground material
Gummy, fibrous, or oily materials not amenable to grinding	cut, shred, or otherwise reduce size of pieces; may need to add anhydrous sodium sulfate (1 : 1) to facilitate grinding

The extract solution is decanted or withdrawn following centrifugation or filtered through Whatman No. 41 filter paper or equivalent (precleaning will be required for use with extracts of samples containing very low concentrations of residue) in a Buchner funnel under reduced pressure. The extraction will need to be repeated two or three times with additional portions of solvent (the USEPA method requires that these portions be 100 mL; however, the volume may be reduced to suit the circumstances if the residue is readily soluble in the extraction solvent). If filtration is being used, the final extraction is followed by transfer of the entire sample into the Buchner funnel with rinsing. Filtration should proceed until visible solvent has been removed from the filter-cake in the funnel. Do not attempt to completely dry the filter-cake as this action may result in the loss of the more-volatile analytes.

An attractive alternative if sufficient concentration of analyte(s) is present is to withdraw an appropriate aliquot of the total extract solution to carry through the rest of the procedure so that manipulation of the extract is simplified and losses are minimized.

2.2.1.2 Samples with High Concentrations This refers to samples containing >20 mg kg^{-1}. Accurately weigh 2 g of sample into a 20-mL vial and cap the cleaned mouth of the vial to prevent cross-contamination with other samples. Add 2 g of anhydrous Na_2SO_4 and mix well in the vial. Add 1.0 mL of surrogate spiking solution to the

Table 7 Solvents for ultrasonic extraction of medium to high concentrations of analytes

Analyte(s)	Solvent
Nonpolar compounds, e.g. organochlorine pesticides and PCBs	hexane or other appropriate solvent
Other semivolatile organics	dichloromethane

mixture. Add sufficient solvent to bring the final solvent volume up to 10 mL and sonicate the sample using the 3 mm tapered microtip probe for 2 min (output control level 5 and mode switch on pulse and 50% duty cycle; Table 7).

2.2.2 Determination of the Dry Weight of the Sample

If sample analytical results are to be expressed on a dry weight basis, a second portion of the sample should be weighed out (at the same time as the portion undergoing analysis). This 5- to 10-g subsample should be dried in a drying oven at 105 °C overnight and then cooled in a dessicator before weighing.

Caution: Care should be taken to carry out the drying of contaminated samples in an oven contained in a well functioning fume cupboard. Significant contamination of the laboratory may result otherwise, and further low residue work in that laboratory may not be possible until the laboratory has been thoroughly decontaminated.

2.2.3 Concentration of Extract

Extract solutions may have to be concentrated (e.g. using a rotary evaporator under reduced pressure) or taken to dryness (e.g. with care under a stream of dry nitrogen) to facilitate exchange of solvent. Care should be taken to avoid volatile loss of analytes during this step.

3 METHOD DEVELOPMENT AND COMPARISON WITH OTHER EXTRACTION METHODS

3.1 Benchmark Methods for Comparison

Soxhlet extraction is itself a benchmark extraction method. A method suitable to be used for comparison to establish complete extraction is reflux in which the sample in a Soxhlet thimble is placed in a still-pot fitted with a condenser, or in which the sample is itself refluxed in the still-pot. The reflux technique ensures that the extraction takes place at the boiling temperature of the extracting solvent.

Ultrasonic extraction is usually evaluated by comparison with Soxhlet extraction as the benchmark.

3.2 Comparison Issues

Two comparison issues arise:

- does the new method extract as much total residue as the benchmark method;
- does the new method have speciation potential in that it does not extract total residue but extracts identifiable residue species such as unsorbed residues and/or reversibly sorbed residues.

Exhaustive extraction is intended to remove all residue of a particular analyte from the matrix. Both Soxhlet and ultrasonic extraction are methods of this type. Gentler, less energetic methods have the potential to extract selectively. For example, solid-phase extraction from water removes the freely dissolved residues of many organics, whereas residues associated with dissolved organic matter pass through; sediments can be filtered or centrifuged from water in which they are suspended.

Soxhlet and ultrasonic extraction have been compared with a number of other extraction techniques, often with Soxhlet extraction as the reference method. Examples are supercritical fluid extraction (SFE), high-pressure solvent extraction (HPSE), accelerated solvent extraction (ASE), i.e. extraction under increased pressure and temperature, and extraction with pressurized (subcritical) water at temperatures at or above 200 °C. Tables 8 and 9 show several recent comparisons.

3.3 Extraction Variables

These include choice of solvent, temperature, and energy level.

3.3.1 Soxhlet Extraction

The extracting solvent ought to be a good solvent for the analyte(s) and a relatively poor solvent for the matrix and compounds of less interest. The more selective the extraction process, the fewer coextractives will have to be removed in a clean-up step.

The temperature used in Soxhlet extraction is related to the boiling point of the extraction solvent; however, by the time the solvent reaches the extraction thimble, it has been cooled somewhat in the condenser. Only if the Soxhlet extractor is of the type in which the thimble is suspended in the path of the vapor (i.e. there is no vapor by-pass tube, Figure 1d) will the actual temperature of the extraction thimble be maintained near that of the boiling solvent. Extracted material in the still-pot is thus refluxed

Table 8 Comparison of Soxhlet and ultrasonic extraction with other methods

Methods	Analytes	Compared with	Ref.
Soxhlet, sonication	organic micropollutants in marine particulates	ASE, SFE, methanolic saponification	7
Soxhlet	organics in environmental solids	SFE	8
Soxhlet	polycyclic aromatic hydrocarbons from soil and fly ash	SFE	9
Soxhlet	organophosphorus hydraulic fluids in soil	SFE, HPSE	10
Soxhlet, sonication	polychlorinated biphenyls in sewage sludge	Cold digestion-saponification, steam distillation, column elution	11

in the extraction solvent. The material in the extraction thimble is heated somewhat because of its position above the still-pot. Thermally labile analytes will be vulnerable to degradation or charring if the extraction solvent boiling point is too high.

3.3.2 Ultrasonic Extraction

The major determining variable in ultrasonic extraction is the energy imparted to the extracting system by the ultrasonic transducer. Probe-type transducers are susceptible to damage from erosive substrates (e.g. sandy soil) and quickly lose their integrity. Bath-type extraction systems are more reliable, but the energy intensity available within the extraction system is usually much lower than that available with the probe-type set-up. Ultrasonic extraction is thus often not suitable for routine extraction where consistency over a period of time is required.

3.4 Making Use of the Data

Extraction data can be used for several purposes. Simple total residue data are suitable for compositional analysis in foods and agricultural products, pesticide residue analysis for regulatory purposes, and forensic analysis. However, analyses for toxicological or nutritional purposes should, take into account the biological availability of the analytes in the matrix being analyzed. Residues may exist in several forms:

- as freely dissolved entities;
- as entities associated with dissolved material such as dissolved organic matter in natural waters;
- as entities sorbed loosely (reversibly) with, for example, proteins or particulate organic matter in water;
- as irreversibly sorbed entities (includes covalently bound, ionically complexed, or physically entrapped residues).

The toxicologist needs to be able to distinguish between these types of residues to be able to interpret the analytical results intelligently. Chemical kinetics connects all of these residue types and their interconversion is controlled by these kinetics under the conditions prevalent at the time of interest; the freely dissolved residues are usually of acute toxicological importance. Other residues may be of importance if they can be converted to freely dissolved species, such as in the digestive tract. Therefore, total residue information will provide only crude correlation (if any) with toxicological effects, and analytical information allowing description of residues of each type is the most valuable to toxicology. Soxhlet and ultrasonic extraction techniques, therefore, may not be nearly as valuable to toxicologists as they are to others interested in chemical analysis.

Table 9 Recoveries for seven PCBs by six extraction methods in sewage sludge using certified reference material (CRM) 392[11]

Method	Recovery (%)	Relative standard deviation (%)
Soxhlet–Dean–Stark (toluene)	97	6.2
Soxhlet (hexane–acetone 2 : 3)	93	6.8
Cold digestion/saponification (2 M KOH in MeOH; partition with hexane)	104	15
Sonication (hexane–acetone 1 : 1)	88	12
Cyclic steam distillation	80	5.2
Column elution (dichloromethane)	61	NR

4 QUALITY CONTROL AND TROUBLESHOOTING

4.1 Data Quality Objectives

Extraction techniques require quality assurance procedures to establish that the extraction is reproducible and as complete as possible and that no systematic losses are

occurring during sample manipulation. Soxhlet and ultrasonic extraction normally are checked for performance by the use of spiked samples, i.e. samples of the substrate to which a known amount of analyte has been added just prior to extraction.

Comparison of extraction methods is best carried out using standard reference materials (SRMs) or CRMs, if they exist. Sources include the National Research Council of Canada (NRCC), the Marine Analytical Standards Program, Halifax, NS and Ottawa, ON, Canada (toxaphene in carp homogenate and marine sediments), the US National Institute of Standards and Technology (NIST) in Gaithersburg, MD, USA (a variety of specific analyte/matrix combinations including toxaphene in cod liver oil), or the European Community Bureau of Reference (BCR) in Brussels, Belgium (SRMs which have been certified for PCBs). Sediment or soil SRMs are most reliably prepared from matrices contaminated with recalcitrant analytes in the field with concentrations of contaminants being determined by exhaustive extraction in the laboratory under carefully documented conditions.

SRMs are produced according to need, for example to back a regulatory system, or to serve a widespread measurement requirement, usually by request. They must of necessity be uniform such that a portion of the SRM supplied today will give the same answer under the same conditions as another supplied at another time.

When conducting interlaboratory studies relating to the measurement of environmental contaminants, an SRM and/or well characterized control material should ideally be included at the extraction step with every set of samples. The extraction technique is then used to see if the artificially introduced analyte can be recovered in a consistent manner. Surrogate standards are used to check for systematic losses.

To ensure that compounds which are present as part of the sample matrix do not interfere with the determination of the analytes being quantified, uncontaminated matrix control material is also analyzed. In some instances, the requirement for uncontaminated control material represents a great challenge for the analyst because, in many cases, there are often no easily obtainable samples of matrices which are entirely free of the contaminant(s).

4.2 Appropriate Surrogate Standards and Concentrations

A surrogate standard is a compound that is chemically similar to the analyte or analyte group but is not expected

Table 10 Surrogate standards for semivolatile and nonvolatile analyte groups (after USEPA method 3500B)

USEPA method	Analyte group[a]	Suggested surrogate standard[b]
8041	phenols by GC	2-Fluorophenol, 2,4,6-tribromophenol
8061	phthalate esters by GC	Diphenyl phthalate, diphenyl isophthalate, dibenzyl phthalate
8081	OC pesticides by GC	2,4,5,6-Tetrachloro-*m*-xylene, decachlorobiphenyl
8082	PCBs by GC	Decachlorobiphenyl
8091	nitroaromatics by GC	2-Fluorobiphenyl
8100	PAHs by GC	2-Fluorobiphenyl, 1-fluoronaphthalene
8121	chlorinated hydrocarbons by GC	α-2,6-Trichlorotoluene, 2,3,4,5,6-pentachlorotoluene, 1,4-dichloronaphthalene
8151	acid herbicides by GC	2,4-Dichlorophenylacetic acid
8270	semivolatiles by GC/MS	Phenol-d_6, 2-fluorophenol, 2,4,6-tribromophenol, nitrobenzene-d_5, 2-fluorobiphenyl, *p*-terphenyl-d_{14}
8280	PCDDs and PCDFs by HRGC/LRMS	Internal standards added at time of extraction; no surrogates
8290	PCDDs and PCDFs by HRGC/HRMS	Internal standards added at time of extraction; no surrogates
8310	PAHs by HPLC	Decafluorobiphenyl
8325	nonvolatiles by HPLC/PBMS or UV/VIS	Benzidene-d_8, caffeine-$^{15}N_2$, 3,3′-dichlorobenzidene-d_6, bis-(perfluorophenyl)-phenylphosphine oxide

[a] GC, gas chromatography; OC, organochlorine; MS, mass spectrometry; PCDD, polychlorinated dibenzo-*p*-dioxin; PCDF, polychlorinated dibenzofuran; HRGC, high-resolution gas chromatography; LRMS, low-resolution mass spectrometry; HRMS, high-resolution mass spectrometry; HPLC, high-pressure liquid chromatography; PBMS, particle beam mass spectrometry; UV/VIS, ultraviolet/visible.

[b] For analyte groups where no surrogate standard is listed, the analytical laboratory may choose one that fits the criteria given above.

to occur in the sample itself. It is thus expected to behave like the analyte group, to be extracted to the same extent, and to be carried through the procedure in the same way as the analyte group. The surrogate standard should be added to each sample, blank, laboratory control sample, and matrix spike sample just before extraction commences. The recovery of the surrogate standard is determined to identify and quantify extraction efficiency and any other gross sample-processing errors and is evaluated on the basis of its measured concentration falling within acceptance limits. Recommended surrogates are given in Table 10.[12] Extraction efficiency so determined refers to recovery of analyte which has not had the opportunity to associate with the matrix; thus, extraction of surrogate standards does not determine extractability of sorbed or bound residues of the same analyte.

4.3 Matrix Spiking to Determine Extraction Efficiency

A portion of uncontaminated analytical matrix, whether soil or agricultural product, is commonly spiked with the analyte dissolved in a solvent which can be subsequently easily removed. The introduced analyte is then allowed to associate with the matrix for a short period (hours) following which the extraction technique is applied and the residues determined. Spiking is almost always carried out with the analyte in solution, as accurate measurement of the quantity being spiked is made easier when the analyte is in solution. Determination of extraction efficiency E is can be calculated using Equation (1),

$$E = \frac{\text{amount of analyte recovered} \times 100\%}{\text{amount of analyte spiked into sample}} \quad (1)$$

4.4 Coextractives and Contaminants

Extraction of residues from matrices, particularly environmental matrices, is often accompanied by the extraction of other compounds with similar solubilities in the extraction solvent. Such compounds may be present in quantities far in excess of that of the analyte requiring their removal before instrumental analytical techniques can be used. Such clean-up techniques can be of similar difficulty to the extraction step itself.

The extraction solvent must also be free of contaminants as when the extract solution is reduced in volume to allow the next step in the analysis to take place, any contaminant present will also itself be concentrated. Such contaminants can then interfere with the analytical procedure by obscuring instrumental responses arising from the analyte(s) of interest.

5 ENVIRONMENTAL APPLICATIONS

5.1 Applicable Matrices for Extraction by these Methods

Environmental solids such as soil, sediment, finely divided wood, seeds, chopped or ground plant tissue, animal tissue ground with sodium sulfate, or oils supported on a sorbent are suitable for extraction by these methods. Recent work by Rahman et al.[11] compared the extraction capability of six extraction methods for chlorobiphenyls in sewage sludge (CRM 392 containing PCBs CB-28, CB-52, CB-101, CB-118, CB-138, CB-153, CB-180) by the following methods (Table 9):

- Soxhlet–Dean–Starck with toluene;
- Soxhlet with hexane–acetone (2 : 3);
- cold digestion/saponification with 2 mol L^{-1} KOH in methanol followed by partition with hexane;
- sonicated liquid–solid extraction with hexane–acetone (1 : 1);
- cyclic steam distillation;
- column elution with dichloromethane.

Treatment of the CRM with BF_3/methanol and partitioning with dichloromethane gave results in close agreement with the certified values for all congeners. See Table 8 for further comparisons.

5.2 Interface with Toxicology

In environmental and biological applications, Soxhlet extraction has commonly been used to measure residues of toxicants including pesticide residues, industrial chemicals, and pharmaceutical drugs. Soxhlet extraction of necessity generates results that reflect total residue, i.e. no speciation of residues occurs. Ultrasonic extraction similarly cannot easily be used to distinguish between the various species of residues. As a result, these two extraction techniques generate results that are only of very approximate environmental use. They will virtually always overstate the concentration of residue with respect to the determination of toxicological risk (or conversely, lead to understatement of the toxicological potency of the residue where measured toxicity is related to total rather than available residue).

5.3 Confounding Phenomena to be Considered

Residues being quantified are usually present in a variety of forms: unassociated or freely dissolved in the interstitial water in a sediment or soil sample, loosely associated but not adsorbed to the matrix, associated with dissolved organic matter, reversibly sorbed to particulate matter or soil organic matter, or irreversibly sorbed to particulate matter or soil organic matter. It must be determined

whether the extraction method will remove all or just a portion (particular species) of the residue.

6 REGULATORY ISSUES

6.1 Value of Total Residue Data

In the case of analysis of residues at concentrations above 20 mg kg^{-1}, total residue is very useful, because the sample is probably being analyzed for a food nutritional component (AOAC Methods 963.15 or 969.14) or for a gross contaminant such as petroleum hydrocarbons in soil.

At concentrations below 20 mg kg^{-1}, total residue data are useful in that they represent 'all of the residue' of the analyte or analyte group in the matrix. These residues may be unassociated with the matrix, loosely associated with the matrix, reversibly bound to the matrix, or irreversibly bound to the matrix. For best results, a determination of each of these chemical species is required. Total residue determination by itself is just the beginning; residues so determined may in fact have little quantitative toxicological significance.

6.2 Relationship between Solvent-extractable Residue and Toxicology

Solvent-extractable residues at low concentrations can only very approximately be used to predict toxicological responses in natural systems. The most readily extracted residues (the 'available' residues) are probably the ones most directly correlated with toxicological response; however, imitating the conditions present in environmental or physiological samples is very challenging, perhaps impossible. In general, no speciation of residues is effected with either Soxhlet or ultrasonic extraction. Techniques other than the above described extraction procedures may have to be used to generate speciation data. Some speciation data may in some cases be determined by difference if they cannot be ascertained directly. If bound residues are present, however, then more vigorous extraction may be required to generate total residue data; high-temperature subcritical water extraction, microwave-assisted extraction, or ASE may be required.

6.3 New Improved Soxhlet Extraction Methods

The Soxhlet extraction method is a long-established technique that has been improved over the years by the development of apparatus allowing more convenient use of the method. The addition of ground glass joints (Figure 1) was a major step forward and one that allowed analysis of trace and ultratrace residues to be conducted using Soxhlet extraction. Prior to the use of ground glass joints, contamination of the extract with coextractives derived from the cork or rubber stopper material would have prevented such application of Soxhlet extraction. Banks of Soxhlet extractors have been used in routine analytical laboratories facilitating the extraction of multiple samples. An automated Soxhlet extraction method has been established as a USEPA method (Method 3541). In this method, a commercially available three-stage extraction system is used to perform the same task in a shorter time. The stages are the following:

1. The extraction thimble containing the sample is immersed in the boiling solvent. The sample remains in intimate contact with the solvent and the analytes are extracted.
2. The thimble is elevated above the solvent and extracted in the 'normal' mode to free the sample of entrained extract.
3. The solvent is evaporated as would occur in a Kuderna–Danish concentrator.

A more recent development has added focused microwave heating to a Soxhlet extraction apparatus.[13] Although Soxhlet extraction is one of the oldest extraction methods in use today, it continues to evolve.

6.3.1 Advantages

In Soxhlet extraction the operation of the extractor is intuitively easy to grasp. Extraction is performed in a visible manner and is seen as being complete or close to complete by its very nature.

Ultrasonic extraction is convenient and straightforward. Ultrasonic baths provide relatively reliable sources of extraction energy.

6.3.2 Limitations

Soxhlet extraction is solvent intensive and requires laboratory facilities in which flammable and toxic solvents can be used safely. The glassware required is expensive and vulnerable to breakage. Cooling of the condenser of the Soxhlet apparatus requires a constant supply of cooling water. Residues are concentrated in the still-pot and are thus vulnerable to thermal degradation. Only total residue can be extracted by Soxhlet extraction; little or no speciation is possible.

Ultrasonic extraction must be performed under conditions which protect the hearing of laboratory personnel. The ultrasonic generator may generate heat that may affect the extraction process. Ultrasonic probes become pitted when used for the extraction of abrasive matrices such as soil, thus changing the energy generated

and affecting the extraction efficiency of the procedure. Ultrasonic baths provide less intense extraction energy. Extraction of total residue is the aim of ultrasonic extraction; little or no speciation is possible.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
ASE	Accelerated Solvent Extraction
BCR	European Community Bureau of Reference
CRM	Certified Reference Material
HPSE	High-pressure Solvent Extraction
NIST	National Institute of Standards and Technology
NRCC	National Research Council of Canada
OP	Organophosphorus Compound
SFE	Supercritical Fluid Extraction
SRM	Standard Reference Material
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Laser Mass Spectrometry in Trace Analysis

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays • Dyes, Environmental Analysis of • Explosives Analysis in the Environment • Immunoassay Techniques in Environmental Analyses • Industrial Waste Dumps, Sampling and Analysis • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Luminescence in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods • Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Instrumental Methods • Soil Sampling for the Characterization of Hazardous Waste Sites • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Waste Extraction Procedures

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Adulteration Determination • Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Lipid Analyses in Food • Liquid Chromatography in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food • Proteins, Peptides, and Amino Acids Analysis in Food • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General

Particle Size Analysis **(Volume 6)**
Ultrasonic Measurements in Particle Size Analysis

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicide Residues in Biota,

Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Oil Shale and Shale Oil Analysis • Petroleum Residues, Characterization of

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction • Pulp and Paper Matrices

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Archaeological Chemical Analysis • Quality Assurance in Analytical Chemistry

REFERENCES

1. B.D. Hill, 'Determination of *s*-Triazine Herbicides in Manitoba Soils', MSc thesis, University of Manitoba, 1974.
2. A.I. Vogel, *A Textbook of Practical Organic Chemistry*, 3rd edition, Longman, 153, 1970.
3. Kontes, *Chemistry and Life Science Products, Catalog*, 50th edition, Vineland, NJ, 172–173, 1992.
4. United States Environmental Protection Agency, *Method 3540C: Organic Extraction and Sample Preparation, Revision 2, in Methods for Evaluation of Solid Waste*, Vol. 1B, Laboratory Manual, Physical/Chemical Methods, 3rd edition, USEPA, Office of Solid Waste and Emergency Response, Washington, DC, 1986 (updated 1995).
5. 'AOAC Official Methods', 16th edition, Association of Official Analytical Chemists International, Gaithersburg, MD, 1998.
6. United States Environmental Protection Agency, *Method 3550B: Ultrasonic Extraction, Revision 2, in Methods for Evaluation of Solid Waste*, Vol. 1B, Laboratory Manual, Physical/Chemical Methods, 3rd edition, USEPA, Office of Solid Waste and Emergency Response, Washington, DC, 1986 (updated 1996).
7. O.P. Heemken, N. Theobald, W. Wenclawiak, 'Comparison of ASE and SFE with Soxhlet, Sonication, and Methanolic Saponification Extractions for the Determination of Organic Micropollutants in Marine Particulate Matter', *Anal. Chem.*, **69**, 2171–2180 (1997).
8. S.B. Hawthorne, D.J. Miller, 'Direct Comparison of Soxhlet and Low- and High-temperature Supercritical CO_2 Extraction Efficiencies of Organics from Environmental Solids', *Anal. Chem.*, **66**, 4005–4012 (1994).
9. B.W. Wenclawiak, T. Paschke, M. Krappe, 'Supercritical Fluid Extraction of Polycyclic Aromatic Hydrocarbons (PAH) from Native Soil and Fly Ash with Pure and Modified Carbon Dioxide and Dimethyl Ether', *Fresenius J. Anal. Chem.*, **357**, 1128–1132 (1997).
10. M.D. David, J.N. Seiber, 'Comparison of Extraction Techniques, including Supercritical Fluid, High-pressure Solvent, and Soxhlet, for Organophosphorus Hydraulic Fluids from Soils', *Anal. Chem.*, **68**, 3038–3044 (1996).
11. M.S. Rahman, M. Parreño, R. Bossi, A.B. Payá-Pérez, B. Larsen, 'Chlorobiphenyls in Sewage Sludge: Comparison of Extraction Methods', *Fresenius J. Anal. Chem.*, **360**, 556–563 (1998).
12. United States Environmental Protection Agency, *Method 3500C, Soxhlet Extraction, Revision 3, in Methods for Evaluation of Solid Waste*, Vol. 1B, Laboratory Manual, Physical/Chemical Methods, 3rd edition, USEPA, Office of Solid Waste and Emergency Response, Washington, DC, 1986 (updated 1996).
13. L.E. Garcìa-Ayuso, M. Sánchez, A. Fernández de Alba, M.D. Luque de Castro, 'Focused Microwave-assisted Soxhlet: An Advantageous Tool for Sample Extraction', *Anal. Chem.*, **70**, 2426–2431 (1998).

NOTE ADDED IN PROOF

For those considering the use of ultrasonic extraction, recent work published by L.K. Weavers, N. Malmstadt, M.R. Hoffmann, *Environ. Sci. Technol.*, **34**(7), 1280–1285 (2000) describes the ability of the ultrasonic extraction procedure to degrade many organic compounds by pyrolytic and free radical reaction mechanisms: 'Destruction of organic compounds occurs in the cavitation bubble itself or its interfacial sheath [through] direct pyrolysis, hydroxylation, or radical reactions that result from the gas phase pyrolysis of H_2O'.

Supercritical Fluid Extraction of Inorganics in Environmental Analysis

Mehdi Ashraf-Khorassani and Larry T. Taylor
Virginia Polytechnic Institute and State University, Blacksburg, USA

Michael T. Combs
3M Corporate Analytical Technology Center, St. Paul, USA

1 Introduction **1**
2 Solubility of Chelating Agents and Metal Chelates **1**
2.1 β-Diketones 3
2.2 Dithiocarbamates 3
2.3 Organophosphorus Reagents 5
2.4 Other Chelating Agents 6
3 Chelating Agent Studies **7**
3.1 β-Diketones 7
3.2 Carbamates 8
3.3 Crown Ethers 9
3.4 Phosphines 10
3.5 Supercritical Fluid Extraction of Organometallic Compounds 11
4 Conclusion **12**
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**

Supercritical fluid extraction (SFE) of metal ions via in situ chelation and preformed metal chelates using pure or modified CO_2 from several matrices has been accomplished. To understand the extraction results better, the solubilities of different ligands and metal chelates have been measured in supercritical fluids (SFs). In most cases, fluorinated chelates have been found to possess higher solubility in the SF than nonfluorinated ones, which in turn can often enhance extraction efficiency. Metal chelate stability under supercritical conditions plays an important role in the extraction of metals. Also, the matrix has a major effect on extraction efficiency of both hydrated metal ions and metal chelates. In order to obtain sufficient extraction not only should solubility of the metal complex be considered, but also the overall stability of the complex is critical.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

SFE has become an attractive alternative to conventional solvent extraction for the recovery of organic compounds from both environmental and biological matrices. The high diffusivity, low viscosity, and variable solvent strength are some of the attractive features of SF. Carbon dioxide is the most common fluid for SFE. The moderate critical temperature (31 °C) and critical pressure (73 atm (1 atm = 14.696 psi = 0.10132 MPa)), chemical inertness, low cost, and wide availability in pure form are a few of the advantages of supercritical CO_2. Most of the studies in the literature (>95%) regarding SFE have dealt with organic compounds.

Different research groups have investigated extraction of metal ions from environmental matrices using either pure or modified supercritical CO_2 as an extraction fluid. It is known that cationic metal ions are not soluble in pure CO_2 due to weak interactions between the positively charged metal ion and the relatively nonpolar CO_2.[1] Therefore, it could be concluded that extraction and separation of metal ions via supercritical CO_2 are not feasible. However, metal ions may be solubilized, extracted, and separated via binding them to an organic ligand, which results in the formation of a neutral stable metal complex that exhibits solubility in SFs. Rapid complexation kinetics between the metal ion and ligand and a high stability constant for the resulting neutral complex, can obviously enhance the extraction process. The efficiency of extraction is dictated by four different factors:

1. ligand solubility in the SF
2. reaction rate of ligand with metal ion in the SF
3. solubility of metal chelate in the SF
4. stability of ligand and metal chelate in the SF.

2 SOLUBILITY OF CHELATING AGENTS AND METAL CHELATES

Two techniques (dynamic and static) have been used to measure the solubility of ligands and metal chelates in SFs. In the dynamic case, a solute is continuously, but slowly swept with SF. The solubility is measured based on the weight of analyte dissolved in an exact volume of fluid. With this technique, the solubility would be expressed in moles of analyte per liter of SF at a given density. The solubility is also commonly reported as the mole fraction of analyte, wherein moles of SF can be obtained from the volume and density. In this procedure, the solubility may depend on the flow rate of fluid passing by the analyte. Ashraf-Khorassani et al.[2] have demonstrated that solubility of ethyl centralite **(1)** not only depends on

pressure and temperature, but is also a function of CO_2 flow rate through the extraction cell. Their results showed that at low flow rate, solubility was greater because the system was closer to equilibrium conditions, while at higher flow rate solubility was lower.

$(C_2H_5)NH{-}\overset{\overset{O}{\|}}{C}{-}NH(C_2H_5)$

(1)

In the static mode, the cell is loaded with a known amount of solute and SF at a known density. Usually, equilibrium between solute and SF is achieved using either recirculation or diffusion to homogenize the fluid. Analysis in this case is usually accomplished by using a spectroscopic method such as ultraviolet (UV) spectroscopy. This technique is convenient since the analysis is in situ and equilibrium can be obtained easily. However, the effect of the SF on the spectral absorption properties of the analyte and the bandshape of the solute complicate these measurements. In addition, adsorption of the solute onto the UV or IR (infrared) cell window from the saturated solution may occur, yielding incorrect results. In the static diffusion mode after system pressurization is attained, the solute is allowed to diffuse into the SF for a certain period of time in order to establish equilibrium between the solute and SF. After equilibrium is established, the cell is depressurized and the analyte dissolved in the SF is collected in an organic solvent outside the vessel. The solubility of an analyte in this case can be measured using either gravimetric

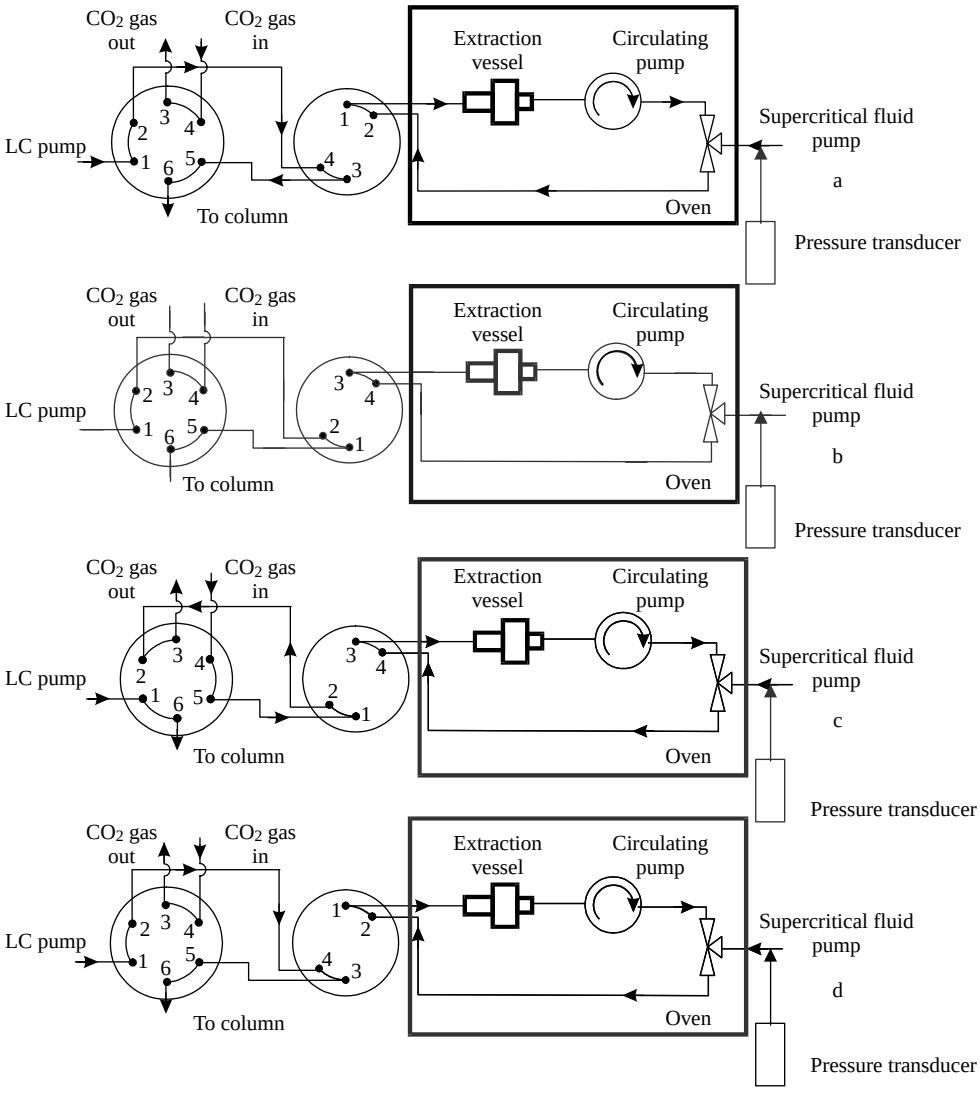

Figure 1 Solubility measurement apparatus.

or spectroscopic methods. Although often used, this technique has several pitfalls including (1) equilibration by diffusion may take more than 1 or 2 h, (2) precipitation of the analyte on the top and inside the extraction tube may occur during depressurization, and (3) the restrictor may plug due to the high concentration of analytes in the decompressing fluid.

Ashraf-Khorassani et al.[3] used on-line SFE/HPLC (high-performance liquid chromatography) to measure the solubility of different metal β-diketonate complexes in supercritical CO_2. Figure 1 shows a schematic of the system used to determine the solubilities. In each measurement, the 0.5 mL extraction vessel was filled with a metal chelate. Next, the system was pressurized to 200 atm using pure CO_2. After pressurization of the system, the three-way valve was closed and a recirculating pump was activated (Figure 1a). The function of the recirculating pump was to ensure complete saturation of SF with the metal chelate of interest. After 30 min of equilibration, the four-port two-position valve with a 1 μL sample loop was rotated such that the solubilized chelate in supercritical CO_2 was directed from the 1 μL loop to the HPLC system (Figure 1b). The HPLC flow then washed the chelate from the injection loop through the column to a variable-wavelength UV absorbance detector operated at appropriate wavelength. By employing a second valve (six-port two-position valve), a stream of carbon dioxide or air could then pass through the sample loop to remove the chromatographic mobile phase from the loop, thus avoiding modification of SF composition (Figure 1c). The four-port two-position valve could then be rotated back to the load position and the procedure repeated (Figure 1d).

This system had several advantages compared with systems used previously. First, solubility measurements were obtained in an organic solvent not under SF conditions, eliminating spectral shifts. Second, complete equilibration between solute and SF was achieved via a recirculation pump. Finally, replicate measurements can be obtained without recharging the vessel with CO_2 or analyte.

Solubility of the metal chelate in supercritical CO_2 is the key factor in SFE of metals. In this respect, different types of ligand have been used for complexation and extraction of trace metals in supercritical CO_2. The literature, however, regarding the solubility of metal chelates in supercritical CO_2 is very limited.

2.1 β-Diketones

In 1991, M'Hamdi et al.[1] measured the solubility of β-diketonate complexes of Cu^{2+}, Y^{3+}, and Ba^{2+} (i.e. the β-diketonate was either acetylacetonate anion (AcAc) **(2)** or hexafluoroacetylacetonate anion (HFA) **(3)**), in supercritical CO_2 at different pressures and temperatures. Solubility measurements were based on the weight of solute lost from the extraction cell and the mass of CO_2 that passed through the extraction system. In this study, HFA complexes showed higher solubility in supercritical CO_2 than the AcAc complexes.

(2) **(3)**

Wai and co-workers used the static mode with a 30 min equilibrium time to measure the solubility of lanthanum and europium 2,2-dimethyl-6,6,7,7,8,8,8-heptafluoro-3,5-octandione (FOD) **(4)**, ($La(FOD)_3$ and $Eu(FOD)_3$) in supercritical CO_2.[4] Analysis of the collected material was obtained via neutron activation analysis. The reported solubility of $La(FOD)_3$ was approximately 2–3 orders of magnitude higher than that obtained for metal–AcAc complexes. The high solubility of $La(FOD)_3$ in supercritical CO_2 was related to the fact that FOD forms thermally stable and highly volatile complexes ($La(FOD)_3 = 5.5 \pm 0.2 \times 10^{-2}$ M and $Eu(FOD)_3 = 7.9 \pm 0.2 \times 10^{-2}$ M).

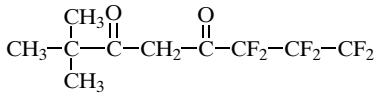

(4)

Ashraf-Khorassani et al.[3] measured solubility of different metal AcAcs and metal HFAs in pure CO_2. In this study solubility of metal HFAs was found to be 2–3 orders of magnitude higher than the solubility of metal AcAcs in keeping with the earlier study of M'Hamdi et al.[1] (Table 1).

2.2 Dithiocarbamates

The first report on solubility of metal diethyldithiocarbamate (DDC) **(5)** and bis(trifluoroethyl)dithiocarbamate (FDDC) **(6)** complexes was published by Laintz et al.[5] They measured the solubility of $Ni(DDC)_2$, $Ni(FDDC)_2$, $Co(DDC)_3$, $Co(FDDC)_3$, $Cu(DDC)_2$, $Cu(FDDC)_2$, Na(DDC), Na(FDDC), $Bi(DDC)_3$ and $Bi(FDDC)_3$) in supercritical CO_2. A high-pressure view cell connected to an ultraviolet/visible (UV/VIS) spectrometer was used to measure the solubility of the metal chelates (Figure 2). In order to assure equilibrium between the analyte and the SF, a magnetic stirrer coated with Teflon™ was placed in

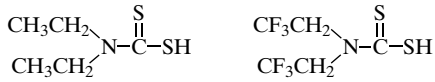

(5) **(6)**

Table 1 Solubility of different metal β-diketone complexes in supercritical CO_2

Metal	Acetylacetone	Hexafluoroacetylacetone	Ref.
Cu^{2+}	[a]6.20×10^{-6}, 119 atm,[b] 150 °C	[c]$(8.7 \pm 0.2) \times 10^{-2}$, 200 atm, 60°C	1
Cu^{2+}	[a]4.75×10^{-5}, 168 atm, 150 °C	[c]$(8.6 \pm 0.2) \times 10^{-2}$, 400 atm, 60 °C	1
Cu^{2+}	[a]7.25×10^{-5}, 218 atm, 150 °C		1
Y^{3+}	[a]1.65×10^{-5}, 119 atm, 150 °C	[a]7.25×10^{-4}, 119 atm, 150 °C	1
Y^{3+}	[a]1.80×10^{-5}, 168 atm, 150 °C	[a]8.35×10^{-4}, 168 atm, 150 °C	1
Y^{3+}	[a]3.40×10^{-5}, 218 atm, 150 °C	[a]9.30×10^{-4}, 218 atm, 150 °C	1
Ba^{2+}		[a]7.90×10^{-5}, 119 atm, 150 °C	1
Ba^{2+}		[a]2.20×10^{-4}, 168 atm, 150 °C	1
Ba^{2+}		[a]2.40×10^{-4}, 218 atm, 150 °C	1
Ni^{2+}		[c]$(8.0 \pm 0.1) \times 10^{-3}$, 200 atm, 60 °C	3
Ni^{2+}		[c]$(9.9 \pm 0.2) \times 10^{-3}$, 400 atm, 60 °C	3
Cr^{3+}		[c]$(>8.0 \times 10^{-2})$, 200 atm, 60 °C	3
Cr^{3+}		[c]$(>8.2 \times 10^{-2})$, 400 atm, 60 °C	3
Cr^{3+}	[c]$(2.0 \pm 0.1) \times 10^{-3}$, 200 atm, 60 °C		3
Cr^{3+}	[c]$(3.5 \pm 0.2) \times 10^{-3}$, 400 atm, 60 °C		3

[a] Mole fraction.
[b] 1 atm = 14.696 psi = 0.10132 MPa.
[c] Mole per liter.

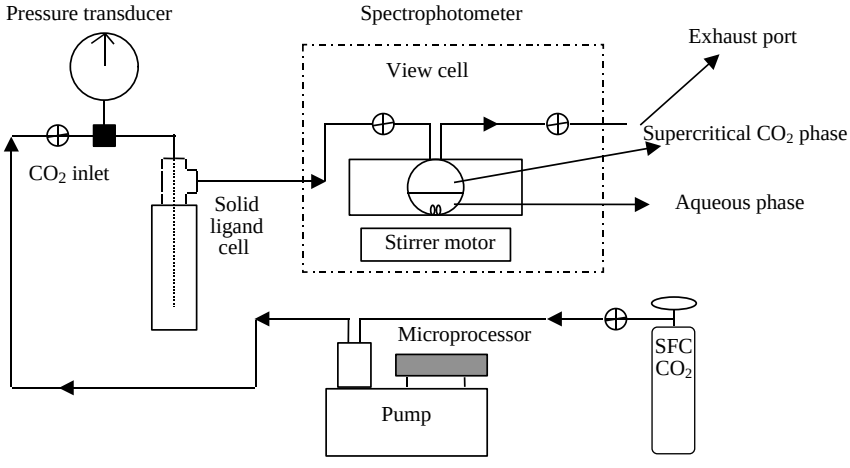

Figure 2 Schematic of solubility measurement device. (Reproduced with permission of the American Chemical Society from ref. 20.)

the view cell. After pressurization of both the view cell and transfer line, the system was equilibrated by stirring for 1 h. UV/VIS spectroscopic measurements were obtained under static supercritical conditions. The FDDC metal chelates were found to have approximately 2–3 orders of magnitude higher solubility than the comparable DDC metal chelates in supercritical CO_2 (Table 2).

Wai et al.[6] in another study used a known amount of metal complex placed in a glass tube where both ends were plugged with glass wool. The sample tube was then placed in an extraction cell and pressurized to 150 atm. The cell was heated for 30 min at 60 °C under static conditions. Following this period, the fluid and dissolved analyte were removed from the extraction cell by decompressing the fluid into a collection vial containing 5 mL of chloroform. After complete depressurization of the extraction cell, the glass tube containing the undissolved analytes was removed from the cell. Next, the empty extraction cell was flushed with excess high-density CO_2. Any metal complex that may have been flushed out of the cell during this process was also collected. The solubility of the metal complex was measured based on the amount of metal complex dissolved in CO_2 of known density and volume equal to that of the extraction cell. Using this technique they measured the solubility of $Hg(FDDC)_2$ and $Hg(DDC)_2$ in both pure and 5% methanol-modified CO_2 (Table 2). The solubility of $Hg(FDDC)_2$ increased from 5.0×10^{-3} M in pure CO_2 to 1.2×10^{-2} M in 5% methanol-modified CO_2.[6]

Table 2 Solubility of different metal–DDC chelates in supercritical CO_2

Chelate	Ligand = DDC	Ligand = FDDC	Ref.
Cu^{2+}	$(1.1 \pm 0.2) \times 10^{-6}$ M, 100 atm,[a] 50 °C	$(9.1 \pm 0.3) \times 10^{-4}$ M, 100 atm, 50 °C	5
Na^{+}	$(1.5 \pm 0.1) \times 10^{-4}$ M, 100 atm, 50 °C	$(4.7 \pm 0.1) \times 10^{-4}$ M, 100 atm, 50 °C	5
Co^{3+}	$(2.4 \pm 0.4) \times 10^{-6}$ M, 100 atm, 50 °C	$(8.0 \pm 0.6) \times 10^{-4}$ M, 100 atm, 50 °C	5
Bi^{3+}	$(1.3 \pm 0.1) \times 10^{-6}$ M, 100 atm, 50 °C	$<10^{-7}$ M, 100 atm, 50 °C	5
Hg^{2+}	$(8.2 \pm 0.6) \times 10^{-6}$ M, 100 atm, 50 °C	$(5.0 \pm 0.4) \times 10^{-3}$ M, 100 atm, 50 °C	6
Hg^{2+}	$(3.0 \pm 0.5) \times 10^{-5}$ M, 150 atm, 50 °C, 5% MeOH	$(1.2 \pm 0.4) \times 10^{-3}$ M, 150 atm, 50 °C, 5% MeOH	6
Ni^{2+}	$(8.5 \pm 1.0) \times 10^{-7}$ M, 100 atm, 50 °C	$(7.2 \pm 1.0) \times 10^{-4}$ M, 100 atm, 50 °C	6
Zn^{2+}	1.1×10^{-6} M, 100 atm, 60 °C	9.5×10^{-4} M, 100 atm, 60 °C	9
Hg^{2+}	6.8×10^{-6} M, 100 atm, 60 °C	3.0×10^{-3} M, 100 atm, 60 °C	9
Hg^{2+}	2.4×10^{-3} M, 200 atm, 60 °C	1.40×10^{-2} M, 200 atm, 60 °C	7

[a] 1 atm = 14.696 psi = 0.01032 MPa.

Ashraf-Khorassani and Taylor[7] used an on-line SFE/HPLC system to measure the solubility of mercury bis(diethyl)dithiocarbamate ($Hg(DDC)_2$) and bis(bis(trifluoroethyl)dithiocarbamate) ($Hg(FDDC)_2$) complexes. Fluorination of the ligand increased the solubility of $Hg(FDDC)_2$ at least an order of magnitude compared with $Hg(DDC)_2$. Results obtained in this study indicated greater solubility than that reported by Wai et al.[6] These differences were attributed to losses of analyte during depressurization of the CO_2.

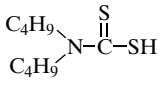

(7)

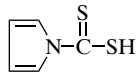

(8)

Wang et al.[8] used the static mode to measure the solubility of zinc dibutyldithiocarbamate (Zn-DBDC) **(7)**, $Zn(DDC)_2$, and Zn pyrrolidine dithiocarbamate (Zn-PDTC) **(8)** complexes (Table 3). In this study, 0.5–0.8 g of test complexing agent contained in a cardboard thimble was placed inside the extraction cell. After 30 min under static conditions, the cell was decompressed and the analyte was trapped in methanol, which was subsequently analyzed spectrophotometrically at 298 nm. The solubility of the zinc chelates in most cases increased with increasing hydrocarbon chain length.

Wai et al.[9] performed a similar study to measure the solubilities of Cu^{2+}, Hg^{2+} and Zn^{2+} complexes with seven different dithiocarbamate ligands (PDTC, DBDC, DDC, P3DC (dipropyldithiocarbamate) **(9)**, P5DC (dipentyldithiocarbamate) **(10)**, DHC (dihexyldithiocarbamate) **(11)**, and FDDC), in supercritical CO_2 at 60 °C and two different pressures (100 and 230 atm). They found that dithiocarbamate ligands with a small Hildebrand solubility parameter value yielded metal complexes that exhibited higher solubilities in supercritical CO_2 (Table 3). For example,

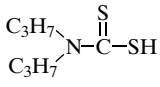

(9)

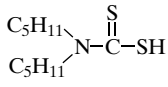

(10)

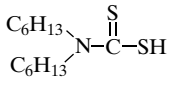

(11)

PDTC with a solubility parameter of 11.43 $cal^{1/2}\,cm^{-3/2}$ (1 cal = 4.1845 J) formed a copper complex ($Cu(PDTC)_2$) which had a solubility of 4.1×10^{-7} M in supercritical CO_2 at 60 °C and 100 atm while FDDC with a solubility parameter of 8.75 $cal^{1/2}\,cm^{-3/2}$ formed a copper complex ($Cu(FDDC)_2$) which had a solubility of 9.1×10^{-4} M under identical conditions.

2.3 Organophosphorus Reagents

Organophosphorus reagents are well known ligands for extraction of actinide series elements. Supercritical CO_2 has shown promise for extraction of uranium from both solid and liquid matrices by complexation with organophosphorus ligands via in situ chelation. The first report on solubility of tributylphosphate (TBP) **(12)** was by Page et al.[10] They determined that 11% of TBP is miscible with CO_2 at 60 °C and 120 atm. Later, Schmitt and Reid.[11] and Lin et al.[12] reported solubility of other phosphine oxides (i.e. TBP, tributylphosphine oxide (TBPO) **(13)**, tri-*n*-octylphosphine oxide (TOPO) **(14)**, and triphenylphosphine oxide (TPPO) **(15)**) in

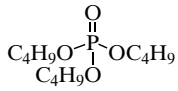

(12)

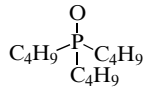

(13)

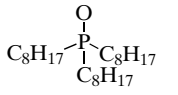

(14)

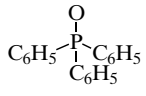

(15)

Table 3 Solubility of dithiocarbamate metal complexes in supercritical CO_2

Metal chelates	Solubility			Solubility			Ref.
	Concentration (M)	Pressure (atm)[a]	Temperature (°C)	Concentration (M)	Pressure (atm)[a]	Temperature (°C)	
$Zn[SC(S)N(C_4H_9)_2]_2$				5.3×10^{-4}	237	55	8
$Zn[SC(S)N(C_2H_5)_2]_2$				3.3×10^{-5}	237	55	8
$Zn[SC(S)NC_4H_8]_2$				5.0×10^{-5}	237	55	8
Zn(PDTC)	3.2×10^{-7}	100	60	9.0×10^{-6}	230	60	9
Zn(P3DC)	7.9×10^{-6}	100	60	1.5×10^{-4}	230	60	9
Zn(BDTC)	8.2×10^{-6}	100	60	6.9×10^{-4}	230	60	9
Zn(P5DC)	1.6×10^{-4}	100	60	3.2×10^{-3}	230	60	9
Zn(HDC)	3.2×10^{-4}	100	60	5.8×10^{-3}	230	60	9
Hg(PDTC)	3.5×10^{-7}	100	60	3.4×10^{-6}	230	60	9
Hg(P3DC)	1.2×10^{-5}	100	60	2.3×10^{-4}	230	60	9
Hg(BDTC)	5.6×10^{-5}	100	60	5.6×10^{-4}	230	60	9
Hg(P5DC)	1.0×10^{-4}	100	60	2.0×10^{-3}	230	60	9
Hg(HDC)	1.6×10^{-4}	100	60	3.8×10^{-3}	230	60	9
Cu(PDTC)	4.1×10^{-7}	100	60	4.0×10^{-6}	230	60	9
Cu(P3DC)	6.3×10^{-6}	100	60	1.2×10^{-4}	230	60	9
Cu(BDTC)	1.3×10^{-5}	100	60	7.2×10^{-4}	230	60	9
Cu(P5DC)	9.0×10^{-5}	100	60	1.8×10^{-3}	230	60	9
Cu(HDC)	2.1×10^{-4}	100	60	2.8×10^{-3}	230	60	9

[a] 1 atm = 14.696 psi = 0.01032 MPa.

supercritical CO_2 at different temperatures and pressures. It was determined that TBPO has high solubility in supercritical CO_2 at 200 atm and 60 °C, while the solubility of TPPO using the same conditions was at least two orders of magnitude lower. Solubility of TOPO was an order of magnitude lower than TBPO and an order of magnitude higher than TPPO.

Smart et al.[13] measured the solubility of bis(2,4,4-trimethylpentyl)monodithiophosphine (Cyanex 302) in supercritical CO_2 at different pressures and 60 °C using a high-pressure view cell. The solubility of Cyanex 302 increased from 8 g L^{-1} to 108 g L^{-1} by varying the pressure from 110–280 atm (1 atm = 0.10132 MPa). Results of these studies demonstrated that phosphorus-containing reagents not only are soluble in supercritical CO_2, but also are stable in supercritical CO_2.

Meguro et al.[14] used a twin equilibrium cell of stainless steel and a glass collection vessel containing 1.5 g of glass wool to measure solubility of TBP, diisodecylphosphoric acid (DIDPA), di-(2-ethylhexyl)phosphoric acid (DEHPA), dihexyl(*N*,*N*-diethylcarbamoyl)methyl phosphate (CMP), and octyl(phenyl)(*N*,*N*-diisobutylcarbamoyl)methylphosphine oxide (CMPO) in supercritical CO_2 over a range of temperatures (30–90 °C) and pressures (7.5–25 MPa). Results showed that solubility of the compounds increased with increasing density of CO_2, showing a linear relationship. Results showed that at 60 °C and 25 MPa TBP, DEHPA, and DIDPA have a solubility of 1.4, 0.81 and 0.090 mol L^{-1}; while, solubility of CMP at 60 °C and 15 MPa was 0.41 mol L^{-1}, and solubility of CMPO at 60 °C and 20 MPa was 0.089 mol L^{-1}.

2.4 Other Chelating Agents

Wang et al.[15] used the static technique to measure the solubility of three different bistriazolo-crown ether ligands **(16, 17)** and **(18)** in both pure and 5% methanol-modified CO_2. Solubilities of these crown ethers increased by an order of magnitude in 5% methanol-modified CO_2 compared with 100% CO_2 (Table 4). These ligands were used for direct in situ chelation of Hg^{2+} and Au^{3+} from both solid and liquid matrices. The solubility of these metal chelates was not measured.

Cowey et al.[16] used dynamic conditions to measure the solubility of 5,7,12,14-tetramethyl-2,3,9,10-dibenzo[*b*,*i*][1,4,8,11]tetraazacyclotetradecine-nickel(II) in supercritical CO_2 and 10% methanol-modified CO_2 at different pressures (160, 250, and 340 atm) and temperatures (40, 50, 60, and 70 °C). The solubility of each analyte was measured by comparing the amount of extracted analyte that was collected with a certain volume of pure or methanol-modified supercritical CO_2. The Ni complex solubility, in 10% methanol-modified CO_2, varied from 7.2×10^{-6} to 5.77×10^{-5} g mL^{-1} depending on the pressure or temperature of the operation (Table 4). They

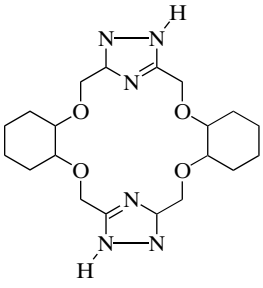

Crown I
(16)

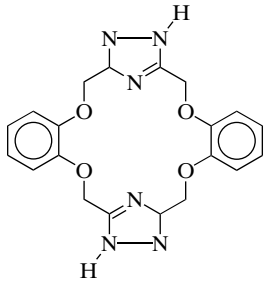

Crown II
(17)

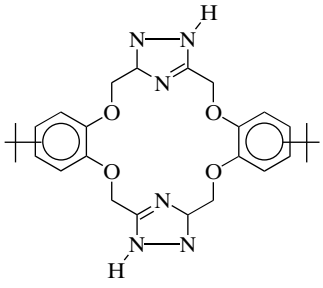

Crown III
(18)

also measured the solubility of ferrocene in pure and 10% methanol-modified CO_2. The ferrocene solubility varied from 4.77×10^{-3} to 1.1×10^{-2} g mL^{-1} depending on the pressure and temperature of the operation (Table 4).

3 CHELATING AGENT STUDIES

The ligand used for SFE of metal ions is of primary importance. The solubility, rate of formation, and stability of metal chelates dictate the success of extraction of metal ions from their environment via SFE. Inorganic SFE of metal ions and chelates will be presented with respect to the ligand used for extraction.

3.1 β-Diketones

Several research groups have demonstrated extraction of metal ions from different matrices using various β-diketones. Extraction of lanthanides and actinide series ions has received considerable attention from several research groups. SFE of lanthanides and actinides was first reported by Lin et al.[3] In their report, the extraction of La^{3+}, Eu^{3+}, Lu^{3+}, and uranyl (UO_2^{2+}) ions was performed from both wet and dry cellulose-based filter paper using pure and methanol-modified CO_2. In each extraction, filter paper was spiked with 10 μg of lanthanide or actinide ion and 20 μL (80 mmol) of FOD was placed in the extraction cell. The chelation and solubilization processes were allowed to occur under static SFE conditions for 10 min. A wet sample was prepared by spiking 10 μL of water onto the filter paper. Following extraction, decompression, and collection of the analyte in chloroform, the lanthanides were back-extracted with 50% HNO_3 solution followed by neutron activation analysis of the acid solution. An extraction efficiency of less than 20% was obtained for La^{3+}, Eu^{3+} and Lu^{3+} using

Table 4 Solubility of miscellaneous ligands and metal chelates in supercritical CO_2

Metal complex and ligand	Solubility		Ref.
Crown I	1.3×10^{-5} M, 200 atm, 60 °C	2.1×10^{-4} M, 200 atm, 60 °C, 5% MeOH	15
Crown II	1.0×10^{-5} M, 200 atm, 60 °C	2.0×10^{-4} M, 200 atm, 60 °C, 5% MeOH	15
Crown III	4.3×10^{-4} M, 200 atm, 60 °C	1.3×10^{-3} M, 200 atm, 60 °C, 5% MeOH	15
Ferrocene	2.6×10^{-3} M, 133 atm, 40 °C		16
Ferrocene	4.9×10^{-3} M, 242 atm, 40 °C		16
Ferrocene	6.5×10^{-3} M, 333 atm, 40 °C		16
5,7,12,14-tetramethyl-2,3 : 9,10-dibenzo[*b*,*i*][1,4,8,11]tetraazacyclotetradecinenickel(II)	1.05×10^{-5} g mL^{-1}, 160 atm, 40 °C, 10% MeOH 1.00×10^{-5} g mL^{-1}, 160 atm, 50 °C, 10% MeOH 8.02×10^{-6} g mL^{-1}, 160 atm, 60 °C, 10% MeOH 7.21×10^{-6} g mL^{-1}, 160 atm, 70 °C, 10% MeOH		16
5,7,12,14-Tetramethyl-2,3 : 9,10-dibenzo[*b*,*i*][1,4,8,11]tetraazacyclotetradecinenickel(II)	1.70×10^{-5} g mL^{-1}, 250 atm, 40 °C, 10% MeOH 2.01×10^{-5} g mL^{-1}, 250 atm, 50 °C, 10% MeOH 2.44×10^{-5} g mL^{-1}, 250 atm, 60 °C, 10% MeOH 3.44×10^{-5} g mL^{-1}, 250 atm, 70 °C, 10% MeOH		16
5,7,12,14-Tetramethyl-2,3 : 9,10-dibenzo[*b*,*i*][1,4,8,11]tetraazacyclotetradecinenickel(II)	2.20×10^{-5} g mL^{-1}, 340 atm, 40 °C, 10% MeOH 2.45×10^{-5} g mL^{-1}, 340 atm, 50 °C, 10% MeOH 2.44×10^{-5} g mL^{-1}, 340 atm, 60 °C, 10% MeOH 5.77×10^{-5} g mL^{-1}, 340 atm, 70 °C, 10% MeOH		16

[a] 1 atm = 14.696 psi = 0.10132 MPa.

pure CO_2 from a wet matrix. However, the recoveries increased to above 65% and 90% from dry and wet matrices, respectively, using 5% methanol-modified CO_2. In a similar fashion the extraction efficiency of uranyl ion spiked onto filter paper was studied. The extraction of $[(UO_2)^{2+}]$ using 5% methanol-modified CO_2 containing FOD at 60 °C and 150 atm from either dry or wet matrix yielded recoveries above 95%.

Lin and co-workers[17,18] have used other fluorinated β-diketones (e.g. HFA, thenoyltrifluoroacetone (TTA) **(19)**, and trifluoroacetylacetone (TAA) **(20)**) for extraction of lanthanides and actinides using both pure and 5% methanol-modified CO_2. The extraction efficiency of La^{3+}, Eu^{3+}, and Lu^{3+} from cellulose-based filter paper using pure supercritical CO_2 saturated with different fluorinated β-diketones was less than 25%. However, the extraction efficiency of the same lanthanides using 5% methanol-modified CO_2 saturated with the same fluorinated ligands varied from 11–95% depending on the ligand.

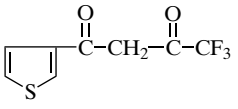

(19)

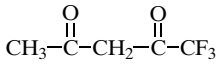

(20)

The effect of the matrix on lanthanide extraction was also investigated. A topsoil sample was spiked with 10 μg each of La^{3+}, Eu^{3+}, and Lu^{3+}.[17] The sample was extracted with neat CO_2 containing a mixture of TBP and one of the fluorinated β-diketones (i.e. HFA, TTA, or FOD). Results showed that the best extraction efficiency of lanthanides (80, 88, and 92%) was obtained using a mixture of TBP and HFA. Furton et al.[19] determined the extraction efficiency of uranium spiked onto solid matrices (e.g. polyester, glasswool, cotton, and Kaolin) via synergistic in situ chelation using pure and modified CO_2. They concluded that SFE yielded higher recovery than conventional liquid extraction. In their study, high extraction recovery of uranium was obtained via 5% ethanol with 0.10 M FOD- and 0.10 M TBP-modified supercritical CO_2.

Ashraf-Khorassani et al.[4] demonstrated extraction of different metal AcAc and metal HFA from an aqueous environment. Both metal AcAc and HFA complexes showed low extraction efficiency from water, although both complexes exhibited high solubility in pure CO_2. These results demonstrated that the extraction of metal chelates is not only solubility dependent, but also depends on the stability of chelates in the specific environment under supercritical conditions.

3.2 Carbamates

By far the greatest amount of research has focused on the use of carbamate ligands for the SFE of metal ions from both solid and liquid matrices. Carbamates, especially fluorinated carbamates, have both the good solubility and the stability needed for successful SFE. Carbamates have also been used for in situ derivatization of free metal ions in a variety of matrices.

Laintz et al.[20] have extracted Cu^{2+} spiked onto Celite. Extraction efficiency of Cu^{2+} using FDDC, however, did not exceed 80% at a CO_2 density of 0.70 g mL^{-1} and 35 °C. Later, Wai et al.[6] reported extraction of mercuric ion (Hg^{2+}) from cellulose-based filter paper using Li(FDDC) as a chelating agent. In this experiment 10 μg of Hg^{2+} was spiked onto filter paper. Then 10 mg of Li(FDDC) was added directly to the extraction vessel, thus creating a large excess of ligand relative to the mercuric ion. The extraction efficiency of Hg^{2+} was less than 12% for a dry sample. The low extraction efficiency of mercuric ion was thought to be due to a strong interaction between Hg^{2+} and the cellulose matrix. However, the addition of a small amount of water (10 μL) onto the matrix increased the extraction efficiency of mercuric ion to 84%. The extraction efficiency of Hg^{2+} was further increased from 84–95% with the use of 5% methanol-modified CO_2. This increase in extraction efficiency was stated to be mostly due to enhanced interaction between Hg^{2+} and the solvent modifier.

Extraction of other heavy metal ions has been achieved using Li(FDDC) as a chelating agent. Liu et al.[21] reported extraction of Co^{3+}, Cd^{2+}, Zn^{2+}, and Cu^{2+} from solid matrices (sand, soil, and filter paper) using supercritical CO_2 saturated with Li(FDDC). Metal ion recovery ranged from 70–100%. Extraction efficiencies were determined using gas chromatography (GC)/atomic emission detection. They reported that metal impurities in the stainless steel extraction vessel could react with the ligand and positively affect the percentage recovery of analyte. They suggested using a non-metal-containing vessel such as polyether ether ketone (PEEK) which cannot affect extraction efficiency of metal ions.

Wai[22] have reported extraction of both divalent and trivalent metal ions from spiked cellulose-based filter paper and sand. Extraction efficiency of greater than 90% was reported for Cd^{2+}, Cu^{2+}, Pb^{2+}, Pd^{2+}, Zn^{2+}, As^{3+}, Au^{3+}, Ga^{3+}, and Sb^{3+} from sand and filter paper using LiFDDC ligand. Wang and Marshall[23] have separated and characterized Cd^{2+}, Zn^{2+}, and Cu^{2+} bound to metallothionein (MT), isolated from rabbit liver samples, using supercritical CO_2 saturated with tetrabutylammonium dibutyldithiocarbamate (TDBDC) ligand. Quantitative recovery of Cd^{2+} was obtained if MT was dissolved in water.

Laintz et al.[20] have been able to extract Cu^{2+} ion quantitatively from an aqueous sample using supercritical CO_2 saturated with Li(FDDC). In their experiment, supercritical CO_2 was passed through a 10 mL stainless-steel vessel filled with solid Li(FDDC), which functioned as a ligand extraction vessel. Next, the supercritical CO_2 saturated with Li(FDDC) was introduced into a high-pressure view cell filled with 5 mL of aqueous Cu^{2+} solution. The reaction of Cu^{2+} ion with Li(FDDC) was monitored at 416 nm (e.g. the charge-transfer absorbance band of $Cu(FDDC)_2$) via UV/VIS spectroscopy as a function of time. The rate of extraction of Cu^{2+} at different densities (constant temperature, 35 °C) and different temperatures (constant density, 0.22 g mL^{-1}) was studied. This was the first report of direct extraction of a metal ion from an aqueous solution using in situ chelation SFE. The extraction efficiency of Cu^{2+} ion was greater than 95% at densities above 0.35 g mL^{-1} and 35 °C. Extraction of Hg^{2+} ion from water was achieved with 5% methanol-modified CO_2 and tetrabutyl-substituted dibenzobistriazolo-crown ether (Crown III) **(18)** at 200 atm and 60 °C.[9] The extraction efficiency of Hg^{2+} was dependent on the amount of ligand placed in the ligand extraction vessel. It was determined that at least 10 mg of ligand is necessary for the extraction of 5 ppm Hg^{2+} from 4.5 mL of contaminated water.

Wai[22] has reported extraction of both divalent and trivalent metal ions with CO_2 from an aqueous matrix using Li(FDDC) as a ligand. An extraction efficiency of greater than 90% was reported for Cd^{2+}, Cu^{2+}, Pb^{2+} and Zn^{2+} from aqueous solution. In their study, after SFE the metal chelates were back-extracted from the collection solvent (chloroform) using nitric acid and analyzed via inductively coupled plasma mass spectrometry (ICP/MS).

Recently, Wang and Marshall[8] used TDBDC ligand for in situ chelation SFE of Zn^{2+}, Cd^{2+}, and Pb^{2+} from aqueous media. Again supercritical CO_2 was passed through a vessel which was filled with the dithiocarbamate. The saturated supercritical CO_2 then passed through the aqueous phase contaminated with heavy metals. Following extraction and collection of the metal chelates, flow injection analysis coupled with atomic absorption spectrometry (FIA/AAS) was used to analyze and quantify the metals. Recoveries of ions from aqueous solution were ≈73% and ≈94% within 5–15 min, respectively. Different dithiocarbamates such as TDBDC, DDC, and PDTC were compared for extraction of Zn^{2+} ion. The extraction efficiency of Zn-DBDC was much greater than Zinc-diethyldithiocarbamate (Zn-DDC) or Zn-PDTC. The high extraction efficiency of Zn-DDC was related to the high solubility of the chelate, since the solubility ratios of Zn-DBDC/Zn-DDC and Zn-DBDC/Zn-PDTC were 20 and 139, respectively.

Ashraf-Khorassani et al.[4] demonstrated extraction of different metal DDC and metal FDDC from an aqueous environment. Metal FDDC complexes exhibited much higher extraction efficiency than metal DDC via supercritical CO_2 using identical extraction conditions. They demonstrated that the extraction of metal DDC complexes is solubility dependent, while extraction of metal ions via in situ chelation is not only solubility dependent, but also depends on the stability of chelate under supercritical conditions. Toews et al.[24] studied the influence of pH on the SFE extraction efficiency of metals and ionizable organic species in water. They concluded that metal extraction by in situ chelation is pH dependent. The pH of water ranged between 2.80 and 2.95 at pressures of 70–200 atm CO_2 and temperatures between 25–70 °C. Therefore, any ligand that forms a stable neutral complex with metal ions at pH 3 should be useful in SFE.

Ashraf-Khorassani and Taylor[7] have also investigated the extraction of mercury ions from a variety of matrices (soil, fly ash, sludge, filter paper, and sand) using DDC and FDDC as chelating agents. Using inert matrices, such as sand, both $Hg(DDC)_2$ and $Hg(FDDC)_2$ produced high recoveries. However, with active matrices (e.g. fly ash and sludge), $Hg(FDDC)_2$ was much more efficient, producing recoveries greater than 95%, while $Hg(DDC)_2$ was less than 60% efficient. In situ chelation of Hg^{2+} using Li(FDDC) as a chelating agent produced greater than 80% recovery from sand, filter paper, soil, and fly ash, but less than 60% from sludge.

3.3 Crown Ethers

Research has also be done using crown ethers as chelating agents for metal-ion extraction. Bistriazolo-crown ethers **(16, 17, 18)** have been used for extraction of Hg^{2+}.[15] Five percent methanol-modified CO_2 saturated with tetrabutyl-substituted dibenzobistriazolo-crown ether **(18)** was found to extract only 78% of Hg^{2+} from sand, while the same fluid in the presence of 10 μL of water improved extraction efficiency to more than 95%. The high extraction efficiency of Hg^{2+} in the presence of water was related to a higher modified fluid–solute interaction. Tetrabutyl-substituted dibenzobistriazolo-crown ether was also used for extraction of similar heavy metal ions (Cd^{2+}, Co^{3+}, Pb^{2+}, Ni^{2+}, Mn^{2+}, Au^{3+}, and Zn^{2+}) from sand and filter paper.[15] Extraction efficiencies of most metal ions other than Au^{3+} from wet matrices, were less than 4% even when using 5% methanol-modified CO_2. The efficiency of extracted Au^{3+} from wet filter paper using 5% methanol-modified CO_2 was 79%.

Another chelating agent such as fluorinated crown ether carboxylic acid (*sym*-difluorobenzene-crown-5-oxyacetic acid) was also used for SFE of lanthanides

and uranyl ion.[3] SFE of lanthanides and uranyl ion using 5% methanol-modified CO_2 saturated with *sym*-difluorobenzo-16-crown-5-oxyacetic acid in the presence of water at 60 °C and 150 atm was less than 7%.

3.4 Phosphines

Different phosphines, including tributylphosphine, have found some use as chelating agents. Phosphines have been used both as primary ligands and also for their synergistic effect when used in conjunction with other ligands.

SFE of lanthanides has been investigated using phosphine ligands. Lin and Wai[17] studied extraction of La^{3+}, Eu^{3+}, and Lu^{3+} from buffered acetate solution at pH 4.0. Maximum extraction efficiencies of lanthanides (ranging from 75 to 89%) were obtained with supercritical CO_2 containing TTA and TBP at 150 atm and 60 °C. Extraction efficiencies of ions with supercritical CO_2 containing only TTA or TBP were poor. Extraction of uranyl and thorium ions from water was also performed.[18] In this study, the addition of 80 μmol TTA to pure supercritical CO_2 increased extraction efficiencies of uranyl and thorium ions from water to 38% and 70%, respectively. The synergistic effect on extraction of each ion by using TBP and TTA mixed with pure supercritical CO_2 was investigated. Extraction efficiencies of ions increased to 70% and 87%, respectively. Addition of 5% methanol as a modifier and TTA as a ligand to the extraction fluid increased extraction efficiencies of both ions to greater than 85%. Similar results were reported in the same paper[18] for the extraction of Th(IV) and U(VI) spiked onto cellulose-based filter paper. The percentage recovery was less than 70% using pure CO_2, but greater than 95% when 5% methanol-modified CO_2 was used. It was concluded that the extraction efficiency of both lanthanides and actinides depends on the structure of the ligand, following the order: TTA ≈ FOD > HFA > TAA > acetylacetone (AA). They also evaluated the synergistic effect of TBP **(12)** for the extraction of lanthanides and actinides using both pure and modified supercritical CO_2 from solid matrices. Significant increases in the extraction efficiency of lanthanides and actinides from sand and filter paper (>90%) were observed when an equimolar amount of TBP and ligand (HFA, TTA, or FOD) were added to the supercritical CO_2. However, the extraction efficiency of the same lanthanides and actinides from the same matrices using supercritical CO_2 saturated with only ligand or TBP was not greater than 65%.

A similar study was reported by Laintz and Tachikawn[25] investigating the synergistic effect of 10, 20, and 30% TBP-modified CO_2 saturated with TTA for the extraction of Sm^{3+}, Eu^{3+}, Gd^{3+}, Dy^{3+}, La^{3+}, Ce^{3+}, Yb^{3+}, and Lu^{3+} from an aqueous environment. Quantitative extraction of Sm^{3+}, Eu^{3+}, Gd^{3+}, and Dy^{3+} (>90%) was achieved using 30% TBP-modified CO_2 saturated with TTA while recovery of La^{3+}, Ce^{3+}, Yb^{3+}, and Lu^{3+} using the same extraction conditions was less than 70%. Extraction of the same lanthanide ions from acidic solution (6 M HNO_3 – 3 M $LiNO_3$) using 30% TBP-modified CO_2 yielded extraction efficiencies of greater than 85% for Sm^{3+}, Eu^{3+}, Gd^{3+}, and Dy^{3+}, while less than 70% recovery of La^{3+}, Ce^{3+}, Yb^{3+}, and Lu^{3+} was obtained. The log of the SFE distribution coefficient was plotted as a function of percent TBP in the extraction fluid. The slope of the line for Sm^{3+}, Gd^{3+}, Eu^{3+}, and Dy^{3+} suggested that not only was TBP forming a 3 : 1 adduct with $Ln(NO_3)_3$ during extraction but it was also forming a 3 : 1 adduct with $Ln(TTA)_3$ during the extraction.

Later, Lin et al.[12] studied the extraction of uranium (UO_2^{2+}) and thorium (Th^{4+}) from nitric acid solutions using different organophosphorus reagents (i.e. TBP **(12)**, TBPO **(13)**, TOPO **(14)**, and TPPO **(15)**). In all extractions, supercritical CO_2 was saturated by passing through a cell filled with an organophosphorus compound. Then the saturated CO_2 with organophosphorus compound was bubbled through the solution in the extraction vessel. Results showed that extraction efficiencies of both UO_2^{2+} and Th^{4+} using TBP are directly dependent on the concentration of nitric acid solution. Extraction efficiencies of both ions decreased from 91% for UO_2^{2+} and 80% for Th^{4+} in 6 M nitric acid solution to 12% for UO_2^{2+} and 20% for Th^{4+} in 0.1 M nitric acid solution. Addition of $LiNO_3$ to the 6 M nitric acid solution increased extraction efficiencies of both ions to 98% and 93%, respectively. Extraction efficiencies of both UO_2^{2+} and Th^{4+} using supercritical CO_2 modified with TBP were greater than 90% regardless of the concentration of nitric acid solution (6 M, 1 M, or 0.1 M).

Meguro et al.[26] developed the SFE method for the efficient separation of U(VI) from nitric acid solution by means of the dynamic extraction procedure using a supercritical CO_2–TBP mixture and demonstrated a potential of SFE in the application to the reprocessing of spent nuclear fuels. More recently[27] they established an equation whereby they were able to measure the distribution ratio of U(VI) from nitric acid solution into supercritical CO_2 containing TBP. The validity of the equation was examined by the measurement of the distribution ratio of the U(VI) as well as nitric acid under a wide range of pressures and temperatures. A simple linear relationship between distribution ratio, D_u, of U(VI) and density, ρ, of CO_2 was derived: $\log D_u = \alpha \log \rho + A + B$, in which α is a proportional constant implying the solvation characteristic of the solute in supercritical CO_2, A is a pressure-independent constant, and B is a variable determined by the distribution equilibrium of nitric acid. This equation would allow the user to enhance

the extraction selectivity of different metal ions. The $\log D_u$ vs $\log \rho$ plots of metals of different kinds exhibited different characteristics.

3.5 Supercritical Fluid Extraction of Organometallic Compounds

In addition to these main classes of ligands, several other studies have been performed on other inorganic species. Different research groups have focused on extraction of organometallic compounds.

Wai et al.[6] reported the extraction efficiency of spiked methylmercury and dimethylmercury onto filter paper. An extraction efficiency of greater than 99% was reported for both compounds using pure CO_2 when filter paper was spiked with 10 μL of water. Ashraf-Khorassani and Taylor[7] reported the extraction efficiencies of the same compounds from different matrices using both pure and methanol-modified CO_2. Their results showed that dimethylmercury can be extracted with an extraction efficiency of greater than 90% from all matrices. However, results for the extraction efficiency of methyl mercury chloride showed that extraction was not only dependent on extraction conditions, but also dependent on the extraction environment and matrix.

Triorganotin compounds have been used for several years in agricultural science, (1) as a biocide to combat fungal growth, (2) in wood preservative formulations to reduce fungal rot, and (3) in marine plants as an antifoulant against barnacles.[28] Diorganotins have routinely been used as heat and light stabilizers in the polymer and food-packing industries.[29] Oudsema and Poole[30,31] studied coupled on-line SFE and supercritical fluid chromatography (SFC) of organotin compounds using 0.3% formic acid-modified supercritical CO_2. Separation of 12 organotin compounds was obtained using a Deltabond methyl column using 0.3% formic acid-modified CO_2. In another experiment, on-line SFE/SFC was used to extract and separate different organotin compounds from marine paint and spiked into a potato or almond matrix. This study was focused mainly on SFC rather than SFE.

Lin et al.[32] studied extraction of six tetraalkyltin and seven additional ionic organotin compounds from spiked topsoil using both pure and 5% methanol-modified CO_2. All tetraalkyltin compounds were extracted from topsoil with recoveries of 90–110% using either fluid at 100 atm and 40 °C. However, recoveries of ionic organotin compounds ranged from 70–90% when Na(DDC) was added to the sample matrix in order to neutralize the change. The addition of 5% methanol as a modifier did not increase extraction efficiencies of ionic organotin compounds. Later, the same group extended their study to extract different ionic organotin compounds from clay and sediments. All soil and sediment samples were treated with diethylammonium DDC prior to extraction. Extractions were performed using both pure and 5% methanol-modified CO_2. Results showed that most dialkyltin, trialkyltin, and tetraalkyltin compounds can be quantitatively extracted from most of the matrices, while the percentage recovery of monoalkyltin compounds from all matrices was less than 20%.[33]

Alzaga and Bayona[34] and Cai and Bayona[35] studied SFE of butyl-, phenyl- and cyclohexyltin compounds from aqueous matrices using C_{18} solid-phase extraction (SPE) discs. The analytes were first concentrated via SPE discs. Next, the organotin compounds that were trapped on the SPE disc were derivatized via Grignard ethylation followed by SFE at 100 atm. Extraction efficiencies of both di- and tributyltin were above 90%, while SFE recovery of monobutyltin was less than 70%.

An extraction efficiency of greater than 80% was obtained for tributyltin from both spiked and certified reference material (CRM) sediment using 347 atm, 60 °C and 20% methanol-modified CO_2.[36] The same group also studied in situ derivatization of butyltin and phenyltin from sediment using hexylmagnesium bromide as a derivatizing agent and supercritical CO_2 as an extraction fluid. Once again, the recovery of monobutyltin was less than 20%, while recovery of di- and tributyltin was greater than 80%. Recoveries of monophenyltin, diphenyltin and triphenyltin were 40, 106, and 114%, respectively.[37] Chau et al.[38] compared extraction efficiencies of butyltin spiked in sediment and a CRM. Their results showed that up to 60% of monobutyltin could be extracted from spiked sediment, while the same analyte from the CRM using pure CO_2 was not detected in the extract. Extraction of the same analyte from the CRM using Na(DDC)-modified CO_2 increased the extraction efficiency of monobutyltin to 148%. The extraction efficiencies of di- and tributyltin from the CRM using Na(DDC)-modified CO_2 were 79 and 85%, respectively.

SFE of alkyltin and alkyllead from solid samples (e.g. soil and polyvinyl chloride (PVC)) using both supercritical and subcritical chlorodifluoromethane was investigated.[39] Quantitative recoveries of triethyllead and triethyltin were obtained, while low recoveries of trimethyltin and trimethyllead were obtained. Addition of Na(DDC) as a complexing agent to the extraction vessel increased extraction efficiencies of the two trimethylmetallic compounds to above 89%. They also demonstrated that the addition of Na(DDC) decreased extraction time by half. In addition, the extraction efficiencies of alkyltin and alkyllead compounds were improved using subcritical chlorofluoromethane instead of supercritical chlorofluoromethane. A higher extraction efficiency of dibutyltin stabilizer from PVC plastic was

also obtained via subcritical chlorofluoromethane than dichloromethane liquid solvent extraction.

Johansson et al.[40] extracted ionic alkyllead from both sediment and urban dust using modified supercritical CO_2. They used different modifiers (methanol, water, and acetone) to optimize extraction efficiencies. Methanol provided the most favorable recoveries. At the optimum extraction conditions (80 °C, 446 atm, and 10% methanol added to the cell) recovery of trimethyllead, triethyllead, and diethyllead was 96, 106, and 80%, respectively.

Recently, Wenclawiak and Krak[41] measured the extraction efficiencies of dimethylarsenic acid (DMA) **(21)** and monomethylarsenic acid (MMA) **(22)** from a solid sample using supercritical CO_2. The DMA and MMA were derivatized under supercritical CO_2 conditions using thioglycolic acid methylester (TGM) **(23)**. After 10 min of static extraction, the valve between the restrictor and extraction vessel was opened and the derivatized analytes were collected in a solvent. Extraction efficiencies of both DMA and MMA were greater than 90% at 400 atm. Temperature did not affect the extraction efficiencies of the derivatized analytes.

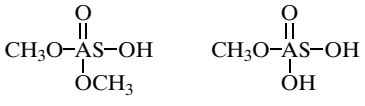

(21) **(22)**

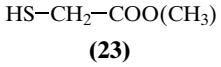

(23)

4 CONCLUSION

Direct extraction of metal ions and preformed metal chelates via SFE from many matrices has been accomplished. Different ligands such as β-diketones, dithiocarbamates, crown ethers, and organophosphorus reagents can be utilized for extraction of metal ions. Results from different studies have demonstrated that fluorinated ligands afford higher metal complex solubility in supercritical CO_2, which makes them more effective for extraction of metal ions. In order to obtain sufficient extraction not only should solubility be considered but also the stability of the complex. In the past few years several different review papers regarding both solubility and extraction behavior of inorganic and organometallic compounds in supercritical CO_2 have been published.[42–44] SFE of inorganic species is a growing field and one that offers a promising future.

ABBREVIATIONS AND ACRONYMS

AA	Acetylacetone
AAS	Atomic Absorption Spectrometry
AcAc	Acetylacetonate
CMP	Dihexyl(*N*,*N*-diethylcarbamoyl)methyl Phosphate
CMPO	Octyl(phenyl)(*N*,*N*-diisobutylcarbamoyl)methylphosphine Oxide
CRM	Certified Reference Material
DBDC	Dibutyldithiocarbamate
DDC	Diethyldithiocarbamate
DEHPA	Di-(2-ethylhexyl)phosphoric Acid
DHC	Dihexyldithiocarbamate
DIDPA	Diisodecylphosphoric Acid
DMA	Dimethylarsenic Acid
FDDC	Bis(trifluoroethyl)dithiocarbamate
FIA	Flow Injection Analysis
FOD	Europium 2,2-dimethyl-6,6,7,7,8,8,8-heptafluoro-3,5-octandione
GC	Gas Chromatography
HFA	Hexafluoroacetylacetonate
HPLC	High-performance Liquid Chromatography
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
MMA	Monomethylarsenic Acid
MT	Metallothionein
PDTC	Pyrrolidine Dithiocarbamate
PEEK	Polyether Ether Ketone
PVC	Polyvinyl Chloride
P3DC	Dipropyldithiocarbamate
P5DC	Dipentyldithiocarbamate
SF	Supercritical Fluid
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction
TAA	Trifluoroacetylacetone
TBP	Tributylphosphate
TBPO	Tributylphosphine Oxide
TDBDC	Tetrabutylammonium Dibutyldithiocarbamate
TGM	Thioglycolic Acid Methylester
TOPO	Tri-*n*-octylphosphine Oxide
TPPO	Triphenylphosphine Oxide
TTA	Thenoyltrifluoroacetone
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Heavy Metals Analysis in Seawater and Brines

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Solid-phase Microextraction in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis

Liquid Chromatography **(Volume 13)**
Supercritical Fluid Chromatography

REFERENCES

1. R. M'Hamidi, J.F. Bocquet, K. Chhor, C. Pommier, 'Solubility and Decomposition Studies on Metal Chelates in Supercritical Fluid for Ceramic Precursor Powder', *J. Supercrit. Fluids*, **4**, 55–59 (1991).
2. M. Ashraf-Khorassani, M.T. Combs, L.T. Taylor, unpublished results.
3. Y. Lin, R.D. Brauer, K.E. Laintz, C.M. Wai, 'SFE of Lanthanides and Actinides from Solid Materials with Fluorinated β-Diketone', *Anal. Chem.*, **56**, 2549–2551 (1995).
4. M. Ashraf-Khorassani, M.T. Combs, L.T. Taylor, 'Solubility of Metal Chelates and Their Extraction from an Aqueous Environment via Supercritical CO_2', *Talanta*, **44**, 755–763 (1997).
5. K.E. Laintz, C.M. Wai, C.R. Yonker, R.D. Smith, 'Solubility of Fluorinated Metal Diethyldithiocarbamates in Supercritical CO_2', *J. Supercrit. Fluids*, **4**, 194–198 (1991).
6. C.M. Wai, Y. Lin, R. Brauer, S. Wang, F. Beckert, 'Supercritical Fluid Extraction of Organic and Inorganic Mercury from Solid Materials', *Talanta*, **40**, 1325–1330 (1993).
7. M. Ashraf-Khorassani, L.T. Taylor, 'Supercritical Fluid Extraction of Mercury(II) Ion via In Situ Chelation and Preformed Mercury Complexes from Different Matrices', *Anal. Chim. Acta*, **379**, 1–9 (1999).
8. J. Wang, W.D. Marshall, 'Recovery of Metals from Aqueous Media by Extraction with Supercritical CO_2', *Anal. Chem.*, **66**, 1658–1661 (1994).
9. C.M. Wai, S. Wang, J. Yu, 'Solubility Parameters and Solubilities of Metal Dithiocarbamates in Supercritical CO_2', *Anal. Chem.*, **68**, 3516–3519 (1996).
10. S.H. Page, S.R. Sumpter, S.R. Goats, M.L. Lee, 'Tri-*n*-butyl phosphate/CO_2 and acetone/CO_2 Phase Behaviors and Utilities in Capillary SFC', *J. Supercrit. Fluid*, **6**, 95–99 (1993).
11. W.J. Schmitt, R.C. Reid, 'The Solubility of Paraffinic Hydrocarbons, and Their Derivatives in Supercritical CO_2', *Chem. Eng. Commun.*, **64**, 155–158 (1988).
12. Y. Lin, N.G. Smart, C.M. Wai, 'Supercritical Fluid Extraction of Uranium and Thorium for Nitric Acid Solutions with Organophosphorus Reagent', *Environ. Sci. Tech.*, **29**, 2706–2708 (1995).
13. N.G. Smart, T.E. Carleson, S. Elshani, S.E. Wang, C.M. Wai, 'Solubility of Chelating Agent and Metal-containing Compounds in Supercritical CO_2', *Talanta*, **44**, 137–149 (1997).
14. Y. Meguro, S. Iso, T. Sasaki, Z. Yoshida, 'Solubility of Organophosphorus Metal Extractants in Supercritical CO_2', *Anal. Chem.*, **70**, 774–780 (1998).
15. S. Wang, S. Elshani, C.M. Wai, 'Selective Extraction of Mercury with Ionizable Crown Ethers in Supercritical CO_2', *Anal. Chem.*, **67**, 919–923 (1995).
16. C.M. Cowey, K.D. Bartle, M.D. Burford, A.A. Clifford, S. Zhu, N.G. Smart, N.D. Tinker, 'Solubility of Ferrocene and Nickel Complex in Supercritical Fluid', *J. Chem. Eng. Data*, **40**, 1217–1221 (1995).
17. Y. Lin, C.M. Wai, 'SFE of Lanthanides with Fluorinated β-Diketones and Tributyl Phosphate', *Anal. Chem.*, **66**, 1971–1975 (1994).
18. Y. Lin, C.M. Wai, F.M. Jean, R.D. Brauer, 'SFE of Thorium and Uranium Ions from Solid and Liquid Materials Lanthanides with Fluorinated β-diketones and Tributyl phosphate', *Environ. Sci. Tech.*, **28**, 1190–1193 (1994).
19. K.G. Furton, L. Chen, R. Jaffe, 'Rapid Determination of Uranium on Solid Matrices by Synergistic In Situ Chelation SFE and UV Absorption Spectroscopy', *Anal. Chim. Acta*, **304**, 203–208 (1995).
20. K.E. Laintz, C.M. Wai, C.R. Yonker, R.D. Smith, 'Extraction of Metal Ions from Liquid and Solid Materials by Supercritical CO_2', *Anal. Chem.*, **64**, 2875–2878 (1992).
21. Y. Liu, V. Lopez-Avila, M. Alcaraz, W.F. Beckert, E.M. Heithmar, 'Determination of Metals in Solid Samples by Complexation SFE/GC/AED', *J. Chromatogr. Sci.*, **31**, 310–316 (1993).
22. C.M. Wai, 'SFE of Trace Metals from Solid and Liquid Materials for Analytical Applications', *Anal. Sci.*, **11**, 165–167 (1995).
23. J. Wang, W.D. Marshall, 'SFE with CO_2 for the Characterization of Cadmium, Copper, Zinc Bound Metallothioneums', *Analyst*, **120**, 623–628 (1995).
24. K.L. Toews, R.M. Shroll, C.M. Wai, 'pH Defining Equilibrium Between Water and Supercritical CO_2. Influence on SFE of Organic and Metal Chelates', *Anal. Chem.*, **67**, 4040–4043 (1995).
25. K.E. Laintz, E. Tachikawn, 'Extraction of Lanthanides from Acidic Solution using Tributyl-phosphate Modified Supercritical CO_2', *Anal. Chem.*, **66**, 2190–2193 (1994).
26. Y. Meguro, S. Iso, H. Takeishi, Z. Yoshida, 'Extraction of Uranium(V) in Nitric acid with Supercritical CO_2 Fluid Containing Tributylphosphorus', *Radiochim Acta*, **75**, 185–191 (1996).
27. Y. Meguro, S. Iso, Z. Yoshida, 'Correlation Between Extraction Equilibrium of Uranium(VI) and Density of CO_2 Medium in a HNO_3 Supercritical CO_2 Tributylphosphate System', *Anal. Chem.*, **70**, 1262–1267 (1998).
28. S.J. Blund, P.A. Cusack, R. Hill, (eds.) *The Industrial Use of Tin Chemicals*, The Royal Society of Chemistry, London, 1986.

29. S.J. Blunden, A. Chapman, in *Organometallic Compounds in the Environment: Principles and Reactions*, eds. P.J. Craig, John Wiley & Sons, New York, 11, 1986.
30. J.W. Oudsema, C.F. Poole, 'On-line SFE/SFC of Organotin with Packed Microbore Columns and Formic Acid-modified CO_2', *Fresenius' J. Anal. Chem.*, **344**, 426–434 (1992).
31. J.W. Oudsema, C.F. Poole, 'Determination of an Organic Stabilizer in a Rigid Poly(Vinyl Chloride) Plastic by On-line SFE/SFC with Formic Acid-modified CO_2 and FID', *J. High Resolut. Chromatogr.*, **16**, 198–200 (1993).
32. Y. Lin, V. Lopez-Avila, M. Alcaraz, W.R. Beckert, 'Determination of Organotin Compounds in Environmental Samples by SFE/GC/AED', *J. High Resolut. Chromatogr.*, **16**, 106–112 (1993).
33. Y. Lin, V. Lopez-Avila, M. Alcaraz, W.R. Beckert, 'Off-line Complexation/SFE/SFC/GC/AED for Determination and Speciation of Organotin Compounds in Soils and Sediments', *Anal. Chem.*, **66**, 3788–3796 (1994).
34. R. Alzaga, J.M. Bayona, 'SFE of Tributyltin and Its Degradation Products from Seawater via Liquid–Solid-phase Extraction', *J. Chromatogr.*, **655**, 51–56 (1993).
35. Y. Cai, J. Bayona, 'Simultaneous Speciation of Butyl, Phenyl, and Cyclohexyltin Compounds in Aqueous Matrices Using Ethylation Followed by Solid-phase Trace Environmental SFE, and GC Determination', *Chromatogr. Sci.*, **33**, 89–97 (1995).
36. J. Dachs, R. Alzaga, J.M. Bayona, P. Quevauviller, 'Development of a SFE Procedure for Tributyltin Determination in Sediments', *Anal. Chim. Acta*, **286**, 319–327 (1994).
37. Y. Cai, R. Alzaga, J.M. Bayona, 'In Situ Derivatization and SFE for the Simultaneous Determination of Butyltin and Phenyltin Compounds in Sediment', *Anal. Chem.*, **66**, 1161–1167 (1994).
38. Y.K. Chau, F.Yang, M. Brown, 'SFE of Butyltin Compounds from Sediment', *Anal. Chim. Acta*, **304**, 85–89 (1995).
39. K. Li, S.F.Y. Li, 'Determination of Alkyllead and Alkyltin Compounds in Solid Samples by Supercritical and Subcritical Fluid Extraction and MEKC', *J. Chromatogr. Sci.*, **33**, 309–315 (1995).
40. M. Johansson, T. Berglof, D.C. Baxter, W. Frech, 'SFE of Ionic Alkyllead Species from Sediment and Urban Dust', *Analyst*, **120**, 755–759 (1995).
41. B.W. Wenclawiak, M. Krak, 'Reactive SFE/SFC of Arsenic Species', *Fresenius' J. Anal. Chem.*, **351**, 134–138 (1995).
42. Y. Lin, N.G. Smart, C.M. Wai, 'SFE and SFC of Metal Chelates and Organometallic Compounds', *Trends Anal. Chem.*, **14**, 123–128 (1995).
43. M. Ashraf-Khorassani, M.T. Combs, L.T. Taylor, 'SFE of Metal Ions and Metal Chelates from Different Environments', *J. Chromatogr. A*, **774**, 37–48 (1997).
44. C.M. Wai, S.F. Wang, 'SFE: Metals as Complexes', *J.Chromatogr. A*, **785**, 369–384 (1997).

Supercritical Fluid Extraction of Organics in Environmental Analysis

Bruce A. Benner, Jr
National Institute of Standards and Technology, Gaithersburg, USA

1 Introduction 1
2 Theory and Instrumentation 1
2.1 What is a Supercritical Fluid? 1
2.2 Thermodynamics and Kinetics of Supercritical Fluid Extraction 2
2.3 Supercritical Fluid Extraction Instrumentation 3
3 Supercritical Fluid Extraction of Environmental Samples 4
4 The Future of Supercritical Fluid Extraction of Environmental Samples 7
Disclaimer 7
Abbreviations and Acronyms 7
Related Articles 8
References 8

Supercritical fluids (SFs) possess several physical properties that make them attractive substitutes for conventional liquid solvents in analytical-scale extractions. Molecular diffusivities of SFs are typically a factor of 10 greater (10^{-4} vs 10^{-5} $cm^2\,s^{-1}$) and their viscosities are a factor of 10 lower (10^{-4} vs 10^{-3} $N\,s\,m^{-2}$) than liquid solvents. These properties improve the rate of mass transfer of a solute from the matrix surface to the bulk extractions fluid – a transfer believed to be the rate-limiting step of most extractions. Under optimized conditions, the improvements in mass transfer can enable a supercritical fluid extraction (SFE) to be completed in less than 1 h compared to ≥12 h for extractions involving conventional solvent techniques. The typical SFE also generates little solvent waste, compared to traditional extraction methods, but limitations have also emerged with respect to the efficiency of SFE as related to the sample composition (matrix effect). This review will discuss the progress made over the last 20 years in SFE as an alternative technique to liquid solvent extractions of environmental samples for the subsequent measurement of organic constituents.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Over the last 25 years, the field of environmental analytical chemistry has gleaned valuable information regarding the levels of global and regional contamination. In the process, analysts have generated significant quantities of potentially hazardous waste, particularly during the extraction of samples. Sample extraction, although generally regarded as less technically demanding than the modern separation and detection techniques, is certainly one of the most important parts of an analytical scheme, and if not chosen carefully, can doom an environmental analysis to failure. An ideal extraction technique should be fast, simple, and cheap. It should quantitatively extract the species of interest while minimizing decomposition, subsequent sample clean-up, and solvent waste.[1] The Soxhlet extraction method, typically used by most environmental analysts, has not changed significantly in nearly 100 years and generally does not fulfill most of the requirements of an ideal extraction method. For example, Soxhlet extraction generates hundreds of milliliters of waste solvent per sample, typically takes 12–48 h to complete, and generally produces an extract that, although quantitative for most analytes associated with the sample (depending on the solvent used), requires clean-up prior to analysis. The cost to dispose of contaminated solvent waste in a responsible manner is usually even greater than the considerable expense of pure solvents employed for extractions of environmental samples. The time required for the extraction and subsequent clean-up, the expense of the pure solvents needed for the extractions, and the expense of disposal of large volumes of waste solvent, combine to make Soxhlet extraction a very expensive sample extraction technique indeed. These considerations have thus motivated research in minimizing the time and considerable waste generated during the conventional liquid solvent extraction of environmental samples (e.g. soil, sediment, solid waste, combustion material, air particulate matter).

There have been a number of reviews discussing the role of SFE in environmental analysis,[2–6] since this was initially believed to be the most important role of the technique. This review will discuss the progress made over the last 20 years in SFE as an alternative technique to liquid solvent extractions of environmental samples for the subsequent measurement of organic constituents.

2 THEORY AND INSTRUMENTATION

2.1 What is a Supercritical Fluid?

The simplest way to define an SF is to refer to the phase diagram for the specific substance. As an example, a phase

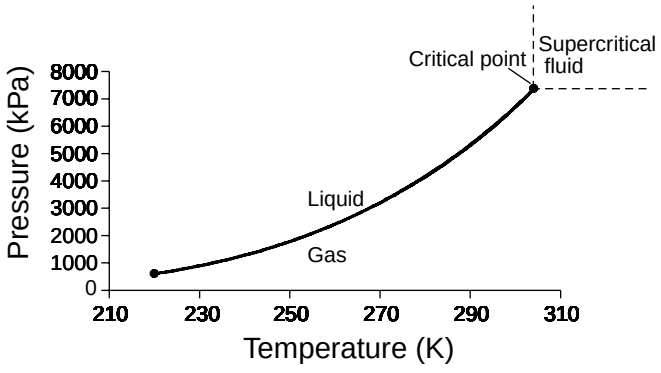

Figure 1 CO_2 phase boundary.

diagram for CO_2 is presented in Figure 1, showing the liquid/gas phase boundary as pressure and temperature are varied. At the pressure of 7397 kPa and temperature of 304 K, the phase boundary between liquid and gaseous CO_2 becomes indistinguishable and this condition defines the critical point for CO_2. Pressures and temperatures above those of the critical point produce an SF, and, in fact, most workers using SF CO_2 for analytical-scale extractions operate at conditions significantly above both the critical pressure and temperature (e.g. pressure $\geq$ 35 000 kPa and temperature $\geq$ 373 K). There is nothing ''magic'' about an SF; in fact, some workers in the fields of supercritical fluid chromatography (SFC) and SFE have suggested renaming the techniques because the ''super-'' prefix may intimidate some potential users not familiar with the rather simple thermophysical definitions. No new name has been generally accepted by those in the SFC and SFE fields, so promoters continue to publish and present work demonstrating the benefits and advantages of techniques based on SFs over traditional solvent-based methods.

One of the most significant benefits of the most commonly used SF, CO_2, is that it does not create a waste disposal or health hazard problem for those using it in extractions. Other practical benefits of this fluid include its relatively low critical temperature and pressure, low reactivity, nonflammability, and low cost for high purity. Supercritical CO_2 is not a universal solvent because its solvent strength under the normal range of pressures (8100–61 000 kPa) allows it to solubilize mainly nonpolar or moderately polar compounds.[1] For extraction of more polar species, one must use another fluid of greater polarity or CO_2 modified with a liquid solvent. For some time now, the use of N_2O as an SF has been discouraged because it is an oxidizing agent, and in at least one case, has caused an explosion during the extraction of a solid sample of high organic content.[7] A number of other fluids, mostly refrigerants,[8] have also been employed as SFs, and specific applications will be detailed below.

2.2 Thermodynamics and Kinetics of Supercritical Fluid Extraction

SFs possess several physical properties that make them attractive substitutes for conventional liquid solvents in analytical-scale extractions. Molecular diffusivities of SFs are typically a factor of 10 greater (10^{-4} vs 10^{-5} $cm^2\,s^{-1}$) and their viscosities are a factor of 10 lower (10^{-4} vs 10^{-3} $N\,s\,m^{-2}$) than liquid solvents.[1] These properties improve the rate of mass transfer of a solute from the matrix surface to the bulk extraction fluid – a transfer believed to be the rate-limiting step of most extractions. Under optimized conditions, the improvements in mass transfer can enable an SFE to be completed in less than 1 h compared to more than 12 h for extractions involving conventional solvent techniques.

Referring to the importance of mass transfer in extractions (see above), the ability of a SF to solubilize a specific analyte or class of analytes, at a given set of conditions, does not necessarily guarantee that SFE will extract all of the species of interest associated with a sample. Competition between the matrix sites and the SF for the analyte is probably a more important consideration in selecting a specific SF or solvent-modified SF for an extraction experiment.[1,6] However, if the analytes are not appreciably soluble in the SF, as demonstrated by spiking experiments on an inert substrate, then it is unlikely that quantitative SFE is possible with that specific set of analytes, sample matrix, and fluid. The solvating strength of SFs can approach those of liquids but only at very high pressures, since this parameter is directly related to fluid density. However, for volatile species, increasing the temperature of extraction can have a greater effect on extraction efficiencies than increasing the pressure.[9,10] Since the solvating power of an SF can be varied by pressure, some have suggested that an SFE could be made selective for a specific compound class by optimizing the pressure and temperature of the extraction.[5,11–14] However, King[5] stated that unless the classes of analytes sought by an analyst differed significantly in their physical properties, selective extraction of the two chemical classes by SFE would be unlikely. The work described by another group[14–16] employed sub- and supercritical water for selective chemical class extractions from environmental solids. Operating at constant pressure, with the knowledge that water's dielectric constant decreases significantly with increasing temperature, they observed chemical class fractionation during SFE of environmental solids at different temperatures, with polar species extracting at lower temperatures (50 °C) and nonpolar solutes extracting at higher temperatures (250 °C). These citations document the relatively few successes in selectively extracting specific species from complicated matrices. The majority of work in SFE

has attempted to extract as much of specific analytes as that extracted by conventional liquid-based methods (e.g. Soxhlet), and has usually required temperature and pressure conditions not conducive to selective extractions.

A model proposed by Pawliszyn[17] predicts that if conditions can be optimized to ensure no retention of the analyte by the matrix, complete extraction of the analyte can be performed by passing two void volumes of the SF through the extraction vessel. He discussed in detail how, for environmental solids, the rate of mass transport of the analyte from the matrix surface to the bulk fluid (desorption and diffusion rate-limiting processes) increases as the extraction temperature increases. An interesting requirement of the model is that the extraction fluid must be soluble in the organic layer presumed to coat most environmental solids, in order for the solute to be extracted into the bulk fluid. Perhaps a more realistic scenario would be that for extraction to take place, some or most of this organic layer would need to be stripped from the particle to expose the solute to the extraction fluid. This would make any selective extraction of an analyte sorbed to the surface of a particle (or pore of a particle) difficult, as the co-extracted organic material would certainly make the extract more complex.

Another model was developed by Bartle et al.,[18] which theorized the extraction kinetics of an idealized sample consisting of spherical particles of similar size in which the analytes of interest are uniformly distributed. Two other assumptions of this model were that the flow of SF past the sample was fast enough to prevent a significant concentration of the analyte from accumulating in the fluid and that the analytes move through the matrix by diffusion. Plotting the kinetic extractions profiles $[\ln(m/m_0)]$ vs time, where m is the mass of the analyte remaining with the sample after time t and m_0 is the mass of the analyte associated with the sample at time zero, of both ideal and real samples, they observed an exponential decline in m, followed by subsequent linear decay over a long period of time. The m_0 or total mass of analyte associated with the sample could then be extrapolated from extracted analyte masses obtained at three different times, one during the exponential decay period and two taken at equal time periods during the linear decay of the mass of analyte associated with the sample.

Clifford et al.[19] refined the model proposed by Bartle et al.[18] which considered the single rate-limiting step of an SFE to be diffusion of the solute from interior pores to the edge of the sample particle. Clifford et al.[19] proposed that reversible desorption of the solute from matrix sites and removal of the solute from the edge of the sample particle into the bulk fluid should also be considered as rate-determining processes in SFE. This was the first SFE model to acknowledge the importance of solute–matrix interactions in directly affecting the efficacy of SFE with some environmental samples. This matrix effect is probably one of the most prominent limitations of SFE, as workers have observed significant differences in SFE extraction efficiencies of the same solutes from similar samples extracted under identical conditions.[20] Specific examples will be discussed below.

2.3 Supercritical Fluid Extraction Instrumentation

Instrumentation requirements for SFE range from simple to complex, depending mostly on the level of automation required for the application. For a simple manual instrument set-up for SFE, as shown in Figure 2, one needs a fluid pump capable of relatively high pressures (20–50 MPa), a sample vessel and associated plumbing (tubing) rated at significantly above the method operating pressures, a valve or flow restrictor, and a collection device, using either a liquid- or solid-phase medium. Commercially available SFE instrumentation typically consists of high-pressure reciprocating or syringe pumps. The reciprocating pumps are typically modified high-pressure liquid chromatographic pumps that allow continuous extraction, whereas syringe pumps are typically filled and pressurized prior to extracting. Syringe pumps can typically operate at higher pressures – as high as 68.9 MPa – than reciprocating pumps, and thus allow for investigations at higher-pressure regimes. Reciprocating pumps offer the advantage of larger extraction volumes than syringe pumps, since the limiting volume for the reciprocating pump would be that of the fluid supply tank. Automated commercial instruments are available for each of these pump systems.

Solid samples are typically packed into high-pressure cylindrical vessels and contained between frits and end-fittings. One manufacturer's instrument extracts the sample cartridge in a chamber whereby the inside and outside surfaces of the extraction vessel are held at the same pressure, negating any need for thick and expensive metal extraction vessels. Most researchers in SFE recommend minimizing dead volume in the sample extraction vessel by adding inert material to the cell after the sample.[6] This maximizes contact between the sample and extracting fluid.

The simplest collection method involves immersing the flow restrictor, employed to maintain the sample vessel at

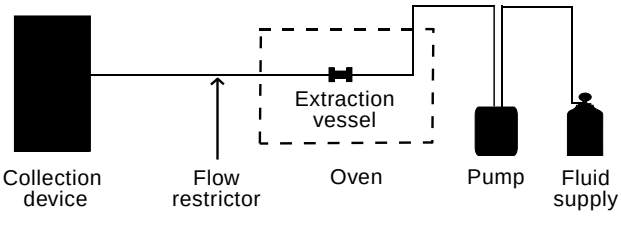

Figure 2 SFE apparatus for manual operation.

supercritical pressures, in a suitable solvent and allowing the extract to dissolve in the solvent as the CO_2 decompresses. The cooling (Joule–Thompson cooling) resulting from the expansion of the CO_2 can improve collection efficiencies of volatile compounds as well as minimizing the decomposition of thermally labile species. Mulcahey et al.[21] discussed potential problems that may arise during collection of an SFE in liquid solvent, including plugging of the restrictor by ice and losses of analytes as aerosols during the decompression of the extraction fluid. They evaluated the use of solid-phase traps for collection of species of different polarities, testing all the different traps for collection of polychlorinated biphenyls (PCBs) extracted by SFE from river sediment. Thompson et al.[22] investigated the effectiveness of nine solvent collection systems for a test mixture consisting of solutes with different polarities. They found that multicomponent solvent mixtures were necessary for quantitative collection of all of the analytes in the test mixture. Two negative aspects of collecting on solid-phase traps are that the researcher must choose both the appropriate packing material for the traps as well as the type and volume of wash solvent. Miller et al.[23] investigated the collection of aliphatic and aromatic hydrocarbons, and PCBs after static SFE. The extracts were collected without a flow restrictor after rapid depressurization into an empty vial, with good recoveries (>90%) obtained.

Automation has probably necessitated some of the more innovative developments in SFE instrumentation, and most systems can process from 8 to 24 samples. These systems typically include a variable-orifice restrictor that is controlled through feedback to the pumps. As a variable-orifice restrictor begins to plug during an extraction and the flow rate lessens, the orifice increases in size, causing expulsion of the plugging material, and the flow rate returns to the set-point. Simple linear restrictors, although inexpensive and easy to obtain, are prone to plugging and are generally not suitable for use in an automated instrument. Automated instruments, in turn, are the most appropriate for method development because multiple extracts can be generated from the same sample by methods using different extraction parameters (e.g. pressure, temperature, static and dynamic extraction times, fluid volume, fluid flow rate). Most automated SFE instruments will also allow access to sample logs, in which specific extraction parameters are recorded for each extracted sample.

3 SUPERCRITICAL FLUID EXTRACTION OF ENVIRONMENTAL SAMPLES

One can classify environmental samples in three general ways. The first and most effectively extracted by SFE would include those samples whose particle surfaces are coated with the contaminating material, such as a fuel, waste oil, or pesticide formulation. In these cases and as long as the fluid employed can solvate the analytes of interest, SFE can be an effective alternative extraction technique to conventional solvent methods. The second sample type has a lower level of contamination and the species are strongly sorbed to surfaces of pores deep within the sample matrix. This sample type is much more difficult to extract by SFE, because, as described in model representations above, the fluid must overcome the interactions between the analyte and pore surface for the analyte to partition into the pore fluid and then diffuse into the bulk fluid before exiting the extraction vessel. This type of sample may also have experienced some aging and weathering, such that the external surfaces of the particle may be quite free of contamination and the bulk of the contaminating species are sequestered in the pores of the particles. The third type of sample includes both pore- and surface-associated analytes with a virtual continuum of relative proportions of each. Generally, the efficacy of SFE with respect to these sample matrices will be proportional to the relative amounts of analyte sorbed to the external surface of the sample particle (easy-to-extract analytes) compared to those deep within the pores (difficult-to-extract analytes) of the sample matrix. Other specific sample characteristics that may have direct effects on the efficiency of SFE include percentage of carbon, clay content, specific surface area, acidity, and pore size distribution.

Most of the environmental SFE applications published in the last 15 years, some examples of which are noted below, have involved the extraction and measurement of aliphatic hydrocarbons,[14,24–30] aromatic hydrocarbons,[9,14,20,31–55] PCBs,[9,14,33,37,41,43,44,56–73] pesticides[59,60,72–89] and herbicides.[74,90–97] The United States Environmental Protection Agency (USEPA) has proposed SFE methods for extraction of total recoverable petroleum hydrocarbons (EPA method 3560), polycyclic aromatic hydrocarbons (PAHs, EPA method 3561), and PCBs and organochlorine pesticides (EPA method 3562), but these methods have yet to be approved.

Burford et al.[24] developed an on-line SFE/gas chromatography (GC) method for determining gasoline and diesel fuel organics from environmental samples, observing comparable recoveries for all the fuel hydrocarbons compared to a sonication method with dichloromethane as the solvent, except for those of gasoline where the SFE-based method yielded higher recoveries than the conventional method. Higher recoveries observed for gasoline-contaminated samples were attributed to more efficient collection of the more volatile hydrocarbons by the SFE method. Similar work was also performed[26] using an off-line method during which

fuel-contaminated soil samples were processed by an SFE method (40 MPa CO_2 at 150 °C for 30 min) and a Soxhlet method using Freon-113 (1,1,2-trichloro-1,2,2-trifluoroethane) followed by analysis by both GC with flame ionization detection and infrared (IR) spectrophotometry (absorption at 2930 cm^{-1}). The SFE and Soxhlet extractions gave comparable results for the fuel-spill soils except for the gasoline- and kerosene-contaminated samples where the SFE method yielded higher recoveries than the solvent-based technique – similarly to results generated using the on-line method.[24] Eckert-Tilotta et al.[26] also discussed the important differences between real-world spill samples where the contaminants can diffuse into the sample matrix, making them difficult to extract, and spiked samples, where the analytes are generally on the surface of the sample, and are easier to extract. Lopez-Avila et al.[28] used SFE to extract petroleum hydrocarbons from soil samples, processing extracts through silica gel to remove polar species and determining the total fuel hydrocarbons by IR spectrophotometry. The results of this nonspecific quantitative method were comparable to a traditional Soxhlet extraction method using Freon-113 as the solvent. An interlaboratory study was performed later[29] that compared the accuracy and precision of the SFE-based methods for extracting petroleum hydrocarbons from soils. Mean recoveries for the SFE extraction method and the two types of IR instruments employed for the study ranged from 75.9 to 107% for three certified reference soils. For a clay soil spiked with petroleum hydrocarbons and containing 30% water, recoveries from the different laboratories were <38%, suggesting the need for drying agents for wet samples.

There have been a number of reports discussing the results of SFE of natural matrix samples for the subsequent measurement of PAHs. Schantz and Chesler[98] reported good recoveries of PAHs from Standard Reference Material ((SRM), available from the National Institute of Standards and Technology (NIST)) 1649 (Urban Dust/Organics) using supercritical CO_2 for 4 h extractions. Hawthorne et al.[40,99,100] have used SRMs for more than 10 years in validating both off-line and on-line SFE approaches. They found that CO_2 and N_2O alone are not able to completely extract PAHs from the urban dust (SRM 1649) and sediment (SRM 1941) SRMs. They also reported that methanol-modified CO_2 and N_2O provided better PAH recoveries than the neat fluids, but the residual methanol created some problems with trapping and analysis. Lopez-Avila et al.[85] used a marine sediment Certified Reference Material (CRM) from the National Research Council of Canada and a PAH-contaminated soil to evaluate SFE experimental conditions. They found that extraction recoveries of PAHs from both materials were poor when neat CO_2 was used as the fluid, and somewhat better, though generally less than quantitative (most recoveries <70% of Soxhlet extraction results), when methanol-modified CO_2 was employed as the fluid. During SFE of SRM 1941 (Organics in Marine Sediment), Pyle and Setty[101] observed recoveries of ≤50% for most of the PAHs measured using CO_2 as the extraction fluid. Porter et al.[102] described off-line SFEs of SRMs 1941 and 1974 (Organics in Mussel Tissue) using 10% methylene chloride-modified CO_2 as the fluid. They reported good extraction recoveries compared with certified concentrations for all of the PAHs measured, and were the first group to report SFE results for the frozen mussel tissue reference material (SRM 1974).

The refrigerant chlorodifluoromethane (R22) has been employed for use in SFE of environmental materials (e.g. air particulate matter, waste sludge, diesel particulate matter).[20,41,51] Hawthorne et al.[41] found that R22 provided faster extraction kinetics and overall better recoveries of PCBs from SRM 1939 [Polychlorinated Biphenyls (Congeners) in River Sediment] than other fluids (CO_2 and N_2O) and modified fluids (methanol-modified CO_2). Paschke et al.[51] reported that R22 yielded excellent recoveries of PAHs from SRM 1650 (Diesel Particulate Matter) compared with NIST's certified results. Two disadvantages of R22 are its potential adverse impacts on stratospheric ozone and global warming.[103]

High-temperature (200 °C) CO_2 extractions of SRM 1649 (Urban Dust/Organics) have yielded significant enhancements in recoveries of PAHs compared with more moderate temperature extractions (50 °C) at the same pressures.[43] Yang et al.[55] found that improved PAH recoveries by SFE at 200 °C from environmental solids were observed for three different matrices (marine sediment, diesel soot, and air particulate matter). They also observed that addition of modifiers did not always improve PAH recoveries compared with pure CO_2 at the same temperatures. Perhaps the most significant conclusion reached by Yang et al.[55] was that improvements in recoveries of PAHs affected by modifiers depended more on the interactions of the modifier with the matrix substrate than on the interactions between the modifier and the analyte (solubility). In this way, the modifier sorbs to the matrix sites, discouraging the readsorption of the analytes to the matrix sites, and thus facilitating analyte transport into the bulk fluid. It should be noted that for all of the temperature/modifier and sample matrix combinations investigated by Yang et al., recoveries of the highest molecular weight PAHs analyzed, indeno[1,2,3-*cd*]pyrene and benzo[*ghi*]perylene (both 276 *m/z* isomers), were usually ≤70% of certified results or those obtained by Soxhlet extraction. One group[44] reported recoveries of phenanthrene, fluoranthene, and pyrene in

SRM 1649 (Urban Dust/Organics) from 150% to nearly 300% of the certified concentrations using different modifier–CO_2 combinations at a moderate temperature (80 °C). Another worker[20] performed the same extractions under the same conditions as Langenfeld et al.[44] and observed good recoveries of the same three PAHs (≥86% of certified concentrations) but not 1.5 to 3 times the certified concentrations as previously reported.[44] Benner[20] applied SFE to four natural matrix SRMs using different fluids [CO_2 (neat and modified), chlorodifluoromethane, and 1,1,1,2-tetrafluoroethane], temperatures (60 °C, 150 °C, and 200 °C), and added water. For two of the four sample matrices (SRMs 1649 and 1944), conditions were optimized to obtain recoveries similar to those achievable with Soxhlet extraction. For SRM 1941a (Organics in Marine Sediment) and SRM 1650 (Diesel Particulate) recoveries for benzo[*ghi*]perylene, the largest of the PAHs studied in this work, were at best ≤50% and ≤20%, respectively, of the certified concentrations. Meyer et al.[49] attempted to extract PAHs from a soil with a high carbon content (a mass fraction of approximately 50% C) and found that 5% toluene-modified CO_2 at 140 °C yielded similar concentrations of analytes to that of Soxhlet extraction (hexane/acetone). Another group[36] suggested that SFE extracted significantly more PAHs from a highly contaminated sediment than Soxhlet extraction (dichloromethane) and somewhat more than that of microwave-assisted extraction (acetone or dichloromethane). Lee et al.[46] optimized a CO_2-based SFE method for extracting 16 priority pollutant PAHs (naturally incurred) from sediments, using a mixture of water, methanol, and dichloromethane as modifiers, extracting at 120 °C for 70 min. Friedrich et al.[38] used triethylamine-modified CO_2 to demonstrate comparable recoveries of PAHs with respect to Soxhlet for a sewage sludge and sediment sample, though extraction times were as long as 2 h, compared with typical SFE extraction times of 30 min to 45 min.

Unlike PAHs, where there is a typical trend of lower recoveries with increasing molecular weight, the less-chlorinated PCB congeners (e.g. three substituted chlorines) are usually more difficult to extract by SFE than the more heavily chlorinated congeners (four to seven substituted chlorines). As an example of this extraction behavior, one group[41] observed under optimized conditions that PCB 28 (2,4,4′-trichlorobiphenyl) was recovered at 63% of the Soxhlet extraction result from a river sediment (SRM 1939) and the tetra- through hepta-congeners were recovered at from 83% to 128% of the Soxhlet values, respectively (chlorodifluoromethane, 40.5 MPa, 100 °C). Bøwadt et al.[61] performed SFE of PCBs from lyophilized fish tissue, achieving comparable recoveries of PCB congeners to those of the traditional extraction technique. The method included the use of coupled columns (SIL-8 and HT-5 phases), which improved congener separation compared with that of single column methods. Hale and Gaylor[66] also performed SFE of lyophilized fish tissues for PCBs, observing results comparable to those achieved by conventional solvent extraction for a number of PCBs from a candidate fish tissue reference material. Hale and Gaylor found that by adding alumina to the collection vial, one eliminated the need for removal of lipids from the extracts. As part of a comparison of Soxhlet and SFE for extracting PCBs from an industrial soil, researchers[64] found that SFE yielded comparable results to that of Soxhlet for 12 PCBs, quantifying the congeners by using the dual column system described above.[61] Schantz et al.[73] compared Soxhlet and SFE for their efficacy in extracting PCBs from two natural matrix sediment SRMs (SRM 1941a and SRM 1944) and a lyophilized mussel tissue SRM (SRM 2974). The SFE and Soxhlet results compared quite favorably for all three SRMs, suggesting that the SFE method should be considered as an alternative method for extracting PCBs from a variety of environmental samples.

Application of SFE to the extraction of pesticides and herbicides was logical, given the reactive nature of those classes of compounds and the relatively gentle extraction conditions typically used in SFE. McNally and Wheeler[93] spiked soils, plant materials, and cell culture media with sulfonylurea herbicides, extracting with SF CO_2 and analyzing the extracts by SFC. Hopper and King[79] built an SFE extractor that was designed to extract pesticides from a wide variety of foods. Recoveries of spiked pesticides were typically ≥90%, but recoveries of incurred pesticides appeared to be matrix-dependent. Lehotay and Ibrahim[82] developed an SFE method for extracting pentachloronitrobenzene pesticides from a variety of vegetables, with an average recovery reported as 90%. Izquierdo et al.[80] found that SFE of soils and cereals spiked with carbamate pesticides resulted in recoveries, at best, of 6% for methiocarb to 80% for aminocarb. Another group[96] compared SFE with liquid-vortex extraction ((LVE), acetonitrile–water–acetic acid) for their ability to extract four pesticides from soil, analyzing the extracts by enzyme immunoassay analysis (EIA) or GC. The SFE/EIA results for atrazine and alachlor were comparable to those obtained by LVE/GC. Lower recoveries were observed by SFE/EIA for simazine, compared with the LVE/GC method.

The majority of SFE environmental applications, a survey of which is described above, involve extraction and measurement of organic species. One group has employed SFE for the extraction of heavy metals[104,105] and radionuclides,[106,107] after mobilization with a chelating agent and dissolution in supercritical CO_2. These

applications can be scaled-up to serve as techniques for mitigating environmental problems or for reprocessing nuclear waste.

In the last five years, pressurized fluid extraction (PFE), a method using conventional liquid solvents at elevated temperatures and pressures, has been introduced,[108] offering the benefits of significant reductions in both time and solvent consumption compared with Soxhlet extraction. Richter et al.[108] described results of PFE for both spiked and native analytes (PAHs and PCBs) from environmental samples and reference materials. Schantz et al.[109] employed six natural matrix SRMs and two natural matrix CRMs for evaluating the effectiveness of PFE in extracting PAHs, PCBs, and chlorinated pesticides. They concluded that PFE yielded results comparable to Soxhlet extraction for all of the materials and analytes investigated and greater recoveries than Soxhlet extraction for the higher molecular weight PAHs from one of the SRMs (SRM 1650, Diesel Particulate Matter). This relatively new extraction technique has probably reduced efforts in developing SFE methods to replace existing Soxhlet extraction methods since PFE uses the same liquid solvents as Soxhlet and its extraction efficiencies have, thus far, been shown to be independent of the sample matrix – unlike SFE.

4 THE FUTURE OF SUPERCRITICAL FLUID EXTRACTION OF ENVIRONMENTAL SAMPLES

During the last 15 years there has been much progress made in the area of sample extraction for environmental analysis, driven mainly by concerns of extraction times and waste solvent. Initially, SFE was regarded as the panacea for sample extraction, but like all emerging techniques, its limitations in extraction efficiencies of specific chemical classes and sample types (matrices) began to emerge. Employing an SFE method as a replacement for an existing method must be done only after comparison with a traditional extraction method. After implementation, the SFE method should still be verified routinely by comparison with a solvent-based extraction technique using a standard material of known composition and homogeneity (e.g. CRMs), since subtle differences in the sample matrix can have detrimental effects on SFE recoveries.

Recent work describing the use of sub- and supercritical water as an extraction fluid[14–16] have excited the SFE, environmental analysis, and environmental mitigation communities. By taking advantage of the significant changes in the dielectric constant of water with respect to temperature, this solvent is able to extract a wide range of chemical classes, and could conceivably be used on a large scale to clean up chemical waste sites contaminated with both organic and inorganic pollutants. On-line SFE applications[110,111] have been shown to be some of the most sensitive methods for qualitative/quantitative analysis of very small samples (less than or equal to micrograms of sample material), and have not received the attention that they deserve from the environmental analytical community. Although on-line SFE would be subject to the same matrix effects as off-line methods, in sample-limited environmental applications the on-line technique may be the only way of obtaining chemical information (qualitative and/or quantitative) from a small sample.

The future of SFE in environmental analysis really depends on the willingness of researchers in the field to continue to investigate applications where SFE shows a clear benefit over conventional (Soxhlet and ultrasonic extraction) and new (PFE) liquid solvent-based methods. This future is not clear and may also depend on the amount of pressure regulatory agencies place on the environmental analytical community to further reduce solvent use in sample extractions.

DISCLAIMER

Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the NIST nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

ABBREVIATIONS AND ACRONYMS

CRM	Certified Reference Material
EIA	Enzyme Immunoassay Analysis
GC	Gas Chromatography
IR	Infrared
LVE	Liquid-vortex Extraction
NIST	National Institute of Standards and Technology
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PFE	Pressurized Fluid Extraction
SF	Supercritical Fluid
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SRM	Standard Reference Material
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Supercritical Fluid Chromatography in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Soil Sampling for the Characterization of Hazardous Waste Sites • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Inorganics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Waste Extraction Procedures

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food • Sample Preparation Analytical Techniques for Food

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Pesticides (New Generation) and Related Compounds, Analysis of

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

REFERENCES

1. S.B. Hawthorne, 'Analytical-scale Supercritical Fluid Extraction', *Anal. Chem.*, **62**, 633A–642A (1990).
2. V. Janda, K.D. Bartle, A.A. Clifford, 'Supercritical Fluid Extraction in Environmental Analysis', *J. Chromatogr.*, **642**, 283–299 (1993).
3. S. Bøwadt, S.B. Hawthorne, 'Supercritical Fluid Extraction in Environmental Analysis', *J. Chromatogr. A*, **703**, 549–571 (1995).
4. S.B. Hawthorne, D.J. Miller, M.D. Burford, J.J. Langenfeld, T. Eckert, P.K. Louie, 'Factors Controlling Quantitative Supercritical Fluid Extraction of Environmental Samples', *J. Chromatogr.*, **642**, 301–317 (1993).
5. J.W. King, 'Fundamentals and Applications of Supercritical Fluid Extraction in Chromatographic Science', *J. Chromatogr. Sci.*, **27**, 355–364 (1989).
6. L.T. Taylor, 'Strategies for Analytical SFE', *Anal. Chem.*, **67**, 364A–370A (1995).
7. D.E. Raynie, 'Warning Concerning the Use of Nitrous Oxide in Supercritical Fluid Extractions', *Anal. Chem.*, **65**, 3127–3128 (1993).
8. T.J. Bruno, *Handbook for the Analysis and Identification of Alternative Refrigerants*, CRC Press, Ann Arbor, 1995.
9. S.B. Hawthorne, D.J. Miller, 'Direct Comparison of Soxhlet and Low- and High-temperature Supercritical Carbon Dioxide Extraction Efficiencies of Organics from Environmental Solids', *Anal. Chem.*, **66**, 4005–4012 (1994).
10. D.J. Miller, S.B. Hawthorne, 'Determination of Solubilities of Organic Solutes in Supercritical Carbon Dioxide by On-line Flame Ionization Detection', *Anal. Chem.*, **67**, 273–279 (1995).
11. S. Kueppers, 'Use of Temperature Variation in Supercritical Fluid Extraction of Polymers for the Selective Extraction of Low Molecular Weight Components from Poly(ethylene terephthalate)', *Chromatographia*, **33**, 434–440 (1992).
12. S.B. Hawthorne, D.J. Miller, 'Extraction and Recovery of Organic Pollutants from Environmental Solids and Tenax-GC Using Supercritical Carbon Dioxide', *J. Chromatogr. Sci.*, **24**, 258–264 (1986).
13. J.M. Levy, R.A. Cavalier, T.N. Bosch, A.F. Rynaski, W.E. Huhak, 'Multi-dimensional Supercritical Fluid Chromatography and Supercritical Fluid Extraction', *J. Chromatogr. Sci.*, **27**, 341–346 (1989).
14. S.B. Hawthorne, Y. Yang, D.J. Miller, 'Extraction of Organic Pollutants from Environmental Solids with Sub- and Supercritical Water', *Anal. Chem.*, **66**, 2912–2920 (1994).
15. Y. Yang, S. Bøwadt, S.B. Hawthorne, D.J. Miller, 'Subcritical Water Extraction of Polychlorinated Biphenyls

from Soil and Sediment', *Anal. Chem.*, **67**, 4571–4576 (1995).
16. Y. Yang, S.B. Hawthorne, D.J. Miller, 'Class Selective Extraction of Polar, Moderately Polar, and Nonpolar Organics from Hydrocarbon Wastes Using Subcritical Water', *Environ. Sci. Technol.*, **31**, 430–437 (1997).
17. J. Pawliszyn, 'Kinetic Model of Supercritical Fluid Extraction', *J. Chromatogr. Sci.*, **31**, 31–37 (1993).
18. K.D. Bartle, A.A. Clifford, S.B. Hawthorne, J.J. Langenfeld, D.J. Miller, R. Robinson, 'A Model for Dynamic Extraction Using a Supercritical Fluid', *J. Supercrit. Fluids*, **3**, 143–149 (1990).
19. A.A. Clifford, M.D. Burford, S.B. Hawthorne, J.J. Langenfeld, D.J. Miller, 'Effect of the Matrix on the Kinetics of Dynamic Supercritical Fluid Extraction', *J. Chem. Soc., Faraday Trans.*, **91**, 1333–1338 (1995).
20. B.A. Benner, Jr, 'Summarizing the Effectiveness of Supercritical Fluid Extraction of Polycyclic Aromatic Hydrocarbons from Natural Matrix Environmental Samples', *Anal. Chem.*, **70**, 4594–4601 (1998).
21. L.J. Mulcahey, J.L. Hedrick, L.T. Taylor, 'Collection Efficiency of Various Solid-phase Traps for Offline Supercritical Fluid Extraction', *Anal. Chem.*, **63**, 2225–2232 (1991).
22. P.G. Thompson, L.T. Taylor, B.E. Richter, N.L. Porter, J.L. Ezzell, 'Trapping Efficiencies of Various Collection Solvents after Supercritical Fluid Extraction', *J. High Resolut. Chromatogr.*, **16**, 713–716 (1983).
23. D.J. Miller, S.B. Hawthorne, M.E.P. McNally, 'Solventless Collection of Analytes by Rapid Depressurization after Static Supercritical Fluid Extraction', *Anal. Chem.*, **65**, 1038–1042 (1993).
24. M.D. Burford, S.B. Hawthorne, D.J. Miller, 'Analysis of Volatile Organics by Supercritical Fluid Extraction Coupled to Gas Chromatography. II. Quantitation of Petroleum Hydrocarbons from Environmental Sample', *J. Chromatogr. A*, **685**, 95–111 (1994).
25. R.W. Current, D.C. Tilotta, 'Determination of Total Petroleum Hydrocarbons in Soil by On-line Supercritical Fluid Extraction/Infrared Spectroscopy Using a Fiber-optic Transmission Cell and a Simple Filter Spectrometer', *J. Chromatogr. A*, **785**, 269–277 (1997).
26. S.E. Eckert-Tilotta, S.B. Hawthorne, D.J. Miller, 'Supercritical Fluid Extraction with Carbon Dioxide for the Determination of Total Petroleum Hydrocarbons in Soil', *Fuel*, **72**, 1015–1023 (1993).
27. W. Li, I.M. Lazar, Y.J. Wan, S.J. Butala, Y. Shen, A. Malik, M.L. Lee, 'Determination of Volatile Hydrocarbons in Coals and Shales Using Supercritical Fluid Extraction and Chromatography', *Energy Fuels*, **11**, 945–950 (1997).
28. V. Lopez-Avila, J. Benedicto, N.S. Dodhiwala, R. Young, W.F. Beckert, 'Development of an Offline SFE [Supercritical Fluid Extraction] – IR Method for Petroleum Hydrocarbons in Soils', *J. Chromatogr. Sci.*, **30**, 335–343 (1992).
29. V. Lopez-Avila, R. Young, R. Kim, W.F. Beckert, 'Interlaboratory Evaluation of an Offline Supercritical Fluid Extraction Infrared Spectrometric Method for Determination of Petroleum Hydrocarbons in Solid Matrices', *J. AOAC Int.*, **76**, 555–564 (1993).
30. B. Liu, B. Lockwood, L.A. Gifford, 'Supercritical Fluid Extraction of Diosgenin from Tubers of *Dioscorea Nipponica*', *J. Chromatogr. A*, **690**, 250–253 (1995).
31. M. Ashraf-Khorassani, M.T. Combs, L.T. Taylor, 'Effect of Moisture on Supercritical Fluid Extraction of Polynuclear Aromatic Hydrocarbons and Phenols from Soil Using an Automated Extractor', *J. High Resolut. Chromatogr.*, **18**, 709–712 (1995).
32. I.J. Barnabas, J.R. Dean, W.R. Tomlinson, S.P. Owen, 'Experimental Design Approach for the Extraction of Polycyclic Aromatic Hydrocarbons from Soil Using Supercritical Carbon Dioxide', *Anal. Chem.*, **67**, 2064–2069 (1995).
33. S. Bøwadt, L. Mazeas, D.J. Miller, S.B. Hawthorne, 'Field-portable Determination of Polychlorinated Biphenyls and Polynuclear Aromatic Hydrocarbons in Soil Using Supercritical Fluid Extraction', *J. Chromatogr. A*, **785**, 205–217 (1997).
34. M.D. Burford, S.B. Hawthorne, D.J. Miller, 'Extraction Rates of Spiked Versus Native PAHs from Heterogeneous Environmental Samples Using Supercritical Fluid Extraction and Sonication in Methylene Chloride', *Anal. Chem.*, **65**, 1497–1505 (1993).
35. J. Dankers, M. Groenenboom, L.H.A. Scholtis, C. Van der Heiden, 'High-speed Supercritical Fluid Extraction Method for Routine Measurement of Polycyclic Aromatic Hydrocarbons in Environmental Soils with Dichloromethane as a Static Modifier', *J. Chromatogr.*, **641**, 357–362 (1993).
36. J.R. Dean, I.J. Barnabas, I.A. Fowlis, 'Extraction of Polycyclic Aromatic Hydrocarbons from Highly Contaminated Soils: a Comparison Between Soxhlet, Microwave and Supercritical Fluid Extraction Techniques', *Anal. Proc.*, **32**, 305–308 (1995).
37. I. Fernandez, J. Dachs, J.M. Bayona, 'Application of Experimental Design Approach to the Optimization of Supercritical Fluid Extraction of Polychlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons', *J. Chromatogr.*, **719**, 77–85 (1996).
38. C. Friedrich, K. Cammann, W. Kleiboehmer, 'Matrix Independent Supercritical Fluid Extraction of Polycyclic Aromatic Hydrocarbons by Employing Binary Modifiers', *Fresenius' J. Anal. Chem.*, **352**, 730–734 (1995).
39. K.G. Furton, C.W. Huang, R. Jaffe, M.A. Sicre, 'High-temperature Supercritical Fluid Extraction of Hydrocarbons from Geological Samples and Comparison to Soxhlet Extraction', *J. High Resolut. Chromatogr.*, **17**, 679–681 (1994).
40. S.B. Hawthorne, D.J. Miller, 'Extraction and Recovery of Polycyclic Aromatic Hydrocarbons from

Environmental Solids Using Supercritical Fluids', *Anal. Chem.*, **59**, 1705–1708 (1987).
41. S.B. Hawthorne, J.J. Langenfeld, D.J. Miller, M.D. Burford, 'Comparison of Supercritical Chlorodifluoromethane, Nitric Oxide and Carbon Dioxide for the Extraction of Polychlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **64**, 1614–1622 (1992).
42. F.M. Lancas, M.H.R. Matta, L.J. Hayasida, E. Carrilho, 'Supercritical Fluid Extraction of Polynuclear Aromatic Hydrocarbons from Coal with Offline CGC [Capillary GC] – MS Analysis', *J. High Resolut. Chromatogr.*, **1991**, 633–635 (1991).
43. J.J. Langenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Effects of Temperature and Pressure on Supercritical Fluid Extraction Efficiencies of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls', *Anal. Chem.*, **65**, 338–344 (1993).
44. J.J. Langenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Role of Modifiers for Analytical-scale Supercritical Fluid Extraction of Environmental Samples', *Anal. Chem.*, **66**, 909–916 (1994).
45. J.J. Langenfeld, S.B. Hawthorne, D.J. Miller, J. Pawliszyn, 'Kinetic Study of Supercritical Fluid Extraction of Organic Contaminants from Heterogeneous Environmental Samples with Carbon Dioxide and Elevated Temperatures', *Anal. Chem.*, **67**, 1727–1736 (1995).
46. H.B. Lee, T.E. Peart, Y. Hong, D.R. Gere, 'Supercritical Carbon Dioxide Extraction of Polycyclic Aromatic Hydrocarbons from Sediments', *J. Chromatogr. A*, **653**, 83–91 (1993).
47. J.M. Levy, L.A. Dolata, R.M. Ravey, 'Considerations of SFE for GC–MS Determination of Polynuclear Aromatic Hydrocarbons in Soils and Sediments', *J. Chromatogr. Sci.*, **31**, 349–352 (1993).
48. V. Lopez-Avila, R. Young, J. Tehrani, J. Damian, S. Hawthorne, J. Dankers, C. Van der Heiden, 'Mini-round-robin Study of a Supercritical Fluid Extraction Method for Polynuclear Aromatic Hydrocarbons in Soils with Dichloromethane as a Static Modifier', *J. Chromatogr. A*, **672**, 167–175 (1994).
49. A. Meyer, W. Kleiboehmer, K. Cammann, 'SFE of PAH from Soils with a High Carbon Content and Analyte Collection via Combined Liquid–Solid Trapping', *J. High Resolut. Chromatogr.*, **16**, 491–494 (1993).
50. A. Meyer, W. Kleiboehmer, 'Supercritical Fluid Extraction of Polycyclic Aromatic Hydrocarbons from a Marine Sediment and Analyte Collection via Liquid–Solid Trapping', *J. Chromatogr. A*, **657**, 327–335 (1993).
51. T. Paschke, S.B. Hawthorne, D.J. Miller, B. Wenclawiak, 'Supercritical Fluid Extraction of Nitrated Polycyclic Aromatic Hydrocarbons and Polycyclic Aromatic Hydrocarbons from Diesel-exhaust Particulate Matter', *J. Chromatogr.*, **609**, 333–340 (1992).
52. G. Reimer, A. Suarez, 'Comparison of Supercritical Fluid Extraction and Soxhlet Extraction for the Analysis of Native Polycyclic Aromatic Hydrocarbons in Soils', *J. Chromatogr. A*, **699**, 253–263 (1995).
53. B. Wenclawiak, C. Rathmann, A. Teuber, 'Supercritical Fluid Extraction of Soil Samples and Determination of Polycyclic Aromatic Hydrocarbons (PAHs) by HPLC', *Fresenius' J. Anal. Chem.*, **344**, 497–500 (1992).
54. B.W. Wenclawiak, T. Paschke, M. Krappe, 'Supercritical Fluid Extraction of Polycyclic Aromatic Hydrocarbons (PAH) from Native Soil and Fly Ash with Pure and Modified Carbon Dioxide and Dimethylether', *Fresenius' J. Anal. Chem.*, **357**, 1128–1132 (1997).
55. Y. Yang, A. Gharaibeh, S.B. Hawthorne, D.J. Miller, 'Combined Temperature-modifier Effects on Supercritical Carbon Dioxide Extraction Efficiencies of Polycyclic Aromatic Hydrocarbons from Environmental Samples', *Anal. Chem.*, **67**, 641–646 (1995).
56. E.G. Alley, G.Z. Lu, 'Analysis of Polychlorinated Biphenyls in Fatty Biological Matrices by Online Supercritical Fluid Extraction and Supercritical Fluid Cleanup', *J. AOAC Int.*, **78**, 1051–1054 (1995).
57. N. Alexandrou, Z. Miao, M. Colquhoun, J. Pawliszyn, C. Jennison, 'Supercritical Fluid Extraction and Cleanup with Capillary GC – Ion-trap Mass Spectrometry for Determination of Polychlorinated Dibenzo-*p*-dioxins and Dibenzofurans in Environmental Samples', *J. Chromatogr. Sci.*, **30**, 351–357 (1992).
58. M. Ashraf-Khorassani, L.T. Taylor, 'Comparison of Modifier Addition to the Matrix Versus Modifier Addition to the Fluid in the Supercritical Fluid Extraction of PCBs from River Sediment', *Am. Lab.*, **27**, 23–28 (1995).
59. S. Atuma, S. Zettermark, L. Hansson, 'Determination of Organochlorine Pesticides and PCBs in Fish Using Offline Supercritical Fluid Extraction, HPLC Separation, and Gas Chromatography', *Organohalogen Compd.*, **23**, 31–34 (1995).
60. J.L. Bernal, J.J. Jimenez, J.M. Rivera, 'Interferences Found in the Determination of Pesticide and PCB Residues by Capillary GC after SFE', *J. Chromatogr. Sci.*, **33**, 500–504 (1995).
61. S. Bøwadt, B. Johansson, P. Fruekilde, M. Hansen, D. Zilli, B. Larsen, J. de Boer, 'Supercritical Fluid Extraction of Polychlorinated Biphenyls from Lyophilized Fish Tissue', *J. Chromatogr. A*, **675**, 189–204 (1994).
62. S. Bøwadt, B. Johansson, F. Pelusio, B.R. Larsen, C. Rovida, 'Solid-phase Trapping of Polychlorinated Biphenyls in Supercritical Fluid Extraction', *J. Chromatogr. A*, **662**, 424–433 (1994).
63. S. Bøwadt, B. Johansson, 'Analysis of PCBs in Sulfur-containing Sediments by Offline Supercritical Fluid Extraction and HR [High-resolution] GC–ECD', *Anal. Chem.*, **66**, 667–673 (1994).
64. S. Bøwadt, B. Johansson, S. Wunderli, M. Zennegg, L.F. de Alencastro, D. Grandjean, 'Independent Comparison of Soxhlet and Supercritical Fluid Extraction for the Determination of PCBs in an Industrial Soil', *Anal. Chem.*, **67**, 2424–2430 (1995).

65. R. Fuoco, P.R. Griffiths, 'Analysis of Polychlorobiphenyls (PCBs) in Contaminated Soil Samples by Supercritical Fluid Extraction – Supercritical Fluid Chromatography Coupled with FTIR Spectrometry', *Ann. Chim.*, **82**, 235–245 (1992).
66. R.C. Hale, M.O. Gaylor, 'Determination of PCBs in Fish Tissues Using Supercritical Fluid Extraction', *Environ. Sci. Technol.*, **29**, 1043–1047 (1995).
67. J.S. Ho, P.H. Tang, 'Optimization of Supercritical Fluid Extraction of Environmental Pollutants from a Liquid–Solid Extraction Cartridge', *J. Chromatogr. Sci.*, **30**, 344–350 (1992).
68. H.R. Johansen, G. Becher, T. Greibrokk, 'Determination of PCBs in Biological Samples Using Online SFE [Supercritical Fluid Extraction] – GC', *Fresenius' J. Anal. Chem.*, **344**, 486–491 (1992).
69. H.B. Lee, T.E. Peart, A.J. Nimi, C.R. Knipe, 'Rapid Supercritical Carbon Dioxide Extraction Method for Determination of Polychlorinated Biphenyls in Fish', *J. AOAC Int.*, **78**, 437–444 (1995).
70. A.G. Mills, T.M. Jefferies, 'Rapid Isolation of Polychlorinated Biphenyls from Milk by a Combination of Supercritical Fluid Extraction and Supercritical Fluid Chromatography', *J. Chromatogr.*, **643**, 409–418 (1993).
71. D. Sterzenbach, B.W. Wenclawiak, V. Weigelt, 'Determination of Chlorinated Hydrocarbons in Marine Sediments at the Part-per-trillion Level with Supercritical Fluid Extraction', *Anal. Chem.*, **69**, 831–836 (1997).
72. B. van Bavel, M. Jaremo, L. Karlsson, G. Lindstrom, 'Development of a Solid-phase Carbon Trap for Simultaneous Determination of PCDD, PCDF, PCB and Pesticides in Environmental Samples Using SFE-LC', *Anal. Chem.*, **68**, 1279–1283 (1996).
73. M.M. Schantz, S. Bøwadt, B.A. Benner, Jr, S.A. Wise, S.B. Hawthorne, 'Comparison of Supercritical Fluid Extraction and Soxhlet Extraction for the Determination of Polychlorinated Biphenyls in Environmental Matrix Standard Reference Materials', *J. Chromatogr. A*, **816**, 213–220 (1998).
74. I.J. Barnabas, J.R. Dean, S.M. Hitchen, S.P. Owen, 'Selective Supercritical Fluid Extraction of Organochlorine Pesticides and Herbicides from Aqueous Samples', *J. Chromatogr. Sci.*, **32**, 547–551 (1994).
75. H. Casabianca, 'Use of Supercritical Fluid Extraction – GC – Atomic-emission Detection for the Analysis of Halogen and Sulfur Pesticides', *Analusis*, **20**, S9–S11 (1992).
76. J.R. Dean, I.J. Barnabas, S.P. Owen, 'Influence of Pesticide–Soil Interactions on the Recovery of Pesticides Using Supercritical Fluid Extraction', *Analyst*, **121**, 465–468 (1996).
77. J.L. Ezzell, E.R. Bruce, 'Supercritical Fluid Extraction of Pesticides and Phthalate Esters Following Solid Phase Extraction from Water', *J. Microcolumn Sep.*, **4**, 319–323 (1992).
78. K.G. Furton, J. Rein, 'Trends in Techniques for the Extraction of Drugs and Pesticides from Biological Specimens Prior to Chromatographic Separation and Detection', *Anal. Chim. Acta*, **236**, 99–114 (1990).
79. M.L. Hopper, J.W. King, 'Enhanced Supercritical Fluid Carbon Dioxide Extraction of Pesticides from Foods Using Pelletized Diatomaceous Earth', *J. Assoc. Off. Anal. Chem.*, **74**, 661–666 (1991).
80. A. Izquierdo, M.T. Tena, d.C. Luque, M. Valcarcel, 'Supercritical Fluid Extraction of Carbamate Pesticides from Soils and Cereals', *Chromatographia*, **42**, 206–212 (1996).
81. A. Koinecke, R. Kreuzig, M. Bahadir, 'Effects of Modifiers, Adsorbents and Eluents in Supercritical Fluid Extraction of Selected Pesticides in Soil', *J. Chromatogr. A*, **786**, 155–161 (1997).
82. S.J. Lehotay, M.A. Ibrahim, 'Supercritical Fluid Extraction and Gas Chromatography – Ion-trap Mass Spectrometry of Pentachloronitrobenzene Pesticides in Vegetables', *J. Assoc. Off. Anal. Chem.*, **78**, 445–452 (1995).
83. S.J. Lehotay, N. Aharonson, E. Pfeil, M.A. Ibrahim, 'Development of a Sample Preparation Technique for Supercritical Fluid Extraction for Multiresidue Analysis of Pesticides in Produce', *J. AOAC Int.*, **78**, 831–840 (1995).
84. S.J. Lehotay, A. Valverde-García, 'Evaluation of Different Solid-phase Traps for Automated Collection and Clean-up in the Analysis of Multiple Pesticides in Fruits and Vegetables after Supercritical Fluid Extraction', *J. Chromatogr. A*, **765**, 69–84 (1997).
85. V. Lopez-Avila, N.S. Dodhiwala, W.F. Beckert, 'Supercritical Fluid Extraction and its Application to Environmental Analysis', *J. Chromatogr. Sci.*, **28**, 468–476 (1990).
86. B. Murugaverl, A. Gharaibeh, K.J. Voorhees, 'Mixed Adsorbent for Clean Up During Supercritical Fluid Extraction of Three Carbamate Pesticides in Tissues', *J. Chromatogr. A*, **657**, 223–226 (1993).
87. Z.V. Skopec, R. Clark, P.M.A. Harvey, R.J. Wells, 'Analysis of Organophosphorus Pesticides in Rice by Supercritical Fluid Extraction and Quantitation Using an Atomic-emission Detector', *J. Chromatogr. Sci.*, **31**, 445–449 (1993).
88. P. Tong, T. Imagawa, 'Optimization of Supercritical Fluid Extraction for Polychlorinated Biphenyls from Sediments', *Anal. Chim. Acta*, **310**, 93–100 (1995).
89. M. Zhou, R.K. Trubey, Z.O. Keil, D.L. Sparks, 'Study of the Effects of Environmental Variables and Supercritical Fluid Extraction Parameters on the Extractability of Pesticide Residues from Soils Using a Multivariate Optimization Scheme', *Environ. Sci. Technol.*, **31**, 1934–1939 (1997).
90. O. Berdeaux, L.F. de Alencastro, D. Grandjean, J. Tarradellas, 'Supercritical Fluid Extraction of Sulfonylurea Herbicides in Soil Samples', *Int. J. Environ. Anal. Chem.*, **56**, 109–117 (1994).

91. V. Janda, G. Steenbeke, P. Sandra, 'Supercritical Fluid Extraction of *s*-Triazine Herbicides from Sediment', *J. Chromatogr.*, **479**, 200–205 (1989).
92. V. Lopez-Avila, N.S. Dodhiwala, W.F. Beckert, 'Developments in the Supercritical Fluid Extraction of Chlorophenoxy Acid Herbicides from Soil Samples', *J. Agric. Food Chem.*, **41**, 2038–2044 (1993).
93. M.E.P. McNally, J.R. Wheeler, 'Supercritical Fluid Extraction Coupled with Supercritical Fluid Chromatography for the Separation of Sulfonylurea Herbicides and their Metabolites from Complex Matrices', *J. Chromatogr.*, **435**, 63–71 (1988).
94. S. Papilloud, W. Haerdi, 'Supercritical Fluid Extraction of Triazine Herbicides from Solid Matrices', *Chromatographia*, **38**, 514–519 (1994).
95. A.M. Robertson, J.N. Lester, 'Supercritical Fluid Extraction of *s*-Triazines and Phenylurea Herbicides from Sediment', *Environ. Sci. Technol.*, **28**, 346–351 (1994).
96. G.K. Stearman, M.J.M. Wells, S.M. Adkisson, T.E. Ridgill, 'Supercritical Fluid Extraction Coupled with Enzyme Immunoassay Analysis of Soil Herbicides', *Analyst*, **120**, 2617–2621 (1995).
97. T. Yarita, Y. Horimoto, A. Nomura, S. Gonda, 'Supercritical Fluid Extraction of Triazine Herbicides from Soil Matrices', *Chromatographia*, **42**, 551–554 (1996).
98. M.M. Schantz, S.N. Chesler, 'Supercritical Fluid Extraction Procedure for the Removal of Trace Organic Species from Solid Samples', *J. Chromatogr.*, **363**, 397–401 (1986).
99. S.B. Hawthorne, D.J. Miller, J.J. Langenfeld, 'Quantitative Analysis Using Directly Coupled Supercritical Fluid Extraction – Capillary Gas Chromatography (SFE – GC) with a Conventional Split – Splitless Injection Port', *J. Chromatogr. Sci.*, **28**, 2–8 (1990).
100. S.B. Hawthorne, D.J. Miller, 'Directly Coupled Supercritical Fluid Extraction – Gas-chromatographic Analysis of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls from Environmental Solids', *J. Chromatogr.*, **403**, 63–76 (1987).
101. S.M. Pyle, M.M. Setty, 'Supercritical Fluid Extraction of High Sulphur Soils, with Use of a Copper Scavenger', *Talanta*, **38**, 1125–1128 (1991).
102. N.L. Porter, A.F. Rynaski, E.R. Campbell, M. Saunders, B.E. Richter, J.T. Swanson, R.B. Nielsen, B.J. Murphy, 'Studies of Linear Restrictors and Analyte Collection via Solvent Trapping after Supercritical Fluid Extraction', *J. Chromatogr. Sci.*, **30**, 367–373 (1992).
103. 'UNEP, 1998, Montreal Protocol: Report of the Refrigeration, Air Conditioning and Heat Pumps Technical Options Committee', United Nations Environment Programme, Nairobi, 1998.
104. K.E. Laintz, C.M. Wai, C.R. Yonker, R.D. Smith, 'Extraction of Metal Ions from Liquid and Solid Materials by Supercritical Carbon Dioxide', *Anal. Chem.*, **64**, 2875–2878 (1992).
105. C.M. Wai, Y. Lin, R. Brauer, S. Wang, W.F. Beckert, 'Supercritical Fluid Extraction of Organic and Inorganic Mercury from Solid Materials', *Talanta*, **40**, 1325–1330 (1993).
106. Y.H. Lin, R.D. Brauer, K.E. Laintz, C.M. Wai, 'Supercritical Fluid Extraction of Lanthanides and Actinides from Solid Materials with a Fluorinated Beta-diketone', *Anal. Chem.*, **65**, 2549–2551 (1993).
107. Y.H. Lin, C.M. Wai, F.M. Jean, R.D. Brauer, 'Supercritical Fluid Extraction of Thorium and Uranium Ions from Solid and Liquid Materials with Fluorinated Beta-diketones and Tributyl Phosphate', *Environ. Sci. Technol.*, **28**, 1190–1193 (1994).
108. B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: a Technique for Sample Preparation', *Anal. Chem.*, **68**, 1033–1039 (1996).
109. M.M. Schantz, J.J. Nichols, S.A. Wise, 'Evaluation of Pressurized Fluid Extraction for the Extraction of Environmental Matrix Reference Materials', *Anal. Chem.*, **69**, 4210–4219 (1997).
110. S.B. Hawthorne, M.S. Krieger, D.J. Miller, 'Analysis of Flavor and Fragrance Compounds Using Supercritical Fluid Extraction Coupled with Gas Chromatography', *Anal. Chem.*, **60**, 472–477 (1988).
111. S.B. Hawthorne, D.J. Miller, M.S. Krieger, 'Coupled SFE – GC: a Rapid and Simple Technique for Extracting, Identifying and Quantitating Organic Analytes from Solids and Sorbent Resins', *J. Chromatogr. Sci.*, **27**, 347–354 (1989).

Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Matthew M. Booth and David H. Powell
University of Florida, Gainesville, USA

1 Introduction **1**
2 History of Gas Chromatography/Quadrupole Mass Spectrometry **2**
2.1 Invention of the Quadrupole Mass Spectrometer 2
2.2 Interfacing Gas Chromatographs to Quadrupole Mass Spectrometers 2
2.3 Types of Quadrupole Mass Spectrometers Used for Gas Chromatography/Mass Spectrometry 3
3 Ionization Methods **6**
3.1 Electron Ionization 6
3.2 Chemical Ionization 7
4 Sample Preparation and Analysis **7**
4.1 Sample Preparation for Volatile Organic Analysis 8
4.2 Sample Preparation for Semi-volatile Organic Analysis 9
4.3 Sample Extract Clean-up 11
4.4 Chemical Derivatization 12
5 Instrumental Analysis Considerations **12**
6 Summary **13**
Abbreviations and Acronyms **13**
Related Articles **13**
References **14**

Gas chromatography (GC) with quadrupole mass spectrometry (MS) is one of the most powerful and widely used analytical techniques. This technology combines the high resolving power of GC with the excellent sensitivity and chemical-structure-informing power of quadrupole MS. The combination of these two techniques provides both the selectivity and sensitivity needed for qualitative and quantitative analysis of organic analytes at trace levels in a wide variety of matrices.

This article provides an overview of the development and combination of these two technologies. Types of quadrupole mass spectrometers that are commonly coupled to gas chromatographs are discussed. Both electron ionization (EI) and chemical ionization (CI) are reviewed. Sample preparation methods are discussed for air, water, and solids. The different approaches employed for highly volatile and semi-volatile substances are presented. The use of chemical derivatives to improve sample response and various instrumental analysis considerations are also covered.

1 INTRODUCTION

When considering analytical informing power, cost-effectiveness, and extent of applications, gas chromatography (GC) with quadrupole mass spectrometry (MS) detection is arguably the most powerful analytical technique available today. It is one of the most commonly used techniques for the trace analysis of organic compounds in water. This article will give an overview of this technology and its applications to the analysis of trace organic compounds in water, soil, and air.

The combination of gas chromatography/mass spectrometry (GC/MS) provides the advantages of the high separation power of GC with chromatographic resolving powers of several hundred thousand theoretical plates per column, and the incredible sensitivity and qualitative selectivity of MS. GC is frequently coupled with selective detectors such as flame ionization, electron capture, flame photometry, etc. These combinations provide only the retention time and the selectivity of the detector as criteria for compound identification and quantitation. The use of a mass spectrometer as the GC detector provides the equivalent of hundreds of detectors in parallel – one at each mass-to-charge ratio measured. This additional dimension of mass-to-charge ratio, in conjunction with retention time, provides an even further increase in selectivity by the connectivity relationships between fragment ions from the same molecular species. Thus, GC/MS provides a selectivity of qualitative identification that is available from few analytical technologies. Further enhancements to selectivity and sensitivity also can be obtained by the use of different ionization techniques, e.g. chemical ionization (CI) or electron ionization (EI).

The importance of the mass spectrometer as a detector for GC can be gleaned from the biannual review of GC published in *Analytical Chemistry*. As observed by Arpino,[1] a 1996 review of GC cited 1300 papers on GC, where 58% employed MS as the detector.[2] In a more recent review in 1998 the percentage using MS detection was 26%.[3] Since GC has been described as ''the world's most widely used analytical technique'', the number of analyses per year performed by GC/MS is truly large[4] (possibly several hundred thousand). The quadrupole

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

mass spectrometer and quadrupole ion trap mass spectrometer are the most common mass spectrometers used in conjunction with GC. According to the noted mass spectrometrist, Professor Fred W. McLafferty of Cornell University, ''Certainly, the number of quadrupoles sold and in use far exceeds the total of all other types of mass spectrometers''. This article will review the history of the development of quadrupole mass spectrometers interfaced to gas chromatographs, and the sample preparation and application of quadrupole GC/MS for the analysis of trace organic compounds in water and other matrices.

2 HISTORY OF GAS CHROMATOGRAPHY/ QUADRUPOLE MASS SPECTROMETRY

2.1 Invention of the Quadrupole Mass Spectrometer

The quadrupole mass spectrometer was conceived and executed first by Paul and Steinwedel[5] in the early 1950s at the University of Bonn. Paul's first efforts involved focusing an ion beam using a hexapolar magnetic field. In a subsequent collaboration with Steinwedel, they used a combination of a four-pole electrical field with a superimposed radiofrequency (RF) field to separate ions with different mass-to-charge ratios. This instrument was developed later into today's linear quadrupole mass spectrometer and was patented in 1956.[6] Paul shared the 1989 Nobel Prize in Physics with Dehmelt for his pioneering work on the development of quadrupole mass spectrometers.

The first commercial quadrupole mass spectrometer was built as a residual gas analyzer by Electronic Associates, Inc. in 1964. It was not until 1967 that Finnigan and others established Finnigan Instruments Corporation with the express purpose of constructing the first commercial quadrupole-based GC/MS instrument (Model 1015). The first instrument was delivered in 1968. Approximately 1 year later, Hewlett-Packard introduced its own quadrupole GC/MS instrument.

Paul and Dehmelt also developed the quadrupole ion trap in the 1950s. However, it remained something of a laboratory curiosity until the early 1980s when Stafford at FinniganMAT developed a scanning technique for the quadrupole ion trap, called the mass selective instability mode, which made the instrument much more analytically useful. In 1983, FinniganMAT introduced the first commercial gas chromatograph/quadrupole ion trap mass spectrometer. Subsequently, other manufacturers have produced quadrupole ion traps interfaced to gas chromatographs.

In 1976 the United States District Court mandated that the United States Environmental Protection Agency (USEPA) needed to analyze for 65 compounds in the nation's wastewater. This list of compounds was subsequently named the ''priority pollutants'' and expanded to 129 compounds that included 114 organics: 25 pesticides and polychlorinated biphenyls (PCBs) and 89 other organics. This action led to the establishment of GC/MS as the mandated method for analysis of the environmental samples under the Clean Water Act Series 600 methods and the Solid Waste Manual SW-846 methods. The USEPA methods for volatile and semi-volatile organics in water (Methods 624 and 625, respectively[7]) were promulgated in December 1979. The adoption by the USEPA of these standardized GC/MS methods using quadrupole mass spectrometers contributed greatly to a phenomenal growth in the use of quadrupole GC/MS. It is estimated that the worldwide market for mass spectrometers was \$20–30 million per year in 1968, which was dominated by magnetic sector mass spectrometers. By 1994, the MS market had grown to \$750 million per year, with quadrupoles accounting for more than 75% of the market.[8]

2.2 Interfacing Gas Chromatographs to Quadrupole Mass Spectrometers

During the early 1950s, gas–liquid chromatography was being developed by James and Martin.[9] Their first report on gas–liquid chromatography appeared in 1952, the same year that they received the Nobel Prize for the development of partition chromatography. The initial reports of a GC/MS interface were by Holmes and Morrel in 1957. The instrument that they used was a slow-scanning magnetic sector mass spectrometer, which severely limited its applicability. Gohlke and McLafferty at the Dow Chemical Company coupled a gas chromatograph to a very fast time-of-flight mass spectrometer in the mid-1950s. In their instrument, the effluent of the GC was split between the mass spectrometer and a thermal conductivity detector, and the flow to the mass spectrometer was controlled by an adjustable needle valve.[10]

Because the first GC systems used packed columns with flow rates of tens of milliliters per minute, interfacing to a mass spectrometer generally required the use of devices to remove some of the carrier gas and enrich the sample. Otherwise, the carrier gas load on the mass spectrometer source would be too high. Four different approaches to GC/MS interfaces were commonly used: effusion through a fritted tube; diffusion through permeable membranes; diffusion using a jet separator; and a modification of this – the open-split interface. The fritted-glass-tube interface was introduced by Watson and Bieman in 1964.[11] The carrier flow from the exit of the gas chromatograph entered the device through a glass constriction to reduce the pressure. The

flow was directed through a tube constructed from glass containing pores of approximately 1 μm in diameter. The outside of the tube was pumped with a mechanical rough pump. The carrier gas effused through the frit faster than the analytes. A constriction on the output to the mass spectrometer helped to control the gas pressure in the mass spectrometer. Enrichment factors (outlet concentration/inlet concentration) were about 50, with efficiencies of 25–40%.

Diffusion-based jet separators were introduced by Ryhage in 1964: he used two jet separators in series.[12] The principle of operation depends upon the difference in momentum between the carrier gas and the analytes eluting from the GC column. In the more modern jet separator pictured in Figure 1, there are two small-diameter tubes that face each other, separated by a gap of less than 1 mm. The region surrounding the tubes is connected to a vacuum pump. Typical carrier gases, such as helium or hydrogen, are much lighter than most analytes eluting from the column and readily diffuse into the vacuum chamber. The analytes eluting from the column, however, are usually much heavier with much higher momenta, and therefore do not diffuse as readily into the vacuum. The gas flow out of the jet separator was thereby much enriched in the heavier components, and the total gas flow was reduced enough to make the gas load manageable in the mass spectrometer source. Enrichment factors were about 100, with efficiencies up to 60%.

In 1966, Llewellyn and Littlejohn[13] introduced a membrane-based separator that depended on the permeation of the analyte through a membrane. The process depended on the solubility and diffusion rate of the analyte through the membrane that separated the outlet of the GC column from the inlet of the mass spectrometer. The membrane was an elastomer in which helium had a very low solubility. Enrichment factors as high as 10^5 were achieved, with efficiencies of 50%. However, one drawback of membrane interfaces is that they degrade the chromatographic resolution by adding to the extra-column band-broadening.

The use of capillary columns in GC enabled the direct coupling of the chromatographic column to the mass spectrometer source. The flow rates for capillary columns, usually of the order of 1 mL min^{-1}, results in much lower gas loads on the mass spectrometer source. This factor, coupled with more efficient differential pumping of mass spectrometer sources, has allowed direct interfacing of the GC column to the MS source in modern GC/MS instruments. This difference means more efficient transfer of eluting peaks from the gas chromatograph to the mass spectrometer and less degradation of chromatographic resolution. Jet separators and open-split interfaces continue to be used in some applications.

2.3 Types of Quadrupole Mass Spectrometers Used for Gas Chromatography/Mass Spectrometry

Three types and configurations of quadrupoles are commonly used for GC/MS. The most common is the simple linear quadrupole mass filter. A combination of three linear quadrupoles connected in series is used less frequently due to cost and complexity, but has the significant advantage of performing tandem MS. The third type that is rapidly growing in popularity is the quadrupole ion trap, which can perform tandem MS without the expense of the triple quadrupole. Each of these instruments is discussed below.

A simple, linear quadrupole mass filter is presented in Figure 2. There are a number of excellent, detailed

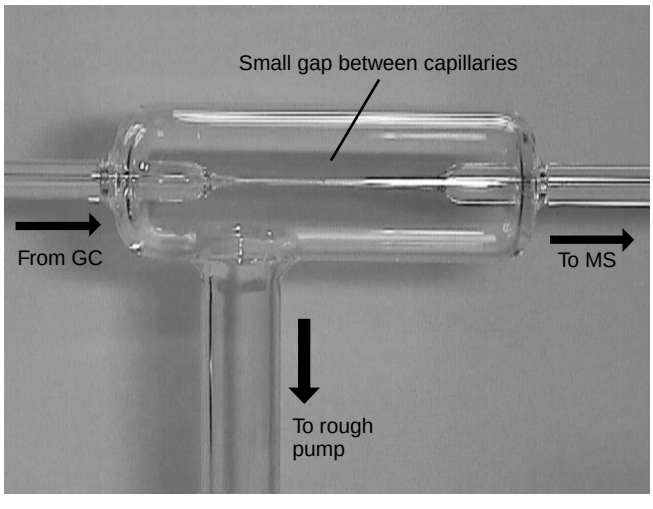

Figure 1 Jet separator interface between gas chromatograph and mass spectrometer.

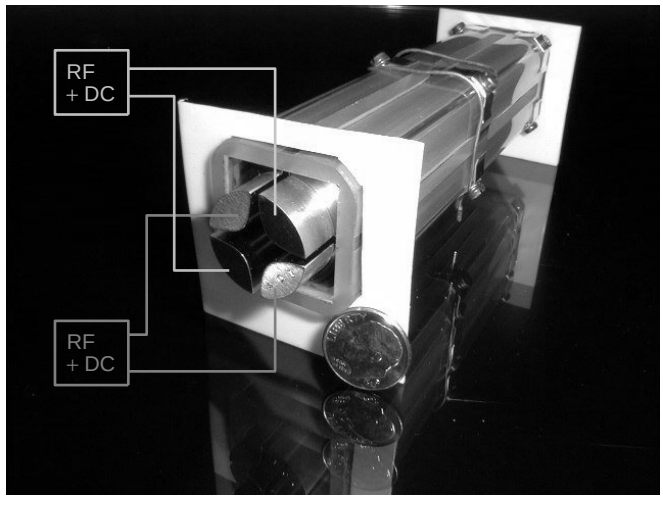

Figure 2 Linear quadrupole mass filter.

descriptions of the operation of quadrupoles,[8,14] and only a brief overview will be given here. The quadrupole mass spectrometer is actually a mass-to-charge filter. True mass spectrometers disperse a beam of ions according to their m/z (e.g. in a magnetic sector instrument the ion beams are separated in space and emerge at different angles from the magnet). Similarly, a time-of-flight mass spectrometer separates the ion beam into different packets of ions that emerge from the analyzer at different times. The quadrupole mass spectrometer only passes a band of ions of a selected m/z, and all the other ions are lost to the surfaces of the analyzer. This difference is due to how the quadrupole operates.

In the quadrupole filter, diagonally opposing pairs of rods are electrically connected. The two pairs of rods are connected to opposite direct current (DC) voltages of the same magnitude but opposite polarities, and their RF voltages are shifted 180° out-of-phase with respect to each other. An ion entering the quadrupole from one end will travel in a zig-zag course toward the opposite end in synchrony with the RF field. For a particular value of the DC and RF voltages, the solutions to the differential equations are such that only a particular m/z ion will travel on a stable trajectory and traverse to the opposite end of the filter where the detector is located. All other ions will undergo unstable oscillations and either collide with the rods or exit the gaps between the rods to collide with the surfaces of the vacuum chamber – these ions will not traverse to the end of the filter. When an ion collides with a surface, it will be discharged and become a neutral molecule that will eventually be pumped away by the vacuum system. The DC and RF voltages and the RF govern the m/z ratio of the ions that the quadrupole will pass; therefore, the device acts as an m/z filter. Changing the magnitudes of the RF and DC voltages but keeping their ratio constant scans the m/z that the quadrupole filter will allow to pass. The frequency of the RF is also kept constant and is usually several hundred kilohertz (e.g. 700 kHz).

One of the features of the quadrupole mass filter is that it can be used as an ion-focusing device if the DC component is turned off. Thus, an RF-only quadrupole

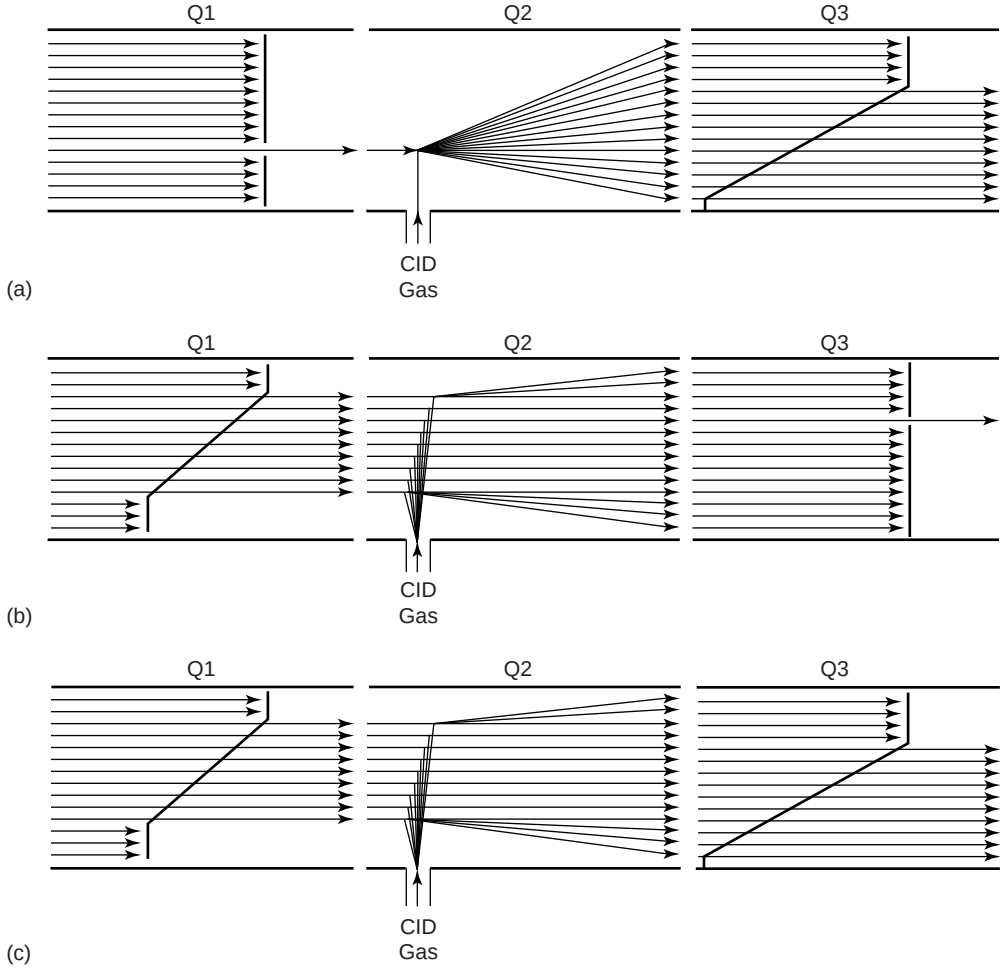

Figure 3 Triple quadrupole mass spectrometer scan modes: (a) product ion scan; (b) precursor ion scan; (c) neutral loss scan.

can be used to transmit ions with a broad range of m/z values and kinetic energies. Yost and Enke[15,16] took advantage of this property to construct an ion collision chamber in their invention of the triple quadrupole mass spectrometer in 1979. In this device, quadrupoles one (Q1) and three (Q3) can be scanned in the usual manner using RF and DC voltages, and the middle quadrupole (Q2) is RF-only, which usually contains a collision gas such as argon at relatively high pressures (in the millitorr region). Ions exiting Q1 are activated by collisions with the argon in Q2. The kinetic energy of these collisions is converted into internal vibrational modes and the ions quickly dissociate into characteristic fragment ions that are analyzed in Q3. This ion activation/dissociation technique is commonly referred to as collision-induced dissociation (CID). The entire process of ion selection, CID, and product ion analysis is referred to as tandem MS and provides considerable qualitative structural information and improvements in signal-to-noise ratios for quantitative target compound analysis.

There are three scan modes that enable the triple quadrupole to be very useful for both qualitative and quantitative analysis (Figure 3). In a product ion scan (Figure 3a), ions are selected by their m/z in Q1, undergo CID in Q2, and are mass analyzed in Q3. This scan mode is particularly useful in providing structural information from the fragmentation pattern of species that do not readily fragment upon ionization in the mass spectrometer source. For example, CI frequently gives a prominent $(M + H)^+$ ion but very little structural information. A product ion scan can provide this information. In a precursor ion scan (Figure 3b), Q3 is set to pass only a particular m/z. Q1 is scanned to provide information on what species fragment to give the species detected in Q3. An example of the application of this scan mode is in the analysis of the priority pollutant phthalate esters in the environment by Hunt et al.[17] These authors performed methane CI on unseparated mixtures of phthalates in environmental sample extracts using a precursor ion scan of m/z 149, which is a characteristic fragment ion for these compounds. The presence of the phthalates in the sample was confirmed by the presence of the characteristic $(M + H)^+$ ions and fragment ions that give rise to m/z 149. In a neutral loss scan (Figure 3c), Q1 and Q3 are scanned at a constant m/z offset that represents the loss of a particular functional group (i.e. the system is set to detect only a particular chemical group that is lost upon CID). As an example, Hunt also found that all but one of a series of priority pollutant chlorinated hydrocarbons underwent the neutral loss of the $Cl^{\bullet}$ radical and HCl. Therefore, performing a neutral loss scan at 35 and 36 is a way of monitoring for the presence of chlorocarbons.

The third type of quadrupole used is the quadrupole ion trap. This device has the great advantage of being able to act as a mass spectrometer and as a device that can be used to store/isolate ions, collisionally activate a selected ion, trap the product ions and scan their m/z ratios. In essence, the ion trap acts as an ''electric field test tube'' and consists of three electrodes, shown separated (Figure 4[18]) and assembled with electrical connections (Figure 5). Two of the electrodes are of hyperboloidal geometry and are called the end-cap electrodes. The end-cap electrodes have small holes in them that can be used to admit ions into the trap or used for the detection of ions ejected from the trap. The third electrode is in the shape of a ring with a hyperbolic cross-section on the inner surface of the ring. Thus, the quadrupole ion trap actually consists of only three electrodes, but in cross-section it looks similar to the end-view of the linear quadrupole mass filter and the equations that govern the motion of the ions are similar for the two devices. Thorough and excellent descriptions of the operation of the ion trap are given in a series of works by March et al.[18–20] The popularity of the ion trap as a mass spectrometer for

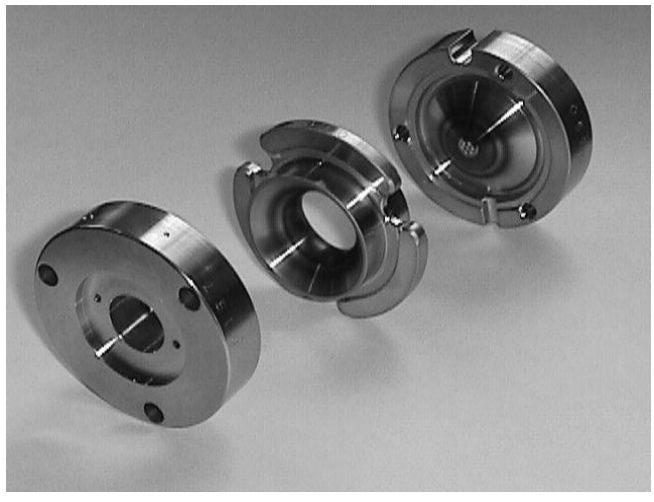

Figure 4 Ion trap with electrodes separated.

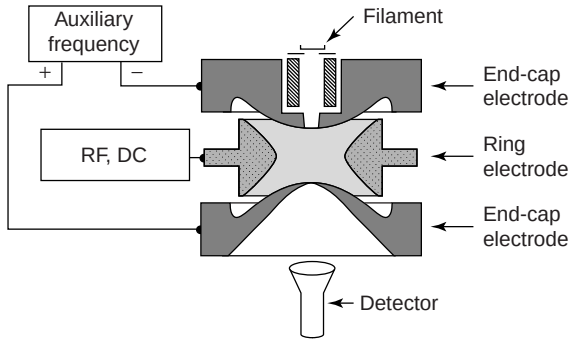

Figure 5 Quadrupole ion trap shown assembled.

separations has grown primarily because it is a relatively low-cost device and is capable of performing tandem MS. It is estimated that more than 4000 ion traps had been sold by 1997 at a total cost of $250 million.[18]

3 IONIZATION METHODS

3.1 Electron Ionization

EI is the oldest and most widely used ionization technique in MS. It has its roots in the development on MS made by J.J. Thomson at the beginning of the 20th century. EI continues to dominate MS because it is the simplest ionization technique to implement. EI deposits enough excess energy in the molecular ion to cause at least partial fragmentation, and extensive libraries of EI mass spectra are readily available. One of the chief disadvantages of EI is that it requires the analyte to be present in the mass spectrometer source as a vapor and is therefore only applicable to molecules that are vaporizable and stable in the gas phase. The minimum vapor pressure required is about 1×10^{-6} Torr. This disadvantage means that EI is not applicable to a wide range of organic compounds, e.g. salts (quaternary amines, carboxylates, etc.), nonionic hydrogen-bonded functional groups (carbohydrates, amino acids, oligonucleotides, etc.) and species of large molecular weight. In spite of these disadvantages, EI is still the most widely used ionization technique in MS because of the strong advantages listed above and because the range of compounds applicable to EI matches well with the wide variety of compounds used commercially. The molecular weights of these compounds primarily cover species below molecular weight 500.

All MS techniques require the formation of an isolated ion in the gas phase. In EI, the molecule to be ionized is heated to provide a pool of molecules in the vapor phase. The pool of gas-phase molecules can come from the effluent of a gas chromatograph, which makes EI an ideal match for GC. Ionization is accomplished by collision between an analyte molecule and an electron accelerated off a heated tungsten wire (the ''filament''). The typical energy of the electrons used for EI is 70 eV, which is generated by a potential difference of 70 V between the filament and the body of the ion source. A diagram of this process is shown in Figure 6. Because the nuclei of atoms are extremely small compared to the electron cloud, the most probable collisions will occur between the 70-eV electron and an electron circling the molecule. This collision results in the ejection of an electron from the molecule and results in the formation of a molecular radical cation, $M^{+\bullet}$. This cation is accelerated out of the ion source by the negative potential on an extracting plate,

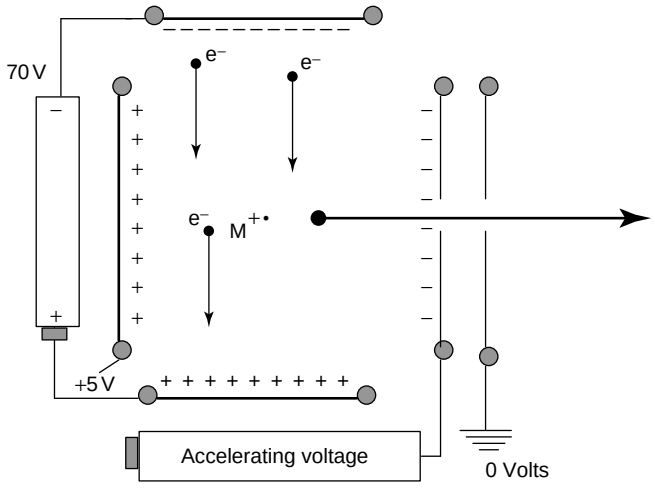

Figure 6 Schematic diagram of an EI source.

by a potential on the entire source, or by a combination of these as illustrated in Figure 6. The EI process can be represented by Equation (1):

$$M^0 + e^-(70\,eV) \longrightarrow M^{+\bullet} + 2e^- \qquad (1)$$

In addition to providing enough energy to ionize the molecule, internal energy is also deposited in the resulting radical cation. This internal energy can be as high as 350 kcal mol^{-1}, which is enough energy to fragment even the strongest chemical bond. If the time required for fragmentation is longer than about 10^{-6} s, the $M^{+\bullet}$ ion will travel through the mass spectrometer and be measured at the m/z of the intact molecular ion. Usually, a portion of the $M^{+\bullet}$ molecules dissociates in the source of the mass spectrometer and is measured as fragment ions that can be used to characterize the chemical structure of $M^{+\bullet}$, because the fragmentations follow well-known rules. The resulting mass spectrum consists of ions that represent the molecular weight of the molecule as well as ions that characterize the chemical structure of the molecule. The EI mass spectrum, therefore, contains a wealth of information about the molecules entering the source of the mass spectrometer.

There are some disadvantages to EI. Only about 1 in 10 000 of the molecules present in the ion source are ionized by the EI process, making it an inefficient ionization process. In addition to this, many molecules do not give an ion that represents the intact molecule (called the molecular ion). Two examples of molecules that give weak or absent molecular ions are long-chain, branched hydrocarbons and alcohols (which usually fragment to lose water). For example, the percent abundances of the molecular ions for 3,3,5-trimethylheptane and *n*-decanol are 0.007% and 0.002%, respectively.[21] This can lead to problems in identifying the compounds that produced

the mass spectrum, because many of the compounds of similar structure have very similar mass spectra.

3.2 Chemical Ionization

As an alternative to electron impact ionization, ionization of the sample can be produced using ionic reactions. This technique, called CI, uses a high pressure of a reagent gas (e.g. methane or ammonia) in the ionization source.[22] The filament produces electrons as in EI (usually set to 100 eV), and these ionize the reagent gas. The reagent gas ions then react with the sample molecules to produce mainly molecular ions (through proton transfer for methane CI). The chemical reaction for the methane CI process is shown by Equations (2–4):

$$CH_4^0 + e^-(100\,eV) \longrightarrow CH_4{}^+ + 2e^- \quad (2)$$

$$CH_4{}^{+\bullet} + CH_4^0 \longrightarrow CH_5{}^+ + CH_3{}^\bullet \quad (3)$$

$$M + CH_5{}^+ \longrightarrow MH^+ + CH_4 \quad (4)$$

The equilibrium of Equation (4) is governed primarily by the difference in proton affinity of the molecule M and methane. The proton affinity is shown by Equation (5):

$$M + H^+ \longrightarrow MH^+ \quad (5)$$

The internal energy deposited in MH^+ depends upon this difference in proton affinities. The extent of fragmentation of MH^+ and the sensitivity to CI, therefore, can be controlled by choice of CI reagent ions and the difference in proton affinities with respect to the target analyte. The most common CI reagent ions are methane, isobutane, and ammonia, although many other reagent ions are used for specific applications. An example of the selectivity afforded by the judicious choice of CI reagent is illustrated by the use of ammonia as the CI reagent ion. Ammonia has a relatively high proton affinity (204 kcal mol^{-1}) so it will only donate its proton to molecules such as aromatic and aliphatic amines. Ammonia does not protonate alcohols, carboxylic acids, and hydrocarbons that do not contain the nitrogen moieties, because their proton affinities are lower than that of ammonia.

The major advantage of CI is that it produces primarily the protonated molecular ion, which makes it fairly easy to determine the molecular weight for the compound. Because EI usually produces a strong fragment ion spectrum and CI produces primarily the protonated molecular ion, these two techniques are complementary in nature. The major disadvantage of this technique is that it requires a better pumping system for the mass spectrometer due to the use of a high pressure of reagent gas, which leads to a higher cost for the instrumentation.

Another important type of CI is negative ion chemical ionization (NICI). In this technique, a reagent gas is used as in conventional CI, but it is used as a moderator for the electrons produced by the filament.[23] These low-energy electrons are readily captured by many organic compounds, most notably halogenated compounds. The chemical reaction for the NICI process is shown by Equation (6):

$$M + e^- \longrightarrow M^- \quad (6)$$

There are two major advantages of NICI. First, it is a highly selective technique, because only those compounds that contain a functional group with a high electron capture cross-section will be ionized by this technique. An application of this is the determination of PCBs in fish samples.[24] The chlorinated biphenyls are efficiently ionized using NICI, whereas the lipids and oils in the fish matrix remain unionized, resulting in a very clean chromatogram. Secondly, NICI is a much more efficient process than either EI or positive ion CI. For compounds that contain multiple electronegative functional groups, the sensitivity can be several orders of magnitude greater under NICI mode than EI or CI.

4 SAMPLE PREPARATION AND ANALYSIS

This section presents the common methods used for the preparation of samples prior to analysis by GC/MS. For water samples, the desired detection limits are frequently less than can be achieved by direct injection of samples into the gas chromatograph. For example, the quantity detectable by a quadrupole mass spectrometer operated in EI mode is often of the order of several nanograms. Direct aqueous injection of 1 μL of water would lead to a concentration detection limit of several nanograms per microliter, i.e. in the parts-per-million range. A detection limit at the parts-per-million level is frequently several orders of magnitude too high for many environmental applications. Also, injection of water can be deleterious to many GC stationary phases over a long period of time. Hydrolysis of labile species can result from reaction with water on the hot surfaces of an injection port, and water is a poor choice of solvent for GC due to its high boiling point and polarity. For these reasons, samples are frequently subjected to extraction followed by concentration prior to analysis by GC/MS. Common preparation procedures are presented below for water, soil, and solid waste.

Sample preparation for GC/MS analysis is commonly divided into two techniques that depend upon the volatility of the species being analyzed. Volatile compounds are often considered as those species that have a boiling point below about 200 °C; semi-volatiles are those

species that are amenable to GC analysis and have boiling points above 200 °C. The upper end of the temperature range accessible to GC includes species whose boiling points are less than approximately 350–400 °C. There are, however, special GC columns designed for higher-temperature analyses. Fused-silica capillary columns coated with 5% diphenyl polysiloxane and 95% dimethyl polysiloxane are the most popular general-purpose columns. These columns have upper temperature limits of 350 °C. Because of the temperature limitations mentioned above and the thermal stabilities of organic compounds, many classes of organic compounds are not amenable to GC/MS analysis. Various authors have estimated that GC is useful for only 20% of organic compounds.[25,26] Examples of substances that cannot be analyzed by GC without prior derivatization include carbohydrates, proteins, fatty acids, salts, and other highly polar or high-molecular-weight species.

4.1 Sample Preparation for Volatile Organic Analysis

4.1.1 Sample Preparation for Volatiles in Water and Soild Samples

Volatile organic compounds typically analyzed by GC/MS include freons (chlorofluorocarbons), chlorinated solvents (such as methylene chloride and perchloroethene), and aromatics (such as benzene and toluene). The analysis of these compounds in environmental samples presents the challenge of preventing losses between sampling and analysis. Samples need to be collected in scrupulously cleaned glass vials with Teflon® septum seals and stored at 4 °C immediately after collection. For water samples, the vial should be filled gently to overflowing then sealed without a headspace gap between the top of the water and the septum. Soil, sediment, and waste samples need to be collected in preprepared glass screw-cap vials containing 5 mL of organic-free water, a clean magnetic stirring bar, and any preservative required (such as sodium bisulfite). Once the sample (typically 5 g) is placed into the vial and sealed with a Teflon® septum and cap, the vial is not opened until analysis.

Volatile organic compounds are most often prepared by purge-and-trap techniques, which are a type of dynamic headspace sampling. An inert gas, such as helium, is bubbled through an aqueous sample of 5–25 mL volume or a soil/sediment sample containing 5 mL of water for 11 min. The partitioning of the volatile organics between the aqueous phase and the vapor phase is governed by Henry's law. Organic compounds of low aqueous solubilities are efficiently transferred to the gaseous stream, which is then passed through a metal tube containing a selection of various adsorbents to trap the organic compounds and remove them from the helium stream. The adsorbents commonly used include silica gel to remove water, charcoal to trap highly volatile species, and various polymers (such as Tenax®, a porous polymer resin based on 2,6-diphenylene-oxide) that trap the remaining compounds. After the purge cycle is completed, the trap is rapidly heated to 180 °C or above (depending on the adsorbants used in the trap) and the analytes are back-flushed onto the GC column. Figure 7 shows a typical purge device configured for water.

Wide-bore GC columns are often used but may require jet separator or open-split interfacing to the mass spectrometer, depending upon the pumping speed of the mass spectrometer. Capillary GC columns are also used but they often require cryogenic focusing at the head of the GC column. In cryofocusing, a short section at the

Figure 7 Purge-and-trap concentrator.

head of the GC column is held at −150 °C by directing a stream of cold gas across the column, e.g. gaseous nitrogen from liquid nitrogen boil-off. After desorption from the trap is complete, this short section of column is rapidly heated to 250 °C in 15 s to desorb the analytes into a tight band at the head of the GC column. Capillary GC has the advantage of enabling direct interfacing to the ion source. A detailed reference for the application of the purge-and-trap technique can be found in the compilation of methods developed by the USEPA for sampling and analysis under Subtitle C of the Resource Conservation and Recovery Act. These methods are published under the title "Test Methods for Evaluating Solid Waste, Physical/Chemical Methods (SW-846)". The method for the purge-and-trap analysis of volatiles in water samples has the method number 5030B, whereas the method number for soild and solid waste samples is 5035. Another useful method for the analysis of volatiles in water has been produced by the US Geological Survey.[27]

Volatile water and soil samples can also be analyzed by static headspace (USEPA Method 5021) and by membrane inlet MS. In the former method, there is a single step of equilibrium partitioning between the liquid phase and the gas phase. A gaseous sample is then withdrawn for MS analysis. In membrane inlet MS, an organic-permeable membrane serves as the inlet to the mass spectrometer, and the gas chromatograph is eliminated. Because there is no separation of the analytes, this technique is more often used with quadrupole ion trap mass spectrometers that are capable of multiple stages of MS in order to provide analyte selectivity.

4.1.2 Sample Preparation for Volatiles in Air Samples

Volatile organic compounds are frequently sampled by drawing an air sample through a metal or glass tube that contains an appropriate sorbent. Synthetic polymers, e.g. Tenax®, can be used alone or in combination with activated carbon, as in USEPA Methods 5041A and 0030 for sampling and analysis of volatiles from hazardous waste incinerators[28] or ambient air.[29] Carbon sorbents have also been used alone[30–32] for the analysis of hazardous air pollutants in ambient air. The high specific surface area of many carbon-based sorbents (100–1000 $m^2\ g^{-1}$) leads to efficient trapping of volatile organic compounds. The target analytes are later thermally desorbed into a purge-and-trap device (see section 4.1.1) or directly cryofocused onto the head of the GC column. An advantage of carbon sorbents over Tenax® is their higher thermal stability and lower background levels of interferences.

Canister-based sampling is currently the preferred method for collecting air samples for volatile organic analyses. In this method (USEPA Method TO14), a whole-air sample is drawn into a stainless-steel canister that has been surface-treated to minimize adsorptive losses.[29] This method allows for a greater range of compounds to be sampled (including many polar organic compounds such as alcohols and ketones). It also allows for multiple analyses from the same air sample, and has much lower background levels than adsorbant tube sampling. For analysis, a portion of the air is removed and concentrated on a cryotrap or a sorbent bed prior to final analysis by GC/MS.

Analytical detection limits for the analysis of volatile organic compounds in air by GC/MS are frequently below 1 ppb by volume where the sample volume ranges from 1 to 5 L. Hydrolysis losses due to reaction with moisture from the air can be troublesome, e.g. hydrolysis of methyl acetate.

For volatiles analyses, the mass spectrometer is usually set to scan from m/z 35 to 260, with a scan time of 0.6–2 s per scan. The lower limit of the scan range is set to avoid the region of air (m/z 28 for N_2 and m/z 32 for O_2) and water that is entrained in the trap (m/z 18 for H_2O).

4.2 Sample Preparation for Semi-volatile Organic Analysis

4.2.1 Sample Preparation for Semi-volatiles in Water and Solid Samples

Semi-volatile organic compounds cannot be purged efficiently from aqueous media. These compounds (e.g. polyaromatic hydrocarbons, chlorinated phenols, and certain pesticides) must be extracted out of the sample using an organic solvent. For the extraction of these compounds from water, three techniques are commonly employed: separatory funnel liquid–liquid extraction; continuous liquid–liquid extraction; and solid-phase extraction. The liquid–liquid extraction techniques use an immiscible organic solvent (such as methylene chloride) to remove nonpolar organic materials from the polar matrix of water. Of course, the partitioning of organic compounds between these two media is governed by the relative solubilities in each medium (water versus organic solvent), so this approach is not useful for highly polar or ionic organic substances such as organic salts. Adjustment of pH can be used to modify the solubilities of organic acids and bases prior to extraction. For example, raising the pH to greater than 12 and then performing the extraction will extract both basic compounds and neutral compounds but will not extract acidic compounds, which will be ionic at high pH. An extraction at high pH is called a base–neutral fractionation. Similarly, extraction at low pH will remove acids and neutrals from the water sample and leave behind the basic compounds, which are acidic at low pH. By using a series of extractions at high then low pH, acidic, neutral and basic compounds can be

extracted into the same organic solvent and can then be combined or analyzed separately.

Separatory funnel extraction has the disadvantage of being somewhat labor intensive. An aliquot of extraction solvent is added to the separatory funnel containing the aqueous sample after the pH is adjusted. The separatory funnel is then shaken for several minutes to ensure good contact between the extraction solvent and the aqueous sample. The two phases are allowed to separate for about 10 min before drawing off each aliquot of extraction solvent. This procedure is generally repeated three times to ensure efficiency of extraction. The extraction solvent is often an organic solvent that is denser than water so that it will sink to the bottom of the separatory funnel and be drawn off through the stopcock; however, the technique can be modified for use with extraction solvents that are less dense than water. One advantage of separatory funnel extraction over continuous liquid–liquid extraction is speed (approximately 1 h for separatory funnel versus 18–24 h for continuous liquid–liquid extraction). This speed factor can be critical when analyzing for unstable compounds, e.g. organochlorine pesticides may dechlorinate under basic extraction conditions. One disadvantage of separatory funnel extractions is the propensity of some samples to form emulsions between the aqueous layer and the organic layer, which can lead to analyte losses if the emulsion is not broken. If the emulsion interface between layers is more than one-third of the size of the solvent layer, mechanical techniques such as stirring, filtration of the emulsion through glass wool, centrifugation, or sonication must be used to complete the phase separation. The USEPA Method 3510C from SW-846 is an excellent reference for liquid–liquid extraction by separatory funnel.

A continuous liquid–liquid extractor is shown in Figure 8. It is designed for extraction solvents with greater density than aqueous media, such as methylene chloride. Continuous liquid–liquid extraction devices are also available for extraction solvents that are less dense than the sample. The advantage of this device is that it continuously provides contact between the sample and fresh solvent that has been distilled from the receiving boiling flask to the condenser and back through the sample. Once the volume of extraction solvent has reached the level of the automatic siphon, this extract is siphoned to the boiling flask and the process is repeated. The USEPA Method 3520C from SW-846 is an excellent resource for this approach. The biggest advantage in using a continuous liquid–liquid extractor over separatory funnels is that emulsions are not formed in this apparatus, thus ensuring high extraction efficiencies.

One of the fastest methods for extracting organic analytes from aqueous media is the use of a solid-phase extraction cartridge. In this approach, the aqueous sample

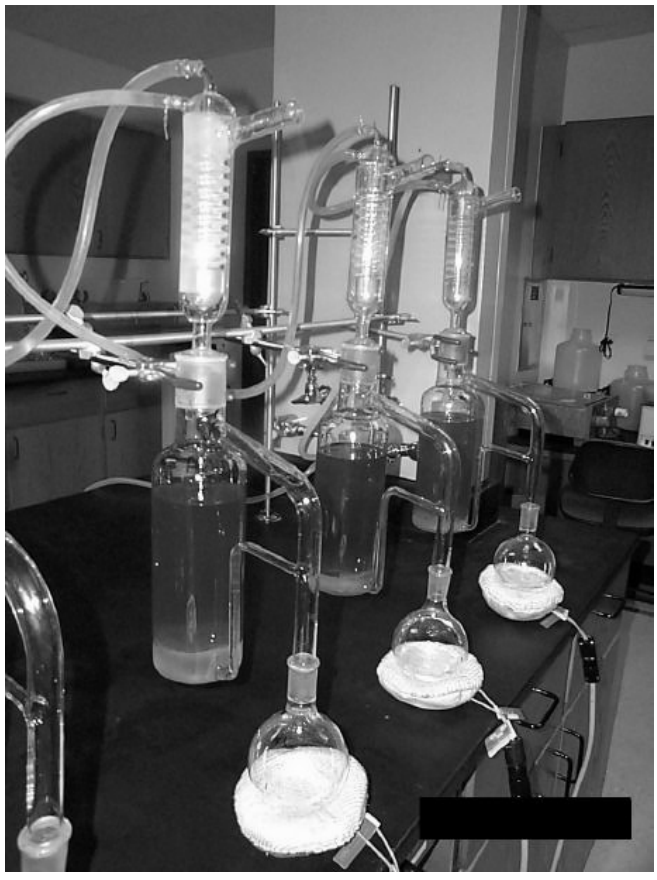

Figure 8 Continuous liquid–liquid extractor.

is passed through a sorbent trap or disk containing an imobilized organic material. The target organic analytes in the aqueous sample partition to the solid phase. After the sample has passed through and the sorbent washed with clean water, the analytes are eluted from the solid phase with a strong organic solvent. One of the chief disadvantages of the solid-phase extraction approach is that many nonpolar organic contaminants are likely to be bound to particulate matter in an aqueous sample and extraction recoveries from real samples are expected to be less than those determined from simply spiking organic-free reagent water. Sample particulates can also clog the solid-phase cartridge. The USEPA Method 3535 from SW-846 is an excellent resource for this approach. Figure 9 shows an example of a disk extraction apparatus.

There are two common sample preparation methods for semi-volatiles in solids, and both involve the use of an organic solvent to extract the organic compounds; one uses a continuous extraction device called a Soxhlet extractor, and the other uses an ultrasonic probe to ensure intimate contact between the sample and the extraction medium. In both approaches, the solid sample is frequently mixed with an equal weight of

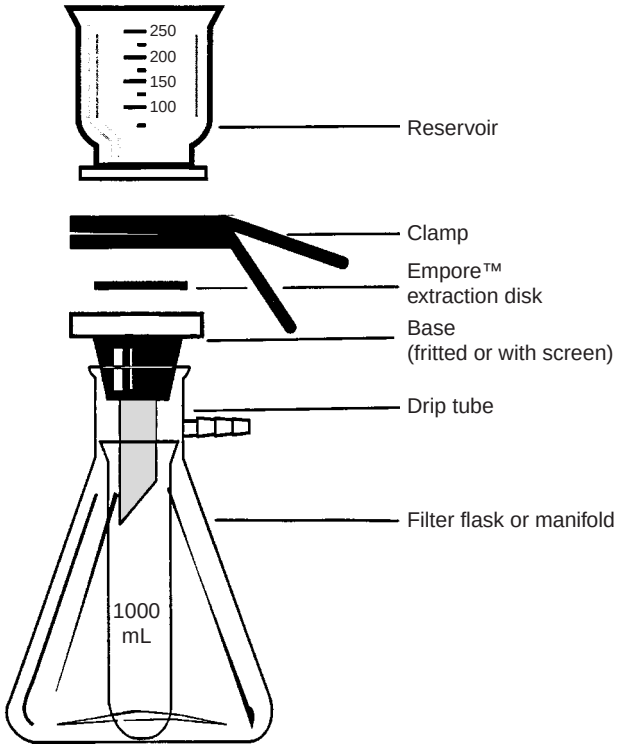

Figure 9 Disk extraction apparatus.

anhydrous sodium sulfate to remove water and ensure good extraction solvent contact with the solid sample. The dried sample is placed in an extraction thimble or between two plugs of glass wool, and extracted using continuously renewed solvent flowing through the sample from the condenser above the Soxhlet extractor. As with the continuous liquid–liquid extractor, an automatic siphon is employed. The chief advantage of the Soxhlet extractor is that it frequently gives higher recoveries than the ultrasonic probe; however, the extraction times are much longer (16–24 h at 4–6 cycles h^{-1}). The extract is then dried with anhydrous sodium sulfate, and further processed as necessary before analysis by GC/MS (concentrated, cleaned-up, etc.). The USEPA Method 3540C from SW-846 is an excellent resource for Soxhlet extraction.

Ultrasonic extraction, although not as rigorous as Sohxlet extraction, does provide adequate recoveries that are usually greater than 50%, with the advantage of being much faster than Soxhlet extraction (triplicate extraction steps with the ultrasonic probe for 3 min each versus 16–24 h with the Soxhlet extractor). In ultrasonic extraction, the sample is mixed with the extraction solvent and subjected to agitation using an ultrasonic probe of usually 300 W power. The shaking provided by the ultrasonic probe ensures intimate contact between the sample and the extraction solvent. One of the drawbacks of using ultrasonic extraction is that the ultrasonic energy used may cause the breakdown of some organic compounds (e.g. organophosphorus pesticides).

It is important to dry the sample thoroughly by mixing with anhydrous sodium sulfate prior to extraction. After each stage of extraction, the extraction solvent is removed by vacuum filtration or centrifugation followed by decanting the extraction solvent. The USEPA Method 3550B from SW-846 is an excellent resource for Soxhlet extraction.

A newer method for extracting organic analytes from solid matrices is the use of solvent extraction under higher temperature and pressures than used with Soxhlet or sonication methods. These methods – using a pressurized fluid reactor (USEPA Method 3545) or a microwave extractor (USEPA Method 3546) – both involve placing the sample as a fine powder in a reaction vessel and adding an extraction solvent (typically methylene chloride, acetone, hexane, or a combination of these). The vessel is then heated to approximately 100 °C and pressurized to 1500 psi. This allows the extraction to take place in less than 15 min using only 15 mL of solvent, versus 16–18 h and 300 mL of solvent using the Soxhlet technique. Reported recoveries for these new techniques are equivalent to those achieved using the Soxhlet technique.

For all analyses for sediments, soils, and other solid media, it is customary to determine the percentage moisture in the sample and report the analyte concentrations on a dry weight basis. This approach can help to remove some of the variability associated with sampling media of high but variable moisture content.

4.2.2 Sample Preparation for Semi-volatiles in Air

A consideration for the analysis of semi-volatiles and nonvolatiles in air is the low analyte concentrations that are of concern. For example, the analysis of polychlorinated dibenzo-*p*-dioxins can have target detection limits below picogram-per-cubic-meter levels (pg m^{-3}). These low detection limits require large sampling volumes of the order of hundreds of cubic meters, which have led to the design of high-volume air samplers.[33] Because semi-volatile analytes can be associated with particulates in the air or in the vapor phase, the sampling train usually involves a combination of quartz-fiber filters to trap particulates and an adsorptive trap such as polyurethane foam. The filter and adsorptive trap are extracted by Soxhlet extraction, concentrated, cleaned up by column chromatography, and analyzed by GC/MS.

4.3 Sample Extract Clean-up

In order to achieve the low detection limits that are required for the analysis of trace organic compounds, both

sample clean-up to remove interfering compounds and extract concentration have to be performed frequently. In the analysis of water, control of the pH during extraction can be considered as a sample clean-up step; however, further sample clean-up can be achieved through column chromatography and/or high-performance liquid chromatography. Column chromatography using silica gel, alumina, or Florisil® adsorbents is frequently used to separate crudely the target analytes according to sample polarity from interferences. An example of this approach is the use of silica gel (silicic acid) to clean up extracts and separate organochlorine pesticides from PCBs in environmental samples.[28] In this clean-up, a hexane sample extract is added to a column containing approximately 1 g of silica gel. Increasingly polar analytes are then eluted by passing increasingly polar extraction solvents through the column. The first fraction is hexane, and it elutes the PCBs and nonpolar pesticides such as dichlorodiphenyltrichloroethane. The second fraction is eluted using 50% (v/v) ethyl ether in hexane and contains the more polar pesticides, e.g. Endosulfan I and II. This is a convenient way to separate the multicomponent PCBs from many of the single-component pesticides that will co-elute on the GC column.

Gel permeation chromatography is also used for sample clean-up; it is based upon molecular size and is most useful for eliminating high-boiling interferences that can contaminate the GC injection port and column. It is applicable to both polar and nonpolar analytes. Typical contaminants that can be removed by gel permeation chromatography include lipids, polymers, copolymers, proteins, natural resins and polymers, cellular components, viruses, steroids, and dispersed high-molecular-weight compounds. A detailed discussion of the application of many sample clean-up methods is contained in the USEPA Test Methods for Evaluating Solid Waste, Physical/Chemical Methods (SW-846).[28]

4.4 Chemical Derivatization

Many compounds cannot be analyzed by GC/MS because they are not amenable to GC. The major reasons for this are that these compounds do not have sufficient volatility (due to molecular weight or polarity) or they are thermally labile and hence decompose under the GC conditions. Although most of these compounds can be analyzed using liquid chromatography/mass spectrometry (LC/MS) it is still desirable to be able to analyze some of them using GC/MS. Instrument availability (GC/MS instruments usually cost much less than LC/MS instruments) and ease of use[34] are determining factors. The most common derivatives are used to make polar compounds more amenable to GC analysis. These include methylation, trimethylsilylation, and acylation. All of these reactions form stable, volatile compounds from carboxylic acids, alcohols, phenols, amines, and thiols, and each has its advantages when analyzing for a given set of compounds. An added benefit to derivatization is that electronegative functional groups can be added to compounds to enhance detectability; at the same time, the compound is modified to make it amenable to GC analysis. An example of this is the pentafluorobenzyl bromide derivative of haloacetic acids, as shown by Equation (7):

$$CCl_3-COOH + C_6F_5CH_2Br \longrightarrow CCl_3-COOCH_2C_6F_5 + HBr \qquad (7)$$

This reaction not only produces a volatile ester derivative from the carboxylic acid, but the addition of five fluorines to the molecule makes the detection limit for this compound in the femtogram range by NICI.

5 INSTRUMENTAL ANALYSIS CONSIDERATIONS

The final instrumental analysis in GC/MS has various requirements for the mass spectrometer, depending upon the analysis being performed. The first requirement is that the scan speed of the mass spectrometer must be sufficient to obtain at least ten scans across a chromatographic peak. This requirement is necessary to define adequately the chromatographic peak profile for quantitation. The scan speed for quadrupole GC/MS is usually 0.5–1 s per scan over the m/z range desired. This scan speed is adequate for most capillary chromatography applications where the peaks are usually not less than 20 s wide; however, for high-speed chromatography, where the chromatographic peaks can be only a few seconds wide, faster mass spectrometers such as time-of-flight instruments must be used. The mass range scanned can extend from m/z 10 to m/z 1000 on some instruments; however, few compounds with molecular weights over 500 are amenable to GC. Many analysts do not scan below m/z 35 to avoid background mass spectral peaks due to N_2 at m/z 28, O_2 at m/z 32, and water at m/z 18.

In the analysis for unknown substances, the mass spectrum can be very useful for qualitative identification by comparison to the extensive libraries of EI mass spectra. Automated search routines are available for comparison of unknown spectra to library spectra. Overlap of unknown peaks from co-eluting GC peaks give mixtures of mass spectra that in the past have been difficult to interpret. New software to deconvolute these overlapping spectra is becoming available.

In the analysis of target compounds GC/MS has been particularly successful, especially in the environmental

areas and for the analysis of drugs. In target compound analysis, frequently three strong and/or unique ions are selected as the characteristic ions to search for in a window where the retention time matches that of a standard. The relative ratios of these ions and their retention times should maximize within one or two mass spectral scans in order to confirm the presence of a target compound. Sensitivity can be enhanced by selected ion monitoring, where only a limited number of target ions are scanned. The mass spectrometer jumps between the set of target ions rather than scanning over the entire mass range. This approach avoids the dead time between ions of different m/z ratios because there is no need to scan over regions where there are no ions of interest. This technique of selected ion monitoring can improve detection limits and signal-to-noise ratios by over an order of magnitude. Tandem MS using quadrupole ion traps or triple quadrupole mass spectrometers is another technique to greatly enhance the signal-to-noise ratios for target compound analysis. In this technique, a particular target ion is selected and then dissociated to give a characteristic fragment ion. Because only these coupled ions are monitored, the chemical noise in the analysis is greatly reduced and the selectivity is greatly enhanced.

Quantitative analysis by GC/MS requires the use of both external standards and internal standards. The variability of ion optics from scan to scan mandates the use of internal standards in order to obtain good quantitation with relative standard deviations that can frequently be in the range of 5–10%. In order to obtain relative standard deviations below 5%, GC/MS frequently requires the use of isotopically labeled internal standards.

6 SUMMARY

GC coupled to quadrupole MS is one of the most powerful analytical techniques available due to its selectivity and sensitivity, but it does have limitations: the compounds must be volatile and stable in the gas phase and be able to survive ionization. However, the ability to identify and quantitate compounds that meet these criteria in complex matrices and mixtures is unmatched.

ABBREVIATIONS AND ACRONYMS

CI	Chemical Ionization
CID	Collision-induced Dissociation
DC	Direct Current
EI	Electron Ionization
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
LC/MS	Liquid Chromatography/Mass Spectrometry
MS	Mass Spectrometry
NICI	Negative Ion Chemical Ionization
PCB	Polychlorinated Biphenyl
RF	Radiofrequency
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Sampling Considerations for Biomonitoring • Soxhlet and Ultrasonic Extraction of Organics in Solids

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas

Chromatography • Sample Preparation for Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Chemical Ionization Mass Spectrometry: Theory and Applications • Gas Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer

REFERENCES

1. P.J. Arpino, in *Selected Topics in Mass Spectrometry in the Biomolecular Sciences*, eds. R.M. Caprioli, A. Malorni, G. Sindona, Kluwer Academic, London, 85–101, 1997.
2. G.A. Eiceman, H.H. Hill, Jr, B. Davani, J. Gardea-Torresday, 'Gas Chromatography', *Anal. Chem.*, **68**, 291R–308R (1996).
3. G.A. Eiceman, H.H. Hill, Jr, J. Gardea-Torresday, 'Gas Chromatography', *Anal. Chem.*, **70**, 321R–339R (1998).
4. W. Jennings, E. Mittlefehldt, P. Stremple, *Analytical Gas Chromatography*, Academic Press, San Diego, 1997.
5. W. Paul, H. Steinwedel, 'A New Mass Spectrometer Without a Magnetic Field', *Z. Naturforch.*, **8A**, 448–450 (1953).
6. W. Paul, H. Steinwedel, 'Apparatus for Separating Charged Particles of Different Specific Charges', German Patent Number 944 900, June 28, 1956.
7. P.H. Dawson, 'Quadrupole Mass Analyzers: Performance, Design, and Some Recent Applications', *Mass Spectrom. Rev.*, **5**, 1–37 (1986).
8. R.E. Finnigan, 'Quadrupole Mass Spectrometers: From Development to Commercialization', *Anal. Chem.*, **66**, 969A–975A (1994).
9. A.T. James, A.J.P. Martin, 'Gas–Liquid Partition Chromatography: The Separation and Microestimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid', *Biochem. J.*, **50**, 679–690 (1952).
10. R.S. Gohlke, 'Time-of-flight Mass Spectrometry and Gas–Liquid Partition Chromatography', *Anal. Chem.*, **31**, 535–541 (1959).
11. J.T. Watson, K. Bieman, 'High-resolution Mass Spectra of Compounds Emerging from a Gas Chromatograph', *Anal. Chem.*, **36**, 1135–1137 (1964).
12. R. Rhyage, 'Use of a Mass Spectrometer as a Detector and Analyzer for Effluents Emerging from High Temperature Gas–Liquid Chromatography Columns', *Anal. Chem.*, **36**, 759–764 (1964).
13. P.M. Llewellyn, D. Littlejohn, 'Dual-stage-membrane Gas Separators With Variable Conductance Means for Varying Their Throughput', *Abstracts of papers presented at the 17th Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy*, 152, 1966.
14. P.E. Miller, M.B. Denton, 'The Quadrupole Mass Filter: Basic Operating Concepts', *J. Chem. Educ.*, **63**, 617–622 (1986).
15. R.A. Yost, C.G. Enke, 'Selected Ion Fragmentation With a Tandem Quadrupole Mass-spectrometer', *J. Am. Chem. Soc.*, **100**, 2274–2275 (1978).
16. R.A. Yost, C.G. Enke, D.C. McGilvery, D. Smith, J.D. Morrison, 'High-efficiency Collision-induced Dissociation in an Rf-only Quadrupole', *Int. J. Mass Spectrom. Ion Phys.*, **30**, 127–136 (1979).
17. D.F. Hunt, J. Shabanowitz, T.M. Harvey, M. Coates, 'Scheme for the Direct Analysis of Organics in the Environment by Tandem Mass Spectrometry', *Anal. Chem.*, **57**, 525–537 (1985).
18. R.E. March, 'An Introduction to Quadrupole Ion Trap Mass Spectrometry', *J. Mass Spectrom.*, **32**, 351–369 (1997).
19. R.E. March, R.J. Hughes, J.F.J. Todd, *Quadrupole Mass Spectometry and its Applications*, Elsevier, Amsterdam, 1976.
20. R.E. March, J.F.J. Todd, *Practical Aspects of Ion Trap Mass Spectrometry*, CRC Press, Boca Raton, FL, 1995.
21. K. Biemann, *Mass Spectrometry – Organic Chemical Applications*, McGraw-Hill Book Company, New York, 1962.
22. T. Maurer, H. Hass, I. Barnes, K.H. Becker, 'Kinetic and Product Study of the Atmospheric Photooxidation of 1,4-Dioxane and its Main Reaction Product Ethylene Glycol Diformate', *J. Phys. Chem. A*, **103**, 5032–5039 (1999).
23. F.G. Kitson, B.S. Larsen, C.N. McEwen, *Gas Chromatography and Mass Spectrometry – A Practical Guide*, Academic Press, San Diego, CA, 1996.
24. T.X. Lu, G.Z. Li, X.Z. Li, W.B. Sun, Q. Wu, 'The TRESR Study of the Naphthoquinone Radicals Photolytically Generated in TX-100 Micelle', *Acta Chim. Sin.*, **56**, 1041–1047 (1998).
25. V.R. Meyer, *Practical High Performance Liquid Chromatography*, John Wiley & Sons, Chichester, 1994.
26. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, John Wiley & Sons, New York, 1979.
27. B.F. Connor, D.L. Rose, M.C. Noriega, L.K. Murtaugh, S.R. Abney, *Methods of Analysis by the US Geological Survey National Water Quality Laboratory – Determination of 86 Volatile Organic Compounds in Water by Gas Chromatography/Mass Spectrometry, Including Detections Less Than Reporting Limits: US Geological Survey Open-File Report 97-829*, US Geological Survey, Denver, CO, 1998.
28. US Environmental Protection Agency, *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods (SW-846)*, US Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, DC, 1996.
29. W.T. Winberry, N.T. Murphy, R.M. Riggan, 'Method TO-1 Determination of Volatile Organic Compounds in Ambient Air Using Tenax Adsorption and Gas Chromatographic Analysis (GC/MS)', in *Compendium of Methods for the Analysis of Toxic Organic Compounds in Ambient Air*, Atmospheric Research and Exposure

Assessment Laboratory, Office of Research and Development, US Environmental Protection Agency, Research Triangle Park, NC, 1-1–1-34, 1988.

30. T.J. Kelly, R. Mukund, S.M. Gordon, M.J. Hays, *Ambient Measurement Methods and Properties of the 189 Title III Hazardous Air Pollutants*, EPA 600/R-94-098, Final Report. NTIS PB95-123923, US Government Printing Office, Washington, DC, 1994.
31. R. Mukund, T.J. Kelly, S.M. Gordon, M.J. Hays, W.A. McClenny, 'Status of Ambient Measurement Methods for Hazardous Air-pollutants', *Environ. Sci. Technol.*, **29**, 183A–187A (1995).
32. J.F. Pankow, W. Luo, L.M. Isabelle, D.A. Bender, R.J. Baker, 'Determination of a Wide Range of Volatile Organic Compounds in Ambient Air Using Multisorbent Adsorption/Thermal Desorption and Gas Chromatography Mass Spectrometry', *Anal. Chem.*, **70**, 5213–5221 (1998).
33. W.T. Winberry, N.T. Murphy, R.M. Riggan, 'Method TO9A – Method for the Determination of Polychlorinated Dibenzo-p-dioxins (PCDDs) in Ambient Air Using High Resolution Gas Chromatography/High-resolution Mass Spectrometry', in *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, Atmospheric Research and Exposure Assessment Laboratory, Office of Research and Development, US Environmental Protection Agency, Research Triangle Park, NC, 9A1–9A94, 1988.
34. S. Steenken, L. Goldbergerova, 'Photoionization of Organic Phosphates by 193 nm Laser Light in Aqueous Solution: Rapid Intramolecular H-transfer to the Primarily Formed Phosphate Radical. A Model for Ionization-induced Chain-breakage in DNA?', *J. Am. Chem. Soc.*, **120**, 3928–3934 (1998).

Trace Organic Analysis by Gas Chromatography with Selective Detectors

Erwin Rosenberg
Institute of Analytical Chemistry, Vienna University of Technology, Vienna, Austria

1 Introduction **1**
1.1 Classification 1
2 Selective Detectors **2**
2.1 Electron Capture Detector 2
2.2 Nitrogen–Phosphorus Detector 4
2.3 Flame Photometric Detector 5
2.4 Photoionization Detector 7
2.5 Electrolytic Conductivity Detector (Hall Detector) 9
2.6 Other Selective Detectors 12
3 Detectors Allowing Selective Recognition **12**
3.1 Gas Chromatography/Mass Spectrometry 13
3.2 Gas Chromatography/Fourier Transform Infrared Spectroscopy 16
3.3 Gas Chromatography/Atomic Emission Detection 17
3.4 Other Atomic Spectroscopic Detectors for Gas Chromatography 19
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

Selective detectors for gas chromatography (GC) allow one to detect selectively after gas chromatographic separation analytes that either contain certain elements or carry functional groups for which the detector is particularly selective (element-selective detectors). Spectroscopic detectors are a class of selective detectors which, in the ideal case, may even allow the identification of the analytes through acquisition of molecule-specific spectra. Selective detectors often allow the detection, quantitation or identification of analytes even in the presence of complex matrices and may thus significantly improve the selectivity of an analytical method.

1 INTRODUCTION

Gas chromatography (GC) is nowadays one of the most universal and important separation techniques in analytical chemistry. Numerous reasons account for this fact: GC has a very wide applicability, ranging from the separation of permanent gases to the determination of petroleum products of extremely high boiling point, exceeding 100 carbon atoms in the molecule, and even highly polar or very involatile compounds can be determined after suitable derivatization (e.g. polysaccharides, organic acids and amines). Additionally, the extremely high separation efficiency of GC, reaching several tens to hundreds of thousands of theoretical plates with modern capillary columns, exceeds that of any other separation technique.

The availability of highly sensitive and selective, or, in some cases, even specific detectors was a further important reason for the success of modern GC. A number of detectors are nowadays available that are selective, or even specific, for certain compound classes or (hetero-)elements contained in a molecule and that, in the case of molecule-specific detectors, even provide positive identification of the gas chromatographically separated compounds.[1–3]

It is therefore appropriate to say that, if a separation can be carried out by GC, no other technique needs to be attempted. This is illustrated by the fact that more than 80% of compounds on the priority pollutant lists of the European Community (EC) and the United States Environmental Protection Agency (USEPA) may be analyzed by GC.[4]

Since the basic detectors of GC have already been discussed in **Instrumentation of Gas Chromatography**, this article will only focus on selective detectors for GC.

1.1 Classification

The classification of GC detectors is usually undertaken on the basis of their operating principle and their selectivity. The following terms and concepts are common: detectors that respond to the concentration of the analyte in the gas vector (grams per cubic meter) are called *concentration-sensitive detectors*. On the other hand, detectors whose response is proportional to the amount of sample (or mass) reaching the detector per unit time (grams per second) are called *mass-flow detectors*.

A different way of classification is via the distinction between universal and selective or specific detectors. A *universal detector* responds to any component present in the mobile phase. A *selective detector* responds only to a more or less restricted group of compounds with similar

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Characteristics of common GC detectors

Detector/Method name	Type	Selectivity	Sensitivity	Linear range
FID	Universal	Responds to all compounds that are ionized in a flame	$10\,pg\,s^{-1}$ C	10^7
TCD	Universal	Responds to all compounds that have a different thermal conductivity to the carrier gas	$1\,ng\,mL^{-1}$ mobile phase	10^6
ECD	Selective	Compounds capturing electrons, e.g. halogens	$0.2\,pg\,s^{-1}$ Cl	10^4
NPD	Selective	N and P	$1\,pg\,s^{-1}$ N $1\,pg\,s^{-1}$ N	10^4
FPD	Selective	P and S	$50\,pg\,s^{-1}$ S $2\,pg\,s^{-1}$ P	10^3 10^4
PID	Selective	Aromatics	$5\,pg\,s^{-1}$ C	10^7
ELCD	Selective	Halogens and S	$1\,pg\,s^{-1}$ Cl $5\,pg\,s^{-1}$ S	10^6 10^4
FTIR spectrometry	Universal[a]	Molecular vibrations	5 ng (of a strong absorber)	10^3
MS	Universal[a]	Characteristic ions	1 ng in full-scan mode 1 pg in ion monitoring mode	10^5
AED	Universal[a]	Tunable to any element	$0.1–50\,pg\,s^{-1}$, depending on the element	

[a] Spectroscopic detectors can be operated in either universal or selective mode.
FID, flame ionization detector; TCD, thermal conductivity detector; ECD, electron capture detector; NPD, nitrogen–phosphorus detector; FPD, flame photometric detector; PID, photoionization detector; ELCD, electrolytic conductivity detector; FTIR, Fourier transform infrared; MS, mass spectrometry; AED, atomic emission detection.

properties. A *specific detector* ideally responds to a single compound (in the case of molecule specific detection) or at least to a limited set of compounds with similar chemical properties. A classification of the most common GC detectors is given in Table 1. It should be mentioned that the general use of the terms selective and specific is somewhat arbitrary and therefore confusing. In order to avoid any ambiguity in this context, selective is the preferred term to distinguish detectors from universal.

In Table 1, the detectors are further characterized by their selectivity, sensitivity and linear dynamic range (LDR). From a practical point of view, *selectivity* can be defined as the inverse concentration ratio of an analyte to which the detector responds selectively and a compound that is not detected with particular selectivity which produce the same signal at the detector. The *sensitivity* is the signal output per unit concentration of a particular analyte in the carrier gas (and in the case of a linear response equals the slope of the calibration graph). Together with the detector noise it determines the minimum detectable concentration of an analyte. A widely accepted value for the minimum detectable amount [= limit of detection (LOD)] is the amount of substance producing a signal-to-noise ratio (S/N) = 3. The *linear range* of the detector response is the part of the calibration graph in which there is a proportional increase of the signal with the analyte concentration in the mobile phase. It is usually a part of the *dynamic range* of the detector where the detector response increases with increasing analyte concentration, but this increase need not necessarily be proportional.

A further point to be considered in the classification of detectors is whether they are of destructive or non-destructive nature. In the latter case, multiple detectors may be used in series.

2 SELECTIVE DETECTORS

2.1 Electron Capture Detector

One of the most popular and useful detectors, particularly for environmental trace analysis, is the electron capture detector (ECD), which was developed in the early 1960s by Lovelock and Lipsky.[5,6] The ECD is a detector with a selectivity towards electrophilic compounds. Halogenated compounds, nitriles, nitrates and conjugated carbonyl compounds can be detected with high selectivity over hydrocarbons, alcohols, non-conjugated aldehydes or ketones. Since many anthropogenic environmental pollutants are halogenated [e.g. halogenated solvents, organochlorine pesticides, polychlorinated biphenyls (PCBs)], the ECD is a highly useful selective and sensitive detector.

Electron capture detection is based on the fact that electrophilic compounds are able to capture slow thermal electrons that are generated from the bombardment of the carrier or make-up gas by β-electron radiation (= fast

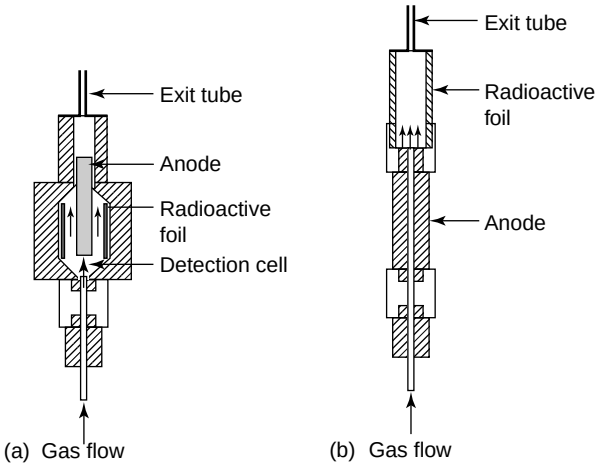

Figure 1 Schematic view of an ECD: (a) coaxial cylinder design and (b) the asymmetric (displaced coaxial cylinder) design.

electrons). This bombardment leads to the liberation of several hundred thermal electrons with low mean energies ($\leq$0.05 eV) per incident electron through repeated elastic and inelastic collisions. The electrons are collected by a counter electrode (anode) to which a positive potential is applied. When an electronegative compound enters the detection region between the β-emitter and the collector electrode, the detector background current, as observed in the presence of pure carrier gas, is reduced. This diminution in detector current is the basis of the quantitative information of the ECD. Typical ECD designs have a ^{63}Ni foil acting as a β-emitter and a collector electrode arranged concentrically. This design allows the reduction of the detector cell volume which is mandatory for use with capillary GC; however, the distance between the ^{63}Ni foil and the anode has to be at least of such dimensions that the β-electrons are completely deactivated by collisions with the carrier gas and transferred into thermal electrons before reaching the anode. This problem is circumvented when the collector electrode is located upstream of the β-emitter foil (Figure 1a and b). However, attention has to be devoted to the detector and particularly the ionization chamber design in order to minimize extracolumn band broadening. Most commercial ECDs have detection cell volumes of 100–400 µL, which still results in additional band broadening. This can be compensated for by the addition of make-up gas at the end of the chromatographic column to maintain column efficiency, but at the price of some loss in sensitivity due to dilution. The detector can be operated either continuously [by applying a direct current (DC) voltage] or discontinuously (in pulsed mode). Pulsed-mode operation is almost exclusively used nowadays, first to minimize space-charge and contact potential effects, and second to increase the LDR of the detector. Commercially available ECDs make use of a modified pulse-sampling technique, called the variable-frequency–constant-current mode. Instead of measuring the cell current at a constant pulse frequency, the cell current is kept constant by variation of the frequency of the pulse applied to the anode for the collection of thermal electrons. Since the frequency changes with the quantity of analyte detected, the detector signal in this case is a voltage proportional to that frequency rather than a current. This mode of operation offers an extended linear range as compared with constant-frequency detection but may exhibit some deviation from linearity in the case of molecules with ultrafast electron attachment rates (e.g. CCl_4, SF_6, CH_3I).

For electron capture detection, the use of nitrogen as carrier gas is a good choice, provided that it is absolutely free of oxygen and other contaminants. Pure He is unsuitable, since it gives rise to the formation of unwanted metastable ions that negatively affect the detector baseline stability. Thus, nitrogen or an argon–methane mixture (5–10% CH_4) is added as make-up gas for the detector when He is used as a carrier gas to maximize column efficiency.

The response of the ECD to organic compounds is highly variable. It is most sensitive for halogen and nitro compounds, and detects also organometallic compounds and conjugated electrophores (e.g. benzophenones, quinones, phthalate esters) (Figure 2). The response of the ECD to halogenated compounds decreases in the order I > Br > Cl $\gg$ F but increases over-proportionally with multiple substitution at the same carbon atom (cf.

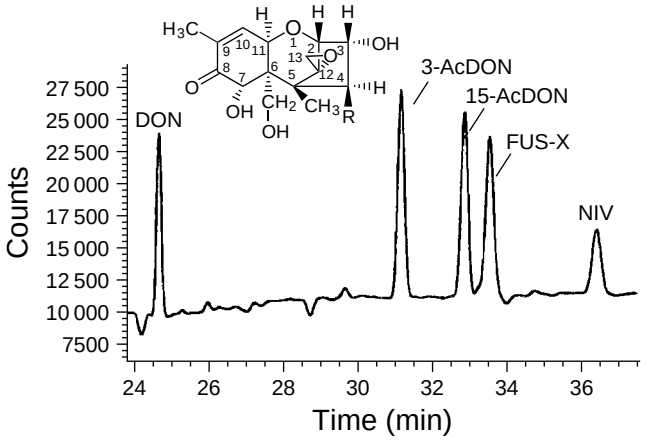

Figure 2 GC/ECD chromatogram of an SFE extract of wheat spiked with the five trichothecene mycotoxins deoxynivalenol (DON, R = OH), 3-acetyldeoxynivalenol (3-AcDON, R = OAc), 15-acetyldeoxynivalenol (15-AcDON; R = OH), fusarenon X (FUS-X; R = OH) and nivalenol (NIV; R = OH), which can be detected owing to the conjugated carbonyl group. [Reproduced by permission of Elsevier Science from R.D. Josephs, R. Krska, M. Grasserbauer, J.A.C. Broekaert, *J. Chromatogr. A*, **795**, 297–304 (1998).]

Table 2 Relative response of the ECD to different organic compounds. [Reproduced by permission of Elsevier Science Publishers from C.F. Poole, S.K. Poole, *Chromatography Today*, Elsevier, Amsterdam, 278, 1991.]

General organic compounds		Halocarbons	
Compound	Relative response	Compound	Relative response
Benzene	0.06	$CF_3CF_2CF_3$	1.0
Acetone	0.50	CF_3Cl	3.3
Di-*n*-butyl ether	0.60	$CF_2{=}CFCl$	100
Methyl butyrate	0.90	$CF_3CF_2CF_2Cl$	170
1-Butanol	1.00	$CF_2{=}CCl_2$	670
1-Chlorobutane	1.00	CF_2Cl_2	3×10^4
1,4-Dichlorobutane	15	$CHCl_3$	3.3×10^4
Chlorobenzene	75	$CHCl{=}CCl_2$	6.7×10^4
1,1-Dichlorobutane	111	CF_3Br	8.7×10^4
1-Bromobutane	280	$CF_2ClCFCl_2$	1.6×10^5
Bromobenzene	450	$CF_3CHClBr$	4.0×10^5
Chloroform	6×10^4	$CF_3CF_2CF_2I$	6.0×10^5
1-Iodobutane	9×10^4	$CF_2BrCF_2CF_2Br$	7.7×10^5
Carbon tetrachloride	4×10^5	$CFCl_3$	1.2×10^6

Table 2). The ECD is most frequently used for the detection of halogenated compounds (Figure 3). Through common derivatization reactions (e.g. acylation with perhalogenated acid anhydrides or acyl chlorides) polar, nonhalogenated compounds may be made accessible to both GC and highly sensitive ECD.

2.2 Nitrogen–Phosphorus Detector

The nitrogen–phosphorus detector (NPD), often also called a thermionic detector, was developed by Kolb and Bischoff in 1974[7] from a predecessor known as the alkali flame ionization detector (AFID).[8] Its construction is very similar to that of a FID. Selectivity for nitrogen- and phosphorus-containing molecules is achieved by the introduction of an alkali metal salt glass above the flame of the detector. In the presence of this thermionic source, molecules containing nitrogen or phosphorus are ionized with much higher yield than hydrocarbons. The ions formed are collected by a counter electrode and the resulting current is measured. The NPD is extremely valuable in both clinical and forensic chemistry where metabolites and drugs of abuse are determined, and also in environmental analysis where it has proven useful for the determination of organophosphorus- and nitrogen-containing pesticides.

The initial design of the NPD by Kolb and Bischoff contains an electrically heated rubidium silicate bead, situated a few millimeters above the detector jet tip and below the collector electrode (Figure 4). The bead is maintained at a negative potential which increases both the stability of the rubidium salt bead and the detector response. The source is operated at a temperature of 600–800 °C by an independent power supply. The flow of hydrogen in the nitrogen-selective mode is too low to support a flame (1–5 mL min^{-1}). Thus, only a plasma is sustained in the region of the alkali metal salt bead by the hydrogen and air flow. For phosphorus-selective detection, a higher flow of ca. 30 mL min^{-1} of hydrogen is used. To suppress the nonselective response by hydrocarbons in the flame mode, the detector jet is grounded and the negative potential of the alkali metal salt bead deflects electrons to ground and away from the collector electrode. In the phosphorus-selective mode, phosphorus-containing fragments are only formed in the region of the alkali metal glass bead, and only

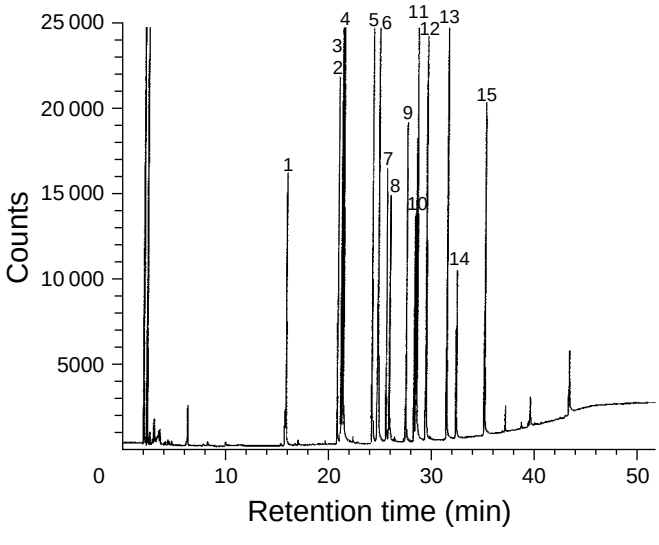

Figure 3 Chromatogram of a mixture of chlorinated phenols derivatized with pentafluorobenzyl bromide. [Reproduced by permission of GIT Verlag from B. Scholz, H. Tretter, in *Buch der Umweltanalytik*, GIT Verlag, Darmstadt, 114, Vol. 4, 1992.]

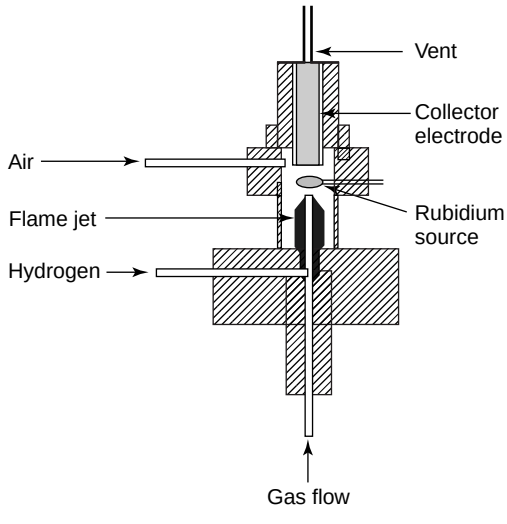

Figure 4 Schematic view of a NPD.

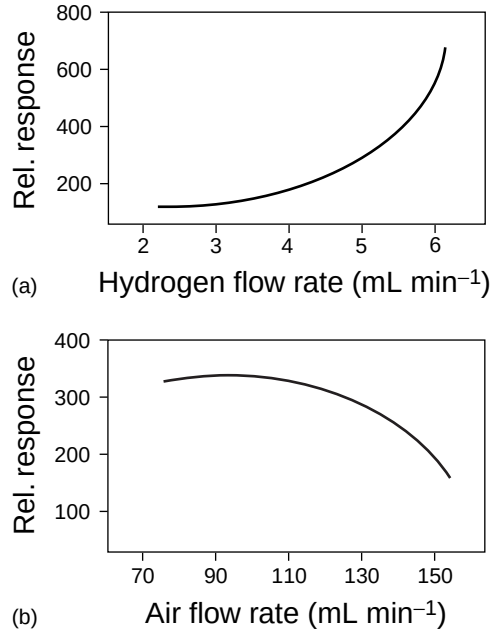

Figure 5 Dependence of the NPD response on the (a) hydrogen and (b) air flow rate, as observed for the pesticide malathion.

electrons generated by this process are able to reach the collector electrode. Operating the detector in the flame mode diminishes its response to nitrogen-containing compounds by more than a factor of 10 as compared with the flameless mode, in which the detector is more selective to phosphorus-containing compounds. Both the selectivity and the sensitivity of the detector depend greatly on the optimization of the operational variables such as hydrogen and air flow rates (Figure 5a and b), bead heating current (temperature), jet potential and choice of carrier gas. The theory of operation of the NPD is not yet completely understood. The presence of rubidium in the alkali metal salt bead seems not to be essential for selectivity, but for a stable response of the detector. Hence the assumption of a surface-catalyzed reaction rather than the evaporation of alkali metal ions from the glass bead seems more appropriate. In any case, negative ions are formed from either nitrogen- and phosphorus-containing analytes (cyanide ions, CN^-, in the former and phosphorus oxoanions, PO^-, in the latter case) which are collected and cause an increase in ion current at the counter electrode.

The high sensitivity of the NPD (allowing the detection of 0.5–10 pg of an N-containing compound and 0.1–1 pg of a P-containing compound), together with a LDR of ca. 10^4 and a selectivity of 10^4–10^5, make it the most useful detector for P- and N-detection. It is comparatively robust and easy to maintain, but the detector performance (sensitivity) tends to degrade after several hundred hours of operation, probably owing to ion-exchange mechanisms at the surface of the alkali metal salt bead. It is therefore good practice to turn off the H_2/air reagent gases and to lower the temperature of the detector when it is not in use for longer periods. For the same reason, a large excess of chlorinated solvents or silylation reagents should be removed prior to injection. Although one is tempted to analyze crude extracts with little sample preparation or cleanup by GC/NPD owing to the excellent selectivity of the detector, this is generally not encouraged owing to the susceptibility of the NPD to contamination. It is also important to mention that not all nitrogen- or phosphorus-containing compounds are detected with equal sensitivity. It appears that a nitrogen–carbon bond is essential for a high response of the NPD in the nitrogen-selective mode, since, for example, esters of nitric acid (e.g. nitroglycerine) are hardly detected.

An example of the application of NPD for the monitoring of drugs of abuse is given in Figure 6.

2.3 Flame Photometric Detector

The flame photometric detector (FPD) is the most widely used detector for the analysis of sulfur- and phosphorus-containing substances in biochemistry, food and environmental analysis. It is based on the observation of element-specific emission of light from sulfur- and phosphorus-containing compounds in a hydrogen-rich flame. Brody and Chaney employed this principle first in 1966 for the construction of the FPD.[9] Since then, this detection principle has been extended to a number of other elements, including tin, germanium, selenium, tellurium, chromium and boron.[10,11]

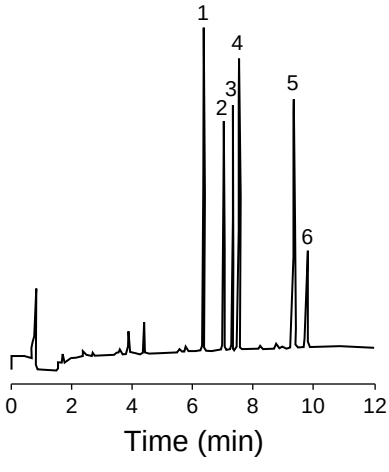

Figure 6 NPD chromatogram of a pesticide sample containing 1, diazinon; 2, methylparathion; 3, malathion; 4, parathion; 5, ethion; and 6, carbophenothion. Splitless injection of 1 ng per component. Column: 12 m × 0.2 mm inner diameter (ID), 0.2 μm film thickness HP 101. [Courtesy of Hewlett-Packard, Palo Alto, CA, USA.]

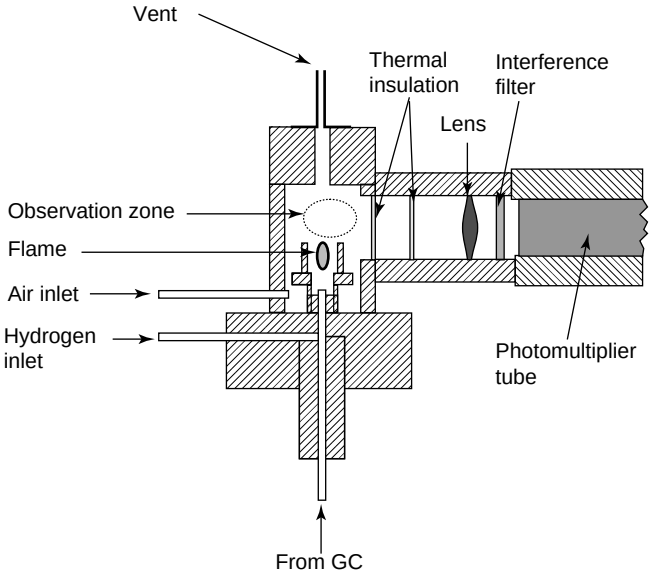

Figure 7 Schematic diagram of the (single) FPD.

Flame photometric detection exploits the fact that sulfur- and phosphorus-containing compounds form excited chemiluminescent species in a hydrogen-rich flame. The excited molecules emit light of characteristic wavelength on relaxation. This radiation passes through an optical filter that transmits only light of characteristic wavelength and is detected by a photomultiplier tube where the photons are converted into an electrical signal.

A typical design of an FPD can be seen in Figure 7. The end of the capillary column is inserted into the base of the FPD jet where the carrier gas is mixed with air or oxygen and combusted in a hydrogen-rich atmosphere at the burner tip. Interfering emissions from hydrocarbons occur mainly directly in the oxygen-rich flame region close to the burner tip, whereas sulfur and phosphorus emissions are more pronounced in the hydrogen-rich upper portion of the flame. Selectivity is usually enhanced by shielding the lower part of the flame by an opaque ring so that the hydrocarbon emission is prevented from reaching the photomultiplier. The extent, intensity and selectivity of the various emitting regions are highly dependent on the burner design and the individual gas flow rates, that is, not only the total gas flow but also the hydrogen/oxygen (or air) ratio, and the detector temperature. Different conditions are found to be optimal for sulfur and phosphorus detection. Sulfur detection is usually more sensitive at lower flow rates of both hydrogen (ca. 50–70 mL min^{-1}) and air (ca. 100 mL min^{-1}), whereas phosphorus requires both higher hydrogen and oxygen flow rates (both at ca. 200 mL min^{-1}). Selectivity of the detector is established by using a 392-nm band-pass filter for sulfur and a 526-nm band-pass filter for phosphorus detection where

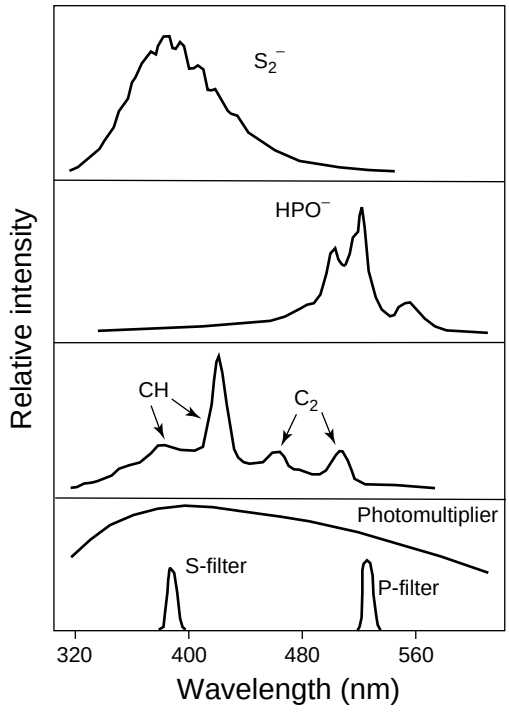

Figure 8 Emission spectra of sulfur, phosphorus and hydrocarbons and relative response function of the photomultiplier and of sulfur- and phosphorus-selective filters.

both elements have their emission maxima, respectively (Figure 8).

The theory of flame photometric detection is known only in outline. It is generally accepted that excited species are formed by several collision reactions in the flame, such as those shown in Equations (1–3):

$$H + H + S_2 \longrightarrow S_2^* + H_2 \quad (1)$$

$$S + S \longrightarrow S_2^* \quad (2)$$

$$S + S + M \longrightarrow S_2^* + M \quad (3)$$

where M is some third molecule. In the case of phosphorus, the phosphorus compounds are first decomposed to PO molecules in the flame, which subsequently are converted into HPO^* species by reactions similar to those in Equations (4) and (5):

$$PO + OH + H_2 \longrightarrow HPO^* + H_2 \quad (4)$$

$$PO + H + M \longrightarrow HPO^* + M \quad (5)$$

It follows from the above equations that a linear dependence can be expected between the amount of phosphorus-containing compounds and the detector response. For sulfur-containing compounds, the dependence of the signal rather follows Equation (6):

$$I = A[S]^n \quad (6)$$

where I is the signal intensity, A an experimental constant, [S] the mass flow rate of sulfur atoms and n an exponential factor. Theoretically, n should assume a value of 2, but in practice values between 1.6 and 2.2. are observed. Various reasons for this deviation are known and include incomplete combustion, signal quenching by both high concentrations of co-eluting hydrocarbons or self-quenching at high analyte concentrations, competing flame reactions that also lead to de-excitation and finally structural effects.

The sensitivity of the FPD is greatly dependent on the operating conditions but is in the 5–50 pg range for sulfur-containing compounds and in the 0.5–5 pg range for phosphorus-containing compounds. The dynamic range is about 10^3 in both modes; however, as explained above, only phosphorus yields a linear response. Selectivity for phosphorus may exceed 10^4 whereas it varies greatly for sulfur compounds and is also strongly dependent on the concentration range. It is not always possible to reach this selectivity (particularly for sulfur-selective detection) because of signal quenching: co-eluting hydrocarbons can quench the detector response owing to a transient change in flame conditions or to de-excitation reactions in the emitting zone of the flame. In this respect, the dual-flame design brings a significant improvement; the typical layout of a dual-flame detector is shown in Figure 9. It consists of a two-burner design where in the first flame the analytes are burnt in a hydrogen-rich flame. The unreacted excess of hydrogen is then mixed again with air to produce a second flame which serves to excite the sulfur or phosphorus species. This two-burner design with one flame for the decomposition and combustion of the analytes eluting from the GC column and one flame for the formation of emitting species largely overcomes the quenching problems discussed above and also the problem of 'solvent flameout', that is, the extinction of the FPD flame by large amounts of the solvent.

Applications in which the FPD has proven to be of particular use are in the petroleum industry (detection of sulfur compounds; Figure 10), for trace sulfur compound analysis and for the detection of organophosphorus pesticides.

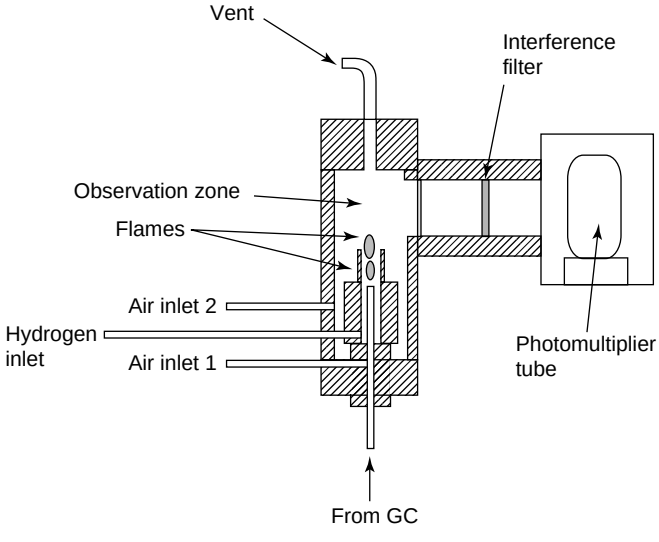

Figure 9 Schematic diagram of the dual FPD.

2.4 Photoionization Detector

The use of far-ultraviolet (UV) ionization (photoionization) detection evolved from initial experiments in the 1960s[12,13] which only eventually led to a commercial instrument available in the mid-1970s, as described by Driscoll and Spaziani[14] after solving problems related to the detector design. The PID is based on the absorption of energy by a molecule which results in electronic excitation from a (discrete) low energy level of the molecule to the higher energy continuum of the ion. The energy of such a transition is in the 5–20 eV range, and thus requires photons in the far-UV range. The ions formed are collected and detected by a collector electrode.

Early detector designs required operation at reduced pressure (typically 1–10 Torr) to maximize the photon intensity, which was not very practical. It was therefore a major breakthrough in the development of the PID when the source and the detection chamber were separated physically to allow independent optimization of both the ion formation and ion collection processes. This feature was also found to improve the stability of the PID, to increase its linear range (from ca. 10^4 to about 10^7) and to lower the background. Still, it

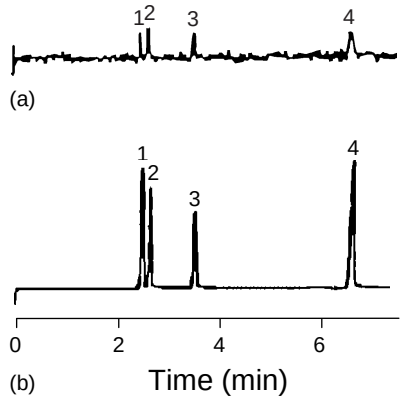

Figure 10 FPD chromatograms of trace sulfur gases at ca. 1 ppm in nitrogen: (1) H_2S; (2) COS; (3) CH_3SH; (4) CS_2. (a) 5-μL on-column injection (10 pg CH_3SH); (b) 100-μL on-column injection (200 pg CH_3SH). FPD with 393-nm filter; column, 100% polydimethylsiloxane, 30 m × 0.53 mm ID, 2.85 μm film thickness; carrier gas, He, 2.75 mL min^{-1}; makeup gas, N_2, 20 mL min^{-1}; H_2, 50 mL min^{-1}; air, 60 mL min^{-1}. [Courtesy of Hewlett-Packard, Palo Alto, CA, USA.]

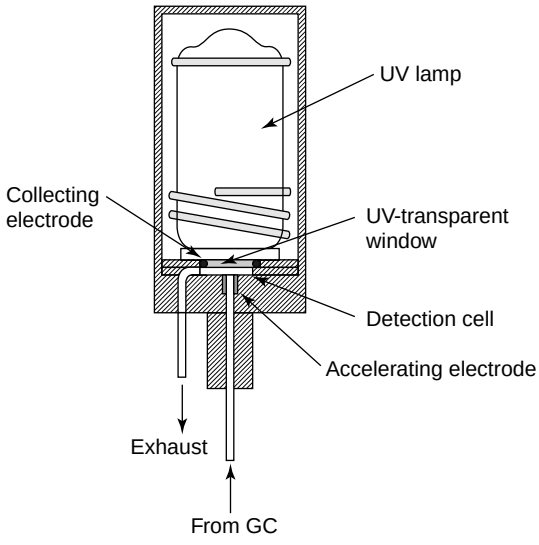

Figure 11 Schematic diagram of a PID.

has to be borne in mind that the ionization efficiency is only about 0.001–0.1%, depending on the analyte. The PID is thus a virtually nondestructive detector and particularly useful for applications where nondestructive detection is required and/or in combination with a second detector. Additionally, it does not require any reagent gases and usually also no makeup gas, which makes it ideally suited for small, portable instruments. The detector selectivity can be tuned to a certain extent by changing the excitation source and thus the wavelength (energy) of exciting radiation. As a drawback of the PID, its compound-dependent response, requiring individual calibration for each analyte, has to be mentioned. A generic design of the PID is shown in Figure 11. A typical PID has two functional units: an excitation source and an ionization chamber. The former consists of a UV source which in most cases is a discharge lamp, filled with an inert gas or gas mixture at reduced pressure that emits monochromatic light of a characteristic wavelength or energy, depending on the choice of the fill gas (e.g. Ar, Xe, Kr) and the window material (e.g. LiF, MgF_2, CaF_2 or SrF_2). Sources of different energies are commercially available (with, e.g. 9.5, 10.0, 10.2, 10.9 and 11.7 eV), but the 10.2 eV source is the most commonly employed. The wavelength (in nanometers) can be calculated from the source energy (in electronvolts) from Equation (7):

$$\lambda(\text{nm}) = \frac{1234.5}{E}(\text{eV}) \tag{7}$$

Thus, the wavelength emitted from a 10.2 eV source is ca. 121 nm.

The discharge compartment is mechanically separated form the ionization chamber by a metal fluoride window which is optically transparent in the far-UV range. The effluent from the GC column passes through the thermostated ionization chamber and between the two electrodes which are positioned on the opposite ends of the ionization chamber. Between these, an electrical field is applied to accelerate and to collect the ions that are formed by photoionization. The current is amplified by an electrometer and read out. Typical ionization chamber volumes are between 40 and 250 μL. The smaller detection cell volumes can still be used with capillary columns without causing unacceptable peak broadening, but the use of a makeup gas to reduce peak broadening or operation at higher column flows ($>2\,\text{mL}\,\text{min}^{-1}$) is recommended. Detector cell volumes between 150 and 250 μL are preferably used with packed columns. Precise thermostating of the detector can minimize baseline drift.

The process of photoionization can be described by a number of reactions in which the analyte A and the carrier gas molecules C interact with each other and with photons of energy $h\nu$. Ionization can occur either *directly* [Equation 8]:

$$\text{A} + h\nu \longrightarrow \text{A}^+ + \text{e}^- \tag{8}$$

or *indirectly* [Equations 9–12]:

$$\text{C} + h\nu \longrightarrow \text{C}^* \tag{9}$$

$$\text{C}^* + \text{A} \longrightarrow \text{A}^+ + \text{e}^- + \text{C} \tag{10}$$

$$\text{A} + h\nu \longrightarrow \text{A}^* \tag{11}$$

$$\text{A}^* \longrightarrow \text{A}'^+ + \text{e}^- \tag{12}$$

(the last two equations indicate that in addition to the detection of ions formed from the original molecule, R^+, detection of ions that result from rearrangement or elimination reactions, R'^+, is also possible). The processes in Equations (13–15) compete with the photoionization of the analytes (*quenching reactions*):

$$\text{A}^+ + \text{e}^- \longrightarrow \text{A} \tag{13}$$

$$\text{e}^- + \text{C} \longrightarrow \text{C}^- \tag{14}$$

$$\text{A}^+ + \text{C}^- \longrightarrow \text{A} + \text{C} \tag{15}$$

It can be understood from the above equations that the detector response depends to a certain degree on the presence of electron-capturing impurities in the carrier gas (e.g. oxygen) and also the choice of carrier gas, which, through collisional de-excitation, can influence the sensitivity.

When using a 10.2 eV photon source, most organic molecules are ionized. Exceptions are the permanent gases, C_1–C_4 hydrocarbons, methanol and acetonitrile (which, therefore, makes them suitable solvents for GC analysis, cf. Table 3) and the chloromethanes. For aromatic compounds, the sensitivity of the PID can be up

Table 3 Ionization potential of some common GC solvents

Solvent	Ionization potential (eV)
Methyl ethyl ketone	9.53
Acetone	9.69
n-Hexane	10.18
n-Pentane	10.35
Methanol	10.85
Dichloromethane	11.35
Chloroform	11.42
Acetonitrile	12.22
Water	12.35

Table 4 Relationship between PID response (with 10.2-eV excitation) and molecular structure

Sensitivity increases
• as the carbon number increases
• for alkanes < alkenes < aromatics
• for alkanes < alcohols < esters < aldehydes < ketones
• for noncyclic compounds < cyclic compounds
• for nonbranched compounds < branched compounds
• for fluorine-substituted < chlorine-substituted < bromine-substituted < iodine-substituted compounds
• for substituted benzenes with ring activators (= electron-releasing groups) and decreases with ring deactivators (= electron-withdrawing groups); exception: halogenated benzenes

to 50 times higher than that of the FID, and for aliphatic compounds, a gain in sensitivity of 5–10-fold over the FID can still be realized. There are collections of data on the response of a great number of compounds relative to benzene,[15] which allows several empirical conclusions to be drawn on the basis of this and other data. Generally, an increase in sensitivity is observed as the carbon number increases. The other relationships are summarized in Table 4.

As an example, the analysis of gasoline components using PID and the different selectivities that can be obtained at different lamp energies are shown in Figure 12(a–d).

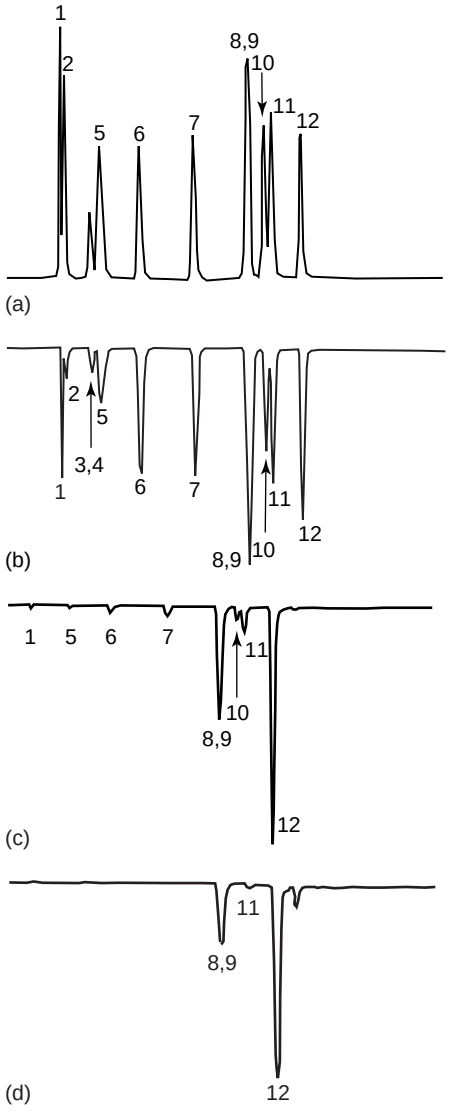

Figure 12 Comparison of the chromatogram of gasoline components obtained by (a) FID and (b–d) PID at different excitation energies: (b) 10.2, (c) 9.5 and (d) 8.3 eV. Peaks: 1 = 1-hexene; 2 = 2,4,4-trimethylpentane; 3 = 1-octene; 4 = *trans*-1,4dimethylcyclohexane; 5 = *cis*-1,4-dimethylcyclohexane; 6 = benzene; 7 = toluene; 8 = *m*-xylene; 9 = *p*-xylene; 10 = cumene; 11 = *o*xylene; 12 = mesitylene. [Redrawn from J.N. Driscoll, in *Detectors for Capillary Chromatography*, Chemical Analysis, eds. H.H. Hill, D.G. McMinn, John Wiley & Sons, New York, 64, Vol. 121, 1992.]

2.5 Electrolytic Conductivity Detector (Hall Detector)

The electrolytic conductivity detector (ELCD) for GC was first reported by Piringer and Pascalau.[16] The ELCD has particularly been utilized for the selective detection of sulfur, halogen and nitrogen compounds. The commercial introduction of this detector was in the mid-1960s, following further development by Coulson.[17]

The basis of this detector is the catalytic combustion of the analytes in a furnace at high temperatures (ca. 900 °C) with the addition of reagent gases. This reaction converts, e.g. halogenated compounds into halo acids which are swept from the reactor into a gas–liquid contactor where they are dissolved in an appropriate solvent (e.g. 1-propanol). The strong acids increase the electrolytic conductivity of the solvent which is then passed through a conductivity cell where its conductivity is measured (see the block diagram in Figure 13).

A number of improvements were introduced to this detector by Hall,[18,19] which led to the design shown

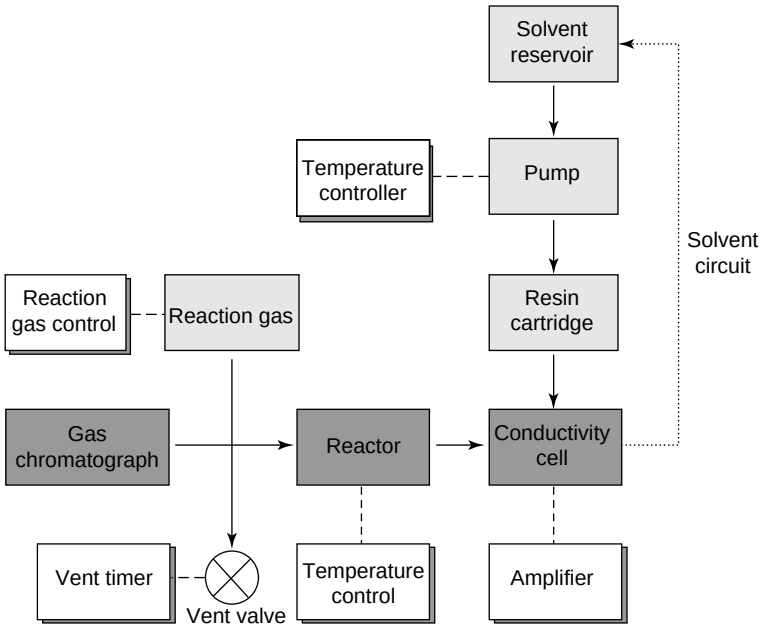

Figure 13 Block diagram of the components of the ELCD.

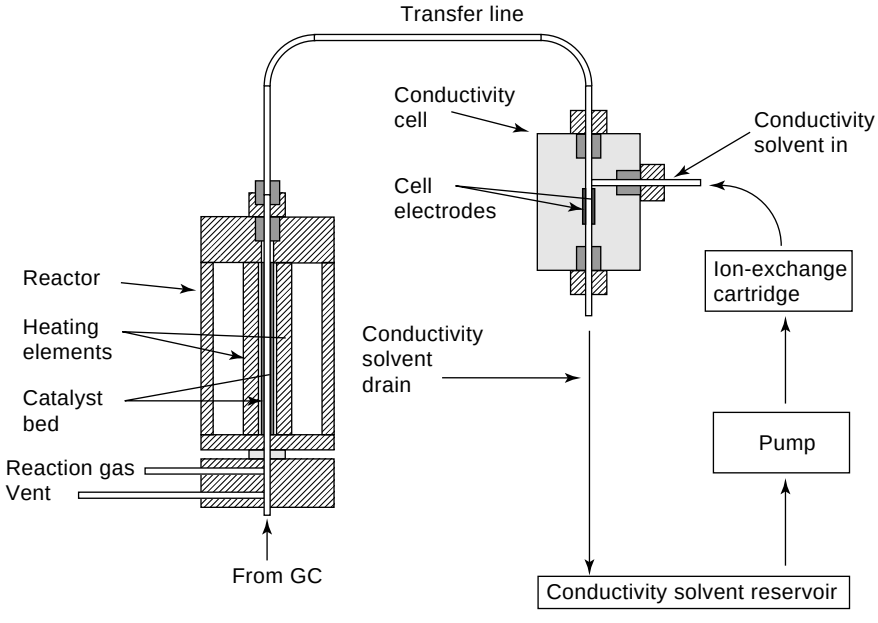

Figure 14 Schematic diagram of the ELCD. [Adapted from J.N. Driscoll, in *Detectors for Capillary Chromatography*, Chemical Analysis, eds. H.H. Hill, D.G. McMinn, John Wiley & Sons, New York, 112, Vol. 121, 1992.]

in Figure 14. Following the reaction cell where the catalytic conversion of the analytes to the detectable species takes place, these are led into a small-volume, concentric cylinder cell for mixing, separation and detection of the detectable reaction products. The solvent and the gaseous reaction products are merged in a small polytetrafluoroethylene (PTFE) tee. The initially heterogeneous gas–liquid mixture forms a smoothly flowing liquid phase which passes by the electrodes of the conductivity cell. The electrolyte stream is driven downwards by the circulation of the collecting solvent through a pump and the pressure that is executed by the carrier gas. Different ways of measuring the conductivity exist of which the differential detector design is the most useful. It employs two conductivity cells in series where the first cell measures the conductivity of

Table 5 Reaction products produced in the catalytic decomposition in the ECD. [Reproduced with permission of John Wiley & Sons, Inc., from *Detectors for Capillary Chromatography*, Chemical Analysis, eds. H.H. Hill, D.G. McMinn, John Wiley & Sons, New York, 113, Vol. 121, 1992]

Compound	Main product(s)	Comments
(a) Under reductive conditions at 850–1000 °C using a nickel reaction tube		
Halogen-containing compounds	HX	HX can be removed by N-mode scrubber and be selectively detected in X mode
Sulfur-containing compounds	H_2S	H_2S can be removed by N-mode scrubber and is only hardly ionized in X mode
Nitrogen-containing compounds	NH_3	NH_3 is selectively detected in N mode and is hardly ionized in X mode
Alkanes	CH_4 and lower alkanes	Products are not detected in any mode
Oxygen-containing compounds	H_2O	H_2O gives little response in X and N modes
(b) Under oxidative conditions at 850–1000 °C using a nickel or alumina reaction tube		
Halogen-containing compounds	HX	HX can be removed by S-mode scrubber
Sulfur-containing compounds	H_2S	SO_2 can selectively be detected in S mode
Nitrogen-containing compounds	N_2 and nitrogen oxides	No response from N_2, little response from nitrous oxides (very low solubility in collecting solvent)
Alkanes	CO_2	CO_2 is hardly ionized in nonaqueous solvents

the solvent while the second measures the conductivity of the solvent plus the reaction products. Measuring the difference of the two conductivities eliminates to a great extent response variations caused by temperature differences and changes/drift in the concentration of the solvent. The solvents mostly employed nowadays are 1-propanol and methanol, either alone or in admixtures with water. Pure water is hardly any longer used as solvent. The lower alcohols, being themselves very weak acids, are used to suppress the dissociation of weak organic acids (e.g. carboxylic acids, phenols) and thus to render the detector selective for the strong inorganic acids formed by the catalytic conversion of the analytes. Furthermore, 1-propanol exhibits a very low solubility for carbon dioxide from the catalytic decomposition so that the detector does not respond to carbon. Most detectors recirculate the conductivity solvent through a closed-loop system. In order to achieve a low background conductivity and to regenerate the solvent, an ion-exchange resin bed is used. The selectivity of the detection can be increased by using an element-specific potentiometric electrode instead of the standard conductivity cell.

The ELCD can be operated in three principal detection modes with the detected species as given:

- halogen mode, HX (where X is usually Cl);
- sulfur mode, SO_2;
- nitrogen mode, NH_3.

The selectivity of the ELCD is controlled by the operating conditions of the detector. When, for example, halogen compounds are to be detected, a nickel catalyst, hydrogen reaction gas, a reactor temperature of 850–1000 °C and 1-propanol are used. Under similar conditions, NH_3 is formed from nitrogen-containing compounds (see Table 5).

The most relevant practical limitations of the ELCD are its poor linearity and peak shape. Poor linearity in most cases is a consequence of neutralization of the conducting solvent when, e.g. its capacity is exhausted. Neutralization problems can be recognized by the typical negative dips before and after the elution of a peak. Peak tailing can often be traced back to contamination of the scrubber or the transfer line, deactivation of the catalyst or the presence of interfering reaction products that are not removed by the scrubber. Further factors affecting detector stability and response are the absolute flow rate of the conductivity solvent and its stability. Practical operation conditions usually represent a compromise since low solvent flow rates increase sensitivity but at the same time increase noise due to pulsation of the solvent delivery system.

The ELCD offers high selectivity and sensitivity under optimized conditions. Detection limits down to 1 pg of chlorine, nitrogen and sulfur can be reached with a dynamic range of 10^3–10^5. The selectivity of the detector varies between 10^4 and 10^9 depending on the heteroelement to be detected and the operating conditions. Compared with the FPD in the sulfur mode, the ELCD has similar or greater sensitivity and selectivity, and compared with the ECD, the ELCD not only has a significantly greater selectivity for chlorine-containing compounds, but also a more uniform response on a molar basis (Figure 15a and b).

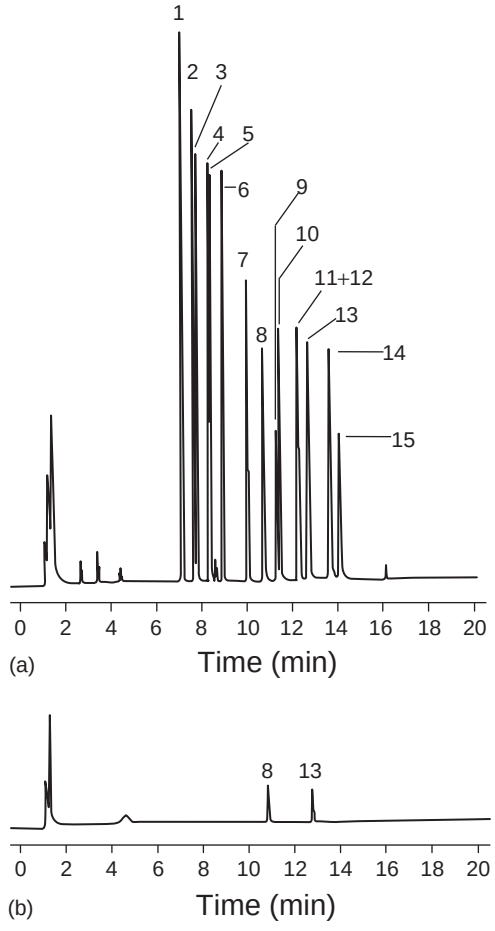

Figure 15 Operation of the ELCD in the sulfur mode (a) without and (b) with a scrubber. ELCD conditions: electrolyte, 40 μL min^{-1} 95% MeOH; reaction tube, 0.030 in ID alumina; reaction gas, 25 mL min^{-1} air; reactor temperature, 950 °C; oven program, 65 °C (1 min), 30 °C min^{-1} to 200 °C, 10 °C min^{-1} to 290 °C (4 min). Peak numbers denote the following chlorinated pesticides: 1, α-BHC; 2, lindane; 3, β-BHC; 4, heptachlor; 5, δ-BHC; 6, aldrin; 7, heptachlor epoxide; 8, endosulfan I; 9, *p*,*p*′-DDE; 10, dieldrin; 11, endrin; 12, *p*,*p*′-DDD; 13, endosulfan II; 14, *p*,*p*′-DDT; 15, endrin aldehyde (only 8 and 13 contain sulfur). [Reproduced by permission of John Wiley & Sons, Ltd., from *Detectors for Capillary Chromatography*, Chemical Analysis, eds. H.H. Hill, D.G. McMinn, John Wiley & Sons, New York, 131, Vol. 121, 1992.]

2.6 Other Selective Detectors

There are a number of other selective detectors that permit element- or compound class-selective detection, but have not been included owing to their less widespread use. These include the sulfur chemiluminescence detector (SCD),[20] which is one of the most selective and sensitive sulfur detectors, the thermal energy analyzer (TEA),[21] designed for the ultratrace determination of nitrosamines, and also based on chemiluminescence detection after low-temperature pyrolysis of the analytes, and the oxygen-sensitive flame ionization detector (O-FID), in which selectivity for oxygen is achieved by selective formation of CO only from oxygen-containing molecules in a first cracking step and reduction of the CO formed to methane which is detected by a conventional FID.[22] A further interesting, although hardly used, detector is the far-UV absorbance detector that can detect all compounds absorbing far-UV radiation in the 120–145 nm range.[23]

3 DETECTORS ALLOWING SELECTIVE RECOGNITION

The detectors discussed so far still have, despite their selectivity, two shortcomings for practical applications: their selectivity is limited in most cases, and they do not provide specific and sufficient information to elucidate the identity of a chromatographically separated compound. This information can, under optimal conditions, be provided by spectroscopic detectors that allow the selective recognition and identification of the separated compounds. The combination of (gas) chromatographic separation and spectroscopic detection is usually referred to as a hyphenated technique. Gas chromatography/mass spectrometry (GC/MS), gas chromatography/Fourier transform infrared (GC/FTIR) spectrometry and gas chromatography/atomic emission detection (GC/AED) are nowadays the most common and most powerful hyphenated techniques used for the identification and structural elucidation of unknown compounds. They are available now as easy-to-use benchtop instruments which constitute the basis for their acceptance in the routine laboratory. The ease of use of these hyphenated techniques is to a great extent due to the development of capillary column chromatography, which is nowadays almost exclusively used. The use of capillary columns with their significantly reduced carrier gas flow rates in many cases eliminates the need for a dedicated interface (with all related problems and shortcomings) and allows a direct connection of the capillary column outlet to the spectrometer, even if this is operating under vacuum as is the case for GC/MS.

The hyphenated techniques (GC/MS, GC/FTIR, GC/AED) are very often used as stand-alone systems. Even more powerful is the combination of two hyphenated techniques [ideally performed on one chromatograph[24]] which is done occasionally and increases the information content of the techniques by combining the data from the two spectroscopic techniques which allows one to confirm or to exclude suspected analytes.[25] Owing to the nondestructive nature of FTIR detection, GC/FTIR may be combined with MS detection in either the serial or parallel mode. When FTIR and MS detection are operated in parallel mode, the column effluent has to be split between the two detectors.

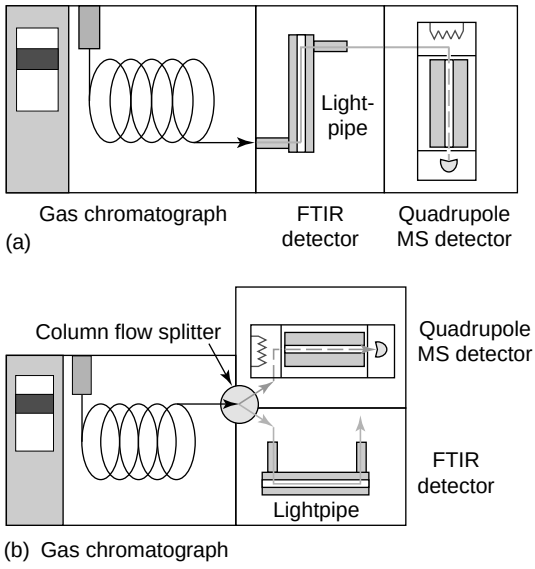

Figure 16 Diagram of a capillary GC/FTIR/MS system coupled (a) in series and (b) in parallel mode.

Owing to the lower sensitivity of FTIR detection compared with MS detection, the split ratio should favor the FTIR spectrometer, e.g. 10:1 (Figure 16a and b). Similarly, the two destructive techniques AED and MS may only be operated in parallel mode; however, owing to the excellent sensitivity of AED for most relevant elements, the split ratio can be close to unity. The properties and characteristics of the individual spectroscopic detectors are discussed in the following sections.

3.1 Gas Chromatography/Mass Spectrometry

The coupling of GC and MS was first described in the mid-1950s. The commercial success of this technique, however, only came with the introduction of capillary column GC in the early 1970s that eliminated the need for a special interface and brought a significant reduction in the pumping capacity necessary to meet the high vacuum requirements for MS detection. At the same time mass analyzers, based on either the quadrupole mass filter or the ion trap principle, became available and allowed the construction of compact, benchtop instruments.[26]

The principle of GC/MS will be explained based on the quadrupole mass filter serving as an example. The sample molecules exit from the GC column and are usually led through a heated transfer line from the gas chromatograph directly into the ionization chamber (ion source) of the mass spectrometer. In the ion source, both molecular and fragment ions are formed and extracted and accelerated into the mass analyzer through the application of a high voltage. In the mass analyzer, the ions are separated according to their mass-to-charge ratio (m/z). The separated ions are detected (after conversion into photoelectrons) by an electron multiplier and the resulting mass spectrum is recorded as a line spectrum of the ion intensity (y-axis) against their m/z ratio (x-axis).

Two different ionization techniques are commonly employed in MS detection for GC: electron impact (EI) and chemical ionization (CI). The most frequently used technique is EI where the molecules are bombarded with electrons with a kinetic energy of 70 eV emitted from a heated rhenium or tungsten filament. The collision of an electron with the molecule forms a radical ion by the expulsion of an electron ($M^{\bullet+}$). This radical ion has an excess energy of about 60 eV, which is enough to dissociate further into fragment ions, radicals and neutral species, thereby producing a characteristic fragment ion pattern. As a rule of thumb, the most labile bonds in the molecule are cleaved and those fragment ions occur with highest abundance that are the most stable. Owing to the low pressure in the ion source region (ca. 10^{-5} Torr $= 1.33 \times 10^{-3}$ Pa), association between molecules and fragment ions does not occur.

Many substance classes, e.g. alkanes, fatty acids and steroids, undergo extensive fragmentation. This can be to such an extent that the molecular ion peak is detected at only very low sensitivity or is not detectable at all. This renders unambiguous identification of homologous compounds difficult since the information on the molecular mass is not available. A further disadvantage is that fragment ions at lower masses ($m/z < 100$) are less characteristic for a molecule and make selective detection difficult.

For the detection of compounds that exhibit strong fragmentation in EI, CI is the preferred method. The latter is a soft ionization technique where the analyte molecules are ionized indirectly through gas-phase ion–molecule reactions. For this purpose, a reagent gas such as methane or isobutane is introduced into the ion source at relatively high pressure (ca. 1 Torr = 133 Pa) and ionized under EI. The ionized reagent gas usually transfers a proton to the analyte upon collision, thereby forming the characteristic $[M + H]^+$ quasi-molecular ion. Owing to the low energy of this reaction, intensive quasi-molecular ions occur which often are the base peak (= the most intense peak) of the spectrum (Figure 17a and b). This allows in most cases the determination of the molecular mass of an unknown. CI is also particularly useful for highly chlorinated compounds (PCBs, dioxins, toxaphenes, organochlorine pesticides) which can be detected in the negative ion mode with excellent selectivity and sensitivity owing to the high electron affinity of these compounds. Over the past years, the operation of GC/MS in the CI mode has greatly gained in simplicity and stability so that GC/CI/MS may nowadays be considered an excellent technique for sensitive quantitative analysis.

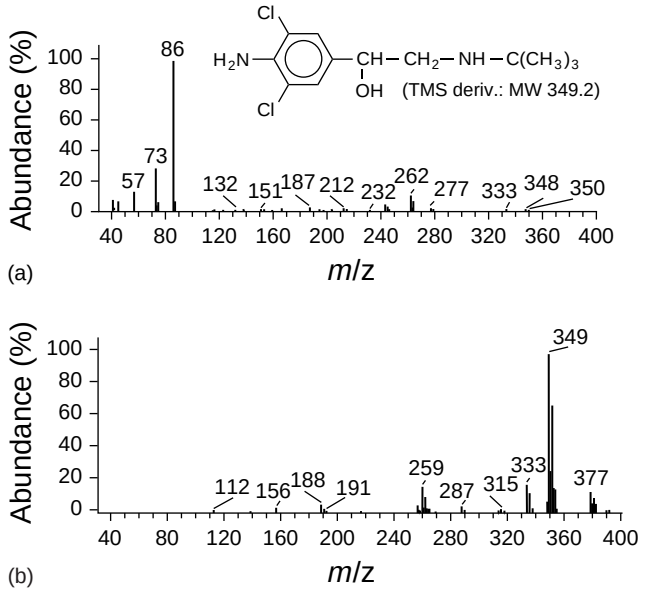

Figure 17 Mass spectra of trimethylsilylated clenbuterol obtained with (a) EI and (b) CI using methane as reagent gas. TMS, trimethylsilyl.

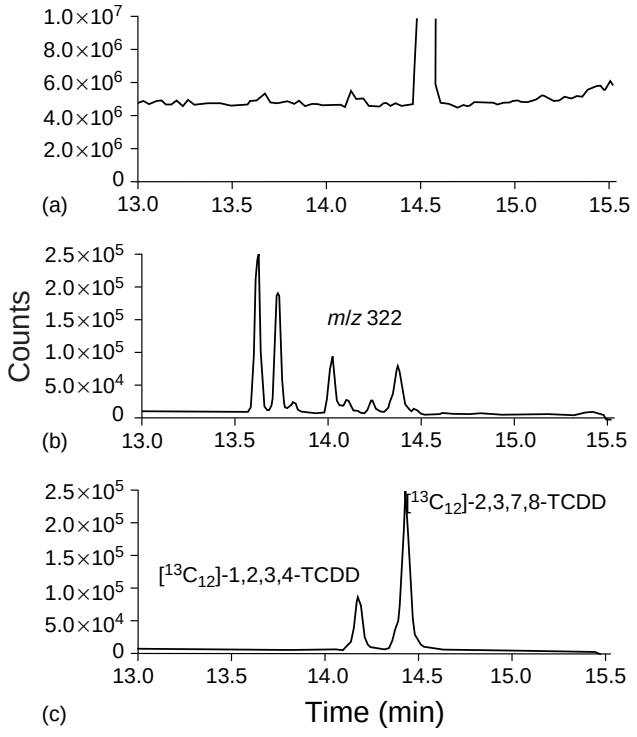

Figure 18 (a) Enlargement from a GC/MS full-scan analysis of a municipal incinerator fly ash for tetrachlorinated dibenzo-*p*-dioxin (TCDD) compounds and (b) SIM chromatogram of the same sample, indicating the presence of some TCDD isomers. (c) SIM chromatogram at *m*/*z* 334 for the same sample to which ^{13}C-labeled TCDDs were added for quality control purposes. [Redrawn from R.E. Clement, E.J. Reiner, in *Detectors for Capillary Chromatography*, Chemical Analysis, eds. H.H. Hill, D.G. McMinn, John Wiley & Sons, New York, 351, Vol. 121, 1992.]

The ions formed in the ion source are accelerated into the mass analyzer where they are separated according to their *m*/*z* values. In the case of a quadrupole mass filter this is achieved by adjusting the voltages and frequency applied to the four rods of the quadrupole. The mass filter thus becomes permeable for ions only of a defined *m*/*z* ratio while ions of other *m*/*z* ratios are not transmitted. A mass spectrum is recorded by variation of the respective voltages with time and is called a scan. The resulting mode of operation, *full-scan mode*, is required when unknown substances are to be identified according to their specific fragmentation pattern. In this mode, the mass spectrometer is scanned continuously while the analytes elute from the column. The summation of all individual ion intensities in a spectrum results in the total ion current which is plotted against the retention time or scan number as a *total ion chromatogram* or *reconstructed ion chromatogram*. The total ion chromatogram is similar to the signal of a universal detector, e.g. the FID, but with greatly different relative intensities. All spectra recorded during a chromatographic run are stored and may be retrieved for identification purposes. Nowadays, this is almost exclusively done by computer-assisted comparison with existing spectral libraries which contain far more than 100 000 reference spectra [e.g. the National Institute of Standards and Technology (NIST) or the Wiley spectral library]. From the stored data, both compound- and class-specific information can be obtained. The latter can be obtained by extracting from the three-dimensional data array (intensity of the ions of different *m*/*z* ratios as a function of time) the intensity of an *m*/*z* ratio characteristic for the compound class monitored (e.g. the molecule ion of a group of isomers or a characteristic fragment ion for a series of homologs). The resulting trace of a fixed *m*/*z* ratio over time is called an *extracted ion chromatogram* (EIC), and it provides compound class selectivity from data acquired in the full-scan mode, but no increase in sensitivity. If highly selective and sensitive quantitative analysis is to be carried out, MS is operated in the selected ion monitoring (SIM) mode. In this mode of operation, the mass spectrometer settings are adjusted in such a way that only a limited number of ions are monitored (typically two to four) by stepwise adjustment of the voltages during one cycle. Hence the time that is available for counting the abundance of a selected few ions is significantly longer than in the full-scan mode (typically by a factor of 50–100). This results in a greatly improved sensitivity and reproducibility and makes quantitative detection in the low-picograms range possible (Figure 18a–c).

An extremely valuable feature of GC/MS is that it allows the use of isotopically labeled internal standards to improve the accuracy and precision of quantitative analysis. These compounds have physical and chemical properties practically identical with those of the original analytes, except that e.g. a number of hydrogen or carbon atoms have been replaced by deuterium or ^{13}C atoms. These isotopically labeled standards are added to the sample before any pretreatment or clean-up step. As sample preparation is performed, both the deuterated and the undeuterated standard behave similarly so that any systematic errors of the method are compensated. Quantitation is then performed in the SIM mode by monitoring selected ion intensities of the analyte and ratioing these to the ion intensities of the isotopically labeled internal standards. The application of this highly useful quantitation technique is restricted only by the limited availability of isotopically labeled standards.

3.1.1 Gas Chromatography with Other Mass Spectrometer Types

Owing to its robustness, its compact construction and competitive price, the quadrupole mass spectrometer is most frequently used in GC/MS instruments. It has, however, the following shortcomings: quadrupole mass filters are low-resolution spectrometers (they are usually operated at 'unit resolution', which means that two neighboring peaks with one unit mass difference can still be separated, but they are not capable of high resolution). Their scanning speed is limited (allowing from <1 to a maximum of 10 scans per second, depending on the scan range and the chosen acquisition parameters). The transmission decreases strongly with increasing m/z ratio and is limited to about 1000 for most commercial instruments (which is still sufficient for most GC/MS applications). The sensitivity in the full-scan mode is ca. 100 times lower than in the SIM mode.

Some of these limitations are overcome by particular mass spectrometer types. The *ion-trap mass spectrometer* has a number of attractive features and has thus gained much in importance over the past years. An ion-trap mass spectrometer can be thought of as a quadrupole mass filter whose rods are closed to a circle. It consists of a central, circular ring electrode and two further electrodes as caps on the top and bottom side of the ion trap to which the respective alternating current (AC) and DC voltages are applied. The ions that are ionized by a conventional EI (or, less frequently, a CI) source enter the detector through an electronic gate which can be opened or closed by application of an appropriate voltage. The ions are stored in the ion trap by the application of a radiofrequency (RF) voltage to the ring electrode. The ion trap detector is operated at a relatively high pressure (10^{-2}–10^{-3} Torr) to dampen the motion of the stored ions and to avoid their immediate loss by ejection out of the ion trap. Analysis of the ions is performed by applying an RF voltage sweep. This successively destabilizes ions of different m/z ratios on their orbit so that they are ejected from the trap and are detected by the electron multiplier. The automatic gain control of modern ion-trap detectors controls the amount of ions in the trap and thus minimizes signal suppression or distortion due to space-charge effects. Ion trap detectors are extremely sensitive and allow, owing to the storage of all ions and successive ejection to the detector during one scan, the acquisition of full-scan spectra in the low-picogram range with a dynamic range of up to 10^6. Further advantages of the ion-trap detector are its compact construction and the less stringent vacuum requirements, allowing reasonably priced instruments to be built. Ion trap detectors furthermore offer the possibility of multi-dimensional MS, MS/MS (or MS^2) and even multiple MS (MS^n). This is achieved by ejecting all ions but those of a selected m/z ratio from the trap and inducing, through the application of an RF voltage sweep, fragmentation of these selected ions. The fragments of this parent ion can be detected in a subsequent scan (Figure 19). If required, this experiment can be repeated to obtain MS^3 mass spectra. This feature significantly increases the structure elucidation capabilities for unknowns by GC/MS, or greatly improves the S/N and thus detectability since the background is virtually suppressed completely. However, it is only possible to obtain daughter ion spectra. Operation in parent-ion or linked scan mode remains reserved for the significantly more complex and more expensive triple quadrupole instruments.

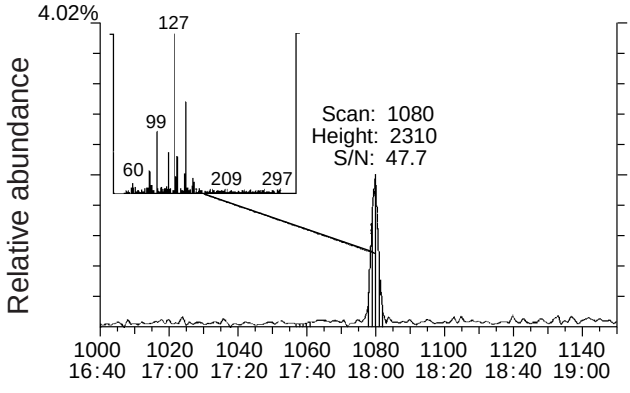

Figure 19 Mass chromatogram of malathion in orange peel extract at the 30 pg (abs.) level, recorded at m/z 127 from collision-induced dissociation in the trap of m/z 173. The inset shows the mass spectrum of the parent compound. [Courtesy of Varian Chromatography Systems, Walnut Creek, CA, USA.]

High-resolution GC/MS (with resolution $R > 5000$) requires the use of significantly more complex instrumentation, such as double-focusing instruments. These consist of an electrostatic analyzer and a magnetic field separator. The electrostatic analyzer provides energy focusing: the ions with a particular m/z ratio but different kinetic energy are deflected towards the same focal point. This significantly improves the mass resolution of the instrument without a decrease in signal (as opposed to reducing the slit width), since the mass resolution of the magnetic separator is critically dependent on the energy and the divergence of the flight path of the entering ions. With a double-focusing instrument, either high resolution or accurate mass determination is possible. This is particularly useful, e.g. in the ultratrace determination of dioxins in complex environmental matrices.

A further type of mass spectrometer, recently introduced for GC detection and capable of both extremely sensitive detection and, in principle, high mass resolution is the *time-of-flight* (TOF) mass spectrometer. In the TOF mass spectrometer, ions are formed in a pulsed ion source and accelerated by the application of a potential V through a field-free flight tube towards the detector. The time t that the ions need to arrive at the detector (in a distance d) is measured and related to the m/z ratio according to Equation (16):

$$t = d\sqrt{\frac{m}{2zeV}} \tag{16}$$

It is evident from this equation that a pulsation or gating of the ion source is required to let only ions of a fixed m/z ratio arrive at the detector at a certain time, and that the mass resolution is dependent on the pulse width. This factor, together with the speed of the counting electronics, has until recently limited the achievable resolution to a few hundred; however, instruments of significantly higher resolution (ca. 5000) are nowadays available. TOF mass spectrometers offer the advantage of a very high sensitivity in the 'scan' mode since all ions generated arrive at the detector, and they excel in the absence of mass discrimination effects where higher masses are transmitted to a lower degree than lower mass ions.

3.2 Gas Chromatography/Fourier Transform Infrared Spectroscopy

Coupled GC/FTIR spectroscopy provides the acquisition of mid-infrared (IR) spectra (typically in the spectral region 400–4000 cm^{-1}) of the peaks as they elute from the chromatographic column. GC/FTIR is a useful alternative and complement to GC/MS detection since it is a non-destructive technique (and may thus be coupled in series with other detectors, e.g. GC/MS), it allows the discrimination of isomers and offers the possibility of functional group-selective detection (through recording of the absorbance at frequencies characteristic of functional groups of the molecule).

FTIR detection for GC is most frequently done with the lightpipe or flow cell approach that was introduced in the early 1980s by Azarraga[27] The lightpipe is a flow cell through which the column effluent passes and that is located in the beam of the FTIR spectrometer (Figure 20). It is a heated, internally gold-coated glass tube with IR-transparent windows (made from, e.g. KBr or ZnSe) which is connected to the chromatographic column via a heated transfer line. The internal gold coating provides maximum reflection of the IR radiation which has to be transmitted through the lightpipe. Typical dimensions of the lightpipe are 1 mm ID and 1–20 cm length, resulting in a detection cell volume of 80–160 µL. It is evident from these figures that the lightpipe volume has to be carefully matched to the GC system and operating parameters. It always represents a compromise between maximum sensitivity (achieved by larger lightpipe dimensions) and minimum distortion of chromatographic resolution (requiring smaller detection cell volumes), particularly when using capillary columns. The simplicity of this approach has made the lightpipe interface the most frequently used for GC/FTIR measurements. Its sensitivity, however, is not adequate for highly sensitive detection at the ultratrace level, since ca. 5–100 ng of a strong IR absorber are required for a good-quality spectrum. The spectra acquired during the elution of a peak are co-added and may be library searched for identification purposes. However, dedicated

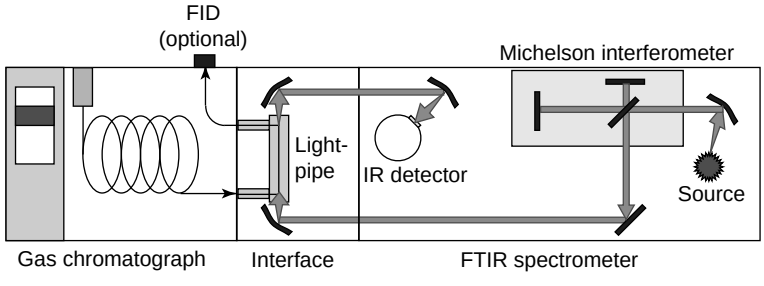

Figure 20 Schematic diagram of a GC/FTIR instrument with a lightpipe interface.

gas-phase libraries are necessary since IR spectra depend strongly on the physical state of the compound.

There exist also other approaches for coupling FTIR detection to GC separation. These are not as experimentally simple, though, and therefore far less popular. They rely on the deposition of the column effluent on a ZnSe plate that is cooled by liquid nitrogen to 77 K. While the analytes elute from the GC column, the ZnSe plate that is mounted on an $x-y$ translation stage is moved with a constant linear velocity. The analytes are thus immobilized along a line on the ZnSe plate and transported into the beam of a dedicated IR microscope that focuses the IR beam to a diameter that matches the sample spot dimensions (ca. 100 μm). The chromatogram in time is thus translated into an (immobilized) chromatogram in space and compounds may be detected in near real time. Owing to the immobilization of the analytes on the ZnSe plate, longer scanning times can be applied and consequently significantly improved detection limits can be achieved (20–100 pg are often sufficient for an IR spectrum). The IR spectra are comparable to those existing in widely available condensed-phase libraries. However, highly volatile compounds are not detectable since they are not efficiently trapped on impacting the liquid nitrogen-cooled ZnSe plate.[28]

Data acquisition in GC/FTIR detection is still highly demanding owing to the large number of data points (a spectrum recorded between 4000 and 400 cm^{-1} with a resolution of 4 cm^{-1} has 900 data points) and the high scanning speed of modern FTIR spectrometers (typically 1–10 scans s^{-1}). To obtain an on-line-signal or a functional group chromatogram, a fast Fourier transformation (FFT) of a limited subset of the data acquired, e.g. with a Michelson-type interferometer, is performed. The FFT of the interferograms with a reduced number of data points is substantially faster, but results in spectra with lower resolution (e.g. 8 or 16 instead of 4 cm^{-1}) which are, however, available practically on-line. Also, functional group chromatograms can be plotted as the absorbance calculated for a spectral window of interest (e.g. in the carbonyl band region) versus time to detect the presence of analytes carrying this functional group (Figure 21a–d).

The most convincing examples of GC/IR analysis originate from the field of natural products analysis: since fragrances and natural substances are very often isomers, having the same molecular formula but different structure, MS is often not able to differentiate these isomers whereas IR detection succeeds (e.g. Figure 22a–c).

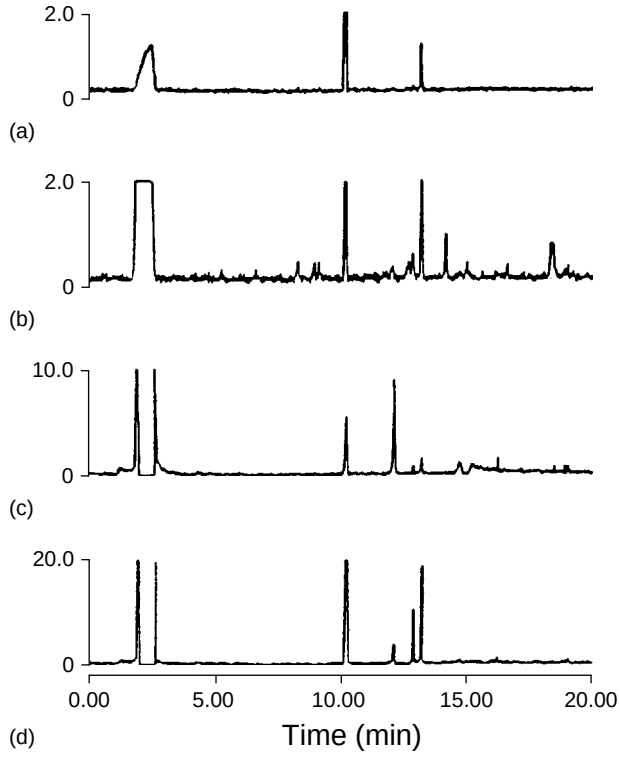

Figure 21 Functional group GC/FTIR of a sediment extract, indicating the presence of carbonyl groups in the detected analytes: (a) and (b) aldehydic Fermi proton resonance vibrations; (c) aliphatic ester carbonyl stretch; (d) aromatic aldehyde carbonyl stretch. (a) 2619–2766 cm^{-1}; (b) 2801–2851 cm^{-1}; (c) 1748–1759 cm^{-1}; (d) 1701–1740 cm^{-1}. The y axes are in arbitrary units. [Reproduced by permission of American Chemical Society from D.F. Gurka, R. Titus, P.R. Griffiths, D.E. Henry, A. Giorgietti, *Anal. Chem.*, **59**, 2362–2365 (1987).]

3.3 Gas Chromatography/Atomic Emission Detection

Of the different atomic spectroscopic techniques to be coupled to GC, only atomic emission spectroscopy has gained greater importance through the availability of commercial instrumentation since the late 1980s.[29,30] The basis was already laid in the mid-1960s through the work of McCormack et al.[31] and Bache and Lisk[32] on plasma emission spectrometric GC detection. Being the most recent development in the group of hyphenated techniques, the AED provides, in contrast to GC/MS and GC/FTIR, *element-specific* rather than *compound-specific* detection and thus ideally complements these detection techniques.

In AED, the chromatographic effluent is introduced into a helium-sustained microwave-induced plasma (MIP). In the plasma, the analytes are completely atomized, partially ionized and excited. On returning to a lower energy level, the atoms or ions emit characteristic radiation which is recorded by a diode-array spectrometer. Different ionization, recombination and energy transfer processes contribute to the excitation of the atoms liberated from the analytes. The use of He as a plasma gas is practical since it is often also applied as a GC carrier gas and is

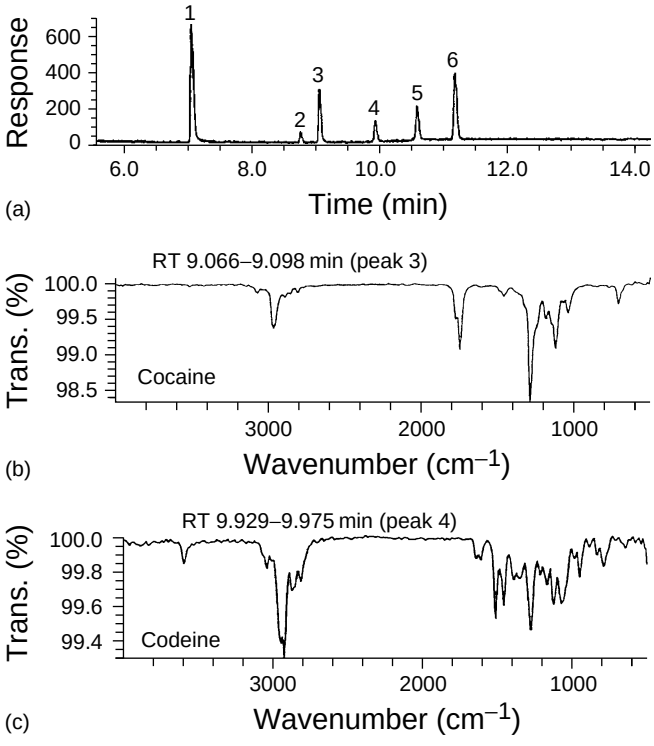

Figure 22 (a) GC/FTIR chromatogram of a mixture of drugs of abuse. Peaks: 1 = caffeine; 2 = methadone; 3 = cocaine; 4 = codeine; 5 = 6-acetylmorphine; 6 = heroin. FTIR spectra of peaks (b) 3 and (c) 4. Injected amounts: 150–375 ng of individual compounds. Trans., transmission. [Courtesy of Hewlett-Packard Co., Palo Alto, CA, USA.]

particularly advantageous in AED, since it has a simpler spectral background and a significantly higher excitation energy than Ar (the first ionization energy is 24.59 eV vs 15.76 eV for argon). Hence practically all elements can be excited efficiently, including the nonmetallic elements. Comparatively low He flow rates (30–300 mL min^{-1}) and low power (ca. 60 W) are required to sustain the plasma, but it is therefore not very resistant to larger sample loads. The plasma is powered by a microwave generator operating at 2.45 GHz and sustained in a 1 mm ID quartz tube which is centered inside the microwave cavity. In the plasma, temperatures of 3500–4000 K are reached, at which the analytes are efficiently atomized and excited. The characteristic element-specific radiation is detected in an axial viewing configuration by a diode-array spectrometer including a holographic grating as dispersion element. The accessible spectral range is from ca. 160 to 800 nm. As the whole spectral range cannot be recorded by the diode-array detector with acceptable resolution, only portions of the emission spectrum in the region of emission lines of interest are recorded at a time (Figure 23). Useful sets of emission lines are, e.g. those of C, H, Cl and Br, which can be measured simultaneously in the 470–496 nm range, or C, S, Se and N, which can be measured in the 170–196 nm range. The recording of elemental chromatograms at emission lines in largely different regions of the spectrum requires separate chromatographic runs. Separate chromatographic runs also have to be made when the required plasma dopant gases that are added to the He in small quantities (also called scavenger gases) are different. The detection of phosphorus, for example, requires only hydrogen as reagent gas and makes its simultaneous detection with C, N and S which require both hydrogen and oxygen as reagent gases impractical, although the emission lines of these elements all lie within the same spectral range.

The sensitivity of AED can vary strongly between elements (cf. Table 6). AED is an excellent method

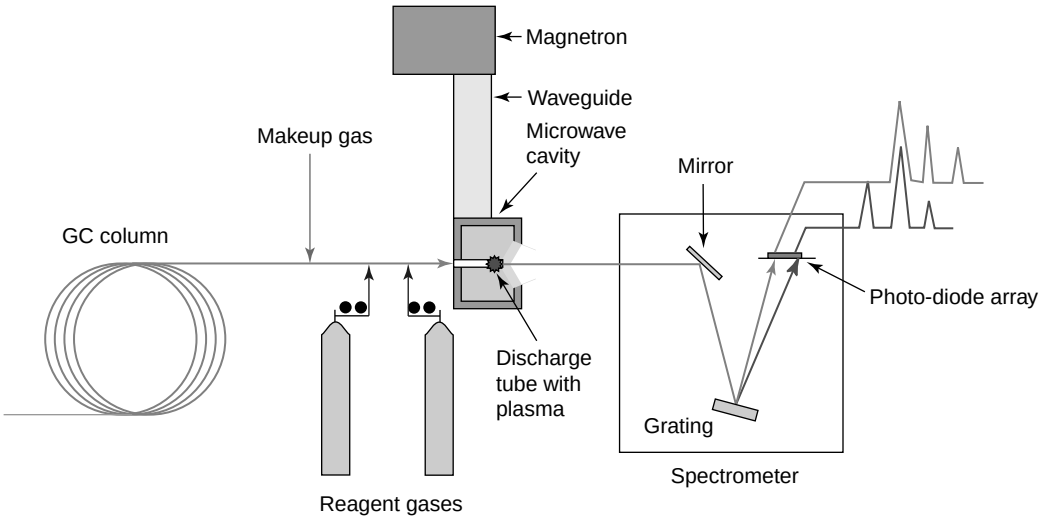

Figure 23 Schematic diagram of a GC/AED instrument.

Table 6 Figures of merit for MIP/AED for GC. [Reproduced by permission of Elsevier Science from P.C. Uden, *J. Chromatogr. A*, **703**, 393–416 (1995).]

Element	Wavelength (nm)	Detection limit [pg s^{-1} (pg)]	Selectivity vs C	LDR
Carbon	247.9	2.7	1	>1000
Carbon	193.1	2.6	1	21 000
Hydrogen	656.3	7.5 (22)	160	500
Hydrogen	486.1	2.2	Variable	6000
Deuterium	656.1	7.4	200	500
Boron	249.8	3.6 (27)	9300	500
Chlorine	479.5	39	25 000	20 000
Bromine	470.5	10	11 500	>1000
Fluorine	685.6	40	30 000	2000
Sulfur	180.7	1.7	150 000	20 000
Phosphorus	177.5	1	5000	1000
Silicon	251.6	7.0	90 000	40 000
Oxygen	777.2	75	25 000	4000
Nitrogen	174.2	7.0	6000	43 000
Aluminum	396.2	5.0	>10 000	>1000
Antimony	217.6	5.0	19 000	>1000
Gallium	294.3	~200	>10 000	>500
Germanium	265.1	1.3 (3.9)	7600	>1000
Tin	284.0	1.6 (6.1)	36 000	>1000
Tin	303.1	(0.5)	30 000	>1000
Arsenic	189.0	3.0	47 000	500
Selenium	196.1	4.0	50 000	>1000
Chromium	267.7	7.5	108 000	>1000
Iron	302.1	0.05	3 500 000	>1000
Lead	283.3	0.17 (0.71)	25 000	>1000
Mercury	253.7	0.1	3 000 000	>1000
Vanadium	292.4	4.0	36 000	>1000
Titanium	338.4	1.0	50 000	>1000
Nickel	301.2	1.0	200 000	>1000
Palladium	340.4	5.0	>10 000	>1000
Manganese	257.6	1.6 (7.7)	110 000	>1000

for sulfur, carbon and metals, but less sensitive for nitrogen and the halogens, especially when compared with the NPD and the ECD, respectively. It offers, however, a number of practical advantages which make its use very convenient for particular applications. The response is in principle compound independent. Owing to the complete atomization of the analytes in the plasma, the signal no longer depends on the molecular structure of the compound to be quantified. This offers the possibility of calibration with a substitute standard when the compounds to be quantified are not available as pure substances or are hazardous. Moreover, empirical formulae may be calculated, which provides complementary and useful information for the interpretation of mass spectral data. The selectivity of AED very often exceeds that of the corresponding element-selective detector such as the NPD or ECD, and also the linear range is often larger. It is also able to differentiate between the halogens chlorine, bromine, iodine and fluorine, which is not possible with any other selective detector. AED is not only able to detect virtually all elements of the periodic table including oxygen, but also some of the stable isotopes, such as ^{2}D, ^{13}C and ^{15}N, based on the slight shift of molecular band position (although at lower sensitivity and fair selectivity over the more abundant isotope).

AED has found broad use in the petroleum industry, where its capability of detecting sulfur-containing, oxygenated and organometallic compounds is particularly useful, and in environmental analysis, where the combination with MS detection greatly improves the interpretability of the data (Figure 24a and b).

3.4 Other Atomic Spectroscopic Detectors for Gas Chromatography

While GC/AED, due to its commercial availability, is probably the most frequently applied atomic spectroscopic detector, a number of other less popular techniques are available for atom spectroscopic GC detection. These include atomic absorption spectrometry (AAS), atomic fluorescence spectrometry (AFS) and inductively coupled plasma mass spectrometry (ICPMS).

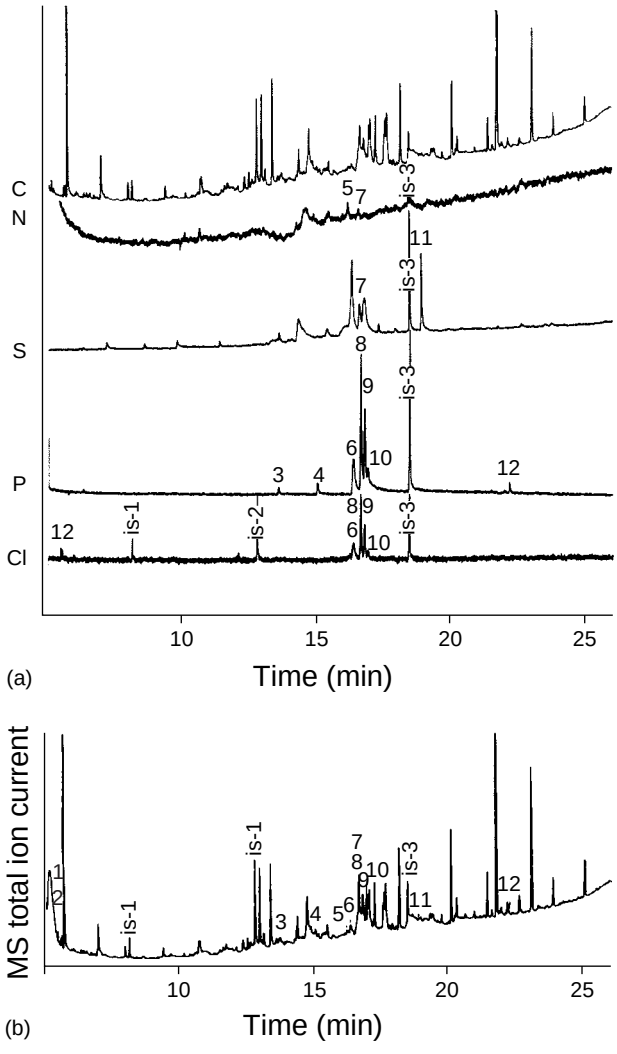

Figure 24 GC/AED/MS traces obtained after injection of a preconcentrated river water sample extract. (a) AED chromatogram of the C, N, S, Cl and P traces. (b) Simultaneously acquired total ion current MS chromatogram. Peaks: is-1 = 5-bromo-2-chlorotrifluoromethylbenzene; is-2 = 1-chlorododecane; is-3 = chloropyriphos; 1 = tetrachloroethane; 2 = trichloropropane; 3 = triisobutyl phosphate; 4 = tributyl phosphate; 5 = atrazine; 6 = tri(2-chloroethyl)phosphate; 7 = *n*-butylbenzenesulfonamide; 8 = tri(2-chloro-1-methylethyl)phosphate; 9 = di(2-chloro-1-methylethyl)-2-chloropropyl phosphate; 10 = 2-chloro-1-methylethyl-bis(2-chloropropyl) phosphate; 11 = sulfur; 12 = octyl diphenylphosphate. [Reproduced by permission of Advanstar Communications, Chester, from H.G.J. Mol, T. Hankmeier, U.A.Th. Brinkman, *LC–GC Int.*, **12**, 108–120 (1999).]

AAS is inherently a metal-specific detection method where GC is a sample introduction technique for either volatile organometallic species (such as organic compounds of lead, tin, mercury, selenium or arsenic) or metal species that can be rendered volatile through derivatization or complex formation. AAS requires the analytes to be efficiently atomized so that the free atoms can be excited. The absorbance of monochromatic radiation characteristic for the element to be detected is measured, and the signal is proportional to the analyte concentration and the optical path length according to the Lambert–Beer law. While the simplest way of coupling GC to AAS is to introduce the GC column outlet directly into the base of an only slightly modified burner of a conventional flame spectrometer, this method suffers from poor sensitivity owing to the short residence time and the comparatively large dilution of the analytes in the detection zone. A more efficient method is the coupling of GC to electrothermal (graphite or quartz furnace) AAS. In this case, a T-shaped quartz tube or a modified graphite furnace is used. The sample is introduced through a heated transfer line into the center of the graphite or quartz furnace. Sometimes the addition of a reactive gas such as hydrogen is useful for enhancing the sensitivity. Detection limits of 10–100 pg of analyte can be reached in favorable cases (e.g. for tin, lead and arsenic). The most interesting applications are from the field of speciation analysis where, e.g. ionic alkyltin compounds are determined after derivatization with a suitable Grignard reagent (Figure 25).

AFS detection for GC combines the selectivity of AAS with the multielement detection capabilities of atomic emission techniques. The theoretically wide applicability

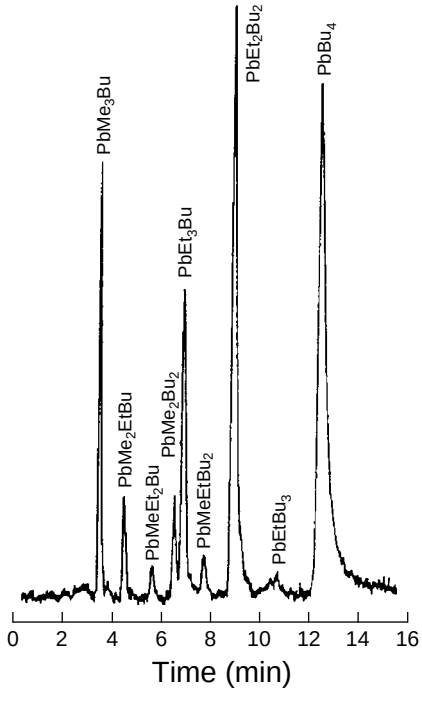

Figure 25 Typical GC/AAS trace of butylated organolead compounds in a rainwater sample, recorded with deuterium background correction. [Courtesy of R.J.A. van Cleuvenbergen, PhD Thesis, University of Antwerp, 1989.]

is in practice limited by the availability of suitable simple and intense excitation sources and by the sparseness of commercial instrumentation. Typical atom reservoirs are the cells also used in AAS with some modifications, e.g. premixed laminar argon sheathed miniature diffusion flames and heated quartz and graphite atomizers. The use of conventional line sources (hollow-cathode and electrodeless discharge lamps) allows detection limits in the 10–100 pg range to be achieved for most metallic elements. Extreme detection limits down to sub-picrogram levels have been reported when using pulsed tunable dye lasers pumped by an Nd : YAG laser or an excimer laser. However, the extremely high costs and the difficulty of operation prevent this technology from being widely applicable.

The inductively coupled plasma (ICP) and particularly also the MIP can advantageously be used as atomizers and ion sources for ultrasensitive elemental MS detection for GC [GC/ICP/MS and GC/microwave-induced plasma mass spectrometry (MIPMS)]. To overcome the pressure difference between the plasma source operated at overpressure and large gas flows and the mass spectrometer which in most cases is a simple low-resolution quadrupole mass analyzer operated at high vacuum, a special interface is required. This consists of a flexible, heated transfer line through which the analytes are introduced into the plasma torch. In the Ar plasma of the ICP, the analytes are efficiently atomized and ionized. These ions are sampled through a two-stage skimmer into the vacuum of the mass analyzer and detected. The advantages of plasma MS detection are the extremely low detection limits that can be achieved (down to the 50–100 fg range), the simultaneous multielement detection ability and the ability to measure isotopic ratios. In practice, the MIP offers significant advantages over the ICP as an ionization source owing to the lower operating power and supporting gas requirements. Additionally, He has a higher ionization potential than Ar, which increases the sensitivity of detection. Finally, spectral interferences through the formation of cluster ions are less pronounced with He. However, owing to the wider availability of ICPMS, the GC/ICPMS set-up is more popular. Typical applications which benefit from the ultratrace sensitivity of ICPMS detection are in the field of speciation analysis (Figure 26).

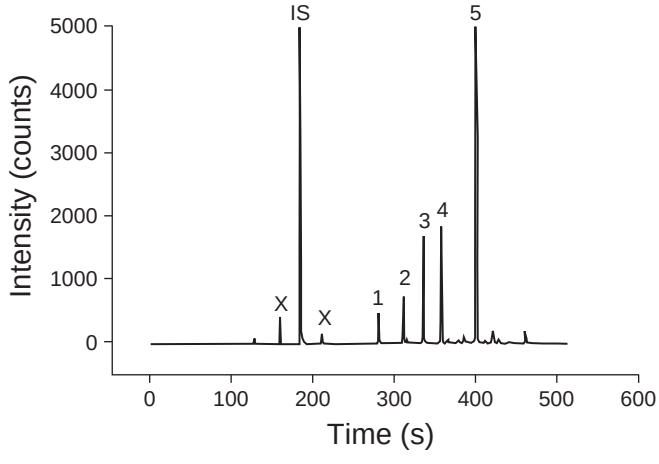

Figure 26 GC/ICPMS trace of propylated organotin compounds in a marine sediment reference material (PACS-1), monitored at the m/z 120 isotope of Sn. Peaks: IS = internal standard; 1 = inorganic tin (as $SnPr_4$); 2 = monobutyltin; 3 = dibutyltin; 4 = tributyltin; 5 = tributylpentyltin (as IS); X = unknown compounds. [Reproduced by permission of Elsevier Science from T. De Smaele, L. Moens, R. Dams, P. Sandra, J. Van der Eycken, J. Vandyck, *J. Chromatogr. A*, **793**, 99–106 (1998).]

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AC	Alternating Current
AED	Atomic Emission Detection
AFID	Alkali Flame Ionization Detector
AFS	Atomic Fluorescence Spectrometry
CI	Chemical Ionization
DC	Direct Current
EC	European Community
ECD	Electron Capture Detector
EI	Electron Impact
EIC	Extracted Ion Chromatogram
ELCD	Electrolytic Conductivity Detector
FFT	Fast Fourier Transformation
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/AED	Gas Chromatography/Atomic Emission Detection
GC/FTIR	Gas Chromatography/Fourier Transform Infrared
GC/MS	Gas Chromatography/Mass Spectrometry
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ID	Inner Diameter
IR	Infrared
LDR	Linear Dynamic Range
LOD	Limit of Detection
MIP	Microwave-induced Plasma
MIPMS	Microwave-induced Plasma Mass Spectrometry
MS	Mass Spectrometry
NIST	National Institute of Standards and Technology

NPD	Nitrogen–Phosphorus Detector
O-FID	Oxygen-sensitive Flame Ionization Detector
PCB	Polychlorinated Biphenyl
PID	Photoionization Detector
PTFE	Polytetrafluoroethylene
RF	Radiofrequency
SCD	Sulfur Chemiluminescence Detector
SIM	Selected Ion Monitoring
S/N	Signal-to-noise Ratio
TCD	Thermal Conductivity Detector
TCDD	Tetrachlorinated Dibenzo-*p*-dioxin
TEA	Thermal Energy Analyzer
TMS	Trimethylsilyl
TOF	Time-of-flight
USEPA	United States Environmental Protection Agency
UV	Ultraviolet

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Organometallic Compound Analysis in Environmental Samples

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. R.E. Sievers (ed.), *Selective Detectors*, Chemical Analysis, Wiley & Sons, New York, Vol. 131, 1995.
2. H.H. Hill, D.G. McMinn (eds.), *Detectors for Capillary Chromatography*, Chemical Analysis, Wiley & Sons, New York, Vol. 121, 1992.
3. M. Dressler, *Selective Gas Chromatographic Detectors*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 36, 1986.
4. D. Barceló, 'Official Methods of Analysis of Priority Pesticides in Water using Gas Chromatographic Techniques', in *Environmental Analysis: Techniques, Applications and Quality Assurance*, ed. D. Barceló, Elsevier, Amsterdam, 149–180, 1993.
5. J.E. Lovelock, S.R. Lipsky, 'Electron Affinity Spectroscopy – A New Method for the Identification of Functional Groups in Chemical Compounds Separated by Gas Chromatography', *J. Am. Chem. Soc.*, **82**, 431–433 (1960).
6. A. Zlatkis, C.F. Poole (eds.), *Electron Capture. Theory and Practice in Chromatography*, Elsevier, Amsterdam, 1981.
7. B. Kolb, J. Bischoff, 'A New Design of a Thermionic Nitrogen and Phosphorus Detector for GC', *J. Chromatogr. Sci.*, **12**, 625–629 (1974).
8. A. Karmen, L. Giuffrida, 'Enhancement of the Response of the Hydrogen Flame Ionization Detector to Compounds Containing Halogens and Phosphorus', *Nature (London)*, **201**, 1204–1205 (1964).
9. S.S. Brody, J.E. Chaney, 'Flame Photometric Detector', *J. Gas Chromatogr.*, **4**, 42–46 (1966).
10. W.A. Aue, C.G. Flinn, 'Chemical Linearization of Chalcogen Response in a Flame Photometric Detector', *J. Chromatogr.*, **158**, 161–170 (1978).
11. C.G. Flinn, W.A. Aue, 'Geometric and Chemical Discrimination in a Non-dispersive Flame Photometric Detector', *J. Chromatogr. Sci.*, **18**, 136–138 (1980).
12. J.E. Lovelock, 'Ionization Methods for the Analysis of Gases and Vapors', *Anal. Chem.*, **33**, 162–178 (1961).
13. D.C. Locke, C.E. Meloan, 'Study of the Photoionization Detector for Gas Chromatography', *Anal. Chem.*, **37**, 389–395 (1965).
14. J.N. Driscoll, F.F. Spaziani, 'New Instrument for Continuous Monitoring of Odorous Sulfur Compounds', *Adv. Instrum.*, **29**(3), 709–712 (1974).

15. M.L. Langhorst, 'Photoionization Detector Sensitivity of Organic Compounds', *J. Chromatogr. Sci.*, **19**, 98–103 (1981).
16. O. Piringer, M. Pascalau, 'A New Detector for Gas Chromatography', *J. Chromatogr.*, **8**, 410–412 (1982) (in German).
17. D.M. Coulson, 'Selective Detection of Nitrogen Compounds in Electrolytic Conductivity Gas Chromatography', *J. Gas Chromatogr.*, **4**, 285–287 (1966).
18. R.C. Hall, 'A Highly Sensitive and Selective Microelectrolytic Conductivity Detector for Gas Chromatography', *J. Chromatogr. Sci.*, **12**, 152–158 (1974).
19. R.C. Hall, 'The Nitrogen Detector in Gas Chromatography', *CRC Crit. Rev. Anal. Chem.*, **7**, 323–380 (1978).
20. J.W. Birks (ed.), *Chemiluminescence and Photochemical Reaction Detection in Chromatography*, VCH, New York, 1989.
21. D.H. Fine, F. Rufeh, D. Lieb, 'Group Analysis of Volatile and Non-volatile *N*-Nitroso Compounds', *Nature (London)*, **247**, 309–310 (1974).
22. W. Schneider, J.C. Frohne, H. Bruderreck, 'Selective Gas Chromatographic Determination of Oxygenated Compounds by Flame Ionization Detection', *J. Chromatogr.*, **245**, 71–83 (1982) (in German).
23. S.A. Borman, 'New Gas Chromatographic Detectors', *Anal. Chem.*, **55**, 726A–730A (1983).
24. H.G.J. Mol, T. Hankemeier, U.A.Th. Brinkman, 'Gas Chromatography of Organic Microcontaminants using Atomic Emission and Mass Spectrometric Detection Combined in One Instrument (GC/AED/MS)', *LC–GC Int.*, **12**, 108–120 (1999).
25. H. Frischenschlager, C. Mittermayr, M. Peck, E. Rosenberg, M. Grasserbauer, 'The Potential of Gas Chromatography with Microwave-induced Atomic Emission Detection (GC/MIP/AED) as a Complementary Analytical Technique in Environmental Screening Analysis of Aqueous Samples', *Fresenius' J. Anal. Chem.*, **359**, 213–221 (1997).
26. F.W. Karasek, R.E. Clement, *Basic Gas Chromatography/Mass Spectrometry: Principles and Techniques*, Elsevier, Amsterdam, 1988.
27. L.V. Azarraga, 'Gold Coating of Glass Tubes for Gas Chromatography/Fourier Transform Infrared Spectroscopy ''Light Pipe'' Gas Cells', *Appl. Spectrosc.*, **34**, 224–225 (1980).
28. P.R. Griffiths, J.A. de Haseth, L.V. Azarraga, 'Capillary GC/FTIR', *Anal. Chem.*, **55**, 1361A–1387A (1983).
29. J.J. Sullivan, B.D. Quimby, 'Evaluation of a Microwave Cavity, Discharge Tube, and Gas Flow System for Combined Gas Chromatography/Atomic Emission Detection', *Anal. Chem.*, **62**, 1027–1033 (1990).
30. B.D. Quimby, J.J. Sullivan, 'Characterization of a Computerized Photodiode Array Spectrometer for Gas Chromatography/Atomic Emission Spectrometry', *Anal. Chem.*, **62**, 1034–1043 (1990).
31. A.J. McCormack, S.C. Wong, W.D. Cooke, 'Sensitive Selective Gas Chromatography Detector based on Emission Spectrometry of Organic Compounds', *Anal. Chem.*, **37**, 1470–1476 (1965).
32. C.A. Bache, D.J. Lisk, 'Determination of Organophosphorus Insecticide Residues using the Emission Spectrometric Detector', *Anal. Chem.*, **37**, 1477–1480 (1965).

Trihalomethanes in Water, Analysis of

Louis Lépine and Roland Gilbert
Institut de Recherche d'Hydro-Québec, Varennes, QC, Canada

1 Introduction **2**
2 Standard Preparation and Sample Handling **3**
2.1 Preparation of Blank Water and Standards 3
2.2 Sample Collection and Preservation 4
3 Direct Aqueous Injection Gas Chromatography **4**
3.1 Instrumentation 5
3.2 Method Parameters 5
3.3 Analytical Performance 6
3.4 Advantages and Limitations 6
4 Liquid–Liquid Extraction Gas Chromatography **7**
4.1 Instrumentation 8
4.2 Method Parameters 9
4.3 Analytical Performance 10
4.4 Advantages and Limitations 10
5 Headspace Gas Chromatography **10**
5.1 Instrumentation 11
5.2 Method Parameters 12
5.3 Analytical Performance 14
5.4 Advantages and Limitations 15
6 Purge-and-trap Gas Chromatography **15**
6.1 Instrumentation 16
6.2 Method Parameters 16
6.3 Analytical Performance 19
6.4 Advantages and Limitations 19
7 Comparison of the Analytical Techniques **19**
Acknowledgments **20**
Abbreviations and Acronyms **21**
Related Articles **21**
References **21**

Trihalomethanes (THMs) are considered to be the major by-products found in water after the chlorination process. They mainly consist of chloroform ($CHCl_3$), bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$) and bromoform ($CHBr_3$). Their content in drinking water is regulated and the maximum contaminant level (MCL) of 100 $\mu g\,L^{-1}$ established in 1979 by the United States Environmental Protection Agency (USEPA) for total THMs, based on a running annual average, still prevails and has been adopted in many other countries. In this article, we describe the four major analytical techniques routinely used for THM analysis. All of these analytical techniques use gas chromatography (GC) with a halogen-specific detector such as the electron capture detector (ECD), electrolyte conductivity detector (ELCD) or mass spectrometry (MS). With high-efficiency capillary columns now commonly available and the high sensitivity of the ECD, all these techniques have detection limits (DLs) at or below 0.1 $\mu g\,L^{-1}$ for each of the four THMs, which is more than adequate for drinking water or beverage applications.

The four analytical approaches differ by the way the THMs are introduced into the chromatographic column. Direct aqueous injection (DAI), where the sample is directly injected into the GC column, is fast and simple. It does not require any other piece of equipment than the gas chromatograph itself. However, since everything in the sample is injected in the column, it is limited to relatively clean matrices. It is generally used as a screening technique when a limited number of samples are analyzed. The second method uses liquid–liquid extraction (LLE) of the THMs with an organic solvent, usually pentane, which is then injected into the GC column. The extraction step offers the possibility of concentrating the THMs from a large-volume sample and isolates them from inorganic salts. However, LLE is by far the most time-consuming procedure and is not well suited for routine analysis of a large number of samples. The last two methods take advantage of the high volatility of the THMs to isolate and concentrate them in the gas phase. In the headspace (HS) technique, the aqueous sample is heated in a sealed vial and an equilibrium is reached between the THMs present in the water and the HS above, from which an aliquot is injected into the GC column. It is simple to use and easily automated. The sensitivity is limited by the partition coefficient of each species between the liquid and gas phases and the size of the gas aliquot injected into the GC column. The DL has been decreased to 1 $ng\,L^{-1}$ by use of a long transfer time to the gas chromatograph and cryogenic trapping in the very first portion of the GC column. Finally, in the purge-and-trap (PT) technique, the THMs are purged out of the sample by an inert gas and concentrated on a solid sorbent, thermally desorbed and injected into the GC column. With a purge efficiency close to 100% for these volatile compounds, all the molecules present in a relatively large water sample (typically 5–25 mL) are injected into the GC column. It is the most sensitive of the four analytical methods and a DL

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

below the 1 ng L^{-1} level has been achieved with cryogenic focusing at the column head.

1 INTRODUCTION

In 1974, Rook[1] in The Netherlands and Bellar et al.[2] in the USA reported the presence of THMs in chlorinated drinking water. Since then, it has been demonstrated that the chlorination process used to disinfect natural water is indeed responsible for the formation of these halogenated organic compounds. It is generally believed that the THMs represent the main group of organohalogens found in chlorinated drinking water. Koch and Krasner[3] estimated that among organohalogen compounds formed during chlorination of water, 77% are THMs, the rest being attributed to haloacetic acids (15%), halonitriles (3%), trichloroacetaldehyde hydrate (4%) and 1% remaining compounds.

By definition, THMs are organic compounds that contain one carbon, one hydrogen and three halogen atoms. They are formed by the classical haloform reaction of chlorine with humic and fulvic acids present in natural waters. Bunn et al.[4] have shown under laboratory conditions that bromine- and iodine-containing THMs are produced because bromide and iodide are oxidized by chlorine into reactive species, which in turn react with humic and fulvic acids in the same way as chlorine. Ishikawa et al.[5] have also shown that bromine-containing THMs are more stable and can be produced by substitution of a chlorine atom in other THMs.

By far the most common THM is chloroform ($CHCl_3$), although lower concentrations of brominated THMs such as bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$) and bromoform ($CHBr_3$) are often detected.[6–8] Iodine-containing compounds reported by a few workers are not usually found.[9,10] For this reason, the term THM usually refers to the four chlorine and/or bromine THMs listed above and will be used as such in this article. Table 1 gives some fundamental properties of these compounds.

Since Rook's discovery in 1974, several studies have been conducted in different countries to measure the level of THMs present in chlorinated municipal waters and Biziuk and Przyjazny[11] presented some of the results in a review on volatile organohalogen compounds. They showed varying concentrations of the four THMs found in chlorinated waters, ranging in most cases from 1 to 150 $\mu g\,L^{-1}$, with a median value of the order of 30–40 $\mu g\,L^{-1}$. In all of these studies, chloroform is always the predominant species. Other references worth mentioning are those of a group of Canadian workers at the Department of Health,[12–15] who showed that the level of THMs produced is greatly dependent on the source and the quality of the raw water. They established a correlation between the total THM level and the total organic content (TOC) or chlorine demand, and also the pH of the raw water. This correlation can explain the wide range of THM concentrations and the seasonal dependence shown in their studies. Also, the highest values were generally observed in treated surface water and the lowest in the treated groundwater. In their most recent paper,[15] they also observed that the level of THMs increased from the treatment plant into the distribution system, probably because of extended reactions between chlorine and the THM precursors remaining in the water. This was also mentioned in a paper by Koukouraki et al.[16]

Several studies have been conducted to assess the potential health risk of long-term exposure to the THMs and other chlorination by-products in drinking water. It has been difficult to obtain conclusive data but the possibility of a causal relationship for rectal cancer and, to a lesser extent, for bladder and colon cancer have been mentioned.[17] It has been also reported that cancer of the liver and kidney developed in rats and mice that ate food and drank for a long time water in which a small amount of chloroform was present.[18] Even though there is still some controversy about the real health risk and the acceptable level of THMs in drinking water, the MCL of 100 $\mu g\,L^{-1}$ established in 1979 by the USEPA[19] for total trihalomethanes (TTHM), based on a running annual average, still prevails and has been adopted in other countries.[15,20] However, a lower MCL of 80 $\mu g\,L^{-1}$ has recently been proposed in USA in the Disinfectants–Disinfection By-products Rule[21] and a Commission of the European Union suggested levels of 40 $\mu g\,L^{-1}$ for $CHCl_3$ and 15 $\mu g\,L^{-1}$ for $CHBrCl_2$.[16]

Table 1 Some fundamental properties of the THMs

Compound	Formula	CAS number	Molecular weight	Boiling point (°C)	Vapor pressure at 20 °C[17] (kPa)	Water solubility at 20 °C ($g\,L^{-1}$)
Chloroform	$CHCl_3$	67-66-3	119.38	61.7	20.1	7.95
Bromodichloromethane	$CHBrCl_2$	75-27-4	163.83	90.1	6.7	6.74
Dibromochloromethane	$CHBr_2Cl$	124-48-1	208.29	120.0	2.0	4.00
Bromoform	$CHBr_3$	75-25-2	252.75	149.5	1.3	3.01

With more data on the long-term risk of THMs and other disinfection by-products, and the evolution of water treatments, it is expected that the MCL could be lowered even further, which justifies the need for very sensitive analytical techniques.

In addition to drinking water, the analysis of THMs has been reported in other types of water and aqueous media. Among them, we should mention the beverage industry,[22] seawater,[23,24] aquatic recreational and therapeutic facilities[24–26] and the power industry.[27,28] In this last field of activities, it has been demonstrated that $CHCl_3$ rapidly decomposes at the high temperature experienced in thermal power plants to give a stoichiometric amount of chloride,[28] which is a corrosion hazard for the materials used in steam–condensate cycles. In many cases it can easily exceed the safe operational limit, especially in nuclear plants, resulting in limited power production and even complete shutdown.

Although the THMs are regulated as a group in most countries, almost all the methods measure them separately and sum the concentration of the four compounds to obtain the required TTHM value. The volatility of these molecules makes them very good candidates for GC analysis and the high sensitivity and selectivity to halogenated compounds of the ECD, ELCD or MS detector permit one to achieve the very low DLs needed. The analytical approaches presented in this article differ in the way the THMs are introduced into the chromatographic column, and how they are isolated from the matrix and in many cases concentrated. Perhaps the simplest approach is the DAI, used successfully in 1977 by Nicholson et al.[29] and later developed by Grob and Habich[30,31] for capillary columns. This technique, which involves no sample pretreatment and needs no special introduction system, has been widely employed for the rapid screening of THMs. Another simple and conventional way to inject these analytes is after LLE with an organic solvent, usually pentane, proposed in the 1970s by various groups.[32–35] It offers the possibility of concentrating in a small volume of solvent the analyte present in a large sample. Owing to the volatile nature of THMs, another efficient way to isolate them is by gas extraction techniques. Among them, the most popular are HS[36] and PT methods.[37] In the former technique, the aqueous sample is heated in a sealed vial and an equilibrium is reached between the THMs present in the water and the HS above, from which an aliquot is injected into the GC column. In the latter, the THMs are purged out of the sample by an inert gas and concentrated on a solid sorbent, and later thermally desorbed and injected into the GC column.

Two very interesting reviews covering these techniques were published by Oliver[17] in 1989 and Biziuk and Przyjazny[11] in 1996. Both of them cover the wider range of the analysis of volatile organics. In this article, the four analytical approaches will be presented in their specific application to the analysis of THMs in drinking water and other aqueous matrices. After discussing the sampling and preservation of samples, along with standard preparation, each of the major analytical techniques will be described, with their performance, advantages and limitations.

2 STANDARD PREPARATION AND SAMPLE HANDLING

The analytical techniques used today are very sensitive with DLs in most cases below the micrograms and even at the nanograms per liter level. Reliable results at those levels can only be achieved by observing good laboratory practices and taking very meticulous care at every step of the analytical protocol, from the preparation of the sampling containers to the analysis by GC, and including the preparation of blank water and standards. Although some specific care might be required depending on the analytical technique used, the major precautions to be taken come from the very nature of the THMs and the composition of the aqueous matrix. Therefore, the same standard preparation and sample handling procedures can generally be applied to all of the analytical approaches.

2.1 Preparation of Blank Water and Standards

THMs are not easily removed by commercial water purification systems. Lépine and Archambault[38] have shown that even with a multiple-stage purification system, detectable amounts of THMs are still present. They also showed that some devices such as an ultraviolet (UV) lamp incorporated in the latest version of some commercial systems can even increase the level of THMs, probably by degradation of halogenated intermediates still present in the purified water. The only way they could obtain perfectly clean blanks (no detectable amount of any of the analytes) was by simply purging the water overnight with an inert gas in a 4-L sparger. The sparger was then kept slightly overpressured with helium to prevent atmospheric contamination. The same principle was later used successfully by other groups[23,39] for low trace analysis and is now one of the methods recommended by the USEPA, which also suggests boiling the water for 15 min before purging for 1 h at 90 °C.[40]

Primary standards of the four THMs, mixed or separately, may be purchased commercially at different concentrations between 200 and 2000 mg L^{-1}. They can also be prepared by the user but only ultrapure analytes and methanol should be used. Standards at the thousand milligrams per liter level can be kept for 1 month if stored in glass vials equipped with a Teflon™-faced

silicone septum filled to the brim (with no HS) and kept at 4 °C with no direct or intense light. Standards in the micrograms or nanograms per liter range must be prepared daily by successive dilution from the primary standard and also kept in vials with no HS. All glassware used in the preparation of nanograms per liter level standards should be checked with blank water prior to first use, specially marked and never used for other applications or higher concentration standards. These manipulations must be performed in an area known to be free of organic vapors and no halogenated organic solvents should be used or stored in the same laboratory. For quality control, a certified standard reference material can be purchased from the US National Institute of Standards and Technology (NIST).

2.2 Sample Collection and Preservation

The sampling procedure described in this section is based on USEPA methods 502.2, 524.2 and 551.1, which are the three methods prescribed by this organisation for analysis of THMs in treated water.[40] The two major concerns at this step are to avoid contamination and loss of volatile compounds by evaporation. Grab samples are collected in screw-capped glass containers equipped with a Teflon™-faced silicone septum. Alternatively, when headspace gas chromatography (HSGC) is used for analysis, the crimped-cap vials used for this technique can be used to bring the sample to the instrument without any further treatment. In this case, a special sampling procedure is required and will be described in section 5.2. Both the containers and septa are prepared prior to use by washing with detergent and rinsing thoroughly with distilled water or equivalent. They are air dried at room temperature, then placed in a 105 °C oven for 1 h, after which they are removed and allowed to cool in an area known to be free of organics. The vials are finally sealed and kept this way until sampling.

If samples, such as chlorinated water, are suspected to contain residual chlorine, a dechlorinating agent should be added into the vial before sampling to inhibit the chlorination reaction during sample storage. Many workers have observed an increase in THM concentration when no preservative was added.[41–43] The preferred dechlorinating agent according to the USEPA is sodium thiosulfate, but ascorbic acid may also be used, as demonstrated by Norin and Renberg.[41] A 3 mg amount of sodium thiosulfate or 25 mg of ascorbic acid are added per 40 mL of sample, but if chlorine is likely to be present at $>5\,mg\,L^{-1}$, its concentration should be determined and the same amount of dechlorinating agent added for each additional $5\,mg\,L^{-1}$ of residual chlorine. In USEPA method 551.1,[40] which uses an LLE procedure, ammonium chloride (6 mg per 60 mL of sample) is also suggested as a dechlorinating agent. Kawakami et al.[42] successfully used hydrazine sulfate and phosphoric acid to preserve samples for TTHM analysis over a period of 9 days. Badawy[43] observed a small loss in the recovery of THMs in the analysis by HSGC at 60 °C when sodium thiosulfate or sodium sulfite was used as dechlorinating agent, presumably because these species can react with an alkyl halide to form an *S*-alkylthiosulfonate and sulfonic acid. No loss in the recovery of THMs was reported by the author when potassium hexacyanoferrate(II) was used as dechlorinating agent.

At the sampling site, the vials must be filled to the brim, without any HS, to avoid contamination from the atmosphere at the sampling station and loss of analyte from the water before analysis. Care should be taken not to flush the dechlorinating agent and that no air bubbles pass through the sample or are trapped when the vial is sealed. The sample is acidified by adding two drops of 1 : 1 HCl and the vial is sealed, Teflon™ face down, and mixed for 1 min. The dechlorinating agent should not be mixed with HCl in the vial before sampling. The only exception is for HSGC analysis when the analytical vials are filled at the sampling site. In this case, only the specified amount of sample is introduced in the vial, leaving the required HS above the water (see section 5.2).

It should be noted that:

1. If only THM analysis is to be performed, the acidification step can be omitted if sodium thiosulfate is used as dechlorinating agent. This exception does not apply if ascorbic acid is used for dechlorination.
2. If the sample foams vigorously when HCl is added, the sample should be discarded. Another set of samples should be taken without adding the HCl, marked as ''not acidified'' and analyzed within 24 h of collection if compounds other than THMs are to be determined.

The samples must be chilled to 4 °C and maintained at that temperature until analysis. If proper dechlorination measures have been taken and there is no HS in the vial, samples can be stored for 14 days at 4 °C before analysis. If LLE is performed on the sample, the organic extracts should be kept in a freezer (<-10 °C) and analyzed within 14 days. The storage area must be free of organic solvent vapors and direct or intense light.

3 DIRECT AQUEOUS INJECTION GAS CHROMATOGRAPHY

The DAI approach is the simplest method of all because it does not require any sample pretreatment or special device for sample introduction into the GC column. It refers to the syringe injection of the water sample directly

into the gas chromatograph, without any extraction or concentration step. Nicholson et al.[29] in 1977 were the first to report the determination of THMs in water using the DAI technique. They injected a 9-μL sample into a 2-mm packed column through a hot injector. They found that the application of this technique to chlorinated water lead to concentrations of THMs significantly higher than the gas sparging technique. Different experiments demonstrated that those higher THM values were the result of the formation in the injector of these compounds by thermal degradation of precursors also present in the sample. Therefore, the results they obtained could be categorized as the total potential haloform that can be formed after chlorination rather than the true concentration in the sample. This phenomenon was studied a year later by Pfaender et al.,[44] who also came to the same conclusion. They showed that a value close to the true THM concentration could be obtained if the sample is injected after a 30-min purge and subtracting this value from the one obtained without purging.

The turning point in the development of the DAI technique was the use by Grob and Habich[30,31] in 1983–84 of a cooled injector for on-column injection into a capillary column. The water sample is injected in a cooled portion of the column, avoiding the thermal degradation of THM precursors in the injector. Grob has described in depth the different parameters that affect this mode of injection for aqueous samples.[31] Basing their work on his development, many workers have used the DAI technique for rapid screening of THMs alone[20,45] or with other halogenated organics[24,46] in drinking water.

For other more complex matrices, Kozlowski and Polkowska[47,48] used an isolation and enrichment technique called thin-layer headspace (TLHS) with self-generation of liquid sorbent. In this version of TLHS, the water sample is heated and the volatile compounds are extracted from the aqueous sample into the gas phase. They are then transferred by a stream of gas into a cooled thermostated tube where water vapor undergoes condensation along with the analytes. This technique produces an aqueous concentrate which is sufficiently pure to be analyzed by DAI. Polkowska et al.[22] successfully applied this technique to the analysis of THMs in commercial beverages, with a DL 10 times lower than that of the DAI technique alone.

A good review of the different aspects of the DAI technique has been published by Turska et al.[49]

3.1 Instrumentation

As already mentioned, the DAI technique requires a gas chromatograph with a cooled injector. Grob[31] pointed out that the best results are obtained with on-column injection, which avoids delayed water elution caused by diffusion and dilution in the injector cavity before the analytes enter the column. Even splitless injection gave much larger peaks than on-column injector. Grab explained this result by the fact that the solubility of the halocarbons is too low for the reconcentration effect to occur in the first portion of the column, which is a key part of this injection mode. As a result, the analytes migrate through the column in a rather large band, resulting in severely broadened peaks.

Column selection is also critical with this injection technique and must fulfill two primary requirements. First, it must elute the water rapidly and completely. This is achieved by nonpolar stationary phases such as methylsilicone. To improve the separation of the analytes, the methylsilicone stationary phase can be modified with small amounts of phenyl or vinyl groups. Second, the volatile analytes need to be sufficiently retained to come out long after the elution of water has ended. The use of the thickest possible stationary phase is the only way to achieve this goal. It provides the necessary strong retention of apolar halocarbons and a high loading capacity, avoiding excessive broadening of the water peak. However, this leads to low efficiency, broad peaks and reduced sensitivity. As a compromise, film thicknesses of 2–5 μm have generally been used for the analysis of THMs by DAI.[20,22,24,45,46] The stationary phase should be immobilized through cross-linking and chemical bonding to the column wall to avoid its hydrolysis by the large amount of water injected. Medium- and wide-bore columns (0.28–0.54 mm) are typical, since the injection of a large volume of water could block temporarily the gas flow in a small diameter column. A column length of 25–30 m is also the most often used, which represents a good compromise between resolution and loss in sensitivity through peak broadening. A retention gap, consisting of a deactivated fused-silica column (1–2 m) is recommended to protect the column from nonvolatile inorganic and organic compounds that might be present in the water sample and that will accumulate at the column inlet and degrade the resolving power of the column. This precolumn can be replaced after a specified number of injections or when performance degradation is observed by broadening of the water peak. The retention gap also serves to narrow the chromatographic bands and permits large-volume injections.

Finally, the ECD is the detector of choice because it combines very high sensitivity and selectivity for halogenated compounds, especially for the THMs which have three halogens per molecule.

3.2 Method Parameters

The optimal column temperature for isothermal runs is just above the boiling point of water, at about 103–104 °C,

as recommended by Grob.[31] At lower temperatures, the elution of water is delayed, resulting in poorer resolution of the lighter halocarbons. At higher temperatures, peak splitting was observed which was more severe with increasing molecular weight, but this phenomenon could not be explained. After the water has eluted from the column, temperature programming can be used for the separation of heavier compounds. The ECD is usually operated at high temperature (300–350 °C) and a makeup gas flow rate of ca. 25–50 mL min^{-1}. This allows for the fast elution of water through the detector cell and a narrower water band.

The calibration of the GC signal is easily performed by external standardization, injecting composite standards prepared in water or in any other suitable solvent. A recovery test should be performed with a fortified sample to check for matrix effects, specially with complex aqueous matrices.

3.3 Analytical Performance

The analytical performance obtained by different groups who have reported the use of DAI for the analysis of THMs in water are summarized in Table 2 along with their chromatographic conditions. The best results were obtained by an injection of 1–2 μL of water sample into a cold on-column injector or a SPI, a 26–30 m × 0.32 mm ID capillary column with a 5-μm film thickness and an ECD. With this configuration, the DL is typically between 0.01 and 0.06 μg L^{-1}. A typical chromatogram obtained under these conditions by Carmichael and Holmes[45] is shown in Figure 1, along with the analysis of a community drinking-water sample. The DLs obtained with a 0.32 mm ID column are an order of magnitude lower than those reported by Temmerman and Quaghebeur[20] using a megabore column (0.53 mm ID) and similar other conditions. Pyle and Gurka[46] obtained even higher DLs using a megabore column but with an ion-trap mass spectrometer, partly because they could not obtain reliable results using injection volumes greater than 0.2 μL, which is 10 times less than the volume injected by the other groups. Finally, the precision reported by the different groups is generally fairly good, with <5% RSD at the 10–20 μg L^{-1} level.

3.4 Advantages and Limitations

The main advantage of DAI is obviously its simplicity. It does not require any piece of equipment other than a gas chromatograph equipped with a cooled injector. Moreover, since no pretreatment or preconcentration step is performed, it eliminates the sometimes laborious step of sample preparation and the risk of either contamination or loss of very volatile analytes prior to the analysis. It is a fast analytical method that can be used for rapid THM screening and quality control in a water treatment plant and can be easily automated with the addition of an autosampler to the chromatograph.

The problem of stationary phase stability with regard to the large volume of water injected, encountered in the early development stage of the technique, has been

Table 2 Reported performance of THM analysis in water by DAIGC

Injection type [volume]	GC conditions	DL[a] (μg L^{-1})	Precision (% RSD) [concentration level]	Ref. [year]
Not mentioned [9 μL]	1.82 m × 2 mm ID Chromosorb® 101, ECD, isothermal 130 or 150 °C	0.8–3.0	1.5–7.5 [50 μg L^{-1}]	29 [1977]
Cold on-column [2 μL]	26 m × 0.32 mm ID × 5 μm PS-255, 1.5 m × 0.32 mm ID retention gap, ECD, isothermal 104 °C	0.02 [$CHCl_3$]	n.s.[b]	31 [1984]
Cold on-column [1 μL]	30 m × 0.53 mm ID × 5 μm DB1, 1.5 m × 0.53 mm ID retention gap, ECD, isothermal 104 °C	0.1–0.5	2–5 [10–100 μg L^{-1}]	20 [1990]
Varian 1093 SPI [1 μL]	30 m × 0.32 mm ID × 1.8 μm DB624, 4 m × 0.52 mm ID retention gap, ECD, 3 min at 104 °C, 10 °C min^{-1} to 150 °C	0.02–0.06	1.0–2.3 [20 μg L^{-1}]	45 [1990]
Finnigan SPI [0.2 μL]	30 m × 0.53 mm ID × 1.5 μm Restek XTI-5, ion-trap MS detection, 3 min at 40 °C, 12 °C min^{-1} to 148 °C	4–5	2–20 [20–50 μg L^{-1}]	46 [1994]
Cold on-column [2 μL]	30 m × 0.32 mm ID × 5 μm DB1, 2 m × 0.32 mm ID retention gap, ECD, isothermal 102 °C	0.01	1.8–3.0	24 [1996]

[a] Based on signal-to-noise ratio (S/N) = 3.
[b] Not specified.
ID, internal diameter; RSD, relative standard deviation; SPI, septum-equipped programmable injector.

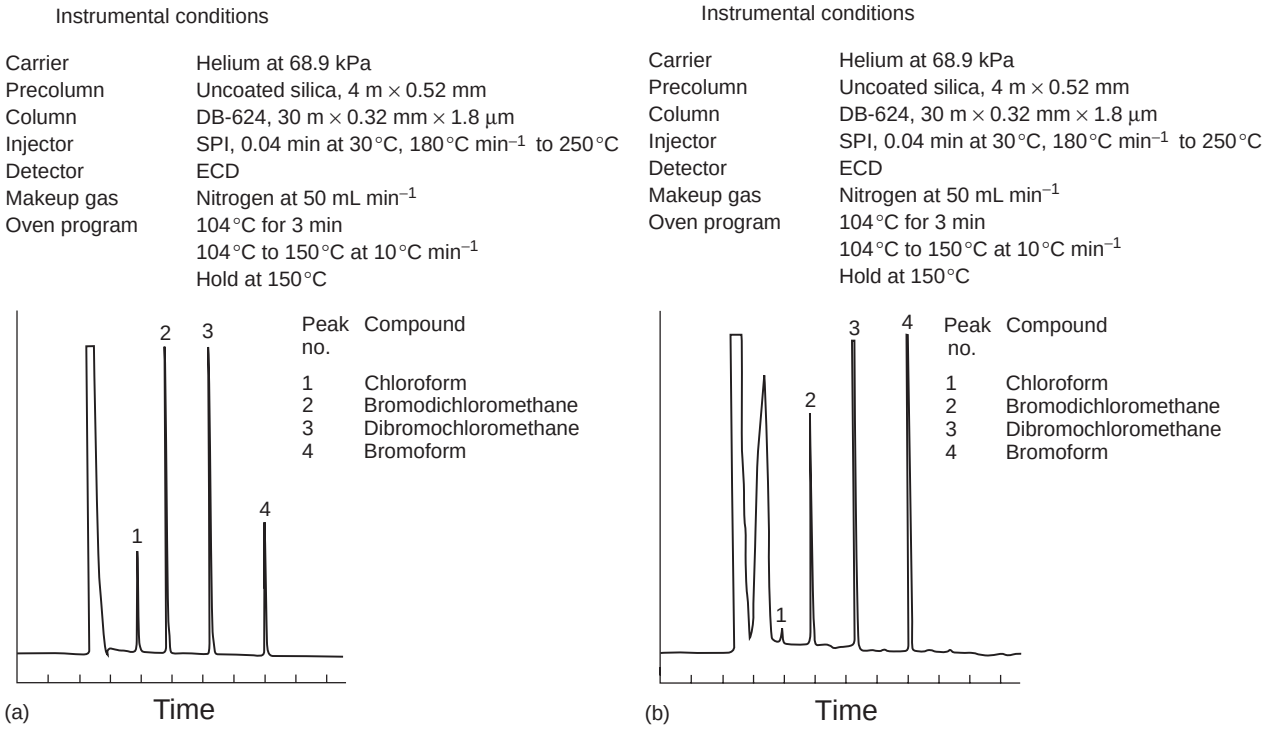

Figure 1 Analysis of THMs by DAIGC/ECD. (a) THMs standard, 60 pg of each compound injected; (b) community drinking water supply. [Reproduced by permission of Varian Limited from Carmichael and Holmes.[45]]

resolved by the advent of modern capillary columns with chemical bonding and cross-linking of the coating. Another problem experienced was the deterioration of column performance due to the deposition at the head of the column of nonvolatile inorganic and organic compounds present in the sample. This can be overcome by the use of a precolumn, which can be easily replaced after a preset number of runs. However, other compounds can nevertheless cause some problems and DAI should be used mainly for relatively clean samples, such as finished chlorinated water. The other fundamental drawback of the technique is the very small amount of sample that can be introduced, which limits its sensitivity. For these reasons, an enrichment and/or isolation technique, such as the previously mentioned TLHS, has to be performed in conjunction with the DAI technique for analyzing low THM concentrations and very complex matrices.

4 LIQUID–LIQUID EXTRACTION GAS CHROMATOGRAPHY

LLE is one of the oldest methods for the isolation of organic compounds from aqueous samples. It has often been used in the past as a reference to validate other analytical approaches and is still one of the methods recommended for THM determination in drinking water by the USEPA.[40] It is based on the principle that THMs, as other organic compounds, tend to partition preferentially from water into an organic solvent. The equilibrium between the two phases can be written mathematically as Equation (1):

$$K_d = \frac{C_s}{C_w} \tag{1}$$

where K_d is the distribution coefficient and C_s and C_w are the concentration of the analyte in the solvent and in the water sample, respectively. The extraction efficiency of the solvent for a given analyte is then represented by Equation (2):

$$E = \frac{100K_d}{K_d + (V_w/V_s)} \tag{2}$$

where E, the extraction efficiency, is expressed in percent and V_s and V_w are the volume of the solvent and the water sample, respectively. We can see from this equation that an efficiency close to 100% is obtained only if K_d is high and the ratio V_w/V_s is low. However, the latter situation is not desirable because it would imply that the volume of solvent is much greater than that of the water sample, which would dilute the analyte. In contrast, to obtain the increased sensitivity by concentrating the analyte in a smaller volume of solvent, the volume

ratio needs to be high, which will result in lower efficiency. In general, a good compromise is achieved with a water-to-solvent ratio of 10–25, with extraction efficiencies >70% at these ratios.[40,50–52] One practice that has proven effective in improving the recoveries is to add salts to the water sample to take advantage of the so-called ''salting-out'' effect. Increasing the ionic strength of the sample lowers the solubility of the THMs in the aqueous phase and increases the distribution coefficient.

The solvent used for extraction has to fulfill two criteria: it has to be immiscible with water and the THMs must be considerably more soluble in it than in water (high K_d). Many solvents have been used by different research groups, the most common being pentane.[16,50–55] It is preferred by most investigators because it is highly volatile and elutes from the GC column before the THMs. However, its high volatility represents a potential risk of solvent losses and has prompted other workers to use less volatile solvents such as hexane,[13,55,56] hexane–diisopropyl ether (1 : 1),[41] methylcyclohexane[52,55] and higher molecular weight alkanes.[52] All these solvents can be obtained relatively pure and the few researchers[52,55] who have systematically compared these different solvents with the same extraction procedure concluded that they give comparable results, and that the choice of a particular solvent rests on the preference of the analyst. The USEPA, in its method 551.1,[40] uses methyl *tert*-butyl ether (MTBE) but proposes pentane as an alternative.

4.1 Instrumentation

As for the DAI technique, the only piece of equipment required with LLE is a gas chromatograph. The best results have been obtained with an on-column or splitless injection and an ECD. The injection volume of the extract is generally between 0.5 and 2 μL. The choice of the column and the temperature program will depend on the presence of interfering species extracted from the sample along with the THMs. In the most recent papers, medium- or narrow-bore capillary columns (0.25–0.32 mm ID) with a standard film thickness (~1 μm) have been used to obtain high resolution and narrow peaks.

The use of an ECD is highly recommended, because of its high selectivity for halogenated compounds and its

Table 3 Reported performance of THM analysis in water by LLEGC

Solvent	Phase ratio V_w/V_s	Extraction efficiency (%)	Injection volume (μL)	GC conditions	DL[a] ($\mu g\,L^{-1}$)	Precision (% RSD)	Ref. [year]
Hexane	30	n.s.[b]	1.0	90 cm × 2.2 mm ID Tenax® 60–80 mesh, ECD, 2 min at 140 °C, 30 °C min^{-1} to 185 °C	0.1–0.5	1–12	13 [1979]
Pentane	10	80–93	n.s.[b]	183 cm × 4 mm ID 10% Squalane on Chromosorb® 80–100 mesh, ECD, isothermal 67 °C	0.1–0.5	6–8	52 [1979]
Pentane	23	72–83	n.s.[b]	183 cm × 4 mm ID 10% Squalane on Chromosorb® 80–100 mesh, ECD, isothermal 67 °C	0.1–0.5	4–6	52 [1979]
Methylcyclohexane	5	86–92	n.s.[b]	183 cm × 4 mm ID 10% Squalane on Chromosorb® 80–100 mesh, ECD, isothermal 67 °C	0.2–0.8	3–5	52 [1979]
Hexane	38	70–83	2.0	15 m × 0.5 mm ID OV17 SCOT, ECD, 3 min at 60 °C, 8 °C min^{-1} to 150 °C	1	2–3	56 [1981]
Pentane	25	72–84	0.5	25 m × 0.2 mm ID SE-54, ECD, isothermal 42 °C	0.1	3–4	50 [1982]
Pentane	23	95–106	1.0	VOCOL 30 m × 0.53 mm ID × 3 μm, ECD, 8 min at 30 °C, 2 °C min^{-1} to 90 °C	0.4–0.9	8.7–18	51 [1992]
Pentane + NaCl	10	85–99	2.0	30 m × 0.25 mm ID × 1 μm DB-1, ECD, initial 15 °C, 2 °C min^{-1} to 50 °C, hold 10 min, 10 °C min^{-1} to 225 °C, hold 15 min	0.03–0.08	3–4	40 [1995]
MTBE + NaCl	17	97–102	2.0	30 m × 0.25 mm ID × 1 μm DB-1, ECD, 22 min at 35 °C, 10 °C min^{-1} to 145 °C	0.005–0.08	0.7–1.3	40 [1995]

[a] Based on S/N = 3.
[b] Not specified.

high sensitivity for the THMs. Also, the low response of this detector to the alkane solvents generally used in LLE is a net advantage of this detector.

4.2 Method Parameters

A volumetric flask has been suggested by some workers[34,41] to perform the extraction. With the water sample to the mark, the required amount of solvent is added and the flask shaken manually. As a result of the narrow neck, the organic extract is easily removed from the top of the aqueous sample since all the suggested solvents have densities lower than that of water. Another convenient and more efficient way to perform the extraction, especially with many samples, is in septum-sealed vials on a mechanical or orbital shaker. The sampling vials in which the samples are stored can even be used for the extraction by removing a certain amount of sample and adding the required amount of solvent. The exact volume of sample extracted is determined by subtracting the weight of the empty vial from the weight with the sample before adding the solvent.

The determination is generally performed by extracting aqueous composite standards with the same procedure as for the samples. These standards must be prepared by spiking a known amount of the analytes directly in the extraction vessel containing blank water. This is necessary to avoid loss of the volatile analytes during transfer of the standard solution. The extraction efficiency should be

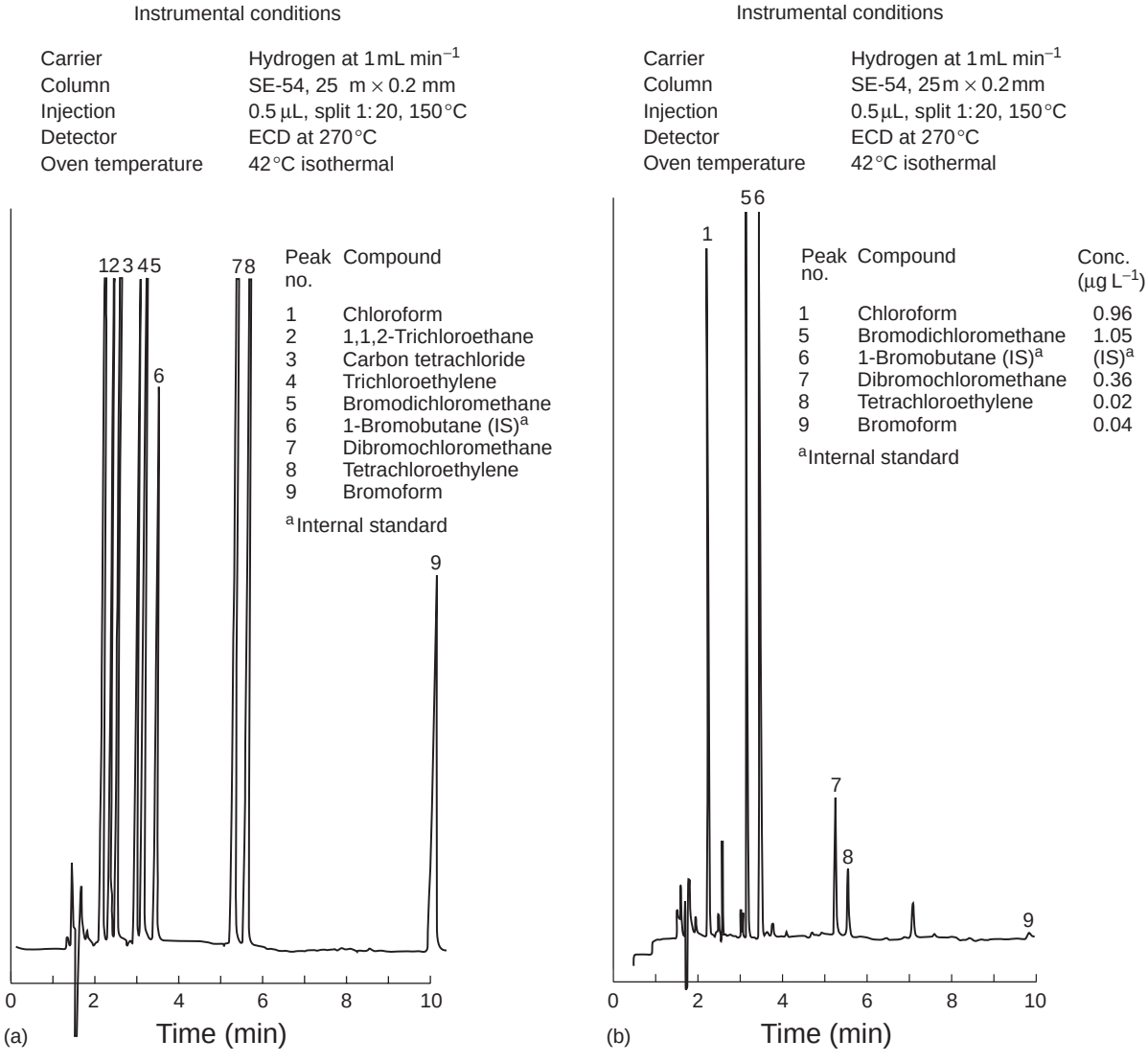

Figure 2 Analysis of organohalides by LLEGC/ECD. (a) Standard mixture of 0.1 µg L^{-1} each in pentane, 2 mL equilibrated with 50 mL of water; (b) organohalides from Lake Constance water supply, sampled in Tübingen, Germany. [Reproduced by permission of Wiley-VCH from Hagenmaier et al.[50]]

checked by extracting fortified samples, especially with complex aqueous matrices or whenever a matrix effect is suspected. A surrogate analyte, not present in the matrix, could also be added to the extraction vessel to monitor the method performance with each sample.

4.3 Analytical Performance

Table 3 presents the extraction efficiencies and the DLs and precisions reported by different analytical groups with the LLE technique. Most of them used pentane as the extraction solvent. The extraction efficiencies are satisfactory, with values between 70 and 90%. The DLs lie in the range $0.1-1\,\mu g\,L^{-1}$. The precision is fairly good with values generally between 2 and 6%. Figure 2(a) and (b) show a typical chromatogram of a pentane extract of a municipal water supply obtained with a $25\,m \times 0.2\,mm$ ID column[50] along with one of the calibration standards. The 10-min run is sufficient to separate the four THMs along with five other organohalide compounds. The best results are those presented in the 1995 edition of USEPA method 551.1, which uses on-column or splitless injection in a narrow-bore capillary column. Their DL is at least one order of magnitude lower than that published earlier, with very good precision. In their method, NaCl is also added to the sample to increase the extraction efficiency. However, it has been noted recently that elevated recoveries of brominated THMs (by synthesis) have been observed owing to the inevitable presence of bromide impurities in the NaCl.[40] For that reason, the authors strongly recommend the use of Na_2SO_4 as a replacement for NaCl.

4.4 Advantages and Limitations

The major advantage of the LLE method is that, as in the DAI approach, it does not need any apparatus other than a gas chromatograph. In addition, it eliminates the problems associated with the injection of a large quantity of water into the GC column and also isolates the analytes from the inorganic impurities present in some matrices, although organic compounds soluble in the extraction solvent might interfere. It has proven effective in concentrating the THMs from a large sample into a small volume of solvent for higher sensitivities.

However, this technique implies a tedious extraction step which cannot be completely automated and requires the intervention of the analyst during the process. It also involves the use of expensive, high-purity solvents, which might be toxic to handle and represent a hazardous waste. The risk of contamination or solvent losses by evaporation during the extraction is also very high and meticulous care in the extraction process, together with good quality control of the process, are mandatory. Finally, the formation of emulsions with certain matrices is yet another problem with this technique.

5 HEADSPACE GAS CHROMATOGRAPHY

The volatility of the THMs and their tendency to partition into the gas phase over aqueous solutions have been used by some analysts to extract these compounds in water samples. The HS analysis technique refers to the determination of the analyte concentration in the gaseous phase being in thermodynamic equilibrium with the sample. Although some authors have used the terms ''dynamic HS'' for techniques where the gas passes continuously over the surface of the water sample or bubbles through the liquid (which will be discussed in the next section), the word HS will be strictly applied in this article to designate the sampling of a static gaseous phase in equilibrium with the sample in a sealed container, as is the practice in the literature.

A comprehensive book by Kolb and Ettre[57] re-edited in 1997 describes in detail the theoretical background and practical aspects of HSGC. In similar terms to those for LLE, the distribution of the analyte between water and gas phase at a given temperature is governed by the partition coefficient K_p, which is a fundamental parameter of an analyte under given conditions and is generally defined in the HS literature by Equation (3):

$$K_p = \frac{C_w}{C_g} \tag{3}$$

where C_w and C_g represent the concentration of the analyte at equilibrium in the water and gas phase, respectively. Knowing that $C = W/V$, a mass balance of the system can be expressed according to Equation (4):

$$\begin{aligned} C_0 V_w &= C_g V_g + C_w V_w = C_g V_g + K_p C_g V_w \\ &= C_g (K_p V_w + V_g) \end{aligned} \tag{4}$$

or Equation (5):

$$C_0 = C_g \left(K_p + \frac{V_g}{V_w} \right. \tag{5}$$

where C_0 is the original concentration of the analyte in the aqueous sample and V_g and V_w are the volume of the gas and water phases, respectively. Introducing the phase ratio β, given by Equation (6):

$$\beta = \frac{V_g}{V_w} \tag{6}$$

Equation (5) becomes Equation (7)

$$C_0 = C_g (K_p + \beta) \tag{7}$$

In a given system and under a given set of conditions, both K_p and β are constant, hence the original concentration in the sample C_0 is proportional to the concentration in the

gas phase at equilibrium, C_g, which can be determined by injecting an aliquot of the gas phase into the gas chromatograph. This is the basis of quantitative HSGC.

The use of the HS technique for THM analysis was reported by Kaiser and Oliver[36] in 1976, using a 60-mL separating funnel immersed in a thermostated water bath and filled with the sample to leave an HS of 2 mL, which was sampled with a gas-tight syringe. A similar set-up was also used by Piet et al.[58] with a 600-mL sample and a 100-mL HS. However, a more convenient way to perform HS sampling is in a septum-sealed vial, where the HS can either be sampled manually with a gas-tight syringe or transferred directly to the GC column. Kolb et al.[59] described the use of an automatic electropneumatic sampling system where the vial containing the sample is pressurized before the HS is injected into the GC column through a standard split injector. This electronically controlled pneumatic dosing system gave a typical reproducibility of 1–1.5% and could be automated for several vials. Otson,[60] a few years later, used a commercial GC autosampler with standard 2-mL vials for performing HSGC of THMs analysis at ambient temperature. Other more recent papers on THM analysis by HS have compared the results obtained with other techniques such as LLE[54,61,62] or PT.[63,64] They showed comparable analytical results in terms of accuracy and precision. The sensitivity was similar to that of LLE but lower than that of the PT technique.

Finally, a variation of HS sampling worth mentioning is a technique called solid-phase microextraction (SPME), proposed by Zhang and Pawliszyn,[65,66] which uses a fused-silica fiber coated with a suitable stationary phase contained inside the needle of a special syringe. During SPME, the volatile compounds concentrate on the fiber, which is lowered into the HS of the vial for a predetermined length of time. Once sampling is completed, the fiber is withdrawn into the syringe needle and immediately inserted into the GC injector, where the analytes are thermally desorbed and injected into the GC column. Although very good results have been obtained with DLs at the 0.01 $\mu g\,L^{-1}$ level and precision <5%,[66,67] more data are needed on the ruggedness of the technique.

5.1 Instrumentation

Today, most chromatographic companies are selling dedicated automatic HS samplers with precise control of the equilibration time and temperature and the pressure in the vial prior to the injection. Some also incorporate additional features, such as vial vibration, which shorten the equilibration time for certain viscous samples. There are three basic types of HS sampling systems: gas-tight syringe, time-controlled balanced pressure and the pressure/loop system. The first case is the simplest of all, where an aliquot of the gas phase is taken with a gas-tight syringe and subsequently injected through the septum of a standard injection port. It can be performed manually or with an automated sampler. The major problem of this approach is that the pressure and the temperature of the sample, and hence the amount injected, are not controlled and actually change in the syringe during the transfer because of expansion through the needle. This problem can be partly overcome in manual injection by the use of a Luer-lock valve. Automated systems usually have a heated syringe but do not have the possibility of locking the pressure. An interesting approach described by Kold and Ettre[57] is the combination of a gas syringe and a heated gas sampling valve. The syringe is used to pull the gas sample through the sample loop of a six-port sampling valve, which is then actuated to inject the aliquot into the column. The amount injected is determined by the temperature and pressure of the sample in the sampling loop, its volume being fixed by the loop dimension.

The other two instrumental approaches are schematized in Figure 3(a) and (b). In the balanced-pressure system, the vial is pressurized by the carrier gas, after equilibrium has been reached, to a pressure equal to the inlet pressure of the column. Next, the carrier gas supply is interrupted by closing valve V1, and the pressurized HS in the vial expands to reach the column. Since both the pressure and time for this operation are set parameters, the amount of sample introduced into the column can be accurately controlled. With this system, it is possible to transfer a large HS volume to the GC column to increase the sensitivity. However, cryogenic refocusing of the analytes at the head of the column is necessary to avoid peak broadening. Large amounts of water vapor would also be transferred and would have to be removed in order to avoid blockage of the column by ice. Kolb et al.[68] showed a set-up incorporating an LiCl trap to remove most of the water before the cooled portion of the column. Finally, a separate gas supply can be used for the pressurization, thus allowing balanced-pressure or increased-pressure injection (above the column inlet pressure) to be performed.

The pressure/loop system (Figure 3b) also starts by pressurizing the gas phase in the vial, usually with the same inert gas as the one used as carrier in the gas chromatograph. The HS is later vented to the atmosphere through the injection loop which is then actuated to put the loop in line with the gas chromatograph and inject its content into the analytical column. The pressure in the loop can be adjusted from atmospheric to the pressurized level by proper setting of the backpressure regulator. With the pressure/loop system, large-volume injection with cryogenic refocusing cannot be performed as easily as with balanced-pressure systems, but could be accomplished by

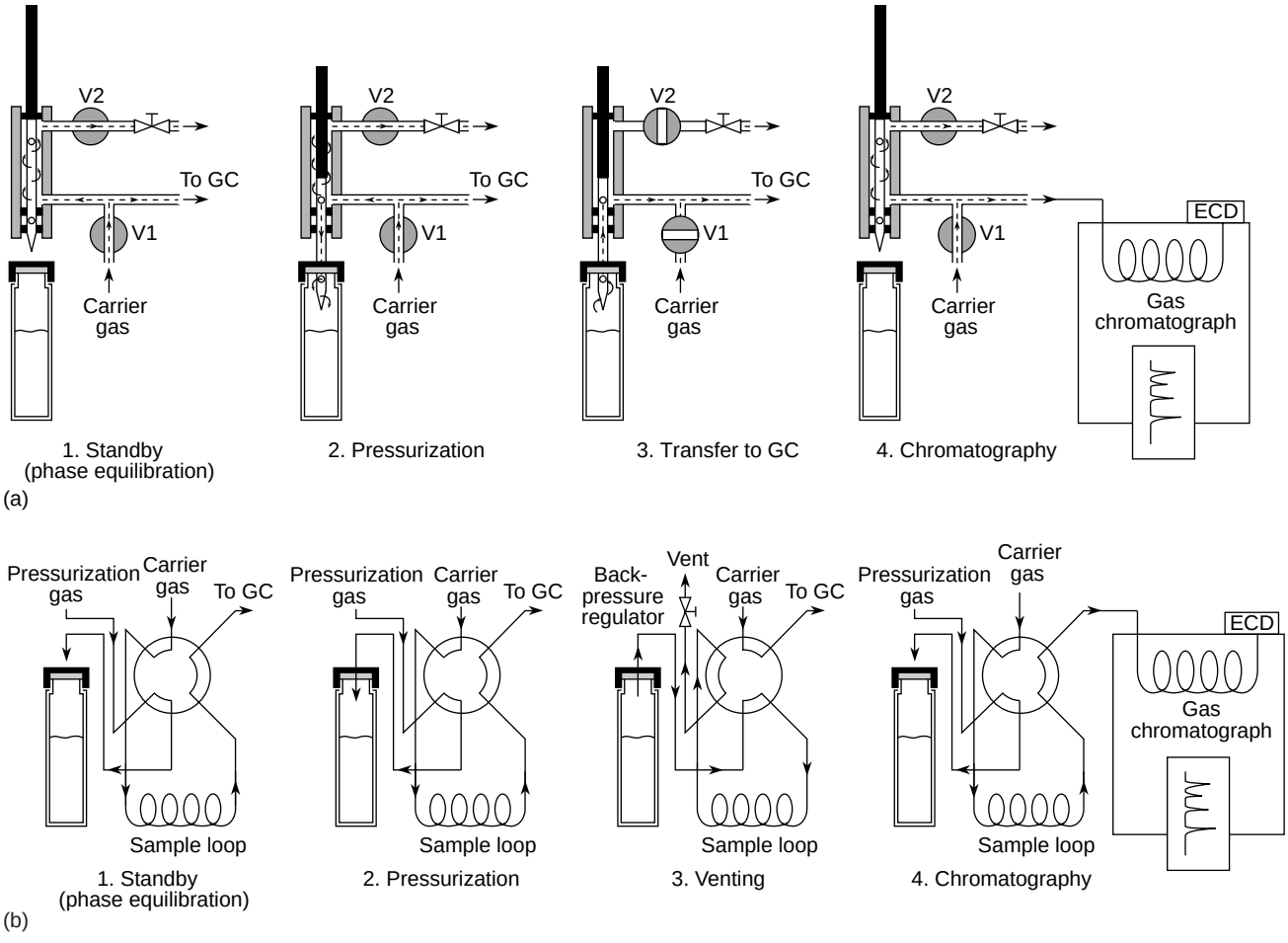

Figure 3 Instrumental set-up involved in HSGC/ECD. (a) Balanced-pressure system; (b) pressure/loop system. [Adapted in part from B. Kolb and L.S. Ettre, *Static HS–GC*, Wiley-VCH, New York, 1997, Copyright John Wiley & Sons, Inc., 1997. Reprinted by permission of Wiley-VCH, Inc. a subsidiary of John Wiley & Sons, Inc.]

the multiple-HS-injection technique, repetitive injection from a single vial or from multiple vials.

The HS approach, which introduces the sample already in a gaseous state, does not cause any constraint on the choice of the GC column. Unless a very high resolution is required because volatiles co-eluting with the THMs are present in the matrix, a medium-to-wide-bore capillary column (0.32–0.53 mm ID) with a slightly polar stationary phase (DB5 or equivalent) is recommended. The high capacity of these columns can accommodate large-volume injection and the higher gas flow facilitates the transfer of the gas aliquot from the HS unit, giving faster analysis. Finally, the high sensitivity of the ECD makes it the best candidate for THM detection.

5.2 Method Parameters

One special feature of HS analysis is that the vials used for analysis can be filled at the collection site. The vials should first be sealed and flushed with an inert gas in order to avoid atmospheric contamination at the sampling site. The water is sampled with a syringe and introduced into the vial through the septum. A second needle is required when the vial is being filled up to release excess pressure. The second needle is left in place one or two seconds after fill-up in order to equilibrate the HS to atmospheric pressure. This is necessary to ensure that the same pressure is present in all the sample vials and in those containing the standard solutions, which will normally be filled at the laboratory under atmospheric pressure. The amount of sample introduced into the vial can be determined by noting the volume on the syringe, or more precisely by the variation in weight of the vial before and after sampling.

In the HS systems, the pressure, temperature and sequence timing must be precisely set in order to obtain reproducible results. Even though only a few seconds would be enough to pressurize the vial, a longer pressurization time (1–3 min) is recommended to achieve

Table 4 Reported performance of THM analysis in water by HSGC

Temperature (°C)	Phase ratio V_g/V_w	Gas injected (μL)	GC conditions	DL[a] (μg L^{-1})	Precision (% RSD) [concentration level]	Ref. [year]
70	2 mL/60 mL	5	1.8 m × 4 mm ID 10% OV-1 glass column, ECD, isothermal 50 °C	0.1–1.0	n.s.[c]	36 [1976]
30	100 mL/600 mL	100	50 m × 0.5 mm ID OV-225 glass column, ECD, isothermal 22 °C until $CHBr_2Cl$ elutes, 4 °C min^{-1} to 80 °C	0.1–0.5	<5 [5–25 μg L^{-1}]	58 [1978]
Ambient	0.25mL/1.75 mL	5	15.3 m × 0.51 mm ID OV-17 SCOT stainless-steel column, ECD, 3 min at 60 °C, 8 °C min^{-1} to 150 °C, hold 6 min	5–10	<5 [10–50 μg L^{-1}]	60 [1981]
90–95	25 mL/100 mL	n.a.[b]	25 m OV-101 capillary column (ID unknown), ECD, cryogenic condensation of 0.2 mL, injection of 0.1 mL splitless, 2 min at −20 °C, 4 °C min^{-1} to 80 °C	0.001	14–22	62 [1983]
n.s.	3 mL/10 mL	100–500	3 m × 3 mm ID Chromosorb® with 20% SF-96 + 20% silicone DC-550 (8 + 2), ECD, isothermal 110 °C	0.1–0.4	n.s.[c]	61 [1984]
60	2 mL/20 mL	n.s.[c]	20.3 cm × 3 mm OD 60–80 mesh Carbopack BSP-800, ELCD, 2 min at 60 °C, 3 °C min^{-1} to 160 °C, hold 10 min	0.2–0.6	6–9	43 [1992]
80	5 mL water (gas phase not specified)	n.a.[b]	50 m × 0.32 mm ID × 2 μm DB5 type, ECD, cryogenic trapping for a 2-min transfer, 5 min at 40 °C, 8 °C min^{-1} to 90 °C	~0.001	n.s.	57 p. 97 [1997]

[a] Based on S/N = 3.
[b] Not applicable.
[c] Not specified.
OD, outer diameter.

proper homogenization of the HS. The second variable adjustable by the operator is the temperature of the six-port valve and the transfer lines from the vial to the valve and from the valve to the gas chromatograph. It should be high enough to prevent any condensation of the sample, but not so high that the septum is burned by a hot needle or that thermal degradation and/or oxidation of sensitive analytes occur. Usually, the temperature is set at or slightly above the thermostating temperature. Finally, the timing sequence and especially the transfer time in a balanced-pressure system need to be precisely controlled in order to obtain reproducible amounts of analytes injected into the column. This is well achieved by the commercial HS instruments now available.

It can be seen in Equation (7) that for a given C_0, the highest C_g and consequently the highest sensitivity will be obtained by minimizing both K_p and β. However, since improvement in sensitivity is directly related to the sum of these two parameters, the impact of working on one of them will depend on their respective absolute values. For example, since for the THMs $K_p \gg \beta$, moving towards lower K_p values will have a greater impact than changing β. Also, the latter parameter can only be reduced to a certain extent because there must be enough HS volume available for transfer into the column (typical volumes injected are 0.5–2 mL). One also needs enough space to ensure that the needle of the HS sampling system will not get into the liquid phase. For these reasons, in a typical 22-mL vial, the sample usually does not exceed about 80% of the vial volume, corresponding to a minimum phase ratio of 0.25. Since the THMs have different K_p values, minimizing β will have a greater impact on $CHCl_3$ than $CHBr_3$ ($K_p = 2.3$ and 12.8, respectively, at 65 °C).[69]

It can be derived from Henry's law that K_p is inversely proportional to the vapor pressure and the activity coefficient of a particular analyte.[57] Therefore, increasing the temperature, and consequently the vapor pressure, will decrease K_p and enhance the sensitivity. Kaiser and Oliver[36] observed an exponential sensitivity increase for the $CHCl_3$ peak from 30 to 90 °C, with a 10-fold increase between these two temperatures. However, they chose to work at 70 °C to minimize the amount of

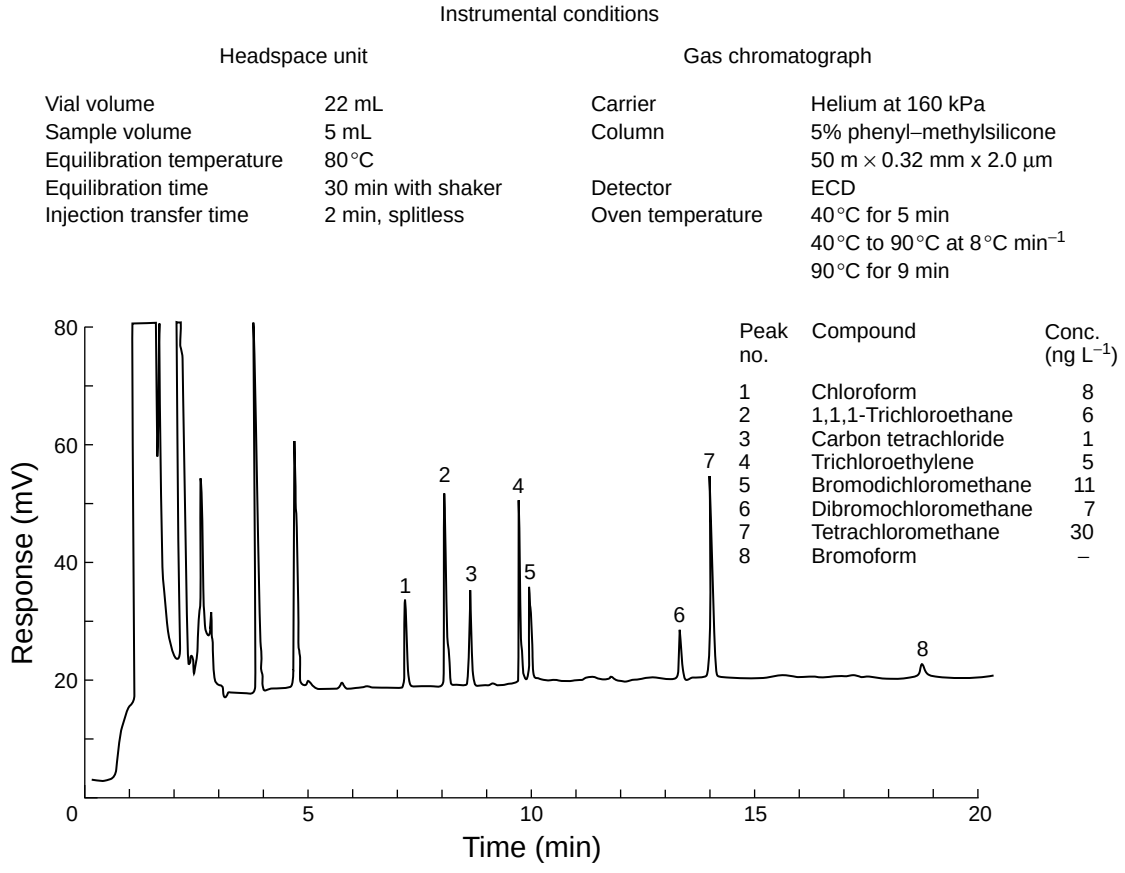

Figure 4 HSGC/ECD analysis of volatile halogenated hydrocarbons in mineral water using a balanced-pressure system with cryofocusing and a water trap. [Reproduced from B. Kolb and L.S. Ettre, *Static HS–GC*, Wiley-VCH, New York, 1997, Copyright John Wiley & Sons, Inc., 1997. Reprinted by permission of Wiley-VCH, Inc. a subsidiary of John Wiley & Sons, Inc.]

water vapor transferred with the analyte, which adversely affects the reproducibility of the results. It should be noted that the influence of temperature is an analyte-specific function and must be evaluated separately for each case. Another way to minimize K_p is to increase the activity coefficient of the analyte by the addition of an electrolyte to the sample, commonly called the ''salting-out'' technique. This effect has been mentioned in a few papers[57,63,64] for the THMs and increases inversely with their respective solubility, the more pronounced sensitivity improvement being with $CHBr_3$. However, it should be mentioned that high salt concentrations (20–50%) are needed to obtain a significant enhancement in sensitivity, which can typically range from one to five times the signal without salt addition. Salting-out also influences the linear range of various compounds and the technique should be investigated for each analyte before being incorporated into routine quantitative analysis.

The determination is generally performed by analyzing aqueous composite standards. To avoid loss of the volatile analytes during transfer, these standard solutions are best prepared by spiking a known amount of the analytes directly in the HS vial containing blank water and prepared with exactly the same procedure as for the samples. It is especially important to have the same volume ratio and initial pressure in the vial for the standards and the samples. Also, since the matrix composition can affect K_p, the recovery of the different analytes obtained with the spiked standards should be compared with those for fortified samples, especially with complex aqueous matrices containing a high salt content. A surrogate analyte, not present in the sample, could also be added to detect changes in matrix composition that could affect K_p.

5.3 Analytical Performance

The analytical performance reported by different groups who have used HS for the analysis of THMs are summarized in Table 4. The precision is satisfactory and is usually <10%. The DL is generally between 0.1 and 1 $\mu g\,L^{-1}$, which is comparable to the DAI or LLE technique. Onodera et al.[61] compared, on the same chromatographic system, HS with LLE and obtained

similar sensitivities for the two techniques. However, the results were obtained using packed columns on rather old systems with a limited amount of the gas phase injected into the column. Comba and Kaiser[62] obtained a DL one to two orders of magnitude lower by transferring a large HS volume and cryogenically condensing it in a separate smaller vial immersed in liquid nitrogen, with subsequent injection into the GC column. Kolb and Ettre[57] showed a chromatogram of mineral water analysis obtained with cryofocusing in a 0.32 mm ID capillary column (Figure 4). The concentration of the THMs, except for $CHBr_3$ which was not quantified, were between 7 and 11 ng L^{-1}, and judging by the S/Ns of the peaks in the chromatogram, the DL can be estimated to be 1 ng L^{-1}. Finally, the linearity was shown by Castello et al.[54] to extend over two to three orders of magnitude, thus simplifying the calibration step for different samples.

5.4 Advantages and Limitations

The major advantage of the HSGC technique is its simplicity of operation. The operator only has to put the vials in the sampler and start the run. With the commercial equipment available, the analysis of many samples can be automated, which makes it possible to run the system unattended and to perform the analysis of a large number of samples around the clock. Also, the fact that extraction is performed in disposable containers simplifies the preparation and cleanup for a series of analyses and minimizes the risk of contamination or carryover between samples. Finally, the isolation of the THMs from most of the water and the inorganic or nonvolatile organic compounds that might be present in the sample makes the chromatography simpler than DAI or LLE.

The only disadvantage of HS is that it requires an additional piece of equipment interfaced to the gas chromatograph to perform the HS extraction. Also, the time required to reach equilibrium adds to the chromatography run time for a particular sample. However, modern equipment can perform equilibration of the next sample concurrently with the chromatographic run. The total analysis time per sample is thus the longer of the equilibration time and the GC run time.

6 PURGE-AND-TRAP GAS CHROMATOGRAPHY

As with the HS technique, PT takes advantage of the high volatility of the THMs to isolate these compounds from the aqueous matrix into a gas phase. Both techniques are, in essence, a gas extraction, where the amount of each compound in the gas phase is dependent on its vapor pressure and its solubility in the sample (K_p). However, instead of isolating a portion of the analytes in a closed HS above the sample, purging the sample constantly sweeps the extracted compounds away to a trap, where they accumulate. Under these circumstances there is no migration of components from the vapor back to the liquid phase. This means that the partial pressure of any individual component above the sample at any time is essentially zero, which encourages an even greater migration of the volatiles into the vapor phase. The amount of analyte extracted from the sample is no longer limited to the equilibrium concentration given by the partition coefficient and, in theory, the totality of a particular volatile compound could be extracted. This can be performed by simply passing a stream of inert gas above the sample, but a much more efficient way of doing it is by passing the gas in the form of tiny bubbles through the liquid phase, thus considerably increasing the interface between the two phases.

The technique was first developed in 1974 by Bellar and Lichtenberg,[37] who designed a purging device made of a specially shaped glass tube equipped at the base with a medium-porosity glass frit, which allowed finely divided gas bubbles to pass through the sample. They showed that slightly soluble (<2% soluble) volatile compounds, such as $CHCl_3$, could be quantitatively extracted from a 5-mL water sample using <150 mL of nitrogen. The extracted volatiles passed through a trap made of a short section of stainless-steel tubing packed with an adsorptive material such as chromatographic-grade porous polymers (Tenax®, Chromosorb®, etc.), silica gel or molecular sieve. The trapped compounds were thermally desorbed and transferred to the GC column for analysis. The trap-and-release process was quantitative for $CHCl_3$ and other compounds having a high retention affinity for the adsorbent used, while water vapor passed through the sorbent bed with no adsorption when porous polymer was used. With a packed GC column and flame ionization detection (FID), they claimed a DL of approximately 0.5 μg L^{-1} for many compounds. THM analysis by the PT approach, along with other halogenated organic compounds, was published by the same authors in 1981 as a USEPA method.[70]

Dressman et al.[52] used ECD and compared the PT with the LLE approach for the analysis of THMs in water. They obtained comparable results and the DLs for all four THMs were between 0.1 and 0.2 μg L^{-1}. Other groups have tried to simplify the technique by eliminating the adsorbent trap and adsorbing the purged organics at the beginning of a packed GC column at close to ambient temperature.[71,72] Their analytical performances were essentially comparable in terms of sensitivity and precision to those of Dressman et al.

The advent of capillary columns with sharper peaks and consequently higher signals permitted the sensitivity of

GC analysis to be improved significantly. However, since the carrier flow rate is considerably reduced with these columns, the desorption process on the PT trap is slower and some sort of analyte refocusing at the head of the column, generally by cryogenic trapping, must be used to prevent peak broadening and avoid losing the benefit of the capillary column. Trussell et al.[73] with MS detection reported DLs "at the low nanograms per liter" level without giving specific numbers, but the precision even at the micrograms per liter level was only fair. The best results were published by Lépine and Archambault,[38] with DLs of 0.01–0.05 ng L^{-1} and precisions of 2.7–4.5% at the 1 ng L^{-1} level. They used ECD with liquid-nitrogen cryotrapping in the first centimeters of a 0.53 mm ID column. They also emphasized the need for special attention to the blank and standard preparation, along with sample handling, which are all critical parts of the analytical protocol at such low concentrations. Connan et al.[23] later used a similar set-up to measure THMs and other halogenated organics at the nanograms per liter level in a 30-mL seawater sample. Another study worth mentioning is that of Zygmunt,[39] who reported a DL of 1 ng L^{-1} and good precision by using a sorbent microtrap instead of cryogenic focusing at the head of the analytical GC column. Finally, Wang et al.[18] used a specially designed sparger and a portable gas chromatograph for monitoring THMs in continuously flowing tap water, with a new analysis performed every 15 min.

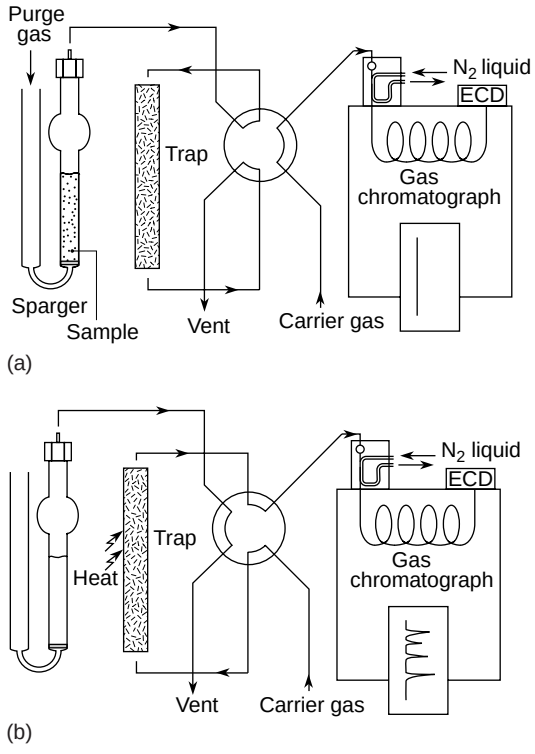

Figure 5 Instrumental set-up involved in PTGC/ECD. (a) Purge mode; (b) desorb and injection mode. [Reprinted with permission from L. Lépine, J.-F. Archambault, *Anal. Chem.*, **64**, 810–814 (1992).[38] Copyright 1992, American Chemical Society.]

6.1 Instrumentation

Commercial units are now available to perform PT analysis and a schematic diagram of a typical instrumental set-up is shown in Figure 5(a) and (b). The PT module is composed of a glass sparger, an adsorption trap with a heating jacket and a six-port valve. Many trap compositions are sold commercially for different applications but, if only THMs are being analyzed, a Tenax® trap is the most often used because this material does not adsorb water and has good adsorption/desorption characteristics for these analytes. The six-port valve and the transfer lines from the trap to the valve and from the valve to the gas chromatograph are heated to avoid condensation. The transfer line from the PT unit to the gas chromatograph is ideally connected directly to the analytical column with a zero-dead-volume (ZDV) connector. A cryofocusing unit is optional but highly recommended when working with capillary columns, especially medium- and narrow-bore columns. Alternatively, the analyte refocusing can be performed by cooling the GC oven to subambient temperature if the chromatograph has this option.

There are no special requirements for the GC column and its choice will be dictated by the type of matrix analyzed in order to obtain adequate resolution of the THMs from interferents. However, larger bore columns allow a higher gas flow, which reduces the desorption time and the transfer of the analytes from the trap to the gas chromatograph. Nonpolar or slightly polar phases such as methylsilicone or 5% phenyl–methylsilicone are the most often mentioned in the literature. Many detectors can be used, but most of the work published has been performed with an ECD because of its great sensitivity and selectivity for halogenated compounds, along with its wide dynamic range.

6.2 Method Parameters

The following procedure is performed in a PT injection analysis. In the purge mode (Figure 5a), the purge gas flows through the sparger and carries the analytes to the trap after which a dry purge of few minutes, where the purge gas flows directly through the trap, may be performed. The latter step is recommended to remove excess water physically adsorbed on the trap. Then the trap is heated and the six-port valve actuated (Figure 5b) so the trap is backflushed by the GC carrier gas to transfer the analytes to the GC column. After the desorption is

complete, the injection is made by increasing abruptly the temperature of the cryogenic unit or the GC oven. The sorbent trap is also heated at a higher temperature between runs to clean it from any remaining compounds and avoid carryover.

The parameters that could be optimized to obtain the best performance of a PT set-up are the temperature of the sparging unit, the sample volume and the purge-gas volume (time × flow rate). For the former parameter, heating the sample would result in more water being transferred with the purged analytes and give less reproducible results.[74] It might also lead to difficulties in the subsequent GC analysis. This is why most studies are performed at room temperature. Using a large sample volume could improve the response but is often limited by the volume of the sparger and would require a long

Instrumental conditions

Purge-and-trap unit

Program	Time (min)	Tenax® trap (°C)	Cryo unit (°C)
Purge gas	Helium at 40 mL min^{-1}		
Purge	8	<30	
Dry purge	4	<30	
Capillary cooldown		<30	–120
Desorb preheat		175	–120
Desorb	4	180	–120
Inject	0.75	180	200
Trap bake	8	225	

Gas chromatograph

Carrier	Helium at 2.7 mL min^{-1}
Column	HP5 (5% diphenyl, 95% dimethylsilicone) 30 m × 0.53 mm x 0.88 μm
Detector	ECD
Oven temperature	30 °C for 5 min 30 °C to 80 °C at 10 °C min^{-1} 80 °C to 180 °C at 20 °C min^{-1} 180 °C for 2 min

Instrumental conditions

Purge-and-trap unit

Program	Time (min)	Tenax® trap (°C)	Cryo unit (°C)
Purge gas	Helium at 40 mL min^{-1}		
Purge	8	<30	
Dry purge	4	<30	
Capillary cooldown		<30	–120
Desorb preheat		175	–120
Desorb	4	180	–120
Inject	0.75	180	200
Trap bake	8	225	

Gas chromatograph

Carrier	Helium at 2.7 mL min^{-1}
Column	HP5 (5% diphenyl, 95% dimethylsilicone) 30 m × 0.53 mm x 0.88 μm
Detector	ECD
Oven temperature	30 °C for 5 min 30 °C to 80 °C at 10 °C min^{-1} 80 °C to 180 °C at 20 °C min^{-1} 180 °C for 2 min

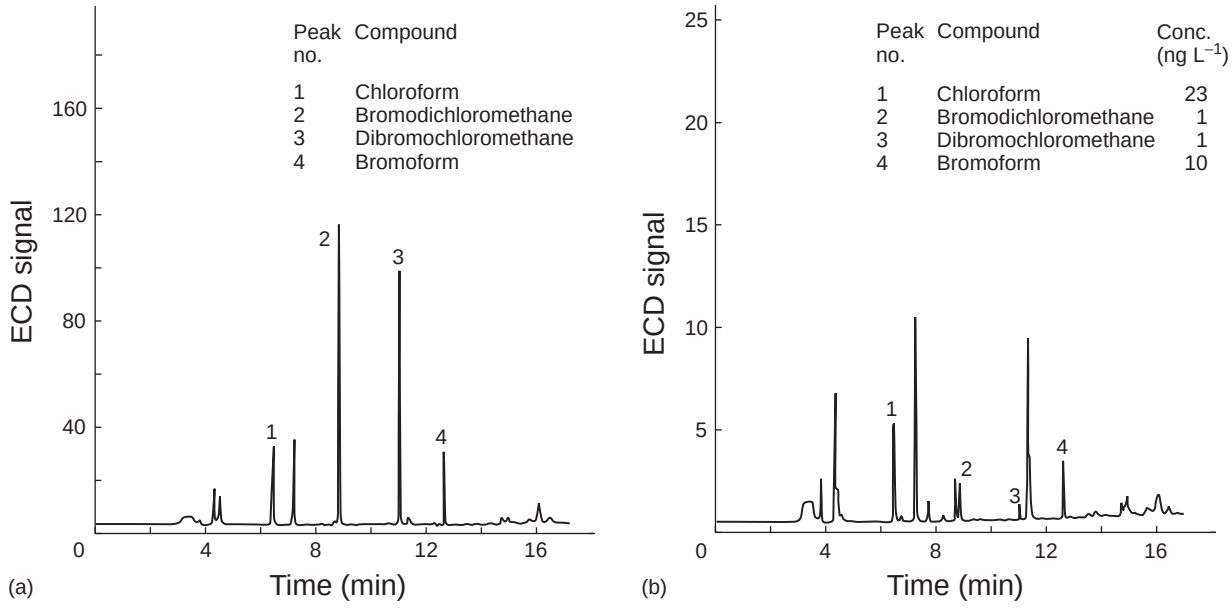

Figure 6 Typical chromatograms of THM analysis by PTGC/ECD. Conditions as in Table 5. (a) Standard solution of 100 ng L^{-1} each in water; (b) St. Lawrence river-water sample taken at Gentilly, Canada. [Reprinted with permission from L. Lépine, J.-F. Archambault, *Anal. Chem.*, **64**, 810–814 (1992).[38] Copyright 1992, American Chemical Society.]

Table 5 Reported performance of THM analysis in water by PTGC

Volume of sample (mL)	Purge time and flow	Adsorbent trap	GC conditions	DL[a] (ng L^{-1})	Precision (% RSD) [concentration level]	Ref. [year]
5	11 min at 20 mL min^{-1}	n.s.[b]	1.83 m × 2 mm ID Tenax®, ELCD, 8 min at 95 °C, 4 °C min^{-1} to 205 °C	100–200	3.5–5.3 [1–100 µg L^{-1}]	52 [1979]
40	30 min at 40 mL min^{-1}	[On-column trapping]	1.8 m × 2.2 mm ID Tenax® GC 60–80 mesh, Hall ELCD, split 1 : 2, 40 °C for trapping, 20 °C min^{-1} to 100 °C, 8 °C min^{-1} to 190 °C, hold 7 min	100–200	2.1–6.8 [16 µg L^{-1}]	71 [1982]
5	10 min at 30–40 mL min^{-1}	[On-column trapping]	2.4 m × 2 mm ID Tenax® GC 60–80 mesh, Hall ELCD, ambient for trapping, 2 min at 60 °C, 10 °C min^{-1} to 180 °C	100	1.8–4.1 [4–16 µg L^{-1}]	72 [1985]
5	8 min at 40 mL min^{-1}	Tenax®	30 m × 0.53 mm ID × 0.88 µm HP-5, ECD, cryofocusing −120 °C, 5 min at 30 °C, 10 °C min^{-1} to 80 °C, 20 °C min^{-1} to 180 °C, hold 2 min	0.01–0.05	2.4–4.5 [1 ng L^{-1}]	38 [1992]
Continuous flow 25 mL min^{-1} for 1 min	1 min at 100 mL min^{-1}	Tenax®	1.83 m × 3 mm ID SP-1000, ECD, isothermal 105 °C	10–100	2.4–9.4 [0.05–5 µg L^{-1}]	18 [1995]
30	30 min at 30 mL min^{-1}	2 cm Tenax® in 30 cm × 0.53 mm ID column at −100 °C	50 m × 0.32 mm ID × 5 µm CP Sil 8, ECD, 5 min at 40 °C, 12 °C min^{-1} to 95 °C, hold 10 min, 15 °C min^{-1} to 160 °C, hold 10 min	0.02–2	2–6 [0–100 ng L^{-1}]	23 [1996]
10	20 min at 30 mL min^{-1}	Tenax® + Carbosieve III S	30 m × 0.32 mm ID × 5 µm DB-1, micro-trap focusing, ECD, 2 min at 50 °C, 10 °C min^{-1} to 200 °C, hold 5 min	1	2.9–4.1 [50 ng L^{-1}]	39 [1996]

[a] Based on S/N = 3.
[b] Not specified.

purge time and large purge-gas volume, which could lead to breakthrough of the more volatile compounds from the sorption trap before the purge is completed. Typically, a 5–25-mL sample is used on most commercial units. Hence the only practical parameters left for optimization are the purge time and flow rate.

Lépine and Archambault[38] optimized the purge time for THM analysis and found that with a flow of $40\,mL\,min^{-1}$, $CHCl_3$ is purged completely in only a few minutes whereas for $CHBr_3$, the least volatile of the group, the purge efficiency is greatly improved with a longer purging time but tends to reach a plateau after 8 min. The purge efficiencies obtained at 8 min for $CHCl_3$, $CHBrCl_2$, $CHBr_2Cl$ and $CHBr_3$ are 100, 99.7, 95.2 and 78.6%, respectively. These values are satisfactory and increasing the purge time only marginally improves the efficiency for $CHBr_3$ and might also cause the breakthrough of $CHCl_3$ (the most volatile of the THMs) from the sorption trap. For the other parameters, they basically used standard values recommended by the manufacturer. A typical chromatogram obtained under these conditions is presented in Figure 6(a) and (b) for a $100\,ng\,L^{-1}$ standard, along with an analysis of a river-water sample showing concentrations between 1 and $23\,ng\,L^{-1}$ for the four THMs. Finally, for working at very low concentrations, a special working procedure should be followed. The Tenax® trap must be conditioned by a 15-min bake every morning, followed by a blank run to verify the cleanliness of the system. In addition, the sparger is purged for at least 2 min before the sample is introduced, to prevent atmospheric contamination.

The external standard calibration technique is generally used by analyzing aqueous composite standards. These standard solutions are treated exactly like the samples and transferred rapidly from their storage vial to the sparger with a syringe, thus minimizing the loss of volatile analytes through volatilization. When analyzing complex aqueous matrices containing a high salt content, the purge obtained with these standards should be compared with that for a real sample by analyzing fortified samples. A surrogate analyte, not present in the matrix, could also be added in the sample to monitor any change in purge efficiency with each sample matrix.

6.3 Analytical Performance

Table 5 gives the analytical performance reported by different groups for THM analysis by the PT technique. The DLs obtained in the 1980s with packed columns were at the $0.1\,\mu g\,L^{-1}$ level, which is comparable to those with DAI, LLE or HS. However, the more recent use of ECDs and capillary columns with some sort of analyte refocusing at the head of the column has improved tremendously the sensitivity of the PT technique, with DLs one to two orders of magnitude below the $1\,ng\,L^{-1}$ level.[23,38] The precision is very good and is generally <6% even at the low nanograms per liter level. Finally, the linearity of the technique has been shown by Lépine and Archambault[38] to extend over four orders of magnitude from 1 to $10^4\,ng\,L^{-1}$, taking full advantage of the large dynamic range of the ECD.

6.4 Advantages and Limitations

The major advantage of the PT technique lies in its superior analytical performance. The sensitivity is unmatched by other techniques and comes from the fact that a relatively large sample can be analyzed and that almost all of the analyte present in the sample is injected into the GC column. The precision is also very good even at concentrations down to the nanograms per liter level.

Like the HS technique, the major disadvantage of PT is the requirement for another piece of equipment to be added to the conventional gas chromatograph. Also, in order to achieve the very low DLs already mentioned, the gas chromatograph should have cryogenic possibilities, either by lowering the whole oven to subambient temperatures or by an add-on unit which could be purchased with the PT unit. Finally, the technique cannot be easily automated, and unattended analysis of a large number of samples can only be achieved through the addition of an external autosampler to the standard PT unit.

One limitation of the technique is that some environmental samples, where surfactants might be present, tend to foam when purged with the inert gas and the frit sparger cannot be used. For such samples, some manufacturers suggest the use of a different sparger made of a glass tube in which the gas is introduced through a stainless-steel needle, giving a stream of large bubbles. However, the purge efficiency is not as good as with the frit sparger.

7 COMPARISON OF THE ANALYTICAL TECHNIQUES

In this article, we have described the four major analytical techniques routinely used for THM analysis in water and a comparison of their major characteristics is given in Table 6. They all have some advantages and drawbacks, which will guide chemical analysts in selecting the technique that best suits their needs, depending on the field of activity and the instrumentation available.

First, we should say that it is one case where the technology largely precedes the regulatory requirements. With high-efficiency capillary columns now commonly available and the high sensitivity of the ECD, all these techniques have DLs at or below $0.1\,\mu g\,L^{-1}$ for each of the

Table 6 Comparison of the discussed GC techniques for THM analysis

Parameter	DAI	LLE	HS	PT
Sensitivity	Good	Good	Excellent	The most sensitive
Precision	Excellent	Excellent	Excellent	Excellent
Matrix analyzed	Only clean aqueous matrix	Moderately clean aqueous matrix	Any aqueous matrix	Any aqueous matrix except with foaming compounds
Specific equipment required	Cold on-column GC injector column	None	HS sampler, optional cryo-trap	PT sampler (cryo-trap recommended)
Simplicity of manipulation	Excellent	Moderate	Excellent	Good
Automation	With GC autoinjector	No	Incorporated within the HS sampler	With a specialized autosampler
Analysis time	Fast, only GC run time	Long, extraction time added to tedious glassware cleanup and preparation time	Moderate–fast, depending on equilibration time vs GC run time	Moderate, purge and desorb time added to the GC run time

four THMs, which is more than adequate for monitoring these analytes in drinking water or beverages. Even in more restrictive applications such as the determination of the background level in natural water samples (lakewater or seawater) or in the nuclear power or semiconductor industry, where high-purity water is needed, this level of detection is sufficient. However, if someone has a special application which requires the very lowest detection achievable, the PT technique is the one to choose. With a purge efficiency close to 100% for these volatile compounds, all the molecules present in a relatively large water sample (typically 5–25 mL) are injected into the GC column, whereas the other approaches have some limitations in that respect. In HS, the amount of analyte introduced into the GC column is limited by the partition coefficient of each analyte between the water and gas phases and the size of the HS aliquot that is injected. However, with cryogenic refocusing at the column head, large loop injection or a long transfer time (depending on the system used) can be used to obtain a DL close to that of the PT technique. In DAI and LLE, the size of the liquid aliquot that the GC column can tolerate (only a few microliters) is the limiting factor.

The complexity of the matrix to be analyzed can also dictate the choice of an analytical technique. With the gas-partition techniques such as HS and PT, since only the volatile compounds are isolated and injected into the GC column, any aqueous matrices can be analyzed. The only exception is the presence of synthetic or natural surfactants, which tend to foam in the PT sparger. In contrast, DAI is limited to relatively clean matrices since everything in the sample is injected onto the column. In LLE, the THMs are isolated from inorganic salts but other organic compounds can be extracted along with the THMs and cause some interference or deteriorate the column performance.

Other more practical aspects, such as the instrumentation required, the simplicity of manipulation, the possible automation and the analysis time, can be considered. If the samples are relatively clean and the sensitivity is not a problem, then DAI is the simplest method. It requires no manipulation of the sample, gives fast results and a large number of samples can be analyzed with an autoinjector on the gas chromatograph. The HS technique also comes close to DAI in terms of simplicity of operation. If the water is introduced into the HS vial at the sampling site, then the only action necessary is to put the vials in the system and start the run. Of course, it requires the purchase of an additional piece of equipment, but the investment can be worthwhile for routine analyses of large numbers of samples, especially when matrix or sensitivity problems may be an issue. The PT method is also easy to perform and needs no manipulation other than taking an aliquot from the sampling vial with a syringe and injecting it into the sparger. However, this operation is performed manually and can only be automated by the addition of a separate autosampler. Finally, LLE is by far the most time-consuming procedure and is not well suited for routine analysis of large numbers of samples. It is, however, a good reference method and may be chosen when the limited number of samples does not justify the purchase of an HS or PT unit, and when DAI is not possible because the gas chromatograph is not equipped with a cool on-column injector or because of matrix problems.

ACKNOWLEDGMENTS

The authors thank Guy Bélanger of Hydro-Québec for supporting financially the production of this article.

ABBREVIATIONS AND ACRONYMS

DAI	Direct Aqueous Injection
DL	Detection Limit
ECD	Electron Capture Detector
ELCD	Electrolyte Conductivity Detector
FID	Flame Ionization Detection
GC	Gas Chromatography
HS	Headspace
HSGC	Headspace Gas Chromatography
ID	Internal Diameter
LLE	Liquid–Liquid Extraction
MCL	Maximum Contaminant Level
MS	Mass Spectrometry
MTBE	Methyl *tert*-Butyl Ether
NIST	National Institute of Standards and Technology
OD	Outer Diameter
PT	Purge-and-trap
RSD	Relative Standard Deviation
S/N	Signal-to-noise Ratio
SPI	Septum-equipped Programmable Injector
SPME	Solid-phase Microextraction
THM	Trihalomethane
TLHS	Thin-layer Headspace
TOC	Total Organic Content
TTHM	Total Trihalomethanes
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
ZDV	Zero-dead-volume

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Selective Detectors

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

REFERENCES

1. J.J. Rook, 'Formation of Haloforms During Chlorination of Natural Waters', *Water Treat. Exam.*, **23**, 234–243 (1974).
2. T.A. Bellar, J.J. Lichtenberg, R.C. Kroner, 'The Occurrence of Organohalides in Chlorinated Drinking Waters', *J. Am. Water Works Assoc.*, **66**, 703–706 (1974).
3. B. Koch, S.W. Krasner, 'Occurrence of Disinfection By-products in a Distribution System', in *Proceedings of Water Quality and Technology Conference, Los Angeles, June 18–22, 1989*, American Water Works Association, Denver, CO, 1203–1230, 1989.
4. W.W. Bunn, B.B. Haas, E.R. Deane, R.C. Kleopfer, 'Formation of Trihalomethanes by Chlorination of Surface Water', *Environ. Lett.*, **10**, 205–213 (1975).
5. T. Ishikawa, T. Sato, Y. Ose, H. Nagase, 'Reaction of Chlorine and Bromine With Humic Substance', *Sci. Total Environ.*, **54**, 185–194 (1986).
6. S.W. Krasner, M.J. McGuire, J.G. Jacangelo, N.L. Patania, K.M. Reagan, E.M. Aieta, 'The Occurrence of Disinfection By-products in US Drinking Water', *J. Am. Water Works Assoc.*, **81**(8), 41–53 (1989).
7. J.M. Symons, T.A. Bellar, J.K. Carswell, J. DeMarco, K.L. Kropp, G.G. Robeck, D.R. Seeger, C.J. Slocum, B.L. Smith, A.A. Stevens, 'National Organics Reconnaissance Survey for Halogenated Organics', *J. Am. Water Works Assoc.*, **67**, 634–647 (1975).
8. R. Otson, D.T. Williams, D. Bothwell, Volatile Organic Compounds in Water at Thirty Canadian Potable Water Facilities', *J. Assoc. Off. Anal. Chem.*, **65**(6), 1370–1374 (1982).
9. W.H. Glaze, J.E. Henderson, G. Smith, 'Analysis of New Chlorinated Organic Compounds Formed by Chlorination of Municipal Wastewater', in *Water Chlorination: Environmental Impact and Health Effects*, ed. R.L. Jolley, Ann Arbor Science, Ann Arbor, MI, 139–159, Vol. 1, 1978.
10. R.F. Thomas, M.J. Weisner, H.J. Brass, 'The Fifth Trihalomethane: Dichloroiodomethane, Its Stability and Occurrences in Chlorinated Drinking Water', in *Water Chlorination: Environmental Impact and Health Effects*, ed. R.L. Jolley, Ann Arbor Science, Ann Arbor, MI, 161–168, Vol. 3, 1980.
11. M. Biziuk, A. Przyjazny, 'Methods of Isolation and Determination of Volatile Organohalogen Compounds in Natural and Treated Waters', *J. Chromatogr. A*, **733**, 417–448 (1996).
12. D.T. Williams, R. Otson, D. Bothwell, K.L. Murphy, J.L. Robertson, 'Trihalomethane Levels in Canadian Drinking Water', *Environ. Sci. Res.*, **16**, 503–512 (1978).
13. R. Otson, D.T. Williams, D. Bothwell, 'Comparison of Dynamic Headspace and Solvent Extraction Techniques for the Determination of Trihalomethanes in Water', *Environ. Sci. Technol.*, **13**(8), 936–939 (1979).

14. R. Otson, D.T. Williams, D.C. Briggs, 'Relationships Between Raw Water Quality, Treatment and Occurrence of Organics in Canadian Potable Water', *Bull. Environ. Contam. Toxicol.*, **28**, 396–403 (1982).
15. D.T. Williams, G.L. Lebel, F.M. Benoit, 'Disinfection By-products in Canadian Drinking Water', *Chemosphere*, **34**, 299–316 (1997).
16. E.E. Koukouraki, R.A. Berkhoudt, E. Diamadopoulos, 'Determination of Volatile Chlorination By-products in Drinking Water in Greece', *Fresenius' Environ. Bull.*, **6**, 160–165 (1997).
17. B.G. Oliver, 'Analysis of Volatile Halogenated and Purgeable Organics', in *Analysis of Traces Organics in the Environment*, eds. B.K. Afghan, A.S.Y. Chau, CRC Press, Boca Raton, FL, 1–29, 1989.
18. H.W. Wang, M.S. Simmons, R.A. Deininger, 'Application of Dynamic Headspace Analysis for Trihalomethanes in Flowing Water', *J. Chromatogr. Sci.*, **33**, 109–115 (1995).
19. US Environmental Protection Agency, 'Regulation for Control of Trihalomethanes', *Fed. Regist.*, **44**(231), 68624–68642 (1979).
20. I.F.M.M. Temmerman, D.J.M. Quaghebeur, 'Analysis of Trihalomethanes by Direct Aqueous Injection (THM/DAI)', *J. High Resolut. Chromatogr.*, **13**, 379–381 (1990).
21. F.W. Pontius, 'An Update of the Federal Drinking Water Regulations', *J. Am. Water Works Assoc.*, **85**(1), 51–62 (1995).
22. Z. Polkowska, J. Namiesnik, J. Czerwinski, B. Zygmunt, 'Volatile Organo-halogen Trace Analysis in Beverages by Thin-layer Headspace Enrichment and Electron-capture Gas Chromatography', *Int. J. Food Sci. Technol.*, **31**, 387–395 (1996).
23. O. Connan, P. Lecorre, P. Morin, 'Purge-and-trap GC Determination of Volatile Halocarbon Compounds in Seawater', *Analusis*, **24**, 55–59 (1996).
24. M. Biziuk, J. Namiesnik, J. Czerwinski, D. Gorlo, B. Makuch, W. Janicki, Z. Polkowska, 'Occurrence and Determination of Organic Pollutants in Tap and Surface Waters of the Gdansk District', *J. Chromatogr. A*, **733**(1–2), 171–183 (1996).
25. D.W. Armstrong, T. Golden, 'Determination of Distribution and Concentration of Trihalomethanes in Aquatic Recreational and Therapeutic Facilities by Electron-capture GC', *LC/GC*, **4**(7), 652–655 (1986).
26. M. Biziuk, J. Czerwinski, E. Kozlowski, 'Identification and Determination of Organohalogen Compounds in Swimming Pool Water', *Int. J. Environ. Anal. Chem.*, **50**, 109–115 (1993).
27. G.J. Hollod, E.W. Wilde, 'Trihalomethanes in Chlorinated Cooling Waters of Nuclear Reactors', *Bull. Environ. Contam. Toxicol.*, **28**, 404–408 (1982).
28. L. Lépine, R. Gilbert, L. Ouellet, 'Thermal Stability of Chloroform in the Steam–Condensate Cycle of CANDU–PHW Nuclear Power Plant', *Can. J. Chem. Eng.*, **70**, 1172–1179 (1992).
29. A.A. Nicholson, O. Meresz, D. Lemyk, 'Determination of Free and Total Potential Haloforms in Drinking Water', *Anal. Chem.*, **49**(6), 814–819 (1977).
30. K. Grob, A. Habich, 'Trace Analysis of Halocarbons in Water; Direct Aqueous Injection with Electron Capture Detection', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **6**, 11–15 (1983).
31. K. Grob, 'Further Development of Direct Aqueous Injection with Electron-capture Detection in Gas Chromatography', *J. Chromatogr.*, **299**, 1–11 (1984).
32. J.E. Henderson, G.R. Peyton, W.H. Glaze, 'A Convenient Liquid–Liquid Extraction Method for the Determination of Halomethanes in Water at the Parts per Billion Level', in *Identification and Analysis of Organic Pollutants in Water*, ed. L.H. Keith, Ann Arbor Science, Ann Arbor, MI, 105–111, 1976.
33. J.P. Mieure, 'A Rapid and Sensitive Method for Determining Volatile Organohalides in Water', *J. Am. Water Works Assoc.*, **69**(1), 60–62 (1977).
34. J.J. Richard, G.A. Junk, 'Liquid Extraction for the Rapid Determination of Halomethanes in Water', *J. Am. Water Works Assoc.*, **69**(1), 62–64 (1977).
35. M.M. Varma, M.R. Siddique, K.T. Doty, A. Machis, 'Analysis of Trihalomethanes in Aqueous Solutions: A Comparative Study', *J. Am. Water Works Assoc.*, **71**(7), 389–392 (1979).
36. K.L.E. Kaiser, B.G. Oliver, 'Determination of Volatile Halogenated Hydrocarbons in Water by Gas Chromatography', *Anal. Chem.*, **48**, 2207–2209 (1976).
37. T.A. Bellar, J.J. Lichtenberg, 'Determining Volatile Organics at Micrograms per Liter by Gas Chromatography', *J. Am. Water Works Assoc.*, **66**, 739–744 (1974).
38. L. Lépine, J.-F. Archambault, 'Parts-per-trillion Determination of Trihalomethanes in Water by Purge-and-trap Gas Chromatography with Electron Capture Detection', *Anal. Chem.*, **64**, 810–814 (1992).
39. B. Zygmunt, 'Determination of Trihalomethanes in Aqueous Samples by Means of a Purge-and-trap System with On-sorbent Focusing Coupled to Gas Chromatography with Electron-capture Detection', *J. Chromatogr. A*, **725**, 157–163 (1996).
40. W.L. Budde, J.W. Munch, 'Methods for the Determination of Organic Compounds in Drinking Water – Supplement III', Report EPA/600/R-95/131, USEPA Office of Research and Development, Cincinnati, OH, August 1995.
41. H. Norin, L. Renberg, 'Determination of Trihalomethanes (THM) in Water Using High Efficiency Solvent Extraction', *Water Res.*, **14**, 1397–1402 (1980).
42. K. Kawakami, Y. Tanaka, Y. Zaitsu, H. Hoshikawa, 'Measurement of Trihalomethanes (THM) in Drinking-water with an Automatic Analyzer', *Water Sci. Technol.*, **28**(11–12), 465–471 (1993).
43. M.I. Badawy, 'Evaluation of Headspace Technique for the Determination of Trihalomethanes in Water', *Bull. Environ. Contam. Toxicol.*, **48**, 630–634 (1992).

44. F.K. Pfaender, R.B. Jonas, A.A. Stevens, L. Moore, J.R. Hass, 'Evaluation of Direct Aqueous Injection Method for Analysis of Chloroform in Drinking Water', *Environ. Sci. Technol.*, **12**(4), 438–441 (1978).
45. D. Carmichael, W. Holmes, 'Screening of Trihalomethanes by Direct Aqueous Injection Using Electron Capture Detection', *J. High Resolut. Chromatogr.*, **13**, 267–269 (1990),Also Reported in Varian Application Note 14, Varian, Walnut Creek Instrumentation Division, Walnut Creek, CA, 1989.
46. S.M. Pyle, D.F. Gurka, 'Volatile Organic Analysis by Direct Aqueous Injection', *Talanta*, **41**(11), 1845–1852 (1994).
47. E. Kozlowski, Z. Polkowska, 'Continuous-flow Thin-layer Headspace (Tlhs) Analysis. 9. Isolation of Trihalomethanes From Complex Mixtures with Autogenous Generation of a Liquid Sorbent Stream', *Chem. Anal. (Warsaw)*, **41**, 173–181 (1996).
48. Z. Polkowska, E. Kozlowski, "Continuous-flow Thin-layer Headspace (Tlhs) Analysis. 10. Preconcentration of Volatile Analytes in Autogenously Generated Stream of Liquid Sorbent', *Chem. Anal. (Warsaw)*, **41**, 183–191 (1996).
49. M. Turska, L. Wolska, B. Zygmunt, J. Namiesnik, 'Direct Injection of Aqueous Samples into Gas Chromatographic Columns', *Chem. Anal. (Warsaw)*, **42**, 787–797 (1997).
50. H. Hagenmaier, G. Werner, W. Jäger, 'Quantitative Determination of Volatile Halogen Hydrocarbons in Water Samples by Capillary Gas Chromatography and Electron Capture Detection', *Z. Wasser Abwasser Forsch.*, **15**(4), 195–198 (1982).
51. C. Garcia, P.G. Tiedra, A. Ruano, J.A. Gomez, R.J. Garciavillanova, 'Evaluation of the Liquid–Liquid Extraction Technique and Application to the Determination of Volatile Halo-organic Compounds in Chlorinated Water', *J. Chromatogr.*, **605**, 251–255 (1992).
52. R. Dressman, A.A. Stevens, J. Fair, B. Smith, 'Comparison of Methods for the Determination of Trihalomethanes in Drinking Water', *J. Am. Water Works Assoc.*, **71**(7), 392–396 (1979).
53. H.C. Hu, P.H. Weiner, 'Modifications to Methods for Volatile Organics Analysis at Trace Levels', *J. Chromatogr. Sci.*, **18**, 333–342 (1980).
54. G. Castello, T. Gerbino, S. Kanitz, 'Sensitivity and Linearity of Headspace and Liquid–Liquid Extraction Techniques for Gas Chromatographic Analysis of Halocarbons in Water', *J. Chromatogr.*, **351**, 165–175 (1986).
55. M.M. Varma, M.R. Siddique, K.T. Doty, A. Machis, 'Analysis of Trihalomethanes in Aqueous Solutions: A Comparative Study', *J. Am. Water Works Assoc.*, **71**(7), 389–392 (1979).
56. R. Otson, D.T. Williams, 'Evaluation of a Liquid–Liquid Extraction Technique for Water Pollutants', *J. Chromatogr.*, **212**, 187–197 (1981).
57. B. Kolb, L.S. Ettre, *Static Headspace/Gas Chromatography: Theory and Practice*, Wiley-VCH, New York, 1997.
58. G.J. Piet, P. Slingerland, F.E. de Grunt, M.P.M. van der Heuvel, B.C.J. Zoeteman, 'Determination of Very Volatile Halogenated Organic Compounds in Water by Means of Direct Headspace Analysis', *Anal. Lett.*, **A11**(5), 437–448 (1978).
59. B. Kolb, P. Pospisil, T. Borath, M. Auer, 'Head Space Gas Chromatography with Glass Capillaries Using an Automated Electropneumatic Dosing System', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**(6), 283–287 (1979).
60. R. Otson, 'Automatic Liquid Injector for Headspace Gas Chromatography', *Anal. Chem.*, **53**, 929–931 (1981).
61. S. Onodera, C. Siriwong, M. Tabucanon, 'Comparison of Methods for Determination of Trihalomethanes in Drinking Water', *J. Sci. Soc. Thailand*, **10**, 221–237 (1984).
62. M.E. Comba, K.L.E. Kaiser, 'Determination of Volatile Contaminants at the ng L^{-1} Level in Water by Capillary Gas Chromatography with Electron Capture Detection', *Int. J. Environ. Anal. Chem.*, **16**, 17–31 (1983).
63. P.L. Wylie, 'Comparing Headspace with Purge-and-trap for Analysis of Volatile Priority Pollutants', *J. Am. Water Works Assoc.*, **80**(8), 65–72 (1988).
64. E.A. Dietz, Jr, K.F. Singley, 'Determination of Chlorinated Hydrocarbons in Water by Headspace Gas Chromatography', *Anal. Chem.*, **51**(11), 1809–1814 (1979).
65. Z. Zhang, J. Pawliszyn, 'Headspace Solid-phase Microextraction', *Anal. Chem.*, **65**, 1843–1852 (1993).
66. Z. Zhang, J. Pawliszyn, 'Sampling Volatile Organic Compounds Using a Modified Solid-phase Microextraction Device', *J. High Resolut. Chromatogr.*, **19**, 155–160 (1996).
67. P.A. Frazey, R.M. Barkley, R.E. Sievers, 'Solid-phase Microextraction with Temperature-programmed Desorption for the Analysis of Iodonation Disinfection Byproducts', *Anal. Chem.*, **70**(3), 638–644 (1998).
68. B. Kolb, G. Zwick, M. Auer, 'A Water Trap for Static Cryo-headspace Gas Chromatography', *J. High Resolut. Chromatogr.*, **19**(1), 37–42 (1996).
69. B. Kolb, M. Aeur, P. Pospisil, 'Methods for the Quantitative Analysis of Volatile Halocarbons from Aqueous Samples by Equilibrium Headspace Gas Chromatography with Capillary Columns', *J. Chromatogr.*, **279**, 341–348 (1983).
70. T.A. Bellar, J.J. Lichtenberg, 'The Determination of Halogenated Chemicals in Water by Purge and Trap Method', Method 502.1, US Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH, 1981.
71. R. Otson, D.T. Williams, 'Headspace Chromatographic Determination of Water Pollutants', *Anal. Chem.*, **54**, 942–949 (1982).
72. M. Mehran, 'Purge-and-column-trap Technique for the Gas Chromatographic Analysis of Halogenated Organic Compounds', *J. Chromatogr. Sci.*, **24**, 546–548 (1985).

73. A.R. Trussell, J.G. Moncur, F.Y. Lieu, R.R. Clark, L.Y.C. Leong, 'New Developments in Dynamic Headspace Analysis of Halogenated Organics', in *Water Chlorination: Environmental Impact and Health Effects*, ed. R.L. Jolley et al., Ann Arbor Science, Ann Arbor, MI, 583–592, Vol. 4, 1983.

74. M.R. Driss, M.L. Bouguerra, 'Analysis of Volatile Organic Compounds in Water by Purge-and-trap and Gas Chromatography Techniques – Operational Parameters Optimization of the Purge Step', *Int. J. Environ. Anal. Chem.*, **45**, 193–204 (1991).

Underground Fuel Spills, Source Identification

Barry K. Lavine, Anthony Moores, and Jason Ritter
Clarkson University, Potsdam, USA

1 Introduction **1**
2 Pattern Recognition **2**
2.1 Definition 2
2.2 Class Membership Problem in Fuel Spill Identification 2
2.3 Principal Component Analysis 3
2.4 Information Content of Principal Components 5
2.5 Implementation of Principal Component Analysis 5
2.6 Soft Independent Modeling of Class Analogy 5
2.7 Pattern Recognition Analysis of Gas Chromatography Profile Data 6
2.8 Genetic Algorithms for Feature Selection and Pattern Recognition 7
2.9 Pattern Recognition Approach to Fuel Spill Identification 10
2.10 Case Studies 10
3 Conclusions **18**
Abbreviations and Acronyms **18**
Related articles **19**
References **19**

Water samples from underground wells or aquifers contaminated by middle distillate fuels exist in one of two forms. Either the water sample collected from the well shows a layer of floating fuel, or the sample contains the dissolved hydrocarbons from the leaking fuel. In the worst case scenario, that of a floating fuel, the fuel layer is collected and analyzed by gas chromatography (GC). The GC profile of the fuel layer is then compared with the GC profile of the different candidate fuels to determine the fuel type of the leaking material. Usually, the gas chromatograms are compared visually, in order to obtain the best match. However, this approach to data analysis is subjective and is not always persuasive in a court of law. Visual analysis of gas chromatograms also has the drawback that it cannot take into account the influence of weathering on the overall GC profile of the fuel.

Pattern recognition (PR) methods offer a better approach to the problem of matching the gas chromatograms of jet fuels because these methods involve less subjectivity in the interpretation of the data. PR methods can identify fingerprint patterns in GC data characteristic of fuel type, even if the fuel samples have been subjected to a variety of conditions. Hence, discriminants can be developed that are relatively insensitive to the changes in the overall GC profile data of the original fuel due to contamination, analytical error, or weathering.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Groundwater is the only source of potable water for more than half of the individual households and communities in south-eastern USA.[1] Its possible contamination by fuels stored in leaking underground tanks or pipelines at or near military airfields has prompted the United States Air Force to fund the development of new methods for the identification of fuel materials recovered from subsurface environments. Burgeoning interest in techniques that can establish the type of fuel responsible for the contamination of an underground well or aquifer is motivated in part by the clean-up costs, legal fees, and fines incurred by the polluter.

Water from underground wells or aquifers contaminated by a leaking fuel exists in one of two forms: either the water sample collected from the well has a layer of floating fuel, or the sample contains dissolved hydrocarbons from the leaking fuel. In the worse case scenario, that of a floating fuel, the fuel layer is usually collected and analyzed by GC. The GC profile of the fuel layer is compared to the GC profile of different candidate fuels to determine the type of fuel responsible for the contamination of the well.

Usually, the gas chromatogram of the fuel spill and a number of suspected sources are compared visually, in order to obtain a best match. However, this approach to fuel spill identification is subjective and suffers from the drawback that it does not take into account the influence of weathering on the overall GC profile. Weathering[2,3] can cause oxidation of aliphatic hydrocarbons, evaporation of lower-molecular-weight alkanes and alkenes, or dissolution of aromatic and other water-soluble compounds in the fuel. (The net result is that differences exist between the gas chromatograms of weathered and unweathered fuels.) Changes in GC operating conditions caused by small variations in the temperature-programming rate of the capillary column can also be a serious problem, complicating the identification of a fuel.[4,5] Finally, changes in the GC profile due to variations in the resolution of the column over time

can also have a deleterious effect. Thus, evidence based on visual analysis of gas chromatograms is not always persuasive in a court of law, especially in cases involving an unweathered fuel identified as the source of a fuel spill, because of differences between gas chromatograms of weathered and unweathered fuels.

The nature of the complex mixture that constitutes a jet fuel suggests that a systematic comparison of gas chromatograms be made, in order to ensure that differences between GC profiles of various fuel types are significant. This is why PR methods[6–9] are used to match the gas chromatograms of fuel spills to suspected sources instead of direct data-based methods such as visual analysis or library search techniques.[10] PR methods involve less subjectivity in the interpretation of GC data than visual methods. The application of multivariate data analysis methods (of which PR is one example) to the fuel spill identification problem is also advantageous for other reasons. First, PR methods can identify features (i.e. GC peaks) in the chromatograms that are characteristic of fuel type even though the fuel samples in the data set have been subjected to a variety of conditions. Hence, discriminants can be developed from the data that are relatively insensitive to changes in the overall GC profile of the original fuel due to weathering, contamination, or analytical error. Second, the discriminatory information sought often consists of subtle variations in relative peak intensities distributed across several peaks in the gas chromatograms. PR methods are especially well suited for extracting this type of information from the large amount of qualitative and quantitative data present in the gas chromatograms of processed fuels. Third, PR methods rely heavily on graphics for the presentation of results. For example, the visual presentation of gas chromatograms as points in a principal component map allow the scientist or engineer to uncover hidden relationships within the data, for example to find distinct sample subgroups in a data set. Clearly, the scientist or engineer can play a more interactive role in the data analysis when using PR methods to analyze the data, which will result in a definite advantage, if one is interested in a careful analysis of the data.

2 PATTERN RECOGNITION

2.1 Definition

PR methods were originally developed to solve the class membership problem. In a typical PR study, samples are classified according to a specific property using measurements indirectly related to the property of interest. An empirical relationship or classification rule is developed from a set of samples for which the property of interest and the measurements are known. The classification rule is then used to predict this property in samples not part of the original training set. The set of samples for which the property of interest and measurements are known is called the training set. The set of measurements that describe each sample in the data set is called a pattern. The determination of the property of interest by assigning a sample to its respective category is called recognition – hence the term ''pattern recognition'' – because recognition is accomplished using the set of measurements that characterize each sample in the data set.

For PR analysis, each fuel sample is represented as a data vector $x = (x_1, x_2, x_3, \ldots, x_j, \ldots, x_n)$, where component x_j is a measurement, such as the area of the jth GC peak. In other words, each fuel sample can be considered as a point in an n-dimensional measurement space. The dimensionality of the measurement space corresponds to the number of measurements that are available for each sample. A basic assumption is that the distance between pairs of points in this measurement space is inversely related to the degree of similarity between the corresponding samples. Therefore, points representing samples from one class will cluster in a limited region of the measurement space distant from the points corresponding to the other class. PR is a set of methods for investigating data represented in this manner, in order to assess its overall structure, which is defined as the overall relationship of each sample to every other sample in the data set.

2.2 Class Membership Problem in Fuel Spill Identification

The class membership problem in fuel spill identification can be formulated as follows. Given a collection of potential source fuels which are characterized by a set of measurements made on each source sample, is it possible to classify the various fuel samples according to fuel type by using a set of measurements believed to be related to the property of interest? Can the classification rule, developed from a set of samples for which the fuel type and measurements are known, be used to predict the fuel type of spill samples that were not part of the original training set? Can we determine the components of the spill sample most related to fuel type, bearing in mind that the composition of a spill sample may differ significantly from the source due to contamination and or weathering? For PR analysis, each fuel type (i.e. class) is represented by a large number of fuel samples; the measurement variables, which are areas of selected GC peaks, are obtained from the gas chromatograms of the fuel samples.

The PR approach to fuel spill identification assumes that the fuel type is uniquely defined by the fuel's chemical composition. For example, jet fuels are manufactured only to meet certain specifications (flash point, vapor pressure, freezing point, etc.). Their composition is only loosely governed by military and civilian fuel specifications. Nevertheless, the chemical composition of a fuel dictates its physical properties. (The composition of a complex mixture uniquely defines its properties.) Hence, it is logical to assume that all fuels can be classified by type on the basis of their GC profiles because a gas chromatogram of a fuel contains information about its chemical composition. In the context of fuel spill identification, PR can clearly be defined as an analytical approach that utilizes mathematical, statistical, and other methods of formal logic to determine by indirect means the properties of fuels that otherwise would be very difficult to measure. Indirect observation of these properties (e.g. fuel type) is feasible provided that information about the chemical composition of the fuel is available because the properties of substances such as fuels are uniquely defined by their chemical composition.

The data analysis approach used to identify fuel type – indirect observation of the properties of a substance – actually employs the data to suggest the classification model. This rather unusual approach is possible because GC profile data usually contain a large number of interrelated measurement variables. To understand why GC profile data sets are redundant, consider JP-4, which is a mixture of alkanes, alkenes, and aromatic hydrocarbons. The gas chromatograms of JP-4 fuels (Figure 1) are characterized by a large number of early-eluting peaks that are large in area. There are only a few late-eluting peaks, but their size is small. Clearly, there is a strong negative correlation between the size of the early- and late-eluting peaks of JP-4. Furthermore, many of the alkane and alkene peaks are correlated, which is not surprising as alkenes are not constituents of crude oil but instead are formed from alkanes during the refining process. Finally, the specific property of a fuel most likely to be reflected in a high-resolution gas chromatogram is its distillation curve, which does not require all 85 peaks for characterization. Hence, the 85-peak gas chromatograms of JP-4 fuels do not necessarily require 85 independent axes to define the position of the data points in the pattern space. The data can be described in terms of a few latent variables that are linear combinations of the original measurement variables. Although these latent variables cannot be measured, they can be generated directly from the data using a technique called principal component analysis (PCA).

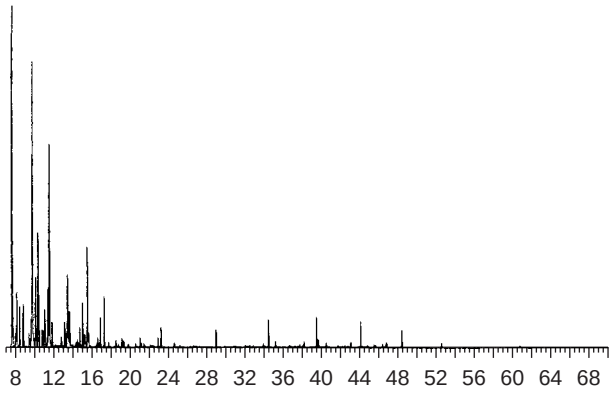

Figure 1 A gas chromatogram of a JP-4 fuel sample.

2.3 Principal Component Analysis

PCA[11,12] is the most widely used multivariate analysis technique in science and engineering. It is a method for transforming the original measurement variables into new, uncorrelated variables called principal components. Using this procedure is analogous to finding a set of orthogonal axes that represent the directions of greatest variance in the data. (The variance is defined as the degree to which the data points are spread apart or scattered in the n-dimensional measurement space.) Often, only two or three principal components are necessary to explain all of the information present in (p-dimensional, $p \gg 3$) compositional data. Hence, PCA can be applied to high-dimensional data to affect dimensionality reduction, in order to identify and display structure, classify samples, or identify outliers.

Dimensionality reduction or data compression is possible using PCA because of correlations between measurement variables. Figure 2 shows a plot of 15 hypothetical JP-4 fuel samples in a two-dimensional measurement space. The coordinate axes of this measurement space are the flash point and freezing point of the JP-4 fuel samples. There appears to be a relationship between the two measurement variables, flash point and freezing point, of the JP-4 fuel samples. This relationship suggests that the flash point and the freezing point are correlated because fixing the value of the flash point limits the range of values possible for the freezing point. If these two measurement variables were uncorrelated, the enclosed rectangle in Figure 2 would be fully populated by JP-4 fuel samples. Because information is defined as the scatter of points in a measurement space, it is evident that correlations between measurement variables decrease the information content of this space. In other words, the data points are restricted to a small region of the measurement space and may even reside in a subspace in the case of strongly correlated measurement variables (Figure 3).

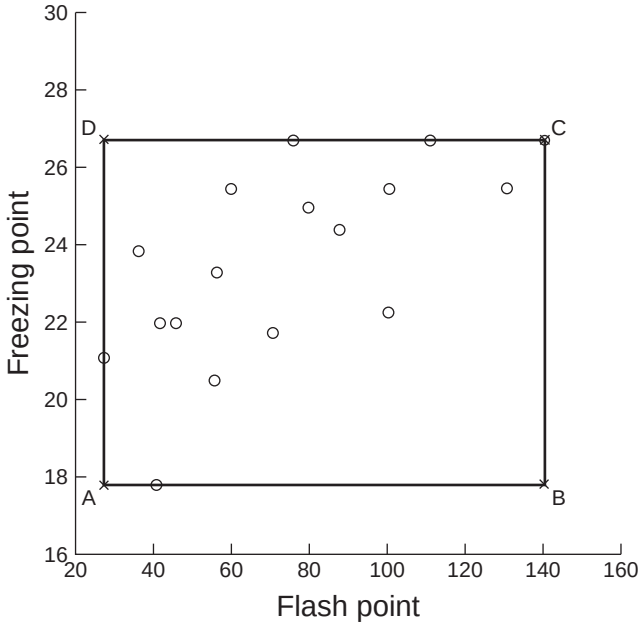

Figure 2 Fifteen hypothetical JP-4 fuel samples projected onto a two-dimensional measurement space. Because the freezing point and flash point are correlated, the data points are restricted to a small region of the measurement space. (Adapted from John Mandel, 'The Regression Analysis of Collinear Data', *J. Research of the NBS*, **190**(6), 465–476 (1985).)

Variables that are highly correlated, i.e. have a great deal of redundancy, are said to be colinear.

Colinearity between measurement variables is a strong indication that a new set of basis vectors can be found that will be better at conveying the information content present in data than axes defined by the original measurements. This new basis set is linked to variation in the data. In other words, a new coordinate system based on variance can be developed to display the data – principal components define the variance-based

$$X = \begin{pmatrix} 1.0 & 1.4 & 2.4 \\ 1.2 & 1.1 & 2.3 \\ 0.6 & 2.0 & 2.6 \\ 1.9 & 1.5 & 3.4 \\ 0.7 & 1.1 & 1.8 \\ 1.6 & 0.1 & 1.7 \\ 0.0 & 1.2 & 1.2 \end{pmatrix}$$

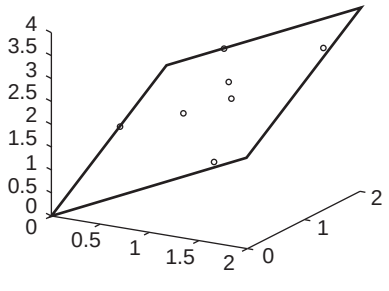

Figure 3 In the case of strongly correlated measurement variables, the data points may reside in a subspace of the original measurement space. (Adapted from Richard G. Brereton, *Multivariate Pattern Recognition in Chemometrics*, illustrated by case studies, p. 108, with kind permission from Elsevier Science, Amsterdam, 1992.)

axes of this new coordinate system. The largest principal component is formed by determining the direction of largest variation in the original measurement space and modeling it via a line fitted by linear least-squares which passes through the center of the data. The second largest principal component lies in the direction of next largest variation; it passes through the center of the data and is orthogonal to the largest principal component (Figure 4). The third largest principal component lies in the direction of next largest variation; it passes through the center of the data and is orthogonal to the first and second largest principal components, and so forth. Each principal component captures a different source of variation or information in the data. (Variation is the same as scatter, and scatter equals information.) The maximum number of principal components or variance-based axes that can be extracted from the data is the smaller of either the number of samples or number of measurements in the data set, as this number defines the largest number of independent variables in our data.

PCA takes advantage of the fact that a large amount of data is generated in fuel spill identification studies. The data have a great deal of redundancy and therefore a great deal of colinearity. Because the measurement variables are correlated, 85-peak gas chromatograms do not necessarily require 85 independent orthogonal axes to define the position of a sample point. Using PCA, the original measurement variables, which constitute a correlated axis system, can be converted into an orthogonal system, which removes correlation by forcing the new axes to be independent. This requirement dramatically reduces the dimensionality of the data because only a few independent axes are necessary to

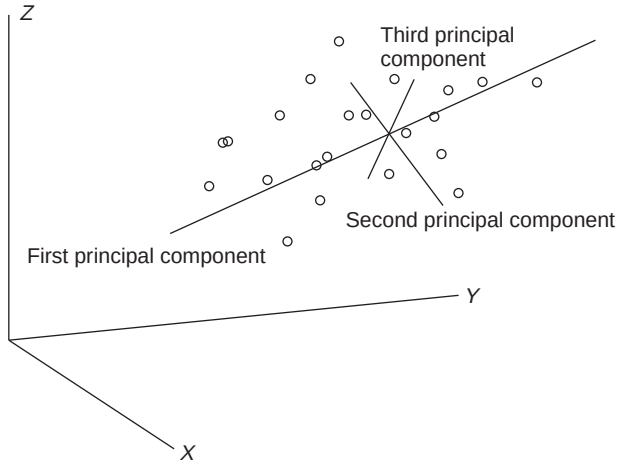

Figure 4 Principal component axes. The third principal component only describes noise in the data. (Reprinted from S. Brown, 'Chemical Systems Under Indirect Observation: Latent Properties and Chemometrics', *Applied Spectroscopy*, **49**(12), 14A–30A (1995), with kind permission from SAS.)

describe the data. The gas chromatograms of a set of fuel spill samples may lie in a subspace of the original measurement space, and a plot of the two or three largest principal components of the data can help us to visualize the relative position of the gas chromatograms in this subspace. Thus, it is possible to identify important relationships in the data – that is, find similarities and dissimilarities within a set of data, by examining principal component plots of the data.

2.4 Information Content of Principal Components

A measure of the amount of information that is conveyed by each principal component is its variance. For this reason, the principal components are usually arranged in order of decreasing variance. The most informative principal component is the first and the least informative is the last. Hence, one would expect that only the first few principal components should convey information about the effect of interest, if the data is collected with due care, because most of the information or variance in the data should be about the effect that we seek to study. However, the situation is not always so straightforward. Each principal component describes some amount of signal and some amount of noise in the data because of accidental correlation between signal and noise. The larger principal components primarily describe signal, whereas the smaller principal components essentially describe noise. When so-called smaller principal components are deleted, noise is being discarded from the data, but so is a small amount of signal. However, the enhancement obtained in signal-to-noise more than compensates for the biased representation of the signal that results from discarding principal components, which contains a small amount of signal but a large amount of noise. Hence, plotting the data in a coordinate system defined by the two or three largest principal components usually provides more than enough information about the overall structure of the data while simultaneously enhancing signal-to-noise.

2.5 Implementation of Principal Component Analysis

The procedure for implementing PCA is as follows. First, the samples or data vectors are arranged in the form of a table or matrix. Each row of the matrix is a sample, and each column of the matrix is a measurement variable. It is essential that each variable encodes the same information. For example, if measurement 5 is the area of a GC peak for benzene in sample 1, it must also be the area of a GC peak for benzene in sample 2, sample 3, . . . sample N.

The data matrix is usually centered about the mean, that is the mean of a column is subtracted from each element of that column. The effect of mean centering on a data matrix is to adjust the means of the columns to zero. To ensure that each variable has an equal weight (and therefore an equal effect) in the analysis, the data can be autoscaled. In other words, the data are first mean centered and then each element of a column is divided by the standard deviation of that column. Autoscaling adjusts the measurements so that each variable has a mean of zero and a standard deviation of one. Autoscaling removes any inadvertent weighting of the variables that would otherwise occur due to differences in magnitude among the measurement variables.

The principal components can be computed directly from the data by singular value decomposition.[13] This algorithm generates the score and loading matrix of the data, and the eigenvalues of the principal components. The score matrix defines the coordinates of the data points in the principal component space, whereas the loading matrix defines the relationship between the original measurement variables and the new basis vectors describing variation. In other words, the principal components of the data can be reconstructed from the original measurement variables using information contained in the loading matrix. The eigenvalues tell how much variation in the data is contained in each principal component. The principal component representing the direction of largest variance in the data has the largest eigenvalue, the principal component representing the direction of next largest variance in the data has the second-largest eigenvalue, and so forth. The amount of information contained in a principal component relative to the original measurement space, i.e. the fraction of the total cumulative variance explained by the principal component, is equal to the eigenvalue of the principal component divided by the sum of all eigenvalues.

2.6 Soft Independent Modeling of Class Analogy

PCA is central to many of the more popular data analysis methods in PR. For example, a classification method based on PCA called SIMCA (soft independent modeling of class analogy) has been developed by Wold et al.[14–16] In SIMCA, a PCA is performed on each fuel class in the data set, and a sufficient number of principal components are retained to account for most of the variation within each class. The number of principal components retained for each class model will typically be different. Deciding on the number of principal components that should be retained for each class is crucial. Retention of too few principal components can distort the information content in the model about fuel type, whereas retention of too many principal components results in a lower signal-to-noise due to retention of noise-laden principal components. Using a procedure called cross-validation,[17] it is possible to determine the

model size directly from the training set data. That is, we can use cross-validation to find the number of principal components needed to adequately describe the signal in the data while simultaneously reducing the noise by not including the so-called secondary eigenvectors or noisy principal components in the class model.

The modeled variance, i.e. the variance explained by the class model, describes the signal. The noise in the data is described by the residual variance, that is the variance not accounted for by the principal component model. (The residual variance is explained by the so-called secondary principal components, which have been truncated or omitted.) Samples are classified in SIMCA by comparing the residual variance of a sample to the average residual variance of the samples comprising the class. In other words, an unknown fuel sample is projected onto a principal component map defined by the samples of a particular fuel class, such as JP-4. If a so-called unknown lies close to the JP-4 fuel samples in the principal component map, then it is considered to be a member of the JP-4 fuel class. However, an unknown is not considered to be a member of JP-4 if it is distant from the JP-4 fuel samples in the map.

The comparison of residual variances, i.e. the residual variance of an unknown with the average residual variance of the samples that comprise the class, provides a direct measure of the similarity of a sample to a particular fuel class. It is a measure of the sample's goodness of fit for a given principal component model. In order to provide a quantitative basis for this comparison, an F-statistic can be used[18] to compare the residual variance of a sample with the mean residual variance of the class. The F-statistic can be used to calculate an upper limit for the residual variance of those samples which belong to the class. The final result is a set of probabilities of class membership for each sample.

The advantages of using SIMCA to classify fuel spill data are threefold. First, an unknown is only assigned to the class for which it has a high probability. If the residual variance of a sample exceeds the upper limit for every fuel class in the training set, the sample would not be assigned to any of the fuel classes in the data set. It is either an outlier or comes from an unknown fuel class, that is, one not represented in the data set. Conventional classification methods[19] will forcibly assign the sample to one of the existing fuel classes in the training set, which would be a mistake. Second, some of the fuel classes in the data set may not be well separated. Hence, a future sample might be assigned to two or more groups by SIMCA, which would make sense, given the similar composition of the fuel samples, which comprise the overlapping fuel classes. Third, SIMCA is sensitive to the quality of the data used to generate the principal component models in the training set. As a result, there are diagnostics to assess the quality of the data. The modeling power[20] describes how well a variable helps the principal components to model variation, and discriminatory power[21] describes how well a variable helps the principal components to classify samples in the data set. Variables with low modeling power and low discriminatory power should be deleted as they only contribute noise to the principal component models in the data set.

The SIMCA method can work with as few as 10 samples from each class, and there is no restriction on the number of measurement variables, which is an important consideration in fuel spill identification problems, because the number of measurement variables often exceeds the number of fuel samples. Most classification techniques would break down in these situations because of problems arising from colinearity and chance classification.[22,23]

2.7 Pattern Recognition Analysis of Gas Chromatography Profile Data

PR analysis of GC profile data is performed in three distinct steps: peak matching, feature selection, and classification.

2.7.1 Peak Matching

Peak matching or peak identification is crucial in preparing gas chromatograms for PR analysis, because PR methods require each object in a data set to be represented as a data vector. Hence, translation of gas chromatograms into data vectors is the first step in any PR study of GC profile data – this requires the determination of the relative location of individual peaks corresponding to the same compound or compounds in the gas chromatograms that constitute the data set. Peak matching is best accomplished using a GC detector that can provide molecular information about the sample, such as a mass-selective detector. Unfortunately, mass-selective detectors are not as linear or sensitive as flame ionization detectors, which is an important consideration because each gas chromatogram is represented as a data vector or point in an n-dimensional measurement space. (Similarity between fuel spill samples is based on a Euclidean distance metric between pairs of points in the measurement space.) Hence, the linearity and sensitivity of the detector are important issues, which is the reason for the flame ionization detector being so widely used in fuel spill identification studies. However, special peak matching software is usually required to translate the gas chromatograms into data vectors when a flame ionization detector is used.

2.7.2 *Feature Selection*

Feature selection is also crucial in a PR study. It is important that we identify and delete peaks from the data set that contain information about experimental artifacts or other systematic variations in the data not related to the legitimate chemical differences between the various fuel classes in the data set. For fingerprinting experiments of the type that are being considered, it is inevitable that relationships will exist among sets of conditions used to generate the data and the patterns that result. One must realize this in advance when approaching the task of analyzing such data. Therefore, the problem is to use information contained in the gas chromatograms characteristic of the source profile without being swamped by the large amount of qualitative and quantitative information contained in the gas chromatogram due to experimental conditions. If the basis of classification for samples in the training set is other than desired group differences, unfavorable classification results will be obtained for a prediction set despite a linearly separable training set. The existence of these complicating relationships is an inherent part of fingerprint-type data. Hence, the goal of feature selection is to increase the signal-to-noise ratio of the data by discarding measurements on chemical components that are not characteristic of the source profile of the fuels represented in the data set.

2.7.3 *Classification*

Classification is usually accomplished by methods such as SIMCA which can effectively handle the attendant problems of colinearity and chance classification that occur with data that has a low object-to-descriptor ratio. Hence, PR is about reasoning, using the available information about the problem to uncover the structure within the data. Mean centering, autoscaling, peak matching, feature selection, and classification are an integral part of this reasoning process. Each plays a role in uncovering the structure of the data.

2.8 Genetic Algorithms for Feature Selection and Pattern Recognition

To use gas chromatograms to discriminate between different fuel types, it is necessary to examine the variability of large numbers of gas chromatograms. PR techniques such as PCA are ideally suited for this task, because they can display variability between a large number of chromatograms and show major clustering trends in large chromatographic data sets. Feature selection, i.e. selecting peaks that are characteristic of the fuel type, is a major goal in all of these studies.

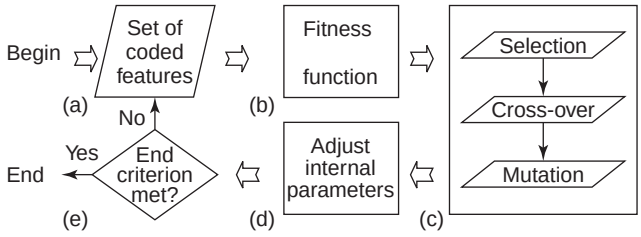

Figure 5 Block diagram of a GA developed for PR analysis. (Reprinted from B.K. Lavine, et al., 'A Genetic Algorithm for Pattern Recognition Analysis of Pyrolysis Gas Chromatographic Data', *Journal of Analytical and Applied Pyrolysis*, **50**, 47–62, Copyright 1999, with kind permission from Elsevier Science.)

Using a genetic algorithm (GA), it is possible to identify peaks in gas chromatograms that contain information about fuel type. A block diagram of a GA developed for fuel spill identification is shown in Figure 5. The GA selects features that optimize the separation of the fuel classes in a plot of the two largest principal components of the data. (In other words, the features selected by the GA yield principal component plots, which show sample clustering on the basis of fuel type.) Because the two largest principal components capture the bulk of the variance in the data, the peaks chosen by the GA contain information primarily about differences between the fuel classes in the data set. (As variance is synonymous with information, feature selection via this criterion causes information about class differences to emerge as the dominant source of variation in the features selected for PR analysis.) In addition, the algorithm focuses on those classes and/or samples that are difficult to classify as it trains using a form of boosting.[24] Samples that consistently classify correctly are not as heavily weighted in the analysis as samples that are difficult to classify. Over time, the algorithm learns its optimal parameters in a manner similar to a neural network.[25] The PR GA integrates aspects of strong and weak learning to yield a ''smart'' one-pass procedure for PR. The various components of the PR GA are described below.

2.8.1 *Population*

Selected features subsets are coded as binary strings called chromosomes. Each chromosome describes a unique set of features. A particular feature is present in a chromosome or binary string only if the corresponding bit in the string is set to 1. The length of each chromosome is equal to the number of features in the data set. The number of chromosomes in the initial population is ϕ. (Usually, ϕ is set to 100.) The chromosomes or binary strings comprising the initial population (i.e. the population at generation 0) are generated at random to minimize potential bias.

2.8.2 Fitness Function

With each generation, the GA computes class and sample weights (SW). These weights are an integral part of the fitness function, which is based on scoring the features according to their ability to optimize the separation of the different sample types in a plot of the two largest principal components of the data. The PCA routine serves as an embedded information filter, significantly reducing the size of the search space, because it restricts the search to feature sets whose principal component plots show clustering on the basis of fuel type. (If a plot of the two largest principal components for a set of features yields well-separated classes, one can only conclude that the bulk of the variance encoded by the set of GC peaks is about discrimination.) The fitness function of the GA contains both normalization and scoring components.

Normalization is used to adjust SW and class weights (CW) to preserve the following property: the sum of the CW is equal to 100, and the sum of the SW in a class is always equal to the CW. This facilitates the tracking and scoring of the chromosomes between generations. The normalization functions are given in Equations (1) and (2):

$$\mathrm{CW}(c) = 100\,\frac{\mathrm{CW}(c)}{\sum_{c}\mathrm{CW}(c)} \tag{1}$$

$$\mathrm{SW}_c(s) = \mathrm{CW}(c)\,\frac{\mathrm{SW}_c(s)}{\sum_{s\in c}\mathrm{SW}_c(s)} \tag{2}$$

where $\mathrm{CW}(c)$ is the weight of class c, and $\mathrm{SW}_c(s)$ is the weight of sample s in class c. Prior to the first generation, the user initializes the CW, with the SW being uniformly distributed in a class.

Scoring is performed on each principal component plot, which is generated for each chromosome after the subset of features coded in the chromosome has been extracted. A principal component plot is scored using the K-nearest neighbors. For a given sample point, Euclidean distances are computed between it and every other point in the principal component plot. These distances are arranged from smallest to largest, and a poll is taken of the point's K-nearest neighbors. For the most rigorous classification, K equals the number of samples in the class to which the point belongs. The sample hit count (SHC), or the number of like-nearest neighbors is $0 \le \mathrm{SHC}(s) \le K_c$. The fitness is computed using Equation (3):

$$\sum_{c}\sum_{s\in c}\frac{1}{K_c}\mathrm{SHC}(s)\mathrm{SW}(s) \tag{3}$$

The mean sample hit rate (SHR) for each sample over the entire population of chromosomes or feature subsets is computed via Equation (4), and is used to drive the boosting routine, where ϕ is the size of the population.

$$\mathrm{SHR}(s) = \frac{1}{\phi}\sum_{i=1}^{\phi}\frac{\mathrm{SHC}_i(s)}{K} \tag{4}$$

To understand scoring, consider a data set with two classes that have been assigned equal weights: class 1 has 10 samples, and class 2 has 20 samples. For uniformly distributed SW, class 1 samples will have a weight of 5 and class 2 samples will have a weight of 2.5, because each class has a weight of 50 and the SW in each class are uniformly distributed. Suppose a sample in class 1 has, as its nearest neighbors, seven class 1 samples in a principal component plot developed from a particular feature subset. Hence, $\mathrm{SHC}(c)/K_c = 7/10$, and the contribution of this sample to the fitness function for the particular feature subset equals 0.7×5 or 3.5. Multiplying SHC/K_c by $\mathrm{SW}(s)$ for each sample and summing up the corresponding product for the 30 samples in the data set yields the value of the fitness function for this particular set of features.

2.8.3 Reproduction

Selection, cross-over, and mutation operators are applied to the chromosomes. Fit strings are retained and selected for breeding, a process called selection, which is the first step toward population reorganization. The fit feature subsets are then broken up, swapped, and recombined, creating new feature subsets, which are introduced into the population of potential solutions. This process is called cross-over. In this study, the selection and cross-over operators are implemented by ordering the population of strings, i.e. potential solutions, from best to worst, while simultaneously generating a copy of the same population and randomizing the order of the strings in this copy with respect to their fitness. A fraction of the population is then selected as per the selection pressure which is set at 0.5. The top half of the ordered population is mated with strings from the top half of the random population, guaranteeing that the best 50% are selected for reproduction, while every string in the randomized copy has a uniform chance of being selected. (This is due to the randomized selection criterion imposed on strings from this population.) If a purely biased selection criterion were used to select strings, only a small region of the search space would be explored. Within a few generations, the population would consist of only copies of the best strings in the initial population.

For each pair of strings selected for mating, two new strings are generated using a variation of three-point cross-over (Figure 6). As in the case of simple three-point cross-over, the length of each new string is the same as the dimensionality of the data. Unlike simple

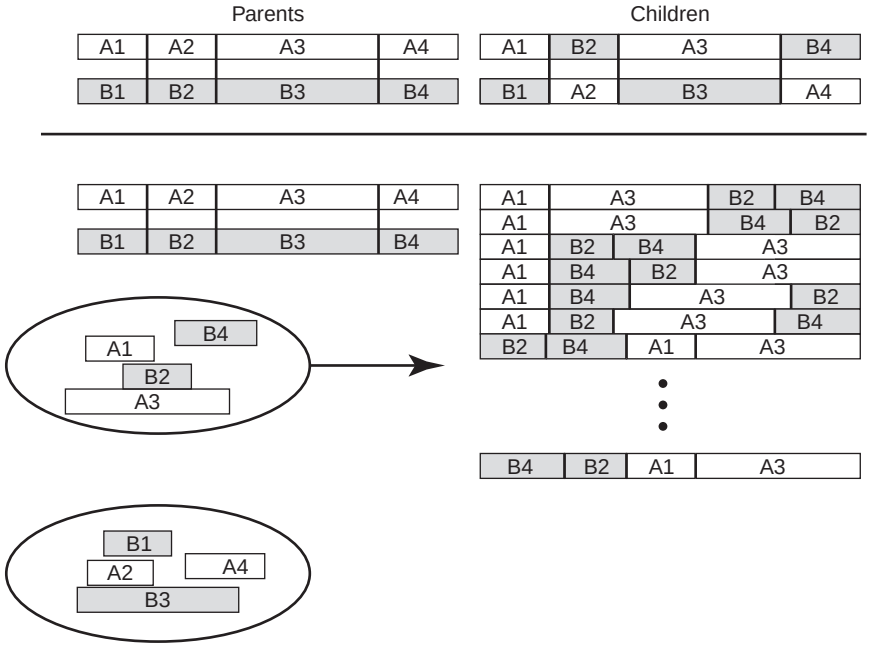

Figure 6 Three-point cross-over with mixing. Instead of swapping alleles and simultaneously preserving their position, four chromosome fragments are distributed and recombined at random with four factorial unique possibilities. (Reprinted from B.K. Lavine, et al., 'A Genetic Algorithm for Pattern Recognition Analysis of Pyrolysis Gas Chromatographic Data', *Journal of Analytical and Applied Pyrolysis*, **50**, 47–62, Copyright 1999, with kind permission from Elsevier Science.)

three-point cross-over, our cross-over operator is not compelled to preserve order among the exchanged string fragments, which safeguards the loss of information or features in the population. It will become less likely for the population variability to fall below a critical value due to the additional degree of freedom provided by the reordering. Furthermore, our variation of three-point cross-over may be useful in searching for good string arrangements. For example, if the current population has bad ordering, where features with a high synergism are spaced at great distances, simple cross-over would probably destroy potentially important allele packets. However, there is a chance to obtain good allele ordering by using a cross-over operator with a reordering algorithm embedded in it.

In the last step of reproduction, a mutation operator is applied to the new strings. The mutation probability of the operator is usually set at 0.01, so 1% of the feature subsets are selected at random for mutation. A chromosome marked for mutation has a single random bit flipped, which allows the GA to explore other regions of the parameter space. If the GA finds a better point, the genes from this point can invade the population, with the optimization continuing in a new direction.

The resulting population of strings, both the parents and children, are sorted by fitness, and the top ϕ strings are retained for the next generation. Because the selection criterion used for reproduction exhibits bias for the higher-ranking strings, the new population is expected to perform better on average than its predecessor. The aforementioned reproductive operators, however, also assure a significant degree of diversity in the population, because the cross-over points and reordering of exchanged string fragments of each chromosome pair are selected at random.

2.8.4 Adjustment of Internal Parameters

The GA is able to concentrate its efforts on classes and samples which are more difficult to classify by boosting their weights (Figure 7). There are two stages in boosting. In the first or learning stage, CW are adjusted relative to each other, in order to achieve an optimal configuration. The class hit rate (CHR), which is the average of the mean SHR for the samples in the class, is computed (Equation 5). Those classes with lower CHR values will be weighted more heavily than classes that score well. The change in the CW, which is computed using a neural network (Equation 6), is monitored throughout the run (where P is the momentum parameter for the neural network and is assigned a value by the user). If the average change in the weights is greater than some tolerance, the GA is said to be learning its optimal CW. Once the tolerance is reached, the CW are fixed and the SW in each class become uniformly distributed according to the CW. This initiates the second stage. The momentum,

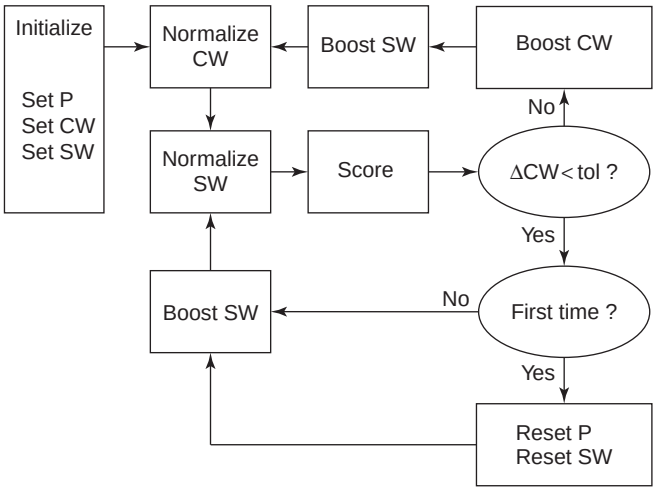

Figure 7 Block diagram of the boosting algorithm used to adjust the weights of difficult classes and/or samples. (Reprinted from B.K. Lavine, et al., 'A Genetic Algorithm for Pattern Recognition Analysis of Pyrolysis Gas Chromatographic Data', *Journal of Analytical and Applied Pyrolysis*, **50**, 47–62, Copyright 1999, with kind permission from Elsevier Science.)

which controls the rate at which the sample and CW are changed (Equations 6 and 7), is initially assigned a value of 0.8 while the GA is learning, but P is adjusted to 0.4 once the CW become fixed. These values have been chosen in part because they facilitate learning by the GA but do not cause a particular sample or class to dominate the calculation, which would result in the other samples or classes not contributing to the fitness function.

$$\mathrm{CHR}_g(c) = \text{Average } (\mathrm{SHR}_g(s): \forall_{s \in c}) \tag{5}$$

$$\mathrm{CW}_{g+1}(c) = \mathrm{CW}_g(c) + P(1 - \mathrm{CHR}_g(c)) \tag{6}$$

$$\mathrm{SW}_{g+1}(s) = \mathrm{SW}_g(s) + P(1 - \mathrm{SHR}_g(s)) \tag{7}$$

2.8.5 End Criterion

During each generation, class and SW are updated using the class and SHR from the previous generation. ($g+1$ is the current generation, whereas g is the previous generation.) The aforementioned procedure, which involves evaluation, reproduction, and boosting of the potential solutions, is repeated until a specified number of generations are executed or a feasible solution is found.

2.9 Pattern Recognition Approach to Fuel Spill Identification

Computer-based PR methods can be used to automate the classification of hydrocarbon contaminants from underground wells or aquifers. The approach involves the development of discriminants, which can model the chromatographic information that distinguishes the various fuel types from other suspected hydrocarbon sources. PR methodology attempts to increase the power of automated data interpretation to the point that weathering of fuels is not a limiting factor in the analysis.

The PR approach to fuel spill identification is based on four underlying principles. First, each fuel type has a unique chemical composition. Second, the aforementioned fuel class has a characteristic GC profile distinguishable from other sources of hydrocarbons. Third, information from the gas chromatograms can be encoded in such a way to be resistant to noise and interference. Fourth, the entire approach does not require the identity of the compounds responsible for the fouling involved, thus making the development of discriminants amenable to automated implementation.

2.10 Case Studies

PR methods have been applied to a large number of problems in fuel spill identification. Although the procedures selected for a given problem are highly dependent upon the nature of the classification problem, it is still possible to develop a general set of guidelines for applying PR techniques to real data sets. In this section a framework for solving the class membership problem, in the context of fuel spill identification, is presented by way of two recently published studies. In both of these studies, fuel spill samples were classified into their respective categories on the basis of their GC profiles.

2.10.1 Identification of Underground Fuel Spills

The first study involves the application of GC and PR methods to the problem of classifying weathered and unweathered jet fuels.[26] The training set consisted of 271 fuel samples representing six different types of jet fuels (JP-4, Jet-A, JP-7, JPTS, JP-5, and AVGAS) obtained from Wright Patterson Laboratory, OH, and Mulkiteo Energy Management Laboratory, WA. These two laboratories examine batches of fuels purchased for use by the Department of Defense to verify the authenticity of manufacturer's claims that purchased fuels meet the designated specifications. The fuel samples used in this study were splits from regular quality control standards collected over a 5-year period. The splits constituted a representative sampling of the fuels.

The fuel samples were stored in sealed containers at $-20\,^{\circ}\mathrm{C}$ prior to analysis by GC. The gas chromatograms of the neat jet fuel samples were used as a training set (Table 1). The prediction set consisted of 31 gas chromatograms of weathered jet fuel (Table 2). Of the 31 weathered fuel samples 17 were collected from sampling wells as a neat oil phase found floating on top of the well

Table 1 Training set

Number of samples	Fuel type
55	JP-4 (fuel used by USAF fighters)
77	Jet-A (fuel used by civilian airliners)
32	JP-7 (fuel used by SR-71 reconnaissance planes)
36	JPTS (fuel used by TR-1 and U-2 aircraft)
49	JP-5 (fuel used by Navy fighters)
22	AVGAS (common aviation jet fuel)

Table 2 Prediction set

Number of samples	Fuel type
11	JP-4 recovered from soil
8	JP-4 recovered as a neat oily phase
3	JP-4 weathered in a laboratory
5	JP-5 recovered as a neat oily phase
2	JPTS recovered as a neat oily phase
2	AVGAS recovered as a neat oily phase

water. Of the 31 fuel samples 11 were recovered fuels extracted from the soil near various fuel spills. The other three fuel samples had been subjected to weathering in the laboratory.

Prior to GC analysis, each fuel sample was diluted with methylene chloride. The diluted fuel samples were injected onto a capillary column using a split injection technique. High-speed GC profiles were obtained with a high efficiency fused silica capillary column 10 m long with an internal diameter of 0.10 mm and coated with 0.34 μm of a bonded and cross-linked 5% phenyl-substituted polymethylsiloxane stationary phase. The column was temperature programmed from 60 °C to 270 °C at 18 °C per minute using an HP-5890 gas chromatograph equipped with a flame ionization detector, a split/splitless injection port, and an HP-7673A autosampler. Gas chromatograms representative of the six fuel types are shown in Figure 8. The high-speed gas chromatograms were digitized using an HP-3357 laboratory automation system. The gas chromatograms were also peak matched using a computer program[27] that correctly assigned peaks by first computing the Kovat's retention index[28] for compounds eluting off the GC column. As the *n*-alkane peaks are the most prominent features present in the gas chromatograms of these fuels,[29] it was a simple matter to compute Kovat retention indices, that is to place the GC peaks

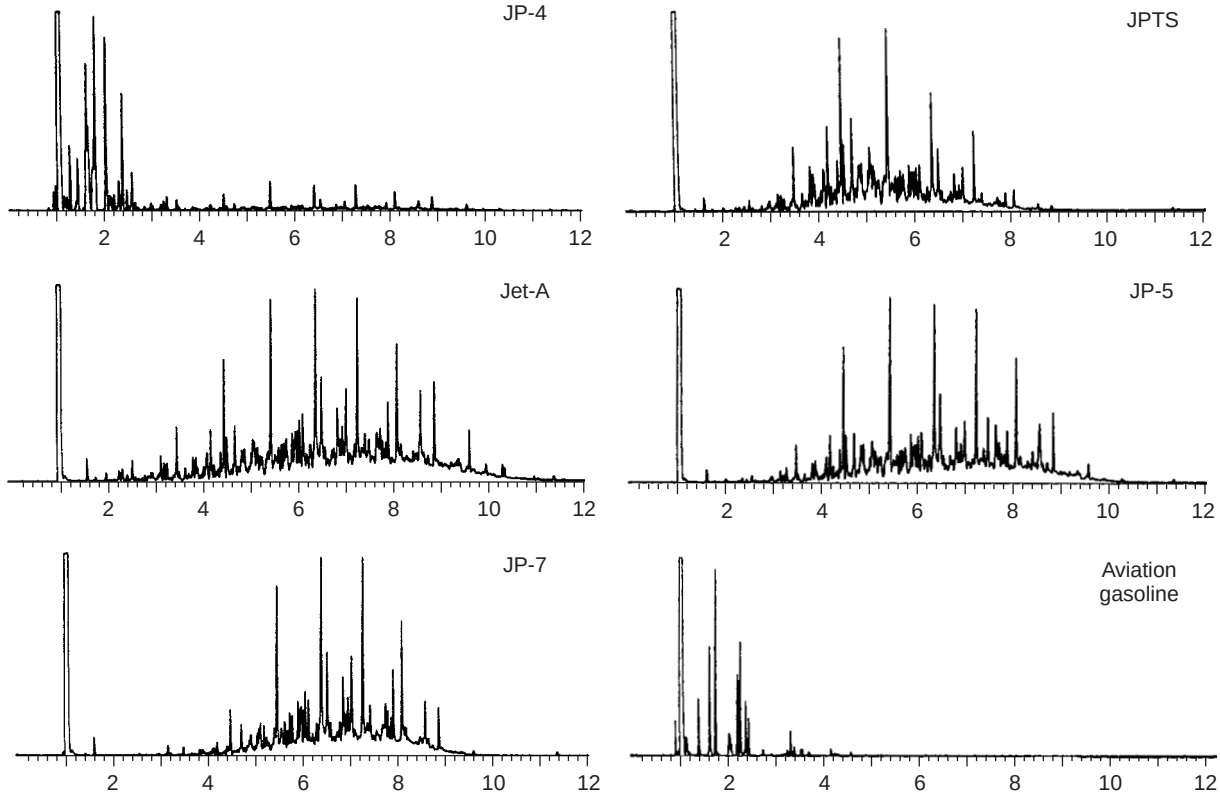

Figure 8 High-speed GC profiles of JP-4, Jet-A, JP-7, JPTS, JP-5, and AVGAS fuels. (Reprinted from B.K. Lavine, A.J. Moores, H.T. Mayfield, A. Faruque, 'Fuel Spill Identification by Gas Chromatography/Genetic Algorithms/Pattern Recognition Techniques', *Analytical Letters*, **31**(15), 2805–2820 (1998), by courtesy of Marcel Dekker.)

on a retention index scale rooted on predominant peaks observed in the gas chromatograms. The peak-matching program then analyzed the GC data in three distinct steps. First, a template of unique GC peaks for the set of 271 gas chromatograms was constructed by examining integration reports, and adding peaks to the template that did not match the retention indices of previously observed peaks. Second, a preliminary data vector was produced for each gas chromatogram by matching it against the template. If the peak was present, its normalized area from the integration report was assigned to the corresponding element of that vector. If the peak was not present, it was assigned a value of zero. Tailing peaks were excluded from the analysis and poorly resolved peaks were coalesced. Third, the number of times a particular feature was found to have a nonzero value was calculated. Those GC peaks below a user-specified number of nonzero occurrences were deleted from the data set, whereas peaks that passed the nonzero frequency criterion were retained. The peak-matching procedure yielded a final cumulative reference index containing 56 features though not all peaks were present in all gas chromatograms. Hence, for PR analysis, each gas chromatogram was initially represented as a 56-dimensional data vector, $x = (x_1, x_2, x_3, \ldots, x_j, \ldots, x_{56})$, where x_j is the area of the jth peak normalized to constant sum using the total integrated peak area.

As outliers can obscure relationships present in data, outlier analysis was performed on each fuel class in the training set prior to PR analysis using the generalized distance test[30] at the 0.01 significance level which was implemented via SCOUT.[31] Two JP-4, three Jet-A, two JP-7, four JPTS, two JP-5, and two AVGAS fuel samples were found to be outliers and were removed from the database. The set of data – 256 gas chromatograms of 56 peaks each – were standardized and autoscaled so that each variable (peak) had a mean of zero and a standard deviation of one within the entire set of 256 gas chromatograms.

PCA was used to examine the training set data, in order to obtain information about trends present in the GC data. Figure 9 shows a plot of the scores for the two largest principal components of the 56 GC peaks obtained from the 256 neat jet fuel samples. Each fuel sample or gas chromatogram is represented as a point in the principal component map. The JP-7 and JPTS fuels are well separated from each other and from the gas chromatograms of JP-4, Jet-A, AVGAS, and JP-5 fuels in the two-dimensional map, suggesting that information characteristic of fuel type is present in the high-speed gas chromatograms of the neat jet fuels.

The overlap of Jet-A and JP-5 fuel samples in the principal component map is not surprising because of their similar physical and chemical properties, such as flash point, freezing point, vapor pressure, and distillation curve.[32] Mayfield and Henley in an earlier study[33] observed that gas chromatograms of Jet-A and JP-5 fuels are more difficult to classify than gas chromatograms of other types of processed fuels, because of the similarity in their overall hydrocarbon composition. Nevertheless, Mayfield and Henley did conclude that fingerprint patterns characteristic of fuel type are present within gas chromatograms of Jet-A and JP-5.

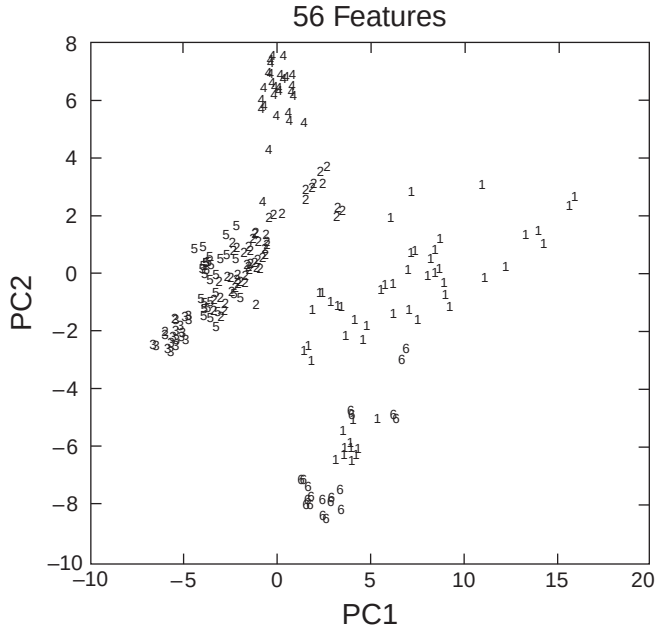

Figure 9 A plot of the two largest principal components of the 56 GC peaks for the 256 neat jet fuel samples: 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5, and 6 = AVGAS. (Reprinted from B.K. Lavine, A.J. Moores, H.T. Mayfield, A. Faruque, 'Fuel Spill Identification by Gas Chromatography/Genetic Algorithms/Pattern Recognition Techniques', *Analytical Letters*, **31**(15), 2805–2820 (1998), by courtesy of Marcel Dekker.)

A GA described in section 2.8 was applied to the 256 gas chromatograms to uncover features characteristic of the source profile of each fuel type. It identified features by sampling key feature subsets, scoring their principal component plots, and tracking those samples or classes that were difficult to classify. The boosting routine of the GA used this information to steer the population quickly to an optimal solution. After 100 generations, the GA identified 20 GC peaks, whose principal component plot showed clustering of the fuel samples according to fuel type (Figure 10). The principal component map of these 20 GC peaks suggests that information is present within the gas chromatograms of the jet fuels characteristic of fuel type. As PCA does not directly utilize class information in developing a map of the data, Figure 10 should be viewed as a conservative estimate of the differences in the hydrocarbon composition of the fuels as reflected by the 20 GC peaks selected by the GA.

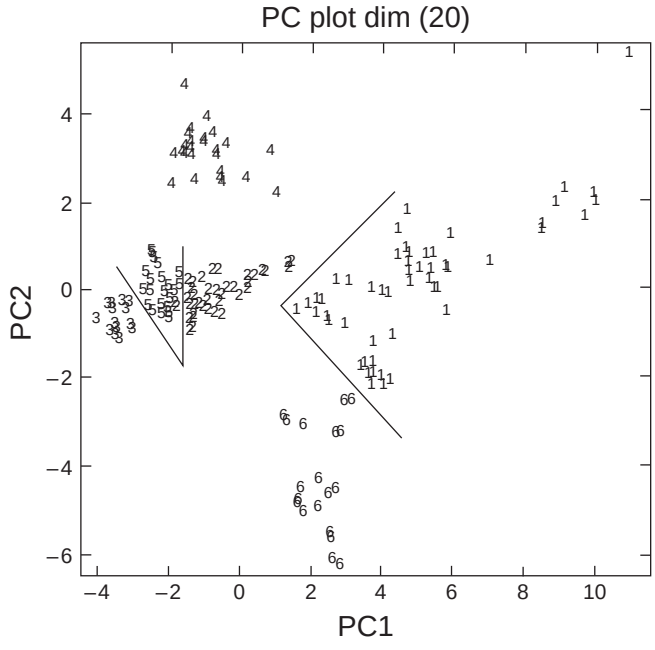

Figure 10 A plot of the two largest principal components of the 20 GC peaks selected by the GA. Clustering on the basis of fuel type is evident in this plot: 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5, and 6 = AVGAS. (Reprinted from B.K. Lavine, A.J. Moores, H.T. Mayfield, A. Faruque, 'Fuel Spill Identification by Gas Chromatography/Genetic Algorithms/Pattern Recognition Techniques', *Analytical Letters*, **31**(15), 2805–2820 (1998), by courtesy of Marcel Dekker.)

Table 3 Training set results[a]

Fuel type	Number in class	Right	Wrong	Right (%)
JP-4	53	53	0	100
Jet-A[b]	74	72	2	97.2
JP-7	30	29	1	96.7
JPTS	32	32	0	100
JP-5[c]	47	44	3	93.6
AVGAS	20	20	0	100
Total[d]	256	250	6	97.6

[a] Outliers were removed from the training set.
[b] Misclassified as JP-5.
[c] Misclassified as Jet-A.
[d] Bootstrapped error rate = 0.022 and cross-validated error rate = 0.023.

A six-way classification study involving JP-4, Jet-A, JP-7, JPTS, JP-5, and AVGAS fuel samples was also undertaken using SIMCA. Principal component models were developed from the 20 GC peaks for each of the six fuel classes in the training set. The number of principal components retained for each fuel class was determined by cross-validating on the total number of misclassifications. In other words, the number of principal components for each fuel class was adjusted until the optimal classification success rate was achieved for the training set. Results from the six-way classification study involving the 256 neat jet fuel samples are summarized in Table 3. The SIMCA classifier has a high recognition rate, i.e. a low apparent error rate, as well as a low bootstrapped and cross-validated error rate. Evidently, the 20 GC peaks selected by the GA contain information about fuel type.

A prediction set of 31 gas chromatograms (Table 2) was used to test the recognition ability of the 20 GC peaks and the discriminant, which these peaks supported. The gas chromatograms of the weathered fuels in the prediction set were run a few months before the gas chromatograms of the neat jet fuels were run. Of the 31 fuel samples in the prediction set 29 were correctly classified using the principal component models developed from the gas chromatograms of the neat jet fuels. Two JP-5 weathered fuel samples were classified as Jet-A, which is not surprising because of the similarity of Jet-A and JP-5 gas chromatograms. Notwithstanding, the high classification success rates obtained for the weathered fuel samples indicate that information about fuel type is also present in their gas chromatograms.

The results of this study are significant for two reasons. First, the change in composition that occurs after a jet fuel is released into the environment constitutes a major problem in fuel spill identification. These changes can be attributed to evaporation of lower-molecular-weight hydrocarbons, dissolution of water-soluble compounds, and microbial degradation. However, the weathered fuel samples used in this study were recovered from subsurface environments, where evaporation and dissolution are retarded. Hence, the predominant weathering factor is probably microbial action, which does not appear to have a pronounced effect on the GC profiles of the jet fuels. This suggests that the weathering of jet fuels in subsurface environments is greatly retarded in comparison to surface spills, thereby preserving the fuel's identity for a longer period of time.

Second, poorly resolved peaks were coalesced during the preprocessing step, which diminished both the resolution and information content of the gas chromatograms. Nevertheless, the GA was able to extract enough information from the profiles of the neat jet fuels to correctly classify gas chromatograms of the recovered fuels. This would suggest that simpler and cheaper GC instruments should be used to identify jet fuels. Cheaper and simpler instruments makes economic sense, and chemometric techniques such as GAs ensure that information present in chromatograms from these simpler, less informative instruments is extracted as completely as possible. This is the philosophy behind the approach to fuel spill identification described in this chapter, which we believe is

feasible because of the redundant nature of fuel spill data. To understand why GC data of fuels are redundant, consider again JP-4, which is a mixture of alkanes, alkenes, and aromatic hydrocarbons. The gas chromatograms of JP-4 fuels are characterized by a large number of early-eluting peaks that are large in area. There are only a few late-eluting peaks and their area is small. Clearly, a strong negative correlation exists between the early- and late-eluting peaks of JP-4 fuels. Furthermore, many of the alkane and alkene peaks are correlated, which should not come as a surprise because alkenes are not natural constituents of crude oil, but instead are formed from alkanes during the refining process. Finally, the specific property of a fuel that is most likely to be reflected in a high-resolution gas chromatogram is its distillation curve, which does not require a large number of peaks for full characterization. Clearly, an increase in resolution will increase the redundancy, but not necessarily the information content of the chromatographic data.

2.10.2 Solid-phase Microextraction/Gas Chromatography

Although GC has been used in the laboratory to type fuel spills, the method can also be adapted to the field by taking advantage of a sampling technique called solid-phase microextraction (SPME).[34–36] This micro-scale analytical extraction technique is designed for use with GC (i.e. SPME/GC). It utilizes a fused silica fiber coated with a polymer (e.g. polydimethylsiloxane) to extract organic compounds from aqueous samples. The fiber is attached to a modified syringe, and the constituents of the fuel are extracted by introducing the syringe into the headspace of the sample and depressing the plunger for a specified period of time. (When the plunger is depressed, the polymer-coated fiber is in contact with the sample.) The extraction efficiency of the SPME sampling method is a function of the partition coefficient of the compounds in the polymer coating. Following extraction, the plunger is retracted, and the SPME device is inserted directly into the injection port of the gas chromatograph. The plunger is then depressed, exposing the polymeric coating to the high temperature of the injector port, ensuring thermal desorption of the absorbed molecules. The absorbed molecules are immediately released into the carrier gas stream, where they are swept onto the column for separation and subsequent identification.

Using an SPME sampling device, it is possible to directly sample a layer of floating fuel in a water sample brought to the surface by a drilling rig. Hence, fuel spill identification can be performed on-site via a portable gas chromatograph equipped with a flame ionization detector and appropriate PR software. To test the suitability of SPME for the proposed application, 180 jet fuel samples representing JP-4, Jet-A, JP-7, JPTS, JP-5, and JP-8 were analyzed by headspace SPME/GC (Table 4). The fuel samples, which were obtained from Wright Patterson Air Force Base or Mulkiteo Energy Management Laboratory, were splits from quality control standards used by these laboratories to verify the authenticity of manufacturer's claims. These control standards constituted a representative sampling of the fuels. Thirteen fuel samples (seven JP-4, four JP-5, and two Diesel) recovered from the subsurface environment near Air Force and Navy airfields were also analyzed by headspace SPME/GC.

Table 4 Neat jet fuels

Number of samples	Fuel type
44	JP-4 (fuel used by USAF fighters)
48	Jet-A (fuel used by civilian airliners)
12	JP-7 (fuel used by SR-71)
15	JPTS (fuel used by TR-1 and U-2 aircraft)
19	JP-5 (fuel used by Navy jets)
42	JP-8 (fuel used by USAF fighters in NATO)[a]

[a] JP-8 will replace JP-4 as the principal United States Air Force fuel.

The following experimental protocol was used to obtain a GC profile of the jet fuel's volatile components. Four milliliters of a neat jet fuel were placed in a 40 mL volatile organic analysis (VOA) bottle. A microstirring bar was also placed in the VOA bottle prior to the introduction of the fuel sample to permit stirring of the sample during the SPME sampling period of 10 min. In a previous study 10 min was found to be sufficient to obtain a representative sampling of the volatile compounds present in a jet fuel.[37] The SPME procedures were performed manually using a commercial SPME holder. A single extraction fiber (100-μm polydimethylsiloxane, SUPELCO) was used to sample the headspace of each fuel sample in the training set. Previous studies in our laboratory have shown that recovered fuels cannot be analyzed by direct SPME sampling because of irreversible fiber damage caused by excessive swelling of the polydimethylsiloxane coating, which is a direct result of the fiber being in intimate contact with the fuel layer.

Gas chromatograms of the fuels were obtained using a high-efficiency fused silica capillary column 10 m in length with an internal diameter of 0.10 mm. The stationary phase was a 0.34-μm bonded and cross-linked 5% phenyl-substituted polymethylsiloxane phase. The GC column was temperature programmed using an HP-5890 gas chromatograph equipped with a flame ionization detector. The column oven was initially set at 40 °C for 3 min, and the column was then temperature programmed at 10 °C per minute to a final temperature of 250 °C and a final hold time of 0 min. The GC was operated in the split

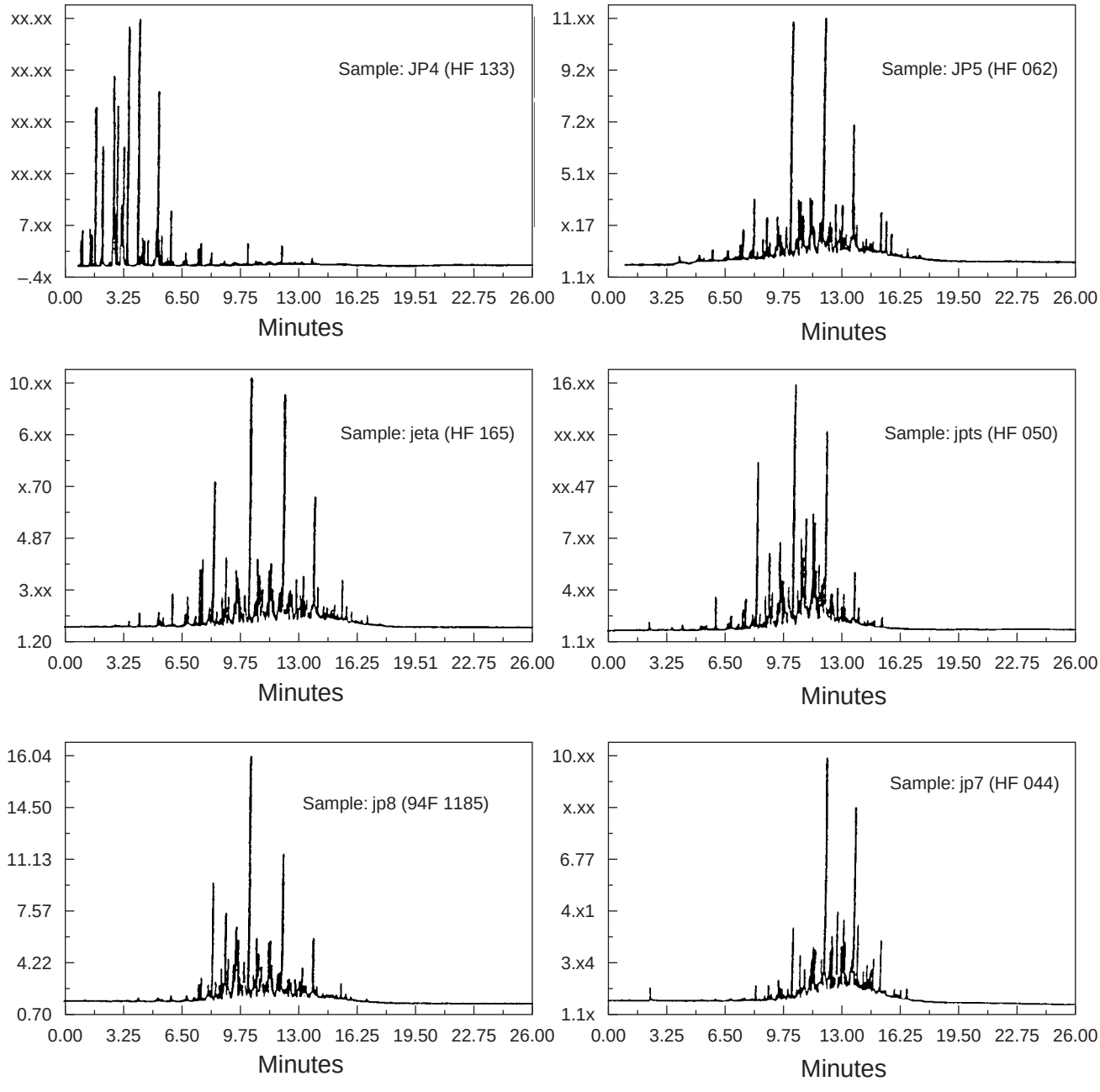

Figure 11 Headspace SPME/GC profiles of neat jet fuels. (Reprinted with the kind permission of the American Chemical Society.)

mode (30 : 1 split ratio) because of the high solubility of the fuel components in the polymer coating of the SPME. Gas chromatograms representative of the six fuel types in this study are shown in Figure 11.

The gas chromatograms were peak matched using a FORTRAN program called SETUP,[27] which analyzed the GC data in four distinct steps. Further details about SETUP can be found in the previous case study on identification of underground fuel spills. SETUP yielded a final cumulative reference file containing 84 features although not all peaks were present in all gas chromatograms. For PR analysis, each headspace chromatogram was initially represented by an 84-dimensional data vector, $x = (x_1, x_2, x_3, \ldots, x_j, \ldots, x_p)$, where x_j is the area of the jth peak normalized using the total integrated peak area for each chromatogram as the normalization factor.

Outliers can adversely influence the performance of PR techniques. Hence, outlier analysis was performed on each fuel class (see Table 4) prior to PR analysis using the

generalized distance test,[30] which was implemented via SCOUT.[31] Twenty-four chromatograms were identified as outliers and were removed from the training set. The outliers included three JP-4 chromatograms, seven Jet-A chromatograms, three JP-7 chromatograms, three JPTS chromatograms, four JP-5 chromatograms, and four JP-8 chromatograms. The large number of outliers in the data set was surprising and is probably due to some difficulties associated with using the SPME sampling device. (In a previous study involving the same neat jet fuel samples only about 3% of the samples in the training set were outliers.[6] The fuels in the previous study had been diluted with methylene chloride prior to their injection onto the same GC column.)

Figure 12 shows a principal component map of the 84 GC peaks obtained from the 156 gas chromatograms that constituted the training set (Table 5). Each chromatogram is represented as a point in the principal component map. The gas chromatograms of the JP-4 fuel samples are linearly separable from the gas chromatograms of the other jet fuels in the principal component map. However, the gas chromatograms of the Jet-A, JP-7, JPTS, JP-5, and JP-8 gas chromatograms overlap, which is not surprising because these are kerosene-based fuels, similar to Jet-A, the fuel used by civilian airliners. Mayfield and Henley[33] reported that gas chromatograms of kerosene-based fuels are more difficult to classify than gas chromatograms of other types of jet fuels, because of the similarity in their overall hydrocarbon composition.

Clearly, feature selection is crucial for increasing the signal-to-noise ratio of the data because measurements on chemical components that are not characteristic of the source profile of the fuels represented in the data set will be discarded. In this study, we used a GA (see section 2.8) to uncover features characteristic of the source profile of each fuel type. The GA identified features by sampling key feature subsets, and scoring their principal component plots. As it trained, the GA focused on those classes and/or samples that were difficult to classify. The boosting routine used this information to steer the population quickly to

Table 5 Training set

Fuel type	Number of chromatograms
JP-4	41
Jet-A	41
JP-7	9
JPTS	12
JP-5	15
JP-8	38
Total	156

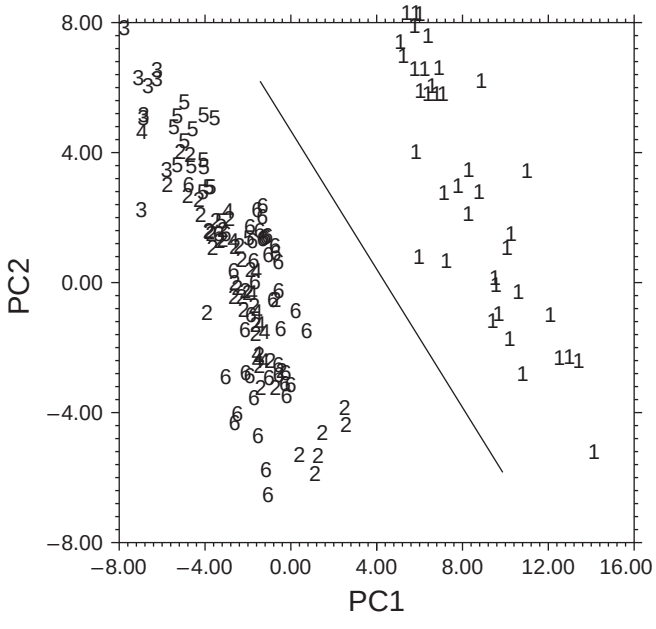

Figure 12 A principal component map of the 84 GC peaks obtained from the 156 neat jet fuel gas chromatograms. The two largest principal components explain 65% of the total cumulative variance. 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5, and 6 = JP-8. JP-4 is linearly separable from the other fuels in the map. (Reprinted with the kind permission of the American Chemical Society.)

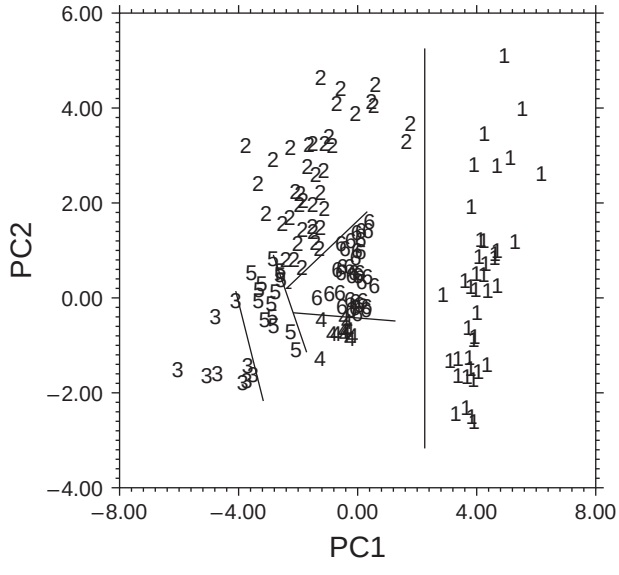

Figure 13 A principal component map of the 22 GC peaks obtained from the 156 neat jet fuel gas chromatograms. The two largest principal components explain 47% of the total cumulative variance. 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5, and 6 = JP-8. The fuels are linearly separable in the map. (Reprinted with the kind permission of the American Chemical Society.)

an optimal solution. After 100 generations, the GA identified 22 peaks. The principal component map of the 22 chromatographic peaks shows clustering of the jet fuel samples according to fuel type (Figure 13). As the data points are well separated by fuel type in the principal component map, it follows that a similar separation exists in the 22-dimensional measurement space. This conclusion is reinforced by the results of *K*-nearest neighbor,[38] which was also used to analyze these 22 chromatographic peaks (Table 6). The *K*-nearest neighbor classification method categorized the fuel samples in the training set according to their proximity to other samples (of pre-assigned categories) in the 22-dimensional measurement space. It is evident on the basis of *K*-nearest neighbor and the principal component map that information is contained in the 22 peaks characteristic of fuel-type.

The 22 GC peaks identified by the GA served as a starting point for the SIMCA study. The training set data was divided into six classes according to fuel type: JP-4, Jet-A, JP-7, JPTS, JP-5, and JP-8. PCA was performed on each fuel class in the training set to separate the information and noise sources. Truncation of the noninformative sources of variation led to a principal component model for each fuel class. The model size or complexity for each fuel class was determined by cross-validation.[15,17,18] For each fuel class, one principal component was used to model the data.

The gas chromatograms in the training set were then fitted to the principal component models, and the residual (the sum of the square differences between an original gas chromatogram and one reproduced by a principal component model) was computed for every gas chromatogram and model. Each gas chromatogram in the training set was classified on the basis of its goodness of fit via an F-test, which was performed using the residual variance of the sample together with the residual variance of the class.[18] The F-test determined whether a sample was inside or outside its assigned class and outside the other classes in the data. A gas chromatogram was assigned to the fuel class for which it had the lowest variance ratio. If the variance ratio exceeded the critical F-value for the class, then it was not assigned to it. Results from the SIMCA six-way classification study involving the training set samples are summarized in Table 7. The recognition rate for each principal component model was favorable.

Table 6 The *K*-nearest neighbor (NN) classification results for the training set

Class	Number	1-NN	3-NN	5-NN	7-NN
JP-4	41	41	40	40	40
Jet-A	41	41	40	40	40
JP-7	9	9	9	9	8
JPTS	12	12	12	12	12
JP-5	15	13	12	11	10
JP-8	38	38	38	38	38
Total	156	154	151	150	148

Table 7 Results from SIMCA classification

Class	Number in class	Right	Wrong	Percentage right
JP-4	41	41	0	100
Jet-A	41	38	3	93
JP-7	9	9	0	100
JPTS	12	10	2	83.33
JP-5	15	15	0	100
JP-8	38	31	7	81.6
Total	156	144	12	92.3

To test the predictive ability of these GC peaks and the principal component models associated with them, a prediction set of 20 neat jet fuel gas chromatograms and 13 weathered jet fuel gas chromatograms was employed (Table 8). For the neat jet fuel samples, we employed a different polydimethylsiloxane SPME fiber or varied the sampling time of the fiber over the headspace of the fuel, in order to assess the robustness of the SPME sampling technique. The 13 weathered jet fuel samples were obtained from sampling wells as a neat oily phase found floating on top of water. The sampling wells were located at Tyndall or Keywest Naval Air Station in Florida. The approach to fuel spill identification taken here, of classifying the gas chromatograms of weathered jet fuels using discriminants developed from the gas chromatograms of neat jet fuels, was taken because the physical and chemical interactions of jet fuel components with a subsurface environment are not yet fully understood.[29,39]

Every neat jet fuel sample in the prediction set was correctly classified. This result suggests that our SPME sampling technique is robust because variations in sampling time or fiber type did not adversely influence the classification of the neat jet fuels. The recovered fuels posed a more challenging identification problem. Of the 13 recovered jet fuels, 11 were correctly classified. Nine of the 11 were uniquely classified,[18] which would suggest a strong classification of the data. (The two diesel fuel samples were not assigned to any fuel class in the training set, which was construed as a correct classification because diesel fuels were not included in the training set.) However, two of the four JP-5 fuels were classified as Jet-A, which is not surprising as Jet-A and JP-5 are similar in their chemical composition. In a previous study,[6,9] we were able to correctly classify all of the weathered fuels using chromatographic data obtained by directly injecting fuels diluted with methylene chloride into the

Table 8 Prediction set

Sample	Fuel type	Source
HF112, fiber A	JP-4	[a]
HF114, fiber A	JP-4	[a]
HF120, fiber A	JP-4	[a]
HF144, fiber A	JP-4	[a]
HF119, fiber A	JP-4	[a]
HF136, fiber A	JP-4	[a]
HF112, fiber B	JP-4	[a]
HF114, fiber B	JP-4	[a]
HF120, fiber B	JP-4	[a]
HF144, fiber B	JP-4	[a]
HF119, fiber B	JP-4	[a]
HF136, fiber B	JP-4	[a]
HF112, 3 min	JP-4	[b]
HF112, 2 min	JP-4	[b]
HF112, 5 min	JP-4	[b]
HF112, 7 min	JP-4	[b]
HF112, 17 min	JP-4	[b]
HF112, 12 min	JP-4	[b]
HF112, 15 min	JP-4	[b]
HF112, 20 min	JP-4	[b]
D-26 valve Pit 2	Diesel	[c]
D-26 valve Pit 1	Diesel	[c]
Pit1UNK	JP-5	[d]
Pit1UNK	JP-5	[d]
Pit2UNK	JP-5	[d]
Pit2UNK	JP-5	[d]
T9W-PF012	JP-4	[e]
T9W-PF009	JP-4	[f]
T9W-PF011	JP-4	[f]
T9W-PF010	JP-4	[f]
T9W	JP-4	[f]
PF007	JP-4	[f]
PF008	JP-4	[f]

[a] Different 100-μm polydimethylsiloxane SPME fibers (than used in the training set) were employed to sample the neat JP-4 fuels in the prediction set.
[b] Sampling time of fiber was varied.
[c] Recovered diesel fuel sample from an excavation pit at Tyndall.
[d] Recovered JP-5 from a sampling pit at Keywest Naval Airstation.
[e] Third skim from a sampling well at Tyndall.
[f] Second skim from a sampling well at Tyndall.

gas chromatograph. Furthermore, the recognition rates obtained for samples in the training set were higher.[6,9] Evidently, there is a loss in classification power when sampling the headspace of the jet fuels using the SPME technique because only the most volatile components are sampled by the fiber.

3 CONCLUSIONS

The combination of GC and PR (GC/PR) is an excellent technique for the identification of underground fuel spills. The method is relatively rapid, sample preparation is minimal, and the instrumentation and software required for analysis is inexpensive and readily available. The chromatographic reproducibility and the discriminatory power necessary for tracing an underground fuel spill to its source can be achieved using conventional GC without resorting to complex clean-up procedures required by so-called traditional methods. Although spectrochemical techniques such as infrared and fluorescence spectroscopy have also been used to characterize other petroleum distillates, they do not show sufficient discriminatory power for jet fuels because the chemical composition of a jet fuel is primarily middle distillates, i.e. C_9 to C_{18} alkanes and alkenes. (Aromatics are only minor constituents.) In addition, contamination of a spill sample by salt, sand, water, or organic material can be a serious problem when using spectrochemical methods because there is no prior separation step. Hence, infrared and fluorescence spectroscopy generally do not yield spectra that would be characteristic of fuel type for these types of materials.

The GC/PR method described in this article is based on the concept of passive tagging, i.e. the composition of the fuel spill itself is used to identify the source of the spill. However, there is a better approach to this problem – so-called active tagging – where a known chemical label is introduced into a processed fuel. A suitable tag or label would be required to possess the following attributes. First, the tag should be resistant to weathering. Second, the tag does not alter the properties of the fuel. Third, the tag would not be harmful to the environment. Fourth, we must be able to detect the tag in the complex mixture that constitutes a jet fuel. At present, there are no suitable chemical tags for jet fuels. Perhaps, this situation will change in the near future.

ABBREVIATIONS AND ACRONYMS

CHR	Class Hit Rate
CW	Class Weights
GA	Genetic Algorithm
GC	Gas Chromatography
PCA	Principal Component Analysis
PR	Pattern Recognition
SHC	Sample Hit Count
SHR	Sample Hit Rate
SIMCA	Soft Independent Modeling of Class Analogy
SPME	Solid-phase Microextraction
SW	Sample Weights
VOA	Volatile Organic Analysis

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Diesel Fuels Analysis • Full Range Crudes, Analytical Methodology of

Chemometrics **(Volume 11)**
Chemometrics • Signal Processing in Analytical Chemistry • Soft Modeling of Analytical Data

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Data Reduction in Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

REFERENCES

1. S.Z. Cohen, S.M. Creeger, R.F. Carsel, C.G. Enfield, in *Treatment and Disposal of Pesticide Wastes*, eds. R.F. Kruager, J.N. Seiber, ACS Symposium Series No. 259, American Chemical Society, Washington, DC, 297–325, 1984.
2. A.P. Bentz, 'Oil Spill Identification', *Anal. Chem.*, **48**(6), 454A–472A (1976).
3. J.S. Mattson, C.S. Mattson, M.J. Spencer, S.A. Starks, 'Multivariate Statistical Approach to the Fingerprinting of Oils by Infrared Spectroscopy', *Anal. Chem.*, **49**, 297–302 (1977).
4. J.A. Pino, J.E. Murray, P.C. Jurs, B.K. Lavine, A.M. Harper, 'Application of Pyrolysis/Gas Chromatography/Pattern Recognition to the Detection of Cystic Fibrosis Heterozygotes', *Anal. Chem.*, **57**, 295–302 (1985).
5. B.K. Lavine, X. Qin, A. Stine, H.T. Mayfield, 'Application of Pattern Recognition Techniques to Problems in Advanced Pollution Monitoring', *Proc. Control Qual.*, **2**, 347–355 (1992).
6. B.K. Lavine, H.T. Mayfield, P.R. Kromann, A. Faruque, 'Source Identification of Fuel Spills by Pattern Recognition Analysis of High Speed Gas Chromatograms', *Anal. Chem.*, **67**, 3846–3852 (1995).
7. B.K. Lavine, A.J. Moores, H.T. Mayfield, A. Faruque, 'Genetic Algorithms Applied to Pattern Recognition Analysis of High Speed Gas Chromatograms of Aviation Turbine Fuels Using an Integrated Jet-A/JP-8 Database', *Microchem. J.*, **61**, 69–78 (1999).
8. B.K. Lavine, 'Environmental Applications of Pattern Recognition Techniques', *Chemolab*, **15**, 219–230 (1992).
9. B.K. Lavine, A. Stine, H.T. Mayfield, 'Gas Chromatography–Pattern Recognition Techniques in Pollution Monitoring', *Anal. Chim. Acta*, **277**, 357–367 (1993).
10. W.A. Warr, 'Computer-assisted Structure Elucidation', *Anal. Chem.*, **65**, 1045A–1052A (1993).
11. I.P. Jolliffe, *Principal Component Analysis*, Springer-Verlag Press, New York, 1986.
12. J. Edward Jackson, *A User's Guide to Principal Components*, Wiley Interscience, New York, 1991.
13. G. Golub, C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1971.
14. S. Wold, 'Cross Validatory Estimation of the Number of Components in Factor and Principal Component Models', *Technometrics*, **20**(4), 397–405 (1978).
15. S. Wold, 'Pattern Recognition by Means of Disjoint Principal Component Models', *Patt. Recog.*, **8**, 127–139 (1976).
16. G. Blomquist, E. Johnson, B. Soderstrom, S. Wold, 'Data Analysis of Pyrolysis Chromatograms by Means of SIMCA Pattern Recognition', *J. Chromatogr.*, **173**, 19–32 (1979).
17. M. Stone, 'Cross Validatory Choice and Assessment of Statistical Predictions', *J. R. Stat. Soc.*, **36**, 111–121 (1974).
18. S. Wold, M. Sjostrom, 'SIMCA: a Method for Analyzing Chemical Data in Terms of Similarity and Analogy', in *Chemometrics: Theory and Application*, ed. B.R. Kowalski, American Chemical Society, Washington DC, 243–280, Vol. 52, 1977.
19. P.A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York, 1975.
20. B.R. Kowalski (ed.), *Chemometrics: Mathematics and Statistics in Chemistry*, NATO ASI Series, D. Reidel Publishing Co., Dordrecht, 1984.
21. C. Albano, W. Dunn III, U. Edlund, E. Johansson, B. Norden, M. Sjostrom, S. Wold, 'Four Levels of Pattern Recognition', *Anal. Chim. Acta*, **2**, 429–443 (1978).
22. B.K. Lavine, P.C. Jurs, D.R. Henry, 'Chance Classifications by Nonparametric Linear Discriminants', *J. Chemomet.*, **2**, 1–10 (1988).
23. B.K. Lavine, D.R. Henry, 'Monte Carlo Studies of Nonparametric Linear Discriminant Functions', *J. Chemomet.*, **2**, 85–90 (1988).
24. Y. Freund, 'A Decision Theoretic Generalization of On-line Learning and an Application to Boosting', *Inf. Comput.*, **121**, 256–270 (1996).
25. P.D. Wasserman, *Neural Computing*, Van Nostrand Reinhold, New York, 1989.
26. B.K. Lavine, A.J. Moores, H.T. Mayfield, A. Faruque, 'Genetic Algorithms in Analytical Chemistry', *Anal. Lett.*, **31**(15), 2805–2822 (1998).
27. H.T. Mayfield, W. Bertsch, 'SETUP: a Chromatogram Transducing Program', *Comput. Appl. Lab.*, **1**, 130–137 (1983).

28. E. Kovats, in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, 230, Vol. 1, 1965.
29. W.E. Coleman, J.W. Munch, R.P. Streicher, H.P. Ringhand, W.F. Kopfler, 'GC-MS Studies of Jet Fuels', *Arch. Environ. Contam. Toxic.*, **13**, 171–180 (1984).
30. S.J. Schwager, B.H. Margolin, 'Generalized Distance Test for Outlier Analysis', *Ann. Stat.*, **10**, 943–953 (1982).
31. M.A. Stapanian, F.C. Garner, K.E. Fitsgerald, G.T. Flatman, J. Nocerino, 'Finding Suspected Causes of Measurement Error in Multivariate Environmental Data', *J. Chemomet.*, **7**, 165–176 (1993).
32. *Handbook of Aviation Fuel Properties*, Coordinating Research Council, Inc., Atlanta, GA, 1983.
33. H.T. Mayfield, M. Henley, *Monitoring Water in the 1990s: Meeting New Challenges*, eds. J.R. Hall, G.D. Glayson, American Chemical Society for Testing and Materials, Philadelphia, PA, 578–597, 1991.
34. Z. Zhang, M.J. Yang, J. Pawliszyn, 'Solid Phase Microextraction: a New Solvent Free Alternative for Sample Preparation', *Anal.Chem.*, **66**(17), 844A–852A (1994).
35. C.L. Arthur, L.M. Killam, S. Motlagh, M. Lim, D.W. Potter, J. Pawliszyn, 'Analysis of Substituted Benzene Compounds in Groundwater Using SPME', *J. Environ. Sci. Technol.*, **26**, 979–983 (1992).
36. C.L. Arthur, L. Killam, K.D. Buchholz, J. Pawliszyn, 'Environmental Analysis of Organic Compounds in Water Using Solid Phase Microextraction', *J. High Res. Chromatogr.*, **15**, 741–744 (1992).
37. B.K. Lavine, 'Solid Phase Microextraction Applied to the Problem of Fuel Spill Identification', Final Report for Summer Faculty Research Program, Armstrong Laboratory, Panama City, FL, 1995.
38. B.R. Kowalski, C.F. Bender, 'Pattern Recognition. A Powerful Approach for Interpreting Chemical Data', *J. Am. Chem. Soc.*, **94**, 5632–5640 (1972).
39. J.C. Spain, C.C. Sommerville, L.C. Butler, T.J. Lee, A.W. Bourquin, 'Degradation of Jet Fuel Hydrocarbons in Aquatic Communities', *USAF Report ESL-TR-83-26*, AFESC, Tyndall AFB, FL, 1983.

Volatile Organic Compounds in Groundwater, Probes for the Analysis of

B.M. Patterson and G.B. Davis
CSIRO Land and Water, Wembley, Australia

A.J. McKinley
University of Western Australia, Nedlands, Australia

1 Introduction 1
2 Theory of Volatile Organic Compound Probes 3
3 Operating Principle of Volatile Organic Compound Probes 5
4 Environmental Application 5
4.1 Calibration and Response Times 5
4.2 Interferences 6
5 Field Evaluation 6
5.1 Site Descriptions 6
5.2 Construction and Emplacement of the Monitoring Installation 6
5.3 Sampling of Volatile Organic Compound Probes and Groundwater 7
5.4 Comparison of Volatile Organic Compound Probe Data to Conventional Analysis 7
6 Advantages and Disadvantages of Volatile Organic Compound Probes 10
7 Conclusions 10
Acknowledgments 10
List of Symbols 10
Abbreviations and Acronyms 10
Related Articles 11
References 11

An automated system using a semipermeable polymer tubing [volatile organic compound (VOC) probe] coupled to an electrochemical sensor has been used for monitoring of VOCs in groundwater. The system operates by purging a VOC probe, which is in contact with groundwater, with gas at a low flow rate. The gas concentration of VOCs exiting the VOC probe can then be related to an external aqueous concentration according to Henry's law. The system enables semicontinuous in situ monitoring of total VOC concentrations in gaseous and aqueous environments.

The in situ VOC probes provided a reliable monitoring technique under long-term (between 1 and 3 months) aggressive and dynamic field conditions. The probes provided data that were comparable to conventional techniques without the labor-intensive sample collection and processing associated with conventional techniques, such as purge and trap or microextraction. Only the total VOC concentration can be determined using the electrochemical sensor; however, coupling the VOC probe to a gas chromatograph does enable compositional analysis to be determined.

1 INTRODUCTION

Groundwater pollution by organic compounds has been reported extensively in the USA and Europe.[1] In the USA, 90 000 cases of groundwater pollution were reported in the period 1989–1990.[2] Groundwater contamination results from leakage of organic fluids from underground storage tanks, accidental spills, inappropriate effluent disposal practices and leaching from landfills. The United States Environmental Protection Agency (US EPA) estimated that in 1992 that 1.6 million underground storage tanks were leaking.[3] The Organization for Economic Co-operation and Development has estimated the world market for environmental equipment and related services will increase from US $200 billion a year in 1990 to US $300 billion by the year 2000.

Petroleum products contain VOCs such as benzene, toluene, ethylbenzene, and xylenes (BTEX) which represent a substantial threat to groundwater quality because of their toxicity and aqueous solubility.[4,5] Other VOCs that are potential groundwater contaminants include chlorinated solvents such as trichloroethene (TCE) and tetrachloroethene (PCE) which have been used as industrial degreasers and dry-cleaning solvents.

These compounds are sparingly soluble in water, but have aqueous solubilities orders of magnitude greater than the maximum drinking water standards. For example, benzene (which is a known human carcinogen) has an aqueous solubility of 1790 mg L^{-1} and a maximum drinking water guideline value of 1 μg L^{-1}.[5] Small spills or leaks of these organic compounds have the potential to contaminate large volumes of groundwater.[6]

There are major difficulties involved in determining organic compound concentrations in groundwater. Access to groundwater samples requires the installation of wells or boreholes. Bores inappropriately constructed or installed using solvent-based glues, for instance can result in groundwater samples that are not representative of the aquifer chemistry. Also, replacing stagnant water

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

present within the borehole screen and casing with fresh groundwater from the surrounding aquifer by pumping boreholes (purging) has been shown to be critical to ensure representative groundwater samples.[7]

Another major difficulty is the reliability of sampling procedures and analytical techniques. Conventional sampling techniques frequently use synthetic polymers in the construction of groundwater monitoring and sampling equipment, such as submersible and gas-operated pumps,[8] and for example polyethylene tubing in pump discharge lines. Studies by Barcelona et al.,[9] Reynolds et al.,[10] and by Parker and Ranney[11,12] confirmed that VOCs are absorbed into polymeric material, including Teflon®. The only materials shown not to absorb VOCs were borosilicate glass and stainless steel. Therefore the use of polymeric tubing as pump discharge lines when sampling for organic compounds is of particular concern. Absorption of organic compounds into polymeric pump discharge lines would result in reduced groundwater concentrations. Desorption from contaminated polymeric pump discharge lines could result in contamination of clean groundwater, and give a false positive result. Exposure of groundwater samples to the atmosphere during sampling, storage, and analysis can also cause significant loss of VOCs by volatilization.[13]

Other potential problems with conventional sampling techniques are the potential for degradation of the more reactive organic compounds during transportation of the groundwater samples to the laboratory for processing. The use of on-site extraction techniques[14] can reduce this problem.

In situ on-site monitoring of organic compounds in groundwater has the potential to overcome a number of difficulties and costs associated with conventional sampling, processing, and analysis. In situ sampling can be more rapid and efficient and, where no groundwater pumping is required during monitoring, little disturbance of groundwater occurs.

A number of in situ monitoring techniques have been developed in recent years. Fiber optic chemical sensors (FOCS)[15–18] have been used to measure organic compounds dissolved in groundwater. These FOCS pass light of an appropriate wavelength to the measurement point by optical fiber. The fiber is terminated with a chemical sensor; interaction of the sensor with the target organic molecule changes fluorescence, absorption, or reflectance of the light. Optical couplers at the surface are able to separate reflected light from the excitation light, and this is analyzed using a spectrometer at the surface. These devices are reported to be able to determine individual organic compounds in complex mixtures. However, individual FOCS are required for each compound of interest, which detracts from the usefulness of these devices for in situ measurement in groundwater, where several organic compounds are often present together. Although these in situ monitoring techniques eliminate a number of sampling problems, the equipment needed for these systems is expensive and not readily transportable.

Another type of in situ monitoring device, termed a diffusion cell or dialysis bag, consists of a semipermeable membrane and relies on the transport and partitioning characteristics of gases and organic compounds across the membrane, into a gas phase or solvent to separate them from the aqueous environment. This system eliminates the need to collect a water sample.

Johnson[19] used hexane-filled cellulose dialysis bags suspended in a pond to monitor polychlorinated biphenyl concentrations. Under near steady state conditions, linear uptake of Arochlor™ 1248 into hexane within the suspended dialysis bag over 32 days resulted in a 100-fold increase in concentration in the hexane. However, the limitations of this technique included the inability to monitor VOCs and provide quantitative results.

Barber and Briegel[20] used diffusion cells consisting of coiled Teflon® tubing installed in backfilled boreholes with nylon access lines to the ground surface to measure methane dissolved in groundwater. Cells were purged with an inert gas that equilibrated with the dissolved methane in the groundwater. Patterson et al.[21] and Davis et al.[22] used a similar system to monitor dissolved oxygen concentrations during remediation of contaminated groundwater and soil.

There are a number of laboratory-based analytical techniques used to analyze VOCs. The most commonly used procedure is purge and trap. This process involves purging the VOCs from the water with an inert gas (hydrogen or helium) and trapping them onto an absorbent material. Purge and trap techniques offer extremely good sensitivity because all the VOCs extracted from generally large sample volumes are transferred to the chromatographic column. However, losses of VOCs from breakthrough of VOCs through Tenax™ traps have been reported. Bianchi et al.[23] observed approximately 15% loss of benzene due to breakthrough when purging through a 90×5 mm tube containing Tenax™ TA at a flow rate of $100\,mL\,min^{-1}$ for 60 min. Barnung and Grahl-Nielsen[24] also observed similar losses when purging through a 100×4.8 mm tube containing Tenax™ gas chromatography (GC) at a flow rate of $40\,mL\,min^{-1}$ for 12.5 min.

Microextraction has also been used to extract VOCs from water samples.[14] This technique is similar to conventional solvent extraction except that only small volumes of solvent are used and the evaporative concentration step is omitted. This technique is more cost effective than purge and trap, allowing twice as many samples to be analyzed in the same time; however, this technique is not as sensitive.

Other laboratory-based analytical techniques include membrane inlet mass spectrometry (MIMS).[25,26] This process involves pumping a water sample past a membrane attached to an inlet system of mass spectrometers, and organic compounds from the water sample preferentially diffuse through the membrane into the mass spectrometer where they are detected. LaPack et al.[27] investigated the effects of membrane thickness, temperature, and membrane inlet geometries on the MIMS process. Harland and Nicholson,[26] used MIMS to determine volatile chlorinated hydrocarbons in environmental samples from four field sites. Maden and Hayward[28] investigated the relative performance of different polymer sheets (silicone, latex, PVC, Teflon®, polyurethane, polyimide, polyethylene, and nitrile) for MIMS and concluded that there were several alternatives to the conventional silicone membrane material, depending on the aqueous matrix.

Solid-phase microextraction (SPME) involves inserting a fused silica fiber coated with a polymer into an aqueous sample and the organic compounds in the water diffuse into the polymer until equilibration is achieved between the fiber and water sample. The fiber is then removed from the water and inserted into a gas chromatographic injection port where the organic compounds are thermally desorbed. The coating on the fiber can range in thickness from 7 to 100 μm and includes poly(dimethylsiloxane) (a silicone) and polyacrylate polymers. This technique has been used by Arthur et al.[29] to analyze BTEX compounds in groundwater samples. Improvements to this technique have included: automation of SPME using a modified gas chromatographic autosampler;[30] headspace SPME analysis of volatile and semivolatile compounds from air, water and soil matrices;[31] on-line monitoring using SPME;[32] and coupling SPME to high-speed isothermal GC.[33]

Membrane extraction with a sorbent interface (MESI)[34] uses a 4 cm hollow fiber silicone tube (0.305 mm internal diameter, 0.470 mm outside diameter) that was placed in the headspace of a water sample and the inside of the tubing was flushed with hydrogen at a flow rate of 2.0 mL min^{-1}. Organic compounds that diffused through the tubing and into the inner gas space were trapped onto a fused silica fiber coated with 100 μm of silicone. The fused silica fiber was then periodically heated to thermally desorb the organic compounds for GC.

Hollow fiber stripping analysis (HFSA) was used by Zander and Pingert[35] and is a similar technique to MESI. This process was a modification to the more conventional closed loop stripping (CLS). Zander and Pingert[35] used microporous polypropylene hollow fibers to partition organic compounds from a water phase into a gas phase, where the organic compounds were trapped on activated carbon. The organic compounds were then stripped from the activated carbon with a solvent and analyzed by GC.

All of these techniques use GC analysis with either flame ionization detection (FID), electron capture detection (ECD) or mass spectrometry (MS) detection. Other analytical techniques have been developed that use similar extraction processes; however, they use specialized sensors to detect the VOCs. Stetter and Cao[36] passed a carrier gas (typically air) through the inner space of a silicone tube in contact with an aqueous solution of chlorobenzene. The gas flow rate through the tubing ranged from 110 to 516 mL min^{-1}. The chlorobenzene in the gas phase was detected using a chlorinated organic vapor sensor. A similar system was developed by Pawliszyn[37] using a solvent rather than a gas to purge the organic compounds from the polymer tubing. The chemical gradient produced from the solvent extraction was measured using a concentration gradient sensor and the response was related to the aqueous concentration.

Although these techniques provide quantitative data of organic concentrations of collected water samples, they do not overcome the problems associated with groundwater sampling (bore construction problems, volatilization, adsorption).

An automated system using a semipermeable polymer coupled to electrochemical sensors has been developed and patented[38] for the monitoring of VOCs in groundwater. This system was initially designed by the Commonwealth Scientific and Industrial Research Organization (CSIRO) in conjunction with the Cooperative Research Centre for Waste Management and Pollution Control. The system enables semicontinuous in situ monitoring of total VOC concentrations in gaseous and aqueous environments. The system operates by purging the polymer tubing with gas at a low flow rate. In aqueous environments, the gas concentration exiting the VOC probe can then be related to an external aqueous concentration according to Henry's law.

2 THEORY OF VOLATILE ORGANIC COMPOUND PROBES

When polymers are in contact with VOC-contaminated water, the organic compounds partition between the aqueous and polymer phases. At equilibrium, the ratio between the concentration in the polymer and the concentration in the aqueous phase has been termed the polymer–water partitioning coefficient (K_{pw}). Partitioning coefficients of selected VOCs between hollow silicone tubing and water over the microgram to milligram per liter aqueous concentration range have been determined.[39] It was found that VOCs were concentrated within the polymer tubing wall between 75 and 2700 times compared

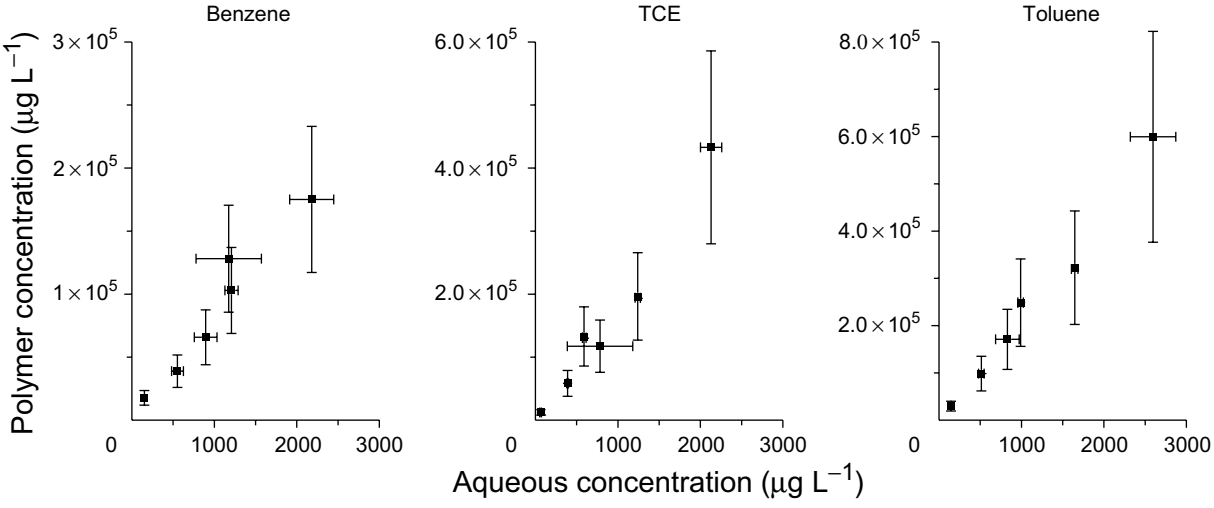

Figure 1 Equilibrium partitioning of benzene, TCE, and toluene between water and silicone polymer tubing. Error bars represent 95% confidence intervals.

to the aqueous phase. Over the aqueous concentration range tested, K_{pw} values were independent of concentration. Examples of equilibrium partitioning between the polymer and aqueous phases are shown for benzene, TCE, and toluene in Figure 1.

VOCs are also partitioned strongly into the gas phase within the polymer tube, and at equilibrium the gas concentration within the polymer tube could be related to an external aqueous concentration according to Henry's law. Due to the high storage of VOCs within the polymer, a similar relationship between the external aqueous phase and the internal gas phase of the polymer was found to still hold when the polymer was purged with a gas stream for a short time at a low flow rate. This gas purging technique has been shown to be a reliable technique over the microgram to milligram per liter aqueous concentration range for the organic compounds tested.[39] However, the gas–water partitioning coefficients (K_{gw}) at least doubled as the temperature was increased from 10 to 25 °C, indicating a significantly greater partitioning into the gas phase inside the tubing with temperature increases, as may be expected from the Henry's law relationship.

Although the relationship between gas concentration and aqueous concentration is temperature dependent, the effect of temperature can be estimated using the van't Hoff equation, and for groundwater monitoring large temperature fluctuations are unlikely. Even so, monitoring the purged gas concentration exiting the

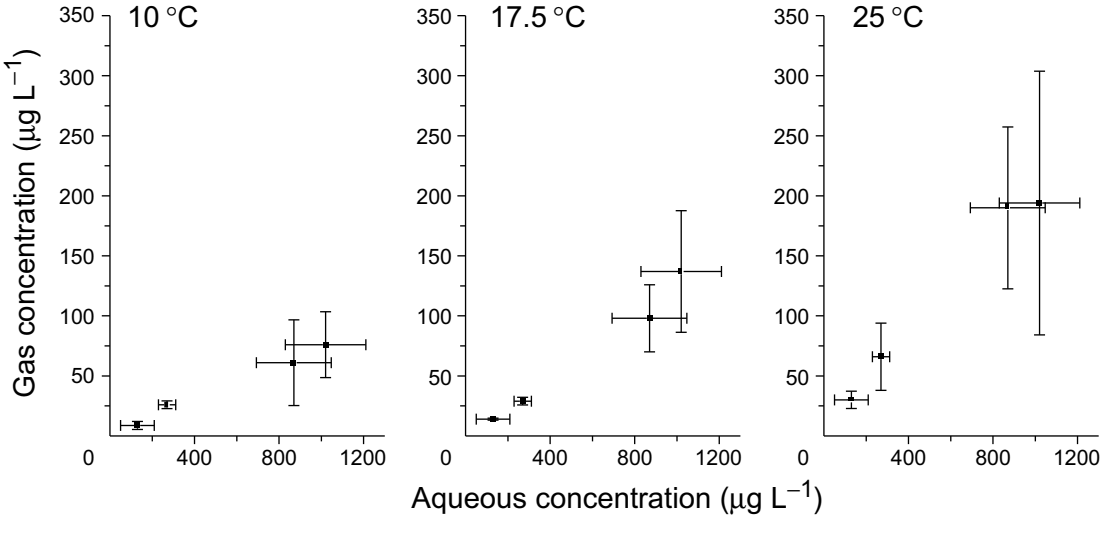

Figure 2 Plots of purged toluene gas concentrations exiting the silicone tube against various aqueous concentrations external to the silicone tube for 10, 17.5, and 25 °C. Error bars are 95% confidence intervals.

polymer tube and the aqueous temperature provides a technique to monitor the aqueous concentration external to the polymer. Examples of equilibrium partitioning at different temperatures for toluene between the internal gas phase of the polymer and the external aqueous phase are shown in Figure 2.

3 OPERATING PRINCIPLE OF VOLATILE ORGANIC COMPOUND PROBES

The VOC unit consists of a probe and a sensor unit. The probe consists of a length of coiled silicone tubing that is intermittently purged with air at 5 mL min^{-1}. The probe also contains a thermistor to monitor environmental temperature. The probe is immersed either in an aqueous, soil, or gas environment. The probe is connected to the sensor unit via a cable (up to 30 m in length) containing influent and effluent gas access lines and a thermistor wire. The access lines are constructed of semi-rigid nylon and found not to significantly sorb/desorb organic compounds for the length of access lines used. For cable lengths greater than 30 m, stainless steel access lines could be used.

The sensor unit contains a solenoid valve, data logger and a gas hydrocarbon sensor. With the solenoid valve, air either passes through or bypasses the probe before reaching the sensor. The solenoid valve is intermittently operated to purge the probe. The data logger controls the opening of the solenoid valve and stores the thermistor and sensor data. A diagram of a VOC probe and sensor unit is shown in Figure 3. From the data logger readings and an appropriate calibration curve, a value for the VOC concentration external to the probe is calculated.

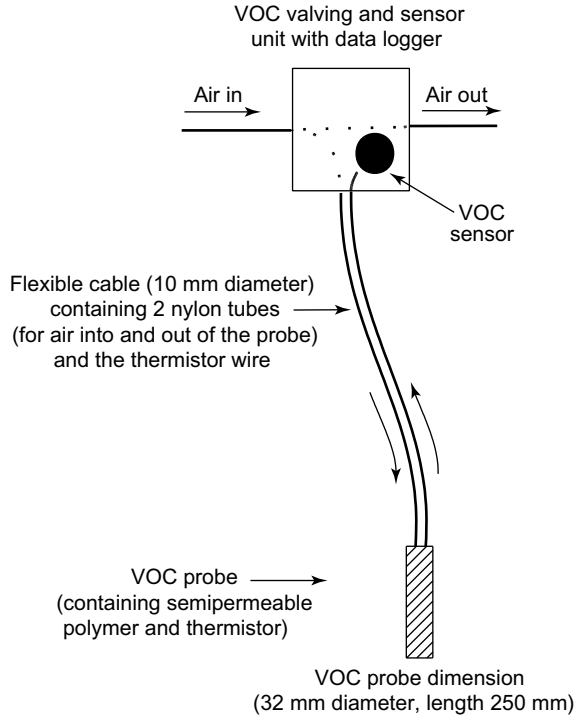

Figure 3 Diagram of VOC probe and sensor unit.

4 ENVIRONMENTAL APPLICATION

4.1 Calibration and Response Times

Calibration of the VOC probe and sensor unit has been carried out in the laboratory. Comparison of measured benzene concentrations in the aqueous phase external to the probe with sensor response data from the VOC probe and sensor unit is shown in Figure 4. Although the response of the VOC probe and sensor unit was not linear, a calibration curve could be established.

The nonlinear response was attributed to the sensor used in the sensor unit, not the VOC probe. A similar nonlinear response was observed when gas samples were monitored directly (no VOC probe used). Also, when gas chromatography/mass spectrometry (GC/MS) analysis was used, linear responses were observed (see Figure 2).

The sensor used in the sensor unit also gave different responses for different classes of organic compounds. Therefore, a single-component calibration curve may not give quantitative analyses for aqueous samples containing a single noncalibrated compound or multicomponents. However, in situ (re)calibration of the sensor unit for quantitative monitoring can be undertaken by collecting groundwater samples from the same location as the VOC probes.

Based on the calibration curve data, the analytical range of the VOC probe and sensor unit for aqueous

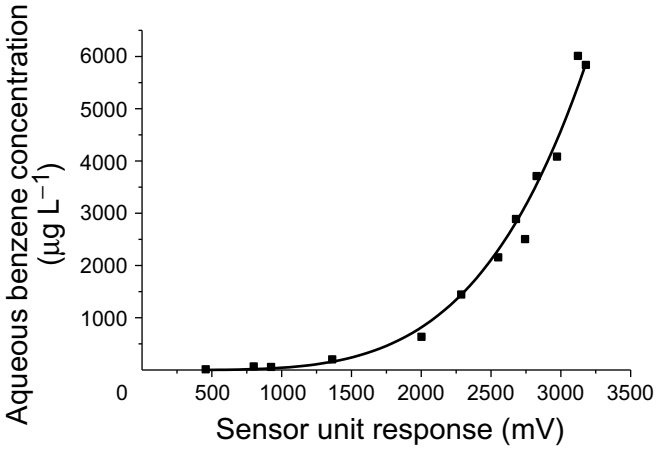

Figure 4 Typical benzene calibration curve.

benzene concentrations would be from a detection limit of approximately 200 μg L^{-1} to approximately 5000 μg L^{-1}.

Response times to step changes in aqueous concentrations of VOCs ranged between 1.7 and 5.2 h.[39] The response time (t_{90}) was defined as the time required to achieve 90% of the stable aqueous organic concentration after a step change in the aqueous organic compound concentration. These response times were determined from laboratory experiments where the water bath was relatively well mixed.[39] In groundwater systems, mixing of waters is relatively poor, relying only on groundwater flow and diffusion. Therefore, these estimated response times would be minimum values for groundwater systems, and more typical response times may be significantly greater. However, fluctuations in groundwater concentrations would also be slow, of the order of days to weeks depending on groundwater flow velocities.

4.2 Interferences

Due to the semipermeable nature of the polymer in the VOC probe and the relatively high gas–water partitioning of the VOCs, the VOC probe and sensor unit detects a relatively narrow range of compounds, thus reducing matrix contamination. Therefore, samples can be analyzed without further clean-up. The only compound that significantly interferes with the VOC probe and sensor unit would be methane; however, the VOC sensor used is approximately two orders of magnitude more sensitive to the VOCs compared with methane.

Temperature does effect the partitioning of VOCs into the probe; however, for groundwater monitoring, large temperature fluctuations are unlikely. Changes in VOC probe and sensor unit responses due to temperature changes are accounted for by monitoring the external temperature and recalculating the aqueous concentration using the van't Hoff equation.

5 FIELD EVALUATION

To assess the VOC probe and sensor unit under field environmental conditions, VOC probes were installed and used to monitor organic concentrations in groundwater at two field sites in Western Australia. Groundwater samples from these field sites were also collected and analyzed using conventional groundwater sampling and analysis techniques. This enabled a comparison of the two techniques and thus an evaluation of the validity of the new technique. At one site, the VOC probes were operated manually to obtain compositional data of groundwater contamination. At the second site, an automated VOC probe and sensor unit was used to obtain total VOC concentrations.

5.1 Site Descriptions

Concentrations of BTEX in groundwater were investigated near a petrol service station in the Perth metropolitan area, Western Australia. At this site a leaking underground storage tank had resulted in contamination of the groundwater with dissolved BTEX compounds. The contaminated groundwater plume was over 420 m long, 0.5–3 m thick and 20–40 m wide. The groundwater at this site was anoxic. A monitoring installation was emplaced approximately 80 m down-gradient from the contamination source. At this location, the depth to groundwater was 3.9 m below ground surface, with a maximum plume thickness of 1.5 m. Groundwater velocity at the site was approximately 150 m $year^{-1}$. Details of the site hydrogeology and the contaminant plume are given by Davis et al.[40,41]

The second site at a fuel storage facility in Western Australia was contaminated with dissolved BTEX compounds. The water table was located approximately 3.2 m below ground surface. Further details of the site hydrogeology and soil/groundwater contamination are given by Johnston et al.[42]

5.2 Construction and Emplacement of the Monitoring Installation

At the service station site, four prototype VOC probes were bundled together to form a multilevel (multidepth) monitoring installation. The centre of each probe was 25 cm apart vertically, which enabled monitoring over a 1.0 m depth interval. Small (4.7 mm outside diameter, 3.1 mm inside diameter) stainless steel minipiezometers with short (5 cm) slotted aluminum screens were also included with the probes in the multilevel installation and screened at the same depth as the centre of the probes. This was done to allow conventional sampling and analysis of groundwater for comparison with data from the VOC probes.

At the fuel storage facility, prototype VOC probes were used during an in situ air-sparging trial. At the main monitoring location, seven VOC probes were bundled together to form a multilevel (multidepth) monitoring installation. The centre of each VOC probe was spaced 1.0 m apart vertically. Additionally, small-diameter (3.18 mm outside diameter, 2.00 mm inside diameter) nylon minipiezometers with short (20 cm) slotted aluminum screens were included in the multilevel installation and screened at a similar depth to the VOC probes. Again, this was to allow more conventional sampling and analysis of groundwater for comparison with data from the probes.

To emplace the installations at each site, a hole was drilled using 50 mm solid augers. A 40 mm PVC casing was then sludged into position and the multilevel installation

was lowered into the temporary-cased bore. The casing was then withdrawn, letting sand collapse around the multilevel installation.

5.3 Sampling of Volatile Organic Compound Probes and Groundwater

At the service station site the multilevel installation was left for 3 months to allow groundwater chemistry to restabilize before sampling commenced. Groundwater and VOC probes were then sampled in triplicate approximately twice a week for a period of 3 months.

VOC probes were sampled manually by passing high-purity nitrogen gas through each probe via the access line at a rate of 5 mL min^{-1}. Probes were purged for 5 min before sampling. During this time, effluent gas from each probe was discharged to the atmosphere. To sample the organic compounds in the effluent gas purged from the probes, absorbent traps (containing Tenax™ TA) were connected to the end of the access line for 4 min, and the organic compounds in the effluent gas were collected on the adsorbent traps. The effluent gas was sampled in triplicate. Traps were then transported back to the laboratory where the organic compounds were desorbed from the traps using a Dani tube desorber and analyzed by GC/MS. Details of the analytical technique are given in Patterson et al.[14]

At the fuel storage facility, the multilevel installations were left for 1 week before the air sparging trial commenced. Groundwater was then conventionally sampled 4–20 times per week during the 4-week air-sparging trial using the technique described above. An automated prototype VOC probe and sensor unit (connected by a 15 m cable) was used to monitor total VOC concentrations for each of the seven probes every 3 h. Probes were purged with air for 15 min prior to recording the sensor response.

Groundwater samples from both sites were collected from the minipiezometers by applying suction using a glass syringe. Stagnant water within the minipiezometers was first purged using the same technique. Groundwater samples were extracted on-site using a microextraction technique and returned to the laboratory for analysis by GC/MS.[14]

5.4 Comparison of Volatile Organic Compound Probe Data to Conventional Analysis

At the service station site, BTEX concentrations measured in the gas purged from within the VOC probes and groundwater temperature data were used to estimate aqueous concentrations of the BTEX compounds in groundwater external to the probes. Lateral movement of the groundwater plume[40,41] between the time of installation and monitoring resulted in detection of groundwater contamination at only the top monitoring location (4.0 m below ground surface). Various VOCs were measured in the gas collected from the probe at this location. The main compounds identified in the effluent gas were benzene, toluene, ethylbenzene, *m*- and *p*-xylene (not chromatographically separated), and *o*-xylene. Results from the 3 months of field monitoring are shown in Figure 5. Benzene data from the VOC probe showed good agreement with data obtained using conventional groundwater monitoring ($r = 0.861$). For toluene, ethylbenzene, *m*- and *p*-xylene and *o*-xylene, results for the early part of the field experiment (between days 120 and 140) indicated that the VOC probe was predicting groundwater concentrations reasonably well. However, between days 140 and 240 the VOC probe was giving higher groundwater concentrations.

Possible reasons for the observed variations in results may be the temperature/K_{gw} calibration, delays in response of the VOC probe under poorly mixed groundwater conditions, or the two techniques may have been measuring slightly different groundwaters due to slightly different monitoring locations.

As changes in groundwater temperature over the 3 month field trial were relatively minor, with the temperature decreasing from 24.0 to 21.2 °C, possible errors associated with the temperature/K_{gw} calibration would not explain the differences in the observed results. Also, if temperature effects were the reason for the observed results, then the results for benzene should also show the same trend as the other compounds. However, benzene did not show increasing concentrations with time compared to conventional monitoring.

Another possible reason for variations between the VOC probe data and traditionally measured groundwater data for the less-volatile VOCs may be 'memory' effects of the polymer due to a combination of the greater K_{pw} values for toluene, ethylbenzene, *m*- and *p*-xylene, and *o*-xylene (K_{pw} values between 210 and 610)[39] compared to benzene ($K_{pw} = 75$)[39] and poorly mixed groundwater conditions. With greater K_{pw} values there is a greater mass of the particular organic compound stored within the polymer. Under poorly mixed conditions such as in groundwater, the greater mass may result in greater inertia in the system when there is a change in aqueous concentrations. Therefore, VOC probes may tend to underestimate concentrations when groundwater concentrations are increasing, and overestimate concentrations when groundwater concentrations are decreasing. However, this possibility would not explain the observed field data, as the VOC probe data was higher than traditionally measured groundwater data during the period when groundwater concentrations were generally increasing.

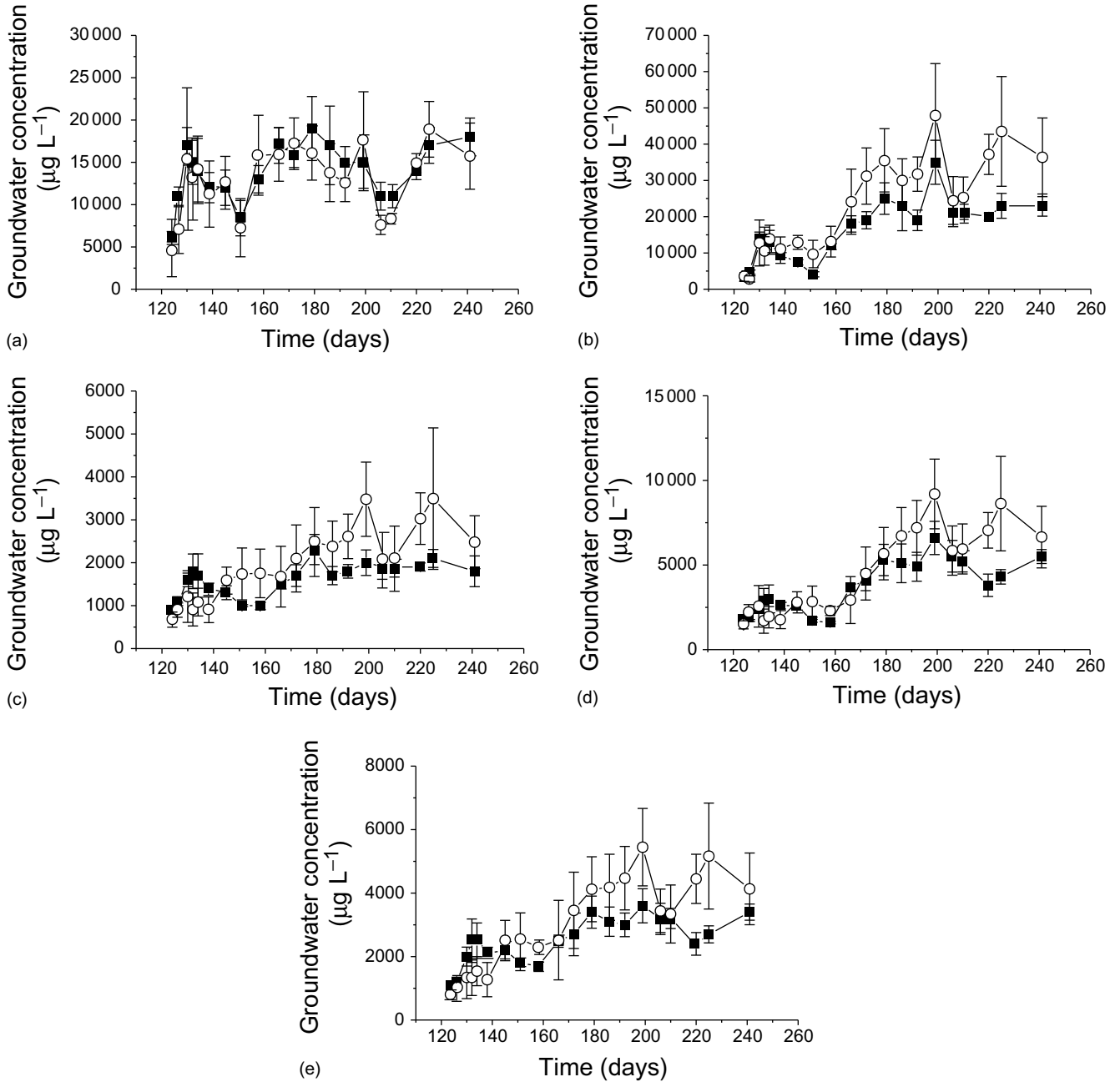

Figure 5 Groundwater concentrations of (a) benzene, (b) toluene, (c) ethylbenzene, (d) *m*- and *p*-xylene, and (e) *o*-xylene, determined using (■) conventional monitoring techniques and (○) manually purged VOC probes. Error bars are 95% confidence intervals.

Another cause for the discrepancies between the probe and the conventional samples for toluene, xylenes, and ethylbenzene may be incomplete desorption from the adsorbent traps.

A further reason for the observed results is that, although the VOC probes and minipiezometers were in close proximity, they were sampling slightly different groundwater. VOC probes covered a depth interval of 20 cm, whereas the minipiezometers only covered a depth interval of 5 cm. Groundwater concentrations at this location varied significantly with depth. No groundwater contamination was detected in groundwater collected from the minipiezometer at the next depth interval (25 cm deeper). Also, the groundwater contamination plume has been observed to move laterally due to changes in flow direction as a result of seasonal recharge.[41] At this site the benzene plume is larger than the plumes of the other organic compounds, due to lower

degradation and lower retardation of benzene.[43] This may also explain why variations for benzene were not as significant between the VOC probe results and groundwater samples, as was the case for the other organic compounds.

At the fuel storage facility, during the air-injection phase of the air sparging trial, desaturation of the aquifer caused by air injection resulted in air entry into a number of the small-diameter minipiezometers, and these groundwater samples could not be collected. Therefore only snapshot pictures of VOC concentrations could be determined by conventional sampling of these minipiezometers when sparging was halted. Also, the extent and distribution of VOCs could not be estimated accurately using these minipiezometers when air was being injected, due to possible groundwater movement away from the sparging well during air injection and the corresponding return when air injection ceased.

The VOC probes had the advantage that monitoring could be carried out during the air injection phase of the trial, because they rely on an equilibrium between the probe and the external aqueous or gaseous environment. This provided more continuous data. Also, because the process was automated, a greater frequency of data collection was possible. From this monitoring, small timescale trends could be observed which would be difficult to infer from conventional monitoring data.

Plots of total VOC concentrations estimated from VOC probes and conventional sampling with GC/MS analysis are shown in Figure 6 for selected depths. For the GC/MS data, total VOC concentrations were estimated from the summation of BTEX concentrations. The BTEX compounds were the most dominant compounds detected in the groundwater. Other organic compounds detected, such as trimethylbenzenes and two-ring aromatic compounds, contributed to less than 5% of the total detected mass. The response from the sensor unit was calibrated to the total VOC concentrations at the 10 m below-ground location using the GC/MS data.

In general, groundwater VOC concentrations determined from the VOC probes showed good agreement with data obtained from the conventional technique. Once the air sparging trial commenced, both techniques showed that VOC concentrations decreased rapidly at the shallow depths (to 10 m below ground surface) at the main monitoring location. VOC probe data also showed that at a number of locations VOC concentrations decreased during air injection, but rebounded once air injection ceased. This phenomenon was not observed using conventional monitoring techniques, as sampling could not occur during air injection.

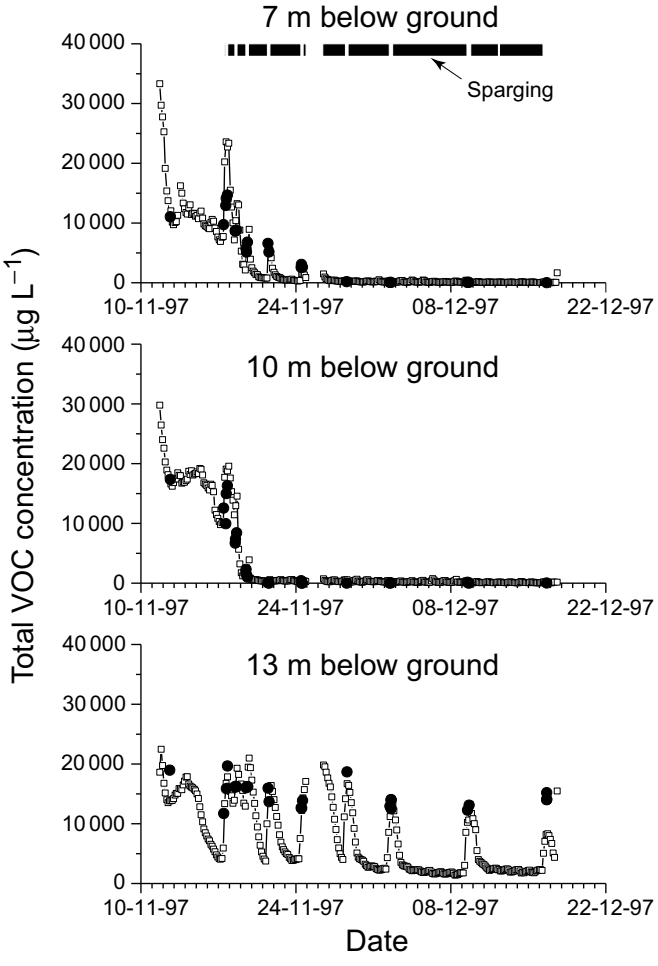

Figure 6 Plots of total VOC concentrations estimated from ( ) VOC probes and (●) conventional sampling with GC/MS analysis for three depths (7, 10, and 13 m below ground surface) at location DP4. All the VOC probes were calibrated using conventional sampling/analysis data from groundwater samples collected 10 m below ground surface. The bar on the 7 m below-ground plot indicates when air injection was on (blackened) and off (white spaces).

This rebound effect provides good evidence of the lateral movement of contaminated groundwater around the probe locations and indicates this monitoring location is likely to be close to the edge of the sparging-effected zone.

The VOC probes demonstrated conclusively the benefits of in situ on-line monitoring of contaminated groundwater. Processes such as groundwater movement during air sparging and rebound once air injection ceased were dramatically demonstrated using the VOC probes at location DP4, 13 m below ground (see Figure 6). However, based on conventional monitoring during periods when air injection was not occurring, it could be assumed that air sparging had no effect on groundwater at the 13 m below-ground location.

6 ADVANTAGES AND DISADVANTAGES OF VOLATILE ORGANIC COMPOUND PROBES

The field experiments highlighted a number of advantages of these new sampling techniques. VOC probes provided a robust long-term sampling process under aggressive field conditions. As the system only consists of a semipermeable polymer tube attached to access lines, no maintenance of the system was required.

Also, due to the semipermeable nature of the polymer, high-molecular-weight organic compounds (such as humic acids) and inorganic compounds are excluded from diffusing through the polymer, acting as an in situ sample clean-up step prior to sampling and analysis. As a result, little matrix contamination of the samples occurs so that samples can be analyzed without further clean-up.

Also, because of the simplicity and small size of the probes (20 cm), the system can be used for vertical-profile monitoring of groundwater contamination, especially where the depth to groundwater is greater than 8–9 m, where suction sampling of groundwater from simple minipiezometers fails.

Another advantage of this technique is the increase in sensitivity compared to conventional techniques. For this technique, in situ concentration of the organic compounds occurs within the polymer. Therefore, larger volumes of purged gas could be collected to increase sensitivity.

The disadvantage with the VOC probe and sensor unit is the sensor used to monitor VOC concentrations is not yet compound specific, therefore compositional data can only be obtained by collecting the purged gas from the VOC probe and analyzing the gas using conventional techniques such as GC/MS.

7 CONCLUSIONS

Recently developed VOC probes were assessed under long-term (between 1 and 3 months) field environmental conditions by comparing data obtained from these probes to data obtained by conventional groundwater sampling and analysis techniques. The in situ VOC probes provided a reliable monitoring technique under aggressive and dynamic field conditions. The probes provided data that were comparable to conventional techniques without the labor-intensive sample collection and processing associated with conventional techniques.

The field investigation also demonstrated the advantages of in situ on-line monitoring. Monitoring could be carried out during the air injection phase of the air-sparging trial, enabling groundwater movement and rebound to be demonstrated. Also, with this automated technique, a greater frequency of sampling was possible, enabling small-timescale trends to be observed accurately. Also, because of the length of the probes (20 cm), they can be used for monitoring of changes in vertical-profiles of groundwater contamination, especially where the depth to groundwater is greater than 8–9 m, and suction sampling of groundwater fails.

ACKNOWLEDGMENTS

The authors wish to thank John Rayner, Terry Power, David Briegel, Jack Smith and Dr Chris Barber from CSIRO Land and Water, and Dr Cynthia Joll from Curtin University for assistance with this work. Thanks also to Dr Mark Imisides for critically reviewing the manuscript. Financial support from the Cooperative Research Centre for Waste Management and Pollution Control and BP Oil is also gratefully acknowledged.

LIST OF SYMBOLS

K_{gw}	Gas–Water Partitioning Coefficient
K_{pw}	Polymer–Water Partitioning Coefficient

ABBREVIATIONS AND ACRONYMS

BTEX	Benzene, Toluene, Ethylbenzene, and Xylenes
CLS	Closed Loop Stripping
CSIRO	Commonwealth Scientific and Industrial Research Organization
ECD	Electron Capture Detection
FID	Flame Ionization Detection
FOCS	Fiber Optic Chemical Sensors
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HFSA	Hollow Fiber Stripping Analysis
MESI	Membrane Extraction with a Sorbent Interface
MIMS	Membrane Inlet Mass Spectrometry
MS	Mass Spectrometry
PCE	Tetrachloroethene
SPME	Solid-phase Microextraction
TCE	Trichloroethene
US EPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Electrochemical Methods • Pervaporation, Analytical • Sampling Considerations for Biomonitoring

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

REFERENCES

1. V.I. Pye, R. Patrick, J. Quarles, *Ground Water Contamination in the United States*, University of Pennsylvania Press, Philadelphia, 1983.
2. R.C. Borden, C.A. Gomez, M.T. Becker, 'Geochemical Indicators of Intrinsic Bioremediation', *Ground Water*, **33**, 180–189 (1995).
3. G.M. Cole, *Assessment and Remediation of Petroleum Contaminated Sites*, Lewis Publishers, Chelsea, Michigan, 1994.
4. E.J. Calabrese, C.E. Gilbert, H. Pastides, *Safe Drinking Water Act: Amendments, Regulations and Standards*, Lewis Publishers, Chelsea, Michigan, 1989.
5. NHMRC/ARMCANZ, *Australian Drinking Water Guidelines*, Canberra, ACT, National Health and Medical Research Council, Agriculture and Resource Management Council of Australia and New Zealand, 1996.
6. G.B. Davis, S. Appleyard, 'Organic Pollutants from Urban Developments Over the Unconfined Groundwater System in Perth WA', in *Groundwater and Land-use Planning*, Fremantle Western Australia, September 1996, eds. C. Barber, G.B. Davis, Centre for Groundwater Studies, Perth, 56–65, 1996.
7. C. Barber, G.B. Davis, 'Representative Sampling of Ground Water from Short-screened Boreholes', *Ground Water*, **25**, 581–587 (1987).
8. M.R. Scalf, F.J. McNabb, W.J. Dunlap, R.L. Cosby, J. Fryberger, *Manual of Groundwater Sampling Procedures*, NWWA/EPA Series, National Water Well Association, Worthington, Ohio, 1981.
9. M.J. Barcelona, J.A. Helfrich, E.E. Garske, 'Sample Tubing Effects on Groundwater Samples', *Anal. Chem.*, **57**, 460–464 (1985).
10. G.W. Reynolds, J.T. Hoff, R.W. Gillham, 'Sampling Bias Caused by Materials Used to Monitor Halocarbons in Groundwater', *Environ. Sci. Technol.*, **24**, 135–142 (1990).
11. L.V. Parker, T.A. Ranney, 'Sampling Trace-level Organics with Polymer Tubings: Part 1. Static Studies', *Ground Water Monitor. Remed.*, **17**(4), 115–124 (1997).
12. L.V. Parker, T.A. Ranney, 'Sampling Trace-level Organics with Polymer Tubings: Part 2. Dynamic Studies', *Ground Water Monitor. Remed.*, **18**(1), 148–155 (1998).
13. J.F. Pankow, L.M. Isabelle, J.P. Hewetson, J.A. Cherry, 'A Syringe and Cartridge Method for Down-hole Sampling of Trace Organics in Groundwater', *Ground Water*, **22**(3), 330–339 (1984).
14. B.M. Patterson, T.R. Power, C. Barber, 'Comparison of Two Integrated Methods for the Collection and Analysis of Volatile Organic Compounds in Groundwater', *Ground Water Monitor. Remed.*, **13**(3), 118–123 (1993).
15. F.P. Malanovich, 'Detecting Chloroorganics in Groundwater', *Environ. Sci. Technol.*, **20**, 441–442 (1986).
16. S.M. Klainer, J.D. Koutsandreas, A. Lawrence, 'Monitoring Groundwater and Soil Contamination by Remote Fibre Spectroscopy', *ASTM Spec. Tech. Publ.*, **963**, 370–380 (1988).
17. S.M. Barnard, D.R. Walt, 'Fibre-optic Organic Vapor Sensor', *Environ. Sci. Technol.*, **25**, 1301–1304 (1991).
18. S.A. Merschman, D.C. Tilotta, 'Fibre-optic Sensor for Aromatic Compounds in Water Based on Solid-phase Microextraction and Ultraviolet Evanescent Wave Absorption Spectroscopy', *Appl. Spectrosc.*, **52**(1), 106–111 (1998).
19. G.D. Johnson, 'Hexane-filled Dialysis Bags for Monitoring Organic Contaminants in Water', *Environ. Sci. Technol.*, **25**, 1897–1902 (1991).
20. C. Barber, D. Briegel, 'A Method for the In Situ Determination of Dissolved Methane in Groundwater in Shallow Aquifers', *J. Contam. Hydrol.*, **2**, 51–60 (1987).
21. B.M. Patterson, T.R. Power, G.B. Davis, C. Barber, 'An In Situ Device to Measure Oxygen in the Vadose Zone and in Groundwater', CSIRO Division of Water Resources Divisional Report 95/0 (Confidential), 1995.
22. G.B. Davis, C.D. Johnston, B.M. Patterson, C. Barber, M. Bennett, 'Estimation of Biodegradation Rates Using Respiration Tests During In Situ Bioremediation of Weathered Diesel NAPL', *Ground Water Monitor. Remed.*, **18**(2), 1–10 (1998).
23. A. Bianchi, M.S. Varney, J. Phillips, 'Modified Analytical Technique for the Determination of Trace Organics in Water Using Dynamic Headspace and Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **467**, 111–128 (1989).
24. T.N. Barnung, O. Grahl-Nielsen, 'Determination of Benzene and Naphthalene in Water by Purge and Trap Isolation and Capillary Column Chromatography', *J. Chromatogr.*, **446**, 271–278 (1989).

25. L.B. Westover, J.C. Tou, J.H. Mark, 'Novel Mass-spectrometric Sampling Device Hollow Fibre Probe', *Anal. Chem.*, **46**, 568–571 (1974).
26. B.J. Harland, P.J. Nicholson, 'Continuous Measurement of Volatile Organic Chemicals in Natural Waters', *Sci. Total Environ.*, **135**, 37–54 (1993).
27. M.A. LaPack, J.C. Tou, C.G. Enke, 'Mass Spectrometry for the Direct Trace Analysis of Volatile Organic Compounds in Air and Water', *Anal. Chem.*, **62**, 1265–1271 (1990).
28. A.J. Maden, M.J. Hayward, 'Sheet Material for Use as Membranes in Membrane Introduction Mass Spectrometry', *Anal. Chem.*, **68**, 1805–1811 (1996).
29. C.L. Arthur, L.M. Killam, S. Motlagh, M. Lim, D.W. Potter, J. Pawliszyn, 'Analysis of Substituted Benzene Compounds in Groundwater Using Solid-phase Microextraction', *Environ. Sci. Technol.*, **26**, 979–983 (1992).
30. C.L. Arthur, L.M. Killam, K.D. Buchholz, J. Pawliszyn, 'Automation and Optimization of Solid-phase Microextraction', *Anal. Chem.*, **64**, 1960–1966 (1992).
31. Z. Zhang, J. Pawliszyn, 'Headspace Solid-phase Microextraction', *Anal. Chem.*, **65**, 1843–1852 (1993).
32. S. Motlagh, J. Pawliszyn, 'On-line Monitoring of Flowing Samples Using Solid Phase Microextraction/Gas Chromatography', *Anal. Chim. Acta*, **284**, 265–273 (1993).
33. T. Górecki, J. Pawliszyn, 'Solid Phase Microextraction/ Isothermal GC for Rapid Analysis of Complex Organic Samples', *J. High Resolut. Chromatogr.*, **18**, 161–166 (1995).
34. M.J. Yang, M. Adams, J. Pawliszyn, 'Kinetic Model of Membrane Extraction with Sorbent Interface', *Anal. Chem*, **68**, 2782–2789 (1996).
35. A.K. Zander, P. Pingert, 'Membrane-based Extraction for Detection of Tastes and Odors in Water', *Water Res.*, **31**, 301–309 (1997).
36. J.R. Stetter, Z. Cao, 'Gas Sensor and Permeation Apparatus for the Determination of Chlorinated Hydrocarbons in Water', *Anal. Chem.*, **62**, 182–185 (1990).
37. J. Pawliszyn, 'Chemical Sensing Using Concentration Gradient Transients Produced During Diffusive Transport of Analytes', *Anal. Chem.*, **64**, 1552–1555 (1992).
38. C. Barber, B.M. Patterson, G.B. Davis, T.R. Power, 'Environmental Monitoring of Organic Compounds', US Patent 5 773 713, CRC-WMPC Ltd, 1998.
39. B.M. Patterson, 'Polymers for Monitoring and Remediation of Organic Compounds in Groundwater', PhD Thesis, University of Western Australia, 1998.
40. G.B. Davis, C.D. Johnston, J. Thierrin, T.R. Power, B.M. Patterson, 'Characterizing the Distribution of Dissolved and Residual NAPL Petroleum Hydrocarbons in Unconfined Aquifers to Effect Remediation', *J. Aust. Geol. Geophys.*, **14**(2/3), 243–248 (1993).
41. G.B. Davis, C. Barber, T.R. Power, J. Thierrin, B.M. Patterson, J.L. Rayner, W. Qinglong, 'The Variability and Intrinsic Remediation of a BTEX Plume in Anaerobic Sulfate-rich Groundwater', *J. Contam. Hydrol.*, **36**, 265–290 (1999).
42. C.D. Johnston, J.L. Rayner, B.M. Patterson, G.B. Davis, 'Volatilization and Biodegradation During Air Sparging of Dissolved BTEX-contaminated Groundwater', *J. Contam. Hydrol.*, **33**, 377–401 (1998).
43. J. Thierrin, G.B. Davis, C. Barber, B.M. Patterson, F. Pribac, T.R. Power, M. Lambert, 'Natural Degradation Rates of BTEX Compounds in a Sulfate Reducing Groundwater Environment', *Hydrolog. Sci. J.*, **38**(4), 309–322 (1993).

Waste Extraction Procedures

Barton P. Simmons
California Environmental Protection Agency, Berkeley, USA

1 Introduction 1
2 History 1
3 Environmental Applications 2
3.1 Hazardous Waste Classification 2
3.2 Waste Treatment Standards 2
3.3 Site Assessment 2
4 Toxicity Characteristic Leaching Procedure 2
4.1 Sample Preparation 3
4.2 Zero-headspace Extraction 3
4.3 Bottle Extraction 3
4.4 Analysis of Extracts 3
5 Synthetic Precipitation Leaching Procedure 3
6 Waste Extraction Test 3
6.1 Sample Preparation 3
6.2 Extraction Procedure 4
7 Other Extraction Procedures 4
7.1 Batch Extractions 4
7.2 Sequential Extractions 4
7.3 Multiple Extractions 5
8 Comparisons Among Extraction Tests 5
8.1 Municipal Landfill Disposal Scenario 5
8.2 Monofill Scenario 5
8.3 Industrial Co-disposal Scenario 5
9 Limitations of Extraction Procedures 5
9.1 Comparisons with Field Data 5
9.2 Oxoanion-forming Elements 6
9.3 Multiphase Wastes 6
Abbreviations and Acronyms 6
Related Articles 6
References 6

Waste extraction procedures (EPs) were developed to identify wastes which have a potential for groundwater or surface water contamination. Procedures are used for both organic and inorganic substances to comply with specific regulatory requirements, for example, to identify wastes which are subject to regulation as hazardous wastes. In the United States, the toxicity characteristic leaching procedure (TCLP) is the most widely used procedure. Volatile organics are extracted using a zero-headspace extractor, while semivolatile organics and inorganic substances are extracted using a bottle extractor. Because the typical batch EP does not measure a fundamental property of a waste, the result is highly dependent on steps in the protocol, particularly the choice of extraction solution. Although batch extraction tests like the TCLP are the most common procedures, other procedures using multiple pHs and multiple liquid : solid ratios have been developed to overcome the limitations of batch tests.

1 INTRODUCTION

EPs have historically been used for estimating the availability of inorganic and organic species from soil. The growth of environmental regulations created the need for waste EPs which could identify ''hazardous'' wastes by their potential to contaminate surface water or groundwater. The primary application of waste EPs has been to predict concentrations in leachate if the waste were disposed of in a municipal landfill, but applications have been made to other scenarios. Comparisons among the tests have shown the need for improved procedures and improved modeling of the leaching process.

2 HISTORY

Major environmental laws of the 1970s, principally the US Resource Conservation and Recovery Act (RCRA), created a system to manage hazardous wastes from ''cradle to grave.'' Central to the US waste management system is a set of criteria for the identification of hazardous wastes.[1] The criteria typically used for hazardous waste identification are ignitability, corrosivity, reactivity, and toxicity. Of these, toxicity is technically the most difficult to define, and the most contentious, since it significantly defines the wastes which are regulated as hazardous wastes. The United States Environmental Protection Agency (USEPA) has depended on a combination of listing hazardous wastes and ''characteristics'' for hazardous waste identification. Since historically one of the major environmental impacts of mismanaged waste was groundwater contamination, the USEPA has used waste EPs to identify wastes which are regulated because they may pose a threat to groundwater. The initial test, the EP, has been replaced by the TCLP.[2] California has developed a similar extraction test, the waste extraction test (WET), for a similar purpose.[3] These tests were all developed with a scenario of disposal in a municipal landfill, where wastes could be exposed to municipal landfill leachate.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

3 ENVIRONMENTAL APPLICATIONS

Extraction tests have been applied to a variety of environmental problems; examples of applications are shown in Table 1. Hazardous waste classification is the most important, since it has a major impact on how wastes will be treated, transported, and disposed of. The major variables which affect the performance of extraction tests are listed in Table 2. Of these, the choice of extraction solution is probably the most important, since inorganic ions may be complexed or chelated by agents in the solution.

3.1 Hazardous Waste Classification

The USEPA criteria for hazardous waste identification include toxicity characteristic criteria using the TCLP. If the TCLP extract of a waste exceeds an applicable toxicity characteristic limit, the waste is classified as hazardous, and is subject to regulation.

California and some (non-US) countries have adopted similar waste classification procedures, based on other extraction test systems. Under RCRA, a state system must generally be at least as stringent as the federal system. As a result, most states, with the exception of California, have adopted the TCLP for waste identification. This article compares the TCLP and WET systems.

Table 1 Objectives of extraction tests

Objective	Description
Hazardous waste classification	Extract with a specified protocol and compare with hazardous waste criteria
Evaluation of waste treatments	Compare extractable concentrations with treatment standards
Fate and transport modelling	Estimate partition coefficients or other properties
Risk assessment	Use extractable concentrations to estimate potential exposures to humans, plants or animals

Table 2 Primary extraction variables

Extraction test variable	Example: TCLP
Sample preparation	Reduce size to <9.5 mm
Extraction solution	Acetate buffer
Liquid-to-solid ratio	20 : 1
Extraction time	18 h
Filtration of extraction solution	Filtration through 0.6–0.8-μm sieve
pH	4.9 or 2.9, depending on alkalinity of waste

3.2 Waste Treatment Standards

The USEPA has also adopted the TCLP for determining compliance with waste-specific treatment standards, primarily for inorganic constituents. These standards set extractable limits for treated wastes prior to disposal in a landfill.[4]

3.3 Site Assessment

EPs have been used as part of the risk assessment process for site mitigation. In this process, an extraction test is used to estimate the concentration of regulated substance which may be extracted from contaminated soil or waste. For example, the synthetic precipitation leaching procedure (SPLP) has been used to estimate how much would be extracted by rainfall if contaminated soil or waste were left in place. This result is compared with either generic risk-based criteria or site-specific criteria based on multipathway risk assessment. Alternatively, site groundwater or site leachate can be used to estimate site-specific extraction. Several states have adopted the SPLP to decide whether contaminated soil can be left in place.

4 TOXICITY CHARACTERISTIC LEACHING PROCEDURE

The TCLP was developed by the USEPA to identify wastes which could contaminate groundwater if the waste were disposed of in a municipal landfill. Accordingly, the extraction solution, an acetate buffer, was originally designed to mimic the effects of municipal solid waste leachate, which typically contains organic acids from the decomposition of garbage and other municipal wastes. Because some alkaline wastes can effectively neutralize the acetate buffer, an acidic acetate buffer, pH 2.9, is used in an attempt to compensate for the effect of alkaline waste. The procedure is presented as Method 1311 in SW-846, *Test Methods for Evaluating Solid Waste*.[2] Since the EP is used for regulatory compliance, and the result is highly dependent on the actual procedure, regulatory agencies do not allow deviation from the published procedure, and it is not considered a candidate for performance-based measurement systems. A summary of the procedure is provided in Table 3; USEPA Method 1311 should be used for procedural details.

For regulatory purposes, the TCLP is not used if total analysis indicates that regulatory limits could not be exceeded. With solid samples, for example, the TCLP is normally not required if total results are less than 20 times the corresponding regulatory limit.

Table 3 Steps in the TCLP

Subsample the waste
Determine the percentage solids
If solids are <0.5%, filter with a 0.6–0.8-μm glass fiber filter, and discard solids
If solids are >0.5%, filter as above
If necessary, mill solids to <9.5 mm diameter
Extract solids
If initial filtrate and extract are compatible, combine
Analyze liquids

4.1 Sample Preparation

For liquid wastes containing less than 0.5% solids, the initial filtrate is defined as the TCLP extract. For solid samples, or liquid/solid mixtures, the steps given in sections 4.2 to 4.4 should be followed.

4.2 Zero-headspace Extraction

If the extract is to be analyzed for volatile organic compounds (VOCs), a zero-headspace extractor is used to minimize the losses of volatile compounds.

4.3 Bottle Extraction

Analytes which do not require the zero-headspace extractor (elements, semivolatile organics, and nonvolatile organics) are extracted with a bottle extractor using a rotating extraction device.

4.4 Analysis of Extracts

If the initial filtrate and the final filtrate are miscible, they are combined and analyzed. If they are not compatible, they are analyzed separately and the results are mathematically combined to give a volume-weighted concentration. The analysis of the liquid TCLP extracts is generally done using the same techniques as for other liquid samples. Oily waste samples may produce TCLP extracts composed partially of oil, and will require digestion or extraction that is appropriate for oil-containing samples.

For regulatory purposes, the final concentrations are compared with the corresponding regulatory limits. For waste classification in general, the variability in results, usually dependent primarily on variability in the waste, is addressed by comparing the upper 90% confidence limit of the mean concentration with the regulatory limit.[5] Because the confidence limit of the mean decreases with increasing sample size, the larger the number of TCLP results, the less ambiguous the classification.

5 SYNTHETIC PRECIPITATION LEACHING PROCEDURE

The SPLP was designed to simulate the extraction of substances from soil or waste by percolating rainfall. For soils from east of the Mississippi River, the SPLP uses a pH 4.2 solution to mimic the effect of ''acid rain.'' Unlike the TCLP, it does not have a strict regulatory application, and it is primarily used in decisions regarding the remediation of contaminated sites. Typically, the SPLP results are used in a fate and transport model to predict potential impacts on groundwater or surface water, using either screening models or site-specific models.

The SPLP is very similar to the TCLP, with the exception of the extraction solution. Table 4 lists the solutions used for various applications.

6 WASTE EXTRACTION TEST

The WET was designed for the same application as the TCLP; namely, to simulate the extraction of regulated substances by municipal solid waste leachate. The most significant difference between the WET and the TCLP is the extraction solution. The citrate buffer used in the WET is much more aggressive than the acetate solution used in the TCLP for some elements. Citrate is known to chelate some elements with much greater affinity than the acetate complexing agent. Citrate is naturally occurring, and has been found to be an important chelating agent for toxic elements and transuranic elements.[6]

6.1 Sample Preparation

The WET has similar sample preparation procedures to the TCLP and SPLP; however, for solid samples, the samples are milled to less than 10-mm diameter

Table 4 SPLP (for details, see Method 1312, ref. 2)

Extraction solution	Composition	pH	Application
Extraction fluid #1	Dilute sulfuric and nitric acids (60/40)	4.2	Soils from east of the Mississippi River, and all wastes and wastewaters
Extraction fluid #2	Dilute sulfuric and nitric acids (60/40)	5	Soils from west of the Mississippi River
Extraction fluid #3	Reagent water	Not specified	Cyanide and volatiles

prior to extraction. Elemental metals, unless powdered, finely divided, or friable, are not regulated. Fragments of elemental metal are therefore removed during the sample preparation step.

6.2 Extraction Procedure

The WET EP is similar to the TCLP and SPLP, although more flexibility is allowed in the choice of extraction apparatus. Unlike the TCLP, the WET does not include a zero-headspace extractor, and has no precautions for the loss of VOCs during the filtration and extraction steps.

7 OTHER EXTRACTION PROCEDURES

A great variety of EPs have been developed by government environmental agencies and other standard-setting organizations. Generally, they may be classified as (1) batch EPs, like the TCLP; (2) sequential procedures, in which the extraction solution is replaced with fresh solution at a specified interval; or (3) multiple EPs which are designed to estimate fundamental leaching parameters. Procedures may or may not use agitation. In addition to extraction tests, leaching tests can be done with a continuous supply of extraction solution. Some extraction tests are done with leachate from the landfill in which it is proposed the waste be placed.

7.1 Batch Extractions

Table 5 lists examples of batch extraction tests which have been developed by governments for regulatory compliance. Batch tests have the advantages of reduced cost and time, but suffer from the limitation that the test conditions may not simulate field conditions.

7.2 Sequential Extractions

Sequential extractions involve periodically replacing a portion of the extraction solution or replacing the solids with a fresh portion of waste. The intent is to simulate the longer-term release of constituents or to simulate equilibrium conditions. Because of the time involved in sequential extractions, they have not been used widely for regulatory applications.

Table 5 Other EPs[7]

Extraction test	Sample size	Extraction solution	Liquid : solid ratio	Agitation	Duration
EP Toxicity Test (USEPA Method 1310)	100 g	Deionized water + 0.5 N acetic acid (max. 2.0 mequiv g^{-1})	20 : 1	Unspecified; continuous	24–28 h
LEP MOE (Ontario)	50 g	Distilled water, acetic acid (2.0 mequiv g^{-1})	20 : 1	End-over-end (10 rpm)	24 h
French Leach Test, AFNOR (France)	100 g	Demineralized water	10 : 1	Roller or shaker	16 h
EE Environment Canada	Variable; fill 90% of bottle	Distilled water	4 : 1	National Bureau of Standards extractor	7 days
ASTM D3987	700 g	Distilled water (ASTM Type IV)	4 : 1	Shaking	48 h
MBLP Environment Canada	>1 g	High purity water, silicate water, brine, or repository water	10 : 1	Rolling and rocking	Variable: 28 days to several years
Official Norm Mexicana NOM-CRP-002-ECOL/1993 Mexico	100 g	Acetate buffer, pH 4.93 or acetate buffer, pH 2.88 (for alkaline wastes)	20 : 1	Rotating extractor	18 h
Dutch Availability Test, NEN 7341[8]	16 g	Two steps: demineralized water plus nitric acid at pH 7, then pH 4	50 : 1	Magnetic stirrer	3 h + 3 h

LEP, leachate extraction procedure; EE, equilibrium extraction; ASTM, American Society for Testing and Materials; MBLP, multiple batch leaching procedure.

7.3 Multiple Extractions

Some multiple procedures have been designed to estimate fundamental leaching parameters, such as availability, solubility, and diffusion.[9] Availability is the maximum quantity of a constituent that can be released into solution under aggressive conditions, such as those in the WET. Solubility is usually measured over a range to estimate the solubility as a function of pH for a specific waste. Diffusion is important for monolithic wastes, such as wastes which have been stabilized with cement, asphalt, or other material, to estimate the rate of release of a constituent.

8 COMPARISONS AMONG EXTRACTION TESTS

As shown in Table 5, extraction tests can be categorized by extraction solution:

1. deionized, distilled, or other purified water,
2. acetic acid or acetate buffer,
3. citrate,
4. other solutions.

Based on the stability of chemical complex formation, the general expected order of aggressiveness of extraction solutions for most metals is: citrate > acetate > purified water. In practice, this is generally the case; citrate tests (e.g. WET) extract more than acetate-based tests (e.g. TCLP), which in turn extract more than purified-water-based systems. Since most organic analytes are neutrally charged, they are relatively insensitive to the pH of the EP. The extraction of organic analytes is primarily determined by their water solubility, but is also affected by agitation, temperature, and the liquid : solid ratio.

8.1 Municipal Landfill Disposal Scenario

The scenario of waste disposal in a municipal landfill is the basis of the TCLP and the WET. These tests assume exposure to landfill leachate, typically assuming a pH of about 5. The complexing ability of the leachate is modeled by acetate in the TCLP and citrate in the WET.

The actual field conditions at municipal solid-waste facilities can vary considerably. Major factors which affect leaching are as follows.

1. The waste streams: municipal solid-waste facilities may accept some quantities of industrial wastes, as well as biosolids from wastewater treatment. Recycling programs can affect the composition of solid waste which is delivered to the facility.
2. Precipitation: the quantity and composition of leachate is driven by the infiltration of precipitation. Landfills located in semi-arid or arid areas may be ''tombs'' which generate little, if any, leachate during the dry season.
3. Landfill design: leachate may be recovered and recycled within the landfill or disposed of in a wastewater treatment facility. The design and performance of liners may also affect the quality of leachate.

8.2 Monofill Scenario

Some industrial wastes are disposed of in a dedicated landfill, or monofill. Monofill leachate may vary considerably from municipal solid-waste leachate, particularly in pH and absence of organic acids. Because of the variety of wastes which could be placed in a monofill, it is not feasible to design a generic extraction test for all industrial wastes. The SPLP has been used to simulate extraction with percolating rainfall, but has not been used widely for waste classification.

8.3 Industrial Co-disposal Scenario

Co-disposal of waste with treated or untreated industrial waste can expose wastes to landfill conditions determined by co-disposed waste. For example, lime-treated waste can generate leachate with a pH approaching 13, which can release arsenic and other elements in significant amounts. If a waste has been de-listed as a result of treatment, the waste can go to a landfill which is not designed for hazardous wastes.

9 LIMITATIONS OF EXTRACTION PROCEDURES

The identified limitations of published procedures have been revealed by comparison with field data and by comparison with column lysimeter data. The limitations are both element specific and waste specific.

9.1 Comparisons with Field Data

Because batch extraction tests are typically conducted over 18–48 h and generate one datum point, they are not expected to quantitatively predict the long-term leaching of regulated substances. The typical regulatory application of a test is qualitative, that is, whether a waste is classified as hazardous or not. Some extraction tests have not accurately identified wastes with potential for groundwater contamination. For example, the TCLP has failed to identify wastes which have contaminated groundwater with arsenic.[10]

9.2 Oxoanion-forming Elements

The TCLP and WET were designed using acetate and citrate, respectively, to simulate the complexing of regulated elements by municipal solid-waste leachate. Several elements, e.g. antimony, arsenic, molybdenum, selenium, and vanadium, combine with oxygen under field conditions to create oxoanions. Oxoanions do not complex like cations, and as a result may not be extracted by the batch extraction tests.[11] Of these oxoanions, arsenic may the most important because of its occurrence in wastes and its risk assessment as a human carcinogen.

9.3 Multiphase Wastes

Multiphase industrial wastes pose a challenge to analytical procedures in general, and extraction tests are no exception. Modeling the extraction of multiphase wastes has not been incorporated into the extraction process. As an example, the TCLP uses an initial filtration to separate the solids from liquids. A sample composed of an aqueous phase and a nonaqueous phase may be separated arbitrarily into an initial filtrate, which may be one- or two-phase, and a ''solid'' which may in fact contain nonaqueous liquid which did not pass through the filter. One consequence is variability in the final result and incompatible liquids in the final filtrate. The development of extraction tests for multiphase liquids is needed, along with models for dense nonaqueous phase liquids, in order to provide reliable estimates of the movement of contaminants from nonaqueous phases into groundwater.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
EE	Equilibrium Extraction
EP	Extraction Procedure
LEP	Leachate Extraction Procedure
MBLP	Multiple Batch Leaching Procedure
RCRA	Resource Conservation and Recovery Act
SPLP	Synthetic Precipitation Leaching Procedure
TCLP	Toxicity Characteristic Leaching Procedure
USEPA	United States Environmental Protection Agency
VOCs	Volatile Organic Compounds
WET	Waste Extraction Test

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices

REFERENCES

1. Title 40 Code of Federal Regulations, Part 261 ''Identification and Listing of Hazardous Waste'', US Government Printing Office, Washington, 1980.
2. US Environmental Protection Agency, 'Test Methods for Evaluating Solid Waste, SW-846', 3rd edition, Office of Solid Waste, Washington, 1996.
3. 'California Code of Regulations', Title 22, section 66261.24, Sacramento, 1985.
4. Title 40 Code of Federal Regulations, Part 268 ''Land Disposal Restrictions'', US Government Printing Office, Washington, 1986.
5. US Environmental Protection Agency, 'Test Methods for Evaluating Solid Waste, SW-846', 3rd edition, Office of Solid Waste, Chapter 9, Washington, 1996.
6. A.J. Francis, C.J. Dodge, J.B. Billow, 'Biodegradation of Metal Citrate Complexes and Implications for Toxic-metal Mobility', *Nature*, **356**, 140–142 (1992).
7. 'Compendium of Waste Leaching Tests', Environment Canada, Report EPS 3/HA/7, May 1990.
8. A.-M. Fallman, 'Performance and Design of the Availability Test for Measurement of Potentially Leachable Amounts from Waste Materials', *Environ. Sci. Technol.*, **31**(3), 735–744 (1997).
9. D.S. Kosson, H.A. van der Sloot, T.T. Eighmy, 'An Approach for Estimation of Contaminant Release During Utilization and Disposal of Municipal Waste Combustion Residues', *J. Haz. Mat.*, **47**, 43–75 (1996).
10. 'Columbia Falls Aluminum v. US Environmental Protection Agency', United States Court of Appeals for the District of Columbia, Washington, DC, WL 151176, DC Cir., 1998.
11. K. Hooper, M. Iskander, G. Silvia, F. Hussein, J. Hsu, M. Deguzman, Z. Odion, Z. Ilejay, F. Sy, M. Petreas, B. Simmons, 'Toxicity Characteristic Leaching Procedure Fails to Extract Oxoanion-forming Elements that are Extracted by Municipal Solid Waste Leachates', *Environ. Sci. Technol.*, **32**(23), 3825–3830 (1998).

Water Analysis: Organic Carbon Determinations

Kenneth Mopper
Washington State University, Pullman, USA

Jianguo Qian
MQ Scientific Inc., Pullman, USA

1 Introduction 1
2 Operational Definitions and Artifacts 2
3 Sampling Precautions 3
4 Measurement of Organic Carbon 3
4.1 Wet Oxidation Methods 3
4.2 High-temperature Combustion Methods 5
4.3 Problems Associated with High-temperature Combustion Methods and Applications to Seawater 5
Acknowledgment 8
Abbreviations and Acronyms 8
Related Articles 8
References 8

The determination of total organic carbon and dissolved organic carbon (TOC and DOC) is one of the most important parameters in water quality and environmental analysis today. TOC is composed of particulate organic carbon, DOC and volatile organic carbon, (POC, DOC and VOC). In most waters, DOC is the dominant component of TOC. POC consists of living and nonliving organic particles and can occasionally become quantitatively important. POC, DOC and VOC are operationally defined by the methods used to separate them. POC and DOC are usually separated from each other by filtration, which can introduce errors due to contamination and filter clogging. Other major errors in TOC and DOC analyses are related to contamination and losses during sampling, sample storage and sample manipulation. In particular, removal of inorganic carbon by either acidification with sparging or by acidification with evaporation to dryness results in variable loss of VOC.

Commonly used TOC and DOC methods fall into two approaches: wet oxidation and high-temperature combustion (HTC). In both approaches, organic matter is oxidized to CO_2, which is then usually determined by nondispersive infrared (NDIR) absorbance. Two wet oxidation methods are in common use: wet chemical oxidation (WCO) and ultraviolet oxidation (UVO). In WCO, a strong chemical oxidant, usually persulfate, is added to the aqueous sample and the digestion is usually carried out batchwise in a reactor at an elevated temperature. In contrast, UVO digestion is commonly carried out in a quartz coil surrounding a mercury vapor lamp. Thus, UVO methods can be readily automated using flow-injection systems. Two types of HTC methods are currently in use: dry combustion and direct aqueous injection. In dry combustion methods, the sample is acidified, dried and the residue is combusted at high temperature in a sealed tube, usually in the presence of a catalyst. This method is advantageous for the analysis of large samples. In direct aqueous injection HTC methods, samples are injected into a high-temperature column (600–900 °C), which usually contains a catalyst. In recent years, the latter method has become widely used for the analysis of seawater and other waters because it appears to have a greater oxidation efficiency and precision than most of the other methods, and it can be readily automated. However, automated HTC instruments have problems in the analysis of saline samples, in particular, salt deposits in the sample injection system, memory (or carry-over) effects, and system blank evaluation. Recently, improvements in the HTC injection system and column design have addressed these problems.

1 INTRODUCTION

The determination of TOC and DOC is an important and routine measurement in environmental, industrial and regulatory studies. It is used for estimating chemical oxygen demand (COD) resulting from anthropogenic activities, including dispersal of municipal sewage effluents, industrial wastes, agricultural runoff, and eutrophication of water bodies. TOC and DOC analyses are also used for monitoring organics in process water (i.e. cooling water, condensate and boiler-feed waters in power plants), in high-purity water generating systems, in drinking water, in desalination plants, in contaminated groundwater near waste-disposal sites, in sterilization water used in biotechnology and pharmaceutical processes, in semiconductor reclaim water, in raw and processed effluent water from food, paper and textile plants, and in a variety of regulatory and remediation operations. TOC and DOC are also important components of the global carbon cycle. Accurate measurements of these carbon pools are critical in oceanic and climatic models, as well as in evaluating food-web dynamics. For many of the above applications, real-time or near real-time evaluation of the organic carbon (OC) content of the water is often essential. Owing

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

to the heterogeneity and structural complexity of natural organic matter, its quantification has mainly relied on the determination of the OC content (DOC or TOC, if unfiltered).

Carbon constitutes about 50% of most organic molecules and it is usually determined by oxidation to CO_2 by various techniques. The techniques that have been employed to determine TOC or DOC in fresh and seawaters, include WCO, photo-oxidation, dry combustion, and direct injection-HTC. Although the oxidation of organic matter to CO_2 would seem to be straightforward, all current techniques have problems.[1] For example, there are differences in the oxidation efficiencies (i.e. the ability to oxidize completely OC to CO_2) as well as the ease of automation among these different techniques. There is also a lack of certified standards and accepted recalcitrant reference compounds for interlaboratory comparisons and for assessing oxidation efficiencies of the techniques. For most techniques, it is difficult to assess the blank, both instrumental and method. Improper assessment of the blanks, together with blank variability, adversely affect the precision and accuracy of TOC and DOC analyses. Poor reproducibility and poor accuracy can also arise from contamination during sampling, sample manipulations (e.g. transfers, filtration and acidification), incomplete sparging of inorganic carbon and sample storage.[2] In this article, both sampling and instrumental aspects of TOC and DOC analysis are discussed.

2 OPERATIONAL DEFINITIONS AND ARTIFACTS

The commonly used units for TOC and DOC are mg L^{-1}, ppm, and μM, where 1 mg L^{-1} is about 1 ppm (w/w), and 83 μM. In this article, the ppm and μM notations are used.

The total carbon content of a sample is the sum of all forms of dissolved inorganic carbon (DIC) and TOC. DIC is composed of the dissolved species within the carbonate system, i.e. carbon dioxide, carbonic acid, bicarbonate ion and carbonate ion. The relative proportion of these species depends on pH and on biological and physical processes. TOC is composed of POC, DOC and VOC matter. In most natural waters, DOC is the dominant component of TOC, usually 50 to >95%. POC consists of living and nonliving particles and can become quantitatively important (e.g. 10% or more of TOC) in surface waters during periods of high biological activity, in lakes and rivers during periods of high land runoff, and in sewerage and industrial effluents.

POC, DOC and VOC are operational definitions since their relative abundance depends on the physical methods used to separate them. POC and DOC are usually separated from each other by filtration. However, filtration can introduce major errors. The pore size range most frequently used is 0.2–1.0 μm. Consequently, varying amounts of submicron detritus, microalgae, bacteria, viruses and colloidal substances pass through the filter and are measured as DOC. For samples containing a high particle load, the effective pore size may dramatically decrease during filtration as the pores become progressively clogged. Clogging results in a change in the size distribution of particles retained and passed by the filter and also in an increase in back pressure with time. Too large an overpressure (or underpressure) during filtration will lead to disruption of particles and fragile organisms with the concomitant leaching of their organic-rich dissolved contents into the filtrate. Other sources of error from filtration include adsorption of DOC onto the filter pad and contamination from the air and the filter itself, if inadequately cleaned.[1,3,4] To avoid filtration problems, the filtration step is sometimes omitted and only TOC is analyzed. As POC often represents <5%, and usually <1%, of the TOC in many types of water samples (e.g. seawater, lake water, drinking water, cooling water, high-purity water, etc.), TOC is often equal to DOC within the analytical precision.

Two approaches are commonly used to separate TOC from DIC.[5] In one approach, TOC is calculated as the difference between total carbon and DIC, the latter being determined as CO_2 released by the sample by either acidification followed by vigorous bubbling with organic- and CO_2-free gas (sparging) or by acidification and gentle heating. This approach for estimating TOC is best suited to freshwater samples where the carbonate concentration is low enough (i.e. similar to the TOC concentration) to ensure good accuracy and precision. However, in the case of seawater and alkaline lake water, DIC is present at about 15–50 times higher concentration than TOC. Thus, for these samples, determination of TOC by difference usually leads to a large analytical error. In the other approach, DIC is removed as above and the remaining TOC in the aqueous sample is determined by oxidation to CO_2, i.e. TOC is determined ''directly'' as opposed to ''by difference''. For both approaches, complete removal of DIC is critical to obtain accurate and precise TOC values.[6]

Removal of DIC by either acidification with sparging or, in the case of dry combustion techniques, by acidification with evaporation to dryness (in an oven, desiccator, or freeze-drier), results in loss of VOC. The extent of this loss depends on the intensity of the DIC elimination step, with the least loss (usually <1% of TOC) with quick sparging and the most loss (about 5–15% of TOC) by oven drying.[1] Thus, the VOC fraction is operationally defined by the analytical procedure used to strip or remove it from the sample, as opposed to that fraction

which is volatile under ambient environmental conditions (wind and turbulence). A few commercial TOC systems are designed to trap and analyze the VOC fraction.

Because some (usually unknown) fraction of VOC is lost during the DIC removal step, TOC is also operationally defined, and is equal to DOC plus POC plus residual VOC. However, since the VOC fraction lost is usually minor (<1% of TOC), the error introduced by this loss is also usually minor. For humic-rich waters, an additional problem resulting from the acidification step is precipitation of humic acid, which may result in an underestimation of the DOC content.[5,7]

3 SAMPLING PRECAUTIONS

Probably the largest source of error in TOC and DOC analyses is related to sampling and sample storage.[1,2,8] In addition to the errors discussed in the previous section (i.e. from filtration and DIC removal), contamination during sampling and analysis is a major concern.[9] Contamination can arise from incompletely cleaned samplers and sample bottles or from leaching of organic substances (e.g. plasticizers) from new plastic or PVC (polyvinyl chloride) sampling and storage bottles.[4,10] In sampling of water bodies, samplers are often used that, by design, are initially open and then are closed at the desired sampling depth. As these samplers are lowered, they pass through the organic-rich surface microlayer and this material can become strongly adsorbed onto the inside walls of the sampler and contaminate the sample. Thus, samplers that are initially closed, then opened at depth are preferred.

After the sample is retrieved, subsequent sample handling steps are also critical, especially for low TOC samples. All transfers should be performed in an organic-free atmosphere or at least upwind from any major airborne contamination sources, in particular the operator and fumes. The latter is especially a problem when sampling from research vessels which are notoriously rich in exhaust and galley fumes. After collecting the uncontaminated sample, it must be stabilized as soon as possible, since biological processes can cause large changes in the DOC concentration even within the first hour of collection.[11] Stabilization is usually performed by quick freezing followed by storage at $-20\,^{\circ}C$, acidification (pH 2–2.5) followed by refrigeration at $<4\,^{\circ}C$, or addition of a strong bacteriostatic agent (e.g. $HgCl_2$). Of course, any chemical addition made to the samples increases the risk of contamination. Thus, considerable care is required to obtain reliable DOC and POC results. The specific precautions and guidelines needed to overcome sampling and preservation artifacts vary with the sampling protocol used and the type of samples being studied. Detailed studies of some of these precautions can be found in the literature.[2,8,11–15] Some specific recommendations and procedures are given in Table 1.

4 MEASUREMENT OF ORGANIC CARBON

Attempts have been made to use the innate ultraviolet (UV) absorption (250–400 nm) of organic matter to estimate DOC, but this approach appears to be limited to specific geographic areas or water bodies where the organic composition of the sampled water does not change appreciably;[16,17] that is, the fraction of non-UV absorbing DOC remains approximately constant.[18,19] Another approach to measuring the organic content of water involves estimating its oxidizability, i.e. COD and total oxygen demand (TOD). However, these methods are mainly useful for organic-rich wastewaters and freshwaters, and are not suitable for uncontaminated drinking water and most natural waters.[20]

The most commonly used TOC and DOC methods fall into two general approaches: wet oxidation and HTC. In both approaches, the organic matter is oxidized to CO_2 which is then determined by any number of techniques: ion chromatography-conductometry, coulometry, titrimetrically using a pH-sensitive indicator, potentiometrically using a CO_2 electrode, gravimetrically, pyrolysis-GC/MS (gas chromatography/mass spectrometry), inductively coupled plasma atomic emission spectrometry (ICP/AES), by conversion to CH_4 followed by flame ionization detection (FID), and by NDIR absorbance. The last technique is most widely used because it is sensitive and interferences can be readily eliminated. Details of CO_2 detection techniques are not addressed in this article.

4.1 Wet Oxidation Methods

Two wet oxidation methods are in common use: WCO and UVO. In WCO, a strong chemical oxidant is added to the aqueous sample after the DIC had been removed by acidification and sparging. Oxidants used include potassium peroxide, dichromate in sulfuric acid, silver-catalyzed dichromate (Walkey–Black method), perchlorate, silver-catalyzed persulfate (also called peroxydisulfate), platinum-catalyzed persulfate, and potassium persulfate. Persulfate-based methods are the most commonly used. The oxidative digestion of OC is usually performed batchwise using either a sealed glass ampoule or a reusable reaction chamber. Reaction times and temperatures are usually in the range of 0.1–4 h and $90–160\,^{\circ}C$, depending on the nature of the sample and oxidant. The CO_2 formed is measured in several ways (see above), but most frequently by NDIR.

Table 1 Recommended sample handling and treatment procedures prior to TOC and DOC analysis

Parameter	Recommendation
1. Sampling	Sampling should avoid contamination, for example, from contaminated air from exhaust stacks, automobile fumes, etc. Always collect and handle samples upwind from possible contamination sources. The precleaned sample container and cap should be rinsed at least three times with sample before filling.
2. Sample container type	Use precleaned borosilica glass bottles, or Teflon™-coated plastic bottles to collect and store samples. Some types of plastic containers (e.g. high-density polyethylene) may also be used if carefully cleaned.[15]
3. Cleaning glass sample containers	Muffle glassware for about 8 h at 450 °C. When cool, store in a sealed, clean container. Prior to use, rinse inside with 10% HF using a squirt bottle (and using gloves and goggles). Thoroughly flush out the HF with fresh low-carbon deionized water (i.e. direct from deionized water source). Fill the glassware to rim with fresh low-carbon deionized water and cover with caps or aluminum foil. The glassware can be stored this way at least 2 days. Decant the deionized water just before use.
4. Cleaning Teflon™-coated plastic sample containers	Wash inside of the Teflon™-coated plastic bottle with acetonitrile. Thoroughly flush out the acetonitrile with the fresh low-carbon deionized water and dry at 40–60 °C in a clean oven. (This is a critical step as any trace of acetonitrile left will contaminate samples and result in an elevated DOC concentration). Fill the washed plastic bottles to rim with fresh low-carbon deionized water and cap them, decanting the deionized water just before use.
5. Sample storage	Samples in glass bottles should be acidified and stored at 2–4 °C. Samples in plastic bottles should be stored at −20 °C or colder. For the latter, acidification is not necessary during storage, but highly recommended immediately after thawing.
6. Sample acidification	Use low-carbon or carbon-free 20–40% H_3PO_4 (~85% H_3PO_4 diluted by low-carbon deionized water) to acidify samples to a pH between 2–2.5.
7. Cleaning autosampler vials	Glass vials are cleaned as in step 3 above. If Teflon™ or Teflon™-backed silicon rubber septa are used, they should be cleaned as in step 4 above. The cleaned septa should be stored in a sealed clean container.
8. Sample sparging	Sparging is needed to completely remove DIC from the acidified sample. This step is critical for accurate determination of TOC and DOC. The sparging gas should be carbon free, such as ''zero'' air, ultrahigh purity oxygen or nitrogen etc. The sparging time depends on gas flow rate, bubble size and sample volume. For example, for a 100 mL sample and a gas flow rate of 100–150 mL min^{-1}, the sparging time may be 8–10 min. For an automated HTC analyzer (such as the MQ1001 TOC analyzer, MQ Scientific), sparging is performed with the carrier gas automatically just before the sample is injected into the combustion column.

The advantages of WCO are that it is analytically simple and relatively inexpensive. However, WCO has some problems. For example, manual WCO methods are relatively labor-intensive and time-consuming. In addition, WCO has been reported to have a lower oxidation efficiency (usually 5–15% lower) compared to other methods, for example photooxidation and HTC.[21–24] Several possibilities have been proposed to explain this apparent incomplete oxidation. For example, because samples are processed in batches, the reagent concentration and strength decrease during the course of the digestion as the easily oxidized substances are converted to CO_2. Consequently, the oxidative strength of the residual reagent may become too low to completely oxidize the remaining, more resistant carbon fraction in the sample.[25] In the case of seawater and estuarine water, chloride interferes strongly with persulfate methods, greatly lowering the oxidative strength of the reagent.[26] Thus, in either case, very high levels of oxidant must be added in order to ensure maximal oxidative efficiency and maximal CO_2 production.[26] But large reagent additions can cause high and variable blanks due to contamination, which in turn leads to lowered precision and higher detection limits, especially for manual WCO methods. One study indicated that at least some of the apparent discrepancy in oxidation efficiency for WCO methods is probably due to improper assessment of the blanks, especially for earlier studies.[23] Another possible cause of the lower oxidation efficiency of WCO methods is the production of volatile reduced organic compounds during the digestion step. These compounds would be lost to the gas phase and go undetected by most CO_2 detection techniques.[27]

Many of the problems associated with WCO are overcome by using UVO methods. Most of these methods rely on flow-injection analysis, as opposed to batch digestion used in WCO. Acidified and sparged samples are injected into an aqueous carrier stream, which is usually buffered and contains an oxidant (e.g. persulfate) and sometimes a catalyst (e.g. titanium dioxide, Hg(II),

or platinum) that greatly enhances the UV oxidation rate and efficiency. The digestion is usually carried out in a quartz coil surrounding a medium-pressure mercury vapor lamp. Since oxidative free radicals are continuously produced by photochemical reactions in the UV-irradiated reaction coil, loss of oxidative efficiency during digestion is not a serious problem, as is possible with WCO. The reaction time is typically 5–30 min at 40–90 °C. The CO_2 produced is stripped from the flow stream, dried and detected (usually by NDIR).

In general, UVO instruments are more complex than WCO set-ups. However, since UVO methods can be readily automated, sample manipulation is greatly reduced, which in turn reduces the potential for contamination and improves the sample throughput and precision relative to manual WCO methods. Automation also facilitates near real-time analysis of TOC and DOC in field and industrial applications. Problems in UVO methods can result from a decrease in UV output due to lamp aging, and from differences in lamp-to-lamp performance. Also, ozone venting of the lamp can be a problem for some experimental set-ups. The oxidation efficiency of most UVO methods (about 95%) appears to be between that of WCO and HTC, with the latter assumed to be 100%. Colloidal material such as complex polysaccharides[21] and algal exudates,[22] particulate matter, and some sulfur- and nitrogen-containing organic compounds have been reported to be incompletely degraded by UVO methods, with the apparent exception of the titanium dioxide-catalyzed UV oxidation method.[28] The latter, although promising, has not been widely tested.

4.2 High-temperature Combustion Methods

Two types of HTC method are currently in use: dry combustion and direct aqueous injection. In dry combustion methods, the sample is acidified, dried and the residue is combusted at high temperature in a furnace or in a sealed tube, with or without the presence of a catalyst, such as copper oxide or platinum.[10] The CO_2 formed is purified in a series of traps prior to quantification. The initial drying step is critical and has been performed in several ways: oven drying at 60 °C, vacuum drying at room temperature, and freeze-drying.[1,10] The dry combustion method has several major disadvantages. In particular, the volatile organic fraction is completely lost. Furthermore, the method is time-consuming and does not lend itself to automation or field use. Contamination is a major problem as the dry salts are strong absorbers of organic vapors. To avoid the latter problem, specially designed freeze-dryers must be used and sample transfer steps must be conducted in a clean environment and by an experienced analyst.[1,10] The main advantage of this method is that large samples can be analyzed resulting in a large analytical signal. This method is particularly useful for stable-isotope analysis of DOC because large samples are needed to obtain a sufficiently large analytical signal.[10]

The disadvantages of dry combustion methods are largely overcome by using direct aqueous injection HTC methods. Acidified and sparged samples (Table 1) are injected into a high-temperature column (e.g. 600–900 °C), which usually contains catalytic materials such as platinum, cobalt oxide, copper oxide or silver oxide. It is generally believed that a catalyst is needed to ensure complete conversion of reduced carbon, including carbon monoxide, to CO_2. At the time of injection, the aqueous sample is rapidly volatilized creating a momentary pressure pulse in the system which needs to be damped.[29] The combustion products are swept through the column with a carrier gas, usually high-purity oxygen, although other gases can also be used.[29] The CO_2 that is produced in the combustion column is then purified and dried through a series of traps prior to detection, usually by NDIR.

4.3 Problems Associated with High-temperature Combustion Methods and Applications to Seawater

DOC is the largest pool of reduced carbon in the sea (about 10^{18} g-C), a mass comparable to the total mass of carbon in the atmospheric CO_2 reservoir, or to that stored in terrestrial forests or soil humus matter.[30] Despite the large size of the DOC pool in the oceans, the concentration of DOC is low, about 40 μM (0.5 ppm) in deep waters and 80–90 μM (about 1 ppm) in surface waters. As a result of both this low concentration and the high salt content, the accurate, precise and artifact-free determination of DOC and TOC in seawater has been a major problem despite over 30 years of effort.[1] The necessity for accurate and precise determination of DOC in natural waters has increased sharply over the last few years because several large-scale models and major field studies of marine systems have focused on carbon cycling and its relation to global fluxes, climatic fluctuations and food-web dynamics. This need has been partly met through recent concerted efforts, which involved several workshops and community-wide interlaboratory comparisons.[12,31]

Direct injection high-temperature catalytic oxidation has become widely used for the analysis of TOC and DOC in seawater and other natural waters because this method appears to have greater oxidation efficiency, accuracy and precision than most of the other methods.[1] A large number of experimental HTC systems have been built and several commercial instruments are available.[6,12] However, many of these instruments have problems, particularly in the analysis of seawater. These problems are mainly related to salt deposits in the sample injection

system, memory (or carry-over) effects, system blank evaluation, and motion and vibration sensitivity of NDIR detectors.

The autosampler and injector components of most commercial systems are generally not designed to work with samples with high salt concentrations (e.g. seawater and brines). Many of these injection systems use a sliding metallic or Teflon™ valve, which is susceptible to salt abrasion. In addition to this problem, the high salt content of seawater and brines can cause mechanical and clogging problems in other components of some fully automated TOC instruments. As a consequence, the autoinjection features have sometimes been by-passed or avoided in favor of more reliable manual syringe injection techniques,[6] which are more labor-intensive. Most shipboard TOC analyses are currently being performed by manual injection, often requiring two or three operators working in shifts.

Manual injection systems also have some problems. Most manual systems employ a septumless injection port. The use of this semi-open injector can result in a major disturbance in the gas flow and an increased potential for contamination (e.g. from a contaminated needle). Using a septum-type injector does not solve these problems, as the septum itself can introduce contamination.[32] Another potential problem with manual injection arises from variations in the injection delay interval (i.e. the time between injections), which results in blank variability and consequently in poor precision and accuracy.[33,34] Thus, for manual injection systems, obtaining high precision for very-low-carbon samples depends on the skill level of the operator. The reason for the injection delay effect is not known. However, this problem can be minimized with the use of an autosampler and autoinjector that employs a short and reproducible injection delay time. In addition, most TOC and DOC analyzers use an autosparging system with a constant sample sparging time to ensure complete DIC removal, which is critical. The use of a constant sparging time also ensures that any minor organic contamination that might be present in the sparging gas would be incorporated reproducibly within the blank.[6]

Memory effects may arise because most HTC column designs have a ''cold'' zone at the top, where vaporized OC together with sea salts can deposit on the wall in the dead volume above the column packing.[33,35] During subsequent injections, these deposits may wash off or ''flake off'' randomly and become oxidized, giving rise to an unusually large peak.[35] Owing to the random nature of this memory effect, it is difficult or impossible to correct for it since it may affect standards and samples differently. Discarding unexpectedly small or large peaks will improve the precision, but this may have a negative impact on the accuracy, especially if the large peaks are due to residual OC deposits from prior samples. Thus, the overall impact may be lowered recoveries and less than 100% oxidation efficiencies for even thermally labile standards.[34,35]

Problems arising from sample injection and memory effects have been recently addressed through major improvements in the HTC injection system and column design.[34] The improvements include coupling a self-rinsing, loop autoinjector directly to the HTC column via a platinum tube so that the injection process is closed to the atmosphere, thereby increasing the precision and eliminating potential contamination during manual injection. In order to address memory effects, the cold zone at the top of the column was essentially eliminated, which allowed the sample injection to take place entirely within the column's ''hot'' zone. This modification eliminated effects due to deposits ''flaking off'' from the walls. In addition to this modification, a small amount of deionized water was automatically injected immediately after each sample injection to keep salt deposits from building up in the connecting tubing. The deionized water injection also helped to reduce memory effects from sample-to-sample carry-over and to increase the lifetime of the autoinjection valve by washing out residual salts.

The accuracy of TOC and DOC analyses is strongly dependent on evaluation of the system blank, which in turn is dependent on conditioning of the instrument, especially the HTC column. Column conditioning is often accomplished with repeated injections of low-carbon distilled water. Hundreds of injections have been reported to be required to obtain a low and stable blank for Pt-based HTC columns,[29,36] which is problematic using manual injection. Once the HTC column is conditioned, accurate assessment of the instrument blank depends on the purity of the injected ''zero'' carbon water. Of the various methods that have been used for producing ''zero'' carbon water,[25] closed loop injection of the condensed water that is trapped after the HTC column appears to be the most effective.[36]

Most HTC columns are packed with alumina beads containing Pt, supposedly to catalyze the complete oxidation of OC to carbon dioxide. However, these columns are relatively expensive and the catalytic sites may be subject to poisoning or deactivation by salt deposition.[33,37] It is not clear what function, if any, the catalyst plays in the oxidation process. In fact, several recent studies have shown that complete oxidation can be achieved with HTC columns without a catalyst, for example using pure quartz beads.[14,34] The latter is less expensive and has a lower blank.[34]

From the above, it is apparent that the precision and accuracy of HTC methods are affected by various factors including the presence or absence of a cold zone (and dead volume) at the head of the column,

Table 2 Advantages and disadvantages of TOC and DOC methods for aqueous samples. Modified from MacKinnon[38]

Parameter	Manual WCO (persulfate)	Automated WCO (persulfate)	UV oxidation (persulfate enhanced)	Dry or sealed tube combustion (Pt catalyzed)	Direct injection HTC
Oxidation efficiency (relative to HTC)	Generally 85–95%, but may be close to 100% depending on proper blank evaluation and sufficient reagent excess	Generally 85–95%, but may be close to 100% depending on proper blank evaluation and sufficient reagent excess	About 95%, but may be close to 100% if TiO_2 catalyzed, (Abdullah and Eek[28])	100%	100%
Changes in oxidation efficiency with time	Oxidative efficiency can decrease during digestion, especially for high TOC or chloride-containing samples	Oxidative efficiency can decrease during digestion, especially for high TOC or chloride-containing samples	Possible, due to degradation of lamp	No	No
Precision	Variable	Good	High	High with proper precautions	High with precautions taken to eliminate memory effects
Blank	High and difficult to assess	High and difficult to assess	Low	Low with proper precautions	Difficult to assess, but low after column conditioning;
Contamination	Variable	Low	Low	Low with proper precautions	Low
Interferences with oxidation	Chloride	Chloride	None	None	None
Memory or carry-over effects	None	None if reaction vessel is properly rinsed	None	None	None if injection system and column are properly designed
Sample preparation	Extensive	Minimal	Minimal	Extensive	Minimal
Automation	Possible	Yes	Yes	No	Yes
Real-time analysis	No	Yes	Yes	No	Yes
Shipboard use	Possible, but not recommended	Yes	Yes	No	Yes
Complete oxidation of colloids and particles	Yes	Yes	No	Yes	Yes
Loss of volatile organics	Moderate	Moderate	Minor	Complete	Minor
Cost of equipment	Low	High	High	Moderate	High

memory and carry-over effects, contamination and salt deposition during injection, variable injection delay time, and sample sparging time.[20,33,34] In addition, variations in carrier gas flow rate and column temperature can have a strong impact on the precision.[34] Most of these sources of error are eliminated or minimized by using a closed self-rinsing autoinjection and autosparging system directly coupled to a noncold zone, low dead volume combustion column.[34] These recent improvements in HTC instrument design should benefit other fields besides oceanography, especially where high precision, low detection limits, salts, or field portability are important. A summary of the advantages and disadvantages of the major TOC and DOC methods is given in Table 2.

ACKNOWLEDGMENT

This article was supported in part by grants from the US National Science Foundation, Division of Ocean Sciences (OCE-9315821 and OCE-9711206).

ABBREVIATIONS AND ACRONYMS

COD	Chemical Oxygen Demand
DIC	Dissolved Inorganic Carbon
DOC	Dissolved Organic Carbon
FID	Flame Ionization Detection
GC/MS	Gas Chromatography/Mass Spectrometry
HTC	High-temperature Combustion
ICP/AES	Inductively Coupled Plasma Atomic Emission Spectrometry
NDIR	Nondispersive Infrared
OC	Organic Carbon
POC	Particulate Organic Carbon
PVC	Polyvinyl Chloride
TOC	Total Organic Carbon
TOD	Total Oxygen Demand
UV	Ultraviolet
UVO	Ultraviolet Oxidation
VOC	Volatile Organic Carbon
WCO	Wet Chemical Oxidation

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Flow-injection Techniques in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Luminescence in Environmental Analysis • Sampling Considerations for Biomonitoring

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis • Infrared Spectroscopy in Process Analysis

REFERENCES

1. P.J. Wangersky, 'Dissolved Organic Carbon Methods: A Critical Review', *Mar. Chem.*, **41**(1–3), 61–74 (1993).
2. J.H. Sharp, E.T. Peltzer, M.J. Alperin, G. Cauwet, J.W. Farrington, B. Fry, D.M. Karl, J.H. Martin, A. Spitzy, S. Tugrul, C.A. Carlson, 'Procedures Subgroup Report', *Mar. Chem.*, **41**(1–3), 37–49 (1993).
3. A.R. Abdel-Moati, 'Adsorption of Dissolved Organic Carbon (DOC) on Glass Fibre Filters During Particulate Organic Carbon (POC) Determination', *Water. Res.*, **24**(6), 763–764 (1990).
4. E.T. Peltzer, N.A. Hayward, 'Spatial and Temporal Variability of Total Organic Carbon along 140 °W in the Equatorial Pacific Ocean in 1992', *Deep-Sea Res. II*, **43**(4–6), 1155–1180 (1996).
5. T. Fukushima, A. Imai, K. Matsushige, M. Aizaki, A. Otsuki, 'Freshwater DOC Measurements by High-Temperature Combustion: Comparison of Differential (DTC-DIC) and DIC Purging Methods', *Water Res.*, **30**(11), 2717–2722 (1996).
6. E.T. Peltzer, P.G. Brewer, 'Some Practical Aspects of Measuring DOC-Sampling Artifacts and Analytical Problems with Marine Samples', *Mar. Chem.*, **41**(1–3), 243–252 (1993).
7. L.A. Kaplan, 'Comparison of High-Temperature and Persulfate Oxidation Methods for Determination of Dissolved Organic Carbon in Freshwaters', *Limnol. Oceanogr.*, **37**(5), 1119–1125 (1992).
8. K. Hiraki, M. Kimura, Y. Suzuki, H. Hattori, Y. Nakaguchi, H. Tsubota, 'Determination of Dissolved Organic Carbon in Northwest Pacific Ocean. The Application of Ultra-Low-Temperature Long-Time Storage', *Kinki Daigaku Rikogakubu Kenkyu Hokoku*, **32**, 19–24 (1996).
9. J. Van Leeuwen, M. Drikas, D. Bursill, B. Nicholson, 'Contamination of Samples for DOC Analysis', *Water: Off. J. Austral. Water Wastewater Assoc.*, **24**(1), 12 (1997).
10. B. Fry, E.T. Peltzer, C.S. Hopkinson, Jr, A. Nolin, L. Redmond, 'Analysis of Marine DOC using a Dry Combustion Method', *Mar. Chem.*, **54**, 191–201 (1996).
11. W. Chen, P.J. Wangersky, 'High-Temperature Combustion Analysis of Dissolved Organic Carbon Produced in Phytoplankton Cultures', *Mar. Chem.*, **41**(1–3), 167–171 (1993).
12. J.H. Sharp, R. Benner, L. Bennett, C.A. Carlson, S.E. Fitzwater, E.T. Peltzer, L.M. Tupas, 'Analysis of Dissolved Organic Carbon in Seawater: The JGOFS EqPac Methods Comparison', *Mar. Chem.*, **48**, 91–108 (1995).
13. L.A. Kaplan, 'A Field and Laboratory Procedure to Collect, Process, and Preserve Freshwater Samples for Dissolved Organic Carbon Analysis', *Limnol. Oceanogr.*, **39**(6), 1470–1476 (1994).
14. P.M. Williams, J.E. Bauer, K.J. Robertson, D.M. Wolgast, M.L. Occelli, 'Report on DOC and DON Measurements made at Scripps Institution of Oceanography, 1988–1991', *Mar. Chem.*, **41**(1–3), 271–281 (1993).
15. L.M. Tupas, B.N. Popp, D.M. Karl, 'Dissolved Organic Carbon in Oligotrophic Waters: Experiments on Sample Preservation, Storage and Analysis', *Mar. Chem.*, **45**, 207–216 (1994).

16. N.S. Bolan, S. Baskaran, S. Thiagarajan, 'An Evaluation of the Methods of Measurement of Dissolved Organic Carbon in Soils, Manures, Sludges, and Stream Water', *Commun. Soil Sci. Plant Anal.*, **27**(13 & 14), 2723–2737 (1996).
17. G.M. Ferrari, M.D. Dowell, S. Grossi, T. Cristina, 'Relationship between the Optical Properties of Chromophoric Dissolved Organic Matter and Total Concentration of Dissolved Organic Carbon in the Southern Baltic Sea Region', *Mar. Chem.*, **55**(3/4), 299–316 (1996).
18. A. Brandstetter, R.S. Sletten, A. Mentler, W.W. Wenzel, 'Estimating Dissolved Organic Carbon in Natural Waters by UV Absorbance (254 nm)', *Z. Pflanzenernaehr. Bodenkd.*, **159**(6), 605–607 (1996).
19. J. Pages, J.-P. Torreton, R. Sempere, 'Dissolved Organic Carbon in Coral-Reef Lagoons, by High Temperature Catalytic Oxidation and UV Spectrometry', *C.R. Acad. Sci., Ser.IIa: Sci. Terre Planetes*, **324**(11), 915–922 (1997).
20. Ye.N. Tarasov, 'Correlation of Organic Carbon with Different Kinds of Oxidizability in the Open Waters of Lake Baikal', *Hydrobiol. J.*, **8**, 53–57 (1972).
21. W. Chen, P.J. Wangersky, 'A High-Temperature Catalytic Oxidation Method for the Determination of Marine Dissolved Organic Carbon and Its Comparison with the UV Photo-Oxidation Method', *Mar. Chem.*, **42**, 95–106 (1993).
22. W. Chen, P.J. Wangersky, 'Production of Dissolved Organic Carbon in Phytoplankton Cultures as Measured by High-Temperature Catalytic Oxidation and Ultraviolet Photo-Oxidation Methods', *J. Plankton Res.*, **18**(7), 1201–1211 (1996).
23. G. Cauwet, 'HTCO Method for Dissolved Organic Carbon Analysis in Seawater: Influence of Catalyst on Blank Estimation', *Mar. Chem.*, **47**, 55–64 (1994).
24. A.I. Agatova, V.V. Sapozhnikov, N.I. Torgunova, 'Comparative Determination of Dissolved Organic Matter in Different Seas by Wet Combustion and High Temperature Catalytic Oxidation', *Okeanolog.* (Moscow), **36**, 470–477 (1996).
25. E.T. Pelzer, B. Fry, P.H. Doering, J.H. McKenna, B. Norman, U.L. Zweifel, 'A Comparison of Methods for the Measurement of Dissolved Organic Carbon in Natural Waters', *Mar. Chem.*, **54**, 85–96 (1996).
26. J.H. McKenna, P.H. Doering, 'Measurement of Dissolved Organic Carbon by Wet Chemical Oxidation with Persulfate: Influence of Chloride Concentration and Reagent Volume', *Mar. Chem.*, **48**, 109–114 (1995).
27. H. Ogawa, N. Ogura, 'Comparison of Two Methods for Measuring Dissolved Organic Carbon in Sea Water', *Nature*, **356**(6371), 696–698 (1992).
28. M.I. Abdullah, E. Eek, 'Automatic Photocatalytic Method for the Determination of Dissolved Organic Carbon (DOC) in Natural Waters', *Water. Res.*, **30**(8), 1813–1822 (1996).
29. A. Skoog, D. Thomas, R. Lara, K.-U. Richter, 'Methodological Investigations on DOC Determinations by the HTCO Method', *Mar. Chem.*, **56**(1–2), 39–44 (1997).
30. J.I. Hedges, 'Global Biogeochemical Cycles: Progress and Problems', *Mar. Chem.*, **39**(1–3), 67–93 (1992).
31. R. Benner, V. Von Bodungen, J. Farrington, J. Hedges, C. Lee, F. Mantoura, Y. Suzuki, P.M. Williams, 'Measurement of Dissolved Organic Carbon and Nitrogen in Natural Waters: Workshop Report', *Mar. Chem.*, **41**(1–3), 5–10 (1993).
32. Y. Suzuki, E. Tanoue, H. Ito, 'A High-Temperature Catalytic Oxidation Method for the Determination of Dissolved Organic Carbon in Seawater: Analysis and Improvement', *Deep-Sea Res.*, **39**(2), 185–198 (1992).
33. E.M. Perdue, F. Mantoura, J. Ertel, C. Lee, K. Mopper, G. Peyton, E. Tanoue, P.M. Williams, O. Zafirou, P. Coble, 'Mechanisms Subgroup Report', *Mar. Chem.*, **41**(1–3), 51–60 (1993).
34. J. Qian, K. Mopper, 'Automated High-Performance, High-Temperature Combustion Total Organic Carbon Analyzer', *Anal. Chem.*, **68**(18), 3090–3097 (1996).
35. E. Tanoue, 'Vertical Distribution of Dissolved Organic Carbon in the North Pacific as Determined by the High-Temperature Catalytic Oxidation Method', *Earth Planet. Sci. Lett.*, **111**, 201–216 (1992).
36. R. Benner, M. Strom, 'A Critical Evaluation of the Analytical Blank Associated with DOC Measurements by High-Temperature Catalytic Oxidation', *Mar. Chem.*, **41**(1–3), 153–160 (1993).
37. J.E. Bauer, M.L. Occelli, P.M. Williams, P.C. McCaslin, 'Heterogeneous Catalyst Structure and Function: Review and Implications for the Analysis of Dissolved Organic Carbon and Nitrogen in Natural Waters', *Mar. Chem.*, **41**(1–3), 75–89 (1993).
38. M.D. MacKinnon, 'The Measurement of Organic Carbon in Sea Water' in *Marine Organic Chemistry*, eds. E.K. Duursma, R. Dawson, Elsevier, Amsterdam, 415–443, 1981.

X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Philip A. Russell
Oxford Instruments, Oxford, UK

1 Introduction **1**
2 History **2**
3 Sample Preparation **3**
3.1 Sample Preparation Using Filtration and Centrifugation 4
3.2 Sample Presentation to the Instrument 4
3.3 Ion-exchange Papers 5
3.4 Preconcentration (Extraction) Techniques 5
3.5 Oxidation State – Speciation Using Preconcentration Methods 5
3.6 Sample Preparation for Total-reflection X-ray Fluorescence 5
4 Environmental Applications **6**
4.1 Analysis of Natural Waters 6
4.2 Nonhazardous Liquid Waste 7
4.3 Liquid Hazardous Waste 8
4.4 Environmental Samples Related to the Nuclear Industry 9
4.5 Novel Applications 10
5 Interference **10**
5.1 Chemical Interference 10
5.2 Instrument Interferences 10
6 Analysis Techniques **11**
6.1 Calibration Techniques 11
6.2 Fundamental Parameter Techniques 13
6.3 International Standard Methods 13
7 Regulations **13**
7.1 Natural Water 14
7.2 Nonhazardous Liquid Waste 14
7.3 Liquid Hazardous Waste 14
8 Quality Control and Troubleshooting **16**
9 Method Development **17**
9.1 Energy-dispersive X-ray Fluorescence 17
9.2 Wavelength-dispersive X-ray Fluorescence 19
9.3 Direct ''As-received'' Analysis 19
9.4 Analysis Using Total-reflection X-ray Fluorescence 22
9.5 Analysis of Filters 22
10 Comparison with Other Spectroscopic Methods **23**
10.1 Comparison of Technology 23
10.2 Business Issues 24
10.3 Regulatory Issues 24
11 Future Developments **24**
11.1 Ultratrace Analysis Using Microsample X-ray Analysis 24
Abbreviations and Acronyms **24**
Related Articles **25**
References **25**

X-ray fluorescence (XRF) spectrometry is an instrumental technique used for multielement analysis. It is one of only three techniques available to the modern analyst, the others being mass spectrometry and optical emission spectroscopy, the latter usually being the inductively coupled plasma (ICP) variant. The determination of environmental sensitive elements in the range P–U in aqueous and/or organic matrices provides challenges to all multielement techniques. With recent improvements in instrumentation and methods design (simplified sample preparation compared with other techniques), XRF, in particular for highly contaminated liquid waste, is becoming the preferred technique. XRF is shown to give the most cost effective and best overall speed of analysis across a range of diverse matrix types. Potable drinking water to highly toxic industrial sludge/waste can be analyzed for either precise ultralow level (parts per billion) or screening purposes, respectively. Energy-dispersive X-ray fluorescence (EDXRF) has provided the best all-round performance with the introduction of dedicated instruments and standard methods for the analysis of environmentally sensitive elements in all forms of liquid matrices.

1 INTRODUCTION

XRF spectroscopy involves the measurement of element-characteristic X-rays emitted from an atom after bombardment by X-ray photons from an X-ray source.

The process of XRF is twofold – first the bombarding of an atom with X-rays from a suitable source causing the ejection of electrons from the inner shells of atoms, and second the detection of characteristic X-rays produced

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

from the cascading of electrons from outer shells to fill vacancies in inner shells.

XRF spectrometry encompasses many different forms but is based upon two principal techniques:

- wavelength-dispersive X-ray fluorescence (WDXRF);
- energy-dispersive X-ray fluorescence (EDXRF).

The differences between these techniques can be summarized in the following way. WDXRF uses X-ray excitation over a wide energy range in the form of high-voltage (20–100 kV) and -power (200–4000 W) X-ray tube sources with high-resolution detection using X-ray diffraction crystals to separate the elements. EDXRF uses narrow energy range X-ray excitation in the form of X-ray sources to low-voltage/power X-ray tubes; energy-restricting filters and/or secondary excitation targets with broad energy range detection using proportional counters or semiconductor Si(Li) or germanium detectors. The two techniques can often be viewed as complementary. Newer derivations of EDXRF are total-reflection X-ray fluorescence (TXRF) and synchrotron radiation X-ray fluorescence (SRXRF).

XRF spectrometry is used for the qualitative and quantitative elemental determinations of solids and liquids. The element range can be from atomic number (Z) > 11 (Na) for liquid matrices. More typically, $Z = 15$ (phosphorus) and above is a practical limitation for the quantitative measurement of environmentally important elements in liquid media owing to the affects of sample support films used.

XRF's true power is not so much in the dedicated analysis of a particular matrix, but its ability to provide a technique for the analysis of a very wide and diverse range of liquid sample types, i.e. methods for ultratrace analysis in waters through to industrial liquid hazardous waste (LHW). The advantages of XRF lie in the technique's ability to measure simultaneously a wide range of elements often presented in their ''as-received'' liquid form. Where required, preconcentration techniques for elements in liquid are usually simple.

Detection limits for both ''as-received'' liquid samples are usually in the 1–10 mg kg^{-1} range and as much as six orders of magnitude lower for samples prepared by preconcentration, the latter depending upon the instrument type employed. Accuracy and precision vary according to matrix constraints, elements determined and methods of sample preparation used. Accuracy of $<\pm 0.3\%$ relative and precision of $<0.2\%$ routinely can be achieved for the analysis of clean ''as-received'' environmental samples. However, as sampling, homogeneity and sample preparation are usually the controlling factors in the analysis of environmental samples, more typical accuracies of $\pm 1\%$ relative and precision levels of 2% can be expected.

Matrix and interelement interference can be problematic when measuring samples of highly variable forms in environmental samples. Speciation analysis can only be carried out through indirect methods of analysis.

Handbooks by Van Grieken and Markowwicz[1] and Lachance and Claisse[2] provide an overview of XRF theoretical principles and techniques. These will be of particular interest to those wishing to obtain a wider appreciation of XRF capability.

Burkhalter[3] outlined the long established use of XRF in the analysis of water for metal contaminants using preconcentration techniques. XRF is a well-established technique for geochemical analysis, is used extensively for soil analysis and, as a more recent adjunct, for sludge and wastewater. Reviews by Bacon et al.[4,5] and Ellis et al.[6,7] provide excellent recent summaries of all aspects of XRF instrumentation and applications.

2 HISTORY

The analysis of environmental samples using XRF has a relatively short history. The first analyses were based upon WDXRF. Relatively simple matrices such as metals present on filter papers were seen as ideal samples to be measured. The use of preconcentration methods using coprecipitation and newly developed ion-exchange papers in the late 1960s and early 1970s encouraged many workers to develop dedicated methods. Ellis et al.[8] provide a review of these methods.

The introduction and development of commercial EDXRF instruments in the early 1970s provided a cost-effective qualitative and quantitative measurement capability. The early EDXRF instruments were not capable of measuring complex, highly variable matrices and thus were somewhat restricted in their capability. However, their potential to determine metals and particularly heavy elements was clear. Powerful multielement sequential WDXRF instruments were seen as the only viable quantitative XRF technique owing to its ability to resolve complex overlapping X-ray lines; however, it remained expensive, and not the ideal tool for the analysis of environmental liquid samples.

The introduction of inductively coupled plasma atomic emission spectroscopy (ICPAES) in the early 1980s provided an ideal technique for the analysis of complex samples for all environmentally toxic metals even though the techniques were expensive and time-consuming. Many methods for analysis of solid waste, water and wastewater samples using ICPAES and the earlier developed atomic absorption spectroscopy (AAS) technique

were developed in the 1980s and integrated into standard analytical protocols such as United States Environmental Protection Agency (USEPA) SW846.[9] These protocols are based upon digestion techniques, becoming more refined with the introduction of microwave digestion. The introduction of such protocols reduced the need for further development in the uses of XRF and not until the late 1980s and early 1990s were instruments developed to compete with the ICP-dominated market.

The increasing use of modern personal computers in XRF instruments and improvements, especially in EDXRF instrumentation, in the mid-1980s gave researchers the opportunity to develop new methods for the analysis not only of waters and wastewater but also LHW samples. The development of TXRF in the late 1980s provided an analytical tool capable of achieving detection limits for metals in water equal to modern inductively coupled plasma mass spectrometry (ICPMS) instruments.

In the early 1990s, EDXRF instruments were developed to provide analytical methods that were capable of determining environmentally sensitive elements such as Hg, As and Cd at regulatory levels in hazardous waste (liquid and solid) without the need for expensive, time-consuming sample preparation.

Currently, EDXRF is providing the main analytical tool for the analysis of hazardous waste in many laboratories. The promulgation of performance-based measurement techniques and methods by organizations such as the USEPA and the American Society for Testing and Materials (ASTM) has allowed XRF, in general, to be accepted as a viable cost-effective alternative to traditional ICPAES and AAS methods used in many laboratories for the analysis of liquid waste samples.

The analysis of a few elements, e.g. Be and B, must be carried out by atomic absorption/emission techniques owing to the limitations of XRF; however, this is compensated by the ability XRF spectrometry to perform the analysis of halogens (excluding F) and S.

Competition in the late 1990s between XRF and ICP as the only multielement techniques available will provide, over the coming years, methods that are based upon the performance needs of both industry and regulators in terms of cost, accuracy and appropriateness.

3 SAMPLE PREPARATION

As with all analytical instrumental techniques, sample preparation plays a crucial role in obtaining data and results that can be viewed as significant. A series of papers by Tatro[10] describing the current sample preparation methods employed by the USEPA for hot-plate and microwave digestion provides an excellent background to the current state of the art in most modern laboratories using AAS and ICP instruments. Wheeler[11] and Van Grieken and Markowicz[1] provide reviews of many X-ray sample preparation techniques, especially those specific to trace metal analysis in liquids.

Sample preparation techniques used by XRF can be separated into two main categories.

- Digestion and wet chemical techniques. These techniques share a common history with optical absorption and emission spectroscopic techniques such as AAS and ICP.
- XRF-specific sample preparation techniques. These techniques are based upon minimizing sample preparation.

Environmental analysis of samples requires the understanding of the limitations and sources of error intrinsic within any sample preparation regime. When comparing techniques used for the analysis of ''environmentally sensitive'' elements, e.g. As, Cd and Hg, especially at trace and ultratrace levels, the sources of error must be related to the quality of information required. Regulatory requirements are based on total element content, and techniques that do not supply this information directly need to be evaluated before data can be reported. Elements such as cadmium, mercury, lead, thallium and other toxic volatile metals are easily lost from samples that undergo extensive chemical preparation and heating. When comparing methods based on detection limits only, it is easy to eliminate techniques that can provide valuable and accurate data but not the ultimate in detection or element coverage.

XRF benefits from the potential ability to analyze samples in their ''as-received'' form. The most common method of sample preparation used for the determination of elements at trace level, i.e. $1-1000\,\mathrm{mg\,kg^{-1}}$, involves placing ''as-received'' solution in a measuring cup. A thin ($1.5-6\,\mu\mathrm{m}$) polymer film supports the liquid sample. Solutions that are in a pure aqueous or organic form can be determined directly. Samples containing particles and/or multiphase solutions must be pretreated to produce a homogeneous sample prior to presentation to an instrument.

For solid samples such as dried sludge or filter-cake residue taken from liquefied material, the method of pelletizing can be employed. This method can involve grinding or milling of a sample to produce a homogeneous fine powder. The resultant powder is then pressed into a pellet to produce a self-supporting disk that can be placed into the instrument for analysis. In applications where samples are sufficiently fine they can be pressed with no other preparation. Other techniques have been

developed to homogenize samples such as sludges and slurries to produce a methodology that will allow samples to be presented to an instrument in a uniform and repeatable way.

The analysis of environmentally sensitive elements in waters and sludges creates some challenges for the XRF technique. It is often the case that the most environmentally toxic elements require detection limits at levels typically below those that can be achieved by ''as-received'' sample preparation techniques. For this reason methods of preconcentration are required.

Details of the above sample preparation techniques will be discussed in the following sections starting with filtration and the analysis of the resultant essentially solid sample, then moving on to preconcentration techniques involving element extraction.

3.1 Sample Preparation Using Filtration and Centrifugation

Knowledge of the differing sensitivity of elements across the XRF spectrum, i.e. from low- to high-Z elements, dictates the amount of filtrate needed to achieve a usable detection limit for the elements of interest. A filtrate can be collected for use in two ways. The first is to collect a large quantity that can be subsequently dried and treated as a bulk sample. The second is to minimize the amount collected and to treat the sample as a thin-film coating on the surface of the filter used.

For bulk samples such as filtered sludges where the solid component is large enough, samples can be either transferred to sample cups where a thin polymer film supports the loose material or, if the sample is capable of compaction, pressed into a solid pellet. In the latter case many samples containing large amounts of organic matter are difficult to compress into pellets, hence loose material in a cup is the only option.

Suggested schemes for the analyses of sludge using XRF would be as follows:

- Centrifugation of the sample to separate solids from the liquid phase. The two phases can then be measured separately.
- Solids can be compacted into a pellet or simply placed in a sample cup. Compaction of the solid sample is difficult if the sample is still damp. Accurate measurement of the solid phase assumes that the particle size of the solid is relatively small, i.e. $<200\,\mu m$. It is often necessary to homogenize the sample prior to pelletizing. This is carried out using a laboratory blender.

The analysis of sludge as pellets or as loosely compacted sample (contained in a sample cup) uses methods similar to those used for solids and soils analyses. The performance of analyses when measuring pressed pellets is better than that of loose powders owing to consistency of sample preparation and their increased density.

Where the particulate content is low, typically 47 mm diameter ashless filter papers are used for laboratory-based analysis. Depending upon the requirements of the analysis, usually a paper capable of achieving particulate entrapment of $>0.45\,\mu m$ is acceptable and may be mandated by methods, especially those used in the water industry.

The filtrates will adhere to one surface of the filter paper, producing a filter cake on the surface. The thickness or total mass of the filter cake is important. If the mass of the filter cake is too high then interelement effects will begin to dominate. It is important to understand that light elements will be affected more by the mass of the filtrate than heavier elements. Therefore, any analysis regime must first establish what elements are required. The lightest element required dictates what the masses should be as these have the lowest sensitivity. Typically 20 mg is a good maximum to achieve. For potable drinking and natural waters this level is reasonable and for preconcentration from 1 L should rarely be exceeded. Actual methods of measuring such filter papers may vary from one instrument to another. These will be discussed further in section 9.

3.2 Sample Presentation to the Instrument

Figure 1(a–c) show typical set-ups of sample presentation used for the determination of analytes in water and wastewater applications. This shows the main types of holders used to retain environmental samples for presentation to spectrometers. All XRF instrument

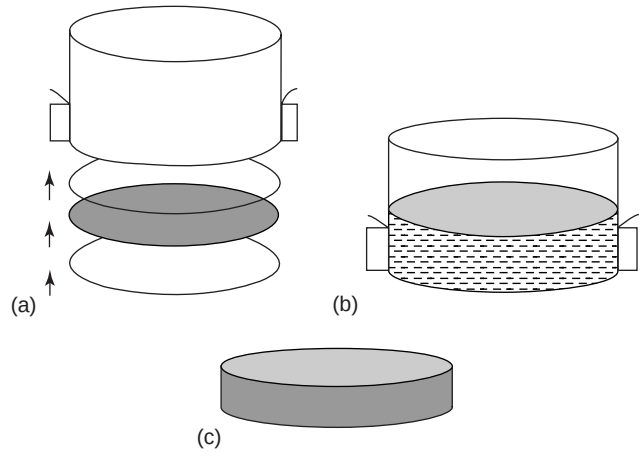

Figure 1 Sample presentation. (a) Filter held between two films; (b) ''as-received'' sample into vented cup; (c) filter cake pressed pellet.

manufacturers provide appropriate holders for samples, in addition to holders being available from laboratory suppliers of XRF consumables.

3.3 Ion-exchange Papers

Methods are based upon filters impregnated with copolymers using sulfonic acid active sites for cation exchange and quaternary ammonium active sites (chloride ion form) for anion exchange. A disadvantage in their broad use is that their efficiency is controlled by pH; however, this can be used to advantage in obtaining oxidation state speciation information, which is valuable for the interpretation of element toxicity. For a number of very important analytes, e.g. chromium, arsenic, antimony and selenium, the use of pH can be used to separate the more toxic oxidation states.

Ellis et al.[8] used commercially available cation-exchange papers in the form of sulfonic acid copolymers and Varshal et al.[12] used cellulose papers impregnated with aminocarboxylic acid to form the cation-exchange paper. Examination of the use of ion-exchange paper regimes typically shows that the quantitative extraction of elements using strong acid cation-exchange papers is limited by pH. Ellis et al.[8] showed the extraction of Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Ag, Cd, Sb, Hg, Tl and Pb, but the species existing as anions [i.e. As(III), Se(IV) and Sb(III)] were not extracted.

The majority of analytes required to be determined fall into the cation group of elements. Anion-exchange papers are used for more specific separations where information on speciation is sought.

A typical procedure for the use of cation exchange papers would be as follows:

- Adjust the sample pH to the appropriate level, typically 3–5, and add any required buffer.
- Prefilter the sample (100 mL–1 L) through a 0.4–0.5-μm ashless filter to remove suspended particulates.
- Filter the sample through the cation-exchange paper. This can be either multiple passes or through a cascade of papers depending upon the expected concentration levels of elements to be captured and the loading capacity of the filters. Flow rates are usually of the order of 30–50 mL min^{-1} for a 25–47-mm diameter paper.

Ion-exchange resins are used, but their use in ''direct analysis'' of the resin is limited owing to the dilution factor of the resin itself and the effect of S or Cl functional groups in attenuating the fluorescence signal, especially from lighter elements. Typically they are used to remove an interfering matrix, e.g. seawater, from the elements of interest.

3.4 Preconcentration (Extraction) Techniques

This method of concentrating environmentally important analytes from waters has been used extensively in XRF. Linder et al.,[13] Ellis et al.[8] and Prange et al.[14] used these methods for different types of XRF instrumentation. The method usually benefits from a single element, usually if not always iron, being present at relatively high concentrations compared with the analytes of interest. Thus the technique is often referred to as ''coprecipitation''. The most effective method of coprecipitation uses the sodium salt form of dibenzyl dithiocarbamate (DBDTC). This carbamate form has been shown to achieve good recoveries for a wide range of analytes and their various valency states. A notable exception is the recovery of arsenic and antimony in the valency 5 state and Cr in the valency 3 state. However, for arsenic, Del Carmen et al.[15] showed that the use of varying pH could achieve an analysis in both valency forms, thus providing valuable information on the more toxic As(III) form.

Coprecipitation techniques will produce a filter cake similar to those of the particulate filter method (see section 3.1) and thus can usually be quantified in a similar way. These will be discussed more fully in section 9 on method development.

3.5 Oxidation State – Speciation Using Preconcentration Methods

Most regulations require total element concentrations to be measured for quality control (QC) of prescribed elements. However, specific species of elements such as Cr, Se and As are far more biologically toxic than others. For this reason, speciation analysis can provide critical information for the environmental analyst. ICP coupled methods are ideally suited to this type of analysis in that they are capable of determining the different species within a sample. However, destructive sample preparation techniques such as acid digestion will remove this information so even the most ideal techniques can be limited for this type of analysis. In XRF, methods to determine speciation rely upon the sample preparation stage to separate species prior to analysis. This is carried out using pH control during the preconcentration stage of sample preparation and the selective use of appropriate concentration techniques best suited for the species required. This is best seen in Table 1, where the most useful element speciation methods are summarized.

3.6 Sample Preparation for Total-reflection X-ray Fluorescence

Sample preparation for TXRF spectroscopy is different from that of conventional EDXRF and WDXRF. TXRF

Table 1 Methods of speciation

Method	Element species	Summary of details	Ref.
NaDBDTC	As(III)	Adjust sample to pH 2. Form precipitate with NaDBDTC and collect on filter paper. Measure directly using XRF	Del Carmen et al.[15]
Fe(III) coprecipitation	As(V)	Coprecipitate with Fe(III) at pH 5. Collect on filter paper. Measure directly using XRF. Best suited to WDXRF owing to the presence of Fe	Del Carmen et al.[15]
APDC[a]	Se(IV) only	Adjust sample to pH 5. Form precipitate with APDC and collect on filter paper. Measure directly using XRF	Castilla y Cortazar and Llopis Cosin[16]
Activated carbon	Se(IV + 0)	Mix with activated carbon. Collect on filter paper. Measure directly using XRF	Castilla y Cortazar and Llopis Cosin[16]

[a] APDC, ammonium pyrrolidinedithiocarbamate.

Table 2 TXRF preparation

Elements	Matrix	Ref.
S, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Pb, Se, Rb, Sr, Mo, Cd, Ba	Trace metals in rainwater	Stossel and Prange[17]
V, Mn, Fe, Co, Ni, Cu, Zn, Se, Mo, Cd, Hg, Pb, U	Seawater	Prange et al.[14]
Cu, Hg, Pb	Potable water	Holynska et al.[18]
K, Ca, Cr, Mn, Fe, Ni, Cu, Zn, Pb	Metals in biofilms in river water	Friese et al.[19]

requires all samples initially to be in solution form. Presentation of the sample to the spectrometer is by way of evaporation on quartz or Perspex™ (Plexiglas™) supports. Thus dissolution methods are used as for AAS and ICP to provide a wide range of preparations. Table 2 gives a selection of studies using TXRF for the preparation of environmental samples.

Practitioners of TXRF have developed many techniques for the preparation of environmental samples, but they suffer from the same drawbacks of other dissolution technologies in that the sample sizes are very small and the methods have the potential for analyte loss. An important area of research and development with respect to TXRF is in the determination of metals in seawater.[14] Analysis is based on the separation of the metals from any salt matrix and final concentration of trace metals as complexes using, for example, dithiocarbamate. Complexes are usually taken up in an organic media and pipetted on to TXRF quartz or Plexiglas™ supports, where they are evaporated to dryness ready for analysis.

4 ENVIRONMENTAL APPLICATIONS

Applications for the analysis of liquids cover a broad field that can be broken down into a number of key areas:

- natural waters – e.g. potable or drinking water, river and seawater;
- wastewater or nonhazardous liquid waste – i.e. industrial or domestic waste and agricultural muck used for land improvement; these can include sewage sludge, runoff or groundwater;
- LHW, i.e. aqueous and/or organic liquid; sludge or other samples potentially containing tens of percent solids.

4.1 Analysis of Natural Waters

Laboratory analysts typically require techniques capable of achieving detection limits 10 times lower than any regulatory limit. For potable water analysis, this requirement can only be achieved using a preconcentration technique when using conventional EDXRF or WDXRF. These techniques are similar if not the same as those used for preconcentration prior to using atomic spectroscopy techniques such as AAS and ICP. This analysis of essentially clean water includes river water, seawater and groundwater, the last from relatively noncontaminated sites.

There are two components that must be considered with respect to the analysis of natural water: suspended solids and dissolved elements. Requirements to determine the constituents of these two phases are often necessary. Particulate filtration of varying volumes of these waters using commercially available ashless filters will achieve the necessary concentration of analytes to allow direct analysis of the filtrates by XRF. A typical scheme for the preparation of samples for analysis of natural waters would be to filter 0.1–1 L of water through a series of

Table 3 Preconcentration techniques

Method	Analytes	Refs.
Ion exchange	Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Ag, Cd, Sb, Hg, Tl, Pb	Ellis et al.[8]
Coprecipitation	Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Ag, Cd, Sb, Hg, Tl, Pb	Ellis et al.,[8] Prange et al.,[14] Holynska et al.,[18] Panayappan et al.[20]
Suspended solids	All elements	Soils analysis literature

filters. The first filter to collect suspended particulates using a filter capable of trapping particulates down to 0.45 μm is followed by filters to trap dissolved elements using cation- and/or anion-exchange papers/resins.

This analysis regime has the drawback in that cation- and anion-exchange papers are pH sensitive. Section 3.3 describes the use of such filters. Practitioners are advised to seek the latest information from manufacturers of such papers to address any problems of pH sensitivity, paper capacity and known chemical interferences. The above regime using a series of ion-exchange filters does, however, leave some doubt as to what may pass through all of the papers. Elements not trapped (concentrated) on either the particulate filter paper or ion-exchange papers will be at very low levels. This leaves one area still to be determined using XRF.

Workers in this field[8,13,14] conclude that the use of the preconcentration method using DBDTC gives the best overall performance for the extraction of metals from clean aqueous solutions. Subsequent reviews and studies have confirmed this conclusion.

Improvements in XRF technology and the introduction of the TXRF technique have made modifications to this method. Studies by Prange et al.,[14] Stossel and Prange[17] and Holynska et al.[18] have refined this technique for use with TXRF.

Varshal et al.[12] described a methodology for XRF analysis of natural waters using cellulose filters impregnated with conformationally flexible aminocarboxylic groups known as DETATA.

The analysis of seawater is treated as a special case. Owing to the high concentration of NaCl present, the metals must first be separated from the salt matrix using chromatographic techniques. Prange et al.[14] described a technique for TXRF using chelation of the metals and selective chromatographic absorption of the metal complexes on lipophilized silica gel carriers with subsequent elution by chloroform–methanol mixtures. The final organic eluate is then dried on TXRF sample supports.

The analysis of natural waters is a key TXRF application. The use of WDXRF is restricted owing to the destruction of filters at high X-ray power typically used by such instruments.

Table 3 provides a review of the techniques that can be used for the analysis of natural waters in conjunction with XRF.

The majority of laboratories require instrument detection limits to be an order of magnitude below their working detection limit. Detection limits for preconcentration methods are difficult to assess owing to the range of XRF instrument configurations available. Concentration from volumes can vary from a few hundred to thousands of milliliters. The ultimate performance of any instrument and preconcentration method must therefore be based upon the minimum mass that can be detected on a filter, measured in micrograms present. Once this figure is determined then the volume required to achieve a measurement significantly above the detection level can be calculated. Figure 2 gives a summary of the typical range of detection limits quoted for XRF instruments. Note: where the literature uses a mass per volume measurement the data have been converted into micrograms present.

XRF instrumentation has improved over the years in terms of count-rate capability and computing power. The peak-to-background ratio that determines ultimate detection limits is controlled primarily by the sample matrix. For this reason, little improvement has been made in achieving lower mass detection limits on filter media since the first publications on these techniques.

4.2 Nonhazardous Liquid Waste

Nonhazardous liquid waste includes materials such as sewage sludge and agricultural sludge used for land improvement and industrial discharges to waterways. These materials are classified as nonhazardous, but they can contain high levels of heavy, potentially toxic metals, and therefore require monitoring.

Traditional methods of sample preparation used for sewage sludge or other materials containing high levels of organic matter are to dry and then ash at 500 °C to reduce the bulk and to concentrate metals for subsequent analysis. This process has a large drawback in that volatile metals such as mercury, cadmium, thallium, selenium, etc. can and often are lost in the process and that it can take many hours.

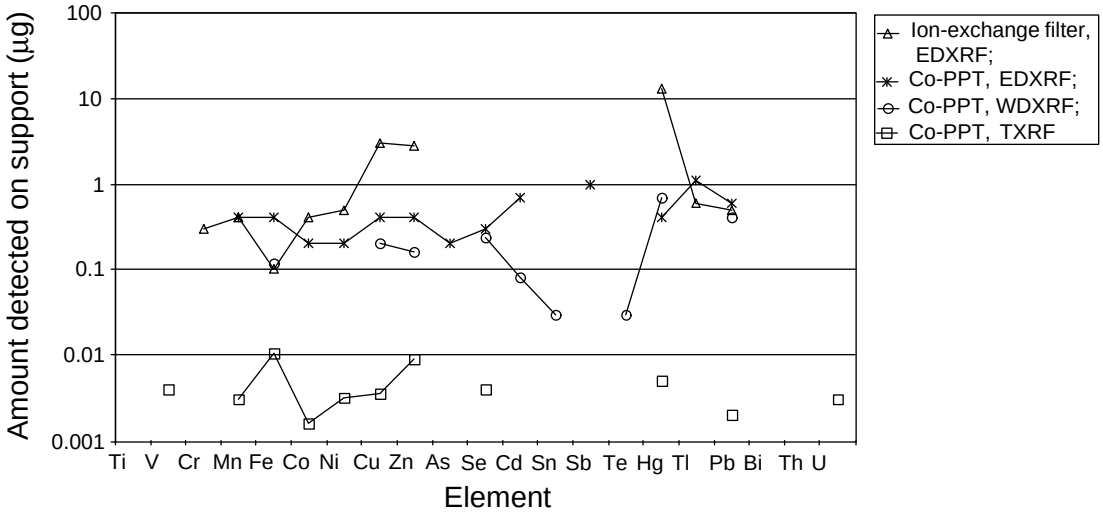

Figure 2 Detection limits for preconcentration methods.

Direct analysis of sludge material requires first that the sample is homogeneous and that it will remain stable during the time of analysis. Sludge material is difficult to stabilize during the analysis times typically employed by XRF. The solid phase will often settle out during an analysis. However, for rapid qualitative scanning or quantitative screening this may not be a problem for heavier elements such as those required for monitoring.

For samples containing high levels of organic solids the sample preparation is usually as described in section 3.1. These samples can be split into two components, i.e. solid and liquid, with direct analysis of the solid and either direct analysis of the liquid or use of a preconcentration technique to achieve appropriate detection of the elements of interest. XRF provides a number of analytical schemes that can be employed for this type of analysis. The ability of XRF to measure samples in their ''as-received'' form provides some distinct advantages to the analyst.

Supernatant or liquid-phase and discharge waters can be treated using the scheme for clean water analyses but may have a high organic content which can cause inteference. However, the requirement to achieve very low detection limits for this type of analysis is not as critical as it is for potable drinking and natural water. Direct analysis of the liquid phase will, in many cases, achieve an acceptable result. The liquid is placed directly into a liquid sample cell incorporating a suitable support film.

Nonhazardous waste in the form of solvents, oils and organic–aqueous mixed solutions often require analyses at very low total metal concentrations. For these type of analyses the preconcentration methods employed for aqueous solutions are not applicable. Organic liquids and solvents requiring determination at levels near or below the routine detection limit of XRF will need preconcentration. The use of chromatographic techniques such as liquid–liquid extraction can be used but in general this area of analysis is not an XRF application.

4.3 Liquid Hazardous Waste

Industrial needs for the measurement of LHW for both regulatory and disposal process monitoring requirements has led to the development of two dedicated XRF methods. Both methods were developed to provide cost-effective alternatives to more traditional ICPAES and AAS methods. These methods are promulgated through the ASTM and are designed to analyze fuel-quality waste (FQW) used principally in the cement industry and similar types of material. The methods are ASTM D5839[21] (approved in 1995) and D6052[22] (approved in 1997).

- The ASTM D5839 method is designed for the analysis of metals only in LHW that can be concentrated through the evaporation of the solvent/liquid fraction. A sample is evaporated to dryness and then mixed with high-purity carbon. Calibration is carried out using certified oil standards prepared in the same manner as unknown samples. Calibrations are linear for the concentration ranges required for regulatory and process control of LHW as an FQW. The method uses dilution into a common matrix, i.e. carbon, to achieve acceptable analytical performance. In this method carbon replaces the light element organic matrix, but the disadvantages of measuring heavy elements in a light matrix are still present and can make this technique prone to low recoveries owing

Table 4 Detection limits for LHW analysis using D6052

Element	Field-portable 3σ LLD ($mg\,kg^{-1}$)	High-resolution 3σ LLD ($mg\,kg^{-1}$)	Element	Field-portable 3σ LLD ($mg\,kg^{-1}$)	High-resolution 3σ LLD ($mg\,kg^{-1}$)	Element	Field-portable 3σ LLD ($mg\,kg^{-1}$)	High-resolution 3σ LLD ($mg\,kg^{-1}$)
P	N/A	200	Zn	12	11	Sn		6
S	N/A	100	Se		5	Sb	<4	5
Cl	N/A	100	As	9	5	I		7
V		8	Br		5	Ba		26
Cr	35	8	Mo		1	Hg	12	7
Ni	25	17	Rh		2	Tl		4
Cu	21	11	Cd	<4	3	Pb	8	4

LLD, lower limit of detection.

to inadequate ''infinite thickness'' of the sample with respect to the high-energy X-rays.

- The ASTM D6052 method is capable of element determination from P to U. Using a <200-μm Al_2O_3 powder a LHW sample is mixed at a ratio of 3 : 1 (Al_2O_3 : sample). This mixture is then measured as a loose powder by EDXRF. Measurements are compared with a set of calibrations based upon traceable mixed-element standards as used for atomic emission/absorption spectroscopic calibrations.

The latter technique overcomes the principal problems that have historically limited the effective use of EDXRF in the analysis of liquids containing elements at highly variable concentrations in potentially multiphasic and/or turbid forms.[23]

The principal advantages of the technique are summarized as follows:

- The Al_2O_3 is activated, with most elements determined, especially potentially volatile metals, firmly bound to the support matrix thus reducing the risk of analyte loss.
- The diluting matrix is relatively heavy compared with other matrices used in similar XRF sample preparation techniques, i.e. carbon[21] and lithium tetraborate/metaborate.[24] This reduces infinite thickness problems in addition to reducing the spectral background.
- Calibrations are straight lines within the recommended concentration ranges, these being:
  - 0–5000 $mg\,kg^{-1}$ for metals;
  - 0–5 wt% for P, S, Cl;
  - 0–1 wt% for Br;
  - 0–2000 $mg\,kg^{-1}$ for I.
- Eliminates problems of phase separation and solids settling during analysis.
- The large sample size used, 5 g, increases representative analysis.

D6052 has been investigated by Piorek[25] for use in conjunction with a field-portable XRF analyzer. This gave results within the acceptable level of performance for a screening technique capable of field use, i.e. ±20% relative accuracy.

Table 4 shows the detection limits published for the D6052 technique.[22,23]

The analysis of LHW is an area that XRF has made many attempts to conquer. The concentrations of prescribed elements are often present at levels easily determinable by XRF. Two measurement techniques can be used to determine these elements:

1. Fundamental parameters (FPs) or standardless analysis. This technique is employed in all commercially available EDXRF and WDXRF instruments. It is based upon the characterization of a matrix in terms of its overall matrix composition and theoretical interelement X-ray line relationships. The XRF literature provides many examples of the use of FP; however, each instrument manufacturer has its own interpretation and exact method of set-up. Guidance from individual instrument manufacturers should be obtained in assessing the suitability of this technique for the elements of interest.
2. Calibrations using certified aqueous and organic/oil standards are the favoured technique used with the standard methods described above.

4.4 Environmental Samples Related to the Nuclear Industry

The determination of long-lived radionuclides and those of low energy, such as uranium, plutonium, thorium and technetium, have been reported to be suitable for analysis using XRF. McMahon[26] describes the various merits of different XRF instrumentation for their analysis.

4.5 Novel Applications

The use of XRF in the analysis of environmental samples is one of practical application. Most researchers use or have developed techniques that can be readily applied. Of a more esoteric nature is the use of synchrotron radiation with EDXRF and the use of energy-dispersive X-ray (EDX) analysis with microprobes. Kaplan et al.[27] used synchrotron excitation EDXRF and EDX to study the transportation of metal contaminants held in the mobile-solid (colloidal) phase within interstitial pores between particles of soil from a contaminated site. The techniques used provide further examples of environmental application of XRF although few researchers have access to such facilities.

Ultrathin-film supports (0.15 μm) have been developed for the analysis of parts per billion levels of elements in small samples (5–50 μL).[28] This method is applicable to very clean aqueous solutions containing no more than 100 ppm total element content (Na–U), hence its usefulness in many environmental applications is limited.

5 INTERFERENCE

Interference present when using methods of preconcentration or advanced methods of preparation, i.e. not ''as-received'' type analysis, can be separated into two main areas: chemical interference which occurs during the preparation of the sample and instrumental interference that is specific to the XRF technique.

5.1 Chemical Interference

Chemical interference usually occurs owing to the inability of all elemental species present within a sample to be quantitatively extracted/concentrated under a single pH regime. Techniques cited in this article provide information on the selectivity of elements and their species, in addition to typical recoveries achieved from the analysis of real samples. Sample preparation methodologies cited detail those elements and valency states that are not selected owing to incorrect pH or other reasons.

Depending on the method employed and the element range to which it is suited, any chemical interferences described in the literature must be accounted for. The main influences or limitations are due to pH. It is generally not possible to recover all valency forms for the range of environmentally sensitive elements in one preparation. Although some preparations are more specific and will capture (recover) differing sets of elements, it is recommended that methods providing a wide elemental selectivity be used. Tables 1–3 show preparation methods and a range of elements and valency states each is capable of selecting.

It is not possible to give a complete listing of influences observed using ion-exchange papers as these are continuously being improved. Any practitioner selecting methods of preconcentration using ion exchange should obtain advice and information from instrument manufacturers and suppliers of such exchange papers. Information provided by manufacturer product listings should be independently verified prior to use.

5.2 Instrument Interferences

Instrument interferences are those that occur between elements at the atomic level, i.e. interelement, or as a limitation of the instrument used to separate and quantify the individual element data from an XRF spectrum. Interelement effects are of two types, absorption and enhancement. The dominant effect is absorption. The majority of instrument manufacturers provide details on these influences and tables or software to show their relationships. Figure 3 shows a spectrum of Cd, Sn and Te and the Cd absorption edge position. In this example Cd is capable of absorbing Te as its absorption edge is just below that of Te Kα lines; however, Cd will only absorb Sn Kβ as its K absorption edge is between that of the two Sn K series lines. For trace element analysis in waters these interelement influences are usually very small and can be ignored. When using filter papers the sample can be considered as being infinitely thin, i.e. the corrections required are based on thin-film techniques where absorption and enhancement effects are also considered to be negligible. Thin-film techniques assume no or insignificant interelement effects. These effects are

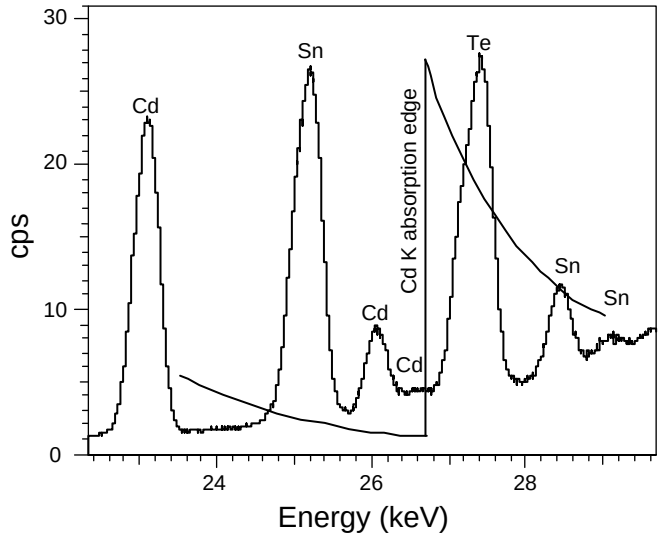

Figure 3 Interelement effects.

applicable to all XRF techniques, i.e. WDXRF, EDXRF and TXRF.

There are a number of ways in which data can be extracted from EDXRF spectra:

- Using a region of interest (ROI): this technique is suited to elements that are not overlapped by others, or the overlap is minimal. This method of data extraction is the most accurate and will give the highest precision of measurement. For more complex spectra, using a set of regions and overlap factors, data can be extracted. This is similar to the use of ROIs, but will allow more complex overlaps to be resolved. However, highly complex spectra where many overlaps exist will be poorly resolved when using this type of spectrum processing.
- Pure element spectrum lines: pure element spectrum lines are used to fit to a spectrum using various peak-fitting mathematical models, e.g. least-squares fitting. This process requires the setting up of the instrument with a set of pure element spectra, either measured from real samples or computer generated. Each line will be needed to fit to the spectrum to allow deconvolution to work. This method potentially will give the best element resolution of highly complex spectra. The precision of measurement is in both cases (ROI or element spectrum), influenced by the number of overlaps being corrected or the element lines active in the spectrum processing. Pure element spectral lines for L series spectra will need to be made from thin-film samples if they are to be used on filters. Variations in interline sensitivities between thin and bulk samples will change. For this reason, many manufacturers supply spectral lines that are split into their sub-line series, i.e. Lα, Lβ, Lγ series.

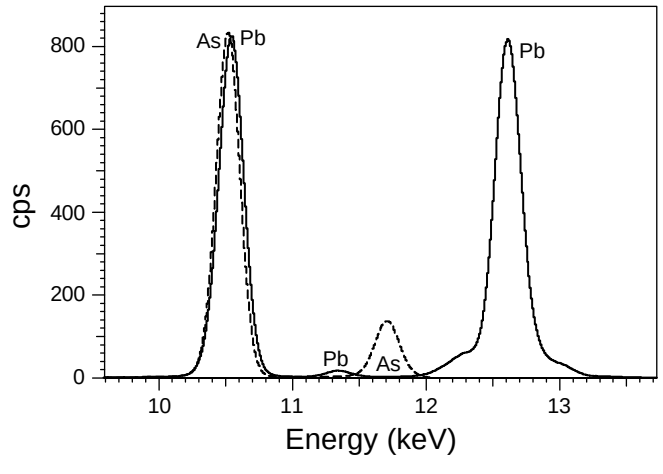

Figure 4 Spectral overlap.

For EDXRF instruments, the main interference is spectrum line overlap. This occurs when one element's line series overlaps that of its neighbors, or a low-energy line series of a heavy element overlaps the line series of a lighter element, e.g. L line overlapping K line spectrum. Figure 4 shows an example of this problem using the classical As Kα and Pb Lα line series overlap. Modern spectrometers use complex mathematical spectrum deconvolution algorithms to extract data from a spectrum. Owing to the use of mathematical algorithms, errors in spectrum processing can be high. The greater the spectrum line overlap, the greater the potential error will be.

WDXRF uses either scanning or fixed monochromators. For fixed-channel systems (simultaneous), the spectral information can only be corrected for overlap if the overlapping element is also present as a fixed-channel. For scanning instruments (sequential), extra scanning points can be programmed to measure overlapping elements and overlap factors applied.

6 ANALYSIS TECHNIQUES

XRF provides two ways of quantification. The first uses a suite of matrix matching standards to produce a calibration line for each element of interest. The second uses a technique specific to XRF referred to as FP or sometimes known as standardless analysis.

6.1 Calibration Techniques

Calibrations can be formed in two ways. The first is to select a group of standards to cover the range of matrix variation and element concentrations to be determined. This can often result in many standards being required. For environmental samples/standards these are often not available. Other methods of producing standards have been used; these are based on methods used for other instrumental techniques, e.g. standard additions, spiking real samples and producing synthetic standards with matching matrices.

Another form of standards calibration is to reduce all standards and samples to a common matrix form. This is essentially what is carried out for atomic emission spectroscopy when samples are usually reduced to an aqueous phase. In XRF the requirement to digest a sample is absent. Instead, a sample can be mixed with a solid medium and either fused into a glass bead or mixed in some way to form a powder. The advantage of these latter forms of standards calibration is that synthetic standards can then be employed widely with no matrix variations existing between standards and samples.

The analysis of environmental samples demands a lot of thought from an analyst. In terms of waste, by definition,

there is no such thing as a standard waste. For this reason, the analyses of waste materials requires calibration to be carried out on samples that can be reduced to a common form. In the last 5 years a number of methods have been promulgated to provide the analyst with ways of calibrating XRF instruments for the analysis of a wide range of liquid waste matrices.[21–23]

Any analysis carried out using these methods should be validated on the basis of final data quality objectives (DQOs) using protocols formulated by organizations such as the USEPA and ASTM. The determination of volatile toxic metals should be given great emphasis when deciding on a method of sample preparation. Any heating of the sample must be considered as a possible source of analyte loss and should be avoided.

Matrix correction techniques used in XRF are shown below. Different manufacturers of XRF instruments provide different solutions depending on instrument geometry. The discussion below attempts to provide an overview of the available techniques used:

1. measurements of high-energy background (Compton);
2. reduction of background using polarized X-rays;
3. use of constant mass for samples;
4. use of internal standard.

One of the most common methods of matrix correction is the measurement and subsequent ratioing of an analyte line to the characteristic Compton Kα anode line in the spectrum. The Compton anode line is generated by the scattering of the X-ray tube anode lines (present in the primary excitation spectrum generated by the tube) scattering off the sample and being detected in the detector. This scattering process is one of energy loss, thus the Compton tube anode line is observed in the spectrum at a point of lower energy than that of the anode energy. The magnitude of the Compton peak is inversely proportional to the average density or atomic weight of the sample. The gross integral Compton line can be used to correct for matrix effects from the position of the Compton line down through lines in lower energy until the first major element line is reached. In many matrix types this is usually to the iron absorption edge at 7.11 keV.

Using polarization to reduce background is highly beneficial in achieving low detection limits for low-Z matrices and improving spectral processing. A drawback, however, is that matrix information is present in the background and matrix correction is still required. Therefore, instruments employing polarization must generate an ''extra'' spectrum with background, preferably a Compton peak. Typically these instruments require higher power X-ray tubes producing 200 W–2 kW.

In order to reduce the number of parameters which will influence an XRF analysis, samples with low-Z matrices or samples that have varying densities should be weighed to a constant mass. This is particularly important where elements are to be measured in matrices that do not provide infinite thickness with respect to the analyte lines of interest. Figure 5 demonstrates the problem of infinite thickness. The use of constant mass is particularly important when calibrations are to be used.

The use of an internal standard at constant mass relative to the mass of sample is a well-known technique for the correction of matrix effects. This can be particularly useful when dealing with low-Z analytes where obtaining sufficient intensity for background measurements in this region of the spectrum (1–5 keV) is more difficult. A number of international standards for the use of XRF employ this technique (e.g. ISO 14596 for the analysis of S in petroleum products).

When using WDXRF simultaneous instruments, although this type of instrument is unsuited to general environmental analysis, a Compton background channel should be fitted where wide matrix variations are expected.

EDXRF and WDXRF can also employ empirical and theoretical ''alpha'' models based on measured intensities and standard concentrations, respectively. These are used to correct for interelement effects where the matrix is relatively well defined in terms of the major elements and where sufficient standards are available.

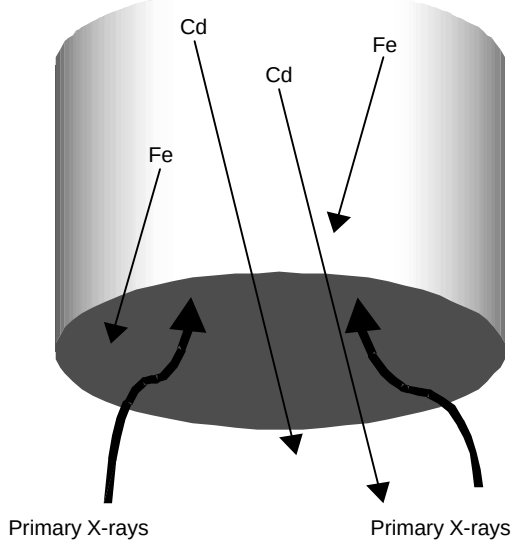

Figure 5 Infinite thickness. The sample is infinitely thick with respect to Fe but not Cd, i.e. the sample is capable of absorbing Fe but not Cd.

Table 5 Standard "liquid" methods using XRF

Name	Organization	Elements	Matrix	Technique
D2622[29]	ASTM	S	Petroleum products	WDXRF
D4294[30]	ASTM	S	Petroleum products	EDXRF
ISO 14596[31]	ISO[a]	S	Petroleum products	WDXRF
D5839[21]	ASTM	Trace metals	LHW	EDXRF
D6052[21]	ASTM	Traces and majors, P–U	LHW	EDXRF
SW846 9075[9]	USEPA	Cl	LHW	EDXRF
SW846 6200[9]	USEPA	Metals	Waste	Field-portable EDXRF
D5059[32]	ASTM	Pb	Gasoline	EDXRF

[a] ISO, International Standards Organization.

6.2 Fundamental Parameter Techniques

Most conventional EDXRF and WDXRF manufacturers provide methods for the analysis of "unknown" samples by means of semiquantitative FP analysis. It is only possible to provide general considerations on the use of this technique as different manufacturers have distinct setting-up requirements and limitations.

Unlike using calibration standards, FP require no matrix-matching standards. However, for FP to calculate all the effects that each element has on another, knowledge of the complete sample matrix is required. As the name suggests, FP relate the physical XRF excitation and detection characteristics of an instrument to the theoretical interelement affects within a sample. This creates a relationship between measured fluorescence intensity of an element in a matrix and its concentration. With the application of corrections such as Compton and/or Rayleigh scattering ratios, for nondetected matrix effects (i.e. organic or aqueous components) a final concentration for each element can be determined. For this reason, the major composition of a sample must be defined either in terms of elements present and/or as a balance. The balance is usually the organic or aqueous component not detectable by XRF.

6.3 International Standard Methods

Table 5 gives a list of the currently available internationally recognized standard reference methods used for the determination of various elements by XRF in liquid matrices for environmental monitoring and regulatory compliance.

Standard methods for the analysis of S and Pb in petroleum products are covered by a series of methods, some of which are prescribed through legislation as mandatory in some countries. Although these methods do not necessarily come under the umbrella of water or wastewater they are designed to ensure that petroleum products meet local environmental legislation and thus are included in this article. All the petroleum methods use calibration techniques of ratio to a background and some, as in the case of ISO 14596, include an internal standard element. These methods are not designed for the analysis of waste material and should not be used for that purpose. Details of the method requirements should be obtained from the appropriate organization.

7 REGULATIONS

Regulations governing the disposal, use and monitoring of waters and wastewater differ from one country to another. For each main section described in this article, i.e. natural water, nonhazardous liquid waste and LHW, different regulations apply.

Table 6 USEPA drinking water regulations

Regulation authority	USEPA	
Date of regulation	1997	
Name of regulation	Safe Drinking Water Act	
Notes	As at 50 μg L^{-1} and Pb at 15 μg L^{-1} currently proposed. Clean sample handling required	
Element	MCLG[a] (mg L^{-1})	MCL[a] (mg L^{-1})
Be	0.004	0.004
F	4.0	4
Cr	0.1	
Cu	1.3	
As		0.05
Se	0.05	0.05
Cd	0.005	0.005
Sb	0.006	0.006
Ba	2.0	2.0
Hg	0.002	0.002
Tl	0.0005	0.002
Pb	zero	0.015

[a] MCL, maximum contamination level; MCLG, maximum contamination level goal.

7.1 Natural Water

The regulations are principally concerned with final drinking water quality. Therefore, this article will concern itself only with those regulations. Any practitioners carrying out analysis of such water should consult their local authority for up-to-date information. Table 6 describes the regulations currently used by the USEPA. This does not represent an exhaustive list of elements but provides an outline of elements required for determination and some typical concentration levels required. Different quality criteria will be used for different world Wide and cross-government authorities [e.g. World Health Organization (WHO) and European Environment Agency]. In general, only TXRF is suitable for the routine analysis of natural waters (including potable drinking water).

7.2 Nonhazardous Liquid Waste

Two areas of analysis need to be considered:

- disposal: this may be into a local river or into the sewerage system;
- the monitoring of such systems.

Unlike methods used for natural waters, samples that fall into this category can be highly variable, i.e. from relatively clean river water, sewage sludge used for agricultural land improvement through muck spraying through to industrial effluent. Table 7 shows a selection of government regulators and regulations indicating elements and monitoring levels required. In general, most regulators are moving away from establishing prescriptive unilateral regulations towards a licensing system that reflects local conditions and final water usage. For a European perspective on waste management with respect to discharge to sewers or watercourses, practitioners are directed to a review carried out by Croner's Waste Management.[33]

7.3 Liquid Hazardous Waste

This area of analysis can also be split into two areas:

- one of assessing the toxicity of any hazardous waste;
- deciding on a suitable disposal route.

The choice of disposal route requires that waste will meet both legislative requirements in terms of maximum

Table 7 Nonhazardous water regulations

Regulation authority	England and Wales Environment Agency	EEC	EEC	UK and North Sea countries	USEPA
Name of regulation	Water Industry Act	Directive 76/464/EEC	Surface Waters Directive SI 1997 No. 2560		Clean Water Act 40 CFR 264 Ground Water Monitoring
Notes	Guide limits	List I (Black list)	List II (Grey list)	Red list – disposal into coastal waters	
Levels:					
Be	×				
V	×				
Cr	×				$0.05\,mg\,L^{-1}$
Ni	×				
Cu	×				
Zn	×				
As	×		25 and $50\,mg\,L^{-1}$ [a]		$0.05\,mg\,L^{-1}$
Se	×				$0.01\,mg\,L^{-1}$
Ag	×				$0.05\,mg\,L^{-1}$
Cd		×		×	$0.01\,mg\,L^{-1}$
Ba					$1.0\,mg\,L^{-1}$
Sn	×				
Sb	×				
Tl					
Hg		×		×	$0.002\,mg\,L^{-1}$
Pb	×				$0.05\,mg\,L^{-1}$
Total[b]	$<10\,mg\,L^{-1}$	0 unless specifically licensed	0 unless specifically licensed	0 unless specifically licensed	

[a] Inland freshwater and coastal seawater, respectively.
[b] Indicates that total prescribed elements shown with × cannot exceed this value.

Table 8 Examples of LHW regulations

Regulation authority	UK Environment Agency	USEPA–RCRA–BIF[a]	
Disposal route	SLFs	SLFs	Underground injection – California
Maximum levels, where cited	Approximate threshold levels vary from 20 to 1000 mg kg^{-1} depending on the toxicity of the element concerned	Approximate threshold levels vary from 20 to 1000 mg kg^{-1} depending on the toxicity of the element concerned	Chlorinated solvents 10 000 mg kg^{-1} As 500 mg L^{-1} Cr(VI) 500 mg L^{-1} Cd 100 mg L^{-1} Pb 500 mg L^{-1} Hg 20 mg L^{-1} Ni 134 mg L^{-1} Se 100 mg L^{-1} Tl 130 mg L^{-1}
Elements	Be, V, Cr, Mn, Co, Ni, Cu, Zn, As, Se, Mo, Ag, Cd, Sn, Sb, Ba, Hg, Tl, Pb	Be, Al, V, Cr, Mn, Co, Ni, Cu, Zn, As, Se, Mo, Rh, Ag, Cd, Sn, Sb, Ba, Hg, Tl, Pb	Cl, As, Cd, Se, Cr, Hg, Tl

[a] BIF, Boiler and Industrial Furnace Regulation; RCRA, Resource Conservation and Recovery Act; SLF, secondary liquid fuel.

pollutant concentration and that any process control requirements are also met. A number of international agreements on the transportation of LHW are now in force. In 1994, the United Nations Basal Convention on Movement of Hazardous waste banned all shipments from the Organization for Economic Cooperation and Development (OECD) countries to those outside the OECD and in 1997 this was extended to include the ban on movement for recycling purposes. In general, material that is considered to be hazardous must be disposed of in its country of origin. Disposal routes for LHW will depend on local legislation and requirements of disposal industries. For these materials only one main form of disposal is available, i.e. incineration. Legislation controlling this process is generally based on a licensing system, that is, incineration facilities are licensed to dispose of waste up to certain concentration levels for each regulatory element. These levels are based upon stack

Table 9 Government environment Web sites

Country	Web site
USA	http://www.epa.gov/
UK	http://www.environment-agency.gov.uk/
European Environment Agency	http://www.eea.eu.int/
France	http://www.ademe.fr/
Germany	http://www.bmu.de/(english/)
Australia	http://www.erin.gov.au/
Canada	http://www.doe.ca/
Japan	http://www.eic.or.jp/(eanet/)
Malaysia	http://www.pop.jaring.my/(jas/)
New Zealand	http://www.mfe.govt.nz/
Taiwan	http://www.epa.gov.tw/(english/)
Singapore	http://www.gov.sg/(env/)
Switzerland	http://www.admin.ch/buwal/d/index.htm
Sweden	http://www.environ.se/
Finland	http://www.vyh.fi/moe/moe.htm
Denmark	http://www.mem.dk/ukindex.htm
Czech Republic	http://www.cizp.cz/
Austria	http://www.ubavie.gv.at/
Ireland	http://www.epa.ie/
Luxembourg	http://www.mev.etat.lu/home.html
Netherlands	http://www.minvrom.nl/environment/home.htm
Northern Ireland	http://www.nics.gov.uk/ehs/index.htm
Norway	http://www.sft.no/
Portugal	http://www.dga.min-amb.pt/welcome.html
Scotland	http://www.sepa.org.uk/
Belgium	http://www.vmm.be/
Iceland	http://www.mmedia.is/umhverfi/english/

emissions but incineration facilities use in-bound fuel specifications to control this process. Table 8 shows a list of typically controlled elements and an indication of regulatory concentration levels. The actual control of elements within this category is based upon their disposal route. LHW use as a SLF is controlled under a user licensing system in the UK and Europe, the contents of which are usually commercially sensitive. Disposal through high-temperature incineration for highly toxic LHW is controlled, from the regulatory standpoint, through stack emission monitoring and individual licensing depending on the capability of the facility. Underground liquid injection, where permitted, is controlled by specific local legislation.

A listing of environmental agencies and their Internet World Wide Web sites is shown in Table 9. These sites offer a range of information and are excellent sources for first contact points in obtaining the latest details and data about local issues and legislation.

8 QUALITY CONTROL AND TROUBLESHOOTING

QC of methods used for the analysis of environmental samples should follow standard good laboratory practice (GLP) guidelines as used for any instrumental analysis technique. Any method developed should provide a history of validation based upon DQOs. The DQO will vary according to the methods usage, i.e. regulatory control, screening only, process control and analytical performance requirements. The use of DQO strategies is described in standards developed through organizations such as the USEPA and ASTM. These are very large documents covering the complete range of waste management activities. Practitioners should limit their research to those sections that define the analytical requirements or capability of a method they wish to employ.

XRF instruments are intrinsically stable when compared with other multielement instrumental techniques. Calibrations are developed as a ''one-off'', with recalibration often not required for years. For this reason, QC requirements are based mainly around establishing and maintaining sample preparation and presentation criteria.

In general, methods developed for use on XRF instruments require two forms of QC monitoring:

1. The instrument should be shown to be stable prior to measurement of any sample. This is carried out by the measurement of a QC monitor. This measurement must be carried out using the same instrument conditions as used for the sample. Comparing the QC monitor concentration with that of the time of calibration (day zero) is usually sufficient to confirm stability. This type of measurement is usually carried out on a daily or shift basis.
2. Using a blank sample and a suitable, either synthetic sample (of known composition), or standard, prepare real samples for analysis. Performance of the blank and ''standard'' in terms of the DQO for each element will confirm the method's suitability for use. These types of measurements are usually carried out on a weekly basis.

If a measurement fails any QC requirement then a number of checks can be made. The following is a brief outline of some of the most commonly occurring problems that can occur assuming the sample has been reanalyzed and freshly prepared to check for obvious errors:

- When using an autosampler, check whether the sample was correctly positioned and measured.
- Check any diagnostic evidence reported with the analysis, e.g. high instrument detector resolution, unusually high errors of analysis, for EDXRF spectrum processing giving poor fitting, indicating a failure of a detector component.
- Sample chamber atmosphere not set correctly, e.g. He gas not turned on or empty cylinder. Observed as a low count rate for light elements but correct for higher Z elements.
- Contamination on X-ray tube or detector components. Observed as a low count rate but all other diagnostic information satisfactory.

For analyses that give results differing from those provided through either standard organizations or from analysis by other techniques, the following tests should be carried out:

- For FP methods a comparison of the elements specified in the analytical method and those observed qualitatively must be made. If a major element, not specified in the method, is observed then it must be added and the sample reprocessed where possible. This indicates that further method development should be carried out.
- When poor detection of minor or trace elements occurs in an EDXRF spectrum, check for the presence of an unexpected major or minor element, not originally specified in the method. This is usually observed by degradation in spectrum processing. This indicates that further method development should be carried out.

- When poor detection of minor or trace elements occurs in a WDXRF spectrum, check that minor or major element spectrum overlaps of trace elements are capable of being resolved by the diffraction crystal used in the monochromator or goniometer, e.g. Ba on Ti or Pb on As. Add any overlap corrections as required in the method. This indicates that further method development should be carried out.
- Observation of elements not expected to be present. When measuring low-*Z* matrices and thin samples such as filters, the high background scatter and poor stopping power of the sample may on occasions excite some of the metals or materials used in the construction of the instrument or any sample holder, e.g. Cu and Zn from brass or bronze and Cr, Fe or Ni from stainless steel. These minor impurities or contaminants can be eliminated by the measurement of appropriate blanks. Consideration of this issue is particularly important for thin samples and those of insufficient infinite thickness.
- The result of a measurement is shown to be incorrect.
    - For calibration methods this may occur if a matrix correction is not appropriately applied. The sample is of a sufficiently different matrix type that the calibration used is inappropriate – this will be apparent if major element and especially Compton peak count rates compared with the calibration are outside the calibration range.
    - For FP analysis the stoichiometry should be matched to the sample. Mixed stoichiometries such as the presence of metal chlorides and oxides or carbonates and oxides are difficult to resolve and can lead to inaccuracies if a metal were present in the chloride form but specified in the method as oxide. Non-detectable elements such as C, H and N present as CO_2, H_2O, etc. need to be calculated using matrix correction procedures such as Compton–Rayleigh ratio or entering information provided from other measurement techniques.

9 METHOD DEVELOPMENT

The development of a method will be dependent upon the type of instrument and the instrument manufacturer's recommendations. The three main types of instruments referred to in this article, i.e. EDXRF, WDXRF and TXRF, require different strategies. The development of a method is split into two components, the sample preparation and the set-up of the instrument – the latter will depend upon the manufacturer's recommendations and type of instrument.

9.1 Energy-dispersive X-ray Fluorescence

Four types of system are in use today:

1. direct or filtered direct excitation of a sample with detection using an Si(Li) detector;
2. direct excitation of different metals targets to produce a secondary excitation to excite the sample with detection using an Si(Li) detector;
3. scattering crystals to produce polarized primary beam excitation with combinations of secondary beam excitation and Bragg reflection crystals to optimize peak-to-background performance using a detection system similar to those above, i.e. Si(Li) detector;
4. either isotope sources or low-powered X-ray tubes with a combination of primary and secondary beam filtration and a sealed gas proportional detector or for hand-held field analysis a silicon PIN diode semiconductor.

Si(Li) detectors are of two main types: those which use liquid nitrogen for cooling and those which do not. Of the latter a number of different types are available, the most common using a Peltier cooled detector. Other types are available but at relatively high cost, e.g. those that use the Joule–Tompson effect for cooling or a Stirling engine.

EDXRF uses a number of key parameters to optimize the excitation of elements within a sample and their subsequent detection. The main components are as follows:

- X-ray tube, anode types of Rh, Ag, Pd, Mo being the most common;
- primary beam filter placed between the tube and the sample;
- secondary target placed between the tube and the sample;
- polarizing crystal scatterer placed between the tube and the sample;
- Bragg reflector between the sample and the detector;
- type of detector – Si(Li), PIN diode or gas (Ne, Ar, Kr or Xe) filled proportional.

The key parameters and combinations used are shown in Table 10.

Depending on the combinations of tube anode excitation voltage (kilovolts) and primary beam filtration, different ranges of elements can be excited. In general the following rules of thumb can be used to classify elements into ranges of instrument excitation parameters. For all forms of EDXRF the following general rules will apply:

Table 10 EDXRF instrument parameters

Excitation type	Range of elements		Excitation conditions	Detection	Detector parameters
Direct excitation	F–Cl	K lines	5–10 kV tube voltage, no primary beam filtration, vacuum or He atmosphere	Si(Li)	Optimum resolution
Direct excitation	K–Zn Sn–Hf	K lines L lines	12–20 kV voltage, Al primary beam filters (of various thickness), air path possible	Si(Li)	Best count rate versus resolution
Direct excitation	Fe–Mo Ta–Pu	K lines L lines	30–45 kV voltage, anode primary beam filters (of various thickness), air path possible	Si(Li)	Best count rate versus resolution
Direct excitation	Zr–Nd (Tb) Th–Pu	K lines L lines	50–60 kV voltage, thick transition metal (Co–Zn) primary beam filter (of various thickness), air path possible	Si(Li)	Maximum count rate
Secondary target excitation	F–Mg	K lines	Metal secondary target Al	Si(Li)	Maximum count rate
Secondary target excitation	Ca–Mn	K lines	Metal secondary target Co	Si(Li)	Maximum count rate
Secondary target excitation	Fe–Rb/Sr W–Pb	K lines L lines	Metal secondary targets Zr or Mo	Si(Li)	Maximum count rate
Direct excitation with special tube anodes	Ge–I W–U	K lines L lines	Typically 50 kV, anode primary beam filter	Use of Bragg reflectors and Si(Li)	Maximum count rate
Use of polarizing targets	Fe–Mo Gd–U	K lines L lines	B_4C + Fe filter, 35–40 kV	Si(Li)	Best count rate versus resolution
Use of polarizing targets	Rh–Pr	K lines	Al_2O_3 + Ta filter, 50–60 kV	Si(Li)	Best count rate versus resolution
Benchtop EDXRF	Mg–Ba Sn–U	K lines L lines	Manufacturer specific tube or source excited, 4 to 25 kV or ^{55}Fe, ^{109}Cd, ^{244}Cm or ^{241}Am isotopes depending on elements required. Combinations of primary filter and absorption edge detector secondary filters	Sealed gas proportional detectors (Ne, Ar or Xe)	Specific detector for elements required
Hand held, isotope source	K–U		^{55}Fe, ^{109}Cd, ^{244}Cm or ^{241}Am isotopes depending on elements required	Silicon PIN diode	Best count rate versus resolution
TXRF, direct excitation	Ca–Ba Sn–U	K lines L lines	Usually 2 tubes, Mo and W, 50 kV, high wattage (~1.5–3 kW), anode primary filters	Large-area Si(Li)	Best resolution. Usually low count rate

Table 11 WDXRF instrument parameters

Excitation type	Range of elements		Excitation conditions	Detection
Direct excitation	Na–In Sn–Pu	K lines L lines	20–40 kV tube voltage, no primary beam filtration, vacuum or He atmosphere	Appropriate crystal, gas proportional (sealed or flow) detector/monochromator
Direct excitation	Ru–Ba	K lines	50–100 kV tube voltage, primary beam filtration, vacuum or He atmosphere	Appropriate crystal, sealed gas proportional or scintillation detector/monochromator
Direct excitation	Na–In Sn–Pu	K lines L lines	20–40 kV tube voltage, no primary beam filtration, vacuum or He atmosphere	Variable multiple crystal/slit goniometer, scintillation counter detector
Direct excitation	Ru–Ba	K lines	50–100 kV tube voltage, no primary beam filtration, vacuum or He atmosphere	Variable multiple crystal/slit goniometer, scintillation counter detector

1. Tube voltage required for an element:
   (a) element absorption edge kV × 1.5 if no primary beam filtration is used;
   (b) element absorption edge kV × 2 if primary beam filtration is used.
2. Primary beam filter element used:
   (a) non-X-ray tube anode material filter – absorption edge of filter (e.g. Al) less than one-third of the voltage applied;
   (b) anode filter – only when the voltage used is greater than the K absorption edge of the X-ray tube anode material and when the tube K lines are required for excitation;
   (c) thick transition metal filter used for elimination of tube anode K lines and background continuum in the energy range 15–30 kV.

9.2 Wavelength-dispersive X-ray Fluorescence

Two types of system are available. Both use an end-window, usually high-power, X-ray tube, typically Rh, to excite the sample:

- sequential WDXRF: this uses a series of Bragg diffracting crystals mounted in a goniometer to scan the available wavelengths to detect analytes present;
- simultaneous WDXRF: this uses the same excitation but a series of fixed monochromators each mounted with a Bragg crystal at an angle capable of receiving a single-analyte X-ray line.

The key parameters used are shown in Table 11.

For WDXRF, the excitation conditions are usually set to the maximum X-ray tube voltage with detection of elements relying on the selectivity of the monochromators or goniometer used. Primary beam filters and selective voltage can be used and the rules that apply to EDXRF are equally applicable here. Variation of the wattage or X-ray flux hitting the sample may be important for volatile matrices such as organic solvents and it is common to reduce the current on high-powered (3–4 kW) WDXRF instruments to prevent sample damage.

9.3 Direct "As-received" Analysis

To develop a methodology for the direct analysis of liquids (e.g. aqueous, organic solutions and sludges) an understanding of the effects of matrix form in terms of primary excitation and detection is required. Low-Z matrices generate high-background scattered radiation. Conversely, high-Z matrices produce little background scatter. The excitation and subsequent detection of high-Z elements in low-Z matrices is very efficient. Achieving the best detection limit requires minimizing the scattered background produced by the low-Z matrix and maximizing the sensitivity of the high-Z element(s) of interest.

The sensitivity of heavy (high-Z) elements in light matrices is dependent upon the density and thickness of the sample measured. The infinite thickness of a sample, with respect to the X-ray lines of the elements of interest, must be controlled or in some way measured in order to obtain reliable quantitative analysis. Figure 5 and Equations (1) and (2) show the effect of varying infinite thickness or sample depth (thickness) and its potential affects on quantitative results. In Figure 5 it is clear that the sample can absorb Fe K but not Cd K X-rays (i.e. Cd K X-rays are capable of passing through). This is referred to as the infinite thickness of a sample with respect to an element. To measure quantitatively the amount of Fe and Cd present in the sample it is important to realize that the effective volume or, as the diameter of a sample is usually constant, thickness of sample measured by the detector for each element is different. This is due to the different energies of the fluorescence X-rays emitted from each element. The effective volume/thickness of sample measured for each element is controlled by the density of the sample matrix and its ability to absorb characteristic fluorescence X-rays from the elements present. If the potential volume/thickness that an element can be measured from is greater than the volume/thickness of sample then the relationship between fluorescence measured and element concentration is no longer constant. The fluorescence measured will increase up to the point of infinite thickness for each element present. Therefore, if a sample is not infinitely thick with respect to the elements of interest, errors will occur in any quantitative measurements.

Equation (1) defines the relationship between infinite thickness and analyte element:

$$I = I_0 \, e^{-\left(\frac{\mu}{\rho}\right)\rho t} \tag{1}$$

where I is the absorbed radiation, I_0 is the incident radiation, μ is the mass absorption coefficient, ρ is the matrix density and t is the matrix thickness.

For practical purposes this equation can be solved for t to show the following form (Equation 2):

$$t = \frac{6.9}{\mu\rho} \sin\theta \tag{2}$$

where θ is the takeoff angle between the sample plane and the detector.

The effects of infinite thickness are of particular importance to the analysis of low-Z matrices such as water, wastewater and LHW. The majority of elements

of interest for environmental analysis in waters and wastewaters are those of the transition metals and elements with $Z > 30$. The typical volumes of sample measured in spectrometers are in the region of 5–15 mL. These samples are presented to the instrument in sample cups of typical height less than 3 cm. In these matrix types the transition metals and heavier elements are not infinitely thick with respect to the solution matrix. The presence of lighter elements such as the alkaline earths and dissolved calcium reduces the infinite thickness for the typically required environmentally sensitive elements. However, it is often the case that samples presented in this manner will require some measurement that relates infinite thickness (or lack of) to that of the concentration of heavy elements present.

For samples where infinite thickness is a problem the following techniques of sample preparation can be employed:

- Weigh samples to a constant mass into sample cups. This is a particularly useful technique especially if samples vary in density.
- Add an internal standard (element). The element chosen must be affected by the matrix and infinite thickness problems in the same way as those elements to be determined. A downside to this technique is the additional stage required and that, typically, elements that can be used are elements of possible interest.
- Reduce the sample size such that the effect of infinite thickness is of no consequence. Typically, a thin layer of sample is used. This can be useful if the sample size available is very small; however, the sensitivity for elements of interests will be significantly reduced. High-energy background in the range 20–50 keV will dominate spectra where the sample is supported by a low-Z matrix, e.g. cellulose filter paper. This limits the method unless instruments are equipped with features that reduce this background component, e.g. polarized X-rays. Counting times will need to be increased to overcome reduced sensitivity.
- Add a heavy element to reduce infinite thickness – referred to as adding a heavy absorber. This is similar to above but the amount used is significant and will dilute the sample by a factor of two or more. This will reduce sensitivity and degrade detection limits. This method is better suited to WDXRF than EDXRF instruments.

For samples where infinite thickness is a problem the following instrumental techniques can be employed:

- Calculation of effect of infinite thickness by inclusion of sample thickness in the analysis equations. This is of particular importance if an FP method of analysis is used.
- Measurements of a section of the spectrum that can be used to correct for lack of infinite thickness, e.g. X-ray tube Compton and/or Rayleigh lines.

The effects of matrix-induced background can be controlled by both sample preparation techniques and instrumental components. The only method of sample preparation used to reduce the effect of background scatter is to add a heavy element to the sample. This has the effect of reducing background scatter and infinite thickness depth with respect to heavier elements that may be required for analysis. This does, however, add complexity to the sample preparation and reduces the measuring precision owing to the errors associated with the additional stages of preparation.

Recent improvements in EDXRF spectrometer design have come from a need to reduce the background component. This has led to the development of devices to reduce both background beneath spectrum lines and high-energy scatter. A number of papers describing spectrometers using these devices have been published, e.g. by Yokhin and Tisdale,[34] Haschke et al.[35] and Heckel.[36] As the detection limit requirements for environmental analysis are being pushed lower, this area of research has been shown to produce the most benefits. The effect of using such devices is to maximize sensitivity and minimize background. Using a polarized primary beam, the background present under an element peak can be reduced with no significant sacrifice to the sensitivity as long as sufficiently high-powered X-ray tubes are used. Heckel[36] used a primary beam scattering device to ''focus'' excitation close to elements of interest, thus increasing the sensitivity over a narrow region of the spectrum. This reduces the amount of unwanted X-rays (usually background continuum) entering the detector. Thus detectors of moderate count-rate capability can be used efficiently for the collection of spectral data, i.e. for those elements of environmental interest.

- For EDXRF instruments, increasing the power of X-ray tubes or the count rate of detectors will improve detection limits, but usually at a financial cost and only by small increments. For any increase in X-ray tube power, a corresponding improvement in the detection system is also required. This is a consequence of using one detector, having a finite count-rate capability, to collect all spectral data. In general, one element cannot be selectively detected differently from its neighbors.
- For WDXRF, elements are detected selectively. Detectors are capable of millions of counts per

second, so at detection levels this performance is not limiting. Simply increasing the count rate by increasing the X-ray tube power will have a corresponding effect on the LLD, i.e. doubling the count-rate will produce a $\sqrt{2}$ improvement on LLD. Typically increasing X-ray tube power is not difficult or financially a burden for this type of instrumentation; however, sample damage or destruction will occur if the tube power is too high and will ultimately be a limiting factor.

For analysis of environmental samples using WDXRF, the sequential instrument is generally the one chosen. This is due to the number of elements required and provides the maximum flexibility.

Presentation of the sample to these instruments must be done with care when using ''as-received'' liquid or semiliquid samples. Reliance on a thin X-ray film to support these types of samples must be a consideration when using high-power excitation conditions. Unlike EDXRF, WDXRF instruments are very difficult to clean if a sample should leak into the sample-measuring chamber. WDXRF instruments are provided with safety windows to minimize liquid spilling into the spectrometer.

9.3.1 Choice of Sample Support Films

There are many support films commercially available and also films supplied by individual XRF instrument manufacturers. The choice of film is based upon two criteria:

1. elements to be determined;
2. chemical resistance required.

Unfortunately it is often the case that these two requirements are mutually exclusive; however, for typically required environmentally sensitive elements, i.e. $Z > 20$, a polyester-based film is clearly suitable. This type of film is available through any XRF consumable supplier and a number of instrument manufacturers in various thicknesses (usually 4–6 μm) and sizes.

Where light elements are required, e.g. P, S and Cl, a polyester-based film has some drawbacks:

- the oxygen content of such films attenuates the X-ray beam;
- the film produces a high background.

In this case a polypropylene film provides superior X-ray transmission and lower background over polyester and is chemically resistant to most aqueous and organic solvents encountered. However, experience has shown that leaching of halogenated solvents through polypropylene film will occur, especially for aromatic compounds. This is usually seen by a relaxation and crinkling of the film, which is exacerbated as the temperature within the sample chamber increases. The general-purpose use of this film is not recommended for this reason and because it is significantly more expensive than any other.

Before a sample support film is used it should be assessed. Leaving a sample supported by the film, on blotting paper, for not less than its expected residence time in the sample chamber will confirm its suitability. If the sample leaks, as evidenced on the blotting paper, then an alternative film must be used. It is necessary to evaluate all possible sample types before making a final choice as a change of film may invalidate a calibration. Instrument manufacturers and suppliers of X-ray sample support films provide valuable information on the suitability of films.

9.3.2 Calibration Methods

Where elements can be detected at measurable levels within natural and wastewaters, and generally in the analysis of LHW, a method using a set of calibration standards can be employed. This is usually restricted to the analysis of LHW, as the lower limits of quantitation are relatively high. Direct ''as-received'' analysis of natural and nonhazardous wastewater often requires limits of quantitation below the detection limit of the technique.

Pure and mixed analyte calibration standards are available through any reputable chemical supplier. These can be used to produce a synthetic set of standards to calibrate a spectrometer. For simple matrix types such as river water or nonhazardous liquids, only waste calibration standards can be used, with an appropriate correction for matrix variation such as infinite thickness.

For the analysis of LHW, synthetic calibrations based upon pure or mixed analyte standards are rarely satisfactory. The methods developed under the auspices of ASTM, e.g. D6052 and D5839, should be used.

Filter cake material present as solid residue from industrial filters and pressed into pellets can be measured against a calibration. Typically this uses soil and other geological standards. Calibrations should be set up to correct for interelement effects and provide a means of correcting for overall matrix effects using a ratio to Compton peak for elements in the Fe–Mo (K series lines) and Pr–Bi (L series lines) range. Practitioners requiring this type of analysis are advised to treat samples as soils and review those XRF methods that apply.

9.3.3 Fundamental Parameter Methods

The analysis of LHW, using the FP approach, has been shown to be successful by a number of manufacturers.

This relies on an accurate measurement of the sample matrix, typically organic, using proprietary software and instrument-specific algorithms. The ultimate performance of such techniques is still a matter for some debate within the XRF scientific community. The use of this technique is clearly an advantage where standards cannot be synthesized or obtained. The speed and adequate detection limit capability of XRF for the measurement of LHW have already made this method of analysis acceptable within an industrial environment. The ±20% accuracy requirements for LHW make the use of FP almost ideal for their analysis. When using FP analysis for LHW samples it is important to specify stoichiometry. For the analysis of most liquids and sludges it is usual to represent all elements in there elemental form. The balance or mean Z of the matrix is then represented as H_2O or C. However, for ''as-received'' samples, settling of particulates and separation of multiple liquid phases during analysis will remain a problem and limit the success of any such measurement. The analysis of filter cake in the form of pressed pellets or dried solid requires a different approach to FP analysis. These sample types are usually given a stoichiometry for the major elements as stable oxides. The water content is then determined by ashing and added to the analysis as a fixed H_2O content. The benefit of separately determining the water content is that the errors produced within an analysis due to the need to estimate the nondetected H_2O fraction are removed. The sample is measured ''as-received'', thus avoiding potential volatile element losses.

The ability to re-process XRF data to obtain more accurate results after the addition of non-XRF detected element data is a method that can be used in a FP analysis.

The use of the FP approach for natural and nonhazardous wastewater analysis is of little appeal. Performance criteria make such methods difficult to validate and with the ease of producing standards for use in calibrations, FP analysis for such matrix types is of little use.

9.4 Analysis Using Total-reflection X-ray Fluorescence

Unlike conventional EDXRF, TXRF relies on no interelement effects existing. For this reason, conventional bulk sample FP-type analysis is not applicable. Once the liquid sample/standard is dried on the analysis support (quartz or Perspex™/Plexiglas™), all samples will appear essentially the same to the instrument. Calibrations are developed using internal standards.

The reasons for addition of an internal standard during sample preparation are twofold, first to ensure that the sample recovery is within acceptable levels and second to act as a reference point for sensitivity corrections and subsequent concentration calculations. A reference point element, typically Co for transition metals analysis, is used in the following way:

- The sensitivity relationship between the reference element and neighboring elements measured under the same instrument conditions is precalculated, usually as part of instrument setting-up routines.
- The intensity versus concentration relationship of the reference element is defined in terms of the original concentration in the sample, often micrograms per kilogram or parts per billion, not its concentration or mass on the support. When measuring a sample containing the reference element it is then possible to calculate the concentration of the other elements present using the sensitivity information.

The above scheme allows complex samples containing many elements to be measured without the need to provide large numbers of standards.

9.5 Analysis of Filters

Using the filters prepared in any of the ways referred to in section 3, conventional EDXRF or WDXRF can be employed to carry out the final determination. Using a set of conditions such as those described in Tables 10 and 11 or as recommended by the instrument manufacturer, the environmentally sensitive elements can be measured.

As the samples are relatively homogeneous and of the same dimensions (i.e. the same filter papers are used for all measurements, typically 25- or 47-mm diameter) analysis is carried out by reference to a calibration developed from a set of reference standards. Standards can be either purchased through standards reference organizations, of which a selection of those providing environmental standards are shown in Table 12, or manufactured to individual requirements.

Using certified standard solutions (purchased through any reputable chemical supplier) containing the elements of interest at concentration levels that can be preconcentrated on the filter medium used, a series of standards can be made. First a series of solutions containing a mixture of the elements required is produced. These solutions are then preconcentrated using the technique chosen for sample preparation. The concentrations of the solutions used should be such that when preconcentrated on the filter the series of filters will hold masses for each element that will span from the detection limit to an estimated value representing the highest mass expected. It should be noted that the sample volume used can be varied to provide higher levels of preconcentration if lower levels of detection are required.

Standards prepared from solution should be tested for extraction recoveries before using. This can be done using

Table 12 Standards organizations

Name of standards organization	Contact address
NIST (National Institute of Standards and Technology)	100 Bureau Drive, Gaithersburg, MD 20899-0001, USA
Conostan (Continental Oil Company)	Conostan Division, Conoco Specialty Products, Inc., 1000 South Pine, P.O. Box 1267, Ponca City, OK 74602-1267, USA
LGC (Laboratory of the Government Chemist)	Queens Road, Teddington, Middlesex TW11 0LY, UK
National Institute of Occupational Health	P.O. Box 8149, Dep, N-0033 Oslo, Norway

radiotracers as described by Prange et al.[14] or through calibration validation using certified water standards.

Typically the filtered area of paper is kept as small as possible but not so large that the filtrate mass will build up to produce interelement effects within the sample. Typically masses in the tens of milligrams range spread over 25–47-mm diameter filters should not be exceeded.

Calibrations are linear across the range of concentrations required for this analysis. Corrections for any interelement effects are not required.

The analysis of unknown samples must include volume data. The analysis of a filter will produce mass data. If the volume of the sample is the same as the volume used for calibration, then the result is directly comparable, e.g. 100 μg measured on a filter will equate to 100 μg/volume of calibration standard. If the volumes of calibration standards and samples are different, then the measured sample mass must be divided by the sample volume. This will provide a corrected concentration in terms of mass/volume.

When measuring filter papers it is vital that both the area excited by the X-ray beam and that seen by the detector cover the filtrate. Failure to achieve this would result in lower than expected count rates for the elements present and thus low recovery figures.

Filters must be held in such a way as to ensure that the entire filter surface is measured. It is often the case that the sample is not homogeneously distributed across the filter. As the cross-section of the X-ray beam does not have a uniform field of flux the excitation efficiency will vary. For this reason it is routine to spin filter samples, thus ensuring that the variations in filtrate density and X-ray beam flux are normalized. For filters larger than the X-ray beam a spinning mechanism that rotates the sample offset from the X-ray beam is required. This will ensure that the entire surface of the filter is covered, producing an average signal. Variations in filtrate density are equally accounted for using an offset spinning mechanism.

The only correction required for filter analysis, assuming the filtrate mass does not generate significant interelement effects, is a ratio to Compton for elements in the Fe–Mo K series lines) and Pr–Bi (L series lines) range. This will correct for variations in the overall matrix.

For the analysis of elements with $Z > 20$ (Ca), an air path environment can be used with EDXRF. WDXRF requires vacuum or for $Z > 20$ a helium atmosphere if samples are to be measured ''damp''.

10 COMPARISON WITH OTHER SPECTROSCOPIC METHODS

10.1 Comparison of Technology

Two main techniques, AAS and ICPAES, dominate elemental analysis. XRF competes with ICPAES for the analysis of environmental samples for the appropriate analysis of elements with $Z > 11$. Ion chromatography (IC) for the analysis of S and halogens is also widely used in environmental analysis and is particularly used for fluorine analysis in liquid samples for which XRF is not suitable.

ICPMS is used for ultratrace-level analysis (micrograms per kilogram), particularly of potable drinking water; however, this level of performance is not required for nonhazardous and hazardous liquid waste. TXRF provides comparable performance although it is limited to the analysis of elements typically with $Z > 20$.

An advantage over spectrophotometric methods such as AAS and ICPAES is XRF's ability to measure the halogens and S in addition to prescribed trace contaminants such as heavy metals. In LHW materials halogens and S compounds are common. At the levels of analysis required by industry for incinerator process control or determining disposal routes, XRF provides a rapid semi-quantitative to quantitative method for S and halogens.

Pedersen and Higgins[37] and Kendall et al.,[38] showing comparisons of EDXRF versus IC, ICPAES and AAS (flame and furnace) techniques for the analysis of

LHW, demonstrated the advantages in using EDXRF methodologies for the regulatory and process control conformance analysis of FQW.

The remaining environmentally sensitive elements not able to be determined by XRF in liquid samples, i.e. Be and B, will drive laboratories to continue to use AAS, ICPAES and ICPMS. However, the cost savings and speed of analysis provided by XRF will tend to minimize the need for every sample to be measured for these light analytes.

Where XRF has a role to play is in the rapid routine analysis of toxic LHW where levels are high enough to perform close to ''as-received'' analysis and still provide analytical techniques that will allow one-off measurements of clean and nonhazardous liquid waste and water.

The analysis of halogens, P and S by IC produces problems for analysts in that sample sizes are very small (typically < 0.2 g). Obtaining results from one such sample and relating its significance to what is usually a heterogeneous bulk waste such as a 40 000-L tanker is difficult.

10.2 Business Issues

The LHW industry segment has seen rapid growth, mainly for EDXRF instruments. The combination of both qualitative and quantitative analysis, capable of measurements within 30 min, is replacing the more traditional methods of AAS and ICPAES within the European and non-US market. Owing to the more ''methods-regulated'' and controlled industry within the USA this change has been slower.

10.3 Regulatory Issues

Specific methods, often prescribed for regulatory requirements, means that ICPAES, ICPMS and AAS, with their superior detection limits, will continue to dominate the ultratrace-level analysis of contaminants in drinking and natural waters. Although TXRF provides the same capability for most elements of interest, lack of internationally recognized methods limit its appeal.

LHW analysis is driven by performance-based methodology where ultralow detection limits are not required. Here, speed, accuracy of $\pm 20\%$ and low cost are the requirements of what is essentially a screening process.

11 FUTURE DEVELOPMENTS

11.1 Ultratrace Analysis Using Microsample X-ray Analysis

Microsample X-ray analysis (MXA™)[28] is a new technique developed initially for the silicon wafer fabrication industry where ultrapure acids and washing waters required rapid cheap analysis techniques. The MXA™ technique uses an ultrathin film, typically 0.15 μm thick, to support dried 5–50-μL samples for analysis. The sample is supported on a hydrophilic area in the center of the film. A sample is slowly dried over a 30–40-min interval and placed in an XRF instrument. The technique relies on there being very little scattered background radiation generated by the very thin support film. The X-ray beam is highly collimated to only excite the small (1–3-mm diameter) sample spot thus further reducing scattered X-ray background entering the detection system.

The limitation of this method is that to reduce scattered radiation or background introduction into the spectrum, very low total analyte concentrations are necessary. The general use of this method is therefore limited. However, as a method to analyze elements present in high-purity aqueous solutions, this technique has many advantages.

Element species not captured using other forms of trace analysis preconcentration may be detectable using the MXA™ technique applied to the supernatant liquid. Applying clean-up techniques to remove alkali metals and Ca, typically found in natural waters, such as those developed for TXRF, may allow this technique to become more widely applicable to environmental analysis.

This technique provides an alternative to the use of TXRF, currently used in the silicon wafer fabrication industry, with lower costs but poorer detection limits.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
APDC	Ammonium Pyrrolidinedithiocarbamate
ASTM	American Society for Testing and Materials
BIF	Boiler and Industrial Furnace Regulation
DBDTC	Dibenzyl Dithiocarbamate
DQO	Data Quality Objective
EDX	Energy-dispersive X-ray
EDXRF	Energy-dispersive X-ray Fluorescence
FP	Fundamental Parameter
FQW	Fuel-quality Waste
GLP	Good Laboratory Practice
IC	Ion Chromatography
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ISO	International Standards Organization
LHW	Liquid Hazardous Waste
LLD	Lower Limit of Detection

MCL	Maximum Contamination Level
MCLG	Maximum Contamination Level Goal
MXA™	Microsample X-ray Analysis
OECD	Organization for Economic Cooperation and Development
QC	Quality Control
RCRA	Resource Conservation and Recovery Act
ROI	Region of Interest
SLF	Secondary Liquid Fuel
SRXRF	Synchrotron Radiation X-ray Fluorescence
TXRF	Total-reflection X-ray Fluorescence
USEPA	United States Environmental Protection Agency
WDXRF	Wavelength-dispersive X-ray Fluorescence
WHO	World Health Organization
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Heavy Metals Analysis in Seawater and Brines • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Proton-induced X-ray Emission in Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Soil Instrumental Methods

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Inductively Coupled Plasma/Optical Emission Spectrometry

Mass Spectrometry **(Volume 13)**
Inorganic Substances, Mass Spectrometric in the Analysis of

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Energy Dispersive, X-ray Fluorescence Analysis • Sample Preparation for X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration • Traceability in Analytical Chemistry

REFERENCES

1. R.E. Van Grieken, A.A. Markowicz (eds.), *Practical Spectroscopy Series, Volume 4. Handbook of X-ray Spectroscopy, Methods and Techniques*, Marcel Dekker, New York, 1993.
2. G.R. Lachance, F. Claisse, *Quantitative X-ray Fluorescence Analysis, Theory and Applications*, John Wiley & Sons, New York, 1995.
3. P.G. Burkhalter, 'Trace Metal Water Pollutants Determined by X-ray Fluorescence', National Technical Information Service, Report Number NRL-76, Naval Research Laboratory, Washington, DC, 1973.
4. J.R. Bacon, A.T. Ellis, A.W. McMahon, P.J. Potts, J.G. Williams, 'Atomic Spectrometry Update – Atomic Mass Spectrometry and X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **9**, 285R–306R (1994).
5. J.R. Bacon, S.R.N. Chenery, A.T. Ellis, A. Fisher, A.W. McMahon, P.J. Potts, J.G. Williams, 'Atomic Spectrometry Update – Atomic Mass Spectrometry and X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **10**, 281R–310R (1995).
6. A.T. Ellis, P.J. Potts, H. Holmes, G.L. Oliver, C. Streli, P. Wobrauschek, 'Atomic Spectrometry Update – X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **11**, 409R–442R (1996).
7. A.T. Ellis, P.J. Potts, H. Holmes, G.J. Oliver, C. Streli, P. Wobrauschek, 'Atomic Spectrometry Update – X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **12**, 461R–490R (1997).
8. A.T. Ellis, D.E. Leyden, W. Wegscheider, B.B. Jablonski, W.B. Bodnar, 'Preconcentration Methods for the Determination of Trace Elements in Water by X-ray Fluorescence Spectrometry', *Anal. Chim. Acta*, **142**, 73–87 (1982).
9. 'SW-846 – Test Methods for Evaluating Solid Waste, Physical/Chemical Methods', US Environmental Protection Agency, Washington, DC, 1996.
10. M.E. Tatro, 'Sample Preparation Techniques for Trace Metal Analysis', *Environ. Testing Anal.*, March/April, 24–37 (1998).
11. B.D. Wheeler, 'Accuracy in X-ray Spectrochemical Analysis as Related to Sample Preparation', *Spectroscopy*, **3**, 3 (1973).

12. G.M. Varshal, T.K. Velyukhanova, V.I. Pavlutskaya, N.P. Tarshinova, A.A. Formanovskyi, F. Seregina, A.M. Shilnikovg, I. Tsysin, Yu.A. Zolotov, 'Detata-filters for Metal Preconcentration and Multi-element Determination in Natural Waters', *Int. J. Environ. Anal. Chem.*, **57**, 107–124 (1994).
13. H.R. Linder, H.D. Seltner, B. Schreiber, 'Use of Dibenzyl Dithiocarbomate as Co-precipitant in the Routine Determination of 12 Heavy Metals in Pharmaceuticals by X-ray Fluorescence Spectroscopy', *Anal. Chem.*, **50**(7), 896–898 (1978).
14. A. Prange, A. Knochel, W. Michaelis, 'Multi-element Determination of Dissolved Heavy Metal Traces in Sea Water by Total-reflection X-ray Fluorescence Spectrometry', *Anal. Chim. Acta*, **172**, 79–100 (1985).
15. M. Del Carmen, Millan-Gonzalez, I. Castilla-Cortazar, 'The Feasibility of Using Speciation Techniques for Arsenic in Waters and Soils', *Agrochimica*, **40**(5–6), 241–246 (1996).
16. M.J. Castilla y Cortazar, J.V. Llopis Cosin, 'Determination of Selenium Species (IV and IV + 0) in Waters for Agricultural Uses', *Agrochimica*, **39**(5–6), 234–239 (1995).
17. R. Stossel, A. Prange, 'Determination of Trace Elements in Rainwater by Total-reflection X-ray Fluorescence Spectrometry', *Anal. Chem.*, **57**, 2880–2885 (1985).
18. B. Holynska, B. Ostachowicz, D. Wegrzynek, 'Simple Method for the Determination of Cu, Hg and Pb in Potable Water with Preliminary Pre-concentration by Total Reflection X-ray Fluorescence Spectrometry', *Spectrochim. Acta, Part B*, **51**, 769–773 (1996).
19. K. Friese, M. Mages, K. Wendt-Potthoff, T.R. Neu, 'Determination of Heavy Metals in Biofilms from the River Elbe by Total-reflection X-ray Fluorescence Spectrometry', *Spectrochim. Acta, Part B*, **52**, 1019–1025 (1997).
20. R. Panayappan, D.L. Venezky, J.V. Gilfrich, L.S. Birks, 'Determination of Soluble Elements in Water by X-ray Fluorescence Spectrometry after Preconcentration with Polyvinylpyrrolidone–Thionalide', *Anal. Chem.*, **50**(8), 1125–1126 (1978).
21. 'D5839-95, Standard Test Method for Trace Element Analysis of Hazardous Waste Fuel by Energy-dispersive X-ray Fluorescence Spectrometry', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 11.04, 1996.
22. 'D6052-97, Standard Test Method for the Preparation and Elemental Analysis of Liquid Hazardous Waste by Energy-dispersive X-ray Fluorescence', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 11.04, 1997.
23. P.A. Russell, R. James, 'Determination of Toxic Elements in Liquid Hazardous Waste Using High-resolution Energy-dispersive X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **12**, 25–32 (1997).
24. H. West, J. Cowley, R. Wills, 'Analysis and Characterization of Water Treatment Plant Sludges by X-ray Fluorescence Spectrometry', *Analyst*, **120**, 1267 (1995).
25. S. Piorek, 'On-site, In-situ Characterization of Contaminated Soil and Liquid Hazardous Waste with Field Portable X-ray Analyzer – A Cost Effective Approach', in *Conference Proceedings, Field Screening Europe*, Kluwer Academic Publishers, Dordrecht, 329–333, 1997.
26. A. McMahon, 'Application of Analytical Methods Based on X-ray Spectroscopy to the Determination of Radionuclides', *Sci. Total Environ.*, **130/131**, 285–295 (1993).
27. D.I. Kaplan, D.B. Hunter, P.M. Bertsch, S. Bajt, D.C. Adriano, 'Application of Synchrotron X-ray Fluorescence Spectroscopy and Energy Dispersive X-ray Analysis to Identify Contaminant Metals on Groundwater Colloids', *Environ. Sci. Technol.*, **28**(6), 1186–1189 (1994).
28. 'Microsample X-ray Analysis', Brochure, Process Analytics, Orem, UT, 1997.
29. 'D2622-94, Standard Test Method for the Determination of Sulfur Products by X-ray Spectrometry', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 5.02, 1998.
30. 'D4294-90, Standard Test Method for Sulfur in Petroleum Products by Energy-dispersive X-ray Spectroscopy', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 5.02, 1998.
31. 'ISO 14596 – Petroleum Products, Determination of Sulfur Content – Wavelength Dispersive X-ray Fluorescence Spectrometry', International Organization for Standardization, Geneva, 1998.
32. 'D5059-92, Standard Test Methods for Lead in Gasoline by X-ray Spectroscopy', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 5.03, 1998.
33. Croner's Waste Management, 'Disposal', Croner Publications, Kingston-upon-Thames, UK, 2-353–2-421, 1997.
34. B. Yokhin, R.C. Tisdale, 'High-sensitivity Energy-dispersive XRF Technology, Part 2: Advances in Instrumentation', *Am. Lab.*, **25**(14), 36H–37H (1993).
35. V.M. Haschke, M. Privik, U. Kowalczyk, J. Broda, 'Possibilities of Rapid Element Analysis of Residual Waste by Energy Dispersive X-ray Fluorescence', *VGB Kraftwerkstechnik*, **6**(74), 549–551 (1994) (in German).
36. J. Heckel, 'Using Barkla Polarized X-ray Radiation in Energy Dispersive X-ray Fluorescence Analysis (EDXRF)', *J. Trace Microprobe Tech.*, **13**(2), 97–108 (1995).
37. B.A. Pedersen, G.M. Higgins, 'EDXRF: a Performance-based Method for Environmental Analysis', *Am. Lab.*, February, 33H–33P (1998).
38. D.S. Kendall, S.D. Schoenwald, M. Siao, S. Hendricks, 'The Determination of Sulfur and Chlorine in Used Oil by X-ray Fluorescence, ICP and Ion Chromotography', *Hazardous Waste Hazardous Mater.*, **12**(4), 373–380 (1995).

Portable Instrumentation: Introduction

Michel Goedert
Hewlett-Packard Laboratories, Palo Alto, USA

The section about *Field-portable Instrumentation* is a collection of papers about portable and transportable analytical instruments and chemical sensors available for analyzing gases and vapors in the field. We also would like it to be a reflection on field measurements and how to improve their quality. All the basic techniques described are mature and have well-established theoretical bases that can be found in other sections of the Encyclopedia. The reader will also find in this section several papers reviewing more advanced methods that are used for measurements of gases and vapors above and below the surface of the earth. In this way, we hope to present a more general approach to the complex problem of atmospheric air pollution. The description of optical analytical techniques is an important issue to field measurements and is treated in detail elsewhere in the Encyclopedia. We encourage the reader to check the other sections since these techniques are an important approach used for global measurements. A paper such as the one written by F. Rathgeb and G. Gauglitz about **Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments**, just to name one of them, provides useful information about the use of optical methods applied to field measurements.

The analysis of gas and vapors in the field is an important part of the remediation process. In the paper prepared by M. Goedert, **Field-portable Instrumentation**, the author gives a general review of field measurements using portable and transportable instruments. This review paper should give the reader an overview of the techniques available in the field. An important part of this article covers direct-reading instruments such as gas sensors, electrochemical devices, ionization detectors and colorimetric detectors. Portable instruments include miniaturized gas chromatographs and mass spectrometers and combinations of these techniques. The author also reviews a powerful, and lesser-known hyphenated technique that is a combination of gas chromatography (GC) and ion mobility spectrometry. It offers less resolution than mass spectrometry (MS) but operates at atmospheric pressure and is truly portable. In the past, ion mobility spectrometry has found applications in the battlefield to detect warfare agents. Civilian applications are presented in this paper.

Chemical sensors are an important part of portable instruments used in the field and are treated in detail in this section. Sensors are generally required to be inexpensive, but must be reliable. In addition they should be user friendly and not require highly trained personnel to operate them. Chemical sensors are not as versatile as analytical instruments such as gas chromatographs or mass spectrometers, but they can usually provide a simple response to simple questions: Is this environment free of such or such toxic compounds? In their paper entitled **Solid-state Sensors for Field Measurements of Gases and Vapors**, F. Dickert and P. Lieberzeit review in detail the principle of operation and applications of several solid-state sensors. They show that chemical sensors are now solidly entrenched in the arsenal of tools available for field measurements, and can provide useful information about the environment for only a fraction of the cost of other analytical instruments.

Electrochemical sensors are an important part of the chemical sensor family. J. Cox and S. Holmstrom review them in their paper **Electrochemical Sensors for Field Measurements of Gases and Vapors**. These sensors are constructed using well-established engineering methods to produce small, compact devices with a high degree of robustness and reliability, freedom from orientation effects, intrinsically low cost and suitability for volume production. They are not very selective, but are sensitive enough to detect gases at the sub-ppm level and when used in array they provide a cost effective and powerful way to monitor combinations of gases in the environment.

An application of remote measurements using chemical sensors has been prepared by J. Hahn and addresses preliminary system engineering for a chemical-sensing network. The title of this paper is **Chemical-sensing Networks: Satellite-based**. It shows the potential of wireless measurements combined with sensors in the field. Sensors and earth-based receiver/transmitters would be linked to low-earth orbiting satellites and the chemical sensors periodically interrogated. The critical technical factors as to orbital analysis, telemetry, data rates, data structures, digital modulation techniques, antennae selection and so on are examined. More specifically the author describes: (1) design of the fiber optic network that will convey subterranean signals to and from the receiver/transmitters on the earth's surface, (2) selection of long-lived, low-power sources for the remotes, (3) application of a pattern recognition algorithm to the frequency shift data that are conveyed to the central processing site's computer at the satellite hub to facilitate the chemical identification of hazardous compounds, (4) satellite transponder access, and (5) the ability of the network to detect chemicals above and below ground.

More traditional chemical instrumentation than chemical sensors is available for field measurements and is

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

reviewed. In the article **Field-based Analysis of Organic Vapors in Air**, R. Berkley describes the results of several evaluations of portable gas chromatographs in the field organized by the US Environmental Protection Agency and other governmental agencies. He shows that field measurements are seldom accurate, mostly because the instruments do not operate well in nonstable environmental conditions. At the same time he emphasizes that analyses performed in a laboratory where the conditions are stable are also subject to important sources of errors due to potential sample degradation. The latter source of errors is difficult to estimate and certainly more difficult to correct than field instrumentation problems.

Is it possible to miniaturize analytical instruments even further? The answer is yes; the miniaturization of analytical instruments is now possible and has already been demonstrated. Portable gas chromatographs and mass spectrometers are becoming smaller and provide faster analysis than their laboratory counterparts. When a decrease in the versatility of the analysis is acceptable, the quality of data provided by portable units is comparable to the quality of measurements obtained with laboratory instruments. W. Higdon and M. Goedert have prepared a technical paper reviewing the subject of **Microelectromechanical Systems Technology Applied to the Miniaturization of Field Instrumentation**. Microfabrication uses the same tools and processes as microelectronics technology to prepare passive micro-electromechanical components for analytical instruments. Microelectromechanical systems (MEMS) applications to analytical chemistry instrumentation started in the 1980s at Stanford University with the design of a gas chromatograph on a chip for the NASA Mars Lender project and eventually led to the development of a commercial instrument.

In this section, the authors review techniques such as lithography, chemical and plasma etching, and evaporation of metal used to prepare MEMS parts. The advantages of this technology are the miniaturization of the parts and the analytical gain in performance that sometimes results from the reduction of the volume of the parts. For instance, fast GC analyses with small-bore columns would not be possible if the internal volume of the detector were not reduced. Analytical applications showing the gain resulting from MEMS are also reviewed in this paper.

The next category of measurement considered is mass dispersion methods. Progress in miniaturization of analytical laboratory instruments has made possible the development of portable mass spectrometers. G. Matz, W. Schröder and T. Kotiaho have prepared a review of the technology and applications in their paper entitled **Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements**. MS, especially combined with gas chromatographic separation, is the most powerful, but also the most expensive and difficult-to-use, analytical tool available. MS is the only tool providing potentially a complete positive identification from the typical fragment patterns. It is fast (less than one second) and accurate (few percent or better). Its dynamic range is large (several decades) and sensitive (sub-ppm).

Sample collection and injection in a gas chromatograph or a mass spectrometer is a difficult problem. An interesting solution, used for the isolation and preconcentration of organic samples from a variety of matrices, is described by L. Müller, T. Górecki and J. Pawliszyn in the paper **Solid-phase Microextraction in Analysis of Pollutants in the Field**. A microextraction is performed by a polymeric coating immobilized on a silica fiber and is based on partitioning of an analyte between the sample matrix and the fiber. The fiber is mounted in a device the size of a hypodermic syringe. Solid-phase microextraction (SPME) is applicable to solid, liquid and gaseous samples and integrates sampling with sample preparation. This attractive approach offers the advantage of solving several important analytical issues in a small package: separation of the sample of interest from the matrix, pre-concentration, and finally transportation and storage before analysis. The separation from the matrix can be done by immersion of the fiber in the sample or by headspace extraction above the sample. When sampling in the field using SPME there are two possibilities: (a) the sampling/sample preparation as well as quantification are performed in the field using SPME and a portable GC; (b) only sampling/sample preparation is performed in the field and the protected fiber with extracted components is then transferred to the laboratory for analysis.

One application paper presented here is about radon measurements. Radioactivity is the spontaneous transformation of the nucleus of an atom by the emission of corpuscular or electromagnetic radiation. Radioactive contaminants are also distinguished by the specialized and very sensitive methods available for the detection of radioactivity. Measurements of a few thousand atoms per liter are not uncommon. The various portable instruments used for measuring radon, thoron and daughter products are reviewed here by F. Cua in the paper entitled **Radon, Indoor and Remote Measurement of**.

When the source of contamination is underground, a different set of equipment is available to perform in situ measurements. One technology that is gaining widespread acceptance for rapid environmental site characterization involves the use of cone-penetrometer systems. This is another example of in situ on-line measurements of determination of water, soil and soil-gas from the subsurface. This technique is not ''portable'' but ''transportable'' and is described in the review paper

written by P. Doskey and E. Cespedes entitled **Cone-penetrometer-deployed Samplers and Chemical Sensors**. It requires a mix of heavy equipment to get under the surface and small, reliable and sensitive types of sensors and transfer lines to perform the in situ analysis. This paper describes the most recent progress made for on-line analytical and in situ chemical detection techniques into the cone penetrometer. Results from site characterizations are provided and compared with laboratory measurements.

What about global measurements? Pollutants are diffusing through the atmosphere and a three-dimensional representation of the problem is very often more useful than a punctual measurement. The paper **Aircraft-based Flux Sampling Strategies** prepared by R. Desjardins, J. MacPherson and P. Schuepp partially answers this question. It may represent a totally different category of transportable instrumentation, but indicates an approach that is worth mentioning about global measurements in the field. It describes the instrumentation used on flux aircraft, the type of data collected and their validity applied to environmental research. Gas exchanges are measured rapidly over large areas and provide a direct measurement of mass and energy exchange at the land/atmosphere interface. The amount of information that is potentially collectable from a specially equipped aircraft is considerable and is extremely useful for getting spatial information about ecosystems and possibly measurement of contaminated clouds. This paper reviews briefly the sensors used and discusses some of the problems associated with the sampling strategy. Positioning of the aircraft can be accurately measured with global positioning systems. The air temperature, carbon dioxide, vapor and ozone concentration can be determined with specific sensors. For trace analyses gas chromatographs have been used with success.

The use of mobile analytical instrumentation in the field offers numerous advantages over bringing the sample to the laboratory for analysis. Sample transportation is limited and consequently the sample degradation is minimized. Another advantage is that rapid identification is possible and suspicious results can be checked by immediate duplication of the analysis. Instruments used in the field should be ideally small and light, use a minimum of consumable fuel (electrical power and gases), operate for long periods without attention and be sensitive enough to detect reliably all contaminants. Unfortunately only a few of the field instruments available today meet these requirements. However, there are promising emerging technologies or analytical methods available today that have been developed for military and space exploration applications that may change the design of analytical instrumentation.

ABBREVIATIONS AND ACRONYMS

GC	Gas Chromatography
MEMS	Microelectromechanical Systems
MS	Mass Spectrometry
SPME	Solid-phase Microextraction

Aircraft-based Flux Sampling Strategies

R.L. Desjardins
Agriculture and Agri-Food, Ottawa, Canada

J.I. MacPherson
National Research Council of Canada, Ottawa, Canada

P.H. Schuepp
McGill University, Montreal, Canada

1 Introduction **1**
2 Techniques of Flux Measurement **1**
2.1 Eddy-covariance Technique 1
2.2 Relaxed Eddy-accumulation Technique 2
3 Instrumentation **2**
3.1 Sensors 2
3.2 Flux Measurements 5
3.3 Other Aircraft-based Systems 6
4 Obtaining Representative Flux Measurements **6**
4.1 Footprint Considerations 6
4.2 Flight Patterns 6
5 Applications **9**
5.1 Ecosystem Characterization 9
5.2 Development of Spectrally Based Algorithms 10
5.3 Unmixing Flux Measurements 11
6 Summary **11**
Acknowledgments **12**
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**

One of the essential elements of plant growth is carbon dioxide assimilation and water vapor loss. Measuring the exchange of these gases can provide an accurate picture of plant growth, health, and ultimate yield. This report describes the instrumentation used on the Canadian flux aircraft and the type of data collected for measuring gas exchange over large areas. It presents flux measurements of carbon dioxide, sensible heat (H) and latent heat (LE) using the eddy-covariance technique. This technique provides the most direct measurements of mass and energy exchange at the land–atmosphere interface. Flux measurements obtained over wetlands near James Bay, the boreal forest in northern Saskatchewan, grasslands in Kansas, agricultural crops in California, and over the city of Fresno in California are presented as examples of the potential of this technique to characterize transfer processes over complex ecosystems. The accuracy of aircraft-based flux measurements is examined using data obtained with other aircraft during wing-to-wing formation flights and with several tower-based systems during tower fly-by. Finally, examples of the use of these data for interpreting satellite data and for characterizing the photosynthetic response of a wide range of vegetation are presented.

1 INTRODUCTION

Information on the magnitude and dynamics of the CO_2 and energy exchange of the Earth's major ecosystems is essential if we are to predict the response of these ecosystems to global changes. Several large-scale projects have been carried out in recent years over a wide range of ecosystems such as grasslands,[1] wetlands,[2] mixed farmlands and forests,[3] and boreal forests.[4] Aircraft-based systems have been an important research tool during these experiments, for model evaluation, for the description of the spatial variability of surface fluxes, and for quantifying the effects of spatial variability on area-averaged surface fluxes.[5]

The mobility of aircraft is a well-recognized asset to quantify fluxes for large areas over the entire depth of the mixed layer.[6–9] This article reviews the current state of the art of aircraft-based flux systems. Although aircraft-based flux measurements are available from many research groups, as a matter of convenience, the examples used to illustrate the basic flux sampling strategies are taken from the authors' research. This paper describes the instrumentation used on flux aircraft, the type of data collected and their quality control. It discusses the accuracy of aircraft flux measurements and presents aircraft–aircraft as well as aircraft–tower comparisons. Finally, flux measurements obtained over a wide range of ecosystems are presented as examples of the kind of information that can be obtained using aircraft-based systems.

2 TECHNIQUES OF FLUX MEASUREMENT

2.1 Eddy-covariance Technique

The eddy-covariance technique provides the most direct measurement of the gas and energy exchange at the land–atmosphere interface, without disturbing the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

environment under study. This technique integrates the flux, F_c over a mosaic of different sources and sinks.[10] Based on Reynold's approximation, it computes the mean products as

$$F_c = \overline{\rho_a WC} = \overline{\rho_a}(\overline{W'C'} + \overline{W}\,\overline{C}) \approx \overline{\rho_a}\left(\frac{1}{N}\sum_{t=1}^{N} W'_t C'_t\right) \quad (1)$$

where the mixing ratio of a scalar, C, equals the sum of the mean mixing ratio, $\overline{C}$, and the deviation from the mean, C'. It is also equal to ρ_c/ρ_a where ρ_c is the density of the gas of interest and ρ_a is the density of dry air. Similarly, the vertical air velocity W is the sum of the mean vertical wind, $\overline{W}$, and the fluctuation from the mean, W'. All these values are measured at equal intervals of time over a certain time period. Reasonably stationary conditions, negligible horizontal advection and absence of chemical reaction involving the gas of interest within the air column below the measuring system, are required for accurate flux measurements.[11] Ignoring these conditions, which may lead to flux divergence with height, may result in significant errors in surface flux estimates from aircraft-based systems.[12,13]

The need for correcting the density of the gas of interest for temperature and water vapor fluctuations is now well recognized for flux measurements of gases using infrared (IR) gas analyzers. Depending on the configuration of the measuring system, the density correction will involve both, only one or none of the above.[14] This correction, which is a function of $\overline{\rho_c}/|F_c|$, results in an error of less than 10% for $\overline{\rho_c}/|F_c| \leq 70\,\mathrm{s\,m^{-1}}$. It can either be done using the so-called Webb correction[15] or by converting the density measurements to the mixing ratio of the gas of interest with respect to dry air.[16] The latter is obtained by measuring the pressure, temperature and humidity of the air at the same time as the density of the gas.

2.2 Relaxed Eddy-accumulation Technique

The relaxed eddy-accumulation technique[17] is a modification of the eddy-covariance technique, which has been developed for aircraft-based flux systems for gases for which no fast-response sensors are available. It involves the partitioning of the gas sample into two separate reservoirs, based on whether the real-time wind-measuring system senses ascending or descending air. The gas sampled can be analyzed later by a slow-response analyzer, such as a gas chromatograph. The conceptual simplicity of the technique is attractive, because the vertical flux estimate $\overline{W'C'}$ for a trace gas is a simple function of the standard deviation of the vertical gust velocity σ_w, the difference of the mean concentration of upward and downward moving gas samples, $(\overline{C^+} - \overline{C^-})$, and an empirical coefficient A:

$$\overline{W'C'} = A\sigma_w(\overline{C^+} - \overline{C^-}) \quad (2)$$

Eddy accumulation systems have been flown on the National Research Council (NRC) Twin Otter[18–20] and the Electra operated by the National Center for Atmospheric Research (NCAR).[21]

3 INSTRUMENTATION

3.1 Sensors

3.1.1 Position

The versatility of an aircraft to act as a sensor platform is enormous (Figure 1). It is suitable to make flux measurements as low as 10 m and as high as the top of the mixed layer (1–2 km). Aircraft can also cover large horizontal areas. Various methods have been used over the years to record the aircraft position, from utilizing the aircraft's basic very high-frequency omni-range/distance measuring equipment (VOR/DME), to Loran-C, to very low frequency (VLF)Omega, to Inertial Navigation Systems (INSs), to the Global Positioning System (GPS). The INS and the current ring-laser gyro inertial reference systems (IRSs) are primarily used on flux aircraft for three-axis inertial velocity measurement, but their position outputs are also recorded, although these can drift by up to $2\,\mathrm{km\,h^{-1}}$ as a result of the Schuler oscillation.[22]

The GPS represents the current standard for position measurement. When operated in the nondifferential mode, accuracy is limited to about 50 m because of "selected availability", an intentional degradation of the GPS signals to restrict the full accuracy to authorized military users. However, centimeter accuracies can be achieved if the system is operated in the differential mode incorporating recorded signals from an accompanying ground station. In this mode, GPS also provides accurate measurement of the aircraft height above mean sea level. Flux aircraft carry radio altimeters and laser altimeters to measure the absolute height above the terrain. These measurements can be combined with pressure–altitude or GPS altitude data to produce records of the height and roughness of the terrain over which the flux measurements are being made.

3.1.2 Air Motion

The most complex and demanding set of measurements from an aircraft is that used to calculate the three-axis wind components. Wind relative to the Earth is determined by the difference between two vectors: the

Figure 1 In major field programs examining the interaction between the atmosphere and the vegetated surface, both flux aircraft and flux towers are used, the aircraft providing spatial and the tower temporal coverage.

airflow relative to the sensors and the motion of the sensors relative to the Earth. As the transit speed of the sensors is typically 10 times the wind speed, the two vectors nearly cancel, requiring very high accuracy in both for good wind resolution. For turbulence measurement, high-frequency response is required as well. The airflow vector is determined from the pressure distribution over a hemisphere pointed into the airstream. On some research aircraft, the hemisphere is the nose radome, sampled through five pressure ports.[23] Many aircraft, including the Twin Otter in Figure 1, use smaller probes extended forward from the nose on booms. The Rosemount 858AJ probe has a hemisphere of 25 mm diameter with five pressure ports. The National Oceanic and Atmospheric Administration (NOAA) Long-EZ aircraft uses a 130 mm hemisphere employing nine pressure ports.[24]

Two primary methods are currently used to measure the motion of the sensors relative to the Earth. Many aircraft utilize commercial INS/IRS, which determine their motion, heading, and attitude angles using high-precision accelerometers and ring-laser gyros. Position and velocity drifts, and resulting wind errors, are corrected during post-flight processing using Kalman filters[25] or spline fits to positions given by GPS. Some aircraft rely entirely on GPS for low-frequency information, matching this with high-frequency data from relatively low-cost accelerometers and mechanical gyros. Commercial GPS systems are now capable of accurate direct measurement of an aircraft's attitude and velocity in addition to position. After computing the vector difference between the instantaneous airflow and inertial velocities, most well-calibrated research aircraft should be able to measure the true wind component to an accuracy of 0.2–0.3 $m\,s^{-1}$ over a frequency range from 0 Hz to at least 10 Hz.

3.1.3 Scalar Sensors

Several fast response sensors are used to measure air temperature, CO_2, H_2O, and ozone concentration on flux aircraft. The placement of some of these sensors on the Twin Otter is shown in Figure 2. The measurements

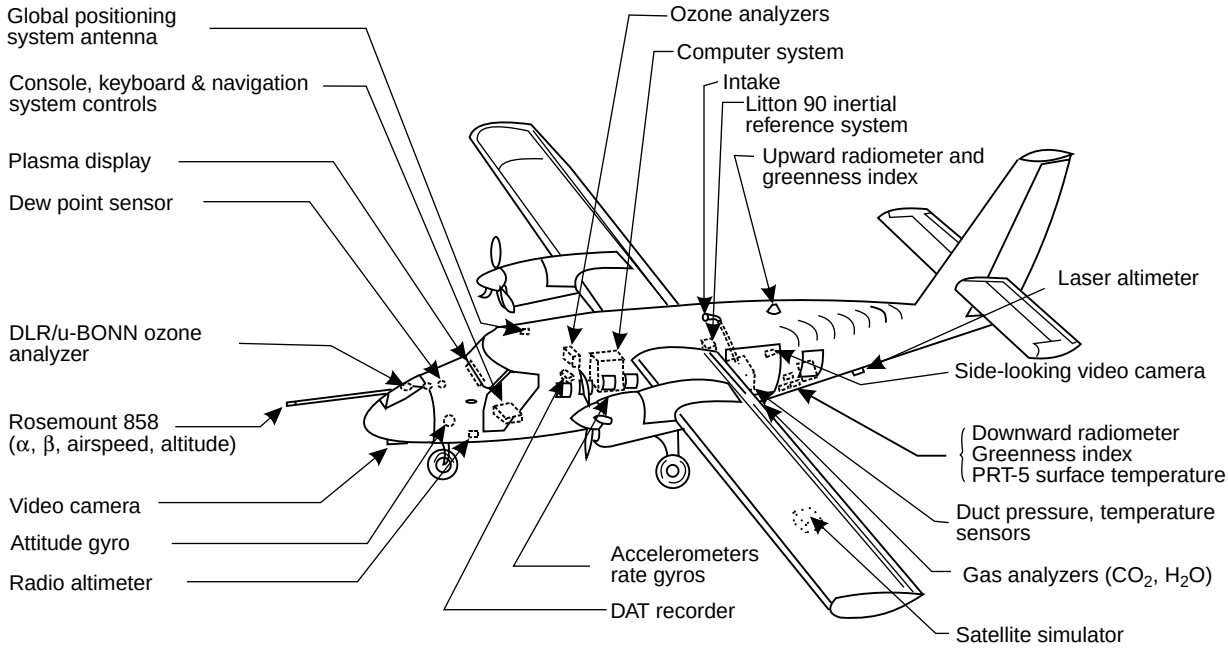

Figure 2 NRC Twin Otter research aircraft instrumented for flux measurements.

are fed to an onboard data processor that applies the calibration constants for each sensor.

For air temperature, the sensor most frequently used is a Rosemount fast response heated (or unheated) probe mounted on the outside of the plane, typically on the nose. The probe has two time constants, 0.016 s for the thermistor and 1.15 s for the housing.[26] Many operators use special high-frequency microbead thermistors installed in either Rosemount housings or in casings of their own design.[27] Several sensors have been used to measure carbon dioxide and water vapor fluctuations.[28–31] In the Twin Otter, an open-path IR gas analyzer developed by Agriculture and Agri-Food Canada had been used for about 13 years.[16] It has a 100% response at 15 Hz, but is not suited to the measurement of absolute concentrations because of the drift in the absolute value. This instrument has recently been replaced by a new open-path IR gas analyzer that provides both the fluctuations and the average concentration with more accuracy. A slightly slower response LI-COR LI-6262 CO_2/H_2O analyzer is also used in the Twin Otter as a back-up unit. In other aircraft, LI-6262, Lyman alphas and other kinds of IR gas analyzers have also been used.[28,30]

Several ozone analyzers have also been deployed on aircraft.[31–33] A fast-response chemiluminescent sensor has frequently been mounted on the nose of the Twin Otter aircraft.[32,33] A Scintrex LOZ-3 ozone detector, which detects the chemiluminescence produced when ozone encounters a surface wetted with a specially formulated solution of Eosin-Y™, has also been tested.[34] The data from this sensor, which has a time constant of the order of 1 s, needs to be corrected for high frequency attenuation for the flux measurements.[26]

For trace gases for which fast-response sensors are not available, gas chromatographs have been used with success on air sampled with an eddy accumulation system to measure the fluxes of agrochemicals and volatile organic compounds (VOCs).[19,20]

3.1.4 Radiation Sensors

The incident solar radiation can be measured with several kinds of pyranometers. It is important to continuously correct the radiometer readings for its mounting alignment and orientation changes due to variations in the heading, and the pitch and roll attitudes of the aircraft. Reflected solar radiation is also measured with a pyranometer. Surface temperature is measured using an IR radiometer. Net radiation (Rn) can be measured directly or it can be calculated from the net short- and long-wave contributions.[35,36] The net short wave is the difference between incident and reflected solar radiation, and the net long wave is the difference between the downward and upward long-wave radiation. The downward long-wave radiation can be estimated from an empirical relationship based on air temperature.[37] The upward long-wave radiation is estimated from the surface temperature measured using an IR thermometer and assuming an emissivity of 0.98. By taking the aircraft attitude angles into account, estimated Rn agreed within a few percent with Rn measured by the various tower-based systems

during BOREAS (Boreal Ecosystem Atmospheric Study) 1994.[36]

3.1.5 *Spectrometers*

Most flux aircraft now carry upward and downward looking spectrometers. The Skye Industries greenness indicator, which is based on the ratio of near-IR (730 nm) to red (660 nm) radiation, indicates the amount of green vegetation beneath the aircraft (Figure 2). The upward facing sensor allows for a possible normalization of the incident radiation to remove errors due to clouds. Another instrument used to measure vegetation characteristics is the downward looking Exotech 100BX satellite simulator. This simultaneously measures reflected radiation over four wavelength bands, and can be configured to simulate two modes of LandSat operation, i.e. thematic mapper (TM) and multispectral sensor (MSS), as well as the Système Pour l'Observation de la Terre (SPOT) satellite. Such sensors, when mounted on a low-flying flux aircraft, only provide information on a narrow strip directly under the aircraft, but these can be compared and integrated with wider scenes provided by various satellites. In most large-scale experiments many kinds of spectral data are collected. Wide-angle sensing spectrometers such as the ESTAR (electronically scanned thinned array radiometer) on the National Aeronautics and Space Administration (NASA) P-3 in the Southern Great Plains (SGP) 1997 Hydrology Experiment, measured surface characteristics that can be directly superimposed on the footprint sensed by the flux aircraft.[38] This combination of data should result in much better correlations between flux measurements from aircraft and surface characteristics observed with airborne spectrometers.

3.2 Flux Measurements

3.2.1 *Data Acquisition*

Much progress has been achieved in the area of data acquisition. Improvements in computer speed, storage capacity, and ruggedness have made them much more reliable in the demanding environment of an aircraft, and have allowed the computation and display of an ever-increasing list of products, including real-time flux estimates. Signals to be recorded in flight originate in a host of analog and digital devices. In the Twin Otter, for example, the input/output (I/O) boards that interface to the input and output devices include RS232 serial boards, Arinc 429, DR11W, analog to digital, digital to analog, digital I/O, small computer systems interface (SCSI) and video graphics array (VGA) graphics. The analog signals are conditioned with a two-phase filtering process. First the signals pass through hardware and signal conditioning where each signal is low-pass filtered at 100 Hz, amplified and biased. These signals are then digitally sampled at a rate of 256 Hz. A digital 47-point 10-Hz low-pass filter is applied and the data are then recorded at 32 Hz on a 4 mm Digital Audio Tape (DAT). The 10 Hz low-pass filter implies that, at a true airspeed of 60 m s^{-1}, the minimum resolvable wavelength is about 6 m.

3.2.2 *Flux Calculations*

Three sets of fluxes are normally computed for the flux calculation. These are computed from the raw, linearly detrended and high-pass filtered time series. The high-pass filtering routine uses a third-order algorithm with an adjustable breakpoint. If set at 0.005 Hz, at a flight speed of 55 m s^{-1} this corresponds to a wavelength of approximately 11 km. In using the eddy-covariance technique to compute fluxes, the data must be adjusted for the transport time for a parcel of air to pass from the noseboom, where the air velocity is measured, to the other sensors. This adjustment is particularly important for runs at low altitude, where the spacing of the sensors can become a significant fraction of the typical eddy size.[16]

After comparing the cospectrum for vertical wind and water vapor using the LI-COR and the Agriculture Canada open path IR analyzer, a substantial underestimation at high frequency was observed by the LI-COR. By using a method to reduce high-frequency losses[26] and assuming the LI-COR had a time constant of 0.2 s, both cospectral estimates matched very well.[34] This correction is now also done for the H flux by assuming the 0.016 s and 1.15 s time constants for the temperature sensor and its housing.[34] As expected, this correction is more important for low-level runs. A similar correction has been done with the vertical wind CO_2 cospectrum. However, it has not been possible to validate this correction using the Agriculture Canada open-path CO_2 analyzer because there is evidence that the CO_2 fluctuations from open-path CO_2 analyzers, which are about 10 times smaller than those of water vapor, are sometimes affected by aircraft motion.[39]

3.2.3 *Data Quality Control*

In a typical field campaign, the aircraft-based flux system records about 100 parameters, 32 times per second for about 100 h. This represents about 1 billion measurements. Quality control of the data should be done during the data collection. This is best done by calculating the fluxes in real time and displaying them to the flight crew. Being able to observe the instantaneous flux measurements reduces the likelihood of collecting data with faulty sensors. A more comprehensive check on the quality of the data can be done after each flight by running the data through spike removal programs as well as examining

the spectra and cospectra for anomalies.[40] A procedure for controlling the quality of the data for both tower-based and aircraft-based flux systems has recently been described.[41]

Errors can also result because the flux sampling criteria are not met due to a poor sample size, a result of either too short a run or an insufficient number of repeated passes.[42] Nonstationary records are particularly troublesome. They should be omitted from the analysis because the calculation of the random and systematic errors are not valid and the computed fluxes are sensitive to the choice of averaging length.[43] The practice of comparing the flux estimates from raw, linearly detrended and high-pass filtered time series can provide a useful screening for the presence of long-wave contributions to the flux estimates.[44]

3.3 Other Aircraft-based Systems

There are many other research groups who are using aircraft-based flux systems. For example, airborne flux measurements of mass and energy were obtained over wetlands in Alaska from the NASA Electra platform.[45] Four flux aircraft were operated during BOREAS-94: the University of Wyoming King Air, the NOAA Long-EZ, the NCAR Electra and the Canadian Twin Otter.[46] The emphasis with all these aircraft was on spatial and temporal coverage. Durand et al. have studied the atmospheric boundary layer over complex terrain.[47] The vertical structure of the marine atmospheric boundary layer has also been studied using two French research aircraft, a Fokker 27 and a Fairchild Merlin IV.[48] Recently, the short- and long-term effects of clearing native vegetation for agricultural purposes have been studied in Australia using two research aircraft which belong to Flinders University.[49] The Institute of Atmospheric Physics in Germany used three instrumented motorized gliders, flying in a coordinated stack, to make simultaneous measurements at three levels in the atmospheric boundary layer.[50] There are many other groups who have either used or are using aircraft-based flux systems.

4 OBTAINING REPRESENTATIVE FLUX MEASUREMENTS

4.1 Footprint Considerations

The flux footprint is the upwind surface area contributing to the flux density measured at a certain height. Determination of the flux footprint is particularly important for flux measurements from heterogeneous source distributions. Figure 3 compares the distances from which 80% of the flux contribution originates, when the fluxes

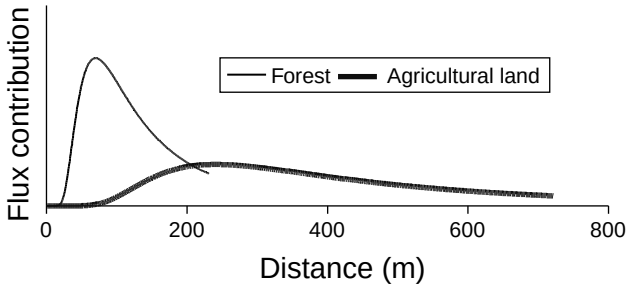

Figure 3 Relative flux contribution over a forest (12 m tall) and agricultural land for unstable conditions as detected by an aircraft at an altitude of 35 m (80% flux contribution is shown).

are measured at 35 m above the vegetation, above a rough surface such as a forest and a smooth surface such as agricultural land. These curves are for unstable conditions; for stable conditions the area contributing to the flux is considerably larger. Such information is very important when one relates fluxes to surface features such as greenness indices and/or vegetation types. Following some of the earlier publications in footprint prediction of scalar flux,[51,52] there has been much interest in this area of research.[53–55] For example, the footprint models over the boreal forest have been fine-tuned on the basis of controlled trace-gas release from the canopy.[56]

4.2 Flight Patterns

The choice of flight pattern (Figure 4) depends on measurement requirements and environmental conditions. Cross-wind profiling at several altitudes within the mixed layer, upwind and downwind of the area of interest, serves to quantify advection and flux divergence with height.[12,13] Profiling and L-shaped patterns are also used to integrate fluxes over an area, and to extrapolate fluxes to the surface for terrain where low-level flights are not possible. Grid flights, which typically consist of about 8 legs of 16 km, spaced 2 km apart, are often flown at low levels. By repeating the whole grid pattern in the opposite direction, each pair of runs on each leg is flown at approximately the same average time, thereby minimizing temporal effects. This type of pattern helps to relate flux values to surface features.[57–60]

Straight transects are frequently flown past towers to evaluate the representativeness of tower-based measurements. Such runs are sometimes extended over long distances to quantify low-frequency contributions to flux and to compare regional flux elements with spectral data to develop spectrally based algorithms. In some cases, straight runs past a tower have been replaced by a circular pattern with a circumference of about 10 km, with comparable results.[36]

Many other flight patterns have resulted in data that have not yet been fully explored. For example, a flight

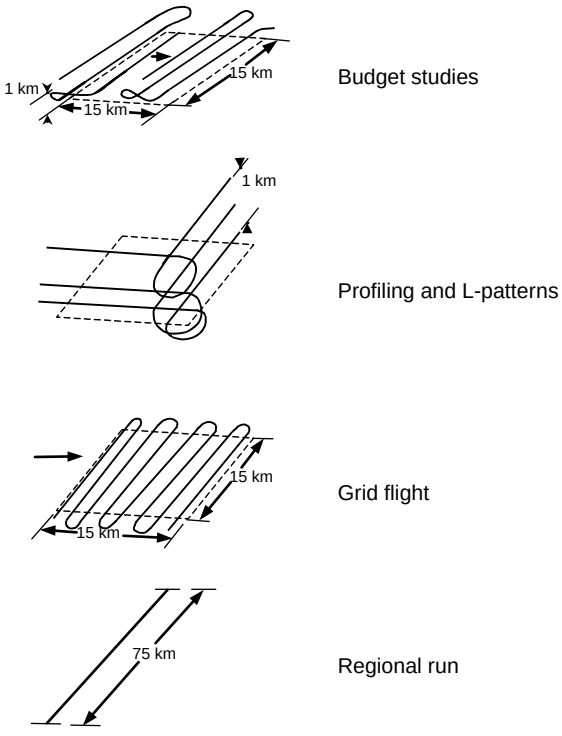

Figure 4 Examples of the flight patterns normally used to obtain flux estimates.

leg over an interstate highway showed large negative fluctuations in ozone due to vehicular emissions of NO.[61] Flights upwind and downwind of a large city have shown that the ozone concentrations downwind of the city were increased by up to 15 ppbv in the first 1500 m.[62] Flights parallel to windbreaks have been used to understand their effects on energy exchange. Flights at night have confirmed that fluxes are close to zero at the top of the nocturnal boundary layer and flights near the top of the mixed layer under daytime conditions are helping to parameterize entrainment. Flights over a large lake in the boreal forest very early in the morning provided insight into the transport of carbon dioxide, water vapor, and

Figure 5 The Wyoming King Air as seen from the Twin Otter during an intercomparison run.

ozone by local circulations at the time of transition from night-time to day-time boundary layer structure.[63]

4.2.1 Aircraft–Aircraft Comparison

Comparisons between measurement systems in flight help to establish confidence in the accuracy of flux instruments, with long transects in close formation as the flight pattern of choice[64] (Figure 5). One advantage of this maneuver is that the aircraft analyze the same air parcels. An alternative to comparison between aircraft with different speeds, is the approximately coincident but independent grid sampling by both aircraft over the same region.[39] Figure 6 shows good agreement between the Wyoming King Air and the Twin Otter for H during BOREAS.[39] A consistent bias was observed between aircraft for the LE flux. During aircraft-to-aircraft comparison in SEMAPHORE (Structure des Echanges Mer–Atmosphère, Propriétés des Hétérogénéités Océaniques: Recherche Expérimentale) just the opposite was observed.[65] Better agreement between LE fluxes than H fluxes indicates that sensor

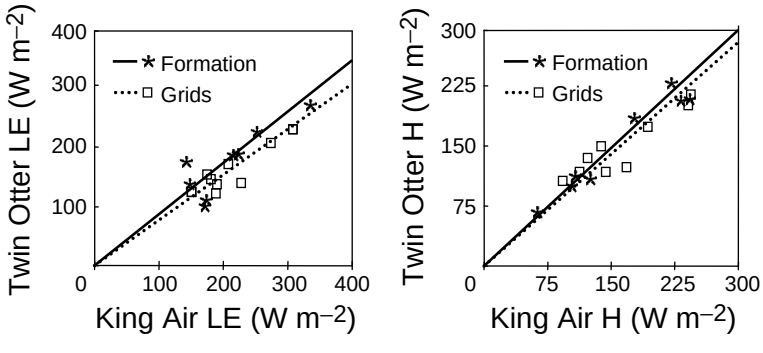

Figure 6 Comparison of the H and LE fluxes as measured by two aircraft during formation and grid flights during BOREAS.

problems are probably the main cause of the disagreements. Many other aircraft intercomparisons were carried out during BOREAS[46] and SGP-97.[35]

4.2.2 *Aircraft–Tower Comparison*

Many comparisons have been made between tower-based and aircraft-based flux systems (Figure 7).[36,66–69] Several factors complicate these comparisons. The aircraft run must be short enough to make sure that the aircraft flux footprint is comparable to the tower flux footprint, and yet the aircraft system must sample a sufficient number of transporting eddies to obtain a reliable flux estimate.[42] Even if such a comparison is difficult to do, it is possible to demonstrate that both systems measure comparable fluxes.

Aircraft- and tower-based flux measurements obtained during the SGP experiment in 1997 are shown in Figure 8. The tower-based flux measurements were collected 1.5 m above grassland, and the airborne flux measurements 30 m above the surface along a 14.7-km track close to the tower. This is not an ideal situation for a comparison as the track was heterogeneous in terms of surface cover.[34] Hence, to make the comparison more realistic, H fluxes were calculated for a 4-km segment close to the tower, as well as for the whole run. This segment consisted primarily of grasslands, similar to the vegetation near the tower. The aircraft-based fluxes are the average of 4–6 passes, whereas the tower data are averaged over 0.5 h periods during which the aircraft fluxes were being measured. For the segments close to the tower, the aircraft H flux was 0.97 of that of the tower H flux. The differences were not significant. For the whole run the H flux observed by the aircraft was 1.43 times the H flux measured by the tower.[34] This is not surprising, considering that a large amount of the track was either dry bare soil or mature wheat.

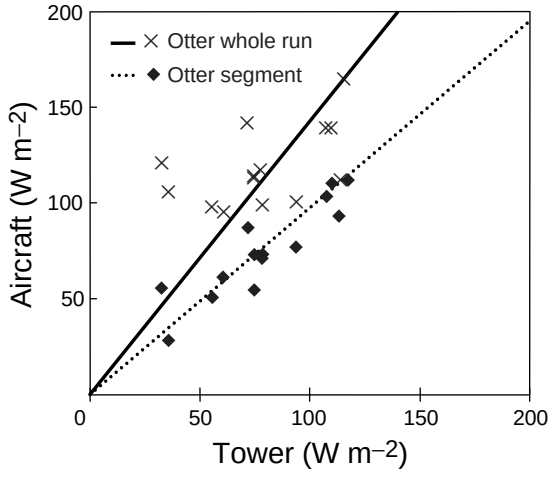

Figure 8 Comparison of the segment and whole run H flux as measured by the Twin Otter at an altitude of 30 m versus a 1.5 m tower, during the SGP experiment in 1997.

Although no methods exist to provide absolute checks on the accuracy of flux measurements, the closure of the energy budget, i.e. the comparison of the sum of H and LE fluxes against Rn minus soil heat flux (G), lends considerable strength to eddy-covariance measurements. Figure 9 shows the sum of H and LE measured by the tower and the aircraft, versus Rn − G measured by the tower during the SGP project in 1997. The aircraft data is again limited to the segment near the tower. The aircraft (H + LE) ratio is 0.86 of (Rn − G), whereas the tower (H + LE) is 0.80 (Rn − G). An underestimation of 14–20% of the H and LE fluxes has frequently been observed over agricultural land around midday when

Figure 7 Twin Otter aircraft during a flux run near a tower-based system over a wheat field in Manitoba, Canada.

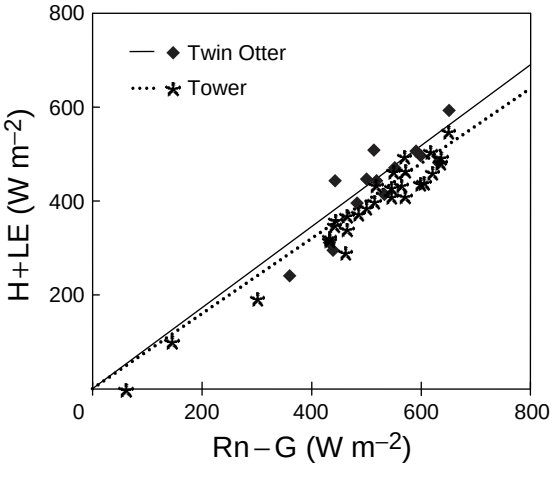

Figure 9 Heat fluxes (H + LE) measured by aircraft- and tower-based systems versus tower-based net (Rn − G).

using the eddy-covariance technique.[68,70] It probably represents the proportion of the exchange that takes place in forms other than turbulent flux and hence not measurable by the eddy-covariance technique. This percentage is lower over rough surfaces such as a forest.[71]

5 APPLICATIONS

5.1 Ecosystem Characterization

5.1.1 Wetlands

Wetlands are important sinks of carbon. They cover an area of $5–6 \times 10^5\,km^2$ worldwide, and consist of sedges, grasses, ericaceous shrubs, sphagnum mosses, and lichens. Carbon dioxide is assimilated by plants and stored in the peat layers, thereby providing a natural sink for atmospheric carbon dioxide. With the ever-increasing possibility of climate change, it is very important to understand the role and response of wetlands to climate change. Aircraft measurements collected around midday over a 25-day period suggested considerable environmental control on the magnitude of CO_2 fluxes over wetlands.[67] Figure 10 shows the CO_2 fluxes measured at an altitude of 35 m over three regions near James Bay during the Northern Wetlands study. Even though these areas are completely different, the CO_2 flux values are primarily a function of the air temperature. A high temperature results in relatively low CO_2 absorption, and vice versa.

5.1.2 Boreal Forests

As part of BOREAS, flux towers were installed at relatively homogeneous locations for the major species in the boreal forest in order to make long-term CO_2 flux measurements.[72] To test whether tower-based fluxes could be scaled-up to obtain regional flux estimates, concurrent aircraft-based flux measurements were collected near these towers by flying a 16×16 km grid pattern at a height of approximately 35 m over the vegetation. Tower-based flux measurements and the relationship between aircraft- and tower-based measurements were used to estimate CO_2 fluxes for the grid area on days for which grid average fluxes were available. Figure 11 demonstrates that even though the boreal forest is a very complex ecosystem, it appears possible to scale up tower-based CO_2 fluxes collected over five of the major species to obtain a reasonably accurate area average of the fluxes.[36]

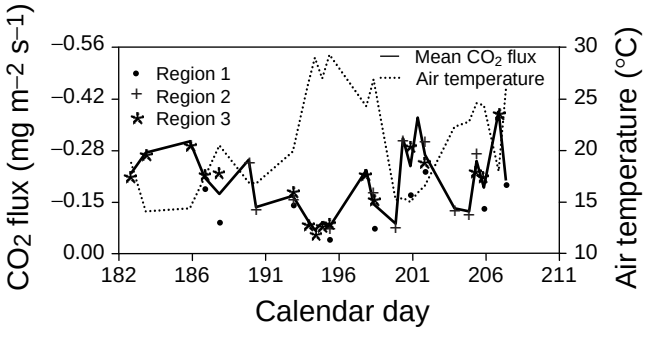

Figure 10 CO_2 fluxes over the Northern Wetlands near James Bay as a function of air temperature during the 1990 field campaign. The different symbols represent different transects covering at least 75 km.

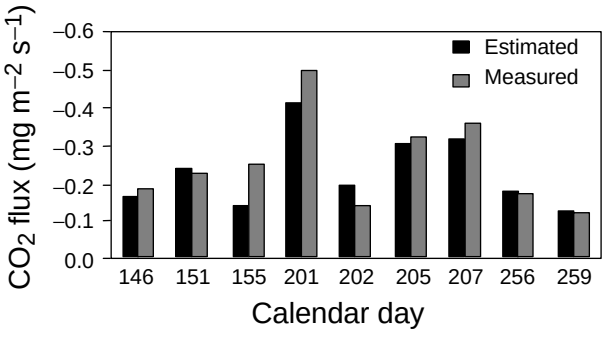

Figure 11 Measured and estimated CO_2 flux for the grid flights during BOREAS in 1994, at the southern study area near Prince Albert, Saskatchewan.

5.1.3 Grasslands

Airborne measurements were used during FIFE (First International Satellite Land Surface Climatology Project Field Experiment) to quantify the spatial variations in CO_2 and H_2O fluxes over a 15×15 km grassland area. The west side of the site was dry and warm whereas the most active vegetation was in the south-east (SE) quadrant. The July 28 footprint-adjusted H_2O fluxes, which are superimposed on the vegetation index picture obtained by the Soviet COSMOS satellite on August 9th, show that the evapotranspiration rate increases as the intensity of the red color increases, i.e. with the amount of green vegetation (Figure 12). The water vapor flux values are the average evapotranspiration of two passes over a 3.8 km segment.[58] A water vapor flux of 100 W m^{-2} is equivalent to an evapotranspiration rate of about 0.15 mm h^{-1}.

5.1.4 Agricultural Regions

Measurements of the CO_2 and H_2O fluxes were made over mixed farmland in California at an altitude of 32 m (Figure 13).[33] These values are 1-km averages with a 3-km running mean. Values of NDVI (normalized difference vegetation index) measured directly below

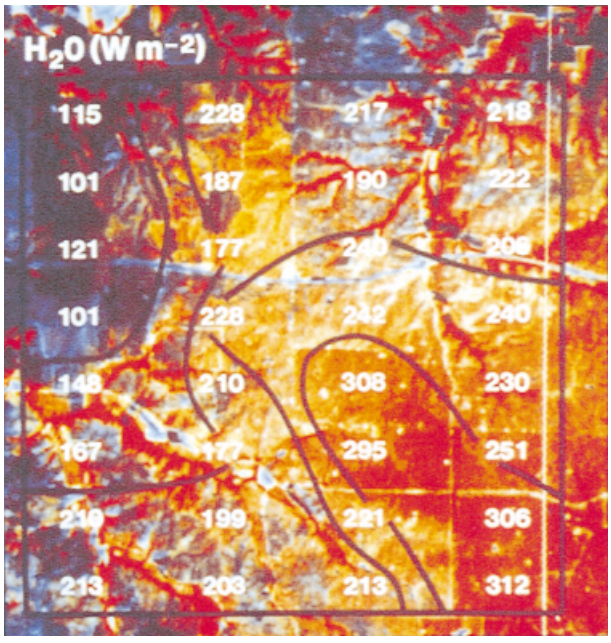

Figure 12 Representation of the amount of green vegetation over the FIFE site obtained from the Soviet satellite COSMOS 1939 on August 9 1989. The footprint-adjusted H_2O flux measurements obtained during a grid flight on July 28 1989 are superimposed.

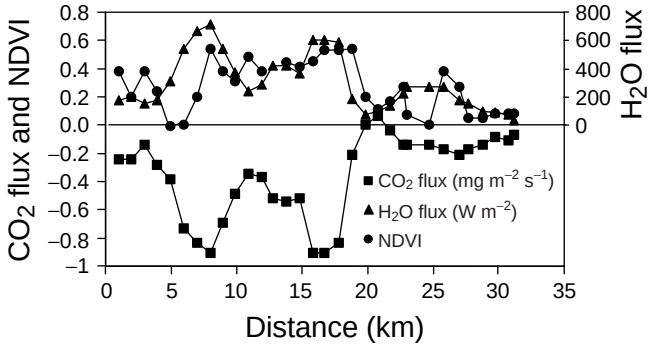

Figure 13 CO_2 and H_2O fluxes, as well as NDVI, all measured at 32 m above an agricultural region in the San Joaquin Valley on July 30 1991.

the aircraft are also shown, to give an indication of the greenness of the vegetation overflown by the aircraft. A value of zero indicates bare soil while a value of 0.6 is typical of very green vegetation. Both the CO_2 uptake and the water vapor flux are large for green vegetation, so that fluxes are well correlated with the vegetation conditions. Such measurements demonstrate the potential to determine the water-use-efficiency of crops, i.e. the rate of growth of crops per unit of water used. The correlation increases considerably if the run is repeated several times, minimizing the effect of intermittent transfer associated with turbulence. The correlation also increases if one corrects for a horizontal offset between the flux values and NDVI because the run was flown into the wind. High correlation between maps of land use and energy exchange characteristics have also been shown in grid flights.[57,60]

5.1.5 Urban Regions

Aircraft make it possible to obtain data over areas where towers would not be practical. Although difficult to arrange, aircraft flights over urban regions can provide valuable information on the amount of CO_2 emitted from fossil fuel use or urban heat island or oasis effects. Figure 14 shows the 5-km mean CO_2 flux measured along a 45-km section of a flight track that crossed the city of Fresno, California, during the afternoon of July 29 1991,[33] based on two passes flown at an altitude of 150 m. The track extended from about 15 km north-west (NW) of the city to approximately 15 km SE of the city. As expected, the rural areas are a sink for CO_2 whereas the city is a source. The magnitude of the source appears to be relatively small because some of the CO_2 emitted is absorbed by the trees in the city.

5.2 Development of Spectrally Based Algorithms

Extensive use of aircraft-based flux measurements will probably come from exploring their suitability as a ground truthing tool for a wide range of satellite-based estimates.[73] Figures 12 and 13 clearly show that CO_2/H_2O fluxes, which are a measure of photosynthesis and transpiration by vegetation, are highly correlated with surface characteristics such as the amount of green vegetation (NDVI). Some refinements are required before these measurements can be directly used, such as removing the effects of other variables. In the case of Figure 15, for example, the solar radiation was decreasing during the latter part of the run. This resulted in smaller CO_2 fluxes

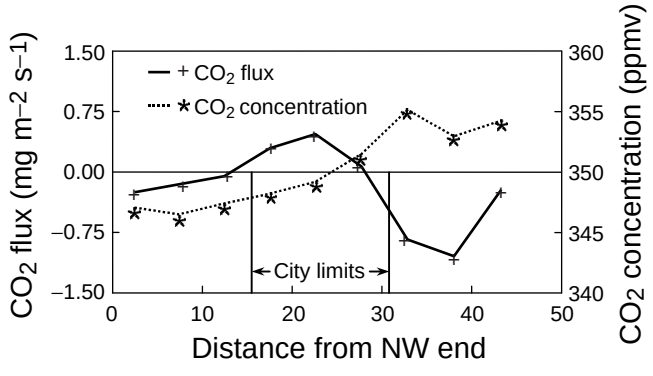

Figure 14 CO_2 concentration and CO_2 flux measured at 150 m along a track that passed over the city of Fresno, California, on July 29 1991.

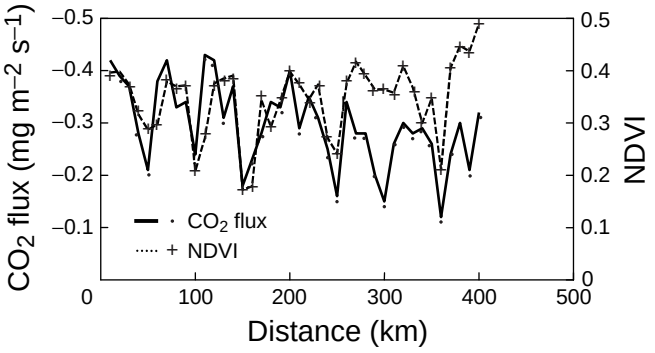

Figure 15 CO_2 flux and the NDVI for a transect from Prince Albert, Saskatchewan, to Thompson, Manitoba.

over the last 200 km. The measurements demonstrate the potential of aircraft-based flux measurements for developing spectrally based algorithms for estimating CO_2 exchanges. The possibility of using other remote sensing measurements, such as radiative surface temperature and NDVI, to estimate the H flux over the boreal forest has also recently been demonstrated with moderate success.[74]

5.3 Unmixing Flux Measurements

A technique for characterizing a wide range of species with fewer towers has been recently explored.[75,76] This unmixing method superimposes the flux footprints for many sections of a grid on a land-cover map from a Landsat TM image to determine the fraction of each type of vegetation in the footprint area. This fraction of vegetation is converted into a fraction contributing to the CO_2 flux for each vegetation type, based on estimated contribution to the CO_2 flux for a wide range of species. A CO_2 flux value is then obtained by the least-squares technique for each of the species selected, and the values corrected by forcing the flux estimates to agree with tower data for the species for which tower data are available. Figure 16 shows the results for July 24 1994, from the BOREAS data set. It shows the mean CO_2 flux value obtained by the Twin Otter aircraft and the estimates at the time of the grid flight for eight vegetation types, i.e. wet conifer, dry conifer, mixed coniferous and deciduous, disturbed, fen, regeneration and visible burn. Case (a) is from a least-squares fit without any constraint; the other cases apply constraints forcing the estimated flux from the aircraft data to match the tower data for that species. The technique provides information for vegetation types for which no tower data are available.

6 SUMMARY

The interest in the use of aircraft-based flux measurements in environmental research has grown enormously since

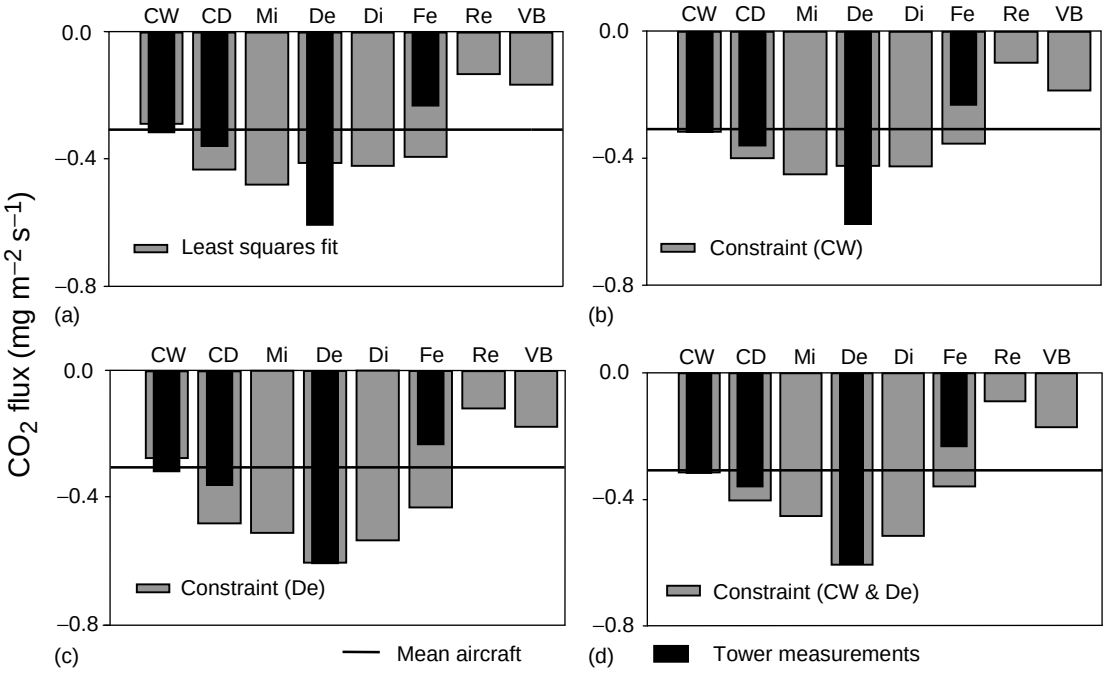

Figure 16 Unmixed CO_2 flux for July 24, 1994, using (a) linear least-squares method, (b) conifer wet forcing, (c) deciduous forcing, and (d) conifer wet and deciduous forcing (CW = conifer wet, CD = conifer dry, Mi = mixed, De = deciduous, Di = disturbed, Fe = fen, Re = rejuvenated, VB = visible burn).

the late 1980s. An aircraft is the only means for obtaining flux average over a wide area. It combines the advantages of rapid deployment to points of interest with the ability to make measurements over large areas and a wide range of altitudes.

This report briefly describes the sensors used on aircraft-based flux systems. It discusses various sampling strategies for obtaining flux measurements over a variety of ecosystems. It also briefly discusses two promising approaches that remain to be fully explored – the development of spectrally based algorithms and the unmixing of fluxes from complex ecosystems.

Although no absolute check exists on the accuracy of aircraft-based flux measurements, comparison with tower-based measurements shows that over grasslands both systems give comparable results. Based on energy budget considerations, however, the fluxes from both systems appear underestimated by about 15–20%. This underestimation, which has frequently been reported with the eddy-covariance technique, appears to be largest for smooth surfaces under high radiation and low wind conditions.

The examples presented should document the potential use and value of aircraft-based flux measurements. Much remains to be further explored, particularly in the area of ground truthing for the wide range of remote sensing observations. It is also important to develop sampling strategies that can provide estimates of diurnal cycles of fluxes.

ACKNOWLEDGMENTS

The assistance of the many scientists and technicians who contributed to data collection and analysis is gratefully acknowledged. The financial support of the following agencies during the collection of the aircraft data is also gratefully acknowledged. These are NRC of Canada, NASA, Agriculture and Agri-Food Canada, Atmospheric Environment Service of Canada, California Air Resources Board, National Science and Engineering Research Council and Canadian Institute for Research in Atmospheric Chemistry.

ABBREVIATIONS AND ACRONYMS

BOREAS	Boreal Ecosystem Atmospheric Study
DAT	Digital Audio Tape
ESTAR	Electronically Scanned Thinned Array Radiometer
FIFE	First International Satellite Land Surface Climatology Project Field Experiment
G	Soil Heat Flux
GPS	Global Positioning System
H	Sensible Heat
INS	Inertial Navigation System
I/O	Input/Output
IR	Infrared
IRS	Inertial Reference System
LE	Latent Heat
MSS	Multispectral Sensor
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
NW	North-west
Rn	Net Radiation
SCSI	Small Computer Systems Interface
SE	South-east
SEMAPHORE	Structure des Echanges Mer–Atmosphère, Propriétés des Hétérogénéités Océaniques: Recherche Expérimentale
SGP	Southern Great Plain
SPOT	Système Pour l'Observation de la Terre
TM	Thematic Mapper
VGA	Video Graphics Array
VLF	Very Low Frequency
VOC	Volatile Organic Compound
VOR/DME	Very High-frequency Omni-range/Distance Measuring Equipment

RELATED ARTICLES

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Field-based Analysis of Organic Vapors in Air

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

REFERENCES

1. P.J. Sellers, F.G. Hall, G. Asrar, D.E. Strebel, R.E. Murphy, 'An Overview of the First ISLSCP Field Experiment', *J. Geophys. Res.*, **97**, 18 345–18 372 (1992).
2. W.A. Glooschenko, N.T. Roulet, L.A. Barrie, H.I. Schiff, H.G. McAdie, 'The Northern Wetlands Study (NOWES): An Overview', *J. Geophys. Res.*, **99**, 1423–1428 (1994).
3. J.C. André, J.P. Goutorbe, A. Perrier, 'HAPEX-MOBILHY – Hydrologic Atmospheric Pilot Experiment for the Study of Water Budget and Evaporation Flux at the Climatic Scale', *Bull. Am. Meteorol. Soc.*, **67**, 138–144 (1986).
4. P.J. Sellers, F. Hall, H. Margolis, B. Kelly, D. Baldocchi, G. den Hartog, J. Cihlar, M. Ryan, B. Goodison, P. Crill, K. Ranson, D. Lettenmaier, D. Wickland, 'The Boreal Ecosystem Atmosphere Study (BOREAS): An Overview and Early Results from the 1994 Field Year', *Bull. Am. Meteorol. Soc.*, **76**, 1549–1577 (1995).
5. C.B. Field, R.E. Murphy, D.S. Schimel, R.A. Schiffer, P.D. Try, F.P. Bretherton, R.E. Dickinson, R.G. Ellington, P.J. Sellers, K.F. Huemmrich, B.W. Meeson, C.O. Justice, F.P. Hall, J.M. Milack, G. Asrar, N.T. Roulet, 'Remote Sensing of the Land Surface for Studies of Global Change: Models–Algorithms–Experiments', *Remote Sens. Environ.*, **51**, 3–26 (1995).
6. R.L. Desjardins, E.J. Brach, P. Alvo, P.H. Schuepp, 'Aircraft Monitoring of Surface Carbon Dioxide Exchange', *Science*, **216**, 733–735 (1982).
7. D.H. Lenschow, 'Aircraft Measurements in the Boundary Layer', in *Probing the Atmospheric Boundary Layer*, ed. D.H. Lenschow, American Meteorological Society, Boston, MA, 39–56, 1986.
8. B. Bénech, P. Durand, A. Druilhet, 'A Case Study of a Non-homogeneous Boundary Layer (Autan 84 Experiment)', *Ann. Geophys.*, **5B**, 451–460 (1987).
9. T.L. Crawford, R.J. Dobosy, R.T. McMillen, C.A. Vogel, B.B. Hicks, 'Air-surface Exchange Measurement in Heterogeneous Regions: Extending Tower Observations with Spatial Structure Observed from Small Aircraft', *Global Change Biology*, **2**, 101–111 (1996).
10. R.L. Desjardins, 'Review of Techniques to Measure CO_2 Flux Densities from Surface and Airborne Sensors', in *Monograph on Advances in Bioclimatology*, ed. G. Stanhill, Springer-Verlag, 1–41, 1991.
11. J.A. Businger, 'Evaluation of the Accuracy with Which Dry Deposition can be Measured with Current Micrometeorological Techniques', *J. Clim. Appl. Meteorol.*, **25**, 1100–1124 (1986).
12. A.K. Betts, R.L. Desjardins, J.I. MacPherson, 'Budget Analysis of the Boundary Layer Grid Flights During FIFE-1987', *J. Geophys. Res.*, **97**, 18 533–18 546 (1992).
13. M. Frech, P. Samuelsson, M. Tjernström, A.M. Jochum, 'Boundary Layer Budgets over the NOPEX Area', *J. Hydrology*, **212**, 155–171 (1999).
14. E. Pattey, R.L. Desjardins, F. Boudreau, P. Rochette, 'Impact of Density Fluctuations on Flux Measurements of Trace Gases; Implications for the Relaxed Eddy-accumulation Technique', *Boundary-layer Meteorol.*, **59**, 195–203 (1992).
15. E.K. Webb, G.I. Pearman, R. Leuning, 'Correction of Flux Measurements for Density Effects due to Heat and Water Vapour Transfer', *Quart. J. Roy. Meteorol. Soc.*, **106**, 85–100 (1980).
16. R.L. Desjardins, J.I. MacPherson, 'Water Vapor Flux Measurements from Aircraft', in *The Measurement and Parameterisation of Land-surface Evaporation Fluxes*, eds. T. Schmugge, J.C. André, Springer-Verlag, 245–260, 1990.
17. J.A. Businger, S.P. Oncley, 'Flux Measurement with Conditional Sampling', *J. Atmos. Oceanic Technol.*, **7**, 349–351 (1990).
18. J.I. MacPherson, R.L. Desjardins, 'Airborne Tests of Flux Measurements by the Relaxed Eddy Accumulation Technique Intercomparison', Proceedings of the 7th AMS Symposium on Meteorological Observations and Instrumentation, New Orleans, 6–11, 1991.
19. T. Zhu, D. Wang, R.L. Desjardins, J.I. MacPherson, 'Aircraft-based Volatile Organic Compounds Flux Measurements with Relaxed Eddy Formation', *Environ. Sci. Technol.*, **33**, 1969–1979.
20. T. Zhu, R.L. Desjardins, J.I. MacPherson, E. Pattey, G. St-Amour, 'Aircraft Measurements of the Concentration and Flux of Agrochemicals', *J. Env. Sci.*, **32**, 1032–1038 (1998).
21. W.A. Cooper, S. Shertz, 'An Intermittent-sampling System for Measuring Fluxes from Aircraft', Ninth AMS Symposium on Meteorological Observations and Instrumentation, Charlotte, 189–194, 1995.
22. K.R. Britting, *Inertial Navigation Systems Analysis*, Wiley-Interscience, 249, 1971.
23. E.N. Brown, C.A. Friehe, D.H. Lenschow, 'The Use of Pressure Fluctuations on the Nose of an Aircraft for Measuring Air Motion', *J. Clim. Appl. Meteorol.*, **22**, 171–180 (1983).
24. T.L. Crawford, R.J. Dobosy, 'A Sensitive Fast-response Probe to Measure Turbulence and Heat Flux from Any Airplane', *Boundary-layer Meteorol.*, **59**, 257–278 (1992).
25. B.W. Leach, J.I. MacPherson, 'Application of Kalman Filtering to Airborne Wind Measurement', *J. Atmos. Oceanic Technol.*, **8**, 51–65 (1991).
26. J. McCarthy, 'A Method for Correcting Airborne Temperature Data for Sensor Response Time', *J. Appl. Meteorol.*, **12**, 211–214 (1973).
27. C. Friehe, D. Khelif, 'Fast-response Aircraft Temperature Sensors', *J. Atmos. Oceanic Technol.*, **9**, 784–795 (1992).
28. D.R. Auble, T.M. Meyers, 'An H_2O and CO_2 Open Path Gas Analyzer for Use with Eddy Flux Systems', *Boundary-layer Meteorol.*, **59**, 243–256 (1990).
29. E. Brach, R.L. Desjardins, G. St-Amour, 'Open Path CO_2 Analyzer', *J. Phys. E. Sci. Instrum.*, **14**, 1415–1419 (1981).

30. A.L. Buck, 'The Variable-path Lyman-alpha Hygrometer and its Operating Characteristics', *Bull. Am. Meteorol. Soc.*, **57**, 1113–1118 (1976).
31. D.H. Lenschow, R. Pearson, Jr, B.B. Stankov, 'Measurements of Ozone Vertical Flux to Ocean and Forest', *J. Geophys. Res.*, **87**, 8833–8837 (1982).
32. R.W. Schmidt, A.M. Jochum, N. Entstrasser, 'Airborne Ozone Flux Measurements Using a New Fast Response Detector', Proc. 7th AMS Symp. on Meteorological Observations and Instrumentations, 243–246, 1991.
33. J.I. MacPherson, R.L. Desjardins, P.H. Schuepp, 'Aircraft-measured Ozone Deposition in the San Joaquin Valley of California', *Atmos. Env. J.*, **29**, 3133–3145 (1995).
34. J.I. MacPherson, 'NRC Twin Otter Operations in the 1997 Southern Great Plains Experiment', Report LTR-FR-146, National Research Council of Canada, 1998.
35. R.J. Dobosy, J.I. MacPherson, 'Intercomparison between Two Flux Airplanes at SGP97', Preprint Volume, 14th American Meteorology Conference on Hydrology, Dallas, Texas, 137–140, 1999.
36. R.L. Desjardins, J.I. MacPherson, L. Mahrt, P.H. Schuepp, E. Pattey, H. Neumann, D. Baldocchi, S. Wofsy, D. Fitzjarrald, H. McCaughey, D.W. Joiner, 'Scaling up Flux Measurements for the Boreal Forest Using Aircraft–Tower Combinations', *J. Geophys. Res.*, **102**, 29 125–29 134 (1997).
37. W.C. Swinbank, 'Long Wave Radiation from Clear Skies', *J. R. Meteorol. Soc.*, **89**, 339–348 (1963).
38. T. Jackson, 'Southern Great Plains 1997 (SGP97) Hydrology Experiment Plan', Version 3/3/97.
39. J.I. MacPherson, 'NRC Twin Otter Operations in BOREAS', National Research Council of Canada Report LTR-FR-129, 1996.
40. F. Chahuneau, R.L. Desjardins, E. Brach, R. Verdon, 'A Micrometeorological Facility for Eddy Flux Measurements of CO_2 and H_2O', *J. Atmos. Oceanic Technol.*, **6**, 193–200 (1989).
41. D. Vickers, L. Mahrt, 'Quality Control and Flux Sampling Problems for Tower and Aircraft Data', *J. Atmos. Oceanic Technol.*, **14**, 512–526 (1997).
42. L. Mahrt, 'Flux Sampling Errors for Aircraft and Towers', *J. Atmos. Oceanic Technol.*, **15**, 416–429 (1998).
43. J. Mann, D.H. Lenschow, 'Errors in Airborne Flux Measurements', *J. Geophys. Res.*, **99**, 14 519–14 526 (1994).
44. P. Caramori, P.H. Schuepp, R.L. Desjardins, J.I. MacPherson, 'Structural Analysis of Airborne Flux Estimates Over a Region', *J. Climate*, **7**, 627–640 (1994).
45. J.A. Ritter, J.D.W. Barrick, G.W. Sachse, G.L. Gregory, M.A. Woerner, C.A. Watson, G.F. Hill, J.E. Collins, 'Airborne Flux Measurements of Trace Species in an Arctic Boundary Layer', *J. Geophys. Res.*, **97**, 16 601–16 625 (1992).
46. R.J. Dobosy, T.L. Crawford, J.I. MacPherson, R.L. Desjardins, R.D. Kelly, S.P. Oncley, D.H. Lenschow, 'Intercomparison Among Four Flux Aircraft at BOREAS in 1994', *J. Geophys. Res.*, **102**, 29 101–29 111 (1997).
47. P. Durand, A. Druilhet, B. Bénech, A. Hamman, 'Mean and Turbulent Structure of Non-homogeneous Boundary Layer: Experimental Study of Two Local Winds During Autan 84', *Ann. Geophys.*, **11**, 502–517 (1993).
48. A. Réchou, P. Durand, A. Druilhet, B. Bénech, 'Turbulence Structure of the Boundary Layer below Marine Clouds in the SOFIA Experiment', *Ann. Geophys.*, **13**, 1075–1086 (1995).
49. S. Chambers, 'Short- and Long-term Effects of Clearing Native Vegetation for Agricultural Purposes', PhD Thesis, The Flinders University of South Australia, 1998.
50. A.M. Jochum, M.E. Reinhardt, H. Willeke, 'Three DFVLR Powered Gliders – A Meteorological System for the Investigation of the Atmospheric Boundary Layer', Proceedings of the 6th AMS Symposium on Meteorological Observations and Instrumentation, New Orleans, 117–120, 1987.
51. M.Y. Leclerc, G.W. Thurtell, 'Footprint Predictions of Scalar Fluxes Using a Markovian Analysis', *Boundary-layer Meteorol.*, **53**, 247–258 (1990).
52. P.H. Schuepp, M.Y. Leclerc, J.I. MacPherson, R.L. Desjardins, 'Footprint Prediction of Scalar Fluxes from Analytical Solutions of the Diffusion Equation', *Boundary-layer Meteorol.*, **50**, 355–373 (1990).
53. T.W. Horst, J.C. Weil, 'Footprint Estimation for Scalar Flux Measurements in the Atmospheric Surface Layer', *Boundary-layer Meteorol.*, **59**, 279–296 (1992).
54. P.H. Schuepp, J.I. MacPherson, R.L. Desjardins, 'Adjustment of Footprint Correction for Airborne Flux Mapping Over the FIFE Site', *J. Geophys. Res.*, **92**, 18 455–18 466 (1992).
55. H.P. Schmid, 'Source Areas for Scalars and Scalar Fluxes', *Boundary-layer Meteorol.*, **67**, 293–318 (1994).
56. S.K. Kaharabata, P.H. Schuepp, S. Ogunjemiyo, S. Shen, M.Y. Leclerc, R.L. Desjardins, J.I. MacPherson, 'Footprint Consideration in BOREAS', *J. Geophys. Res.*, **102**, 29 113–29 124 (1997).
57. C.M. Mitic, P.H. Schuepp, R.L. Desjardins, J.I. MacPherson, 'Spatial Distribution and Co-occurrence of Surface-atmosphere Energy and Gas Exchange Processes Over the CODE Grid Site', *Atmos. Env. J.*, **29**, 3169–3180 (1995).
58. R.L. Desjardins, P.H. Schuepp, J.I. MacPherson, D.J. Buckley, 'Spatial and Temporal Variation of the Fluxes of Carbon Dioxide and Sensible and Latent Heat Over the FIFE Site', *J. Geophys. Res.*, **97**, 18 467–18 476 (1992).
59. S. Ogunjemiyo, P.H. Schuepp, J.I. MacPherson, R.L. Desjardins, 'Analysis of Flux Maps from Twin Otter Grid Flights in BOREAS 1994', *J. Geophys. Res.*, **102**, 29 135–29 145 (1997).
60. P.H. Schuepp, C.M. Mitic, R.L. Desjardins, J.I. MacPherson, 'Spatial Distribution of Ozone Uptake as a Function of Land Use and Energy Exchange Characteristics', *Atmos. Env.*, **29**, 3171–3182 (1995).
61. D.H. Lenschow, R. Pearson, Jr, B.B. Stankov, 'Estimating the Ozone Budget in the Boundary Layer by Use of

Aircraft Measurements of Ozone Eddy Flux and Mean Concentration', *J. Geophys. Res.*, **86**, 7291–7297 (1981).
62. H.H. Janzen, R.L. Desjardins, J.M.R. Asselin, B. Grace (eds.), *The Health of Our Air: Toward Sustainable Agriculture in Canada*, Research Branch, Ottawa, Agriculture and Agri-Food Canada Catalogue No. A53-1981/1998E, 98 pp., 1999.
63. J. Sun, R.L. Desjardins, L. Mahrt, J.I. MacPherson, 'Transport of Carbon Dioxide, Water Vapor, and Ozone by Turbulence and Local Circulations', *J. Geophys. Res.*, **103**, 25 873–25 885 (1998).
64. J.I. MacPherson, R.L. Grossman, R.D. Kelly, 'Intercomparison Results for FIFE Flux Aircraft', *J. Geophys. Res.*, **97**(D17), 18 499–18 514 (1992).
65. D. Lambert, P. Durand, 'Aircraft to Aircraft Intercomparison During SEMAPHORE', *J. Geophys. Res.*, **103**, 25 109–25 123 (1998).
66. R.L. Desjardins, R.L. Hart, J.I. MacPherson, P.H. Schuepp, S. Verma, 'Aircraft and Tower-based Fluxes of Carbon Dioxide, Latent and Sensible Heat', *J. Geophys. Res.*, **97**, 18 477–18 485 (1992).
67. R.L. Desjardins, J.I. MacPherson, P.H. Schuepp, H. Hayhoe, 'Airborne Flux Measurements of CO_2 and H_2O Over the Hudson Bay Lowlands', *J. Geophys. Res.*, **99**, 1551–1561 (1994).
68. R.L. Desjardins, J.I. MacPherson, H. Neumann, G. den Hartog, P.H. Schuepp, 'Flux Estimates of Latent and Sensible Heat, Carbon Dioxide and Ozone Using an Aircraft–Tower Combination', *Atmos. Env. J.*, **29**, 3147–3158 (1995).
69. P. Durand, H. Dupuis, D. Lembert, B. Benech, A. Druilhet, K. Katsaros, P. Taylor, 'Comparison of Sea Surface Flux Measured by Instrumented Aircraft and Ship During SOFIA and SEMAPHORE Experiments', *J. Geophys. Res.*, **103**, 25 125–25 136 (1998).
70. R.L. Desjardins, 'Carbon Dioxide Budget of Maize', *Agric. For. Meteorol. J.*, **36**, 29–41 (1985).
71. E. Pattey, R.L. Desjardins, G. St-Amour, 'Mass and Energy Exchange Over a Black Spruce Forest During Dry Periods of BOREAS 1994', *J. Geophys. Res.*, **102**, 28 967–28 976 (1997).
72. T.A. Black, G. den Hartog, H.H. Neumann, P.D. Blanken, P.C. Yang, C. Russell, Z. Nesic, X. Lee, S.G. Chen, R. Staebler, M.D. Norak, 'Annual Cycles of Water Vapor and Carbon Dioxide Fluxes in and above a Boreal Aspen Stand', *Global Change Biol.*, **2**, 219–229 (1996).
73. R.L. Desjardins, R. Pelletier, P.H. Schuepp, J.I. MacPherson, H. Hayhoe, J. Cihlar, 'Measurements and Estimation of CO_2 Fluxes from Aircraft and Satellite-based Systems', *J. Geophys. Res.*, **95**, 25 549–25 558 (1995).
74. L. Mahrt, J. Sun, J.I. MacPherson, N.O. Jensen, R.L. Desjardins, 'Simple Formulation of Surface Fluxes from Canopies: Applications to BOREAS', *Boundary-layer Meteorol.*, **102**, 29 641–29 649 (1997).
75. J.M. Chen, S.G. Leblanc, J. Cihlar, R.L. Desjardins, J.I. MacPherson, 'Extending Aircraft and Tower-based CO_2 Flux Measurements to a Boreal Region Using a LandSat TM Land Cover Map', *J. Geophys. Res.*, **104**, 16 859–16 878 (1999).
76. J.M. Chen, R.L. Desjardins, S.G. Leblanc, J. Cihlar, J.I. MacPherson, P.H. Schuepp, E. Pattey, 'Estimating CO_2 Fluxes for a Wide Range of Cover Types', Preprint from the Conference of Agriculture and Forest Meteorology, New Mexico, November 2–6, 411–414, 1998.

Chemical-sensing Networks: Satellite-based

John C. Hahn
Indianapolis, USA

1 Introduction 1
2 Orbital Analysis 3
3 A Frequency Plan Configuration that Ensures Low Levels of Channel Interference 4
4 Selection of Antennae 5
5 Data Structure and Associated Telemetry Format 6
6 Determination of the Most Practical Transponder Access Technique for this Application 7
6.1 Random Access 7
6.2 Code Division Multiple Access 7
6.3 Adaptive Polling with Random Burst Transmissions 8
6.4 Summary of Techniques 9
7 Selection of Modulation Techniques Appropriate to this Service 9
8 Choosing a Long-lived, Reliable, Power Source for the Remotes 9
9 Feasibility Analysis of the Prescribed Wide Area Network Realization 9
Abbreviations and Acronyms 11
Related Articles 11
References 11

This paper addresses preliminary system engineering for a disposable, remote, chemical-sensing network. This network could interconnect a satellite hub, low-earth orbiting (LEO) satellites, remote sensors, and a fiber optic (OF) interface that interlinks remote chemical sensors, situated underground in aquifers or directly beneath hazardous chemical/waste sites, with remote transmitter/receivers on the earth's surface. Some critical technical factors: orbital analysis, telemetry, data rates, data structures, digital modulation techniques, antennae selection, remote power considerations, satellite transponder access, and other hardware tradeoffs, are examined in an effort qualitatively to optimize the configuration of a wide area network (WAN). This paper demonstrates the viability of the proposed network, which is suited to the detection of hazardous chemicals in the atmosphere or underground.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

In this article, a new approach to chemical-sensing networks is addressed. This conceptual design will implement surface acoustic wave (SAW) sensors, in association with sensor arrays, fiber-optic links, earth-based receiver/transmitters, and LEO satellites – all working in tandem within a WAN. All major subsystems and technical considerations of this remote chemical-sensing network and their attendant preliminary system engineering will be discussed. The first of these considerations is: how does one most effectively sense the presence of chemical agents or hazardous chemicals in a local area and communicate their presence with adequate notice to appropriate authorities and/or participants, thereby permitting them to take the necessary precautions? The current consensus appears to be that there are three generally accepted chemical sensor techniques: electrochemical sensing, ion mobility spectroscopy (IMS), and SAW chemical sensing.

In the electrochemical sensing method chemical vapors are ionized via a radioactive source. These ionized vapors then pass through an electromagnetic field, where different chemicals are distinguished according to their differences in electric charge and resulting characteristic residence times in the electromagnetic field.

IMS distinguishes between chemicals via differences in charge and mass of ionized vapors in their flight paths.

SAW sensors distinguish between chemicals via shifts in the resonance frequencies. Each sensor has a cylindrical shape, and is impregnated on its periphery with a polymer coating, which is a radio frequency (RF) transmission medium. Currently there exist approximately eight polymer coatings. When each unexposed coating has an RF signal propagate along its length, it demonstrates a characteristic resonance frequency. However, when a foreign substance contacts the polymer coating, the resonance frequency will change – the magnitude of frequency shift depending upon the amount of foreign chemical coming in contact with the coating and the identity of that chemical. An empirically determined database must be called upon to properly evaluate the chemical patterns resulting from the presence of each foreign chemical on a SAW sensor. The pattern-recognition algorithm that is used to identify the foreign chemicals considers, among other things, the polymer coating types employed in each SAW sensor array and the degree of frequency shift that has been observed.[1–4] The pattern recognition algorithm will then correlate the polymer coating types

with the degree of resulting resonance frequency shift for each foreign chemical at a particular sensor array, using primarily this information to compare to a chemical database at the satellite WAN central processing site. It is then at this central processing site, that the foreign chemical is identified. The central processing site for such a LEO WAN is typically located at the satellite network's ground station, alternatively referred to as the ground segment. The LEO satellites of this network may do routine computations, but the ground station serves as the control center for the entire chemical-sensing network. The respective satellites relay information to the central processing site at the ground station, but all important mission operations are orchestrated through the ground station.

If the presence of potentially hazardous foreign chemicals has been sensed, how can this information be most effectively disseminated to the appropriate interested parties? Three common communication options would be:

(1) transmission of chemical information from remote chemical sensors to a land-based transponder via a terrestrial communication link;
(2) transmission of chemical information from remote chemical sensors to an airborne transponder; and
(3) transmission of chemical information from remote chemical sensors on the surface of the earth to LEO satellite(s).

Qualitatively speaking, option (1) appears to be the least attractive. Large side lobes of remote antenna patterns would enhance the probability of detection and compromise in hostile, or business confidential, environments, and the prospect of ''drop outs'' on the communication link would be significant because of line-of-sight obstructions between remote sensors and the earth-based transponder – especially so in the case of a mobile, earth-based transponder, where obstructions to the signal path will be regularly encountered as the transponder moves.

Obstructions to the line-of-sight and fading due to the multipath effects should render this option the least viable of the three. On the other hand, the airborne transponder of option (2) would be less susceptible to line-of-sight obstruction and multipath fading, since the communication links would not be propagating as near to the earth as in option (1). Finally, option (3) would appear to be the most versatile, as it affords the least prospect of line-of-sight obstruction and multipath fading – owing to its more vertically directed communication links. Although in peacetime an airborne transponder would be quite appropriate, an aircraft flying in a periodic fashion overhead would evoke suspicion in the minds of hostile observers. A LEO satellite is harder to detect, thereby providing more confidentiality to its users, and also providing no prospect for endangerment to human life. It is appropriate to mention at this point that, although the types of satellite-based, remote-sensing techniques henceforward advocated for commercial and/or confidential applications in this article might be enhanced to the point of rendering them militarily secure, addressing such enhancements was not the objective of this author and, as a result, is beyond the intended scope of this article. The major points to be offered here will address a satellite-based WAN for, at most, confidential business applications.

As the third option is the most practical of the three, the remainder of this article will focus on remote-sensing configurations of the LEO satellite variety. In order to ensure success of the LEO satellite option, as a minimum the following preliminary system engineering considerations must be addressed:

(1) execution of an orbital analysis to determine the number of satellites, the orbital characteristics of each satellite, velocity, orbital period, stabilization error, etc., the desired availability of the satellites to the remote sensors, and the maximum antenna coverage attainable per satellite;
(2) determination of both the associated bandwidth allocations for transmit and receive data and frequency plans for transmit and receive channels that ensure levels of cochannel and adjacent channel interference at or below the thermal noise floor;
(3) selection of antenna types for both the satellites and remote sensors in accordance with design objectives;
(4) definition of the data structure and its associated telemetry format;
(5) determination of the optimal method of accessing the LEO satellite transponders, providing the maximum of performance and security that is consistent with cost constraints of a disposable, economical, remote chemical-sensing network;
(6) selection of modulation techniques that are amenable to low-cost, low-power remote sensors, i.e. selection of spectrally efficient modulation techniques for power-limited transmission from remote to transponder and noncoherent detection of signals at remote sensors which have been transmitted from the LEO transponder;
(7) selection of an economical power source that provides sufficient power for the remote chemical sensors; and
(8) verification of feasible and acceptable performance of communication links in this proposed remote

chemical sensor network via computed figures of merit, resulting from a link budget analysis.

2 ORBITAL ANALYSIS

For the sake of convenience a polar, circular orbit of 250 km will be arbitrarily chosen. Since the circular orbit is a special subset of elliptical orbit types, Kepler's Laws[5,6] for the period of satellite revolution and velocity of elliptical orbit are used, yielding an orbital velocity of 7.755 km s^{-1}, or 17 340.4 mph. Similarly,the period of revolution is 1.4917 h, 89.50 min. Using the fact that the maximum possible antenna coverage from an altitude of 250 km is at that antenna beamwidth which provides a solid angle that is tangent to the earth's circumference at right angles to the earth's radius, yields a maximum antenna beamwidth of 148.4 degrees. In addition, we will specify that the satellite be 3-axis stabilized with a stabilization error of $+/-5$ degrees, for the sake of convenience (although at this writing the stabilization error can be less than $+/-1$ degree[7]). Then, in order to account for the stabilization error uncertainty, 10 degrees must be added to the computed antenna beamwidth – for a total required beamwidth of 158.4 degrees. This result indicates that satellite transmit and receive antenna patterns should approach hemispheric antenna patterns if they are to achieve the maximum antenna coverage. An antenna beamwidth of 148.4 degrees subtends an arc of 31.57 degrees on the surface of the earth, referenced to the earth's center. Unfortunately, although this design configuration would afford the maximum realizable antenna beamwidth, we compute from the orbital geometry that it would require 11.4 LEO satellites at an orbital altitude of 250 km, which must naturally be rounded to 12 LEO satellites orbiting at 250 km altitude. This implies that the revised, narrower antenna beamwidth subtends an arc of 30.0 degrees on the surface of the earth, referenced to the earth's center. This arc is a radial section of the earth's surface, that is 3339.51 km across. For an orbit of 250 km altitude, the orbital velocity is 7.755 km s^{-1}, which corresponds to a linear velocity on the earth's surface of 7.4625 km s^{-1}. This says that a LEO satellite is made available to each of the 12 antenna coverage zones on the earth for a duration of 447.5 s, or 7.458 min. That is, in order for each of the 100 remote sensors (or remote sensor arrays, in the case of the SAW sensor arrays) to be polled for chemical information while the LEO satellite is overhead in their coverage zone, all chemical information must be transmitted during the 447.5 s when that LEO satellite is available to the coverage zone. If a 1% duty cycle is used for each transmitting sensor, or sensor array, then the total time of transmission for all 100 sensors reduces to 4.475 s, and the time of transmission for each of the 100 sensors is reduced to 0.044 75 s.

Therefore, if 100% satellite availability is desired in this application, i.e. a LEO satellite must be within communication range of the chemical sensor network at all times, a minimum of 11.4 satellites would be needed to achieve maximum antenna coverage – according to application of orbital geometry, where the slant range is a maximum of 1803.2 km at the maximum beamwidth of 148.4 degrees, and the slant ranges at distal edges of the beamwidth are tangent to the surface of the earth. Each of the antenna beamwidths for these 11.4 satellites would subtend a corresponding arc of 31.57 degrees, relative to the center of the earth, translating to an arc that spans 3514.27 km across the surface of the earth. This naturally would be rounded to 12 satellites. When 12 satellites are implemented, the antenna beamwidth narrows to 141.02 degrees, subtending an arc of 30 degrees relative to the earth's center, which corresponds to an arc that is 3339.51 km across.

This orbital geometry can be more easily explained if the reader now refers to Figure 1, which depicts an orbital diagram and some orbital parameters for such a mission. In Figure 1, angle ASD represents the required antenna beamwidth for the LEO satellite antenna. This antenna pattern should approach hemispheric in type. In keeping with the knowledge that 11.4 LEO satellites would ensure that angles SAC and SDC are at right angles to the earth radii at points A and D, and the slant range

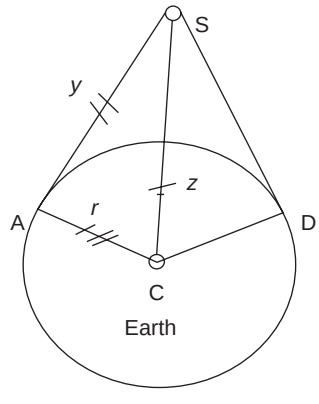

Figure 1 Some orbital characteristics for the proposed chemical-sensing network LEO satellite, $\angle ACD = 30\,deg = 2\angle ACS$; $\angle ASD = 141.0\,deg = 2\angle ASC$ = required antenna beamwidth; $\angle SAC = \angle SDC = 94.5\,deg$; h = altitude = 250 km; $z = h + r = 6378 + 250 = 6628$ km y = slant range = 1715.62 km; time of transmission = 0.044 75 s; 12 LEO satellites at 250 km; 12 LEO satellite coverage zones each span 3339.51 km arcs on the earth's surface; time to traverse LEO coverage zone = 447.51 s; overhead bit set + chemical sensor data set: (a) for electrochemical sensor, 18 bits; (b) for IMS sensor, 32 bits; (c) for SAW sensor, 116 bits.

would be 1803.2 km, a more realizable 12 satellites should yield a slightly narrower required antenna beamwidth of 141.0 degrees. This compares to the required antenna beamwidth in the former case of 148.4 degrees. For the 11.4 satellite scenario recall that the arc of the earth spanned by the LEO satellite's antenna coverage zone was 31.57 degrees, relative to the earth's center. Then a slightly narrower antenna beamwidth would reasonably span 30 degrees. Similarly, the narrower beamwidth and coverage zone in the 12-satellite case would necessitate a slightly higher angle at points A and D than 90 degrees (94.5 degrees). Figure 1 depicts the required antenna beamwidth, but also the longest side, SC, of the two symmetric triangles. These two triangles coincide roughly to the LEO antenna pattern's beamwidth, which is 141.0 degrees in the 12-satellite case. The half angles of these two triangles, angles ASC and DSC, are 70.5 degrees each. This common side has a length equal to the earth radius, plus the satellite altitude of 250 km. The second angle of each triangle is 1/2 the angle subtended by the beamwidth (referenced to the center of the earth), or 15.0 degrees. The third angle of the triangle is the sum of the first two angles subtracted from 180 degrees, or 94.5 degrees. Using the law of cosines, the slant range comes to 1715.62 km. With the slant range now known, the elevation angle[5] is computed, having a value of 1.00 degrees.

To recapitulate on the dynamics of data transmission to the available LEO satellite, it is available to the antenna coverage zone for 447.5 s. For this 1% duty cycle application, all 100 sensors will transmit their data to the LEO over a total period of 4.475 s. Each sensor is randomly allotted a 4.475-s time slot, in which to transmit its overhead bits and chemical data sensor set. Chaotic circuits at the ground station would then randomly select a 0.044 75-s interval within this 4.475-s time slot for transmission of data to the available LEO overhead. This fact implies that 99% of any remote's allotted time slot with the overhead LEO satellite is unused, but if a chemical emergency ensues at another chemical remote, not next in the polling sequence but somewhere else in the same LEO's coverage zone, the polling sequence will be adapted to accommodate such emergencies. Prospective methods as to how such chemical emergencies might be addressed effectively by the chemical-sensing network are elaborated upon further in sections 5 and 6. At this point, it is best to say briefly that an emergency transmission will be permitted soon after the currently transmitting sensor has finished its transmission.

In order to mitigate the possibility of remote collisions, i.e. two remotes sending communications to the proximal LEO transponder at the same time, it is recommendedq that an inactive period be established immediately after the conclusion of each sensor transmission. For example, since the time of transit for any sensor transmission would require a worst case 0.0057 s to contact the LEO overhead, a 0.3 s inactive time window after each sensor transmission is allotted (before the polling sequence is resumed). This 0.3 s of idle time is not expected to compromise the effective operation of the network. On the other hand, contaminated sensors should have higher prospects of communicating their emergency status to the available LEO, without costly time delays. The only apparent consequences of this modification would appear to be a moratorium of 30 s duration on polling requests to be relayed by the available LEO satellite, during its 447.5-s period when it is overhead in one of the 12 satellite coverage zones, and a reduction in delay times prior to communication with the available LEO satellite – regarding chemical emergencies that may exist in the vicinity of the notifying sensor.

Only the chemical sensors that have been exposed to chemical emergencies would be empowered to communicate with the LEO satellite at a time not dictated by the polling sequence. Such sensors would be required to have a pre-empt bit value of ''1'', in order to send chemical data to the available LEO satellite on an out-of-turn basis. (See sections 5 and 6 for more elaboration on data structures and telemetry formats, and LEO transponder access.) Those sensors which have been exposed to hazardous chemicals would simply enter a queue of remotes needing to transmit hazardous chemical information to the nearest LEO transponder. In the event that there are no other sensors in queue with hazardous chemical data to transmit, that remote would wait to be polled by the LEO transponder during that transponder's next inactive period. (This presumes that the ground segment has the capability to adapt its polling sequence as per a remote's chemical environment, and that demand assignment multiple access (DAMA) is chosen as the transponder access method.) In the unlikely event of widespread chemical contamination, to the extent that sensors are unable to communicate with the available LEO for prolonged periods, their data may be buffered at the sensors for later transmission to the next LEO satellite to pass over that coverage zone.

3 A FREQUENCY PLAN CONFIGURATION THAT ENSURES LOW LEVELS OF CHANNEL INTERFERENCE

This frequency plan was chosen because, after appropriate interference analysis and consultation of National

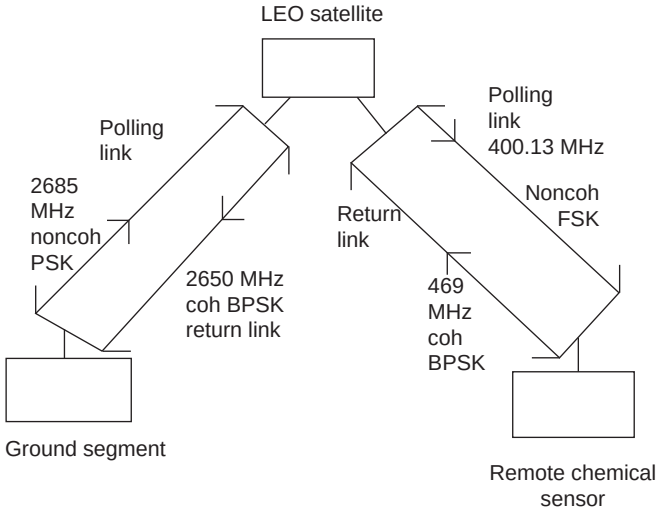

Figure 2 Fundamental satellite links for this WAN. Note that each remote SAW sensor array is either attached on the earth's surface to its nearest receiver/transmitter, or connected to it from underneath the earth's surface via an OF link.

Telecommunications and Information Administration (NTIA) Tables,[8] it was demonstrated that all potential adjacent channel and cochannel interference resides at or below the thermal noise floor for this specific frequency plan. As depicted in Figure 2, the polling link from the ground station, i.e. ground segment, to the LEO satellite is transmitted at a frequency of 2685 MHz, and modulated in a noncoherent, frequency-shift keyed (FSK) manner. The polling link originates at the ground segment, from which a noncoherent FSK digitally modulated signal is transmitted at the modulated carrier frequency of 2685 MHz to the LEO satellite. At the LEO satellite the noncoherent FSK signal is then downconverted to 400.13 MHz, and transmitted to the remote chemical sensors, either near or on the earth's surface. The implementation of noncoherent FSK on the polling link enables the use of an economical, simple envelope detector as the remote's receiver.

On the return link, from remote, to LEO satellite, to ground segment, the remote chemical sensor initiates transmission of a modulated carrier frequency of 469 MHz. A spectrally efficient digital modulation technique is best suited to this power-limited transmission on the return link from the diminutive remote to the LEO, where coherent binary phase-shift keying (BPSK) is recommended for this return link service. At the LEO satellite the return link signal is then upconverted to 2650 MHz, before it is directed back to the ground segment.

4 SELECTION OF ANTENNAE

In this section the antenna types for both the LEO satellites and each of the remote sensors need to be chosen. In the case of the LEO satellite, an antenna pattern approaching hemispheric provides the satellite its broadest possible access to chemical sensors positioned on the earth's surface – within a coverage disk that is 3339.51 km in diameter. On the other hand, the remote sensors may be equipped with antennae that exhibit more directive antenna patterns, so that the remote antenna beamwidth and its associated solid angle is narrower, thereby resulting in less loss of power and a higher quality of signal reception at the satellite receiver in the face of power-limited transmissions from the chemical remote sensors. However, if 100% satellite availability is desired, a more directive LEO satellite antenna pattern would also necessitate more satellites. For example, if the new satellite antenna pattern were one fourth the beamwidth of the currently considered antenna, or 35.25 degrees, four times more satellites, i.e. 48, would be required instead of 12. The cost for a disposable, economical, LEO satellite network for these additional 36 LEO satellite deployments would be prohibitive. Nevertheless, sample computations indicate that the link margin could be as much as 10.8 dB higher with a transmit antenna that transmits at one fourth the beamwidth of the nearly hemispheric antenna pattern (141.0 degrees).

Since the link margin in the worst case, from hub to available LEO satellite, to the remote sensor on the polling link, is 4.89 dB at 10E(-6) bit error rate (BER) in the satellite WAN's current satellite configuration, it would appear that a more directive LEO antenna pattern is not absolutely necessary for this application. Prospective LEO satellite antennae for this economical, remote-sensing application must exhibit simplicity of design, insignificant ground planes, compact size, diminutive weight, circular polarization, and an approximately hemispheric antenna pattern. Such an antenna is a quadrifilar helix antenna, according to Johnson and Jasik.[9] Further, circular polarization is, in general, a desired trait where communication via satellites is concerned.[9] Conversely, as mentioned previously, a more directive remote antenna pattern will result in both a clearer reception at the satellite transponder and a more powerful transmission of the remote's transmitted signal, which is especially attractive in a scenario such as this, where the remote sensor's transmission is power-limited. An axial mode antenna is the antenna of choice for the remote's power-limited service. This axial mode antenna would be used to both receive and transmit signals from the remote. It has

no ground plane, is compact and lightweight – while offering a fairly directive antenna pattern, whose maximum gain is realized along the helical axis.[9]

5 DATA STRUCTURE AND ASSOCIATED TELEMETRY FORMAT

Three data structures will be addressed here, one for each of three chemical sensor options. In the case of the electrochemical sensor, four bits are needed for each sensor data set. This includes two bits to indicate whether the threshold for any of four nerve agents, or similarly behaving neural irritants, have been exceeded. Similarly, the other two bits will indicate whether the safety threshold for any of the four most common blister agents, or similarly behaving skin irritants, have been exceeded, in addition to the 14 overhead bits, for a sum total of 18 bits.

IMS is the second of the chemical sensor techniques that is commonly used to assess the presence of foreign chemicals in the environment, and requires eighteen bits for its sensor data set. Nine bits are allocated for nerve agents, or similarly behaving irritants, and the other nine bits for blister agents, or similarly behaving skin irritants. The nine bits pertaining to chemicals affecting nerves will give some indication as to the identity of the nerve agents or irritants, in addition to informing as to the severity of the exposure – thereby implying what protective action should be taken by humans in the immediate vicinity of these dangerous chemicals. The nine bits that are allocated for the blister agents, or similarly behaving skin irritants, are utilized in an analogous manner as those for nerve agents, with the same 14 additional overhead bits used as aforementioned with the electrochemical sensor.

A 4-SAW sensor array requires 60 bits for its sensor data set, 15 for each SAW sensor. For each of the SAW chemical sensors, the fifteen bits communicate in a binary format the magnitude of frequency shifts that result when a sensor's polymer coating becomes impregnated with a foreign chemical. These 15 bits would then allow binary codes to quantify frequency shifts of up to 2**(15), or 32 768 Hz. These bits are again in addition to 14 overhead bits allocated for each of four SAW sensors in the array, for a total of 116 bits ($[4 \times 15] + [4 \times 14]$). For each of the three chemical sensors enumerated in the prior paragraph, overhead bits are allocated, in addition to those included in the sensor data sets. There are 14 overhead bits which add to the data sensor sets for both the electrochemical sensor and the ion mobility spectroscope, and 14 overhead bits that add to the sensor data sets for each of four chemical sensors in the 4-SAW array. This comes to a total of 18 bits (overhead bits and data sensor set combined) for the electrochemical sensor, 32 bits (18 in the IMS sensor data set, plus 14 overhead bits) for the ion mobility spectroscope, and 116 bits for the 4-SAW array (15 frequency shift bits for each SAW sensor, plus 14 overhead bits for each SAW sensor). The foregoing data structures for each of three sensor types will have 0.044 75 seconds to transmit (see section 2), implying that the data rates for the electrochemical sensor, ion mobility spectroscope, and SAW sensor array are 402 bps, 715 bps, and 2592 bps respectively. Overhead and total bit allocations are as follows:

(1) electrochemical sensor total bits (14 total overhead bits and 18 total bits):

   (a) one start and one stop bit (two total bits),
   (b) a numerical identifier, designating each of the 100 remote sensors in the remote chemical-sensing network with a unique numeric value (seven total bits),
   (c) as no polymer coating will be applied to the electrochemical sensor, four bits will occupy overhead bit positions 10 through 13, where bits 1001 indicate this sensor,
   (d) a pre-empt bit ''1'', to alert that a sensor has detected hazardous chemicals, to pre-empt a sensor in the polling sequence, that is not in chemical contact currently with a hazardous chemical (one bit),
   (e) additional to overhead bits, below/above threshold for poisons hazardous to human nervous systems (two bits),
   (f) additional to overhead bits, below/above threshold for hazardous chemicals that cause blistering of human tissue (two bits);

(2) IMS overhead bits with same breakdown as for the electrochemical sensor's overhead bits (with same fourteen overhead bits and bits 1010 to indicate the IMS sensor), 14 overhead bits and 32 total bits:

   (a) nine bits for chemicals hazardous to human nervous systems, indicating levels of concentration and probable chemical identity, and
   (b) nine bits for chemicals causing blisters to human tissue, indicating levels of concentration and probable chemical identity;

(3) 14 overhead bits for each of the four SAW sensors in a 4-SAW array (56 total overhead bits and 116 total bits):

   (a) one start and one stop bit (two bits),

(b) a numerical identifier for each of the 100 sensors proposed for this WAN (seven bits),
(c) identity of the polymer coating used for each of the four SAW sensors in the array, of eight presumed polymer coating selections, where four bits will be used here,
(d) pre-empt bit ''1'', indicating a contaminant-exposed sensor, to pre-empt a sensor that
(e) is not in current chemical contact with a hazardous chemical (one bit), and
(f) 60 total frequency shift bits, 15 for each of four SAW sensors in the array.

6 DETERMINATION OF THE MOST PRACTICAL TRANSPONDER ACCESS TECHNIQUE FOR THIS APPLICATION

Three prospective modes of satellite transponder access will be evaluated qualitatively:

(1) random access,
(2) code division multiple access (CDMA), and
(3) adaptive polling with random burst transmissions from the remote chemical sensors to the LEO satellite transponder – a variant of DAMA.

On the basis of fundamental cost/benefit analyses, all three transponder access techniques were evaluated as to their relative merits.

6.1 Random Access

Option (1) was analyzed first, and its merits will now be assessed. In a disposable, economical remote chemical-sensing network, this option has several disadvantages, in both commercial and military senses. From the standpoint of bandwidth requirements, in order to ensure that the incidence of simultaneous transponder ''collisions'' by two or more sensor-transmitted signals is 10% or less, the bandwidth requirements are generally as much as twenty times the information bandwidth. Here the information bandwidth would be conservatively twice the data rate. ''Collision'' in this sense refers to two or more transmitted signals arriving at a satellite transponder at the same time and frequency. This necessary increase in bandwidth for a random-access transponder application also would require proportionately greater power at the satellite, adversely affecting its compact size and low cost. Additionally each random-access remote sensor will continue to transmit to the LEO satellite overhead at random intervals, when the satellite is within range. Such a quality is not an asset in a military application. In summarizing, the random-access technique may be acceptable for some commercial services, but its excessive power requirements and frequent transmissions render it undesirable for services where confidential transmission, low prospect of detection, or economy of cost and size are principal considerations.

6.2 Code Division Multiple Access

In Option (2), the feasibility of using CDMA for this application will now be addressed.[10] To initiate this performance comparison, the author consulted two manufacturers of spread spectrum multiple access (SSMA) transmit and receive equipment. Each of these manufacturers stated that direct sequence (DS) spread spectrum would be more expensive than frequency hopped spread spectrum (FHSS), whereas the random burst transmit/receive technique would be cheaper to implement than either of the spread spectrum methods. Additional research in popular technical references served to corroborate these engineering consultations.[11,12] The frequency hopping approach calls for a frequency synthesizer that is adept at switching carrier frequencies very quickly, which serves as a complicating factor for its hardware design. However, noncoherent detection that can be realized with simple envelope detectors is sufficient to perform the receiver function for frequency hopped (FH) CDMA.[13] Before the received signal can be demodulated at the envelope detector, its receiver must acquire and track the pseudorandom noise (PN) code that has defined the sequence of hopping during the signal's transmission – and finally dehop the FH signal to its baseband frequency.[13]

The FH CDMA method, although using more bandwidth over time than a non-SSMA technique, will, for an instant at each frequency hop slot, occupy the same information bandwidth as would one 116-bit word of a non-FH SAW sensor word of chemical data – where each ''slot'' is about the width of the information bandwidth necessary to modulate a sensor's data emission (overhead bits plus its sensor data set). That is to say, at any instant, the modulated FH CDMA signal's bandwidth is typically similar in magnitude to non-CDMA transmission at the same frequency. On the other hand, a DS CDMA signal must be allocated much more bandwidth continuously than its FH counterpart; it therefore requires commensurately more power than does FH CDMA. Similarly, a high hopping speed of the carrier frequency is essential to the FHSS method, but not at all for the direct sequence spread spectrum (DSSS) method. In a similar fashion to its FH counterpart, the PN code for DSSS must first be acquired, tracked, and despread before the demodulation process can begin at the receiver. The demodulation process at the DS receiver typically entails

coherent demodulation of a phase-shift keyed (PSK) digital signal. In summarizing the CDMA method, it would appear that the greater power requirement of DS (with its associated higher bandwidth requirement), and its more complex detection apparatus, make it more expensive to implement. The high-speed frequency synthesizer adds some additional expense to the FH method also, exceeding the cost of a typical receive/transmit approach without spread spectrum. Finally, in order to render a CDMA communication system more ''confidential'' or ''secure'', it should be made resistant to jamming, difficult to intercept, and difficult to detect. However, owing in part to the additional cost for hardware to perform PN code acquisition and tracking and signal encryption, spread spectrum communication systems are appreciably more expensive than their non-CDMA counterparts.

6.3 Adaptive Polling with Random Burst Transmissions

The adaptive polling system with burst transmissions, as would be generated by proposed chaotic circuits for this communication system, is simply a variant of DAMA and, as such, is more economical, more simply designed, and smaller in size and weight than would be either Option (1) (Random Access) or Option (2) CDMA, which would exhibit higher costs and weights than Option (3).

The general application of adaptive polling with random bursts of data, whose random transmissions are induced via chaotic circuits[11,12,14–16] is conceived to operate as follows. Each of the one hundred remote chemical sensors envisioned in the LEO satellite's antenna coverage area, will be designated with a permanent numerical identifier. Each sensor's order in the polling sequence is determined via random selection of its numerical identifier with the aid of chaotic circuits or some other means of random number generation, at the central processing site of the ground segment. It is via this random process that polled transmissions of chemical data will be transmitted. At the conclusion of each polled transmission of remote chemical data to the LEO satellite, as earlier stated in section 2, there is a 0.3-s window of opportunity through which yet-to-be-polled remotes may transmit their chemical data from contaminant-exposed sensors to the available LEO satellite. Such sensors must include a ''1'' pre-empt bit in their overhead bit sets, if the immediately following chemical sensor data is not to be discarded at the LEO satellite. For the sake of illustration, if a ''0'' pre-empt bit accompanies an overhead bit set, its associated chemical sensor data set will not be transmitted to the nearest LEO satellite on an out-of-turn basis. This chemical sensor data set will instead be transmitted, with its ''0'' pre-empt bit, only when its sensor has been polled in the polling sequence.

To further illustrate this point, let us presume hypothetically that sensor #1 has been polled, and is transmitting nonhazardous chemical data to the LEO satellite overhead. At the same instant, however, sensor #29 in the same satellite coverage zone has detected hazardous chemicals in its vicinity. Sensor #29 therefore initiates contact of a pre-emptive nature with the available LEO satellite, to pre-empt the polling process, and thereby postpone the transmission of nonhazardous chemical data to the LEO – in favor of hazardous chemical data, which is instead transmitted to the LEO, and then to the central processing site at the ground control station – during the 0.3-s window subsequent to sensor #1's transmission to the available LEO.

As soon as sensor #1 completes its transmission of nonhazardous chemical data to the LEO overhead, sensor #29 will begin its conveyance of hazardous chemical data to the same LEO. From the LEO this hazardous chemical data is relayed to the hub, or ground control station, where the central processing facility identifies the chemical, its concentration, and what, if any, protective measures and/or disposal actions should be taken in the vicinity of sensor #29. As soon as sensor #29's chemical data has been conveyed to the ground station for further processing, the available LEO satellite in the same satellite coverage zone resumes the polling sequence at the sensor following sensor #1 in that satellite's randomized polling sequence.

An available LEO satellite's designed objective is to ensure that no single sensor is ignored in the LEO satellite's quest for hazardous chemical data as it passes over a satellite coverage zone. In the foregoing illustration the pre-empt bit, a constituent bit of the overhead bit set for one of the chemical sensor data sets in the satellite coverage zone, becomes in essence a priority classifier. If the priority of the transmitting sensor is ''0'', the polling sequence can be modified by a sensor whose pre-empt bit value, i.e. priority value, is ''1''. When such an event occurs, the chemical data transmission associated with the ''0'' pre-empt bit is completed, at which time ''1'' pre-empt bit's chemical data is transferred, to be followed by a resumption of the polling sequence. On the other hand, if a remote chemical sensor with pre-empt bit value ''1'' is transmitting, its transmission cannot be interrupted by another sensor of priority one until transmittal of its hazardous chemical data is complete. This prioritization of the polling sequence, according to the value of a chemical sensor data set's pre-empt bit value, is the quality that renders this polling process ''adaptive''.

Presuming this WAN's design precepts are economy, disposability, performance, simplicity of design, and reliability, each LEO satellite will be designed in such a way as to yield worldwide coverage with 100% satellite availability, but with the minimum number of LEO

satellites. This goal appears to have been achieved here, and the critical orbital parameters for such a mission, have been earlier portrayed both in section 2 and on Figure 1 of this article. Other considerations essential to achieving the desired communications, and drawing clear conclusions from the communication of these chemical data at the ground station's central processing site, will be discussed in the succeeding three sections.

6.4 Summary of Techniques

In summary, the random-access technique has many drawbacks associated with its potential application to this remote, chemical-sensing WAN, most evident of which are the high power and bandwidth requirements, high cost, unwieldy size as a result of high power needs, and resultant high weight of communications equipment that would be needed at both the receiver/transmitters for remote chemical sensors and at the 12 LEO satellites. These disadvantages render the random-access transponder the least attractive of the three candidate transponder access methods. Although having some advantages from the standpoints of confidentiality, difficulty of detection, low probability of intercept, and resistance to jamming, the complexity of CDMA receivers will pose problems with regard to cost. More importantly, however, this transponder would be used in peacetime, rather than military, applications. In addition to excessive cost associated with spread spectrum equipment for this application, the high bandwidth required by the DS CDMA transponders may result in undesired weight and size. Therefore, the DAMA, adaptive polling transponder would without a doubt be the most attractive of the three transponder candidate types for a disposable, chemical-sensing, nonmilitary WAN application.

7 SELECTION OF MODULATION TECHNIQUES APPROPRIATE TO THIS SERVICE

The most critical cost constraint in the configuration of this remote, chemical-sensing network, is the matter of determining how best to limit the cost of remote, chemical sensors – while at worst realizing only mildly discernible degradations in performance. The remote sensor will necessarily be power-limited, and its detector be of simple envelope detector type. Our selection of modulation techniques should complement both these remote design criteria. A spectrally efficient digital modulation scheme would be prudent for the remote transmitter, owing to its power-limited quality. Similarly, an envelope detector provides economical receiver function. Fortunately, both these remote design criteria are realizable when coherent BPSK[5,17] data are sent from each chemical remote's transmitter, and noncoherent FSK transmissions from the available LEO satellite are received at each polled, remote sensor's envelope detector in that satellite coverage zone.

8 CHOOSING A LONG-LIVED, RELIABLE, POWER SOURCE FOR THE REMOTES

After consultations with various vendors of lithium/manganese dioxide batteries, it was concluded[18] that for a duty cycle of less than 50%, in this case, 1%, 3 W power can be delivered by just such a small battery. Vendor performance curves have verified that use of the lithium/manganese dioxide batteries to power remote chemical sensors is indeed reasonable. These power levels for remote transmissions will provide more than sufficient link margins of 4.8 dB or more for this application, according to link results in Table 1.

9 FEASIBILITY ANALYSIS OF THE PRESCRIBED WIDE AREA NETWORK REALIZATION

All engineering analyses pertaining to this remote, chemical-sensing satellite configuration indicate that such a WAN, as described both in the abstract overview of the Introduction and in more elaboration on the foregoing pages, is entirely realizable. More importantly, such a chemical, remote-sensing WAN has been modeled, and should perform effectively, even with the imposition of ambitious cost and performance constraints. These results show that for confidential, commercial and/or public applications this network is more than acceptable. In a military scenario, in the presence of potential jamming and low-power transmissions, it is entirely possible that this WAN, working in tandem with chaotic circuits

Table 1 Link margin results

Sensor type	Link type	Link margin [1.0 E (−5)]	Link margin [1.0 E (−6)]
IMS	Return	13.76 dB	12.92 dB
IMS	Polling	11.32 dB	10.48 dB
SAW	Return	8.16 dB	7.32 dB
SAW	Polling	5.73 dB	4.89 dB
SAW	Return	*8.16 dB	*7.32 dB
SAW	Polling	*5.72 dB	*4.88 dB

* Please note that only link margins with accompanying asterisks exhibit OF links, in addition to RF satellite uplinks and satellite downlinks. All link margins without asterisks have RF wireless links, but no OF links.

Table 2 Equations for the computation of link budgets and link margins and orbital parameters

C/No = {Pt Gt Gr (Lamda)**2}/[((4piR)**2) k Ts (fade) (backoff)]	1
1/(C/No) total = 1/(C/No)up + 1/(C/No)dn	2
1/(C/No) total = 1/(C/No)up + 1/(C/No)dn + 1/(C/No)of	3
Eb/No = C/(No Rd)	4
Pe = Q[(2Eb/No)**1/2] for coherent BPSK digital modulation	5
Pe = 1/2 exp (−Eb/2No) for noncoherent FSK digital modulation	6
(Eb/No)act − (Eb/No)BER = link margin in [dB] (for BER = 1.0 E −5)	7
(Eb/No)act − (Eb/No)BER = link margin in [dB] (for BER = 1.0 E −6)	8
slant range, y: $y = [\text{Re}^{**}2 + (\text{Re} + h)^{**}2 - 2(\text{Re} + h)\cos < \text{SCA}]^{**}1/2$ $y = 1715.623$ km by the Law of Cosines where angle SCA = <SCA, as appears in Figure 1 Re = earth radius [km], h = satellite altitude [km], <SCA = 15.00 deg	9
elevation angle, ϕ: $\phi = [(\text{Re} + h)/y] \sin <\text{SCA} = 1.00$ deg	10
coverage zone: coverage zone = $2(\pi)$Re/(Nsat) coverage zone = 3339.51 km where Nsat = number of LEO satellites = 12	11
orbital velocity, Vo: Vo = $[u(2/\text{Re} - 1/a)]^{**}1/2$ Vo = 7.755 km s^{-1} = 17 340.4 mph for 1.61 km/mi where a = length of the semimajor axis in an elliptical orbit, $a = (\text{Re} + h)$ for a circular orbit and $e = 0$, the orbit eccentricity u = GM = 3.986 10E14 m**3/s**2, the gravitational constant	12
orbital period, P: $P = 2(\pi)[(a^{**}3)/u]^{**}1/2 = 2(\pi)\{[(6.628\ 10E6\ \text{m})^{**}3]/[3.986\ 10E14\ \text{m}^{**}3/\text{s}^{**}2]\}^{**}1/2$ = 1.4917 h = 89.50 min	13

Table 3 Numerical values for variables in Table 2

P_{hub} = transmit power of hub (or ground segment) = 8.0 W
P_{sat} = transmit power of LEO satellite = 5.0 W
P_{rem} = transmit power of chemical remote sensor = 3.0 W
G_{hub} = antenna gain of hub = 38.5 dBi = 7079.5
G_{sat} = LEO satellite antenna gain = 4 dBi = 2.5
G_{rem} = remote antenna gain = 3.0 dBi = 2.0
Y = slant range = 1715.62 km = 1.7156 × 10 E 6 m = 1.7156 × 10**6 m
k = 1.38 × 10**(−23) = 1.38 E (−23) J K^{-1}
T_{s} = system noise temperature at the receiving end of a link (be it hub, satellite, or remote)
T_{rem} = 2610 K = noise temperature of a remote chemical sensor in this application
T_{sat} = 650 K = noise temperature of a LEO satellite
T_{hub} = 300 K = noise temperature at hub, i.e. ground segment or ground station
R_{d} = data rate for either the electrochemical sensor (402 bps), IMS (715 bps), or the SAW sensor (2592 bps)
number of LEO satellites = 12
h = altitude of LEO satellites = 250 km
velocity of LEO satellite orbit = 7.755 km s^{-1} = 17 340.4 mph
period of LEO orbit = 5370.1 s = 89.50 min = 1.4917 h
effective satellite antenna beamwidth = 141.0 degrees
satellite coverage zone = arc that is 3339.51 km across, for each of 12 antenna coverage zones at the earth's surface
LEO satellite time of traversal = time to cross a coverage zone = 447.51 s
= the time required to traverse an orbital arc that is 3470.4 km across at at a 250 km altitude
time of data transmission to LEO per remote, given 100 remotes and 1.0% duty cycle = 0.04475 s
λ = wavelength of electromagnetic transmission in meters
Fade = degrading effect of multipath or other fading on transmission = 3.0 dB = 2.0
BO = backoff = attenuation relative to peak power at the LEO satellite transponder to minimize the prospect of nonlinearity propagation at the receiver = 3.0 dB = 2.0

for enhanced security of operation, could satisfy some military communications standards. However, determination as to whether the communication methods outlined above are appropriate for military application would require large amounts of time and planning, money, engineering in the design, development, and test of computer simulations and/or hardware prototypes. Quantifying and evaluating precisely to which standards of military security the proposed WAN might conform is certainly a worthy topic for follow-on engineering research and analysis, but, once more, is beyond the scope of this article.

The two figures and three tables of this document also serve to confirm the feasibility of the proposed remote, chemical-sensing WAN. For example, Table 1 demonstrates that the satellite links with SAW sensors, with IMS sensors, or underground SAW sensors connected via OF links to receiver/transmitters on the earth's surface,[19] all provide more than adequate link margins (at least 1.5 dB) of 4.88 dB or more. Consequently, the abovementioned WAN for remote, chemical-sensing, as described approximately herein, appears to be viable. Table 2 offers some essential equations used in the feasibility modeling of the proposed WAN, such as those equations used in the computation of link budgets and link margins – replete with probability of bit error equations for noncoherent FSK digital modulation on the polling link and coherent BPSK digital modulation on the return link. Table 3 tabulates the majority of numerical values required in the solution of the equations in Table 2, e.g. powers of transmission, antenna gains, noise temperatures, data rates on links, presumed levels of fading and backoff, slant range, and Boltzmann's constant. The equations of Table 2 and the numerical data from Table 3 were then implemented in the computation of link budgets, and the resulting link margins offered in Table 1. It has been presumed that these links are designed for commercial applications, with prospects for the occurrence of jamming interferences viewed as remote, and, as a result, concluding that a link margin of 1.5 dB or greater is sufficient.

ABBREVIATIONS AND ACRONYMS

BER	Bit Error Rate
BPSK	Binary Phase-shift Keying
CDMA	Code Division Multiple Access
DAMA	Demand Assignment Multiple Access
DS	Direct Sequence
DSSS	Direct Sequence Spread Spectrum
FH	Frequency Hopped
FHSS	Frequency Hopped Spread Spectrum
FSK	Frequency-shift Keyed
IMS	Ion Mobility Spectroscopy
LEO	Low-earth Orbiting
NTIA	National Telecommunications and Information Administration
OF	Fiber Optic
PN	Pseudorandom Noise
PSK	Phase-shift Keyed
RF	Radio Frequency
SAW	Surface Acoustic Wave
SSMA	Spread Spectrum Multiple Access
WAN	Wide Area Network

RELATED ARTICLES

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Radon, Indoor and Remote Measurement of • Solid-state Sensors for Field Measurements of Gases and Vapors

REFERENCES

1. Technical Correspondence with Dr. David Venezky of the Naval Research Lab on the Topic of SAW Devices as Chemical Sensors, Washington, DC, 1992.
2. Technical Correspondence with Dr. Hank Wohltjen and Dr. N. Lynn Jarvis of Microsensor Systems, Inc., Bowling Green, Kentucky, 1992.
3. 'A Smart Sensor Utilizing a Surface Acoustic Wave Sensor Array and Pattern Recognition for Selective Trace Organic Vapor Detection (Parts I and II)', NRL Memorandum Reports 6916 and 6917, Naval Research Lab, Washington, DC, November 19, 1991.
4. 'Final Report Miniature CW Agent Detector', Microsensor Systems, Inc., Contract Number N60921-91-A359, Springfield, VA, January, 1992.
5. W.L. Pritchard, J.A. Sciulli, *Satellite Communication Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
6. J.R. Wertz (ed.), *Spacecraft Attitude Determination and Control*, 4th edition, D. Reidel, Dordrecht, 36–47, 1985.
7. Discussions with Engineers at NASA's Goddard Space Flight Center, Greenbelt, MD, on NASA's Magnitudes for Mission Frequency Uncertainties and Stabilization Errors, June, 1998.
8. Tables of Frequency Allocations and Other Extracts From Manual of Regulations and Procedures for Federal Radio Frequency Management, Dept. of Commerce, NTIA, Washington, DC, September, 1986.

9. R.C. Johnson, H. Jasik, *Antenna Engineering Handbook*, 2nd edition, McGraw-Hill, New York, 1984.
10. A.S. Tannenbaum, *Computer Networks*, 3rd edition, Prentice-Hall, Englewood Cliffs, NJ, 1996.
11. G.M. Bernstein, M.A. Lieberman, 'Secure Random Number Generation Using Chaotic Circuits', *IEEE J. Circuits Systems*, September (1990).
12. L. Letham, D. Hoff, A. Folmsbee, 'A 128K EPROM Using Encryption of Pseudorandom Numbers to Enable Read Access', *IEEE J. Solid State Circuits*, October (1986).
13. R.C. Dixon, *Spread Spectrum Systems*, 2nd edition, John Wiley and Sons, New York, 1984.
14. L.M. Pecora, T.L. Carroll, 'Synchronization in Chaotic Systems', *Am. Phys. Soc.*, **64**(8) (1990).
15. E. Ott, C. Grebogi, J.A. Yorke, 'Controlling Chaos', *Am. Phys. Soc.*, **64**(11) (1990).
16. W.L. Ditto, L.M. Pecora, 'Mastering Chaos', *Scientific American*, August (1993).
17. K. Feher, *Digital Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
18. Technical Correspondence with Duracell, Inc., Bethel, Connecticut, on Lithium/Manganese Dioxide Batteries and their Power Output via Digital Facsimile, 1994.
19. C. Hentschel, *Fiber Optics Handbook*, Hewlett-Packard Company, Boeblingen, Germany, 1989.

Cone-penetrometer-deployed Samplers and Chemical Sensors

Paul V. Doskey
Argonne National Laboratory, Argonne, USA

Ernesto R. Cespedes
US Army Corps of Engineers Waterways Experiment Station, Vicksburg, USA

1 Introduction 1
2 Cone Penetrometer Testing 3
3 Samplers 3
3.1 Groundwater 3
3.2 Soil 4
3.3 Soil Gas 5
4 On-line Analytical Techniques 7
4.1 Isolation Techniques 7
4.2 Preconcentration Techniques 10
4.3 Transfer Techniques 11
5 Chemical Detection Techniques 12
5.1 Chemical Sensors 12
5.2 Gas Chromatography 25
5.3 Mass Spectrometry 25
6 Technology Demonstration and Validation 26
6.1 Samplers 26
6.2 On-line Analytical Techniques 27
6.3 Chemical Detection Techniques 29
7 Conclusions 30
Acknowledgments 30
Abbreviations and Acronyms 31
Related Articles 32
References 32

The development of rapid, cost-effective, accurate methods for environmental site characterization and monitoring has been a very active area of research during the past two decades. One approach that has recently received widespread attention involves the use of cone penetrometer technologies for in situ detection and mapping of environmental contaminants in subsurface soils and groundwater. The most recent technological developments include the incorporation of on-line analytical and in situ chemical detection techniques into the cone penetrometer. This article describes a variety of sampling and chemical sensing technologies that have been or are currently being adapted to cone penetrometer systems for the detection of subsurface contaminants, including petroleum, oils, and lubricants (POLs), volatile organic compounds (VOCs), toxic metals, explosives/energetics, and radioactive wastes. For example, cone penetrometer devices that incorporate in situ isolation techniques to purge VOCs from groundwater eliminate the analysis of the sample media in the laboratory. A high level of agreement was found between an in situ purging technique and the conventional method of bailing a sample from a groundwater sampler, transferring the sample to a container, and analyzing the sample in the laboratory. Completion of a site characterization with this probe was estimated to require approximately 15% of the time needed for conventional methods, significantly reducing costs. The laser-induced fluorescence (LIF) sensor, which measures POLs, represents the first and most mature cone penetrometer chemical detection technology. Technology evaluations estimated that costs of site characterizations using the LIF probe are approximately 50% lower than costs of site characterizations using conventional methods for collecting soil cores, subsampling the cores, and analyzing the subsamples in the laboratory.

1 INTRODUCTION

The traditional methods for determining the location and levels of subsurface chemical contamination in widely varying types of soil and groundwater are time-consuming and costly and often result in less than adequate site characterization. Conventional environmental site investigations require installation of many monitoring wells and subsequent analysis of discrete soil and groundwater samples by using standard laboratory methods. Samples are often collected without any a priori knowledge about the exact location and extent of contaminant plumes. Zones or plumes of contamination can be completely missed or over- or underestimated. Effective site characterization is often hampered by the inability to select optimal locations for monitoring wells. The locations of the monitoring wells are usually based on information gleaned from site historical data, groundwater hydrology, or indirect chemical screening by measuring volatile contaminants in soil gas or some combination of the three. Because of uncertainties in the available information, placement of monitoring wells is at best an inexact science; historical data are often incomplete or inaccurate; knowledge of groundwater hydrology at the site may not provide the level of detail required to understand site characteristics; and interpretations of soil gas measurements may

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

be complicated by erratic movement of vapor in the soil due to impervious layers and changes in atmospheric temperature and pressure. Consequently, many wells are not properly positioned and, therefore, yield information of marginal utility. Furthermore, the ability to resolve horizontal and vertical features in the distribution of chemical contaminants is limited by the spacing between wells and the vertical spacing between samples. Accurately delineating the boundaries of contaminant plumes and defining small-scale vertical structure in the distribution of contamination have important implications with respect to site remediation. The more precisely the area of contamination is defined, the less likely it is that clean material will be unnecessarily removed or subjected to costly remediation procedures. For detailed spatial mapping of contaminated areas, they must often be sampled and analyzed in an interactive manner. Such an approach can be prohibitively costly and labor intensive when conventional techniques are used. In addition, in a number of subsurface environments, the delay from sample collection to interpretation of laboratory analyses can severely hamper the response time and can hinder contaminant containment, possibly resulting in a much larger extent and expense of cleanup.[1]

As a result of these shortcomings, much current research has focused on the development of rapid, cost-effective in situ techniques to monitor both organic and inorganic pollutants. These new methods provide a thorough mapping of contaminated areas by being able to sample the subsurface with high spatial resolution. This mapping makes it possible to rapidly interpret the extent and location of contaminants and efficiently plan and direct cleanup operations. Furthermore, in situ methods minimize the risk of sampling artifacts and can be performed in hostile environments without exposing personnel to toxic contaminants.

One technology that is gaining widespread acceptance for rapid environmental site characterization involves the use of cone penetrometer systems. The electronic cone penetrometer test (CPT) was developed to provide detailed geotechnical information on the subsurface. By combining the results of the CPT with a detailed chemical characterization of the groundwater, soil, and soil gas, the migration of subsurface contaminants can be investigated.[2–8] Many sampling and sensing technologies have been interfaced with the cone penetrometer to provide a chemical characterization of the subsurface. These sampling and sensing technologies provide near-real-time quantitative or qualitative data or both of subsurface chemical contamination with high spatial resolution. The devices include samplers that collect water, soil, and soil gas from the subsurface. The medium is typically transferred to an inert container and returned to the laboratory for analysis. For some analytes, particularly VOCs, these transfer steps can lead to incomplete recoveries of the chemical substances and erroneous results. On-line analytical techniques that include an in situ isolation of the chemical substance from the subsurface media in the cone penetrometer and a direct transfer of the analyte to an instrument or chemical sensor at the surface are being developed to eliminate these transfer steps. In a more direct approach, chemical sensors have been interfaced with the cone penetrometer to perform an in situ analysis of the subsurface media, thereby eliminating the isolation and transfer steps of the on-line analytical techniques.

The Site Characterization and Analysis Penetrometer System (SCAPS) was initially developed by the US Army during the late 1980s and at that time was the first cone penetrometer system specifically designed for rapid environmental site characterization. The SCAPS (Figure 1) is based on a custom-engineered 20-ton truck capable of hydraulically pushing an instrumented probe to a maximum depth of 50 m. The truck houses two separate, protected work spaces to allow access to contaminated sites with minimal risk to the work crew. One of the work spaces contains the penetrometer tool, the push pipe, and hydraulic controls for leveling the truck and advancing the penetrometer tool. The other work space houses optical systems, chemical analysis equipment, and the equipment for digital data acquisition, processing, and display. The penetrometer unit can be equipped with a number of sensors and samplers designed to detect a variety of classes of subsurface contaminants, including POLs, VOCs, metals, explosives,

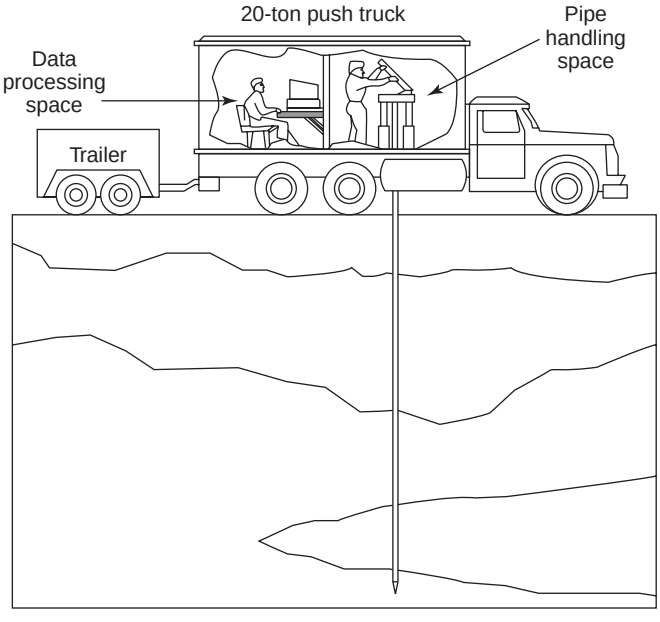

Figure 1 Typical CPT system for environmental site characterization and monitoring. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

and radioactive wastes.[9] Incorporated into the SCAPS rig, and also into a number of other cone penetrometer units, is equipment for decontaminating the push rods as they are being retracted that consists of rubber wipers and steam cleaning equipment mounted under the push room. As the push rods are retracted following a cone penetrometer push, the rods are scrubbed and steam cleaned, and the wastewater is collected for off-site disposal. Also incorporated into a number of cone penetrometer probes is equipment that allows for grouting of the hole during probe retraction to prevent any migration of contaminants. The SCAPS trucks are currently being operated by the United States Department of Defense (USDOD), the United States Department of Energy (USDOE), the United States Environmental Protection Agency (USEPA), and private industry to conduct site characterization tasks.[10] The various sensors and samplers used on SCAPS and similar cone penetrometer systems are described in subsequent sections.

2 CONE PENETROMETER TESTING

The CPT is used to gather geotechnical information on the subsurface. Mechanical cone penetrometers were developed in the early 1900s to identify strong soil layers for the design of pilings and building foundations. The mechanical CPT provides a discontinuous record of geotechnical properties at shallow depths. The electronic cone penetrometer was developed in the 1960s to examine accurately soil stratigraphy at depths approaching 90 m.[2,6,8,11–13] The electronic cone penetrometer, which consists of an instrumented conical probe (cone) and 1-m sections of metal pipe (rods), is advanced with hydraulic jacks from a skid- or truck-mounted rig that can weigh between 20 and 40 tons. The maximum thrust the rods can withstand is 45 000 lb force (200 kN). This thrust will advance the cone to 30 m in stiff material and as much as 80 m in looser soils. The probe is advanced in 1-m intervals at 1–2 cm s^{-1} and is interrupted at each interval to attach additional rods to the top of the string of rods in the cone penetrometer truck. The electronic CPT represents a continuous record of the resistance to penetration.

The cone consists of a hollow steel rod that is 3.0–4.1 cm in outer diameter (OD) [2.5-cm inner diameter (ID)] with a 60° conical point and a friction sleeve that is 13.3 cm long. Several sensors located in the cone are connected to an automatic data recording system in the cone penetrometer truck by an electrical cable that runs through all of the rod sections. Strain gauges located behind the tip and friction sleeve measure the displacement associated with the resistance to penetration. The tip resistance is a measure of the total stress or intergranular pressure plus water pressure, and the friction resistance is the sum of friction and adhesion. The ratio of these two resistances is used to determine soil type and to derive a stratigraphic log of the CPT. Soil classification by using the CPT is empirically based.[14–16] The classification differentiates gravel, sand, silt, clay, some intermediate soil types and some overconsolidated soils. The American Society for Testing and Materials (ASTM) standard D 3441-86 is used to ensure uniformity with regard to the cone characteristics, calibration, and testing and reporting procedures.[17]

The cone can contain several other sensors for measuring parameters that are useful in interpreting the migration of subsurface contaminants. Dynamic and equilibrium pore water pressures are measured by using a porous stone filter in the cone that is connected to a pressure transducer.[18,19] A probe to determine soil moisture has also been developed.[20] Sensors to measure the inclination are used for correcting depth to the lateral drift of the cone. Electrical conductivity sensors and nuclear and seismic probes have also been used in the cone penetrometer. Soil resistivity can be used as a rough measure of the contamination of the soil.[19,21] The hydraulic conductivity, hydrostatic head, independent soil classifications, and soil permeability can be derived from measurements obtained from these additional sensors.

3 SAMPLERS

Many samplers have been interfaced with the cone penetrometer to perform an in situ collection of groundwater, soil, and soil gas. Following sample collection, the soil and groundwater probes are usually brought to the surface, and the sample is transferred to a container and analyzed on-site or transported off-site to a laboratory for analysis. Soil gas probes directly transfer the gas from the probe through tubing to a container or gas analyzer at the surface. Careful attention must be paid to the materials that are used in the device and in the sample container so that the media is not contaminated or the analytes are not sorbed to the device or container. Groundwater and soil gas are typically sampled through a porous element in the probe. Soil cores are collected in liners located in the rods of the penetrometer. The soil cores can be retrieved and subsampled in the field, or the ends of the liner can be capped and the entire core returned to the laboratory.

3.1 Groundwater

Several different designs have been used for cone penetrometer groundwater samplers.[8] One of the common features of the designs is a porous element through which the groundwater must pass before it enters a sample chamber or container. The filter is typically covered by a

retractable shield that protects it from contamination during the cone penetrometer sounding. Groundwater enters many of the devices under ambient hydrostatic pressure; however, some samplers are capable of applying a slight vacuum to aid in the collection of the sample.

Wellpoint batch water (WBW) samplers have been used for many years and are designed to collect large volumes of groundwater.[22–26] The WBW sampler designed by Hogentogler & Co., Inc. (Columbia, MD, USA) has a stainless-steel point that is attached to a rod section containing a slotted stainless-steel screen. The screen can be either exposed or protected by a sleeve. A sampler with an unprotected screen can be used at several depths; however, the filter may become contaminated during the push. A sampler with a protective sleeve is pushed to the desired depth, and the rods are retracted to expose the filter. Water samples are retrieved with a bailer, or the sample chamber can be pressurized and the sample retrieved from another line, or the entire probe can be retrieved and the sample poured out of the tip.

Two designs of the HydroPunch™ (QED Environmental Systems, Inc., Ann Arbor, MI, USA) have been used for groundwater sampling.[27] The sample chamber of the HydroPunch I™ has a capacity of 500 mL and is constructed of stainless steel and Teflon®.[28,29] A shielded section of perforated stainless-steel pipe is used as an intake. A disposable polypropylene screen covers the water intake in the sample chamber. To retrieve the sample the rods are retracted so that two Teflon® check valves close and retain the sample in the chamber. At the surface, the upper check valve is replaced with a stopcock and a disposable tube, the sampler is turned upside down, and the sample is transferred to an appropriate container. The sample chamber of the HydroPunch II™ has a capacity of 1250 mL and is constructed of stainless steel and Viton®.[27] Unlike the HydroPunch I™, this sampler is capable of collecting pools of organic contaminants that are less dense than water. The sample intake consists of a shielded 125-μm, 120-mesh stainless-steel screen. Poly(vinyl chloride) or polypropylene screens can also be used with the sampler. A bailer is lowered through the rods and into the sample chamber to retrieve the sample. At the surface, the water is transferred from the bailer to another container. A sampler similar to the HydroPunch™ has been designed that uses a 250-μm polypropylene porous filter as the water intake [Applied Research Associates, Inc. (ARA), South Royalton, VT, USA]. The sample can be retrieved with a Teflon® or stainless-steel bailer or can be pumped to the surface with a double-valve pump.

The Hydrocone™ (In Situ Technology, Orlando, FL, USA) consists of a sampler body with a shielded, telescoping porous tip.[30,31] The sampler is pushed to depth and then retracted, exposing the porous frit to the groundwater. The rate at which the sample chamber is filled is monitored by a water level sensor and controlled by maintaining a balance between the natural hydrostatic pressure and a variable argon gas back-pressure. This process also minimizes the amount of volatilization that may occur for some organic analytes during sample collection. The Hydrocone™ is retrieved, and the sample is decanted from the sample chamber. The Hydro Trap™ sampler (Fugro Geosciences, Inc., Houston, TX, USA) is designed similarly to the Hydrocone™ and uses a shielded stainless-steel screen as a porous element.

The BAT™ sampler (Figure 2a–c)[32] has a stainless-steel drive point with a stainless-steel body and retractable sleeve.[33–35] The probe is driven to the desired depth, and the stainless-steel sleeve is retracted, exposing a filter tip to the groundwater. Filter tips of high-density polyethylene (HDPE), sintered ceramic, and porous polytetrafluoroethylene (PTFE) were originally designed for the system,[34] although Type 316 stainless-steel filters have also been used.[36] The sample chamber has a double-ended hypodermic needle connection. A sampling probe containing a pre-evacuated vial with a septum-lined cap is lowered to the sample chamber and pierced by the needle, forming a leak-tight connection with the sample chamber. The sample is collected under in situ hydrostatic pressure. The vial is retrieved, making it unnecessary to transfer the sample to another container.

The Cone-Sipper™ (ARA) is able to collect samples at multiple depths during a single penetrometer push and is able to accommodate the electronic cone.[37] A specially designed two-stage filter that is exposed during the penetrometer sounding removes coarse- and fine-grained materials from the water (Figure 3). Water is drawn into lower and upper sample chambers by using a combination vacuum/pressure line. The two chambers are separated by a ball check valve. During the pressurization cycle, the check valve prevents water from returning to the lower chamber while water is being drawn from the upper chamber to the surface through Teflon® tubing. The water can be collected in vials at the surface, or the Teflon® sample line can be directly connected to the purge vessel of a purge-and-trap concentrator to provide an on-line analytical capability. The sampler is cleaned between sample collections by pressurizing the device with an inert compressed gas and purging the system out through the filter. Distilled water is then forced through the system in an additional cleansing step.

3.2 Soil

The Gouda™ and the MOSTAP™ samplers were originally developed to be used with the cone penetrometer to obtain soil cores for geotechnical analysis and are currently used to collect soil cores for chemical analysis.

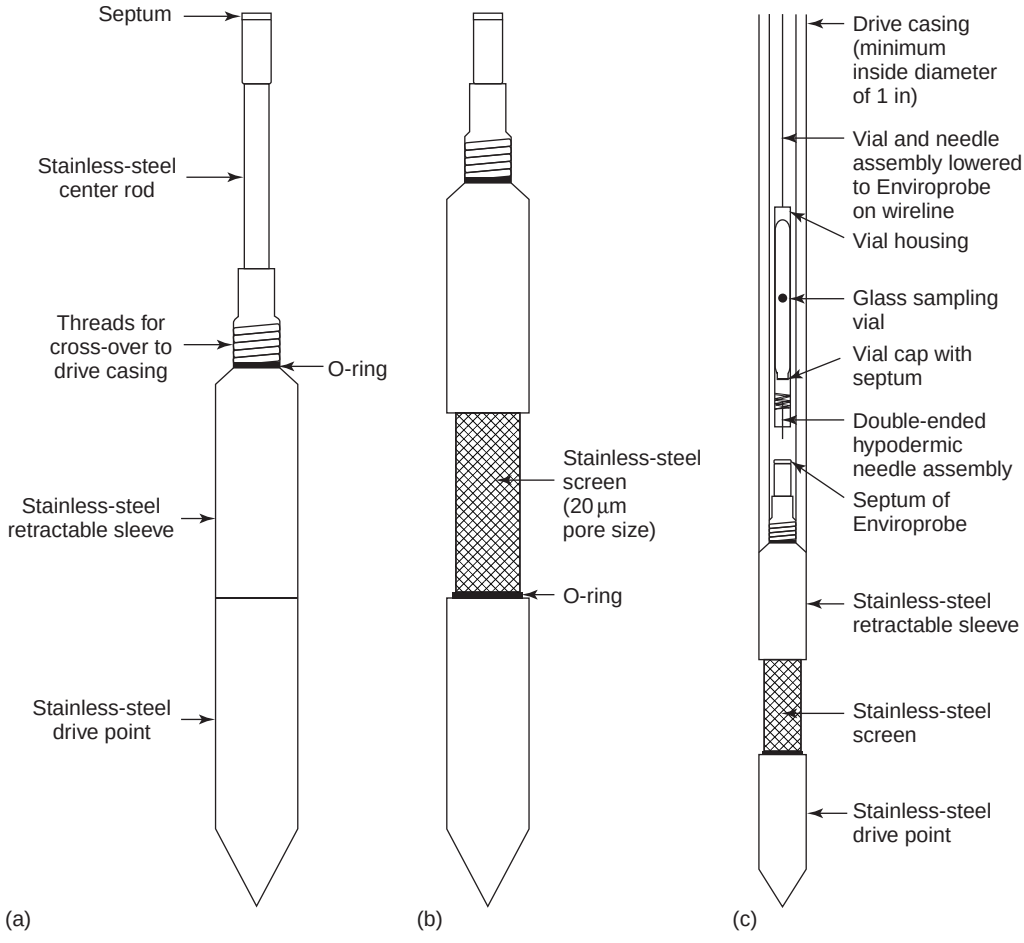

Figure 2 Schematic of BAT™ sampler for groundwater collection.[32] (a) Closed position; (b) open position; (c) sample collection configuration. (Reprinted by permission of the National Groundwater Association.)

The basic differences between the two devices are their disengagement mechanisms and the size of the soil cores that they are able to retrieve. The Gouda™ sampler isolates a soil core that is 25×195 mm ($\sim 95\,cm^3$), and the MOSTAP™ sampler collects a soil core that is 38×520 mm ($\sim 1924\,cm^3$).

The GEOSTAR™ soil samplers (Hogentogler & Co., Inc.) can retrieve soil cores that are 24×175 mm and 25×200 mm (Figure 4). The samplers are completely sealed with O-rings while the probe is pushed to depth. The rods are then retracted about 18 cm to withdraw the hardened stainless-steel tip into the body of the sampler, expose the cutting edge, and lock the device into the open position. The cone penetrometer is then advanced to collect the soil core. A vent in the top of the sample barrel is open during the push and then closed when the sample is retrieved to hold the sample in the liner. Fugro Geosciences, Inc. (Houston, TX, USA) also use a soil sampler that has a stainless-steel split barrel for retaining the soil core.

A soil sampler developed by ARA retains a soil core that is approximately 35×600 mm. The soil sampler is pushed to the desired depth and a tip release bar is lowered by a cable through the rods. The sampler is retracted to relieve the pressure at the tip and then the cable is pulled up to release the tip. The release bar is then pulled up and dropped on to the tip assembly. This action releases the latches and engages the tip release bar. The cable is then pulled up and the latches are disengaged from the tip release bar. The sampler is then advanced by pushing on the cone penetrometer rods to fill a metal or plastic sample tube. A split metal tube may also be used in the sampler. Upon retrieval of the sample barrel, the sample tube may be capped, or the sample may be extruded from the tube. Samples obtained with the split sample tube may be subsampled. An additional benefit of the sampler is the ability to stack the sample barrels to double the length of the core that can be collected.

3.3 Soil Gas

Soil gas is sampled within the vadose zone of the subsurface. The gas enters a port in the sampler

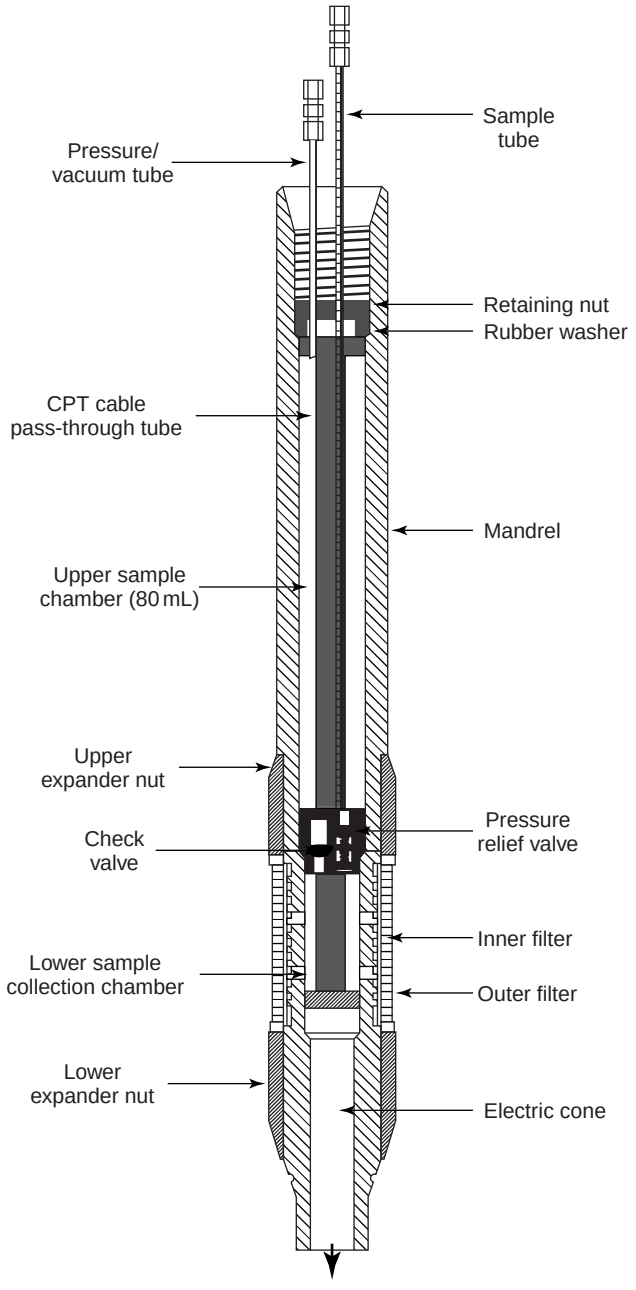

Figure 3 Schematic of Cone-Sipper™ (ARA, South Royalton, VT) probe for groundwater collection. The Cone-Sipper™ was developed jointly by the Westinghouse Savannah River Company (Aiken, SC) and ARA and is manufactured by Vertek (South Royalton, VT), a subsidiary of ARA, under license to Westinghouse Savannah River Company.

and is drawn through tubing to the surface by applying a vacuum. For VOC analysis, the sample is collected in a polymeric bag, glass bulb, or stainless-steel canister. The sample may also be concentrated on a solid sorbent or pass directly into a gas analyzer.

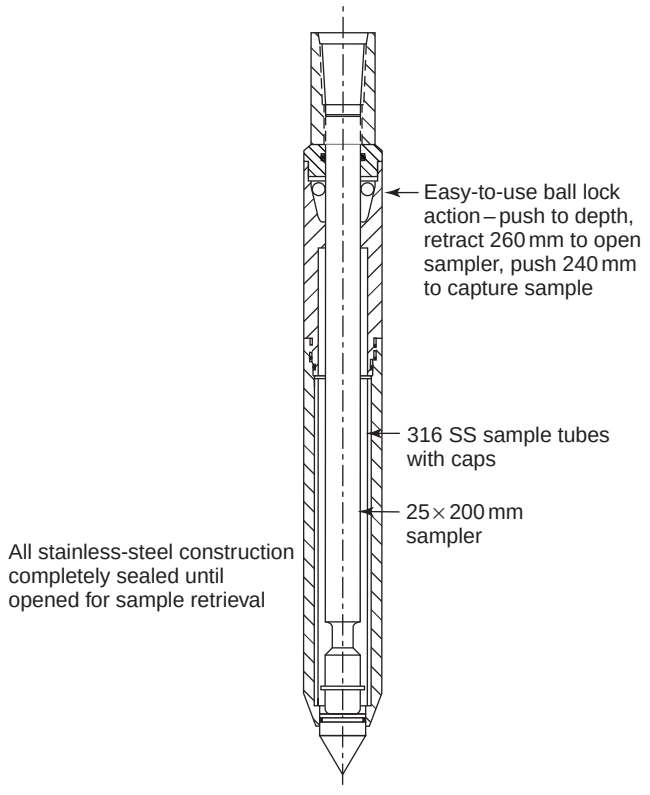

Figure 4 Schematic of GEOSTAR™ sampler (Hogentogler & Co., Inc., Columbia, MD) for the collection of soil cores.

The Cone Penetrometer/Vadose Zone Vapor (CP/VZV™) probe consists of a hollow barrel covered with a porous material that is positioned 46 cm above the tip of the cone.[38] The hollow barrel serves as a gas collection zone that is connected to an organic vapor meter or a field-portable gas chromatograph at the surface by Teflon® tubing. The BAT™, Cone-Sipper™, and HydroPunch™ devices can also be used for the collection of soil gas. Soil gas is efficiently collected in the preevacuated vial of the BAT™ sampler. In the Cone-Sipper™, a vacuum pump is attached to the vacuum line of the probe, and a sampling container or gas analyzer is attached to the vent tube to collect or analyze the sample. The pressurized line allows the sampler to be purged in between sample collections. The HydroPunch™ can also be operated similarly for the collection of soil gas.[39] Samplers that use a two-line vapor sampling design[13] to purge the vapor transfer lines between sample collections have also been developed (Hogentogler & Co., Inc.).

The multiport sampler (MPS), developed by the United States Army Corps of Engineers Waterways Experiment Station (WES) and shown in Figure 5, has 12 vertically stacked sampling modules that are independently operated from the surface to collect multiple vapor samples during a single penetration.[40]

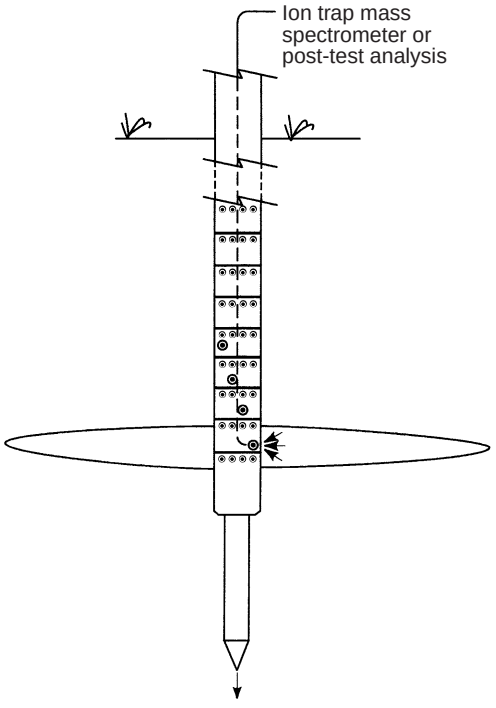

Figure 5 The MPS for the collection of soil gas. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

The MPS is advanced to the desired sampling depth, a module is selectively opened, and analyte is drawn through a sidewall port. An applied vacuum is used to open the ports and draw the sample through nylon tubing to the surface directly to an analytic instrument or to a solid sorbent to concentrate the sample. Replicate samples can be collected at a single depth, and different ports can be used during depth profiling to prevent cross-contamination of samples. In addition to the sampling modules, the MPS is capable of real-time soil stratigraphy mapping and also of grouting through the probe tip as the MPS is retracted to prevent cross-layer contamination. The MPS has been successfully used to sample vapors from soils contaminated with chlorinated organic solvents to determine the relative concentration of the contaminants in different soil strata.

4 ON-LINE ANALYTICAL TECHNIQUES

On-line analytical techniques have been interfaced with the cone penetrometer to quantify VOC and semivolatile organic compound (SVOC) contamination of the subsurface. These techniques include isolation, preconcentration, transfer, and chemical detection steps. This section will focus on the first three steps of the method, and chemical detection techniques will be discussed in section 5.

In situ isolation techniques have been developed to eliminate the transfer of groundwater, soil, and soil gas from in situ sampling probes to a container prior to chemical analysis. Erroneous results have been found, particularly for VOCs, when this transfer step is included in the sampling technique.[41] Following the isolation step, the organic contaminants can be concentrated at depth, analyzed in situ by a chemical sensor or instrument, or transferred through tubing to a chemical sensor or instrument in the cone penetrometer truck. Laboratory methods have been developed to isolate organic analytes from groundwater and soil by water purging and thermal desorption (TD) techniques, respectively. Cryogenic preconcentration, solid sorbent preconcentration methods, solid-phase microextraction, and membrane extraction techniques have all been used to preconcentrate the analyte before analysis. The analytes may also be transferred directly from the isolation probe to an instrument or chemical sensor at the surface. This technique allows for multiple sample collection from a single penetrometer push provided that the porous element can be cleaned between sample collections; however, a critical aspect of the on-line techniques is the quantitative transfer of analytes from the sample probe to the instrument or chemical sensor at the surface through tubings that can approach 100 m in length.

4.1 Isolation Techniques

Several samplers have been developed for the cone penetrometer that use TD to isolate organic analytes from soil. The thermal desorption sampler (TDS™) incorporates a TD system into a custom-designed soil sampler.[42,43] The TDS™ consists of a series of concentric steel cylinders with gas channels and piston chambers that are sealed by O-rings (Figure 6a and b). The probe is activated from a center actuator rod that is held in place by locking lugs. The device is pushed to depth, the lugs are pneumatically released, the piston is retracted to open the sample chamber, and the probe is advanced to capture a small plug of soil of known volume in the chamber. The chamber is heated to 150 °C by a Nichrome-wrapped ceramic heater that is protected with a stainless-steel sleeve. A thermocouple is used to monitor the temperature of the sample chamber. Preheated purge gas enters the sample chamber through the annular space between the outer housing and the actuator rod housing. The volatilized organics are transferred through the center of the actuator rod. The TDS™ was evaluated in the laboratory with soils to which chlorinated aliphatic hydrocarbons, aromatic hydrocarbons, and chlorinated aromatic hydrocarbons were added. Sand, silt, and clay soils containing 10 and 20% water were used in the laboratory tests. Analyte recoveries were all greater than 80%,

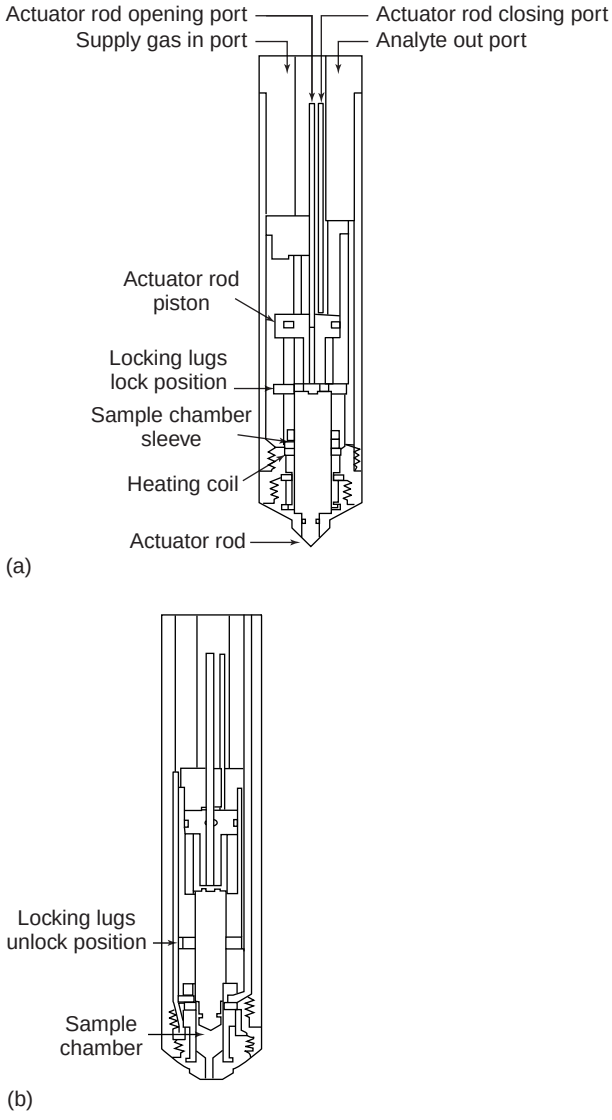

Figure 6 Schematic of the TDS™ for the isolation of VOCs from soil. (a) push configuration; (b) ready-to-sample configuration. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

but they were significantly lower in silt containing 20% moisture than in clay containing 10% moisture or in a mixture of sand and silt containing 10% moisture. Recoveries for benzene, trichloroethene (TCE), toluene, and chlorobenzene were 87, 87, 95, and 89%, respectively, and reflected the efficiency of drying the soil during TD. At 100 °C, the sampler effectively dried soils with moisture contents up to 17%. The drying efficiency was 95% for three soils containing 10% moisture and was 43% for silt containing 20% moisture. The baseline detection limit of the TDS™ is estimated to be 5 ppb for VOCs when using solid sorbent preconcentration with gas chromatography (GC). The estimated analyte carryover is 4%.

The Thermal Extractor for the Cone Penetrometer (TECP™) has been developed as a TD device for SVOCs in soil.[44] Cartridge heaters are used to heat the soil adjacent to the surface of the probe to approximately 250 °C. The sample ports are covered with stainless-steel disks (20 μm). The SVOCs are volatilized from the soil during a 4-min TD step as a vacuum is applied to the sample ports to facilitate transfer of the analytes. Similar to the TDS™, recoveries of SVOCs from the TECP™ are highly dependent on the efficiency of heating the soil adjacent to the probe. Laboratory investigations with soil containing a moisture content of 22% with the cartridge heater temperature set at 450 °C indicated that the temperature of the 20-cm long probe was 400 °C. Under these conditions, a maximum temperature of 250 °C was attained in the soil at a distance of 2 mm from the probe. The zone of collection is estimated to be approximately 2 cm^3. The volume of soil that is analyzed by the TECP™ has to be estimated because, unlike the TDS™, a known volume of soil is not captured by the device. Laboratory evaluation of the TECP™ using dry sand containing a mixture of polycyclic aromatic hydrocarbon (PAH) compounds, organochlorine pesticides, and polychlorinated biphenyl (PCB) compounds at 12, 15, and 28 ppm, respectively, indicated that recoveries for these SVOCs were 60–90%. The TD efficiencies decreased as the moisture content increased because of a decrease in the heating efficiency of the probe. At a moisture content of about 20%, recoveries were 35–50% from sand. Recoveries of 20–40% were achieved with a fortified soil sample containing 10% moisture and were a few percent lower in a coal tar sample containing 13% moisture. Estimated detection limits of the device are approximately 500 ppb by cryogenic preconcentration TD/GC/mass spectrometry (MS).

ARA developed the Dynamic Thermal Desorption™ (DTD™) probe to desorb VOCs from soils of the vadose zone.[45] The DTD™ probe contains a heating sleeve and a specially designed sampling port. The sampling port is covered by a screen consisting of a lengthwise slotted steel sleeve that is backed by a sintered steel filter, identical with the filter of the Cone-Sipper™. The slotted steel sleeve protects the sintered steel filter during the penetration and the sintered filter keeps fine material from blocking the port. The heater is composed of a heating sleeve, a heating wire, and a thermal insulator. The heated sleeve warms the soil ahead of the sample port to volatilize the VOCs. The heated soil gas is continuously drawn through the sample port and into the probe by vacuum.

Several devices have been developed to isolate VOCs from groundwater by an in situ purging method. Design considerations for in situ purge devices (ISPDs™) have been presented.[46,47] An in situ device for purging

VOCs from groundwater in monitoring wells has been developed in which water enters the sample chamber in the device through a sintered metal filter in the base of the chamber and a normally closed solenoid valve.[47] The top of the chamber is vented to the atmosphere. Several different methods were evaluated to detect the level of water in the sample chamber and initiate the purge cycle. The filling sequence is initiated by opening the valve on the top of the chamber. The level-sensing circuit is made up of three electrodes: one electrode at the base of the sample chamber, a second electrode slightly above the base, and a third electrode near the top of the chamber. A potential difference is applied between the base and the other two electrodes and the current flow is monitored. The valve is closed once a current is detected. Other level-sensing devices that were proposed included pressure switches, optical monitoring, and mechanical floats. The sample is purged by pulling a vacuum on the system through a sorbent tube to trap the volatile compounds.

The ISPD™ integrates a miniaturized purge device with the cone penetrometer.[46,48] The purge device is machined as four separate Type 316 stainless-steel sections (Figure 7). The sample section contains the purge chamber, purge line (with 60-μm porous sintered nickel frit), conductivity level sensor, and connection for the transfer line. The insert sleeve contains the mixing and purge line inserts. The mixing insert positions the purge line at the base of the water in the sample section. The inlet section aligns the mixing insert connection to two miniature electrically actuated two-way valves. A hole large enough to house the connections for the CPT runs through the purge device. A proprietary frit (Bladon International, Oak Brook, IL, USA) is used as the groundwater inlet. The valving is controlled from the surface and consists of miniaturized valves that are also connected to a conductivity detector that closes a valve when a 5-mL volume of groundwater has entered the sample chamber. The purge cycle is initiated and the analytes are transferred to the surface through stainless-steel tubing that is heated to 50 °C. The analytes are either collected on a sorbent cartridge or transferred directly into an instrument for analysis. The advantages of this technique are that (1) a known volume of groundwater is purged and (2) the sampler can be used for depth profiling. Laboratory testing and preliminary field testing of the device have been completed.[48] Laboratory testing of the purge device was performed with a mixture of halogenated aliphatic hydrocarbons and aromatic hydrocarbons at levels of 1–1000 ng L^{-1}. Recoveries were all >90%, with <1% carryover between samples. Evaluations of the device in a monitoring well indicated that the sampling precision and carryover were 14% and <5%, respectively, for groundwater containing 420 ppb of 1,1-dichloroethane. In another field test, good agreement

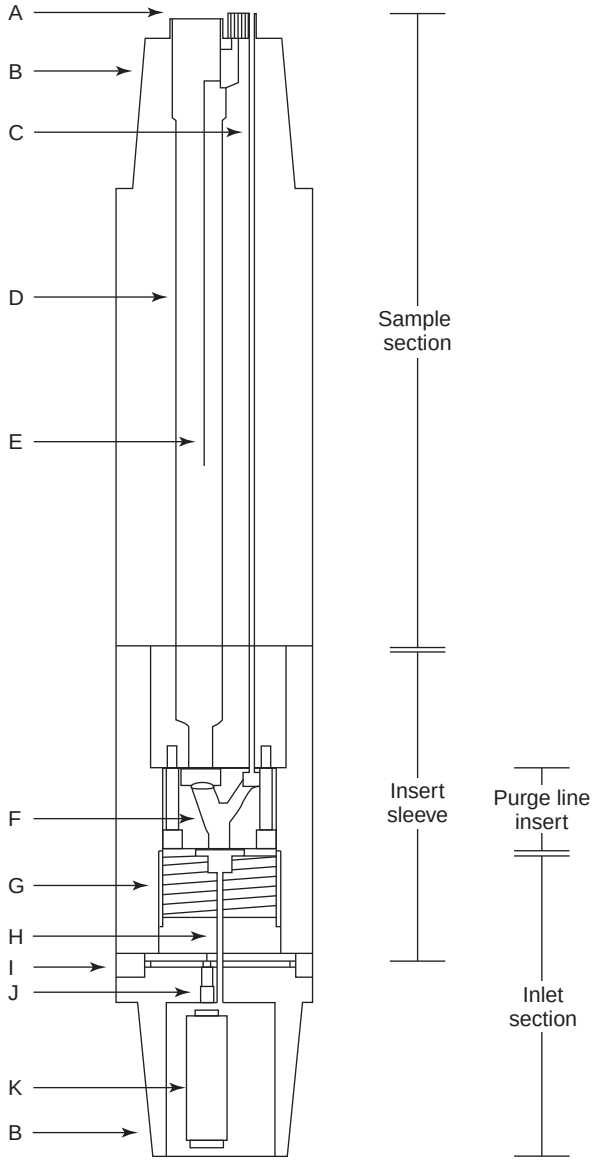

Figure 7 Schematic of ISPD™ for the isolation of VOCs from groundwater.[48] A, transfer line connection; B, threading to cone penetrometer rods; C, purge gas line; D, sample chamber; E, level sensor; F, base of water column; G, inlet; H, conduit to inlet; I, inlet frit; J, valve manifold; K, valve assembly.

was found for samples collected from a nearby monitoring well with a bailer and for samples analyzed in situ by the ISPD™ in a cone penetrometer, with the exception of *cis*-1,2-dichloroethene, which was much higher in samples analyzed by the ISPD™.

The cone penetrometer in situ purging (CISP™) probe developed by ARA is similar to the ISPD™.[49] The CISP™ probe obtains a sample by allowing a fixed volume of groundwater to enter a stainless-steel chamber in the probe through a normally closed solenoid valve. The sample is driven from the soil into the probe by ambient

hydraulic pressure and is purged with nitrogen. A two-stage filter is used at the inlet, which has a slotted sleeve and sintered screen that is the same as the inlet used with the Cone-Sipper™. A conductivity level sensor is used to initiate the purge cycle.

The Hydrosparge™, developed in a joint effort between Oak Ridge National Laboratory (ORNL) and WES, is a miniaturized purge device that is designed to be used with the HydroPunch™ and PowerPunch™ samplers (GeoInsight, Clayton, CA, USA) to isolate VOCs from groundwater.[50] The PowerPunch™ is based on the design of the HydroPunch™ and is used to insert small-diameter monitoring wells by using the cone penetrometer truck.[51] The device is capable of rapidly obtaining large volumes of water because of its adjustable screen length. Water is allowed to enter the cavity of the sampler, and the level is monitored with a conductivity sensor. The Hydrosparge™ is lowered into the sampler after the water level has become stable. The analytes are purged from the groundwater and transferred to the surface through Teflon® tubing.

Other techniques have been used in the laboratory for the isolation of VOCs from soil and groundwater but they have not been interfaced with the cone penetrometer. These techniques include the spray-and-trap method[52] and headspace analysis. The spray-and-trap method employs dynamic stripping of the VOCs by gradually spraying the water sample into a recipient vessel through a surface expansion nozzle. The time required for complete extraction of VOCs is about 4 min and is less than the 11 min recommended for complete extraction by the purge-and-trap technique. In addition, unlike the purge-and-trap method, the spray-and-trap method can efficiently isolate VOCs from waters containing surfactants. Headspace analysis has been applied to soil and groundwater samples, but an in situ device that employs this isolation technique has not been developed. The technique relies on an accurate determination of the VOC concentration in the gaseous headspace and the use of Henry's law to determine the concentration of the VOC in the aqueous phase. The Henry's law constant is the ratio of the equilibrium concentration of an analyte in air above an aqueous phase to its equilibrium concentration in the aqueous phase. The Henry's law constant is approximated as the ratio of a compound's vapor pressure and its aqueous solubility. Consequently, an analysis of the concentration of an analyte in the gaseous headspace can be used to derive the concentration of that analyte in the aqueous phase from the Henry's law constant. For groundwater samples, the analysis is straightforward; however, for soil samples, the soil would have to be immersed in an aqueous phase and vigorously mixed prior to sampling of the headspace. This procedure may be too complicated to interface with the cone penetrometer.

4.2 Preconcentration Techniques

Organic compounds in soil, soil gas, and groundwater are typically preconcentrated before analysis. Several methods have been used for this purpose and include cryogenic and solid sorbent preconcentration, solid-phase microextraction, and membrane extraction techniques. These techniques can be coupled with the in situ isolation techniques to provide an on-line analytical capability for the cone penetrometer.

Organic analytes are cryogenically preconcentrated at subambient temperatures on solid substrates such as glass, fused silica, or quartz wool and on solid, fused-silica, or porous-silica glass beads. For VOCs and SVOCs, the solid substrate is typically cooled to −190 and −60 °C, respectively. These techniques typically use a two-position valve to direct gas flows during the trapping and desorption steps. A resistive heater is used to heat the substrate at rates of ~400 °C min^{-1} for efficient TD. A cryointegration technique has been developed that does not use sample loops or mechanical valves in the sample flow path.[53,54] The sample is pulled through the cold trap with a vacuum pump that can operate from 20 torr to the pressure of the carrier gas. The trap is cooled from −100 to −200 °C. The sample is drawn through the cold trap and is then desorbed while being backflushed. The sample trap is heated by capacitive discharge. Bare metal trap tubes consisting of a 15-cm length of 0.30-mm ID Cu (30%)–Ni (70%) alloy and porous polymer adsorbent-lined trap tubes have been used. The adsorbent-lined trap tubes are used for the most volatile analytes. Organic analytes that are isolated from groundwater by purge-and-trap methods or from moist soils by TD in an inert gas may also contain significant amounts of water that may plug the cryogenic trap. One of the drawbacks of the cryointegration technique is that it is limited to a sample volume of about 2 mL for water-saturated air. Consequently, gas streams may require a Nafion® dryer to remove the water prior to cryogenic preconcentration. Cryogenic preconcentration devices are large and have not been miniaturized for use in the cone penetrometer; however, the devices can be directly interfaced to an instrument or chemical sensor at the surface as part of an on-line analytical method.

Solid sorbent methods have been used extensively to preconcentrate VOCs in gas streams. Typical sorbents include charcoal, Tenax®, graphitized carbon black, and mixtures of materials in the same sorbent bed. The hydrophobic nature of these sorbents prevents water from being collected in most applications, with the exception of the TD of some wet soils. Graphitized carbon black has been found to be suitable for a wide range of volatile organohalogens in soil gas[55] and Tenax® has been used to collect VOCs that have

been desorbed from soil.[43] The ORNL arrayed sampler for collection of VOCs in soil gas[56] uses multisorbent cartridges containing Carbotrap™, Carbotrap™-C, and Carbosieve™ S-III (Supelco, Bellefonte, PA, USA). The sorbent can be located in either a sorbent tube in the cone penetrometer,[49] as in, for example, the Downhole Trap Module (DTM™) (ARA), or at the surface directly interfaced to an instrument in an on-line analytical technique.[46,48] The sorbent is usually thermally desorbed while backflushing with an inert gas. For applications where the sorbent contains an excessive amount of water, the sorbent can be extracted with a polar solvent, and an aliquot of the extraction solution can be analyzed.

The use of solid-phase microextraction techniques for the analysis of VOCs in soil and water by headspace analysis and for the analysis of water directly has been described in detail.[57,58] Briefly, a fiber is coated with an organic phase that is similar to those used as stationary phases for GC columns. The fiber is allowed to equilibrate with the media and is analyzed by inserting it into the heated injection port of a gas chromatograph. The method has to be calibrated to determine the concentration of the analyte in the solid or aqueous phase. The method has not been interfaced with a cone penetrometer, but it may find some application for the headspace analysis of media isolated in the cone penetrometer.

Membrane extraction methods use membranes in the form of disks or tubings. Disk technology has been used for the extraction of SVOCs from water and air.[59–62] PTFE disks enmeshed with 8-μm diameter C_{18}-bonded silica have been used for the extraction of SVOCs. The analytes are eluted from the disk with a solvent. Silicone-rubber tubing has been used in membrane introduction mass spectrometry (MIMS) for selective transport of analytes into a mass spectrometer.[63] Water that may interfere with the analysis is retained by the membrane. The technique has been used with success for the analysis of VOCs in water. In the laboratory, silicone-rubber membranes have also been suspended in the headspace above liquids and in the aqueous phase as an inert gas is passed through the tubing to extract the analytes from the media.[64] This technique could possibly be used as part of an on-line analytical technique in, for example, the HydroPunch™ sampler, to extract VOCs from groundwater.

4.3 Transfer Techniques

Analyte transfer techniques are used to transport the analytes in an inert gas from the cone penetrometer to a chemical sensor or instrument at the surface. These techniques are necessary when the incorporation of preconcentration devices or chemical sensors and instruments into the cone penetrometer rods is not possible. The most critical aspect of the approach is the quantitative transfer of the analytes through tubing that can approach 100 m in length. The transfer line must be continuous and strung through all of the rods that are assembled for the penetrometer sounding. One end of the line is connected to the isolation device in the penetrometer and the other end is connected to the chemical sensor or instrument in the cone penetrometer truck. The positioning of the rods in the rack of the cone penetrometer truck requires that the transfer line be bent in a radius of 15 cm or less as it is strung from the end of one rod into the next rod. Consequently, the tubing must be rugged enough so that repeated bending and straightening will not crimp or break the transfer line.

Transfer lines of fused silica or glass-lined stainless-steel tubing are routinely used in the trace analytical laboratory to interface sample concentrators with analytical instruments. Passivated metal tubings [ProSteel® (J&W Scientific, Folsom, CA, USA); Ultimetal® (Chrompack, Raritan, NJ, USA)] are now used to fabricate capillary columns for GC where ruggedness is a concern. For the analysis of VOCs in ambient air, samples are stored in passivated metal or fused-silica-lined canisters. Polymeric tubing such as Teflon® would be advantageous for use in cone penetrometer applications because of its flexibility. The transfer line materials must be inert to the analyte of interest. The transfer line must also be heated to a temperature that can maintain the analytes in the vapor state and prevent the condensation of water in the tubing. Condensation may sorb water-soluble analytes and restrict the gas flow. For VOCs, heating the tubing to a temperature above the dew point of water at the subsurface temperature may be sufficient. For SVOCs, the tubing must be heated much higher ($>200\,^{\circ}C$) to maintain the analytes in the vapor phase.

Various metal (e.g. stainless steel, aluminum, and nickel) and polymeric tubings [e.g. Viton® (polyvinylidene fluoride hexafluoropropylene), nylon (polyamide), Tefzel® (ethylenetetrafluoroethylene), PEEK (polyether ether ketone), and also Teflon® PFA (perfluoroalkoxy), Teflon® PTFE, and Teflon® FEP (fluorinated ethylene polypropylene)] have been evaluated for their ability to transport VOCs in the gas phase.[42,43,46,65] The results indicate that stainless steel is superior to nickel, aluminum, and the polymeric tubings. Polymeric tubings contain varying levels of organic contamination and the outgassing of these contaminants increases when the material is heated. In addition, the transfer efficiencies of the VOCs through polymeric tubing decrease as the aqueous solubility of the analyte decreases. The analytes partition into the polymer and can be recovered; however, volumes of purge gas that greatly exceed the interior volume of the tubing are necessary to recover the VOCs completely and to minimize sample carryover. Additional

purging of the transfer line will decrease the speed of sample collection and complicate the preconcentration of VOCs at the surface by on-line analytical techniques. Metal tubing apparently contains active sorption sites that are efficiently deactivated by water molecules. Transfer efficiencies for VOCs through stainless-steel tubings are nearly 100% in moist gas streams; however, aluminum and nickel tubing retain some VOCs under these conditions. Silcosteel® tubing (Supelco, Bellefonte, PA, USA), a type of glass-lined stainless-steel tubing, may be a suitable material for a cone penetrometer transfer line because of the inertness such tubing has exhibited in laboratory applications and because of its flexibility. Silcosteel® tubing is now used in a transfer line for SVOCs,[43] but the flexibility and ruggedness of tubings having ODs of 0.159 and 0.318 cm have to be evaluated. Passivated metal tubing may be suitable as a transfer line material if it can be produced with a wider ID than a capillary column so that the gas flows that are needed for rapid transfer of analytes can be accommodated.

Parker Hannifin Corp. produces a commercially available heated transfer line (Temptrace®; Parflex Division, Ravenna, OH, USA). Heat-traced lengths of type 316 stainless-steel tubing (0.318 cm OD) have been used in cone penetrometer applications.[48] The tubing has a 24 W ft^{-1}, 120 V, braided constant-wattage heater and aluminum heat-transfer tape that allow the tubing to be heated to 150 °C. The product contains nonhygroscopic fiberglass insulation and a Mylar® moisture barrier and has a urethane jacket. The nominal OD of the heat-traced tubing is about 1.27 cm, which makes it narrow enough to fit within the cone penetrometer rods (2.54 cm ID) that are most commonly used.

ARA developed a unique heated transfer line by using a self-regulating parallel resistance heat trace that is located immediately next to Teflon® PFA tubing.[49] The temperature of the heat trace is monitored by a thermistor located midway along the line. The heat trace and the sample transfer line are encased in a braided fiberglass insulator to prevent heating of the purge gas line and to reduce the temperature so that the line can be handled by the cone penetrometer operator. The heated portions of the line can connect with a second Teflon® PFA tube for supplying the purge gas and electrical leads. The wires and tubes are packaged in a protective plastic tubing to prevent excessive wear and tangling. The line can be heated to 150 °C.

A unique transfer line has been developed that can be heated to 280 °C for the efficient transfer of SVOCs, particularly PAHs, PCBs, and organochlorine pesticides.[44] The transfer line uses Silcosteel® tubing (0.159 cm OD) and is heavily insulated so that it can be safely handled. The tubing is held within high-temperature sleeving and is coated with aluminum foil glass cloth. The tubing is insulated and supported by a bundle of Teflon® tubing. The outer jacket is made with tape and silicone adhesive.

5 CHEMICAL DETECTION TECHNIQUES

A variety of chemical detection technologies have been used to measure subsurface contaminants.[66,67] These technologies include sensors to provide in situ chemical detection and instrumentation such as gas chromatographs and mass spectrometers that are used in on-line analytical techniques with the cone penetrometer.

5.1 Chemical Sensors

A chemical sensor is a transducer consisting of a physical layer and a chemically selective layer that is capable of providing direct information about the chemical composition of its environment.[68] Chemical detection can be based on the mass of the analyte or its optical, electrochemical, or radiochemical properties.

5.1.1 Mass Sensors

Piezoelectric crystals are sensitive to mass and lack chemical specificity; however, they have been developed as chemical vapor monitors by applying a nonvolatile coating to the crystal that selectively interacts with chemical substances.[69,70] Several reviews are available on the application of these devices for the measurement of organic vapors.[71–75] In general, the devices lack specificity for individual chemicals, but arrays of detectors coated with different materials have been used with some success to obtain molecular information;[76–80] however, detection is typically by chemical compound class.

A portable acoustic wave sensor (PAWS) system has been developed for the real-time analysis of soil gas samples in the field and the in situ monitoring of contaminants in vadose-zone monitoring wells.[81] The system consists of a pump to facilitate gas flow across the sensor, a Nafion® dryer to eliminate water from the gas stream, and a three-way valve. The piezoelectric crystal is coated with polyisobutylene. The instrument was evaluated in the laboratory for its ability to quantify TCE and was found to have a detection limit of 8 ppmv, with a dynamic range of 55 000 ppmv. The response was reversible with a 1-s recovery time. Minimum detection limits of 1–10 ppmv were established for many VOCs and were reduced to 10–100 ppbv with a preconcentrator. The system was evaluated in the field in monitoring wells containing tetrachloromethane (CCl_4). At levels of 100–20 000 ppmv, the results were within 2% of those determined with a photoacoustic infrared (IR) system.

The detection limit for CCl_4 was 10 ppmv. Calibrations were stable and were within 1.5% before and after the field measurements. The system has not been adapted for use in the cone penetrometer.

5.1.2 Optical Sensors

Optical sensors are interfaced with the cone penetrometer with fiber optics that are strung through the penetrometer rods to transmit light to and from the sample media. The fiber-optic chemical sensor (FOCS™) uses a spectroscopic measurement of the signal.[82,83] The basic technique of a FOCS™ can involve (1) a direct spectroscopic measurement of the analyte, (2) reversible binding of the analyte with indicator molecules, or (3) chemical and biochemical reactions of the analyte.[67]

5.1.2.1 Direct Spectroscopic Measurement

Laser-induced Fluorescence Sensor. The first fiber-optic spectroscopic sensor for in situ detection of chemical contaminants by using cone penetrometers was based on measuring the fluorescence of POLs when excited by pulsed ultraviolet (UV) laser energy. That first sensor was developed for SCAPS by the WES[84] and was adapted from a sensor designed by the United States Navy's Naval Command, Control, and Ocean Surveillance Center; Research, Development, Test and Evaluation Division (NRaD), for seaborne applications. The first-generation SCAPS LIF sensor, shown in Figure 8(a) and (b), consists of a probe with dual fibers for excitation and emission, and a flush-mounted 6.35-mm diameter sapphire window for transmitting the laser illumination to the outside of the penetrometer probe. The UV energy source for this system consists of a pulsed nitrogen laser with an output of 1.4 mJ, 0.8-ns pulse width, 10-Hz repetition rate, at 337 nm.[1,85] This excitation source represented a mature, commercially available technology that could be readily incorporated into a field LIF system at the time that this effort was initiated in 1989. Another advantage of the 337-nm excitation source, identified at that time, was that fiber attenuation of energy at this wavelength is relatively modest, permitting transmission of the excitation energy over fiber lengths of 100 m or more. This permits the LIF sensor system to be used to the full push capability of the standard 20-ton penetrometer system (i.e. 50 m).

The system uses a spectrograph coupled to an intensified photodiode array and optical multichannel analyzer (OMA) detection system. Because this system uses a diffraction grating to disperse the fluorescence emission signal over a 1024-element diode-array detector, reading out an entire spectral signature from the fluorescing fingerprint signal is possible with a resolution of 0.5 nm or better in approximately 16 ms. In practice, the system usually integrates the emission from 10–20 laser shots, which requires a total analysis time of 1–2 s to collect a fluorescence emission spectral signature. In addition to the optical sensors, the probe also incorporates geophysical sensors that provide continuous soil classification and stratigraphy during the penetrometer push. The LIF probe is also equipped with a grouting system that is used to seal the hole as the probe is being retracted, in order to prevent migration of contaminants across soil layers.

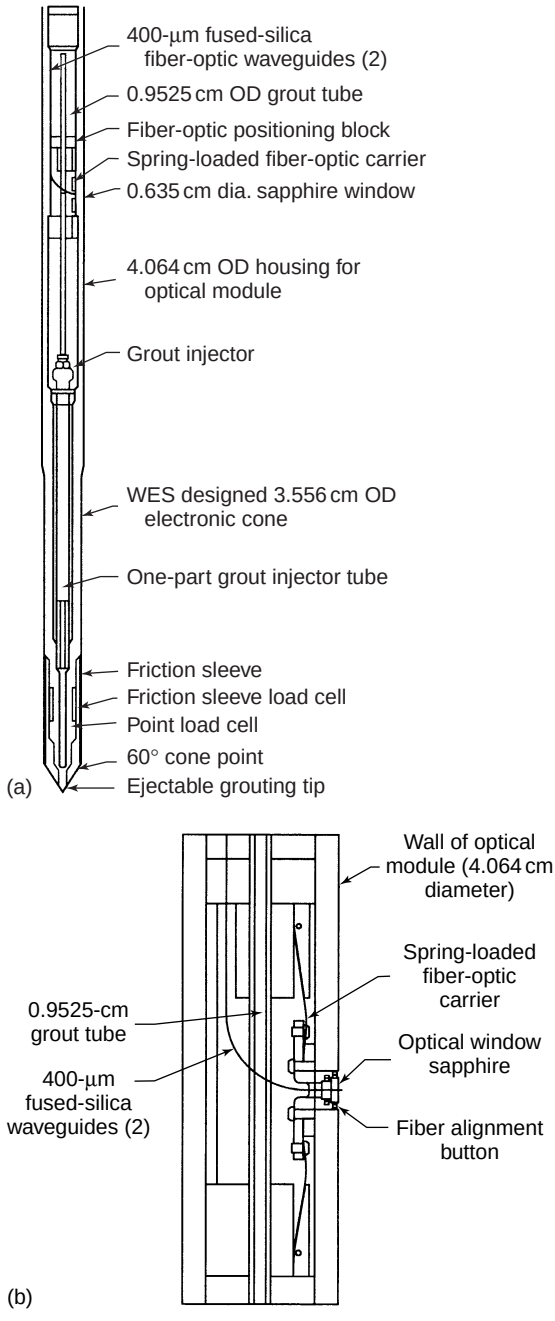

Figure 8 (a) Cross-sectional view of the SCAPS LIF probe, with (b) a close-up of the optical window. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

When advanced at the 2 cm s^{-1} standard push rate, the LIF probe collects contaminant information and soil classification data with a spatial resolution of 2 cm. Figure 9 presents an example of the data that are collected by the SCAPS LIF probe. These plots, which are displayed in real time during the penetrometer push, show the geophysical sensor information that is used to compute the chart of soil classification vs depth, and also a summary of the LIF data (fluorescence intensity and wavelength at peak vs depth). This particular penetrometer push indicates an area of significant POL contamination in a layer 4.2–4.4 m below the ground surface. In addition to the data displayed in Figure 9, the SCAPS data acquisition system collects and stores the complete fluorescence spectrum (1024 points from 270 to 800 nm) for each laser pulse (or average of pulses). After a representative number of penetrometer pushes, the geophysical and fluorescence data are used to develop a model of the subsurface contaminant plume, as shown in Figure 10. The first-generation SCAPS LIF probe was demonstrated in the field in 1990[85] and has been extensively used to characterize USDOD and USDOE sites. In addition, this technology has been licensed to private industry, and a number of commercial firms are currently using similar cone penetrometer systems. The SCAPS LIF system has undergone extensive field tests and demonstrations at a number of sites and has proven to be an effective screening tool for mapping subsurface plumes of POL contaminants.[86]

A limitation of the 337-nm excitation source identified early in the SCAPS development is that it does not induce fluorescence in single-ring aromatic compounds such as the benzene, toluene, ethylbenzene, and xylene (BTEX) components of fuels. In fact, the lightest weight polycyclic aromatic compound that is excited appears to

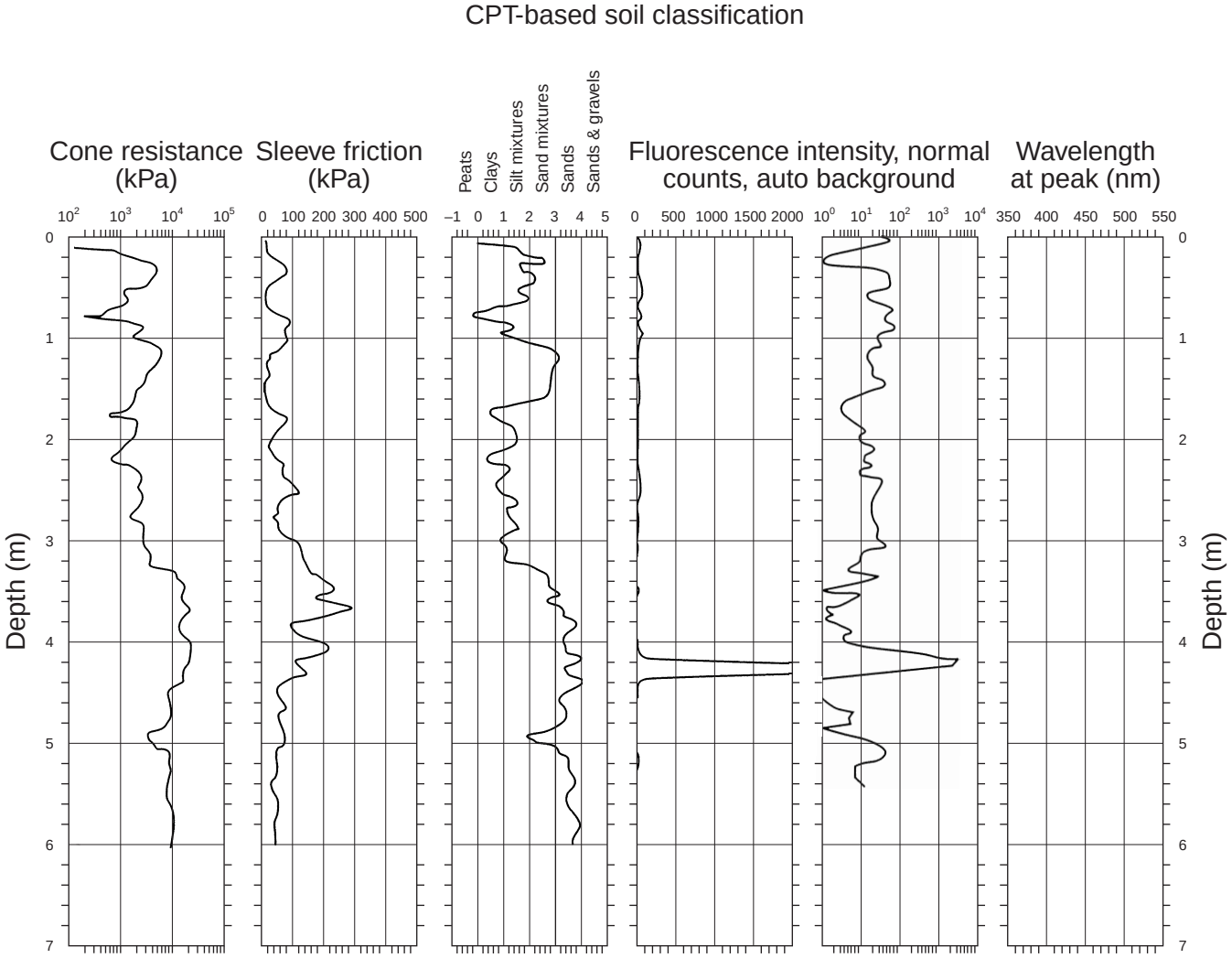

Figure 9 Typical data output of the SCAPS LIF probe at a petroleum-contaminated site. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

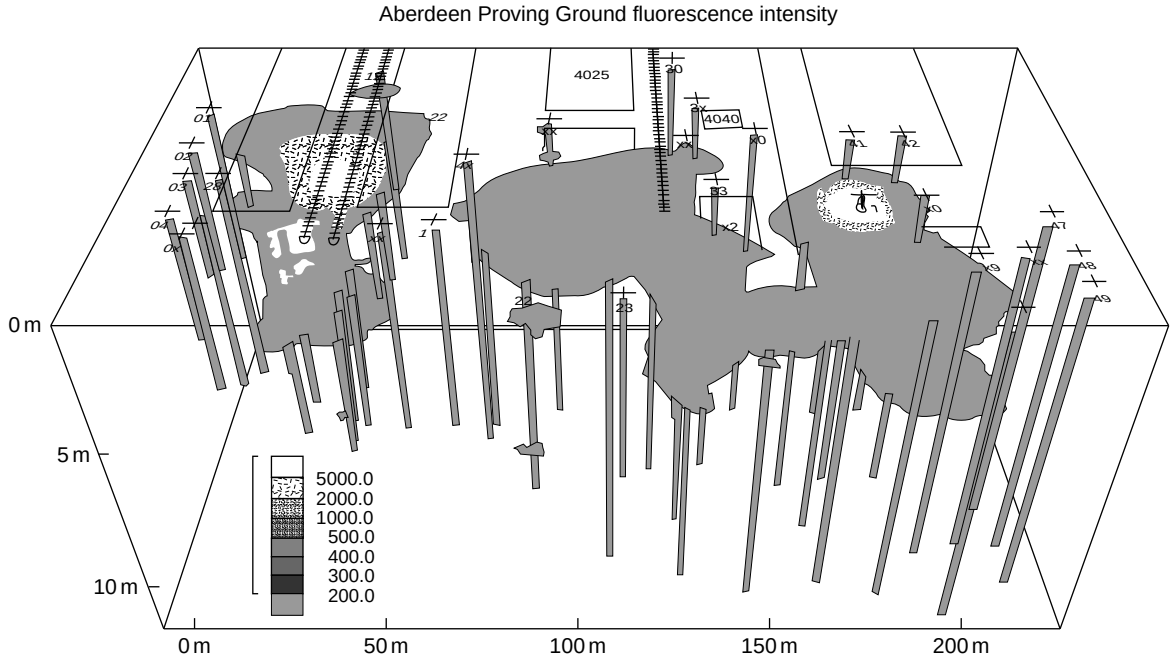

Figure 10 Three-dimensional visualization of LIF probe data, showing underground plume of POLs. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

be the disubstituted naphthalenes. In general, the 337-nm source is most effective for inducing fluorescence in PAHs with three rings or more.[87] This lack of capability of the nitrogen-based LIF sensor to detect BTEX and light fuels (e.g. the jet fuels JP-4 and JP-5 and unleaded gasoline) has led to a number of efforts to test shorter wavelength sources for incorporation into SCAPS.[88] One of the second-generation SCAPS LIF probes resulting from these efforts is a tunable LIF SCAPS probe currently designated as the Rapid Optical Screening Tool (ROST™). The ROST™ was originally developed for the US Air Force's Armstrong Laboratory at Tyndall Air Force Base (AFB) by North Dakota State University (NDSU) and is currently produced commercially by Fugro Geosciences, Inc. The system was initially designed as a field-portable LIF sensor to analyze liquid samples containing aromatic hydrocarbons and fuels in monitoring well applications. Subsequently, the sensor was installed in a cone penetrometer probe and was field-tested in a limited number of sites.[89] The system makes use of a frequency-doubled Nd : YAG laser to pump a dye laser the output of which is then frequency-doubled to generate UV light in the range 230–300 nm. In this system, the Nd : YAG laser can also be used directly with harmonic frequency conversion to produce 266-, 355-, and 532-nm outputs. The advantages of these alternative excitation wavelengths include the capability to excite lighter weight aromatic compounds (one- and two-ring aromatic compounds) and to select the most appropriate excitation wavelength for the particular contaminant of interest and the site-specific conditions. One penalty suffered by using shorter wavelength excitation light in the cone penetrometer is that energy losses in the transmission fiber increase dramatically at wavelengths shorter than 290 nm. This loss may preclude effective transmission of short-wavelength light energy over the length of fiber (100 m) required to reach the maximum push depth of a 20-ton penetrometer system. It does appear to be possible, however, to transmit sufficient quantities of short-wavelength light energy over shorter lengths of fiber that could be used to make measurements to depths of 30 m.

Unlike the fixed-wavelength LIF system, the NDSU system uses a monochromator, photomultiplier tube (PMT), and digital storage oscilloscope to detect fluorescence. A major difference in this approach, compared with the OMA detector used by the original SCAPS LIF system, is that in the dynamic or push mode, the detector measures the time decay or lifetime of the fluorescence signal at a single wavelength selected by the monochromator. Full spectral-time decay data matrices can be generated by moving the monochromator in 10-nm increments and recording the time-dependent fluorescence signal at each wavelength. With the present (50-Hz) laser system, this procedure requires approximately 2 min to collect a wavelength–time matrix (WTM) consisting of 21 different wavelengths with a spectral resolution of 10 nm.[90]

Another example of a SCAPS LIF probe that incorporates shorter wavelength sources is shown in

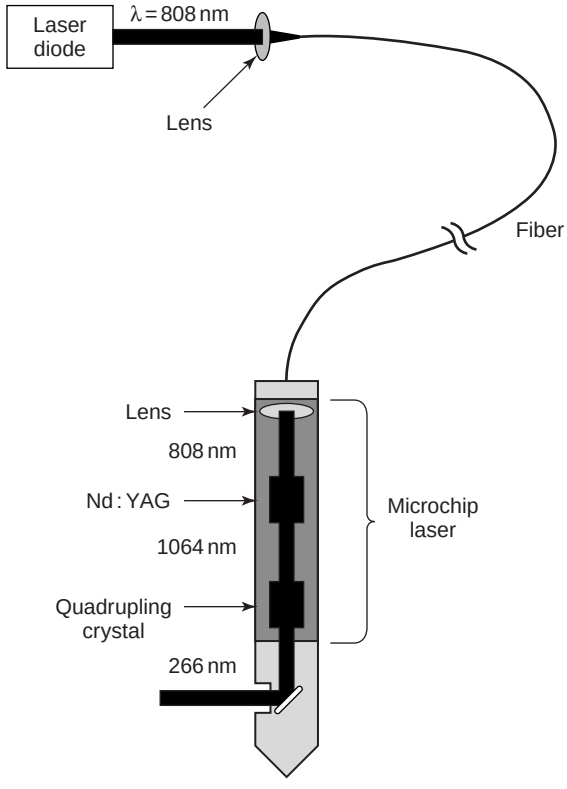

Figure 11 Schematic of LIF probe (226 nm wavelength) that incorporates a diode-pumped microlaser. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

Figure 11. This probe, recently designed and fabricated by NRaD, incorporates a diode-pumped Nd : YAG microchip laser developed by Massachusetts Institute of Technology/Lincoln Laboratory (MIT/LL) to generate 266-nm laser energy in the probe (thereby eliminating the fiber-optic delivery system). Initial field tests by the US Navy indicate the potential to detect jet fuel concentrations as low as 100 ppm.

Recently, Tufts University field-tested a variation of the SCAPS LIF probe to perform parallel LIF spectroscopy.[91] The Tufts system uses multiple-wavelength excitation resulting from Raman shifting of a pulsed Nd : YAG laser operating at 266 nm. The Raman shifter generates multiple wavelengths (both shorter and longer than the 266-nm pump energy), which are separated and launched into individual fibers to separate sapphire windows in the cone penetrometer probe. The fluorescence signals excited by the various excitation fibers are collected by separate fibers and routed to a charge-coupled device (CCD) spectrometer for analysis. The resulting excitation/emission matrix (EEM) is used to identify the POL type and concentration.

In addition to the LIF probes described in this section, a number of commercial firms, such as ARA, have developed and deployed LIF probes based on adaptations of the initial SCAPS LIF sensor design.

Laser-induced Breakdown Spectroscopy Sensor. The laser-induced breakdown spectroscopy (LIBS) sensor is used to detect heavy metals in contaminated soils and groundwater and is based on a simple plasma spectrochemical approach.[92] A high-peak-power laser pulse (typically a Q-switched Nd : YAG laser that produces approximately 10-ns laser pulses with up to 200 mJ pulse energy) is focused on a solid, liquid, or gaseous sample and produces a spark (optical breakdown). The spark emission light consists of unique spectral signatures from excited atoms, radicals, and ions generated from the sample. The plasma emission is collected by optical fibers or lenses and passed through a wavelength dispersive device such as a monochromator or acousto-optical tunable filter (AOTF). The wavelength-resolved light is then detected by a linear diode array, CCD detector, or a single point detector. The information on wavelength and light intensity is passed to a computer, and the data are analyzed by using sophisticated chemometrics software. The output can be used to identify the contaminant and determine its concentration.

The LIBS technique possesses many desirable attributes: (1) compact and lightweight sensors that can be carried by a single person; (2) inherent versatility such that, with the use of optical fibers, any solid, liquid, or vapor sample can be analyzed; (3) no sample preparation; (4) analysis is quick, particularly with the use of high-repetition-rate lasers, array detectors, and high-speed computers; (5) rugged and reliable sensors because all of the electro-optical components are solid-state, and the fiber optics can easily be replaced if necessary; and (6) less-expensive LIBS sensors are possible because of chip lasers and less expensive array detectors.

Figure 12 shows a typical LIBS set-up used in the laboratory for collecting data from liquid and soil samples. An Nd : YAG laser source delivers 80- to 100-mJ pulses at 10 Hz that are focused on an approximately 100-μm spot on the sample. The resulting emissions are collected and analyzed by a spectrometer with an intensified diode-array or CCD detector. Contaminated soil samples are placed on a turntable to expose an undisturbed surface to each laser shot. This arrangement mimics the actual cone penetrometer operation, in which the probe is moving while the laser is firing. An example of LIBS spectra of Yuma sand that has been spiked with various concentrations of Pb salts is shown in Figure 13. The variation in response at a wavelength of about 406 nm is proportional to the Pb concentration. Detection limits ranging from 0.5 to 11 ppm by weight for Pb, Hg, Cr, Cd, and Zn in sand have been achieved with this system.[93]

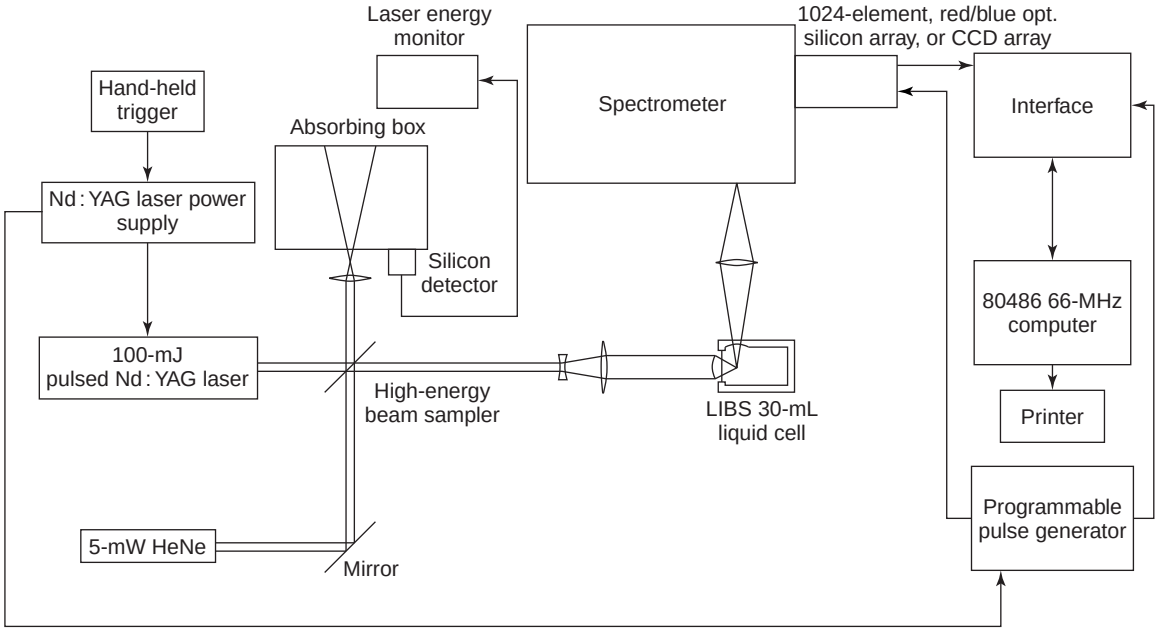

Figure 12 Laboratory arrangement of LIBS instrumentation. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

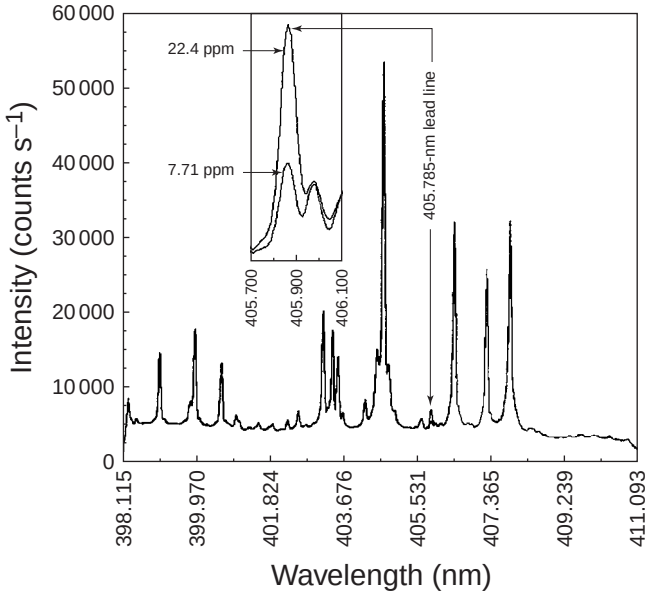

Figure 13 Spectra of lead-contaminated sand collected with a LIBS probe. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

Research groups at WES and NRaD have field-tested two different cone-penetrometer-based LIBS systems. The first system uses optical fibers to deliver the laser energy (up to 25 mJ at 10 Hz) from the laser in the truck through a sapphire window in the probe to the soil surface. The resulting emissions are collected by a fiber bundle and routed to the spectrometer and computer on the SCAPS truck. The second system, shown in Figure 14(a) and (b), uses a probe-mounted miniature Nd : YAG laser to deliver 80-mJ pulses at 0.33 Hz to the soil surface. The extent of the subsurface Pb contamination found at an Air Force firing range was recently mapped by using the downhole laser LIBS system and is shown in Figure 15.

X-ray Fluorescence Sensor. A cone-penetrometer-based X-ray fluorescence (XRF) system has been developed jointly by the Naval Research Laboratory and WES for detecting metal contaminants in soils and groundwater. The XRF sensors operate by detecting characteristic atomic emissions that result from exciting samples with X-rays. X-rays, which have higher energy than the binding energy of an atomic core electron, will excite core electrons to an empty state above the atomic states. The resulting empty core states are refilled by other electrons within the atom after emitting energy (fluorescing) in the form of lower-energy X-rays. Because the electron energy states producing the fluorescence are entirely within the atom, the X-rays are produced with a constant and well-known energy that is characteristic of individual atomic species.[94] As a result, XRF is sufficiently sensitive to detect many metals at levels of about 10 ppm and does not suffer from matrix effects (where the signal level depends on the exact nature of the chemical bonds involving the target heavy metal and neighboring species) as much as some of the more traditional chemical analytical techniques.

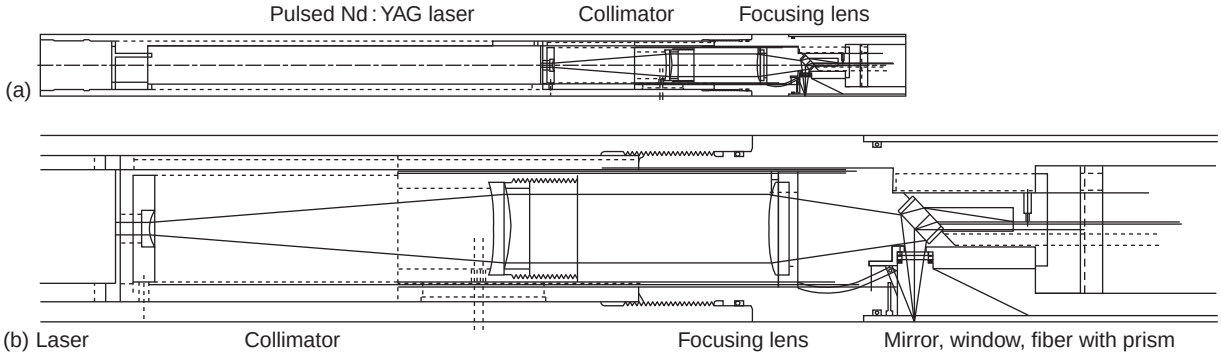

Figure 14 Schematic of a LIBS probe that incorporates a miniature Nd : YAG laser. (a) 1 : 4 scale including entire laser; (b) 3 : 4 scale with laser output only. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

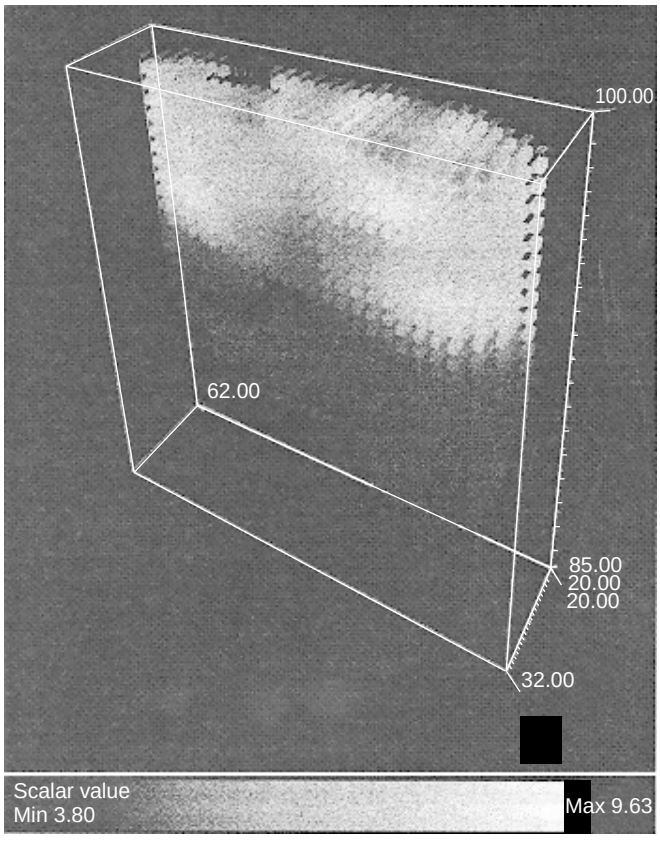

Figure 15 Three-dimensional visualization of LIBS probe data showing subsurface lead contamination at a small-arms firing range. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

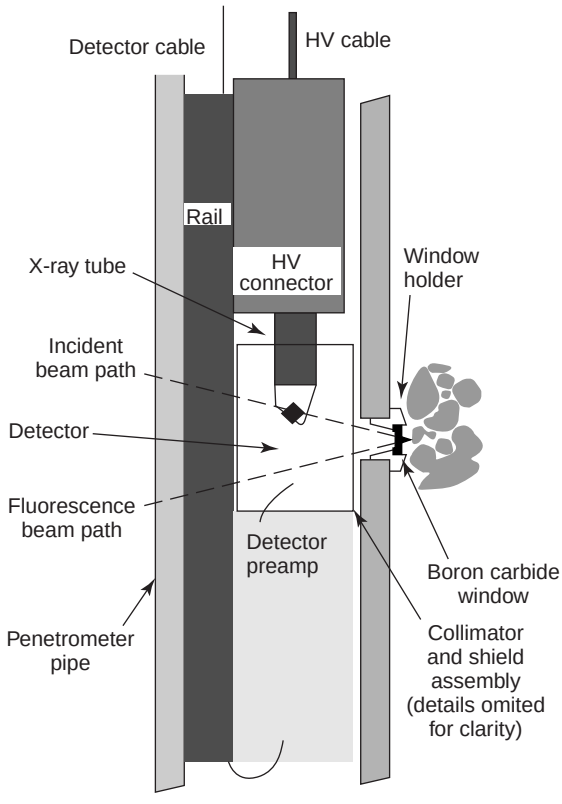

Figure 16 Schematic of XRF probe. HV = high voltage. (Courtesy of the Naval Research Laboratory.)

Implementation of an XRF sensor in the cone penetrometer has been made possible by recent advances in the miniaturization of X-ray source tubes, detectors, and the cooling system required for their operation. Figure 16 shows a schematic of the SCAPS XRF probe, which incorporates a miniature X-ray tube, a boron carbide window to transmit the X-ray excitation and collect the resulting fluorescence, and an energy-dispersive detector. The high-voltage power supply required to operate the X-ray tube is located uphole in the SCAPS truck. The probe has been thoroughly evaluated in the laboratory[95] and has undergone limited field testing.

Ultraviolet Fluorescence Probes. An in situ ultraviolet fluorescence (UVF) probe has been proposed to detect single-ring aromatic compounds in groundwater.[96] Unlike LIF, where the UV radiation is generated by

a laser, the UVF probe uses a mercury vapor lamp as the UV source. Preliminary evaluation of the method with a benchtop version of the probe has been made. A miniaturized lamp and PMT have been proposed for use in a cone penetrometer, which would preclude the use of fiber optics. The volatile aromatic hydrocarbon concentration is estimated from the ratio of the fluorescence of the contaminant–water mixture to the fluorescence of the contaminant–water mixture after it has been purged with an inert gas. Using the ratio of the fluorescence signals is necessary because the method is nonspecific, and fluorescence from humic substances would interfere. These substances are nonvolatile and cannot be purged from water; however, turbidity was found to interfere with the analysis. Therefore, the sample would have to be filtered before the measurement. The detection limit for a laboratory benchtop device was 500 ppb. Lower detection limits are expected for a miniaturized version.

Recently, a patent was awarded[97] for a cone penetrometer UVF probe to detect POL contaminants. In this design, a UV lamp located in the probe housing provides excitation energy, which is reflected through a UV-transparent window in the housing to the soil by one or more reflecting elements in the housing. The reflecting element or elements allow longitudinal offset of the lamp from the window, providing space for a high-intensity UV lamp. Fluorescence emitted from contaminants in the soil is transmitted through the window and collected by a fiber-optic link which, in turn, transmits it to the instrumentation at the surface. A concentrating reflector between the lamp and the housing and a focusing lens between the lamp and the reflecting element collect and focus excitation energy from the lamp. A filter between the lamp and the first reflecting element eliminates unwanted wavelengths from the excitation energy, and a filter between the window and the fiber-optic link removes unwanted wavelengths from the emission spectrum. The design should improve sensitivity, especially for the lighter POL contaminants, and eliminates the need for complicated and expensive laser excitation equipment at the surface.

Fiber-optic Raman Sensors. Raman sensing is a spectroscopic technique that is useful for detecting nonfluorescing chemicals such as solvents and other VOCs. Detection is based on measuring the wavelength shift of monochromatic light scattered from the sample. The measured wavelength shift (i.e. the Raman shift) is characteristic of the molecular structure of the sample. The drawback to this type of sensing is that the Raman signal is approximately six orders of magnitude weaker than the signal produced by LIF. Nevertheless, the Raman technique is suitable for detecting accumulations of chemical contaminants as light nonaqueous-phase liquids (LNAPLS) or dense nonaqueous-phase liquids (DNAPLS), which tend to collect in subsurface pools of almost pure product. The NRaD has developed and field-tested a SCAPS fiber-optic Raman sensor (FORS) probe for detecting pools of VOCs. The FORS system is very similar to the original SCAPS LIF system described previously. The FORS system employs a nitrogen laser coupled to a single delivery fiber and a sapphire window to excite the soil or groundwater outside the probe. In order to maximize the relatively weak Raman signals, seven collection fibers are used, instead of the single fiber used in the LIF probe. Each fiber is equipped with a filter module consisting of two graded index (GRIN) lenses to collimate light from the fiber and filters. The excitation fiber has a band-pass filter module to isolate the excitation laser line from emissions from the fiber. Each collection fiber has a long-pass filter module to reject the laser line while allowing the Raman-shifted signal from the sample to be transmitted. Figure 17(a–e) shows an example of data collected by the FORS system. The spectra shown are those of toluene in a sand matrix. Figure 17(a) shows the Raman lines of toluene superimposed on a fluorescence background, and also Raman emissions from the fibers. Because the fiber background has been well characterized, it can be subtracted, as shown in Figure 17(b). The fluorescence background signal can be eliminated either by taking the derivative of the second spectrum (Figure 17c) or by fast Fourier transform (FFT) filtering (high-pass filtering in the frequency domain), as shown in Figure 17(d). For comparison purposes, a Raman spectrum of neat toluene is shown in Figure 17(e).

To improve the sensitivity of FORS probes, a number of probes have been developed that make use of surface-enhanced Raman scattering (SERS), in which the analyte of interest is deposited on a specially prepared rough metal surface. The effect of SERS is to enhance the Raman signal by up to eight orders of magnitude. The NRaD is currently developing a SCAPS FORS probe equipped with organic-coated SERS substrates for selective detection of BTEX and chlorinated solvents; the detection capability is much improved over the original SCAPS FORS probe.[98]

Fiber-optic Infrared Reflectance Probe. Remote fiber-optic Fourier transform infrared (FTIR) spectroscopy is a direct spectroscopic technique that can be used to detect trace amounts of organic contaminants in subsurface media.[99,100] A fiber-optic infrared reflectance (FOIR) probe has been developed that illuminates the soil with a broad-band optic source through a sapphire window in the cone penetrometer.[101] The diffuse reflectance from the soil is collected through the window and transmitted through an optical fiber to the surface, where it is spectrally analyzed with an FTIR spectrometer

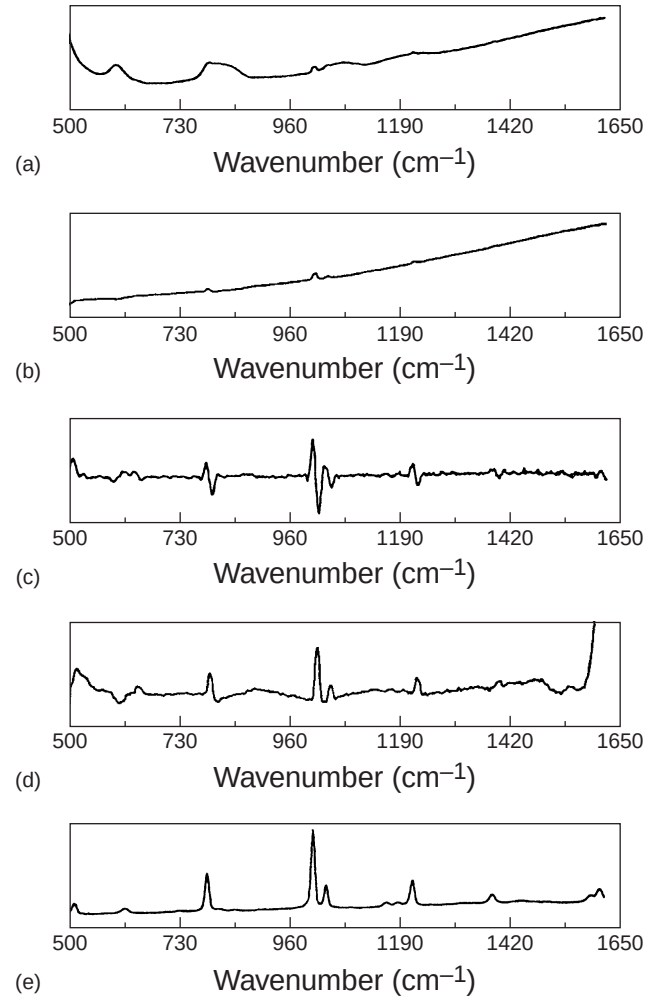

Figure 17 Raman spectra collected with a FORS probe. Parts (a–e) are described in the text. (Courtesy of Naval Command, Control, and Ocean Surveillance Center, Research, Development, Test and Evaluation Division.)

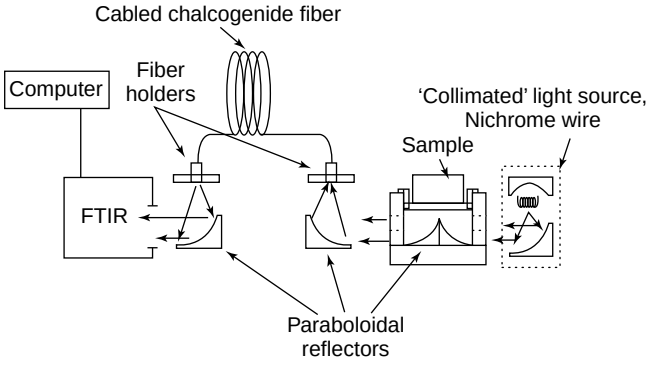

Figure 18 Schematic of a FOIR probe.[101]

(Figure 18). The broad-band optic source is a Nichrome wire representing a blackbody source at a temperature of 1000 K. Arsenic/sulfur-based chalcogenide optical fibers were found to be more rugged and chemically suitable than fluoride-based fibers. The IR diffuse reflectance was measured with a mercury cadmium telluride detector having a wavelength cutoff of 19 μm. The spectral content was obtained with a KVB/Analet FTIR spectrometer with a resolution setting of 4 cm^{-1}. A lithium niobate wafer was placed in front of the detector to filter radiation with wavenumbers below 2225 cm^{-1}. The operating wavenumber range for the system is 6000–2225 cm^{-1}. The method was tested in the laboratory by spiking sand with levels of TCE from 250 to 1000 ppm.[101] The C–H absorption bands of the chlorinated hydrocarbons occur in the wavenumber region from 3100 to 2850 cm^{-1}. An H–S absorption band at 3200 cm^{-1} from the chalcogenide fiber was also observed and used as an internal reference to reduce the matrix dependence of the measurement. Laboratory experiments showed that the minimum observable concentration of TCE in sand was approximately 250 ppm. The technique has also been used to measure a mixture of benzene and TCE on sand.[100] Diesel fuel marine (DFM) has also been measured with this technique. The band depth was linearly related to the DFM concentration at levels up to 1 wt%; however, calibrations at higher concentrations required the use of band depth and spectral shape analysis. Increasing the sensitivity of the technique has been proposed by using a relatively narrow-band detector such as indium antimonide and by increasing the optical power delivered to the detector.[101] Better transmission of power may be achieved by (1) using a more powerful optical source, (2) using lower-loss fibers, and (3) improving the collection efficiency at the fiber facet; however, increasing the power may promote vaporization of the analyte and water. The water absorption bands may interfere because of their proximity to the C–H stretch of the hydrocarbons.

5.1.2.2 Indicator Coatings FOCSs™ have also been developed in which the analyte binds reversibly to an indicator coating on the fiber. The coating is designed to have a chemistry that is specific to the analyte. The chemistry results in a change of the refractive index of the sample that can be related to the analyte concentration. Optical fibers have also been coated with fluorescent dyes. These techniques have not been interfaced with the cone penetrometer but they have been deployed in monitoring wells. Typical sensitivities for substances in the gas phase are in the range of parts per million by volume (ppmv), with sensitivities as low as 100 ppm for substances in the aqueous phase. FOCSs™ have been developed to detect hydrocarbons in leaking underground fuel storage tanks,[102] aromatic hydrocarbons,[103,104] and volatile organohalogen compounds.[105]

5.1.2.3 Chemical Reactions A FOCS™ system has been interfaced with the cone penetrometer for the in situ detection of TCE in soil gas.[106,107] The detection system is based on the Fujiwara reaction, in which an organohalogen reacting with pyridine in basic solution produces a characteristic color change. Sensitivity to various organohalogens is determined by the amount of tetrabutylammonium hydroxide in the reagent and generally increases as the number of chlorine atoms in the molecule increases. The FOCS™ system consists of three components: the reagent, the electro-optical system, and the sensor cell (Figure 19).[108] The sensor cell is a flexible tube of semipermeable material that is connected at both ends by stainless-steel capillaries. One capillary contains the input optical fiber originating from the source lamp and the reagent fill tube, and the other capillary contains the output optical fiber to the detector and reagent drain tube. Organohalogens are transferred through the membrane and react to produce a color change. The intensity of the color change is dependent on the concentration of the analyte and is monitored with optical fibers. The pumping system consists of a computer-controlled syringe pump and six-way valve for transferring the chemical reagent. The sensing reaction is irreversible; however, the solvent-delivery system has the capacity to make 60 measurements. The optical system consists of a portable fluorimeter that includes a source lamp, fiber optics to carry the probe and modulated beam to and from the sensor cell, a filter, and a silicon photodiode that uses a phase-sensitive, differential-absorption detection technique. The sensitivity of the device for TCE is about 200 ppbv, with an upper limit of >500 ppmv. The system has also been commercialized (PurSense™ THM-1; Purus, Inc., San Jose, CA, USA) for the detection of both TCE and trihalomethanes in water and air.[109] The commercialized sensor is linear up to 1 mg L^{-1}. Detection limits for analytes in the gas phase are about 100 ppbv. The operating cycle of the trace analysis is 60 s for reagent flushing and 90 s for measurement. The chemistry is selective, and interferences from polar aqueous-phase analytes containing both halogens and oxygen atoms are low because of the high aqueous solubilities of these substances. The system is more selective when used in the headspace mode than it is when used in the determination of halogens in the aqueous phase directly. Direct detection in the aqueous phase is complicated by fouling of the membrane with inorganic salts, nonvolatile organic compounds, and microorganisms.

The WES has developed a cone penetrometer probe that uses chemically treated tapes that are exposed to the environment outside of the probe to detect the presence of explosive contaminants. When the tape comes into contact with 2,4,6-trinitrotoluene (TNT), a reaction causes the tape to change color, which is detected via a fiber by a spectrophotometer mounted on the cone penetrometer truck.[110] The tape is continuously fed during the penetrometer push from a slot near the tip

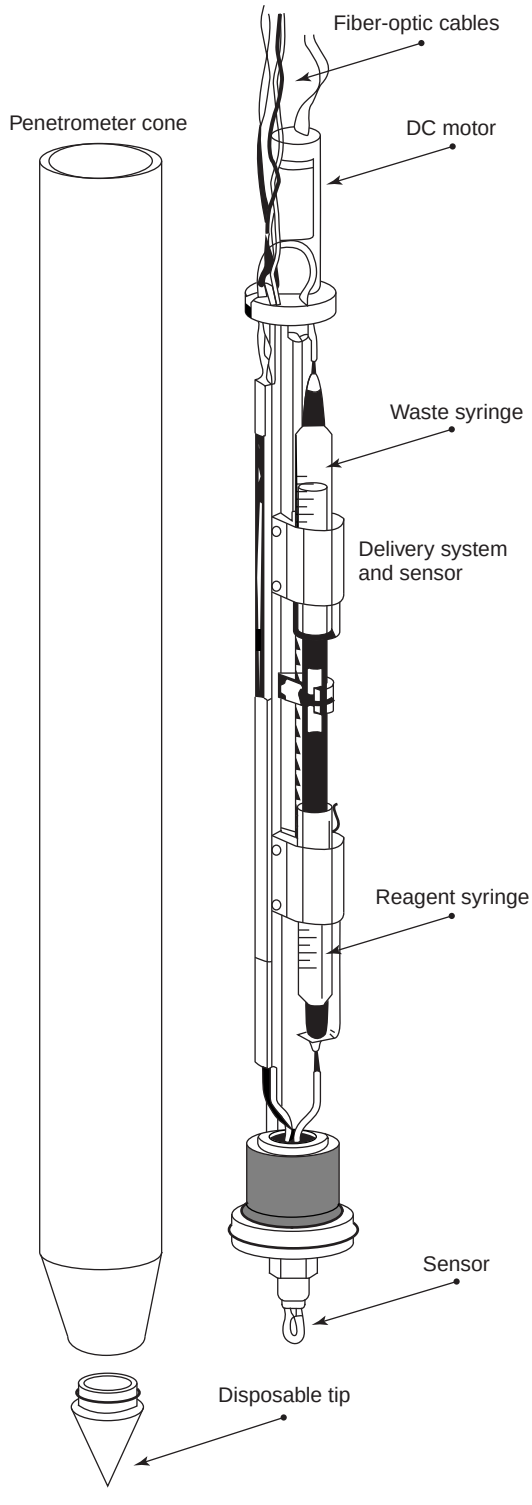

Figure 19 Diagram of a FOCS™ for the detection of TCE in soil gas.[108]

and passes over a sapphire window where color changes are transmitted by the fiber-optic link. Some of the shortcomings of the colorimetric method include rather poor sensitivity and a tendency for the tape to break during the push.

5.1.3 Electrochemical Sensors

Electrochemical sensors have many properties that make them attractive for in situ applications in the cone penetrometer, including (1) small size, (2) simple circuit requirements, (3) no chemical waste, (4) low power requirements, and (5) powerful analytical properties, including high sensitivity, modest selectivity, and a fully reversible response. A thorough discussion on the design and mechanism of operation of electrochemical gas sensors operating in the amperometric mode has been presented.[111] For proper amperometric operation of these sensors, the analyte must be electrochemically active and must come into actual physical contact with the active electrode of the sensor. For the electrochemical sensing of gases, the added requirement is that the analyte be in the vapor phase. Amperometric instrumentation and methods have been developed for the selective detection of several important environmental contaminants.[112] The WES and Transducer Research Inc. (TRI) have developed a number of SCAPS probes equipped with electrochemical sensors for detecting explosives, chlorinated solvents, and VOCs.

The traditional methods to measure TNT and other explosives in environmental samples include sample collection, preparation, and sophisticated laboratory analytical procedures.[113] Field-assay kits are available; however, the sample must be collected and prepared for analysis with chemical reagents that require disposal.[114] Recently, an analytical method for the measurement and quantitation of soil-bound TNT and other explosive contamination has been developed that uses electrochemical sensors operating in the amperometric mode.[115] Gas sensors cannot measure TNT directly because of its low vapor pressure. Therefore, it was necessary to develop a method that increases the vapor pressure of TNT or appropriate signature compounds. The thermal catalysis of nearly all organic compounds over heated noble metal surfaces in air generates electrochemically active vapors. The distribution of thermal products depends primarily on the temperature and nature of the catalytic surface, which can be precisely controlled, and the compound undergoing pyrolysis. The thermal decomposition of TNT, hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX), and other explosives over heated noble metal surfaces generates characteristic products that can be efficiently detected by electrochemical gas sensors. The signature vapors generated by thermal catalysis of a number of different explosives have similar characteristics. TNT, RDX, and most other explosives contain nitro (NO_2) groups, and the catalytic degradation of explosives not only generates carbon dioxide, carbon monoxide, and water, but also nitrogen- and oxygen-containing compounds according to Equation (1):

$$C_7H_5(NO_2)_3 + O_2 \longrightarrow CO_2 + H_2O + CO + NO_y(NO, NO_2, HONO, \text{etc.}) \quad (1)$$

Several of these thermal decomposition products are readily detectable with electrochemical sensors. A SCAPS explosives probe that incorporates electrochemical sensors for the in situ measurement of explosive contamination and geophysical sensors for determining soil layering has been jointly developed by the WES and TRI.[116] The probe collects information on soil classification during the penetrometer push and collects contaminant information during penetrometer retraction. Figure 20 illustrates the main components of the

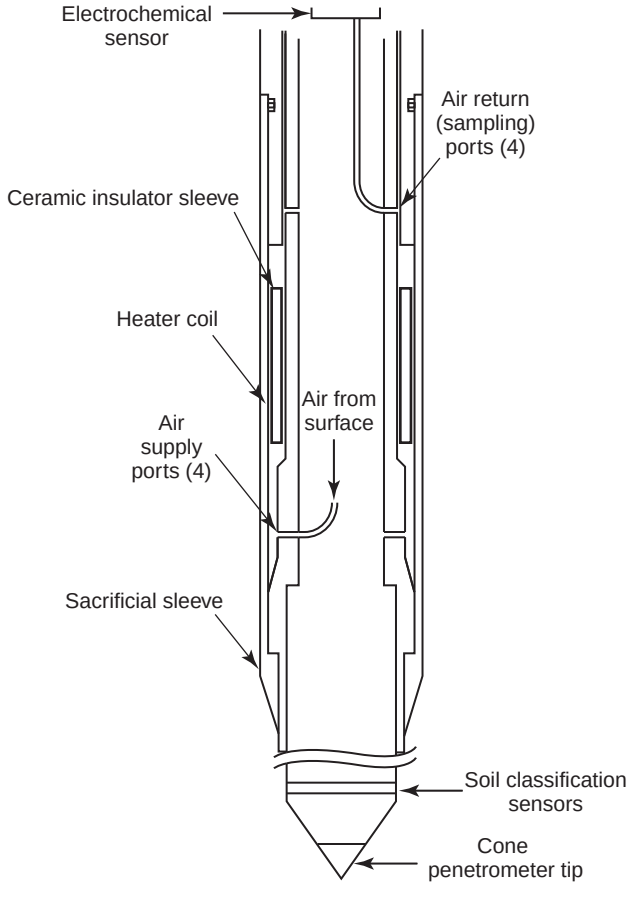

Figure 20 Schematic of an electrochemical sensor probe for detecting explosives in soils. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

SCAPS explosives probe. The probe incorporates an external pyrolyzer system that transforms explosive compounds into electroactive vapors (NO, NO_2, and CO). A pneumatic system transports the pyrolyzed vapors from the soil to the electrochemical sensors inside the probe that generate electrical signals proportional to the concentration of the electroactive vapors. The probe's umbilical (1) allows the chemical sensor signal to be monitored continuously at the surface, (2) ensures positive flow of clean air through the vapor sampler, (3) supplies power to the pyrolyzer during analysis, (4) interfaces the geophysical sensors to onboard SCAPS computers providing real-time soil classification data, and (5) supplies grouting fluids to the tip of the probe. A sacrificial sleeve is used to protect the vapor-sampling ports and the pyrolyzer unit during the penetrometer push. The sleeve is removed during retrieval by soil friction, thereby exposing both the pyrolyzer and the air/vapor sampler system. Laboratory tests have demonstrated sensitivities of better than 1 ppm by weight for TNT in sand.[117]

A probe similar to that used for explosives detection has been developed by the WES and TRI for the detection of VOCs in the vadose zone [RCl probe[116]]. The probe is equipped with electrochemical sensors that detect general classes of VOCs and some halogen-containing compounds (with no known interferents). The probe does not require an external heater because the analytes of interest are present in the vapor phase.

A novel solid-state electrochemical diode detector has been developed for the selective detection of halogenated VOCs and a metal-oxide semiconductor (MOS) gas sensor for the detection of aliphatic and aromatic hydrocarbons and common oxygenated hydrocarbons such as alcohols and ketones in the vadose zone or in the headspace of groundwater.[118] The solid-state electrochemical diode detector consists of a tiny bead of a high-temperature solid-state electrolyte deposited on a noble metal filament. A current is passed through the bead to heat it to several hundred degrees Celsius. Halogenated organics that contact the bead surface decompose, releasing halogen ions that are transported through the bead and cause a measurable increase in conductivity. The response is nonlinear and must be calibrated at several concentrations. The MOS sensor consists of a ceramic tube coated with tin oxide that is heated with a filament to several hundred degrees Celsius. The detector exhibits some selectivity to aliphatic and aromatic hydrocarbons and common oxygenated solvents such as alcohols and ketones. The MOS sensor does not respond to fully chlorinated compounds such as CCl_4 and responds weakly to many other chlorinated hydrocarbons. The sensitivity of the solid-state electrochemical diode detector for CCl_4 is about 2 ppbv and that of the MOS sensor for benzene is about 200 ppbv.

5.1.4 Radiochemical Sensors

The WES, in collaboration with the USDOE, developed a SCAPS spectral γ-probe for passive detection of radioactive wastes.[119] The probe, shown in Figure 21, consists of a detector crystal that is optically coupled to a PMT, power supply, microprocessor, and high-voltage supply. When a γ-ray from radioactive sources in the soil strikes the crystal, it emits a photon, which is detected by the PMT and converted into an electrical voltage. The amplitude of this voltage is proportional to the energy of the photon, which, in turn, is proportional to the energy of the incident γ-radiation. This analog signal passes through a high-speed analog-to-digital converter (ADC) and is digitized. From the ADC, the resulting value is passed on to a microprocessor, which sorts the value within a histogram of counts versus energy. Characteristic peaks in the γ-spectrum indicate the presence of γ-emitting radionuclides. The probe samples 512 channels of γ-radiation in the energy range 100 keV–3 MeV at a rate of one complete spectrum every 2 s. The detector is controlled by an onboard microprocessor, which transmits the digital values along a four-conductor logging cable to a microcomputer for storage and display. The detector offers a choice of three different 1×3 in crystal types: NaI, CsI, and bismuth-germanium-oxide (BGO); however, NaI currently appears to be the only one of the three with sufficient resolution for high-performance spectral analysis. In addition to the γ-sensor, the SCAPS probe is also equipped with geophysical sensors to provide real-time soil classification.

This probe was initially field-tested at the USDOE's Fernald Site near Cincinnati, Ohio, in September 1994. First, the probe was lowered into test drums that had been prepared with known amounts of U-238. The drums were tested for 30 min each. Real-time software developed by Pacific Northwest National Laboratory (PNNL) identified suspected contaminants and their concentrations. Next the SCAPS truck was moved to a contaminated field location, where several pushes were made at predetermined locations. Operators could view soil classification and gross γ-counts versus depth during the push. The probe was also parked at selected depths for a more detailed spectral investigation.

After the first-generation spectral γ-probe had been field-tested, certain improvements and modifications to the probe were identified, and a second-generation spectral γ-probe was fabricated, which has higher resolution and is less expensive than the first-generation probe. The resolution of the NaI detectors (about 14%) was not sufficient to observe some spectral peaks of interest. The required probe sensitivity is such that it would be capable of locating and identifying U-238 in concentrations as low as 35 pCi g^{-1} in soil. The first-generation probe

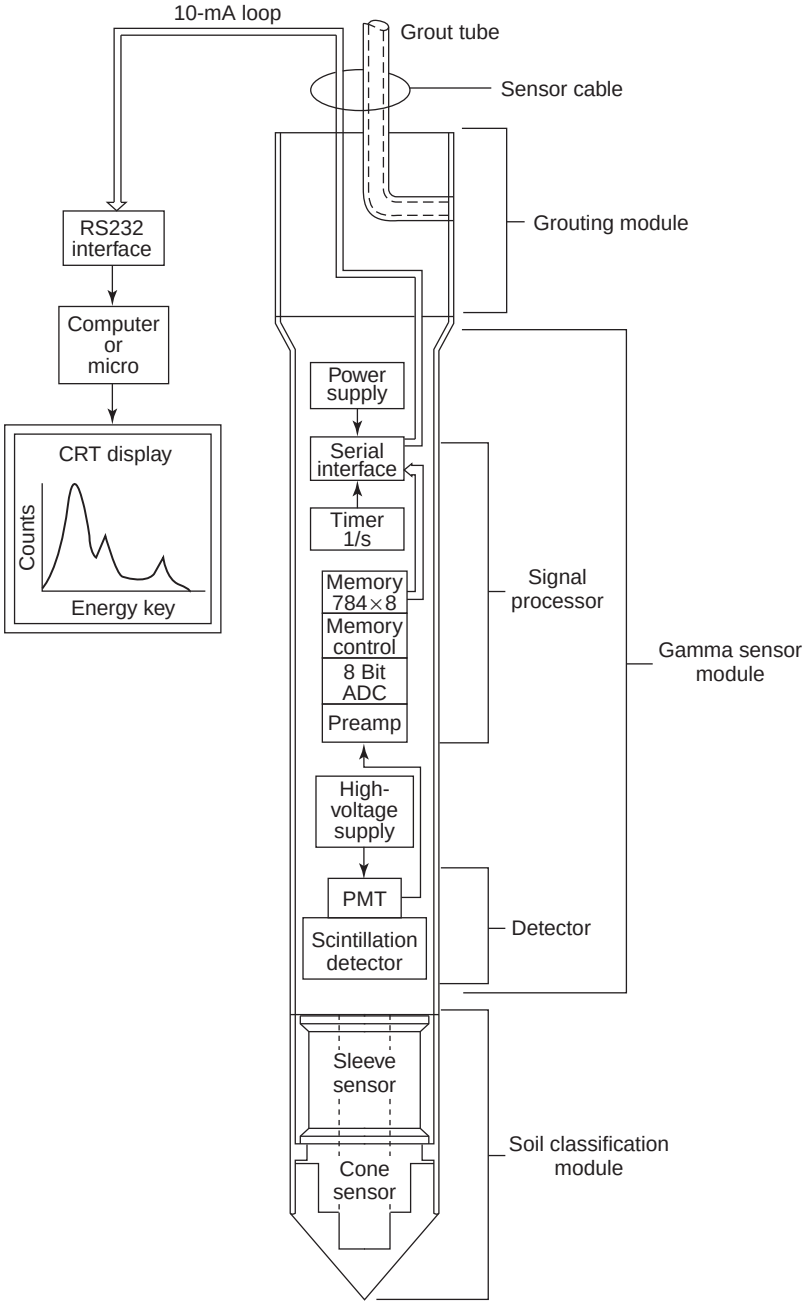

Figure 21 Schematic of radiochemical sensor probe. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

had a lower limit of detection for U-238 of between 60 and 80 $pCi\,g^{-1}$. The improved probe incorporates NaI detectors with a resolution of better than 9%, a higher resolution PMT, and better optical coupling between the crystal and PMT. In addition, the first-generation probe had a considerable redundancy of costly electronics, because each probe contained an ADC, high-voltage supply, power supply, and microprocessor section. To reduce the cost and redundancy of the improved probe, all of the electronics, with the exception of the detector and preamplifier, were relocated into the SCAPS truck. Thus, each probe can share a common high-voltage supply, power supply, ADC, and microprocessor section. This design allows the operators to adjust gain levels from an amplifier located in the SCAPS truck, rather than inside the probe, and reduces the cost of the technique because the downhole equipment could be lost in the event of a rod failure.

5.2 Gas Chromatography

A GC system consists of an injection system, an analytical column, and a detector. The smallest field-portable instruments, with the exception of the Cone-GC™,[118] have not been miniaturized for incorporation into the cone penetrometer; however, field-portable and transportable gas chromatographs have been used to analyze samples collected from the subsurface[43] and have been interfaced with cone penetrometer isolation devices.[46,48] Injection systems typically include a module for analyte preconcentration and water management. Capillary columns allow the complete resolution of complex mixtures in about 30 min. Detectors typically include electrolytic conductivity, electron-capture, flame ionization, photoionization, and thermal conductivity types. Sensitivities of these detectors range from 10 ppt to 1 ppm. Comparison of the retention times of analytes and standard materials under identical chromatographic conditions provides a tentative identification of the contaminant; however, a more positive identification can be made by analyzing the sample on columns of widely different polarities. Detector specificity also permits some tentative identifications to be made. Consequently, GC systems that have multicolumn and multidetector capabilities are very useful in subsurface contaminant applications.

ARA has proposed interfacing the CISP™ and DTD™ probes with a high-speed GC system[120] to measure subsurface fuel contamination.[45,49] High-speed GC techniques are being developed for the rapid screening of gasoline-to-diesel-range organic compounds in soils and soil gas.[121] In a high-speed screening mode, BTEX can be detected in 10–20 s, with sensitivities in the parts per billion range, by using a photoionization detector. In a high-resolution advanced field analysis mode, fingerprint chromatograms are obtained to determine fuel type in about 200 s with a flame ionization detector. The advantage of using high-speed GC over conventional chromatography is that analysis times can be reduced from 30 min to about 100 s. For example, by combining high-speed column switching and pressure-adjustable selectivity, a complete separation of 20–30 VOCs has been accomplished in <1 min. The technique relies on the injection of a very small gas plug from a cryointegration inlet. Sample injections are typically limited to about 2 mL of water-saturated air. Further work on injection systems needs to be done to apply the method to subsurface analysis.

A GC system has been developed that can fit within a cone penetrometer (Cone-GC™) to monitor solvents selectively in the vadose zone or in the headspace of groundwater (Figure 22).[118] The system has a 2.5 cm OD by 91 cm long GC probe assembly that can fit inside a conventional cone penetrometer, a 30 m long connecting cable assembly, and a power/data acquisition assembly that allows user interaction with the GC probe. An automated valving system is used to collect the sample in either a 0.33-mL loop or a sorbent tube and complete the injection into the gas chromatograph. The analytical column is a 0.159 cm OD packed column that is 91 cm long. A novel solid-state electrochemical diode detector is used for the selective detection of halogenated solvents, and a MOS gas sensor is used for the detection of aliphatic and aromatic hydrocarbons and common oxygenated hydrocarbons such as alcohols and ketones. Typical analysis times are <10 min; however, the fast analysis times of the Cone-GC™ are limited by the low chromatographic resolution and may not be suitable to resolve complex mixtures of contaminants.

5.3 Mass Spectrometry

MS has been used extensively to monitor subsurface contamination because of its ability to quantify organic compounds accurately and simultaneously to provide positive identification of the analytes.[66,122] Field-portable and transportable systems can be installed in

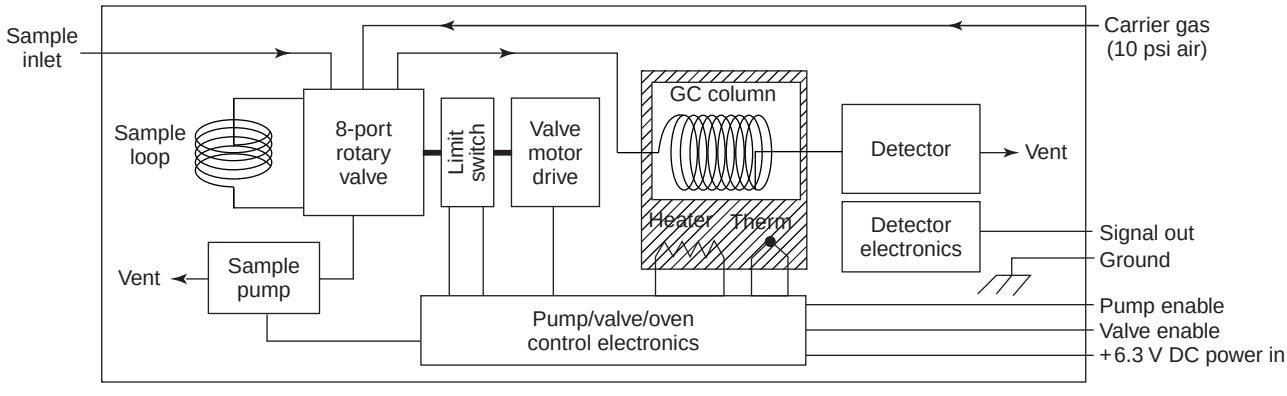

Figure 22 Diagram of the Cone-GC™ probe system,[118] developed by Microsensor Systems, Inc. (Bowling Green, KY, USA).

the cone penetrometer truck for ease of analysis in the field. Mass spectrometer systems have been directly interfaced to sampling devices in the cone penetrometer and also used in the field for the analysis of samples collected by conventional sampling devices in the cone penetrometer. The mass analyzers that are most commonly used in field instruments are the linear quadrupole and the quadrupole ion trap.[122] The linear quadrupole has moderate to low resolution capability. The ion trap is 10–100 times more sensitive than the linear quadrupole; however, the instrument may suffer from air and water ion–molecule reactions that complicate the mass spectra.

Direct-sampling ion-trap mass spectrometry (DSITMS) has been developed to provide faster and less costly methods for the analysis of organic contaminants in environmental samples.[123,124] The analytes can be introduced directly into the mass spectrometer with no sample preconcentration and little or no sample preparation. The analytes are monitored in real time or near real time at low parts per billion levels. Analysis times range from instantaneous to approximately 3 min. Several sample introduction systems have been developed for the DSITMS system that include a soil/water purge module, a TD module for sorbent tubes, and a direct air-sampling module for real-time monitoring of VOCs. Detection limits for VOCs in water and soils are about 1–5 and 10–30 ppb, respectively. Air samples can be quantified down to 1 ppb. Sample preconcentration can extend the detection limits down to 50–100 pg. The system has been used to analyze sorbent tubes from the TDS™ and has the potential to be interfaced with some of the cone penetrometer isolation devices to provide an on-line analytical capability. For example, DSITMS has been interfaced with the Hydrosparge™ for the rapid analysis of groundwater by using the cone penetrometer.

A field-transportable TD mass spectrometer has been used to characterize PCBs, PAHs, and VOCs in subsurface samples.[125,126] Analysis of VOCs in groundwater, PCBs in soil, and PAHs in soil required about 10, 10–15, and 20–30 min per sample, respectively, using a short GC column sampling probe as a fast GC system. Interfacing the TD mass spectrometer with the TECP™ has significantly reduced the analysis times. Analysis times are approximately 7 min per sample for SVOCs and VOCs using ion fingerprint detection algorithms. Detection limits are in the parts per million range for soil.

6 TECHNOLOGY DEMONSTRATION AND VALIDATION

As with any new technology, before the novel techniques described in this article can gain widespread acceptance and begin to replace conventional methods, the following questions must be addressed:

1. Are there significant and quantifiable improvements in the speed, cost, sensitivity, and accuracy of the technique over conventional methods?
2. What is the quality of the data in widely varying soil and groundwater types?
3. How rugged, operable, and user friendly are the technologies under actual field conditions?
4. Will the innovative methods be accepted by the appropriate regulatory agencies?

A large number of technology demonstrations have been and are currently being conducted to transfer the innovative cone penetrometer technologies from the laboratory development and testing phases to their implementation in the field. For example, the USDOE began the Rapid Commercialization Initiative (RCI) to assist inventors of environmental technologies, including cone penetrometer devices, in bringing their devices to the commercialization phase.[127] The RCI assists the private sector and state and federal agencies in (1) finding appropriate sites for the demonstration and testing of mature environmental technologies, (2) verifying the performance and cost of the technologies, and (3) facilitating and expediting the issuing of permits. The following discussion contains a brief synopsis of several field demonstrations and validations of cone penetrometer chemical technologies. The data from these demonstrations were made available to the authors but do not include the entirety of the data available from field demonstrations of cone penetrometer devices. Other devices that were described in sections 3, 4, and 5 are also in the laboratory and field validation phases, and the reader is encouraged to contact the investigators directly for additional information on their validation, regulatory acceptance, and commercialization.

6.1 Samplers

The cone penetrometer samplers are designed to perform an in situ collection of groundwater, soil, and soil gas. Following sample collection, the probe is usually brought to the surface, and the sample is transferred to a container and analyzed on-site or transported off-site to a laboratory for analysis. Comparisons between groundwater samples collected by the BAT™ and HydroPunch™ devices with samples collected from nearby monitoring wells by using bailers have usually been favorable.[3,32,36,128–130] The groundwater that is isolated in monitoring wells represents a wider subsurface strata than the narrow depth sampled by the in situ devices; therefore, some discrepancies are expected. The sealed container of the BAT™ sampler and the pressurized chamber of the

Hydrocone™ are able to minimize the loss of VOCs more effectively than the HydroPunch™; however, losses of aromatic and chlorinated aromatic hydrocarbons to the HDPE filter of the BAT™ sampler have been observed.[32,36,130] A major shortcoming of the WBW, BAT™, HydroPunch™, and Hydrocone™ groundwater probes is that they cannot accommodate the electronic cone. Also, by incorporating a retractable sleeve to protect the porous element during the penetration, they are restricted to a single sample collection during each sounding. Data from field demonstrations of the Cone-Sipper™ are available.[37] The advantages of the Cone-Sipper™ over the other groundwater samplers include its multiple-depth-sampling capability, ability to accommodate the electronic cone, and its adaptability to an on-line analytical technique.

Soil coring devices operate effectively; however, careful attention must be paid to subsampling the soil cores and transferring the samples to containers before analysis to preserve the integrity of the sample, particularly for VOCs.[41] Significant losses of VOCs have been observed when subsamples were transferred by spoon to glass vials.[131] Soil residues retained on the threads of the vials prevented the caps from sealing properly and permitted VOCs to volatilize from the sample container. Quantitative results for VOCs were obtained by using open-barrel plastic syringes to subsample the soil core and to transfer the sample to a glass vial.

The comparability of the soil gas concentrations determined by soil gas probes to the actual soil gas concentrations are unknown. VOCs are suspected of being volatilized from soil by the vacuum that is applied to transfer the soil gas sample through a transfer line to a container or gas analyzer at the surface.[55,107,132] Consequently, the soil gas concentrations may be dependent on the ability of the soil to supply gas at a rate comparable to the rate at which it is withdrawn to the surface. Direct comparisons with soil gas concentrations determined with, for example, the passive dual planar membrane organics monitor (PDPMOM) or its predecessor, the TerraTrog™ system[133–135] could possibly be used to validate the soil gas probes. They are soil gas implant systems that contain permeable Teflon® or polyimide membranes through which the soil gas enters a sample chamber. An inert gas is used to purge the sample chamber and transfer the soil gas to the surface. Soil gas enters the chamber because of the difference in partial pressure of the analyte in the soil gas and the inert purge gas. This collection method is more passive than the soil gas probes even when it is used in its dynamic purging mode because it does not reduce the pressure at the particle surface. Soil gas probes can be used as effective screening tools for site characterizations when used in concert with analyses of groundwater and soil at the site.

6.2 On-line Analytical Techniques

Cone penetrometer devices that incorporate in situ isolation techniques to purge VOCs from groundwater were developed to eliminate the transfer of groundwater from the in situ sampling probes to a container prior to chemical analysis and to minimize losses of analytes (section 4.1). The groundwater probes include the ISPD™, the CISP™ probe, and the Hydrosparge™. The ISPD™ and the CISP™ probe have been tested in the laboratory and have undergone preliminary field tests; however, further testing of these devices is necessary. Field tests of the ISPD™ and the CISP™ probe have revealed that the design of the sample inlet is critical to the successful operation of these purge devices. The purging of water to quantitatively recover VOCs is an accepted methodology; however, the use of this technique in a groundwater probe is complicated by the necessity to filter the water as it enters the probe. The filter of the ISPD™ has a tendency to clog. The CISP™ probe incorporates a two-stage filter that has worked more effectively. The Hydrosparge™ is deployed within the HydroPunch™ and PowerPunch™ samplers. The water inlets of these samplers are filters with large surface areas that facilitate the collection of groundwater; however, the HydroPunch™ and PowerPunch™ samplers cannot incorporate the electronic cone and do not have the multiple-depth-sampling capabilities of the ISPD™ and CISP™ probe.

The Hydrosparge™ has been undergoing extensive technology demonstrations since 1996 to obtain regulatory acceptance. The initial field demonstrations were funded by the Environmental Security Technology Certification Program (ESTCP) and were conducted at Aberdeen Proving Ground (APG), Maryland, during June 1996 and near the McClellan AFB Annex, Davis, California, during November and December 1996.[136] During these demonstrations, the Hydrosparge™ results were validated against groundwater samples that were bailed from the HydroPunch™ wells after the Hydrosparge™ data had been collected. Laboratory analysis of the bailed samples was conducted by USEPA Method 8260A, while samples from the Hydrosparge™ were analyzed by an on-line analytic technique in the field using a Teledyne™ DSITMS. Figure 23 shows the Hydrosparge™ measurements of tetrachloroethene (PCE), TCE, and dichloroethene (DCE) plotted against the results from the laboratory analyses of the samples that were bailed from the HydroPunch™. The plot indicates a high degree of agreement between the in situ purging technique and the conventional method of bailing a sample from a groundwater sampler, transferring the sample to a container, and analyzing the sample in the laboratory. Table 1 summarizes some of the results from the field demonstration.

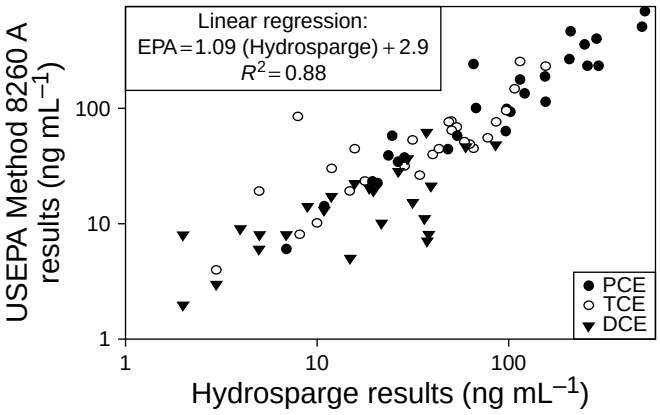

Figure 23 Hydrosparge™ validation data, ESTCP demonstration, Davis, CA, 1996. (Courtesy of the US Army Corps of Engineers Waterways Experiment Station.)

Table 1 Concentrations of VOCs in groundwater[a] collected from conventional monitoring wells and sampled with a cone penetrometer using a HydroPunch™

Well	Hydrosparge™ (ng mL^{-1})			USEPA Method 8260A (ng mL^{-1})		
	DCE	TCE	PCE	DCE	TCE	PCE
03-HP1[b]	16	51	257	22	75	340
03-HP2	27	50	535	28	74	640
03-HP3	<5	<5	<5	<5	<5	<5
CMW03[c]	5	5	117	6	19	170
05-HP1	<5	8	29	<5	8	37
05-HP2	<5	<5	7	<5	<5	6
05-HP3	32	35	158	15	26	110
CMW05	30	16	216	36	44	440
06-HP1	40	52	157	21	63	180
06-HP2	61	109	213	45	140	250
06-HP3	<5	<5	<5	<5	<5	<5
06-HP4	86	159	519	48	220	480
CMW06	7	12	25	8	30	57

[a] The groundwater was bailed and analyzed by USEPA Method 8260A and then purged in situ with the Hydrosparge™ and analyzed with a DSITMS.
[b] Groundwater isolated near conventional monitoring well using a cone penetrometer with HydroPunch™ sampler.
[c] Groundwater isolated in a conventional monitoring well.

The Hydrosparge™ demonstrations at APG also yielded valuable cost comparisons between the two techniques. Using conventional site characterization cost and time estimates provided by the APG site manager and actual SCAPS cost data, it is estimated that a SCAPS investigation using the Hydrosparge™ probe saved over $300 000 and was completed in approximately 15% of the time that conventional methods would have required.

Cone penetrometer devices that incorporate in situ isolation techniques to thermally desorb VOCs from soil were developed to perform an in situ analysis of the VOCs and SVOCs in soil and eliminate the collection of soil cores and the analysis of subsamples in the laboratory. The devices include the TDS™, the DTD™ probe, and the TECP™. Of the three probes, the TDS™ has undergone the most extensive field testing. Recoveries of analytes by each of these probes are directly correlated with the moisture content of the soil and are related to their ability to dry the soil during the TD step. Recoveries generally decrease with increasing soil moisture. A distinct advantage of the TDS™ over the other devices is its ability to desorb thermally a known volume of soil; however, the mass of the soil sample is still unknown.

Results obtained with the TDS™ were validated by collecting soil samples adjacent to the TDS™ sampling locations with a cone-penetrometer-deployed split-spoon sampler. A subsample of soil from the split-spoon sampler was extracted, immediately placed in methanol in accordance with the method of Hewitt,[137] and analyzed by an off-site laboratory using USEPA Method 8260A. Another subsample was placed directly in the probe, thermally desorbed, and analyzed onboard the SCAPS truck with a DSITMS. The TDS™ measured 21 discrete samples during eight pushes (Table 2). The maximum

Table 2 Results from field investigation near McClellan AFB, California, to validate TDS™

Sample	Depth (m)	TDS™[a] (ng g^{-1})	USEPA Method 8260A[b] (ng g^{-1})	TD[c] (ng g^{-1})
TD01-S1-05	1.5	0	7.00	nd[d]
TD01-S2-15	4.6	0	19.0	nd
TD01-S3-20	6.1	0.3	4.50	nd
TD01-S4-25	7.6	34.4	15.0	6.40
TD01-S5-36	11	28.2	15.0	nd
TD02-S1-26	7.9	17.4	0	24.4
TD03-S1-26	7.9	12.0	60.0	34.0
TD04-S1-30	9.1	28.8	42.0	50.8
TD04-S1-30D	9.1	–	26.0	nd
TD05-S1-26	7.9	15.3	35.0	47.6
TD05-S2-28	8.5	35.0	39.0	37.2
TD05-S3-30	9.1	27.6	47.0	48.0
TD05-S4-32	9.8	27.0	66.0	78.0
TD06-S1-30	9.1	22.6	52.0	23.6
TD06-S2-32	9.8	13.5	92.0	62.0
TD06-S3-34	10	32.6	23.0	8.80
TD07-S1-26	7.9	13.5	25.0	36.4
TD07-S2-28	8.5	13.8	28.0	15.2
TD07-S3-30	9.1	33.2	71.0	16.8
TD08-S1-26	7.9	15.0	28.0	39.2
TD08-S2-28	8.5	12.6	72.0	41.6
TD08-S3-30	9.1	25.0	47.0	29.6

[a] In situ TD of soil using the cone penetrometer with TDS™.
[b] Soil sample collected using the cone penetrometer with split-spoon sampler and analyzed by USEPA Method 8260A.
[c] Soil sample collected using the cone penetrometer with split-spoon sampler and thermally desorbed with the TDS™.
[d] nd = Not determined.

sampling depth and distance to the water table were 11 and 10 m, respectively, below the ground surface. PCE was the primary volatile contaminant detected by the TDS™. PCE was found at levels below the detection limit in a few samples; however, only one sample analyzed by the off-site laboratory indicated the presence of TCE in the soil. Concentrations of PCE were generally low, ranging from undetectable to 100 ng g^{-1} in both the TDS™ samples and the verification samples collected with the split-spoon sampler (Table 2). Concentrations of VOCs in many of the soil verification samples were less than the lower reporting limit for USEPA Method 8260A and are, therefore, reported as estimated (*J*) values. In general, results from subsamples of soil from the split-spoon sampler that were thermally desorbed and analyzed with the DSITMS are within a factor of two of the subsamples analyzed by USEPA Method 8260A.

6.3 Chemical Detection Techniques

Chemical sensors have been interfaced with the cone penetrometer to perform an in situ analysis of the subsurface media. The analyte isolation and transfer steps are eliminated, and thus these techniques provide a rapid analysis of the subsurface media. The SCAPS LIF sensors represent the first and most mature cone penetrometer chemical detection technology. The LIF sensor has undergone extensive field testing, demonstration, and validation since 1991 and has received regulatory acceptance, has been licensed and commercialized, and is being routinely used in site characterization and remediation operations. By the end of 1996, the US Government-owned SCAPS units equipped with LIF sensors had already completed over 200 site investigations. The SCAPS LIF demonstration and validation project and its approach to obtaining regulatory acceptance have become the model for other cone-penetrometer-based site characterization technologies.

During August 1994, the USEPA conducted an evaluation of the SCAPS LIF sensors (both the nitrogen laser and the tunable-laser-based systems) at three POL-contaminated sites as part of the Superfund Innovative Technology Evaluation Program. The USEPA evaluated the SCAPS operating procedures, costs, and data quality and compared them with the most appropriate conventional method of drilling, sampling, and laboratory analysis. After the conclusion of the three field evaluations, the USEPA published an innovative technology evaluation report,[86] which concluded that the SCAPS technique using the LIF sensor generated screening-level data faster than the conventional methods and with little or no generation of waste. The SCAPS operation also required less manpower than conventional methods, was less expensive ($20 000 to characterize all of the three sites, compared with $43 000 for drilling and sampling), generated almost 1200 more data points than the conventional approach, and provided both contaminant and stratigraphic data in real time. One identified drawback of the SCAPS technology was the need for laboratory analysis of critical samples because the SCAPS LIF sensors provided screening-level data rather than definitive quantitative data. Detection thresholds for the SCAPS LIF technologies are estimated in the innovative technology evaluation report to be 60 mg kg^{-1} for total petroleum hydrocarbons and 19 mg kg^{-1} for volatile petroleum hydrocarbons. The USDOE[138] and the Naval Facilities Engineering Command[139] evaluated the SCAPS LIF technology and estimated costs that were up to 50% lower than site characterizations using conventional methods. The SCAPS LIF sensor program has actively pursued regulatory acceptance by conducting extensive field demonstrations, together with laboratory verifications for numerous Federal, state, and local regulatory agencies. The NRaD in San Diego, California, has been very active in pursuing regulatory acceptance for the SCAPS LIF through a series of well-designed and publicized demonstrations that have involved representatives from the USEPA, the California Environmental Protection Agency, the Western Governors Association, and the Interstate Technology Regulatory Cooperative. Demonstrations of the SCAPS that included extensive verification sampling were conducted at the Hydrocarbon National Test Site located in Port Hueneme, California, and at arid sites at Sandia National Laboratories in New Mexico during 1995. The results of these demonstrations culminated with the SCAPS LIF technology being first certified in 1996 by the California Environmental Protection Agency, followed thereafter by multistate certification, and recognition as a screening-level tool by the USEPA.[86]

Another SCAPS technology that recently transitioned to the ESTCP field demonstration and validation program is the electrochemical sensor probe for detecting explosives in soils. Initial field evaluations at the Louisiana Army Ammunition Plant (LAAP), the Volunteer Army Ammunition Plant (VAAP) in Tennessee, the USDOE's Pantex facility in Amarillo, Texas, and the Longhorn Army Ammunition Plant (LHAAP) in Texas demonstrated the sensor's capability to detect low (parts per million) to moderate (10%) levels of explosives (TNT, RDX, and HMX). Table 3 shows a qualitative comparison of the sensor results with those from soil samples that were analyzed using USEPA Method 8330 to validate the results. This contingency analysis involves evaluation of the frequency of the sensor's true positive and negative response as determined by results from the validation sample. A true-positive sensor response is assigned whenever the validation sample indicates the presence of an

Table 3 Contingency analysis for explosives sensor response and verification soil samples

Site	True-positive	False-positive	True-negative	False-negative
VAAP	8	2	1	2
LAAP	4	0	0	0
Pantex	4	0	3	1
LHAAP	4	1	5	0
Total	20	3	9	3
Total (%)	87	25	75	13

explosive contaminant at a location where the sensor yields a response above background. Similarly, a true-negative response is assigned where both the validation sample and the explosive sensor yield no response above background. A compilation of data from the field tests indicates that the explosives sensor yielded 87 and 75% true-positive and true-negative responses, respectively (Table 3). The false-positive responses observed may be due to the presence of nonexplosive nitrogen species present in the soils or may be attributed to inhomogeneity of TNT in environmental samples.[140] Additional electrochemical sensors have been incorporated into the SCAPS explosives probe to mitigate the effects of interferences from nonexplosives.

Technologies using the SCAPS for detecting toxic metals in the subsurface have also recently been transitioned from laboratory and limited field evaluations to the ESTCP for full demonstration and validation. A number of field demonstrations of the fiber-optic LIBS probe [NRaD[141]], the downhole laser LIBS probe [WES[142]], and the XRF probe [Naval Research Laboratory, WES[143]] using side-by-side evaluations have been performed. Preliminary field evaluations have demonstrated that these technologies can detect low levels of metals in soils including Cr, Pb, Cu, and Fe. Under the ESTCP project, a direct comparison (performance and cost) among the three technologies will be performed, and the best technology will be transitioned to government and commercial SCAPS users.

7 CONCLUSIONS

The interfacing of chemical technologies with the cone penetrometer has made it possible to perform characterizations of many contaminated sites in a less intrusive, more cost-effective manner than performing the same investigation by the conventional method of drilling, sampling, and chemical analysis in the laboratory. The cone penetrometer is evolving from a device used only to gather geotechnical information about the subsurface to a tool that can also collect samples of groundwater, soil, and soil gas for chemical analysis and can provide in situ chemical detection. The most recent technological developments include the incorporation of on-line analytical and in situ chemical detection techniques into the cone penetrometer. It is apparent that a large number of probes have been invented that are capable of measuring inorganic and organic contamination in screening and quantitative modes over a wide range of concentrations; however, it is also apparent that many of these probes have not undergone the rigorous field demonstrations, validations, and implementation in site characterizations that will be necessary to obtain regulatory acceptance of the methodologies. To gain regulatory acceptance, data from the probes must be validated against data from verification samples that are collected and analyzed by accepted methods. The field validation should be performed at a variety of sites containing different soil and groundwater types to evaluate the effects of the sample matrix on the mechanical and chemical performances of the probe. Results from site characterizations performed by conventional methods of drilling, sampling, and laboratory analysis should then be compared at the same site with the results of characterizations performed with the cone penetrometer technologies to quantify improvements in the accuracy, speed, and cost of the site characterization.

ACKNOWLEDGMENTS

We thank our colleagues for providing valuable information on the cone penetrometer devices. The technical reviews by Dr Nancy A. Marley and Ms Candace M. Rose of Argonne National Laboratory and by Dr William M. Davis of the US Army Corps of Engineers Waterways Experiment Station are greatly appreciated. The assistance of Ms Eileen Brazelton of Argonne National Laboratory with editing is also greatly appreciated. The writing of this article was supported by the US Department of Energy under contract W-31-109-Eng-38 as part of the Characterization, Monitoring, and Sensor Technology Integrated Program of the Office of Technology Development, Assistant Secretary for Environmental Management (PVD), and by the Strategic Environmental Research and Development Program (SERDP), Washington, DC (ERC). The authors also wish to acknowledge SERDP, the Environmental Security and Technology Certification Program (ESTCP), the Army Environmental Center, and the Army's Installation Restoration Research Program (IRRP) for funding the SCAPS sensor and sampler development and demonstration projects described in this article.

This article is reprinted by permission of John Wiley & Sons, Ltd. from the *Encyclopedia of Environmental Analysis and Remediation*, edited by Robert. A. Meyers, copyright 1988 by John Wiley & Sons, Ltd.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
AFB	Air Force Base
AOTF	Acousto-optical Tunable Filter
APG	Aberdeen Proving Ground
ARA	Applied Research Associates, Inc.
ASTM	American Society for Testing and Materials
BGO	Bismuth-germanium-oxide
BTEX	Benzene, Toluene, Ethylbenzene, and Xylene
CCD	Charge-coupled Device
CISP	Cone Penetrometer In Situ Purging
CPT	Cone Penetrometer Test
CP/VZV	Cone Penetrometer/Vadose Zone Vapor
DCE	Dichloroethene
DFM	Diesel Fuel Marine
DNAPLS	Dense Nonaqueous-phase Liquids
DSITMS	Direct-sampling Ion-trap Mass Spectrometry
DTD	Dynamic Thermal Desorption
DTM	Downhole Trap Module
EEM	Excitation/Emission Matrix
ESTCP	Environmental Security Technology Certification Program
FEP	Fluorinated Ethylene Polypropylene
FFT	Fast Fourier Transform
FOCS	Fiber-optic Chemical Sensor
FOIR	Fiber-optic Infrared Reflectance
FORS	Fiber-optic Raman Sensor
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GRIN	Graded Index
HDPE	High-density Polyethylene
HMX	Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine
ID	Inner Diameter
IR	Infrared
ISPD	In Situ Purge Device
LAAP	Louisiana Army Ammunition Plant
LHAAP	Longhorn Army Ammunition Plant
LIBS	Laser-induced Breakdown Spectroscopy
LIF	Laser-induced Fluorescence
LNAPLS	Light Nonaqueous-phase Liquids
MIMS	Membrane Introduction Mass Spectrometry
MIT/LL	Massachusetts Institute of Technology/ Lincoln Laboratory
MOS	Metal-oxide Semiconductor
MPS	Multiport Sampler
MS	Mass Spectrometry
NDSU	North Dakota State University
NRaD	United States Navy's Naval Command, Control, and Ocean Surveillance Center; Research, Development, Test and Evaluation Division
OD	Outer Diameter
OMA	Optical Multichannel Analyzer
ORNL	Oak Ridge National Laboratory
PAH	Polycyclic Aromatic Hydrocarbon
PAWS	Portable Acoustic Wave Sensor
PCB	Polychlorinated Biphenyl
PCE	Tetrachloroethene
PDPMOM	Passive Dual Planar Membrane Organics Monitor
PEEK	Polyether Ether Ketone
PFA	Perfluoroalkoxy
PMT	Photomultiplier Tube
PNNL	Pacific Northwest National Laboratory
POL	Petroleum, Oils, and Lubricant
PTFE	Polytetrafluoroethylene
RCI	Rapid Commercialization Initiative
RDX	Hexahydro-1,3,5-trinitro-1,3,5-triazine
ROST	Rapid Optical Screening Tool
SCAPS	Site Characterization and Analysis Penetrometer System
SERS	Surface-enhanced Raman Scattering
SVOC	Semivolatile Organic Compound
TCE	Trichloroethene
TD	Thermal Desorption
TDS	Thermal Desorption Sampler
TECP	Thermal Extractor for the Cone Penetrometer
TNT	2,4,6-Trinitrotoluene
TRI	Transducer Research Inc.
USDOD	United States Department of Defense
USDOE	United States Department of Energy
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UVF	Ultraviolet Fluorescence
VAAP	Volunteer Army Ammunition Plant
VOC	Volatile Organic Compound
WBW	Wellpoint Batch Water
WES	United States Army Corps of Engineers Waterways Experiment Station
WTM	Wavelength–Time Matrix
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Laser-induced Breakdown Spectroscopy, Elemental Analysis • Laser- and Optical-based Techniques for the Detection of Explosives • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment

Environment: Water and Waste cont'd **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements • Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

Atomic Spectroscopy **(Volume 11)**
Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Ablation in Atomic Spectroscopy • Laser-induced Breakdown Spectroscopy

X-ray Spectrometry **(Volume 15)**
Portable Systems for Energy-dispersive X-ray Fluorescence

REFERENCES

1. S.H. Lieberman, G.A. Theriault, S.S. Cooper, P.G. Malone, R.S. Olsen, P.W. Lurk, 'Rapid, Subsurface, in situ Field Screening of Petroleum Hydrocarbon Contamination Using Laser Induced Fluorescence over Optical Fibers', *Proceedings of the 2nd International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 57–63, 1991.
2. R.A. Klopp, 'In-situ Penetration Testing for Delineation of Ground Water Contaminant Plumes', *Proceedings of the Third National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 329–343, 1989.
3. M. Smolley, J.C. Kappmeyer, 'Cone Penetrometer Tests and Hydropunch™ Sampling an Alternative to Monitoring Wells for Plume Definition', *Proceedings of the Conference on New Field Techniques for Quantifying the Physical and Chemical Properties of Heterogeneous Aquifers*, National Water Well Association, Dublin, OH, 717–724, 1989.
4. M. Smolley, J.C. Kappmeyer, 'Cone Penetrometer Tests and HydroPunch® Sampling: A Screening Technique for Plume Definition', *Ground Water Monit. Rev.*, **11**, 101–106 (1991).
5. N.A. Berzins, 'Use of the Cone Penetration Test and BAT™ Ground Water Monitoring System to Assess Deficiencies in Monitoring Well Data', *Ground Water Manage.*, **11**, 327–340 (1992).
6. C.Y. Chiang, K.R. Loos, R.A. Klopp, 'Field Determination of Geological/Chemical Properties of an Aquifer by Cone Penetrometry and Headspace Analysis', *Ground Water*, **30**, 428–436 (1992).
7. P.D. Kuhlmeier, T.E. Sturdivant, 'Delineation of Lithology and Ground Water Quality in a Complex Fluvial Estuarine Depositional Zone', in *Current Practices in Ground Water and Vadose Zone Investigations*, eds. D.M. Nielsen, M.N. Sara, American Society for Testing and Materials, Philadelphia, PA, 183–198, 1992.
8. B. Manchon, 'Introduction to Cone Penetrometer Testing and Ground Water Samplers', Presented at the *Workshop of the Fifth National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 1991.
9. E.R. Cespedes, B.H. Miles, S.H. Lieberman, 'Development of Optical Sensors for the Site Characterization and Analysis Penetrometer System (SCAPS)', in *Optical Sensing for Environmental Monitoring*, SP-89, Air and Waste Management Association, Pittsburgh, PA, 621–632, 1994.
10. Advanced Sciences, Inc., 'Review of Instrumentation and Sensors: Cone Penetrometer Applications', Task 6 Report, Contract No. DE-AC05-87OR21706, U.S. Department of Energy, Martin Marietta Energy Systems, Inc., Oak Ridge, TN, 1994.
11. S.S. Cooper, P.G. Malone, R.S. Olsen, G.B. Mohrman, 'Use of an Instrumented Cone Penetrometer in Monitoring Land Disposal Sites', *Proceedings of the 5th National Conference on Hazardous Wastes and Hazardous Materials*, Hazardous Materials Control Research Institute, Silver Spring, MD, 424–427, 1988.
12. C.Y. Chiang, C.C. Stanley, L. Hekma, G.F. Boehm, 'Characterization of Ground Water and Soil Conditions by Cone Penetrometry', *Proceedings of the Conference on Petroleum Hydrocarbons and Organic Chemicals in Ground Water: Prevention, Detection, and Restoration*, National Water Well Association, Dublin, OH, 141–157, 1989.
13. A.R. Holguin, S.H. Edelman, 'Advances in Cone Penetrometer Testing for Environmental Characterization of Soil and Ground Water', *Proceedings of the 1994*

Federal Environmental Restoration III & Waste Minimization II Conference and Exhibition, Hazardous Materials Control Resources Institute, Rockville, MD, 135–144, Vol. I, 1994.
14. P.K. Robertson, R.G. Campanella, 'Interpretation of Cone Penetration Tests. Part I: Sand', *Can. Geotech. J.*, **20**, 718–733 (1983).
15. P.K. Robertson, R.G. Campanella, 'Interpretation of Cone Penetration Tests. Part II: Clay', *Can. Geotech. J.*, **20**, 734–745 (1983).
16. P.K. Robertson, 'Soil Classification Using the Cone Penetrometer Tests', *Can. Geotech. J.*, **27**, 151–158 (1990).
17. ASTM, 'Standard Test Method for Deep, Quasi-static, Cone and Friction-cone Penetration Tests of Soil', in *Soil and Rock, Building Stones: Geotextiles*, Designation: D 3441-86, American Society for Testing and Materials, Philadelphia, PA, 409–414, Vol. 4.08, 1988.
18. P.K. Robertson, J.P. Sully, D.J. Woeller, T. Lunne, J.J.M. Powell, D.G. Gillespie, 'Estimating Coefficient of Consolidation from Piezocone Tests', *Can. Geotech. J.*, **29**, 539–550 (1992).
19. A.I. Strutynsky, R.E. Sandiford, D. Cavaliere, 'Use of Piezometric Cone Penetration Testing with Electrical Conductivity Measurements (CPTU-EC) for the Detection of Hydrocarbon Contamination in Saturated Granular Soils', in *Current Practices in Ground Water and Vadose Zone Investigations*, eds. D.M. Nielsen, M.N. Sara, American Society for Testing and Materials, Philadelphia, PA, 169–182, 1992.
20. R. Knowlton, 'TDR and Fiber Optic Sensors for Cone Penetrometer Applications', *Proceedings of the 2nd Annual Site Characterization and Analysis Penetrometer System (SCAPS) Sensor Development Workshop*, SERDP-96-2, US Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, 243–258, 1996.
21. R.G. Campanella, I. Weemees, 'Development and use of an Electrical Resistivity Cone for Ground Water Contamination Studies', *Can. Geotech. J.*, **27**, 557–567 (1990).
22. A.J. Rutter, J.R. Webster, 'Probes for Sampling Ground Water for Gas Analysis', *J. Ecol.*, **50**, 615–618 (1962).
23. R.C. Reeve, E.J. Doering, 'Sampling the Soil Solution for Salinity Appraisal', *Soil Sci.*, **99**, 339–344 (1965).
24. P.H. John, M.A. Lock, M.M. Gibbs, 'Two New Methods for Obtaining Water Samples from Shallow Aquifers and Littoral Sediments', *J. Environ. Qual.*, **6**, 322–324 (1977).
25. W.D. Harrison, T.E. Osterkamp, 'A Probe Method for Soil Water Sampling and Subsurface Measurements', *Water Resour. Res.*, **17**, 1731–1736 (1981).
26. K.R. Patton, 'A Cost-effective Screening Technique Used in the Characterization of Agricultural Compounds in Ground Water', *Ground Water Manage.*, **1**, 831–839 (1990).
27. K.E. Cordry, 'HydroPunch™ II – The Second Generation. A New In-situ Ground Water Sampling Tool', *Ground Water Manage.*, **5**, 715–723 (1991).
28. K.E. Cordry, 'Ground Water Sampling Without Wells', *Proceedings of the 6th National Symposium and Exposition on Aquifer Restoration and Ground Water Monitoring*, National Water Well Association, Dublin, OH, 262–271, 1986.
29. R.W. Edge, K. Cordry, 'The Hydropunch™: An in situ Sampling Tool for Collecting Ground Water from Unconsolidated Sediments', *Ground Water Monit. Rev.*, **9**, 177–183 (1989).
30. P. Fierro, Jr, J.E. Mizerany, 'Utilization of Cone Penetrometer Technology as a Rapid, Cost-effective Investigative Technique', *Proceedings of the Seventh National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 473–485, 1993.
31. S.R. Archabal, J.R. Hicks, M.C. Reimann, 'Application of Hydropunch® and Cone Penetrometer Techniques to Subsurface Investigation at a Solvent-contaminated Site', *Proceedings of the Ninth National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 371–386, 1995.
32. D.A. Zemo, T.A. Delfino, J.D. Gallinatti, V.A. Baker, L.R. Hilpert, 'Field Comparison of Analytical Results from Discrete-depth Ground Water Samplers: BAT™ Enviroprobe and QED Hydropunch™', *Ground Water Manage.*, **11**, 341–355 (1992).
33. B.-A. Torstensson, 'A New System for Ground Water Monitoring', *Ground Water Monit. Rev.*, **4**, 131–138 (1984).
34. S. Haldorsen, A.M. Petsonk, B.-A. Torstensson, 'An Instrument for in situ Monitoring of Water Quality and Movement in the Vadose Zone', *Proceedings of the NWWA Conference on Characterization and Monitoring of the Vadose (Unsaturated) Zone*, National Water Well Association, Worthington, OH, 158–172, 1985.
35. B.-A. Torstensson, A.M. Petsonk, 'A Hermetically Isolated Sampling Method for Ground-water Investigations', in *Ground-water Contamination: Field Methods*, eds. A.G. Collins, A.I. Johnson, ASTM STP 963, American Society for Testing and Materials, Philadelphia, PA, 274–289, 1988.
36. B.S. Mines, J.L. Davidson, D. Bloomquist, T.B. Stauffer, 'Sampling of VOCs with the BAT Ground Water Sampling System', *Ground Water Monit. Rev.*, **13**, 115–120 (1993).
37. B. Pemberton, 'MetLab Phase II Sample Analysis Report', SRT-ESS-96-280, Westinghouse Savannah River Co., Aiken, S.C., 1996.
38. S.T. Litherland, T.W. Hoskings, R.L. Boggess, 'A New Ground Water Survey Tool: The Combined Cone

Penetrometer/Vadose Zone Vapor Probe', *Proceedings of the NWWA/API Conference on Petroleum Hydrocarbons and Organic Chemicals in Ground Water – Prevention, Detection and Restoration*, National Water Well Association, Dublin, OH, 322–330, 1985.

39. M.D. Varljen, 'Combined Soil Gas and Ground Water Field Screening Using the HydroPunch® and Portable Gas Chromatography', *Proceedings of the Seventh National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 499–509, 1993.
40. D.A. Leavell, L.T. Lee, Jr, 'Design, Development, and Operation of the Multiport Sampler', Technical Report GL-95-16, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, 1995.
41. R.L. Siegrist, P.D. Jenssen, 'Evaluation of Sampling Method Effects on Volatile Organic Compound Measurements in Contaminated Soils', *Environ. Sci. Technol.*, **24**, 1387–1392 (1990).
42. K.F. Myers, J.M. Brannon, R.A. Karn, C.B. Price, D.Y. Eng, A.B. Strong, S.S. Cooper, 'Laboratory Evaluation of a Volatile Organic Compound Analysis System for the Site Characterization and Analysis Penetrometer System', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 177–184, Vol. 1, 1995.
43. K.F. Myers, J.M. Brannon, R.A. Karn, C.B. Price, D.Y. Eng, A.B. Strong, S.S. Cooper, 'Laboratory Evaluation of a Volatile Organic Compound Analysis System for the Site Characterization and Analysis Penetrometer System', Technical Report IRRP-95-3, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, 1995.
44. A. Gorshteyn, S. Smarason, A. Robbat, Jr, 'Speciation of Subsurface Contaminants by Cone Penetrometry Gas Chromatography/Mass Spectrometry', *Environ. Sci. Technol.*, **33**, 2474–2480 (1999).
45. S.P. Farrington, W.L. Bratton, 'A Thermal Desorbing CPT Probe for Sampling Volatile Organic Compounds', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 530–539, 1997.
46. P.V. Doskey, J.H. Aldstadt, J.M. Kuo, M.S. Costanza, M.D. Erickson, 'In Situ Measurement of Volatile Organic Compounds in Ground Water by Methods Coupled to the Cone Penetrometer', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 151–155, Vol. 1, 1995.
47. C.R. Morlock, 'An In Situ Sparging Sampler for Volatile Organic Hydrocarbon Analysis', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 953–960, Vol. 2, 1995.
48. J.H. Aldstadt, C.H. Batson, A.F. Martin, M.D. Erickson, M.S. Costanza, M.B.J. Foster, T.R. James, P.V. Doskey, 'Development and Preliminary Evaluation of a Cone Penetrometer-based Method for In Situ Determination of Volatile Organic Compounds in Ground Water', *Field Anal. Chem. Technol.*, **1**, 239–247 (1997).
49. W.L. Bratton, S.P. Farrington, 'Measuring Fuel Contamination Using Cone Penetration Testing and High Speed Gas Chromatography', in *Proceedings of the Conference on Industry Partnerships to Deploy Environmental Technology*, U.S. Department of Energy's Morgantown Energy Technology Center, Morgantown, WV, 355–365, 1996.
50. W.M. Davis, J.F. Powell, J. Furey, C.V. Thompson, M. Wise, 'Rapid Subsurface Detection of Volatile Organic Compounds in Ground Water', *Preprints of Papers*, 213th ACS National Meeting, San Francisco, CA, Division of Environmental Chemistry, Inc., American Chemical Society, Washington, DC, 47–49, 1997.
51. K. Cordry, 'PowerPunch™: A Self Completing Direct Push Well', *Proceedings of the Ninth National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 363–369, 1995.
52. G. Matz, P. Kesners, 'Spray and Trap Method for Water Analysis by Thermal Desorption Gas Chromatography/Mass Spectrometry in Field Applications', *Anal. Chem.*, **65**, 2366–2371 (1993).
53. M.A. Klemp, M.L. Akard, R.D. Sacks, 'Cryofocusing Inlet with Reverse Flow Sample Collection for Gas Chromatography', *Anal. Chem.*, **65**, 2516–2521 (1993).
54. M. Klemp, A. Peters, R. Sacks, 'High-speed GC Analysis of VOCs: Sample Collection and Inlet Systems', *Environ. Sci. Technol.*, **28**, 369A–376A (1994).
55. P.V. Doskey, M.S. Costanza, M.C. Hansen, W.T. Kickels, 'A Solid Sorbent Method for the Collection and Analysis of Volatile Halogenated Organic Compounds in Soil Gas', *J. Chrom. A*, **738**, 73–81 (1996).
56. R.A. Jenkins, C.E. Higgins, T.M. Gayle, G.W. Allin, R.R. Smith, 'Field Experiences with a Multisorbent Arrayed Sampler for In-situ Collection of Vadose Zone Volatile Organic Compounds', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 1018–1024, Vol. 2, 1993.
57. M. Chai, C.L. Arthur, J. Pawliszyn, R.P. Belardi, K.F. Pratt, 'Determination of Volatile Chlorinated Hydrocarbons in Air and Water with Solid-phase Microextraction', *Analyst*, **118**, 1501–1505 (1993).
58. M.J. Yang, J. Pawliszyn, 'Multiplex Gas Chromatography with a Hollow Fiber Membrane Interface for Determination of Trace Volatile Organic Compounds in Aqueous Samples', *Anal. Chem.*, **65**, 1758–1763 (1993).
59. D.F. Hagen, C.G. Markell, G.A. Schmitt, D.D. Blevins, 'Membrane Approach to Solid-phase Extractions', *Analyt. Chim. Acta*, **236**, 157–164 (1990).

60. C. Markell, D.F. Hagen, V.A. Bunnelle, *LC–GC*, **9**, 332–337 (1990).
61. C. Markell, D.F. Hagen, 'An Overview of Solid Phase Extraction Disk Technology as a Sample Preparation Method for Semivolatiles in Water', Presented at the *American Water Works Association Water Quality Technology Conference*, Orlando, FL, 1991.
62. W.K. Reagen, A.-D. Vo, 'Method for On-site Analysis of Pesticides in Ground Water-unfiltered SPE Disc', *AT-ONSITE^{SM}*, **1**, 27–33 (1995).
63. S.J. Bauer, R.G. Cooks, 'MIMS for Trace-level Determination of Organic Analytes in On-line Process Monitoring and Environmental Analysis', *Am. Lab.*, October, 36–51 (1993).
64. M.J. Yang, S. Harms, Y.Z. Luo, J. Pawliszyn, 'Membrane Extraction with a Sorbent Interface for Capillary Gas Chromatography', *Anal. Chem.*, **66**, 1339–1346 (1994).
65. P.V. Doskey, J.H. Aldstadt, J.M. Kuo, M.S. Costanza, 'Evaluation of an In Situ, On-line Purging System for the Cone Penetrometer', *J. Air Waste Manage. Assoc.*, **46**, 1081–1085 (1996).
66. S.E. Carpenter, P.V. Doskey, M.D. Erickson, P.C. Lindahl, 'Performance Specifications for Technology Development: Application for Characterization of Volatile Organic Compounds in the Environment', Technical Report ANL/ER/TM-3, Argonne National Laboratory, Argonne, IL, 1994.
67. E.N. Koglin, E.J. Poziomek, M.L. Kram, 'Emerging Technologies for Detecting and Measuring Contaminants in the Vadose Zone', in *Handbook of Vadose Zone Characterization & Monitoring*, eds. L.G. Wilson, L.G. Everett, S.J. Cullen, Lewis Publishers, Ann Arbor, MI, 657–700, 1995.
68. J. Janata, A. Bezegh, 'Chemical Sensors', *Anal. Chem.*, **60**, 62R–74R (1988).
69. H. Wohltjen, R. Dessy, 'Surface Acoustic Wave Probe for Chemical Analysis. I. Introduction and Instrument Description', *Anal. Chem.*, **51**, 1458–1464 (1979).
70. H. Wohltjen, 'Mechanism of Operation and Design Considerations for Surface Acoustic Wave Device Vapour Sensors', *Sens. Actuators*, **5**, 307–325 (1984).
71. M.D. Ward, D.A. Buttry, 'In Situ Interfacial Mass Detection with Piezoelectric Transducers', *Science*, **249**, 1000–1007 (1990).
72. A.R. Katritzky, R.J. Offerman, 'The Development of New Microsensor Coatings and a Short Survey of Microsensor Technology', *CRC Crit. Rev. Anal. Chem.*, **21**, 83–113 (1989).
73. A. Mierzwinski, Z. Witkiewicz, 'Piezoelectric Detectors Coated with Liquid-crystal Materials', *Talanta*, **34**, 865–871 (1987).
74. D.S. Ballantine, Jr, H. Wohltjen, 'Surface Acoustic Wave Devices for Chemical Analysis', *Anal. Chem.*, **61**, 704A–715A (1989).
75. G.C. Guilbault, J.M. Jordan, 'Analytical Uses of Piezoelectric Crystals: A Review', *CRC Crit. Rev. Anal. Chem.*, **19**, 1–28 (1988).
76. W.P. Carey, B.R. Kowalski, 'Chemical Piezoelectric Sensor and Sensor Array Characterization', *Anal. Chem.*, **58**, 3077–3084 (1986).
77. S.A. Batterman, E.T. Zellers, 'Assessment of Subsurface VOCs Using a Chemical Microsensor Array', Final Report, US Department of Energy, Office of Environmental Restoration and Waste Management, Office of Research and Development, Technology Development, Washington, DC, 1993.
78. J.W. Grate, S.J. Martin, R.M. White, 'Acoustic Wave Microsensors, Part I', *Anal. Chem.*, **65**, 940A–948A (1993).
79. J.W. Grate, S.J. Martin, R.M. White, 'Acoustic Wave Microsensors, Part II'., *Anal. Chem.*, **65**, 987A–996A (1993).
80. J.W. Grate, S.L. Rose-Pehrsson, D.L. Venezky, M. Klusty, H. Wohltjen, 'Smart Sensor System for Trace Organophosphorus and Organosulfur Vapor Detection Employing a Temperature-controlled Array of Surface Acoustic Wave Sensors, Automated Sample Preconcentration and Pattern Recognition', *Anal. Chem.*, **65**, 1868–1881 (1993).
81. G.C. Frye, D.W. Gilbert, C. Colburn, R.W. Cernosek, T.D. Steinfort, 'Above-ground and In Situ Field Screening of VOCs Using Portable Acoustic Wave Sensor (PAWS) Systems', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 715–726, Vol. 2, 1995.
82. M.A. Arnold, 'Fiber-optic Chemical Sensors', *Anal. Chem.*, **64**, 1015A–1025A (1992).
83. W.R. Seitz, 'Chemical Sensors Based on Immobilized Indicators and Fiber Optics', *CRC Crit. Rev. Anal. Chem.*, **19**, 135–173 (1988).
84. S.S. Cooper, P.G. Malone, 'Device for Measuring Reflectance and Fluorescence of In-situ Soil', (to US Army), US Pat. 5,128,882, 1992.
85. S.E. Apitz, L.M. Borbridge, G.A. Theriault, S.H. Lieberman, 'Remote In Situ Determination of Fuel Products in Soils: Field Results and Laboratory Investigations', *Analysis*, **20**, 461–474 (1992).
86. US Environmental Protection Agency, 'Site Characterization Analysis Penetrometer System (SCAPS): Innovative Technology Evaluation Report', EPA No. 540/R-95/520, Office of Research and Development, Washington, DC, 1995.
87. S.H. Lieberman, 'Proceedings of the SCAPS Workshop on Sensors for the U.S. Army Environmental Center', Denver, CO, 1–7, 1993.
88. G.A. Theriault, R. Newberry, J.M. Andrews, S.E. Apitz, S.H. Lieberman, 'Fiber Optic Fluorometer System Based on a Dual Wavelength Laser Excitation Source', *Proc. SPIE*, **1796**, 115–123 (1992).

89. W.L. Bratton, J.D. Shinn, J.L. Bratton, 'Case Studies of Measuring Fuel Contamination Using In-situ Fluorescence Techniques', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 431–442, Vol. 1, 1993.
90. G.D. Gillispie, R.W.St. Germain, 'Performance Characterization of the Rapid Optical Screen Tool (ROST™)', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air & Waste Management Association, Pittsburgh, PA, 478–489, Vol. 1, 1995.
91. J. Lin, S.J. Hart, T.A. Taylor, J.E. Kenny, 'Laser Fluorescence EEM Probe for Cone Penetrometer Pollution Analysis', *Proc. SPIE*, **2367**, 70–79 (1994).
92. A.W. Miziolek, E.R. Cespedes, 'Spectroscopic Analysis of Heavy Metal Contamination of the Environment', *Optics Photon. News*, **7**, 39–41 (1996).
93. D.A. Alexander, D.A. Poulain, 'Quantitative Analysis of the Detection Limits for Heavy-metal-contaminated Soils by Laser-induced Breakdown Spectroscopy', Final Report for US Army Corps of Engineers, Contract No. DACA39-95-K-0053, University of Nebraska, Lincoln, NE, 1996.
94. W.T. Elam, R.R. Whitlock, J.V. Gilfrich, 'Use of X-ray Fluorescence for In-situ Detection of Metals', *Proc. SPIE*, **2367**, 59–69 (1994).
95. W.T. Elam, J.W. Adams, K.R. McDonald, J.V. Gilfrich, 'Design of an X-ray Fluorescence Sensor for the Cone Penetrometer', *Field Anal. Chem. Technol.*, **2**, 97–102 (1998).
96. C.R. Morlock, 'In Situ Ultraviolet Fluorescence Probe for Detection of Volatile Organic Hydrocarbons in Ground Water', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 408–415, Vol. 1, 1995.
97. J.H. Ballard, S.S. Cooper, J.C. Morgan, W.R. Lawrence, B.E. Reed, 'Probe Device for Detecting Contaminants in Subsurface Media', (to US Army), US Pat. 5,548,115, 1996.
98. P.A. Mosier-Boss, R. Newberry, S.H. Lieberman, 'Development of a Cone Penetrometer Deployed Solvent Sensor Using a SERS Fiber Optic Probe', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 588–599, 1997.
99. K.J. Ewing, G. Nau, F. Bucholtz, I.D. Aggarwal, G. Robitaille, 'Fiber Optic Infrared Reflectance Sensor for Detection of Solvents in Soil', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, Vol. 1, 144–150, 1995.
100. K.J. Ewing, G. Nau, F. Bucholtz, I.D. Aggarwal, 'IR Fiber Optic Chemical Sensors for Hazardous Waste Detection', *Proc. SPIE*, **2504**, 68–74 (1995).
101. S.T. Vohra, F. Bucholtz, G.M. Nau, K.J. Ewing, I.D. Aggarwal, 'Remote Detection of Trichloroethylene in Soil by a Fiber-optic Infrared Reflectance Probe', *Appl. Spectros.*, **50**, 985–990 (1996).
102. H. Arman, S.M. Klainer, J.R. Thomas, 'An On-line Fiber Optic Chemical Sensor (FOCS™) System for Monitoring Above and Below Ground Hydrocarbon Storage Tanks', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 352–361, Vol. 1, 1993.
103. R.P. De Filippi, T.J. Cody, Jr, 'A Monitoring System for Hydrocarbon Leakage from Petroleum and Chemical Product Sites', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 1241–1251, Vol. 1, 1993.
104. M.R. Shahriari, J. Liu, G.H. Sigel, Jr, M. Liva, 'Ruggedized Fiber Optic Environmental Sensors Based on Porous Optical Materials', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 175–182, Vol. 1, 1993.
105. M. Butler, S. Klainer, J. Tussey, J. Thomas, 'A Two Channel Portable Fluorimeter for Use with Fiber Optic Chemical Sensors (FOCS™)', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 183, Vol. 1, 1993.
106. F.P. Milanovich, P.F. Daley, K. Langry, B.W. Colston, S.B. Brown, S.M. Angel, 'A Fiberoptic Sensor for Continuous Monitoring of Chlorinated Hydrocarbons', *Proceedings of the 2nd International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 43–48, 1991.
107. J. Rossabi, B. Colston, Jr, S. Brown, F. Milanovich, L.T. Lee, Jr, 'In Situ, Subsurface Monitoring of Vapor-phase TCE Using Fiber Optics', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 1165–1173, Vol. 2, 1993.
108. J. Rossabi, B. Colston, Jr, S. Brown, F. Milanovich, L.T. Lee, Jr, 'In Situ, Subsurface Monitoring of Vapor-phase TCE Using Fiber Optics', Technical Report, WSCR-MS-247, Westinghouse Savannah River Co., Aiken, SC, 1993.
109. J.C. Wells, P.G. Blystone, M.D. Johnson, W.R. Haag, B.W. Colston, Jr, 'A Commercialized Fiber Optic Sensor for the Detection of Volatile Chlorinated Hydrocarbons', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 345–351, Vol. 1, 1993.
110. C.E. Grey, S.S. Cooper, P.G. Malone, 'Method and Apparatus for In-situ Detection and Determination of Soil Contaminants', (to US Army), US Patent 5,246,862, 1993.

111. Z. Cao, W.J. Buttner, J.R. Stetter, 'The Properties and Application of Amperometric Gas Sensors', *Electroanalysis*, **4**, 253–266 (1992).
112. W.R. Penrose, L. Pan, J.R. Stetter, W.M. Ollison, 'Sensitive Measurement of Ozone Using Amperometric Gas Sensors', *Anal. Chim. Acta*, **313**, 209–219 (1995).
113. J. Yinon, S. Zitrin, *Modern Methods and Applications in Analysis of Explosives*, John Wiley & Sons, Chichester, UK, 1993.
114. T.F. Jenkins, M.E. Walsh, 'Development of Field Screening Methods for TNT, 2,4-DNT, and RDX in Soil', *Talanta*, **39**, 419–428 (1992).
115. W.J. Buttner, M. Findlay, W. Vickers, W.M. Davis, E.R. Cespedes, S.S. Cooper, J.W. Adams, 'In situ Detection of Trinitrotoluene and Other Nitrated Explosives in Soils', *Anal. Chim. Acta*, **341**, 63–71 (1997).
116. J.W. Adams, W.M. Davis, E.R. Cespedes, W.J. Buttner, M.W. Findlay, 'Development of Cone Penetrometer Electro-chemical Sensor Probes for Chlorinated Solvents and Explosives', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 667–670, 1997.
117. E.R. Cespedes, S.S. Cooper, W.M. Davis, W.J. Buttner, W.C. Vickers, 'In Situ Detection of TNT Contamination Using Electrochemical Sensors in Cone Penetrometer System', *Proc. SPIE*, **2367**, 33–42 (1994).
118. N.L. Jarvis, H. Wohltgen, 'Miniature GC for In-situ Monitoring of VOCs within a Cone Penetrometer', Final Technical Report, Contract No. DE-AR21-94MC31187, US Department of Energy, Morgantown Energy Technology Center, Morgantown, WV, 1996.
119. J.C. Morgan, J.W. Adams, J.H. Ballard, 'Field Demonstration of the SCAPS Cone Penetrometer Gamma Probe for Rad-waste Detection', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 600–611, 1997.
120. R. Sacks, M. Akard, 'High-speed GC Analysis of VOCs: Tunable Selectivity and Column Selection', *Environ. Sci. Technol.*, **28**, 428A–433A (1994).
121. R. Sacks, M. Klemp, M. Akard, 'High-speed Capillary Column GC for Rapid Screening of Gasoline to Diesel Range Organic Compounds', *Field Anal. Chem. Technol.*, **1**, 97–102 (1996).
122. W.C. McDonald, M.D. Erickson, B.M. Abraham, A. Robbat, Jr, 'Developments and Applications of Field Mass Spectrometers', *Environ. Sci. Technol.*, **28**, 336A–343A (1994).
123. M.B. Wise, C.V. Thompson, M.V. Buchanan, R. Merriweather, M.R. Guerin, 'Direct Sampling Ion Trap Mass Spectrometry', *Spectroscopy*, **8**, 14–22 (1993).
124. M.B. Wise, C.V. Thompson, M.R. Guerin, R.A. Jenkins, 'Development and Testing of a Field Transportable Direct Sampling Ion Trap Mass Spectrometer', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 543–556, Vol. 1, 1993.
125. A. Robbat, Jr, T.-Y. Liu, B. Abraham, 'Evaluation of a Thermal Desorption Gas Chromatograph/Mass Spectrometer: On-site Detection of Polychlorinated Biphenyls at a Hazardous Waste Site', *Anal. Chem.*, **64**, 358–364 (1992).
126. A. Robbat, Jr, T.-Y. Liu, B. Abraham, 'On-site Detection of Polycyclic Aromatic Hydrocarbons in Contaminated Soils by Thermal Desorption Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **64**, 1477–1483 (1992).
127. E. Lightner, Personal communication, US DOE, Germantown, MD, 1997.
128. T.L. Lammons, T.A. Mann, A.M. Pelletier, 'An Assessment Tool for Sampling Ground Water in the Undisturbed Saturated Zone', *Proceedings of the Third National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 345–356, 1989.
129. C.L. Bergren, R.C. Tuckfield, N.M. Park, 'Suitability of the Hydropunch™ for Assessing Ground Water Contaminated by Volatile Organics', *Proceedings of the Fourth National Outdoor Action Conference on Aquifer Restoration, Ground Water Monitoring, and Geophysical Methods*, National Water Well Association, Dublin, OH, 387–401, 1990.
130. D.A. Zemo, T.A. Delfino, J.D. Gallinatti, V.A. Baker, L.R. Hilpert, 'Field Comparison of Analytical Results from Discrete-depth Ground Water Samplers', *Ground Water Monit. Rev.*, **15**, 133–141 (1995).
131. A.D. Hewitt, 'Comparison of Site Characterization for Trichloroethylene by Headspace Gas Chromatography and Purge-and-trap Gas Chromatography Mass Spectrometry', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 195–202, Vol. 1, 1993.
132. T.M. Christy, S.C. Spradlin, 'The Use of Small Diameter Probing Equipment for Contaminated Site Investigation', *Ground Water Manage.*, **11**, 87–101 (1992).
133. D.P. Lucero, 'Soil Gas Sampler Implant for Monitoring Dump Site Subsurface Hazardous Fluids', *Proceedings of the 6th National Conference on Hazardous Wastes and Hazardous Materials*, Hazardous Materials Control Research Institute, Silver Spring, MD, 548–553, 1989.
134. D.P. Lucero, 'A Soil Gas Sampler Implant for Monitoring Dump Site Subsurface Hazardous Fluids', *Haz. Mater. Control*, **3**, 36–44 (1990).
135. D.P. Lucero, R.H. Ilgner, R.R. Smith, R.A. Jenkins, 'Subsurface Fluids Screening by an Analytical System Employing a Diffusion-limited and Implantable Sampling Module Deployable with a Cone Penetrometer', in *Field Analytical Methods for Hazardous Wastes and*

Toxic Chemicals, Air and Waste Management Association, Pittsburgh, PA, 518–529, 1997.

136. W.M. Davis, J.F. Powell, K. Konecny, J. Furey, C.V. Thompson, M. Wise, G. Robitaille, 'Rapid In-situ Determination of Volatile Organic Contaminants in Ground Water Using the Site Characterization and Analysis Penetrometer System', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 464–469, 1997.
137. A.D. Hewitt, 'Comparison of Methods for Sampling Vadose Zone Soils for Determination of Trichloroethylene', *J. AOAC Int.*, **77**, 458–463 (1994).
138. J.D. Schroeder, S.R. Booth, L.D. Trocki, 'Cost Effectiveness of the Site Characterization and Analysis Penetrometer System', Technical Report LA-UR-91-4016, Los Alamos National Laboratory, Los Alamos, NM, 1991.
139. Naval Facilities Engineering Command, 'Evaluation of Potential Savings that May Be Achieved Using the Site Characterization and Analysis Penetrometer System at Navy-owned Petroleum Contaminated Sites', Technical Memorandum TM-2025-ENV, Port Hueneme, CA, 1–14, 1994.
140. T.F. Jenkins, C.L. Grant, G.S. Brar, P.G. Thorne, T.A. Ranney, P.W. Schumacher, 'Sampling Error Associated with Collection and Analysis of Soil Samples at TNT-contaminated Sites', *Field Anal. Chem. Technol.*, **1**, 151–163 (1997).
141. G.A. Theriault, S. Bodensteiner, S.H. Lieberman, 'A Cone Penetrometer Probe for the In-situ Detection of Metals in Soils', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 690–701, 1997.
142. B.H. Miles, J. Cortes, E.R. Cespedes, 'Laser Induced Breakdown Spectroscopy (LIBS) Detection of Heavy Metals Using a Cone Penetrometer: System Design and Field Investigation Results', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 671–689, 1997.
143. W.T. Elam, J. Adams, K.R. Hudson, B. McDonald, D. Eng, G. Robitaille, I. Aggarwal, 'Field Demonstration of the SCAPS XRF Metals Sensor', in *Field Analytical Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 681–689, 1997.

Electrochemical Sensors for Field Measurements of Gases and Vapors

James A. Cox and Scott D. Holmstrom
Miami University, Oxford, USA

1 Introduction **1**
2 Background **2**
2.1 History 2
2.2 Classes of Electrochemical Sensors 2
2.3 Analytical Performance 3
2.4 Contributions of Materials Science to Sensor Development 4
3 Scope **5**
3.1 Environmental Monitoring 5
3.2 Biomedical Applications 6
3.3 Sensors in Process Monitoring and the Work Place 8
3.4 Other Applications 9
4 Selected Applications of Electrochemical Gas Sensors **10**
4.1 Amperometric Sensors 10
4.2 Potentiometric Sensors 12
4.3 Conductivity Sensors 13
4.4 Comparison to Related Devices 14
5 Summary **15**
Abbreviations and Acronyms **15**
Related Articles **15**
References **16**

Sensors that utilize signals produced by redox of a targeted species (analyte) at an electrode under controlled potential, the conductance changes produced by contact of a chemical species with an electrolyte, or a change in the equilibrium potential of an electrode upon contact with an analyte are being employed in studies of gaseous samples of environmental, biomedical, and industrial interest. Changes in the work function of chemically modified field effect transistors (FET) also is a strategy related to both conductivity and potentiometry. Electrochemical sensors are fundamentally advantageous over other types in that they directly convert information from the chemical to the electrical domain, rather than requiring another transducer. Yet, few have found commercial application. The common types that are employed for field measurements of gaseous samples include devices to quantify oxygen, hydrogen sulfide, sulfur dioxide, oxides of nitrogen, hydrogen cyanide, hydrogen chloride, and chlorine. The limitation is not sensitivity; by this measure electrochemistry is among the best of the analytical measurement methods. Amperometric sensors accrue selectivity by control of the electrode potential along with control of the process by which the analyte partitions from the gas to the electrolyte. The latter is the primary means by which conductivity and potentiometry systems achieve sensitivity. These methods are not as selective as those in which a chemical reaction is incorporated in the sensor. The most selective electrochemical sensors employ an enzymatic reaction as part of signal production, but these are applicable to only a few analytes, such as alcohols, of importance in the gas phase.

1 INTRODUCTION

A challenge in designing chemical sensors that are suited to operate as stand-alone devices is that an attendant separation method is not convenient, so the detection system must be highly selective. An additional problem is that any chemical step that is associated with the measurement, such as conversion of the targeted substance (analyte) to a species amenable to producing a signal, must operate in a small, isolated environment without a reagent reservoir. In that the output of a sensor must be an electrical signal, electrochemical systems have less-complex designs than other sensors where the conversion of information from the chemical to the electrical domain requires two steps. Moreover, the attendant chemistry, if any, typically is simple. Yet, commercial electrochemical sensors are available for only a limited number of analytes, including H_2S, O_2, CO, SO_2, NO, HCN, HCl, NO_2, and Cl_2,[1] primarily because of the problem of attaining sufficient selectivity.

In this article, the scope of field applications of electrochemical sensors for gaseous chemical species is described. Field studies are broadly defined to include measurements both at remote sites and in environments where the stand-alone characteristic of these devices favors their use over conventional, bench-top instrumentation. An emphasis is placed on the direction of research in this field and on the analytical figures-of-merit that can be achieved.

Electrochemical sensors generally fall into one of three categories. Potentiometric devices have as their analytical signal the potential across a cell. Because the time required to develop an equilibrium potential is related to the resistance of the layer between the electrodes, the electrodes are separated by an ionic conductor, which can be a solid or a liquid. Amperometric sensors provide a

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

signal from redox at the indicator electrode surface under controlled potential. In the third category, the signal is the conductivity of the electrolyte which changes in response to the analyte. Other types include galvanic cells and devices in which the signal is related to the work function of a coating, especially on a FET.

Electrochemical gas sensors often involve a two-step process, the partitioning of the analyte from the gas phase onto the electrolyte[2] and the signal-generation step. The electrolyte can be a solid, a liquid, or a combination of the two. Its function is to provide an ionically conducting connection between the electrodes. With liquid-containing electrolytes, the partitioning is by dissolving the gas-phase analyte into the liquid. Most often, the signal generation of such sensors is by redox at an indicator (working) electrode. Unless the sensor design has a three-phase boundary among the gas, liquid, and electrode, the response time is slowed by mass transport of the liquid-phase analyte to the electrode. With solid electrolytes, a variety of signal-generation processes are employed. Here, partitioning is by adsorption. In cases such as the commercial SnO_2-based CO sensor, which operates at about 400 °C, the adsorption step is directly coupled with an oxygen-atom exchange in the semiconductor that leads to a rapid change in the Fermi level and, hence, the threshold voltage of the device.[3] When redox is the signal-generation step, the response time and sensitivity is limited by mass transport in the solid, as detailed later. Solid-state systems often are based on a chemical reactions between the analyte and sensing material, a process that changes the resistance and/or work function of the device.

2 BACKGROUND

2.1 History

Electrochemical gas-phase sensors first became important with the development of the oxygen electrode. This sensor evolved from the simple measurement of dissolved oxygen by measurement of the current due to its reduction in a polarographic cell except that a platinum indicator electrode was used. To minimize interference from surface-active species in the sample, the Pt was coated with a membrane; however, Clark advanced the technology by isolating a film of supporting electrolyte from the sample with an oxygen-permeable membrane.[4,5] As shown in Figure 1, this now-common design monitors analytes in either gas or liquid samples. The selectivity is accrued by the relative rates of diffusion of the analyte and the interferents through the choice of the membrane and the selection of the appropriate combination of the supporting electrolyte and applied potential. For example, by using a mixture of 0.1 M KCl and 0.01 M $ZrOCl_2$ as the supporting electrolyte, a mercury indicator electrode, an applied potential of −1.25 V vs Ag/AgCl, and an anion-exchange membrane as the barrier, the system serves as a sensor for NO_3^- and presumably for NO_2 in the gas phase.[6]

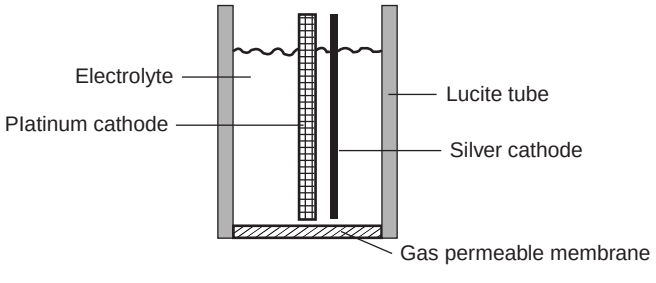

Figure 1 Clark-type amperometric sensor.

A significant improvement in the general design in Figure 1 was the replacement of the film of liquid electrolyte with a solid polymer electrolyte (SPE). The use of SPEs originated in fuel cell and battery technology, but adaptation to sensors yielded robust systems with long-term stability if a reservoir is provided to keep the SPE wet. The SPEs are ion-exchange polymers such as Nafion® (DuPont) and Tosflex® (Tosoh), which are perfluorinated cation- and anion-exchange materials, respectively. Recent applications have employed porous electrode materials that were supported on the membrane, which exposes the electrode/SPE interface directly to the gaseous sample, thereby providing rapid response.[7]

Solid-state gas sensors that are based on metal oxide electrolytes date to the early 1960s.[8] An oxide such as ZrO_2 that is doped to provide mobile O^{2-} sites comprises a common form. One approach to measurement is to have the analyte perturb the O^{2-} population (e.g. chemisorption and decomposition of an alcohol yields H atoms that combine with O^{2-}), thereby changing the conductivity. The electrolytes also can be configured to provide amperometric or potentiometric measurement. A major application of these sensors is the determination of oxygen, especially where operation at elevated temperature is required.

2.2 Classes of Electrochemical Sensors

The process that generates the signal is the basis of the classification method in this chapter. With amperometric sensors, the signal is the current that is produced upon oxidation or reduction of the analyte at an electrode surface. In general, the current can be limited by the rate of the electron-transfer reaction or the rate of transport of the analyte to the electrode surface; however,

the former is not useful for analytical determinations. The rate of electron transfer is a function of applied potential. By applying a potential that is well-positive of the standard reduction potential, E^0, the condition in which an oxidation is limited by mass transfer often can be achieved. For reduction to serve as the current-generating step, the applied potential is well-negative of E^0. It should be noted that many electrochemical reactions that are predicted on the basis of E^0 are too slow in the accessible potential range to provide a useful current. In this case, either a suitable electrocatalyst must be employed or some other means of sensing must be used.

Amperometric sensors operate under conditions in which the mode of mass transport to the electrode surface is diffusion. In this case the flux of the analyte, f, is given by Equation (1):

$$f = D\left(\frac{\partial C}{\partial x}\right)_{x=0} \tag{1}$$

where D is the diffusion coefficient, C is the concentration of the analyte, and x is the distance from the electrode surface. For reliable measurements, the sensor must operate at steady-state. Because C is zero at $x = 0$ at the operating potential described above and the slope of the plot of C vs x near the electrode is nearly linear, the partial derivative in Equation (1) is proportional to the sample concentration. The analytical current, i, is given by Equation (2):

$$i = nFAf \tag{2}$$

where n is the number of electrons transferred per mole; F, the Faraday, and A, the electrode area. So the concentration is related to current by Equation (3):

$$i = \frac{nFADC}{\delta} \tag{3}$$

where δ is the boundary layer containing the gradient. This treatment is valid for a single phase, so in a gas sensor, where partitioning between phases can occur, it is necessary to assume that C in Equation (3) (i.e. the concentration in the electrolyte) is directly proportional to the concentration of the analyte in the gas phase in order to achieve linear calibration curves.

Potentiometric sensors are available in a wide variety of configurations, obviating a simple treatment of the signal-generating process. In the simplest case, the measured signal is the potential of an indicator electrode vs a reference electrode under equilibrium conditions, and the potential is related to concentration by the Nernst equation. For example, if $HCl_{(gas)}$ is partitioned into an aqueous solution (receiver) contacted by a reference electrode and a Ag/AgCl electrode Equation (4),

$$\text{Reference//receiver, AgCl/Ag} \tag{4}$$

the $HCl_{(gas)}$ is indirectly measured as the Cl^- produced. The system can operate as a concentration cell by selecting Ag/AgCl in contact with a standard concentration of Cl^- as the reference. The measured potential, E, is given by Equation (5):

$$E = \left(\frac{RT}{F}\right) \log\left(\frac{[Cl^-]_{\text{receiver}}}{[Cl^-]_{\text{standard}}}\right) \tag{5}$$

A second simple system is configured as a pH electrode in contact with a thin film of solution; exposure to a gas with acid–base properties such as CO_2 changes the pH of the film. More complicated treatment is required when, for example, mixed potentials are developed, solid solution formation occurs, or the measured signal is the work function of a material. These will be treated in the description of applications in section 4. It should be noted that a common feature is that the measured signal is linear with respect to the logarithm of concentration, although over narrow ranges and at low concentrations the relationship approaches linearity without the logarithmic function.

Detection by conductivity is used in certain gas sensors that employ solid electrolytes. A report by Chakraborty et al.[9] is illustrative of this class of sensor. A glass consisting of Fe_2O_3–Sb–TeO_2 showed semiconducting properties with charge-exchange occurring between mixed-valence iron sites. The d.c. conductivity in the temperature range 463–598 K was investigated in atmospheres containing various amounts of O_2. A decrease in conductivity upon exposure to O_2 was attributed to a decrease in Fe^{2+} from Equation (6):

$$O_{2(\text{adsorbed})} + Fe^{2+} \longrightarrow O_2^- + Fe^{3+} \tag{6}$$

which attenuated charge-exchange. Not considered was an ionic conductivity mechanism with O^{2-} as the charge carrier. As noted above, O_2^- often is considered the charge carrier in metal oxide systems such as SnO_2 and ZrO_2. In this regard, sensing of fuel gases with such devices has been attributed to consumption of O_2^- by the analyte.

2.3 Analytical Performance

Important factors in evaluating gas sensors include the detection limit (DL), sensitivity, linear dynamic range (LDR), response and recovery times, selectivity, short-term stability (repeatability), and long-term stability. The DL is the minimum concentration of the analyte that can be observed with a stated degree of certainty. Operationally, it is the concentration that yields a blank-corrected signal that is a factor, k times the standard deviation (SD) of the blank; generally, $k = 3$. Assuming that the blank uncertainty is random, that the measured

SD is representative of the population distribution, and that two measurements are required (namely, the blank and the sample) which is typically used, $k = 3$, provides detection at the 98.3% confidence level.[10] Published values of DL are useful as guides to the range over which a sensor may be used, but practical measurements at the DL generally are not possible. The DL is not a significant factor for sensors that monitor major components, but it is crucial for monitoring toxic materials.

The response time is generally reported as T_{90} or T_{50}, which are the times to reach 90% and 50% of the steady-state signal, respectively.[11] The limiting factor can be the sampling step or the measuring step. The former occurs with a barrier between the sample and the indicator electrode. For example, in many designs, the analyte must diffuse through a membrane and/or through a layer of electrolyte prior to reaching the indicator electrode. The time constant in this case is related to $(\Delta x)^2/D$ where Δx is the thickness of the barrier. Typical D-values are in the ranges ca. $10^{-1}\,cm^2\,s^{-1}$ in the gas phase, $10^{-5}\,cm^2\,s^{-1}$ for an aqueous system, and $10^{-8}\,cm^2\,s^{-1}$ or smaller in a solid. Values between liquid and solids are observed in SPEs such as Nafion®. The most rapid responses are observed with porous barriers where the pores are air-filled. Here, a gas/electrolyte/electrode three-phase boundary exists, and the response time becomes limited by the electrode response. A detailed analysis of a commercial CO detector was performed.[11] The diffusion path consisted of 0.2 cm through a sintered disk, 0.4 cm through air, 0.03 cm through a backing membrane, and 0.01 cm through an electrode membrane; the respective D-values in these media were 0.15, 0.21, 7×10^{-3}, and $2 \times 10^{-3}\,cm^2\,s^{-1}$, respectively. The T_{90}-value under these conditions was limited primarily by the electrode response with some evidence of a contribution from diffusion. This result suggests that in any system with a direct air-path to the electrode, T_{90} will be related to the electrode response, whereas in a system that involves mass transport of the analyte through an SPE with liquid-filled pores or through a solid, diffusion will be the factor that limits response time.

With an amperometric sensor, D also influences the sensitivity, which is the slope (m) of a linear calibration curve (see Equation 3). Because the units of a calibration curve are variable, the value of m does not have much value in terms of comparing various sensors for a given analyte. For such comparisons, a preferable parameter is the analytical sensitivity, which is m/SD (where SD is the standard deviation). LDRs also must be compared with caution in that a consensus definition is not available. Often the LDR is taken as the region of the calibration curve where the departure from linearity is less than 5%. Reports of selectivities and lifetimes also are not subject to defined standards.

2.4 Contributions of Materials Science to Sensor Development

Recent advances in electrochemical sensor technology have resulted primarily from new materials. An early improvement was the substitution of an SPE for an aqueous solution in the general design of an amperometric sensor (Figure 1). Nafion® is the most widely used SPE. The backbone is tetrafluoroethylene, and the structure contains side-chains of perfluorinated vinyl ethers with terminal sulfonic acid functionalities that provide strongly acidic cation-exchange sites. It is highly ordered and stable in aqueous solution under a wide range of acid–base and redox environments. The sulfonic acid sites generally are considered to exist in inverse micelles that are connected by neutral channels, but some studies have considered the ionic sites to be distributed along continuous channels.[12] Ionic conductance under dialysis conditions occurs with an activation energy of about $25\,kJ\,mol^{-1}$, which is evidence for the inverse micelle structure in Nafion® wetted with aqueous solution.[13]

A limitation on the use of Nafion® as the electrolyte is the influence of humidity on the response. Lee et al.[14] addressed this problem systematically. A sensor for CO was used where the current due to its oxidation at a Pt electrode provided the signal; a Nafion® SPE provided the current path across the cell. The sensitivity and the background current were functions of the humidity of the sample, in part because swelling as the humidity increases causes an increase in the conductivity of the SPE. Moreover, because the oxidation of CO is dependent upon water (Equation 7),

$$CO + H_2O \longrightarrow CO_2 + 2H^+ + 2e^- \qquad (7)$$

they surmised that the rate of oxidation of CO increases with water content. However, other factors such as the rate of diffusion of H^+ away from the indicator electrode and the diffusion of CO within the Nafion® can be important.

Solids prepared by sol-gel chemistry are alternatives to Nafion® as electrolytes for gas sensors. One attractive feature is that they are suited as hosts for a variety of chemical species such as enzymes and complexing agents, whereas Nafion® and other SPEs host materials primarily by ion-exchange reactions. The products of sol–gel processes are polymers with repeating oxygen-bridged metal systems such as silica, vanadia, alumina, and titania. A typical process is the hydrolysis of tetraethyl orthosilicate (TEOS) to a colloidal sol followed by condensation into a loosely crosslinked gel. With aging, during which water and ethanol are expelled, a glass-like solid termed a xerogel is formed. Unless the material is fused, the xerogel is porous and contains residual water both in the pores and

bound to the backbone. An important characteristic is that during the gelation and drying steps, species larger than the pore diameters, which range from a few Angstroms (1 Å = 0.1 nm) to the nanometer scale (depending on the preparation procedure), are encapsulated by a self-templated mechanism; hence, the activities of substances such as enzymes are the same in xerogels as in homogeneous aqueous solutions.[15] Detailed descriptions of sol–gel processes are widely available.[15–17]

With CO as the analyte, an amperometric sensor with a silica xerogel as the electrolyte provided a sensitivity that was independent of humidity over the range 9–76%,[18] which is in marked contrast to the above-described system with Nafion® as the SPE. The sensitivity was unchanged even though the water content of the xerogel ranged from 20 to 50%. Apparently, the water content was sufficiently high to yield a pseudo-first-order rate of oxidation. It is interesting that the diffusion of CO through the xerogel was also constant over this range. In this regard, a three-phase boundary was not present, so the CO diffused through a 5-μm layer of the xerogel to reach the indicator electrode. A similar result was observed with a silica-based amperometric sensor for gaseous hydrogen peroxide.[19]

Materials science also influences the design of electrocatalytic systems. Conducting composites that are doped with catalysts are emerging as important electrode materials. Sampath and Lev[20] described a composite electrode made from a combination of carbon powder, Pd, and an organically modified silicate. The latter was a product of sol–gel chemistry, but with methyltrimethoxysilane as the precursor. Various enzymes such as glucose oxidase and lactate oxidase were encapsulated in the composite. In contact with liquid samples, T_{90}-values of 15 s were obtained, and stable responses for up to 120 days were observed. Although not used for gas sensing, extension to samples such as alcohols (using an appropriate enzyme and SPE) can be envisioned. Other new materials as electrocatalytic indicator electrodes include metal-doped glassy carbon; conducting composites containing platinum group metal oxides, metal phthalocyanines, or porphyrins; and xerogels or conducting organic polymers with dispersed platinum group metal centers. Such electrodes are described in a review.[21]

3 SCOPE

In this section, the areas of application of electrochemical gas sensors are described. Emphasis is placed on describing the most important analytical performance characteristics for each of the sample types.

3.1 Environmental Monitoring

Perhaps the most challenging area of field applications of electrochemical gas sensors is environmental monitoring. Here, trace level determination is the dominant theme

Table 1 Selected electrochemical sensors for gas-phase analytes in the environment

Analyte	Class[a]	Limit of detection[b]	Threshold limit values[c]	Interferences[d]	Ref.
SO_2	A	7 μL m^{-3}	2 ppm	H_2O, H_2S	25
NO, NO_2	A	50 μL m^{-3}	25 ppm, 3 ppm	NRI	25
Halogenated hydrocarbons	C	≈100 ppb	50 ppm (CH_3Cl)	NRI	26
Cl_2	A	0.75 ppbv	0.5 ppm	SO_2, H_2S	27
HCl	P	2.1 ppbv	5 ppm C	SO_2, H_2S	27
O_3	A	0.5 ppbv	0.1 ppm C	NO_2	28
NO_2	A	1 ppb	3 ppm	O_3	28
O_3	A	≈ppb	0.1 ppm C	NI	29
O_3	A	224 ppbv	0.1 ppm C	Cl_2, NO_2	30
NO_2	A	2 ppbv	3 ppm	SO_2	31
NO_2	A	0.2 ppm	3 ppm	NI	32
NO_2	P	30.3 ppthv	3 ppm	NRI	33
NO_2	A	0.01 ppmv	3 ppm	NI	34
NO	A	0.2 ppmv	25 ppm	NI	34
NO_2	A	NR, <1000 ppm	3 ppm	NI	35
NO	A	NR, <1000 ppm	25 ppm	NI	35

[a] A = Amperometric; P = Potentiometric; C = Conductivity.
[b] NR = DL not reported; limit of detection estimated from text of references 26, 29, 33, and 35.
[c] C = Ceiling limit; no TLV (threshold limit value) reported.
[d] NRI = no significant interferences reported, but investigated; NI = interferences not investigated.

in that the analytes that are monitored typically are hazardous. Exacerbating the challenge of environmental analysis is the restriction on the type of methodology that can be employed. While mobile laboratories and grab sampling of stable species allow the use of sophisticated measurement technology, there is a need for the further development of remote sensors. For example, long-path infrared spectroscopy has been used for on-site measurement of the ratio of CO to CO_2 emission from vehicles on roadways,[22,23] but using this technology to continuously monitor air quality at multiple sites has prohibitive costs. Electrochemical sensing is relatively inexpensive, particularly if coupled with micromachining technology by which lithography can be used for large-scale production of platforms that contain the instrumentation and data transmission electronics.[24]

DL and long-term stability are the most important of the analytical performance criteria for remote application of environmental sensors. The former needs to be related to values such as the Occupational Health and Safety Standards. Table 1 is a summary of reported DLs and interferents for selected electrochemical gas sensors that are under development, along with values for permissible exposure limits.

The need for remote sensing of environmental samples is most compelling for analytes that lack the stability for grab sampling with subsequent transport to an analytical laboratory. Ozone, which is among the strongest oxidizing agents, provides an example. Ozone is of environmental significance in part because it causes plant damage as well as degradation of plastics and rubbers. Moreover, it is considered a cause of respiratory disorders.[36]

Electrochemical determinations of ozone typically are based on its reduction at a gold electrode (Equation 8).

$$O_3 + 2e^- + 2H^+ \longrightarrow O_2 + H_2O \qquad (8)$$

At a pH less than one, the onset of reduction occurs at approximately 1.0 V vs saturated calomel electrode (SCE). A direct adaptation of this reaction to amperometric sensing was reported by Schiavon et al.[30] Here, the interface between the indicator electrode and the gas sample was a Nafion® ion-exchange membrane that was wetted by contact to a 0.1 M $HClO_4$ reservoir (Figure 2). The sensor was operated at 0.5 V vs SCE. Chlorine and oxides of nitrogen interfere in that they are reduced at this potential. The system provided a LDR of four orders-of-magnitude, a DL of 224 ppbv, and a stability of 2 months.

The DL of the above sensor is above the ceiling level of 0.1 ppmv for O_3, but for many applications detection at the threshold level is required. One route to improving the DL, albeit at the expense of simplicity of instrumentation, without changing the signal-generating

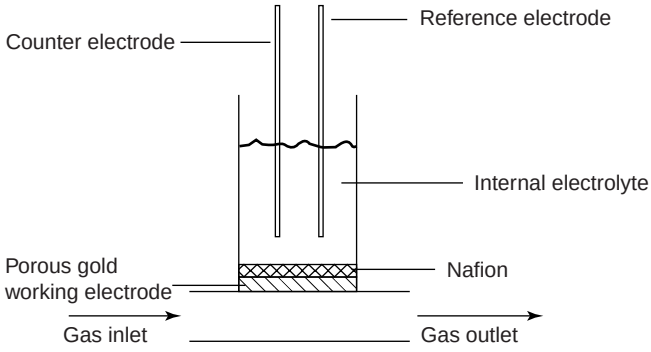

Figure 2 Amperometric sensor based on a SPE. The redox site is the three-phase boundary of the sample gas, electrolyte, and indicator electrodes.

chemistry was to employ pulse voltammetry at a gold electrode.[29] Here, the sensor design was similar to that in Figure 1; a gas permeable membrane was placed over a lens paper that was used to immobilize an electrolyte film to which the electrodes were contacted. The ozone was sampled from the air into a solution that contacted the sensor via a frit, so a reservoir to keep the lens paper wetted was not necessary. Rather than apply the reduction potential continuously, a value where no reaction occurs was applied to allow O_3 to accumulate for 5 s. Next, the electrolysis potential was applied, and the current was measured after 100 ms. After several cycles, a steady-state current was obtained. The T_{90} value was 7 s with a 1-mil dimethylsilicone gas permeable membrane. This mode resulted in a 50-fold improvement in the DL relative to constant potential application. The strategy of employing a preconcentration step does not address the problem of selectivity. Any component of the sample that is electroactive at the measurement potential (in the range 0.5 to 0.7 V vs SCE in these studies) will interfere.

The above reports illustrate the challenges presented by field studies of environmental samples. Section 4 includes descriptions of other studies aimed at development of ozone sensors and of sensors for other analytes of interest in environmental science.

3.2 Biomedical Applications

Most biomedical analysis is performed in a dedicated laboratory; an important exception is monitoring of glucose in blood. Moreover, analysis of gaseous samples is less common than studies on liquid samples. Field determinations of gaseous biomedical samples primarily involve emergency or law enforcement measurements of analytes related to respiration, including the indirect determination of ethanol in blood. In contrast to environmental monitoring, trace analysis and unattended measurements by remote sensors are not very important; instead, linear response, reproducibility, response and

recovery times, and freedom from interference are high priorities in terms of analytical performance criteria.

A representative example of a sensor applicable to respiration studies is the report by Tierney and Kim[37] of amperometric sensors for the determinations of NO and O_2 and a potentiometric sensor for the determination of CO_2. Of importance in this study is the employment of micromachining technology in devising sensors with fast response times. Specifically, porous platforms are used in conjunction with porous electrodes to enhance the formation of a three-phase boundary, which allows mass transport to occur in air rather than water. A deleterious effect of this facile mass transport is that the electrode reaction rather than mass transport is current-limiting, so the calibration curve is not linear. Also, the redox of some analytes yields products that poison the electrode surface and/or the SPE; in these cases, diffusional barriers sometimes are employed to lower the rate of electrolysis, which, in turn, increases the long-term stability. With O_2 as the analyte, the T_{90} is 300 ms and a calibration curve, albeit nonlinear, is obtained over the range 10–100% O_2. Interference of such species as N_2O and halothane, which are the two anesthetics that have known electrochemical activity, is not observed at the operating potential, −0.55 V vs Ag/AgO. Nitric oxide was determined in an analogous manner[37] except that the sensing reaction was given by Equation (9):

$$NO + 2H_2O \longrightarrow NO_3^- + 4H^+ + 3e^- \qquad (9)$$

A calibration curve was obtained over the range 20–100 ppmv, and the DL, which was calculated on the basis of $k = 2$, was 5 ppmv. The T_{90} was 330 ms, and the time to return to the baseline upon flushing was 550 ms.

For the potentiometric measurement of CO_2, Tierney and Kim[37] modified the Severinghaus-type sensor, which utilizes pH measurement in an electrolyte film separated by an air-gap from the sample. Iridium oxide was used for the pH measurement, and the micromachined porous system was used to enhance mass transport. Moreover, carbonic anhydrase enzyme was included in a commercially available hydrogel to reduce response times for the potentiometric sensor; in this regard, the use of the enzyme circumvents the slow hydrolysis of CO_2 that otherwise limits the response time. A calibration curve yielded a linear correlation between potential and $\log(\%CO_2)$ over the range 1–10% CO_2. The T_{90} was 1690 ms, and the system was stable for up to 8 h.

An amperometric sensor for ethanol in breath was developed on the basis of amperometry in a cell with a SPE.[38] The indicator electrode was porous gold that was supported on Nafion®. The gold was exposed directly to the gas sample. A solution of 1 M NaOH was in contact with the backside of the Nafion®. This allowed pre-oxidation of gold to a catalytically active oxide at potentials below that required for the production of the analytical signal. The oxidation of ethanol yielded a product that fouls the electrode surface. Hence, pulsed electrochemical detection (PED) was used to obtain the optimum performance. In PED, the electrode potential was pulsed in a sequence that provided for the following:

1. stripping the gold oxide and accumulated products by reduction;
2. forming the catalytic gold oxide; and
3. causing the oxidation of ethanol.

The 95% response time was estimated as 1 s, and the signal returned to the baseline in 20 s when the system was flushed with nitrogen. The DL of the sensor, 1 ppmv, is sufficient for the biomedical or the forensic determination of ethanol in breath. As described, the apparatus is not amenable to field measurements, but if combined with a micromachined platform and the use of micro-electronics to apply, measure, and transmit electrical signals, a portable, low-cost system can be projected.

A summary that delineates the scope of electrochemical sensors for gaseous samples of interest in biomedical

Table 2 Selected electrochemical sensors for gases of importance in biomedical analysis

Analyte	Class[a]	Limit of detection[b]	Threshold limit values	Interferences	Ref.
O_2	A	NR	–	NRI	37
NO	A	≈5 ppm	25 ppm	NI	37
CO_2	P	NR	5000 ppm	NI	37
NO	A	80 μM	25 ppm	NRI	39
NO	A	10^{-7} M	25 ppm	NRI	40
CH_3COO^-	P	0.6 μg mL^{-1}	–	NRI	41
O_2	P	7 mmHg	–	H^+, ionic strength	42
CO	A	5 ppmv	25 ppm	NI	18
C_2H_5OH	A	1 ppmv	1000 ppm	NRI	38

[a] A = Amperometric; P = Potentiometric; C = Conductivity.
[b] NR; limit of detection estimated from text of references 37, 39, 40, and 42.

Table 3 Selected electrochemical sensors for gases in industrial settings

Analyte	Class[a]	Limit of detection[b]	Threshold limit values[c]	Interferences	Ref.
C_2H_4	A	40 ppb	SA	NO, NO_2, SO_2	44
C_2H_2	A	20 ppb	SA	NRI	45
H_2S	A	NR	10 ppm	NI	46
H_2S	A	≈2 ppmv	10 ppm	NI	43
CH_3SH	A	≈3 ppmv	0.5 ppm	NI	43
CH_3SSCH_3	A	≈2 ppmv	0.6 ppm	NI	43
C_6H_5OH	A	≈1 ppm	5 ppm	Other phenolic compounds	47
H_2O_2	A	10.2 ppbv	1 ppm	NI	19
PH_3/AsH_3	A	0.01 ppm	0.3, 0.05 ppm	NRI	48
$C_{12}H_{27}O_4P$	P	1 ppb	0.2 ppm	NI	49
H_2	C	≈0.001%	SA	O_2	50
CO_2	P	100 ppm	5000 ppm	O_2	51
N_2H_4	C	1 ppb	0.01 ppm	NRI	52
NH_3	A	2 μmol mL^{-1}	25 ppm	NI	53

[a] A = Amperometric; P = Potentiometric; C = Conductivity.
[b] NR; limit of detection estimated from text of references 43, 47, 49–53.
[c] SA = simple asphyxiant.

analysis in the field is in Table 2. Detailed descriptions of selected systems are in section 4.

3.3 Sensors in Process Monitoring and the Work Place

Owing to the enormous number of industrial settings and potential analytes in the workplace, a simple set of desired characteristics for these sensors cannot be defined. Applications range from monitoring a reactant or product on the order of percent composition to the determination of a toxic by-product on the order of ppb. Also, the operating conditions such as temperature, relative humidity, and pH vary greatly. The scope of applications and approaches being used for development of sensors for such uses is illustrated below and in Table 3.

Sulfur-containing compounds, a class that includes some of the most toxic compounds known, are common by-products of industrial processes. For example, hydrogen sulfide gas has a ceiling limit of 15 ppmv (American Conference of Governmental Industrial Hygienists, 1996); the required DL for H_2S is well below this value. An amperometric sensor for H_2S was developed using a carbon indicator electrode that was doped with the electro-oxidation catalyst, cobalt phthalocyanine and using a sodium phosphate–agarose hydrogel as the electrolyte.[43] With the indicator electrode at 0.4 V vs Ag/AgCl, a linear amperometric response to H_2S over the range 2–18 ppmv was obtained. Extension to other sulfur compounds and the details of the mechanism are in section 4.1.

Hydrogen peroxide has a TLV of 1 ppmv. A variety of electrochemical sensors have been developed for the detection of trace levels of gaseous hydrogen peroxide. One example is shown in Figure 3. A diffusion scrubber

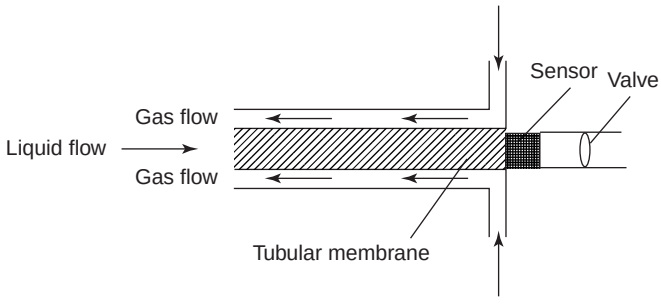

Figure 3 Example of a diffusion scrubber. A valve that controls the flow of the buffer allows preconcentration of components of the flowing gas into the buffer by adjusting the gas–liquid contact time.

is employed to introduce the hydrogen peroxide into a bulk electrolyte solution.[54] In this design, the analyte gas and the recipient buffer are flowing in a countercurrent fashion. The DLs obtained with this sampler depend on solution volume, flow rate, and preconcentration time. With a 9-min enrichment, the DL was below 1 ppbv.[55]

A more compact design of an amperometric sensor for H_2O_2 is based on the diffusion of hydrogen peroxide gas into a silica solid electrolyte made by base-catalyzed sol–gel chemistry.[19] A 10-μm film of silica coated coplanar disks of Pt, which served as the indicator and auxiliary electrodes, and Ag/AgCl, which was the reference electrode. The wires were sealed by epoxy in a 3.5-mm housing. The silica contained $MgCl_2$, which acted as an internal humidistat and provided ionic conductivity. The samples were the gas phase above solutions of various concentrations of H_2O_2 in water. The concentrations were calculated from thermodynamic data. The signal was obtained by applying a potential pulse after a prescribed

preconcentration time. The DL was 10 ppbv, and the LDR was 96–5400 ppbv.

Species such as H^+, O_2, and H_2 that influence the rate of corrosion or are concerns regarding the potential for explosions are important in certain industries. Monitoring is complicated if it must be performed at high temperature. A high-temperature potentiometric sensor for H_2 was developed that employed a solid electrolyte of 9% Y_2O_3 in zirconia and a fluorocarbon-coated Pt or Pd indicator electrode.[50] A linear correlation between potential and log(%H_2) was shown from 0.001 to 100% H_2 at 300 °C. In the presence of O_2 above 0.002%, the potential of the sensor response increased linearly with log(%O_2). When ion-beam sputtered electrodes were employed with a thickness of 150 nm, a response time on the order of 200 s was reported.

Sensing O_2 at high temperature also is of concern in monitoring combustion. An O_2 sensor based on conductivity changes in $SrTiO_3$ at 1000 °C was developed to monitor the exhaust of internal combustion engines.[56] The response time was 5 ms whereas with common ZrO_2-based systems it is 100 ms.

A potentiometric sensor capable of monitoring CO_2 at high temperatures was based on a solid electrolyte of Li_2CO_3–Li_3PO_4–$LiAlO_2$ with $LiCoO_2$–Co_3O_4 serving as a reference electrode.[51] Various CO_2 concentrations were introduced at a rate of 100 $cm^3\,min^{-1}$. A steady-state response in one minute was reported at 400 °C. Although the DL was not reported, a linear response with respect to the logarithm of CO_2 concentration was shown for 10^2–10^4 ppmv. The sensor was stable for one month. Oxygen was a significant interferent; a linear dependence of potential on the logarithm of the O_2 level was demonstrated.

Monitoring ethylene is important in several industries, including agriculture where it is important in ripening of produce. An amperometric sensor for ethylene was developed through the chemical deposition of gold onto a Nafion® SPE.[44] The gold electrode faced the sample gas; a solution of 0.5 M H_2SO_4 was an internal electrolyte in contact with the opposite side of the Nafion® membrane. Relative responses reported for acetylene, NO, NO_2, and SO_2 were 7.4, 1.8, 0.3, and 2.5, respectively, but these species are not expected to be present at significant concentrations in agricultural facilities. No responses were reported for ethanol, acetaldehyde, and carbon monoxide. A similar sensor was developed for the detection of acetylene in industrial settings.[45] Rather than using a constant potential detection as was done for ethylene, PED was performed to minimize the poisoning of the electrode surface by the oxidation products. No significant interferents were reported using PED for NO, NO_2, SO_2, and ethylene.

The above citations suggest that for measurements at room temperature the common strategy is to employ amperometric sensors with a SPE, which often is employed in conjunction with a porous electrode to develop a three-phase boundary. At elevated temperature, potentiometry and conductivity associated with inorganic electrolytes such as metal oxides are the typical bases of sensing.

3.4 Other Applications

The detection of analytes for purposes other than those previously categorized includes forensics. One of the largest applications for gas-phase sensors in forensics is for the determination of alcohol in breath. Electrochemical sensors for ethanol often employ enzymes as catalysts. An example of such a sensor is one developed for the determination of primary alcohols consisting of carbon chain lengths from two to six.[57] In this case, alcohol dehydrogenase (ADH) that is immobilized on an acetylcellulose membrane is the catalyst. The catalyst is separated from the sample solution by a gas-permeable membrane (Figure 4). This membrane serves both to prevent the enzyme from leaching into the sample solution and to provide an air-gap between the sample and electrode. The latter improves the selectivity because only volatile substances reach the electrode surface. The liquid layer between the membrane and the electrodes contains an electron-mediator. The mechanism for the generation of the analytical signal is given by Equation (10), which is enzyme-catalyzed and Equation (11), which occurs at the indicator electrode.

$$C_2H_5OH + 2Fe(III) \longrightarrow CH_3CHO + 2Fe(II) + H_2 \qquad (10)$$

$$Fe(II) \longrightarrow Fe(III) + e^- \qquad (11)$$

The current produced by Equation (11) was directly proportional to the concentration of C_2H_5OH over a wide range of conditions. Among the factors in obtaining

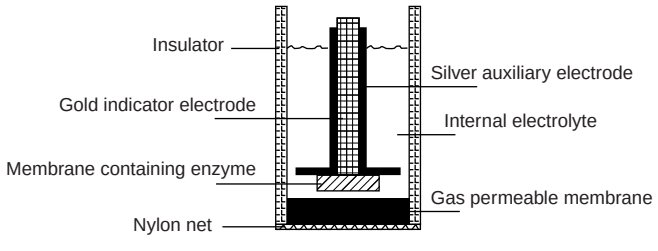

Figure 4 Amperometric gas sensor employing an enzymatic reaction. The enzyme, which is immobilized in a membrane, converts the analyte of interest to a suitable product for reaction at the electrode surface.

a wide LDR and low DL was the choice of the gas-permeable membrane. A comparison was made between polytetrafluoroethylene (PTFE) and a silicone rubber membrane. With the PTFE membrane, the upper limit of the linear range was 0.08 mM, whereas with the silicone rubber membrane, a linear range up to 1.8 mM was reported. The selectivity of this sensor was good. No response to 0.1 mM solutions of methanol, acetic acid, ascorbic acid, and glucose was observed, but a 0.1 mM acetaldehyde solution gave a 2% response relative to ethanol. It was demonstrated that the sensor functioned continuously for more than two weeks; however, after nine days, the sensitivity dropped to 85% of its original response. Although the data were reported for the determination of ethanol vapors above a solution, extension to its measurement in breath is straightforward.

An electrochemical sensor that was applied to the measurement of ethanol in breath has been reported.[38] This amperometric system utilized a SPE-supported porous gold indicator electrode, which faced the sample. The measurement was by PED. The response (95% of steady-state) and recovery times were 1 and 20 s, respectively. The DL, 1 ppmv, is more than a factor of a hundred lower than the level needed to measure ethanol in breath.

Electrochemical sensors for gases and vapors in forensic science for analytes other than ethanol are not common at present. Yet, numerous other applications can be envisaged. The measurement of any analyte that can be converted to an electroactive compound with a high vapor pressure can be accomplished. For example, many enzymatic reactions yield carbon dioxide or ammonia which subsequently can be determined by electrochemical sensors.

4 SELECTED APPLICATIONS OF ELECTROCHEMICAL GAS SENSORS

In the previous section, reports that illustrated the type of field applications where electrochemical gas sensors are employed were emphasized. Here, examples of more fundamental studies that can be applied directly to monitoring in the field are provided. Consistent with this objective, the present section is organized according to the chemistry of the process that provides the signal.

4.1 Amperometric Sensors

Measuring the current at constant potential as a function of concentration is the dominant strategy for electrochemical sensing at other than elevated temperature. This approach benefits from the high signal-to-background ratio that is developed because the background generally is very low. The basic sensor design is very simple (see Figures 1 and 2). The electrolyte that bridges the three electrodes is usually a solid such as an ion-exchange polymer or a material prepared by sol–gel chemistry. A limitation of amperometric sensors is that the vast majority of chemical species that are thermodynamically predicted to undergo redox at an electrode in an accessible potential range are electrochemically silent because of unfavorable electron-transfer kinetics. Hence, investigations on electrocatalysis often are required prior to the development of an amperometric sensor for a given analyte. A related approach is to use enzymatic reactions to convert the analyte to a substance that is electrochemically active, a strategy that also provides a degree of selectivity to the resulting sensor.

Enzymatic reactions prior to amperometric detection have been reported for a variety of compounds. An important example is an amperometric sensor for phenol vapors. In one design, the diffusion of molecular oxygen as well as phenol vapors across a membrane separating the sample gas from the electrode system was used.[47] The biosensor consisted of two layers: the barrier membrane composed of a carbon material modified with PTFE and a Nafion® membrane that contained the enzyme, tyrosinase. The design allowed the enzyme layer to contact an electrolyte solution. The enzymatic reaction is shown in Equations (12–14):

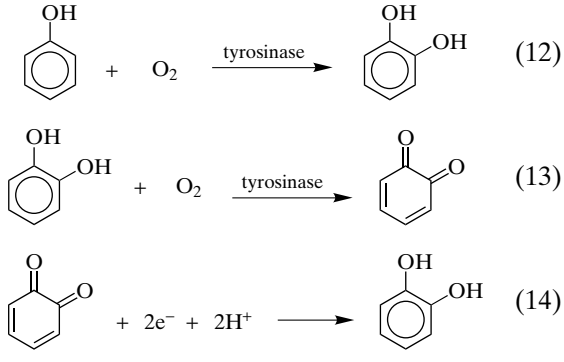

The analytical signal for the sensor is produced from the electrochemical reduction of the *o*-quinone. The activity was retained for more than 20 days. The standards were prepared by headspace methods. That is, solutions of known concentration of phenol were prepared and sealed. After equilibration between the liquid and gas phases in the vials, the vapor was sampled by puncturing septa with a syringe. A DL of approximately 1 ppmv was estimated from the calibration curve. The sensor suffered from a response time on the order of several minutes.

A more general approach to electrocatalysis in amperometric sensing is to use electron-transfer mediation. Here, a redox couple that undergoes facile electron-transfer at

an electrode is selected as a catalyst Equations (15) and (16). Generally, it is immobilized on the electrode surface, but it also can be immobilized in the SPE in contact with the indicator electrode. Assume the mediator couple is

$$M + e^- \rightleftharpoons N \quad (15)$$

and that the analyte, A, is not electroactive even though thermodynamically

$$B + e^- \rightleftharpoons A \quad (16)$$

If the E^0 values for Equations (15) and (16) predict that

$$M + A \longrightarrow N + B \quad (17)$$

the electrochemical oxidation of A can be achieved by immobilizing the mediator couple (M, N) at the electrode surface and controlling the electrode potential so that M is the primary species. Now, introducing A to the sensor causes the momentary conversion of some of the M to N. At that electrode potential N is immediately re-oxidized to M, yielding a current. Clearly, the oxidation current is related to the concentration of A. This indirect oxidation of A follows the same quantitative relationship of current to concentration as the direct electrolysis.

An example of a mediated oxidation is the use of cobalt phthalocyanine as the mediator in a sensor for the determination of sulfur-containing compounds as described in section 3.3.[43] The proposed reaction when H_2S is partitioned from a gaseous sample into a basic electrolyte is shown in Equations (18) and (19):

$$2HS^- + 2Co^{2+} \rightleftharpoons 2Co^+ + H_2S_2 \quad (18)$$

$$Co^+ \rightleftharpoons Co^{2+} + e^- \quad (19)$$

The mechanism of methanethiol detection was the same with the exception that it was obtained at a higher pH. The general mechanism of mediation was explained for the oxidation of an analyte; however, the analogous reactions serve to promote reduction. For example, the detection of dimethyldisulfide was based on Equations (20) and (21):

$$2H^+ + CH_3SSCH_3 + 2Co^+ \rightleftharpoons 2Co^{2+} + 2CH_3SH \quad (20)$$

$$Co^{2+} + e^- \rightleftharpoons Co^+ \quad (21)$$

where the reduction of Co^{2+} established the surrogate current for the reduction of CH_3SSCH_3.

Immobilized Fe(III,II) centers, which were in the form of phenanthroline complexes served as electron mediators for the oxidation of NO_2.[31] The interesting aspect of the sensor design involved immobilizing the centers by electropolymerization of a film of the mediator. The reaction primarily responsible for the current is probably given by Equation (22):

$$Fe(III) + N(IV) \longrightarrow Fe(II) + N(V) \quad (22)$$

although the authors did not list this as the operative mechanism. Detection of NO_2 down to 2 ppbv was reported and the calibration curve was linear up to 25 ppbv. Provided that the electrode was kept in contact with the electrolyte, the performance was constant for one month. Interferents listed were Cl_2 and SO_2. Of importance is that N_2O did not provide a signal at the operating potential, and NO only gave a signal when it constituted nearly 100% of the sample gas.

The simultaneous determination of NO and NO_2 was achieved with a commercial design (Back Cell®) in which gold deposited on porous alumina comprised the indicator and auxiliary electrodes.[35] The direct oxidation of these species generated the current. At 0.3 V vs a Pt quasi-reference electrode, the sum of these two species was determined. At −0.3 V, NO_2 alone was measured. By subtraction, concentrations of NO and NO_2 up to 7000 ppmv and 9000 ppmv, respectively, were determined, but DLs were not reported. The T_{90} was 1.51 s, and the baseline recovery time (to 90% of the original value) was 1.27 s.

An amperometric sensor for simultaneous measurement of NO and NO_2 that did not rely on subtraction was one based on protonic conduction through α-zirconium acid phosphate.[34] The pellicular zirconium phosphate served as a solid-state proton conductor. Gold or graphite was the indicator electrode. Current–potential curves indicated that the reduction of NO_2 and oxidation of NO Equations (23) and (24) occurred at different potentials.

$$NO_2 + H^+ + e^- \longrightarrow HNO_2 \quad (23)$$

$$NO + H_2O \longrightarrow HNO_2 + H^+ + e^- \quad (24)$$

Graphite provided the superior sensitivity with NO_2 as the analyte. A DL of 0.01 ppmv and a linear range up to 15 ppmv were reported for measurements performed at constant potential. A DL of 0.2 ppmv and an upper linear range of 15 ppmv were reported for the determination of NO at a gold electrode, but in this case, a pulsed potential was employed to help discriminate the NO signal from that produced by NO_2.

For biomedical applications, a more rapid response to NO is important in that it has a half-life in the presence of O_2 of 5–6 s. The same reactivity, when coupled with the envisaged future need for field measurements of NO, presents a compelling case for further improvement of sensors for this gas. A promising approach is the use of a bilayer of Nafion® and cellulose acetate as a barrier between the sample and the electrode surface.[39]

The hydrolyzed cellulose acetate layer, which is in direct contact with the sample, serves as a size-selective barrier, whereas Nafion®, which is an anionic polymer, acts as an anion exclusion membrane. A steady-state response to NO was achieved in 20 s. Perhaps the substitution of Nafion® with a porous membrane will improve the response time to a practical value for determinations of NO in biomedical samples.

Phosphine (PH_3) and arsine (AsH_3) are extremely toxic gases produced by several industrial processes. An amperometric sensor for these species was developed using gold that was plated onto a PTFE membrane as the indicator electrode.[48] An inner solution of sulfonic acid was the electrolyte; it contained the reference and auxiliary electrodes. The sensor was selective over ethanol, methanol, isopropyl alcohol, acetone, and toluene. The gas phase directly contacted the gold electrode. The DLs were 0.01 ppmv, and the LDRs extended to 1.1 ppmv for arsine and phosphine. The sensor also had a long lifetime; after 6 months of continuously monitoring air, the sensor still maintained more than 80% of its original response.

Electrochemical sensing of hydrogen peroxide can be based on either its reduction or its oxidation. For studies of gases, the oxidative approach is preferred because O_2 reduction generally overlaps that of hydrogen peroxide. An amperometric sensor was developed that employed a silica xerogel prepared by base-catalyzed sol–gel chemistry, which ensured that the pH of the pore liquid in the gel promoted formation of HO_2^-, which is the electroactive species.[19] A thin film of silica was cast on a three-electrode system in which platinum served as the indicator and auxiliary electrodes and silver served as the reference electrode. A LDR spanning three orders of magnitude and a DL of 10.2 ppbv were reported. The hydrogen peroxide vapors were introduced to the sensor via the headspace method (described above). The sensor also was applied as a gas-phase monitor of H_2O_2 released from the reaction of glucose and glucose oxidase enzyme (GOx) in O_2-saturated solution.

Amperometric sensing of ozone is related to that of hydrogen peroxide. As discussed in section 3.1, electrochemical O_3 sensors typically employ oxidation at a gold electrode to produce the signal. An alternative design employed a silver iodide disk as a solid electrolyte pressed between platinum gauze and silver electrodes.[28] In this case, the response is from a galvanic reaction, not electrolysis at an applied potential. The sensor design is shown in Figure 5. The suggested mechanism for the response is shown in Equations (25–27):

$$2AgI + O_3 \longrightarrow Ag_2O + I_2 + O_2 \qquad (25)$$

$$I_2 + 2e^- \longrightarrow 2I^- \qquad (26)$$

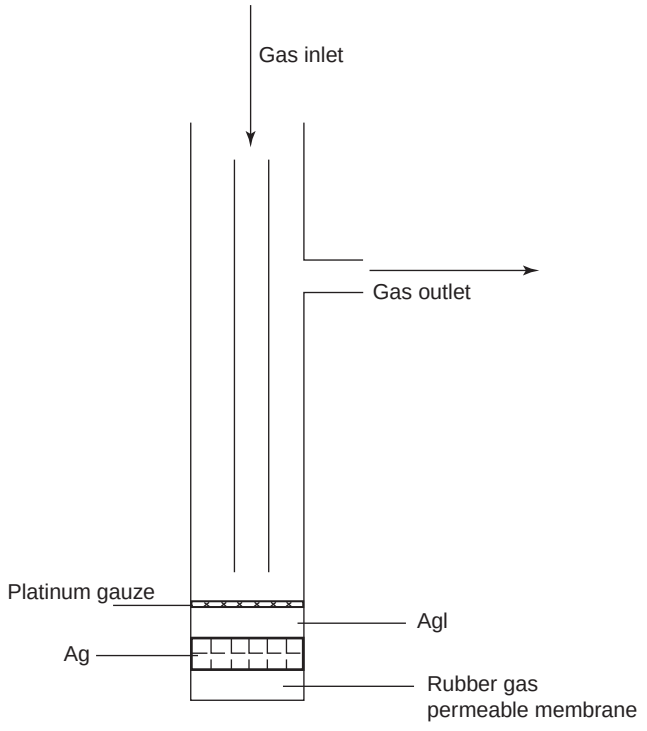

Figure 5 Potentiometric sensor for the determination of ozone (adapted from reference 28). Silver iodide serves as the solid electrolyte in this sensor design.

$$Ag \longrightarrow Ag^+ + e^- \qquad (27)$$

A DL of 1 ppbv was reported, and a linear response was obtained for ozone levels up to 0.5 ppmv. The sensor responds to NO_2 as well as O_3, so they are mutual interferents.

A galvanic cell for the determination of NO_2 was reported using a nitrite sodalite electrolyte sandwiched between silver and gold electrodes.[33] The sensor had a linear range of 3.03×10^{-3} to 6.39×10^{-2} atm; the DL was not reported. An important feature of the sensor was that the response was independent of the oxygen content of the sample gas.

Galvanic sensors have limited appeal relative to amperometric systems based on a constant applied potential. As in any battery that is producing current, reagents at both electrodes must be replenished. But perhaps a greater problem is the DL. The anodic and cathodic currents must be the same, so the potential will vary to values that provide a reactive species. At low concentrations of the analyte, the reactive species will be an interferent.

4.2 Potentiometric Sensors

Potentiometry is an equilibrium method; no current passes through the cell during the measurement. The

potential difference between two electrodes constitutes the measured signal. To relate this signal to the concentration of an analyte, one of the half-cells is of a composition that provides an invariant potential (i.e. it is a reference electrode). In one measurement scheme, the other half-cell includes a redox electrode. For equilibrium to be reached, both the oxidized and the reduced form of the half-cell species need to be present. Moreover, for meaningful analytical data to be obtained, the concentrations in both oxidation states must be high enough to avoid a mixed-potential, which is a situation where more than one redox couple contributes to the potential.

Mixed potentials at redox electrodes are related to the relative contributions of exchange currents from various redox couples. For a simple reaction, $A + e^- \rightleftharpoons B$, the exchange current is related to the currents from the pair of reactions, $A + e^- \longrightarrow B$ and $B \longrightarrow A + e^-$, which are identical at equilibrium. The analogous limitation of a membrane-based ion selective electrode (ISE) is the contribution of an interferent, *j*, to the potential developed by the analyte, *i*, in the expression used to quantify the measurements Equation (28):

$$E = L + \left(\frac{S}{z}\right) \log \left(a_i + k_{ij} a_j^{z/x}\right) \tag{28}$$

where L is a constant; S, the slope which at 25 °C ideally is 59 mV; z, the charge of the analyte; a_i, the activity of the analyte; a_j, the activity of the interferent; x, the charge of the interferent; and k_{ij}, the selectivity coefficient. Generally, ISEs (a class that includes the glass pH electrode) provide lower DLs than redox electrodes. They are of limited scope in gas analysis. Here, the analyte must be partitioned from the gas phase into a liquid film that contacts the ISE. Among the analytes for which ISEs can be used are substances that alter the pH of the liquid film and halides.

The use of a pH electrode for the determination of acetic acid in the gas phase was reported.[41] The electrode was contacted to a sodium acetate solution that, in turn, was separated from the sample by a Teflon® membrane. The pH changed in response to incursion of acetic acid vapors. A limitation of this approach is that any component of the gaseous sample which shifts the acetate/acetic acid equilibrium will interfere.

Potentiometric measurements of chloride can be made with a silver wire coated with a layer of silver chloride. The potential of this Ag/AgCl electrode is dependent on the activity of chloride ions present. Gaseous HCl can be determined by its dissolution into a basic buffer, thereby forming chloride ion.[27]

A sensor for O_2 was developed on the basis of the mixed potential that was developed in conjunction with the corrosion of a cobalt wire.[42] The potential response was from the oxidation of cobalt and the reduction of oxygen on different sites of same electrode. For determining oxygen in air or in the headspace above liquids, sampling was through a gas permeable silicone rubber tubing. The response to the partial pressure of O_2 was evaluated using a set of four sensors. A linear plot of E vs $\log(P_0)$, where P_0 is the partial pressure of O_2, over the P_0 range 0.0097–0.94 provided a slope of 112 ± 4 mV (decade)$^{-1}$. Absolute drift rates of the sensors varied with time. For initial times of 0, 72, and 96 h, the drift rates were 2.1 ± 0.4, 0.78 ± 0.22, and 0.25 ± 0.11 mV h^{-1}, respectively. The drift limits the utility of this sensor for some field applications.

A potentiometric sensor for the determination of tributyl phosphate vapor (TBP) based on a specific charge-transfer interaction with iodine has been reported.[49] The electrode was prepared in a two-step process. The first step involved the electrochemical synthesis of a poly(cyclophosphazenebenzoquinone) (PPBQ) matrix. In the second step, I_2 was incorporated into the matrix. Electron-sharing between the I_2 and the oxygen on the P=O site of TBP was reported to perturb the electrical properties of PPBQ, which in the absence of TBP are related to the interaction of the I_2 with the PPBQ. The potentiometric response reflects a change in the work function, so the system behaves like a FET. A plot of E vs $\log(C_{TBP})$ is a straight line with a slope of 35.9 mV (decade)$^{-1}$. The linear range is from 0.27 to 8.8 ppbv TBP.

Overall, solid-state potentiometric devices that provide a response on the basis of work function changes represent an emerging area of electrochemical gas sensors. In many cases, the actual application of this chemistry is to use conductivity changes as the signal, as described in the next section.

4.3 Conductivity Sensors

Most sensor designs in this category monitor the conductance change due to a charge-transfer reaction or the interaction of an analyte with a semiconducting material. A complicating factor is that in certain electrolytes (e.g. conducting organic polymers), conductivity can reflect mobility of electrons, ions, or a combination of the two.

Hydrazine sensors are under investigation because it is a highly toxic material at trace levels, and at elevated levels it can promote explosions. In one strategy for its detection, a platform of an electronically conducting polymer, poly(3-hexylthiophene), was deposited onto a quartz substrate by spin coating.[52] The concentration of the gas was related to the increased resistance of the film. The sensor was calibrated to allow the determinations of both dosage of hydrazine (ppb-h) and instantaneous concentration. The dose was determined

by measuring the change in resistance over a specified time interval; essentially, the data were integrated. The concentration was computed from this value by dividing the dose by the time interval over which a measurement was taken. The resistance varied exponentially with dosages up to 10 ppb-h. This sensitivity allowed 0.1 ppb-h doses and 1-ppb concentrations of hydrazine to be measured. The reproducibility of the calibration curve for 100 sensors was ±20%. Real-time monitoring of a 69-ppb hydrazine stream was measured within 5% of the calibrated value for 10 min, and the response changed less than 5% over an 8-hour period in air. The sensor was shown to be selective through exposure to suspected interferents. These interferents included aniline, dimethylamine, ethylenediamine, and ammonia. The sensor was exposed for 8 h to 16, 80, 80, and 200 ppm-h, respectively, of these compounds; these values are the TLV for these compounds. These exposures yielded signals that corresponded to hydrazine-equivalent doses of 187, 58, 35, and 9 ppb-h, respectively.

An interesting approach to conductivity sensing was to employ Langmuir–Blodgett films of phthalocyanine deposited on a silicon chip containing a gold interdigitated electrode as a platform. By this method, a sensor for the determination of chlorine, bromine, and iodine was developed.[58] Included in the design were a micro-heater and a diode for control and monitoring of the sensing film temperature. The primary steps causing the conductivity change were the adsorption of the gas on the film surface and its subsequent diffusion into the film. Evidence for the importance of diffusion into the film was that thickening of the film caused an increase in the signal at the expense of response time. The analyte gas acted as a π-electron acceptor, thereby forming a redox couple. The positive charge that was produced was delocalized over the phthalocyanine ring, which caused an increase of the conductivity. Hence, this was a p-type semiconductor. The sensor had a linear response at room temperature from 20 to 100 ppmv Cl_2. In order to re-establish the baseline response, the temperature was elevated to a point that resulted in the chlorination of the phthalocyanine ring. It was observed that the maximum signal was attained at different temperatures for each of the analytes. The explanation was that at low temperatures the acceptor holes of doped semiconductors are connected to acceptor sites, which is termed ''freeze out''. As the temperature is increased, more holes are ionized; they enter the valence band as free holes. The coulombic field of the adsorbed species makes it more difficult for these temperature-induced holes to escape and produce an increase in conductivity. The higher the electron affinities of species from the gas phase, the higher temperature that is needed for maximum conductance. A possible limitation on the use of this sensor is the irreversible halogenation of the phthalocyanine ring that occurs. Not only will it limit the lifetime but also it establishes the upper limit of the operating temperature of the sensor.

An oxygen sensor employing $SrTiO_3$ as the solid electrolyte with a response time fast enough to monitor internal combustion engines has been reported.[56] The $SrTiO_3$ layer was sputtered onto an Al_2O_3 substrate that contained a heater and temperature sensor. The response to oxygen disappeared below 600 °C; moreover, control to ±25 °C was needed for appropriate accuracy.

One of the most promising approaches in sensing is to employ FETs that have coatings for which the work function changes in response to an analyte gas.[49,59,60] Formally, the sensing can be by potentiometry or by conductivity. These devices are considered in section 4.4.

4.4 Comparison to Related Devices

As mentioned above, FET-based sensors are emerging as important approaches to field measurements of chemical species in gases. One factor driving this field is the compatibility with micromachined assemblies, which allows not only miniaturization but also the deployment of multiple units on the same chip. The latter, in turn, allows the use of chemometric methods such as multivariant analysis to extract information on mixtures, thereby providing enhanced selectivity. An excellent summary of this general topic is present by Domanský et al.[59] Specifically, FET arrays were used to determine mixtures of H_2 and NH_3 in air using a combination of Pd and polyaniline layers to devise chemically sensitive FETs. The work function change in the presence of the analytes leads to an analytical response function similar to that in potentiometry; that is, the potential is measured vs the logarithm of the concentration of the analyte. Using a related device, but on analytes partitioned into a liquid, NH_3 and SO_2 were determined with a pH-sensitive insulated-gate FET.[60]

Surface acoustic wave (SAW) devices also are closely related to electrochemical sensors. A SAW device consists of two separate interdigital transducers and two interlocked electrodes on a piezoelectric medium. The first interdigitated transducer produces the SAW by disturbing the piezoelectric medium. The second interdigitated transducer converts the acoustic wave to an electric signal, commonly referred to as a delay line. A coating on the quartz crystal of the piezoelectric device can be deposited in an attempt to obtain selectivity for a given analyte. The adsorption of a species onto the surface or changes in temperature and/or pressure results in a change in the velocity of the wave. This change in velocity is measured as a change in frequency provided that the sensor is part of an oscillator circuit. The relative simplicity of the device is one of the greatest merits for

employing SAW devices as gas sensors. Because only the surface is used, the device can be printed on a silicon chip. Such production allows for low cost and ease of miniaturization.

One of the major limitations in SAW sensors is the selectivity of the deposited film. Any adsorbed species yields a response. A second SAW device can be employed that has an unmodified surface or one that has been passivated. Here, the difference in response constitutes the analytical signal.

An example of a possible application is a set of SAWs with coatings of tri-substituted triazine derivatives as sensors for chemical warfare agents.[61] A variety of lanthanum coordination compounds were tested for their response to the nerve agent sarin (GB) and the stimulant, dimethyl methylphosphonate (DMMP).[62] The mechanism for adsorption of the test vapors is that the La(III) that is bound to aminopolycarboxylic acid residues also binds to the oxygen atoms surrounding the phosphorus atom of the analyte vapors. The most effective complex was La(III)2-bis(carboxymethyl)aminohexadecanoic acid. A quadratic relationship was observed between the frequency shift and the concentration of the analyte from 1 to 10 ppmv. DLs of 0.1 ppmv were reported. The signal reached 80% of steady-state response in 4 s, and the baseline recovered in 13 s with DMMP as the analyte. The respective times for GB were 8 and 11 s.

Of these related systems, FETs perhaps are the more promising. Selectivity with SAW sensors is established only by adsorption whereas with FET-based devices, adsorption can operate in conjunction with selective chemical reactions such as decomposition.

5 SUMMARY

Electrochemical sensors for field measurement gases are attractive primarily because they have design features that are compatible with mass production by modern micromachining and microelectronics technology. By these methods, inexpensive platforms can be made that combine the necessary instrumentation to develop and transmit signals related to concentration. However, to date, application to chemical sensing has lagged behind physical measurement, e.g. pressure and temperature. As summarized in Arenas et al,[1] a number of electrochemical gas sensors are commercially available, but the list of analytes to which they are applied is limited. In contrast, research on electrochemical sensing has demonstrated that in addition to the small, gaseous species that are commonly determined (O_2, CO, H_2S, CO_2, and NO_x, for example), electrochemical gas sensors for analytes such as ethanol, various sulfur compounds, phenol, or tributyl phosphate have the response rates, sensitivities, and stabilities to merit widespread application. Devices based on FET designs are especially promising. Finally, an attribute of sensors based on solid electrolytes is that they can be used in harsh environments, including the monitoring of gases at high temperature.

ABBREVIATIONS AND ACRONYMS

ADH	Alcohol Dehydrogenase
DL	Detection Limit
DMMP	Dimethyl Methylphosphonate
FET	Field Effect Transistors
GB	Sarin
GOx	Glucose Oxidase Enzyme
ISE	Ion Selective Electrode
LDR	Linear Dynamic Range
NI	Interferences Not Investigated
NR	Detection Limit Not Reported
NRI	No Significant Interferences Reported, but Investigated
PED	Pulsed Electrochemical Detection
PPBQ	Poly(cyclophosphazenebenzoquinone)
PTFE	Polytetrafluoroethylene
SA	Simple Asphyxiant
SAW	Surface Acoustic Wave
SCE	Saturated Calomel Electrode
SD	Standard Deviation
SPE	Solid Polymer Electrolyte
TBP	Tributyl Phosphate Vapor
TEOS	Tetraethyl Orthosilicate
TLV	Threshold Limit Value

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Electrolytes, Blood Gases, and Blood pH

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment • Formaldehyde, Environmental Analysis of

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Organic Analysis in Environmental Samples by Electrochemical Methods • Phenols Analysis in Environmental Samples • Sampling Considerations for Biomonitoring

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

Industrial Hygiene **(Volume 5)**
Carcinogens, Monitoring of Indoor Air

Industrial Hygiene cont'd **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Ion-selective Electrodes: Fundamentals

REFERENCES

1. R.V. Arenas, K.R. Carney, E.B. Overton, 'Portable, Multigas Monitors for Air Quality Evaluation', *Am. Lab.*, **25**, 25–31 (1993).
2. Y. Fang, J. Leddy, 'Voltammetry in Gas Phase Environments', *J. Electroanal. Chem.*, **384**, 5–17 (1995).
3. A.J. Crocker, 'Metal-Oxide-Semiconductor Gas Sensors', in *Techniques and Mechanisms in Gas Sensing*, eds. P.T. Moseley, J.O.W. Norris, D.E. Williams, Adam Hilger, New York, Chapter 12, 1991.
4. L.C. Clark Jr., U.S. patent 2 913 386, 1959.
5. I. Fatt, *Polarographic Oxygen Sensors*, CRC Press, Boca Raton, 1976.
6. J.A. Cox, G.R. Litwinski, 'Voltammetric Ion Selective Electrode for the Determination of Nitrate', *Anal. Chem.*, **51**, 554–556 (1979).
7. G. Bontempelli, N. Comisso, R. Toniolo, G. Schiavon, 'Electroanalytical Sensors for Nonconducting Media Based on Electrodes Supported on Perfluorinated Ion-Exchange Membranes', *Electroanalysis*, **9**, 433–443 (1997).
8. T. Seiyama, N. Yamazoe, 'Recent Advances in Gas Sensors in Japan', in *Fundamentals and Applications of Chemical Sensors*, eds. D. Schuetzle, R. Hammerle, ACS Symposium Series 309, American Chemical Society, Washington, D.C., Chapter 2, 1986.
9. S. Chakraborty, H. Satou, H. Sakata, 'Direct Current Conductivity and Oxygen Gas-Sensing Properties of Iron-Antimony-Tellurite Glasses', *J. Appl. Phys.*, **82**, 5520–5525 (1997).
10. J.D. Ingle, S.R. Crouch, *Spectrochemical Analysis*, Prentice-Hall, Englewood Cliffs, 174, 1988.
11. P.R. Warburton, M.P. Pagano, R. Hoover, M. Logman, K. Crytzer, Y.J. Warburton, 'Amperometric Gas Sensor Response Times', *Anal. Chem.*, **70**, 998–1006 (1998).
12. S.J. Sonderheimer, N.J. Bunce, C.A. Fyfe, 'Structure and Chemistry of Nafion®-H: A Fluorinated Sulfonic Acid Polymer', *Rev. Macromol. Chem. Phys.*, **C26**, 353–413 (1986).
13. J.A. Cox, R. Gajek, G.R. Litwinski, J. Carnahan, W. Trochimczuk, 'Optimization of Ion Exchange Membrane Structures for Donnan Dialysis', *Anal. Chem.*, **54**, 1153–1157 (1982).
14. S.B. Lee, A. Cocco, D. Keyvani, G.J. Maclay, 'Humidity Dependence of Carbon Monoxide Oxidation Rate in a Nafion®-Based Electrochemical Cell', *J. Electrochem. Soc.*, **142**, 157–160 (1995).
15. B.C. Dave, B. Dunn, J.S. Valentine, J.I. Zink, 'Sol–Gel Encapsulation Methods for Biosensors', *Anal. Chem.*, **66**, 1120A–1127A (1994).
16. L.L. Hench, J.K. West, 'The Sol–Gel Process', *Chem. Rev.*, **90**, 33–72 (1990).
17. C.J. Brinker, G.W. Scherer, *Sol–Gel Science: The Physics and Chemistry of Sol–Gel Processing*, Academic Press, San Diego, 1990.
18. M.E. Tess, J.A. Cox, 'Humidity-Independent Solid-State Amperometric Sensor for Carbon Monoxide Based on an Electrolyte Prepared by Sol–Gel Chemistry', *Anal. Chem.*, **70**, 187–190 (1998).
19. S.D. Holmstrom, J.A. Cox, 'Solid-State Voltammetric Determination of Gaseous Hydrogen Peroxide Using Nanostructured Silica as the Electrolyte', *Electroanalysis*, **10**, 597–601 (1998).
20. S. Sampath, O. Lev, 'Inert Metal-Modified, Composite Ceramic-Carbon, Amperometric Biosensors: Renewable, Controlled Reactive Layer', *Anal. Chem.*, **68**, 2015–2021 (1996).
21. J.A. Cox, M.E. Tess, T.E. Cummings, 'Electroanalytical Methods Based on Modified Electrodes: A Review of Recent Advances', in *Reviews in Analytical Chemistry*, ed. M. Zangen, Freund Publishing House, London, England, Vol. XV(3), 173–223, 1996.
22. G.A. Bishop, J.R. Starkey, A. Ihlenfeldt, W.J. Williams, D.H. Stedman, 'IR Long-Path Photometry: A Remote Sensing Tool for Automobile Emissions', *Anal. Chem.*, **61**, 671A–677A (1989).

23. Y. Zhang, D.H. Stedman, G.A. Bishop, P.L. Guenther, S.P. Beaton, 'Worldwide On-Road Exhaust Emissions Study by Remote Sensing', *Environ. Sci. Technol.*, **29**, 2286–2294 (1995).
24. Q. Wu, K. Lee, C. Liu, 'Development of Chemical Sensors Using Microfabrication and Micromachining Techniques', *Sensors and Actuators B*, **13–14**, 1–6 (1993).
25. P. Bruno, M. Caselli, M. Della Monica, A. Di Fano, 'Simultaneous Determination of SO_2, NO and NO_2 in Air by Differential Pulse Polarography', *Talanta*, **26**, 1011–1014 (1979).
26. Z. Cao, J.R. Stetter, 'A Selective Solid-State Gas Sensor for Halogenated Hydrocarbons', *Sensors and Actuators B*, **5**, 109–113 (1991).
27. W. Matuszewski, M.E. Meyerhoff, 'Continuous Monitoring of Gas-Phase Species at Trace Levels with Electrochemical Detectors: Part 2 Detection of Chlorine and Hydrogen Chloride', *Anal. Chim. Acta*, **248**, 391–398 (1991).
28. S. Suzuki, K. Nagashima, 'A Galvanic Solid-State Sensor for Monitoring Ozone and Nitrogen Dioxide', *Anal. Chim. Acta*, **144**, 261–266 (1982).
29. R.B. Smart, R. Dormond-Herrera, K.H. Mancy, 'In Situ Voltammetric Membrane Ozone Electrode', *Anal. Chem.*, **51**, 2315–2319 (1979).
30. G. Schiavon, G. Zotti, G. Bontempelli, G. Farnia, G. Sandoná, 'Amperometric Monitoring of Ozone in Gaseous Media by Gold Electrodes Supported on Ion Exchange Membranes (Solid Polymer Electrolytes)', *Anal. Chem.*, **62**, 293–298 (1990).
31. M. Bonakdar, J. Yu, H.A. Mottola, 'Continuous-Flow Performance of Carbon Electrodes Modified with Immobilized Fe(II)/Fe(III) Centers', *Talanta*, **36**, 219–225 (1989).
32. F. Opekar, 'An Amperometric Solid-State Sensor for Nitrogen Dioxide Based on a Solid Polymer Electrolyte', *Electroanalysis*, **4**, 133–137 (1992).
33. M.R.M. Jiang, M.T. Weller, 'A Nitrite Sodalite-Based NO_2 Gas Sensor', *Sensors and Actuators B*, **30**, 3–6 (1996).
34. G. Alberti, F. Cherubini, R. Palombari, 'Amperometric Solid-State Sensor for NO and NO_2 Based on Protonic Conduction', *Sensors and Actuators B*, **37**, 131–134 (1996).
35. M.J. Tierney, H.L. Kim, M. Madou, T. Otagawa, 'Microelectrochemical Sensor for Nitrogen Oxides', *Sensors and Actuators B*, **13–14**, 408–411 (1993).
36. EPA, *Air Quality Criteria for Ozone and Related Photochemical Oxidants*, Dec. 1993, External Review Draft, USEPA 600/AP-93/004, 1993.
37. M.J. Tierney, H.L. Kim, 'Electrochemical Gas Sensor with Extremely Fast Response Times', *Anal. Chem.*, **65**, 3435–3440 (1993).
38. G. Schiavon, N. Comisso, R. Toniolo, G. Bontempelli, 'Pulsed Amperometric Detection of Ethanol in Breath by Gold Electrodes Supported on Ion Exchange Membranes (Solid Polymer Electrolytes)', *Electroanalysis*, **8**, 544–548 (1996).
39. F. Pariente, J.L. Alonso, H.D. Abruña, 'Chemically Modified Electrode for the Selective and Sensitive Determination of Nitric Oxide (NO) in Vitro and in Biological Systems', *J. Electroanal. Chem.*, **379**, 191–197 (1994).
40. B. Fabre, S. Burlet, R. Cespuglio, G. Bidan, 'Voltammetric Detection of NO in the Rat Brain with an Electronic Conducting Polymer and Nafion® Bilayer-Coated Carbon Fibre Electrode', *J. Electroanal. Chem.*, **426**, 75–83 (1997).
41. S.S.M. Hassan, M.A. Ahmed, K.H.A. Mageed, 'Potentiometric Gas Sensing Probe for the Selective Determination of Acetate Ion', *Anal. Chem.*, **66**, 492–496 (1994).
42. R.K. Meruva, M.E. Meyerhoff, 'Potentiometric Oxygen Sensor Based on Mixed Potential of Cobalt Wire Electrode', *Anal. Chim. Acta*, **341**, 187–194 (1997).
43. J.P. Hart, A.K. Abass, 'A Disposable Amperometric Gas Sensor for Sulphur-Containing Compounds Based on a Chemically Modified Screen Printed Carbon Electrode Coated with a Hydrogel', *Anal. Chim. Acta*, **342**, 199–206 (1997).
44. L.R. Jordan, P.C. Hauser, G.A. Dawson, 'Amperometric Sensor for Monitoring Ethylene', *Anal. Chem.*, **69**, 558–562 (1997).
45. L.R. Jordan, P.C. Hauser, 'Electrochemical Sensor for Acetylene', *Anal. Chem.*, **69**, 2669–2672 (1997).
46. F. Opekar, S. Bruckenstein, 'Determination of Gaseous Hydrogen Sulfide by Cathodic Stripping Voltammetry after Preconcentration on a Silver Metalized Porous Membrane Electrode', *Anal. Chem.*, **56**, 1206–1209 (1984).
47. A. Kaisheva, I. Iliev, S. Christov, R. Kazareva, 'Electrochemical Gas Biosensor for Phenol', *Sensors and Actuators B*,**44**, 571–577 (1997).
48. N. Nakano, S. Ogawa, 'Preparation of Thin Gold-Film Electrode for an Electrochemical Sensor for Phosphine and Arsine', *Sensors and Actuators B*, **21**, 51–55 (1994).
49. J. Li, J. Janata, M. Josowicz, 'Application of Poly(cyclophosphazene) for Potentiometric Detection of Tributyl Phosphate Vapor', *Electroanalysis*, **8**, 778–783 (1996).
50. N. Hara, D.D. Macdonald, 'Development of Dissolved Hydrogen Sensors Based on Yttria-Stabilized Zirconia Solid Electrolyte with Noble Metal Electrodes', *J. Electrochem. Soc.*, **144**, 4152–4157 (1997).
51. Y.C. Zhang, H. Tagawa, S. Asakura, J. Mizusaki, H. Narita, 'Solid-State CO_2 Sensor with Li_2CO_3–Li_3PO_4–$LiAlO_2$ Electrolyte and $LiCoO_2$–Co_3O_4 as Solid Reference Electrode', *J. Electrochem. Soc.*, **144**, 4345–4350 (1997).

52. D.L. Ellis, M.R. Zakin, L.S. Bernstein, M.F. Rubner, 'Development of Electrically Conductive Poly(3-hexylthiophene) as a Thin Film Sensor for Hydrazine Vapor', *Mat. Res. Soc. Symp. Proc.*, **293**, 159–163 (1993).
53. J.A. Cox, K.S. Alber, 'Amperometric Gas-Phase Sensor for the Determination of Ammonia in a Solid-State Cell Prepared by a Sol–Gel Process', *J. Electrochem. Soc.*, **143**, L126–L128 (1996).
54. H.L. Huang, P.K.Dasgupta, 'Electrochemical Sensing of Gases Based on Liquid Collection Interfaces', *Electroanalysis*, **9**, 585–591 (1997).
55. H. Huang, P.K. Dasgupta, Z. Genfa, J. Wang, 'A Pulse Amperometric Sensor for the Measurement of Atmospheric Hydrogen Peroxide', *Anal. Chem.*, **68**, 2062–2066 (1996).
56. J. Binder, 'New Generation of Automotive Sensors to Fulfill the Requirements of Fuel Economy and Emission Control', *Sensors and Actuators A*, **31**, 60–67 (1992).
57. M. Hikuma, M. Takeda, H. Matsuoka, I. Karube, 'Reagentless Enzyme-Based Sensor Using a Gas-Permeable Membrane for Determination of Alcohols', *Anal. Chim. Acta*, **306**, 209–215 (1995).
58. H. Wang, J.B. Lando, 'Gas-Sensing Mechanism of Phthalocyanine Langmuir–Blodgett Films', *Langmuir*, **10**, 790–796 (1994).
59. K. Domanský, D.L. Baldwin, J.W. Grate, T.B. Hall, J. Li, M. Josowicz, J. Janata, 'Development and Calibration of Field-Effect Transistor-Based Sensor Array for Measurement of Hydrogen and Ammonia Gas Mixtures in Humid Air', *Anal. Chem.*, **70**, 473–481 (1998).
60. S. Alegret, J. Bartroli, C. Jiménez, M. del Valle, C. Dominguez, E. Cabruja, A. Merlos, 'Flow-Through pH-ISFet as Detector in Automated Determinations', *Electroanalysis*, **3**, 349–354 (1991).
61. A.R. Katritzky, J.N. Lam, H.M. Faid-Allah, 'Synthesis of New Microsensor Coatings and Their Response to Test Vapors', *Talanta*, **38**, 535–540 (1991).
62. M.S. Nieuwenhuizen, J.L.N. Harteveld, 'Development of a Surface Acoustic Wave Gas Sensor for Organophosphorus Nerve Agents Employing Lanthanide Compounds as the Chemical Interface', *Talanta*, **41**, 461–472 (1994).

Field-based Analysis of Organic Vapors in Air

Richard E. Berkley
US Environmental Protection Agency, Research Triangle Park, NC, USA

1 Introduction **2**
2 Characteristics of Laboratory- and Field-Based Analyses **2**
2.1 Advantages of Laboratory-based Operation 2
2.2 Disadvantages of Laboratory-based Operation 2
2.3 Advantages of Field-based Operation 2
2.4 Disadvantages of Field-based Operation 3
2.5 Use of Laboratory- and Field-based Analysis Together 3
3 Detectors **3**
3.1 Universal Detectors 3
3.2 Selective Detectors 4
4 Commercially Available Deployable Gas Chromatographs **5**
4.1 Characteristics and Capabilities 5
4.2 Evolutionary Trends in Commercially Available Instrumentation 5
5 Field Evaluation Studies **5**
5.1 Vilnius, Lithuania 6
5.2 Northern Delaware 6
5.3 Crosby, Texas 7
5.4 Bayamón, Puerto Rico 8
5.5 Lawrence, Kansas 11
5.6 Reference Method 11
5.7 Source Generator 11
5.8 Procedure 11
5.9 Audit Standards 11
5.10 Plume Sampling 12
6 Conduct of Field Evaluation Studies **13**
6.1 Organization 13
6.2 Site Selection 13
6.3 Existing Sources 13
6.4 Artificial Sources 13
6.5 Reference Analytical Method 13
6.6 Audit Standards 14
6.7 Data Reporting 14
7 High-speed Laboratory Analysis Along with Field-based Analysis **14**
7.1 High-speed Gas Chromatography 14
7.2 High-speed Mass Spectrometry 14
7.3 High-speed Gas Chromatography/Mass Spectrometry Analysis 15
8 Conclusion **15**
8.1 Present State of Field-based Analysis 15
8.2 Future Prospects 16
Abbreviations and Acronyms **16**
Related Articles **17**
References **17**

Volatile organic compounds (VOCs) exist in the vapor phase in ambient air at partial pressures ranging from less than 1 part per billion to a few parts per million. Some of them are toxic and potentially harmful to the environment and to human health. The threshold concentrations at which individual compounds become harmful is often not well defined, and less is known about mixtures. Accurate assessment of health risk requires analytical methods that are accurate, affordable, convenient, fast, and sensitive enough to detect each compound at a concentration below the threshold of harmful effect. The usual approach to identification and quantitation of them has been to collect samples and submit them to laboratory analysis. United States Environmental Protection Agency (USEPA) Method TO-14 is a standard procedure for analyzing a group of 41 potentially harmful nonpolar and semipolar organic compounds, which is in widespread use and has been found to detect and measure target compounds reliably under most conditions. However, more rapid delivery of data is often needed, and samples can deteriorate during storage pending analysis. Alternative analytical methods that produce immediate data are needed. Analysis of samples at the time and place of collection has seldom been done because of the lack of suitable equipment and the vicissitudes of operating it under field conditions. Portable or field-deployable gas chromatographs (PGCs) have, with cause, been regarded as less accurate than laboratory equipment, and a perception remains that analyses done in the field are less useful than those done in the laboratory. In fact, laboratory- and field-based analyses each have inherent advantages and disadvantages, and using them together can finesse the disadvantages of both. A high-speed gas chromatograph (GC) using a time-of-flight mass spectrometer (TOFMS) detector can mitigate the problem of delayed data.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

VOCs exist in the vapor phase in ambient air at partial pressures ranging from less than 1 part per billion (ppb) to a few parts per million (ppm). Some of them originate from natural processes, such as plant and animal metabolism, but most come from chemical manufacturing, combustion, and related processes. Some VOCs are toxic and potentially harmful to the environment and to human health. The threshold concentrations at which individual compounds become harmful is often not well defined, and less is known about mixtures. Accurate assessment of health risk requires analytical methods that are accurate, affordable, convenient, fast, and sensitive enough to detect each compound at a concentration below the threshold of harmful effect. There are many VOCs in air at trace concentrations, and the more sensitive the analytical method, the greater is the number that will be detected. The usual approach to their identification and quantitation has been to collect samples and submit them to laboratory analysis. A typical procedure includes (1) collection of whole air samples in passivated canisters or separation of VOCs from air on a sorbent cartridge, (2) delivery of samples for analysis, either by passing whole air from a canister sample through a cryogenic trap to separate VOCs from the air matrix or by thermal desorption of VOCs from a cartridge into an inert gas stream which then passes through a cryogenic trap, followed by (3) vaporization from the cryotrap into the carrier stream of a GC which may be equipped with any of a variety of specific and nonspecific detectors. The USEPA Method TO-14, developed by Pleil and McClenny,[1] is a standard procedure for analyzing a group of 41 nonpolar and semipolar organic compounds which are potentially harmful and often found in ambient air as combustion products, engine exhaust, or fugitive emissions from various human activities. Method TO-14 is in widespread use and has been found to detect and measure target compounds reliably under most conditions, but there are circumstances in which more rapid delivery of data would be desirable and conditions under which samples awaiting analysis might become compromised. Alternative methods of analysis that produce immediate data are needed. In the past, analysis of samples at the time and place of collection has seldom been done because of the lack of suitable equipment and the vicissitudes of operating it under field conditions. PGCs have, with cause, been regarded as less accurate than laboratory equipment, and a perception remains that analyses done in the field are less useful than those done in the laboratory. In fact, both have inherent strengths and inherent weaknesses.

2 CHARACTERISTICS OF LABORATORY- AND FIELD-BASED ANALYSES

2.1 Advantages of Laboratory-based Operation

Laboratory instruments are usually near state-of-the-art and are operated by competent personnel in a sheltered environment, which inspires confidence in their reliability. Because the operating environment is stable, detection limits (DLs) can readily be minimized. It is desirable to operate at maximum sensitivity for several reasons: (1) it may not be known at what level the harmful effect of a compound becomes significant; (2) if the method is continuously detecting trace levels, there is a positive indication of proper function, which is not the case if nothing at all is reported until some threshold concentration is exceeded; (3) there are worldwide ambient background concentrations of certain compounds, e.g. tetrachloromethane (CAS 56-23-5) at ~0.13 ppb, which indicate proper function if the method is sensitive enough.[2]

2.2 Disadvantages of Laboratory-based Operation

Large monitoring projects produce backlogs of samples which endure long periods of storage, during which samples may deteriorate, especially if highly reactive or corrosive compounds were collected. A sample analyzed accurately in the laboratory may no longer be the sample that was collected in the field. Although sample integrity during storage in passivated canisters has been demonstrated in the absence of highly reactive compounds, artifact formation can be caused, for example, by HCl.[3–5] Samples are analyzed one at a time, and analyses can be delayed by equipment malfunction. Meanwhile, sample collection is blind, guided only by theoretical inferences and human olfaction, and canisters or cartridges awaiting analysis are not available for use. A faster way to obtain data is needed, one which can point to places which need more thorough investigation and also provide timely estimate of risk to field personnel.

2.3 Advantages of Field-based Operation

Portable instruments, which analyze VOCs in ambient air without storing or preconcentrating samples, can minimize sampling errors and produce data in near real-time, so that sampling protocols can be modified to use results as they are obtained. Sample ''storage'' is vanishingly brief, so that there is little or no time for samples to become corrupt. Since PGCs are more easily transported than large numbers of sampling devices, they can produce larger volumes of data at lower cost. Also, when analysis is performed on-site, data are immediately available, and more thorough investigation can be done faster at less cost.

2.4 Disadvantages of Field-based Operation

Many of the disadvantages of on-site analysis arise from the instability of field environments. Sunshine or rain falling directly on a PGC can wreak havoc and perhaps terminate function. GCs in general, whether laboratory or portable, cannot perform while subjected to significant fluctuations in ambient temperature, since that affects both column temperature and detector sensitivity. Every field operator has experienced the frustration of seeing careful PGC placement foiled by inclement weather and fervently wished for costlier, heavier, bulkier but more robust equipment. There are other disadvantages: many PGC columns are isothermal, so they are not able to analyze as many compounds per run as temperature-programmed laboratory instruments. Design compromises intended to reduce cost and enhance portability and the rate of sample throughput can limit performance. The short columns used to minimize analysis times sacrifice resolution for speed. PGCs can be calibrated for only the limited number of compounds which their (usually) specific detectors can see at low parts per billion levels, and they almost always rely heavily on retention time for compound identification. Finally, they require skilled operators. For these reasons, laboratory-based analysis remains more popular, although it must be conceded that data from good analyses of bad samples can be as useless as data from bad analyses of good samples.

2.5 Use of Laboratory- and Field-based Analysis Together

It has become widely appreciated that the integrity of sample-handling procedures, no matter how sedulously designed, cannot be taken for granted. On-site and laboratory-based analyses can be used together so that their respective advantages and disadvantages complement each other.[6–10] PGCs can be useful for showing where intensive sampling is needed, so that laboratory-based analysis can be applied with maximum efficiency, and they can validate data from laboratory analyses because their data quality is not affected by sample degradation during storage. If laboratory and field methods consistently produce similar results, then probably both are correct. If they differ, that difference may indicate what went wrong and what can be done about it.

3 DETECTORS

A variety of detectors are available for PGC. There are two kinds: *universal detectors* respond to almost everything that elutes from the column whereas *selective detectors* respond to certain classes of compounds while ignoring all others.

3.1 Universal Detectors

Universal detectors are those which respond to any change in the composition of the carrier stream.

The *thermal conductivity detector* (TCD) consists of a pair of filaments or thermistors which form half of a Wheatstone bridge. The carrier stream and a reference stream flow past each of them at the same rate. When both streams are identical, the filaments or thermistors are cooled at the same rate, the bridge stays in balance and the voltage across it is zero. Any change in the thermal conductivity of either stream unbalances the bridge, resulting in a net voltage which is directed to a recording device. Obviously stable flow rates are critical to proper function. Since the thermal conductivity of a gas consists of a combination of translational, rotational, and vibrational energy modes, the introduction of molecules other than the carrier gas will change it. Among commonly used carrier gases helium is monatomic, having only translational energy, so any molecular species eluting in a helium carrier stream must produce a response. TCDs are not very sensitive, although Hewlett-Packard/Microsensor Technology (MTI) produces chromatographs with a micromachined TCD which has a volume of 240 nL and can detect airborne vapors at concentrations as low as ~1 ppm. An instrument of that sensitivity is primarily useful for industrial hygiene application, and preconcentration of samples by three orders of magnitude is necessary to use it for ambient air analysis.

A *flame ionization detector* (FID) feeds the carrier stream into a hydrogen–air flame in which eluents are burned, producing ions to be collected on electrodes. The resulting current is amplified and sent to a detector. A moderate concentration of water in the carrier stream will not be detected, since it is the principal product of the flame. Other substances that do not burn are detectable only to the extent that they ionize. The response to a compound is roughly proportional to the number of its carbon atoms. FIDs are more sensitive than TCDs but are not the most sensitive detectors. They are frequently used without chromatography as a "total hydrocarbon" monitor. Their relative insensitivity and their need for hydrogen and air supply limit their use in the field, although some commercially available PGCs are equipped with them.

An *argon ionization detector* (AID) ionizes atoms of argon carrier gas with a radioactive source. The argon ions then react with analyte molecules to ionize them, after which they are collected on electrodes. Because of the relatively high energy of the process (11.7 eV), most compounds respond. AIDs are relatively sensitive and are useful for analysis of highly volatile "light end" compounds which, having fewer carbon atoms to oxidize,

are least detectable with an FID, and, having fewer vibrational/rotational modes, are least detectable with a TCD. Since there are fewer isomers of low-molecular-weight compounds (butane and smaller), all of them can be separated with even the short columns used in PGCs. For larger molecules, it becomes increasingly difficult to distinguish toxic analytes from the multitude of relatively innocuous background compounds that co-elute.

A *high-energy photoionization detector* (PID) ionizes eluting compounds with an 11.7-eV ultraviolet lamp. It is a universal detector for the same reason that an AID is universal, and its performance is similar. Its function is analogous to that of lower energy PIDs as described below.

3.2 Selective Detectors

PGCs generally use short columns to facilitate short analysis times, so their chromatographic resolution is low, but even laboratory GCs with full-length columns cannot resolve all the trace-level (<5 ppb) compounds in an air sample. Typically, when a sample containing trace levels is analyzed on a temperature-programmed column using a mass spectrometer detector, about 200 peaks appear, each containing 3–5 compounds that can be recognized and an unknown number of others which escape tedious scrutiny. A detector that responds only to compounds of interest is clearly useful. Selective detectors respond to specific classes of compounds but not to other compounds that co-elute with them.

The *PID* consists of an ultraviolet lamp which irradiates the carrier stream. Eluent molecules having ionization potentials lower than the radiation energy are ionized to produce an ion current. Detector selectivity is determined by radiation energy. High-energy lamps (11.7 eV) are virtually universal detectors. Lower energy lamps (10.6 eV and below) ionize many double-bonded compounds, such as aromatic hydrocarbons and olefins, but not saturated compounds such as hydrocarbons and halocarbons. Toxic and carcinogenic compounds with double bonds are readily detected by low-energy PIDs. However, not all low-energy PIDs are alike. Some of them use medium-pressure lamps, others use low-pressure lamps and both have advantages. Medium-pressure lamps are physically more robust than low-pressure lamps and emit a relatively broad band of wavelengths. They run hot, so they can be used with heated columns. Low-pressure lamps emit an extremely narrow resonance fluorescence line with intensity far exceeding that of a medium-pressure lamp. Detectors equipped with low-pressure lamps are extremely sensitive. The disadvantage is that heating them causes the emission line to broaden and lose intensity. As a result, PIDs based on low-pressure lamps are used for isothermal analysis of volatile compounds. They are generally less robust than medium-pressure lamps, and their service life is shorter. The intensity of light emitted by low-pressure lamps is so great that second photon absorptions cause ionization of compounds with ionization potentials slightly above lamp energy. A low-pressure PID-equipped PGC can be calibrated for compounds with higher ionization potentials, but it is advisable not to do that when analyzing trace levels in ambient air. For example, tetrachloromethane can be detected at concentrations of 1 ppm and above, but it elutes near benzene (CAS 71-43-2) on many columns. Benzene, being formed in all manner of combustion processes, is ubiquitous and detectable at sub-parts per billion concentrations, so it could be erroneously identified by a microprocessor as tetrachloromethane at a concentration high enough to precipitate panic. The development of PID detectors has been reviewed.[11]

The *surface acoustic wave* (SAW) detector is based on the ability of molecules adsorbed on a piezoelectric crystal to change the velocity of a sound wave propagated along its surface. Vibration by the sound wave causes the crystal to generate an electric current which can be amplified and recorded. Deposition of polymers on the surface can alter the velocity at which sound is propagated so as to tune the crystal to respond to three or four compounds in preference to others that might co-elute with them. SAW detectors for a variety of compounds are under development.[12] In general they ignore all but very high concentrations of coeluting compounds.

The *electron capture detector* (ECD) uses a radioactive source to ionize the carrier gas, producing an electron current which is continuously monitored. When a compound with high electron affinity arrives in the detector it captures electrons, and the resulting decrease in electron current is detected. ECDs are extremely sensitive to perhalo compounds such as tetrachloromethane, tetrachloroethylene (CAS 127-18-4) and hexachlorobutadiene (CAS 87-68-3). They are less sensitive to compounds having fewer halogens. Since they are highly selective, they are potentially useful for detection of any compounds with high electron affinity.

Mass spectrometers function as both universal and specific GC detectors. They have not been used much in portable instruments because of size and cost, although field-deployable gas chromatograph/mass spectrometer (GC/MS) instruments are now commercially available. The use of mass spectrometers as GC detectors is discussed in section 7.3.

A variety of other specific detectors, including flame photometric, electrolytic conductivity, far-ultraviolet and thermionic ionization, can be used with some of the commercially available PGCs.

4 COMMERCIALLY AVAILABLE DEPLOYABLE GAS CHROMATOGRAPHS

Early PGCs were crude, insensitive devices, hardly capable of analyzing anything of significance. They typically featured syringe injection onto a packed column leading to a TCD connected to a galvanometer or an external strip-chart recorder. When the Photovac (PVK) 10A10 appeared, it was claimed to detect less than 1 ppb of benzene in 1 mL of air. It differed from predecessors only in that it had a low-pressure PID.[13] Its simple design was followed by instruments from PVK and other vendors which automatically collect samples and inject them into capillary GC columns and use many different detectors. Data are processed by on-board microprocessor-based algorithms. Such innovations have made the task of on-site analysis more practical and more accurate for both identification and quantitation.

4.1 Characteristics and Capabilities

Field-based analytical instruments can be characterized as portable or deployable. Truly *portable* instruments are virtually self-contained. They are battery powered and have an internal carrier gas supply. An operator can lift the instrument with one hand, not necessarily while in use, although some PGCs are almost capable of hand-held operation. A *field-deployable* instrument requires line power and usually an external carrier gas supply also. It may be too heavy for a single operator to shift conveniently. Some are laboratory instruments that have been reduced in size for convenience and hardened for transport. Some currently and potentially available portable GCs are listed in Table 1 and field-deployable GCs are listed in Table 2.

4.2 Evolutionary Trends in Commercially Available Instrumentation

There is considerable variety in technical detail among PGCs. The current market is reminiscent of the early days of the automobile when steam power and battery power vied with internal combustion power rolling on tires, solid or fragile, along roads that were scarcely passable, burning fuel that was marginally available. Evolution of the automobile subsequently converged upon the most practical technical designs, and PGCs will do likewise. At present, emulation of data from concurrent collocated laboratory-analyzed samples is prominent among criteria for evaluation of PGCs. In future, it may become more common for laboratory analyses to be judged by how well they emulate data produced by field-deployable instruments, since they analyze fresh samples.

5 FIELD EVALUATION STUDIES

Beginning in the mid-1980s, the USEPA conducted a series of field evaluation studies of PGCs. The study design has evolved from early attempts to determine

Table 1 Commercially available portable GCs

PVK Voyager	36 × 31 × 12 cm 6 kg	110 V AC 12 V DC	Nitrogen carrier	Autosample (loop) or syringe port
	Isothermal column at 30–80 °C	Precolumn backflush	Detectors: PID, ECD	Data to external computer
Microsensor Systems Inc. Model 301	8 × 36 × 31 cm 5.5 kg	110 V AC 12 V DC	Scrubbed ambient air carrier	Autosample (sorbent trap)
	Isothermal at 65 °C	No column backflush	Detector: SAW	Internal processor, external computer or printer
STX Scentograph	17 × 34 × 50 cm 12 kg	110 V AC 12 V DC	Argon, helium carrier	Autosample to sorbent bed or syringe port
	Isothermal to 180 °C	Column backflush	Detectors: AID, ECD, TCD, PID	External computer
MTI Inc. Model P200	15 × 25 × 35 cm 5.4 kg	110 V AC 12 V DC	Helium, nitrogen, argon carrier	Autosample (loop)
	Isothermal between 30 and 200 °C	No column backflush	Detector: 240-nL TCD	External printer or computer
Inficon Hapsite	46 × 43 × 18 cm 16 kg	24 V DC (GC/MS) 110 V AC (service module)	Nitrogen carrier	Direct sample inlet
	Isothermal column 40–100 °C	No column backflush	Detector: Getter pumped MSD	Internal processor or external computer

AC, alternating current; DC, direct current; STX, sentex; MSD, mass selective detector.

Table 2 Commercially available field-deployable GCs

HNU Systems Model 311	38 × 56 × 25 cm 25 kg	110 V AC	Any GC carrier gas	Autosample (loop) or syringe port
	Isothermal or programmable to 200 °C at 15 °C s^{-1}	Column backflush	Detectors: any two of PID, ECD, FUV or FID	Internal printer or external computer
SRI Instruments Model 8610C	32 × 37 × 32 cm 12–25 kg	110 V AC or 220 V AC	Electronic pressure control. Helium, nitrogen, scrubbed ambient air	Purge and trap, autosample (sorbent), or syringe port
	Programmable to 400 °C	No column backflush	Detectors: any three of PID, FID, ELCD, Dry ELCD, FPD, TID, ECD, TCD, NPD, HID	External printer or computer
Viking SpectraTrak 573	31 × 46 × 61 cm 34 kg	110 V AC	Helium, nitrogen, hydrogen carrier	Autosample (sorbent or loop) or syringe port
	Programmable to 400 °C at 50 °C min^{-1}	No column backflush	Detector: Hewlett–Packard 5973 MSD	Embedded pentium® computer
CMS MiniCAMS	25 × 31 × 22 cm 9 kg	110 V AC	Nitrogen, hydrogen, air carriers	Autosample (sorbent or loop) or syringe port
	Isothermal or ballistically programmable to 350 °C	No column backflush	Detectors: FPD, pulsed FPD, PID + FID or PID + XSD or FPD + FID	Internal microprocessor
Chromolytics Microfast	15 × 36 × 20 cm	110 V AC	Hydrogen carrier	Autosample (sorbent or loop)
	Programmable to 260 °C at 15 °C s^{-1}	Precolumn backflush	Carrier: Dual FID	External computer

FUV, far-ultraviolet detector; ELCD, electrolytic conductivity detector; FPD, flame photometric detector; TID, thermionic ionization detector; NPD, nitrogen–phosphorus detector; HID, helium ionization detector; XSD, halogen-sensitive detector.

whether the PGC could detect anything at all toward a systematic comparison of PGC data with data obtained by other means. The earliest studies were conducted in the southern USA at various locations from Texas to Virginia. Gradually USEPA Method TO-14 has become the usual reference standard method for comparison with PGC data. Some of the studies are described below.

5.1 Vilnius, Lithuania

In May and June 1989, two PVK 10S70 PGCs were operated at a roadside site near Vilnius in a joint study of automobile exhaust by the USEPA and the Voeikov Main Geophysical Observatory in Leningrad. Whole air samples were also collected in passivated canisters which were returned to the USA and analyzed by Method TO-14. Time-integrated canister samples were collected over a 2-h period, while the PGC data were periodic grab samples taken during that time. Not enough canisters could be transported to dedicate part of them to simultaneous grab sampling with PGCs. Although not identical, the PGC results were in close agreement with each other and in substantial agreement with TO-14 data.[8]

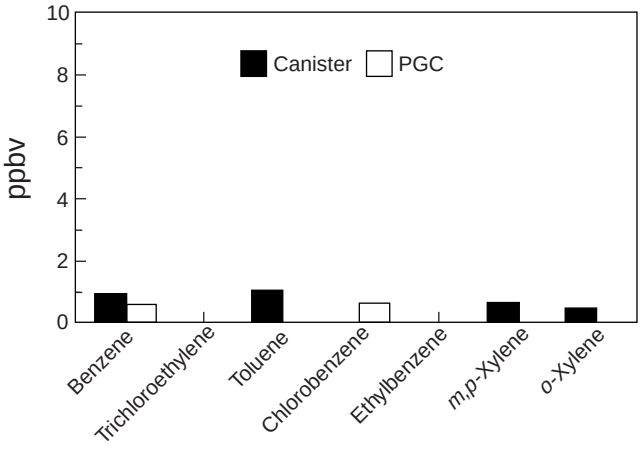

Figure 1 Concentrations of benzene, trichloroethylene (CAS 79-01-6), toluene, chlorobenzene (CAS 108-90-7), ethylbenzene (CAS 100-41-4), *m*- and *p*-xylene (CAS 108-38-3 and CAS 106-42-3) and *o*-xylene (CAS 95-47-6) found at Grantham Lane.

5.2 Northern Delaware

Also in 1989, collocated canister grab samples were taken simultaneously with PVK 10S70 samples at various industrial sites in Northern Delaware. Figures 1–4 show progressively increasing agreement between the two methods as concentrations increase.[7–9]

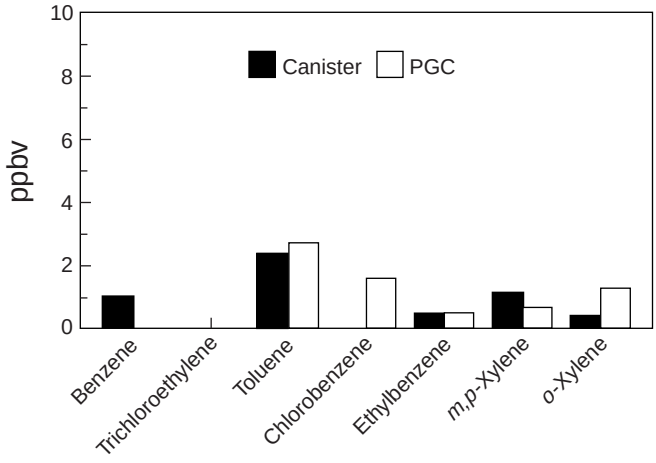

Figure 2 Concentrations of VOCs found at the Halby Waste Lagoon.

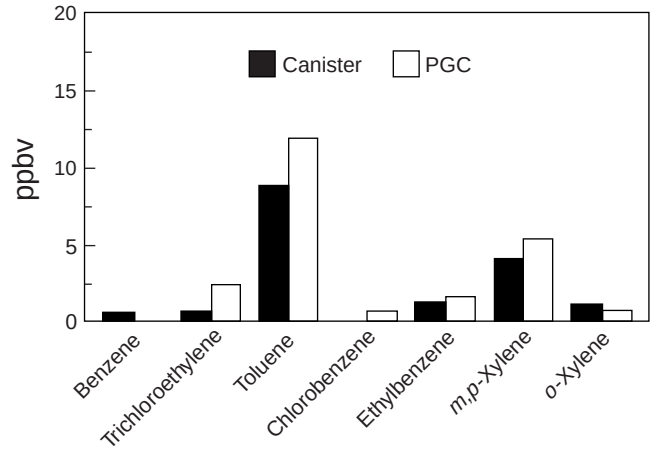

Figure 3 Concentrations of VOCs found at the Pigeon Point Waste Disposal Facility.

5.3 Crosby, Texas

In January 1992, six commercially available PGCs were evaluated at a Superfund site during start-up of bioremediation. This site was an abandoned and flooded gravel pit into which refinery waste had been dumped. Floating organic material on the surface of the pond had been burned periodically, leaving a layer of about 4 m of organic sludge overlain by a like depth of water that was clean enough to support some aquatic life, but volatile solvents were leaching out of the sludge into groundwater and threatening nearby water wells. Bioremediation was done by vigorous mixing of water and sludge while oxygenating and fertilizing to encourage some of the ambient bacteria. In a pilot study, the air used to supply oxygen had purged out substantial amounts of VOCs which led to high airborne concentrations in the

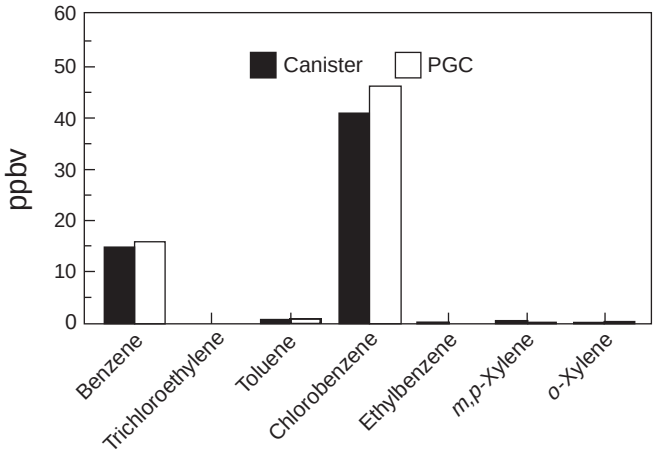

Figure 4 Concentrations of VOCs found near a chemical plant.

surrounding area. During actual bioremediation, pure oxygen was used instead of air, and there was very little airborne contamination. The PGCs were located near the perimeter of the pond, where concentrations of VOCs proved to be slightly above ambient background levels. Collocated grab samples were collected in Summa-polished canisters during PGC sampling. They were analyzed by Method TO-14 using a MSD and served as reference standards to assess the accuracy of data reported by the PGCs.

The PGCs included a PVK 10S Plus with a 10.6-eV PID, Microsensor Systems 301 with a SAW, STX Scentograph with an 11.7-eV AID, HNU Model 311 with a 10.2-eV PID and SRI 8610 with a 10.2-eV PID and an ELCD. A PVK 10S70 owned by the USEPA was also used. All units were operated inside a shed where the temperature was maintained at about 21 °C. It was located 6 m laterally and 4 m above the edge of the pond. Each unit used its own sample pump to import outside air through 1/8 in o.d. stainless-steel tubing. Calibration checks were performed periodically using mixtures prepared by dynamic dilution of commercial standards and stored in 6-L Summa-polished canisters. Grab samples were collected by opening the valve of an evacuated canister while holding it within 75 cm of the gathered ends of the sampling tubes while samples were collected. Canister samples were analyzed by Method TO-14 using gas chromatography with mass selective detection (GC/MSD). Canister data were taken to indicate true identities and concentrations of the compounds present. DLs for the PGCs, shown in Table 3, were calculated using data acquired during field calibrations. They were not as low as DLs determined under laboratory conditions but were typical of what can be expected under field conditions. Figures 5 and 6 contrast data from two grab samples with corresponding PGC data. Labeled vertical bars indicate PGC data and

Table 3 DLs (ppb) for PGCs calculated from field calibrations

Instrument	Benzene	Toluene	Tetrachloroethylene	Chlorobenzene
PVK 10S Plus	0.5	1.2	0.5	1.5
MSI 301	6.7	20.5	INT	INT
STX Scentograph	3.8	4.3	3.4	7.6
HNU 311	2.7	3.6	4.9	4.2
SRI 8610 PID	0.4	0.3	0.2	0.2
SRI 8610 ELCD	NR	NR	2.4	5.2

INT = interference. Another peak or an elevated baseline made it impossible to calculate DL.
NR = no response to ELCD.

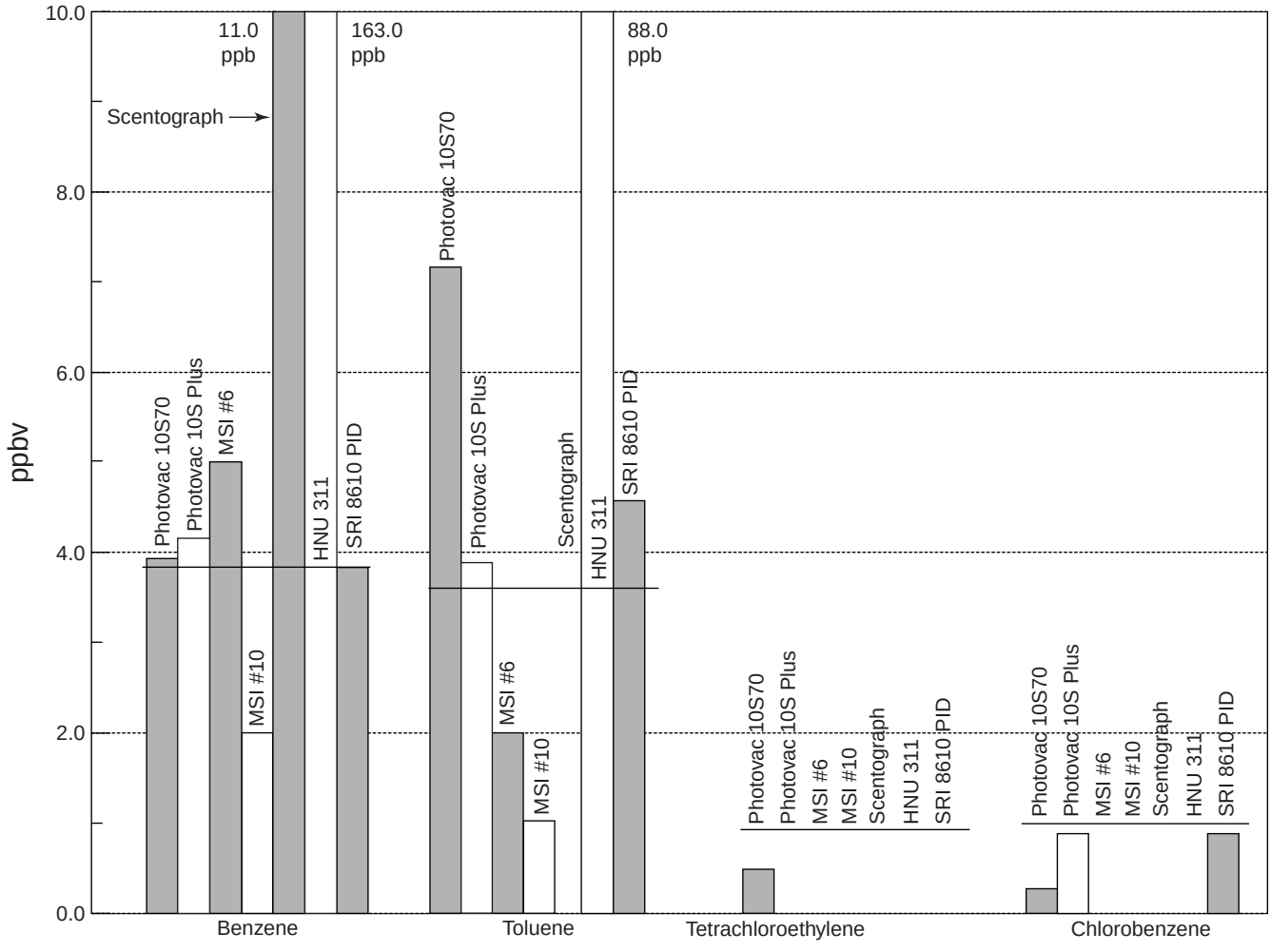

Figure 5 Concentrations of benzene, toluene, tetrachloroethylene (CAS 127-18-4) and chlorobenzene found at the French Limited Superfund Site.

canister data are shown as horizontal lines. Analyte levels were near the DLs in Table 3, and the agreement between PGCs was close enough to show that several of them could provide reasonable estimates of the compounds for which they were calibrated. A detailed report of the study has been published.[14]

5.4 Bayamón, Puerto Rico

During June 1993, a field study was conducted near San Juan to compare data produced by two PGCs with data from samples which were collected simultaneously in 6-L Summa-polished canisters and analyzed by Method

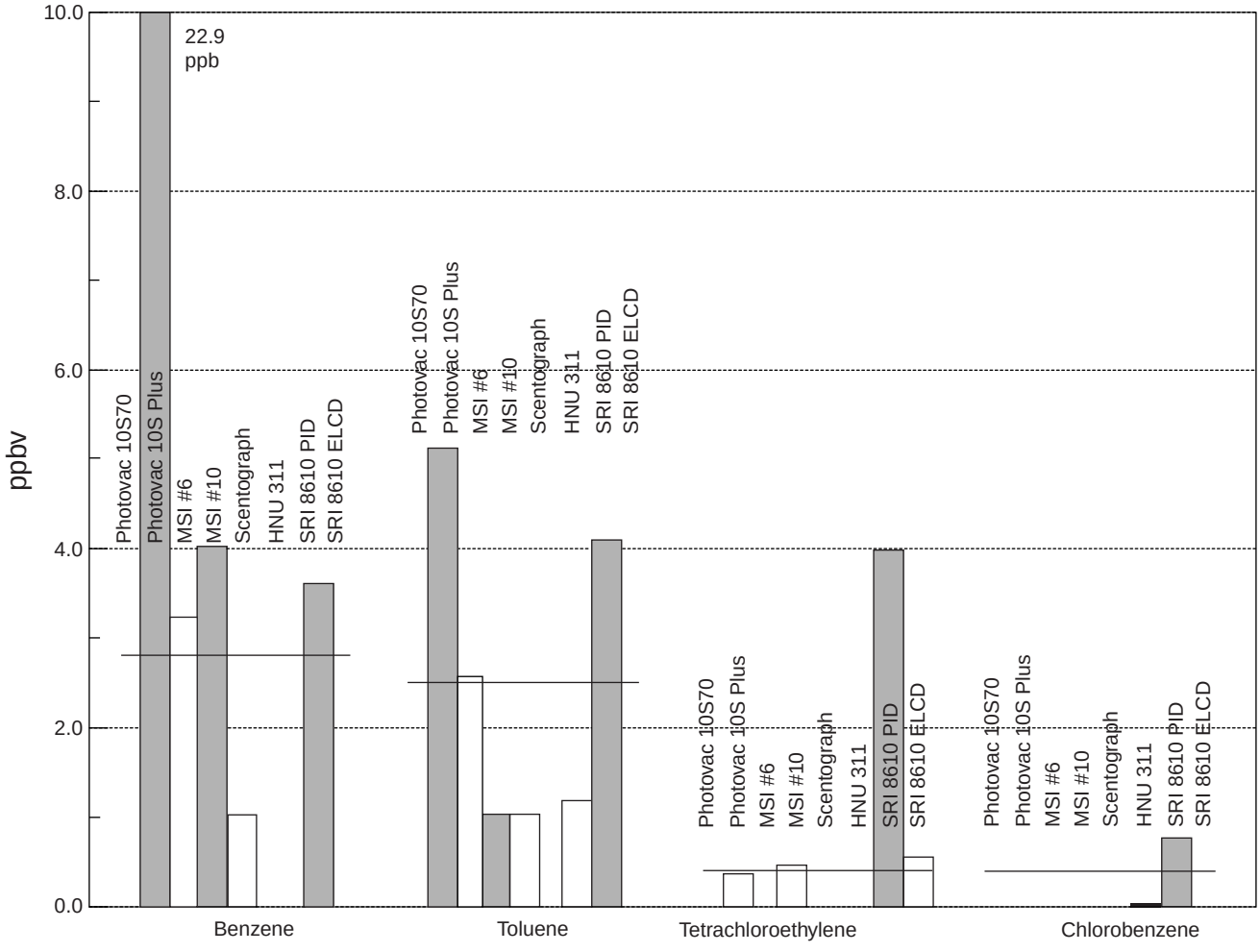

Figure 6 Concentrations of VOCs found at the French Limited Superfund Site.

TO-14. They were the CMS Series 2000 Field MiniCAMS with an FID and a solid-sorbent preconcentrator, and the HNU Model 311 with a PID which periodically sampled the continuously pumped air stream by diverting the contents of a sample loop. In this study, each commercial instrument was operated by a factory representative, and vendors were responsible for calibration, maintenance, and troubleshooting their own equipment. The site was at a monitoring station operated by the Junta de Calidad Ambiental (JCA) (Environmental Quality Board) at the jail in Bayamón, within the San Juan Metropolitan Area and downwind of the Cataño industrial area. JCA had two small shelters with air conditioning and electric power. The PGC operators collaborated to rent a recreational vehicle to which electric power from a JCA building was connected so it could be used as a mobile laboratory during the study. A schematic of the area around the site is shown in Figure 7. Ambient concentrations were lower than expected. Canister samples of ambient air collected during the study were analyzed by Method TO-14 using GC/MSD. Canisters were filled by holding the valve within 20 cm of the group of PGC inlets and opening it just far enough to fill the canister while the PGCs were sampling.

Occasionally, each instrument analyzed an audit standard which had been certified before the study by GC with FID and ECD. After the study it was analyzed again by GC/MSD. There was little change in the audit standard between analyses before and after the study. The standards included all 41 compounds on the TO-14 target list, although the PGCs were calibrated for only four of them. Each PGC performed several analyses of the audit standard, and field audit data were in close agreement with laboratory analyses for both PGCs.

PGCs were calibrated for benzene, toluene (CAS 108-88-3), *m*- and *p*-xylene (CAS 108-38-3 and CAS 106-42-3) and *o*-xylene (CAS 95-47-6). Agreement between PGC and canister data, although not exact, was close. Concentrations of analytes varied between 0 and 25 ppb. An understanding of how well the methods agreed can

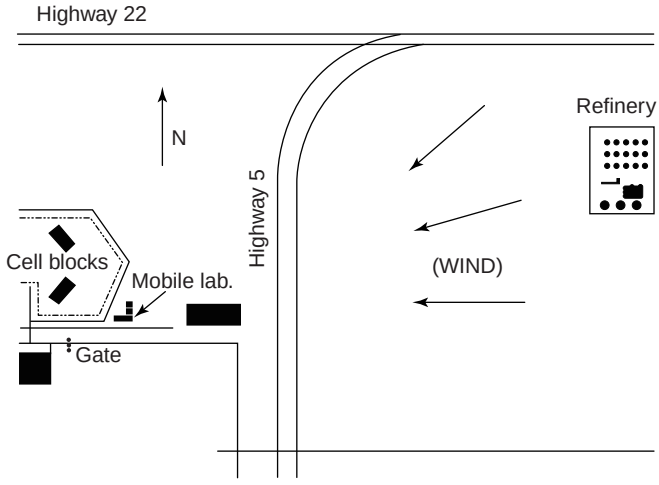

Figure 7 Map of the area between the refinery and the Bayamón Cárcel. (Reproduced by permission of InfoScience Services Corporation.)

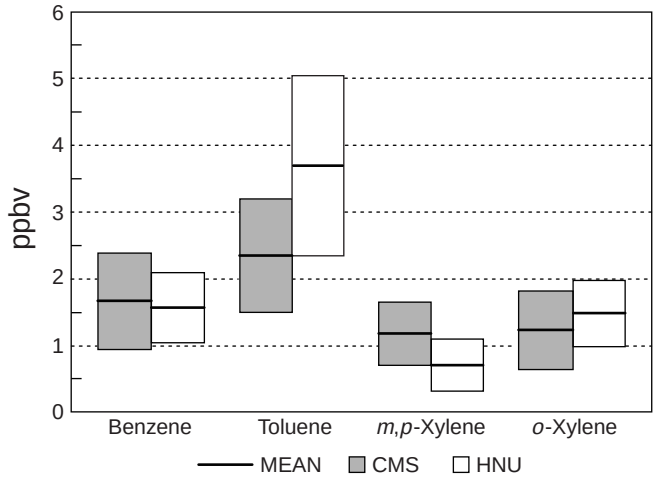

Figure 8 Average values (horizontal lines) and standard deviations (SDs) (vertical bars) of absolute values of differences between canister/TO-14 data and corresponding PGC data in the Bayamón Cárcel field study. (Reproduced by permission of InfoScience Services Corporation.)

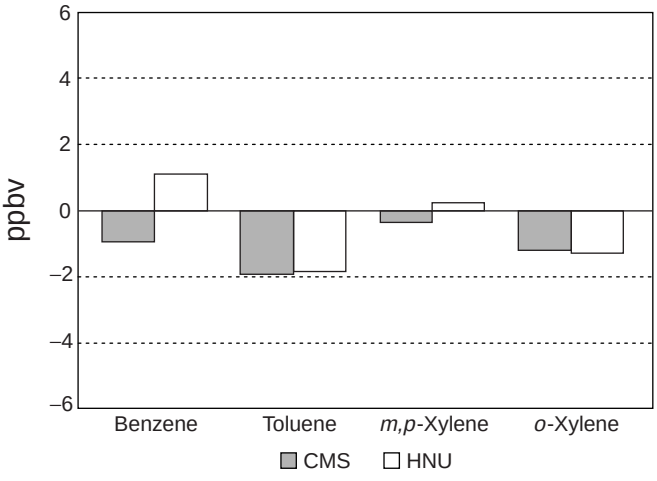

Figure 9 Average values of differences between canister/TO-14 data and corresponding PGC data in the Bayamón Cárcel field study. (Reproduced by permission of InfoScience Services Corporation.)

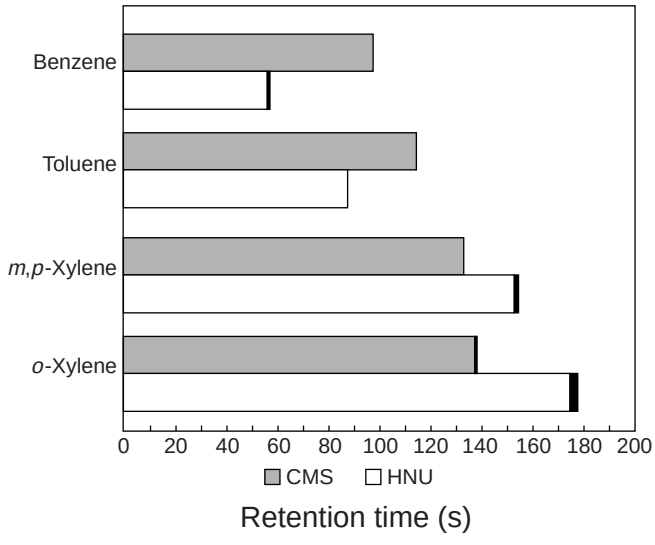

Figure 10 Average retention times and SDs of retention times for PGCs used in the Bayamón Cárcel field study. (Reproduced by permission of InfoScience Services Corporation.)

be obtained by averaging the absolute values of the differences between the PGC and TO-14 results for each compound for all comparison runs. Average absolute differences are represented by the elevations of the thick horizontal lines in Figure 8. The SD of the absolute values of the differences is shown by the bar extending above and below the line. Differences were finite, real, and relatively small, except at lowest concentrations. The averages are not much larger than their SDs, which suggests that the differences are essentially random and do not contain a large component of systematic error. The differences, averaged without taking absolute values so that positive and negative values tend to cancel each other, are shown in Figure 9. A slight negative bias is apparent.

If the peak recognition algorithm of a PGC has retention time windows set too wide, then interferences become more likely. If the variation in retention time for a certain compound is substantially larger than for the analytes eluting before and after it, then intermittent misidentification of an interfering compound could be the cause. The retention time stabilities of each analyte on each PGC were evaluated by comparing the SD of all retention times for each compound with the mean

retention time of that compound. Retention times are contrasted with SDs of retention times in Figure 10. The width of the tiny rectangular boxes at the ends of the bars represents the SD of retention times, which were small enough (invisible in most cases) to suggest that chromatographic conditions were stable during the study, and that the peak identification algorithms were finding only a single compound within each peak identification window. A detailed report of the study has been published.[15]

5.5 Lawrence, Kansas

In August 1995 a field study was conducted on the intramural playing fields of the University of Kansas. PGC vendors were invited to participate at their own expense. An artificial plume of analytes was used, so that control over identities and concentrations of analytes could be exercised, adequate use could be made of audit standards and field blanks, and as many participants as chose to attend could be accommodated. The six participants in this study included four PGC vendors, a contractor operating the reference Canister/Method TO-14 method, and a contractor operating a mobile triple quadrupole mass spectrometer (TAGA) mounted on a bus. The PGCs were the PVK 10S Plus, with a low-pressure 10.6-eV PID, the STX Scentograph with an 11.7-eV AID, the MTI M200 and the MSA-Baseline (MSA) Ekho-M. The MTI was operated by Louisiana State University Institute for Environmental Studies (LSU). It had a microbore column and a micromachined 240-nL TCD and used a separate preconcentrator. The Ekho-M chromatograph originated in the Siberian Department of the Russian Academy of Sciences as an explosives detector and was under development as an air monitor by MSA. Its column consisted of a hexagonal bundle of ~1000 parallel microbore capillaries, each 20 cm in length. The bundle was 2 mm in diameter. This unique design permitted analysis of a relatively large sample without overloading the column. An AID was used.

5.6 Reference Method

Reference samples were collected in Summa-polished 6-L canisters. A mechanical flow controller was used to fill canisters at a uniform rate during the sample collection period. In some cases, two canisters were filled on either side of the PGCs to check uniformity of the plume. In other cases, especially during turbulence, one canister was filled at the inlet of each PGC. Southwest Research Institute (SWR) analyzed the canister samples by Method TO-14. The TAGA was placed downwind of the sampling devices to confirm the presence of the plume before sampling began.

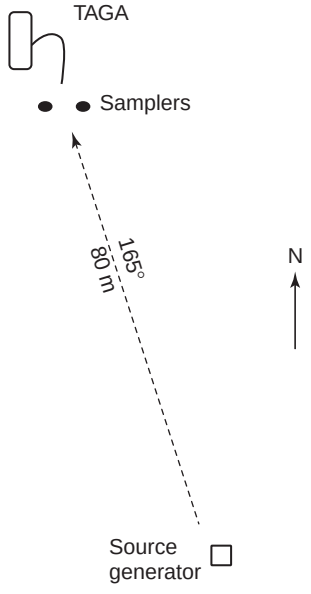

Figure 11 Positioning of sampling devices relative to the source generator in the Lawrence, KS, field study. (Reproduced by permission of InfoScience Services Corporation.)

5.7 Source Generator

Instead of sampling near an industrial or hazardous waste site, a source generator was used to create a plume of airborne vapors at a pristine site. The plume was generated by using a paint sprayer to project a fine spray of a solvent mixture out the top of a metal tube approximately 15 cm in diameter and 2 m tall. Samples were collected at least 70 m downwind, as shown in Figure 11, where it was expected that mixing of vapors with air would be reasonably thorough. Plume composition was changed between sampling periods.

5.8 Procedure

Participants were required to calibrate their own instruments. PGC sampling began after the TAGA confirmed the arrival of the plume. Since the PGCs had sampling periods of different lengths, starting times were staggered to coincide their mid-points.

5.9 Audit Standards

Audit standards at nominal concentrations of 2.5, 7.5, 15, 25 and 40 ppb of each compound were prepared by dynamic dilution of a commercially prepared, nominally 10 ppm mixture of TO-14 target compounds in nitrogen. Each PGC operator was required, without knowing the concentration in the canister, to analyze at least one audit standard at each level. Data from these analyses were plotted to determine linearity of response and accuracy of calibration. Audit standards were also

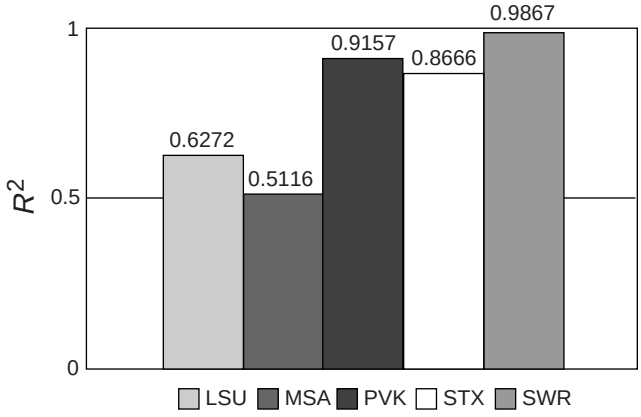

Figure 12 Averages of correlation coefficients for all compounds reported by each participant in the Lawrence, KS, field study. (Reproduced by permission of InfoScience Services Corporation.)

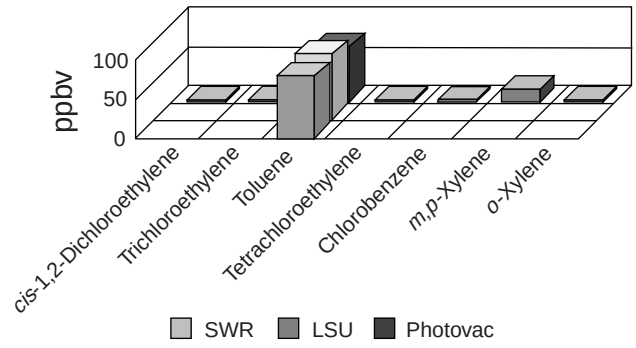

Figure 13 PGC and reference (SWR) data on *cis*-1,2-dichloroethylene (CAS 156-59-2), trichloroethylene, toluene, tetrachloroethylene, chlorobenzene, *m*- and *p*-xylene and *o*-xylene, collected during stable air flow in the Lawrence, KS, field study. (Reproduced by permission of InfoScience Services Corporation.)

submitted to analysis by SWR. The PGCs which relied on preconcentration (LSU, MSA, STX) showed more scatter than the one which did not (PVK). Results for all compounds analyzed by all PGCs and the reference method were summarized in terms of averaged correlation coefficients. Figure 12 displays averages of the correlation coefficient values for scatter plots of all compounds reported by each method. The closer a value approaches unity, the higher the degree of correlation and the better the fit to linear plots. Data produced by MTI and MSA were more scattered than data produced by PVK and STX or by the SWR canister/TO-14 reference method. That difference is consistent with the greater experience of PVK and STX, and the laboratory basis of the analyses performed by SWR. Also, PVK, having an exceptionally sensitive detector, does not preconcentrate samples. Although the precision may be slightly impaired by the additional preconcentration step, it appears to be more affected by operator experience. Analysis of audit standards was the most successful part of the study. There was little to choose between PVK, STX, and SWR with regard to linearity of scatter plots. The lower correlation coefficients registered by MTI and MSA reflect the considerably shorter time that each of them has practised sample collection by preconcentration. They will undoubtedly profit from further experience.

5.10 Plume Sampling

Inclement weather disrupted much of the study. Surface wind varied capriciously, thunderstorms appeared at intervals, and temperatures above 32 °C prevailed between storms. As a result, the reference method often did not function properly. Canister grab samples are no more effective than PGCs for characterizing an unmixed plume. This outcome was disappointing, but not surprising. A study based on use of a plume generator must depend on fair weather, which was forthcoming only sporadically. Analyses performed during periods of relative calm and consistent wind velocity produced data that generally resembled those shown in Figure 13. Analyses done during turbulent weather showed no agreement between portable chromatographs and no agreement with the reference method, as shown in Figure 14. PGC retention time stabilities for *o*-xylene (CAS 95-47-6) are illustrated in Figure 15. PVK retention times were more variable because the operator changed the carrier flow rate before recalibration. When SDs of retention times were calculated separately for runs pertaining to each calibration, they were significantly smaller, and the average of those SDs, displayed in Figure 16, is labeled ''10S Plus Re-averaged''. Detailed reports of the study have been published.[16,17]

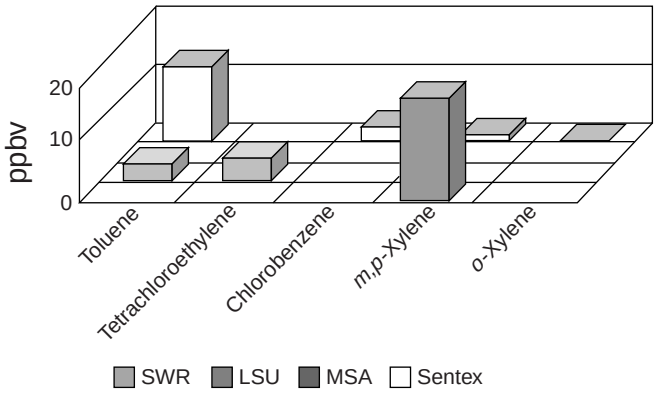

Figure 14 PGC and reference (SWR) data collected during unstable air flow in the Lawrence, KS, field study. (Reproduced by permission of InfoScience Services Corporation.)

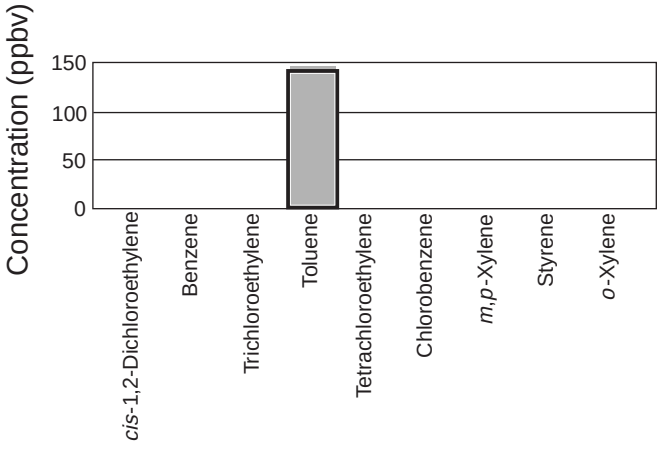

Figure 15 LSU data (solid bar) contrasted with TAGA data (hollow bar) for *cis*-1,2-di-chloroethylene, benzene, trichloroethylene, toluene, chlorobenzene, *m*- and *p*-xylene, styrene (CAS 100-42-5) and *o*-xylene in the Lawrence, KS, field study. (Reproduced by permission of InfoScience Services Corporation.)

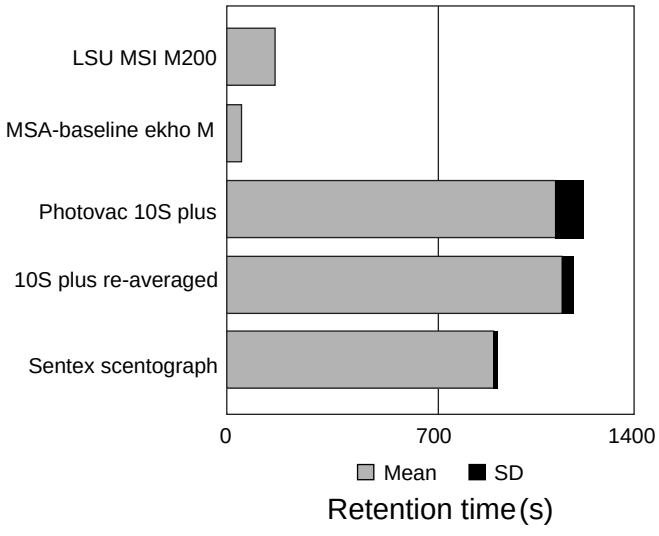

Figure 16 Average retention times and retention time stabilities for *o*-xylene for PGC participants in the Lawrence, KS, field study. (Reproduced by permission of InfoScience Services Corporation.)

6 CONDUCT OF FIELD EVALUATION STUDIES

6.1 Organization

A field evaluation study needs a coordinator to organize the efforts of participants and make necessary arrangements. It requires cooperation among a group of persons which, at minimum, must approximate the largest number of operations to be carried out simultaneously. In field studies where multiple PGCs are compared, the number of them which can be evaluated is limited by the number of operators, which is limited in turn by available shelter. Extensive on-the-spot assistance by vendor technical representatives has proven crucial to the success of field comparison studies. It is best to have PGC vendors participate at their own expense using their own instruments and performing their own calibrations. Vendors gain valuable experience from participation in field studies. They learn how their instruments perform in a field environment, and they find out what competitors can do.

6.2 Site Selection

An ideal site for a comparison study would be located near a source of elevated concentrations of compounds that the participating PGCs can detect, such as high traffic areas, industrial complexes, and hazardous waste sites. Sponsorship by a local property owner or government authority is necessary. Issues of jurisdiction and security for the participants must be satisfactorily managed.

6.3 Existing Sources

A facility manager who suspects that his site emits significant quantities of toxic VOCs will be reluctant to entertain people who are bent on proving it. Neighbors, provided their own houses are in order, may be less reluctant. For a field study, several important vapors should be present at concentrations in the range 5–50 ppb. Such ideal conditions are difficult to find.

6.4 Artificial Sources

An alternative to finding a site with adequate amenities and suitably polluted air is to use a pristine site and pollute it temporarily with a source generator. The advantages are (1) knowing which analytes will be present, (2) being able to control their concentrations, and (3) being exposed to them only during sampling and not at other times. The disadvantages are that there are relatively few compounds which are suitably volatile and sufficiently innocuous to be released into the air at high concentrations, and the forbidden compounds are precisely those which are most interesting.

6.5 Reference Analytical Method

Concurrent collocated grab samples can be collected by holding the inlet valve of the canister adjacent to the PGC intake tubes and opening it slightly so that the canister fills completely during the sampling period of the PGCs. If the sample collection intervals of PGCs differ, then PGC start times can be staggered so that the mid-points of all the sampling periods coincide. A mechanical flow

controller can be used to assure a constant sampling rate for the canister during the longest sampling period. Having all PGCs and the reference method sample air from a common manifold through which air is drawn by a pump could be advantageous, but different PGCs sample at different flow rates during different intervals, so nothing can prevent confusion if concentrations change during a sampling period.

6.6 Audit Standards

Audit standards can be provided in canisters for analysis by each PGC at several concentrations in the 0–50 ppb range. Each participant should analyze each concentration as many times as possible, and the standards should be laboratory-analyzed before and after the study. Accuracy and linearity of response can be assessed from audit data independent of meteorological interferences. Audit standard data reveal problems with instrument and operator performance, including calibration errors, systematic sample collection errors, and poor precision, and they guarantee useful data in spite of bad weather.

6.7 Data Reporting

The coordinator is responsible for receiving data from participants and processing them. The volume of data will be large, and data reporting should be planned in advance. Where possible, data should be submitted as computer files in specified format. Otherwise, they will arrive in a variety of formats. Almost anything is possible, from numbers scribbled on paper towels to spreadsheet files on disk, and the organizational schemes of those files will be limited only by the fertile imaginations of participants. Each will be rational and no two alike. A prudent coordinator will impose uniformity on data reports in advance. Otherwise, unlimited time will be spent teasing order out of chaos. Data from the reference method will not be available until a month or two later. When it comes, the coordinator must try to remember what was done 2 months earlier while matching reference data with PGC data. Having aged 2 months will not render him omniscient.

7 HIGH-SPEED LABORATORY ANALYSIS ALONG WITH FIELD-BASED ANALYSIS

Laboratory GCs using fused-silica capillary columns have defined the state of the art for airborne VOC analysis. They have generally been optimized for maximum chromatographic resolution, so that chromatograms run 30–80 min and display as many peaks as possible. Sample delivery to the instrument, cooling the GC oven between runs, and data processing also take time. A faster method could enhance the usefulness of laboratory data which would not be immediately available but would appear much sooner, and might be partially reported to the field crew while still on-site. Sample storage would be minimized. Occasional system failures further delay data. A reference standard method that produced laboratory data as fast as PGCs would be valuable for evaluation studies and could form part of a dual analytical method more powerful than its laboratory- and field-based components.

7.1 High-speed Gas Chromatography

Gas chromatography is the most widely used analytical technique for analysis of VOCs in ambient air.[18] Starting in 1965, reports began to appear about high-speed or fast gas chromatography (FGC).[19–25] Recently, the design and application of a FGC system to air monitoring has been reported.[26–29] Even specific detectors may not completely eliminate interferences. The duration of a chromatogram can be drastically reduced by increasing column head pressure to speed up carrier velocity (100–200 cm s^{-1}) through a short column. Shortening columns and increasing the carrier flow rate reduces the resolution but need not proportionally decrease the separation. Injecting the sample in a narrower bandwidth (<10 ms) produces peaks which are narrower and therefore better separated at the same resolution.[27,28,30–32] On the other hand, many GC detectors cannot collect data rapidly enough to describe such narrow peaks accurately, so special fast detectors must be used. Also, increasing carrier gas velocity reduces GC column capacity and limits its sample size. On the other hand, narrower peaks are taller and thus more detectable if the detector can trace them accurately. The difficulties of reproducibly performing narrow bandwidth injections and detecting narrow peaks have long inhibited use of FGC, but recently, a reliable narrow bandwidth injection system has been marketed by Chromatofast Inc., and various kinds of fast detectors have become available. Chromatograms of maximum duration are no longer necessarily optimal. Indeed, lengthening chromatograms to maximize chromatographic resolution and minimize interferences between analytes and background hydrocarbons is futile.

7.2 High-speed Mass Spectrometry

By most criteria, mass spectrometers would be ideal GC detectors. They are both universal and specific. In most cases they can identify analytes unequivocally, even in the presence of interferences. Their output can be searched using mass spectral databases, and they can be used for quantitative analysis. The problem has always been that mass spectrometers were too slow to be compatible with

capillary GCs. Mass spectral scans, until recently, could be done perhaps three times per second without undue sacrifice of mass spectral resolution, unless the scan range was narrow. Collimation to focus and improve mass spectral resolution wasted much of the ion beam and curtailed response. Early-eluting capillary GC peaks are typically about 1 s wide and could not be adequately traced by the three data points that a mass spectrometer scanning at maximum practical speed could produce. It was necessary to sacrifice chromatographic resolution in order to ''see'' early peaks. The marriage of these two powerful analytical techniques, although successful, was not a happy one.

Recently, Leco Corporation has marketed the Pegasus II, a TOFMS that was originally designed by Holland and co-workers specifically as a detector for FGC.[33–35] A schematic diagram of the high-speed TOFMS is shown in Figure 17. It can acquire and record up to 500 mass spectra per second. Use of prototype high-speed TOFMSs has been demonstrated for a variety of analytical applications.[35–38] A TOFMS differs from other mass spectrometers in that it does not ''scan''. Instead, the ions are ejected from the ion source periodically by a positive pulse which projects them along the axis of a ''drift tube''. Instead of a series of tiny collimation apertures, they pass through a hole about 2 cm in diameter at the entrance to the drift tube. At the far end they encounter a positive electrical field or ''reflectron'' which reflects them back at a small angle toward an array detector located near the ion source. Ions having more or less than average velocities along the axis of the drift tube penetrate the reflectron to greater or lesser depths, respectively, so that all ions of the same mass emerge from the reflectron at the same time with their thermal energy spread minimized and their mass spectral resolution enhanced. Ions

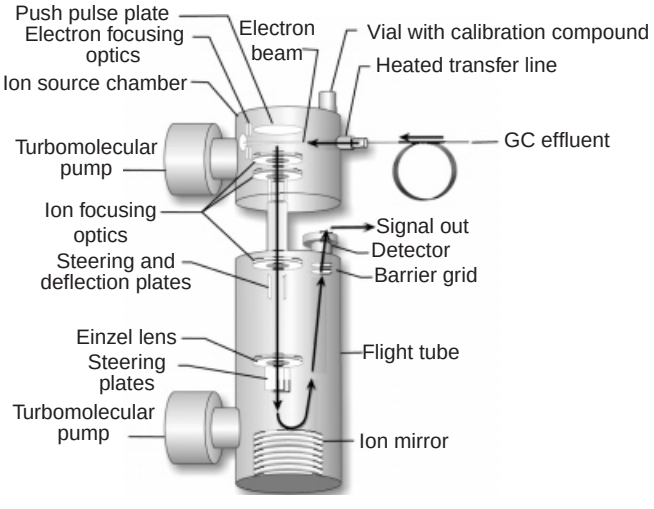

Figure 17 Schematic diagram of the Leco Pegasus II TOFMS. (Reproduced by permission of Leco Corporation.)

are pulsed from the ion source at a constant rate of 5 kHz, and the resulting mass spectra are averaged before being written to disk at an operator-selected rate up to 500 Hz. Each mass spectrum written to disk is an average of 10 or more individual spectra, depending on how fast they are recorded. The Pegasus II is fast enough to cope with the output of any FGC. Since all the ions forming each mass spectrum are ejected from the ion source simultaneously, the individual mass spectra are not biased by changes in concentration in the ion source which occur during the time taken by an ordinary mass spectral scan. Averaged spectra recorded at more than 30 Hz are practically free from bias, and peak deconvolution, being much simplified, can be done with simple algorithms, which makes it possible to process GC/MS data in real time.

Occasionally TOFMSs have been adapted to perform as GC detectors without benefit of supporting software. The volume of raw data produced by FGC coupled with a TOFMS is enormous. To process manually data acquired in 1 day into publishable information can take months. The Pegasus II incorporates software which controls the GC and TOFMS system, records and processes data, and compiles reports. Data are written to disk in real time and data processing takes about as long as acquisition and may be done concurrently.

7.3 High-speed Gas Chromatography/Mass Spectrometry Analysis

The combination of a modified Chromatofast injection accessory with a Varian 3400 high-speed GC and a Leco Pegasus II TOFMS has been used to demonstrate rapid analysis of whole air grab samples collected in 6-L passivated canisters by a procedure analogous to USEPA Method TO-14.[38] The Chromatofast withdrew 3 mL min^{-1} of air from a sample stream. During a period of 40 s, 2 mL of air were passed through the cryotrap. Then the trap was heated by capacitive discharge to inject the sample in a narrow bandwidth (<10 ms) on to a 12 m × 0.25 mm column coated with 1.4 μm of DB-624. The turn-around time between injections was less than 10 min. The GC oven was programmed from 50 to 170 °C at 35 °C min^{-1}. About 200 s were required for all compounds on the TO-14 target list to elute. Individual ion chromatograms for seven ions from the same chromatogram are shown in Figure 18.

8 CONCLUSION

8.1 Present State of Field-based Analysis

On-site analysis is presently in a state of rapid development. A wide variety of technical approaches

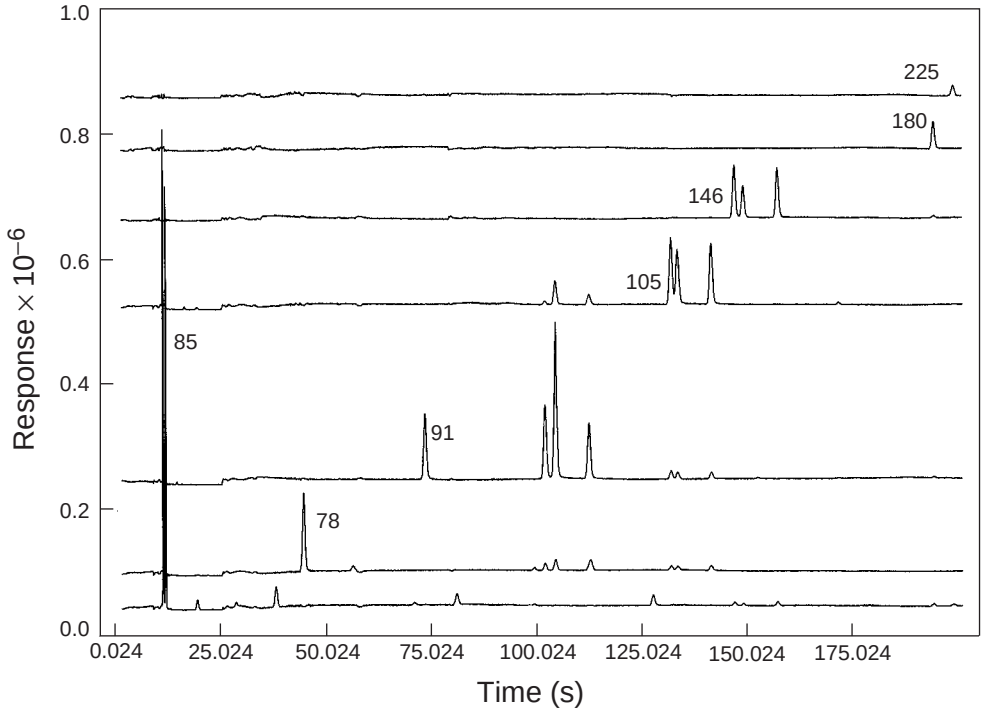

Figure 18 Ion chromatograms for dichlorodifluoromethane (CAS 75-71-8) (*m*/*z* 85), benzene (*m*/*z* 78), toluene and xylenes (*m*/*z* 91), 4-ethyltoluene (CAS 622-96-8), 1,3,5-trimethylbenzene (CAS 108-67-8), 1,2,4-trimethylbenzene (CAS 95-63-6) (*m*/*z* 105), *m*-dichlorobenzene (CAS 541-73-1), *p*-dichlorobenzene (CAS 106-46-7) and *o*-dichlorobenzene (CAS 95-50-1) (*m*/*z* 146), 1,2,4-trichlorobenzene (CAS 120-82-1) (*m*/*z* 180) and hexachlorobutadiene (CAS 87-68-3) (*m*/*z* 225).

have been applied. Skilled and knowledgeable operators are needed for proper operation of most instruments. Comparison with laboratory analysis remains the standard of PGC performance for the time being. Individual instruments are generally limited to specific classes of compounds, and complete characterization of a complex site may require the use of several types. The use of PGCs to pilot collection of samples for laboratory analysis is becoming accepted, but total reliance upon PGCs for site characterization awaits more extensive and reassuring experience.

Laboratory- and field-based methods of analysis can complement each other. The widely held perception that laboratory-based analyses are more ''accurate'' because they are done in a stable environment using state of the art equipment is misleading. The futility of accurately analyzing samples flawed by blind collection or lengthy storage is too seldom considered. Likewise, enthusiasts for field-deployable analysis can be tempted to ignore the effects of operating in environments where even laboratory equipment could not perform properly. Ideally, fast laboratory- and field-based methods should be used together.

8.2 Future Prospects

Reliability of data obtained from on-site instruments is not yet generally accepted, but instruments are becoming more sophisticated and user-friendly. As the quality of data continues to improve and consistent performance continues to be demonstrated, on-site data will be increasingly accepted. Eventually on-site data will become the standard by which the integrity of samples analyzed in the laboratory after prolonged storage will be judged. Two principal benefits are expected from faster analyses. The first is quicker access to data. The second is a reduced sample storage time, resulting in better sample integrity and quicker cycling of sample containers. Both laboratory- and field-based analyses have important advantages and disadvantages. Used together, their respective disadvantages cancel, while their advantages reinforce each other.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
AID	Argon Ionization Detector
DC	Direct Current
DL	Detection Limit
ECD	Electron Capture Detector
ELCD	Electrolytic Conductivity Detector
FGC	High-speed or Fast Gas Chromatography

FID	Flame Ionization Detector
FPD	Flame Photometric Detector
FUV	Far-ultraviolet Detector
GC	Gas Chromatograph
GC/MS	Gas Chromatograph/Mass Spectrometer
GC/MSD	Gas Chromatography with Mass Selective Detection
HID	Helium Ionization Detector
JCA	Junta de Calidad Ambiental
LSU	Louisiana State University Institute for Environmental Studies
MSA	MSA-Baseline
MSD	Mass Selective Detector
MTI	Microsensor Technology
NPD	Nitrogen–Phosphorus Detector
PGC	Portable or Field-deployable Gas Chromatograph
PID	Photoionization Detector
PVK	Photovac
SAW	Surface Acoustic Wave
SD	Standard Deviation
STX	Sentex
SWR	Southwest Research Institute
TAGA	Mobile Triple Quadrupole Mass Spectrometer
TCD	Thermal Conductivity Detector
TID	Thermionic Ionization Detector
TOFMS	Time-of-flight Mass Spectrometer
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
XSD	Halogen-sensitive Detector

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Selective Detectors

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Field-portable Instrumentation cont'd **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry • Time-of-flight Mass Spectrometry

REFERENCES

1. 'Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air', EPA-600/4-84-017, Environmental Protection Agency, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC, June 1988.
2. P. Warneck, *Chemistry of the Natural Atmosphere*, Academic Press, San Diego, CA, 270, 1988.
3. K.D. Oliver, J.D. Pleil, W.A. McClenny, 'Sample Integrity of Trace Level Volatile Organic Compounds in Ambient Air Stored in Summa-polished Canisters', *Atmos. Environ.*, **20**, 1403–1411 (1986).
4. A.R. Gholson, J.F. Storm, R.K.M. Jayanty, R.G. Fuerst, T.J. Logan, M.R. Midgett, 'Evaluation of Canisters for Measuring Emissions of Volatile Organic Air Pollutants from Hazardous Waste Incineration', *J. Air Pollut. Control. Assoc.*, **39**, 1210–1217 (1989).
5. B.B. Kebbekus, G. Dai, W. Ji, C.-H. Liao, L.-C. Hung, 'Methods for Analysis of VOC in Ambient Air: Comparison between Tenax and Canister Collected Samples', in *Proceedings of the 1991 US EPA/AWMA International Symposium, Durham, NC*, Air and Waste Management Association, Pittsburgh, PA, 255–260, 1991.
6. R.E. Berkley, 'Field Evaluation of Photovac 10S50 Portable Photoionization Gas Chromatograph for Analysis of Toxic Organic Pollutants in Ambient Air', EPA/600/D-88/088, Environmental Protection Agency, Research Triangle Park, NC, 1988.
7. R.E. Berkley, J.L. Varns, W.A. McClenny, J. Fulcher, 'Field Evaluation of Photovac 10S70 Portable Gas Chromatograph', in *Proceedings of the 1989 EPA/AWMA International Symposium on Measurement of Toxic and Related Air Pollutants*, Air and Waste Management Association, Pittsburgh, PA, 19–26, 1989.
8. R.E. Berkley, J.L. Varns, J. Pleil, 'Comparison of Portable Gas Chromatographs and Passivated Canisters for Field Sampling Airborne Toxic Organic Vapors in the United States and the USSR', *Environ. Sci. Technol.*, **25**(8), 1439–1444 (1991).
9. R.E. Berkley, K.D. Oliver, K. Kronmiller, 'Performance Optimization of Photovac 10S70 Portable Gas Chromatograph', in *Proceedings of the 1990 EPA/AWMA International Symposium on Measurement of Toxic and*

Related Air Pollutants, Air and Waste Management Association, Pittsburgh, PA, 849–854, 1990.
10. R.E. Berkley, 'Evaluation of Emission Sources and Hazardous Waste Sites Using Portable Chromatographs', in *Proceedings of the Second International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, ICAIR Life Systems, Cleveland, OH, 253–262, 1991.
11. R.E. Berkley, 'Overview of Field-deployable Gas Chromatographic Analysis of Airborne Toxic Organic Vapors', *J. On-site Anal.*, **2**, 313–320 (1998).
12. S.L. Rose-Pehrsson, J.W. Grate, D.S. Ballantine, Jr, P.C. Jurs, 'Detection of Hazardous Vapors Including Mixtures Using Pattern Recognition Analysis of Responses from Surface Acoustic Wave Devices', *Anal. Chem.*, **60**, 2801–2811 (1988).
13. R.E. Berkley, 'Evaluation of Photovac 10S50 Portable Photoionization Gas Chromatograph for Analysis of Toxic Organic Pollutants in Ambient Air', EPA/6004-86/041, PB87-132858, Environmental Protection Agency, Research Triangle Park, NC, 1986.
14. R.E. Berkley, M. Miller, J.C. Chang, K. Oliver, C. Fortune, J. Adams, 'Evaluation of Portable Gas Chromatographs', in *Proceedings of the 1993 EPA/AWMA International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Air and Waste Management Association, Pittsburgh, PA, 682–693, 1993.
15. R.E. Berkley, M. Colón, J. González, I. Droz, 'Field Comparison of Portable Gas Chromatographs with Method TO-14', *J. On-site Anal.*, **1**, 161–170 (1994).
16. R.E. Berkley, M. Davis, M. Hansen, D. Lane, 'Performance Comparison of Field-deployable Gas Chromatographs with Canister/TO-14 Analyses', in *Proceedings of the 1996 EPA/AWMA International Conference on Measurement of Toxic and Related Air Pollutants*, Air and Waste Management Association, Pittsburgh, PA, 707–719, 1996.
17. R.E. Berkley, M. Davis, M. Hansen, D. Lane, 'Design of a Field Comparison Study for Portable Gas Chromatographs', *J. On-site Anal.*, **3**, 88–104 (1998).
18. J.D. Pleil, K.D. Oliver, W.A. McClenny, 'Ambient Air Analyses Using Nonspecific Flame Ionization and Electron Capture Detection Compared to Specific Detection by Mass Spectroscopy,' *J. Air Pollut. Control. Assoc.*, **38**, 1006–1010 (1988).
19. D.H. Desty, A. Goldup, W.T. Swanton, 'Performance of Coated Capillary Columns,' in *Gas Chromatography*, eds. N. Brenner, J.E. Callen, M.D. Weiss, Academic Press, New York, 105–135, 1962.
20. D.H. Desty, 'Capillary columns: Trials, Tribulations and Triumphs,' in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, 199–228, Vol. 1, 1965.
21. J.C. Sternberg, 'Extra Column Contributions to Chromatographic Band Broadening,' in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, 203–270, Vol. 2, 1966.
22. G. Gaspar, P. Arpino, G. Guiochon, 'Study in High Speed Gas Chromatography,' *J. Chromatogr. Sci.*, **15**, 256–261 (1977).
23. G. Gaspar, R. Annino, C. Vidal-Madjar, G. Guiochon, 'Influence of Instrumental Contributions on the Apparent Column Efficiency in High Speed Gas Chromatography', *Anal. Chem.*, **50**, 1512–1518 (1978).
24. C.P.M. Schutjes, E.A. Vermeer, J.A. Rijks, C.A. Cramers, 'Increased Speed of Analysis in Isothermal and Temperature Programmed Capillary Gas Chromatography by Reduction of the Column Inner Diameter,' *J. Chromatogr.*, **253**, 1–16 (1982).
25. A. van Es, J. Janssen, R. Bally, C. Cramers, J. Rijks, 'Sample Introduction in High Speed Capillary Gas Chromatography: Input Band Width and Detection Limits,' *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **10**, 273–279 (1987).
26. R.F. Mouradian, S.P. Levine, R.D. Sacks, 'Evaluation of a Nitrogen-cooled, Electrically Heated Cold Trap Inlet for High Speed Gas Chromatography,' *J. Chromatogr. Sci.*, **28**, 643–648 (1990).
27. R.F. Mouradian, S.P. Levine, R.D. Sacks, M.W. Spence, 'Measurement of Organic Vapors at sub-TLV Concentrations Using Fast Gas Chromatography,' *Am. Ind. Hyg. Assoc. J.*, **51**(2), 90–95 (1990).
28. R.F. Mouradian, S.P. Levine, H. Ke, 'Measurement of Volatile Organics at Part per Billion Concentrations Using a Cold Trap Inlet and High Speed Gas Chromatography,' *J. Air Pollut. Control. Assoc.*, **41**, 1067–1072 (1991).
29. H. Ke, S.P. Levine, R.F. Mouradian, R. Berkley, 'Fast GC for Environmental and Industrial Health Air Monitoring,' *Am. Ind. Hyg. Assoc. J.*, **53**, 130–137 (1992).
30. H. Ke, S.P. Levine, R.F. Mouradian, R.E. Berkley, 'High Speed Chromatography for Air Monitoring,' in *Proceedings of 1991 EPA/AWMA International Symposium on Measurement of Toxic and Related Air Pollutants*, Air and Waste Management Association, Pittsburgh, PA, 558–563, 1991.
31. H. Ke, S.P. Levine, R.F. Mouradian, R.E. Berkley, 'Fast Gas Chromatography for Air Monitoring: Limits of Detection and Quantitation', *Am. Ind. Hyg. Assoc. J.*, **53**, 130–137 (1992).
32. H. Ke, S.P. Levine, R. Berkley, 'Analysis of Complex Mixtures of Vapors in Air by Fast Gas Chromatography', *Air Waste Manag. Assoc. J.*, **42**, 146–162 (1992).
33. R. Grix, R.E. Tecklenburg, Jr, J.T. Watson, J.F. Holland, G.E. Yefchak, B.D. Gardner, J. Allison, C.G. Enke, H. Wollnik, 'A Time-of-Flight Mass Spectrometer for High-speed Chromatography', in *Proceedings of 39th ASMS Conference on Mass Spectrometry and Allied*

Topics, Nashville, TN, May 19–24, American Society for Mass Spectrometry, Santa Fe, NM, 154–155, 1991.

34. B.D. Gardner, J.F. Holland, 'High Speed, High Throughput GCMS Using Time-of-Flight Mass Spectrometry with Time Array Detection', in *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics, Portland, OR, May 12–16*, American Society for Mass Spectrometry, Santa Fe, NM, 1119, 1996.
35. J. Grainger, V. Green, J. Barr, C. McClure, Z. Liu, D.G. Patterson, Jr, J.F. Holland, B.D. Gardner, 'Fast Gas Chromatography/High Resolution Mass Spectrometry and Fast Gas Chromatography/Time-of-Flight Mass Spectrometry Analysis of PCBs, PAAHs, CDDs and Threat Assessment Standards', in *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics, Portland, OR, May 12–16*, American Society for Mass Spectrometry, Santa Fe, NM, 860, 1996.
36. B.D. Gardner, R.E. Berkley, J.F. Holland, 'Rapid Analysis of TO-14 Airborne Contaminants by Fast GC/TOFMS', in *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics, Portland, OR, May 12–16*, American Society for Mass Spectrometry, Santa Fe, NM, 513, 1996.
37. J. Song, B.D. Gardner, J.F. Holland, R.M. Beaudry, 'Rapid Analysis of Volatile Flavor Compounds in Horticultural Produce Using SPME and GC/Time-of-Flight Mass Spectrometry', in *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics, Portland, OR, May 12–16*, American Society for Mass Spectrometry, Santa Fe, NM, 547, 1996.
38. R.E. Berkley, B.D. Gardner, J.F. Holland, 'Analysis of Method TO-14 Target Analytes Using a Cryofocusing High-speed Gas Chromatograph Interfaced to a High-speed Time-of-Flight Mass Spectrometer', in *Proceedings of the 1997 EPA/AWMA International Symposium on Measurement of Toxic and Related Air Pollutants*, Air and Waste Management Association, Pittsburgh, PA, 346–355, 1997.

Field-portable Instrumentation

Michel G. Goedert
Hewlett-Packard Laboratories, Palo Alto, USA

1 Introduction 2
2 Techniques of Collection for Vapors and Gaseous Samples 3
2.1 Regulations and Government Agencies 3
2.2 Ambient Air Quality Standards 3
2.3 Atmospheric Dispersion Modeling 3
2.4 Classification of Contaminants 4
2.5 Grab Sampling 4
2.6 Integrated Sampling 6
3 Direct-reading Instruments for Environmental Measurements 11
3.1 Introduction to Direct-reading Instruments 11
3.2 Chemical Gas Sensors 12
3.3 Detector Tubes and Other Colorimetric Direct-reading Indicators 23
3.4 Ionization Detectors 25
4 Gas Chromatography for Environmental Measurements 29
4.1 Basic Components of a Gas Chromatograph 29
4.2 Classification of Gas Chromatographs Used for Field Measurements 30
4.3 Categories of Analysis 32
4.4 Trace Analysis of Gaseous Samples 32
4.5 Solid Extraction Thermal Desorber Unit 33
4.6 Transfer Lines 33
4.7 Airborne Pollutants 33
4.8 Collection and Analysis of Airborne Pollutants 34
4.9 Theory of Packed and Fused-silica Capillary Column Operation 35
4.10 Parameters Influencing a Separation 38
4.11 Computer Optimization of the Separation 40
4.12 Sample Injection Devices 42
4.13 Types of Capillary-column Inlets 44
4.14 Gas Chromatography Detectors – Basic Terminology 47
4.15 Gas Chromatography Detector Selection 47
4.16 Portable and Transportable Gas Chromatographs 53
5 Mass Dispersion Methods for Environmental Measurements 59
5.1 Mass Spectrometry 59
5.2 Mass Analyzers 60
5.3 Portable Gas Chromatography/Mass Spectrometry Detectors 65
5.4 Ion Mobility Spectrometry 68
Abbreviations and Acronyms 77
Related Articles 78
References 79

The analysis of gas and vapors in the field is an important part of the remediation process. The purpose of this article is to familiarize the reader with some of the analytical instruments available today to determine the composition of gaseous samples. Although this section deals with portable instruments, it was found necessary to mention on-site measurements since this is an important part of field measurements. On-site measurements are done with conventional analytical instruments operating from recreational vehicles. The samples have to be collected and the first part of this article gives an overview of some of the techniques, such as canister and solid adsorbents, used for this purpose. Identification of compounds collected on-site or directly in the field can be done by gas chromatography (GC) and mass spectrometry (MS). A comprehensive description of the hardware used with these techniques is provided.

An important part of the portable section covers direct-reading instruments such as gas sensors, electrochemical devices, ionization detectors and colorimetric detectors. These instruments provide only a limited amount of information and, most of the time, require a prior knowledge about the content of the sample. However, they are small, portable and cheaper than laboratory analytical instruments and therefore they have found a market.

Portable instruments include miniaturized gas chromatographs and mass spectrometers. These instruments do not offer the high performance of their laboratory counterparts, but they give more information about the sample than direct-reading instruments. The next step in process measurement sophistication is to combine these techniques. The most popular hyphenated technique is certainly the coupling of a gas chromatograph with a mass spectrometer, but there are also powerful and lesser-known techniques such as GC combined with ion mobility spectrometry (IMS). IMS has found a market with military applications for the detection of chemical warfare agents. It offers less resolution than MS but operates at atmospheric

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

pressure and is therefore truly portable. When combined with GC it offers some interesting features which will be described.

1 INTRODUCTION

In recent years, the area of environmental laws governing emissions has been one of the fastest-growing areas of legislation. Its primary aim has been to protect the environment by exercising control over emissions from industrial processes. From country to country, different agencies monitor compliance, and failure to meet the standards can result in prosecution and heavy penalties. Sources of emissions can be broadly split into two areas: mobile (vehicles) and stationary (plants). The volume of legislation governing automotive exhaust emissions has been on the increase for some time. Most of this legislation involves the measurement of oxygen, carbon monoxide, carbon dioxide, unburned hydrocarbons, and, to some extent, nitrogen oxides. An instrument used to monitor mobile sources for field measurements has some requirements: it must be small and light, use a minimum of consumable fuel (electrical power and gases), and operate for hours, even days, without need of recharging.

Of primary interest among stationary sources of emissions are the emissions from flues or stacks. The composition of flue gas depends on the fuel or chemical process causing the emission, but the main components are emitted as a result of fossil fuel burning. The gases to be analyzed consist mostly of oxygen, unburned hydrocarbons, oxides of carbon, sulfur, and nitrogen in varying concentrations. The majority of legislation governing emissions is concerned with carbon monoxide, carbon dioxide, sulfur dioxide, nitric oxide, and nitrogen dioxide.

In many instruments designed specifically to monitor a single chemical, interferences are minimized, and concentrations within the healthy range – permissible exposure limit (PEL) and threshold limit value (TLV) – can be reliably and accurately measured. When used correctly and appropriately, measurements made with these dedicated instruments can often be used in lieu of traditional instruments, such as gas chromatographs.

Generally, instruments dedicated to a single chemical are less expensive and easier to use in the field than are instruments that detect many different chemicals, such as GC and broad-spectrum infrared (IR) units. The instruments with broader measurement capabilities can also be dedicated to the measurement of a single chemical, but this is generally not cost-effective. IR instruments dedicated to the measurement of a single chemical are constructed more simply than those with broader monitoring capabilities. One available variation is a single-instrument housing that can be adapted to monitoring different chemicals by replacement of the sensors. The choice of instrument depends on the situation to be monitored; in general, for atmospheres for which the types of contaminants are well characterized, the more specific the monitor the better. On the other hand, with a number of unknowns the more appropriate instrument will be a nonspecific general-survey instrument.

The intent of this article is to provide a guide in choosing non-optical techniques used for the analysis of gaseous samples. The most commonly available technologies are described here, along with descriptions of commercial products using those technologies in practical applications, if available. Most of these techniques are mature, and their theoretical bases well established. On the other hand, the hardware used for field measurements is under continuous development, so information is valid only up to the time of writing. It is important to check the current state-of-the-art information in specialized journals and contact vendors to get updated information about products.

Measurements can be performed off-line by collecting the samples and analyzing them later, generally in the laboratory. Several approaches are described in detail. Sample collection can be discrete (grab sampling, use of canister) or done over time (solid sorbent, impingers).

When the decision is made to take a field measurement, the proper analytical technique must be selected. Although many approaches are available, the final decision is based on what is already known about the sample. If the components of the gaseous sample are known and the user wants to determine only the presence and the concentration level of the analytes, the use of cartridge tubes or any other technique based on colorimetry is perfectly acceptable. These tubes are inexpensive, sensitive, and give visual confirmation of the level of the gases present. However, they are irreversible and cannot be used as permanent monitoring devices. Electrochemical sensors are used when an unknown element of the gaseous sample must be determined. These sensors, described below, are relatively selective and sensitive enough to detect gas or combinations of gases at the sub-parts per million level. They can be operated in an alarm (monitoring) condition and are reversible.

Portable monitors, such as the flame ionization detector (FID) and photoionization detector (PID), are also described; they give some useful information regarding the family of compounds of the analytes. They are sensitive and reliable but provide only a little information about the nature of the sample itself. They are easy to use and give the go/no-go information necessary for a quick decision. Several commercial units are described.

The next category of measurements uses instruments that separate the components of the mixture, at least

partially, before detecting them and thus allows quantification of the sample components. Gas chromatographs fall into this category. GC is a mature technology that separates the various constituents of a gaseous mixture before detecting them. GC is not an absolute method of identification.

Since GC identification is based only on the comparison of retention times and not on the physical properties of the analytes, multiple compounds may be eluting at the same time and not be identified. Therefore, it is not possible to detect with 100% certainty all the components of a totally unknown sample. However, GC provides a reasonable amount of information and is so convenient that it is commonly used. It is fast, inexpensive, easy to use, and, when combined with other analytical techniques, such as MS, provides an accurate method of identification. The components of a gas chromatograph, from the injection to the separation to the detection of the sample, are described. Portable units are becoming smaller, more sophisticated, and more versatile. However, the quality of data provided with portable units is still significantly lower than the quality of measurements obtained with laboratory instruments. Some of the portable units available will be described, and the pros and cons of those techniques will be identified.

The next category of measurement is mass dispersion methods. Two of these methods are discussed: MS and IMS. IMS was developed to detect toxic gases in chemical warfare applications. IMS is more qualitative than quantitative. IMS is sensitive (parts per billion level), fast (within seconds), and ideal for detecting suspected gases, but it is not as good at identifying totally unknown mixtures. IMS instruments are portable and easy to use, even by unskilled operators. Since interest in this technique is growing for civil applications, it will be described here. MS, especially combined with GC separation, is the most powerful but also the most expensive and difficult to use analytical tool available. MS provides positive identification, even in complex situations. It is fast (<1 s) and accurate. Its dynamic range is large (in decades) and sensitive. This technique is now becoming portable, thanks to the technological improvements of portable vacuum technics (better material, turbomolecular pumps). It is usually coupled with GC to simplify the interpretation of spectra.

2 TECHNIQUES OF COLLECTION FOR VAPORS AND GASEOUS SAMPLES

2.1 Regulations and Government Agencies

The analysis of air, soil, and water samples is still, and will continue to be, an important component of the overall effort to resolve the environmental problems associated with undesirable and unhealthy chemicals. Environmental laws and regulations and government agencies are obviously major participants in the activities related to the protection of the environment.[1] Any analytical work performed in conjunction with a federal or state regulation or government agency must conform to the guidelines that it provides. A working knowledge of the guidelines and of the capabilities of the various analytical techniques available is essential for any analyst involved in environmental analysis.[2]

At the federal level, the United States Environmental Protection Agency (USEPA) is the most recognized regulatory organization. From the regulatory perspective, the primary function of the USEPA is to interpret laws passed by Congress and to formulate operating practices. Other prominent federal agencies, such as the Department of Energy (DOE), the US Army Corps of Engineers (COE), the Department of the Interior (DOI), the Department of Labor (DOL), the Occupational Safety and Health Administration (OSHA) and the National Institute of Occupational Safety and Health (NIOSH) are also charged with the protection of the environment. Detailed information on a particular government agency or regulation may be obtained directly from the agency.

2.2 Ambient Air Quality Standards

At the federal level, the quality of the air is measured by comparing pollutant concentrations with the National Ambient Air Quality Standards (NAAQS). At the state and local level, it is measured by comparing the pollutants with NAAQS, as well as with state ambient air quality standards (SAAQS) and local ambient air quality standards, if any. Recent legislation (the 1977 and 1990 federal Clean Air Act amendments) has added other parameters, such as esthetics (visibility) and risk (air toxicity), to the evaluation of air quality.

2.3 Atmospheric Dispersion Modeling

Modeling is used to determine what happens to pollutants emitted into the atmosphere. When a pollutant is emitted, it mixes with the existing air, and its concentration decreases because of atmospheric turbulence. This process is termed *atmospheric dispersion*. Most of the point-source atmospheric models that are used in the regulatory process are termed *Gaussian dispersion* models. Non-point-source models are used to access regional pollutants (e.g. urban air shed modeling) or vehicle emissions (e.g. freeway or traffic impact modeling). Summaries of preferred air quality models such as SCREEN, SCREEN3, and ISC have been described.[3]

More information can be found in the USEPA guidelines and model handbooks.

2.4 Classification of Contaminants

Contaminants may be divided, depending on physical characteristics, into gases, vapors, and particulate matter. The terms *gases* and *vapors* are often used interchangeably, although they are not identical. At ordinary temperature and pressure, both gases and vapors will diffuse rapidly. Gases are generally noncondensable at room temperature, whereas vapors are considered to be condensable, since they are derived from volatile liquids. Therefore, under ordinary conditions, gases remain in the gaseous state even when present at high concentrations. Vapors, on the other hand, may condense at high concentrations and coexist in both gas and aerosol forms. Particulate matter can be further differentiated as dust, fumes, smoke and soot, and liquid particles. Dusts are solid particles formed from inorganic or organic materials reduced in size by mechanical processes. These particles range in size from the visible to the submicroscopic, but the principal concern of industrial hygienists is with those under 10 μm in size, because such particles stay suspended in the atmosphere for a long period of time and can reach the deepest parts of the lungs.

Fumes are fine particles formed from solid materials by evaporation, condensation, and gas-phase molecular reactions. These particles range in size from 0.1 to <0.01 μm. Smoke and soot are products of incomplete combustion of organic materials. The size of smoke particles is usually <0.5 μm. Liquid particles are produced by condensation and atomization from the gaseous state. Droplets formed from atomization are usually >5 μm in diameter.

There are two methods for collecting gaseous samples. In grab sampling, an actual sample of air is taken in a flask, bottle, bag, or other suitable container; in continuous or integrated sampling, gases or vapors are removed from the air and concentrated by passage through an absorbing or adsorbing medium. Grab sampling usually involves the collection of instantaneous or short-term samples, normally within a few seconds or minutes. This type of sampling is acceptable when peak concentrations are sought or when concentrations are relatively constant. An important feature of grab samples is that their collection efficiency is normally 100%. However, sample decay does occur for various reasons, such as reaction or adsorption on the inner surfaces of the collector.

Grab sampling is of questionable value when the concentration of the contaminant varies with time or is low, when the measurement device is barely sensitive enough to measure a sample, or when a time-weighted average (TWA) exposure is desired.

Integrated sampling consists of extracting the gas or vapor and concentrating it by solution in an absorbing liquid, reaction with an absorbing solution, or collection on a solid adsorbent. Collection efficiency is frequently $<100\%$; therefore, individual efficiency percentages must be determined for each case.

Guidelines for the selection of sampling devices and analytical procedures are available.[4–6] Secondary sources are also available in books[7] or in journals such as the *American Industrial Hygiene Association Journal* and *Applied Occupational and Environmental Hygiene*.

Generally, nonreactive and nonabsorbing gaseous substances may be collected as grab samples. Gases and vapors that are water-soluble or that react rapidly with absorbing solutions can be collected in simple gas-washing bottles.

Gaseous substances that are volatile and less soluble or that react slowly with absorbing solutions require more liquid contact and may require more elaborate sampling devices. Insoluble and nonreactive gases and vapors are collected by adsorption on activated charcoal, silica gel, or another suitable adsorbent.

2.5 Grab Sampling

The preparation steps required for trace analysis of most air or gaseous samples differ from those required for solid or aqueous samples in that they are performed partially in the field; the gaseous samples themselves are never really transported to the laboratory. The organic contaminants contained in these samples are usually extracted or removed from the gaseous matrix and concentrated during the sampling procedure.

In its simplest design, a measured volume of the sample is drawn into the sampling system or directly through the measuring instrument by some type of pump. The simplest sampling system consists of only a collection medium. The analytes are retained on or in the collection medium, which is transported to the laboratory for preparation and analysis.

Gaseous samples can be collected in rigid glass or metal flasks, in soft plastic bags, or in combinations of these with aluminum foil in sizes up to 120 L. For plastic bag sampling, the sample is introduced into the bag by a hand- or battery-operated pump or a squeeze bulb. Pyrex gas-collecting tubes of 300 mL capacity with a capillary standard taper stopcock at each end may be used when the atmosphere sampled contains components incompatible with plastic bag materials. The flask may be evacuated in the laboratory and then opened in the environment to be sampled, or the flask may be evacuated in the field with a pump. Specially passivated stainless-steel pressure canisters are used for short-period sampling of reactive gases and vapors.

Figure 1 Plastic bag used for the collection of gaseous samples. This bag can also be used to prepare calibration samples by injecting a known quantity of chemicals with a syringe. (Courtesy of Calibrated Instruments Inc., Hawthorne, NY.)

2.5.1 Sample Bags

When low detection limits are not required, sample bags are used to collect gaseous samples and are transported to the laboratory for analysis without intermediate preparation steps. Transportation of gaseous samples in this manner is useful only when the analytes of interest are stable and when other methods of collection are not applicable. For instance, bags are used when sorbents have a low breakthrough volume or when collection of TWA samples is required.[8] These bags can be made of plastics but are more often made of materials such as Tedlar, Mylar, Scotchpak, and Teflon®. Figure 1 shows a plastic bag equipped with a septum injector.

The sampling apparatus consists of a 20-L (5-gal) open-top container with a sealable lid, a Teflon® stopcock, Teflon® bulkheads, Teflon® tubing, and a sampling pump.[8] The bag is attached to a low-flow exhaust outlet and to an adjustable pump. The sampling period determines the size of the sample bag and the flow rate of the pump. Most bags have a septum for injection using a syringe.

Personal bags are often used to collect gases and vapors for high-resolution analysis of environmental samples and occupational sampling. This method is commonly used for collecting air samples in open fields and from vapor wells. Bulk air samples are collected in small bags carried by an operator. Personal sampling requires that the collection device be small and light. A pump and a battery are provided with the device and must be carried away by the operator. Sampling bags are used when other methods are not applicable, for instance, when sorbent tubes have a low breakthrough volume or when there is no alternative method. Sampling bags are also used when a mixture of gases is incompatible. Sampling bags come in various materials and shapes but most have a valve to allow for injection filling (with a syringe) using a pump and a septum. Sample losses from bags are caused by leaks, chemical reaction of the sample with the bag walls, adsorption of the sample into the bag material, and permeation of the sample through the walls of the bag.

2.5.2 Evacuated Vessels

In many cases, the sampler simply wants to collect a quick sample of air from a given area. Evacuated flasks and cans can be used for this type of collection.

Evacuation-type devices are sealed vessels in which a vacuum has been created, with only one inlet. These devices can be used to collect gross gas contamination. The concentration of contaminants must be high enough to be detected in a small-sample volume.

Evacuated cans, similar to those that hold shaving cream, are available commercially. By pushing in the top of the can, the seal is broken, allowing air to enter. Vacu-Sampler (MDA Scientific, Lincolnshire, IL, USA) is an example of a container completely evacuated in the factory (Figure 2). When the sample is collected, a volume of 120 cm^3 of vapor or gas can be obtained. Before use, bulk air sampler bags should be cleaned, tested for leaks, and preconditioned by flushing with the gas to be sampled (if known) or zero air.

2.5.3 Stainless-steel Canisters

Collecting ambient air samples in canisters provides a convenient method of integrating ambient samples over a period of time, such as 24 h. Collection can be unattended and provides enough samples to allow detection of small quantities of pollutants. Two sampling modes are used: passive and pressurized. In passive sampling, an evacuated canister is opened to the atmosphere through a flow restrictor adjusted so that the differential pressure causes the canister to fill over a designated period of time. For pressurized sampling, an initially evacuated canister is filled by the action of a flow-controlled pump from a near-atmospheric vacuum to a positive pressure that should not exceed $1.2 \times 10^{-5}\,N\,m^{-2}$ (25 psi). Pressurized sampling is used when long-term integrated samples or higher-volume samples are required. Commercial canister units are sold with a filter that traps particulate material. Figure 3 shows a canister used for sampling ambient gas and vapor. Canisters are generally not used for

Figure 2 The Vacu-Sampler can is under vacuum, but when the top of the can is pushed, the seal is broken, allowing air to enter. (Courtesy of MDA Scientific, Inc.)

the collection of polar compounds, such as methanol, ammonia, and hydrogen chloride, because these adsorb on the walls of the vessel. The list of volatile organic compounds (VOCs) that can be collected is defined in USEPA canister method TO-141.[9]

2.6 Integrated Sampling

Gas and vapor samples may be collected in a solvent with wash bottles, impingers, and absorbers; on adsorbents (e.g. activated charcoal); by condensation; and in large plastic bags that are filled using a battery-operated pump. After sampling, the contents may be analyzed in the field or laboratory by nondispersive IR (carbon monoxide), GC (hydrocarbons, chlorinated solvents), and other methods. Absorber characteristics vary, depending on the gas or vapor to be collected. Simple bubbling devices, such as impingers and Drechsel bottles, are adequate for readily soluble gases such as hydrogen chloride, hydrogen fluoride, and sulfur dioxide. For less easily absorbable materials, such as chloride and nitrogen dioxide, multiple contact washing is required (as with fritted glass absorbers). Sometimes it is desirable to burn

Figure 3 Canister for sampling ambient gas and vapor for VOCs. Samples can be collected in two modes: passive (the canister is under vacuum) and pressurized (the canister is filled using a pump). (Courtesy of Graceby Andersen Instruments, Inc., Smyma, GA.)

the gas or vapors in a furnace and sample the oxidation products, e.g. for chlorinated hydrocarbons.

Adsorption tubes are the method of choice for insoluble or nonreactive vapors. Commonly used adsorbents include activated charcoal, silica gel, and molecular sieves. Gas adsorption traps are sometimes preceded in the device by one or two water vapor adsorption stages containing calcium chloride, calcium sulfate, or silica gel, all of which have excellent water vapor adsorption characteristics and poor adsorption capacity for most organic molecules. In the laboratory, the collected vapors may be desorbed thermally or stripped from the adsorbent with carbon disulfide and the recovered vapors quantified by GC using a suitable detector.

Adsorption tubes used for personal integrated sampling of many organic gases and vapors contain two interconnected chambers in series, filled with gas adsorption charcoal. The first chamber, containing 100 mg of charcoal, is separated from a backup section, containing 50 mg of carbon, by a plastic foam plug. Sampling can be conducted for as long as 8 h without saturating the first chamber when occupational exposures are at or below the TLV. The contents of the two chambers are analyzed separately to determine whether the first-stage adsorbent has become saturated and has lost an excessive amount of the sample to the second stage.

GC column packing materials (e.g. Tenax®) are also used in adsorbent traps for field sampling of organic vapors. They are particularly useful for sampling high-boiling compounds. The resealed trap is returned to the laboratory for analysis. Absorption of inorganic

atmospheric constituents by liquid coatings on solid supports has been used to collect nitrogen dioxide on triethanolamine-coated molecular sieves.

Chemically active compounds may react with each other or with oxygen in the air after adsorption and make it difficult or impossible to recover and quantify the adsorbed gases and vapors. In such instances, the best collection method may be to react the contaminants chemically to a stable derivative or to condense them at low temperatures using a mixture of dry ice and acetone or liquid nitrogen as the coolant. Direct-reading instruments combine sampling and analytical functions and usually display results rapidly. Many in this category are also capable of storing continuous readings and displaying averages for selected time intervals on command.

Organic compounds and other pollutants present in air can exist either as gases or as particulate matter. These compounds can originate from various manufacturing and mechanical processes as a result of the incomplete combustion of the fuel material or as a result of the spontaneous volatilization of material from a chemical spill, uncontained waste, or uncontained feedstocks. Vegetation sources also release a significant volume of organic compounds into the atmosphere.

Pollutants that exist primarily in the gaseous state are described as VOCs. These are compounds with boiling points typically around or below 100 °C and vapor pressures of $>10^{-1}$ Torr. The semivolatile organic compounds (SVOCs) have vapor pressures in the range of $10–10^{-5}$ N m^{-2} ($10^{-1}–10^{-7}$ Torr) and can exist in the gaseous phase or in the particle-bound phase, either as an adsorbed material or as a liquid droplet or mist. These adsorbed compounds can remain attached to the particles as they are dispersed through the atmosphere, or they can undergo chemical reaction in the atmosphere. Alcohols, aldehydes, ketones, organic acids, esters, and organic nitrates have been identified as components of organic smog aerosols, which are essentially liquids.

2.6.1 Solid Sorbent Sampling

Generally charcoal, silica gel, and Chromosorbs are used for occupational air sampling and Tenax® and carbon molecular sieves are preferred for environmental samples. Since the levels expected are much lower in environmental samples, the collection period is much longer. Typically, environmental samples are collected on a 24-h average. Sampling time and volume must be determined experimentally, based on the overall concentration and the number of components.

Solid sorbents are specific for groups of compounds, and one sorbent will not work with all compounds.[8] Most solid sorbents do not differentiate among compounds during collection, so unwanted compounds may be collected along with the target compounds. On the other hand, because of the wide range of vapor pressures and breakthrough volumes for organic gases of interest, there is no one solid sorbent that can collect all gases and vapors simultaneously. Sometimes sorbents are coated with reagents to enhance collection of a specific compound.

The biggest concern in collecting material on a sorbent tube is whether breakthrough can occur. Breakthrough occurs when the front section of a tube is saturated and enough compound accumulates in the backup section for it to begin to exit the tube with the airstream.[10]

When selecting sorbents, the potential for water absorption must be considered. Both the compound being sampled and the water compete for the absorbent's surface, resulting in concentrations that appear lower than they actually are. Table 1 lists factors affecting the collection behavior of solid sorbents.[11]

2.6.1.1 Charcoal Sorbent Charcoal is one of the most commonly used sorbents, since it is useful for sampling a wide variety of organic gases and vapors, including several different compounds at a time. Ordinary charcoal is activated by steam at 800–900 °C, causing it to form a porous structure. Charcoal tubes are generally analyzed using the solvent carbon disulfide to collect adsorbed contaminants. The adsorbing capability of activated charcoal varies from batch to batch. Charcoal tube collection efficiency for various hydrocarbons may be affected by such variables as sampling rate, vapor concentration and totally adsorbed hydrocarbon mass. Nonpolar compounds, preferentially sampled on charcoal, displace polar compounds in charcoal media. Competitive adsorption also occurs among polar compounds. Factors that may contribute to the affinity of the molecules for charcoal

Table 1 Factors affecting the collection behavior of solid sorbents

Factor	Effects
Temperature	Reduced adsorption at high temperature; breakthrough volume is reduced
Humidity	Water vapor is adsorbed by polar sorbents and reduces breakthrough volumes
Flow rate	Varies with type of sorbents; at higher flow, reduces breakthrough volumes
Concentration	Breakthrough volumes decrease at highest concentrations
Mixture	When several compounds are present, the compound most strongly held will displace the other compounds, down the length of the tube
Nature of sorbent	Decreases in sorbent particle size are proportional to increases in sampling efficiency and drops in pressure
Size of tube	Breakthrough volumes are proportional to the sorbent volumes

include hydrogen bonding, molecular size, volatility, and the dipole moment.

Breakthrough volumes are variable and a function of the carbon, temperature, humidity, storage times, and pollutant.[11] Since charcoal tubes have been shown to be affected by high humidity, sample volume should be reduced when high humidity is present.

2.6.1.2 Silica Gel Sorbent Silica gel is considered a more selective sorbent than activated charcoal, and gases and vapors are more easily desorbed from it. It is the adsorbent recommended for collecting organic amines, both alkyl and aromatic, such as aniline and *o*-toluidine.

Factors that affect the dynamic adsorption of materials on silica gel include the size range of the gel particles, tube diameter and length, temperature during sampling, concentration of contaminants being sampled, air humidity, and duration of sampling. Since the polarity of the adsorbed compounds determines the binding strength of silica gel, compounds of higher polarity will displace compounds of lower polarity. Therefore, when attempting to collect relatively nonpolar compounds, the presence of co-existing polar compounds may interfere with collection on silica gel.

Silica gel shows a decrease in breakthrough capacity with increasing humidity. Under high humidity conditions, the sample is lost because of saturation with water vapor.[12]

2.6.1.3 Molecular Sieves The carbon molecular sieve is a spherical, macroporous structure, the result of pyrolysis of the synthetic polymeric or petroleum pitch precursors. The diameter of the micropores and their number are responsible for the differences in tube retention volume, adsorption coefficient, and equilibrium sorption capacity for a given compound.[13] The limiting factor with molecular sieves is humidity. Carbon molecular sieves are most commonly used to collect environmental samples of highly volatile nonpolar organic compounds.

2.6.1.4 Porous Polymeric Sorbents Porous polymers are another class of sorbent used for air sampling and include Tenax® GC, Porapaks, Chromosorbs 101–108, and XAD tubes. Their wide variety offers a high degree of selectivity for specific applications. Limitations include displacement of less volatile compounds, especially by carbon dioxide, irreversible absorption of some compounds, such as amines and glycols, oxidation, hydrolysis, and polymerization reactions of the sample, chemical changes in the contaminant in the presence of reactive gases and vapors, such as nitrogen oxides, sulfur dioxide and inorganic acids, artifacts arising from reaction and thermal desorption, limited retention capacity, thermal instability, and limitations of sampling volume, flow, and time.[12]

The Porapaks are a group of porous polymers that exhibit a wide range of polarity. The least polar member, Porapak P, is used in GC columns, and the most polar, Porapak T, separates water and formaldehyde. Chromosorbs 101–108 are similar to the Porapaks. Chromosorb 101 is the least polar and Chromosorb 104 the most polar. XAD resins include a number of different porous polymer types. XAD-2, the most commonly used, is equivalent to Chromosorb 102 and is used to collect anisidine and tetraethyllead.

2.6.1.5 Tenax® Tenax® is one of the most widely used porous polymers, especially for environmental sampling. It has been used for studies of VOC levels in indoor air. Tenax® is a polymer of 2,6-diphenyl-*p*-phenylene oxide and can be used to collect organic bases, neutral compounds, and high-boiling compounds. Tenax® is used mostly for sampling low concentrations of volatile compounds.

Tenax® GC has a high thermal stability and can withstand temperatures of up to 350 °C, which permits it to be used for thermal desorption. Thermal desorption is desirable because the entire sample is introduced into the analytical system, whereas solvent extraction dilutes the sample and allows only a portion of it to be injected. One limitation of thermal desorption is that the sample can be injected only once. Tenax® has other properties besides its temperature stability that make it useful for concentrating compounds of medium volatility. It is relatively inert and has low, but not zero, affinity for water vapor. Because of its extensive use, its advantages and limitations have been well characterized. It is commonly used for multicomponent qualitative analysis by gas chromatography/mass spectrometry (GC/MS) and quantitative analysis with standards.[14]

However, retention or breakthrough volumes for a variety of highly volatile compounds on Tenax® are low.[15] Other limitations of Tenax® include a laborious cleanup procedure to control blank problems, a short useful half-life after sample collection, and a tendency to decompose during sampling to produce acetophenone and benzaldehyde; the humidity in the air being sampled may affect sample retention; high concentrations of or a large number of compounds may exceed the retention capacity of the Tenax® bed; and artifact formation may occur owing to chemical reactions during sampling and/or thermal desorption. Tenax® is not effective for low molecular weight hydrocarbons (C_4 and below) and midrange (C_5–C_{12}) highly polar compounds.

Tenax® also reacts with strong oxidizing agents, such as chlorine, ozone, nitrogen oxides, and sulfur oxides, to form benzaldehyde, acetophenone, and phenol. Another

documented chemical transformation with Tenax® is that oxidizers react with organics, such as styrene, to produce compounds such as benzaldehyde and chlorostyrene.

2.6.1.6 Other Sorbents Other sorbents occasionally used for integrated gas and vapor sampling include alumina gel and Florisil. Alumina gel, a form of oxide, is rarely used as an air-sampling sorbent except for special applications such as those using formaldehyde. Alumina gel permits a several thousandfold concentration of pollutants. It selectively absorbs polar and higher molecular weight compounds, with the degree of polarity determining the binding strength of a compound on this gel. Alumina gel has been used to collect polar compounds (such as alcohol, glycols, and ketones) and aldehydes, such as formaldehyde. Like all polar sorbents, it has a high affinity for water and thus high-humidity samples are difficult to handle.

Florisil, based on silicic acid, is used for polychlorinated biphenyl (PCB) collection and also for some pesticides. Table 2 lists types and uses of solid sorbents.[12]

2.6.2 Impingers for Solvent Extraction

Impingers are suitable for collecting nonreactive gases and vapors that are highly soluble in the absorbing liquid and also those that react rapidly with a reagent in the solution, such as occurs in the neutralization of strong acids and bases. These devices are usually made of glass with an inlet tube connected to a stopper fitted into a graduated vial. The inlet tube rests slightly above the vial bottom. A measured volume of absorber liquid is placed into the vial, and the unit is connected to a pump by flexible tubing. The function of these absorbers is to provide sufficient contact between the sampled air and the liquid surface to provide complete absorption of the gas or vapor. Figure 4(a) and (b) shows how impingers are set up to collect personal samples.

Problems with the use of impingers include condensation of material in the sampling lines and losses by adsorption or volatilization from the equipment.

Table 2 Types of solid sorbents

Solid sorbent	Characteristics
Activated charcoal	Very large surface area; reactive surface; high adsorptive capacity; breakthrough capacity is a function of the source of the charcoal, its particle size, and packing configuration in the sorbent bed
Silica gel	Less reactive than charcoal; polar; hygroscopic
Porous polymers	Less surface area and less reactive surface than charcoal; adsorptive capacity is lower; reactivity is lower
Molecular sieves	Zeolites and carbon molecular sieves retain adsorbed species according to their molecular size; water may displace organics
Coated sorbents	These are sorbents on which a layer of reagent has been deposited; adsorptive capacity is determined by the capacity of the reagent to react with a particular analyte

(a) (b)

Figure 4 Personal sampling impingers and bubblers. These devices are usually made of glass and filled with the absorbing liquid. (Courtesy of Gilian Environmental.)

Impingers are bulky and are susceptible to breakage. For these reasons, occupational sampling with impingers is being phased out and replaced by sorbent methods. However, there are still applications, such as in high humidity or when only certain samples (e.g. formaldehyde) are collected, for which the impinger approach is necessary.[16]

2.6.3 Badges

Another passive colorimetric sampling device is the spot plate, or badge. These have indicator strips or buttons designed to change color when a critical accumulation of a given gas is reached. Badges should not be used to warn employees of levels in excess of the PEL. Real-time instruments with audible alarms should be used for these situations. An example of a badge indicator is the ChromAir (Gilian Environmental, West Caldwell, NJ, USA) (Figure 5). This passive monitor badge is for indoor air quality and OSHA compliance screening. The passive monitor relies on the principle of diffusion. The contaminant gas or vapor being monitored diffuses through a series of six cells on the front of the badge, each having an increased resistance to diffusion. After diffusing through the cells, the gas or vapor reacts with an indicator layer containing a specific reagent for the contaminant, and the cells change color. When sampling is completed, the exposure dose is clearly indicated on the opposite side of the badge. ChromAir badges are available for ammonia, acetone, chlorine, carbon monoxide, sulfur dioxide, nitrogen dioxide, hydrogen sulfide, carbon disulfide, methanol, ethanol, hydrazine, and formaldehyde. The sensitivity of these badges depends on the application and is typically in the parts per million range. To obtain the average concentration in parts per million, the exposure dosage (parts per million) is divided by the exposure time in hours.

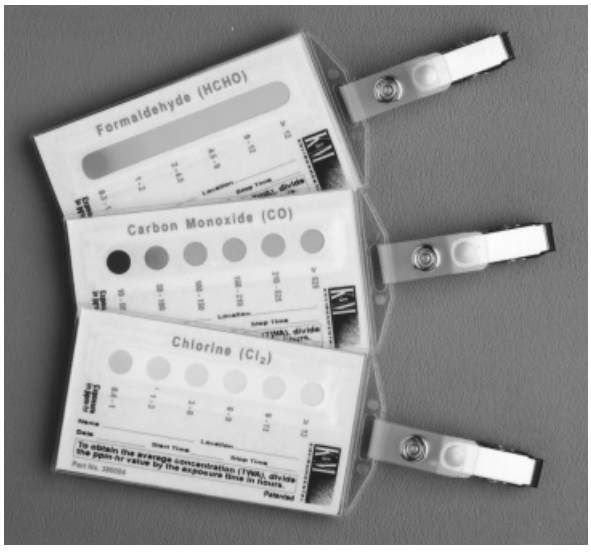

Figure 5 ChromAir passive monitoring badges from Gilian Environmental.

2.6.4 Passive Collection of Gases and Vapors

Passive monitors are lightweight badge assemblies that rely on natural wind currents rather than pumps to move contaminated air to the collection surface. Most of these units are dosimeters that accumulate an average dose and give an integrated measurement. Some units are specific for a given gas, such as chlorine, ammonia, sulfur dioxide, hydrogen sulfide, nitrogen dioxide, carbon monoxide, and mercury. Others can sample multiple gases such as organic vapors. The obvious advantage of passive monitors is their ease and simplicity of use. When used for personal monitoring, they do not interfere with worker activity and are unlikely to affect the behavior pattern of the wearer, whereas wearing a pump with its tubing and media might. The most commonly used passive monitors contain a solid sorbent similar to that used in sorbent tubes. There are also devices that use liquid absorption solutions and passive devices that change color proportional to the concentration present.

Passive devices rely on two basic collection principles: diffusion and permeation. With diffusion-controlled monitors, the mass uptake of the monitor is controlled by the length and diameter of the badge cavity and the physicochemical properties of the contaminants. When permeation is used for sampling, the mass uptake of the monitor is controlled by the physicochemical characteristics of the membrane and the contaminants. The mass uptake is a direct function of the badge permeation sampling rate, the ambient concentration, and the sampling time.

In diffusion devices, the most common passive devices, molecules pass through a barrier or draft shield that minimizes the effect of air currents on a stagnant air layer; they are then collected on an adsorbent material such as charcoal. A concentration gradient is created within the cavity of stagnant air, and the amount of gas or vapor transferred is proportional to the ambient concentration. Each compound has a unique diffusion coefficient for each type of badge. The diffusion coefficient necessary to calculate the final concentration after analysis in the laboratory limits sampled materials to those for which this value is established.

On permeation dosimeters, the gaseous contaminants dissolve in a polymeric membrane and are then transferred to a collection medium, such as a solution. Permeation across the membrane is controlled by the solubility of the gas or vapor in the membrane material and by the rate of its diffusion across the membrane under a concentration gradient. Factors influencing permeation include thickness and uniformity of the membrane, affinity of the membrane for the contaminant, swelling or

shrinking of the membrane, and possible etching by corrosive chemicals. The efficiency of these devices depends on finding a membrane that is easily permeated by the contaminant of interest and not by the others.

As for diffusion monitors, sampling rates must be determined for each analyte and type of permeation monitor. One difficulty in making such determinations is the need to use thin and fragile membranes to obtain practical sampling rates. The long response times required to reach equilibrium, the sampling rate, and the degree of permeation may be affected by changes in temperature and ambient humidity.

The accuracy and precision of passive monitors depend on sampling time, air currents, and temperature and humidity effects. In the presence of mixtures, passive monitors may be affected by competition among compounds for adsorptive sites, resulting in displacement of one compound by another or preferential adsorption of one compound over another. Passive monitors must be handled with care, since a penetration in the membrane would cause variations in the amount of contaminant collected. Sampling in a dusty environment may also be a problem if the gas being collected is capable of adsorbing onto the dust.

Collection of organic vapors is one of the most common uses for passive monitors. Organic vapor badges contain twice as much charcoal as a typical charcoal tube. The charcoal is usually impregnated in some sort of a matrix. Regardless of the design of the organic vapor monitor, the analytical procedure is generally the same: desorption with carbon disulfide and analysis by GC. The affinity of various compounds for the activated carbon in a badge varies and, because diffusion coefficients exist only for certain gases and vapors, the use of badges is limited to these compounds regardless of the wide applicability of the badges for sampling organic vapors. The accuracy of an organic vapor monitor scan varies, depending on the concentrations being sampled and on the exposure period. The difference in the design of the individual monitors, type of charcoal used, quality control, type of adsorption pad, particle size of the charcoal, amount of charcoal present, surface area of the adsorption material, and membrane or draft shield selection create each monitor's unique performance with respect to the compounds collected.

3 DIRECT-READING INSTRUMENTS FOR ENVIRONMENTAL MEASUREMENTS

Direct-reading instruments for gases and vapors operate on many different principles. They include detector tubes and instruments such as the FID and PID. Chemical sensors are part of another family of direct-reading sensors that is rapidly displacing some of these techniques. Chemical sensors are currently undergoing intense research and development. These detectors are generally not specific but can be calibrated for a single vapor or known mixture. Gas chromatographs, mass spectrometers, and ion mobility spectrometers represent a distinct class of instruments in that they specifically address the issue of separation as well as detection, and as such they will be discussed in later sections. This section reviews direct-reading instruments. A comprehensive review of direct-reading gas and vapor sensors, electrochemical devices, ionization detectors, optic sensors, and instruments is available.[17]

3.1 Introduction to Direct-reading Instruments

Chemical sensors are a fast-growing part of the direct-reading instrument area. They are found almost everywhere when a single-point measurement such as oxygen, carbon monoxide, or carbon dioxide must be performed. This type of sensor often operates in multisensor combinations. For instance, oxygen meters are often paired with combustible gas detectors. Oxygen measurements are needed not only to evaluate the presence of the concentrations necessary to sustain life but also to evaluate correctly the readings of combustible gas indicators (CGIs). Chemical sensor applications are found almost everywhere: in environmental measurements of air and water, in industrial applications for food and pharmaceuticals, in clinical applications, and in many other applications. This great interest in chemical sensors is due to their small size, low cost, low power requirements, and capability of being integrated into portable instruments. This discussion is limited to gas detection.

Indicator (or detector) tubes are direct-reading industrial hygiene air analysis instruments. They are small, light, hand-operated, and safe in all atmospheres, and generate immediate readouts. In addition, an indicator tube is the simplest and most economical air analysis method available for many common air contaminants. Tube detectors for >160 chemicals are currently available, some in more than one concentration range. The most popular types are offered by commercial sources. Most of the chemicals measured by indicator tubes are included in the current tabulation of American Conference of Governmental Industrial Hygienists (ACGIH) TLVs. Color-developing reagents give point measurements that are almost instantaneous and integrated and that are collected for a long time. The approach chosen is usually dictated by the regulator and by the sensitivity of the measurement device. The instant repartition of specific contaminants may be lost when integrated measurements are chosen over a long period of time, but this

approach is better at identifying contaminants that are harmful to the exposed community.

One of the most important developments in air-sampling technology has been the commercial appearance of passive dosimeters that measure the exposure to low concentrations of a broad list of airborne volatile substances. Many of these dosimeters use the principle of diffusion to a nonspecific adsorbent with subsequent laboratory analysis. These direct-reading devices use permeation through a plastic film barrier to a compound-specific chemical that bonds to color-developing reagents.

Instruments such as the PID and FID respond to a broad class of organic and inorganic gases and vapors and are used widely for survey measurements. The only specificity of these instruments is determined by the class of molecules they detect. For instance, FID meters detect only organic compounds and do not respond to constituents commonly found in air and water. The threshold of specificity of the PID meter is controlled partially by the selectable level of energy emitted by the lamp.

3.2 Chemical Gas Sensors

Chemical sensors transform chemical information into electrical signals. These sensors are characterized by small dimensions and relatively low costs and lead naturally to portable instruments for survey and personal measurements. These sensors are generally classified according to the mode of transduction, e.g. electrochemical, optical, or acoustic. Some well-established chemical sensors include electrolyte potentiometric sensors for oxygen, catalytic sensors for combustible gases, and semiconductor oxide sensors. These find use in the detection of both the combustible gases and the toxic gases associated with chemical systems.

New techniques for chemical sensors are emerging, such as field effect transistors, surface acoustic wave (SAW) devices, fiber-optic gas sensors, and pattern recognition methods used with arrays of sensors in so-called electronic noses that evaluate flavors.[18] Solid electrolyte sensors for oxygen and semiconductor gas sensors, based on the response characteristics of a wide range of inorganic oxides and of organometallic materials, are also part of this new generation of sensors. Some of the designs are described below.

Electrochemical sensors with aqueous electrolytes, operating on fuel cell principles, have been adopted widely to meet the growing demand for reliable low-cost gas-measuring devices in safety and process control applications. They are available for the detection of organic and inorganic gases such as ammonia, sulfur dioxide, hydrogen sulfide, chlorine, carbon monoxide, isocyanates, nitrogen compounds, oxygen, phosgene, formaldehyde, ozone, and hydrogen cyanide. General information about chemical sensors and electroanalytical methods is available.[19–22] Applications of real-time measurements using electrochemical sensors have been published.[8]

The greatest advantage of electrochemical sensors is that they provide a direct relationship between the concentration of gas being measured and the electrical signal at the output (current or voltage, depending on the type of sensor). Also, they can be constructed using well-established engineering methods to produce small, compact devices with a high degree of robustness and reliability, freedom from orientation effects, intrinsically low cost, and suitability for volume production. Such sensors can operate within an ambient temperature range of -20 to $55\,^{\circ}C$ without the need for external heating. Their power requirements are, therefore, low; some designs are available that are completely self-powered, so that additional power is required only for extrasensor functions such as alarm monitoring, recording, and transmitting data.

A wide range of gas concentrations may be measured using electrochemical gas sensors. Thus commercially available systems to monitor oxygen levels range, e.g. from nearly 100% in medical and gas-purity applications to around 20% for general air quality systems, to just a few percent in flue gas analysis, down to a few parts per million in some gas purity and process control applications. Toxic gases can be measured from around 20% down to a few parts per billion.

Electrochemical sensors do not suffer from variations in ambient humidity, a problem for many semiconductor and solid-state devices. Even through cross-interferences can pose problems, as with any chemical sensor, electrochemical sensors can achieve high specificity when suitably designed and operated.

3.2.1 Principle of Operation of Electrochemical Sensors

The following discussion is limited to potentiometric and amperometric sensors, since they are the most commonly used with portable instruments.

3.2.1.1 Potentiometric Sensors If a cell signal reflects a change in the potential of the sensor, it is classified as potentiometric. Potentiometric devices have been well studied and can serve as sensors for a variety of gases such as ammonia, hydrogen cyanide, carbon dioxide, hydrogen sulfide, and sulfur dioxide. Potentiometric devices include ion-selective electrodes (of which the pH electrode is a well-known example). These operate on the basis of the Nernst equation, for which electrode potential is a function of concentration (Equation 1):

$$E = E^* + \frac{RT}{z} \log C \tag{1}$$

where E is the potential of the sensing electrode (SE), referred to as the standard hydrogen electrode (SHE), in volts, E^* is the standard electrode potential of the SE, in volts, R is the ideal gas constant ($8.3143\,J\,K^{-1}\,mol^{-1}$), T is the absolute temperature, in kelvin, C is the ratio of the product of the activities of the product species of the electrochemical reaction, written as an oxidation reaction, to the product of the reactant species of that electrochemical reaction (in dilute solutions, activities may be approximated by concentrations), and z is the number of electrons transferred in the electrochemical reaction.

Potentiometric sensors for gases monitor the direct chemical reaction of the gas with the electrolyte, thereby changing the potential of the SE from the initial potential to produce an electrochemical couple with one of the electrolyte ions. The potentiometric sensor observes the potential difference between the SE and another electrode. This difference occurs when the chemical species of choice is detected. The output logarithmically depends on the concentration of the species being detected. The recovery time of the sensor is a function of the concentration of the sample. The higher the concentration, the longer is the recovery time. At a high concentration, the sensor may never recuperate.

3.2.1.2 Amperometric Sensors Amperometric sensors are important in portable instrument design because they are relatively small, inexpensive, and lightweight, and use little power to generate significant signals. These sensors exhibit fast responses, detect parts per million levels of electrochemically active gases and vapors, can be engineered to have significant selectivity, and operate over a wide range of temperatures. Their accuracy is also considered better than that of other methods such as potentiometry. Amperometric sensors are further subdivided into two classes: potentiostatic and coulometric.

In potentiostatic amperometry, an indicating electrode is polarized against a reference electrode (RE) to a constant potential by passing a polarizing current. The sample flows through the cell, and the electric current proportional to the concentration of the species to be measured is recorded. A potentiostatic sensor uses a constant-voltage circuit, i.e. the SE is held at a fixed voltage.

Electrochemical sensors can be categorized as galvanic or electrolytic. A galvanic cell consists of two electrodes and an electrolyte undergoing a spontaneous chemical reaction, which develops an electrical potential difference between the electrodes. The term galvanic refers to any device that can convert chemical energy to electrical energy, be it in a gas sensor, battery, fuel cell, or any other device. Electrolytic is a term applied to the conversion of electrical charges into chemical energy, as in electroplating or charging a battery. Potentiostatic amperometry involves galvanic devices, but requires an external variable current supply to maintain a fixed electrode potential; coulometric amperometry involves electrolytic devices. The process that occurs in these sensors is electrooxidation or electroreduction of the species to be analyzed, resulting in a current. Some require oxygen to function and others do not.

A typical potentiostatic amperometric sensor consists of several major parts: filter, membrane, working (or sensing) electrode, electrolyte, counter electrode, and RE. The gaseous species of interest is transported (by pump or diffusion) across the membrane to the SE of the cell; it then migrates to the electrolyte boundary, dissolves in the electrolyte, diffuses to the electrode surface, and reacts electrochemically. The products of the reaction diffuse away from the electrocatalytic surface.

As a way of understanding this method of detection, a typical potentiostatic curve is shown in Figure 6. Oxidation and reduction current are represented on the $+y$ and $-y$ axes, respectively, while voltage is plotted along the x axis. Below the reaction potential of a given electrooxidizable gas there is no reaction; therefore, the current is zero (except for a small charging effect). As the reaction potential is approached and exceeded, the current rises sharply until it reaches a maximum limited only by the diffusion of the gas to the reaction site. The magnitude of this current is directly proportional to the number of electrons per mole, the Faraday constant, the diffusion coefficient, the surface area of the reaction site, and the concentration of the gas; it is inversely proportional to the diffusion gradient resulting from the varying concentrations of reacting gas that range from the highest concentration found in the electrolyte, to zero concentration at the electrode surface. There is a linear relationship between the current generated and the concentration of the gas. An even further increase in voltage results in

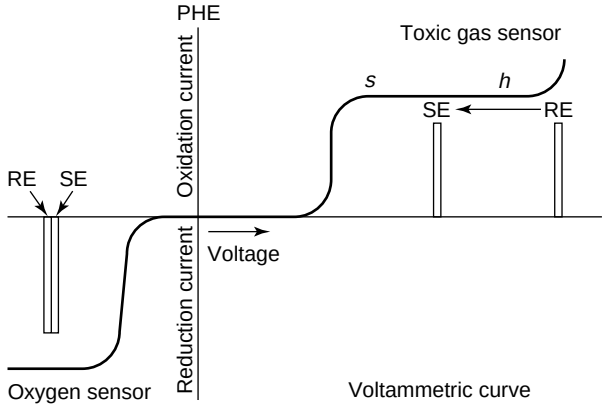

Figure 6 Basic principles of operation of a potentiostatic sensor. (Courtesy of Interscan Corp.)

a high current because of electrooxidation of water to oxygen gas, and no detection takes place.

For the device to work, the sensing (working) electrode must be held at a potential within the region of the limited diffusion current. This is represented by the SE in Figure 6. An RE with a lower or higher oxidation potential than that of the SE is used to maintain the SE in the region of the limited diffusion current. The potential of the RE must not change with the passage of current. Depending on whether the potential of the RE is higher or lower than that of the SE, it will have a voltage that is biased in either the negative or positive direction. The height of the diffusion plateau from the voltage axis is linear with the concentration of the sample present. This is an example of an externally biased sensor. Typical features of potentiostatic sensors include fast response, linear response in a broad concentration range, nonlinear dependence on the flow rate, and temperature dependence of the signal.

Commercial coulometric monitors are constant-current devices. Coulometric monitors measure the quantity of electricity (in coulombs) that passes through a solution during the occurrence of an electrochemical reaction. Therefore, measurement of the amount of electrical energy transferred across an electrode–solution interface (number of electrons) in terms of the coulombs required to carry the reaction of a specific substance to completion is called coulometry. During coulometric sensing, the gas being monitored is consumed by electrolysis during passage through the sensor. The current is controlled by the feed rate of the sample and corresponds to the charge passed in a given time unit (Equation 2):

$$\text{Current} = \frac{\text{Charge}}{\text{Time}} \tag{2}$$

It is generally a wet-chemical titration method in which one of the reactants is generated in the test cell by electrolysis of the solution. These detectors can be made specific by adjusting the concentration, pH, and composition of the electrolyte. An important feature of the coulometric sensor is that the current is, within certain limits, independent of changes in the working electrode; hence the electrode characteristics are unimportant. Electrodes may be constructed from platinum foil, wire, or mesh. The accuracy of the sample flow rate controls the accuracy of measurement and must be kept as constant as possible; the signal is, however, independent of the temperature. Coulometric cells are reusable after being cleaned and recharged with a fresh electrolyte solution.

3.2.2 Basic Description of a Fuel Cell Electrochemical Sensor

In its simplest form, a fuel cell electrochemical sensor consists of two similar gas diffusion fuel cell electrodes, a concentrated ionically conducting aqueous electrolyte, a low-impedance external electrical circuit, and a diffusion barrier (Figure 7).[19] In clean air, when reactant gases are absent, both electrodes assume the same potential and no current flows in the external circuit. If an electrochemically oxidizable gas is present, it diffuses to the SE first and causes its potential to shift in the negative direction. The resulting potential difference between the SE and the other electrode, the so-called counter electrode, then causes a current to flow in the external circuit, which is sustained by electrochemical oxidation of the reactant gas at the SE and matched by an equivalent amount of oxygen reduction at the counter electode.

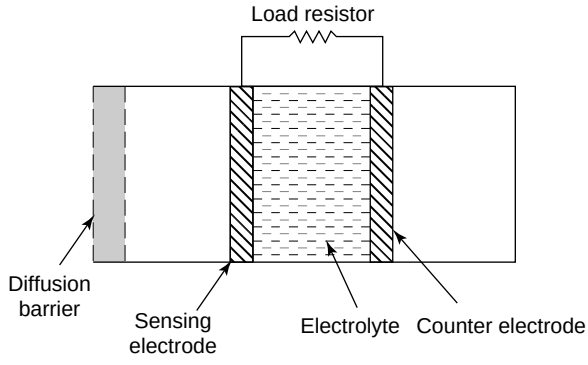

Figure 7 Schematic diagram of the basic elements of a fuel cell gas sensor.[19]

A carbon monoxide sensor with an acid electrolyte is represented by Equations (3–5):

Sensing electrode:

$$2CO + 2H_2O \longrightarrow 2CO_2 + 4H^+ + 4e^- \tag{3}$$

Counter electrode:

$$O_2 + 4H^+ + 4e^- \longrightarrow 2H_2O \tag{4}$$

Cell reaction:

$$2CO + O_2 \longrightarrow 2CO_2 \tag{5}$$

Carbon monoxide diffuses via the diffusion barrier to the SE, where it is oxidized to carbon dioxide and protons, consuming water and releasing electrons to the external circuit. The carbon dioxide product is rejected by the acidic electrolyte and diffuses out of the cell via the diffusion barrier. Protons produced at the SE migrate to the counter electrode and recombine with electrons and oxygen to form water.

This electrochemical power source is converted to a sensor by the inclusion of the diffusion barrier at the SE of the cell. This barrier is fundamentally important to the operation of the sensor and is designed to restrict access of the reactant gas so severely that it becomes completely oxidized at the SE.

A reactant molecule present in ambient air passes through the following stages in undergoing detection by an electrochemical sensor:

1. mass transport in the gas phase to the diffusion barrier;
2. diffusion, under a concentration gradient, through the sensor barrier;
3. gas-phase diffusion from the barrier to the electrode, through any porous electrode support and into the electrolyte structure itself;
4. diffusion through an aqueous electrolyte film to the catalyst surface;
5. adsorption from the liquid phase onto the catalyst surface, followed by surface migration to the catalyst active sites;
6. electron transfer between catalyst surface and adsorbate, at the active center, to form products;
7. desorption of products and their subsequent transport back into the gas phase and ambient air, and/or into the bulk electrolyte.

3.2.2.1 Diffusion Barrier The nature of the diffusion barrier exerts a profound influence on the characteristics of the sensor. Two broad categories of diffusion barrier have been used, identified according to the diffusion process through the barrier.

In the first type, the rate-controlling process is one of diffusion in solution, sometimes designated activated diffusion. It is characterized by a high exponential response to temperature. The two variants of activated diffusion barriers involve diffusion in solution through an electrolyte film and through a solid polymer membrane.

In the second type, the gas remains in the gas phase as it diffuses through a porous barrier. It is characterized by a relatively low temperature coefficient and also exists in two variants involving either microprobes or micropores. The most common barrier employed for macropore diffusion uses a capillary or pore size that is much greater than the mean free path of the gas molecules undergoing diffusion. Here, intermolecular collisions dominate the diffusion process, resulting in a sensor output that has a considerably lower temperature coefficient than do solid-state membranes. It should be mentioned that the sensor will have a $T^{1/2}$ temperature effect.

For barriers in which the pore diameter is much smaller than the mean free path of the reactant gas molecules that occur in the micropore barriers, the molecules pass through the pore, ricocheting off the pore walls with substantially no intermolecular collisions.[23] This diffusion process has fundamentally different properties from the macropore process and leads to different sensor properties. The diffusion rate is determined by the rate of capture of molecules and the molecule velocity. The rate of capture varies as T^{-1} and the velocity as $T^{1/2}$, the combination resulting in proportionality to $T^{-1/2}$. The capture rate, and hence the diffusion rate, vary as d^{-1} (where d is the pore diameter), in contrast to the macropore case, for which diffusion rates are independent of pore size.

Of practical significance is that macropore membrane barriers can be easily fabricated and do not suffer the problems associated with solid membranes. By combining both categories of porous diffusion barriers, the negative and positive temperature coefficients can be offset to produce a sensor with negligible temperature coefficient.[24]

3.2.2.2 Electrodes The electrochemical reactions occur at the electrode–liquid interface, where the solid catalyst surface is in contact with the gaseous reactant dissolved in the liquid electrolyte. To maximize the area of contact, gas diffusion electrodes are employed.[25,26] These consist of finely divided catalyst powders with large surface areas, mixed with a polytetrafluoroethylene (PTFE) and supported on a highly porous PTFE backing membrane. This design controls catalyst wetting and provides gas channels to the wet electrode interior.

3.2.2.3 Electrolyte The electrodes of an electrochemical sensor are separated by a high ionic conductivity, negligible electronic conductivity medium. All commercial liquid electrolyte fuel cell sensors employ aqueous solutions of strong mineral acids, alkali metal hydroxides, or neutral salts that are absorbed in a separator material between the electrodes, which consists of plastic or glass mats as used in conventional electrochemical power source systems. The electrolyte should have a high boiling point and a low freezing point (typically $< -20\,^{\circ}C$).

The choice of electrolyte can affect the electrocatalytic activity of the electrode and can provide a means of controlling selectivity. Some gases are completely inactive, electrochemically, in alkaline media, but react readily in acidic media, while other gases react conversely.

The work in the field of solid ion conductors in the 1950s led to increasing interest in all solid electrolyte cells. A solid-state cell in which the electrolyte was a zirconia-based ceramic was developed in the 1960s for the measurement of oxygen partial pressures. Zirconia sensors have been used in industry ever since, and the technology is now considered mature. A major application for zirconia sensors operating on the electrochemical principle is the monitoring of the air : fuel ratio in fossil fuel combustion systems; this monitoring, combined with closed-loop control, permits maximum efficiency and minimal pollution. Zirconia sensors are versatile and can be tailored to the application; the response time to a change in oxygen gas concentration can be rapid (e.g. 50 ms for oxygen partial pressures > 100 Pa at $700\,^{\circ}C$);

the sensor can be highly selective to oxygen. Potentiometric devices can measure oxygen pressure levels over an extremely wide range (10^{6}–10^{-25} Pa); alternatively, amperometric devices deliver a linear response, which is preferred in some applications. It is likely that zirconia sensors will continue to dominate the oxygen-sensing market for applications requiring operations at moderate and elevated temperatures.[27]

3.2.2.4 Two-electrode and Three-electrode Designs In its simplest form, a fuel cell sensor comprises two electrodes connected by a load resistor. Such a device is self-powered, and the voltage developed across the load resistor R_{load} by the system constitutes the signal from the sensor. The response time of the cell increases proportionally with the current. The lowest possible value is desirable, and an alternative approach consists of using a potentiostatic circuit that runs the cell at nearly zero current. The equivalent circuit is shown in Figure 8(a).

A larger problem with the two-electrode approach derives from polarization at the counter electrode, particularly when an oxygen reduction counter electrode is used. Such a counter electrode has the advantage of unlimited capacity for ambient air monitoring but has the disadvantage of significant polarization. The SE potential must follow the counter electrode potential fairly closely. Changes in potential of the SE can cause interfering currents from side reactions caused by other electrochemical couples.

A practical consequence is that the useful concentration range of such two-electrode sensors is limited. However, they are simple and self-powered and, therefore, are used successfully commercially.

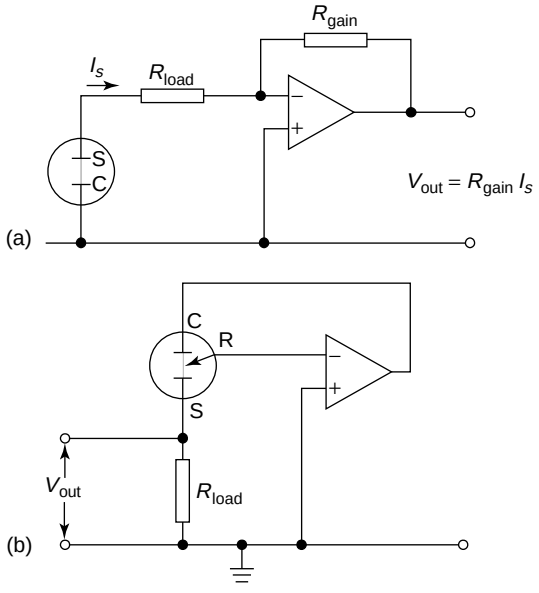

Figure 8 (a) A two-electrode potentiostatic circuit. (b) A three-electrode potentiostatic circuit. S, sensing electrode; C, counter electrode; R, reference electrode.

The effect of polarization at the counter electrode can be completely eliminated electronically using three-electrode sensors (Figure 8b).

In a three-electrode sensor, all three electrodes may be constructed of the same material. The way in which they are connected to the circuit determines whether they are sensing, counter, or reference electrodes. The SE interacts with the gas to be monitored. The counter electrode acts as an electron sink for anodic sensing reactions by completing the circuit for the sensor. The third electrode acts as an RE. No appreciable current passes through the RE, because it is connected to the high impedance input of an amplifier. The RE provides a bias reference for either an applied or an internal potential, so that the SE is at a potential within the region of the limited-diffusion current. Using a potentiometric feedback between the RE and the SE maintains the latter at a constant potential.

As with a two-electrode circuit, the signal is determined by the voltage generated across R_{load}, except that an amplifier with negative feedback is used to maintain a constant potential relative to the RE at the ground side of the load resistor. This potential is independent of the counter electrode potential. With this arrangement, the change in potential of the SE is due solely to the current generated at the SE by the reactant gas; this current is used as the signal.

Three-electrode circuits also allow the SE potential to be biased with respect to its rest potential, which can be beneficial in promoting desired reactions that do not occur at the RE potential.

3.2.2.5 Construction of a Typical Gas Electrochemical Sensor There are two basic types of sensor designs available commercially, depending on the electrolyte (alkaline or acidic) and the operating principle (metal–air battery or fuel cell).

The construction of an alkaline metal–air, capillary-type oxygen sensor is shown in Figure 9. Conventional nickel–cadmium battery hardware is used for the containment can, which is filled with a lead anode material. This containment can also act as the counter electrode electrical output. A sealing grommet holds the oxygen-reducing SE assembly, which inserts into the top of the anode can after electrolyte charging. Cell closure and sealing are effected by crimping the anode can over the grommet, thus avoiding electrical contact with the cathode.[28]

Various modifications of sensors are possible. For example, one manufacturer may choose to acidify slightly the hydrogen peroxide electrolyte in a sulfur dioxide conductimetric sensor to decrease the effect of certain

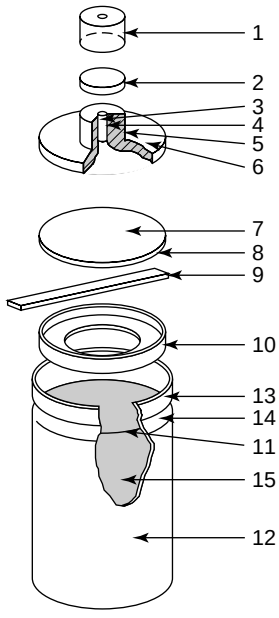

Figure 9 Schematic diagram of the construction of an oxygen–lead, alkaline electrolyte electrochemical sensor. 1, end cap; 2, antibulk flow membrane; 3, capillary; 4–6, capillary plates; 7 and 8, SE; 9, SE contact; 10, sealing grommet; 11, separator; 12–14, container, anode contact; 15, anode. (Courtesy of City Technology, Ltd.)

interferents. The gas exposure path can be modified to obtain differing gas flow rates that reach the electrode, thus affecting the magnitude of the signal. In units with active sampling, high flow rates are used to pull gases to the sampling electrode to obtain the highest sensitivity. Control of the electrode exposure area is important, and in practical designs as much catalyst as possible is packed into this area (by using material with high surface area) to ensure adequate reacting capability and high sensitivity. Each chemical species has its own unique interaction at an electrode, but for almost all the electrochemical cells, the primary variables are working electrode potential, analyte concentration, current, and time.

The construction materials of the sensor also influence its operating characteristics. Choosing construction materials and sensor geometry is critical and has a profound influence on the accuracy, precision, response time, sensitivity, background signal, noise, stability, lifetime, and selectivity of the resulting sensor. For example, selection of a gold rather than a platinum electrocatalyst for the SE allows for the selective determination of hydrogen sulfide in the presence of carbon monoxide.

Membranes are usually chosen for their ability to protect the SE. However, if the membrane has a low permeability to air, the sensor will have a slow response time. Materials used for the membrane construction are typically Teflon® and high-density plastics such as polypropylene, because such materials must be compatible with reactive gases and corrosive electrolytes.

It is important to select the proper electrolyte for each sensor, since the electrolyte composition can affect the solubility and the rate of diffusion of the reactant gas to the electrode (catalyst) surface. Electrolyte composition can also alter the chemical being monitored before it reaches the electrode surface. For example, the use of an acidic electrolyte to detect ammonia causes the formation of an ammonium ion that may not be as electrochemically active as ammonia under the conditions of the cell. The electrolyte profoundly influences the response characteristics observed for sensors with strongly acidic and basic electrolytes.

Another means of chemically controlling the sensor properties is by altering the composition of the electrocatalyst. Each catalyst formulation has unique properties. The reactivity of platinum with carbon monoxide oxidation has been found to be 10^3–10^6 times better than that of gold. Although both reactions occur on metal, one is orders of magnitude faster than the other.

The selectivity of the sensor can also be improved by controlling the electrochemical potential of the working electrode. For instance, proper selection of a gold electrode potential allows the determination of nitrogen dioxide in the presence of nitric oxide. Response time, linearity, zero drift, repeatability, sensor stability, and even sensor life depend on sensor design and methodology.

As noted, gas sensors are often covered with a membrane that is selectively permeable to a given contaminant. This membrane minimizes the likelihood of poisoning the electrodes by an electroactive and surface-active species. The resolution of these systems is also enhanced if other contaminants (interferents) that also undergo electron transfer at the electrode can be excluded by tuning the circuitry. The appropriate choice of electrocatalyst can help in achieving selectivity in some applications. However, although choosing a membrane or an electrolyte for its selectivity is a good idea, it does not eliminate all the possible interferents, since virtually any compound for which a similar type of detection method is used is a potential interferent.

3.2.3 Multiple Gas Instruments Based on Gas Electrochemical Sensors

Multiple gas instruments that are able to monitor several gases at the same time are available. Advantages to this approach are almost simultaneous monitoring of various gases at a low concentration level without need for a separation column, the low cost associated with the use of electrochemical sensors, relatively fast access to the information, and portability caused by the small

dimensions of the sensors. The principal goals of such a system are to control the sampling conditions and to limit cross-sensitivities to other compounds.[29]

Figure 10 shows a typical arrangement for a multisensor system equipped with a water trap, particle filter, pump, and in-line SO_x–NO_x filter. Moisture condensing from the gas stream can cause blockages and flooding, and absorbs sulfur dioxide and nitrogen dioxide from the sample. Any water, therefore, must be removed as soon as it condenses. The gas stream must also be cooled to the ambient temperature. The particle filter should trap solid particles but not remove any gas the user wants to measure. The choice of the materials in contact with sampled gas should have low absorption properties.

Figure 11 shows the Minigas Multigas monitor from Neotronics of North America, Inc. The Minigas is a confined space gas detector that can display up to four gases:

- oxygen (0 up to 35%);
- flammable [percent lower explosive level (LEL) 0–99%];
- up to two toxic gases: carbon monoxide, hydrogen sulfide, sulfur dioxide, or chlorine.

The sensitivity of the Minigas for toxic gases varies with the nature of the gas. The unit can detect 0.1 ppm of chlorine and 1 ppm of the other gases. The sampling is either by diffusion or with an optional internal pump.

The Minigas can operate on nonrechargeable batteries (40 h minimum with four AA batteries) or with a rechargeable battery (nicad, 12–15 h; NiMH, 21 h minimum). The dimensions of this instrument are depth 1.75 in (4.6 cm) × width 2.75 in (7.0 cm) × length 7.12 in (18.5 cm). It weighs less than 31 oz (831 g) with the zinc rechargeable battery version. The Minigas is equipped with an alphanumeric liquid display which simultaneously gives the reading of gas concentrations and is programmed to give an audible alarm.

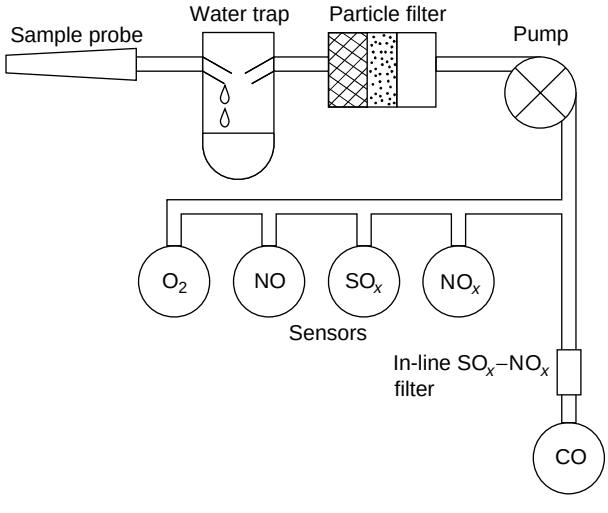

Figure 10 Typical arrangement of a multisensor instrument equipped with sensors for oxygen, nitrous oxide, sulfur dioxide, nitrogen dioxide, and carbon monoxide. In addition the instrument is equipped with a water trap, particle filter, pump, and selective trap to capture SO_x and NO_x compounds and increase the selectivity of the carbon monoxide sensor. (Courtesy of City Technology Ltd.)

Figure 11 View of the Minigas Multigas monitor from Neotronics of North America. The Minigas Multigas personal monitor simultaneously monitors and digitally displays up to four gases. (Courtesy of Neotronics of North America, Inc., Flowery Branch, GA.)

3.2.4 Typical Performance of Gas Electrochemical Sensors

Typical performance data for electrochemical sensors are given in Table 3. Signaling the presence of a hazardous atmosphere is often more important than measuring its precise content, at least for portable monitoring instruments. Safety sensors do not normally need to have a range capability far in excess of the legal exposure limits. It is apparent from Table 3 that these sensors are small and portable.

3.2.4.1 Typical Linearity Curves For low concentrations, i.e. below a few percent of reactant gas, the sensor outputs are linear with respect to reactant concentration (Figure 12a). Figure 12(b) shows the error that results from assuming a linear response over the concentration range on calibrating in air at 20.9%. At concentrations above a few percent, sensors become increasingly nonlinear and require some partial electronic compensation.

Table 3 Typical performances of electrochemical sensors[a]

Characteristic	Oxygen	Carbon monoxide	Sulfur dioxide	Nitric oxide	Nitrogen dioxide
Sensor type	AO2	A3E/D	3SF	3NF/F	3ND
Dimensions[b] (mm)	H, 39; W, 29.3	H, 22; W, 42.5	H, 17.8; W, 42.5	H, 17.8; W, 42.5	H, 17.8; W, 42
Operating life in air (years)	2	3	2	3	2
Maximum range	0–100%	0–4000 ppm	0–5000 ppm	0–5000 ppm	0–1000 ppm
Resolution	0.1%	1 ppm	1 ppm	1 ppm	0.5 ppm
Response time (90%) (s)	<20	<40	<30	<20	<35
Approximate weight (g)	37	31	22	22	22

[a] Courtesy of City Technology Ltd., Portsmouth, UK.
[b] H, height; W, width.

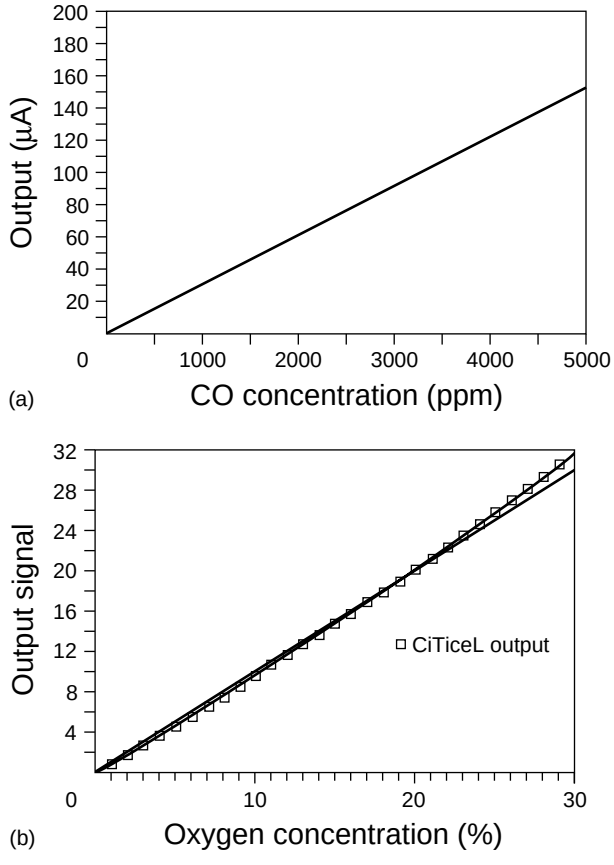

Figure 12 (a) Output signal as a function of concentration of a capillary barrier sensor for 0–5000 ppm of carbon monoxide. (b) Output signal as a function of concentration of a capillary barrier sensor for 0–25% oxygen.[19]

3.2.4.2 Design Parameters Affecting the Selectivity of an Electrochemical Sensor Two key features of sensors are that they be highly specific and exhibit minimal effects from cross-interfering gases. Three possible approaches to reaching these goals are to develop specific electrode catalyst and electrolyte systems, control the operating potential of the SE, and use chemical filters to remove interfering gases selectively. Each method has limitations in practice. For instance, electrocatalysts can be developed that exclude some potential cross-interfering compounds, but it is difficult to develop specific catalysts that are sufficiently reactive toward the less reactive gases and do not respond to the more reactive cross-interfering substances.

The selectivity of the sensor can also be improved by controlling the electrochemical potential of the working electrode. For instance, proper selection of a gold electrode potential allows the determination of nitrogen dioxide in the presence of nitric oxide. Response time, linearity, zero drift, repeatability, sensor stability, and even sensor life depend on sensor design and methodology.

Chemical filters placed in the main gas stream of the input sensor remove cross-interfering gases by selective adsorption, chemical oxidation, or reduction. These reactions can be specific. The main disadvantage of this technique is its short filter life; as the filter is exposed to the full gas stream and reaches exhaustion, it must be replaced or regenerated periodically. An alternative and much more efficient method is to place the filter between the electrode and the diffusion barrier.[28] In this design, the filter material is exposed to the relatively much smaller amount of interferent gas diffusing to the electrode.

Most applications involve ambient air monitoring, for which the background gas is essentially nitrogen plus oxygen and remains constant. However, if the composition changes, recalibration of the sensor to the new environment becomes necessary. This is called the background effect.

3.2.4.3 Limitations of Electrochemical Sensors Operating temperature of these sensors is typically from −20 to 50 °C. This is not a problem for most applications, but the user should be aware that lower temperatures tend to result in longer response times. When the sensors are used outdoors in cold climates, ambient temperatures should be monitored to ensure that they do not fall below the lower limits.

Electrochemical sensors designed to measure specific toxic gases may have cross-sensitivities to other compounds. Response specificity is determined by the semipermeable membrane selected, the electrode material, and the retarding potential (the potential used to retard the reaction of gases other than those being sampled). The primary gases that interfere are those that other electrochemical sensors are designed to monitor, because they are easily electrooxidizable or electroreducible. Gases that cannot be oxidized or reduced, such as methane, do not interfere.

Interference can be eliminated with selective chemical filters or by prescrubbing the sampled atmosphere. Both of these methods have been used by manufacturers for some applications. However, not all interfering gases can be eliminated, and the sampling professional must be familiar with those that may pose a problem.

Poor stability of a potentiostatic sensor is primarily due to the effect of temperature on the inherent background current present in every sensor. This current is due to internal electrooxidation or electroreduction. Most passive electrochemical sensors should not be exposed to the air when not in use, as the water loss from evaporation through the porous membrane will shorten the sensor life.

Sensor longevity is also a function of the gases being tested. Since most chemical sensors depend on a specific chemical reaction to generate a signal, the reactions involved in these sensing mechanisms are often not sufficiently reversible, so that the sensitive material (electrolyte) is used up during the life of the sensor. Sensor lifetime has many determinants, including heat, humidity, dirt, and cumulative gas exposure; the importance of maintenance should not be underestimated. Sensors need periodic recalibration; costs associated with calibration represent the majority of expenses associated with these sensors.

3.2.5 Thermochemical Sensors

Gases and vapors have certain thermal properties that can be exploited during analysis. For the instruments used in industrial hygiene applications, one of two thermal properties, conductivity or heat of combustion, is measured.

The thermal conductivity detector (TCD) is a simple device that operates on the principle that a hot body loses heat at a rate that depends on the composition of the surrounding gas,[8] i.e. the ability of the surrounding gas to conduct heat away from the hot body can be used as a measure of the composition of the gas. In usual practice, a TCD consists of an electrically heated element or sensing device whose temperature at constant electrical power depends on the thermal conductivity of the surrounding gas. The resistance of the sensing device is used as a measure of its temperature. TCDs are universal detectors, responding to all compounds that have a different thermal conductivity than the ambient air. They have large dynamic ranges, of the order of 10^5, and limits of detection of the order of 10–100 ppm for most analytes. TCDs require good temperature and flow control.

The heat-of-combustion detectors, which make up the largest single class of direct-reading instruments for analyzing airborne gases and vapors, measure the heat released during combustion or reaction of the contaminant gas of interest.[8] The released heat is a particular characteristic of combustible gases and may be used for quantitative detection. The heat mechanism in combustion instruments employs catalysts – catalytically heated filaments or oxidation catalysts. This mechanism may measure either resistance change or temperature change via thermocouples or thermistors.

The sensor contains pairs of elements, each containing a platinum wire coil embedded within a catalytic bead bed. An active detector element oxidizes combustible gases, while an inert reference element compensates for changes in ambient conditions. Both elements are in a Wheatstone bridge circuit and the out-of-balance signal produced is proportional to the concentration of combustible gases.

Each heat-of-combustion detector is specific to some combustible gases. Some specificity can be introduced by manipulating the temperature to combust interfering gases or by carefully selecting the oxidation catalyst. Combustible gas instruments also use a semiconductor as the sensitive element. Semiconducting properties can be altered by adding doping impurities. These oxide coatings can be designed to accelerate selectively the oxidation of certain compounds, thus providing more specificity than does a conventional catalytic combustion sensor. Many instruments can detect both combustible gases and toxic gases. When specificities are needed, sensors with the appropriate selective characteristics are installed. The combustible gas sensor has a broad response to various combustible gases and is left unmodified. Selectivity is made possible through different mixtures of oxides and selected operating temperatures.

3.2.5.1 Limitations of Thermochemical Sensors

Thermochemical sensors for combustible gas measurements are used primarily for occupational measurements near sources of combustible material. The CGI was the first direct-reading instrument to be developed. Their initial use was to detect explosive methane in mines. A variety of CGI designs are on the market, ranging from simple sensors to sophisticated microprocessor-based units with multiple detectors. Prolonged exposure to certain substances (e.g. silicones, chlorinated hydrocarbons, and lead- or sulfur-containing compounds) results in an

irreversible decrease in the sensitivity of thermochemical sensors. Some compounds (e.g. halogenated hydrocarbons) can temporarily inhibit performance, but sensors normally recover after a period of operation in clean air.

3.2.6 Other Types of Chemical Sensors

3.2.6.1 Semiconductor-type Gas Sensors Among other types of chemical sensors,[30] the first type of sensor based on adsorption of gases on metal oxide semiconductors was developed by Figaro Engineering over 30 years ago. The gas to be detected is adsorbed on the surface or reacts with the reactive surface oxygen (adsorbed oxygen) of the oxide semiconductor. This induces a change in its surface space-charge layer, which is then converted into a change in electrical resistance of the polycrystalline elements. The gas recognition is carried out through the surface chemical processes; more specifically, the adsorption or reaction sites play the role of receptor for the gases. On the other hand, the conversion of such an interaction to an electrical resistance change is carried out through the physical processes that cause the transport of electrons. This is determined not only by the semiconducting properties of the oxide used but also by the microstructure of the element, such as the grain size of semiconductor particles, and the geometry of the connection between particles.

The metal-oxide semiconductor (MOS) is another category of gas sensor. It is a device in which the conduction between the source and drain is modified by a charge or potential on the gate electrode. Because they are able to detect changes in charge or potential, MOS transistors have been investigated as possible chemical sensors. For example, if ions are absorbed or adsorbed onto the gate, the charge on the gate is modified, which leads to a detectable signal. This can occur by a number of different mechanisms. Either the gas can interact chemically with the gate material and change the carrier concentration on the gate, and hence its Fermi energy, or the gas can induce dipoles in the gate that alter the potential gradients at the gate–oxide interface, or a change in a chemical cell can occur between the gate and the oxide.

Semiconductor gas sensors are based on devices that are manufactured using silicon technologies.[30] Various types of sensors have been developed using either this or MOS transistor technology, in which the material used to fabricate the base is palladium instead of polysilicon[31] and thus has a high selectivity for hydrogen. Sensor technology is based on the ion-sensitive field effect transistor (ISFET), for which an ion-sensitive membrane provides selectivity.[32]

These types of solid-state sensors detect a variety of gases: hydrogen,[33] ammonia,[34] unsaturated hydrocarbons,[35] saturated hydrocarbons,[36] carbon monoxide,[37] hydrogen sulfide,[38] alcohol,[39] arsine,[40] and oxygen.[41] Many reviews of these technologies are available.[42,43]

3.2.6.2 Surface Acoustic Wave Sensors for Gas Sensing Devices that exploit the transmission of SAWs have been used in electronic applications for more than two decades. The principle of operation of an SAW for gases is conceptually simple.[44] An SAW – a periodic deformation to the normal surface – is transmitted across the surface of a solid, which is presented to the atmosphere to be analyzed. If the solid is a piezoelectric crystal (PZX), then the wave can be induced electrically. The transmission of the SAW is sensitive to the mass of material adsorbed on the surface and may be damped by the environment. It is altered by the interaction of gaseous analyte molecules with the surface, usually via an intermediate (coating) layer; the process is monitored by measuring changes in the frequency, amplitude, or phase of the transmitted wave.

There are many reports of SAWs being proposed as gas detectors, and these reports identify the problems associated with the choice of coating, the specificity, the coating thickness, and the obtaining of reproducible amounts of polymer coating on the surface of the sensing element. Applications of SAW devices for the detection of nitrogen dioxide,[45] hydrogen,[46] sulfur dioxide, and hydrogen sulfide[47] have been investigated. Sensitivity is sample dependent and is typically at the level of a few parts per million. Strong effects on the performance of the detectors at ambient temperature have been reported.

SAWs have been used to sense a variety of organic gaseous species such as nitromethane, methanol, propanol, benzene, tetrachloromethane, *N*,*N*-dimethylacetamide, 1-butanol, 2-butanol, diethyl sulfide, 1,2-dichloroethane, isooctane, methanesulfonyl fluoride, and tributyl phosphate.[48]

3.2.6.3 Optical Fiber Gas Sensing The development of inexpensive, high-quality optical fibers for the communications industry has provided the essential component for the development of optical fiber sensors.[20] The physical principle underlying optical fibers is that of total internal reflection. If a material with a higher refractive index (the core) is surrounded by a material with a lower refractive index (the cladding), then repeated total internal reflections result in optical energy traveling along the core. This structure can be planar for waveguides or cylindrical for optical fibers. The evanescent wave is important for sensing applications, because it can interact with analytes outside the light-guiding core by absorption, luminescence, or scattering, thereby modifying the characteristics of the guided light.

There are several advantages for sensor applications, such as small size and flexibility (typically a 200-μm plastic-clad silica fiber can be bent around a 1-cm radius mandrel). The current development of polyimide, amorphous carbon, and metallic coatings means that fibers can operate at temperatures of up to 400 °C. These fibers operate remotely up to several kilometers without significant degradation of performance. In addition, fibers can be multiplexed, meaning that expensive optical equipment can be shared. They are intrinsically safer when flammable or explosive reagents are present.

Most of the disadvantages of optical fiber gas sensors are associated with the multiphase chemical-sensing technique. For instance, the response time of the sensor, which is determined by the mass transport in the reagent phase, is directly related to the thickness of the immobilized reagents; response times of the order of minutes are not unusual. Another common problem is that the sensitivity of the sensor can change when the chemical properties of the immobilized reagents change.

Optical fiber sensors can be subdivided into two types: extrinsic and intrinsic. For extrinsic sensors, the optical fiber merely acts as a light-guiding link between the measurement point and a remote spectrometer. With this approach, the optical properties of the analyte are modified by an added reagent and measured directly. For intrinsic sensors, the fiber, probably in some modified form, is the sensing element. The refractive index or the effect of temperature on the optical transmission can be used for this type of measurement. The examples that follow are illustrative of the capabilities of optical fiber sensing.

Methane has been detected at a level of 700 ppm by an extrinsic sensor and a remote spectrometer at a distance of 1 km from the source. Measurements over even longer distances (up to 5 km) have been demonstrated, showing a decrease in sensitivity but still giving useful data.[49]

Irreversible sensors using immobilized reagents have detected hydrogen cyanide in air within 1 min at parts per million levels.[50] The irreversibility limits its potential applications.

One example of a sensor that is based on luminescence quenching is that for oxygen under partial pressure in the range 0–20% oxygen with an accuracy of 0.4%.[51] Another gas that has been detected by fluorescence quenching is sulfur dioxide in the range 0.01–6%.[52] Nitrogen dioxide, hydrogen chloride, and chlorine have been detected in the 1–10 ppm range using luminescence quenching of a tetraphenylporphine Langmuir–Blodgett film.[53] One important limitation is again the irreversibility of the sensor.

For intrinsic sensors, interaction with the analyte or an immobilized analyte-sensitive reagent can occur only within the waveguide or in its vicinity (by an evanescent wave interaction). Sensors based on the principle of direct propagation of light through the reagent phase have been shown to be reversible up to around 10 ppm for ammonia gas and 5 ppm for hydrogen chloride gas.[54]

Evanescent wave sensors using planar and cylindrical optical waveguides have been described. Ammonia concentrations as low as 60 ppm were detectable within 1 min.[55] In this application, a coated capillary tube was used instead of a solid glass rod to increase the number of reflections with the outer surface and hence the sensitivity. Methane has also been detected without the addition of any further reagent at a concentration level of 5% in air.[56] Another interesting evanescent wave sensor is an integrated optical device designed to sense hydrogen at a level of 20 ppm but with a response time of several minutes.[57]

3.2.6.4 Surface Plasmon Resonance Sensor Surface plasmon resonance can also provide the basic transduction mechanism for optical gas sensors. A surface plasmon is a particular form of electromagnetic wave that propagates along the surface of a metal. It can be excited optically by light undergoing total internal reflection at the surface of a glass substrate onto which the metal film has been deposited (Figure 13a and b). With the proper choice of metal, usually silver or gold with a thickness of a few nanometers, excitation occurs at a particular angle of incidence, leading to a sharp dip in the intensity of the reflected beam at that particular angle. This angle depends on the surface plasmon's resonant frequency and is sensitive to variations in the refractive index of the medium immediately adjacent to the metal surface. The sensitivity of the resonant frequency to variations in the refractive index of the neighboring medium falls

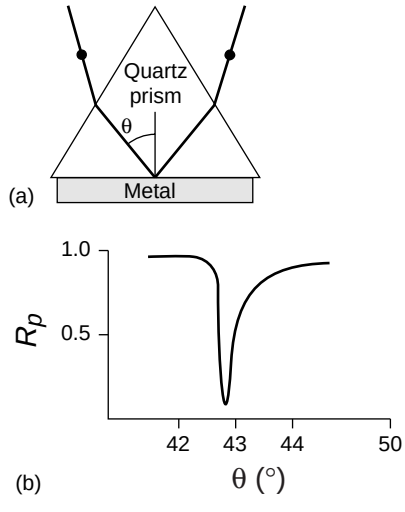

Figure 13 (a) Arrangement for optical excitation of surface plasmons; (b) reflectivity vs. angle of incidence for a surface plasmon resonance.[58]

exponentially as a function of distance from the metal surface, having the same form as and a similar decay constant to the evanescent field interaction discussed previously. This principle of analysis has been applied to anesthetic gas concentrations with a detectable limit of 10 ppm and a response time of 1 s.[58]

3.3 Detector Tubes and Other Colorimetric Direct-reading Indicators

Detector tubes and other direct-reading colorimetric indicators are sampling media that change color when exposed to contaminated gases or vapors. A typical sampler is a glass tube filled with a solid granular material that has been coated with a chemical substance that reacts to change when contaminated air moves through the tube. The primary use of these devices is for occupational sampling, since they are generally not sensitive enough to detect the low levels of contaminants measured in environmental detection.

Three types of colorimetric indicators are used for determining contaminant concentration in air: liquid reagents, chemically treated papers, and glass detector tubes that contain solid indicating chemicals.[59]

Liquids reagents determine alkaline gases by measuring the volume of air required to produce a color change. This method is somewhat inconvenient and requires skill but is capable of good accuracy. Convenient laboratory procedures using liquid reagents have been simplified and packaged for field use. Reagents are supplied in sealed ampoules or tubes and are diluted at the time of use. Unstable mixtures are freshly prepared by mixing the ingredients at the time of the measurement.

The TDI/MDI Analyzer Kit (Sensidyne, Inc., Clearwater, FL, USA) is based on liquid reagent technology. It provides a rapid method for field determination of toluene diisocyanate (TDI) and methylene bis(4-phenylisocyanate) in air. A sample is drawn through a special absorbing solution using the BDX 55 pump and a midget impinger at 2.8 L min^{-1} (0.1 ft^3 min^{-1}) for 10 min. The solution is transferred to a test tube, and a series of reagents are added to produce a blue–red color. The color is compared to a color reference card graduated in parts per million by volume. Results within 0.01–0.35 ppm limits can be obtained in about 30 min. This approach is labor intensive and not geared for fast analysis but should be mentioned since it is the basis of all colorimetric measurements.

Chemically treated papers are used to detect and determine the nature of toxic gases. The method is simple, convenient, and compact. Papers may be used wet, dry, or freshly prepared, as needed. Special chemical chalks or crayons are also used to sensitize ordinary papers to gases. Semiquantitative determinations may be made by hanging the paper in contaminated air. The accuracy of such a procedure is limited by the fact that the volume of air sample is rather indefinite and the degree of color change in the paper is influenced by air currents and temperature. More accurate quantitative results may be obtained by using a sampling device capable of passing a measured volume of air. Visual evaluation of the stains on the paper may be made by comparison with color charts or by photoelectric instruments.

Tape-based instruments make up the next level of sophistication of chemically treated papers that can provide quantitative results. These instruments can be purchased as dedicated systems or as instruments in which different cassettes can be inserted depending on the nature of the gases to be monitored. In either case, the instrument will sample only one chemical at a time and has poor selectivity. An example of a toxic gas monitor using a gas detector cassette tape is shown in Figure 14.

During exposure to a contaminant, the reaction produces a continuous characteristic stain on the paper tape. After exposure, the tape is moved incrementally to an optical reader. The reflected optical density of the stain, which is proportional to the concentration of the contaminant in the air, is compared with the reference reflectance and displayed on a meter. The advantage of the optical reader is that the system is calibrated optically and does not require gases for the calibration of the instrument. The disadvantages of this approach are the intermittent nature of this type of detection and measurement as used with paper tapes and the relatively long response time of the system (typically 60 s).

The model FP-250A from RKI Instruments, Inc. (Hayward, CA, USA) uses cassette tapes impregnated

Figure 14 Gas monitor that uses a gas detection cassette. (Model FP-250A from RKI Instruments, Inc.)

with various reagents to obtain a selective detection. Gases such as hydrogen sulfide, phosphine, and arsine can be identified, but potential interferences will not be detected if both gases are present at the same time. Consequently, this type of instrument is best adapted to monitor the level of toxic gases of a known nature such as phosphine. The minimum detection level also depends on the gas monitored. Arsine, for instance, is detected at the 2.5 ppb level in 20 s, whereas it takes 60 s to detect silane at a 250 ppb level with the same tape.

Detector tubes are available for short-term measurements (grab sample measurements) or long-term measurements (dosimeters). Short-term measurements take typically 1–2 min, whereas long-term collection can take 4–8 h, using either active or passive methods. Active methods involve the use of a portable battery-powered pump or hand-powered pump. Passive methods, such as the badges described earlier, rely on diffusion.

3.3.1 Applications of Detector Tubes

There has been a great expansion in the development and use of detector tubes, and more than 400 different types are now available. Detector tubes are convenient for quantitative and qualitative evaluations of toxic hazards in industrial atmospheres and process control and for the rapid evaluation of spills of hazardous materials.[60,61] Detector tubes are also used for detecting explosive hazards, checking compressed breathing air, and confirming carbon monoxide in exhaled breath or in gas released from a sample of blood. They are used for law enforcement to measure alcohol in the breath and signs of arson. Several manuals provide comprehensive descriptions and listings.[62–64]

Detector tubes are simple, fast, inexpensive and easy to use, but are also subject to many limitations and potential errors. The results may be dangerously misleading, so the sampling procedure should always be supervised and the results interpreted by an adequately trained operator.

3.3.2 Operating Procedure for Detector Tubes

The use of detector tubes is extremely simple. After its two sealed ends have been broken open, the glass tube is exposed to a known quantity of air, using either a calibrated squeeze bellows or a piston pump. The observer then reads the concentration in the air by examining the exposed tube. More recently developed types of tubes interpret the length of the stain produced on the indicator gel rather than color changes. The stain length is measured against a calibration scale that can be printed either directly on the tube or on the provided chart. Selection of the detector tubes depends on the chemical for which monitoring will be done and on the concentration range of interest.

3.3.3 Specificity and Sensitivity of Detector Tubes

Many tubes exhibit cross-sensitivity to other gases, and the manufacturers' literature should be consulted before use. Most tubes are not specific. In the presence of mixtures, the tube reading should be interpreted by a trained operator. The lack of specificity of some tubes may be an advantage for detection of substances not indicated by the manufacturer, but such use requires specific knowledge of the identity of the reagent and of the proper correction to the calibration scales. Most of the tubes are generally designed for detection of relatively high gas concentrations (a few parts per million).

3.3.4 Equipment

Figure 15 shows two models of pump sold by National Dräger. The Accuro bellows pump and short-term detector tubes form a portable sampling unit for use in measuring concentrations of various gases and vapors. Dräger detector tubes are available for measuring approximately 350 air contaminants and for gas analysis. The pump delivers 100 mL of sample air with each pump stroke. After a prescribed number of pump strokes, the stain length or the discoloration of the tubes gives a direct measure of the gas or vapor concentration. Calibration scales are printed directly on most types of tubes. The Accuro is a modular system. For large-volume measurements, the Accuro slides into the electronically programmable Accuro 2000. The Quantimeter 1000 is a programmable, battery-operated bellows pump with the same flow characteristics as the hand-operated pump and is intrinsically safe. The complete Dräger Accuro Deluxe Pump Kit with spare parts and tube opener weighs approximately 1.5 kg

Figure 15 Dräger Accuro, Accuro 2000, and Quantimeter 1000. The figure shows the tubes operating in a manually and electronically programmable pump. (Courtesy of National Dräger, Inc.)

(3.3 lb). The detector tubes are usually specific for particular gases and vapors. This specificity is achieved not only by the use of specific and stable reagents but also by the use of precleansing layers placed in front of the actual reactive layer to absorb selectively interfering components that may be contained in the gas or vapor sample. The reading deviations for many of the detector tubes are not more than ±25% from the true value. The long-term diffusion detector tubes operate on the principle of gaseous diffusion to give a long-term TWA measurement without a pump. The contaminant gas diffuses into the tube by means of the concentration gradient between the ambient atmosphere and the interior of the tube.

Dräger recently introduced a chemical measuring system (CMS) based on a chip, which is in effect the gas detection sensor and allows multiple measurements on one plastic carrier (Figure 16). The reactive preparation necessary for detection is kept in a hermetically sealed glass capillary until needed. The housing of the chip also protects the capillaries from potential external mechanical influences. Prelayers filter out potential interferences from other gases to ensure that the result of the measurement is substance-specific.

Figure 16 The Dräger CMS. The optical reader is shown with a chip containing 10 measurement capillaries filled with a substance-specific reagent system. (Courtesy of National Dräger, Inc.)

Figure 16 shows the analyzer and the chip containing 10 microchannels filled with small quantities of reagents. Small quantities of reagent ensure a quick and reproducible reaction with the contaminant gas or vapor (typically 1–2 min). The analyzer records the reaction products optoelectronically. The measurement principle of the CMS is based on a dose measurement that depends on concentration, i.e. a dynamic dose measurement. The measurement time is not constant but adapts itself automatically to the concentration, i.e. the higher the concentration, the shorter is the measuring time, thereby providing information about possible gas dangers more quickly. For instance, the standard measuring time of nitrogen dioxide at a concentration of 5 ppm is 30–35 s; for a fivefold value this drops to about 10–12 s.

A pump system with mass flow sensor, which is part of the analyzer, draws a constant mass of air; the final results are, therefore, totally independent of external ambient pressure conditions. The analyzer operates on batteries and has an autonomy of 7 h.

The Dräger CMS unit can analyze ammonia, benzene, carbon dioxide, carbon monoxide, chlorine, hydrochloric acid, hydrogen cyanide, hydrogen sulfide, nitrogen dioxide, nitrous fumes, perchloroethylene, and sulfur dioxide. Chips with other specificities are under development.

3.4 Ionization Detectors

General-survey instruments are used to monitor ambient conditions whenever immediate results are needed, even though the identity of the compound present may be unknown or a mixture of compounds may be present. The primary detectors used are the PIDs and FIDs. General descriptions of the fundamental theory behind their operation are available.[65–67]

These instruments are used for general surveying on hazardous waste sites and for industrial hygiene monitoring.[68] They screen air, soil, water, and drum bulk samples and establish priorities for laboratory analysis. They can also be used to determine if decontamination procedures are effective. They are often used in a qualitative manner to see if volatiles are present, but they also can be used quantitatively if properly calibrated. Other applications include leak detection, perimeter monitoring, and continuous monitoring using periodic checks. Another application is boundary line sampling for total hydrocarbons to detect airborne releases to the environment from an industrial operation or hazardous waste site. All these instruments are calibrated for known gases and respond to different compounds with different levels of sensitivity. A response factor is provided with these units, but it requires that the identity of the gas to be analyzed is known.

3.4.1 Flame Ionization Detector

In a conventional FID used in GC, a gaseous sample is pyrolyzed in a hydrogen flame, producing ions and electrons. When hydrocarbons in the sample are introduced to the detection zone, ions are produced by the reaction shown in Equation (6):

$$RH + O \longrightarrow RHO^+ + e^- \longrightarrow H_2O + CO_2 \quad (6)$$

where RH is the carbon compound. Ionized carbon fragments move under the influence of an electric field present between the electrodes surrounding the flame. As the positive ions are collected, a current corresponding to the collection rate is generated on the input electrode. The current is directly related to the hydrocarbon number R present in the sample. The FID response is proportional to the number of carbon atoms present in the sample. Typically, the response of the FID to a C_6 hydrocarbon is six times the response to methane, which has only one carbon atom.

The flame ionization detector for general survey (FIDGS) is different from the FID for GC in design and performance. Instead of being mixed with the fuel and carried away to the jet, the sample is fed into the air sheath around the flame, and thus the sample reaches the flame only by a process of diffusion. This FIDGS is not selective. All hydrocarbons produce almost the same response. The basic design of the FIDGS is demonstrated by the Foxboro (East Bridgewater, MA, USA) model shown in Figure 17.

Instrument characteristics, such as sensitivity, are generally given as methane equivalents with a correcting factor.[65] Water and carbon dioxide molecules are also present and result from the ionization process and from the recombination of oxygen, hydrogen, and carbon molecules, but they do not affect the response of the FIDGS. This detector is insensitive to water, inert gases, and inorganic compounds and has a negligible response to carbon monoxide and carbon dioxide. It is only when the oxygen level is displaced by a high concentration of other gases that the response varies and possibly extinguishes the flame, i.e. when there is <14% oxygen.

Typically, the FIDGS has a wide dynamic range of four to five orders of magnitude, which can be expanded even further by diluting the sample. The dilution kit can also be used to enrich oxygen-deficient samples by adding ambient air that is rich in oxygen, as is modeled by Foxboro's design.

The fuel (H_2) is provided in a small tank under pressure (typically several atmospheres), and oxygen is supplied from ambient air. To keep the background noise low, purified air must be used; ambient air is cleaned by drawing it through a charcoal trap and is then pumped into the detector. The sample is introduced into the ionization chamber by means of a pump.

Figure 18 shows a commercial FIDGS package from PE Photovac (Markham, Ontario, Canada). This model is hand carried and has a built-in hydrogen cylinder. The unit displays the measurement results digitally at

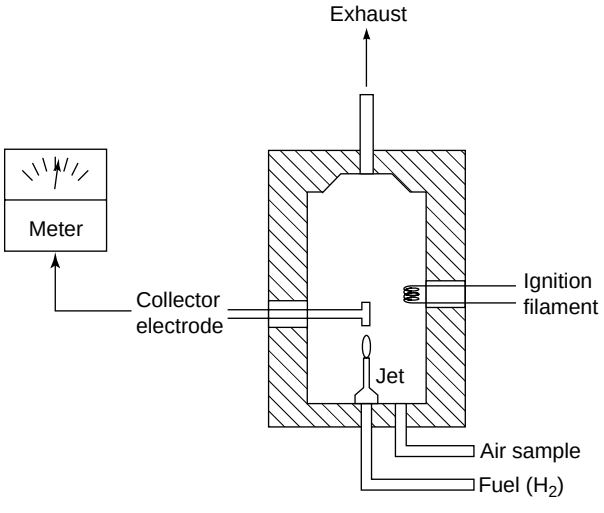

Figure 17 Basic schematic of a flame ionization detector used for general survey (FIDGS). The sample is fed into the air sheath and gives a response that is less dependent on the number of carbons than is the conventional detector used in GC. (Courtesy of Foxboro.)

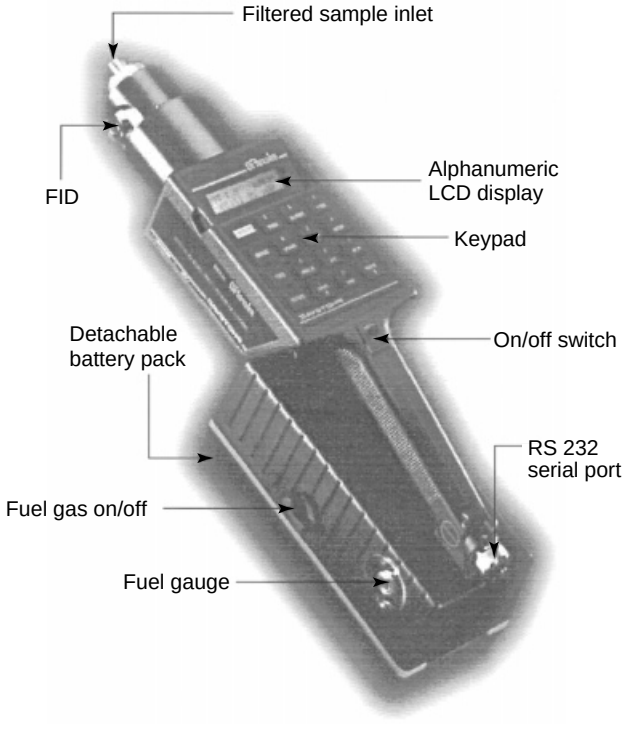

Figure 18 MicroFID from PE Photovac. The unit has a built-in hydrogen cylinder. (Courtesy of PE Photovac.)

Table 4 Micro-FID specifications[a]

Feature	Specifications
Size	43.5 cm (17.1 in) long, 9.8 cm (3.85 in) wide, 18.8 cm (7.4 in) high
Weight	3.7 kg (8.1 lb)
Keypad	16-key silicone with tactile feedback
Display	2-line, 16-character LCD with alphanumeric and bar graph readout
Serial output	RS-232, 9600 baud, for connection with printer or PC
Audio output	On alarm, low battery and FID flameout
Battery capacity and type	15 h, sealed lead–acid (snap-on replaceable pack)
Analog output	0–1 V, full-scale
Hydrogen cylinder discharge time	>11 h
Operating concentration range	0.1–50 000 ppm methane equivalent (two ranges)
Accuracy	Methane (after calibration with 0 ppm air and 100 ppm methane span gas): within ±10% of actual methane concentration (0.1–2000 ppm range)
Response time	<3 s (to 90% full-scale)
Detection limit	0/5 ppm methane
Intrinsic safety	Class I; division 1; groups A, B, C and D

[a] Data from PE Photovac.

concentrations ranging from 0.1 to 50 000 ppm. The readings are displayed in parts per million of methane and require a correcting factor for other gases. The characteristics of the MicroFID from PE Photovac are given in Table 4. This instrument has an autonomy of >10 h on the same batteries and hydrogen tank. When used in combination with a PC, the MicroFID becomes a powerful sensor and complies with EPA Method 21 (fugitive emissions monitoring).

Common applications for a FIDGS include, but are not limited to, the following:

- industrial hygiene surveys and short-term health and safety monitoring;
- leak detection and fugitive emissions monitoring;
- detection of soil gases and headspace of water samples;
- detection of arson;
- solvent vapor extraction (SVE) systems in groundwater redemption surveillance.

3.4.2 Photoionization Detector

Photoionization is a flameless ionization technique. A PID consists of an ultraviolet (UV) lamp of a specific energy and an ionization chamber (Figure 19). Compounds passing through the chamber are excited by photons of UV energy from the lamp and ionized. The resulting ions are collected by an electric field; the current measured gives an accurate representation of the amount of compound present.[69]

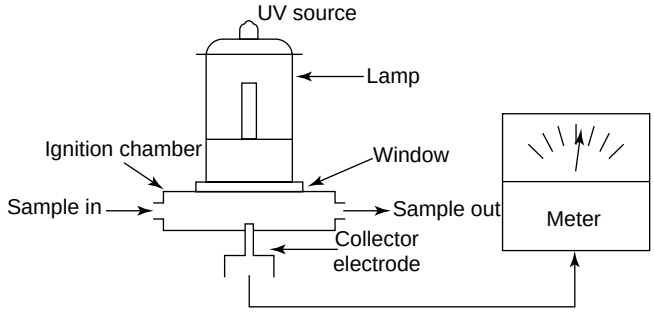

Figure 19 Typical PID. Ions are produced by photoionization initiated by the absorption of a high-energy photon by a molecule or atom. (Courtesy of Foxboro.)

Whether a compound can be detected by a PID depends on the lamp energy and the energy required to ionize the sample (the ionization potential is the energy required to remove an electron from the sample). If the lamp energy is greater than the compound's ionization potential, the PID will detect it.

Lamps of 9.6, 10.0, 10.6, and 11.8 eV are usually available. The 11.8-eV lamp permits the detection of many compounds, whereas the lower-energy lamps allow more selectivity, theoretically, by not responding to undesired compounds with a higher ionization potential. However, since the radiation from the UV source is not entirely monochromatic, compounds with higher ionization potentials will still be detected to some degree with lower-energy specified PID lamps as evidenced by PID/GC analysis of equal concentrations of compounds with both high and low ionization potentials.

Figure 20 shows a Model 2020 hand-held PID unit from PE Photovac. It requires compressed calibration gases. The only maintenance operations are the replacement of the water/particle inlet filter and occasional cleaning of the UV lamp. Table 5 gives the specifications of the Model 2020 PID. This PID is characterized by its small size and low weight. The operating range of the PID is limited to 2000 ppm. The PID is used for the detection of VOCs, hazardous wastes, and fugitive emissions, and all operations related to emergency response. Another application is industrial hygiene, for which it is used to monitor toxic gases and vapors in a workplace environment. The unit can operate automatically for 10 h. The internal data logger retains a complete record of the day's activity.

3.4.3 The Foxboro Portable Toxic Organic Vapor Analyzer

The TVA-1000 is a compact portable instrument for the simultaneous analysis of organic and inorganic

Figure 20 Hand-held Model 2020 PID from PE Photovac. (Courtesy of PE Photovac.)

compounds contained in the air (Figure 21). The instrument uses individual or concurrent PID and FID in the same package. The combination of detectors with fundamentally different responses provides unique response ratios for additional qualitative information about the analytes.

Each unit is factory-calibrated with methane (in the case of the FID) or isobutene (in the case of the PID). However, both detectors respond to many different compounds with different levels of sensitivity. To adjust the analyzer reading to the compound of interest, a correction factor must be entered into the analyzer. The result is compound-specific (corrected) readings. The specifications of the TVA-1000 are given in Table 6.

3.4.4 Comparison of the Flame Ionization Detector and Photoionization Detector

The FID and the PID both have advantages and disadvantages and complement rather than compete with each other.[70] The FID has a wider operating range than does the PID (0–50 000 compared with 0–2000 ppm) and is, therefore, well suited for measuring high concentrations. The FID is also less susceptible to water vapor interference than is the PID.

The PID does not require hydrogen to operate and is the detector of choice if fuel gas is limited or not available.

Table 5 Specifications of the hand-held PID model 2020[a]

Feature	Specifications
Weight	0.7 kg (1.75 lb)
Size	24.8 cm (9.75 in) long, 7.6 cm (3 in) wide, 6.6 cm (2.6 in) high
Detector	Photoionization, with quick-change electrodeless discharge tube 10.6 eV (standard); 11.7, 10.0, 9.5, and 8.4 eV available
Keyboard	6-key silicone with tactile feedback
Display	(1) 4-character, 1.2 cm (0.5 in) high backlit LCD for numeric; (2) 2-line, 16-character backlit LCD for alphanumeric status readout
Battery power	7.2 V, field-replaceable nickel–cadmium or AC operation from charge
Battery charge–discharge time	4 h charge/10 h discharge
Data-logging memory	1000 data points (16K)
Serial output	RS-232, 9600 baud
Audio output	Audio alarm on user-settable concentration limits (TWA, STEL, peak)
Inlet filter–water trap	Replaceable, Teflon®–polypropylene, 0.1 mm
Inlet flow rate	>300 mL min^{-1}
Operating temperature range	0–40 °C (32–105 °F)
Operating humidity range	0–100% RH (noncondensing)
Operating concentration range	0.5–2000 ppm isobutene equivalent (two ranges)
Response time	<3 s (to 90% full-scale)
Accuracy	±10% or ±2 ppm
Intrinsic safety	Class I; division 1; groups A, B, C and D; intrinsically safe for use in zone 1 locations; Eex ibm IIC T4

[a] Data from PE Photovac.

Thus the overall PID package is typically smaller than one with an FID. It can also operate as a selective detector, depending on the ionization lamp selected. The PID is also sensitive to aromatic and chlorinated compounds and can measure some inorganic compounds that the FID does not detect at all (e.g. ammonia, carbon disulfide, arsine, phosphine, carbon tetrachloride, chloroform, ethylamine, formaldehyde, and hydrogen sulfide).

The two detectors operating simultaneously in a single instrument offer a net advantage for survey monitoring by providing information quickly about organic and inorganic vapors.

The responses of the two detectors relative to each other help identify the compounds being analyzed. For instance, the PID does not respond to methane at all, but the FID responds well. A high FID reading with virtually

Figure 21 The Model TVA-1000 portable toxic vapor analyzer, from Foxboro combines an FID and a PID in the same package and can operate both detectors simultaneously. (Courtesy of Foxboro.)

Table 6 Specifications of the Model TVA-1000[a] portable toxic vapor analyzer

Feature	Photoionization	Flame ionization
Safety certification	Factory mutual; class I; division 1; groups A, B, C, and D; CENELEC Eex ibd IIC T4	
Data logging	Onboard	
Readout	Bar graph and 4-digit LCD	
Operating range	0–2000 ppm	0–50 000 ppm
Response time	<3.5 s	
Minimum detecable level	100 ppm benzene	200 ppb hexane
Alarms	Low, high, STEL	
Sample flow rate	1000 mL min^{-1}	
Power supply	Rechargeable nickel–cadmium battery	
Fuel	Not required	99.5% hydrogen
Portable operation time	8 h	
Approximate mass	5.5 kg (12 lb)	
Nominal dimensions	33 cm × 25 cm × 8 cm (13.5 in × 10.3 in × 3.2 in)	
Analog output	0–2 V	
Repeatability	±1%	
Automatic autoranging	Yes	
Diagnostics	Yes	

[a] Courtesy of Foxboro Co.

no PID response may indicate the presence of a naturally occurring methane, commonly found at excavation sites. The PID responds well to some inorganic gases that the FID does not detect. A high PID reading with no FID reading may suggest the presence of an inorganic compound.

4 GAS CHROMATOGRAPHY FOR ENVIRONMENTAL MEASUREMENTS

The introduction of gas–liquid chromatography in the 1950s and its subsequent enhancements coincided roughly with the increased demand for quick, reliable, and sensitive analytical methods for the determination of the many chemical compounds that contaminate the environment. GC proved invaluable for many of these applications and continues to play a major role. GC is a physical method of separation that relies on differences in solubility of components distributed between a mobile and a stationary phase. In the GC procedure of interest here, a sample is vaporized and injected into a mobile phase such as hydrogen, helium, or nitrogen and passed into a column containing the stationary (solid or liquid) phase. The components of the sample demonstrate varying affinities for the stationary phase, thus slowing (to varying degrees) their passage through the column. In the perfect case, each sample component has a different retention time in the column and is thereby separated from all others. The magnitude of the signal that each component yields as it emerges from the column and enters the detector is a measure of its concentration.

Two objectives are paramount for any analysis using GC: the data must be accurate and precise and they must be produced in a minimum of time. Sampling and separation/measurement are more important in some cases than in others. In some instances, it is more difficult to obtain a representative sample than to analyze it. GC can reduce the difficulty in obtaining a representative sample because the instrument is small enough to be brought to the field, thus eliminating errors caused by the degradation of samples during transport.

This section details some practical hints about chromatographic hardware, choice of columns, selection of injection systems, and choice and performance of detectors. Several portable gas chromatographs are described. Coupled techniques such as GC/MS are described later.

4.1 Basic Components of a Gas Chromatograph

A basic chromatograph may be extremely simple (Figure 22).[66] It consists of a supply of carrier gas (commonly helium, nitrogen, or hydrogen) with appropriate flow regulation. In this example, the sample is introduced by means of a hypodermic syringe through a rubber septum cap into the carrier gas stream, where it is rapidly

vaporized at the head of the column before passing through the column, where it is separated. The column, in turn, is housed in a suitable temperature-controlled oven. Finally, a detector is provided to indicate when the components are eluting from the column. This detector signal, when recorded as a function of time, produces a chromatogram.

The time required for each of the components to elute is called the retention time, and the magnitude of the detector signal (peak height and peak area) depends on the amount of component present in the sample (Figure 23). In this figure, the unretained peak (usually air) is defined as t_M, and all other retention times are corrected from t_M since this time is a characteristic of only the void volume of the column and has nothing to do with chromatography (Equation 7):

$$t'_R = t_R - t_M \tag{7}$$

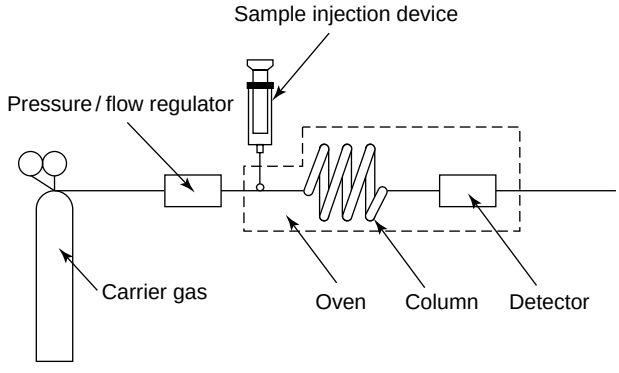

Figure 22 Schematic of a basic gas chromatograph.[66]

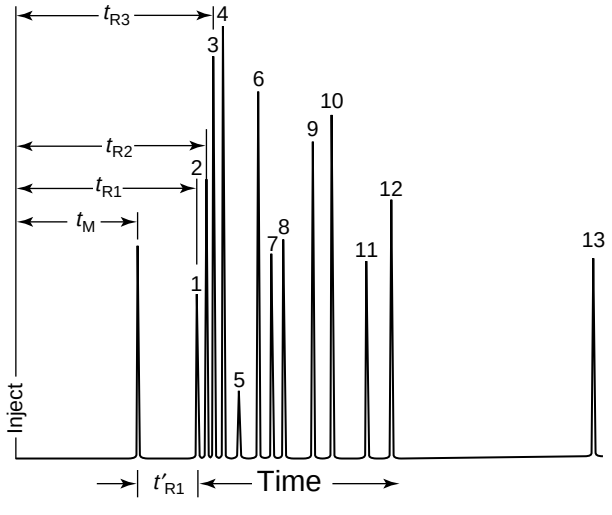

Figure 23 Typical capillary column chromatogram with retention times of air sample components 1, 2, and 3, labeled t_{R1}, t_{R2}, and t_{R3}, respectively.[66]

4.2 Classification of Gas Chromatographs Used for Field Measurements

The GC field measurement is only one part of the overall analytical procedure, which includes sample collection, sample extraction, sample concentration, sample derivatization (conversion of a chemical to another chemical), sample separation (in the separation column), and the data collection and reduction necessary to interpret the chemical information. The selection of the field instrument to be used must be based on the understanding of the complexity of the application. The instrument can be hand-portable, portable or transportable. Each type has its own features.

Portability implies that the total weight and size of the instrument are within some reasonable limits, i.e. easily carried by the average person. Versatility roughly depends on the instrument's weight and size; the smaller the instrument, the more specialized it will be. Thus the difference between portable and nonportable gas chromatographs depends on the manufacturer's chosen trade-offs between size and analytical capability. Analytical capability depends on the features available – type of columns, maximum temperature of the overall analytical system, type of detectors, and type of injectors. Therefore, when comparing portable gas chromatographs, it is important to understand the trade-offs that the manufacturer makes between analytical capability and portability. The analysis of gaseous samples is the simplest of field measurements, since most of the time sample preparation is not required. Hand-portable gas chromatographs fit this category and are oriented exclusively to the analysis of gases. They are light, have limited analytical capabilities (but can be sensitive), and are tailored to give a quick answer to a specific application.

A typical hand-portable gas chromatograph is a self-contained unit that includes its own gas supply and a battery pack for autonomous operation. Figure 24 shows an example of field measurement using a hand-portable instrument during an emergency situation. In this application the instrument must identify the presence of a volatile chemical without disturbing the site. No front-end sample preparation is necessary and only the vapors coming from the suspected source of contamination are analyzed. A typical analysis with a sub-parts per million sensitivity can be accomplished in a matter of minutes, but this requires some knowledge about the nature of the sample to be analyzed. There is a high probability of a miss if the hand-portable instrument is not tuned to the compound present.

The field analysis of medium/boiling-temperature volatile samples by GC is the next step of difficulty. This analysis requires that the pneumatic lines of the instrument be heated to a temperature of up to 200 °C

Figure 24 Field intervention using a hand-portable instrument. (Courtesy of PE Photovac.)

to avoid condensation of samples. The instrument has more features than a hand-portable instrument, but fewer than a laboratory instrument. For instance, it can accept liquid samples, and the column can be temperature-programmed to reduce the analysis time (but at the expense of the size and weight of the instrument).

Some basic sample extraction or sample preparation is also possible before analysis. The instrument can also be run directly at the field site, provided electrical power is available [alternating current (AC) line or portable power generator]. The portable gas chromatograph carries its own gas supply. Figure 25 shows the field operation of a portable gas chromatograph from HNU Systems, Inc. (Newton Highlands, MA, USA). This instrument is smaller than a laboratory instrument and can be carried by an operator in the back of a station wagon. This portable instrument provides a broader knowledge about the sample to be measured than does a hand-portable gas chromatograph, but it is bulkier and, therefore, more difficult to use for spot checks in the field.

The most elaborate and more expensive approach to quality measurements in the field is to use conventional laboratory instruments operating in a mobile laboratory. A mobile laboratory can be equipped with most of the analytical tools found in a conventional laboratory. Laboratory analytical instruments, such as gas chromatographs

Figure 25 Portable instrument used for field intervention. (Courtesy of HNU Systems, Inc.)

and mass spectrometers, are housed in commercial vehicles that have been modified to accept these instruments and to allow sample preparation. Depending on the vehicle, it can be modified to accommodate wet chemistry, analytical evaluation, and data handling at the same time. Such vehicles also provide a shelter and climate control for the operator and the equipment, and can be designed to operate in extreme weather conditions.

Figure 26 shows a mobile laboratory modified by Centrum Analytical Laboratories (Redlands, CA, USA). This van offers more than 110 ft^2 of laboratory space. The vehicle is air-conditioned and can provide up to 16 kV A of electrical power via diesel generators. Gas cylinder storage is provided for the transportation of compressed gases. This particular unit can accommodate a sample

Figure 26 Mobile laboratory unit from Centrum, Inc. This vehicle can be equipped with the same type of analytical equipment used in a conventional laboratory. This view shows the vehicle with some analytical equipment. (Courtesy Centrum, Inc.)

preparation unit, several analytical instruments, and a computing facility, including Ethernet connections. Its full-blown version can handle gaseous, liquid, and solid samples and provide valuable field measurements with the highest-quality results. It can also be remotely connected to the outside world via wireless communication.

4.3 Categories of Analysis

Analysis can be categorized by the range of concentration of the component being analyzed – major components (5–10%), minor components (100 ppm–5%), and trace analysis (1 ppb–100 ppm). Only the last category will be considered. Trace analysis for which the sample concentration is <100 ppm is certainly the most difficult to perform and is routinely required for environmental analysis. Methods of trace analysis require the highest level of instrumentation sensitivity and sample preparation to enhance the concentration level of the analyte. Analytical procedures for air or gaseous sample matrices are required for three types of analysis: air emission methods for stationary or point sources such as incinerators, power plants, and various industrial processes; ambient air monitoring methods; and methods for monitoring indoor air contamination in the workplace. Air emissions and ambient air monitoring fall within the jurisdiction of state agencies and several offices of the USEPA. Indoor air monitoring falls into the realm of industrial hygiene monitored by OSHA. The basic principles for the sample preparation and analysis of industrial hygiene samples are the same as those applied to outdoor air pollution.[71]

4.4 Trace Analysis of Gaseous Samples

With the sophisticated analytical instruments available, it is easy to lose sight of the importance of sample preparation techniques. Even the most powerful of the analytical instruments will be useless with a sample that is improperly collected or prepared.

Gases and vapors diffuse rapidly, mix freely with the general atmosphere, and can reach equilibrium in a short time. They therefore offer fewer difficulties in air sampling than do aerosols. For air sampling purposes, contaminants can be grouped with respect to solubility and vapor pressure. Many gases and vapors of hygienic significance are water-soluble and can be collected in aqueous media with or without a dissolved reacting chemical to suppress the vapor pressure of the solute. Gases and vapors that are not water-soluble but are soluble or reactive in other agents can be adsorbed in a suitable solvent. Gases and vapors that are neither soluble nor reactive may be collected on adsorbent (e.g. activated charcoal, silica gel, or molecular sieves) in either active or diffusive (passive) samplers. Adsorbents have become the sample collection medium of choice for all gases and vapors because of their convenience and generally high collection efficiency.

As indicated previously, analysis is the final step of a complex process that depends on the nature of the sample, the matrix in which the sample is found, the level of concentration of the sample, and the availability of a field instrument. In the most general case, the steps preceding analysis are identified as sample collection, sample concentration, sample extraction, and sample derivatization. It is only after all or some of these steps are completed that the sample can be analyzed. Sample collection can be accomplished just before the analysis by drawing the sample through a sampling line connected to the gas chromatograph. The probe itself can be the gas chromatograph.[72] Tedlar bags or appropriate containers (glass or inert metal) can also be used to collect samples that are then transported to a laboratory for analysis. The use of evacuated stainless-steel canisters to sample VOCs is a standard procedure in the industry.[73] These approaches are applicable to clean samples, such as hydrocarbons in air. The drawback of sample storage is the risk of sample degradation, either by adsorption on the walls of the container or by selective adsorption in the trap, creating a change in the relative concentration of the compounds present in the sample.

When the sample is part of a complex matrix, whether gaseous, liquid, or solid, the extraction stage consists of isolating the vapor sample from its matrix. This extraction stage is usually accomplished thermally (headspace, purge and trap, thermal desorber, pyrolyzer). It is seldom done directly at the collection point, since extraction equipment is not readily transportable to the field. Care should be taken in evaluating the analysis as the extraction process can degrade the sample and introduce a bias.

Solid adsorbents are useful for trapping a gaseous sample before analysis. This trapping approach also offers the advantage of concentrating the sample at the time it is stored, by passing a known quantity of air through the adsorbent and retaining only the sample of interest. The analyte collected on a solid adsorbent can then be desorbed with an organic solvent. This sample-rich solvent can then subsequently be injected into the gas chromatograph. The analyte may also be concentrated by partitioning it into a liquid phase coated on a support material. Impingers can be used to concentrate analytes in a gaseous matrix by passing the sample through an appropriate solvent that retains the analytes but allows the gas to escape. The solvent can be further concentrated or injected directly into the gas chromatograph. The adsorption is usually sample- and volume-dependent. All these devices must be carefully prepared for the samples of interest when quantitative measurements are performed. Sorbents and impingers were reviewed earlier; solid adsorbents have been described in the literature.[74]

Derivatization is an additional preparation step that is used less frequently. It consists of chemically converting an analyte into another chemical species to improve stability, sensitivity, or instrument performance. One popular application of derivatization techniques is to convert a compound that does not respond to an electron capture detector (ECD) into a product with electronegative properties and a favorable ECD response.[75] Another example of derivatization is the conversion of carbon monoxide or carbon dioxide to methane before introduction to the FID.

The following general discussion considers aspects of sample preparation procedures so that the analyst can evaluate practical applications and identify the logic behind the various modifications to the procedures.

4.5 Solid Extraction Thermal Desorber Unit

The use of solid sorbent sampling is illustrated with an application of a field-operable unit for VOC analysis, developed by Louisiana State University.[76] The unit includes two traps and provides an enrichment factor up to 1000 : 1. The first trap provides the concentration of analytes on a large breakthrough trap (the concentration trap), desorption of analytes on a low volume desorption plug trap (the focusing trap), desorption of analytes into the analyzer instrument in a narrow plug, and back-flushing of the system between analyses to prevent cross-contamination.

Figure 27 illustrates the VOC sample processing device. The concentrating trap can be exposed to large volumes of gases, up to 2 L, on a low thermal mass concentrating trap packed with hydrophobic adsorbents. Breakthrough volumes are, of course, compound-dependent, but with dual-bed series, packed traps are typically in the 1–2 L range for common VOC analytes when using Tenax® and carbon molecular sieve adsorbents. Water in the sample is not significantly retained by these sorbents, and thus its concentration is reduced in the extracts. Since the desorption plug from the concentrating trap is several milliliters in volume (typically 5–10 mL), the analytes in this plug are reconcentrated on a small and low thermal mass focusing trap into a final volume of approximately 300 μL. This small concentrated plug of analytes, typically 1 μL and generally devoid of large quantities of water, is then passed to an appropriate analyzer for qualitative and quantitative determinations.

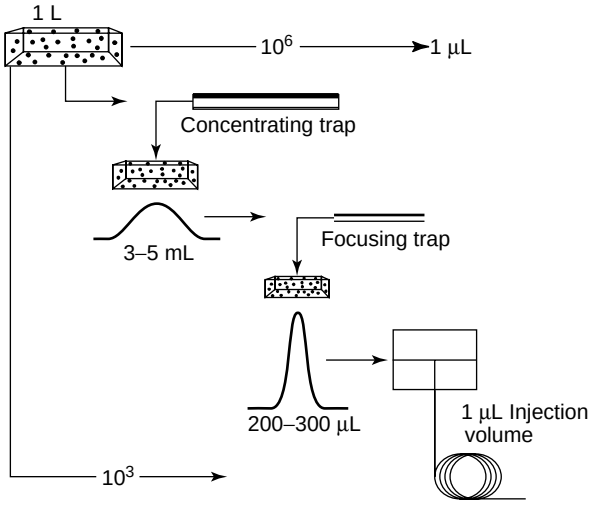

Figure 27 Pneumatic configuration of a two-stage preconcentration unit. The first trap is used for concentrating several liters of air and the second trap for focusing the sample in a plug of a few hundred microliters. (Courtesy of Analytical Specialists Inc.)

The unit manufactured by Analytical Specialists Inc. (Baton Rouge, LA, USA) is microprocessor-controlled and can be operated remotely. The total amount of power necessary to operate the unit and desorb the samples from the traps is approximately 50 W. Furthermore, since it is a dual-trap system, selective adsorbent chemistry can be used to enhance performance in special applications.

4.6 Transfer Lines

The use of fluidic (no moving mechanical parts) sampling and column-switching valves in GC has been reviewed.[77] Enviroprobe, the air probe developed by the University of Utah and commercialized by FemtoScan (Salt Lake City, UT, USA), is an example of this approach (Figure 28).[78] The Enviroprobe operates with a small vacuum pump and provides the functions of collection, injection (from 20 ms), and separation [column of 100 μm inner diameter (ID), 1 m long]. The unit output is connected to another analytical instrument, here a portable mass spectrometer.

4.7 Airborne Pollutants

Ash or fly ash, the solid debris that remains after incineration processes, is another class of solid sample material that is of environmental concern. These airborne and potentially carbonaceous particles contribute to visibility reduction and have complex chemical compositions that include carcinogenic and mutagenic organic compounds. The chemical composition and structure of particle organic matter can provide important information regarding origin, distribution, and fate of respirable aerosols of small dimensions (typically 10 μm). Although current regulations for these types of small particle matter (PM_{10}) are concerned only with total particle densities over a period of time (typically 24 h), future regulations must address the obvious link between their chemical composition (e.g. heavy metal, carcinogen, or allergen content) and associated health risks.[79] Permits for PM_{10}

Figure 28 View of a transfer line (Enviroprobe) connected to a portable mass spectrometer. The unit integrates several functions of sample collection, sample injection, and sample separation. (Courtesy of FemtoScan Corp.)

emissions for specific particles must be obtained from the USEPA.[80] Another subset of particle matter is $PM_{2.5}$, particles with mean aerodynamic diameters $< 2.5\,\mu m$. The USEPA is considering adding a $PM_{2.5}$ standard, primarily to protect vision.

Numerous organic compounds may be adsorbed on particles as they enter the atmosphere and contribute to air, soil, and water pollution. These organic compounds are difficult to extract from particulate airborne pollutants.

Many of the same compounds that must be determined in soil and water samples must also be determined in airborne pollutants. The major procedural difference required for the analysis of air samples is concerned with the collection of the sample itself. The GC portion of the analysis for a given class or group of compounds is essentially the same as that described for the analysis of extracts prepared from soil and water samples.

4.8 Collection and Analysis of Airborne Pollutants

Methods for particle collection and separation are still empirical. A field procedure will be used as an example of a fieldable system that encompasses sample collection, sample transfer, and sample analysis units, including a gas chromatograph and a mass spectrometer operating in a mobile laboratory unit.[81]

Atmospheric particles with an aerodynamic diameter $< 10\,\mu m$ (PM_{10}) are obtained by drawing an aerosol through an impaction jet and depositing larger particles on a plate immediately in front of the nozzle. The impactor is contained in a stainless-steel delivery tube to channel the air flow to a filter holder and a particle counter. Particles are collected in a clean filter. The sampling port is raised approximately 1.5 m (5 ft) above the roof of a transportable laboratory module (Figure 29).

Aerosol concentration and particle-size distribution measurements (parallel to aerosol collection on quartz fiber filters) are performed at 10-min intervals over eight size-resolved channels covering the 0.3–10-μm diameter range using a multichannel particle counter.

Samples are positioned inside a special glass reaction tube lined with a ferromagnetic foil. VOCs and SVOCs are desorbed by flash desorption at a temperature of 315 °C and nonvolatile organic matter is pyrolyzed with a Curie-point desorption–pyrolysis reactor at a

Figure 29 Mobile laboratory module on a three-quarter-ton pick-up truck used for the collection of PM_{10} particles. (Courtesy of the Center for Micro Analysis and Reaction Chemistry, Salt Lake City, UT.)

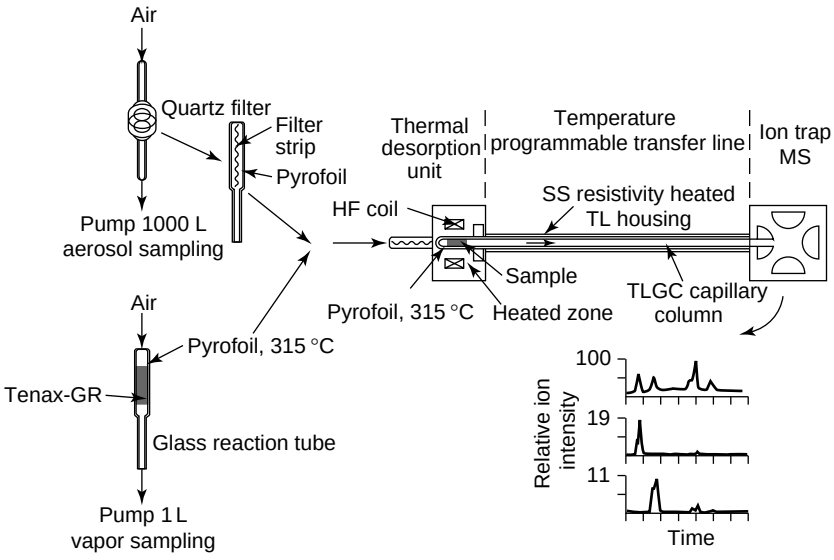

Figure 30 Schematic representation of the PM_{10} (aerosol) and VOC (vapor) sampling modules and the desorption GC/MS system used for analyses.[82]

temperature of 650 °C. A schematic representation of the unit is shown in Figure 30.

Volatile analytes from the reactor are transferred into the mass analyzer via a short capillary column (transfer line) that can be temperature-programmed up to 20 °C s^{-1}. The capillary is inserted into the ion trap mass spectrometer.

4.9 Theory of Packed and Fused-silica Capillary Column Operation

The introduction of inert fused-silica columns markedly changed the practice of GC, allowing high-resolution chromatography in most laboratories. A further decrease in the use of packed columns occurred after the introduction of megabore capillary columns of 0.53 mm ID that, with their increased sample capacity, served as direct replacements for the packed column. These developments, in conjunction with the emergence of immobilized or cross-linked stationary phases especially tailored for fused-silica capillary columns, have been responsible for the greater acceptance of capillary GC.

The use of packed columns represents 20% of GC analysis for laboratory applications. The percentage is even smaller for field applications. Packed columns are employed primarily for preparative applications, for fixed gas analysis, and in those separations for which high resolution is not always desirable. Packed columns will continue to be used for GC methods that were previously validated on packed columns and for which the time and cost of revalidation on capillary columns would be prohibitive.

The introduction of commercially available fused-silica capillary columns in late 1979 followed an investigation of the applicability of fused silica in fiber-optic technology to chromatographic columns.[82] The inherent strength and flexibility of fused silica make it easier to use and less fragile than silicate glass capillary columns. In addition, fused silica provides a more inert surface for improved capillary performance and less adsorption of active components.

Uncoated fused-silica capillary tubing is susceptible to moisture in the air. Water molecules attack the silicon–oxygen bonds, forming silanol groups and opening cracks that propagate into fractures. Protecting the silica from moisture and scratches is the main reason for the outer coating of either polyimide or some other equally resistant material.

A good understanding of the use of capillary columns begins with an understanding of the terminology. Perhaps the most important terms to become familiar with are selectivity, resolution, and efficiency. Further information on the theory of capillary GC and its practical implications can be found in the literature.[66,67,83]

4.9.1 Column Selectivity

The critical step in developing a GC method is the selection of a stationary phase for which each sample component has a different affinity, thus ensuring a separation in the column. Analysis on a large number of

phases has been reported through the years and has been reviewed in the literature.[84,85] By far the best method for classifying stationary phases was introduced in 1966[86] and later modified.[87] In this procedure a group of compounds (benzene, butanol, 2-pentanone, nitropropane, and pyridine) containing different functional groups are chromatographed on the phase to be classified, and their Kováts retention indices I are recorded. The differences in the indices for the compounds from those recorded on the squalane stationary phases are the Rohrschneider indices for that phase and are a measure of the phase's selectivity to the particular functional group contained in the probes. For the probes listed in Table 7,[66] the constants are denoted by x', y', z', u', and s'. Some of the more popular phases, along with their modified Rohrschneider indices, are listed in the table. It should be noted that a casual inspection quickly reveals that there is no significant difference in the selectivities of the older SE-30 phase and the newer highly purified methylsiloxane OV-1.

4.9.2 Kováts Retention Indices

A uniform expression for retention data that is relatively free of instrumental contribution is given by the Kováts retention index system.[88]

The retention index I of a compound is descriptive of its elution position between two saturated hydrocarbons. The retention index of an *n*-alkane is assigned a value equal to 100 times its carbon number. Thus, for instance, the I values of *n*-octane, *n*-decane, and *n*-dodecane are 800, 1000, and 1200, respectively; by definition, they are applicable to any column (packed or capillary) and any liquid phase and are independent of every chromatographic condition, including column temperature. For all compounds other than *n*-alkanes, the chromatographic conditions, such as the stationary phase, film thickness, and column temperature, must be specified.

The I value of a component can be determined by spiking a mixture of *n*-alkanes with the component(s) of interest and analyzing the resulting mixture via chromatography under specific conditions. A plot of log-adjusted retention times versus retention indices is generated, and the retention index of the solute under consideration is determined by extrapolation. To eliminate graphical errors, the retention index of an analyte a at an isothermal column temperature and for a stationary phase S can be calculated from Equation (8):

$$I_a^S = 100N + 100n \left(\frac{\log t'_{R,a} - \log t'_{R,N}}{\log t'_{R,(N+n)} - \log t'_{R,N}} \right) \tag{8}$$

where $t'_{R,a}$ is the adjusted retention time of the component under consideration, $t'_{R,N}$ is the adjusted retention time of the *n*-alkane eluting before the component, $t'_{R,(N+n)}$ is the adjusted retention time of the *n*-alkane eluting after the component, and N is the carbon number of the *n*-alkane with retention $t'_{R,N}$.

The concept of a retention index has been extended to programmed GC.[89] The calculation of temperature-programmed I values, as shown in Equation (9), involves a linear function of unadjusted retention times rather than the logarithmic function used to calculate isothermal I values. The temperature-programmed index of analyte a on liquid phase S and at temperature program p is given by Equation (9):

$$I_{a,p}^S = 100N + 100n \left(\frac{t'_{R,a} - t'_{R,N}}{t'_{R,(N+n)} - t'_{R,N}} \right) \tag{9}$$

where $t'_{R,a}$, $t'_{R,N}$, and $t'_{R,(N+n)}$ are the retention temperatures of the component and the *n*-alkanes that bracket the component. Retention indices normalize instrumental variables in gas chromatographs, allowing retention data generated on different systems to be compared. Libraries of Kováts indices for large numbers of compounds on different liquid phases (polar and nonpolar) are available and allow the chromatographer to identify unknown compounds rapidly.[90] The dual-column approach permits the simultaneous gathering of data on two stationary-phase columns of different polarities. One advantage of this approach is greater confidence in the identification, since compounds that co-elute on one of the columns may well be separated on the other.

Table 7 McReynold-modified Rohrschneider constants for commonly used stationary phases[a]

Stationary phase	x'	y'	z'	u'	s'
SE-30	15	53	44	64	41
OV-1	16	55	44	65	42
OV-17	119	158	162	243	202
OV-25	178	204	208	305	280
Carbowax 20M	322	536	368	572	510

[a] From Annino and Villalobos.[66]

4.9.3 Relative Retention and Selectivity

Relative retention α is a measure of the selectivity of the column. It is defined as the ratio of adjusted retention times (Equation 10):

$$\alpha = \frac{t'_{R,j}}{t'_{R,i}} \tag{10}$$

where $t'_{R,j} > t'_{R,i}$. The selectivity factor α is a measure of the physicochemical interactions between the solutes and the stationary phase.

In the older literature, retention was reported as relative retention with respect to some arbitrary standard. However, it is much more sensitive to a number of variables, such as temperature and flow rate, than is the Kováts retention index and has been abandoned as a reporting tool. However, it is still being used as the selectivity variable in the second critical step of designing a chromatographic method, i.e. calculation of the necessary length needed for a given separation.

In capillary chromatography, another measure of column separation power is the separation number, or Trennzahl (*TZ*). The Trennzahl is defined as the resolution of two consecutive members of a homologous series differing by one methylene unit. The Trennzahl is related to resolution *R* by Equation (11):

$$TZ = \frac{R}{1.177} - 1 \tag{11}$$

Practically, the separation number is a measure of the number of peaks that could be separated between the two consecutive homolos, or the effective peak number. The calculation of Trennzahl is illustrated in Figure 31(a) and (b).[91] The separation number for the capillary column calculated from the C_{12} and C_{13} alkanes is more than three times greater than that for the packed column. The capillary column is capable of resolving 21 peaks between C_{12} and C_{13}, the packed column only six.

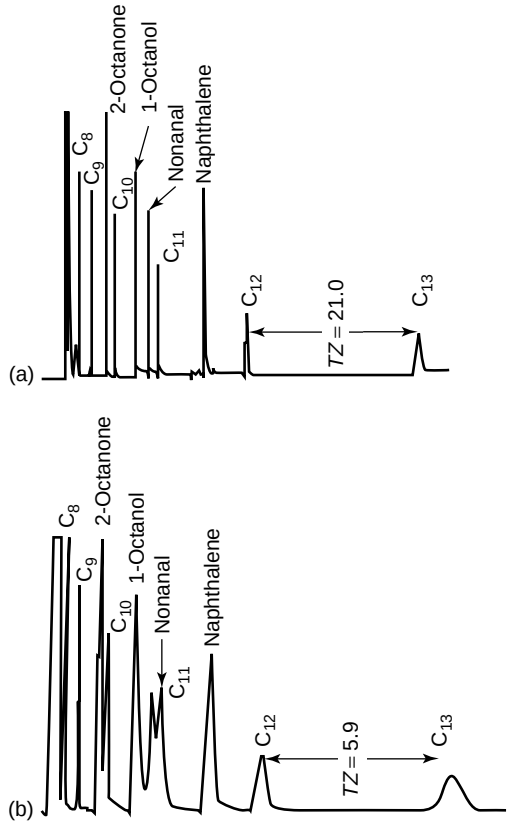

Figure 31 Separation numbers, calculated for (a) capillary and (b) packed columns. Capillary: 10 m × 0.25 mm coated with methylsilicone, helium 3.5 cm s^{-1}. Packed: 1.2 m × 2 mm coated with 3% OV-101 on 100–200-mesh Chromosorb W-HP, helium 30 mL min^{-1}. (Courtesy of Agilent Technologies.)

4.9.4 Column Efficiency

The simplest form of the fundamental equation underlying the relationship between *h* as a function of the longitudinal diffusion and its resistance to mass transfer at low mobile phase velocities is often referred to as the Golay equation[92] (Equation 12):

$$h = \frac{B}{\bar{\mu}} + C\bar{\mu} \tag{12}$$

where *h* is the column efficiency or height equivalent to a theoretical plate (HETP), *B* is the contribution to plate width due to longitudinal diffusion of the solute in the gas phase, *C* is the contribution to plate width due to resistance in mass transfer of the solute between the mobile and stationary phase, and $\bar{\mu}$ is the average linear velocity of the mobile phase. An expanded form of this equation that includes terms for the resistance to mass transfer from the gas phase and extra column contributions is available.[66] Longitudinal diffusion is always present and acts to broaden the solute peak the longer the component stays in the column. Resistance to mass transfer is a function of the partition ratio *k* and describes the diffusion of a component into and out of the stationary liquid phase.

A measure of the solute plug broadening as it travels the length of the column is defined in terms of the solute retention time and the standard deviation of the chromatographic peak, as the theoretical plate number *n*. It can be calculated by Equation (13):

$$n = \left(\frac{t_R}{\sigma}\right)^2 \tag{13}$$

where t_R is the retention time and σ is the standard deviation of the peak.

Column efficiency *h* is usually normalized to column length, n/L. The efficiency of a column can also be expressed as the relationship of the HETP, *h*. This is a measure in millimeters of the column length occupied by one theoretical plate (Equation 14):

$$h = \frac{L}{n} \tag{14}$$

For a wall-coated open-tubular (WCOT) capillary column, the minimum theoretical efficiency attainable h_{min}

is given by Equation (15):

$$h_{\min} = \sqrt{\frac{1 + 6k + 11k^2}{3(1+k)^2}} \tag{15}$$

where k is the capacity factor. This is the theoretical column efficiency obtained for perfect stationary phase coating under ideal conditions. Many manufacturers, as a measure of their column coating efficiency, compare the calculated value with that obtained experimentally on their column.

4.9.5 Column Resolution

The decision as to what resolution is required is an arbitrary one and in the hands of the analyst. Once a number has been picked for the resolution, the problem then revolves around the following questions: how many plates are needed to achieve this resolution? and how long a column is needed to obtain the required number of plates? A quantitative expression for the degree of peak separation is Equation (16):

$$R = \frac{2(t_{R,j} - t_{R,i})}{w_{b,j} + w_{b,i}} \tag{16}$$

where R is the resolution, t_R is the retention time of components i and j, and w_b is the peak width at the baseline. A resolution of 1.5 gives essentially baseline resolution of symmetrical peaks of equal height.

4.9.6 Length of Column Required for Separation

The number of plates required for a given resolution of two solutes was derived as follows according to Equation (17):[93]

$$n = 16R^2 \left(\frac{\alpha}{\alpha - 1}\right)^2 \left(\frac{k+1}{k}\right)^2 \tag{17}$$

Thus the required column length is given by Equation (18):

$$L_{\text{reqd}} = hn_{\text{reqd}} \tag{18}$$

Unfortunately, the prediction of an experimentally obtainable value of h is not very easy, and the application of Equation (18) is therefore not as simple as it appears.

4.10 Parameters Influencing a Separation

The major factors that influence h, and thus the required column length, are a function of several interrelated column, extracolumn, and operational parameters. Major factors that influence column efficiency are column ID, column length, the type of stationary phase and its film thickness, the type of carrier gas, the carrier gas velocity, and the column temperature, in addition to other instrumental limitations, such as the sample injection width, dead volumes between injector and column and detector, detector volume, data acquisition time constants, and data rates.

Fortunately, a computer model is now available that not only computes L_{reqd} from the inputted values of the column and instrument variables, the identity of the stationary phase, and the Kováts indices of the sample components, but also generates the predicted chromatogram. Use of such a program saves invaluable laboratory method development time.[94] Examples of computer-aided column selection and optimization are described later.

4.10.1 Choice of Carrier Gas

Profiles of h versus u for three carrier gases with a thin-film capillary column are displayed in Figure 32. The greatest efficiency is obtained with nitrogen at a lower linear velocity, but speed of analysis must be sacrificed. The increasing portion of the curve is steeper for nitrogen, meaning that efficiency and resolution drop quickly when working off u_{opt}. On the other hand, for a slight loss in the number of theoretical plates, a reduced analysis time is possible with helium and hydrogen, because u_{opt} occurs at a higher linear velocity. Moreover, the mass transfer contribution or rising portion of a curve is less steep with helium or hydrogen, which permits working over a wider range of linear velocities without a substantial sacrifice of resolution.

When comparing these carrier gases at a linear velocity corresponding to equal values of plate height, note that the lighter carrier gas (H_2) can elute solutes at

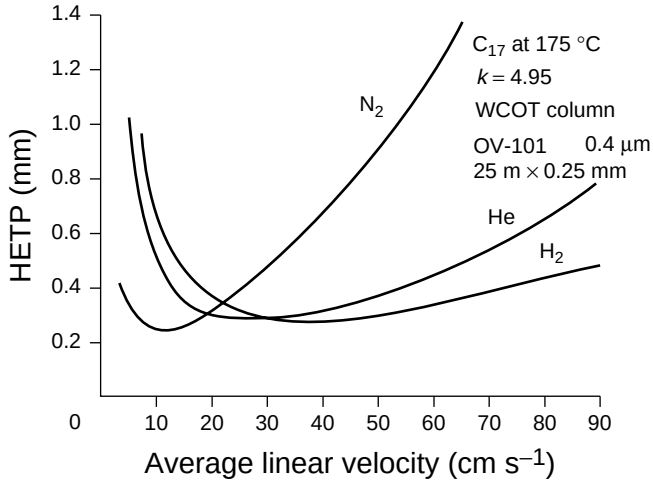

Figure 32 Profile of HETP versus linear velocity for helium, hydrogen, and nitrogen. (Courtesy of Agilent Technologies.)

lower temperatures during temperature programming with narrower band profiles, since higher velocities can be used. Thus hydrogen or helium is recommended over nitrogen as a carrier gas.

The effect of the carrier gas on chromatographic resolution is illustrated in Figure 33, in which the flow is slightly off the optimum. The greatest efficiency (the minimal HETP) is obtained with nitrogen; however, this minimal value occurs only over a narrow range of carrier gas velocities with its efficiency decreasing sharply with increasing average linear velocity.

4.10.2 Effect of Column Temperature on Separation

Column temperature has a dramatic effect on separation. The logarithm of the capacity factor k is inversely proportional to the temperature. The lower the column temperature, the greater is the capacity factor, thus increasing the time that the component spends in the stationary liquid phase. Increased retention in the stationary liquid phase allows greater advantage to be taken of the selective properties of that stationary phase. However, the chromatographer should be aware that, as a result of this selectivity change due to temperature, large changes in retention can occur that, in some cases, change the order of the elution.

4.10.3 Stationary Phase Film Thickness and Column Internal Diameter

Changes in either the stationary phase film thickness or the column ID change the phase ratio β, and greatly influence the capacity factor. An increase in film thickness (or decrease in the phase ratio) results in an increase in the capacity factor and, therefore, increases in component retention and resolution. This is illustrated in the analysis of gasoline shown in Figure 34. Increased resolution,

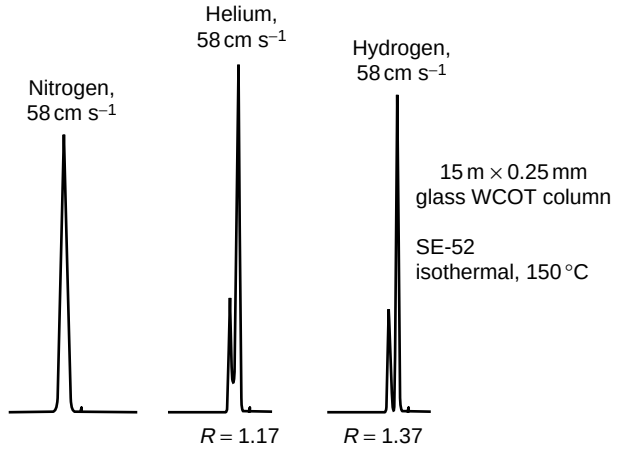

Figure 33 Effect of carrier gas on the resolution of *n*-heptadecane and pristane. (Courtesy of Agilent Technologies.)

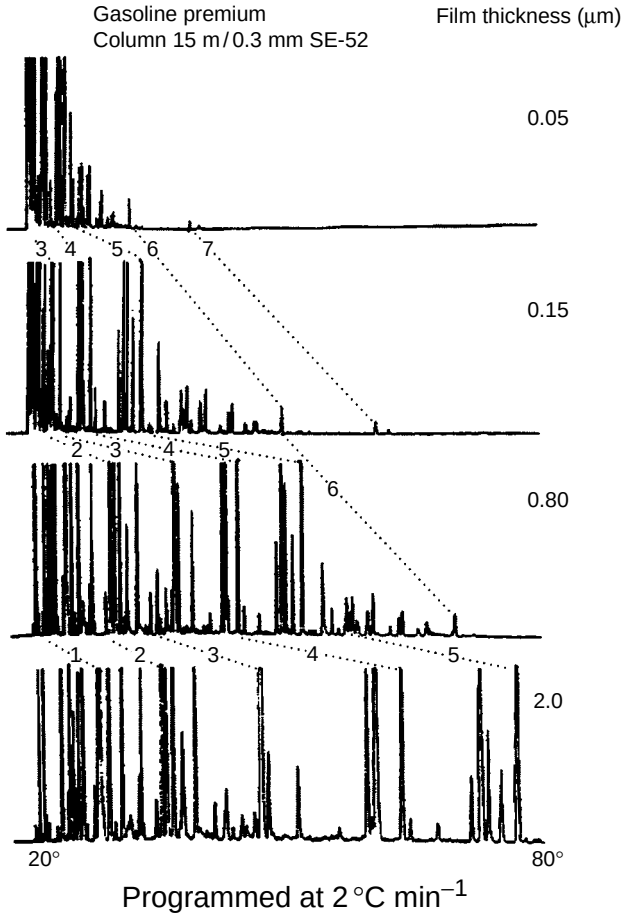

Figure 34 Analysis of gasoline on columns with varying stationary phase film thicknesses.[96]

however, is accompanied by a dramatic increase in analysis time.[95]

Analysis time, measured as the retention time of the last eluting peak from a column of length L, is given by Equation (19):

$$t_{R,x} = L\frac{1 + k_x}{\bar{\mu}} \tag{19}$$

where k_x is the capacity factor of the last-eluting peak. The capacity factor is given by Equation (20):

$$k_x = \frac{t_{R,x} - t_M}{t_M} \tag{20}$$

where t_M is the retention time of an unretained compound (such as air) and $\bar{\mu}$ is the average linear velocity of the mobile phase. Reducing the column length will decrease the analysis time at the expense of resolution (since n is decreased). This may not be a severe handicap, since most chromatographic separations do not require baseline separation. Column temperature has a dramatic effect on separation. The logarithm of k is inversely proportional

to the temperature. The lower the temperature, the greater is the capacity factor, thus increasing the time that the component spends in the stationary liquid phase. Reducing the ID of the column increases the efficiency and resolution of the separation but reduces the sample capacity.

4.10.3.1 Sample Capacity Sample capacity is the ability of the column to tolerate high concentrations of solute. Degradation of chromatographic performance is seen when the column capacity is exceeded. This situation is commonly referred to as overload and is indicated by peak broadening and asymmetry as shown in Figure 35.

The relationships between sample capacity, film thickness, and phase ratio are shown in Figure 36 as log plots of film thickness versus phase ratio and sample capacity relative to a 250-μm ID column with a 1-μm film. These curves are for the commonly used commercial FSWCOT column of ID between 50 and 530 μm. By extrapolating the appropriate curve for the desired column diameter and film thickness, both the phase ratio of the column and the sample capacity can be determined. For example, a 3-μm film on a 530-μm ID column (phase ratio of 45) has 5.5 times the sample capacity of the 250-μm ID column.

4.10.4 Fast Capillary Gas Chromatography

A major trend in GC in recent years has been fast separations using small-bore columns.[96] In high-speed gas chromatography (HSGC), separations are measured in seconds rather than minutes, and this speed presents a challenge for sample introduction, temperature programming, and detection.[97] One of the primary objectives for developing HSGC is to provide near-real-time monitoring for field applications. Practical limits for HSGC

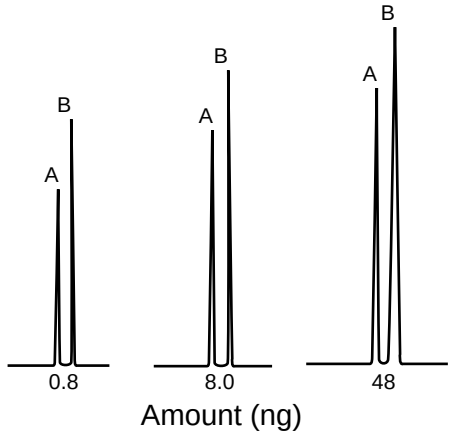

Figure 35 Peak shape of (A) dodecanol and (B) *n*-pentadecane with increasing amounts of sample on the column. Column: 10 m × 0.10 mm × 0.17 mm cross-linked fused-silica wall-coated open-tubular (FSWCOT). Sample capacity is reached at 8 ng/component.[92] (Courtesy of Agilent Technologies.)

using currently available instrumentation have been published.[98]

Several papers have discussed the advantages and disadvantages of narrow-bore WCOT columns for high-speed capillary GC.[99–101] For HSGC without significant loss in resolution, it has been recommended that the column ID should be reduced to 0.1 mm or less. However, in many cases there is no advantage to using 0.1-mm ID columns over an optimized operation of the more conventional 0.25-mm ID column.[94] The most serious limitations to the use of reduced-diameter columns are associated not only with sample capacity but also with the various limitations of current instrumentation.

Within the limits of current instrumentation, 100-μm ID columns of 5–10 m length can be operated with minor modifications.

4.11 Computer Optimization of the Separation

Chromatographic separation optimization can be broken into two major areas: stationary phase selection and optimization of the analysis conditions. Several tools are now available to aid the chromatographer in making a decision with few or no laboratory experiments. Pro ezGC (Analytical Innovations, Inc., Kettering, OH, USA) and Dry Lab (LC Resources, Walnut Creek, CA, USA) are models that use the temperature and pressure programming approach to GC analysis. OnLine's (Duxbury, MA, USA) GC Lab is an isothermal model that uses the appropriate lengths of serially coupled columns containing the stationary phase of widely different selectivities. One example using Pro ezGC to optimize the run time on a given liquid phase is given below.

In the following computer-aided optimization example, Pro ezGC is used to optimize the temperature separation of the mixture of 20 phenols and pesticides on cross-linked 5% phenylmethyl siloxane DB-5. The list of the components is given in Table 8. The first simulation shown in Figure 37(a) and (b) was run with a hypothetical column of 60 m × 0.25 mm ID × 0.25 μm film thickness coated with DB-5. An arbitrary temperature of 100 °C and a flow velocity of 45 cm s^{-1} hydrogen were used. This run was isothermal with no attempt to optimize the separation.

Obviously, it is not necessary to be a chromatographic expert to see that the total run time can be shortened considerably by programming the column temperature, by shortening the column, and by selecting another flow velocity. Program simulation gives the user the capability to study the separate effect of each of these parameters and to optimize all the conditions simultaneously with a minimum of experimental chromatographic work. The operator defines the simulated runs with parameters similar to those used with a gas chromatograph, but

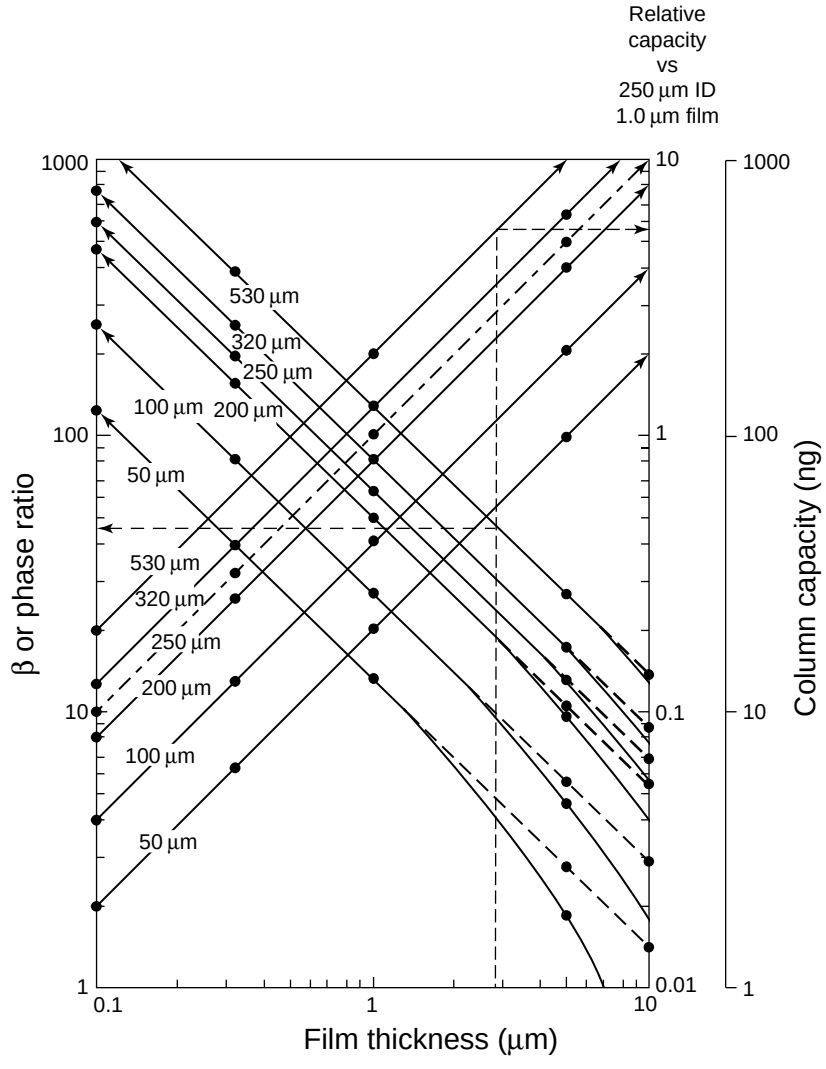

Figure 36 Logarithmic relationship between film thickness, phase ratio, and sample capacity.[92] (Courtesy of Shell Development Co., Houston, TX.)

yielding the tremendous advantage of obtaining the results in minutes instead of days.

The boundary conditions of the simulation are as follows:

Length of column: 7.5, 15, 30, and 60 m
Liquid phase: DB-5
Flow velocity: 45–120 cm s^{-1}, in increments of 15 cm s^{-1}
Initial temperature: 100–200 °C, in increments of 25 °C
Temperature programming conditions: 5–45 °C min^{-1}, in increments of 5 °C min^{-1}
Final temperature: 300 °C

This range of conditions requires 1080 simulations, which are performed in just a few seconds on a laptop computer.

The optimized conditions of the run are as follows:

Resolution: 1.5
Column: 15 m × 0.25 mm ID × 0.25 µm film thickness, DB-5
Initial temperature: 150 °C
Temperature rate: 30 °C min^{-1} up to 300 °C
Inlet pressure: 54.6 kPa (8 psi)

Figure 37 shows the results of this optimization. The total separation for the baseline resolution is done in <4 min. There are applications for which baseline separation is not necessary.

By accepting a minimal resolution of 0.8 instead of 1.5, the shorter calculated run time could be reduced by 50% (<2 min). No attempt was made to shorten this separation any further.

Table 8 Computer simulation of the separation of phenols and pesticides on DB-5

Component number	Component name
1	Phenol
2	2-Chlorophenol
3	2,4-Dimethylphenol
4	2-Nitrophenol
5	2,4-Dichlorophenol
6	Dichlorvos
7	4-Chloro-3-methylphenol
8	2,4,6-Trichlorophenol
9	2,4-Dinitrophenol
10	4-Nitrophenol
11	2-Methyl-4,6-dinitrophen
12	Ethroprop
13	Chloroproham
14	Trifluralin
15	Atrazine
16	Pentachlorophenol
17	Diazinon
18	Disulfoton
19	Alachlor
20	Oxyfluorofen

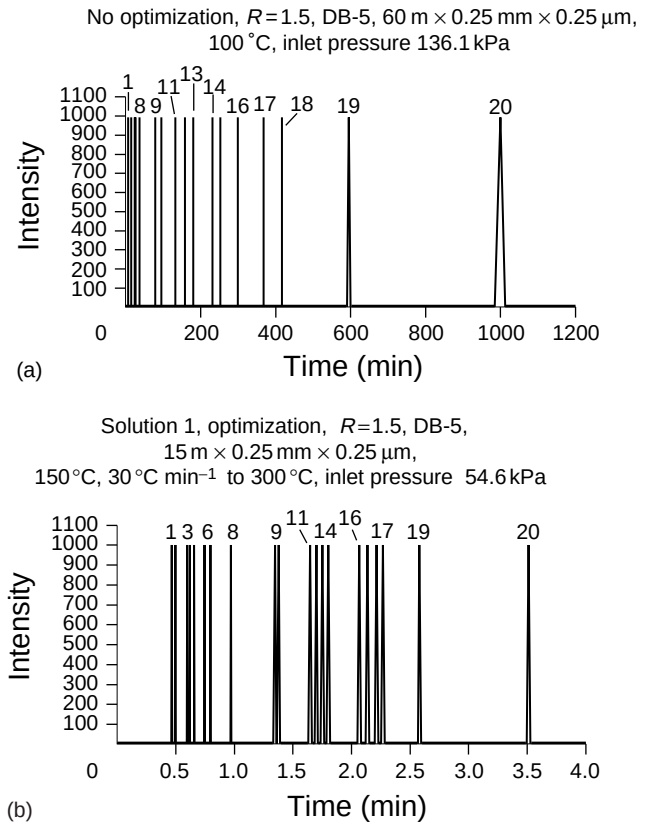

Figure 37 Simulated separation of 20 phenols and pesticides using the application Pro ezGC. (a) Experimental conditions have been arbitrarily set; (b) global optimization of the chromatographic separation. Baseline separation is achieved in <4 min.

4.11.1 General Comments

The computer tools that are now available, although currently limited to gas–liquid applications, offer the tremendous potential for operating gas chromatographs much more efficiently using less skilled personnel. A program such as Pro ezGC aids the analyst in minimizing run time by optimizing the programming conditions of flow and temperature. The GC Lab simulation tool gives the user the capability of optimizing the overall selectivity of the column (probably the most important step in method development) by optimizing the selectivity of two independent segments of column individually coated with different liquid phases. Also, optimization may be focused either on designing for minimal time analysis or minimum detection limits. Most of the simulation programs allow for the addition of compounds to the library, either through in-house laboratory work or by purchasing commercial libraries containing many compounds in a variety of stationary phases.

The simulation tools allow the analyst to experiment with various methods, in essence developing a GC method at the computer keyboard. The parameters of the proposed method are obtained quickly and can then be used to design the proper laboratory experiments for its validation. All of this saves the valuable laboratory time normally required in the initial method development stage. The experimental results, however, are only as good as the chromatographic model that is used to predict them. In the case of fast chromatography, for instance, the dead volumes upstream and downstream of the column contribute to peak broadening, and accurate quantitative analysis is sacrificed as peak widths decrease to below the minimum of 20 points per peak needed for accurate and precise quantitative determination. One of the simulation tools (GC Lab) allows for user inputs of these extra column contributions. It has been verified that when this is done, the predicted chromatograms generated with the model match those obtained in the laboratory.[102,103]

As with all automated tools, the user should be conscious of their limitations, regarding them as somewhat imperfect mathematical tools. It is only by iteratively correcting to the model using results obtained in the laboratory that accurate predictions of experimental results are generated by the computer.

4.12 Sample Injection Devices

The main function of a GC inlet is to provide the accurate, reproducible, and predictable introduction of a sample into the column. The sample can be gaseous, liquid, or solid, and several injection modes can be used. There is no single inlet that can satisfy all analytical requirements.

Inlets are usually divided into packed-column inlets and capillary-column inlets. Sample introduction has been discussed at length in the literature.[104–106]

Inlets must be selected carefully for each analysis and used to optimize chromatographic efficiency, analytical accuracy, and the reproducibility of results. Because capillary columns have higher efficiency and lower sample capacities than do packed columns, inlet performance is much more vital to obtaining accurate results with capillary rather than with packed-column systems. Once an inlet has been selected based on sample, column type, and analysis goals, all inlet variables must be set appropriately to achieve optimal results. The type of analysis and the composition of the sample itself are the primary factors that must be considered when setting or determining the interrelated variables related to the injection itself (type of injector, injected volume, injection temperature) and to the chromatographic analysis (column selection and column temperature).

4.12.1 Effect of Inlet on Peak Broadening

A basic functional requirement of GC inlets is that they introduce the sample into the column as a narrow band with a composition that is identical with that of the original sample. The inlet should produce peaks (width-of-injection function) that are much narrower than the expected peak width resulting from the column band-broadening process. More efficient columns require narrower initial peak widths. This can be accomplished by using inlets that generate narrow peaks initially, or by using subsequent focusing techniques. Initial bandwidths are broadened by band broadening in time or by band broadening in space. Band broadening in time is caused by the slow transfer of sample vapor from the inlet to the column. The initial peak width is equal to the time it takes for the sample to be transferred to the column. Band broadening in space is a direct consequence of the migration and spreading of a liquid sample within the column, either after the cool on-column injection of the sample into the column or after recondensation of the sample in the column. The condensed liquid, which starts by occupying only a few centimeters of column, becomes too thick to be stable and spreads over a greater length of column. The carrier gas pushes the plug further into the column, creating a flooded zone. The solute material is spread over the full length of the flooded zone, creating an initial peak width that equals the length of the flooded zone.

4.12.2 Sample Focusing as a Way to Reduce Peak Broadening

Several focusing techniques are used to narrow peak broadening in time and space:[108] stationary phase focusing, solvent focusing, thermal focusing, and retention gap.

4.12.2.1 Stationary Phase Focusing Stationary phase focusing is the most frequently used focusing technique and is possible only in temperature-programmed analysis. Retention times of solutes are an exponential function of temperature. As the initial temperature of the column is lowered, the speed at which solutes travel down the column slows drastically. As a vaporized sample moves from the inlet to the column, it comes in contact with the stationary phase and is trapped in a narrow zone. The lower the temperature, the more effective the focusing.

4.12.2.2 Solvent Focusing As the condensed solvent starts to evaporate, solutes with volatility similar to that of the solvent tend to concentrate and focus on the solvent tail. This solvent focusing, or the ''solvent effect'', yields narrow peaks for these early-eluting compounds, as shown in Figure 38.

4.12.2.3 Thermal Focusing Thermal focusing relies on the condensation of gases in a tube or at the head of the column. Peaks narrow as the solute volume is reduced during condensation. Solutes do not migrate into the chromatographic system until the temperature is raised, then they are vaporized again.

Sometimes cryogenic temperatures (cryogenic focusing) are used to focus peaks from inlets or auxiliary sampling devices that generate peaks broadened in time. Thermal focusing narrows bandwidths effectively only when the column temperature is approximately 150 °C below the boiling points of the solutes. In this sense, thermal focusing does not rely on chromatographic processes. It requires only a surface on which vapors can condense.

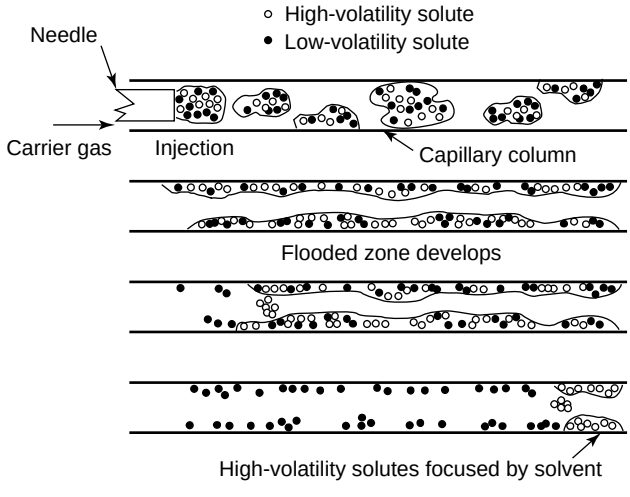

Figure 38 The solvent effect.[107] (Copyright 1997, Agilent Technologies. Reproduced with permission.)

4.12.2.4 Retention Gap A retention gap is an empty piece of column that accommodates the condensed sample but does not retain solvent or solutes once they have been vaporized. The primary function of the retention gap is to reduce the length of the flooded zone created whenever solvent is condensed in the column. An equally important function is to protect the column from nonvolatile sample components, especially when doing cool on-column injection (Figure 39).

4.13 Types of Capillary-column Inlets

Capillary-column inlets that use hypodermic syringes to introduce the samples fall into two categories: vaporizing and nonvaporizing (cold) injections.[105]

4.13.1 Vaporizing

Capillary direct, or vaporizing, injections are used with some capillary columns of ID>0.5 mm and are made by substituting a special insert inside a packed-column inlet to protect the column from nonvolatile sample components.

Split mode injection uses a vaporizing inlet that vents most of the sample in the split mode and transfers most of it to the column in the splitless mode. It is also column protecting but can cause discrimination and decomposition of the sample. Split injection is used for general analysis, whereas splitless injection is most frequently used for trace analysis.

The programmed temperature vaporizer (PTV) inlet offers a mixture of injection possibilities, including cool sample introduction, split or splitless modes, and sample concentration (solvent elimination mode); it is column protecting. Because of this flexibility, PTV inlets are useful for both general and trace analyses.

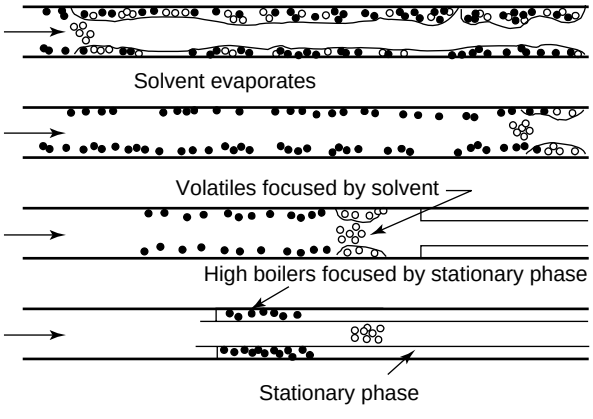

Figure 39 Visual representation of on-column injection into a retention gap. Both solvent and stationary phase focusing occur.[107] (Copyright 1997, Agilent Technologies. Reproduced with permission.)

4.13.2 Nonvaporizing

Cool on-column inlets give high accuracy and reproducibility, do not lead to sample decomposition, have the least solute discrimination among all the inlets, and work by depositing the sample directly into the column. Unlike vaporizing sample introduction techniques, here the sample is not exposed to high temperatures during injection or transfer to the column. Cool on-column inlets are used for the analysis of samples with a wide boiling-point range or those that are thermally sensitive for trace analysis.

4.13.3 Types of Capillary-column Inlets Using Mechanisms Other than Syringes

Other sample introduction devices include gas and liquid sampling valves, headspace autosamplers, thermal desorbers, purge-and-trap samplers, pyrolyzers, and transfer lines. Depending on the sampling device and the GC column used, auxiliary sampling devices can be connected directly to the column or to an existing inlet.

4.13.3.1 Valves Valves can consistently introduce fixed volumes of gas or liquid samples and are simple to automate. Valves are versatile, can operate at elevated temperatures (several hundred degrees), have a fast response time (down to fractions of a second), and can be small enough to be mounted into a portable instrument. Valves are frequently used for sampling gases and liquids in moving streams (process and on-line monitoring).

4.13.3.2 Sample Introduction Based on Gas–Liquid Extraction The principle of sample introduction based on a gas–liquid extraction is fairly simple. Once an equilibrium has been achieved, an analyte distributes itself between two immiscible phases (here the gas and the liquid sample phases) at some constant ratio. The two modes of gas–liquid extraction are static and dynamic. In the static mode, the analysis begins with the addition of an accurate volume of sample into the vial. The sample volume must be selected to allow for the appropriate volume of gas phase or headspace in the remaining volume of the vial.

In the dynamic or purge-and-trap mode, extraction of the sample is effected by passing a constant flow of gas through the aqueous sample. The analytes are continuously purged from the sample and carried by the gas flow to a medium that can collect and concentrate the analytes.

4.13.3.3 Microextraction Microextraction has been shown to be an efficient method for the introduction of samples into a high-speed gas chromatograph and is applicable to liquid–liquid extraction.[108]

4.13.3.4 Headspace Autosamplers (Static Gas–Liquid Extraction) Headspace analyzers perform static gas–liquid extractions and are used to determine volatiles in liquids, solids, or complex matrices. Once the sample has been added, the vial is sealed with a septum cap, placed in a temperature-controlled environment, and allowed to equilibrate for a certain period of time. After the equilibrium has been achieved, a sample of the gaseous phase in the vial is removed by a gas-sampling syringe and injected into the gas chromatograph. Headspace autosamplers inject a portion of the gas that is in equilibrium with a sample into a thermostated sealed vial.[109] Headspace analysis is used for the analysis of residual solvents, fragrances, and volatile pollutants in soil and water. Reproducible headspace analysis requires excellent control of all the parameters affecting the vapor pressure.

4.13.3.5 Purge-and-trap Automatic Samples (Dynamic Gas–Liquid Extraction) Purge-and-trap combines dynamic headspace trapping with thermal desorption. Volatile components are continuously purged out of a water sample (dynamic headspace), trapped on an adsorbent, and then desorbed quickly for introduction into the gas chromatograph (thermal desorption). Purge-and-trap is used mainly for analysis of environmental pollutants in water and for analysis of volatiles in beverages.

4.13.3.6 Thermal Desorbers Thermal desorbers are used in environmental sampling and are complementary to headspace analysis and analytical pyrolysis. Volatile sample components that are contained in a solid sample or have been absorbed on a solid adsorbent are thermally liberated in the sample in a stream of carrier gas and carried to the column. The advantages and disadvantages of sorbent sampling, active and passive techniques, solvent and thermal desorption, grab-sampling containers, cryogenic traps, and cold-trap concentration for the analysis of C_1-C_4 hydrocarbons can be found in the literature.[110] This technique is used for monitoring hazardous gases in the workplace and for environmental air analysis.

4.13.3.7 Pyrolyzers Pyrolyzers are used to thermally cleave non-volatile samples into volatile fragments, which are then analyzed by GC. Temperatures in the 500–1000 °C range are normally used for the analysis of polymers, fibers, microorganisms, and geological samples.

4.13.3.8 Transfer Lines (Ambient Vapor Sampling Inlet Module) The ambient vapor sampling (AVS) inlet module,[79] mentioned earlier, combines injection of gas with crude GC separation. Its main application is for field MS and for IMS measurements.

Figure 40(a) and (b) shows a schematic view of the Enviroprobe. The AVS is a valveless approach[111] to injecting variable quantities of gases by modulating the time during which the carrier gas exchanges the carrier gas flow in tube two for the sample flow in the opposite direction. This approach permits resolution in milliseconds for the injection pulse width.[72] The GC separation is performed in the transfer line between the injector and the detector. This approach minimizes the upstream dead volumes and at the same time reduces the sample adsorption sites.

4.13.3.9 Cryofocusing Trapping and Fast Desorption Cryofocusing of samples, followed by fast thermal desorption, at a rate of several thousand degrees Celsius per second, is an approach used to generate narrow injection pulses. One successful approach, originated at the University of Michigan,[112] uses a cryofocusing capillary dimension trap for sample collection and capacitive discharge for heating the trap rapidly.[113] This approach is now commercialized by Chromatofast (Ann Arbor, MI, USA) and is shown in Figure 41(a–c).

Figure 41(a) shows the sample collection and cryofocusing mode. In this mode, valve V_1 is open, V_2 is closed,

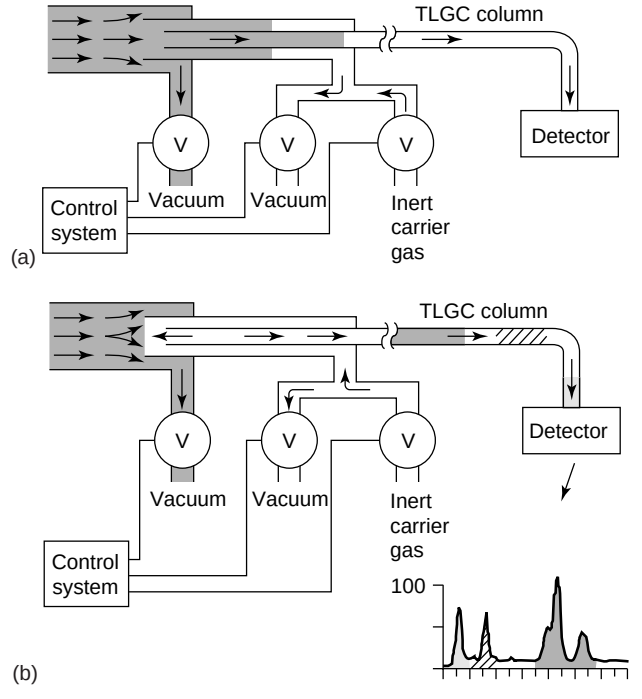

Figure 40 Schematic view of the Enviroprobe (AVS) from FemtoScan. (a) In the injection mode, the two vacuum lines are opened and the sample gas is injected into the column. (b) By turning off the second vacuum line (middle connection), the inert carrier gas is forced through the column and creates a barrier between the supply of sample gas and the column. (Courtesy of FemtoScan Corp.)

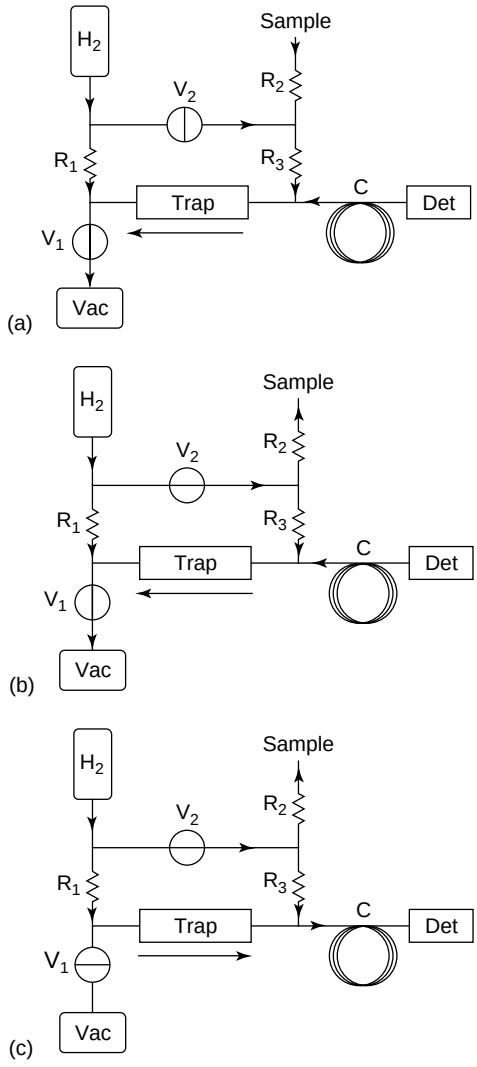

Figure 41 Schematic diagram of the reverse sampling cryofocusing instrument for high-speed GC. (a) Sample collection mode; (b) flush mode; (c) analyze mode. V_1 and V_2 are micro-solenoid valves; R_1–R_3 are capillary flow restrictors; Trap is the trap chamber which contains the capillary dimension trap tube; C is the column; Det is the detector; Vac is the vacuum pump; and H_2 is the source of hydrogen carrier gas. Arrows in the figure indicate flow direction. (Courtesy of Chromatofast.)

and the cold trap is maintained at a low trapping temperature (between −20 and −160 °C, depending on the samples to be trapped). With the vacuum pump in the system, sample vapor is introduced and flows through the restrictors R_2 and R_3, then into the cold trap, where it is cryofocused. This reverse flow means the sample is introduced into the trap end closest to the separation column.

Figure 41(b) shows the backflush mode. Both valves are open, and flows through the column and the cold trap are from right to left. Therefore, any components that have not yet eluted from the column are backflushed toward the trap. If the trap is cooled to trapping temperatures, the backflushed components are cryofocused again for a subsequent injection.

Figure 41(c) shows the injection and separation mode. Valve V_1 is closed and V_2 is open, causing the flow through the cold trap to be from left to right. The trap is then rapidly heated with a capacitive discharge power supply. This serves to flash vaporize the sample and deliver it to the column as a narrow vapor plug. Injection bandwidths are around 5 ms.

This inlet system can interface to conventional split inlets for liquid samples or can be used to sample air directly or from containers such as Tedlar bags. For air sampling, an adjustable amount of preconcentration can be accomplished by adjusting sampling times. This injector feature also uses the reverse sampling facility. The sample is collected and reinjected from the same end of the trap, which helps minimize injection bandwidth, minimize degradation of thermally labile compounds, and limit sample contamination or carryover from the split inlet. In addition, this inlet design prevents the sample from coming into contact with the valves, preventing contamination and reactivity of the sample with valve inner surfaces.

Figure 42 shows an application of fast injection to the determination of solvents from paint. A total of 15 solvents were separated in <1 min on a DB-1 column.[175] A conventional GC analysis for these compounds required a 1–2-h cycle time. This application analyzes the air in paint booths at automotive plants.

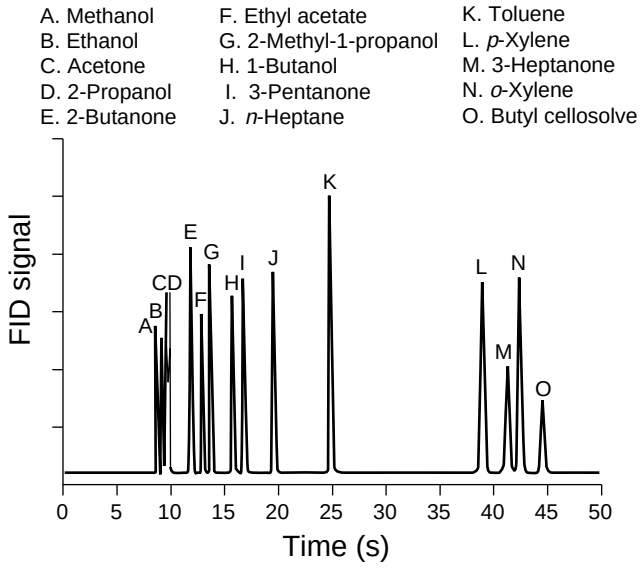

Figure 42 Example of fast GC injection and separation, showing the separation of paint booth solvents. Experimental conditions: column, DB-1, 8 m × 0.25 mm ID, 0.25 µm film thickness; temperature program, 35 °C for 0.1 °C min^{-1}, 50 °C min^{-1} to end of analysis; column flow, 4 mL min^{-1}. (Courtesy of Chromato-FAST Inc.)

4.14 Gas Chromatography Detectors – Basic Terminology

Whatever detector is used, there are three major response characteristics of detectors: sensitivity, selectivity, and dynamic range. To be useful in a practical sense, stability and repeatability must also be considered.

4.14.1 Sensitivity

This is the response per amount of sample, i.e. the slope of the response–amount curve. The minimum amount on the curve is defined as the minimum detectable quantity (MDQ). The MDQ is practically defined as the amount of sample for which the peak height is twice the noise height (signal-to-noise ratio = 2).

4.14.2 Selectivity

Selectivity identifies which categories of compounds give a detector response. Some detectors, such as the TCD, sense almost everything and are considered universal. Others, such as the ECD, respond only to certain types of components and are useful when the detected component is present in a complex, nondetected matrix. Selectivity is usually expressed as the ratio of the injected quantities of two compounds that give the same output response. For instance, the selectivity of the ECD for anhydrides relative to hydrocarbons is 10 000 : 1.

4.14.3 Dynamic Range

Dynamic range is the range of sample concentrations over which a relative incremental change in sample concentration produces a similar relative change in the signal. This range can be quite large for certain detectors (seven orders of magnitude for the FID). For others, such as the flame photometric detector (FPD), it can be limited.

4.15 Gas Chromatography Detector Selection

The selection of the GC detector to be used is determined by the information desired. The compound structure is identified by a mass spectrometric detector (MSD) or an infrared detector (IRD). High sensitivity is achieved by the ECD, the nitrogen–phosphorus detector (NPD), and the MSD. High selectivity is provided by the ECD, NPD, FPD, and the MSD in the selected-ion monitoring (SIM) mode. The electrolytic conductivity detector (ELCD) and the PID also offer high selectivity for particular species. A general description of these detectors follows. GC detectors are covered in detail in the literature.[65,67,115]

4.15.1 Flame Ionization Detector

The FID consists of an oxygen-rich hydrogen flame that burns organic molecules, producing ionized molecular fragments. The resulting ions are collected by a polarized collector (Figure 43). The FID is selective for materials that ionize in an air–hydrogen flame. Few ions are formed until an organic compound elutes into the flame. The resulting large increase in ions is collected by a polarized collector. The FID is characterized by a large dynamic range (10^7), high sensitivity ($15\,\mathrm{mC\,g^{-1}}$), low detection limit ($10^{-12}\,\mathrm{g\,s^{-1}}$), and small dead volume (nanoliters). For these reasons, the FID is the most popular GC detector today. With optimized operating parameters and a good analytical column, it is possible to determine components in as little as 20 pg or about 5 ppb of sample gas.

The response of the FID depends on the flow rate of all three gases – carrier gas, hydrogen, and air – with hydrogen being the most critical. The flame is usually optimized by successive injections until the signal reaches its maximum.

4.15.2 Photoionization Detector

The PID consists of a UV lamp and an ionization chamber. The lamp is filled with a gas (usually argon or hydrogen) and produces emission lines characteristic of the gas when excited. The radiation passes through a metal fluorite window and into the ionization chamber or cell. There the sample absorbs the radiation and ionizes. Electrodes collect the ions; the current measured is proportional to the sample concentration (Figure 44).

Most of the energy is emitted at one frequency; the lamps are characterized according to the energy of this major emission line. The lamp is selected based on the ionization potential of the compound being measured and on the selectivity and sensitivity required. Lamps of 8.3, 9.5, 10.2, and 11.7 eV are available. Using a

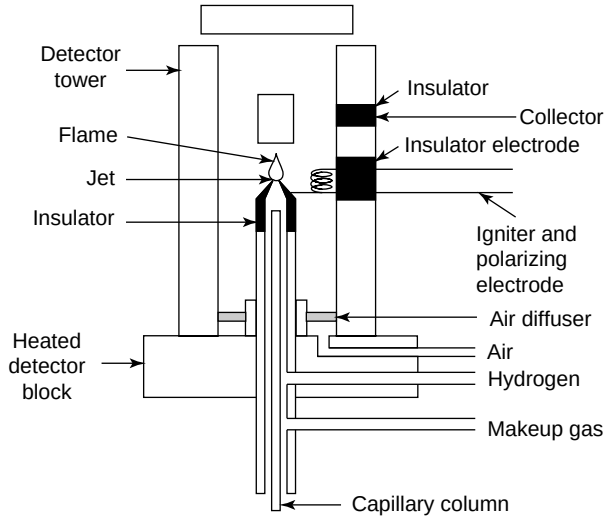

Figure 43 Schematic of an FID.

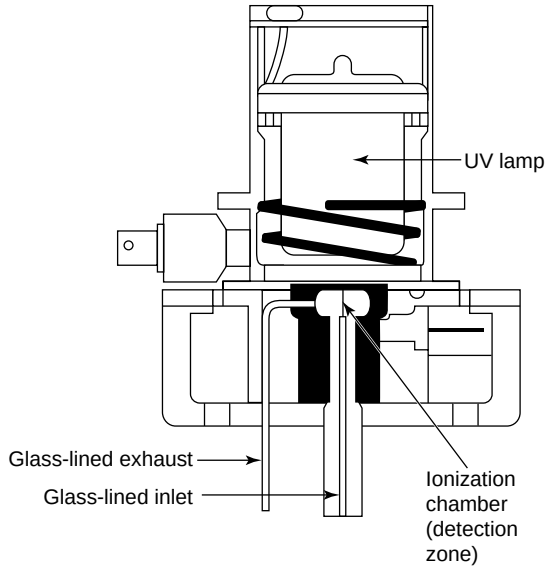

Figure 44 Principal elements of a PID. (Courtesy of HNU Systems, Inc.)

10.2-eV lamp, the PID does not respond to C_1-C_4 hydrocarbons (ionization potential 10.5–12.98 eV) but does respond to C_5 hydrocarbons and above. Likewise, it does not respond to permanent gases (IP > 10.5 eV). It does respond, however, to aromatics, olefins, ketones, amines, etc. Thus for complex samples such as gasoline, PID chromatograms are much easier to interpret than those obtained with the FID.

The PID is a concentration detector with a dynamic range of 10^7 and an MDQ of 2–3 pg benzene. Detector cell volume is an important specification for the capillary column. The available cell volume today is <30 μL, which may be acceptable for normal-speed capillary chromatography but not for high-speed variations.

4.15.3 Electron Capture Detector

The cavity of the ECD contains two electrodes with the effluents of the column passing between them. The cathode is coated with β-emitting radioactive material. The emitted high-energy electrons produce large amounts of positive ions and low-energy (thermal) electrons in the carrier gas. These electrons are normally collected by the positively charged electrode, thus producing a standing current of typically $10^{-8}-10^{-9}$ A. However, when a material with an affinity for thermal electrons is present in the carrier gas, it captures some of these electrons, and consequently, the standing current that is observed is reduced. The essential elements of the ECD are summarized in Figure 45.

Instead of applying a steady direct current (DC) polarization voltage, a pulse of 1–3 μs at 50 V is applied to the electrodes every 150 μs. This pulse collects only free

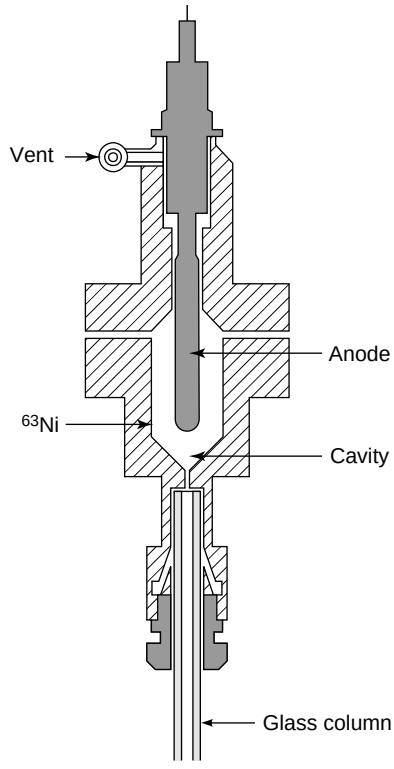

Figure 45 Cross-section of an ECD. (Copyright 1997, Agilent Technologies. Reproduced with permission.)

electrons, since the heavier negative ions cannot respond in this small interval and are swept out of detection by the carrier gas. The most common carrier gas selections for ECD are nitrogen and 5% methane in argon.

The ECD is highly selective. It responds to electrophilic species such as halogenated compounds. Pesticides give an excellent response with MDQs at the picogram level and over a dynamic range of 10^4.

For successful ECD operation, it is important that the carrier and purge gases are clean and dry (99.9995%). Moisture, oxygen (leaks), or other contaminants must be avoided for stable operation of the detector. The column must be preconditioned before use. Finally, the use of a radioactive source is regulated by many governments and usually requires a license to be portable.

4.15.4 Nitrogen–Phosphorus Detector

The NPD (also called a thermoionic detector) uses a jet and collector similar in appearance to the FID. However, this detector is made selective for nitrogen- and phosphorus-containing compounds by a source of alkali metal salt positioned above the jet (Figure 46). This thermoionic source efficiently ionizes nitrogen- and phosphorus-containing organic molecules. Ions are collected, and the resulting current is measured for the chromatogram.

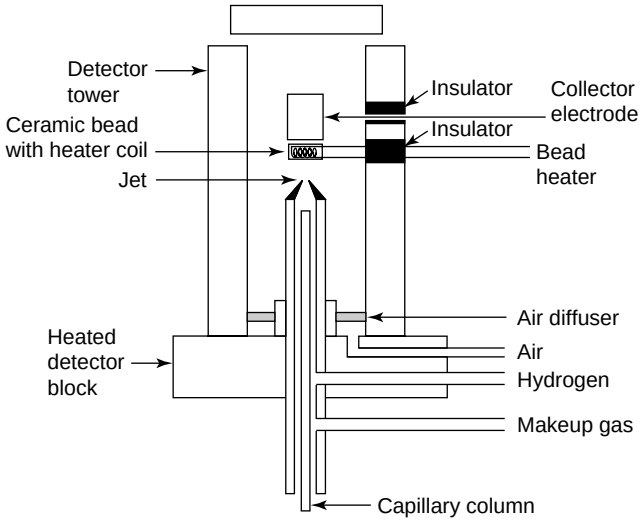

Figure 46 Design of a commercially available NPD.[67]

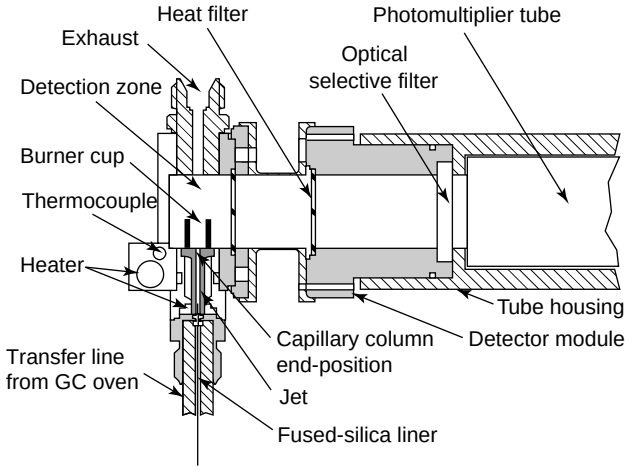

Figure 47 Diagram of a commercially available FPD. The detection zone (chemiluminescent zone) is above the flame; since the flame is shielded by a burner cup, the photomultiplier can see only the area above the flame.[114] (Copyright 1997, Agilent Technologies. Reproduced with permission.)

The NPD has proven valuable in pesticide analysis, where its selectivity replaces costly extraction or concentration steps. It can detect a few picograms of nitrogen- and phosphorus-containing organic compounds. The NPD has a dynamic range of 10^4 and has a selectivity of 10 000 : 1 to 100 000 : 1 by weight of nitrogen and phosphorus to weight of carbon. It does not detect inorganic nitrogen, such as N_2 and ammonia.

4.15.5 Flame Photometric Detector

An FPD is essentially a flame emission spectrometer with its design optimized as a GC detector. This is an element-specific detector that responds to a particular molecular species. It is used primarily for the determination of sulfur- or phosphorus-containing compounds.

The sample contained in the column is fed to a flame, where the individual atoms are excited to higher electronic states by the energy of the flame. These excited atoms and molecular fragments subsequently return to the ground state with the emission of characteristic atomic or molecular band spectra. In the case of sulfur, the band emission from the S_2 species at 394 nm is used, while the 526-nm emission is used for phosphorus.

A narrow band-pass filter is used to isolate the appropriate analytical wavelength, and its intensity is measured by a photomultiplier (Figure 47). The FPD response to sulfur and phosphorus is a factor of 10^4 greater than its response to hydrocarbons, and the MDQ is at the subnanogram level. The response to sulfur concentrations is not linear. It should be proportional to the square of the sulfur concentration; in practice, a factor of 1.8 provides a better fit to the calibration curve.

Figure 48(a) and (b) illustrates the selectivity that results when the FPD is used with capillary columns for the analysis of sulfur compounds in petroleum naphtha. Interference from the concentrated hydrocarbon background is not observed.[114]

One potentially serious problem with the FPD is the undesired light absorption that can occur in the flame. This can happen by hydrocarbon quenching from collision when a high concentration of carbon dioxide occurs in the flame at the same time as the sulfur atom species. Also, self-quenching can occur at high concentrations of the heteroatom species. In this case, collisional energy absorption or reabsorption of the photon by an inactivated species could effectively prevent the photon from reaching the photomultiplier.

The flame conditions are critical to successful operation. Gas flows are also critical to optimizing selectivity and sensitivity. Temperature can also have a significant effect on sensitivity. Finally, condensation of water, especially with halogenated solvents or samples, will cause corrosion in the detection zone or fogging of the window leading to the photomultiplier.

In summary, the FPD has a number of limitations. These include its response dependence on the $O_2 : H_2$ ratio, the H_2 flow rate, the type of sulfur compound, and even the length of time the flame has been burning. In addition, the best results are obtained by generating a calibration curve for each compound in the mixture.

4.15.6 Thermal Conductivity Detector

The TCD responds to any compounds with a thermal conductivity that is different from that of the carrier gas. Hydrogen and helium are recommended carrier gases because of their exceptionally high thermal conductivity differences from all compounds. The principle of

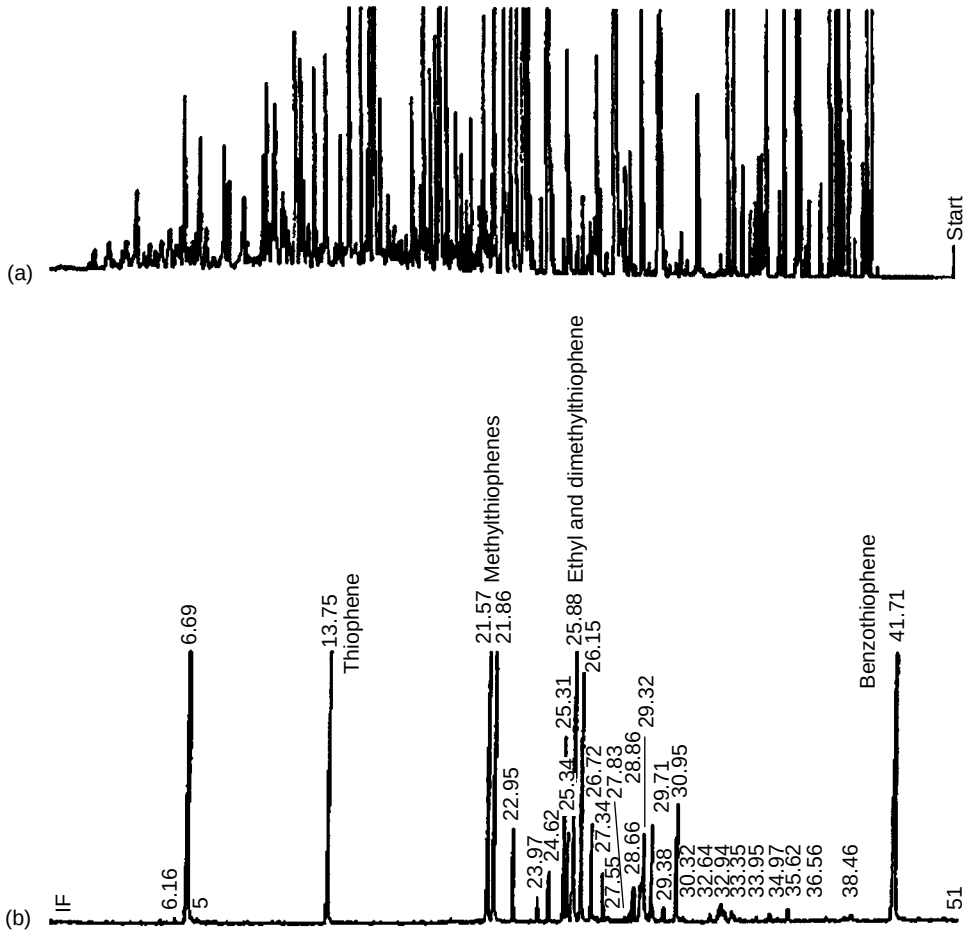

Figure 48 Sulfur compounds in the complex hydrocarbon mixture of petroleum naphtha using (a) FID and (b) FPD.[115] (Courtesy of Agilent Technologies.)

operation of a TCD is fairly simple: when the analyte is present in the carrier gas, the thermal conductivity drops and less heat is lost to the cavity wall. Under constant-temperature operation the voltage supply to the filament drops. This change is recorded and measured.

In practice, the TCD is extremely temperature sensitive. One can show that a difference of $10^{-6}\,^{\circ}C$ is detectable in a single-filament detector; for this reason, a TCD is usually composed of two identical cells physically closed from each other and part of a Wheatstone bridge. One of the cells sees only pure carrier gas and the other is connected at the output of the column (Figure 49). Although this procedure reduces the variations in signal caused by changes in block temperature and any other common mode signal, such as variations in carrier gas flow rate, it does not eliminate their effect on the final signal. Careful thermal design is of critical importance.

An alternative solution to the stability problem has been offered.[116,117] It has been shown that it is possible to produce a stable and fairly rapidly responding TCD by using only one thermal conductivity filament

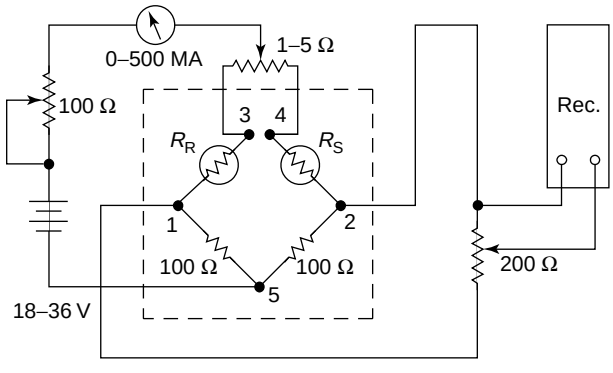

Figure 49 Wheatstone bridge for the differential measurement of thermal conductivity. (Courtesy of Gow-Mac Instrument Co., Benthehem, PA.)

and alternately switching (at 10 Hz) the column and reference flow of carrier gas and effluents from the column (Figure 50). This pneumatic modulation has the advantage of almost completely eliminating the common mode signal and producing a signal with much lower

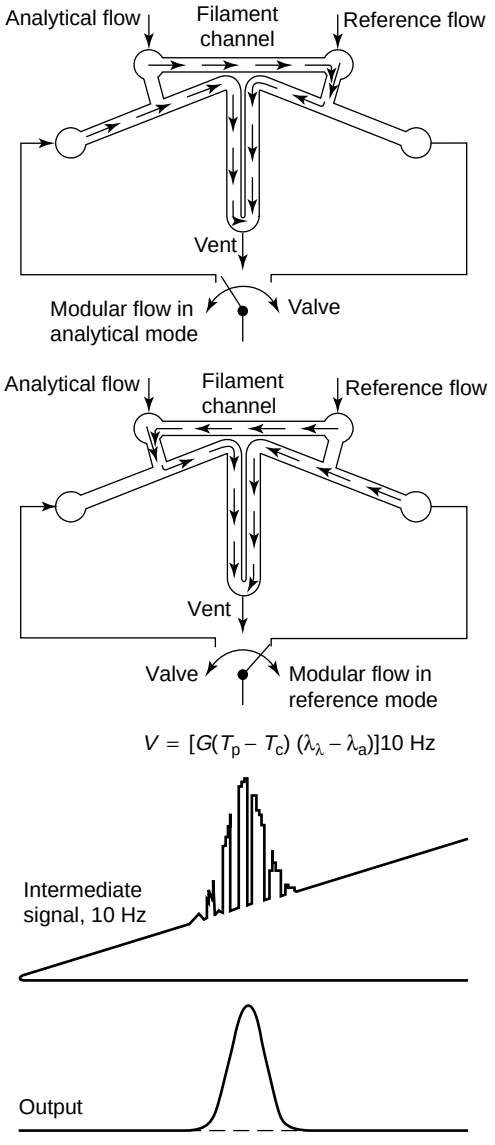

Figure 50 Diagram of a flow-modulated TCD that uses only one filament. The analytical and reference gas flows are switched at a rate of 10 times per second. This design results in a detector with a short warm-up time and low drift. (Courtesy of Agilent Technologies.)

baseline noise, drift, and wander. Assuming that a minimum of 10–20 data samples are required to define a peak accurately, pneumatic modulation restricts the minimum peak width to approximately 1 s and is not suited for fast chromatography unless makeup gas is used.

The dynamic range of the TCD is $>10^4$ and its detection limit is <1 ppm. Samples that react with the filament can reduce the TCD's sensitivity and create instability in the baseline. Nonmodulated designs are particularly sensitive to common mode signals such as flow and temperature variations, and require long stabilization times. Generally, it is difficult to operate a nonmodulated design for high-sensitivity field operations. The pulse-modulated single-filament TCD showing the inertness and dynamic range of the detector is shown in Figure 51.

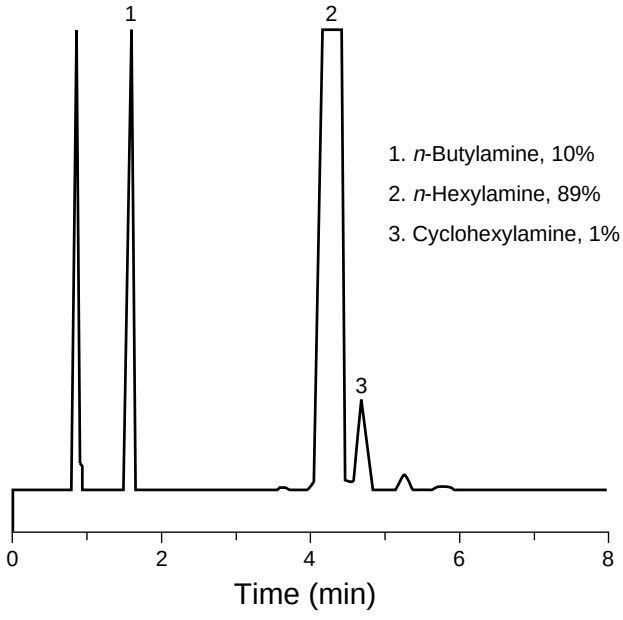

Figure 51 TCD separation of amines: 1 mL injected and split 20 : 1. Column: 10 m × 0.53 mm methylsilicone. Helium carrier gas at 21 cm s^{-1}. Isothermal at 80 °C.[92] (Courtesy of Agilent Technologies.)

4.15.7 Electrolytic Conductivity Detector

The ELCD can be used to detect halogen-, sulfur-, or nitrogen-containing compounds selectively. This is done by mixing the effluents with a reaction gas (oxidizing or reducing, depending on the analysis) in a reaction tube (usually nickel). The resulting products are then mixed with ionizing solvents, producing a conductive solution (Figure 52). This conductivity is measured and recorded.

In the case of halogen analysis, the column effluent is reduced with H_2 at 850 °C in a nickel reaction tube, producing strong acids (e.g. hydrogen chloride from chlorinated components). This gaseous product is then dissolved in *n*-propanol, and the change in solvent conductivity is measured. Other reduction products besides the hydrogen halides do not ionize and, therefore, do not cause an increase in conductivity in the slightly acidic *n*-propanol solvent.

The sensitivity for chlorinated compounds is <1 pg and the dynamic range is 1×10^6. For sulfur, the sensitivity is 2 pg and the dynamic range is 1×10^4. For nitrogen, the sensitivity is 4 pg and the dynamic range is 1×10^4. Selectivity is 100 000 : 1 by weight of halogen relative to carbon and 10 000 : 1–100 000 : 1 for sulfur and nitrogen, depending on the compound and element.

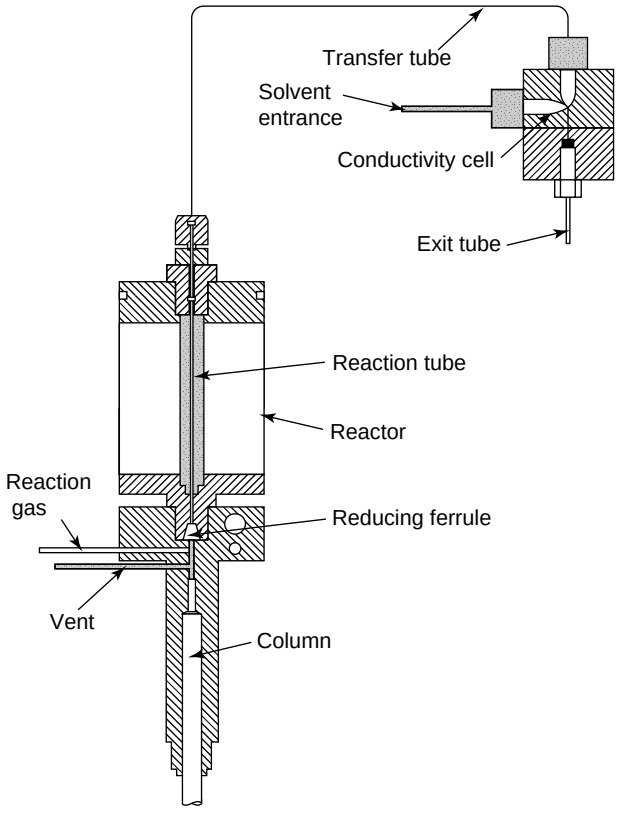

Figure 52 Design of a commercially available ELCD. (Courtesy of OI Corp., College Station, TX.)

Higher reactor temperatures (950 °C) are necessary for compounds that are more difficult to reduce, such as PCBs. Using different solvents and gases, the ELCD has been made to detect heteroatoms other than halogens. Good results have been reported for sulfur and nitrogen selectivity.

4.15.8 Summary

Figure 53 shows how common detectors compare with respect to sensitivity and dynamic range. An MDQ below the nanogram level (in parts per million) can routinely be detected. An MDQ at the picogram level (in parts per trillion) is at the limit of detectability of most detectors operating without preconcentration. The dynamic range of GC detectors is in excess of 4–5 decades.

Table 9 summarizes the main characteristics of GC detectors in terms of MDQ, dynamic range, and selectivity. All these detectors can be adapted to field operations but some of them will require more attention than others when used to maximum sensitivity. The main difficulties for the user are the time it takes to stabilize the detector and its susceptibility to environmental parameter variations, such as ambient temperature.

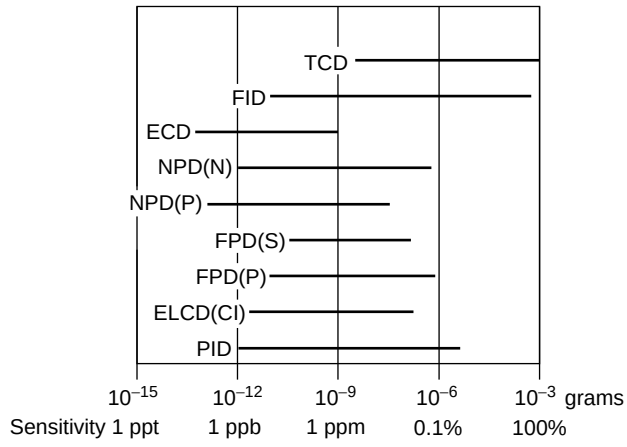

Figure 53 Comparison of various common GC detectors.[115] (Courtesy of Agilent Technologies.)

Other detectors such as mass spectrometers and ion mobility spectrometers are used for field operation and will be discussed below. The AED has field applications for which unequivocal analytical information is needed.[118] It has been successfully tested in various quantitative applications, such as the evaluation of lead, sulfur, and oxygenates in gasoline. The AED has also been used to identify base neutral pollutants, herbicides and PCBs. It has also been used successfully in unfriendly territory for treaty verification.

The appropriate detector is determined by the class of compounds to be analyzed. Some of the applications of the various detectors are summarized in Table 10.

The FID and the PID have long been used in hand-portable units, and their ruggedness has been demonstrated under difficult conditions. The NPD is similar in design to the FID; it can be operated easily and provides valuable information.

The TCD, at least in the flow modulation mode of operation (one filament), can also provide reliable data and can stabilize in a matter of minutes. When operated in a differential mode at the maximum sensitivity, the dual-filament TCD detector has a tendency to drift endlessly. Since in this mode the TCD relies on the equality of resistance between the reference and detection filaments, it is also susceptible to aging and becomes more sensitive to common mode signal variation after the resistance of one filament changes irreversibly. This may happen when the detection cells see a corrosive sample. Again, the single-filament TCD is insensitive to common mode signal variations.

The ECD has many potential field applications but may not be practical, since it can be easily contaminated; for this reason it requires a lot of attention when operating in the low range of detectability and may not be practical. The FPD is difficult to operate quantitatively in the field

Table 9 Summary of main characteristics of GC detectors

Name	Type	Selective for	Typical minimum detection level (S/N = 2)	Linear dynamic range
FID	Selective	Material that ionizes in air–H_2 flame	5 pg C s^{-1}	10^7
TCD	Universal	Anything with thermal conductivity different from carrier	400 pg mL^{-1} carrier	$>10^4$
ECD	Selective	Gas-phase electrophores	0.1 pg Cl s^{-1} (varies with structure)	10^4
PID	Selective	Compounds ionized by UV	2 pg C s^{-1}	10^7
Thermoionic	Selective	N, P, heteroatoms	0.4 pg N s^{-1} 0.2 pg P s^{-1}	10^4
ELCD	Selective	Halogens, N, S	0.5 pg Cl s^{-1} 2 pg S s^{-1} 4 pg N s^{-1}	10^6 10^4 10^4
FPD	Specific	P, S	20 pg S s^{-1} 0.9 pg P s^{-1}	10^3 10^4
FTIR	Universal	Molecular vibration	1000 pg strong absorber	10^3
MSD	Universal	Tunable for any species	10 pg to 10 ng (depending on SIM vs scan)	10^5
AED	Universal	Tunable for any element	0.1–20 pg s^{-1} (depending on element)	10^4

AED, atomic emission detector.

Table 10 Choice of GC detectors based on class of compounds to be analyzed[a]

Compounds	Detectors
Fixed gases	TCD, PID
Hydrocarbons	FID, PID
Halogen compounds	ECD, ELCD
Nitrogen compounds	NPD, ELCD
Sulfur compounds	FPD, ELCD
Phosphorus compounds	FPD, NPD
Oxygen compounds	IRD
Isomers	IRD

[a] The FID detects the organic portion of the compounds; the MSD is universal and gives structural information.

because it requires a calibration for each compound; also, the sensitivity of the FPD varies with the time the flame has been lit. Finally, the ELCD has some potential applications for field measurements but is bulky and not designed for field operation.

4.16 Portable and Transportable Gas Chromatographs

This section gives general information about field gas chromatographs and describes some of the instruments available in 1996. The intent is not to promote or criticize the models selected but to provide the reader with a tool for selecting an instrument by understanding the pros and cons of each.

The market for portable gas chromatographs is growing steadily, and manufacturers are developing innovative ways to reduce the size and weight of GC instruments. Each new design represents a trade-off between functionality and portability.

High boiling-temperature compounds require the GC column to operate above ambient temperature and, therefore, necessitate larger amounts of electrical power than can be realistically supplied by batteries. The most common trade-off for portability is to provide analysis of only VOCs and to run the column and the detector just above ambient temperature. This approach allows the instrument to be operated from battery packs for several hours.

Portable instruments usually eliminate interactive interfaces, such as keyboards and displays, and provide only computer interfaces to reduce weight and size. However, this compromise places the burden on the user to provide a computer device or network.

One method of reducing the chromatograph to its simplest functionality is to limit its analytical capabilities. The column used is a narrow-bore column of length >1 m, which is long enough to give a crude separation but no baseline separation of complex mixtures. The column is heated to several hundred degrees in a crude oven and has enough separation ability to serve as the front end of a mass spectrometer or an ion mobility spectrometer for which the peaks can be identified without baseline separation.[73,119–121] In this type of trade-off, the burden of analytical separation shifts from

Table 11 Portable gas chromatographs from PE Photovac

Feature	Voyager	SnapShot
Size	27 cm (10.6 in) wide, 39 cm (15.4 in) long, 15 cm (5.9 in) high	12.7 cm (5 in) wide, 35.6 cm (14 in) long, 23 cm (9 in) high
Weight	6.8 kg (15 lb)	4.4 kg (9.7 lb)
Detector(s)	Dual ECD–PID, PID with 10.6 eV	Photoionization with 10.6 eV RF-excited electrodeless discharge tube
Keyboard	4 Command keys, 4 soft keys	Keypad with 19 key silicone with tactile feedback
Display	128 × 64 Element graphical LCD with backlighting	7-line, 16-character LCD for alphanumeric display
Battery capacity	Standard battery, 6 h discharge and 5 h charge; extended life battery, 9 h discharge and 5 h charge; Ni–Cd (replaceable packs)	4–6 h; snap-on, rechargeable lead–acid pack
Power	10–18 V DC 115 or 240 V AC; adapter provided	N/A
Serial output	RS232, 9600 baud for connection to printer or PC	RS-232 (300–19 200 baud) for tabular and graphical printouts
Alarm output	Internal audio, 86 dB; alarm LED	Internal audible indicator for alarm
Intrinsic safety	Class I; division 1; groups A, B, C, and D	Class I; division 1; groups A, B, C, and D; intrinsically safe for use in Zone 1 locations; Eex ibm IIC T4
Chromatography	3-Column set with isothermal wide-bore capillary columns in a precolumn–backflush configuration; auto and syringe inject; exclusive PE Photovac all-metal valve array	Wide-bore capillary columns
Minimum detectability	10 ppb for benzene, halohydrocarbons, oxygenates, and hydrogen sulfide	0.1 ppm for benzene and ethylene oxide
Maximum oven temperature	Isothermal operation at 80 °C	Isothermal operation at 40 °C
Operating environment conditions	0–40 °C (50–105 °F), 0–100% RH (noncondensing)	0–40 °C (50–105 °F), 0–100% RH (noncondensing)

the gas chromatograph column to the sophisticated detector and data interpretation.

4.16.1 PE Photovac

PE Photovac has two notable portable instruments: the SnapShot GC, which has a low weight 4.5 kg (<10 lb), and the Voyager, which is a combined gas chromatograph and total VOC analyzer with built-in computation and data-logging capabilities. The Voyager is also low in weight, 6.8 kg (15 lb). The main features of these two instruments are summarized in Table 11. The column used in both of these instruments operates isothermally at <100 °C and is, therefore, used primarily for VOCs. Raw data are stored during analysis and are available later for data reduction. Each of these instruments can be connected to a network for either remote control or data retrieval. Both have a small canister for the supply of carrier gas and a battery pack that allows typically 4–9 h of field operation; both have internal sampling pumps for automatic headspace analysis of water or soil.

An exploded view of the SnapShot instrument is shown in Figure 54. This instrument is application-driven and operates with interchangeable modules. Factory-programmed application modules include Compound Library, Analyzer Set-Up, and Analytical Method Software for Automatic Compound Identification and Measurement.

The Voyager portable GC is shown in Figure 55. This instrument has a number of unique features, including a built-in three-column set for analysis of VOCs with a wide range of boiling points and molecular weights. A dual PID–ECD system allows the determination of both chloroalkane and chloroalkene solvents of importance in environmental monitoring. Inorganic gases such as hydrogen sulfide, carbon disulfide, nitric oxide, arsine, and phosphine can also be detected. Confirmation analysis can also be performed by injection of the sample on each of the three columns, each of a different polarity. The Voyager GC, like the SnapShot GC, has Intrinsic Safety Classification, which makes it useful for hazardous confined-space entries to refineries and chemical plants.

4.16.2 MTI Analytical Instruments

MTI Analytical Instruments (Fremont, CA, USA) uses silicon wafer fabrication technology to micromachine

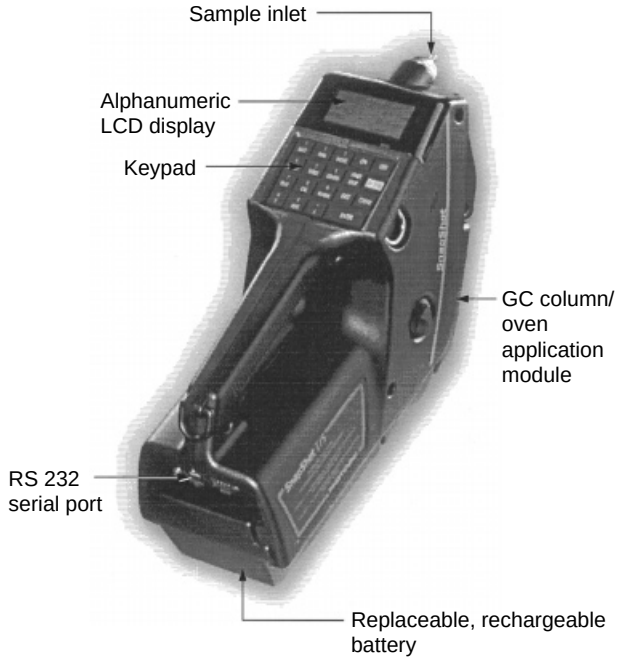

Figure 54 The SnapShot hand-portable gas chromatograph. (Courtesy of PE Photovac.)

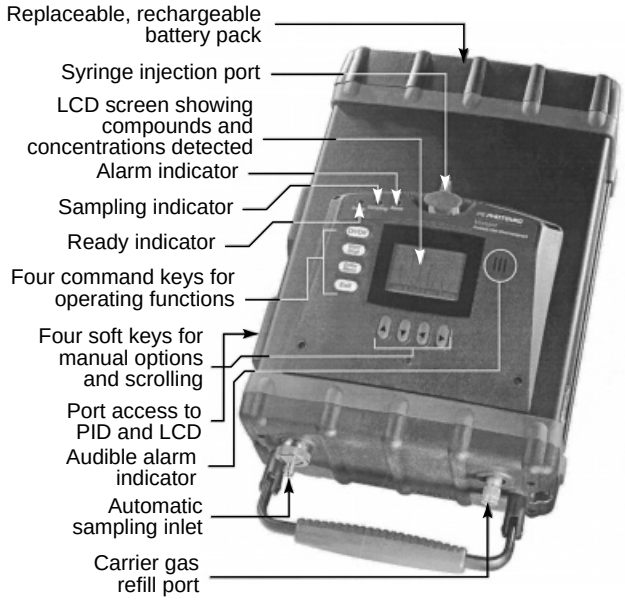

Figure 55 Portable gas chromatograph, model Voyager GC. (Courtesy of PE Photovac.)

parts of its portable gas chromatograph. This technology combines precision etching (micromachining) and insulator-to-silicon (anodic) or silicon-to-silicon (fusion) multilayer bonding with standard integrated circuit processing techniques, such as photolithography, diffusion, ion implementation, and thin-film deposition.

MTI Analytical Instruments combines the varied aspects of silicon micromachining technologies to fabricate a gas injector and a TCD (Figure 56).[122] The gas injector is a hybrid combination of micromachining technology and conventional machining.[123] The small physical dimensions of the valves and restrictor channels result in the sample reaching the column as a sharp plug.

The small internal volume of these micromachined parts has allowed MTI to design a portable gas chromatograph that has little extra column dead volumes. The instruments can, therefore, use short, narrow-bore capillary columns and perform a typical gas analysis in 20–30 s,[124] whereas several minutes are needed for conventional gas chromatographs (Figure 57).

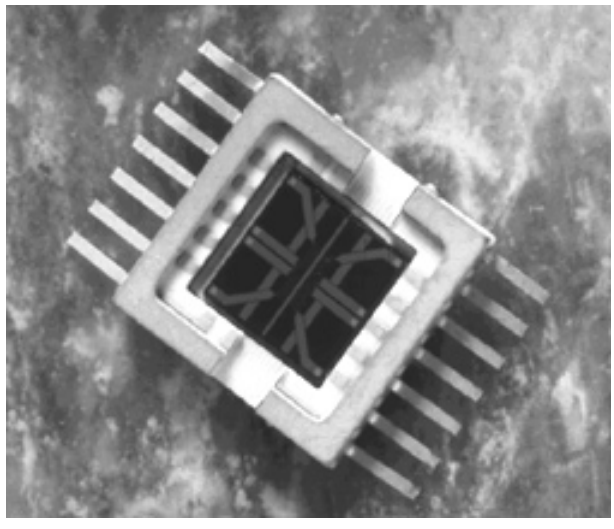

Figure 56 Micromachined TCD. (Courtesy of Agilent Technologies.)

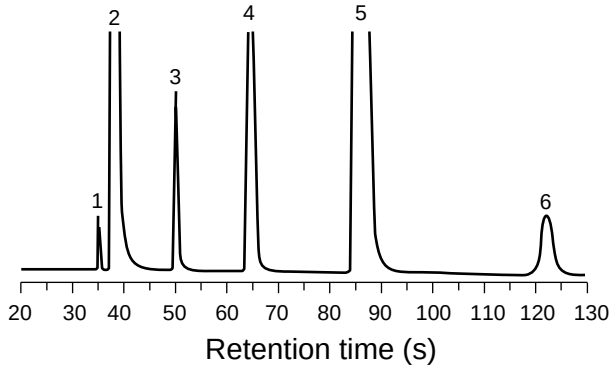

Figure 57 High-speed chromatogram of a gaseous sample obtained with an M200. Peaks: 1, helium; 2, hydrogen; 3, oxygen; 4, nitrogen; 5, methane; 6, carbon monoxide.[125] Column: molecular sieve, 5 Å, 10 m × 0.32 mm ID; 30 µm film thickness, pressure 29.7 psi; carrier gas, argon; temperature, 110 °C; injection time, 10 s; and run time, 100 s. (Reprinted with permission of Agilent Technologies.)

Table 12 Portable gas chromatograph features of Models P200 and P200H from MTI

Feature	P200	P200H
Size	15 cm (6 in) high, 36 cm (14 in) wide, 36 cm (14 in) deep	15 cm (6 in) high, 36 cm (14 in) wide, 41 cm (16 in) deep
Weight	10.4 kg (23 lb)	13.2 kg (29 lb)
Samples	Only gases with boiling points up to 150 °C (typically C_1-C_9)	Only gases with boiling points up to 220 °C (typically C_1-C_{12})
Detector	Silicon micromachined thermal conductivity with 240 nL internal volume; linear dynamic range $10^5 \pm 5\%$	
MDQ	1 ppm for most compounds	
Repeatability	±2% RSD at constant temperature and pressure	
Column heater range	30–180 °C isothermal operation	50–180 °C isothermal operation
Carrier gas	Internal carrier gas supply via refillable 300-mL tank; 1800 psi DOT certified for 40 h continuous operation, externally connectable supply of He, N_2, or Ar	
Power	Internal 12-V DC rechargeable lead–acid battery or 12-V DC power supply; 6–8 h; 30 W maximum	Internal 12-V DC rechargeable lead–acid battery or 12-V DC power supply; 4 h; 60 W maximum
External outputs/inputs	Analog output 0 ± 1 V DC and 0 ± 10 V DC, RS-232 for instrument control and data collection	
Keyboard/display	None, requires an external computer	
Environmental	Operating temperature range 0–50 °C; humidity range 0–95% noncondensing	
Chromatography	Small-bore columns	

Table 12 lists some of the features of the low- and high-temperature versions of the P200. Each version has battery packs and gas supplies that make field operation possible for several hours (depending, of course, on the amount of power drawn from the battery pack). The instruments accept gaseous samples only and can operate in isothermal conditions. The two modules differ in stationary phase (column material), operating temperature of the column, and injected volume. The sample transfer line and the injector operate at lower temperatures than the column and the detector. This design may create cold spots and condense some high boiling-temperature compounds. Therefore, it should be considered primarily for VOCs. It cannot be used for the injection of a liquid sample, but vapors with high boiling points up to 220 °C can be analyzed.

Figure 58 shows an inside view of the P200 Micro GC. The sample inlet, at the front of the instrument, is heated to 110 °C to prevent water condensation. A gaseous sample is pumped into the instrument with a built-in pump and analyzed in up to four GC modules simultaneously. A rechargeable carrier gas tank provides up to 40 h of operation. The rechargeable battery pack can operate a maximum of 8 h before needing to be recharged.

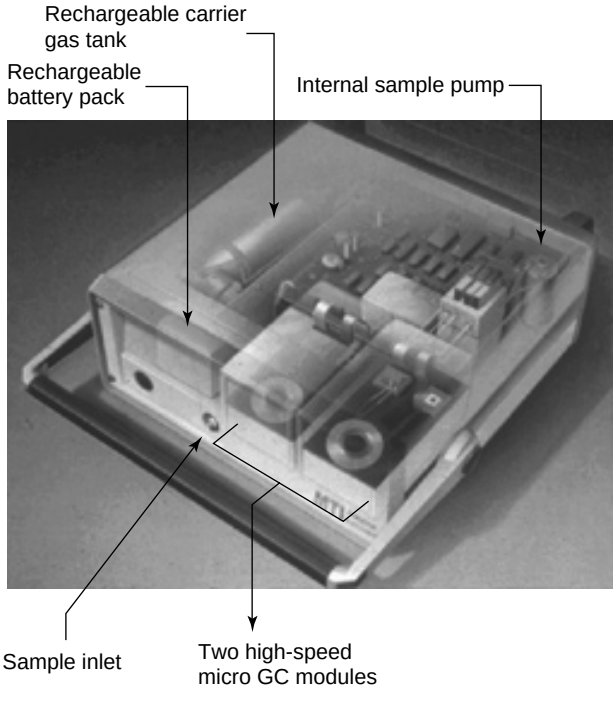

Figure 58 Portable instrument, Model P200, from MTI. (Courtesy of Agilent Technologies.)

4.16.3 Model 311D from HNU Systems Inc.

HNU offers a full VOC/SVOC gas chromatograph for operation in the field or in a laboratory. This instrument includes a column oven with multiramp temperature programming or isothermal operation at up to 200 °C, heated syringe injector and detectors (one or two), sample pump with automatic gas sample valve, or a Tenax® sample concentrator/injector. Detector options include PID, far-UV, ECD, TCD, FPD, and FID. Detectors can be configured singularly or in series, with single or multiple columns in the oven. The gas chromatograph includes a built-in carrier gas supply and requires an external AC to operate. The features of the 311D

Table 13 Main characteristics of Model 311D field gas chromatograph from HNU

Feature	Specifications
Size	56.4 cm (22.5 in) wide, 24.6 cm (9.8 in) deep, 39.1 cm (15.2 in) high
Weight	25 kg (55 lb) including metal transportation case
PID	Lamps, 11.7 (11.8) eV, 10.2 (10.6) eV, 9.5 eV; minimum detectable level 10.2 eV: <1 ppb benzene; linear dynamic range, 10^7; detector volume, <100 µL
ECD	Source, 10 mCi ^{63}Ni; minimum detectable level, 1.8 ppb SF_6; linear dynamic range, 10^4; detector volume, 270 µL
Far-UV detector	Source, 10.2 eV (124 nm) UV lamp; minimum detectable level, 1.8 ppm dibromomethane, 2 ppm CFCs; linear dynamic range, 10^4; detector volume, 40 µL
FPD	Compounds detected, sulfur, phosphorus; Filters, rare earths; minimum detectable level, 100 ppb H_2S; linear dynamic range, 10^5
TCD	Type, differential bridge; minimum detectable level, 100 ppm O_2; linear dynamic range, 10^4
FID	Minimum detectable level, <1 ppm *n*-pentane; linear dynamic range, 10^6
Power	110, 115, or 230 V AC; 3 A (110 V AC)
Output	RS-232 for GC control via PC (Windows software included); analog output, 0–1 and 0–10 V; chromatogram, analysis, TWA, and STEL reports with built-in printer
Alarms	Two programmable alarm levels for each compound
Chromatography	Packed or capillary, single, or multiple columns
Inlet	Heated liquid–gas multipurpose syringe injector port; sample pump with 10-port, precolumn backflush sample valve; sample pump with Tenax® sample concentration/injector
Carrier gas	Onboard supply for 12–24 h of operation
Calibration	Automatic, at user-defined internals
Column oven	Isothermal operation to 200 °C in 1 °C increments; 1–15 °C min^{-1}; heat-up, 50–200 °C in approximately 10 min; cool-down, 200 to 70 °C in <10 min
Operating environmental conditions	Ambient temperature, −10 to 35 °C; ambient humidity, 20–95% (noncondensing)

are given in Table 13, and the instrument is shown in Figure 59. In the figure, the gas chromatograph is controlled from an external PC. The carrier gas supply is in the detachable instrument cover. The heated injector is used for syringe injections of liquids or gases. Built into the GC is a sample pump and either a gas sample valve or Tenax® gas sample concentrator/injector for single or continuous analysis. The concentrator improves the lower detection levels for most compounds.

Figure 59 Transportable gas chromatograph, Model 311D, shown with a laptop computer used for data handling and control. (Courtesy of HNU Systems, Inc.)

The performance of this gas chromatograph has not been degraded for portability; the performance of the 311D in the field is similar to that of a laboratory gas chromatograph. One way to check the performance of an instrument is periodically to inject a test mixture and compare the resulting chromatogram with a standard.[125,126] The measured retention times and relative peak heights should not vary from the standard by more than some agreed-on value. For example, according to HNU Systems, the repeatability of retention times (which is a measure of temperature and flow stability) is better than 0.5%.

4.16.4 Agilent Technologies 6890 Gas Chromatography

Mobile laboratories are equipped with conventional nonportable instruments or combinations of instruments, such as GC/MS. Field applications requiring versatile, high-quality chromatographic data rely on laboratory instruments operating from mobile shelters. The 6890 GC from Agilent Technologies (Palo Alto, CA, USA) is one such instrument used in the field.

The 6890 GC is at the extreme end of the spectrum of field instruments (Figure 60). It is not hand-portable, and requires a bench and several kilowatts of electrical power. This instrument is not designed to operate in extreme temperature and dust conditions; thus one could easily question its choice for field applications. However, it is versatile, given the range of detectors and the nature of samples it can analyze. The specifications for this model are shown in Table 14. A major advantage of the 6890 GC is that it operates with specific detectors, including MSD and AED, and thus provides positive identifications

Figure 60 Model 6890 gas chromatograph. The instrument is seen equipped with an autosampler. In a more elaborate version, the gas chromatograph is equipped with a mass spectrometer or an elemental analyzer. (Copyright 1997, Agilent Technologies. Reproduced with permission.)

that cannot be achieved with conventional GC detectors for gas, liquid, and solid samples. Since this instrument must operate inside a shelter, it can be connected to sophisticated data reduction and communication tools, providing an integrated solution to high-quality analysis.

4.16.5 General Comments

GC is a simple technology that is inexpensive and powerful. This technique has been used for several decades and still has an attractive future. The development of portable gas chromatographs is relatively recent and shows that the field is growing, although the performance limitations of these instruments still require the field operation of conventional instruments. Portable gas chromatographs available today do not yet offer the same versatility as laboratory instruments. However, by operating directly at the site, they can give valuable on-the-spot qualitative analytical information.

The major advantage of portable gas chromatographs lies in their ability to make measurements directly in the field, thus eliminating the transport and manipulation of samples. If used properly, they can provide limited but immediate information about the nature and concentration of compounds. The more specialized portable instruments provide unskilled users with tools adapted to their applications. The Photovac SnapShot offers modules for specific applications and is a good example of a clever compromise between portability and versatility.

''Just enough GC'' is an interesting approach applied to the design of portable instruments. This concept, combined with the wireless networking of analytical instruments in the field, opens the market for totally new types of portable instruments. The portable gas chromatograph is a sensor under interrogation by a remote unit.

Portable instruments are limited by the range of samples that they can handle. For instance, it is not possible today to analyze semivolatile compounds with a portable unit operating on batteries, since the amount of power necessary to heat the oven and the column cannot be supplied by battery packs. Only alternative thermal designs to the conventional heating of the column by a convection oven may provide a solution to this limitation.

Table 14 Specifications of 6890 Series gas chromatograph from Agilent Technologies

Feature	Specifications
Size	58.5 cm (23 in) wide, 50 cm (20 in) deep, 50 cm (20 in) high
Weight	50 kg (110 lb)
FID	Maximum temperature, 450 °C; minimum detectable limit, <5 pg C s^{-1} propane; linear dynamic range, 10^7
NPD	Maximum temperature, 400 °C; minimum detectable limit, <0.4 pg N s^{-1}, <0.2 pg P s^{-1}; linear dynamic range, 10^5; selectivity, 35 000 : 1 g N g^{-1} C^{-1}, 75 000 : 1 g P g^{-1} C^{-1}
TCD (single filament)	Maximum temperature, 400 °C; minimum detectable limit, <400 pg propane mL^{-1} helium carrier; linear dynamic range, 10^5
ECD	Maximum temperature, 400 °C; source, 15 mCi ^{63}Ni; minimum detectable level, <0.04 pg s^{-1} lindane; linear dynamic range, 10^4
MSD	Contact Hewlett-Packard
AED	Contact Hewlett-Packard
Power	110, 115, or 230 V A, 2250 V A, 50/60 Hz
Output	RS-232, HP-IB communication port; 2 analog outputs
Column	Packed or capillary
Oven	Oven size: 28 cm × 31 cm × 16 cm (11 × 12 × 6 in); temperature range, from −80 (with liquid nitrogen) up to 450 °C; maximum temperature rate, 120 °C min^{-1}; ambient rejection temperature $<100:1$
Inlet	Maximum temperature, 450 °C; split/splitless, cold on-column; heated multipurpose liquid–gas sample injector for all types of packed and capillary columns (heated from ambient to 200 °C); autosampler
Operating environmental conditions	Ambient temperature, 5–40 °C; ambient relative humidity, 5–80% (noncondensing); on-board compensation for ambient temperature and atmospheric pressure variations

The GC hardware is also a limiting factor. When conventional GC hardware is used with portable instruments it increases the power demand and size of the instrument to unacceptable levels. Most of the injectors and detectors (with only a few exceptions, such as the microthermal conductivity hardware developed by MTI Analytical Instruments) have been designed for laboratory instruments. By microfabricating its detector with technologies borrowed from the semiconductor industry, Agilent Technologies has demonstrated that small-volume detectors can be developed that are well adapted to small-bore columns. The result is faster analysis of field samples.

The major challenge faced by the manufacturer of field gas chromatographs is to determine the maximum range of compounds that can be analyzed by a given instrument. HNU, with the Model 311D, can operate at up to 200 °C and accepts a broad range of columns, including packed columns. HNU also offers a range of specific detectors for analysis of gases, hydrocarbons, and halogen compounds. The result is versatile instruments that can analyze PCBs, perform static headspace analysis, or monitor landfill gas. The price to pay for this versatility includes their larger size and weight and the fact that they need greater electrical power.

At the upper end of price and size, the 6890 GC from Agilent Technologies is by far the least portable and field-operable instrument, but it offers several valuable advantages when high-quality data are needed. Since this instrument must operate in a mobile laboratory, it can operate with a broad range of nonconventional detectors such as MSDs and AEDs. The analytical tools provided give the user the capability to get the same quality and versatility of measurement for gas, liquid, and solid samples as in the laboratory. Although the instrument operates in the field, there is still a need to transport samples to the mobile laboratory, which can be remote from the measurement site with the risk of sample degradation that this implies.

5 MASS DISPERSION METHODS FOR ENVIRONMENTAL MEASUREMENTS

Two popular techniques illustrate analysis based on analyte mass or ion size: MS and IMS.

The former method is widely accepted by government agencies such as the USEPA. It is applicable to most of the compounds found in air, soil, and water at a detection level of a few parts per billion; differentiation of molecules is based on their mass. IMS uses ion size to differentiate between analytes. This is a truly hand-portable technique in which the chemistry has been adapted to a specific application. Both techniques are used with GC to effect a separation before analysis.

5.1 Mass Spectrometry

MS is a microanalytical technique that requires only a few nanomoles of sample to obtain characteristic information regarding the structure and molecular weight of the analyte. It is a destructive technique in which the sample is consumed during analysis. In all cases, some form of energy is transferred to the analyte molecules to effect ionization. In most cases the nascent molecular ions of the analyte ''explode'' in a predictable fashion into a variety of fragment ions; the resulting fragmentation pattern constitutes the mass spectrum.[127] In principle, the mass spectrum of each compound is unique and can be used as a chemical ''fingerprint'' to characterize the sample.

Figure 61 gives a synopsis of the entire process of analysis by MS, culminating in the bar graph mass spectrum often published the literature.[127]

For analysis by a conventional mass spectrometer, the sample must have a vapor pressure of $>1.3\,\mathrm{N\,m^{-2}}$ ($>10^{-2}$ mmHg), because molecules of the sample migrate by diffusion from the inlet system into the ionization chamber. The inlet may be either a direct probe or a GC inlet. As the neutral molecules randomly diffuse throughout the ion source to the ionization chamber,

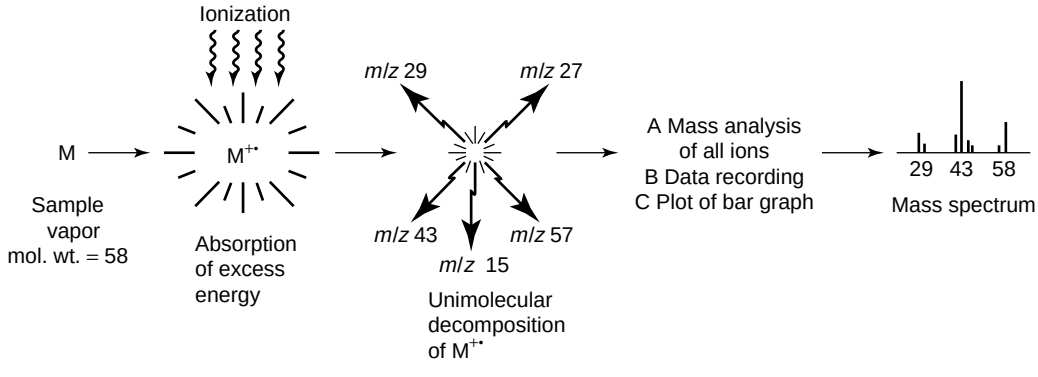

Figure 61 Synopsis of analysis by MS. This example illustrates the fragmentation of the original vapor sample into fragment ions.[128] M, molecules of the sample in the vapor state.

only a few hundredths of a percent of them are ionized. This fraction is referred to as the ionization efficiency.

The most common ionization process, electron ionization (EI), is achieved by bombarding the sample with electrons at 70 eV energy. The ionization process in general is nothing more than the transfer of energy to the neutral molecule in the vapor stage, giving the neutral molecule sufficient energy to eject one of its own electrons and, therefore, become charged with a residual positive charge. This process produces a molecular ion with a positive charge (Figure 61). This molecular ion still has considerable energy, and much of that energy can be dissipated by fragmentation of its chemical bonds.

This decomposition of various chemical bonds leads to the production of fragment ions whose mass is equal to the sum of the atomic masses of the group of atoms retaining the positive charge during the decomposition process. It is important to realize that not all of the molecular ions decompose into fragment ions. Molecules producing a molecular ion that is stable tend to survive or not to fragment, and an intense molecular-ion peak is recorded. During the analysis of a compound having a molecular ion that is unstable, nearly all of the molecular ions decompose into fragment ions, and in these cases the mass spectrum contains only a small peak for the molecular ion.

The task now is to analyze all the ions created in the ionization chamber according to mass. Ions have an electrical charge that permits them to be focused by various ion optics potentials to form a well-collimated ion beam. This beam of ions is accelerated out of the ionization chamber and into a mass analyzer.

5.2 Mass Analyzers

The mass analyzer can be any one of several types: quadrupole (which includes quadrupole filter and ion trap), Fourier transform mass spectrometry (FTMS), ion cyclotron resonance (ICR), time-of-flight (TOF), and magnetic sector. Regardless of which is used, the result is the same. The ions are analyzed according to their abundance along a mass scale. During the process of data recording and reduction, the data can be arranged in tabular form or in bar graph format. Additional details concerning the design and operating principles of mass spectrometers in general can be found in the literature.[127–129]

5.2.1 Time-of-Flight

In TOF mass spectrometers, ions are given a kinetic energy with a high-voltage pulse. Another important parameter is the duty cycle of the instrument, defined as the ratio of the time when the TOF is operating to the period between pulses. For instance, in orthogonal acceleration TOF, the duty cycle is the ratio of fill time of the sample in the deflector to the period of the repetition pulses. The duty cycle is a measure of the efficiency of measurement of the instrument. The higher the duty cycle number, the more efficient is the TOF.

Ions then drift through a tube to a detector (Figure 62). The length of the tube (between a fraction of a meter and several meters) determines the resolution of the TOF. The longer the drift tube, the better is the resolution. Although all ions leave the source together, ions with smaller mass-to-charge ratios accelerate to a higher velocity than heavy ions of the same charge and will reach the detector first. Identification is done by measuring the transit time of the ion. Since the typical drift time for the largest ions is $<50\,\mu s$, the accelerating pulse could be repeated at approximately 20 kHz.

In a TOF system, a burst of ions (with a distribution of masses) that is accelerated to a constant energy will eventually be resolved in space (ions of similar mass will be grouped together), because each ion is travelling at a different velocity. The $w = zeV = (1/2)mv^2$ equation predicts an inverse quadratic relationship between v and m/z.

In this instrument, some ions reach the detector at different times, and the m/z value of an ion is determined by its time of arrival at the detector. Ions of low mass reach the detector before those of high mass. Voltages within the system, except for the ''pulser'', are constant (typically from a few hundred volts to 1000 V DC). This contrasts with the other mass filter types, in which nonselected ions are filtered out by changing magnetic or electric fields.

There are two principal modes of operation in time-of-flight mass spectrometry (TOFMS): pulsed operation and continuous operation. In the continuous mode, ions are produced nearly all the time, thereby increasing the sensitivity of the TOF by at least one order of magnitude. The ionization process takes place for at least 70% of the TOF cycle (typically 100 μs for a mass range of 1000 Da) and is shut off only during the period of ion acceleration (typically a few microseconds). In the pulsed mode, the cycle is initiated by ionizing the molecules for typically 1 μs and then by pulsing the ions into the acceleration region for 1–2 μs, where mass separation is effected. The larger bundle of ions formed in the continuous mode

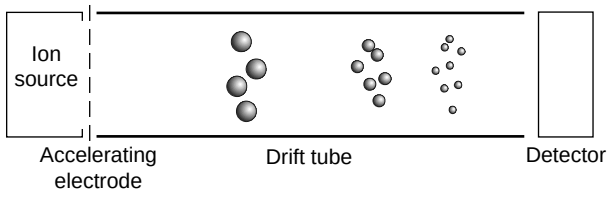

Figure 62 Basic description of a TOF mass spectrometer, showing the main components of the analyzer. (Courtesy of Agilent Technologies.)

leads to increased sensitivity but decreased resolution compared to the pulsed mode of operation.

TOFMS requires a data acquisition system capable of digitizing at a rate of about 100 MHz to maintain a timing resolution of 10 ns in the instrument. The data must be processed further at similar rates. Although TOFMS is well suited to acquiring full spectra, it is also well suited for SIM by integrating the ion current at each given m/z value for several TOF cycles.

Manufacturers are highly motivated to reduce the size of drift tubes to make their products more portable. With the use of V-shaped and linear reflectors, which are well suited for field operations, Kore Technology (Cambridge, UK) has been able to reduce the length of its Model T-CAT drift tube to 0.2 m.[130]

5.2.2 *Quadrupole Filter*

The quadrupole mass filter consists of four conductive rods that are used to shape an electrodynamic field. This nonmagnetic mass spectrometer employs a combination of DC and radiofrequency (RF) potential as a mass filter. Ideally, the four rods arranged symmetrically (Figure 63) should have the shape of a hyperbola in cross-section.[131,132] This can be accomplished, as in the case of the Hewlett-Packard Model HP5973A MSD, by using a single piece of 20-cm long quartz, with a portion of the inner surface made conductive. The magnitude of the field determines which ion will pass through the quadrupole to the detector.

The entire mass spectrum is scanned as voltages are swept from a pre-established minimum to a maximum value (approximately 400 V DC and 2000 V RF), but at

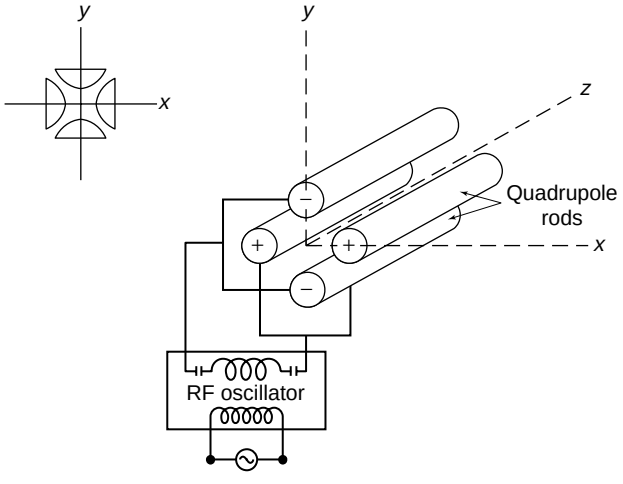

Figure 63 Diagram of a quadrupole mass filter showing the arrangement of quadrupole rods and the electrical connection to an RF generator. A DC potential (not shown) is also impressed on the rods. The inset illustrates the ideal quadrupole electric field dispersion. (Courtesy of Agilent Technologies.)

a constant DC : RF ratio. This ratio defines the operating stability, sensitivity, and resolution of the instrument.[127]

The quadrupole features reasonable sensitivity (picogram range) and scanning rate (several thousand daltons per second), which are adequate for use with small-bore GC columns that produce peaks in the hundreds of milliseconds range. Another advantage of the quadrupole is that it produces mass spectra with a constant mass resolution, i.e. the separation of the peaks at m/z 400 and 401 is the same as for m/z 30 and 31.

A quadrupole can be operated in one of two modes: scanning or SIM. In the scanning mode, voltages are swept from a pre-established minimum to a maximum value at a constant DC : RF ratio, therefore scanning the sample over the overall mass range of the analyzer. In the SIM mode, the DC : RF ratio is set constant to a value that sets a measurement to a predetermined m/z ratio. In the SIM mode, the analyzer acts as a narrow-band filter and rejects all the other masses. To summarize, the scanning mode leads to more versatility (no knowledge about the exact masses to be monitored is required) but less sensitivity than the SIM mode.

The quadrupole mass spectrometer is more capable of maintaining quality output at higher pressures than other types of mass spectrometers; this may be an advantage in GC/MS where a constant flow of carrier gas is supplied to the analyzer. The only harmful effect of higher pressures (10^{-4} Torr) is that once the mean free path is less than the tube length, molecular collisions result in loss of ions to the poles and thus loss of sensitivity linearly proportional to the increase in pressure.

The quadrupole is well suited for SIM because selected ions from any region of the mass spectrum can be monitored without altering the optimum conditions in the ion source or mass analyzer. Furthermore, the parameters (superimposed RF and DC fields) that control the mass scale can be changed rapidly with good response and are well stabilized through the mass range. These features are particularly desirable for computer-controlled operation of the MS for ion monitoring.

Quadrupole MS systems produce EI spectra that can be compared with accepted classical library spectra for positive identification of either target compounds or unknowns. A major advantage of this feature is the predictability of data. Even if the spectrum of an unknown is not in a library, the spectrum can be interpreted by using isotope ratio calculations and well-established fragmentation patterns.

One limitation is that the sweep time cannot be reasonably reduced to below 10 ms, thereby limiting the rate of acquisition of chromatographic signals to a peak width of approximately 100 ms. When the ratio of peak width to sweep time is too small (i.e. 10 : 1), the spectra can be distorted by the slope of the GC peak, for

which heavier masses are augmented relative to lighter ones on the GC peak's upslope, and diminished on the downslope.

5.2.3 Ion Trap

In the ion trap, ionization and mass filtering occur in the same place.[127] A conceptual schematic of the ion trap detector (ITD) is shown in Figure 64.[133] The top end cap, a ring electrode, and bottom end cap are electrically insulated from one another. Sample molecules enter the central cavity of the ion trap from the inlet. A filament produces a beam of electrons that penetrates into the cavity and interacts with the molecules in the cavity. An electrical field is generated by an AC voltage applied to the ring electrode. This electrode is between caps on each end of the cavity. Ions of selected mass-to-charge ratios are trapped within the electrical field. Ions over a specified range of m/z values are trapped by appropriately selected RF and DC voltages.[134]

A mass spectrum is produced by ramping the electrical field, which ejects ions from the cavity according to their mass-to-charge ratio, thus producing a mass spectrum. The ion trap is scanned from low to high mass by increasing the voltage on the ring electrode.

The ion trap has some advantages over most other forms of mass spectrometer in terms of size, weight, and pumping requirements. These advantages make the ion trap attractive for field applications.

Under favorable conditions, almost 100% of the parent ions can be converted to product ions, although a 10–50% conversion is more typical. The analogous conversion in most beam-type MS experiments is one to three orders of magnitude lower, implying significant reductions in detection limits.

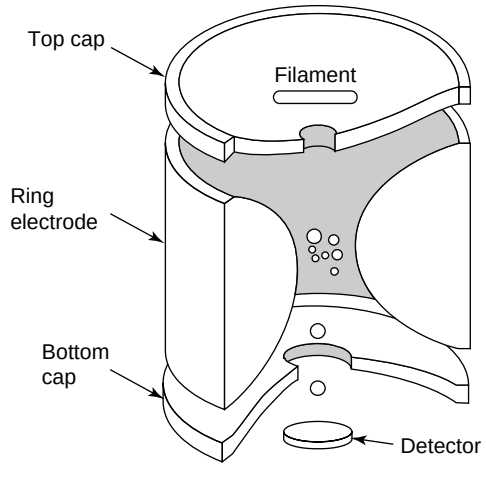

Figure 64 Exploded view of an ion trap. (Courtesy of Agilent Technologies.)

The major drawbacks of the ion trap as a mass analyzer arise from the much greater ion densities that can be encountered in an ion trap instrument relative to those encountered in ion beams in most analytical mass spectrometers. High densities of ions of like charge can adversely affect the ion trap experiment in several ways, such as imposing a limit on the number of ions that can be stored, thereby limiting considerably the dynamic range.

5.2.4 Ion Cyclotron Resonance Mass Spectrometer

Although a product based on ICR technology is not yet available, the technology is attractive and therefore will be discussed. One prototype of such an analyzer was developed by Lawrence Livermore National Laboratory (LLNL) (Livermore, CA, USA).

ICR is the operating principle of the Penning ion trap. Ions are created inside the trap, and analysis and detection are also done there. Both features reduce ion loss and allow the ions to be sampled many times. A traditional Penning ion trap is a rectangular box that sits inside a powerful electromagnet (Figure 65a). A gas is introduced into the box and ionized by an electron beam. Two opposing ends of the box receive a positive electrical charge; the four sides, forming a square tube along the magnetic field direction, receive a less positive charge. The positively charged ions are repelled by the positively charged ends, thus the electrical voltage prevents the ions from leaving the box along the direction of the magnetic field. If the positive ions try to leave the box across the direction of the magnetic field, i.e. toward any of the less positive sides, they are deflected by the magnetic field into cyclical orbits, called cyclotron motion.

The frequency of this cyclical motion is determined by the magnetic field and the ratio of the ions' electrical charge to their mass. Because the values of the magnetic field and the ions' electrical charge are known, the mass of a trapped ion can be determined from the frequency of its orbit. This method of determining an ion's molecular weight and thus its identity is known as ion cyclotron resonance.

Fourier transform algorithms have been applied to this technique, thereby improving its performance. For FTMS, the ICR instrument is operated in a pulsed mode. In Figure 65(a), a burst of electrons forms an assortment of ions in the cell. In the presence of a constant and uniform magnetic field (Figure 65b), all ions move in circular orbits with a characteristic cyclotron frequency w_c, which depends only on the ions' m/z value (Equation 21):

$$w_c = \frac{zeB}{2\pi m} \tag{21}$$

In Figure 65(b) a frequency-swept chirp signal is applied to the cell. Ions that experience oscillating fields equal

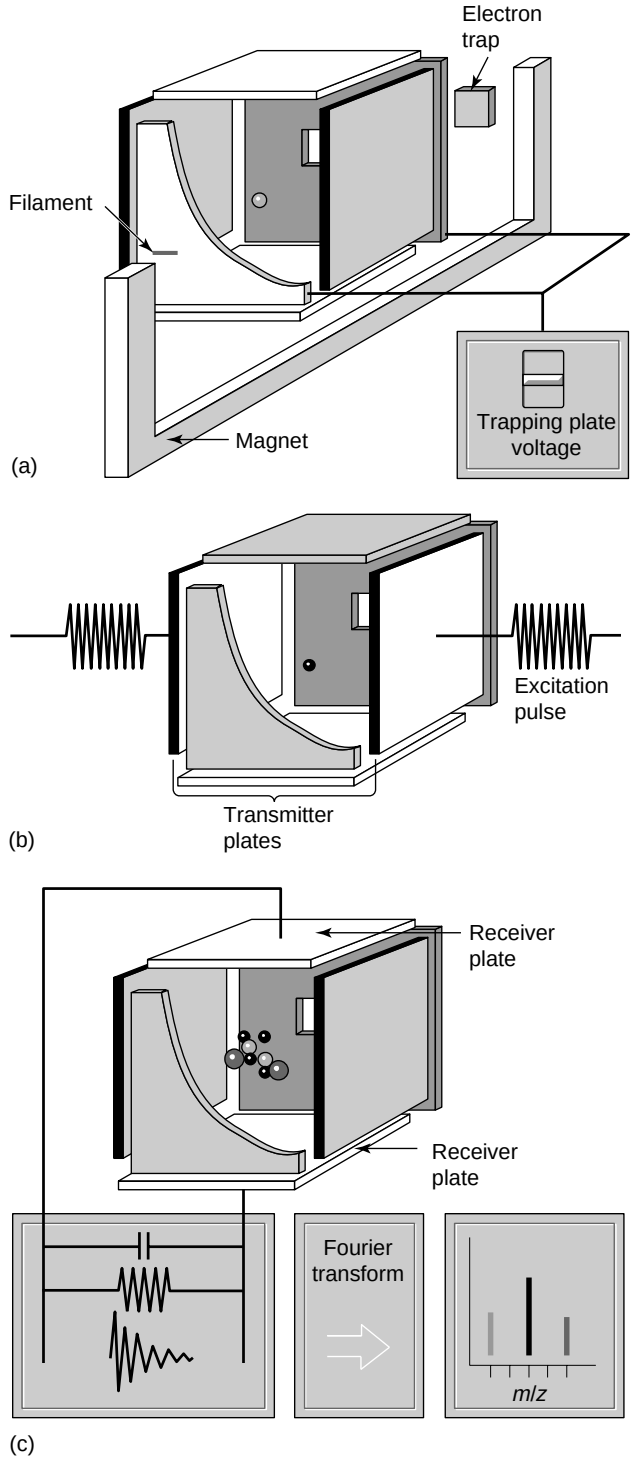

Figure 65 Schematic diagram of an ICR/MS system. (a) Basic configuration of the instrument, showing the trapping plates, the direction of the magnetic field, and the introduction of the ions into the cubic cell. (b) A frequency sweep is applied to the transmission plate (horizontal plates) forcing the ions into oscillation. (c) Ions move coherently between the receiver plates (vertical plates) where they are collected. Fourier transform is then applied to the signal. (Courtesy of Agilent Technologies.)

in frequency to their cyclotron frequency absorb energy from the circuit generating the chirps. During this process, these ions move to orbits of larger radii and become in phase with the exciting field. In Figure 65(c) the translationally excited ions move coherently between the receiver plates and establish an image current in the external conducting network attached to these plates. Because the chirp contains all the excitation frequencies, the ions transmit a complex RF signal that contains the frequency components characteristic of all of the ions present. The resulting total time-dependent image current is subjected to Fourier transformation, which resolves these components and produces a database for the mass spectrum. For low-resolution GC/MS, the pressure requirements are comparable to those of other common MS types, but to reach levels of performance identical to quadrupole or ion trap, high vacuums are likely to be needed.

FTMS is unique in that increased measurement times increase both sensitivity and resolution. This advantage derives from the fact that in FTMS the ions are not consumed during the detection process. Furthermore, all the ions at all masses are detected simultaneously in a process that could be viewed as frequency-array detection. The FTMS procedure provides an integrating effect that optimizes the signal-to-noise ratio and minimizes the problems due to a changing partial pressure of the analyte in the ion source, as in GC/MS applications.

At the present time there are not enough data to conclude that this technique is going to be a potential competitor for field applications. LLNL developed an award-winning prototype of an FTMS instrument based on ICR. The prototype of weight < 12.3 kg (33 lb) uses a battery, fits into a small attaché case, and has an accuracy of $1:10^3$. The permanent magnet is designed to optimize the magnetic field. It is cylindrical, 7.6 cm (3 in) across and 5.7 cm (2.3 in) high, with a center bore of slightly less than 2.5 cm (1 in). A thin-walled vacuum chamber containing the ion trap fits inside the hole. A small ion pump, along with a miniature cryogenic pump, maintains the tube's vacuum pressure. A miniature piezoelectric inlet valve admits gas to be sampled into the tube. There is also a pulse electron beam that ionizes the gas in the trap.[135] This unit is claimed to have a mass range of 3–1000 Da and a detection limit of 20–100 atoms, but it requires a high vacuum, in the region of 10^{-8} Torr, to operate with this kind of sensitivity, making it less attractive for a portable instrument coupled with a gas chromatograph.

5.2.5 Magnetic Sector

The magnetic sector mass spectrometer uses a magnetic field to separate ions (Figure 66). Sample ions are accelerated out of the ion source and through a magnetic

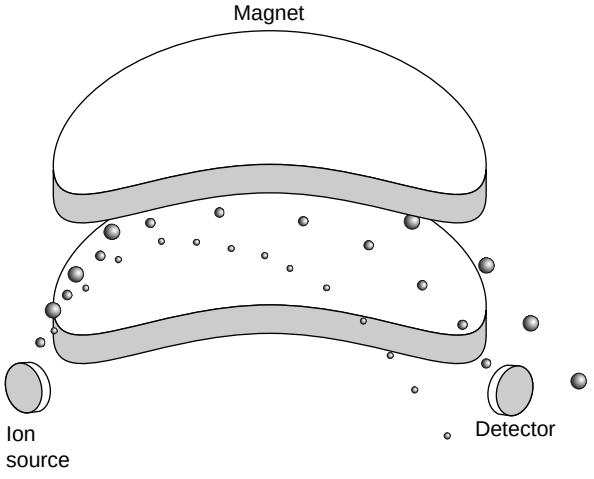

Figure 66 Magnetic sector mass spectrometer.[133] (Courtesy of Agilent Technologies.)

field. The mass analyzer of this type of instrument is a sector-shaped magnetic field that disperses the total ion beam from the ion source into discrete ion beams of individual m/z values through a process of momentum dispersion and direction focusing.

The m/ze of any ion traversing a fixed radius (fixed magnet geometry) can be related to the magnetic field and accelerating potential by Equation (22):

$$\frac{m}{ze} = r^2 \frac{B^2}{2V} \tag{22}$$

It is common practice not to deal with the absolute charge on the electron, and for convenience it is assumed to be unity. Thus Equation (22) is often found in the simplified form shown in Equation (23):

$$\frac{m}{z} = r^2 \frac{B^2}{2V} \tag{23}$$

The radius of curvature of an ion trajectory is proportional to its momentum (mv). Therefore, the total ion beam is dispersed into many individual ion beams, each having its characteristic radius of curvature in the region of the magnetic field. The imposed accelerating potential V and magnetic field B permit those ions of a specific mass m_1 to follow a trajectory to the detector, whereas the heavier ions of mass m_2 ($m_2 > m_1$) follow a trajectory that leads to collisions with the walls of the analyzer. Because the instrument has a fixed radius of curvature, either the magnetic field or the accelerating potential must be scanned to collect the mass spectrum.

Magnetic scanning is advantageous in that the entire mass range may be scanned while the instrument maintains optimum accelerating potential. A minor disadvantage is related to the quadratic dependence of m/z on the value of magnetic field B . As the mass spectrum is scanned, the peaks at the high mass end appear closer together than do those at the low-mass end.

The combination of magnetic field and the forward velocity of the ions exerts a lateral force on the ion, causing it to follow a curved path. The radius of the curve depends on the m/z ratio of the ion, the strength of the magnetic field, and the charge and velocity of the ion (the last two values depend on ion source behavior). For a given strength, only a single m/z ion will curve at the radius needed to reach the detector opening. All others will miss and be pumped away by the vacuum system.

Advances in GC continue to impose demands for faster scan rates on the mass spectrometer in GC/MS applications. The scan rate of a magnetic instrument is limited by the reluctance of the electromagnet. Laminated magnets permit more rapid changes in the magnetic field so that the mass range can be scanned at rates approaching 0.1 s per decade. Two major disadvantages of magnetic MS for field applications are the weight associated with the magnet and the potential effect of the magnetic field surrounding the instrument. If scan speeds exceed a certain rate, the magnetic field can change significantly during the ion transit period (the time required for an ion to traverse the magnet), causing ions to follow a distorted path to the detector. This results in losses in both resolution and sensitivity.

5.2.6 Cycloidal Mass Spectrometry

Cycloidal (trochoidal) MS is another emerging technology that has applications with residual gas analyzers. This technique could also potentially be used as an analytical tool, although no data are yet available to back this theory up. The technique is briefly discussed.

When a charged particle is moving in a plane that is perpendicular to a uniform magnetic field (Figure 67) that is in turn perpendicular to a uniform electrical field, the particle follows a cycloidal trajectory. The ions' path has a

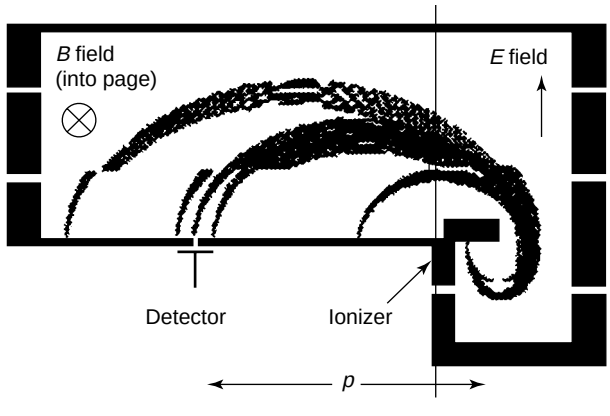

Figure 67 Basic description of a cycloidal mass spectrometer showing the trajectory of the ions inside the analyzer.

repetition distance, or cycloidal pitch, that is independent of ion energy and the angle of injection into the crossed fields. Given a fixed electrical field E and fixed magnetic field B, this pitch depends only on the mass-to-charge ratio of the ion. If p represents the pitch, this may be expressed by Equation (24):

$$P = 2\pi \frac{E(m/z)}{B^2} \tag{24}$$

If an ion source produces ions that are forced to enter such a crossed set of fields through a narrow slit, ions of a given mass-to-charge ratio move with a trajectory that comes back to the initial position after moving a horizontal distance equal to the pitch. Again, this is independent of the initial entrance angle of the ion into the field and the initial energy of the ion. The cycloidal mass spectrometer uses fixed electric fields and a collector array to collect a wide band of mass-selected ions simultaneously. The dimensions of the analyzer are potentially small (centimeters).

Small trochoidal mass spectrometers such as the MG2100, from Industrial Scientific Corporation (Oakdale, PA, USA) are suitable for the measurement of residual gas composition in ultrahigh vacuum systems and could potentially be used in applications for environmental analysis. A potential drawback of the FTMS technique is the need for powerful magnets and the interaction of high magnetic fields with instrumentation. There is no published information on the effect of a downgraded vacuum (due to the coupling of a gas chromatograph) on the performance of this instrument and it is, therefore, too early to speculate on potential GC/MS applications.

5.3 Portable Gas Chromatography/Mass Spectrometry Detectors

Over the past few years, GC/MS has emerged from the laboratory and into the field. In spite of its complexity and size limitations, the sheer analytical power of this hybrid separation and analysis technique has helped it earn its place in the field. A number of workers have demonstrated that transportable GC/MS systems can be moved to a field site and operated by bringing samples to the instrument either manually or via long, heated, analytical transfer lines. The next step is to develop field-portable equipment with analytical performances similar to those achieved with laboratory instrumentation.

There is a broad range of commercially available benchtop instruments, including quadrupoles and ion traps. Several specialized MS instruments operating in a mass range up to 1000 Da have been developed for applications for which transportability is a prime requirement. Well-known examples include the Bruker Franzen (Bremen, Germany) MEM and EM650 systems, originally developed for military applications involving chemical agent detection, and the Viking Spectra Trak system, primarily designed for environmental applications. All the instruments used for field analytical work require only unit resolution.

5.3.1 Portable Gas Chromatography/Mass Spectrometry Detector Based on a Quadrupole Mass Analyzer

Figure 68 shows a view of the Viking Spectra Trak 572 (Viking Instruments Corporation, Chantilly, VA, USA). This instrument is self-contained and requires only the power from an AC line. This transportable GC/MS is configured for GC injection with a capillary direct transfer line. The miniaturized GC oven accommodates standard capillary columns and provides temperature programming and a split or splitless injector. Other options are available and should be checked with Viking Instruments Corp. Figure 69 shows the overall instrument without a cover. The mass analyzer is a fixed-geometry monolithic quadrupole from Hewlett-Packard (Model HP5973A MSD). Figure 69 also shows the vacuum pumps, the GC oven, and the sample introduction hardware. The main instrument features are summarized in Table 15.

5.3.2 The Hapsite

The Hapsite instrument is designed to cover mass ranges from 1 to 300 Da and, therefore, is useful only for the analysis of VOCs. The mixture is first separated by a GC column before being analyzed by a quadrupole filter. One of the particularities of this instrument is that it is self-contained and operates on batteries. The instrument includes the vacuum pump, the carrier gas supply, an internal gas canister for calibration, rechargeable batteries, and all the electronics to control the instrument and acquire and transfer the data to a foreign site.

Figure 68 Spectra Trak 572 portable GC/MS instrument. (Courtesy of Viking Instrument Corp.)

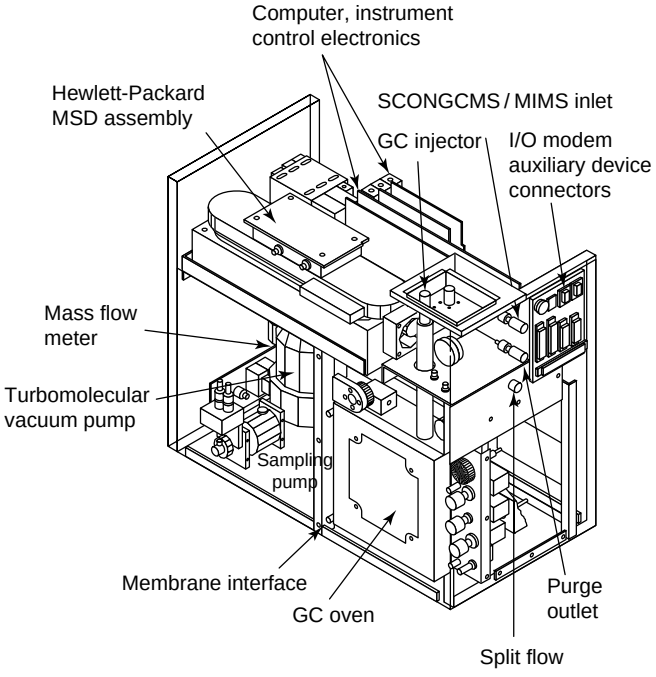

Figure 69 Cutaway view of Spectra Trak 572 chassis. (Courtesy of Viking Instruments Corp.)

The vacuum system is the part of the instrument that typically consumes the most power. This subsystem is also the heaviest portion of the instrument. The Hapsite portable GC/MS instrument manufactured by INFICON (East Syracuse, NY, USA) uses a commercial bulk getter pumping system. In this vacuum subsystem vanadium–iron–zirconium alloy sintered pellets are heated to 200 °C to pump chemically most gases and vapors. The lack of pumping capacity for noble gases represents a problem because of argon accumulation due to small air leaks in the vacuum system. A small secondary ion getter system is added to provide sufficient noble gas pumping. This instrument requires a separate service unit with an AC supply, vacuum pump, and carrier gas supply to prepare the portable unit before field operations. The specifications of the instrument are summarized in Table 16.

5.3.3 Time-of-flight Mass Analyzer

The performance of the TOFMS analyzer is a direct function of flight-tube length. The resolving power of a commercially available instrument with a 2-m flight tube is approximately 500. Various improvements have been developed to reduce the length of the instrument without limiting the flight path. Kore Technology has patented and developed a spectrometer geometry that starts with a ring-shaped ion source with a dimension of 20 cm (8 in) and a mass range of 1000 Da (Figure 70).[130]

Table 15 Specifications of the SpectraTrak 571 portable GC/MS system from Viking

Feature	Published specifications
Mass range	1.6–700.0 Da; mass stability ±0.15 Da over 12 h
MS tuning	Autotuning using self-contained PFTBA supply
Dynamic range	10^6; linear over four orders of magnitude
Scan rate	Up to 1800 Da s^{-1} at 10 points Da^{-1}
Sensitivity	Scan mode, 100 pg hexachlorobenzene; SIM mode, 1 pg hexachlorobenzene
SIM dwell time	Adjustable from 10 to 9999 ms for each selected m/z value
Ionization mode	70 eV EI
Vacuum system	70 L s^{-1} dual-stage turbomolecular vacuum pump with integrated molecular drag pump as second stage; a compact, lightweight, oil-free diaphragm pump provides rough pumping from atmospehric pressure
Size	46 cm (18 in) high × 61 cm (24 in) wide × 29 cm (11.5 in) deep
Weight	Approximately 34 kg (75 lb)
Internal power consumption	1200 W at cold start; typical continuous operation, 650 W, 110/220 V AC, 50/60 Hz
GC column	Capillary column up to 0.32 mm ID and up to 60 m long
GC oven	Maximum temperature 325 °C; multiple temperature ramps up to 25 °C min^{-1}
GC injector	Split/splitless with adjustable split ratio
External communications	Modem; communication software enables remote operation, system diagnostics, file transfer

PFTBA = perfluotributylamine

Table 16 Specifications of the portable Hapsite GC/MS instrument

Feature	Published specifications
Mass range	1–300 Da
Scan rate	1000 Da s^{-1} at 10 points Da^{-1}
Ionization mode	70 eV EI
Vacuum systems	15 L s^{-1} NEG pump
Operating conditions	0–45 °C
Size	46 cm (18 in) long × 43 cm (17 in) × 18 cm (7 in) high
Weight	Approximately 16 kg (35 lb) with batteries
Internal power consumption	24 V, 25 W at normal conditions
GC column	Supelco SPB1, 30 m × 0.32 mm ID × 1.0 μm film, isothermal operations
SIM channels	10
External communications	Ethernet

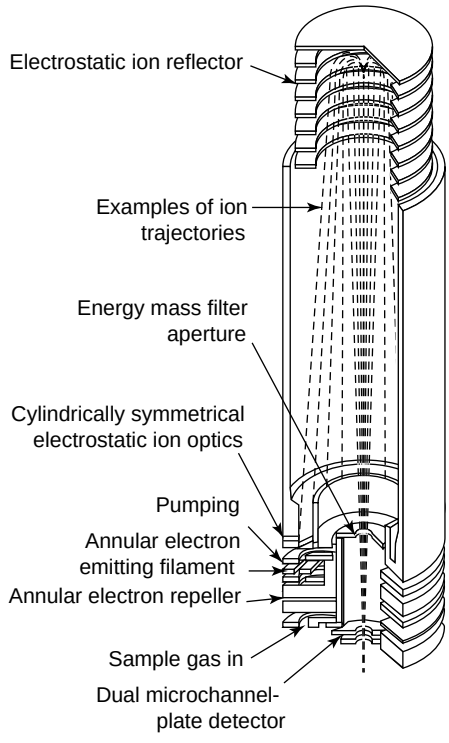

Figure 70 Cutaway view of a ring-shaped ion source. (Courtesy of Kore Technology, Ltd.)

The ions are extracted perpendicular to the ring and deflected inward so that, as they proceed through the spectrometer, each annular bunch of ions converges to a relatively small size by the time it reaches the detector. This allows a relatively large source to be used with a small detector giving parts per billion detection limits. In addition, an aperture can be placed at the crossover to allow mass or energy selection and enhance mass resolution. The specifications of the Kore TOF mass analyzer are given in Table 17.

Table 17 Specifications of the Kore TOF mass analyzer

Feature	Specifications
Mass analyzer	Converging annular time of flight
Ultimate sensitivity	No concentrator, 1 ppm; membrane module, 10 ppb (preliminary); trap and desorb, 1 ppb (preliminary)
Spectrometry source pressure	5×10^{-7} mbar (during analysis)
Mass resolution	0.2 Da full width at half-height maximum (FWHM) at 100 Da
Mass accuracy	Better than 0.04%
Mass range	1–1000 Da
Dynamic range	6 decades (dual acquisition)
Inlet option	Leak valve
Computer	IBM PC or compatible
Running time	Continuous analysis, 3 h; pumping only, 2.5 days
Main case and contents	Weight, 20.9 kg (46 lb) (including batteries); size, 53.1 cm × 32.8 cm × 21.3 cm (21 cm × 13 cm × 8 in)
Auxiliary case and contents	Weight, 11.8 kg; size, 45.2 cm × 32.5 cm × 12.9 cm

5.3.4 General Comments

Miniaturized spectrometry is a powerful analytical technique that has found wide use in laboratory and applied monitoring areas. Applications have included planetary orbiting missions (Mars *Viking* Lander GC/MS), earth orbiting satellites (*Atmospheric Explorer* satellites and others), manned spacecraft respiratory experiments (*Skylab*), submarine atmosphere monitors, and commercial, medical, and industrial process monitors. Research efforts in the late 1990s have demonstrated that miniaturized systems can be built with today's technology.

To miniaturize a mass spectrometer, some trade-offs must be accepted. For instance, it is generally accepted that sensitivity falls off rapidly with increasing mass range and decreasing size and that it is difficult to get it back by increasing gas load, implying a correspondent higher pumping speed and larger pump.

In addition to understanding the parametric relationships for each type of instrument, a practical knowledge of its performance is also essential. Some general comments regarding the various analyzers can be made. For instance, the ion trap is a proven sensitive analyzer with a dynamic range that is more limited than some other systems. The ion trap requires a collision gas to achieve good operation, which places a high gas load on the instrument, affecting the size and type of vacuum pump. The quadrupole mass spectrometer does not have these limitations and operates at lower pressure than the ion trap. It has a larger dynamic range, but is somehow less sensitive and is not as fast as the ion trap.

Although the TOF system may have appeal, there are duty cycle issues that must be understood before its net performance can be determined. TOF operation also puts a lot of burden on the electronics and vacuum system to operate at its optimum. The drift tube also must be folded to reduce the overall dimensions of the instrument. This design may degrade the performance of the instrument.

The ICR instrument must be operated in the Fourier transform mode, which requires highly sophisticated electronics to achieve superior performance. Like other magnetic sector instruments, ICR systems also have electronics problems due to the size of the magnet and to the effect of the magnetic field on surrounding instruments.

Magnetic sectors and trochoidal mass spectrometers also have a duty cycle limitation, but using an array

detector (focal plane detector) could, at least in theory, overcome this limitation. The overall mass range of the instrument could easily be monitored simultaneously. The major disadvantage for field applications is the weight of the magnet.

5.3.5 Example of Field Measurement by Gas Chromatography/Mass Spectrometry

A confidential site in the northeastern United States was characterized as having shallow soil PCB contamination.[136] Previous data suggested the presence of Aroclor 1248 and 1260 in the formerly earthen containment area. Field measurements performed with a portable GC/MS system from Viking Instruments confirmed the identities of PCBs (Figure 71a). Field GC/MS analysis also allowed easy resolution of the Aroclor mixtures detected in many samples, allowing for accurate PCB quantification (Figure 71b). The quality of the measurements is equivalent to or better than those performed in the laboratory, since the elimination of sample transportation reduces the risk of sample deterioration. Turnaround is typically hours instead of days, which can significantly reduce costs.

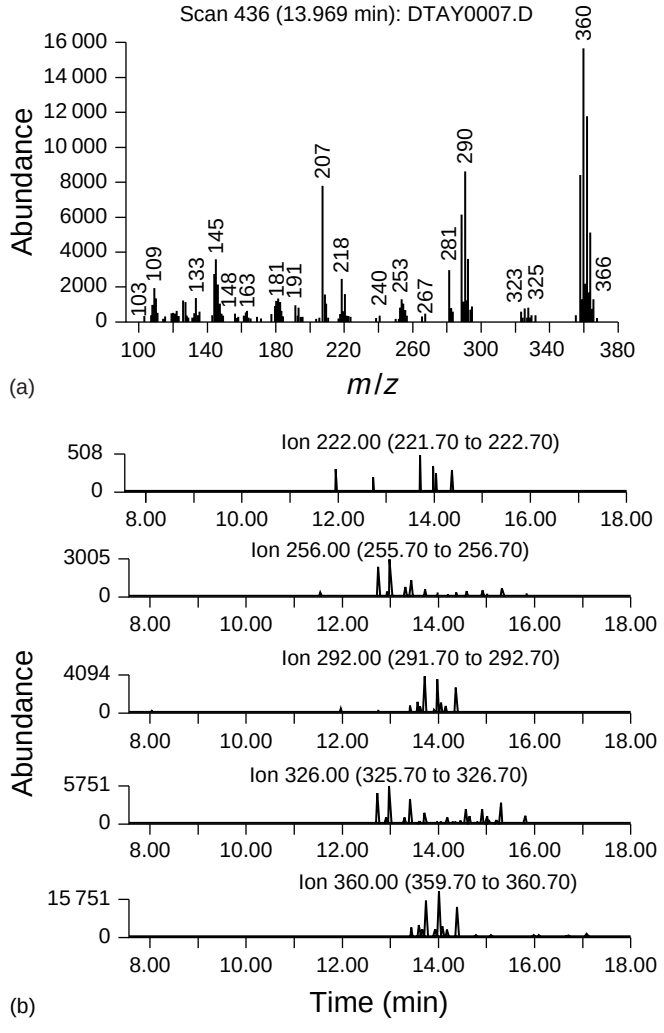

Figure 71 Field measurements by GC/MS. (a) On-site GC/MS determination of PCBs. (b) Extracted ion profiles of Aroclor mixtures, illustrating accurate PCB quantitation. (Courtesy of Viking Instruments Corp.)

5.4 Ion Mobility Spectrometry

IMS, like MS, falls in the category of mass dispersion methods. IMS was first described as an analytical method for trace analysis in 1974;[137] since then, its potential for field analysis has received some recognition, especially in the military, for the detection of chemical warfare agents such as nerve gases and blister agents. In addition to these applications, IMS has also been recognized as useful in a number of narrow technical niches for the chemical analysis of vapors. Several broad categories in the development of IMS have been identified:

- a selective sensor in which the chemistry of ionization can be tailored or manipulated for the detection of a specific compound or classes of compounds;
- a stand-alone analyzer for samples of predictable composition;
- a potential catalyst for improving the performance of GC for crude separation.

When the experimental conditions are suitable (low moisture and elevated temperature, ionization mechanism), all chemicals introduced into an IMS system can be ionized, which produces a spectrum of responses. In short, the conversion of molecules to ions at ambient pressure followed by ion characterization is nearly as comprehensive as MS.

The IMS detector has some unique advantages over GC detectors, including MSDs, but also has some limitations, the most serious being its limited dynamic range. In addition, IMS should not be compared with TOFMS. While certain functional similarities exist between TOFMS and IMS, comparisons can create unrealistic expectations for IMS performance. IMS peaks are typically rather broad compared with the range of possible drift times, and the amount of qualitative information available from IMS is less than that from MS and IR spectrometry but more than that from FID and ECD. On the positive side, IMS detectors have been engineered to be truly hand-portable stand-alone units, with a chemistry adapted to specific applications in order to enhance their sensitivity to some compounds. A brief introduction to IMS follows, with emphasis on field applications and equipment. An excellent source of information for this technology is available.[138]

Figure 72 Warfare application of IMS. The operator is using the CAM instrument for the decontamination of a military vehicle. (Courtesy of Graseby Ionics Ltd.)

5.4.1 Description

IMS has received a fresh examination in the last few years.[139–141] The interest in this technique is due to its intrinsic response features (excellent detection limits) when the experimental conditions are suitable (low moisture and elevated temperature) and to practical considerations (size, weight, and power advantages). Interest has also grown, following successful, proven application of IMS technology to significant challenges, including chemical warfare agent sensing, contraband screening, and toxic vapor monitoring.

Figure 72 shows a chemical agent monitor (CAM) unit developed by Graseby Ionics, Ltd. (Watford, Herts, UK). Its simplicity of design, atmospheric pressure operation, universality of response, and ion-resolving capability have enticed both scientists and instrument manufacturers to miniaturize and modify IMS detectors to provide analytical determination of targeted compounds in nonlaboratory environments.

The IMS technique is illustrated schematically in Figure 73.[142] The sample is introduced by means of a carrier gas into a reaction region, where it is ionized. Although positive and negative gas-phase ions for IMS have been produced by a variety of methods, including photoionization, laser multiphoton ionization, thermoionic emission, and corona spray, the most common ionization source is radioactive ^{63}Ni foil.[142] In the presence of sample molecules these ions undergo ion–molecule reactions to generate product ions. The nature of the product ions depends on the type of ion–molecule reaction that occurs, and early models for their formation can be found.[143,144] Once formed, product ions are extracted from the ion source region under the influence of an electrical field, and ions are drawn toward an ion shutter. The ion shutter is periodically pulsed, and ions are injected into the drift region.

In the drift region, the ions travel toward the collector still under the influence of the electric field imposed on the entire drift tube. Because ions with different cross-sections travel with different velocities, the ions are separated and arrive at the collector at different times. When ions of different types and mobilities exist in the drift tube, the ions can be separated through differences in their mobilities.

5.4.1.1 Drift Velocity The ion mobility is characteristic of a particular ion and can provide a means for detecting and identifying vapors. Ion mobilities are determined from ion velocities, which are measured in a

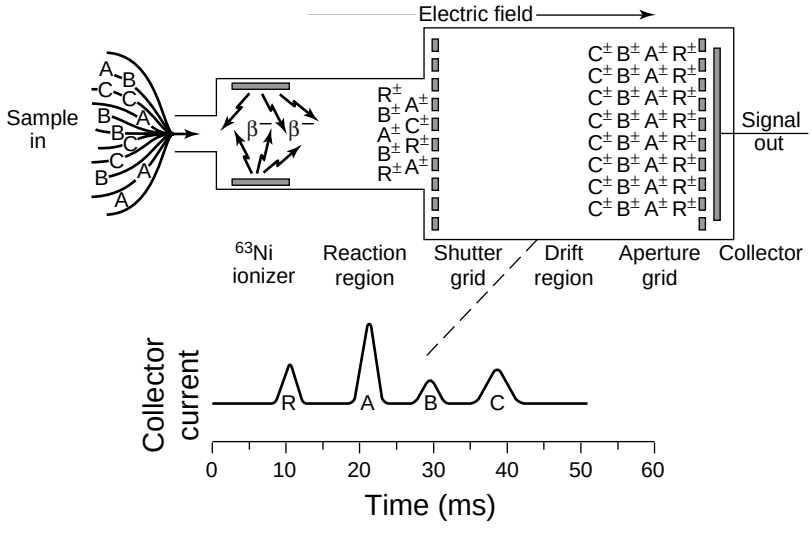

Figure 73 Basic principle of an IMS instrument. The various ions resulting from the ionization process are introduced into the drift region, where they will be separated according to their drift time.[143]

drift tube with supporting electronics. In practice, a vapor sample is introduced into the reaction region of a drift tube where neutral molecules of the vapor undergo ionization, and the resulting ions (product ions) are injected into the drift region for mobility analysis. Mobility K is determined from the drift velocity V_d attained by ions in a weak electric field E of the drift region at atmospheric pressure (Equation 25):

$$V_d = KE \tag{25}$$

Ion velocities are inversely dependent on the effective collisional cross-section of an ion, making IMS a kind of molecular size analyzer. Commonly, electrical fields of $200\,V\,cm^{-1}$ and mobilities of $1–2\,cm^2\,V^{-1}s^{-1}$ result in ion speeds of 200 to $1000\,cm\,s^{-1}$ or drift times of 5–20 ms in 4–20-cm long drift regions at ambient pressure. As the reactant ions and product ions are drawn toward and collide with the detector plate, current is registered and a signal is generated as a mobility spectrum (Figure 74a and b). When ions of different types and mobilities exist in the drift tube, the ions can be separated through differences in their mobilities (Figure 74b).

5.4.1.2 Ionization Process Vapors from nearly all chemical classes or functional groups can be ionized through atmospheric-pressure chemical ionization (APCI) processes. However, a given substance can be converted into different product ions, depending on the ionization mechanism. In addition, various product ions can be formed from a substance at a given concentration and temperature, but most of these are fragile clusters that do not survive transit even from the source region to the shutter. Hence in an IMS spectrum, only the stable ions survive and the bulk of all ions decompose, sometimes into stable ions. In practice, higher clusters are formed in the source and are rapidly depopulated to protonated monomers and proton-bound dimers in the drift region.

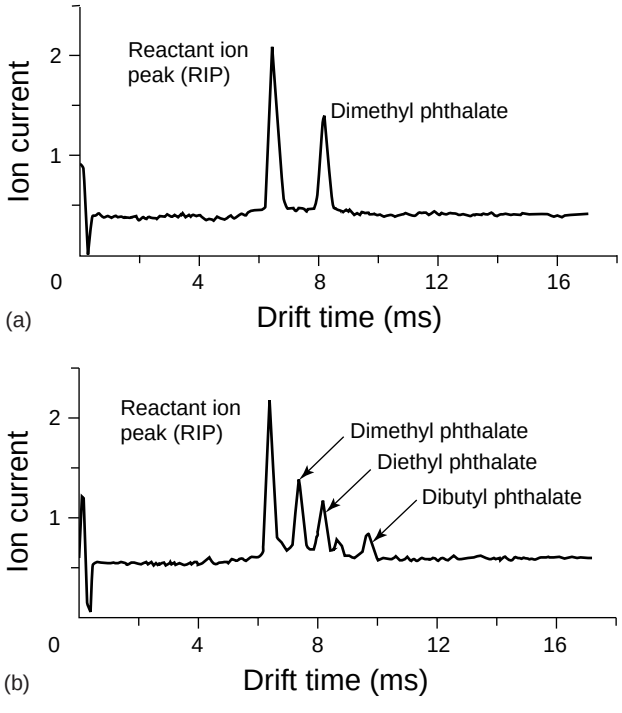

Figure 74 An ion mobility spectrum is a plot of detector response versus drift time. (a) A reactant ion peak (RIP) and a product ion peak are evident. (b) Productions are shown for two compounds.[139]

A fortuitous consequence of proton or electron transfers is that molecules form ions that retain much of the original shape and size of the molecules. In addition, product ions rarely dissociate or fragment, since the energetics of the APCI process are weak; mobility spectra thus often comprise only one ion or a few ions with a direct structural association to the original neutral molecule. The extensive fragmentation of molecules in EI ion sources for mass spectrometers does not occur as a general rule in IMS.

Ions are generated in a purified nitrogen or air carrier gas containing a trace of water vapor, by the action of β-particles emitted from a ^{63}Ni foil. The primary N_2^+ ion resulting from the ionization of nitrogen gas is too short-lived to appear in the IMS trace but begins a series of ion–molecule reactions, principally with nitrogen and trace amounts of water.

In the absence of a sample, the resulting stable ion clusters have been identified as $(H_2O)_nNH_4^+$, $(H_2O)_nNO^+$, and $(H_2O)_nH^+$.[145,146] These background ions are called reactant ions because they undergo further ion–molecule reactions with neutral gas-phase analytes to produce analyte product ions. The formation of product ions occurs in the ionization region predominantly by collisions between the reactant ions and the sample molecules. In addition, product ions may be formed through association reactions in which an adduct is made by a weak attachment of a reactant ion to a neutral vapor. A trace organic molecule injected into the carrier gas undergoes reactions with these ions and electrons, forming reaction product ions that show both quasimolecular and simple dissociated ions in their mobility spectra. This method is extremely sensitive and reasonably quantitative. It is capable of detecting 10^{-12} g or less of a compound and can provide identification through the characteristic positive and negative mobility spectra.

Positive and negative ions can be independently observed merely by the choice of the electrical field polarity. Positive product ions are formed mainly through a proton-transfer reaction. Early models for the formation of product ions of positive polarity have been summarized.[143] By monitoring positive reactant ions, one can gain essentially the same information from IMS that is obtained from an FID.[147]

The formation of negative ions in IMS is closely related to those chemical events that occur in an ECD. Early reviews of the ECD form the foundation for the current understanding of negative ion chemistry in IMS.[148–150] Two general observations have evolved: the data contained in positive mobility spectra bear a close similarity to those obtained in chemical ionization MS, and the data contained in negative mobility spectra correspond to those obtained with GC/ECD. Interpretation of the data based on this model is available.[137]

Ions formed in the reaction region are injected as a short pulse (200–300 μm) into a drift tube, where they acquire constant velocities under the influence of a weak electrical field and collisions with a counterflowing neutral drift gas (air or nitrogen). Various ions injected into the IMS drift region will arrive at the Faraday cup detector at different times, as determined by their mass, charge, and collisional cross-section with a drift gas. Consequently, different ion species, even if characterized by the same mass, may arrive at the detector at different times.[151]

An important variation of IMS ionization that can be used to enhance the sensitivity or selectivity of the technique for particular classes of compounds involves the modification of reactant ion populations by adding carefully controlled concentrations of dopants to the drift gas.[152] For example, Cl^- reactant ions have been used to increase sensitivity to explosives, NH_4^+ reactant ions enhance selectivity to amines, and $(CH_3COCH_3)_2H^+$ reactant ions are used to ionize organophosphonates selectively.[153,154]

In summary, IMS comprises and is governed by two distinct, independent or separate events: gas-phase ionization in air at atmospheric pressure through collisional charge exchanges or ion–molecule reactions and ion characterization using mobilities of gas-phase ions in a weak electrical field.

5.4.2 Instrumentation for Ion Mobility Spectrometry

It is convenient to divide the IMS system into three major sections: the sample introduction system, which is the interface of the instrument with the ambient atmosphere, the drift tube, where ionization and the separation and detection of the ions take place, and the detection device, where the signal is acquired, analyzed, and displayed.

5.4.2.1 Sample Introduction The use of a nonporous dimethylsilicone membrane barrier to the interface between the IMS system and the ambient atmosphere is the most common way to introduce samples into the IMS instrument, but in addition to membranes, other sample-to-drift tube connections have been made by GC inlets, electrospray, thermal desorption, and pyrolysis. The membrane retards the flux of water and ammonia molecules present in the sampled ambient air, thus limiting the number of polar molecules available for cluster formation. Excessive clustering reduces the resolution of IMS and may lead to erratic behavior of the ion signals. Membrane interfaces are especially important in a portable instrument when the drift tube operates near ambient temperature, as clustering presents a serious problem under this condition. The main advantage of employing a membrane interface is that it makes ambient-temperature operation possible and can lead to a considerable reduction in power consumption and thus in size and weight, which is required for handling instruments. The disadvantages of using a membrane are diminished sensitivity and an increase in the response time of the instrument. Other ways of increasing the selectivity and sensitivity of IMS rely on preseparation by GC and will be discussed later.

Reagent gases can also be selected as a means of increasing the sensitivity and specificity of IMS by preferentially ionizing the target compound or by enhancing the separation of the target ions from the reactant ions without affecting the response time of the instrument. For instance, adding a reagent gas that has a proton affinity just slightly below that of the target eliminates most of the interfering ion signals, and the presence of the target compound is easily detected. For example, monitoring isopropylamine (proton affinity = 218.6 kcal mol^{-1}) using ethylamine (proton affinity = 217.0 kcal mol^{-1}) as a reagent gas should exclude interferences from all compounds that have proton affinities below that of ethylamine.

5.4.2.2 The Drift Tube The first patented commercial IMS analyzers were created by Frankin-GNO, Inc.[155,156] The drift tube was made of circular elements biased at different electrical potentials to maintain a uniform electric field gradient of 200–300 V cm^{-1} along the central axis of the tube. These elements were alternating conducting and insulating guard rings.

Molecules were ionized by a polished β-emitting radioactive ^{63}Ni source inserted into the first guard ring. The shutter grid was of the Bradury–Nielson type, in which a potential difference is placed between a set of interdigitated thin wires with a voltage difference applied to each of the independent wires, thus creating a strong electrical field (of about 600 V cm^{-1}) perpendicular to the axis of the field gradient across the drift tube.[157] A collector plate, somewhat like a Faraday cup, was placed close to the collector to increase detection efficiency and to filter out the noise caused by the pulse on the gate, thus improving the signal-to-noise ratio.

The basic design has been modified in several ways.[158] The drift tube is now segmented with a unidirectional flow (Figure 75). With this modification, the sample is introduced into the reaction region just downfield of the

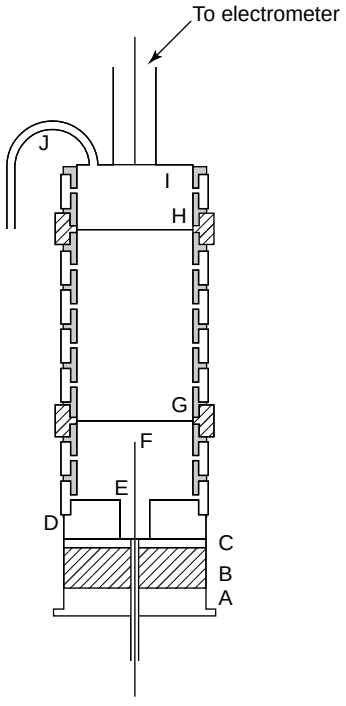

Figure 75 Schematic drawing of a drift tube.[158] A, Detector base; B, macro insulator; C, high-voltage supply; D, ^{63}Ni foil; E, capillary column; F, entrance gate; G, exit gate; H, collector; I, drift gas inlet; J, exit of the drift gas.

ionization source and is swept by the drift gas through the source. The unidirectional flow drastically reduces both memory effects and residence or response time and clarifies ionization chemistry, thus improving the performance of the cell drastically. The ions are moved down the length of the tube by the electrical field, while the neutral sample molecules that are not ionized spend only a short time in the reaction region before being swept out.

5.4.2.3 Detection Devices and Signal Processing The ion mobility spectrum is a plot of the signal intensity (ion current) as a function of the drift time, taking zero as the time the shutter-grid gate is opened to allow ions to enter the drift region. The most common and simple detection method used to measure the ion current in the IMS drift tube is a collector plate, somewhat like a Faraday cup. The output signal from the collector is, therefore, an analog signal that is digitized and displayed as a digital signal.

There are several basic modes of operation: signal averaging (up to hundreds of spectra are averaged to increase the signal-to-noise ratio), hardware monitoring, and Fourier transform.[159] Another method of registering a signal is through the use of two shutters with boxcar integration. In this technique, a ion shutter near the source region is operated at a fixed frequency of approximately 30 Hz for a conventionally sized drift tube, and the shutter pulse width may be 100–300 μs. After the first shutter is opened, a delay is measured and the second shutter, near the detector, is opened for a short interval and then closed. Only ions with a drift time corresponding to the delay are allowed to pass to the detector. This delay is gradually increased to a maximum, and the spectrum is generated during the sweep of the delay time.

5.4.3 Quantitative Limitations of Ion Mobility Spectrometry

IMS has been shown to be a selective and extremely sensitive detection technique for monitoring vapors or air pollutants. However, systems with membrane-type inlets, such as the CAM, are subject to serious limitations with regard to both qualitative and quantitative performance. Moreover, an ion mobility spectrometer, as with any secondary ionization detector, exhibits a nonlinear response with respect to the concentration of analytes introduced into the ionization chamber. Such a response increases exponentially up to the depletion of the reactant ions that is observed as the disappearance of the RIP peak from IMS trace.

As shown in Figure 76, standard IMS systems are easily overloaded with analyte and have a narrow linear dynamic range.[160] Each IMS trace in Figure 76 was obtained at a different concentration of dimethyl methylphosphonate (DMMP) with a 4-s sampling time in the positive ion mode. DMMP vapor concentrations of 20 and 60 ppb produce only protonated monomer peaks. However, proton-bound dimer and trimer signals for DMMP appear above 130 ppb. Furthermore, the intensities of the protonated monomer and the proton-bound dimer are proportional to the DMMP concentration between 20 and 280 ppb and between 130 and 1700 ppb, respectively. The intensity of the trimer shows a narrow linear response range. In other words, high sample concentrations produce cluster ions that not only reduce the useful linear range but also produce more complex IMS traces.

5.4.4 Hand-held Ion Mobility Spectrometry Instruments

The widest application of hand-held IMS instruments is the detection of chemical warfare agents (Table 18). The CAM made by Graseby Ionics, Ltd. and manufactured under license in the USA by Environmental Technologies Group, Inc. (Baltimore, MD, USA), is slightly larger than a cordless telephone. Weighing less than 1.8 kg, with 38 cm as its longest dimension, this CAM represents an exceptional work of engineering.[139] The drift tube, sampling and flow systems electronics, signal processing, and a simple bar display have been compressed into a

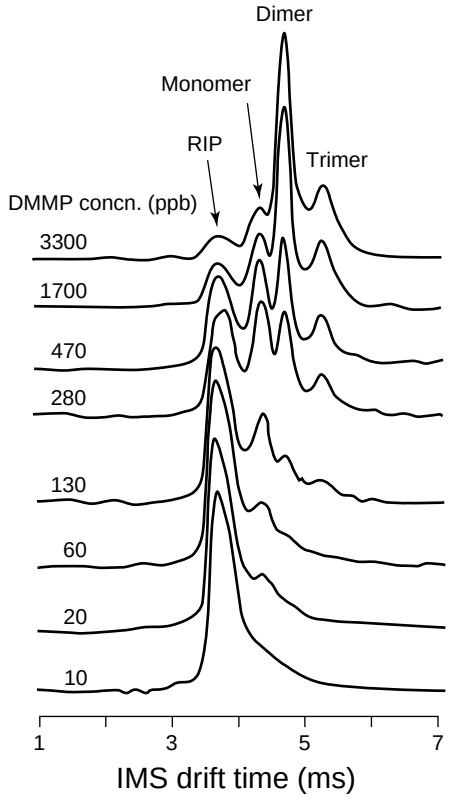

Figure 76 Variations in IMS spectra due to changes in the DMMP sample concentration.

Table 18 Major parameters of the military IMS unit commercialized by Graseby Ionics Ltd.

Feature	Specifications
Detection	Nerve and blister agents
Sensitivity	NATO requirements
Power supply	Single 6-V battery (sealed Li/SO_2 system)
Battery life	Minimum 6 h continuously; typically 10 h normal use at 20 °C
Environment	DEFSTAN 00-1 cat. A2 to C1
Temperature range	−30 to 55 °C operating (−55 to 70 °C storage)
Dimensions	38 cm (15.2 in) longest dimension
Weight	1.5 kg (3.3 lb)

hand-held module. A civilian spin-off of the CAM, called the airborne vapor monitor (AVM), is also produced by Graseby Ionics, Ltd. A recent development in hand-held IMS instruments is the inclusion of a GC column as a preconcentrator that is interfaced with the AVM commercialized by FemtoScan (Figure 77).

The production of commercial applications is, to a large extent, a spin-off of the military chemical agent detector and monitor program. When discussing industrial and environmental applications of IMS, it is convenient to

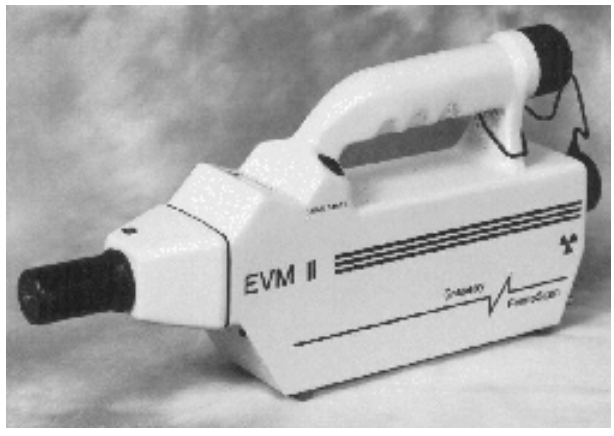

Figure 77 Civilian GC/IMS hand-held unit, Model EVM II. This environmental vapor monitor is used for the detection of mixtures of toxic vapors and gases. (Courtesy of FemtoScan Corp.)

differentiate between fixed-point monitors and hand-held instruments. The former are used to detect, measure, and monitor a specific compound or a number of chemicals at a fixed location on a continuous basis. The hand-held instruments are generally used as monitors for a short duration at a given spot and are easily transportable from point to point. Hence the fixed-point monitors are useful for obtaining the concentration versus time profile at a given point, whereas the hand-held instruments are useful for obtaining concentration versus location profiles. Hand-held instruments are ideally suited for detecting leaks or hot spots at which the chemical of interest has accumulated.

The major environmental and industrial applications of IMS instruments include monitoring toxic chemicals, stack gases, and chemicals that are considered hazardous to humans or to the environment. To date, few detailed accounts of the performance of IMS have been made public. Applications of IMS can be found in reviews[161–163] (Table 19). A comprehensive list of chemicals that can be monitored by IMS is available;[164] it includes compounds such as toluidine, pyridine, diaminotoluene, dimethyl sulfoxide, glutaraldehyde, methyl salicylate, acetic acid, acetic anhydride, halothane, enfluarane, isoflurane, and nicotine.

5.4.5 Partial Review of Published Ion Mobility Spectrometry Applications

The use of IMS for field applications has been described in a number of symposia related to waste monitoring. Some key environmental and industrial applications of IMS have been reviewed.[140] Nickel carbonyl, which is highly toxic with a TLV of 50 ppb, has been targeted.[165] Another application of IMS is monitoring the toxic

Table 19 Compounds of environmental interest suitable for IMS with typical detectable limits[a]

	Compound	Minimum detectable limit (ppm)	Compound	Minimum detectable limit (ppm)
Acid and stack vapors	HCN	0.1	HCl	0.1
	NH_3	0.1	NO_2	0.1
	HI	0.1	HF	0.1
	Phosgene ($COCl_2$)	0.1	Nitric acid (HNO_3)	0.1
Other chemicals	Acetaldehyde (CH_3CHO)	0.1	Aliphatic amines	0.005
	Aromatic amines	0.005	Aniline	0.005
	Methylenedianiline	0.01	Chlorine (Cl_2)	0.1
	TDI	0.005	Nitrobenzene	0.005
	1,6-Diisocyanatohexane	0.1		
	PCBs	0.4	Vinyl acetate	0.1
Solvents	Furans	0.1	Alcohols	0.1
	Cyclohexanone	0.01	Acrylonitrile	0.01
	Phenols	0.1	Toluidine	0.1
	Chlorinated solvents	1	Acetonitrile	0.01
	Ketones	0.01	Esters	0.01
	Ethers	0.1		
Semiconductor gases	PCl_3	0.1	NF_3	0.1
	Dry etchant gas	0.1		

[a] From Environmental Technologies Group, Inc.

chemicals hydrazine and monomethylhydrazine (MMH) at levels below TLVs above 10 ppb.[166] An advanced environmental application for commercial IMS is the monitoring of hydrogen fluoride at parts per million levels.[167] Another application is the monitoring of highly toxic (TLV 5 ppb) TDI.[168] Unlike industrial and environmental applications, the use of IMS-based technology to detect contraband explosives or drugs does not require quantitative measurements. Because of the high sensitivity of IMS, its specificity, and the fact that it operates at atmospheric pressure, it is apparently a leading candidate for detection of hidden contraband materials.[169] The major problem is getting enough molecules of the target compound into the detector. This problem is most severe with chemicals that have low vapor pressures.

IMS is currently being successfully applied to the problem of the on-line trace detection of plastic and other explosives in airports and other facilities.[170] The methods of sample retrieval consist primarily of batch sampling for particle residue on a filter card for introduction into the IMS system. The sample is desorbed using air as the carrier into the IMS instrument, where negative ions of the explosives can be detected, some as an adduct with a reagent ion such as Cl^-. Based on studies and tests conducted by different airport authorities, this method seems to work well for low vapor pressure explosives, such as trinitrotoluene (TNT). These are highly adsorptive and can be found in nanogram quantities on contaminated surfaces.

An ion mobility spectrometer is also capable of detecting minute amounts of narcotics with few false alarms and a minimum of operator intervention. Sampling rates of 10 suitcases or passengers per minute have been reported.[171]

5.4.6 Recent Developments in Gas Chromatography/Ion Mobility Spectrometry Techniques

In GC, constituents of a mixture are physically separated by the column and identified from their elution time. With the exception of the mass spectrometer, GC detectors do not provide any information about the chemical identities of the compounds. The most important feature that GC brings to IMS is the individual delivery of compounds to the IMS system, thus reducing the problem of interferences and providing near-ideal ionization conditions. The GC part of the system is usually unsophisticated compared with a laboratory gas chromatograph, and provides only crude separation of the components.

The hand-held IMS instrument shown schematically in Figure 78 is a combination of a successful military IMS system with an inlet developed for sampling ambient atmospheric vapors.[172] The addition of a short capillary column between a specialized valveless inlet and the ion mobility spectrometer allows for the separation of vapors in less than a few seconds.[173]

The pumps, used originally in hand-held IMS instruments to control drift gas and sampling flows, are used to lower the pressure in the drift tube. This results in a pressure drop across the column, and the air is drawn through the column as the carrier gas. The configurations shown in Figure 78 permit only isothermal column temperatures and constitute a prototype hand-held GC/IMS system.

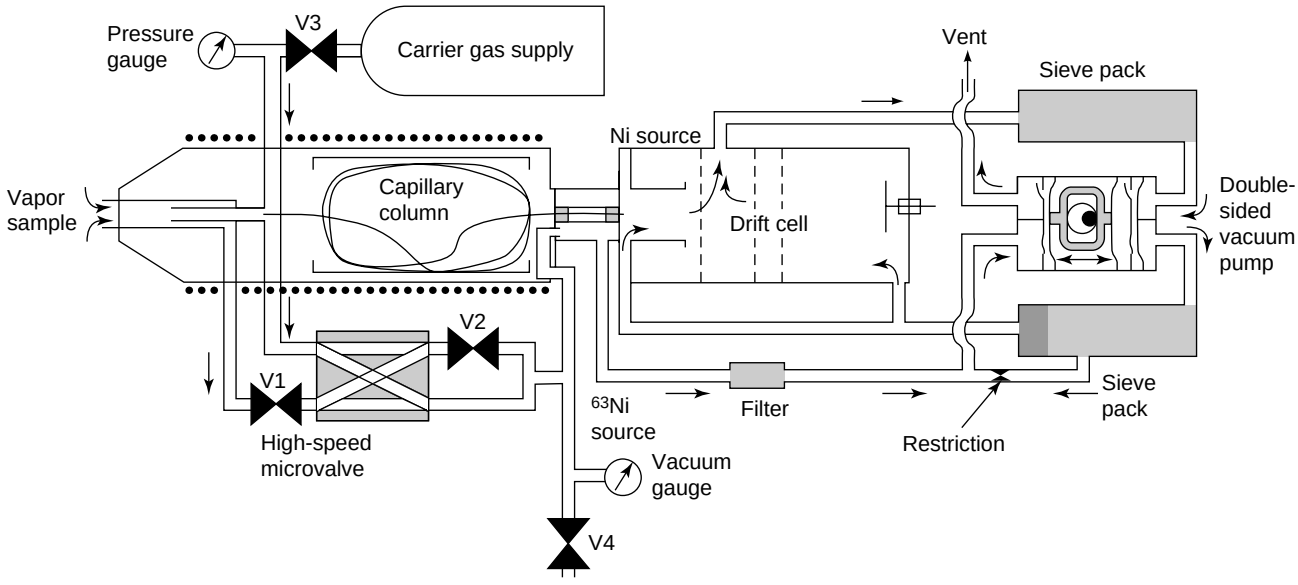

Figure 78 Schematic diagram of a hand-held GC/IMS instrument.[160]

The ability to produce an effective hand-portable instrument depends on achieving several goals: limited sample preparation and pre-separation (short chromatographic column, for instance), lightweight design (simplified design, use of air instead of nitrogen), and high-speed response (just enough chromatography, no baseline separation). The efficiency of the GC column in this example is only 2000–5000 theoretical plates, yet it is suitable for delivering the phosphonate esters individually to the ion source and reducing interference problems.

Figure 79 illustrates some advantages of operating an IMS system in the AVS transfer line gas chromatography (TLGC) mode, compared with the membrane inlet for air analysis of an equimolar mixture of model alkylphosphonates.[160] The IMS trace of the ternary mixture of alkylphosphonates without GC pre-separation (i.e. in the case of a membrane-type inlet) reveals several mixed-ion peaks [e.g. DMMP + diethyl methylphosphonate (DEMP) and DEMP + diisopropyl methylphosphonate (DIMP), etc.] in addition to protonated monomer, proton-bound dimer and trimer formation by particular components (Figure 79, trace A). The IMS trace of the same mixture is presented as trace B, but it was obtained by an average of 32 scans recorded after AVS injection into a 2-m long capillary column. Pre-separation by TLGC introduces each component separately into the ion source. Therefore, the relative intensities of particular component signals reflect the composition of the original mixture, since no mixed-ion peaks or competitive ionization processes are observed. All individual IMS traces recorded after sampling the analyte mixture are presented in Figure 79, traces C, in the form of a pseudo-three-dimensional waterfall graph. Furthermore, a reconstructed transfer line gas chromatogram of this ternary mixture, obtained by summing all ion current intensities at drift times between 5 and 7 ms, is shown in Figure 79 as trace D. It is clear that a 2-m long capillary column provides sufficient resolving power to produce just enough separation of the ternary mixture components within 30 s, simplifying ionization processes and dramatically improving qualitative and quantitative capabilities of the CAM for analysis of mixtures.

The use of short, fast separation columns reduces many problems associated with analysis of complex mixtures. However, under real-world conditions, co-elution of compounds is inevitable in IMS and can complicate the interpretation of spectra. Therefore, part of the optimization should include the chromatographic separation.

5.4.7 Recent Progress in Ion Mobility Spectrometry Technology

The limited linear response of IMS to one or two orders of magnitude is one of the limitations to using IMS as a GC detector. To extend this dynamic range there has been some attempt to operate the IMS inlet on a dynamic dilution sampler.[174] In this concept, the sample intake is pulsed with a fast valve (down to 20 μs) into the inlet against a flow of clean air or nitrogen. The amount of sample entering the inlet is governed by the pulse rate or the pulse width of the valve. In an effective and practical inlet, either of these would be controlled by the RIP intensity from the mobility spectrum. The valve is fully

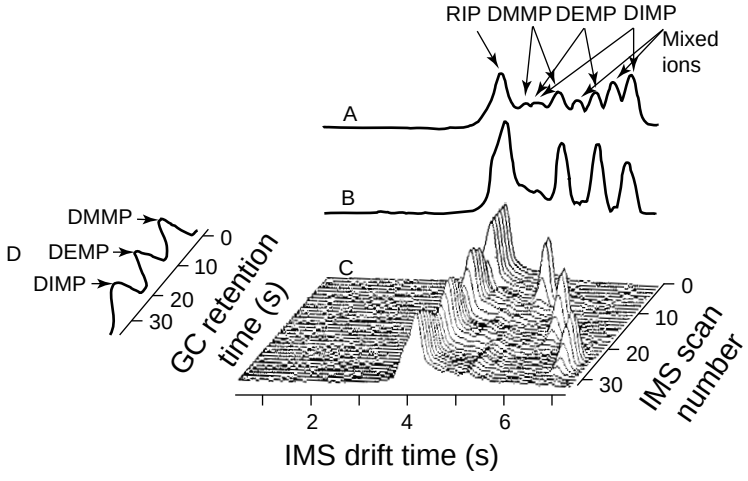

Figure 79 Comparison of a direct and an AVS GC/IMS inlet mode. A, direct inlet IMS trace of three methylphosphonates (without TLGC pre-separation). B, sum of the 32-ion mobility scans from the GC/IMS traces of the methylphosphonates shown in C; C, quasi-three-dimensional waterfall display of a GC/IMS profile. D, chromatogram of the GC run shown in C.[160]

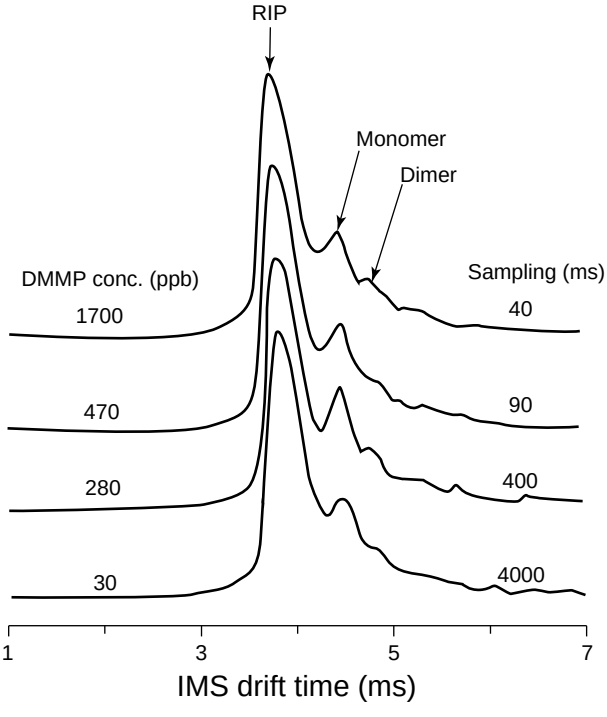

Figure 80 Variations in IMS traces with DMMP sample concentration. Compare with Figure 76. The IMS overload is controlled by varying the sample size as a function of AVS pulse length (AAC approach).[160]

open until the RIP intensity falls below a preset value, which then triggers pulsing of the valve.

Figure 80 illustrates the application of the servo-inlet concept. The IMS traces were been obtained for various concentrations of DMMP in air and at different sampling pulse durations [called automatic attenuation control (AAC)]. These spectra should be compared with those in Figure 76, where the sample was injected without dilution. In Figure 80, only the protonated monomer is the major peak for DMMP vapor concentrations up to 1700 ppb. This is achieved by specifying the depletion level of the RIP peak (selected in this application as 50% of its background level), and a sampling time of the AVS that should be shorter than the intrinsic broadening caused by the GC/IMS system. Based on these spectra, the dynamic range of IMS is extended over four decades.

5.4.8 General Comments About Industrial Applications of Ion Mobility Spectrometry

IMS is a sensitive detection method with a broad response that has found favor in certain applications, owing to the advantages of some IMS drift tube configurations. The industrial products based on IMS techniques are still limited, probably because the use of these devices requires a meticulous understanding of the measurement process. Both the target chemical and the matrix must be evaluated when considering IMS for an application. IMS limitations and possible solutions to some of its identified problems have been published[163] and are summarized below:

- The chemical compound to be detected must be in the vapor state and must be ionized in the reaction tube of the IMS system. Some materials, including small hydrocarbons, are not easily ionized with a radioactive source when using water-reactant ions, but may be seen using other ionization techniques, such as charge exchange ionization.

- The IMS response varies with the gas pressure in the drift tube. Small atmospheric pressure variations can disrupt measurements and require frequent calibration when IMS is used at high sensitivity.
- Ion clouds are formed by stealing charge from the reactant ion, which may significantly change the concentration. This problem is usually corrected by doping the carrier gas stream with a compound that has a proton or electron affinity slightly below that of the compound to be analyzed.
- IMS peaks, and particularly the RIP, change position with variation in humidity because of differences in water clustering. In a portable IMS this can be controlled by using a recycling loop of air that is continuously dried by scrubbers.
- The limited dynamic range of one to two orders of magnitude is a serious limitation. In addition, the signature of a compound varies with the concentration. Working at a constant concentration (dynamic dilution) close to that used for calibration is an attractive but not yet practical approach to achieving quantitative results over a broad dynamic range.
- Water and ammonia, even at low concentrations, can create a change in the performance of the IMS system, which can take a long time to recover.

Although IMS was begun over 25 years ago as an analytical technique, the technology is still in a nascent stage of development. The current implementations of IMS have shown technology ''fences'' rather than fundamental limitations. For example, high-resolution IMS drift tubes of a small size have been demonstrated in Germany and, recently, the injection pulse width has been recognized as critical, opening up a broad area of research and potentially new products. In conclusion, significant technical improvements can be expected in the near future and, as the price of IMS instrumentation decreases, a broader use of this technology for field industrial applications will also be seen.

ABBREVIATIONS AND ACRONYMS

AAC	Automatic Attenuation Control
AC	Alternating Current
ACGIH	American Conference of Governmental Industrial Hygienists
AED	Atomic Emission Detector
APCI	Atmospheric Pressure Chemical Ionization
AVM	Airborne Vapor Monitor
AVS	Ambient Vapor Sampling
CAM	Chemical Agent Monitor
CGI	Combustible Gas Indicator
CMS	Chemical Measuring System
COE	Army Corps of Engineers
DC	Direct Current
DEMP	Diethyl Methylphosphonate
DIMP	Diisopropyl Methylphosphonate
DMMP	Dimethyl Methylphosphonate
DOE	Department of Energy
DOI	Department of the Interior
DOL	Department of Labor
ECD	Electron Capture Detector
EI	Electron Ionization
ELCD	Electrolytic Conductivity Detector
FID	Flame Ionization Detector
FIDGS	Flame Ionization Detector for General Survey
FPD	Flame Photometric Detector
FSWCOT	Fused-silica Wall-coated Open-tubular
FTMS	Fourier Transform Mass Spectrometry
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HETP	Height Equivalent to a Theoretical Plate
HSGC	High-speed Gas Chromatography
ICR	Ion Cyclotron Resonance
ID	Inner Diameter
IMS	Ion Mobility Spectrometry
IR	Infrared
IRD	Infrared Detector
ISFET	Ion-sensitive Field Effect Transistor
ITD	Ion Trap Detector
LEL	Lower Explosive Level
LLNL	Lawrence Livermore National Laboratory
MDQ	Minimum Detectable Quantity
MMH	Monomethylhydrazine
MOS	Metal-oxide Semiconductor
MS	Mass Spectrometry
MSD	Mass Spectrometric Detector
NAAQS	National Ambient Air Quality Standards
NIOSH	National Institute of Occupational Safety and Health
NPD	Nitrogen–Phosphorus Detector
OSHA	Occupational Safety and Health Administration
PCB	Polychlorinated Biphenyl
PEL	Permissible Exposure Limit
PID	Photoionization Detector
PTFE	Polytetrafluoroethylene
PTV	Programmed Temperature Vaporizer
PZX	Piezoelectric Crystal
RE	Reference Electrode
RF	Radiofrequency
RIP	Reactant Ion Peak

SAAQS	State Ambient Air Quality Standards
SAW	Surface Acoustic Wave
SE	Sensing Electrode
SHE	Standard Hydrogen Electrode
SIM	Selected-ion Monitoring
SVE	Solvent Vapor Extraction
SVOC	Semivolatile Organic Compound
TCD	Thermal Conductivity Detector
TDI	Toluene Diisocyanate
TLGC	Transfer Line Gas Chromatography
TLV	Threshold Limit Value
TNT	Trinitrotoluene
TOF	Time-of-flight
TOFMS	Time-of-flight Mass Spectrometry
TWA	Time-weighted Average
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
VOC	Volatile Organic Compound
WCOT	Wall-coated Open-tubular

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Automotive Emissions Analysis with Spectroscopic Techniques • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Chemical-sensing Networks: Satellite-based • Cone-penetrometer-deployed Samplers and Chemical Sensors • Electrochemical Sensors for Field Measurements of Gases and Vapors • Field-based Analysis of Organic Vapors in Air

Field-portable Instrumentation cont'd **(Volume 5)**
Microelectromechanical Systems Technology Applied to the Miniaturization of Field Instrumentation • Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements • Radon, Indoor and Remote Measurement of • Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

Food **(Volume 5)**
Sample Preparation Analytical Techniques for Food

Forensic Science **(Volume 5)**
Ion Mobility Spectrometry in Forensic Science • Mass Spectrometry for Forensic Applications

Industrial Hygiene **(Volume 5)**
Industrial Hygiene: Introduction • Aerosols and Particulates Analysis: Indoor Air • Carcinogens, Monitoring of Indoor Air • Chromatographic Techniques in Industrial Hygiene • Direct Reading Instruments for the Determination of Aerosols and Particulates • Dust, Measurement of Trace Elements in

Industrial Hygiene cont'd **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Process Instrumental Methods **(Volume 9)**
Chromatography in Process Analysis • Mass Spectrometry in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Sampling and Sample Preparation in Process Analysis

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Elevation Modeling and Displacement Mapping using Radar Interferometry • Elevation Modeling from Satellite Data • Satellite and Sensor Systems for Environmental Monitoring

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Gas Chromatography/Mass Spectrometry • Time-of-flight Mass Spectrometry

REFERENCES

1. R. Erbes, *A Practical Guide to Air Quality Compliance*, 2nd edition, John Wiley & Sons, New York, 1996.
2. N. Stoloff, *Regulating the Environment: an Overview of Federal Environmental Laws*, Oceana, Dobbs Ferry, NY, 1991.
3. M. Katz, 'Advances in Analysis of Air Contaminants. A Critical Review', *J. Air Pollut. Control Assoc.*, **30**, 528–557 (1980).
4. National Institute for Occupational Safety and Health, *NIOSH Manual of Analytical Methods*, 2nd edition, DHEW (NIOSH) Pub. No. 75–121, 1975; 3rd edition, DHEW (NIOSH) Pub. No. 84–100, 1984 (revised 1990); 4th edition, DHHS (NIOSH) Pub. No. 94–113, 1994.
5. Occupational Safety and Health Administration, *OSHA Analytical Methods Manual*, OSHA Analytical Laboratories, Salt Lake City, UT, 1985; Suppl., 1991, 1993.
6. Health and Safety Executive, *Methods for the Determination of Hazardous Substances*, HSE Occupational Medicine and Hygiene Laboratory, Sheffield, 1981–1993.
7. Intersociety Commitee's Methods for Air Sampling and Analysis, *Methods of Air Sampling and Analysis*, 3rd edition, Lewis, Chelsea, MI, 1988.
8. S. Ness, *Air Monitoring for Toxic Exposures*, Van Nostrand Reinhold, New York, 1991.
9. USEPA, *Compendium Method T0141: The Determination of Volatile Organic Compounds (VOCs) in Ambient Air Using Summa® Passivated Canister Sampling and Gas Chromatographic Analysis*, USEPA, Washington, DC, 1988.
10. L. Pagnotto, *Air Sampling Instruments*, 6th edition, American Conference of Governmental Industrial Hygienists, Cincinnati, OH, 1983.
11. *Fundamentals of Analytical Procedures in IH*, AIHA, Akron, OH, 363–373, 1987.
12. S. Crisp, *Ann. Occup. Hyg.*, **23**, 47–76 (1980).
13. W.R. Betz, S.G. Maroldo, G.D. Wachob, 'Characterization of Carbon Molecular Sieves and Activated Charcoal for Use in Airborne Contaminant Sampling', *Am. Ind. Hyg. Assoc. J.*, **50**(4), 181–187 (1989).
14. P.M. Eller (ed.), National Institute for Occupational Safety and Health, *NIOSH Manual of Analytical Methods*, NIOSH, Cincinnati, OH, 1984.
15. R.H. Brown, C.J. Purnell, 'Collection and Analysis of Trace Organic Vapor Pollutants in Ambient Atmospheres. The Performance of a Tenax-GC Adsorbent Tube', *J. Chromatogr.*, **178**, 79–90 (1979).
16. R.R.A. Miksch, 'Personal Monitoring Employing Knudsen Diffusion: Development and Application to the Measurement of Formaldehyde in Indoor Air', Air Technology Laboratories, Fresno, CA.
17. M.L. Woebkenberg, C.S. McCammon, *Air Sampling Instruments: Direct-reading Gas and Vapor Instruments*, 8th edition, ACGIH, Cincinnati, OH, 430–510, 1995.
18. E. Kress-Rogers (ed.), *Handbook of Biosensors and Electronic Noses*, CRC Press, New York, 1997.
19. P. Moseley, J. Morris, D.E. Williams (eds.), *Techniques and Mechanisms in Gas Sensing*, Adam Hilger, New York, 1991.
20. S. Yamauchi (ed.), *Chemical Sensor Technology*, Elsevier, New York, Vol. 4, 1992.
21. H. Strobel, W. Heineman, *Chemical Instrumentation: Systematic Approach to Instrumental Analysis*, 3rd edition, John Wiley & Sons, New York, Chapt. 27–32, 1989.
22. J. Wang, *Analytical Electrochemistry*, VCH Publishers, New York, 1994.
23. A. Foust, *Principles of Unit Operations*, John Wiley & Sons, New York, 1960.
24. A. Tantram, Y. Chan, 'Gas Sensor [Electrochemical]', City Technology Ltd., *US Pat.*, 4 324 632, 1982.
25. A. Tantram, A. Tseung, 'Structure and Performance of Hydrophobic Gas Electrodes', *Nature (London)*, **221**, 167–168 (1969).
26. J. Giner, C. Hunter, 'Mechanism of Operation of the Teflon-bonded Gas Diffusion Electrode: Mathematical Model', *J. Electrochem. Soc.*, **116**, 1124–1130 (1969).
27. B.C.H. Steele, J. Drenna, R.K. Slotwinski, N. Bonanos, P. Butler, 'Factors Influencing the Performance of Zirconia-based Oxygen Monitors', *Adv. Ceram.*, **3**, 286–309 (1981).
28. A. Tantram, J.M. Kent, A.G. Palmer, 'Utility Gas Sensor', City Technology Ltd., *US Pat.*, 4 132 616, 1979.
29. A. Tantram, Y. Chan, 'Utility Gas Sensor [Electrochemical Carbon Monoxide Detector]', City Technology Ltd., *US Pat.*, 4 633 704, 1987.
30. W. Mokwa, in *Chemical Sensor Technology*, ed. S. Yamauchi, Chemical Sensor Technology Series, Elsevier, New York, Vol. 4, 1992.
31. I. Lundström, S. Shivaraman, C. Svensson, L. Lundkvist, 'A Hydrogen-sensitive MOS Field-effect Transistor', *Appl. Phys. Lett.*, **26**, 55–57 (1975).
32. J. Janata, 'Chemical Selectivity of Field-effect Transistors', *Sens. Actuators*, **12**, 121–128 (1987).
33. K.I. Lundström, M.S. Shirvaman, C.M. Svensson, 'A Hydrogen-sensitive PD-gate MOS Transistor', *J. Appl. Phys.*, **46**, 3876–3881 (1975).

34. T. Fare, A. Spetz, M. Armgarth, I. Lundström, 'Quasi-static and High-frequency C(V)-response of thin Platinum Metal-oxide Silicon Structures to Ammonia', *Sens. Actuators*, **14**, 369–386 (1988).
35. F. Winquist, I. Lundström, 'Thin Metal-film Oxide Semiconductor Structures with Temperature-dependent Sensitivity for Unsaturated-hydrocarbons', *Sens. Actuators*, **12**, 255–261 (1987).
36. T.L. Poteat, B. Lalevic, 'Pd-MOS Hydrogen and Hydrocarbon Sensor Device', *IEEE Electron. Device Lett.*, **2**, 82–84 (1981).
37. G.J. Maclay, K.W. Jelley, S. Nowroozi-Esfahani, M. Formosa, 'The Response of MOS Sensors with Ultrathin Palladium Gates to Carbon Monoxide and Methane', *Sens. Actuators*, **14**, 331–348 (1988).
38. Z. Weixin, Z. Yibing, 'Temperature Characteristics of an H_2S-sensitive P% Less than D%-gate MOS Transistor', *Sens. Actuators*, **15**, 85–93 (1988).
39. U. Ackelid, M. Armgarth, A. Spetz, I. Lundstrom, 'Ethanol Sensitivity of Palladium-gate Metal-oxide–Semiconductor Structures', *IEEE Electron Device Lett.*, **7**, 353–355 (1986).
40. W. Mokwa, K. Dobos, G. Zimmer, 'Palladium-gate MOS Devices for Arsine Detection', *Sens. Actuators*, **12**, 333–339 (1987).
41. W.P. Kang, J.F. Xu, B. Lalevic, T.L. Poteat, 'Sensing Behavior of Pd–$SnO_{(x)}$ MIS Structure Used for Oxygen Detection', *Sens. Actuators*, **12**, 349–366 (1987).
42. J. Janata, M. Josowicz, D.M. DeVaney, 'Chemical Sensors', *Anal. Chem.*, **66**, 207R–228R (1994).
43. D.C. Johnson, M.D. Ryan, G.S. Wilson, 'Dynamic Electrochemistry: Methodology and Application', *Anal. Chem.*, **60**, 147R–162R (1988).
44. C.G. Fox, J.F. Alder, in *Techniques and Mechanisms in Gas Sensing*, eds. P. Moseley, J. Norris, D.E. Williams, Adam Hilger, New York, Chapt. 13, 1991.
45. M.S. Nieuwenhuizen, A.J. Nederlof, 'Surface Acoustic Wave Gas Sensor for Nitrogen Dioxide Using Phthalocyanines as Chemical Interfaces. Effects of Nitric Oxide, Halogen Gases, and Prolonged Heat Treatment', *Anal. Chem.*, **60**, 236–240 (1988).
46. A. D'Amico, A. Palma, E. Verona, 'Palladium-surface Acoustic-wave Interaction for Hydrogen Detection', *Appl. Phys. Lett.*, **41**, 300–301 (1982).
47. A. Bryant, M. Poirier, G. Riley, D.L. Lee, J.F. Vetelino, 'Gas Detection Using Surface Acoustic Wave Delay Lines', *Sens. Actuators*, **4**, 105–111 (1983).
48. S.J. Martin, K.S. Schweizer, A.J. Ricco, T.E. Zipperian, 'Gas Sensing with Surface Acoustic Wave Devices', in *Transducers '85: Proceedings of International Conference on Solid State Sensors and Actuators*, New York, IEEE, Philadelphia, PA, 71–73, 1985.
49. K. Chan, H. Ito, H. Inaba, T. Furuya, '10 km-long Fibre-optic Remote Sensing of CH_4 Gas by Near Infrared Absorption', *Appl. Phys. B*, **38**, 11–15 (1985).
50. A.E. Bentley, J.F. Alder, 'Optical Fibre Sensor for Detection of Hydrogen Cyanide in Air. I: Reagent Characterization and Impregnated Bead Detector Performance', *Anal. Chim. Acta*, **222**, 63–73 (1989).
51. M.E. Lippitsch, J. Pusterhofer, M.J.P. Leiner, O.S. Wolfbeis, 'Fibre-optic Oxygen Sensor with the Fluorescence Decay Time as the Information Carrier', *Anal. Chim. Acta*, **205**, 1–6 (1988).
52. O. Wolfbeis, A. Sharma, 'Fibre-optic Fluorosensor for Sulphur Dioxide', *Anal. Chim. Acta*, **208**, 53–58 (1988).
53. R.B. Beswick, C.W. Pitt, 'Optical Detection of Toxic Gases Using Fluorescent Porphyrin Langmuir–Blodgett Films', *J. Colloid Interface Sci.*, **124**, 146–155 (1988).
54. S. Muto, A. Ando, T. Ochiai, H. Ito, H. Sawada, A. Tanaka, 'Simple Gas Sensor Using Dye-doped Plastic Fibers', *J. Appl. Phys.*, **28**, 125–127 (1989).
55. J.F. Giuliani, H. Wohltjen, N.L. Jarvis, 'Reversible Optical Waveguide Sensor for Ammonia Vapors', *Opt. Lett.*, **8**, 54–56 (1983).
56. H. Tai, H. Tanaka, T. Yoshino, 'Fiber-optic Evanescent-wave Methane-gas Sensor Using Optical Absorption for the 3.392-μm Line of a He–Ne Laser', *Opt. Lett.*, **12**, 437–439 (1987).
57. K. Nishizawa, in *Proceedings of 4th International Conference on Optical Fiber Sensors*, Tokyo, 131, 1986.
58. C. Nylander, B. Liedberg, T. Lind, 'Gas-detection by Means of Surface–Plasmon Resonance', *Sens. Actuators*, **3**, 79–88 (1982).
59. E. Campbell, H. Miller, *Chemical Detectors, a Bibliography for the Industrial Hygienist with Abstracts and Annotations*, LAMS-2378, Los Alamos Scientific Laboratory, Los Alamos, NM, Vol. I, 1961; Vol. II, 1964.
60. A.A. Grote, W.S. Kim, R.E. Kupel, 'Establishing a Protocol from Laboratory Studies to be Used in Field Sampling Operations', *Am. Ind. Hyg. Assoc. J.*, **39**, 880 (1978).
61. V. Brown, in *Proceedings of the HazTech International Conference, August 11–15, 1986, Denver, Colorado*, Colorado Ground Water Association, Denver, CO, 125–136, 1986.
62. *Dräger-tube Handbook*, National Dräger, Pittsburgh, PA, 1997.
63. *Precision Gas Detector System Manual*, Sensidyne/Gastec, Largo, FL, 1997.
64. *Direct Reading Colorimetric Tubes–A Manual of Recommended Practices*, American Industrial Hygiene Association, Fairfax, VA, 1997.
65. M. Dressler, in *Journal of Chromatography Library*, Elsevier, Amsterdam, Vol. 36, 1986.
66. R. Annino, R. Villalobos, *Process Gas Chromatography Fundamentals and Applications*, Instrument Society of America, Research Triangle Park, NC, 1992.
67. R.L. Grob, *Modern Practice of Gas Chromatography*, 3rd edition, Wiley-Interscience, New York, 1995.

68. J.N. Driscoll, J.H. Becker, 'Industrial Hygiene Monitoring with a Variable Selectivity Photoionization Analyzer', *Am. Lab.*, 69–76 (1979).
69. R.T. Smyth, D. Bingham, 'The Use of Nonspecific PID Monitors for Remedial Investigations and Field Studies', in *HazTec International Conference Proceedings, Denver, Colorado, Aug. 11–15, 1986*, Colorado Ground Water Association, Denver, CO, 1986.
70. I.N. Lee, S.S. Que Hee, C.S. Clark, 'Additivity of Detector Responses of a Portable Direct-reading 10.2 eV Photoionization Detector and a Flame Ionization Gas Chromatograph for Atmospheres of Multicomponent Organics: Use of PID/FID Ratios', *Am. Ind. Hyg. Assoc. J.*, **48**(5), 437–441 (1987).
71. C.W. Bayer, 'Advances in Trapping Procedures for Organic Indoor Pollutants', *J. Chromatogr. Sci.*, **32**, 312–316 (1994).
72. N.S. Arnold, W.H. McClennen, H.L.C. Meuzelaar, 'Vapor Sampling Device for Direct Short Column Gas Chromatography/Mass Spectrometry Analyses of Atmospheric Vapors', *Anal. Chem.*, **63**, 299–304 (1991).
73. G.F. Evans, T.A. Lumpkin, D.L. Smith, M.C. Somerville, 'Measurements of VOCs from the TAMS Network', *J. Air Waste Manage. Assoc.*, **42**(10), 1319–1323.
74. M. Harper, 'Methods of Characterizing Sorbents for Air Sampling Purposes', *Stud. Surf. Sci. Catal.*, **80**, 267–274 (1993).
75. D.W. Lehmpuhl, J.W. Birks, 'New Gas-chromatographic Electron-capture Detection Method for The Determination of Atmospheric Aldehydes and Ketones Based on Cartridge Sampling and Derivatization with 2,4,6-Trichlorophenylhydrazine', *J. Chromatogr. A*, **740**, 71–81 (1996).
76. R.W. Sherman, M.K. Solecki, E.S. Collard, T.H. McKinney, L.H. Grande, E.B. Overton, 'Development of a Field Portable Concentrator/Purge and Trap Device for Analysis of VOC in Ambient Air and Water Samples', in US Environmental Protection Agency, Res. Dev., Report EPA/600/D-89/189, *Field Screening Methods Hazard. Waste Site Invest.*, 279–282, 1988.
77. R. Annino, 'The Application of Fleuric Devices to Gas-chromatographic Instrumentation', *Adv. Chromatogr.*, **26**, 67–96 (1987).
78. W.H. McClennen, N.S. Arnold, H.L.C. Meuzelaar, 'Apparatus and Method for Sampling', *US Pat.*, 4 970 905, 1990.
79. US Environmental Protection Agency, Technical Assistance Document, 'Recommended Recordkeeping Systems for Air Pollution Control Equipment, Part 1: Particle Matter Controls', EPA-340/1-86-021, Stationary Source Compliance Division, Washington, DC, 1987.
80. R.E. Erbes, 'Draft Model Title V Permits for PM10 Emissions Form Lime Manufacturing Plants', in *A Practical Guide to Air Quality Compliance*, 2nd edition, John Wiley & Sons, New York, 1996.
81. J.P. Dworzanski, H.L.C. Meuzelaar, W. Maswadeh, X. Nie, P. Cole, N. Arnold, in *Proceedings of the 3rd International Symposium on Field Screening Methods for Hazardous Waste and Toxic Chemicals. Las Vegas, Feb. 1993*, Air and Waste Management Association, Pittsburgh, PA, 517–541, 1993.
82. R.D. Dandeneau, E.H. Zerenner, 'Investigation of Glasses for Capillary Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**, 351–356 (1979).
83. F.I. Onuska, F.W. Karasek, *Open Tubular Column Gas Chromatography in Environmental Sciences*, Plenum Press, New York, 1984.
84. L. Blomberg, 'Current Aspects of the Stationary Phase in Gas Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **5**, 520–533 (1982).
85. J.K. Haken, 'Developments in Polysiloxane Stationary Phases in Gas Chromatography', *J. Chromatogr.*, **300**, 1–77 (1984).
86. I. Rohrschneider, 'Eine Methode zur Charakterisierung von Gaschromatographischen Trennflussigkeiten. [A Method of Characterization of the Liquids Used for Separation in Gas Chromatography]', *J. Chromatogr.*, **6**, 22 (1966).
87. W.O. McReynolds, 'Characterization of Some Liquid Phases', *J. Chromatogr. Sci.*, **8**, 685–692 (1970).
88. E. Kovats, *Adv. Chromatogr.*, **1**, 229–247 (1966).
89. H. Van den Dool, P.D. Kratz, *J. Chromatogr.*, **11**, 463–471 (1963).
90. J.F. Sprouse, A. Varano, 'Development of a GC Retention Index Library ', *Am. Lab.*, **16**, 54–68 (1984).
91. K.J. Hyver, P. Sandra (eds.), *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard, Palo Acto, CA, ref. doc. 5950–3562, 1989.
92. M. Golay, in *Gas Chromatography 1957 (East Lansing Symposium)*, eds. V.J. Coates, H.J. Noebels, I.S. Fagerson, Academic Press, New York, 1–13, 1958.
93. J.H. Purnell, *J. Chem. Soc.*, **54**, 1268 (1960).
94. R. Annino, R. Villalobos, 'Eliminating Wasteful Cut and Try Procedures in Gas–Chromatography Method Development Using Computer Modeling', *J. Testing Eval.*, **29**(3), 143–148 (1996).
95. K. Grob, G. Grob, 'Practical Capillary Gas Chromatography. A Systematic Approach', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**, 109–117 (1979).
96. A.G. Loudon, J. Kolbe-Boelke, J. Adler, J. Stach, 'The Use of On-site Analysis for the Protection of Personnel', *Analusis*, **23**(2), M22–M24 (1995).
97. S. Dagan, A. Amirav, 'Fast, High-temperature and Thermolabile GC MS in Supersonic Molecular-beams', *Int. J. Mass Spectrom. Ion Processes*, **133**(2), 187–210 (1994).
98. R. Annino, 'Practical Limits of High Speed Chromatography: Evaluation with a Computer Based Theoretical Model of Several Commercially Available Gas Chromatographs', *J. High Resolut. Chromatogr.*, **19**, 285–290 (1996).

99. S.A. Van, J. Janssen, R. Bally, C. Cramers, J. Rijks, 'Sample Introduction in High Speed Capillary Gas Chromatography; Input Band Width and Detection Limits', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **10**, 273–279 (1987).
100. P. Sandra, 'Fast Capillary Gas Chromatography', *LC-GC*, **5**, 236–246 (1987).
101. K.J. Hyver, R.J. Phillips, 'Considerations in Enhancing Resolution, Speed, and Sensitivity in Capillary Gas-chromatography and Gas-chromatography Mass-spectrometry', *J. Chromatogr.*, **399**, 33–46 (1987).
102. R. Villalobos, R. Annino, 'The Computer-aided Optimization of Capillary Columns for Minimum Time Analysis and Minimum Detectability', *J. High Resolut. Chromatogr.*, **12**, 149–160 (1989).
103. R. Villalobos, R. Annino, 'Computer-aided Selection and Optimization of Chromatographic Columns and Conditions for Multicolumn Analysis Procedures', *J. High Resolut. Chromatogr.*, **13**, 764–773 (1990).
104. K. Grob, *On-column Injection in Capillary Gas Chromatography*, Hüthig, New York, 1987.
105. P. Sandra (ed.), *Sample Introduction in Capillary Gas Chromatography*, Hüthig, New York, Vol. 1, 1985.
106. K. Grob (ed.), *Classical Split and Splitless Injection in Capillary Gas Chromatography*, 2nd edition, Hüthig, New York, 1988.
107. M.S. Klee, 'Gas Chromatography Inlet: an Introduction', Ref. 5958–9468, Hewlett-Packard, Palo Alto, CA, 1990.
108. T. Gorecki, J. Pawlazyn, 'Sample Introduction Approaches for Solid-phase Microextraction Rapid GC', *Anal. Chem.*, **67**(18), 3265–3274 (1995).
109. B.V. Ioffe, A.G. Vitenberg, *Head-space Analysis and Related Methods in Gas Chromatography*, John Wiley & Sons, New York, 1984.
110. K. Des Tombe, D.K. Verma, L. Stewart, E.B. Reczek, 'Sampling and Analysis of Light-Hydrocarbons C_1-C_4 – A Review', *Am. Ind. Hyg. Assoc. J.*, **52**, 136–144 (1991).
111. D.R. Deans, 'A New Gas Sampling Device for Gas-Chromatography', *J. Chromatogr.*, **43**, 43 (1984).
112. M.A. Klemp, M.L. Akard, R.D. Sacks, 'Cryofocusing Inlet with Reverse Flow Sample Collection for Gas-chromatography', *Anal. Chem.*, **65**, 2516–2521 (1993).
113. L.A. Lanning, R.D. Sacks, R.F. Mouradian, S.P. Levine, J.A. Foulke, 'Electrically Heated Cold Trap Inlet System for Computer-controlled High-speed Gas-chromatography,' *Anal. Chem.*, **60**, 1994–1996 (1988).
114. 'Detectors for Gas Chromatography, a Practical Primer', Ref. 5958–9433, Hewlett-Packard, Palo Alto, CA, 1987.
115. C. Haigh, 'Gas Chromatography Analysis of Sulfur Compounds in Naptha by Capillary Gas Chromatography', Hewlett-Packard Application, Publication No. 43-5954-7615, Hewlett-Packard, Palo Alto, CA, 1986.
116. D.E. Clouser, J.S. Craven (to Hewlett-Packard), *US Pat.*, 4 254 654, 1981.
117. J.S. Craven, D.E. Clouser, 'A Fresh Design for Thermal Conductivity Detector', Presented at the 30th Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Cleveland, OH, March 7, 1979.
118. P.C. Uden, K.J. Slatkavitz, R.M. Barnes, R.L. Deming, 'Empirical and Molecular Formula Determination by Gas Chromatography/Microwave-induced Plasma Atomic Emission Spectrometry', *Anal. Chim. Acta*, **180**, 401–416 (1986).
119. E.U. Ehrmann, H.P. Dharmasena, K. Carney, E.B. Overton, 'Novel Column Heater for Fast Capillary Gas-chromatography', *J. Chromatogr. Sci.*, **34**, 533–539 (1996).
120. R.V. Arenas, K.R. Carney, E.B. Overton, 'Portable, Multigas Monitors for Air-quality Evaluation. 2. Survey of Current Models', *Am. Lab.*, **25**(11), 25–31 (1993).
121. A.P. Snyder, C.S. Harden, A.H. Brittain, M.G. Kim, N.S. Arnold, H.L.C. Meuzelaar, 'Portable Hand-held Gas Chromatography/Ion Mobility Spectrometry Device', *Anal. Chem.*, **65**, 299–306 (1993).
122. M.W. Bruns, 'Silicon Micromachining and High Speed Gas Chromatography', in *Proceedings of the 1992 International Conference on Industrial Electronics, Control, Instrumentation and Automation. Power Electronics and Motion Control*, **3**, 1640–1644 (1992).
123. S.J. Santy, 'Portable Gas Chromatograph', Presented at the Gas Quality Measurement Symposium, Institute of Gas Technology, Chicago, IL, June 10–12, 1991.
124. R. Siemers, D. Heigel, A. Spilkin, 'Rapid Micro-GC Analysis for Permanent Gases', *Am. Lab.*, **23**(March), 44L–44R (1991).
125. M. Goedert, G. Guichon, 'Sources of Error in Measurement of Retention Times', *Anal. Chem.*, **42**, 962–968 (1970).
126. M. Goedert, G. Guichon, 'High Precision Measurements in Gas Chromatography. A Study of Systematic Errors on the Determination of Retention Times', *Anal. Chem.*, **45**, 1188–1196 (1973).
127. J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Raven Press, New York, 1997.
128. J.B. Farmer, in *Mass Spectrometry*, ed. C.A. McDowell, McGraw-Hill, New York, 7–44, 1963.
129. J. Roboz, *Introduction to Mass Spectrometry*, Wiley-Interscience, New York, 1968.
130. S.J. Mullock, 'Ion Gun and Mass Spectrometer Employing the Same Ion Gun', *Eur. Pat.*, 6 873 381, 1995.
131. P.H. Dawson, B.Q. Yu, 'The 2nd Stability Region of the Quadrupole Mass Filter. 1. Ion Optical-properties', *Int. J. Mass Spectrom. Ion Processes*, **56**, 25–50 (1984).
132. W. Paul, M. Raether, 'Das elektrische Massenfilter', *Z. Phys.*, **140**, 262–271 (1955).
133. *Mass Spectrometry Fundamentals*, Hewlett-Packard Manual Part Number G1626-90002, Hewlett-Packard, Palo Alto, CA, 1993.

134. J. Todd, G. Lawson, R. Bonner, in *Quadrupole Mass Spectrometry and its Applications*, ed. P.H. Dawson, Elsevier, New York, 181–224, 1976.
135. D. Dietrich, R. Keville, 'Miniature Mass Spectrometer', *Sci. Tech. Rev.*, (Nov./Dec.), 30–31 (1995).
136. J. Christenson, B. Eckenrode, D. Gallis, 'Field-proven Transportable GC/MS: Real-world Case Studies and Success Stories', in *Proceedings of the Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Publication No. VIP-47, Air and Waste Management Association, Pittsburgh, PA, 1322, Vol. 2, 1995.
137. F.W. Karasek, 'Plasma Chromatography', *Anal. Chem.*, **46**(8), 710A–720A (1974).
138. G.A. Eiceman, Z. Karpas, *Ion Mobility Spectrometry*, CRC Press, Boca Raton, FL, 1994.
139. G.A. Eiceman, 'Advances in Ion Mobility Spectrometry: 1980–1990', *CRC Crit. Rev. Anal. Chem.*, **22**, 17–36 (1991).
140. R.H. St. Louis, H.H. Hill, Jr, 'Ion Mobility Spectrometry in Analytical Chemistry', *CRC Crit. Rev. Anal. Chem.*, **21**, 321–355 (1990).
141. H.H. Hill, Jr, W.F. Siems, R.H. St. Louis, 'Ion Mobility Spectrometry', *Anal. Chem.*, **62**, 1201A–1209A (1990).
142. G.E. Spangler, D.N. Campbell, K.N. Vora, J.P. Carrico, 'Developments in Ion Mobility Spectrometry', *ISA Trans.*, **23**(1), 17–28 (1984).
143. P. Kebarle, 'Ion Thermochemistry and Solvation from Gas Phase Ion Equilibria', *Annu. Rev. Phys. Chem.*, **28**, 445–476 (1977).
144. G. Nicol, J. Sunner, P. Kebarle, 'Kinetics and Thermodynamics of Protonation Reactions: $H_3O^+ + (H_2O)_h + B = BH^+(H_2O)_b + (h - b + 1)H_2O$, where B is a Nitrogen, Oxygen or Carbon Base', *Int. J. Mass Spectrom. Ion Processes*, **84**, 135–155 (1988).
145. A. Good, D.A. Durden, P. Kebarle, 'Ion–molecule Reactions in Pure Nitrogen and Nitrogen Containing Traces of Water at Total Pressures 0.5–4 Torr. Kinetics of Clustering Reactions Forming $H^+(H_2O)_n$', *J. Chem. Phys.*, **52**, 212–221 (1970).
146. M. Shahin, *J. Chem. Phys.*, **45**, 2600–2605 (1965).
147. F.W. Karasek, H.H. Hill, Jr, S.H. Kim, S. Rokushika, 'Gas-chromatographic Detection Modes for Plasma Chromatograph', *J. Chromatogr.*, **135**, 329–339 (1977).
148. F.W. Karasek, G.E. Spangler, in *Electron Capture–Theory and Practice in Chromatography*, eds. A. Zlatkis, C.F. Poole, Elsevier, Amsterdam, 377–406, 1981.
149. E.P. Grimsrud, 'Gas-phase Ion Chemistry within the Electron-capture Detector', *Mass Spectrom. Rev.*, **10**, 457–489 (1991).
150. E.P. Pellizari, *J. Chromatogr.*, **98**, 323 (1974).
151. S.H. Kim, G.E. Spangler, 'Ion Mobility Spectrometry Mass-spectrometry of 2 Structurally Different Ions Having Identical Ion Mass', *Anal. Chem.*, **57**, 567–569 (1985).
152. G.E. Spangler, M.D. Lutherville, J.N. Cox, 'Utility Ion Mobility Spectrometer System with Improved Specificity', *US Pat.*, 4 551 624, 1985.
153. G.E. Spangler, J.P. Carrico, D.N. Campbell, 'Recent Advances in Ion Mobility Spectrometry for Explosives Vapor Detection', *J. Test. Eval.*, **13**(3), 234–240 (1985).
154. G.E. Spangler, D.E. Campbell, J.P. Carrico, presented at the 1983 Pittsburgh Conference and Exposition on Analytical Chemistry, Atlantic City, NJ, March 1983.
155. F.W. Karasek, 'The Plasma Chromatograph', *Res. Dev.*, **21**, 34–37 (1970).
156. M.J. Cohen, F.W. Karasek, 'Plasma Chromatography TM – New Dimension for Gas Chromatography and Mass Spectrometry', *J. Chromatogr. Sci.*, **8**, 330–337 (1970).
157. N.E. Bradury, R.A. Nielsen, *Phys. Rev.*, **49**, 388–393 (1936).
158. H. Hill, D. McMinn (eds.), *Detectors for Capillary Chromatography*, John Wiley & Sons, Chichester, 1992.
159. F.J. Knorr, R.L. Eatherton, W.F. Siems, H.H. Hill, Jr, 'Fourier Transform Ion Mobility Spectrometry', *Anal. Chem.*, **57**, 402–406 (1985).
160. J.P. Dworzanski, M.G. Kim, A.P. Snyder, N.S. Arnold, H.L.C. Meuzelaar, 'Performance Advances in Ion Mobility Spectrometry Through Combination with High-speed Vapor Sampling, Preconcentration and Separation Techniques', *Anal. Chim. Acta*, **293**, 219–235 (1994).
161. H.H. Hill, G. Simpson, 'Capabilities and Limitations of Ion Mobility Spectrometry for Field Screening Applications', *Field Anal. Chem. Technol.*, **VI**(N3), 119–134 (1997).
162. J. Roehl, 'Environmental and Process Applications for Ion Mobility Spectrometry', *Appl. Spectrosc. Rev.*, **26**, 1–57 (1991).
163. R.H. St. Louis, H.H. Hill, Jr, 'Ion Mobility Spectrometry in Analytical Chemistry', *CRC Crit. Rev. Anal. Chem.*, **21**, 321–355 (1990).
164. Technical Brochures, 'TDI/AVM Monitor, PUMA, Airbone Vapor Monitor (AVM), AVM Material List and E.IMS', Graseby Ionics, Watford, Herts, 1996.
165. W.M. Watson, C.F. Kohler, 'Continuous Environmental Monitoring of Nickel Carbonyl by Fourier-transform Infrared Spectrometry and Plasma Chromatography', *Environ. Sci. Technol.*, **13**, 1241–1243 (1979).
166. R.C. Young, J.C. Travis, 'A Feasibility Study of an Ion Mobility Spectrometer (IMS) for Hydrazine Vapor Detection', NASA Technical Memo 102 150, Resource Technology, JFK Space Center, FL, 1989.
167. A.T. Bacon, R. Getz, J. Reategui,'Ion-mobility Spectrometry Tackles Tough Process Monitoring', *Chem. Eng. Prog.*, **87**, 61–64 (June 1991).
168. J.L. Brokenshire, V. Dharmarajan, L.B. Coyne, J. Keller, 'Near Real Time Monitoring of TDI Vapour Using Ion Mobility Spectrometry (IMS)', *J. Cell. Plast.*, **26**, 123–142 (1990).

169. A.H. Lawrence (ed.), *Cargo Inspection Technologies, San Diego, CA, 25–27 July*, in *SPIE Proc. Ser.*, 1994-07-25, 1994.
170. L.L. Danylewych-May, 'Modification to the Ionization Process to Enhance the Detection of Explosives by IMS', in *Proceedings of the 1st Int. Symp. on Expolsives Detection Technology*, N93-21856 07-03, FAA, 672–686, 1992.
171. J.B. Schultz, 'Detection of Concealed Explosives and Narcotics Using Ion Mobility Spectrometry', *ICAO J.*, **45**(June), 19–20 (1990).
172. N.S. Arnold, J.P. Dworzanski, H.L.C. Meuzelaar, W.H. McClennen, 'Challenges of Developing a GS/IMS Based Personal Chemical Hazard Detector', *Proceedings of Int. Symp. on Field Screening Methods for Hazardous Wastes and Toxic Chemicals, Las Vegas, 1995*, Publication No. VIP-47, Air and Waste Management Association, Pittsburgh, PA, 1038–1045, 1995.
173. N. Arnold et al., in *Proceedings of the Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Las Vegas, Nevada, 1038, Vol. 2, 1995.
174. W.C. Blanchard, A.T Bacon, 'Utility Ion Mobility Spectrometer', *US Pat.*, 4 797 554, 1989.
175. M. Klemp, Chromab FAST, personal communication, 1997.

Microelectromechanical Systems Technology Applied to the Miniaturization of Field Instrumentation

Michel Goedert and William Higdon
Hewlett-Packard Laboratories, Palo Alto, USA

1 Introduction 1
2 Growing Importance of Microelectromechanical Systems 2
3 Basic Microelectromechanical Systems Technology 2
3.1 Photolithography 3
3.2 Silicon Etching 3
3.3 Material Deposition 5
3.4 Bonding of Microelectromechanical Systems Structures 6
4 Example of Processing of Microelectromechanical Systems Component 7
5 Applications of Microelectromechanical Systems to Portable Instrumentation 9
5.1 Gas Sensors 10
5.2 Portable Gas Chromatographs 12
5.3 Miniaturization of Mass Spectrometers 20
6 Conclusion 21
Abbreviations and Acronyms 21
Related Articles 21
References 23

This paper explores the miniaturization of analytical instruments using emerging microelectromechanical systems (MEMS) technology. MEMS provides an enabling technology base for miniaturization of many kinds. It has grown in parallel with the growth of the semiconductor industry since the discovery of silicon piezoresistance at the Bell Laboratories in the early 1950s, and borrows many principles and techniques from silicon ''wafer fabs''. The technology is one of several that promises to revolutionize our lives in the 21st century.

We review some of the processes developed to fabricate MEMS parts and give examples of components developed with this technology. Finally, we describe how this technology has been used to develop a portable gas chromatograph (GC) and other chemical instrumentation for the detection of gases and vapors in the field. We show, with experimental results, how analytical instrumentation can benefit from MEMS technology.

1 INTRODUCTION

MEMS technology has emerged since the late 1950s as an outgrowth of the semiconductor industry. Research laboratories in academia and industry have studied potential commercial MEMS applications and technologies; however, progress has been slow and sometimes arduous compared to the fast-paced electronics revolution. MEMS exploits many more physical, chemical, and optical effects than does pure electronics, and the design and fabrication effort for MEMS is correspondingly greater.

The piezoresistive effect in silicon was discovered at Bell Laboratories in 1954, allowing silicon to be used for sensing pressure and force. In the 1960s the first silicon-diaphragm pressure sensors were introduced, and by 1979 they had penetrated the automobile market as manifold absolute pressure (MAP) sensors. Since then other MEMS devices have made strides into automotive, medical, ink-jet printer nozzle, and industrial markets.

In the early 1970s NASA (National Aeronautics and Space Administration) funded Stanford University in California to develop a miniature high speed GC intended for use on the two Mars Viking missions in 1975 and 1976. NASA wanted a small lightweight analytical instrument that could operate with low resource consumption and a sensitive detector, but be rugged enough to withstand the gravitational forces of a rocket launch and a Mars landing. A breadboard instrument was developed to demonstrate the feasibility of this technology applied to measurement of atmospheric gases. For the first time a design integrated a gas injector, a small column and a detector in one silicon wafer. The GC-on-a-chip was born.[1]

This GC[2] constituted a technology breakthrough and a new trend in the future development of analytical instruments. On the down side, the ''GC chip'' itself was only a part of the overall GC instrument; data handling equipment, gas supply and control were all external to the silicon wafer.

The Stanford GC project led in 1983–1998 to the development of portable GCs for industrial applications by MTI Analytical Instruments; those products were acquired by Hewlett-Packard (Palo Alto, CA) in 1998, and their development continues.

These miniature GC systems illustrate how MEMS technology can provide miniaturization and unique performance advantages, and how silicon micromachining applied to analytical instrumentation provides advantages over precision machining when small dimensions are crucial, as is often the case in medical, analytical or biological applications.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Less visible than the GC on the chip, but economically more successful because of higher volume of production, other silicon micromachined components such as inkjet heads[3] and sensors used by the automotive industry[4,5] were also introduced at about the same time.

In a complex electromechanical environment such as a car, the need for effective, accurate, reliable and low-cost sensors is pressing. This is why the automotive industry accounts for the largest share of the market for MEMS silicon sensors.[6] The automotive industry recognized many years ago that a smarter car was inevitable to meet strict US Federal requirements to protect passengers and the environment. Low-powered semiconductor devices and rapid advancements in computers made it possible to place sensors throughout the automobile to monitor its wellbeing. These sensors continuously send status reports to the onboard computer which controls actuators to maintain predetermined specifications. Cars now have dozens of MEMS transducers monitoring everything from engine combustion and transmission temperature, to oil and water pressure, to accelerometer devices that inflate air bags. Fuel injector nozzles are MEMS devices that enable the engine to operate at peak performance. These devices have made it possible to keep the cost of automobiles at a reasonable level while dramatically increasing their performance and longevity.[7]

Many manufacturing facilities use MEMS technology in the multiplicity of sensors throughout their processes. Air conditioning systems use MEMS flow sensors and temperature sensors to control the environment.[8] Navigational gyroscopes and global positioning systems (GPS) are also MEMS devices.[9] Manufacturers of analytical instruments are also benefiting from sensor development and are providing better and cheaper control and measurements of fluids. Unfortunately, MEMS device development is much slower than microelectronics in general, making them bottlenecks for even further performance and cost improvements of systems.[10]

2 GROWING IMPORTANCE OF MICROELECTROMECHANICAL SYSTEMS

The total USA market for MEMS in 1996 was approximately $1 billion in revenues, an increase of 12% over the revenues for 1995. Revenues are expected to grow steadily through the turn of the century. The MEMS military and aerospace market is by far the smallest of the market segments. The significance of this market is reflected in the Department of Defense's (DoD) current initiatives in funding many MEMS research activities. The aerospace portion consists of sales generated from accelerometers and pressure sensors used in commercial aviation in applications such as altimeters, barometers, collision avoidance systems, navigation systems, and cabin pressure. This segment represented $62.2 million in revenues and 5.8% of the overall MEMS market revenues in 1996. The MEMS automotive market is a significant component of the MEMS industry and is one of the market drivers associated with the establishment of MEMS as a viable commercial technology. Most of this segment consists of sales from airbag accelerometers and pressure sensors, and MAP sensors. In 1996, the segment represented $354 million in revenues – 33% of total market revenues.

Along with automotive, industrial, and ink-jet markets, the MEMS medical market has facilitated MEMS success to date. The bulk of the medical market is made up by the sales of disposable blood pressure sensors. The USA MEMS medical market totaled $164.4 million in revenues in 1996, representing 15.3% of the total MEMS market. The information technology (IT) market comprises ink-jet printer cartridges and MEMS displays. In 1996, the IT portion totaled revenues of $400 million.[11]

3 BASIC MICROELECTROMECHANICAL SYSTEMS TECHNOLOGY

MEMS technology plays an important role in the very practical science of miniaturization, but is often associated only with silicon. Mechanical properties of silicon are unique and well established,[12] but other materials, such as insulators, can also be used for MEMS fabrication. MEMS technology, either silicon or insulator based, when applied to portable chemical instrumentation, is a unique tool used to produce three-dimensional (3-D) structures most often from two-dimensional photographic patterns. These parts can have etched cavities, microchannels, or other fluid conduits, and are formed by adding layers of materials that are fused together. The results are structures that can be designed to sustain vacuum or high pressure over a wide span of subambient or elevated temperatures. Thin layers of metals can also be evaporated on insulated suspended structures to provide sensing elements used, for instance, for the fabrication of gas sensors. MEMS techniques can be complex and costly to implement, but are able to produce unique, miniaturized and very versatile parts.

To appreciate fully the advantages and limitations of MEMS, some basic aspects of integrated circuit (IC) technology will first be reviewed. We will describe the most important IC processes used to produce MEMS structures, i.e. photolithography, wet and dry etching, material addition, and material bonding. We will also describe processes used most exclusively for MEMS. For instance, screen printing has a great value for low-volume applications such as dedicated sensors. Another

emerging application is LIGA, the German acronym for X-ray lithography, electrodeposition (galvanoformung), and molding (abformtechnik), which is a sophisticated approach to producing structures by sculpting them in 3-D in thick layers of photosensitive material. It is based on high-aspect-ratio lithography and plating.

This review of MEMS techniques is far from complete; our intent is to show the possibilities of this technology with a limited number of practical examples. The interested reader will find some excellent books on MEMS. Madou[13] gives useful information about a broad range of applications. Reviews of MEMS technology trends[14] or applications are periodically published in specialized journals such as *Sensors and Actuators B* (published by Elsevier). In addition many internet sites contain good descriptions of IC fabrication and MEMS fabrication.

3.1 Photolithography

Photolithography was first developed by a Frenchman named Joseph-Nicéphore Niépce in 1822. The word ''lithography'' comes from the Greek words *lithos* (stone) and *graphein* (to write). The process we know today was first used in 1950 to produce ICs on silicon wafers. With the advent of ICs, the process of patterning smaller and smaller features onto silicon and gallium arsenide (GaAs) wafers led to ever more sophisticated technologies that are still evolving today. Millions of devices can perform an array of functions at high speed on a few square millimeters of silicon. However, the major difference is that semiconductors use two-dimensional photolithography (on the x- and y-axis), while MEMS devices typically use a third z, axis to be able to create 3-D features.

Photolithography is the first basic tool of MEMS processes. In its principle it is simple: ultraviolet (UV) radiation shines through a piece of glass covered with a predetermined opaque chrome pattern (called a ''photo mask plate'') onto a thin film surface of photosensitive organic polymeric (called ''photoresist'').[15] The pattern, which is on the mask plate, is reproduced on the exposed area of the silicon wafer (Figure 1).

The photoresist thus exposed changes its molecular character from the nonexposed areas and is then susceptible to development by solvents. The solvents remove only the exposed areas in the case of a ''positive resist'' or the nonexposed area of a ''negative resist''. After this step, the bare oxide layer on the silicon wafer is exposed and ready for the next step of the process. The unexposed areas retain a protective layer of photoresist that shields them from the next process step. The thin layer of silicon oxide is removed where it is exposed, leaving bared silicon. Features with resolutions of less than 0.25 μm can be achieved.

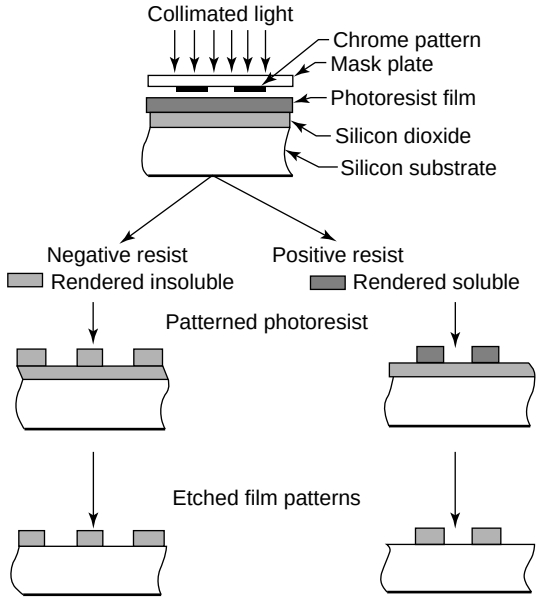

Figure 1 Basic principle of photolithography. Negative resists become less soluble in developer when they are exposed to radiation, and positive resists become more soluble after exposure and thus allow ''positive'' or ''negative'' patterns.

3.2 Silicon Etching

Another important step of MEMS processing technology is the etching after the masking operation. Etching is done to create cavities, holes or modify the surface of structures. Etching can be done using chemicals (wet etching), plasma (dry etching) or both. Wet etching can be isotropic, meaning that the etch rate is independent of the crystal orientation of the silicon, or anisotropic (orientation dependent) with different rates in different crystal directions within a single-crystal piece of silicon according to a set of directions which is determined by the orientation of silicon used.

When ordering silicon, the crystal orientation must be specified. The most common orientation used in the IC industry is the ⟨100⟩ orientation but MEMS uses ⟨110⟩ wafers as well. Some definitions are necessary to understand the difference between these orientations. Silicon has a diamond-lattice crystal structure. Each silicon atom has four nearest neighboring atoms to which it is covalently bonded. The principal axes in a crystal can also be used to develop a notation to define specific directions and planes (Figure 2), termed ''Miller indices''.[16] These are a series of small integer numbers enclosed in caret, brackets, parentheses, and braces. For example, [111] denotes a specific direction, whereas ⟨111⟩ denotes the family of all eight directions equivalent to [111]. A (100) notation denotes a particular lattice plane, and {100} denotes all the planes crystallographically equivalent to (100).

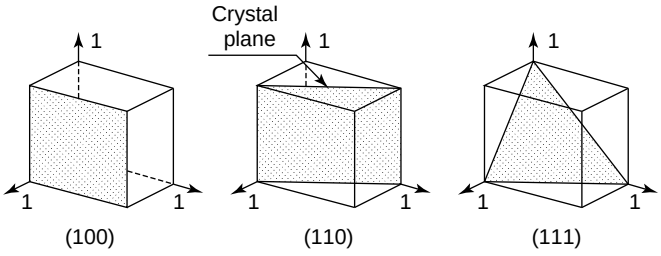

Figure 2 Schematic representation of Miller indices in a cubic lattice system. Crystalline silicon shares with diamond the crystal structure called interlocking face-centered cubic. The axes of the unit cube provide a rectilinear coordinate system that makes it possible to specify directions and planes within the crystal.

The processing characteristics and some material properties of silicon wafers depend on the orientation. The {111} planes have the highest density of atoms on the surface, so crystals grow most easily on these planes. Mechanical properties such as tensile strength are higher for ⟨111⟩ directions. The module of elasticity also shows orientation dependence. Processing characteristics such as oxidation are similarly orientation dependent. For instance, {111} planes oxidize faster than {100} planes, because they have more atoms per unit surface area available for the oxidation reaction to occur.

3.2.1 Wet Etching

Wet etching still predominates over dry etching in many situations. Wet etching of silicon results in atomically smooth surfaces and sharp edges that cannot be obtained with dry etching. It is also cheaper and is usually considered environmentally friendly. Figure 3 illustrates schematically the difference between isotropic and anisotropic etching on a ⟨100⟩ oriented wafer. In isotropic etch (Figure 3a) the silicon is etched uniformly in all directions and at the end the cross-section of the structure is roughly semicircular. Note also that the silicon is removed under the mask (undercut) and therefore the final dimension of the structure will be larger than the dimension defined by the mask. With anisotropic etching, the lateral etching is much smaller than the vertical etching. The cross-section of the structure at the end looks like Figure 3(b). Etching stops when the two ⟨111⟩ planes meet and the undercut is almost negligible. The angle between the top and the side of the channel is a constant (54.74°) and corresponds to the angle between ⟨100⟩ and ⟨111⟩ planes. There are two important consequences. The first is that with this etching approach it is not possible to etch parallel walls with a ⟨100⟩ oriented silicon wafer. The second consequence is that the maximum depth of the channel is set by the opening of the mask. Finally, by using ⟨110⟩-oriented silicon, straight-walled grooves are formed (see Figure 3c). In this way 0.6-μm wide openings on 1.2-μm centers and 600-μm deep have been made.[17] The

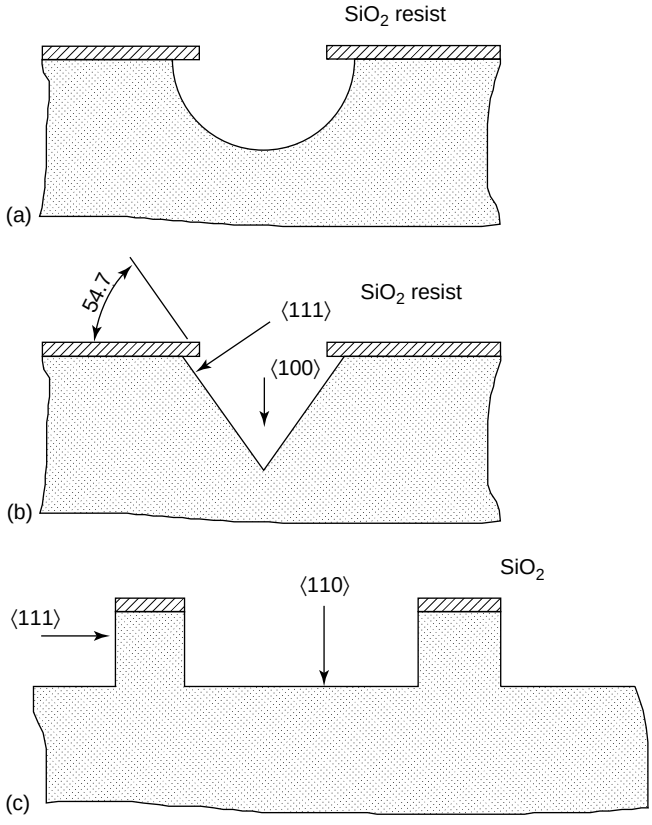

Figure 3 Basic description of (a) isotropic etch and (b) anisotropic etch on silicon (⟨100⟩ orientation) and (c) anisotropic etching on silicon (⟨110⟩ orientation).

choice of crystal orientation, therefore, is generally not left to the discretion of the crystal grower, but is a device design consideration which depends on the structure.

Table 1 gives a brief summary of properties of some chemical etchants applied to silicon.[12]

HF acid and KOH are two popular etchants used for the microfabrication of parts. Etching with HF gives isotropic results (same etching rate in all directions) with an etch rate of approximately 3.0 μm min^{-1}. KOH is also orientation dependent, but exhibits a high anisotropic etching ratio (400 : 1) of ⟨110⟩ to ⟨111⟩ and, for this reason, is useful for groove etching. A disadvantage of HF is that it etches SiO_2 at a rate that limits its use as a mask when deep etching is considered. In structures requiring long etching times, silicon nitride (Si_3N_4) is the preferred masking material.

3.2.2 Dry Etching

The major advantage of dry etching is that it is directional for most materials. Dry etching gives more control, in terms of anisotropy, but is also more expensive and complex to implement than wet etching. Two dry etching methods are commonly used: reactive ion etching (RIE)

Table 1 Chemical etching of silicon[12]

Etchant (diluent)	Typical composition (mL)	Temp. (°C)	Etch rate ($\mu m\ min^{-1}$)	Anisotropic etch rate ratio	Masking film
HF	25	22	0.7–3.0	1:1	SiO_2 (300 Å min^{-1})
HNO_3	30				
(Water, CH_3COOH)	80				
KOH	50 g	50	1	400:1	SiO_2 (14 Å min^{-1})
(Water, isopropyl)	100				

HF, hydrofluoric; KOH, potassium hydroxide–water–isopropyl alcohol.

and ion milling. Both methods are somewhat directional. With RIE, the surface is bombarded with high-energy ions creating radicals which react with surface atoms. With ion milling, higher energy neutral atoms physically sputter material from the surface. When the bombardment is perpendicular to the surface, the ratio of the depth to the width of etching, defined as aspect ratio, is generally higher than 50:1. A magnetically confined gas discharge is obtained between a thermoionic cathode and a concentric-anode cylinder. To extract the ion beam from the discharge region, a potential difference is applied between a pair of grids with aligned holes. The ions are injected into the working space as a well-collimated energetic beam. The beam is made neutral by extracting low-energy electrons from an auxiliary thermoionic cathode so that it can be used for sputtering insulators as well as conductors.[18]

Reactive plasma etching is another technique which offers several important advantages, when dry etching is considered. In contrast to sputter etching, large differences in etch rates for different materials can be obtained. With this technique, the substrate is exposed to plasma of a reactive gas. Since the ions are not highly energetic, sputtering is not in itself an important surface-removal mechanism. Rather, the reactive gas components appear to be first absorbed on the surface, where they are dissociated by electrons or ion bombardment from the plasma. Then they react with the film material, forming a gaseous product molecule. The reaction products are desorbed as gases and pumped away. Anisotropic etching can occur because the impingement rate on the horizontal surface is much greater than at the side walls at the low pressures used, where the mean free path of the molecules is typically much larger than the depth to be etched. Because of the chemical nature of the process, a high degree of control over the relative etch rates for the film, resist, and substrate can be obtained by the choice of suitable materials.[19]

3.3 Material Deposition

Material deposition used for MEMS processes includes the addition of a wide range of organic and inorganic materials. Additive processes involve surface modification such as ion implantation, annealing or deposition. Solids can be deposited from a liquid, plasma, gas, or a solid state. In the deposition methods from the gas phase two categories can be distinguished, a direct line-of-sight deposition technique called physical vapor deposition (PVD) and a diffuse-convective mass transfer technique such as chemical vapor deposition (CVD). PVD techniques include evaporation, sputtering, molecular beam epitaxy, and laser ablation deposition. Plasma-enhanced CVD, atmospheric pressure CVD, low-pressure CVD, very low-pressure CVD, metallorganic CVD and spray pyrolysis are some of the CVD techniques used with MEMS.[13]

Among PVD techniques, thermal evaporation is one of the oldest thin film deposition techniques. Evaporation is based on boiling off a heated material in a vacuum.[20] Evaporation is used to produce films of high purity and known structure at a high rate of deposition (typically 0.5 $\mu m\ min^{-1}$ for Al).

Sputtering is preferred over evaporation for several reasons:[20] there is almost no limitation in the choice of materials to be deposited, in-line sputtering equipment is available when successive depositions have to be done, in situ cleaning can be done with sputter etch, the adhesion is excellent, and the deposition can be done over larger areas. The major negative aspect is the high cost and complexity of the equipment.

CVD is another useful tool used for the microfabrication of MEMS structures. CVD is used to produce low-temperature film such as oxides and silicon nitride films. One of the challenges of any surface-micromachined MEMS structure is to limit the residual stress of the film. If we take, for instance, the fabrication of a conventional flow sensor, the heating and sensing elements are suspended on a bridge of silicon nitride to limit the thermal losses between the sensing element and the thermal ground. If the residual stress is not negligible at the operating temperature, the supporting structure will buckle and break. CVD is a very useful tool and enables the user to control the residual stress of the film during deposition.

3.3.1 Silk-screening or Screen Printing

Screen printing falls in hybrid technology rather than MEMS but is worth mentioning because it presents a cost-effective means of depositing a wide variety of films on planar substrates for low-volume production. A paste or ink is pressed onto a substrate through openings in the emulsion of a stainless steel screen.[21] The paste consists of a mixture of the material of interest, an organic binder and a solvent. The organic vehicle determines the flow properties of the paste. The bonding agent provides adhesion of particles to one another and to the substrate. The active particles make the ink a conductor, a resistor, or an insulator. The lithographic patterns in the screen emulsion are transferred onto the substrate by forcing the paste through the mask openings with a squeegee.

Inks have been developed for sensor applications. For example, pastes incorporating Pt, Pd, and Sb dopants have been developed for the construction of high temperature (>300 °C) gas sensors for reducing gases.[22] Thick metal films have been deposited on alumina to form the active material in low-temperature (<180 °C) gas sensors.[23] For biosensor applications, thick-film technology based on pastes that can be deposited at ambient temperature is crucial. Special grades of polymer-based paste have been developed for electrodes,[24] but are applicable to other MEMS structures such as chemical sensors.

Screen printing forms an excellent alternative to thin-film technology when size does not matter but cost in small production does. Biomedical applications fall into this category. Thin-film technology offers better resolution. The finish of the film is also smoother and more reproducible.

3.3.2 X-ray Lithography, Electrodeposition and Molding Technology Applied to Microelectromechanical Systems

The LIGA technique involves high-energy X-ray lithography of thick photoresist from micrometers to centimeters and development to make a resist mold. By applying galvanizing techniques the mold is filled with a metal. The resist structure is removed and metal product results. Alternatively, the metal part can serve as a mold itself for precision plastic injection molding.[25,26] Several types of plastic molding processes have been tested, including reaction injection molding, thermoplastic injection molding, and hot embossing. The so-formed plastic part, just as the original resist structure, can serve as a mold again for fast and cheap production since it does not rely on a new X-ray exposure. LIGA enables new building materials and a wider dynamic range of dimensions and possible shapes to be used.

LIGA relies on wet chemical plating technology to deposit metals on plastic or silicon.[27] Plating allows thick deposition of metal (up to hundreds of micrometers) on almost any conductive and nonconductive substrate. Lists of plating baths and operating conditions are available in the literature.[28]

X-ray lithography is used with LIGA processes because of its superior performance over optical lithography. X-ray lithography uses a shorter wavelength and a large depth of focus and is well adapted for use with 3-D microstructures for these reasons. Reproducibility is excellent and results are independent of substrate type and surface reflection. The aspect ratio obtainable, defined as the ratio of depth to width of the structure, is typically 100. The major drawback of LIGA is the high cost of the equipment necessary to produce the mold.

A typical LIGA process is shown in Figure 4 as an illustration of a nonsilicon process used to produce MEMS structures.[29] A thick photoresist of up to 1 mm is applied on to a conductive substrate and is then exposed to X-ray radiation from a synchrotron source through a high contrast X-ray mask (Figure 4a). The exposed photoresist is then developed, resulting in the first copy of the high-aspect ratio structure (Figure 4b). The whole structure is electroplated beyond the thickness of the developed photoresist (Figure 4c). After separation from the mask, a complementary metallic mold is formed (Figure 4d). A conductive plate is attached and plastic is injection molded into the metal part (Figure 4e). The metal is separated from the molded structure (Figure 4f). The molded structure is used to electroplate replicas of the original metal part, which in turn can be used to make multiple copies of the injection-molded parts (Figure 4g). The LIGA process is expensive and difficult to develop. However, it should offer the advantage of producing large quantities of parts at low cost. The dimensions of parts obtained by this process in research and development facilities exceed 1 mm with a resolution of a few micrometers. The industrial applications of LIGA techniques have still to be demonstrated.

3.4 Bonding of Microelectromechanical Systems Structures

Most MEMS structures require some type of wafer bonding or cavity sealing, either with glass or with another silicon wafer. We will describe some of the current processes most used in the industry.

3.4.1 Bonding Silicon to Silicon or to Pyrex®

The use of silicon to develop MEMS parts for pneumatic applications requires that these parts be encapsulated. There are several approaches used today, one to bond silicon to silicon and the other to bond silicon to Pyrex®. One technique used is field-assisted thermal bonding.[30]

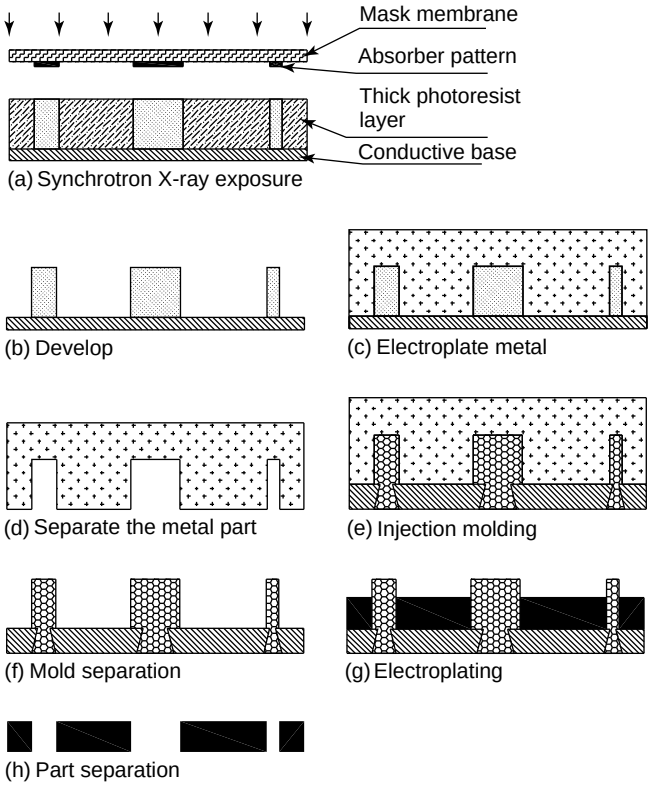

Figure 4 Typical LIGA process. (a) Deposition of a thick layer of photoresist and synchrotron X-ray exposure, (b) develop, (c) electroplating operation, (d) complementary mold, (e) injection molding, (f) mold separation, (g) electroplating, (h) part separation.[29]

A glass slide (Pyrex® #7070 or #7740 from Corning) is placed over a polished wafer, the assembly is heated to about 400 °C and a high voltage (~1000 V) is applied to the glass. For wafers of small dimensions the metal contact may be a single point probe located near the periphery of the wafer. Since the negative electrode is applied to the glass, ionic conduction causes a drift of positive ions away from the glass–silicon interface into the bulk of the glass. The depletion of positive ions at the interface results in a high electric field across the air gap between the two plates. Electrostatic forces effectively clamp the pieces locally, conforming the two surfaces to obtain the irreversible bonding. A variant of the anodic bonding between two silicon wafers can also be achieved by depositing a thin layer of borosilicate glass on the surface of one of the wafers. Bonding happens by applying a dc voltage of approximately 50 V through the substrates at a temperature of 450–550 °C.

Silicon wafers can also be directly bonded to silicon with a technique called silicon fusion bonding. This direct bonding is based on a chemical reaction between OH-groups present on the surface of the native or grown oxides covering the wafers. Wafer bonding may be achieved by placing the surfaces of the two wafers in close contact and inserting them in an oxidizing ambient at a temperature greater than 800 °C. Higher temperatures (above 1000 °C) are usually required to obtain voidless and high strength bonding.[31]

Another attractive technique is eutectic bonding. The Au–Si eutectic bonding takes place at a temperature of 363 °C, which is well below the melting temperature of the individual component of the alloy. The bond is irreversible and strong.[32] There are other eutectic alloys applicable and their choice is determined by the maximum temperature that the device can sustain without damage.

4 EXAMPLE OF PROCESSING OF MICROELECTROMECHANICAL SYSTEMS COMPONENT

The typical process steps of MEMS are shown in Figure 5. The final MEMS structure could be part of an ink-jet nozzle. More examples of this technology can be found in the literature.[13,20,33]

The first step (Figure 5a) consists of growing an oxide layer, usually silicon oxide, on both sides of the silicon wafer. This layer of oxide will act as a masking layer during the etching process. Silicon dioxide is produced by heating the silicon wafer to between 900 and 1150 °C, in a stream of water vapor, or pure oxygen at atmospheric pressure.

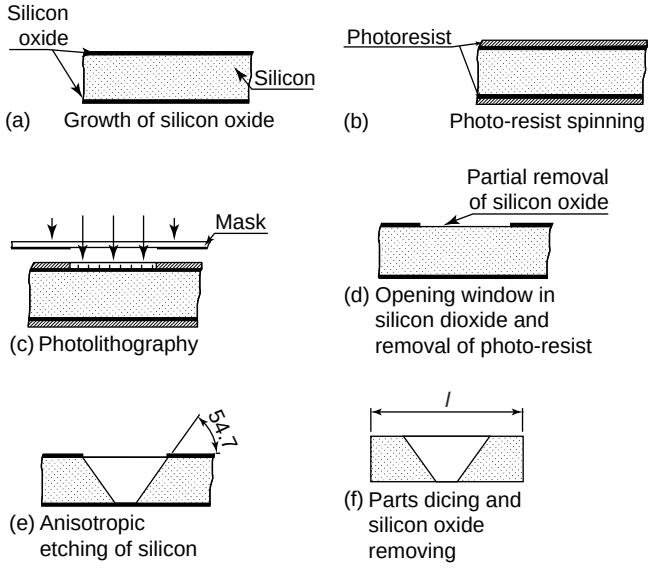

Figure 5 Main steps of a basic MEMS process. (a) Growth of a thin layer of silicon oxide on the silicon wafer (typically 1000 Å). The wafer is heated in a steam of water vapor, (b) photoresist spinning (typically 1 μm), (c) photolithography, (d) opening of a window in the silicon dioxide with acid and removal of photoresist, (e) chemical etching of structure either with acid or with plasma, (f) dicing parts with a diamond saw.

The thickness of the oxide varies between 20 Å (native oxide) to several thousands of angstroms. Other thin layers of material such as silicon nitride can be deposited.

During step 2 (Figure 5b), the photoresist is deposited, usually on to both sides of the wafer. Photoresist chemistry is a science in itself with a multitude of variations to achieve specific results. The process engineer can achieve different results by manipulating variables from the use of either negative or positive resist, light exposures to create over and under development doses, as well as using various resist thickness from microns to millimeters. Usually the photoresist is spun at speeds varying from 1500 to 8000 rpm using centrifugal force to flow the liquid evenly across the surface of the wafer. Acceptable variance in thickness for a 1-μm thick film of photoresist is typically ±10 Å (±0.1%) from wafer to wafer. Once the photoresist has been applied, the wafer is ''soft-baked'' at a temperature between 75 and 100 °C for approximately 10 min. This prebake process drives off some of the solvents and stress relieves the wafer. It also promotes adhesion of the photoresist to the wafer.

Photolithography is done during step 3 (Figure 5c). After the soft-bake process, the resist-coated wafer is placed onto a projection aligner. Using the wafer ''flat'' to assure proper positioning, the wafer is precisely aligned to within a few micrometers of the mask. The UV lamp illuminates, passing through the unpatterned areas of the photomask and exposing the photoresist film on the wafer. The pattern on the photomask is transferred to the photoresist, changing the molecular structure of the polymeric material in those areas exposed to the light. The areas that have been exposed will now react with the development solvent. These areas will soften and dissolve in the development solvent, opening the designated region down to the surface of the wafer. The unexposed portion of the wafer will remain solid to protect the wafer during either the etching process or adding of material. The wafer is ''hard baked'' to drive off more solvents and increase the hardness of the photoresist. The patterned wafer is then developed with solvents exposing the oxide layer.

The silicon dioxide is etched and the residual photoresist removed during step 4 (Figure 5d). The oxide on the surface of the wafer is usually etched with acid only at the location exposed during the previous step. All the other locations are still protected by hard photoresist. Now that the windows on the oxide layer have been opened, the remaining photoresist needs to be removed by using either wet or dry processing called ''stripping''.

There are several ways to strip photoresist from the wafer using commercial resist strippers. H_2SO_4 (sulfuric acid) based products are commonly used to wet strip the photoresist from the wafer without attacking the wafer's surface, while dry stripping is an oxygen plasma process. ''Ashing'' the wafer after the resist strip helps to remove any remaining photoresist residue. This process is accomplished by putting the wafer into a plasma chamber and exposing the wafer to a mild oxygen plasma treatment. The dry stripping process removes all the remaining photoresist on the wafer, leaving the pattern formed during processing. This dry stripping process has become more popular, as it poses fewer disposal problems with toxic, flammable, and dangerous chemicals.

At this point, the silicon wafer is ready for etching.

Step 5 (Figure 5e) shows an example of anisotropic etch on ⟨100⟩ surface orientation. In this example the hole does not go through the silicon.

If there are several masks in the process, steps 2 to 5 are repeated to create the final device.

Step 6 (Figure 5f), the final step, consists of cutting the wafer into individual components called ''die'' by using a diamond bladed dicing saw. The MEMS part is ready to be packaged.

The above process is typical for developing patterns on the surface of a silicon wafer. However, other processes and materials are being used to develop precision structures in plastics, stainless steel, copper, or any material that can be etched with a solution or plasma process. Photolithography enables the creation of multiple small features within a device and repeat this pattern across the surface of the material of interest. The reason silicon wafers are most widely used is because the expertise and documented history created over many years make them a known quantity in the design of new devices. Complicated structures can be designed more easily when using known processes that create known results. However, in the MEMS world, there are many researchers developing new process methods for various materials. As an example we should mention ''Lab-on-a-chip'' technology developed for biological instrumentation which relies on microfluidics advanced sensors.[34] Microfluidics chip technology enables an increase in the number of chemical reactions that can be performed concurrently. They are expected to affect optimization of reactions in combinatorial chemistry, drug screening, and polymerase chain reaction (PCR) DNA amplification where speed, precision and cost of reagents are important. The development of a ''Lab-on-a-chip'' relies on the use of miniaturization technology in plastic parts which is equivalent to what has been done with silicon for parts.

Another emerging MEMS technology used to form devices is a German process called ''LIGA''. This process has been described in section 3. In addition, there are other micromachined techniques that have demonstrated usefulness for certain specific applications: deep reactive-ion etching (DRIE) of silicon, microelectrodischarge machining (μEDM),[35] laser micromachining,[36] and polymer stereolithography.[37]

How is MEMS different from semiconductor processing? Semiconductor fabrication is a two-dimensional process (surface processing) that removes and replaces precisely patterned materials on an X and Y planar surface. MEMS is a 3-D process (bulk processing) that removes and replaces precisely patterned materials, but also creates Z axis etching to create bridges and suspended beams by using undercut or etching techniques. Not only can MEMS technology create 3-D structures, it can also use the integration of semiconductor technology to create smart mechanical devices. Complementary metal oxide semiconductor (CMOS) technology can add electronic capability to a MEMS mechanical structure to create a new dimension of device development. The pervasiveness of automotive passive restraint systems has emphasized the need for improving system reliability while simultaneously reducing the cost and size of the system. An accelerometer that utilizes silicon micromachining and a CMOS process, along with low cost plastic packaging, supports the requirements of the next generation of automotive passive restraint systems.[38]

5 APPLICATIONS OF MICROELECTROMECHANICAL SYSTEMS TO PORTABLE INSTRUMENTATION

During the 1990s there have been numerous attempts in the military and in the private sector to develop microstructures based on MEMS for field applications. One of the most obvious reasons is to reduce the size and weight of these microstructures and to build truly portable systems. The results of space exploration, for instance, have emphasized that cost effectiveness and highly focused mission objectives were necessary to increase the number of missions. MEMS is one of the promising key enabling technologies to create cost-effective, miniaturized, robust, and functionally focused spacecraft for both robotic and human exploration programs. Just to cite a few of them, examples of MEMS devices under development include microgyroscopes, microseismometers, microhygrometers, quadrupole mass spectrometers (MSs), and micropropulsion engines.[29,39,40]

Analytical chemistry instrumentation is also on the verge of a revolution. More sensitive detection techniques permit smaller sample sizes and when combined with new concepts based on micromachining and microfluidics, facilitate the control and handling of gases and fluid samples at low cost with batch processing of parts (Figure 6). All these factors are contributing to the renewed interest in and vast potential impact of microinstrumentation. Microsensor technology has paved the way for a new generation of vastly more powerful microsensing systems.

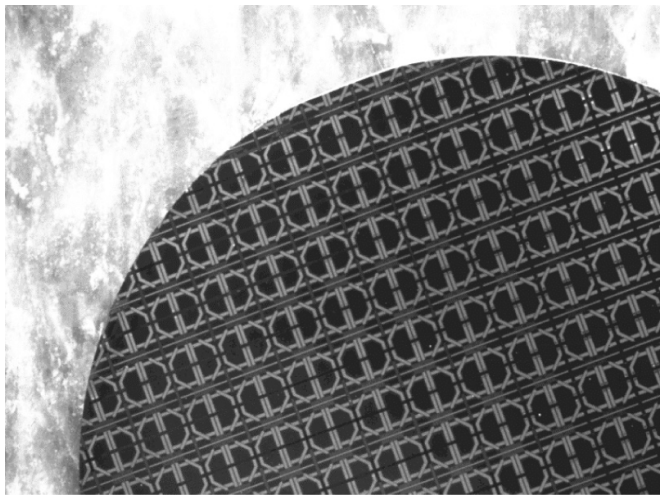

Figure 6 Example of batch processing of 4-inch wafers yielding a large number of MEMS parts. (Reproduced by courtesy of Hewlett-Packard Company.)

There are many examples of successful applications of MEMS to microsensors as we will see in the next section on gas sensors. These microsensors and microinstruments, are already here and are going to change the way we think about chemical analysis and chemical analysis laboratories.[41–43] In the section on portable GCs we will give a detailed description of an instrument based on MEMS.

Although beyond the scope of this article we should also mention that miniaturization is not only for gas phase analysis but is also applicable to liquid phase analysis and biochemical measurements. Miniature biomedical instruments include moving sample, wash, and calibration fluid through small pneumatic connections, controlling the temperature of the fluid, getting the sample into and out of the instrument, providing fresh viable reagents, and detection. All these operations are performed on a lab-on-the-chip and have to be done at low cost. Most of the technology developed to-date has been around the use of plastic injection and hot stamping[44] rather than silicon micromachining.

Optical sensors are also an important part of the industry that MEMS have transformed. New concepts and practical issues for accomplishing complete miniature sensor modules for chemical and biochemical applications, which are at the same time small, highly sensitive and stable, are being developed. The novelty of the approach is mainly based on using different types of smart planar optical transducer (SPOT) chips, which not only carry out the task of conventional integrated optical transducers, but also perform additional functions for which peripheral equipment is otherwise needed. The emphasis is on pointing out and discussing the most important aspects for refractometric modules. Different approaches

and sensor types are compared with respect to fulfilling the requirements for practical applications. New results concerning the use of multilayer waveguides and the implementation of novel modules are possible.

Whatever the type of application there are several important key issues which have to be resolved to guarantee the success of the idea:[45]

- importance of a proper theoretical modeling of the sensor chips
- availability of microfabrication technologies suitable for realizing sensors based on a variety of substrate materials, such as glasses, polymers and semiconductors
- fluidic interfacing of the MEMS parts is probably the most critical issue to be solved when going to production and often can make or break a project
- possibility of mass-producing low-cost chips.

In the following we will limit our discussion to the applications of MEMS that apply to the analysis of gases and vapors using gas sensors and gas chromatography.

5.1 Gas Sensors

The detection of gases with chemical sensors based on MEMS technology is a very dynamic market. According to a recent market survey published by Frost & Sullivan[46] the industrial gas sensor market for the USA was $329 million in 1992, advancing to $427 million in 1994. It was expected that the steady growth should continue, propelling the market size to $1.03 billion in 2000. The compound annual growth rate for the period 1994 to 2000 is forecast to be 15.9%. These estimates and forecasts apply to the basic devices offered by manufacturers of gas sensors and often include a limited amount of integral electronics included in the sensor package.

The European industrial gas sensors market increased from $141.7 million in revenue in 1994 to $154.3 million in 1997. Growth of both unit shipments and revenues were witnessed over this time following the economic improvements in Europe and advancements in the manufacturing and performance of gas sensors. Continued growth of the market is expected in the forthcoming years supported by low, but increasing prices and increasing applications. Frost & Sullivan expects the total European industrial gas sensors market to increase to $221.0 million in 2004, from almost 3 million units shipped. The main source of new applications for gas sensors is expected to be new European legislation governing some industries. There will be several other factors driving the market in the coming years.

One factor contributing to the worldwide growth is the continuing increase in government regulations concerning personal safety, confined spaces in the workplace, continuous emissions, ambient air, and the environment. Even where government regulations do not apply, public awareness about these considerations is increasing.

Other factors have played, and will continue to play, an important part in driving the gas sensor market. Advances in technology, together with competition among the various approaches to gas sensing, have provided a hospitable environment for increasing use of sensors. Since more than one approach is available to detect many industrial gases, competition among these approaches forces manufacturers to enhance the performance of their products. The lack of selectivity of solid state sensors, in general, is making for a very fast growing market for the more selective but more expensive wet electrochemical gas sensors and the newly emerging inexpensive optochemical techniques,[47] but the subject is out of the scope of this article about MEMS.

A family of sensors based on surface loading which are attractive for environmental measurements are the surface acoustic wave sensors (SAWs). SAWs are MEMS structures which have been under intense investigation during the 1990s as a basis for chemical gas sensors. These are resonators, reflective delay lines, and dispersive structures containing two interdigitated transducers coated with a polymer film (Figure 7). Application of an alternating voltage to the input transducer generates a strain field in the underlying quartz caused by its piezoelectric properties. The strain field launches a surface acoustic wave that travels along the substrate, interacting with the overlayer film, before being converted back into an

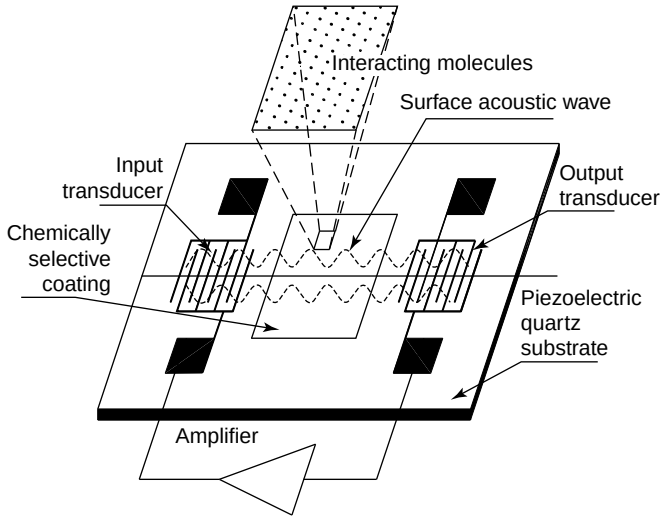

Figure 7 Schematic view of a SAW. Two interdigitated transducers are on the top of quartz substrate and are coated with an absorptive polymer film. The signal transmission is delayed by the adsorption of gas molecules on the surface and the corresponding signal is proportional to the concentration of the gas.[48]

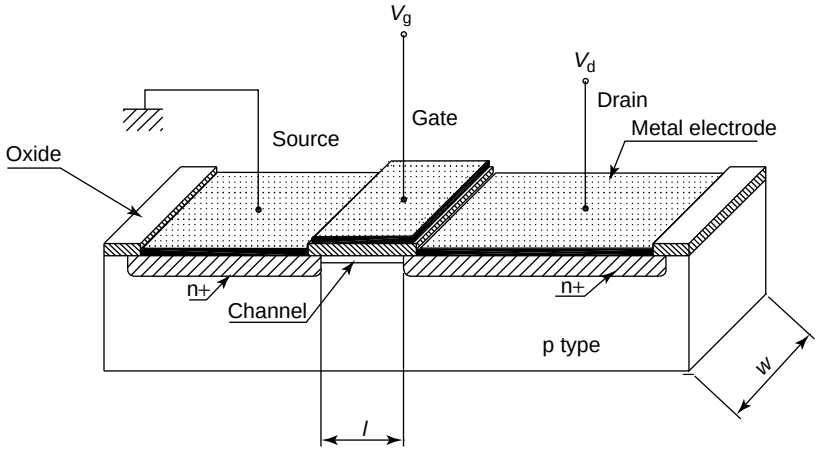

Figure 8 Schematic diagram of a MOS transistor used for the detection of volatile organic vapors. The presence of volatile organic compounds (VOCs) change the characteristics of the transistor. V_g is the gate voltage, V_d is the applied junction bias, w is the width and l is the length of the channel.

electrical signal by the output transducer. The wave/film chemical or physical interaction with the molecules of gas to be detected results in a perturbation of the wave propagation properties, specifically the wave velocity, v and the wave attenuation α (the rate of wave diminution with the distance). This sensitivity to thin film properties makes SAW devices well suited for chemical sensors to monitor gas and vapor species with a typical sensitivity of a few parts per million.

SAWs are also sensitive to a broad range of perturbations in addition to surface mass loading. Changes in the electrical, mechanical, or rheological properties of a thin film or other medium in contact with the device surface can induce a response. Extrinsic variables such as temperature, pressure, and electric field can perturb SAWs as well. Such multiparameter sensitivity is a double-edged sword, while providing a multitude of physical mechanisms to probe interfacial chemical interactions, it also creates responses to unintended perturbations, with confusing or misleading results.

SAW chemical sensors offer considerable promise for the future, particularly for those applications where their sensitivity to surface mass changes, or the ability to provide two-dimensional (velocity and attenuation) responses to multiple perturbations, offers a competitive edge that outweighs the costs of making measurements at radiofrequencies. SAW radio sensors make it possible to read measurement values from a remote location[49] using a radio emitter. The decisive advantage of these SAW sensors lies in their passive operation with simple radiofrequency electronics, and in the possibility of wireless installation at particularly inaccessible locations. Despite a fairly good understanding of the fundamentals of SAW chemical sensors, they have yet to be widely applied to commercial problems.[50]

A sensor for the detection of gases needs to be able to produce an electrical response which is dependent on the concentration of gaseous species present. One type of detector which satisfies this criterion and is also capable of being integrated into the associated electronics is the gas-sensitive metal oxide semiconductor (MOS) transistor. This transistor is a device in which the conduction between the source and drain is modified by a charge or potential on the gate electrode (Figure 8). Basically, the device is a capacitor with one electrode being the gate and the other the channel. The conduction of the channel is controlled by the charge present on the gate. Hence anything which alters either the potential of the gate or the charge at the gate will alter the conduction of the channel and this change can be detected as an electrical signal which can then be processed. Gases absorbed or adsorbed onto the gate change either the potential or the charge on the gate. This can occur by a number of different mechanisms. Either the gas can chemically interact with the gate material and change the carrier concentration of the gate and hence the Fermi energy, or else the gas can induce dipoles in the gate which will alter the potential gradients at the gate–oxide interface, or else a chemical cell can be formed between the gate and the oxide. All these mechanisms will alter the potential or barrier height at the gate–insulator interface and hence modify the channel conduction.

Catalysis gas sensors form another type of semiconductor used for vapor and gas detection. They are characterized by reduced power consumption as low as 100 mW, and are well suited to portable gas detection devices. Sensors with a base material of tin dioxide and 4 wt % of Pd dopant are fabricated by means of thick film technology. Both the composition parameters

of the sensing layer and the heater are optimized in order to sense propane/butane in air with the threshold level of 100 ppm. The original method of tin dioxide preparation for the sensing layer and for the thick film heater being stable at temperatures up to 600 °C has been demonstrated.[51]

Cyrano Sciences Inc. (Pasadena, CA) is developing an array sensor called ''electronic'' nose to be released late 1999. The hand-held system is lightweight, hand portable and operates on batteries (Figure 9).

The sensing mechanism is based on the use of a carbon loaded polymer that expands like sponge when it comes in contact with a vapor, increasing the resistance of the composite. An array of 32 channels with different responses produces a distinct response signature. This unique response results in an electrical fingerprint that is used to characterize any odor. The typical response time of the array is 10 s.

Figure 9 Portable ''electronic nose'' for use in the field to detect volatile compounds. The device incorporates a LCD (liquid crystal display) and a flexible sniffing tube. Rapid response time provides an effective and accurate measure of vapors present. (Reproduced by courtesy of Cyrano Sciences, Inc.)

5.2 Portable Gas Chromatographs

The conventional approach to off-site gas analysis measurement relies on the collection of the samples in the field and measurements in the laboratory. Despite its apparent simplicity, this approach presents numerous drawbacks. The process of collecting a gas sample in a vessel and transporting it to the laboratory for analysis delays the results and quite often causes the sample to degrade, making the analysis inaccurate or not representative of the mixture from which the sample had been taken. Portable instrumentation, on the other hand, allows on-site measurements directly at the sampling source and thus provides limited risk of sample degradation, better accuracy, potentially greater sample throughput and almost immediate results more economically than with traditional technology.

What is the analytical price to pay in order to gain portability? Many conventional analytical instruments operate at or near their theoretical limits and, before reducing the size of the instrument, it is necessary to understand the effects of the ''scaling factor'' on the performance of the instrument. For instance, if the speed of a GC analysis is to be reduced from minutes to seconds, the internal volume of the detector must be scaled down proportionally, but the sensitivity can suffer. The next question regarding size reduction is the access to the right technology, but research in instrument miniaturization has demonstrated that the development of portable GCs based on MEMS is possible. This interrelationship of MEMS components is what led to the ''GC-on-a-chip'' designed in the late 1970s/early 1980s at Stanford University[1,2] and its spin-off the P-200 GC developed by Microsensor Technology Inc. (MTI) now commercialized by Hewlett-Packard (Figure 10).

The GC-on-a-chip, as it was then called, was constructed on a 2″-(5 cm) wafer with a miniature sample injector, a microcolumn and a microthermal conductivity detector (MTCD). Micrometer-dimensioned features were created by using IC process techniques such as photolithography to pattern, and chemical etching technologies to create precision shaped structures. These 3-D structures were then bonded into place and connected to several other microstructures to form at that time the world's smallest and fastest GC. The sample injector incorporated a 10-μm wide sample loop and injected the sample composition onto a 1.5-m long capillary column, wound into a shallow groove of approximately 200-μm wide and 40-μm deep coated with a thin film of liquid phase. The column was wound into a spiral to enable its placement on the wafer. Once separated, the sample elements were analyzed by a microfabricated TCD.[52] The chip was fabricated on a separate wafer and the thin-film sensor deposited on the top of a thin layer of sputtered

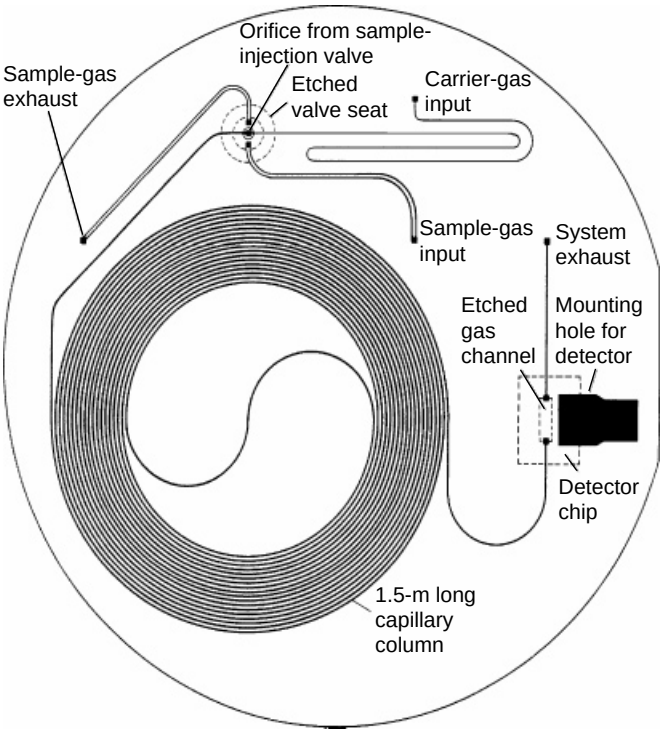

Figure 10 A later version of the GC-on-a-chip developed at Stanford University.[1] The silicon GC is a complete analytical instrument built on a 5-cm wafer. The capillary column of the GC is a spiral 1.5-m long. Etching a groove in a wafer and bonding the wafer to glass forms the column. The thermal conductivity detector (TCD) and the gas-sampling valve are external to this system and not shown on this figure.

Pyrex®. The chip was mechanically clamped on the end of the column. The interface between the sample injector, the GC column and the dual-detector cavity were all silicon micromachined. A Pyrex® cover plate, electrostatically bonded to the silicon substrate, ensured that the GC structure was leakproof. The GC-on-a-chip could separate a simple gas mixture composed of nitrogen, pentane, and hexane in less than 5 s.[2]

Before microbore fused-silica capillary columns first entered the market, the Stanford team made an attempt to etch the column directly on a silicon wafer. These micromachined GC devices were the only available components that could provide fast analysis. High-speed chemical separations (analyses requiring less than 120 s) were eventually realized with the interfacing of silicon micromachined components and later with commercially available quartz microbore capillary columns. By the early 1990s, refinements were made to improve manufacturability and reduce costs and MTI began to produce both portable and laboratory Micro GC instruments.

The principal market for these small high-speed instruments was for the analysis of ambient VOCs. Analysis times were typically 60 s or less depending on the composition being analyzed. These instruments were focused on three major markets: (1) chemical/petrochemical, (2) natural gas and (3) fixed gases. Using the same modular construction, the Micro GC could be configured to have up to four (Quad) GC modules in a single instrument. Since each module had its own injector, heated microbore column, and TCD, the instrument could perform up to four separate but parallel analyses, permitting complete analysis of complex compositions in an extremely short period of time.

This portable GC combines the varied aspects of silicon micromachining and conventional technologies to fabricate a gas injector and a TCD.[53] The gas injector for instance is a hybrid approach of micromachining technology and conventional machining,[42] and provides a sharp plug injection function necessary to avoid peak broadening with small bore columns.

Figure 11 shows quantitative results of a comparison between the Quad series refinery gas analyzer (RGA) and a conventional laboratory instrument. The test was done by measuring the relative standard deviation of the retention times for the various components of a refinery gas sample. The standard deviation for each of the peaks is smaller for the portable instrument than for the conventional GC and shows that miniaturization does not affect the performance of the portable instrument.

A typical Micro GC instrument incorporates up to four GC modules. Each module consists of a silicon micromachined gas injection system (*Micro Injector*), a four filament silicon micromachined TCD (*Micro TCD*), and a heated microbore capillary column connecting the injector to the detector. These devices are encapsulated between two pieces of thermal insulating foam and packaged in a $4 \times 4 \times 6$ inch ($100 \times 100 \times 150$ mm) metal enclosure. In addition to the modules, the instrument also contains a vacuum pump and a pressure-regulating manifold for each module. The portable series also includes a 0.3-L carrier gas cylinder and a 12-V dc battery power system for field use.

Figure 12 is an open view of the MTI-P series Portable GC showing all the elements of the GC. The sample inlet, at the front of the instrument, is heated to 110 °C to prevent water condensation. A gaseous sample is introduced to the instrument through the sample inlet port(s) using the built-in vacuum pump and is distributed through the separate GC modules simultaneously.

5.2.1 Micro Injector

The *Micro Injector* is constructed of four laminated layers of materials which form a single unified functional component that measures approximately $1 \times 1\frac{3}{16} \times \frac{1}{8}$ inch ($25 \times 30 \times 3$ mm) thick. Part of the injector assembly (Figure 13) controls the microfluidic movement of the

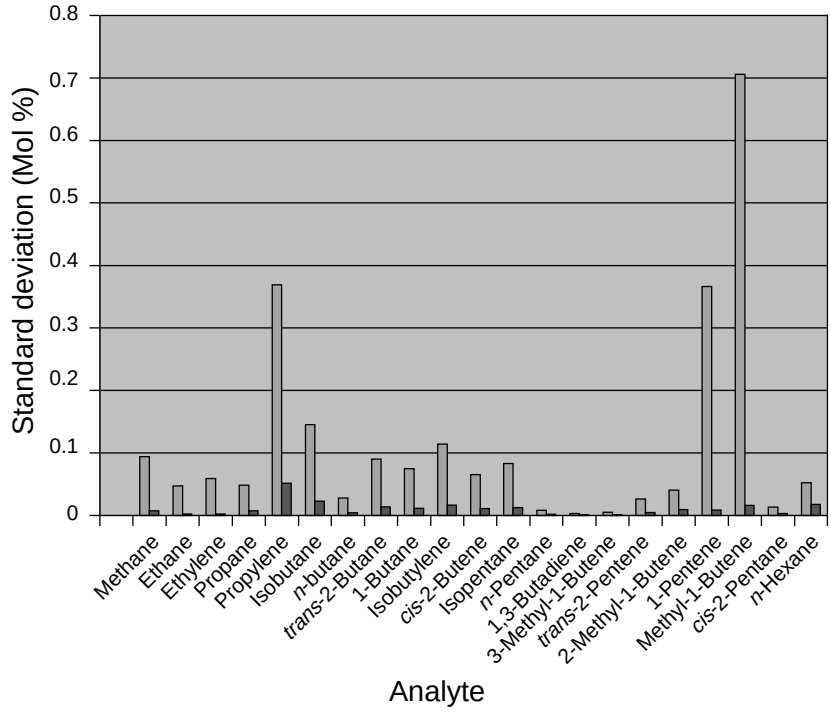

Figure 11 Bar graph showing the relative standard deviation of retention times measured on a conventional GC (in gray) and on the Micro-GC, P-200 series (in black), for a standard mixture of components of a plant stream. The increased speed of the analysis performed by P-200 Micro-GC does not show any degradation of the results. (Reproduced by courtesy of Hewlett-Packard Company.)

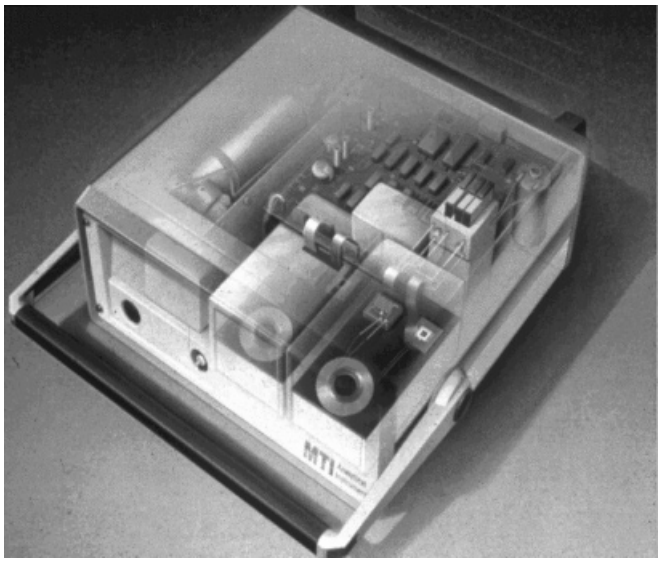

Figure 12 Open view of the portable dual micro GC developed by MTI. The illustration shows a P-200 GC with two GC modules in the lower right corner of the instrument. The GC modules contain an injector, a heated capillary microbore column and a MTCD encapsulated in high-temperature insulation. (Reproduced by courtesy of Hewlett-Packard Company.)

Figure 13 The *Micro Injector* precisely controls the fluidic movement of carrier and sample gas throughout the Micro GC. The 32-μL serpentine sample loop supplies the column with preselected sample volume of 1–10 μL for separation. The body of the Injector consists of two Pyrex® layers, a Kapton layer and an etched silicon layer sandwiched together as a unit. (Reproduced by courtesy of Hewlett-Packard Company.)

gases throughout the complete analytical process. While in the standby mode, the injector receives and stores

the sample gas in its serpentine sample loop. With precise micromachined valves, it injects a predetermined amount of sample gas into the carrier gas stream of the GC. The MEMS etched silicon layer of the injector includes normally open and normally closed valves, a flow restrictor channel, and several sizes of gas channels linking the various mentioned devices together. The injected quantity of gaseous sample is variable up to 31 μL.

This MEMS design of the injector has several distinct benefits:

- small sample paths introducing little band broadening
- high speed of operation necessary for fast analysis
- four valves integrated in the injector that can be operated independently.

The flow-through sample valve is shown in Figure 14 whereby the sample gas enters the valve body from one of the outer ports, flows around the sealed center port, and exits the opposite outer port to vent. When the diaphragm lifts, opening the center port, the sample gas is pushed through the center port and onto the column for separation. Precise opening time and constant inlet pressure enable the GC to repeat accurate sample injections every time.

The normally open/normally closed valves are of the same construction except there is only one outer port and the center port is either normally opened or closed. A separate pilot solenoid valve pneumatically pressurizes the upper half of the diaphragm, forcing it down onto the valve seat and sealing the gas flow. The diameter of the valve body is 2500 μm with a 300 μm valve seat for a total dead volume of less than 10 nL. Opening and closing of the valve occurs in less than 5 ms with the 15-μm movement of the diaphragm. With this type of precise control, these microvalves can introduce accurate amounts of sample gas onto the column with extremely high levels of repeatability and accuracy. The length of time the valve stays open determines the precise amount of sample volume that is pushed onto the column for gas separation. The valves can be operated above 100 °C for millions of cycles without any signs of wear or fatigue.

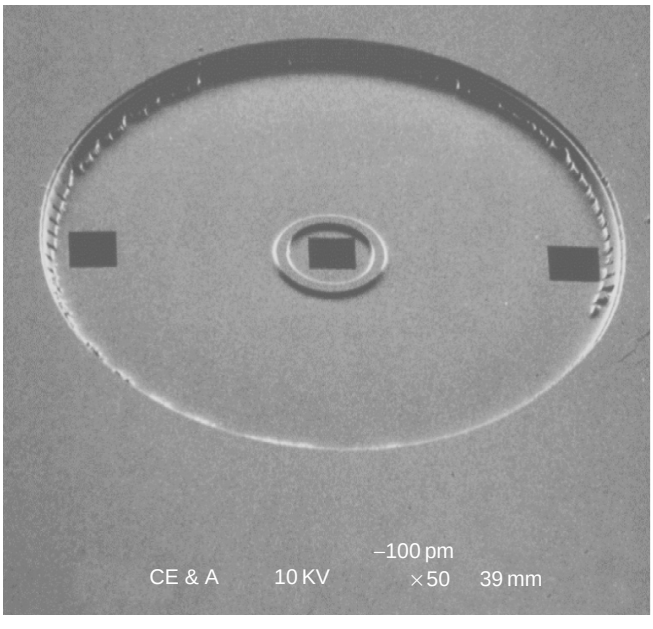

Figure 14 Details of a valve showing the sealing ring etched in silicon. The P-200 has several valves based on the same design. The typical dimensions of these MEMS parts are less than 1 mm. The center port is sealed with a Kapton diaphragm actuated by an external pilot valve. The dead volume of these valves is less than 10 nL. (Reproduced by courtesy of Hewlett-Packard Company.)

Valves are an important part of MEMS used for the control of fluids in analytical instruments and the hybrid approach chosen for the valve here is specific to the design of this instrument. There are several other micromachined valve designs besides pneumatic driven valves using separate solenoid pilot valves. Some combine the valve actuator with the microvalve design based on a variety of principles including electrostatic force,[54] liquid expansion,[55] bimetallic expansion,[56] thermal expansion[57] and, for liquid control applications, piezoelectric actuation.[58] An interesting application of MEMS components applied to gas control is a proportional microvalve developed by Hewlett-Packard Laboratories.[59] This valve is thermally actuated by a metallic actuator and is totally integrated, meaning that the actuator and the body of the valve are part of the same assembly. This mode of actuation precludes operations at elevated temperatures, but offers the advantage of proportional flow control of gases starting as low as a fraction of a microliter per minute up to liters per minute at pressure up to 200 psi (1.3×10^6 Nm^{-2}).[60]

5.2.2 *Microthermal Conductivity Detector* (MicroTCD)

The *MicroTCD* used in the HP Micro GC is another example showing that MEMS technology is very well adapted to solve analytical problems. Small-bore columns are used to perform fast analyses, but when the internal diameter (ID) of the column decreases, peak width increases when all other instrumental parameters are unchanged. To avoid broadening peaks in the chromatogram, the internal volume of the detector should be of the same order as the ID of the column. MEMS technology is particularly well adapted to solve this type of problem and gives the designer the ability to fabricate 3-D structures with small internal volumes. The *MicroTCD*, for instance, has an internal volume of 180 nL, which is perfect for microbore columns. Higher sensitivity is an inherent characteristic of concentration in microdetectors

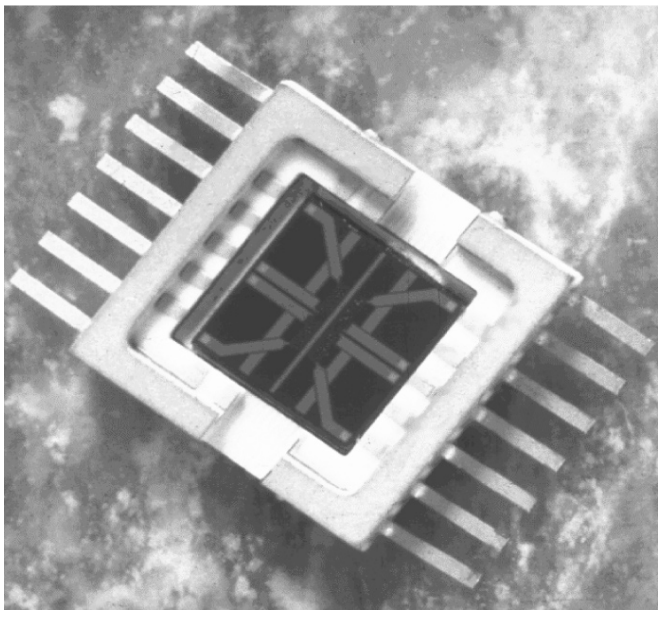

Figure 15 MTCD used in the P-200 GC. This detector has two filaments per channel. The two channels lie on each side of the center single gold line. The electrical connections for each filament can be seen coming from the corner bonding pad to the channel. The pneumatic connection (not shown) enters at one end of the channel and flows across the filament exiting at the other end of the channel. (Reproduced by courtesy of Hewlett-Packard Company.)

over mass detectors when the internal volume is decreased. Van Es from the University of Eindhoven, for instance, has shown that the sensitivity of the MTCD for n-C7 becomes comparable to the flame ionization detector when columns of 100 μm or less ID are used.[61]

The *MicroTCD* is shown in Figure 15. It has two parallel pneumatic channels (reference and detection) running through the center of the silicon die, with two nickel filaments suspended over each channel. The four 0.8-μm thick nickel filaments are electrically connected to form a Wheatstone bridge resistor circuit. These filaments are the sensing portion of the detector and because of their small thermal inertia can be sampled 100 times per second. These nickel filaments are suspended across a 250-μm wide by 50-μm deep trapezoid-shaped channel and are located in the center of each channel that was formed during the anisotropic etching part of the micromachining process. The other half of the two channels is machined into the Pyrex® cover plate, aligned and anodically bonded to the silicon detector die. The anodic bond seals the two parts together, completing the formation of both channels with two filaments suspended in the middle of each channel.

A magnified view of a detector filament cross-section is shown in Figure 16. The filament suspended above the silicon channel is entirely supported by a silicon dioxide

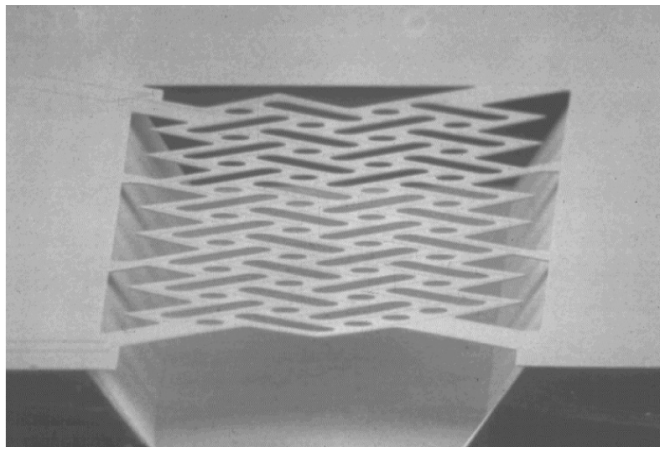

Figure 16 The TCD has four filaments of which one is shown in a cross-sectional view. Sputtered onto the silicon dioxide membrane suspension is a layer of 0.8-μm thick nickel to form a resistor to measure the thermal absorption of the gases that flow through the 50 × 250 μm wide gas channel. (Reproduced by courtesy of Hewlett-Packard Company.)

membrane. This filament is made of an evaporated 0.8 μm thick nickel layer, sealed with a thin film of silicon nitride for corrosion protection. The thermal conductivity changes in the environment are detected by measuring the changes in current being applied to the filaments in order to maintain a constant filament temperature. The higher the difference in thermal conductivity, the greater the signal.

5.2.3 Microbore Capillary Column

A major trend of GC has been to reduce analysis times by using small bore columns to provide near-real-time monitoring. Separation times in high-speed gas chromatography (HSGC) are in the range of seconds rather than minutes and such speed can only be achieved by using small bore columns and limiting peak broadening. Practical limits for fast analysis using currently available instrumentation are understood.[62] Some of the working solutions consist of getting a sharp sample introduction plug inside the instrument, programming the column temperature at high rates and reducing internal dead volumes.[63,64] The early attempt by the Stanford team to use a MEMS approach to integrate an injector, a column and a detector to a silicon wafer was an excellent approach to eliminating the causes of peak broadening due to the interconnection of the components, and at the same time produced etched channels of small dimensions to provide fast separation. They were successful in reducing internal volumes, but overall provided a GC system with poor analytical performance. The main reason is that the cross-section of an etched column is far from being circular (actually,

it is closer to being a semicircle when the channel is isotropically etched) and therefore the thickness of the liquid film inside the column is not constant, thus creating a broadening of the peaks and resulting in a low column efficiency. For practical reasons, the length of a column that can be etched in a silicon wafer can hardly exceed a few centimeters and therefore limits the analytical separation to only a very simple analysis. For these reasons, it was decided to abandon the etched column and choose conventional small bore capillary columns.

Several papers in the literature discuss the use of narrow-bore wall coated open tubular (WCOT) columns for high-speed capillary GC.[63–67] The main conclusion is that the ID of the column should be reduced to at least 0.32 mm but not less than 0.1 mm to achieve a fast analysis without significant loss of resolution. The most serious problems to overcome to reduce even further the ID of the column below 0.1 mm are the difficulty in dealing with the limited sample capacity of the column and with the practical reduction of the internal pneumatic volumes in series with the GC system.

A typical microbore capillary column used in Micro-GC is shown in Figure 17. These columns are made of quartz glass, have ID tubing between 0.1 mm and 0.53 mm, up to 50 m long. The exterior is coated with a thin film of polymeric polyimide to protect the glass tube from fracturing. The internal wall of the tube is coated with a thin layer of polar and nonpolar material or solid adsorbent formulated to separate specific gases as they flow through the column.

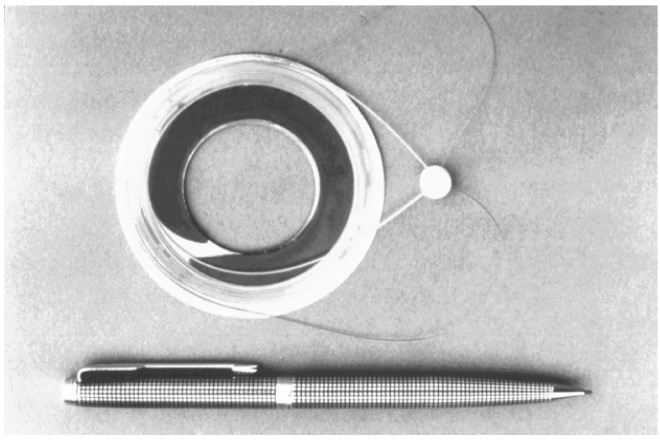

Figure 17 The microbore capillary column is a quartz tube coated with a thin film of polymeric material called polyimide to protect it from breaking. The column spool is configured with a thin film resistive heater and two sets of tubing to carry the sample and carrier gas. The sample column is an 8-m × 0.32-mm tube internally lined with a phase material called OV-1. The phase material is specifically selected to separate a particular combination of hydrocarbon gases. (Reproduced by courtesy of Hewlett-Packard Company.)

5.2.4 Field Applications of a Portable Gas Chromatograph, Micro Gas Chromatograph from Hewlett-Packard

The P series Micro GC from Hewlett-Packard is an illustration of how MEMS technology can elegantly solve some old analytical problems.

5.2.4.1 Natural Gas Analysis A typical natural gas analysis performed off-site consists of collecting the sample and bringing it to a laboratory to be analyzed for the British thermal unit (BTU) value by volume. This is slow, time-consuming and very often inaccurate. The Micro GC ensures accurate and rapid data analysis directly at the well head site. The overall analysis is done directly in the field with an accuracy of ±2% relative standard deviation ±1 BTU in 1000.

The Micro GC also performs extended analysis of natural gas samples. Extended analysis is the separation of natural gas without the grouping of C_6+ (hexane and beyond). The BTU content of a sample is calculated from the individual BTU values of C_6(hexane), C_7(heptane), C_8 (octane) and C_9 (nonane), instead of using an average C_6+ value (hexane, heptane and octane). This analysis generally takes 35 min with a conventional GC but is performed in less than 140 s with the P-200 series Micro GC.

The results of the analysis with a P-200 equipped with two analytical modules are shown on Figure 18. The first module contains an OV-1 (8 m × 0.15-mm ID) micropacked column for baseline separation of nitrogen, methane, carbon dioxide, ethane and propane. The complete separation of these samples is done in less than 80 s (Figure 18). The second module contains a PLOT U (8 m × 0.320-mm ID) capillary column to separate isobutane and normal butane, isopentane and normal pentane, hexane, heptane, and octane. The analysis of these components is done in less than 130 s.

5.2.4.2 Analysis of Landfill Gas Landfill gas generates unpleasant odors, greenhouse gases and VOCs as well as by-products, which can cause explosions and fires, and kill vegetation. Consequently, regulatory agencies are imposing increasingly stringent requirements on the amount of landfill gas that can be released into the environment, often requiring collection systems (wells) to be installed.

A landfill gas analyzer separates and analyzes H_2, CO, C_1 (CH_4-methane), O_2 and measures the anaerobic efficiency of the bacteria in the landfill. Two modules are used for this determination. The analysis of the permanent gases H_2, O_2, N_2, CO and CH_4 is done in less than 45 s using a 10 m, 0.32-mm ID, 30 μm d_f, microbore molecular sieve 5A PLOT column (Figure 19a, Channel A). The fixed gases and light hydrocarbons are separated using a 4 m, 0.32-mm ID PoraPLOT-U, 10 μm d_f column to

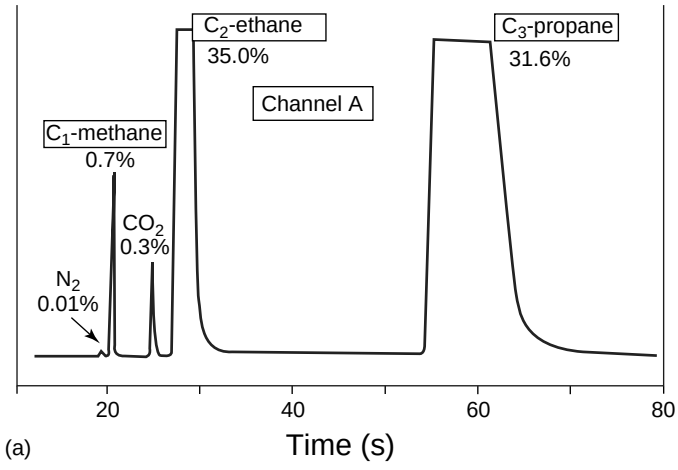

(a)

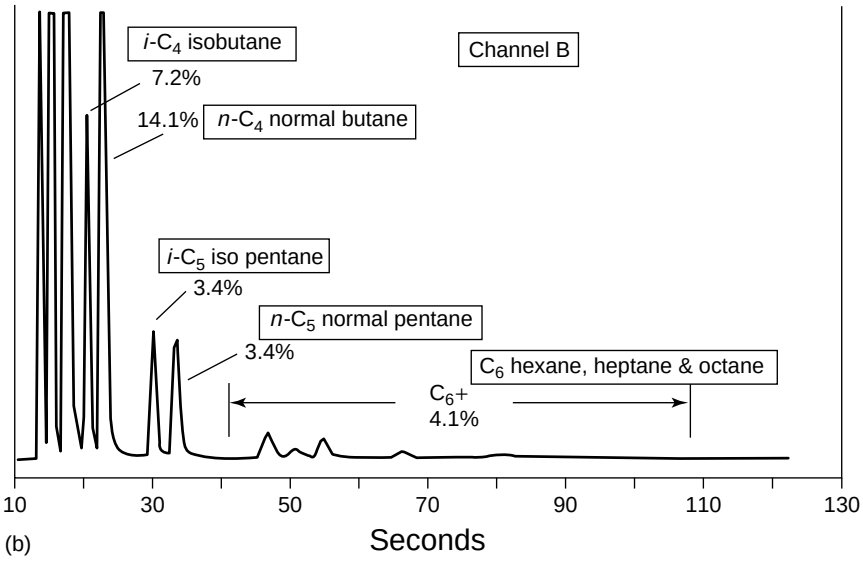

(b)

(c)

Figure 18 Field application of the P-200 GC analyzing 1100 BTU natural gas. Channel A of the instrument is equipped with an OV-1 column and separates C1 through C3 in 80 s (a). Channel B is using a PLOT U column and separates components from C4 through C8 in 130 s (b). The operator performing the separation in the field is shown in (c). (Reproduced by courtesy of Hewlett-Packard Company.)

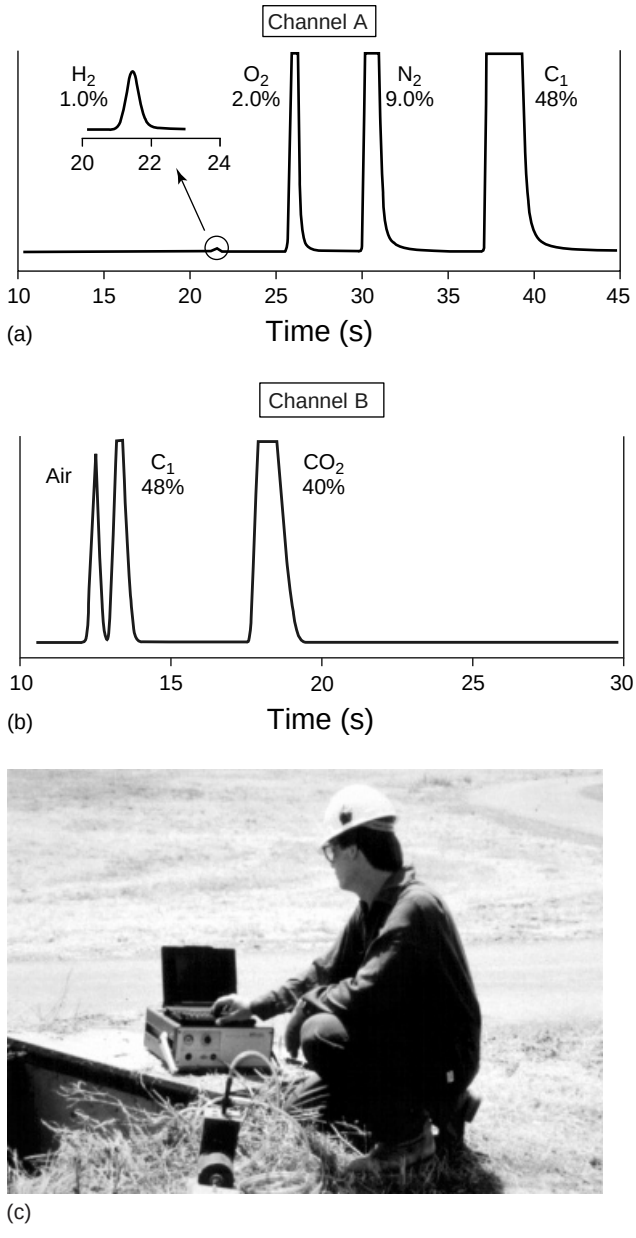

Figure 19 Landfill gas analysis is used to measure the anaerobic oxidation efficiency in nature's decomposition process. The P-200 GC is equipped with two GC modules (Channel A and Channel B). Channel A monitors the oxygen and nitrogen using a 4 m × 0.32 mm ID molecular sieve column for separation at 50 °C (a). Channel B separates air, carbon dioxide and methane on a 0.25 m × 0.5 mm ID HayeSep A Micropack column at 40 °C (b). The actual measurement with this portable instrument is shown in (c). (Reproduced by courtesy of Hewlett-Packard Company.)

separate CO_2, CH_4, and C_3, C_4, C_5 isomers. The analysis of these gases is done in less than 25 s (Figure 19b, Channel B). Typically, CO_2 and CH_4 values range from 40 to 55% and O_2 and N_2 values are in the range of 2 to 10%. The Micro GC's detection limit for H_2S is about 10 ppm.

When using or selling gas as an energy source, the operator can calculate the BTU content to determine pricing. BTU information is also valuable for deciding where to place new landfill wells and when to start harvesting gases, e.g. the methanogenic state of the landfill.

5.2.4.3 Refinery Gas Analysis The petrochemical industry uses the Micro Quad series GC (which houses up to four independent column modules) to perform an analysis of a complex mixture of refinery gases, called refinery gas analysis (RGA). Configured with dual carrier gas inlets (helium and argon), the sample is pulled into the instrument and simultaneously distributed to each of the four Micro GC modules and analyzed in less than 3 min. Each of these modules is independently temperature and pressure controlled, and has a precolumn to protect the analytical column and reduce analysis time.

The first column module (Figure 20a, Channel A) consists of a 0.32-mm ID 10-m molecular sieve 5A PLOT, operating at 110 °C, which uses argon as its carrier gas and has a 0.7-m PoraPLOT-U precolumn for backflush-to-vent. Channel A specifically analyzes He, H_2, O_2, N_2, CH_4 (methane), and CO in 130 s.

The second column module (Figure 20b, Channel B) consists of a 0.32-mm ID, 8-m PoraPLOT-U operating at 80 °C that uses helium as a carrier gas and uses a 1.0-m PoraPLOT-Q precolumn for backflush-to-vent. Channel B specifically analyzes methane, CO_2, ethylene, ethane, acetylene, hydrogen sulfide, propane/propylene, propyne and C_4 plus (butane) in 100 s.

The third column module (Figure 20c, Channel C) has a 0.32-mm ID 10-m Al_2O_3 column and uses helium as a carrier gas. Channel C separates propane, propylene, C_4 *n*-butane and isobutylene, and C_5 pentane hydrocarbons in 130 s.

The fourth column module (Figure 20d, Channel D) uses a 0.15-mm ID 10-m OV-1 column and uses helium as a carrier gas with backflush-to-vent. The OV-1 analyzes C_4 isobutene through C_6+ normal hexane in 85 s.

These examples show that there are values associated with the implementation of miniaturized components. Smaller components increase speed to analysis and reduce the volume of carrier gas needed. Less sample gas is also needed to obtain the same result, but faster. Reduction in component size enables the instrument to be smaller and occupy less costly bench space. When smaller and lighter components are available enabling lightweight portable instruments to be designed, these instruments can be easily carried into the field to perform on-site analysis. Speed of analysis enables the user to acquire and analyze more samples in a given time period, saving money and increasing the amount of available data. The disadvantage of miniaturization is lack of the versatility found in laboratory instruments. Portable

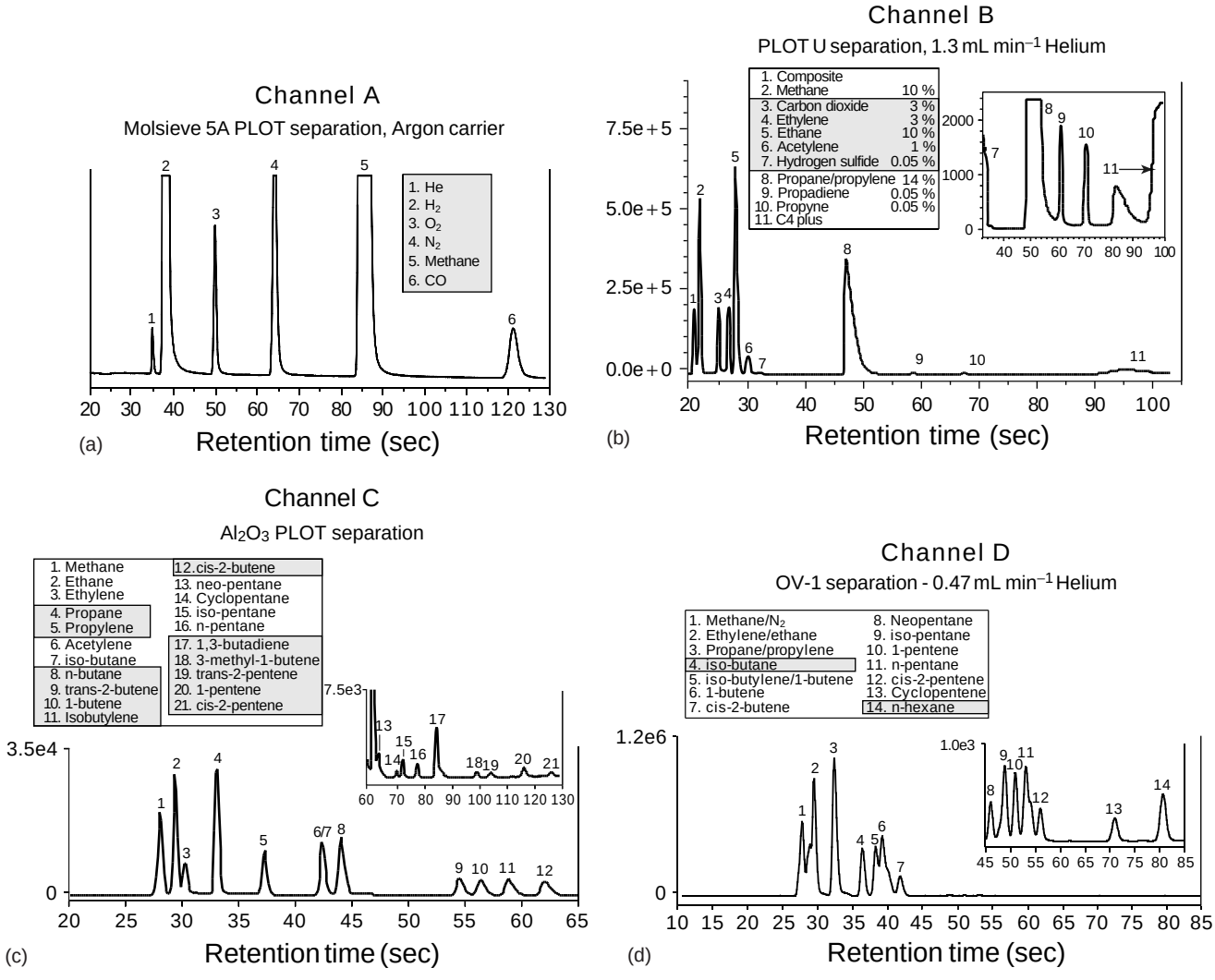

Figure 20 The refinery gas analysis utilizes four independent Micro GC modules to analyze simultaneously a complex gas sample in less than 160 s. The same analysis performed with a conventional GC instrument would take from 30 to 60 min. (Reproduced by courtesy of Hewlett-Packard Company.)

instruments have to be specialized for use in a few specific applications.

5.3 Miniaturization of Mass Spectrometers

Finally, we should also mention briefly the work done in the miniaturization of MSs. Mass spectrometry, when combined with GC, continues to be a standard method for the identification and characterization of gaseous chemical species. It is widely used in environmental, medical, military, and industrial applications, and would see even broader use if the equipment were truly portable for field measurements. To date, most of the equipment used for high-performance MS is based on conventional technology and is adapted to laboratory measurements.

There have been a few attempts to apply MEMS technology to portable MS and we will report preliminary results available for two of them. The Applied Physics Laboratory has developed a small, powerful, time-of-flight MS in a collaborative program with the Johns Hopkins Medical Institution and the University of Maryland. Miniaturization of the equipment required the development of new techniques for ion formation in the ion source region and the use of kinetic-energy spread correction schemes based on ion reflectrons. Sampling and ionization schemes for solids, liquids, and vapors have led to promising conceptual designs for microorganism identification, environmental monitoring, and law enforcement analytical tools.[68]

The Jet Propulsion Laboratory (JPL) has initiated a program to miniaturize and ruggedize a quadrupole mass filter using MEMS techniques to make them suitable for space missions.[69] The JPL program objective is to

achieve a mass range of 1 to 300 atomic mass units (amu) with a resolution of 0.5 amu. This compares favorably with the 15–212 amu range and a 1-amu resolution of the MS used on the Viking I and II spacecraft, which landed on Mars in 1976. The LIGA technique will be used to create quadrupole rods 2.2 mm in length. To achieve the stated dynamic range and resolution, pole dimensions must be controlled to within 0.1%. Although this geometrical accuracy is yet to be demonstrated, it is within the theoretical limit of the LIGA technique.

6 CONCLUSION

The microsystems industry is now a rapidly growing sector of MEMS that has emerged since the late 1960s. The onset was mainly a technological spin-off from microelectronics/IC technology. Silicon micromachining gave the main triggering technology push as a very promising process technology with distinctive features. Sensor applications gave the first market opportunities and batch-organized processing technology adapted from the microelectronics industry provided the key to high quality at low cost. Today, MEMS has matured into a separate industry sector with its own market and manufacturing infrastructure with a use of materials other than silicon. Microsystems based on MEMS are used for applications ranging from low-cost, high-volume automotive applications to high-cost, low-volume instrumentation applications for the aerospace industry. Sensors used by the automotive industry continue to be the prime market for this technology and are used more and more to reduce production costs and provide more sophisticated and reliable instruments.

In this review of applications to chemical instrumentation we have shown how MEMS technology has been used to produce components which have a unique performance. There are now enough documented applications to indicate that this technology is real and will eventually change the industry. We have described how MEMS technology has been applied to produce components of small internal sizes and thus improve the performance of the instrument.

ABBREVIATIONS AND ACRONYMS

BTU	British Thermal Unit
CMOS	Complementary Metal Oxide Semiconductor
CVD	Chemical Vapor Deposition
DRIE	Deep Reactive-ion Etching
GaAs	Gallium Arsenide
GC	Gas Chromatograph
GPS	Global Positioning Systems
HF	Hydrofluoric
HSGC	High-speed Gas Chromatography
IC	Integrated Circuit
ID	Internal Diameter
IT	Information Technology
JPL	Jet Propulsion Laboratory
KOH	Potassium Hydroxide–Water–Isopropyl Alcohol
LIGA	X-ray Lithography (X-ray Lithography), Electrodeposition (Galvanoformumg), and Molding (Abformtechnik)
MAP	Manifold Absolute Pressure
MEMS	Microelectromechanical Systems
MOS	Metal Oxide Semiconductor
MS	Mass Spectrometer
MTCD	Microthermal Conductivity Detector
NASA	National Aeronautics and Space Administration
PCR	Polymerase Chain Reaction
PVD	Physical Vapor Deposition
RGA	Refinery Gas Analyzer
RIE	Reactive Ion Etching
SAW	Surface Acoustic Wave Sensor
SPOT	Smart Planar Optical Transducer
TCD	Thermal Conductivity Detector
UV	Ultraviolet
VOC	Volatile Organic Compound
WCOT	Wall Coated Open Tubular
3-D	Three-dimensional
μEDM	Microelectrodischarge Machining

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Airborne Instrumentation for Aerosol Measurements • Automotive Emissions Analysis with Spectroscopic Techniques • Diode Laser Spectroscopic Monitoring of Trace Gases • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments • Photoacoustic Spectroscopy in Trace Gas Monitoring

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Flow-injection Techniques in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental

Analyses • Industrial Waste Dumps, Sampling and Analysis • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Soil Instrumental Methods • Solid-phase Microextraction in Environmental Analysis • Supercritical Fluid Extraction of Inorganics in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Underground Fuel Spills, Source Identification • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Chemical-sensing Networks: Satellite-based • Cone-penetrometer-deployed Samplers and Chemical Sensors • Electrochemical Sensors for Field Measurements of Gases and Vapors • Field-based Analysis of Organic Vapors in Air • Field-portable Instrumentation

Field-portable Instrumentation cont'd **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements • Radon, Indoor and Remote Measurement of • Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

Food **(Volume 5)**
Food Analysis Techniques: Introduction

Forensic Science **(Volume 5)**
Ion Mobility Spectrometry in Forensic Science

Industrial Hygiene **(Volume 5)**
Industrial Hygiene: Introduction • Aerosols and Particulates Analysis: Indoor Air • Carcinogens, Monitoring of Indoor Air • Chromatographic Techniques in Industrial Hygiene • Direct Reading Instruments for the Determination of Aerosols and Particulates

Industrial Hygiene cont'd **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Pesticides in Water: Sampling, Sample Preparation, Preservation

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Oil Shale and Shale Oil Analysis • Petroleum Residues, Characterization of

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Chemical Reagents and Derivatization Procedures in Drug Analysis • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Robotics and Laboratory Automation in Pharmaceuticals Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers • Inverse Gas Chromatography in Analysis of Polymers

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction • Chromatography in Process Analysis • Flow and Sequential Injection Analysis Techniques in Process Analysis • Mass Spectrometry in Process Analysis

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Gas Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. J. Angell, S. Terry, P. Barth, 'Silicon Micromechanical Devices', *Sci. Am.*, **248**(4), 44–55 (1983).
2. S.C. Terry, J.H. Jerman, J.B. Angell, 'A Gas Chromatograph Air Analyzer Fabricated on a Silicon Wafer', *IEEE Trans. Electron Devices*, ED-26, 1880–1886 (1979).
3. K.E. Petersen, 'Fabrication of an Integrated Planar Silicon Ink-jet Structure', *IEEE Trans. Electron Devices*, ED-26, 1918–1920 (1979).
4. G.T.A. Kovacs, K. Petersen, M. Albin, 'Silicon Micromachining – Sensors to Systems', *Anal. Chem.*, **68**, A407–A412 (1996).
5. P. Ohlckers, H. Jakobsen, *IEEE Colloquium on Assembly and Connections in Microsystems (Ref. No.1997/004)*, IEEE London, UK, 1997.
6. Report No. 915-40: *World Emerging Sensor Technologies – High Growth Markets Uncovered*, Frost & Sullivan, Mountain View, CA, 1993.
7. D.S. Eddy, D.R. Sparks, 'Application of MEMS Technology in Automotive Sensors and Actuators', *Proc. IEEE*, **86**(8), 1747–1755 (1998).
8. W. Buff, M. Rusko, E. Goroll, J. Ehrenpfordt, T. Vandahl, S.C. Schneider, M. Levy, B.R. McAvoy, 'Universal Pressure and Temperature SAW Sensor for Wireless Applications', *Proc. IEEE Ultrasonics Symposium*, **1**, 359–362 (1997).
9. J.F. Arrigo, K.J. Page, P.M. Chau, N.C. Tien, 'The Application of Reconfigurable Processors to MEMS Calibration and Signal Processing in Inertial Navigation Systems', *Proceedings 2nd Annual IEEE International Conference on Innovative Systems in Silicon*, USA, 1997.
10. P. Ohlckers, A. Hanneborg, M. Nese, 'Batch Processing for Micromachined Devices', *J. Micromech. Microeng.*, **5**(2), 47–56 (1995).
11. *Microelectromechanical Systems (MEMS)*, Frost & Sullivan, Mountain View, CA, 1997.
12. K.E. Petersen, 'Silicon as a Mechanical Material', *Proc. IEEE*, **79**(5), 420–457 (1982).
13. M. Madou, *Fundamentals of Microfabrication*, CRC Press, Boca Raton, 1997.
14. K.J. Gabriel, 'Microelectromechanical Systems (MEMS)', *IEEE Aerospace Conference Proceedings*, Snowmass at Aspen, CO, USA, 1997.
15. W.M. Thompson, C.G. Wilson, M.J. Bowden, *Introduction to Microlithography*, American Chemical Society, Washington DC, 1994.
16. R. Laudise, *The Growth of Single Crystals*, Prentice Hall, Englewood Cliffs, NJ, 1970.
17. K.E. Bean, 'Anisotropic Etching of Silicon', *IEEE Trans. Electron, Devices*, ED-25, **10**, 1185–1193 (1978).
18. H.R. Kaufman, *Technology of Electron-bombardment Ion Thrusters*, Advances in Electronics and Electron Physics 36, Academic Press, New York, 265–373, Vol. 36, 1974.
19. R.G. Poulsen, 'Plasma Etching in IC Manufacture – A Review', *J. Vac. Sci. Technol.*, **1**, 266–274 (1977).
20. I. Brodie, J. Muray, *The Physics of Microfabrication*, Plenum Press, New York, London, 1984.
21. H.D. Goldberg, R.B. Brown, D.P. Liu, M.E. Meyerhoff, 'Screen Printing: A Technology for the Batch Fabrication of Integrated Chemical-sensor Arrays', *Sens. Actuators, B (Chemical)*, **21**(3), 171–183 (1994).
22. R.W. Hower, J.H. Shin, G.S. Cha, R.K. Meruva, M.E. Meyerhoff, R.B. Brown, *Sensors & Actuators, B, Chemical*, Elsevier, Stockholm, Sweden, 168–172, Vol. 33, 1996.
23. V. Ferrari, D. Marioli, A. Taroni, E. Ranucci, P. Ferruti, 'Gravimetric Chemical Sensors in Thick-film Technology with Hybrid Electronics', *Conference Proceedings. SAA '96 National Meeting on Sensors for Advanced Applications*, Brescia, Italy, 3–10, 1997.
24. E. Milella, F. Musio, M.B. Alba, 'Polypyrrole LB Multilayer Sensitive Films for Odorants', *Thin Solid Films Conference*, 908–910, 1996.
25. R. Ruprecht, W. Bacher, J. Hausselt, V. Piotter, 'Injection Molding of LIGA-similar Microstructures Using Filled and Unfilled Thermoplastics', *Micromachining and Microfabrication Process Technology, Proc. SPIE*, Austin, TX, 146–157 (1995).
26. J. Mohr, W. Bacher, P.S.M. Bley, U. Wallrabe, 'The LIGA Process. A Tool for the Fabrication of Microstructures Used in Mechatronics. *Proceedings 1st Japanese-French Congress of Mechatronics*, Besancon, France, 1–7, 1992.
27. T. Akin, 'Silicon Micromachining and Microfabrication Techniques for Integrated Sensor and Actuator Systems', *Proceedings of MELECON '94, IEEE Region 8*, USA, 1994.
28. K. Najafi, C.H. Mastrangelo, 'Solid-state Microsensors and Smart Structures', *Proc. IEEE Ultrasonics Symposium*, 1993.
29. W.C. Tang, 'Micromechanical Devices at JPL for Space Exploration', *1998 IEEE Aerospace Conference Proceedings*, Aspen, CO, 461–470, 1998.
30. G. Wallis, D.I. Pomerantz, 'Field Assisted Glass-metal Sealing', *J. Appl. Phys.*, **40**, 3946–3949 (1969).
31. H. Ohashi, J. Ohura, T. Tsukakoshi, M. Shimbo, 'Improved Dielectrically Isolated Device Integration by Silicon-wafer Direct Bonding Techniques', *IEEE International Electron Devices Meeting, Technical Digest, IEDM '86*, Los Angeles CA, 211–213, 1986.
32. A.L. Tiensuu, J.A. Schweitz, S. Johansson, 'In Situ Investigation of Precise High Strength Micro Assembly Using Au–Si Eutectic Bonding', *8th International Conference on Solid-state Sensors and Actuators (Transducer '95)*, Stockholm, Sweden, 236–239, 1995.

33. S.M. Sze, *VLSI Technology*, McGraw-Hill Book Company, New York, 1983.
34. S. Marshall, 'Fundamental Changes Ahead for Lab Instrumentation', *R&D Magazine*, 18–27 (1999).
35. K. Takahata, S. Aoki, T. Sato, 'Fine Surface Finishing Method for 3-dimensional Micro Structures', *Proc. IEEE Micro Electro Mech. Sys. Workshop*, San Diego, CA, 73–78, 1996.
36. D. Bauerle, 'Laser-chemical Processing: Recent Developments', *Appl. Surface Sci.*, **106**, 1–10 (1996).
37. K. Ikuta, K. Hirowatari, 'Real Three-dimensional Micro Fabrication Using Stereo Lithography and Metal Molding', *Proc. IEEE Micro Electro Mech. Sys. Workshop*, Fort Lauderdale, FL, 42–47, 1993.
38. L. Spangler, C.J. Kemp, 'ISAAC-integrated Silicon Automotive Accelerometer', *8th International Conference on Solid State Sensors & Actuators & Eurosensors*, 585–588, 1995.
39. W.C. Tang, 'MEMS Applications in Space Exploration', *Proceedings of the SPIE – The International Society for Optical Engineering*, 1997.
40. W.C. Tang, *Proceedings 1997. Design Automation Conference, 34th DAC*, Acm New York, NY, USA, Anaheim, CA, USA, 1997.
41. M.W. Bruns, 'High-speed Portable Gas-chromatography – Silicon Micromachining', *Erdol & Kohle Erdgas Petrochemie*, **47**, 80–84 (1994).
42. S.J. Santy, 'Portable Gas Chromatograph', *Gas Quality Symposium*, Institute of Gas Technology, DesPlaines, Chicago, IL, 1991.
43. K. Petersen, *International Electron Devices Meeting. Technical Digest (Cat. No. 96CH35961)*, IEEE New York, NY, USA, San Francisco, CA, USA, 239–242, 1996.
44. H. Becker, W. Dietz, 'Microfluidic Devices for u-TAS Applications Fabricated by Polymer Hot Embossing', *Proceedings of SPIE – Microfluidic Devices and Systems*, Santa Clara, CA, 177–182, 1998.
45. R.E. Kunz, *Sensors & Actuators, B, Chemical*, Elsevier, Zurich, Switzerland, 13–28, Vol. 38, 1997.
46. *Profiles of US Industrial Chemical and Gas Sensor Manufacturers*, Frost & Sullivan, Mountain View, CA, 1995.
47. M. Trinkel, W. Trettnak, F. Freininger, R. Benes, O.S. Wolfbeis, 'Optochemical Sensor for Ammonia Based on a Lipophilized pH Indicator in a Hydrophobic Matrix', *Int. J. Environ. Anal. Chem.*, **67**, 237–251 (1997).
48. G. Frye, D.W. Gilbert, C. Colburn, R.W. Cernosek, T. Steinfort, 'Above-ground and In Situ Field Screening of VOCs Using Portable Acoustic Wave Sensor (PAWS) Systems', in *Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, Las Vegas, Nevada, 715–726, 1995.
49. L. Reindl, G. Scholl, T. Ostertag, C.C.W. Ruppel, W.E. Bulst, F. Seifert, M. Levy, S.C. Schneider, B.R. McAvoy, 'SAW Devices as Wireless Passive Sensors', *IEEE Ultrasonics Symposium Proceedings*, 363–367, 1996.
50. A.J. Ricco, 'SAW Chemical Sensors. An Expanding Role with Global Impact', *Electrochem. Soc. Interface*, **3**(4), 38–44 (1994).
51. J.H. Kim, J.S. Sung, Y.M. Son, A.A. Vasiliev, V.V. Malyshev, E.A. Koltypin, A.V. Eryshkin, D.Y. Godovski, A.V. Pisyakov, S.S. Yakimov, 'Propane/Butane Semiconductor Gas Sensor with Low Power Consumption', *Sens. Actuators, B*, **B44**, 452–457 (1997).
52. J.H. Jerman, *A Miniature, Thin-film Thermal Conductivity Detector for an Integrated Gas Chromatograph*, Stanford Electronics Laboratories, Stanford University, Stanford, CA, 1981.
53. M.W. Bruns, 'Silicon Micromachining and High Speed Gas Chromatography', *Proceedings of the 1992 International Conference on Industrial Electronics, Control, Instrumentation and Automation*, 1640–1644, 1992.
54. J.K. Robertson, K.D. Wise, 'An Electrostatically-actuated Microvalve for Semiconductor Gas Flow Control', *Technical Digest Solid-state Sensor and Actuator Workshop*, 148–151, 1996.
55. M. Zdeblick, 'A Planar Process for an Electric-to-fluidic Valve', PhD Thesis, Stanford University, 1988.
56. J.D. Busch, D. Johnson, 'Prototype Microvalve Actuator', *IEEE Micro Electro Mechanical Systems Workshop*, 40–41, 1990.
57. H. Jerman, 'Prototype Microvalve Actuator', *IEEE Solid State Sensor and Actuators Workshop*, 1990.
58. J.G. Smits, 'A Piezoelectric Micropump with Three Valves Working Peristaltically', *Sens. Actuators*, **15**, 153–167 (1990).
59. G. Gordon, P. Barth, US Patent 5 058 856, May, 1991.
60. P. Barth, C. Beatty, L. Field, J. Baker, 'A Robust Normally-closed Silicon Valve', *Solid-state Sensor and Actuator Workshop*, Hilton Head, SC, 248–250, 1994.
61. A.J.J. Van Es, 'High Speed Narrow Bore Capillary Gas Chromatography', PhD Thesis, Technishe Universiteit Eindhoven, 1990.
62. A. Van Es, J. Janssen, R. Bally, J. Rijks, 'Sample Introduction in High Speed Capillary Gas Chromatography; Input Band Width and Detection Limits', *HRC & CC J. High Resolut. Chromatogr.*, **10**(5), 273–279 (1987).
63. K.J. Hyver, R.J. Philips, 'Considerations in Enhancing Resolution, Speed and Sensitivity in Capillary Gas Chromatography and Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **399**, 33–46 (1987).
64. P. Sandra, 'Fast Capillary Gas Chromatography', *LC/GC*, **5**, 236–246 (1987).
65. A.G. Loudon, J. Kolbe-Boelke, J. Adler, J. Stach, 'The Use of On-site Analysis for the Protection of Personnel', *Analysis*, **23**(2), M22–M24 (1995).
66. R. Annino, 'Practical Limits of High-speed Chromatography – Evaluation with a Computer-based Theoretical-model of Several Commercially Available Gas Chromatographs', *J. High Resolut. Chromatogr.*, **19**(5), 285–290 (1996).

67. S. Dagan, A. Amirav, 'Fast, High Temperature and Thermolabile GC/MS in Supersonic Molecular Beams', *Int. J. Mass Spectrom. Ion Proc.*, **133**(2), 187–210 (1994).
68. W.A. Bryden, R.C. Benson, S.A. Ecelberger, T.E. Phillips, T.J. Cornish, R.J. Cotter, 'Tiny-TOF/MALDI Mass Spectrometry for Particulate Drug and Explosives Detection', *Proc. SPIE Int. Soc. Opt. Eng.*, **2511**, 153–164 (1995).
69. D.V. Wiberg, M.H. Hecht, O.J. Orient, A. Chutjian, K. Yee, S. Fuerstenau, 'A LIGA Fabricated Quadrupole Array for Mass Spectrometry', *Ext. Abs., High Aspect Ratio Micro Structure Tech. Forum*, Madison, WI, 1997.

Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Gerhard Matz and Wolfgang Schröder
Technical University Hamburg-Harburg, Hamburg, Germany

Tapio Kotiaho
VTT Chemical Technology, FIN-02044 VTT, Finland

1 Introduction 1
2 Instrumentation 2
2.1 Inlet Systems 2
2.2 Mass Separation 3
2.3 Mixture Separation by Gas Chromatography 3
3 Sampling 4
3.1 Gaseous Samples 4
3.2 Water Samples 5
3.3 Soil Samples 5
4 Mobile Mass Spectrometry and Gas Chromatography/Mass Spectrometry Systems 5
4.1 On-line Mass Spectrometry 8
4.2 Mobile Gas Chromatography/Mass Spectrometry Systems 10
5 Quality Assurance/Quality Control Procedures and Comparison of On-site/In situ Measurements with other Methods/Fixed Laboratory Methods 11
6 Experience Gained from Analytical Campaigns with Mobile Systems 13
7 Conclusion 14
Abbreviations and Acronyms 15
Related Articles 15
References 15

The potential of mass spectrometry (MS) and gas chromatography/mass spectrometry (GC/MS) techniques for on-site analysis of compounds in a variety of matrixes such as air, water, soil and surfaces is discussed. The fundamental principles of operation including different mass separation procedures, modes of injection and separation methods for mixtures are presented here with a special emphasis on equipment suitable for measurement in the field. Commercially available systems as well as prototypes are described in detail. The power of MS and/or GC/MS analyses on-site compared with conventional methods carried out in the laboratory is demonstrated in a detailed review of the literature dealing with measurement campaigns and experiences when applying mobile MS and GC/MS. In conclusion, mobile mass spectrometers have reached a high standard of quality and the operational areas in which they are utilized are likely to increase considerably in the future.

1 INTRODUCTION

The first incentive to operate mobile mass spectrometers came from space research to investigate the gas and surface material of Venus (Pioneer program)[1,2] and Mars (Viking Mars mission[3,4]) in the early 1970s. Also in research on the atmosphere of the earth, mobile MS experiments (e.g. on balloons and in rockets) were accomplished.[5,6] For vapor sniffing a small three-stage membrane quadrupole mass spectrometer has been constructed to detect explosives, chemical effluents and drugs.[7] In addition to this research equipment, military applications and instruments for detecting chemical warfare agents on the battlefield have been developed, and several hundred of these specially designed systems are installed in reconnaissance vehicles.

The requirement to detect chemical substances in the field safely almost inevitably led to the measurement principle of MS. If organic compounds are to be detected, a large variety of substances with different molecule sizes and functionality have to be taken into account. Suitable detectors must be able to supply a multiplicity of detailed item information. Apart from mass spectrometers, basically, infrared (IR) spectroscopy performs this task well since it achieves an extremely high degree of interaction with the analyte. However, only structures of molecules which reach a state of resonance with electromagnetic radiation can be detected. Accordingly, only fractions of the molecule are described. If there is a restriction with regard to evaporable materials, MS is the most successful chosen detection procedure. All basic modules of the molecule structure are taken into account. The molecular ion determines the molecular weight of the material, and fragments, which usually consist of characteristic parts of the molecule, permit identification of the substance simply from the typical fragment patterns. Furthermore, the sensitivity of current mobile mass spectrometers is superior to that of commercial mobile IR instruments. For mixtures, chromatographic separation prior to MS detection facilitates identification and quantification. This is usually achieved by coupling gas chromatography (GC)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

with MS (GC/MS). Identification is carried out by spectrum comparison with commercially available spectrum databases.[8,9]

As a result of the military developments, progress in microelectronics and miniaturization of laboratory MS devices, mobile MS systems are now also available in the civil sector. They are used for analyzing contaminated air, soil, water and waste, in worker protection as well as in chemical accidents.

The use of mobile instrumentation, in particular mobile mass spectrometers, for environmental measurements offers many clear advantages compared with analysis performed in stationary laboratories. The main advantages are

- sample handling and transportation steps are minimized when the analysis is performed on-site/in situ;
- possibilities of the change of sample composition are minimized since samples are analyzed in real-time or near-real-time;
- rapid and reliable identification and quantification capability in various types of emergency situations (e.g. fires, chemical spills);
- unexpected or unusual analytical results can be confirmed rapidly;
- reduced total analytical costs during remediation of a contaminated site.[10–12]

This article therefore aims to present the current level of development of the MS or GC/MS technique with mobile devices and their efficiency as an example of the overall possible and necessary improvement in analytical techniques.

2 INSTRUMENTATION[10,13–19]

In the mass spectrometer ions are separated in electrical and/or magnetic fields and their masses are measured (Figure 1). Common devices ionize the molecules by electron impact. Emitted from a glow cathode, the electrons are accelerated to 70 eV. The mass separation process takes place in a vacuum chamber evacuated to at least 0.01 Pa (10^{-7} bar) to allow a collision-free flight path through the analyzer.

Frequently used inlet systems implemented in mobile devices and mass separation processes are described below. Since signal overlay by the presence of many compounds in a mixture prevent identification and quantification, gas chromatographic columns are used in some devices. Their mode of operation is also briefly described.

The description of the operating principles of the machine is limited to constructions that are actually used in mobile devices or would possibly be suitable.

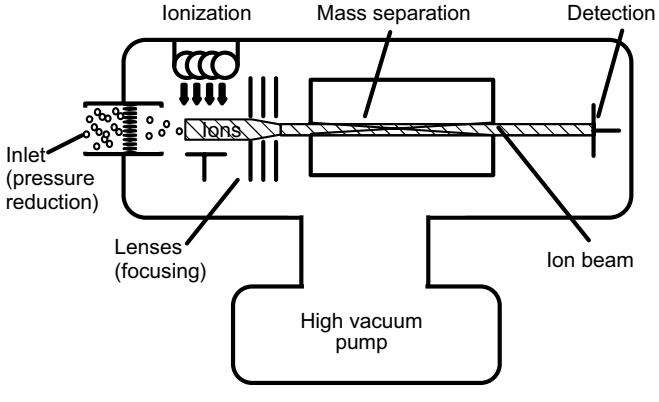

Figure 1 Principal structure of mass spectrometers.

2.1 Inlet Systems

The function of the inlet system is to transfer gaseous substances into the vacuum section of the device according to the efficiency of the vacuum pumps. The easiest method of implementation is by means of direct inlet systems using defined leaks formed by a thin hole or a capillary, inlets using differential pumping (Figure 2a) (e.g. jet separator) or by means of the membrane separator (Figure 2b).[17,20,21] Since direct inlets are more easily contaminated upon entry of non-volatile materials or particles than those which use a membrane, they are normally only used for gas analysis or coupled with GC. In differential inlets high-volume sample transfer through the inlet line is achieved. The risk of contamination compared with restriction capillary inlets is smaller owing to the short contact time of the sample to inlet parts.

In the membrane separator the sample gas is sucked through a thin silicone tube leading through the vacuum whereby a fraction of the gas passes through the silicone membrane into the vacuum by permeation

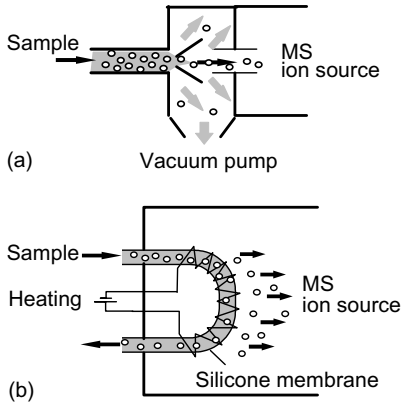

Figure 2 Inlet systems for mass spectrometers; (a) coupling with differential pumping; (b) using a membrane separator.

processes. In addition, flat membrane layers are used. Owing to the hydrophobic nature of silicone, organic components are introduced into the vacuum while polar compounds (mainly water) are discriminated. Also semi-volatile materials (up to benzopyrene or polychlorinated biphenyls (PCBs)) can be analyzed whilst heating the membrane. The advantage of membrane separators is their extremely good durability. However, they may only be used for volatile compounds if they are not heated. If they are heated this involves higher machine expenditure and high levels of energy consumption.

2.2 Mass Separation

During mass isolation we basically distinguish between two procedures: separation in time (SIT) and separation in space (SIS). In the case of SIT the separation system works like a variable filter which allows different masses to pass temporally. The quadrupole[18] (Figure 3) and ion trap systems[22,23] (Figure 4) are examples of this method. The advantage of SIT separation systems is their compactness. The ion trap arrangement offers the additional possibility of holding individual ions and letting them fragment again in the mass spectrometry/mass

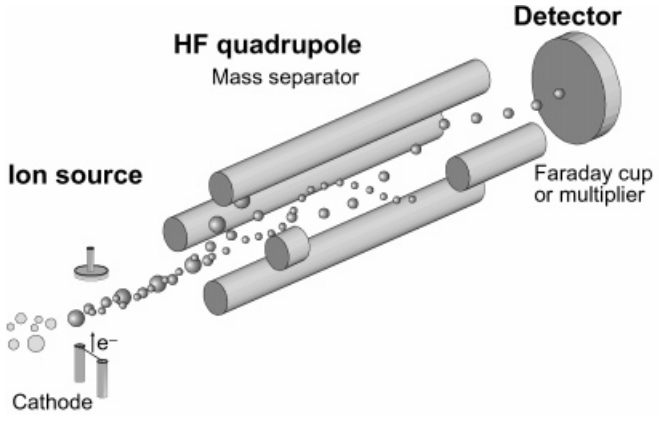

Figure 3 Quadrupole mass spectrometer.

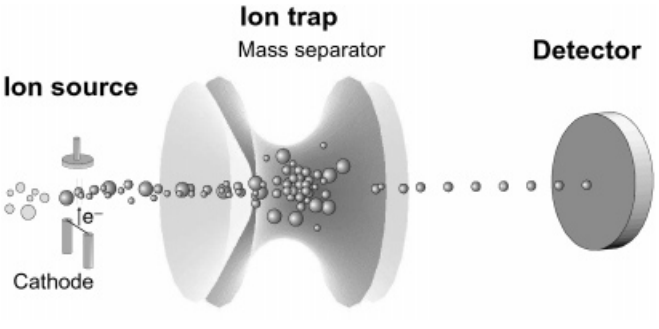

Figure 4 Ion trap mass spectrometer.

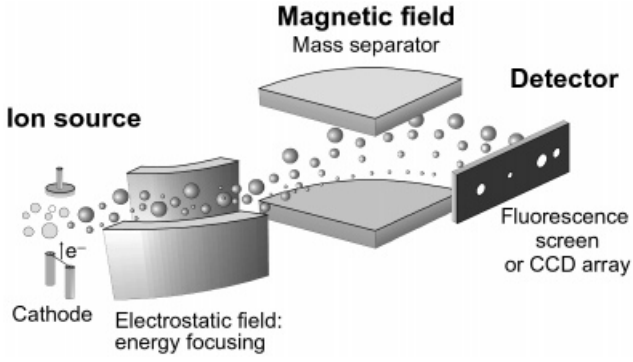

Figure 5 Nonscanning magnetic field mass spectrometer.

spectrometry (MS/MS) mode.[24] Nonscanning magnet sector field devices with detector array[25] (Figure 5) and time of flight (TOF) mass spectrometers[26] (Figure 6) are examples of SIS. In a magnetic mass spectrometer the moving energy of the accelerated ion cloud is focused by an electrostatic field and the masses are separated in the magnetic field on discrete ion flight paths to be finally measured by a set of detectors. In the TOF mass spectrometer after acceleration, the ions are separated according to their mass-related different TOF. The advantage of these devices is that they are very fast (full spectra within a few milliseconds for SIS compared with a few 100 ms for SIT).

2.3 Mixture Separation by Gas Chromatography[27]

The gas chromatograph consists of an injector in which the sample is sucked in or evaporated, and a capillary (inner diameter 0.1 to 0.5 mm) through which the carrier gas passes, moving the compounds through the column while separation takes place. A part of the sample is injected into the gas flow and conducted through the capillary. Interaction with a polymer coating on the inner wall of the capillary causes the compounds to separate. Analytes may be volatile organic compounds (VOCs) as well as semivolatile organic compounds (SVOCs). They are isolated according to their different evaporating ability and within certain limits remain briefly or longer according to their different polarity in the polymer. This makes it possible to separate even complex mixtures. Heating the capillary results in faster movement of the compounds through the column. Thus, at high temperatures the run times of semi-volatile compounds may be substantially shortened. However, separation under a constant temperature (isothermal operation) takes place within compound groups of a small range of evaporating ability. More volatile compounds pass unseparated ahead and low-volatiles remain on the injector side. Increasing the temperature continuously (temperature-programmed GC) causes all

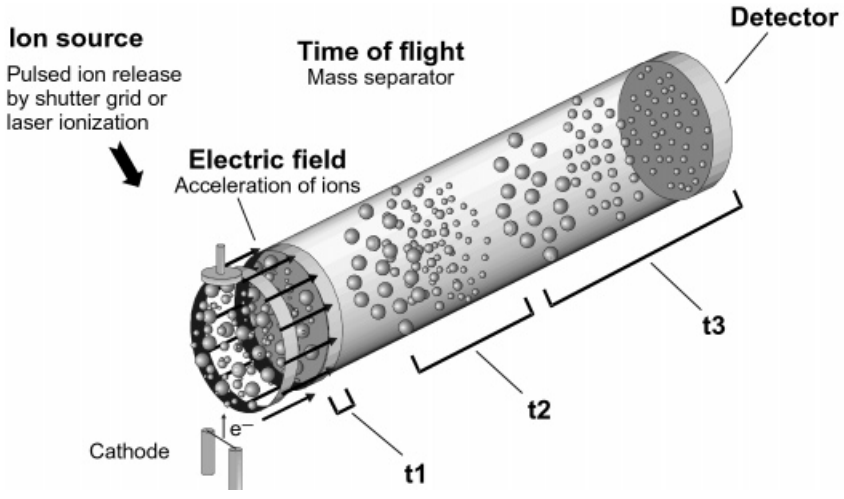

Figure 6 Time-of-flight mass spectrometer.

injected substances to pass through the capillary with optimal separation power at any range of volatility. Regarding mobile applications it is important to mention that temperature programming requires a much higher amount of power than isothermal heating.

Generally, injection temperatures range from 200–300 °C. Similar temperatures are possible as a maximum for heating the separation film on the inner surface of the capillary. Usually nitrogen or helium is used in mobile devices as the carrier gas. In addition, the use of carbon dioxide may have certain advantages.[28–33] Since modern GC phases have become extremely stable towards high temperatures and chemical reactions as a result of polymer cross-linking and covalent linkage to the capillary wall (chemical bonding), air can also be used as a carrier gas without degradation of the phase at high temperatures.[34,35] Filtered ambient air is generally used in such applications.

The duration of the chromatographic run depends on the slope of the temperature program and the length of the separation capillary. Lengths between less than 1 m and 15 m are the norm in mobile gas chromatographs. Separation in very short capillaries is called transfer-line chromatography[36] because the restrictor transfer line to the MS is used for sample gas feed and fast separation in one. The run times in transfer-line GC vary between seconds and some minutes. Especially for fast GC/MS, where co-elution of compounds with similar volatility frequently occurs, evaluation procedures are available to deconvolute even strongly overlapping signals.[37,38] Thus, spectra from clean compounds are generated and used for clear identification based on MS database comparisons. Fast GC/MS analysis of PCB mixtures has proven that from GC runs of 5 min the information of separations that usually last 40 min may be extracted.[39]

Procedures allowing on-line chromatography without injection by impressing a modulation signal at the injector side of the capillary, is briefly mentioned here.[40–41] This technique seems to fit in very well with the demands on fast on-line measurements. For example, it is supposed to be a future alternative to conventional GC in chemical analysis on board spacecraft.[42]

3 SAMPLING

Sampling techniques with a low level of instrument expenditure and sample preparation, which are feasible in the field, are described. The summary is subdivided into applications for the on-line measuring technique and the procedures suitable for the different matrices occurring in the environment.

3.1 Gaseous Samples

For gas sampling plastic foil bags made of inert material (Tedlar™) are used.[43,44] High-grade steel cans with a passivated surface against adsorption,[45] and glass vials[46] are also used. They are evacuated before measurement and the sample is gathered locally by opening a valve. Usually subsets are transferred into the analyzer with gas-tight glass syringes.

Sampling with simultaneous enrichment is performed by enrichment on an adsorbent. Tenax™ is the most commonly used material[47] since humidity from the air is not enriched and therefore does not adversely affect the thermal desorption and injection. If extremely volatile materials, such as vinyl chloride which barely adsorbs on Tenax™, are to be enriched, traps with different adsorbents are used,[48] e.g. a Tenax™ packing is followed

by modified activated charcoal layers with a higher surface activity.

Gaseous samples can also be analyzed directly without a separate sampling step by introducing the gaseous sample into the ion source of a mass spectrometer via a direct inlet, such as valves or capillaries, or via a membrane inlet.[10] These methods offer an excellent means for on-line monitoring of gaseous streams or ambient air as well as industrial emissions.

3.2 Water Samples[49–51]

Of the different sampling procedures for water, sampling with amber-colored bottles is the most usual method used. This original sample is filled without headspace and properly sealed to reduce the risk of loss of highly volatile compounds during transport.

For headspace analysis this sample is filled into another gas-tight vial with a defined headspace, where the volatile substances are in equilibrium with the liquid phase. The compounds are taken either directly out of the gas space from the bottle using a gas-tight syringe for direct injection or enriched by adsorption to an adsorptive layer inserted into the headspace (commonly known as solid-phase micro extraction, (SPME)[52]). Adsorption layers can be silicone[53,54] or active charcoal layers[55] which enrich VOC from gas space. The injection occurs by thermal desorption. Dynamic procedures are also used for enrichment of VOC. In 'purge and trap'[56] the volatiles are purged by gas flow and trapped on adsorbent. In 'spray and trap'[57,58] the liquid sample is distributed by spray evaporation into a counter-current gas stream from which compounds are trapped, with the advantage that foaming of the sample is avoided.

Another simple sampling method for on-line monitoring of VOCs from environmental water samples/streams is membrane extraction.[17] When a membrane inlet is used an environmental sample can be pumped directly through a tubular membrane extraction unit, i.e. a separate sampling step is not required. This method has been used successfully for on-line monitoring of VOCs with a transportable mass spectrometer mounted in a boat[59] and for industrial wastewater monitoring.[60]

The enrichment of SVOC can take place on compact filters with an adsorptive surface onto which the sample is directly sucked (solid-phase extraction, SPE) using a plastic syringe[61] or filter disk.[62,63] The enriched compounds are eluted from the dried collection media by solvents and the extracts are injected. In addition, direct enrichment using an injector liner filled with small amounts of sorbent has been described.[64]

Liquid/liquid extraction (LLE) is the most frequently used laboratory procedure applied with accordingly smaller solvent quantities in the form of micro extraction[65,66] which is also used in the field.

3.3 Soil Samples

Direct thermal desorption with piston-type heaters pressed onto the soil surface has been described as a semiquantitative pre-screening procedure.[67–69] Thermal desorption of very small subsets (50–100 mg) without previous sample handling (e.g. homogenization) can also be used as a fast screening procedure.[70–72] However, owing to the small amount of analyzed material both procedures have the disadvantage of being less representative[73] and strongly influenced by matrix effects.[74] It appears to be better to collect discrete samples in the range of 5–50 g and analyze after simple cleanup steps.[75]

Fast, volumetric soil sampling can be performed with disposable plastic syringes (5–10 mL) with the tips cut off.[76,77] The syringe is filled by penetrating the opened cylinder into the soil. The material is transferred to a glass vial by pushing down the plunger. Sampling in polyethylene bags[78] is a simple procedure suitable when sampling highly contaminated material.

Field-suited procedures have been described in literature for the detection of VOCs and SVOCs. VOCs are detected by means of headspace analysis,[77–79] 'purge and trap',[80–83] extraction[84] or also by purging without sample treatment.[85] A modified membrane inlet mass spectrometric technique called purge-and-membrane mass spectrometry (PAM/MS) has recently been developed for the analysis of VOCs from soil samples.[86] In this method, VOCs are purged from soil samples with an inert gas and the stream is directed via a membrane inlet into a mass spectrometer for analysis.

For SVOCs, extraction is the simplest processing routine. To avoid time-consuming drying of the sample whilst at the same time reaching high yields, polar solvents such as methanol or acetone are used. Polar solvents bind the water content and overcome the binding forces between soil and the pollutant effectively.[87] The sample may be hand shaken,[88,89] applied to an ultrasonic bath[90] or extracted with a sonicator.[91]

4 MOBILE MASS SPECTROMETRY AND GAS CHROMATOGRAPHY/MASS SPECTROMETRY SYSTEMS

In principle, all routine laboratory mass spectrometers, except for large TOF mass spectrometers, double-focusing magnet mass spectrometers and Fourier transform systems, may be operated on-line[92,93] or in a mobile laboratory.[94–97] Systems presented here are mentioned in publications dealing with realistic on-site and field analysis or come from manufacturers whose area of application emphasizes the unique mobile measuring technique by advertisement (e.g. at Pittcon 1997 and 1998).

Table 1 On-line mass spectrometers

Cited in text as	Company	Address	Internet address	Model	Operation mode	Construction details[a]	Analyses[a]
MS1	VG/Fisons Instruments	32 Commerce Center, Cherry Hill Drive, Danvers, MA 01923, USA	http://www.vgscientific.com/scinew/welcome.htm	Leaktrace	direct		
MS2	Leybold/Inficon	Two Technology Place, East Syracuse, NY 13057, USA	http://www.inficon.com/	Transpector, QAS 100, MSQ 104, QX 200	direct		99
MS3	Ametek	Station Square, Paoli, PA 19301, USA	http://www.ametek.com/ametek/default.asp	Dycor QuadLink, System 1000	direct		
MS4	Balzers/Pfeiffer	24 Trafalgar Square, Nashua, NH 03063-1988, USA	http://www.pfeiffer-vacuum.com/welcome.html	QMS 200	direct		
MS5	Kogan, A.I. Ioffe Inst. (Russia)			Magnetic Mattauch-Herzog type	direct, (valve)	100, 101	
MS6	Home build Jet Propulsion Lab. device USA	quadrupole from Ferran Scientific	http://www.jpl.nasa.gov/	16 rod quadrupole array	direct, (valve)	102	
MS7	Teledyne	1274 Terra Bella Ave, Mountain View, Ca 94043, USA		3DQ Discovery	direct	103	104
MS8	Balzers/Pfeiffer	24 Trafalgar Square, Nashua, NH 03063-1988, USA	http://www.pfeiffer-vacuum.com/welcome.html	QMG 511	direct	105	
MS9	Finnigan MAT	Barkhausenstr. 2, D-28197 Bremen, Germany	http://www.finnigan.de/	ITS-40	direct	106	107, 108, 109
MS10	Balzers/Pfeiffer	24 Trafalgar Square, Nashua, NH 03063-1988, USA	http://www.pfeiffer-vacuum.com/welcome.html	QMG 421C, QMG 5M, IKR020, QMS 420, QMG 420	MIMS		97, 106, 111–116
MS11	Finnigan MAT	Barkhausenstr. 2, D-28197 Bremen, Germany	http://www.finnigan.de/	ITS-40, TSQ 4500, Magnum	MIMS		117, 118
MS12	Leda Mass (Spectra International)	380 Woodview Ave, Morgan Hill, CA 95037, USA	http://supersite.net/semiN2/spectra-rga/home.htm	Leda Mass	MIMS		119–121

(*continued overleaf*)

Table 1 (*continued*)

Cited in text as	Company	Address	Internet address	Model	Operation mode	Construction details[a]	Analyses[a]
MS13	VG/Fisons Instruments	32 Commerce Center, Cherry Hill Drive, Danvers, MA 01923, USA	http://www.vgscientific.com/scinew/welcome.htm	Survey	MIMS		
MS14	Monitor Group, (Div. of L.B. Foster)	290 East Union Rd., Cheswick, PA 15024, USA	http://www.lbfoster.com/	MG 2100	MIMS		
MS15	Hemond, Massachusetts Inst. of Technol. USA			Cycloid-MS	MIMS	122, 123	
MS16	Bruker-Franzen	Fahrenheitstr. 4, D-28359 Bremen, Germany	http://www.bruker.ca/products/ms/mobile.html	EM 640	MIMS	124	125
MS17	Kore Technology	Cambridge Science Park, CB4 4WF, UK	http://www.kore.co.uk	T-CAT	MIMS	126	104
MS18	Gard, Univ. of California, USA			ATOFMS	direct	127	
MS19	Perkin-Elmer/Sciex	SCIEX, 71 Four Valley Drive Concord, Ontario, Canada	http://www.iso9000directory.com/sciex.htm	TAGA 6000E/8000	direct	128–131	132–136
MS20	Bruker-Franzen	Fahrenheitstr. 4, D-28359 Bremen, Germany		TurboTOF	direct		137
MS21	V & F Analysentechnik	Maderspergerstr. 18, A-6060 Absam, Austria		Airsense 500	direct	138	139
MS22	Atomika	Bruckmannring 40, D-85764 Oberschleißheim/München, Germany	http://www.atomika.com/	IMR-MS 100	direct		140, 141
MS23	Hewlett Packard	2 Choke Cherry Road, Rockville, MD 20850, USA	http://www.hp.com/pressrel/mar98/02mar98q.htm	HP 4440	direct		142

[a] References.

4.1 On-line Mass Spectrometry

Table 1 lists the on-line mass spectrometers currently available. The simplest mass spectrometers are the leak detectors for gases with capillary vacuum inlet (MS1–4). The main application for these devices is to search for leakages in tanks or other closed systems. This is normally done by detecting a leaking out tracer gas (e.g. a fluorocarbon). Two other mass spectrometers, small, portable prototypes with magnetic and quadrupole array mass separation with valve inlet systems well suited for gas detection have been described (MS5–6) as well as mass spectrometers with differential inlet systems (MS7–9).

An ion trap mass spectrometer (MS7) has been evaluated by the United States Environmental Protection Agency (USEPA).[103] With this system, mounted in a van, VOCs may be measured on-line in MS/MS mode. In this mode ions of the mother spectrum are further fragmented by collision with a reaction gas. From these fragments the daughter spectrum is recorded, displaying structure details of the target ion from the other spectrum. MS/MS procedures enable compound identification even in complex mixtures without chromatographic separation. Measuring with MS/MS requires a rather complicated measuring strategy and is used when target compounds are known before the measurements start.

Membrane inlet mass spectrometers (MIMS) are listed under MS10–17. Particularly good portability is a feature of the devices dealt with in MS14 and MS15, which can be carried as luggage or as backpack devices problem-free by the analyst during the measurement. Two mobile TOF mass spectrometers have been designed, one with a very compact membrane inlet (MS17) and the other with a rather sophisticated direct inlet system (MS18). The latter operates with laser ionization to analyze aerosols as positive and negative ions simultaneously in two different sections of the mass spectrometers. Systems where special ionization techniques are used (atmospheric pressure ionization (API), atmospheric pressure chemical ionization (APCI) and ion molecule reaction (IMR), see below for details) to enhance the limit of detection limit (DL) are mentioned under MS19, MS21 and MS22.

The DL largely depends on the operation mode of the mass spectrometers. The lowest DLs are achieved in the selected ion monitoring (SIM) operation, since only few signals which are typical for the process are determined. This way substances up to a few ppb in air[115,110] and 500 ppq in water[117] were detected on-line. The disadvantage of SIM is that only target compounds are recorded and the device is ''blind'' to components which generate other mass fragments.

Membrane inlets are preferably used to analyze VOCs. If the sample supply as well as the membrane is heated (e.g. 250 °C in the case of MS16[125,143]) SVOCs such as polycyclic aromatic hydrocarbons (PAHs) are also detectable more or less on-line in short enrichment and desorption cycles.[124]

The DL can be lowered further by longer sample enrichment on a trap and thermal desorption injection. MIMS with pulse-like thermal desorption of the membrane[120,121] or collecting in a cryo minitrap[144,145] are described.

Very effective ionization methods such as API[146] and APCI[147] lead likewise to a higher sensitivity of the devices. This is combined with a high volume direct inlet which requires a high pumping speed which means large pumps and a high power consumption. Therefore these instruments (e.g. MS19) are limited to operation in large containers or mobile laboratories. On the other hand, polar and aggressive substances may be detected which predestines this device type for on-line measurement for the description of reactive trace compounds in the atmosphere[127,148,149] and polar indoor air pollutants.[136]

Other MS devices (MS20–22) are used as automated exhaust analyzing equipment e.g. in engine test stands. MS20 is a very fast TOF mass spectrometer with laser ionization while MS21 and MS22 are supplied with charge transfer or very smooth chemical ionization IMR.[140,141,150] The IMR ionization succeeds almost without fragmentation. On the basis of molecular ions a very fast simultaneous on-line detection of several components of the exhaust gas is achieved.

In the fast-growing market of gas sensor arrays (electronic noses[151]) new developments of mobile mass spectrometers (MS10, MS23 and GC/MS6 modified without GC) are used as highly specific on-line array detectors. Mass fragments of the emitted VOC are continuously recorded by scanning 100 to 200 masses and each single mass is considered as a highly specific detector. Using pattern recognition the resulting array is interpreted and compared with the reference pattern for identification of the sample. Examples of applications are the differentiation of beverages and foodstuffs.[116,142,152]

On-line mass spectrometers are used for process control in chemical or biological reactors with emission monitoring and control as well as efficient detectors for searching volatile pollutant emissions in the field. Looking for known compounds or mixtures, one or several target ions can be traced next to each other still allowing high time resolution. If unknown materials are affected the identification is often still possible in full-scan mode. If the identification is uncertain, the devices may take over warning functions and prove the absence of toxic materials (e.g. in worker protection[133,153,154]).

Table 2 GC/MS systems

Cited in text as	Company	Address	Internet address	Model	Operation mode	Construction details[a]	Analyses[a]
GC/MS1	University of Utah, Meuzelaar		http://www.utah.edu/marc/roving.htm	Modified HPMSD	direct	155, 156	157, 158
GC/MS2	University of Utah, Meuzelaar		http://www.utah.edu/marc/roving.htm	Modified Finnigan Ion Trap, "Minitmass"	direct	36, 159	
GC/MS3	Sinha	Jet Propulsion Lab., Pasadena, CA, USA		Magnetic Mattauch-Herzog MS	direct	25, 160	
GC/MS4	Inficon	East Syracuse, NY 13057, USA	http://www.techstuff.com/inficon.htm	Hapsite	MIMS		161
GC/MS5	Bruker-Franzen	Fahrenheitstr. 4, D-28359 Bremen, Germany	http://www.gulflink.osd.mil/foxnbc/n05_s03.htm# mm1	MEM	MIMS	35	68–70, 104, 161–166
GC/MS6	Bruker-Franzen	Fahrenheitstr. 4, D-28359 Bremen, Germany	http://www.bruker.ca/products/ms/mobile.html	EM 640	MIMS	168, 169	35, 37, 103, 125, 152, 164, 170, 171
GC/MS7	Viking	3800 Cocorde Parkway, Chantille, VA 20151, USA		Spectra Trak 572, 600, 620, 672	MIMS, direct	172, 173	14, 103, 133, 164, 170–176
GC/MS8	VG/Fisons Instruments	32 Commerce Center, Cherry Hill Drive, Danvers, MA 01923, USA	http://www.vgscientific.com/scinew/welcome.htm	MD 800	direct		177

[a] References.

4.2 Mobile Gas Chromatography/Mass Spectrometry Systems

Table 2 lists GC mass spectrometers currently available. The advantage of on-line MS to measure fast is almost preserved, if separation occurs on very short GC capillaries, in runs from few seconds, duration. The shortening of the chromatographic capillary to ca. 10% of the usual capillary length (of ca. 30 m to a few meters) only leads to a reduction of the separation power to ca. 30%.[178,179] Thus, despite the reduction of GC run time down to a duration between seconds and few minutes it is still possible to achieve separation powers with 10^3 to 10^5 theoretical plates.[155] Most common inner diameters for these short capillaries are below 0.2 mm (microbore capillaries). They are a direct connector between the vacuum part of the mass spectrometer and the injector. It is for this reason that we speak of 'transfer-line chromatography'.[36] Devices of this type have been constructed from parts of laboratory mass spectrometers (GC/MS1,2) or newly designed as a compact size magnet mass spectrometer (GC/MS3). Fast injection, which is essential for rapid separation, is handled by multiport valves[25] or with a special valveless ambient vapor sampling (AVS) inlet[36] (see Figure 7). The AVS system comprises three concentric tubes whose internal flow controls the vapor sampling process. The innermost tube is the transfer line column to the mass spectrometers. The intermediate tube is cut to extend ca. 2 cm beyond the end of the inner tube. In sampling mode all gas flows are directed towards the transfer line column. Injection is interrupted by switching the flow of the intermediate tube in counter-current direction. Thus, clean gas supplies the transfer line column and the sample gas is prevented from entering the mass spectrometers. Transfer-line mass spectrometers are most commonly used for the detection of VOCs. If the capillary is heated and injection occurs at 275 °C using the AVS it is also possible to analyze SVOCs (e.g. benzopyrene).[36] Transfer-line mass spectrometers are prototypes which are small in size and portable so that they can easily be carried by the analyst.[180]

A truly portable and commercially available GC mass spectrometer is the Hapsite (GC/MS4) with an analyzer weight of 16 kg. It consists of two parts: a pumping station and the analyzer module. For operation the mass spectrometer is evacuated once by an external pump. It is then possible to operate the analyzer module for up to three hours by battery by means of the membrane inlet, vacuum pumping by thermal getter and ion getter pumps and low temperature ($T < 80$ °C) of the gas ways. VOCs are sucked from a sniffer probe into a sample loop, injected by a valve and separated on a 30 m capillary GC column. This column is divided into a short pre-column and a separation column. After injection carrier gas backflush through a y-splitter between these columns prevents contamination of the GC mass spectrometer system. The capillaries may be heated isothermally up to 80 °C yielding for a maximum GC run time of about 10 min. The system has recently been evaluated by the USEPA.[161]

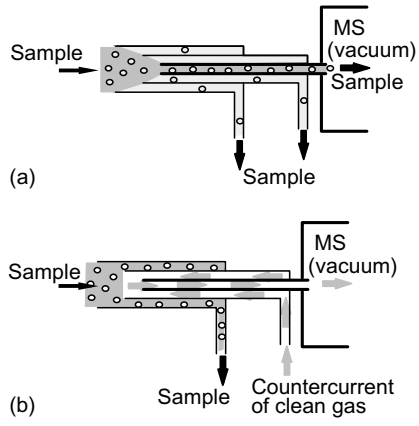

Figure 7 The AVS valveless injection system.

Transportable GC/MS units with a weight of 60 kg and above (GC/MS5–8) compared with portable ones offer almost the same analytical specifications as laboratory equipment. Such instruments are compact and can be transported in a van. GC/MS methods such as extract injection, thermo-desorption and additional on-line monitoring guided by analytical software are applicable. The inlet systems of transportable MS units are heated. Therefore all substances from VOCs to SVOCs may be analyzed as in conventional GC.

The mobile environmental monitor (MEM) (GC/MS5) is a quadrupole mass spectrometer which was modified from military applications during the 1980s.[181,182] It can handle strong vibrations such as those which occur during transportation in an off-road vehicle and remain fully operable. It is capable of recording spectra of 1–400 m/z every 0.5 s. The MS inlet membrane can be heated up to a temperature of 220 °C, the recipient is maintained at 180 °C and pumped solely by an ion getter pump. Three probes are available that can be easily mounted for quick exchange with plug-in connectors. Carrier gas is channelled through the probes by a pump located behind the membrane separator of the MS instrument.

Ambient air enters the injector and passes the GC column serving as the mobile phase for GC. Two of the three probes are provided with 5–15 m GC capillary columns and can be operated in a temperature program mode of up to 240 °C. One of the GC probes contains a thermal desorber for injection of adsorbent-enriched VOC and injection of extracts containing VOCs and SVOCs as well. From the other GC probe evaporable compounds

from untreated samples may be injected using a thermal desorption device which heats the sample between two pneumatically driven heaters.[183] The third probe is a sniffer sampling device to analyze gaseous compounds in the surroundings of the mass spectrometer. SVOCs are thermally desorbed by pressing the heated head of the probe against the surface of the sample. Separation is performed by means of transfer-line chromatography. The system was evaluated by USEPA in 1988.[69]

An improvement of the MEM towards higher compactness and a weight minimization to 65 kg is the quadrupole MS EM 640 (S) (GC/MS6). The EM 640 has undergone intensive USEPA testing during their Environmental Technology Verification Program (ETV).[103] A modification of this instrument, called the EM 640S, has a direct inlet together with a turbo-molecular pump and a roughing pump integrated in the system. Both versions are capable of recording spectra of 1–640 m/z every 0.5 s. SIM operation is also possible. A carrier gas control unit enables operation of the GC capillaries with externally attached gases with split/splitless injection or with a thermal desorber for adsorbents. The maximum temperature of the GC temperature program is 270 °C. The updated software permits multitasking. The probes consist of two compact GC modules with GC separation capillaries of different separation phase and lengths of up to 15 m. The GC probes can be easily mounted for quick exchange with plug-in connectors. The system was evaluated by USEPA in 1997.[103,171]

The GC/MS systems (GC/MS7) designed by the company Viking are accommodated in a suitcase or a sturdy housing. They are extremely compact units with a weight of ca. 60 kg. The device designed from a Hewlett Packard MSD is evacuated with an internal turbo-molecular pump and an additional mechanical pump attached externally over a vacuum tubing. Spectra within the mass range of 10–650 m/z can be recorded every 0.3 s. The different software-controlled inlet systems offer a high standard of convenient operation. In addition to a membrane inlet for gases the system has a direct inlet GC equipped with one split/splitless extract injector, an integrated enrichment unit for gaseous samples and a thermal desorber for adsorbent traps. GC capillaries up to 25 m and oven temperature up to 280 °C are used. The Spectra Trak 672 model was tested thoroughly in the context of the USEPA ETV Program.[103,176]

Within the EUREKA Project EU 674 new technologies for the application of mobile analytical laboratories are being developed. The VG/Fisons MD 800 (GC/MS8) is part of the device's equipment which is a commercial laboratory GC mass spectrometer being modified for mobile application by mounting it into a feed framework. Tests have been performed, but a report has not yet been presented.

Mobile GC/MS systems allow analyses to be carried out in the field with a quality which is hardly inferior to that of laboratory instruments. Because of the various analysis possibilities in on-line screening of VOC to SVOC GC/MS runs, these devices are designed for applications with highly toxic materials (e.g. chemical warfare agents) or scenarios which have a broad range of materials (e.g. in fire plumes, chemical emergencies or on unexplored chemical waste dumps).

5 QUALITY ASSURANCE/QUALITY CONTROL PROCEDURES AND COMPARISON OF ON-SITE/IN SITU MEASUREMENTS WITH OTHER METHODS/FIXED LABORATORY METHODS

The use of proper quality assurance/quality control (QA/QC) procedures together with analysis performed in a mobile laboratory and/or with mobile instruments is as important as when the analysis is performed in a stationary laboratory.[12] One could also easily argue that the QA/QC protocols are even more important in a mobile laboratory. This is partly due to the fact that the field conditions are assumed to be not as well controlled as the conditions in a stationary laboratory; for example the possible risk of contamination of instruments, blanks and samples may be higher in a mobile laboratory.[12] In addition, the users and developers of mobile instruments are responsible for using proper QA/QC methods in order to produce high quality data at all times which will accelerate the acceptance of the use of mobile instruments in environmental monitoring. This is very important since governmental offices and research institutes which stipulate environmental regulations are often reluctant to accept innovative techniques in place of the traditional well-known and thoroughly tested stationary laboratory instruments and methods.

The current literature on mobile mass spectrometers shows that in most cases the same customary QA/QC methods, such as initial and continuing calibrations, laboratory duplicates, labeled internal standards, matrix spikes and various types of blanks, as those used in stationary laboratories are also used with mobile mass spectrometers.[12,35,170,184–187] Some research groups are also using a few added QA/QC methods such as:

1. automatic daily tuning of the mass spectrometer;
2. using only disposable equipment for sampling and sample handling;
3. developing analytical methods adapted to the needs of on-site conditions;
4. analyzing a reference sample twice a day;

Table 3 Comparison of analytical results obtained by on-site GC/MS and off-site laboratory GC/MS results in analysis of VOCs from air[178] (concentrations in ppbv)

Compound	Sample 1 on-site	Sample 1 off-site	Sample 2 on-site	Sample 2 off-site	Sample 3 on-site	Sample 3 off-site
1,1,1-Trichloroethane	1.9	1.9	nd	nd	20	71
Trichloroethene	1.1	nd	3.2	1.8	43	30
Tetrachloroethene	12	4.4	8.7	3.4	19	9.0
Benzene	2.7	2.6	1.5	nd	2.8	2.5
Toluene	5.0	2.6	4.3	2.5	9.7	5.2
Xylenes	2.0	0.6	5.0	1.7	7.5	3.5
Styrene	1.0	nd	7.4	4.3	2.0	nd

nd = not detected.

5. keeping a control chart of these measurements in order to assist identification of possible instrumental problems or other errors.[35,185–186]

A continuous method for the mass spectrometer function check has also been introduced. In this method the signal of argon (m/z 40) being one component of the carrier gas is continuously followed in order to control the mass scale tuning and the sensitivity of the mass spectrometer.[35,185–186] An important aspect of the QA/QC protocols is their proper use and this will happen only if the personnel using mobile instruments have adequate knowledge and experience in using the instruments. This can only be achieved through proper training of the personnel as is currently the case, for example, with firefighters in Hamburg, Germany.[11]

Comparison of the quality of the data obtained using mobile mass spectrometers with the data obtained using stationary laboratory mass spectrometers or with the data obtainable using other mobile instruments is a very important part of the validation of the performance of mobile mass spectrometers.[35,67,68,76,104,108,171,172,174,184,186,188–192] These comparisons include determination of VOCs from air, soil gas, soil and water,[76,104,108,174,188,189] measurement of PCBs from soil and sediment,[68] measurement of PAHs from soil[35,67,184,186] measurement of phenols from soil[184,191] and measurement of pesticides from soil and sediment.[190] In addition, the USEPA has performed a thorough validation of two different commercial mobile GC mass spectrometers in the analysis of VOCs from water, soil and soil gas.[171,176] In most of these studies the data obtained using mobile mass spectrometers have been compared with the data obtained using stationary laboratory benchtop GC/MS instruments; however, comparisons with various GC methods and with a high-performance liquid chromatography (HPLC) method has also been made. Normally the agreement between the various methods has been very good; only in the case of analysis of VOCs has any significant discrepancy been observed. A simple explanation for the discrepancy is that owing to the large volatility of VOCs they are easily lost from the samples unless very careful sample storage, transportation and preparation methods are used. However, the USEPA instrument validation studies have shown that in the analysis of VOCs from various matrixes, mean accuracy is better than 40% and precision values better than 50% can be easily obtained.[171,176] In the following a few examples of these comparison studies are discussed in more detail.

In Tables 3 and 4 the results obtained in the analysis of VOCs from air and soil are presented, respectively. The on-site results presented in Table 3 were obtained by collecting the samples to charcoal sorbent tubes, thermally desorbing the analytes into a mobile GC/MS instrument and utilizing modified USEPA TO1 and TO2 methods.[174] The off-site results in Table 3 were obtained by collecting the samples to charcoal sorbent tubes, extracting the analytes by CS_2, injecting the liquid into a benchtop GC/MS instrument and using modified National Institute for Occupational Safety and Health (NIOSH) 1500, 1501 and 1003 methods. From the results presented it is easy to notice that the agreement between the data is good even though slightly different analytical methods were used. For some analytes the off-site results show slightly lower quantitative values, probably due to the evaporation of the analytes during the sample handling. The loss of VOCs due to sample preparation,

Table 4 Comparison of analytical results obtained by on-site HS/GC, on-site PT/GC/MS and off-site PT/GC/MS in analysis of trichloroethene in soil samples[76] (trichloroethene concentration in μg/g)

Sample	HS/GC, on-site	PT/GC/MS, on-site	PT/GC/MS, off-site
Bore hole 4/a	0.16	0.14	0.014
Bore hole 4/b	9.6	13	0.046
Bore hole 4/c	130	33	0.095
Bore hole 4/d	0.40	0.63	<0.0038
Bore hole 8/a	1.6	3.6	0.016
Bore hole 8/b	0.30	0.53	0.028
Bore hole 8/c	0.58	0.82	0.20

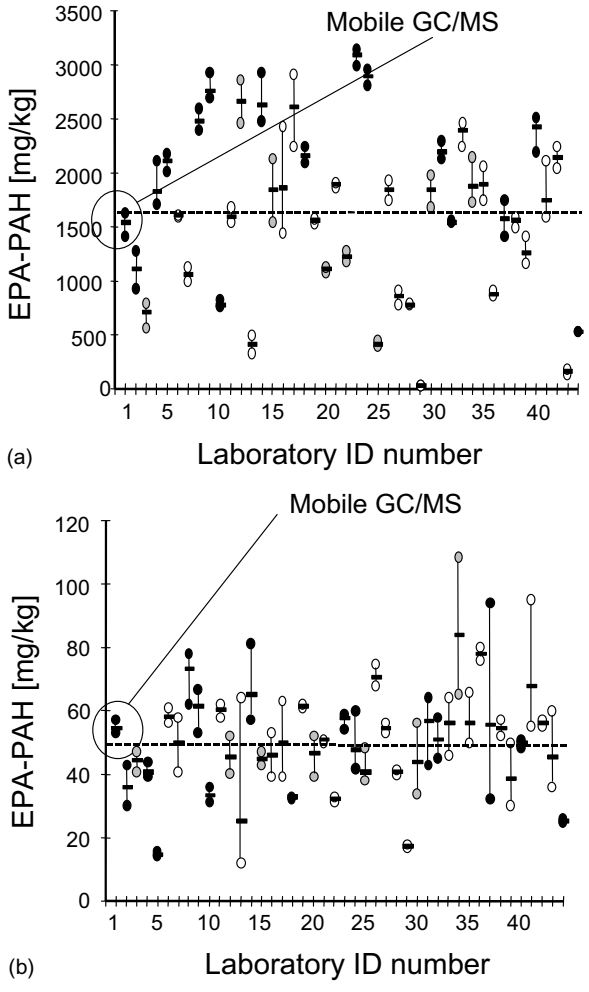

Figure 8 Summary of an interlaboratory comparison of 44 laboratories in analysis of PAHs from soil utilizing the mobile GC/MS method (sample i.d.1) and the stationary laboratory GC/MS: ●; GC/flame ionization detection: ◍; HPLC: ○ methods. Mean concentration: ----- all laboratories; ▬ each laboratory. A sum of 16 PAH compounds as defined in the USEPA method 610(1984) is presented.[35] (Reproduced from 'Fast GC/MS Field Screening for Excavation and Bioremediation of Contaminated Soil', G. Matz, W. Schröder, *Field Analyt. Chem. Technol.*, Copyright © (1996 John Wiley & Sons Inc.).)

storage and transportation is very clearly seen in the results presented in Table 4.[76] In the table quantitative results for trichloroethene measured from soil samples by on-site headspace gas chromatography (HS/GC) and on-site and off-site purge-and-trap gas chromatography/mass spectrometry (PT/GC/MS) are presented. Note that the agreement between the analytical results obtained on-site by HS/GC and PT/GC/MS is very good but the off-site PT/GC/MS results differ considerably from the on-site results. The only reasonable explanation for the difference was the loss of VOCs during the sample handling steps, which is a phenomenon observed quite frequently.[76,108,189] This fact is very clear proof that in some cases on-site measurements can provide higher quality data than the off-site stationary laboratories.

In the analysis of SVOCs the agreement between the results obtained using mobile mass spectrometers and the results obtained by various analytical methods in stationary laboratories are very good.[35,67,68,184,186,190,191] As a representative example of the very good agreement, the results of an interlaboratory comparison study in the analysis of PAHs from soil samples are presented in Figure 8.[35] There were 44 different laboratories participating in this study. The results obtained by the mobile GC/MS instrument agreed well with the results obtained by the other laboratories. Data obtained in the field are very close to the mean concentration of all the laboratories, i.e. the value which is most probably the correct concentration.

6 EXPERIENCE GAINED FROM ANALYTICAL CAMPAIGNS WITH MOBILE SYSTEMS

The actual efficiency of on-site analysis largely depends on the qualifications of the users or on their practice in the particular application. Complex sample preparation steps from the laboratory must be simplified and shortened. The sequence of QA/QC analyses must be a fixed constituent of the analytical process.[185] Standardized analytical methods[193] and software-guided methodologies from the sampling to the result of analysis[11] may offer assistance.

The diverse information supplied, particularly by mobile GC/MS systems, in the rapid sequential determination from different matrixes and types of analyte requires very good organization of the resulting data. Presentation of the results must be documented in such a way as to allows immediate access, including current status of the QA/QC analyses and all device parameters. Great effort towards developing non-standard software is necessary here.

The mobile GC mass spectrometer supplies important information to the sampling point at a high repetition rate, and this must be used immediately for good reason. So the analytical team easily becomes a decision-maker. Since the sampling, the work in excavation or handling of chemical emergencies is directly controlled by the result of the analysis, cooperation with the drilling team, geologist, environmental protecting agency or head of the fire fighting team must be clearly determined before the first analysis.

Analysis campaigns with mobile GC/MS have already demonstrated the advantages of analytical investigations in the field.[35] Excavation following the borders of contamination results in a drastic reduction in toxic material

which has to be disposed of, inhomogeneity of the investigated site can be described as deciding between sampling strategies and the success of bioremediation can be better proved by statistically based evaluation from multisampling and many analyses. With round robin tests and measurement campaigns parallel to laboratory investigations it has turned out that with VOCs more accurate results were obtained by local analysis[76,164,175] owing to problematic sample storage. Although, generally, mobile mass spectrometers are more expensive than appropriate laboratory instruments, savings of up to 70% can be made in relation to laboratory analysis by careful cost balancing over the whole analysis campaign.[104,175] Taking into account that 80% of the analysis error is derived from taking, transporting or storing the sample,[194] the application of mobile systems appears as the best means of more reliable environmental analyses.

7 CONCLUSION

In this article various aspects including sampling, analytical methods, quality assurance/control and instrumentation, of on-site/in situ environmental measurements using mobile mass spectrometers have been presented. It has also been shown that mobile mass spectrometers have produced valuable data in many different kinds of applications, e.g. monitoring of environmentally significant compounds from water, air and soil, and identification and quantification of possible toxic compounds produced in fires or released during chemical spills. The quality of the data obtainable using mobile mass spectrometers has also proved to be as good as obtainable with the traditionally used stationary laboratory GC/MS methods or with other chromatographic methods. In addition, mobile mass spectrometers currently commercially available are briefly presented.

The future will no doubt bring advances to all the sectors mentioned above but from the point of view of a research scientist the most interesting advances will surely take place in the instrumentation sector. There are at least three different very important mass spectrometer development areas, which are also heavily dependent on the development of interrelating scientific work such as the development of electronics, computer hardware and software, vacuum technology and battery technology.

One is the improvement of the sensitivity of the mass spectrometers. Looking back on the history of the development of mass spectrometers, we perceive that typically one order of magnitude better sensitivity has been obtained per decade of development of new instruments.[195] This same trend is also likely to continue in the future.

Another trend is that more and more versatile instruments have been produced. This is especially true when considering the development of quadrupole ion traps. This advance will also continue in respect to types of mass spectrometers other than the ion traps. One of the latest trends in the development of all types of instruments, including mass spectrometers, is miniaturization of them and building them on a silicon chip. For example, at the 46th ASMS Conference on MS and Allied Topics there was a special session about miniaturizing mass spectrometers [Session ''Miniaturizing Mass Spectrometers'', 46th ASMS Conference, Orlando, Florida, May 31–June 4, 1998]. In the session, designs of miniature magnetic sector, quadrupole and time-of-flight instruments were presented. On the other hand, mass spectrometer developers know that there are physical limitations which accompany miniaturization such as coulomb interactions limiting the ions' focusing, mass resolution and sensitivity, limits of the lowest detectable current, range of analytes dependent on heating and the power consumption. This development is moving quickly forward since already in this session data obtained with a quadrupole mass spectrometer made on a silicon chip has been presented.[196] The mass spectrometric development will produce mobile mass spectrometers that are capable of continuous long-term monitoring of environment, and truly miniaturized, i.e. hand-portable, and easy to use mobile mass spectrometers, which can be used for identification and quantification of unknown compounds in environmental accidents, fires, chemical spills and other situations where rapid reliable analytical results are required.

Improvement of the current, and development of new, sampling methods is another important research area in future. It is believed that especially sampling methods such as membrane inlets and other direct introduction methods will become more popular owing to their ease of use and the fact that with direct methods all the possible changes of the samples caused by the sample handling steps are minimized.

When all the foreseen development steps have been taken into consideration we can conclude that the use of mobile mass spectrometers will certainly increase considerably in the future in universities, research institutes and industrial laboratories. All the comparisons between mobile and laboratory GC/MS point out the advantages of field analysis with respect to quality, reliability, speed of analysis and costs. A definite milestone for the rapid increase in use of mobile mass spectrometers for environmental analysis will be the time when governmental or intergovernmental (e.g. European Union) agencies accept, introduce and demand standard methods for environmental analysis which are based on mobile mass spectrometers. In addition, the range of application areas of mobile mass spectrometers will

increase, for example one certain new application area will be direct measurement of biological warfare agents from various matrixes. In conclusion, developers, users and manufacturers of mobile mass spectrometers can look forward to a bright future.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
AVS	Ambient Vapor Sampling
DL	Detection Limit
ETV	Environmental Technology Verification
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HS/GC	Headspace Gas Chromatography
IMR	Ion Molecule Reaction
IR	Infrared
LLE	Liquid/Liquid Extraction
MEM	Mobile Environmental Monitor
MIMS	Membrane Inlet Mass Spectrometer
MS	Mass Spectrometry
MS/MS	Mass Spectrometry/Mass Spectrometry
NIOSH	National Institute for Occupational Safety and Health
PAH	Polycyclic Aromatic Hydrocarbons
PAM/MS	Purge-and-membrane Mass Spectrometry
PCB	Polychlorinated Biphenyl
PT/GC/MS	Purge-and-trap Gas Chromatography/Mass Spectrometry
QA/QC	Quality Assurance/Quality Control
SIM	Selected Ion Monitoring
SIS	Separation In Space
SIT	Separation In Time
SPE	Solid-phase Extraction
SPME	Solid-phase Micro Extraction
SVOC	Semivolatile Organic Compound
TOF	Time of Flight
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Field-based Analysis of Organic Vapors in Air

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Gas Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. L. Colin, D.M. Hunten, 'Pioneer Venus Experiment Descriptions', *Space Science Reviews*, **20**, 451–525 (1977).
2. N.C. Thomas, W.E. Crosmer, D. Nowak, 'A Study of the Feasibility of Mechanical Pumps for Use With the Pioneer-Venus Probe Mass Spectrometer Inlet System', AMES Research Center, National Aeronautics and

Space Administration, Report NASA CR114581 BSR 4036, 1973.
3. A.O. Nier, W.B. Hanson, M.B. McElroy, A. Seiff, N.W. Spencer, 'Entry Science Experiments for Viking 1975', *ICARUS*, **16**, 74–91 (1972).
4. D.R. Rushneck, A.V. Diaz, D.W. Howarth, J. Rampacek, K.W. Olson, W.D. Dencker, P. Smith, L. McDavid, A. Tomassian, H. Bulota, 'Viking Gas Chromatograph-Mass Spectrometer', *Rev. Sci. Instrum.*, **49**, 817–834 (1978).
5. U.v. Zahn, K. Mauersberger, 'Small Mass Spectrometer with Extended Measurement Capabilities at High Pressures', *Rev. Sci. Instrum.*, **49**, 1539–1542 (1978).
6. E. Sagawa, T. Itoh, 'Mass Spectrometric Observation of SO_2 in the Stratosphere', *Geophys. Res. Lett.*, **4**, 29–32 (1977).
7. J.E. Evans, J.T. Arnold, 'Monitoring Organic Vapors', *Environ. Sci. Technol.*, **9**, 1134–1137 (1975).
8. 'Wiley Mass Spectral Database', 5th edition, John Wiley & Sons, New York, 1989.
9. 'NIST/EPA/NIH Mass Spectral Database', (NIST '98), Version 1.6d, US Department of Commerce, 1998.
10. T. Kotiaho, 'On-site Environmental and In situ Process Analysis by Mass Spectrometry', *J. Mass Spectrom.*, **31**, 1–15 (1996).
11. G. Matz, W. Schröder, A. Harder, A. Schillings, P. Rechenbach, 'Fast On-site GC/MS Analysis of Hazardous Compound Emissions from Fire and Chemical Accidents', *Field Anal. Chem. Technol.*, **1**(4), 181–194 (1997).
12. E.N. Koglin, L.R. Williams, 'Testing, Performance Validation and Quality Assurance/Quality Control of Field-Portable Instrumentation', *Trends Anal. Chem.*, **13**, 294–299 (1994).
13. F.W. Karasek, 'Basic Gas Chromatography – Mass Spectrometry, Principles and Techniques', Elsevier, Amsterdam, 201, 1988.
14. M.R. Walsh, M.A. LaPack, 'On-line Measurements using Mass Spectrometry', Proc. Forty-ninth Annual Symp. on Instrum. for the Process Industry, 97–107, Instrument Society of America, 1994.
15. M. Linscheid, 'Instrumental Developments in Organic Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **337**, 648–661 (1990).
16. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', *Anal. Chem.*, **12**, 599–651 (1996).
17. F.R. Lauritsen, T. Kotiaho, 'Advances in Membrane Inlet Mass Spectrometry (MIMS)', *Rev. Anal. Chem.*, 1–44 (1997).
18. J.V. Hinshaw, 'Bench-top Mass Spectrometric Detectors', *LC-GC Int.*, Jan. 1995, **8**, 22–26 (1995).
19. J. Allison, R.M. Stepnowski, 'The Hows and Whys of Ion Trapping', *Anal. Chem.*, **59**, 1072A–1088A (1987).
20. T. Kotiaho, F.R. Lauritsen, T.K. Choudhury, R. Graham, G.T. Tsao, 'Membrane Introduction Mass Spectrometry', *Anal. Chem.*, **63**, 875A–883A (1991).
21. S. Bauer, 'Membrane Introduction Mass Spectrometry, an Old Method that is Gaining New Interest Through Recent Technological Advances', *Trends Anal. Chem.*, **14**, 202–213 (1995).
22. S.A. McLuckey, G.J. Van Berkel, D.E. Goeringer, G.L. Glish, 'Ion Trap Mass Spectrometry of Externally Generated Ions', *Anal. Chem.*, **66**, 689A–696A (1994).
23. R.E. March, 'Ion trap mass spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **118**, 71–135 (1992).
24. T.L. Sheehan, 'GC-MS-MS: The Next Logical Step in Benchtop GC-MS', *International Laboratory*, March/'97, 11A–11D (1997).
25. M.P. Sinha, G. Gutnikov, 'Development of a Miniaturized Gas Chromatograph-Mass Spectrometer with a Microbore Capillary Column and an Array Detector', *Anal. Chem.*, **63**, 2012–2016 (1991).
26. R. Cotter, 'Time-of-flight Mass Spectrometry for the Analysis of Biological Molecules', *Anal. Chem.*, **64**, 1027–1039 (1992).
27. P.J. Baugh, *Gas Chromatography*, IRL Press at Oxford University, Oxford, 1993.
28. T. Maeda, H. Tatematsu, F. Morishita, 'Capillary Gas Chromatography Using Carbon Dioxide as the Carrier Gas', *Anal. Science*, **7**, (Suppl., Proc. Int. Congr. Anal. Sci., 1991, Pt. 1), 219–222 (1991).
29. V.G. Berezkin, I.V. Malyukova, V.R. Alishoev, 'The Influence of the Carrier Gas on Retention in Capillary Gas–Solid Chromatography', *J. High Resolut. Chromatogr.*, **19**, 272–276 (1996).
30. V.G. Berezkin, I.V. Malyukova, V.R. Alishoev, 'Dependence of Retention of Organic Compounds on Carrier Gas Pressure in Capillary Adsorption Chromatography', *J. Chromatogr., A*, **699**, 155–159 (1995).
31. Y. Shen, M.L. Lee, 'High Speed Solvating Gas Chromatography Using Packed Capillaries Containing Sub-5μm Particles', *J. Chromatogr., A*, **778**, 31–42 (1997).
32. Y. Shen, M.L. Lee, 'High-Speed Gas Chromatography Using Packed Capillary Columns', *J. Microcolumn Sep.*, **9**(1), 21–27 (1997).
33. Y. Shen, M.L. Lee, 'Separation of Polar Compounds by Solvating Gas Chromatography with Carbon Dioxide Mobile Phase', *Chromatographia*, **46**, 587–593 (1997).
34. C. Henry, 'Taking the Show on the Road, Portable GC and GC/MS', *Anal. Chem.*, (News & Features), 195A–200A (1997).
35. G. Matz, W. Schröder, 'Fast GC/MS Field Screening for Excavation and Bioremediation of Contaminated Soil', *Field Anal. Chem. Technol.*, **1**, 77–85 (1996).
36. N.S. Arnold, W.H. McClennen, H.L.C. Meuzelaar, 'Vapor Sampling Device for Direct Short Column Gas Chromatography/Mass Spectrometry Analyses of Atmospheric Vapors', *Anal. Chem.*, **63**, 299–304 (1991).
37. B. Nölke, G. Baykut, 'Analysis of Volatile Organic Compounds Using a Mobile GC/MS', Field Screening

Methods for Hazardous Wastes and Toxic Chemicals, Proc. of an International Symposium, Air and Waste Management Association, Las Vegas, Nevada, **2**, 911–918, 1995.
38. P. Hindmarch, C. Demir, R.G. Brereton, 'Deconvolution and Spectral Clean-up of Two-component Mixtures by Factor Analysis of Gas Chromatographic-Mass Spectrometric Data', *Analyst*, **121**, 993–1001 (1996).
39. Y.V. Gankin, A. Gorshteyn, S. Smarason, A. Robbat, 'Time-Condensed Analyses by Mass Spectrometry', *Anal. Chem.*, **70**, 1655–1663 (1998).
40. J.B. Phillips, D. Luu, J.B. Pawliszyn, 'Multiplex Gas Chromatography by Thermal Modulation of a Fused Silica Capillary Column', *Anal. Chem.*, **57**, 2779–2787 (1985).
41. M.J. Yang, J. Pawliszyn, 'Multiplex Gas Chromatography: A Practical Approach for Environmental Monitoring', *Trends Anal. Chem.*, **15**, 273–278 (1996).
42. J.R. Valentin, 'Multiplex Gas Chromatography', *LC-GC*, **2**, 32–41 (1989).
43. J.M. Andino, J.W. Butler, 'A Study of the Stability of Methanol-Fueled Vehicle Emissions in Tedlar™ Bags', *Environ. Sci. Technol.*, **25**, 1644–1646 (1991).
44. Y. Wang, T.S. Raihala, A.P. Jackman, R.S. John, 'Use of Tedlar™ Bags in VOC Testing and Storage: Evidence of Significant VOC Losses', *Environ. Sci. Technol.*, **30**, 3115–3117 (1996).
45. D.A. Brymer, L.D. Ogle, C.J. Jones, D.L. Lewis, 'Viability of Using SUMMA Polished Canisters for the Collection and Storage of Parts per Billion by Volume Level Volatile Organics', *Environ. Sci. Technol.*, **30**, 188–195 (1996).
46. C. Morris, R. Berkley, J. Bumgarner, 'Preparation of Multicomponent Volatile Organic Standards Using Static Dilution Bottles', *Anal. Lett.*, **16**, 1585–1593 (1983).
47. S.F. Patil, S.R. Lonkar, 'Evaluation of Tenax™ TA for the Determination of Chlorobenzene and Chloronitrobenzenes in Air Using Capillary Gas Chromatography and Thermal Desorption', *J. Chromatogr., A*, **684**, 133–142 (1994).
48. W.A. McClenny, K.D. Oliver, E.H. Daughtrey, 'Analysis of VOCs in Ambient Air Using Multisorbent Packings for VOC Accumulation and Sample Drying', *Air & Waste Management Assoc.*, **45**, 792–800 (1995).
49. C.J. Koester, R.E. Clement, 'Analysis of Drinking Water for Trace Organics', *Crit. Rev. Anal. Chem.*, **24**, 263–316 (1993).
50. L. Starostin, Z. Witkiewicz, 'Environmental Water Samples Preparation for Chemical Analysis', *Chem. Anal.*, **39**, 263–279 (1994).
51. D. Barcelo, M.C. Hennion, 'Sampling of Polar Pesticides from Water Matrices', *Anal. Chim. Acta*, **338**, 3–18 (1997).
52. C.L. Arthur, K. Pratt, S. Motlagh, J. Pawliszyn, R.P. Belardi, 'Environmental Analysis of Organic Compounds in Water using Solid Phase Micro Extraction', *J. High Resolut. Chromatog.*, **15**, 741–744 (1992).
53. T. Gorecki, J. Pawliszyn, 'Field-portable Solid-phase Microextraction/Fast GC System for Trace Analysis', *Field Anal. Chem. Technol.*, Proc. of an International Symposium, Las Vegas, Nevada, **1**, 277–284 (1997).
54. T. Gorecki, J. Pawliszyn, 'Sample Introduction Approaches for Solid Phase Microextraction/Rapid GC', *Anal. Chem.*, **67**, 3265–3274 (1995).
55. D. Djozan, Y. Assadi, 'A New Porous-layer Activated-charcoal-coated Fused Silica Fiber: Application for Determination of BTEX Compounds in Water Samples using Headspace Solid-phase Microextraction and Capillary Gas Chromatography', *Chromatographia*, **45**, 183–189 (1997).
56. S.M. Abeel, A.K. Vickers, D. Decker, 'Trends in Purge and Trap', *J. Chromatogr., Science*, **32**, 328–338 (1994).
57. G. Matz, P. Kesners, 'Spray and Trap Method for Water Analysis by Thermal Desorption Gas Chromatography/Mass Spectrometry in Field Applications', *Anal. Chem.*, **65**, 2366–2371 (1993).
58. G. Baykut, A. Voigt, 'Spray Extraction of Volatile Organic Compounds from Aqueous Systems into the Gas Phase for Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **64**, 677–681 (1992).
59. B.J. Harland, P.J. Nicholson, 'Continuous Measurement of Volatile Organic Chemicals in Natural Waters', *Sci. Total Environ.*, **135**, 37–54 (1993).
60. T. Kotiaho, R. Kostiainen, R.A. Ketola, T. Mansikka, I. Mattila, V. Komppa, T. Honkanen, K. Wickström, J. Waldvogel, O. Pilviö, 'Development of a Fully Automatic Membrane Inlet Mass Spectrometric System for On-line Industrial Waste Water Monitoring', *Proc. Control and Qual.*, **11**, 71–78 (1998).
61. I. Rodriguez, R. Cela, 'Combination of Solid-phase Extraction Procedures with Gas Chromatographic Hyphenated Techniques for Chlorophenol Determination in Drinking Water', *Trends Anal. Chem.*, **16**, 463–475 (1997).
62. D.L. Mayer, J.S. Fritz, 'Semi-micro Solid-phase Extraction of Organic Compounds from Aqueous and Biological Samples', *J. Chromatogr., A*, **773**, 189–197 (1997).
63. J. Schülein, D. Martens, P. Spitzauer, A. Kettrup, 'Comparison of Different Solid Phase Extraction Materials and Techniques by Application of Multiresidue Methods for the Determination of Pesticides in Water by High-Performance Liquid Chromatography (HPLC)', *Fresenius' J. Anal. Chem.*, **352**, 565–571 (1995).
64. E. Baltussen, H.-G. Janssen, P. Sandra, C.A. Cramers, 'A Novel Type of Liquid Liquid Extraction for the Preconcentration of Organic Micropollutants from Aqueous Samples: Application to the Analysis of PAH's and OCP's in Water', *J. High Resolut. Chromatogr.*, **20**, 395–399 (1997).

65. C.S. Barrio, E.R. Melgosa, J.S. Asensio, J.G. Bernal, 'Extraction of Pesticides from Aqueous Samples: a Comparative Study', *Mikrochimica Acta*, **122**, 267–277 (1996).
66. T. Welsch, H. Block, 'Separation and Enrichment of Traces of Explosives and their By-products from Water by Multiple Micro Liquid Extraction for their Determination by Capillary Gas Chromatography', *Fresenius' J. Anal. Chem.*, **357**, 904–908 (1997).
67. A. Robbat, T.Y. Liu, B.M. Abraham, 'On-site Detection of Polycyclic Aromatic Hydrocarbons in Contaminated Soils by Thermal Desorption Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **64**, 1477–1483 (1992).
68. A. Robbat, T.-Y. Liu, B.M. Abraham, 'Evaluation of a Thermal Desorption Gas Chromatograph/Mass Spectrometer: On-site Detection of Polychlorinated Biphenyls at a Hazardous Waste Site', *Anal. Chem.*, **64**, 358–364 (1992).
69. T.M. Trainor, F.H. Laukien, 'Design and Performance of a Mobile Mass Spectrometer Developed for Environmental Field Investigation', USEPA, Res. Dev., (Rep.) EPA, EPA/600/D-89/189, Field Screening Methods Hazard. Waste Site Invest., 299–310, Washington DC, 1988.
70. B. Wiezorek, A. Schiller, G. Baykut, K.P. Wanczek, 'An Infrared Pyrolyzer for Analysis of Solid Samples by Gas Chromatography/Mass Spectrometry', *Rev. Sci. Instrum.*, **63**, 5607–5612 (1992).
71. D.K. Banerjee, M.R. Gray, 'Analysis of Hydrocarbon-contaminated Soil by Thermal Extraction/Gas Chrommatography', *Environ. Sci. Technol.*, **31**, 646–650 (1997).
72. M.J. Geerdink, C. Erkelens, J.C. van Dam, J. Frank, K.Ch.A.M. Luyben, 'Fast Screening of Contaminated Soil Samples Using Thermal Desorption Mass Spectrometry', *Anal. Chim. Acta*, **315**, 159–166 (1995).
73. G.U. Fortunati, M. Pasturenzi, 'Quality in Soil Sampling', *Quimica Analytica*, **13**, S5–S20 (1994).
74. I. Scheunert, M. Mansour, 'Binding of Organic Pollutants to Soil Organic Matter', *Int. J. Environ. Anal. Chem.*, **46**, 189–199 (1991).
75. M.J. Gardner, 'Sample Heterogeneity and Analytical Precision: How Much Sample Should I Take for Analysis?', *Anal. Proc. Anal. Comm.*, **32**, 115–116 (1995).
76. A.D. Hewitt, 'Comparison of Methods for Sampling Vadose Zone Soils for Determination of Trichloroethylene', *J. AOAC Int.*, **77**, 458–463 (1994).
77. A.D. Hewitt, P.H. Mlyares, D.C. Leggett, T.F. Jenkins, 'Comparison of Analytical Methods for Determination of Volatile Organic Compounds in Soils', *Environ. Sci. Technol.*, **26**, 1932–1938 (1992).
78. G.A. Robbins, R.D. Bristol, V.D. Roe, 'A Field Screening Method for Gasoline Contamination using a Polyethylene Bag Sampling System', *Groundwater Monitoring Review*, **9**, 87–97 (1989).
79. O.R. West, R.L. Siegrist, T.J. Mitchell, R.A. Jenkins, 'Measurement Error and Spatial Variability Effects on Characterization of Volatile Organics in the Subsurface', *Environ. Sci. Technol.*, **29**, 647–656 (1995).
80. A.P. Bianchi, M.S. Varney, J. Phillips, 'Analysis of Volatile Organic Compounds in Estuarine Sediments using Dynamic Headspace and Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **542**, 413–450 (1991).
81. O.C. Amaral, L. Olivella, J.O. Grimalt, J. Albaiges, 'Combined Solvent Extraction/Purge and Trap Method for the Determination of Volatile Organic Compounds in Sediments', *J. Chromatogr.*, **675**, 177–187 (1994).
82. T.C. Voice, B. Kolb, 'Comparison of European and American Techniques for the Analysis of Volatile Organic Compounds in Environmental Matrices', *J. Chromatogr. Sci.*, **32**, 306–311 (1994).
83. M.J. Charles, M.S. Simmons, 'Recovery Studies of Volatile Organics in Sediments Using Purge/Trap Methods', *Anal. Chem.*, **59**, 1217–1221 (1987).
84. M.D.F. Askari, M.P. Maskarinec, S.M. Smith, P.M. Beam, C.C. Travis, 'Effectiveness of Purge-and-trap for Measurement of Volatile Organic Compounds in Aged Soils', *Anal. Chem.*, **68**, 3431–3433 (1996).
85. Y. Yokouchi, M. Sano, 'Trace Determination of Volatile Organic Compounds in Soil Based on Thermal Vaporization Followed by Tenax™/GC Trapping and Capillary Gas Chromatography/Mass Spectrometry', *J. Chromatogr., A*, **555**, 297–301 (1991).
86. R. Kostiainen, T. Kotiaho, I. Mattila, T. Mansikka, M. Ojala, R.A. Ketola, 'Analysis of Volatile Organic Compounds in Water and Soil Samples by Purge-and-membrane/Mass Spectrometry', *Anal. Chem.*, **70**, 3028–3032 (1998).
87. F.M. Dunnivant, A.W. Elzerman, 'Determination of Polychlorinated Biphenyls in Sediments, Using Sonication Extraction and Capillary Column Gas Chromatography–Electron Capture Detection with Internal Standard Calibration', *J. AOAC Int.*, **71**, 551–556 (1988).
88. C. Keuchel, R. Niessner, 'Rapid Field Screening Test for Determination of 2,4,6-Trinitrotoluene in Water and Soil with Immunofiltration', *Fresenius' J. Anal. Chem.*, **350**, 538–543 (1994).
89. R.J. Van Delft, A.S.M. Doveren, A.G. Snijders, 'The Determination of Hydrocarbons in Soil Using a Miniaturized Extraction Method and Gas Chromatography', *Fresenius' J. Anal. Chem.*, **350**, 638–641 (1994).
90. A. Eschenbach, M. Kästner, R. Bierl, G. Schaefer, B. Mahro, 'Evaluation of a New, Effective Method to Extract Polycyclic Aromatic Hydrocarbons from Soil Samples', *Chemosphere*, **28**, 683–692 (1994).
91. 'Test Methods for Evaluation Solid Waste' 3rd edition, USEPA EPA SW-846, Method 3550, 10, Washington DC, 1990.
92. T.N. Corso, C.K. van Pelt, T. Brenna, 'Gas Chromatograph Injection Liner for Continuous Analyte Admission into a Mass Spectrometer', *Anal. Chem.*, **70**, 1030–1032 (1998).
93. M. Koch, 'On the Conversion of a Mass Selective Detector from Gas Chromatography/Mass Spectrometry Application to Stand-alone, On-line, Real-time

Mass Spectrometry Application', *Rev. Sci. Instrum.*, **65**, 2808–2818 (1994).
94. P.H. Hemberger, J.E. Alarid, D. Cameron, C.P. Leibmann, T.M. Cannon, M.A. Wolf, R.E. Kaiser, 'A Transportable Turnkey Gas Chromatograph/Ion Trap Detector for Field Analysis of Environmental Samples', *Int. J. Mass Spectrom. Ion Processes*, **106**, 299–313 (1991).
95. M. Heyl, 'Globetrotting with a Modular GC/MS Laboratory', *Field Anal. Chem. Technol.*, **1**, 3–12 (1996).
96. M.J. Duane, S. Facchetti, 'On-site Environmental Water Analyses by ICP-MS', *Sci. Tot. Environ.*, **172**, 133–144 (1995).
97. V.T. Virkki, R.A. Ketola, M. Ojala, T. Kotiaho, V. Komppa, A. Grove, S. Facchetti, 'On-site Environmental Analysis by Membrane Inlet Mass Spectrometry', *Anal. Chem.*, **67**, 1421–1425 (1995).
98. K. Jiao, A. Robrat, 'Performance-based Field Methods for Analyzes of Substituted Phenols by Thermal Desorption Gas Chromatography/Mass Spectrometry', *J. AOAC Int.*, **79**, 131–142 (1996).
99. M. Betz, 'In-situ-Messung von Deponiegasen mittels mobilem Massenspektrometer', *Technisches Messen*, **55**, 177–180 (1988).
100. V.T. Kogan, A.K. Pavlov, Y.V. Chichagov, Y.V. Tuboltsev, G.Y. Gladkiv, A.D. Kazanskii, V.A. Nikolaev, R. Pavlichkova, 'Design and Testing of a Portable Magnetic Mass Spectrometer', *Field Anal. Chem. Technol.*, **1**, 331–342 (1997).
101. V.T. Kogan, A.D. Kazanskii, A.K. Pavlov, 'Portable Mass Spectrometer For Environmental Monitoring', *Instruments and Experimental Techniques*, **38**, 106–111 (1995).
102. O.J. Orient, A. Chutjian, V. Garkanian, 'Miniature, High-resolution, Quadrupole Mass-spectrometer Array', *Rev. Sci. Instrum.*, **68**, 1393–1397 (1997).
103. 'GCMS Demonstration Plan', USEPA, Environmental Technology Evaluation Program, ETV, [http://www.epa.gov/etvprgrm/02/gcmsdpl.htm], Washington DC, 1997.
104. M.B. Wise, C.V. Thompson, R. Merriweather, M.R. Guerin, 'Review of Direct MS Analysis of Environmental Samples', *Field Anal. Chem. Technol.*, **1**, 251–276 (1997).
105. M.A. DesJardin, S.J. Doherty, J.R. Gilbert, M.A. LaPack, J. Shao, 'Better Understanding of Plant and Pilot Plant Operations Using On-line Mass Spectrometry', *Process Control and Quality*, **6**, 219–227 (1995).
106. L.E. Dejarme, S.J. Bauer, R.G. Cooks, F.R. Lauritsen, R. Kotiaho, T. Graf, 'Jet Separator/Membrane Introduction Mass Spectrometry for On-line Quantitation of Volatile Organic Compounds in Aqueous Solutions', *Rapid Commun. Mass Spectrom.*, **7**, 935–942 (1993).
107. M.B. Wise, M.V. Buchmann, C.V. Thompson, M.R. Guerin, 'Environmental Field Measurements with an Ion Trap Mass Spectrometer', Proc of the 40th ASMS Conference on Mass Spectrometry and Allied Topics, Washington DC, 500–502, 1992.
108. M.B. Wise, C.V. Thompson, M.V. Buchanan, R. Merriweather, M.R. Guerin, 'Direct Sampling Ion Trap Mass Spectrometry', *Spectroscopy*, **8**, 14–22 (1993).
109. M.B. Wise, C.V. Buchmann, D.E. Thompson, S.A. Goeringer, R.I. Crutcher, M.L. Hileman, M.L. Emery, G.T. Alley, 'Consideration for the Size and Power Reduction of a Field Transportable Ion Trap Mass Spectrometer', Proc. of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, 438a–438b, 1993.
110. R.A. Ketola, T. Mansikka, M. Ojala, T. Kotiaho, R. Kostiainen, 'Analysis of Volatile Organic Sulfur Compounds in Air by Membrane Inlet Mass Spectrometry', *Anal. Chem.*, **69**, 4536–4539 (1997).
111. M. Ojala, R. Ketola, T. Mansikka, R. Kostiainen, 'Detection of Volatile Organic Sulfur Compounds in Water by Headspace Gas Chromatography and Membrane Inlet Mass Spectrometry', *J. High Resolut. Chromatogr.*, **20**, 165–169 (1997).
112. M. Ojala, R.A. Ketola, V. Virkki, H. Sorsa, T. Kotiaho, 'Determination of Phenolic Compounds in Water using Membrane Inlet Mass Spectrometry', *Talanta*, **44**, 1253–1259 (1997).
113. M.A. LaPack, J.C. Tou, C.G. Enke, 'Membrane Extraction Mass Spectrometry for the On-line Analysis of Gas and Liquid Process Streams', *Anal. Chem.*, **63**, 1631–1637 (1991).
114. F.R. Lauritsen, S. Gylling, 'On-line Monitoring of Biological Reactions at Low Parts-per-trillion Levels by Membrane Inlet Mass Spectrometry', *Anal. Chem.*, **67**, 1418–1420 (1995).
115. R.A. Ketola, M. Ojala, H. Sorsa, T. Kotiaho, R.K. Kostiainen, 'Development of a Membrane Inlet Mass Spectrometric Method for Analysis of Air Samples', *Anal. Chim. Acta*, **349**, 359–365 (1997).
116. R.A. Ketola, J. Heikkonen, S. Pieponen, F.R. Lauritsen, T. Kotiaho, 'Classification of Cola Beverages on the Basis of Mass Spectra Measured by Membrane Inlet Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**, 1011–1017 (1998).
117. M. Soni, S. Bauer, J.W. Amy, P. Wong, R.G. Cooks, 'Direct Determination of Organic Compounds in Water at Parts-per-quadrillion Levels by Membrane Introduction Mass Spectrometry', *Anal. Chem.*, **67**, 1409–1412 (1995).
118. 'Emission Measurement: Measurement of Polycylic Aromatic Hydrocarbons (PAH) in Stationary Industrial Plants – Dilution Method (RWTÜV-Method) Gaschromatographic Determination', VDI-Norm 3873, Part 1, Beuth Verlag, Berlin, 1992.
119. T. Meyer-Jens, G. Matz, H. Märkl, 'On-line Measurement of Dissolved and Gaseous Hydrogen Sulfide in Anaerobic Biogas Reactors', *Appl. Microbiol. Biotechnol.*, **43**, 341–345 (1995).

120. G. Matz, F. Lennemann, 'On-line Monitoring of Biotechnological Processes by Gas Chromatographic/Mass Spectrometric Analysis of Fermentation Suspensions', *J. Chromatogr., A*, **750**, 141–149 (1996).
121. G. Matz, M. Loogk, F. Lennemann, 'On-line Gas Chromatography – Mass Spectrometry for Process Monitoring Using Solvent-free Sample Preparation', *J. Chromat. A*, **819**, 51–60 (1998).
122. H.F. Hemond, 'A Backpack-portable Mass Spectrometer for Measurement of Volatile Compounds in the Environment', *Rev. Sci. Instrum.*, **62**, 1420–1425 (1991).
123. M.C. Hadka, R.K. Dickinson, 'Rapid Screening of Organic Contaminants Using a Mobile Mass Spectrometer in the Field', USEPA, Res. Dev., (Rep.) EPA, 600/D-89/189, Field Screening Methods Hazard. Waste Site Invest., Washington DC, 423–432, 1988.
124. A. Walte, *On-line Analyze gasförmiger und partikelgebundener Aromaten in Abgasen mit einem mobilen Massenspektrometer*, Beuth Verlag, Berlin, 1994.
125. A. Walte, G. Matz, W. Münchmeyer, 'On-line Abgasanalyze von Aromaten mit mobilem GC-MS', *WLB Wasser Luft und Boden*, **3**, 54–57 (1995).
126. A.J. White, M.G. Blamine, C.A. Corlett, B.W. Griffiths, D.M. Martin, S.B. Spencer, S.J. Mullock, 'Development of a Portable Time-of-flight Membrane Inlet Mass Spectrometer for Environmental Analysis', *Rev. Sci. Instrum.*, **69**, 565–571 (1998).
127. E. Gard, J.E. Mayer, B.D. Morrical, T. Dienes, D.P. Fergenson, K.A. Prather, 'Real-time Analysis of Individual Atmospheric Aerosol Particles: Design and Performance of a Portable ATOFMS', *Anal. Chem.*, **69**, 4083–4091 (1997).
128. R.W. Bell, G.B. De Brou, 'Mobile Ambient Air Monitoring', *Proc. 33rd Int. Ind. Waste Conf.*, **33**, 103–122 (1986).
129. D.A. Lane, 'Mobile Mass Spectrometry', *Environ. Sci. Technol.*, **16**, 38A–46A (1982).
130. S.N. Ketkar, J.G. Dulak, W.L. Fite, J.D. Buchner, S. Dheandhanoo, 'Atmospheric Pressure Ionization Tandem Mass Spectrometric System for Real-time Detection of Low-level Pollutants in Air', *Anal. Chem.*, **61**, 260–264 (1989).
131. J. Sunner, M.G. Ikonomou, P. Kebarle, 'Sensitivity Enhancements Obtained at High Temperatures in Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **60**, 1308–1313 (1988).
132. T.J. Kelly, D.V. Kenny, 'Continuous Determination of Dimethylsulfide at Part-per-trillion Concentrations in Air by Atmospheric Pressure Ionization Mass Spectrometry', *Atmos. Environ.*, **25A**, 2155–2160 (1991).
133. A. Humphrey, S. Hawthorn, S. Schuetz, P. Solinski, D. Mickunas, 'Air Monitoring and Sampling Techniques Used in Support of a Cleanup Action at an Inactive Drum Recycling Facility', Field Screening Methods for Hazardous Wastes and Toxic Chemicals. Proc. of an International Symposium, Air and Waste Management Association, Las Vegas, NV, **1**, 338–345, 1995.
134. D.B. Mickunas, V. Kansal, R.D. Turpin, T.H. Pritchett, 'Ambient Air Monitoring of a SARA Title III Facility Using the TAGA 6000E MS/MS', *J. Hazard. Mater.*, **43**, 45–54 (1995).
135. B.A. Thomson, T. Sakuma, J. Fulford, D.A. Lane, N.M. Reid, J.B. French, 'Fast In Situ Measurement of PCB Levels in Ambient Air to ng m^{-3} Levels using a Mobile Atmospheric Pressure Chemical Ionization Mass Spectrometer System', *Adv. Mass Spectrom.*, **8B**, 1422–1428 (1990).
136. N.S. Karellas, G.B. De Brou, A.C. Ng, 'Real-time Detection of Airborne PCP by Mobile APCI MS/MS', 40th ASMS Conference on Mass Spectrometry, Washington, DC, 989–990, 1992.
137. H. Nagel, R. Frey, U. Boesl, *On-line Analysis of Formaldehyde and Acetaldehyde in Non-Stationary Engine Operation Using Laser Mass Spectrometry*, SAE Technical Paper Series, No. 961084, The Engineering Society for Advancing Mobility, 1996.
138. J. Villinger, W. Federer, 'Dynamischer Vielkanal Gasanalysator in der Motorentwicklung', *MTZ*, **52**, 520–525 (1991).
139. J. Villinger, W. Federer, A. Dornauer, A. Weissnicht, M. Hönig, T. Mayr, 'Dynamic Monitoring of Differentiated Hydrocarbons in Direct Engine Exhaust: A Versatile Tool in Engine Development', SAE Technical Paper Series, No. 960063, The Engineering Society for Advancing Mobility, 1996.
140. W. Lindinger, J. Hirber, H. Paretzke, 'An Ion/Molecule-Reaction Mass Spectrometer used for On-line Trace Gas Analysis', *Int. J. Mass Spectrom. Ion Processes*, **129**, 79–88 (1993).
141. T. Wanke, J. Vehlow, 'IMR-MS On-line Measurements in the Exhaust Gas of a Municipal Solid Waste Incineration Pilot Plant (TAMARA)', *Chemosphere*, **34**, 345–355 (1997).
142. C. Kai Meng, P.L. Wylie, Cynthia Cai, 'Classifying Citrus Products with a Chemical Sensor', *Food Testing & Analysis*, (October–November 1998), 17–18 (1998).
143. G. Matz, W. Münchmeyer, G. Kibelka, 'Membrane Introduction Mass Spectrometry for Monitoring of Aromatic Compounds in Exhaust Gases', Proceedings 44th ASMS Conf. on Mass Spectrometry and Allied Topics, **866**, 1996.
144. M.A. Mendes, R.S. Pimpim, T. Kotiaho, M.N. Eberlin, 'A Cryotrap Membrane Introduction Mass Spectrometry System for Analysis of Volatile Organic Compounds in Water at the Low Parts-per-trillion Level', *Anal. Chem.*, **68**, 3502–3506 (1996).
145. M.J. Yang, Y.Z. Harms, J. Pawliszyn, 'Membrane Extraction with a Sorbent Interface for Capillary Gas Chromatography', *Anal. Chem.*, **66**, 1339–1346 (1994).

146. A. Hirabayashi, M. Sakairi, Y. Takada, H. Koizumi, 'Recent Progress in Atmospheric Pressure Ionization Mass Spectrometry', *Trends Anal. Chem.*, **16**, 45–51 (1997).
147. F.L. Eisele, H. Berresheim, 'High-pressure Chemical Ionization Flow Reactor for Real-time Mass Spectrometric Detection of Sulfur Gases and Unsaturated Hydrocarbons in Air', *Anal. Chem.*, **64**, 263–288 (1992).
148. O. Möhler, Th. Reiner, F. Arnold, 'A Novel Aircraft-based Tandem Mass Spectrometer for Atmospheric Ion and Trace Gas Measurements', *Rev. Sci. Instrum.*, **64**, 1199–1207 (1993).
149. A.A. Viggiano, 'In situ Mass Spectrometry and Ion Chemistry in the Stratosphere and Troposphere', *Mass Spectrom. Rev.*, **12**, 115–137 (1993).
150. D. Bassim, P. Tosi, R. Schlögl, 'Ion-Molecule-Reaction Mass Spectrometer for On-line Gas Analysis', *J. Vac. Sci. Technol. A*, **16**(1), 114–122 (1998).
151. M.A. Craven, J.W. Gardner, P.N. Bartlett, 'Electronic Noses – Development and Future Prospects', *Trends Anal. Chem.*, **15**, 486–493 (1996).
152. W. Münchmeyer, A. Walte, G. Matz, J. Lehne, 'Fast Analysis of Complex Mixtures with Membrane Inlet Mass Spectrometry and Neuronal Networks', Poster 356, Proc. of the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, 1997.
153. L.J. Luskus, H.J. Kilian, 'Evaluation of a Portable Mass Spectrometer for Measuring Monomethyl Hydrazine and Nitrogen Tetroxide in Air', USAF School of Aerospace Medicine, Brooks Air Force Base, TX, 1–13, 1976.
154. K. Khan, 'Mass Spectrometers Poll Plant Pollution', *INTECH*, 40–42 (1993).
155. H.L.C. Meuzelaar, 'Man-portable GC/MS, Opportunities, Challenges and Future Directions', Field Screening Methods for Hazardous Wastes and Toxic Chemicals, Proc. of an International Symposium, Air and Waste Management Association, Las Vegas, NV, **1**, 35–54, 1993.
156. H.L.C. Meuzelaar, D.T. Urban., N.S. Arnold, 'Development and Testing of a Man-portable Gas Chromatography/Mass Spectrometry System for Air Monitoring', Second Intern. Symposium:, 'Field Screening Methods for Hazardous Wastes and Toxic Materials', Air and Waste Management Association, Las Vegas, 289–297, 1991.
157. W.H. McClennen, C.L. Vaughn, P.A. Cole, S.N. Sheya, D.J. Wager, T.J. Mott, J.P. Dworzanski, N.S. Arnold, H.L.C. Meuzelaar, 'Roving GC/MS: Mapping VOC Gradients and Trends in Space and Time', *Field Anal. Chem. Technol.*, **1**, 109–116 (1996).
158. W.H. McClennen, N.S. Arnold, H.L.C. Meuzelaar, 'Field-portable Hyphenated Instrumentation: The Birth of the Tricorder?', *Trends Anal. Chem.*, **13**, 286–293 (1994).
159. W.H. McClennen, N.S. Arnold, K.A. Roberts, H.L.C. Meuzelaar, J.S. Lighty, E.R. Lindgren, 'Fast, Repetitive GC/MS Analysis of Thermally Desorbed Polycyclic Aromatic Hydrocarbons (PAHs) from Contaminated Soils', *Combust. Sci. Tech.*, **74**, 297–309 (1990).
160. M.P. Sinha, G. Gutnikov, 'Role of a Microbore Capillary Column in the Development of a Portable Gas Chromatography/Mass Spectrometer', *J. Microcol. Sep.*, **4**, 405–410 (1992).
161. 'Portable Gas Chromatograph/Mass Spectrometer', Leybold-Inficon Hapsite: Measurement of Chlorinated Volatile Organic Compounds in Water', Environmental Technology Verification Program ETV, USEPA, Draft Report, Washington DC, 1998.
162. A.G. Loudon, J. Kölbe-Boelke, J. Adler, J. Stach, 'The Use of On-site Analysis for the Protection of Personnel', *Analusis*, **23**, 22–24 (1995).
163. D. Noble, 'Back into the Storm. Reanalyzing Health Effects of the Gulf War', *Anal. Chem.*, **66**, 805–808 (1994).
164. W.C. McDonald, M.D. Erickson, B.M. Abraham, A. Robbat, 'Developments and Applications of Field Mass Spectrometers', *Environ. Sci. Technol.*, **28**, 336A–343A (1994).
165. B.M. Abraham, T.-Y. Liu, A. Robbat, 'Data Comparison Study Between Field and Laboratory Detection of Polychlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons at Superfund Sites', *Hazard. Waste Hazard. Mater.*, **10**, 461–473 (1993).
166. P. Kowlaski, 'On-site Solution: Mobile Mass Spectrometry for EPAs SITE Program', *Environmental Lab.*, **23**, (1992).
167. A. Robbat, G. Xyrafas, 'Evaluation of a Field-based, Mobile, Gas Chromatograph/Mass Spectrometer for the Identification and Quantification of Volatile Organic Compounds on EPAS Hazardous Substance List', USEPA, Res. Dev., (Rep.) EPA, EPA/600/D-89/189, Field Screening Methods Hazard. Waste Site Invest., 343–348, Washington DC, 1988.
168. G. Baykut, 'Modular Sampling and Inlet Systems for Mobile Environmental Mass Spectrometry', *Trends Anal. Chem.*, **14**, 10–23 (1995).
169. G. Baykut, B. Nölke, H.-P. Vetters, G. Weiss, 'Soil, Water and Air Analyses Using a New Mobile Mass Spectrometer with Modular Sampling and Inlet Systems', *AT-ONSITE*, **1**, 34–41 (1994).
170. G. Baykut, J. Franzen, 'Mobile Mass Spectrometry, A Decade of Field Applications', *Trends Anal. Chem.*, **13**, 267–275 (1994).
171. 'Bruker-Franzen Analytical Systems', Environmental Technology Verification Report: Field Portable Gas Chromatograph/Mass Spectrometer EM640, USEPA, (EPA/600/R-97/149), Washington DC, 1997.
172. B.A. Eckenrode, 'The Application of an Integrated Multifunctional Field-portable GC/MS System', *Field Anal. Chem. Technol.*, **2**, 2–30 (1998).

173. B. Eckenode, R. Starzynski, D. Brennan, R. Drew, B. Andresen, B. Ruiz, G. Coutts, J. Bushman, A. Alcaraz, J. Cornish, 'Design Concepts and Performance Criteria in the Development of a Compact, Person-portable GC/MS system', Field Screening Methods for Hazardous Wastes and Toxic Chemicals, Proc. of an International Symposium, Las Vegas, NV, **1**, 589–596, 1995.
174. S.P. Schuetz, P.J. Solinski, D.B. Mickunas, A.M. Humphrey, R.D. Turpin, 'Comparison of Data Quality Produced by an On-site Field GC/MS and an Off-site Permanent Laboratory GC/MS: Support of a Cleanup Action at an Inactive Drum Recycling Facility', *J. Hazard. Mater.*, **43**, 67–75 (1995).
175. T.J. Kuehn, D.A. Bell, 'Economic Impacts of Field GC/MS Analysis', *Air and Waste Management Association*, **1**, 643–660 (1993).
176. 'Viking Instruments Corporation,Environmental Technology Verification Report:Field Portable Gas Chromatograph/Mass Spectrometer Spectra Trak 672', USEPA, (EPA/600/R-97/148), Washington DC, 1997.
177. Internet address: http://www.eureka.be/cgi-bin/query35.sh?67
178. J.V. Hinshaw, 'Selecting the Right Column Length', *LC-GC*, **6**(2), 82–84 (1993).
179. T.J. Bahowick, R.E. Synonev, 'Correlation of Quantitative Analysis Precision to Retention Time Precision and Chromatographic Resolution for Rapid, Short-column Analysis', *Anal.Chem.*, **67**, 631–640 (1995).
180. H.L.C. Meuzelaar, W.H. McClennen, J.P. Dworzanski, S.A. Sheya, 'Hyphenated Techniques: The Next Generation of Field-portable Analytical Instruments?', Field Screening Methods for Hazardous Wastes and Toxic Chemicals, Proc. of an International Symposium, Las Vegas, NV, **1**, 38–47, Washington DC, 1995.
181. M. Weiss, G. Matz, J. Franzen, 'A Novel Sample Separating Inlet for Direct Environmental Analysis', Proc. 10th International Mass Spectroscopy Conference, Swansea 1985, 845–846, 1985.
182. G. Matz, B. Odernheimer, 'Fast, Selective Detection of TCDD using the Mobile Mass Spectrometer MM1', *Chemosphere*, **15**, 2031–2034 (1986).
183. W. Schröder, G. Matz, J. Kübler, 'Preservatives on Waste Wood: Fast Detection with GC/MS, GC/ECD and Ion Mobility Spectrometry', *Field Anal. Chem. Technol.*, **2**, 287–297 (1998).
184. E. Denly, D. Hopper, 'Field Chemistry for PAHs and VOCs Applied to a Risk-based Soil Cleanup at a Landfill', Field Analytical Methods for Hazardous Wastes and Toxic Chemicals, January 29–31, 1997, Air and Waste Management Association, Las Vegas, NV, 752–770, 1997.
185. G. Matz, W. Schröder, 'New Strategies for Quality Assurance in Fast Soil Analysis in the Field', *Proceedings 'Contaminated Soil '95'*, eds. W.J. van den Brink, R. Bosman, F. Arendt, Kluwer Academic Publishers, Dordrecht, 1995.
186. G. Matz, W. Schröder, J. Flachowsky, 'On-site Investigation of Contaminated Soils by GC-MS and EDXRF-Techniques', in *Contaminated Soil '93*, eds. F. Arendt, G.J. Annokkée, R. Bosman, W.J. van den Brink, Kluwer Academic Publishers, Dordrecht, 657–664, 1993.
187. R.J. Mesley, W.D. Pocklington, R.F. Walker, 'Analytical Quality Assurance', *Analyst*, **116**, 975–990 (1991).
188. W.M. Davis, M.B. Wise, J.S. Furey, C.V. Thompson, 'Rapid Detection of Volatile Organic Compounds in Groundwater by In Situ Purge and Direct-sampling Ion-Trap Mass Spectrometry', *Field Anal. Chem. Technol.*, **2**, 86–96 (1998).
189. C.P. Ardito, T.A. Duval, D.W. Reaber, 'Characterizing a VOC Site Through On-site Analysis of Soil Gas', *Volatile Organic Compounds in the Environment*, ASTM STP 1261, eds. W. Wang, J. Schnoor, J. Doi, American Society for Testing and Materials, 203–218, 1996.
190. A. Robbat, Jr, C. Liu, T.-Y. Liu, 'Field Detection of Organochlorine Pesticides by Thermal Desorption Gas Chromatography/Mass Spectrometry', *J. Chromatogr., A*, **625**, 277–288 (1992).
191. K. Jiao, A. Robbat, Jr, 'Performance-based Field Methods for Analysis of Substituted Phenols by Thermal Desorption Gas Chromatography/Mass Spectrometry', *J. AOAC Int.*, **79**, 131–142 (1996).
192. P.M. Williamson, 'On-site Analysis of Landfill Gas by GC/MS', *Field Screening Methods Hazard. Wastes Toxic Chem., Proc. Int. Symp.*, **2**, 897–902, Air and Waste Management Association, 1995.
193. G. Matz, 'Determination of VOCs in Fire and Chemical Accidents by Mobile GC/MS', in *Current Protocols in Field Analytical Chemistry*, Unit 1B2, ed. V. Lopez-Avila, J. Wiley & Sons, New York, 1998.
194. J.L. Jack, 'On-site Analysis, USEPA Monitoring and Measurement Technologies Program (MMTP): Part 1', *Am. Environ. Lab.*, **12/93**, 27–29, 1993.
195. K.L. Busch, 'Desorption Ionization Mass Spectrometry', *J. Mass Spectrom.*, **30**, 233–240 (1995).
196. S. Fuerstenau, G. Soli, T. Cunningham, B. Hancock, B. Pain, M. Sinha, 'Solid State Ion Detectors for Charge Detection Mass Spectrometry', *46th ASMS Conference on Mass Spectrometry and Allied Topics*, Air and Waste Management Association, Orlando, FL, May 31 – June 4, 1998.

Radon, Indoor and Remote Measurement of

Florence T. Cua
Princeton, USA

1 Introduction **1**
2 Sampling Strategies **2**
2.1 Radon Compared with Radon Progeny 2
2.2 Sampling Time 2
3 Theory and Operating Principles of the Detectors **5**
3.1 Passive Detectors 5
3.2 Active Radon Detectors 7
4 Comparison Studies Among the Radon, Indoor and Remote Monitors **10**
Abbreviations and Acronyms **10**
Related Articles **10**
References **11**

The existing radon and radon progeny measurement instruments are described in this article. The various methods and instruments used depend on whether radon or radon/thoron progeny or daughter products are being measured, the duration of the measurement, whether short-term or long-term and the type of radiation being detected (alpha, beta, or gamma). The modes of detection used include the following: activated charcoal radon monitors, electret ion chamber, registration of nuclear tracks in solid-state materials, liquid scintillation, ionization chambers, scintillation detectors with zinc sulfide, ZnS(Ag), alpha particle spectrometers with silicon diodes, surface barrier or diffused junction detectors, and gamma-ray spectrometry with NaI(Tl) scintillation crystals or germanium lithium (GeLi) semiconductor detectors, fiber optics sensors and glass. The advantages and disadvantages will be described for various portable instruments used for measuring radon, thoron, and their daughter products. Comparison studies among the radon monitors will be included.

The reader is referred in particular to George[1] and National Council on Radiation Protection (NCRP) Report 97[2] for additional information.

1 INTRODUCTION

Radon, the inert gaseous element created from the radioactive decay of ^{226}Ra and ^{232}Th, itself decays with a short half-life (3.8 days for ^{222}Rn and 55.6 s for ^{220}Rn) to produce a series of short-lived daughter products. Measurement techniques generally determine the radon concentration (Bq m^{-3} or pCi L^{-1}) in air or water or the total energy of the alphas released by the complete decay of the radon daughter products (RDPs); the quantity is usually expressed in terms of the working level (WL).

Table 1[3,4] shows the radiometric properties of ^{222}Rn, ^{220}Rn, and their short-lived daughter products. Details can be found in NCRP Report 58[3] and Shleien.[4] Radon and its daughter products, when inhaled, deliver an significant alpha radiation dose to bronchial tissue that has been implicated in radiogenic lung cancer. Radon also has an affinity for bone tissues.

Several quantities are almost uniquely used in the measurement of radon and the daughter products:

1. The WL is the total amount of energy which would be eventually released by the alpha particles emitted by short-lived RDPs. It is defined as the combination of radon daughters in one liter of air that will result in the emission of 1.3×10^5 MeV of potential alpha energy (2.08×10^{-5} J m^{-3}). The WL is the common unit for expressing radon daughter dose rate, originally in uranium mines, but now in environmental exposures as well. Under equilibrium conditions, 3700 Bq m^{-3} (100 pCi L^{-1}) of radon will produce 1.0 WL in approximately four hours. The quantity is historically related to its use with measurements made of the air in uranium mines.
2. The equilibrium equivalent concentration (EEC) is the radon concentration, in equilibrium with the short-lived daughters, that would have the same potential alpha energy concentration (PAEC) per unit volume as the existing mixture. Numerically, it is calculated from measured concentrations of the daughters (Equation 1):

$$\mathrm{EEC} = 0.105[\mathrm{A}] + 0.516[\mathrm{B}] + 0.379[\mathrm{C}] \quad (1)$$

where EEC is expressed in Bq m^{-3} or pCi L^{-1} and [A], [B], and [C] are the respective concentrations of the ^{218}Po, ^{214}Pb, and ^{214}Bi daughters in the same units. Equations (2–4) may be used to calculate the PAEC in WLs from a particular ^{222}Rn concentration in EEC.

$$\mathrm{WL} = \frac{\mathrm{EEC\ (in\ Bq\,m^{-3})}}{3700} \quad (2)$$

$$\mathrm{WL} = \frac{\mathrm{EEC\ (in\ pCi\,L^{-1})}}{100} \quad (3)$$

$$\mathrm{Jm^{-3}} = \frac{\mathrm{EEC\ (in\ Bq\,m^{-3})}}{1.8 \times 10^8} \quad (4)$$

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Radiometric properties of ^{222}Rn(a) and ^{220}Rn(b) and its short-lived daughter products

Radionuclide	Half-life	Energy (MeV)	Gamma and X-rays
A. Rn-222	3.82 d	5.5 (alpha)	0.5
Po-218	3.05 min	6.0 (alpha)	–
Pb-214	26.8 min	0.7, 1.0 (beta)	0.01–0.8
Bi-214	19.7 min	0.5–3.3 (beta)	0.01–2.4
Po-214	63 μs	7.7 (alpha)	0.8
B. Rn-220	55 s	6.3 (alpha)	0.55
Po-216	0.15 s	6.8 (alpha)	0.8
Pb-212	10.6 h	0.3, 0.6 (betas)	0.01–0.3
Bi-212	60.6 min	6.1 alpha 1/3 0.4–2.2 beta 2/3	0.01–1.8
Po-212	3×10^{-7} s	8.8 (alpha)	–
Tl-208	3.1 min	0.7–1.8 (beta)	0.01–2.6

Note: The betas are maximum energies. (Please see Shleien et al.[4] for electrons and individual betas.)

See the Figure 10.1 in NCRP 97[2] for the contributions of RaA(^{218}Po), RaB(^{214}Pb) and RaC(^{214}Bi) to the PAEC as a function of time from initially pure ^{222}Rn. The primary reference is Evans[18] and the secondary reference is NCRP 97.[2] Also, see Figure 10.2 in NCRP 97[2] for the build-up and decay of alpha activity of RaA(^{218}Po) and RaC′(^{214}Po) from individual isolated RaA, RaB(^{214}Pb) and RaC(^{214}Bi), each having an initial decay rate of 0.17 Bq (10 disintegrations per minute). The primary reference is Holaday et al.,[19] and the secondary reference is NCRP 97.[2] The author recommends NCRP 97[2] for a very good background on measurement of radon and radon daughters in air.

3. The cumulative exposure is expressed in terms of the working level month (WLM), where WLM = WL × (hours exposed/170), since uranium miners were considered to have worked 170 hours in a month. 1 WLM = exposure to 1 WL for 170 h. A more detailed discussion of these quantities and units appears in NCRP Report 97.[2]

Another concern about airborne radon and its daughter products is that the high electrical charges of the newly created daughters, especially ^{210}Po, cause a significant fraction to become attached to dust, water and other aerosol particles in the air. They may also be deposited on the surfaces of walls, ceilings and objects in the vicinity (''plateout'').

The plateout radioactivity is thus removed from consideration when radiation doses from inhaled radioactivity are to be estimated unless it is resuspended. Some of it may, however, be ingested. Also, several techniques to measure this activity and relate the results to the total exposure from radon over an extended period have been developed.

A measure of the fraction of radon daughters attached to aerosol particles is the ''equilibrium factor'', F, the ratio of the PAEC in the existing mixture to that which would exist if all short-lived daughters were in equilibrium with the radon present (=EEC/[^{222}Ra]). The value of F depends on the ambient environmental conditions such as humidity, temperature, the amount of turbulence, infiltration and exfiltration. It also changes with time as the relative abundance of the various radon daughters changes due to radioactive decay. The NCRP has recommended values of 0.4 for indoor environments, and 0.7 for outdoor environments when actual measured values are not available. The United States Environmental Protection Agency (USEPA) uses a range of 0.3–0.7 for their equilibrium ratio (ER) and assumes the midrange value of 0.5 for indoor environments.

2 SAMPLING STRATEGIES

2.1 Radon Compared with Radon Progeny

Air concentration measurements are made much more frequently of radon than of the radon progeny. Although sampling and analysis for radon progeny are more difficult and more costly than that of the parent radon gas, they are required for certain situations where a more precise assessment of the radiation exposure is needed.[1] Measurement of radon air concentrations must consider that three isotopes, ^{219,220,222}Ra may exist. However, the ^{219}Ra produced by ^{235}U in the soil with its half-life of 4 s and ^{220}Rn from thorium, half-life of 55 s, are usually not present.

In some cases, both the radon and the daughter concentrations are measured. One method is the two-filter method.[4] Ambient air, containing a mixture of radon and daughters, is drawn through the first filter to remove the radon progeny. The radon containing air passes into a decay chamber, where the daughters produced are collected on a second filter. The concentration of radon (on the first filter) and that of the radon progeny (on the second filter) can be counted separately by alpha spectroscopy, scintillation counting, or a passive detector.[5]

2.2 Sampling Time

There are three time regimes for sampling. It should be noted that in all three cases, the results are averaged over the sampling time. The sampling time can vary from seconds in the case of sniffing to seconds to minutes for grab sampling to days, months or longer for integrated time measurements.

- *Instantaneous (''grab'' sampling[5])*: air is pumped into a vessel. Metered filling to obtain time-averaged results may extend the sampling period. The analysis is performed after the sampling is finished.
- *Continuous (''sniffing'')*: air is drawn through a filter to remove the progeny, then through a detector volume usually at a constant volume rate. Analysis is performed during the sampling time.
- *Integrated*: a detector or monitor is placed in the sampling area for an extended time. The analysis is performed after the sampling time and the results produced as an integrated exposure over the sampling time or an average concentration over the sampling time.

2.2.1 Grab Sampling

The advantages of grab sampling for radon gas are the quick delivery of results and that the sampling equipment is usually very portable.[6] The disadvantages are that the sample period is too short to use as a short-term measurement (i.e. less than 48 h). Also the results must be corrected for the presence of any thoron (^{220}Rn). As a practical matter, the sample cells can become so contaminated with the daughters that keeping them out of use for a time is necessary to reduce the background activity. The lower limit of detection (LLD) for most grab sampling methods is a few tenths of a pCi L^{-1}; the expected precision is quite low (~25%) unless counting times on the order of several hours are used (~10% at or above 4.0 pCi L^{-1}). The individual sample cells may need calibration as well as the analysis systems. Individual background levels for each cell may need to be established. In addition, spiked samples and duplicate samples may be necessary to ensure acceptable quality control on the results.

Grab sampling is typically conducted by collecting the air to be studied in a 5–15-liter container that is sent to a laboratory for analysis. The radon can be concentrated for easier or more accurate analysis by passing the air sample through an activated charcoal trap submerged in liquid nitrogen. At this low temperature, radon is retained on the charcoal while the air passes through. The trap is then warmed in the presence of helium and the radon transferred to a counting cell or ionization chamber for analysis. There is usually a wait of four hours or more to allow the collected radon to come to equilibrium with its short-lived daughters although recent advances in the development of computer-based analysis systems have somewhat reduced the waiting time. Instruments used to measure WLs usually draw air through a filter to collect the daughter products, which are then analyzed.

Grab sampling can be used in a ''sniffing'' mode, say as a diagnostic tool for finding entry routes of radon gas, by measuring the sample almost immediately after it is taken. The technique provides a quick, but imprecise, measurement to identify relative differences between multiple measurements.

Practical concerns about grab WL devices include the following: the air pump must be well regulated, requiring routine calibration of the pumps and flowrates with a complete calibration annually. Typical airflow rates range from 2 lpm to 25 lpm, but they can be lower or higher. The sampling period must be precisely controlled and the delay time between the end of sampling and start of analysis (counting) accurately recorded. Care must be taken so that the exposed face of the filter faces the detector. The complex operating procedure requires a trained operator. The technique typically has a high sensitivity – ~0.0005 WL with a precision of 10% at 0.02 WL or higher – and can produce relatively quick results. The technique generally cannot be used for short-term or long-term measurements that are required by agencies such as the USEPA.

Examples of two of the more common protocols for grab sample WL analysis are the following:[6]

2.2.1.1 Kusnetz Count for 10 min, 40–90 min after the end of sampling; the WL averaged over the sampling time is calculated from an empirical algorithm usually by a computer.

Both filters are actually measuring progeny concentrations – the first filter measures the actual progeny concentration, and the second uses the build-in of the progeny from pure radon to measure the radon concentration.

The Kusnetz method and modifications are very well explained in NCRP 97.[2]

2.2.1.2 Tsivoglou Count from 2–5 min, 6–20 min, and 21–30 min; the WL averaged over the sampling time is calculated along with individual activities of the daughter products ^{214}Pb, ^{214}Po, and ^{218}Po.

2.2.2 Continuous Sampling

In continuous sampling, the air is drawn or allowed to diffuse first through a filter to remove the radon daughters then through a sensitive volume where the analysis is performed. There is relatively good precision with most devices. Continuous sampling devices can track hourly variations (or more often depending upon the device) and are frequently controlled by computers that record the results and parameters for downloading or analysis at a later time. The results can reveal evidence of tampering or improper ventilation. Typical sampling periods range from 48 h to months if the filter is maintained or replaced periodically. The continuous sampling techniques can be used to measure and record variations in the WLs with

time as well as the individual radon and thoron decay products.

In active systems, the air is drawn through the sensitive volume by a pump; in passive systems, the air is allowed to diffuse into the volume, a process that requires about 4 h to come to equilibrium. The typical sampling period is 48 h, of which the data from the last 44 contiguous hours are used.

Continuous radon monitors are relatively costly and require a trained operator. Some units do not have good precision around the 4.0 pCi L^{-1} concentration used as an action level by the USEPA. Also, since the air is in the chamber for a short period, the short-lived thoron is counted along with radon. Consequently, the readings must be used qualitatively only unless one compensates for the presence of the thoron. The four-hour ramp-up time for passive systems may be a significant drawback. Continuous monitoring in dusty, smoky, and high-humidity environments should be avoided or carefully maintained. If the results from WL monitors are converted to pCi L^{-1}, the ER or equilibrium factor, F, must be measured or assumed.

The LLD of the typical continuous sampling device is 1.0 pCi L^{-1} or less, with a precision of $\pm 10\%$ at concentrations of 4.0 pCi L^{-1} or greater. The devices are usually calibrated annually, with semi-annual crosschecks. If scintillation cells are used as detectors, the background counting rates must be monitored. The devices that measure WL have sensitivities of 0.01 WL or less with a precision of $\pm 10\%$ at 4.0 pCi L^{-1} or higher concentrations. The background level of the WL instruments should be determined at least after every 168 h of use, taking care to draw clean or aged air or nitrogen through the filter.

2.2.3 Integrating Sampling

Integrating samples provide a single value of either radon concentration or WL averaged over the exposure period. They can also provide a measurement of total exposure (usually in WLM) for the period of that exposure. The length of the period can be as short as a few days or can be of the order of many years. Although there are active devices, the most commonly used monitors are of the passive type. Such devices include canisters or other vessels containing activated charcoal, alpha track detectors, and electret-based ionization chambers. Although the analysis of these devices can require sophisticated equipment and trained personnel, the deployment is using performed by non-technical and untrained individuals, usually with a brief set of instructions. Integrating dosimeters are, by far, the most commonly used devices for measuring radon concentrations or WLs in private houses and other buildings. Testing periods are generally classified as either long-term or short-term.

Charcoal canisters, alpha track detectors, and charcoal liquid scintillation devices are available for purchase. Electret ion chamber detectors are generally available only from laboratories. They are exposed to the air in the home for a specified period and then sent to a laboratory for analysis. Both short-term and long-term passive devices are generally inexpensive. Some of these devices may have features that offer more resistance to test interference or disturbance than others. Professional radon testers may use any of these devices to measure the home's radon level.[6]

2.2.3.1 Short-Term Testing Short-term tests remain in place from 2 to 90 days, depending on the device. Common passive devices for short-term testing include alpha track detectors, charcoal canisters, charcoal liquid scintillation detectors, and electret ion chambers. Because radon levels vary from day to day and season to season, due to such things as weather changes, opening and closing of doors, the type of air conditioning/heating systems and the lifestyle of the family, a short-term test is less likely than a long-term test to provide an adequate measure of the annually averaged radon level used by most public health agencies for decision making purposes. A short-term monitoring period is more likely to detect an unrepresentative peak or valley in the radon concentration than the long-term test. However, the quick results provided by the short-term tests make them very popular for use in real estate transactions and initial sampling.

2.2.3.2 Long-Term Testing Long-term tests remain in place for more than 90 days. Alpha track and electret ion chamber detectors are commonly used for this type of testing. A long-term test provides results that are more representative of an annual mean value for radon concentrations or WL. In the USA, long-term tests are frequently used to confirm initial short-term results between 4 pCi L^{-1} and 10 pCi L^{-1}, because the USEPA recommends remedial action if the average annual radon concentration in occupied areas is greater than 4 pCi L^{-1}.

2.2.4 Measurement Location and Conditions

Short-term measurements in buildings are generally made under closed building conditions, i.e. windows, outside vents, and external doors closed, and internal–external air exchange systems off. These conditions should be present at least 12 h before the beginning of the tests. The measurements should be made in areas away from drafts and extreme temperature conditions. The measurement location should be at least 90 cm from windows and other openings, 30 cm from external walls, 50 cm from the floor, and 10 cm from other objects. If possible, it should be made in the general breathing zone area (2–2.5 m from the floor).

3 THEORY AND OPERATING PRINCIPLES OF THE DETECTORS

3.1 Passive Detectors

3.1.1 Activated Charcoal Detectors

This detector typically consists of a cylindrical "salve" canister, containing 25–200 g of activated charcoal although there are several different configurations. The charcoal has a high affinity for radon that adsorbs on its active sites. The canister is prepared and sealed until the exposure period begins. It is placed in the test area with the top removed for a period of three to four days. During this time, the radon diffuses into the charcoal and is adsorbed. At the end of the exposure period, the canister top is replaced, sealing the canister again. A period of at least three hours but not exceeding seven days is allowed before the canister is analyzed. Analysis is done by placing the canister before a gamma analysis detection system, usually consisting of a NaI(Tl) scintillation crystal and a multichannel analyzer calibrated to measure the intensity of the gammas from one or more of the daughter products, usually ^{214}Pb and ^{214}Bi. The LLD for these systems ranges from 3.7 to 7.5 Bq m^{-3} (0.1 to 0.2 pCi L^{-1}) for an exposure period of four days. Occasionally, other detectors, such as high purity intrinsic germanium (HPGe) semiconductor detectors are used instead of NaI(Tl).

Some charcoal detectors have a diffusion barrier that slows the diffusion of the radon and filters out the daughter products. The barrier also slows the diffusion of atmospheric water vapor that competes with radon for the charcoal sites and extends the period of integration to about seven days. According to Cohen,[6] a typical collector with 25 g of charcoal, at intermediate humidity and temperature, collects 0.068 Bq per Bq m^{-3} of radon in the air (68 pCi per pCi L^{-1}) after one week's exposure.

A second activated charcoal detection method uses collectors containing 1–2 g of charcoal that is analyzed by liquid scintillation counting (alpha/beta analysis). The radon is eluted from the charcoal into an organic solvent in which radon is highly soluble at low temperatures then into the scintillation cocktail for analysis in a liquid scintillation analyzer. Each ^{222}Rn decay results in the release of two alpha particles and two beta particles each of which may be detected by the analyzer. The LLD is about 5 Bq m^{-3} (0.14 pCi L^{-1}).

The activated charcoal detector is one of the least expensive devices for radon measurement. Each canister costs about $1 (US) and the charcoal can be regenerated to drive off the radon and organic gases and recycled. A gamma analysis system generally costs ~$20 000 (US). Since the deployment and retrieval of the detectors requires only brief instructions and no training, the detectors are very popular for screening of private houses and buildings. The USEPA estimates that over 50% of screening measurements in the USA use the charcoal detector. The detectors are placed by homeowners, house inspectors, or real estate agents, then recovered and mailed to a processor for analysis.

Since the radon adsorption processes are temperature-dependent and compete with water vapor and other molecules, these detectors are used only in indoor environments that experience moderate environmental fluctuations. The correction for humidity is determined by obtaining the weight gain of the charcoal during the exposure period. It is rarely larger than ±6%. Since the average correction is only about 3%, the step is usually omitted. The correction for temperature is 1.5% per degree Fahrenheit, with an uncertainty of 5%.

Because the response of the detectors is controlled by the diffusion of the radon, they are not true integrating devices if radon concentration peaks and valleys are dramatic. Their use is limited to short-term integrated time sampling, so they may miss changes in radon concentration due to rainfall or long-term barometric pressure variations. They must also be analyzed within a very short time after exposure.

Another potential problem is that of deliberate tampering with the sampling process, by such actions as increasing the area ventilation rate by opening windows or doors or activating an exhaust fan when the canister is exposed, locating the detector in a nonrepresentative area or covering the detector for part of the exposure period. None of these deceptive practices will be detected by the results of the analysis.

3.1.2 Electret Ionization Chambers[8]

These detectors are variations on the common ionization type detector (see below) in which the radon decay radiation particles create charged pairs in the air of the active volume of the detector. The charges collected by one electrode are measured to determine the total activity averaged over the exposure time. In these passive devices, the electrical potential is supplied by a Teflon®-based electret rather than an active electrical power supply.

The electret ionization chamber consists of a conducting plastic chamber that contains the electret. When the detector is positioned and the chamber opened to the ambient air, radon diffuses through a filter into the chamber. There it decays, with the radiation produced by it and its daughter products generating ion pairs in the air. The electrons produced in the ionization events are drawn to the electret and consequently reduce the potential difference (voltage) on the surface of the electret. This change in the electret surface voltage is measured by a separate calibrated voltmeter and corresponds to an activity integrated over the period of exposure. The device is

used until the electret surface potential is depleted. The electret is then replaced.

The electret ion chambers may be used in either indoor or outdoor environments since they are less sensitive to environmental fluctuations such as temperature and humidity. They can be configured for short-term measurements of two to seven days as well as for long-term measurements of three to six months by using electrets of varying thickness. If the electret voltage is not depleted, the same unit can be used to make sequential measurements by measuring sequential drops. The devices do not need to be analyzed immediately after exposure.

One disadvantage is that because the detectors are sensitive to all types of natural background radiation, a correction must be made. Another is that the electret of the device can be damaged by touching the surface by contamination or by physical impact.

The are two manufactured electret-based ion chambers, Rad Elec's E-PERM and the Ra Dome™ chamber. The short-term E-PERM devices have sensitivities of 7 Bq m^{-3} (0.2 pCi L^{-1}) for a three-day exposure. The Ra Dome™ is a long-term device that measures 150 Bq m^{-3} in 90 days.

3.1.3 Alpha Track Etch Detectors

This group of passive detectors is based on the phenomena that charged particles, especially alpha particles such as those emitted during the decay of radon and its daughters, leave a track of disrupted chemical bonds as they penetrate into solid materials and give up their energy. In certain materials, these tracks can be made visible by chemical etching because they etch at a significantly faster rate than the unaffected material. The areal density of the tracks can be correlated to the total exposure of the material to radon and its daughters (tracks cm^{-2} per Bq d^{-1} m^{-3}).

Cellulose nitrate (LR-115) and allyl diglycol carbonate (CR-39) films are two of the most commonly used materials. The chemistry of the etching process, usually in an NaOH solution, and the calibration for radon exposure for these materials is probably most established for these two materials. The track density is determined by manual scanning under a microscope or with the use of a computer-based optical scanning program.

The permanence of the track density image and the capability of the detectors for integrating measurements over extremely long times are the method's chief advantages. The tracks left by beta and gamma radiation are not imaged by the etch process because they are not large enough to exceed the threshold for imaging in the etch process. Therefore, ambient background corrections are not required. The cost per detector is relatively low and the simplicity and robustness of the detector have resulted in its widespread use for long-term integrating measurements of radon concentrations. They are easily distributed and, in the USA, they can be returned for processing by the postal mail system.

The drawbacks include the need for labor-intensive chemical or electrochemical processing and the variations in the material response due to thickness, acquired background and temperature variations. The detector calibration is also sensitive to the altitude of the measurement because the alpha particle range increases as the ambient air density decreases. At higher altitudes, the average energy the alphas deposit in the detector thus increases.

The track etch material, which varies in surface area, is usually mounted in the bottom of a chamber with one end open to the ambient air. The tracks are created by that fraction of the alphas emitted during the decay of the radon in the chamber and the daughter products (^{218}Po and ^{214}Po) produced that impinge on the detector with enough kinetic energy to create a track large enough to be imaged after etching. The chamber must be configured so that areas of persistent electrical static charge are eliminated as a source of interference with the accumulation of the tracks. Cohen has found that the calibration factor for nuclear track detectors has been found to vary unpredictably with a standard deviation of 16%, presumably due to electric charge effects associated with use of plastic cups.[7] For radon concentration measurements, a filter may remove the ambient daughter products so that only radon diffuses into the chamber. For WL measurements the opening is unfiltered and the ER is either determined or assumed (usually to be 0.5).

Track etch is not as sensitive as other techniques. For example, reasonably accurate environmental measurements require exposure periods of three months to one year. The range of detection is 0.2–20 000 pCi L^{-1} months (0.004 to 400 WLM at an ER of 0.5) for a coffee-cup-sized detector or 0.4 to 600 000 pCi L^{-1} months (0.008 to 1200 WLM) for a pill-box-size detector. Typical calibration factors[7] are 0.001 tracks per Bq m^{-3} day (0.0372 per pCi L^{-1} day). A 90-day exposure at 37 Bq m^{-3} (1 pCi L^{-1}) gives an average of 3.3 tracks in addition to 0.4 tracks from background, a total of 3.7, with a standard deviation of 1.9, and a percentage uncertainty of 58%.

The precision, especially at low concentrations, can be improved by increasing the number of microscopic fields counted. For example, if the background is 4 tracks per field where the total counted tracks is 8 per field then counting 1 field gives a precision error of ±75%, counting 3 fields, ±50%, and 10 fields, ±25%.

The track etch technique is not confined to plastic detectors. Recent advances have been made, primarily in Europe, in applying the procedure to glass such as

that in aged mirrors and picture frames, and with less success, windows. If the glass has been exposed over a period of years, then the track density can be correlated to the integrated exposure for the individuals occupying the same rooms. Although there are many influences on the accuracy of this technique, the promise of obtaining at least a first approximation of the long-term exposure of individuals has captivated several researchers.

Another technique involving aged glass involves the measurement of the ^{210}Po activity imbedded in the surface of the glass by the recoil energy of the parent. The glass surface is etched with hydrofluoric acid to remove the ^{210}Po which is then measured directly by mass spectrometry or by alpha liquid scintillation analysis. Jones[9] found the technique useful for exposures greater than 259 pCi L^{-1}-years with a correlation of 0.97 in a controlled laboratory experiment.

3.1.4 Thermoluminescent Dosimeters

Thermoluminescent (TL) materials are commonly used to measure integrated radiation dose in a wide variety of different situations. The technology can also be used to evaluate radon concentrations and WL measurements.

TL is the ability of certain materials, such as lithium fluoride, to retain certain changes in the electron energy distributions induced by absorption of radiation. A small but constant fraction of electrons of the electron-hole pairs created by the absorption of the radiation energy are trapped for extended periods of time by interstitial imperfections in the lattice structure. When the material is heated to a few hundred degrees centigrade by an external source, the electrons are driven from these traps and the subsequent recombination produces a visible light photon in the blue-green area of the electromagnetic spectrum (wavelength $\sim$ 400 nm). These photons are multiplied by a photomultiplier tube or photodiode after the TL material is heated and a stream of photons produced. The charge (coulombs) produced by the photoreceptor is proportional to the radiation dose delivered to the TL material.

Thermoluminescent dosimeters (TLDs) record all radiation doses so the radon results must be corrected for background radiation. The TLDs are mounted in a detector assembly that filters the daughters from the ambient air then causes the daughters from the radon gas decay to attach to the surface of the TLD. The TLD, usually a rectangular chip of a few millimeters on a side, is removed from the detector assembly and placed in a separate reader where it is heated under controlled conditions and the light released measured.

In one detector configuration, the air entering the measurement chamber passes first through silica gel to remove the moisture then through a filter that removes the daughter products then into a funnel-shaped chamber. In the chamber, a battery-induced electrical field attracts the daughters to the surface of the TLD. The TLD is mounted at the narrow end of the funnel.

Occasionally a TLD is used in conjugation with another detector in a radon measurement device. One example is the personal radon monitor (PRM) described by Litt et al.[10,11] The PRM is a sensitive, passive integrating device used to monitor individual human exposures. Radon diffuses into the monitor through a 4.8 mm thick conducting foam barrier that keeps out the progeny. The radon detector is a 9 $\times$ 9 mm CR-39 nuclear track film. Gamma ray exposure data are obtained from CaF_2 TLD chips placed beneath the CR-39. The CR-39 and CaF_2 TLD are covered with thin aluminized Mylar that nullifies any charge artifacts. Some attenuation of the alpha particle energies occurs in the Mylar; however, the Mylar is necessary to produce a uniform track distribution. Each PRM has provision for triplicate CR-39 film and TLD exposures. The PRM housing is a cylinder, 7.5 cm in diameter and 3.0 cm in height, and is worn on a belt.

3.2 Active Radon Detectors

3.2.1 Scintillation Detectors

There are several basic types and configurations of scintillation detectors. The scintillation process begins when a fraction of the kinetic energy lost in the scintillator by the ionizing radiation is converted to light that is then observed by a photosensitive device such as a photomultiplier tube or a photodiode. This device generates an electrical pulse, quite small but usually proportional to the amount of light observed. Electrical amplifiers produce the much larger but still proportional signal that is counted and perhaps analyzed.

The scintillators and the photosensitive devices must be chosen so the maximum efficiency for observing the light is obtained. Usable scintillation light is usually produced in the blue-green wavelengths, around 400 nm, since most of the photosensitive devices are most efficient at those wavelengths.

Several different scintillation detectors can be used for radon measurements. Silver activated zinc sulfide [ZnS(Ag)], one of the earliest known scintillating substances, is used in powder form to coat the inside surface of measuring cells made of material transparent in the blue-green wavelengths of the visible spectrum. Radon-containing air is drawn into the cell, perhaps through a filter to remove the existing daughters. The cell can be sealed before counting or continuously pumped through the cell during counting. The vessel is placed in a dark, light-tight environment and observed by the photosensitive monitor. A significant fraction of the alphas

produced by the decay of the radon and its daughters impinge on the ZnS; the resultant light flashes are seen and recorded. The more energetic betas may also be counted. The thin ZnS layer is effectively insensitive to gamma radiation. The system is calibrated with known activities of radon and the vessel volume is determined accurately. The cells, frequently referred to as Lucas cells, after the creator of the original cell, are available in many different cell volumes and configurations. The cell size is generally prescribed by the size of the active area of photomultiplier tube; the larger the volume, the lower the sensitivity and the LLD.

This system is used in several counting modes. A grab sample of filtered radon-containing air allowed to come to equilibrium (a four-hour delay before counting) is a very popular method for determining radon concentration. If only relative differences in radon concentrations are desired, a sample taken can be counted without delay in a ''sniffer'' type mode.

The most popular technique for determination of the WL is to collect the radon daughters on a high efficiency filter and then count the emitted alphas without discrimination (sorting by alpha energy) during or after the end of sampling. A number of different sampling and counting intervals can be used to determine the individual concentrations of ^{218}Po, ^{214}Pb and ^{214}Bi. One very popular method for determining the WL, developed by Kusnetz and later refined by George, consists of a single gross alpha counting interval usually between 40 and 90 min after a 5–10-min sample period. The LLDs for ^{218}Po, ^{214}Pb, ^{214}Bi, and the PAEC for a five-minute sample at a flow rate of 10 L min^{-1} are 20 Bq m^{-3} (5 pCi L^{-1}), 7 Bq m^{-3}, 20 Bq m^{-3}, and 50 nJ m^{-3} (0.0025 WL) respectively.

One commercially available system is manufactured by the Pylon Electronic Development Co. The TEL 520 is a trace level radon gas detector. The sensor is a 14 000 mm^2 high efficiency ZnS(Ag) detector dimensionally optimized for 6.00–7.69 MeV alpha particles with an efficiency of approximately 4 cpm pCi^{-1} coupled to a specially selected stable high-gain photomultiplier tube and unique scaler circuitry. It may be used with two filters to measure both radon and thoron concentrations.

One of the problems with these systems is that the radon and daughter products tend to be difficult to flush from the detector cells. Thus, the background counting level gradually increases as the cell is reused. Although the rate of increase can be reduced by flushing the cells with a radon-free gas the increasing background raises the LLD of the system lowering its sensitivity with continuous use of the cells.

Another commonly used scintillation material is NaI doped with an impurity (Tl). This detector is most sensitive to gamma radiation and essentially insensitive to alphas and betas. The vessel used with this system is usually fabricated of material, such as aluminum or polyethylene, that has a low attenuation for the gammas and X-rays produced during the decay of radon and the daughters. With this system, the radon is pumped into the vessel and the vessel is placed in a specific physical orientation with respect to the photomultiplier tube assembly. The fraction of emitted gammas from the radon decay that pass through the detector may deposit some or all of their energy into the scintillation crystal. The resulting light flashes are counted and analyzed. NaI(Tl) detection systems are typically used to analyze the passive activated charcoal canister monitors for integrated radon concentration determinations.

This system is also sensitive to gammas from background radiation so it is usually housed with a substantial shield of lead or concrete to reduce the background interference. Although NaI(Tl) is the most efficient and most cost-effective material for this detector configuration, other scintillator materials such as CsI could be used.

Liquid scintillators are also useful for radon detection. They have the advantage of higher counting efficiency since the radon and the daughters are intimately mixed with the scintillator material. The disadvantages include the need for wet chemistry to prepare the samples and limitations due to the solubility of the cocktail solutions. Usually the radon-bearing air is passed through an organic solvent in which radon is highly soluble at low temperatures. The mixture is introduced into a vial containing a liquid scintillator. The vial is then sealed and placed in an optical cavity where the light flashes produced as each ^{222}Rn decay are followed by two alpha particles and two beta particles, and are observed by a photosensitive detector and counted and perhaps analyzed. One technique, due to Van Dijk and De Jong,[12] used 4 g of silica gel (size distribution, 1–3 mm) cooled over liquid nitrogen to trap the radon. After absorption, the tube with silica gel was warmed in ice–water for 5 min, after which the contents were added to a counting vial containing 20 mL of a toluene-based scintillation liquid.

The use of liquid scintillation for analysis of activated charcoal passive monitors has been discussed in another section. Liquid scintillation may also be used to measure radon concentrations in water.

Kim[20] has investigated the use of fiber optic scintillation glass. They have developed a detector constructed of a bundle of cesium s doped glass fibers in a light-tight chamber. The radon-containing air is pumped into the chamber and diffuses into the fiber bundle. There the large effective surface area of the fiber bundle detects the alphas released with high efficiency.

3.2.2 Internal Gas Ionization Chamber Counters

In these systems, the radon and daughters are introduced into a chamber in which an electrical field has

been established between two electrodes. The radiation emitted in the decay process ionizes the air molecules, generating electrons and positively charged ions, which are drawn to respective electrodes in response to the Coulomb forces present. The resultant current can be measured as electrical pulses (for the alpha particles since the magnitude is large enough), averaged over a period as a current, or collected as a total charge for a period. Since the currents generated by the detector volumes are quite small, the measurement systems tend to be sophisticated and complex. From any of these parameters, the average radon activity in the active volume can be determined if the air sample is filtered or the WL if the ambient radon daughters are not filtered out and the ER is determined or assumed.

Ionization chambers are among the most accurate and sensitive methods for radon concentration measurements. They are also among the most expensive and complex. Thus, they are typically used as primary laboratory calibration instruments for radon measurements and for evaluating secondary instruments for traceability purposes. A typical LLD for these instruments is about 0.7 mBq (0.019 pCi) for a 17-hour count.

3.2.3 Semiconductor Detectors

Semiconductor detectors have the capability to measure deposited alpha energies with the highest accuracy of any of the radon detectors. Most of the systems using semiconductor detectors use this characteristic to discriminate the alphas from the radon and the daughter decays. These instruments can discriminate radon from thoron as well as from the daughters. The diffused junction detector consists of a piece of high purity silicon in which a ''depletion'' area is created by the diffusion of an impurity that easily attracts free or conduction electrons. Thus, the area is depleted from the normal number of charge carriers available to it. The depletion zone is thinly coated with a conductive layer of a few nanometers and an electric field is established across it by imposing a potential difference of a few thousand volts. The alpha particles emitted by the radon and daughter decay penetrate the conductive layer expending their kinetic energy in the depletion zone with the creation of charged pairs of electrons and positively charged ''holes''. The electrons have enough energy to move about the silicon structure as conduction electrons and are collected usually as an electrical pulse. Surface

Table 2 Instruments and methods for the estimation of ^{222}Rn concentrations

Detection system	Sampler volume or weight	Sampler filling method	Measurement period	LLD (95%)	Sensitivity
Ionization chambers 21 pulse	Up to 1 l	Evacuation or flow through	1000 min	0.74 mBq	
Scintillation cell	0.1 l	Evacuation	60 min	1.1 mBq	
	0.45 l	Evacuation	60 min	3.7 mBq	
	0.14 l	Flow through	60 min	3.7 mBq	
Continuous scintillation cell	0.1 l	Flow through			1 kBq m^{-3}
	1.5 l	Flow through			40 Bq m^{-3}
Electrostatic collection					
ZnS to film	0.03 l	Diffusion	170 h		1.5 kBq m^{-3}
ZnS to PMT	0.9 l	Diffusion	40 min	26 Bqm^{-3}	
ZnS to PMT	3 l	Diffusion	40 min	7.4 Bqm^{-3}	
SS to MCA	10 l	Diffusion	40 min	0.7 Bqm^{-3}	
SS to MCA	2 l	Diffusion	10 min		40 Bqm^{-3}
PERM(LiF)	1.5 l	Diffusion	7 d	1.1 Bqm^{-3}	
EGARD(LiF)	1 l	Diffusion		3.3 kBqm^{-3}d	
^{222}Rn only					
ZnS to PMT	1.9 l	Diffusion	60 min	2.2 Bqm^{-3}	
Charcoal adsorption	150 g				1.1 Bqm^{-3}
	80 g		72 h	11 Bqm^{-3}	
	25 g		50 h(?)		40 Bqm^{-3}
Filtration					
Single filter		10 min–3 h filtration		1 Bqm^{-3}	
Two filter	140-l decay chamber	Continuous filtration			4 Bqm^{-3}
Two filter	600-l decay chamber	Continuous filtration			0.4 Bqm^{-3}
Two filter	100-l decay chamber	Continuous filtration		0.4 Bqm^{-3}	
Nuclear track detector			3 months	500 Bqm^{-3} day	

Reproduced by permission of Eric E. Kearsley, Ph.D. of the National Council of Radiation Protection and Measurements for Table 7.1 of NCRP Report No. 97, Measurement of Radon and Radon Daughters in Air(1988) on February 22, 1999.

barrier detectors, a slightly different configuration, are also used, as are photodiodes.

The instruments usually employ an onboard computer for the analysis and data storage as well as to perform the measurements on a preset schedule. They are frequently used as sniffers. One device, developed by the US Department of Energy Environmental Measurements Laboratory (EML), has a sensitive volume of 6 L. Air is drawn into the cylindrical volume and the charged daughters are attracted to the surface of the detector for spectroscopic analysis. The system can measure 150 Bq m^{-3} of ^{222}Rn or ^{220}Rn in less than 10 min and stores the data, along with the temperature and the relative humidity measurements, in the computer. The energy discrimination capabilities of the detector allow the device to measure the short-lived ^{218}Po in a sniffing mode since it can respond to concentration changes in a short time. The sensitivity for ^{218}Po is 0.08 cpm Bq^{-1} m^{-3} (3 cpm pCi^{-1} L^{-1}).

WL measurements are made by drawing air through a filter detector assembly. The alpha emissions from radon and thoron daughters are measured and the data is stored on an onboard computer. One portable, battery-operated device has a small internal pump with a flow rate around 0.15 L min^{-1}. The computer-scheduled sampling program is typically one hour sampling intervals over periods that can range from one to seven days. The LLD is about 0.0005 WL. This unit has a separate computer-based reader and programmer.

Table 2 summarizes the instruments and methods for the estimation of ^{222}Rn concentrations.[2]

4 COMPARISON STUDIES AMONG THE RADON, INDOOR AND REMOTE MONITORS

Several comparison studies of the different types of detectors and measurement techniques have been published. Martz et al.[13] found the correlation coefficient of 0.9980 (and a linear slope of 1.052) between one type of alpha track monitor results and diffusion barrier charcoal canister (DBCC) results, averaged over a six-month period. The correlation coefficient between the 60 paired 1-wk DBCC measurements and the integrated Pylon AB-5 continuous scintillation chamber measurements is 0.9879 and the slope is 1.05.

Future comparison studies should be made among the TLD and solid state nuclear track detector (SSNTD) radon and RDPs measuring devices that would include Kotrappa et al.[14]

An international climatic test program for integrating radon detectors is described in Hopper et al.[15]

The National Institute of Standards and Technology (NIST) primary ^{222}Rn measurement system was described in the paper authored in 1990 by Colle et al.[16] It suggested upgrading, expanding and replacing the primary radon measurement system, consisting of pulse ionization chambers and ancillary gas handling and gas purification equipment. A secondary measuring system[17] for ^{222}Rn-in-air, based on measurements of the photon emission rate of radon samples contained in spherical 35-mL glass ampoules with a NaI(Tl) well counter, was calibrated against the NIST[21] pulse-ionization-chamber national radon measuring system with an overall uncertainty of approximately 2%.[17] An international intercomparison of marine atmospheric ^{222}Rn measurements in Bermuda was done in October 1991.[22–24] NIST polyethylene-encapsulated $^{226}Ra/^{222}Rn$ emanation standards were used for calibration of electret-based integral radon monitors.[25]

ABBREVIATIONS AND ACRONYMS

DBCC	Diffusion Barrier Charcoal Canister
EEC	Equilibrium Equivalent Concentration
EML	Environmental Measurements Laboratory
ER	Equilibrium Ratio
LLD	Lower Limit of Detection
NCRP	National Council on Radiation Protection
NIST	National Institute of Standards and Technology
PAEC	Potential Alpha Energy Concentration
PRM	Personal Radon Monitor
RDP	Radon Daughter Product
SSNTD	Solid State Nuclear Track Detector
TL	Thermoluminescent
TLD	Thermoluminescent Dosimeter
USEPA	United States Environmental Protection Agency
WL	Working Level
WLM	Working Level Month

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants

Industrial Hygiene **(Volume 5)**
Direct Reading Instruments for the Determination of Aerosols and Particulates

REFERENCES

1. A.C. George, 'State-of-the-Art Instruments for Measuring Radon/Thoron and their Progeny in Dwellings – A Review', *Health Physics*, **70**(4), 451–463 (1996).
2. NCRP Report 97, 'Measurement of Radon and Radon Daughters in Air', National Council on Radiation Protection and Measurements, Bethesda, MD, USA, 1988.
3. NCRP Report 58, 'A Handbook of Radioactivity Measurements Procedures', 2nd edition, National Council on Radiation Protection and Measurements, Bethesda, MD, 1985.
4. B. Shleien, L.A. Slaback, Jr, B.K. Birky, *Handbook of Health Physics and Radiological Health*, 3rd edition, Williams and Wilkins, Philadelphia, 1998.
5. EPA Home Buyer's and Seller's Guide to Radon (http://www.epa.gov).
6. 'Radon Measurement Proficiency Course', Manual, Eastern Regional Radon Training Center Rutgers, The State University of New Jersey, New Brunswick, 1997.
7. B.L. Cohen, 'Comparison of Nuclear Track and Diffusion Barrier Charcoal Adsorption Methods for Measurement of ^{222}Rn Levels in Indoor Air', *Health Physics*, **50**, 828ff (1986).
8. P. Kotrappa, L.R. Stieff, et al. Electret Ion Chamber from Rad Elec Inc. http://www.guidance-inc.com/~radelec/tech.htm.
9. R.F. Jones, 'Retrospective Determination of Radon Exposure to Glass Using Liquid Scintillation Counting'. Doctoral Thesis. Rutgers, the State University, New Brunswick, NJ, 1993.
10. B.R. Litt, J.M. Waldman, N.H. Harley, P. Chittaporn, 'Validation of a Personal Radon Monitor for Use in Residential Rn-222 Exposure Studies', *Health Physics*, **61**(6), 727–735 (1991).
11. N.H. Harley, P. Chittaporn, M.H. Roman, J. Sylvester, 'Personal and Home Rn-222 and Gamma Ray Exposure Measured in 52 Dwellings', *Health Physics*, **61**(6), 737–744 (1991).
12. W. Van Dijk, P. De Jong, 'Determining the Rn-222 Exhalation Rate of Building Materials Using Liquid Scintillation Counting', *Health Physics*, **61**(4), 501–509 (1991).
13. D.E. Martz, J.L. George, G.H. Langner, 'Comparative Performance of Short Term Diffusion Barrier Charcoal Canisters and Long Term Alpha Track Monitors for Indoor Rn-222 Measurements', *Health Physics*, **60**(4), 497–505 (1991).
14. P. Kotrappa, S.K. Dua, N.S. Pimpale, P.C. Gupta, K.S.V. Nambi, A.M. Bhagwat, S.D. Soman, 'Passive Measurements of Radon and Thoron Using TLD or SSNTD on Electrets', *Health Physics*, **43**(3), 399–404 (1982).
15. R.D. Hopper, F. Steinhausler, M. Ronca-Battista, 'IAEA/EPA International Climatic Test Program for Integrating Radon Detectors', *Health Physics*, **77**(3), 303–308 (1999).
16. R. Colle, J.M.R. Hutchinson, M.P. Unterweger, 'The NIST Primary Radon-222 Measurement System', *J. National Institute of Standards and Technology*, **95**(2), 155–165 (1990).
17. J.M.R. Hutchinson, J. Cessna, R. Colle, P.A. Hodge, 'An International Radon-in-Air Measurement Intercomparison Using a New Transfer Standard', *Appl. Radiat. Isot.*, **43**(1/2), 175–189 (1992).
18. R.D. Evans, 'An Engineers Guide to the Elementary Behavior of Radon Daughters', *Health Physics*, **17**, 229 (1969).
19. D.A. Holaday, D.E. Rushing, R.D. Coleman, P.F. Woolrich, H.L. Kusnetz, W.F. Bale, 'Control of Radon and Daughters in Uranium Mines and Calculations on Biologic Effects', USPHS Report 494, US Government Printing Office, Washington, 1957.
20. Kim Il Sik, 'The Application of Trivalent-cerium-doped Silicate Glass Scintillator to the Detection of Ionizing Radiation', PhD Thesis, Rutgers University, New Brunswick, 1993.
21. R. Colle, '^{222}Rn and ^{226}Ra Measurement Standards and Calibration Activities at NIST', Seminar presented at LPRI, Saclay, France, 1993.
22. R. Colle, M.P. Unterweger, P.A. Hodge, J.M.R. Hutchinson, S. Whittlestone, G. Polian, B. Ardouin, J.G. Kay, J.P. Friend, B.W. Blomquist, W. Nadler, T.T. Dang, R.J. Larsen, A.R. Hutter, 'An International Intercomparison of Marine Atmospheric ^{222}Rn Measurements in Bermuda', *J. Geophy. Res.*, **100**(D8), 16 617–16 638 (1995).
23. R. Colle, M.P. Unterweger, P.A. Hodge, J.M.R. Hutchinson, 'An International Marine-atmospheric ^{222}Rn Measurement Intercomparison in Bermuda Part I: NIST Calibration and Methodology for Standardized Sample Additions', *J. Res. NIST.*, **101**(1), 1–19 (1996).
24. R. Colle, M.P. Unterwerger, J.M.R. Hutchinson, S. Whittlestone, G. Polian, B. Ardouin, J.G. Kay, J.P. Friend, B.W. Blomquist, W. Nadler, T.T. Dang, R.J. Larsen, A.R. Hutter, 'An International Marine-atmospheric ^{222}Rn Measurement Intercomparison in Bermuda Part II: Results for the Participating Laboratories', *J. Res. NIST.*, **101**(1), 21–46 (1996).
25. R. Colle, P. Kotrappa, J.M.R. Hutchinson, 'Calibration of Electret-based Integral Radon Monitors Using NIST Polyethylene-Encapsulated ^{226}Ra/^{222}Rn Emanation (PERE) Standards', *J. Res. NIST.*, **100**(6), 629–639 (1995).

Solid-phase Microextraction in Analysis of Pollutants in the Field

L. Müller, T. Górecki, and J. Pawliszyn
University of Waterloo, Canada

1 Introduction **2**
1.1 Suitability of Solid-phase Microextraction for Field Analysis 3
2 Solid-phase Microextraction Coupled to Field Portable Fast Gas Chromatography **4**
3 Selected Field Applications of Solid-phase Microextraction **5**
3.1 Field Analysis by Solid-phase Microextraction Coupled to Gas Chromatography 5
3.2 Field Applications for Indoor Air Monitoring 5
3.3 In Situ Analysis of Groundwater and Soil Gas 8
4 New advances in Solid-phase Microextraction Field Samplers **9**
4.1 Storage Investigations 12
4.2 Field Applications 13
5 Conclusions **14**
Acknowledgments **15**
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

Solid-phase microextraction (SPME) is a modern sampling/sample preparation method, used for isolation and preconcentration of organic molecules from a variety of matrices. SPME uses a short piece of a fused silica fiber coated with a polymeric stationary phase. The fiber is mounted in a device resembling a syringe. During transport, storage and manipulation, the fiber is retracted into the needle of the device. During extraction and desorption of the analytes, the fiber is exposed. Analytes present in a sample partition into or onto the coating, depending on its type. The process continues until equilibrium is reached between the coating and the sample. From then on, longer extraction times do not result in larger amounts of analyte extracted. Once the extraction is finished, the fiber is retracted back into the needle, and the device is transferred to a gas or liquid chromatograph for analyte separation and determination. When gas chromatography (GC) is used, the analytes are thermally desorbed from the fiber in a GC injector. Coupling of SPME with high-performance liquid chromatography (HPLC) requires a special interface.

Two distinct SPME coating types are available commercially. Coatings of the first type, including poly(dimethylsiloxane) (PDMS) and poly(acrylate) (PA), extract analytes by absorption. This process is non-competitive, therefore in most cases the amount of an analyte extracted by such coatings from a sample is independent of the matrix composition. No saturation or displacement effects occur. The amount of an analyte extracted from a sample is linearly dependent on its initial concentration, provided that several important variables, including (but not limited to) temperature, extraction time and mass transfer conditions, are kept constant. Coatings of this type usually perform very well for compounds of medium to low volatility.

Coatings of the second type, including poly(dimethylsiloxane)/divinylbenzene (PDMS/DVB), Carbowax™/divinylbenzene (CW/DVB) and Carboxen™/poly(dimethylsiloxane) (CX/PDMS), extract analytes by adsorption. This process is limited to the surface of the coating. It is competitive, which means that a molecule with higher affinity to the coating can displace a molecule with lower affinity. Since the number of active sites on the surface of any coating is limited, a linear response for those coatings can be expected only when the concentrations of all the compounds that can be extracted by the coating from a sample are low. Adsorption-type coatings are particularly suitable for volatile analytes, for which they offer much better sensitivity than PDMS or PA.

SPME is very well suited for field applications, especially when the analysis is carried out on site. The fiber can be exposed directly to the medium analyzed, for example lake water or ambient air, without the need to collect a sample and without knowing the exact volume of the sample the fiber is exposed to. Analysis can then be performed using field portable instrumentation. Manual operation of the device is very simple. The fibers are reusable, which makes the cost of analysis low. Separation of volatile components sampled by SPME can be very fast when a dedicated system (available commercially from SRI Instruments) is used.

Alternatively, SPME can be used to sample in the field, and then transported for the analysis to the laboratory. Modified devices have to be used for this purpose to avoid analyte losses during transport and storage, as well as contamination of the samples. Transporting fibers is much easier than transporting glass or metal containers with water or air samples.

The main disadvantage of SPME in the field is its lack of robustness. The needle can be easily bent, and the fiber can be broken when handled without sufficient care.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

New designs of field portable SPME devices address those issues. Also, it might be difficult to accurately control in the field all the experimental parameters that affect the amount of analyte extracted.

1 INTRODUCTION

Field analysis is rapidly gaining more and more importance, responding to the need for immediate results in environmental monitoring, as reviewed by V. Lopez-Avila et al.[1] Immediate remediation for industrial spillages or environmental accidents is possible only if the results are provided in real time. Performing an interactive sampling, where only relevant locations are sampled, can save time and resources. Also, the costs of a field investigation are lower compared to a traditional procedure, by cutting the expenses of transport and storage of bulky and heavy environmental samples, by shortening the time of the entire analysis and by focusing the investigation only on significant sampling sites. Results are more representative of the real sample composition if the analysis is performed directly in the field, so that volatile losses or sample degradation are minimized.

To date, several types of analytical instrumentation can be transported to the field for on-site investigations. The promise of cheaper and faster analysis is pushing the optimization of field analytical methods, from immunochemical assays and chemical sensors to GC.[1] Even the sensitivity and identification capabilities of mass spectrometry (MS) can be applied directly in the field with the recent introduction of portable instruments.[2]

For field investigations of volatile compounds, portable high-speed gas chromatographs can be utilized. Fast GC, which utilizes short capillary columns and high carrier gas flow-rate, can reduce the time of the instrumental analysis by one order of magnitude, producing therefore even more immediate results and reducing the cost of analysis.[1]

SPME is a powerful tool for the extraction/sample preparation of environmental samples, widely described in literature.[3–5] A microextraction, defined as extraction of a very small portion of analytes, is performed by a polymeric coating immobilized on a silica fiber and exposed to the sample. For this purpose, the fiber is mounted at the end of a small diameter stainless steel tube, which, in turn, is located inside a fine steel needle. The fiber can be retracted into the needle or exposed from it by manipulating the plunger of a syringe-like device. When the fiber does not need to be exposed, it can be protected inside the needle. The device is commercially available from Supelco (Bellefonte, PA), together with a variety of fiber assemblies with different coatings.

The SPME extraction process is based on partitioning of an analyte between the sample matrix (gas, liquid or solid) and the fiber. Typically, only a fraction of the total amount of the analyte present in the sample is extracted. Exhaustive or near-exhaustive extraction can be achieved only under specific conditions, when the volume of the sample is small and the affinity of the compound towards the coating (measured by the fiber/sample partition coefficient K_{fs}) is high. The amount of an analyte n extracted by absorption-type SPME fibers at equilibrium in a two-phase system can be described by Equation (1):[3]

$$n = \frac{K_{fs} V_s V_f C_0}{V_s + K_{fs} V_f} \tag{1}$$

where V_f is the fiber coating volume, V_s is the sample volume, and C_0 is the initial concentration of the analyte of interest in the sample. This equation indicates that in a two-phase system there is a direct relationship between the initial analyte concentration in the sample and the amount extracted, even though other factors influence the efficiency of extraction at equilibrium, such as the affinity of the fiber for the target compound, and the sample volume. As in all equilibrium processes, several factors can affect the amount of an analyte extracted by the fiber, from the temperature and the pH to the organic and salt content in the sample.[3]

If an adsorption process is applied for extraction, a different theory is necessary to describe the extraction process.[6] The following considerations pertain to PDMS/DVB, CW/DVB and Carbowax™/template resin coatings. No comprehensive theory has been developed yet for the CX coating, for which capillary condensation plays an important role.

Based on Langmuir's theory of the adsorption process,[6] Equation (2) can be used to describe the processes involved in SPME extraction with solid coatings:

$$n = C_{fA}^{\infty} V_f = \frac{K_A C_{0A} V_s V_f (C_{f\max} - C_{fA}^{\infty})}{V_s + K_A V_f (C_{f\max} - C_{fA}^{\infty})} \tag{2}$$

where n is the amount of analyte adsorbed by the porous polymeric coating at equilibrium,

- C_{fA}^{∞} is the concentration of analyte A on the fiber at steady state;
- V_f is the volume of the fiber coating;
- K_A is the adsorption equilibrium constant of analyte A;
- C_{0A} is the initial concentration of analyte A in the sample;
- V_s, V_f are the volumes of the sample and of the fiber, respectively;

$C_{f\max}$ is the concentration of active sites on the surface (corresponding to the maximum achievable analyte concentration on the surface).

The general form of Equation (2) is very similar to that of Equation (1), which describes the extraction process in a two-phase system when absorption is the extraction mechanism. The main difference between Equations (1) and (2) is the presence of the fiber concentration term $(C_{f\max} - C_{fA}^{\infty})$ in the numerator and denominator of Equation (2) (also, note that the meaning of K_A is entirely different than that of K_{fs}: K_A is adsorption equilibrium constant, while K_{fs} is the partition coefficient). For very low analyte concentrations on the fiber, it can be assumed that $C_{f\max} \gg C_{fA}^{\infty}$. For this condition to be fulfilled, analyte concentration in the sample and/or its affinity towards the coating must be very low. When these requirement(s) are met, a linear dependence should be observed. If, however, the amount of an analyte on the fiber is not negligible compared to the total number of active sites, the dependence cannot be linear any more.[6]

SPME fibers can be exposed to solid, liquid or gaseous matrices, using three different extraction modes:

- direct immersion of the fiber in the sample matrix (liquid or gaseous matrices);
- extraction of the headspace above the sample (liquid or solid samples);
- direct immersion with the fiber protected with a high-molecular weight cutoff membrane eliminating adverse effects when the sample is very dirty (liquid samples).[3]

A variety of available polymeric coatings with different chemical properties allow the technique to be used for a broad range of compounds widely differing in polarity and volatility (from amines and phenols to volatile organic compounds (VOCs) and inorganic substances).[7] To date, hundreds of applications have been developed with SPME,[8] even though the power of this technique in areas including industrial hygiene and environmental and process monitoring has not been completely explored yet.

1.1 Suitability of Solid-phase Microextraction for Field Analysis

Several features of SPME make this technique very well suited for field investigations. The feature that differentiates SPME from all the other sampling techniques can be explained from the principle of the extraction-partitioning process described by Equation (1). If a large volume sample is extracted, e.g. atmospheric air or lake water, or K_{fs} is very small (low affinity of the fiber coating toward the analyte), the term V_s in the denominator of Equation (1) is much larger than the $K_{fs}V_f$ term, and the latter can be neglected. The final relationship that describes the amount absorbed by the SPME fiber in such a case is described by Equation (3):

$$n = K_{fs}V_f C_0 \tag{3}$$

It follows from Equation (3) that under such conditions the amount of analyte extracted by the fiber at equilibrium, in a two-phase system, is not dependent on the sample volume. This means that collection of a known volume of a sample is not necessary in the sampling procedure, and the fiber can be exposed directly to the sample in the field, without removing it from the environment. The amount of analyte extracted is negligible in comparison to the total amount present in a large volume of a sample. The system is not disturbed by the SPME extraction. For instance, the scents released by a flower can be sampled by simply exposing the fiber to the inside of the flower bulb without removing it from the field, and a real field sample can be analyzed. The above is not always true: if the amount of scents removed from the headspace of the flower is not negligible, the flower will react by increasing their production and the normal situation will be disturbed.

Other features make SPME very suitable for field use. It is pen-shaped, easy to handle and to operate by any operator. It can be exposed even to locations difficult to reach for sample collection (e.g. groundwater or deep soil). It is reusable and the fiber can be easily changed. If a field portable instrument is available, the fiber can be immediately desorbed in the hot injector of a GC for chromatographic investigation, since the extraction and sample preparation processes are combined into a single step. SPME can also be coupled to fast GC: the lack of solvent in the whole process is an important requirement for high-speed GC.[9,10]

The possibility of automation of the extraction technique makes it potentially useful for semi-continuous monitoring of industrial processes[11] and for industrial hygiene purposes.[12]

The advantages of SPME for field sampling can also be exploited when the use of portable instrumentation is not possible. After field sampling, the SPME device, with the analytes stored on the fiber, can be transported to the laboratory for further investigation. A special sampler is required for this purpose in order to prevent analyte losses during storage and transportation. Since the SPME fiber is fragile, care should taken during sampling in field conditions. New SPME field samplers were designed in order to satisfy all the features required by field applications, such as robustness of the device, protection of the fiber during sampling and minimization of analyte losses during storage.

2 SOLID-PHASE MICROEXTRACTION COUPLED TO FIELD PORTABLE FAST GAS CHROMATOGRAPHY

High-speed GC is a powerful analytical technique that dramatically reduces the time of chromatographic separation, by up to one order of magnitude. Fast GC typically utilizes short narrow-bore capillary columns and high carrier-gas flow rates.[10] Good separation can be achieved by fast GC if narrow injection bands are produced. Band broadening can be minimized by avoiding the use of solvents, by reducing the detector inner volume and by introducing the sample to the column very rapidly.

SPME coupled to high-speed GC is a good combination to perform rapid and cost-effective investigations in the field, even of complex organic samples. SPME is particularly suited for fast GC, as it is solvent-free, and the thin coatings can provide very fast desorption of analytes at high temperatures. Some instrumental modifications were performed recently in order to achieve successful fast separations.[9,10]

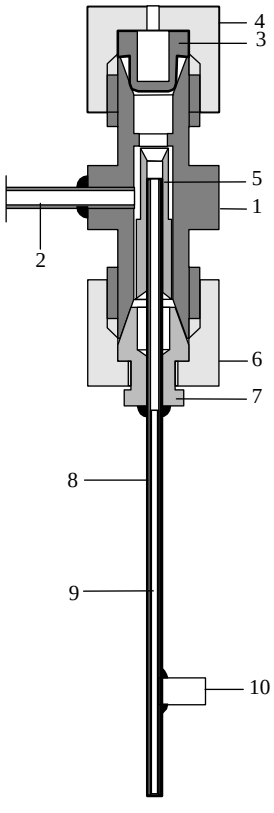

Figure 1 Dedicated injector to couple SPME with fast GC, allowing heating rates of up to 4000 °C s^{-1}. 1: modified Swagelok fitting; 2: stainless steel tubing; 3: molded septum; 4: nut; 5: needle guide; 6: nut; 7: drilled-thru plug; 8: stainless steel tubing; 9: 0.53-mm i.d. fused silica capillary; 10: electrical contact.

A portable system was optimized for field investigations using SPME coupled to fast GC and was commercialized by SRI Instruments (model 8610C, SRI Instruments, Torrance, CA). The instrument was tested in combination with a flame ionization detector (FID), a photoionization detector (PID), and a dry electrolytic conductivity detector (DELCD). A dedicated injector, presented in Figure 1, was mounted on the portable system in order to use SPME for high-speed separation. The injector guarantees very fast thermal desorption of the analytes from the SPME fiber.

The injector for high-speed GC should produce as narrow an injection band as possible. Regular injector ports have too large internal volumes (e.g. split/splitless injector), since they usually have to accommodate large volumes of gaseous samples or vapors produced by solvent injection. Thermal focusing for separation improvement is not convenient for fast separations, since temperature programming is impractical for high-speed GC. Hence, an injector port with a small internal volume was required for this application. Also, very fast thermal desorption from the SPME fiber was required to produce a narrow injection band and achieve effective separation. In the dedicated injector for SPME coupled to fast GC, the injector port was maintained cold during needle introduction and was rapidly heated only when the fiber was exposed to the carrier gas stream. The desorption area of the injector was heated by capacitive discharge that allowed heating rates as fast as 4000 °C s^{-1}, and very narrow injection bands were observed, as required by fast GC. The injector (Figure 1) was based on a modified Swagelok fitting (1) which was drilled through in the middle, and the carrier gas was supplied by a piece of 1/16 inch tubing (2), soldered to it. The inlet of the injector was closed with a molded septum (3) and the original SRI nut with needle guide (4). The hole inside the fitting was enlarged to accommodate the brass needle guide (5). The back end of the injector was closed with a nut (6) and drilled-thru plug (7). The plug had a piece of gauge 19 steel tubing (8) soldered into it. Inside the tubing was a segment of 0.53 mm i.d. fused silica capillary (9). The tubing (8) assured mechanical stability and durability, and served as the heating element at the same time. The capacitative discharge power supply was connected to the injector body and to the end of tubing (8) through contact (10). The GC column was connected to the injector with a zero-dead volume butt connector. Tubing (8) was thermally insulated along its entire length (not shown).

Neither the FID nor the DELCD detector required any modifications before being used for fast GC. On the other hand, the relatively large internal volume of the PID caused a very significant peak broadening. To overcome this problem, a simple insert was designed and introduced into the internal cavity of the detector, reducing the

internal volume of the PID from approximately 130 μL to 25 μL. The insert was tightly fitted into the detector cavity to avoid creation of void volumes. An excellent improvement in resolution for separation of BTEX (benzene, toluene, ethylbenzene, xylenes) was achieved in this way.[13]

The system was optimized in the laboratory with standards of BTEX (benzene, toluene, ethylbenzene and *o,m,p*-xylenes; 100 μg L^{-1} each compound in water) and a complete separation was performed within 15 s. Purgeables A (trichlorofluoromethane, 1,1-dichloroethene, dichloromethane, 1,1-dichloroethane, trichloromethane, tetrachloromethane, trichloroethene, 1,2-dichloropropane, 2-chloroethyl vinyl ether, 1,1,2-trichloroethane, tetrachloroethene, dibromochloromethane, chlorobenzene; 200 μg L^{-1} each compound in water) were separated in 2 minutes with good precision using the DELCD.[9,10] The SRI system was then carried to the field for trace analysis of trichloroethylene (TCE) in soil (see section 3.1) and airborne formaldehyde for industrial hygiene monitoring (see section 3.2).

3 SELECTED FIELD APPLICATIONS OF SOLID-PHASE MICROEXTRACTION

3.1 Field Analysis by Solid-phase Microextraction Coupled to Gas Chromatography

Several applications of SPME and high-speed GC for field analysis have been reported. The portable SRI model 9300B instrument, described in the previous section and now commercially available as model 8610C, was applied in the field for various matrices and different pollutants: field determination of traces of TCE in soil samples,[13] VOC in groundwater[14] and indoor measurements of airborne formaldehyde[15] (see section 3.2).

SPME/fast GC field determinations were performed by quantifying TCE in soil (clay) samples. By collecting core samples from depths up to 5 m, the migration of the pollutant was investigated. Soil samples were collected and extracted with methanol. An aliquot of the extract obtained was used to spike pure water, and the aqueous solution was then vigorously stirred in a sealed vial to equilibrate with the headspace above the water. A PDMS/DVB 65 μm SPME fiber was exposed to the headspace of the static water sample for the extraction. The described procedure was performed directly in the field. After extraction, the SPME fiber was desorbed in the portable fast GC coupled to PID system for on-site instrumental investigation. The entire process of SPME extraction of spiked water samples, desorption and instrumental analysis took 3 min to complete, allowing over 500 samples to be quantified in 10 days. This successful study is an effective application of SPME coupled to fast GC for rapid field investigations.

Groundwater samples were analyzed by SPME/fast GC in the field to evaluate VOC content.[14] The commercially available instrumentation (SRI model 8610C) was used for this purpose. Of great interest is the validation performed by comparison with traditional liquid–liquid extraction (LLE) followed by instrumental analysis by gas chromatography/electron capture detector (GC/ECD) with a Hewlett Packard model HP 5890 GC in the laboratory. A comparison of the quantitative results obtained by the two methods showed a scattered difference between 3% and 5%, demonstrating the reliability of field investigations performed by SPME/high-speed GC.

3.2 Field Applications for Indoor Air Monitoring

Monitoring of air is becoming more and more important not only for environmental pollution evaluations, but also for industrial hygiene applications.[12] SPME can be a useful tool for industrial hygiene studies since it is suitable for field evaluations and is easy to automate. It can potentially be used therefore for semi-continuous monitoring of indoor/outdoor air pollutant levels. An automated procedure is necessary to monitor levels of hazardous and undesirable exposures workers are subjected to in their working day. Proper monitoring equipment for this purpose should be portable, should give immediate and continuous results and should alert to excessive exposure.[12]

When sampling air, SPME can be used in two ways. In the conventional approach, the fiber is simply exposed to the atmosphere analyzed by depressing the plunger of the SPME device. The analytes reach the fiber coating mainly via advection and (to a lesser extent) diffusion. The fiber is usually exposed for a time long enough for the analytes to reach equilibrium. Analyte equilibration times depend on their partition coefficients and mass transfer conditions. Alternatively, SPME can be used as a time-weighted average (TWA) sampler for gas phase analytes. In this approach, the fiber is retracted a known distance into the needle during the sampling period. In contrast to the conventional approach, the analytes are not allowed to reach equilibrium with the fiber. Analyte sampling rate is controlled by a number of factors, including the diffusion coefficient and the concentration gradient inside the needle. Analysis by this method yields the concentration of an analyte averaged over the entire sampling period (e.g. eight hours in industrial hygiene).

TWA concentration is defined according to Equation (4):

$$\overline{C}_{t_n} = \frac{C_1 t_1 + C_2 t_2 + C_3 t_3 + \cdots + C_n t_n}{t_1 + t_2 + t_3 + \cdots + t_n} \qquad (4)$$

where $\overline{C}_{t_n}$ is the TWA concentration, C_1 is the analyte concentration observed for time t_1, and so on, until time t_n. This definition can be used to determine TWA concentration from discrete measurements (grab sampling). Such measurements should cover the entire period of interest. Conventional SPME can be relatively easily utilized for this type of measurement provided field portable instrumentation is used. The fiber has to be exposed to the outside atmosphere and equilibrated with the analytes, followed by immediate desorption and GC analysis. Once the analytes are removed from the fiber, the entire cycle has to be repeated for as many times as required to cover the period of interest.

An alternative approach to determining TWA concentration is with one sampling session where $t_1 = t_n$. For this approach to be successful, one has to make sure that the amount of analyte reaching the trapping medium at any given moment is directly proportional to the analyte concentration outside of the sampler at the same moment. It is also very important that analyte molecules trapped inside the sampler are not re-released to the surrounding atmosphere and do not affect further uptake of analyte molecules. When the latter conditions are fulfilled, the trapping medium is called a ''zero sink''.[16,17,18]

The following presentation is based on a paper by Martos and Pawliszyn.[19] TWA sampling by SPME can be accomplished by leaving the fiber inside the needle during the sampling session. Figure 2(a) shows the SPME fiber position for TWA sampling, while Figure 2(b) presents analyte concentration gradient inside the needle. In Figure 2, C_{BULK} is the analyte concentration in the bulk of the atmosphere examined, C_{FACE} is the concentration at the face of the sampling device, and $C_{SORBENT}$ is the concentration at the gas/sorbent interface.

Since the dimensions of the SPME needle are very small, analytes can reach the fiber only by diffusion. From Fick's law, analyte transport in such a system can be described by Equation (5):

$$\frac{dJ}{dt} = -\frac{DA}{L}\,dc \tag{5}$$

where J is the weight (ng) of analyte passing through a cross-section A (cm^2) during a time t (min); D is analyte diffusion coefficient in air ($cm^2\ min^{-1}$), L is the diffusion path length (cm), and C is analyte concentration (ng cm^{-3}). We can now define the sampling rate R ($cm^3\ min^{-1}$) by Equation (6):

$$R = D\left(\frac{A}{L}\right. \tag{6}$$

Hence (Equation 7)

$$\frac{dJ}{dt} = -R\,dc \tag{7}$$

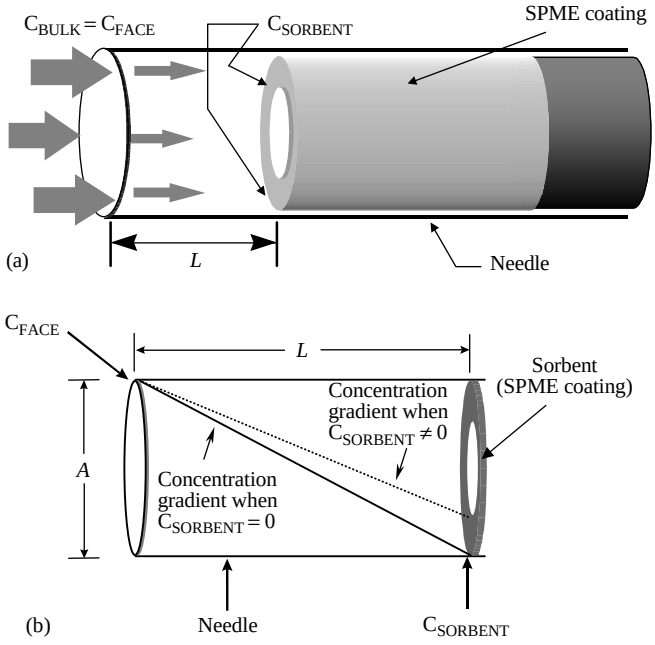

Figure 2 (a) Positioning of the fiber inside the needle used for TWA sampling with SPME. (b) Schematic representation of the concentration gradient of the analyte between the opening of the needle and the fiber coating for a zero and a non-zero sink situation. L is the path length and A is the surface area of the opening.

Equation (7) can be integrated with the following limits to concentration and time to give Equation (8):

$$J = -R\int_{t_1}^{t_2}\int_{C_F}^{C_S} dc\,dt = -Rt\int_{C_F}^{C_S} dc \tag{8}$$

where $C_F = C_{FACE}$, and $C_S = C_{SORBENT}$. After integration, Equation (9) is obtained:

$$\frac{J}{t} = R(C_F - C_S) \tag{9}$$

where J/t is the average sampling rate over the period of time from t_1 to t_2. For a zero sink, $C_S = 0$, and $J = M$ (mass of analyte trapped by the coating), so that Equation (10) has the following form:

$$\frac{M}{t} = R \cdot \overline{C}_F \tag{10}$$

where $\overline{C}_F$ is the average concentration of an analyte at the face of the SPME needle during the sampling period. It follows from the above equation that the rate of analyte uptake is directly proportional to the sampling rate (being a function of the geometric dimensions of the sampler and analyte diffusion coefficient) and analyte concentration.

Equation (10) can be rearranged to allow the determination of TWA concentration $\overline{C}_{(t)}$ of an analyte from the mass of the analyte extracted by the SPME fiber

(Equation 11):

$$\frac{M}{Rt} = \overline{C}_{(t)} \tag{11}$$

Sampling rate R can be easily determined by exposing the SPME device with the fiber retracted into the needle to an atmosphere containing a constant concentration of the analyte of interest. In such a case, from Equation (9) we can derive Equation (12):

$$R = \frac{M}{t\mathrm{C}_F} \tag{12}$$

Once R is known, TWA concentration of an analyte in air can be easily determined by SPME from the amount of analyte extracted by the retracted fiber in a given time t.

SPME fiber assembly is very flexible as a passive sampler, since the length of the diffusion path L can be easily changed by repositioning the fiber inside the needle. When high concentrations of an analyte are to be analyzed for prolonged periods of time, the fiber can be retracted deeper into the needle, and vice versa. If the sampling rate R is determined with the fiber at L and the sampling took place with the fiber at L', the left side of Equation (10) has to be multiplied by a factor L'/L to obtain correct results.

The assumption that SPME coating acts as a zero sink is not always fulfilled. Analytes with relatively low affinity to the coating may reveal certain non-negligible vapor pressure at the gas–coating interface, according to Equation (13):

$$\mathrm{C}_S = X_A \frac{P^\circ}{P_t} \frac{MW}{24.45} \frac{298}{T} \tag{13}$$

where X_A is the mole fraction of analyte in the PDMS coating, P° is the pure analyte vapor pressure, P_t is the total gas pressure at the interface, MW is the analyte molecular weight, T is the temperature in Kelvin and 24.45 is the molar volume of gas at 298K. When $\mathrm{C}_S \neq 0$, a reduction in the concentration gradient occurs, as depicted by the dotted line in Figure 2(b). The zero sink assumption is usually fulfilled for highly efficient sorbent materials and for chemisorption, as is the case with on-fiber analyte derivatization.

Figure 3 presents the results of an experiment in which continuous exposure of the SPME fiber assembly with retracted needle for a certain time to a standard gas atmosphere was compared to intermittent exposure, in which the fiber assembly was alternately exposed to the standard gas and pure air. In the latter case, the total time for which the assembly was exposed to the standard gas was equal to the continuous exposure time. Should the zero sink assumption be fulfilled, the amount of analyte from the intermittent exposure should be the same as from continuous exposure. It follows from Figure 3 that

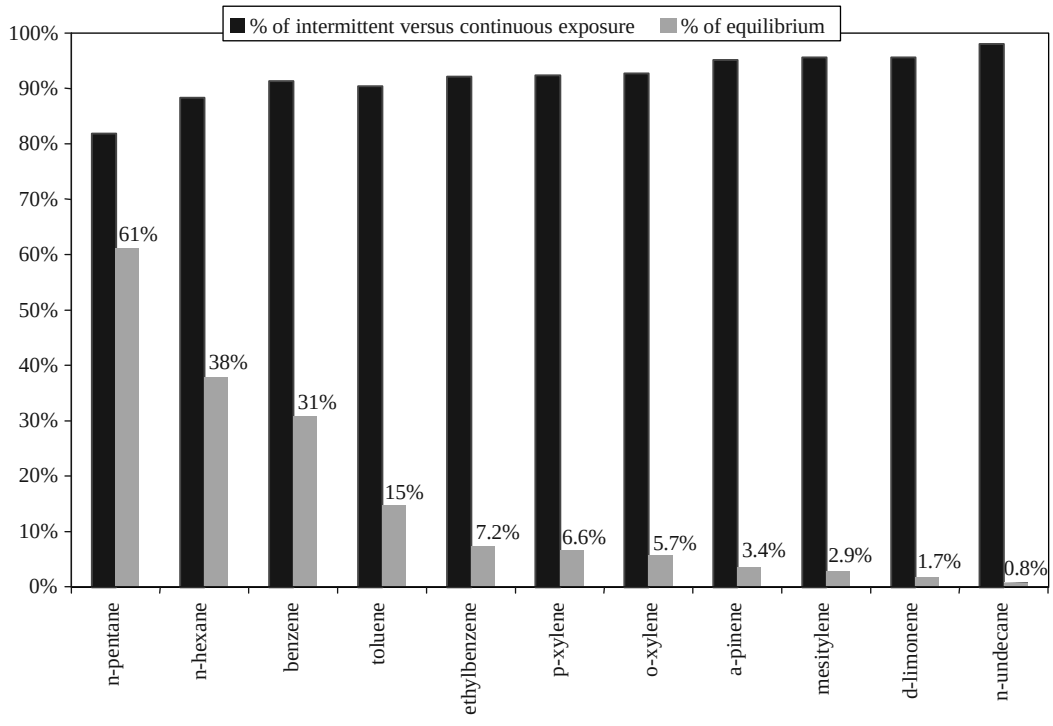

Figure 3 Amounts of analytes trapped by the 100 μm PDMS fiber after intermittent exposure relative to continuous exposure (black bars; total exposure time 15 min in both cases) and the fraction of equilibrium amount trapped by the fiber after the continuous 15 min exposure (gray bars).

PDMS coating does not act as a zero sink for analytes with relatively low affinity to this coating. Assuming that a 10% loss is acceptable, the useful range starts with compounds characterized by fiber/gas partition coefficients (K_{fg}) greater than 1000. Short-term (<30 min) TWA sampling is possible for such compounds.

Figure 3 also presents what fraction of equilibrium amount was loaded on the fiber after 15 min continuous retracted fiber exposure. The data indicate that when the amount extracted is greater than ~5% of the equilibrium amount, analyte mass uptake rate becomes significantly affected. With this in mind, it is possible to use this method to sample *n*-undecane for 75 min with $L = 0.3$ cm, and proportionally longer with larger L.

Applicability of this method for field use was proven in a study where styrene ($K_{fg} \approx 3100$) was determined in an industrial facility. The fiber was retracted 3 mm into the needle, and the sampling time was 30 min. The results were compared to those obtained using 30 min charcoal tube samples. Concentrations of 54 μg L^{-1} (25 °C) versus 56 μg L^{-1} (25 °C) were determined for the charcoal tubes and SPME, respectively.[15]

PDMS coating is not an ideal sorbent for passive sampling of volatile and/or polar analytes. Sorbent capacity for such compounds is low, and the non-zero analyte vapor pressure at the surface of the coating affects the mass uptake rate. One way to remediate problems of this nature is to derivatize the analyte on the coating to a stable product with low vapor pressure and high affinity to the coating. Such an approach was verified with on-fiber derivatization of formaldehyde using *o*-(2,3,4,5,6-pentafluorobenzyl) hydroxylamine (PFBHA) with the PDMS/DVB fiber in the retracted position.[20] Experiments in which the fiber doped with PFBHA was alternately exposed to a standard formaldehyde gas mixture and pure air indicated that a cumulative amount of the derivatization product (oxime) was formed with no measurable losses caused by intermittent exposure to pure air. Thus, the system fulfilled the 'zero sink' criteria. No significant losses of derivatizing agent were observed when the doped fiber was stored inside the needle of the SPME device, whose tip was inside an empty 1.8 ml vial sealed with a Teflon-lined septum during transportation to the field. After sampling, the fibers were carried to the laboratory for instrumental analysis. In another experiment, the doped fiber was exposed to a standard mixture containing 636 ppbv formaldehyde for 1007 min ($L = 3.0$ cm). TWA concentration found with this system was within 7% of the expected concentration.

The device was tested in the field against the standard of the National Institute for Occupational Safety and Health (NIOSH) Method 2541. The fiber was retracted to 3.0 cm, and the sampling time was 420 min. Sampling was carried out in three separate locations. The TWA concentrations found for the SPME sampler were 109, 152 and 57 ppbv, compared to 102, 160 and 51 ppbv, respectively, for the reference method. The excellent agreement between the results obtained by the two methods indicates that passive sampling of formaldehyde using SPME can be a very successful alternative to currently used techniques, as it offers a number of unique advantages: the devices are small, lightweight, reusable and cheap, and the manipulations are minimal and very simple. Once the sampling is finished, the results can be obtained within minutes by thermally desorbing the derivatization product in a portable GC injector for chromatographic analysis.

On-fiber derivatization combined with portable fast-GC, previously described in section 3.1, can also be used for grab sampling of formaldehyde. In this case, the fiber is directly exposed to the atmosphere examined for sampling. The sampling time is selected according to formaldehyde concentration in the air. Very short sampling times (a few seconds) are used for high concentrations. When the expected formaldehyde concentration is low, the sampling time can be extended to a few minutes. In general, the sampling time is selected in such a way that the amount of analyte trapped is within the linear part of the uptake curve.

Grab sampling was carried out simultaneously with TWA measurements described above. The doped fiber was exposed for 30 s to the air and then immediately desorbed in the GC port. The entire procedure of fiber loading/sampling/sample preparation/separation and detection took 3 min with the portable system. Formaldehyde concentrations found in the same locations were 106, 140 and 93 ppbv, respectively. The agreement with the results of TWA sampling was therefore excellent taking into account the completely different nature of the two types of measurements: grab sampling represents momentary concentrations, while TWA sampling reflects concentrations averaged over a long period of time. The limit of detection for the grab sampling method was of the order of 10 ppbv (30 s sampling), which is well below the limits of detection of other formaldehyde grab sampling methods.

3.3 In Situ Analysis of Groundwater and Soil Gas

Volatile content in groundwater and soil should be measured on-site, because composition changes can occur during sampling, pumping or excavation. SPME can be successfully applied to monitor underground pollution as an alternative to other methods that showed some limitations in field applications.[21] A device designed for this purpose (Figure 4) was developed by Nilsson et al.[21] and tested in the field for the determination of VOCs in

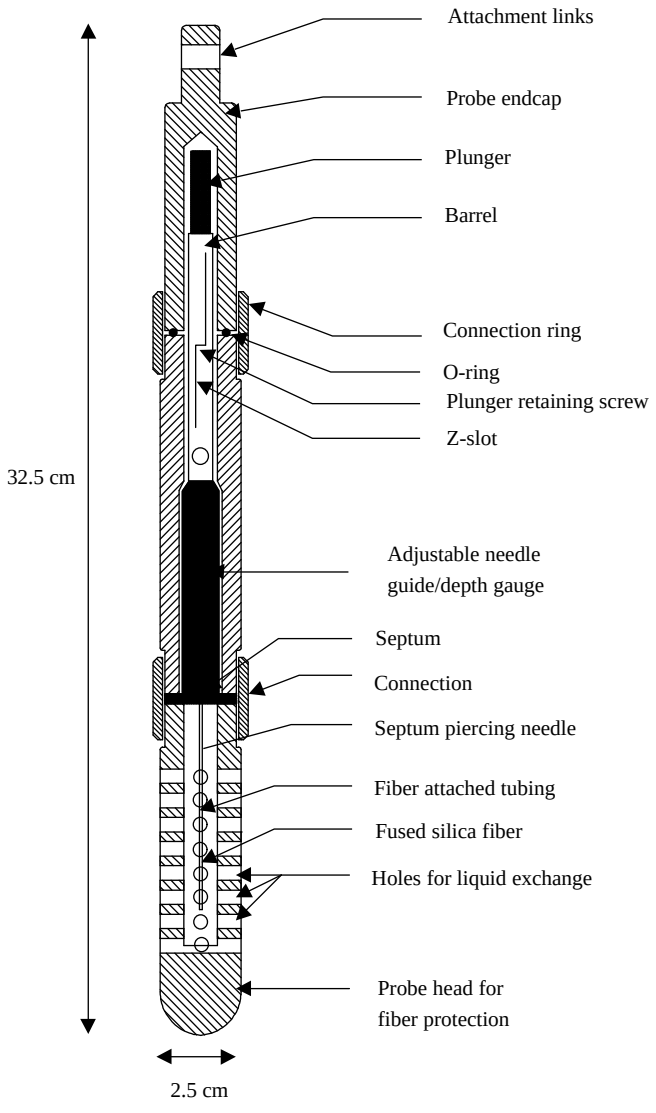

Figure 4 Design of a prototype SPME probe for in situ sampling of groundwater and soil gas. (Adapted by permission of Gordon and Breach Publishers from Nilsson et al.[21])

groundwater and soil gases. The SPME sampling probe was introduced into the well, to reach groundwater for direct sampling. Monitoring of groundwater headspace or underground gases at different depths was accomplished by placing the SPME sampler at different depths, or over the headspace of the water. After extraction by exposure of the fiber, the probe was retracted to the surface and the SPME fiber was immediately desorbed in the GC of a mobile laboratory. Results obtained by underground sampling with the SPME probe were compared to concentrations found for the same samples collected with traditional procedures and analyzed by SPME in a mobile laboratory. VOC levels were evaluated in the underground waters of several wells. The results provided by in situ SPME extraction of groundwater were, in general, higher than those obtained with traditional sampling methods, where the pumping of the sample to the surface causes volatile compound losses. Toluene and naphthalene were detected by underground soil gas investigation and their presence was confirmed by extracting the same samples with Tenax tubes. To sample the volatile content of the soil, during excavation, an alternative SPME probe was designed to fit directly inside the head of a cone penetrometer. In this application of SPME, different ground levels can be monitored.

4 NEW ADVANCES IN SOLID-PHASE MICROEXTRACTION FIELD SAMPLERS

If a portable instrument is available, the SPME fiber can be immediately desorbed after sampling in the GC injector, and storage of the sample on the fiber is not necessary. If the instrument is not movable to the field, the sample has to be carried to the laboratory. Collection and transportation of bulky and/or heavy samples of water, soil or air can be avoided by using SPME for field extractions. The fiber can be exposed to the matrix directly where the sample is located, and only the SPME sampler, with the fiber sealed in an appropriate way, has to be transported to the laboratory. In this case the minimization of sample losses from the SPME fiber is crucial for the reliability of the results. A dedicated SPME sampler for field applications has to be designed to preserve the integrity of the sample during the transportation between the sampling location and the laboratory. The purpose of a recent study on the optimization of SPME field samplers was to achieve no losses during storage, even of volatile compounds, adsorbed/absorbed onto the SPME fiber by the time the sample is analyzed.[22,23] An SPME device for field sampling should also be simple, easy to handle and to transport, robust, inexpensive and, in some cases, disposable, in order to avoid any cross-contamination due to multiple extractions of dirty matrices. It is important that the fiber is protected during the sampling procedure, so that it does not break when it is exposed to the matrix in field.

To date, several approaches have been studied to prevent analyte losses from the SPME fiber during storage.[24] Various silicone rubber septa were tested as sealing materials for the needle during storage in the SPME sampler. By sealing the tip of the needle, the fiber was supposed to be stored in a closed environment. However, this method did not take into consideration the possible partitioning of the analytes released from the fiber during storage into the septum itself. Taking into account that a septum is made of nearly the same material as the PDMS fiber, the significance of such a

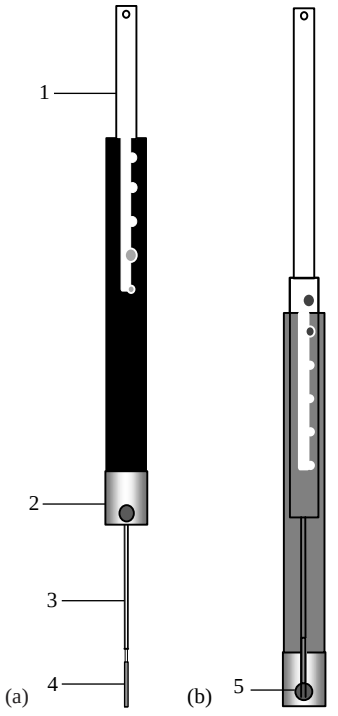

Figure 5 Schematic diagram of an SPME field sampler commercialized by Supelco. (a) Sampling position with fiber exposed to the sample. (b) Storage position with the fiber retracted inside the needle and the tip of the needle sealed by a septum located inside the nosepiece. 1: plunger; 2: aluminum nosepiece; 3: needle; 4: SPME fiber; 5: thermogreen LB-2 septum (Supelco).

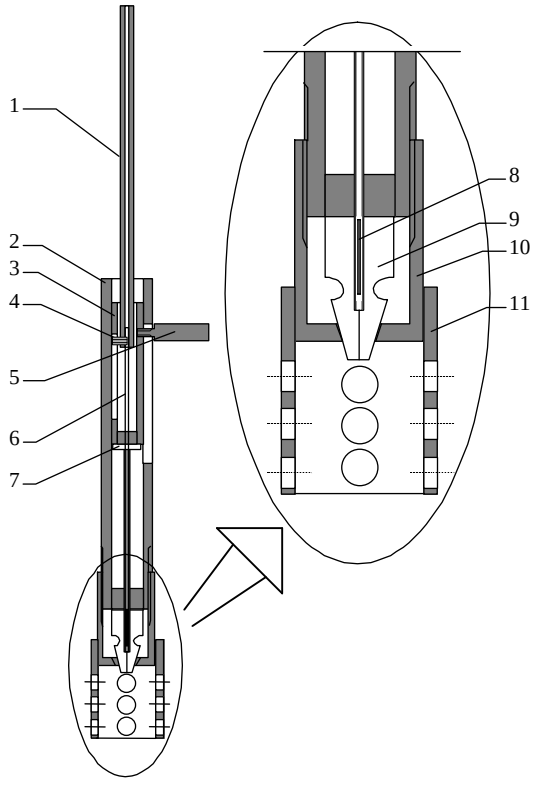

Figure 6 SPME field sampler with a two-leaf closure. 1: plunger; 2: aluminum body; 3: movable barrel; 4: set screw; 5: thumb screw; 6: inner tubing; 7: SPME fiber assembly; 8: SPME fiber; 9: two-leaf closure; 10: finger-tight nut; 11: nylon shield.

phenomenon should not be underestimated, especially for volatile analytes. Also, losses due to permeation of volatiles through the septum are possible. A field SPME sampler, commercialized by Supelco, is presented in Figure 5.[22] The fiber is exposed to the sample by pushing the plunger (1) of the device during the extraction process, as presented in Figure 5(a). Sealing the device for sample storage is achieved by retracting the tip of the needle (3) into the sealing septum (5) (Thermogreen Septum LB-2, Supelco) placed in the aluminum nosepiece (2) of the device (Figure 5b). In this way, the SPME fiber (4), carrying the analytes, is stored inside the sealed needle. The lifetime of this device is related to the longevity of the fiber, which is not replaceable.

Some alternative devices were designed and built for field sampling (Figures 6 and 7).[23] The samplers were designed with the goal to avoid the use of rubber septa as the sealing material, to eliminate or minimize contamination or losses during storage. Figure 6 presents a field sampler that does not use a septum to seal the needle. Instead, the tip of the needle is closed by squeezing a two-leaf closure. The body of the sampler is made of aluminum. Moving the barrel (3), whose position can be fixed by thumbscrew (5), can change the length to which the needle is exposed. The SPME fiber (8) can be exposed from the needle by depressing plunger (1), guided by a Z-slot. In the storage position, the tip of the needle is sealed by a nylon two-leaf closure (9) squeezed around it. The closure is sealed by tightening a finger-tight nut (10). Commercial fiber assembly (7) was used in this prototype after removing the colored hub from the internal tubing (6). Should the fiber need replacement, it can be done by introducing the inner tubing from a new fiber assembly through the needle opening and fastening it to the plunger with screw (4). Between-fiber reproducibility is strictly related to the volume of the stationary phase on the fibers. The dependence of the volume on the length of the fiber is linear, while the dependence on the thickness of the coating is quadratic. Therefore, it is very important that the stationary phase thickness and length are reproducible. This feature is guaranteed by the manufacturer.

To sample, the opening between the two halves of the closure (9) is enlarged by unscrewing the nut (10), and the needle is pushed out. To protect the fiber during sampling, a removable nylon shield (11), with large holes

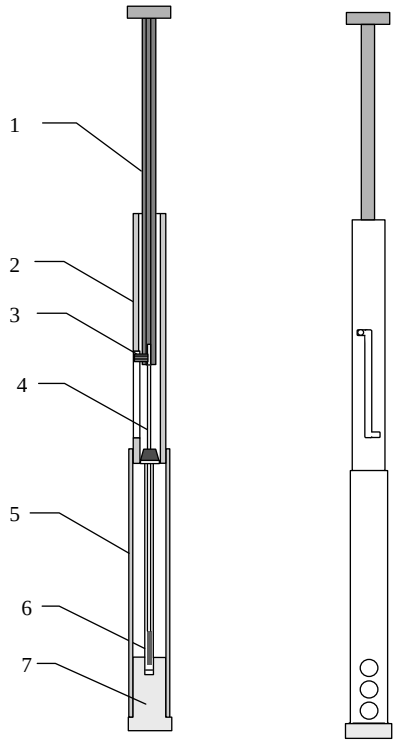

Figure 7 Disposable SPME field sampler with Teflon cap. 1: plunger; 2: nylon body; 3: set screw; 4: SPME fiber assembly; 5: nylon shield; 6: SPME fiber; 7: Teflon cap.

through which air or water can circulate, can be optionally mounted on the nut (10). With this device, no cap has to be removed from the body of the sampler to expose the fiber, so the sampling procedure is more rapid and simple. More advanced versions of the prototype were subsequently prepared, with two new leaf closures made of different materials (Delrin and KEL-F) with a cleaner cut between the two halves. The choice of the material is very important because it cannot be too hard so that a good seal between the two surfaces is achieved (which excludes any metal), but also cannot be too soft (e.g. Teflon) to prevent premature wear caused by the needle movement. Any kind of coating commercially available can be used with this device. The device was tested for storage with CX/PDMS 75 µm, PDMS 100 µm and PDMS/DVB 65 µm fibers.[23]

Figure 7 shows a prototype of a rugged and disposable field sampler made of nylon. The needle tip is sealed during transport with a Teflon cap (7). The opening in the cap fits tightly the outer diameter of the needle. A nylon shield (5) can be mounted to protect the SPME fiber (6) during sampling. The plunger (1), guided by a small screw (3) that moves in a Z-slot, regulates the exposure of the SPME fiber during sampling. Any commercial SPME fiber assembly can be mounted on this device. Since Teflon is inert, it is preferred over silicone rubber as the sealing material for the SPME needle during storage. No interactions between the analytes desorbed from the fiber and the cap should occur during typical storage times. On the other hand, Teflon is soft, therefore the tightness of the cap fit can decreases with use. Besides, the cap can get lost when it is removed from the needle. The device was tested with PDMS 100 µm, PDMS/DVB 65 µm and CX/PDMS 65 µm coatings.[23]

A different concept was also explored for field analysis purposes. An innovative field sampler (Figure 8) was designed and built by modification of a commercially available gas-tight syringe with a special valve for gas sampling (SampleLock 50 µL, Hamilton, model 80956). The syringe has a metal valve (3) that, when it is in the

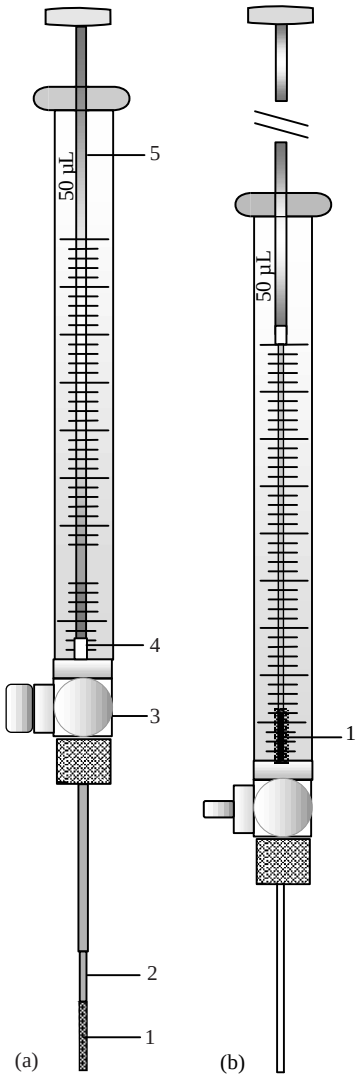

Figure 8 Gas-tight valve syringe modified for SPME field applications. (a) Sampling, (b) Storage with the valve in the closed position. 1: SPME fiber; 2: inner tubing of the commercial fiber assembly; 3: metal gas-tight valve; 4: Teflon tip of the plunger; 5: plunger.

closed position, seals the air sample inside the barrel even under pressure, as specified by the manufacturer. Some modifications of the valve syringe were performed in order to adapt it for use with SPME fibers. A small hole, of a diameter closely matching that of the inner tubing (2) of the commercial fiber assembly, was drilled through the Teflon tip (4) of plunger (5). The stainless steel inner tubing of the fiber assembly was cut at a right length so that it was completely contained inside the barrel when the plunger indicated 50 µL. The tubing was mounted by pressing it into the hole in the plunger tip. Since no glue was used, the fiber could be easily changed or replaced. When the SPME fiber (1) was exposed to the sample for extraction (Figure 8a), the valve was open, the hole inside the valve was aligned with the needle opening, and the inner tubing with the fiber could easily go through. To store the analytes during transport to the laboratory (Figure 8b), the plunger was withdrawn up to the 50 µL mark, so that a fixed amount of air was collected inside the barrel, and the SPME fiber (1) was maintained in the sample environment during storage. Should losses of analytes from the fiber occur via equilibration between the fiber and the air inside the syringe barrel, all the analytes should be injected anyway, since also the air inside the barrel is introduced into the GC column by pushing the plunger. To minimize losses due to desorption of volatiles from the fiber the volume of air in the syringe should be as small as possible. The danger of adsorption of less volatile compounds onto the glass walls of the barrel potentially exists during prolonged storage, but the interaction for the most volatile compounds should be negligible and deactivation (by silanization) of the glass surface of the barrel should prevent any undesirable adsorption.

4.1 Storage Investigations

A very important feature for an SPME field sampler is the sealing capacity during storage and transport of the sampler to the laboratory. Storage capacity of the new field samplers, described in section 4, was investigated by storing several VOCs at different temperatures on selected coatings, as described by L. Müller et al.[23] Gaseous standards of methylene chloride, chloroform, 1,1,1-trichloroethane, benzene, toluene, tetrachloroethylene, 1,1,2,2-tetrachloroethane and 1,3-dichlorobenzene covered a wide range of volatility of the investigated compounds. Air standards (30 µg L^{-1}), prepared freshly for each experiment in 1 L glass bulbs, were extracted by exposing the SPME fiber for 3 minutes under static conditions. The following fibers were tested for storage capacity: PDMS 100 µm, PDMS/DVB 65 µm and CX/PDMS 75 µm. The 100 µm PDMS fiber is particularly suitable for investigations on the sealing capacity of field samplers, because it has the lowest affinity towards very volatile analytes. (Standards of 60 µg L^{-1} were used for experiments with PDMS 100 µm, due to a lower response of this fiber toward the analytes investigated.) The PDMS/DVB mixed coating is a new generation fiber characterized by high affinity towards volatile and certain polar compounds. Analytes are extracted by this coating via adsorption, therefore they do not have to diffuse into the bulk of the coating material, as is the case with 100 µm PDMS fibers that extract analytes by absorption.[6]

CX/PDMS 75 µm is a relatively new fiber created specifically for volatiles. The molecular sieve Carboxen™ fraction of the coating acts as a trap for volatiles, which can be desorbed only at high temperatures. From the physicochemical characteristics, the best storage efficiency should be expected for this fiber, which in theory is the best for field investigations. However, its applications are limited to very volatile compounds only.

At the scheduled time (±3 s), after storage, the fibers were desorbed in the GC injector and the analysis by GC coupled to FID (Varian 3400) was carried out. An SPB-5 column (30 m × 0.25 mm × 1 µm film thickness) from Supelco was used for separation of the analytes with the following GC parameters: oven temperature program 40 °C for 2 min, 15 °C min^{-1} to 150 °C hold for 1 min; injector temperature 250 °C; carrier gas pressure (H_2) 30 psi. Increasingly longer storage times were investigated, from 5 min to 24 h. The dependence of storage efficiency on temperature was evaluated by keeping the samplers at different temperatures: room temperature (24 °C), fridge temperature (4–8 °C) and dry ice temperature (−70 °C). Results presented in Table 1 show the percentage of the analytes retained by the SPME fiber after 24 hours storage, evaluated by comparison with the amount extracted and immediately desorbed in the GC injector. Two values are given, one for the most volatile, and one for the least volatile analyte. If not otherwise specified, the first number corresponds to chloroform and the second number to 1,1,2,2-tetrachloroethane. As expected, cold storage significantly reduces losses of volatile analytes from the fiber. However, certain drawbacks occur due to the low temperature of the device when storing the field sampler in dry ice (−70 °C). They include mechanical problems with the movement of the inner tubing through the septum of the fiber assembly, as the septum becomes very hard at low temperatures. For this reason, after storing any sampler in dry ice, a few minute warm-up at room temperature was required before injection, especially with metal devices. It can be noticed that refrigerated storage (4–8 °C) can generate some unexpected results, e.g. storage percentages of more volatile compounds higher that those of less volatile analytes. This behavior can be explained assuming that significant contamination could have occurred in the refrigerator, where high

Table 1 Summary of results of storage investigations of the prototype field samplers, expressed as percentage of the analyte retained by the fiber after 24 h of storage[a]

	PDMS/DVB 65 μm			CX/PDMS 75 μm			PDMS 100 μm		
	24 °C	4–8 °C	−70 °C	24 °C	4–8 °C	−70 °C	24 °C	4–8 °C	−70 °C
Leaf device (Figure 6)	0–89	75–75	14–81	52–113	86–61	74–88	0–18	0–54	85–90
Teflon capped device (Figure 7)	7–82	124–85	88–92[b]	51–81	72–55	71–70	6–72	38–89	103–91
Supelco field sampler (Figure 5)	–	–	–	32–85	70–54	88–92	0–8	0–37	67–80
Valve syringe (Figure 8)	–	–	–	–	–	–	58[c]–77[d]	86[c]–88[d]	–

[a] The two values in the table describe the percentages for the most volatile and the least volatile compound investigated (chloroform and 1,1,2,2-tetrachloroethane respectively, if not otherwise specified).
[b] This value corresponds to percentage of tetrachloroethylene; less volatile compounds were not quantifiable because of the presence of interference peaks.
[c] This value corresponds to methylene chloride.
[d] This value corresponds to 1,3-dichlorobenzene.

concentrations of solvent vapors were present. This means that good sealing during storage is important not only to avoid losses of analytes from the fiber, but also to prevent contamination from the outer environment where the sampler is kept during transport. This effect is especially noticeable for the CX/PDMS fiber, with the highest affinity towards the most volatile compounds.

As expected, the most volatile analytes are retained to a lesser extent by the SPME fiber during storage. The 100 μm PDMS fiber, which showed the worst storage capacity of all the devices investigated, could not retain any chloroform for 24 h of refrigerated storage with any of the samplers investigated, with the exception of the Teflon capped sampler, which still retained 38% of the compound on the fiber. The performance of the valve syringe was undoubtedly the best, especially considering that a PDMS fiber was used for the investigations: storage of 86% of methylene chloride after 24 h in the refrigerator was an excellent result. Even at room temperature, the device showed much better performance compared to the other designs. Only a 100 μm PDMS fiber was used for storage investigations with the valve syringe device, since it was assumed that if the device works well with the coating that performs the worst, it should work even better with the remaining coatings. Low storage capacity of the Supelco field sampler, evident for long storage times, confirmed that silicone rubber septa are not well suited to prevent the release of volatile compounds from the fiber.

According to expectations, the 75 μm CX/PDMS was the coating that showed the best storage capacity itself, no matter how the needle was sealed. For instance, even after 24 h of storage on dry ice, over 70% of chloroform was still on the fibers with all the devices investigated. Other storage investigations with this coating reported in literature showed that a 3-day storage, even of very volatile compounds, can be achieved without considerable losses from the fiber.[22]

The 65 μm PDMS/DVB fiber was not tested with the Supelco field sampler, because the device was not commercially available with this coating. In general, decreasing the temperature of storage, as shown by experimental data, consistently enhanced storage capacity for PDMS/DVB. Even though contamination occurred during storage due to the presence of organic solvents in the refrigerator, good storage efficiency could be achieved with this coating with the Teflon capped device (Figure 7) when the sampler was kept on dry ice for 24 h (Table 1).[23]

4.2 Field Applications

The authors performed some field evaluations of the recently developed prototypes of the field samplers.[23] Since the valve syringe (Figure 8) was the experimental device that showed the best storage efficiency during laboratory experiments, it was compared to the field sampler commercialized by Supelco in a field investigation of air samples. Analysis of indoor air in an organic laboratory of the Chemistry Department of the University of Waterloo (ON, Canada) was performed with 100 μm PDMS fibers (for this experiment the CX/PDMS fiber in the commercial field sampler was replaced with the PDMS fiber). For field sampling, the air inside the cabinet where organic wastes were stored was extracted by exposing the SPME fibers of the two devices under static conditions for 3 min. The fibers were immediately desorbed and analyzed by GC coupled to FID to obtain reference chromatograms. Extractions were then repeated, and the devices were stored for 150 min at 4 °C. Storage capacity was evaluated as percentage, by comparing the area counts of the

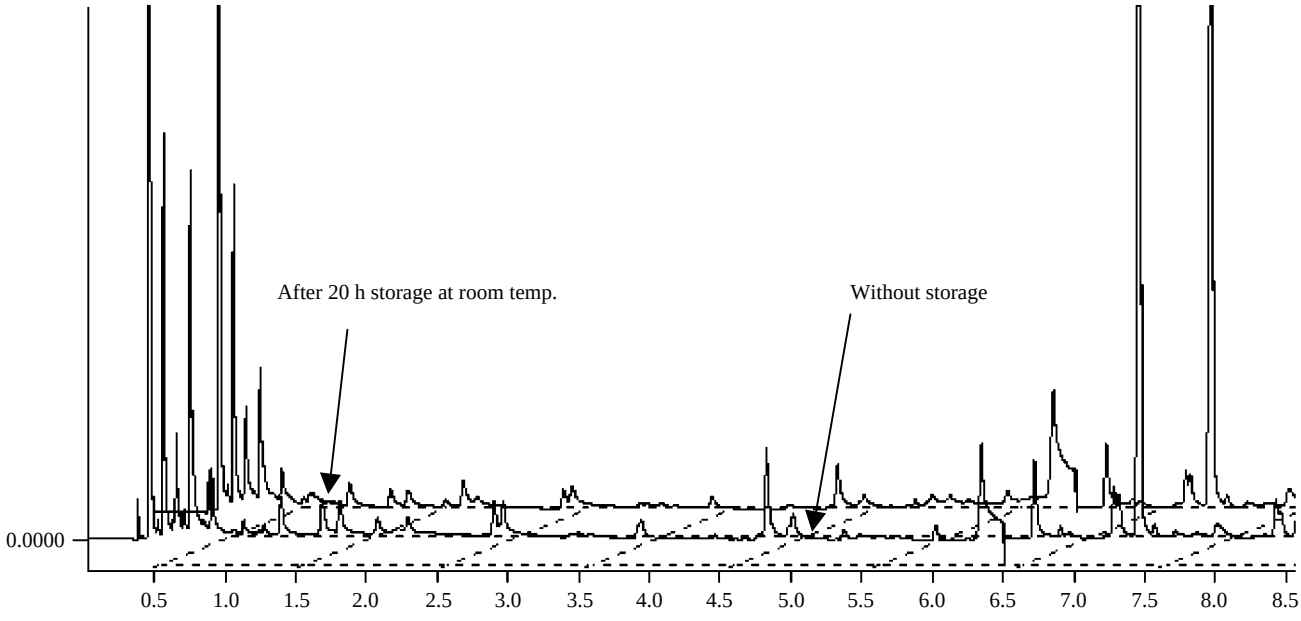

Figure 9 Chromatograms of the scents released by a lilac flower obtained by on-site SPME sampling with the valve syringe device and a CX/PDMS 75 μm fiber.

peaks in the chromatograms before and after storage. This experiment confirmed the superiority of the valve syringe for the storage of volatiles. Compounds from the volatility range between chloroform and toluene were found on the fiber after 150 min at 4 °C in percentages between 53.6% and 60.5% for the valve syringe device, and between 21.1% and 49.8% for the Supelco field sampler. Taking into account that the experiment was carried out with PDMS fibers, the result obtained for the valve syringe was very good.

Another field sampling session with the 50 μL valve syringe was performed with a 75 μm CX/PDMS fiber, which is more suitable for field investigations of volatiles. Scents released by a lilac bush were extracted by wrapping a lilac flower directly on the plant with a plastic bag, and exposing the fiber to the headspace of the flower for 3 min. One sampler was then stored for 20 h at room temperature before desorption in a GC/FID instrument, while the other was immediately carried to the laboratory and analyzed (3 min were required to reach the laboratory). A comparison of the chromatograms obtained with and without storage is presented in Figure 9. Even after 20 h losses of analytes were minimal, as the relative area counts were close to 100%.

The success of this sampling experiment for volatile compounds was determined both by the performance of the sampler, and by the suitability of the CX/PDMS coating to this kind of analysis. The collection of a small portion of air inside the gas-tight syringe improves the sensitivity of the method, and guarantees the injection of almost all the analytes extracted by the fiber during field sampling. Besides, the valve syringe proved to be practical and easy to use for field investigations where retraction of the fiber after sampling had to be fast and easy. The syringe needle was removed during storage, so that contamination due to adsorption of chemicals on the steel walls of the needle was minimized and the device was easier to store and transport.

5 CONCLUSIONS

Several field applications of SPME were recently described in the literature. They utilized either a portable instrument for fast GC, or dedicated devices for SPME field sampling. An increasing number of field applications is expected in the future due to the particular suitability of this technique for field investigations, for a variety of reasons described herein that make SPME an excellent tool for field investigation. Recent development of portable fast GC instrumentation that allows fiber desorption directly in the field further decreased the time of the entire analytical procedure, already reduced by using SPME as sampling and sample preparation technique in one single step rather than in many, as is the case with traditional techniques.

The opportunity of sampling in the field and transporting the device to the laboratory for instrumental investigation seems attractive, if appropriate devices are used. The commercially available device is well suited for sample storage provided a coating most appropriate

for the analytes examined is used. It performs well for volatile compounds when equipped with the Carboxen™ coating. The pen-like device with a leaf closure is a very elegant concept. To perform according to expectations, more sophisticated methods would have to be used for the manufacture of the leaf closure, to assure a complete tightness once the leaves are squeezed by the nut. Once this is achieved, this device would probably be the most convenient for field sampling. On the other hand, the nylon holder with the Teflon cap could easily be manufactured as a low-cost, disposable device. The sealing method used in this holder proved to be superior to sealing with a silicone septum, and the entire device could be manufactured by injection molding. The cap itself could be made of a cheaper material, e.g. high-density polypropylene. It could be strapped to the device to prevent losing it in the field.

Excellent storage, even of very volatile compounds, was achieved using a modified gas-tight syringe equipped with a valve that sealed the environment inside the syringe barrel where the SPME fiber was placed during storage. The device would be ideally suitable for sampling and storage of volatile analytes. Concurrent injection of a small amount of air can potentially increase the sensitivity for the most volatile analytes. The drawback of this device is the relatively high price of the syringe itself. For less demanding applications, the cheaper devices described above would probably be more suitable.

Very promising applications, from indoor air monitoring to groundwater and flower scent analysis, open the frontiers of field sampling by SPME.

ACKNOWLEDGMENTS

Gordon and Breach Publishers and the OPA (Overseas Publishers Association), together with the authors of reference 21, are acknowledged for permission to use some of their material in this chapter. The financial support of Supelco, Varian and Natural Sciences and Engineering Research Council of Canada is also gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

BTEX	Benzene, Toluene, Ethylbenzene, Xylenes
CW/DVB	Carbowax™/Divinylbenzene
CX/PDMS	Carboxen™/Poly(dimethylsiloxane)
DELCD	Dry Electrolytic Conductivity Detector
FID	Flame Ionization Detector
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detector
HPLC	High-performance Liquid Chromatography
LLE	Liquid–Liquid Extraction
MS	Mass Spectrometry
NIOSH	National Institute for Occupational Safety and Health
PA	Poly(acrylate)
PDMS	Poly(dimethylsiloxane)
PDMS/DVB	Poly(dimethylsiloxane)/Divinylbenzene
PFBHA	*o*-(2,3,4,5,6-Pentafluorobenzyl) hydroxylamine
PID	Photoionization Detector
SPME	Solid-phase Microextraction
TCE	Trichloroethylene
TWA	Time-Weighted Average
VOC	Volatile Organic Compound

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants • Formaldehyde, Environmental Analysis of

Environment: Water and Waste cont'd **(Volume 4)**
Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Field-based Analysis of Organic Vapors in Air

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Industrial Hygiene cont'd **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Instrumentation of Gas Chromatography

REFERENCES

1. V. Lopez-Avila, H.H. Hill, 'Field Analytical Chemistry', *Anal. Chem.*, **69**(12), 289R–305R (1997).
2. T. Kotiaho, 'On-site Environmental and In Situ Process Analysis by Mass Spectrometry', *J. Mass Spectrom.*, **31**, 1–15 (1996).

3. J. Pawliszyn, *Solid-phase Microextraction. Theory and Practice*, Wiley-VCH, New York, 1997.
4. Z. Zhang, M.J. Yang, J. Pawliszyn, 'Solid-phase Microextraction: A New Solvent Free Alternative for Sample Preparation', *Anal. Chem.*, **66**, 844A–853A (1994).
5. A. Boyd-Boland, M. Chai, Y. Luo, Z. Zhang, M. Yang, J. Pawliszyn, T. Górecki, 'New Solvent-free Sample Preparation Technique Based in Fiber and Polymer Technologies', *Environ. Sci. Technol.*, **28**, 569A–574A (1994).
6. T. Górecki, X. Yu, J. Pawliszyn, 'Theory of Analyte Extraction by Selected Porous Polymer SPME Fibers', *Analyst*, **124**, 643–649 (1999).
7. R. Eisert, K. Levsen, 'Solid-phase Microextraction Coupled to Gas-chromatography: A New Method for the Analysis of Organics in Water', *J. Chromatogr. A*, **733**(1 + 2), 143–157 (1996).
8. J. Pawliszyn (ed.), *Applications of Solid-phase Microextraction*, The Royal Society of Chemistry, Cambridge, 1999.
9. T. Górecki, J. Pawliszyn, 'Solid-phase Microextraction/ Isothermal GC for Rapid Analysis of Complex Organic Samples', *J. High Resol. Chromatogr.*, **18**, 161–166 (1995).
10. T. Górecki, J. Pawliszyn, 'Sample Introduction Approaches for Solid-phase Microextraction/Rapid GC', *Anal. Chem.*, **67**, 3265–3274 (1995).
11. R. Eisert, K. Levsen, 'Development of a Prototype System for Quasi-continuous Analysis of Organic Contaminants in Surface or Sewage Water Based on In-line Coupling of Solid-phase Microextraction to Gas-chromatography', *J. Chromatogr. A*, **737**(1), 59–65 (1996).
12. M. Harper, C.R. Glowacki, P.R. Michael, 'Industrial Hygiene', *Anal. Chem.*, **69**, 307R–327R (1997).
13. T. Górecki, J. Pawliszyn, 'Field-portable Solid-phase Microextraction System for Trace Analysis', *Field Analyt. Chem. Technol.*, **1**(5), 277–284 (1997).
14. D. Basset, H. Groenevelt, B. Parker, private communication (1998).
15. P.A. Martos, J. Pawliszyn, 'Calibration of Solid-phase Microextraction for Air Analyzes Based on Physical Chemical Properties of the Coating', *Anal. Chem.*, **69**, 206–215, (1997).
16. W.J. Lautenberger, E.V. Kring, J.A. Morello, 'A new Personal Badge Monitor for Organic Vapors', *Am. Ind. Hyg. Assoc. J.*, **41**(10), 737–747 (1980).
17. R.H. Brown, R.P. Harvey, C.J. Purnell, K.J. Saunders, 'A Diffusive Sampler Evaluation Protocol', *Am. Ind. Hyg. Assoc. J.*, **45**, 67–75 (1984).
18. J. Namiesnik, T. Górecki, E. Kozlowski, 'Passive Dosimeters – An Approach to Atmospheric Pollutants Analysis', *The Science of the Total Environment*, **38**, 225–258 (1984).
19. P.A. Martos, J. Pawliszyn, 'Time-weighted Average Sampling with Solid-phase Microextraction Device: Implications for Enhanced Personal Exposure Monitoring to Airborne Pollutants', *Anal. Chem.*, **71**(8), 1513–1520 (1999).
20. P.A. Martos, J. Pawliszyn, 'Sampling and Determination of Formaldehyde Using Solid-phase Microextraction with On-fiber Derivatization', *Anal. Chem.* **70**(11), 2311–2320 (1998).
21. T. Nilsson, L. Montanarella, G. Baglio, R. Tilio, G. Bidoglio, S. Facchetti, 'Analysis of Volatile Organic Compounds in Environmental Water Samples and Soil Gas by Solid-phase Microextraction', *Intern. J. Environ. Anal. Chem.*, **69**(3), 217–226 (1998).
22. R. Shirey, V. Mani, R. Mindrup, 'On-site analysis', *Am. Environ. Lab.*, **1–2**, 21–22 (1998).
23. L. Müller, T. Górecki, J. Pawliszyn, 'Optimization of Solid-phase Microextraction for Field Applications', *Fresenius' J. Anal. Chem.*, in press.
24. M. Chai, K. Arthur, J. Pawliszyn, R. Belardi, K. Pratt, 'Determination of Volatile Chlorinated Hydrocarbons in Air and Water with Solid-phase Microextraction', *Analyst*, **118**, 1501–1505 (1993).

Solid-state Sensors for Field Measurements of Gases and Vapors

Franz L. Dickert and Peter A. Lieberzeit
Vienna University, Vienna, Austria

1 Introduction 1
2 Device Principles 2
2.1 Electronic Devices 2
2.2 Pellistors 4
2.3 Optical Devices 4
2.4 Mass-sensitive Devices 5
3 Sensitive Layers 6
4 Selected Sensing Applications 7
4.1 Electrochemical Sensors 7
4.2 Field Effect Devices 13
4.3 Optical Sensors 14
4.4 Mass-sensitive Sensors 17
5 Conclusion 18
6 Commercial Sensors 18
Abbreviations and Acronyms 19
Related Articles 19
References 20

Remote and in-field analysis is developing into a crucial aspect of modern analytical chemistry. Chemical sensors, which are small measuring devices consisting of a recognition layer, a transducer and the measurement electronics, have proven to be highly suitable tools for this purpose. Although usually they are constructed for the detection of very limited amounts of chemical compounds and often do not reach the detection limits of modern analytical instruments, they show very desirable advantages: the systems are easy to operate, mostly rugged, small, good value and can be manufactured by established technological methods. Due to their size and ruggedness they can even be operated in harsh environments and often they can be constructed to be suitable for remote measurements, where several sensors are interrogated by a central analytical site.

Devices being used include optical fibers, field effect transistors (FETs), electrochemical cells (be they potentiometric or amperometric) and mass-sensitive components of all kinds. The chemically-sensitive coatings range from bare transducer surfaces to dyes reacting with analyte or sterically-specific binding sites (host–guest complexes etc.). A seeming drawback is the fact that only few sensors react specifically towards one defined component but normally show a sensitivity pattern towards chemically-related substances. By using an array of several components followed by modern methods of data evaluation (such as neural networks) a group of analytes can be detected and characterized simultaneously. This article introduces device and recognition layer principles as well as covering field sensors published during the last one and a half decades based on all the different kinds of transducers and sensor layers. Some of the devices (such as the lambda probe) are already marketed, others are developed to a stage where field measurements have already been carried out, but which are not market-ready.

1 INTRODUCTION

Public sensitivity towards environmental pollution has greatly increased during the last few years. This has resulted in a growing demand for analytical tools suitable for the fast and economic determination of contaminants in soil, water and air. Additionally, remote and in-field measurements are becoming more and more important.[1] Different strategies can be used for fulfilling these needs: the first possibility is to miniaturize classical analytical instrumental techniques such as gas chromatography (GC),[2] mass spectrometry (MS) and electrophoretic systems and also detectors such as flame analyzers,[3] and the second is the use of chemical sensors. The latter have established themselves as an important field of analytical chemistry and are defined as follows:[4] ''Chemical sensors are (small) devices capable of continuously recognizing concentrations of chemical constituents in liquids or gases and converting this information in real time to an electrical or optical signal''. Hence the aim of a sensor is to translate chemical information into a measurable quantity, usually an electronic one. Figure 1 shows the scheme of a chemical sensing system.[5] It consists of a sensitive layer, a transducer and a data processing unit. The sensitive layer is exposed to the analyte and is able to interact with it. Upon this interaction, one of the layer's physical properties changes (e.g. polarity, optical absorbance, fluorescence behavior, mass).[6] These changes are detected by the transducer, which converts them to an electrical signal. Finally, the data obtained are processed and stored.

Usually a chemical sensor does not react specifically towards a defined compound but shows a certain sensitivity pattern towards a class of chemical substances. This can be overcome in two different ways: either the samples are preseparated (which actually contradicts the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

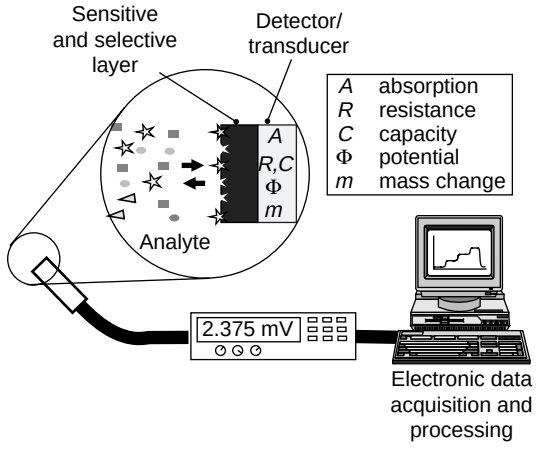

Figure 1 Schematic set-up of a chemosensor.

premise that a sensor should be an apparatus easy to handle even for persons not having special training in chemistry) or an array consisting of sensors with different selectivity patterns is used.[7–9] Preferably the number of sensors in an array exceeds the number of analytes to be measured simultaneously by at least one. The resulting sensor patterns can be evaluated by modern methods of data analysis[10] such as partial least squares or neural networks. A comparison and evaluation of different pattern recognition systems were given by Shaffer et al.[11]

Another great advantage of sensors for field measurements is the fact that many sensor principles allow the construction of remote sensing systems. This can result in a multisensor system controlled by one center of analysis, e.g. for waste monitoring along a river.

Summarizing, chemical sensors have several advantages: they are usually very small devices being highly suitable for all kinds of remote analysis applications. Owing to their simplicity compared with classical analytical instruments, chemosensors are rugged, do not require too much maintenance and can also be operated by non-professional personnel. Sensing systems allow on-line and in-time monitoring of specified analytes and thus are highly suitable for all kinds of applications, where it is important to trace a specified analyte over a longer period of time. As a result of miniaturization, several different sensors can be integrated on one device, making multianalyte measurements and elimination of cross-sensitivity possible, and most sensor devices are suitable for field use as a matter of principle. Further, sensor systems are usually constructed so as to be good value so that they can be sold at a reasonable price to a broad spectrum of customers.

Certainly there are also some inherent limitations. First, chemical sensors usually do not react specifically with one analyte but show certain sensitivities towards closely related substances. This makes the elimination of "cross-talk" by other analytes a tedious task.[12] Second, classical analytical instrumentation also usually offers the possibility of dealing with a wide variety of different analytes, i.e. they offer higher flexibility (although at a much higher price). Third, unfortunately, aging of the sensitive layers still imposes limitations on chemical sensing.

2 DEVICE PRINCIPLES

A wide variety of devices are suitable as transducers in chemical sensing, which can be classified according to the signal generation principles. One can distinguish between different kinds of electrochemical (potentiometric, amperometric, etc.), optical and mass-sensitive devices as well as pellistors, which will be discussed in the following sections.

2.1 Electronic Devices

Some of the commercially most successful sensor devices such as the Figaro sensor,[13] the lambda probe[14] and the chemically sensitive field effect transistor (ChemFET) operate with components measuring an electrochemical or electronic quantity, e.g. changes in surface charge, in chemical potential or in current.

2.1.1 Potentiometric Sensors

Several of very popular chemical sensors are based on potentiometric devices, whose measurand is the potential difference between a measuring electrode and a reference caused by different analyte concentrations. An early integrated device for this purpose is the glass electrode for pH measurements (Figure 2), which today is standard laboratory equipment and actually would not now be called a sensor owing to its size. Continuous development led to a highly integrated device, the FET, which will be discussed separately. Other examples of potentiometric sensors are all kinds of ion-selective electrodes with either solid (e.g. LaF_3) or liquid membranes. The most widely used potentiometric sensor based on ionic conductivity in a solid is the lambda probe.[15] It consists of ZrO_2/Y_2O_3 where charge is transported via oxygen ions. The electrolyte separates the sample and reference chambers and the signal is obtained in this concentration cell according to Nernst's law.

2.1.2 Conductometric Devices

A widely applied conductometric sensor is the so-called Taguchi or Figaro sensor, which was first used commercially in 1968[16] for the monitoring of ambient combustible gases in households. It consists of SnO_2,

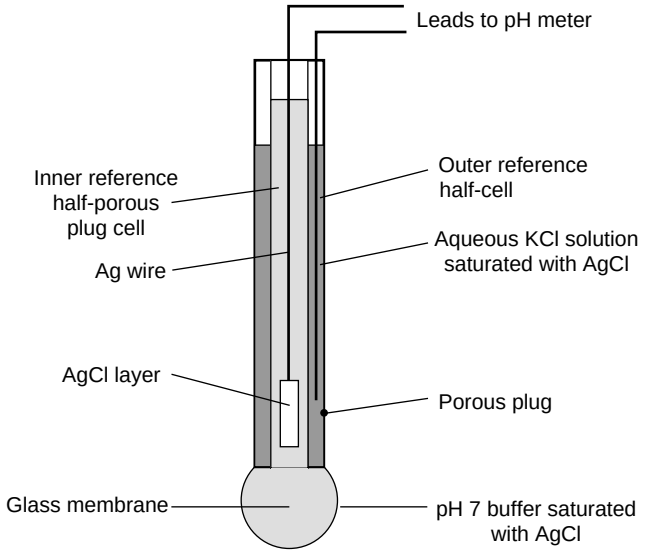

Figure 2 Schematic set-up of a glass electrode with included reference.

usually doped with a noble metal such as Pt or Pd. In the crystal lattice oxygen defects are generated, resulting in an n-type semiconductor. When exposed to the atmosphere the ceramic material adsorbs oxygen, which is reduced by conducting electrons of the semiconductor, resulting in a change in surface resistivity. Ambient combustible gases react with the oxygen present on the device, thus reversibly decreasing its surface concentration, which leads to an increase in conductance due to an increase of electrons in the matrix. These kinds of sensors also are highly field-capable: although they have to be heated (usually they operate at temperatures of about 350–500 °C), they can easily be miniaturized, thus dramatically lowering their power consumption. This also allows the feasible combination of several devices to give an array without the need for too much space, which offers the possibility of simultaneous multicomponent sensing.

2.1.3 Amperometric Devices[17]

Amperometric devices use the current accompanying an electrochemical reaction as a measurand. One of the best known examples of this kind of sensor is the so-called Clark electrode for the detection of gaseous oxygen. It consists of a platinum and a silver electrode (working as anode) and KOH as electrolyte. Oxygen is reduced and silver is oxidized. Although it is a powerful sensor, the Clark cell also shows the inherent disadvantage of all amperometric sensors: the silver electrode is dissolved during the measurement (and precipitates as silver chloride), thus allowing only a restricted operational time. To economize on silver, an auxiliary Pb electrode can additionally be used, although this does not alter the fact that the electrode material is consumed during measurements. Nonetheless, this type of sensor can be used for discardable one-way systems owing to the usually low costs.[18] Other examples of amperometric cells include biochemical systems, e.g. for glucose.[19]

2.1.4 Field Effect Devices

One can distinguish between two principal types of electronic devices, namely active and passive devices. Active devices directly yield a comparably huge current by using comparably low control voltages, i.e. they usually act as amplifiers yielding a more powerful output than the input signal. Field effect devices exist as both active and passive components. Capacity diodes based on a metal–insulator–semiconductor design are passive whereas FETs are active devices. Two acronyms are often used with respect to field effect devices, namely MOS capacitor and diode, where MOS means metal oxide semiconductor and MIS meaning metal insulator semiconductor. The two abbreviations denote the sandwich structure of which the device is made and are synonymous.

MIS capacitors and diodes consist of a semiconducting material (usually Si) on which a thin insulating layer is generated by oxidation (often SiO_2). Finally, a metal coating, usually Pd or Pt, is applied. When using a capacitor, the voltage dependence of the capacity under changing ambient conditions is observed; in the case of a diode the current–voltage characteristic is measured. The signal is obtained by adsorption and dissociation of analytes on the metal surface. Additionally, there is an equilibrium between atoms adsorbed there and those adsorbed on the boundary between metal and insulator where they become polarized leading to a shift in voltage (Figure 3).

Bergveld introduced the first ion-sensitive field effect transistor (ISFET) in 1970,[20] which can be regarded as the beginning of chemical sensors based on FETs. The principle of a FET[21] is shown in Figure 4. It

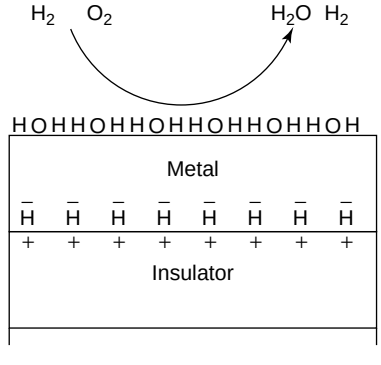

Figure 3 Reactions and equilibria occurring in metal–insulator devices.

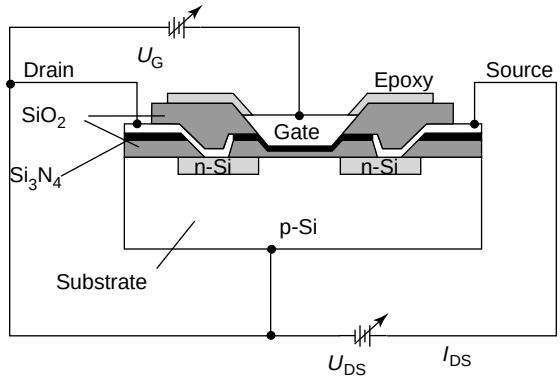

Figure 4 Schematic set-up of a FET.

normally consists of a p-type semiconductor with two n-doped regions near to each other, the source and the drain which are contacted by the respective electrodes. The third electrode (the gate) is mounted above the region between the source and drain. By variation of the potential at the gate electrode, the current between the other two electrodes is influenced, the chemically sensitive layer is applied to the gate electrode. The great advantage of FETs is their fast reaction towards changes in ionic or dipolar properties and the fact that they provide directly an electronic signal that is easy to measure. Additionally, they react very fast to changes in their chemical environment. Nonetheless, they have an inherent limitation: the signal generated shows long-term drift if the isolation materials are not perfectly manufactured. Existing temperature drift can be compensated for by combining at least two devices on one semiconductor substrate, where one FET is tightly encapsulated and used as a reference field effect transistor (REFET). Existing external limitations can be overcome by proper encapsulation and electronic isolation and also by using a REFET. In an array one can additionally measure the difference signal between the sensor and reference, thus improving the signal-to-noise ratio.

FETs can also be used for gas sensing, e.g. for the detection of hydrogen. In this case the gate is coated with a noble metal (e.g. Pt or Pd), where the sensing mechanisms are the same as for MIS devices. Variation of the metal coating results in enhanced selectivity towards different analytes. A novel approach is the use of a suspended-gate field effect transistor (SGFET), where a thin air gap (about 100 nm) is placed below the gate electrode, where the adsorption takes place.

Another field of application for FETs is biosensing: here usually a bioactive layer is deposited on an ISFET (e.g. enzymes such as peroxidase). Usually the actual measurand is a change in hydrogen ion concentration via a pH-sensitive layer. A wide variety of biosensors can be found in the literature.[22]

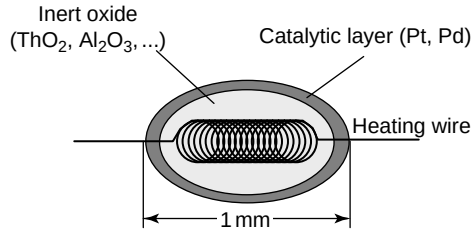

Figure 5 Pellistor.

2.2 Pellistors

Pellistors are used for the measurement of reducing gases and consist of a ceramic material (mostly ThO_2 or Al_2O_3) doped with a catalyst (Pt or Pd) and a thermoresistive element[23] (see Figure 5). The entire device is heated. Ambient reducing gases are adsorbed on the surface and react catalytically with oxygen. The resulting combustion heat changes the sensor temperature, resulting in a variation of resistivity. Pellistors are easy to produce and cheap and very reliable but have the major drawback of not being selective.

2.3 Optical Devices

In contrast to the devices introduced up to now, the chemical information in the case of an optical sensor is not directly transformed into an electronic quantity but is obtained as a change in an optical property, such as absorption or fluorescence. Wolfbeis edited a comprehensive introduction to the field of fiber-optic chemical sensing.[24] Optical methods making use of all parts of the electromagnetic spectrum are well established in analytical chemistry and for structure determination. The breakthrough in the use of these methods for chemical sensing applications came with the development of waveguide fibers. These were originally designed for telecommunications technology and made it possible to transport light over long distances using a very small cross-sectional area. For sensing applications this offers the major advantage of bringing the device to the potential analyte instead of the other way round. The distances covered by fibers commercially available in large quantities at a suitable price can reach up to 1000 m. The next major improvement in optical chemical sensing was the development of the inexpensive light-emitting diode (LED) for wavelengths reaching from the ultraviolet (UV) to the near-infrared (NIR) region. Two different types of optical sensors can be distinguished: plain fiber sensors, making use of the direct influence of the analyte on the optical information, and indicator-mediated sensors, where the analyte interacts with a dye (forms an equilibrium), which results in a change of the dye's optical behavior.[25,26] Fiber-optical systems additionally offer the possibility that a bundle of fibers reaching different spots of

analytical interest can be combined and interrogated by a single controlling device, thus leading to a comparably easy possibility of supervising a larger area via a central basis station. This also makes complicated sampling procedures redundant, which especially in the case of environmental monitoring can be very time-consuming. Optical sensors have a wide variety of advantages:

- light can be transmitted over long distances, thus bringing the measuring device to the analyte;
- if several fibers optimized for different analytes are combined in a bundle, simultaneous monitoring of several different compounds is possible;
- optical methods usually do not consume analyte and leave it chemically unaffected;
- fiber optics are produced from materials that usually are highly stable and thus can also be operated in harsh environments;
- the light can be guided to different locations, thus offering the possibility of comparably easy measurement of spatial distributions of an analyte.

Some of the drawbacks are as follows:

- indicators tend to undergo photobleaching or bleeding, which is especially adverse for measurements in the liquid phase;
- ambient light can cause undesirable interferences.

Although almost all methods of optical spectroscopy are suitable for chemical sensing applications, mostly fluorescence and absorption spectroscopy are used, preferably in the visible and near-UV range of the spectrum, even though devices operating in the NIR and middle-infrared (MIR) regions are becoming increasingly important.

A special class of chemical sensors is based on surface plasmon resonance (SPR).[27] Such a system consists of a prism brought to optical contact with a thin noble metal layer. When measuring the reflectance of this metal as a function of the angle of the incident beam of light, a sharp minimum occurs. At that distinct angle SPR is induced, which leads to a change in light reflectivity. By application of a suitable chemically sensitive layer to the metal surface, the effect can be used for chemical sensing; especially biosensors based on this effect have been reported. Thus SPR phenomena, which are oscillations in the metal electron gas, are strongly dependent on the properties in the direct vicinity of the metal and therefore react towards chemical changes in their environment.

2.4 Mass-sensitive Devices

Monocrystalline piezoelectric materials such as quartz (actually the most important material in this field) and lithium tantalate can be used very advantageously in chemical sensing. On applying an alternating current to a crystal plate of such a material cut at a specific angle, a mechanical oscillation is induced, and both surface and bulk waves are observed (depending on the cutting angle). It was shown by Sauerbrey in 1959[28] that the resonance frequency of such an oscillator is highly dependent on the mass present on the electrodes. Hence these devices act like small and highly sensitive balances. In combination with chemically sensitive layers suitable for the selective inclusion of a certain analyte, they form powerful sensors.

The most widely used mass-sensitive devices are quartz microbalances (QMBs) and the surface acoustic wave (SAW) resonator. A QMB (see Figure 6) consists of an AT-cut quartz plate on which an electrode is applied on each side. When an alternating current is applied, a thickness shear wave is generated in the substrate, the resonance frequency of which is mainly determined by the thickness of the quartz plate. Commercial QMBs are available for frequencies up to 20 MHz; above that limit the plates would become too thin to be mechanically stable. In sensor applications usually resonance frequencies around 10 MHz are used. Despite their relatively low fundamental frequency, QMBs are a powerful tool in sensing applications, especially when ruggedness and comparably low sensitivity towards

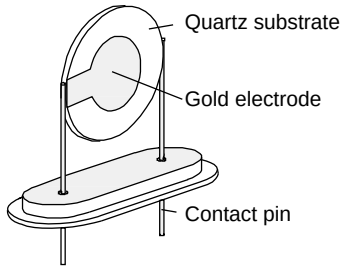

Figure 6 QMB.

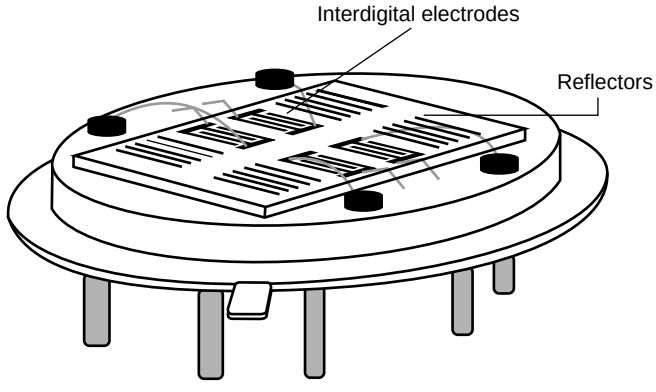

Figure 7 SAW.

ambient parameters (e.g. high damping by liquids) is needed.

SAWs, on the other hand, are produced from ST-cut quartz plates where an interdigital transducer (IDT: comb electrode; see also Figure 7) is generated by a lift-off process. Application of an alternating current creates a surface wave in the quartz, which is detected by a second IDT on the substrate. SAWs can be operated in frequency ranges up to 2.5 GHz and analytical applications have been reported up to 1 GHz,[29] which means greatly increased sensitivity towards mass loading compared with QMBs. The drawback is that normal SAWs cannot be used in the liquid phase owing to excessive damping of the surface wave. A possibility of overcoming this problem is the use of the so-called shear wave resonator.[30]

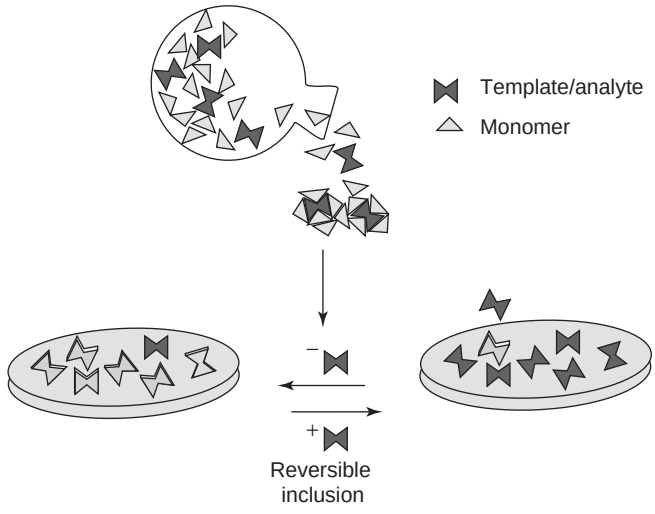

Figure 8 Molecular imprinting.

3 SENSITIVE LAYERS

Many of the devices introduced above cannot distinguish between different chemical species (e.g. mass-sensitive devices, some optical devices) and thus need a layer whose interaction with the analyte is then transformed into the electronic or optical signal, respectively. In general there are two different types of sensitive layers: the first interacts with the analyte forming an adduct or something similar due to chemical interactions or reactions, whereas the second offers spatially fitting cavities for analyte inclusion which can additionally offer binding sites for polar interactions such as hydrogen bonds (figuratively there are convex interaction sites, where the analyte is somehow captured by a surface, be it external or internal, and concave sites, where the analyte is encompassed in a cavity of the matrix). The interactions used for sensing applications include dipolar interactions, hydrogen bonds, $\pi-\pi$ interactions, donor–acceptor interactions and discrimination via steric properties.

Affinity layers usually consist of a medium holding the respective receptor. The best known examples of such materials are all kinds of pH indicators, which are widely applied in sensing, but also other dyes showing, e.g. solvatochromism. Other examples of affinity layers are polymers containing polar groups or side chains, e.g. for humidity monitoring.

Two principal approaches are used for the synthesis of cavities tailored for the spatial requirements of a certain analyte: supramolecular chemistry and molecular imprinting. In supramolecular chemistry, molecules containing a cavity with a size suitable for the incorporation of a special analyte are synthesized. Examples of molecules offering cavities of suitable size are differently substituted cyclodextrins, paracyclophanes, calixarenes and calixresorcinarenes, all of which have already been reported in sensing applications.[31] The synthesis of this class of compounds is often assisted by computer modeling to obtain hints for faster manufacturing. Nonetheless, their synthesis is rather sophisticated and time-consuming. For increased chemical and mechanical stability, these cavitands can be embedded in a supporting polymer.

The other innovative approach to the synthesis of potential sensor materials is the method of molecular imprinting[32,33] (see Figure 8). Here cavities are generated in a highly cross-linked polymer by mixing its precursor compounds with the analyte, forming a kind of synthetic antibody. Polymerization takes place in the presence of the template, which does not participate in the formation process. When synthesis is completed, the template is removed from the material, leaving behind adapted cavities and diffusion pathways. The great advantages of this method are both the compatibility with technological processes, which makes it very promising for mass fabrication, and the speed with which the respective layers are obtained. Large amounts of cross-linker are used for synthesis to stiffen the polymeric material, thus preventing the system from collapsing after removal of the template molecule. This also results in mechanically and chemically stable materials which show advantageous features for use in harsh environments.

For molecularly imprinted polymers, it is usually difficult to do preliminary computer calculations, but fortunately this is not required as the steric properties of the template exert a direct influence on the shape of the hollows. In the initial stage of research, the template was linked to one of the monomer compounds either by covalent bonds or by polar interactions, which adds chemical recognition to the sterically fitting cavities. Nonetheless, there are also possibilities of generating suitable hollows determined only by van der Waals interactions, thus offering purely steric recognition sites. For the synthesis

of molecularly imprinted sensor materials one does not need to know the exact chemical composition of later analytes. This leads to the possibility of designing sensors even for substances which are not exactly known, or for complex mixtures, thus pushing forward the frontiers of chemical sensing.

4 SELECTED SENSING APPLICATIONS

Many papers on sensors are published every year and many of them are related to environmental sensing applications. In this article we want to concentrate more on sensing systems for environmental and ambient parameters than on sensors specifically designed for industrial or technological process control. In the following sections the sensors will be arranged according to their transducing principle rather than according to the analytes of interest. Nonetheless, some general articles and reviews will be included, such as those dealing with environmental sensing.[34–39] A review on air pollution and its detection methods was published by Fox in 1987,[40] and environmental gas sensing has been reported.[41–43] Fox's paper covers mainly different publications on pollution monitoring by a variety of analytical methods and also includes an extensive list of references. An electrochemical sensor array for the simultaneous determination of water quality parameters (pH, conductivity, temperature, redox potential and dissolved oxygen concentration) has been reported and evaluated by a group from Southampton.[44] In this case the sensor approach is illustrated by combining several measuring devices suitable for very different analytes or physical quantities. In combination with an appropriate method for data evaluation, a broad range of ambient parameters can be simultaneously monitored and evaluated in real time.

A special task of in-field sensing is the monitoring of traces of explosives, which is needed, for example, for the detection and removal of land mines. A short overview of the development of sensors for military uses (also on the battlefield) has been given[45] and a review of the detection of clandestine explosives has been published.[46]

4.1 Electrochemical Sensors

Miniaturization is a key step in the development of chemical sensors, but this was difficult to achieve for electrochemical applications owing to a lack of suitable reference electrodes. A paper by Suzuki et al.[47] deals with the production of an integrated three-electrode system for electrochemical sensing to overcome this problem. They reported the design of a liquid-junction Ag/AgCl electrode suitable for miniaturization and showed the design of an integrated chip including working, counter and reference electrodes in thin-film technology. This system's ability as a tool for electrochemical analysis is demonstrated by the example of a glucose sensor. The device is a great leap towards a good-value dischargeable tool suitable for industrial mass production.

4.1.1 Conductivity Sensors

As already stated, the so-called Taguchi or Figaro sensor was one of the first sensors to become commercially available. This type of sensor usually consists of a semiconducting metal oxide (e.g. SnO_2) which adsorbs oxygen from the ambient atmosphere. Theory[48] indicates that electrons are transferred from negatively charged sites on the surface to these adsorbed atoms, thus leading to a decrease in surface conductivity (i.e. oxygen is reduced by conducting electrons). Reducing gases can react with the adsorbed oxygen, thus causing an increase in surface charge leading to a higher conductivity. As all the reactions occurring are reversible, devices operating with this principle meet the requirements of a chemosensor. An introduction to tin dioxide sensing was given by Takahata,[49] where not only are laboratory processes for the production of thin- and thick-film sensors described but also outlines for the solving of a basic problem in using this type of sensor are given. The fundamental drawback mentioned is the increase in the material's surface charge density with time and the problems caused by false alarms as a consequence. The effect has both reversible and irreversible causes, which can be attributed to elevated temperatures (above 500 °C) according to the analyte's heat of combustion and ambient humidity. Obviously tin dioxide from inner parts of the sensor is deposited on the outside, leading to a decrease in catalyst surface concentration. A way to overcome this difficulty is to dope the sensor material with vanadium. In addition to the introduction of the drift effects mentioned, the article also gives a short overview of other types of analytes and the resulting applications based on tin dioxide sensors.

The problems of increasing sensitivity with time (so-called 'hypersensitivity' leading to false alarms) and the working conditions of SnO_2 sensors in the field were comprehensively discussed by Nakamura.[50] Decreasing device resistance is mainly traced back to effects caused by multiple exposure to combustible gases resulting in longer periods at temperatures higher than 550 °C. As a result, Nakamura observed three effects, namely a decrease in water desorption, substantial changes in the tin dioxide crystal structure and a decrease in Pd dispersion. All these parameters lead to a denser packing of the ceramic material and variations of the surface structure, thus aggravating analyte diffusion to the reactive parts in the matrix.

Hübner and Obermeier reported basic studies on semiconducting materials suitable for carbon monoxide

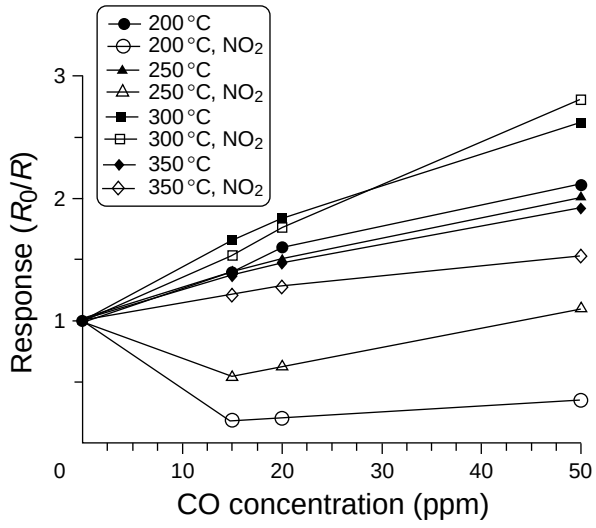

Figure 9 Avoiding cross-sensitivity of CO sensor towards nitrogen oxide by varying the measuring temperature.

sensing,[51] and later Barbi et al.[52] reported sensors for the monitoring of CO in urban environments [for carbon dioxide, see Kaneyasu et al.[53]]. Carbon monoxide is a highly toxic pollutant occurring mainly as a consequence of incomplete combustion processes. The authors reported grain-size sensors suitable for the detection of CO at concentrations down to a few parts per million. Additionally, cross-sensitivity towards NO_2 is discussed, which contrary to CO is an electron acceptor and thus shows reverse effects on surface charge. Figure 9 shows the solution to this problem by choosing an appropriate temperature. Another sensor system for the monitoring of ambient traces of CO and NO to regulate the atmosphere inside a car for the prevention of passenger casualties has been proposed by Wiegleb and Heitbaum[54] based both on indium and tin oxides. Their sensor set-up includes both ceramics on one component. Figure 10(a) and (b) presents the results obtained for the two devices. Torvela et al. studied the influence of humidity on thick-film tin dioxide gas sensors for the use of emission control in power stations.[55]

Gaseous CO_2 can also be determined by metal oxide semiconducting devices. Hanada et al. described a system consisting of an La-doped tin dioxide sensor combined with a reference and a filter for the removal of alcohol.[56] Comparison of data obtained by field measurements with infrared (IR) methods showed the high potential of the device.

A sensor array consisting of three different metal oxide sensors for the simultaneous detection of methane and ammonia in the presence of ambient humidity has been reported.[57] Data evaluation is done by the means of artificial neural networks and the relative errors for

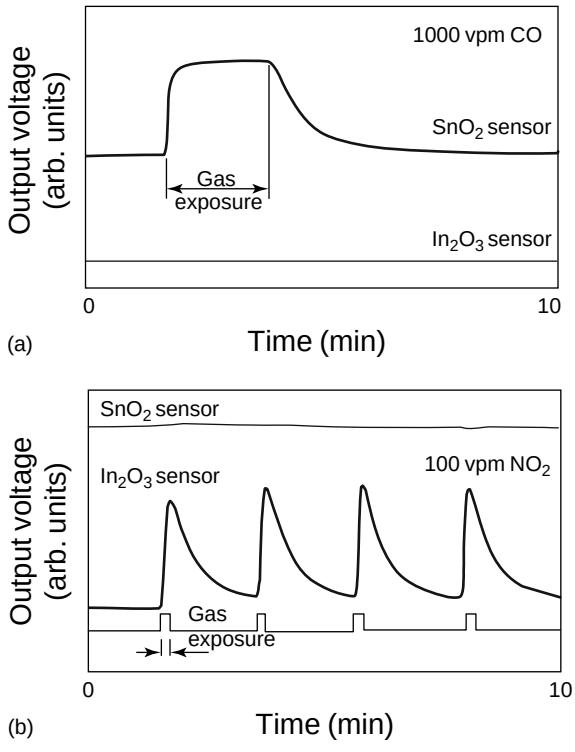

Figure 10 Dual sensor set-up for the parallel determination of (a) CO and (b) NO_2.

the quantitative determination of the analyte are below 5%. As the monitoring devices are located on the same substrate, interferences caused by material defects, etc. are compensated for. Here again it can be seen that by methods of modern data analysis applied to sensors with different sensitivity patterns all analyte concentrations of interest can be determined without a sensor reacting specifically and exclusively to one analyte.

A possibility for designing tin dioxide thin films with enhanced sensitivity based on layers fabricated by the sol–gel process, which is a very innovative method for the synthesis of ceramic material, has been introduced together with mechanistical studies of the material.[58]

Perovskite-type thick-film sensors reacting towards environmentally hazardous gases have been proposed and field tested by Martinelli et al.[59] They used printed thick films of $SmFeO_3$ and $LaFeO_3$ as sensitive layers, both being p-type semiconductors. All materials used are characterized according to structural and sensing properties. The perovskite-type materials were combined with tin dioxide devices to form an array which was tested in an Italian pollution measurement station where the signals could be compared with those obtained by classical analysis. Figure 11 shows the signals for NO obtained both by the $LaFeO_3$ sensor and the reference station during a 2 day measurement period. The excellent correlation between the two data sets can

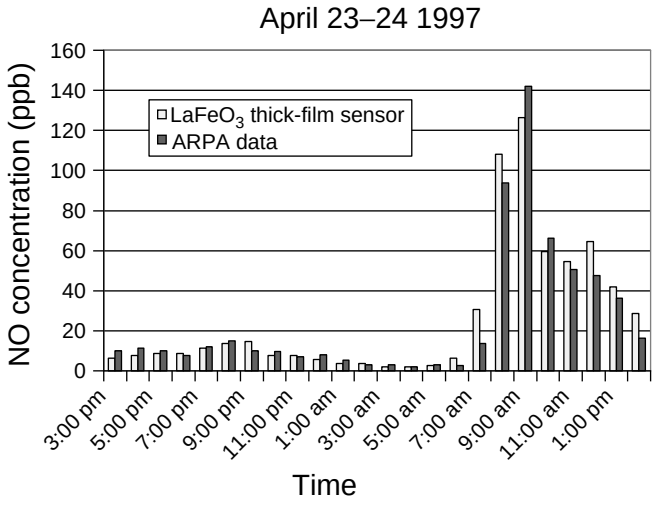

Figure 11 Comparison between NO sensor and analytical reference method during a field test.

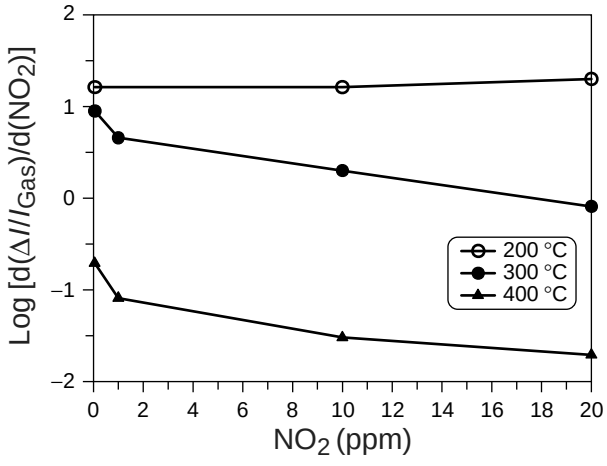

Figure 12 Sensor curves for an NO_2 sensor based on titanium–tungsten oxides at different temperatures.

clearly be seen. Nitrogen dioxide is also an environmental pollutant requiring thorough monitoring. Nelli et al.[60] reported a sub-parts per million NO_2 sensor based on titanium–tungsten oxides suitable for the detection of traces down to concentrations below 1 ppm (see Figure 12), and discussed the effects of ambient humidity.

Ozone has become a substance of interest for environmental analysis owing to its high toxicity and the fact that its concentration in atmospheric layers near ground level has been increased mainly due to NO_x emissions. Takada et al. introduced a sensor system for O_3 based on thin-film In_2O_3 semiconductors.[61] Thin-film technology has to be used owing to the low stability of ozone: in a thick-film sensor it would decompose before reaching the core of the sensing material. The influence of

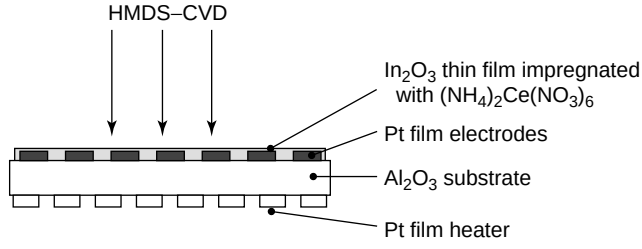

Figure 13 Set-up of O_3 sensor published by Takada et al.[61]

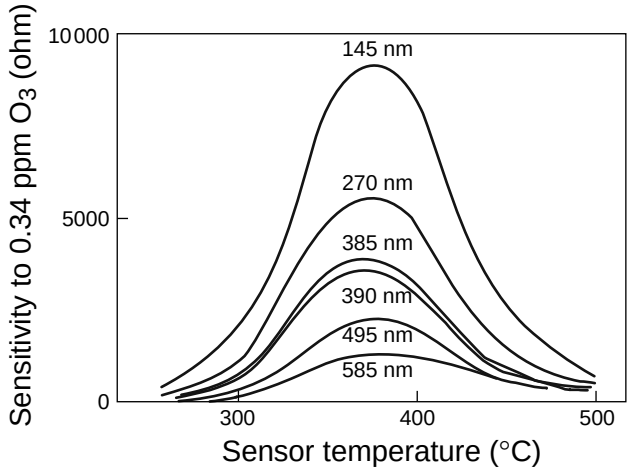

Figure 14 Response of O_3 sensor by Takada et al.[61] depending on layer height.

reducing gases on the sensor signal can be eliminated by coating the entire system with a thin layer of SiO_2. Figure 13 shows a sensor set-up without this 'protective' layer, and in Figure 14 the sensitivity of sensors with several film thicknesses is plotted versus temperature. The results clearly indicate an ideal operating range for the device and that increasing height of the sensitive layers strongly decreases the sensor response. The sensors obtained underwent a long-term test and field testing with the results shown in Figure 15. WO_3 has also proven to be a suitable material for ozone sensing, permitting analysis down to the parts per billion level.[62] Figure 16 shows a comparison between the solid-state sensor used and a classical UV analyzer during a field measurement. The very good agreement between the two methods can easily be seen and proves the sensor's suitability for this analytical task.

Another analyte of environmental interest is ammonia in air. Although NH_3 potentially reacts with sulfur and nitrogen oxides, an excess amount in air is a potential hazard. One of the main sources of atmospheric ammonia is livestock buildings. A sensor based on Cu-doped SnO_2 thick films for the monitoring at such sites has been reported.[63] Figure 17 shows the response

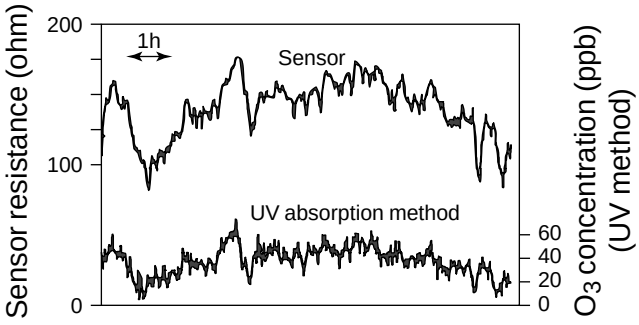

Figure 15 Results of O_3 sensor field measurements.

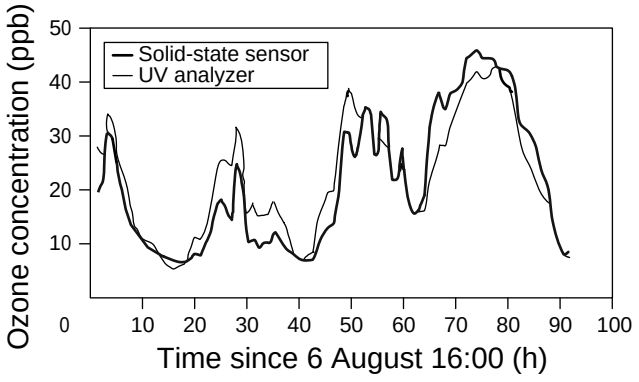

Figure 16 Validation of ozone sensor based on WO_3 with UV.

of this sensor towards ammonia in humid air, where the humidities correspond to values typically found in livestock buildings. The authors also reported field testing of the device which has proven to be useful for a prolonged period of time. Another promising material under development for this purpose is molybdenum oxide.[64]

A further application of tin dioxide sensors, where commercial testing devices are reported to be available, is in odor sensing,[65] where the sensors were optimized for the detection of organic sulfides. A sensor system for the monitoring of H_2S in urban environments based on Ag-doped tin dioxide devices has been published and tested by Lantto and Mizsei.[66]

An approach for the detection of organic contaminants in water has been reported by Straková et al.[67] For the determination, a standard SnO_2 sensor was used, the analytes being extracted via a silicone-rubber membrane. The authors reported the determination of several oxygenated, chlorinated and aromatic compounds. Whereas all sensors introduced up to then directly measured the analyte, in this case it is separated from the matrix by the means of the gas-permeable but watertight membrane. Thus sensor kinetics are influenced by two steps, namely separation of the analyte and matrix and reaction of the analyte with the sensitive element.

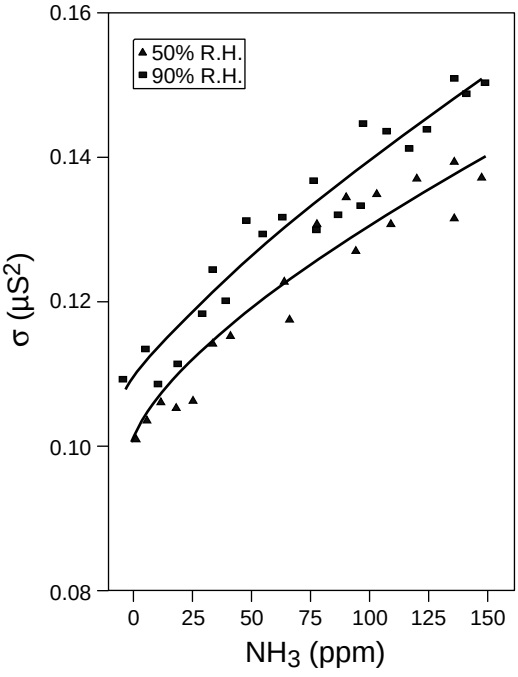

Figure 17 Response of ammonia sensor in highly humid environments.

4.1.2 Potentiometric Sensors

Some potentiometric sensors for different analytes such as oxygen, chlorine, SO_2 and SO_3 were introduced by Kleitz et al.[68] The best known potentiometric sensors are the lambda sensor and ion-selective electrodes, as already stated above. To complement the lambda probe, Meixner et al. made suggestions:[69] they proposed nitrogen oxides and carbon monoxide as additional analytes because they doubt whether oxygen alone is a suitable analyte for the determination of the optimal ratio between fuel and air for combustion. They used $SrTiO_3$ for the lambda probe, Al_2O_3/V_2O_5 for NO_x and Ga_2O_3 for CO and reported good laboratory results. The background for this proposition is the fact that a deviation of the ideal lambda value by as little as 0.001–0.002% decreases the catalyst efficiency from 98% to 85%. The adjustment is especially difficult during starting, braking and accelerating a car.

To complement the Sr titanate devices, Schreiner and Härdtl introduced ceramics based on the formula $SrTi_{1-x}Fe_xO_3$, being highly suitable for oxygen sensing applications as the signal is comparably insensitive to temperature.[70] Generally this is a crucial aspect as engine temperature is highly dependent on the time, style and environment in which it is operated.

A concise overview of the research on sensors for SO_2 and SO_3 was given by Worrell[71] and Adachi and Imanaka.[72] Worrell introduced K_2SO_4, Na_2SO_4 and

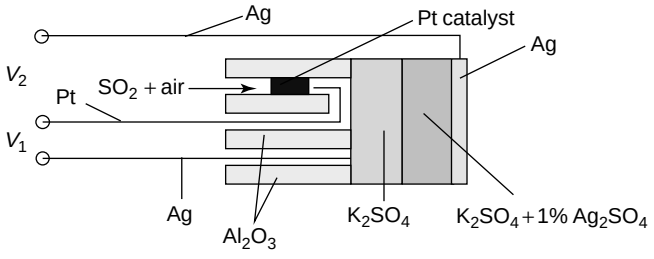

Figure 18 Potentiometric sensor for SO_2 introduced by Worrell.[71]

Li_2SO_4 as potential solid electrolyte materials. Figure 18 shows such a cell with an included reference electrode suitable for measuring different amounts of sulfur dioxide. Another class of promising sensors for this purpose is thin-film sulfate electrolyte sensors based, e.g. on Nasicon (a polymeric Na-selective material) membranes (see Figure 19). A hand-held system for the determination of different inorganic ions via interchangeable sensor tips has been reported by Ross et al.[73] The device has proven its reliability in the analysis of drinking water, whereas other applications require further investigation. The ion-selective planar sensor strip can easily be changed, so a variety of analytes can be determined consecutively. The same group has additionally introduced a miniaturized flow-through system for the on-line detection of nitrate ions in water samples.[74] Figure 20 shows a block diagram of the device; the authors reported an operational range between 10^{-5} and 10^{-2} mol L^{-1}.

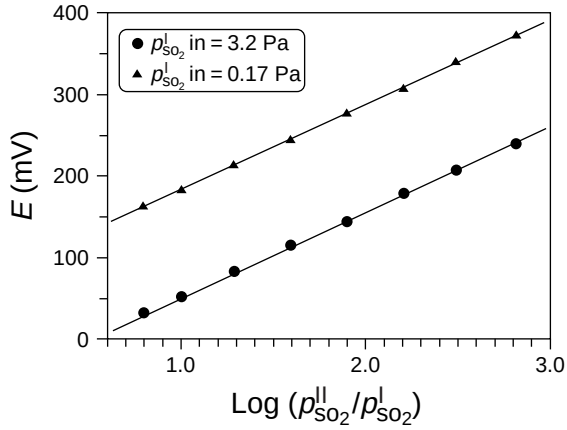

Figure 19 Sensor response to SO_2 for a potentiometric SO_2 sensor.

Another approach for ion detection in wastewaters is the use of chalcogenide glasses as chemical sensors.[75] Among the ions detected are copper, iron, mercury and cadmium, which are all interesting for the analysis of, e.g. industrial wastewaters. Figure 21 shows the response of the copper-selective sensor during a field measurement in industrial wastewater control. In another paper,[76] a thallium-selective electrode was described and in Figure 22 one can see the sensor curve obtained for natural waters doped with different amounts of thallium.

Also dealing with water quality determination, Sakai et al. introduced a multichannel potentiometric sensor

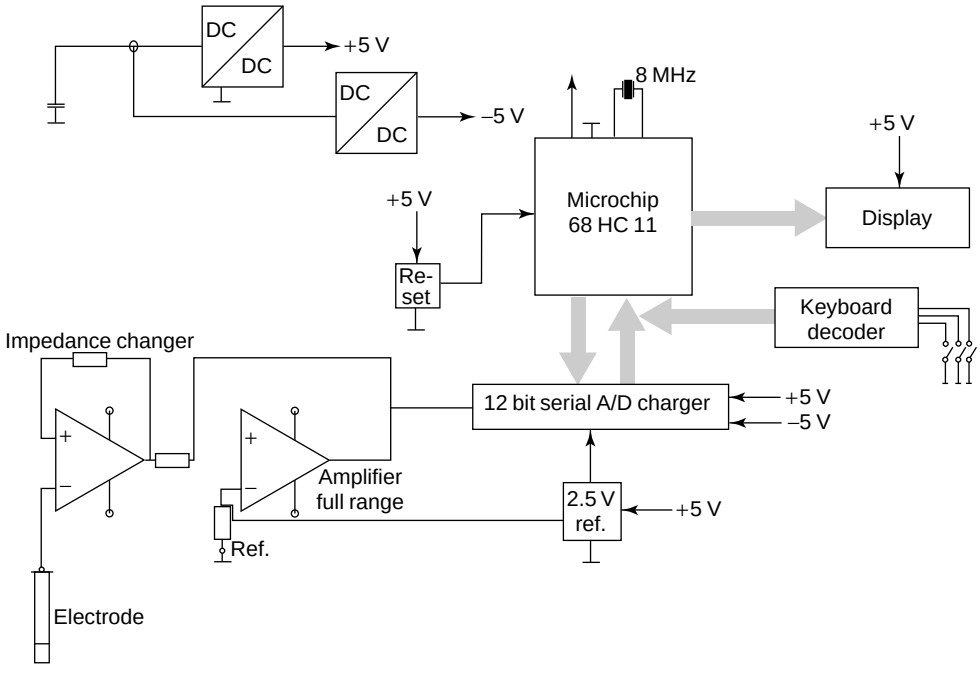

Figure 20 Block diagram of a sensor system for ionic contaminants in water with changeable electrodes suggested by Ross et al.[73]

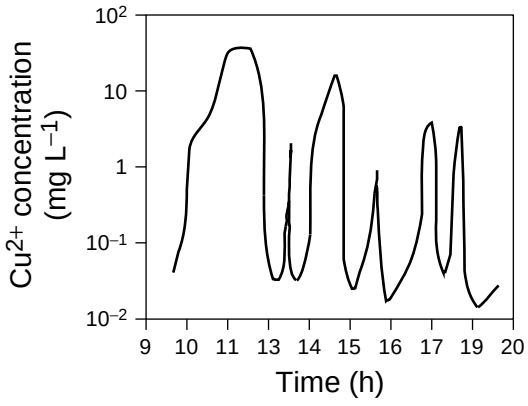

Figure 21 Field-measurement data for a copper-selective potentiometric sensor.

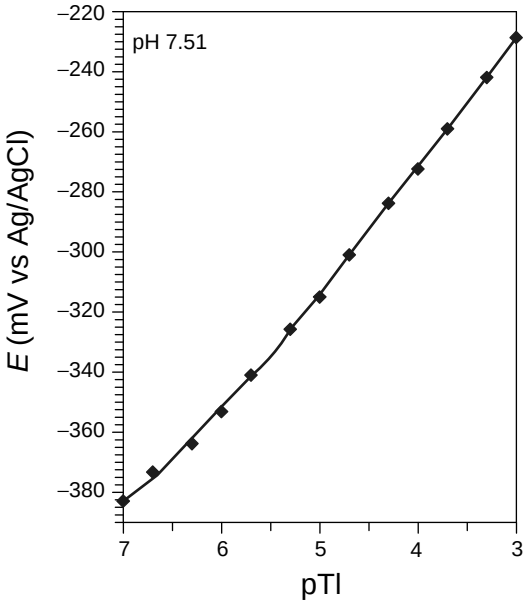

Figure 22 Calibration curve for a thallium-selective potentiometric sensor obtained in natural water doped with thallium.

for the determination of river water quality.[77] They used a set-up with eight different sensitive lipid layers and the results obtained in eight-dimensional space were analyzed and visualized via principal component analysis. This method usually yields parameters whose physical meaning is unknown but they are nevertheless suitable for qualitative predictions and quantitative rankings.

4.1.3 Amperometric Sensors

Amperometry uses the electric current associated with an electrochemical conversion when a defined voltage is applied. This measurand can be absolutely determined, which means that no reference system has to be defined and used, and depends linearly on analyte concentration, which eases data evaluation. Parellada et al. gave an overview of amperometric field sensing devices based on biorecognition principles.[78] An amperometric sensor for the detection of nitrogen oxides has been suggested by Tierney et al.[79] They use a Back Cell™ design where the analyte does not diffuse to the electrodes through the polymeric electrolyte but from the porous ceramic back side of the device. Diffusion through these pores is substantially faster than through a polymeric matrix, making the response much more rapid than for classical devices. The measurements take place at the so-called triple points which represent the boundaries between the electrolyte, the electrode and the gas present in the pores. Discrimination between NO and NO_2 is achieved electrochemically: both are oxidized on gold electrodes at similar potentials, but nitrogen dioxide also shows reductive behavior in the range of potential used. Figure 23(a) and (b) depicts the calibration curves for both of the analytes in air, proving the potential of the sensor.

Amperometry offers a promising approach in the determination of H_2S in aquatic resources.[80] This highly toxic substance is abundant in all systems, where sulfidic

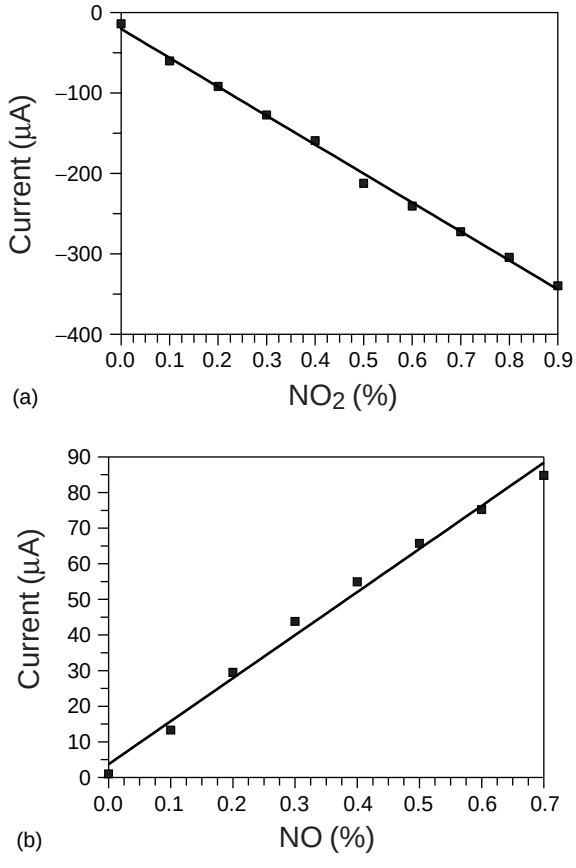

Figure 23 Sensor curves for (a) NO_2 and (b) NO for an amperometric NO_x-sensor developed by Tierney et al.[79]

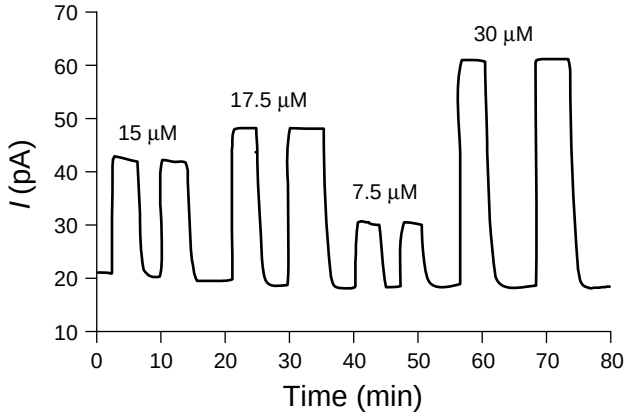

Figure 24 Amperometric sensor response towards H_2S in water.

sources (be they natural or artificial) are present. The system resembles the Clark electrode, making use of hexacyanoferrate as oxidizing agent for the sulfide. All electrodes are located at the tip (diameter 30–50 µm) of a tube. Figure 24 shows the results for this sensor obtained during a flow-through experiment.

A summary of the development of biosensors for environmental applications was given by Renneberg and Gründig.[81] They introduced a wide variety of potential analytes and sensors developed for their detection, but also stated that there are only a few systems for market-ready field applications owing to stability problems with the enzymes used. A paper dealing especially with the detection of organophosphorus pesticides was presented by Gibson et al.[82]

4.1.4 Other Electrochemical Sensors

An overview of electrochemical sensors with solid polymer electrolytes which are of potential interest for environmental applications was given by Opekar and Štulik.[83]

A promising approach for carbon dioxide sensing based on a coulometric device was proposed by Trapp et al.[84] Titration of CO_2 is done by cathodically produced hydroxide ions, whose excess after reaching the end-point is detected via a pH-sensitive electrode (which in fact means potentiometric sensing). The authors reported the possibility for detecting 200–20 000 ppm of carbon dioxide, which complies with the requirements in ambient sensing of this analyte. The sensor's main drawback is its limited operational time.

A sensor system for the field measurement of redox potentials has been introduced by Teasdale et al.[85] After discussing problems of electrode stability and calibration, they introduced a set-up with 15 platinum wire electrodes for the spatially resolved determination of estuary redox potential, be it in the water or at the water/sediment

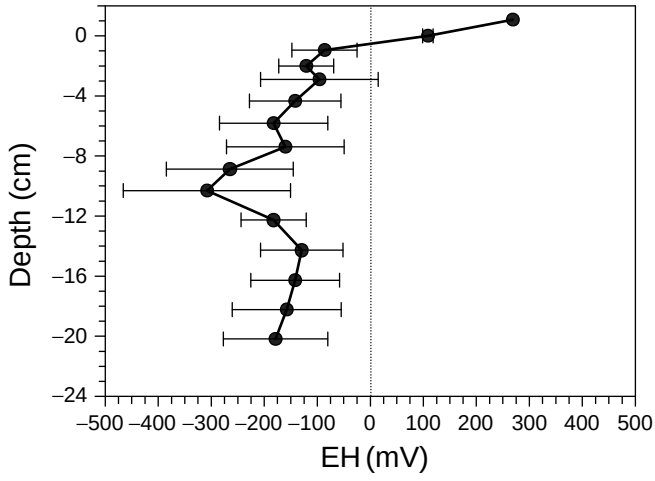

Figure 25 Sediment depth profile of the electrochemical potential.

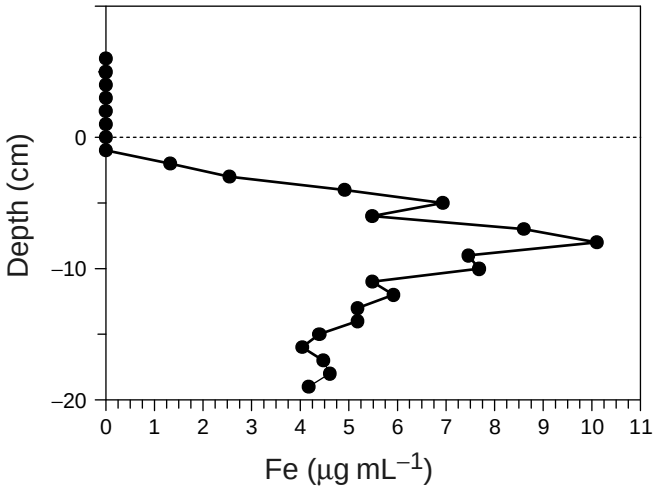

Figure 26 Sediment depth profile of iron concentration.

boundary layer. This measurand can be used as an indicator for pollution of the ecosystem being monitored. Additionally, concentration profiles over the boundary layer of analytes present both in water and in sediment are a topic of major scientific interest. Figures 25 and 26 show both the average redox potential profile and the iron concentration of a sediment site obtained with this apparatus.

4.2 Field Effect Devices

Lundström et al.[21] gave an introduction to sensing via field effect devices and described a wide variety of (laboratory) applications. Some examples of ion-selective sensors (mostly pH) are given there and will not be discussed further in this article. Additionally, some gas-sensing applications are introduced, including analytes such as hydrogen, ethanol, hydrogen sulfide, ammonia,

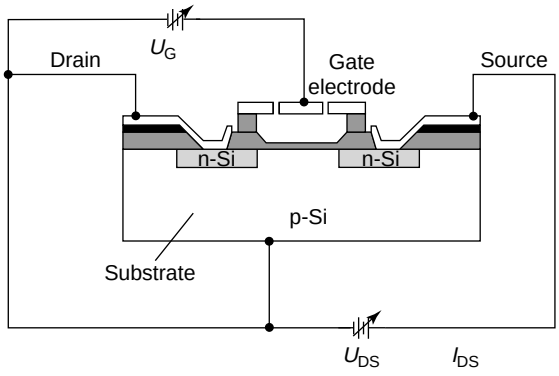

Figure 27 Schematic set-up of an SGFET.

some hydrocarbon gases and carbon monoxide. Many of them are more within the range of laboratory methods than field testing. One exception is again the sensing of ozone:[86] an SGFET coated with potassium iodide was used for its monitoring. The theory of gas FETs indicates that the electronic work function at an interface between the semiconductor and a layer different from it can be influenced by gas adsorption. The SGFET devices (Figure 27) make use of the interface between the lower side of the electrode and the air gap that can be found there. The response of this sensor during a field measurement in the Bavarian Alps compared with an industrial wet chemical ozone analyzer can be seen in Figure 28. The values obtained show excellent correlation between the data obtained by the sensor and the reference system, which again proves the suitability of chemosensors for environmental applications.

A pH-sensitive ISFET for the in-field detection of acid rain was introduced by Wakida et al.[87] They found that tris(2-ethylhexyl)amine is an ideal gate material for the desired purpose, as is confirmed by the data depicted in Figure 29, where sensor results for field samples are compared with those obtained with a glass electrode.

Starodub et al. reported a sensor system based on electrolyte–insulator–semiconductor devices combined with multienzyme layers (containing, e.g. glucose oxidase, urease, cholinesterase) for the detection of toxic compounds,[88] namely heavy metal ions and organophosphorus pesticides. Quantitative analysis is performed by the measurement of the residual enzyme activities after exposure to the analyte.

4.3 Optical Sensors

As already stated above, two types of optical sensing applications exist, namely those making use of chemically sensitive layers and the others relying directly on the optical properties of the analytes. Articles about the development of fiber-optic chemical sensors can be found elsewhere.[89,90]

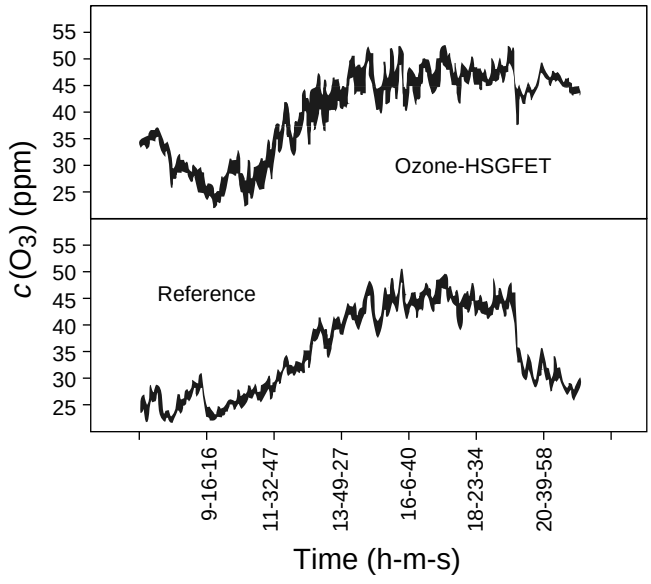

Figure 28 Sensor response of SGFET-based ozone sensor compared with a reference system during field measurements in the Bavarian Alps.

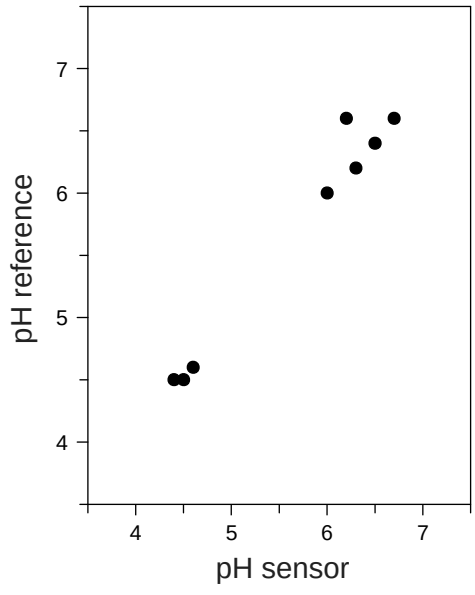

Figure 29 Comparison of a pH-sensitive ISFET coated with tris(2-ethylhexyl)amine with a glass electrode on natural rain samples.

A field application for the detection of polycyclic aromatic hydrocarbon (PAH) compounds has been suggested.[91] Six different analytes were investigated with a time-resolved fiber-optic fluorescence system with an optical pathlength of up to 50 m (for experimental apparatus, see Figure 30a and b). The qualitative information is obtained via both spectral and signal decay

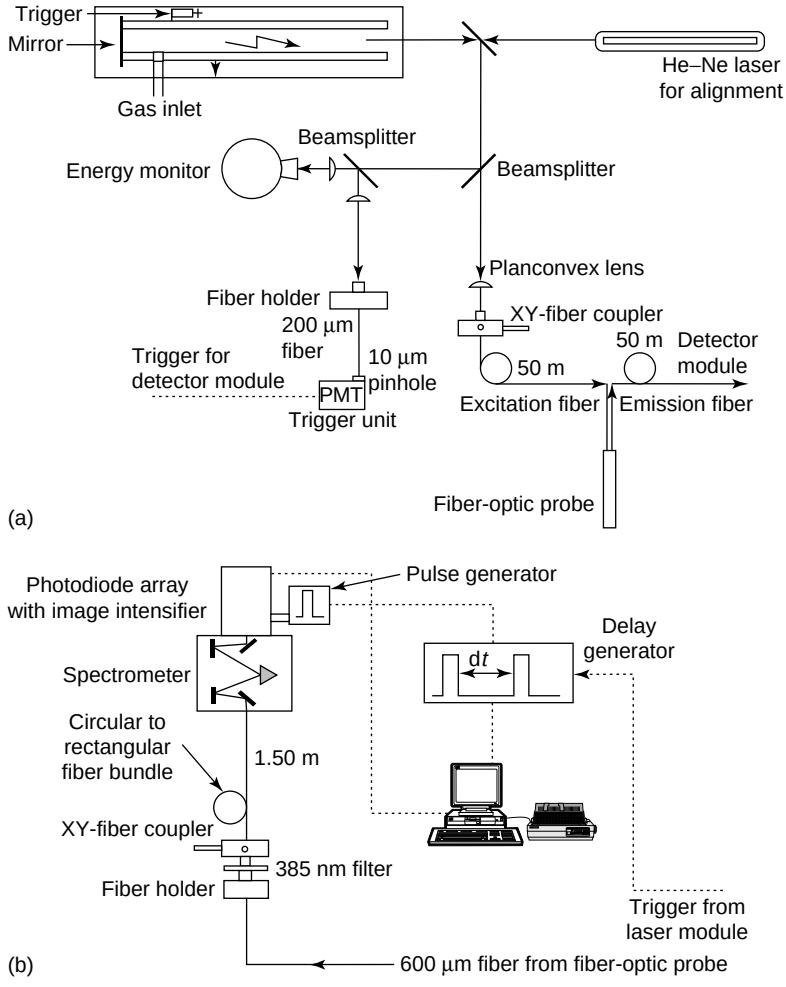

Figure 30 Fiber-optic sensor system for remote sensing of groundwater contaminants.

information (due to interaction with oxygen). The authors also examined the influence of humic acids, stating that despite their huge fluorescence signal at least a semiquantitative determination of the PAHs is easily possible. They can be detected down to the range of nanograms per liter over a fiber length of 50 m. Another system making use of fluorescence effects has been published by Frense et al.[92] where changes in the fluorescence of algal chlorophyll is measured. As the cells react towards several herbicides and pesticides they can (after immobilization) be used for the detection of these analytes.

Fluorescence systems are widely used as a result of the method's inherent advantages, which clearly outweigh the fact that only a limited number of analytes are accessible by this method: fluorescence spectrometry is usually a highly sensitive method allowing very low detection limits and the effects are very fast, can be measured accurately and often allow a discrimination between the analyte signal and the background, i.e. additional matrix substances.

A system consisting of three fibers for the detection of hydrocarbon vapors for pipeline leakage detection and tank, storage and transport monitoring has been introduced[93] (see Figure 31). High-intensity Hg lamps and LEDs are used as optical sources. The nylon jacket and sometimes also the cladding of the sensing fiber are removed, thus exposing either directly the silica core or the silicone cladding directly exposed to the hydrocarbon analyte. The sensor works on quenching principles, therefore the light intensity in the reference fiber is much higher than that in the sensing fiber. Field data for the sensor response towards gasoline are given in Figure 32.

Larjava et al. introduced a Fourier transform infrared (FTIR) gas analyzer for the simultaneous determination of different industrial boiler exhaust gases.[94] They reported the construction of a field testing device with integrated data evaluation for this purpose. Another system for the analysis of industrial exhaust gases is a fiber-optic laser spectrometer operating in the IR range

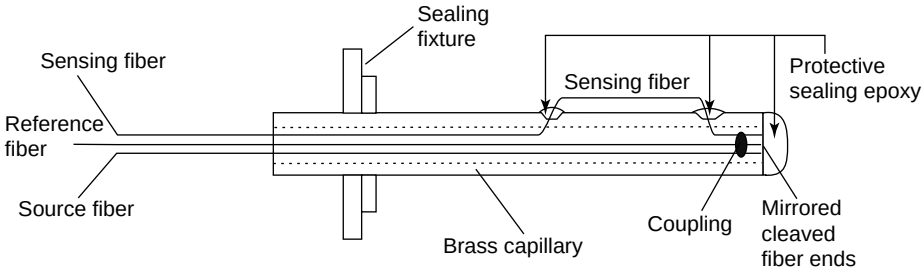

Figure 31 Fiber-optic probe for the detection of hydrocarbons in the environment (leak detection in pipelines, etc.).

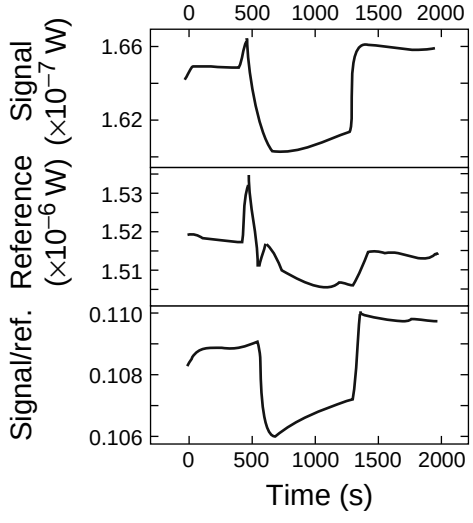

Figure 32 Response of hydrocarbon sensor.

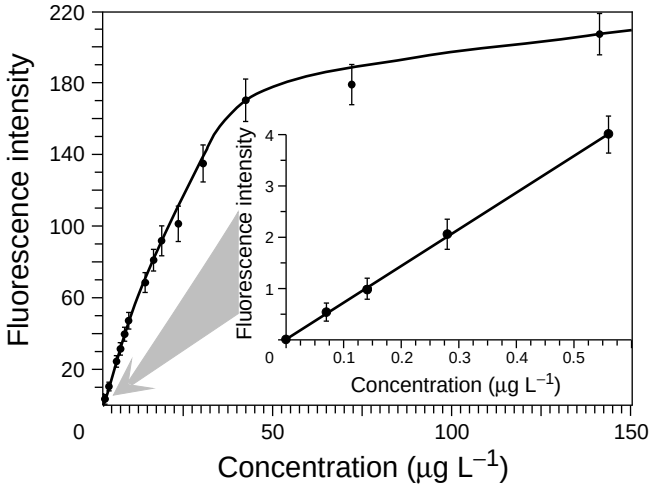

Figure 33 Fluorescent signal of PAH imprinted sensor layer.

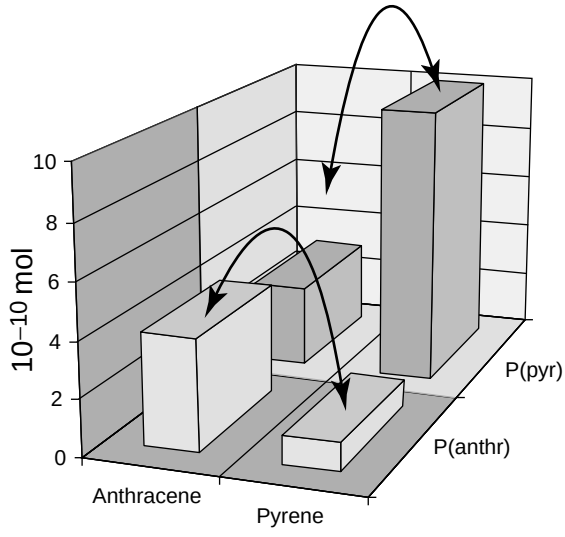

Figure 34 Selectivity pattern of polymers imprinted with different PAHs. P(pyr), pyrene-imprinted polymer layer; P(anthr), anthracene-imprinted polymer layer.

for the detection of ammonia vapors being emitted from reactors to minimize NO_x in exhaust gases.[95] This system makes it possible to detect ammonia selectively even under conditions where many interfering gases such as water vapor are present. Another successful sensor based on NIR laser diodes for the determination of different ambient gases including humidity has been reported.[96]

A hand-held fiber-optic system for the detection of organic solvents in water was introduced by Dickert et al.[97] Here a diphenyl phthalide dye was embedded behind a gas-permeable membrane, which extracts the organic solvents from the water. The dye shows hypochromic effects and therefore the change in absorbance can be used to quantify analyte concentration. The whole system was included in a market-ready housing.

PAHs are a class of highly toxic substances occurring at all kinds of incomplete combustion processes. A possibility of their direct detection via fluorescence spectrometry by using molecularly imprinted polymers has been described.[98] The sensors produced provide both good selectivity between different PAHs and low detection limits, as can be seen in Figures 33 and 34. Concentration limits for the fluorimetric measurements are down to 1 µg L^{-1} and even lower. Another system for the monitoring of anthropogenic compounds in groundwater

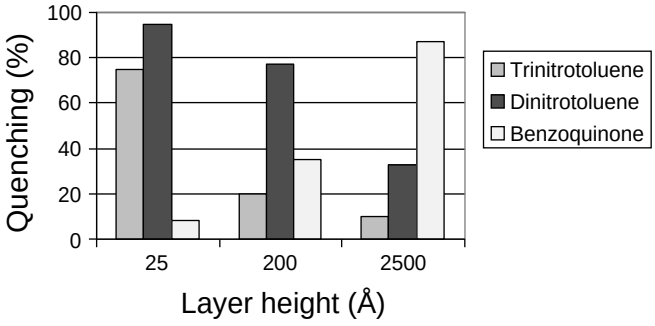

Figure 35 Response of sensors coated with different polymer layer heights.

under development was proposed by Leonard,[99] who used fibers coated with a fluorescence dye.

Yang and Swager described a sensor for the detection of trinitrotoluene (TNT) vapor based on fluorescence quenching.[100] As sensitive layers, fluorescent, shape-persistent porous pentiptycene polymers are used. Figure 35 shows the fluorescence quenching of the different analytes of interest for different film thicknesses, thus proving the analytical capabilities of this system.

A possibility for the detection of chlorine in air, based on a one-way sensor strip, was described by Ralfs and Heinze.[101] The authors suggested *o*-toluidine embedded in a silicone–polycarbonate copolymer matrix with external reading in the visible and NIR region (350–1100 nm). The absorbance of the *o*-toluidine increases when ambient chlorine is present, and as a result analyte concentrations down to 0.05 ppm are detectable with this system. The low price makes it possible to use these sensors as discardable probes (the cost per piece is around US$ 0.1), thus highlighting the system's suitability for fast analysis.

Another system based on NIR measurements was published by Zimmermann et al.[102] Here optical fibers are coated with siloxane membranes suitable for the incorporation of nonpolar organic substances. These are extracted from aqueous media. Detection takes place by the absorption of evanescent NIR radiation. A great advantage is the fact that a long fiber can be used as a sensitive medium, therefore increasing sensitivity. A novel optical sensing system can be realized by surface-enhanced Raman scattering (SERS) taking place on colloidal gold or silver particles. An overview of this innovative technique, including some aspects of environmental analysis, was given by Vo-Dinh.[103]

Lanthanide-enhanced fluorescence spectroscopy,[104] where ligands influence a rare earth ion's fluorescence behavior, has only recently been established as a potential sensing method. Fluorescence of these ions (often Eu^{3+}) is a comparably slow process (from 100 ns up to microseconds), so it can be separated from background signals fairly easily. This effect was utilized for the detection of nerve agents in water, where an optical sensor was constructed by molecularly imprinting a polymer with an Eu^{3+} complex containing a hydrolysis product of soman.[105] The authors reported high sensitivity and selectivity, making field measurements possible. Sensors of this kind are not only of military interest but also show promise for pesticide sensing, as some organophosphorus nerve agents are closely related to agrochemical substances.

4.4 Mass-sensitive Sensors

Mass-sensitive devices have already been applied in both the gaseous and liquid phase. Additionally, both single sensors and arrays have been reported [for some fundamental aspects of the layer design for sorption sensor arrays, see Lau et al.[106]]. Dickert et al.[107] introduced QMB and SAW coatings based on mixtures of linear and globular molecules for the detection of diesel vapors. A sensor array based on six QMB devices with a resonance frequency of 10 MHz for the detection of organic vapors was described by Hierlemann et al.[108] The sensors were coated with differently substituted polysiloxanes, widely applied as stationary phases in GC. The authors used 1-propanol, 2-propanol, tri- and tetrachloromethane, trichloroethylene and toluene and their mixtures as analytes. Different methods of data analysis were applied and compared with each other.

Another paper dealing with the separation of environmentally sensitive analytes was published by Dickert et al.[109] In this case the detection of the chemically very

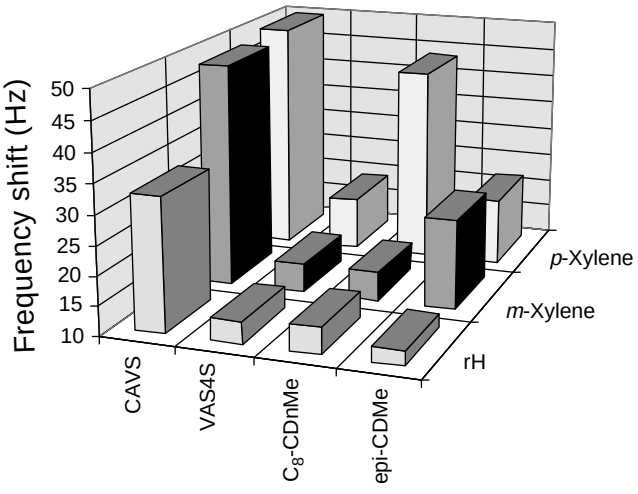

Figure 36 Selectivity pattern of mass-sensitive sensor array towards *p*-xylene and *m*-xylene. CAVS, VAS4S, two types of calix-resorcin-arenes; C_8-CDMe, permethylated dioodo-octane-crosslinked beta-cyclodextrin; Epi-CDMe, permethylated epichlorhydrine-crosslinked beta-cyclodextrin.

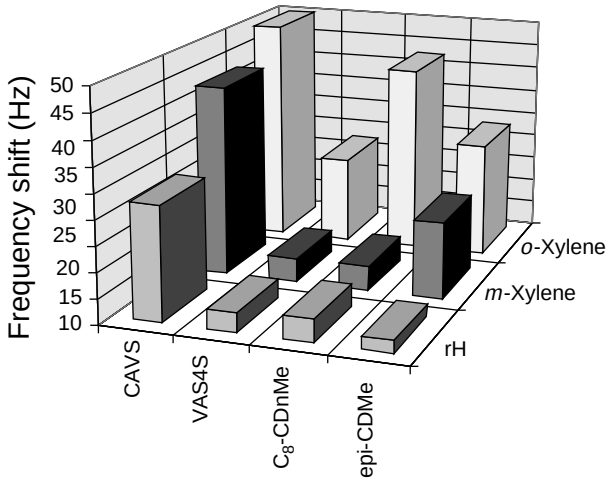

Figure 37 Selectivity pattern of mass-sensitive sensor array towards *p*-xylene and *m*-xylene. CAVS, VAS4S, two types of calix-resorcin-arenes; C_8-CDMe, permethylated dioodo-octane-crosslinked beta-cyclodextrin; Epi-CDMe, permethylated epichlorhydrine-crosslinked beta-cyclodextrin.

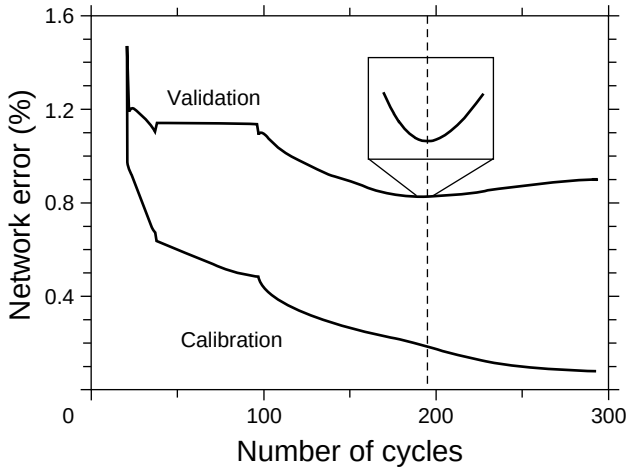

Figure 38 Calibration and validation curve for the neural network used for xylene detection.

similar isomers of xylene under different amounts of ambient humidity was the topic of interest. An array of four QMBs (principal frequency 10 MHz) was used, and the sensitive layer materials were cavitands instead of GC phases. Figures 36 and 37 present the different sensors' responses to the analytes and Figure 38 shows the calibration and validation curves for the neural network used. The great advantage of this sensor system is its ability to discriminate between chemically very similar compounds (i.e. the three different isomers of xylene) and ambient humidity.

An important class of analytes for in-field sensing is chemical warfare agents and several sensor systems for their detection are under development. An SAW array containing 14 devices coated with different polymers such as fluoropolyol, poly(ethylenimine), ethyl cellulose and poly(epichlorohydrin) has been reported by Rose-Pehrsson et al.[110] The entire device consists not only of the sensor array and the system of multicomponent data evaluation but also of preconcentration tubes.

The optimization of a sensor array for the analysis of organic compounds in water has been published.[111] The main importance of this system can be found in the monitoring of contaminants caused by shipwrecking of tankers or other vessels. The authors introduced a wide variety of both sensor coatings and potential analytes using QMB thickness shear wave resonators.

5 CONCLUSION

Chemical sensors show a variety of inherent advantages giving them great potential for use in environmental analytical chemistry. These positive features include their ruggedness, the fact that the devices and systems are usually not too complicated (but nonetheless highly developed) analytical instruments, which can also be operated by non-professionally trained personnel, and their price. Some sensor systems have already proven to be powerful analytical tools in everyday use, e.g. the Figaro sensor and the lambda probe. The wide variety of methods (electronic, optic, mass-sensitive, etc.) and chemical interactions (host–guest, steric recognition, functional recognition, etc.) used open up an enormous range of potential measuring devices. Especially gas-sensing field applications have already been reported, but current research also deals extensively with sensor analyses in the liquid phase.

6 COMMERCIAL SENSORS

Although a wide variety of sensors (be they physical or chemical) for all kinds of measurands is commercially available, there are not many examples of market-ready chemosensors apart from the lambda probe and the Figaro sensor. An overview over commercially available multisensor systems (electronic noses) was given by Strike et al.,[9] including the composition of the respective sensor arrays.

A variety of sensors for atmospheric pollutants are based on amperometric detection.[112] A typical cell is shown in Figure 39. These devices are offered for the detection of several oxides (sulfur, nitrogen, carbon)

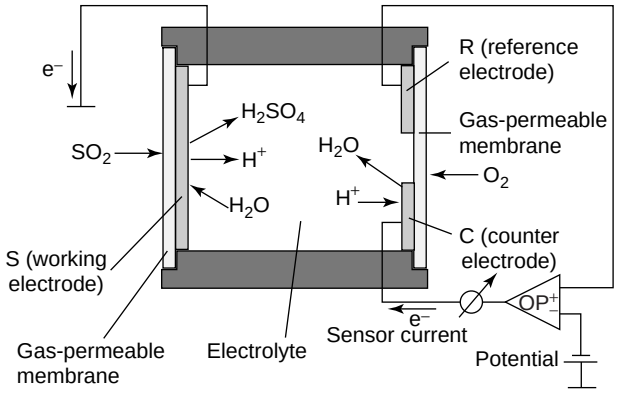

Figure 39 SO_2 sensor as an example of commercially available amperometric gas sensors.

and oxygen. Another company offering a wide variety of amperometric gas sensors is Sensoric,[113] selling devices suitable for a range of analytes from atmospheric pollutants to compounds such as gaseous hydrides and phosgene.

A chemical nose containing six 10 MHz QMB resonators is sold by HKR Sensorsysteme.[114] This array is built for the discrimination of complex species, the resonators being covered with six different GC layers. Potential analytes are all kinds of mixtures, e.g. present in food analysis. An electronic nose combining both tin dioxide and mass-sensitive devices is commercially available.[115]

ABBREVIATIONS AND ACRONYMS

ChemFET	Chemically Sensitive Field Effect Transistor
FET	Field Effect Transistor
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
IDT	Interdigital Transducer
IR	Infrared
ISFET	Ion-sensitive Field Effect Transistor
LED	Light-emitting Diode
MIR	Middle-infrared
MIS	Metal Insulator Semiconductor
MOS	Metal Oxide Semiconductor
MS	Mass Spectrometry
NIR	Near-infrared
PAH	Polycyclic Aromatic Hydrocarbon
QMB	Quartz Microbalance
REFET	Reference Field Effect Transistor
SAW	Surface Acoustic Wave
SERS	Surface-enhanced Raman Scattering
SGFET	Suspended-gate Field Effect Transistor
SPR	Surface Plasmon Resonance
TNT	Trinitrotoluene
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Luminescence in Environmental Analysis • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Electrochemical Sensors for Field Measurements of Gases and Vapors • Field-based Analysis of Organic Vapors in Air

Field-portable Instrumentation cont'd **(Volume 5)**
Microelectromechanical Systems Technology Applied to the Miniaturization of Field Instrumentation

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Chemometrics **(Volume 11)**
Chemometrics

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals • Microbalance, Electrochemical Quartz Crystal

REFERENCES

1. V. Lopez-Avila, H.H. Hill, 'Field Analytical Chemistry', *Anal. Chem.*, **69**, 289R–305R (1997).
2. U. Lehmann, H. Plehn, O. Krusemark, J. Müller, 'A Miniaturized Gas Chromatograph for Autonomous and Longtime Measurements', in *Proc. Sensor99, Nürnberg*, AMA Service, Wunstorf, 155–158, Vol. 1, 1999.
3. S. Zimmermann, B. Riepenhusen, J. Müller, 'Micromachined Flame Analyzers: Flame Ionization Detector (FID); Atomic Emission Flame Spectrometer (AES)', in *Proc. Sensor 99, Nürnberg*, AMA Service, Wunstdorf, 409–413, Vol. 1, 1999.
4. M. Otto, in *Analytical Chemistry: The Authentic Text to the FECS Curriculum Analytical Chemistry*, eds. R. Kellner, J.-M. Mermet, M. Otto, H.M. Widmer, Wiley-VCH, Weinheim, 359–373, 1998.
5. (a) F. Dickert, O. Schuster, 'Piezoelektrische Chemosensoren – von der Adsorption zur molekularen Erkennung mit Wirt-Gast-Chemie', *Chem. Unserer Zeit*, **28**, 147–152 (1994); (b) F. Dickert, 'Chemosensoren für Gase und Lösungsmitteldämpfe', *Chem. Unserer Zeit*, **26**, 138–143 (1992).
6. (a) F.L. Dickert, A. Haunschild, 'Sensor Materials for Solvent Vapor Detection – Donor–Acceptor and Host–Guest Interactions', *Adv. Mater.*, **5**, 887–895 (1993); (b) F.L. Dickert, U.P.A. Bäumler, G.K. Zwissler, 'Supramolecular Structures and Chemical Sensing', *Synth. Metals*, **61**, 47–52, 1993.
7. R. Müller, 'Multisensor Signal Processing', in *Sensors: A Comprehensive Survey*, eds. W. Göpel, J. Hesse, N. Zemel, VCH, Weinheim, 313–330, Vol. 1, 1989.
8. (a) U. Weimar, 'Elektronische Nasen: Gestern, Heute, Morgen', *Sensoren und Messtechnik: Vorträge der ITG Fachtagung vom 9. bis 11. März 1998 in Bad Nauheim*, VDE-Verlag, Berlin, 207–227, 1998; (b) T. Doll, A. Köster, H. Gerlach, F. Derbel, H.R. Tränkler, I. Eisele, 'Konfigurations-, Auswerte- und ''intelligente'' Selbstüberwachungskomponenten des modularen Gas-Multisensorsystems IGM', in *Sensoren und Messtechnik: Vorträge der ITG Fachtagung vom 9. bis 11. März 1998 in Bad Nauheim*, VDE-Verlag, Berlin, 231–247, 1998.
9. D.J. Strike, M.G.H. Meijerink, M. Koudelka-Hep, 'Electronic Noses – A Mini-review', *Fresenius' J. Anal. Chem.*, **364**, 499–505 (1999).
10. B.K. Lavine, 'Chemometrics', *Anal. Chem.*, **70**, 209R–228R (1998).
11. R.E. Shaffer, S.L. Rose-Pehrsson, R.A. McGill, 'A Comparison Study of Chemical Sensor Array Pattern Recognition Algorithms', *Anal. Chim. Acta*, **384**, 305–317 (1999).
12. F.L. Dickert, A. Haunschild, P. Hofmann, 'Cholesteric Liquid Crystals for Solvent Vapor Detection – Elimination of Cross Sensitivity by Band Shape Analysis and Pattern Recognition', *Fresenius' J. Anal. Chem.*, **350**, 577–581 (1994).
13. 'General Technical Information', Figaro USA, Glenview, IL; also available at http://www.figarosensor.com/products/common.pdf.
14. E.M. Logothetis, 'Automotive Oxygen Sensors', in *Chemical Sensor Technology*, ed. N. Yamazoe, Kodansha, Tokyo and Elsevier, Amsterdam, 89–104, Vol. 3, 1991.
15. (a) H. Dueker, K.H. Friese, W.D. Friese, 'Ceramic Aspects of the Bosch Lambda Sensor', SAE Paper 75 0223, SAE Automotive Engineering Congress, Detroit, February 1975, 807–824, 1975; (b) G.V. Velasco, J.P. Schnell, 'Gas Sensors and Their Applications in the Automotive Industry', *J. Phys. E (Sci. Instrum.)*, **16**, 973–977 (1983).
16. A. Chiba, 'Development of the TGS Gas Sensor', in *Chemical Sensor Technology*, ed. S. Yamauchi, Kodansha, Tokyo and Elsevier, Amsterdam, 1–18, Vol. 4, 1992.
17. J. Wang, F. Lu, L. Angnes, J. Liu, H. Sakslund, Q. Chen, M. Pedrero, L. Chen, O. Hammerich, 'Remarkably Selective Metallized Carbon Amperometric Biosensors', *Anal. Chim. Acta*, **305**, 3–7 (1995).
18. H. Hinkers, C. Sundermeier, R. Lürick, F. Walfort, K. Cammann, M. Knoll, 'Amperometric Microelectrode Array in Containment Technology', *Sens. Actuators B*, **26–27**, 398–400 (1995).
19. (a) D. Müller, H. Wilke, N. Kolytsheva, 'Activated Platinum Electrodes as Transducer for a Glucose Sensor Using Glucose Oxidase in a Photopolymer Membrane', *Fresenius' J. Anal. Chem.*, **357**, 534–538 (1996); (b) Y. Liu, H. Liu, J. Qian, J. Deng, T. Yu, 'Entrapment of Both Glucose Oxidase and Peroxidase in Regenerated Silk Fibroin Membrane', *Fresenius' J. Anal. Chem.*, **355**, 78–82 (1996); (c) I.G. Casella, M. Gatta, M.R. Guascito, T.R.I. Cataldi, 'Highly-dispersed Copper Microparticles on the Active Gold Substrate as an Amperometric Sensor for Glucose', *Anal. Chim. Acta*, **357**(1–2), 63–71 (1997); (d) J. Li, L.S. Chia, N.K. Goh, S.N. Tan, H. Ge, 'Mediated Amperometric Glucose Sensor Modified by the Sol–Gel Method', *Sens. Actuators B*, **40**, 135–141 (1997).

20. P. Bergveld, 'Development of an Ion-sensitive Solid-state Device for Neurophysiological Measurements', *IEEE Trans. Biomed. Eng.*, **BME-17**, 70–71 (1970).
21. I. Lundström, A. van der Berg, B.H. van der Schoot, H.H. van der Flekkert, M. Armgarth, C.J. Nylander, 'Field Effect Chemical Sensors', in *Sensors. A Comprehensive Survey*, eds. W. Göpel, J. Hesse, N. Zemel, VCH, Weinheim, 467–528, Vol. 2, 1991.
22. H.L. Schmidt, W. Schuhmann, F.W. Scheller, F. Schubert, 'Specific Features of Biosensors', in *Sensors. A Comprehensive Survey*, eds. W. Göpel, J. Hesse, N. Zemel, VCH, Weinheim, 717–818, Vol. 3, 1992.
23. A.R. Baker, 'Improvements in, or Relating to, Electrically Heatable Filaments', *UK Patent*, 892 530, 1962.
24. O.S. Wolfbeis, *Fiber Optic Chemical Sensors and Biosensors*, CRC Press, Boca Raton, FL, 1991.
25. F.L. Dickert, M. Vonend, H. Kimmel, G. Mages, 'Dyes of the Triphenylmethane Type as Sensor Materials for Solvent Vapors', *Fresenius' Z. Anal. Chem.*, **333**, 615–618 (1989).
26. F.L. Dickert, A. Haunschild, P. Hofmann, G. Mages, 'Molecular Recognition of Organic Solvents and Ammonia: Shapes and Donor Properties as Sensor Effects', *Sens. Actuators B*, **6**, 25–28 (1992).
27. J. Homola, S.S. Yee, G. Gauglitz, 'Surface Plasmon Resonance Sensors: Review', *Sens. Actuators B*, **54**, 3–15 (1999).
28. G. Sauerbrey, 'Verwendung von Schwingquarzen zur Wägung dünner Schichten und zur Mikrowägung', *Z. Phys.*, **155**, 206–222 (1959).
29. F.L. Dickert, P. Forth, W.-E. Bulst, G. Fischerauer, U. Knauer, 'SAW Devices – Sensitivity Enhancement in Going from 80 MHz to 1 GHz', *Sens. Actuators B*, **46**, 120–125 (1998).
30. (a) J.B. Briot, S. Ballandras, E. Bigler, G. Martin, 'Gravimetric Sensitivity of Transverse Waves Trapped by Metal Gratings on Thin Quartz Plates', in *Proceedings of 1997 IEEE International Frequency Control Symposium, Orlando, Florida*, IEEE Cat. No. 97CH36016, IEEE, Piscataway, NJ, 207–212, 1997; (b) M. Tom-Moy, R.L. Baer, D. Spira-Solomon, T.P. Doherty, 'Atrazine Measurements Using Surface Transverse Wave Devices', *Anal. Chem.*, **67**, 1510–1516 (1995).
31. (a) F.L. Dickert, U. Geiger, K. Weber, 'Modified Cyclodextrines as Mass-sensitive Coatings for Solvent Vapor Detection', *Fresenius' J. Anal. Chem.*, **364**, 128–132 (1999); (b) F.L. Dickert, M. Tortschanoff, K. Weber, M. Zenkel, 'Process Control with Mass-sensitive Chemical Sensors – Cyclodextrin Modified Polymers as Coatings', *Fresenius' J. Anal. Chem.*, **362**, 21–24 (1998); (c) F.L. Dickert, U.P.A. Bäumler, H. Stathopulos, 'Mass Sensitive Solvent Vapor Detection with Calix[4]resorcinarenes: Tuning Sensitivity and Predicting Sensor Effects', *Anal. Chem.*, **69**, 1000–1005 (1997).
32. (a) G. Wulff, 'Molecular Imprinting in Cross-linked Materials by Means of Matrix Molecules – On the Way to Synthetic Antibodies', *Angew. Chem., Int. Ed. Engl.*, **34**(17), 1812–1832 (1994); (b) D. Kriz, O. Ramström, K. Mosbach, 'Molecular Imprinting – New Possibilities for Sensor Technology', *Anal. Chem.*, **69**, 345A–350A (1997); (c) D. Kriz, O. Ramström, A. Svensson, K. Mosbach, 'Introducing Biomimetic Sensors Based on Molecularly Imprinted Polymers as Recognition Elements', *Anal. Chem.*, **67**, 2142–2144 (1995).
33. (a) F.L. Dickert, P. Forth, P. Lieberzeit, M. Tortschanoff, 'Molecular Imprinting in Chemical Sensing – Detection of Aromatic and Halogenated Hydrocarbons as well as Polar Solvent Vapors', *Fresenius' J. Anal. Chem.*, **360**, 759–762 (1998); (b) F.L. Dickert, O. Hayden, 'Imprinting with Sensor Development – On the Way to Synthetic Antibodies', *Fresenius' J. Anal. Chem.*, **364**, 506–511 (1999).
34. T. Seiyama, T. Nakahara, T. Takeuchi, 'Overview of Gas Sensors for Environmental Use', *Stud. Environ. Sci.*, **59**, 233–272 (1994).
35. A.M. Dietrich, J.N. Jensen, W.F. da Costa, 'Chemical Species', *Water Environ. Res.*, **68**(4), 391–406 (1996).
36. A.A. Suleiman, G.G. Guilbault, 'Piezoelectric Crystal Detectors for Environmental Pollutants', *Stud. Environ. Sci.*, **59**, 273–303 (1994).
37. K.R. Rogers, E.J. Poziomek, 'Fiber Optic Sensors for Environmental Monitoring', *Chemosphere*, **33**(6), 1151–1174 (1996).
38. R. Narayanaswamy, 'Chemical Transducers Based on Fiber Optics for Environmental Monitoring', *Sci. Total Environ.*, **135**(1–3), 103–113 (1993).
39. C. Camara, C. Perez-Conde, M.C. Moreno-Bondi, C. Rivas, 'Fiber Optical Sensors Applied to Field Measurements', *Tech. Instrum. Anal. Chem.*, **17**, 165–193 (1995).
40. D.L. Fox, 'Air Pollution', *Anal. Chem.*, **59**, 280R–294R (1987).
41. N. Yamazoe, N. Miura, 'Environmental Gas Sensing', *Sens. Actuators B*, **20**, 95–102 (1994).
42. N. Yamazoe, 'The State of the Art of Gas Sensing for Environmental Protection', in *Technical Digest of the Seventh International Meeting on Chemical Sensors, Beijing, 1998*, 2–5, 1998.
43. W. Göpel, U. Guth, R. Götz, M. Rösch, 'Exhaust Gas Sensing with Solid Electrolytes: Quo Vadis', in *Proc. Sensor 99 Nürnberg*, 125–130, Vol. 2, 1999.
44. J.K. Atkinson, A.W.J. Cranny, W.V. Glasspool, J.A. Mihell, 'An Investigation of the Performance Characteristics and Operational Lifetimes of Multi-element Thick Film Sensor Arrays Used in the Determination of Water Quality Parameters', *Sens. Actuators B*, **54**, 215–231 (1999).
45. A.W. Czarnik, 'A Sensor for Landmines', *Nature (London)*, **394**, 417–418 (1998).

46. M.S. Nieuwenhuizen, 'Detection of Clandestine Explosives', in *Organic Energetic Compounds*, ed. P.L. Marinkas, Nova Science Publishers, Commack, 251–287, 1996.
47. H. Suzuki, T. Hirakawa, S. Sasaki, I. Karube, 'An Integrated Three-electrode System with a Micromachined Liquid-junction Ag/AgCl Reference Electrode', *Anal. Chim. Acta*, **387**, 103–112 (1999).
48. G. Heiland, D. Kohl, 'Physical and Chemical Aspects of Oxidic Semiconductor Gas Sensors', in *Chemical Sensor Technology*, ed. T. Seiyama, Kodansha, Tokyo and Elsevier, Amsterdam, 15–38, Vol. 1, 1988.
49. K. Takahata, 'Tin Dioxide Sensors – Development and Applications', in *Chemical Sensor Technology*, ed. T. Seiyama, Kodansha, Tokyo and Elsevier, Amsterdam, 39–55, Vol. 1, 1988.
50. Y. Nakamura, 'Stability of the Sensitivity of SnO_2-based Elements in the Field', in *Chemical Sensor Technology*, ed. T. Seiyama, Kodansha, Tokyo and Elsevier, Amsterdam, 71–82, Vol. 2, 1989.
51. H.P. Hübner, E. Obermeier, 'Reactively Sputtered Tin Dioxide Thin-film Gas Sensors: Correlation between Fabrication Parameters and Co-sensitivity', *Sens. Actuators B*, **17**, 351–354 (1989).
52. G.B. Barbi, J.P. Santos, P. Serrini, P.N. Gibson, M.C. Horrillo, L. Manes, 'Ultrafine Grain-size Tin-dioxide Films for Carbon Monoxide Monitoring in Urban Environments', *Sens. Actuators B*, **24–25**, 559–563 (1995).
53. K. Kaneyasu, Y. Setoguchi, S. Sonoda, T. Nakahara, I. Aso, N. Nakagaichi, 'A Carbon Dioxide Gas Sensor Based on Solid Electrolyte for Air Quality Control', in *Technical Digest of the Seventh International Meeting on Chemical Sensors, Beijing*, 184–186, 1998.
54. G. Wiegleb, J. Heitbaum, 'Semiconductor Gas Sensor for Detecting NO and CO Traces in Ambient Air of Road Traffic', *Sens. Actuators B*, **17**, 93–99 (1994).
55. H. Torvela, H. Harkoma, S. Leppävuori, 'Detection of the Concentration of CO Using SnO_2 Gas Sensors in Combustion Gases of Different Fuels', *Sens. Actuators*, **17**, 369–375 (1989).
56. M. Hanada, H. Koda, K. Onega, T. Onouchi, 'Development of CO_2 Sensor Using Lanthanum Doped Tin-dioxide Semiconductor Gas Sensor', in *Proceedings of Sensor 1995, Nürnberg*, ACS Organizations, Wunstorf Steinhude, 445–450, 1995.
57. G. Huyberechts, P. Szecówka, J. Roggen, B.W. Licznerski, 'Simultaneous Quantification of Carbon Monoxide and Methane in Humid Air Using a Sensor Array and an Artificial Neural Network', *Sens. Actuators B*, **45**, 123–130 (1997).
58. Z. Jin, H.-J. Zhou, Z.-L. Jin, R.F. Savinell, C.-C. Liu, 'Application of Nano-crystalline Porous Tin Oxide Thin Film for CO Sensing', *Sens. Actuators B*, **52**, 188–194 (1998).
59. G. Martinelli, M.C. Carotta, M. Ferroni, Y. Sadaoka, E. Traversa, 'Screen-printed Perovskite-type Thick Films as Gas Sensors for Environmental Monitoring', *Sens. Actuators B*, **55**, 99–110 (1999).
60. P. Nelli, L.E. Depero, M. Ferroni, S. Groppelli, V. Guidi, F. Ronconi, L. Sangaletti, G. Sbereveglieri, 'Sub-ppm NO_2 Sensors Based on Nanosized Thin Films of Titanium–Tungsten Oxides', *Sens. Actuators B*, **31**, 89–92 (1996).
61. (a) T. Takada, 'Ozone Detection by In_2O_3 Thin Film Gas Sensor', in *Chemical Sensor Technology*, ed. T. Seiyama, Kodansha, Tokyo and Elsevier, Amsterdam, 59–70, Vol. 2, 1989; (b) T. Takada, K. Suzuki, M. Nakane, 'Highly Sensitive Ozone Sensor', *Sens. Actuators B*, **13–14**, 404–407 (1993).
62. R.L. Jones, R.A. Cox, D.E. Williams, K.F.E. Pratt, P. McGeehin, J. O'Sullivan, 'Tungsten Oxide Sensors Employed for the Atmospheric Ozone Measurements: Comparison with UV Analyzers', in *Proc. Sensor99, Nürnberg*, AMA Service, Wunstorf, Germany, 159–164, Vol. 1, 1999.
63. G. Huyberechts, M. van Muylder, M. Honoré, J. Desmet, J. Roggen, 'Development of a Thick Film Ammonia Sensor for Livestock Buildings', *Sens. Actuators B*, 18–19, 296–299 (1994).
64. D. Mutschall, K. Holzner, E. Obermeier, 'Sputtered Molybdenum Oxide Thin Films for NH_3 Detection', *Sens. Actuators B*, **35–36**, 320–324 (1996).
65. T. Nakahara, H. Koda, 'Tin Dioxide Gas Sensor – A New Approach to Odor Sensing', in *Chemical Sensor Technology*, ed. N. Yamazoe, Kodansha, Tokyo and Elsevier, Amsterdam, 19–32, Vol. 3, 1991.
66. V. Lantto, J. Mizsei, 'H_2S Monitoring as an Air Pollutant with Silver-doped SnO_2 Thin Film Sensors', in *Proceedings of Third International Meeting on Chemical Sensors, Cleveland, Ohio*, 39–43, 1990.
67. M. Straková, E. Matisová, P. Šimon, J. Annus, J.M. Lisý, 'Silicone Membrane Measuring System with SnO_2 Gas Sensor for On-line Monitoring of Volatile Organic Compounds in Water', *Sens. Actuators B*, **52**, 274–282 (1998).
68. M. Kleitz, E. Siebert, P. Fabry, J. Fouletier, 'Solid State Electrochemical Sensors', in *Sensors. A Comprehensive Survey*, eds. W. Göpel, J. Hesse, N. Zemel, VCH, Weinheim, 341–428, Vol. 2, 1991.
69. H. Meixner, J. Gerblinger, U. Lampe, M. Fleischer, 'Thin-film Gas Sensors Based on Semiconducting Metal Oxides', *Sens. Actuators B*, **23**, 119–125 (1995).
70. H.-J. Schreiner, K.H. Härdtl, 'Temperaturunabhängige Sauerstoffsensoren zur Regelung von Verbrennungsvorgängen', in *Sensoren und Messtechnik: Vorträge der ITG Fachtagung vom 9. bis 11. März 1998 in Bad Nauheim*, VDE-Verlag, Berlin, 109–116, 1998.
71. W.L. Worrell, 'Solid Electrolyte Sensors for SO_2 and/or SO_3', in *Chemical Sensor Technology*, ed. T. Seiyama, Kodansha, Tokyo and Elsevier, Amsterdam, 15–38, Vol. 1, 1988.

72. G.-Y. Adachi, N. Imanaka, 'Development of an SO_x-sensor Based on Metal Sulfates', in *Chemical Sensor Technology*, ed. N. Yamazoe, Kodansha, Tokyo and Elsevier, Amsterdam, 131–145, Vol. 3, 1991.
73. B. Ross, G. Steiner, M. Kiesshauer, M. Bradter, K. Cammann, 'Instrument with Integrated Sensors for a Rapid Determination of Inorganic Ions', *Sens. Actuators B*, **26–27**, 380–383 (1995).
74. O. Wassmus, K. Cammann, 'Nitrate Sensor System for Continuous Flow Monitoring', *Sens. Actuators B*, **18–19**, 362–364 (1994).
75. A.V. Legin, E.A. Bychkjov, Yu.G. Vlasov, 'Analytical Applications of Chalcogenide Glass Chemical Sensors in Environmental Monitoring and Process Control', *Sens. Actuators B*, **24–25**, 309–311 (1995).
76. M.S. Miloshova, B.L. Seleznev, E.A. Bychkov, 'Chalcogenide Glass Chemical Sensors for Determination of Thallium in Natural and Wastewater', *Sens. Actuators B*, **18–19**, 373–375 (1994).
77. H. Sakai, S. Iiyama, K. Toko, 'Evaluation of Water Quality and Pollution Using Multichannel Sensors', in *Technical Digest of the Seventh International Meeting on Chemical Sensors, Beijing*, 873–875, 1998.
78. J. Parellada, A. Narváez, M.A. López, E. Domínguez, J.J. Fernández, V. Pavlov, I. Katakis, 'Amperometric Immunosensors and Enzyme Electrodes for Environmental Applications', *Anal. Chim. Acta*, **362**, 47–57 (1998).
79. M.J. Tierney, H.-O.L. Kim, M. Madou, T. Otagawa, 'Microelectrochemical Sensor for Nitrogen Oxides', *Sens. Actuators B*, **13–14**, 408–411 (1993).
80. (a) P. Jeroschewski, K. Haase, A. Trommer, P. Gründler, 'Galvanic Sensor for Determination of Hydrogen Sulfide', *Electroanalysis*, **6**, 769–772 (1994); (b) P. Jeroschewski, K. Haase, A. Trommer, P. Gründler, 'Galvanic Sensor for the Determination of Hydrogen Sulfide/Sulfide in Aqueous Media', *Fresenius' J. Anal. Chem.*, **346**, 930–933 (1993); (c) P. Jeroschewski, S. Braun, 'A Flow Analysis System with an Amperometric Detector for the Determination of Hydrogen Sulfide in Waters', *Fresenius' J. Anal. Chem.*, **354**, 169–172 (1996); (d) P. Jeroschewski, 'Amperometrischer Mikrosensor für *In Situ*-messungen von Schwefelwasserstoff in Gewässern und Sedimenten', in *Umweltdiagnostik mit Mirkosystemen*, ed. G. Henze, M. Köhler, J.P. Lay, Wiley-VCH, Weinheim, 158–180, 1999.
81. R. Renneberg, B. Gründig, 'Biosensoren für die Umweltanalytik', *Nachr. Chem. Tech. Lab.*, **42**, 45–49 (1994).
82. T. Gibson, G. Nelson, J. Rippeth, 'Novel Sensor for Detection and Monitoring of OP Pesticides', in *Proc. Sensor99, Nürnberg*, 111–116, Vol. 2, 1999.
83. F. Opekar, K. Štulik, 'Electrochemical Sensors with Solid Polymer Electrolytes', *Anal. Chim. Acta*, **385**, 151–162 (1999).
84. T. Trapp, B. Ross, K. Cammann, E. Schirmer, C. Berthold, 'Development of a Coulometric CO_2 Gas Sensor', *Sens. Actuators B*, **50**, 97–103 (1998).
85. P.R. Teasdale, A.I. Minett, K. Dixon, T.W. Lewis, G.E. Batley, 'Practical Improvements for Redox Potential (EH) Measurements and the Application of a Multiple-electrode Redox Probe (MERP) for Characterizing Sediment In Situ', *Anal. Chim. Acta*, **367**, 201–213 (1998).
86. (a) T. Doll, I. Eisele, 'Gas Detection with Work Function Sensors', *SPIE*, **3539**, 96–105 (1998); (b) A. Fuchs, M. Bögner, T. Doll, I. Eisele, 'Room Temperature Ozone Sensing with KI Layers Integrated in HSGFET Gas Sensors', *Sens. Actuators B*, **48**, 296–299 (1998).
87. S. Wakida, M. Ohnishi, M. Yamane, K. Higashi, J. Liu, X. Wu, Z. Zhang, 'Studies of Acid Rain Monitoring Sensors Based on Semiconducting Devices (VIII); pH-sensitive Field-effect Transistors Using Tertiary Amines as Sensing Materials', in *Technical Digest of the Seventh International Meeting on Chemical Sensors, Beijing*, 879–881, 1998.
88. N.F. Starodub, N.I. Kanjuk, A.L. Kukla, Yu. M. Shirshov, 'Multi-enzymatic Electrochemical Sensor: Field Measurements and their Optimization', *Anal. Chim. Acta*, **385**, 461–466 (1999).
89. S.M. Klainer, J.R. Thomas, J.C. Francis, 'Fiber-optic Chemical Sensors Offer a Realistic Solution to Environmental Monitoring Needs', *Sens. Actuators B*, **11**, 81–86 (1993).
90. S.M. Klainer, K. Goswai, D.K. Dandge, S.J. Simon, N.R. Herron, D.L. Eastwood, L.A. Eccles, 'Environmental Monitoring Applications of Fiber Optic Chemical Sensors (FOCS)', in *Fiber Optic Chemical Sensors and Biosensors*, ed. O.S. Wolfbeis, CRC Press, Boca Raton, FL, 83–122, Vol. 2, 1991.
91. U. Panne, R. Niessner, 'A Fiber-optical Sensor for Polynuclear Aromatic Hydrocarbons Based on Multidimensional Fluorescence', *Sens. Actuators B*, **13–14**, 288–292 (1993).
92. D. Frense, A. Müller, D. Beckmann, 'Detection of Environmental Pollutants Using Optical Biosensor with Immobilized Algae Cells', *Sens. Actuators B*, **51**, 256–260 (1998).
93. E.F. Carome, G. Fischer, V. Kubulins, 'Fiber-optic Sensor System for Hydrocarbon Vapors', *Sens. Actuators B*, **13–14**, 305–308 (1993).
94. K.T. Larjava, K.E. Tormonen, P. Jaakkola, A.A. Roos, 'Field Measurements of Flue Gases from Combustion of Miscellaneous Fuels Using a Low-resolution FTIR Gas Analyzer', *J. Air Waste Manage. Assoc.*, **47**, 1284–1290 (1997).
95. H. Ahlberg, S. Lundqvist, R. Tell, T. Andersson, *Spectrosc. Eur.*, **6**, 22 (1994).
96. E. Mágori, R. Strzoda, E. Chemisky, B. Borchert, M. Fleischner, H. Meixner, I. Eisele, 'Spektroskopische

Gasdetektion im Nachen INfrarot (NIR) mit DFB-Laserdioden', in *Sensoren und Messtechnik: Vorträge der ITG Fachtagung vom 9. bis 11. März 1998 in Bad Nauheim*, VDE-Verlag, Berlin, 117–126, 1998.
97. F.L. Dickert, S. Schreiner, G.R. Mages, H. Kimmel, 'Fiber-optic Sensor for Organic Solvents in Wastewater', *Anal. Chem.*, **61**, 2306–2309 (1989).
98. F.L. Dickert, H. Besenböck, M. Tortschanoff, 'Molecular Imprinting Through van der Waals Interactions: Fluorescence Detection of PAHs in Water', *Adv. Mater.*, **10**, 149–151 (1998).
99. K.M. Leonard, 'Development of a Fiber-optic Chemical Sensor for Multicontaminant Monitoring of Environmental Systems', *Sens. Actuators B*, **24–25**, 458–461 (1995).
100. J.-S. Yang, T.M. Swager, 'Porous Shape Persistent Fluorescent Polymer Films: An Approach to TNT Sensory Materials', *J. Am. Chem. Soc.*, **120**, 5321–5322 (1998).
101. M. Ralfs, J. Heinze, 'Disposable Optochemical Sensor for the Determination of Chlorine Concentrations in the ppb Range', *Sens. Actuators B*, **44**, 257–261 (1997).
102. B. Zimmermann, J. Bürck, H.-J. Ache, 'Studies on Siloxane Polymers for NIR-evanescent Wave Absorbance Sensors', *Sens. Actuators B*, **41**, 45–54 (1997).
103. T. Vo-Dinh, 'SERS Chemical Sensors and Biosensors: New Tools for Environmental and Biological Analysis', *Sens. Actuators B*, **29**, 183–189 (1995).
104. J. Meyer, U. Karst, 'Zeitverzögerte Fluoreszenzspektroskopie mit Lanthanidkomplexen – Prinzipien und Anwendungen', *Nachr. Chem. Tech. Lab.*, **47**, 1116–1119 (1999).
105. A.L. Jenkins, O.M. Uy, G.M. Murray, 'Polymer-based Lanthanide Luminescent Sensor for Detection of the Hydrolysis Product of the Nerve Agent Soman in Water', *Anal. Chem.*, **71**, 373–378 (1999).
106. K.-T. Lau, J. Micklefield, J.M. Slater, 'The Optimization of Sorption Arrays for Use in Ambient Conditions', *Sens. Actuators B*, **50**, 69–79 (1998).
107. F.L. Dickert, M.E. Zenkel, W.-E. Bulst, G. Fischerauer, U. Knauer, 'Fullerene/Liquid Crystal Mixtures as QMB- and SAW-coatings – Detection of Diesel- and Solvent-vapors', *Fresenius' J.Anal. Chem.*, **357**, 27–31(1997).
108. A. Hierlemann, U. Wiemar, G. Kraus, M. Schweizer-Berberich, W. Göpel, 'Polymer-based Sensor Arrays and Multicomponent Analysis for the Detection of Hazardous Organic Vapors in the Environment', *Sens. Actuators B*, **26–27**, 126–134 (1995).
109. F.L. Dickert, O. Hayden, M.E. Zenkel, 'Detection of Volatile Compounds with Mass-sensitive Sensor Arrays in the Presence of Variable Ambient Humidity', *Anal. Chem.*, **71**, 1338–1341 (1999).
110. S.L. Rose-Pehrsson, D. Di Lella, J.W. Grate, 'Smart Sensor System and Method Using a Surface Acoustic Wave Vapor Sensor Array and Pattern Recognition for Selective Trace Organic Vapor Detection', *US Patent*, 5 469 369, 1992.
111. R. Lucklum, S. Rösler, J. Hartmann, P. Hauptmann, 'On-line Detection of Organic Pollutants in Water by Thickness Shear Mode Resonators', *Sens. Actuators B*, **35–36**, 103–111 (1996).
112. 'Product Information', Endress + Hauser Gastec, Gerlingen, 1995.
113. 'Product Information', Unitronic, Düsseldorf, 1999.
114. 'Product Information', HKR Sensorsysteme, Munich, 1995.
115. 'Product Information', 'MOSES II', Lennartz Electronic, Tübingen, 1997.

Food Analysis Techniques: Introduction

Robert J. McGorrin
Oregon State University, Corvallis, USA

1 Introduction 1
2 Analytical Challenges for Food Assays 1
3 Sample Preparation 1
4 Macrocomponent Analyses 2
5 Microcomponent Analyses 3
6 Physical Methods 3
7 Special Applications 3
Abbreviations and Acronyms 4
References 4

1 INTRODUCTION

Food is a complex, heterogeneous mixture of vastly numerous and diverse biochemical substances. Food chemists are concerned with the composition and properties of food products, beverages, confections, and flavors and the chemical and physical changes they undergo during post-harvest handling, processing, storage, and preparation. Measuring the quality, safety, sensory properties, nutrition, and stability of food products is the primary concern of analytical chemists working within the food processing industry, and in academic and government laboratories affiliated with food programs. The Nutrition Labeling and Education Act of 1990 (NLEA) set new labeling requirements for all packaged foods sold in the USA,[1] which often necessitates the determination of both macro food components and trace nutrients by analytical techniques. Besides quality assurance, the goals of food analysis may be directed towards basic research (healthy food ingredients, bioactive compounds for disease prevention) or product development (convenient preparation, better tasting).

The range of food products which may require analysis encompasses foods which may be respiring (meats, fruits and vegetables) but not growing, or others capable of growth (seeds, eggs). Food products are prepared from disrupted components such as flour; fruit and vegetable purees and juices; and isolated carbohydrates, proteins, and vegetable oils. Dairy chemists utilize milk as a feedstock for production of cheese and cultured dairy products, which require analysis to ascertain the fermentation endpoints for optimum flavor and texture. Foods are usually highly hydrated, but their moisture contents may range from more than 95 percent to less than 4 percent.

Appearance, color, flavor, and texture are critical aspects for the sensory quality of foods. Alternatively the instability of foods, which limits their shelf life, can be initiated by chemical reactions, enzymatic reactions, or physical forces. Analysis of the degree of emulsification or moisture content, for example, may allow a prediction of the shelf life limit to be made for a salad dressing or a potato chip snack, respectively. Changes in texture are the result of toughening or softening, loss or gain in solubility or water-holding capacity. Flavor alteration may result from development of oxidative rancidity, undesirable browning reactions, or desirable compounds produced through cooking. Typical color changes include darkening or bleaching and the development of desirable colors, or production of off-colors. In most cases, food assays have been developed to enable measurement of food product attributes, depending on the specific application. Precise characterization of foods enables new product development and quality control, while troubleshooting product problems and customer complaints.

2 ANALYTICAL CHALLENGES FOR FOOD ASSAYS

The isolation and measurement of individual chemical species in food systems usually poses a difficult challenge. Despite the availability of modern techniques of separation and identification, such as gas chromatography/mass spectrometry (GC/MS), high-performance liquid chromatography/tandem mass spectrometry (HPLC/MS/MS), etc., rarely is it possible to load a syringe with a food sample, and directly inject it into an analytical instrument to obtain a sensible result! Because of the diversity of food products and the variety of analytical methods depending on the intended application, procedures for preparation of different food sample categories need to be developed, evaluated, and collaboratively studied among various laboratories as an integral part of any analytical method.

3 SAMPLE PREPARATION

Sample preparation and extraction remain among the most time-consuming and error-prone steps in the food analysis process. However they are critical procedures because food scientists need to isolate and concentrate a wide range of analytes from complex and varied matrices.

The chemical analysis of foods and raw materials is a multistep process with three key stages: sampling,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

homogenization, and sample preparation.[3] *Sampling* involves the procurement of a representative sample for the laboratory. The sampling process may require dividing a food into edible and nonedible portions, and making selections based on the analytical objective, e.g. selecting only for ripened rather than over-ripe fruit, or for baked goods with correct degree of browning. A *homogeneous* composited sample is prepared by chopping or grinding, followed by blending or mixing. This ensures that a sample is obtained which is representative of the overall food mixture, from which several subsamples can be taken (e.g. imagine sampling a pizza which contains mushroom, sausage, and green pepper components). Care must be taken to avoid contaminating the sample or exposing it to excessive heat, which can cause loss of volatile flavor components or accelerate decomposition of labile food components such as vitamins. Finally, *sample preparation* is the physical and chemical manipulation of the test portion prior to analytical measurement. Sample preparation typically involves pH adjustment, extraction with organic solvents, solid-phase extraction, cleanup (filtration, liquid–liquid partitioning), and concentration. As an alternative, headspace concentration may be performed to capture volatile components as an alternative to solvent extraction. In this encyclopedia, food sample preparation topics are discussed in subsequent articles by M.J. Lichon, **Sample Preparation for Food Analysis, General**; and A. Chaintreau, **Sample Preparation, Headspace Techniques**.

New instrumental techniques that are easily automated and integrated with other analytical systems are key to productivity and data integrity for many laboratories. Automated instruments are helping to drive interest in improved sample extraction and preparation methods. These techniques are not only environmentally sound, because they incorporate low-volume or solvent-free extractions, but in most cases, they can increase the efficiency of the extraction. Examples include supercritical fluid extraction, accelerated solvent extraction, solid-phase extraction, solid-phase microextraction (SPME), and thermal desorption (TDS). Discussion of these instrumental methods can be found in the article by M. Rothaupt, **Sample Preparation Analytical Techniques for Food**. Modular systems are now readily available that automate these procedures and bridge them to analytical instruments, such as gas chromatography (GC), GC/MS or high-performance liquid chromatography (HPLC).

4 MACROCOMPONENT ANALYSES

Proximate analyses (fat, protein, moisture, carbohydrate, and ash contents) have been typically performed using classical wet chemistry methods to ensure the suitability of food products.[4] New developments in instrumental techniques are providing rapid, automated alternatives to classical methods.[4,5] Current regulatory requirements for food manufacture obligate not just the analysis of fat content, but also the quality of the fat, i.e. saturated, monounsaturated, polyunsaturated, monoglycerides, diglycerides, cholesterol, etc. Classical total protein determinations were historically based on nitrogen content (the Kjeldahl technique); however modern electrophoretic and chromatographic methods are being developed which enable rapid and specific amino acid and peptide assays. Bulk carbohydrate values are required to be listed as specific sugars: simple, complex, starch, dietary fiber, and crude fiber. Moisture content has been redefined to specify water as being ''free'', ''bound'', or ''hydrated'', and all of these modes require a specific analytical measurement. These fundamental aspects of food analysis are described in subsequent articles by S.S. Nielsen, **Proximate Assays in Food Analysis**; D. Marini, **Lipid Analyses in Food**; S. Dierckx, K. Boeve, J. Van Camp, and A. Huyghebaert, **Proteins, Peptides, and Amino Acids Analysis in Food**; H.N. Englyst, **Starch Analysis in Food**; H.N. Englyst, M.E. Quigley, and G.J. Hudson, **Dietary Fiber Analysis as Non-starch Polysaccharides**; and D. Reid, **Water Determination in Food**.

HPLC is a very powerful and versatile technique for analyzing and purifying biomolecules, and consequently is a well-established mode for food and beverage analysis. Ongoing advances in column supports, bonding chemistry, and packing materials have enabled increased speed and efficient separation of a wide range of organic and inorganic food analytes at low parts per million and high parts per billion levels. Related column techniques that involve separation by either ionic mobile phases or charged fields are ion chromatography (IC) and capillary electrophoresis (CE) respectively. Typical food analytes quantified by these techniques include sugars, aspartame, phospholipids, vitamins, caffeine, organic acids, ionic species, peptides, and proteins. More sensitive fluorescence detectors for HPLC instruments enable lower limits of detection for vitamins, proteins, aspartame, and mycotoxins. In this encyclopedia, these topics are covered in articles by L.M.L. Nollet, **Liquid Chromatography in Food Analysis**; P.F. Cancalon, **Electrophoresis and Isoelectric Focusing in Food Analysis**; and S. Nakai and Y. Horimoto, **Fluorescence Spectroscopy in Food Analysis**.

Spectroscopic techniques are highly desirable for analysis of food macrocomponents because they often require minimal sample preparation, provide rapid analysis, and have the potential to run multiple tests on a single sample. These advantages particularly apply to nuclear magnetic resonance (NMR), infrared, and near-infrared

spectroscopy. The latter technique is routinely used as a quality assurance tool to determine compositional and functional analysis of food ingredients, process intermediates, and finished products. In this encyclopedia, these instrumental methods are discussed by J.S. de Ropp and M.J. McCarthy, **Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials**; R. Mehrotra, **Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis**; and B.G. Osborne, **Near-infrared Spectroscopy in Food Analysis**.

5 MICROCOMPONENT ANALYSES

Volatile compounds are among the samples most often analyzed in foods, flavors, and aroma systems. The combination GC/MS has been used since the 1960s to characterize flavor mixtures.[6] More recent developments in capillary column technology and integrated headspace sampling systems have shortened the assay time, and increased limits of detection for potent flavor components.[3] This topic is discussed by Y. Chen and C.-T. Ho, **Flavor Analysis in Food**.

Additional trace components in foods include colorants, pigments, vitamins and minerals. Recent interest in so-called ''functional foods'', which impart health benefits beyond basic nutrition, relies on the development of new assays with increased sensitivity and specificity to measure efficacious phytochemical components in foods. In this encyclopedia, techniques for analysis of vitamins are described by L. Ye and R.R. Eitenmiller, **Vitamins Analysis in Food**.

Atomic spectroscopic methods including atomic absorption, atomic emission, and elemental mass spectrometry are routinely used to measure elements (e.g. sodium and calcium) and heavy metals (e.g. lead and cadmium) in food products. Inductively coupled plasma mass spectrometry (ICP/MS) offers the advantage of providing simultaneous multielement measurement, which facilitates analysis of the large numbers of elements in food samples. For further discussion, see S.A. Baker and N.J. Miller-Ihli, **Atomic Spectroscopy in Food Analysis**.

6 PHYSICAL METHODS

Various physical properties are important to measure and characterize attributes related to food texture, viscosity, emulsion stability, and particle size. ''Texture'' is typically associated with solid or semi-solid foods (for example cream cheese, chocolate, jelly) and can be measured using rheological tests which qualify shear rate, compression, fracturability, and viscoelastic properties. However, fluid foods such as sugar syrups, vegetable oils, and wines can also be characterized by these techniques. Measurement and understanding of physical attributes enables control and adjustment of food sensory properties including ''mouthfeel'', crispness, stickiness, chewiness, and creaminess. Instruments which measure these properties include viscometers, pressure/force transducers (Instron), and rheometers. These tools facilitate food process uniformity, standardization, and quality assurance and are fully discussed in the section by M.A. Rao, **Viscosity of Food: Measurement and Application**.

The measurement of gas, liquid, solid, or botanical ''particles'' in foods relates to quantifying their particle size distribution. Examples of applications for particle size analysis include control of grittiness from lactose crystals in ice cream, creaminess of margarine emulsions related to water droplet size, and separation stability of salad dressings due to oil droplet size and distribution in the continuous phase. Size distribution of powdered food ingredients such as starches, spray-dried milk, or encapsulated flavors may relate to speed of water dispersion or perceived grittiness in the mouth. Several instrumental techniques are used, which include microscopy, laser light scattering, field-flow fractionation, NMR, and ultrasonic spectroscopy. Particle size techniques are further described in the article by M.M. Robins, **Particle Size Analysis in Food**.

7 SPECIAL APPLICATIONS

Enzymes, microbes and antibodies can be used as biosensors to rapidly measure the presence of amylases or proteases in foods, which can lead to starch thinning or bitter flavors, respectively. Enzyme-linked assays have been used to measure vitamins, antibiotic residues, microbial toxins, various sugars, ethanol, and carbon dioxide. In certain cases, enzyme activity assays are useful food safety probes to measure whether adequate processing of a food product has been achieved. This is further discussed by J.R. Powers, **Enzyme Analysis and Bioassays in Food Analysis**.

Analytical techniques have been developed for determination of authenticity of suspect food products and ingredients, where price differentials prompt economic fraud. Modern examples include substitution of corn syrup for honey; organic acids and sugars for apple juice; and benzaldehyde for almond extract. Combinations of chromatographic, NMR, isotopic, and enzyme-linked methods have been used to screen suspected food ingredients. (See E. Wilhelmsen, **Adulteration Determination**).

Analytical measurement of toxicological residues in foods including pesticides and aflatoxins is described by M. Volante and P. Branca, **Pesticides, Mycotoxins and Residues Analysis in Food**. Techniques discussed include

chromatographic, spectrographic, and immunochemical methods for assaying these compound classes.

Modern analytical instrumental techniques, when applied as probes for basic research, enable interactions to be studied with in vivo situations. As an example, two-dimensional nuclear magnetic resonance (2-D NMR) techniques have been applied to probe the metabolic processes in plants, and the interaction between plants and soils. This enables in vivo measurements of metabolite concentrations, enzyme kinetics, stable isotope tracers for biochemical pathway elucidation, and structural characterization of organic matter. Recent advances are described elsewhere in this encyclopedia by T.W.-M. Fan and A.N. Lane, **Nuclear Magnetic Resonance in Analysis of Plant Soil Environments**.

ABBREVIATIONS AND ACRONYMS

CE	Capillary Electrophoresis
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS/MS	High-performance Liquid Chromatography/Tandem Mass Spectrometry
IC	Ion Chromatography
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
NMR	Nuclear Magnetic Resonance
SPME	Solid-phase Microextraction
TDS	Thermal Desorption
2-D NMR	Two-dimensional Nuclear Magnetic Resonance

REFERENCES

1. Federal Register, 58, 003 632–690, 2066–2964, Jan. 6, 1993.
2. L.M. Nollet (ed.), *Handbook of Food Analysis*, Marcel Dekker, New York, Vols. 1 and 2, 1996.
3. S.S. Nielsen (ed.), *Food Analysis*, 2nd edition, Aspen Publishers, Gaithersburg, MD, 1998.
4. M.E. Reisman, 'Sample Preparation: Breaking the Bottleneck', *Food Quality*, **10**, 22 (1998).
5. G.J. Demenna, G.J. Brown, 'Analytical Techniques for Today's Food Industry: A Primer', *Food Testing & Analysis*, **1**, Aug./Sept., 32 (1995).
6. G.A. Reineccius, 'Gas Chromatography and Mass Spectrometry in Quality Control and Research', *Food Sci. Technol.*, **45**, (1991).
7. R. Marsili (ed.), *Techniques for Analyzing Food Aroma*, Marcel Dekker, New York, 1997.

Adulteration Determination

Eric C. Wilhelmsen
Wilhelmsen Consulting, Milpitas, USA

1 Introduction 1
2 Regulatory Responses 2
3 Modern Definitions and Types of Adulteration 3
3.1 Misbranding 3
3.2 Economic Adulteration 4
3.3 Filth and Other Foreign Substances as Adulteration 5
3.4 Modern Examples 5
4 Strategies for Determining Adulteration 6
4.1 Specific Marker for an Adulterant 6
4.2 Deviation in a Component 7
4.3 Matrix of Specific Components 8
4.4 Pattern or Profile Recognition 8
5 Economics of Testing 11
6 Specific Methods by Products 12
6.1 Apple Juice 13
6.2 Botanicals 13
6.3 Coffee 14
6.4 Dairy Products 15
6.5 Essential Oils 16
6.6 Maple Syrup and Honey 16
6.7 Meat Products 17
6.8 Olive Oil 18
6.9 Orange Juice 19
6.10 Other Juices 20
7 Future Trends 22
Abbreviations and Acronyms 22
Related Articles 22
References 23

Determining adulteration in food is accomplished by comparing analytical data with historical or control data. Without a proper reference set, the determination of adulteration is impossible. Adulteration is best shown by detecting a foreign component that is characteristic of a specific adulterant. Sometimes it is necessary to use deviations in a parameter, but typically these approaches are less sensitive owing to the natural variation in food. Often practitioners are interested in the more difficult converse, that is, authentication of a product or ingredient. In practice, authenticating a product or ingredient is equivalent to not finding any adulteration. Clearly, authentication is a negative, the absence of adulteration. It can never be entirely proven. The quality of a determination of authenticity is clearly linked to number and types of tests performed. The decision not to perform a test that yields additional information increases the potential of erroneously accepting a sample.

This proof of the negative challenge has led to competition between the practitioners of authentication and those who would adulterate their products or ingredients. This competition is readily apparent. Each time a useful new method has been introduced, previously undetected adulterations have surfaced. Unfortunately, the new methods generally do not supersede older methods and represent additional authentication costs. Fortunately, the vast majority of the food industry is not involved and generally cooperates in policing the marketplace.

1 INTRODUCTION

The history of food authentication closely parallels the early history of food chemistry. Sadly, as people gained a better understanding of the composition of foods, this information was used to improve the formulas for adulterating food. However, early efforts in food chemistry also allowed detection of fraud and adulteration. It is probable that the war between authentication practitioners and those who would deceive the buyer is much older than the historical record.

It is likely that human greed has prompted adulteration for almost as long as food has been traded. Watered wine was familiar to the writers of the Old Testament of the Bible.[1] A stele of King Hummurabi of Babylon from around 1750 BC prohibits the sale of weak and overpriced beer.[2] The extensive trade of Roman times was not without its problems, including complaints about misbranded wine.[3] In 1660, an Act of Parliament was passed in England "for the better ordering of wine retail, and for preventing abuses in the mingling, corrupting and vitiating wines, and for selling, limiting the prices of the same".[4] From the seventeenth century forward the record becomes more extensive. For some, Frederick Accum's publication *A Treatise on Adulterations of Food* was an important turning point. It aroused concern and indignation about the quality and safety of the food supply. A modern textbook[5] cites Accum's treatise and lists most of the following adulterations as not uncommon from this period:

- Annatto – Turmeric, rye, barley, wheat flour, calcium sulfate and carbonate, salt, and Venetian red (ferric oxide, which in turn was sometimes adulterated with red lead and copper) were used as adulterants;

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

- Pepper, black – This product was adulterated with many foreign substances including gravel, leaves, twigs, stalks, pepper dust, linseed meal, and plants other than pepper;
- Pepper, cayenne – To restore color lost from bleaching by light, pigments such as vermillion (α-mercury sulfide), ocher (mixtures of metallic oxides especially iron and clay) and turmeric were used;
- Essential oils – Oil of turpentine, other oils, and alcohol;
- Vinegar – Sulfuric acid;
- Lemon juice – Sulfuric and other acids;
- Coffee – Roasted grains, occasionally roasted carrots or scorched beans and peas; also baked horse liver;
- Tea – Spent, redried leaves and leaves of other plants;
- Milk – The addition of water was the main form of adulteration, hence the name for the village water pump, the ''black cow''; other common adulterants included chalk, starch, turmeric for color, gums, and soda. Occasional adulterants included gelatin, dextrin, glucose, preservatives (borax, boric acid, salicylic acid, sodium salicylate, potassium nitrate, sodium fluoride, and benzoate), and colors including annatto, saffron, caramel, and some sulfonated dyes;
- Beer – Watering was undoubtedly occurring. In addition, ''Black extract'', a water extract of the poisonous berries of *Cocculus indicus* was apparently added to impart flavor, narcotic properties, and toxicity to the beverage;
- Wine – Various colorants, flavors, aging agents, preservatives and antacids were added;
- Sugar – Sand, dust, lime, pulp and coloring matters were added;
- Butter – Excessive salt and water, potato flour, and curds;
- Chocolate – Starch, ground sea biscuits, tallow, brick dust, ocher, Venetian red, and potato flour;
- Bread – Alum and flour made from products other than wheat;
- Confectionery products – Colorants containing lead and arsenic.

Once the seriousness of the adulteration problems became evident in the 1800s, many forces were marshaled to remedy the situation. Legislation was passed. Scientists focused their talents on food chemistry. These forces greatly improved the situation and continue to operate today, but adulteration has still not been eliminated.

In 1922 the United States Department of Agriculture investigated a claim that a chocolate coating was made by illegal substitution of coconut oil for cocoa butter and milk fat. In 1923, convictions were handed down in Federal Court in the United States for adulterating olive oil with peanut oil. In 1926, a butter fraud was detected where a low-grade butter, lard and buttermilk mixture was sold as butter. In 1936, sardines packed in ''olive oil'' were found to be packed in mixed oil. In 1938, a Nebraska firm shipped butter low in butterfat. The list goes on and on.

Today, food adulteration is primarily economic. Consumers purchase a product and receive something else of lesser value, usually a mixture of the authentic product and something to extend it. Examples include corn syrup in honey, hydrolyzed inulin in apple juice, grapefruit juice in orange juice, and other food oils in olive oil. These substitutions are all easy to detect with modern methods. More examples and details are included in section 3.4 below. Sadly, one can only wonder about what types of deception are still defeating the screening methods.

2 REGULATORY RESPONSES

Regulatory response is usually started by passing a law or act. Such laws allow various types of enforcement action including standards, definitions and test procedures to be defined. Around the world, different nations have taken different approaches. Sometimes these approaches are viewed as nontax trade barriers because they are too restrictive or are arbitrary. These barriers classify product from other geographic regions as out-of-specification. Such barriers are mentioned here only because their expressed purpose is to prevent adulteration.

The first food laws in the United Kingdom were enacted in 1860 and 1875. Canada enacted its first in 1875. The first comprehensive food law in the United States was passed in 1906. This law prohibited the use of certain harmful chemicals in foods and prohibited the interstate commerce of misbranded or adulterated foods. In the United States, concerns about adulteration and false health claims led to passing the Food, Drug and Cosmetic Act (FDCA) in 1938, which was later revised in 1958 with the Food Additives Amendment of the Federal FDCA. This amendment documented the principle of establishing the safety of chemicals before allowing use, which is now a widely accepted practice all over the world. In Europe, the European Union is working towards a harmonized set of regulations. This is still very much a work in progress. Internationally, Codex Alimentarius has developed more than 220 commodity standards, 40 codes of practice and limits for more than 500 food additives. Both these organizations are moving toward more horizontal regulations that cut across commodities. These laws, acts and other forms of legislation provide

general guidance for the most part and yield definitions such as those discussed in the next section.

Sometimes organizations develop definitions. For example in the United States, the Code of Federal Regulations included standards of identity that define the preparation and composition of some products and ingredients. If a product does not conform to this specific definition, it is not the product in question. Nonconforming products need to be labeled carefully to avoid misbranding.

In some cases this legislation has led to more specific enforcement tools. The German Richterwerte und Schwankungsbreiten bestimmter Kennzahlen (RSK) values are a specific example. These values are intended as guidelines for verifying the nonadulteration of fruit juices and other products produced from them. The RSK values are intended to include the natural variation in these commodities. In practice, there continues to be discussion about what was included and not included in the original data. The Dutch and Japanese have also developed their own reference values.

In some cases, these reference values have been supported by specific methods. Around the world, different organizations have supplied certified methods. AOAC (Association of Official Analytical Chemists) International is one certifying organization. The Technical Committee for Juice and Juice Products (TCJJP) has submitted juice authentication methods to AOAC through its membership. In Europe the International Federation of Fruit Juice Producers has developed a compendium of certified methods.

At the bottom line, a practitioner seeking to authenticate a food product must understand his/her goal. The requirements to prove adulteration legally vary with country. The decision to buy or not buy a lot of a commodity can require a lower level of proof than required to obtain judgement in court or gain a criminal conviction. In the United States, federal prosecutors have been successful in prosecuting some cases.[6] For example, they obtained:

- A $100 000 fine and 5 year prison sentence for the former president and chief executive officer of an orange juice company that put more than 40 million gallons of adulterated orange juice on the United States market over 11 years.
- Fines and forfeitures totaling $120 000 for a seafood company and two of its principals for adding water to scallops to increase their net weight and thus net profits since scallops are priced by weight.
- A $2.18 million fine for an established baby-food manufacturer for selling a product labeled ''100 percent'' apple juice but which actually contained only sugars, water, and flavorings.

3 MODERN DEFINITIONS AND TYPES OF ADULTERATION

For the purposes of this article, food will be considered to be adulterated if:

- Any valuable constituent has been in whole or in part omitted;
- Any substance has been substituted wholly or in part therefore;
- Damage or inferiority has been concealed in any manner; or
- Any substance has been added thereto or mixed therewith so as to increase its bulk or weight or reduce its quality or strength, or make it appear better or of greater value than it is.

This definition is based on the United States FDCA but has general applicability. The vast majority of this article focuses on the procedures and practices used to detect one or more of these criteria.

3.1 Misbranding

Misbranding comes in many forms. It is the mislabeling of a product. It is one of the faults associated with the sale and distribution of an adulterated product. Properly labeled, many misbranded products would be legal but they might lack the consumer appeal of the mislabeled product. Declaring an inferior grade of olive oil to be extra virgin is an example of misbranding that would fall in the third definition of adulteration. To label a product 100% juice that is in reality 10% hydrolyzed inulin syrup (HIS) is another example of misbranding. This type of misbranding would fall into the second class of adulteration, where something is added. Anytime a product does not conform to the information on the label, misbranding has occurred.

When a food processor purchases an ingredient from a supplier, it is generally purchased with the intent to use it in a product where it will be identified on the ingredient list as the material purchased. If a food processor buys an ingredient such as apple juice concentrate from a supplier, it is usually with the expectation that they will be able to use the same name on their ingredient list. If the concentrate in question is not pure apple juice concentrate, any product made from the concentrate would be misbranded because it misidentifies the ingredient. Therefore a user can be guilty of misbranding with no knowledge of the infraction. However, in most of the world, intent would be considered before invoking criminal penalties.

Misbranding also extends to ingredients that are misnamed. Such products fit into the third class of adulterated foods where inferiority is being concealed

but are otherwise not very interesting from an analytical point of view. To illustrate this point we can consider a set of examples. There is a growing interest in healthy and natural foods. Sugar and other traditional sweeteners such as corn syrup are undesirable. To overcome these objections, formulators have normally used fruit juice concentrates, rice syrup and other sweeteners. More recently, there have been a number of new sweetening ingredients that are breaking new ground and finding acceptance in the marketplace. Some examples include ''agave nectar'', a fructose syrup made from the root of the agave plant which has no juice component, ''evaporated cane juice'', a partially refined sucrose from sugar cane that is not a juice product, and ''a natural extract from corn and barley malt'', an exotic form of glucose syrup. Given that these ingredients have been purchased in good faith by the processors who labeled them as the ingredients they purchased, are the producers guilty of misbranding? Potentially yes, if some regulatory agency such as the United States Food and Drug Administration determines that these ingredients were improperly labeled when they were sold. For these examples, no authoritative statement has been made, so a processor needs to decide whether the names being used are appropriate. If these names are challenged by a regulatory authority, the marketing advantage obtained by using these exotic names could be lost, with other damaging consequences. From the analytical perspective of determining adulteration in foods, the problems of misnaming a food and the resulting misbranding are not particularly germane. Such problems are for lawyers and regulators who are attempting to maintain a level playing field where consumers can purchase products that are fairly labeled.

Misbranding and adulteration do not extend to the problem of quality. Ginseng root is reported to lack potency until it is over 6-years old. However, ginseng powder in the absence of a standard of identity or other regulatory definition would refer to the powder from *Panax giseng*, irrespective of potency. At present, methodologies are still in development for authenticating botanicals, but there are many problems in dealing with them. However, in the broadest sense, important and valuable components are missing, so this ginseng could be considered adulterated. Many other products could lack valuable components as well. Processed orange juice will have less vitamin C than fresh orange juice. Yet this loss of vitamin C is not considered a form of adulteration.

3.2 Economic Adulteration

At present this is the most significant class of adulteration problems. It can involve either removing or diluting components, or it can involve adding components. In most cases misbranding also occurs. The key feature of economic adulteration is the use of an ingredient of lesser value or cost than the authentic product. This provides a greater volume of product that was cheaper to produce and can still be sold at an equivalent or almost equivalent price.

Water is the adulterant of choice. It is generally cheaper than any food product. A slight dilution can provide a handsome increase in the profit margin of a food company if their customers do not notice the reduction in component concentration. Water has been added to many beverage products to increase the volume. Sometimes other components are added to mask the addition of water. Various sugars and organic acids can be used to mask water addition, depending on the product.

The general rule for detecting economic adulteration is to monitor an expensive component that is present in consistent amounts. Monitoring specific anthocyanins is an appropriate way to detect the addition of plum juice in dark cherry juice. The presence of naringin in orange juice usually points to the addition of grapefruit juice. The postmanufacturing addition of orange pulp wash is another form of economic adulteration. The legal status of this practice varies from location to location. However, no jurisdiction considers pulp wash to be synonymous with orange juice.

Economic adulteration can also involve removing components. It is possible to refine a product to such an extent that it becomes something else. For example low-quality grape or apple juice can be treated extensively with activated carbon or ion-exchange resins. These treated juices will become colorless and almost flavorless sugar solutions. The organic acids, the polyphenols and some minerals will have been removed. The materials will no longer be juice. These materials would make a very good base for a fruit-flavored punch. How far down this path should a processor be allowed to go and still call the product a juice? Enforcement in this area is still very difficult.

Economic adulterations generally pose no health risk to the consumers. They do, however, distort the marketplace. A reputable processor cannot compete when a competitor can get 10% more profit per unit of production by selling adulterated product. The lack of health risk makes obtaining resources to investigate economic adulteration hard to justify. Government regulators are generally focused on food safety issues. Food processors are reluctant to spend money to monitor the purity and quality of ingredients if there is only a limited enforcement threat.

A new motivation has arisen to influence United States processors. Plaintiff lawyers have become more aggressive in pursuing class action suits on behalf of the consumers who have been deceived by one or more food

companies. The adverse publicity associated with the courts certifying a plaintiff class has generally prompted companies to settle out of court. Such actions are often pursued after the news media break a story about some adulteration or other food problem.

3.3 Filth and Other Foreign Substances as Adulteration

Although clearly a form of adulteration, filth and other foreign substances are much less significant in today's marketplace. The adulterations of the 1800s discussed above included more of these types of problem. Such problems are generally characteristic of less developed marketplaces, where accepted norms of production and quality are still developing. These conditions prevail in the markets for many botanicals and other herbs used in dietary supplements. An early study on authenticating coffee seems appropriate to recall.[7] The work of Arthur Hill Hassell on coffee in 1850 is cited as an example of using a microscope, an innovation in his day, to detect adulterants such as chicory, roasted wheat, and burnt sugar in coffee. Perhaps purchasers of botanicals should apply the same direct approach.

When other foreign substances are discussed, economic adulterants are generally not considered. Two recent examples will serve to illustrate the distinction. In 1981 in Spain, industrial grade rapeseed oil was substituted for edible olive oil. In this incident, more than 20 000 people were made ill and as many as 800 people died.[8]

Another example is the so-called antifreeze wine problem which came to light in 1985.[9] The intention was to use ethylene glycol to give the impression of greater body. It also enhanced the sweetness of wines, transforming inexpensive wines into expensive tasting wines. Ethylene glycol was also invisible to the normal testing methods of the day. Generally Austrian wines are credited with this scandal but the practice seems to have been more widespread.

The problem with these types of adulteration is that authentication practitioners do not normally apply tests that would detect these materials. One can only speculate about when a new scandal will break. It only takes one individual with flexible ethics to damage a whole industry.

3.4 Modern Examples

Adulteration continues to be a problem. Most of the food industry seems content to produce authentic food, but each time a significantly new methodology is applied to detecting adulteration in food, a new scandal is uncovered. Often the companies hurt most are the honest competitors who did not even profit from the adulteration. It should be noted that generally all the facts do not emerge when these events are uncovered. Often, there are different perspectives. No matter how carefully examples are presented, some bias will occur.

Cane sugar and corn syrup have a long history of use in foods. Prior to the application of stable isotope ratio analysis (SIRA), these materials were more difficult to detect in food products where they did not belong. In practice it is possible to determine whether a product contains sugar produced by Calvin cycle plants (C-3), by Hatch–Slack (C-4) plants or by a plant using the unusual Crassulacean acid metabolism (CAM).[10] This methodology detected substantial adulteration of orange juice, apple juice, honey and maple syrup with these inexpensive corn and cane sweeteners in the 1970s and 1980s.

In the mid-1980s, a company was producing a 100% juice product relying on a mixed fruit concentrate as a major component. Unfortunately, it eventually came out that a major component of the mixed fruit concentrate was a fruit syrup made by strong ion-exchange treatment of a low-quality juice product. This low-cost component provided a substantial and unfair cost advantage. This cost advantage prompted the investigation by competitors that led to the discovery of the nonjuice component.

In the late 1980s, low juice levels were detected in cranberry juice cocktails. Finished products did not contain the necessary 25% juice. The detection of additional citric and malic acids and of grape skin pigments was pivotal.[11] Some samples were determined to have as little as 5% juice.

In a case that still plagues the Gerber Products Company, who had absolutely no involvement, another major baby-food provider was detected selling artificial apple juice in 1988. The baby-food manufacturer did not actually make the artificial product. They purchased it as a concentrate from another manufacturer. Nevertheless, substantial fines and criminal penalties were imposed. This artificial product lacked components such as chlorogenic acid. It also contained D-malic acid from the use of racemic malic acid in the formulation.

In the mid-1990s a family-owned honey and syrup making operation netted $500 000 from its bogus products. This company was investigated because consumers complained that the products did not taste authentic.

In late 1995 and 1996, apple juice concentrate was entering the Unites States that contained HIS in the place of some of the apple juice concentrate. While the product, which is a high-fructose syrup, was not considered to be a health risk, it did prompt sampling and testing nationwide. Many companies voluntarily pulled product from stores shelves. Class action litigation followed. The HIS was detected with a capillary GC (gas chromatography) method that had not yet received great acceptance.[12]

In 1997 at the Food Authenticity – Issues and Methodologies (FAIM) workshop in La Baule, France, a wide range of adulteration concerns were raised.[13] These included:

- inappropriate strains of bacteria in cultured dairy products claiming to contain special probiotic species;
- nonmilk fat in butter;
- the use of ultrafiltration to increase the yield of cheese through the inclusion of whey proteins;
- cheeses made from milk of different species than claimed (e.g. goat cheese including cow's milk);
- soft wheat adulteration of durum wheat semolina and pasta;
- orange juice adulteration with other sugars, other fruit, colors, flavors and pulpwash;
- beet sugar in honey;
- virgin olive oil contaminated with other refined oils such as sunflower oil;
- mixing of species within meat products;
- mislabeling of fish species.

In 1998 and 1999, juice drink products were observed to contain less than their declared juice content. Other drinks have contained undeclared juices in lieu of the declared ingredients. Nut confections have been found to contain traces of undeclared nut meats from other species and have apparently caused severe reactions and death.

The war goes on between authenticators and those individuals who would profit through deception. Authentication practitioners continue to seek newer and more efficient methods. Others seek new ways to squeeze extra profits out of their sales, even if the customer does not receive quite the product they wanted. Unfortunately, as in most wars, it is the innocent bystanders who suffer most. When an adulteration scandal breaks, the whole industry is suspect. The ethical producers are painted with the same brush, lose good will, and generally lose revenues.

4 STRATEGIES FOR DETERMINING ADULTERATION

All strategies for determining adulteration rely on comparing an unknown to a body of data. Such comparisons can be as simple as detecting a foreign component, to as complex as matching spectra to a model adulterated sample. One of the biggest challenges in this process is selecting the data for comparison. If this control group is inappropriate, any determinations are likely to be in error.

There are many recurring themes in adulteration. The chief adulterants are water, sweeteners, or inexpensive cousins. The choice of water is obvious. Dilution of liquid products like milk or juice is an old problem. Sweeteners add bulk and are generally less expensive than other food components. Today, a creative formulator has many sweetener choices. One is generally similar in composition to the product that is to be extended. And finally, a product may have a distant cousin that is much less expensive. Grapefruit dilution of orange juice is an obvious example. Given the recurring themes in adulteration, it should not be surprising that there are recurring themes in determining adulteration. The next five subsections examine some of these themes.

The peer-reviewed literature contains a significant amount of compositional data but not all of it is relevant to authenticating food products. This data requires work to extract but is clearly a reasonable place to start. When one must deal in many commodities, it may be necessary to rely on the efforts of others to collect this data. Generally, it is even more prohibitive to generate one's own data. There are references that tabulate the composition of various food products individually. There are some that tabulate the results for a variety of commodities. Are these results valid? Were the test materials authentic? Do they represent the population of interest? Every practitioner must determine whether a specific data set is appropriate for each determination of authenticity.

In many parts of the world various data compendiums are accepted as a regulatory standard for juice. The previously mentioned RSK values are one such example. The values of the Association Francaise de Normalization and the Criteria for Authenticity in the Netherlands are two more. These compendiums are all of limited value because they do not present the data for individual samples and only present summary data such as the mean and some measure of dispersion. Without actual case-by-case data, many correlations are lost, greatly reducing the information that can be extracted regarding the composition of these commodities. The normative database of the TCJJP is an attempt to overcome this limitation.

There is a final problem associated with selecting data for comparison. How can this data be validated? What is the true value for any compositional parameter? This is a vexing problem when various databases disagree. It is complicated further by the potential for data from adulterated samples to be included in various compendiums. If one orders a simple commodity like apple juice, how many different products are likely to be received? Which ones should be used to generate data?

4.1 Specific Marker for an Adulterant

This is the simplest case and a theme that appears often. Typically there is a marker compound that does

not belong in the food product being tested. This marker is a well-documented constituent of a potential adulterant. This approach assumes knowledge about the food product. The comparison relies on the knowledge that the marker is not found in the food product. Given that the marker does not belong in the food, it is expected never to be present at levels at or above the detection limit. Therefore any confirmed detection is indicative of adulteration. The need for confirmation is related entirely to the expected action of the practitioner's client. Litigation will demand a higher standard of proof than a decision not to buy.

Interpretation of these specific markers is straightforward. If the marker is above the level of quantitation it can be used to estimate the level of adulteration based on a typical concentration in the adulterant. Clearly, other markers and components might be desirable to build a stronger case for litigation, but one well-chosen marker should be enough to establish that the product has been modified.

HPLC (high-performance liquid chromatography) analysis of the anthocyanin pigments in a product can provide this type of evidence. The anthocyanins of grape and grape-skin extracts are very different from those found in strawberry. In strawberry one finds anthocyanins based on the anthocyanidins, pelargonidin and a little cyanidin. The anthocyanidins are the aglycones obtained after the removal of the sugar side chains from the anthocyanins. In contrast, grape products include many more, including mavidin, peonidin, delphinidin, cyanidin, petunidin, and pelargonidin. Raspberry only contains cyanidin. Plum contains cyanidin and peonidin. Radish contains pelargonidin. Clearly these compounds have great power for sorting out the presence or absence of the various red fruit and vegetable products used in food. These markers have been used to detect deception in many red products.

Another example is D-malic acid, which is not synthesized by plants. Plants normally produce only L-malic acid. D-malic acid can be assayed by chiral chromatography[14] or with an enzymatic method. The enzymatic approach is based on D-malate dehydrogenase, an enzyme that oxidizes D-malic acid to pyruvate and carbon dioxide in the presence of an appropriate cofactor. For the chromatographic procedure, samples have been analyzed on a polystyrene divinylbenzene copolymer column with 16-mM L-valine/8-mM copper(II) acetate adjusted to pH 5.5 with NaOH as mobile phase and detection at 330 nm. This method was suitable for determining the adulteration of apple juice with D, L-malic acid.

The theme of enantiomeric purity of organic acids has also been applied to grape juice.[15] Grape juices with added synthetic malic and tartaric acids were subjected to ultrafiltration after decolorization with charcoal and separated by HPLC on a TSK™ gel QAE-2SW™ anion-exchange column. Polarized photometric detection (PPD) at 520 nm and UV (ultraviolet) detection at 333 nm were used. UV detection gave positive peaks for each acid but PPD detection gave a positive peak for each L-isomer, a negative peak for the D-isomer and no peak for the racemate. The coupled PPD/UV detection system did not require the physical separation of L- and D-isomers. The method is suitable for the routine examination of grape juice adulteration at the 5% level.

The strategy of a single marker is very cost effective if the suspected adulterant is known. However, if the adulterant is not present, little information has been obtained towards authenticating the unknown. There are still many other possible adulterants that need to be eliminated and a practitioner may wish to consider another strategy even though it might add to his or her costs.

4.2 Deviation in a Component

Sometimes unique markers that indicate the presence of an adulterant are unknown or nonexistent. In these cases another strategy or theme is necessary. Adulterants often change the ratio or ranges of common components. In these cases, an observed value outside of the normal range can provide the necessary evidence.

Statistical analysis is more complex than just achieving a confirmed detection. The mean and distribution of a control population are compared to the value associated with the unknown. A likelihood that the unknown is a member of the population is easily calculated by using a normal approximation which is usually adequate but the specific distribution should be examined before selecting a cutoff point. Two standard deviations above or below the mean is a typical choice. Ultimately, a practitioner needs to decide whether to accept or reject a sample. Statistical analysis provides a rationale for the decision. Multiple questionable samples from a single source can be used to increase the sensitivity of this approach. Five samples each with a likelihood of being real of less than 5% is an unlikely event and would certainly provide a reason to investigate further.

Proline and sorbitol levels in apple juice are examples of these types of indices. On a single strength basis, proline is generally 2–5 ppm in apple juice. When it is greater than 10 ppm, it becomes likely that pear juice was added. Clearly, the more proline the greater the suspicion. Sorbitol can provide similar information. In apple juice, sorbitol levels are typically 0–1%. Sorbitol values much greater than 1% again indicate the probable presence of pear juice. SIRA with ^{13}C also shows deviation. Authentic apple juice will generally yield a value of -26 to -24 parts per thousand. Anything greater than -22 parts per thousand indicates that corn or cane sugars have been added.

In both of these examples, the experts argue about the exact values that should be accepted or rejected. Ultimately, the discussion revolves around the authenticity of various unusual samples that affect the distribution and range of the authentic population. The only way to resolve these differences is to accumulate large databases of samples where authenticity is not questioned.

One recurring approach that is used to overcome the database limitation is to employ an internal reference. For example, in pineapple juice or partially reconstituted pineapple concentrate the ^{13}C levels in the sugars were compared with those in the organic acids when these components were separated on a simple cation-exchange resin, AG50W-X8™, which removes amino acids and other cationic substances (retained on column), an anion-exchange column, AG1-X8 which retains organic acids and other anionic substances, while sugars were removed by washing with water.[16] The sugars, and malic and citric acids were analyzed and the ^{13}C isotope ratios were determined using MS (mass spectrometry) coupled to an elemental analyzer. This procedure is appropriate for pineapple because one of the major economic challenges has been in obtaining enough acidity in product from Thailand which has been the low-cost source since the late-1970s.

A second example is the use of honey protein to look for adulteration.[17] Protein was isolated from honey by tungstic acid precipitation, and the $\delta^{13}C$ value of the protein and the honey was determined by the AOAC Official Method 31.158-31.161. A difference of more than 1.0% suggests adulteration of the honey. This theme was also followed, using malic acid as a control in apple juice.[18]

This internal control approach is used for data besides stable isotope ratios. Practitioners examine fructose : glucose ratios, ratios of various flavonoids, and D-malic acid as a fraction of total malic acid. Nevertheless, this approach is still limited to confirming the presence or absence of a specific adulterant or family of adulterants. In contrast to the specific marker approach, these procedures do confirm that at least one attribute of the sample is within the normal range, but there is still much room for adulterants to go undetected.

4.3 Matrix of Specific Components

The matrix approach promoted by Brause et al.[19] is the approach taken by most practitioners for authenticating juices. It can be applied to other matrices as well. In this approach an array of tests is performed that are designed to detect the full array of expected adulterants. Such an array usually generates enough data for most unknown adulterants to yield one or more suspicious values. This is not always true. Apple juice concentrate adulterated with hydrolyzed inulin is an example of an adulterant missed by people who were slow to adopt the capillary chromatography procedure.[20]

Two practical matrices are presented in Table 1 and in Table 2. Table 1 presents the matrix for apple juice. It provides a rationale for detecting pear juice, beet sucrose, beet invert sugar, high-fructose corn syrup (HFCS), cane invert sugar, cane sucrose, hydrolyzed inulin, and totally synthetic ingredients such as those sold in the infant apple juice scandal in 1988. Similarly, Table 2 presents a matrix for orange juice. In orange juice, the likely adulterants are somewhat different. Grapefruit juice and pulp wash need to be considered. Matrices such as these can be run for about $500 per sample at a contract laboratory which will provide an interpretation. However, different contract laboratories will have their own versions of these matrices. Some will add methods. Some will use different values. There is still some art in the authentication process.

This approach still has the potential to miss an adulterant. Fundamentally, this approach is looking for a specific set of adulterants and has the potential to miss an adulterant selected to avoid detection by this screening procedure. However, this is more complicated than it sounds. A successful adulterant must have much of the general composition of the food product in question to avoid detection and still be cheap enough to justify the effort and risk. This concern prompts some professionals to hold back some information about their matrix to leave some uncertainty for the would-be adulterators.

4.4 Pattern or Profile Recognition

This strategy or approach is the new frontier in authentication and is still evolving. Instead of using a shortlist of analyses like the matrix approach, one or more profiling methods are used. A profiling method is characterized as one that generates large amounts of information that can characterize a product. Often much of this information is not understood directly, but is instead just considered descriptively. This is not unlike the whirls and patterns in a fingerprint.

Chromatographic and spectroscopic methods have great potential. The key is in developing specific procedures that yield a profile that includes the information necessary to distinguish authentic from adulterated samples. One of the major challenges in this area is the lack of a standard data format or a standard procedure for generating the profiles.

Near-infrared (NIR) spectroscopy falls into this strategy. It has been applied to a wide range of foods and ingredients.[21] Usually, the spectra look so similar that chemometric analysis tools are required to extract the information and make possible discrimination between authentic and adulterated samples. Many practitioners

Table 1 Apple juice matrix for authenticating apple juice and detecting some common adulterants[a]

	Authentic apple juice	Pear added	Beet sucrose added	Beet invert added	HFCS added	Cane invert added	Cane sucrose added	Totally synthetic	HIS added
% Fructose	5–8%	5–8%	<5%	>5%	>5%	>5%	<5%	<6%	>5%
% Glucose	1–4 %	1–4%	1–3%	>4%	>4%	>4%	1–3%	>3%	<3%
Fructose : glucose ratio	>1.6	>1.6	Variable	<1.6	<1.6	<1.6	Variable	1.0	>1.6
% Sucrose	0–5%	0–5%	Could be >5%	0–3%	0–3%	0–3%	Could be >5%	Variable	<3%
Chlorogenic acid	>30 ppm	>30 ppm	Lower but variable	Lower but variable	Lower but variable	Lower but variable	Lower but variable	None	Lower but variable
Total malic acid (mg/100 mL)	200–900	200–900	200–700	200–700	200–700	200–700	200–700	200–900	200–700
L-Malic acid	>90% of total	>90% of total	>90% of total	>90% of total	>90% of total	>90% of total	>90% of total	50% of total (racemic)	>90% of total
Proline	2–5 ppm	2–5 ppm	≫10 ppm	2–5 ppm	2–5 ppm	2–5 ppm	2–5 ppm	None	2–5 ppm
Sorbitol	0–1%	≫1%	0–1%	0–1%	0–1%	0–1%	0–1%	None	0–1%
$\delta^{13}C$ (SIRA)	-25 ± 1 ppt	-25 ± 1 ppt	-25 ± 1 ppt	-25 ± 1 ppt	>-22 ppt	>-22 ppt	>-22 ppt	Depends on sugar source	-25 ± 1 ppt
Oligosaccharides by HPLC/PAD	No extra peaks	No extra peaks	No extra peaks	2 extra peaks	1 extra peak	2 extra peaks	No extra peaks	2 extra peaks	?
Oligosaccharides by capillary GC	No extra peaks	No extra peaks	No extra peaks	2 extra peaks	2 extra peaks	2 extra peaks	No extra peaks	2 or more extra peaks	2 extra peaks

[a] Reproduced with permission from Allan Brause, ACSC, Columbia, MD.
HFCS, high-fructose corn syrup; PAD, pulsed ampherometic detection.

Table 2 Orange juice matrix for authentication of orange juice and detecting some common adulterants

	Authentic orange juice	Pulp wash added	Cane sucrose added	Beet sucrose added	Beet medium invert added	Beet invert added	Grapefruit added	Apple juice added
% Fructose	1.8–2.8%	Variable	Lower	Lower	Higher	Same	Same	Higher
% Glucose	1.8–2.8%	Variable	Lower	Lower	Higher	Same	Same	Variable
% Sucrose	3.5–5.5%	Variable	Higher	Higher	Lower	Same	Same	Lower
% Total sugars	7.2–10.8%	Variable	Same	Same	Same	Same	Same	Same
Glucose/fructose	1.00 ± 0.10	Can vary	Same	Same	Same	Same	Same	<1
%Total sugars/°Brix	61–91%	Can vary	Same	Same	Same	Same	Same	Same
Sucrose/total sugars	30–60%	Can vary	Increase to >60%	Increase to >60%	Can be <30%	Same	Same	Lower
Abs_{443}/Abs_{325}	>0.100	Decreases	No extra peaks	Varies	Varies	Varies	Varies	Varies
Fluorescence	1–2 peaks	2–3 Peaks increase abs. 300–306 nm	1–2 Peaks	1–2 Peaks	1–2 Peaks	1–2 Peaks	1–3 Peaks	1–2 Peaks
Sodium	<50 ppm	Maybe >50 ppm	Variable	Variable	Variable	Variable	Same	Same
Potassium	>1400 ppm	Variable	Can be lower	Can be lower	Can be lower	Can be lower	Same	Same
$\delta^{18}O$ if from concentrate	$+14 \pm 1$ ppt	Variable	Lower	Lower	Lower	Lower	Same	Lower
$\delta^{18}O$ if fresh	+3 to +7 ppt	Variable	Lower	Lower	Lower	Lower	Same	Lower
Naringin	None	None	None	None	None	None	0.10 ppm	None
Hesperidin/narirutin	>2.00	<2.00	Same	Same	Same	Same	Lower	Same
Benzoate	None	Maybe up to 100 ppm	None	None	None	None	May be present	None
Isocitric acid	>44 ppm	Variable	Lower	Lower	Lower	Lower	Higher	Lower
$\delta^{13}C$ (SIRA)	<−23.5 ppt	<−23.5 ppt	Much higher	Same	Same	Same	Same	Same
Sorbitol	None	None	None	None	None	None	None	Present
Carotenoids	Low β-carotene High β-cryptoxanthin	Same as pure orange juice	Lower	Lower	Lower	Lower	Lower	Lower
Oligosaccharides by HPLC/PAD	No extra peaks	No extra peaks	No extra peaks	No extra peaks	4 extra peaks	2 extra peaks	No extra peaks	Variable
Oligosaccharides by capillary GC	No extra peaks	No extra peaks	No extra peaks	No extra peaks	2 extra peaks	2 extra peaks	No extra peaks	No extra peaks
$\delta^{2}H$ by NMR	>103.5 ppm except Brazil: >102.5	Same unless sugar also added	Higher	Lower	Lower	Lower	Same	Same

are uncomfortable with such approaches because it is difficult to check what the statistics indicate manually.

GC yields complex chromatograms that can be used to authenticate products. In one example, 56 pure butter samples from six countries were analyzed.[22] Model adulterants included soya oil, maize oil and lard. Samples were injected onto a 3% OV-1™ packed column with He as carrier gas. The results were evaluated by partial least squares (PLS) regression and neural networks. The results from the two procedures were compared. Whereas neural networks were only used for classification of the samples, PLS also generated quantitative results and was able to extract significant patterns in small calibration sets.

The analysis of vanilla flavors and extracts is a challenging area of authentication. Vanillin, the most important flavor component, is readily synthesized and consumers will pay a premium for natural product. Efforts continue to apply site-specific natural isotopic fractionation study by NMR (nuclear magnetic resonance) to this problem.[23] Analytical improvements have been made but further progress is desired. The key compounds involved are vanillin and *p*-hydroxybenzaldehyde.

An alternative approach is also being developed using a technique described as Global Fingerprint Verification™. The partial results of an unpublished trial conducted with the Flavor Extract Manufacturers' Association (FEMA) are shown in Figure 1 (see Table 3). It is readily apparent from the false color image that Veritek Verification Technologies, Inc. have a technology that shows great promise as a pattern recognition system for categorizing food products. In this illustration, many of the various types of vanilla are correctly categorized with a single dye. In practice many orthogonal dyes would be used to generate a profile that is very specific and sensitive to a wide range of changes. Statistical analysis of the data showed complete discrimination. The automated system routinely performs comparisons in four replicates with only a negligible increase in costs. This technology has already found great success in pharmaceuticals and some success in the food industry.

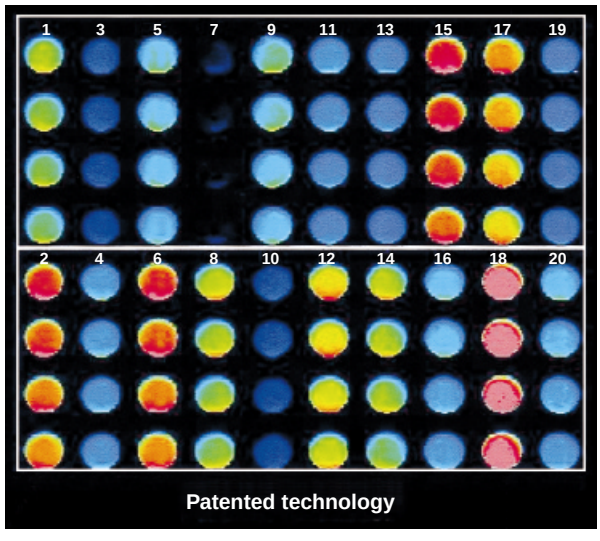

Figure 1 FEMA origin of vanilla extract. Omega for this test is: 0.35. See Table 3 for details of samples.

Table 3 Components of vanilla extract in Figure 1

Sample identification[a]	Mean fluorescent unit
1. Indonesian vanilla beans, #17	5.50
2. Java vanilla beans, #38	10.33
3. Madagascar vanilla beans, #18	1.94
4. Madagascar vanilla beans, #26	5.30
5. Tonga vanilla beans, #30	5.27
6. Java, #14	10.66
7. Bourbon beans (Madagascar origin), #1	0.57
8. Vanilla beans Bourbon, #40	7.76
9. Comoros, #11	5.60
10. Bourbon vanilla beans, #21	2.48
11. Madagascar vanilla beans, #24	4.54
12. Java, #13	9.42
13. Comoros vanilla beans, #27	4.09
14. Bali vanilla beans, #29	7.96
15. Bali vanilla beans, #36	9.43
16. Comoros, #12	5.89
17. Indonesian vanilla beans, #20	8.34
18. Java, #6	13.57
19. Madagascar vanilla beans, #19	3.90
20. Tonga, #16	6.50

[a] #s refer to the FEMA code.

Other profiles are mentioned in the product sections of section 6. These techniques all share the characteristic of collecting large amounts of data and collecting information about the whole product, not just one or two components. This distinction has the potential to break the cycle of adding a new method every time there is a new breakthrough in adulterating products. There will be a learning curve for both the practitioners and those who make decisions based on the advice of the practitioners.

5 ECONOMICS OF TESTING

The economics of testing are complicated. Testing is expensive. No food processor wants to incur the cost. Most food processors believe that they deal with reputable suppliers and therefore have little reason to test. Regulators must respond to political pressure. In the United States this means that they are intent upon pursuing food safety issues. Regulators do not have time and resources to chase an economic adulteration. Consumers want the product to be inexpensive, healthy

and taste wonderful. At the bottom line, no one wants to test. It is amazing that the marketplace is as fair as it is.

For regulators, academic and trade association scientists, there is very little choice. If they receive money to monitor the market, they need to follow through. Granted that such players have liberty in seeking funding, but in practice their decisions about what and how much to test are determined by external factors. The situation for a food processor is somewhat different.

An intentional adulteration by a food processor is not an interesting case for self-monitoring. It is a willful fraud. Someone has made a conscious decision to try to sell inappropriate product. Any testing done by such a processor will be to test the quality of the deception and will be very limited. Such a fraud will decrease the cost of goods and should translate into a lower price, a higher profit margin, or a larger marketing budget. The harsh reality of this unfair competition should prompt competitors to watch each other's products. This type of surveillance should guarantee that adulteration will eventually be discovered.

Unfortunately, if a flexibly minded adulterator is not too greedy, he or she is virtually impossible to detect. Every method and approach has a limit of detection. If a product consistently hugs the boundaries of authentic material, it is likely that the product has been modified. A more intensive investigation is warranted. It is possible that a geographic or varietal variation is responsible but this has not always been the case. Adulteration is increasingly sophisticated and in a low margin business, even a little help can greatly increase profitability.

If processors do not want to monitor their competitors and deal with the problems in their marketplace, they need to insure that regulators or other third party players such as a trade association have the resources to monitor the marketplace. However, the processor must bear in mind that this is a two-edged proposition. They need to be confident that their own products will not be found suspect. There are many examples where this has occurred for specific commodities. Generally this approach works for commodities that are largely vertical, i.e. where there is little risk of buying adulterated ingredients. The goal is to insure that market participants do not intentionally adulterate their products.

For the case of unintentional adulteration, the situation is more interesting for a food processor. If a food processor unknowingly purchases an adulterated ingredient, the product made will be adulterated even though there will have been no intent. The food processor must grapple with the following questions to determine how best to protect the business:

- How likely is the purchase of an adulterated ingredient?
- How likely is any unintentional adulteration to be detected?
- What is the cost of a detected adulteration event?
- How effective is a testing program in preventing unintentional adulteration?

With answers to these questions, a food processor can estimate the extent of exposure and consider how to proceed. Unfortunately, answers to these questions are not exact. If an ingredient is in short supply and expensive, the risk of purchasing an adulterated lot is probably higher than if the ingredient is abundant and cheap. If the supplier is a long-term business partner that you have learned to trust, the likelihood of a problem decreases. However, that trust should be based on facts and not on a friendly smile. Once time purchases are the most likely to cause problems, especially when the entire history of the lot is not known.

The impact of a major unintentional adulteration scandal to a food processor is generally substantial but need not destroy the franchise. The food processor generally needs to recall the affected product, which is an expensive process. In addition the processor needs to act to regain the public trust by explaining how this bad thing happened and why it will not happen again. For a food processor, the potential loss in consumer confidence is probably the most damaging aspect of this turn of events.

It takes a good and extensive testing program to seriously reduce the risk of purchasing adulterated ingredients. However, when a sound testing program is coupled with an approved supplier list and long-standing business relationships are built up, the risk can be greatly reduced at a much lower lost.

Testing and authentication monitoring are like a safety program. It is difficult to eliminate all accidents but the money spent on safety is usually wisely spent if the safety program is well designed and well implemented. A testing and authentication program needs to be an integrated part of the business. It is a tool for purchasing. It is a tool for marketing and sales. However, a poorly designed program is just a way to spend money and may yield no benefits. No company seems immune unless it is entirely vertical. The challenge is to place a value on the various costs, benefits and potential risks and seek an optimal solution.

6 SPECIFIC METHODS BY PRODUCTS

Although it is impossible to cover all commodities where adulteration is a potential problem, owing to space limitations here, the following overviews should provide some access to the literature. The intent is

to illustrate themes used in detecting adulteration. Ashurst and Denis provide a good overview of many methods, including stable isotope analysis, NMR, infrared (IR) spectroscopy, oligosaccharide analysis, enzymatic analyses, DNA and PCR (polymerase chain reaction), electrophoretic, antibody, and various statistical methods for using this data.[24] The intent in the following subsections is to provide numerous recent examples.

6.1 Apple Juice

Apple juice has been used as an example throughout this article. It is well characterized and has been involved in several adulteration scandals. The authentication matrix presented in Table 1 includes seven possible adulterants. However, it does seem appropriate to include a few more references and procedures.

A capillary GC method for the detection of fingerprint oligosaccharides, resulting from the adulteration of fruit juices and other foods with commercial sweeteners, has significantly changed the practice of authentication.[25] After lyophilization and derivatization with Tril-Sil Z™, the food samples were analyzed on an HP-5™ column. The limit of detection for several different sweeteners is about 5%. At this point, no practitioner of authentication should be unaware of this method.

Given the importance of racemic malic acid in the detection of apple juice adulteration, one more method does not seem out of place.[26] L-Malic acid is the predominant acid in apple juice and no D-malic acid should be present; the presence of D-malic acid indicates adulteration of the juice with synthetic malic acid, the racemic mixture that contains 50% D-malic acid. Apple juice samples were extracted with a C_{18} cartridge, filtered and chromatographed on two reversed-phase LC (liquid chromatography) columns connected in series, with 0.2 M phosphate buffer as mobile phase and detection at 214 nm to determine the total malic acid. L-malic acid was determined with the use of an enzymic test kit (Boehringer–Mannheim, Indianapolis, IN, USA), with dilution of the sample by a factor of 10 and absorbance readings at 365 nm as modifications.

It is important to recognize that not everything that is tried works. Sometimes a data set will contain enough information to make all of the desired discriminations. In this example,[27] a pattern recognition program was used in conjunction with a spreadsheet program for classification of authentic apple juice samples of different varieties and geographical origins. The data analysis included HPLC determination of sugars and organic acids, determination of titratable acidity, measurement of degrees Brix, enzymatic determination of L-malic acid and determination of stable carbon isotope ratio. The method successfully classified samples but was not sensitive enough for authentication purposes. In most instances, adulterants were detected at concentrations over 20%.

And last, before leaving apple juice, it seems appropriate to illustrate that an alternative procedure can accomplish the same ends. Research has identified other markers to detect the presence of pear in apple or apple in pear.[28] In this procedure, fruit was extracted with methanol, evaporated and resuspended in water before a column cleanup procedure. Juices could be applied directly to the cleanup column. Various fractions were eluted from this column and analyzed on a column of ODS (octadecylsilane)-Hypersil™. Isorhamnetin glucoside (limit of detection 10 ng), which was undetectable in apples, is reported as an indicator of pear. Phloretin glucoside and phloretin xyloglucoside (limit of detection 7 ng) is reported as an indicator of apple in pear juice.

6.2 Botanicals

Although perhaps not typical food components, many botanicals are entering the diet in dietary supplements and increasingly are being added to mainstream foods. Where results have been published, the approaches for authentication are similar. For many of these materials, methods are marginal at best. However, some procedures have been published and others are under development. Authentication and potency are major issues. In many cases standards are difficult to obtain and there is much discussion about the active components.

For botanicals, the most common deception is the replacement of the desired material with another plant material with the same general appearance. It is not easy to distinguish between various finely ground plant materials. However, at this point it is often difficult to distinguish between low-potency botanicals and mislabeled product. This has prompted some work.

Researchers reported the evaluation of a commercial system, REMED HS, for screening Chinese herbal medicines for adulteration by neutral and basic Western medicines.[29] The system involved online sample purification and extraction, separation by HPLC, scanning UV detection, and evaluation of the spectral and retention data by reference to a library consisting of data for 555 drugs and metabolites. The technique was applied to the analysis of five herbal medicines. Detection limits for most adulterants were 100–300 ng mL^{-1}.

For many botanicals, enantiomeric composition is a useful tool for detecting adulteration.[30] In this example, essential oils were extracted from seeds of dill, and annual and biennial caraway with hexane, hydrodistillation or a combination of both methods. Analysis was by GC with a cyclodextrin in the stationary phase separating authentic samples of (+)- and (−)-limonene, (+)- and (−)-carvone,

and (−)-*cis*/*trans*)-carveol. The enantiomeric composition of dill and annual and biennial caraway seeds were similar, except that dill had a higher (−)/total limonene ratio and contained no (−)-*cis*-carveol.

It is not surprising that regular capillary GC is useful in authenticating botanicals, given the high resolving power of the technique. In this particular example,[31] solid-phase extraction was the preferred method of sample preparation for analyzing espomheira santa, an anti-ulcer drug, because it gave a cleaner sample and consumed less solvent. Resulting solutions were analyzed by GC on a column coated with OV-17-OH and FID (flame ionization detection). This method allows the detection of friedelan-3-ol and friedelin which can be used as markers for the authentic anti-ulcer drug.

In another GC example,[32] supercritical extraction was selected as the method of choice for preparing the samples. Guava leaves were cut in pieces, mixed and ground with granulated celite, and extracted in a supercritical fluid extraction cell using CO_2. The extracted material was trapped by passing the CO_2 effluent through propan-2-ol. The extracted samples were analyzed by capillary GC on a DB1 column.

Materials such as saffron or black pepper are valuable and are thus more likely to be adulterated. Saffron has been characterized by microscopy and by color reactions (under the microscope) with phloroglucinol, H_2SO_4 and molybdophosphate.[33] Saffron can also be differentiated from some substitutes because it is not fluorescent under UV radiation. TLC (thin-layer chromatography) of extracts in aqueous 80% ethanol is carried out on silica gel with butanol–acetic acid–H_2O (4 : 1 : 1) as mobile phase. The three main spots (retention factor = 0.29, 0.43 and 0.56) are caused by the principal pigments of saffron.

Papaya seeds have been used to adulterate black pepper. This form of adulteration can be detected by measuring benzyl glucosinolate (glucotropaeolin).[34] Samples of papaya seed–black pepper are de-fatted by extraction with 2,2,4-trimethylpentane–propan-2-ol (9 : 1) and then air dried. The remaining sample is then extracted with boiling water, sinigrin is added as internal standard, and the solution is treated with barium and mercury acetates. After centrifuging, the supernatant liquid is cleaned up on a microcolumn of DEAE Sephadex A-25. The papaya marker is eluted with 20-mM pyridine acetate and determined by GC. There is also an enzyme procedure based on the release of glucose.

6.3 Coffee

Coffee is another valuable commodity. The economies of many central American nations are strongly tied to this commodity. The value of coffee is linked to its geographic origins and the variety. It should come as no surprise that methods for characterizing coffee and coffee solids have been developed.

NIR and mid-IR spectroscopy has been reported to be useful in authenticating varieties of coffee and in detecting adulteration. In one study,[35] arabica and robusta coffee beans were ground and lyophilized. NIR spectra were recorded in reflectance mode at ambient temperature over 400–2498 nm at 2-nm intervals. Mid-IR spectra were obtained by FTIR (Fourier transform infrared) spectroscopy over 800–4000 cm^{-1} at 8 cm^{-1} resolution. Using discriminant analysis, varieties could be distinguished. In another study,[36] most of the same researchers compare several data processing techniques using IR spectroscopy and chemometrics to detect added glucose, xylose and fructose, or added glucose, starch and chicory. An artificial neural net shows the greatest promise for extracting the classification information. Another related procedure shows promise for authentication.[37] Extracts of arabica, robusta and their blends were either freeze dried or dried on glass-fiber filter paper under reduced pressure, and NIR spectra of the dried material were recorded from 1100 to 2498 nm. The spectra were analyzed by principal-components and factorial discriminant analysis. Of 65 samples, three were misclassified when presented on filter paper, and nine were misclassified when presented freeze dried. Finally, using a different type of sample preparation,[38] 28 samples of roasted coffee beans were ground and mixed with KBr. IR spectra were recorded at 800–4000 cm^{-1}, with a resolution of 8 cm^{-1}, with a diffuse reflectance accessory. Differences between the spectra from arabica and robusta were observed around ~1744 and ~1150 cm^{-1}. Considerable grouping of the spectra, according to the species, occurred. Conventional discrimination analysis applied to the principal component scores was able to give 100% discrimination between the two species. The applicability of the method is limited by the small samples used but it should be useful for detecting adulteration in commercial coffee.

Coffee and coffee solids have been adulterated with various materials. Detection methods have focused on simple sugars, a recurring theme in authentication testing. For example,[39] commercial soluble coffee can be adulterated with coffee husks or parchment. These adulterants can be detected by using xylose as a tracer. Xylose was determined in 700 commercial coffees using anion-exchange chromatography with PAD. The cutoff limit was a total xylose concentration of 0.4% above which the coffee was considered adulterated. Xylose concentration ranged from 0.01 to 3.15%. Of the samples analyzed, 81 were regarded as adulterated with xylose. Total glucose and free manitol were also examined as tracers for adulteration with coffee husks, or parchment

and coffee husks, respectively. Free mannitol levels >0.30% were a sign of adulteration. Free mannitol ranged from 0.01 to 1.55% and total glucose ranged from 0.46 to 42.5%. Adulteration has also been detected using enzyme assays for these sugars[40] or by using GC.[41] All three approaches have found use in many other matrices.

The use of stable isotope techniques is another recurring theme. Stable isotope methods have been applied to coffees and coffee extracts by extracting the caffeine.[42] Coffee beans were extracted with boiling water and the extracts were treated with lead carbonate or MgO and filtered. Caffeine was extracted with $CHCl_3$. Most of the solvent was evaporated and the caffeine was precipitated with light petroleum and crystallized from methanol. The $^{13}C:^{12}C$ and $^{15}N:^{14}N$ ratios of the crystals were determined by isotope ratio MS after combustion and the site-specific $^{2}H:^{1}H$ ratios were determined by NMR after combustion and reduction of H_2O with Zn. Principal component and discriminant analysis permitted the differentiation of natural and synthetic caffeine and of American and African coffee.

6.4 Dairy Products

Dairy products are a valuable source of protein and calcium. Special dairy products have been developed in many cultures to take advantage of the taste and functional differences of milks from different species. These specialty products are again subject to adulteration from milk from other animals, usually bovine milk. In addition, milks are often adulterated with other fats that are less costly than milk fat. Whey and whey proteins have found their way into some products and ingredients as well. There is also the problem of dilution with water. For these commodities, again, the same methodologies and approaches have been applied.

Capillary GC has been used to examine the triglyceride profiles of Mexican cheeses.[43] Cheese was ground for 5 min and extracted with hexane and injected onto a fused-silica column coated with SPB-50. Although the proportions of triglycerides in different cheeses determined on SPB-50 were similar, the concentrations of individual triglycerides and the total content of triglycerides varied between the cheeses. Adulteration of one sample with vegetable oil was detected. A similar procedure has been applied to detect refined animal and vegetable fat in milkfat.[44] The detection of cholesta-3,5-diene and stigmasta-3,5-diene in samples of milk fat indicated adulteration by beef tallow and vegetable oil, respectively.

Other researchers have suggested the use of the dodecanoic-to-decanoic acid ratio for detecting species adulteration in cheese.[45] The dodecanoic-to-decanoic acid ratio distinguished between cow, goat and sheep cheeses (mean values 1.16, 0.46 and 0.58, respectively). The degree of adulteration of goat or sheep cheese with cow's milk can be calculated from the increase in this ratio. In a similar manner, the measurement of the ratio of butyric acid to octadecanoic acid can be used to detect the presence of vegetable oil. This ratio is typically high from November to April, and by taking this into account, the minimum detectable concentration of vegetable fat in milk fat can be reduced from 10.5 to ~8.2%.[46]

Researchers have reviewed other approaches that have been used to measure and profile triglycerides in milk to authenticate various products.[47,48] Methodologies have included HPLC, IR spectroscopy, and isotopic analysis. Differential scanning calorimetry has also been considered for detecting animal fats in butter.[49]

The detection of dilution with water is a very old problem for dairy products. The village pump added to the milk supply for many years. Various approaches can be used to detect water addition, including density, refractometry, and other intrinsic properties. Of these intrinsic properties, freezing point depression seems to have international acceptability.[50] Alternatively, a purchaser can use IR and measure the protein, butterfat, and other milk solids as a basis for purchase. Water content is not particularly important under these conditions.

The problem of species adulteration continues to be the focus of research. Recently, pyrolysis MS was applied to the problem.[51] In order to measure the adulteration of sheep's and goat's milk with cow's milk, portions of the pure milks and binary mixtures were applied to Fe/Ni foils and dried at 50 °C for 30 min. For pyrolysis MS, the foils were heated at 100 °C for 5 s before Curie-point pyrolysis at 530 °C for 3 s (temperature rise time 0.5 s). Data were collected over the range m/z 51–200 and normalized to the total ion count. Cluster analysis of the spectra of the pure milks showed that the major source of variation between the different milks was not the variable fat content. Chemometric models were able to predict concentrations of cow's milk of ≤20% with errors of ≤±1.5% in samples on which they had not been trained.

Most researchers have focused on proteins to detect species adulteration. Different separation methods have been selected but the process is the same, the key being to identify a protein that does not belong in the product. As an example,[52] buttermilk powders, nonfat dry milk, total milk proteins, sodium caseinate, whey proteins and fat globule membrane (FGM) proteins were examined with SDS/PAGE (sodium dodecylsulfate/polyacrylamide gel electrophoresis). Protein bands were stained manually for 15 min in 0.1% Coomassie brilliant blue R dissolved in ethanol–acetic acid–H_2O (3:1:6). Band densities were quantified by scanning at 600 nm in a Bio-Rad™ video densitometer. The limit of detection for the FGM protein

band was 420 ng. Two unique FGM proteins were clearly visible in authentic buttermilk powders. SDS/PAGE using the automated system required for <3 h. Other researchers have used capillary electrophoresis instead of SDS/PAGE.[53,54] These procedures gave good separation of the serum proteins and caseins. Root square deviation (RSD) for migration times and peak areas were 0.085% and 2–4%, respectively. Electrophorograms of goat, cow and sheep milk were compared and were sufficiently different to permit the detection of 1% adulteration from another species. The method could also detect the effect of heat damage on casein, and detect the addition of dried powdered milk to fresh milk. To detect cow's milk in goat's milk, a researcher showed the absence of αS1-casein in goat's milk providing a sensitive and quantitative measure of the cow's milk.[55]

Speciation has also been effected by HPLC.[56] The presence of 1–3% of cow's milk in samples of sheep, goat's, mare's and human milk was detected by subjecting whey or casein samples to HPLC on a reversed-phase Nucleosil™ 300-5C5 column with aqueous 0.1% trifluoroacetic acid and an acetonitrile gradient. Detection of species-specific milk proteins was by absorbance at 280 nm or by fluorescence detection at 340 nm. For sheep and goat's milk, the presence of >1% of cow's milk was shown by a peak associated with bovine β-lactoglobulin A, which was detected at 290 nm in chromatograms of whey samples. The method was also applied to yogurt and cheese samples.

Whey protein is a special problem in dairy products. It has already been mentioned several times. It can often be detected in the protein profiles mentioned above. In addition, whey addition to nonfat dry milk powder can be detected by an elevation of the thiol to milk protein ratio.[57] Whey proteins can also be detected by another compositional variance, determined by changes in the amino acid profile after hydrolysis.[58]

6.5 Essential Oils

Essential oils are expensive and are extensively used in flavors and cosmetics. Many of these oils are well characterized and the compositions are known. Authentic or natural oils are more valuable than synthetic oils, given the importance of origin in labeling. This has prompted the development of many chromatographic procedures. Given that many of these oils contain enantiometrically pure constituents, enantioselective GC is useful in authenticity control for some economically important essential oils.[59] German researchers report a HPLC approach for detecting ethyl *p*-dimethylaminobenzoate or menthyl anthranilate in lemon oil by HPLC.[60] Other researchers have used multidimensional GC coupled with MS and IR to characterize eucalyptus oil.[61] A multidimensional chiral system was used to characterize rose oils.[62] The method was applied to the chirality evaluation of *cis*/*trans* rose oxides, linalol and citronellol from both authentic Bulgarian and Turkish rose oils and commercially available oils. The enantiomeric ratios of genuine constituents of the oils are discussed as indicators of naturalness. Mentha oils have been examined for illegal adulteration using TLC and chiral GC.[63] Other researchers report fingerprints and composition information for a variety of oils including cardamom, lemon grass, nutmeg, etc.[64–67]

Using their high-field NMR, researchers developed a method for detecting adulterated bitter almond and cinnamon oils based on measuring the site-specific deuterium content of benzaldehyde,[68] which is indicative of its source, viz. synthetic (ex-toluene and ex-benzaldehyde), natural (ex-kernels from apricots, peaches and cherries and ex-bitter almonds) and semisynthetic (ex-cinnamaldehyde extracted from cinnamon). Benzaldehyde is the main component of bitter almond oil and samples were analyzed directly. A purified cinnamaldehyde sample was isolated from cinnamon oil and transformed to benzaldehyde by refluxing with 3% aqueous K_2CO_3.

In an approach similar to that for benzaldyde, adulteration of natural mustard oil with synthetic allyl isothiocyanate can be detected by site-specific deuterium NMR.[69] The four monodeuterated isotopomers of allyl isothiocyanate were observed by ^{2}H-NMR, and site-specific hydrogen isotope ratio determinations were performed. Different results were obtained for the natural and the synthetic compounds. The ^{13}C, ^{15}N and ^{34}S isotopic deviations and other isotopic parameters were subjected to principal components analysis and the results indicated the geographical origin of natural mustard oils.

6.6 Maple Syrup and Honey

Maple syrup and honey are relatively unique. They are both mainly sugar but are valued highly for their taste. Maple syrup is mainly sucrose and has been adulterated variously with cane sugar, beet sugar, and other sweeteners. Honey is a mixture of fructose, glucose and a little sucrose. Products differ somewhat geographically but are often valued above and beyond these differences. Honey is often identified having been made from a particular blossom which does impart character to the honey. Given the high sugar content of these two commodities, it is relatively difficult to detect sugar addition if the right materials are used and are used at moderate levels.

It should come as no surprise that isotopic procedures have been dominant. Stable isotope techniques are useful for examining the origins of sugars. Maple syrup can be analyzed by deuterium site-specific nuclear isotope

fractionation NMR after the conversion of sugar to alcohol. Overall $^{13}C:^{12}C$ isotope ratios can be determined by SIRA/MS. With this data, added sugar could be determined at levels of 5–20% of total solids.[70] The $\delta^{13}C$ values for maple syrup, cane sugar, and corn syrup are −24.12, −11.29 and −11.85 parts per thousand, respectively.[71] Similarly, $^{13}C:^{12}C$ isotope ratio can be used to test honey for the presence of corn or maize syrup.[72]

An alternative to the isotopic approaches for detecting added sugars is oligosaccharide analysis, which can be performed in a number of ways. In one example,[73] maple syrups were diluted with HFCS or beet medium invert sugar (BMIS). Samples were prepared by ion exchange through a column comprising AG50W-X8™ cation-exchange resin and a column of AG1-X8 anion-exchange resin followed by a C_{18} Sep-Pak™ cartridge and a filter. Samples were analyzed by HPLC on a CarboPac PA1™ column equipped with sodium acetate with a gradient of sodium hydroxide. PAD was used. Fingerprint oligosaccharides were shown to be present in the inexpensive sweeteners (HFCS and BMIS) which were not present in pure maple syrup. The detection limit was 5% for adulteration of maple syrup with these inexpensive sweeteners. In another paper, researchers used a sample cleanup step to concentrate oligosaccharides from honey and other syrups prior to HPLC analysis.[74] Honey or syrup was diluted with water, and the higher saccharides (which are absent from authentic honey) were concentrated by medium-pressure LC on carbon–Celite with aqueous 7% ethanol as wash solution and aqueous 50% ethanol as mobile phase. The eluted sugars were determined by HPLC on an Ultrasphere™ ODS column with H_2O as mobile phase and a differential refractometric detector. This method was also used for analysis of hydrolyzates of corn, potato, wheat and rice starch. Syrup material from C-3 or C-4 plants could be detected at levels of 1% in honey by this method. Using a similar isolation procedure, oligosaccharide bound galactose is also suggested as marker of adulteration of honey with beet sugar.[75]

Returning to another common theme, capillary GC has been used to examine the oligosaccharides of honey for authentication.[76] Honey and sweetener samples were diluted with H_2O to a Brix value of 5.5. Samples (250 μL) were lyophilized at 25 °C for 10 h and derivatized with 0.5-mL Tri-Sil Z by heating to 80 °C for 1 h. The TMS (trimethylsilyl) derivatives were analyzed on a column coated with DB-5 and FID. The method was used to determine the adulteration of honey with invert syrups (beet or cane). The detection limit for invert sugar addition was 5%.

Flavonoids from honey have been considered as a tool for characterizing types of honey.[77] Honey (~50 g) was mixed with 5 parts of H_2O (acidified to pH 2 to 3 with HCl) until completely fluid, and then filtered through cotton. The filtrate was passed through a column of Amberlite™ XAD-2, which was then washed with acid H_2O, followed by neutral H_2O (pH 7), before elution of the phenolic fraction with methanol. The eluate was concentrated under reduced pressure, dissolved in methanol and further purified by passage through a column of Sephadex LH-20. The second fraction, containing the flavonoids, was concentrated, redissolved in methanol, and analyzed by HPLC on a LiChrocart™ RP-18 column. Sixteen flavonoids were identified. Later work characterized the flavonoids found in Portuguese heather honey.[78]

6.7 Meat Products

Meat is another important class of commodity. As economic status rises, meat consumption rises. However, not all meat is valued equally and there are ways that meat has been extended, including the use of vegetable protein and the proteins from other species. Different species have been mixed, diluting an expensive commodity with a less expensive commodity. The analytical problems concerned with meats are complicated by the extreme changes that occur during processing and cooking. Nevertheless, there are a wide variety of procedures that have been developed.

The most common approach to detecting other species is with an antibody in an immuno-assay of some type. For example,[79] troponin antigen was prepared and antisera were raised in rabbits. The antisera were used in an agar gel immunodiffusion test to identify the species of fresh, cooked and decomposed cattle and buffalo meat. Rabbit anticattle troponin serum was species specific and showed no cross-reactivity. It could detect up to 10% adulteration of fresh cattle meat with other meat and decomposed meat. Rabbit antibuffalo troponin serum had to be adsorbed on glutaraldehyde to make it species specific. It could detect as little as 1% adulteration in fresh buffalo meat and decomposed meat. Neither sera could detect cooked meats.

Other researchers used lactate dehydrogenase (LDH) to detect turkey adulteration in beef or pork.[80] The sample was homogenized with 0.15-M NaCl in 0.01-M sodium phosphate buffer at pH 7.2 and centrifuged. The supernatant was filtered and diluted 1 : 100 to 1 : 10 000 in 0.1-M sodium phosphate buffer at pH 7.2 containing 0.15-M NaCl, 1% BSA (bovine serum albumin) and 0.2% Tween 20. Portions were incubated for 1 h at 37 °C in microtiter plate wells that had been coated with LDH D5E mAb. After washing, biotinylated LDH B3C mAb was added for 1 h at 37 °C and washed again. Finally, an incubation with avidin–horseradish

peroxidase conjugate was accomplished in 30 min at 37 °C. Peroxidase activity was measured using 2,2′-azino-bis-(3-ethylbenzothiazoline)-6-sulfonic acid and H_2O_2 with detection at 405 nm. Linear relationships between the turkey LDH concentration and turkey content in beef and pork were found. The method detected 1% adulteration of beef or pork. Other researchers have reported enzyme-linked immunosorbent assays for distinguishing many different types of meat including unprocessed beef, sheep, horse, kangaroo, pig, camel, buffalo, and goat.[81–83] None of these assays claim to work for cooked meat.

Alternatively, researchers have used protein profiles by electrophoresis or HPLC to detect adulteration. For example,[84] scampi flesh was homogenized in aqueous 2% SDS (sodium dodecylsulfate) and heated at 60 °C for 30 min. After centrifugation, the SDS extracts were separated by SDS/PAGE at 30 mA on 12% total acrylamide gels (2.67% crosslinker) with 4% stacking gels (2.67% crosslinker) and the gels were Ag stained. SDS/PAGE gave profiles characteristic of the shellfish species, enabling adulteration of *Nephraps norvegicus* with Pacific scampi or tropical shrimp to be detected in either raw or cooked meat. In a similar system,[85] the identity of cooked meats was determined by extracting samples with 1% SDS containing 1 mM dithiothreitol and 10 mM EDTA (ethylenediaminetetraacetic acid). The samples were clarified by centrifugation, filtered and subjected to SDS/PAGE. After electrophoresis, proteins were fixed with methanol and acetic acid and stained with Coomassie blue. The meats from mammals and birds of different species produce characteristic patterns of bands, and adulteration by even 10% of meat from a different species could be detected with certainty in binary mixtures. In a third denaturing electrophoretic system, researchers measured duck and goose liver in foie gras.[86]

HPLC has also been used to detect adulteration in chicken or turkey.[87] Samples were blended with water and filtered before analysis on a Hi-Pore™ RP-304 in aqueous 0.1% trifluoroacetic acid. Elution was accomplished with a gradient from 37% to 60% acetonitrile–water (950 : 50). The peaks specific to the meat components were identified and quantified. The response was rectilinear from 5 to 100% of either chicken or turkey. The method gave reproducible results and other meats did not interfere.

Although protein profiles and immunoresponse have been more common, other approaches have been used. The adulteration of beef products by $\geq$5% of pork can be detected by determination of eicosadienoic acid (C_{20} : 2), which is present in pork at $\leq$0.7%, but in beef at 0.05%.[88] The proportions of other fatty acids can be measured for the same purpose. Analysis was effected on a column coated with DB-Wax.

Before leaving meat, there is one more problem that merits examination. Feral or range animals often command a premium price. Canthaxanthin, a carotenoid, is present in farmed salmon or rainbow trout. Astaxanthin is present in wild fish. Either can be extracted from skin or muscle tissue into acetone and ethyl ether. The concentrated extract can be fractionated on Silica gel 60 by TLC with *t*-butyl alcohol–light petroleum (boiling range 40–60 °C) (1 : 7) as mobile phase. All operations were carried out in dim light and with minimum exposure to air. Retention factors were 0.14, 0.43, 0.51 and 0.91 for astaxanthin, astaxanthin monoacyl ester, canthaxanthin diacyl ester and astaxanthin diacyl ester, respectively.[89]

6.8 Olive Oil

Olive oil has been traded for millennia. Its value is determined by both origin and quality. The olive oil marketplace has been rocked by various scandals. Consequently, researchers have developed a number of strategies for authenticating this commodity. Common adulterations include adding other types of oil or manipulating the oil to improve the grade. Extra virgin olive oil is considered the most desirable.

As an example of a profiling approach, a combination of FT (Fourier transform)–Raman spectroscopy and multivariate analysis procedures was proposed for detecting vegetable oils (soybean, corn and raw olive residue oils) in virgin olive oil.[90] The Raman spectra were recorded for 3250–100 cm^{-1}. The spectral data were processed using multivariate procedures. Six genuine virgin oil samples with different chemical compositions were selected to prepare the calibration standards containing 1–10% trilinolein. The method was tested with olive oil samples containing 1, 5 or 10% vegetable oil. Data analysis by regression on principal components produced the correct discrimination between genuine and adulterated oils for all the test samples and identified the correct level of contamination for 93% of the test samples.

Similarly, using NIR spectroscopy (over the range 400–2500 nm), extra virgin olive oil adulterated with sunflower, rapeseed or soya-bean oil could be detected.[91] Using smoothed data over a 10-nm interval in the range 1100–2498 nm, and a combined first derivative, PLS and principal component analysis method, a 10-term equation was developed to predict the extent of oil adulteration. Other IR approaches have also been successful.[92,93]

Other tools, including NMR, HPLC, GC and pyrolysis MS, have been used to detect adulteration. The same approaches are observed in most important food commodities. For example, the NMR spectra were recorded at 600.13 MHz. Multivariate analysis, including a principal components analysis and a variables selection method, was used to differentiate between the oils. Cluster analysis was used to reveal any natural groupings of the oils.

Diacylglycerols, linolenic acid, water, acetic acids, phenols and sterols could all be detected, so the method may be useful for authentication and quality control of virgin olive oil.[94]

By HPLC, various groups examined different markers. For example, when olive oil was diluted with THF (tetrahydrofuran) and with methanol, it could be analyzed by HPLC on a Spherisorb ODS column with 0.05-M sodium perchlorate–methanol (1 : 9) and amperometric detection to measure tocopherols and tocotrienols as markers of authenticity.[95] This reversed-phase HPLC method did not require saponification before analysis. Alternatively, an authenticity factor was developed for detecting linolenic-rich oils in olive oil by HPLC.[96] Researchers have examined the triglycerides of olive oil by HPLC as a means of authentication.[97–99]

Mass spectroscopy has also been applied to olive oil. Not surprisingly it can be quite effective in detecting corn oil and, in some approaches, other seed oils. The combination of pyrolysis MS with a neural network correctly assessed the authenticity of the olive oil samples.[100] In another study,[101] the $^{13}C:^{12}C$ ratio was determined in individual fatty acids by GC combustion/isotope mass spectroscopy. Maize, rapeseed or groundnut oil were saponified with NaOH, cooled, acidified to pH 3 with HCl and the free fatty acids extracted with hexane. Hexane was removed from the samples under N_2 and a 14% BF_3/methanol complex was added and heated to form the methyl esters. The esters were dissolved in hexane and separated by GC on an SGE BPX-70 column. The column was interfaced with an isotope ratio mass spectrometer via a combustion interface that consisted of a ceramic furnace, packed with CuO/Pt catalyst, at 850 °C. Consistent differences were detected in the $\delta^{13}C$ values of oils from different geographical regions. Not surprisingly, direct study of triglycerides and fatty acid profiles by GC can be used to authenticate olive oil.[102]

6.9 Orange Juice

Orange juice is another commodity whose market has been scandalized by adulteration. The number of legal and legitimate products made that are designed to look and taste like orange juice is considerable. Some of these products contain little or no juice. Some use citrus extracts for flavor and others do not. Premiums are paid based on geographical origin. There are many methods for detecting various types of adulteration in orange juice, including C, H or O isotope ratio analysis; site-specific natural isotope fractionation with NMR; HPLC; tracer studies; the Petrus method involving UV and visible spectrophotometry; ICP/AES (inductively coupled plasma/atomic emission spectroscopy); and electrochemical detection.[103] Researchers have supplied various statistical and pattern recognition tools. A matrix for this commodity has already been presented. However, this commodity is at the cutting edge in terms of new methods for authentication and in adulteration.

As we have seen for several other commodities, pyrolysis MS is being applied to authentication problems. Researchers have done the same for orange juice.[104] Orange juice was centrifuged. The supernatant was applied to an Fe–Ni foil and dried for 30 min at 50 °C. The sample was analyzed by pyrolysis MS. Data were normalized as a percentage of the total ion count in order to correct for the effects of sample size. Artificial neural networks, radial basis functions and PLS linear regression were evaluated as techniques for spectral deconvolution. Each method yielded excellent predictions for the level of sucrose adulteration, with a limit of detection of 1 g L^{-1} of added sucrose. However, PLS regression using eight latent variables for predictions yielded the best results. Furthermore, the inputs to multilayer perceptrons could be reduced by the use of principal components analysis from 150 masses to 8 principal component scores without any loss in the predictive ability of the model.

HPLC is an important tool for authenticating orange juice and most liquid food products. In the matrix presented above, it can be used to generate sugar profiles, measure the various organic acids and determine the flavonoids. Other more specific methods have also been developed for various authentication purposes. Specific examples include a study to measure the various saponified orange juice carotenoids.[105] Carotenoids were isolated from orange juice concentrate by precipitation with Carrez reagent [zinc cyanoferrate(II)] and were recovered in a small volume of acetone. After extraction into petroleum ether and a solvent change to diethyl ether, the saponification was carried out with 6 mL of 10% methanolic KOH in the dark. The ether layer was washed with water until free of alkali and evaporated to dryness. The residue was dissolved and analyzed by HPLC with diode-array detection on a nonendcapped C_{30} reversed-phase column in methanol–water–methyl tertiary butylether as mobile phase (18 : 1 : 1) and gradient elution to 19 : 0 : 1 in 12 min, and then to 43 : 0 : 7 in 20 min, to 3 : 0 : 1 in 30 min and 1 : 0 : 1 in 50 min. Thirty-nine carotenoid pigments were separated and identified. The six largest carotenoid peaks (at 430 nm) were identified as auroxanthin A, mutatoxanthin A, mutatoxanthin B, lutein, zeaxanthin and isolutein.

Using PAD, researchers have detected beet sugar adulteration of orange juice.[106,107] Orange juice or BMIS were analyzed by LC on two columns of Carbo Pac PA1™ connected in series, with gradient elution with NaOH and sodium acetate with PAD. Results indicated the presence of several oligosaccharides in BMIS that were either in low concentration or nonexistent in orange juice;

determination of these carbohydrates allowed detection of as little as 5% of BMIS as an adulterant sugar in orange juice. A switching valve was suggested to enable simple sugars to go to waste after the first column and so prevent overloading of the column and detector for processing large numbers of samples. This modification reduces the analytical time by 50% without any sacrifice in information or detection limits.

Grapefruit juice has found its way into orange juice at low levels. It is generally much cheaper. For detecting grapefruit juice addition to orange juice, HPLC analysis of naringin is suggested.[108] Hesperidin is suggested to detect orange juice in grapefruit. For analysis, unfiltered juice was cleaned up on a polyamide column; elution was effected with methanol. The solution was evaporated to dryness, and the residue was taken up in 70% methanol or HPLC on a Hypersil™ ODS. Elution was carried out, either isocratically or with a gradient of methanol or acetonitrile in 1% acetic acid.

Pulp wash is another common adulterant of orange juice. Two approaches are reported for detecting it. The first is a little-used HPLC procedure where methoxylated flavonoids are used as markers of pulp wash.[109] The methoxylated flavonoids sinensetin, nobiletin, 3,3′,4′,5,6,7,8-heptamethoxyflavone, 4′,5,6,7-*tetra*-*O*-methylscutellarein and tangeretin in the samples were determined by HPLC on an ODS column with aqueous 40% acetonitrile as mobile phase and fluorescence detection with excitation at 330 nm and emission at 430 nm. An iterative program was used to compute the composition of the binary mixture from experimental results and known contents of flavonoids in original juices and aqueous extracts originating from different countries. A spectroscopic method is more commonly used but is not useful for mixed juice products.[110] Orange juice samples were analyzed for pulp wash and dilution by measuring the absorbance sum at 443, 325 and 280 nm and the 443–325 nm absorbance ratio, and by fluorimetry at 310–423 nm with excitation at 230–340 nm (details given). The effect of adulteration of samples by dilution or by addition of pulp wash on the measured parameters is significant.

Capillary zone electrophoresis measurements of the minor anions in orange juice and orange pulp wash were made.[111] Orange juice was diluted to 10° Brix and centrifuged. After filtration the anions were analyzed on an electrophoresis system with a fused silica column and an applied voltage of 20 kV. Range and mean results for orange juice, orange pulp wash and processing plant water could not be used to detect adulteration of orange juice with orange pulp wash.

Repeating another familiar pattern, isotope techniques have found use in authenticating orange juice. Site-specific isotopic content and stable isotope ratio have both found utility in detecting sugar addition.[112] In an alternative approach,[113] freeze-dried orange juice was mixed with hydrogen peroxide and centrifuged. The supernatant was mixed with trimethylsilylpropionic-2,2,3,3-d_4 sulfonic acid Na salt (internal standard) and ^{1}H NMR spectra were acquired at 400 MHz. The water signal was presaturated for 3 s with a low-power continuous-frequency radio wave. The free induction decays were multiplied with an exponential factor corresponding to 0.3 Hz in the Fourier-transformed spectrum and the spectra were normalized. The partial linear fit (PLF) data preprocessing tool was used to correct for variations in resonance positions between different spectra. With pattern recognition software, the method could discriminate between authentic and nonauthentic samples such as those to which beet sucrose or BMIS had been added.

Capillary GC analysis of oligosaccharides is another recurring theme that applies to orange juice.[114] This procedure is based upon the detection of traces of maltotriose (DP3) and maltotetraose (DP4) which are absent from the sugar profiles of citrus fruits. Samples were centrifuged, filtered and preconditioned on C_{18} Sep-Pak™ cartridges. The samples were evaporated to dryness and the residue was dissolved in water–acetonitrile (7 : 13). The DP3 and DP4 carbohydrates were separated by LC on an aminopropyl silica gel column with H_2O–acetonitrile (7 : 13) as mobile phase. The fractions containing DP3 and DP4 carbohydrates were evaporated to dryness and the sugars were converted into oxime-TMS derivatives and extracted into 2-mL isooctane which was injected on to DB-1 column. The detection limit was $\sim$1 mg L^{-1}. The method can be applied to citrus fruit juices adulterated with sugar syrups but not crystalline sugar products.

NIR shows promise as a fingerprinting method for detecting adulteration of orange juice.[115] Authentic and adulterated orange juice samples were vacuum dried at 45 °C and 5.3 KPa for 9 h, then manually ground and sieved to 300 μm. The dried samples were placed in cells (50-mm diameter × 10-mm depth) with a quartz side-window, and were scanned. The reflectance spectra recorded over 1100–2498 nm at 2-nm intervals. Application of principal component and factorial discriminant analysis to these spectra can detect the adulteration of orange juice with an average accuracy of 90%.

6.10 Other Juices

As a final commodity group, other juices were selected by the ease with which they can be adulterated. Fortunately, the simple adulterants are easily detected. As before,

there are many recurring themes. It is a matter of application and data.

As a first example in this last subsection, researchers have developed a procedure to compare the ^{13}C content of the sugars in a juice to that found in the fruit pulp.[116] This is an internal standard approach and greatly improves the sensitivity to detecting adulteration in fruit juices. Samples were centrifuged to yield a pulp and supernatant or sugar fraction. The fractions were combusted to produce carbon dioxide and isotopic analysis was performed.

Another example with infra-red seems appropriate to illustrate the growing trend to deal mathematically with the water rather than dry samples. In this example,[117] spectral measurements of raspberry purées prepared from fruit gathered in the 1993, 1994 and 1995 growing seasons were made using a FTIR spectrometer at 8-cm^{-1} resolution with 256 interferograms. Single beam ATR (attenuated total reflectance) spectra were collected for each sample, transformed to absorbance units using a background spectrum for water, and truncated to 235 data points in the region 899–1802 cm^{-1}. This method follows the trend in being dependent on chemometrics for interpretation.

Wine has been used as an example in several places. It seems appropriate to mention a current concern. Red wines can be adulterated with other anthocyanin pigments. One study outlines a simple procedure for detecting elderberry pigments.[118] For this procedure, red wine was extracted with a C_{18} Sep-Pak™ cartridge and washed with 1% formic acid. The anthocyanins were eluted with methanolic 1% formic acid and the eluant was concentrated under vacuum. The residue was dissolved and filtered before HPLC analysis on an ODS Hypersil™ column with gradient elution with aqueous 0.6% $HClO_4$–methanol (4 : 1 to 12 : 13 over 32 min) and detection at 520 nm. The anthocyanins, cyanidin 3-sambubioside-5-glucoside, cyanidin 3-glucoside and cyanidin 3-sambubioside, were present in the elderberry extract, with the latter two co-eluting. The presence of these peaks in red wine indicated that it had been adulterated with elderberry pigments.

Pineapple is one of the more important minor fruits. It has also been adulterated substantially from time to time. One of the big problems is obtaining juice with the correct balance of sugar to acid. Such problems are easily remedied with either sugar or acids. Pineapple is a CAM plant, which makes the stable isotope methods less discriminating because of the intermediate $\delta^{13}C$ values of pineapples. However, researchers have reported an HPLC procedure for the free sugars and detecting the addition of corn or cane sugars by detecting the presence of maltose.[119] For this procedure, samples were blended with methanol and the resulting slurry was heated in a magnetic mixer for 20 min and centrifuged. The solutions were evaporated to small volume, diluted with water and applied to a Sep-Pak™ C_{18} cartridge. Part of the filtrate was mixed with acetonitrile and the solution was filtered. Samples prepared in this manner were analyzed on a Bondapak/carbohydrate column with aqueous 75% acetonitrile as mobile phase and UV detection. In natural samples, the ratio of I/II was found to be ~1 and the maltose concentration was zero. The method was used for the analysis of suspected cases of adulteration of pineapple products.

Given the importance of the organic acids and titratable acidity to pineapple flavor, the addition of citric acid or malic acid has been a more common form of adulteration. HPLC has been used to measure these acids and can often detect adulteration.[120] After using a sample preparation procedure similar to the above procedure for pineapple sugars, samples were analyzed by HPLC on a C_{18} column with detection at 214 nm. The main compounds identified by HPLC were citric and L-malic acid, with smaller amounts of oxalic, quinic and succinic acids. For most samples, the sum of these acids was close to the total titratable acid and the ratio citric acid : malic acid was close to 2 : 1.

The last two references consider berries. Berries are expensive and therefore prone to adulteration with less expensive materials. In the first case,[121] adulteration of cranberry juice products by oenocyanin, a pigment from grapes, was detected by comparing HPLC profiles of cranberry anthocyanins with those from oenocyanin. The HPLC procedure used a reversed-phase styrene–divinylbenzene column in 10% acetic acid in water. Gradient elution was from 0% to 90% methanol–water–acetic acid (6 : 3 : 1). Both solvents were buffered to pH 1.6 with 3% H_3PO_4. Detection was at 530 nm.

In the second case,[122] blackcurrant adulteration by blackberries was detected by differences in the flavonol glycosides and flavonol glucuronides. Blackberries or blackcurrants were homogenized with methanol, filtered and concentrated. The residual solution was cleaned up on a polyamide column, from which flavonol glycosides were eluted with methanol, and flavonol glucuronides with methanolic 0.5% NH_3. Each eluate was concentrated and filtered before HPLC analysis on a Hypersil™ ODS column. Ion chromatography of the flavonol glycoside eluate was necessary if adulteration by redcurrants was suspected. Quercetin 3-*O*-β-glucuronide is absent from blackcurrants and is thus an indicator of adulteration by blackberries, an addition of 1% of blackberry juice being detectable. Quercetin 3-*O*-β-glucuronide is also present in grapes, but addition of grape juice can be identified by the detection of tartaric acid.

7 FUTURE TRENDS

The future of adulteration determination is likely to be a mix of more of the same and new methods. Adulteration is likely to be a concern as long as there are those who believe they can make extra money by providing deceitful alternative products. New methods based on profiling techniques are likely to become increasingly important as authenticators seek more discrimination power and lower costs. The world's growing population and shrinking global marketplace will force scientists, regulators and food processors to standardize definitions for various commodities. These definitions need to be statistically based. The key for their development is networks that generate and share data for authentic samples. Without this data, outliers are indistinguishable from adulterated samples. It is only a matter of time until a new fraud is exposed that embarrasses one or more major food companies.

The increasing demand for foods that contain additional phytochemicals and other nutrients offers a new frontier for the deceitful and for those who strive to maintain a fair and honest marketplace. Elixirs and tonics are again entering the marketplace. New methods are required. New databases need to be developed. It will be interesting to see where the resources are found to meet these needs. Technical expertise to make and propose products is moving much faster than the ability to verify and test. However, it is unlikely that we will see a return of the truly dangerous forms of adulteration seen in past centuries.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
ATR	Attenuated Total Reflectance
BMIS	Beet Medium Invert Sugar
BSA	Bovine Serum Albumin
CAM	Crassulacean Acid Metabolism
C-3	Calvin Cycle Plants
C-4	Hatch–Slack
DP3	Maltotriose
DP4	Maltotetraose
EDTA	Ethylenediaminetetraacetic Acid
FAIM	Food Authenticity – Issues and Methodologies
FDCA	Food, Drug and Cosmetic Act
FEMA	Flavor Extract Manufacturers' Association
FGM	Fat Globule Membrane
FID	Flame Ionization Detection
FT	Fourier Transform
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HFCS	High-fructose Corn Syrup
HIS	Hydrolyzed Inulin Syrup
HPLC	High-performance Liquid Chromatography
ICP/AES	Inductively Coupled Plasma/Atomic Emission Spectroscopy
IR	Infrared
LC	Liquid Chromatography
LDH	Lactate Dehydrogenase
MS	Mass Spectrometry
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
ODS	Octadecylsilane
PAD	Pulsed Ampherometic Detection
PCR	Polymerase Chain Reaction
PLF	Partial Linear Fit
PLS	Partial Least Squares
PPD	Polarized Photometric Detection
RSD	Root Square Deviation
RSK	German Richterwerte und Schwankungsbreiten Bestimmter Kennzahlen
SDS	Sodium Dodecylsulfate
SDS/PAGE	Sodium Dodecylsulfate/Polyacrylamide Gel Electrophoresis
SIRA	Stable Isotope Ratio Analysis
TCJJP	Technical Committee for Juice and Juice Products
THF	Tetrahydrofuran
TLC	Thin-layer Chromatography
TMS	Trimethylsilyl
UV	Ultraviolet

RELATED ARTICLES

Food **(Volume 5)**
Lipid Analyses in Food • Near-infrared Spectroscopy in Food Analysis

Chemometrics **(Volume 11)**
Chemometrics

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Isotope Ratio Mass Spectrometry

REFERENCES

1. The Holy Bible, Authorized King James Version, Isaiah 1 : 22.
2. M.J. Dennis, 'Establishing Food Authenticity', *Chem. Ind.*, **24**(Dec 15), 997–1001 (1997).
3. M.J. Dennis, 'Establishing Food Authenticity', *Chem. Ind.*, **24**(Dec 15), 997–1001 (1997).
4. F. Hallgarten, *Wine Scandal*, George Weidenfeld & Nicolson, London, 12, 1986.
5. O.R. Fennema, *Food Chemistry*, 3rd edition, Marcel Dekker, New York, 5, 1996.
6. P. Kurtzweil, 'Fake Food Fight', *FDA Consumer Magazine*, March–April (1999).
7. M.J. Dennis, 'Establishing Food Authenticity', *Chem. Ind.*, **24**(Dec 15), 997–1001 (1997).
8. B.T. Hunter, 'Detecting Adulterated Food', *Consumers' Res. Magazine*, **78**(12), 8–10 (1995).
9. F. Hallgarten, *Wine Scandal*, George Weidenfeld & Nicolson Limited, London, 32–41, 1986.
10. D.A. Kruegar, 'Stable Isotope Analysis', *Analytical Methods of Food Authentication*, eds. P.R. Ashurst, M.J. Dennis, Blackie Academic and Professional, 14–35, 1998.
11. V. Hong, R.E. Wrolstad, 'Detection of Adulteration in Commercial Cranberry Juice Drinks and Concentrates', *J. Assoc. Off. Anal. Chem.*, **69**, 208–213 (1986).
12. N.A. Low, D.A. Hammond, 'Detection of Inulin in Apple Juice by Capillary Gas Chromatography with Flame Ionization Detection', *Fruit Processing*, **4**, 135–139 (1996).
13. FAIM: Food Authenticity – Issues and Methodologies, European Commission-funded 'Concerted Action' No AIR3-CT94-2452, FAIM Euroconference, La Baule, France, June 4–6, 1997.
14. T.A. Eisele, 'Determination of D-malic Acid in Apple Juice by Liquid Chromatography: Collaborative Study', *J. AOAC Int.*, **79**(1), 50–54 (1996).
15. A. Yamamoto, A. Matsunaga, E. Mizukami, K. Hayakawa, M. Miyazaki, 'Enantiomeric Purity Determination by High-performance Liquid Chromatography with Coupled Polarized Photometric/UV Detection. Analysis of Adulterative Addition [to Fruit Juices] of Synthetic Malic and Tartaric Acids', *J. Chromatogr. A*, **667**(1–2), 85–89 (1994).
16. E. Jamin, J. Gonzalez, G. Remaud, N. Naulet, G.G. Martin, 'Detection of Exogenous Sugars or Organic Acids Addition in Pineapple Juices and Concentrates by ^{13}C Isotope Ratio (IR) MS Analysis', *J. Agric. Food. Chem.*, **45**(10), 3961–3967 (1997).
17. J.W. White, K. Winters, 'Honey Protein as Internal Standard for Stable Carbon Isotope Ratio Detection of Adulteration of Honey', *J. Assoc. Off. Anal. Chem.*, **72**(6), 907–911 (1989).
18. E. Jamin, J. Gonzalez, G. Remaud, N. Naulet, G.G. Martin, D. Weber, A. Rossmann, H.L. Schmidt, 'Improved Detection of Sugar Addition to Apple Juices and Concentrates Using Internal Standard Carbon-13 IRMS', *Anal. Chim. Acta.*, **347**(3), 359–368 (1997).
19. A.R. Brause, J.M. Raterman, L.W. Doner, E.C. Hill, 'Detection of Adulteration in Apple Juice and Orange Juice by a Chemical Matrix Method', *Fluess-Obst.*, **53**(1), 15–16, 21–23 (1986).
20. N.A. Low, D.A. Hammond, 'Detection of Inulin in Apple Juice by Capillary Gas Chromatography with Flame Ionization Detection', *Fruit Processing*, **4**, 135–139 (1996).
21. G. Downey, 'Authentication of Food and Food Ingredients by Near Infrared Spectroscopy', *J. Near-infrared Spectrosc.*, **4**(1–4), 47–61 (1996).
22. M. Lipp, 'Determination of the Adulteration of Butter Fat by its Triglyceride Composition Obtained by GC. A Comparison of the Suitability of PLS and Neural Networks', *Food-Chem.*, **55**(4), 389–395 (1996).
23. G.S. Remaud, G.S. Martin, Y.L. Martin, G.J. Martin, 'Detection of Sophisticated Adulteration of Natural Vanilla Flavors and Extracts: Application of the SNIF–NMR Method to Vanillin and *p*-hydroxybenzaldehyde', *J. Agric. Food. Chem.*, **45**(3), 859–866 (1997).
24. P.R. Ashurst, M.J. Dennis, *Analytical Methods of Food Authentication*, Blackie Academic and Professional, London, 14–35, 1998.
25. N.H. Low, 'Using GC to Detect Food Adulteration', *Hewlett Packard Peak*, (2), 2–4 (1997).
26. J.R. Elkins, J.R. Heuser, 'Detection of Adulteration in Apple Juice by L-malic/Total Malic Acid Ratio: Collaborative Study', *J. AOAC Int.*, **77**(2), 411–415 (1994).
27. L.S. Pilando, R.E. Wrolstad, 'Effectiveness of Pattern Recognition, Sugar, Non-volatile Acid, and Carbon-13/Carbon-12 Analyses for Detecting Adulteration in Apple Juice', *J. Food. Compos. Anal.*, **5**(1), 10–24 (1992).
28. B. Wald, R. Galensa, 'Detection of Adulteration of Apple Juice and Pear Juice', *Z. Lebensm. Unters. Forsch.*, **188**(2), 107–114 (1989).
29. S.J. Lai, S.R. Binder, H. Essien, K.C. Wen, 'Identification of Western Medicines as Adulterants in Chinese Herbal Medicines Using a Broad-spectrum Drug Screening HPLC System', *J. Liq. Chromatogr.*, **18**(14), 2861–2875 (1995).
30. H.J. Bouwmeester, J.A.R. Davies, H. Toxopeus, 'Enantiomeric Composition of Carvone, Limonene, and Carveols in Seeds of Dill and Annual and Biennial Caraway Varieties', *J. Agric. Food. Chem.*, **43**(12), 3057–3064 (1995).
31. J.H.Y. Viegas, F.M. Lancas, N.R. Antoniosi, 'High Temperature Capillary GC Analysis of Phytopreparations of '*Espinheira Santa*' (*Maytenus ilicifolia* M. and *Maytenus*

aquifolium M.–Celastraceae), a Brazilian Antiulcer Plant', *Chromatographia.*, **40**(5–6), 341–344 (1995).
32. L. Sagrero-Nieves, J.P. Bartley, A. Provis-Schwede, 'Supercritical-fluid Extraction of the Volatile Components from the Leaves of *Psidium guajava* L. (Guava)', *Flavor Fragrance J.*, **9**(3), 135–137 (1994).
33. R. Oberdieck, 'Contribution to the Knowledge and Analysis of Saffron (*Crocus sativus* L.)', *Dtsch. Lebensm. Rundsch.*, **87**(8), 246–252 (1991).
34. C.L. Curl, G.R. Fenwick, 'On the Determination of Papaya Seed Adulteration of Black Pepper', *Food. Chem.*, **12**(4), 241–247 (1983).
35. G. Downey, R. Briandet, R.H. Wilson, E.K. Kemsley, 'Near- and Mid-infrared Spectroscopies in Food Authentication: Coffee Varietal Identification', *J. Agric. Food. Chem.*, **45**(11), 4357–4361 (1997).
36. R. Briandet, E.K. Kemsly, R.H. Wilson, 'Approaches to Adulteration Detection in Instant Coffees Using Infrared Spectroscopy and Chemometrics', *J. Sci. Food. Agric.*, **71**(3), 359–366 (1996).
37. G. Downey, J. Boussion, 'Authentication of Coffee Bean Variety by Near-infrared Reflectance Spectroscopy of Dried Extract', *J. Sci. Food. Agric.*, **71**(1), 41–49 (1996).
38. E.K. Kemsely, S. Rualt, R.H. Wilson, 'Discrimination Between *Coffea arabica* and *Coffea canephora* Variant *robusta* Beans Using Infrared Spectroscopy', *Food Chem.*, **54**(3), 321–326 (1995).
39. J. Prodolliet, M. Bruelhart, M.B. Blanc, V. Leloup, G. Cherix, C.M. Donnelly, R. Viani, 'Adulteration of Soluble Coffee with Coffee Husks and Parchments', *J. AOAC. Int.*, **78**(3), 761–767 (1995).
40. A. Berger, M. Bruelhart, J. Prodolliet, 'Detection of Adulteration in Pure Soluble Coffee by Enzymic Sugar Determination', *Lebensm.–Wiss. Technol.*, **24**(1), 59–62 (1991).
41. F. Lutman, 'Detection of Adulteration of Roasted and Ground Coffee with Barley', *Boll. Chim. Unione Ital. Lab. Prov., Parte. Sci.*, **33**(S4–S5), 573–580 (1982).
42. D. Danho, N. Naulet, G.J. Martin, 'Deuterium, Carbon and Nitrogen Isotopic Analysis of Natural and Synthetic Caffeines. Authentication of Coffees and Coffee Extracts', *Analusis*, **20**(3), 179–184 (1982).
43. A. Pena, L. Nevarez, C. Labastida, S. Capella, 'Triglyceride Characterization of Different Mexican Cheeses by High-temperature Capillary Gas Chromatography (HT-CGC)', *J. High Resolut. Chromatogr.*, **20**(10), 572–574 (1997).
44. C. Mariani, S. Venturini, E. Fedeli, G. Contarini, 'Detection of Refined Animal and Vegetable Fats in Adulteration of Pure Milkfat', *J. Am. Oil. Chem. Soc.*, **71**(12), 381–384 (1994).
45. J.L. Iverson, A.J. Sheppard, 'Detection of Adulteration in Cow, Goat and Sheep Cheeses Utilizing Gas–Liquid Chromatographic Fatty Acid Data', *J. Dairy Sci.*, **72**(7), 1707–1712 (1989).
46. J.R. Fox, A.H. Duthie, S. Wulff, 'Effect of the Seasonal Variation of Fatty Acids in Milk Fat on the Sensitivity of a Test for Vegetable Fat Adulteration', *J. Dairy Sci.*, **72**(8), 1981–1985 (1989).
47. M. Lipp, 'Review of Methods for the Analysis of Triglycerides in Milk Fat: Application for Studies of Milk Quality and Adulteration', *Food Chem.*, **54**(2), 213–221 (1995).
48. M. Collomb, M. Spahni, 'Adulteration of Milk Products. Review of the Analytical Criteria for the Detection of Vegetable and Animal Fats in Milk Fat', *Mitt. Geb. Lebensmittelunters. Hyg.*, **82**(6), 615–662 (1991).
49. E. Coni, M. Di-Pasquale, P. Coppolelli, A. Bocca, 'Detection of Animal Fats in Butter by Differential Scanning Calorimetry: a Pilot Study', *J. Am. Oil. Chem. Soc.*, **71**(8), 807–810 (1994).
50. F. Harding, 'Milk Adulteration – Freezing Point Depression', *J. Soc. Dairy Technol.*, **43**(3), 61 (1990).
51. R. Goodacre, 'Use of Pyrolysis Mass Spectrometry with Supervised Learning for the Assessment of the Adulteration of Milk of Different Species', *Appl. Spectrosc.*, **51**(8), 1144–1153 (1997).
52. E.L. Malin, J.J. Basch, J.J. Shieh, B.C. Sullivan, V.H. Holsinger, 'Detection of Adulteration of Buttermilk Powder by Gel Electrophoresis', *J. Dairy Sci.*, **77**(8), 2199–2206 (1994).
53. N. De-Jong, S. Visser, C. Olieman, 'Determination of Milk Proteins by Capillary Electrophoresis', *J. Chromatogr. A.*, **652**(1), 207–213 (1993).
54. F.T.A. Chen, J.H. Zang, 'Determination of Milk Proteins by Capillary Electrophoresis', *J. AOAC Int.*, **75**(5), 905–909 (1992).
55. M.M. Furtado, 'Detection of Cow's Milk in Goat's Milk by Polyacrylamide Gel Electrophoresis', *J. Dairy Sci.*, **66**(9), 1822–1824 (1983).
56. W. Urbanke, W. Luf, E. Brandl, 'Use of HPLC for Control of the Adulteration of Milk and Milk Products of Different Species', *Z. Lebensm. Unters. Forsch.*, **195**(2), 137–142 (1992).
57. S.D. Hill, R.L. Richter, C.W. Dill, 'Sulfur-based Method to Detect Adulteration of Non-fat Dry Milk with Whey', *Cult. Dairy Prod. J.*, **23**(1), 14–18 (1988).
58. R. Greenberg, H.J. Dower, 'Detection of Added Whey Protein Concentrate in Non-fat Dry Milk by Amino-acid Analysis', *J. Agric. Food. Chem.*, **34**(1), 30–32 (1986).
59. W.A. Konig, C. Fricke, Y. Saritas, B. Momeni, G. Hohenfled, 'Adulteration or Natural Variability? Enantioselective Gas Chromatography in Purity Control of Essential Oils', *J. High Resolut. Chromatogr.*, **20**(2), 55–61 (1997).
60. O. Phillip, H.D. Isengard, 'A New Method for Authenticity Control of Lemon Oils by HPLC', *Z. Lebensm. Unters. Forsch.*, **201**(6), 551–554 (1995).
61. K.A. Krock, N. Ragunathan, C.L. Wilkins, 'Multi-dimensional Gas Chromatography Coupled with Infrared

and Mass Spectrometry for Analysis of Eucalyptus Essential Oils', *Anal. Chem.*, **66**(4), 425–430 (1994).
62. P. Kreis, A. Mosandl, 'Chiral Compounds of Essential Oils. XII. Authenticity Control of Rose Oils, Using Enantioselective Multi-dimensional Gas Chromatography', *Flavor Fragrance J.*, **7**(4), 199–203 (1992).
63. P. Kreis, A. Mosandl, H.G. Schmarr, 'Mentha Oils. Enantioselective Analysis of Menthyl Acetate in Mentha Oil Quality Evaluation', *Dtsch. Apoth. Zeitung.*, **130**(47), 2579–2581 (1990).
64. S. Zhang, Y. Lan, X. Qin, 'Gas Chromatographic Analysis for Volatile Oil in Imported Cardamom Seed (*Amomum Xanthioides*) and Adulterants (Hong Ke Sha, Cao Dou Kou, Hong Dou Kou, Zhu Mu Sha)', *Yaowu Fenxi Zazhi*, **9**(4), 219–222 (1989).
65. Analytical Methods Committee Royal Society of Chemistry, 'Application of Gas–Liquid Chromatography to the Analysis of Essential Oils. XI. Monographs for Seven Essential Oils', *Analyst*, **109**(10), 1343–1360 (1984).
66. M.J. Prager, M.A. Miskeiwicz, 'Gas Chromatographic–Mass Spectrometric Analysis, Identification and Detection of Adulteration of Natural and Concentrated Lemon Oils', *J. Assoc. Off. Anal. Chem.*, **65**(1), 166–171 (1982).
67. M.J. Prager, M.A. Miskeiwicz, 'Gas Chromatographic–Mass Spectrometric Analysis, Identification and Detection of Adulteration of Perfumery Products from Bitter-orange Trees', *J. Assoc. Off. Anal. Chem.*, **64**(1), 131–138 (1981).
68. G. Remaud, A.A. Debon, Y.I. Martin, G.G. Martin, G.J. Martin, 'Authentication of Bitter Almond Oil and Cinnamon Oil: Application of the SNIF-NMR Method to Benzaldehyde', *J. Agric. Food. Chem.*, **45**(10), 4042–4048 (1997).
69. G.S. Remaud, Y.L. Martin, G.G. Martin, N. Naulet, G.J. Martin, 'Authentication of Mustard Oils by Combined Stable Isotope Analysis (Site-specific Natural Isotope Fractionation–NMR and Isotope Ratio MS)', *J. Agric. Food. Chem.*, **45**(5), 1844–1848 (1997).
70. G.G. Martin, Y.L. Martin, N. Naulet, H.J.D. McManus, 'Application of Deuterium SNIF–NMR and Carbon-13 SIRA–MS Analyses to Maple Syrup: Detection of Added Sugars', *J. Agric. Food. Chem.*, **44**(10), 3206–3213 (1996).
71. M.F. Morselli, K.L. Baggett, 'Mass Spectrometric Determination of Cane Sugar and Corn Syrup in Maple Syrup by Use of Carbon-13–Carbon-12 Ratio: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **67**(1), 22–24 (1984).
72. L.R. Croft, 'Stable-isotope Mass Spectrometry in Honey Analysis', *Trends Anal. Chem.*, **6**(8), 206–209 (1987).
73. J.G. Stuckel, N.H. Low, 'Maple Syrup Authenticity Analysis by Anion-exchange Liquid Chromatography with Pulsed Amperometric Detection', *J. Agric. Food. Chem.*, **43**(12), 3046–3051 (1995).
74. J. Lipp, H. Zieglar, E. Conrady, 'Detection of High Fructose and Other Syrups in Honey Using High-pressure Liquid Chromatography', *Z. Lebensm. Unters. Forsch.*, **187**(4), 334–338 (1988).
75. J.W. White, R.W. Meloy, J.L. Probst, W.F. Huser, 'Detection of Beet Sugar Adulteration of Honey', *J. Assoc. Off. Anal. Chem.*, **69**(4), 652–654 (1986).
76. N.H. Low, W. South, 'Determination of Honey Authenticity by Capillary Gas Chromatography', *J. AOAC Int.*, **78**(5), 1210–1218 (1995).
77. F. Ferreres, F.A. Tomas-Barberan, M.I. Gil, F. Tomas-Loente, 'HPLC Technique for Flavonoid Analysis in Honey', *J. Sci. Food. Agric.*, **56**(1), 49–56 (1991).
78. F. Ferreres, P. Andrade, F.A. Tomas-Barberan, 'Flavonoids from Portuguese Heather Honey', *Z. Lebensm. Unters. Forsch.*, **199**(1), 32–37 (1994).
79. Y. Saisekhar, P.M. Reddy, 'Use of Troponin for Species Identification of Cattle and Buffalo Meats', *J. Food. Sci. Technol.*, **32**(1), 68–70 (1995).
80. C.H. Wang, D. M. Smith, 'Lactate Dehydrogenase Monoclonal Antibody Immunoassay for Detection of Turkey Meat in Beef and Pork', *J. Food. Sci.*, **60**(2), 253–256 (1995).
81. Anon., 'Beef, Horse or Kangaroo', *Meat. Hyg.*, 5427–5428 (1987).
82. R.M. Patterson, R.G. Whittaker, T.L. Spencer, 'Improved Species Identification of Raw Meat by Double Sandwich Enzyme-linked Immunosorbent Assay', *J. Sci. Food. Agric.*, **35**(9), 1018–1023 (1984).
83. R.G. Whitaker, T.L. Spencer, J.W. Copland, 'Enzyme-linked Immunosorbent Assay for Species Identification of Raw Meat', *J. Sci. Food. Agric.*, **34**(10), 1143–1148 (1983).
84. A. Craig, A.H. Ritchie, I.M. Makie, 'Determining the Authenticity of Raw Reformed Breaded Scampi (*Nephrops norvegicus*) by Electrophoretic Techniques', *Food. Chem.*, **52**(4), 451–454 (1995).
85. A. Zerifi, C. Labie, G. Benard, 'SDS [Sodium Dodecylsulfate] – PAGE [Polyacrylamide Gel Electrophoresis] Technique for the Species Identification of Cooked Meat', *Fleischwirtschaft*, **71**(9), 1060–1062 (1991).
86. M. Bonnefoi, G. Benard, C. Labie, 'Gel Electrophoresis: a Qualitative Method for Detection of Duck and Goose Liver in Canned Foie Gras', *J. Food. Sci.*, **51**(5), 1362–1363 (1986).
87. S.H. Ashoor, M.A. Osman, 'Liquid-chromatographic Quantitation of Chicken and Turkey in Unheated Chicken–Turkey Mixtures', *J. Assoc. Off. Anal. Chem.*, **71**(2), 403–405 (1988).
88. L. Matter, 'Identification of Animal Species: [Detection of] Pork in Beef Products by High-resolution Capillary Gas Chromatographic Analysis of Fatty Acids', *GIT Fachz. Lab.*, **34**(5), 664 (1990).
89. S. Scalia, M. Isaksen, G.W. Francis, 'Improved Thin-layer Chromatographic Method for the Detection of Canthaxanthin in Salmonidae', *Arch. Lebensmittelhyg.*, **40**(6), 130–132 (1989).

90. V. Baeten, M. Meurens, M.T. Morales, R. Aparicio, 'Detection of Virgin Olive Oil Adulteration by Fourier-transform Raman Spectroscopy,' *J. Agric. Food. Chem.*, **44**(8), 2225–2230 (1996).
91. I.J. Wesley, F. Pacheo, A.E.J. McGill, 'Identification of Adulterants in Olive Oils', *J. Am. Oil Chem. Soc.*, **73**(4), 515–518 (1996).
92. I.J. Wesley, R.J. Barnes, A.E.J. McGill, 'Measurement of Adulteration of Olive Oils by Near-infrared Spectroscopy', *J. Am. Oil Chem. Soc.*, **72**(3), 289–292 (1995).
93. Y.W. Lai, E.K. Kemsley, R.H. Wilson, 'Potential of Fourier Transform Infrared Spectroscopy for the Authentication of Vegetable Oils', *J. Agric. Food. Chem.*, **42**(5), 1154–1159 (1994).
94. R. Sacchi, M. Patumi, G. Fontanazza, P. Barone, P. Fiordiponti, L. Mannina, E. Rossi, A.L. Segre, 'A High-field Proton Nuclear Magnetic Resonance Study of the Minor Components in Virgin Olive Oils', *J. Am. Oil. Chem. Soc.*, **73**(6), 747–758 (1996).
95. F. Dionisi, J. Prodolliet, E. Tagliaferri, 'Assessment of Olive Oil Adulteration by Reversed-phase High-performance Liquid Chromatography–amperometric Detection of Tocopherols and Tocotrienols', *J. Am. Oil. Chem. Soc.*, **72**(12), 1505–1511 (1995).
96. A.H. El-Hamdy, N.K. El-Fizga, 'Detection of Olive Oil Adulteration by Measuring its Authenticity Factor Using Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. A.*, **708**(2), 351–355 (1995).
97. R.V. Flor, Le-Tiet-Hecking, B. D. Martin, 'Development of High-performance Liquid Chromatography Criteria for Determination of Grades of Commercial Olive Oils. I. Normal Ranges for the Triacylglycerols', *J. Am. Oil Chem. Soc.*, **70**(2), 199–203 (1993).
98. E. Salivaras, A.R. McCurdy, 'Detection of Olive Oil Adulteration with Canola Oil from Triacylglycerol Analysis by Reversed-phase High-performance Liquid Chromatography', *J. Am. Oil Chem. Soc.*, **69**(9), 935–938 (1992).
99. M. Tsimidou, R. Macrae, 'Detection and Quantitative Determination of Adulteration of Olive Oil: Fractional Crystallization as a Preliminary Stage Prior to HPLC', *Int. Anal.*, **1**(2), 29, 31–32, 34 (1987).
100. R. Goodacre, D.B. Kell, G. Bianchi, 'Food Adulteration Exposed by Neural Networks', *Analysis Europa*, **5**, 35–37 (1995).
101. S. E. Woodbury, R.P. Evershed, J.B. Griffith, P. Farnell, 'Detection of Vegetable Oil Adulteration Using Gas Chromatography Combustion–Isotope Ratio Mass Spectrometry', *Anal. Chem.*, **67**(15), 2685–2690 (1995).
102. M. Tsimidou, R. Macrae, I. Wilson, 'Authentication of Virgin Olive Oils Using Principal Component Analysis of Triglyceride and Fatty Acid Profiles. II. Detection of Adulteration with Other Vegetable Oils.', *Food. Chem.*, **25**(4), 251–258 (1987).
103. W.W. Widmer, P.F. Cancalon, S. Nagy, 'Methods for Determining the Adulteration of Citrus Juices', *Trends Food. Sci. Technol.*, **3**(11), 278–286 (1992).
104. R. Goodacre, D. Hammond, D.B. Kell, 'Quantitative Analysis of the Adulteration of Orange Juice with Sucrose Using Pyrolysis Mass Spectrometry and Chemometrics', *J. Anal. Appl. Pyrolysis*, **40**, 41 135–41 158 (1997).
105. R. Rouseff, L. Raley, H.J. Hofsommer, 'Application of Diode Array Detection with a C-30 Reversed Phase Column for the Separation and Identification of Saponified Orange Juice Carotenoids', *J. Agric. Food. Chem.*, **44**(8), 2176–2181 (1996).
106. K.W. Swallow, N.H. Low, D.R. Petrus, 'Detection of Orange Juice Adulteration with Beet Medium Invert Sugar Using Anion-exchange Liquid Chromatography with Pulsed Amperometric Detection', *J. Assoc. Off. Anal. Chem.*, **74**(2), 341–345 (1991).
107. D.R. White, Jr, P.F. Cancalon, 'Detection of Beet Sugar Adulteration of Orange Juice by Liquid Chromatography–Pulsed Amperometric Detection with Column Switching', *J. AOAC. Int.*, **75**(3), 584–587 (1992).
108. F. Siewek, R. Galensa, V. Ara, 'High-performance Liquid Chromatographic Detection of Adulteration of Orange and Grapefruit Juice', *Ind. Obst. Gemueseverwert.*, **70**(1), 11–12 (1985).
109. B. Heimhuber, R. Galensa, 'Method for the Analysis of Mixtures of Various Citrus Fruit Juices', *Lebensmittelchemie*, **45**(2), 36–37 (1991).
110. D.R. Petrus, J.A. Attaway, 'Spectral Characteristics of Florida Orange Juice and Orange Pulpwash. Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **68**(6), 1202–1206 (1985).
111. K.W. Swallow, N.H. Low, 'Capillary Zone Electrophoretic Analysis of the Minor Anions Present in Orange Juice and Orange Pulpwash', *J. Agric. Food. Chem.*, **42**(12), 2808–2811 (1994).
112. G.G. Martin, V. Hanote, M. Lees, Y.L. Martin, 'Interpretation of Combined ^{2}H SNIF–NMR and ^{13}C SIRA–MS Analysis of Fruit Juices to Detect Added Sugar', *J. AOAC Int.*, **79**(1), 62–72 (1996).
113. J.T.W.E. Vogels, L. Terwel, A.C. Tas, F. Van den Berg, F. Dukel, J. Van der Greef, 'Detection of Adulteration in Orange Juices by a New Screening Method Using Proton NMR Spectroscopy in Combination with Pattern Recognition,' *J. Agric. Food Chem.*, **44**(1), 175–180 (1996).
114. V. Cornell, W. Ooghe, P. Sandra, 'Detection of Sugar Addition to Citrus Juices by Capillary GC', *J. High Resolut. Chromatogr.*, **18**(5), 286–288 (1995).
115. M. Twomey, G. Downey, P.B. McNulty, 'The Potential of NIR Spectroscopy for the Detection of the Adulteration of Orange Juice', *J. Sci. Food Agric.*, **67**(1), 77–84 (1995).
116. A. Rossman, J. Koziet, G.J. Martin, M.J. Dennis, 'Determination of the Carbon-13 Content of Sugars and Pulp from Fruit Juices by Isotope-ratio Mass Spectrometry

(Internal Reference Method). A European Interlaboratory Comparison', *Anal. Chim. Acta*, **340**(1–3), 21–29 (1997).

117. E.K. Kemsley, J.K. Holland, M. Defernez, R.H. Wilson, 'Detection of Adulteration of Raspberry Purees Using Infrared Spectroscopy and Chemometrics', *J. Agric. Food. Chem.*, **44**(12), 3864–3870 (1996).
118. P. Bridle, C. Garcia Viguera, 'A Simple Technique for the Detection of Red Wine Adulteration with Elderberry Pigments', *Food. Chem.*, **55**(2), 111–113 (1996).
119. M. Montana Camara, C. Diez, M. Esperanza-Torija, 'Free Sugars Determination by HPLC in Pineapple Products', *Z. Lebensm. Unters. Forsch.*, **202**(3), 233–237 (1996).
120. M.M. Camara, C. Diez, M.E. Torija, M.P. Cano, 'HPLC Determination of Organic Acids in Pineapple Juices and Nectars', *Z. Lebensm. Unters. Forsch*, **198**(1), 52–56 (1994).
121. M.L. Hale, F.J. Francis, I.S. Fagerson, 'Detection of Oenocyanin in Cranberry Juice Cocktail by HPLC Anthocyanin Profile', *J. Food. Sci.*, **51**(6), 1511–1513 (1986).
122. B. Wald, R. Galensa, K. Herrmann, 'Detection of Adulteration of Blackcurrant Products by Blackberry Juice by Determination of Flavonoids by HPLC', *Fluess Obst.*, **53**(7), 349–353 (1986).

Atomic Spectroscopy in Food Analysis

Scott A. Baker and Nancy J. Miller-Ihli
US Department of Agriculture, Beltsville, USA

1 Introduction **1**
2 Sampling and Sample Preparation Methods for Foods **2**
2.1 Sample Collection and Processing 2
2.2 Sample Preparation 3
3 Measurement Techniques **7**
3.1 Atomic Absorption Spectrometry 8
3.2 Inductively Coupled Plasma Atomic Emission Spectrometry 10
3.3 Inductively Coupled Plasma Mass Spectrometry 11
4 Food Composition Laboratory Methods **13**
4.1 Sample Preparation 13
4.2 Flame Atomic Absorption Spectrometry 14
4.3 Graphite Furnace Atomic Absorption Spectrometry 14
4.4 Inductively Coupled Plasma Atomic Emission Spectrometry 14
4.5 Inductively Coupled Plasma Mass Spectrometry 15
5 Speciation **16**
6 Quality Assurance, Method Validation, and Reference Materials **17**
Disclaimer **21**
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

Since food is the primary source of essential elements for humans, the accurate and precise analysis of food materials is critical. The methods best suited to meet this task are atomic spectroscopic methods such as atomic absorption, atomic emission, and elemental mass spectrometry (MS). Methods commonly used in the generation of food composition data include flame atomic absorption spectrometry (FAAS), graphite furnace atomic absorption spectrometry (GFAAS), inductively coupled plasma atomic emission spectrometry (ICP/AES), and more recently, inductively coupled plasma mass spectrometry (ICP/MS). FAAS and ICP/AES offer similar detection limits (DLs) (ng mL^{-1} levels), whereas, GFAAS and ICP/MS can provide sub-ng mL^{-1} detection capability. The choice of a method often depends on detection capability, but ease of use, speed of analysis, and cost must be considered. ICP/AES and ICP/MS offer the advantage of providing simultaneous multielement measurements, making them well suited for the analysis of large numbers of elements in food samples. If one or a few elements are to be determined, less expensive atomic absorption spectrometry (AAS) methods might be more suitable. The most critical stage in the development of analytical methods is sample preparation. Samples can be prepared using numerous procedures, but the most useful for a wide range of analytes and sample matrices are based on dry ashing or wet ashing of the sample. In addition to total element determinations, speciation measurements are important to determine the exact form of the element that is present in the sample. Since the bioavailability of an element is dependent on its chemical form (speciation), the development of reliable methods for identification and quantification of trace element species is critical. The coupling of chromatographic separation and atomic spectroscopic detection has proven useful for elemental speciation measurements.

1 INTRODUCTION

Since food is the primary source of essential elements for humans, the accurate and precise analysis of food materials is critical for consumers and health care professionals to make choices and recommendations based on the nutrient content of foods. Elements of interest include both major elements (e.g. calcium, sodium, potassium, magnesium, phosphorus, and chloride) and minor elements (e.g. iron, chromium, manganese, copper, zinc, cobalt, iodine, selenium, and molybdenum). The National Research Council has established the recommended dietary allowance (RDA) for seven elements (calcium, phosphorus, magnesium, iron, zinc, iodine, and selenium).[1] Estimated safe and adequate daily dietary intakes have also been issued for five additional elements (copper, manganese, fluoride, chromium, and molybdenum) and estimates of requirement for three electrolytes (sodium, potassium, and chloride). Several additional trace elements are considered essential (arsenic, nickel, silicon, boron, and cobalt) or potentially essential (cadmium, lead, lithium, tin, and vanadium) but required levels have not been quantified.[1]

In an attempt to provide nutritional information to consumers, the Nutrition Labeling and Education Act (NLEA) of 1990 mandated, for the first time, that labels on almost all processed food must contain the levels of 14 nutrients.[2] Three of the mandatory nutrients

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

are elements (calcium, iron and sodium). Thirty-four additional nutrients, 12 of which are elemental (potassium, phosphorus, magnesium, zinc, iodine, selenium, copper, manganese, fluoride, chromium, molybdenum, and chloride), which affect human health may currently be labeled voluntarily.

A review of the state of knowledge of food composition data demonstrates that there is a significant lack of data for specific food groups and for several of the elements of interest. This is due largely to the lack of suitable analytical methodology that can provide rugged, accurate elemental analyses of food matrices and the cost required to do large-scale sampling to obtain representative values. Because of mounting interest among health professionals and the public in the maintenance or promotion of good health and disease prevention, rather than diagnosis and curing of disease, reliable analytical data are increasingly important. This information can be used to assess the nutrient content of consumed foods and also for the establishment of the essentiality of elements that could play a role in the promotion of good health. The generation and compilation (such as in the US Department of Agriculture's Nutrient Database for Standard Reference, SR-12) of reliable analytical data is critical as knowledge of the role of trace elements in human health is expanding. In addition, the role of trace element speciation in relation to absorption and metabolism has stirred greater interest in the development of techniques capable of providing not only total elemental content of foods, but also species-specific information. In many cases, speciation information is of primary importance, since only selected forms of the element are biologically active and absorption can be significantly affected depending on the chemical form.

The methods most suitable for the rapid and accurate determination of the elemental content of foods are atomic spectroscopic methods such as AAS and atomic emission spectrometry (AES). Challenges in the determination of metals and minerals in foods include the wide range of concentrations present, ranging from ng g^{-1} levels (e.g. chromium and cobalt) to percent levels (e.g. sodium and potassium), combined with the analyte and matrix specificity of many official methods. In addition, interest in identifying the essentiality of low-level trace elements in foods demands very sensitive and selective analytical methods. Methods commonly used in the generation of food composition data include FAAS, GFAAS, flame atomic emission spectrometry (FAES), ICP/AES, and more recently, ICP/MS. Although FAES can be and often is successfully used for the determination of sodium and potassium in foods, it is not discussed in this contribution because of its relatively limited applicability combined with the fact that these elements are also easily measured with more robust multielement techniques, such as ICP/AES and ICP/MS.

In this overview on the use of atomic spectroscopy in the analysis of foods, the characteristics and applicability of the various analytical methods will be discussed. The strengths and limitation of the various techniques will be highlighted and recently reported applications will be featured. In addition, sample preparation considerations will be discussed since this is often the most critical part of the complete analytical method. A section is also dedicated to trace element speciation because of the importance of this field in nutritional science, and also because of the many innovative methods and interesting applications that have been reported in recent years. The importance of quality assurance (QA) and the preparation/characterization of reference materials (RMs) will also be discussed. This is because successful implementation of new technology and new analytical methods depends largely on the effective use of QA quality control (QC) strategies, which, in turn, rely on the development and availability of well-characterized RMs.

2 SAMPLING AND SAMPLE PREPARATION METHODS FOR FOODS

2.1 Sample Collection and Processing

The suitability of the sample provided for analysis is an often-overlooked aspect in analytical determinations. Because proper selection and treatment of food samples is crucial, attention given to sampling must be at least as great as that given to the analyses. Too frequently, costly and time-consuming analyses are carried out on samples that have been improperly collected or handled, leading to suspect data and a waste of valuable resources.[3] For a detailed discussion on sampling as it relates to the production of food composition data, the reader is referred to an excellent book by Greenfield and Southgate.[3]

It is important that careful consideration be given to what information is to be provided by the analysis. For example, is it important to gain information on the variability of trace element content of food due to the season, climate, location, etc. or is information on the average content of the food (e.g. for food label values) of primary interest? Often work completed at the Food Composition Laboratory (FCL) (US Department of Agriculture, Beltsville Human Nutrition Research Center, Beltsville, MD, USA) is in support of the Nutrient Data Laboratory, which is responsible for publication of the US Department of Agriculture's Nutrient Database for Standard Reference (SR-12). This publication serves as the primary database for US food composition

data; therefore, the goal is to include data that are representative of the national food supply. Because of the expense of large-scale national samplings, a feasible and cost-effective alternative is to pick up samples from several different locations across the country with brands selected based on frequency of consumption, in order to gain as representative a sample as possible. Samples are brought into the laboratory and, depending on the form of the sample and the information desired, a decision is made either to analyze multiple samples or to combine samples and remove separate aliquots for analysis. The complexity and homogeneity of the sample will dictate what steps are needed to ensure that a representative sample is provided for analysis. Most often, a preliminary evaluation of the homogeneity of a select number of analytes in the material serves as a basis for selection of sample size and number of replicate samples that must be analyzed. Analyte concentration can be reported on an as received (fresh weight) or dry weight (moisture corrected) basis. To obtain the concentration on a dry weight basis, the measured concentration of a moisture-containing sample is divided by one minus the sample moisture. For example, if the sample moisture is 10%, the measured concentration is divided by 0.9 (1 − 0.1) to obtain the concentration on a dry weight basis. Moisture values are typically determined by drying samples (e.g. in a vacuum oven) until a constant weight is obtained.

Tools are generally available that facilitate sample homogenization without trace element contamination. Otherwise, specialized tools can often be custom produced. For cutting of food samples, titanium or high-purity glass knives can be used. Large-scale mixing devices (e.g. Robot-Coupe, Jackson, MS, USA) equipped with titanium blades and a plastic-lined bowl to avoid contamination have been used at the FCL for the blending and homogenization of QC samples including a diet RM (RM 8431 – Mixed Diet) sold by the National Institute of Standards and Technology (NIST).[4,5] Plastic trays and storage containers are most often used to minimize trace element contamination. Elements that have a high potential for contamination include Al, Ca, Cr, Cu, Fe, Zn, and Sn. Sample storage containers should be selected based on potential for introducing contaminants while considering cost effectiveness. Although Teflon® containers present the lowest risk of contamination, less expensive polyethylene and polypropylene containers are generally adequate. Containers made of quartz of borosilicate glass are less desirable. During sample preparation (particularly for low-temperature (<100 °C) wet ash preparations), glassware can be silanized to minimize the likelihood of contamination. The coating will be retained on the surface of the glassware to temperatures of approximately 350 °C, providing excellent protection from contaminants being leached from the glassware due to the interaction of the sample matrix with the glass. In addition, silanization protects from loss of analyte due to adsorption on container walls.

During the sample processing stage, contamination can generally be controlled with good laboratory practices. In an ideal situation, samples can be processed in high-energy particulate air (HEPA) filtered clean rooms designed to have minimal contamination. Clean hoods and laminar air flow units can also prove useful and the analyst should consider common sense practices, such as wearing clothing and shoe protectors to minimize particulate generation, selecting powder free gloves, cleaning glassware and plasticware with dilute, high-purity acid, and covering gas cylinders that can produce metal contamination.

2.2 Sample Preparation

It is imperative that high-purity reagents be used for sample preparation to minimize trace element contamination. Sub-boiling distilled or doubly distilled acids are good choices for sample treatment when ultratrace determinations are required. For example, sub-boiling distilled nitric acid (Seastar Chemicals, Sidney, BC, Canada) contains very low levels of contamination: $<50\,pg\,mL^{-1}$ for Ca, $<20\,pg\,mL^{-1}$ for Fe; $<10\,pg\,mL^{-1}$ for K, Cr, Cu, Na, Mg, Ni, and Zn, and $<1\,pg\,mL^{-1}$ for Mn, Co, and V. Typical contamination levels in hydrochloric acid (Ultrex II, J.T. Baker, Phillipsburg, NJ, USA) are $<500\,pg\,mL^{-1}$ for Ca and Fe, $<200\,pg\,mL^{-1}$ for Cr and Ni, $<100\,pg\,mL^{-1}$ for Na and K, $<50\,pg\,mL^{-1}$ for Mg, V, and Zn, and $<10\,pg\,mL^{-1}$ for Co, Cu, and Mn. In addition to acid purity, all reagents used throughout the sample preparation and analysis provide potential sources of contamination and must be carefully evaluated.

Environmental considerations are equally as important as reagent purity when selecting a sample preparation method. If possible, samples should be handled in clean air conditions under laminar flow to prevent contamination. Potential sources of contamination include the analyst (particulate from skin or clothing), sample container and delivery material (e.g. pipets), air, gloves, metallic heating blocks, muffle furnaces, and various other sources. Due consideration must be given to the many potential sources of contamination if reliable quantitative results are to be obtained.

Most commonly, samples are introduced into atomic spectrometers as solutions. Depending on the sample matrix and analytes of interest, one of several procedures can be used to get samples (or analytes) into solution. These include extraction, dry ashing, and wet ash digestion. Wet ash digestions can either be done at atmospheric pressure on a hot plate or heating block, or at high pressure using a digestion bomb or by

microwave heating. These techniques will be discussed more thoroughly in the following sections. In addition, direct analysis can be accomplished in some instances either by simple dilution of the sample (e.g. with milk samples) or by slurry (or solids) analysis techniques. Sample preparation for direct and slurry analysis will be discussed in section 2.2.4.

2.2.1 Extraction

In some cases, a simple analyte extraction or leaching procedure using acidic solutions can be effective for food and biological samples.[6,7] The success of these methods depends on the analytes of interest and the sample matrix. Azouzi et al. have reported a procedure for the room-temperature leaching of several elements from mussel samples using a solution of 1.6 M HNO_3, 1.2 M HCl, and 0.1 M H_2O_2.[6] Samples of 200 mg were ultrasonicated in 10 mL of the acid mixture for 120 min, and determinations were carried out by either FAAS or GFAAS. The authors reported quantitative recoveries for Ca, Cu, Fe, K, Mg, Mn, Na, V, and Zn and partial recoveries (50–80%) for Cd, Co, Cr, Rb, and Se. Relative errors were lower than 9% when analyzing a certified reference material (CRM 278 Mussel Tissue).

Tetramethylammonium hydroxide (TMAH) and tertiary amines have been employed for the alkaline extraction of elements from biological and food samples.[8–11] Mixed amines have been demonstrated to be useful for the complete extraction of Ca, Cu, Mg, Mn, Na, K, Zn, Mo, Rb, Se, Cr, and Zn from several RMs with detection by ICP/AES and ICP/MS.[8] Samples (0.2 g) and 10 mL of 9.5% v/v mixed amines were magnetically stirred for 30 min at room temperature and then centrifuged to remove the solids. Materials studied include the NIST Standard Reference Material (SRM) 1577b Bovine Liver, 1566a Oyster Tissue, 1570a Spinach Leaves, 1567 Wheat Flour, and 1548 Total Diet. Unfortunately, low recovery values were observed for Al and Fe in all of the matrices and for Ca in spinach and apple leaves. The authors concluded that the low recoveries were probably the result of insoluble silicate compounds often retained in the inorganic matrix of plant tissues.

The use of TMAH extraction has also been reported for the determination of iodine in food samples with ICP/MS detection.[9,10] Iodine is very susceptible to loss during sample preparation either by dry ashing, where an alkaline agent is necessary, or wet ashing, which requires addition of a strong oxidizing agent (e.g. $HClO_4$). As an alternative sample preparation technique, quantitative extraction of iodine from foods (e.g. milk powder, egg powder, and cod muscle) has been demonstrated using TMAH at elevated temperature (90 °C) for 1 h in a sealed vessel.[9] Samples of 0.2–0.5 g were used with 5 mL of water and 1 mL of 25% TMAH. The procedure was not successful for the extraction of iodine from an algae powder (~75% recovery) unless particle sizes were reduced to less than 300 μm. A similar TMAH extraction procedure was used for the determination of iodine in milk powders, bovine liver, and infant food.[10] In comparing this extraction method with microwave-assisted digestion, it was concluded that although both methods provided accurate results, the extraction procedure was more time-consuming (2–3 h compared with 1 h for microwave digestion) and could not guarantee total extraction of iodine in all cases.

Although a simple extraction might be useful in some instances, it is not generally recommended. This is because recoveries are dependent on both the analyte and food matrix, making the technique of little utility in a laboratory that deals with a range of sample types. Better results can be expected if methods are developed using more reliable and complete dry ash and wet ash procedures.

2.2.2 Dry Ashing

Preparing samples by dry ashing is commonly used in the analysis of foods.[12–14] Although the method is generally time-consuming, usually taking a day or more, it requires little attention from the analyst since samples are prepared in a batch mode. Ashing aids (e.g. magnesium nitrate) and modification of the sample matrix are often used to eliminate potential losses and/or speed up the ashing process; however, these can lead to contamination of the sample and poorer DLs. A benefit of dry ashing, compared with wet ash digestions, is that the resulting ash can be dissolved in a small amount of diluent. The small dilution factors possible with dry ashing offer the potential for improved DLs.

Results from a Nordic Committee on Food Analysis (NMKL) collaborative study on the determination of several metals (Pb, Cd, Zn, Cu, Fe, Cr, and Ni) in foodstuffs using dry ashing and AAS (flame and furnace) demonstrate the utility of the technique.[12] In this method, 10–20 g samples (liver paste, apple sauce, minced fish, wheat bran, milk powder, and composite diets) were dried and then ashed overnight at 450 °C. Then, 1–3 mL of H_2O was added and the temperature maintained at 450 °C for several hours. This process was repeated until complete combustion was obtained (i.e. white/gray ash present). Hydrochloric acid (5 mL of 6 M) was then added and the solution was evaporated to dryness. HCl is generally not recommended for GFAAS analyses because of the formation of volatile chlorides that can be lost during thermal pretreatment steps. The residue was dissolved in 0.1 M HNO_3 for AAS (flame and furnace) determinations. It was concluded by the authors that

no loss of Pb, Cd, Zn, Cu, Fe, and Cr occurred during the dry ashing since the measured values agreed with reference values; however, Ni determinations showed poor precision, probably owing to contamination.

2.2.3 Wet Ashing

Wet ash sample decomposition can be performed on a heating plate under ambient pressure,[15,16] in a pressurized digestion vessel,[16–19] or with open-[20,21] and closed-vessel microwave heating.[22–25] In a study involving the determination of Mn and Cr by AAS, Tinggi et al. used wet ashing performed on a heating plate to decompose effectively a variety of food samples (breads, cereals, meats, dairy products, vegetables, and fruits).[15] In this method, 5–15 g of moist foods (or 1–2 g of dry) were placed in 150-mL flasks. An acid mixture of HNO_3–H_2SO_4 (12 + 2 mL) was added and left overnight. The next day, glass funnels were added to the flasks for refluxing and the mixture was heated. The appearance of white fumes of sulfur trioxide was used to indicate completion of the digestion. It should be noted that the authors also investigated the addition of $HClO_4$ to the digestion mixture, but found it unnecessary to achieve complete digestion. Good reproducibility was reported with the HNO_3–H_2SO_4 mixture, except for samples containing high levels of carbohydrate or fat. In these cases, it was necessary to use smaller sample sizes (1–2 g of wet food), increased digestion times, and additional HNO_3. The accuracy of the method was verified by the measurement of NIST SRMs (SRM 1549 Non-fat Milk, SRM 1573 Tomato Leaves, and SRM 1577a Bovine Liver). Although some workers have reported difficulties with the determination of Cr in samples with appreciable silica contents (e.g. lettuce), necessitating the addition of HF, these authors reported that no statistical difference in measured concentrations was found when HF was added to the digestion mixture. Although these results indicate the suitability of open-vessel wet ash digestion for the determination of trace elements in some instances, there are drawbacks. They are time-consuming, subject to analyte contamination and loss, and might not be vigorous enough to ensure complete decomposition of some matrices. For this reason, high-temperature and/or high-pressure methods are often preferred.

Microwave sample digestion is a growing field and the number of analysts using this sample preparation tool has increased considerably in the last decade. An excellent source of information on the principles and application of microwave sample preparation is the book by Kingston and Jassie.[26] *It must be noted that as a safety issue, specially designed laboratory microwave ovens rather than household microwave ovens must be used for procedures since corrosive fumes can severely damage the devices and lead to an explosion*. Although microwave digestion can be a rapid and reliable means of sample decomposition with relatively minor risk of contamination, care must be taken when analyzing samples with a variety of matrices. This is because, depending on the carbohydrate content of the samples, pressures will build up at different rates and require venting at different times. For this reason, microwave digestion systems that are capable of monitoring temperature and pressure in each vessel, such as the Perkin-Elmer Multiwave (Perkin-Elmer Corp., Norwalk, CT, USA), are essential if different types of foods are to be analyzed at the same time. Use of this system for the decomposition of a wide range of food samples has been reported.[27] Samples (nuts, eggs, cream, cheese, etc.) of 0.4–0.7 g were placed in high-pressure quartz digestion vessels and 5 mL of concentrated HNO_3 was added. Vessels were sealed and heated according to the following program: stage 1, ramp from 100 to 600 W in 5 min; stage 2, hold at 600 W for 5 min; stage 3, 1000 W for 10 min; and stage 4, 0 W for 15 min. Samples were then opened and diluted in water for analysis. Digested samples were analyzed by ICP/AES for NLEA mandatory and voluntary elements. Results for several food SRMs (milk powder, spinach, corn, total diet, etc.) were in good agreement with certified values.

Wu et al. investigated a microwave digestion method using HNO_3 or a mixture of HNO_3 and H_2O_2 for the preparation of plant and grain RMs for ICP/MS measurements.[23] The use of HNO_3 as a digestion medium is typically preferred for ICP/MS measurements since this acid introduces few spectral interferences. Although $HClO_4$ and H_2SO_4 are also used as oxidants, there are drawbacks to their use including the production of a number of spectral interferences and they are also an additional potential source of contamination. Perchloric acid can also be extremely dangerous, whereas H_2SO_4 can form sparingly soluble sulfates that can precipitate from solution. The method used by these researchers involved weighing 0.5 g (dry mass) of sample in a Teflon® digestion vessel and adding 2.5 mL (grain samples) or 5.0 mL (plant samples) of concentrated HNO_3. The mixture was then allowed to sit overnight prior to microwave digestion. A maximum of 12 samples were digested according to the following program: heat to 165 °C within a ramp time of 10 min at 1000 W; hold for 20 min at 165 °C under a maximum pressure of 13.8 bar (200 psi). After digestion, samples were diluted in water to a final HNO_3 concentration of 2% v/v. For the HNO_3–H_2O_2 digestion, 2–5 mL of H_2O_2 (30% v/v) was added to the digestion mixture prior to microwave digestion. The presence of H_2O_2 helped to maintain a higher temperature under the pressure limit (13.8 bar) and reduced the carbon content in the digestates, but its impurities limited the determination of some elements at low levels. For grain

RMs, the majority of recoveries for the 26 elements determined were within 85–115%. For plant SRMs, the majority of recoveries were within 70–115%, but were only 40–80% for Ni, Th, Ti, and U. The authors attributed the low recoveries to siliceous material that was not completely decomposed during digestion.

Some interesting modifications of traditional microwave digestion technology can be found in the literature. Deaker and Maher reported a low-volume microwave digestion method for the determination of Se in biological tissues by GFAAS.[22] Small samples (0.025–0.1 mg) and 1 mL of HNO_3 were placed in 7-mL screw-top Teflon® vessels. Two of these vessels were than placed into larger 120-mL digestion vessels. Twelve of these traditional, larger vessels could be loaded into the microwave cavity at once. The optimized microwave program was 600 W for 2 min, 0 W for 2 min, and 450 W for 45 min. Quantitative recoveries of selenium from several RMs (oyster tissue, albacore tuna, fish flesh, etc.) were obtained using this method. Recently, Levine et al. evaluated a high-pressure, high-temperature focused microwave digestion system (BM-1S/II, Plazmatronika, Wroclaw, Poland) for the determination of 18 elements in different RMs including four food-related materials (peach leaves, orchard leaves, oyster tissue, and bovine liver) by ICP/AES and GFAAS.[25] Compared with traditional microwave digestion systems, the focused system employing tetrafluorometoxil polymer digestion vessels (TFM-PTFE) can accommodate higher temperatures (350 °C) and pressures (up to 160 bar). This should result in more complete destruction of organic matrix constituents using only HNO_3 as the digestion medium. By comparing GFAAS background absorbance and liquid chromatographic analysis of nitrobenzoic acid decomposition products, the authors concluded that the new system provided more complete destruction of the sample matrix. The high-pressure, high-temperature system also produced more accurate results for metals in the RMs (average error 7.1% in mean recovery values compared with 11.8% for the traditional system).

2.2.4 Direct Analysis and Slurry Sampling

Direct analysis (including simple dilution) of food samples offers a number of potential advantages. Advantages include reduced sample preparation time, decreased potential for analyte loss, reduced risk of contamination from reagents because less sample handling and fewer reagents are required, elimination of hazards associated with the use of strong acids and other corrosive reagents, and the potential for improved sensitivity since samples are not as dilute. Two categories of sample preparation for direct analysis are considered in this section. One involves the simple dilution of samples that are, more or less, solubulized (e.g. for beverages). The other method involves the preparation of a slurry or suspension of solid samples in a liquid diluent.

Arruda et al. reported a method for the direct analysis of milk for Al using GFAAS.[28] Samples (5 mL) were diluted to 15 mL with 0.2% v/v HNO_3 for analysis. Determinations using this sample preparation method were compared with results obtained by both dry ashing and wet ashing. Good agreement was reported and the accuracy of the method was verified by analyzing a milk RM. A procedure for the direct determination of major (Ca, K, Mg, Na, and P) and minor (Zn, Al, Ba, Cu, I, Mn, Mo, Pb, Rb, Se, and Sr) elements in milk by ICP/AES and ICP/MS using mixed tertiary amines has also been published.[29] Typically, the analysis of milk by direct nebulization of aqueous solutions in inductively coupled plasma (ICP) spectrometric methods is plagued by poor sensitivity and accuracy. The poor performance has been attributed to milk's fatty nature and the existence of large droplets that cannot be completely vaporized in the plasma. In addition, the chemical form of the analyte element (existing in aqueous phase, casein micelles, etc.) may also be critical.[29] In the method presented, 0.1 g of powdered milk (or 1 mL of liquid milk) was dissolved in 10.0 mL (or 9.0 mL) of 10% v/v mixed amine reagent at pH 8 for ICP/AES measurements. For ICP/MS measurements, 0.2 g of powdered milk (or 2.5 mL of liquid milk) was dissolved in 49.5 mL (or 47.0 mL) of 5% mixed amine reagent at pH 8. It was stated that the mixed amine appeared to promote the total or partial dissolution of casein micelles, increasing the metal concentration in the aqueous phase. This led to improved performance compared with earlier milk analysis by ICP/AES, which were plagued by analyte insolubility in alkaline solutions or problems in the nebulization and atomization of milk slurry particles. The accuracy of the proposed method was verified by the analysis of milk SRMs. The values obtained agreed with the certified values at the 95% confidence level.

Flow injection analysis can be useful for the direct analysis of food and beverage samples. A review on the application of flow injection in atomic spectrometry by Burguera and Burguera demonstrates the versatility of the technique, such as for on-line dilution, dissolution, liquid–liquid extractions, and ion-exchange preconcentration.[30] Flow injection was used by Baxter et al. for the multielement analysis of wines by ICP/MS.[31] Wine samples (5.0 mL) were diluted 1 : 1 with sample diluent (1.0% HNO_3) prior to analysis. Injections of 0.5 mL were introduced into a 1.0-mL carrier stream composed of 5.0% ethanol and 0.5% HNO_3. The method was verified by spike recovery experiments where all values were within $100 \pm 25\%$ and by analysis of provisional RMs produced by the Community Bureau of Reference (CBR).

The effect of ethanol concentration on the analyte signal was investigated and it was reported that only minor variations of ±10% were observed, making the method applicable to a range of alcohol contents in wines. The data were used to determine the origin of various Spanish and English wines.

Solid samples can be analyzed directly by preparing them as slurries or suspensions. These samples can then be nebulized into a flame (for FAAS) or plasma (for ICP/AES/MS), or alternatively, deposited in a graphite tube for GFAAS or electrothermal vaporization (ETV) ICP/AES/MS measurements. The benefit of the latter techniques employing graphite tubes is that residence times are much longer, leading to improved atomization/vaporization and fewer problems related to particle size effects. Slurry sample introduction for GFAAS and ETV/ICP/AES/MS is straightforward and can be automated using commercial autosamplers. This is especially advantageous for routine determinations. Numerous applications of slurry GFAAS and ETV/ICP/AES/MS can be found in the literature[32–35] and specific applications will be discussed later.

Most food samples must be ground into smaller particles prior to analysis. This can be accomplished by freeze-drying samples and then shattering them in a ball-mill device or using wet grinding. A wet grinding method used in this laboratory[36] involves adding samples and Teflon® beads (10 : 1 mass ratio of beads to sample) to polyethylene bottles. Approximately 10–15 mL of water or dilute HNO_3 are added to the bottles and the mixture is shaken for 2 h.

A general procedure for slurry preparation is provided here; however, readers are encouraged to consult the literature[32,33] for a more complete treatment of factors (e.g. particle size, density, analyte distribution, mode of mixing) that must be considered for optimizing slurry preparations. Samples (typically 1–50 mg) can be weighed directly into autosampler cups to avoid problems that can result from the transfer of small quantities of sample. Although a variety of diluents have been used, the authors have found a mixture containing 5% v/v sub-boiling distilled HNO_3 and 0.005% Triton X-100™ surfactant (Rohm and Haas registered trademark for octylphenoxypolyethoxyethanol; Sigma Chemicals, St Louis, MO, USA) to be useful for a range of sample types. Nitric acid facilitates the extraction of analyte into the liquid phase of the slurry; whereas, Triton X-100™ serves as a wetting agent and assists in particle dispersion. If particle coagulation is not problematic and/or analyte extraction is not desirable (e.g. in homogeneity characterizations), deionized, distilled water can be used as the diluent. In some cases, it might be necessary to prepare larger masses and/or volume slurry preparations. This can be done by placing sample and diluent in polyethylene tubes, then vortexing the mixture while a smaller aliquot is removed for the autosampler cup.

3 MEASUREMENT TECHNIQUES

Several factors should be considered when selecting an appropriate measurement technique. These factors include detection capability, ease of use, availability of instrumentation, speed of analysis, and cost. One must consider the relative importance of the factors for their particular application; however, detection capability is often a primary concern. A good general rule in selecting a method is to use the simplest method possible that meets the analysis requirements. For example, if the goal of an analysis is to determine $\mu g\,mL^{-1}$ (ppm) levels of metals in a beverage sample, the analyst should check to determine if it can be analyzed by simple dilution and measurement by flame AAS or simultaneous multielement ICP/AES (if more than a few elements are to be determined). Most trace element determinations in foods are accomplished by a handful of atomic spectroscopic techniques. These techniques are FAAS, GFAAS, ICP/AES, and, more recently, ICP/MS. Since a detailed description (including fundamentals and applications) of each of the methods is provided elsewhere in this Encyclopedia, only a brief description of the techniques is provided here. Basic principles of the methods and applications related to the determination of metals in foods will be reviewed in this section. A comparison of some of the important attributes associated with the measurement techniques under consideration can be found in Table 1.

Table 1 Comparison of measurement techniques

Technique	Multielement	DL ($ng\,mL^{-1}$)	Matrix effects	Selectivity	Precision (%)	Accuracy	Cost
FAAS	No	1–100	Small	Good	0.1–1	Good	Low
GFAAS	No	0.01–1	Small	Good	1–5	Good	Moderate
ICP/AES	Yes	1–100	Small	Moderate	0.1–1	Good	Moderate
ICP/MS	Yes	0.001–0.1	Moderate	Good[a]	0.2–2	Good	High

[a] Isotopic selectivity.

3.1 Atomic Absorption Spectrometry

AAS is the most widely used technique for the determination of metals. A detailed discussion of AAS principles, instrumentation, and applications can be found in several books devoted to the subject.[37,38] In this technique, the absorption of radiation by neutral, ground-state atoms is measured to determine the amount of analyte present in the sample. Radiation sources for AAS include the hollow-cathode lamp (HCL) and the electrodeless discharge lamp (EDL), whereas atomization is accomplished with a flame (typically air–acetylene or nitrous oxide–acetylene) or an electrothermal atomizer [electrothermal atomization atomic absorption spectrometry (ETAAS), also known as GFAAS]. AAS determinations are usually made by FAAS when the concentration of analyte is high enough, or by GFAAS when concentrations are low. When the detection capability is adequate, FAAS should be selected whenever possible because it is less time-consuming and also less sensitive to interference effects (e.g. from background absorbance) than GFAAS. Advantages of AAS are the potential for very high sensitivity (ng mL^{-1} DLs for FAAS and sub-ng mL^{-1} for GFAAS), instrumental simplicity, relatively low cost, and excellent and well-documented performance characteristics (good precision and accuracy). A disadvantage is that AAS is primarily a single-element technique, although continuum-source and sequential AAS systems have been developed. In addition, the linear range is typically less than two orders of magnitude, requiring the use of alternative, less sensitive wavelengths or some other means of reducing sensitivity (e.g. mini-gas flows during furnace atomization).

3.1.1 Flame Atomic Absorption Spectrometry

FAAS is a well-established technique for the analysis of trace metals in food samples. It is estimated that 80% of all currently available trace element food composition data are the result of FAAS analyses after either wet ashing or dry ashing sample pretreatment. FAAS is a simple and robust tool for the analysis of a wide range of elements and calibration can typically be accomplished using a 3–5-point calibration curve made with aqueous standards. Because of the robustness (i.e. maturity) of the technique, relatively few new FAAS methods are being published in the analytical literature. This is not to say that interesting work is not being done in this area and several of the recently reported applications of FAAS in the analysis of foods[6,15,39–41] or food-related materials[42] are presented here.

The determination of Ca, Cu, Fe, Mg, and Zn in mussels by FAAS was recently reported.[6] Samples were prepared for analysis by ultrasonic extraction in a mixture of HNO_3, HCl, and H_2O_2 and analyzed using FAAS at 422.6 nm (Ca), 324.8 nm (Cu), 248.8 nm (Fe), 285.2 nm (Mg), and 213.9 nm (Zn) with an air–acetylene flame. The method provided a DL, defined as $3\sigma_{blank}$, of 0.081, 0.012, 0.059, 0.002, and 0.007 μg mL^{-1} for Ca, Cu, Fe, Mg, and Zn, respectively. Relative standard deviation (RSD) values ranged from 0.6 to 7.1% ($n = 8$ analyses). In discussing the figures of merit for this and other techniques, it is important to remember that these values are dependent on a number of variables including the instrumental measurement, the sample preparation procedure including reagent contamination levels, and the sample matrix.

The essentiality of Mn and Cr for humans is well established; however, requirements or levels of absorption for adults have not been clearly determined. These questions can only be answered when accurate and reliable data on concentrations of these elements in foods are available. A method for the determination of Mn by FAAS (and Mn and Cr by GFAAS) at 279.5 nm was reported using an air–acetylene flame.[15] Wet ashing was used to prepare a wide variety of samples for analysis. Relatively high levels of Mn (>10 μg g^{-1}) were found in most cereal products, whereas meats, dairy products, vegetables, and fruits contained low levels of the element. The accuracy of the FAAS method was verified by analyzing NIST SRM 1573 Tomato Leaves and NIST SRM 1577a Bovine Liver; recoveries were 95 and 98%, respectively. The precision (%RSD) of these measurements was 3% ($n = 5–6$) for Mn and the DL, based on a 2.0-g sample in a 20-mL final volume, was 0.15 μg g^{-1} (solution DL 0.015 μg mL^{-1}).

Although the importance of Se in the human diet is well established, the range between safe intake and toxicity is narrow, making accurate and reliable methods of analysis critical. Se determinations by traditional FAAS suffer from a relatively poor DL (around 1 μg mL^{-1}); however, hydride generation (HG) sample introduction can be used to achieve sub-parts per bilion DLs for this element.[41] The determination of Se in dairy foods and drinks by HG AAS has been reported and the data used to estimate the daily dietary intake of Se in the normal Spanish diet from these food sources.[41] Samples were wet ashed with concentrated HNO_3 and $HClO_4$, then Se^{6+} was reduced to hydride forming Se^{4+} by addition of concentrated HCl. The standard additions method was used for calibration. RSD values ($n = 7$ measurements of three random samples) ranged from 5% to 13% and the DL (3σ) was 0.210 ng mL^{-1}. The accuracy of the method was checked by spike recovery for three random samples and ranged from 97% to 106%. Of the dairy products studied, the highest levels of Se were found in aged cheese with concentrations ranging from 86 to 366 ng g^{-1} (wet-weight basis). The mean Se concentration was 164 ng g^{-1}. The highest level of Se in various alcoholic

and nonalcoholic drinks was found in pineapple juice (mean concentration of 2.88 $\mu g\,L^{-1}$).

3.1.2 Graphite Furnace Atomic Absorption Spectrometry

GFAAS provides sub-parts per billion detection capability using microliter-sized samples deposited in a graphite tube. The graphite tube is resistively heated to temperatures of up to 2800 °C to produce a dense cloud of atomic vapor. Traditionally, samples have been introduced into the graphite furnace as solutions; however, direct solid and slurry sampling have also been employed.[43] The technique is a versatile and highly sensitive means of analysis for a wide range of sample types. The success of GFAAS can largely be attributed to the development of the stabilized temperature platform furnace (STPF) concept.[43,44] The STPF approach employs a platform furnace so that the analyte is atomized in a more isothermal environment. STPF also provides for matrix modification for analyte stabilization, rapid furnace heating, pyrolytic graphite-coated tubes to reduce chemical interferences, good background correction (such as the Zeeman effect), no gas flow during atomization, and integrated absorbance measurements for quantification. Employing these conditions can significantly improve the quality of GFAAS analyses, allowing the technique to be used routinely for the determination of trace elements. Most often, calibration (even for slurry samples) can be performed with aqueous standards. The wide range of applications that have been reported in the literature demonstrates the versatility of the method.

Chromium is one of the essential trace elements in the human body and is typically found at low levels in foods. Soil contamination is an important factor contributing to the total Cr concentration in plant tissues because of uptake of this element. Because excessive levels of Cr (particularly Cr^{6+}) in the human diet can cause adverse effects, monitoring of foods grown in areas of suspected contamination is necessary. The determination of Cr in wine and grapes by GFAAS was reported by a team of French and Spanish investigators since these foods could be a significant source of the element in a typical diet.[45] Wine samples were analyzed directly, whereas, grapes were microwave digested in a mixture of HNO_3, H_2O_2, and $HClO_4$ prior to analyses. Cr was measured at 357.9 nm and the furnace program included a 1000 °C char step (no modifier) and an atomization temperature of 2600 °C. Calibration with aqueous standards was used for quantification. The suitability of analyzing wine samples directly was assessed by comparing concentrations determined with those obtained by dry ashing the samples. The accuracy of the GFAAS method was validated by spike recovery experiments (mean recovery 99.3% for four determinations in two different samples). The precision of the method (%RSD) was between 1.8 and 9.0% for six determinations in five different samples and a DL of 1.8 pg (0.09 $ng\,mL^{-1}$) was reported. Chromium concentrations were determined in a total of 79 wine samples and 12 grape samples. The levels were in the ranges 7–90 $\mu g\,L^{-1}$ in red wines, 6.6–43.9 $\mu g\,L^{-1}$ in white wines, 7.3–14.7 $\mu g\,L^{-1}$ in rosé wines, and 10.5–36.0 $\mu g\,L^{-1}$ in champagnes. Concentrations in grape samples were in the ranges 2.5–20.9 $\mu g\,kg^{-1}$ for white varieties and 2.4–64.6 $\mu g\,kg^{-1}$ for red varieties. The daily contribution from wine consumption in France was estimated to be 4.06 μg per resident, which is less than 10% of the daily dietary intake.

Phifer at the US Food and Drug Administration developed a method for the determination of Cr (357.9 nm) and Mo (313.3 nm) in medical foods (enteral formulas) by GFAAS.[46] Samples were prepared for analysis by dry ashing. Both Cr and Mo were atomized off the wall of the graphite tube and neither analyte required the use of a chemical modifier. Optimum charring temperatures were 1600 and 1650 °C for Cr and Mo, respectively. Atomization temperatures of 2400 °C and 2650 °C proved optimum for Cr and Mo, respectively. Calibration with aqueous standards was performed and the method was validated by analyzing four NIST SRMs (SRM 1572 Citrus Leaves, SRM 1566a Oyster Tissue, SRM 1577a Bovine Liver, and SRM 1568 Rice Flour). Results for all of the materials analyzed were within the certified concentration ranges. The reproducibility of the method (%RSD) determined for one medical product, containing $0.31 \pm 0.02\,\mu g\,g^{-1}$ Cr and $0.63 \pm 0.03\,\mu g\,g^{-1}$ Mo, was 6.8% for Cr and 4.8% for Mo ($n = 5$ measurements). Reported DLs (3σ) were 0.36 $ng\,mL^{-1}$ for Cr and 1.0 $ng\,mL^{-1}$ for Mo.

GFAAS has frequently been employed for the determination of Se in food samples.[13,20,47] A method for the determination of Se in infant and enteral formulas by dry ash sample preparation and GFAAS detection has been developed.[13] Samples were ashed with the aid of HNO_3 and magnesium nitrate, then HCl (5 + 1) was added to dissolve all of the ash and reduce Se^{6+} to Se^{4+}. Se^{4+} was subsequently reduced to Se^0 with ascorbic acid and collected on a membrane filter. Filters were then microwave digested in HNO_3, diluted in water, and analyzed by GFAAS. The reason for the Se conversion was to isolate it from the dissolved ash, a step that was essential since only deuterium background correction was available. Nickel nitrate was used as a chemical modifier and the optimum charring and atomization temperatures were 1100 and 1800 °C respectively. Several RMs (Mixed Diet, Corn Bran, Milk Powder, etc.) were analyzed with the proposed method and measured concentrations were within the concentration range for each of the samples. Measured Se concentrations in

12 different infant and enteral formulas ranged from $2.9 \pm 2\,\mu g\,kg^{-1}$ ($n = 6$ determinations) for a nonfortified infant soy formula to $191 \pm 7\,\mu g\,kg^{-1}$ ($n = 4$ determinations) for a fortified infant milk-based powder. A DL (3σ) for Se of $0.44\,ng\,mL^{-1}$ was reported.

As mentioned previously, GFAAS can also be used for the analysis of slurry samples. Recent applications of slurry sampling GFAAS for the analysis of food include the determination of Cr in tomato leaves,[33] Cr, Co, and Ni in vegetables,[34] Mo, Al, Cr, and Mn in milk powder,[48] Se in fish samples,[49] and Cd, Cu, Fe, Pb, and Se in fruit.[50] Because of the relatively large amount of solid being vaporized in the graphite tube, matrix effects can be problematic. The presence of organic material, not completely removed during the charring step, affects the volatility of some elements, increases background levels, and leads to the build-up of carbonaceous material in the furnace. To alleviate these problems, chemical modifiers can be used to stabilize volatile elements during furnace heating and good background correction methods (e.g. Zeeman or Smith–Hieftje) should be employed. Effective methods for removing carbonaceous material and thus avoiding its deleterious effects include the use of an oxygen or air ashing step[33,50] in the furnace program or addition of H_2O_2[34,48] to the slurry diluent.

Chromium, cobalt, and nickel are typically found at very low concentrations in vegetables, the primary sources of these elements in the human diet. Because concentrations are low, very sensitive detection methods such as GFAAS must be employed for their determination. Carlosena et al. compared several sample preparation methods (dry ashing, slurry preparation, and continuous-flow microwave-assisted digestion) for the determination of Cr, Co, and Ni in vegetables.[34] For slurry analyses, samples were ground in a household grinder and then a mortar and pestle before they were dried for several hours. A 0.1–0.15-g portion of the sample was then mixed with 2.5 mL of 2% v/v HNO_3 and 2.5 mL of 20% v/v H_2O_2 and ultrasonicated for 10 min prior to injection. Although palladium nitrate and magnesium nitrate were investigated as chemical modifiers, it was found that no additional modifier was needed since the presence of H_2O_2 was effective in removal of the total carbonaceous residue. Optimum charring and atomization temperatures for the analytes were 1600 and 2650 °C for Cr, 1500 and 2600 °C for Co, and 1400 and 2650 °C for Ni. Results from the slurry method using aqueous calibration standards were compared with those from the other sample preparation procedures and with certified values for NIST SRM 1572 Citrus Leaves. It was concluded that the slurry method was accurate and precise (<8% RSD) for the determination of Co and Ni, but produced low values (about 80% recovery) for Cr. Reported DLs (3σ) were $84.5\,ng\,g^{-1}$ ($1.7\,ng\,mL^{-1}$) for Cr, $28.0\,ng\,g^{-1}$ ($0.56\,ng\,mL^{-1}$) for Co, and $42.0\,ng\,g^{-1}$ ($0.84\,ng\,mL^{-1}$) for Ni. Levels of the elements in vegetables (lettuce, endive, cauliflower, cabbage, leek, and watercress) ranged from <DL to $0.15 \pm 0.01\,\mu g\,g^{-1}$ for Cr, <DL to $0.68 \pm 0.02\,\mu g\,g^{-1}$ for Co, and 0.31 ± 0.03 to $3.26 \pm 0.13\,\mu g\,g^{-1}$ for Ni. The highest concentration of each of the elements was found in lettuce samples.

A method for the analysis of fruit slurries for Cd, Cu, Fe, Pb, and Se using GFAAS has also been reported.[50] In this method, portions of 1.5–2.5 g of homogenized fresh sample were placed in polyethylene bottle and 10 g of zirconia spheres and 10.0 mL of 1.0% w/v Triton X-100™ were added. The bottles were shaken for about 60 min, then samples were separated from the beads and diluted in water for analysis. It was reported that the grinding procedure ensured that 90% of the particles had diameters of <25 μm. The relatively fine particle size and low density of the samples allowed for manual shaking of the samples to be used to ensure that particles did not settle out prior to sample injection. Optimum charring and atomization temperatures for the analytes were 700 and 2000 °C for Cd ($NH_4H_2PO_4$ as modifier), 1000 and 2300 °C for Cu (no modifier), 1400 and 2400 °C for Fe [$Mg(NO_3)_2$ as modifier], 500 and 2100 °C for Pb (no modifier), and 1000 and 2200 °C for Se [$Ni(NO_3)_2$ as modifier]. Aqueous calibration was employed for all determinations and the accuracy of the method was validated using NIST SRM 1572 Citrus Leaves, spike recovery, and comparison with results obtained using microwave digested samples. DLs (3σ) were $0.3\,ng\,g^{-1}$ ($0.024\,ng\,mL^{-1}$) for Cd, $3.5\,ng\,g^{-1}$ ($0.28\,ng\,mL^{-1}$) for Cu, $15.0\,ng\,g^{-1}$ ($1.2\,ng\,mL^{-1}$) for Fe, $0.5\,ng\,g^{-1}$ ($0.04\,ng\,mL^{-1}$) for Pb, and $10.0\,ng\,g^{-1}$ ($0.8\,ng\,mL^{-1}$) for Se. Precision values (%RSD) for 10 replicate determinations on each of five different samples ranged from 3.5 to 7.0%. The method was used for the analysis of eight types (five samples per type) of fruit samples (strawberry, medlar, banana, papaya, kiwi, mango, custardapple, and pineapple). Mean determined concentrations (fresh weight of edible portion) ranged from 0.3 to $50\,ng\,g^{-1}$ for Cd, from 0.337 to $1.094\,\mu g\,g^{-1}$ for Cu, from 2.00 to $5.50\,\mu g\,g^{-1}$ for Fe, from 0.050 to $0.396\,\mu g\,g^{-1}$ for Pb, and from 10 to $20\,ng\,g^{-1}$ for Se.

3.2 Inductively Coupled Plasma Atomic Emission Spectrometry

ICP/AES is a powerful technique for the determination of trace elements in foods. An excellent source of information on the fundamentals and some applications of ICP/AES is the book edited by Montaser and Golightly.[51] In this technique, samples are introduced into a high-temperature (7000–8000 K) argon plasma where they are efficiently vaporized, atomized, and

excited. The emission of excited atoms is then measured to determine the quantity of analyte present in the sample. One of the major advantages of ICP/AES over AAS techniques is that it is capable of simultaneous multielement analysis, offering the potential for greater sample throughput. In addition, ICP/AES offers a linear dynamic range of five orders of magnitude or greater (compared with two orders of magnitude for AAS) making it useful for determining both macro and micro constituents of the sample. The DLs for ICP/AES are similar to those for FAAS and typically poorer than those for GFAAS by one or two orders of magnitude; however, the use of axially viewed ICP/AES systems has narrowed the gap. Because of the multielement capability and generally adequate detection capability, ICP/AES is becoming increasingly popular for the determination of trace elements in foods.

Negretti de Brätter et al. used ICP/AES for the determination of Al, Ca, Cu, Fe, K, Mg, Na, P, and Zn in total diet samples.[18] One of the purposes for the study was to compare mineral and trace element intake for Venezuelan populations where a high incidence of stomach cancer exists and for populations where there is a low rate of incidence. Homogenized samples were digested using a high-pressure wet digestion method with HNO_3 and H_2O_2 (2 + 1). Analytes were measured simultaneously at the following wavelengths; 396.15 nm (Al), 324.75 nm (Cu), 766.49 nm (K), 589.59 nm (Na), 213.86 nm (Zn), 393.37 nm (Ca), 259.94 nm (Fe), 279.08 nm (Mg), and 214.91 (P). Analysis of NIST SRM 1548 Total Diet was used to confirm the accuracy of the proposed method. Results for all of the elements except for K were within the certified concentration range. Results for K were 11% low compared with the mean certified concentration and it was assumed that this was caused by absorption of K by the Teflon® cover of the digestion vessel. The DLs (3σ) for the ICP/AES method were 37 $\mu g\,kg^{-1}$ (Zn), 2400 $\mu g\,kg^{-1}$ (P), 7.3 $\mu g\,kg^{-1}$ (Fe), 1.64 $\mu g\,kg^{-1}$ (Mg), 2.5 $\mu g\,kg^{-1}$ (Cu), 0.72 $\mu g\,kg^{-1}$ (Ca), 12 $\mu g\,kg^{-1}$ (Al), 40.5 $\mu g\,kg^{-1}$ (Na), and 490 $\mu g\,kg^{-1}$ (K). Results for the mean daily intake of the population investigated for the high stomach cancer incidence area ($n = 77$) as compared with the low incidence area ($n = 33$) showed significantly higher daily intakes for Na (2082 versus 1471 mg), K (1190 versus 731 mg), and P (640 versus 381 mg) and significantly lower intakes for Ca (925 versus 1379 mg) and Cu (2.78 versus 4.66 mg). No differences were obtained for Al, Fe, Mg, and Zn.

A Spanish group recently used ICP/AES for the characterization of green coffee varieties.[52] Eleven metals (Zn, P, Mn, Fe, Mg, Ca, Na, K, Cu, Sr, and Ba) were chosen as chemical descriptors and pattern recognition techniques, such as principal component analysis and cluster analysis, were applied to characterize arabica and robusta coffees. Samples were wet ashed using a mixture of HNO_3 and H_2SO_4 prior to analysis. Elements were determined at the following wavelengths: 213.86 nm (Zn); 214.91 nm (P); 257.61 nm (Mn); 259.94 nm (Fe); 279.55 nm (Mg); 393.37 nm (Ca); 589.59 nm (Na); 766.49 nm (K); 324.75 nm (Cu); 407.77 nm (Sr); and 455.40 nm (Ba). Significant differences were found in the metallic content of the two coffee varieties. The largest differences were found with P, Mn, and Cu, making these elements suitable for characterizing the two coffee varieties. Determined concentrations were 1410–2200 $\mu g\,g^{-1}$ for P, 14.5–47.9 $\mu g\,g^{-1}$ for Mn, and 14.3–76.9 $\mu g\,g^{-1}$ for Cu.

The determination of metals in food and biological samples with an axially viewed ICP/AES system using both a conventional concentric nebulizer and a microconcentric nebulizer has been reported.[53] Seventeen elements were determined in three microwave-digested RMs (CBR CRM 63 Milk Powder, CRM 185 Bovine Liver, and CRM 278 Mussel Tissue). Elements were determined at the following wavelengths: 328.07 nm (Ag); 396.15 nm (Al); 193.70 and 188.98 nm (As); 614.17 nm (Ba); 317.93 nm (Ca); 228.80 nm (Cd); 228.62 nm (Co); 267.72 nm (Cr); 327.40 nm (Cu); 259.94 nm (Fe); 769.90 nm (K); 279.81 nm (Mg); 257.61 nm (Mn); 819.48 nm (Na); 231.60 nm (Ni); 220.35 nm (Pb); 346.45 nm (Sr); and 213.86 nm (Zn). Measured DLs were typically better with the conventional nebulizer than the microconcentric nebulizer, but the magnitude of the difference was element dependent. For the elements studied, DLs (3σ) ranged from 0.05 $ng\,mL^{-1}$ (Mn) to 30.48 $ng\,mL^{-1}$ (Ca) for the conventional nebulizer and from 0.13 $ng\,mL^{-1}$ (Mn) to 173.66 $ng\,mL^{-1}$ (Na) for the microconcentric nebulizer. Results for the analysis of RMs were in good agreement (t-test used) with the certified values, except for a few cases. It was concluded that either nebulizer could be used to obtain both accurate and precise results; however, the microconcentric nebulizer gave poorer sensitivity and was more susceptible to a deterioration in analytical performance when high-salt matrices were analyzed.

3.3 Inductively Coupled Plasma Mass Spectrometry

In the last decade, ICP/MS determinations have started to play a significant role in the generation of food composition data. The technique involves the detection of ions extracted from a high-temperature argon plasma. ICP/MS provides multielement, isotopic data with sub-$ng\,mL^{-1}$ DLs with the same convenience of FAAS or ICP/AES and much faster than GFAAS. As interest in low-level trace element concentrations continues to grow, the role of ICP/MS in the analysis of food samples will continue to expand. An excellent source of information on ICP/MS is a recently published book edited by Montaser.[54] In addition, Crews has published a review article that

deals specifically with the determination of trace elements in foods using ICP/MS detection.[55] The article discusses a range of topics, including instrumentation, interferences, matrix effects, applications, and speciation measurements that utilize ICP/MS detection. Advantages of ICP/MS include excellent sensitivity and selectivity, simple spectra, wide linear dynamic range (6–7 orders of magnitude), and the ability to acquire rapidly multielement and isotopic information. The disadvantages of ICP/MS include costs (although they are becoming increasingly competitive) and a relatively low tolerance to total dissolved solids and susceptibility to matrix effects. Although mass spectra are much simpler than emission spectra, some spectral interferences (isobaric overlap, overlap from polyatomics, and doubly charged species) do exist and must be accounted for. In addition, a number of nonspectral interferences can occur, such as transport effects, ionization interferences, and ion sampling effects.[55] Internal standards are frequently employed for ICP/MS determinations to account for nonspectral interferences; however, care must be used in selecting a suitable internal standard. Ideally, the internal standard should be similar in both mass and ionization potential to the analyte or analytes of interest.

The high sensitivity of ICP/MS is illustrated in Table 2, which gives DLs for several elements obtained for wet ashed total diet samples. The reference on which the table is based[55] also contains data from the analysis of several RMs (mussel tissue, lobster hepatopancreas, oyster tissue, peach leaves, and mixed diet) that were digested using HNO_3 in a high-pressure digestion vessel. In general, the results were in good agreement with the certified values. For the determination of V, Cr, Zn, As, and Se, more accurate results were obtained when 8% nitrogen was added to the nebulizer gas flow to reduce polyatomic interferences. The DLs were generally poorer when nitrogen was added (a factor of 10 or less), but were adequate for the multielement determinations.

Table 2 DLs[a] ($ng\,mL^{-1}$ in solution and $\mu g\,g^{-1}$ fresh weight) for wet ashed total diet samples obtained with ICP/MS detection [adapted from Crews[55]]

Element	DL ($ng\,mL^{-1}$)[a]	DL ($\mu g\,g^{-1}$)[a]
B	0.5	0.05
Cr	0.1	0.01
Mn	0.1	0.01
Fe	10	1.0
Co	0.05	0.005
Ni	0.5	0.05
Cu	2.0	0.2
Zn	3.0	0.3
Se	0.02	0.002
Mo	0.1	0.01
Sn	0.5	0.05
Na	120	60
Mg	2	1
P	80	40
K	500	250
Ca	600	300

[a] 0.5 g of sample (fresh weight) diluted to a final volume of 250 mL for Na, Mg, P, K, and Ca determinations and diluted to 50 mL for the remaining elements.

A very powerful method that can be used with ICP/MS (and any other MS technique) detection is isotope dilution (ID). In ID analysis, the sample is spiked with a known amount of an enriched isotope of the analyte. The altered isotope ratio is then determined in the sample to calculate the amount of analyte present in the original sample. ID is recognized as a definitive analytical method and the accuracy of the method depends on the ability to precisely measure isotope ratios. Since another isotope of the element represents the ideal internal standard, results are expected to be accurate even when nonspectral interferences are present. Park and Suh used ID ICP/MS to determine Cr, Fe, Cd, and Pb in rice flour.[56] For the determination of Fe and Cr, the ICP was operated in the ''cool plasma'' mode to suppress the background molecular ion intensity ($^{40}Ar^{12}C$ interferes with ^{52}Cr and $^{40}Ar^{16}O$ interferes with ^{56}Fe). However, because the plasma was less robust, spectral interferences from $^{40}Ca^{16}O$ and $^{40}Ca^{16}O^{1}H$ on ^{56}Fe and ^{57}Fe were severe. The Ca was removed from the samples using a microcolumn loaded with 8-hydroxyquinoline. For ID/ICP/MS determinations, the isotopes measured were $^{52\text{ and }53}Cr$, $^{56\text{ and }57}Fe$, $^{111\text{ and }112}Cd$, and $^{206\text{ and }208}Pb$. Microwave digestion using HNO_3 only was not adequate for the determination of Fe and Cr in rice flour RMs. An additional treatment with HF and $HClO_4$ was necessary for quantitative recoveries of Fe and Cr in the samples. DLs (3σ) of $0.003\,\mu g\,g^{-1}$ ($0.04\,ng\,mL^{-1}$) for Cr, $0.069\,\mu g\,g^{-1}$ ($0.92\,ng\,mL^{-1}$) for Fe, $0.0003\,\mu g\,g^{-1}$ ($0.004\,ng\,mL^{-1}$) for Cd, and $0.0003\,\mu g\,g^{-1}$ ($0.004\,ng\,mL^{-1}$) for Pb were reported.

ETV coupled to ICP/MS has been used for the analysis of slurried food and biological samples.[11,35,57] Unlike GFAAS where samples must be vaporized and atomized in the graphite tube, ETV work requires efficient vaporization and subsequent transport of the analyte to the ICP/MS system. Compared with GFAAS, ETV/ICP/MS is more prone to matrix effects (such as space–charge effects and ionization suppression) that result from large amounts of material being introduced to the ICP/MS system. In addition, transport of the analyte from the furnace to the ICP is often enhanced in the presence of the sample matrix, making quantification on the basis of aqueous standards difficult. To deal with these issues, several steps can be taken. Use of internal standardization, ID, and/or addition of a physical carrier

to the graphite tube can be used to improve the accuracy of ETV/ICP/MS measurements. To ensure the reliability of internal standardization for quantification, an internal standard with a similar mass, ionization potential, and vaporization characteristics compared with the analytes to be determined should be selected. To meet these criteria, several internal standards might have to be used which is both inconvenient and also limits the number of analytes that can be determined. The transient nature of the ETV signal and the scanning nature of quadrupole ICP/MS instruments typically only allow for the determination of no more than five elements (isotopes) simultaneously.

ID can also be used for quantification for elements that have multiple isotopes. Recently, it has been used for the determination of Cu, Zn, Cd, and Pb in marine RMs by slurry ETV/ICP/MS.[57] Enriched isotopes of ^{65}Cu, ^{67}Zn, ^{111}Cd, and ^{204}Pb were spiked into the slurries as solutions. Samples were charred at 400 °C and vaporized at 2300 °C for the analyses. Owing to the volatility of Zn and Cd, a modifier (NH_4NO_3) was required to stabilize these analytes during the charring step. Determined concentrations of the analytes in three RMs (DORM-2 Dogfish Muscle, DOLT-1 Dogfish Liver, and NIST SRM 1566a Oyster Tissue) were in good agreement with certified values. The precision (%RSD) of the measurements was better than 14% ($n = 7$). The relatively high RSD values probably resulted from heterogeneity of the materials at the levels analyzed and/or nonrepresentative sampling of the slurry samples. DLs (3σ), which were estimated from standard addition curves, were in the range of 5–50 ng g^{-1} (0.05–5 ng mL^{-1}) for Cu, 200–500 ng g^{-1} (2–5 ng mL^{-1}) for Zn, 8–20 ng g^{-1} (0.08–0.2 ng mL^{-1}) for Cd, and 20–50 ng g^{-1} (0.2–0.5 ng mL^{-1}) for Pb.

Physical carriers such as Pd, C, $Mg(NO_3)_2$, and NaCl are commonly used to improve analyte transport in ETV/ICP/MS. This principle is identical with what is observed when slurry samples and aqueous samples are compared – enhanced signals are observed with the slurry samples because the sample matrix improves analyte transport. The benefit of using physical carriers for quantification in ETV/ICP/MS is that the transport can be normalized because the physical carrier rather than the sample matrix controls the efficiency of analyte transport, providing similar sensitivities for both aqueous standards and slurry samples. The use of Pd as a physical carrier in quantitative ETV/ICP/MS measurements has been investigated in this laboratory.[35] When combined with oxygen ashing (750 °C for 40 s with 300 mL min^{-1} air) for the removal of carbonaceous material, the addition of 1 μg of Pd as a physical carrier allowed the determination of V, Mn, Ni, Cu, and Pb in NIST SRM 1548 Total Diet, SRM 1566a Oyster Tissue and LUTS-1 Lobster Hepatopancreas using aqueous standardization. The use of larger quantities of Pd (up to 10 μg) led to space–charge effects, which reduced the signal intensities by as much as 80%.

4 FOOD COMPOSITION LABORATORY METHODS

Techniques and methods used for the determination of elements of nutritional interest in the FCL of the US Department of Agriculture are discussed in this section. Instrumental techniques commonly employed in this laboratory include FAAS, GFAAS, ICP/AES, and ICP/MS. Method selection depends on several factors, including the number of analytes to be determined and the required detection capability. Specific methods for analytes typically measured by each of the techniques are provided.

4.1 Sample Preparation

Both dry and wet ash sample preparations are commonly used in this laboratory. In most instances, either method can be used successfully [e.g. see Miller-Ihli[58]]; however, there are particular instances where one method is preferred over the other. For example, if large numbers of samples are to be analyzed, wet ashing is preferred because the samples can be simultaneously digested using a heating block. With dry ashing, the muffle furnace can accommodate only a relatively small number of samples at one time. A benefit of dry ashing is that the sample does not have to be diluted as much, offering the potential for greater sensitivity.

4.1.1 Dry Ashing

Approximately 0.5–2 g of homogenized sample is weighed into acid-cleaned quartz or borosilicate tubes and placed in a muffle furnace (Lindberg, Wattertown, WI, USA). The temperature is slowly increased to 480 °C and kept there overnight. If the sample ash is white, the sample is diluted to a final volume of 15 mL using 5% v/v sub-boiling distilled HNO_3 (Seastar, Sidney, BC, Canada). If the ash is not white, the sample is treated with 1 mL of concentrated HNO_3, taken to dryness on a heating block and returned to the muffle furnace at 480 °C overnight. If necessary, the nitric acid treatment is repeated until a white ash is obtained. The sample ash is diluted to a final volume of 15 mL using 5% v/v sub-boiling distilled HNO_3 (a mixture of 5% v/v HNO_3 and 5% v/v HCl is used for ICP/AES determinations) and stored in acid-cleaned polyethylene test tubes until analyzed. Additional dilutions might be required depending on the analyte and the measurement technique. It should be noted that when analyzing samples by GFAAS, 600 μg

of $Mg(NO_3)_2$ are added as a modifier (ashing aid) during the dry ashing procedure.

4.1.2 Wet Ashing

Approximately 0.5–2 g of homogenized sample is placed into acid-cleaned silanized quartz or borosilicate tubes and 1–2 mL of water are added to make a slurry. To the slurry, 1 mL of sub-boiling distilled HNO_3 (Seastar) is added. Test tubes are placed on a Multiblock heater (Lab-Line Instruments, Melrose Park, IL, USA) and heated overnight at 80 °C. The next day, samples are treated with 1 mL of 50% H_2O_2 (Fisher, Fairlawn, NJ, USA) added dropwise, and the mixture is heated at 100 °C for 5–6 h. The peroxide treatment is repeated during this time until the digests are clear. The maximum amount of peroxide added should never exceed 5 mL. Digests are then heated overnight at 80 °C. A 1-mL volume of concentrated HCl (Ultrex II, J.T. Baker) is then added to the tube (for improved extraction of Fe) and the mixture is heated at 80 °C for 3–4 h. For GFAAS determinations, HCl is not used because volatile chlorides can form and lead to poor recoveries. Samples are filtered using ashless 7-cm No. 41 filter papers (Whatman, Maidstone, UK), diluted to a final volume of 15 mL using 5% v/v HNO_3 (a mixture of 5% v/v HNO_3 and 5% v/v HCl is used for ICP/AES determinations), and stored in acid-cleaned polyethylene test tubes until analyzed. Additional dilutions might be required depending on the analyte and the measurement technique.

4.2 Flame Atomic Absorption Spectrometry

FAAS determinations are made with a Perkin-Elmer Model 5100 PC spectrophotometer equipped with both a deuterium arc lamp for background correction in the ultraviolet (UV) region (180–300 nm) and a tungsten lamp for background correction in the visible region (300–900 nm). Elements typically measured by FAAS in this laboratory are Ca, Cu, Fe, Mg, Mn, and Zn.[58] An air–acetylene flame is used in all cases. Most often, the FAAS technique is applied only when a small number ($\leq$3) of analytes are being determined, since more rapid multielement techniques (ICP/AES and ICP/MS) are available. Instrumental conditions for FAAS determinations are provided in Table 3. Note that Ca was determined with the use of 8-hydroxyquinoline as a protecting agent to prevent the formation of refractory solute species. No modifiers (releasing agents, ionization suppressors, etc.) are used for the other analytes.

4.3 Graphite Furnace Atomic Absorption Spectrometry

When the detection capabilities of FAAS and ICP/AES are not adequate, GFAAS is frequently used for the trace determination of elements of nutritional interest. Commonly, Cr levels are low ($<1\ \mu g\ g^{-1}$) in food samples and require the use of GFAAS. For GFAAS determinations, a Perkin-Elmer 5100 PC spectrophotometer with Zeeman background correction is used. The system includes an AS-60 autosampler and a USS-100 ultrasonic slurry mixer for slurry measurements. Pyrolytic graphite-coated platform-containing tubes are used. Peak-area measurements are used for quantification and calibration is typically performed with aqueous standards. For the analysis of foods and other carbonaceous materials, oxygen ashing is used to remove organic material prior to atomization. Instrumental conditions for the determination of Cr by GFAAS are provided in Table 4.

4.4 Inductively Coupled Plasma Atomic Emission Spectrometry

ICP/AES is routinely used for the determination of a suite of elements (Na, K, Ca, Co, Cu, Cr, Fe, Mg, Mn, Ni, P, V, and Zn) that are of nutritional interest.[58] A typical ICP/AES system is a Leeman Labs (Lowell, MA, USA) PS3000 ICP, which is a combination simultaneous

Table 3 Operating conditions for FAAS determinations [adapted from Miller-Ihli[58]]

Element	Wavelength (nm)	HCL current (mA)	Slit	Gas flow rates ($mL\ min^{-1}$)		Range of standard concentrations ($\mu g\ mL^{-1}$)
				Acetylene	Air	
Ca	422.7	10	0.7	3.5	10.0	0.5–4.0
Cu	324.8	15	0.7	2.6	10.0	0.1–2.0
Fe	248.3	30	0.2	2.5	10.0	0.2–4.0
Mg	285.2	6	0.7	2.5	10.0	0.1–2.0
Mn	279.5	20	0.2	2.8	10.0	0.1–2.0
Zn	213.9	15	0.7	2.4	10.0	0.1–2.0
Sample uptake rate		$5\ mL\ min^{-1}$				
Replicates		3				
Measurement time		0.5 s				

Table 4 Operating conditions for the GFAAS determination of Cr

Wavelength	357.9 nm
HCL current	25 mA
Slit	0.7
Modifier	60 μg $Mg(NO_3)_2$
Range of standard concentrations	5–50 ng mL^{-1}
Replicates	3
Measurement time	8 s

Furnace program:

Temperature (°C)	Ramp time (s)	Hold time (s)	Ar gas flow rate (mL min^{-1})	Step
200	40	40	300	Drying
750	20	40	300 (air)	Oxygen ashing
20	1	50	300	Ar purge
2400	0	10	10	Atomization
2700	1	5	300	Clean-out

and sequential high-resolution, échelle-based system. Ca, Co, Cu, Cr, Fe, Mg, Mn, Ni, P, V, and Zn are typically measured in the simultaneous mode, whereas, Na and K are measured sequentially. Operating conditions for ICP/AES determinations are provided in Table 5. Optimization of the plasma viewing location (vertical and horizontal) is based on the use of a 10 μg mL^{-1} solution of Mn (257.6 nm) as the peaking element. In this routine, the mirror that focuses the image of the plasma on the entrance slit is moved to optimize the emission signal. A ''peak optics'' routine, which adjusts the aperture plate to locate the 296.73-nm Hg reference line, runs automatically every 20 min. Background correction is employed using off-line wavelengths selected on the basis of a wide range of sample scans.

Table 5 Operating conditions for ICP/AES determinations [adapted from Miller-Ihli[58]]

ICP:	
Frequency	40 MHz
Rf power	1.0–1.1 kW
Coolant gas flow rate	11–12 L min^{-1}
Auxiliary gas flow rate	0.3–0.5 L min^{-1}
Nebulizer pressure	40–45 psi
Nebulizer type	Hildebrand grid nebulizer
Measurement parameters:	
Mode	Simultaneous except for Na and K (measured in sequential mode)
Sample uptake rate	1.5 mL min^{-1}
Integration time	0.3 s
Replicates	3

Element	Wavelength[a] (nm)	Range of standard concentrations (μg mL^{-1})
Na	589.6 I	5–500
K	766.5 I	5–500
Ca	317.9 II	5–500
Co	228.6 II	0.1–10
Cu	324.8 I	0.1–10
Cr	267.7 II	0.1–10
Fe	259.9 II	1–100
Mg	279.1 II	5–500
Mn	257.6 II	0.1–10
Ni	231.6 II	0.1–10
P	214.9 I	5–500
V	310.2 II	0.1–10
Zn	213.9 II	0.1–10

[a] I and II designate whether spectral lines originate from neutral atom or singly ionized state, respectively.

4.5 Inductively Coupled Plasma Mass Spectrometry

A method for the determination of 13 elements (Na, Mg, P, K, Ca, V, Cr, Fe, Mn, Co, Ni, Zn, and Cu) in food samples using ICP/MS detection has recently been developed in this laboratory.[59] A Perkin-Elmer SCIEX (Thornhill, ON, Canada) Elan 6000 ICP/MS instrument, equipped with a standard cross-flow nebulizer and Scott-type, double-pass Ryton® spray chamber, is used. Operating conditions for the ICP/MS measurements are given in Table 6. Because matrix effects related to high levels of matrix elements can be significant when analyzing food digests (signal suppressions of 50% are common), an internal standard is used. Ga (50 ng mL^{-1}) has proven suitable as an internal standard for the full suite of elements studied. A correction equation for the isobaric interference of $^{40}Ca^{16}O^{1}H$ on ^{57}Fe is determined. The correction factor is obtained by measuring a 50 μg mL^{-1} solution of Ca and calculating the ratio of the signals at *m*/*z* 60 ($^{43}Ca^{16}O^{1}H$) and *m*/*z* 43 (Ca). The extent of the CaOH interference is measured at *m*/*z* 60 and 43, as opposed to *m*/*z* 57 and 40, because of the inability to measure the Ca signal at *m*/*z* 40 owing to the very large Ar background. The correction value is typically around 3.5%.

Table 6 Operating conditions for ICP/MS determinations [adapted from Baker and Miller-Ihli[59]]

ICP:	
Frequency	40 MHz
Rf power	1.2 kW
Plasma gas flow rate	15 L min^{-1}
Auxiliary gas flow rate	0.9 L min^{-1}
Nebulizer gas flow rate[a]	0.9–1.0 L min^{-1}
Measurement parameters:	
Scan mode	Peak hopping
Lens scanning[b]	Enabled
Detector mode[c]	Dual – pulse and analog
Dwell time	20 ms
Sweeps per reading	60
Replicates	3

Element	Isotope	Range of standard concentrations[d]
Na	23	0.5–30 μg mL^{-1}
Mg	24	0.5–30 μg mL^{-1}
P	31	0.5–30 μg mL^{-1}
K	39	0.5–30 μg mL^{-1}
Ca	43	0.5–30 μg mL^{-1}
V	51	5–500 ng mL^{-1}
Cr	53	5–500 ng mL^{-1}
Mn	55	5–500 ng mL^{-1}
Fe	57	5–500 ng mL^{-1}
Co	59	5–500 ng mL^{-1}
Ni	60	5–500 ng mL^{-1}
Cu	63	5–500 ng mL^{-1}
Zn	66	5–500 ng mL^{-1}
Ga	69	5–500 ng mL^{-1}

[a] Adjusted to maximize ion signals for 10 ng mL^{-1} tuning solution of Mg, Rh, Ce, Ba, and Pb. Adjusted for CeO^+/Ce^+ <3% and Ba^{2+}/Ba^+ <2%.

[b] Lens scanning (AutoLens™) calibration performed with a 10 ng mL^{-1} solution of Be, Co, and In containing 25 μg mL^{-1} of Na, K, and Ca.

[c] Detector cross-calibration (Dual Detector™ mode) performed using a multielement solution containing levels of analytes that produced signals in pulse counting mode of approximately 200 000–1 000 000 cps (typically 200 ng mL^{-1}–2 μg mL^{-1}).

[d] Standards for minor elements (5–500 ng mL^{-1}) also contained 30 μg mL^{-1} of Na, Mg, K, P, and Ca for improved matrix matching.

5 SPECIATION

For essential elements to exert an effect on human health, they must be available from food both for absorption and subsequent utilization by the body (i.e. they must be bioavailable). It is now generally accepted that the total amount of an element present in food tells little of how efficiently it will be absorbed and used. A major influence on the bioavailability of an element is the chemical form or speciation of that particular element. Measurement of chemical speciation can refer to the determination of oxidation states, metal–ligand complexes, and mineral compounds.[60] An excellent overview of methods for trace metal speciation and problems associated with the techniques is given in a book edited by Krull.[61] The combination of chromatographic separation and atomic spectroscopic detection provides a powerful, versatile tool for trace element speciation measurements. In recent years, there have been a large number of publications dealing with chemical speciation measurements in foods and biological samples using chromatogaphy combined with atomic spectroscopy. These include the determination of metal–carbohydrate complexes in fruits and vegetables by size-exclusion high-performance liquid chromatography (HPLC) with ICP/MS detection,[62] multielement speciation of tea infusion using size-exclusion chromatography (SEC) ICP/MS,[63] and Se speciation measurements in enriched garlic, onion, and broccoli samples.[64]

Methods for the speciation of Se compounds have often been described in the literature. Emteborg et al. developed a method for the separation and detection of selenomethionine, selenite, selenate, and selenocystine using microbore ion chromatography (IC) coupled to GFAAS.[65] Automated collection of 20-μL fractions in autosampler cups was used for interfacing the chromatographic separation with GFAAS. With a column flow rate of 80 μL min^{-1}, the system provided a time resolution of 15 s, which was adequate for the separation of the selenium compounds. For comparison of chromatographic resolution, IC direct injection nebulization ICP/AES was also used. According to the authors, the trade-off in chromatographic resolution and time consumption in the detection step (approximately 2 min per fraction) using IC/GFAAS was compensated for by a high degree of simplicity and the high specificity and sensitivity of GFAAS. DLs of 2.8–4.1 ng mL^{-1} (42–61 pg absolute) were reported and the repeatability ranged from 4.3 to 6.2% RSD ($n = 4$) for the four forms of selenium studied.

Direct coupling of HPLC to ICP/MS is a convenient and sensitive method for Se speciation measurements.[64,66,67] As part of a study of Se bioavailability from the human diet, Crews et al. developed an HPLC/ICP/MS method for the determination of Se species in in vitro gastrointestinal extracts of cooked cod.[66] The method used pepsin and HCl (pH 2.0) to simulate gastric digestion in the stomach, and pancreatin, amylase, and bile salts for simulating digestion in the small intestine (pH 6.9). Total Se was measured in cooked cod and in extracts by ICP/MS, following digestion in high-pressure digestion vessels. For speciation measurements, the HPLC effluent was introduced directly to the cross-flow nebulizer of the ICP/MS system at a flow rate of 0.75 mL min^{-1}. HPLC/ICP/MS allowed the separation of four Se standards (selenomethionine, selenocystine, sodium selenite, and sodium selenate) in less than

12 min with a reproducibility of retention times of better than 5% RSD. The total concentration of Se (measured as ^{82}Se) in cooked cod was determined to be $1.52 \pm 0.07\,\mu g\,g^{-1}$ and the amount of Se extracted in the gastric supernatant and gastrointestinal supernatant was 50.5 and 61.3%, respectively. HPLC/ICP/MS analysis of the extract revealed two Se species, selenite and an unidentified form that was believed to be an organic form because of the similarity in retention time to selenomethionine. Approximately 12% of the total Se was present as selenite. It should be noted that the instrumental reponse (signal per mass of ^{82}Se) was species specific: the inorganic forms gave approximately equal responses but the responses for selenomethionine and selenocystine were significantly lower (40 and 55% respectively). The authors suggested that this could be due to either suppression of the Se signal for the organic standards or retention of the Se from the organic standards by the chromatography system. What this means is that quantification of individual species can only be achieved if standards for that particular form are used.

In the future, there is no doubt that the role of speciation measurements in food samples will continue to expand. As analytical methodology is developed and refined to meet the requirements of a particular analysis, the role of trace element species in human health can be more clearly defined. Atomic spectroscopic detection, especially very sensitive MS techniques, will certainly play a dominant role in the field of chemical speciation measurements.

6 QUALITY ASSURANCE, METHOD VALIDATION, AND REFERENCE MATERIALS

General aspects of QA are highlighted here. There are three general aspects to validating analytical data: (1) development of a method of analysis; (2) validation of the method; and (3) ensuring that the method remains valid throughout its use.[68]

Development of a reliable analytical method requires that the strengths, limitations, and ruggedness of the analytical aspects be documented by appropriate research and development. Potential matrix effects must be identified for each sample matrix. In some instances, interferences are analyte specific requiring that different instrumental parameters or sample preparation methods may be necessary for both different matrices and for different elements. Such analytical variables must be identified and taken into consideration in the development of an accurate analytical procedure.

Once the method has been developed it must be validated, ensuring that the method produces data which are in agreement with the ''true value'' of the analyte in the samples.[69] This is accomplished by doing calibration with method of additions (to check for matrix interference effects) or by running a well-characterized QC material or CRM, to see if the method is producing accurate results.[70] There are several sources of RMs suitable for food analyses, including the International Atomic Energy Agency (IAEA), the National Research Council of Canada (NRCC), and NIST. Perhaps the best known is NIST, which prepares geological, botanical, inorganic, and organic CRMs which they refer to as SRMs and which come with certificates ''certifying'' specified analyte contents and providing information about the particle size of the material and information on its handling, processing, and packaging.

Table 7 contains data for a wide range of commercial RMs suitable for elemental analyses of food samples. Reference values for as many as 24 elements are available for a total of 34 food-related CRMs, which is a significant improvement compared with 5 years ago. More materials are needed, however, which are more similar to foods as consumed (not freeze-dried) to serve as QC materials for large studies.

Large-scale production of CRMs is very expensive and it is extremely time-consuming to certify a candidate material. In the case of foods, CRMs are needed with certified values for proximates, organic constituents (e.g. vitamins), and also trace elements. Because of the wide range of needs for the variety of analytes of interest, it is typically not feasible to make a single material (e.g. diet material) which will be certified over a wide range of analytes. It does not make sense, for example, to have a single material certified for proximates (ash, fat, and moisture) as well as trace elements because of the cost of certifying a material. It is simply too expensive to justify the use of several grams of a material certified for a range of trace elements to be used for moisture determinations. As a result, it is critical to have a range of suitable CRMs for the variety of analytes of interest.

Most laboratories involved in large-scale studies with hundreds of determinations utilize secondary QC samples for ongoing QC. The CRMs are used for method validation and then in-house QC samples are developed which usually consist of a particularly large amount of an individual sample or several samples that are pooled together. Reference concentrations for the various analytes of interest are typically determined through cooperation with at least one additional laboratory. After being characterized, the QC samples are used in several ways. They may be analyzed in conjunction with unknown samples to verify sample preparation procedures and calibration. QC samples can be labeled or analyzed

Table 7 List of food-related RMs (all concentrations in $\mu g\,g^{-1}$ dry weight unless indicated otherwise)

FCL ID	RM	Manufacturer[a]	Al	As	B
R-0010	Rice Flour (SRM 1568)	NIST		0.41 ± 0.05	
R-0011	Rice Flour (SRM 1568a)	NIST	4.4 ± 1.0	0.29 ± 0.03	
R-0020	Wheat Flour (SRM 1567)	NIST		(0.006)	
R-0021	Wheat Flour (SRM 1567a)	NIST	5.7 ± 1.3	(0.006)	
R-0030	Total Diet (SRM 1548)	NIST	(33)		(2.5)
R-0040	Carrot Powder (ARC/CL-CP)	ARC/CL			10.1 ± 0.58
R-0050	Skim Milk Powder (ARC/CL-MP)	ARC/CL			
R-0060	Pork	ARC/CL			
R-0070	Wheat Flour (ARC/CL-WF)	ARC/CL			
R-0080	Potato	ARC/CL			
R-0090	Total Diet (ARC/CL-HDP)	ARC/CL			
R-0100	Skim Milk (ARC/CL-SM)	ARC/CL			
R-0110	Mixed Diet RM 8431	NIST	4.39 ± 1.07	0.924 ± 0.344	
R-0120	Nonfat Milk Powder SRM 1549	NIST	(2)	(0.002)	
R-0130	Milk Powder IAEA A-11	IAEA	(1.3)	(0.048)	(1.7)
R-0140	Oyster Tissue SRM 1566a	NIST	202.5 ± 12.5	14.0 ± 1.2	
R-0150	Bovine Liver SRM 1577a	NIST	(2)	0.047 ± 0.006	
R-0151	Bovine Liver SRM 1577b	NIST	(3)	(0.05)	
R-0160	Dogfish Muscle DORM1	NRCC		17.7 ± 2.1	
R-0170	Dogfish Liver DOLT1	NRCC		10.1 ± 1.4	
R-0180	Dogfish Muscle DORM2	NRCC	10.9 ± 1.7	18.0 ± 1.1	
R-0190	Dogfish Liver DOLT2	NRCC	25.2 ± 2.4	16.6 ± 1.1	
R-0210	Bovine Muscle Powder (SRM 8414)	NIST	1.7 ± 1.4	0.009 ± 0.003	0.6 ± 0.4
R-0220	Nondefatted Lobster Hepatopancreas [LUTS-1 (as bottled)]	NRCC		2.83 ± 0.13	
R-0230	Nondefatted Lobster Hepatopancreas [LUTS-1 (dry weight)]	NRCC		19.0 ± 0.9	
R-0240	Lobster Hepatopancreas (TORT-1)	NRCC		24.6 ± 2.2	
R-0250	Spinach Leaves (SRM 1570)	NIST	870 ± 50	0.15 ± 0.05	(30)
R-0251	Spinach Leaves (SRM 1570a)	NIST	310 ± 11	0.068 ± 0.012	37.6 ± 1.0
R-0260	Apple Leaves (SRM 1515)	NIST	286 ± 9	0.038 ± 0.007	27 ± 2
R-0270	Citrus Leaves (SRM 1572)	NIST	92 ± 15	3.1 ± 0.3	
R-0280	Peach Leaves (SRM 1547)	NIST	249 ± 8	0.060 ± 0.018	29 ± 2
R-0290	Fish Tissue (MA-B-3/TM)	IAEA	20.9 [14.8–36.5]	2.11 [1.42–2.51]	
R-0303	Water Sample (SRM 1643d) ($\mu g\,mL^{-1}$)	NIST	0.01276 ± 0.0035	0.05602 ± 0.00073	0.1448 ± 0.0052
R-0310	Infant Formula (SRM 1846)	NIST			

FCL ID	Number	Ca	Cd	Co	Cr	Cu	F
R-0010	1568	140 ± 20	0.029 ± 0.004	0.02 ± 0.01		2.2 ± 0.3	
R-0011	1568a	118 ± 6	0.022 ± 0.002	(0.018)		2.4 ± 0.3	
R-0020	1567	190 ± 10	0.032 ± 0.007			2.0 ± 0.3	
R-0021	1567a	191 ± 4	0.026 ± 0.002	(0.006)		2.1 ± 0.2	
R-0030	1548	1740 ± 70	0.028 ± 0.004			2.6 ± 0.3	
R-0040		1700 ± 45	0.0686 ± 0.0031				
R-0050		13 040 ± 260	(<0.005)		0.0079 ± 0.0032	0.58 ± 0.06	
R-0060		150 ± 6	0.022 ± 0.0042			2.68 ± 0.283	
R-0070		208 ± 4	0.039 ± 0.004		0.028 ± 0.005	2.35 ± 0.18	
R-0080		91 ± 6	0.035 ± 0.0016		0.098 ± 0.018	3.87 ± 0.18	
R-0090		2860 ± 124	0.021 ± 0.003			3.18 ± 0.19	
R-0100		13 000 ± 291	(<0.005)			0.59 ± 0.148	
R-0110	8431	1940 ± 140	0.042 ± 0.011	0.038 ± 0.008	0.102 ± 0.006	3.36 ± 0.33	
R-0120	1549	13 000 ± 500	0.0005 ± 0.0002	(0.004)	0.0026 ± 0.0007	0.7 ± 0.1	(0.20)
R-0130	A-11	12 900 ± 800	(0.526)	0.005 ± 0.001	(0.257)	0.838 ± 0.165	(0.26)
R-0140	1566a	1960 ± 190	4.15 ± 0.38	0.57 ± 0.11	1.43 ± 0.46	66.3 ± 4.3	(240)
R-0150	1577a	120 ± 7	0.44 ± 0.06	0.21 ± 0.05		158 ± 7	
R-0151	1577b	116 ± 4	0.50 ± 0.03	(0.25)		160 ± 8	
R-0160	DORM1		0.086 ± 0.012	0.049 ± 0.014	3.60 ± 0.40	5.22 ± 0.33	
R-0170	DOLT1		4.18 ± 0.28	0.157 ± 0.037	0.40 ± 0.07	20.8 ± 1.2	
R-0180	DORM2		0.043 ± 0.008	0.182 ± 0.031	34.7 ± 5.5	2.34 ± 0.16	

Table 7 (*continued*)

FCL ID	Number	Ca	Cd	Co	Cr	Cu	F
R-0190	DOLT2		20.8 ± 0.5	0.24 ± 0.05	0.37 ± 0.08	25.8 ± 1.1	
R-0210	8414	145 ± 20	0.013 ± 0.011	0.007 ± 0.003	0.071 ± 0.038	2.84 ± 0.45	(0.22)
R-0220	LUTS-1	203 ± 33	2.12 ± 0.15	0.051 ± 0.006	0.079 ± 0.012	15.9 ± 1.2	
R-0230	LUTS-1	1360 ± 220	14.2 ± 1.0	0.34 ± 0.04	0.53 ± 0.08	107 ± 8	
R-0240	TORT-1	8950 ± 580	26.3 ± 2.1	0.42 ± 0.05	2.4 ± 0.6	439 ± 22	
R-0250	1570	13 500 ± 300	(1.5)	(1.5)	4.6 ± 0.3	12 ± 2	
R-0251	1570a	15 270 ± 410	2.89 ± 0.07	0.39 ± 0.05		12.2 ± 0.6	
R-0260	1515	15 260 ± 150	(0.014)	(0.09)	(0.3)	5.64 ± 0.24	
R-0270	1572	31 500 ± 1000	0.03 ± 0.01	(0.02)	0.8 ± 0.2	16.5 ± 1.0	
R-0280	1547	15 600 ± 200	0.026 ± 0.003	(0.07)	(1)	3.7 ± 0.4	
R-0290	MA-B-3/TM	0.00349 [0.00318–0.00360]			0.64 [0.45–0.94]	3.08 [2.85–3.57]	
R-0303	1643d	31.04 ± 0.05	0.00647 ± 0.00037	0.02500 ± 0.00059	0.1853 ± 0.00020	0.0205 ± 0.0038	
R-0310	1846	3670 ± 200				5.04 ± 0.27	

FCL ID	Number	Fe	I	K	Mg	Mn	Mo
R-0010	1568	8.7 ± 0.6		1120 ± 20		20.1 ± 0.4	(1.6)
R-0011	1568a	7.4 ± 0.9	(0.01)	1280 ± 8	560 ± 20	20.0 ± 1.6	1.46 ± 0.08
R-0020	1567	18.3 ± 1.0		1360 ± 40		8.5 ± 0.5	(0.4)
R-0021	1567a	14.1 ± 0.5	(0.001)	1330 ± 30	400 ± 20	9.4 ± 0.9	0.48 ± 0.03
R-0030	1548	32.6 ± 3.6		6060 ± 280	556 ± 27	5.2 ± 0.4	(0.27)
R-0040		12.6 ± 0.79		10 200 ± 350	350 ± 9	4.90 ± 0.17	0.0495 ± 0.0024
R-0050		4.53 ± 0.61	1.74 ± 0.07	17 700 ± 590	1230 ± 24	0.45 ± 0.07	0.49 ± 0.03
R-0060		52.84 ± 5.456			(1010)	0.30 ± 0.024	0.034 ± 0.016
R-0070		51 ± 3.79		2200 ± 190	562 ± 10	12.87 ± 0.34	0.25 ± 0.028
R-0080		22.0 ± 2.0		14 100 ± 100	747 ± 16	8.1 ± 0.32	0.212 ± 0.049
R-0090		30.4 ± 0.9		9420 ± 300	785 ± 25	12.9 ± 0.58	0.262 ± 0.035
R-0100		4.54 ± 0.855			1230 ± 27	0.44 ± 0.104	0.49 ± 0.032
R-0110	8431	37.0 ± 2.6		7900 ± 420	650 ± 40	8.12 ± 0.31	0.288 ± 0.029
R-0120	1549	1.78 ± 0.10	3.38 ± 0.02	16 900 ± 300	1200 ± 30	0.26 ± 0.06	(0.34)
R-0130	A-11	3.65 ± 0.76	(1.5)	17 200 ± 1000	1100 ± 80	0.373 ± 0.081	(1.3)
R-0140	1566a	539 ± 15	4.46 ± 0.42	7900 ± 470	1180 ± 170	12.3 ± 1.5	
R-0150	1577a	194 ± 20		9960 ± 70	600 ± 15	9.9 ± 0.8	3.5 ± 0.5
R-0151	1577b	184 ± 15		9940 ± 20	601 ± 28	10.5 ± 1.7	3.5 ± 0.3
R-0160	DORM1	63.6 ± 5.3		15 900 ± 1000	1210 ± 130	1.32 ± 0.26	
R-0170	DOLT1	712 ± 48		10 100 ± 1000	1100 ± 150	8.72 ± 0.53	
R-0180	DORM2	142 ± 10				3.66 ± 0.34	
R-0190	DOLT2	1103 ± 47				6.88 ± 0.56	
R-0210	8414	71.2 ± 9.2	0.035 ± 0.012	15 170 ± 370	960 ± 95	0.37 ± 0.09	0.08 ± 0.06
R-0220	LUTS-1	11.6 ± 0.9		948 ± 72	89.5 ± 4.1	1.20 ± 0.13	
R-0230	LUTS-1	77.8 ± 6.0		6360 ± 480	601 ± 28	8.02 ± 0.86	
R-0240	TORT-1	186 ± 11		10 410 ± 400	2550 ± 250	23.4 ± 1.0	1.5 ± 0.3
R-0250	1570	550 ± 20		35 600 ± 300		165 ± 6	
R-0251	1570a			29 030 ± 520	(8900)	75.9 ± 1.9	
R-0260	1515	(80)	(0.3)	16 100 ± 200	2710 ± 80	54 ± 3	0.094 ± 0.013
R-0270	1572	90 ± 10	1.84 ± 0.03	18 200 ± 600	5800 ± 300	23 ± 2	0.17 ± 0.09
R-0280	1547	218 ± 14	(0.3)	24 300 ± 300	4320 ± 80	98 ± 3	0.060 ± 0.008
R-0290	MA-B-3/TM	95.4 [87.4–107.2]		0.00932 [0.0090–0.0100]	0.00113 [0.00104–0.00120]	2.62 [2.22–3.03]	
R-0303	1643d	0.0912 ± 0.0039		2.356 ± 0.035	7.989 ± 0.035	0.03766 ± 0.00083	0.1129 ± 0.0017
R-0310	1846	63.1 ± 4.0		7160 ± 380	538 ± 29	(0.4)	

FCL ID	Number	Na	Ni	P	Pb	S
R-0010	1568	6.0 ± 1.5	(0.16)			
R-0011	1568a	6.6 ± 0.8		1530 ± 80	(<0.010)	1200 ± 20
R-0020	1567	8.0 ± 1.5	(0.18)			
R-0021	1567a	6.1 ± 0.8		1340 ± 60	(<0.020)	1650 ± 20
R-0030	1548	6250 ± 260	(0.41)	3240 ± 40	(0.05)	2580 ± 260

(*continued overleaf*)

Table 7 (*continued*)

FCL ID	Number	Na	Ni	P	Pb	S
R-0040		485 ± 24				
R-0050		4870 ± 350	0.059 ± 0.022		0.016 ± 0.003	
R-0060					0.089 ± 0.013	
R-0070			0.153 ± 0.009	2090 ± 140	0.018 ± 0.007	
R-0080		895 ± 42	0.193 ± 0.043	2370 ± 100	0.026 ± 0.0028	
R-0090		7870 ± 570	0.271 ± 0.038		0.043 ± 0.008	
R-0100					0.016 ± 0.003	
R-0110	8431	3120 ± 160	0.644 ± 0.151	3320 ± 310		
R-0120	1549	4970 ± 100		10 600 ± 200	0.019 ± 0.003	3510 ± 50
R-0130	A-11	4420 ± 330	(0.93)	9100 ± 1020	(0.27)	(18 000)
R-0140	1566a	4170 ± 130	2.25 ± 0.44	6230 ± 180	0.371 ± 0.014	8620 ± 190
R-0150	1577a	2430 ± 130		11 100 ± 400	0.135 ± 0.015	7800 ± 100
R-0151	1577b	2420 ± 60		11 000 ± 300	0.129 ± 0.004	7850 ± 60
R-0160	DORM1	8000 ± 600	1.20 ± 0.30		0.40 ± 0.12	
R-0170	DOLT1	7260 ± 730	0.26 ± 0.06		1.36 ± 0.29	
R-0180	DORM2		19.4 ± 3.1		0.065 ± 0.007	
R-0190	DOLT2		0.20 ± 0.02		0.22 ± 0.02	
R-0210	8414	2100 ± 80	0.05 ± 0.04	8360 ± 450	0.38 ± 0.24	7950 ± 410
R-0220	LUTS-1		0.200 ± 0.034		0.010 ± 0.002	
R-0230	LUTS-1		1.34 ± 0.23		0.069 ± 0.011	
R-0240	TORT-1	36 700 ± 2000	2.3 ± 0.3	8790 ± 210	10.4 ± 2.0	12 200 ± 1000
R-0250	1570		(6)	5500 ± 200	1.2 ± 0.2	
R-0251	1570a	18 180 ± 430	2.14 ± 0.10	5180 ± 110	(200)	(4600)
R-0260	1515	24.4 ± 1.2	0.91 ± 0.12		0.470 ± 0.024	(1800)
R-0270	1572	160 ± 20	0.6 ± 0.3	1300 ± 200	13.3 ± 2.4	4070 ± 90
R-0280	1547	24 ± 2	0.69 ± 0.09	1370 ± 70	0.87 ± 0.03	(2000)
R-0290	MA-B-3/TM	0.00216 [0.00200–0.00231]		0.00853 [0.00753–0.00920]	4.62 [3.85–5.13]	
R-0303	1643d	22.07 ± 0.64	0.0581 ± 0.0027		0.01815 ± 0.00064	
R-0310	1846	2310 ± 130		2610 ± 150		

FCL ID	Number	Se	Sn	V	Zn
R-0010	1568	0.4 ± 0.1			19.4 ± 1.0
R-0011	1568a	0.38 ± 0.04	(0.005)	(0.007)	19.4 ± 0.5
R-0020	1567	1.1 ± 0.2			10.6 ± 1.0
R-0021	1567a	1.1 ± 0.2	(0.0033)	(0.011)	11.6 ± 0.4
R-0030	1548	0.245 ± 0.005	(3.6)		30.8 ± 1.1
R-0040					6.9 ± 0.15
R-0050		0.082 ± 0.0072			41.7 ± 1.00
R-0060		0.0394 ± 0.031			103.80 ± 3.005
R-0070		0.057 ± 0.00545			14.6 ± 0.71
R-0080					9.0 ± 0.32
R-0090		0.181 ± 0.017			28.9 ± 1.3
R-0100		0.082 ± 0.0077			41.68 ± 1.056
R-0110	8431	0.242 ± 0.030			17.0 ± 0.6
R-0120	1549	0.11 ± 0.01	(<0.02)		46.1 ± 2.2
R-0130	A-11	0.0339 ± 0.0072	(0.122)	(0.10)	38.9 ± 2.3
R-0140	1566a	2.21 ± 0.24	(3)	4.68 ± 0.15	830 ± 57
R-0150	1577a	0.71 ± 0.07		0.099 ± 0.008	123 ± 8
R-0151	1577b	0.73 ± 0.06		(0.123)	
R-0160	DORM1	1.62 ± 0.12			21.3 ± 1.0
R-0170	DOLT1	7.34 ± 0.42			92.5 ± 2.3
R-0180	DORM2	1.40 ± 0.09	(0.023)		25.6 ± 2.3
R-0190	DOLT2	6.06 ± 0.49	(0.13)		85.8 ± 2.5
R-0210	8414	0.076 ± 0.010		(0.005)	142 ± 14
R-0220	LUTS-1	0.641 ± 0.054			12.4 ± 0.8
R-0230	LUTS-1	4.30 ± 0.36			82.9 ± 5.4
R-0240	TORT-1	6.88 ± 0.47	0.139 ± 0.011	1.4 ± 0.3	177 ± 10
R-0250	1570				50 ± 2
R-0251	1570a	0.117 ± 0.0009		0.57 ± 0.03	82 ± 3

Table 7 (*continued*)

FCL ID	Number	Se	Sn	V	Zn
R-0260	1515	0.050 ± 0.009	(<0.2)	0.26 ± 0.03	12.5 ± 0.3
R-0270	1572	(0.025)	(0.24)		29 ± 2
R-0280	1547	0.120 ± 0.009	(<0.2)		17.9 ± 0.4
R-0290	MA-B-3/TM	1.46 [1.35–1.70]			109.2 [106–111.9]
R-0303	1643d	0.01143 ± 0.00017		0.0351 ± 0.0014	0.07248 ± 0.00065
R-0310	1846	(0.08)			

[a] Manufacturers: ARC/CL, Agricultural Research Center of Finland, Central Laboratory.
Values in parentheses are noncertified reference values; values in square brackets are the range of reference values for this material.

''blind'' (unidentified to the analyst). The results of the QC samples are compared with the predetermined ''true'' values and poor accuracy or precision typically leads to rejection of the results for samples run in that series.

Accuracy is concerned with correctness. If a measurement process produces the correct results it is considered to be accurate and the measured value is also accurate. The scatter of values reflects the measurement precision (or reproducibility). In setting up a QC program, it is important to focus on both accuracy and precision and have an effective means of measuring both. It has been stated that a measurement is accurate when the reported value does not differ from the true value and that error in reported values occurs as a result of bias and imprecision. An accurate method is one capable of providing precise and unbiased results (within acceptable limits). Accurate and precise are relative terms.[69]

The DL, defined as three times the standard deviation of the blank (3σ), represents the lowest concentration which can be reliably determined (33% RSD) using a particular method. In FCL we prefer to work at a quantitation limit of five times the DL (which provides a measurement precision of 6.6% RSD). Consider that RM 8431 Mixed Diet will be used as a control material for a study using FAAS and that we are interested in Cu and Mn. If a 1-g sample (dry weight) is diluted to 15 mL using a wet ash digestion, the expected solution concentrations will be Cu $0.224\,\mu g\,mL^{-1}$ and Mn and $0.541\,\mu g\,mL^{-1}$. The flame solution DL for both of these elements is $0.001\,\mu g\,mL^{-1}$ [from Welz[37]], so both will be measured well above both the DL and the quantitation limit. It is critical to realize that analyte concentration may not be equivalent for both unknown samples and CRMs, and that unknown samples are seldom as homogeneous as a carefully prepared CRM. Individual analysts and the food composition community in general have committed to further development of both commercial CRMs and in-house QC materials representing a wide range of food matrices (based on proximate content) so that more representative materials are available for QA and QC.[70] Experimental design must take into consideration the statistical information provided by the data so that the analyst can easily determine the confidence they may have in their data.

DISCLAIMER

Mention of trademark or proprietary products does not constitute a guarantee or warranty of the product by the US Department of Agriculture and does not imply their approval to the exclusion of other products that may also be suitable.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
ARC/CL	Agricultural Research Center of Finland, Central Laboratory
CBR	Community Bureau of Reference
CRM	Certified Reference Material
DL	Detection Limit
EDL	Electrodeless Discharge Lamp
ETAAS	Electrothermal Atomization Atomic Absorption Spectrometry
ETV	Electrothermal Vaporization
FAAS	Flame Atomic Absorption Spectrometry
FAES	Flame Atomic Emission Spectrometry
FCL	Food Composition Laboratory
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HCL	Hollow-cathode Lamp
HEPA	High-energy Particulate Air
HG	Hydride Generation
HPLC	High-performance Liquid Chromatography
IAEA	International Atomic Energy Agency
IC	Ion Chromatography
ICP	Inductively Coupled Plasma

ICP/AES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
ID	Isotope Dilution
MS	Mass Spectrometry
NIST	National Institute of Standards and Technology
NLEA	Nutrition Labeling and Education Act
NMKL	Nordic Committee on Food Analysis
NRCC	National Research Council of Canada
QA	Quality Assurance
QC	Quality Control
RDA	Recommended Dietary Allowance
RM	Reference Material
RSD	Relative Standard Deviation
SEC	Size-exclusion Chromatography
SRM	Standard Reference Material
STPF	Stabilized Temperature Platform Furnace
TMAH	Tetramethylammonium Hydroxide
UV	Ultraviolet

RELATED ARTICLES

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Adulteration Determination • Sample Preparation for Food Analysis, General

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

REFERENCES

1. *Recommended Dietary Allowances*, 10th Edition, National Research Council, National Academy Press, Washington, DC, 1989.
2. *Fed. Regist.*, **58**(3), 632–690, 2066–2964 (1993).
3. H. Greenfield, D.A.T. Southgate, *Food Composition Data, Production, Management, and Use*, Elsevier Applied Science, Barking, Chapter 7, 1992.
4. N.J. Miller-Ihli, W.R. Wolf, 'Characterization of a Diet Reference Material for 17 Elements', *Anal. Chem.*, **58**, 3225–3230 (1986).
5. W.R. Wolf, M. Ihnat, in *Biological Reference Materials: Availability, Uses, and Need for Nutrient Analyses*, ed. W.R. Wolf, Wiley-Interscience, New York, Chapters 6 and 10, 1985.
6. H. El Azouzi, M.L. Cervera, M. de la Guardia, 'Multielemental Analysis of Mussel Samples by Atomic Absorption Spectrometry After Room Temperature Sonication', *J. Anal. At. Spectrom.*, **13**, 533–538 (1998).
7. H. Minami, T. Honjyo, I. Atsuya, 'A New Solid–Liquid Extraction Sampling Technique for Direct Determination of Trace Elements in Biological Materials by Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **51**, 211–220 (1996).
8. J.A. Nóbrega, Y. Gelinas, A. Krushevska, R.M. Barnes, 'Determination of Elements in Biological Materials by Inductively Coupled Plasma Atomic Emission and Mass Spectrometry After Extraction with a Tertiary Amine Reagent', *J. Anal. At. Spectrom.*, **12**, 1239–1242 (1997).
9. P.A. Fecher, I. Goldmann, A. Nagengast, 'Determination of Iodine in Food Samples by Inductively Coupled Plasma Mass Spectrometry After Alkaline Extraction', *J. Anal. At. Spectrom.*, **13**, 977–982 (1998).
10. G. Rädlinger, K.G. Heumann, 'Iodine Determination in Food Samples Using Inductively Coupled Plasma Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **70**, 2221–2224 (1998).
11. D. Pozebon, V.L. Dressler, A.J. Curtius, 'Determination of Trace Elements in Biological Materials by ETV–ICP–MS After Dissolution or Slurry Formation with Tetramethylammonium Hydroxide', *J. Anal. At. Spectrom.*, **13**, 1101–1105 (1998).
12. L. Jorhem, 'Determination of Metals in Foodstuffs by Atomic Absorption Spectrophotometry After Dry Ashing: NMKL Interlaboratory Study of Lead, Cadmium, Zinc, Copper, Iron, Chromium, and Nickel', *J. AOAC Int.*, **76**, 798–813 (1993).
13. K.K. Cook, 'Determination of Selenium in Infant Formula and Enteral Formula by Dry Ash Graphite Furnace Atomic Absorption Spectrophotometry with Deuterium Background Correction', *J. AOAC Int.*, **79**, 1162–1166 (1996).
14. M.L. Cervera, J.C. Lopez, R. Montoro, 'Arsenic Content of Spanish Cows' Milk by Dry Ashing Hydride Generation Atomic Absorption Spectrometry', *J. Dairy Sci.*, **61**, 83–89 (1994).
15. U. Tinggi, C. Reilly, C. Patterson, 'Determination of Manganese and Chromium in Foods by Atomic Absorption Spectrometry After Wet Digestion', *Food Chem.*, **60**, 123–128 (1997).
16. H.A.M.G. Vaessen, C.G. van de Kamp, B. Szteke, 'Aluminum Determination in Food Matrices IUPAC Check Sample Survey of Analytical Performance', *Z. Lebensm.-Unters. Forsch.*, **194**, 456–460 (1992).
17. E.H. Larsen, M.B. Ludwigsen, 'Determination of Iodine in Food-related Certified Reference Materials Using Wet Ashing and Detection by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 435–439 (1997).
18. V.E. Negretti de Brätter, P. Brätter, A. Reinicke, G. Schulze, W.O.L. Alvarez, N. Alvarez, 'Determination of

Mineral and Trace Elements in Total Diet by Inductively Coupled Plasma Atomic Emission Spectrometry: Comparison of Microwave-based Digestion and Pressurized Ashing Systems Using Different Acid Mixtures', *J. Anal. At. Spectrom.*, **10**, 487–491 (1995).

19. D. Amarasiriwardena, A. Krusherska, M. Argentine, R.M. Barnes, 'Vapour-phase Acid Digestion of Micro Samples of Biological Material in a High-temperature, High-pressure Asher for Inductively Coupled Plasma Atomic Emission Spectrometry', *Analyst*, **119**, 1017–1021 (1994).
20. E.J. Gawalko, T.W. Nowicki, J. Barr, R. Tkachuk, S. Wu, 'Comparison of Closed-vessel and Focused Open-vessel Microwave Dissolution for Determination of Cadmium, Copper, Lead, and Selenium in Wheat, Wheat Products, Corn Bran, and Rice Flour by Transverse-heated Graphite Furnace Atomic Absorption Spectrometry', *J. AOAC Int.*, **80**, 379–387 (1997).
21. H. Garraud, M. Robert, C.R. Quetel, J. Szpunar, O.F.X. Donard, 'Focused Microwave-assisted Digestion of Biological Reference Materials for the Determination of Trace Metals by ICP/MS', *At. Spectrosc.*, **17**, 183–185 (1996).
22. M. Deaker, W. Maher, 'Low Volume Microwave Digestion for the Determination of Selenium in Marine Biological Tissues by Graphite Furnace Atomic Absorption Spectroscopy', *Anal. Chim. Acta*, **350**, 287–294 (1997).
23. S. Wu, X. Feng, A. Wittmeier, 'Microwave Digestion of Plant and Grain Reference Materials in Nitric Acid or a Mixture of Nitric Acid and Hydrogen Peroxide for the Determination of Multi-elements by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 797–806 (1997).
24. J. Ziaziaris, J.L. Kacprzak, 'Microwave Digestion–Inductively Coupled Plasma Atomic Emission Spectrometric Determination of Boron in Raw Noodles', *J. AOAC Int.*, **78**, 874–877 (1995).
25. K.E. Levine, J.D. Batchelor, C.B. Rhoades, Jr, B.T. Jones, 'Evaluation of a High Pressure, High Temperature Microwave Digestion System', *J. Anal. At. Spectrom.*, **14**, 49–59 (1999).
26. H.M. Kingston, L.B. Jassie, *Introduction to Microwave Sample Preparation*, ACS Press, Washington, DC, 1988.
27. K.W. Barnes, 'A Streamlined Approach to the Determination of Trace Elements in Foods', *At. Spectrosc.*, **19**, 31–39 (1998).
28. M.A.Z. Arruda, M.J. Quintela, M. Gallego, M. Valcárcel, 'Direct Analysis of Milk for Aluminum Using Electrothermal Atomic Absorption Spectrometry', *Analyst*, **119**, 1695–1699 (1994).
29. J.A. Nobrega, Y. Gelinas, A. Krushevska, R.M. Barnes, 'Direct Determination of Major and Trace Elements in Milk by Inductively Coupled Plasma Atomic Emission and Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 1243–1246 (1997).
30. J.L. Burguera, M. Burguera, 'Flow Injection for Automation in Atomic Spectrometry', *J. Anal. At. Spectrom.*, **12**, 643–651 (1997).
31. M.J. Baxter, H.M. Crews, M.J. Dennis, I. Goodall, D. Anderson, 'The Determination of the Authenticity of Wine from Its Trace Element Composition', *Food Chem.*, **60**, 443–450 (1997).
32. N.J. Miller-Ihli, 'A Systematic Approach to Ultrasonic Slurry GFAAS', *At. Spectrosc.*, **13**, 1–6 (1992).
33. N.J. Miller-Ihli, 'Influence of Slurry Preparation of the Accuracy of Ultrasonic Slurry Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **9**, 1129–1134 (1994).
34. A. Carlosena, M. Gallego, M. Valcárcel, 'Evaluation of Various Sample Preparation Procedures for the Determination of Chromium, Cobalt, and Nickel in Vegetables', *J. Anal. At. Spectrom.*, **12**, 479–486 (1997).
35. R.W. Fonseca, N.J. Miller-Ihli, 'Analyte Transport Studies of Aqueous Solutions and Slurry Samples Using Electrothermal Vaporization ICP/MS', *Appl. Spectrosc.*, **49**, 1403–1410 (1995).
36. N.J. Miller-Ihli, 'Slurry Sampling for Graphite Furnace Atomic Absorption Spectrometry', *Fresenius' J. Anal. Chem.*, **337**, 271–274 (1990).
37. B. Welz, *Atomic Absorption Spectrometry*, VCH, Weinheim, 1985.
38. J.C. Van Loon, *Selected Methods of Trace Metal Analysis*, Wiley, New York, 1985.
39. M. Plessi, D. Bertelli, A. Monzani, 'Determination of Aluminum and Zinc in Infant Formulas and Infant Foods', *J. Food. Compos. Anal.*, **10**, 36–42 (1997).
40. F. Corradini, G. Franchini, A. Marchetti, C. Preti, 'Determination of Alkali and Alkaline Earth Metals in Aceto Balsamico Tradizionale di Modena by Flame Atomic Absorption Spectrometry', *J. AOAC Int.*, **79**, 1167–1173 (1996).
41. J.P. Díaz, M. Navarro, H. López, M.C. López, 'Determination of Selenium Levels in Dairy Products and Drinks by Hydride Generation Atomic Absorption Spectrometry: Correlation with Daily Dietary Intake', *Food Addit. Contam.*, **14**, 109–114 (1997).
42. J. Kuligowski, K.M. Halperin, 'Stainless Steel Cookware as a Significant Source of Nickel, Chromium, and Iron', *Arch. Environ. Contam. Toxicol.*, **23**, 211–215 (1992).
43. W. Slavin, *Graphite Furnace AAS: A Source Book*, Perkin-Elmer, Norwalk, CT, 1984.
44. W. Slavin, D.D. Manning, G. Carnrick, 'The Stabilized Temperature Platform Furnace', *At. Spectrosc.*, **2**, 137–145 (1981).
45. C. Cabrera-Vique, P.-L. Teissedre, M.-T. Cabanis, J.-C. Cabanis, 'Determination of Levels of Chromium in French Wine and Grapes by Graphite Furnace Atomic Absorption Spectrometry', *J. Agric. Food Chem.*, **45**, 1808–1811 (1997).
46. E.C. Phifer, 'Determination of Chromium and Molybdenum in Medical Foods by Graphite Furnace Atomic

Absorption Spectrophotometry', *J. AOAC Int.*, **78**, 1497–1501 (1995).
47. A. Alegria, R. Barbera, R. Farré, A. Moreno, 'GFAAS Determination of Selenium in Infant Formulas Using a Microwave Digestion Method', *Nahrung*, **38**, 382–385 (1994).
48. P. Viñas, N. Campillo, I. López-García, M. Hernández-Córdoba, 'Electrothermal Atomic Absorption Spectrometric Determination of Molybdenum, Aluminum, Chromium, and Manganese in Milk', *Anal. Chim. Acta*, **356**, 267–276 (1997).
49. G.S.B. Januzzi, F.J. Krug, M.A.Z. Arruda, 'Application of the Slurry Technique to the Determination of Selenium in Fish Samples by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **12**, 375–378 (1997).
50. C. Cabrera, M.L. Lorenzo, M.C. Lopez, 'Electrothermal Atomic Absorption Spectrometric Determination of Cadmium, Copper, Iron, Lead, and Selenium in Fruit Slurry: Analytical Application to Nutritional and Toxicological Quality Control', *J. AOAC Int.*, **78**, 1061–1067 (1995).
51. A. Montaser, D.W. Golightly (eds.), *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*, VCH, New York, 1987.
52. M.J. Martin, F. Pablos, A.G. Gonzalez, 'Characterization of Green Coffee Varieties According to Their Metal Content', *Anal. Chim. Acta*, **358**, 177–183 (1998).
53. M. De Wit, R. Blust, 'Determination of Metals in Saline and Biological Matrices by Axial Inductively Coupled Plasma Atomic Emission Spectrometry Using Microconcentric Nebulization', *J. Anal. At. Spectrom.*, **13**, 515–520 (1998).
54. A. Montaser (ed.), *Inductively Coupled Plasma Mass Spectrometry*, Wiley-VCH, New York, 1998.
55. H.M. Crews, 'Inductively Coupled Plasma Mass Spectrometry (ICP/MS) for the Analysis of Trace Element Contaminants in Foods', in *Progress in Food Contaminant Analysis*, ed. J. Gilbert, Blackie, London, 147–186, 1996.
56. C.J. Park, J.K. Suh, 'Determination of Trace Elements in Rice Flour by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 573–577 (1997).
57. Y.-C. Li, S.-J. Jiang, 'Determination of Cu, Zn, Cd, and Pb in Fish Samples by Slurry Sampling Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **359**, 205–212 (1998).
58. N.J. Miller-Ihli, 'Trace Element Determinations in Foods and Biological Samples Using Inductively Coupled Plasma Atomic Emission Spectrometry and Flame Atomic Absorption Spectrometry', *J. Agric. Food Chem.*, **44**, 2675–2679 (1996).
59. S.A. Baker, N.J. Miller-Ihli, 'Trace Element Determinations in Food and Biological Samples Using Inductively Coupled Plasma Mass Spectrometry', *At. Spectrosc.*, **20**, 167–173 (1999).
60. H.M. Crews, 'Speciation of Trace Elements in Foods, with Special Reference to Cadmium and Selenium: Is it Necessary?', *Spectrochim. Acta, Part B*, **53**, 213–219 (1998).
61. I.S. Krull (ed.), *Trace Metals Analysis and Speciation*, Elsevier, Amsterdam, 1991.
62. J. Szpunar, P. Pellerin, A. Makarov, T. Doco, P. Williams, R. Lobinski, 'Speciation of Metal–Carbohydrate Complexes in Fruit and Vegetable Samples by Size Exclusion HPLC/ICP/MS', *J. Anal. At. Spectrom.*, **14**, 639–644 (1999).
63. K.E. Ødegärd, W. Lund, 'Multi-element Speciation of Tea Infusion Using Cation-exchange Separation and Size-exclusion Chromatography in Combination with Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 403–408 (1997).
64. H. Ge, X.-J. Cai, J.F. Tyson, P.C. Uden, E.R. Denoyer, E. Block, 'Identification of Selenium Species in Selenium-enriched Garlic, Onion, and Broccoli Using High-performance Ion Chromatography with Inductively Coupled Plasma Mass Spectrometry', *Anal. Commun.*, **33**, 279–281 (1996).
65. H. Emteborg, G. Bordin, A.R. Rodriguez, 'Speciation of Selenium in Aqueous and Biological Matrices by Microbore Ion Chromatography Coupled with Electrothermal Atomic Absorption Spectrometry via Ultra Low Volume Fraction Collection', *Analyst*, **123**, 893–989 (1998).
66. H.M. Crews, P.A. Clarke, D.J. Lewis, L.M. Owen, P.R. Strutt, A. Izquierdo, 'Investigation of Selenium Speciation in In Vitro Gastrointestinal Extracts of Cooked Cod by High-performance Liquid Chromatography–Inductively Coupled Plasma Mass Spectrometry and Electrospray Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1177–1182 (1996).
67. C. Casiot, J. Szpunar, R. Lobinski, M. Potin-Gautier, 'Sample Preparation and HPLC Separation Approaches to Speciation Analysis of Selenium in Yeast by ICP–MS', *J. Anal. At. Spectrom.*, **14**, 645–650 (1999).
68. J.M. Harnly, W.R. Wolf, N.J. Miller-Ihli, 'Quality Assurance Analysis of Inorganic Nutrients in Foods', in *Modern Methods of Food Analysis*, eds. K.K. Stewart, J.R. Whitaker, AVI Publishing, Westport, CT, Chapter 6, 1984.
69. J.K. Taylor, *Quality Assurance of Chemical Measurements*, Lewis Chelsea, MI, 1987.
70. D.L. Anderson, R.G. Downing, G.V. Iyengar, 'Trace Elements in Food Reference Materials: Compositional and Analytical Perspectives', *Fresenius' J. Anal. Chem.*, **352**, 107–110 (1995).

Dietary Fiber Analysis as Non-starch Polysaccharides

Hans N. Englyst, Michael E. Quigley, and Geoffrey J. Hudson
Englyst Carbohydrate Services, Eastleigh, UK

1 Introduction 2
2 Historical Background and the Definition of Dietary Fiber 3
3 Applications and Health Implications 3
3.1 Scope and Field of Application 3
3.2 Food Tables 4
3.3 Food Labeling 4
3.4 Physiological Studies and Health Implications 4
4 Methods and Principles of Measurement 4
4.1 Apparatus and Reagents 5
4.2 Sample Preparation 6
4.3 Isolation and Hydrolysis of Non-starch Polysaccharides 7
4.4 Determination of Constituent Sugars by Gas–Liquid Chromatography 8
4.5 Determination of Constituent Sugars by High-performance Liquid Chromatography 10
4.6 Determination of Constituent Sugars by Colorimetry 11
5 Quality Control and Troubleshooting 12
5.1 Troubleshooting for the Common Hydrolysis Steps 12
5.2 Troubleshooting for the Gas–Liquid Chromatography Procedure 12
5.3 Troubleshooting for the High-performance Liquid Chromatography Procedure 13
5.4 Troubleshooting for the Colorimetry Procedure 13
6 Method Development 13
6.1 Isolation of Nonstarch Polysaccharides 13
6.2 Optimization of Nonstarch Polysaccharide Hydrolysis Conditions 13
6.3 Response Factors in the Chromatographic Assays 13
6.4 Factors Affecting the Reducing Sugar Assay 14
7 Method Validation 14
7.1 Comparison of the Colorimetric, Gas–Liquid Chromatography and High-performance Liquid Chromatography Assays 14
7.2 International Collaborative Trials 15
7.3 Certification of Reference Materials 15
8 Comparison with the Prosky Procedure 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

The original hypothesis on fiber, that a diet with a high content of unrefined plant foods was linked to a low incidence of Western diseases, has been largely confirmed. The plant cell wall encapsulates and thus controls the release of nutrients, including free sugars and starch, from the plant tissue and thereby influences the glycemic response. This ability to influence digestion and absorption, and thus the physiological effects of other nutrients, is an important property of dietary fiber. The other main properties of a high-fiber diet are increased fermentation and fecal bulk, and binding of potentially toxic substances, e.g. lectins, thus affecting large gut physiology and health. Naturally high-fiber diets have low contents of fat and sodium, and are a good source of vitamins and minerals. The national dietary guidelines recommend a diet rich in fruit, vegetables and whole-grain cereals, and thus naturally rich in plant cell-wall material, the major constituents (approximately 90%) of which are nonstarch polysaccharides (NSP). Food labeling for dietary fiber should help the consumer in the choice of the recommended unfortified, high-fiber diet. Analysis of NSP represents a good measure of endogenous plant cell-wall material for most plant foods and NSP values thus provide a very good marker for a high-fiber diet.

The Englyst procedure for the measurement of dietary fiber as NSP involves enzymatic hydrolysis of starch, precipitation of NSP in ethanol, acid hydrolysis of the NSP and measurement of the released constituent sugars by gas–liquid chromatography (GLC), high-performance liquid chromatography (HPLC) or colorimetry. Values for total, soluble and insoluble NSP may be obtained using any of the end-point techniques. The detailed information obtained from the chromatographic methods, which identify and quantify the individual constituent sugars, is particularly useful in studies of the relation between intakes of NSP and health; values for the constituent NSP sugars have been published for a wide range of foods.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The original hypothesis on fiber was that a diet with a high content of unrefined plant foods (rich in cell-wall material) was linked to a low incidence of some of the so-called Western diseases.[1–3] The hypothesis has been largely confirmed and this is the background for the recommendations in national dietary guidelines for the consumption of a diet rich in fruit, vegetables and whole-grain cereals, and thus naturally rich in plant cell-wall material. The major constituents (approximately 90%) of this material are NSP and, therefore, the plant cell-wall NSP content of foods is a very good marker for the fiber-rich diet recommended in the guidelines. In line with this, in 1994 the EC Scientific Committee for Food stated in a report of their deliberations on the definition of dietary fiber for nutrition labeling:[4]

> There was lengthy discussion as to the material that should be defined as fibre for the purposes of nutrition labelling. In particular the inclusion or not of Resistant Starch in the definition of fibre was argued extensively. The Committee decided that the material to be considered as fibre for the purposes of nutrition labelling should be confined to non-starch polysaccharides of cell-wall origin.

and in 1998, the FAO/WHO presented a list of recommendations on the role of carbohydrates in nutrition and stated:[5]

> That the analysis and labeling of dietary carbohydrates, for whatever purpose, be based on the chemical divisions recommended. Additional groupings such as polyols, resistant starch, non-digestible oligosaccharides and dietary fiber can be used, provided the included components are clearly defined.

In line with this, methods for the measurement of free sugars, starch, NSP and, more recently, short-chain carbohydrates including inulin have been developed.[6–11]

The Englyst procedure for the measurement of dietary fiber as NSP has evolved from the principles laid down by McCance and Widdowson and later by Southgate.[10,12] The procedure involves enzymatic hydrolysis of starch, precipitation of NSP in ethanol, acid hydrolysis of the NSP and measurement of the released constituent sugars by GLC, HPLC or colorimetry. Values for total, soluble and insoluble NSP may be obtained using any of the end-point techniques. The detailed information obtained from the chromatographic methods, which identify and quantify the individual constituent sugars, is particularly useful in studies of the relation between intakes of NSP and health; values for the constituent NSP sugars have been published for a wide range of foods.[13,14] The spectrum of the constituent sugars is characteristic for various types of plant NSP and may indicate the origin of the NSP measured and the physiological effects. The values for wholemeal bread and wholemeal wheat products are characterized by high levels of insoluble NSP in the form of cellulose (measured as insoluble NSP glucose (Glc)) and arabinoxylans. Wholemeal wheat NSP are fermented slowly and incompletely, and exert a considerable effect on fecal bulk. White bread contains only 30% as much NSP as wholemeal bread, and these NSP are largely soluble and has little effect on fecal bulk. Oats contain a greater proportion of soluble NSP compared with wheat products, and the main fraction is a β-glucan, which is measured as soluble NSP Glc. This is associated with a greater effect on cholesterol metabolism, which is in agreement with the health claims for oat products. Fruit and vegetables have high levels of soluble fiber, and the main fraction in these foods is pectin, which is measured as soluble NSP uronic acids (Uac). In general, cereal products contain more xylose (Xyl) than arabinose (Ara), while fruits and vegetables contain less Xyl than Ara, which is measured mainly in the soluble fraction. High values for Uac indicate a diet rich in fruits and vegetables. When the detailed information on the constituent sugars is not required, values for total,

Table 1 Nutritional classification of plant NSP

Type	Origin in the diet	Encapsulate and moderate release of other nutrients	Marker for a naturally high-fiber diet[a]
Plant cell wall NSP			
Present in the plant cell wall (= dietary fiber)	Unrefined plant foods	Yes	Yes
Isolated from the plant cell wall (Cellulose, pectin)	Food additives	No	No
Nonplant cell wall NSP (Gums, mucilages, algal polysaccharides, chemically modified polysaccharides)	Food additives	No	No

[a] A diet rich in fruit, vegetables and whole-grain cereal products as recommended in national dietary guidelines.

soluble and insoluble dietary fiber may be obtained by the more rapid colorimetric end-point procedure.[10] The colorimetric version can be completed within 8 working hours, and it is suitable for food labeling and for quality control. The Englyst procedure for the measurement of dietary fiber as NSP has been thoroughly tested in large international collaborative trials.[15,16] Values for dietary fiber measured as NSP by this technique are used in the McCance and Widdowson UK food tables.[17–19]

The nutritional classification of plant NSP is given in Table 1.

2 HISTORICAL BACKGROUND AND THE DEFINITION OF DIETARY FIBER

The link between the ingestion of unprocessed foods and good health is chronicled as far back as Hippocrates (4th century BC), who recommended the eating of wholemeal bread for its ''salutary effects upon the bowels''. The term dietary fiber was used first by Hipsley, and the dietary fiber hypothesis, which was put forward by Cleave, Burkitt and Trowell more than 2000 years after Hippocrates, was related to the health benefits consequent upon the ingestion of a diet rich in unrefined plant foods and was specifically related to the plant cell wall.[1–3,20] The plant cell wall encapsulates and thus controls the release of nutrients, including free sugars and starch, from the plant tissue and thereby influences the glycemic response. This ability to influence digestion and absorption, and thus the physiological effects of other nutrients, was seen at an early stage to be an important property of dietary fiber.[21] The other main properties associated with dietary fiber were its ability to increase fecal bulk and reduce transit time, and its ability to adsorb and eliminate cholesterol and potentially toxic substances, thus affecting large gut physiology and health. On the basis of these original concepts, any material characterized as dietary fiber should have the physicochemical characteristics of the cell-wall material in plant foods. No type of starch (including that escaping digestion in the human small intestine), inulin or any of the naturally occurring or semisynthetic mono-, di- or oligosaccharides that are not digested and absorbed in the human small intestine has the inherent properties of plant cell-wall material, and none should be included in the definition or measurement of dietary fiber. The structural integrity of plant cell walls may hinder the access of digestive enzymes and cause starch to escape digestion, but there is no reason to classify that starch, or starch escaping digestion in the small intestine for whatever reason, as dietary fiber.

Dietary fiber and starch are chemically distinct compounds, and Trowell[22] made it clear that the definitions proposed earlier[3,23] were not intended to include starch when he stated in his comment on definitions that, ''derived from plant cell walls and not digested by human alimentary enzymes, starch was excluded from all these definitions and was never named at that time or subsequently as a constituent of dietary fiber''. He concluded, ''At the present time there is considerable international agreement concerning the principal constituents of dietary fiber. They are all polysaccharides, mainly cellulose, hemicelluloses and pectic substances, conveniently designated nonstarch polysaccharides''. In 1987 it was proposed by Englyst et al. that dietary fiber should be defined and measured, for the purpose of food labeling, as NSP.[24]

Naturally high-fiber diets are low in free sugars, salt and fat, and provide a good source of a range of nutrients, including vitamins, minerals and antioxidants. All the properties of a high-fiber diet, including those related to the structural properties such as the encapsulation of nutrients within plant cell walls, have been implicated in the protection from a variety of diseases.[22,25]

It is clear that any definition used as the basis for measurement of dietary fiber must provide values that serve as a reliable marker for a high-fiber diet if they are to be meaningful in terms of the evidence that exists in support of the dietary fiber hypothesis.

3 APPLICATIONS AND HEALTH IMPLICATIONS

3.1 Scope and Field of Application

Dietary fiber defined and measured as NSP, as described by Englyst et al., is in keeping with the original concept of dietary fiber. Measurement of NSP provides a reliable index of plant cell-wall material and gives the analyst a clear objective.[1–3,24]

The methods have been used successfully for the analysis of a very wide range of foods. The information obtained by the GLC and HPLC procedures is valuable for interpreting the results of physiological and epidemiological studies where disease may be related to the type of dietary fiber. When detailed information is not required, however, values for NSP can be obtained by colorimetry. In comparative studies of a wide range of food products, good agreement has been demonstrated between the GLC, HPLC and colorimetric versions of the Englyst procedure.

NSP are not affected by normal food processing, so the amounts of fiber in processed foods and mixed diets can be calculated simply from knowledge of the amount in the raw product, although possible losses to the boiling water must be taken into account. This independence of the

method from food processing makes values for dietary fiber measured as NSP suitable for food labeling and easy to defend.

3.2 Food Tables

In addition to the FAO/WHO, the compilers and the users of food tables and databases in general are currently requesting specific measurements of dietary carbohydrates that will allow total carbohydrates to be calculated as the sum of the parts rather than, as has been done in the past, calculating carbohydrates ''by difference''.[5,26] The calculation is not possible, however, when techniques are used that yield values for a mixture of components, and many of the current ''fiber'' techniques include noncarbohydrate components of the sample material.

Different countries have chosen methods that include different components and therefore yield different values for fiber. The material measured as ''fiber'' by nonspecific gravimetric procedures may be a mixture of NSP, starch and noncarbohydrate components in unspecified proportions that may vary as the result of food processing; such values are impossible to interpret in chemical terms and cannot be used in energy calculations. Like other values, dietary fiber must be based on the measurement of chemically identified components.

3.3 Food Labeling

The purpose of food labeling for dietary fiber is to help the consumer in choosing the high-fiber diet recommended in nutritional guidelines.

NSP represent the principal component (approximately 90%) of plant cell walls, and their measurement provides a good measure of endogenous plant cell-wall material for most plant foods.[10,24,27] NSP values thus provide a reliable marker for the high-fiber diet embodied in the dietary fiber hypothesis and recommended in national dietary guidelines.

Fiber is included as a carbohydrate for nutritional labeling in some countries, while in others it is listed separately. We suggest that when dietary fiber is measured as NSP, it is properly listed under carbohydrate for nutritional labeling purposes.

3.4 Physiological Studies and Health Implications

The digestion of plant polysaccharides has been studied in some detail using this technique, and all the NSP have been demonstrated to reach the human large intestine where they may be fermented at different rates and to different extents.[28–32] The gases carbon dioxide, hydrogen and, in about 50% of people, methane, and the short-chain fatty acids propionate, acetate and butyrate in various proportions are the major end-products of fermentation by the large gut microbiota. The gases are absorbed and lost via the lungs or passed as flatus. The short-chain fatty acids are absorbed and represent a source of energy, e.g. butyrate is absorbed and used as the main source of energy for the colonic epithelia.

Although NSP are not digested and absorbed in the human small intestine, their presence can influence the rate and extent of digestion and absorption of other dietary carbohydrates. Starch and sugars trapped within whole plant cells or a dense food matrix may be digested and absorbed more slowly because access of the hydrolytic enzymes is physically hindered (this is discussed in more detail in the article **Starch Analysis in Food**).

4 METHODS AND PRINCIPLES OF MEASUREMENT

The Englyst procedure measures dietary fiber as NSP, using enzymatic chemical methods, and has evolved from the principles laid down by Southgate in 1969.[12] Starch is completely removed enzymatically and NSP are measured as the sum of the constituent sugars released by acid hydrolysis. The sugars may be measured by GLC or HPLC to obtain values for individual monosaccharides or a single value for total sugars may be obtained by colorimetry.[10] Values may be obtained for total, soluble and insoluble NSP, and a small modification allows cellulose to be measured separately.

The Englyst procedure allows the measurement of total, soluble and insoluble NSP in plant food products within an 8 h working day using the colorimetric end-point, or within 1.5 working days with the chromatographic procedures. Preparation of the hydrolyzate for

Table 2 Principle of measurement of NSP

1. Dry/defat sample if necessary
2. Disperse [dimethyl sulfoxide (DMSO)] and hydrolyze starch enzymatically
3. Precipitate NSP in acidified aqueous ethanol
4. Disperse and hydrolyze NSP with sulfuric acid
5. Measure released constituent sugars by:
   - Option 1: GLC[a]
   - Option 2: HPLC[a]
   - Option 3: Colorimetry[b]

[a] Values for individual constituent sugars of insoluble and total NSP. Soluble NSP calculated as the difference between insoluble and total NSP.

[b] Single values for insoluble and total NSP. Soluble NSP calculated as the difference between insoluble and total NSP.

analysis is virtually a single-tube procedure and no special skill or equipment is needed for the colorimetric version.

The procedure as described here provides the following options. (1) GLC procedure: measures NSP as the sum of neutral sugars obtained by GLC and Uac measured separately. (2) HPLC procedure: measures NSP as the sum of neutral sugars and Uac. (3) Colorimetric procedure: measures NSP as reducing sugars. (4) Colorimetric procedure with separate measurement of Uac.

The principle is set out in Table 2.

4.1 Apparatus and Reagents

High-purity reagents and distilled, deionized water, or water of equivalent purity, should be used throughout the method.

4.1.1 Reagents Common to Gas–Liquid Chromatography, Colorimetric and/or High-performance Liquid Chromatography Procedures

All sugars used for standards should be dried to constant mass under reduced pressure with phosphorus pentoxide before use.

Acidified Ethanol Absolute and 85% (v/v). Add 1 mL of 5 mol L^{-1} hydrochloric acid per liter of ethanol.

Benzoic Acid, Saturated Prepare a saturated solution of benzoic acid at room temperature. Add 0.5 g of benzoic acid per 100 mL of water; some benzoic acid should remain undissolved after overnight stirring. The saturated solution is stable at room temperature for long periods.

Dimethylphenol Solution Dissolve 0.1 g of 3,5-dimethylphenol in 100 mL of glacial acetic acid.

Dimethyl Sulfoxide

Enzyme Solution I Take 2.5 mL of heat-stable amylase (EC 3.2.1.1: Termamyl; Englyst Carbohydrate Services, Cat. No. 7367-723), make to 200 mL with pre-equilibrated sodium acetate buffer, mix and keep it in a 50 °C water-bath. Prepare the solution immediately before use.

Enzyme Solution II Take 1.2 g of pancreatin (Sigma, Cat. No. P1500) in a 50-mL tube, add 12 mL of water, vortex-mix initially and then mix for 10 min with a magnetic stirrer. Vortex-mix again, then centrifuge at low speed for 10 min. Take 10 mL of the (cloudy) supernatant, add 2.5 mL of pullulanase (EC 3.2.1.41: Promozyme; Englyst Carbohydrate Services, Cat. No. 7367-727) and vortex-mix. Prepare the solution immediately before use and keep it at room temperature.

Glass Balls 2.5–3.5 mm diameter (Merck, Cat. No. 332124G).

Sand Acid-washed, 50–100 mesh (Merck, Cat. No. 330945E).

Sodium Acetate Buffer 0.1 mol L^{-1}, pH 5.2. Dissolve 13.6 g of sodium acetate trihydrate and make up to 1 L with water. Adjust to pH 5.2 with 0.1 mol L^{-1} acetic acid. To stabilize and activate enzymes, add 4 mL of 1 mol L^{-1} calcium chloride to 1 L of buffer.

Sodium Chloride–Boric Acid Solution Dissolve 2 g of sodium chloride and 3 g of boric acid in 100 mL of water.

Sodium Phosphate Buffer 0.2 mol L^{-1}, pH 7. Adjust 0.2 mol L^{-1} Na_2HPO_4 to pH 7 with 0.2 mol L^{-1} NaH_2PO_4.

Sulfuric Acid 12 mol L^{-1} (Merck, Cat. No. 19321 6Y).

Sulfuric Acid 2 mol L^{-1}. Add 5 mL of 12 mol L^{-1} sulfuric acid to 25 mL of water. Allow to cool to room temperature before use.

Sulfuric Acid 2.4 mol L^{-1}. Add 5 mL of 12 mol L^{-1} sulfuric acid to 20 mL of water and mix.

4.1.2 Reagents Used Only in the Gas–Liquid Chromatography Procedure

Acetic Anhydride

Ammonium–Sodium Tetrahydroborate Solution A solution of 6 mol L^{-1} ammonium hydroxide, containing 200 mg mL^{-1} of sodium tetrahydroborate ($NaBH_4$). Prepare immediately before use.

Bromophenol Blue Solution 0.4 g L^{-1} (Merck, Cat. No. 21006 4K).

Gas–Liquid Chromatography Internal Standard Solution 1 mg mL^{-1}. Weigh, to the nearest 1 mg, 500 mg of allose. Dissolve in water, add 250 mL of saturated benzoic acid and make up to 500 mL with water to give a 1 mg mL^{-1} solution. The solution is stable at room temperature for several months.

Glacial Acetic Acid

Gas–Liquid Chromatography Stock Sugar Mixture Weigh, to the nearest 1 mg, 0.52 g of rhamnose (Rha), 0.48 g of fucose (Fuc), 4.75 g of Ara, 4.45 g of Xyl, 2.3 g of mannose (Man), 2.82 g of galactose (Gal), 9.4 g of Glc

and 2.79 g of galacturonic acid (GalA) (or 3.05 g of GalA monohydrate). Dissolve together in water, add 500 mL of saturated benzoic acid and make up to 1 L with water The solution is stable at room temperature for several months.

1-Methylimidazole

4.1.3 Reagents Used Only in the High-performance Liquid Chromatography Procedure

High-performance Liquid Chromatography Neutral Sugars Internal Standard Solution Weigh, to the nearest 1 mg, 10 000 mg of deoxygalactose. Add 10 mg of thiomersal ($C_2H_5HgSC_6H_4COONa$) as preservative and make up to 1 L with water.

High-performance Liquid Chromatography Neutral Sugars Stock Sugar Mixture Weigh, to the nearest 1 mg, 0.52 g of Rha, 0.48 g of Fuc, 4.75 g of Ara, 4.45 g of Xyl, 2.3 g of Man, 2.82 g of Gal and 9.4 g of Glc. Dissolve together in water, add 10 mg of thiomersal as preservative and make up to 1 L with water. Store at 4 °C.

High-performance Liquid Chromatography Uronic Acids Internal Standard Solution Weigh, to the nearest, 1 mg, 0.0454 g of mannuronic acid lactone. Dissolve in water, add 1 mg of thiomersal as preservative and make up to 100 mL with water. Store at 4 °C.

High-performance Liquid Chromatography Uronic Acids Stock Sugar Mixture Weigh, to the nearest 1 mg, 0.0930 g of GalA (or 0.102 g of GalA monohydrate) and 0.0233 g of glucuronic acid (GlcA). Dissolve together in 2 mol L^{-1} sulfuric acid and make up to 100 mL with 2 mol L^{-1} sulfuric acid. Store at 4 °C.

Pectinase Solution Add 25 volumes of water to 1 volume of pectinase (EC 3.2.1.15; Englyst Carbohydrate Services, Cat. No. 7367-726.

4.1.4 Reagents Used Only in the Colorimetric Procedure

Color Reagent Dissolve 10 g of 3,5-dinitrosalicylic acid and 300 g of sodium/potassium tartrate in approximately 300 mL of water plus 400 mL of 1 mol L^{-1} NaOH. Dissolve by stirring (overnight) and make up to a final volume of 1 L with water. Sparge for 10 min with helium or nitrogen, or degas using an ultrasonic bath. Store in well-capped opaque bottles and keep for 2 days before use. The reagent is stable at room temperature for at least 6 months.

Dimethylglutaric Acid Solution 0.5 mol L^{-1}. Add 98.5 g of 5 mol L^{-1} NaOH to 8.0 g (weighed to the nearest 1 mg) of dimethylglutaric acid (DMG) and make up to 100 mL with water at room temperature. [Note, since the purity of DMG varies between lot numbers, it is essential that the pH is checked. A portion of the DMG solution should be diluted 1 : 1 (v/v) with 2 mol L^{-1} sulfuric acid; the pH should be 3.75 (±0.15) at room temperature. If the pH is greater than 3.9, add 1 mL of water to the stock solution and check the pH again. If the pH is less than 3.6, add 1 mL of 5 mol L^{-1} NaOH to the stock solution and check the pH again. Repeat as necessary.]

Color Stock Sugar Mixture Make the color stock sugar solution by weighing, to the nearest 1 mg, 10.185 g of Ara, 5.145 g of Glc and 2.16 g of GalA (or 2.36 g of GalA monohydrate). Dissolve together in water, add 500 mL of saturated benzoic acid and make up to 1 L with water. The solution is stable at room temperature for several months.

Pectinase Solution Add 9 volumes of water to 1 volume of pectinase (EC 3.2.1.15; Englyst Carbohydrate Services, Cat. No. 7367-726).

4.2 Sample Preparation

All samples must be finely divided (to pass a 0.5-mm mesh) so that representative subsamples may be taken. Foods with a low water content (<10 g per 100 g of sample) may be milled, and foods with a higher water content may be homogenized wet or milled after freeze-drying. Analysis of three sub-samples, A, B and C, allows separate values to be obtained directly for total NSP, insoluble NSP and cellulose, respectively. Soluble NSP is determined as the difference between total and insoluble NSP. Portions A and B are treated identically throughout the procedure, except in steps 3 and 4. The third portion, C, is needed only if a separate value for cellulose is required; carry it through steps 1, 2 and 3 of the procedure then go to step 5.2.

Step 1.1 Weigh, to the nearest 1 mg, between 50 and 1000 mg, depending on the water and NSP content of the sample (to give not more than 300 mg of dry matter, e.g. 300 mg is adequate for most dried foods but smaller amounts should be used for bran and purified fiber preparations) into 50–60-mL screw-topped glass tubes. Add 300 (±20) mg of acid-washed sand and approximately 15 glass balls to each. If the sample is dry (85–100 g of dry matter per 100 g of sample) and contains less than 10 g of fat per 100 g of sample, proceed to step 2; otherwise, go to step 1.2. All analyses should be done in duplicate.

Step 1.2 Add 40 mL of acetone, cap the tubes and mix several times over 30 min. Centrifuge at 1000 *g* for 10 min to obtain a clear supernatant and remove as much

of the supernatant liquid as possible without disturbing the residue. Vortex-mix vigorously to ensure that the residue is dispersed thinly around the bottom 5 cm of the tube. Place the rack of tubes in a pan of water at 75 °C in a fume-cupboard. Remove the tubes singly and vortex-mix vigorously at frequent intervals until the tubes and residues are dry.

4.3 Isolation and Hydrolysis of Nonstarch Polysaccharides

4.3.1 Dispersion and Enzymatic Hydrolysis

Pre-equilibrate sufficient acetate buffer at 50 °C (8 mL required per sample).

Step 2.1 Add 2 mL of DMSO to the dry sample, cap the tube and immediately mix the contents using a vortex-mixer, treating each tube in turn. It is essential that all the sample is wetted and no material is encapsulated or adhering to the tube wall before proceeding. When DMSO has been added to all the tubes, vortex-mix three or four times during 5 min. Vortex-mix and immediately place two tubes in a boiling water-bath. Remove after 20 s, vortex-mix and immediately replace the tubes in the bath. Repeat this for subsequent pairs of tubes until all the tubes are in the bath, then leave them for 30 min from that time. During this period, prepare enzyme solutions I and II (see section 4.1; the volumes given are suitable for 24 samples).

Step 2.2 Remove one tube at a time, vortex-mix, uncap and immediately add 8 mL of enzyme solution I (kept at 50 °C), cap the tube, vortex-mix thoroughly, ensuring that no material adheres to the tube wall, and replace it in the boiling water-bath. Leave the tubes there for 10 min, timed from the last addition of enzyme. Transfer the rack of tubes to the 50 °C water-bath. After 3 min, add 0.5 mL of enzyme solution II to each tube and mix the contents thoroughly to aid distribution of the enzyme throughout the sample. Replace the tubes in the 50 °C water-bath and leave them there for 30 min. Mix the contents of each tube continuously or after 10, 20 and 30 min. Transfer the rack of tubes to the boiling water-bath and leave them there for 10 min.

4.3.2 Precipitation and Washing of the Residue for Measurement of Total Nonstarch Polysaccharides

Only sample portion A is given this treatment.

Step 3.1 Cool the samples by placing in ice–water. Add 0.15 mL of 5 mol L^{-1} hydrochloric acid and vortex-mix thoroughly two or three times during 5 min with samples being replaced in the ice–water. Add 40 mL of acidified absolute ethanol and mix well by repeated inversion, then leave in ice–water for 30 min. Centrifuge at 1500 *g* for 10 min to obtain a clear supernatant liquid. Remove as much of the supernatant liquid as possible, without disturbing the residue, and discard it.

Step 3.2 Add approximately 10 mL of acidified 85% ethanol to the residue and vortex-mix. Make up to 50 mL with acidified 85% ethanol and mix thoroughly by repeated inversion. Centrifuge and remove the supernatant liquid as above. Repeat this stage using 50 mL of absolute ethanol.

Step 3.3 Add 30 mL of acetone to the residue and vortex-mix thoroughly to form a suspension. Centrifuge and remove the supernatant liquid as described in step 3.1.

Step 3.4 Place the rack of tubes in a pan of water at 75 °C in a fume-cupboard or a TurboVap (Zymark) at 65 °C. Remove the tubes singly and vortex-mix vigorously at frequent intervals, to ensure that the residue in each tube is finely divided, until each tube and residue appears dry. Place the rack of tubes in a fan oven at 80 °C for 10 min to remove any last traces of acetone. It is essential that the residues and tubes are completely free of acetone.

4.3.3 Extraction and Washing of the Residue for Measurement of Insoluble Nonstarch Polysaccharides

Only sample portion B is given this treatment.

Step 4.1 After the treatment with enzymes in step 2, add 40 mL of sodium phosphate buffer. Place the capped tubes in a boiling water-bath for 30 min. Mix continuously or a minimum of three times during this period. Remove the tubes and equilibrate to room temperature in water. Centrifuge and remove the supernatant liquid as described in step 3.1.

Step 4.2 Add approximately 10 mL of water and vortex-mix. Make up to approximately 50 mL with water and mix well by repeated inversion. Centrifuge and remove the supernatant liquid as described in step 3.1. Repeat this stage using 50 mL of absolute ethanol. Proceed as described for steps 3.3 and 3.4.

4.3.4 Acid Hydrolysis of the Residue from Enzymatic Digestion

Step 5.1 Add 5 mL of 12 mol L^{-1} sulfuric acid to one tube and immediately vortex-mix vigorously: ensure that all the material is wetted. Repeat this for each tube in turn. Once the acid has been added to all the tubes, vortex-mix again and place all the tubes in a water-bath at 35 °C.

Leave the tubes at 35 °C for 30 min with vigorous vortex-mixing after 5, 10 and 20 min to disperse the cellulose. Add 25 mL of water rapidly and vortex-mix. Place into a boiling water-bath and leave for 1 h, timed from when boiling recommences; mix after 10 min. Cool the tubes in tap water.

Step 5.2 A modification allowing the separate measurement of cellulose and noncellulosic polysaccharides (NCP). To portion C, after steps 1–3, add 30 mL of 2 mol L^{-1} sulfuric acid, and mix. Place in a boiling water-bath and leave for 1 h, timed from when boiling recommences, stirring continuously or after 10 min. The value for cellulose is obtained as the difference between Glc (measured by GLC or by Glc oxidase) for sample portions A and C. NCP is calculated as the difference between total NSP and cellulose.

4.3.5 Breaks in the Procedure

The procedure may be halted at either of the following stages: (1) after precipitation, washing and drying the starch-free residue in steps 3 and 4; the residue may be stored for long periods; or (2) after the hydrolysis with 2 mol L^{-1} sulfuric acid in step 5; the hydrolyzate may be kept at 4 °C for 48 h.

4.4 Determination of Constituent Sugars by Gas–Liquid Chromatography

This assay includes the measurement of neutral sugars by GLC and the separate measurement of Uac by colorimetry.

4.4.1 Measurement of Neutral Nonstarch Polysaccharide Constituents by Gas–Liquid Chromatography

For preparation of the standard sugar mixture, mix 1.0 mL of the GLC stock sugar solution and 5 mL of 2.4 mol L^{-1} sulfuric acid. Treat 2×1.0 mL of this standard sugar mixture for calibration in parallel with the hydrolyzates from step 5 of the procedure. Prepare the alditol acetate derivatives for chromatography as follows. Add 0.50 mL of GLC internal standard (IS) (1 mg mL^{-1} allose) to 1.0 mL of the cooled hydrolyzates from step 5 and to 2×1 mL of the standard sugar mixture; vortex-mix. Place the tubes in ice–water, add 0.4 mL of 12 mol L^{-1} ammonium solution and vortex-mix. Test that the solution is alkaline (add a little more ammonium solution if necessary but replace the ammonium solution if more than 0.1 mL extra is required), then add approximately 5 μL of the antifoam agent octan-2-ol and 0.1 mL of the ammonium–sodium borohydride solution and vortex-mix. Leave the tubes in a heating block or in a water-bath at 40 °C for 30 min, then remove and add 0.2 mL of glacial acetic acid and mix again. Transfer 0.5 mL into a 30-mL glass tube and add 0.5 mL of 1-methylimidazole and 5 mL of acetic anhydride. Vortex-mix, then leave the tubes for 10 min for the reaction to proceed (the reaction is exothermic and the tubes will become hot). Add 0.9 mL of absolute ethanol, vortex-mix and leave for 5 min. Add 10 mL of water, vortex-mix and leave for 5 min. Add 0.5 mL of bromophenol blue solution. Place the tubes in ice–water and add 5 mL of 7.5 mol L^{-1} potassium hydroxide; a few minutes later, add a further 5 mL of 7.5 mol L^{-1} potassium hydroxide, cap the tubes and mix by inversion. Leave until the separation into two phases is complete (10–15 min) or centrifuge for a few minutes. Draw part of the upper phase into the tip of an automatic pipette; if any of the blue lower phase is included, allow it to separate, then run it out of the tip before transferring a portion of the upper phase alone to a small (autoinjector) vial. Inject 0.5–1 μL of the alditol acetate derivatives.

4.4.2 Gas–Liquid Chromatography Conditions

The following conditions are used: injector temperature, 275 °C; column temperature, 220 °C; detector temperature, 275 °C; carrier gas, helium; flow-rate, 8 mL min^{-1}. Under these conditions, a GLC system fitted with flame ionization detector, autoinjector and computing integrator, using a Supelco SP-2330 wide-bore capillary column (30 m × 0.75 mm ID) or a Supelco SP-2380 wide-bore capillary column (30 m × 0.53 mm ID) will allow accurate determination of the individual sugars in the GLC standard sugar mixture within 8 min. Carry out conventional GLC measurement of the neutral sugars. At the beginning of each batch of analyses, equilibrate with the isothermal elution conditions for at least 1 h. Do several calibration runs to check that the response factors are reproducible.

4.4.3 Calibration

Calibration ratios are shown in Table 3 for the combination of the GLC standard sugar mixture and GLC IS

Table 3 Calibration details for GLC

Sugar	Actual (mg mL^{-1})	Recovery (%)	Apparent (mg mL^{-1})	Calibration ratio
Rha	520	52	1000	1
Fuc	480	96	500	0.5
Ara	4750	95	5000	5
Xyl	4450	89	5000	5
Man	2300	92	2500	2.5
Gal	2820	94	3000	3
Glc	9400	94	10 000	10
Allose (IS)				3

(allose). The ''Actual'' column shows the amount of each sugar in the mixture and the ''Apparent'' column shows the values to be used for calibration taking into account the recovery of NSP constituents. The ''Calibration ratio'' column gives the ratio of sugars to the IS after the addition of allose to the standard sugar mixture as described in the text. (The experimental evidence for the recovery values has been published and is discussed in detail in section 6.3.)

4.4.4 Calculation of Neutral Sugars

The amount of each individual sugar (expressed as grams of polysaccharide per 100 g of sample) is calculated according to Equation (1):

$$\text{Sugar} = \frac{A(\text{t}) \times W(\text{i}) \times 100 \times R(\text{f}) \times 0.89}{A(\text{i}) \times W(\text{t})} \qquad (1)$$

where $A(\text{t})$ and $A(\text{i})$ are the peak areas of the sample and the IS, respectively, $W(\text{i})$ is the weight (in milligrams here 15: total hydrolyzate 30 mL × 0.5 mg allose) of the IS; $W(\text{t})$ is the weight (in milligrams) of the sample, $R(\text{f})$ is the response factor for individual sugars obtained from the calibration run with the sugar mixture and IS (allose) treated in parallel with the samples and 0.89 is the factor for converting experimentally determined values for monosaccharides to polysaccharides.

4.4.5 Measurement of Uronic Acids by Colorimetry

Prepare the standard solutions as follows. The GLC standard sugar mixture in 2 mol L^{-1} sulfuric acid contains, for the purpose of calibration, 500 μg mL^{-1} of GalA. To prepare the uronic acid standard solutions, transfer 0.5, 2.0 and 3.0 mL of this sugar mixture into separate tubes and make up to 10 mL with 2 mol L^{-1} sulfuric acid to give standards of 25, 100 and 150 μg mL^{-1} of GalA.

Place in separate tubes (40–50-mL capacity) 0.3 mL of blank solution (2 mol L^{-1} sulfuric acid), 0.3 mL of each of the standard solutions and 0.3 mL of the sample hydrolyzates, diluted if necessary (with 2 mol L^{-1} sulfuric acid) to contain no more than 150 μg mL^{-1} of Uac (e.g. no dilution for flour, 1:2 for bran, 1:5 for most fruits and vegetables). Add 0.3 mL of sodium chloride–boric acid solution and mix. Add 5 mL of concentrated sulfuric acid and vortex-mix immediately. Place the tubes in a heating block at 70 °C for 40 min. Remove the tubes and cool to room temperature in water (the tubes may be kept in the water for up to 1 h). Add 0.2 mL of dimethylphenol solution and vortex-mix immediately. After 15 min measure the absorbance at 400 and 450 nm in the spectrophotometer against the blank solution. The timing for measurement of the absorbance of standards and samples should be identical. In practice, this is achieved by adding the color reagent at 1-min intervals. Subtract the absorbance reading at 400 nm from that at 450 nm, to correct for interference from hexoses. A straight line should be obtained if the differences in absorbance for the standards are plotted against concentration. Only the 100 μg mL^{-1} standard is required for routine analysis, and it may be kept at 5 °C for several weeks.

4.4.6 Calculation of Uronic Acids

The amount of Uac (expressed as grams of polysaccharide per 100 g of sample) is calculated according to Equation (2):

$$\text{Uac} = \frac{A(\text{t}) \times V(\text{t}) \times D \times C \times 100 \times 0.91}{A(\text{s}) \times W(\text{t})} \qquad (2)$$

where $A(\text{t})$ is the difference in absorbance of the sample solution, $V(\text{t})$ is the total volume of sample solution (in milliliters, here 30), D is the dilution of the sample solution, C is the concentration of the standard (here 0.1 mg mL^{-1}), $A(\text{s})$ is the difference in absorbance of the 100 μg mL^{-1} standard, $W(\text{t})$ is the weight (in milligrams) of the sample and 0.91 is the factor for converting experimentally determined values for monosaccharides to polysaccharides.

4.4.7 Calculation of Nonstarch Polysaccharides

The amounts of total, soluble and insoluble NSP (in grams per 100 g of sample) are calculated as follows:

$$\text{Total NSP} = \text{neutral sugars calculated for portion A} + \text{Uac calculated for portion A}$$

$$\text{Insoluble NSP} = \text{neutral sugars calculated for portion B} + \text{Uac calculated for portion B}$$

$$\text{Soluble NSP} = \text{Total NSP} - \text{Insoluble NSP}$$

It is recommended that a sample is taken for determination of the dry matter as the loss in weight after overnight incubation at 104 °C. Results may then be expressed as grams of polysaccharides per 100 g of dry matter.

4.4.8 Breaks in the Gas–Liquid Chromatography Procedure

The procedure may be halted at either of the following stages: (1) after acidification of the reduced samples (see section 4.4); the samples may be stored at room temperature for 2–3 days; (2) the acid hydrolyzate from step 5 may be kept at 5 °C for several weeks before the measurement of Uac.

4.5 Determination of Constituent Sugars by High-performance Liquid Chromatography

This assay includes dilution of one subsample of the hydrolyzate and direct measurement of the neutral sugars, and the separate measurement of Uac in a second subsample of the hydrolyzate subjected to treatment with pectinase.

4.5.1 Measurement of Neutral Sugars by High-performance Liquid Chromatography

For preparation of the HPLC standard sugar mixture, mix 1.0 mL of the HPLC stock sugar mixture and 5 mL of 2.4 mol L^{-1} sulfuric acid. Treat 2×0.15 mL of this standard sugar mixture for calibration of HPLC in parallel with the hydrolyzates from step 5 of the procedure.

To 0.15 mL of hydrolyzate or the HPLC standard sugar mixture add 5 mL of the HPLC neutral sugars IS solution (deoxygalactose) and mix well. Inject 25 μL for analysis.

4.5.2 High-performance Liquid Chromatography Conditions

A Dionex Model PAD 2 detector may be used with the following pulse potentials and durations: $E_1 = 0.05$ V ($t_1 = 300$ ms); $E_2 = 0.60$ V ($t_2 = 120$ ms); $E_3 = -0.60$ V ($t_3 = 60$ ms). The response time is 1 s and the detector output is set at 1000 nA.

A Dionex AG-5 guard column, an inert high-pressure valve and a CarboPac PA-1 column are placed in series. The AG-5 column and inert high-pressure valve are used to retain and bypass sulfate ions around the analytical column. Elute with the following: 23% (v/v) solution 1 (20 mmol L^{-1} NaOH) from 0 to 3.5 min; a gradient from 23% to 1% (v/v) solution 1 from 3.5 to 4.5 min; and 1% (v/v) solution 1 from 4.5 to 30 min at a flow rate of 1 mL min^{-1}. Re-equilibrate with the starting conditions for at least 6 min between runs. Add 300 mmol L^{-1} NaOH at a flow-rate of 0.5 mL min^{-1} to the column effluent before the pulsed amperometric detection (PAD) cell to minimize baseline drift and increase the analytical signal. Saturate the eluent with helium (Dionex Eluent De-gas module) to minimize CO_2 absorption. Under the conditions described here, sulfate ions are retained for 80 s on the AG5 guard column and column switching (via the inert high-pressure valve) is applied after 60 s to prevent sulfate ions from reaching the analytical column. Sulfate ions are purged from the guard column within 19 min, well within the total run time. To regenerate the PA-1 column at the end of the chromatographic run, wash with 100 mmol L^{-1} NaOH–600 mmol L^{-1} sodium acetate for 1 h at a flow-rate of 1 mL min^{-1}. Then, wash with 1 mol L^{-1} NaOH (1 mL min^{-1}) for 1 h but, to avoid contamination of the internal reference solution, do not allow this solution to pass through the detector.

Table 4 Calibration details for HPLC

Sugar	Actual (mg mL^{-1})	Recovery (%)	Apparent (mg mL^{-1})	Calibration ratio
Rha	520	52	1000	1
Fuc	480	96	500	0.5
Ara	4750	95	5000	5
Xyl	4450	89	5000	5
Man	2300	92	2500	2.5
Gal	2820	94	3000	3
Glc	9400	94	10 000	10
Deoxygalactose (IS)				2

The calibration ratios are shown in Table 4 for the combination of the standard sugar mixture and deoxygalactose, the HPLC IS: the ''Actual'' column shows the amount of each sugar in the mixture, and the ''Apparent'' column shows the values to be used for calibration taking into account the recovery of NSP constituents.

The amount of each individual sugar (in grams per 100 g of sample) is calculated according to Equation (3):

$$\text{Sugar} = \frac{A(\text{t}) \times W(\text{i}) \times 100 \times R(\text{f}) \times 0.89}{A(\text{i}) \times W(\text{t})} \qquad (3)$$

where A(t) and A(i) are the peak areas of the sample and the HPLC IS, respectively, W(i) is the weight of the IS if added to the whole sample (in milligrams, here 25), W(t) is the weight of the sample (in milligrams), R(f) is the response factor for individual sugars obtained from a calibration run with the sugar mixture treated in parallel with the samples and 0.89 is the factor for converting experimentally determined values for monosaccharides to polysaccharides. It is recommended that a sample is taken for analysis of the dry matter as the loss in weight after overnight incubation at 104 °C. Results may then be expressed as grams of polysaccharides per 100 g of dry matter.

4.5.3 Measurement of Uronic Acids by High-performance Liquid Chromatography

To 0.5 mL of hydrolyzate or calibration mixture, add 0.5 mL of the HPLC Uac IS solution (mannuronic acid lactone), 0.5 mL of DMG solution (8 g per 100 mL) and 2.1 mL of 1 mol L^{-1} NaOH. After vortex-mixing, the pH must be between pH 3.5 and 4.0 (if the pH is not correct, prepare fresh DMG solution and repeat). Add 0.1 mL of pectinase solution, vortex-mix and place the tubes in a water-bath at 50 °C for 20 min. Cool the tubes, remove 0.5 mL, add 0.1 mL of phenol red indicator solution (1 mg mL^{-1}) and add sufficient (approximately

2 mL) freshly prepared 0.017 mol L^{-1} NaOH until the pH is between 7 and 8. Inject 25 µL onto the chromatographic column. Baseline separation of GalA, (GlcA) and mannuronic acid is achieved isocratically within 13 min using 25% solution 2 (100 mmol L^{-1} NaOH, 600 mmol L^{-1} sodium acetate) and 75% water at a flow-rate of 1.0 mL min^{-1}.

Use the following ratios for the calibration mixture for the calculation of the response factors: 0.5 for GalA, 0.125 for GlcA and 0.25 for mannuronic acid, to take into account losses during hydrolysis. The values for the calculation of response factors take into account 7% loss of both GalA and GlcA, and the hydrolysis of mannuronic acid lactone to mannuronic acid.

The amount of Uac (in grams per 100 g of sample) is calculated according to Equation (4):

$$\text{Uac} = \frac{A(\text{t}) \times W(\text{i}) \times 100 \times R(\text{f}) \times 0.91}{A(\text{i}) \times W(\text{t})} \quad (4)$$

where A(t) and A(i) are the peak areas of the sample and the IS, respectively, W(i) and W(t) are the weight of the IS (in milligrams; here 15) and the weight of the sample (in milligrams), respectively, R(f) is the response factor for individual Uac obtained from the calibration mixture and 0.91 is the factor for converting the experimentally determined values for monosaccharides to polysaccharides.

4.6 Determination of Constituent Sugars by Colorimetry

4.6.1 Preparation of the Standard Sugar Mixture

Take 0.5 mL of the colorimetry stock sugar mixture into a glass tube, add 2.5 mL of 2.4 mol L^{-1} sulfuric acid and mix to give 3 mL of 3 mg mL^{-1} color sugars standard solution in 2 mol L^{-1} sulfuric acid. The colorimetric reaction is linear up to 3 mg mL^{-1} sugar. The absorbance of the test samples should not exceed that of the standard.

4.6.2 Measurement of Total Reducing Sugars

Place in separate glass tubes 0.5 mL of the standard sugar solution, 0.5 mL of the hydrolyzate from step 5 and, in each of two tubes (blanks 1 and 2), 0.5 mL of 2 mol L^{-1} sulfuric acid. Add 0.5 mL of DMG solution and vortex-mix. Check the pH of one drop of blank 1; it should be between 3.5 and 4. If it is different from this, check the preparation of the 2 and 12 mol L^{-1} sulfuric acid and the DMG solution. If the pH is correct, add 0.1 mL of diluted pectinase solution, vortex-mix and place all the tubes into a water-bath at 50 °C for 20 min. Cool the tubes to room temperature, add 0.1 mL of 3 mol L^{-1} sodium hydroxide, vortex-mix and leave for 5 min. Add 1 mL of the dinitrosalicylate reagent to each tube and vortex-mix. Place all the tubes together into a briskly boiling water-bath for 5 min. Remove the rack of tubes and cool to room temperature in water. Add 10 mL of water (at room temperature) and mix well by inversion (do not use a vortex-mixer at this stage). Measure the absorbance at 530 nm against blank 2.

Note: sample blanks may be prepared by diluting the hydrolyzates as described above, replacing the color reagent with water, and reading the absorbance against water. Alternatively, and more conveniently, the absorbance of the undiluted hydrolyzate can be measured against water and the value divided by 24.4 (the dilution of the hydrolyzate after addition of the color reagent and water). The absorbance of the test sample is then calculated by subtracting this value. When the hydrolyzate is colorless and the NSP content is more than 5%, the sample blank is not required.

4.6.3 Calculation of Nonstarch Polysaccharides

The amount of total NSP (portion A) and of insoluble NSP (portion B), in grams of polysaccharide per 100 g of sample, is calculated according to Equation (5):

$$\text{NSP} = \frac{A(\text{t}) \times V(\text{t}) \times D \times F \times C \times 100 \times 0.89}{A(\text{s}) \times W(\text{t})} \quad (5)$$

where A(t) is the absorbance of the sample solution (minus the absorbance of the hydrolyzate blank if measured), V(t) is the total volume of the sample solution (in milliliters, here 30), D is the dilution of the sample solution ($D = 1$ if no dilution), F is the factor correcting the difference between the composition of monosaccharides in the standard sugar mixture and that in NSP of various types of plant foods [for the calculation of NSP in cereals (except oats) $F = 0.95$, in fruit and non-starchy vegetables $F = 1.05$ and in starchy vegetables, oat products and unknown samples $F = 1$; using the standard sugar mixture as specified and these factors makes corrections for the 2–4% hydrolytic losses], C is the concentration (in mg mL^{-1} sugars) of the standard, A(s) is the absorbance of the standard, W(t) is the weight (in milligrams) of sample taken for analysis and 0.89 is the factor for converting experimentally determined monosaccharides to polysaccharides. The amount of soluble NSP is calculated as the difference between total NSP and insoluble NSP. It is recommended that a sample is taken for analysis of the dry matter as the loss in weight after overnight incubation at 104 °C. Results may then be expressed as grams of polysaccharides per 100 g of dry matter.

4.6.4 Correction Based on Separate Measurement of Uronic Acids

When a sample, e.g. pectin, has a high content of Uac, a more accurate value for NSP may be obtained if Uac are measured separately. The color standard sugar mixture contains 12.5% (w/v) Uac, and it has been determined (data not shown) that this leads to a 17% underestimation of NSP when the sample contains only Uac (using $F = 1$). Correction for the underestimation is straightforward if a separate value for Uac is obtained (see above). If the sample contains 12.5% monomeric Uac, the color standard sugar mixture is entirely appropriate and no correction is required; otherwise, the correction required is an increment or decrement to the NSP value obtained by colorimetry equivalent to 17% of the uronic acid content that is in excess of or less than 12.5% of the sample, respectively.

The percentage of the NSP value to be corrected for is the difference (Δ) between the value for Uac (Z; expressed as a percentage of the NSP value, X) and 12.5, and is calculated according to Equation (6):

$$\Delta = \left(\frac{Z}{X} \times 100 \quad - 12.5 \right. \tag{6}$$

and the correction factor Y is calculated using Equation (7):

$$Y = 0.17 \left(\frac{X}{100} \times \Delta \right. \tag{7}$$

which may be reduced to $0.17Z - 0.021X$, and then the value for total NSP after correction is $X + Y$.

4.6.5 Breaks in the Colorimetry Procedure

The procedure may be halted at either of the following stages: (1) after precipitation, washing and drying the starch-free residue (see step 3.4); the residue may be stored for long periods. (2) After the hydrolysis with $2\,\text{mol}\,\text{L}^{-1}$ sulfuric acid; the hydrolyzate may be kept at 4 °C for 48 hours.

5 QUALITY CONTROL AND TROUBLESHOOTING

Two kits are available from Englyst Carbohydrate Services to help ensure that accurate analytical results are obtained. The kit for the colorimetric procedure contains color reagent and a second kit contains a solution of allose as IS for the GLC procedure. Both kits contain the required enzymes, sugar solutions and reference materials and are rigorously tested in-house. The use of all four reference materials as suggested below may be used as part of a complete quality control procedure.

Reference sample 1 (white flour) This sample is used to check the efficiency of the starch hydrolysis and washing steps where there is small amounts of NSP in the presence of large amounts of starch.

Reference sample 2 (haricot bean) This sample contains all the constituent sugars of NSP and is used to check the efficiency of the starch hydrolysis and washing steps for samples that may cause handling problems, such as aggregation.

Reference sample 3 (high-amylose starch resistant to α-amylose) This sample is taken through the entire procedure for total NSP and is used to check the efficiency of the starch hydrolysis and washing steps.

Reference sample 4 (cellulose) This sample is subjected to direct acid hydrolysis only and is used to check the acid hydrolysis steps.

5.1 Troubleshooting for the Common Hydrolysis Steps

1. Variation between replicate analyses may be due to inaccurate pipetting (test/calibrate dispensers by weighing replicates of water) or to incomplete removal of acetone in step 3.4.
2. If values for Glc are too high for samples of known composition and/or variable for replicates, this may be due to incomplete wetting of sample with DMSO in step 2.1. Mix vigorously immediately after addition of DMSO.
3. If values for Glc and Uac for samples of known composition are too low and/or are variable for replicates, this may be due to incomplete wetting of samples with $12\,\text{mol}\,\text{L}^{-1}$ sulfuric acid in step 5. Vortex-mix vigorously before and after addition of sulfuric acid and at intervals during the incubation.

5.2 Troubleshooting for the Gas–Liquid Chromatography Procedure

1. Extra peaks on the chromatogram may be due to incomplete reduction of monosaccharides. Ensure an alkaline pH before adding $NaBH_4$. Replace old $NaBH_4$; do not compensate for loss of activity by adding more $NaBH_4$.
2. Variation between replicate analyses may be caused by nonreproducible pipetting of the IS or hydrolyzates. Test/calibrate dispensers by weighing 1 mL replicates of water.
3. If the response factors are not reproducible, this may be due to inaccurate pipetting of the sugar mixture and/or IS. Test/calibrate dispensers by weighing replicates of water.

5.3 Troubleshooting for the High-performance Liquid Chromatography Procedure

If retention times vary during chromatography, regenerate the analytical columns as described in section 4.5.

5.4 Troubleshooting for the Colorimetry Procedure

1. Variation between replicate analyses may be due to inaccurate pipetting (test/calibrate dispensers by weighing replicates of water) or to incomplete removal of acetone in step 3.4.
2. If no color is produced for standards and/or samples, this may be due to an error in the preparation of the sulfuric acid or sodium hydroxide solutions. Make new reagents. Test that the pH of the solution is between 7 and 8 before adding the color reagent solution.

6 METHOD DEVELOPMENT

6.1 Isolation of Nonstarch Polysaccharides

Isolation of NSP from starch and free sugars is achieved by dispersal and enzymatic hydrolysis of starch and subsequent precipitation of NSP. The enzymatic hydrolysis of starch with amylase, pullulanase and amyloglucosidase[33] was initially achieved with an overnight incubation. The introduction of glass balls and sand to prevent agglutination and the use of a heat-stable amylase has reduced the length of incubation to less than 1 h.[10,11]

Oligosaccharides may be linked to polysaccharides by noncovalent bonds, such as calcium bridges forming complexes that are insoluble in 80% ethanol at pH 5.2 as used in earlier versions of the NSP procedure.[11,33,34] Some free GalA is insoluble in 80% ethanol in the presence of sodium acetate buffer (0.1 mol L^{-1}, pH 5.2). Acidifying the 80% ethanol to pH 2 solubilizes all free GalA and prevents coprecipitation of oligosaccharides, thus avoiding their artifactual inclusion in the measurement of NSP. Values obtained for the total NSP content of a polysaccharide mixture (duplicate analysis on four different occasions) were 91.2% (±2.8%) (w/w) and 91.9% (±2.1%) (w/w) when the precipitation in ethanol was done at pH 5.2 and 2, respectively.

6.2 Optimization of Nonstarch Polysaccharide Hydrolysis Conditions

Cellulose and some uronic acid-containing materials require dispersal in 12 mol L^{-1} sulfuric acid before hydrolysis is possible. In the Englyst procedure as it was described in 1982, NSP were completely dispersed and hydrolyzed by treatment with 12 mol L^{-1} sulfuric acid for 1 h at 35 °C, followed by treatment with 1 mol L^{-1} sulfuric acid for 2 h at 100 °C.[33] Subsequently, it was shown that the second step could be completed within 1 h using 2 mol L^{-1} sulfuric acid. The 1 h of treatment with 12 mol L^{-1} sulfuric acid was incorporated because it was found that some samples aggregate during the drying with acetone, hindering the access of the 12 mol L^{-1} sulfuric acid and requiring prolonged treatment to ensure complete dispersion and hydrolysis. It has been shown that dispersal in 12 mol L^{-1} sulfuric acid is complete within 10 min if samples are in powder form.[11,35] In 1994, sand and glass balls were incorporated into the procedure to prevent aggregation and to ensure that samples are finely divided after the drying procedure. With this modification, the treatment with 12 mol L^{-1} sulfuric acid was reduced to 30 min without affecting the results.

As a test for completeness of depolymerization, HPLC with PAD was used to detect any oligosaccharides remaining in solution after the acid hydrolysis steps.[10] No residual oligomers were detected after hydrolysis of cellulose and other noncellulosic neutral sugar-containing NSP, indicating complete hydrolysis. Oligosaccharides were found, however, after hydrolysis of uronic acid-containing polymers, such as pectin, indicating incomplete depolymerization. This incomplete hydrolysis is not important if the uronic acid content is to be determined by the modified Scott procedure as used in the GLC analysis of NSP, because this method measures both free Uac and uronic acid polymers in solution. When using the colorimetric or HPLC procedures, however, prolonged treatment (3 h) with 2 mol L^{-1} sulfuric acid is required to obtain maximal values. Such a treatment is time-consuming and results in further undesirable losses. Treatment with pectinase after adjustment of the pH of the hydrolyzate to between 3.5 and 4 was introduced and this resulted in complete recovery of Uac within 20 min, with the added benefit of no hydrolysis loss. These modifications have improved the analysis of NSP by the colorimetric procedure and allowed measurement of the Uac by HPLC, which had not been previously described.

6.3 Response Factors in the Chromatographic Assays

In the Englyst procedure, NSP are hydrolyzed by treatment with 12 mol L^{-1} sulfuric acid for 30 min at 35 °C, followed by treatment with 2 mol L^{-1} sulfuric acid for 1 h at 100 °C. The treatment with 12 mol L^{-1} sulfuric acid is essential for dispersal but results in extensive sulfation of the monosaccharides, requiring subsequent hydrolysis with dilute acid at high temperature. This, in turn, causes some destruction of the monosaccharides released. The conditions chosen for hydrolysis are therefore necessarily a compromise between adequate release, desulfation and

Table 5 Calibration mixture to account for incomplete recovery of sugars after acid hydrolysis of polymers

	Constituent sugar (g per 100 g of dry sample)								
	Rha	Fuc	Ara	Xyl	Man	Gal	Glc	Uac	IS
Loss during treatment with 12 mol L^{-1} sulfuric acid	4	1	2	4	3	<0.5	<0.5	<0.5	–
Loss during treatment with 2 mol L^{-1} sulfuric acid	3	3	3	7	3	3	3	7	–
Underestimation due to incomplete desulfation	–	–	–	–	2	3	3	–	–
Underestimation due to incomplete hydrolysis	41	–	–	–	–	–	–	–	–
Overall recovery (%)	52	96	95	89	92	94	94	93	–
Calibration mixture (g L^{-1})	0.52	0.48	4.75	4.45	2.3	2.82	9.4	2.79	2
Calibration ratio	1	0.5	5	5	2.5	3	10	3	2

destruction of the monosaccharides. These parameters have been extensively investigated and a full report is given elsewhere.[11] Table 5 shows a summary of the results and shows how the composition of the sugar calibration mixture used in the GLC and HPLC procedures (used for calibration of response factors) takes into account the incomplete recoveries.

6.4 Factors Affecting the Reducing Sugar Assay

A colorimetric procedure was introduced by Englyst and Hudson as a faster, more convenient alternative to the GLC procedure for obtaining values for total NSP.[36] This colorimetric technique, however, consistently gave values that were slightly higher for cereal products and marginally lower for some fruits and vegetables that contained large amounts of Uac than the values obtained by GLC.[15] The procedure was subsequently modified in 1994, incorporating four changes.[10]

Treatment with pectinase was introduced to complete the hydrolysis of the uronic acid-containing material and this has improved the recovery for fruits and vegetables. Advanced Maillard Reaction Products (MRP), such as 5-hydroxymethylfurfural and 2-furaldehyde, formed during cooking, have reducing aldehyde groups that react with the color reagent. MRP exhibit a time-dependent increase in absorbance, whereas virtually maximum values are obtained for the NSP constituent monosaccharides at 5 min. The absorbance values for 5-hydroxymethylfurfural at 5 min are only 30% of those obtained at 15 min, the incubation time in the original procedure. The time of reaction with the color reagent has been reduced to 5 min, which has minimized the interference from products formed during food processing.

Mineral ions have been reported to increase the intensity of the color formed upon heating dinitrosalicylic acid with reducing groups. The extent of the interference is variable and dependent upon the concentrations of both mineral ions and reducing sugars. We showed that mineral ions from the salts of potassium, calcium, phosphorus, sodium, iron, zinc, copper and manganese and traces of cobalt, aluminum and molybdenum were soluble in acidified 80% (v/v) ethanol at pH 2, and the use of acidified ethanol has completely removed this source of interference.[10]

The fourth modification was the introduction of factors to take into account the different composition of the NSP of fruits, vegetables, cereal products, starchy legumes and oat products: 43 vegetables, 40 fruits, 52 cereal products (other than oats) and 17 starchy legume and oat product samples were analyzed by the GLC procedure. On the basis of the monosaccharide composition and knowledge of the color yields of individual monosaccharides, we determined that factors of 1.05 for fruits and vegetables and 0.95 for cereals are appropriate. No correction is required for starchy legumes or oat products when using the color sugar mixture described here (section 4.6).

In summary, sources of potential interference in the measurement of NSP have been identified and eliminated, making the NSP procedure more accurate and robust. Virtually identical values for NSP have subsequently been obtained for all samples measured by the GLC and the colorimetric end-points and the improvements have been reflected by reduced interlaboratory variation as observed in the EU Community Bureau of Reference (BCR) detailed methods study (see section 7.3).

7 METHOD VALIDATION

7.1 Comparison of the Colorimetric, Gas–Liquid Chromatography and High-performance Liquid Chromatography Assays

Three end-point techniques are available as alternatives to obtain values for NSP. The chromatographic

procedures yield detailed information about the composition of the dietary fiber but require more sophisticated equipment. We have demonstrated that the GLC and HPLC procedures give identical values for both neutral sugars[37] and Uac.[37,38] The colorimetric procedure has been demonstrated in interlaboratory studies to give values for total, soluble and insoluble NSP that are identical with values obtained by the chromatographic procedures. The colorimetric assay is ideally suited for food labeling purposes or where values for individual constituent sugars are not required.

7.2 International Collaborative Trials

The Englyst procedure has been the subject of a series of international collaborative trials organized by the UK Ministry of Agriculture, Fisheries and Food (MAFF).[15] This has culminated in the publication of the GLC and the colorimetry techniques as MAFF Approved Methods.[39,40]

In the MAFF IV study, 37 laboratories from 11 countries took part to compare the accuracy and precision of the Englyst GLC and colorimetric procedures and the Association of Official Analytical Chemists (AOAC) Prosky gravimetric method. All methods performed reasonably well for both repeatability and reproducibility but differences were seen in the results. The Prosky procedure gave values that were, on average, 19% higher than those obtained by the Englyst procedure, and the reasons are explained in detail in section 8. The results in Table 6 show an overall summary of the results obtained by the Englyst GLC, Englyst colorimetry and Prosky procedures.

7.3 Certification of Reference Materials

As the result of a large international trial of methodology, following rigorous study of stability of the test materials, five BCR certified reference materials (CRMs) are available for the Englyst GLC and colorimetry NSP procedures:[16] (1) dried haricot bean powder, CRM 514; (2) dried carrot powder, CRM 515; (3) dried apple powder, CRM 516; (4) full fat soya flour, CRM 517; and (5) dried powdered bran breakfast cereal, CRM 518. These CRMs can be used to check the performance of the analytical method and as quality control of analytical measurements for nutritional labeling.

Table 6 Repeatability and reproducibility of the Prosky procedure and the Englyst GLC procedure[a]

Method	Number of laboratories	Fiber content (mean)	Repeatability *r*-95 (mean)	Reproducibility *R*-95 (mean)
Englyst GLC				
Total	22	8.92	1.24	2.67
Soluble	22	3.83	1.36	2.01
Insoluble	22	5.09	0.95	1.92
Prosky (gravimetric)				
Total	16	10.82	2.11	5.34
Soluble	14	2.50	1.11	2.33
Insoluble	14	8.12	1.5	3.11

[a] The data are taken from a MAFF trial [15]

8 COMPARISON WITH THE PROSKY PROCEDURE

The UK MAFF commissioned a study in 1996 to determine the reason for the differences in values obtained for dietary fiber by the widely used Prosky procedure and by the Englyst NSP procedures, the two major contenders for fiber labeling for foods.[41]

Analytical values used with intake values to calculate the amounts of ''fiber'' provided in the UK diet and the results are given in Table 7.

Equal ''fiber'' intakes are obtained for the Meat products group, as expected, and only small differences are observed for the Fruit products and Nuts groups. However, the intakes calculated with the Prosky values for the other seven food groups are, on average, 37% higher than those calculated using the NSP values, ranging from 20% (Green vegetables) to 85% (Other vegetables) higher for single food groups.

The results of the study showed:

- MRP produced by heating various combinations of carbohydrates and proteins are measured as dietary fiber by the Prosky procedure.
- Some starch escapes digestion in vivo because it is inaccessible to enzymes (resistant starch types 1 and 2;

Table 7 Calculated intakes of ''fiber'' using Prosky values and Englyst values

Food group	Food intake ($g\,day^{-1}$)	''Fiber'' intake ($g\,day^{-1}$)	
		Prosky	Englyst
Bread	110	4.1	3.2
Other cereals	90	3.6	2.9
Meat products	44	0.2	0.2
Green vegetables	37	1.2	1.0
Potatoes	133	3.2	2.5
Other vegetables	73	2.4	1.3
Canned vegetables	35	1.0	0.7
Fresh fruit	65	1.2	0.9
Fruit products	43	0.3	0.2
Nuts	2	0.2	0.1
Total	632	17.4	13.0

see the article **Starch Analysis in Food**), but this starch is made available to enzymatic hydrolysis in vitro by the milling and heating steps in the Prosky procedure, and therefore is not included in Prosky fiber values. In contrast, starch that is digestible in foods that are normally eaten hot may retrograde during the cooling of samples for analysis to a form that is not digestible by the enzymes used in the Prosky procedure and is hence included in the measurement.

- Fat, even at levels below that requiring extraction according to the protocol, can be included in the Prosky values.
- Organic acids may be included in the Prosky values, and the proportion included is increased in the presence of calcium.

The stated aim of the Prosky procedure is to measure the sum of indigestible polysaccharides and lignin but this is not achieved, since the procedure does not measure inulin or all resistant starch. Further, the Prosky procedure includes artifacts, such as MRP, and highly variable amounts (up to 40% of the material measured) of unidentified material. This, together with the conceptual difference in the definition of dietary fiber, is the reason for the difference between values obtained by the Englyst procedure and those obtained by the Prosky procedure. Note that because, in general, the Prosky procedure does include NSP, Prosky values will tend to be equal to or greater than values for NSP, and the Prosky value may be further increased by food processing.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
Ara	Arabinose
BCR	Community Bureau of Reference
CRM	Certified Reference Material
DMG	Dimethylglutaric Acid
DMSO	Dimethyl Sulfoxide
Fuc	Fucose
Gal	Galactose
GalA	Galacturonic Acid
GLC	Gas–Liquid Chromatography
Glc	Glucose
GlcA	Glucuronic Acid
HPLC	High-performance Liquid Chromatography
IS	Internal Standard
MAFF	Ministry of Agriculture, Fisheries and Food
Man	Mannose
MRP	Maillard Reaction Products
NCP	Noncellulosic Polysaccharides
NSP	Nonstarch Polysaccharides
PAD	Pulsed Amperometric Detection
Rha	Rhamnose
Uac	Uronic Acids
Xyl	Xylose

RELATED ARTICLES

Food **(Volume 5)**
Starch Analysis in Food

REFERENCES

1. D.P. Burkitt, H. Trowell, (eds.), *Refined Carbohydrate Foods and Disease. Some Implications of Dietary Fiber*, Academic Press, New York, 1975.
2. T.L. Cleave, G.D. Campbell, N.S. Painter, *Diabetes, Coronary Thrombosis and the Saccharine Disease*, John Wright, Bristol, 1969.
3. H. Trowell, 'Ischemic Heart Disease and Dietary Fiber', *Am. J. Clin. Nutr.*, **25**, 926–932 (1972).
4. Commission of the European Communities, Directorate Industry, III/E/1, Scientific Committee for Food, Document III/5481/93-EN-FINAL, Brussels, 21 October 1993.
5. FAO/WHO, 'Carbohydrates in Human Nutrition. Report of a Joint FAO/WHO Expert Consultation, Rome, 14–18 April 1997'. FAO Food and Nutrition Paper No. 66, 1–140, FAO, Rome, 1998.
6. H.N. Englyst, G.J. Hudson, 'The Classification and Measurement of Dietary Carbohydrates', *Food Chem.*, **57**, 15–21 (1996).
7. H.N. Englyst, M.E. Quigley, G.J. Hudson, 'Definition and Measurement of Dietary Fiber', *Eur. J. Clin. Nutr.*, **49**(Suppl. 3), S48–S62 (1995).
8. M.E. Quigley, G.J. Hudson, H.N. Englyst, 'Determination of Resistant Short-chain Carbohydrates (Non-digestible Oligosaccharides) using Gas–Liquid Chromatography', *Food Chem.*, **65**(3), 381–390 (1999).
9. K.N. Englyst, H.N. Englyst, G.J. Hudson, T.J. Cole, J.H. Cummings, 'Determination of Rapidly Available Glucose in Foods. A Measurement that Reflects the Glycemic Response', *Am. J. Clin. Nutr.*, **69**, 448–454 (1999).
10. H.N. Englyst, M.E. Quigley, G.J. Hudson, 'Determination of Dietary Fiber as Non-starch Polysaccharides with Gas–Liquid Chromatographic, High-performance Liquid Chromatographic or Spectrophotometric Measurement of Constituent Sugars', *Analyst*, **119**, 1497–1509 (1994).
11. H.N. Englyst, M.E. Quigley, G.J. Hudson, J.H. Cummings, 'Determination of Dietary Fiber as Non-starch Polysaccharides by Gas–Liquid Chromatography', *Analyst*, **117**, 1707–1714 (1992).
12. D.A.T. Southgate, 'Determination of Carbohydrates in Foods. II. Unavailable Carbohydrates', *J. Sci. Food Agric.*, **20**, 331–335 (1969).

13. H.N. Englyst, S.A. Bingham, S.A. Runswick, E. Collinson, J.H. Cummings, 'Dietary Fiber (Non-starch Polysaccharides) in Fruit, Vegetables and Nuts', *J. Hum. Nutr. Dietet.*, **1**, 247–286 (1988).
14. H.N. Englyst, S.A. Bingham, S.A. Runswick, E. Collinson, J.H. Cummings, 'Dietary Fiber (Non-starch Polysaccharides) in Cereal Products', *J. Hum. Nutr. Dietet.*, **2**, 253–271 (1989).
15. R. Wood, H.N. Englyst, D.A.T. Southgate, J.H. Cummings, 'Determination of Dietary Fiber in Foods – Collaborative Trials. IV. Comparison of Englyst GLC and Colorimetric Measurement with the Prosky Procedure', *J. Assoc. Publ. Anal.*, **29**, 57–141 (1993).
16. A.W. Pendlington, N. Meuree-Vanlaethem, A. Brookes, 'The Method Specific Certification of the Mass Fraction of Dietary Fiber in Lyophilised Haricot Beans, Carrot, Apple, Full Fat Soya Flour and Bran Breakfast Cereal Reference Materials, CRMs 514, 515, 516, 517 and 518', Office for Official Publications of the European Communities, Luxembourg, 1996.
17. B. Holland, I.D. Unwin, D.H. Buss, *Cereals and Cereal Products; Composition of Foods*, Royal Society of Chemistry, London, 1988.
18. B. Holland, I.D. Unwin, D.H. Buss, *Vegetables, Herbs and Spices. The Fifth Supplement to McCance and Widdowson's Composition of Foods*, Ministry of Agriculture, Fisheries and Food, London, and Royal Society of Chemistry, Cambridge, 1991.
19. B. Holland, A.A. Welch, I.D. Unwin, D.H. Buss, A.A. Paul, D.A.T. Southgate, *McCance and Widdowson's the Composition of Foods*, 5th edition, Ministry of Agriculture Fisheries and Food, London, and Royal Society of Chemistry, Cambridge, 1991.
20. E.H. Hipsley, '"Dietary Fiber" and Pregnancy Toxaemia', *Br. Med. J.*, **ii**, 420–422 (1953).
21. G.B. Haber, K.W. Heaton, D. Murphy, 'Depletion and Disruption of Dietary Fiber. Effects on Satiety, Plasma Glucose and Serum Insulin', *Lancet*, **i**, 679–682 (1977).
22. H. Trowell, 'Dietary Fiber; a Paradigm', in *Dietary Fiber, Fiber-Depleted Foods and Disease*, eds. H.C. Trowell, D. Burkitt, K.W. Heaton, Academic Press, London, 1–20, 1985.
23. H.C. Trowell, D.A.T. Southgate, T.M.S. Wolever, A.R. Leeds, M.A. Gassull, D.J.A. Jenkins, 'Dietary Fiber Redefined', *Lancet*, **i**, 967 (1976).
24. H.N. Englyst, H. Trowell, D.A.T. Southgate, J.H. Cummings, 'Dietary Fiber and Resistant Starch', *Am. J. Clin. Nutr.*, **46**, 873–874 (1987).
25. D.A.T. Southgate, H.N. Englyst, 'Dietary Fiber: Chemistry, Physical Properties and Analysis', in *Dietary Fiber, Fiber-Depleted Foods and Disease*, eds. H. Trowell, D. Burkitt, K. Heaton, Academic Press, London, 31–55, 1985.
26. P.E. Koivistoinen, N.-G. Asp, H.N. Englyst, G.J. Hudson, L. Hyvönen, H. Kallio, P.P. Salo-Väänänen, 'Memorandum on Terms, Definitions, and Analytical Procedures of Protein, Fat and Carbohydrates in Food for Basic Composition Data: Issues and Recommendations', *Food Chem.*, **57**, 33–35 (1996).
27. R.R. Selvendran, J.A. Robertson, 'The Chemistry of Dietary Fiber: an Holistic View of the Cell Wall Matrix', in *Dietary Fiber: Chemical and Biological Aspects*, eds. D.A.T. Southgate, K. Waldron, I.T. Johnson, G.R. Fenwick, Special Publication No. 83, 27–43, Royal Society of Chemistry, Cambridge, 1990.
28. H.N. Englyst, S. Hay, G.T. Macfarlane, 'Polysaccharide Breakdown by Mixed Populations of Human Fecal Bacteria', *FEMS Microbiol. Ecol.*, **45**, 163–171 (1987).
29. J.H. Cummings, H.N. Englyst, 'Fermentation in the Human Large Intestine and the Available Substrates', *Am. J. Clin. Nutr.*, **45**, 1243–1255 (1987).
30. H.N. Englyst, J.H. Cummings, 'Digestion of the Carbohydrates of Banana (*Musa paradisiaca sapientum*) in the Human Small Intestine', *Am. J. Clin. Nutr.*, **44**, 42–50 (1986).
31. H.N. Englyst, J.H. Cummings, 'Digestion of Polysaccharides of Potato in the Small Intestine of Man', *Am. J. Clin. Nutr.*, **45**, 423–431 (1987).
32. H.N. Englyst, J.H. Cummings, 'Digestion of the Polysaccharides of Some Cereal Foods in the Human Small Intestine', *Am. J. Clin. Nutr.*, **42**, 778–787 (1985).
33. H.N. Englyst, H.S. Wiggins, J.H. Cummings, 'Determination of the Non-starch Polysaccharides in Plant Foods by Gas–Liquid Chromatography of Constituent Sugars as Alditol Acetates', *Analyst*, **107**, 307–318 (1982).
34. U. Schlemmer, 'Studies of the Binding of Copper, Zinc and Calcium to Pectin, Alginate, Carrageenan and Gum Guar in HCO_3–CO_2 Buffer', *Food Chem.*, **32**, 223–234 (1989).
35. C. Hoebler, J.-L. Barry, A. David, J. Delort-Laval, 'Rapid Acid Hydrolysis of Plant Cell Wall Polysaccharides and Simplified Quantitative Determination of Their Neutral Monosaccharides', *J. Agric. Food Chem.*, **37**, 360–367 (1989).
36. H.N. Englyst, G.J. Hudson, 'Colorimetric Method for Routine Measurement of Dietary Fiber as Non-starch Polysaccharides. A Comparison with Gas–Liquid Chromatography', *Food Chem.*, **24**, 63–76 (1987).
37. M.E. Quigley, H.N. Englyst, 'Determination of Neutral Sugars and Hexosamines by High-performance Liquid Chromatography with Pulsed Amperometric Detection', *Analyst*, **117**, 1715–1718 (1992).
38. M.E. Quigley, H.N. Englyst, 'Determination of the Uronic Acid Constituents of Non-starch Polysaccharides by High-performance Liquid Chromatography with Pulsed Amperometric Detection', *Analyst*, **119**, 1511–1518 (1994).
39. V39, 'Dietary Fiber: Englyst Procedure for Determination of Dietary Fiber as Non-starch Polysaccharides: Measurement of Constituent Sugars by Gas–Liquid

Electrophoresis and Isoelectric Focusing in Food Analysis

Paul F. Cancalon
Florida Department of Citrus, Lake Alfred, USA

1 Introduction 1
2 Gel Electrophoresis 2
2.1 Polyacrylamide Gel Electrophoresis 2
2.2 Isoelectric Focusing 2
2.3 Two-dimensional Gel Electrophoresis 2
2.4 Agarose Electrophoresis 2
2.5 Nondenaturing Gel Electrophoresis 3
2.6 Isotachophoresis 3
2.7 Food Applications 3
3 Capillary Electrophoresis 3
3.1 Capillary Electrophoresis Separation Techniques 4
3.2 Detection 5
3.3 Application to Food Analysis 6
4 Conclusion 14
Abbreviations and Acronyms 14
Related Articles 15
References 15

Electrophoresis is based on the differential migration of charged components in an electric field. Until recently, gel electrophoresis and particularly sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) were the main electrophoretic techniques used in food analysis. Recently, capillary electrophoresis (CE) has become widely adopted for the monitoring of many food components.

Such rapid development is due to several factors, but especially to speed and versatility. Most often separation can be performed in a few minutes with minimal sample preparation. A new method can be started by filling the capillary with the proper buffer, and new modular instruments can be rapidly fitted with the appropriate detector. Various CE procedures have been developed since the introduction of this technique. Capillary zone electrophoresis (CZE), the basic CE method, consists of a capillary tubing of fused silica with a diameter of 50–150 μm. The tube is filled with a buffer and subjected to an electric field, and under these conditions the positive charged species migrate toward the cathode at a rate determined by their mass-to-charge ratios. However, a net movement of the buffer toward the cathode, electroosmotic flow (EOF), is generated within the tube by the ionized capillary wall silanol groups. As a result, all the molecules present in the buffer proceed toward the cathode, the neutral molecules migrate as a single group, and each negative species moves toward the cathode at a rate determined by the difference between the EOF and its anodic attraction. The other main CE techniques are micellar electrokinetic chromatography (MEKC), where an anionic surfactant is added to the buffer to form negatively charged micelles, isotachophoresis (ITP), in which the sample is sandwiched between electrolytes of high and low mobility and the analytes migrate in order of decreasing mobility, and isoelectric focusing (IEF), which is one of the main techniques used in protein gel electrophoresis: amphoteric electrolytes are allowed to migrate in a capillary where they form a pH gradient. In such a gradient, proteins move until they reach their isoelectric point (pI). Many detection techniques have been applied to CE, including direct and indirect ultraviolet (UV) absorption, fluorescence, electrochemical detection and even mass spectrometry (MS).

Most food components can be examined with these techniques, particularly proteins, ions, organic acids, sugars and various plant chemicals.

1 INTRODUCTION

In order to maintain quality, food destined for human consumption is submitted to a considerable number of analyses from the raw ingredient to the final product. The analytical data must be able to assess the authenticity of the food material but should also reveal the possible addition of adulterants. Most of the chemical and physical methods that were developed have been applied to food analysis as it can be seen in the Official Methods of Analysis of the Association of Official Analytical Chemists (AOAC) International. Hundreds of methods can be found using techniques from simple titration to complex nuclear magnetic resonance (NMR) analyses.[1] Analytical methods for food products were reviewed by Chang et al.[2] Of the 26-page review only one page is devoted to electrophoresis techniques, which shows that before the development of CE, electrophoretic procedures did not account for a major part of food analytical methods. Reviews of methods of food analysis have been published by Pare and Belanger[3] and by Linden.[4]

Electrophoretic techniques are all based on the differential migration of charged components when submitted to an electric field.[5,6] The practical aspects of electrophoresis have been reviewed in great detail

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

by Westermeier.[6] The principles and applications of electrophoresis to the analysis of food and agricultural products were reviewed by Autran,[7] who described four different modes of electrophoretic techniques:

1. mobile front electrophoresis
2. zone electrophoresis
3. IEF
4. ITP.

Only the last three have present-day applications.

2 GEL ELECTROPHORESIS

These methods separate components mainly according to size, the small molecules migrating faster than the larger ones. They are very powerful and are widely used in medical research for protein and nucleic acid analysis. However, since in most cases these techniques are not automated, they require skilled analysts and are usually time-consuming. Detection has been a major problem in gel analysis. Separated materials are trapped in a matrix and not easily quantified. Preferential staining of the proteins embedded in the gel with dyes or silver has been the main visualization method. Densitometric analysis of the stained proteins after removal of the excess dye has been the main quantitation procedure.

However, staining is linear only at low concentrations, it can vary between proteins and it is not exactly reproducible since many factors during separation, staining and destaining may affect the final gel (for a review, see Chen and Evangelista[8]). Gels, however, are very convenient when radioactive compounds are monitored. Another technique of detection which has been particularly useful is immunoblotting. With this technique, molecules separated by an electrophoretic method are transferred from a gel to an immobilized membrane where reactions with ligands such as antibodies are performed. Reaction with its antibody can reveal the presence of a particular protein with great specificity or identify a protein in a one- or two-dimensional (2-D) gel.[7]

2.1 Polyacrylamide Gel Electrophoresis

SDS PAGE is used for the analysis of peptides and proteins. The proteins bind to a negatively charged detergent (SDS) and are allowed to migrate in a polymeric gel.[7] Since all SDS–protein complexes have similar mass-to-charge ratios, separation depends on the size of the protein.[9] The pores of the gel can be controlled by modifying the acrylamide concentration and the level of cross-linkage of the polymer, which can be the same all over the gel or increase linearly or even exponentially. It is therefore possible to select a gel for a particular protein size range.

Gel electrophoresis is done either by disk electrophoresis, where a single analytical separation is performed in a cylindrical tube of gel or by slab electrophoresis, where separation occurs in a sheet of gel that can contain many individual wells. Electrophoresis can be continuous when the same gel and buffer are used or discontinuous when a stacking gel is used before the electrophoresis.[10,11] Gel electrophoresis has been until recently the technique of choice for the analysis of peptides and proteins.[5–7,9]

2.2 Isoelectric Focusing

In this technique, a pH gradient is formed in the gel by allowing the migration in the gel of amphoteric electrolytes, available commercially in many pH ranges. They are electrophoresed in the gel where they form a pH gradient.[7] When submitted to an electric field, charged proteins move into the pH gradient until they reach their pI where they become neutral, stop and accumulate as a sharp peak that can be stained or used in 2-D slab gel.

2.3 Two-dimensional Gel Electrophoresis

Many proteins have identical or similar molecular weight (MW) or pI and are not separated by the methods examined so far. O'Farrell[12] combined the IEF and SDS PAGE methods to separate proteins according to their pI and MW. The proteins are first separated on a disk gel by IEF. The gel is then deposited on a slab gel and the proteins are allowed to separate again according to their MW.

In the final 2-D slab gel, dozens of proteins may appear for a given MW and 1000 proteins can be revealed. This method is extremely powerful, but is very time-consuming and difficult to perform. We were able to detect minor changes in the protein composition of fish nervous tissues separated first by IEF between pH 4 and 8 followed by SDS PAGE between 10 and 200 kDa.[13]

2.4 Agarose Electrophoresis

Agarose can also be used to make gels; such gels at a concentration of 1% have very large pores and are used for large proteins. Agarose gels are the main medium for the separation of DNA and RNA fragments. Pulsed field electrophoresis is used for nucleic acid separation.[14] These methods have received only limited application in the area of food analysis; however, the recent introduction of genetically modified organisms in many food products has drastically increase the demand for DNA analysis of food components.

2.5 Nondenaturing Gel Electrophoresis

SDS and other chemicals such as urea and mercaptoethanol used in SDS PAGE produce the dissociation of the proteins into their individual subunits. The separation of native proteins can be done under conditions that preserve their conformation, subunit interactions and even biological activity. The original Ornstein–Davis buffer system is such a nondenaturing electrophoretic buffer.[10,11]

2.6 Isotachophoresis

This electrophoretic system consists of a leading and a terminating electrolyte containing anions of greater and lower mobilities than that of the sample to be analyzed. During separation, a potential gradient is formed and zones containing individual substances are formed. The gel must allow the nonrestrictive migration of the protein, and polymer with only 5% acrylamide is often used. Proteins or ampholytes (spacers), with mobilities intermediate between those of two proteins being examined, can be added to improve separation.[7]

2.7 Food Applications

Most of the methods examined so far are relatively complex and their use in food analysis has been mostly limited to research problems. Electrophoretic analysis of food, not including CE, has been reviewed by Autran[15] and dealt mostly with protein analysis. The produce analyzed ranged from the egg content of pasta to milk proteins and wheat composition. A review on the use of electrophoretic techniques for wine study has been published by Carnevillier et al.[16] Proteins from specific animal species have been identified by electrophoresis; this is particularly important when one type of food is expensive and subject to fraud. Huang et al.[17] identified red snapper by SDS PAGE, IEF and 2-D gel methods; the best results were obtained on 10–12.5% gel and an ampholyte mixture of 20% pH 3–10 and 80% pH 4–6.5. More recently, Hsieh et al.[18] developed an IEF method for the identification of red snapper; they found that bands in the area (pH 3.5, 9.5) can be used as a marker for this fish.

Cereal genetics have been investigated by electrophoresis. Dougherty et al.[19] examined 14 cultivars of hard red winter wheat and found a correlation between some gliadin and glutenin proteins and loaf volume. Furthermore, 2-D electrophoresis by the same group[20] showed that image analysis of the 2-D gels provided a reproducible method for quantitating wheat proteins. A similar study was performed by Picard et al.,[21] who analyzed durum wheat leaf proteins by 2-D electrophoresis and statistically analyzed the resulting gels to allow line identification. Corn zeins were examined by Drzewiecki[22] by discontinuous PAGE. Changes in strawberry proteins labeled with [^{35}S]methionine were examined using SDS PAGE.[23] Various other types of food proteins have been also examined by electrophoresis. Santoro analyzed wine must proteins by IEF and SDS PAGE.[24] Two consecutive series of stacking and running gels were used by Kubis and Gros to examine rabbit skeletal muscle.[25] SDS PAGE (5%) and Western immunoblotting were used by Huff-Lonergan et al.[26] to follow the post mortem degradation of titin and nebulin, two proteins associated with meat tenderness, in beef muscle. 2-D gel electrophoresis and antiserum reactions were used to isolate ADP-glucose pyrophosphorylase from tomato.[27] Similarly, the isozymes of several citrus enzymes were followed by King et al.[28] using PAGE analysis. Chalcone isomerase was examined by Fouche and Dubery[29] by SDS PAGE, native PAGE analysis and IEF. King et al.[30] separated citrus isozymes by PAGE using buffers best suited for each enzyme. The adulteration of goat cheese with cow milk was detected by Addeo et al.[31] using IEF of β-lactoglobulins. Antibiotic residues were detected in milk at the parts per billion level with a 2% agarose gel with a tris(hydroxymethyl)aminomethane (Tris) buffer (pH 8), followed by bioautographic detection with *Bacillus stearothermophilus*.[32] Pulsed field gel electrophoresis was used by Izquierdo-Cana et al.[33] to examine the electrophoretic karyotypes of 74 strains of *Saccharomyces* from fermenting wine musts.

3 CAPILLARY ELECTROPHORESIS

CE is a recent technique that has become widely adopted in the last few years with the availability of commercial instruments.[34] Camilleri[35] has published a history of CE showing the exceptional growth of this technique during the 1990s. CE has become a major analytical tool in the areas of pharmaceutical and food analysis. Such rapid development is due to several factors, but especially to speed and versatility. Most often separation can be performed in a few minutes with minimal sample preparation. The lack of precision noted in early instruments has been overcome and in many cases the precision and repeatability are similar to those obtained with high-performance liquid chromatography (HPLC) and superior to those provided by gel electrophoresis. Newer instruments have become particularly versatile with the introduction of modular apparatus, where a CE unit can be rapidly fitted with the appropriate detector chosen from a wide array of instruments available: detection by UV, diode-array, visible, fluorescence, conductivity and even mass spectrometric methods are

now possible.[36] The rise in the use of CE in food analysis has been documented in several reviews. Early work was reviewed by Zeece[37] and more recent work by several groups.[38–41]

3.1 Capillary Electrophoresis Separation Techniques

The various CE procedures developed since the introduction of this technique have been extensively reviewed in many publications.[42–48] Most of these techniques represent CE adaptation of procedures used in gel electrophoresis and described previously.

3.1.1 Capillary Zone Electrophoresis

CZE can be viewed as the basic CE method from which all others are derived.[41–50] CZE involves the use of a capillary tube usually made of fused silica with a diameter of 50–150 μm. The tube is filled with a buffer and subjected to an electric field, and under these conditions the positively charged species migrate toward the cathode at a rate determined by their mass-to-charge ratios. However, an EOF or net movement of the buffer toward the cathode is generated within the tube by the ionized capillary wall silanol groups. As a result, all the molecules present in the buffer proceed toward the cathode; the neutral molecules migrate as a single group transported by the EOF, and each negative species moves toward the cathode at a rate determined by the difference between the EOF and its anodic attraction.

3.1.2 Capillary Wall Modifications

In CZE, the EOF controls the movement of all the transported molecules, so many efforts have been made to manipulate the process.[51,52] The EOF can be slowed, stopped or even reversed. These EOF modifiers act by interacting with the capillary wall silanol groups. These compounds can be solvent additives that do not require modification of the capillary itself; other compounds are adsorbed on the wall while others are chemically bonded to the silanol groups. These modifications have allowed the development of separation using reversed polarity and are very useful for the separation of chemicals such as proteins that have a tendency to bind to capillary walls.

3.1.3 Micellar Electrokinetic Chromatography

Many chemicals are either neutral or hydrophobic and cannot be readily separated by CZE. In MEKC,[53,54] an anionic surfactant is added to the buffer to form negatively charged micelles. An analyte, depending on its hydrophobicity, will be partitioned between the water and the micelles. The mobility of the analyte will be determined by its own velocity in water, that of the micelles, the EOF and the partition coefficient between water and micelles.

3.1.4 Chiral Electrophoresis

Chiral separation works on a principle similar to that of MEKC. Molecules such as a cyclodextrin (CD) can form complexes with other molecules depending on their spatial configuration. The migration velocity is dependent on the stability of the complexes and the migration rates of the various species. Separation of enantiomers is achieved if one configuration has more affinity for the CD than the other. Chiral CE has been extensively used in the analysis of pharmaceutical drugs (reviews are available[55,56]).

3.1.5 Isotachophoresis

In ITP, the sample is sandwiched between electrolytes of high and low mobility and the analytes migrate in order of decreasing mobility until a steady state is reached.[57] This method has been useful for concentrating dilute samples.[58] The actual instrumentation can be somewhat complex and can involve switching valves, since various buffers have to be injected.

3.1.6 Isoelectric Focusing Electrophoresis

As seen previously, IEF is one of the main techniques used in protein gel electrophoresis, it is an essential part of 2-D gel electrophoresis and it has been adapted to CE. Amphoteric electrolytes are allowed to migrate in a capillary where they form a pH gradient. In such a gradient, proteins move until they reach their pI and accumulate as a sharp peak that can be subsequently eluted. Methods where the pH gradient and the separation are run consecutively or simultaneously have been developed.[59,60] Even 2-D analyses have now been developed for CE.

3.1.7 Polymer Sieving Electrophoresis

As for IEF, the techniques developed for disk and slab gels have been adapted to CE. Polyacrylamide and agarose gels have been used to separate proteins, carbohydrates and nucleotides. These filled capillaries can provide interesting results but they are relatively expensive and deteriorate rapidly. The capillary can also be filled with linear polymers such as noncross-linked acrylamide, which perform similar separations and have the advantage of being replaceable.[61]

3.1.8 Electrochromatography

In the last few years, a new technique has been implemented that combines CE and HPLC.[62,63] The

EOF is used as an electroosmotic pump to move the solvent rather than pressure as in HPLC. Typically, a 25-μm capillary is packed with a very fine material used in reversed-phase HPLC separation. A state-of-the-art review of capillary electrochromatography has been published by Stevenson et al.[64] summarizing the potentials and drawbacks of this technique, which is still being developed and has not received large numbers of applications.

3.2 Detection

In early CE, detection was a major problem (for a review, see Albin et al.[65]). Since the tube itself is used as the detection cell, the light path is very small, and the sensitivity is limited. In the last few years, considerable progress has been made and now CE detection is usually equivalent to that of other techniques. Many types of detection have been adapted to CE[66–71] and with the introduction of modular equipment, easily replaceable detectors are available for the same instrument and in some cases can be used in series during an analysis.

3.2.1 Ultraviolet/Visible Absorption

UV detection was the original CE detection method.[67] Detectors using single and multiple wavelengths were first introduced. Diode-array and high-speed scanning have allowed the entire UV spectrum to be examined simultaneously. Many modifications of the detection system have been implemented to increase sensitivity. Changes in the capillary geometry such as local enlargement (bubbles) or Z cells have been used to increase the path, and UV lasers have been shown to enhance the detection limit (DL) up to fourfold. Many molecules such as carbohydrates, organic acids and inorganic ions do not absorb readily at the usual UV wavelengths. Indirect UV detection[69] has been extensively used in CE analyses. With this method, a molecule absorbing strongly in the UV region is added to the buffer. During separation the nonabsorbing compounds displace the chromophore and create a zone of lower UV absorbance. As a result, a negative peak is created, that can be recorded and quantified.

3.2.2 Fluorescence

Fluorescence detectors are now available for CE, and instruments with laser-induced fluorescence (LIF) have been shown to be particularly sensitive.[70,71] Some compounds can fluoresce naturally but, in most cases, modifications of the molecules are necessary. Following derivatization of analytes such as amino acids (AAs) and peptides, detection at the attomole level has been reported. As with UV methods, indirect fluorescence detection is also possible.

3.2.3 Electrochemical Detection

Electrochemical detection in CE was found to be technically complex, since the detector had to be separated from the capillary to avoid electric field interference.[72–74] The detection had to be off- or end-column and until recently commercial instruments were not available. Conductivity detection in CE has been shown to be most useful for the analysis of organic and inorganic ions. Amperometric and particularly pulsed amperometric detection (PAD) are still not available at the commercial level although they would be extremely useful for sugar and carbohydrate analysis in food monitoring, but many laboratory-made instruments have been produced and provided excellent results.

3.2.4 Derivatization

Derivatization of compounds with poor or no chromophores can greatly improve the detection of CE-separated chemicals.[75,76] In CE most derivatizations are performed before the electrophoresis; however, some post- and even on-electrophoresis analyses have been performed. Added groups usually generate fluorescent or UV-absorbing molecules. Amines, aldehydes, ketones and carboxyl groups are most often targeted, and therefore derivatization has been mainly used to improve the detection of AAs, proteinaceous compounds and carbohydrates.

3.2.5 Capillary Electrophoresis and Mass Spectrometry

In recent years, it has become possible to couple CE with electrospray MS using various interface techniques.[77] These techniques are most useful for the identification and characterization of biological molecules. Protein and drug chemistry have so far been the main beneficiaries of these technical developments.

3.2.6 Immunoassays and Affinity Electrophoresis

Finally, methods involving specific binding are also used in CE analysis. Immunoreaction with LIF is one of the most often used techniques where the antibody–fluorescently labeled antigen complex is revealed.[78] Other types of receptor–ligand complexes such as enzyme–substrate complexes have also been used.[79–81] This is done by producing either a mobile complex with specific migration properties or by immobilizing one of the components of the complex.

3.3 Application to Food Analysis

CE has proved to be effective in analyzing most of the molecules found in biological material. Any method that allows the analysis of the components of plant and animal tissues could also be used for food analysis, although the border between studies related to food and plant and animal physiology is somewhat blurred. Therefore, applications in other areas such as pharmaceutical and biomedical, plant and animal physiology are also relevant and may provide useful information for the examination of food components. Perrett[82] has reviewed the use of CE in biomedical and pharmaceutical research.

3.3.1 Carbohydrates

Several reviews on the analysis of saccharides by CE have been published.[83–86] New methods for the analysis of oligosaccharides have been examined by Grimshaw[87] and a detailed protocol for the CE analysis of oligosaccharides has even been reported by Linhardt[88] in *Methods in Enzymology*. However, relatively little work had been done in the area of food carbohydrate analysis, although simple sugars and oligo- and polysaccharides are very common and are extensively analyzed in food monitoring.

Saccharides absorb only in the far-UV region (at about 190 nm) and can be detected by direct UV detection. However, the sensitivity is not very good and many other compounds also absorb in this area of the spectrum, but since UV detectors are readily available, the far-UV region has been used to detect many chemicals. Direct UV analysis of sugars was performed using copper chelation.[89] More recently, direct UV detection at 192 nm was used by Zhong et al.[90,91] to separate pectins, an important food product, with a 50 mM phosphate buffer (pH 7). A similar separation was obtained with citrus pectins using a 30 mM borate buffer (pH 9.3) containing 5% acetonitrile with direct UV detection at 192 nm.[92]

The separation of borate complexes was one of the first types of CE analysis developed for saccharides. In the past, this method was used to examine pea oligosaccharides at 195 nm.[93] Indirect detection and borate complexation (200 mM) have been combined to analyze sucrose at the limit of 2 nM with *p*-nitrophenol as a chromophore at 400 nm.[94] A different approach was taken by Brewster and Fishman,[95] who measured a colored reaction product generated by the formation of complexes between starch and iodine with separation and detection in the visible region.

Indirect UV detection has been used extensively to analyze carbohydrates. The methods usually require a high pH to insure ionization of the sugars, but the sensitivity decreases at very high pH because of competition with hydroxide ions. The conditions for the indirect UV analysis of sugars were examined by Xu et al.[96] using sorbate, riboflavin and 3,4-dimethoxycinnamic acid as chromophores. DLs 25 times better than those published with similar techniques were obtained with a riboflavin–NaOH buffer (pH 12.3). Specifically in food analysis, fruit juice sugars from apple, orange and grape were measured by Klockow et al.[97] and Klockow-Beck[98] by indirect UV detection, with a 6 mM sorbate buffer (pH 12.2). The same group[99] compared results obtained by their CE method and by HPLC/PAD. CE DLs were 0.2–0.3 mM whereas the PAD limits were as low as 0.5–1 μM. However, both methods gave accurate and reproducible results. Indirect UV detection was used to examine sugars (glucose, fructose and sucrose) in citrus juices; with sorbate at pH 11.5, sugars can be analyzed in 12 min.[100] The results for citrus juices were similar to those obtained by HPLC/PAD,[101] although the detection level of HPLC was much lower. However, the main advantage of CE over HPLC is that the method does not require any preparation apart from dilution and filtration. Similarly, sugars from soft drinks and uronic acids were separated by Zemann et al.[102] in 4 min with 6 nM sorbate buffer (pH 12.1) using reversed polarity with hexadimethrine bromide. The use of tryptophan (Trp) and *N*-benzylcinchonidium as a chromophore has been shown to improve the sensitivity greatly over the use of sorbate.[103]

Saccharide detection can be greatly improved by attaching a probe to the compound to be analyzed. In addition to enhancing detection, the added tag can change the saccharide charge and modify the mobility, but derivatization increases the analytical time. Acidic monosaccharides were separated as sulfanilic acid or 7-aminonaphthalene-1,3-disulfonic acid derivatives.[104] This method was applied by the same group to the analysis of sialogangliosides.[105] Mono- and oligosaccharides were derivatized with 8-aminonaphthalenetrisulfonic acid and separated by reversed polarity at pH 2.5.[106] The oligosaccharide composition of various beers was examined by Guttman et al.[107] after labeling with the fluorophore 8-aminopyrene-1,3,6-trisulfonate. Birch pulp xylan was analyzed by direct UV absorption at 232 nm after reduction of the acidic oligosaccharides with borate–NaOH buffer.[108] The authors further analyzed the acidic oligosaccharides with the same buffer and the neutral oligosaccharides after derivatization with 6-aminoquinoline using a similar buffer with detection at 245 nm. The MW of the analytes was determined by MS.[109] Cellulose degradation products were separated by Motellier et al.[110] using indirect UV detection, with 4 mM nicotinic acid as a chromophore and 15 mM creatinine on an uncoated capillary.

For many years, PAD has been the detection technique of choice for carbohydrate analysis by HPLC, where it

has been extensively used in food analysis to monitor sugar adulteration.[101] However, no such detector is yet available commercially for CE. All the studies reported with PAD were performed with laboratory-made equipment, and detectors using copper electrodes were found to be well suited for CE analysis of sugars. Colon et al.[111] used a copper microelectrode to examine glucose and fructose in carbonated soft drinks. A similar method was used by Ye and Baldwin,[112] who determined the saccharide composition of apple juice with a DL at or below the femtomole level. Weber and Lunde[113] used CE/PAD to examine carbohydrates in a large number of biological materials such as bovine fetuin and coagulation factors; an 8 mM sodium borate buffer (pH 10.4) provided the best separation. Separation of polysaccharides according to molecular size was achieved by reversing the EOF with cetyltrimethylammonium bromide (CTAB).[114] A copper(I) oxide electrode with 0.1 M NaOH solution was used by Huang and Kok[115] to separate sugars at the micromole level. Sugars, sugar alcohols and sugar acids in fruit juices and soft drinks were examined by Goto et al.[116] using amperometric detection with a copper electrode. The addition of cationic surfactant to the alkaline buffer doubled the sensitivity and a DL of 2.5 fmol was reported.

An interesting on-column enzymatic analysis of glucose was performed by Jin et al.[117] Two enzymatic reactions were combined to generate fluorescent compounds and measure glucose at the nanomolar level in volumes as small as 200 nL.

3.3.2 Amino Acids and Proteins

Many techniques have been developed for AA analysis by CE and detailed reviews been published,[118,119] but so far the use of CE to examine AAs in food has been limited. As for carbohydrates, UV detection is not an ideal choice for AA analysis since they absorb only in the very far-UV region. Toldra and Aristoy[120] examined the essential AA composition of dry cured ham with citrate buffer (pH 2.5) and monitoring at 200 nm. A similar analysis was done on Cheddar cheese.[121] Recently, Klampfl et al.[122] analyzed free underivatized AAs by direct UV absorption at 185 nm with a 10 mM NaH_2PO_4 buffer containing 30 mM octanesulfonic acid (pH 2.36). They used the method to analyze the AA composition of fresh and canned orange juice and several beers. Three aromatic AAs, namely, phenylalanine (Phe), tyrosine (Tyr) and Trp, absorb significantly in the UV region and were examined in citrus juice with a method developed for the general analysis of citrus juice by UV detection. Their presence in significant quantity was found to be associated with freshness.[123,124] AA derivatives have also been examined. Tomlinson et al.[125] examined the reaction products of 5-hydroxymethylfurfural and glycine as a model of the Maillard reaction. The same group[126] examined the colored products from the Maillard reaction with a phosphate buffer (pH 6.5), and showed a better separation of the products than with an HPLC method. A similar study was performed by Ames et al.[127] 5-Hydroxymethyl-2-furaldehyde and 2-furaldehyde were estimated in fruit juices by MEKC with direct sample injection.[128]

Derivatization represents an effective method for AA detection. Many types of reagents have been tried, with pre- and postcapillary derivatization and even within the capillary. Most molecules were UV absorbing or fluorophores. Dansylated AAs were separated by Skocir et al.[129] from corn and seaweed, using MEKC with a 20 mM borax buffer (pH 9.2) containing 102 mM SDS. A very sensitive analysis of fluorescamine-labeled AAs and peptides in single nerve cells was described by Shippy et al.[130] Electrochemical detection should improve the CE analysis of AAs.

Ye and Baldwin[131] used CE with electrooxidation at a Cu electrode to examine AA mixtures at the 1–10-fmol level. The method was applied to the quantitation of aspartame in various soft drinks. A similar study was performed by Sabah and Scriba[132] to examine the degradation of aspartame with a phosphate buffer at pH 2.7 and a borate buffer at pH 9.35. Amperometric detection was used by Guo et al.[133] to examine AAs without derivatization with a strong alkaline buffer (pH 12). They used the method to analyze cytochrome *c* hydrolyzates.

Proteins have been extensively analyzed by CE, and a large number of reviews have been published since the mid-1990s.[8,134–138] A review on the use of CE for the analysis of proteins and nucleic acids was published by Karger et al.[134] and the principles of peptide analysis by CZE were reviewed by Wheat.[135] Dolnik reviewed the used of CZE to analyze proteins.[136] A complete review of protein analysis by CE has been published by Chen and Evangelista,[8] Regnier and Lin[137] and Schwartz et al.,[138] who reviewed the CE separation of proteins and peptides. Detection is mostly done by direct UV absorption, but only three AAs (Phe, Tyr and Trp) absorb significantly in the UV region, and quantitation of different proteins may be difficult, since they may contain different proportions of these three AAs.

Many studies have been reported on the use of CE for food protein analysis. Recio et al.[139] reviewed the use of CE for milk products analysis, including protein polymorphism, heat treatment, proteolysis and adulteration. The caseinomacropeptide from enzyme-treated acid casein and rennet whey were analyzed at pH 2.5 in 11 min, by a method that allows the determination of the kinetics of rennet coagulation.[140]

Rennet whey solids in milk and buttermilk powder were further estimated for caseinomacropeptide by CZE using 6 M urea with a 0.15 mM citrate buffer[141] and casein phosphopeptides were examined by Adamson and Reynolds.[142] Paterson et al.[143] separated lactoglobulin by CZE, the best separation being achieved with 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), buffer (pH 8) containing 0.1% Tween and the same group[144] compared CZE (150 mM borate buffer (pH 8.5) containing 0.05% Tween) with other methods of analyzing milk proteins. Cattaneo et al.[145,146] also used CZE to analyze ewe milk. Milk proteins were extensively analyzed by Recio et al.[147] Whey proteins were also analyzed[147] using an uncoated capillary with a 100 mM borate buffer (pH 8.2) and 30 mM sodium sulfate. In order to detect adulteration of fresh milk with powdered milk, Recio and Olieman[148] estimated heat-denatured proteins in casein on a hydrophilic column with a citrate buffer containing 6 M urea and detection at 214 nm. The same group[149] compared CE, HPLC and gel isoelectrofocusing for the study of milk heat treatment CE separation was performed on an uncoated capillary with 100 mM borate buffer (pH 8.2) with 30 mM sodium sulfate. The authors concluded that CZE provided a better separation than HPLC. The same group[150] studied the proteolysis of casein on a hydrophilic coated capillary with a citric acid–citrate buffer (pH 3) containing 6 M urea and 0.5 g L^{-1} methylhydroxyethylcellulose. Cow, goat and ewe milk mixtures were analyzed on a coated column with a citrate buffer (pH 3) at 40 °C. It was concluded that as little as 8% cow milk could be detected in goat or ewe milk. McSweeney and Fox[151] used CE to follow cheese ripening.

Cereal protein composition is another area that has been extensively examined by CE. Werner[152] used CE to analyze gliadins and glutenins in wheat and determine varietal origins. He used a coated column with an aluminum buffer (pH 2.3) or a capillary filled with a polymer and a buffer containing glycerol and methanol. Wheat gliadins were analyzed by Bietz[153] with a phosphate buffer (pH 2.5) containing a linear hydrophilic polymer. Wheat glutenins[154] were analyzed by CE with a commercial buffer containing 5% methanol and the separation was similar to that obtained by SDS PAGE. The authors also compared the reproducibility of their CE method for wheat identification.[155] Lookhart and Bean[156] analyzed oat and rice proteins by CZE with phosphate buffer (pH 2.5) containing 0.05% hydroxypropylmethylcellulose and concluded that CE performed as well as or better than PAGE or HPLC. The same authors[157] analyzed wheat, oat and rice proteins on an uncoated column with a similar buffer containing 20% acetonitrile. The best separation of rice prolamins was obtained when 26 mM laurylsulfobetain was added to the buffer. Gliadins were analyzed by Capelli et al.[158] to discriminate wheat cultivars, using a buffer containing 40 mM aspartic acid with 7 M urea and 0.5% hydroxyethylcellulose. Gliadins were also examined by AbdelAal et al.[159] Varietal differentiation of *Vicia* seeds was examined from their prolamine component by CE with a 200 mM borate buffer (pH 8.3) containing 8 mM SDS.[160] Basha[161] examined peanut proteins using a 0.3% borate buffer (pH 8.3) with an uncoated column and monitoring at 214 nm.

A general procedure for the analysis of food proteins (egg white, egg yolk and milk proteins) was developed by Chen and Tusak[162] using untreated silica with a borate buffer (pH 10). The method was applied to egg white and yolk and also milk caseins and whey proteins. Separation of fermentation broth was achieved by Strege and Lagu[163] using a polyacrylamide-coated capillary. Wine proteins were analyzed by Luguera et al.[164] on an uncoated capillary with a 100 mM Tris–HCl buffer (pH 8) at 12 kV and detection at 214 nm. The method proved to be more sensitive than PAGE methods. Deng et al.[165] separated tobacco seed proteins in 12 min with a 100 mM phosphate buffer containing 20% acetonitrile and 0.05% hydroxypropylmethylcellulose. Kajiwara and Hemmings analyzed ginseng polypeptides by CE.[166]

Meat authenticity is another problem that has been investigated by CE, particularly for fish tissue. CZE at pH 7.4 was used by LeBlanc et al.[167] to examine fish sarcoplasmic proteins. Several species of fish could be differentiated in less than 10 min with a 50 mM phosphate buffer, and it was concluded that this CE method was superior to conventional electrophoretic and HPLC procedures for the screening and monitoring of fish products. A similar study was performed by Gallardo et al.[168] using CZE with 30 mM phosphate buffer (pH 2.44) and detection at 214 nm. Eight flatfish were examined and each species showed specific protein profiles that could be used for identification.

3.3.3 Lipids

Lipids are hydrophobic, often neutral and therefore difficult to analyze with conventional CE. To overcome technical problems, two approaches have been pursued: MEKC and CE in nonaqueous solvents. Nonaqueous CE has recently been reviewed by Miller and Khaledi.[169] Previously, Blomberg and Andersson[170] examined the use of CE in lipid analysis and described the separation of a mixture of isoprenyl pyrophosphate with indirect LIF detection. Nonionic aromatic compounds were examined by Shi and Fritz[171] with a borate buffer containing 50 mM tetraheptylammonium bromide and 42% acetonitrile. MEKC was used by Ingvardsen et al.[172] to separate phospholipids with a buffer containing cholate and propanol. Free fatty acids have been separated by CZE with

indirect UV detection using 5 mM diethyl butyrate as a chromophore (240 nm) in water–ethylene glycol ethyl ether containing 0.5 M trimethylammonium propanesulfonate at 45 °C. The method was applied to butter fatty acid separation after bromination of the double bonds.[173] Indirect UV detection at 264 nm with 2.5 mM anthraquinone-2-carboxylic acid and 40 mM Tris in *N*-methylformamide–dioxane was used by Drange and Lundanes[174] to separate hydrogenated fish oil fatty acids. Skin fatty acids (stearate, oleate, palmitate and myristate) were separated by Neubert et al.[175] with indirect UV detection, using 5 mM dinitrobenzoic acid, 10 mM phosphate buffer (pH 8) and 50% 1-propanol at 45 °C. The same group[176] separated phospholipids in a nonaqueous buffer of acetonitrile–2-propanol–hexane (57 : 38 : 5) with 20 mM ammonium acetate, with characterization by MS. Kleinhenz and Harper[177] examined short-chain cheese fatty acids. Hydroperoxides of unsaturated fatty acids were separated by MEKC with a buffer containing 70% 60 mM phosphate buffer (pH 6.3), 30% ethanol and 1.1 g L^{-1} 23-lauryl ether (Brij™) 35 and detected in the UV region.[178] MEKC with cholate was also used to analyze steroids by CE.[179] Recently, Heinig et al.[180] compared the analysis of saturated and unsaturated fatty acids from margarine and thistle oil by CE and HPLC. The CE method used indirect UV detection with 10 mM *p*-hydroxybenzoate, 5 mM Tris, 40 mM Brij™ 35 and 50% acetonitrile. The authors concluded that the CE method was faster and easier to use but the HPLC procedure was more reproducible and sensitive.

3.3.4 Inorganic and Organic Ions

Numerous CE methods have been developed to analyze anions, cations and organic acids. The analytical procedures for these molecules are often grouped together since these ions can be separated by similar methods. In the last few years, a large number of reviews on the CE separation of ions have been published. Haddad published a review on the CE separation of inorganic ions[181] in comparison with ion chromatographic techniques and concluded that the two methods should complement each other. A similar review was also presented by Buchberger.[182] Macka and Haddad[183] reviewed extensively the methods used for the CE analysis of metal ions. The formation of chelates and complexes was examined by Erim[184] and Buchberger.[185] The detailed procedures for the CE analysis of ions were explained by Jones.[186] The separation of inorganic anions and cations has also been reviewed by Mazzeo,[187] who examined the various techniques of inorganic ion analysis. Fritz[188] reviewed the CE analysis of inorganic anions and metal cation analysis.

Most ions absorb only in the very far-UV region and in most cases indirect detection has been used for ion analysis. Chromate was one of the first chromophores used. Mazzeo[187] optimized anion analysis using reversed polarity with 5 mM sodium chromate–3.5 mM tetradecyltrimethyl ammonium hydroxide (TTAOH)–10 mM 2-(*N*-cyclohexylamino)ethanesulfonic acid (CHES) (pH 9.0) at 254 nm. Cousins et al.[189] compared the effectiveness of several carrier electrolytes ranging from chromate to benzoate for the separation of small anions. Inorganic anions were also separated in nonaqueous media containing methanol and dimethylformamide.[190] While anions are better separated with reversed polarity, cations can be run using direct polarity, with a complexing agent to improve separation; chemicals such as ethylenediaminetetraacetic acid (EDTA) interact with the cations and modify their mobility. Conditions for the indirect UV analysis of metal ions were examined by Chen and Cassidy.[191] They examined the separation of 26 metal ions under various parameters such as the UV-absorbing probe, the pH, the concentration of the complexing agents and the nature of the capillary column. Alkali and alkaline earth metals in mineral water at the 0.1 mM level were studied by Simunicava et al. using tartrate, 18-crown-6 and benzimidazole as a visualization co-ion.[192] A similar separation of metal cations was obtained by Morin et al.[193] using imidazole and benzylamine buffers. They also explored in detail complexing agents that allow the separation of metal cations with similar mobility; of the chemicals they tried, they were most satisfied with phenylethylamine, benzylamine, *p*-toluidine and 4-methylbenzylamine. The same group examined minor cations in matrix samples.[194] Indirect UV detection at 214 nm was used by Padarauskas et al.[195] to analyze simultaneously anions and cations. Each species is injected from a different end of the capillary and they migrate in opposite directions towards the detector; under these conditions three anions and five cations were separated in 4 min. Finally, conductivity and suppressed conductivity detection have been used. Avdalovic et al.[196] separated many inorganic anions at the parts per million level. The same group,[197] still using suppressed conductivity detection, separated anions from fluorides to bromide with a 2 mM sodium tetraborate electrolyte with 5 mM sulfuric acid as regenerant. Using conductivity detection, Jones et al.[198] separated 37 anions ranging from bromide to gluconate in about 12 min with a buffer containing 50 mM CHES, 20 mM lithium hydroxide and 0.03% Triton™ X-100 (pH 9.2).

Many reports on the use of CE to analyze inorganic ions in food products have been published. Indirect photometric detection with an imidazole–sulfuric acid buffer was also used by Yang et al.[199] to measure K, Na, Ca and Mg in apple and orange juice A method has been developed[200] to quantify K, Na, Ca, Mg and Mn in solid food: following digestion with HNO_3–H_2O_2,

the minerals were separated with 500 mM imidazole, 130.6 mM 2-hydroxyisobutyric acid and 50 mM 18-crown-6–20% methanol (pH 4.5). The method was applied to the determination of the composition of various teas.

Baechmann et al.[201] used indirect fluorimetric detection to examine alkaline earth metal ions in beverages, such as cola. Low MW ions in musts and wines were examined by Dedieu et al.[202] Several groups have analyzed nitrites and nitrates in food products. Stransky et al.[203] examined nitrates in milk by ITP. The concentrations of nitrate and nitrite in vegetables were measured with indirect UV detection.[204] Nitrites and nitrates were also examined in a large variety of foods ranging from cured meat to cheese and fruit juice with 0.5 mM OFM Anion-BT–1000 ppm chloride at 210 nm.[205] The same reagent diluted in 5 mM chromate was used by Trenerry to measure the sulfite level of many vegetables, processed foods and sea foods.[206] Calcium and magnesium in wheat flour were examined by Kajiwara et al.[207] using CZE. Mineral water anions and cations were analyzed simultaneously by Kobayashi et al.[208] with a borate buffer (pH 9) with 0.5 mM tetradecyltrimethylammonium bromide (TTAB). Anionic impurities in glycerol were determined by capillary ITP[209] while inorganic anions and cations were examined at the single-cell level by Bazzanella et al.[210] Chloride, sulfate and nitrates were analyzed by Swallow and Low[211] at a level as low as 0.2 ppm using indirect UV detection. However, the authors were unable to use this technique to detect the addition of pulpwash (an orange juice by-product) to orange juice.

In CE, organic acids can be often separated with anions or at least can be followed by similar methods. Ion CE was described in detail by Jones[186] and Stover reviewed the CE analysis of organic acids and ions.[212] Like inorganic ions, organic acids absorb only in the very low-UV region and detection had to be done by indirect methods and more recently by conductivity detection. The formation of organic acids (lactic, acetic, phosphoric, citric, propionic, butyric) during lactic fermentation of various vegetables was examined by ITP and detected by a conductivity detector.[213] Buchberger et al.[214] studied fermenting acids in silage using indirect UV detection with sorbate or direct UV detection at 185 nm. Acetic, lactic and butyric acids were measured. ITP has been used by Kvasnicka et al.[215] to monitor metals, organic and inorganic acids and volatile fatty acids during sugar production. Lalljie et al.[216] analyzed organic acids in juice produced during sugar refining and in chicory root extract.

Hop bitter acids have been analyzed by various groups. Szucs et al.[217] were able to analyze directly iso-α-acids in beer by sample stacking of large injected volumes. They obtained similar separation by microemulsion electrokinetic chromatography.[218] Organic acids in beer were also examined by CZE.[219] Inorganic and organic anions, ranging from chloride to pyroglutamate, have been analyzed in beer and wort in less than 7 min using indirect UV detection at 350 nm and reversed polarity with a buffer containing 5 mM 2,6-pyridinedicarboxylic acid and 0.5 mM CTAB.[220] CE (phthalate buffer, pH 5.6) and HPLC were compared by Levi et al.[221] to study organic acids in wine, and they found that the two techniques gave similar results. In wine, molecules ranging from chloride to lactic acid were separated in 10 min.[222] Isocitric, citric, malic and tartric acids were separated from orange juice in 14 min with inverse polarity and a buffer containing phthalic acid, CTAB and 20% methanol at pH 7.[223] Lazzarelli et al.[224] used chromate and phthalate with a modified OFM Anion-BT buffer to examine organic acids (citric and malic) and inorganic acids in fruit juices and particularly orange juice. Citrus organic acids, namely citric, malic and isocitric acid, were examined with several methods.[225,226] Isocitric acid is present in orange juice in very small amounts compared with the other two acids. Furthermore, some of these acids are bound and must be hydrolyzed before analysis. Hydrolysis was performed with a procedure similar to that used for the enzymatic analysis of isocitric acid. A 5 mL volume of 4 M sodium hydroxide solution was added to 10 mL of juice to raise the pH to about 13.5. After 10 min, 5 mL of 4 M HCl were added to the solution to decrease the pH to 5.5. With indirect UV detection and monitoring at 205 nm, organic acids were first examined with direct polarity. Using 5 mM phthalate (pH 3.5) at 20 °C and 30 kV, under these conditions standards of these three acids were well separated but in the juice only citric and malic were separated.[225] Under different conditions (5 mM phthalate, pH 3.6, with 3 mM $CaCl_2$ at 15 °C and 30 kV), isocitric acid in juice could be seen, but the best separation with indirect UV detection was achieved with reversed polarity and a buffer containing 5 mM phthalate, 0.5 mM TTAB and 50 mM *N*-tris(hydroxymethyl)-2-aminoethanesulfonic acid (TES) (pH 8) at −15 kV and 20 °C. With this method, citric and malic acids were easily measured but isocitric acid was difficult to resolve and a separate run was necessary. Conductivity detection gave more precise and more reliable data and the three acids could be measured in a single run. The following conditions were used: 130 mM histidine, 130 mM MES, 0.7 mM TTAB, 0.03% Triton™ X-100, 10 ppm EDTA and 15% acetonitrile, pH 6, 21 °C and −18 kV. The results obtained by conductivity detection for the three organic acids were similar to those reported by three different laboratories using conventional enzymatic methods.[226]

3.3.5 Vitamins and Plant Chemicals

Since ascorbic acid is water soluble and absorbs very well in the UV region, it has been monitored by

many groups. Extraction with metaphosphoric acid and stabilization with dithiothreitol were used by Thompson and Trenerry[227] with MEKC and a sodium deoxycholate buffer to assess total L-ascorbic acid in fruits and vegetables. The dithiothreitol reduces the dehydroascorbic acid to ascorbic acid and allows the estimation of the total amount of vitamin C. The authors used a similar method to measure ascorbic acid in beers, wines and fruit beverages.[228] Cahill and Wightman[229] also measured ascorbate and catecholamine by CE. L- and D- ascorbic acid have been analyzed by Davey et al.[230] in parsley and mushroom after extraction with metaphosphoric acid with 200 mM borate buffer (pH 9) at 260 nm. The method has a limit of detection of 84 fmol per injection. On comparing HPLC and CE methods, the authors found CE to be more suitable for the analysis of low vitamin concentrations in certain complex matrices, such as mushroom. A complex ITP/CZE method was used by Prochazkova et al.[231] to measure trace amounts of ascorbic acid in human body fluids with a sensitivity of $0.09-0.15\,mg\,L^{-1}$. Vitamin C is routinely measured by CZE as part of a general analysis of citrus juices (see below); from the multiwavelength electropherogram the 280 nm trace is extracted and ascorbic acid measured. The results are similar to those obtained by HPLC.[124]

Huopalahti and Sunell[232] monitored simultaneously vitamins B_1, B_2 and B_6 by CZE. Up to nine water-soluble vitamins (vitamins C, B_1, B_3, B_6 and B_{12}, folinate, orotic acid, pantothenate and nicotinamide) were separated by Jegle[233] with a phosphate buffer (pH 7). L- and D- folinic acids were separated by Cellai et al.[234] using CD as a chiral active component. Choline was examined in foods and infant formula with a 5 mM 1-methylimidazole buffer (pH 4.5).[235] Niacin in yeast spread was extracted by autoclaving in calcium hydroxide and separated by CE with a 1:1 mixture of 0.02 M sodium tetraborate and 0.02 M disodium hydrogenorthophosphate (pH 9.2).[236] The results were similar to those obtained by the traditional colorimetric procedure. The authors[237] measured niacin in various foods with a buffer containing 0.02 M phosphate buffer (pH 7) with 15% acetonitrile. Hsieh and Kuo[238] separated retinoids by MEKC. Fourteen vitamins and vitamin cofactors were investigated by MEKC in foods such as orange juice with a buffer containing 100 mM phosphate, 500 mM taurine, 75 mM cholate and 2% propanol.[239] Pedersen-Bjergaard et al.[240] separated fat-soluble vitamins by hydrophobic interaction electrokinetic chromatography. Vitamins A, D and E were separated on an uncoated fused-silica capillary with a buffer of acetonitrile–water (80:20) and 80 mM tetradecylammonium bromide.

The CE analysis of plant chemicals has seen a major increase in the last few years and is one of the main areas of CE use for food products. A review of the CE analysis of natural products has been published by Issaq.[241] Many of these natural compounds absorb in the UV region and can be easily examined with most instruments. Sugarcane flavonoids have been examined by CZE with a borate buffer (pH 9.5) and 20% methanol.[242] Pietta's group reported several studies on the analysis of flavonoids. They used MEKC and diode-array ultraviolet detection (DAD/UV)[243,244] to separate flavonoids. They further[245] used MEKC with a 25 mM buffer (pH 8.25) with 50 mM SDS to analyze components from the root of *Eleutherococcus senticosus* and to detect adulteration in *Arnica montana* and *chamissonis*.[246]

Morin et al.[247] studied the complexation of borate and flavonoids. Shihabi et al. compared the HPLC and CE separation of isoflavones;[248] the HLPC method required a complex isocratic elution whereas the CE separation was done on a simple capillary with a borate buffer at pH 8.6 and took less than 8 min. A comparison between CE and HPLC for the analysis of honey flavonoids has been reported by Delgado et al.[249] and by Ferreres et al.,[250] who used MEKC, with a borate buffer (pH 8) containing 50 mM SDS and 10% methanol, to differentiate various types of honey by their flavonoid content. The influence of temperature on the CE separation of flavonols was studied by Fernandez de Simon et al. using a boric acid buffer (pH 9.6).[251]

Shepherd and McGhie[252] used CE with a tetraborate buffer to measure chlorogenic acid, apigenin, isoorientin and quercitin and the method was used to monitor the activity of polyphenol oxidase. Gil et al. developed an MEKC method to separate methylated flavones[253] and analyzed the phenolic compounds in red wine[254] with a borate buffer at pH 9.5 in a fused-silica column. From the results they were able to differentiate wines from different areas. Citrus seed limonoids (limonin and nomilin) were quantified by MEKC with a 12 mM tetraborate buffer containing 35 mM SDS.[255]

The composition of Chinese herbal tea including the flavonoid constituents was examined by Sheu and Chen[256] by HPLC and CE. Rhubarb aloe-emodin, emodin and rhein were measured by MEKC with a buffer containing 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS), SDS and acetonitrile at pH 10.96.[257] Ferulic and coumaric acids and esters and feruloyl glucosides from wheat bran were analyzed by Donaghy and McKay[258] using CZE. Flavonoids and cinnamic and phenolic acids were separated from blackcurrant bud extracts with a buffer containing 50 mM phosphate and 100 mM boric acid (pH 7) at 50 °C and 20 kV with detection at 232 nm.[259] Liang et al.[260] used an MEKC method to analyze the flavonoid content of four species of *Epimedium* with a buffer containing 30 mM borate and 12 mM SDS (pH 10.5) at 22 kV. The same group analyzed in detail the migration properties of flavonoids.[261,262]

Isoflavones from Lodoideae and Leguminosae were studied by CE/MS with an ammonium acetate buffer (pH 9.5).[263,264] Chrysanthemum coumarins were examined by CE with 0.2 M boric acid–0.05 M borax (11 : 9) (pH 8.5),[265] while Chen and Sheu used MEKC to separate coumarins.[266] Coumarin derivatives were followed by Bogan et al.[267] Fourteen plant phenolic compounds were separated by Jen et al.[268] with a borate buffer (pH 9) and detection at 210 nm; the DLs ranged from 20 to 100 $\mu g\,L^{-1}$. Tea polyphenols were separated by Larger et al.[269] with a buffer containing 20 mM SDS, 50 mM phosphate and 50 mM tetaborate with 10% acetonitrile (pH 6) and detection at 278 nm. Hiermann and Radl[270] separated aromatic acids from *Epilobium angustifolium* L. with a 20 mM phosphate buffer (pH 7.0) at 40 °C; peaks were detected at 280 nm. We routinely examine citrus juices for hesperidin, neohesperidin, narirutin and naringin to assess the purity of the products (see below).[123,124]

Anthocyanins are another group of plant chemicals that have received attention from CE analysts. In 1996, Bridle et al.[271] developed a method to detect anthocyanins at 580 nm, using a 150 mM sodium borate buffer (pH 8) at 25 °C. The same system was used by this group to examine anthocyanins in strawberries and elderberries[272] and the results were compared with those obtained by HPLC, which gave better separation for the latter fruit. The same chemicals were separated from blackcurrant with a strongly acidic phosphate buffer (pH 1.5) containing 30% acetonitrile.[273] Watanabe et al.[274] used CE to separate safflower pigments. The yellow water-soluble pigment could be analyzed with a 300 mM borate buffer (pH 9) and both yellow and red pigments could be separated by MEKC with 2% butyl acrylate–butyl methacrylate–methacrylic acid polymeric acid, and caffeine and its metabolites were analyzed by MEKC.[275]

3.3.6 Additives, Toxins, Drugs, Pesticides

The monitoring of food products requires the constant examination of many drugs, food additives and degradation products to indicate whether a product is fit for consumption. The analysis of food additives by CE has been reviewed by Schuster,[276] pollutants by Dabek–Zlotorzynska,[277] pharmaceutical drugs by Shihabi[278] and pesticides by El Rassi.[279] Alkaloids can be responsible for various food poisonings and have been followed by CE. Ergovaline in seeds was measured by Ma et al.[280] with a phosphate buffer (pH 3.5) containing 50% methanol. Liu and Sheu[281] used CE to compare alkaloid compositions in samples of *Ephedrae herba* and *Phellodendri cortex*. Alkaloids from various plants were analyzed by Unger et al.[282] using field amplified sample injection of a 70% methanol solution into a buffer of 200 mM ammonium acetate (pH 3.1)–methanol (1 : 1)[283] and further studied the influence of the alkaloid structure on mobility.[284] Vandepeer et al.[285] used MEKC buffers to examine alkaloids and vanillin benzoate in various drinks and confectionery. Ergot alkaloids have been separated with a phosphate buffer (pH 2.4) containing clyclodextrins, urea and poly(vinyl alcohol).[286] Strychnine and brucine from *Strychnos nux-vomica* were quantified with 10 mM phosphate buffer–methanol (9 : 1) at pH 2.4.[287] Yang and Smetena[288] separated nicotine and other alkaloids from tobacco with a 120 mM phosphate buffer (pH 2.7). The Chinese drugs tetrandrine and fangchinoline were examined by Yang et al.[289]

A CE immunofluorescence method was used by Blais et al.[290] to determine chloramphenicol in milk: bound and unbound forms of the chloramphenicol conjugate were titrated with LIF detection. Sorbic acid and benzoic acid in foods and beverages were separated by Pant and Trenerry[291] with a 0.02 M phosphate buffer (pH 9.2) and 0.05 M SDS by MEKC. We routinely monitor sorbate and benzoate in citrus juice with a direct CZE analysis using a borate buffer.[123,124] Dinelli et al.[292] compared CE, HPLC and immunoassay for the detection of the herbicide terbutylazine in tap water. CE gave the best sensitivity and the shortest retention time; however, CE and HPLC require prior extraction of the sample. The same group used MEKC[293,294] to quantify sulfonylurea herbicides at the parts per billion level. Nine herbicides were separated in 16 min by Farran and Ruiz[295] with an MEKC buffer. Wigfield et al.[296] analyzed paraquat and diquat in potatoes on a silinazed column; the method was shown to be superior to an HPLC method. Herbicides were also separated by Schmitt et al. by CZE.[297] Nemoto and Lehotay[298] used a 50 mM ammonium acetate buffer (pH 4.75) to follow pesticides extracted from soybean by pressurized liquid extraction. Alcantara et al.[299] analyzed the pesticides phloxine B and urarine in coffee by CZE with a borate buffer (pH 8.5) and UV and fluorescence detection. Results similar to those obtained by HPLC were reported.

Using 1-naphthol-3,6-disulfonic acid as a chromophore, Buscher et al.[300] monitored the enzymatic hydrolysis of inositol phosphates (such as phytic acid from grain) during fermentation. Phytic acid was analyzed in plant material with a fluorinated ethylene–propylene copolymer capillary associated with two conductivity cells.[301] Mycotoxin fumonisins were examined by Maragos[302] after derivatization with fluorescein isocyanate. The two isomers of the neurotoxin 2-(*N*-oxalyl)-L-2,3-diaminopropionic acid (ODAP) were measured in grass peas with a phosphate buffer (pH 7.8) with direct UV detection at 195 nm.[303] Sciacchitano and Hirshfield used CE to identify the *Clostridium botulinum* type E neurotoxin gene in smoked

fish[304] and fungal metabolites were examined by Nielsen et al. with MEKC.[305]

Toxic degradation products in food are monitored very closely and CE is now playing a role in these analyses. Mopper and Sciacchitano[306] measured histamine in fish with a citrate buffer (pH 2.5), and showed that the method was superior to the official AOAC fluorimetric method, particularly at low concentrations. The authors also reported distinct electropherograms for various fish species. Buzy et al.[307] used a buffer containing morpholine at pH 4 to estimate shellfish poisoning toxins. A 50 mM phosphate buffer (pH 2.44) was used by Gallardo et al.[308] to quantify histamine in fish and fish products and to identify fish species.[309] Zhao et al.[310] separated domoic acid in fish products with a buffer of 22.5 mM borate (pH 9.2) and 20 mM β-CD and Bouaicha et al.[311] used MEKC to monitor okadaic acid. Separation of nitrosamines was achieved with a CD-modified MEKC and a combined packed and open-tubular column.[312]

5-(hydroxymethyl)-2-furaldehyde (HMF) in jam and soft drinks and furosine in dry milk were examined by Corradini et al. using MEKC; a mean concentration of 4.6–74.2 mg of HMF was detected in soft drinks.[313] Furosine as an indicator of heat treatment in milk and cheese was measured by Tirelli and Pellegrino[314] with a 100 mM phosphate buffer (pH 7).

Indirect UV detection was used by Trevaskis and Trenerry[315] to determine oxalic acid, nitrate and molybdate in vegetables. Artificial sweeteners can easily be detected in the UV region and several analyses have been published. Liu and Li[316] compared CE and HPLC for the analysis of *Stevia* sweeteners. As seen previously, Sabah and Scriba[132] examined aspartame. MEKC with a sodium deoxycholate buffer was used by Thompson et al.[317,318] to measure many artificial sweeteners in soft drinks. Using the same technique, the authors[319] also examined synthetic colors in confectionery and cordials using an MEKC buffer. Food colorants were also

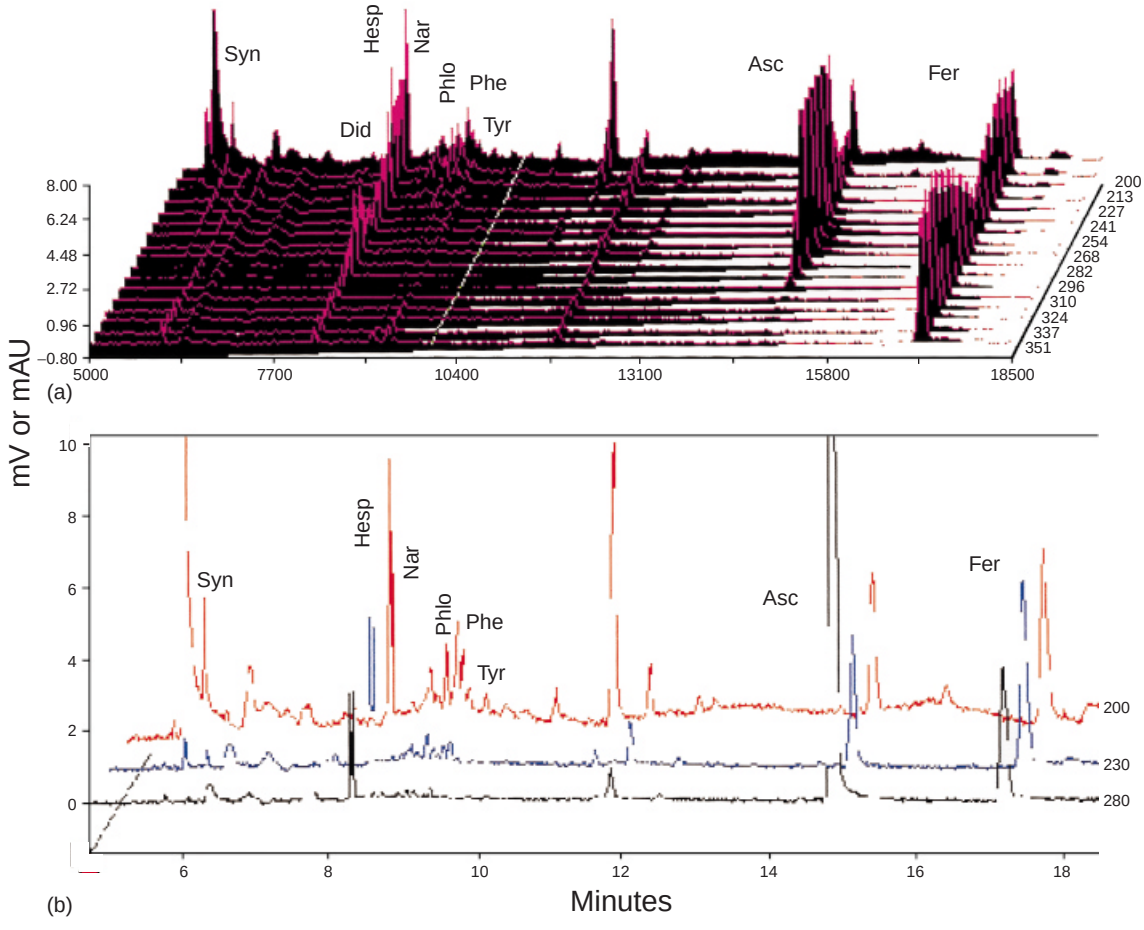

Figure 1 (a) Three-dimensional electropherogram of orange juice. The third dimension is produced by stacking electropherograms produced at many wavelengths; as a result the spectrum of each separated compound is produced. Syn = synephrine; Did = didymin; Hesp = hesperidin; Nar = narirutin; Phlo = phlorin; Phe = phenylalanine; Tyr = tyrosine; Asc = ascorbic acid, Fer = ferulic acid (internal standard). (b) 200, 230 and 280 nm traces extracted from the three-dimensional electropherogram.

examined by Razee et al.[320] using CD and by Liu et al.[321] Eleven phenolic antioxidants were separated by Abrantes et al.[322] by MEKC with a buffer containing 15 mM SDS, 50 mM phosphate and 25 mM borate (pH 7).

3.3.7 Food Monitoring

Several studies have shown that a single CE analysis can be used to separate and quantify chemically unrelated components of the same matrix. For example, the quality of green tea was monitored by Horie and Kohata[323] using a sodium borate buffer containing 50 nM SDS at pH 8.4. Detection was effected at 194 nm for all compounds except ascorcic acid, which was examined at 270 nm. From the data it was possible to estimate the quality of the teas. Various types of teas were also examined by Larger et al.,[269] who analyzed xanthine, catechin and flavonol by MEKC with a buffer containing 20 mM SDS, 50 mM phosphate, 50 mM sodium tetraborate and 10% acetonitrile (pH 6). Each type of tea has a specific profile that allows its identification. Brandy and oak extracts were analyzed by CE and HPLC.[324] The CE analysis was done with a 0.1 M borate buffer containing 5% ethanol (pH 9.2). It was concluded that CE and HPLC are complementary and the CE method could be used to assess the quality and authenticity of brandies.

In the past, the monitoring of citrus juices was done by a series of analyses using HPLC and spectrophotometric methods. A CE method to analyze simultaneously many citrus juice components in a single procedure has been reported.[123,124] Following filtration, the samples are analyzed, in 20 min, with an uncoated capillary using a 35 mM sodium borate buffer (pH 9.3) containing 5% (v/v) acetonitrile at 21 kV and 23 °C. The main compounds monitored on a regular basis are a biogenic amine (synephrine), some flavonoids (didymin, hesperidin, narirutin, neohesperidin and naringin), the polyphenol phlorin, three AAs absorbing in the UV region (Trp, Phe and Tyr), ascorbic acid, an unidentified peak generated by heat and storage and two preservatives (sorbate and benzoate) that can be added to citrus products (Figure 1). Each compound can subsequently be quantified. The didymin, narirutin and phlorin peaks are used with an artificial neural network to assess the volume of added pulp wash, a by-product of juice preparation. This last procedure is equivalent to the quantitation of a blend of two similar juices. This CE method allows the rapid monitoring of citrus juices, giving information on the quality, freshness and possible adulteration of the product. Similar procedures could be used to monitor other fruit juices and quantify diverse juice blends.[325]

4 CONCLUSION

Food analysis is largely dominated by HPLC methods and most electrophoretic techniques have seen only limited use in this area. However, studies published in the last few years have shown a large increase in the use of CE for the analysis of food components. Food examination by CE should continue to expand since many procedures developed for other biological materials could and probably will be applied to edible products. Several reasons can be found to explain the rise of this technique. Most CE separations can be performed in a few minutes with little sample preparation. As opposed to HPLC, CE requires small volumes of solvents and, contrary to other electrophoretic techniques, can be easily automated. It allows many compounds of different chemical nature to be separated in a single analysis, hence CE is most useful for analyzing rapidly large numbers of samples. Further developments should be mainly in the area of food quality monitoring.

ABBREVIATIONS AND ACRONYMS

AA	Amino Acid
AOAC	Association of Official Analytical Chemists
Brij™	23-Lauryl Ether
CAPS	3-(Cyclohexylamino)-1-propanesulfonic Acid
CD	Cyclodextrin
CE	Capillary Electrophoresis
CHES	2-(*N*-Cyclohexylamino)ethanesulfonic Acid
CTAB	Cetyltrimethylammonium Bromide
CZE	Capillary Zone Electrophoresis
DAD/UV	Diode-array Ultraviolet Detection
DL	Detection Limit
EDTA	Ethylenediaminetetraacetic Acid
EOF	Electroosmotic Flow
HMF	5-(Hydroxymethyl)-2-furaldehyde
HPLC	High-performance Liquid Chromatography
IEF	Isoelectric Focusing
ITP	Isotachophoresis
LIF	Laser-induced Fluorescence
MEKC	Micellar Electrokinetic Chromatography
MES	2-(*N*-Morpholino)ethanesulfonic Acid
MS	Mass Spectrometry
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
ODAP	2-(*N*-Oxalyl)-L-2,3-diaminopropionic Acid

PAD	Pulsed Amperometric Detection
PAGE	Polyacrylamide Gel Electrophoresis
Phe	Phenylalanine
pI	Isoelectric Point
SDS	Sodium Dodecyl Sulfate
TES	*N*-Tris(hydroxymethyl)-2-aminoethanesulfonic Acid
Tris	Tris(hydroxymethyl)aminomethane
Trp	Tryptophan
TTAB	Tetradecyltrimethylammonium Bromide
TTAOH	Tetradecyltrimethyl Ammonium Hydroxide
Tyr	Tyrosine
UV	Ultraviolet
2-D	Two-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors • High-performance Liquid Chromatography of Biological Macromolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Proteoglycan and Acidic Polysaccharide Analysis

Environment: Water and Waste **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Proteins, Peptides, and Amino Acids Analysis in Food • Vitamins Analysis in Food • Water Determination in Food

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Particle Size Analysis **(Volume 6)**
Sieving in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Gel Electrophoresis in Protein and Peptide Analysis

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction

Pharmaceuticals and Drugs **(Volume 8)**
Combinatorial Chemistry Libraries, Analysis of

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography • Micellar Electrokinetic Chromatography

General Articles **(Volume 15)**
Archaeological Chemical Analysis

REFERENCES

1. 'Food Composition; Additives; Natural Contaminants', in *Official Methods of Analysis of AOAC International*, ed. P. Cunniff, 16th edition, AOAC International, Arlington, VA, Vol. II, 1995.
2. S. Chang, E. Holm, J. Schwarz, P. Rayas-Durate, *Food. Anal. Chem.*, **67**, R127–R153 (1995).
3. J.R.J. Pare, J.M.R. Belanger (eds.), *Instrumental Methods in Food Analysis*, Elsevier, Amsterdam, 1997.
4. G. Linden (ed.), *Analytical Techniques for Foods and Agricultural Products*, VCH Publishers, New York, 1996.
5. D.M. Gersten (ed.), *Gel Electrophoresis: Proteins Essential Techniques*, John Wiley & Sons, New York, 1996.
6. R. Westermeier (ed.), *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations*, 2nd edition, John Wiley & Sons, New York, 1997.
7. J.C. Autran, 'Electrophoresis', in *Analytical Techniques for Foods and Agricultural Products*, ed. G. Linden, VCH Publishers, New York, 121–148, 1996.
8. F.T.A. Chen, R.A. Evangelista, 'Protein Analysis by Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 173–197, 1996.
9. K. Weber, M. Osborn, 'The Reliability of Molecular Weight Determinations by Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis', *J. Biol. Chem.*, **244**, 4406–4412 (1968).
10. B.J. Davis, 'Disk Electrophoresis II. Method and Application to Human Serum Proteins', *Ann. N.Y. Acad. Sci.*, **121**, 404–427 (1964).
11. L. Ornstein, 'Disc Electrophoresis I. Background and Theory', *Ann. N.Y. Acad. Sci.*, **121**, 321–349 (1964).
12. P.H. O'Farrell, 'High Resolution Two Dimensional Electrophoresis of Proteins', *J. Biol. Chem.*, **250**, 4007–4021 (1975).
13. P. Cancalon, S.T. Brady, R.J. Lasek, 'Slow Transport in a Nerve with Embryonic Characteristics, the Olfactory Nerve', *Dev. Brain Res.*, **38**, 275–285 (1988).
14. U. Romling, T. Huer, B. Tummler, 'Bacterial Genome Analysis by Pulsed Field Gel Electrophoresis Techniques', in *Advances in Electrophoresis*, eds.

A. Chrambach, M.J. Dunn, B.J. Radola, John Wiley & Sons, New York, 355–408, 1994.
15. J.C. Autran, 'Recent Perspectives on the Genetics, Biochemistry and Functionality of Wheat Proteins. Trends', *Food Sci. Technol.*, **4**, 358–362 (1993).
16. V. Carnevillier, M. Feuillat, C. Charpentier, 'L'Electrophorese, Principle et Applications en Oenologie', *Rev. Oenolo.*, **81**, 35–38 (1997).
17. T.-S. Huang, M.R. Marshall, C.-I. Wei, 'Identification of Red Snapper (*Lutjanus campechanus*) Using Electrophoretic Techniques', *J. Food Sci.*, **60**, 279–283 (1995).
18. Y.-H. Hsieh, F.-C. Chen, M. Nur, 'Rapid Species Identification of Cooked Red Snapper Using Isoelectric Focusing', *J. Food Sci.*, **62**, 15–19 (1997).
19. D.A. Dougherty, R.L. Wehling, M.G. Zeece, J.E. Partridge, 'Evaluation of Selected Baking Quality Factors of Hard Red Winter Wheat Flours by Two Dimensional Electrophoresis', *Cereal Chem.*, **67**, 564 (1990).
20. G. Yang, R.L. Wehling, M.G. Zeece, J.E. Partridge, D.R. Shelton, 'Characterization of Hard Red Winter Wheat Storage Proteins by Two-dimensional Electrophoresis and Their Correlations with Selected Quality Parameters', *Cereal Chem.*, **72**, 568–570 (1995).
21. P. Picard, M. Bourgoin-Greneche, M. Zivy, 'Potential of Two Dimensional Electrophoresis in the Identification of Closely Related Durum Wheat Lines', *Electrophoresis*, **18**, 174–181 (1997).
22. J. Drzewiecki, 'Discontinuous PAGE of Zeins for Purity Assessment of Maize Hybrids, and Cluster Analysis of PAGE Data from Inbred Lines', *Plant Varieties Seeds*, **9**, 149–158 (1996).
23. P.M. Civello, A.R. Chaves, M.C. Anon, 'Changes in Protein Composition During Strawberry (*Fragaria times ananassa* Duch.) Fruit Ripening', *J. Food Biochem.*, **20**, 135–153 (1996).
24. M. Santoro, 'Fractionation and Characterization of Must and Wine Proteins', *Am. J. Enol. Viticult.*, **46**, 250–254 (1995).
25. H.-P. Kubis, G. Gros, 'A Rapid Electrophoretic Method for Separating Rabbit Skeletal Muscle Myosin Heavy Chains at High Resolution', *Electrophoresis*, **18**, 64–66 (1997).
26. E. Huff-Lonergan, T. Mitsuhashi, F.C. Parrish, R.M. Robson, 'Sodium Dodecyl Sulfate–polyacrylamide Gel Electrophoresis and Western Blotting Comparisons of Purified Myofibrils and Whole Muscle Preparations for Evaluating Titin and Nebulin in Postmortem Bovine Muscle', *Anim. Sci.*, **74**, 770–785 (1996).
27. B.-Y. Chen, H.W. Janes, 'Multiple Forms of ADP-Glucose Pyrophosphorylase from Tomato Fruit', *Plant Physiol.*, **113**, 235–241 (1997).
28. B.J. King, L.S. Lee, S. Rackemann, P.T. Scott, 'Electrophoretic Conditions for High Resolution Citrus Isozymes in Polyacrylamide Gel Electrophoresis', *Electrophoresis*, **16**, 32–38 (1995).
29. S.D. Fouche, I.A. Dubery, 'Chalcone Isomerase from *Citrus sinensis*: Purification and Characterization', *Phytochemistry*, **37**, 127–132 (1994).
30. B.J. King, L.S. Lee, R.G. Rackeman, P.T. Scott, 'Electrophoretic Conditions for High Resolution Citrus Isozymes in Polyacrylamide Gel Electrophoresis', *Electrophoresis*, **16**, 32–38 (1995).
31. F. Addeo, L. Moio, L. Chianese, A. Di Luccia, 'Detection of Cow Milk in Milk or Goat Cheese by Isoelectric Focusing of Beta-lactoglobulins: Applications and Limitations', *Ital. J. Food Sci.*, **1**, 45–52 (1989).
32. H.C. Cutting, W.M. Kiessling, F.L. Bond, J.E. McCarron, K.S. Kreuzer, J.A. Hurbut, J.N. Safos, 'Agarose Gel Electrophoretic Detection of Six β-Lactam Antibiotic Residues in Milk', *J. AOAC Int.*, **78**, 663–667 (1995).
33. P.M. Izquierdo-Cana, J.F. Ubeda-Iranzo, A.I. Briones-Perez, 'Study of the Karyotype of Wine Yeasts Isolated in the Region of Valdepenas in Two Consecutive Vintages', *Food Microbiol.*, **14**, 221–225 (1997).
34. D. Perrett, 'Manufacturers and Suppliers of CE Equipment and Components', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 487–490, 1998.
35. P. Camilleri, 'History and Development of Capillary Electrophoresis', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 1–22, 1998.
36. P. Thibault, J. Dovichi, 'General Instrumentation and Detection Systems Including Mass Spectrometric Interfaces', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 23–89, 1998.
37. M. Zeece, 'Capillary Electrophoresis: A New Analytical Tool for Food Science', *Trends Food Sci. Technol.*, **3**, 6–10 (1992).
38. P.F. Cancalon, 'Food Analysis by Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 583–606, 1997.
39. J. Lindeberg, 'Capillary Electrophoresis in Food Analysis', *Food Chem.*, **55**, 73–94 (1996).
40. J. Lindeberg, 'Capillary Electrophoresis in Food Analysis', *Food Chem.*, **55**, 95–101 (1996).
41. S. Swedberg, 'Capillary Electrophoresis: Principles and Applications', in *Instrumental Methods in Food Analysis*, eds. J.R.J. Pare, J.M.R. Belanger, Elsevier, Amsterdam, 367–394, 1997.
42. D.R. Baker, *Capillary Electrophoresis*, John Wiley & Sons, New York, 1995.
43. P.G. Righetti, *Capillary Electrophoresis in Analytical Biotechnology*, CRC Press, Boca Raton, FL, 1995.
44. K.D. Altria (ed.), *Capillary Electrophoresis Guidebook, Principles, Operation, and Applications*, Methods in Molecular Biology, Vol. 52, Humana Press, Totowa, NJ, 1996.

45. H. Shintani, J. Polonski (eds.), *Handbook of Capillary Electrophoresis Applications*, Blackie Academic and Professional, London, 1997.
46. J.P. Landers (ed.), *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, FL, 1997.
47. M.G. Khaledi (ed.), *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, John Wiley & Sons, New York, 1998.
48. P. Camilleri (ed.), *Capillary Electrophoresis, Theory and Practice*, 2nd edition, CRC Press, Boca Raton, FL, 1998.
49. L. Krivankova, P. Bocek, 'Continuous Free-flow Electrophoresis', *Electrophoresis*, **19**, 1064–1074 (1998).
50. R.P. Oda, J.P. Landers, 'Introduction to Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 1–47, 1997.
51. A.M. Dougherty, N. Cooke, P. Shieh, 'Capillary Surface Modifications in Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 675–715, 1997.
52. C.S. Lee, 'Improved Capillary Electrophoresis Separations Associated with Controlling Electroosmotic Flow', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 717–739, 1997.
53. K.R. Nielsen, J.P. Foley, 'Micellar Electrokinetic Chromatography', in *Capillary Electrophoresis. Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 135–182, 1998.
54. J.R. Mazzeo, 'Micellar Electrokinetic Chromatography', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 49–73, 1997.
55. A. Guttman, 'Capillary Electrophoresis Separation of Enantiomers by Cyclodextrin Array Chiral Analysis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 75–100, 1997.
56. B. Chankvetadze, *Capillary Electrophoresis in Chiral Analysis*, John Wiley & Sons, New York, 1998.
57. L. Krivankova, P. Bocek, 'Capillary Isotachophoresis', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 251–276, 1998.
58. D.S. Burgi, R.L. Chien, 'On-line Sample Preconcentration for Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Pess, Boca Raton, FL, 479–493, 1997.
59. R. Rodriguez-Diaz, T. Wehr, M. Zhu, V. Levi, 'Capillary Isoelectric Focusing', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 101–138, 1997.
60. J.E. Wiktorowicz, 'Capillary Isoelectric Focusing', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 223–250, 1998.
61. P. Shieh, N. Cooke, A. Guttman, 'Capillary Gel Electrophoresis', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 185–221, 1998.
62. L.A. Colon, K.J. Reynolds, R. Alices-Maldonado, A.M. Fermier, 'Advances in Capillary Electro-chromatography', *Electrophoresis*, **18**, 2162–2174 (1997).
63. K.A. Kelly, M.G. Khaledi, 'Capillary Electro-chromatography', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 277–299, 1998.
64. R. Stevenson, K. Mistry, I.R. Krull, 'CEC'98: State-of-the-art', *Am. Lab.*, **30**(16), 16A–16L (1998).
65. M. Albin, P. Grossman, S. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**, 489A–497A (1993).
66. K. Kitagishi, 'Detection', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 84–103, 1997.
67. S.L. Pentoney, J.V. Sweedler, 'Optical Detection Techniques for Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 379–423, 1997.
68. L. Cruz, S.A. Shippy, J.V. Sweedler, 'Capillary Electrophoresis Detectors Based on Light', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 303–353, 1998.
69. H. Poppe, X. Xu, 'Indirect Detection in Capillary Electrophoresis', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 375–404, 1998.
70. M. Saito, Y. Kurosu, 'Recent Advances in Fluorescence Technology', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 104–117, 1997.
71. C.E. MacTaylor, A.C. Ewing, 'Critical Review of Recent Developments in Fluorescence Detection for Capillary Electrophoresis', *Electrophoresis*, **19**, 1234 (1998).
72. P.D. Voegel, R.P. Baldwin, 'Electrochemical Detection in Capillary Electrophoresis', *Electrophoresis*, **18**, 2267–2278 (1997).
73. C. Haber, 'Electrochemical Detection in Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 425–447, 1997.
74. B. Rhoden Bryant, F.D. Swanek, A.G. Ewing, 'Electrochemical Detection in High-performance Capillary Electrophoresis', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 355–374, 1998.

75. H.A. Bardelmeijer, J.C.M. Waterval, H. Lingeman, R. Van't Hof, A. Bult, W.J.M. Underberg, 'Pre-, On- and Post-column Derivatization in Capillary Electrophoresis', *Electrophoresis*, **18**, 2214–2227 (1997).
76. H.A. Bardelmeijer, H. Lingeman, C. deRuiter, W.J.M. Underberg, 'Derivatization in Capillary Electrophoresis', *J. Chromatogr. A*, **807**, 3–26 (1998).
77. J.F. Banks, 'Recent Advances in Capillary Electrophoresis/Electrospray/Mass Spectrometry', *Electrophoresis*, **18**, 2255–2266 (1997).
78. R.A. Evangelista, F.-T.A. Chen, 'Capillary Electrophoresis-based Immunoassays', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 219–239, 1996.
79. J.E. Wiktorowcz, 'Affinity Analysis by Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 270–282, 1997.
80. N.M. Schultz, L. Tao, D.J. Rose, R.T. Kennedy, 'Immunoassays and Enzyme Assays Using Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 379–423, 1997.
81. I.J. Colton, J.D. Carbeck, J. Rao, G.M. Whitesides, 'Affinity Capillary Electrophoresis: A Physical-organic Tool for Studying Interactions in Biomolecular Recognition – Review', *Electrophoresis*, **19**, 367–382 (1998).
82. D. Perrett, 'Capillary Electrophoresis in Biomedical and Pharmaceutical Research', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 441–485, 1998.
83. Z. El Rassi, 'Recent Developments in Capillary Electrophoresis of Carbohydrate Species', *Electrophoresis*, **18**, 2400–2408 (1997).
84. M.V. Novotny, 'Capillary Electrophoresis of Carbohydrates', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 729–765, 1997.
85. J.D. Olechno, J.A. Nolan, 'Carbohydrate Analysis by Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 297–345, 1997.
86. Z. El Rassi, Y.S. Mechref, 'Analysis of Carbohydrates', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 273–362, 1998.
87. J. Grimshaw, 'Analysis of Glycoaminoglycans and Their Oligosaccharide Fragments by Capillary Electrophoresis', *Electrophoresis*, **18**, 2408–2414 (1997).
88. R. Linhardt, 'Capillary Electrophoresis of Oligosaccharides', *Methods Enzymol.*, **230**, 265–280 (1994).
89. A. Bazzanella, K. Bachmann, 'Separation and Direct UV Detection of Sugars by Capillary Electrophoresis Using Chelation of Copper(II)', *J. Chromatogr. A*, **799**, 283–288 (1998).
90. H.-J. Zhong, M.A.K. Williams, R.D. Keenan, D.M. Goodall, C. Rolin, 'Separation and Quantitation of Pectins Using Capillary Electrophoresis: A Preliminary Study', *Carbohydr. Polym.*, **32**, 27–32 (1997).
91. H.J. Zhong, M.A.K. Williams, D.M. Goodall, M.E. Hansen, 'Capillary Electrophoresis Studies of Pectins', *Carbohydr. Res.*, **308**, 1–8 (1998).
92. P.F. Cancalon, Z. Li, 'CE Analysis of a Heat Generated Compound and of Pectins in Citrus', in *Annual Citrus Processors Meeting*, Florida Department of Citrus, Lake Alfred, FL, Vol. 48, 8–9, 1997.
93. A. Arentoft, S. Michaelsen, H. Sorensen, 'Determination of Oligosaccharides by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **652**, 517–524 (1993).
94. J. Plocek, J. Chmelik, 'Separation of Disaccharides as Their Borate Complexes by Capillary Electrophoresis with Indirect Detection in Visible Range', *Electrophoresis*, **18**, 1148–1152 (1997).
95. J. Brewster, M. Fishman, 'Capillary Electrophoresis of Plant Starches as the Iodine Complexes', *J. Chromatogr. A*, **693**, 382–387 (1995).
96. X. Xu, W.T. Kok, H. Poppe, 'Sensitive Determination of Sugars by Capillary Zone Electrophoresis with Indirect UV Detection Under Highly Alkaline Conditions', *J. Chromatogr. A*, **716**, 231–240 (1995).
97. A. Klockow, A. Paulus, V. Figueiredo, R. Amado, H. Widmer, 'Determination of Carbohydrates in Fruit Juices by Capillary Electrophoresis and High Performance Liquid Chromatography', *J. Chromatogr. A*, **680**, 187–200 (1994).
98. A. Klockow-Beck, *Diss. Abstr. Int.*, **57**, 576 (1996).
99. A. Klockow, A. Paulus, 'Sugar Analysis of Fruit Juices by Capillary Electrophoresis', *Am. Lab.*, **28**(14), 28C–28I (1996).
100. P.F. Cancalon, 'CE Analysis of Organic Acids and Sugars in Citrus Products by Indirect UV Detection', presented at the High Performance Capillary Electrophoresis Meeting, 1996.
101. P.F. Cancalon, 'Changes in the Saccharide Composition of Citrus Juice and Anatomical Fractions During Fruit Maturation', *Proc. Fl. State Hort. Soc.*, **107**, 253–256 (1994).
102. A. Zemann, D.T. Nguyen, G. Bonn, 'Fast Separation of Derivatized Carbohydrates by Coelectroosmotic Capillary Electrophoresis', *Electrophoresis*, **18**, 1142–1147 (1997).
103. B. Lu, D. Westerlund, 'Indirect UV Detection of Carbohydrates in Capillary Zone Electrophoresis by Using Tryptophan as a Marker', *Electrophoresis*, **17**, 325–332 (1996).
104. Y. Mechref, Z. El Rassi, 'Capillary Zone Electrophoresis of Derivatized Acidic Monosaccharides', *Electrophoresis*, **15**, 627–634 (1994).
105. Y. Mechref, G. Ostrander, Z. El Rassi, 'Capillary Electrophoresis of Carboxylated Carbohydrates, 1. Selective Precolumn Derivatization of Gangliosides with UV

Absorbing and Fluorescent Tags', *J. Chromatogr. A*, **695**(1), 83–95 (1995).
106. A.J. Mort, E.M.W. Chen, 'Separation of 8-Aminonaphthalene-1,3,6-trisulfonate (ANTS)-labeled Oligomers Containing Galacturonic Acid by Capillary Electrophoresis: Application to Determining the Substrate Specificity of Endopolygalacturonases', *Electrophoresis*, **17**, 379–383 (1996).
107. A. Guttman, S. Brunet, N. Cooke, 'Capillary Electrophoresis of Carbohydrates in the Biopharmaceutical and Food and Beverage Industries', *LC-GC*, **14**, 788–792 (1996).
108. A. Rydlund, O. Dahlmann, 'Rapid Analysis of Unsaturated Acidic Xylooligosaccharides from Kraft Pulps Using Capillary Zone Electrophoresis', *J. High Resolut. Chromatogr.*, **20**, 72–76 (1997).
109. A. Rydlund, O. Dahlmann, 'Oligosaccharides Obtained by Enzymatic Hydrolysis of Birch Kraft Pulp Xylan: Analysis by Capillary Zone Electrophoresis and Mass Spectrometry', *Carbohydr. Res.*, **300**, 95–102 (1997).
110. S. Motellier, C. Richet, P. Merel, 'Analysis of Cellulose Degradation Products by Capillary Electrophoresis', *J. Chromatogr. A*, **804**, 363–370 (1998).
111. L. Colon, R. Dadoo, R. Zare, 'Determination of Carbohydrates by Capillary Zone Electrophoresis with Amperometric Detection at a Copper Microelectrode', *Anal. Chem.*, **65**, 476–481 (1993).
112. J. Ye, R. Baldwin, 'Determination of Carbohydrates, Sugar Acids and Aldihols by Capillary Electrophoresis and Electrochemical Detection at the Copper Electrode', *J. Chromatogr.*, **687**, 141–148 (1994).
113. P.L. Weber, S.M. Lunte, 'Capillary Electrophoresis with Pulsed Amperometric Detection of Carbohydrates and Glycopeptides', *Electrophoresis*, **17**, 302–309 (1996).
114. W.H. Zhou, R.P. Baldwin, 'Capillary Electrophoresis and Electrochemical Detection of Underivatized Oligo- and Polysaccharides with Surfactant-controlled Electroosmotic Flow', *Electrophoresis*, **17**, 319–324 (1996).
115. X.J. Huang, W.T. Kok, 'Determination of Sugars by Capillary Electrophoresis with Electrochemical Detection Using Cuprous Oxide Modified Electrodes', *J. Chromatogr. A*, **707**, 335–342 (1995).
116. M. Goto, K. Tanaka, Y. Esaka, B. Uno, 'Development of a Rapid and Sensitive Determination Method of Carbohydrates and Their Related Compounds by Capillary Electrophoresis with Amperometric Detection at a Copper Electrode', *Bunseki Kagaku*, **45**, 95–100 (1997).
117. Z. Jin, R. Chen, L.A. Colon, 'Determination of Glucose in Submicroliter Samples by CE/LIF Using Precolumn or On-column Enzymatic Reactions', *Anal. Chem.*, **69**, 1326–1331 (1997).
118. H. Issaq, K. Chan, 'Separation and Detection of Amino Acids and Their Enantiomers by Capillary Electrophoresis: A Review', *Electrophoresis*, **16**, 467–480 (1994).
119. J.T. Smith, 'Developments in Amino Acid Analysis Using Capillary Electrophoresis', *Electrophoresis*, **18**, 2377–2392 (1997).
120. F. Toldra, M. Aristoy, 'Availability of Essential Amino Acids in Dry-cured Ham', *Int. J. Food Sci. Nutr.*, **44**, 215–219 (1993).
121. M. Strickland, B.C. Weimer, J.R. Broadbent, 'Capillary Electrophoresis of Cheddar Cheese', *J. Chromatogr. A*, **731**, 305–313 (1996).
122. C.W. Klampfl, W. Buchberger, M. Turner, G.S. Fritz, 'Determination of Underivatized Amino Acids in Beverage Samples by Capillary Electrophoresis', *J. Chromatogr. A*, **804**, 349–355 (1998).
123. P. Cancalon, C. Bryan, 'Use of Capillary Electrophoresis for Monitoring Citrus Juice Composition', *J. Chromatogr. A*, **652**, 555–562 (1993).
124. P.F. Cancalon, 'Analytical Monitoring of Citrus Juices by Using Capillary Electrophoresis', *J. AOAC Int.*, **82**, 95–106 (1998).
125. A. Tomlinson, J. Landers, I. Lewis, S. Naylor, 'Buffer Conditions Affecting the Separation of Maillard Reaction Products by Capillary Electrophoresis', *J. Chromatogr. A*, **652**, 171–177 (1993).
126. A. Tomlinson, J. Mlotkiewicz, I. Lewis, 'Application of Capillary Electrophoresis to the Separation of Coloured Products of Maillard Reactions', *Food Chem.*, **49**, 219–223 (1994).
127. J.M. Ames, A. Arnoldi, L. Bates, M. Negroni, 'Analysis of the Methanol-extractable Nonvolatile Maillard Reaction Products of a Model Extrusion-cooked Cereal Product', *J. Agric. Food Chem.*, **45**, 1256–1263 (1997).
128. D. Corradini, C. Corradini, 'Separation and Determination of 5-Hydroxymethyl-2-furaldehyde in Fruit Juice by Micellar Electrokinetic Capillary Chromatography with Direct Sample Injection', *J. Chromatogr. A*, **624**, 503–509 (1992).
129. E. Skocir, M. Prosek, M. Strazisar, T. Golob, A.M. Plestenjak, 'Quantitative Determination of Some Amino Acids in Maize Seed and Seaweed by Micellar Electrokinetic Chromatography', *J. Microcol. Sep.*, **9**, 451–457 (1997).
130. S. Shippy, J. Jankowsky, A. Timperman, J. Sweedler, 'Neuropeptide Determination Utilizing a Novel Laser Induced Fluorescence Scheme for CE', in *Pittsburgh Conference Abstracts*, 1060, 1995.
131. J.N. Ye, R.P. Baldwin, 'Determination of Amino Acids and Peptides by Capillary Electrophoresis and Electrochemical Detection at a Copper Electrode', *Anal. Chem.*, **66**, 2669–2674 (1994).
132. S. Sabah, G.K.E. Scriba, 'Determination of Aspartame and Its Degradation and Epimerization Products by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **16**, 1089–1096 (1998).
133. Y. Guo, L. Colon, R. Dadoo, R. Zare, 'Analysis of Underivatized Amino Acids by Capillary Electrophoresis

Using Constant Potential Amperometric Detection', *Electrophoresis*, **16**, 493–497 (1995).
134. B. Karger, Y. Chu, F. Foret, 'Capillary Electrophoresis of Proteins and Nucleic Acids', *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 579–610 (1995).
135. T. Wheat, 'Principles and Practice of Peptide Analysis with Capillary Zone Electrophoresis', in *Peptide Analysis Protocols*, eds. B.M. Dunn, M.W. Pennington, Humana Press, Totowa, NJ, 65–83, 1995.
136. J.T. Dolnik, 'Capillary Zone Electrophoresis of Proteins', *Electrophoresis*, **18**, 2353–2361 (1996).
137. F.E. Regnier, S. Lin, 'Capillary Electrophoresis of Proteins', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 683–728, 1998.
138. H.E. Schwartz, A. Guttman, A. Vinther, 'Separation of Peptides and Proteins', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 363–396, 1998.
139. I. Recio, L. Amigo, R. Lopez-Fandino, 'Assessment of the Quality of Dairy Products by Capillary Electrophoresis of Milk Products', *J. Chromatogr. B*, **697**, 231–242 (1997).
140. J. Otte, L. Midtgaard, K. Qvist, 'Analysis of Caseinomacropeptide(s) by Free Solution Capillary Electrophoresis', *Milchwissenschaft Milk Sci. Int.*, **50**, 75–79 (1995).
141. J. Van Riel, C. Olieman, 'Determination of Caseinomacropeptide with Capillary Zone Electrophoresis and Its Application to the Detection and Estimation of Rennet Whey Solids in Milk and Buttermilk Powder', *Electrophoresis*, **16**, 529–533 (1995).
142. N. Adamson, E. Reynolds, 'High Performance Capillary Electrophoresis of Casein Phosphopeptides Containing 2–5 Phosphoseryl Residues: Relationship Between Absolute Electrophoretic Mobility and Peptide Charge and Size', *Electrophoresis*, **16**, 525–528 (1995).
143. G. Paterson, J. Hill, D. Otter, 'Separation of Betalactoglobulin A, B and C Variants of Bovine Whey Using Capillary Zone Electrophoresis', *J. Chromatogr. A*, **700**, 105–110 (1995).
144. N. Kinghorn, C. Norris, G. Paterson, D. Otter, 'Comparison of Capillary Electrophoresis with Traditional Methods to Analyse Bovine Whey Proteins', *J. Chromatogr. A*, **700**, 111–123 (1995).
145. T.M.P. Cattaneo, F. Nigro, P.M. Toppino, V. Denti, 'Characterization of Ewe's Milk by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **721**, 345–349 (1996).
146. T.M.P. Cattaneo, F. Nigro, G.F. Greppi, 'Analysis of Cow, Goat and Ewe Milk Mixtures by Capillary Electrophoresis (CZE), Preliminary Approach', *Milchwissenschaft*, **51**, 616–619 (1996).
147. I. Recio, E. Molina, M. Ramos, M. Defrutos, 'Quantitative Analysis of Major Whey Proteins by Capillary Electrophoresis Using Uncoated Capillaries', *Electrophoresis*, **16**, 654–658 (1995).
148. I. Recio, C. Olieman, 'Determination of Denatured Serum Proteins in the Casein Fraction of Heat-treated Milk by Capillary Zone Electrophoresis', *Electrophoresis*, **17**, 1228–1233 (1996).
149. I. Recio, M. De Frutos, A. Olano, M. Ramos, 'Protein Changes in Stires Ultra-high-temperature-treated Milks, Studied by Capillary Electrophoresis and High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **44**, 3955–3959 (1996).
150. I. Reci, L. Amigo, M. Ramos, R. Lopez Fandino, 'Application of Capillary Electrophoresis to the Study of Proteolysis of Caseins', *J. Dairy Res.*, **64**, 221–230 (1997).
151. P.L.H. McSweeney, P.F. Fox, 'Chemical Methods for the Characterization of Proteolysis in Cheese During Ripening', *Lait*, **77**, 41–76 (1997).
152. W. Werner, 'Ferguson Plot Analysis of High Molecular Weight Glutenin Subunits by Capillary Electrophoresis', *Cereal Chem.*, **72**, 248–251 (1995).
153. J.A. Bietz, 'Fractionation of Wheat Gluten Proteins by Capillary Electrophoresis', in *Gluten Proteins, Proceedings of the 5th International Workshop*, Association for Cereal Research, Detmold, 404–413, 1994.
154. K.H. Sutton, J.A. Bietz, 'Variation Among High Molecular Weight Subunits of Glutenin Detected by Capillary Electrophoresis', *J. Cereal Sci.*, **25**, 9–16 (1997).
155. J.A. Biets, G.L. Lookhart, 'Wheat Varietal Identification by Capillary Electrophoresis: An Inter-laboratory Comparison of Methods', *Food Sci. Technol., Lebensm. Wiss. Technol.*, **30**, 210–213 (1997).
156. G. Lookhart, S. Bean, 'Rapid Differentiation of Oat Cultivars and of Rice Cultivars by Capillary Zone Electrophoresis', *Cereal Chem.*, **72**, 312–316 (1995).
157. G.L. Lookhart, S.R. Bean, 'Improvements in Cereal Protein Separations by Capillary Electrophoresis: Resolution and Reproducibility', *Cereal Chem.*, **73**, 81–87 (1996).
158. L. Capelli, F. Forlani, F. Perini, N. Guerrieri, N. Cerletti, P.G. Righetti, 'Wheat Cultivar Discrimination by Capillary Electrophoresis of Gliadins in Isoelectric Buffers', *Electrophoresis*, **19**, 311–318 (1998).
159. E.S.M. AbdelAal, D.A. Salama, P. Hucl, F.W. Sosulski, W. Cao, 'Electrophoretic Characterization of Spring Spelt Wheat Gliadins', *J. Agric. Food Chem.*, **44**, 2117–2123 (1996).
160. C. Lucchese, G. Dinelli, A. Bonetti, A. Lovato, 'Specific and Varietal Differentiation in *Vicia* L. and *Vicia villosa* by Means of IEF/PAGE and Capillary Electrophoresis', *Seed Sci. Technol.*, **25**, 253–262 (1997).
161. S.M. Basha, 'Separation of Peanut Proteins by Capillary Electrophoresis', *J. Agric. Food. Chem.*, **45**, 400–402 (1997).
162. F.-T. Chen, A. Tusak, 'Characterization of Food Proteins by Capillary Electrophoresis', *J. Chromatogr. A*, **685**, 331–337 (1994).
163. M. Strege, A. Lagu, 'Capillary Electrophoretic Separations of Biotechnology-derived Proteins in *E. coli*

Fermentation Broth', *Electrophoresis*, **16**, 642–646 (1995).
164. C. Luguera, V. Monero-Arribas, E. Pueyo, C. Polo, 'Capillary Electrophoretic Analysis of Wine Proteins. Modifications During the Manufacture of Sparkling Wines', *J. Agric. Food Chem.*, **45**, 3766–3770 (1997).
165. B. Deng, J. Lu, J. Yan, 'Separation of Alcohol-soluble Proteins from Tobacco Seeds by Capillary Electrophoresis', *J. Chromatogr. A*, **803**, 321–327 (1998).
166. H. Kajiwara, A.M. Hemmings, 'Capillary Electrophoretic Analysis of Ginseng Polypeptide', *Electrophoresis*, **19**, 1270–1274 (1998).
167. E. Leblanc, S. Singh, R. Leblanc, 'Capillary Zone Electrophoresis of Fish Muscle Sarcoplasmic Proteins', *J. Food Sci.*, **59**, 1267–1270 (1994).
168. J. Gallardo, C. Sotelo, C. Perez-Martin, 'Use of Capillary Zone Electrophoresis for Fish Species Identification. Differentiation of Flatfish Species', *J. Agric. Food Chem.*, **3**, 1238–1244 (1995).
169. J.L. Miller, M.G. Khaledi. 'Nonaqueous Capillary Electrophoresis', in: *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 525–555, 1998.
170. L. Blomberg, P. Andersson, 'Capillary Electrophoresis for Lipid Analysis', *Inform*, **5**, 1030–1037 (1994).
171. Y. Shi, J. Fritz, 'Capillary Zone Electrophoresis of Neutral Organic Molecules in Organic – Aqueous Solution', *J. High Resolut. Chromatogr.*, **17**, 713–718 (1994).
172. L. Ingvardsen, S. Michaelsen, H. Sorensen, 'Analysis of Individual Phospholipids by High-performance Capillary Electrophoresis', *J. Am. Oil Chem. Soc.*, **71**, 183 (1994).
173. W. Buchberger, K. Winna, 'Determination of Free Fatty Acids by Capillary Zone Electrophoresis', *Mikrochim. Acta*, **122**, 5–52 (1996).
174. E. Drange, E. Lundanes, 'Determination of Long-chained Fatty Acids Using Non-aqueous Capillary Electrophoresis and Indirect UV Detection', *J. Chromatogr. A*, **771**, 301–309 (1997).
175. R. Neubert, K. Raith, J. Schiewe, 'Capillary Zone Electrophoresis in Skin Fatty Acid Analysis', *Pharmazie*, **52**, 212–215 (1997).
176. K. Raith, R. Wolf, J. Wagner, R. Neubert, 'Separation of Phospholipids by Nonaqueous Capillary Electrophoresis with Electrospray Ionisation Mass Spectrometry', *J. Chromatogr. A*, **802**, 185–188 (1998).
177. J.P. Kleinhenz, W.J. Harper, 'Comparison of Direct Capillary Electrophoresis and Capillary Gas Chromatographic Methods for the Analysis of Lower Molecular Weight Fatty Acids in Swiss Cheese', *Milchwissenschaft Milk Sci. Int.*, **52**, 200–204 (1997).
178. O. Schmitz, S. Gab, 'Separation of Isomeric Hydroperoxides of Unsaturated Fatty Acids by Capillary Electrophoresis', *J. Chromatogr. A*, **767**, 249–253 (1997).
179. M.W. Hsiao, S.T. Lin, 'Separation of Steroids by Micellar Electrokinetic Capillary Chromatography with Sodium Cholate', *J. Chromatogr. Sci.*, **35**, 259–264 (1997).
180. K. Heinig, S. Hissner, S. Martin, C. Vogt, 'Separation of Saturated and Unsaturated Fatty Acids by Capillary Electrophoresis and HPLC', *Am. Lab.*, **30**, 24–29 (1998).
181. P.R. Haddad, 'Comparison of Ion Chromatography and Capillary Electrophoresis for the Determination of Inorganic Ions', *J. Chromatogr. A*, **770**, 281–290 (1997).
182. W. Buchberger, 'Inorganic ions', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 531–549, 1997.
183. M. Macka, P.R. Haddad, 'Determination of Metal Ions by Capillary Electrophoresis', *Electrophoresis*, **18**, 2482–2501 (1997).
184. F.B. Erim, 'Metal Chelation', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 517–530, 1997.
185. W. Buchberger, 'Capillary Electrophoresis of Metal Complexes', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 509–516, 1997.
186. W.R. Jones, 'Capillary Ion Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 156–212, 1997.
187. J.R. Mazzeo, 'Capillary Electrophoresis of Inorganic Ions', in *High-performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, John Wiley & Sons, New York, 825–852, 1998.
188. J.S. Fritz, 'Determination of Inorganic Anions and Metal Cations', in *Capillary Electrophoresis, Theory and Practice*, 2nd edition, ed. P. Camilleri, CRC Press, Boca Raton, FL, 249–272, 1998.
189. S. Cousins, P. Haddad, W. Buchberger, 'Evaluation of Carrier Electrolytes for Capillary Zone Electrophoresis of Low-molecular-mass Anions with Indirect UV Detection', *J. Chromatogr. A*, **671**, 397–402 (1994).
190. H. Salimi-Moosavi, R. Cassidy, 'Capillary Electrophoresis of Inorganic Anions in Nonaqueous Media with Electrochemical and Indirect UV Detection', *Anal. Chem.*, **67**, 1067–1073 (1995).
191. M. Chen, R. Cassidy, 'Separation of Metal Ions by Capillary Electrophoresis', *J. Chromatogr. A*, **640**, 425–431 (1993).
192. E. Simunicova, D. Kaniansky, K. Loksikova, 'Separation of Alkali and Alkaline Earth Metal and Ammonium Cations by Capillary Zone Electrophoresis with Indirect UV Absorbance Detection', *J. Chromatogr. A*, **665**, 203–209 (1994).
193. P. Morin, C. Francois, M. Dreux, 'Capillary Electrophoresis of Alkali and Alkaline Earth Cations with

Imidazole or Benzylamine Buffers', *J. Liq. Chromatogr.*, **17**, 3869–3888 (1994).
194. X. Cahours, P. Morin, M. Dreux, 'Quantitative Determination of Inorganic Minor Cations in Sodium-, Calcium-, Magnesium-matrix Simulated Samples by Capillary Electrophoresis', *J. Chromatogr. A*, **810**, 209–220 (1998).
195. A. Padarauskas, V. Olsauskaite, G. Schwedt, 'Simultaneous Separation of Inorganic Anions and Cations by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **800**, 369–375 (1998).
196. N. Avdalovic, C. Pohl, R.D. Rocklin, J. Stillian, 'Determination of Cations and Anions by Capillary Electrophoresis Combined with Suppressed Conductivity Detection', *Anal. Chem.*, **65**, 1470–1475 (1993).
197. M. Harrold, J. Stillian, L. Bao, R. Rocklin, N. Avdalovic, 'Capillary Electrophoresis of Inorganic Anions and Organic Acids Using Suppressed Conductivity Detection – Strategies for Selectivity Control', *J. Chromatogr. A*, **717**, 371–383 (1995).
198. W.R. Jones, J. Soglia, M. McGlynn, C. Haber, J. Reineck, C. Krstanovic, 'Capillary Ion Electrophoresis with Conductivity Detection', *Am. Lab.*, **28**, 25 (1996).
199. Q. Yang, M. Jimidar, T. Hamoir, J. Smeyers-Verbeke, D. Massart, 'Determination of Alkali and Alkaline Earth Metals in Real Samples by Capillary Ion Analysis', *J. Chromatogr. A*, **673**, 275–285 (1994).
200. Q. Yang, C. Hartmann, J. Smeyers-Verbeke, J.D.L. Massart, 'Method Development and Validation for the Determination of Mineral Elements in Food and Botanical Materials for Capillary Electrophoresis', *J. Chromatogr. A*, **717**, 415–425 (1995).
201. K. Baechmann, J. Boden, I. Haumann, 'Indirect Fluorimetric Detection of Alkali and Alkaline Earth Metal Ions in Capillary Zone Electrophoresis with Cerium(III) as Carrier Electrolyte', *J. Chromatogr.*, **626**(2), 259–265 (1992).
202. F. Dedieu, G. Nouadje, P. Puig, 'Contribution of Capillary Electrophoresis to the Analysis of Low Molecular Weight Ions in Musts and Wines', *Rev. Oenol. Technol. Vitivini. Oenolo.*, **72**, 7–10 (1994).
203. Z. Stransky, V. Dostal, Z. Cvak, V. Dulova,'Isotachophoretic Determination of Nitrates in Milk', *Acta Univ. Palacki. Olomuc., Fac. Rerum Nat.*, **102**, 121–128 (1991).
204. M. Jimidar, C. Hartmann, N. Cousement, D. Massart, 'Determination of Nitrate and Nitrite in Vegetables by Capillary Electrophoresis with Indirect Detection', *J. Chromatogr. A*, **706**, 479–492 (1995).
205. P.A. Marshall, V.C. Trenerry, 'The Determination of Nitrite and Nitrate in Foods by Capillary Ion Electrophoresis', *Food Chem.*, **57**, 339–345 (1996).
206. V.C. Trenerry, 'The Determination of Sulphite Content of Some Foods and Beverages by Capillary Electrophoresis', *Food Chem.*, **55**, 299–303 (1996).
207. H. Kajiwara, A. Sato, S. Kaneko, 'Analysis of Calcium and Magnesium Ions in Wheat Flour by Capillary Zone Electrophoresis', *Biosci. Biotechnol. Biochem.*, **57**, 1010–1011 (1993).
208. J. Kobayashi, M. Shirao, H. Nakazawa, 'Simultaneous Determination of Anions and Cations in Mineral Water by Capillary Electrophoresis with a Chelating Agent', *J. Liq. Chromatogr. Relat. Technol.*, **21**, 1445–1456 (1998).
209. T. Meissner, F. Eisenbeiss, B. Jastorff, 'Determination of Anionic Trace Impurities in Glycerol by Capillary Isotachophoresis with Enlarged Sample Load', *J. Chromatogr. A*, **810**, 201–208 (1998).
210. A. Bazzanella, H. Lochmann, A.D. Tomos, K. Bachmann, 'Determination of Inorganic Cations and Anions in Single Plant Cells by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **809**, 231–239 (1998).
211. K. Swallow, N. Low, 'Capillary Electrophoretic Analysis of Minor Anions Present in Orange Juice and Orange Pulpwash', *J. Agric. Food Chem.*, **42**, 2808–2811 (1195).
212. F.S. Stover, 'Organic Acids and Organic Ions', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 550–579, 1997.
213. J. Karovicova, J. Polonsky, M. Drdak, P. Simko, V. Vollek, 'Capillary Isotachophoresis of Organic Acids Produced by Selected Microorganisms During Lactic Acid Fermentation', *J. Chromatogr. A*, **638**, 241–246 (1993).
214. W. Buchberger, C.W. Klampfl, F. Eibensteiner, K. Buchgraber, 'Determination of Fermenting Acids in Silage by Capillary Electrophoresis', *J. Chromatogr. A*, **766**, 197–203 (1997).
215. F. Kvasnicka, G. Parking, C. Harvey, 'Capillary Isotachophoresis as a New Tool in Sugar Factory Analysis', *Int. Sugar J.*, **95**, 451–458 (1993).
216. S. Lalljie, J. Vindevogel, P. Sandra, 'Quantitation of Organic Acids in Sugar Refinery Juices with Capillary Zone Electrophoresis and Indirect UV Detection', *J. Chromatogr. A*, **652**, 563–569 (1993).
217. R. Szucs, J. Vindevogel, P. Sandra, L. Verhagen, 'Sample Stacking Effects and Large Injection Volumes in Micellar Electrokinetic Chromatography of Ionic Compounds: Direct Determination of Iso-acids in Beer', *Chromatographia*, **36**, 323–329 (1993).
218. R. Szucs, J. Vindevogel, E. Everaert, L. Decooman, P. Sandra, D. Dekeukeleire, 'Separation and Quantification of All Main Hop Acids in Different Hop Cultivars by Microemulsion Electrokinetic Chromatography', *J. Int. Brew.*, **100**, 293–296 (1994).
219. K. de Vries, 'Determination of Organic Acids in Beer by Capillary Electrophoresis', *J. Am. Soc. Brew. Chem.*, **51**, 155–157 (1993).
220. T. Soga, M. Wakaura, 'Determination of Inorganic and Organic Anions in Beer and Wort by Capillary Electrophoresis', *J. Am. Brew. Chem.*, **55**, 44–46 (1997).

221. V. Levi, T. Wehr, K. Talmadge, M. Zhu, 'Analysis of Organic Acids in Wines by Capillary Electrophoresis and HPLC', *Am. Lab.*, January, 29–31 (1993).
222. R. Cotter, M. Benvenuti, J. Krol, 'The Use of Capillary Ion Analysis in the Analysis of Organic Acids Impotable Liquids', in *Pittsburgh Conference Abstracts*, 1365, (1995).
223. N. Lopez Martinez, A. Rodriguez Roldan, 'Aplicacion de la Electrophoresis Capilar Micelar al Analisis Cuantitativo de Acidos Carboxilicos', *Alimentaria*, **243**, 71–74 (1993).
224. S. Lazzarelli, R. Bazzarini, A. Trifiro, G. Saccani, S. Gherardi, 'Elettroforesi Capillare per il Dosaggio di Ioni Organici ed Inorganici nei Succhi di Frutta', *Ind. Bev.*, **25**, 209–216 (1996).
225. P.F. Cancalon, 'Analysis of Organic Acids by Capillary Electrophoresis', presented at the 45th Annual Citrus Processors Meeting, 1994.
226. P.F. Cancalon, 'Analysis of Organic Acids in Citrus Juice by Capillary Electrophoresis', in *Pittsburgh Conference Abstract*, 1175, 1997.
227. C. Thompson, V. Trenerry, 'A Rapid Method for the Determination of Total L-Ascorbic Acid in Fruits and Vegetables by Micellar Electrokinetic Capillary Chromatography', *Food Chem.*, **53**, 43–50 (1995).
228. P. Marshall, V. Trenerry, C. Thompson, 'The Determination of Total Ascorbic Acid in Beers, Wines, and Fruit Drinks by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. Sci.*, **33**, 426–432 (1995).
229. P. Cahill, R. Wightman, 'Simultaneous Amperometric Measurement of Ascorbate and Catecholamine Secretion from Individual Bovine Adrenal Medullary Cells', *Anal. Chem.*, **67**, 2599–2605 (1995).
230. M.W. Davey, G. Bauw, M. Van Montagu, 'Analysis of Ascorbate in Plant Tissues by High-performance Capillary Zone Electrophoresis', *Anal. Biochem.*, **239**, 8–19 (1996).
231. A. Prochazkova, L. Krivankova, P. Bocek, 'Quantitative Trace Analysis of L-Ascorbic Acid in Human Body Fluids by On-line Combination of Capillary Isotachophoresis and Zone Electrophoresis', *Electrophoresis*, **19**, 300–304 (1998).
232. R. Huopalahti, J. Sunell, 'Use of Capillary Electrophoresis in the Determination of B Vitamins in Pharmaceutical Products', *J. Chromatogr.*, **636**, 133–135 (1993).
233. U. Jegle, 'Separation of Water-soluble Vitamins via High-performance Capillary Electrophoresis', *J. Chromatog. A*, **652**, 495–501 (1993).
234. L. Cellai, C. Desiderio, R. Filippetti, S. Fanali, 'Capillary Electrophoresis Quantitation of *l*-L-Folinic Acid in the Presence of Its Inactive *d*-L-Form', *Electrophoresis*, **14**, 823–825 (1993).
235. N. Carter, V.C. Trenerry, 'The Determination of Choline in Vitamin Preparations, Infant Formula and Selected Foods by Capillary Zone Electrophoresis with Indirect Ultraviolet Detection', *Electrophoresis*, **17**, 1622–1626 (1996).
236. C.M. Ward, V.C. Trenerry, I. Pant, 'The Application of Capillary Electrophoresis to the Determination of Total Niacin in Concentrated Yeast Spreads', *Food Chem.*, **58**, 185–192 (1997).
237. C.M. Ward, V.C. Trenerry, 'The Determination of Niacin in Cereals, Meat and Selected Foods by Capillary Electrophoresis and High Performance Liquid Chromatography', *Food Chem.*, **60**, 667–674 (1997).
238. Y.Z. Hsieh, K.L. Kuo, 'Separation of Retinoids by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **761**, 307–313 (1997).
239. S. Buskov, P. Moller, H. Sorensen, J.C. Sorensen, S. Sorensen, 'Determination of Vitamins in Food Based on Supercritical Fluid Extraction Prior to Micellar Electrokinetic Capillary Chromatographic Analyses of Individual Vitamins', *J. Chromatogr. A*, **802**, 233–241 (1998).
240. S. Pedersen-Bjergaard, K.E. Rasmussen, T. Tilander, 'Separation of Fat Soluble Vitamins by Hydrophobic Interaction Electrokinetic Chromatography with Tetradecylammonium Ions as Pseudostationary Phase', *J. Chromatogr. A*, **807**, 285–295 (1998).
241. H.J. Issaq, 'Capillary Electrophoresis of Natural Products', *Electrophoresis*, **18**, 2438–2452 (1997).
242. T. McGhie, 'Analysis of Sugarcane Flavonoids by Capillary Zone Electrophoresis', *J. Chromatogr.*, **634**, 107–112 (1993).
243. P. Pietta, P. Mauri, R. Facino, M. Carini, 'Analysis of Flavonoids by MECC with Ultraviolet Diode Array Detection', *J. Pharm. Biomed. Anal.*, **10**, 1041–1045 (1992).
244. P. Pietta, P. Mauri, L. Zini, C. Gardana, 'Optimization of Separation Selectivity in Capillary Electrophoresis of Flavonoids', *J. Chromatogr. A*, **680**, 175–179 (1994).
245. P. Pietta, P. Mauri, C. Gardana, L. Zini, 'Micellar Electrokinetic Chromatographic/Ultraviolet Diode Array Analysis of *Eleutherococcus senticosus*', *Phytochem. Anal.*, **5**, 305–308 (1994).
246. P. Pietta, P. Mauri, A. Bruno, I. Merfort, 'MEKC as an Improved Method to Detect Falsifications in the Flowers of *Arnica montana* and *A. chamissonis*', *Planta Med.*, **60**, 369–372 (1994).
247. P. Morin, F. Villard, M. Dreux, 'Borate Complexation of Flavonoid-*O*-glycosides in Capillary Electrophoresis II. Separation of Flavonoid-3-*O*-glycosides Differing in Their Sugar Moiety', *J. Chromatogr.*, **628**, 161–169 (1993).
248. Z. Shihabi, T. Kute, L. Garcia, M. Hinsdale, 'Analysis of Isoflavones by Capillary Electrophoresis', *J. Chromatogr. A*, **680**, 181–185 (1994).
249. C. Delgado, F. Tomas-Barderan, T. Talou, A. Gaset, 'Capillary Electrophoresis as an Alternative to HPLC for Determination of Honey Flavonoids', *Chromatographia*, **38**, 71–78 (1994).

250. F. Ferreres, M. Amparo-Blanquez, M. Gil, F. Tomas-Barberan, 'Separation of Honey Flavonoids by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **669**, 264–274 (1994).
251. B. Fernandez de Simon, I. Estrella, T. Hernandez, 'Flavonoid Separation by Capillary Electrophoresis. Effect of Temperature and pH', *Chromatographia*, **41**, 389–392 (1995).
252. K. Shepherd, T. McGhie, 'Novel Capillary Electrophoresis Technique for the Study of Plant Phenolic Enzymic Oxidation Mechanisms', *J. Agric. Food Chem.*, **43**, 657–661 (1995).
253. M.I. Gil, F. Ferreres, F. Tomas-Barberan, 'Micellar Electrokinetic Capillary Chromatography of Methylated Flavone Aglycones', *J. Liq. Chromatogr.*, **18**, 3007–3019 (1995).
254. M.I. Gil, C. Garcia-Viguera, P. Bridle, F. Tomas-Barberan, 'Analysis of Phenolic Compounds in Spanish Red Wines by Capillary Zone Electrophoresis', *Z. Lebensm.-Unters. Forsch.*, **200**, 278–281 (1995).
255. V.E. Moodley, D.A. Mulholland, M.W. Raynor, 'Micellar Electrokinetic Capillary Chromatography of Limonoid Glucosides from Citrus Seeds', *J. Chromatogr. A*, **768**, 143–152 (1995).
256. S. Sheu, H. Chen, 'Simultaneous Determination of Twelve Constituents of I-tzu-tang, a Chinese Herbal Preparation, by High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr. A*, **704**, 141–148 (1995).
257. Y. Zong, C. Che, 'Analysis of Rhubarb by Micellar Electrokinetic Capillary Chromatography', *J. Nat. Prod. Lloydia*, **58**, 577–582 (1995).
258. J. Donaghy, A. McKay, 'Measurement of Feruloyl/*p*-Coumaroyl Esterase by Capillary Zone Electrophoresis', *World J. Microbiol. Biotechnol.*, **11**, 160–162 (1995).
259. J.B. Fernandes, D.W. Griffiths, H. Bain, 'The Evaluation of Capillary Zone and Micellar Electrokinetic Capillary Chromatographic Techniques for the Simultaneous Determination of Flavonoids, Cinnamic and Phenolic Acids in Blackcurrant (*Ribes nigrum*) Bud Extracts', *Phytochem. Anal.*, **7**, 97–103 (1996).
260. H.R. Liang, H. Siren, P. Jyske, M.L. Reikkola, P. Vuorela, H. Vuorela, R. Hiltunen, 'Characterization of Flavonoids in Extracts from Four Species of *Epimedium* by Micellar Electrokinetic Capillary Chromatography with Diode-array Detection', *J. Chromatogr. Sci.*, **35**, 117–125 (1997).
261. H.R. Liang, H. Vuorela, P. Vuorela, R. Hiltunen, M.L. Riekkola, 'The Statistical Evaluation of Migration Parameters of Flavonoids in Capillary Electrophoresis with Reference to Structural Descriptors', *J. Liq. Chromatogr. Relat. Technol.*, **21**, 625–643 (1998).
262. H.R. Liang, H. Vuorela, P. Vuorela, M.L. Riekkola, R. Hiltunen, 'Prediction of Migration Behaviour of Flavonoids in Capillary Zone Electrophoresis by Means of Topological Indices', *J. Chromatogr. A*, **798**, 233–242 (1998).
263. M.A. Aramendia, V. Borau, I. Garcia, C. Jimenez, F. Lafont, J.M. Marinas, A. Porras, F.J. Urbano, 'Determination of Isoflavones by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry', *J. Mass Spectrom.*, S153–S157 (1995).
264. M.A. Aramendia, I. Garcia, F. Lafont, J.M. Marinas, 'Determination of Isoflavones Using Capillary Electrophoresis in Combination with Electrospray Mass Spectrometry', *J. Chromatogr. A*, **707**, 327–333 (1995).
265. R.J. Ochocka, D. Rajzer, P. Kowalski, H. Lamparczyk, 'Determination of Coumarins from *Chrysanthemum segetum* L. by Capillary Electrophoresis', *J. Chromatogr. A*, **709**, 197–202 (1995).
266. C.T. Chen, S.J. Sheu, 'Separation of Coumarins by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **710**, 323–329 (1995).
267. D.P. Bogan, R.D. Thornes, M. Tegtmeier, E.A. Schafer, R. O'Kennedy, 'Direct Determination of 7-Hydroxycoumarin and 7-Hydroxycoumarin-glucuronide in Urine by Using Capillary Electrophoresis', *Analyst*, **121**, 243–247 (1996).
268. J.F. Jen, Y.H. Hsu, M.R. Lee, 'Separation of Plant Phenolic Compounds by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **734**, 375–380 (1996).
269. P.J. Larger, A.D. Jones, C. Dacombe, 'Separation of Tea Polyphenols Using Micellar Electrokinetic Chromatography with Diode Array Detection', *J. Chromatogr. A*, **799**, 309–320 (1998).
270. A. Hiermann, B. Radl, 'Analysis of Aromatic Plant Acids by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **803**, 311–314 (1998).
271. P. Bridle, C. Garcia-Viguera, F. Tomas-Barberan, 'Analysis of Anthocyanins by Capillary Zone Electrophoresis', *J. Liq. Chromatogr. Relat. Technol.*, **19**, 537–545 (1996).
272. P. Bridle, C. Garcia-Viguera, 'Analysis of Anthocyanins in Strawberries and Elderberries. A Comparison of Capillary Zone Electrophoresis and HPLC', *Food Chem.*, **59**, 299–304 (1997).
273. C.T. daCosta, B.C. Nelson, S.A. Margolis, D. Horton, 'Separation of Blackcurrant Anthocyanins by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **799**, 321–327 (1998).
274. T. Watanabe, N. Hasegawa, A. Yamamoto, S. Nagai, S. Terabe, 'Separation and Determination of Yellow and Red Safflower Pigments in Food by Capillary Electrophoresis', *Biosci. Biotechnol. Biochem.*, **61**, 1179–1183 (1997).
275. Y.P. Zhao, C.E. Lunte, 'Determination of Caffeine and Its Metabolites by Micellar Electrokinetic Capillary Electrophoresis', *J. Chromatogr. B*, **688**, 265–274 (1997).
276. R. Schuster, 'Capillary Zone Electrophoresis Analysis of Additives in Food Samples', in *Handbook of*

Capillary Electrophoresis Applications, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 607–616, 1996.

277. E. Dabek-Zlotorzynska, 'Environmental Pollutants', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonski, Blackie Academic and Professional, London, 639–656, 1996.
278. Z.K. Shihabi, 'Therapeutic Drug Monitoring by Capillary Electrophoresis', *J. Chromatogr. A*, **807**, 27–36 (1998).
279. Z. El Rassi, 'Capillary Electrophoresis of Pesticides', *Electrophoresis*, **18**, 2465–2481 (1997).
280. Y. Ma, K. Meyer, D. Afzal, 'Isolation and Quantification of Ergovaline from *Festuca arundinacea* (Tall Fescue) Infected with the Fungus *Acremonium coenophilum* by High Performance Capillary Electrophoresis', *J. Chromatogr. A*, **652**, 535–538 (1993).
281. Y. Liu, S. Sheu, 'Determination of Quaternary Alkaloids from *Phellodendri cortex* by Capillary Electrophoresis', *J. Chromatogr.*, **634**, 329–333 (1993).
282. M. Unger, D. Stockigt, D. Belder, J. Stockigt, 'General Approach for the Analysis of Various Alkaloid Classes Using Capillary Electrophoresis and Capillary Electrophoresis Mass Spectrometry', *J.Chromatogr. A*, **767**, 263–276 (1997).
283. M. Unger, J. Stockigt, 'Improved Detection of Alkaloids in Crude Extracts Applying Capillary Electrophoresis with Field Amplified Sample Injection', *J. Chromatogr. A*, **791**, 323–331 (1997).
284. M. Unger, 'Capillary Zone Electrophoresis of Alkaloids: Influence of Structure on Electrophoretic Mobility', *J. Chromatogr. A*, **807**, 81–87 (1997).
285. J. Vandepeer, V.C. Trenerry, G. Keogh, 'Some Examples of the Analysis of Complex Samples by Micellar Electrokinetic Capillary Chromatography with Photodiode Array Detection', *Electrophoresis*, **17**, 1361–1366 (1996).
286. K. Frach, G. Blaschke, 'Separation of Ergot Alkaloids and Their Epimers and Determination in Sclerotia by Capillary Electrophoresis', *J. Chromatogr. A*, **808**, 247–252 (1998).
287. Y.-Y. Zong, C.-T. Che, 'Determination of Strychnine and Brucine by Capillary Electrophoresis', *Planta Med.*, **61**, 456–458 (1998).
288. S.S. Yang, I. Smetena, 'Evaluation of Capillary Electrophoresis for the Analysis of Nicotine and Selected Minor Alkaloids from Tobacco', *Chromatographia*, **40**, 375–378 (1995).
289. J.J. Yang, H. Long, H.W. Liu, A.J. Huang, Y.L. Sun, 'Analysis of Tetrandrine and Fangchinoline in Traditional Chinese Medicines by Capillary Electrophoresis', *J. Chromatogr. A*, **811**, 274–279 (1998).
290. B. Blais, A. Cunningham, H. Yamazaki, 'A Novel Immunofluorescence Capillary Electrophoresis Assay System for the Determination of Chloramphenicol in Milk', *Food Agric. Immunol.*, **6**, 409–417 (1994).
291. I. Pant, V.C. Trenerry, 'The Determination of Ascorbic Acid and Benzoic Acid in Foods and Beverages by Micellar Electrokinetic Capillary Chromatography', *Food Chem.*, **53**, 219–226 (1995).
292. G. Dinelli, A. Vicari, A. Bonetti, P. Catizone, 'Comparison of Capillary Electrophoresis, HPLC, and Enzyme Immunoassay for Terbutylazine Detection in Water', *J. Agric. Food Chem.*, **43**, 951–955 (1995).
293. G. Dinelli, A. Vicari, A. Bonetti, 'Separation of Sulfonylurea Metabolites in Water by Capillary Electrophoresis', *J. Chromatogr. A*, **700**, 195–200 (1995).
294. G. Dinelli, A. Vicari, V. Brandolini, 'Detection and Quantitation of Sulfonylurea Herbicides in Soil at the ppb Level by Capillary Electrophoresis', *J. Chromatogr. A*, **700**, 201–207 (1995).
295. A. Farran, S. Ruiz, 'Simultaneous Separation of Neutral and Acidic Herbicides by Micellar Electrokinetic Capillary Chromatography', *Int. J. Environ. Anal. Chem.*, **58**, 121–129 (1995).
296. Y. Wigfield, K. McCormack, R. Grant, 'Simultaneous Determination of Residues of Paraquat and Diquat in Potatoes Using High-performance Capillary Electrophoresis with Ultaviolet Detection', *J. Agric. Food Chem.*, **41**, 2315–2318 (1993).
297. P. Schmitt, A. Garrison, D. Freitag, A. Kettrup, 'Separation of s-Triazine Herbicides and Their Metabolites by Capillary Zone Electrophoresis as a Function of pH', *J. Chromatogr. A*, **723**, 169–177 (1996).
298. S. Nemoto, S.J. Lehotay, 'Analysis of Multiple Herbicides in Soybeans Using Pressurized Liquid Extraction and Capillary Electrophoresis', *J. Agric. Food Chem.*, **42**, 2109–2199 (1998).
299. J.P. Alcantara Licudine, N.L. Bui, M.K. Kawate, Q.X. Li, 'Analysis of Phloxine B and Uranine in Coffee by High-performance Liquid Chromatography and Capillary Zone Electrophoresis After Solid Phase Extraction Cleanup', *J. Agric. Food Chem.*, **46**, 1005–1011 (1998).
300. B. Buscher, H. Irth, E. Andersson, U. Tjaden, J. van der Greef, 'Determination of Inositol Phosphates in Fermentation Broth Using Capillary Electrophoresis with Indirect UV Detection', *J.Chromatogr. A*, **678**, 145–150 (1994).
301. P. Blatny, F. Kvasnicka, E. Kenndler, 'Determination of Phytic Acid in Cereal Grains, Legumes, and Feeds by Capillary Isotachophoresis', *J. Agric. Food Chem.*, **43**, 129–133 (1995).
302. C. Maragos, 'Capillary Zone Electrophoresis and HPLC for the Analysis of Fluorescein Isothiocyanate-labeled Fumonisin B1', *J. Agric. Food Chem.*, **43**, 390–394 (1995).
303. A. Arentoft, B. Greirson, 'Analysis of 3-(*N*-Oxalyl)-L-2,3-diaminopropionic Acid and Its Alpha-isomer in Grass Pea (*Lathyrus sativus*) by Capillary Zone Electrophoresis', *J. Agric. Food Chem.*, **43**, 942–945 (1995).
304. C.J. Sciacchitano, I.N. Hirshfield, 'Molecular Detection of *Clostridium botulinum* Type E Neurotoxin Gene in

Smoked Fish by Polymerase Chain Reaction and Capillary Electrophoresis', *J. AOAC Int.*, **79**, 861–865 (1996).
305. M.S. Nielsen, P.V. Nielsen, J.C. Frisvad, 'Micellar Electrokinetic Capillary Chromatography of Fungal Metabolites – Resolution Optimized by Experimental Design', *J. Chromatogr. A*, **721**, 337–344 (1996).
306. B. Mopper, C. Sciacchitano, 'Capillary Zone Electrophoretic Determination of Histamine in Fish', *J. Assoc. Off. Anal. Chem.*, **77**, 881–884 (1994).
307. A. Buzy, P. Thibault, M. Laycock, 'Development of a Capillary Electrophoresis Method for the Characterization of Enzymatic Products Arising from the Carbamoylase Digestion of Paralytic Shellfish Poisoning Toxins', *J. Chromatogr. A*, **688**, 301–316 (1994).
308. J.M. Gallardo, C.G. Sotelo, R.I. Perez-Martin, 'Determination of Histamine by Capillary Zone Electrophoresis Using a Low pH Phosphate Buffer: Application in the Analysis of Fish and Marine Products', *Z. Lebensm. Unters. Forsch. A, Food Res. Technol.*, **204**, 336–340 (1997).
309. J.M. Gallardo, C.G. Sotelo, C. Pineiro, R.I. Perez-Martin, 'Use of Capillary Zone Electrophoresis for Fish Species Identification. Differentiation of Flatfish Species', *J. Agric. Food Chem.*, **43**, 1238–1244 (1995).
310. J.Y. Zhao, P. Thibault, M.A. Quilliam, 'Analysis of Domoic Acid and Isomers in Seafood by Capillary Electrophoresis', *Electrophoresis*, **18**, 268–276 (1997).
311. N. Bouaicha, M.C. Hennion, P. Sandra, 'Determination of Okadaic Acid by Micellar Electrokinetic Chromatography with Ultraviolet Detection', *Toxicon*, **35**, 273–281 (1997).
312. C. Ng, C. Ong, H. Lee, S. Li, 'Capillary Electrophoretic Separation of Nitrosamines Using Combined Open-tubular and Packed Capillary Columns', *J. Chromatogr. Sci.*, **32**, 121–125 (1994).
313. C. Corradini, I. Nicoletti, G. Cannarsa, D. Corradini, L. Fizzoferrato, V. Vivanti, in *Proceedings of Euro-food Chem*, VIII2, 299–302, 1995.
314. A. Tirelli, L. Pellegrino, 'Determination of Furosine in Dairy Products by Capillary Zone Electrophoresis: A Comparison with the HPLC Method', *Ital. J. Food Sci.*, **7**, 379–385 (1995).
315. M. Trevaskis, V.C. Trennery, 'An Investigation into the Determination of Oxalic Acid in Vegetables by Capillary Electrophoresis', *Food Chem.*, **57**, 323–330 (1996).
316. J. Liu, S.F.Y. Li, 'Separation and Determination of *Stevia* Sweeteners by Capillary Electrophoresis and High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **18**, 1703–1719 (1995).
317. C. Thompson, V. Trenerry, B. Kemmery, 'Micellar Electrokinetic Capillary Chromatographic Determination of Artificial Sweeteners in Low-joule Soft Drinks and Other Foods', *J. Chromatogr. A*, **694**, 507–514 (1995).
318. C. Thompson, V. Trenerry, B. Kemmery, 'Determination of Cyclamate in Low Joule Foods by Capillary Zone Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr. A*, **704**, 203–210 (1995).
319. C. Thompson, V. Trenerry, 'Determination of Synthetic Colors in Confectionery and Cordials by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **704**, 195–201 (1995).
320. S. Razee, A. Tamura, T. Masujima, 'Improvement in the Determination of Food Additive Dyestuffs by Capillary Electrophoresis Using Beta-cyclodextrin', *J. Chromatogr. A*, **715**, 179–188 (1995).
321. H.W. Liu, T. Zhu, Y.N. Zhang, S.Z. Qi, A.J. Huang, Y.L. Sun, 'Determination of Synthetic Colourant Food Additives by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **718**, 448–453 (1995).
322. S. Abrantes, M.R. Philco, A.P. Damant, L. Castle, 'Application of Capillary Electrophoresis in Migration Studies of Food Contact Materials – Part 1: Substituted Phenolic Additives', *J. High Resolut. Chromatogr.*, **20**, 270–274 (1997).
323. H. Horie, K. Kohata, 'Application of Capillary Electrophoresis to Tea Quality Estimation', *J. Chromatogr. A*, **802**, 219–223 (1998).
324. M.R. Bronze, L.F.V. Boas, A.P. Belchior, 'Analysis of Old Brandy and Oak Extracts by Capillary Electrophoresis', *J. Chromatogr. A*, **768**, 143–152 (1997).
325. P.F. Cancalon, 'Capillary Electrophoresis and Fruit Juice Monitoring', in *Methods to Detect Adulteration of Fruit Juice Beverages*, eds. S. Nagy, R.L. Wade, AGScience, Auburndale, FL, 181–197, 1995.

Enzyme Analysis and Bioassays in Food Analysis

J.R. Powers
Washington State University, Pullman, USA

1 Introduction **1**
2 Enzymology **1**
2.1 Overview 1
2.2 Kinetics 2
2.3 Factors that Affect Enzyme Reaction Rate 2
2.4 Methods of Measurement 3
2.5 Applications 4
3 Immunology **10**
3.1 Overview 10
3.2 Methods 10
3.3 Applications 12
4 Analytical Microbiology **13**
4.1 Overview 13
4.2 Applications 13
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

The objectives for carrying out analysis of food products include food safety, food quality, meeting regulatory requirements and the control of food processes. Biological entities used in food analyses include enzymes, microbes and antibodies. Based on categories of food analysis general principles of enzyme, immunological and whole organism (microbial) analyses are described. Applications of the use of enzymes for substrate analysis, with examples of their application, are presented. A brief description is given of the use of enzymes as preparative tools in food analysis and of biosensors that use enzymes as elements. Immunological analysis applicable to food systems is given with an emphasis on applications related to food safety. Microbial assays used in food analysis are described including vitamin analysis and protein quality.

1 INTRODUCTION

The major purposes for carrying out analyses of food products include assessing food safety and quality, meeting regulatory requirements, and providing information for the control of food processes. Biological tools used in food analyses include enzymes, microbes and antibodies. These biological entities can play a role in accomplishing each of the principal purposes of food analysis. Based on these categories of analysis, general principles of enzyme, immunological and whole organism (microbial) analyses are given. Applications of the use of enzymes for substrate analysis, with examples of their application, are presented. A brief description is given of the use of enzymes as preparative tools in food analysis, and of biosensors that use enzymes as elements. Immunological analyses applicable to food systems are given, with an emphasis on applications related to food safety. Microbial assays used in food analysis are described, including vitamin analysis, detection of antibiotics and assessment of protein quality.

2 ENZYMOLOGY

2.1 Overview

Enzymatic analysis is the measurement of analytes with the aid of added enzymes or the measurement of endogenous enzyme activity to give an indication of the state of a biological system including those related to food. Enzyme catalysis can take place under relatively mild conditions, allowing for measurement of relatively unstable compounds not amenable to some other techniques. In addition, the specificity of enzyme reactions can allow for measurement of components of complex mixtures without the time and expense of complicated separation techniques.

There are several uses of enzyme analyses in food science and technology. In several instances, enzyme activity is a useful measure for adequate processing of a food product. Enzyme activity assays are also used by the food technologist to assess potency of enzyme preparations used as processing aids.

Food scientists also use commercially available enzyme preparations to measure constituents of foods that are enzyme substrates. A corollary of the use of commercially available enzymes is to measure enzyme activity as a function of enzyme inhibitor content in a food. For example, organophosphate insecticides are potent inhibitors of the enzyme acetylcholinesterase, hence the activity of this enzyme in the presence of a food extract is a measure of this category of insecticide concentration in the food. Also of interest is the measurement of enzyme activity associated with food quality. Another use of enzyme assays to determine food quality is estimation of protein nutritive value by monitoring the activity of added proteases on food protein samples. Enzymes can also be used to measure the appearance of degradation

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

products in foods during storage. Also, enzymes are used as preparative tools in food analysis.

In order to carry out enzyme analyses in foods successfully, an understanding of certain basic principles of enzymology is necessary. After a brief overview of these principles, examples of the use of enzymatic analyses in food systems are given.

2.2 Kinetics

Enzymes are proteins that are biological catalysts. A catalyst increases the velocity of a thermodynamically possible reaction. The enzyme does not modify the equilibrium constant of the reaction, and the catalyst is not consumed in the reaction. Because enzymes affect rates (velocities) of reactions, some knowledge of *enzyme kinetics* is needed to use enzymes effectively in analysis. To measure the rate of an enzyme-catalyzed reaction, typically the enzyme is mixed with the substrate under specified conditions (pH, temperature, ionic strength, etc.) and the reaction is followed by measuring the amount of product that appears, or by measuring the disappearance of substrate. Consider Equation (1) as a simple representation of an enzyme-catalyzed reaction:

$$S + E \rightleftharpoons ES \longrightarrow P + E \qquad (1)$$

where S is the substrate, E is the enzyme, ES is the enzyme–substrate complex and P is the product.

The formation of the enzyme–substrate complex is very rapid and is not normally seen in the laboratory. The brief time in which the enzyme–substrate complex is initially formed is on the millisecond scale and is called the *presteady state period*. The rate of change of the product concentration following the presteady state period gives us the *initial velocity* (v_0). After the presteady state period, a *steady state period* exists in which the concentration of the enzyme–substrate complex is constant. A time course needs to be established experimentally by using a series of points or a continuous assay to establish the appropriate timeframe for the measurement of the initial velocity.

The rate of an enzyme-catalyzed reaction depends on the concentration of the enzyme and also on the substrate concentration. With a fixed enzyme concentration, increasing substrate concentration will result in an increased velocity. As substrate concentration increases further, the increase in velocity slows until, with a very large concentration of substrate, no further increase in velocity is noted. The velocity of the reaction at this very large substrate concentration is the *maximum velocity* (V_m) of the reaction under the conditions of that particular assay. The substrate concentration at which one-half V_m is observed is defined as the *Michaelis constant* (K_m), an important characteristic of an enzyme. It is an indication of the relative binding affinity of the enzyme for a particular substrate. The lower the K_m, the greater the affinity of the enzyme for the substrate. Equation (2) is the Michaelis–Menten equation, the equation for a right hyperbola.

$$v_0 = \frac{V_m[S]}{K_m + [S]} \qquad (2)$$

2.2.1 Order of Reaction

The velocity of an enzyme-catalyzed reaction increases as substrate concentration increases, as noted above. A *first order* reaction with respect to substrate concentration is obeyed when substrate concentration is small ($[S] \ll K_m$). This means that the velocity of the reaction is directly proportional to the substrate concentration in this region. When the substrate concentration is further increased, the velocity of the reaction no longer increases linearly, and the reaction is *mixed order*. If substrate concentration is increased further, the velocity asymptotically approaches the maximum velocity (V_m). Under these conditions, the velocity is independent of substrate concentration. Note that at large substrate concentrations ($[S] \gg K_m$), the velocity is directly proportional to enzyme concentration ($V_m = k_2[E_0]$). Thus, where $[S] \gg K_m$, the rate of the reaction is *zero order* with respect to substrate concentration (is independent of substrate concentration) but first order with respect to enzyme concentration.

Thus, if one is interested in measuring the amount of enzyme in a reaction mixture, if possible substrate concentrations should be used so that the observed velocity approximates to V_m. Conversely, if one is interested in measuring substrate concentration by measuring initial velocity, substrate concentrations less than K_m must be used, in order to have a rate directly proportional to substrate concentration.

2.3 Factors that Affect Enzyme Reaction Rate

The velocity of an enzyme-catalyzed reaction is affected by a number of factors, including enzyme and substrate concentrations, temperature, pH, ionic strength, and the presence of inhibitors and activators. The velocity of an enzyme-catalyzed reaction will depend on the enzyme concentration in the reaction mixture, and doubling the enzyme concentration should double the rate of the reaction. It is critical that the substrate concentration is saturating during the entire period the reaction mixture is sampled and the measured amount of product formed or substrate disappearing is linear over the period during which the reaction is sampled. The activity of the enzyme is obtained as the slope of the linear part of the line of a plot of product or substrate concentration versus time. If a large number of samples are to be assayed, a single aliquot is often taken at a single time. This will

give good results only if the time at which the sample is taken falls on the linear portion of a plot of substrate concentration or product concentration versus time of reaction. The plot becomes nonlinear if the substrate concentration falls below the concentration needed to saturate the enzyme, if the increase in concentration of product produces a significant amount of back reaction, or if the enzyme loses activity during the time of the assay. Normally, an experiment is designed in which enzyme concentration is estimated such that no greater than 5–10% of the substrate has been converted to product within the time used for measuring the rate. A better method of estimating rates is to measure initial rates of the reactions, in which the change in substrate or product concentration is determined at times as close as possible to time of initiation of the reaction.

Sometimes it is not possible to carry out enzyme assays at $[S] \gg K_m$. The substrate may be very expensive, relatively insoluble, or K_m may be large (i.e. $K_m >$ 100 mM). When substrate concentration is much less than K_m, the substrate term in the denominator of the Michaelis–Menten equation can be ignored and $v = (V_m[S])/K_m$, which is the equation for a first-order reaction with respect to substrate concentration. Under these conditions, a plot of product concentration versus time gives a nonlinear plot. A plot of $\log([S_0]/[S])$ versus time gives a straight line relationship. The slope of the line of the log plot is directly related to the enzyme concentration. When the slope of a series of these log plots is further plotted as a function of enzyme concentration, a straight line relationship should result. If possible, the reaction should be followed continuously or aliquots removed at frequent time intervals.

2.3.1 Effect of Temperature

Temperature can affect observed enzyme activity in several ways. Most obvious is the fact that temperature can affect the stability of the enzyme and also the rate of the enzyme-catalyzed reaction. Other factors in enzyme-catalyzed reactions that may be considered include the effect of temperature on the solubility of gases that are either products or substrates of the observed reaction, and the effect of temperature on the pH of the system. A good example of the latter is the common buffering species Tris (tris(hydroxymethyl)aminomethane), for which the pK_a changes 0.031 per 1 °C of change.

Temperature affects both the stability and the activity of the enzyme. At relatively low temperatures, the enzyme is stable. However, at higher temperatures denaturation dominates and a markedly reduced enzyme activity is observed. The velocity is expected to increase as the temperature is increased. The velocity approximately doubles for every 10 °C rise in temperature. The temperature optimum is not a unique characteristic of the enzyme. The optimum applies instead to the entire system because type of substrate, pH, salt concentration, substrate concentration and time of reaction can affect the observed optimum. For this reason, investigators should fully describe the system in which the effects of temperature on observed enzyme activity are reported.

2.3.2 Effect of pH

The observed rate of an enzyme-catalyzed reaction is greatly affected by the pH of the medium. Enzymes have pH optima and commonly have bell-shaped curves for activity versus pH. The observed effect of pH on enzyme activity is a manifestation of the effects of pH on enzyme stability and rate of substrate to product conversion, and may also be due to changes in ionization of the substrate.

To establish the pH optimum for an enzyme reaction, the reaction mixture is buffered at different pHs and the activity of the enzyme is determined. To determine pH–enzyme stability relationships, aliquots of enzyme are buffered at different pH values and held for a specified period of time (e.g. 1 h). The pH of the aliquots is then adjusted to the pH optimum, and each aliquot is assayed. These studies are helpful in establishing conditions for handling the enzyme, and may be useful in establishing methods for controlling enzyme activity in a food system. Note that pH stability and the pH optimum for the enzyme activity are not true constants. These may vary with a particular source of enzyme, the specific substrate used, the temperature of the experiment, and the buffering species used in the experiment. In using enzymes for analysis, it is not necessary for the reaction be carried out at the pH optimum for activity, or even at a pH at which the enzyme is most stable. However, it is critical to maintain a fixed pH during the reaction (i.e. use buffer) and to use the same pH in all studies to be compared.

2.3.3 Inhibitors

An *enzyme inhibitor* is a compound which, when present in an enzyme-catalyzed reaction medium, decreases the enzyme activity. Enzyme inhibitors can be categorized as *irreversible* or *reversible* inhibitors. Enzyme inhibitors include inorganic ions, such as Pb^{2+} or Hg^{2+} that can react with sulfydryl groups to inactivate the enzyme, compounds that resemble substrate, and naturally occurring proteins that specifically bind to enzymes (e.g. protease inhibitors found in legumes).

2.4 Methods of Measurement

For practical enzyme analysis, it is necessary to be familiar with the methods of measurement of the reaction. Any physical or chemical property of the system that relates to

substrate or product concentration can be used to follow an enzyme reaction. A wide variety of methods are available to follow enzyme reactions, including absorbance spectrometry, fluorimetry, manometric methods, titration, isotope measurement, and viscosity. An example of the use of spectrophotometry as a method for following enzyme reactions is use of the spectra of the pyridine coenzyme NADH (nicotinamide adenine dinucleotide (reduced form)) and NADPH (nicotinamide adenine dinucleotide phosphate (reduced form)), in which there is a marked change in absorbance at 340 nm upon oxidation/reduction. Many methods depend on the increase or decrease in absorbance at 340 nm when these coenzymes are products or substrates in a coupled reaction.

An example of using several methods to measure the activity of an enzyme is in the assay of α-amylase activity.[1] α-Amylase cleaves starch at α-1,4 linkages in starch and is an endoenzyme. This reaction can be monitored by a number of methods, including reduction in viscosity, increase in reducing groups upon hydrolysis, reduction in color of the starch–iodine complex, and polarimetry. However, it is difficult to differentiate between the activities of α-amylase and β-amylase using a single assay. β-Amylase cleaves maltose from the nonreducing end of starch. While a marked decrease in viscosity of starch or reduction in iodine color would be expected to occur because of α-amylase activity, β-amylase can also cause changes in viscosity and iodine color if in high concentration. In order to establish whether α-amylase or β-amylase is being measured, the change in the number of reducing groups must be determined as a basis of comparison. Because α-amylase is an endoenzyme, hydrolysis of a few bonds near the center of the polymeric substrate will cause a marked decrease in viscosity, while hydrolysis of an equal number of bonds by the exoenzyme β-amylase will have little effect on viscosity. If options are available in methodology, a method should be selected that is able to monitor the reaction continuously, is sensitive, and is specific for the enzyme-catalyzed reaction.

2.4.1 Coupled Reactions

Enzymes can be used in assays via coupled reactions. *Coupled reactions* involve the use of two or more enzyme reactions so that a substrate or product concentration can be readily followed. In using a coupled reaction, there is a *measuring reaction* (Equation 3) and an *indicator reaction* (Equation 4). For example

$$\mathrm{S1} \xrightarrow{\mathrm{E1}} \mathrm{P1} \qquad (3)$$

$$\mathrm{P1} \xrightarrow{\mathrm{E2}} \mathrm{P2} \qquad (4)$$

The role of the indicating enzyme (E2) is to produce P2, which is readily measurable, and hence an indication of the amount of P1 produced by E1. Alternatively, the same sequence can be used in measuring S1, the substrate for E1. When a coupled reaction is used to measure the activity of an enzyme (e.g. E1 above), it is critical that the indicating enzyme E2 not be rate limiting in the reaction sequence. The measuring reaction must always be rate determining. Consequently, E2 activity should be much greater than E1 activity for an effective assay. Coupled enzyme reactions can have problems with respect to pH of the system if the pH optima of the coupled enzymes are quite different. It may be necessary to allow the first reaction (e.g. the measuring reaction catalyzed by E1 above) to proceed for a time, and then arrest the reaction by heating, to denature E1. The pH is adjusted, the indicating enzyme E2 added, and the reaction completed. If an end-point method is used with a coupled system, the requirements for pH compatibility are not as stringent as for a rate assay, because an extended time period can be used to allow the reaction sequence to go to completion.

2.5 Applications

As described above, certain information is needed prior to using enzyme assays analytically. In general, knowledge of K_m, the time course of the reaction, the enzyme specificity for substrate, the pH optimum and pH stability of the enzyme, and the effects of temperature on the reaction and stability of the enzyme are desirable. This information is available from the literature in many instances. However, a few preliminary experiments may be necessary, especially in the case of experiments in which velocities are measured. A time course to establish linearity of product formation or substrate consumption in the reaction is a necessity. An experiment to show linearity of velocity of the enzyme reaction to enzyme concentration is recommended.

2.5.1 Substrate Assays

The following is not an extensive compendium of methods for the measurement of food components by enzymatic analysis. Instead, it is meant to be representative of the types of analyses possible. The reader can consult handbooks published by the manufacturers of enzyme kits, for example, Boehringer-Mannheim,[2] the review article by Whitaker,[3] (see Table 1) and the series by Bergmeyer[4] for a more comprehensive guide to enzyme methods applicable to foods.

2.5.1.1 Sample Preparation Because of the specificity of enzymes, sample preparation prior to enzyme analysis is often minimal, and may involve only extraction and removal of solids by filtration or centrifugation.

Table 1 Analytical determination of compounds that serve as substrates. (Adapted from Whitaker[3])

Compound	Enzyme	Detection[a]
Acetate	Acetate kinase	S; $NAD^+ \longrightarrow NADH$, in coupled assay
Acetylcholine	Choline kinase	R; ^{32}P-phosphoryl choline
L-Amino acids	L-Amino acid oxidase	S; H_2O_2, coupled with peroxidase
D-Amino acids[b]	D-Amino acid oxidase	S; H_2O_2, coupled with peroxidase E; NH_3
Ammonia	Glutamate dehydrogenase	S; $NADH \longrightarrow NAD^+$
L-Aspartate	Glutamate oxaloacetate transaminase	S; $NADH \longrightarrow NAD^+$, in coupled assay
L-Asparagine	Asparaginase	S; $NADH \longrightarrow NAD^+$ in coupled assay through aspartic acid
Carbon dioxide	Carbonic anhydrase	E; H^+
Cellulose	Cellulase	S; glucose in coupled reaction
Cholesterol	Cholesterol esterase	S; H_2O_2 with peroxidase
	Cholesterol oxidase	S; H_2O_2 with peroxidase
Citrate	3-Hydroxybutyrate dehydrogenase	S; $NADH \longrightarrow NAD^+$
Isocitrate	Isocitrate dehydrogenase	S; $NADP^+ \longrightarrow NADPH$
Creatine	Creatine kinase	S; $NADH \longrightarrow NAD^+$, in coupled assay
Creatinine	Creatininase	S; $NADH \longrightarrow NAD^+$, in coupled assay involving creatine (above)
Ethanol	Alcohol dehydrogenase	S; $NAD^+ \longrightarrow NADH$
D-Fructose	Hexokinase	S; $NADP^+ \longrightarrow NADPH$, in coupled assay
D-Galactose	Galactose oxidase	S; H_2O_2 with peroxidase
	Galactose dehydrogenase	S; $NAD^+ \longrightarrow NADH$
D-Glucose	Glucose oxidase	S; H_2O_2 with peroxidase
	Hexokinase	S; $NADP^+ \longrightarrow NADPH$, in coupled assay
D-Gluconic acid	Gluconokinase	S; $NADP^+ \longrightarrow NADPH$, in coupled assay
L-Glutamate	Glutamate decarboxylase	E; CO_2 electrode
	Glutamate dehydrogenase	S; $NAD^+ \longrightarrow NADH$
L-Glutamine	Glutaminase	S; $NAD^+ \longrightarrow NADH$, in coupled reaction
Glycerol	Glycerol kinase	S; $NAD^+ \longrightarrow NADH$, in coupled assay
Glycogen	Amyloglucosidase	S; $NADH \longrightarrow NAD^+$, in coupled assay through glucose
L-Lactic acid	L-Lactate dehydrogenase	S; $NAD^+ \longrightarrow NADH$
D-Lactic acid	D-Lactate dehydrogenase	S; $NAD^+ \longrightarrow NADH$
Lactose	β-Galactosidase	S; H_2O_2 with peroxidase in glucose determination
Maltose	α-Glucosidase	S; H_2O_2 with peroxidase in glucose determination
Mannose	Hexokinase[c]	S; $NADP^+ \longrightarrow NADPH$, in coupled assay
Nitrate	Nitrate reductase	E; NH_3 electrode
Nitrite	Nitrite reductase	E; NH_3 electrode
Oxalate	Oxalate decarboxylase	S; CO_2 with indicator
Peroxide	Peroxidase	S; oxidized dye
L-Phenylalanine	L-Phenylalanine ammonia-lyase	S; $NADH \longrightarrow NAD^+$, in coupled assay
Pyruvate	L-Lactate dehydrogenase	S; $NADH \longrightarrow NAD^+$
Starch	Amyloglucosidase	S; $NADH \longrightarrow NAD^+$, in coupled assay through glucose
Sucrose	Invertase	S; H_2O_2 with peroxidase in glucose determination
Urea	Urease	E; NH_3 electrode

[a] E, electrometric method; F, fluorescent method; R, radioactive method; S, spectrophotometric method; NAD^+, nicotinamide adenine dinucleotide (oxidized form); $NADP^+$, nicotinamide adenine dinucleotide phosphate (oxidized form).
[b] Determines the D forms of alanine, methionine, phenylalanine, proline, tryptophan, tyrosine, and leucine nonspecifically.
[c] Fructose-6-phosphate and glucose-6-phosphate will also be determined by this assay method.

Nevertheless, owing to the wide variety of foods that might be encountered by the analyst using enzyme assays, a check should be made of the extraction and enzyme reaction steps by standard addition of known amounts of analyte to the food and extract, and measurement of recovery of that standard. If the standard additions are fully recovered, this is a positive indication that the extraction is adequate, that the sample does not contain interfering substances that require removal prior to the enzymatic analysis. In some cases, interfering substances are present but can be readily removed by precipitation or adsorption. For example, polyvinylpolypyrrolidone (PVPP) powder can be used to decolorize juices or red wines. With the advent of small syringe minicolumns (e.g. C_{18}, silica, and ion-exchange cartridges), it is also relatively easy and fast to attain group separations to remove interfering substances from a sample extract.

2.5.1.2 Total Change/End-point Methods While substrate concentrations can be determined in rate assays when the reaction is first order with respect to substrate concentration, ($[S] \ll K_m$), substrate concentration can

also be determined by the total change or end-point method. In this method, the enzyme-catalyzed reaction is allowed to go to completion, so that concentration of product can be measured and directly related to substrate. An example of such a system is the measurement of glucose using glucose oxidase and peroxidase, as described below.

In some cases, an equilibrium is established in an end-point method, in which there is a significant amount of substrate remaining in equilibrium with the product. In these cases, the equilibrium can be altered. For example, in cases in which a proton-yielding reaction is used, alkaline conditions (increase in pH) can be used. Trapping agents can also be used, in which the product is effectively removed from the reaction, and the reaction goes to completion by mass action. Examples include the trapping of ketones and aldehydes by hydrazine. In this way, the product is continually removed and the reaction pulled to completion. The equilibrium can also be displaced by increasing cofactor or coenzyme concentration.

Another means of driving a reaction to completion is a regenerating system.[4] For example, in the measurement of glutamate, with the aid of glutamate dehydrogenase, the reactions in Equations (5) and (6) can be effected:

$$\text{glutamate} + \text{NAD (nicotinamide adenine dinucleotide)} + H_2O \overset{\text{dehydrogenase}}{\rightleftharpoons} \alpha\text{-ketoglutarate} + \text{NADH} + NH_4^+ \quad (5)$$

$$\text{pyruvate} + \text{NADH} + H^+ \xrightarrow{\text{lactate dehydrogenase}} NAD^+ + \text{lactate} \quad (6)$$

In this system, NADH is recycled to NAD^+ via lactate dehydrogenase until all the glutamate to be measured is consumed. The reaction is stopped by heating to denature the enzymes present, a second aliquot of glutamate dehydrogenase and NADH is added, and the α-ketoglutarate (equivalent to the original glutamate) measured via decrease in absorbance at 340 nm. An example in which the same equilibrium is displaced in the measurement of glutamate is given by Equations (7) and (8):

$$\text{glutamate} + \text{NAD} + H_2O \overset{\text{glutamate dehydrogenase}}{\rightleftharpoons} \alpha\text{-ketoglutarate} + \text{NADH} + NH_4^+ \quad (7)$$

$$\text{NADH} + \text{INT} \xrightarrow{\text{diaphorase}} NAD^+ + \text{formazan} \quad (8)$$

Iodonitrotetrazolium chloride (INT) is a trapping reagent for the NADH product of the glutamate dehydrogenase-catalyzed reaction. The formazan formed is measurable colorimetrically at 492 nm.

2.5.1.3 Measurement of Sulfite Sulfite is a food additive that can be measured by several techniques, including titration, distillation followed by titration, gas chromatography, and colorimetric analysis. Sulfite can also be specifically oxidized to sulfate by the commercially available enzyme sulfite oxidase (SO); see Equation (9):

$$SO_3^{2-} + O_2 + H_2O \xrightarrow{\text{SO}} SO_4^{2-} + H_2O_2 \quad (9)$$

The H_2O_2 product can be measured by several methods including use of the enzyme NADH-peroxidase (Equation 10):

$$H_2O_2 + \text{NADH} + H^+ \xrightarrow{\text{NADH-peroxidase}} 2H_2O + NAD^+ \quad (10)$$

The amount of sulfite in the system is equal to the NADH oxidized, which is determined by a decrease in absorbance at 340 nm. Ascorbic acid can interfere with the assay, but can be removed by using ascorbic acid oxidase.[5]

2.5.1.4 Colorimetric Determination of Glucose The combination of enzymes glucose oxidase and peroxidase can be used specifically to measure glucose in a food system.[6] Glucose is preferentially oxidized by glucose oxidase to produce gluconolactone and hydrogen peroxide. The hydrogen peroxide plus *o*-dianisidine in the presence of peroxidase produces a yellow color that absorbs at 420 nm (Equations 11 and 12). This assay is normally carried out as an end-point assay. There is stoichiometry between the color formed and the amount of glucose in the extract, which is established with a standard curve. Because glucose oxidase is quite specific for glucose, it is a useful tool in determining the amount of glucose in the presence of other reducing sugars.

$$\beta\text{-D-glucose} + O_2 \xrightarrow{\text{glucose oxidase}} \delta\text{-gluconolactone} + H_2O_2 \quad (11)$$

$$H_2O_2 + \textit{o}\text{-dianisidine} \xrightarrow{\text{peroxidase}} H_2O + \text{oxidized dye (colored)} \quad (12)$$

2.5.1.5 Starch/Dextrin Content Starch and dextrins can be determined by enzymatic hydrolysis using amyloglucosidase, an enzyme that cleaves α-1,4 and α-1,6 bonds of starch, glycogen, and dextrins, liberating glucose. The glucose formed can be subsequently determined enzymatically. Glucose can be determined by the previously described colorimetric method, in which glucose is oxidized by glucose oxidase and coupled to

a colored dye via reaction of the glucose oxidase product (hydrogen peroxide) with peroxidase. An alternative method of measuring glucose is by coupling hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PDH) reactions (Equations 13 and 14):

$$\text{glucose} + \text{ATP (adenosine triphosphate)} \xrightarrow{\text{HK}} \text{glucose-6-phosphate} + \text{ADP (adenosine diphosphate)} \quad (13)$$

$$\text{glucose-6-phosphate} + \text{NADP}^+ \xrightarrow{\text{G6PD}} \text{6-phosphogluconate} + \text{NADPH} + \text{H}^+ \quad (14)$$

The amount of NADPH formed is measured by absorbance at 340 nm and is a stoichiometric measure of the glucose originating in the dextrin or starch hydrolyzed by amyloglucosidase. The amount of starch determined by this method is calculated using Equation (15):

$$c = V\text{MW} \times \Delta A_{340}\, \varepsilon b v 1000 \quad (15)$$

where c is the starch in sample solution (g L^{-1}), V is the volume (mL) of reaction mixture, MW is the molecular weight of starch (because this method measures glucose derived from starch use 162.1, $\text{MW}_{\text{glucose}} - \text{MW}_{\text{water}}$), ε is the absorption coefficient of NADPH at 340 nm ($6.3\ \text{L mmol}^{-1}\ \text{cm}^{-1}$), b is the light pathlength of cuvette (1 cm), v is the volume of sample (mL), and ΔA_{340} is the absorbance at 340 nm ($=A_{340,\text{sample}} - A_{340,\text{reagent blank}}$).

Note that HK catalyzes the phosphorylation of fructose as well as glucose. The determination of glucose is specific because of the specificity of the second reaction catalyzed by G6PDH, in which glucose-6-phosphate is the substrate (Equation 14).

This assay sequence can be used to detect the dextrins of corn syrup used to sweeten a fruit juice product. A second assay would be needed, however, without treatment with amyloglucosidase, to account for the glucose in the product. The glucose determined in that assay would be subtracted from the result of the assay in which amyloglucosidase is used.

The same HK, G6PDH sequence used to measure glucose can also be used to measure other carbohydrates in foods. For example, lactose and sucrose can be determined via specific hydrolysis of these disaccharides by β-galactosidase and invertase respectively, followed by the use of the earlier described HK, G6PDH sequence.

2.5.1.6 Determination of D-Malic Acid in Apple Juice

Two stereo isomeric forms of malic acid exist. L-Malic acid occurs naturally, while the D form is normally not found in nature. Synthetically produced malic acid is a mixture of these two isomers. Consequently, synthetic malic acid can be detected by a determination of D-malic acid. One means of detecting the malic acid is through the use of the enzyme decarboxylating D-malate dehydrogenase (DMD).[7] Decarboxylating DMD catalyzes the conversion of D-malic acid according to Equation (16):

$$\text{D-malic acid} + \text{NAD}^+ + \text{H}^+ \xrightarrow{\text{DMD}} \text{pyruvate} + \text{CO}_2 + \text{NADH} \quad (16)$$

The reaction can be followed by the measurement of NADH photometrically. Because CO_2 is a product of this reaction and escapes, the equilibrium of the reaction lies to the right and the process is irreversible. This assay is of value because the addition of synthetic D/L-malic acid can be used to increase the acid content of apple juice and apple juice products illegally.

ATP bioluminescence is based on the reaction catalyzed by the enzyme luciferase from the North American firefly (*Photinus pyralis*).[8] The chemical energy of ATP drives the luciferase-catalyzed oxidative decarboxylation of luciferin with the resultant emission of one photon of light for each ATP (Equation 17).

$$\text{luciferin} + \text{ATP} + \text{O}_2 \xrightarrow{\text{luciferase}} \text{decarboxyluciferin} + \text{CO}_2 + \text{AMP (adenosine monophosphate)} + \text{PPi (inorganic pyrophosphate)} + \text{light} \quad (17)$$

Thus, a measure of the amount of ATP present is achieved. Because living cells contain ATP this reaction can be used as a means of detecting biomass including microbes. The hygiene status of food contact surfaces as a measure of sanitation and cleaning is a critical quality assurance function, because microbes can be a direct risk. In addition, food product residues can provide nutrients that allow initially small numbers of microbes to grow, compromising the next batch of food processed through the equipment. Because ATP is present in living cells including the food product residue, ATP luminescence can be used to monitor the cleaning, the removal of food residue, and the sanitizing of food equipment to destroy microorganisms. Swab sampling of surfaces followed by ATP luminescence assay has been commercialized and the procedure can be completed in less than one minute to assess cleaning efficiency and remedial action can be taken. Microbial contamination of food product can also be measured by this reaction if ATP of the food itself is eliminated. The use of ATPase prior to extraction of microbial cell ATP and measuring the liberated ATP has been shown to be an effective means of sterility testing in fruit juice.[9]

2.5.2 Enzymes as Preparative Tools

Enzymes can be used in a number of ways to prepare food samples for analyses. A good example of the preparative use of enzymes in food analysis is in the determination of fiber in foods. Dietary fiber is made up of plant polysaccharides and lignin that are not digested by human enzymes. Dietary fiber can be estimated by selectively solubilizing digestible carbohydrates, lipids and proteins by a combination of chemical and enzyme treatment. Enzymes used for this purpose include amylases to solubilize starch and proteases to digest proteins. Undigested materials are collected, dried and the residue weighed, ashed and the weight of the ash subtracted from the residue weight. Alternatively, after solubilization of digestible materials by enzyme treatment, fiber components in the residue are acid hydrolyzed, the resulting monosaccharides are measured, and the sum of the monosaccharides in the hydrolysate represents fiber. An example of the latter approach is that of Englyst and Cummings.[10] In this method, samples are mixed with dimethylsulfoxide and heated to gelatinize starch. Starch and protein are digested with pancreatin (a pancreatic preparation that contains several enzymes, including amylase, proteases and lipase) and pullulanase, which cleaves α-1,6 linkages in starch. Starch hydrolysis is completed by adding amyloglucosidase. Undigested carbohydrate is precipitated by addition of ethanol. The insoluble residue (fiber) is collected by centrifugation, washed with ethanol and acetone, and dried. The residue is acid hydrolyzed and then neutral sugars are measured by gas chromatography and uronic acids (from pectin) are measured colorimetrically. A general problem encountered in the analysis of fiber by treatment of samples with polysaccharide-hydrolyzing enzymes, is that some microbial preparations (e.g. amyloglucosidase) may be contaminated with β-glucanase (cellulase) resulting in an underestimation of cellulose in the fiber.[11]

The polymerase chain reaction (PCR) is an established DNA amplification methodology that takes advantage of a heat-stable DNA polymerase as a preparative tool. The technique uses three steps: DNA preparation, DNA amplification, and DNA detection. In this process DNA is released from bacteria via lysis after a pre-enrichment step, a targeted DNA region is cyclically amplified with the polymerase using specific primers and the sample DNA is separated by electrophoresis and detected. PCR has the advantage of quickly and specifically identifying pathogenic microbes. For example, PCR using four sets of oligonucleotide primers for target genes for enterotoxigenic, enteroinvasive, enterohemorrhagic and total *Escherichia coli* has been used for detection of fewer than 10^2–10^3 cells per gram of these types of microbe in artificially contaminated ground beef.[12] The protocol, including pre-enrichment steps, takes less than 1.5 days, compared to enrichment and selective media methods that take 4–5 days.[13]

2.5.3 Enzyme Activity Assays

2.5.3.1 Peroxidase Activity Peroxidase is found in most plant materials and is reasonably stable to heat. A heat treatment that will destroy all peroxidase activity in a plant material is usually considered to be more than adequate to destroy other enzymes and most microbes present. In vegetable processing, therefore, the adequacy of the blanching process can be monitored by following the disappearance of peroxidase activity.[14] Peroxidase catalyzes the oxidation of guaiacol (colorless) in the presence of hydrogen peroxide to form tetraguaiacol (yellow brown) and water (Equation 18). Tetraguaiacol has an absorbance maximum around 450 nm. Increase in absorbance at 450 nm can be used to determine the activity of peroxidase in the reaction mixture.

$$H_2O_2 + \text{guaiacol} \xrightarrow{\text{peroxidase}} \underset{\text{(colored)}}{\text{tetraguaiacol}} + H_2O \qquad (18)$$

2.5.3.2 Lipoxygenase It has been pointed out that lipoxygenase may be a more appropriate enzyme to measure the adequacy of blanching of vegetables than peroxidase.[15] Lipoxygenase refers to a group of enzymes that catalyze the oxidation by molecular oxygen of fatty acids containing a *cis,cis*,1,4-pentadiene system, producing conjugated hydroperoxide derivatives (Equation 19):

$$(-CH{=}CH{-}CH_2{-}CH{=}CH-) + O_2 \xrightarrow{\text{lipoxygenase}} (-C(OOH){-}CH{=}CH{-}CH{=}CH-) \qquad (19)$$

(conjugated)

A variety of methods can be used to measure lipoxygenase activity in plant extracts. The reaction can be followed by measuring loss of fatty acid substrate, oxygen uptake, occurrence of the conjugated diene at 234 nm, or the oxidation of a cosubstrate such as carotene.[16] All these methods have been used and each has its advantages. The oxygen electrode method is widely used and replaces the more cumbersome manometric method. The electrode method is rapid and sensitive and gives continuous recording. It is normally the method of choice for crude extracts, but secondary reactions involving oxidation must be corrected for or eliminated. Zhang et al. have reported the adaptation of the O_2 electrode method for the assay of lipoxygenase in green bean homogenates without extraction.[17] Owing to the rapidity of the method (<3 min including the homogenization),

on-line process control, using lipoxygenase activity as a control parameter for optimization of blanching of green beans, is a real possibility. The formation of conjugated diene fatty acids with a chromophore at 234 nm can be followed continuously. However, optically clear mixtures are necessary. Bleaching of carotenoids has also been used as a measure of lipoxygenase activity. However, the stoichiometry of this method is uncertain and all lipoxygenases do not have equal carotenoid bleaching activity. Williams et al. have developed a semiquantitative spot test assay for lipoxygenase in which I^- is oxidized to I_2 in the presence of the linoleic acid hydroperoxide product and the I_2 is detected as an iodine–starch complex.[15]

2.5.3.3 Alkaline Phosphatase Assay Alkaline phosphatase is a relatively heat-stable enzyme found in raw milk. The thermal stability of alkaline phosphatase in milk is greater than the nonspore-forming microbial pathogens present in milk. The phosphatase assay is applied to dairy products to determine whether pasteurization has been done properly, and to detect the addition of raw milk to pasteurized milk. The common phosphatase test is based on the phosphatase-catalyzed hydrolysis of disodiumphenyl phosphate liberating phenol.[18] The phenol product is measured colorimetrically after reaction with CQC (2,6-dichloroquinonechloroimide) to form a blue indophenol. The indophenol is extracted into *n*-butanol and measured at 650 nm. This is an example of a physical separation of product to allow the ready measurement of an enzyme reaction. A fluorimetric assay has been suggested and has been commercialized for measurement of alkaline phosphatase, in which the rate of fluorophore production can be monitored directly without butanol extraction used to measure indophenol when phenylphosphate is used as substrate.[19] The fluorimetric assay was shown to give greater repeatability compared to the standard assay, in which phenylphosphate is used as substrate and was capable of detecting 0.05% raw milk in a pasteurized milk sample.

2.5.3.4 α-Amylase Activity Amylase activity in malt is a critical quality parameter. The amylase activity in malt is often referred to as diastatic power and refers to the production of reducing substances by the action of α- and β-amylases on starch. The measurement of diastatic power involves the digestion of soluble starch with a malt infusion (extract) and following the increase in reducing substances by measuring reduction of Fehling's solution or ferricyanide. Specifically measuring α-amylase activity (often referred to as dextrinizing activity) in malt is more complicated and is based on using a limit dextrin as substrate. Limit dextrin is prepared by action of β-amylase (free of α-amylase activity) on soluble starch. The β-amylase clips maltose units off the nonreducing end of the starch molecule until an α-1,6-branch point is encountered. The resulting product is a β-limit dextrin that serves as the substrate for the endo-cleaving α-amylase. A malt infusion is added to the previously prepared limit dextrin substrate and aliquots are removed periodically to a solution of dilute iodine. The α-amylase activity is measured by the changed color of the starch–iodine complex in the presence of excess β-amylase used to prepare the limit dextrin. The color is compared to a colored disc on a comparator. This is continued until the color is matched to a color on a comparator. The time to reach that color is the *dextrinizing time* and is a measure of α-amylase activity, a shorter time representing a more active preparation.

Because α-amylase is an endoenzyme, when it acts on a starch paste the viscosity of the paste is dramatically reduced, greatly influencing flour quality. Consequently, α-amylase activity is of great importance in whole wheat. Wheat normally has small amounts of α-amylase activity, but when wetted in the field, preharvest sprouting (pregermination) can occur, with a dramatic increase in α-amylase activity. Preharvest sprouting cannot be easily detected visually, so measurement of α-amylase activity can be used as a sensitive estimate of preharvest sprouting. The *falling number method* is a procedure in which ground wheat is heated with water to form a paste, and the time it takes for a plunger to fall through the paste is recorded.[20] Accordingly, the time in seconds (the falling number) is inversely related to the α-amylase activity and the degree of preharvest sprouting. This method of measuring enzyme activity is a good example of using change in physical property of a substrate as a means of estimation of enzyme activity.

2.5.3.5 Rennet Activity Rennet, an extract of bovine stomach, is used as a coagulating agent in cheese manufacture. Most rennet activity tests are based on noting the ability of a preparation to coagulate milk. For example, 12% nonfat dry milk is dispersed in a 10-mM calcium chloride solution and warmed to 35 °C. An aliquot of the rennet preparation is added and the time of milk clotting is observed visually. The activity of the preparation is calculated in relation to a standard rennet. As opposed to coagulation ability, rennet preparations can also be evaluated for proteolytic activity by measuring the release of a dye from azocasein (casein to which a dye has been covalently attached). In this assay, the rennet preparation is incubated with 1% azocasein. After the reaction period, the reaction is stopped by adding trichloroacetic acid. The trichloroacetic acid precipitates the protein that is not hydrolyzed. The small fragments of colored azocasein produced by the hydrolysis of the rennet are left in solution and absorbance is read at

345 nm.[21,22] This assay is based on the increase in solubility of a substrate upon cleavage by an enzyme.

2.5.3.6 Meat Cooking Indicators Concerns regarding the safety of food products has prompted the development of rapid accurate methods for verification of proper processing of food products. In the case of measuring microbiological safety of meat products, a number of enzymes endogenous to the meat have been examined as indicators of adequate cooking to inactivate food pathogens. Several bovine muscle enzymes have been studied as potential indicators of adequate heating.[23] This work compared the inactivation of the food pathogens *E. coli* O157 : H7 and *Salmonella senftenberg* in ground beef to endogenous muscle enzymes acid phosphatase, lactate dehydrogenase, peroxidase, phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenase, and triose-phosphate isomerase. The temperature dependence of the inactivation rate constant of the triose-phosphate isomerase was similar to those of the two pathogens, suggesting that this enzyme could be used as an endogenous time–temperature indicator in beef products.

2.5.4 Biosensors/Immobilized Enzymes

The use of immobilized enzymes as analytical tools is currently receiving increased attention. An immobilized enzyme in concert with a sensing device is an example of a biosensor. A biosensor is a device composed of a biological sensing element (e.g. enzyme, antibody, etc.) coupled to a suitable transducer (e.g. optical, electrochemical, etc.). Immobilized enzymes, because of their stability and ease of removal from the reaction, can be used repeatedly, thus eliminating a major cost in enzyme assays. The most widely used enzyme electrode is the glucose electrode in which glucose oxidase is combined with an oxygen electrode to determine glucose concentration.[24] When the electrode is put into a glucose solution, the glucose diffuses into the membrane where it is converted to gluconolactone by glucose oxidase with the uptake of oxygen. The oxygen uptake is a measure of the glucose concentration. Glucose can also be measured by the action of glucose oxidase with the detection of hydrogen peroxide, in which the hydrogen peroxide is detected amperometrically at a polarized electrode.[25] A large number of enzyme electrodes (biosensors) have been reported in the literature. For example, a glycerol sensor, in which glycerol dehydrogenase was immobilized, has been developed for the determination of glycerol in wine.[26] NADH produced by the enzyme was monitored with a platinum electrode. Immobilized glucose oxidase has been combined with glucoamylase and an oxygen electrode in a flow injection analysis (FIA) system for the evaluation of starch damage in wheat flour.[27] Starch damage is a measure of intensity of milling, and in some cases results in poor-quality finished products. Starch damage was estimated after pretreatment of flour slurries with fungal α-amylase for 15 min and the resulting filtrate was injected into the FIA system containing the immobilized enzymes. The method was demonstrated to give results in 20 min that agreed well with time-consuming traditional methods.

3 IMMUNOLOGY

3.1 Overview

Immunological methods are useful in food analysis because of their specificity and sensitivity. Immunochemical techniques can be used to analyze individual components to relate composition to food quality, to detect contaminants in foods and to monitor processing.

Immunological techniques are dependent on the interactions of antibodies with antigens. Antibodies are serum proteins (immunoglobulins) produced by animals in response to immunization with compounds (antigens) which are foreign to the animal. A wide range of compounds, including polysaccharides and proteins, can serve as antigens. Antibodies produced in response to an antigen have the ability specifically to bind with that antigen due to combining sites on the immunoglobulin with a portion of the antibody. That portion is the antigenic determinant. The specificity of combination between antibody and antigen is similar to that observed in other protein–ligand interactions such as enzyme–substrate.[28]

The size of antigenic determinants has been estimated to be no larger than six or seven sugar residues.[29] A protein would be expected to have many antigenic determinants, resulting in the production of a number of antibodies each with a specificity for a particular antigenic site on the protein. Haptens are small molecules (less than 1000 Da) that are unable to cause the production of antibodies when injected into an animal, and must be linked to a protein such as bovine serum albumin to elicit antibodies to that small molecule. Antibodies made against the conjugate will bind to hapten found in samples.

An immunochemical assay is a procedure in which the noncovalent binding between an antigen and antibody is used as the mode of specificity for the analysis. There are a number of ways to observe or monitor the antibody–antigen noncovalent interaction.

3.2 Methods

3.2.1 Agglutination

The ability of immunoglobulins to bind more than one ligand enables antibodies to agglutinate cells or other

particles that have multiple antigens on their surface. Agglutination can be used as an indirect technique for determining antigens. A series of dilutions of sample to be tested is made into tubes to which a fixed number of red blood cells or latex particles coated with antigen (or antibody) and antibody preparation are added. The particle-bound antigen competes with the antigen in the sample for the antibody. The settling pattern of the cells or particles is observed and antigen amount is estimated based on noting which sample dilution has caused agglutination in comparison to a series of known amounts of antigen.

3.2.2 Immunodiffusion

A number of immunochemical methods have been developed that use gels as media for the precipitation of antibody–antigen complexes. Agar or agarose gels are usually used for these methods, since the opaque antibody–antigen precipitate formed can be easily distinguished from the translucent background. Immunodiffusion methods are divided into single and double immunodiffusion techniques.

In most single immunodiffusion techniques, the antigen is allowed to diffuse into a gel in which antibodies have been dispersed. The antigen may be added to the top of a tube containing the gel and allowed to diffuse in a single dimension or it may be added to a well cut into the gel and allowed to diffuse in two dimensions. As the antigen encounters the antibody dispersed in the gel it will form a precipitin band. The precipitation is reversible in an excess of antigen and the precipitin band will appear to move through the gel as it progressively dissociates as more antigen diffuses into the precipitin area and reforms at the diffusing antigen front.

In double diffusion methods, antigens and antibodies are placed in different parts of the gel and allowed to diffuse toward one another. The diffusion can be done in one dimension by placing the antigen and antibody at opposite ends of a tube containing agar or in two dimensions by putting antigens and antibodies in separate wells cut into a layer of agar. The rates of diffusion of the antibody and antigen are dependent on their concentrations and diffusion coefficients, and precipitin bands of different antibody–antigen complexes can occur at different regions of the gel. Therefore, both single and double diffusion techniques have the ability to resolve mixtures of antigens.

Radial immunodiffusion is an example of a single immunodiffusion method that can be used to measure the amount of antigen in a solution and is based on the relationship of rate of diffusion of the antigen to antigen concentration. The antiserum made against the antigen is uniformly dispersed in an agar gel, and dilutions of a known amount of antigen to be measured are placed in wells of the gel. As the antigens diffuse into the gel, precipitin rings are formed around the wells. The diameter of the rings will increase with time due to the continuous dissolution and reformation of precipitate in response to the changing ratio of antibody to antigen. After an appropriate time, the diameters of the precipitin rings are measured and the diameters of the rings around wells of known antigen concentration are plotted as a function of the log of the antigen concentration. This curve is used to estimate the concentrations of the unknown solutions from the diameter of their precipitin rings.

The Ouchterlony method is a double diffusion method in two dimensions and is widely used to determine the relationship of antigens to one another. In this method, the antiserum is added to a well in the center of an agar gel plate and antigen solutions are placed in surrounding wells equidistant from the antiserum well. If two adjacent wells have the identical antigen, the precipitin bands formed between the two wells will fuse into a continuous band of identity. If a mixture of antigens is placed in a well and allowed to diffuse toward a well containing antibodies made against each antigen, several precipitin bands will form due to the independent precipitation of each antibody–antigen complex. If two different antigens are placed in adjacent wells, and allowed to diffuse toward the antibody mixture, the precipitin bands will form a cross in a reaction of nonidentity. If two antigens that share similar or common determinants are placed in adjacent wells they will cross-react and give a reaction of partial identity in which there is a partial fusion of precipitin bands and the formation of a spur extending toward the antigen well with the heterologous antigen. This double diffusion method provides a means of screening samples for multiple antigens and of determining their relationships with other antigens, and is of use in food analysis for determining the origin of food components.

3.2.3 Radioimmunoassay

Radioimmunoassays (RIAs) are methods that use an isotopic label on the antigen or antibody as a means of measuring the antibody–antigen complex. The sensitivity of this type of assay is limited primarily by the ability to prepare antibodies or antigen with high specific radioactivity and by the binding constant of the antibody to the antigen.

Most RIAs use labeled antigen and are based on the competition for antibody between radiolabeled antigen and antigen in the sample. As the amount of unlabeled antigen in the sample increases, a smaller amount of radioactivity is bound in the antibody–antigen complex. Thus, after interaction is allowed to take place the antibody–antigen complex is separated from free

antigen and the amount of bound or free radioactivity is determined. By using known amounts of unlabeled antigen, it is possible to construct standard curves by plotting the ratio of bound to free radioactive antigen or the percentage inhibition of bound radioactive antigen as a function of antigen concentration.

3.2.4 Enzyme-linked Immunosorbent Assay

Enzyme-linked immunosorbent assay (ELISA) is a solid phase assay in which a reactant (either the antigen or antibody) is adsorbed onto the surface of a test tube or microtiter well, and one of the reacting species (most commonly the antibody) is conjugated with an enzyme. Enzyme immunoassay techniques use the enzymatic conversion of substrate to a colored product as a way of measuring the amount of antigen present. A major advantage of enzyme immunoassay, compared to RIA, is the lack of a requirement to use radioactive compounds. In addition, the sensitivity of an enzyme immunoassay can be greater than RIAs because of the catalytic amplification generated by bound enzyme.

3.3 Applications

Applications of immunological methods to food analysis have been largely focused around food safety issues including the detection or measurement of microbes and microbial toxins, pesticide and drug residue analysis, detection of food allergens, estimation of adequate processing for microbial inactivation, and (to a lesser extent), detection of adulteration. The application of immunological techniques to the analysis of food has included the detection of food pathogens and toxins produced by these microbes. Serotyping using immunodiffusion methods has long played a role in identifying bacterial strains and toxins.

The rapid assay of *Salmonella* species has been accomplished through the use of selective enrichment media and ELISA.[30,31] These methods allow the detection of *Salmonella* in chicken at less than one microbe per 25 g of chicken sample. The use of latex agglutination in food systems for pathogen detection has been reported by Matar et al.[32] These workers describe the rapid detection of the pathogen *Listeria monocytogenes* by using antibody to the *L. monocytogenes* toxin listeriolysin O, coated onto polystyrene-modified latex beads. The method is sensitive and rapid (48 h), whereas culture techniques require more than 4 days.

A major application of immunoassays in foods has been the detection of microbial toxins. A significant effort has been directed at the rapid detection of mycotoxins in foods. Mycotoxins are fungal metabolites that are produced by many species including *Aspergillus, Penicillium* and *Fusarium*, and can be found in a number of food materials, including corn, wheat, barley, peanuts, apples, cottonseed, and milk (due to consumption of contaminated feed). Mycotoxins can produce a variety of symptoms in humans and animals, including cancer. Consequently, strict regulations have been established for several of these toxins. Methods of mycotoxin analysis include physical methods such as thin-layer chromatography, gas chromatography and mass spectrometry. Because of interferences in many food matrices, extensive sample clean-up is needed to use these techniques, which are time-consuming and costly.

An alternative to the separation and analysis of mycotoxins is the application of immunological methods. Because mycotoxins are small molecules, they are first conjugated to a carrier protein, such as bovine serum albumin, to be made immunogenic. Antibodies have been made against the major mycotoxins, and both RIAs and ELISA techniques have been successfully applied and commercialized.[33] One particularly interesting method has been described in which antibodies specific for several different mycotoxins, aflatoxin B_1, fumonisin B_1 and zearalenone, were immobilized in differentiated lines on nitrocellulose membranes, and a competitive assay was performed in which free mycotoxins from the sample compete with enzyme-labeled mycotoxin for the immobilized antibodies. The amount of color from the bound enzyme is inversely related to the mycotoxin in the sample, and the intensity is assessed using a camera, video monitor and microcomputer.[34]

Contamination of shellfish with paralytic shellfish poisoning (PSP) toxins has little effect on the shellfish; but numbness, tingling, and muscular paralysis or death can result in humans who ingest these toxins. PSP toxin analysis is an example in which a bioassay using a whole animal has been used but can be replaced by an immunoassay. The whole animal assay is simply a measurement of the death of a mouse after intraperitoneal injection of a sample extract. A commercially available competitive ELISA test for PSP toxins using saxitoxin–enzyme conjugate has been systematically tested in mussels and scallops and when compared to the mouse bioassay was found to be 150 times more sensitive.[35]

The detection of economic adulteration of several animal products has been successfully addressed by using immunological methods. The detection of cows' milk at $<1\%$ has been accomplished in goats' and ewes' milk and cheeses through the use of competitive ELISA, using antibodies to bovine caseins.[36,37] Similarly, an antiserum against an autoclaved porcine muscle extract was successfully used in a competitive ELISA to detect $<1\%$ of pork in cooked beef and mutton.[38]

Food allergies involve abnormal immune responses by certain individuals to specific food proteins.[39] A relatively small number of food sources (peanuts, soybean, milk, tree nuts, fish, crustacea, molluscs and wheat) are responsible for most food allergies. Because avoidance diets are the principal means by which food allergic individuals prevent adverse reactions, the individual must depend on proper labeling and thus depend on the ability of food processors to detect trace amounts of food allergens reliably in food products.[28] A challenge for the food industry is that, in cleaning between products, a residue of allergenic food might be left on processing equipment and be carried into a food product in which the allergenic food is not listed on the food label. Thus, a rapid quality assurance method of detection of allergenic proteins is critical. Mills et al. have devised a dipstick method for the detection of peanut, in which antibody made against conarachin, a major storage protein found in peanuts, is used.[40] Antibodies to purified conarachin are bound to the surface of a paddle-type dipstick, which when immersed into a liquid sample binds conarachin. The dipstick is rinsed and biotinylated anticonarachin is added to ''sandwich'' the captured peanut protein. The dipstick is rinsed again and immersed into a solution containing avidin–peroxidase conjugate. Avidin strongly binds to the biotin-labeled antibody. Finally the dipstick is rinsed, peroxidase substrate added, and a positive sample found by the appearance of the colored product, which absorbs at 450 nm. The dipstick assay detected as little as 0.1% peanut contamination in chocolate with an assay time of <40 min using food samples extracted in aqueous buffer and clarified by centrifugation.

3.3.1 Biosensors Based on Immunological Interaction

Biosensors based on the antibody–antigen interaction have been developed and applied to the analysis of a number of food constituents. One type of immunological biosensor that has been commercialized takes advantage of surface plasmon resonance (SPR), which allows direct analysis without labeling of antigen or antibody. SPR occurs when surface plasmon rays are excited at a metal–liquid interface. Light is directed at the surface and reflected from the side not in contact with a liquid sample. The surface in contact with the liquid sample has either antibody or antigen attached, and the formation of the antibody–antigen complex is detected as a change in the SPR signal. This type of biosensor has been used for detection of the pathogen *E. coli* O157 : H7 and antibiotics in milk, as well as measurement of the vitamins folic acid and biotin.[41–44] This technology offers specific rapid sensitive analyses and is expected to play a role in food quality assurance systems by providing real-time data in a processing environment.

4 ANALYTICAL MICROBIOLOGY

4.1 Overview

It is possible to use microbes to measure many substances that are essential nutrients or inhibitors of microbial metabolic processes. Microbiological assays for the measurement of nutrients are based on the dependence of microbial growth on the presence of limiting amounts of the nutrient being measured. The fundamental methodology in many microbiological assays is similar: addition of the test substance (sample) and growth stimulation or inhibition is measured. The growth response can be measured in several ways including acid production, oxygen uptake, carbon dioxide production, optical density, turbidity, weight, area of colony, or inhibition of spore germination.

An ideal microbe for use in analytical assays should be easily cultivated, nonpathogenic, sensitive to the nutrient to be measured, grow rapidly, and have a readily measurable growth response. Bacteria are most often used as the test organism in microbial assays because of their rapid growth and ease of cultivation.[45]

4.2 Applications

4.2.1 Vitamin Assays

A number of water-soluble vitamins can be readily measured by using the organisms listed in Table 2.[46] In general, these methods entail an aqueous extraction of the food, autoclaving and clarification. The sample and media containing all nutrients except the vitamin being measured are mixed, followed by aseptic inoculation with the appropriate microorganism. Parallel standards containing the limiting vitamin are assayed to prepare a standard

Table 2 Water-soluble vitamins that can be measured by analytical microbiology[46]

Analyte	Microorganism	AOAC method
Cobalamin (B_{12})	*Lactobacillus leichmannii* (ATCC No. 7830)	952.20
Folic acid	*Streptococcus faecalis* (ATCC No. 8043)	944.12
Pantothenic acid	*Lactobacillus plantarum* (ATCC No. 8014)	945.74
Riboflavin (B_2)	*Lactobacillus casei* subsp. *rhamnosus* (ATCC No. 7469)	940.33
Niacin/niacinamide	*Lactobacillus plantarum* (ATCC No. 8014)	944.13
Pyridoxine, pyridoxal, pyridoxamine (B_6)	*Saccharomyces uvarum* (ATCC No. 9080)	961.15

AOAC = Association of Official Analytical Chemists.

curve. After an incubation period, the determination of growth response is made by titration of the acid produced (72-h incubation) or turbidity of the media (16–24-h incubation). An exception to the above scheme is the measurement of the B_6 vitamers. In this case, samples are autoclaved in aqueous HCl and the vitamers separated by cation-exchange chromatography and stepwise elution prior to inoculation of fractions with *Saccharomyces uvarum* and measurement of growth response by turbidimetry. Another approach to assay of vitamins using microbes is to add radiolabeled glucose to the media and measure growth by monitoring the amount of $^{14}CO_2$.[47] Thiamine, pyridoxine, and pantothenic acid analysis has been successfully carried out using this method.

4.2.2 Antibiotic Assays

Antibiotics are used to control diseases in cattle and these antibiotics may enter the food supply via milk or meat. The presence of antibiotics in the food supply is of concern because these compounds can cause allergic reactions, might create an environment favorable to the development of antibiotic resistance in microbes and, in the case of milk, affect starter cultures in the manufacture of fermented dairy products. Approximately 5–10% of the population is hypersensitive to penicillin or other antibiotics and suffers allergic reactions (skin rashes, hives, asthma, anaphylactic shock) at concentrations as low as 1-ppb penicillin. There is also concern that small amounts of antimicrobial agents may significantly shift the resistance patterns in the microbes found in the gastrointestinal tract of man. Antibiotics are also known to interfere with the manufacture of several fermented dairy products. Concentrations of 1 ppb delay starter activity for cheese and yogurt. Antibiotics also decrease the acid and flavor production associated with butter manufacture, reduce the curdling of milk, and cause improper ripening of cheeses.

Testing methods for antibiotic residues in animal products have received significant attention from food industry regulatory agencies and animal production agriculture, and have included use of microbial inhibition assays as well as enzyme inhibition and immunological assays.

The use of *Bacillus stearothermophilus* var. *calidolactis* spore suspensions has been extensively used for testing for antibiotics in milk.[48] In one method, a small paper disc is used to wick up a milk sample by capillary action, and the wetted disc is placed on an agar media surface previously seeded with *Bacillus* spores. The sample is incubated at $64 \pm 2\,^{\circ}C$ for ca. 2.5 h during which time well-defined zones of clearing around a positive standard control disc to which penicillin has been added will be seen. The sample is compared to the positive standard. This test can be extended to include treatment of the milk sample with penicillinase followed by the disc assay, allowing identification of samples that contain penicillin or penicillin in combination with nonpenicillin inhibitory substances.

Another method also utilizes *Bacillus stearothermophilus* var *calidolactis*, in which a nutrient mixture including bromcresol purple is incubated with a milk sample for 2.5–3 h at 65 °C.[49] If no antibiotic is present, acid produced by the microbe causes the bromcresol purple to turn yellow. The presence of an antibiotic prevents acid formation and the purple color of the indicator remains unchanged.

A more equipment-intensive alternative for β-lactam antibiotic analysis is based on the irreversible binding of β-lactams to the cell wall of a microorganism (*B. stearothermophilus*).[50] A suspension of microbial cells and radiolabeled penicillin are mixed with the sample. If the sample has β-lactam antibiotic present, it will compete with the labeled penicillin for binding sites on the bacterial cell wall. The amount of ^{14}C bound is measured using a gas counter and the reduction of the counts is related to the amount of competing antibiotic in the sample. Counts lower than the control are considered positive. An extension of this analysis is the use of two microorganisms and several radiolabeled antibiotics (both ^{14}C and ^{3}H).[51] The combination of more than one microbe binder and two radiolabels allows screening of seven antibiotic types, including tetracycline, penicillin, erythromycin, streptomycin, sulfamethazine, chloramphenicol, and novobiocin. The method can screen for all seven of the antibiotics in 15 min but requires a scintillation counter.

4.2.3 Protein Quality

Protein quality is often assayed using an expensive 28-day rat feeding trial. Microbiological evaluation of protein quality can reduce cost and assay time. Protozoans have received a great deal of attention because they are capable of using insoluble proteins. The heterotrophic protozoan *Terahymena pyriformis* has been extensively explored for use as a test organism for the measurement of protein quality evaluation. *T. pyriformis* has the same requirement for essential amino acids as for normal rat growth and the ability to digest proteins; thus, growing cultures can degrade intact proteins and the ability of *T. pyriformis* W to utilize proteins for growth is similar to the growing rat.[52] The technique for biological evaluation of protein with *Tetrahymena* has been established.[53] Growth of the organism was originally done by counting cell numbers using a hemocytometer with graduated rulings. This method of measuring growth is tedious and is a limitation, and automatic cell counters have been used.[54,55]

ABBREVIATIONS AND ACRONYMS

ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
CQC	2,6-Dichloroquinonechloroimide
DMD	D-Malate Dehydrogenase
ELISA	Enzyme-linked Immunosorbent Assay
FIA	Flow Injection Analysis
G6PDH	Glucose-6-phosphate Dehydrogenase
HK	Hexokinase
INT	Iodonitrotetrazolium Chloride
NAD	Nicotinamide Adenine Dinucleotide
NAD^+	Nicotinamide Adenine Dinucleotide (Oxidized Form)
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate (Oxidized Form)
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
PCR	Polymerase Chain Reaction
PPi	Inorganic Pyrophosphate
PSP	Paralytic Shellfish Poisoning
PVPP	Polyvinylpolypyrrolidone
RIA	Radioimmunoassay
SO	Sulfite Oxidase
SPR	Surface Plasmon Resonance

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis • Monosaccharides and Sugar Alcohol Analysis

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Glucose Measurement • Immunochemistry • Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Food **(Volume 5)**
Dietary Fiber Analysis as Non-starch Polysaccharides • Pesticides, Mycotoxins and Residues Analysis in Food • Starch Analysis in Food • Vitamins Analysis in Food

Nucleic Acids Structure and Mapping **(Volume 6)**
Polymerase Chain Reaction and Other Amplification Systems

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Antibiotics, Pharmaceutical Analysis of

Kinetic Determinations **(Volume 12)**
Enzymatic Kinetic Determinations

REFERENCES

1. P. Bernfeld, 'Amylases, Alpha and Beta', *Methods Enzymol.*, **1**, 149–158 (1955).
2. Boehringer-Mannheim, *Methods of Biochemical Analysis and Food Analysis*, Boehringer-Mannheim GmbH, Germany, 1987.
3. J.R. Whitaker, 'Analytical Uses of Enzymes', in *Food Analysis. Principles and Techniques, Volume 3, Biological Techniques*, eds. D. Gruenwedel, J.R. Whitaker, Marcel Dekker, New York, 297–377, 1985.
4. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, Academic Press, New York, 1983.
5. H.O. Beutler, 'A New Enzymatic Method for Determination of Sulfite in Food', *Food Chem.*, **15**, 157–164 (1984).
6. E. Raabo, T.C. Terkildsen, 'On the Enzymatic Determination of Blood Glucose', *Scand. J. Clin. Lab. Invest.*, **12**, 402–407 (1960).
7. H.O. Beutler, H.B. Wurst, 'A New Method for the Enzymatic Determination of D-Malic Acid in Foodstuffs. Part I: Principles of the Enzymatic Reaction', *Dtsch. Lebensm.-Rundsch.*, **86**, 341–344 (1990).
8. A.L. Kyriakides, P.D. Patel, 'Luminescence Techniques for Microbiological Analysis of Foods', in *Rapid Analysis Techniques in Food Microbiology*, ed. P.D. Patel, Blackie Academic & Professional, London, 196–231, 1994.
9. J.G.H.M. Vossen, H.D.K.J. Vanstaen, 'A New Method for the Rapid Detection of Microbial Contamination of Fruit Juices', in *Recent Developments in Food Analysis*, eds. W. Baltes, P.B. Czedik-Eysenberg, W. Pfannhauser, Verlag Chemie, Weinheim, 229–233, 1981.
10. H.N. Englyst, J.H. Cummings, 'Non-starch Polysaccharides (Dietary Fiber) and Resistant Starch', in *New Developments in Dietary Fiber*, eds. I. Furda, C.J. Brine, Plenum Press, New York, 205–225, 1990.
11. L. Prosky, 'Dietary Fiber', *J. AOAC Int.*, **81**, 151–152 (1998).
12. P.W. Lett, J.P. Southworth, D.D. Jones, A.K. Bej, 'Detection of Pathogenic *Escherichia coli* in Ground Beef Using Multiplex PCR', *Food Testing Anal.*, **1**, 34–38 (1995).
13. A.D. Hitchins, P.A. Hartman, E.C.D. Todd, 'Coliforms – *Escherichia coli* and Its Toxins', in *Compendium of Methods for the Microbiological Examination of Foods*, eds. C. Vanderzant, D.F. Splittstoesser, American Public Health Association, Washington DC, 325–369, 1992.
14. USDA, Enzyme Inactivation Tests (frozen vegetables), Technical Inspection Procedures for the Use of USDA

Inspectors, Agricultural Marketing Service, US Department of Agriculture, Washington DC, 1975.
15. D.C. Williams, M.H. Lim, A.O. Chen, R.M. Pangborn, J.R. Whitaker, 'Blanching of Vegetables for Freezing – Which Indicator Enzyme to Use', *Food Technol.*, **40**, 130–140 (1986).
16. K. Surrey, 'Spectrophotometric Method for Determination of Lipoxidase Activity', *Plant Physiol.*, **39**, 65–70 (1964).
17. Q. Zhang, R.P. Cavalieri, J.R. Powers, J. Wu, 'Measurement of Lipoxygenase Activity in Homogenized Green Bean Tissue', *J. Food Sci.*, **56**, 719–722 (1991).
18. G.K. Murthy, D.H. Kleyn, T. Richardson, R.M. Rocco, 'Phosphatase Methods', in *Standard Methods for the Examination of Dairy Products*, 16th edition, ed. G.H. Richardson, American Public Health Association, Washington DC, 413, 1992.
19. R. Rocco, 'Fluorometric Determination of Alkaline Phosphatase in Fluid Dairy Products: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **73**, 842–849 (1990).
20. AACC, *Approved Methods of Analysis*, 9th edition, American Association of Cereal Chemists, St. Paul, MN, 1995.
21. G.L. Christen, R.T. Marshall, 'Selected Properties of Lipase and Protease of *Pseudomonas fluorescens* 27 Produced in 4 Media', *J. Dairy Sci.*, **67**, 1680–1687 (1984).
22. S.M. Kim, J.F. Zayas, 'Comparative Quality Characteristics of Chymosin Extracts Obtained by Ultrasound Treatment', *J. Food Sci.*, **56**, 406 (1991).
23. A. OrtaRamirez, J.F. Price, Y.C. Hsu, G.J. Veeramuthu, J.S. Cherry Merritt, D.M. Smith, 'Thermal Inactivation of *Escherichia coli* O157 : H7, *Salmonella senftenberg*, and Enzymes with Potential as Time–Temperature Indicators in Ground Beef', *J. Food Prot.*, **60**, 471–475 (1997).
24. G.G. Guilbault, G.J. Lubrano, 'Enzyme Electrode for Glucose', *Anal. Chim. Acta*, **60**, 254–256 (1972).
25. Y. Shimizu, K. Morita, 'Microhole Array Electrode as a Glucose Sensor', *Anal. Chem.*, **62**, 1498–1501 (1990).
26. K. Matsumoto, 'Simultaneous Determination of Multicomponent in Food by Amperometric FIA with Immobilized Enzyme Reactions in a Parallel Configuration', in *Flow Injection Analysis (FIA) Based on Enzymes or Antibodies*, ed. R.D. Schmid, GBF Monographs, 193, Vol. 14, 1990.
27. R. Haginoya, K. Sakai, T. Komatsu, S. Nagao, K. Yokoyama, T. Takeuchi, R. Matsukawa, I. Karube, 'Determination of Damaged Starch and Diastatic Activity in Wheat Flour Using a Flow-injection Analysis Biosensor Method', *Cereal Chem.*, **74**(6), 745–749 (1997).
28. S.L. Hefle, 'Immunoassay Fundamentals', *Food Technol.*, **49**, 102–107 (1995).
29. E.A. Kabat, 'The Nature of an Antigenic Determinant', *J. Immunol.*, **97**, 1–11 (1966).
30. H.Z. Tian, T. Miyamoto, T. Okabe, Y. Kuramitsu, K. Honjoh, S. Hatano, 'Rapid Detection of *Salmonella* spp. in Foods by Combination of a New Selective Enrichment and a Sandwich ELISA Using Two Monoclonal Antibodies Against Dulcitol 1-Phosphate Dehydrogenase', *J. Food Prot.*, **59**, 1158–1163 (1996).
31. L.S. Van-Poucke, 'Salmonella-TEK, a Rapid Screening Method for *Salmonella* Species in Food', *Appl. Environ. Microbiol.*, **56**, 924–927 (1990).
32. G.M. Matar, P.S. Hayes, W.F. Bibb, B. Swaminathan, 'Listeriolysin O-Based Latex Agglutination Test for the Rapid Detection of *Listeria monocytogenes* in Foods', *J. Food Prot.*, **60**, 1038–1040 (1997).
33. J.J. Pestka, M.N. Abouzied, Sutikno, 'Immunological Assays for Mycotoxin Detection', *Food Technol.*, **49**, 120–128 (1995).
34. M.N. Abouzied, J.J. Pestka, 'Simultaneous Screening of Fumonisin B1, Aflatoxin B1, and Zearalenone by Line Immunoblot: A Computer Assisted Multianalyte Assay system (CAMAS)', *J. AOAC Int.*, **77**, 495–501 (1994).
35. E. Usleber, M. Donald, M. Straka, E. Martlbauer, 'Comparison of Enzyme Immunoassay and Mouse Bioassay for Determining Paralytic Shellfish Poisoning Toxins in Shellfish', *Food Addit. Contam.*, **14**, 193–198 (1997).
36. W. Richter, I. Krause, C. Graf, I. Sperrer, C. Schwarzer, H. Klostermeyer, 'An Indirect Competitive ELISA for the Detection of Cows' Milk and Caseinate in Goats' and Ewes' Milk and Cheese Using Polyclonal Antibodies Against Bovine Gamma Caseins', *Z. Lebensm.-Unters. Forsch.*, **204**, 21–27 (1997).
37. G. Anguita, R. Martihn, T. Garcia, P. Morales, A.I. Haza, L. Gonzales, B. Sanz, P.E. Hernandez, 'A Competitive Enzyme-linked Immunosorbent Assay for Detection of Bovine Milk in Bovine and Caprine Milk and Cheese Using a Monoclonal Antibody Against Bovine Beta-casein', *J. Food Prot.*, **60**, 64–66 (1997).
38. W.N. Sawaya, 'Detection of Pork in Processed Meat by an Enzyme-linked Immunosorbent Assay Using Antiswine Antisera', *J. Food Sci.*, **55**, 293–297 (1990).
39. J.A. Nordlee, S.L. Taylor, 'Immunological Analysis of Food Allergens and Other Food Proteins', *Food Technol.*, **49**, 129–132 (1995).
40. E.N.C. Mills, A. Potts, G.W. Plumb, N. Lambert, M.R.A. Morgan, 'Development of a Rapid Dipstick Immunoassay for the Detection of Peanut Contamination of Food', *Food Agric. Immunol.*, **9**, 37–50 (1997).
41. M. Medina, 'SPR Biosensor: Food Science Applications', *Food Testing Anal.*, **3**, 14–16, 34 (1997).
42. M. Medina, 'Hygromycin B Antibody Production and Characterization by a Surface Plasmon Resonance Biosensor', *J. Agric. Food Chem.*, **45**, 389–394 (1997).
43. C. Mellgren, A. Sternesjo, P. Hammer, G. Suhren, L. Bjorck, W. Heeschen, 'Comparison of Biosensor, Microbiological, Immunochemical and Physical Methods for Detection of Sulfamethazine Residues in Raw Milk', *J. Food Prot.*, **59**, 1223–1226 (1996).
44. BIAcore, *BiaQuant for Rapid Vitamin Analysis*, BIAcore, Piscataway, NJ, 1997.

45. W. Hewitt, *Microbiological Assay: An Introduction to Quantitative Principles and Evaluation*, Academic Press, New York, 1977.
46. AOAC, *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th edition, Arlington, VA, 1990.
47. M.N. Voigt, R.R. Eitenmiller, 'A Liquid Scintillation Technique for Microbial Vitamin Analysis', *J. Food Sci.*, **44**, 1780–1781 (1979).
48. J.W. Messer, J.E. Leslie, G.A. Houghtby, J.T. Peeler, J.E. Barnett, '*Bacillus stearothermophilus* Disc Assay for Detection of Inhibitors in Milk: Collaborative Study', *J. AOAC*, **65**, 1208–1214 (1982).
49. W.N. Kelley, 'Qualitative Ampule and Multitest for Beta-lactam Residues in Fluid Milk Products: Collaborative Study', *J. AOAC*, **65**, 1193–1207 (1982).
50. S.E. Charm, R.K. Chi, 'Rapid Screening Assay for Beta-lactam Antibiotics in Milk: Collaborative Study', *J. AOAC*, **65**, 1186–1192 (1982).
51. S.E. Charm, R.K. Chi, 'Microbial Receptor Assay for Rapid Detection and Identification of Seven Families of Antimicrobial in Milk: Collaborative Study', *J. AOAC*, **71**, 304–316 (1988).
52. D.L. Hill, *The Biochemistry and Physiology of Tetrahymena*, Academic Press, New York, 1972.
53. J.A. Stott, H. Smith, G.D. Rosen, 'Microbiological Evaluation of Protein Quality with *Tetrahymena pyriformis* W. 3. A Simplified Assay Procedure', *Br. J. Nutr.*, **17**, 227–235 (1963).
54. D.J. Teunisson, 'Elutriation and Coulter Counts of *Tetrahymena pyriformis* Grown in Peanut and Cottonseed Media', *Appl. Microbiol.*, **21**, 878–887 (1971).
55. G.M. Evancho, H.D. Hurt, P.A. Devlin, R.E. Landers, D.H. Ashotn, 'Comparison of *Tetrahymena pyriformis* W and Rat Bioassays for the Determination of Protein Quality', *J. Food Sci.*, **42**, 444–448 (1977).

Flavor Analysis in Food

Yong Chen and Chi-Tang Ho
Rutgers University, New Brunswick, USA

1 Introduction **2**
2 Isolation and Concentration Techniques **2**
2.1 Direct Solvent Extraction of Aqueous Samples 2
2.2 Steam Distillation of Samples Followed by Solvent Extraction 3
2.3 Simultaneous Steam Distillation and Extraction 3
2.4 High-vacuum Distillation of Lipid Samples 4
2.5 Direct Solvent Extraction of Solid Samples 4
2.6 Extraction with Liquid or Supercritical Carbon Dioxide 4
2.7 Headspace Analysis 5
2.8 Direct Thermal Desorption Methods 6
2.9 Solid-phase Microextraction 7
3 Separation Techniques **8**
3.1 Pre-gas Chromatography Separation 8
3.2 Gas Chromatography Separations 10
4 Identification of Flavor Compounds **12**
4.1 Gas Chromatography Retention Indices 12
4.2 Mass Spectrometry 12
4.3 Fourier Transform Infrared Spectroscopy 13
5 Quantification of Volatiles **13**
6 Gas Chromatography/Olfactometry Analysis **13**
Abbreviations and Acronyms **15**
Related Articles **15**
References **16**

Flavor compounds impart food aroma and taste, which are the major factors influencing consumer's perception of the quality of food. Understanding the nature of flavor compounds and their effects on human organoleptic responses is important and flavor analysis is an inseparable part of the process.

Analysis of flavor compounds, however, is complicated by several factors. While taste can basically be categorized as being salty, sour, bitter and sweet, the sensations caused by aroma compounds are much more complex. This results from the presence of numerous volatiles in foods and the ability of humans to differentiate these odorous components. The number of volatile and semivolatile compounds in foods ranges from 50 to 250 compounds in fresh fruits and vegetables to more than double these numbers in thermally processed foods such as roast beef and coffee. The odor-active compounds are usually found in very low concentrations, ranging from a few hundred parts per million (ppm) for strongly flavored food products to less than 10 ppm for weakly flavored foods. In addition, many of these flavor compounds are highly reactive and thermally labile. All of these aspects make flavor analysis a challenging task. Various methods have been developed to cope with the complexity and to resolve practical problems related to food flavors. The first few steps usually involve isolation and concentration of volatiles and semivolatiles from their original food matrix. A number of different techniques, ranging from conventional solvent extraction and distillation to the newly developed direct thermal desorption (DTD) and solid-phase microextraction (SPME) are available for isolating flavors from diverse food systems. Each of these techniques has its own applications with advantages and limitations. No single method will give a flavor isolate truly representative of the flavor profile in the food. Method selection should be based on volatility, thermal stability and reactivity of the flavor compounds and their associated food system. Thereafter separation of the concentrated flavor extract is traditionally conducted by gas chromatography (GC). The combined use of GC with other analytical instruments increases the capacity of basic GC separation. If applicable, pre-GC treatments of flavor extract using liquid chromatography, acid–base separation and chemical trapping of certain volatiles should also be employed to enhance the performance of GC separation. A cyclodextrin-based chiral stationary phase (CSP) is the first choice for separating enantiomers. The follow-up identification of flavor compounds can be carried out by integrating data about retention index (RI) and mass spectrometry (MS) of specific components. However, information obtained by gas chromatography/mass spectrometry (GC/MS) reveals little insight into the actual flavor present in a food system, because many volatiles detected instrumentally contribute little or nothing to the sensory character of a food product. Sensory techniques, coupled with a technique known as gas chromatography/olfactometry (GCO) make it possible to elucidate the relative impacts of various volatiles on the flavor characteristic of a food. CharmAnalysis and aroma extract dilution analysis (AEDA) are two popular methods to measure the potency of flavor components. However, these two methods are inconsistent with psychophysical views. An alternative method named Osme overcomes this problem. It is a quantitative bioassay method used to measure the response to odorants on a time–intensity scale. To

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

quantify volatile compounds, the traditional internal standard method is frequently used, but it is crude and often referred to as ''semiquantitative'' owing to its inherent limitations. For accurate quantification, a stable isotope dilution assay (IDA) should be used to target specific compounds. This article intends to describe briefly these analytical techniques and discuss their advantages and disadvantages and their applications.

1 INTRODUCTION

Flavor represents a major attribute of food and a deciding factor in the public's measurement of the quality of food. The analysis of flavor is complicated by several factors. The number of flavor compounds in food ranges from 50 to 250 compounds in fresh fruits and vegetables to more than double these numbers in processed foods such as roast beef and coffee. The odor-active compounds are usually found in very low concentrations, ranging from a few hundred ppm for strongly flavored food products to less than 10 ppm for weakly flavored foods.[1] In addition, many of the flavor compounds are highly reactive and thermally labile. All these factors make the analysis of flavor a great challenge in analytical chemistry. This section briefly reviews the common methodology used in the isolation and characterization of flavor compounds in foods. Several monographs have been published on this subject.[2–5]

2 ISOLATION AND CONCENTRATION TECHNIQUES

Flavor components in food are generally very complex, ranging from 50 to 250 chemicals and even more in thermally processed items. Yet some of the key flavor compounds might exist at extremely low levels beyond the detection threshold of available instruments. On the other hand, human olfactory organs are exceptionally sensitive to odor and may detect as little as 10^{-19} mol of an odorant. It is necessary to isolate and concentrate the analytes from the origin matrix before any analytical process. Great care should be taken before and during isolation and concentration. Flavor compounds differ greatly in physical and chemical properties. Some are very volatile whereas others such as vanillin have a very high boiling point. Some flavors are thermally labile and vulnerable to thermal degradation to produce artifacts. Some are extremely unstable or reactive when separated from food and concentrated for analysis. To liberate the flavor for isolation, some type of crushing, blending, grinding or homogenization is usually applied to the sample. These different processing methods affect the flavor profile to some degree. pH is a factor to be controlled since it influences the GC profile by changing the volatility of flavor compounds. In some cases, enzymes can be used to facilitate the release of odorants. For instance, amylase can be employed in a starch-based food matrix to free the entrapped volatiles. In some situations, enzymes contained in foods need to be inactivated to prevent alteration of the flavor profile during a lengthy isolation process. Inactivation of enzymes in aqueous samples is often realized via the use of heat. However, heat imposes a risk of thermal artifact formation. Another major source of artifacts comes from oxidation during isolation and concentration.

2.1 Direct Solvent Extraction of Aqueous Samples

Direct solvent extraction utilizes the fact that most flavor compounds exhibit substantial solubility in organic solvents. It is the only method that is reasonably efficient in isolating components of both high and low water solubility[6] and is commonly used to extract flavor compounds from beverages, fruit juices and wines and any materials in liquid form or that can be made into liquid form.

The process is accomplished in two steps: extraction and concentration. Separating funnels or commercial liquid/liquid (L/L) extractors may be used for the extraction. Solvent selection is important and the relevant information can be obtained from Weurman,[7] Sugisawa[8] and Leahy and Reineccius.[9] In general, solvents with a low boiling point are suitable. Nonpolar solvents should be used for samples containing alcohol. Among the commonly used solvents, methylene chloride was found to be satisfactory for general purposes, particularly for flavor compounds with an enolone structure.[10] Other solvents often used include diethyl ether, diethyl ether–pentane mixtures, hydrocarbons and Freons.[7] Safety should be carefully practised during solvent handling owing to the toxicity and potential carcinogenicity of some solvents. Particulate contained in the sample, if any, can be filtered conveniently by a syringe filter such as those recommended for high-performance liquid chromatography (HPLC) sample preparation. Emulsion formation is another frequently encountered problem, particularly in the presence of nonvolatile solutes. Methods such as gentle shaking, centrifugation, slow stirring and pH adjustment may help to curb this pitfall.

A continuous L/L extractor may also be used for extracting flavors from aqueous systems. Continuous-batch L/L extraction generally turns out to be a more efficient method and requires less labor than using a batch separating funnel method but the cost is greater and it requires a longer extraction time.[11] Extraction

with liquid carbon dioxide or supercritical carbon dioxide is also included in this category and is discussed in section 2.6.

Following extraction, removal of the solvent or the concentration of the extract is necessary in most cases before chromatographic separation and identification. This is usually carried out by distillation in a special apparatus combined with the use of sodium sulfate. One of the major problems associated with direct solvent extraction is the impurities extracted and concentrated together with the desired analytes. These nonvolatile components come from natural or Maillard pigments, lipids and alkaloids. These same components may leave a residue in the GC injector. Solvent extraction is generally applied only to fat-free foods.

2.2 Steam Distillation of Samples Followed by Solvent Extraction

Distillation followed by solvent extraction is a common sample preparation technique. It takes advantage of the volatility of flavor components and nonvolatility of the major food constituents. Steam distillation typically involves a source of steam, the sample vessel and a series of cold traps. In direct distillation, the sample is normally placed in a round-bottomed flask and dispersed in water. Heat is applied directly to the flask to distill the steam-distillable components. However, overheating should be avoided to prevent scorching of the sample. Stirring may be applied to curb bumping incurred from the possible presence of particulates. An antifoaming agent may be added if foaming poses a potential problem, but this will result in undesirable peaks in later GC analysis. Indirect steam distillation overcomes some of these problems. The sample is not heated directly, so less sample decomposition occurs. The steam comes externally from an electrically heated steam generator or is produced in a round-bottomed flask heated by a mantle. The steam and volatiles are usually condensed in a series of traps cooled with a succession of coolants ranging from ice–water, a slurry of dry ice and acetone or methanol to liquid nitrogen. Compared with the direct technique, indirect steam distillation is more rapid and easier.

Vacuum steam distillation is a choice if sample decomposition remains a concern. In this case, an inert gas should be bled into the system to aid in agitation. A number of cold traps should be fitted in-line to protect the pump from water vapor and the sample from pump oil vapors. The higher boiling components do not distill as efficiently as they do under atmospheric pressure.

The distillate, obtained from distillation, is a very dilute solution containing water, which should be removed using Na_2SO_4 or anhydrous $MgSO_4$. Extraction of flavors from the distillate with an organic solvent followed by concentration is necessary.

2.3 Simultaneous Steam Distillation and Extraction

As the name simultaneous steam distillation and extraction (SDE) suggests, the distillation and extraction processes needed for sample preparation proceed simultaneously in this technique, which is accomplished via an apparatus specially designed by Likens and Nickerson.[12] Its unique features and related advantages have made this method a popular approach for flavor preparation. Although several modified versions have been made since its origin, the basic principle remains the same. One latest version is shown in Figure 1. A three-necked vertical neck flask (500 ml to 5 L capacity) is filled with the sample slurry to half of its capacity. The sample flask is routinely heated by a heating mantle or a heated oil bath to facilitate distillation with or without stirring. When necessary, agitation is used to prevent bumping and the pH should be adjusted prior to distillation. On the other side of the apparatus, a round or pear-shaped flask of 10–50 mL capacity containing solvent is warmed to vaporize the liquid solvent. Inside the condensers, the vaporized extracting solvent and the sample steam mix

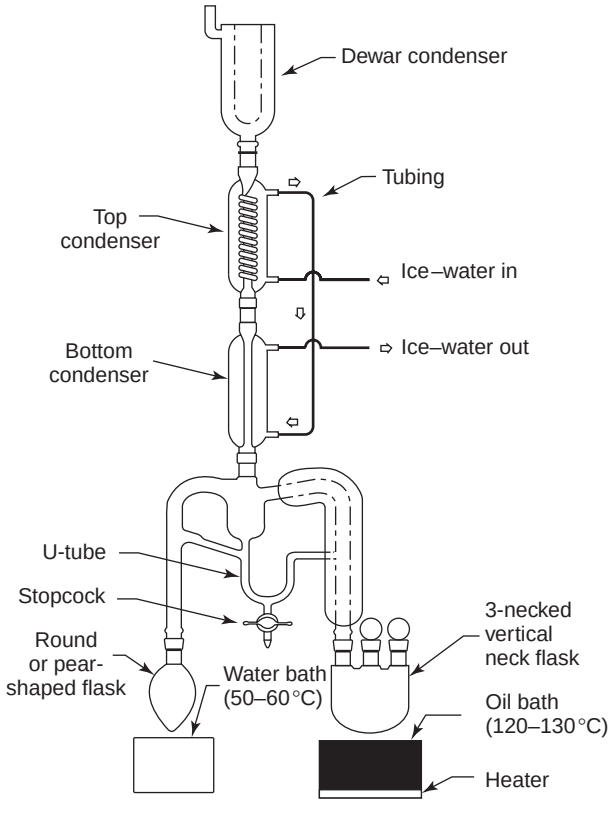

Figure 1 Simultaneous SDE apparatus. (Reprinted from Marsili[5] by courtesy of Marcel Dekker Inc.)

thoroughly and condense under the action of a constant flow of cold water. Consequently, the flavor components from the sample steam are efficiently extracted into the solvent solution.

Various solvents have been used. Schultz et al.[13] found hexane to be excellent, except for lower boiling water-soluble compounds, for which diethyl ether was considerably better. Methylene chloride also gained popularity, but pentane–diethyl ether mixtures are currently favored by most researchers. No matter what solvents are used, boiling chips should be added to both flasks to ensure smooth boiling. The SDE process is normally performed for 1–3 h. When it is completed, the system is cooled and the solvent is collected and dried over an agent such as anhydrous sodium sulfate and concentrated by slow distillation or by nitrogen stream flushing.

One major problem found with SDE is thermal degradation. To cope with this problem, vacuum versions of SDE are available to reduce the temperature. However, this might have an adverse effect on flavor recovery[9] and poses difficulty in operation since one must balance the boiling of two flasks, prevent solvent and volatile losses and hold the pressure constant.[10] Chung and Cadwallader[14] compared SDE under both atmospheric and vacuum conditions for analyzing volatiles from cooked blue crab and found that atmospheric SDE generated more odor-active compounds than vacuum SDE, probably the result of thermal degradation. Besides thermally generated chemicals, other possible sources of artifacts include antifoam agents, vacuum greases on glass joints, contaminated water supplies, solvent impurities and rubber tubing.[11] However, steam distillation remains a popular method for flavor isolation owing to its various advantages such as efficiency, the small volume of solvent required, high recovery and a single operation to accomplish extraction and concentration.

2.4 High-vacuum Distillation of Lipid Samples

Lipid samples are those containing a high percentage of lipids, which include vegetable and nut oils, coffee oil, cocoa butter and lard. Solvent extraction of flavor from this type of samples generally should not be conducted owing to the co-extraction of both flavors and lipids. Although either steam distillation followed by solvent extraction or Likens–Nickerson extraction can be employed, the recovery is not satisfactory.[10] Since aroma compounds are typically lipophilic, extraction of aroma compounds from these food resources requires different techniques.

High-vacuum distillation appears to be a better choice. It involves the direct transfer of volatiles from a food sample to a cold condenser. This requires very short distances between the condenser and the food sample and the use of high-vacuum systems ($<10^{-3}$ Torr).[11] Several designs are available for this purpose.[15,16] A falling-film molecular still, designed on the principle of vaporization of the flavor from a heated thin film of the oil under high vacuum, can be used when large amounts of lipid materials are present.[10] The most commonly found artifacts via this method are silicone compounds from the high-vacuum grease used on glass joints. The use of silicone O-rings avoids this problem.[15]

2.5 Direct Solvent Extraction of Solid Samples

Direct solvent extraction is also a simple and convenient technique for sample preparation, suitable mostly for dry products such as spices, grain flours, chocolate, coffee and other food items of low moisture content. A Soxhlet extractor is often used for the extraction. A portion of the ground sample can be placed in a Soxhlet thimble and extracted with an organic solvent of low boiling point, which can be either diethyl ether or methylene chloride. A series of extraction cycles are necessary to obtain more analytes. At the end, the solvent with extracted analytes should be combined and concentrated. However, nonvolatile compounds such as lipids, caffeine and pigments will also be collected together with the desired analytes. Consequently, techniques such as steam distillation or high-vacuum stripping should be conducted to separate the volatile from the nonvolatile compounds to prevent column clogging if the latter are present in large amounts. Another problem with Soxhlet extractors is the large amount of solvent needed, especially if a large quantity of sample is to be extracted, given the solvent-to-sample ratio needed of around 2 : 1 by volume, which is generally unfavorable. Large solvent volumes necessitate extreme solvent purity to minimize artifacts or solvent contaminants.[11] Column chromatography provides an alternative means to extract solid samples. In this case, the dry solid sample can be ground with an inert bulking material. The blend is then packed into a glass chromatography column and eluted with organic solvent.[17]

2.6 Extraction with Liquid or Supercritical Carbon Dioxide

The drive to search for safer and more environmentally friendly extraction and separation methods for flavor analysis resulted in the development of an extraction technique involving liquid or supercritical carbon dioxide as solvent.[18] Although extraction with liquid CO_2 and extraction with supercritical CO_2 are two different processes, the principles are the same. Liquid CO_2 extraction takes place at a subcritical state (temperatures <

31 °C, pressures < 74 bar), whereas supercritical extraction proceeds at temperatures above 31 °C and pressures above 74 bar[19,20] However, supercritical CO_2 has been stated to be a more versatile solvent than liquid CO_2 owing to the wider range of extraction conditions,[21] and therefore presumably a wider range of applications. The physical properties of compressed CO_2 and the required operating conditions confer on the separation technique a number of advantages compared with more conventional extraction and separation processes. First, CO_2 is nontoxic and nonflammable, so it is safer to handle and friendly to the environment. The extracted compounds can be smelled and tasted owing to the absence of organic solvents. The relatively mild temperature and pressures employed in the process reduce the possibility of artifacts produced by thermally labile compounds. In addition, the extracted material is protected from oxidation because of the inert atmosphere in which extraction and separation are conducted. Characteristics of supercritical fluids suitable for extraction purposes and their comparison with conventional methods have been reported by Brogle[22] and King et al.[23]

The advantages of liquid or supercritical CO_2 extraction and separation techniques make them ideal for the food and flavor industries and their use has increased during recent years.[20,24–26] They are used regularly for the extraction of hops in the brewing industry, for the decaffeination of coffee and tea, for the dealcoholization of beverages and in the extraction of oleoresins from spices.[20,27,28] Flavors and fragrances extracted by this method were of higher quality and more true to the source than those from other methods such as solvent extraction, steam distillation, hydrodiffusion and vacuum distillation.[29,30]

When using supercritical carbon dioxide, it is necessary to balance temperature, pressure and flow rate, which requires complex instrumentation.

2.7 Headspace Analysis

Volatile organic compounds contributing to flavor are present in foods in a dynamic equilibrium, which results in the partition of these compounds between the food matrix itself and the gas phase immediately surrounding it. When a food is enclosed in a sealed vessel, its flavor volatiles will partition into the headspace gas of the container, and the many individual compounds comprising the flavor will eventually attain a gas-phase equilibrium concentration. This equilibrium is profoundly influenced by the solubility of the flavor compounds in the food matrix, their vapor pressures, temperatures and many other unpredictable factors such as their enzymatic release from nonvolatile precursors. Static headspace analysis may be chosen when chemical compounds exist as gases at room temperature or have a sufficiently high vapor pressure to evaporate and produce a gas-phase solution. Experimentally, the food sample is placed into a headspace vial that is septum sealed and often warmed to facilitate volatile vaporization. A time period is required to establish equilibrium between the sample phase and the gaseous phase. Subsequently, an aliquot of the headspace gas is withdrawn with a syringe and injected into a gas chromatograph injection port. Since there is no heat involved, the introduction of artifacts is greatly limited. Analyses of volatiles emitted by plants or food products provide data representing fresh flowers, fruits and vegetables as they truly are. In addition, this technique is simple and reliable and remains the cornerstone of many quality assurance/quality control (QA/QC) protocols in the food industry.

However, the static headspace method is also associated with some shortcomings. It has a relative low sensitivity owing to the limited volume of headspace gas that can be injected into the GC column and the low volatility of some flavor constituents. Some organoleptically significant compounds exist at trace levels and may not be detected because of the detection threshold capabilities of GC and GC/MS. Attempts to increase the sensitivity by injecting a larger volume of headspace gas results in loss of chromatographic resolution. Furthermore, oxygen and water vapor often present in the headspace can cause rapid degradation of expensive GC stationary phases and mass spectrometer filaments.

In order to enrich the headspace gas prior to GC analysis, various techniques have been investigated and together they are termed as dynamic headspace analysis since they all involve a means of concentrating headspace aroma volatiles. Broadly, these techniques include cryotrapping, steam distillation, solvent extraction and adsorption (purge and trap). Detailed information about these techniques can be found in Chang et al.,[31] Hartman et al.[32] and Wampler.[33]

The physical principle of these sample preparation methods is to strip the volatile flavor compounds from the headspace above the foods by a stream of air or inert gas.[34] The volatiles are trapped by a collection trap while the purging gas filters through. Different from the static headspace method, where an equilibrium between the sample matrix and the surrounding gaseous phase is established before sampling, in the dynamic headspace method the headspace gas along with the volatile analytes is constantly purged and removed, while the analytes are retained. The process not only causes more of the volatile compounds dispersed in the sample matrix to leave the sample and pass into the trap, but also results in a higher concentration of the analytes available for GC/MS

analysis. In fact, a dynamic headspace apparatus is usually referred to as a ''sample concentrator''. The trapping step may involve adsorption on to a high-surface-area absorbent material (e.g. activated carbon or Tenax® TA) or cold trapping by condensing or freezing the analytes in the trap. The trapped analytes are then desorbed into the GC and GC/MS system for analysis. Figure 2 shows the apparatus used for one of the dynamic headspace techniques called purge and trap.

The dynamic headspace technique does not require the use of solvents. It is an easy sample preparation process and can be automated. In addition, the concentration process increases the sensitivity. Volatiles at parts per billion (ppb) levels can be analyzed routinely and it is possible to analyze components even at the level of parts per trillion (ppt).[35] Furthermore, selectivity and simplicity can be enhanced by using a specific sorbent targeting for specific analytes in the dynamic process.

2.8 Direct Thermal Desorption Methods

DTD is the technique of sparging the volatile compounds from a sample matrix and transferring them directly on to a GC column for analysis.

The development of DTD has gone through several stages. In earlier research, samples were stationed inside a liner with the help of glass wool. The liner was then placed directly in the injection port. The volatiles from the sample were thermally desorbed on to a packed column held at room or subambient temperature followed by GC analysis. This process is associated with a few drawbacks, including finger burn and broad peaks. A later development is called external closed-loop inlet device (ECID), which enables the sample to be heated in its own block instead of inside the injection port. A detailed description is available in Grimm et al.[36] Although ECID has been adapted for use with both packed and capillary columns, its application to capillary columns turned out not to be successful.[36]

The most recent advancement of DTD is called short-path thermal desorption (SPTD). It employs a thermal desorption unit that can be operated either manually or automatically. The device (Figure 3) sits directly on top of the injection port of most GC systems, where it is utilized for the direct desorption of both volatile and semivolatile samples into the GC injection port and column. The sample is placed in a glass-lined stainless-steel tube (GLT) with or without the presence of an adsorbent resin such as Tenax® TA or activated carbon inside the tube. When ready for analysis, the tube is affixed on one end with an injection needle and the other end is screwed into the SPTD unit. During injection, the tube with the needle is released down

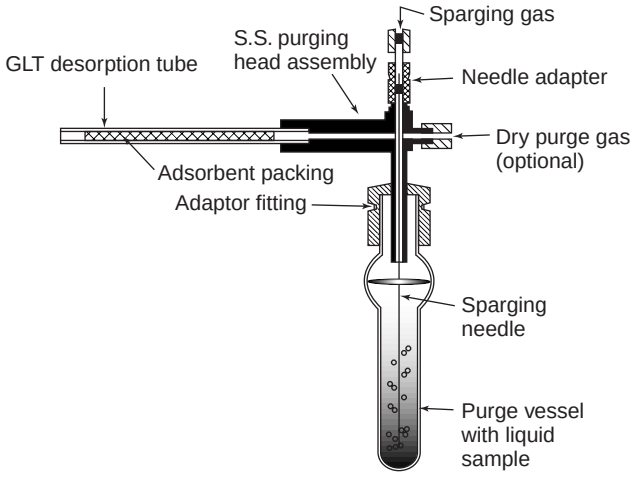

Figure 2 Purge-and-trap apparatus for the isolation and concentration of volatile flavor compounds from liquid sample matrices. (Reprinted from Ho and Manley[3] by courtesy of Marcel Dekker Inc.)

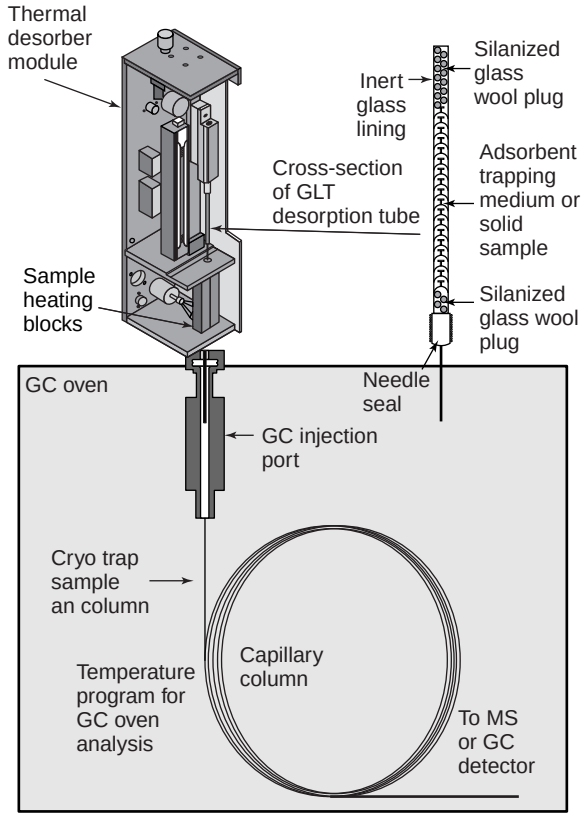

Figure 3 Schematic representation of an SPTD module mounted on a gas chromatograph, illustrating the features and operating principles of the system. (Reprinted from Ho and Manley[3] by courtesy of Marcel Dekker Inc.)

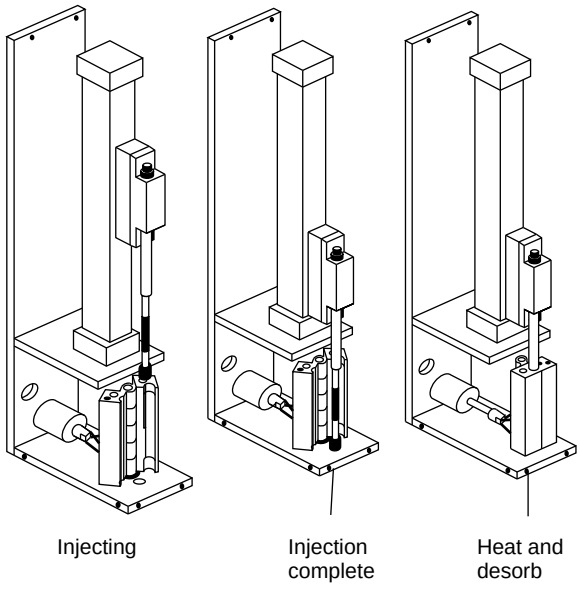

Figure 4 Graphical representation of a thermal desorption unit in different operating modes. (Reprinted with permission from the ''Catalog of Short Thermal Desorption Model TD-4'', Scientific Instrument Service.)

into the heating block of the unit and positioned in proper alignment with the GC injection port (Figure 4). The desorption temperature, time and equilibration are controlled by an electronic console processor. When injection is complete, the flows are readjusted according to the method of analysis. The hinged heating blocks close around the desorption tube, which is heated to the desired temperature to desorb the volatiles from the sample. Under the flow of carrier gas, the desorbed volatiles are purged into the GC injection port and on to the front of the GC column that is held at room or subambient temperature. The volatiles are thus trapped continuously into the GC column until the completion of desorption. Subsequently, GC temperature programming is initiated to volatilize the condensed organics for elution, separation and identification.

Compared with more conventional methods for sample isolation and concentration, DTD is a thermal extraction system of high sensitivity. In addition to being solvent free, the sample requirement is small and tedious sample cleanup is eliminated. More relevant information can be obtained in Hartman et al.[32] Owing to the ''short path'' of sample flow, SPTD overcomes the shortcomings of previous desorption systems via the elimination of transfer lines. It provides several unique advantages over other desorption systems. First, it enables the sample to be heated rapidly. Second, the desorbed component can be easily and efficiently transferred into the injection port of the GC system from a GLT, reducing the possible degradation that is frequently encountered in other thermal desorption devices. As a result, artifact production is minimal.

DTD, including SPTD, can be used widely in food analysis.[32,37] In general, the samples should be solid or semisolid with moisture contents lower than 5% and volatile flavor levels in the range $1–10^6$ ppm (w/w). The moisture content must be low enough so as not to cause freezing and plugging of the GC column when operating in the cryofocusing mode, especially when utilizing narrow-bore capillary columns. Sample sizes in the range 1–500 mg are common in this technique. Before placing the sample inside the tube, sample homogenization is often needed to compensate for flavor variation. This can be accomplished using a laboratory grinding mill with the help of a thermally preconditioned chromatography support (60–80-mesh Chromosorb®).[32] Samples successfully analyzed for flavors or off-flavors using DTD techniques include spices, coffee, candy, food ingredients, peanuts, vanilla beans, chewing gum, flowers, plant materials, natural products and packaging materials.[38]

2.9 Solid-phase Microextraction

SPME is a newly developed sample extraction and concentration technique used for volatile and semivolatile compound analysis. It is based on the principle of partitioning of organic components between a bulk aqueous or vapor phase and thin polymeric films coated on fused-silica fibers.

The SPME device, as shown in Figure 5, is like a syringe with a spring-loaded plunger and a barrel with a detent to allow the plunger to be held in an extended position

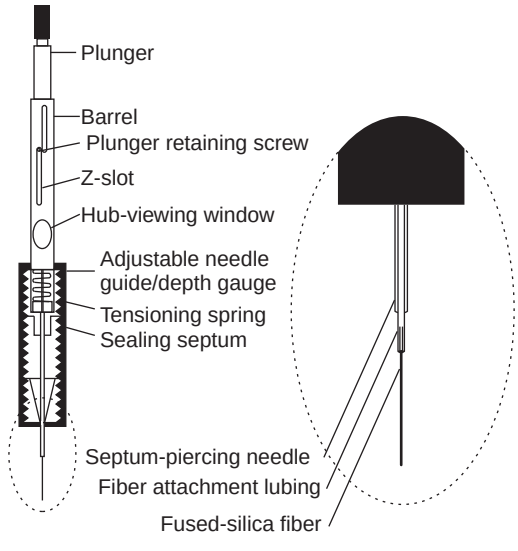

Figure 5 Graphical representation of an SPME device. (Reprinted from Marsili[5] by courtesy of Marcel Dekker Inc.)

during the extraction and the injection/desorption period. Inside the barrel there is a stainless-steel needle, which encloses another length of stainless-steel tubing fitted tightly to a short piece of solid-core fused-silica fiber. The bottom part of the fiber is coated with a thin layer of stationary phase. The most commonly used coating polymeric materials include polyacrylate, Carbowax™, divinylbenzene and polydimethylsiloxane. Therefore, there are different SPME fibers coated with these materials for extracting compounds of different size, polarity and volatility.

The sample being analyzed is usually placed in a vial or other suitable container, which is sealed with a septum-type cap to facilitate equilibrium between the headspace and the sample matrix. Before sampling, the fiber should be cleaned because it could possibly be contaminated by aroma chemicals from the air. This can be done by inserting the fiber-attached needle into a GC injection port for a few minutes to desorb thermally any possible contaminant. During sampling, the SPME needle pierces the septum and the fiber is extended out from the needle and immersed directly into an aqueous sample or into the headspace above a liquid or solid sample. Organic compounds in the sample are subsequently adsorbed on the fiber. After a suitable sampling time, the fiber is withdrawn into the needle, which is then removed from the septum and inserted directly into the injection port of a gas chromatograph. Inside the injection port, the absorbed compounds are thermally desorbed and transferred directly to the column for analysis. Care should be taken during the whole process to prevent the brittle fiber from breaking. For high accuracy and precision, full equilibrium is not necessary but a consistent sampling time and other sampling parameters are essential.

The selection of the stationary phase is based on the targeted analytes. Two factors should be considered: polarity of the analytes and volatility and size of the analytes. Polar phases such as polyacrylate and Carbowax™ coatings should be chosen for polar analytes. Fibers with a thicker film (100 μm) are better for volatiles, but also can be used for less volatile compounds with a longer extraction time. Porous fibers can also retain small analytes and are ideal for C_2-C_6 analytes. Thinner film fibers (7 and 30 μm polydimethylsiloxane) are better for larger molecules. Selecting the most appropriate fiber certainly influences the SPME recovery. Sample agitation or stirring, minimizing headspace volume, salt addition and adjustment of pH and temperature are other means of choice for facilitating equilibrium and improving SPME recovery.

The simple SPME requires no solvents or complicated apparatus. Eliminating most drawbacks of conventional extraction methods, it can concentrate volatile and non-volatile compounds in both liquid and gaseous samples for analysis by GC, GC/MS or HPLC. The process can be automated to obtain consistent data of high precision and accuracy. Uniquely, it can be used to extract analytes directly in the field without collecting a sample.

As a result of its various advantages, SPME has quickly gained wide application as an analytical tool in various areas, including flavor analysis, since its commencement. Most recently it was used to analyze the aroma volatiles of a specific rose variety in a weightless environment. It has been successfully used to evaluate flavor and aroma volatiles in fruits,[39,40] wine[41] and other beverages.[42,43] These applications demonstrate the utility of the technique in trapping highly volatile, lower molecular weight and thermally labile flavor compounds where solvent use and distillation methods have proven to be detrimental to product aromagrams.

3 SEPARATION TECHNIQUES

3.1 Pre-gas Chromatography Separation

3.1.1 Liquid Chromatography

The flavor mixtures obtained from food samples usually contain numerous chemical components. This complexity often makes it difficult to identify the flavor components occurring in the system. It is desirable to simplify the mixture to less complex fractions prior to further instrumental analysis. On the other hand, some flavor compounds exist at extremely low levels and large amounts of sample are needed to obtain enough of these trace elements. For the purpose of structure elucidation, the targeted flavor compound needs to be pure and in a relatively large amount in order to conduct nuclear magnetic resonance (NMR) spectroscopy. These aspects necessitate the use of some conventional separation techniques such as column chromatography, alumina fractionation and thin-layer chromatography (TLC) in flavor analysis. Takeoka et al.[44] used neutral alumina to fractionate the extract from pineapple and identify 26 constituents, including seven sulfur compounds for the first time. In studies of the thermal degradation of thiamine, medium-pressure liquid chromatography based on silica gel was used to preseparate concentrates into different fractions before analysis by high-resolution gas chromatography (HRGC) and HRGC/MS and NMR spectroscopy.[45] Fractionation by silica gel TLC has found its use in pinpointing off-flavor compounds from spray-dried skim-milk powder[46] and identifying volatile components in sesame seed oil.[47] Coulibaly and Jeon[48] studied various solid-phase materials including activated carbons, C_{18} reversed-phase silica, Florisil and silica gel

for isolation of relatively nonvolatile flavor compounds from ultra-high temperature (UHT) milk. The results of their study indicated that the combination of activated carbon and C_{18} reversed-phase silica is the most effective method for isolating flavor constituents for GC and GC/MS analysis.

The use of HPLC in flavor analysis is largely related to the separation and identification of compounds that are nonvolatile and thermally unstable and usually impart taste perception. These compounds include sugar, nonsugar sweeteners, bitter amino acids and peptides, citrus limonoids and flavonone glycosides. Since HPLC is normally conducted at room temperature, heat-induced artifact formation is greatly reduced. The nature of HPLC also enables it to analyze aqueous samples directly without the need to extract flavor chemicals into an organic phase as is necessary in GC. Fractions obtained via HPLC can be subjected to sensory evaluation given that the solvent used is water or ethanol.

The application of HPLC can be enhanced when it is coupled with GC and GC/MS. The role of HPLC in this powerful analytical system is to fractionate a complex flavor mixture so as to simplify later GC/MS separation and identification, which has to be conducted for each HPLC fraction. This is tedious, but it is necessary, especially for very complex flavor systems. Direct transfer of HPLC fractions into a capillary gas chromatograph designed by Grob et al.[49] made it more convenient to use. To investigate the character impact of compounds in roasted coffee, hundreds of roasted coffee flavor components were fractionated by column chromatography and preparative HPLC, and then subjected to GC, GC/MS and simultaneous GC/sniffing. The sulfur-containing flavor components 3-methyl-2-butene-1-thiol, 1,3-dimercapto-3-methylbutanol and 3-mercapto-3-methylbutyl formate were found to contribute to the flavor of roasted coffee.[50]

3.1.2 Acid–Base Separation

Some flavor compounds exhibit a certain degree of acidity or alkalinity. They can be ionized to be soluble in aqueous solutions. This method takes advantage of the differential solubility of ionized and nonionized species in aqueous and organic solvents.[51] Ionization of ionizable compounds is generally achieved by controlling the pH of the aqueous solvent. This method is normally conducted after obtaining a flavor isolate mixture in order to isolate selectively the compounds of interest.

This method is relatively simple and does not require sophisticated equipment or expensive reagents. After fractionation, the number of flavor constituents is reduced for column analysis, thus allowing a more specific column choice and tailoring of GC operating conditions.[51] However, this technique is commonly associated with incomplete separation and poor recovery. To compensate for this, large volumes of solvents and multiple extraction are necessary, which will result in tedious work and dilution of flavor compounds, followed by a greater loss of volatiles during the subsequent step of concentration. Moreover, artifact formation is possible under both alkaline and acidic conditions.[51]

Acid–base separation can be combined conveniently together with purge-and-trap techniques for separation purposes. This method modifies the purge-and-trap method to make it more selective; thus it is called the selective purge-and-trap method. It enables flavor chemicals that are minor in quantity but significant in flavor impact to be singled out, thus overcoming the interference by major constituents or by compounds with the same chromatographic properties. This method can be used for the analysis of important flavor volatiles such as pyrazine, pyridines, pyrroles, oxazoles and thiazoles.[52,53] Kuo et al.,[54] Joo and Ho[55] and Yoo and Ho[56] have used this technique for the analysis of volatile pyrazines. In their experiments, as shown in Figure 6, the sample was dispersed in a purge vessel and the suspension was constantly sparged with a stream of nitrogen gas. The purged gas along with the volatiles passed through an HCl trap. Owing to their weakly basic properties, pyrazines were trapped as salts in the acid trap. At the end of purging, the pyrazine-containing acid solution was titrated with NaOH to become basic, and then an alkaline solution was either extracted by a liquid extraction method to isolate the pyrazines or again purged with N_2 gas to trap the pyrazines in a Tenax® TA tube. The results proved that this method not only quantifies minor or trace pyrazines from a complex background but also it is much less laborious and time-consuming than other traditional isolation and fractionation methods.

3.1.3 Analysis of Aldehydes and Dicarbonyls Using Chemical Trapping Agents

Volatile aldehydes and dicarbonyls, occurring in foods or generated via various mechanisms, pose characteristic

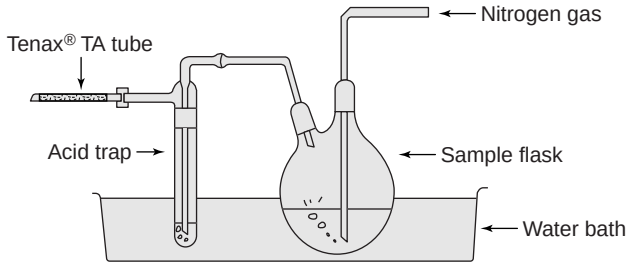

Figure 6 Apparatus for selective purge-and-trap isolation and fractionation. (Reprinted with permission from Kuo et al.[54] Copyright 1989 American Chemical Society.)

flavor notes or serve as precursors for the formation of other important flavor compounds. They can play both positive and negative roles by imparting to foods good flavor or off-flavor. The importance of carbonyls also resides in their potential toxicity.[57]

These aspects make the analysis of aldehydes and dicarbonyls in foods necessary. Many methods designed for carbonyl elucidation have been reported. Most of them are based on the concept of using so-called chemical trapping agents, which react with the carbonyl compounds to produce derivatives that can be readily detected. Some of the commonly used agents include 2,4-dinitrophenylhydrazine,[58] *N*-benzylethanolamine[59] and benzyloxyamine.[60] One problem in using these derivatization agents is that the strongly acidic and high-temperature conditions required may alter the carbonyl compounds of interest or produce additional carbonyl compounds.[61] Moreover, the formation of both *syn* and *anti* forms of aldehyde derivatives is inevitable, making peak identification extremely difficult. Fortunately, a new derivatization method with cysteamine (aminoethanethiol) as the trapping agent was developed recently by Shibamoto et al.[62,63] According to Yasuhara and Shibamoto,[64] this method has several advantageous features compared with other derivatization methods: (1) only one derivative is formed from one aldehyde; (2) the reaction takes place in very mild conditions, rapidly and with almost quantitative yield; (3) the derivatives, thiazolidines, can be separated perfectly with a fused-silica capillary column and detected selectively with a nitrogen–phosphorus detector; and (4) excess of cysteamine does not interfere with GC analysis. It has been used to analyze volatile carbonyl compounds in foods, beverages, cigarette smoke and automobile exhaust.[61] Many volatile carbonyls at trace levels in the headspace of food samples have also been satisfactorily analyzed by this method.[64,65] A detailed description of the experimental procedure is available in Yasuhara and Shibamoto.[64]

3.2 Gas Chromatography Separations

3.2.1 Standard

In GC, the components of a vaporized sample are separated as a result of partitioning between a mobile gaseous phase and a stationary phase held in a column. The stationary phase in GC is a liquid supported on an inert solid matrix. The excellence of GC has made it one of the most important and widely used column chromatographic methods. The basic process involves the single column separation of a volatile mixture followed by detection at a single "universal" detector. The sample is introduced as a gas at the head of the column and those components that have a finite solubility in the stationary liquid phase distribute themselves between this phase and the gas according to the law of equilibrium. Elution is then accomplished by forcing an inert gas such as helium through the column. The rate of movement of the various components along the column depends on their tendency to dissolve in the stationary liquid phase.

Given its analytical power, it is not surprising that GC has become crucial to the specialized field of aroma investigation. Even in its most essential form, which enables a single column separation of a volatile mixture followed by detection with a single detector, GC can provide several pieces of key information regarding the composition, the elution times, the relative concentrations and the relative volatilities of the various components in the sample mixture under specific operating conditions. However, an obvious limitation is that a basic GC process would fail to reveal the definitive identification, structural and aroma character related to the peaks in a chromatogram. In addition, its ability to separate volatile components is restricted by the complexity of possibly numerous volatile components present. It is not uncommon to have a total component population of hundreds or thousands for a flavor analysis at a target concentration level of as low as 10 ppt.[66,67] Moreover, the actual separation capacity depends on a number of factors, including column efficiency, system performance, column temperature limits and system operating parameters.

3.2.2 Multidimensional

One way to enhance the separation performance by a basic GC process, referred to as single-dimension gas chromatography (SDGC), is to combine different techniques together. A series of hardware configurations are loosely bound under the broad term known as multidimensional gas chromatography (MDGC). According to Bertsch,[68] two columns of different selectivity and a device to transfer selectively a portion of a chromatographic effluent from one column into another are needed for a technique to qualify as an MDGC technique. However, Wright[67] favored the notion that in a broader sense the technique can be called MDGC as long as it provides increased qualitative information about peak identification, functionality or other characteristics than could otherwise be provided by SDGC alone. Under this concept, MDGC includes GC combined with MS, Fourier transform infrared (FTIR) spectroscopy, atomic emission detection (AED), inductively coupled plasma (ICP), etc. For the purpose of volatile separation, MDGC usually denotes a two-dimensional GC system. A simplified version of the system is shown in Figure 7. This basic system consists of an injection device, a first column for preliminary separation, a flow-switching device connecting the first column and the second column, a first detector for

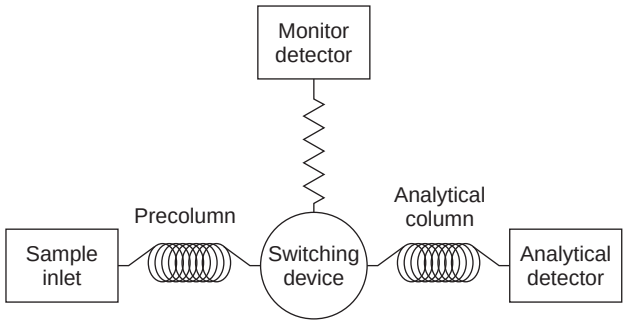

Figure 7 Schematic diagram illustrating the key elements of a general MDGC separation system. (Reprinted from Marsili[5] by courtesy of Marcel Dekker Inc.)

profiling the first column separation and a second detector for profiling the second column separation. The first column and detector are often referred to as the precolumn and monitor detector and the second column and detector are called the analytical column and analytical detector.

Among these components, it is the flow-switching device that permits the tremendous increase in separation power of MDGC. It can transfer those portions of the precolumn effluents that are adequately separated or are of no direct interest to the monitor detector, while it directs those portions of the precolumn effluent that are incompletely resolved or are of particular interest into the analytical column for additional separation. By doing this, a number of chromatographic separation difficulties resulting from sample complexity can be overcome. In the system, the precolumn de facto serves as an extension of the sample preparation or sample cleanup steps,[66] so a larger sample volume is allowed for injection compared with SDGC. The peaks appear on the monitor detector as a reference to guide the selective transfer of the effluent of interest to the analytical column/detector, which usually achieves the desired separation. Separation enhancement is positively related to the difference of polarity or selectivity between the precolumn and the analytical column.[69] In many cases, MDGC permits a particular separation to a lower detection limit.

When concentrating the volatile organic compounds from food products, any large amount of potential interfering compounds is inevitably collected and introduced into the GC system. Under SDGC conditions, some of the most significant aroma peaks may be covered up by the surrounding mass of peaks of little or no significance associated with aroma. MDGC can be utilized to overcome this shortcoming. Nishimura[70] was able to identify six new compounds in ginger by AEDA combined with GC and a modified multidimensional GC/MS system. Homatidou et al.[71] used the Freon 11 extraction method and a combination of MDGC and GC/MS to study the volatiles of orange-fleshed *Cucumis melo* L. var. Cantaloupensis and reported a total profile of identified compounds. In addition, MDGC is often used for enantiomer analysis. Chiral γ-lactones from the raw flavor extract of strawberries and of some commercially available fruit-containing foods and beverages were directly stereoanalyzed by MDGC.[72]

3.2.3 Chiral

Like many other chemical molecules, flavor compounds can also exist as enantiomers. They pose various flavor characteristics of different intensity and with different specificities toward odorant receptors in human subjects. It becomes necessary to differentiate enantiomers partially owing to their distinct impact on flavor characteristics and flavor perception. However, analysis of chiral flavor compounds is not an easy task. Various methods, including chiroptical methods, NMR spectroscopy, electrophoresis and chromatography have been attempted. These methods each have their own advantages and limitations, and thus are suitable only for specific purposes in different disciplines.[73] For the analysis of flavors and fragrances, chromatographic methods are the most suitable. Among the common chromatographic methods, TLC is mainly used for achiral routine analysis whereas HPLC and supercritical fluid chromatography (SFC) offer the advantages of large sample throughput for preparative enantiomer separation.

For analytical purposes, however, GC is the first choice. The key step in the analysis resides in the separation of enantiomeric pairs, which can be accomplished by using the so-called indirect method or direct method. The indirect method requires derivatization of enantiomers to volatile diastereomers. The characteristics and application of a number of derivatization agents were summarized by Schreier et al.[74] This method can be applied to a wide range of chiral compounds such as alcohols, aldehydes, ketones, carboxylic acids, esters and even hydrocarbons. However, the lack of absolute enantiomeric purity of the chiral derivatization agent makes this method inferior to the later developed direct method that is based on the use of a CSP of high enantiomeric purity. According to the separation mechanism, CSP can be divided into three main classes: amide phases, metal complex phases and cyclodextrin phases. They are applicable in different scenarios with their own special features. From the standpoint of flavor analysis, a cyclodextrin-based CSP is the most useful and has become increasingly popular in recent years. Cyclodextrins are cyclic glucans (cyclomaltooligoses) with at least 6–12 D-glucopyranose units in an α-1,4-glycosidic connection. α-, β- and γ-cyclodextrins with 6, 7 and 8 glucose units are usually used in analytical chemistry. In recent years, a wide variety of CSPs based on these materials have become available. Almost every chiral flavor

compound has been made amenable to enantiomer separation by GC with excellent results. The separation of enantiomers requires more sophisticated instrumentation compared with routine flavor analysis. Preferably an MDGC system composed of an achiral precolumn and a chiral main column should be used to separate the enantiomers better.[72,75] Often the MDGC system can be coupled with an MS detector or FTIR spectrometer for enantiomer identification.[76–78]

It is noteworthy that selecting the most suitable CSP for a particular enantiomeric separation is difficult. This is complicated by the commercial availability of various kinds of CSP and the fact that at different temperatures, the elution order of enantiomers can be reversed.[73] Nevertheless, understanding the enantiomer composition of chiral compounds is very important in the study of biochemical pathways and the biosynthesis of flavor compounds. It also plays a crucial role in the differentiation of natural and synthetic flavors and fragrances, which are important in quality control and in the composition and preparation of fragrances and perfumes.

4 IDENTIFICATION OF FLAVOR COMPOUNDS

4.1 Gas Chromatography Retention Indices

In GC, the time elapsed from sample injection to the appearance of a peak maximum is called the retention time (RT). It is the chromatographic property of a chemical, since each compound shows a characteristic emergence on the chromatogram. The RT of an unknown can be compared with that of a standard for identification. However, the RT of a compound is influenced by a number of factors such as column type, carrier gas flow rate and temperature programming conditions. It is almost impossible to reproduce the same RT for the same compound even under the same running conditions. The need to create a standardized system to express GC retention data resulted in the birth of an RI system, known as retention indices or Kovats indices, so called after its designer, Kovats.[79] In this system, the RT of a solute is compared with the RTs of a homologous series of *n*-paraffin hydrocarbons and the RI (I) is calculated by Equation (1):

$$I = 100N + 100(N+1)\frac{\log t_1\mathrm{R(A)} - \log t_1\mathrm{R}(N)}{\log t_1\mathrm{R}(N+1) - \log t_1\mathrm{R}(N)} \quad (1)$$

where N and $N+1$ = carbon numbers of paraffin hydrocarbons eluting before and after an unknown solute A and $t_1\mathrm{R}(N)$ and $t_1\mathrm{R}(N+1)$ = RT of paraffin hydrocarbons whose carbon numbers are N and $N+1$, respectively.

The other available RI system is an ethyl ester index system.[80] GC RIs are used alone or in conjunction with other analytical means for the identification of the unknown. The calculated RI of the unknown is compared with the literature database for the same column type. In GCO analysis, the aroma quality of a volatile component combined with more than one RI value of different column types is considered as equivalent to identification by GC/MS.[81] RIs together with GC/MS can have a much better performance for identification than GC/MS alone.[82]

4.2 Mass Spectrometry

In flavor analysis, MS is one of the most powerful and effective methods in terms of identification. Together with GC, it provides both qualitative and quantitative information about the aroma from a complex food matrix. The GC column effluent is first ionized by electron bombardment to form a spectrum of ion fragments and then MS selectively allows fragments of a certain mass to charge ratio (m/z) to be analyzed under preset conditions. Quadrupole mass spectrometers provide a relatively lower resolution although they are more easily interfaced to various inlet systems. On the other hand, by manipulating an electrical field with different geometries, an ion trap is created in an ion trap spectrometer. In this ion trap, fragments with specific m/z can be trapped, concentrated and stored prior to subsequent mass analysis. As a result, ion trap spectrometry is inherently more sensitive than other mass detection methods. Typically, full-scan spectra are possible even though the analytes are at picogram levels. The ion-storage ability of the ion trap has a useful application in advanced MS ion-preparation techniques (e.g. tandem MS), which greatly increases the analytical power of the instrument.[83]

The mass spectrum obtained is a characteristic of the analyte. For identification, the mass data of each unknown are analyzed via computer matching with spectral databases of known compounds. The mass spectrum of an individual compound is unique. Nevertheless, a certain class of compounds do exhibit some common spectral features that are useful in identity differentiation. The feature of a spectrum results from the fragmentation of a molecule, which characterizes itself from other molecules and from class to class. Flavor constituents ranging from aldehydes, ketones, alcohols and acids to heterocyclic compounds display their own characteristic patterns of fragmentation. More detailed information is available in Porter and Baldas,[84] Ho and Carlin,[85] Schwarzenbach and Gubler[86] and Mottram.[87]

The mass spectra derived from an ion trap mass spectrometer are slightly different from that obtained from a quadrupole mass spectrometer. The ion trap mass spectra usually have a profound $(M + 1)^+$ ion which make them more difficult to compare with most of the commercially available databases which were generated using quadrupole mass spectrometers.

4.3 Fourier Transform Infrared Spectroscopy

FTIR spectroscopy, based on the fact that there is a correlation between the wavenumbers at which a molecule absorbs infrared radiation and its structure, is a well-known analytical tool. Its power in elucidating a chemical structure by providing information about the functional groups within a molecule arises from the observation that functional groups tend to absorb infrared radiation in the same wavenumber range regardless of the structure adjacent to the functional groups. It is also able to provide quantitative information, such as the concentration of a molecule in a sample, on the basis of Beer's law.[88] However, FTIR spectroscopy by itself cannot necessarily be used to reveal the complete picture of an unknown. The development of infrared technology has advanced to combine it with different techniques such as GC, HPLC, ultraviolet/visible spectroscopy, MS and NMR spectroscopy, to enhance the overall analytical performance. The combination of GC and FTIR spectroscopy has proven to be very useful in analyzing complex mixtures quickly and accurately. GC/FTIR spectroscopy and GC/MS have been used to investigate citrus flavor decomposition processes[89] and to study flavors in products such as kiwi fruit,[90] shiitake mushrooms[91] and teas.[92] Using GC, FTIR spectroscopy and MS together, Chang et al.[93] identified a total of 74 compounds from bread crumbs and crust, of which 11 had bread-like odors. This research indicates that FTIR spectroscopy, especially when combined with other advanced analytical techniques, is complementary in identifying flavor compounds.

5 QUANTIFICATION OF VOLATILES

Quantification of volatiles in flavor analysis is routinely accomplished by using internal standards no matter which analytical technique is involved. In general, the chemical used as the internal standard should be stable under all conditions during analysis. It should have similar chemical and physical properties to the analyte and not react with the analyte to produce artifacts. The internal standard should be added to the sample matrix before any sample treatment so that it is processed simultaneously with the sample volatiles. For a liquid sample, the internal standard solution should be miscible with the sample matrix to ensure proper dispersion. In analyzing flavor components from solid material using a purge-and-trap or a thermal desorption technique, similar volatility of the internal standard to that of the analyte is desirable, and the introduction of an internal standard into the sample tube just before purging is advised to avoid preferential volatilization.

A combination of internal standard and appropriate response factors can be used for accurate volatile analysis and simultaneous determination of nonvolatiles.[94] However, bias is difficult to avoid because it is impossible for an internal standard to have chemical and physical properties close to those of all chemical compounds in a mixture. In addition, the use of a traditional internal standard method is limited when quantifying trace levels of volatiles or compounds that are labile or polar.[95] Overall, the traditional internal standard method is crude and often referred to as ''semiquantitative''. For accurate quantification, the stable IDA can be used to target specific compounds. In this method, the isotope-labeled internal standard possesses the same chemical and physical properties as the analyte. Loss of analyte during the whole analysis period is also encountered by the standard. As a result, the accuracy is greatly enhanced. The advantage of IDA over other quantitative approaches has made it increasingly popular in both academia and industry.[96] Several articles have provided detailed information about this method.[96,97]

6 GAS CHROMATOGRAPHY/ OLFACTOMETRY ANALYSIS

Although flavor analysis by GC/MS provides important information about the identity and chemical nature of volatile constituents obtained from food products, the identification of the multitude of volatiles reveals little direct insight into the actual flavor present in a food system. One reason is that many of the volatiles contribute little or nothing to the sensory character to the food product. In many cases the impact flavor compounds occur in small numbers and at low concentrations, and might not be detected by GC/MS. The discrepancy between the analytical contents and the sensory intensity of aroma chemicals led to the development of GCO, providing a link between complex chemical profiles and sensory perception. It is defined by Acree[98] as the collection of techniques utilizing humans as detectors on gas chromatographs or gas chromatographs as olfactometers to deliver pure doses of odorants to human subjects. The technique involves the sniffing of GC effluents to interpret sensory characteristics for chromatographically separated flavor compounds. The

concept behind the device is to split a portion of the effluent separated by a GC column to a ''sniff port'' while the majority of the effluent flows to a chemical detector for analysis. The effluents coming to the ''sniff port'' are sensed by a ''sniffer''. In order to reduce the discomfort caused by sniffing hot carrier gases, Dravnieks and O'Donnell[99] improved the original design by including humidifier air to cool the effluent and to reduce nasal dehydration. Years later, a higher resolution GCO system with additional features was described by Acree et al.[100] who have pioneered GCO work in the last 10–15 years.

The objective of GCO is to identify flavor compounds and to elucidate their relative impact and the characteristic flavor of a food. This is accomplished by evaluating GCO data obtained via measuring two variables: perceived odor activity and chromatographic RT.[101] RT by itself is rarely sufficient to identify an unknown chemical. It needs to be combined with other properties of the eluting chemical such as mass or Fourier transform spectra, RT on a different GC phase or odor character in order to associate chemical structure with GCO data.[101] One way of enhancing the usefulness of retention data from GCO is to convert RTs to RIs and compare the RIs with those in databases for flavor compounds. On the other hand, odor activity is a more complicated variable. It is characterized by odor quality identification, intensity assessment and potency quantification.[101] Several terms such as ''aroma value'', ''unit flavor base'', ''odor unit'', and ''odor activity value'' (OAV) have been used to describe the relationship of a compound's concentration to its odor threshold in order to assess its odor significance.[102] The potency of a flavor component relative to some other flavor components is a comparison of their dose levels at the same intensity response. It is a useful parameter to sort out the small number of contributing odorants from the very large number of odorless volatiles found in most natural products, fragrances and processed foods.[103–105]

The measurement of potency is most popularly carried out by two methods: AEDA[106] and CharmAnalysis.[100] Both methods are able to produce quantitative estimates of potency or relative potency for the various compounds that elute from a gas chromatograph.[98] The process is typified by a series of extract dilutions usually at a ratio of 1:2 or 1:3. Each dilution is sniffed until no significant odor is detected and the presence or absence of odor is recorded for each dilution. In AEDA the number of dilutions necessary to eliminate the presence of an odor at a particular RT is used to estimate the potency of a specific odor. In CharmAnalysis, a chromatogram completed with peaks and quantified peak areas (called Charm values) is generated to quantify potency. The generation of a chromatogram involves a number of steps including a series of chromatographic runs. First, it is necessary to characterize RIs for the chromatographic system. The following GCO runs of actual sample extracts at a series of dilutions, with an internal standard added, are conducted in a randomized order, to prevent any possible bias. The olfactometer delivers the effluent to a ''sniffer'' who responds to the detection of an odor by depressing a button and holding it down till the odor changes in character or disappears and then selects a descriptive word for the odor detected. All the information thus obtained is incorporated into a software called CharmAnalysis. The RT intervals of each odor are plotted to produce a square peak at a fixed height. Charm chromatographs are constructed by summing the square peaks produced at a given RI from all the samples in a dilution series and then plotting peak areas at a specified RI versus dilution value.[95] The added square areas are integrated into peak areas (charm values) by use of an algorithm[95,98] partially for the purpose of reducing the disparity stemming from the chromatographic properties of an aroma compound.[95] The peak areas, often referred to as Charm values, are believed to be proportional to the concentration of the odor-active compounds.[100] Detailed information about how to conduct a CharmAnalysis is available in Acree et al.[100]

AEDA shares a similarity with the CharmAnalysis technique. In AEDA, the GC effluent is split evenly between a flame ionization detector and a sniff port. GC injection is usually performed via the on-column technique that minimizes any chemistry associated with the injection process. RIs are also utilized to characterize the chromatography usually carried out on two different stationary phases. Both a polar and a nonpolar stationary phase are often used to chromatograph fractionated extracts in order to achieve maximum resolution of both polar and nonpolar constituents. Serial dilution usually proceeds in a continuous twofold pattern. The ratio of the concentration of the compound in the initial extract to its concentration in the most dilute extract in which the odor is detected by GCO is defined as flavor dilution (FD) factor which corresponds to the Charm value. The higher the FD factor of an odorant, the greater is its odor impact. FD values are plotted against the RI to obtain an FD chromatogram.

The primary difference between these two GCO techniques resides in the data obtained. The Charm value is based on peak area whereas the FD factor represents peak height. CharmAnalysis software was designed to request information that allows for peak reconstruction. AEDA is a simpler approach in that the GCO data acquisition is obtained solely based on the presence or absence of an odor determined by an individual.

As a powerful tool in analyzing aroma contributing components, GCO has found wide application in flavor analysis and is often used to detect trace amounts of

character impact compounds or off-odors in foods. It was used recently to investigate the characteristic lychee (*Litchi chinesis* Sonn.) fruit aroma and it was found that at least 10 compounds, including geraniol, guaiacol, *cis*-rose oxide, isobutyl acetate and furaneol, are significant odor contributors. It was concluded that the aroma of lychee is the result of interaction between compounds with floral, nutty, citrus and fruity aromas.[107] Grosch[108] has summarized the use of GCO and the character-impact odorants identified in olive oil, butter, Swiss cheese, meat, bread, beer and green tea and the study of off-flavors from several food items. A series of work using GCO to study the aromas of different types of raw and heated milk and fermented dairy products was reviewed by Friedrich and Acree.[109] Other food items analyzed by GCO involving either CharmAnalysis or the AEDA technique include grape and grape juice,[110,111] chocolate and cocoa[112] and lobster meat.[113] The list can be extended to numerous other products.[95,98] Examples of the combined use of AEDA and headspace analysis include application to mushrooms.[114]

One problem associated with AEDA and CharmAnalysis is their inconsistency with psychophysical views since both of these methods assume that the response to an odorous stimulus is linear, and that all compounds have identical response slopes with increasing concentration.[102,115] Psychophysical views are based on the principles of Stevens' law, which states that the response to a stimulus is an exponential rather than a linear event.[102] A relatively new GCO method named Osme[116] was developed and it conforms to this law. It is a quantitative bioassay method used to measure the response to odorants on a time–intensity scale.[115] Unlike AEDA and CharmAnalysis, Osme is not based on an odor detection threshold.

The Osme method has been used to identify character-impact compounds in wine, hop oils, and beer[115] and in apple flavor analysis.[102] Piggott[117] reported that Osme was more effective than odor-detection threshold techniques.[115]

The aforementioned GCO techniques all have their own characteristics and are useful in determining the quality and potency of odorants in foods. However, limitations are still present. Information is obtained for individual compounds presented to the nose outside the food matrix,[102] which influences the odor properties of these compounds. Bias can be introduced since findings by different GCO methods and their data analyses can be different in terms of which compounds are most important in a sample.[118] Therefore, aroma reconstitution to confirm GCO results by sensory comparison with the original samples is necessary. This has been done for several food items including apple,[119] cheddar cheese[120] and strawberry juice.[121]

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
AEDA	Aroma Extract Dilution Analysis
CSP	Chiral Stationary Phase
DTD	Direct Thermal Desorption
ECID	External Closed-loop Inlet Device
FD	Flavor Dilution
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GCO	Gas Chromatography/Olfactometry
GLT	Glass-lined Stainless-steel Tube
HPLC	High-performance Liquid Chromatography
HRGC	High-resolution Gas Chromatography
ICP	Inductively Coupled Plasma
IDA	Isotope Dilution Assay
L/L	Liquid/Liquid
MDGC	Multidimensional Gas Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OAV	Odor Activity Value
QA/QC	Quality Assurance/Quality Control
RI	Retention Index
RT	Retention Time
SDE	Steam Distillation and Extraction
SDGC	Single-dimension Gas Chromatography
SFC	Supercritical Fluid Chromatography
SPME	Solid-phase Microextraction
SPTD	Short-path Thermal Desorption
TLC	Thin-layer Chromatography
UHT	Ultra-high Temperature

RELATED ARTICLES

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Adulteration Determination • Atomic Spectroscopy in Food Analysis • Sample Preparation Analytical Techniques for Food • Sample Preparation, Headspace Techniques

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Mass Spectrometry **(Volume 13)**
Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications

REFERENCES

1. V.A. Yaylayan, 'Flavor Chemistry', in *Encyclopedia of Food Science and Technology*, ed. Y.H. Hui, John Wiley & Sons, New York, 955–964, 1992.
2. D.D. Bill, C.J. Mussinan, *Characterization and Measurement of Flavor Compounds*, ACS Symposium Series No. 289, American Chemical Society, Washington, DC, 1985.
3. C.-T. Ho, C.H. Manley, *Flavor Measurement*, Marcel Dekker, New York, 1993.
4. C.J. Mussinan, M.J. Morello, *Flavor Analysis*, ACS Symposium Series No. 705, American Chemical Society, Washington, DC, 1997.
5. R. Marsili, *Techniques for Analyzing Food Aroma*, Marcel Dekker, New York, 1997.
6. R.G. Buttery, L.C. Ling, 'Methods for Isolating Food and Plant Volatiles', in *Biotechnology for Improved Foods and Flavors*, eds. G.R. Takeoka, R. Teranishi, P.J. Williams, A. Kobayashi, ACS Symposium Series No. 637, American Chemical Society, Washington, 240–248, 1996.
7. C. Weurman, 'Isolation and Concentration of Volatiles in Food Odor Research', *J. Agric. Food Chem.*, **17**, 370–384 (1969).
8. H. Sugisawa, 'Sample Preparation: Isolation and Concentration', in *Flavor Research, Recent Advances*, eds. R. Teranishi, R. Flath, H. Sugisawa, Marcel Dekker, New York, 11–49, 1971.
9. M. Leahy, G. Reineccius, 'Comparison of Methods for the Isolation of Volatile Compounds from Aqueous Model Systems', in *Analysis of Volatiles. Methods. Applications*, ed. P. Schreier, Walter de Gruyter, New York, 19–47, 1984.
10. T.H. Parliment, 'Solvent Extraction and Distillation Techniques', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 1–26, 1997.
11. H.B. Heath, G. Reineccius, *Flavor Chemistry and Technology*, Van Nostrand Reinhold, New York, 1986.
12. S.T. Likens, G.B. Nickerson, 'Detection of Certain Hop Constituents in Brewing Products', *Proc. Am. Soc. Brew. Chem.*, 5–13 (1964).
13. T. Schultz, R.A. Flath, R. Mon, S. Eggling, R. Teranishi, 'Isolation of Volatile Components from a Model System', *J. Agric. Food Chem.*, **25**, 446–449 (1977).
14. H.Y. Chung, K.R. Cadwallader, 'Aroma Extract Dilution Analysis of Blue Crab Claw Meat Volatiles', *J. Agric. Food Chem.*, **42**, 2867–2870 (1994).
15. P.E. Brandt, B.H. Jensen, 'Quantitative Recovery of Volatiles from Fat and Oils by Combined High Vacuum Degassing and Thin Film Molecular Distillation', *J. Am. Oil Chem. Soc.*, **52**, 278–281 (1975).
16. B. Johnson, G. Waller, A. Burlingame, 'Volatile Components of Roasted Peanuts: Basic Fraction', *J. Agric. Food Chem.*, **19**, 1020–1024 (1971).
17. G.A. Reineccius, P.G. Keeney, W. Weisberger, 'Factors Affecting the Concentration of Pyrazines in Cocoa Beans', *J. Agric. Food Chem.*, **20**, 202–206 (1972).
18. P. Maheshwari, Z.L. Nikolov, T.M. White, R. Hartel, 'Solubility of Fatty Acids in Supercritical Carbon Dioxide', *J. Am. Oil Chem. Soc.*, **69**, 1069–1076 (1992).
19. S.S.H. Rizvi, A.L. Benado, J.A. Zollweg, J.A. Daniels, 'Supercritical Fluid Extraction: Fundamental Principles and Modeling Methods', *Food Technol.*, **40**(6), 55–65 (1986).
20. E.S. Carbonell, 'Extraction of Flavors with Supercritical Carbon Dioxide', *Cereal Foods World*, **36**, 935–937 (1991).
21. K. Kerrola, H. Kallio, 'Volatile Compounds and Odor Characteristics of Carbon Dioxide Extracts of Coriander (*Coriandrum sativum* L.) Fruits', *J. Agric. Food Chem.*, **41**, 785–790 (1993).
22. H. Brogle, 'CO_2 as a Solvent: Its Properties and Applications', *Chem. Ind. (London)*, **19**, 385–390 (1982).
23. M.-F. King, B.L. Hamilton, M.A. Matthews, D.C. Rule, R.A. Field, 'Isolation and Identification of Volatiles and Condensable Material in Raw Beef with Supercritical Carbon Dioxide Extraction', *J. Agric. Food Chem.*, **41**, 1974–1981 (1993).
24. W.G. Schultz, J.M. Randall, 'Liquid Carbon Dioxide for Selective Aroma Extraction', *Food Technol.*, **24**, 94–98 (1970).
25. C.C. Chen, R.T. Rosen, C.-T. Ho, 'Chromatographic Analyses of Gingerol Compounds in Ginger (*Zingiber officinale* Roscoe) Extracted by Liquid Carbon Dioxide', *J. Chromatogr.*, **360**, 163–173 (1986).
26. S.F. Li, S. Hartland, 'A New Industrial Process for Extracting Cocoa Butter and Xanthines with Supercritical Carbon Dioxide', *J. Am. Oil Chem. Soc.*, **73**, 423–429 (1996).
27. D.S. Gardner, 'Industrial Scale Hop Extraction with Liquid Carbon Dioxide', *Chem. Ind. (London)*, **19**, 402–405 (1982).
28. N.K. Sinha, D.E. Guyer, D.A. Gage, C.T. Lira, 'Supercritical Carbon Dioxide Extraction of Onion Flavors and Their Analysis by Gas Chromatography/Mass Spectrometry', *J. Agric. Food Chem.*, **40**, 842–845 (1992).
29. J.A. Merkle, D.K. Larick, 'Conditions for Extraction and Concentration of Beef Fat Volatiles with Supercritical Carbon Dioxide', *J. Food Sci.*, **59**, 478–483 (1994).
30. B. Meyer-Warnod, 'Natural Essential Oils: Extraction Processes and Application to Some Major Oils', *Perfum. Flavor.*, **9**, 93–104 (1984).
31. S.S. Chang, F.M. Vallese, L.S. Huang, O.A.L. Hsieh, D.B.S. Min, 'Apparatus for the Isolation of Trace Volatile Constituents from Foods', *J. Agric. Food Chem.*, **25**, 450–455 (1977).
32. T.G. Hartman, J. Lech, K. Karmas, J. Salinas, R.T. Rosen, C.-T. Ho, 'Flavor Characterization Using Adsorbent

Trapping/Thermal Desorption or Direct Thermal Desorption/Gas Chromatography and Gas Chromatography/Mass Spectrometry', in *Flavor Measurement*, eds. C.-T. Ho, C.H. Manley, Marcel Dekker, New York, 37–60, 1993.
33. T.P. Wampler, 'Analysis of Food Volatiles Using Headspace/Gas Chromatographic Techniques', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 27–58, 1997.
34. M. Güntert, G. Krammer, H. Sommer, P. Werkhoff, 'The Importance of the Vacuum Headspace Method for the Analysis of Fruit Flavors', in *Flavor Analysis: Developments in Isolation and Characterization*, eds. C.J. Mussinan, M.J. Morello, ACS Symposium Series No. 705, American Chemical Society, Washington, DC, 38–60, 1998.
35. L. Lépine, J.-F. Archambault, 'Parts-per-Trillion Determination of Trihalomethanes in Water by Purge and Trap Gas Chromatography with Electron Capture Detection', *Anal. Chem.*, **64**, 810–814 (1992).
36. C.C. Grimm, S.W. Lloyd, J.A. Miller, A.M. Spanier, 'The Analysis of Food Volatiles Using Direct Thermal Desorption', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 59–79, 1997.
37. T.G. Hartman, S.V. Overton, J.J. Manura, C.W. Baker, J.N. Manos, 'Short Path Thermal Desorption: Food Science Applications', *Food Technol.*, **45**(7), 104–105 (1991).
38. J. Manura, 'Direct Thermal Analysis Using the Short Path Thermal Desorption System', *Mass Spectrom. Source*, **12**(1), 22–27 (1991).
39. E. Ibanez, E. Ramos, J. Tabera, G. Reglore, 'Analysis of Volatile Fruit Components by Headspace SPME', *Food Chem.*, **63**, 281–286 (1998).
40. G. Paliyath, M.D. Whiting, M.R. Stasiak, D.P. Murr, B.S. Clegg, 'Volatile Production and Fruit Quality during Development of Superficial Scald in Red Delicious Apples', *Food Res. Int.*, **30**, 95–103 (1997).
41. N. Gandini, R. Riguzzi, 'Headspace Solid-phase Microextraction Analysis of Methyl Isothiocyanate in Wine', *J. Agric. Food Chem.*, **45**, 3092–3094 (1997).
42. X.G. Yang, T. Peppard, 'Solid-phase Microextraction for Flavor Analysis', *J. Agric. Food Chem.*, **42**, 1925–1930 (1994).
43. Z. Penton, 'Flavor Volatiles in Fruit Beverages with Automated SPME', *Food Testing Anal.*, **2**(3), 16–18 (1996).
44. G.R. Takeoka, R.G. Buttery, R. Teranishi, R.A. Flath, M. Guntert, 'Identification of Additional Pineapple Volatiles', *J. Agric. Food Chem.*, **39**, 1848–1851 (1991).
45. M. Guntert, H.J. Bertram, R. Emberger, R. Hopp, H. Sommer, P. Werkhoff, 'Thermal Degradation of Thiamin (Vitamin B1): A Comprehensive Survey of the Latest Studies', in *Sulfur Compounds in Foods*, eds. C.J. Mussinan, M.E. Keelan, ACS Symposium Series No. 564, American Chemical Society, Washington, DC, 199–223, 1994.
46. H. Shiratsuchi, M. Shimoda, K. Imayoshi, K. Noda, Y. Osajima, 'Off-flavor Compounds in Spray-dried Skim Milk Powder', *J. Agric. Food Chem.*, **42**, 1323–1327 (1994).
47. M. Shimoda, H. Shiratsuchi, Y. Nakada, Y. Osajima, 'Identification and Sensory Characterization of Volatile Flavor Compounds in Sesame Seed Oil', *J. Agric. Food Chem.*, **44**, 3909–3912 (1996).
48. K. Coulibaly, I.J. Jeon, 'Solid-phase Extraction of Less Volatile Flavor Compounds from Ultrahigh-temperature Processed Milk', *J. Agric. Food Chem.*, **40**, 612–616 (1992).
49. K. Grob, D. Frohlich, B. Schilling, H.P. Neukom, P. Nageli, 'Coupling of High Performance Liquid Chromatography with Capillary Gas Chromatography', *J. Chromatogr.*, **295**, 55–61 (1984).
50. W. Holscher, O.G. Vitzthum, H. Steinhart, 'Prenyl Alcohol-source for Odorants in Roasted Coffee', *J. Agric. Food Chem.*, **40**, 655–658 (1992).
51. G.A. Reineccius, S. Anandaraman, 'Analysis of Volatile Flavors', in *Food Constituents and Food Residues: Their Chromatographic Determination*, ed. J.F. Lawrence, Marcel Dekker, New York, 195–293, 1984.
52. C.-T. Ho, G.J. Hartman, 'Formation of Oxazolines and Oxazoles in Strecker Degradation of DL-Alanine and L-Cysteine with 2,3-Butanedione', *J. Agric. Food Chem.*, **30**, 793–794 (1982).
53. R.F. Hurrell, 'Maillard Reaction in Flavor', in *Food Flavors. Part A. Introduction*, eds. I.D. Morton, A.J. MacLeod, Elsevier, Amsterdam, 399–437, 1982.
54. M.-C. Kuo, Y.G. Zhang, T.G. Hartman, R.T. Rosen, C.-T. Ho, 'Selective Purge-and Trap Method for the Analysis of Volatile Pyrazines', *J. Agric. Food Chem.*, **37**, 1020–1022 (1989).
55. K. Joo, C.-T. Ho, 'Quantitative Analysis of Alkylpyrazines in Regular and Low-fat Commercial Peanut Butter Preparations', *Biosci. Biotechnol. Biochem.*, **61**, 171–173 (1997).
56. S.S. Yoo, C.-T. Ho, 'Characteristics of the Volatile Flavor Components from Peanut Butters Isolated by Selective Purge-and-Trap Method', *Food Sci. Biotechnol.*, **7**, 188–192 (1998).
57. R.E. Albert, A.R. Sellakumar, S. Laskin, M. Juschner, N. Nelson, C.A. Snyder, 'Gaseous Formaldehyde and Hydrogen Chloride Induction of Nasal Cancer in the Rat', *J. Natl. Cancer Inst.*, **68**, 597–603 (1982).
58. D.F. Smith, T.E. Kleindienst, E.E. Hudgens, 'Improved High Performance Liquid Chromatographic Methods for Artifact-free Measurements of Aldehydes in the Presence of Ozone Using 2,4-Dinitrophenylhydrazine', *J. Chromatogr.*, **483**, 431–436 (1989).
59. E.R. Kennedy Jr, R.H. Hill, 'Determination of Formaldehyde in Air as An Oxazolidine Derivative by Capillary

Gas Chromatography', *Anal. Chem.*, **54**, 1739–1742 (1982).
60. R.M. Le Lacheur, L.B. Sonnenberg, P.C. Singer, R.F. Christman, M.J. Charles, 'Identification of Carbonyl Compounds in Environmental Samples', *Environ. Sci. Technol.*, **27**, 2745–2753 (1993).
61. A. Yasuhara, K. Kawada, T. Shibamoto, 'Gas Chromatographic/Mass Spectrometric Method for Analysis of Trace Carbonyl Compounds in Foods and Beverages', *J. Agric. Food Chem.*, **46**, 2664–2670 (1998).
62. T. Hayashi, C.A. Reece, T. Shibamoto, 'A New Analytical Method for Volatile Aldehydes', in *Characterization and Measurement of Flavor Compounds*, eds. D.D. Bills, C.J. Mussinan, ACS Symposium Series No. 289, American Chemical Society, Washington, DC, 61–78, 1985.
63. K. Umano, T. Shibamoto, 'Analysis of Headspace Volatiles from Overheated Beef Fat', *J. Agric. Food Chem.*, **35**, 14–18 (1987).
64. A. Yasuhara, T. Shibamoto, 'Determination of Volatile Aliphatic Aldehydes in the Headspace of Heated Food Oils by Derivatization with 2-Aminoethanethiol', *J. Chromatogr.*, **547**, 291–298 (1991).
65. A. Yasuhara, T. Shibamoto, 'Quantitative Analysis of Volatile Aldehydes Formed from Various Kinds of Fish Flesh during Heat Treatment', *J. Agric. Food Chem.*, **43**, 94–97 (1995).
66. D.W. Wright, K.O. Mahler, J. Davis, 'The Application of Multidimensional Techniques to the Analysis of Sub-ppb Levels of Vinyl Chloride Monomer in Polyvinylchloride', *J. Chromatogr. Sci.*, **20**, 291–295 (1992).
67. D.W. Wright, 'Application of Multidimensional Gas Chromatography Techniques to Aroma Analysis', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 113–141, 1997.
68. W. Bertsch, 'Methods in High Resolution Gas Chromatography. 1. Two Dimensional Techniques', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **1**, 85–90 (1978).
69. G. Schomburg, H. Husmann, E. Huebinger, N.A. Koenig, 'Multidimensional Capillary Gas Chromatography/Enantiomeric Separations of Selected Cuts Using a Chiral Second Column', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **7**, 404–410 (1984).
70. O. Nishimura, 'Identification of the Characteristic Odorants in Fresh Rhizomes of Ginger (*Zingiber officinale* Roscoe) Using Aroma Extract Dilution Analysis and Modified Multidimensional Gas Chromatography/Mass Spectroscopy, *J. Agric. Food Chem.*, **43**, 2941–2945 (1995).
71. V.I. Homatidou, S.S. Karvouni, V.G. Dourtoglou, C.N. Poulos, 'Determination of Total Volatile Components of *Cucumis melo* L. variety Cantaloupensis', *J. Agric. Food Chem.*, **40**, 1385–1388 (1992).
72. A. Mosandl, U. Hener, U. Hagenauer-Hener, A. Kustermann, 'Stereoisomeric Flavor Compounds. 33. Multidimensional Gas Chromatography Direct Enantiomer Separation of γ-Lactones from Fruits, Foods, and Beverages', *J. Agric. Food Chem.*, **38**, 767–771 (1990).
73. A. Bernreuther, U. Epperlein, B. Koppenhoefer, 'Enantiomers: Why They Are Important and How to Resolve Them', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 143–207, 1997.
74. P. Schreier, A. Bernreuther, M. Huffer, *Analysis of Chiral Organic Molecules*, Walter de Gruyter, Berlin, 1995.
75. A. Bernreuther, N. Christoph, P. Schreier, 'Determination of the Enantiomeric Composition of Gammalactones in Complex Natural Matrices Using Multidimensional Capillary Gas Chromatography', *J. Chromatogr.*, **481**, 363–367 (1989).
76. A. Bernreuther, P. Schreier, 'Multidimensional Gas Chromatography/Mass Spectrometry: A Powerful Tool for the Direct Chiral Evaluation of Aroma Compounds in Plant Tissues. II. Linalool in Essential Oils and Fruits', *Phytochem. Anal.*, **2**, 167–170 (1991).
77. A. Mosandl, D. Juchelka, 'Advances in the Authenticity Assessment of Citrus Oils', *J. Essent. Oil Res.*, **9**, 5–12 (1997).
78. G. Full, A. Bernreuther, G. Krammer, P. Schreier, 'Foodstuff Analysis. On-line HRGC/FTIR for Direct Enantiodifferentiation of Aroma Compounds from Complex Matrices', *Labo*, **30**, 7–8 (1991).
79. E.S. Kovats, 'Gas-Chromatographische Charakterisierung Organischer Verbindung. Teil 1. Retentions Indices Aliphatischer Halogenide, Aklohole, Aldehyde und Ketone', *Helv. Chim. Acta.*, **41**, 1915–1932 (1958).
80. T. Shibamoto, 'Retention Indices in Essential Oil Analysis', in *Capillary Gas Chromatography in Essential Oil Analysis*, eds. P. Sandra, C. Bicchi, Hüthig, Heidelberg, 259–274, 1987.
81. I. Blank, 'Gas Chromatography/Olfactometry in Food Aroma Analysis', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 293–330, 1997.
82. R. Richmond, E. Pombo-Villar, 'Gas Chromatography/Mass Spectrometry Coupled with Pseudo-Sadtler Retention Indices, for the Identification of Components in the Essential Oil of *Curcuma longa* L', *J. Chromatogr.*, **760**, 303–308 (1997).
83. C.K. Huston, 'Ion Trap Mass Spectrometry for Food Aroma Analysis', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 209–235, 1997.
84. Q.N. Porter, J. Baldas, *Mass Spectrometry of Heterocyclic Compounds*, Wiley-Interscience, New York, 1985.
85. C.-T. Ho, J.T. Carlin, 'Formation and Aroma Characteristics of Heterocyclic Compounds in Foods', in *Flavor Chemistry: Trends and Developments*, eds. R. Teranishi, R.G. Buttery, F. Shahidi, ACS Symposium Series No. 388, American Chemical Society, Washington, DC, 92–104, 1989.

86. R. Schwarzenbach, D. Gubler, 'Detection of Heterocyclic Aromatic Amines in Food Flavors', *J. Chromatogr.*, **624**, 491–495 (1992).
87. D.S. Mottram, 'Flavor Compounds Formed During the Maillard Reaction', in *Thermally Generated Flavors*, eds. T.H. Parliament, M.J. Morello, R.J. McGorrin, ACS Symposium Series No. 543, American Chemical Society, Washington, DC, 104–126, 1994.
88. B. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton, FL, 1996.
89. G. Fischbock, R. Kellner, W. Pfannhauser, 'Capillary GC/FTIR and GC/MS Studies of Citrus Flavor Decomposition Processes', in *Frontiers of Flavor*, ed. G. Charalambous, Elsevier, Amsterdam, 147–167, 1988.
90. W. Pfannhauser, R. Kellner, G. Fischbock, 'GC/FTIR and GC/MS Analysis of Kiwi Flavors', in *Flavors and Off-flavors'89*, ed. G. Charalambous, Elsevier, Amsterdam, 357–373, 1990.
91. B.A. Charpentier, M.R. Sevenants, R.A. Sanders, 'Comparison of the Effect of Extraction Methods on the Flavor Volatile Composition of Shiitake Mushrooms (*Lentinus edodes*) via GC/MS and GC/FTIR', in *The Shelf Life of Foods and Beverages*, ed. G. Charalambous, Elsevier, Amsterdam, 413–433, 1986.
92. M. Kawakami, S.N. Ganguly, J. Banerjee, A. Kobayashi, 'Aroma Composition of Oolong Tea and Black Tea by Brewed Extraction Method and Characterizing Compounds of Darjeeling Tea Aroma', *J. Agric. Food Chem.*, **43**, 200–207 (1995).
93. C.Y. Chang, L.M. Seitz, E.IV. Chambers, 'Volatile Flavor Compounds of Breads Made from Hard Red Winter Wheat and Hard White Winter Wheat', *Cereal Chem.*, **72**, 237–242 (1995).
94. T.S. Chamblee, B.C. Clark, G.B. Brewster, T. Radford, G.A. Iacobucci, 'Quantitative Analysis of the Volatile Constituents of Lemon Peel Oil. Effects of Silica Gel Chromatography on the Composition of Its Hydrocarbon and Oxygenated Fractions', *J. Agric. Food Chem.*, **39**, 162–169 (1991).
95. B.S. Mistry, T. Reineccius, L.K. Olson, 'Gas Chromatography/Olfactometry for the Determination of Key Odorants in Foods', in *Techniques for Analyzing Food Aroma*, ed. R. Marsili, Marcel Dekker, New York, 265–292, 1997.
96. C. Milo, I. Blank, 'Quantification of Impact Odorants in Food by Isotope Dilution Assay: Strength and Limitations', in *Flavor Analysis: Developments in Isolation and Characterization*, eds. C.J. Mussinan, M.J. Morello, ACS Symposium Series No. 705, American Chemical Society, Washington, DC, 250–259, 1998.
97. W. Grosch, 'Detection of Potent Odorants in Foods by Aroma Extraction Dilution Analysis', *Trends Food Sci. Technol.*, **4**, 68–73 (1993).
98. T.E. Acree, 'Bioassays for Flavor', in *Flavor Science*, eds. T.E. Acree, R. Teranishi, American Chemical Society, Washington, DC, 1–20, 1993.
99. A. Dravnieks, A. O'Donnell, 'Principles and Some Techniques of High Resolution Headspace Analysis', *J. Agric. Food Chem.*, **19**, 1049–1056 (1971).
100. T.E. Acree, J. Barnard, D. Cummingham, 'A Procedure for the Sensory Analysis of Gas Chromatographic Effluents', *Food Chem.*, **14**, 273–286 (1984).
101. T.E. Acree, 'GC/Olfactometry', *Anal. Chem.*, **69**, 170A–175A (1997).
102. A. Plotto, J.P. Mattheis, D.S. Lundahl, M.R. McDaniel, 'Validation of Gas Chromatography Olfactometry Results for 'Gala' Apples by Evaluation of Aroma-active Compound Mixtures', in *Flavor Analysis: Developments in Isolation and Characterization*, eds. C.J. Mussinan, M.J. Morello, ACS Symposium Series No. 705, American Chemical Society, Washington, DC, 290–302, 1998.
103. G. Blekas, H. Guth, 'Evaluation and Quantification of Potent Odorants of Greek Virgin Olive Oils', in *Food Flavors: Generation, Analysis, and Process Influence*, ed. G. Charalambous, Elsevier, Amsterdam, 419–427, 1995.
104. H. Guth, 'Identification of Character Impact Odorants of Different White Wine Varieties', *J. Agric. Food Chem.*, **45**, 3022–3026 (1997).
105. P. Schieberle, 'Heat-induced Changes in the Most Odor-active Volatiles of Strawberries', in *Trends in Flavor Research*, eds. H. Maarse, G.G. van der Heij, Elsevier, Amsterdam, 345–352, 1994.
106. F. Ullrich, W. Grosch, 'Identification of the Most Intense Odor Compounds Formed During Autoxidation of Linoleic Acid', *Z. Lebensm. Unters. Forsch.*, **184**, 277–282 (1987).
107. P.K.C. Ong, T.E. Acree, 'Gas Chromatography Olfactory Analysis of Lychee (*Litchi chinesis* Sonn.)', *J. Agric. Food Chem.*, **46**, 2282–2286 (1998).
108. W. Grosch, 'Determination of Potent Odorants in Foods by Aroma Extract Dilution Analysis (AEDA) and Calculation of Odor Activity Values (OAVs)', *Flavour Fragrance J.*, **9**, 147–158 (1994).
109. J.E. Friedrich, T.E. Acree, 'Gas Chromatography Olfactometry (GC/O) of Dairy Products', *Int. Dairy J.*, **8**, 235–241 (1998).
110. P.A. Braell, T.A. Acree, R.M. Butts, P.G. Zhou, 'Isolation of Nonvolatile Precursors of β-Damascenone from Grapes Using Charm Analysis', in *Biogeneration of Aromas*, eds. T.H. Parliament, R. Croteau, ACS Symposium Series No. 317, American Chemical Society, Washington, DC, 75–84, 1986.
111. H.H. Baek, K.R. Cadwallader, E. Marroquin, J.L. Silva, 'Identification of Predominant Aroma Compounds in Muscadine Grape Juice', *J. Food Sci.*, **62**, 249–252 (1997).
112. P. Schnermann, P. Schieberle, 'Evaluation of Key Odorants in Milk Chocolate and Cocoa Mass by Aroma Extract Dilution Analyses', *J. Agric. Food Chem.*, **45**, 867–872 (1997).
113. K.R. Cadwallader, Q. Tan, F. Chen, S.P. Meyers, 'Evaluation of the Aroma of Cooked Spiny Lobster Tail Meat

by Aroma Extract Dilution Analysis', *J. Agric. Food Chem.*, **43**, 2432–2437 (1995).

114. R. Lizárraga-Guerra, H. Guth, M.G. López, 'Identification of the Most Potent Odorants in Huitlacoche (*Ustilago maydis*) and Austern Pilzen (*Pleurotus* sp.) by Aroma Extract Dilution and Static Headspace Samples', *J. Agric. Food Chem.*, **45**, 1329–1332 (1997).
115. R. Miranda-Lopez, L.M. Libbey, B.T. Watson, M.R. McDaniel, 'Odor Analysis of Pinot Noir Wines from Grapes of Different Maturities by a Gas Chromatography/Olfactometry Technique (Osme)', *J. Food Sci.*, **57**, 985–993 (1992).
116. M.R. McDaniel, R. Miranda-Lopez, B.T. Watson, N.J. Micheals, L.M. Libbey, 'Pinot Noir Aroma: A Sensory/Gas Chromatographic Approach', in *Flavors and Off-flavors '89*, ed. G. Charalambous, Elsevier, Amsterdam, 23–36, 1990.
117. J.R. Piggott, 'Relating Sensory and Chemical Data to Understand Flavor', *J. Sensory Stud.*, **4**, 261–272 (1990).
118. N. Abbott, P. Etievant, S. Issanchou, D. Langlois, 'Critical Evaluation of Two Commonly Used Techniques for the Treatment of Data from Extract Dilution Sniffing Analysis', *J. Agric. Food Chem.*, **41**, 1698–1703 (1993).
119. H. Young, J.M. Gilbert, S.H. Murray, R.D. Ball, 'Casual Effect of Aroma Compounds on Royal Gala Apple Flavors', *J. Sci. Food Agric.*, **71**, 329–336 (1996).
120. C. Dacremont, Z. Vickers, 'Concept Matching Technique for Assessing Importance of Volatile Compounds for Cheddar Cheese Aroma', *J. Food Sci.*, **59**, 981–985 (1994).
121. P. Schieberle, T. Hofmann, 'Evaluation of the Character Impact Odorants in Fresh Strawberry Juice by Quantitative Measurements and Sensory Studies on Model Mixtures', *J. Agric. Food Chem.*, **45**, 227–232 (1997).

Fluorescence Spectroscopy in Food Analysis

Shuryo Nakai and Yasumi Horimoto
University of British Columbia, Canada

1 Introduction 2
2 Protein Structure 2
2.1 Intrinsic Fluorometry 3
2.2 Extrinsic Fluorometry 3
3 Toxins 4
3.1 Mycotoxins 4
3.2 Toxins of Paralytic Shellfish Poisoning and Diarrhetic Shellfish Poisoning 5
4 Bacteria 5
4.1 Bioluminescence 5
4.2 Immunofluorescent Assays 5
5 Heating 6
6 Fat Oxidation 6
7 Vitamins 7
7.1 Water-soluble Vitamins 7
7.2 Fat-soluble Vitamins 8
8 Additives 8
8.1 Antibiotics 8
8.2 Aspartame 9
8.3 Salicylates 9
9 Amino Acids 9
10 Enzymes 9
10.1 Alkaline Phosphatase 10
10.2 Proteolytic Enzymes 10
11 Miscellaneous 11
Acknowledgment 11
Abbreviations and Acronyms 11
Related Articles 12
References 12

The recent application of fluorescence spectroscopy to food analysis is reviewed and future trends in fluorometry are discussed. For food proteins, two techniques, i.e. intrinsic fluorometry and extrinsic fluorometry, are contrasted. Changes in the fluorescence intensity due to tryptophan and the anisotropy were measured in food proteins, including β-lactoglobulin, α-lactalbumin and lysozyme, to estimate changes in molecular structure during environmental alterations. Major applications of extrinsic fluorometry to food proteins are surface hydrophobicity measurements using hydrophobic probes. For this purpose, this approach is currently the most popular method in food protein chemistry probably because of the simplicity of the analytical technique. Advantages and disadvantages of various fluorescence probes and their applications were compared. Toxins are one of the most appropriate applications of sensitive fluorometry. Examples of the application shown here are mycotoxins and toxins of shellfish poisoning. For the former, capillary electrophoresis was introduced as a new tool. For the latter, the methods of derivatization are critical and therefore compared. For enumerating bacterial infection, bioluminescence using luciferase and direct epifluorescent filter technique (DEFT) are being used. Further, immunofluorescence is useful for specifically detecting pathogenic bacteria such as Salmonella, Listeria *and enterotoxigenic* Escherichia coli. *Fluorescence caused by heating foods and fat oxidation was measured to assess their intensity. The thiobarbituric acid (TBA) reaction for measuring fat oxidation by colorimetry has been replaced frequently by fluorometry to improve the accuracy and specificity. Vitamins have been other popular analytes for fluorometry. Water-soluble and fat-soluble vitamins are discussed separately. High-performance liquid chromatography (HPLC) combined with fluorometric detectors is popular for the simultaneous analysis of multivitamins or multiforms of vitamins. Normal-phase HPLC of fat-soluble vitamins eliminated the need for solvent extraction and, in some cases, even the saponification process. As food additives, antibiotics (although they may be considered contaminants), aspartame and salicylates are discussed in this chapter. In amino acids analysis, reversed-phase high-performance liquid chromatography (RPHPLC) with fluorometric detectors has replaced the traditional ion exchange–ninhydrin colorimetric detector systems because of simpler, quicker elution and higher sensitivity of detection. In enzyme chemistry, alkaline phosphatase (ALP) determinations of the adequacy of milk pasteurization and proteolytic enzyme activity are the most frequently reported methods in the recent literature. A fluorometric substrate is used for the former, which converts to a fluorescent form upon loss of a phosphate radical due to the action of ALP. For the latter, the same reagent as in amino acid analysis emits fluorescence upon reaction with an α-amino group yielded from proteins by proteolysis.*

During the past one or two decades, fluorometry has in many cases replaced traditional colorimetry because of its higher sensitivity and selectivity. However, the modern automated, quicker separation technique achieved by RPHPLC has, in some cases, allowed fluorometric detectors to be replaced by ultraviolet (UV) detectors. Although HPLC fluorometric detector systems will remain

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

as sensitive, versatile methods for analytes at very low concentrations, such as toxins and antibiotics, a substantial enhancement of sensitivity is achieved by using laser-induced fluorometry. This new trend is worth considering as a replacement for traditional radiolabeling techniques.

1 INTRODUCTION

Food analysis has exploited the characteristic advantages of fluorometry, i.e. high sensitivity and specificity. However, the advent of automated liquid chromatography (LC) with superior selectivity using more economical UV detectors is slowly replacing fluorometric analysis of components such as vitamin A in food, mainly because the amount of sample is not a crucial constraint in food analysis. Despite this, a combination of chromatography and fluorometry is still advantageous for detecting extremely low concentrations of contaminants such as toxins, pathogenic microbes, and antibiotics. Another important application of fluorometry in foods is the analysis of structure changes in proteins. Structure analysis is essential and, therefore, fluorometry is a very valuable tool for elucidating structure–function relationships of food proteins, such as the capacities of emulsifying, foaming, gelation, and film-formation.

The high selectivity of spectrofluorometry is due to use of two spectra, i.e. excitation and emission spectra, while the high sensitivity of the technique is a result of measuring radiation against absolute darkness, unlike UV–visible spectrometry, where each of two radiations is compared with its own deviation, i.e. the light reduced by the absorption caused by a sample against the bright reference blank without absorption. As a result, 100–1000 times higher sensitivity than that of spectrophotometry can be expected by using spectrofluorometry. However, spectrofluorometers with the two-spectra system are more expensive than UV–visible spectrophotometers with a single-spectrum system. Since the classic work of Udenfriend in 1969,[1] many compounds in foods have been analyzed by fluorometry, namely proteins, peptides, amino acids, vitamins, carbohydrates, enzymes, steroids, some inorganic compounds, and toxins.

In 1989, the monograph entitled *Fluorescence Analysis in Foods* was published.[2] This book covered mainly fluorescence microscopy and image analysis applied for analyzing mostly raw materials used in food processing, such as grains, meat, and fish, including the detection of contaminants such as fish bones. These techniques are also important for maintaining high quality of food products.

This chapter deals with two major categories, namely spectrophotofluorometry and its use in detectors in HPLC. Spectrofluorometry is a powerful instrumental analysis for measuring mainly structure changes in food proteins, whereas HPLC with fluorometric detection is currently utilized in analyzing toxins, pathogenic microbes, heating effects, fat oxidation, vitamins, additives, amino acids, and enzyme activity in foods.

Use of laser-induced fluorescence spectroscopy for analysis of ochratoxin A (section 3.1) and vitamin (section 7) exemplifies recent advancement in laser-induced fluorescence spectroscopy with a detectability comparable to that of radiolabeling methods. Considering the dramatic progress made in modern analytical techniques, only those topics which have appeared in recent literature have been reviewed in this chapter rather than covering all possible areas of application in fluorescent analysis of foods.

2 PROTEIN STRUCTURE

The recent most thorough coverage of fluorescence spectroscopy in proteins and enzymes was made in the book edited by Brand and Johnson.[3] Readers are also recommended to refer to the article **Fluorescence Spectroscopy in Peptide and Protein Analysis** in this encyclopedia. However, in this article, the topics center on the methods that are popular in the analysis of food proteins. They involve fluorescence intensity, polarization, and energy transfer based on intrinsic as well as extrinsic fluorescences, and surface hydrophobicity using extrinsic probes.

Absorption of light by a molecule raises the energy level from the ground state to an excited state. The excited molecule returns to the ground state by emitting light or by losing energy in the form of heat (vibrational relaxation) or chemical change. The ratio of the energy emitted to the total energy absorbed is the quantum yield. The characteristic time delay between absorption and emission of a photon, usually 1–10 ns, is the fluorescence lifetime. Fluorescence can be quenched by interactions with other molecules or its environment. Energy transfer measures the transfer efficiency within the distances from 0.5 to 10 nm between energy donor and acceptor molecules. Fluorescence of a ligand, e.g. a dansyl group, increases upon binding to protein. This effect is caused either by a change in the polarity of the microenvironment, or by an energy transfer between tryptophan residue(s) in the protein and an acceptor in the ligand. When polarized light is used for excitation, the plane of polarized light can be altered by a fluorophore. The extent of depolarization is expressed as polarization, $p = (I_v - I_h)/(I_v + I_h)$, or anisotropy, $r = (I_v - I_h)/(I_v + 2I_h)$, where I_v and I_h are vertical and horizontal

fluorescence intensities respectively. The fluorescence anisotropy is a measure of the inequality in molecular movement in solution.

For food proteins, an excellent review was published by Strasburg and Ludescher,[4] who have provided a comprehensive explanation of the principles of the different fluorescent technologies that have been utilized in food analysis.

2.1 Intrinsic Fluorometry

Tryptophan is the most widely used intrinsic probe in proteins. The emission peak is at about 350 nm in water but shifts to about 315 nm in nonpolar media, such as within hydrophobic regions in the interior of the folded proteins. Protonated amino and carbonyl groups, as well as amide bonds, readily quench tryptophan fluorescence, with a quantum yield typically in the range 0.1–0.5. Changes in fluorescence intensity due to variation in quencher proximity provide a sensitive indicator of alterations in protein conformation, such as those resulting from protein–ligand binding, protein–protein associations or protein denaturation. Changes in intrinsic anisotropy can be utilized to monitor shape variations derived from protein unfolding, mass changes resulting from aggregation, or conformational transitions associated with a change in a particular protein domain. When proteins do not contain tryptophan, tyrosine residue(s) can be used instead as in the case of tryptophan residues.[4]

The two tryptophan residues, i.e. Trp19 and Trp61, in β-lactoglobulin are inaccessible to water by being buried in hydrophobic environments in the protein interior.[5] Changes in the fluorescence intensity and emission wavelength have been used to monitor the denaturation. The protein undergoes reversible conformational changes at 50 °C, thereby affecting one Trp residue, while a second irreversible change occurs at 70 °C affecting another residue. Shifts in pH affected the anisotropy,[6] and the relaxation times for rotational motion were in good agreement with those expected for the monomer, dimer, and octamer. The supramolecular associative properties of β-lactoglobulin were investigated by measuring concentration dependence of the accessibility of backbone exchangeable amide proton by H/D exchange in nuclear magnetic resonance (NMR) and by their intrinsic fluorescence features.[7] The weak forces involved in this association were different from those stabilizing the native dimer.

Binding of retinol to β-lactoglobulin has been measured using retinol fluorescence through energy transfer from the intrinsic Trp.[8] Similarly, the binding with phosphatidylcholine has been reported.[9] Tryptophan fluorescence of chicken lysozyme and α-lactalbumin in native and modified forms and fragments of these proteins has been investigated using the iodide ion as a quencher.[10] Exposure of Trp residues in both proteins has been determined. Effects of heating through a Maillard reaction on β-lactoglobulin–lactose mixtures have been investigated by monitoring the formation of fluorosine and the decrease in the Trp fluorescence quantum yield.[11]

2.2 Extrinsic Fluorometry

Rhodamine maleimide is being used to label the calcium-binding protein calmodulin to investigate the binding with channel protein using fluorescence polarization.[12] This interaction is likely to play an important role in the mechanism of pale, soft exudative pork.

Another covalent extrinsic probe used is Iaedans [5-({[(2-iodacetyl)amino]ethyl}amino) naphthalene-1-sulfonic acid], which binds with sulfhydryl groups in muscle proteins.[13] By observing the variations in fluorescence intensity, conformational transitions of the proteins due to environmental changes can be assessed.

Most other studies have been conducted by using noncovalently binding extrinsic probes to investigate the relationships with protein functionality.[14,15] 8-Anilinonaphthalenesulfonate (ANS) has been the most popular hydrophobic probe for the determination of surface hydrophobicity of proteins. Occasionally, its dimeric form, bis-ANS, which has a greater quantum yield in nonpolar environments as a result of binding with proteins stronger than the monomer ANS, is used for the same purposes. These effects permit observation of depolarization by energy transfer among the bound fluorophores, which can be used to estimate the distribution of the ligands among the protein molecules.

Cardamone and Puri[16] stated that ANS binding and the resultant K_a measured by a Scatchard or Kloz plot and, to a lesser extent, quantum yield may be used as a measure of the relative surface hydrophobicity of proteins. Titration of protein solutions containing increasing concentrations of the fluorescent probe can provide information on both the number and the affinity of binding sites. This may be useful in determining whether a high intensity of fluorescence results from the presence of many binding sites of only moderate hydrophobic character, or from the existence of a high-affinity site with considerable hydrophobic character.

The application of fluorescence probes to investigate the proximity of the probe binding site(s) and aromatic groups has also been assessed by measuring the efficiency of energy transfer in casein micelles.[17] Transfer of excitation energy from aromatic chromophores to bound probe at an adjacent hydrophobic site causes quenching of the intrinsic fluorescence in the aromatic side chain. This transfer efficiency has also been measured as an index of surface hydrophobicity.

Another group of anionic fluorescence probes is the fatty acid analog type, including *cis*-parinaric acid (CPA)

and *trans*-parinaric acid, which has been used as a probe for proteins and biological membranes.[18] The parinaric acids are among the few known nonaromatic fluorophores. Their similarity to native fatty acids, nonfluorescence in water, and good Stokes' shift characteristics are among the advantages of their use to probe for hydrophobic regions that may be important in protein–lipid interactions in food systems. Good correlations were obtained between the relative hydrophobicity values of proteins determined by CPA fluorescence and the properties related to protein–lipid interactions, such as interfacial tension and emulsifying activity.[19] This is an aliphatic hydrophobicity that can be differentiated from the aromatic hydrophobicity by using aromatic fluorophores, such as ANS.[20] The latter may be useful in investigating protein–aromatic interactions, such as those with retinol and aromatic flavor compounds. ANS and CPA were both used for the elucidation of heat-induced gelation of myosin.[21]

A limitation in using anionic probes such as ANS and CPA to determine protein hydrophobicity includes a possibility that electrostatic as well as hydrophobic interactions may contribute to the interaction with probes.[22] The use of neutral, e.g. prodan (6-propionyl-2-dimethylaminonaphthalene),[23] or uncharged probes, e.g. 1,6-diphenyl-1,3,5-hexatriene (DPH),[24] may circumvent this problem. However, the nonpolar nature of DPH restricts its solubility in aqueous systems and thus limits its use as a probe for protein solutions, although it is not the absolute deterrent.[24]

Binding with hydrophobic probes was utilized to investigate the effects of heating milk proteins[25] and soy proteins.[26] Not only ANS binding, but also intrinsic tryptophan fluorescence, was used for the same purpose in α-lactalbumin,[27] β-lactoglobulin,[28] and gluten.[29] The effects of heating, the addition of dissociating agents, and S–S cleavage on the emulsifying properties of soy proteins were studied by using CPA hydrophobicity.[30] Similarly, the effects of pH and NaCl on the emulsifying and foaming properties of cowpea globulin have been investigated.[31] The importance of surface hydrophobicity in rennet coagulation of milk proteins was discovered by using fluorometry.[17,32]

3 TOXINS

Fluorometry is one of the most valuable instrumental analytical techniques for determining causes of food poisoning by analyzing toxins. More powerful tools such as NMR, gas chromatography (GC)/mass spectrometry (MS), supercritical fluid chromatography (SFC)/MS, atmospheric pressure ionization (API)/LC/MS, MS/MS, GC/Fourier transform infrared (FTIR), and SFC/FTIR can also be used for measuring and/or identifying toxins.[33] However, HPLC with fluorometric detection is still the analytical tool of choice for many researchers, probably because of lower installation costs and the versatility of the instrument for multianalytical purposes.

3.1 Mycotoxins

Aflatoxins are mycotoxins (fungal toxins) are named from "*A. fla*"*vus* derived "toxin". They are toxic secondary metabolites produced by fungi, primarily of the genus *Aspergillus*, e.g. *Asp. flavus* and *Asp. parasiticus*. These toxins are found as natural contaminants in foods and feeds, including peanuts, corn, and other grains. Aflatoxins are converted metabolically in the bodies of humans and domestic animals to potent liver carcinogens.[34] Because of their strong carcinogenic nature, aflatoxins are regulated at much lower levels than any other mycotoxins, usually at <20 ppb. Cole and Cox[35] reported 16 aflatoxins. However, only aflatoxins B_1 and B_2, which exhibit a blue fluorescence; G_1 and G_2, which show a yellow–green fluorescence; and M_1, which is found in milk as a metabolite produced in animals fed aflatoxin B_1, are routinely monitored in foods and feeds. The fungi can produce aflatoxins in their natural habitat under stress conditions or in storage under moist conditions (85%) at 25–40 °C.

Aflatoxin B_1 is the most prevalent, toxic, and carcinogenic of all the aflatoxins. Because of the potential health hazards for humans, analysis of aflatoxins in various foods is important from the food safety aspects. Efforts to minimize aflatoxin contamination in susceptible commodities are the subject of many food research projects.

Thin-layer chromatography (TLC) has long been the most popular method for analysis of aflatoxins. However, owing to higher accuracy in quantification, HPLC, especially RPHPLC, has replaced TLC. Native fluorescence at $\lambda_{ex} = 360$ nm for all four aflatoxins ($\lambda_{em} = 440$ nm for aflatoxins B_1 and B_2 and 470 nm for aflatoxins G_1 and G_2) has been used for detection of aflatoxins after separation using HPLC. A UV detector at 365 nm is also utilized to monitor the effluents from the chromatography column, but a fluorescence detector is required to improve the accuracy of detecting aflatoxins B_2 and G_2.

The weak native fluorescence of aflatoxins may be intensified by using, for instance, trifluoroacetic acid. Acid-catalyzed precolumn derivatization converts aflatoxins B_1 and G_1 to the corresponding saturated hydroxy derivatives B_{2A} and G_{2A}. Another method of enhancing fluorescence is the postcolumn derivatization of aflatoxin B_1, which is achieved by the action of iodine or bromine, thereby increasing the fluorescence 25-fold.

Kok et al.[36] described the optimum operating conditions of the electrochemical KOBRA™ cell to generate bromine for the same purpose and applied this technique to the determination of aflatoxins in cattle feed. A reversed-phase (RP) C_{18} column was used, with a mobile phase of water–methanol–bromide including nitric acid. The aflatoxins were monitored using a fluorescence detector at $\lambda_{ex} = 360\,nm$ and $\lambda_{em} > 420\,nm$. The detection limits were reported to be 0.04 ng for aflatoxins B_1 and G_1, and 0.02 ng for B_2 and G_2. Aflatoxin M_1 in goat dairy products was analyzed using HPLC with a fluorescence detector after the sample had been cleaned by using an immunoaffinity column.[37] The mycotoxin was detected in about 30% of samples from dairy farms at levels ranging from 4 to 37 ng l^{-1} for milk and 19–160 ng kg^{-1} in cheeses.

Capillary electrophoresis with a laser-induced fluorescence detector was used to determine ochratoxin A in roasted coffee, corn, and sorghum.[38] This fungal toxin is one mycotoxin that is routinely monitored due to its high toxicity and widespread occurrence in foods and feeds. The extraction and isolation procedures are carried out using a silica column and an immunoaffinity cleanup column analogous to other chromatographic methods for aflatoxins.

Capillary electrophoresis has also been used in corn samples to analyze fumonisin B_1, a mycotoxin produced by species of *Fusarium*. As this toxin did not have a UV chromophore, it was fluorescein-labeled.[39] Zearalenone is another toxin produced by *Fusarium* molds, which grow on corn that is immature or has a high moisture content at harvest. After separating by RPHPLC on a C_{18} column, it was determined fluorometrically.[40] Despite the higher costs of instrumentation, HPLC/MS is more advantageous due to greater sensitivity and selectivity, without the need for derivatization as in fluorometry.

3.2 Toxins of Paralytic Shellfish Poisoning and Diarrhetic Shellfish Poisoning

Paralytic shellfish poisoning (PSP) toxins include saxitoxin, neosaxitoxin and gonyaitoxins, which show neither UV absorption nor fluorescence. Oxidation under alkaline conditions, e.g. with $NaOH-H_2O_2$, converts most of PSP toxins to fluorescent forms, which can be used in postcolumn derivatization after eluting from RPHPLC columns. The tolerance levels of PSP are 40–80 µg per 100 g, and HPLC can detect individual PSP toxins below the 1 µg per 100 g level.

Diarrhetic shellfish poisoning (DSP) toxins, including okadaic acid and its derivatives, are named dinophysistoxins (DTX), pectenotoxins (PTX), and yessotoxin (YTX). To make fluorometric detection feasible, two derivatizations have been proposed using the reactions with 9-anthryldiazomethane (ADAM) and 4-bromomethyl-7-methoxycoumarin (Br-Mme). Since ADAM derivatives require an additional cleanup process and are unstable, the Br-Mme derivatization has been more popular, thus being extensively used. A convenient one-vial procedure using Br-Mme for esterification of okadaic acid and DTX-1 to give a fluorescent coumarin ester at $\lambda_{ex} = 325\,nm$ and $\lambda_{em} = 390\,nm$ has been developed.[41] For HPLC of the Br-Mme derivatives, the use of a C_{18} RP column and isocratic elution with acetonitrile–water (70 : 30) has been suggested. Contamination levels of 0.03–2.5 mg DSP kg^{-1} have been reported.[42] Proposed tolerance levels are 20–60 µg per 100 g and the application of HPLC to the determination of the acidic components of DSP complex allows the detection of 10 µg of okadaic acid and/or DTX-1 in 100 g of shellfish meat.

4 BACTERIA

According to the literature, fluorometry to detect bacteria-infected foods either relies on their bioluminescence or utilizes immunofluorescence assays.

4.1 Bioluminescence

Bioluminescence is used in food analysis to detect bacterial adenosine triphosphate (ATP). In the presence of ATP and magnesium ions, luciferin is oxidized to oxyluciferin, which releases a photon by the action of the enzyme luciferase. The number of photons of light emitted is proportional to the amount of ATP initially present in the sample.[43] Luciferase reactions are rapid and the emitted light can be measured by a luminometer and the results converted to the number of bacterial cells per ml of sample. A number of automated luminometer systems are commercially available for assessing bacterial ATP,[44] and sensitivities range from 10^3 to 10^4 cells ml^{-1}.

Another fluorescence system to enumerate viable bacteria is the DEFT.[43] The monomeric fluorochrome acridine orange binds with double-stranded DNA, thereby emitting a green fluorescence. However, when cells are actively dividing, the fluorescence changes to orange as the binding is to single-stranded RNAs produced as a result of bacterial division. In the DEFT procedure, bacterial cells in food samples are pretreated with trypsin, filtered through a polycarbonate membrane, stained with acridine orange and then the number of orange cells on the filter is counted under a fluorescence microscope.

4.2 Immunofluorescent Assays

Antibodies can be labeled with fluorescein isothiocyanate (FITC) to identify bacteria. *Salmonella* serotypes in foods

were identified by this method,[45] which was also applied for the detection of *Listeria monocytogenes* in food samples.[46]

A solid-phase fluorescent capillary immunoassay was developed to detect *Escherichia coli* O157 : H7 in foods.[47] A soft glass capillary tube served as the solid support, to which heat-killed *E. coli* O157 : H7 were adsorbed. Polyclonal anti-*E. coli* O157 : H7 antibody conjugated with biotin was used, and the bound antigen–antibody complex was detected using avidin labeled with Cy5, a fluorescent cyanine dye. Any *E. coli* O157 : H7 in the sample would compete with the formation of this complex, thereby reducing fluorescence. This assay was tested for sensitivity with spiked ground beef and apple cider samples. The minimum detectable number of cells present in the initial inoculum was approximately 1 colony forming units (CFU) per 10 g of ground beef when the samples were enriched by incubating for 7 h at 37 °C. In the case of apple cider, *E. coli* cells were collected using immunomagnetic beads, and then the fluorescent immunoassay was performed as in the case of beef samples.[47]

5 HEATING

Protein-bound Maillard compounds were hydrolyzed with proteinase or acid and separated by ion-exchange chromatography. Effluents were detected directly with a fluorescence detector.[48] Fluorescent peptides resulting from casein/lactose interaction in milk by heating were determined by RPHPLC.[49] Heterocyclic aromatic amines are mutagenic compounds that are formed during heating of meat and fish. These substances are produced by the reaction of creatine with amino acids and carbohydrates. One of those compounds, i.e. 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) is fluorescent and can be separated by HPLC.[50]

Fluorescence was the technique that provided the best models for anisidine value, oligomers, iodine value, and vitamin E in deteriorated oil after repeated frying cycles.[51] Algal phycobiliprotein, fluorescent phycoerythrin, can be applied as the detecting component of a time–temperature integrator for monitoring thermal processing of foods.[52]

6 FAT OXIDATION

The methods to measure lipid oxidation are divided into two categories: primary changes and secondary changes.[53] The ideal method to measure lipid oxidation is the direct measurement of initial products of oxidation, i.e. hydroperoxides. Recently, much research has been conducted into the association between oxidation and clinical disease. Therefore, research has been focused on the determination of hydroperoxides rather than secondary changes.[54] The analytical procedure to be used for this purpose should be sensitive, precise, and specific.

Lipid oxidation is one of the major causes of quality deterioration in meat and cooked meat products during storage. Gray and Monahan[53] reviewed the measurement of lipid oxidation in meat and meat products. Several methods have reported the extent of lipid oxidation in freeze-dried meat and fish using fluorescence techniques. Recently, Hasegawa et al.[55] reported a procedure for evaluating oxidative deterioration in dried fish by measuring protein-bound fluorescence.

Detection of lipid hydroperoxides by HPLC has been reviewed by Yang.[54] Two postcolumn detection methods were compared, i.e. the chemiluminescence of luminol and the spectrophotometric determination of I_3^-. Miyazawa et al.[56] measured lipid hydroperoxides in meats and fish using the fluorometric peroxygenase assay. Hydroperoxide is reduced to alcohol in equimolar concentration in a reaction catalyzed by peroxygenase during hydroxylation of the substrate, 1-methylindole.[56] The method can determine total lipid hydroperoxides in a sample without extraction.

The flow injection analysis (FIA) of Akasaka et al.[57] was highly sensitive. After sample injection, lipid hydroxides were reacted with diphenyl-pyrenylphosphine (DPPP) in a stainless steel coil, and the fluorescence intensity of hydroperoxide showed good linearity between 0.4 and 79 pmol with a detection limit of 0.2 pmol.[57]

The primary products from lipid oxidation decompose to form secondary products. Among these secondary products, malondialdehyde (MDA) has been used as an index of oxidation in foods.[58] The 2-thiobarbituric acid (TBA) method is most popular for the measurement of MDA, despite a criticism for lack of specificity and sensitivity. Recently, chromatographic methods (GC or HPLC) have provided more specific and sensitive determination of MDA than the TBA method. Samples of homogenized butter, margarine, oil, fish, and meat tissue were treated with 5% trichloroacetic acid to remove protein. The supernatant was incubated in 0.28% TBA at 90 °C for 30 min. The MDA/TBA adduct thus produced was detected by measuring fluorescence (λ_{ex} = 515 nm, λ_{em} = 543 nm) after eluting from the RPHPLC column.[59]

Even though the chromatographic method is specific and sensitive, the TBA method is still frequently preferred over chromatographic methods because of its simplicity. The TBA method was reviewed by Hoyland and Taylor.[60] Raharjo et al.[61] reported an improved

TBA method for measuring lipid peroxidation in beef. Substances interfering with the MDA/TBA reaction were removed by using a Sep-Pak C_{18} cartridge. The method was adequately fast, specific, and sensitive.

Raharjo and Sofos[58] reviewed procedures for the analysis of MDA in muscle tissue. Several modified TBA procedures were discussed in their review article. Jo and Ahn[62] developed a sensitive and reliable fluorometric analysis of TBA reactive substances in turkey meat. This method is useful for samples with low-level lipid oxidation.

7 VITAMINS

Vitamin determination in food analysis is very important in efforts to minimize nutrient losses before and after processing,[63] to fortify foods with vitamins,[64] or to use vitamins as food additives such as antioxidants.[65]

Microbiological, biological, and enzymatic assays have been applied to determine vitamin concentration.[66] In recent years, HPLC in the RPHPLC mode or normal-phase mode has been widely applied for determination of both fat-soluble and water-soluble vitamins.[66] HPLC provides a wide range of separation and detection methods. Fluorescence detection provides improved sensitivity and selectivity compared with UV detection as already discussed. Postcolumn derivatization methods have frequently been employed in the HPLC determination of vitamins in foods.

Rahn[67] submitted excellent summary tables for fluorescent analysis of vitamins in foods. According to Rahn, there are two types of fluorescence that can be used for analysis, i.e. natural fluorescence of vitamins A, B_2, folic acid, and E, and the fluorescence of vitamins C and K_3 derived from condensation with *o*-phenylenediamine (OPD). His table includes vitamins A, B_1, B_2, B_c (folic acid), B_3 (nicotinamide), B_5, B_6, B_{12}, C, D, E, K_1 (phylloquinone), and K_3 (menadione). Another table summarizes the results of laser-induced molecular fluorescence.[68] Several vitamins and amino acids could be detected at the parts per trillion level, indicating a tremendous improvement of several orders of magnitude over the detection limit achieved by conventional light sources.

7.1 Water-soluble Vitamins

Water-soluble vitamins (B_1, B_2, and C) have been separated by RPHPLC and measured by using fluorescence detection.[66] Vitamin B_1 consists of a substituted pyrimidine linked by a methylene group to a substituted thiazole.[63] In analysis, thiamine is usually treated with a strong oxidizing agent to form fluorescent thiochrome. Vitamin B_2 occurs naturally in three principal forms, i.e. riboflavin (RF), flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD).[69] FMN and FAD are converted to RF, and total riboflavin (TRF) is assayed fluorometrically. However, this method is tedious, and time-consuming, and needs laborious sample preparation.[69]

Sims and Shoemaker[70] reported simultaneous determination of thiamine and RF using HPLC with fluorescence detection ($\lambda_{ex} = 370\,nm$, $\lambda_{em} = 435\,nm$ for thiochrome, $\lambda_{ex} = 370\,nm$, $\lambda_{em} = 520\,nm$ for RF). Fernando and Murphy[71] also measured thiamine and RF in soybeans and tofu using wavelengths $\lambda_{ex} = 364\,nm$, $\lambda_{em} = 436\,nm$ for thiochrome, and $\lambda_{ex} = 436\,nm$, $\lambda_{em} = 535\,nm$ for RF. Russell and Vanderslice[72] studied non-degradable extraction and simultaneous quantification of RF, FMN, and FAD by HPLC. The method was successfully applied to a variety of foods, including raw and cooked meats, dairy products, eggs, and cereal products. The results of these HPLC analyses generally agreed with the published values for TRF. Russell and Vanderslice[69] modified the standard fluorometric determination of RF ($\lambda_{ex} = 440\,nm$, $\lambda_{em} = 565\,nm$) to permit more convenient quantification by FIA.

An HPLC method with fluorometric detection ($\lambda_{ex} = 333\,nm$, $\lambda_{em} = 375\,nm$) was applied for analysis of vitamin B_6 in foods and feeds.[73] High recovery and precision were reported in comparison with a microbiological assay using *Saccharomyces uvarum*. Simultaneous determination has been reported for the measurement of thiamine, RF, and pyridoxine in medical foods[74] and in infant formula.[75] A fluorometric detector was programmed to detect B_2 and B_6 simultaneously, and then thiochrome. RPHPLC was used for thiamine, RF, and pyridoxine in medical foods.[74] Vitamin B_1 was postcolumn derivatized to thiochrome. Vitamin B_{12} was labeled with a fluorescent reagent [6,7-dimethoxy-1-methyl-2(1*H*)-quinoxaline-3-propionylcarboxylic acid hydrazide] and determined over a range of $0.2–10\,\mu g\,L^{-1}$ of fluid samples of foods.[76]

The reduced form and the oxidized form of vitamin C are referred to as ascorbic acid and dehydroascorbic acid, respectively. The total vitamin C activity is the sum of both forms.[77] Precolumn or postcolumn derivatization can enhance the sensitivity of dehydroascorbic acid measurements by fluorescence detection. Ascorbic acid has been measured based on the rapid oxidation of ascorbic acid, which then reacts with OPD to form a fluorescent quinoxaline. Total vitamin C was measured by HPLC with precolumn derivatization with OPD.[77] This method could analyze various foods, including complex matrices of canned corn, potatoes, green beans, potato chips, and cereals, by avoiding the effect of interference

matrix. Simultaneous determination of ascorbic, dehydroascorbic, isoascorbic, and dehydroisoascorbic acids in meat-based food products was reported using RPHPLC with a postcolumn fluorescence detector.[78] Similarly, HPLC/fluorescence methods were developed for vitamin K_1[79] and folate in foods.[80]

7.2 Fat-soluble Vitamins

Numerous HPLC procedures have been reported for fat-soluble vitamins. They are extracted with nonpolar solvents after saponification and separated by either normal or RPHPLC.[66] The procedures have been simplified by eliminating extensive sample preparation and interference.

A simple and rapid method for the determination of retinol in whole milk has been described.[81] It involves sample preparation, saponification, and extraction in the same tube, thereby avoiding the risky and time-consuming evaporation of solvent. The retinol is then separated by HPLC with fluorescence detection ($\lambda_{ex} = 344$ nm, $\lambda_{em} = 472$ nm). The method is rapid, sensitive, and precise for milk samples of small sizes.

Vitamin E is found in nature in the form of a number of tocopherols and tocotrienols. Analytical procedures involve extraction, saponification, and chromatography of the nonsaponifiable matter by TLC or GC.[66] LC with fluorescence detection has been widely utilized for the determination of vitamin E activity in various foods. Ang et al.[82] used RPHPLC with fluorescence detection ($\lambda_{ex} =$ 275 and 280 nm, $\lambda_{em} = 330$ nm) for measurement of the level of different forms of tocopherol in chicken breast and leg muscles and to assess the effect of cooking on the stability of these tocopherols. Razagui et al.[83] reported a spectrofluorometric method to measure α-tocopherol for detecting nutritional supplement products ($\lambda_{ex} = 298$ nm, $\lambda_{em} = 330$ nm). The procedure was reported to be very accurate and without interference.

An intercomparison of the methods for vitamin determination was organized in 1993 involving 18 European laboratories.[84,85] HPLC with a fluorescence (sometimes UV) detector was used to analyze retinol, β-carotene and α-tocopherol in milk powder, pork muscle, and haricot vert beans. Relative standard deviations of reproducibility (RSDR) were 10–16% except for β-carotene in haricot vert beans (52%). RSDR of vitamins B_1, B_2, B_6 and C were 9–18%, except for vitamin B_2 (28–74%) and B_6 (18–51%).

Normal-phase HPLC with a silica column was used to analyze vitamins A and E and β-carotene in dairy foods[86] and vitamin E in beef muscle.[87] Lipids extracted with organic solvents were directly applied to the column without saponification for the former, while saponified extracts were applied for the latter.

8 ADDITIVES

Food additives are used to improve flavor, color, texture, and shelf life. Among the food additives, fluorometric analysis of antibiotics, aspartame and salicylates has appeared in recent literature. Antibiotics might not belong to this category in a strict sense, but as they have been inadvertently contained in foods, it was decided to discuss them here.

8.1 Antibiotics

Antibiotics have been used mainly in dairy herds and fish farming. Antibiotics can cause allergic reactions in humans, therefore it is important to detect them in order to prevent health hazards from contaminated foods.

Rapid screening tests using commercial test kits are available for the detection of antibiotics.[88] They are microbial inhibition assays (Delvotest P, Delvotest SP, and AIM'96) and enzyme receptor tests (SNAP, LacTek, Charm II, Delvo-X-Press BLII, BetaScreen 4P, Parallux, EZ-Screen and Ridscreen). The limitation of these methods is sometimes their lack of specificity. Among these test kits, BetaScreen 4P and Parallux use fluorescence as the detection mechanism. BetaScreen adopts an enzyme-linked immunosorbent assay (ELISA)-based test using fluorescence and ALP.[88,89] A novel qualitative enzyme immunoassay based on fluorescence detection was used to determine β-lactam antibiotics in milk.[90] The assay is specific to six β-lactams. Parallux is based on a solid-phase fluorescence immunoassay (SPFIA) and can detect up to six β-lactam antibiotics.[88]

LC is one of the most reliable techniques for detecting traces of antibiotics. Payne et al.[90] determined eprinometin marker residue in edible bovine tissue using HPLC with fluorescence. It is rapid, sensitive, and selective. Sarafloxacin residues in chicken eggs were analyzed by on-line microdialysis and HPLC with programmable fluorescence detection.[91] Ang and Luo[92] described the precolumn derivatization procedure using HPLC with fluorescence detection for rapid determination of ampicillin in raw and pasteurized milk. Amoxicillin and ampicillin in milk were determined simultaneously by HPLC after derivatization.[93] Four fluoroquinolones in milk were measured by HPLC with fluorescence detection.[94] The limit of detection was 1.8 μg penicillin G kg^{-1}. Similarly, spectinomycin and penicillin G were determined by a rapid detection assay.[95] Also, tetracycline and oxytetracycline were determined in meat and fish in a similar fashion.[96] Oxytetracycline was detected by this method at the level of 5 ng g^{-1} in meat and fish, and 1 ng g^{-1} in eggs. The coefficient of determination (relative standard

deviations) at 50 ng g^{-1} and 200 ng g^{-1} ranged from 1.6% to 3.1%.

Sulfonamides (sulfadiamine, sulfathiazole, sulfethazine, sulfamonomethoxine, sulfamethoxazole, sulfadimethoxine) were analyzed using LC with fluorescence.[97] The method was rapid and simple. Laloux et al.[98] also determined sulfonamides in milk after fluorescamine derivatization. Gerhardt et al.[99] reported a procedure for analyzing streptomycin and dihydrostreptomycin in milk. These antibiotics were detected by fluorescence following postcolumn derivatization with 1,2-naphthoquinone 4-sulfonic acid.

Aquaculture has magnified concern about drugs used in fish farming. LC with fluorescence has been used for the detection of different antibiotics, e.g. quinolones and sulfonamide. Pfenning et al.[100] determined residues of flumequine and nalidixic, oxolinic, and piromidic acids in salmon and shrimp using LC with simultaneous UV and fluorescence detection. Amoxicillin in catfish and salmon tissues was determined by RPHPLC with precolumn formaldehyde derivatization.[101] Gradient LC with fluorescence detection was also used to determine sarafloxacin residues in channel catfish muscle tissue.[102] Sulfonamides in salmon tissue were measured by LC with postcolumn derivatization and fluorescence detection.[103]

8.2 Aspartame

Aspartame, *N*-L-α-aspartyl-L-phenylalanine 1-methyl ester, is a low-calorie sweetener which is 200 times sweeter than sugar. Aspartame is used to sweeten or enhance fruit flavor in a variety of foods and beverages. Aspartame is composed of two amino acids, aspartic acid and the methyl ester of phenylalanine. Upon digestion, aspartame breaks down into aspartic acid, phenylalanine and methanol.

Speedy and simple analysis of aspartame is required for quality control purposes. Many analytical procedures have been employed to determine the concentration of aspartame and its decomposition products, which have been formed during manufacturing and storage of food products. One of these analytical methods is a spectrophotometric method.[104] However, LC is most commonly used for the determination of aspartame.[105] Sample preparation is minimal in labor for liquid samples, while solid samples are extracted with water and then injected onto the column. In another paper, it was reported that the detection limit by measuring native fluorescence was 0.01 mg l^{-1} of diet soft drinks.[106]

8.3 Salicylates

Salicylates, which are esters or salts of salicylic acid, are natural compounds made by a variety of plants. They compete with aspirin for the same cellular binding sites in the body, providing similar actions to aspirin. They relieve pain and fever, but at the same time they can cause health risks such as intestinal bleeding or allergic reactions. A diet rich in fruits and vegetables may help prevent heart attacks as well as some cancers by providing healthy doses of salicylates. Therefore, there is an interest in the salicylate levels of both natural foods and food additives.

TLC has been used for qualitative analysis of salicylates in food. LC with UV detection or fluorescence detection has been used to monitor salicylates in food.[107] While HPLC with UV detection suffered from interference problems, HPLC with fluorescence detection was sensitive and specific.

9 AMINO ACIDS

To analyze the amino acid composition of proteins, the protein hydrolysate has been applied to a cation-exchange column and ninhydrin derivatives of amino acids in the effluent are measured colorimetrically using a spectrophotometric detector. However, RPHPLC with a C_{18} or C_8 column before or after derivatization of amino acids with *o*-phthaldialdehyde (OPA) is convenient for detection using a fluorometric detector.[108] This is because of simpler solvent systems required for eluting the amino acids from the column and quicker elution (less than 30 min) than ion-exchange chromatography. The phenylthiohydantoin (PTH)/RPHPLC can accurately quantify 5 pmol of 50-kDa protein, 2–3 times more sensitively than ion-exchange/ninhydrin detector systems.

Progress has recently been made in RPHPLC after precolumn derivatization. Rapid derivatization and direct injection without cleaning is now feasible by using 6-aminoquinolyl-*n*-hydroxysuccinidyl carbamate to derivatize for fluorometric detection.[109] By using phenylisothiocyanate for derivatization, a UV detector that is more popular on HPLC instruments can be used with similar detectability to fluorometric detectors.[110] For the first time, detection of femtomolar amounts in various foods has been achieved by fluorometric detection of ninhydrin derivatives of amino acids.[111]

10 ENZYMES

Despite the great potential of a broader application of fluorometry to enzyme assays, ALP and protease are the only enzymes reported in the recent literature concerning food analysis.

10.1 Alkaline Phosphatase

Proper pasteurization is a crucial process in the dairy industries. ALP (EC 3.1.3.1) is one of the indigenous enzymes found in raw milk.[112] ALP activity has been used as an indicator for determining the adequacy of pasteurization of milk. ALP in milk is inactivated at temperatures slightly higher than necessary to kill non-spore-forming pathogenic bacteria.[113] Therefore, the presence of ALP activity indicates either inadequate pasteurization or recontamination of already pasteurized milk with raw milk, thus causing a health hazard.

Different methods have been proposed to quantify ALP activity. Colorimetric methods have been adopted as an official method in the dairy industry. They use phenyl phosphates, such as *p*-nitrophenyl phosphate, or phenolphthalein monophosphate as a substrate, and ALP activity is then measured spectrophotometrically. The disadvantages of these approaches are complicated sample preparations such as protein precipitation, extraction, dialysis, and isolation of enzymatically formed products prior to analysis.

A fluorometric method has been developed lately.[114] The method was adopted by the Association of Official Analytical Chemists (AOAC) in 1990.[114] A fluorometric substrate, Fluorophos®, loses a phosphate radical and converts to a highly fluorescent form of Fluoroyellow®, when acted upon by ALP. The rate of fluorophore formation is monitored and the enzyme activity is expressed in $mU\,L^{-1}$. The method is rapid (3 min), sensitive, precise, and easy to carry out. It avoids interference and nonspecificity, and does not require long and complicated sample preparation. It has been applied to dairy products other than whole milk, e.g. skim milk, chocolate milk, and cream.[114]

Considerable research has been conducted using this new Fluorophos® method. It can detect ALP activity below the US legal limits in dairy products.[115] ALP activity in pasteurized milk has been determined at various combinations of temperatures and time. Eckner[116] reported a good correlation between ALP inactivation and pathogen inactivation. The Fluorophos® method was favorably compared to the Australian standard method.[117] Most of the papers published so far have reported that the Fluorophos® method is fast, sensitive, and reliable.

ALP has also been widely used in ELISA, blotting techniques and in situ hybridization. The advantage of ATP is the good stability of the enzyme as well as the conjugates, high sensitivity, and the availability of fluorogenic and chromogenic substrates.[118] The ELISA has been widely employed in many areas of food analysis. Fluorescent avidin and streptavidin are extensively used in the ELISA. Examples of applications can be found in many books and reports.[119]

In a similar manner to ALP, acid phosphatase was also used to indicate the heating of cooked, boneless, nonbreaded broiler breast and thigh meat.[120]

10.2 Proteolytic Enzymes

In the dairy industry, proteolytic activity is a useful predictor of shelf life. During storage of raw milk, psychrotrophic bacteria become the dominant flora and produce heat-resistant extracellular enzymes. These enzymes can degrade fat and proteins, thereby adversely affecting the quality and shelf life of milk and dairy products.[121] In cheeses, proteolysis during ripening greatly affects texture and flavor.[122,123] It is extremely important to find the extent of proteolysis as results of cheese maturity, acceleration of cheese ripening, and/or selection of starter culture.[122]

The degree of proteolysis in cheeses has been determined by assessing the amount of nitrogen-soluble compounds using various precipitants to remove insoluble proteins. Various methods used to extract soluble nitrogenous compounds were reviewed in recent papers.[123] However, these methods are tedious and time-consuming. More rapid methods have been suggested to determine proteolysis by analyzing free amino acids. For the quantification of amino acids, chromatographic techniques similar to those as already discussed have been used. Fluorescamine was used for a postcolumn derivatization while OPA has been used for both pre- and postcolumn derivatizations.[124]

In milk and dairy products, several methods are used to measure proteolysis.[125,126] They are chemical (Kjeldahl) and spectrophotometric, such as UV, Hull, Lowry, dye binding, ninhydrin, trinitrobenzenesulfonic acid (TNBS), and fluorometric, such as OPA and fluorescamine. *o*-Phthalaldehyde reacts with the α-amino group of free amino acids and small peptides under alkaline conditions in the presence of a reducing agent and forms a fluorescent compound. The OPA reaction is more rapid and convenient than the TNBS method.[127] The postcolumn OPA derivatization has also been applied to quantify free amino acids in cheeses.[128]

A fluorescamine method used to determine proteolytic activity is based on the use of the fluorescent dye, which reacts specifically with proteins and amino acids of microbial cells. The fluorescamine method was compared with other methods, i.e. absorbance at 280 nm, TNBS assay, and the Lowry method.[129] Each method was applied to the trichloroacetic acid soluble fraction of milk protein, which had been preliminarily digested with trypsin. The fluorescamine method was the most reliable and sensitive method among those four methods.

Proteolytic activity has also been determined using an FIA method. The amount of fluorescein released from enzymatic cleavage of the labeled protein used as a substrate is found to be directly proportional to the proteolytic activity. The method is rapid, sensitive, and suitable for routine determination.[130]

11 MISCELLANEOUS

Histamine and histidine in tuna were determined by HPLC with a fluorescence detector after derivatization with OPA.[131] Biogenic amines in food were determined by precolumn derivatization with 2-naphthyloxycarbonyl chloride.[132] Eight heterocyclic carcinogenic amines in satay, a local delicacy in Singapore, were determined using HPLC by measuring their natural fluorescence.[133] Nitrated polycyclic aromatic hydrocarbons in foods were determined by on-line reduction and HPLC.[134]

Ethanol contents in low ethanol concentration drink and vinegar were determined by measuring intrinsic fluorescence of alcohol dehydrogenase.[135] Fluorometric HPLC determination of ethylenediaminetetraacetate in foods was proposed by reacting with Zr and Alizarin Red S.[136] Fluorometric determination of connective tissue in beef was discussed with regard to the relative direction of fluorescence measurement.[137] The effect of pH on fluorescence formation related to fish deterioration was discussed.[138]

The application of fluorescence polarization spectroscopy in food research has been reviewed by Marangoni.[139] It was utilized in the determination, using microviscosity, of isotropic fluids such as edible fats and oils, and oil-in-water emulsion cores. The same method was also used for structural order determination of liposomes and biological membranes.

When Udenfriend's book[1] was consulted again, it was found that the applications mentioned in this article had not fully covered the analysis of some carbohydrates (e.g. hexoses, pentoses, sialic acid), aldehydes, ketones, steroids, and inorganic constituents (e.g. calcium, magnesium, zinc, aluminum, selenium, fluoride, ammonia, hydrogen peroxide) of foods. Thus, it is worth noting that many components in foods, other than those discussed in this article, can also be analyzed by fluorometry.

ACKNOWLEDGMENT

This work was financially supported by a Research Grant from the Natural Sciences and Engineering Council of Canada.

ABBREVIATIONS AND ACRONYMS

ADAM	9-Anthryldiazomethane
ALP	Alkaline Phosphatase
ANS	Anilinonaphthalenesulfonate
AOAC	Association of Official Analytical Chemists
API	Atmospheric Pressure Ionization
ATP	Adenosine Triphosphate
Br-Mme	4-Bromomethyl-7-methoxycoumarin
CFU	Colony Forming Units
CPA	*cis*-Parinaric Acid
DEFT	Direct Epifluorescent Filter Technique
DPH	1,6-Diphenyl-1,3,5-hexatriene
DPPP	Diphenyl-pyrenylphosphine
DSP	Diarrhetic Shellfish Poisoning
DTX	Dinophysistoxins
ELISA	Enzyme-linked Immunosorbent Assay
FAD	Flavin Adenine Dinucleotide
FIA	Flow Injection Analysis
FITC	Fluorescein Isothiocyanate
FMN	Flavin Mononucleotide
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
MDA	Malondialdehyde
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OPA	*o*-Phthaldialdehyde
OPD	*o*-Phenylenediamine
PhIP	2-Amino-1-methyl-6-phenyl-imidazo[4,5-*b*]pyridine
PSP	Paralytic Shellfish Poisoning
PTH	Phenylthiohydantoin
PTX	Pectenotoxins
RF	Riboflavin
RP	Reversed-phase
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RSDR	Relative Standard Deviations of Reproducibility
SFC	Supercritical Fluid Chromatography
SPFIA	Solid-phase Fluorescence Immunoassay
TBA	Thiobarbituric Acid
TLC	Thin-layer Chromatography
TNBS	Trinitrobenzenesulfonic Acid
TRF	Total Riboflavin
UV	Ultraviolet
YTX	Yessotoxin

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food • Vitamins Analysis in Food

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Surfaces **(Volume 10)**
Photoluminescence in Analysis of Surfaces and Interfaces

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Surface Measurements using Absorption/Luminescence • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

REFERENCES

1. S. Udenfriend, *Fluorescence Assay in Biology and Medicine*, Academic Press, New York, Vol. II, 1969.
2. L. Munk, *Fluorescence Analysis in Foods*, Longman Scientific & Technical, Harlow, UK, 1989.
3. L. Brand, M.L. Johnson (eds.), 'Fluorescence Spectroscopy', *Methods Enzymol.*, **278** (1997).
4. G.M. Strasburg, R.D. Ludescher, 'Theory and Applications of Fluorescence Spectroscopy in Food Research', *Trends Food Sci. Technol.*, **6**, 69–75 (1995).
5. O.E. Mills, 'Effect of Temperature on Tryptophan Fluorescence of β-Lactoglobulin B', *Biochim. Biophys. Acta*, **434**, 324–332 (1976).
6. P. Wahl, S.N. Timasheff, J.C. Auchet, 'Polarized Fluorescence Decay Curves for β-Lactoglobulin A in Various States of Association', *Biochemistry*, **8**, 2945–2949 (1969).
7. S. Iametti, L. Scaglioni, S. Maxxini, G. Vecchio, F. Bonomi, 'Structural Features and Reversible Association of Different Quaternary Structures of β-Lactoglobulin', *J. Agric. Food Chem.*, **46**, 2159–2166 (1998).
8. R.D. Fugate, P.-S. Song, 'Spectroscopic Characterization of β-Lactoglobulin–Retinol Complex', *Biochim. Biophys. Acta*, **625**, 28–42 (1980).
9. E.M. Brown, R.J. Carrol, P.E. Pfeffer, J. Sampugna, 'Complex Formation in Sonicated Mixtures of β-Lactoglobulin and Phosphatidylcholine', *Lipids*, **18**, 111–118 (1983).
10. M.A.M. Edwards, S.E. Silva, 'Exposure of Tryptophanyl Residues in α-Lactalbumin and Lysozyme. Quantitative Determination by Fluorescence Quenching Studies', *Radiat. Environ. Biophys.*, **25**, 113–122 (1986).
11. V. Moreaux, I. Birlouez-Aragon, 'Degradation of Tryptophan in Heated β-Lactoglobulin–Lactose Mixtures is Associated with Intense Maillard Reaction', *J. Agric. Food Chem.*, **45**, 1905–1910 (1997).
12. H.C. Yang, M.M. Reedy, C. Burke, G.M. Strasburg, 'Calmodulin Interaction with the Skeletal Muscle Sarcoplasmic Reticulum Calcium Channel Protein', *Biochemistry*, **33**, 518–525 (1994).
13. Z. Grabarek, P.C. Leavis, J. Gergely, 'Calcium Binding to the Low Affinity Sites in Troponin C Induces Conformational Changes in the High Affinity Domain', *J. Biol. Chem.*, **261**, 608–613 (1986).
14. S. Nakai, E. Li-Chan, G.E. Arteaga, 'Measurement of Surface Hydrophobicity', in *Methods of Testing Protein Functionality*, ed. G.M. Hall, Blackie Academic & Professional, London, 226–259, 1996.
15. S. Nakai, 'Measurement of Protein Hydrophobicity', in *Current Protocols in Food Analytical Chemistry*, R. Wrolstad (Editor-in-Chief), John Wiley & Sons, New York, 2000.
16. M. Cardamone, N.K. Puri, 'Spectrofluorimetric Assessment of the Surface Hydrophobicity of Proteins', *Biochem. J.*, **282**, 589–593 (1992).
17. C.A. Gatti, P.H. Risso, S.M. Zerpa, 'Study of the Inhibitory Effect of Hydrophobic Fluorescent Markers on the Enzyme Coagulation of Bovine Casein Micelles: Action of TNS', *Food Hydrocol.*, **12**, 393–400 (1998).
18. L.A. Sklar, B.S. Hudson, R.D. Simoni, 'Conjugated Polyene Fatty Acids as Fluorescent Probes: Binding to Bovine Serum Albumin', *Biochemistry*, **15**, 5100–5108 (1977).
19. A. Kato, S. Nakai, 'Hydrophobicity Determination by a Fluorescence Probe Method and its Correlation with Surface Properties of Proteins', *Biochim. Biophys. Acta*, **624**, 13–20 (1980).
20. S. Hayakawa, S. Nakai, 'Relationships of Hydrophobicity and Net Charge to the Solubility of Milk and Soy Proteins', *J. Food Sci.*, **50**, 486–491 (1985).
21. C. Boyer, S. Joandel, A. Ouali, J. Celioli, 'Determination of Surface Hydrophobicity of Fast and Slow Myosins from Rabbit Skeletal Muscles: Implication in Heat-induced Gelation', *J. Sci. Food Agric.*, **72**, 367–375 (1996).
22. F. Greene, 'Interactions of Anionic and Cationic Fluorescent Probes with Proteins: The Effect of Charge', *J. Protein Chem.*, **3**, 167–180 (1984).
23. C.A. Haskard, E.C.Y. Li-Chan, 'Hydrophobicity of Bovine Serum Albumin and Ovalbumin Determined Using Uncharged (PRODAN) and Anionic (ANS^-) Fluorescent Probes', *J. Agric. Food Chem.*, **46**, 2671–2677 (1998).

24. T. Tsutsui, E. Li-Chan, S. Nakai, 'A Simple Fluorometric Method for Fat-binding Capacity as an Index of Hydrophobicity of Proteins', *J. Food Sci.*, **51**, 1268–1272 (1986).
25. S. Iametti, F. Bonomi, 'Monitoring the Surface Hydrophobicity of Milk Proteins: A Real-time Study on Heat-induced Modifications', in *Int. Dairy Fed. Spec. Issue*, **9303** ('Protein and Fat Globule Modifications by Heat Treatment, Homogenization and Other Technological Means for High Quality Dairy Products'), 111–116 (1993).
26. U. Kalapathy, N.S. Hettiarachchy, K.C. Rhee, 'Effect of Drying Methods on Molecular Properties and Functionalities of Disulfide Bond-cleaved Soy Proteins', *J. Am. Oil Chem. Soc.*, **74**, 195–199 (1997).
27. L. Eynard, S. Iametti, R. Perla, B. Francesco, 'Surface Hydrophobicity Changes and Heat-induced Modifications of α-Lactalbumin', *J. Agric. Food Chem.*, **40**, 1731–1736 (1992).
28. S. Iametti, S. Cairoli, B. De Gregori, F. Bonomi, 'Modifications of High-order Structures Upon Heating of β-Lactoglobulin: Dependence on the Protein Concentration', *J. Agric. Food Chem.*, **43**, 53–58 (1995).
29. N. Guerrieri, E. Alberti, V. Lavelli, P. Cerletti, 'Use of Spectroscopic and Fluorescence Techniques to Assess Heat-induced Modifications of Gluten', *Cereal Chem.*, **73**, 368–374 (1996).
30. I. Nir, Y. Feldman, A. Aserin, N. Garti, 'Surface Properties and Emulsification Behavior of Denatured Soy Proteins', *J. Food Sci.*, **59**, 606–610 (1994).
31. R.E. Aluko, R.Y. Yada, 'Some Physicochemical and Functional Properties of Cowpea (*Vigna unguiculata*) Isoelectric Protein Isolate as a Function of pH and Salt Concentration', *Int. J. Food Sci. Nutr.*, **48**, 31–39 (1997).
32. B. Lieske, 'Influence of Preliminary Treatments on Structural Properties of Casein Micelles Affecting the Rennetability', *Lait*, **77**, 201–209 (1997).
33. M.M. Mossoba, *Spectral Methods in Food Analysis*, Marcel Dekker, New York, 1999.
34. W.F. Busby, G.N. Wogan, *Aflatoxin, Chemical Carcinogens*, 2nd edition, ed. C.E. Searle, American Chemical Society, Washington, DC, 945–1136, 1985.
35. R.J. Cole, R.H. Cox, *Handbook of Toxic Fungal Metabolites*, Academic Press, New York, 937, 1981.
36. W.Th. Kok, Th.C.H. Van Neer, W.A. Traag, L.G.M.Th. Tuinstra, 'Determination of Aflatoxins in Cattle Feeds by Liquid Chromatography and Post-column Derivatization with Electrochemically Generated Bromine', *J. Chromatogr.*, **367**, 231–236 (1986).
37. C. Finoli, A. Vecchio, 'Aflatoxin M_1 in Goat Dairy Products', *Microbiol. Aliments Nutr.*, **15**, 47–52 (1997).
38. S. Corneli, C.M. Maragos, 'Capillary Electrophoresis with Laser-induced Fluorescence Methods for the Mycotoxin Ochratoxin A', *J. Agric. Food Chem.*, **46**, 3162–3165 (1998).
39. C.M. Maragos, 'Detection of Mycotoxin Fumonisin B_1 by a Combination of Immunofluorescence and Capillary Electrophoresis', *Food Agric. Immunol.*, **9**, 147–157 (1997).
40. E. Rosenberg, R. Kriska, R. Wissiak, V. Kmetov, R. Josephs, E. Razzazi, M. Grasserbauer, 'High-performance Liquid Chromatographic–Pressure Chemical Ionization Mass Spectrometry as a New Tool for the Determination of the Mycotoxin Zearalenone in Food and Feed', *J. Chromatogr. A*, **819**, 277–288 (1998).
41. J.L. Shen, G. Ganzlin, B. Luckas, *Proceedings of Symposium on Marine Biotoxins Paris, January, 1991*, ed. J.M. Fremy, Editions CNEVA, Maisons-Alfort, 101, 1991.
42. J.L. Shen, C. Hummert, B. Luckas, 'Sensitive HPLC–Fluorometric and HPLC–MS Determination of Diarrhetic Shellfish Poisoning (DSP)-toxins as 4-bromomethyl-7-methoxycoumarin (BrMmc) Esters', *Fresenius' J. Anal. Chem.*, **357**, 101–104 (1997).
43. J.D. Dziezak, 'Rapid Methods for Microbiological Analysis of Foods', *Food Technol.*, **41**(7), 56–73 (1987).
44. D.Y.C. Fung, 'Rapid Methods and Automation for Food Microbiology', in *Instrumental Methods for Quality Assurance in Foods*, eds. D.Y.C. Fung, R.F. Mattews, Marcel Dekker, New York, 1–38, 1991.
45. B.M. Thomason, 'Current Status of Immunofluorescent Methodology for *Salmonellae*', *J. Food Prot.*, **44**, 381–384 (1981).
46. M.A. Khan, A. Seaman, M. Woodbine, 'Immunofluorescent Identification of *Listeria monocytogenes*', *Zentralbl. Baketeriol. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. Reihe A*, **239**, 62–69 (1977).
47. J. Czajka, C.A. Batt, 'A Solid Phase Fluorescent Capillary Immunoassay for the Detection of *Escherichia coli* O157:H7 in Ground Beef and Apple Cider', *J. Appl. Bacteriol.*, **81**, 601–607 (1996).
48. T. Henle, U. Schwarzenbolz, A.W. Walter, H. Klostermeyer, 'Protein-bound Maillard Compounds in Foods: Analytical and Technological Aspects', *Royal Soc. Chem. Spec. Publ.*, **223**, 178–183 (1998).
49. F.J. Morales, S. Jimenez-Perez, 'Monitoring of Heat-induced Proteolysis in Milk and Milk-resembling Systems', *J. Agric. Food Chem.*, **46**, 4391–4397 (1998).
50. M. Murkovic, M. Friedrich, W. Pflannhauser, 'Heterocyclic Aromatic Amines in Fried Poultry Meat', *Z. Lebenm.-Unters. Forsch. A*, **205**, 347–350 (1997).
51. S.B. Engelsen, 'Explorative Spectrometric Evaluation of Frying Oil Deterioration', *J. Am. Oil Chem. Soc.*, **74**, 1495–1508 (1997).
52. A. Orta-Ramirez, D.M. Smith, J.E. Merrill, 'Denaturation of Algal Phycobiliproteins Can be Used as a Thermal Process Indicator', in *Proceedings 4th International Maritime Biotechnology Conference*, eds. Y. Le

Gal, H.O. Halvorson, Plenum, New York, 245–249, 1998.
53. J.I. Gray, F.J. Monahan, 'Measurement of Lipid Oxidation in Meat and Meat Products', *Trends Food Sci. Technol.*, **3**, 315–319 (1992).
54. G.C. Yang, 'Detection of Lipid Hydroperoxides by High-performance Liquid Chromatography Coupled with Post-column Reaction', *Trends Food Sci. Technol.*, **3**, 15–18 (1992).
55. K. Hasegawa, Y. Endo, E. Fujimoto, 'Oxidative Deterioration in Dried Fish Model Systems Assessed by Solid Sample Fluorescence Spectrophotometry', *J. Food Sci.*, **57**, 1123–1126 (1992).
56. T. Miyazawa, M. Kashima, K. Fujimoto, 'Fluorometric Peroxygenase Assay for Lipid Hydroperoxides in Meats and Fish', *J. Food Sci.*, **58**, 66–70 (1993).
57. K. Akasaka, T. Takamura, H. Ohrui, H. Meguro, K. Hashimoto, 'Highly Sensitive Flow Injection Analysis of Lipid Hydroperoxides in Foodstuffs', *Biosci. Biotechnol. Biochem.*, **60**, 1772–1775 (1996).
58. S. Raharjo, J.N. Sofos, 'Methodology for Measuring Malonaldehyde as a Product of Lipid Peroxidation in Muscle Tissues: A Review', *Meat Sci.*, **35**, 145–169 (1993).
59. P. Bergamo, E. Fedele, M. Balestrieri, P. Abrescia, L. Ferrara, 'Measurement of Malondialdehyde Levels in Food by High-performance Liquid Chromatography with Fluorometric Detection', *J. Agric. Food Chem.*, **46**, 2171–2176 (1998).
60. D.V. Hoyland, A.J. Taylor, 'A Review of the Methodology of the 2-Thiobarbituric Acid Test', *Food Chem.*, **40**, 271–291 (1991).
61. S. Raharjo, J.N. Sofos, G.R. Schmidt, 'Improved Speed, Specificity and Limit of Determination of an Aqueous Acid Extraction Thiobarbituric Acid–C_{18} Method for Measuring Lipid Peroxidation in Beef', *J. Agric. Food Chem.*, **40**, 2182–2185 (1992).
62. C. Jo, D.U. Ahn, 'Fluorometric Analysis of 2-Thiobarbituric Acid Reactive Substances in Turkey', *Poultry Sci.*, **77**, 475–480 (1998).
63. S.R. Tannenbaum, V.R. Young, 'Vitamins and Minerals', in *Food Chemistry*, ed. O.R. Fennema, Marcel Dekker, New York, 477–543, 1985.
64. J. Giese, 'Vitamin and Mineral Fortification of Foods', *Food Technol.*, **49**(5), 109–112, 114, 116, 118, 120–122 (1995).
65. J.G. Elliott, 'Application of Antioxidant Vitamins in Foods and Beverages', *Food Technol.*, **53**(2), 46–48 (1999).
66. J.F. Gregory, III, 'Methods of Vitamin Assay for Nutritional Evaluation of Food Processing', *Food Technol.*, **37**(1), 74–80 (1983).
67. R.O. Rahn, 'Fluorescence and Phosphorescence', in *Food Analysis, Volume 2, Physicochemical Techniques*, eds. D.W. Gruenwedel, J.R. Whitaker, Marcel Dekker, New York, 293–341, 1984.
68. J.H. Richardson, 'Sensitive Assay of Biochemicals by Laser-induced Molecular Fluorescence', *Anal. Biochem.*, **83**, 754–762 (1977).
69. L.F. Russell, J.T. Vanderslice, 'Comments on the Standard Fluorometric Determination of Riboflavin in Foods and Biological Tissues', *Food Chem.*, **43**, 79–82 (1992).
70. A. Sims, D. Shoemaker, 'Simultaneous Liquid Chromatographic Determination of Thiamine and Riboflavin in Selected Foods', *J. AOAC Int.*, **76**, 1156–1160 (1993).
71. S.M. Fernando, P.A. Murphy, 'HPLC Determination of Thiamine and Riboflavin in Soybeans and Tofu', *J. Agric. Food Chem.*, **38**, 163–167 (1990).
72. L.F. Russell, J.T. Vanderslice, 'Non-degradative Extraction and Simultaneous Quantitation of Riboflavin, Flavin Mononucleotide, and Flavin Adenine Dinucleotide in Foods by HPLC', *Food Chem.*, **43**, 151–162 (1992).
73. J.V. Schoonhoven, J. Schrijver, H.V.D. Berg, G.R.M.M. Haenen, 'Reliable and Sensitive High-performance Liquid Chromatographic Method with Fluorometric Detection for the Analysis of Vitamins B_6 in Foods and Feeds', *J. Agric. Food Chem.*, **42**, 1475–1480 (1994).
74. G.W. Chase, W.O. Landen, Jr, A.G.M. Soliman, R.R. Eitenmiller, 'Method Modification for Liquid Chromatographic Determination of Thiamine, Riboflavin and Pyridoxine in Medical Foods', *J. AOAC Int.*, **76**, 1276–1280 (1993).
75. G.W. Chase, W.O. Landen, Jr, R.R. Eitenmiller, A.G.M. Soliman, 'Liquid Chromatographic Determination of Thiamine, Riboflavin, and Pyridoxine in Infant Formula', *J. AOAC Int.*, **75**, 561–565 (1992).
76. F. Watanabe, K. Abe, S. Takenaka, T. Fujita, Y. Nakano, 'Method for Quantitation of Total Vitamin B_{12} in Foods Using a Highly Fluorescent Vitamin B_{12} Derivative', *J. Agric. Food Chem.*, **45**, 4661–4663 (1997).
77. K.Y. Dodson, E.R. Young, A.G.M. Soliman, 'Determination of Total Vitamin C in Various Food Matrixes by Liquid Chromatography and Fluorescence Detection', *J. AOAC Int.*, **75**, 887–891 (1992).
78. M. Ali, E.T. Philippo, 'Simultaneous Determination of Ascorbic, Dehydroascorbic, Isoascorbic and Dehydroisoascorbic Acids in Meat-based Food Products by Liquid Chromatography with Postcolumn Fluorescence Detection: A Method Extension', *J. AOAC Int.*, **79**, 803–808 (1996).
79. E. Jakob, I. Elmadfa, 'Application of a Simplified HPLC Assay for the Determination of Phylloquinone (Vitamin K_1) in Animal and Plant Food Items', *Food Chem.*, **56**, 87–91 (1996).
80. M.D. Lucock, M. Green, M. Priestnall, I. Daskalakis, M.I. Lenene, R. Hareley, 'Optimization of Chromatographic Conditions for the Determination of Folates in Foods and Biological Tissues for Nutritional and Clinical Work', *Food Chem.*, **53**, 329–338 (1995).
81. S.K. Jensen, 'Retinol Determination in Milk by HPLC and Fluorescence Detection', *J. Dairy Res.*, **61**, 233–240 (1994).

82. C.Y.W. Ang, G.K. Searcy, R.R. Eitenmiller, 'Tocopherols in Chicken Breast and Leg Muscles Determined by Reverse Phase Liquid Chromatography', *J. Food Sci.*, **55**, 1536–1539 (1990).
83. I.B. Razagui, P.J. Barlow, K.D.A. Taylor, 'A Spectrofluorimetric Procedure for the Determination of α-Tocopherol in Nutritional Supplement Products', *Food Chem.*, **44**, 221–226 (1992).
84. P.C.H. Hollman, J.H. Slangen, P.J. Wagstaffe, U. Faure, D.A.T. Southgate, P.M. Finglas, 'Intercomparison of Methods for the Determination of Vitamins in Food. 1. Fat-soluble Vitamins', *Analyst*, **118**, 475–480 (1993).
85. P.C.H. Hollman, J.H. Slangen, P.J. Wagstaffe, U. Faure, D.A.T. Southgate, P.M. Finglas, 'Intercomparison of Methods for the Determination of Vitamins in Food 2. Water-soluble Vitamins', *Analyst*, **118**, 481–488 (1993).
86. A.K. Hewavitharana, A.S. Van Brakel, M. Harnett, 'Simultaneous Liquid Chromatographic Determination of Vitamins A, E and β-Carotene in Common Dairy Foods', *Int. Dairy J.*, **6**, 613–624 (1996).
87. Q. Liu, K.K. Scheller, D.M. Schaefer, 'Technical Note, A Simplified Procedure for Vitamin E Determination in Beef Muscle', *J. Animal Sci.*, **74**, 2406–2410 (1996).
88. P. Russell, 'Antibiotic Determination', *Milk Industry Int.*, **4**, 17–20 (1997).
89. A. Sternesjo, G. Johnson, 'A Novel Rapid Enzyme Immunoassay (Fluorophos® Betascreen) for Detection of β-Lactam Residues in Ex-farm Raw Milk', *J. Food Prot.*, **61**, 808–811 (1998).
90. L.D. Payne, V.R. Mayo, L.A. Morneweck, M.B. Hicks, T.A. Wehner, 'HPLC–Fluorescence Method for the Determination of Eprinomectin Marker Residue in Edible Bovine Tissue', *J. Agric. Food Chem.*, **45**, 3501–3506 (1997).
91. R.J. Maxwell, E. Cohen, D.J. Donoghue, 'Determination of Sarafloxacin Residues in Fortified and Incurred Eggs Using On-line Microdialysis and HPLC/Programmable Fluorescence Detection', *J. Agric. Food Chem.*, **47**, 1563–1567 (1999).
92. C.Y.W. Ang, W. Luo, 'Rapid Determination of Ampicillin in Bovine Milk by Liquid Chromatography with Fluorescence Detection', *J. AOAC Int.*, **80**, 25–30 (1997).
93. W. Luo, E.B. Hansen, Jr, C.Y.W. Ang, J. Deck, 'Simultaneous Determination of Amoxicillin and Ampicillin in Bovine Milk by HPLC with Fluorescence Detection', *J. Agric. Food Chem.*, **45**, 1264–1268 (1997).
94. J.E. Roybal, A.P. Pfenning, S.B. Turnipseed, C.C. Walker, J.A. Hurlbut, 'Determination of Four Fluoroquinolones in Milk by Liquid Chromatography', *J. AOAC Int.*, **80**, 982–987 (1997).
95. M.B. Medina, 'Particle Concentration Fluorescence Assay for Rapid Detection of Trace Levels of Antibiotics', *ACS Symp. Ser.*, 636 (Veterinary Drug Residues), American Chemical Society, Washington, DC, 132–143, 1996.
96. Gilson Ref 800351A, *Gilson Solutions for the Control of Food Quality*, Gilson, Inc., Middleton, WI, 13, 1998.
97. C.E. Tsai, F. Kondo, 'Liquid Chromatographic Determination of Fluorescent Derivatives of Six Sulfonamides in Bovine Serum and Milk', *J. AOAC Int.*, **78**, 674–678 (1995).
98. J. Laloux, J.M. Romnee, C. Marin, D. Vanwijnsberghe, 'Determination of Sulfonamides in Milk Using Fluorescamine Derivatization and HPLC Separation', *Milchwissenschaft*, **51**, 517–520 (1996).
99. G.C. Gerhardt, C.D.C. Salisbury, J.D. MacNeil, 'Analysis of Streptomycin, and Dihydrostreptomycin in Milk by Liquid Chromatography', *J. AOAC Int.*, **77**, 765–767 (1994).
100. A.P. Pfenning, R.K. Munns, S.B. Turnipseed, J.E. Roybal, D.C. Holland, A.R. Long, S.M. Plakas, 'Determination and Confirmation of Identities of Flumequine and Nalidixic, Oxolinic, and Piromidic Acids in Salmon and Shrimp', *J. AOAC Int.*, **79**, 1227–1235 (1996).
101. C.Y.W. Ang, W. Luo, E.B. Hansen, Jr, J.P. Freeman, H.C. Thompson, Jr, 'Determination of Amoxicillin in Catfish and Salmon Tissues by Liquid Chromatography with Precolumn Formaldehyde Derivatization', *J. AOAC Int.*, **79**, 389–396 (1996).
102. J.R. Meinertz, V.K. Dawson, W.H. Gingerich, B. Cheng, 'Liquid Chromatographic Determination of Sarafloxacin Residues in Channel Catfish Muscle Tissue', *J. AOAC Int.*, **77**, 871–875 (1994).
103. T.A. Gehring, L.G. Rushing, H.C. Thompson, Jr, 'Determination of Sulfonamides in Edible Salmon Tissue by Liquid Chromatography with Postcolumn Derivatization and Fluorescence Detection', *J. AOAC Int.*, **80**, 751–755 (1997).
104. E. Diçkaya, M. Çagin, A. Telefoncu, 'Enzymatic Method for the Spectrophotometric Determination of Aspartame', *Food Chem.*, **50**, 95–97 (1994).
105. J. Prodolliet, M. Bruelhart, 'Determination of Aspartame and its Major Decomposition Products in Foods', *J. AOAC Int.*, **76**, 275–282 (1993).
106. K. Wróbel, K. Wróbel, 'Determination of Aspartame and Phenylalanine in Diet Soft Drinks by High-performance Liquid Chromatography with Direct Spectrofluorometric Detection', *J. Chromatogr.*, **A773**, 163–168 (1997).
107. D.P. Venema, P.C.H. Hollman, K.P.L.T.M. Janssen, M.B. Katan, 'Determination of Acetylsalicylic Acid and Salicylic Acid in Foods, Using HPLC with Fluorescence Detection', *J. Agric. Food Chem.*, **44**, 1762–1767 (1996).
108. J. Haginaka, J. Wakai, 'Fluorimetric Determination of Amino Acids by High-performance Liquid Chromatography Using a Hollow-fiber Membrane Reactor', *J. Chromatogr.*, **396**, 297–305 (1987).
109. M. Ward, 'Amino Acid Analysis Using Precolumn Derivatization with 6-aminoquinolyl-*N*-hydroxysuccinimyl Carbamate', in *The Protein Protocols Handbook*, ed. J.M. Walker, Humana Press, Totowa, NJ, 461–465, 1996.

110. G.B. Irvine, 'Amino Acid Analysis Using Precolumn Derivatization with Phenylisothiocyanate', in *The Protein Protocols Handbook*, ed. J.M. Walker, Humana Press, Totowa, NJ, 467–472, 1996.
111. T. Henle, 'The Amino Acid Analyzer – Still Powerful in Food Analysis and Research', Curr. Status Future Trends Anal. Food Chem., in *Proceedings 8th European Conference on Food Chemistry*, eds. G. Sontag, W. Pfannhauser, Austrian Chemical Society, Vienna, Austria, 70–73, Vol. 1, 1995.
112. F.V. Koskikowski, 'Enzyme Behavior and Utilization in Dairy Technology', *J. Dairy Sci.*, **71**, 557–573 (1988).
113. H.D. Kay, W.R. Graham, 'The Phosphatase Test for Pasteurized Milk', *J. Dairy Res.*, **6**, 191–203 (1935).
114. R.M. Rocco, 'Fluorometric Determination of Alkaline Phosphatase in Fluid Dairy Products: Collaborative Study', *J. AOAC Int.*, **73**, 842–849 (1990).
115. C.J. Painter, B.L. Bradley, Jr, 'Residual Alkaline Phosphatase Activity in Milks Subjected to Various Time–Temperature Treatments', *J. Food Prot.*, **60**, 525–530 (1997).
116. K.F. Eckner, 'Fluorometric Analysis of Alkaline Phosphatase Inactivation Correlated to *Salmonella* and *Listeria* Inactivation', *J. Food Prot.*, **55**, 960–963 (1992).
117. R.G. Black, M. Kuzyk, J. Duggan, 'Evaluation of a Fluorometric Assay for Alkaline Phosphatase in Fluid Dairy Products', *Aust. J. Dairy Technol.*, **47**, 64–67 (1992).
118. G.C. Howard (ed.), *Methods in Nonradioactive Detection*, Appleton & Lange, Norwalk, CT, 1993.
119. M.R.A. Morgan, C.J. Smith, P.A. Williams (eds.), *Food Safety and Quality Assurance. Application of Immunoassay Systems*, Elsevier Science Publishers Barking, UK, 1992.
120. C.E. Davis, 'Fluorometric Determination of Acid Phosphatase in Cooked Boneless, Nonbreaded Broiler Breast and Thigh Meat', *J. AOAC Int.*, **81**, 887–906 (1998).
121. J.M. Cox, 'The Significance of Psychrotrophic Pseudomonads in Dairy Products', *Aust. J. Dairy Technol.*, **48**, 108–113 (1993).
122. P.F. Fox, 'Proteolysis during Cheese Manufacturing and Ripening', *J. Dairy Sci.*, **72**, 1379–1400 (1989).
123. P.F. Fox, L.H. McSweeney, 'Proteolysis in Cheese during Ripening', *Food Rev. Int.*, **12**, 457–509 (1996).
124. R.Y. Yada, R.L. Jackman, J.L. Smith, A.G. Marangoni, 'Analysis: Quantitation and Physical Characterization', in *Food Proteins. Properties and Characterization*, eds. S. Nakai, H.W. Modler, VCH, Weinheim, Germany, 348–350, 1996.
125. J.M. Wallace, P.F. Fox, 'Rapid Spectrophotometric and Fluorimetric Methods for Monitoring Nitrogenous (Proteinaceous) Compounds in Cheese and Cheese Fractions: A Review', *Food Chem.*, **62**, 217–224 (1998).
126. Y. Ardo, H. Melsel, 'Methods for Direct Measurement of Peptide Bond Cleavage in Cheese', *Bull. Int. Dairy Fed.*, **261**, 10–13 (1991).
127. L. Lemieux, R. Puchades, R.E. Simard, 'Free Amino Acids in Cheddar Cheese: Comparison of Quantitation Methods', *J. Food Sci.*, **55**, 1552–1554 (1990).
128. D. Gonzalez de Llano, M. Ramos, M.C. Polo, 'Gel Filtration and High-performance Liquid Chromatographic Analysis of Phosphotungstic Acid Soluble Peptides from Blue Cheeses', *Chromatographia*, **23**, 764–766 (1987).
129. K.K.H. Kwan, S. Nakai, B.J. Skura, 'Comparison of Four Methods for Determining Protease Activity in Milk', *J. Food Sci.*, **48**, 1418–1421 (1983).
130. D. Narinesingh, A. Pope, T.T. Ngo, 'Rapid, Sensitive Fluorimetric Assay for Proteases Using Fluorescein-labelled Albumin Coupled to Fractogel Activated with 2-Fluoro-1-methylpyridinium Salt in Flow-injection Analysis', *Anal. Chim. Acta*, **230**, 131–136 (1990).
131. V. Fratini, C. Lionetti, 'Histamine and Histidine Determination in Tuna Fish Samples Using High-performance Liquid Chromatography. Derivatization with *o*-Phthalaldehyde and Fluorescence Detection or UV Detection of Free Species', *J. Chromatogr. A*, **809**, 241–245 (1998).
132. J. Kirschbaumm, I. Bush, H. Bruchner, 'Determination of Biogenic Amines in Foods by Automated Pre-column Derivatization with 2-Naphthyloxycarbonyl Chloride (NOC-Cl)', *Chromatographia*, **45**, 263–268 (1997).
133. J. Wu, H.K. Lee, M.K. Wong, C.N. Ong, 'Determination of Carcinogenic Heterocyclic Amines in Satay by Liquid Chromatography', *Environ. Monit. Assess.*, **44**, 405–412 (1997).
134. S. Schlemitz, W. Pfannhauser, 'Analysis of Nitro-PAHs in Food Matrices by Online Reduction and High-performance Liquid Chromatography', *Food Addit. Contam.*, **13**, 969–977 (1996).
135. S. de Marcos, J. Galban, R. Albajez, J.R. Castillo, 'Enzymic Determination of Ethanol Based on the Intrinsic Fluorescence of Alcohol Dehydrogenase', *Anal. Chim. Acta*, **343**, 117–123 (1997).
136. A.M.G. Campana, F.A. Barrero, M.R. Ceba, 'Sensitive Spectrofluorimetic Method for the Determination of Ethylenediaminetetraacetic Acid and Its Salts in Foods with Zirconium and Alizarin Red S in a Micellar Medium', *Anal. Chim. Acta*, **329**, 319–325 (1996).
137. H.J. Swatland, N.T. Madsen, T. Nielsen, 'Fluorometry of Connective Tissue in Beef, Relative to Direction of Measurement', *Food Sci. Technol.*, **29**, 536–541 (1996).
138. S.P. Aubourg, 'Effect of pH on Fluorescence Formation Related to Fish Deterioration', *Z. Lebensm.-Unters.-Forsch. A*, **207**, 268–272 (1998).
139. A.G. Marangoni, 'Steady-state Fluorescence Polarization Spectroscopy as a Tool to Determine Microviscosity and Structural Order in Food Systems', *Food Res. Int.*, **25**, 67–80 (1992).

Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Ranjana Mehrotra
National Physical Laboratory, New Delhi, India

1 Introduction **1**
2 Instrumentation **2**
2.1 Dispersive Spectrometer 2
2.2 Fourier Transform Infrared Spectrometer 2
3 Sampling Techniques **3**
3.1 Transmission Methods 3
3.2 Reflectance Methods 5
3.3 Fibre Optic Probe 6
4 Quantitative Analysis **6**
4.1 Beer–Lambert Plot 6
4.2 Classical Least Square Method 6
4.3 Inverse Least Square Method 7
4.4 Partial Least Square/Principle Component Analysis 7
4.5 Calibration 7
4.6 Error Analysis 8
5 Food Applications **8**
5.1 Analysis of Water 8
5.2 Analysis of Sugars 8
5.3 Analysis of Oils and Fats 9
5.4 Analysis of Milk and Dairy Products 10
5.5 Analysis of Beverages and Juices 12
5.6 Analysis of Grains 13
5.7 Analysis of Meat 14
6 Gas Chromatography/Infrared Spectroscopy **14**
6.1 Gas Chromatography/Infrared Instrumentation 14
6.2 Applications of Gas Chromatography/Infrared Spectroscopy 14
7 Advantages and Limitations **15**
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Infrared (IR) radiation is usually defined as that electromagnetic radiation whose frequency is between 20 and 14 500 cm^{-1}. Within this region of the electromagnetic spectrum, chemical compounds absorb IR radiation provided there is a dipole moment change during a normal molecular vibration, molecular rotation, and molecular rotation–vibration or from combinations, differences and overtones of the normal molecular vibrations. The frequencies and intensities of the IR bands exhibited by a chemical compound uniquely characterize the material and its IR spectrum can be used to identify and quantify a particular substance in an unknown sample. IR spectroscopy in both the mid-infrared (MIR) and near-infrared (NIR) regions is emerging as a qualitative as well as quantitative analytical technique with great potential in a wide range of applied areas. The prime advantage of IR spectroscopy in food applications is that once the instrument has been calibrated, several constituents in a sample can be measured rapidly and simultaneously. No sample manipulation is required. It is also applicable to on-line analysis in the food processing industries.

1 INTRODUCTION

All molecules are made up of atoms linked by chemical bonds. The movement of atoms and chemical bonds can be compared with that of a system comprised of springs and balls in constant motion. Their motion can be understood as being composed of two components, the stretching and bending vibrations. The frequencies of these vibrations depend on the nature of the particular bonds, such as C–H/C–O bonds.

IR radiation is usually defined as that electromagnetic radiation whose frequency is between 20 and 14 500 cm^{-1}. Within this region of the electromagnetic spectrum, chemical compounds absorb IR radiation provided there is a dipole moment change during a normal molecular vibration, molecular rotation, and molecular rotation–vibration or from combinations, differences and overtones in the normal molecular vibrations. The frequencies and intensities of the IR bands exhibited by a chemical compound uniquely characterise that material and its IR spectrum can be used to identify and quantify the particular substance in an unknown sample.

The IR region can be divided into three segments, the far-infrared (FIR) region, MIR region and NIR regions. Each region of the spectrum plays a different role in analysis according to the different character of the transitions involved in each case.

The FIR region extends from 20 to 400 cm^{-1}. This region contains the bending vibrations of C, N, O and F with heavier atoms and additional bending

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

motions in cyclic or unsaturated systems. This region is useful for studying isomers, organometallics and organic compounds whose atoms are heavy and bonds are weak.

However, the region between 400 and 4000 cm^{-1} is the most attractive for chemical analysis. This MIR region corresponds to fundamental transitions in which one vibrational mode is excited from its lowest energy state to its first excited state. This region includes the frequencies corresponding to the fundamental vibration of virtually all of the fundamental groups of the organic molecules. These spectral lines are typically narrow and distinct, making it possible to identify and monitor a band corresponding to the specific feature. As a result, quantitative calibrations performed in the MIR are usually straightforward and robust.

The NIR region lies between the visible and the MIR region of the electromagnetic spectrum and is defined as the spectral region spanning the wavelength range 4000–14 500 cm^{-1}. This region arises from transitions in which a photon excites a normal mode of vibration from ground state to the second or higher excited vibrational state (overtones) and transitions in which one photon simultaneously excites two or more vibrational modes (combinations). This region is often associated with hydrogen atoms, as in O–H, N–H and C–H systems. The NIR region is mainly used for quantitative analysis. Owing to short wavelength and high energy and hence high penetrating power, the region is very useful for biological samples.

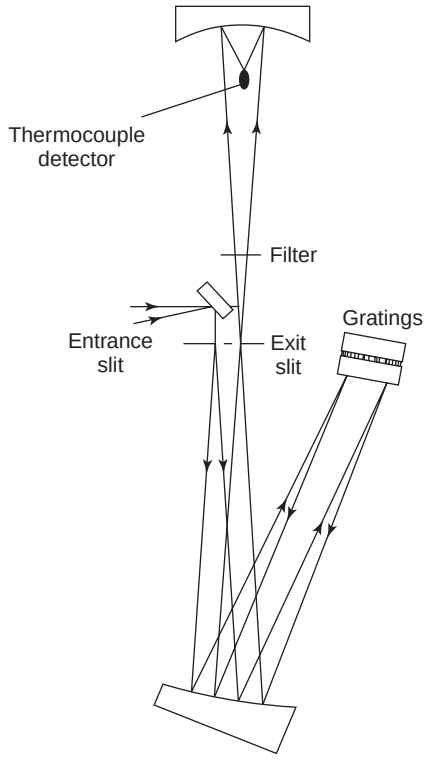

Figure 1 Optical path of a double beam IR spectrometer with a grating monochromator. (Reproduced with permission from B. Stuart, *Biological Applications of Infrared Spectroscopy*, Copyright John Wiley & Sons, Ltd.)

2 INSTRUMENTATION

There are two different types of commercial instrument available for IR measurements: the dispersive/filter-based and the Fourier transform spectrometer. Each of these instruments works in a distinctly different manner.

2.1 Dispersive Spectrometer

Since 1940, dispersive instruments have been available for measuring IR spectra in the MIR and NIR regions. A dispersive instrument consists of a source, monochromator and a detector (Figure 1). The radiation source in most IR instruments is a globar lamp, which is made of silicon carbide in the MIR region and is a tungsten halogen lamp in NIR region. A dispersive monochromator utilizes either a prism made from materials such as sodium chloride, potassium bromide, cesium fluoride, or a grating. The purpose of a prism or grating is to enable the separation of continuous blackbody radiation into its component wavelengths. The grating used in dispersive IR spectrometers consists of parallel grooved lines ruled into a reflective metal surface. The detector mainly used in the MIR region is deuterated triglycine sulfate (DTGS) and in the NIR region is silicon (Si) and lead sulfide (PbS).

Some instruments are dedicated to specific applications. In these instruments, a number of optical interference filters are used to select discrete wavelengths that can be impinged on the sample. The filters are selected to transmit wavelengths that are known to be absorbed by the sample constituents. The instrument inserts filters one at a time into the light beam to direct individual wavelengths of radiation onto the sample.

The resolution of a dispersive/filter-based IR spectrometer is a function of the size of the entrance and exit slits of the monochromator. Both transmittance and reflectance measurements are possible with this type of instrument.

2.2 Fourier Transform Infrared Spectrometer

The typical Fourier transform infrared (FTIR) spectrometer has three basic components: a source, a detector and a Michelson interferometer. The sources used in FTIR in the MIR region are a ceramic element and a silicon carbide rod, both operating at approximately 1100 °C. The ceramic element is a spiral wound nichrome wire coated with a ceramic material, which is heated when

current is passed. The second source, silicon carbide, is quite rugged and efficient, but because of large size and power dissipation, water-cooled or air-cooled housings are necessary. Most of the FTIR spectrometers have a silicon carbide element as source. In the NIR region a tungsten halogen lamp is mostly used. Two different types of detector are mainly used in FTIR in the MIR region: a thermal detector such as DTGS, and a pyroelectric detector and photoconductive detector, such as the mercury cadmium telluride (MCT) detector. The detectors used in the NIR region are silicon (Si) and lead sulfide (PbS).

The unique part of an FTIR spectrometer is the Michelson interferometer shown in Figure 2. IR radiation from the source is made parallel and strikes the beam splitter at an angle of 45°. The beam splitter transmits half the radiation striking it and reflects the other half. The transmitted and reflected beams leave the beam splitter at right angles and strike two perpendicularly placed plane mirrors (see Figure 2), one of which can move in a direction perpendicular to the plane, which return the two beams to the beam splitter. The two beams recombine at the beam splitter and show interference. Fifty percent of the beam reflected from the fixed mirror is transmitted through the beam splitter and the other 50% is reflected back in the direction of the source. The radiation, which emerges from the interferometer at 90° to the input beam and is called the transmitted beam, is then passed to the sample, and subsequently goes to the detector. The moving mirror produces an optical path difference between the two arms of the interferometer. This is called the retardation and is twice the displacement of the moving mirror from the equidistant point. If a monochromatic source like a laser is used, the radiant energy reaching the detector will vary as a cosine function of the retardation. The detector response will reach a maximum every time the retardation is an integral number of wavelengths of the radiation. At this time, the beams from the two mirrors combine at the beam splitter in phase and show constructive interference. If the moving mirror is then moved one quarter of a wavelength of radiation, the retardation is changed by one half of a wavelength. The beams from the two mirrors combine at the beam splitter one half of a wavelength out of phase for the beam going to the detector and show destructive interference. The detector response as a function of the retardation is called the interferogram. The spectrum is generated from the interferogram with the help of Fourier transform.

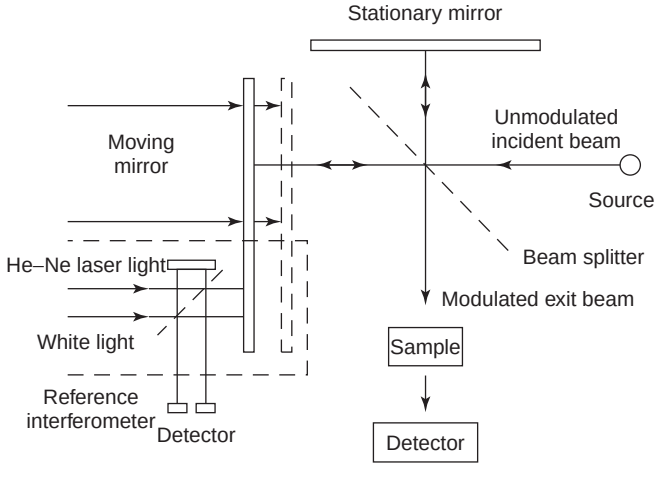

Figure 2 Michelson interferometer. (Reproduced with permission from B. Stuart, *Biological Applications of Infrared Spectroscopy*, Copyright John Wiley & Sons, Ltd.)

The main advantage of the FTIR spectrometer over the dispersive instrument is a high signal-to-noise ratio (S/N). This is because all the radiation frequencies reach the detector at the same time and the intensities of all the wavelength elements are analyzed simultaneously. Another advantage is that the spectrum can be recorded in less time.

3 SAMPLING TECHNIQUES

IR spectroscopy is a versatile analytical technique for qualitative and quantitative analysis of solids, liquids and gaseous samples.

3.1 Transmission Methods

Transmission spectroscopy is the oldest basic technique for sample analysis in the IR region. The method of analysis is based on the absorption of the IR beam by a sample at specific wavelengths. Solids, liquids and gases can be analyzed by transmission spectroscopy.

3.1.1 Solid Samples

Solid samples can be analyzed in the following forms:

- KBr Disc – A very popular technique for analysis of solids is the KBr disc. In this technique, a few milligrams of sample are finely ground with a mortar and pestle and then mixed with 50 to 100 parts of dry KBr powder. The mixture is placed in a special device and compressed into a disc at high pressure. This sinters the mixture and produces a clear transparent disc.
- Mulls – This method for solid samples involves grinding the sample and then suspending it in 1–2 drops of a mulling agent. This is followed by further grinding until a smooth paste is obtained. The paste so obtained is spread between two IR transmitting

windows. The most commonly used mulling agent is Nujol (liquid paraffin).

- Films – Spectra of solids not soluble in a suitable solvent are best obtained in the form of capillary films. A large drop of the neat liquid is placed between two IR transmitting windows, which are then squeezed together and mounted in the spectrometer in a suitable holder. Plates need not have high polish, but must be flat to avoid distortion of the spectrum.

  Thin films can often be prepared by casting a thin film from a solution of the sample and evaporating off the solvent. The film may be cast onto an IR transparent window or onto a suitable support from which it can be readily peeled. The film so obtained can be analyzed by using a film holder.

3.1.2 Liquid Samples

The easiest samples to run on an IR instrument are those in the liquid state. Slightly viscous samples can be squeezed between the two-polished IR transmitting plates and run as thin films. If the liquid is not viscous, fixed or variable pathlength can be used. Fixed pathlength sealed cells are useful for volatile liquids but cannot be disassembled for cleaning. Variable pathlength cells (Figure 3) are demountable so that the windows can be cleaned. An insoluble polymeric film is used as a spacer between two IR plates. Spacers are available in a variety of thicknesses allowing one to use various pathlengths. These cells are filled with the help of a syringe. An important consideration in the choice of IR cells is the type of window material. The material must be transparent to the incident IR radiation. Some commonly used IR materials are given in Table 1.

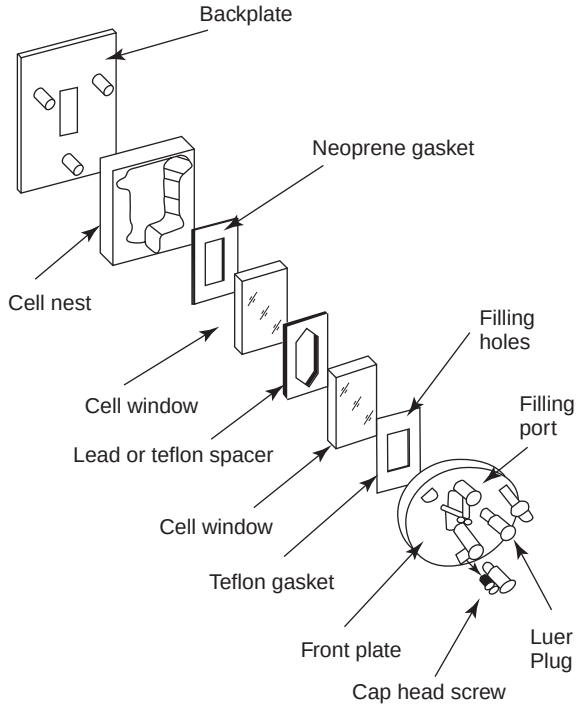

Figure 3 Variable pathlength liquid cell.

3.1.3 Gas Samples

Gas cells are essentially gastight containers fitted with IR transparent windows to enable the radiation to enter and exit the container, and a means for introducing, pressurizing and evacuating the gas. The gas cells are mainly constructed from borosilicate glass or stainless

Table 1 Properties of some IR spectrometer window materials

Material	Range (cm^{-1})	Refractive index	Properties
NaCl	40 000–600	1.52	Soluble in water, slightly soluble in alcohol, fair resistance to mechanical and thermal shock
KBr	43 500–400	1.54	Soluble in water and alcohol, hygroscopic, good resistance to mechanical and thermal shock
CaF_2	77 000–900	1.4	Insoluble in water, resists most acids and bases, does not fog, good for high pressure work
BaF_2	66 666–800	1.45	Insoluble in water, soluble in acids and NH_4Cl, does not fog, sensitive to thermal and mechanical shock
KCl	33 000–400	1.47	Similar properties to NaCl but less soluble, hygroscopic
CsBr	42 000–250	1.66	Soluble in water and acids, hygroscopic
CsI	42 000–200	1.74	Soluble in water and alcohol, hygroscopic
ZnSe	20 000–500	2.43	Insoluble in water, because in low absorbance in IR ideal for ATR work
Sapphire	50 000–1540	1.76	Insoluble in water, nontoxic
Ge	5000–500	4.01	Insoluble in water, soluble in hot H_2SO_4 and aqua, good for ATR
Si	8333–33	3.42	Insoluble in water, hard and inert

ATR, attenuated total reflectance.

steel with windows at each end. The gas density is many times lighter than liquid and solid at standard temperature and pressure. Therefore, transmission spectroscopy of gases requires cells with a longer pathlength than those used for liquid or solid analysis. For higher concentrations, a 10-cm long gas cell is generally used. For low concentrations, a long pathlength gas cell is used (few meters). Long pathlength is achieved in a short space by using a multipass cell in which the IR beam is bounced through the sample several times in order to obtain the desired pathlength. A typical gas cell is shown in Figure 4.

3.1.4 Cell Thickness Measurement

Cell thickness measurement can be done by an interference fringe method. This method is ideally suited to cells whose windows have a high polish. If an empty cell with parallel windows is placed in the spectrophotometer and a wavelength range scanned, an interference pattern is obtained which is as close as possible to the 100% line. The cell thickness is calculated from the expression, Equation (1),

$$b = \frac{n}{2}\eta_d(\nu_1 - \nu_2) \tag{1}$$

where n is the number of fringes (peaks or troughs) between two wavenumbers ν_1 and ν_2, and η_d is the refractive index of the sample material. If the measurements are made in wavelength, Equation (2) is:

$$b = \frac{n\lambda_1\lambda_2}{2}\eta_d(\lambda_1 - \lambda_2) \tag{2}$$

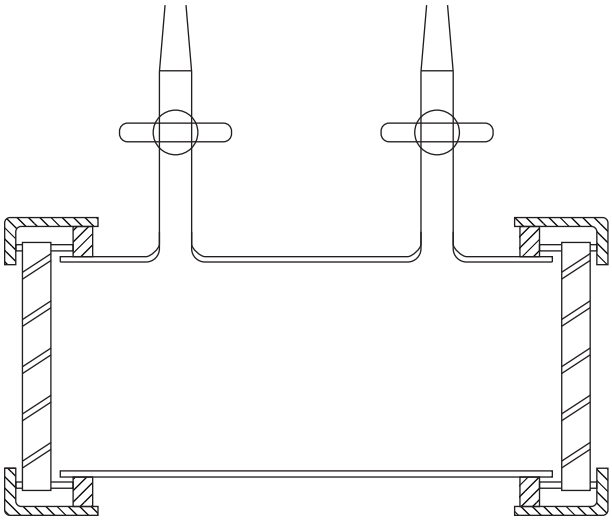

Figure 4 Gas cell.

where λ_1 is the starting wavelength and λ_2 the ending wavelength. The fringe method also works well for measurement of film thickness.

3.2 Reflectance Methods

Reflectance techniques are mainly used for samples which cannot be analyzed by conventional transmittance methods.

3.2.1 Specular Reflectance

When a beam of IR radiation is focused on to the sample, two types of reflectance may occur, namely specular and diffused. Specular reflectance occurs when the reflected angle of IR radiation equals the angle of incidence. The amount of light reflected depends on the angle of incidence, the refractive index, surface roughness and absorption properties of the sample. This technique is mainly used for surface measurements from the shiny surface of the sample.

3.2.2 Diffuse Reflectance

Diffuse reflectance mainly occurs from rough surfaces. As shown in Figure 5, the source radiation penetrates one or more particles in the sample and then is diffusely reflected in all directions. This radiation is collected and measured by a spectrophotometer. In diffuse reflectance the spectrum is processed using the Kubelka and Munk function, this changes the reflectance spectrum into one resembling a linear absorbance spectrum given as, Equation (3),

$$F(R_x) = \frac{(1 - R_x)^2}{2R_x} = \frac{c}{k} \tag{3}$$

where R_x is the reflectance of a thick scattering layer, c is the concentration and k is the molar absorption coefficient.

Many food samples give diffuse reflectance spectra. These include powders, fibers or matt surface samples. The spectrum is very sensitive to the particle size, which

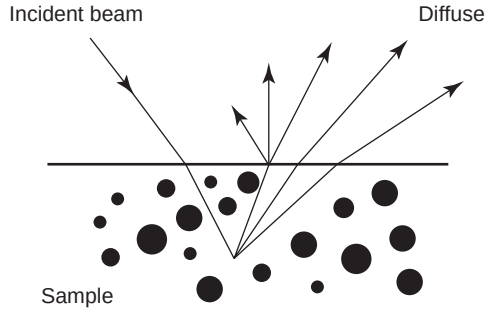

Figure 5 Representation of diffuse reflectance.

affects the scattering of radiation. If the particle size is not uniform, spectral distortion may occur.

3.2.3 Attenuated Total Reflectance

Attenuated total reflectance (ATR) spectroscopy has been one of the most popular sampling techniques employed for biological samples. ATR spectroscopy utilizes the phenomenon of total internal reflectance (Figure 6). A beam of IR radiation entering a crystal will undergo total internal reflection when the angle of incidence at the interface between the sample and crystal is greater than the critical angle. The critical angle is a function of the refractive indices of the two surfaces. The beam penetrates a fraction of a wavelength beyond the reflecting surface and when a material that absorbs radiation is in very close contact with the reflecting surface, the beam loses energy at the wavelength where the material absorbs. The resultant attenuated radiation is measured as a function of wavelength that gives rise to the absorption spectral characteristics of the sample.

In an ATR crystal, a standing wave is established near the surface of the crystal. The amplitude of the electric field decreases exponentially with distance from the surface of the ATR. The distance at which the field reaches 1/e of its initial magnitude is defined as the penetration depth d_p. The magnitude of d_p depends on the wavelength of the radiation, λ, the refractive index of the ATR crystal, n_p, the refractive index of the sample, n_s, and the angle of incidence of the beam at the surface of the IRE, ϕ, according to Equation (4)

$$d_p = \frac{\lambda}{2\pi} n_p (\sin^2 \phi - n_{sp}^2)^{1/2} \tag{4}$$

where $n_{sp} = n_s/n_p$ ($n_s < n_p$).

The crystals used in ATR cells are made from materials which have low solubility in water and are of high refractive index such as zinc selenide (ZnSe), germanium (Ge) and thallium/iodide (KRS-5). Solid samples, like powder, can be put directly onto the ATR crystal and compressed to ensure good surface contact. It is also possible to deposit materials onto the surface of the ATR crystal by applying drops and then evaporating the solvent. Liquid samples can be poured directly into the ATR trough.

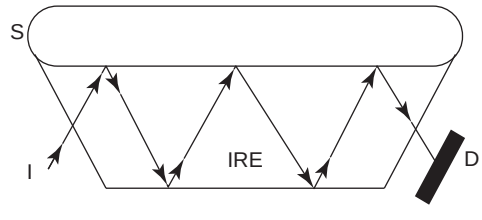

Figure 6 Horizontal ATR crystal accessory. S, sample; D, detector; I, IR beam and IRE, internal reflection element.

3.3 Fibre Optic Probe

An MIR fibre optic ATR probe is the most exciting new sampling technique available for IR scientists. This is extremely useful for analyzing samples that are at a remote location or for analyzing samples that are not of the optimum shape or size to fit into the sample compartment. NIR fiber optic transmittance and diffuse reflectance probes are also available for on-line process analysis.

4 QUANTITATIVE ANALYSIS

4.1 Beer–Lambert Plot

Quantitative measurements in absorption spectroscopy are based on the fundamental relationship between absorbance, pathlength and concentration. The relationship known as Beer's law is given by Equation (5)

$$A(\nu) = k(\nu) \times b \times c \tag{5}$$

where $A(\nu)$ is the absorbance at frequency ν, which is proportional to the product of the pathlength of the IR beam through the sample b and concentration of the sample c; $k(\nu)$ is the proportionality constant and is defined as the absorptivity of the substance being analyzed at frequency ν and is a measure of the substance's ability to absorb IR radiation of that frequency and to translate it into molecular vibrational energy.

4.2 Classical Least Square Method

When only a single IR component is present in a sample or in other words when an isolated nonoverlapped band can be identified, Equation (1) can be used to develop a quantitative method for determining the component's concentration. However, in complex systems like food products, several compounds are usually found. In this case multicomponent quantitative analysis becomes important. Beer's law has to be modified and written as, Equation (6)

$$A_{\lambda 1} = K_{a,\lambda 1} C_a + E_a \tag{6}$$

where C_a is the concentration of component A, $A_{\lambda 1}$ is the absorbance at wavelength 1, and E_a is the error in the concentration predicted by least squares fit line and actual concentration. According to Equation (6), the concentration is a function of the absorbances at a series of given wavelengths. This method is also called the *K*

matrix method. Using matrix algebra, Equation (6) can be solved for a best fit least square line to the data. Once the equation has been solved for the K matrix, it can be used to predict concentrations of unknown samples. One disadvantage of this technique is that the equations must be calibrated for every component in the mixture. If the concentration of one component is omitted, the predicted absorbance will be incorrect. In other words, the classical least square (CLS) technique can only be applied to systems where the concentration of every component in the mixture is known. In addition, there must be no significant interactions between the different components when they are mixed together. For example, when two liquids are mixed together, there may be some interaction that occurs to create a small amount of new compound. This new compound will have a spectrum of its own, and thus, it will also have a different set of absorptivity coefficients from the other two components. If the concentration of this new compound is not included in the calibration, the calculated K matrix will not reflect the absorptivities of this component and predictions of unknown will be in error. Typically the CLS calibration method is most useful when applied to samples which have minimal intercomponent interactions such as many gas phase mixtures.

4.3 Inverse Least Square Method

As discussed in the previous section, if the concentrations of all the components in the mixtures are not known, the CLS method does not give very good calibration equation and there may be a significant amount of error. In these cases, the inverse least square (ILS) method is used.

Beer's law can be rearranged and written as Equation (7)

$$C_a = A_{\lambda 1} P_{a,\lambda 1} + E_a \tag{7}$$

where C_a is the concentration of component A, $A_{\lambda 1}$ is the absorbance at wavelength 1, and E_a is the error in the concentration predicted by the least squares fit line and actual concentration. According to Equation (7), the concentration is a function of the absorbances at a series of given wavelengths. This method is also called the P matrix method. This method is the best method for almost all quantitative work as knowledge of the sample composition is not required.

This method also has some disadvantages. Owing to the dimensionality of matrix equations, the number of selected wavelengths cannot exceed the number of training samples. This problem can be overcome by measuring many more samples to allow additional wavelengths. However this causes the problem of collinearity and the solution becomes unstable with respect to each component.

4.4 Partial Least Square/Principle Component Analysis

Partial least square (PLS) and principal component regression (PCR) are the most widely used multivariate calibration methods used in food science. Both PLS and PCR methods make use of the inverse calibration approach, where it is possible to calibrate for the desired component while implicitly modeling the other sources of variation. In both methods, the inverse problem is solved by replacing the original variables with linear combinations of the variables known as factors. The difference between the PCR and PLS factor analysis methods lies in the fact that PCR performs the factor analysis in such a way that the factors describe only the variance in the calibration spectra. Variations in the spectra that are not related to the concentration of the analyte may receive large weight in the analysis. The PLS method, however, has been designed to extract factors that simultaneously account for the variance in the spectral data and correlate with analyte concentrations. Thus, better concentration prediction ability is expected for PLS relative to PCR when a portion of the variation in the spectral data does not correlate well with concentration variations.[1] Mathematical descriptions of both methods are given elsewhere.[2]

4.5 Calibration

The first step in calibrating an instrument, for a particular application, is to select a set of calibration or training samples. The samples are representative of the product that has to be analyzed. Training samples should cover the full range of the product and should have a relatively uniform distribution of concentrations across that range. The calibration samples must be analyzed by the traditional reference method normally used for that constituent. The spectral data are obtained for each training sample with the IR instrument at all available wavelengths. All the data are stored in the computer memory and quantitative analysis is performed on the data.

Quantitative analysis of the spectral data is a three-step procedure. The first step is the calibration, in this the regression model that best relates the IR spectral data to the reference data is determined. The second step is the validation, in which known samples are used to ensure that an overfitted solution has not been obtained and to establish the robustness of the calibration. After the calibration has been established, the third step is prediction with an unknown sample.

The greatest error sources in any calibration are generally reference laboratory error, repack error (non-homogeneity) and nonrepresentative sampling in the calibration set. Performing an in-house audit of procedures, equipment and paying particular attention on sample preparation can reduce these error sources.

4.6 Error Analysis

The root mean square error and the multiple correlation coefficient are both ways of calculating how well the calibration equation fits the data. The values of the root mean square difference (RMSD), which is an indication of the average error in the analysis, for each component is, Equation (8)

$$\text{RMSD} = \left[\frac{1}{N}\sum_{i=1}^{N}(A_i - B_i)^2\right]^{0.5} \tag{8}$$

and the square of the correlation coefficient (R^2), which is an indication of the quality of fit is, Equation (9)

$$R^2 = \frac{\sum_{i=1}^{N}(A_i - C)^2}{\sum_{i=1}^{N}(B_i - C)^2} \tag{9}$$

where A_i is the actual concentration of the analyte in the sample i, B_i represents the estimated concentration of the analyte in the sample i, C is the mean of the actual concentrations in the prediction set and N is the total number of the samples used in the prediction set.

Normally a correlation coefficient can have values between 0 and 1. The value close to zero indicates that the calibration is failing to relate the obtained values to the reference values. As the correlation coefficient increases the obtained values become better and better indicators of the reference values until, when the correlation coefficient reaches unity, the obtained values and the reference values are identical in all cases.

5 FOOD APPLICATIONS

IR spectroscopy in both the MIR and NIR regions is emerging as a qualitative as well as quantitative analytical technique with great potential in a wide range of applied areas. IR spectroscopy mainly in the NIR region is widely used in food and agro-based industries. A detailed bibliography has been published in conference proceedings[3,4] and a review article.[5,6]

5.1 Analysis of Water

IR spectroscopy in the NIR region is considered to be a universal approach for determination of moisture content in food products because water has strong absorption in the NIR region providing the sensitivity needed for accurate determination. In addition, changes in the water absorption band are reflective of chemical interactions with other molecular species in the samples, and therefore it can be used for analysis of other food constituents, which do not have their own characteristic absorption bands. The spectrum of water consists mainly of two composite bands, the first overtone at 1450 nm and combination band at 1940 nm at room temperature (Figure 7). These bands are influenced by changes in the hydrogen bonding. The O–H band position and patterns differ from one food product to another. The IR spectrum of water is also a function of temperature. As a function of changing temperature, the relative properties of the individual bands making up these composite bands change, resulting in a shift in the band position. The influence on water absorption caused by variation in temperature has been used to develop a calibration equation with temperature compensation to determine the brix value in peaches.[7]

The changes in chemical composition of food products during aging affect the states of hydrogen bonding of O–H bonds present in the water content. The staling and aging of food products, for example, wheat flour is easily studied using NIR spectroscopy.[8]

5.2 Analysis of Sugars

The sugar cane refinery depends on constant monitoring of in-process and finished product streams in order to maintain process control, plant efficiency and on-going product quality. Testing of sugar cane for sucrose content and other common constituents like glucose and fructose has traditionally been done in sugar industry by a variety of gravimetric and wet chemical methods. While these standard techniques are reliable, they are often time-consuming, operator dependent and involve the use of hazardous reagents. Second, the most commonly used technique in the sugar industry – sucrometry/polarimetry – is not a true indicator of

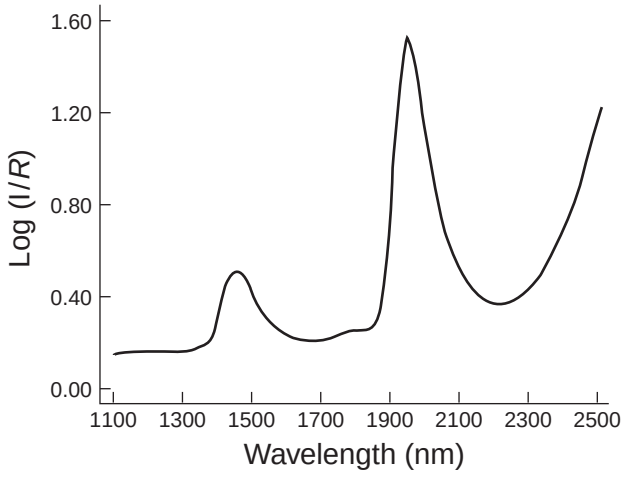

Figure 7 NIR spectrum of water at room temperature.

sucrose content and does not provide the concentration of individual components (sucrose, glucose and fructose) present in the sugar cane juices. More complete analytical methods are desired for better characterization of sugar mixture, which can readily be adopted for on-line work.

Quantitative determination of the main constituents of a sugar mixture can be carried out by IR spectroscopy. Different types of sampling techniques can be adopted for analysis of sugar cane juices and sugar cane solids in both the MIR and NIR regions. The juices can be handled by ATR sampler in the MIR region. Typical spectra of sugar cane juices and syrup[9] obtained in the MIR region using the ATR accessory are shown in Figure 8. It is important to identify the peaks of different sugars by taking the spectra of pure components individually (Figure 9, Table 2). For quantitative analysis, different

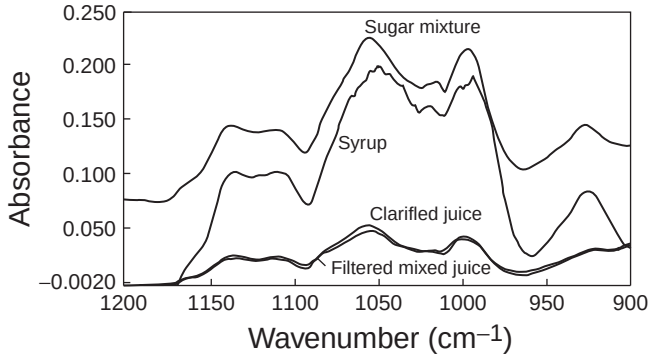

Figure 8 FTIR/ATR spectra of sugar cane juices and sugar cane syrup.

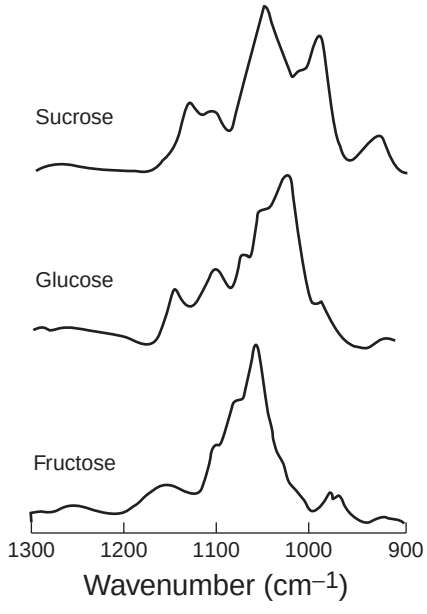

Figure 9 FTIR/ATR spectra of different components of a sugar mixture.

Table 2 Characteristic IR bands of sugars commonly found in food products

Component	Wavenumber (cm^{-1})	Assignment
Sucrose	1050	C–O stretching
	980	C–O deformation
Glucose	1015	C–O stretching
Fructose	1065	C–O stretching

statistical methods can be used to obtain calibrations for synthetic mixtures of sucrose, glucose and fructose in concentration ranges typically encountered in sugar cane juice. The calibrations can be performed on real sugar cane juice samples for better results. In this case, the reference values of the training set can be obtained by traditional chemical methods.

A diffuse reflectance accessory in the NIR region works best for sugar cane solids. Figure 10 shows the spectrum obtained for a prepared cane sample in diffuse reflectance mode.[10] The water absorption bands at 1450 nm and 1930 nm mainly dominate the spectrum. The O–H band absorption is masked beneath the water O–H vibrational signal. The main region used for sugar analysis is 2000–2500 nm. Another region 1600–1900 nm can also be exploited for quantitative work. Particle size plays a very important role while analyzing solid samples. The solid sample should be properly ground and homogenized for accurate results.

5.3 Analysis of Oils and Fats

Studies on oils and fats can be strongly profiled by IR spectroscopic techniques because oils are mainly composed of lipids, which have strong characteristic absorption owing to vibrations of the carbonyl group and hydrocarbon chain in the MIR region.

Several different approaches have been used for qualitative and quantitative analysis of oils and fats. In the

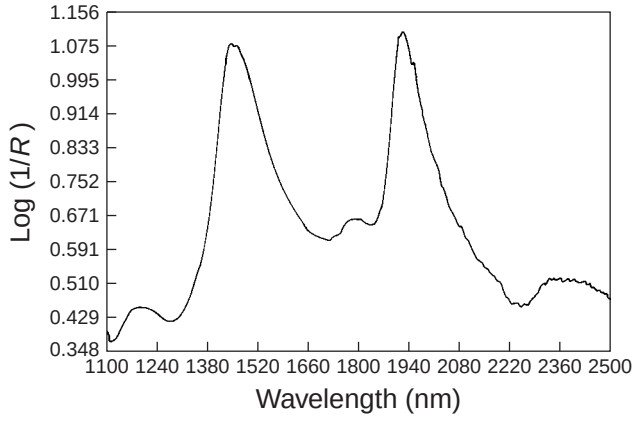

Figure 10 NIR spectra of sugar cane solids (prepared cane).

MIR range the ATR sampling technique is mainly used either with pure samples or their solutions in solvents like ethylether.[11] CaF_2 flow cells with pathlengths of 0.1–1 mm are also applicable. Several accessories have also been designed for analyzing oil on-line using NaCl or KBr discs.[12] A fiber optic probe with chalcogenide fiber equipped with a two reflection ZnSe ATR crystal has also been designed and spectra can be measured simply by immersing the probe in the oil sample.[11]

Beside characterization of different edible oils,[13] detection of adulterations,[14] determination of degree of unsaturation,[15] determination of *cis* and *trans* isomers,[16] fatty acid contents and total solid fats,[17] and monitoring auto-oxidation of oils and fats[18] are some of the interesting applications of IR in the area of oils and fats.

The characterization and discrimination between different oils is based on some characteristic absorptions, mainly asymmetric C–O stretching frequencies of the ester grouping (C–C–O) that appear around 1300 cm^{-1} and asymmetric C–O stretching of the O–C–C grouping that appears around 1000 cm^{-1} (Table 3). FTIR along with chemometric techniques has made the classification of a large number of commercially available vegetable oils and margarines possible.[19] Using a sealed fixed pathlength (NaCl) cell, it is possible to distinguish edible oils from other oils according to their vegetable origin.[13] Analysis was based on determination of *cis* and *trans* double bond content and detection of impurities. Absorbancies corresponding to the *cis* double bond content are predominately responsible for this discrimination. Partially hydrogenated vegetable oil can be distinguished through *trans* C–H bending absorbance. Detection of impurities in sesame oil present in the *trans* form like lignin, sesame, sesamolin and sesamol and so on has been possible using their characteristic absorption in the fingerprint region.

Table 3 Characteristic IR bands of lipids and fats in oil samples

Component	Wavenumber (cm^{-1})	Assignment
Lipids and fats	3010	C–H stretching
	2956	CH_3 asymmetric stretching
	2920	CH_2 asymmetric stretching
	2870	CH_3 symmetric stretching
	2850	CH_2 symmetric stretching
	1750–1710	C=O stretching
	1473–1463	CH_2 scissoring
	1460	CH_3 asymmetrical bending
	1400	CH_2 wagging
	1400–1000	C–O stretching
	968	C=C bending out of plane
	914	C=C bending out of plane

The NIR region has also been widely used in studies of fats and oils. Even though well-resolved bands are very rarely obtained and spectra consist of broad overlapping bands which cannot be assigned to a single chemical. But with the help of chemometric techniques, it is possible to analyze biological samples especially food products, qualitatively and quantitatively, using NIR spectroscopy.

Determination of *cis* and *trans* double bond content and detection of impurities can be attempted by NIR reflectance spectroscopy. Vegetable oils like cottonseed, peanut, soyabean and canola are differentiated using discriminate analysis in NIR region.[20] Wavelength regions from 1600 to 1800 nm and from 2100 to 2200 nm are selected for analysis since most differences in the NIR spectra patterns of oils are evident in these regions. Using the second order spectra, it is possible to separate the spectra of different vegetables. Choosing the same selected wavelengths and using discrimination analysis it is possible to classify unknown samples. The spectral region 1600–2200 nm of oils contains information about the fatty acid composition. The bands assigned to straight carbon chains are *cis* double bonds and are mainly located here. By applying principal component analysis (PCA) on the standardized full NIR spectral data of this region, classification of vegetable oils related to the fatty acid composition can be performed rapidly and easily.

5.4 Analysis of Milk and Dairy Products

Milk, butter, cheese, milk powder and some special milk-based dietary foods like those for babies, constitute the wide range of dairy products where the quantitative analysis of various constituents is desired. The main ingredients of interest to the dairy industry are butterfat and protein. In dried milk powders and special foods, determination of moisture and lactose is also very important. This forces industry to test thousands of samples. IR spectroscopy is now being widely used by the dairy industry as an analytical technique for various constituents. It is now an established AOAC (Association of Official Analytical Chemists) method for determination of fat, protein and total solids in liquid milk samples.

The casein content in raw milk samples has been determined using the FTIR spectroscopic method. A flow cell maintained at 40 °C is used as a sampler, the milk samples are homogenized and absorption spectra are collected in the MIR range. The spectral bands of C–H and amide are correlated with the concentration of casein. PLS analysis and PCR are used for quantitative analysis.[21]

Besides quantification of fat, protein and lactose in milk, IR spectroscopy in the MIR region also finds

application in analysis of milk fat, like determination of the *trans* fatty acid content in solid samples, milk powder and cheese.[22] IR absorption spectroscopy does not find much use since blending and homogenizing the solutions of cheese and powder milk in water has become mandatory. IR reflectance spectroscopy mainly in the ATR mode, is used for these samples.

Reflectance spectroscopy is mainly used in the NIR region. Analysis of raw milk for measuring the composition in terms of fat, protein, lactose and total solids is an established method. Figure 11 shows the spectra of milk and water. As shown in Figure 11, the broad and intensive absorption bands of water overlap the absorption bands of protein and fat in this region. This overlap makes it difficult to correlate specific bands to fat and protein. These bands can be seen after transforming the raw data to second derivatives (Figure 12). Table 4 gives the NIR bands assigned to fat, protein and lactose in milk. Another problem is the hindrance to spectral deformation caused by scattering by fat globules. By applying PCA to the NIR spectra, it has been possible to assign some specific wavelengths to fat, protein, and lactose and to discriminate between the samples.[23]

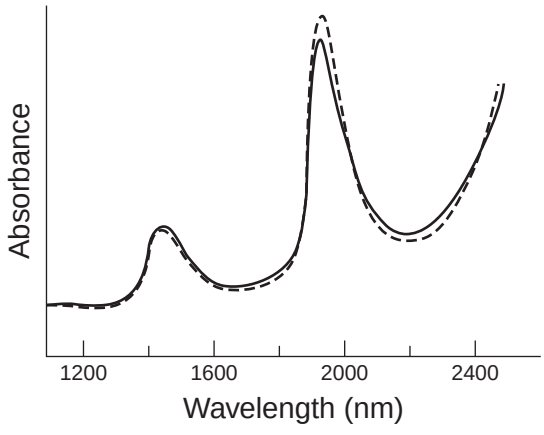

Figure 11 NIR spectra of milk (solid line) and water (dashed line).

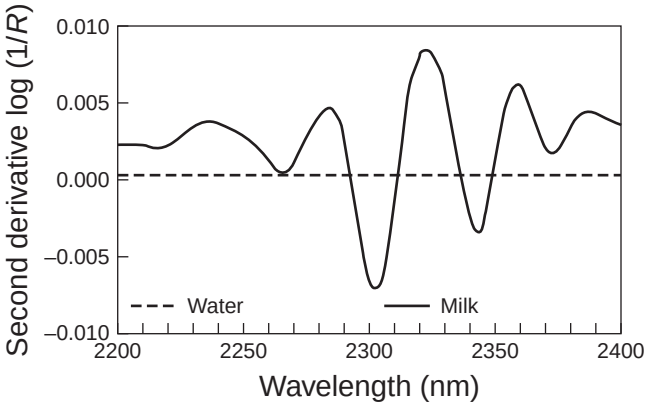

Figure 12 Second derivative spectra of milk and water.

Table 4 Specific bands of fat, protein and lactose in raw milk in NIR region

Component	Wavelength (nm)	Assignment
Fat	1724	C–H stretch 1st overtone
	1752	C–H stretch 1st overtone
	2308	C–H bend 2nd overtone
	2344	C–H_2 bend 2nd overtone
Protein	2050	N–H/amide 2nd or N–H/amide 3rd or combination
	2180	N–H bend 2nd overtone
		C–H stretch/C=O stretch combination
		C=O stretch/amide 3rd combination
Lactose	2094	O–H combination

NIR spectroscopy has also been applied to the analysis of dried products for fat, moisture and protein to obtain the right composition mixture for special products. Because of uniform particle size, diffuse reflectance measurement in the case of spray dried products are much more suitable.[24]

Cheese is another important dairy product. It is difficult to analyze using IR spectroscopy because of differences in processing, physical and chemical composition, and the high level of moisture and milk fat. However, few sampling techniques have been standardized for the reflectance mode to determine moisture, fat and salt content in cheese. A homogenized sample of processed cheese has been analyzed using the ATR technique.[25] The bands at 1735 cm^{-1}, 1244 cm^{-1} and 1165 cm^{-1} (Figure 13, Table 5) assigned to ester groups in fats are used for quantitative determination of fats. For protein content, the 1545 cm^{-1} band is more suitable since the 1640 cm^{-1} band is partly due to proteins and partly due to water.

Reflectance and transmittance measurements on grated cheese samples and also on intact samples using a fiber optic probe can be successfully used for quantitative analysis of fat and moisture content.[26] The composition of butter with respect to moisture and the solids nonfat content can also be determined using NIR spectroscopy.

Another important application of IR spectroscopy is the detection of adulteration. Foreign fat adulteration in dairy products, butter and margarine can be done by studying the fatty acid profile of their NIR spectra.[27] NIR spectroscopy is also useful for characterization of spectral variations induced by the presence of different kinds of contaminants in milk.[28] This can be used to predict the addition of foreign substances to milk.

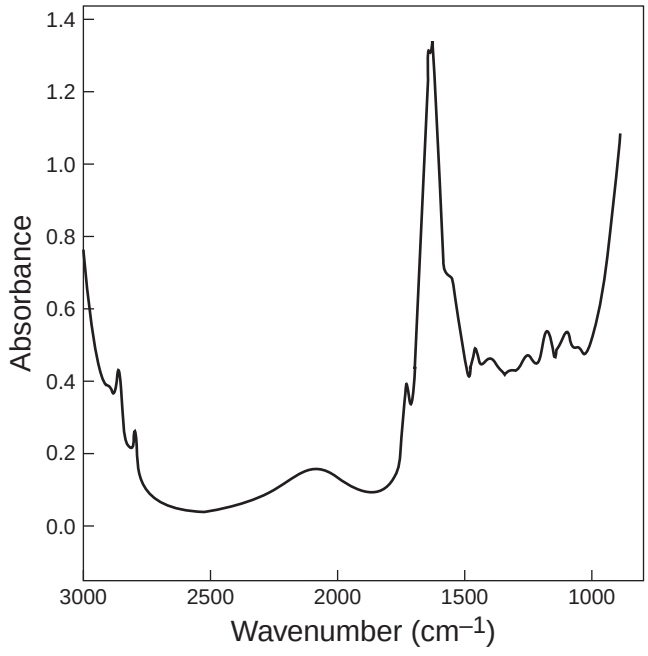

Figure 13 FTIR/ATR spectrum of processed cheese.

Table 5 Characteristic IR bands in cheese samples

Component	Wavenumber (cm^{-1})	Assignment
Fat	1743	C=O
Protein	1634	Amide I
Lactose	1165	C–O stretching
	1082	C–O stretching
Methylene groups	2920	CH_2 stretching
	2854	CH_2 stretching
	1396	CH_2 bending
	1358	CH_2 bending

5.5 Analysis of Beverages and Juices

The IR technique is useful for determination of different constituents in different types of fruit juices and beverages. The percentage of absolute alcohol, methanol and other higher alcohols in alcoholic beverages like beer, wine, whiskies and brandies can be estimated using both MIR and NIR spectroscopy. A procedure has been proposed for ethanol determination in the vapor phase in the MIR region. The method is based on injection of small volumes of untreated samples into a heated pyrex glass reactor in which, at a temperature between 80 and 100 °C, the ethanol is volatilized and introduced by means of an FTIR spectrometer. By measuring the area of the flow injection, the recording obtained from the absorbance of transient signals in the wavenumber range between 1150 and 950 cm^{-1}, allows direct quantification of ethanol without the water background problem or interference from sugars. The method has been applied from low alcohol beverages to wines and spirits.[29] The important bands of interest in the case of alcohols are C–O stretching (1000–1300 cm^{-1}) and O–H stretching (2500–3650 cm^{-1}), which are mainly used for quantitative analysis.

The ATR sampling technique in the MIR region is used for quantification of ethanol and methanol in different varieties of whiskies. Figure 14 shows the spectrum of a mixture of water–methanol–ethanol–acetone, after subtracting the dominant water absorption band, to characterize successfully the peaks that are due to each component.[30] Calibration has been performed with a large number of synthetic samples. The predicted values match closely the reference values obtained by chemical methods with 99.9% accuracy. Besides quantification using PCR or PLS analysis, it is also possible to determine molecular structure from the vapor phase FTIR spectra and set up a discriminating rule that classifies alcoholic and nonalcoholic beverages.[31]

A fiber optic transmission probe working in the NIR region can also be used to estimate ethanol in alcoholic beverages as well as during the timespan of a fermentation process.[32] A silica optical fiber of specific design can be used in-line in a process plant and it is possible to transmit NIR radiation in the 1–2 μm waveband over fiber distances of up to 1–2 km for process analysis and to determine ethanol percentage.[33]

Quantitative analysis of individual sugars and acids in fruit juices has been performed both in the MIR and NIR regions using different sampling techniques. The ATR sampling technique is used for estimation of sugars and organic acids in apple juices in the MIR region. The spectral region 900–1200 cm^{-1} encloses the absorption bands for all sugar components (Figure 9). A typical spectrum of apple juice is shown in Figure 15. Calibration has been performed on synthetic solutions using PCR and PLS algorithms.[34]

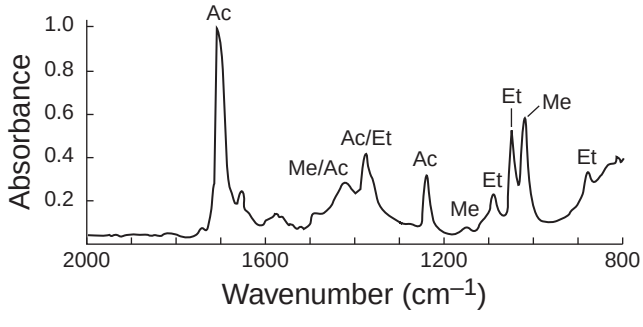

Figure 14 FTIR/ATR spectrum of a mixture of water–methanol–ethanol–acetone after subtracting the water. Ac, acetone, 0.277%, Me, methanol, 0.058%, Et, ethanol, 0.083%.

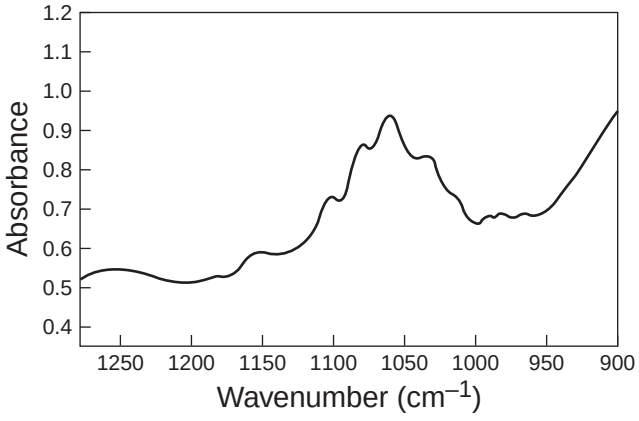

Figure 15 FTIR/ATR spectrum of apple juice.

5.6 Analysis of Grains

Application of IR spectroscopy to the evaluation of essential nutrients in food and agricultural commodities is being extensively explored in the area of grains and cereals. In the MIR region, the ATR sampling method can be used for protein, starch and carbohydrate analysis. The main regions of interest are listed in Table 6. The sample must normally be ground prior to analysis. The requirement for grinding must be related to the heterogeneity of sample and also to the particle size distribution.

However, the main contribution to evaluation of grains is made by NIR spectroscopy. The technique is being used to screen breeding and commercial samples for more desirable concentrations of constituents, which relate to the quality desired or preferred by processors and consumers. In addition to protein and moisture content, NIR spectroscopy is used to determine hardness, sieving test ash, color, starch damage, water absorption and loaf volume in wheat or flours[35] or amylase content in rice.[36] The beer-making potential of barley[35] and the degradability of barley forages[37] is also being judged by NIR spectroscopy. It can determinate the quality of oil seeds and legume grains for degree of saturation and composition.[38]

Initially NIR spectroscopy was applied to determination of moisture, protein, lipid and bran in wheat, but now a classification system based on spectral analysis has been developed which can differentiate between different varieties. Two-dimensional correlation spectroscopy across the MIR (Figure 16) and NIR (Figure 17) regions using correlation between amide I C=O/N–H stretch and phenolic stretch in the MIR with traditional starch C–O–H at 2104 nm in the NIR is used to distinguish winter from spring wheat.[39]

The NIR image processing technique using a charge-coupled device (CCD) camera has been used for discriminating between varieties of rice giving around 90% accuracy. Sampling cells, which can be used

Table 6 Characteristic IR bands present in grains

Component	Wavenumber (cm^{-1})	Assignment
Proteins	3300	N–H stretching
	1650–1660	C=O and N–H stretching
	1550–1600	Amide II
	1260–1350	Amide III
	627	Amide IV
	725	Amide V
Starch and carbohydrates	1000–1400	C–O stretching
Moisture	3300–3400	O–H stretching

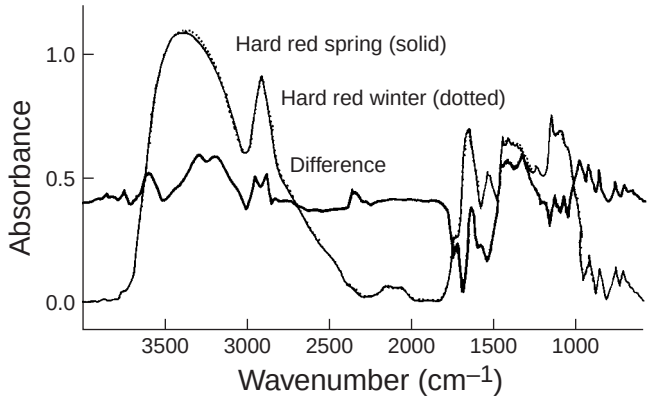

Figure 16 MIR spectra of hard red winter and spring wheats and their difference spectrum. (Reproduced from Barton et al., *J. Near Infrared Spectrosc.*, **4**, 139–152 (1996),[39] NIR Publications.)

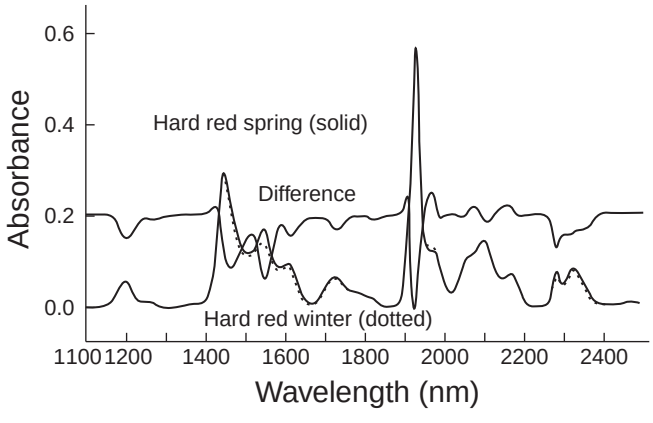

Figure 17 NIR spectra of hard red winter and spring wheats and their difference spectrum. (Reproduced from Barton et al., *J. Near Infrared Spectrosc.*, **4**, 139–152 (1996),[39] NIR Publications.)

to grain a single rice kernel and measure the spectra, are used to predict the illegal mixture over two varieties.[40] To determine moisture, a solvent extraction method has been used which is based on shift in the water band caused by variation in water content in an organic solvent like *N*,*N*-dimethylformamide and dimethylsulfoxide.[41]

5.7 Analysis of Meat

Fat, protein and moisture contents of meat are very important and are commonly being obtained by the extraction–rinsing solvent method. NIR reflectance is a rapid method for determining the composition of meat. The residues from extracts of pork meat after being washed and dried were considered. The second derivatives of the spectra were recorded. The second derivative minimum/maximum peaks of protein are observed at 2408 nm and 2338 nm, while minimum/maximum peaks of fat are at 1212 nm, 1284 nm, 2204 nm and 1188 nm. Moisture shows minimum/maximum peaks at 1734 nm, 1774 nm, 2290 nm, and 1698 nm in the second derivative spectra recorded in the NIR reflectance mode.

In another experiment an 11 filter NIR transmittance analyzer (Meatspec analyzer) is used, in the 825–1075-nm range, to analyze two different types of meat samples, beef and pork.[42] A system based on the use of PCA has been devised for testing the identity and the homogeneity of the samples. The system checks the spectral characteristics of the sample and rejects any sample which shows a deviation from the mean of all samples that is greater than a previously defined specification.

6 GAS CHROMATOGRAPHY/INFRARED SPECTROSCOPY

Gas chromatography (GC) for separation of volatile components of a mixture is an important tool for analytical chemists. FTIR has arisen as an important detector for gas chromatographic elutes. GC elutes are generally samples ranging from a few tens of micrograms to less than a nanogram. An FTIR spectrum interfaced to a gas chromatograph provides structural information about GC elutes. The FTIR offers a real time and nondestructive IR chromatographic detection of elutes.

6.1 Gas Chromatography/Infrared Instrumentation

A gas chromatography/Fourier transform infrared (GC/FTIR) system consists of mainly a gas chromatograph, FTIR spectrophotometer, lightpipe, detector and computer. A schematic diagram is shown in Figure 18.[43]

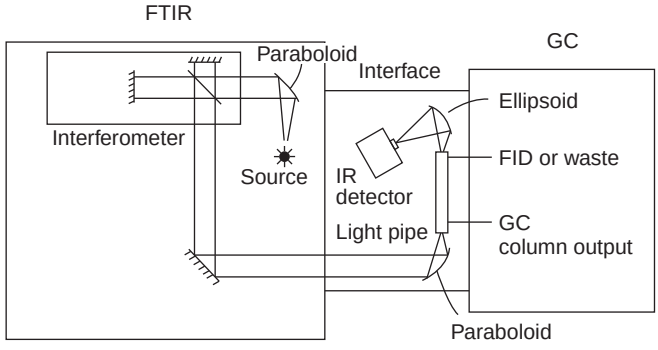

Figure 18 Schematic diagram of a GC/FTIR spectrometry system. (Reproduced with permission from Griffiths and de Haseth,[43] Copyright John Wiley & Sons, Ltd.)

The FTIR spectrophotometer is generally at low resolution (8 cm^{-1}), the bandwidth of gaseous components being generally large owing to contributions from rotational transitions. A stable IR source is generally used since the source energy through the light pipe is cut to about 10%. A narrow band 700–4000 cm^{-1} MCT detector is commonly used. The spectral range is sufficient for detection of most of the compounds. The light pipe, the interface between GC and FTIR, is the most important component and is generally made from quartz. Its dimensions are based upon the width of GC peaks or, in other words, the type of GC column.

A range of GC techniques is applicable to the GC/FTIR system. Capillary column GC gives better chromatographic resolution and efficiency but has lower loading capacity. Use of capillary columns with a 0.5–1 μm film coating is currently very popular.

For higher loading capacity, packed columns are generally used. Certain other modifications have also been used. One of them is a matrix isolation technique in which the GC elutes are trapped in a frozen matrix of argon atoms. IR radiation is focussed upon a frozen sample and the reflection absorption spectrum is collected. The technique is more sensitive and gives high S/N. The chromatogram can also be frozen on an IR window kept at low temperatures and the spectrum can be collected in transmission mode. The data processing system collects information from an interferogram and processes to create the chromatograms.[44]

6.2 Applications of Gas Chromatography/Infrared Spectroscopy

The applications of gas chromatography/infrared (GC/IR) spectroscopy are versatile. The area of food analysis and product safety is of prime importance. A major field for the application for GC/FTIR is in the identification of important food constituents or for trace contaminant analysis.

The GC/FTIR technique is mainly used for identification of aroma and volatile components. Coffee and chocolates are important consumer products. Flavor is their important quality parameter. The aroma from a coffee plant can be analyzed using the GC/FTIR technique.[45] Nitrogen gas is first passed through a chamber containing coffee berries followed by a trap for concentrating the aroma components. A specially designed desorption–injection apparatus is used to deposit the concentrates on to the GC column for identification using GC/FTIR. With this technique, it is possible to identify various compounds in the aroma. Headspace GC is also used for monitoring volatilies in beverages. Heating the sample at 100 °C, the vapor is compressed and introduced onto a gas chromatograph equipped with a flame ionization detector (FID) that analyzes the volatilies from coffee samples. The volatile species including esters, aldehydes, ketones, acids, amines, alcohols and monoterpenes are effectively separated by GC and classified by FTIR using selected wavelength regions. The 1680–1780 cm^{-1} region is useful for esters, aldehydes and ketones. The C–O stretch region, 1300–1000 cm^{-1}, is used for identifying alcohols, esters and the C–N stretch vibration of amines. The 2800–3000 cm^{-1} region is selected for aliphatic hydrocarbons and 3000–3140 cm^{-1} for unsaturated hydrocarbons.[46]

GC/FTIR has tremendous power in discriminating between geometrical isomers. The technique finds use in analysis of essential oils that contain a mixture of geometrical isomers. The technique also has high sensitivity for carbonyl and hydroxyl functionalities present in substances responsible for imparting flavor and fragrance to these essential oils. Using GC/FTIR, more than 260 components are detected in peppermint oil extract.[47]

Another novel and interesting application is to check the freshness of food products. Many components are shown to be produced in trace amounts during staling of food. One such component, ethyl carbamate is considered to be a carcinogen and is produced during the fermentation process. GC/FTIR is a useful analytical technique for its powers of detection in beverages and foods. An internal standard method using ethyl carbamate isotopically labeled with ^{13}C and ^{15}N in less than 1-mg quantities is used. The standard is added to whisky and the validity of technique in detecting <1-mg ethyl carbamate is checked.[48]

7 ADVANTAGES AND LIMITATIONS

The prime advantage of IR spectroscopy is that once the instrument has been calibrated, several food constituents in a sample can be measured rapidly and simultaneously. No sample manipulation is required. It is also applicable for on-line analysis in the food processing industry. Table 7 gives a comparison of the IR technique with other techniques.

There are a few limitations to working with IR spectroscopic methods. The main drawback is that the technique is indirect, based on reference values, it is subjected to errors. Samples of biological origin are especially complex and difficult to handle. Even minute changes in absorbance indicate large concentration changes and the bands representing the components are overlapped by interfering bands. As a result the detection limit of the technique cannot be reduced beyond 0.5–1.0%. The analysis is also affected by the differences in performance and the type of calibration program used in a particular application.

In the area of food applications solid sampling is often the method of choice by users. The diffuse reflectance technique is prone to error. The intensity of diffusely reflected energy is dependent on sample packing, density, sample crystalline structure, refractive index, particle size distribution, and absorptive qualities of the material. In addition, errors introduced in laboratory reference methods, nonhomogeneity of samples introduced during packing or repacking and also the variance in sample population are other noticeable drawbacks associated with the IR technique. All these errors add up to cause a large variation in analysis. Quantitative analysis in IR spectroscopy also depends on how well the Beer–Lambert law is fulfilled. The large concentration ranges and spectral abnormalities introduced by

Table 7 Comparison of IR spectroscopy versus other methods

IR spectroscopy	Chemical method/HPLC/ polarimetry/others
Nondestructive technique	Destructive technique
No sample preparation	Sample preparation required
Short analysis time only 1 min after calibration is done	Close to half an hour to two hours depending upon the method used
Environmental friendly (no use of hazardous chemicals)	Use of chemicals are necessary
No sample blockage	In case of HPLC, suspension particles can block the columns In case of polarimetry, opaque solution prevents the determination
Solids and liquids can be handled by the same instrument	Each component (solid and liquid) analysis involves different equipment
On-line analysis is possible	On-line analysis is not possible

HPLC, high-performance liquid chromatography.

factors like those discussed above cause nonlinearity in Beer's law.

Performing an in-house audit of procedures and equipment and paying particular attention to sample preparation can reduce these sources of error.

ACKNOWLEDGMENTS

The author is thankful to all her colleagues specially Dr Alka Gupta for help rendered during preparation of the manuscript.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
ATR	Attenuated Total Reflectance
CCD	Charge-coupled Device
CLS	Classical Least Square
DTGS	Deuterated Triglycine Sulfate
FID	Flame Ionization Detector
FIR	Far-infrared
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/FTIR	Gas Chromatography/Fourier Transform Infrared
GC/IR	Gas Chromatography/Infrared
HPLC	High-performance Liquid Chromatography
ILS	Inverse Least Square
IR	Infrared
IRE	Internal Reflection Element
MCT	Mercury Cadmium Telluride
MIR	Mid-infrared
NIR	Near-infrared
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Square
RMSD	Root Mean Square Difference
S/N	Signal-to-noise Ratio

RELATED ARTICLES

Food **(Volume 5)**
Near-infrared Spectroscopy in Food Analysis

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Gas Chromatography/Infrared Spectroscopy • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. D.M. Haaland, E.V. Thomas, Paper K05, presented at the meeting of the Federation of Analytical Chemistry and Spectroscopy Societies, Boston, November, 1988.
2. K.R. Beebe, R.J. Pell, M.B. Seasholtz, in *Chemometrics A Practical Guide*, John Wiley & Sons, New York, 1998.
3. M.C. Davies, P.C. Williams (eds.), 'Near Infrared Spectroscopy: Future Waves', 7th Proceedings International Conference Near Infrared Spectroscopy, NIR Publications, Chichester, UK, 1996.
4. G.D. Batten (eds.), 'Leaping Ahead Infrared Spectroscopy', 6th Proceedings International Conference Near Infrared Spectroscopy, Near Infrared Spectroscopy Group, North Melbourne, Australia, 1995.
5. R.H. Wilsons, in *New Phys.-Chem. Technology Characterization Complex Food Systems*, ed. E. Dickinson, Blackie, Glasgow, UK, 177–195, 1995.
6. D.H. McQueen, R. Wilson, A. Kinnunen, *Trends Anal. Chem.*, **14**(10), 482–492 (1995).
7. S. Kawano, A. Hideyuki, 'Development of a Calibration Equation with Temperature Compensation for Determining the Brix Value in Intact Peaches', *J. Near Infrared Spectrosc.*, **3**, 211–218 (1998).
8. G. Osborne, 'Near Infrared Spectroscopic Studies of Starch and Water in Some Proccessed Cereal Foods', *J. Near Infrared Spectrosc.*, **4**, 195–200 (1998).
9. R. Mehrotra, S.K. Dhawan, A. Kumar, B.D. Malhotra, S. Chandra, J.J. Bhagat, 'Quantitative Analysis of Sugar Mixture by Infrared Spectroscopy: Application to Sugar Industries', NPL report, 25–32 (1995).
10. R. Mehrotra, J. Tewari, A. Gupta, S. Chandra, S.P. Varma, 'Analysis of Sugarcane Solids Using Diffuse Reflectance Spectroscopy', NPL/Department of Science and Technology, New Delhi report, 1999.
11. N. Dupuy, L. Duponchel, J.P. Huvenne, B. Sombret, P. Legrand, 'Classification of Edible Fats and Oils by Principal Component Analysis of Fourier Transform Infrared Spectra', *Food Chem.*, **57**, 245–251 (1996).
12. R.T. Sleeter, M.G. Matlock, 'Automated Quantitative Analysis of Isolated (Nonconjugated) *trans* Isomers Using Fourier Transform Infrared Spectroscopy Incorporating Improvements in the Procedure', *J. Am. Oil Chem. Soc.*, **66**, 121–127 (1989).
13. B. Dahlberg, S.M. Lee, S.J. Wenger, J.A. Vargo, 'Classification of Vegetable Oils by FT-IR', *Appl. Spectrosc.*, **51**(8), 1118–1123 (1997).

14. J. Wesley, R.J. Barnes, A.E. McGill, 'Measurement of Adulteration of Olive Oils by Near Infrared Spectroscopy', *J. Am. Oil Chem. Soc.*, **72**, 289–292 (1995).
15. F.R. Van de Voort, J. Sedman, G. Emo, A.A. Ismail, 'Rapid and Direct Iodine Value and Saponification Number Determination of Fats and Oils by Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy', *J. Am. Oil Chem.*, **69**, 1118–1123 (1992).
16. F.R. Van de Voort, 'A Rapid FTIR Quality Control Method for Fat and Moisture Determination in Butter', *Food Res.*, **25**, 193–198 (1992).
17. F.R. Van de Voort, K.P. Memon, J. Sedman, A.A. Ismail, 'Determination of Solid Fat Index by Fourier Transform Infrared Spectroscopy', *J. Am. Oil Chem. Soc.*, **73**, 411–416 (1996).
18. F.R. Van de Voort, A.A. Ismail, J. Sedman, J. Dubois, T. Nicodemo, 'The Determination of Peroxide Value by Fourier Transform Infrared Spectroscopy', *J. Am. Oil Chem. Soc.*, **71**, 921–926 (1994).
19. M.D. Gullien, N. Cabo, 'Infrared Spectroscopy in the Study of Oils and Fats', *J. Sci. Food Agric.*, **75**, 1–11 (1997).
20. K.M. Bewig, A.D. Clarke, C. Roberts, N. Unklesbay, 'Discriminant Analysis of Vegetable Oils by Near Infrared Reflectance Spectroscopy', *J. Am. Oil Chem. Soc.*, **71**, 195–200 (1997).
21. K. Hewavitharana, B.V. Brakel, 'Fourier Transform Infrared Spectroscopic Method for Rapid Determination of Casein in Raw Milk', *Analyst*, **122**, 701–704 (1997).
22. F. Ulberthand, M. Henninger, 'Quantitation of *trans* Fatty Acids in Milk Fat Using Spectroscopic and Chromatographic Methods', *J. Dairy Res.*, **61**, 517–527 (1994).
23. P. Robert, D. Bertrand, M.F. Devaux, 'Multivariate Analysis Applied to Near Infrared Spectrum of Milk', *Anal. Chem.*, **59**, 2187–2191 (1987).
24. R. Frankhuizen, 'NIR Analysis of Dairy Products', in *Handbook of Near Infrared Analysis*, eds. D.A. Burns, E.W. Ciurczak, Marcel Dekker, Vol. 13, 609–641, 1992.
25. R. Mehrotra, A. Gupta, J. Tewari, Y. Anand, S.P. Varma, 'FTIR/ATR Studies of Cheese', manuscript in preparation.
26. M.M. Pierce, L. Randy, R.L. Wehling, 'Comparison of Sample Handling and Data Treatment Methods for Determining Moisture and Fat in Cheddar Cheese by Near Infrared Spectroscopy', *J. Agric. Food Chem.*, **42**, 2830–2835 (1994).
27. T. Sato, S. Kawano, M. Iwamoto, 'Detection of Foreign Fat Adulteration of Milk Fat by Near Infrared Spectroscopy', *J. Dairy Sci.*, **73**, 3408–3413 (1990).
28. N. Pedretti, D. Bertrand, M. Semenou, P. Robert, R. Giangiacomo, 'Application of an Experimental Design to the Detection of Foreign Substances in Milk', *J. Near Infrared Spectrosc.*, **1**, 174–184 (1993).
29. P.A. Ponce, S. Garrigues, M. Guardia, 'Vapor Generation Fourier Transform Infrared Direct Determination of Ethanol in Alcoholic Beverages', *Analyst*, **121**, 923–928 (1996).
30. R. Mehrotra, 'Analysis of Alcohol in Wines and Beers Using ATR Technique', manuscript in preparation.
31. J.A. Perkins, E.J. Hasenoehrland, P.R. Griffiths, 'Expert System based on Principal Component Analysis for the Identification of Molecular Structures from Vapor Phase Infrared Spectra 1. Theory. Identification of Alcohols', *Anal. Chem.*, **63**, 1738–1747 (1991).
32. G. Cavinato, D. Mayes, Z. Ge, J.B. Gallis, 'Non Invasive Method for Monitoring Ethanol in Fermentation Processes Using Fiber Optic Near Infrared Spectroscopy', *Anal. Chem.*, **62**, 1977–1982 (1990).
33. R. Mackison, S.J. Brinkworth, R.M. Belchamber, R.E. Aries, D.J. Cutler, C. Deeley, H.M. Mould, 'A Demonstration of Truly Remote In-line Near Infrared Process Analysis', *Appl. Spectrosc.*, **46**(6), 4987 (1992).
34. J. Tewari, M. Joshi, A. Gupta, R. Mehrotra, S. Chandra, 'Determination of Sugars and Organic Acid Concentration in Apple Juices Using Infrared Spectroscopy', *J. Sci Ind. Res.*, **58**, 19–24 (1999).
35. G. Osborne, T. Fearn, S.G. Stevenson, in *Near-infrared Technology in the Agricultural and Food Industries*, eds. P. Williams, K. Norris, American Association of Cereal Chemists, St. Paul, MN, 185, 1987.
36. Y. Ootake, S. Kokot, 'Discrimination between Glutinous and Non Glutinous Rice by Vibrational Spectroscopy I: Comparison of FT-NIR, PAS and Raman Spectroscopy', *J. Near Infrared Spectrosc.*, **6**, 251–258 (1998).
37. H. Hsu, A. Mc Neil, E. Okine, G. Mathison, R. Siawash, 'Near Infrared Spectroscopy for Measuring In Situ Degradability in Barley Forages', *J. Near Infrared Spectrosc.*, **6**, 129–143 (1998).
38. J.A. Panford, in *Near Infrared Technology in the Agricultural and Food Industries*, eds. P. Williams, K. Norris, American Association of Cereal Chemists, St. Paul, MN, 201, 1987.
39. F.E. Barton, D.S. Himmelsbach, D.D. Archibald, 'Two Dimensional Vibrational Spectroscopy. V. Correlation of Mid and Near Infrared of Hard Red Winter and Spring Wheats', *J. Near Infrared Spectrosc.*, **4**, 139–152 (1996).
40. R.K. Cho, Y.K. Kwon, 'Identification of Rice Variety Using Near Infrared Spectroscopy', *J. Near Infrared Spectrosc.*, **6**, A67–A73 (1998).
41. H. Takamura, N. Endo, T. Matoba, 'Near Infrared Spectroscopic Determination of Moisture Content in Foods: Extraction Method by Organic Solvents', *J. Near Infrared Spectrosc.*, **6**, 235–240 (1998).
42. M.C. Davies, I.A. Cowe, R.P. Withey, G. Eddison, T. Fearn, 'Commodity Testing and Sub-sample Homogeneity System for Meatspec Analyzer', *J. Near Infrared Spectrosc.*, **6**, 69–75 (1998).
43. P.R. Griffiths, J.A. de Haseth, 'GC/FTIR Spectrometry', in *Fourier Transform Infrared Spectrometry*, eds. P.J. Elving, J.D. Winefordner, John Wiley & Sons, New York, 564–610, 1986.

44. R. Namba, 'Industrial Applications of GC/FTIR', in *Practical Fourier Transform Infrared Spectroscopy*, eds. J.R. Ferraro, K. Krishnan, Academic Press, New York, 469–518, 1990.
45. P. Magidman, J.M. Purrel, 'Analysis of the Aroma of Intact Fruit of Coffea Arabica by GC/FTIR', *Appl. Spectrosc.*, **38**(2), 181 (1984).
46. S. Compton, P. Stout, 'Headspace Gas Chromatography Fourier Transform Infrared Spectrometry for Monitoring Volatiles in Commercial Brand Coffee', Application notes 78, Bio-Rad Digilab Division, USA, 1990.
47. S. Bourne, S.L. Hill, J.R. Powell, D.R. Prasad, K. Krishnan, 'High Sensitivity Fourier Transform Infrared Spectroscopy of Gas Chromatographic Fractions Using a Liquid Nitrogen Cold-trapping Technique', Application notes 75, Bio-Rad Digilab Division, USA, 1990.
48. M.M. Mossoba, J.T. Chen, W.C. Brumley, S.N. Page, 'Two Dimensional Liquid Chromatographic Method for Resolution of Prostaglandin Enantiomers', *Anal. Chem.*, **60**, 948 (1948).

Lipid Analyses in Food

Domenico Marini
Direzione Centrale per l'Analisi Merceologica ed il Laboratorio, Chimico, Rome, Italy

1 Introduction **1**
1.1 Preparation of Samples 2
2 Thin-layer Chromatography **3**
2.1 Normal Phase Thin-layer Chromatography 3
2.2 Modified Normal Phase 3
2.3 Reversed Phase 3
3 Gas Chromatography **4**
3.1 Columns 4
3.2 Detectors 4
3.3 Quantification 4
4 High-performance Liquid Chromatography **5**
4.1 Detectors 5
4.2 Columns 5
4.3 Preliminary Resolution with Thin-layer Chromatography 6
4.4 Subsequent Characterization with Gas Chromatography 7
5 Mass Spectrometry/Mass Spectrometry **7**
5.1 Methods of Ionization 7
6 Nuclear Magnetic Resonance Spectroscopy **8**
6.1 Proton Spectroscopy 8
6.2 Spectroscopy with Carbon-13 8
6.3 Two-dimensional Spectroscopy 9
7 Analysis of Fatty Acids **9**
7.1 Gas Chromatographic Analysis 9
7.2 Argentation Thin Layer Chromatography 13
8 Analysis of Triacylglycerols **14**
8.1 Thin-layer and Conventional Column Chromatography 14
8.2 Separations with Column Chromatography 15
8.3 Thin-layer Chromatography with Flame Ionization Detection 16
8.4 Enzymatic Methods 16
8.5 Gas Chromatography 16
8.6 High-performance Liquid Chromatography 17
8.7 Stereospecific Analysis 23
Abbreviations and Acronyms **25**
Related Articles **25**
References **25**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The term lipid encompasses a great host of compounds (e.g. fatty acids (FAs), mono-, di- and triglycerides), that are almost ubiquitously present in foods; their characterization and quantification is then of importance for an analytical chemist.

The first step for an accurate analysis is sample preparation: choosing an appropriate solvent or phases can isolate one particular class of lipid from a sample [a good separation of classes can be accomplised by thin-layer chromatography (TLC)]; obviously, the sample pretreatment will be governed by the determinations to follow. Indeed, nowadays, there is a host of techniques, mainly chromatographic, for lipid analysis. TLC is a rapid way to gather information about the sample (presence of conjugated double and triple bonds, free FAs and of oxygenated FAs), particularly with the development of modified normal phases. For example, silica gel can be impregnated with boric acid, which inhibits isomerization of mono- and diglycerides, or it can be treated with silver ions [argentation, for silver ion thin-layer chromatography (Ag-TLC)] used for resolution of unsaturated FAs. Gas chromatography (GC) and high-performance liquid chromatography (HPLC) are the most common techniques actually employed, because of their versatility (many parameters can be changed, e.g. stationary and mobile phases, detector) and to their capacity to be hyphenated (coupled with another measuring device). Particularly, the advent of the light-scattering detector has allowed greater reproducibility in the HPLC of lipids. Mass spectrometry (MS) is often coupled with GC. However, coupling MS with another MS (MS/MS or tandem MS) is useful for the determination of the structure of long-chain FAs with improved selectivity and sensitivity. Finally, nuclear magnetic resonance (NMR) is able to identify functional groups, even in mixtures, and is a valid aid for the elucidation of structures with double bonds (DBs) by interpretation of the coupling patterns.

Each analytical method relies on appropriate sample preparation (e.g. methylation for the GC analysis of FAs) and a critical approach to evaluate its usefulness.

1 INTRODUCTION

In plants, fats are synthesized from carbohydrates by the plant itself, while animal lipids come from two different sources: exogenous lipids ingested in food and endogenous lipids synthesized by animal tissues. In

animals, the former are nearly all triacylglycerols (TAGs), whereas the latter include a high proportion of polar lipids, chiefly phospholipids.[1]

For numbering the three hydroxyl groups on the glycerol molecule, the convention originally proposed by Hirschmann[2] is universally adopted. An *sn*- (stereospecifically numbered) prefix is included in the names of all glycerols.[3] A single molecular species is identified by listing the *sn*-1, *sn*-2 and *sn*-3 positions in order. An rac prefix indicates that the middle FA in the abbreviation is attached at the *sn*-2 position, while the remaining two acids are equally divided between the *sn*-1 and *sn*-3 positions, yielding a racemic mixture of two enantiomers. A β prefix indicates that the middle FA in the abbreviation is esterified at the β- or *sn*-2 position. For example: *sn*-OLLn is *sn*-1-oleo-2-linoleo-3-linolenin; rac-OLLn is equal proportions of *sn*-OLLn and *sn*-LnOL; and β-OLLn is *sn*-OLLn plus *sn*-LnLO in any proportion.

In most species, saturated FAs tend to occupy mainly position *sn*-1, though substantial amounts of oleic acid can also be found at that position. Unsaturated FAs and certain medium-chain FAs (C_{14}) are mainly concentrated at position *sn*-2.

Long-chain FAs, such as stearic acid, are primarily located at position *sn*-1, and unsaturated FAs (e.g. oleic acid and linoleic acid) at position *sn*-3.

Very-long-chain polyunsaturated FAs occur mainly in the middle position in fish oils, with monoenoic acids concentrated in the outer positions.

The goal in the analysis of lipids is the complete resolution of the lipid classes and the molecular species for the identification and quantification of all components. Since no single analytical method is capable of identifying and quantifying all lipid species, analysis involves a multistep extraction, stepwise chromatographic resolution and sequential MS identification of lipid mixtures of progressively simpler compositions.[4]

TLC can be used to separate different categories of lipids according to their functional groups, while Ag-TLC (with $AgNO_3$) separates TAGs according to the degree of unsaturation. Reversed-phase thin-layer chromatography (RPTLC) sorts according to carbon number (CN). GC assesses the degree of unsaturation of FAs and CN, while GC/MS is used for peak identification and quantification. HPLC is perhaps the preeminent method for separating TAGs.

1.1 Preparation of Samples

Sample preparation for GC, HPLC and MS must include cleaning for chemical and particulate contaminants to provide the component of interest in a solution, free from interfering matrix elements and in appropriate concentration for detection and quantification. This usually can be accomplished by sample purification and preliminary separation by TLC or HPLC.

Lipid samples are normally obtained by extraction with organic solvents but, recently, adsorption cartridges (solid-phase extraction) have been introduced to improve recoveries of the more polar components.

Neutral solvent methods of total lipid extraction result in low recoveries of acidic phospholipids, lysophospholipids and nonesterified FAs, while use of alkaline solvents may result in deacylated products depending upon the temperature and the time of exposure.[3,5–7]

A lipid extract usually contains water and other nonlipid contaminants, which must be removed from the sample prior to derivatization, storage or chromatographic separation. A preliminary separation of lipid classes is also desirable. The small amounts of water in the solvent extract can be removed azeotropically during solvent evaporation, while small amounts of water with any water-soluble components can be removed by passing the wet organic solvent through a small column packed with anhydrous sodium sulfate.

In many cases, it is useful to effect a separation of the neutral and phospholipid fractions by some adsorption chromatographic technique using Fluorosil or silica gel as the adsorbent.[8] Then, the polar and nonpolar lipid classes may be separated using commercial adsorption cartridges. In fact, TAGs and cholesteryl esters can be easily separated from the glycerophospholipids and sphingomyelins by sequential elution with mixtures of chloroform and methanol while complex acidic lipids and neutral phospholipids and ceramides can be resolved with methanol–water and chloroform–methanol mixtures, respectively.[9]

A good separation of the lipid classes is obtained by adsorption TLC. Using a neutral solvent system for development, it is possible to retain the glycerophospholipids and sphingomyelins at the origin, while the neutral lipids are carried up the plate according to their R_F values.[6] When using a polar solvent system, the neutral lipids are carried along together with the solvent front, while the phospholipids are resolved according to their polarity.

The strategy chosen for the analysis of specific lipids depends on the nature of the information needed and can frequently be achieved by adopting some established analytical routine. For example, FA analysis may require little more than a preparation of FA methyl esters from the total sample and a GC analysis. A more difficult problem may call for the analysis of the molecular species in a glycerolipid, which could require determination of the composition, molecular association and positional distribution of the FAs in a glycerolipid molecule.

2 THIN-LAYER CHROMATOGRAPHY

TLC is the simplest and still most widely employed technique in lipid analyses because it provides rapid and complete separations of most neutral and phospholipid classes.

Preliminary information on lipid composition can be easily obtained using TLC on silica gel layers with a solvent system such as hexane–diethyl ether–acetic acid (80:20:2 [by volume (v)]).[10] Detection of all of the lipid classes can be made first under ultraviolet (UV) light (254 nm) or after spraying with the 1′,2′-dichlorofluorescein and examining under UV light. Spraying with sulfuric acid–methanol (50:50) and subsequent heating gives a permanent record.

Generally, seven spot sections can be observed:

1. alkanes/alkenes/squalene/sterol ester/wax ester (R_f 98–94)
2. triglycerides (R_f 60; predominant)
3. free FAs (R_f 39)
4. sterols (R_f 19)
5. diglycerides (R_f 15–21)
6. monoglycerides (R_f 2)
7. phospholipids (R_f 0).

The TLC analysis is useful for a preliminary detection of the presence of compounds with conjugated double/triple bonds (detection under UV light), the presence of free FAs (R_f value comparison), and the presence of oxygenated FAs, unusual lipid classes or other unusual compounds.

Particular spray reagents permit the selective detection of keto, epoxy, hydroxy and cyclopropene groups.[11] Even if the use of instrumental analytical methods is preferred, these TLC techniques may provide more complete information and are not hazardous.

If unusual TLC spots are detected, the lipid can be hydrolyzed, and the resulting nonsaponifiable matter and free-FA fractions can be separated and further analyzed.

2.1 Normal Phase Thin-layer Chromatography

A large variety of solvent systems and qualitative visualization reagents are available for normal-phase TLC, but usually the first approach is the TLC of a lipid sample in two or more solvent systems to examine its uniformity and fraction identity.[6,12,13]

The TLC separation can be increased further by the introduction of finer gel grades. When these are used in combination with smaller plates and thinner layers of adsorbent the technique is known as high-performance thin-layer chromatography (HPTLC).[14] With this method it is possible to separate mixtures of nonpolar lipids, glycosphingolipids and phospholipids in one dimension. More than 20 different lipid subclasses are separated with three to four consecutive developing solvents. (This method was applied successfully to the analysis of glycosphingolipids from erythrocytes and other biological membranes in a simple identification of sphingolipidosis in patients.)

New strategies arise from the development of the TLC plate in two directions at a 90° angle with two different solvent systems.[15] The solvent systems are chosen to enhance different properties of the lipid molecules and the plate is dried between developments. The development in each direction can be carried out with one or more solvent systems and chromatographic conditions, thus giving rise to a large number of variations. The second TLC separation can also be performed following isolation of the components resolved in the first TLC separation.

2.2 Modified Normal Phase

TLC may be used after impregnation of the silica gel with special additives, e.g. boric acid or $AgNO_3$.

2.2.1 Modification with Boric Acid

Silica gel impregnated with boric acid is used to prevent or minimize the isomerization of mono- and diacylglycerols during chromatographic separation. The borate is incorporated into the silica gel at 5–10% and the plates are spread, dried, activated and developed in the usual manner. The boric acid forms complexes with vicinal hydroxyl groups and thus blocks the transfer of fatty acyl chains from one position to another.[16]

2.2.2 Argentation

Argentation TLC (Ag-TLC) is extensively utilized for analytical and semipreparative resolution of methyl and glyceryl esters of saturated and unsaturated FAs.[17,18] The incorporation of 5–20% $AgNO_3$ into the silica gel permits effective complexing of the π DBs with silver ions, which retard the migration of the polyunsaturated FAs.

The separation is based essentially on the number of DBs per molecule and separate bands are obtained for compounds with 0–6 DBs per molecule.

2.3 Reversed Phase

RPTLC systems have been developed using RP-18, RP-8 or CN plates.[19] Resolution of fatty acid methyl esters (FAMEs) by TLC often is complicated by co-migration of certain acyl isomers in heterogeneous mixtures: An RPTLC method that employs 10% [weight/volume (w/v)] $AgNO_3$ in a mobile phase containing acetonitrile–1,4-dioxane–acetic acid (80:20:1 v/v)

allows one-dimensional resolution of a wide range of acyl methyl esters.

This innovation enables improved separation of saturated FAME ranging from C_{12} through C_{22} from unsaturated FAME by TLC.[20]

3 GAS CHROMATOGRAPHY

GC can be considered one of the most valuable and most efficient methods of resolution and quantification of lipids. The development of flexible fused quartz capillary columns has made capillary GC the method of choice and is gradually phasing out packed GC columns from research laboratories. Except for electron capture, the flame ionization detector (FID) has replaced all others for lipid GC.

Quantification in GC (and also in HPLC) is based on the principle that the masses of each of the separate constituents in the mixture are proportional to the area under each peak.

The retention time is reflected by the distance on the chromatogram between the air peak and the position of the band peak. Working with an H_2 flame detector, the air outflow is not detectable; in that case elution of the solvent can be taken as a starting point.

3.1 Columns

Packed or capillary columns may be employed. Packed columns range from 1.5 to 3 m in length, with an internal diameter of 2–4 mm and diatomaceous earth or other support and a narrow range of particle sizes (between 125 and 200 μm). The stationary phase is normally a polyester-type polar liquid (e.g. diethylene glycol succinate) or others such as cyanosilicone, methyl silicone, etc.[21–23] The working temperature ranges from 175 to 200 °C, depending upon the column type.

Capillary columns may be made of glass or fused silica gel. Internal diameter ranges from 0.2 to 0.8 mm, and column length from 25 to 100 m depending upon the type of acid to be separated. The preferred stationary phase is of polyethylene glycol chemically bonded. The bonded phase thickness should be from 0.1 to 0.2 μm.

3.2 Detectors

The quantification of the chromatograms is essentially a function of the detector and is easy with detectors that give a reproducible mass or molar response. It is also helpful to have a sensitive detector response that is linear over the working range for all analytes in the sample.

Since natural glycerolipids usually do not contain chromophores, they are best detected by the universal detection systems such as FID and total ion current in MS, or light scattering for HPLC.

The FID responds essentially to the carbon mass in the solute. It is sensitive and possesses a wide dynamic range. It is not affected by temperature programming in GC or solvent gradients in HPLC and requires minimal calibration except for components of very-high molecular mass in GC.

3.3 Quantification

For quantification, the peak area of each component is reported as a percentage of the total area of all peaks. These results reflect the relative composition of a multicomponent sample, provided all the components are eluted, separately detected and give the same detector response per weight or mole unit. When the detector response is not equivalent, the area percentage of a component is multiplied by a response factor, which must be determined, to reflect the weight or mole percentage of the component reaching the detector.[24,25]

3.3.1 Use of an Internal Standard

Use of an internal standard is the most satisfactory method for the quantification of neutral lipids and FAME.[26] A sample of an appropriate standard (e.g. C_{17} or C_{23} methyl esters for FAs and C_{30} TAG for neutral lipids) equivalent to 10–20% of total peak area is added to the sample before introduction into the chromatograph. After chromatography, the peak areas for the analyte and the internal standard are measured and peak area ratios are determined. The amount of each unknown is calculated in relation to the known concentration of the standard. Using an internal standard, the errors arising at each step are self-compensating, because both standard and unknown are affected proportionally.

3.3.2 Use of External Standard

In external standardization, response curves are established for each unknown by injecting into the chromatograph progressively larger amounts of reference compounds and recording the quantitative response.[27,28] The peak area recorded for the unknown is then compared with the corresponding point on the standard curve and the concentration of the unknown read off from the graph. The standards are prepared in the same solvent as the unknowns and the standard response must be confirmed before and after the analysis of the unknowns.

4 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

Neutral lipid classes and molecular species can be successfully separated by HPLC, even if many problems are associated with the detection of the nonchromogenic lipids in the column effluent. These problems have been largely overcome by the use of the universal light scattering,[29] hydrogen FID[30] and MS[31] detectors.

4.1 Detectors

The development of the light scattering detector[32] has revolutionized HPLC of lipids. Detection is achieved by nebulizing the column effluent in a stream of tepid (30–35 °C) gas. Although the response is nonlinear and the dynamic range much more limited than that of the FID the light or laser light scattering detector responds to all glycerolipids in a reproducible manner.

Both UV absorbing and fluorescent derivatives yield essentially molar responses, which lend readily to quantification of all molecular species over a wide concentration range.[33,34]

Furthermore, wavelength ratios and absorption spectra provide useful, although not definitive, information about the identity or purity of the peak.

The refractive index and short-wave UV detection systems employed in early HPLC of glycerolipids are best suited for qualitative work, although important quantitative applications have been reported.[35] These detectors require extensive calibration and are subject to serious solvent restrictions.

HPLC with a refractive index detector can achieve quantitative analysis of synthetic mixtures of saturated TAGs (with FAs from caprylic to stearic), using an eluent of propionitrile and butyronitrile (80 : 20 v/v).[36]

4.2 Columns

4.2.1 Normal Phase

Normal-phase HPLC is used mainly for lipid class separations, since molecular species have long equilibration times and show poor resolution. Thus, TAGs and mono- and diacylglycerols are readily resolved from each other and from cholesteryl esters and glycerophospholipids.[29] Like the TAGs, the long-chain diracylglycerols can be readily recovered as a lipid class from normal-phase HPLC, with very limited fractionation. The corresponding X-1,3- and X-1,2-diracylglycerols are resolved, with the X-1,3-isomers emerging ahead of the X-1,2-isomers,[37] again as established for normal-phase TLC. Normal-phase HPLC also allows the separation of TAGs containing common and hydroxy FAs.[38]

Furthermore, normal-phase HPLC is well suited for the preparative resolution of the diacyl-, alkylacyl- and alkenylacylglycerols as the acetates, trimethylsilyl and terbutyldimethylsilyl ethers or benzoyl, dinitrobenzoyl, naphthyl, anthroyl and pyrenyl derivatives.[33,34,39,40] The elutions are performed with solvents of low polarity and the elution order of the three (or any two) is the same: alkenylacyl > alkylacyl > diacyl.

1,3- and 1,2-Diacylglycerols can be resolved by normal-phase HPLC,[41] but 1,3-isomers of low molecular weight interfere with 1,2-isomers of high molecular weight. The 1(3)-isomers were eluted ahead of the 2-isomers of monoacylglycerols when run in the underivatized form. Normal-phase HPLC has also been used for the isolation of mono- and digalactodiacylglycerols.

4.2.2 Modified Normal Phase

Normal-phase HPLC separations may be modified by inclusion of various modifiers in the adsorbent phase. HPLC on a silver ion loaded silica gel column (5 μm) gives resolution of TAGs based on degree of unsaturation.[42] The separations obtained for vegetable oils are similar to those realized by Ag-TLC.

HPLC on $AgNO_3$-treated silica gel can reverse separations of isomers, as TAGs with the unsaturated FA in the *sn*-2 position are eluted ahead of the corresponding isomer with the unsaturated FA in the *sn*-1 or *sn*-3 positions.

Improved resolution of both unsaturates and regioisomers is achieved with a 3 μm column.[43] More durable $AgNO_3$–HPLC columns are produced by loading the silver ions on sulfonic acid anion-exchange columns (Nucleosil SSA).[44] Although these columns do not permit resolution of positional isomers, they have been employed for effective resolution of fish oil TAGs.[45] The fractions, however, were still extremely complex and required further fractionation by complementary chromatographic techniques, e.g. reversed-phase high-performance liquid chromatography (RPHPLC).

The seven geometric (cis/trans) isomers of linolenic (18 : 3) acid are resolved by $AgNO_3$–HPLC.[46] For this purpose the temperature must be lowered to 10 °C and the column is eluted with a solvent gradient of dichloromethane and methanol.

Adsorbent cartridges containing cyanopropyl-bonded silica gel can be employed for the separation of neutral lipid classes with acetone modifying the base solvent (hexane–tetrahydrofuran (98 : 2 v/v)).[47]

The separation of monoacylglycerols as the didinitrophenylurethane (DNPU) derivatives by normal-phase HPLC is possible on nitrile-bonded silica using *n*-hexane–1,2-dichloroethane–ethanol (40 : 10 : 1 v/v/v).[48] The peaks emerge in the following order: 20 : 0, 18 : 0,

16:0, 18:1, 16:1, 18:2 and 18:3. The peaks of 1- and 2-monoacylglycerols with the same acyl group show complete overlapping.

The major soybean phospholipids can be separated on β-cyclodextrin-bonded silica using isocratic elution with hexane, isopropanol, ethanol or water–tetramethylammonium phosphate as the mobile phase, and UV detection.[49] The presence of tetramethylammonium phosphate in the mobile phases is critical for good resolution.

4.2.3 Reversed Phase

RPHPLC with short wavelength UV detection (210 nm) can be used for the analysis of TAGs containing highly unsaturated FAs, including α- and γ-linolenic acids.[50,51] This, however, requires increased detector sensitivity, which is achieved by employing a thermostat to stabilize the temperature of the entire HPLC system.

A typical reversed-phase column consists of 5–10 μm alkane-bonded silica gel (C_{18}, C_8, etc.). Solvent temperature programming has been shown to provide advantages for resolution and peak recovery as well as for expanding the range of solvent types or combinations available for separation of those compounds that are difficult to dissolve in the semi-aqueous solvents.[52]

Critical pairs, which result from the approximate equivalence of one DB to a chain shortened by two methylene units, are effectively separated with acetone–acetonitrile (63.6:36.4 v/v) as an isocratic mobile phase. Other solvent systems are also capable of resolving critical pairs and triplets of TAGs. An acetone–acetonitrile system has been adapted to the HPLC resolution of butterfat TAGs, but the greatest improvement results from simply decreasing the sample load.

Excellent resolution of fish oil TAGs on RPHPLC is obtained using a gradient of 10–90% isopropanol in acetonitrile.[53]

The mobile phase for RPHPLC of polar lipids consists of an aqueous component and an organic modifier (5–40% for most applications), with a third component occasionally added as an ion-pairing reagent. The molecular species of intact glycerophospholipids are resolved using methanol–water–acetonitrile along with choline chloride or phosphate salts.[54]

4.2.4 Chiral Phase

Although enantiomeric TAGs cannot be resolved, their random deacylation products can be separated by chiral-phase HPLC following conversion into DNPU derivatives,[48,55] while separations of diracylglycerols can be carried out on *N*-(*S*)-2(4-chlorophenyl)isovaleroyl-D-phenylglycine chemically bonded to γ-aminopropyl silanized silica.[56]

The *sn*-1,2-enantiomers are eluted well ahead of the *sn*-2,3-enantiomers of the monoacid diacylglycerols. Improved separation of enantiomeric diacylglycerols can be obtained with *N*-(*R*)-1-(1-naphthyl)ethylaminocarbonyl-(*S*)-valine liquid phase, which could be employed for the resolution of mixtures of homologous diacylglycerols.[57,58]

Effective resolution of the enantiomers of unsaturated diacylglycerols derived from natural fats and oils by Grignard degradation can be provided by the (*R*)-1-(1-naphthyl)ethylamine liquid phase. The molecular species of the resolved diacylglycerols are then identified by on-line MS.[59,60]

Excellent resolution of the enantiomeric monoacylglycerols and monoalkylglycerols can be also obtained using amide-urea type chiral phases.[61]

A new chemical method suitable for the stereospecific analysis of marine biogenic lipid triacyl-*sn*-glycerols was reported by Ando et al.[62] Triacyl-*sn*-glycerols are partially hydrolized with ethyl magnesium bromide and 1-monoacylglycerol products are resolved into *sn*-1 and *sn*-3 monoacylglycerol fractions by HPLC on a chiral column (50 cm × 4 mm internal diameter), packed with 5 μm particles of *N*-(*R*)-1-(α-naphtyl)ethyl-aminocarbonyl-(*S*)-valine chemically bonded to γ-aminopropyl silanized silica (Sumichiral OA-4100, Sumitomo Chemical Co., Osaka), using a mixture of hexane–1,2-dichloroethane–ethanol (40:12:3 v/v/v) as mobile phase at a constant flow rate of 0.5 ml min^{-1}.

4.3 Preliminary Resolution with Thin-layer Chromatography

Although lipids can be resolved into individual classes and molecular species by HPLC alone when the mixture contains total neutral lipids, it is frequently advantageous to effect a preliminary resolution of lipid classes by TLC to avoid potential overlaps among different homologous series. Thus, the individual phospholipid classes resolved by normal-phase TLC during the initial isolation and purification can be subjected to RPHPLC either in the intact form[63] or following dephosphorylation and conversion into UV-absorbing derivatives.[33,41]

Combining Ag-TLC with RPHPLC, it is possible to increase fractionation of short-chain TAGs from butterfat.[64] Using Ag-TLC with DNPU derivatives of 1,2(2,3)-diacylglycerols combined with chiral-phase HPLC permits the separation of the molecular species of diacylglycerols derived by Grignard degradation. The *sn*-1,2(2,3)-diacylglycerols were resolved from X-1,3-diacylglycerols by borate TLC.[65]

4.4 Subsequent Characterization with Gas Chromatography

HPLC/GC can be also employed for identification of the components resolved by normal, silver ion, reversed-phase or chiral-phase HPLC. Normal-phase HPLC may be employed instead of TLC for the isolation of neutral and polar lipid classes for subsequent GC examination on polar or nonpolar columns following appropriate derivatization.[13]

The diacyl-, alkylacyl- and alkenylacylglycerol subclasses have been completely resolved by combining normal-phase HPLC with polar capillary GC of the trimethylsilyl ethers.[66] Both reversed-phase[67] and silver ion[68,69] HPLC can provide small groups of molecular species of TAGs for improved subsequent GC analysis as intact molecules on polarizable capillary columns. In more conventional approaches, fractions from silver ion HPLC are analyzed by GC for FA composition.[69]

Chiral-phase HPLC provides effective resolution of the enantiomeric mono- and diacylglycerols but fails to resolve completely the component molecular species, which must be identified by other methods, including GC.

5 MASS SPECTROMETRY/MASS SPECTROMETRY

MS/MS is used for the structural analysis of long-chain carboxylic acids. It provides additional information and can also improve selectivity and sensitivity. It is achieved by arranging two or more mass spectrometers in a series. Specificity and sensitivity is further improved by coupling MS/MS to capillary GC and HPLC.

MS/MS is MS of a specific ion selected from the initial mass spectrum. The selected parent ion is allowed to enter a field-free reaction region, where it may undergo unimolecular or collision-induced dissociation (CID). The resulting fragments (daughter ions) are analyzed by a second mass spectrometer, which provides the MS/MS spectrum (MS/CID/MS). Generally the intermediate field-free region is used as a collision chamber where the internal energy of the parent ion is increased through collisions with some target gas. This increased internal energy induces the ion to fragment into daughter ions, which are then analyzed in the second mass analyzer. The daughter ions are characteristic of the structure of the parent ion and the MS/MS allows the determination of that structure.[70–73]

The maximum amount of the original kinetic energy of the parent ion that can be transformed into the internal energy needed for CID is regulated by the need to conserve the total momentum (Equation 1):

$$E_{\max} = \frac{M_2}{M_1 + M_2} E_{\text{ion}} \tag{1}$$

where $E_{\max}$ is the maximum amount of internal energy acquired in the collision; E_{ion} is the original kinetic energy of the parent ion; M_2 is the mass of the target atom or molecule; and M_1 is the mass of the parent ion.

Although the first mass spectrometer separates the mixture from the gas or the liquid chromatograph and the second determines the mass spectra of the molecular ion, the method differs from both GC/MS and HPLC/MS, which are based on chemical separation. In MS/MS, the initial separation is based on physical properties. This approach allows assessment of neutral fragment loss and parent ion scans, which have no counterpart in GC/MS or HPLC/MS. In addition, in MS/MS, the first mass spectrometer serves to filter out the background and improves the signal-to-noise ratio.

5.1 Methods of Ionization

The soft ionization MS techniques, such as fast atom bombardment, field desorption, thermospray and electrospray, have the ability to ionize lipid molecules without causing extensive fragmentation. This generally leads to mass spectra characterized by molecular adduct ions. Ionspray, which can be used with flow rates compatible with HPLC, is similar to electrospray except that the initial nebulization process is assisted by pneumatic or ultrasonic means. Electrospray favours multiply charged ions for proteins, but FAs and their esters usually acquire only one charge per molecule.

The desorption–ionization methods have been most often employed in the past to generate the parent ions of large polar organic molecules. More recently, thermospray and electrospray have also been used for this purpose because they may be coupled to HPLC by appropriate interfacing. The most common method used to obtain fragmentation of the molecular ionic species produced by a ''soft ionization'' process is CID.

Electrospray has a very unique property that distinguishes it from other soft ionization techniques. It can be used to produce CID spectra of singly charged species with greater sensitivity than can be achieved with triple quadrupole mass spectrometer (QqQ) electrospray MS/MS systems.

When analytes are present in the sprayed solution, molecular adducts, typically protonated parent ions $[M + H]^+$, are formed and guided into the reaction region of the mass spectrometer between the exit of

the glass capillary and the first skimmer. If a low voltage of 50–120 V is applied to the capillary exit, the molecular ionic species remain intact and the molecular mass of the analyte is obtained. However, if higher voltages are applied to the capillary exit, extensive and reproducible fragmentation of the molecular adduct ion can be achieved.

The best results in terms of sensitivity, specificity and reproducibility are obtained by using HPLC as a separation device combined with the electrospray CID/MS method instead of relying on the selection of specific masses from mixtures and using the first Q of a QqQ system for the separation.

6 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

In the past, NMR spectroscopy was mainly a tool for the analysis of pure compounds. As a consequence of the now widely available high-resolution spectrometers, the improvement of computer power and the introduction of new pulse sequences, the technique can also be applied with great success to lipids.

6.1 Proton Spectroscopy

The detection of a number of functional groups in seed oils, such as allenes, conjugated DBs, conjugated double–triple bonds, cyclopropane and cyclopropene rings, cyclopentene, epoxy, hydroxy, and terminal vinyl groups, is possible by their characteristic ^{1}H NMR signals.[74–81] The methylene groups adjacent to a keto group show shift values similar to those adjacent to the carboxyl group when located remote from the carboxyl or other functional groups. However, in some special cases, the presence of a keto group can be indicated by its ^{1}H NMR spectrum as shown, e.g. for 4-keto-α-eleostearic acid.

By an appropriate interpretation of the coupling patterns and consideration of the integral values of the signals, the configuration of the DBs in conjugated systems can be elucidated even in mixtures.[82] Furthermore, it is possible to distinguish FAs that contain a DB at the n-3 terminal from other unsaturated FA isomers by means of their terminal methyl signals.[83,84] For example, in the case of a seed oil that contained two n-3 acids (linolenic and α-parinaric acid (18 : 4) (9*c*, 11*t*, 13*t*, 15*c*)), even the separation of their terminal methyl and allylic methylene ^{1}H NMR signals (Figure 1) could be observed in the oil mixture.[85]

By integration of distinct ^{1}H NMR signals, a quantification of unusual FAs in lipid mixtures can be effected in an easy and nondestructive way, as was shown for cyclopropene,[79] n-3 FAs,[83] and epoxy FAs.[86]

6.2 Spectroscopy with Carbon-13

Because of the naturally low content of the ^{13}C isotope ($^{12}C/^{13}C$ NMR = 98.9/1.1), ^{13}C NMR spectroscopy is an

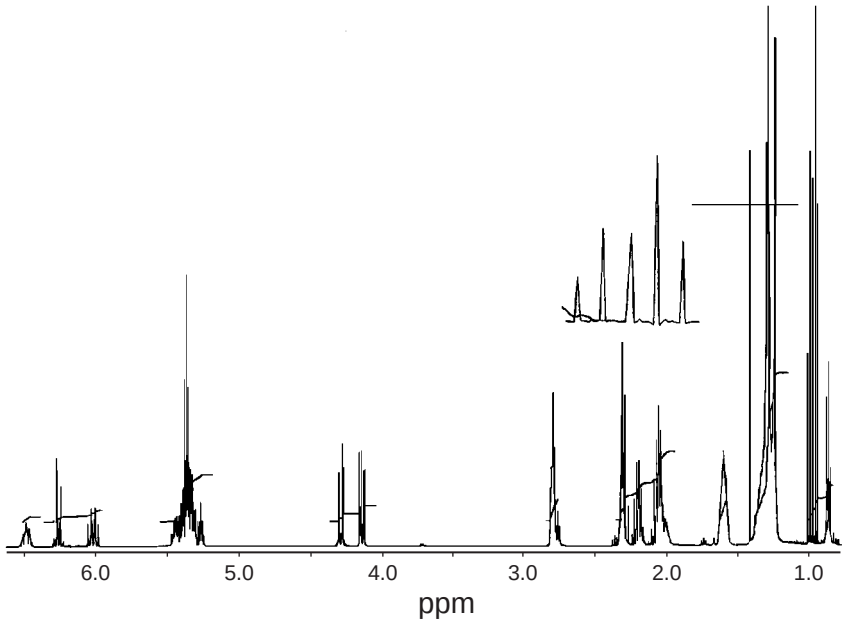

Figure 1 Proton NMR spectrum of a Euphorbiaceae seed oil (500 MHz). Signal groups (ppm) at chemical shift δ 0.88 (CH_3, all FAs); 0.968 (CH_3, linolenic acid); 1.0 (CH_3, parinaric acid); 1.28 (bulk CH_2): 1.58 (CH_2-CH_2-COOR, all FA); 2.05 (allylic CH_2); 2.2 (allylic CH_2, parinaric acid); 2.3 ($-CH_2-COOR$); 2.8 (double allylic CH_2); 4.1–4.3 (glycerol CH_2); 5.25 (glycerol CH); 5.35 (alkenic CH); 5.43–6.5 (conjugated alkenic CH, parinaric acid). (Reproduced by permission of AOCS Press from Spitzer et al.[85])

insensitive method for routine analysis. The detection limit for each FA is 2–5%, depending on its structure and measuring parameters.

When a spectrum is recorded routinely in the broad-band decoupled mode, the signals of the functional groups of FA are weak in comparison to those of the methyl, methylene and methine signals (nuclear Overhauser effect) and may be overlooked.

The chemical shift and the intensity of the ^{13}C NMR signals provide structural information on the individual FAs and lipid classes, and even in some cases the distribution of the different acyl groups on the glycerol backbone and the position of the unsaturated centers and functional groups in the acyl group.[87]

6.3 Two-dimensional Spectroscopy

The measurement of two-dimensional NMR spectra can provide useful structural information. The $^{1}H-^{1}H$ homonuclear and $^{1}H-^{13}C$ heteronuclear correlation NMR spectroscopy techniques enable the identification of the configuration of the conjugated triene system of α-oleostearic acid in cherry oil[82] and the configuration and localization in the FA chain of the conjugated tetraene system of α-parinaric acid in a Euphorbiaceae seed oil.[85]

With high-field ^{1}H NMR (600 MHz), it is possible to effect the simultaneous detection of minor compounds of olive oil, such as aldehydes, diacylglycerols, linolenic acid, water, acetic acid, phenols and sterols by using two-dimensional totally correlated spectroscopy and nuclear Overhauser effect spectroscopy.[88]

7 ANALYSIS OF FATTY ACIDS

7.1 Gas Chromatographic Analysis

FAs are commonly analyzed by GC of their FAMEs, although sometimes it is also possible to effect the GC separation of underivatized FAs.[89] The total gas chromatographic analysis of FAMEs has several steps: FA esterification and the injection, separation, identification and quantification of the derived FAMEs.

7.1.1 Preparation of Fatty Acid Methyl Esters

To be analyzed by GC, the FAs of lipids have to be split off and converted into derivatives with lower boiling points, such as esters of alcohols. The FAs are removed from the glycerol, sterol or alcohol by saponification with sodium hydroxide or potassium hydroxide, then methylated with either boron trifluoride–methanol reagent[90] or with methanolic mineral acids.[91]

Saponification–esterification procedure requires very long reaction times. Direct transesterification of lipids (''alcoholysis'') proceeds more rapidly than saponification, with hydrolysis and esterification taking place in one step, and requiring only one reagent.

Some problems are associated with ester preparation by transesterification, for example:

- incomplete conversion of lipids into FAMEs
- changes in the original FA composition during transesterification
- formation of artefacts that can be wrongly identified as FAs or overlap with FAME peaks in the GC analysis
- contamination of the GC column with traces of the esterification reagent
- incomplete extraction of FAMEs after transesterification
- losses of highly volatile short-chain FAMEs.

The reagents used for methylation are based mainly on methanol. Because nonpolar lipids (triglycerides, cholesterol esters) are not soluble in the transesterification reagents and cannot react in a short time, solvents such as benzene, toluene, chloroform or tetrahydrofuran have to be added to dissolve these lipids.[92,93]

Normally, after transesterification, FAMEs have to be extracted with alkanes such as *n*-pentane, *n*-hexane or isooctane. Water or salt solutions are added to increase recovery of FAMEs.

Because methyl esters of short-chain FAs, such as those present in dairy products, are highly volatile and are partially soluble in water, quantitative recovery is very difficult. For samples that contain short-chain FAs, the preparation of butyl esters,[94] which have higher boiling points and are less water soluble than methyl esters, is recommended.[95]

7.1.2 Acid-catalyzed Transesterification

Common reagents used for acid-catalyzed transesterification are methanolic hydrochloric acid, methanolic sulfuric acid, and boron trifluoride in methanol. All are suitable for lipid transesterification and also for free FA methylation.

However, neither reactions catalyzed by acid nor those using acid boron-trifluoride proceed at ambient temperature; both types of reaction require heating. Boron trifluoride–methanol reagent (12–14% w/v) is most often used for transesterification of all types of lipid.[96] The reaction at 100 °C requires 30 min for phosphoglycerides and triacylglycerides and 90 min for sphingomyelin. (A small sample of the lipid extract (dissolved in chloroform) is placed in a screw-capped tube (Teflon cap liner); 0.5 to 1 ml boron trifluoride–methanol

reagent ($140\,g\,l^{-1}$, containing an adequate amount of butylated hydroxytoluene as antioxidant) is added; the tube is then closed and heated at 90 °C for 2 h.) After the transesterification has been completed, FAMEs are extracted twice by adding *n*-hexane and water.

Anhydrous methanol containing hydrochloric or sulfuric acids has been used for lipid esterification under very different conditions. Complete transesterification with 5% methanolic hydrochloric acid can be carried out by heating the sample in the reagent under reflux for about 2 h. The reaction can also be carried out at 50 °C overnight. A solution of 1–2% (v/v) concentrated sulfuric acid in methanol can be used in the same way for the transesterification of lipid samples.

7.1.3 Base-catalyzed Transesterification

Reagents usually used for base-catalyzed transesterification of lipids are sodium methoxide and potassium hydroxide in methanol.

Transesterification with sodium methoxide proceeds at ambient temperature very rapidly,[95] and the risk of decomposition of polyunsaturated FA is lower. However, it does not esterify free FAs and it does not transesterify sphingolipids. It is possible to use a combined procedure for the conversion of free and glyceride-bound FAs. In the first step, sodium methoxide is used for the transesterification of glycerol-bound FAs, and in the second step methanolic hydrochloric acid is used for the esterification of any free FAs.[97]

Another reagent used for the transesterification of lipids is potassium hydroxide in methanol, but this has the disadvantage of a potential risk of saponification during transesterification.

7.1.4 Sample Injection

In high-resolution GC, sample injection is the most critical step with respect to achieving high accuracy and precision.

7.1.4.1 Split Injection With the split injection, the sample is introduced into a hot injector chamber or injection port. Splitting of the carrier gas (normally in a range between 1 : 10 and 1 : 200) has the advantage that samples with relatively large FAME concentrations can be injected without the risk of overloading the stationary phase, causing degradation of column performance. Injector temperatures for FAME injection are 250–300 °C or even below 250 °C.

The main problem of the classical split injection technique in FAME analysis is discrimination between high-boiling and low-boiling compounds in the sample, owing to changes in the splitting ratio caused by pressure waves, selective evaporation in the syringe needle and incomplete evaporation.[98] Sample discrimination can be avoided by preventing sample discrimination in the syringe needle, vaporizing the sample rapidly and mixing the vaporized sample thoroughly with the carrier gas. This can be achieved by high-speed injection, small sample volumes, low sample concentrations, high injector temperatures (375 °C), good injector insert designs and frequent cleaning of the insert.

7.1.4.2 On-column Injection The problems associated with split injection can be overcome by the use of cold sample injection: on-column injection or programmed-temperature vaporizing injection.

In the on-column technique, the sample is directly introduced into the cool section of the column and no prior vaporization in a hot injector chamber is required. As no sample splitting is performed and as no evaporation of the sample occurs in the syringe needle, sample discrimination can be completely avoided. Consequently, samples with a wide range of component volatilities can be determined with high accuracy and precision.[99,100]

In programmed-temperature vaporization, the sample is introduced into a cold vaporizing chamber that is packed with deactivated glass or quartz wool for homogeneous mixing of sample and carrier gas. After withdrawal of the syringe needle and after evaporation of the solvent, the vaporizing chamber is very quickly heated to 250–350 °C for evaporation of the higher-boiling compounds. This technique can be used in three different operation modes: cold split injection, cold splitless injection and solvent elimination injection.

Splitless injection, in general, is useful for the analysis of very dilute solutions while the solvent elimination technique allows the injection of large sample volumes (up to 250 µl). This technique is critical for the analysis of FAs containing less than 16 carbon atoms because these FAs can be lost together with the solvent through the split exit.[101,102]

7.1.5 Separation of Fatty Acid Methyl Esters

The separation of FAMEs can be carried out routinely by means of high-resolution modern capillary columns (with a length of 50 m and more and 10^5 or more theoretical plates). These are suitable for a wide range of chain lengths (CLs). Packed columns are still acceptable for the analysis of simple FAME mixtures such as vegetable oils, which contain no FA with more than 18 carbon atoms.

7.1.5.1 Stationary phases Separation of FAMEs is carried out with three different types of capillary column: nonpolar, polar and very polar. The polarity of the stationary phase influences retention times of FAMEs, especially those of polyunsaturated types. The column should be chosen according to the sample to be analyzed. In general, the resolution capability for

FAMEs, especially for unsaturated types, is highest in columns with very polar phases.

Complete analyses of all common FAs may be obtained with FAME using polar capillary columns and FID.

The polar capillary columns (consisting of bis-cyanopropylsiloxane) also yield effective separation of all CL and double-bond number homologs, including most positional double-bond isomers; however, only certain cis and trans positional isomers are resolved. Other cis and trans isomers and methyl and hydroxyl substitution isomers are not resolved and longer columns or special liquid phases (e.g. 60–100 m SP-2340 or SP-2560) have been chosen for their separation.[103]

In polar capillary GC of the TAG esters of long-chain FAs, the choice of the column length is critical because the recoveries decrease greatly with increasing column length, although the resolution increases.[104] The longer columns, however, require higher operating temperatures (e.g. 280–300 °C).

The most important very polar stationary phases are composed of 100% cyanoethylsilicone oil (SP-2340, OV-275), 100% cyanopropylsilicone (CP™-Sil 88) or biscyanopropyl–dimethylsiloxane (SP-2330) (68 : 32 v/v).

The main advantage of polar phases compared with nonpolar phases is their high resolution of unsaturated FAMEs. If a critical pair of FAMEs, such as 20 : 3n-3 and 20 : 4n-6, co-elute on a specific phase, resolution can be achieved in most cases by use of a more polar phase.

Given a constant CL, every additional DB increases the retention time. At a constant number of DBs, FAMEs in which the last DB is located farther away from the *w*-end elute before those in which the last DB is located closer to the *w*-end. In the FAs with 18 carbon atoms, the elution order is: 18 : 0, 18 : 1 (n-9), 18 : 1 (n-7), 18 : 2 (n-6), 18 : 3 (n-6), 18 : 3 (n-3), 18 : 4 (n-3).

Geometrical double-bond isomers can be resolved to a certain extent on very polar capillary columns of the cyanosilicone type.

The analysis of trans FAs in foods (occurring mainly in partially hydrogenated oils and fats and in ruminant fats) is becoming of considerable importance because of the implications of these acids for human health.[105] On very polar columns, the C_{18} monoethylenic trans FAs elute before the corresponding monoethylenic cis FAs, with some overlap.[103]

All geometrical isomers of linolenic acid methyl ester on a CP™-Sil 88 coated capillary column can be separated under isothermal conditions within 45 min. As in the monoethylenic FAMEs, the inclusion of cis DBs increased retention times compared with corresponding positions of trans DBs. The elution order of the geometrical isomers of linolenic acid methyl ester is: ttt < ctt < tct < cct < ttc < ctc < tcc < ccc, where t indicates trans and c indicates cis.

Equivalent chain length (ECL) values ranged between 19.45 and 19.89.[106]

The most important stationary phases of intermediate polarity are polyethylene glycol (DB-Wax, SUPELCOWAX-10, Carbowax 20M), acidified polyethylene glycol (FFAP), dimethyl–cyanopropylphenylpolysiloxane (DB-1701) (86 : 14 v/v) and cyanopropyl–phenyl–methyl (25 : 25 : 50 v/v) methylsilicone polymers (OV-225, DB-225, SP-2300). The main nonpolar stationary phases are based on dimethylsilicones (SPB-1, SPB-5) such as dimethyl–diphenylpolysiloxane (95 : 5 v/v) (DB-5, SPB-5, CP™-Sil 8CB) or 100% dimethylpolysiloxane (DB-1, Rt-1, SPB-1, SP-2100, OV-1, OV-101, CP™-Sil 5CB).[107]

Figure 2 shows the separation of FAMEs from a human milk sample on a CP™-Sil-5 cross-linked methylsilicone-coated column.

The nonpolar columns allow also a partial resolution of the saturated and unsaturated FAME, with the unsaturated species emerging slightly earlier than the corresponding saturated species. The nonpolar liquid phases vary in chemical structure and in their interaction with the FAs giving slight differences in relative migration rates, which, along with the film thickness and column length, can be exploited to minimize peak overlap. Generally lipid chemists favor the polar liquid phases for FAME.

The operating temperature constitutes an important practical aspect. Increases in column temperature increase the speed of the analysis but may decrease the peak resolution (the column life-time is reduced and the properties of the liquid phase may change).

Different liquid phases have different maximum operating temperatures. During the chromatography of TAGs, column temperatures exceeding 350 °C are necessary and only a few liquid phases are suitable for this purpose.

Other critical aspects of practical GC concern for TAGs include the method of sample introduction. In usual GC operations, the sample must be evaporated at a temperature that is higher than the column. High-temperature resolution of glyceryl and steryl esters is best performed by admitting the sample directly to the column at its starting temperature, which may be advantageously combined with programmed temperature vaporization.

Using capillary columns, it has been proven that the highest recoveries of components are obtained following on-column injection at or below the column temperature. This is performed using flexible quartz needles with wide bore capillaries,[108] or a special movable (''pop-up'') on-column injection apparatus[109] in which the sample is introduced while the front end of the column is outside the oven. Following injection of the sample, the column

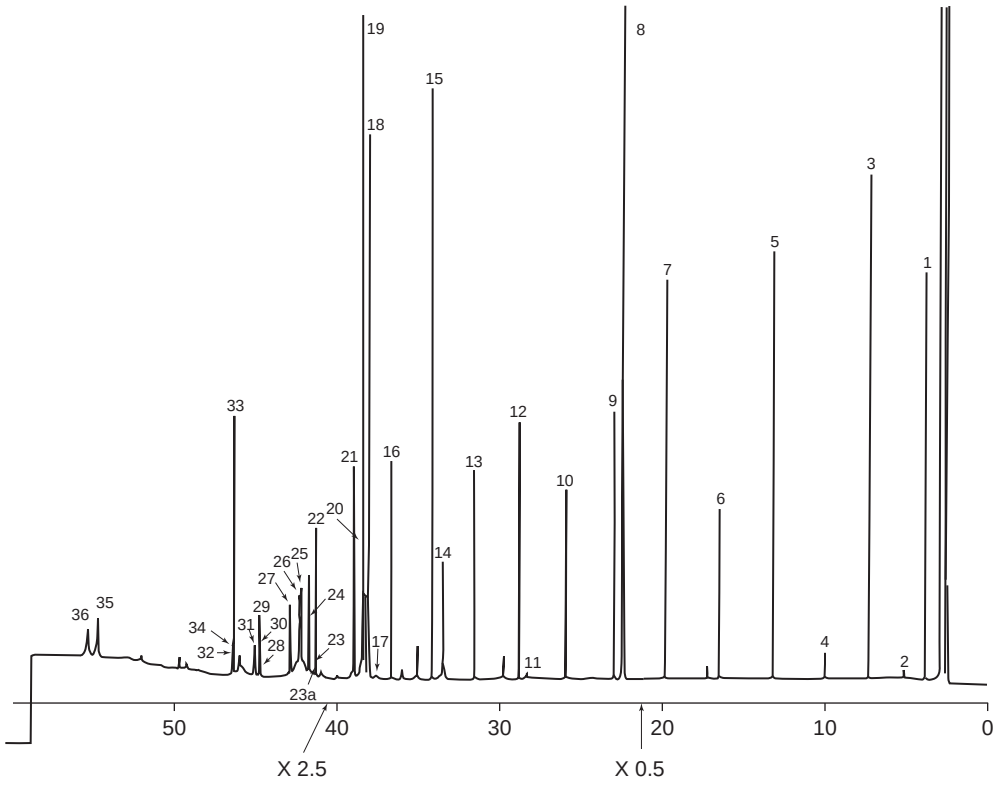

Figure 2 Separation of FAMEs from a human milk sample from Curacao on a CP™-Sil-5 cross-linked methylsilicone-coated capillary column (50 m × 0.20 mm internal diameter). Chromatographic conditions: split injection (split ratio 1 : 20); oven temperature program 60 °C, 4 °C min^{-1} to 200 °C, 50 °C min^{-1} to 285 °C, 285 °C for 15 min; carrier gas helium (0.52 ml min^{-1}). Peaks: 1, 5 : 0; 2, 6 : 0; 3, 7 : 0; 4, 8 : 0; 5, 9 : 0; 6, 10 : 0; 7, 11 : 0; 8, BHT; 9, 12 : 0; 10, 13 : 0; 11, 14 : 1; 12, 14 : 0; 13, 15 : 0; 14, 16 : 1 (n-7); 15, 16 : 0; 16, 17 : 0; 17, 18 : 3 (n-6); 18, 18 : 2 (n-6) + 18 : 3 (n-3); 19, 18 : 1 (n-9); 20, 18 : 1 (n-7); 21, 18 : 0; 22, 20 : 4 (n-6); 23, 20 : 5 (n-3); 23a, 20 : 3 (n-9); 24, 20 : 3 (n-6); 25, 20 : 2 (n-6); 26, 20 : 1 (n-9); 27, 20 : 0; 28, 22 : 5 (n-6); 29, 22 : 6 (n-3); 30, 22 : 4 (n-6); 31, 22 : 5 (n-3); 32, 22 : 0; 33, 24 : 1 (n-9); 34, 24 : 0; 35, 5β-cholestan-3α-ol; 36, cholesterol. (Reprinted from G. van der Steege et al., 'Simultaneous Quantification of Total Medium- and Long-chain Fatty Acids in Human Milk by Capillary Gas Chromatography with Split Injection', *J. Chromatogr. B*, **415**, 1–11, Copyright (1987), with permission from Elsevier Science.)

is returned to the oven held at the desired starting temperature. Capillary columns with split or splitless injection systems, which depend on evaporation with the sample, are not well suited for injection of solutions of compounds with a wide range of molecular masses. Such samples are evaporated and split at different ratios, usually resulting in poor recoveries of the components of higher molecular mass. No difficulties were encountered with determination of plasma total lipid profiles if on-column injection was used.

High-temperature GC on nonpolar columns has been extensively applied in the determination of total lipid profiles, and CN of diacylglycerols and TAGs. The original separations obtained on short packed nonpolar columns have been compared favorably with the separations obtained on capillary columns, including columns made with bonded nonpolar liquid phases.[110] The use of hydrogen as carrier gas and of capillary columns of minimum length have greatly improved the recovery and shortened the retention times of the compounds of higher molecular mass.[25] The nonpolar liquid phases still require derivatization of any polar functional groups.

Trimethylsilyl and terbutyldimethylsilyl derivatives are especially useful; they increase the volatility of the molecules and allow their recovery at lower temperatures and slightly shorter retention times than achieved with the corresponding acetates.

7.1.5.2 Capillary Column Dimensions and Film Thickness

The length of the column is an important parameter for resolution capability. Increases in column length improve the resolution capability but also increase retention times. Relatively short columns with very polar phases often give better resolution of unsaturated FAMEs than relatively long columns with nonpolar phases. Most of the columns used for FAME separation range between 25 and 50 m in length. Excellent separation of all the unsaturated FAMEs from various tissues has been demonstrated

on a 25 m very-polar (CP™-Sil 88) column, whereas on a 50 m nonpolar column based on methylsilicone, separation of all the important unsaturated FAs could not be achieved.[111–113]

Relatively short columns (25 m or shorter) can be used for rapid analysis of relatively simple FAME mixtures. A common vegetable oil containing FAs with CLs between 14 and 18 carbon atoms can be analyzed in less than 5 min on a short column.[114]

The internal diameter and the film thickness of a WCOT capillary column influence its capacity, resolution capability and FAME retention times. Increasing film thickness increases the sample capacity of the column, the resolution capability and the retention times. In most cases, a film thickness of 0.25 μm will provide sufficient resolution capability (although the resolution capability, of course, depends primarily on the type of stationary phase).

Decreasing the internal diameter of the column increases the resolution capability of the column but increases retention times and reduces the column capacity.

If narrow-bore capillary columns (with thin films) are used, the amount of sample introduced into the column must be restricted by splitting the carrier gas at a large split ratio. However, most analyses can be performed on medium-bore capillary columns that have an internal diameter of 0.25 or 0.32 mm.

7.1.6 Identification of Fatty Acid Methyl Esters

The best way to identify FAMEs is to compare their retention times with those of individual purified standards. An alternative way of identifying FAMEs is based on their relative retention times or equivalent chain-length values.[114] The former is defined as the ratio between the retention time of any FAME and that of a reference FAME (usually $C_{16:0}$ or $C_{18:0}$).

The use of equivalent chain-length values (based on Kovats' retention index[115–117]) results in a ready visualization of the position of a particular FAME relative to saturated straight-chain acids, 16 : 0, 18 : 0, 20 : 0, 22 : 0, 24 : 0. Values for FAMEs can then be calculated either from an equation or by plotting on semilogarithmic paper.

7.1.7 Quantification of Fatty Acid Methyl Esters

For detection of FAMEs, the FID is most convenient. The amounts of FAMEs are usually determined via the peak area. (The peak area of a given amount of FAME is not influenced by the peak shape provided that extremely strong leading or tailing does not occur. Electronic digital integration is the most accurate and reproducible means of quantifying peak areas.) Quantification is commonly carried out after calibration of the system with standards containing known amounts of FAMEs. FAME mixtures should be similar in composition to the samples to be analyzed.

A usual method for absolute quantification is the addition of a FAME not present in the sample as an internal standard. The internal standard should be selected according to the sample to be analyzed.

Because in most food samples FAs with 16 and 18 carbon atoms are predominant, $C_{17:0}$ is a regular internal standard in the analysis of FAMEs.[118] If the sample contains $C_{17:0}$, other FAMEs such as $C_{19:0}$, $C_{21:0}$ or $C_{23:0}$ can be used.[119]

7.2 Argentation Thin Layer Chromatography

Ag-TLC is based on the ability of silver ions to form polar complexes reversibly with unsaturated centers in organic molecules such as lipids. It enables separation according to the number, geometrical configuration and position of DBs in molecules.

In Ag-TLC, both precoated and homemade plates can be impregnated with silver ions to give satisfactory separations. The adsorbent is commonly silica gel G.

FAs, usually as the FAMEs, can be separated according to the number and configuration of the DBs.

The choice of mobile phase depends on the separation required. Resolution of methylene-interrupted unsaturated FAs with up to three DBs can be achieved using benzene–hexane as the developing solvent.[120–122] Separation of FAs with up to six DBs is possible with a single development of benzene–ethyl acetate while other methods involve double developments with hexane–diethyl ether–acetic acid mixtures or a two-stage development, first with a polar solvent (chloroform–methanol–water or hexane–diethyl ether mixtures) to separate FAs with three to six DBs, followed by a development suitable for FAs with fewer DBs.[123–125]

Using HPTLC plates and toluene–acetonitrile mixtures as mobile phase, components with two to six DBs and 18, 20 and 22 carbons can be resolved into zones according to degree of unsaturation; within each zone there is separation of FAs apparently according to CL, although the position of the first DB in each ester may be the important factor (Figure 3).[126]

All (n-6) polyunsaturated FAs can be separated, as are (n-3) polyunsaturated FAs. Also geometrical isomers with trans DBs are readily separated from the corresponding cis isomers, the former having greater mobility.[127]

An important problem is the resolution of mixtures containing FAs with more than one DB with all possible combinations of geometrical isomers. For mixtures of C_{18} FA isomers with zero to two DBs,

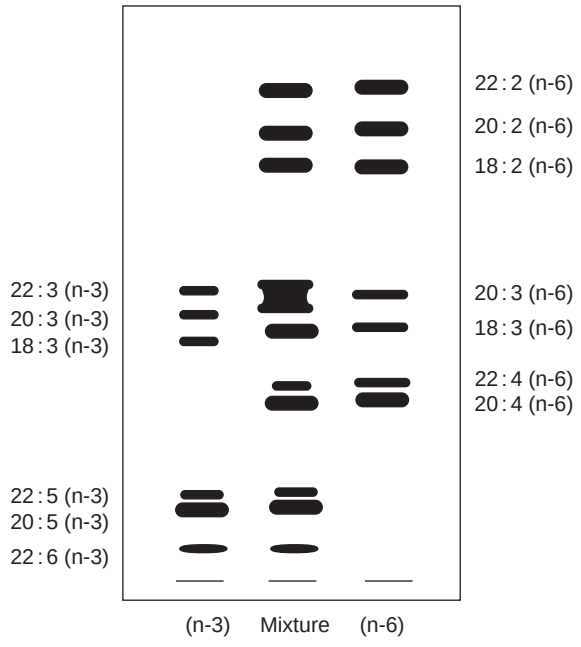

Figure 3 Diagrammatic Ag-TLC separation of (n-3) and (n-6) polyunsaturated FAMEs.[126] The silica gel 60 TLC plate was prepared by spraying with a 10% (w/v) solution of $AgNO_3$ in acetonitrile. The developing solvent was toluene–acetonitrile (97:3 v/v) and bands were visualized by spraying with 3% (w/v) copper acetate–8% (v/v) orthophosphoric acid in water. (Reprinted from R. Wilson, J.R. Sargent, 'High Resolution Separation of Polyunsaturated Fatty Acids by Argentation Thin-layer Chromatography', *J. Chromatogr.*, **623**, 403–407, Copyright (1992), with permission from Elsevier Science.)

cis-monoenes overlapped with trans,trans-dienes under standard conditions.[128] They could be resolved, however, by using silver benzene sulfonate-impregnated plates and by modifying the mobile phase. A similar mixture with the addition of conjugated dienes (trans/trans, cis/trans, cis/cis) could be resolved completely only by using two plates of differing $AgNO_3$ concentrations with large volumes of different developing solvents (petroleum ether–acetone and chloroform–methanol) in open tanks.[129]

Ag-TLC can be also applied to a variety of substituted FAs including unsaturated epoxy, hydroxy, dihydroxy and halohydroxy FAs. One interesting separation is that of methyl 11-(*R*)-hydroxy-9(*Z*),12(*Z*)-octadecadienoate from the unresolved pair of methyl 9-hydroxy-10(*E*), 12(*Z*)- and 13-hydroxy-9(*Z*),11(*E*)-octadecadienoates in a study of the oxidation products from linoleic acid treated with a red algal enzyme preparation.[130] Recently, hydroperoxides of methyl oleate have been converted to hydroxides and separated into cis and trans isomers by Ag-TLC prior to separation of the hydroxy positional isomers by HPLC and identification by GC ion trap MS.[131]

8 ANALYSIS OF TRIACYLGLYCEROLS

8.1 Thin-layer and Conventional Column Chromatography

The most commonly employed methods for analysis of TAGs are Ag-TLC[132–134] and RPTLC.[135]

8.1.1 Silver Ion Thin-layer Chromatography

Since the first separations of TAGs by Ag-TLC,[136] the technique has been used, either on its own or as a preliminary simplification step, to elucidate the structures of many fats and oils, including cocoa butter, cod liver oil, cottonseed oil, evening primrose oil, grape seed oil, lard, maize oil, milk fat, olive oil, orange seed oil (Figure 4),[137] pepper seed oil, soybean oil, sunflower oil and tomato seed oil, as well as for studying the process of interesterification and the changes produced by heating.

This method is based on the ability of alkenic bonds to form π-complexes with silver ions, which permits the separation of substances with differing degrees of unsaturation.[138]

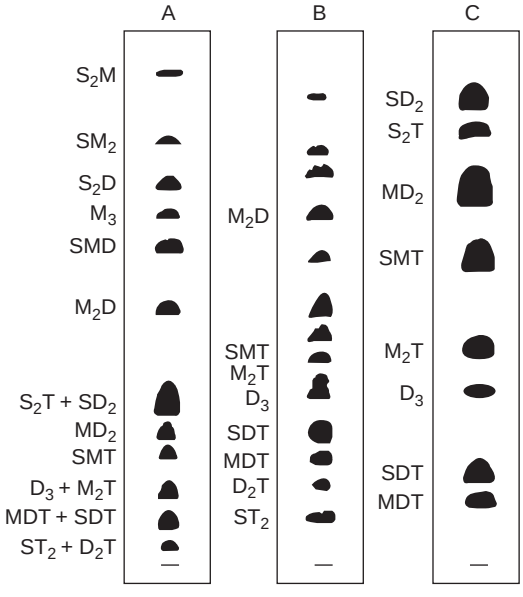

Figure 4 Diagrammatic Ag-TLC separation of TAGs from orange seed oil. The silica gel G layer was impregnated with 0.5% methanolic $AgNO_3$ and plates were developed in open cylindrical tanks. Plate A, mobile phase of petroleum ether–acetone (8 ml, 100:6 v/v); plate B, petroleum ether–acetone (5 ml, 100:8 v/v) followed by petroleum ether–acetone (8 ml, 100:5 v/v); plate C, petroleum ether–acetone (4 ml, 100:7 v/v) followed by petroleum ether–acetone (15 ml, 100:4 v/v). Bands were visualized by exposing plates sequentially to bromine and sulfuryl vapors followed by heating to 180–200 °C. (Reproduced by permission of the authors and *La Rivista Italiana Delle Sostanze Grasse* and redrawn from Tarandjiiska and Nguyen.[137])

Of species with the same total number of DBs, that in which the DBs are more concentrated in individual acyl residues has the slower mobility. TAGs containing linolenate are usually retained more strongly than those containing acyl groups with up to two DBs but with one more DB in total, although the order of species with trienoic acids can vary according to the nature of the mobile phase. The 20 possible species containing FAs with zero to three DBs (i.e. trisaturated to trilinolenate) can be totally resolved.

The thin layer is a silica gel slurry containing 5–25% $AgNO_3$ applied to a plate. The solution of glycerides is applied to the plates in the form of a streak, and the plate is developed using ether–benzene or chloroform–methanol mixtures. Under these conditions, saturated TAGs migrate together with the front, followed by the various groups of TAGs in order of increasing unsaturation.[139,140]

After separation, the TAGs are located by UV light and may be quantitatively recovered by elution of the respective bands; the composition of each TAG can be determined by its location on the chromatogram and analysis of its FAs by applying the usual methods of conversion to FAMEs and GC. A range of developing solvents can be used for resolving TAGs. These include benzene with cyclohexane, cyclooctane, diethyl ether or methanol, petroleum ether–acetone, tetrachloromethane–acetic acid, isopropyl alcohol–chloroform and toluene with chloroform, diethyl ether or hexane. Methanol (1%) in chloroform can be recommended for resolving components with up to six DBs, and increasing the amount of methanol to 5% for separation of more highly unsaturated species on a separate TLC plate.[18]

TAG isomers differing only in the position of the acyl moieties (at the *sn*-2 and *sn*-1/3 positions) can be resolved also, allowing the identification of species with specific FAs at the *sn*-2 position without the need for techniques such as enzyme hydrolysis. SSM–SMS, SSD–SDS and MSM–MMS pairs have been separated with the former member of each pair migrating the least (where D is unsaturated, M is monounsaturated and S is saturated). All these pairs and others (DDS–DSD and DDM–DMD) have been resolved by continuous development with chloroform–methanol in open tanks.[141,142]

8.1.2 Reversed-phase Thin-layer Chromatography

RPTLC was first introduced into TAG analysis by Kaufman et al.[143] and it has been successfully used to resolve TAG mixtures according to their partition number (PN), which is given by Equation (2):

$$PN = CN - 2n \quad (2)$$

where n is the number of DBs in the TAG molecule. The best resolutions have been achieved using stationary phases consisting of long-chain hydrocarbons or liquid paraffin and the quantification usually proceeds by indirect GC or spectrophotometric methods. Today, densitometry is the method of choice for quantification of the compounds separated by TLC. Since charring is most often used for the direct densitometric determination of TAGs, long-chain hydrocarbons and liquid paraffin are unsuitable as the stationary phase. Reverse-phase-chromatographic systems must, therefore, combine good selectivity with the possibility of direct densitometric quantification. Plates are cooled and stored in a desiccator over phosphorus pentoxide until silanization. Silanization of the layer is achieved by placing the plates in a closed chamber over dimethylchlorosilane vapors for a certain length of time. The plates are washed by elution with methanol and dried at 110 °C for 1 h and then stored in closed glass boxes without any special precautions. Densitograms are carried out on a densitometer by zigzag scanning in transmission mode at 450 nm using a 1.2 mm slit.

To ascertain which systems are optimal for TAG analysis, various layer types, mobile phases and detection methods have been tested under identical conditions using a variety of TAG standards.[144] Three different types of layer (plain unimpregnated silica gel TLC, silica gel layers impregnated with $AgNO_3$ by the manufacturer or in the laboratory, and silica gel layers with chemically bonded C_{18} groups) were compared by Masterson et al.[145] using chromatographic standards and applied to analyses of hen's egg yolk, lard and snail extracts.

Ag-TLC plates can be prepared by hand-dipping the silica gel plates into a Desaga glass dipping chamber containing a 2.5% solution of $AgNO_3$ in methanol, followed by activation in an oven at 110 °C for 1 h just before use. Commercially prepared plates (20 cm × 30 cm) impregnated with 10% $AgNO_3$ were also used.

Standards were chosen based on both saturation class (S_3, S_2M, M_3 and D_3) and PN. All standard solutions must be prepared at a concentration of 1.0 g l^{-1} in chloroform.

The best separations are achieved using isopropyl alcohol–chloroform with Ag-TLC plates and mixtures of isopropyl alcohol–chloroform (1.5 : 98.5) with plain silica gel TLC and acetonitrile–methyl ethyl ketone–chloroform (50 : 35 : 15).

8.2 Separations with Column Chromatography

Neutral and polar lipid classes can be separated and isolated for subsequent analysis using column chromatographic modes, creating selectivity in the isolation of compounds by altering the solid/support solvent system so that a unique interaction exists between the compound to be isolated and the functional group of the solid phase.

Compounds that are very diverse in chemical structure may differ greatly in their interactions with the solid-phase surface moieties, whereas compounds that are chemically similar may exhibit only subtle, though still exploitable, differences. Therefore, by varying the solvent environment (e.g. polarity) around the solid phase, lipids can be selectively isolated with a high degree of purity and recovery. The most widely used columns are aminopropyl bonded phase (Bond Elut) columns that may or may not be placed in a vacuum curve to enhance the separation rate. The mixtures used with this type of column to elute TAGs are 1% diethylether and 10% methylene chloride in hexane, and the quantity used to elute the TAGs is 8 ml.[146,147]

8.3 Thin-layer Chromatography with Flame Ionization Detection

TLC with FID is a simple, easy-to-use analytical method that is sufficiently sensitive to produce good separations and is suitable for the quantification of complex lipids. In TLC/FID, the sample is spotted on a quartz rod coated with silica gel, developed chromatographically and finally subjected to detection by FID as the rod is passed through a hydrogen flame.

The response factors for various lipids differ and depend upon such variables as sample volatility, amount of material analyzed, its composition and the rate of passage of the Chromarod through the flame.[148–151]

Quantification of lipids by FID after separation by TLC on Chromarods can be improved by the use of internal standards that reduce rod-to-rod variation but the relative response of different lipid classes to an internal standard has not been found to be constant: it varies with the amount of sample applied, the ratio of sample to internal standard and the scanning speed.[152]

Since latroscan FID sensitivity varies depending upon the substances to be detected, it is necessary to obtain a suitable calibration curve prepared on the basis of the peak area measurements and the amounts of the components.[153]

The latroscan technique has proven highly effective in separating phospholipids, cholesterol, TAGs and FAs, but the system may not always be so effective in separating TAGs.[154,155]

8.4 Enzymatic Methods

Analysis of TAGs has been simplified by the introduction of enzymatic methods, which have been automated to provide analysts with quick, easy and direct procedures. Enzymatic methods are based on determinations of the glycerol portion of the TAG molecules after hydrolysis to remove the FAs. Dosing takes place in the following three stages: TAG emulsion, complete hydrolysis by a lipase and a stearase, and indirect spectrophotometric measurement of the glycerol at 340 or 500 nm.[156]

The TAG can be determined using either a dehydrogenase system in which oxidation of NADH is followed at 340 nm or the oxidase/peroxidase system in which formation of a blue color is followed at 500 nm (Scheme 1).[157,158]

8.5 Gas Chromatography

The first studies on separations of TAGs using GC were carried out by Fryer et al.[159] and Hubber.[160] Kuksis and McCarthy[161] improved the separation technique by employing direct injection onto the column.

In the 1960s, TAG analysis was carried out on columns packed with dimethyl polysilene (OV-1, OV-101, JXR, S-30) or phenyl methyl polysiloxane (OV-17, SE-52). The internal diameter was 2.5–3.5 mm. The column length depended upon the type of separation. For TAGs above C_{40}, 0.5–0.7 m columns were used. Columns longer than 2.0 m are impractical for separating TAGs above C_{48}.

$$\text{TAG} + 3H_2O \xrightarrow{\text{Lipase/stearase}} \text{Glycerol} + 3\text{FFA}$$

$$\text{Glycerol} + \text{ATP} \xrightarrow{\text{Glycerol kinase}} \text{Glycerol 3-phosphate} + \text{ADP}$$

Dehydrogenase system

$$\text{ADP} + \text{Phosphoenolpyruvate} \xrightarrow{\text{Pyruvate kinase}} \text{Pyruvate} + \text{ATP}$$

$$\text{Pyruvate} + \text{NADH} \xrightarrow{\text{Lactate dehydrogenase}} \text{Lactate} + \text{NAD}^+$$

Oxidase/peroxidase system

$$\text{Glycerol 3-phosphate} + O_2 \xrightarrow[\text{oxidase}]{\text{Glycerophosphate}} \text{Dihydroxyacetone phosphate} + H_2O_2$$

$$\text{4-Chlorophenylphenazone} + \text{4-Aminophenazone} + 2H_2O_2 \xrightarrow{\text{Peroxidase}} \text{4-(4-Benzoquinonemonoimino)-phenazone} + \text{HCl} + 2H_2O$$

Scheme 1 Enzymatic assay of TAG. The dehydrogenase system is followed by A_{340}, which decreases as NADH is oxidized. The oxidase/peroxidase system generates a blue color that is followed at 500 nm.

Since the 1970s, capillary GC using apolar columns (OV-1, SE-52, SE-54) has become a routine tool for characterizing the TAG profile of fats.[162] TAG separation on apolar silicone phases is based mainly on molecular mass differences (CN determinations), whereas unsaturation barely contributes to peak retention.[163]

The columns used at the end of the 1990s, have chemically bonded, polarizable stationary phases; their length varies from 5 to 25 m, with an internal diameter ranging from 0.2 to 0.32 mm. Column length must increase as the molecular mass of the TAGs to be separated increases. The stationary phase used has an upper temperature limit of 360 °C. Hydrogen is the most commonly used carrier gas with this type of column, and the hydrogen flow rate decreases significantly with increasing temperature. Using hydrogen as the carrier gas affords several advantages over nitrogen, used in packed columns and in the earlier capillary columns. One such advantage is a lower elution temperature compared with that for nitrogen and helium, leading to shorter analysis times and less thermal decomposition.

On employing isothermal chromatographic conditions to analyze a mixture of TAGs with different vapor pressures or long retention times at high temperature, the amount of the stationary phase and column efficiency (length) are selected so as to achieve good separation of the most volatile components. With TAGs, the main problem is always to reduce the analysis time for isothermal chromatography; however, longer columns are needed to achieve separation of the low-molecular-mass components, while retention times for high-molecular-mass TAGs are very extended, and certain ones may not elute. When programmed temperature chromatography is used, the retention times for a series of compounds with the same CN generally increases linearly with the number of carbon atoms rather than exponentially as occurs in isothermal chromatography.

8.5.1 Stationary Phase

The high-temperature polarizable phenylmethylsilicone liquid phases permit high-temperature GC separation of natural TAGs based on degree of unsaturation.

The phenylmethylsilicone liquid phase becomes sufficiently polar as the temperature increases above 290 °C, as indicated by the longer retention of unsaturated compared with saturated TAGs. As the temperature increases, the molecular species of TAGs are eluted in order of increasing number of DBs per molecule.

Wide-bore capillary columns containing polarizable phenylmethylsilicone can be used for high resolution of palm oil TAG fractions with automated cold on-column injection; polarizable phase capillary GC can be also used in combination with MS for the determination of the elution profiles and for identification of the eluted species of natural TAGs.

Identification of closely eluting or overlapping GC peaks can be obtained by prior Ag-TLC resolution. This part of the analytical system can be applied sequentially to the resolution and quantification of the diacylglycerol moieties of glycerophospholipids and TAGs, and subsequently to their FA moieties.[6] Figure 5 illustrates the application of polarizable capillary GC in combination with Ag-TLC for the resolution of butteroil TAGs.[164]

With a polarizable methylphenylsilicone stationary phase, the CN sequence is retained, but unsaturation now contributes to the fine structure of each CN fraction. For example, the order of elution becomes tinpalmitin, palmitoyl-oleoyl-palmitin and palmitoyl-linolenoyl-palmitin, all of which have saturated acids on the molecule; separation by triplets takes place according to the unsaturation of the middle acid.[165] Vegetable and animal fats can be analyzed using this type of column, separating substances with the same number of carbon atoms but differing by only a single DB.[166]

8.6 High-performance Liquid Chromatography

8.6.1 Reversed-phase High-performance Liquid Chromatography

RPHPLC is the principal chromatographic technique for separating mixtures of TAGs. TAGs are separated based on CL and the degree of unsaturation of the FAs. The PN is used to characterize TAG molecules and determine their elution (Equation 2).[167]

Different workers have developed other parameters analogous to PN for molecularly discriminating between TAGs with the same PN value, mainly to improve identification of chromatogram peaks.[168–170]

Considerable progress has been now made in improving RPHPLC methodology, overcoming the difficulties encountered in separating the TAGs in natural fats.

8.6.1.1 Separation of Triacylglycerols of Similar Partition Number Mixtures of TAGs composed of FAs differing only minimally in CL and degree of unsaturation have the same PN value. Despite the improvements in efficiency achieved using octadecyl silanized columns with a particle size of 3 μm connected in series and elution gradients with acetone as the organic modifier, poorly resolved or even unresolved TAG pairs and groups still exist, especially in highly complex mixtures of natural fats such as milk fat and fish oils.[171]

8.6.1.2 Separation of Triacylglycerols with Marked Differences in Structure Simultaneous analysis of TAGs differing markedly in molecular structure (i.e. short-,

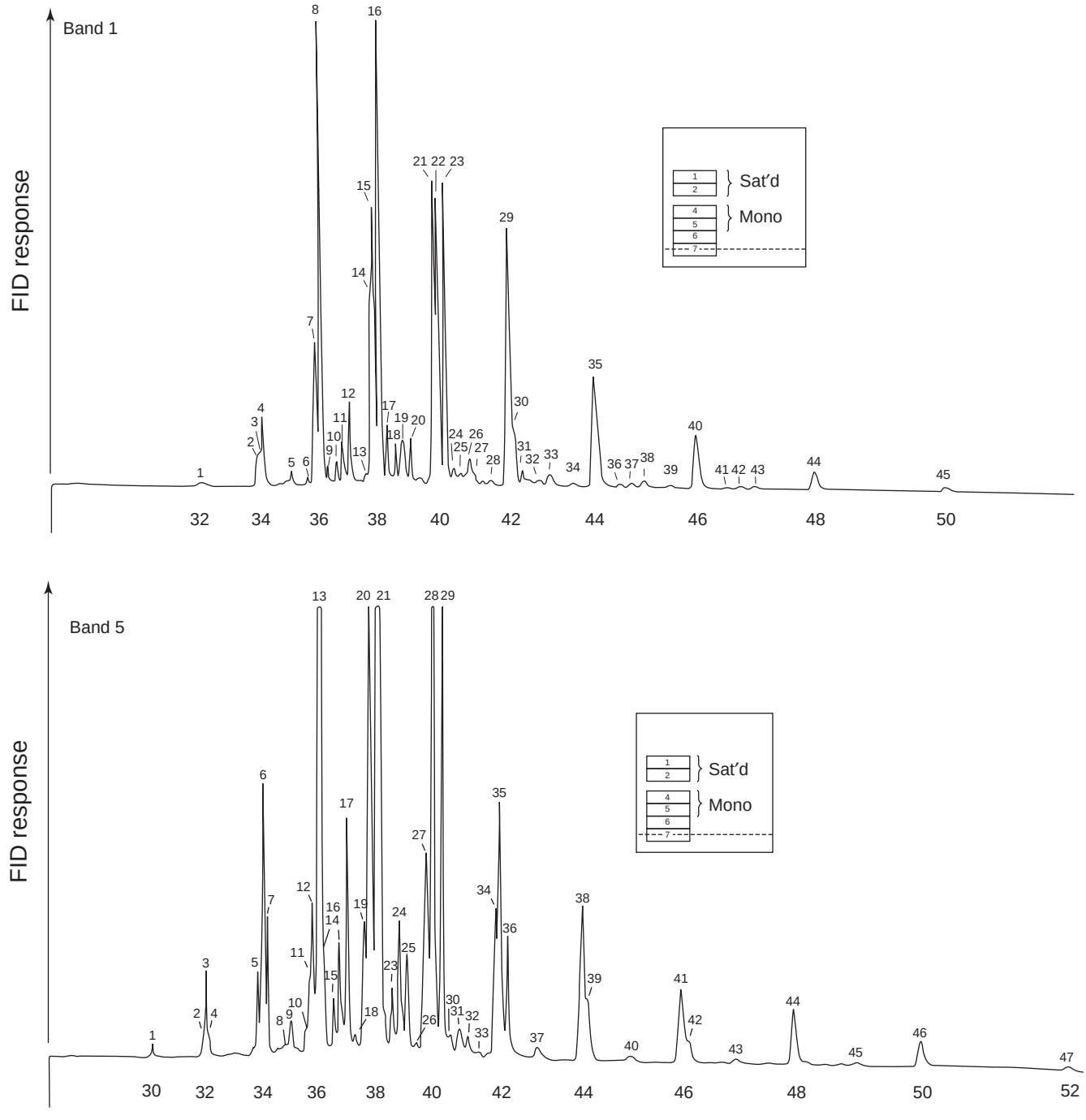

Figure 5 Polar capillary GC profiles of the saturated long chain-length (band 1) and the short-chain cis-monoene (band 5) TAG fractions isolated from R-4 butterfat distillate by Ag-TLC. Inset: 15% Ag-TLC resolution of R-4 distillate in chloroform plus 0.75% ethanol. Peak identification band 1: 7, 8-14-14 + 8-12-16 + 10-14-12; 8, 6-14-16 + 6-12-18; 14, 10-14-14 + 10-12-16; 15, 8-14-16 + 8-12-18; 16, 6-16-16 + 6-14-18; 21, 10-14-16 + 12-14-14; 22, 8-16-16 + 8-14-18; 23, 6-16-18; 29, 10-16-16 + 10-14-18; 30, 8-16-18; 35, 10-16-18 + 12-16-16 + 14-14-16; 40, 14-16-16; 44, 16-16-16; 45, 16-16-18. Peak identification band 5: 6, 12-18 : 1-6 + 14-16 : 1-4; 12, 12-18 : 1-6 + 14-16 : 1-6; 13, 14-18 : 1-4 + 16-16 : 1-4; 20, 18 : 1-14-8 + 12-16 : 1-10; 21, 18 : 1-16-4 + 16 : 1-18-4; 27, 18 : 1-14-8 + 16 : 1-14-10; 28, 18 : 1-16-6; 29, 18 : 1-18-4; 34, 10-181-14 + 12-16 : 1-14; 35, 18 : 1-16-8; 38, 10-18 : 1-16; 41, 16-18 : 1-12 + 14-14-18 : 1; 44, 14-16-18 : 1; 46, 16-16-18.1. GC conditions: column, flexible quartz capillary (25 m × 0.25 mm internal diameter, RSL-300 custom made); carrier gas H_2, on-column injection; temperature 40 °C, ballistically heated to 290 °C, at 10 °C min^{-1} to 330 °C and at 2 °C min^{-1} to 350 °C. (Reprinted from J.J. Myher et al., 'Identification of the More Complex Triacylglycerols in Bovine Milk Fat by Gas Chromatography–Mass spectroscopy using Polar Capillary Columns', *J. Chromatogr.*, **452**, 93–118, Copyright (1988), with permission of Elsevier Science.)

medium- and long-chain TAGs with differing degrees of unsaturation spanning a broad range of PN values) is difficult because of problems in achieving, simultaneously, both good resolution of the least strongly retained TAGs and reasonable analysis times and elution in narrow chromatographic bands for the most strongly retained TAGs.[172]

8.6.1.3 Detection of Triacylglycerols at the Column Outlet
Refractive index and UV detectors were most commonly used in HPLC analyses for fats.[173,174] Consequently, it has been necessary to develop new systems such as light-scattering detectors and combined HPLC/MS. As pure standards are usually unavailable, and incomplete resolution of critical groups makes peak identification difficult on the chromatograms, combining HPLC with MS may furnish valuable information on the structure of TAGs.

The stationary phase yielding the best separations is octadecylsilane supported on spherical particles of silica gel. While good separations have been obtained using a particle diameter of 5 μm, reducing the diameter to 4 and 3 μm improved selectivity between chromatogram peaks for columns of similar size.

Octadecylsilane columns are normally 15 to 30 cm long with an internal diameter of 4–5 mm. Two and three columns have been connected in series to improve separations of the TAGs from natural fats. Increases in column length are limited by the rise in column head pressure and the substantially longer analysis times.

8.6.2 Mobile Phase

Mobile-phase composition is the most important chromatographic factor affecting the separation of TAGs in RPHPLC because of the small variability in the type of stationary phase employed. Most researchers have elected to use a mobile phase comprising an organic base solvent and an organic modifier. Acetonitrile is the most widely used organic base solvent, though *n*-propionitrile has sometimes been preferred.[175]

Organic modifiers are added to improve the solubility of the TAGs in the base solvent of the mobile phase, to bring about changes in mobile-phase polarity and to increase peak selectivity. When choosing an organic modifier, the following factors should be taken into account, in addition to the nature of the solvent:

- whether the mixture consists of single or multiple components
- the percentage of organic modifier in the mobile phase
- whether that percentage is constant or varies over the course of elution.

Under both isocratic and gradient elution conditions, an increase in mobile-phase polarity leads to an increase in retention times and in the selectivity of pairs or groups of TAGs with the same PN: there are linear relationships between the logs of the capacity factor (k) and TAG selectivity and the percentage organic modifier in the mobile phase, irrespective of solvent type and elution conditions.[176]

The choice of organic modifier, single or multiple component composition of the modifier, proportion of modifier to base solvent, and possible variations in that proportion during elution depends mainly upon the type of TAG to be separated and the detection system employed. Although many different solvents have been used as the organic modifier, acetone is most commonly employed because of its effect in improving the selectivity of TAG pairs or groups with the same PN. Acetone cannot be the modifier with UV detectors because of its strong absorbance at the wavelengths absorbed by the TAGs (200–237 nm).

In such cases, hexane, *n*-propanol, ethanol and methyl-*tert*-butylether have been proposed as organic modifiers and have produced good separations of complex mixtures.

Initially, TAG analysis was carried out using a constant proportion of organic modifier in the mobile phase with refractive index detection.[177,178] Those conditions of analysis yielded relatively satisfactory separations of simple mixtures of TAGs, such as most vegetable oils, which have mean composition of 5–7 FAs and 15–20 TAG species. Although the resolution of most such TAGs was good, certain critical pairs remained unresolved, e.g. LLL–LnLO and LnLO–LnLP for PN 42; LLO–LnOO, LnOO–LnLSt and LLP–LnOP for PN 44; LLSt–LOP for PN 46; and OOO–LOSt and OOP–LStP for PN 48 (where L is linoleic acid, Ln is linolenic acid, O is oleic acid, P is palmitic acid and St is stearic acid).

Varying the proportion of organic modifier during elution has been proposed as a solution when analyzing more complex natural mixtures, where there is a much greater likelihood of encountering molecules with the same PN value because of the larger number of FAs than found in vegetable oils and, usually, the much higher number of TAG types.

Gradient elution systems are intended to ensure good solubility of all the TAGs present in the mixture in the mobile phase throughout the run, reduce retention times for highly saturated TAGs and, at the same time, achieve good resolution of critical pairs with the same PN.[179] Employing two 5 μm octodecylsilane columns, a linear gradient of 20 to 80% (v/v) dichloromethane–acetone (3 : 1 v/v) in acetonitrile and light-scattering detection, it is possible to achieve excellent separation of the TAGs in vegetable oils and animal fats (Figure 6).[180]

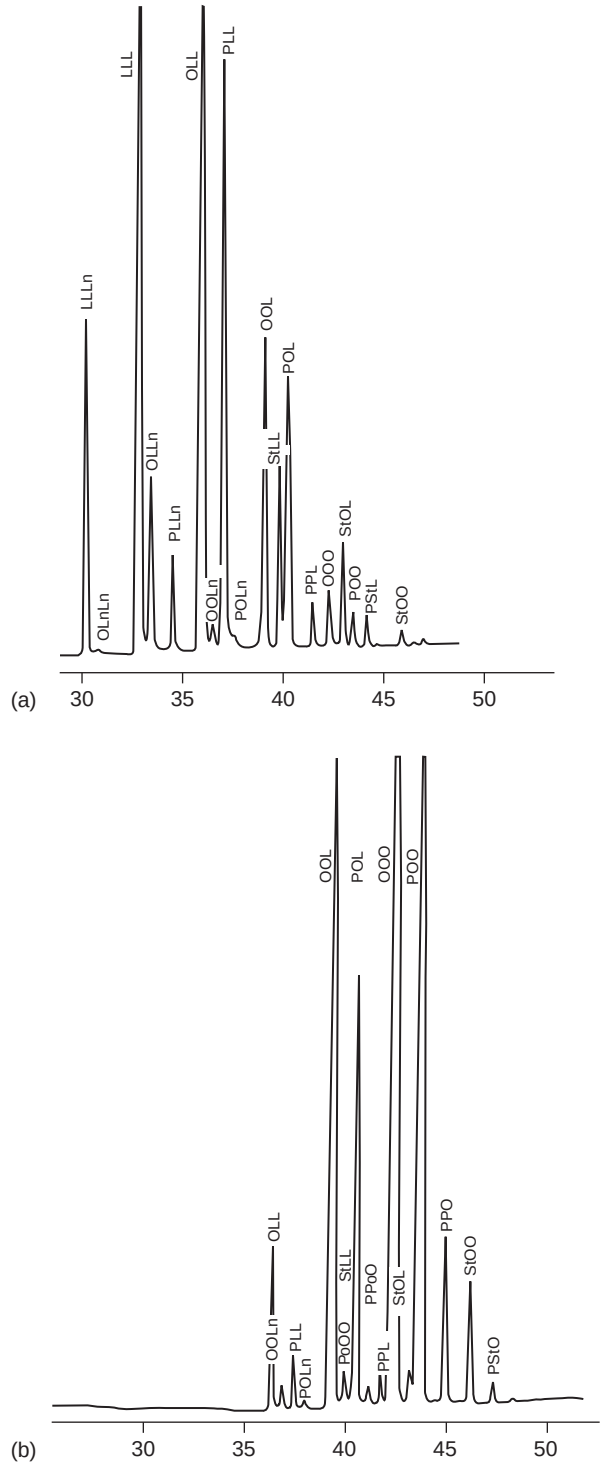

Figure 6 HPLC separation of soybean oil (a) and olive oil (b) TAGs using two 5 µm Lichrosphere 100 CH-100 SUPER columns (250 mm × 4.0 mm internal diameter) connected in series. Column temperature ambient; mobile-phase linear gradient of 20 to 80% (v/v) dichloromethane–acetone (3 : 1 v/v) in acetonitrile for 60 min at a flow rate of 1.0 ml min^{-1}; mass detection. (Reproduced from Perrin and Prevot.[180] Permission sought.)

Despite the enhancement in resolution brought about by elution gradients, complex mixtures of natural fats still contain many unresolved TAG pairs or groups with the same PN, and this is one of the most important unsolved analytical challenges remaining today. However, nonlinear gradients and specific-step gradients have been used, chiefly with a view to improving the separation of critical pairs.[181]

Normally, the mobile-phase flow-rate ranges from 0.5 to 2.0 ml min^{-1}.

Under isocratic elution conditions, small increases in flow rate enhanced the resolution of certain critical pairs.[171,176]

8.6.3 Sample Solvent and Column Temperature

The sample solvent may exert a major effect on the resolution of a chromatographic system. Various solvents reveal substantial differences according to the nature of the injection solvent and the mobile phase. The effect of the solvent on the chromatographic resolution can be explained in terms of the theory put forward by Horvath and Melander.[182] They suggested that in order for a dissolved solute molecule to be carried by the mobile phase it must create room and this gives rise to a reduction in free volume. Chloroform is suggested as the sample solvent when working with mobile phases in which solubility of TAGs, particularly highly saturated TAGs, is poor.

Most RPHPLC analyses of TAGs can be done with the column at ambient temperature; however changes in column temperature will result in changes in chromatographic resolution. Higher column temperatures decrease retention times and selectivity for TAGs, especially for critical pairs with the same PN.

Although lower temperatures result in better separation of the TAGs, the most highly saturated TAGs may precipitate out of the mobile phase, and consequently the preferred working temperatures is at least 30 °C.

Because the choice of column temperature represents a compromise designed to ensure good solubility and good selectivity of critical pairs with the same PN, it is advantageous to use temperature gradients to analyse complex mixtures of TAGs spanning a broad range of PN values.[183] Figure 7 illustrates the separation of butterfat TAGs using a temperature gradient of 10 to 55 °C with a 5 µm octodecylsilane column (250 × 4.0 mm internal diameter) and a mobile phase of acetonitrile–acetone (50 : 50 v/v).[183]

8.6.4 Detectors

Refractive index detectors have been used in most analyses of simple mixtures of TAGs under isocratic

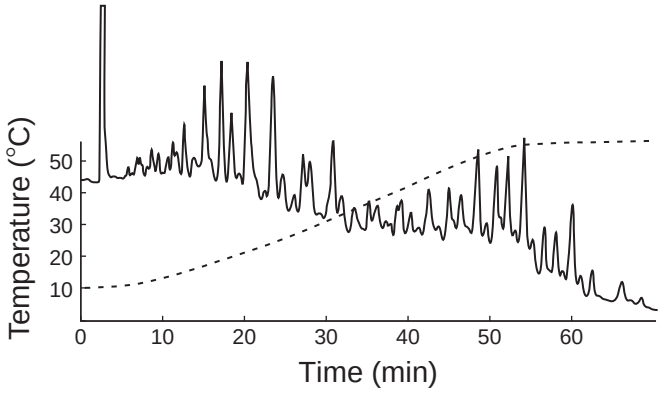

Figure 7 HPLC separation of butterfat TAGs using a single 4 μm Lichrosphere 100 RP-18 column (250 mm × 4.0 mm internal diameter). Column temperature programmed (dashed line) from 10 to 55 °C at a rate of 1 °C min^{-1}; mobile phase 50% (v/v) acetone in acetonitrile at a flow rate of 1.0 ml min^{-1}; mass detection. (Reprinted from C. Maniongui et al., 'Determination of Bovine Butterfat Triacylglycerols by Reverse-phase Liquid Chromatography and Gas Chromatography', *Journal of Chromatography B*, **543**, 81–103, Copyright (1991), with permission of Elsevier Science.)

conditions of elution. However, they have low sensitivity and low stability because of their sensitivity to changes in temperature and pressure and cannot be used with gradient elution because they are affected by the composition of the mobile phase.

Plattner[184] used infrared detectors at 5.75 μm, especially for qualitative purposes. Although this type of detector is suitable for gradient elution, substantial baseline drift may occur.

Absorption in the UV region 200–237 nm provides for highly sensitive detection of TAGs and good detector stability and can be used with gradient elution.[185] Considerable baseline drift may occur when the gradient composition of the mobile phase varies over a broad range.[186] UV detector response in that region of the spectrum is mainly to weakly chromophoric ester bonds with a lower molar extinction coefficient. However, UV absorption can be considerably enhanced by interactions between the carbonyl groups and the DBs.

Moving-fire FIDs are not widely used or available commercially for HPLC analysis and may be subject to shortcomings involving low sensitivity because only a small portion of the solvent eluting from the column can be used for FID detection.[187]

Recently, RPHPLC analyses of TAGs can be performed using mass detectors.[188,189] Mass detection is based on light scattering by solute molecules following nebulization of the column eluate and evaporation of the solvent.[190,191] Mass detection can be used with gradient elution and nonaqueous solvents and does not undergo baseline drift. Composition of the mobile phase, elution rate, nebulizer gas (normally compressed air) flow rate and solvent volatilization temperature all have an important influence on TAG response during mass detection.

Chemical detection by means of post-column reaction derivatization has seldom been used for TAG detection in HPLC. Mass detector also provides valuable information on TAG structure. Chemical ionization mass spectra have successfully identified all the major TAGs together with the sufficiently resolved minor TAGs in various natural fats and oils.

8.6.5 Identification of Molecular Species

One of the most important problems in RPHPLC of natural samples of TAGs is the identification of the molecular species in the peaks on the chromatogram. Wada et al.[167] established a linear relationship between the capacity factor (k) values of chromatographic peaks and the PN values of the TAGs. In addition, aspects of the theoretical prediction and identification of the TAGs in the peaks have been examined.

On the basis of their relative retention times, using an experimental linear relationship between relative retention time and the total number of carbons in saturated TAGs, Herslof et al.[168] estimated theoretically the equivalent carbon number (ECN) for unsaturated TAGs.

This is defined in the same way as the PN (Equation 2) as shown in Equation (3):

$$\mathrm{ECN} = \mathrm{CN} + a'n \tag{3}$$

where n is the number of DBs in the TAG and a' is a constant. The value of the constant a' varies with the chromatographic system. However, it may take on values approaching 2; when $a' = 2$, the PN and ECN have the same value.

The value of a' can be calculated from the linear relationship between $\log k$, CN and n (Equation 4):

$$\log k = a'n + b'\mathrm{CN} + c'n \tag{4}$$

The value of a' is equal to c'/b'.[192]

The parameters PN and ECN have generally been used in RPHPLC to characterize TAG molecules. However, TAGs with the same PN can be differentiated by ECN.

The preceding relationships are calculated for isocratic elution. Consequently the estimation of ECN is not appropriate for gradient elution, as TAGs of very similar molecular characteristics may elute in descending order of ECN. In such cases, it has been proposed that the ECN can be estimated by a second-degree function using CN and n (Equation 5):[193]

$$\mathrm{ECN} = \mathrm{CN} + d'\mathrm{CN}^2 + e'n + f'n^2 \tag{5}$$

Marini and Balestrieri have established a theoretical CN, analogous to the PN and ECN, calculated from the CN and a function depending upon the degree of unsaturation of the FAs.[194]

On analyzing pure standards of simple TAGs and mixed TAGs obtained by transesterification of the former, it is possible to obtain a linear relationship between what is termed equivalent CL (L) and selectivity of the TAGs with respect to triolein (α_0) (Equation 6):[171]

$$L = g' + h' \log \alpha_0 \tag{6}$$

The value of the equivalent CL L for the TAGs in the chromatogram is calculated from the experimental value of α_0 for the peaks; it can then be used to estimate the molar FA composition of the peaks for a natural fat.

The procedure for predicting the TAGs in RPHPLC peaks derived from natural fat based on the ECN is very complicated for large numbers of FA constituents (e.g. more than seven), since the variety of possible molecular species, even taking *sn*-stereospecific positions to be equivalent, is extremely high.

A matrix model using six variables, for the CL and number of DBs for each of the three acids esterifying the glycerol, taking *sn*-stereospecific positions as equivalent, has been developed by Takahashi et al.[192,195,196] and is given in Equation (7):

$$\log k' = i' + j'_1 \mathrm{CL}_1 + j'_2 \mathrm{CL}_2 + j'_3 \mathrm{CL}_3 + z'_1 \mathrm{DB}_1 + z'_2 \mathrm{DB}_2 + z'_3 \mathrm{DB}_3 \tag{7}$$

Applying partition chromatography theory, they found a linear relationship between the log of the relative retention time for TAGs and potential relative retention indices for each constituent FA.

The IUPAC Commission on Oils, Fats and Derivatives undertook development of a method for and a collaborative study in the determination of triglycerides in vegetable oils by liquid chromatography (LC). Three collaborative studies were conducted from 1985 to 1987. Refinements were made in the method after the first collaborative study, and the second and third collaborative studies demonstrated that the method produces acceptable results. Materials studied were soybean oil, almond oil, sunflower oil, olive oil, rapeseed oil and blends of palm and sunflower oils and almond and sunflower oils. Six test samples were analyzed by 18 laboratories from 11 countries in the second study; four test samples were analyzed by 16 laboratories from 12 countries in the third study. The method for determination of trigiycerides (by PN) in vegetable oils by LC was adopted first by AOAC International as an IUPAC–AOC–AOAC method.[197]

Triglycerides in vegetable oils are separated according to ECN by RPHPLC and detected by differential refractometry. Quantification is by area normalization. Elution order is determined by calculating ECN (Equation 2). The ECN can be calculated more accurately by taking the origin of the DBs into account (Equation 8):

$$\mathrm{ECN} = \mathrm{CN} - d_\mathrm{o} n_\mathrm{o} - d_\mathrm{l} n_\mathrm{l} - d_\mathrm{ln} n_\mathrm{ln} \tag{8}$$

where n_o, n_l and n_ln are number of DBs of oleic, linoleic and linolenic acids, respectively, and coefficients d_o, d_l, and d_ln are calculated from reference triglycerides. Under the conditions of this method, ECN approximates (Equation 9):

$$\mathrm{ECN} = \mathrm{CN} - (2.60 \times n_\mathrm{o}) - (2.35 \times n_\mathrm{l}) - (2 - 17 \times n_\mathrm{ln}) \tag{9}$$

With several reference trigiycerides, resolution (α) can be calculated with respect to triolein (Equation 10):

$$\alpha = \frac{\mathrm{RT}'_\mathrm{i}}{\mathrm{RT}'_\mathrm{triolein}} \tag{10}$$

Using reduced a retention time where RT is retention time. $\mathrm{RT}'_\mathrm{i} = \mathrm{RT}_\mathrm{i} - \mathrm{RT}_\mathrm{solvent}$. The retention values for all triglycerides of FAs contained in reference triglycerides (Figure 8) can be determined by a graph of $\log \alpha$ versus n.

8.6.6 Argentation Chromatography

Argentation chromatography separates TAGs according to their degree of unsaturation, unlike RPHPLC in which the separation is more complex and depends on the CL of the constituent FAs in addition to the degree of unsaturation. The basis for TAG separation is the ability of the π-electrons in the DBs on the FA chains to interact with the silver ions on the stationary phase and form stable polar complexes. As the number of DBs in the molecule increases, so does the complex-formation effect and hence retention.[18]

Of all the factors influencing HPLC techniques, the stationary phase is unquestionably the main factor in Ag-TLC. The silver ions may be added to the stationary phase in one of two ways: by impregnating the silica-gel support with a silver salt, normally $AgNO_3$ or by bonding silver ions to the phase by means of an ion-exchange phase.

Silica-gel columns impregnated with $AgNO_3$ were the first to be used. Impregnation with $AgNO_3$ has varied from 5 to 10%. At higher values chromatographic resolution is not improved, and at lower ones the silica-gel adsorption activity increases considerably, resulting in peak tailing.[43,198]

Stationary phases consisting of phenylsulfonic groups chemically bonded to the silica gel appear to be the most advantageous form. Such columns are available in acid or sodium salt form and must first be neutralized by washing with sodium or ammonium hydroxide before treatment with aqueous sodium nitrate.

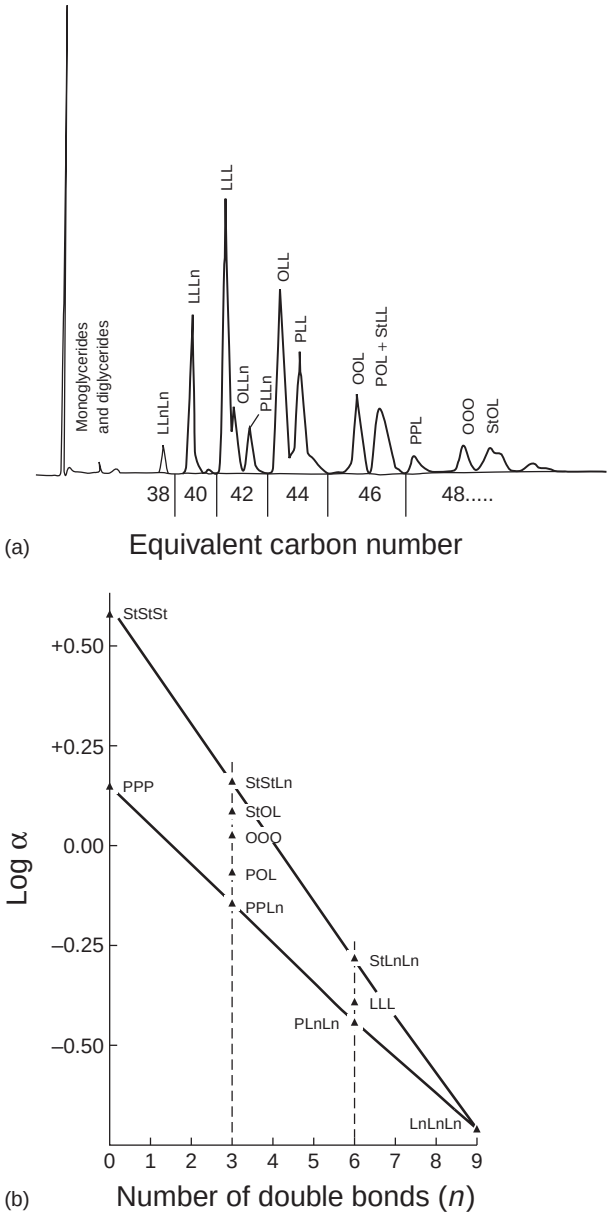

Figure 8 (a) Typical liquid chromatographic separation of soybean oil as a reference material for the identification of triglycerides: O, oleic acid; P, palmitic acid; L, linoleic acid; St, stearic acid; Ln, linolenic acid. Therefore, OOO represents triolein. (b) Retention values from a graph of $\log \alpha$ versus n (Equation 10). The graph determines retention values for all triglycerides of FAs contained in reference triglycerides.[197] (Reproduced with permission from Firestone, *Journal of AOAC International*, **77**, 954–957. Copyright (1994) AOAC International.)

The silver ion content depends upon the number of bonded sulfonic acid groups and the degree of substitution of the initial cation by silver ions.

The composition of the mobile phase exerts a decisive influence on TAG separations because of solute retention on these columns. This occurs by mixed interaction with the more highly polar TAG groups involving the formation of complexes with the silver ions and adsorption by the unbonded polar groups in the support. Consequently, solvents such as benzene, toluene and acetonitrile appear to interact preferentially with the silver ions and interfere with the interactions with the DBs. By comparison, methanol, isopropyl alcohol and acetone may block interactions with the unbonded polar groups on the support.

Mass or light-scattering detectors and moving-flame FID can be used in TAG analysis by argentation HPLC because they do not limit the choice of solvents for the mobile phase, which can be problem with UV detectors and acetone. They also afford good stability and sensitivity when using complex elution gradients with mixtures of three or more solvents.[43]

8.7 Stereospecific Analysis

The physical and chemical properties of fats are related both to FA composition and FA distribution in the TAGs. Consequently, it is necessary not only to identify the acid constituents but also to determine the distribution of those acids in the TAG molecule. This distribution characterizes fats, since there may be several different patterns of distribution on the glyceride molecule for a given FA composition. This change in distribution produces fats with entirely different properties, particularly physical ones.

If we consider the case of a fat composed of only two FAs, A and B, and take positions 1 and 3 on the glycerol molecule to be identical and interchangeable, the two FAs can form six different TAGs (four if no distinction is made between isomers): AAA (A_3), AAB and ABA (A_2B), ABB and BAB (AB_2) and BBB (B_3). With three FA constituents, A, B and C, the number of possible TAGs rises to ten (excluding isomers): A_3, B_3, C_3, A_2B, AB_2, A_2C, AC_2, B_2C, BC_2 and ABC. This rises further to 18 if isomers are included.

The number of possible TAGs climbs rapidly with the number of FA constituents; consequently, the complete analysis of a fat composed of three to five main FAs and a large number of minor FA constituents is practically impossible.

The problem is customarily simplified in various ways, the two most common being to avoid considering isomers and to determine closely related groups of TAGs rather than pure TAGs. The categories or groups established at any given time will depend upon available analytical techniques. Accordingly, initially only four categories of TAGs were determined, based on the number of saturated (S) and unsaturated (U) acyl groups they contained. These groups were designated S_3, S_2U, SU_2 and U_3. With currently existing methods, it is possible to identify additional, more specific categories.

In addition to the preceding simplifications, it is nearly always necessary to establish specific cases, depending upon the analytical method employed, to enable the TAG composition to be calculated from the data obtained.

The goal of analyzing FA distribution is to determine the TAG composition of fats and to develop models of FA distribution on natural glycerides in order to elucidate biosynthetic pathways and thus predict the approximate glyceride composition of a given fat. With advances in analytical techniques, new theories explaining the distribution of acyl remnants have been put forward. Currently, such theories are much more consistent for vegetable fats, synthesized entirely in situ by the plant, than for animal fats, which are partly endogenous and partly exogenous.

The first hypothesis explaining the distribution of FAs in natural TAGs was the theory of the formation of simple TAGs, that is, each chemical species consisting of a single FA only: triolein, tripalmitin, etc. There followed several more theories. The theory of minimal distribution postulates the formation of the smallest possible number of pure TAGs. The rule of even distribution, developed by Hilditch,[199] postulates that FAs are distributed as widely as possible among all TAGs (namely, an acid accounting for 33.3% of the total will be present in all TAGs only once, an acid accounting for between 33.3 and 66.6% will be present between one and two times, and simple TAGs composed of that acid will be formed only when the share contributed by that acid exceeds 66.6%). The random distribution theory holds that FAs are statistically distributed among all the hydroxyl groups on all the glycerol molecules. The random distribution or restricted statistical distribution theory is a modified version of the random distribution theory that was developed by Kartha,[200] to keep the content of the completely saturated glycerides calculated to within the bounds of the experimental range. Finally there are the theories of random distribution at position 2 and random distribution at positions 1 and 3.

The theory of random distribution at positions 1 and 3 was developed independently by Vander Wal[201] and Coleman and Fulton[202] from newer data obtained by hydrolysis using pancreatic lipase and is based on the following assumptions:

- positions 1 and 3 are equivalent
- the FA composition at position 2 and positions 1 and 3 is different and independent
- the acids occupying position 2 as well as those occupying positions 1 and 3 are randomly distributed.

Using these assumptions, it is possible to compute the TAG composition of natural fats from the data on the FA composition at different positions obtained by lipolysis. If the total concentration of an acid, A, in mole % in the TAGs is known and the concentration for that same acid at position 2, A_2, is known, the concentration at positions 1 and 3 will be given by Equation (11):

$$A_1 = A_3 = \frac{3}{2A} - \frac{1}{2A_2} \tag{11}$$

and so on for the remaining acids. From this the molar concentration of the different TAGs can be calculated. AAA is $[A_1][A_2][A_3] \times 10^{-4}$; ABA is $[A_1][B_2][A_3] \times 10^{-4}$; AAB is $2 \times [A_1][A_2][B_3] \times 10^{-4}$; ABC is $2 \times [A_1][B_2][C_3] \times 10^{-4}$; ACB is $2 \times [A_1][C_2][B_3] \times 10^{-4}$; and BAC is $2 \times [B_1][A_2][C_3] \times 10^{-4}$ (where square brackets indicate concentration; i.e. $[A_1]$ is the concentration of A_1).

Enzymatic lipolysis using porcine pancreatic lipase produces glycerol and FAs in a stepwise manner via the intermediate di- and monoglyceridic stages (Equation 12):

$$\begin{aligned} \text{TAGs} &\longrightarrow \text{1,2-Diglycerides} + \text{2,3-Diglycerides} + (\text{FAs}) \\ &\longrightarrow \text{2-Monoglycerides} + (\text{FAs}) \\ &\longrightarrow \text{Glycerol} + (\text{FAs}) \end{aligned} \tag{12}$$

This approach has been applied to a number of vegetable and animal fats and has shown that unsaturated C_{18} acids (oleic, linoleic and linolenic acids) occupy nearly all available position 2 sites, whereas saturated and unsaturated C_{20} and C_{22} acids are present only at positions 1 and 3.

Based on these observations, certain workers have suggested that in vegetable fats position 2 is preferentially acylated by unsaturated acids with 18 carbon atoms and positions 1 and 3 are acylated by the remaining acids and those unsaturated C_{18} acids that are unable to find a free position 2 site.

Within these limitations, acyl groups follow a pattern of statistical or random distribution. Calculating the composition of the basic glyceridic groups, S_3, S_2U, SU_2 and U_3, where S is saturated and U unsaturated (including unsaturated acids with 20 and 22 carbon atoms), shows that the results obtained using random distributions at position 2 and positions 1 and 3 are in better agreement with lipolysis data than the results for the statistical and even distribution theories.

*8.7.1 Stereospecific Analysis of Triacyl-*sn*-glycerols*

HPLC determination of the *sn* positions occupied by FAs on the triacyl-*sn*-glycerol molecule is rather complicated, since most analyses require, first, careful sample preparation and, second, GC analysis of the FAs obtained from the enantiomeric fractions separated by preparatory HPLC.

Sample preparation entails prior partial hydrolysis of the triacyl-*sn*-glycerols to obtain mono- and/or diacyl-*sn*-glycerols, normally by means of the Grignard reaction with ethylmagnesium bromide, followed by formation of derivatives of the mono- and diacyl-*sn*-glycerols and isolation by solid-phase extraction or TLC. HPLC analysis of the mono- and diacyl-*sn*-glycerol derivatives, and in some cases of the derivatives of the triacyl-*sn*-glycerols themselves, has been performed using normal-phase adsorption chromatography, reversed-phase chromatography and chiral-phase chromatography.

ABBREVIATIONS AND ACRONYMS

Ag-TLC	Silver Ion Thin-layer Chromatography
CID	Collision-induced Dissociation
CL	Chain Length
CN	Carbon Number
DB	Double Bond
DNPU	Didinitrophenylurethane
ECL	Equivalent Chain Length
ECN	Equivalent Carbon Number
FA	Fatty Acid
FAME	Fatty Acid Methyl Ester
FID	Flame Ionization Detector
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PN	Partition Number
QqQ	Triple Quadrupole Mass Spectrometer
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RPTLC	Reversed-phase Thin-layer Chromatography
sn	Stereospecifically Numbered
TAG	Triacylglycerol
TLC	Thin-layer Chromatography
UV	Ultraviolet
v	Volume
w/v	Weight/Volume

RELATED ARTICLES

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Liquid Chromatography in Food Analysis • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Chemometrics **(Volume 11)**
Clustering and Classification of Analytical Data

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Chiral Separations by High-performance Liquid Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. W.C. Breckenridge, in *Handbook of Lipid Research*, (ed.), A. Kuksis, Plenum Press, New York, 197, Vol. 1, 1978.
2. H. Hirschmann, 'The Nature of Substrate Asymmetry in Stereoselective reactions', *J. Biol. Chem.*, **235**, 2762–2767 (1960).
3. J.J. Myher, A. Kuksis, 'General Strategies in Chromatographic Analysis of Lipids', *J. Chromatogr.*, **B671**, 3–33 (1995).
4. J. Folch, M. Lees, G.H. Sloane-Stanley, 'A Simple Method for the Determination and Purification of Total Lipids from Animal Tissues', *J. Biol. Chem.*, **226**, 497–509 (1957).
5. E.G. Bligh, W.J. Dyer, 'A Rapid Method of Total Lipids Extraction and Purification', *Can. J. Biochem. Physiol.*, **37**, 911–917 (1959).
6. A. Kuksis, in *Chromatography*, ed. E. Heftman, Elsevier, Amsterdam, B171–B227, 1992.
7. G.J. Nelson, in *Analyses of Fats, Oils and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists' Society, Champaign, IL, 20–59, 1991.
8. N.A. Shaikh, 'Assessment of Various Techniques for the Quantitative Extraction of Lysophospholipids from Myocardial Tissues', *Anal. Biochem.*, **216**, 313–321 (1994).

9. T. Kyrklund, 'Two Procedures to Remove Polar Contaminants from a Crude Brain Lipid Extract by using Prepacked Reversed-phase Columns', *Lipids*, **22**, 274–277 (1987).
10. W.W. Christie, 'Fatty Acids and Lipids: Structure, Extraction and Fractionation into Classes', in *Gas Chromatography and Lipids*, The Oily Press., Ayr, Scotland, 11–42, 1989.
11. K.M. Hosamani, 'A Minor Source of Vernolic, Malvalic and Sterculic Acids in *Phtecollobium dulce* (syn. *Inga dulcis*) Seed Oil', *J. Am. Oil Chem. Soc.*, **72**, 489–492 (1995).
12. S. Ando, M. Saito, in *Chromatography of Lipids in Biomedical Research and Clinical Diagnosis*, ed. A. Kuksis, Elsevier, Amsterdam, 266–310, 1987.
13. R.G. Ackman, in *Analyses of Fats, Oils and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists' Society, Champaign, IL, 270–300, 1991.
14. J.K. Yao, G.M. Rastetter, 'Microanalysis of Complex Tissue Lipids by High-performance Thin-layer Chromatography', *Anal. Biochem.*, **150**, 111–116 (1985).
15. V.P. Pehelkin, A.G. Vereschagin, *Adv. Chromatogr.*, **22**, 87–129 (1992).
16. A.E. Thomas, J.E. Scharoun, H. Ralston, 'Quantitative Estimation of Isomeric Monoglycerides by TLC', *J. Am. Oil Chem. Soc.*, **42**, 789–792 (1965).
17. K.B. Tomer, C.E. Parker, 'Biomedical Application of LC/MS', *J. Chromatogr. B*, **492**, 189 (1989).
18. B. Nikolova-Damyanova, in *Advances in Lipid Methodology One*, ed. W.W. Christie, Oily Press, Ayr, Scotland, 181–237, 1992.
19. C.F. Poole, S.K. Poole, 'Instrumental Thin-layer Chromatography', *Anal. Chem.*, **66**, 27A–37A (1994).
20. T.C. Marquardt, R.F. Wilson, 'An Improved Reversed-phase Thin-layer Chromatography Method for Separation of Fatty Acid Methyl Esters', *J. Am. Oil Chem. Soc.*, **75**, 1889–1892 (1998).
21. V. Ruiz-Gutierrez, E. Montero, J. Villar, 'Determination of Fatty Acid and Triacylglycerol Composition of Human Adipose Tissue', *J. Chromatogr.*, **581**, 171–178 (1992).
22. V. Ruiz-Gutierrez, A. Cert, J.J. Rios, 'Determination of Phospholipid Fatty Acid and Triacylglycerol Composition of Rat Caecal Mucosa', *J. Chromatogr.*, **575**, 1–6 (1992).
23. H.M. Liebich, C. Wirth, B. Jakober, 'Analysis of Polyunsaturated Fatty Acids in Blood Serum after Fish Oil Administration', *J. Chromatogr. B*, **572**, 1–9 (1991).
24. A. Kuksis, L. Marai, J.J. Myher, Y. Itabashi, S. Pind, in *Analyses of Fats, Oils and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists' Society, Champaign, IL, 464–495, 1991.
25. A. Kuksis, J.J. Myher, *Adv. Chromatogr.*, **28**, 267 (1989).
26. B. Truembach, G. Rogler, K.J. Lackner, G. Schmitz, 'Improved Separation of Radioactively Labelled Cellular Phospholipids by HPLC', *J. Chromatogr. B*, **656**, 73–76 (1994).
27. M.P. Purdon, in *Analyses of Fats, Oils and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists' Society, Champaign, IL, 193, 1991.
28. W.C. Picket, C.S. Ramesha, *Meth. Enzymol.*, **187**, 142 (1990).
29. W.W. Christie, in *Advances in Lipid Methodology One*, ed. W.W. Christie, Oily Press, Ayr, Scotland, 239–271, 1991.
30. R.A. Moreau, H.C. Gerard, in *CRC Handbook of Chromatography*, eds. K.D. Mookerjee, N. Weber, CRC Press. Boca Raton, FL, 41, 1993.
31. A. Kuksis, L. Marai, J.J. Myher, Y. Itabashi, S. Pind, in *Analyses of Fats, Oils and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists' Society, Champaign, IL, 214–232, 1991.
32. R.A. Moreau, in *Lipid Chromatographic Analysis*, ed. T. Shibamoto, Plenum Press, New York, 251–272, 1994.
33. M.L. Blank, F. Snyder, in *Lipid Chromatographic Analysis*, ed. T. Shibamoto, Plenum Press, New York, 291–316, 1994.
34. A. Rastegar, A. Pelletier, G. Duportail, L. Freysz, C. Leray, 'Sensitive Analysis of Phospholipid Molecular Species by HPLC Using Fluorescent Naproxen Derivatives of Diacylglycerols', *J. Chromatogr.*, **518**, 157–165 (1990).
35. B.G. Sempore, J.A. Bezard, 'Separation of Monoacylglycerols by Reversed-phase HPLC', *J. Chromatogr.*, **596**, 185–195 (1992).
36. R. Fabien, J.D. Craske, M. Wootton, 'Quantitative Analysis of Synthetic Mixtures of Triacylglycerols with Fatty Acids from Caprylic to Stearic', *J. Am. Oil Chem. Soc.*, **70**, 551–554 (1993).
37. T.A. Foglia, P.D. Vail, T. Iwama, 'HPLC Analysis of 1-Alkyl-2-acyl and 1-Alkyl-3-acyl-sin-glycerols', *Lipids*, **22**, 362–365 (1987).
38. L. Haalck, F. Spener, 'Prediction of Retention Times of Hydroxylated Triacylglycerols in Reversed-phase HPLC', *J. Chromatogr.*, **498**, 410–413 (1990).
39. A.J. Yates, in *Neuromethods 7. Lipids and Related Compounds*, eds. A.A. Bolton, G.B. Baker, L.A. Horrocks, Humana Press, Clifton, NJ, 265–327, 1988.
40. L.Y. Yang, A. Kuksis, J.J. Myher, *Lipids*, **26**, 806 (1991).
41. J. Liu, T. Lee, E. Bobik, Jr, M. Guzman-Harty, C. Hastilow, 'Quantitative Determination of Monoglycerides and Diglycerides', *J. Am. Oil Chem. Soc.*, **70**, 343–347 (1993).
42. E.W. Hammond, *Chromatography for the Analysis of the Lipids*, CRC Press, Boca Raton, FL, 113–168, 1993.
43. B.S.J. Jeffrey, 'Silver Complexation Liquid Chromatography for Fast, High Resolution Separations of Triacylglycerols', *J. Am. Oil Chem. Soc.*, **68**, 289–293 (1991).
44. W.W. Christie, 'Silver-ion Chromatography of Triacylglycerols on Solid Phase Extraction Columns Packed

with a Bonded Sulfonic Acid Phase', *J. Lipid Res.*, **30**, 1471–1473 (1989).
45. P. Laakso, W.W. Christie, 'Combination of Silver Ion and Reversed-phase HPLC in the Fractionation of Herring Oil Triacylglycerols', *J. Am. Oil Chem. Soc.*, **68**, 213–223 (1991).
46. P. Juaneda, J.L. Sebedio, W.W. Christie, 'Complete Separation of the Geometrical Isomers of Linoleic Acid by HPLC with a Silver Ion Column', *J. High Resolut. Chromatogr.*, **17**, 321–324 (1994).
47. A.H. El Hamdy, W.W. Christie, *J. High Resolut. Chromatogr.*, **17**, 55 (1994).
48. T. Takagi, Y. Ando, 'Separation of Monoacylglycerols by HPLC on Nitrile-bonded Phase', *J. Am. Oil Chem. Soc.*, **71**, 459–460 (1994).
49. S.L. Abidi, T.L. Mounts, K.A. Rennick, 'Separation of Major Soybean Phospholipids on β-Cyclodextrin-bonded Silica', *J. Liq. Chromatogr.*, **17**, 3705–3725 (1994).
50. M.J. Wojtusik, P.R. Brown, J.G. Turcotte, 'Separation and Detection of Triacylglycerols by High-performance Liquid Chromatography', *Chem. Rev.*, **89**, 397–406 (1989).
51. K. Aitzetmuller, M. Gronheim, 'Separation of Highly Unsaturated Triacylglycerols by Reversed-phase HPLC with Short Wavelength UV Detector', *J. High Resolut. Chromatogr.*, **15**, 219–226 (1992).
52. E. Frede, H. Thiele, 'Analysis of Milkfat by HPLC', *J. Am. Oil Chem. Soc.*, **64**, 521–528 (1987).
53. L. Marai, J.J. Myher, A. Kuksis, 'Reverse Phase LC/MS of Marine Oil Triacylglycerides', *INFORM*, **3**(4), abstract P6, 500 (1992).
54. G.M. Patton, S.J. Robins, 'HPLC of Molecular Species of Glycerophospholipids in Studies of Lipoproteins and Lipid Transport', in *Journal of Chromatography Library: Chromatography of Lipids in Biomedical Research and Clinical Diagnosis*, ed. A. Kuksis, 311–347, Vol. 37, Elsevier, Amsterdam, 1987.
55. T. Takagi, *Progr. Lipid Res.*, **29**, 277 (1990).
56. Y. Itabashi, T. Takagi, 'High Performance Liquid Chromatographic Separation of Diacylglycerol Enantiomers on a Chiral Stationary Phase', *J. Chromatogr.*, **402**, 257–264 (1987).
57. T. Takagi, Y. Itabashi, 'Rapid Separations of Diacyl- and Diacylglycerol Enantiomers by High Performance Liquid Chromatography on a Chiral Stationary Phase', *Lipids*, **22**, 596–600 (1987).
58. T. Takagi, T. Suzuki, 'Enantiomer Separation of Diacid-diacylglycerol Mixture by HPLC on a Chiral Column', *J. Chromatogr.*, **519**, 237–243 (1990).
59. Y. Itabashi, A. Kuksis, L. Marai, T. Takagi, 'HPLC Resolution of Diacylglycerol Moieties of Natural Triacylglycerols on a Chiral Phase Consisting of Bonded (*R*)-(+)-1-(1-Naphtyl)ethylamine', *J. Lipid Res.*, **31**, 1711–1717 (1990).
60. Y. Itabashi, A. Kuksis, J.J. Myher, 'Determination of Molecular Species of Enantiomeric Diacylglycerols by Chiral Phase High Performance Liquid Chromatography and Polar Capillary Gas–Liquid Chromatography', *J. Lipid Res.*, **31**, 2119–2126 (1990).
61. T. Takagi, Y. Ando, 'Enantiomer Separations of Mixtures of Monoacylglycerol Derivatives by HPLC on a Chiral Column', *Lipids*, **25**, 398–400 (1990).
62. Y. Ando, K. Nishimura, N. Aoyanagi, T. Takagi, 'Stereospecific Analysis of Fish Oil Triacyl-*sn*-glycerols', *J. Am. Oil Chem. Soc.*, **69**, 417–424 (1992).
63. M.G. Wiley, M. Przetakiewicz, M. Takahasi, I. Lowenstein, 'An Extended Method for Separating and Quantifying Molecular Species of Phospholipids', *Lipids*, **27**, 295–301 (1992).
64. J.J. Myher, A. Kuksis, L. Marai, 'Identification of the Less Common Isologous Short-chain Triacylglycerols in the Most Volatile 2.5% Molecular Distillate of Butter Oil', *J. Am. Oil Chem. Soc.*, **70**, 1183–1191 (1993).
65. J.A. Bezard, B.G. Sempore, *INFORM*, **6**, 508 (1995).
66. J.J. Myher, A. Kuksis, S. Pind, 'Molecular Species of Glycerophospholipids and Sphingomyelins of Human Erythrocytes: Improved Method of Analysis', *Lipids*, **24**, 396–407 (1989).
67. T. Oshima, H.-S. Yoon, C. Koizumi, 'Application of Selective Ion Monitoring to the Analysis of Molecular Species of Vegetable Oil Triacylglycerols Separated by Open-tubular Column GLC on a Methylphenylsilicone Phase at High Temperature', *Lipids*, **24**, 535–544 (1989).
68. A. Kemppinen, P. Kalo, 'Fractionation of the Triacylglycerols of Lipase-modified Butter Oil', *J. Am. Oil Chem. Soc.*, **70**, 1203–1207 (1993).
69. W.W. Christie, E.Y. Breechany, K. Stelanov, *Chem. Phys. Lipids*, **46**, 127 (1988).
70. K.L. Busch, G.L. Glish, S.A. Luckey, *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*, VCH Publishers, New York, 1988.
71. J.-L. Le Quere, in *Advances in Lipid Methodology Two*, ed. W.W. Christie, Oily Press, Ayr, Scotland, 215–245, 1993.
72. S.-A. Rossi, J.V. Johnson, R.A. Yost, 'Short Column Gas Chromatography/Tandem Mass Spectrometry for the Detection of Underivatized Anabolic Steroids in Urine', *Biol. Mass Spectrom.*, **23**, 131 (1994).
73. H. Kallio, G. Currie, in *CRC Handbook of Chromatography: Analysis of Lipids*, eds. K.D. Mukherjee, N. Weber, CRC Press, Boca Raton, FL, 435–459, 1993.
74. M.S.F. Lie Ken Jie, C.F. Wong, 'Synthesis and NMR Properties of Positional Isomers of Methyl Allenic Fatty Esters', *Chem. Phys. Lipids*, **61**, 243–254 (1992).
75. H.B.S. Conacher, F.D. Gunstone, G.M. Hornby, F. Padley, 'Glyceride Studies: Part IX: Intraglyceride Distribution of Vernolic Acid and of Five Conjugated Octadecatrienoic Acids in Seed Glycerides', *Lipids*, **5**, 434–441 (1970).

76. O. Suzuki, T. Hashimoto, K. Hayamizu, O. Yamamoto, 'Studies on the Nuclear Magnetic Resonance Spectra of Olefinic Protons of Conjugated Fatty Acid Methyl Esters', *Lipids*, **5**, 457–462 (1970).
77. J.R. Vikery, F.B. Whitfield, G.L. Ford, B.H. Kenneth, 'Xymenynic Acid in *Santalum obtusifolium* Seed Oil', *J. Am. Oil Chem. Soc.*, **61**, 890–891 (1984).
78. F.T. Longone, A.H. Miller, 'Shielding Effects in Substituted Cyclopropanes', *Chem. Commun.*, 447–448 (1967).
79. N.E. Pawlowski, J.E. Nixon, R.O. Sinnhuber, 'Assay of Cyclopropenoid Lipids by Nuclear Magnetic Resonance', *J. Am. Oil Chem. Soc.*, **49**, 387–392 (1972).
80. C.Y. Hopkins, 'Occurrence of Cypentenyl Fatty Acids', *J. Am. Oil Chem. Soc.*, **45**, 778–783 (1968).
81. M. Pollard, 'Nuclear Magnetic Resonance Spectroscopy (High Resolution)', in *Analysis of Oils and Fats*, eds. R.J. Hamilton, J.B. Rossell, Elsevier, London, 401–434, 1986.
82. F. Comes, M. Farines, A. Aumelas, J. Soulier, 'Fatty Acids and Triacylglycerols of Cherry Seed Oil', *J. Am. Oil Chem. Soc.*, **69**, 1224–1227 (1992).
83. R.L. Sacchi, E.P. Medina, S.P. Aubourg, P. Addeo, L. Paolillo, 'Proton Nuclear Magnetic Resonance Rapid and Structure-specific Determination of Omega-3 Polyunsaturated Fatty Acid in Fish Lipids', *J. Am. Oil Chem. Soc.*, **70**, 225–228 (1993).
84. M. Aursand, J.R. Rainuzzo, H. Grasdalen, 'Quantitative High-resolution ^{13}C and ^{1}H NMR of Omega-3 Fatty Acids from White Muscle of Atlantic Salmon', *J. Am. Oil Chem. Soc.*, **70**, 971–981 (1993).
85. V. Spitzer, W. Tomberg, M. Zucolotto, 'Occurrence of α-parinaric Acid in the Seed Oil of *Sebastiana brasilensis* (Euphorbiaceae)', *J. Am. Oil Chem. Soc.*, **73**, 569–573 (1996).
86. V. Spitzer, K. Aitzetmuller, K. Vosman, 'The Seed Oil of *Bernardia pulchella* (Euphorbiaceae) – A Rich Source of Vernolic Acid', *J. Am. Oil Chem. Soc.*, **73**, 1733–1735 (1996).
87. F.D. Gunstone, 'High Resolution ^{13}C NMR Spectroscopy of Lipids', in *Advances in Lipid Methodology Two*, ed. W.W. Christie, Oily Press, Dundee, Scotland, 1–68, 1993.
88. R. Sacchi, M. Pattumi, G. Fontanazza, P. Barone, P. Fiordiponti, L. Mannina, E. Rossi, A.L. Segre, 'A High Field ^{1}H Nuclear Magnetic Resonance Study of the Minor Components in Virgin Olive Oils', *J. Am. Oil Chem. Soc.*, **73**, 747–758 (1996).
89. J. Vicanova, E. Tvrzicka, K. Stulik, 'Capillary Gas Chromatography of Underivatized FA with a Free FA Phase Column and a Programmed Temperature Vaporizer Injector', *J. Chromatogr. B*, **656**, 45–50 (1994).
90. L.D. Metcalfe, A.A. Schmitz, J.R. Pelka, 'Rapid Preparation of Fatty Acid Esters from Lipids for Gas Chromatographic Analysis', *Anal. Chem.*, **38**, 514–515 (1966).
91. P.C. Barnes, C.E. Holaday, 'Rapid Preparation of Fatty Acid Esters Directly from Ground Peanuts', *J. Chromatogr. Sci.*, **10**, 181–183 (1972).
92. G. Lepage, C.C. Roy, 'Direct Transesterification of All Classes of Lipids in a One-step Reaction', *J. Lipid Res.*, **27**, 114–120 (1986).
93. W. Butte, 'Rapid Method for the Determination of Fatty Acid Profiles from Fats and Oils Using Trimethylsulfonium Hydroxide for Transesterification', *J. Chromatogr.*, **261**, 142–145 (1983).
94. S.W. Christophersen, R.L. Glass, 'Preparation of Milk Fat Methyl Esters by Alcoholysis in an Essentially Non Alcoholic Solution', *J. Dairy Sci.*, **5**, 1289–1290 (1969).
95. J.L. Iverson, A.J. Sheppard, 'Butyl Ester Preparation for Gas–Liquid Chromatographic Determination of Fatty Acids in Butter', *J. Assoc. Anal. Chem.*, **60**, 284–288 (1977).
96. C.D. Bannon, G.J. Breen, J.D. Craske, N.T. Hai, N.L. Harper, K.L. O'Rourke, 'Analysis of Fatty Acid Methyl Esters with High Accuracy and Reliability. Part. III – Literature Review and Investing into the Development of Rapid Procedures for the Methoxide-catalyzed Methanolysis of Fats and Oils', *J. Chromatogr.*, **247**, 71–89 (1982).
97. H.T. Badings, C. de Jong, 'Glass Capillary Gas Chromatography of FAME. A Study of Conditions for the Quantitative Analysis of Short- and Long-chain FA in Lipids', *J. Chromatogr.*, **27**, 493–506 (1983).
98. K. Grob, *Classical Split and Splitless Injection in Capillary Gas Chromatography*, Huethig, Heidelberg, 1988.
99. G. Schomburg, *Gas Chromatography*, VCH, Weinheim, 1990.
100. K. Grob, *On-column Injection in Capillary Gas Chromatography*, Huethig-Verlag, Heidelberg, 1987.
101. F. Poy, S. Viusani, F. Terrosi, 'A Universal Sample Injection System for Capillary Column GC using a Programmed Temperature Vaporizer (PTV)', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **5**, 355–359 (1982).
102. F. Poy, L. Cobelli, 'Quantitative Aspects of the Programmed Temperature Vaporization Technique of Sample Introduction in Parallel Capillary Column and Microbore Capillary Column Gas Chromatography', *J. Chromatogr.*, **279**, 689–694 (1983).
103. W.M.N. Ratnayake, R. Hollywood, E. O'Grady, J.L. Beare-Rogers, 'Determination of *cis*- and *trans*-Octadecenoic Acids in Margarines by Gas Liquid Chromatography–Infrared Spectrometry', *J. Am. Oil Chem. Soc.*, **67**, 804–810 (1990).
104. J.J. Myher, A. Kuksis, L.-Y. Yang, *Biochem. Cell Biol.*, **68**, 336 (1990).
105. W.C. Willett, M.J. Stampfer, J.E. Manson, G.A. Colditz, F.E. Speizer, B.A. Rosner, L.A. Sampson, C.A. Hennekens, *Lancet*, **341**, 581 (1993).

106. R.L. Wolff, 'Resolution of Linolenic Acid Geometrical Isomers by a GLC on a Capillary Column Coated with a 100% Cyanopropyl Polysiloxane Film (CP™ Sil 88)', *J. Chromatogr. Sci.*, **30**, 17–22 (1992).
107. N.C. Shantha, G.E. Napolitano, 'Gas Chromatography of Fatty Acids', *J. Chromatogr.*, **624**, 37–51 (1992).
108. A. Kuksis, in *Lipid Chromatographic Analysis*, ed. T. Shibamoto, Marcel Dekker, New York, 177–222, 1994.
109. E. Geeraert, P. Sandra, D. Schepper, 'On-column Injection in the Capillary Gas Chromatographic Analysis of Fats and Oils', *J. Chromatogr.*, **279**, 287–295 (1983).
110. J. Molkentin, D. Precht, 'Comparison of Packed and Capillary Columns for Quantitative Gas Chromatography of Triglycerides in Milk Fat', *Chromatographia*, **39**, 265–270 (1994).
111. G. van der Steege, F.A.J. Muskiet, I.A. Martini, N.H. Hutter, E.R. Boersma, 'Simultaneous Quantification of Total Medium- and Long-chain Fatty Acids in Human Milk by Capillary Gas Chromatography with Split Injection', *J. Chromatogr. B*, **415**, 1–11 (1987).
112. F.A.J. Muskiet, J.J. van Doormaal, I.A. Martini, B.G. Wolthers, W. van der Slik, 'Capillary Gas Chromatography Profiling of Total Long Chain Fatty Acids and Cholesterol in Biological Materials', *J. Chromatogr.*, **278**, 231–244 (1983).
113. M. Volmer, G. Meiborg, F.A.J. Muskiet, 'Simultaneous Capillary Gas Chromatography Profiling of Medium- and Long-chain Fatty Acid Methyl Esters with Split Injection. Collection for Injection Related Discrimination by the Bracketing Method', *J. Chromatogr.*, **434**, 385–394 (1988).
114. R.G. Ackman, W.M.N. Ratnayake, in *The Role of Fats in Human Nutrition*, eds. A.J. Vergroesen, M. Crawford, Academic Press, London, 441, Chapter 11, 1989.
115. E. Kovats, 'Gas-chromatographische Charakterisierung Organischer Verbindungen. Teil 1: Retentionsindices Aliphatischer Halogenide, Alkohole, Aldehyde und Ketone', *Helv. Chim. Acta*, **41**, 1915–1932 (1958).
116. A. Wehrli, E. Kovats, 'Gas-chromatographische Charakterisierung Organischer Verbindungen. Teil 3: Berechnung der Retentionsindices Aliphatischer, Alicyclischer und Aromatischer Verbindungen', *Helv. Chim. Acta*, **42**, 2709–2736 (1959).
117. E. Kovats, 'Gas Chromatographic Characterization of Organic Substances in the Retention Index System', *Adv. Chromatogr.*, **1**, 229–247 (1965).
118. H.T. Slover, E. Lanza, 'Quantitative Analysis of Food Fatty Acids by Capillary GC', *J. Am. Oil Chem. Soc.*, **56**, 933–943 (1979).
119. K. Eder, A.M. Reichlmayr-Lais, M. Kirchgessner, 'Simultaneous Determination of Amounts of Major Phospholipid Classes and their Fatty Acid Composition in Erythrocyte Membranes Using HPLC and GC', *J. Chromatogr.*, **598**, 33–42 (1992).
120. T.L. Hearn, S.A. Sgoutas, J.A. Hearn, D.S. Sgoutas, 'Polyunsaturated Fatty Acids and Fat in Fish Flesh for Selective Species for Health Benefits', *J. Food Sci.*, **52**, 1209–1211 (1987).
121. H. Heckers, F.W. Melcher, 'Isolation of *trans*-Hexadecanoic and *trans*-Octadecanoic Fatty Acid Methyl Esters from Lipids Extracts by Means of Argentation and Reversed-phase Thin-layer Chromatography', *J. Chromatogr.*, **256**, 185–189 (1983).
122. A. Shibahara, K. Yamamoto, K. Shinkai, T. Nakayama, G. Kajimoto, *Biochim. Biophys. Acta*, **1170**, 245 (1993).
123. M.J. Aveldano, M. van Rollins, L.A. Horrocks, 'Separation and Quantitation of Free FA and FAME by RP-HPLC', *J. Lipid Res.*, **24**, 83–93 (1983).
124. P. Schick, C. Levey, *J. Planar Chromatogr.*, **3**, 269 (1990).
125. V.K.S. Shukla, K.C. Srivastava, 'Argentation TLC of Polyunsaturated Fatty Acid Methyl Esters', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **1**, 214–218 (1978).
126. R. Wilson, J.R. Sargent, 'High Resolution Separation of Polyunsaturated Fatty Acids by Argentation Thin-layer Chromatography', *J. Chromatogr.*, **623**, 403–407 (1992).
127. M.M. Mossoba, R.E. McDonald, A.R. Prosser, 'Gas Chromatographic/Matrix Isolation/Fourier Transform Infrared Spectroscopic Determination of *trans*-Monounsaturated and Saturated Fatty Acid Methyl Esters in Partially Hydrogenated Menhaden Oil', *J. Am. Oil Chem. Soc.*, **41**, 1998–2002 (1993).
128. R.G. Ackman, E.J. MacPherson, 'Coincidence of *cis*- and *trans*-Monoethylenic Fatty Acids Simplifies the Open-tubular Gas–Liquid Chromatography of Butyl Esters of Butter Fatty Acids', *Food Chem.*, **50**, 45–52 (1994).
129. D. Chobanov, R. Tarandjiiska, B. Nikolova-Damyanova, 'Quantification of Unsaturated Fatty Acids by Argentation Thin Layer Chromatography', *J. Planar Chromatogr.*, **5**, 157–163 (1992).
130. M. Hamberg, W.H. Gerwick, P.A. Asen, 'Linoleic Acid Metabolism in the Red Alga *Lithothamnion corolloides*: Biosynthesis of 11(*R*)-Hydroxy-9(*Z*),12(*Z*)-octadecadienoic acid', *Lipids*, **27**, 487–493 (1992).
131. R. Bortolomeazzi, L. Pizzale, G. Lercker, 'Chromatographic Determination of the Position and Configuration of Isomers of Methyl Oleate Hydroperoxides', *J. Chromatogr.*, **626**, 109–116 (1992).
132. H.K. Mangold (ed.), *CRC Handbook of Chromatography: Lipids*, CRC Press, Boca Raton, FL, 387–397, Vol. 1, 1984.
133. W.W. Christie, *Lipid Analysis*, 2nd edn, Pergamon Press, Oxford, 144–145, 1982.
134. W.W. Christie, *High-performance Liquid Chromatography and Lipids: A Practical Guide*, Pergamon Press, Oxford, 183–208, 1987.
135. B. Nikolova-Damyanova, B. Amidzhin, 'Separation of Triglyceride Groups by Reversed-phase Thin-layer

Chromatography on Silanized Kieselguhr', *J. Chromatogr.*, **446**, 283–291 (1988).
136. C.B. Barret, M.S.J. Dallas, F.B. Padley, 'The Separation of Glycerides by Thin Layer Chromatography on Silica Impregnated with Silver Nitrate', *Chem. Ind. (London)*, 1050–1051 (1962).
137. R. Tarandjiiska, H. Nguyen, 'Triglyceride Analysis of Orange Seed Oil by Argentation Thin-layer Chromatography', *Riv. Ital. Sost. Grasse*, **65**, 489–492 (1988).
138. J.N. Roehm, O.S. Privett, 'Changes in the Structure of Soybean Triglycerides During Maturation', *Lipids*, **5**, 353–358 (1970).
139. M.H. Jee, A.S. Ritchie, 'Separation of Certain Triglyceride Isomers by Argentation Thin-layer Chromatography with Flame Ionisation Detection by the Iatroscan TH 10', *J. Chromatogr.*, **370**, 214–217 (1986).
140. J. Sherma, *TLC Technical Series 1*, Whatman, Clifton, NJ, 6, 1987.
141. B. Nikolova, H. Nguyen, D. Chobanov, S. Dimov, in *Proceedings FECS 3rd International Congress of Chemistry and Biotechnology of Biologically Active Natural Products 1985*, VCH, Weinheim, 131, Vol. 5, 1987.
142. B. Nikolova-Damyanova, D. Chobanov, S. Dimov, *J. Liq. Chromatogr.*, **16**, 3997 (1993).
143. H.P. Kaufmann, Z. Makus, B. Das, 'Die Papier-chromatographies Auf Dem Fettgebiet XLIV. Mitteilung: Die Pc-Analyse Der Glyceride', *Fette Seifen Anstrichm.*, **63**, 125–132 (1961).
144. B. Fried, J. Sherma, *Thin-layer Chromatography Techniques and Applications*, Marcel Dekker, New York, 1994.
145. C. Masterson, B. Fried, J. Sherma, *J. Liq. Chromatogr.*, **15**, 2967–2980 (1992).
146. M.A. Kaluzny, L.A. Duncan, M.V. Merrit, D.E. Epps, 'Rapid Separation of Lipid Classes in High Yield and Purity using Bonded Phase Columns', *J. Lipid Res.*, **26**, 135–140 (1985).
147. J.J. Agren, A. Julkunen, I. Penttila, 'Rapid Separation of Serum Lipids for Fatty Acid Analysis by a Single Aminopropyl Column', *J. Lipid Res.*, **32**, 1871–1876 (1992).
148. A.J. Fraser, D.R. Tocher, J.R. Sargent, 'TLC/FID Detection and the Quantitation of Marine Neutral Lipids and Phospholipids', *J. Exp. Mar. Biol. Ecol.*, **88**, 91–99 (1985).
149. J.K.G. Kramer, R.C. Fouchard, E.R. Farnworth, 'The Effect of Fatty Acid Chain Length and Unsaturation on the Chromatographic Behavior of Triglycerides on Iatrosca', *Lipids*, **20**, 617–619 (1985).
150. J.K.G. Kramer, B.K. Thomson, E.R. Farnworth, 'Variation in the Relative Response Factor for Triglycerides on Iatroscan Chromarods with Fatty Acid Composition and Sequence of Analysis', *J. Chromatogr.*, **355**, 221–228 (1986).
151. T. Oshima, W.M.N. Ratnayake, R.G. Ackman, 'Cod Lipids, Solvent Systems and the Effect of Fatty Acid Chain Length and Unsaturation on Lipid Class Analysis by Iatroscan TLC/FID', *J. Am. Oil Chem. Soc.*, **64**, 219–223 (1987).
152. P. Mares, T. Rezanka, J. Novak, 'Analysis of Human Blood Triacylglycerols Using Capillary Gas Chromatography, Silver Ion-thin Layer Chromatography Fractionation and Desorption Chemical Ionization–Mass Spectrometry', *J. Chromatogr. B*, **568**, 1–10 (1991).
153. N.C. Shantha, R.G. Ackman, 'Advantages of Total Lipid Hydrogenation Prior to Lipid Class Determination on Chromarods-SIII', *Lipids*, **25**, 570–574 (1990).
154. T. Yasumoto, B. Undertal, T. Aune, B. Hormazabal, O.M. Skulberg, Y. Oshima, in *Proceedings of the 4th International Conference on Toxic Marine Phytoplankton*, Lund, June, 1989, eds. E. Graneli, B. Sundstrom, L. Edler, D.M. Anderson, Elsevier, New York, 436, 1989.
155. C.C. Parrish, A.S.W. de Freitas, C. Bodennec, E.J. Macpherson, R.G. Ackman, 'Lipid Composition of the Toxic Marine Diatom, Nitzschia Pungens', *Phytochemistry*, **30**, 113–116 (1991).
156. S. Chemin-Douhaud, A. Karleskind, 'Application des Enzymes a l'Analyse des Lipides', *Rev. Fr. Corps Gras*, **34**, 19–25 (1987).
157. IUPAC, *Pure Appl. Chem.*, **57**, 899 (1985).
158. S. Chemin-Douhaud, A. Karleskind, 'Determination Rapide de la Repartition en Mono, di et Triglycerides des Huiles et Corps Gras, par Dosage Enzymatique apres Fractionnement par Chromatographie sur Couche Mince', *Rev. Fr. Corps Gras*, **29**, 125–128 (1982).
159. F.H. Fryer, W.L. Ormand, G.B. Crump, 'Triglyceride Elution by Gas Chromatography', *J. Am. Oil Chem. Soc.*, **37**, 589–590 (1960).
160. V.R. Huebner, 'The Analysis of Glycerides by High Temperature Gas–Liquid Partition Chromatography', *J. Am. Oil Chem. Soc.*, **37**, 628–631 (1961).
161. A. Kuksis, M.J. McCarthy, 'Gas–Liquid Chromatographic Fractionation of Natural Triglyceride Mixtures by Carbon Number', *Can. J. Biochem. Physiol.*, **40**, 679–686 (1962).
162. E. Geeraert, P. Sandra, 'Capillary Gas Chromatography of Triacylglycerols in Fats and Oils Using High Temperature Phenylmethylsilicone Stationary Phase. I', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **8**, 415–422 (1985).
163. Y. Itabashi, T. Takagi, 'Gas Chromatographic Resolution on Polar Open Tubular Column of Saturated and Non Saturated Wax Esters Isomers Differing in Combinations of Acyl and Alcoholic Groups', *J. Chromatogr.*, **299**, 351–363 (1984).
164. J.J. Myher, A. Kuksis, L. Marai, P. Sandra, 'Identification of the more Complex Triacylglycerols in Bovine Milk Fat by Gas Chromatography–Mass Spectrometry Using Polar Capillary Columns', *J. Chromatogr.*, **452**, 93–118 (1988).

165. J. Ruiz-Gutiérrez, A. Cert, J.J. Rios, 'Determination of Phospholipid Fatty Acids and Triacylglycerols Composition of Rat Caecal Mucosa', *J. Chromatogr.*, **575**, 1–6 (1992).
166. E. Geeraert, P. Sandra, in *Chromatography of Lipids in Biomedical Research and Clinical Diagnosis*, ed. A. Kuksis, Elsevier Science, Amsterdam, 59, Chapter 2, 1987.
167. S.A. Wada, C. Koizumi, J. Nonaka, 'Analysis of Triglycerides of Soybean Oil by High Performance Liquid Chromatography in Combination with GC', *Yukagaku*, **26**, 95–99 (1977).
168. B. Herslof, O. Podlaha, B. Toregard, 'HPLC of Triglycerides', *J. Am. Oil Chem. Soc.*, **56**, 864–866 (1979).
169. A.H. El Hamdy, E.G. Perkins, 'High Performance Reversed Phase Chromatography of Natural Triglyceride Mixtures: Critical Pair Separation', *J. Am. Oil Chem. Soc.*, **58**, 867–872 (1981).
170. J.P. Goiffon, C. Reminiac, D. Furon, 'Application de la Chromatographie Liquide Haute Performance a l'Analyse des Triglycerides des Corps Gras', *Rev. Fr. Corps Gras*, **28**, 199–207 (1981).
171. L.J.R. Barron, M.T.G. Hierro, G. Santa-Maria, *J. Dairy Res.*, **57**, 517–526 (1990).
172. E. Frede, 'Improved HPLC of Triglycerides by Special Tempering Procedures', *Chromatographia*, **21**, 29–36 (1986).
173. W.P. Edwards, 'The Use of HPLC to Separate TGs in Confectionery Fats using UV Detection', *J. Liq. Chromatogr.*, **8**, 2625–2640 (1985).
174. L.J.R. Barron, G. Santa-Maria, 'Non Aqueous Reverse-phase HPLC Analysis of Triglycerides', *Chromarographia*, **23**(3), 209–214 (1987)
175. K. Aitzetmuller, M. Gronheim, 'Gradient Elution HPLC of Fats and Oils with Light Scattering Detection', *Fat Sci. Technol.*, **5**, 164–168 (1993).
176. Y. Lozano, 'Analyse des Triglycerides de l'Huile d'Avocat par CLHP en Phase Inverse', *Rev. Fr. Corps Gras*, **30**, 333–346 (1983).
177. K. Aitzetmuller, H. Wessels, G. Werner, *Fette Wiss. Technol.*, **90**, 442–447 (1988).
178. L.J.R. Barron, M.V. Cela, G. Santa-Maria, 'Determination of the Triglyceride Composition of Grapes by HPLC', *Chromatographia*, **25**(7), 609–612 (1988).
179. M.A.N. Zeitoun, W.E. Neff, E. Selke, T.L. Mounts, 'Analyses of Vegetable Triglycerides Molecular Species by Reversed Phase High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **14**, 2685–2698 (1991).
180. J.L. Perrin, A. Prevot, 'Utilisation d'un Detecteur a Diffusion de la Lumiere Laseer dans l'Etude des Corps Gras par CLHP II: Analyse des Triglycerides des Huiles et des Graisses', *Rev. Fr. Corps Gras*, **33**, 437–445 (1986).
181. A.J. Palmer, F.J. Palmer, 'Rapid Analysis of Triacylglycerols Using High-performance Liquid Chromatography with Light Scattering Detection', *J. Chromatogr.*, **465**, 369–377 (1989).
182. C. Horvath, W. Melander, 'Liquid Chromatography with Hydrocarbonaceous Bonded Phases; Theory and Practice of Reversed-phase Chromatography', *J. Chromat. Sci.*, **15**, 393–404 (1977).
183. C. Maniongui, J. Gresti, M. Bugaut, S. Gauthier, J. Bezard, 'Determination of Bovine Butterfat Triacylglycerols by Reversed-phase Liquid Chromatography and Gas Chromatography', *J. Chromatogr.*, **543**, 81–103 (1991).
184. R.D. Plattner, 'High Performance Liquid Chromatography of Triglycerides: Controlling Selectivity with Reverse Phase Columns', *J. Am. Oil Chem. Soc.*, **58**, 638–642 (1981).
185. B. Herslof, 'HPLC of Triacylglycerols Using UV Detection', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **4**, 471–473 (1981).
186. J.L. Robinson, M. Tsimidou, R. Macrae, 'Evaluation of the Mass Detector for Quantitative Detection of Triglycerides and Fatty Acid Methyl Esters', *J. Chromatogr.*, **324**, 35–51 (1985).
187. H.V. Drushel, 'Needs of the Chromatographer. Detectors', *J. Chromatogr. Sci.*, **21**, 375–384 (1983).
188. M.F. Caboni, L.S. Conte, G. Lercker, 'Rapid HPLC Analysis of Triacylglycerols by Isocratic Elution and Light Scattering Detection', *Ital. J. Food Sci.*, **4**, 125–132 (1992).
189. A.S. McGill, C.F. Moffat, *Lipids*, **27**, 360–370 (1992).
190. J.M. Charlesworth, 'Evaporative Analyzer as a Mass Detector for Liquid Chromatography', *Anal. Chem.*, **50**, 1414–1420 (1978).
191. A. Stolyhwo, H. Colin, G. Guiochon, 'Use of Light Scattering as a Detector Principle in Liquid Chromatography', *J. Chromatogr.*, **265**, 1–18 (1983).
192. K. Takahashi, T. Hirano, K. Egi, K. Zama, 'A Mathematical Model for the Prediction of Triglyceride Molecular Species by High Performance Liquid Chromatography', *J. Am. Oil Chem. Soc.*, **62**, 1489–1492 (1985).
193. M.T.G. Hierro, M.C. Tomás, F. Fernández-Martìn, G. Santa-Maria, 'Determination of the Triglyceride Composition of Avocado Oil by High-performance Liquid Chromatography Using a Light Scattering Detector', *J. Chromatogr.*, **607**, 329–338 (1992).
194. D. Marini, F. Balestrieri, 'Analisi di Miscele di Trigliceridi Mediante HPLC con Rivelatore UV', *Riv. Sost. Grasse*, **64**, 11–16 (1989).
195. K. Takahashi, T. Hirano, K. Zama, 'A New Concept for Determining Triglyceride Composition of Fats and Oils by Liquid Chromatography', *J. Am. Oil Chem. Soc.*, **61**, 1226–1229 (1984).
196. K. Takahashi, T. Hirano, K. Egi, M. Hatano, K. Zama, 'Supplementary Consideration of the Triglyceride Matrix Model on Reverse Phase HPLC', *J. Am. Oil Chem. Soc.*, **63**, 1543–1546 (1986).
197. D. Firestone, 'Liquid Chromatographic Method for Determination of Triglycerides in Vegetable Oils in Terms of their Partition Numbers: Summary

of Collaborative Study', *JAOAC Int.*, **77**, 954–957 (1994).
198. H. Kallio, T. Vauhkonen, R.R. Linko, 'Thin-layer Silver ion Chromatography and Supercritical Fluid Chromatography of Baltic Herring (*Clupea harengus membras*) Triacylglycerols', *J. Agric. Food Chem.*, **39**, 1573–1577 (1991).
199. T.P. Hildich, P.N. Williams, *The Chemical Constitution of Natural Fats*, 4th edn, Chapman and Hall, London, 378, 1964.
200. A.R.S. Kartha, 'Biosynthesis of Fat in Ripening Coconut: The ''in vivo'' Quantum Synthesis of Triglycerides', *J. Sci. Food Agric.*, **15**, 299–302 (1964).
201. R.J. Vander Wal, 'Calculation of the Distribution of the Saturated and Unsaturated Acyl Groups in Fats from the Pancreatic Lipase Hydrolysis Data', *J. Am. Oil Chem. Soc.*, **37**, 18–20 (1960).
202. M.H. Coleman, W.C. Fulton, *Enzymes of Lipid Metabolism*, Pergamon Press, New York, 127, 1961.

Liquid Chromatography in Food Analysis

L.M.L. Nollet
Hogeschool Gent, Belgium

1 Introduction 1
2 High-performance/Pressure Liquid Chromatography 1
3 Sample Preparation and Derivatization 2
3.1 Sample Preparation 2
3.2 Derivatization 2
4 Nutrient Analysis 2
4.1 Amino Acids 2
4.2 Peptides 3
4.3 Proteins 3
4.4 Lipids and Phospholipids 4
4.5 Carbohydrates 4
4.6 Alcohols 5
4.7 Vitamins 5
4.8 Organic Acids 7
4.9 Organic Bases 7
4.10 Phenolic Compounds 8
4.11 Bittering Substances 8
4.12 Pigments 8
5 Residues 9
5.1 Mycotoxins 9
5.2 Residual Antibacterials 9
5.3 Residues of Growth Promoters 9
5.4 Pesticides 10
5.5 *N*-Nitroso Compounds 11
6 Miscellaneous 11
6.1 Colorants 11
6.2 Preservatives 11
6.3 Synthetic Antioxidants 11
6.4 Sweeteners 12
7 Anions and Cations 12
Abbreviations and Acronyms 12
Related Articles 13
Further Reading 13
References 13

This article gives a general overview of the possibilities for use of the HPLC (high-performance liquid chromatography) technique in food analysis. Different methods of sample preparation, cleanup and, if necessary, derivatization techniques are discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

A general overview of the possibilities for using the HPLC technique in food analysis is presented. Because the subject is extensive this article can only portray generalities. For further information the reader should consult the referenced books and articles.

This article begins by discussing various methods of sample preparation, cleanup and, if necessary, derivatization techniques. In subsequent sections the separation, analysis, and detection of different nutrients, residues, and additives in foods are treated.

2 HIGH-PERFORMANCE/PRESSURE LIQUID CHROMATOGRAPHY

HPLC (high-pressure or high-performance liquid chromatography) instruments comprise different parts: solvent reservoirs, pumping units for isocratic or gradient analysis, injection devices, columns and detectors.

In adsorption or normal-phase liquid chromatography (NPLC) the stationary phase (packings) is polar: silica, alumina or polar-bonded silicas (cyanopropyl-bonded, aminopropyl-bonded or diol-bonded silicas). Different solvents are used as mobile phases such as pentane and hexane.

In reversed-phase liquid chromatography (RPLC) the stationary phase (C_8, C_{18}, phenyl bonded) is apolar. An alkyl group is bonded onto silica gel. A wide variety of stationary phases exist. The mobile phase is hydrophilic. Mixtures of water and organic modifiers such as methanol, acetontrile, and tetrahydrofuran are used.

Three modes exist in ion chromatography (IC):

- Ion exchange – Cation or anion exchangers separate samples by their charges. Cation and anion exchangers are either strong (SCX, strong cation exchanger and SAX, strong anion exchanger) or weak (WCX, weak cation exchanger and WAX, weak anion exchanger).
- Ion exclusion – Nonionic compounds are separated on an ion-exchange packing through differences in acid strength, size, and hydrophobicity.
- Ion pairing – This is a combination of a RP (reversed-phase)-packing with a mobile phase of an aqueous organic mixture and an ion-pairing agent. The agent forms an ion pair with the analyte with increased lipophilicity.

Size-exclusion chromatography (SEC) separates molecules by their size. Synonyms are gel permeation chromatography (GPC) and gel filtration chromatography (GFC).

A wide choice of special phases exist for chiral or enantiomeric separations. Silver IC separates positional and configurational isomers of fatty acids.

In isocratic elution the composition of the mobile phase is constant; in gradient elution the solvent strength is increased.

Hyphenated techniques consist of coupling different techniques such as liquid chromatography (LC) with mass spectrometry (MS) or with nuclear magnetic resonance (NMR).

Detectors monitor the samples. Photometric detectors are based on UV/VIS (ultraviolet/visible) absorption, fluorescence or chemiluminescence. Photodiode array (PDA) and light scattering are other types of photometric detectors.

RI detects the difference between the refractive index of the sample and the reference. Electrochemical (EC) detectors measure concentrations.

3 SAMPLE PREPARATION AND DERIVATIZATION

3.1 Sample Preparation

Two techniques frequently used in sample preparation are solid-phase extraction (SPE) and supercritical fluid extraction (SFE). Another technique, LLE (liquid–liquid extraction), extracts compounds from an aqueous solution in a water-immiscible solvent. In SPE, analytes are absorbed and eluted according to their different affinities between the packing material and the mobile phase in a disposable column. On-line and off-line devices exist. In SFE, carbon dioxide is used to solubilize analytes. Important parameters are time and pressure.

3.2 Derivatization

Some molecules do not have UV (ultraviolet) absorbance characteristics, fluorescent or EC properties. Derivatization makes molecules detectable by adding a chromophore, fluorophore or electroactive product. Derivatization is done on-line or off-line. Precolumn derivatization is reaction before separation and postcolumn derivatization is reaction after separation.

4 NUTRIENT ANALYSIS

4.1 Amino Acids

Of the amino acids,[1–3] 9 are essential and 11 are nonessential. Some are conditionally essential or essential in specific circumstances. In Table 1 the essential and nonessential amino acids are listed. Nearly all foods contain amino acids either in the free form or as units in proteins. Reasons for detection are different and may be:

Table 1 Essential and nonessential amino acids

Essential	Nonessential
Histidine	Alanine
Isoleucine	Arginine
Leucine	Asparagine
Lysine	Aspartate
Methionine	Cysteine
Phenylalanine	Glutamate
Threonine	Glutamine
Tryptophan	Glycine
Valine	Proline
	Serine
	Tyrosine

- demand for information of nutritive value,
- detection of adulteration in foods and beverages,
- detection of potentially toxic amino acids or peptides and
- measurement of meat protein content.

Sample preparation for a total amino acid analysis is performed by an acid or alkaline hydrolysis of the protein into amino acids. A commonly used hydrolysis is 6 M HCl at 110 °C for 18–24 h. This procedure is a compromise for optimal estimation of a great number of amino acids. Alternatives could be: multiple hydrolysis times (24, 48, and 72 h), parallel application of several techniques, eliminating HCl using other organic acids, or using protective agents. An example of a problematic acid analysis is tryptophan. Sample preparation for tryptophan is hydrolysis in a base (e.g. NaOH) or enzymatic hydrolysis.

In many cases no sample pretreatment is necessary for a free amino acid analysis, only filtration and eventually dilution. Complete dissolution of the amino acids is pursued. For solid materials, homogenization of the sample and extraction in a solvent (e.g. 0.1 N HCl) are appropriate steps. An alternative is a hot acid aqueous extraction.

Large macromolecules, especially proteins, are responsible for loss of resolution over time. Deproteinization by precipitation with acids or organic acids, ultrafiltration, SPE with cation exchange or C_{18} resin, or guard columns with same packing as the analytical columns are steps in the sample cleanup. Precipitating agents are perchloric acid (PCA), phosphotungstic acid (PTA), picric acid (PA), trichloroacetic acid (TCA), acetonitrile (ACN), and sulfosalicylic acid (SSA).

Table 2 Derivatization methods for amino acids

	Derivatization agent	Instrumentation
Precolumn	PITC	Picotag
	OPA/FMOC	AminoQuant
	AQC	AccQ Tag
Postcolumn	Ninhydrin	
	OPA	

PITC, phenylisothiocyanate; OPA, *o*-phthalaldehyde; FMOC, 9-fluorenylmethyl chloroformate; AQC, 6-aminoquinolyl-*N*-hydroxy-succimidyl carbamate.

In Table 2 some derivatization methods are shown.

The separation techniques of amino acids fall in different classes:

- Ion exchange chromatography (IEC): the majority of methods use cation exchange resins (sulfonated polystyrene). Many automated amino acid analyzers exist: LKB, Beckman, . . . Ninhydrin or OPA are used as postcolumn derivatizing agents.
- RP is now very popular: precolumn derivatization agents are PITC, OPA, FMOC, and AQC.
- chiral: the number of enantiomeric separation methods of D- and L-stereoisomers of amino acids is growing. Some methods involve precolumn derivatization with a chiral reagent and separation with RPLC, and other separations of underivatized enantiomers on a chiral stationary phase or with chiral mobile-phase modifiers.

After separation, detection can be performed in the VIS (visible), UV or fluorescence mode.

4.2 Peptides

Peptides[4] improve functional properties of food and can be used as antioxidants, antimicrobial agents, surfactants or to give characteristic flavors to foodstuffs.

Reasons of detection of peptides are:

- peptide mapping
- characterization of foods or
- characterization of bitter peptides.

Sample preparation consists of different steps:

- extraction by water, solutions of acetic acid or sodium chloride, or organic solvents and deproteinization by precipitation with the same acids as for amino acids or ultrafiltration,
- separation from peptides or sugars, organic acids, salts, etc. by anion- and cation-exchange resins or dialysis,
- separation of amino acids and peptides by IEC or SEC and
- a possible final step by passing through Sep-Pak™ C_{18} cartridges.

Separation is performed by RPLC, IEC or SEC. In RPLC, silica-based columns are most widely used. The mobile phase consists of a mixture of water and an organic solvent like ACN or methanol. Elution gradient and isocratic separations are performed. The pH is kept lower than three and ion pair forming agents may be added (e.g. trifluoroacetic acid, TFA).

In the ion-exchange mode two types of ion exchangers are used: cation exchangers for neutral and basic peptides and anion exchangers for neutral or acid peptides.

The material of size exclusion columns is silica, alumina or organic materials such as polyesters with OH groups.

UV, fluorescence and MS are used as detection modes. Derivatizing agents are used for enhanced detection such as dansyl chloride or PITC. Fluorescamine is a fluorescent derivatizing agent with an excitation peak at 390 nm and an emission band at 475 nm.

Many methods of separating stereoisomers have been published.

4.3 Proteins

Table 3 gives a summary of the classification of proteins of concern for the food analyst.

The reasons for detection vary:

- some proteins influence functional properties of food,
- cultivar discrimination,
- heat classification of skim milk powder or
- detection of rennet whey buttermilk solids in milk powder.

Proteins have to be solubilized by reduction agents (e.g. β-mercaptoethanol) before determination and the solution clarified by centrifugation or filtration to remove cell debris, lipids, and nucleic acids. Different classes of proteins can be separated by ultracentrifugation or

Table 3 Food proteins

Cereal proteins	Milk proteins
Albumins	Caseins (α-Caseins, β-Caseins, κ-Caseins)
Globulins	
Prolamins (gliadin–zein)	Whey proteins (β-Lactoglobulin, α-Lactalbumin, Bovine serum albumin, Immunoglobulins)
Glutelins (soluble–insoluble)	

by selective precipitation using salts, organic solvents, or heat.

IEC, SEC, RPLC, hydrophobic interaction, metal chelate and bioaffinity, and hydroxyapatite chromatography are all possibilities for separating proteins. Metal chelate or bioaffinity chromatography is based on the affinity of a protein for a specific molecule in the stationary phase. Hydroxyapatite chromatography separates molecules on differential binding to phosphate and calcium sites of hydroxyapatite.

4.4 Lipids and Phospholipids

In Table 4 a simplified classification of lipids and phospholipids[5–7] is shown.

- Fatty acids: fatty acids are the major part of acylglycerols and include saturated, mono-unsaturated and polyunsaturated fatty acids. Essential fatty acids, e.g. linoleic acid and α-linolenic acid, are of great importance. Eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), w-3-polyunsaturated fatty acids, possess health benefits.
- Sterols and sterylesters: phytosterols, as β-sitosterol, are found in groundnut and soybean oil. Cholesterol is associated with health problems.
- Phospholipids: phospholipids such as lecithin and products from soybeans are important in the food industry (e.g. emulsifiers). Sphingomyelin and phosphatidylinositol are other examples of phospholipids.

Lipids are extracted by solvents, centrifuged and filtrated, further washed and finally distillated or evaporated. Free fatty acids can be isolated from a lipid extract by SPE. Lipid classes can be separated by column chromatography, TLC (thin-layer chromatography) or HPLC.

Alternatives for the analysis of lipids by GC (gas chromatography) are HPLC RP and argentation-HPLC (silica gel with silver nitrate). Different detection modes are used: RI, UV, IR (infrared), ELSD (evaporative light-scattering detection, and CI (chemical ionization) MS.

Table 4 Lipids and phospholipids

Fatty acids
Acylglycerols (glycerides):
esters of glycerol and fatty acids
monoacylglycerols, diacylglycerols, triacylglycerols
Sterols and steryl esters:
phytosterols (plants), cholesterol (animals)
Phospholipids:
glycerol + fatty acid(s) + phosphate + organic bases

For free fatty acids analysis GC is the method used traditionally. However, many HPLC methods have been used. Detection is neither sensitive nor selective. To increase these parameters UV or fluorescent derivatives are prepared either precolumn or postcolumn. Examples of derivatizing agents are 1-benzyl-3-*p*-tolutriazine, 1-*p*-nitrobenzyl-3-tolutriazine, or 9-anthryldiazomethane (ADAM) and of derivatives are 2-naphthacylesters, phenacylesters, and nitrophenylhydrazides.

Separation of *cis* and *trans* isomers of unsaturated fatty acids is performed by silver ion or argentation chromatography. Monohydroxy fatty acids are separated by NPLC and RPLC after derivatization with *O*-(*p*-nitrobenzyl)-*N*,*N*′-(diisopropyl)isourea (PNBDI). Many methods have been used to detect dimer and trimer acids, and higher molecular weight molecules formed during thermal oxidation.

Triglycerides can be separated by GC at higher temperatures (350 °C) or by RPLC. In the mobile phase no water is present (NARP, nonaqueous reversed-phase). An ion-exchange column impregnated with silver (silver IC) is also used.

Oxidized fats, caused by heat, light, and metals have undesirable favors and odors and result in health problems. Hydroperoxides and cyclic peroxides are such oxidation products. RPLC separation and ELSD or UV are possible methods to analyze these products.

Phospholipids are separated from other lipids by SPE. In the literature different compositions of solid or mobile phases are proposed. HPLC methods used for phospholipids can be arranged into separation techniques for phospholipid classes and separation techniques for molecular species. NPLC with a mobile phase of ACN, chloroform, or hexane/2-propanol is used to separate phospholipid classes. ELSD is mostly used as a detection system. UV-absorbing and fluorescent derivatization are applied in methods for molecular species characterization.

4.5 Carbohydrates

In Table 5 an overview of different carbohydrates[8] is shown.

Table 5 Carbohydrates

Monosaccharides:	
Pentoses	Xylose, ribose
Hexoses	Glucose, fructose, galactose
Disaccharides	Sucrose, lactose, maltose
Trisaccharides	Raffinose
Oligosaccharides	3–14 units
Polysaccharides	100–1000 units, starch
Polyhydric alcohols	Mannitol, sorbitol, xylitol

In the analysis of carbohydrates a difference is made between total soluble carbohydrates, total carbohydrates, reducing sugars and available carbohydrates. Among the many reasons for detecting sugars are the relationship that sugars have with health or health risks (e.g. dental caries), the determination of energy content and the need to determine adulteration. Other reasons are quality control of drinks and foodstuffs and monitoring of fermentation of alcoholic beverages.

Filtering or removing of proteins and fats by LLE, and extraction by (hot) water or solvents constitute steps in sample preparation. Dialysis is an alternative. The samples are further cleaned up by clarifying agents (e.g. Carrez reagents), solvent precipitation (e.g. mixture of acetate buffer/ACN), and ion-exchange or Sep-Pak™ cartridges (SPE). Precolumn or postcolumn derivatization can improve the detection of the analytes.

Often a RI detection follows separation over an amino-bonded silica, a special carbohydrate column or over an ion-exchange column. RI detection is the most widely used detection method for carbohydrates despite its low sensitivity. After a derivatization step, other possibilities are UV detection, EC detection, and evaporative light scattering detection.

4.6 Alcohols

Alcohols are detected in foods and beverages because knowledge of the alcohol content is needed to establish tax rates and because there may be adulteration by, for instance, glycerol, or methanol content which has toxic properties, and vinification processes and fermentation.

Sample preparation and cleanup of alcohols is carried out by distillation, solvent extraction, ion exchange, or by gel filtration. In most cases, however, samples are simply diluted and filtered through 0.45-μm or 0.2-μm filters.

IEC is a possible choice. The used mobile phase is a diluted water solution of H_2SO_4 (0.004–0.001 N) or H_3PO_4. Detection is in RI, UV, or EC mode.

4.7 Vitamins

There are two groups of vitamins, fat-soluble and water-soluble vitamins. Table 6 shows the different vitamins, possible precursors or vitamin-related compounds.

Different reasons can be listed why vitamins are assayed in foods:

- to provide data about food composition,
- to check vitamin changes during processing, packaging, and storage,
- to evaluate geographical, environmental, and seasonal parameters,
- to monitor fortified foods for added quantities and
- to evaluate contract specifications and regulations.

Table 6 Fat-soluble and water-soluble vitamins

Fat-soluble vitamins	
Vitamin A	Retinoids (retinol, retinyl esters, retinaldehyde), retinyl acetate, retinyl palmitate
Provitamin A	Carotenoids (β-carotene)
Vitamin D	Cholecalciferol (D_3), ergocalciferol (D_2)
Vitamin E	α-,β-, γ- and δ-tocopherol
Vitamin K	Phylloquinone (K_1), menaquinones
Water-soluble vitamins	
Vitamin B_1	Thiamin, phosphorylated forms
Vitamin B_2	Riboflavin, flavin mononucleotide, flavin adenine dinucleotide
Vitamin B_6	Pyridoxine, pyridoxal, pyridoxamine, three 5′-phosphate esters (pyridoxine phosphate pyridoxal phosphate and pyridoxamine phosphate)
Vitamin C	L-Ascorbic acid, dehydroascorbic acid
Nicotinic acid and nicotinamide	
Folacin (folic acid)	
Pantothenic acid	
Biotin	
Vitamin B_{12}	Cyanocobalamin

Different HLPC methods are developed for the analysis of individual vitamins and vitamin-related products and for simultaneous determination of multiple vitamins.

4.7.1 Fat-soluble Vitamins

To extract fat-soluble[9] vitamins from their matrices, and to remove glycerides, an alkaline hydrolysis (saponification) can be performed followed by LLE, enzymatic hydrolysis, alcoholysis, direct solvent extraction or SFE. Further cleanup is done by precipitation of sterols, open-column chromatography, SPE or SEC.

Separation modes are NPLC (vitamins A and E) or RPLC (carotenoids, vitamins D, E and K) or a combination of both modes (two-dimensional HPLC) for determining trace amounts of naturally occurring vitamins in foods. Absorbance of vitamins A, D, E and K and carotenoids or fluorescence of vitamin E are measured. An alternative is EC detection.

4.7.1.1 Vitamin A and Carotenoids Saponification or alkaline hydrolysis, hot or at ambient temperature, is used most to extract vitamin A and carotenoids. Many methods can be found in literature. Either NPLC or RPLC are used.

4.7.1.2 Vitamin D This vitamin can be extracted after saponification by a mixture of solvents. Vitamin D

absorbs in the 260-nm region. For trace amounts of vitamin D, and also vitamin K, separation is performed in a two-dimensional system.

4.7.1.3 Vitamin E Mostly NPLC but also RPLC are used for the determination of vitamin E. Sample preparation steps are hot saponification and solvent extraction. See Figure 1 for chromatograms of vitamin E.[10]

4.7.1.4 Vitamin K NPLC, RPLC and two-dimensional methods are described in the literature.[9,11]

4.7.2 Water-soluble Vitamins

Protein and nonprotein material is removed by enzymatic, acid or alkaline hydrolysis. TCA precipitates proteins. Finally water-soluble vitamins[11–13] are extracted by LLE or SPE. In many cases, adequate protection from sunlight, higher temperatures and oxygen are necessary. Most analysts use RPLC.

UV, fluorescence or EC detection are preferred for vitamin C. Vitamin C, riboflavin, niacin, folacin, pantothenic acid and multivitamins are measured in UV mode. Fluorescence detection is used for vitamin C, thiamin, riboflavin and vitamin B_6. Folacin can also be detected electrochemically.

4.7.2.1 Vitamin C HPLC is a very suitable analysis method for vitamin C. Methods using RP, polymer, ion exchange, ion exclusion, and amino columns and using UV, fluorescence, or EC detection are carried out.

4.7.2.2 Thiamin Thiamin is derivatized to thiochrome esters by ferricyanide to improve fluorescence detection. HPLC methods are frequently used to determine thiamine. An alternative is FIA (flow-injection analysis).

4.7.2.3 Riboflavin and Niacin Either acid or alkaline hydrolysis followed by an enzymatic hydrolysis can be used to extract vitamin B_2 and B_3 vitamers. RP or polymer-based columns are used to separate; UV or fluorescence are used to detect.

In Figure 2 a chromatogram is shown of folate standards.[14]

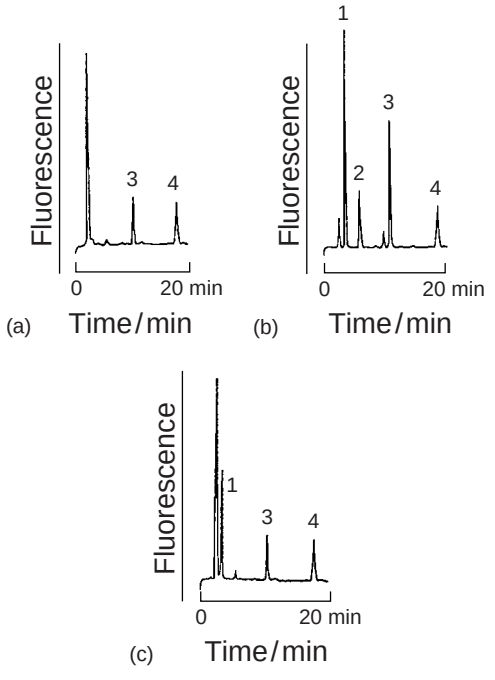

Figure 1 Chromatograms of vitamin E obtained with fluorescence detection (excitation, 285 nm; emission, 310 nm), flow rate of 1.0 mL min^{-1}, injection volume of 50 µL, and mobile phase of 0.5% isopropyl alcohol in hexane. Column: Lichrosorb Si60, 5 µm, 4.6 mm × 25 cm. (a) Extract of unfortified zero control reference material, (b) Standard, and (c) standard reference material. Peaks: 1, All-rac-α-tocopheryl acetate; 2, (*R*,*R*,*R*)-α-tocopherol; 3, γ-tocopherol; 4, δ-tocopherol. (Reprinted from *J. AOAC Int.*, **81**, 584 (1998). Copyright, 1998 by AOAC International.[10])

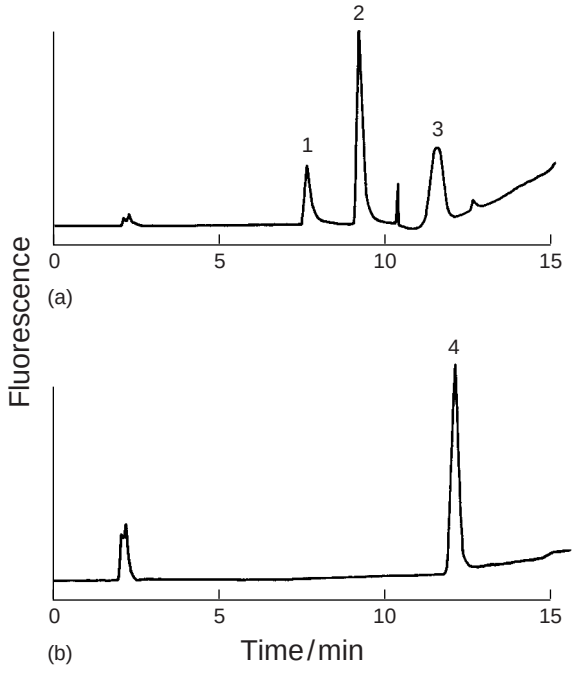

Figure 2 Liquid chromatograms of reduced folate monoglutamate standards. (a) Peak 1, tetrahydrofolate (1 ng), peak 2, 5-methyltetrahydrofolate (3 ng), and peak 3, 5-formyltetrahydrofolate (4 ng). Fluorescence detection wavelengths: excitation, 290 nm; emission, 356 nm. (b) Peak 4, 10-formylfolic acid (2 ng). Fluorescence detection wavelength: excitation, 360 nm; emission, 460 nm. Columns: Hypersil ODS, 150 × 4.6 mm, 3 µm and Spherisorb ODS, 250 × 4.6 mm, 5 µm. Mobile phase: ACN–phosphate buffer (pH 2.2) gradient (9% ACN increased to 24% within 8 min after 4-min lag phase, ACN decreased to 9% at 14 min for equilibration); 0.8 mL min^{-1}. (Reprinted from *J. AOAC Int.*, **80**, 375 (1997). Copyright, 1997 by AOAC International.[14])

4.8 Organic Acids

Organic acids in foods are the result of normal biochemical processes or by addition, hydrolysis or bacterial growth. They are added to foods as acidulants, stabilizers, or preservatives. They give a particular flavor, taste or aroma and are responsible for specific sensory properties. Table 7 gives a list of frequently found organic acids.

Beverages are simply diluted in an appropriate solution or in the mobile phase followed by centrifugation and filtration (0.45 or 0.22 μm). Carbonated drinks must be degassed. SPE or ion-exchange resins further clean up.

Sample preparation steps for solid samples are: extraction by mixtures of water and ethanol (70–80%) or ACN, centrifugation, and filtration. Pre- and post-column derivatization of organic acids is performed on-line or off-line. Derivatization methods for UV-detection use phenacyl, naphthacyl, and *p*-nitrobenzyl groups; for fluorometric detection 4-methyl-7-methoxycoumarin, 4-methyl-6,7-dimethylcoumarin, and 9-methylanthracine are used.

Organic acids are separated by IEC, RPLC or ion-exclusion chromatography. The method most used is ion-exclusion chromatography. In most methods UV, conductivity, or RI detection are used.

4.9 Organic Bases

Organic bases[15] can be grouped by different criteria. In Table 8 different groups of organic bases are tabulated.

- Biogenic amines are naturally present in the central nervous system as neurotransmitters. They are often present in some foods such as cheese and fish and are reported to be responsible for several diseases when used in large amounts. Histamine is associated with scombroid poisoning. Biogenic amines can be formed by degradation processes in food.

Table 7 Organic acids

Monocarboxylic acids		Di-and tricarboxylic acids
Unsubstituted	Formic acid Acetic acid Propionic acid Butyric acid Isobutyric acid	Oxalic acid Oxalacetic acid
Substitued	Glycolic acid Lactic acid Pyruvic acid Glyoxylic acid	
Alicyclic	Quinic acid Shikimic acid	

Table 8 Organic bases

- Amines
  - Biogenic amines
    - Histamine, tryptamine, putrescine, cadaverine, spermidine, spermine
  - Unsulfonated aromatic amines
    - Aniline, benzidine
  - Heterocyclic amines
    - Quinolines, quinoxalines, amino acid pyrolyzates
  - Nitrosamines and volatile alkylamines
    - Methylamine, dimethylamine, butylamine, penthylamine
- Alkaloids
  - Methyl- and hypoxanthines, purines, pyrimidines
    - Methylxanthines: theobromine, caffeine
  - Cinchona alkaloids
    - Quinidine, quinine
  - Spicy food alkaloids
    - Piperine
  - Solanum alkanoids
  - Ergot alkaloids

- Unsulfonated aromatic amines: synthetic food colors contain toxic impurities or can degrade leading to carcinogenic aromatic amines.
- Mutagenic heterocyclic amines may be created by frying or boiling meat.
- Volatile aliphatic amines are responsible for organoleptic changes and capable of forming nitrosamines.
- Methylxanthines: caffeine and theobromine are present in cocoa, coffee, tea and many beverages. Caffeine can become a drug.
- Purines and pyrimidine bases are parameters for evaluating food quality or degradation.
- Cinchona alkaloids: quinine is a widely used drug and flavoring agent in beverages.
- Spicy food alkaloids: piperine, pipetine, and capsaicinoids are found in plants and used in cooking.
- Solanum alkaloids are compounds found in *Solanum* plants.
- Ergot alkaloids are produced by *Claviceps* fungi and may be poisonous.

Considering the wide variety of organic bases, all types of extraction, separation and detection methods are applied. A derivatization can be included in the method.

For biogenic amines and mutagenic heterocyclic amines RPLC is the most suitable method. Precolumn or post-column derivatization are used to enhance absorbance or fluorescence (e.g. AQC, FMOC, and OPA) of biogenic amines.

Frequently ion-pair chromatography or RPLC are used to analyze unsulfonated aromatic amines followed by spectrophotometric detection at 254 and ±510 nm.

RPLC and UV-detection at 254 nm are conditions of purine and pyrimidine analyses. A wide variety of methods exist for the separation of methylxanthines.

Solanum alkaloids are separated by HPLC and UV in the 200 nm region; spicy food alkaloids are separated by HPLC and UV at 280 nm.

Most methods for ergot alkaloids consist of HPLC with fluorimetric detection.

4.10 Phenolic Compounds

In Table 9 a list of phenolic compounds is given.

Phenolic compounds are diverse compounds characterized by a hydroxyl group on a benzene ring. Phenolic compounds give the sensory quality, taste and flavor of fresh foods. They are useful for their pharmacological properties and in perfumery.

In sample preparation an extraction by organic solvents can be followed by an acid hydrolysis (100 °C 2 N HCl > 2 h). Sample cleanup is by LLE, SPE or open column chromatography. Separation is done by RPLC followed by an UV or PDA (photodiode array) detection. The mobile phase is frequently a mixture of water, methanol or ACN, and an acid such as acetic acid or TFA. Often the mobile phase is buffered or an ion-pairing agent is added.

4.11 Bittering Substances

In Table 10 the different components of *Humulus lupulus* L. can be found.[16]

α- and β-acids are extracted by an organic solvent as toluene or methanol. Hop pellets or hop extracts are diluted in methanol or in the mobile phase. Ion-pairing or RPLC separate the components that are detected at 314 nm. Sample pretreatment of iso-α-acids consists of precipitation, LLE and SPE. Separation is in the ion-pairing mode and detection at 270 nm.

4.12 Pigments

In Table 11 the different pigments are tabulated.

- Carotenoids are detected in foodstuffs to evaluate food quality by pigment concentration, to meet regulations on colorants and because of the interest in carotenoids as provitamins and antioxidants. In sample preparation UV-light and heat should be avoided. After solvent extraction lipids and chlorophylls are removed by saponification. Sterols can be removed by precipitating them in different solvents. Carotenoids are separated by either NPLC or RPLC. Detection is in UV/VIS mode or also with MS.
- Chlorophylls: the green color of chlorophyl gives an idea of the degreening or maturing of fruits, of food quality or of color loss during processing. Chlorophylls are possibly natural colorants. All sample preparation and analysis steps must be carried out in darkness. After solvent extraction, filtering and SPE, chlorophylls are separated by NPLC but preferably by RPLC. Detection is by absorbance or fluorescence.
- Anthocyanins: the reasons for detecting anthocyanins in food are the same as for carotenoids and chlorophylls – evaluation of food quality by color and the use of natural colorants. Extraction is done by acid solvent mixtures. Further cleanup is worked out by SPE. A very good choice is RPLC for separation and UV/VIS or PDA for detection.
- Betalains: once again the reasons for detection are identical. Betalains are extracted in water or water–solvent mixtures. Further cleanup has to be

Table 9 Phenolic compounds

Simple phenols
Phenolic acids
Coumarins
Naphthoquinones
Anthraquinones
(Iso)flavonoids
Lignans
Lignins
Tannins

Table 10 Components of *Humulus lupulus* L

Hop *Humulus lupulus* L.	
• Wax	
• Total resins	
–Soft resins	Products of the brewing process:
= α-Acids	⇒ Iso-α-acids
Humulone	Bittering substances
Cohumulone	Beer bitterness
Adhumulone	
= β-Acids	
Lupulone	
Colupulone	
Adlupulone	
–Hard resins	

Table 11 Pigments in foods

Carotenoids
β-Carotene, lutein, isolutein, lycopene, zeaxanthin
Chlorophylls
Chlorophyll a, chlorophyll b, phaeophytin
Anthocyanins (glycosides → anthocyaninides + sugar)
Cyanidin, delphinidin, malvidin
Betalains
Betacyanin, betaxanthin
Myoglobin

done by chromatographic methods. Good separation methods are RPLC or ion-pairing. Detection is in UV/VIS mode.

- Myoglobin gives the color to meat. This is the prominent reason for detection: meat quality. Myoglobin is extracted in water or aqueous buffers. Samples are further cleaned up by fractionating with ammonium sulfate, dialysis, or column chromatography. Almost all modes of LC have been used to analyze myoglobins.

5 RESIDUES

5.1 Mycotoxins

Mycotoxins[17] are fungal metabolites that can be found in foods, especially grains, peanuts and milk. Contamination can occur before and in all steps of food production: in the growing crop, during storage and processing and in the finished food. Mycotoxins are dangerous for human health and may be carcinogens. The different mycotoxins are enumerated in Table 12.

Mycotoxin distribution in raw materials is seldom homogeneous. Good attention must be given to the sampling. After solvent extraction, cleanup is worked out by filtration, liquid–liquid partition, column chromatography and SPE. A precolumn derivatization (e.g. TFA) or postcolumn derivatization (e.g. iodine) step can be included. A good alternative for HPLC as an analysis technique is TLC.

Table 12 Mycotoxins

Mycotoxins	Fungi
Aflatoxins $B_1B_2G_1G_2$	*Aspergillus flavus* *Aspergillus parasiticus*
M_1M_2	metabolic products
Fumonisins	*Fusarium moniliforme* *Fusarium proliferartum*
Zearalenone	*Fusarium* species
Ochratoxin A	*Aspergillus ochraceus* *Penicillium verrucosum*
Patulin	*Aspergillus* species *Penicillium* species
Deoxynivatenol	*Fusarium* species
Trichothecenes T-2 toxin HT-2 toxin Nivolenol Deoxynivolenol	*Fusarium* species
Cyclopiazonic acid	*Aspergillus* species *Penicillium* species
Alternaria toxins	*Alternaria* species
Ergot alkaloids	*Claviceps purpurea*
Sterigmatocystin	*Aspergillus* species *Bipolaris* species
Citrinin	*Aspergillus* species *Penicillium* species
Moniliformin	*Fusarium* species
Roquefortine	*Penicilium roqueforti*

Aflatoxins are extracted with chloroform, methanol, acetone, or ACN. The samples are cleaned up by SPE or immunoaffinity (IA) columns. Separation is in RP mode followed by fluorometric detection.

Extraction of Ochratoxin A is with organic solvents in acid medium followed by SPE or IA cleanup. The separation in RP mode is also followed by a fluorometric detection. GC not HPLC is the most used separation technique for trichothecenes. Determination of zearalenone is performed by RPLC and fluorescence detection.

5.2 Residual Antibacterials

Residues of antibacterials[18] are found in dairy products in low concentrations. Consequences of residues in foods are antibiotic resistance, mutagenic and carcinogenic effects. In Table 13 different classes of antibacterials with a few examples can be found.

Owing to the wide variety of compounds, sample preparation methods are very different. The methods include extraction, cleanup (SPE) and if necessary derivatization.

RPLC is an excellent technique for separating residual antibacterials. Detection is in different modes.

5.3 Residues of Growth Promoters

Growth promoters[19] are used for therapeutic or prophylactic purposes and for a better breeding efficiency or better protein conversion. The European Union has prohibited substances which have a hormonal action (some hormones may be used for therapeutic purposes).

A classification is shown in Table 14.

Sample preparation consists of a simple homogenization followed by centrifugation or enzymatic digestion followed by extraction by organic solvents. Further cleanup is performed by column chromatography, SPE

Table 13 Classification of antibacterials

Aminoglycosides	Apramycin, kanamycin A, B, and C
β-Lactams	Amoxicillin, cloxacillin, ampicillin
Macrolides	Erythromycin, tylosin
Nitrofurans	Nitrofurazone, furazolidone
Peptides	Bacitracin, polymyxin
Quinolones	Nalidixic acid, flumequine
Sulfonamides	Sulfanilamide, sulfamethazine
Tetracyclines	Oxytetracycline, chlortetracycline
Other	

Table 14 Classification of growth promoters

	Estrogenic	Nonestrogenic
Endogeneous steroids	17 β-Estradiol	Progesterone Testosterone Nortestosterone
Exogeneous steroids	17 β-Estradiol (esters)	Testosterone (esters) Trenbolone (esters)
Nonsteroidal molecules	Diethylstilbestrol Hexestrol Zeranol	
β-Agonistic drugs	Clenbuterol Mobuterol	

or MSPD (matrix solid-phase dispersion). HPLC can be used to separate the residues in different columns (NPLC, RPLC and other). Detection is UV, fluorescence or EC detection.

5.4 Pesticides

Pesticides[20] are chemical compounds that are used to control pests. Dependent on the pest, acaricides, fungicides, herbicides, insecticides, etc. are used. More than 1300 pesticides are available. The major groups are summarized in Table 15.

Residues of pesticides in food can have different negative effects on human health. The extraction modes for pesticide residues are SPE, solvent extraction and SFE. Further cleanup methods are liquid–liquid partitioning and distillation. If necessary, precolumn derivatization is performed (e.g. using *p*-nitrobenzylchloride – an UV derivatizing agent for ETU (ethylenethiourea), a degradation product of ethylene-bis-dithiocarbamate) or postcolumn derivatization (e.g. using OPA, a fluorescent derivatizing agent of *N*-methylcarbamate).

Considering the large number of pesticides and the wide variety, most types of columns are used: NPLC, RPLC, IEC, amino-bonded, diol-bonded, cyano-bonded and SEC. Also all types of detection system are chosen in one or another method: UV, fluorescence or EC.

Table 15 Major groups of pesticides

Groups		Degradation products
Urea pesticides	Herbicides or insecticides	ETU
Diuron, chlorotoluron, fluometuron		
Organochlorine pesticides	Insecticides	DDE
Chlordane, heptochlor, aldrin, DDT		
Carbamate pesticides	Insecticides or acaricides	
Aldicarb, carbonyl, pirimicarb		
Organophosphates	Insecticides	
Malathion, parathion, azinphos methyl, organothiophosphate		Metabolites II, III, IV, V, VI, VII

DDE, 1,1-Dichloro-2,2-bis(4-chlorophenyl)ethylene; DDT, 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane.

5.4.1 Fungicide Residues

Fungicides have a wide variety of chemical structures and functional groups and are nearly all multifunctional. In Table 16 fungicides are classified into six categories.

Multiresidues of fungicides are extracted by solvent extraction (ACN or acetone) and further cleaned up by LLE, column chromatography, SPE and others. Samples can be analyzed by RPLC or ion-pairing LC.

5.4.2 Herbicide Residues

The major classes of herbicides can be found in Table 17.

The same extraction procedures and cleanup methods as for organophosphates, organochlorine pesticides and

Table 16 Fungicide classes

Inorganic compounds	Copper and sulfur (compounds)
Organometallic compounds	Organotin and organomercury
Organophosphorus compounds	
Halogenated hydrocarbons	Methylbromide
Organonitrogen compounds	
Aliphatics/aromatics/heterocyclic	
Organic nitrogen–sulfur compounds	

Table 17 Major classes of herbicide

Classes	Example
Triazines	Atrazine
Phenoxyalkanoic acids	(2,4-Dichlorophenoxy) acetic acid
Aryloxyphenoxypropanoic acids	Fluazifop
Bipyridinium compounds	Diquat
Substituted phenols	Bromoxynil
Organophosphates	Glyphosate
Chloroacetamides	Metolachlor
Carbamates	Propham
Urea	Linuron
Dinitrianilines	Trifluralin
Heterocyclic compounds	Terbacil

Table 18 Classification of NOCs

Nitrosamines:
Volatile nitrosamines
Aryl nitrosamines
Nitrosated amino acids
Hydroxylated nitrosamines
Nitrosamides:
Alkyl nitrosamides
Nitrosaurea
Nitrosocarbamate
Nitrosoguanidine
Nitrosated dipeptides
Nitroso sugar amino acids

fungicides are applicable. In some cases precolumn or postcolumn derivatization to enhance UV or fluorescence detection is carried out. All types of columns are used to separate the residues. Detection is in UV or fluorescence modes.

5.5 *N*-Nitroso Compounds

Table 18 shows a classification of NOCs (*N*-nitroso compounds).

NOCs have the functional group N−N=O. They are formed by the reaction between a nitrogen-containing organic compound (amines, amides) and a nitrosating agent (nitrites, nitrogen oxide gases). NOCs are mutagenic, carcinogenic and teratogenic. The best known NOC is *N*-nitrosodimethylamine (NDMA).

Sample preparation has different steps: extraction, cleanup and concentration. Extraction is performed by solvent extraction, steam or vacuum distillation or SPE. Cleanup is achieved by LLE or column extraction. Because many NOCs are nonvolatile or are not separated by GC, HPLC is a good choice. Different approaches are carried out:

- HPLC methods with postcolumn NO detection by chemiluminescence (HPLC/TEA, thermal energy analysis). The stationary phase is silica, phenyl-bonded, cyano-bonded or amino-bonded silica.
- HPLC methods with UV detection of postcolumn or precolumn dinitrosation products other than NO on a RP column.

6 MISCELLANEOUS

6.1 Colorants

Natural (from vegetable, animal or mineral origin) or synthetic (organic or inorganic) colorants[21] provide foods with an attractive appearance. Classification of colorants is not simple but such a classification has been carried out by EEC, International Numbering System, and in the Color Index. Applied colors are erythrosine, amaranth, Ponceau 4R, Allura Red, Sunset Yellow, and tartrazine.

Use of dyes or chemical impurities in colorants can have undesirable side effects. Azo dyes are reduced to aromatic amines. Metals, aromatic amines and others are detected as impurities.

Liquid samples are filtered through 0.45-μm filters and injected. Water-soluble samples are dissolved in (heated) water. Solid foods are extracted in organic solvents such as methanol or petroleum ether, filtered or centrifuged.

Concentration techniques for dyes are wool-dyeing procedures, column chromatography, IEC, ion-pair chromatography, solvent extraction, or SPE.

Food colorants can be separated by NPLC, ion-pair chromatography, or RPLC combined with UV-detection.

6.2 Preservatives

Preservatives are chemical compounds added to food products to prevent or inhibit the activity and growth of microorganisms. Sulfites and sulfur-based compounds are detected by ion-pair HPLC. In most cases the detection is in the EC mode.

Post-harvest preservatives, thiabendazole, biphenyl and *o*-phenyl phenol, used to protect certain fruits, are extracted by ethyl acetate, separated (NPLC/RPLC) and UV detected.

The organic acids formic acid, benzoic acid, sorbic acid and esters or salts of 4-hydroxybenzoic acid are extracted from the matrix by steam distillation or direct extraction. Separation is carried out in normal-phase, RP, ion-pair, ion-exchange or amino-bonded phase. Detection is at 235 nm.

6.3 Synthetic Antioxidants

Synthetic antioxidants[22] inhibit the process of oxidation in foods. Only the primary antioxidants alkyl gallates, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), *tert*-butylhydroquinone, nordihydroguaiaretic acid, and ethoxiquin are discussed here.

The last three are not generally allowed in foods or have toxic effects.

Vacuum or steam distillation are frequently used for isolation of antioxidants from food samples. HPLC is not the only method used to analyze antioxidants. NPLC, RPLC (the majority), SEC and IEC are performed. In most methods discussed, 280 nm is the chosen wavelength.

6.4 Sweeteners

The capacity to produce a sweet taste[23] is inherent in sugars such as sucrose, amino acids like aspartame, proteins such as thaumatin and cyclic (cyclamates) and aromatic organic derivatives (neohesperidin dihydrochalcone).

Some intense sweeteners are extracted from plants (glycyrrhizin, steviosides, thaumatin), others are obtained by organic synthesis (acesulfam-K, sucralose, saccharin). Most of the intense sweeteners are not metabolized by the body.

Bulk sweeteners (simple sugars, sugar alcohols, and starch) are mostly metabolized by the body. Intense sweeteners are nonnutritive or noncaloric sweeteners. Their sweetness is 30 to 3000 times that of sucrose.

- Acesulfam-K is very stable even at high temperatures. HPLC is a very good choice to determine Acesulfam-K in foods. Generally a RP column and a mobile phase consisting of ACN and KH_2PO_4 buffer are used. The sweetener and other compounds (aspartame, saccharin, ...) are detected by UV in the region of 200–230 nm.
- Aspartame: the stability of aspartame in dry products is relatively good. Decomposition occurs in liquid products at higher temperatures and at pH <3.0 or >6.0. A breakdown product is phenylalanine which is inconvenient for individuals suffering from phenylketonuria.

  A very suitable analysis method is RPLC. UV-detection is between 210 and 220 nm. By varying the organic modifier or the pH different of the mobile phase sweeteners, decomposition products or other molecules can be detected simultaneously.
- Cyclamate: Stability of cyclamate in foods is very good over a wide range of parameters. Cyclamate is separated from other components by HPLC. Precolumn (e.g. sodium hypochlorite) or postcolumn derivatization (e.g. methyl violet) has to be done.
- Glycyrrhizin (extract from *Glycyrrhiza glabra* L.) is analyzed on a RP column using ACN–acetic acid–water as a mobile phase and a detector at 254 nm.
- Saccharin is commonly used in different types of food. HPLC is a very good choice for determining saccharin. The chromatographic conditions are the same as for acesulfam-K.
- Stevioside is a glycoside found in the leaves of *Stevia rebaudiana* Bertoni. HPLC conditions are
  - RP or amino-bonded phase column,
  - ACN–water or methanol–5 mM NaOH as mobile phase,
  - detection at 210 nm.

Table 19 Anions and cations in food

Anions:	
Halogens	
BrO_3^-	Baking product
Methylbromide	Fumigating product
I^-	In table salt
Nitrogen	
NO_3^-	In meat production
NO_2^-	Preservative
Sulfur	
SO_3^-	Antioxidant
Cations:	
Alkali metals	
Na^+, K^+	Electrolyte balance High blood pressure (Na^+)
Alkaline earths	
Mg^{2+}	Bone growth
Ca^{2+}	
Transition metals	
Cu^{2+}	Environmental hazards
Ni^{2+}	

7 ANIONS AND CATIONS

Important anions and cations and reasons for their detection are tabulated in Table 19.

In sample preparation of anions the following steps can be distinguished: dilution, filtration, precipitation of proteins (treating with Carrez solutions), and SPE. Steps in the sample cleanup of cations are grinding, dilution and filtration.

Separation modes for anions and cations are IEC or ion-exclusion chromatography. Conductivity detection is common in IC. However, eluents and analyte are conductive and thus it is difficult to distinguish the difference in conductivity between analyte and eluent. A suppression system reduces the conductivity of the eluent: suppression IC. In nonsuppressed IC low ionic strength eluents are used. Indirect detection with the help of a counterion is in the UV/VIS region. CZE (capillary zone electrophoresis) is a good alternative.

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
ADAM	9-Anthryldiazomethane
AQC	6-Aminoquinolyl-*N*-hydroxysuccimidyl Carbamate
BHA	Butylated Hydroxyanisole
BHT	Butylated Hydroxytoluene
CI	Chemical Ionization
CZE	Capillary Zone Electrophoresis

DDE	1,1-Dichloro-2,2-bis(4-chlorophenyl)-ethylene
DDT	1,1,1-Trichloro-2,2-bis(*p*-chlorophenyl)-ethane
DHA	Docosahexaenoic Acid
EC	Electrochemical
ELSD	Evaporative Light-scattering Detection
EPA	Eicosapentaenoic Acid
ETU	Ethylenethiourea
FIA	Flow-injection Analysis
FMOC	9-Fluorenylmethyl Chloroformate
GC	Gas Chromatography
GFC	Gel Filtration Chromatography
GPC	Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
IA	Immunoaffinity
IC	Ion Chromatography
IEC	Ion Exchange Chromatography
IR	Infrared
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
MS	Mass Spectrometry
MSPD	Matrix Solid-phase Dispersion
NARP	Nonaqueous Reversed-phase
NDMA	*N*-Nitrosodimethylamine
NMR	Nuclear Magnetic Resonance
NOC	*N*-Nitroso Compound
NPLC	Normal-phase Liquid Chromatography
OPA	*O*-Phthalaldehyde
PA	Picric Acid
PCA	Perchloric Acid
PDA	Photodiode Array
PITC	Phenylisothiocyanate
PNBDI	*O*-(*p*-nitrobenzyl)-*N*,*N*′-(diisopropyl)-isourea
PTA	Phosphotungstic Acid
RI	Refractive Index
RP	Reversed-phase
RPLC	Reversed-phase Liquid Chromatography
SAX	Strong Anion Exchanger
SCX	Strong Cation Exchanger
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction
SSA	Sulfosalicylic Acid
TCA	Trichloroacetic Acid
TEA	Thermal Energy Analysis
TFA	Trifluoroacetic Acid
TLC	Thin-layer Chromatography
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible
WAX	Weak Anion Exchanger
WCX	Weak Cation Exchanger

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis • Monosaccharides and Sugar Alcohol Analysis

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Flavor Analysis in Food • Fluorescence Spectroscopy in Food Analysis • Lipid Analyses in Food • Pesticides, Mycotoxins and Residues Analysis in Food • Proteins, Peptides, and Amino Acids Analysis in Food • Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Organophosphorus Pesticides in Water and Food, Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Vitamins: Fat and Water Soluble, Analysis of

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Chiral Separations by High-performance Liquid Chromatography • Ion Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography

FURTHER READING

R. Macrae (ed.), *HPLC in Food Analysis*, Academic Press, London, 1988.
L.M.L. Nollet (ed.), *Food Analysis by HPLC*, Marcel Dekker, New York, 1992.
L.M.L. Nollet (ed.), *Handbook of Food Analysis*, Marcel Dekker, New York, 1996.

REFERENCES

1. G.C. Barrett (ed.), *Chemistry and Biochemistry of the Amino Acids*, Chapman and Hall, New York, 1985.
2. J.M. Rattenbury (ed.), *Amino Acid Analysis*, John Wiley and Sons, New York, 1981.
3. G. Sarwar, H.G. Botting, *J. Chromatogr.*, **615**, 1 (1993).
4. C. Schoneisch, S.K. Kwok, G.S. Wilson, S.R. Rabel, J.F. Stobaugh, T.D. Williams, D.G. Van der Velde, *Anal. Chem.*, **65**, 67R (1993).
5. W.W. Christie, *Advances in Lipid Methodology*, Oily Press, Dundee, Vol. 1, 1992.
6. W.W. Christie, *Advances in Lipid Methodology*, Oily Press, Dundee, Vol. 2, 1993.

7. J.L. Sebedio, E.G. Perkins, *New Trends in Lipid and Lipoprotein Analysis*, AOCS Press, Champaign, IL, 1995.
8. G.G. Birch (ed.), *Analysis of Food Carbohydrate*, Elsevier, London, 1985.
9. G.F.M. Ball, *Fat-soluble Vitamin Assays in Food Analysis*, Elsevier Applied Science, London, 1988.
10. G.W. Chase, A.R. Long, *J. AOAC Int.*, **81**, 584 (1998).
11. A.P. De Leenheer, W.E. Lambert, H.J. Nelis (eds.), *Modern Chromatographic Analysis of Vitamins*, 2nd edition, Marcel Dekker, New York, 1992.
12. A. Polesello, A. Rizzolo, *J. Micronutr. Anal.*, **2**, 153 (1986).
13. B. Polesello, A. Rizzolo, *J. Micronutr. Anal.*, **8**, 105 (1990).
14. L.L. Vahteristo, V. Ollilainen, P. Varo, *J. AOAC Int.*, **80**, 373 (1997).
15. M. Popl, J. Faehnrich, V. Tatar, *Chromatographic Analysis of Alkaloids*, Marcel Dekker, New York, 1983.
16. Chi-Tang Ho, C.H. Manley (eds.), *Flavor Measurement*, Marcel Dekker, New York, 1993.
17. V. Betina (ed.), *Chromatography of Mycotoxins: Techniques and Applications*, Elsevier, New York, 1993.
18. I.D. Morton (ed.), *Determination of Veterinary Residues in Food*, Ellis Horwood, Chichester, UK, 1991.
19. C. Creaser, R. Purchase (eds.), *Food Contaminants, Sources and Surveillance*, The Royal Society of Chemistry, Cambridge, 1991.
20. G. Zweig, J. Sherma (eds.), *High-performance Liquid Chromatography of Pesticides*, Academic Press, New York, 1982.
21. A.L.P. Branen, P.M. Davidson, S. Saminen, *Food Additives*, Marcel Dekker, New York, 1990.
22. B.J.F. Hudson (ed.), *Food Antioxidants*, Elsevier, London, 1990.
23. L. O'Brien-Nabors, R.C. Gelardi (eds.), *Alternative Sweeteners*, 2nd Edition, Marcel Dekker, New York, 1991.

Near-infrared Spectroscopy in Food Analysis

Brian G. Osborne
BRI Australia Ltd, North Ryde, Australia

1 Introduction **1**
2 Instrumentation **2**
2.1 Monochromators 3
2.2 Diode Array Spectrometers 3
2.3 Filter Instruments 3
3 Sample Presentation **3**
3.1 Diffuse Transmittance 3
3.2 Diffuse Reflectance 4
3.3 On-line Samplers 4
4 Calibration Development **6**
4.1 Data Preprocessing 6
4.2 Sample Population Structuring 6
4.3 Regression Techniques 6
4.4 Qualitative Analysis 7
5 Food Applications **7**
5.1 Cereals and Cereal Products 7
5.2 Milk and Dairy Products 10
5.3 Meat 10
5.4 Fish 10
5.5 Fruit and Vegetables 10
5.6 Confectionery 11
5.7 Beverages 11
5.8 Authenticity 11
6 The Literature **11**
7 Standard Methods **12**
Acknowledgments **12**
Abbreviations and Acronyms **12**
Related Articles **13**
References **13**

Near-infrared (NIR) spectroscopy is based on the absorption of electromagnetic radiation at wavelengths in the range 780–2500 nm. NIR spectra of foods comprise broad bands arising from overlapping absorptions corresponding mainly to overtones and combinations of vibrational modes involving C–H, O–H and N–H chemical bonds. The concentrations of constituents such as water, protein, fat and carbohydrate can in principle be determined using classical absorption spectroscopy. However, for most food samples, this chemical information is obscured by changes in the spectra caused by physical properties such as the particle size of powders. This means that NIR spectroscopy becomes a secondary method requiring calibration against a reference method for the constituent of interest. As a consequence of the physics of diffuse transmittance and reflectance and the complexity of the spectra, calibration is normally carried out using multivariate mathematics (chemometrics).

NIR spectroscopy is used routinely for the compositional, functional and sensory analysis of food ingredients, process intermediates and final products.

The major advantage of NIR is that usually no sample preparation is necessary, hence the analysis is very simple and very fast (between 15 and 90 s) and can be carried out on-line. One of the strengths of NIR technology is that it allows several constituents to be measured concurrently. In addition, for each fundamental vibration there exists a corresponding series of overtone and combination bands with each successive overtone band approximately an order of magnitude less intense than the preceding one. This provides a built-in dilution series which allows several choices of absorptions of different intensity containing the same chemical information. Finally, the relatively weak absorption due to water enables high-moisture foods to be analyzed.

The major limitation of NIR spectroscopy in food analysis is its dependence on less-precise reference methods.

1 INTRODUCTION

The NIR forms that part of the electromagnetic spectrum in the wavelength range 780 nm to 2500 nm (Figure 1). NIR, like all radiation, behaves as a wave with the properties of *simple harmonic motion* which may be defined in terms of two properties:

- the frequency of vibration
  = no. of times the wave pattern is repeated in 1 s
  = angular velocity/2
- the wavelength $= \dfrac{\text{velocity of light}}{\text{frequency}}$

Chemical bonds between atoms in molecules vibrate and to a first approximation this vibration behaves as a simple harmonic motion. The motion of each atom may be treated as an independent vibration with respect to a fixed centre of mass of the molecule rather like a mass attached to a spring. The vibration frequency is a function of the masses of the two atoms m_1 and m_2 and the strength of the bond k and there is a parabolic relationship between potential energy and interatomic distance.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

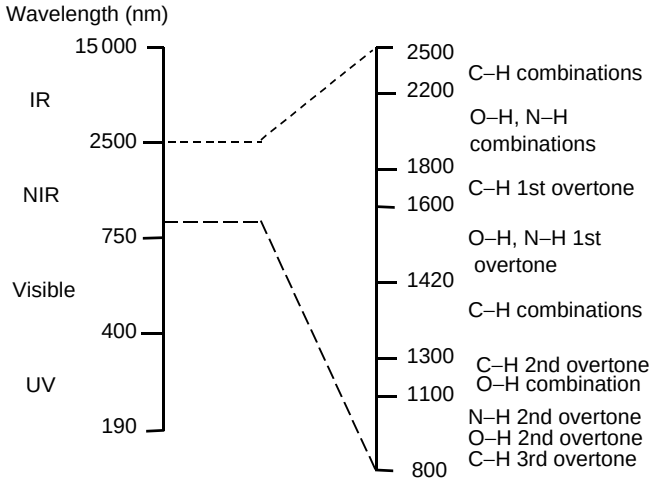

Figure 1 Principal types of NIR absorption bands and their locations.

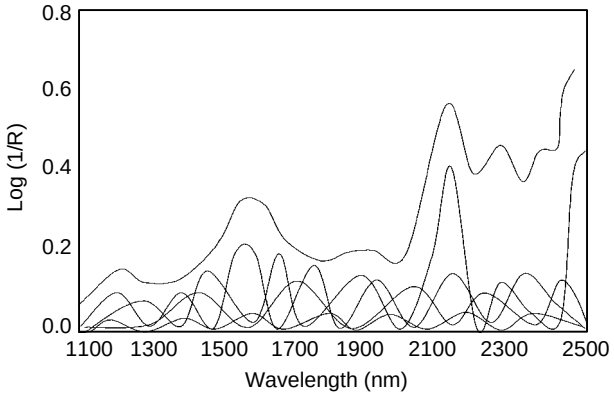

Figure 2 A typical spectrum of a powdered cereal sample (upper curve) and some of its constituent absorptions.

When the frequency of the radiation matches that of the vibrating molecule, there will be a net *transfer of energy* from the radiation to the molecule which can be measured as a plot of energy versus wavelength called a spectrum (Figure 2). However, energy is not transferred continuously but in discrete packets called quanta. In consequence, molecules can only occupy discrete energy levels defined by whole numbers $0, 1, 2, \ldots$ and the energy change required for a transition between two levels is directly related to the frequency of the energy by multiplying by h (Planck's constant). For simple harmonic motion, the absorption or emission of energy can only take place if it matches a transition between any pair of adjacent energy levels. Since almost all the molecules in a sample will normally occupy the lowest energy level 0, this means that the only possible transition is from 0 to 1, called the fundamental.

The simple harmonic model does not adequately describe the vibrational motion of chemical bonds because at one extreme atoms repel one another the closer they approach and at the other extreme the molecule dissociates. For this more realistic model, transitions from energy level 0 to 2, 0 to 3 etc. become possible and these are referred to as first, second etc. *overtones* with frequencies corresponding approximately to twice, three times etc. that of the fundamental. The frequencies of many overtone bands are in the NIR region. The greater the deviation of a particular bond system from harmonicity, the greater the intensity of overtone bands and the most anharmonic bonds are those involving the lightest atom, hydrogen, and a heavier atom such as carbon, nitrogen or oxygen.

While each frequency of vibration of a complex molecule generally involves oscillation of all the atoms, those atoms undergoing large amplitude motions are often limited to a particular functional group, e.g. $-OH$, $-NH_2$ etc. The characteristic frequency is similar for the same functional group where it appears in different molecules. This concept of *characteristic group frequencies* is the basis of analysis by infrared (IR) and, because each fundamental will be associated with its own potential energy curve and its own set of overtones, of analysis by NIR also.

The number of bands a given molecule will produce can be calculated from the number of coordinates required to describe the total momentum of the molecule. $3n$ coordinates describe the location in space of a molecule of n atoms of which 3 describe the translational and 3 the vibrational motion (2 if linear) of the molecule. There are therefore $3n - 6$ ($3n - 5$ if linear) possible *modes of vibration* but some of these may be the same. Furthermore, absorption can only occur if the molecular vibration is accompanied by a change in dipole moment such as with the hydrogen chloride molecule.

Another type of absorption band occurs in the NIR region. *Combination bands* arise by interaction of two or more vibrations taking place simultaneously and the frequency of a combination band is simply the sum of multiples of the relevant fundamental frequencies. Thus, a triatomic molecule with 3 fundamental bands could have 3 first overtones, 3 second overtones etc. and 6 combination bands. It soon becomes obvious that NIR spectra are much more complex than IR spectra. Nevertheless, overtones and combinations can still be assigned to group frequencies.

2 INSTRUMENTATION

Successful application of NIR depends on the correct choice of instrument. There are three different types available (described in detail by Osborne et al.[1])

2.1 Monochromators

Grating monochromators are used to measure the full visible and NIR spectrum and may be used in transmittance or reflectance mode. They are therefore the most versatile instruments. Monochromators are generally used for research or when a wide range of different applications is required. Three different detectors are available: a silicon detector covers the range 400–1100 nm, an indium gallium arsenide covers the range 800–1700 nm and a lead sulfide the range 1100–2500 nm. Some instruments contain both silicon and lead sulfide detectors (giving them a wavelength range 400–2500 nm), some only the lead sulfide. The latter would be restricted to measurement of powdered or granular samples by diffuse reflectance (see section 3.2) or transmittance via fiber-optic probes. At least one model of NIR monochromator has been designed for on-line use via either fiber-optics or a powder sampler.

Another type of dispersive monochromator used in NIR instruments is the acousto-optically tunable filter (AOTF). An AOTF comprises a crystal of TeO_2 through which a plane travelling acoustic wave is generated at right angles to the incident light beam. This causes the crystal to behave as a longitudinal diffraction grating with a periodicity equal to the wavelength of sound across the material. The main advantages of AOTF over grating instruments is their mechanical simplicity (i.e. no moving parts) and their wavelength stability.

2.2 Diode Array Spectrometers

Diode array spectrometers employ an array of IR-emitting diodes. These function as both the light source and the wavelength selection system. Diode array instruments typically cover the range 400–1700 nm. They have the advantages that the measurement is very fast (e.g. one spectrum per second) and noninvasive. These features are particularly useful where a high sample throughput or ultra-rapid on-line measurements are required.

2.3 Filter Instruments

The simplest and cheapest NIR instruments are based on a limited number, usually between six and twenty, of interference filters. These filters are chosen to represent the absorptions used for the most popular applications, e.g. protein, moisture and oil in agricultural samples. Filter instruments are designed for a limited range of routine analyses, either in the laboratory or on-line.

3 SAMPLE PRESENTATION

The wide range of applications of NIR in food analysis is possible because of the different sample presentation techniques. There is a technique available for any type of liquid, slurry, powdered or solid sample.

3.1 Diffuse Transmittance

Radiation interacting with a sample may be absorbed, transmitted or reflected. In the classical spectroscopy experiment, reflection is eliminated so that the proportion of radiation attenuated by the sample may be measured as transmittance. *Beer's law* then defines a proportionality between transmittance and the product of concentration of the absorbing species and path length. For a clear transparent liquid sample such as beer, hot starch melts, wine or vegetable oil, the path length may be fixed by means of a static or flow-through sample cuvette or a pair of fiber-optic probes and a calibration developed using samples of known concentrations. For example, Halsey[2] used standard solutions of ethanol in water to develop a calibration for alcohol content of beer. It should be noted that, owing to the relatively weak intensities of NIR absorption bands, samples such as vegetable oils may be analyzed without dilution in a solvent.

Beer's law is only valid in the absence of light scatter in the sample. Scattering changes the path length through which the radiation passes and, because the amount of scattering varies from sample to sample, the path length cannot be defined. This type of experiment is known as *diffuse transmittance* (Figure 3), the most well-known example of which is liquid whole milk. The fat globules in the milk scatter light in the manner shown and invalidate Beer's law. Diffuse transmittance measurements are usually carried out in the 800–1100 nm region of the spectrum where the weak absorptions enable useful data to be obtained using thicknesses of 1–2 cm of samples such as meat, cheese or whole grain. Near-infrared transmittance (NIT) instruments are particularly applicable to the analysis of whole grains and a typical apparatus is shown schematically in Figure 4. A sample of grain is placed in a hopper from where aliquots are dispensed into the measurement chamber. When analysis is complete, the sample is discharged into a collection tray. This arrangement lends itself well to adaptation for on-line measurements (see section 5.1).

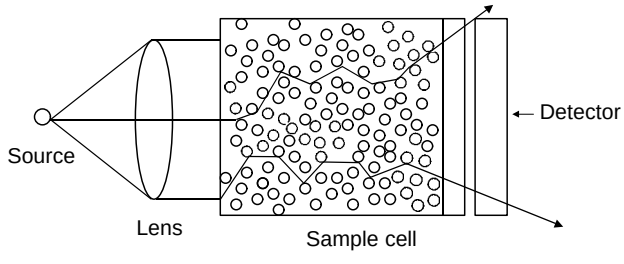

Figure 3 Diffuse transmittance spectroscopy.

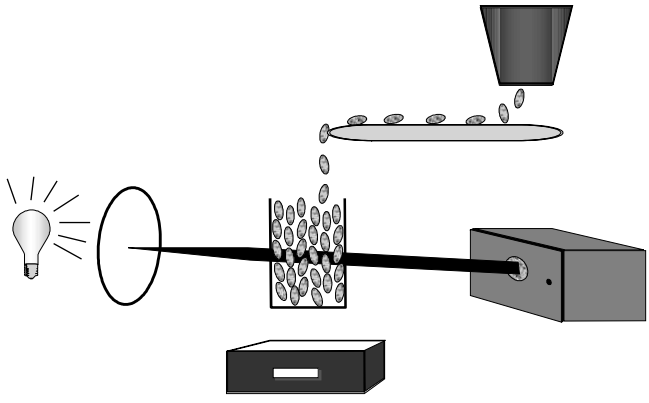

Figure 4 Diffuse transmittance spectroscopy of a whole grain sample.

3.2 Diffuse Reflectance

For a smooth surface such as glass, most of the radiation is reflected from the surface by regular or *specular reflection* and no absorption takes place. In the 1100–2500 nm region, the amount of scattering makes the path length so great that transmittance through 1 cm of most samples is negligible. This situation is called *diffuse reflectance* because most of the incident radiation is reflected. If a matt surface reflects diffusely without penetration into the sample, like specular reflectance no absorption takes place. If, however, some of the radiation penetrates the surface when it reaches each particle it can be reflected, absorbed or transmitted. The net result is that the diffusely reflected radiation (R) can be empirically related to concentration (c) in an analogous way to Beer's law i.e. $\log 1/R = kc$ where k is a factor which incorporates both absorbtivity and path length.

A typical NIR diffuse reflectance experiment is shown in Figure 5. A powdered sample such as wheat meal is packed into a 1 cm-deep sample cell and covered with a quartz window. The sample cell is then placed into the instrument where it is illuminated with NIR radiation and the reflected radiation measured by either a set of detectors set at 45° to the incident beam or a single detector onto which the radiation is focused by means of an integrating sphere. This configuration has been adapted for on-line measurement using a powder sampler for flour or milk powders.

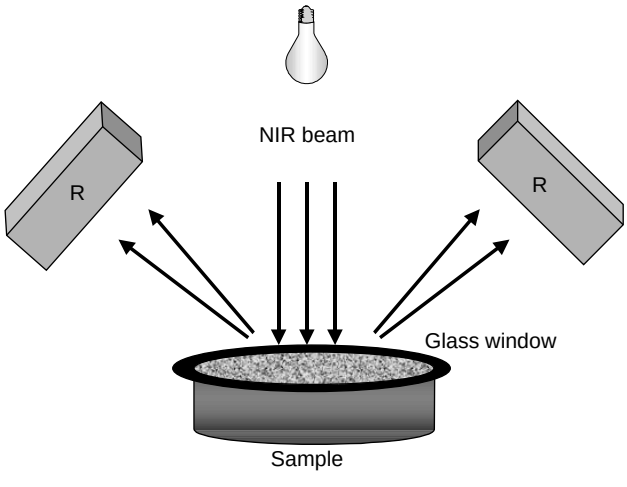

Figure 5 Diffuse reflectance analysis of a powdered sample.

The diffuse reflectance experiment can be adapted for use with liquids by placing a ceramic tile beneath the sample. In this sample presentation mode, known as *transflectance*, the radiation is transmitted through the sample, reflected from the ceramic then transmitted back through the sample before finally reaching the detector. Transflectance is thus a hybrid of transmittance and reflectance. *Interactance*, another hybrid of transmittance and reflectance, involves illumination and detection at laterally separated points on the sample's surface. It is normally accomplished using a fiber-optic probe in which one set of fiber-optic bundles carries the incident radiation and another carries the reflected radiation. This type of arrangement is particularly useful for large samples such as intact fruit.

3.3 On-line Samplers

Ideally, a sensor needs to be sufficiently compact and robust to be placed in the required position on-line. It needs to be relatively low cost, rapid and capable of measuring the desired analyte or property on an intact sample. Finally, it must be capable of accurate and stable calibration. NIR instruments meet these requirements. Since the applications are limited and predefined, filter instruments which fulfil the criteria of compactness, robustness and cost can usually be employed. NIR is inherently a rapid technique (as fast as one spectrum per second with diode array technology) and data from many measurements can be combined into a moving average so as to identify trends.

There are three types of NIR on-line analyzer.

3.3.1 Remote (Noncontact) Sensor

The first dedicated on-line NIR sensor was invented by Edgar and Hindle of Infrared Engineering Co in the UK. This is based on a sensing head, which is referred to as a ''gauge'', remote from the sample. This design has the advantages of a low cost of instrument and simplicity of installation but imposes severe constraints on instrument design as it is susceptible to interference from ambient light variations, dust build-up on the optical surfaces and atmospheric humidity variation. The Infrared Engineering gauge was therefore designed specifically for on-line application and constructed to

minimize such potential interferences. The Infrared Engineering MM55 Gauge is based on a non-contacting sensing head fixed approximately 200 mm from the product flow by mounting to the side of a section of spouting inclined at an angle of 60° from the horizontal. The gauge focuses light from a quartz halogen lamp into a parallel beam which is projected onto the product via a filter wheel which transmits narrow bands of energy at the desired NIR wavelengths. The reflected energy, which has been modulated by interaction with the product, is captured by a detector in the gauge. In this way, measurements on the moving flour stream are taken through a toughened glass window at a rate of up to five per second. Signals from the MM55 gauge are conveyed to an electronic processing unit which contains the calibration equation and produces the analytical result. This is passed to a strip recorder to provide a permanent record. Applications of the MM55 in the food industry include moisture determination on both continuous samples such as powders and discontinuous samples such as whole biscuits, control of protein in flour, monitoring of fat content of potato crisps and sorting of fruit.

Recently, in Australia, a novel application of NIR spectroscopy to the monitoring of bread dough development has been invented by Wesley et al.[3] The invention makes use of the high scan speed (one spectrum per second) of the Perten DA-7000 diode array instrument. In this case, the instrument was inverted over an open mixing bowl and the signal from the dough was recorded continuously without stopping the mixer. In this way, using second derivative data at defined wavelengths assigned to protein and water, an NIR mixing curve was obtained. The NIR mixing curves were in good agreement with curves obtained using measurement of mixer torque over a range of flour types and mixer conditions.

3.3.2 By-pass Sampler

To control the composition of the grist (wheat mixture) during the blending process, it is necessary to perform measurements on the grain itself because the time delay between blending the wheats and producing the final flour is too long for an effective feedback control loop. Whole grain NIR instruments generally employ a hopper and flow-through cell and are thus relatively easy to adapt to on-line use. An on-line version of the Infratec 1225 Whole Grain Analyzer is commercially available.

One of the first near-infrared on-line powder samplers (NIROS) was based on an InfraAlyzer 300B instrument which was split into two parts comprising the electronics unit which would be removed from the immediate vicinity of sampling and the optical unit which was to be modified for on-line sample presentation. This method of sample presentation was selected on the principle that the sample needed to be held stationary against the optics window during the measurement cycle and that the flour surface should be smooth, continuous and firmed under pressure. These conditions were met by taking direct samples from the flour stream, either from gravity-fed spouting or a positive pressure blowline, into a mechanism designed to reproduce the action of a human operator in packing flour into a sample cup.

NIR on-line samplers were first developed to measure the protein content of flour. This is still the most popular application and provides an excellent example of an NIR feedback control system. Dried gluten is commonly used, particularly in Europe, to replace wholly or partly the protein in flour which would otherwise be derived from high-protein wheat in the grist. The success of on-line NIR for monitoring flour protein content therefore led to its incorporation into a closed-loop control system for gluten addition to flour.[4] A mixing screw is installed between the gluten feed and the NIR sampler station from which a feedback signal controls the gluten feeder. The system has proved to be an efficient and accurate method of control of gluten addition to achieve a target protein content in the flour.

3.3.3 Fiber-optic Probe

Fiber-optic probes have the widest range of applications in on-line food analysis. Uses include dairy products, meat, fruit, beer and extruded snack foods.

Extrusion cooking is used to manufacture a wide range of products from cereals, including breakfast cereals, snack foods and pet foods. In these products, texture and density are key quality requirements which are controlled by the processing conditions summarized by the term ''degree of cook''. Conventionally, control is attempted by measurement of inputs such as the moisture content, screw speed, barrel temperature, etc. on the basis that the relationship between these and the product quality is understood. However, a more direct means of control is highly desirable. In contrast to breadmaking, where the product structure is largely controlled by protein development, in extrusion cooking it is the starch component of the cereal-based raw material which is the chief functional ingredient. Therefore, the aim was to use NIR to monitor changes in starch structure. Initially, products were prepared over a wide range of processing conditions and the extrudates freeze-dried and ground to powders prior to NIR spectroscopy. Since successful calibrations based on spectral features related to starch structure were obtained using the powdered samples,[5] the measurement was attempted on-line. A fiber-optic reflectance probe was installed in the die of the extruder and the light was transmitted through the melt and reflected back from a polished steel pillar;

in effect, therefore, a transflectance measurement was used. The spectral characteristics of the different forms of starch found in the hot melts close to the die of the extruder were identical to those of the powdered extrudates. This research has opened the way to an on-line process monitoring and control system for extrusion cooking.

4 CALIBRATION DEVELOPMENT

In NIR analysis, the calibration provides all three basic functions of analytical chemistry i.e. separation, identification and quantification. Therefore, derivation of the calibration equation for each application is the most crucial step in reliable and accurate analysis. There is, however, no unique method for deriving satisfactory calibration equations. The following sections provide an overview of the main techniques used and more detail can be found in Osborne et al.[1]

4.1 Data Preprocessing

Quantitatively the most important factor in the NIR reflectance spectrum is the particle size of the sample. $\text{Log}(1/R)$ increases with increasing particle size because the apparent path length becomes longer. Therefore, for example, the same wheat sample ground to different particle sizes will result in substantially different spectra. However, the effect is not additive but multiplicative, i.e. proportional to $\log(1/R)$. Since in reflectance spectra of foods and agricultural materials $\log(1/R)$ coincidentally increases with increasing wavelength, the effect of particle size appears to be a function of wavelength.[6,7]

Correction for particle size is accomplished empirically by multiple regression mathematics but many spectroscopists prefer to use derivatives. The first derivative

$$\frac{d(\log 1/R)}{d\lambda}$$

is the slope of the spectrum at wavelength λ and is calculated as the difference between $\log 1/R$ at two adjacent wavelengths or, more usually, two segments in which the data are smoothed (averaged) over a small region of the spectrum defined by the segment size. The difference between these wavelengths or segments is called the gap. The second derivative

$$\frac{d^2(\log 1/R)}{d\lambda^2}$$

is the difference between two adjacent first derivatives. Derivative mathematics is described using the notation (derivative order, gap, segment, second smoothing) e.g. (2, 8, 8, 1) means a second derivative with a 16 nm gap and 16 nm segment.

Derivatives have two useful properties:

1. resolution of overlapping bands;
2. deconvolution of background (derivative ratios[6] correct for multiplicative effects).

The effect of particle size can also be overcome by multiplicative scatter correction[8] or standard normal variate and detrend.[9] The multiplicative scatter correction algorithm proceeds by computing the mean spectrum of the sample set and rotating each sample spectrum so that a regression line drawn through it aligns with a regression line drawn through the mean spectrum.

4.2 Sample Population Structuring

The NIR analysis of food and agricultural samples is based on calibration against a reference method using a set of samples which represent all the variability in the population from which they are drawn. The problem becomes one of identifying these samples. One set of criteria would be to obtain analytical data on a large database of samples and select those with the most variation in composition. However, this is an expensive option. The concept of population structuring was introduced by Shenk and Westerhaus.[10] The idea is to assemble a library containing the NIR spectra of a large number of samples and use discriminant analysis techniques to define the boundaries of the library. The spectra of further samples can then be examined to ascertain whether they belong to the population represented by the library and whether their spectra are different to those of existing samples. The spectral library can be used in several ways:

- to define the optimum calibration set;
- to verify that test samples match the calibration;
- to produce locally weighted calibrations for optimum measurement accuracy;
- to identify new samples with which to extend the calibration.

4.3 Regression Techniques

The usual method of calibration of NIR instruments is by multiple linear regression of the reference data on the spectral data. The difficulty in this is twofold. Firstly, NIR spectra contain a large amount of data whereas only a few terms should be included in a regression equation. Secondly, because of the effect of scatter, the raw NIR data are highly intercorrelated and this precludes the use of forward stepwise regression, the most common method of selection of terms in

multiple linear regression. For fixed filter instruments, the wavelengths for each application may be known. If not, there are only a few from which to select and is feasible to employ an all possible combinations strategy. For monochromator or diode array instruments on the other hand, more sophisticated methods are necessary. The use of mathematical pretreatments such as derivatives or multiplicative scatter correction (see section 4.1) would make forward stepwise selection more reliable but it is important to ensure that the selected wavelengths are consistent with known chemical absorptions. This can be done by comparing a plot of correlation versus wavelengths (Figure 6) with the spectrum of the constituent of interest.[11] However, this is not always possible. For example, functional properties do not usually have a defined spectral pattern. An increasingly popular approach, therefore, is to reduce the amount of data by principal components analysis then perform linear regression on the principal components. If the principal components are selected with regard to minimizing the correlation coefficient, the method becomes partial least squares (PLS). The end result is a calibration equation from which the constituent of interest may be calculated from a (usually) linear combination of spectral data. The equation has associated statistics which define the closeness of fit of the actual and predicted values to the least squares line. It is always important to plot a scatter graph as illustrated in Figure 7 in order to detect any aberrant data. Ideally, the scatter plot should contain data points distributed evenly about the line but within the confidence limits as shown in the example.

4.4 Qualitative Analysis

In applications such as authenticity testing (see section 5.8), where the aim is to classify rather than perform a quantitative determination, a qualitative or discriminant

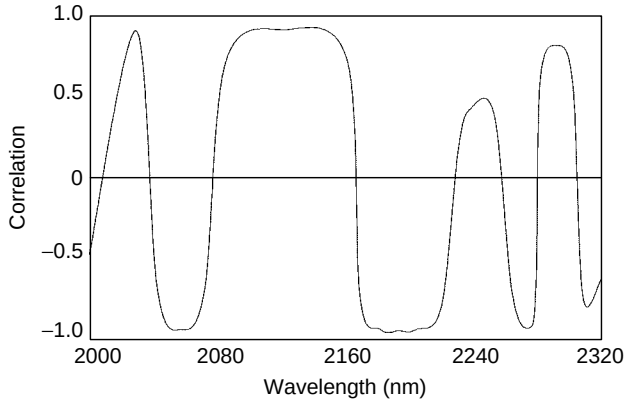

Figure 6 Plot of correlation versus wavelength for a set of calibration samples.

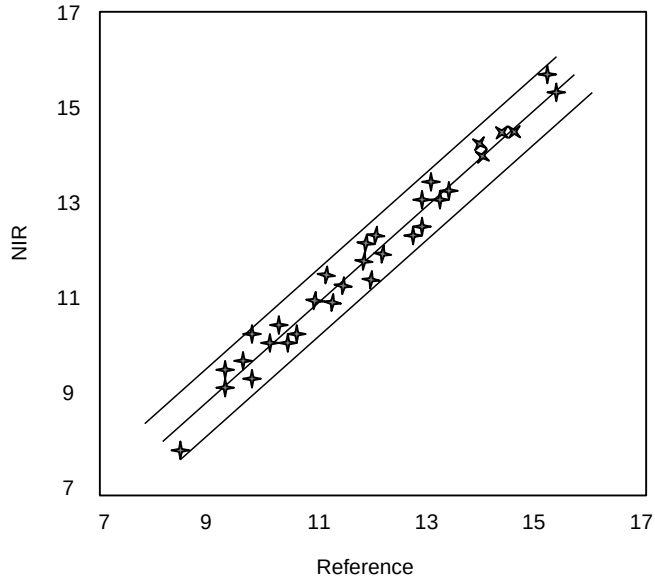

Figure 7 A typical scatter plot of NIR versus reference data.

analysis approach is often used. The mathematical approach to this problem is to develop a model which defines the mean and standard deviation of each sample type in multidimensional space then test to which group an unknown sample belongs. This test is based on a non-Euclidean distance measurement such as the Mahalanobis distance. The Mahalanobis distance is defined according to elliptical contours spreading out from the group centres. The mathematics and applications of discriminant analysis are described by Mark.[12] More sophisticated classification models have been achieved using artificial neural networks (ANNs). ANNs consist of input nodes which receive one piece of spectral data each and distribute it to a hidden layer of nodes where the data are transformed before being distributed again to a set of output nodes. The nodes simply provide the mechanism for distribution of the data through the hidden layer which may be considered a nonlinear function approximation machine. The crucial feature of ANNs is that they are adaptive. One example of the application of ANNs to a food classification problem has been published by Hervas et al.[13]

5 FOOD APPLICATIONS

5.1 Cereals and Cereal Products

NIR has been used for the quality testing of crossbred material from wheat breeding programs since the late 1970s. Because NIR is rapid, low cost per test and requires a relatively small quantity of sample, it is routinely used throughout the world to determine the protein content

of wheat to allow breeders to screen large numbers of lines for this key characteristic. In addition, NIR is widely used to replace some of the chemical tests necessary as part of a quality testing program (Table 1). In particular, NIR hardness and moisture measurements may be used to determine the conditioning requirements prior to test milling while protein and moisture may be determined on the resulting flour.

The ultimate application of NIR analysis in a wheat breeding program would be direct prediction of functional quality including flour yield, damaged starch, water absorption, dough development time, extensibility and loaf volume. Although some breeding programs are using NIR to eliminate lines according to some of these measurements, they have not gained widespread acceptance. The reason appears to lie in the sensitivity of calibration to changes in variety, growing location and season, which necessitates frequent re-calibration. The availability to breeding programs of whole grain instruments provides a renewed incentive to develop the use of NIR to take the place of the entire procedure for quality screening of early generation lines where the numbers and quantity of samples preclude the use of the traditional methods. A recent development in instrumentation is the diode array spectrometer as exemplified by the Perten Model DA-7000. This type of instrument is quick to load as it requires no sample cell, and spectral data are acquired very rapidly; typically, an analytical result can be obtained in 15 s.

NIR is used in Australia to predict optimum fertilizer requirements of cereal crops by analysis of total nitrogen and carbohydrate in plant tissue samples. The tissue test for rapid determination of shoot nitrogen status in cereal crops was first developed for rice. The test has since been extended to encompass wheat shoot nitrogen and fructans. The tissue testing system is based on a plant sample taken by the farmer at a specific growth stage (Zadok scale) which is dried and analyzed by NIR. The nitrogen and fructan results are then entered into a database which is used to determine the appropriate fertilizer recommendation.[14]

The use of NIR technology to help farm managers understand the fertility of their crops and to segregate or blend grain on its protein content prior to delivery is gaining in popularity. In many countries, the price of wheat is determined by its protein content, often with substantial price increments between grades. This has encouraged some farmers to blend wheat on farm according to its NIR protein to increase their deliverable tonnage of higher priced grain. Providing the farm's instrument has been properly calibrated and is well maintained, this strategy should be financially successful. Some farmers are also realizing that NIR has the potential to analyze a large number of samples in a short period of time which could enable them to assess their crop for optimum development or to map differences in fertility across fields.

Table 1 NIR applications in quality testing of wheat breeding lines

Constituent	Wheat	Flour
Protein	✓	✓
Moisture	✓	✓
Hardness	✓	
Water absorption		✓ (indicative)

NIR has been used in Australia, Canada, Europe and the USA to monitor growers' deliveries of wheat and barley for two decades, and some of the first research papers, e.g. Williams,[15] described the calibration of instruments for this purpose. Canada was the first grain-producing country to utilize NIR for all wheat protein testing in its segregation program. As well as increased efficiency of the Canadian wheat segregation program, the adoption of NIR testing has resulted in total cost savings to date of CAN$2.5 m. Furthermore, the replacement of the Kjeldahl system, which involves concentrated sulphuric acid and heavy metal catalyst, by the chemical-free NIR method has enabled Canada to lead the way in environmentally safe large-scale protein testing of grain. The Canadian Grain Commission now uses NIT exclusively for providing quality data in support of the marketing of about 50 million tonnes of grain annually.

By 1980, NIR protein testing had become the official method of the US Federal Grain Inspection Service (FGIS) and since then has been in continuous routine use in wheat segregation in the USA. In the USA, following the Grain Quality Improvement Act 1986, hardness measurements have been introduced as part of the wheat grading system. Wheat hardness has been measured in both meal and whole grain. Since there is no international standard reference method for wheat hardness, a definition based on NIR spectra of ground grain has been proposed by Norris et al.[16]

Delwiche et al. in the USA[17] have carried out a systematic study to develop a wheat classification system for US market class based on NIR protein and hardness measurements on single kernels. This concept is being developed by the FGIS into a practical system for the objective, rapid and sensitive detection of mixtures of wheat classes.

In Europe in the early 1980s, flour millers encountered a need to stimulate the cultivation of higher tonnages of high protein wheat by payment of premiums for protein content. A rapid method of load-by-load testing which produced results acceptable to both farmers and millers became necessary in order to calculate the price on delivery to the mill. NIR was adopted for this purpose in

a number of European countries including UK, Ireland, Germany and Sweden.

Initially, instruments based on reflectance measurements on ground grain were used for testing grain at harvest and these were shown to result in calibrations for protein which were stable over growing locations and seasons and transferable between instruments.[18] By the mid 1980s, NIT instruments had become available. These instruments typically employ a hopper, into which a sample of whole grain is placed, to dispense the sample into the measurement compartment where it is analyzed by transmittance spectroscopy (Figure 4). Usually, 4–20 subsamples are analyzed sequentially and the results averaged to obtain a more representative result for the sample. The ability to perform protein testing on whole grain eliminates the need for grinding, thus simplifying and speeding up the analysis. Protein calibrations for whole wheat and barley have been shown to perform with equal accuracy to those derived for the ground grain.[19] However, the criteria for accuracy at receival are more stringent than those for selection of breeding lines. This is because grower payments are determined according to the NIR protein content. In some countries, the payment scale is not continuous but there are cutoffs between different grades where there may be substantial price differentials. Consequently, although the accuracy of NIR calibrations are normally assessed over a wide range, in this case the scatter around a much smaller range becomes important.

In grain trading it is highly desirable to eliminate between-instrument variation in analysis results so that the farmer receives the correct grading and thus fair payment irrespective of the receival silo delivered to. This can be achieved by networking so that small differences in the response of individual instruments are corrected for in the calibration. All instruments in the network are connected to a central computer by modems, allowing simultaneous and uniform calibration and update of all instruments in the network. In 1995, all the European networks were brought together into ''The European Grain Network'' with the aim of standardizing NIR grain testing throughout the continent.

Paddy rice must be harvested at the optimum maturity to obtain high yield and quality. This is assessed by the moisture content of rice at receival. NIR reflectance instruments are unsuitable for use in rice receival testing because of the high moisture content at which rice is harvested. Following the successful development of NIT calibration for whole rice grain, Australian rice growers have adopted this technology for moisture testing on paddy rice.

The use of NIR technology to determine the protein and moisture contents of both wheat and flour is now routine practice in flour mills worldwide. It is used for testing each delivery of wheat in order to make decisions about acceptance, price and binning; for determination of conditioning time from measurement of hardness; and for analyzing flour to check that it complies with specifications before shipment to the customer. However, the power of NIR goes far beyond mere replacement of wet chemical analysis for raw material and product testing. With the introduction of an NIR instrument into the mill, the laboratory was able for the first time to make a direct contribution to controlling the process in real time. So useful has NIR proved to be for this purpose that in many mills it soon became moved from the laboratory into the mill itself where it could be used off-line by shift millers for process monitoring whenever the mill is running. The success of NIR for this purpose gave it the necessary credibility as a potential on-line sensor in flour mills (see section 3.3).

Many large plant bakeries employ NIR to monitor the quality of their flour and other raw ingredients. Typically, flour would be tested for protein and moisture (possibly, but less commonly, color and water absorption). In biscuit and confectionery bakeries, other ingredients such as sucrose, chocolate etc. may also be analyzed. Having verified the composition of ingredients against their quality specifications, the next point of control of a bakery process is at the dough mixing stage. Detection of faults in the dough avoids energy wastage in baking an out-of-specification product and, for biscuit doughs, may allow the dough to be re-cycled. NIR has been used for compositional analysis of intact biscuit and bread dough pieces and for monitoring the progress of bread dough mixing.

NIR calibrations for compositional analysis of short-sweet biscuit doughs enabled the simultaneous detection of errors in fat of 5% relative to the total amount of fat in the recipe and gross errors in the water, sucrose and flour percentages.[20] NIR has also been used to determine the protein and fat contents in bread dough.

During baking, the moisture content of doughs is reduced from about 18% to 2% for biscuits and 60% to 40% for bread. Monitoring the changes in the moisture content during baking allows the oven performance to be optimized. However, this is not easy because there is a non-uniform distribution of water content through both the dough and the freshly-baked product.

For biscuits, an Infrared Engineering MM55 Gauge (see section 3.3) is commonly used to monitor the moisture contents of biscuits on a conveyer as they leave the oven. This measurement depends on a reproducible relationship between the gauge reading which represents the surface moisture content in the freshly-baked biscuit and the bulk moisture content of the fully equilibrated product in packet and on strategies to compensate for the discontinuous nature of discrete products on a conveyer.

Table 2 NIR applications in the analysis of cereal products

Product	Moisture	Protein	Fat	Starch	Sugars	Fiber
Bread	✓	✓	✓	✓	✓	
Biscuits			✓		✓	
Cake mixes			✓		✓	
Breakfast cereals			✓			✓
Pasta	✓	✓	✓			
Snack foods				✓		✓

NIR is applicable to the analysis of moisture, protein, fat, starch, sugars and fiber in intact cereal foods such as bread, biscuits, cake mixes, breakfast cereals, pasta and snack foods (Table 2).

NIR has been applied to the measurement of the rate of staling of white bread crumb without the use of a reference method.[21] Rate constants were calculated directly from plots of second derivative NIR data versus time. This is a rare example in food research where NIR has been used as a primary method.

5.2 Milk and Dairy Products

Milk analyzers based on mid IR spectroscopy have been available since the 1960s but it was not until much later that NIR began to be used in the dairy industry. Although the mid IR is satisfactory for the proximate analysis of liquid milk, dairy chemists are faced with one of the widest ranges of sample types in the food industry. NIR has a key role in the analysis and process control of dairy products. It offers flexibility in the analysis of protein, moisture, fat and lactose contents in a wide range of dairy products including:

- liquid milk
- dried whole milk, skim milk and whey powders
- cream
- traditional and processed cheese.

Many of these products are emulsions whose sampling for classical chemical analysis is difficult. For example, blending such samples changes their physical characteristics. NIR offers the possibility of on-line analysis which avoids the need for batch sampling and minimizes sampling error by averaging of virtually instantaneous, continuous measurements. Milk powders are analyzed on-line using a powder analyzer which enables the moisture content to be controlled. For most other on-line applications, fiber-optic probes are used. These have the advantages of minimal maintenance owing to the absence of moving parts, robustness to high temperatures and pressures and the survival of clean-in-place protocols involving the use of caustic chemicals. Two examples of on-line NIR fiber-optic measurement in dairy processing are moisture control of cream cheese and processed cheese. Cream cheese is manufactured in a series of set tanks from which product is transferred into a separator. NIR monitoring of the product from the separator enables compensation for the variation in characteristics from different set tanks. NIR measurements of processed cheese have been used for process diagnostics in which a greater understanding of the effect of temperature on the final moisture content has been gained.

The extent of development of NIR applications in the dairy industry is such that it is possible to purchase an instrument which is factory-calibrated for proximate analysis of cheddar cheese. However, these calibrations are only valid for cheddar made to the traditional recipe and sampled directly from the vat; non-traditional and matured cheddars require different calibrations.

5.3 Meat

NIR spectroscopy is widely used in the meat industry for proximate analysis. A special interactance fiber-optic probe has been designed to spear carcasses and determine their fat content. This enables the carcasses to be sorted prior to butchering.

Dedicated instruments are available to determine the protein, fat and moisture contents of ground meat and meat products and factory-set calibrations are available for cooked meat, cooked ham, liver sausage, and pepperoni. Meat samples are minced then blended in a food processor before being packed into an open sample cell. The amount of sample needs to be controlled either by mass or by depth.

Isaksson et al.[22] have used the MM55 on-line gauge (see section 3.3) for proximate analysis of ground beef directly at a meat grinder outlet. NIR has also been used for sensory analysis of sausages and to discriminate between fresh and frozen-then-thawed beef.

5.4 Fish

The analysis of farmed atlantic salmon and rainbow trout enables the optimization of their processing. NIR fiber-optic interactance probes can be used to measure nondestructively the protein, moisture and oil contents of whole fish.[23]

5.5 Fruit and Vegetables

Fresh fruit and vegetables are graded by shape, size and color. Objective, nondestructive methods of sorting enable growers and packers to market a consistent product over an extended season. In addition, high-quality produce can be segregated for high-return markets while

lower-quality produce can be identified for processing or other uses.

The nondestructive sorting of fruit for ripeness by optical spectroscopy was originally carried out in the visible region. However, the availability of fiber-optic interactance probes led to a resurgence of research interest in this application via the determination of sugar content. The reason for developing a direct measure of sugar content of fruit is that appearance is not a reliable guide to sweetness. Japanese researchers[24,25] have demonstrated the successful application of NIR to the on-line determination of sugar content in intact peaches and mandarins and developed an automated fruit sorting machine based on this principle. Australian scientists have extended this concept to tropical fruits such as melons, mangoes and pineapples.

5.6 Confectionery

NIR applications in the confectionery industry include the determination of moisture in granulated sugar and chocolate crumb, protein and oil in cocoa powder and fat in whole chocolate. NIR data have also been shown to correlate highly with sensory data on raw and roasted cocoa beans, chocolate mass and finished chocolate.

5.7 Beverages

When barley is germinated under controlled conditions, it undergoes a series of complex biochemical reactions which result in its conversion to malt. The malt is then mashed with water to produce a liquid called wort which is then fermented to produce beer. Since malting takes about twelve days, barley breeders need a rapid means of predicting the malting quality of barley grain. Calibrations have been developed for nitrogen, lysine, β-glucan and malt hot water extract. As with wheat functional quality, NIR analysis of barley is used to predict the quality of material which is derived by processing of the grain and there is doubt whether tests on ungerminated grain fully account for the interaction of enzymes and substrates during germination. Attempts have been made to improve the accuracy of prediction by analysis of key quality characteristics of the malt or the wort (Table 3).

In the brewing industry, NIR is widely used to monitor the original gravity and alcohol content of beer using on-line flow-through cells. Standard errors of the calibration for alcohol of 0.1–0.2% have been reported[2] using either transmittance or transflectance cells. The Liquidata probe has been used to achieve a process guarantee on any beer to better than $\pm$0.04% alcohol.

NIR is used to monitor fruit quality and determine the alcohol content of wine and dedicated filter instruments for wine analysis are commercially available.

Table 3 NIR applications in quality testing of barley breeding lines

Constituent	Barley	Malt	Wort
Total nitrogen	✓	✓	
Moisture	✓	✓	
Lysine	✓		
β-Glucan	✓		
Extract	✓	✓	✓
Soluble protein			✓
Free α-amino N			✓

Other beverage applications include the moisture content of tea and coffee, sugar content of fruit juices and the sensory quality of tea and coffee blends.

5.8 Authenticity

The study of food authenticity involves establishing whether a sample is genuine in terms of its description, including geographical origin. Many analytical techniques have been used to examine the authenticity of a wide range of different foods. The applications of NIR for authenticity testing of coffee, fruit pulps, milk powders, orange juice, pig carcasses, rice, sausages, sugars, vegetable oils, wheat grain and wheatflour have been reviewed by Downey.[26] These applications are based on the principles of discriminant analysis (see section 4.4) in which the problem is to compare the spectrum of the test sample with a reference library (see section 4.2). The library should contain examples of known authentic and adulterated samples. In the simplest case, the question is simply whether the test sample belongs to one of two populations such as ''Basmati rice'' or ''not Basmati rice''. Often, however, it is required to classify within a series of possible classes (e.g. different vegetable oils or wheat varieties), to identify a particular kind of adulteration (e.g. medium invert beet sugar, pulp wash or zest oil in orange juice) or to quantify an adulteration (e.g. non-dairy fat in milk products).

6 THE LITERATURE

Scientific papers describing the practical applications of NIR in food analysis are spread widely throughout the literature. As evidenced by the key references cited in this article, such papers may be found in journals devoted to analysis, spectroscopy, general food science or more specialized areas of food science. This makes it particularly difficult to search the literature. However, *NIR News* (NIR Publications, Chichester, UK) has since

1990 included a references section in an attempt to overcome this problem.

The only scientific journal devoted exclusively to NIR spectroscopy is *Journal of Near Infrared Spectroscopy* (NIR Publications, Chichester, UK). Volume 4 was dedicated to Karl H Norris, the founder of modern NIR analysis. It contains historical, review or leading edge papers by many of the pioneers of NIR spectroscopy and provides a particularly valuable introduction to the technique.

The proceedings of the series of international conferences on near-infrared spectroscopy (ICNIRS) provide a valuable addition to the literature on the developments in and applications of NIR spectroscopy:

- J. Hollo, K. J. Kaffka, J. L. Gonczy, *Near-Infrared Diffuse Reflectance/Transmittance Spectroscopy*, Akademiai Kiado, Budapest, 1987.
- C. S. Creaser, A. M. C. Davies, *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, Cambridge, 1988.
- M. Iwamato, S. Kawano, *Proceedings of the Second International Near-Infrared Spectroscopy Conference*, Korin, Tokyo, 1989.
- R. Biston, N. Bartiaux-Thill, *Proceedings of the Third International Conference on Near-Infrared Spectroscopy*, Agricultural Research Centre Publishing, Gembloux, Belgium, 1990.
- K. I. Hildrum, T. Isakksson, T. Naes, A. Tandberg, *Near-Infrared Spectroscopy. Bridging the Gap between Data Analysis and NIR Applications*, Ellis Horwood, Chichester, 1992.
- G. D. Batten, P. C. Flinn, L. A. Welsh, A. B. Blakeney, *Leaping Ahead with Near-Infrared Spectroscopy*, Royal Australian Chemical Institute, Melbourne, 1995.
- A. M. C. Davies, P. C. Williams, *Near-Infrared Spectroscopy: The Future Waves*, NIR Publications, Chichester, 1995.

There are only two books currently in print devoted to NIR analysis:

- D. A. Burns, E. W. Ciurczak, *Handbook of Near-Infrared Analysis*, Marcel Dekker Inc., New York, 1992.
- B. G. Osborne, T. Fearn, P. H. Hindle, *Practical NIR Spectroscopy*, 2nd edition, Longman, Harlow, 1993.

7 STANDARD METHODS

The following standard methods of NIR analysis of foods have been adopted:

American Association of Cereal Chemists (AACC)

Method 39-01. General near-infrared instrumentation and techniques. First Approved: 1989.

Method 39-10. Near-infrared reflectance method for protein determination. First Approved: 1982.

Method 39-11. Near-infrared reflectance method for protein – wheat flour. First Approved: 1985.

Method 39-20. Near-infrared reflectance method for protein and oil determination – soybeans. First approved: 1991.

Method 39-21. Near-infrared method for whole-grain analysis. First Approved: 1991.

Method 39-25. Near-infrared method for protein content in whole-grain wheat. First Approved: 1995.

Method 39-70A. Wheat hardness as determined by near-infrared reflectance. First Approved 1986.

Association of Official Analytical Chemists (AOAC)

Official First Action Method 997.06. Protein (crude) in wheat. Whole grain analysis, Near-infrared spectroscopic method. Adopted: 1997.

International Association for Cereal Science & Technology (ICC)

ICC Standard No. 159. Determination of protein by near-infrared reflectance (NIR) spectroscopy. Approved: 1995.

ICC Recommendation No. 202. Procedure for near-infrared (NIR) reflectance analysis of ground wheat and milled wheat products. Approved: 1986.

Royal Australian Chemical Institute (RACI) Cereal Chemistry Division

Method No. 11.01 Determination of protein and moisture in whole wheat and barley by NIR. First Approved: 1998.

ACKNOWLEDGMENTS

The author thanks Mrs S. Ormston for preparation of the original artwork used in this article.

ABBREVIATIONS AND ACRONYMS

AACC	American Association of Cereal Chemists
ANN	Artificial Neural Network
AOAC	Association of Official Analytical Chemists
AOTF	Acousto-optically Tunable Filter
FGIS	Federal Grain Inspection Service
ICC	International Association for Cereal Science & Technology
ICNIRS	International Conferences on Near-Infrared Spectroscopy

IR	Infrared
NIR	Near-Infrared
NIROS	Near-Infrared On-line Powder Sampler
NIT	Near-Infrared Transmittance
PLS	Partial Least Squares
RACI	Royal Australian Chemical Institute

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels

Polymers and Rubbers **(Volume 9)**
Near-infrared Spectroscopy of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. B.G. Osborne, T. Fearn, P.H. Hindle, *Practical NIR Spectroscopy*, Longman, Harlow, 49–78, 1993.
2. S.A. Halsey, 'The Use of Transmission and Transflectance Near-Infrared Spectroscopy for the Analysis of Beer', *J. Inst. Brew.*, **91**, 306–312 (1985).
3. I.J. Wesley, N. Larsen, B.G. Osborne, J.H. Skerritt, 'Non-Invasive Monitoring of Dough Mixing by Near-Infrared spectroscopy', *J. Cereal Sci.*, **27**, 61–69 (1998).
4. T. Fearn, P.I. Maris, 'An Application of Box–Jenkins methodology to the Control of Gluten Addition in a Flour Mill', *Appl. Statist.*, **40**, 477–484 (1991).
5. R.C.E. Guy, B.G. Osborne, P. Robert, 'The Application of Near-Infrared Reflectance Spectroscopy to Measure the Degree of Processing in Extrusion Cooking Processes', *J. Food Eng.*, **27**, 241–258 (1996).
6. K.H. Norris, P.C. Williams, 'Optimisation of Mathematical Treatments of Raw Near-infrared Signal in the Measurement of Protein in Hard Red Spring Wheat. I. Influence of Particle Size', *Cereal Chem.*, **61**, 158–165 (1984).
7. M. Manley, A.E.J. McGill, B.G. Osborne, 'The Effect of Light Scattering on NIR Reflectance and Transmittance Spectra of Wheat', *J. Near Infrared Spectrosc.*, **2**, 93–99 (1994).
8. P. Geladi, D. MacDougal, H. Martens, 'Linearization and Scatter Correction for Near-infrared Reflectance Spectra of Meat', *Appl. Spectrosc.*, **39**, 491–500 (1985).
9. R.J. Barnes, M.S. Dhanoa, S.J. Lister, 'Standard Normal Variate Transformation and De-trending of Near-infrared Diffuse Reflectance Spectra', *Appl. Spectrosc.*, **43**, 772–777 (1989).
10. J.S. Shenk, M.O. Westerhaus, 'Population Structuring of Near Infrared Spectra and Modified Partial Least Squares Regression', *Crop Sci.*, **31**, 1548–1555 (1991).
11. B.G. Osborne, 'Comparative Study of Methods of Linearization and Scatter Correction in Near Infrared Reflectance Spectroscopy', *Analyst*, **113**, 263–267 (1988).
12. H. Mark, 'Qualitative Discriminant Analysis', in *Handbook of Near-Infrared Analysis*, eds. D.A. Burns, E.W. Ciurczak, Marcel Dekker, New York, 329–363, 1992.
13. C. Hervas, A. Garrido, B. Lucena, N. Garcia, E. De Pedro, 'Near-Infrared Spectroscopy for Classification of Iberian Pig Carcasses Using an Artificial Neural Network', *J. Near Infrared Spectrosc.*, **2**, 177–184 (1994).
14. V.B. McGrath, A.B. Blakeney, G.D. Batten, 'Fructan to Nitrogen Ratio as an Indicator of Nutrient Status in Wheat Crops', *New Phytol.*, **136**, 145–152 (1997).
15. P.C. Williams, 'Application of Near Infrared Reflectance Spectroscopy to Analysis of Cereal Grains and Oilseeds', *Cereal Chem.*, **52**, 561–576 (1975).
16. K.H. Norris, W.R. Hruschka, M.M. Bean, D.C. Slaughter, 'A Definition of Wheat Hardness Using Near-Infrared Reflectance Spectroscopy', *Cereal Foods World*, **34**, 696–705 (1989).
17. H. Song, S.R. Delwiche, Y.-R. Chen, 'Neural Network Classification of Wheat Using Single Kernel Near Infrared Transmittance Spectra', *Opt. Engng.*, **34**, 2927–2934 (1995).
18. B.G. Osborne, T. Fearn, 'Collaborative Evaluation of Universal Calibrations for the Measurement of Protein and Moisture in Flour by Near Infrared Reflectance', *J. Food Technol.*, **18**, 453–460 (1983).
19. P.C. Williams, D.C. Sobering, 'Comparison of Commercial Near Infrared Transmittance and Reflectance Instruments for Analysis of Whole Grains and Seeds', *J. Near Infrared Spectrosc.*, **1**, 25–32 (1993).
20. B.G. Osborne, T. Fearn, A.R. Miller, S. Douglas, 'Application of Near Infrared Reflectance Spectroscopy to the Compositional Analysis of Biscuits and Biscuit Dough', *J. Sci. Food Agric.*, **35**, 99–105 (1984).
21. R.H. Wilson, B.J. Goodfellow, P.S. Belton, B.G. Osborne, G. Oliver, P.L. Russell, 'Comparison of Fourier Transform Mid Infrared Spectroscopy and Near Infrared Reflectance Spectroscopy with Differential Scanning Calorimetry for the Study of the Staling of Bread', *J. Sci. Food Agric.*, **54**, 471–483 (1991).

22. T. Isaksson, B.N. Nilsen, G. Togersen, R.P. Hammond, K.I. Hildrum, 'On-line, Proximate Analysis of Ground Beef Directly at a Meat Grinder Outlet', *Meat Sci.*, **43**, 245–253 (1996).
23. J.P. Wold, T. Isaksson, 'Non-destructive Determination of Fat and Moisture in Whole Atlantic Salmon by Near-infrared Diffuse Reflectance Spectroscopy', *J. Food Sci.*, **62**, 734–736 (1997).
24. S. Kawano, H. Watanabe, M. Iwamoto, 'Determination of Sugar Content in Intact Peaches by Near Infrared Spectroscopy with Fiber-Optics in Interactance Probe', *J. Japan Soc. Hort. Sci.*, **61**, 445–451 (1992).
25. S. Kawano, T. Fijiwara, M. Iwamoto, 'Nondestructive Determination of Sugar Content in Satsuma Mandarin using Near Infrared (NIT) Transmittance', *J. Japan Soc. Hort. Sci.*, **62**, 465–470 (1993).
26. G. Downey, 'Authentication of Food and Food Ingredients by Near Infrared Spectroscopy', *J. Near Infrared Spectrosc.*, **4**, 47–61 (1996).

Nuclear Magnetic Resonance in Analysis of Plant Soil Environments

Teresa W.-M. Fan
University of California, Davis, USA

Andrew N. Lane
National Institute for Medical Research, London, UK

1 Available Nuclear Magnetic Resonance Methodology **1**
1.1 Introduction 1
1.2 One-dimensional Nuclear Magnetic Resonance 2
1.3 Two-dimensional Nuclear Magnetic Resonance 4
1.4 Instrumental and Sample Considerations 5
1.5 Solid-state Nuclear Magnetic Resonance 6
2 Applications in the Plant Soil Environment **6**
2.1 Plant Metabolism 6
2.2 Plant–Soil Interactions 11
3 Future Prospects **20**
3.1 Resolution 20
3.2 Detection Limits 21
3.3 New Applications 21
Acknowledgments **21**
Abbreviations and Acronyms **21**
Related Articles **21**
References **22**

This article presents an overview of nuclear magnetic resonance (NMR) methods and their applications in probing the metabolic processes in plants and interaction between plants and soils. The basic principles of both solution- and solid-state NMR are introduced, with an emphasis on useful parameters that can be employed for determining covalent structures and conformations of biomolecules and their environmental derivatives. Both one-dimensional (1-D) and two-dimensional (2-D) NMR techniques are described, and the advantages of improved resolution and information content for the 2-D techniques are demonstrated. The substantial improvement in the sensitivity of inverse detection methods for detecting heteronuclei is also shown. Instrumental considerations as well as sample preparation requirements are described for in vivo and in vitro applications and analysis of small and macromolecules. Applications of multinuclear 1-D and 2-D NMR in plant metabolism and plant–soil interactions are illustrated. These include in vivo measurements of subcellular compartmentation, determination of metabolite concentrations, enzyme or exchange kinetics, in vitro analysis of plant tissue extracts and root exudates, use of stable isotope tracers for analysis of biochemical pathways, and structural and conformational characterization of natural organic matter (NOM) from soils and sediments. The relative advantages and disadvantages of solution- and solid-state NMR techniques for analysis of plant–soil systems are discussed with examples from characterization of NOM, although 1-D ^{13}C solid-state methods have been of choice due to the polydispersive nature and low aqueous solubility of NOM. However, recent advances in 2-D solution-state methods are presented that provide valuable information on structure, conformation, and ligand binding of NOM, which is not practical by solid-state methods. The limitations of NMR analysis for in vivo, in vitro and NOM are noted, along with a discussion on future prospects for NMR analysis in terms of spectral resolution and detection limits achievable.

1 AVAILABLE NUCLEAR MAGNETIC RESONANCE METHODOLOGY

1.1 Introduction

Here we give a brief overview of the principles and techniques of NMR relevant to plant metabolism and plant–soil interactions. More detailed descriptions can be found in a number of monographs[1–6] and elsewhere in this Encyclopedia.

NMR spectroscopy is based on the principle that certain nuclei can interact with a magnetic field. The nuclear property that gives rise to this interaction is spin, which generates a magnetic moment in the nucleus. Not all nuclei have spin; of the elements of most relevance in the biological arena, the isotopes ^{1}H, ^{13}C, ^{15}N, ^{31}P are NMR active. All of these nuclei have a spin quantum number of 1/2, which means that the nuclear magnetic moment can have only two orientations, namely in the same direction as the applied magnetic field, or opposed to it. These two orientations have a different energy, so that absorption of energy is required to change the orientation from the lower to the higher energy state. The energy must be supplied as an oscillating magnetic field, which is the magnetic component of electromagnetic radiation. As the energy difference is very small, the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

absorption (resonance) frequency is low. For magnetic field strengths commonly available, the proton resonates in the range 100 to 600 MHz. Other nuclei resonate at lower frequencies, e.g. ^{31}P at 40%, ^{15}N at 10% and ^{13}C at 25% of the proton frequency.

1.2 One-dimensional Nuclear Magnetic Resonance

All 1-D spectra consist of a number of peaks or bands that can be characterized by a small number of parameters. These are the chemical shift, spin–spin coupling, peak area, and peak width at half height. Each parameter has a physical significance that can be related to chemical bonding, relative amounts of material, and dynamic properties, respectively.

1.2.1 Chemical Shift

In NMR, the resonant frequency of a peak is usually referenced to a standard, and expressed as a ratio to the external field strength in parts per million (ppm). This ratio is defined as the chemical shift, which is sensitive to the local chemical bonding. Therefore, it is possible to identify protons in different chemical groups on the basis of their chemical shifts. Compound identification can be obtained by matching the chemical shifts of the unknown compound to those of the standard. For example, alanine can be assigned by two ^{1}H resonances at 1.47 ppm (methyl) and 3.77 ppm (Cα), respectively (Figure 1a). Note also that the peak areas are in the ratio 3 : 1 reflecting the relative amounts of the methyl and methine protons. A large data base for ^{1}H chemical shifts of biological compounds is available.[7,8]

Chemical groups can also be identified using nuclei other than ^{1}H. In phosphorylated compounds, the ^{31}P-NMR chemical shifts of phosphate monoesters and diesters are quite distinct, as are the shifts of polyphosphates such as in adenosine 5′-triphosphate (ATP). Similarly ^{15}N chemical shifts are characteristic for amino, amido, and imino groups. ^{13}C-NMR is particularly useful for identifying carbonyls, quaternary carbons, and many protonated aliphatic and aromatic carbons because of its large dispersion of chemical shifts (>200 ppm) and the extensive data base for biological compounds.

1.2.2 Spin–Spin Coupling (Scalar Interactions)

The magnetic nuclei can interact with other nuclei in the same molecule by two distinct mechanisms, each providing different information about the chemical structure. The first is the 'scalar interaction', in which a nucleus senses the spin state of another nucleus through the bonding electrons. This spin–spin coupling is most effective over a limited number of bonds (typically up to four covalent bonds), and is manifested by splitting of the peaks as shown in the alanine spectrum in Figure 1(a). The methyl protons are split into two lines, whereas the methine proton is split into four lines having intensities in the ratio 1 : 3 : 3 : 1. In general, a proton that interacts with n other protons is split into $n + 1$ lines, with intensity ratios given by the binomial coefficients. The magnitude of the splitting is the coupling constant, denoted J. Typical ranges of coupling constants include $^{1}J_{CH} \approx 120–200$ Hz, $^{1}J_{CC} \approx 40$ Hz, $^{1}J_{NH} \approx 70–90$ Hz, $^{3}J_{CH} \approx {}^{3}J_{HH} \approx 1–10$ Hz. J depends on a number of factors including the chemical nature of the coupled spins, the number of bonds, the bond multiplicity, and torsion angles. For example, the three-bond coupling constant in such a structure as H*a*CCH*b* depends on the torsion angle about the central C–C bond. Since the coupling constant varies from a lower limit (H*a*H*b* gauche) to an upper limit (H*a*H*b* trans), its value provides a way of determining the dihedral (or torsion) angle. The relationship between the magnitude of the coupling constant and the dihedral angle is given by the Karplus equation (Equation 1):

$$J = A + B\cos(\phi) + C\cos^2(\phi) \qquad (1)$$

where ϕ is the dihedral angle.

The three constants A, B and C have been determined for a wide range of compounds (allowing for variations from electronegativity of substituents on the carbon atoms).

1.2.3 Relaxation

In addition to the through bond interaction, nuclei can also interact through space. Once the nuclei are excited by absorbing radiation, they can be returned to the ground state by an exchange of magnetization between the sample nuclei and their surroundings. This general process is termed relaxation. There are several mechanisms by which the nuclei can relax, but the most important one is dipole–dipole coupling between spins that are close in space. The strength of this interaction depends on the size of the nuclear magnetic moments, the angle between the field direction and the vector between the two nuclei, how fast the molecule is rotating, and the inverse sixth power of their distance (r) from each other. The magnetic moments are known properties of a given nucleus, while in solution, the random motion of the molecules causes the angular part to average to zero. This leaves only two unknown parameters: the distance between the spins and the rate of the motion. If one of the two is known, the other can be determined through measurement of dipole–dipole coupling.

The process by which the nuclei give up their excess magnetic energy to the surroundings is called spin–lattice relaxation, and is characterized by a time constant T_1. T_1 determines how long a system takes to re-establish

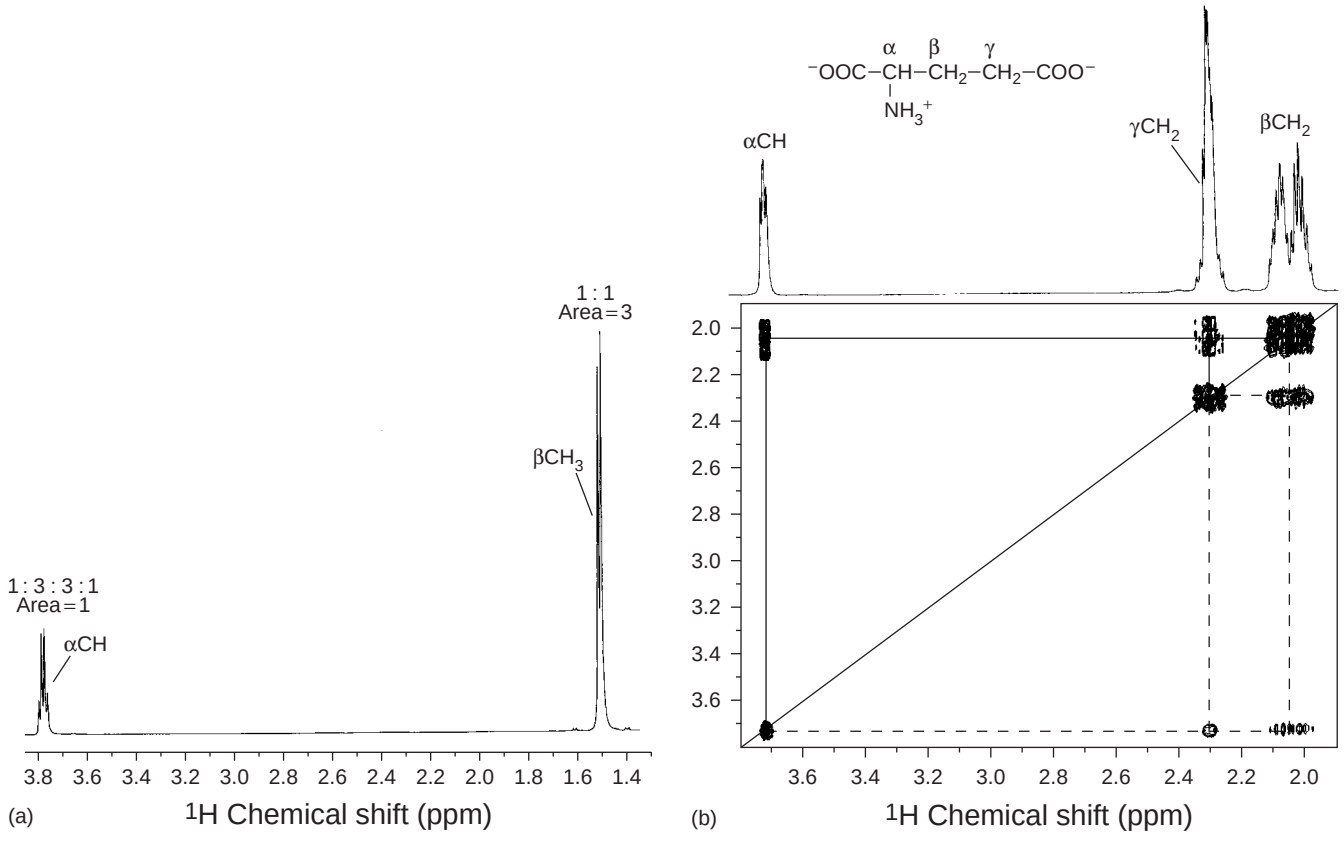

Figure 1 ^{1}H-NMR spectra of L-alanine (a) and L-glutamate (b) in D_2O. All spectra were recorded at 25 °C at 14.1 Tesla. In the 1-D spectrum of alanine (a), the peaks at 1.47 ppm and 3.77 ppm are the βCH_3 and αCH resonances, respectively, while the labels above denote the relative intensities (1 : 3 : 3 : 1 and 1 : 1) and areas of their multiplet components, respectively. The multiplets result from scalar coupling between the αCH and βCH_3 resonances. In plot (b), the 2-D correlation spectroscopy (COSY) (upper triangle) and total correlation spectroscopy (TOCSY) (lower triangle) spectra of glutamate are combined and presented as contour plots. Lines connecting the cross-peaks to the diagonal peaks indicate the presence of scalar coupling. Note that the cross-peak between the αCH and γCH_2 resonances in the TOCSY spectrum is missing in the COSY spectrum. This is because only 2–3-bond couplings are detected in COSY spectra while couplings beyond three bonds are also observed in TOCSY spectra.

the equilibrium condition: typically T_1 values lie in the range of a few milliseconds to a few seconds. The spin–lattice relaxation involves only the component of the magnetization parallel to the polarizing magnetic field (conventionally z). For this reason it is also called longitudinal relaxation. There is a second relaxation process which occurs in the plane perpendicular to the magnetic field (transverse relaxation). Interactions between spins while rotating in the x–y plane cause loss of phase coherence, which involves a change in the entropy instead of the energy of the system. The time constant for this process is T_2 (spin–spin relaxation time). T_2 also determines the breadth of each NMR line with the width at half height (or line width) equal to $1/\pi T_2$.

These relaxation processes are important because they affect the sensitivity of the experiment and the types of NMR experiments that can be performed. The T_1 value is also important where quantitation is desired for the following reasons. The NMR signal is directly proportional to the number of spins in the detection region (i.e. the concentration of a given sample). However, if insufficient time is given between excitation pulses to allow the system to return to equilibrium, then a reduced signal will be obtained depending on the value of T_1. Different nuclei can have different T_1 values so that under conditions of rapid pulsing, the observed signal would not be proportional to the concentration of the spins. This is particularly problematic in ^{13}C-NMR where the ^{13}C nucleus is relaxed by the directly attached proton. As the relaxation rate is proportional to the number of attached protons, the values of T_1 are quite different for CH_3, CH_2, CH and quaternary carbons. Fortunately, T_1 can be directly measured, and the observed intensities corrected in a straightforward manner.[2]

Since dipole-induced T_1 and T_2 processes involve all nuclei that are close in space, it is possible to measure the interaction between pairs of nuclei by cross-relaxation. The dipolar interaction that involves mutual

exchange of magnetization leads to a change in peak intensity, and is limited to spins that are ≤ 5 Å apart. This phenomenon is known as the nuclear Overhauser effect (NOE). The NOE is a measure of the strength of the dipole–dipole interaction between a pair of spins. As this interaction depends on r^{-6}, it provides a means of measuring short distances between pairs of nuclei within a molecule. The combined measures of dihedral angles and interproton distances form the basis of three-dimensional (3-D) structure determination for biomacromolecules.[6] Magnetization can also be transferred by a physical exchange of spins between two sites or molecules, such as in ligand binding (free and bound states) and conformational exchange. These magnetization exchange experiments can be used to measure the kinetics of processes that occur on the timescale of T_1. Examples of this application are the analysis of unidirectional fluxes for ATP synthesis and hydrolysis.[3]

1.2.4 Sensitivity Enhancement

The intrinsic sensitivity of a nucleus is determined by its gyromagnetic ratio (γ), which in turn determines the magnitude of the nuclear magnetic moment and its resonance frequency. Since the proton nucleus (^{1}H) has far greater sensitivity than heteronuclei such as ^{13}C and ^{15}N, the sensitivity of protonated X nuclei can be increased by transferring proton magnetization to the attached X nucleus via the scalar coupling using insensitive nuclear enhancement by polarization transfer (INEPT).[9] This increases the sensitivity by a factor γ_H/γ_X which amounts to 4- and 10-fold for ^{13}C and ^{15}N, respectively.

The X-nucleus peak will also be split by scalar coupling from directly attached protons (^1J) or by long-range (^2J and ^3J) coupling to protons, leading to both reduced sensitivity and resolution. It is therefore common to decouple the protons by broad-band irradiation. This averages the effects of the proton and causes the X resonance to collapse into a singlet, with a concomitant increase (>twofold) in sensitivity and resolution. Irradiation of the protons is also accompanied by NOE for the X nucleus. For ^{13}C, the NOE is always positive, and can enhance the signal up to threefold for small molecules, but for larger molecules, the NOE may be <20%. Since the gyromagnetic ratio of ^{15}N is negative (i.e. it rotates counter clockwise compared with ^{1}H or ^{13}C), the NOE can lead to a fourfold negative enhancement of the signal for small molecules, and a slight reduction in signal strength for large molecules. For intermediate sized molecules, the signal can disappear completely. Therefore, signal enhancement via NOE for ^{15}N needs to be considered carefully.

Since both ^{13}C and ^{15}N are present at low natural abundance (1.1% and 0.37%, respectively), the sensitivity of the experiments can be further improved by isotopic enrichment. The ^{13}C and ^{15}N labels in the enriched molecules can be selectively detected in complex mixtures, using the heteronuclear indirect detection methods described below.[7] Indirect detection of the X nucleus via their attached protons can improve the X nucleus sensitivity to approach that of the proton nucleus. The conventional sensitivity barrier of heteronuclear experiments is thus greatly diminished and their applications to agricultural problems will be much more fruitful.

1.3 Two-dimensional Nuclear Magnetic Resonance

1.3.1 Direct Detection

Although 1-D NMR can contain sufficient information to determine structures and conformational properties of simple molecules, this is not the case for large molecules or mixtures of molecules due to the limitation in spectral resolution and obtainable structure parameters. These problems can be alleviated by two complementary approaches, which can be combined to further improve resolution. The first approach is to increase the magnetic field strength for better signal separation as the latter is directly proportional to the former. The second approach is to increase the number of frequency dimensions. Thus, in 2-D NMR, signals are spread out over a plane composed of two frequencies, instead of along a line as in 1-D NMR. To achieve this, there must be an interaction between pairs of spins. There are two mechanisms by which magnetization can be exchanged between spins, namely the scalar (J-coupled) interaction and the dipolar interaction.[2]

In the basic COSY type of experiments, protons that interact via scalar coupling with other protons are observed as resonating at two frequencies. The spectral representation is usually in the form of a contour plot, where the off-diagonal elements or cross-peaks (symmetric with respect to the diagonal) indicate such scalar interactions (Figure 1b). As scalar coupling is observed between protons separated by up to four covalent bonds, the cross-peaks in a COSY spectrum provide a ''map'' of the covalent connectivity of a molecule. Additional information of this kind can be obtained by TOCSY,[10] also known as HOHAHA (homonuclear Hartmann–Hahn).[11] In this experiment, all spins that form a scalar-coupled system give rise to cross-peaks. Giving glutamate as an example (see Figure 1b for structure), only two sets of cross-peaks are observed in the COSY experiment, i.e. between the two protons of the first two carbons, and between those of the second and third carbons (e.g. Figure 1b). In contrast, in the TOCSY experiment, cross-peaks are also observed between the protons attached to the first

and third carbons (Figure 1b). This can be valuable in crowded spectra where some cross-peaks are obscured by spectral overlap and in tracing covalent bond networks of a given molecule.

Another major 2-D method of correlating spins is by making use of the dipole–dipole interaction or NOE, which is known as nuclear Overhauser effect spectroscopy (NOESY). The NOE is transmitted through space, and its magnitude provides an estimate of the distance between spins, independent of the number of bonds between them. It is therefore extremely valuable for determining the conformation of folded molecules (secondary and tertiary structures) such as polypeptides, polysaccharides and nucleic acids[1,6] or interaction of small molecules with macromolecules such as humics.[12] It also provides a means for observing interactions between subunits in a polymer (quaternary structure), where no scalar coupling interaction can be detected. Moreover, it is very useful for determining the conformations of small, rigid molecules, especially in conjunction with the quantitative analysis of scalar coupling constants. The NOE across the glycosidic linkage can be used for determining the sequence and conformation of oligosaccharides.[13] In small molecules ($M <$ ca. 1 kDa) the $^1H-^1H$ NOE is positive, corresponding to an increase in signal. For macromolecules ($M \gg 1$ kDa), the NOE is negative (decrease in signal). This implies that at some molecular size (or rotational correlation time), the NOE can become zero, irrespective of the proton–proton distance. However, there is an NOE analog of the TOCSY experiment called ROESY (rotating frame Overhauser effect spectroscopy) for which the effect is positive for all molecular sizes.[13,14] This experiment is preferred for oligopeptides and oligosaccharides of $M \approx 1000$ Da or for flexible molecules where the correlation time is short.

The cross-peak patterns and chemical shifts observed in COSY and TOCSY experiments can help define the structures of molecular fragments and even of entire molecules in complex mixtures such as crude extracts of plant tissues without sample fractionation.[7,8,15] A significant limitation is in identifying functional groups containing heteroatoms, and in connecting molecular fragments separated by NMR-inactive heteroatoms such as oxygen. This can be overcome by using heteronuclear 1-D and 2-D NMR techniques. For example, amino and amido groups can be distinguished in 1-D ^{15}N-NMR, as can carbonyl, aromatic, and aliphatic quaternary carbons in ^{13}C-NMR, and phosphate monoesters and diesters in ^{31}P-NMR. It is also feasible to correlate the heteronucleus with protons in 2-D heteronuclear correlation (HETCOR) experiments.[16] However, these direct detection heteronuclear methods suffer from low intrinsic sensitivity, which means a requirement for larger amounts of sample.

1.3.2 Indirect Detection

In addition to having a low intrinsic sensitivity, heteronuclei such as ^{13}C and ^{15}N are present at low natural abundance (1.1% for ^{13}C and 0.37% for ^{15}N), which further decreases the overall sensitivity. Hence, it is highly desirable to obtain ^{13}C and ^{15}N spectra with the sensitivity of the proton. This can be achieved by the "indirect detection" methods. In the simplest 2-D HSQC (heteronuclear single quantum coherence) experiment,[17] the magnetization of the proton is transferred to the hetero (X) nucleus by INEPT, and the X nucleus is then allowed to precess for different amounts of time to label with X frequency. The magnetization is then transferred back to the attached proton for detection, which increases the sensitivity by $(\gamma_H/\gamma_X)^3$. This represents a signal enhancement of 64-fold for ^{13}C and 1000-fold for ^{15}N.[18] HSQC is mainly useful for detecting X nuclei with directly bonded proton(s), which means that carbonyl, tertiary nitrogens, and quaternary carbons and nitrogens are not detected. However, it is possible to correlate protons and an X nucleus over two or more bonds. For example, a long-range HSQC experiment can be performed for ^{31}P or ^{15}N since, for most compounds, the X nucleus is isolated (i.e. no N–N or P–P coupling) so that there is only one or two possible coupling pathways to the protons. Other methods of correlating protons to a heteronucleus have also been reported, which in effect produces an X-edited TOCSY (HeteroTOCSY).[19,20] For carbon long-range coupling, the HMBC (heteronuclear multiple bond correlation) experiment[21] may be applied, which correlates carbons with protons 2–4 bonds away, and is therefore valuable for detecting carbonyl and quaternary carbons.

The X-edited 2-D experiments provide a powerful means for identifying isotopically enriched compounds and the position(s) of their isotopic labels. When spectral resolution permits, the degree of enrichment in proton experiments can be determined from the labeled satellites and the non-labeled central peak. The ratio of the satellite peak area to the total peak area is equal to the fraction of the enrichment. These capabilities will be invaluable in tracer studies where the fate and transformation of both natural products and anthropogenic pollutants in the plant–soil environment are probed.

1.4 Instrumental and Sample Considerations

For small molecules ($M_r \leq 500$), all of the aforementioned experiments can be successfully performed with relatively low concentrations of solutes (in the sub millimole range) on a modern spectrometer operating at 11.75 to 14.1 T (500 and 600 MHz for ^1H, respectively). With ^1H detection, a 1-D NMR spectrum can be obtained within a few minutes at 10 μM in a volume of 0.5–0.6 mL (0.3 mL with special Shigemi tubes), i.e. at ≤5 nmol.

Mixtures of metabolites can be resolved and identified by TOCSY in an overnight run at ca. 100 μM concentration, which is typically obtainable from <10 mg of dry plant tissues.[8] The HSQC can be performed in a few hours at natural abundance at similar solute concentrations for ^{13}C, and with isotopic enrichment a reasonable spectrum can be recorded at much lower concentrations in a similar time.

For large molecules such as proteins, nucleic aids, and humic substances, the sensitivity is generally poorer because the resonances become broader due to a faster spin–spin relaxation (i.e. short T_2 values) such that there is insufficient time to sample the magnetization before it decays. The resonances for humic substances may be additionally broadened due to the association with paramagnetic ions. Therefore, only a subset of the above experiments work for macromolecules, and generally require larger quantities and longer acquisition times.

Another consideration is suppression of the solvent resonance. Tissue extracts can generally be lyophilized and dissolved in deuterated solvents to minimize the contribution of solvent signals in ^{1}H-NMR measurements. However, some measurements need to be performed in non-deuterated solvents (e.g. water) as in the case of in vivo NMR or detection of solvent-exchangeable protons (e.g. amido, amino, or hydroxyl protons). As H_2O is 110 M in protons, while the solutes may be only ca. 100 μM, this creates a large dynamic range problem (ca. 10^6 fold). Although it is possible to suppress the water resonance by irradiation with coherent radiofrequency at the H_2O frequency, in practice this is usually inadequate for dilute solutions, and may introduce spectral artifacts. This problem is greatly alleviated with the more recent solvent suppression methods that utilize pulsed magnetic field gradients along the z-axis to selectively dephase the water signal and provide excellent suppression even in dilute solutions. The most common method of this type is Watergate[22] which is routinely used in biomacromolecular NMR, and can be incorporated into most of the common 2-D and 3-D experiments.

1.5 Solid-state Nuclear Magnetic Resonance

Some samples of interest are either insoluble or insufficiently soluble in common solvents, so they would be better suited for solid-state NMR analysis. The main difference between the liquid- and solid-state methods is that in the solid state, the highly restricted motion leads to a dependence of dipolar interactions on the angle between the nuclear magnetic moments and the magnetic field direction, whereas in the liquid state the Brownian motion of the solutes causes the dipolar interactions to average such that there is no angular dependence. In solids, dipolar interactions can be very strong, amounting to local magnetic fields that cause variations in chemical shift of tens of kHz. Thus, instead of a single sharp line observed in the liquid state, the solid-state resonance is a broad envelope that reflects the many orientations of each nucleus with respect to the magnetic field. Hence, an alternative means of averaging the dipolar interactions and chemical shifts must be used to improve the spectral resolution. Spinning the sample at high speed at the "magic angle" ($\cos^{-1}(1/3)^{0.5} = 54.74°$) causes the dipolar and chemical shift anisotropy interactions to average. The rotation speed has to be higher than the characteristic frequency range of the observed nucleus, which for ^{13}C is a few kHz. Magic angle spinning (MAS) is performed routinely in solid-state ^{13}C-NMR.

For the low natural abundance nuclei such as ^{13}C, the solid-state measurements suffer the same sensitivity problem as the liquid-state analysis. To improve sensitivity, the larger proton magnetization is transferred to the X nuclei through dipolar coupling.[5] This is in practice achieved by Hartmann–Hahn matching, in which strong radiofrequency fields are applied simultaneously to the protons and X nucleus, with intensities proportional to their gyromagnetic ratios. Such magnetization transfer from the abundant protons to the dilute X nuclei is known as cross-polarization (CP) transfer. The efficiency of the transfer is a function of the chemical composition and relaxation properties. In addition, the large reservoir of fast relaxing protons allows more rapid acquisition of the X signal, which means that a higher signal intensity can be recorded in a shorter duration. The combination of CP transfer and MAS (CP/MAS) with ^{13}C-NMR has been widely applied to soils, where chemical shifts have been used to reveal the chemical groups of the organic matter present. However, reliance on the chemical shift information alone can lead to ambiguities in assignments (see below for a more detailed discussion of assignment strategies).

2 APPLICATIONS IN THE PLANT SOIL ENVIRONMENT

2.1 Plant Metabolism

2.1.1 In Vivo Nuclear Magnetic Resonance

Ideally, one would like to measure metabolic processes in situ so that spatially and temporally resolved physiological/biochemical reactions can be determined. This capability is vital to the understanding of all life processes since maintenance of distinct cellular and organ compartments is required for the proper functioning of the metabolic machinery. In vivo NMR represents one of the few analytical tools that possess this capability.

2.1.1.1 Sample Preparation The types of agriculture-related samples that have been studied in vivo by NMR include whole seedlings, excised plant tissues (root tips, mature roots, shoot tips, leaf blades), as well as plant, fungal, and algal cell cultures. Among these, root tips are the most commonly investigated, due to their high metabolic activity and strong NMR signals obtainable. To implement in vivo NMR studies of plant/fungal/algal metabolism, the sample needs to be maintained under physiologically relevant conditions. This entails supply of oxygen and nutrients, removal of waste products, and illumination in the case of photosynthetic tissues. The former two requirements can be met with superfusion systems for whole plants[23] and excised plant tissues[24,25] or airlift systems for plant cell suspensions.[26,27] However, implementing uniform and optimal illumination has been problematic for investigating photosynthetic processes in leaf tissues. For algal suspensions, utilizing fiber-optic cables for light delivery appeared promising[28] but further development is needed for working with suspensions of high cell densities. The latter is important for obtaining NMR spectra of good signal-to-noise ratios. It should also be emphasized that care must be taken to avoid complications from tissue wounding and aging processes in performing in vivo NMR experiments with excised plant tissues. That is, experiments should be conducted after initial stabilization of the excised tissues and before significant tissue deterioration occurs.

2.1.1.2 Information Obtainable In vivo NMR has been utilized to measure both steady-state and dynamic processes in plants, fungi, and algae.[29] The following types of information are available:

- tissue content and changes of NMR-observable inorganic and organic metabolites (see references below);
- intracellular pH (see references below);
- subcellular distribution of inorganic and organic metabolites, most commonly in the cytoplasm and vacuole;[25,30–32]
- enzyme or exchange kinetics.[33]

In vivo observable tissue metabolites in plants, fungi, and algae include a number of organic acids, amino acids, and carbohydrates by ^{1}H-NMR,[30] borate and borate esters by ^{11}B-NMR,[34] storage carbohydrates and triglycerides by natural abundance ^{13}C-NMR,[35] nitrate, ammonium, and some quaternary ammonium compounds by ^{14}N-NMR,[36] sodium ion by ^{23}Na-NMR,[37,38] aluminum–polyphosphate complexes by ^{27}Al-NMR,[39] inorganic phosphate, various phosphate ester (including high-energy nucleoside triphosphates), and polyphosphates by ^{31}P-NMR,[35,40,41] and potassium ion by ^{39}K-NMR.[42] Using stable isotope tracers, ^{13}C- and ^{15}N-labeled metabolites including amino and organic acids, mobile and storage carbohydrates, polyols, and secondary metabolites in plants and associated microbes have been detected in vivo by ^{13}C- and ^{15}N-NMR.[32,35,43,44]

For measuring intracellular pH, ^{31}P-NMR is by far the most frequently used tool because several phosphate metabolites, most notably inorganic phosphate, have pK_a values within 1–1.5 units of the plant cell pH. For example, the chemical shift of inorganic phosphate (p$K_a = 6.8$) is sensitive to pH in the range of 5.5 to 8 and consequently two distinct signals (corresponding to cytoplasmic and vacuolar inorganic phosphates) are typically observed in ^{31}P-NMR spectra of plant tissues in vivo (see Figure 2). From the chemical shifts and appropriate titration curves of inorganic phosphate,[45] the pH of the plant cell compartments can be determined to within ± 0.1 units near pH 7 (poorer at the extreme ends of the titration curve). Relative changes in intracellular pH such as those induced by salinity (Figure 2) can be measured in real-time to within ± 0.05 pH units accuracy. Other metabolites including histidine,[25] ^{13}C-bicarbonate,[46] and fluorinated amino acids[47] are also useful as intracellular pH probes, via ^{1}H-, ^{13}C-, and ^{19}F-NMR analysis, respectively. Also shown in Figure 2 are signals arising from other important phosphate metabolites including nucleoside 5′-triphosphates (NTPs), nicotinamide adenine dinucleotide phosphate (NADP) and its reduced form, uridine 5′-diphosphate glucose (UDPGlc), and sugar phosphates. Phosphodiester signals may be observed in some systems such as sycamore cell suspensions (Figure 3).

The distinction between cytoplasmic and vacuolar inorganic phosphates illustrates the value of in vivo NMR for measuring the subcellular distribution of metabolites. In principle, metabolite compartmentation can be followed by in vivo NMR, provided that the NMR signals of a given metabolite are sufficiently sensitive to the compartment conditions (e.g. pH, ionic strength, paramagnetic ions) such that their location can be differentiated by chemical shifts and/or other NMR parameters (e.g. relaxation behavior). Figure 4 demonstrates an in vivo ^{13}C-NMR analysis of malate and citrate compartmentation in the cytoplasm and vacuole of sycamore cells.[48] The respective localization of histidine and malate in the vacuole and cytoplasm of maize roots was demonstrated via in vivo ^{1}H-NMR analysis.[30] The intra- and extra-cellular localization of sodium ions in plant tissues was achieved with the use of a shift reagent and ^{23}Na-NMR.[37]

In terms of dynamic processes, analysis of unidirectional rates of ATP turnover by the ^{31}P-NMR saturation transfer technique represents a unique and powerful

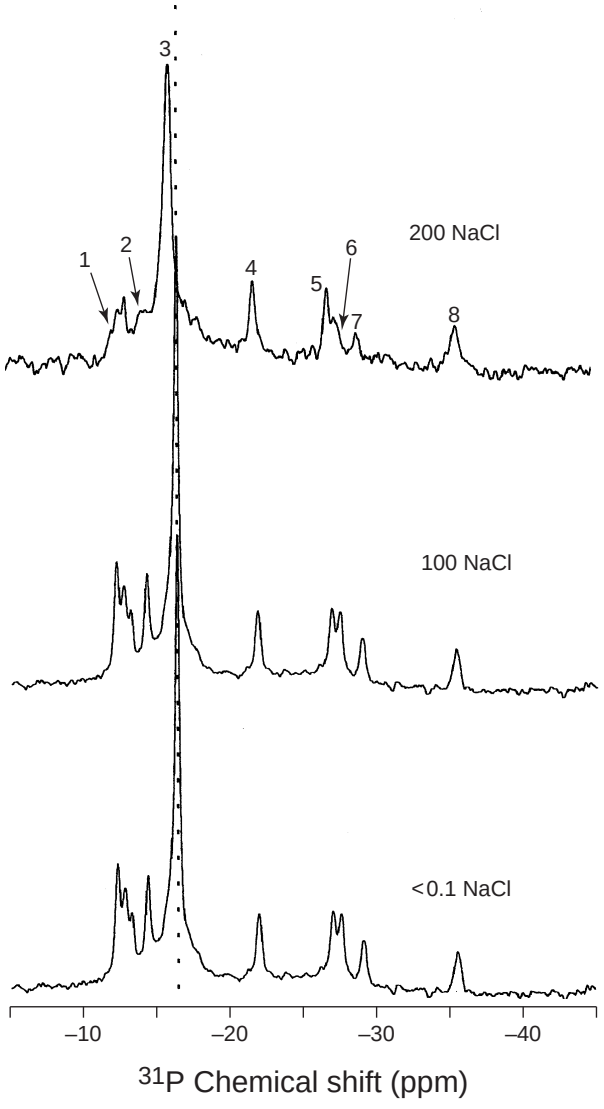

Figure 2 In vivo ^{31}P-NMR spectral changes of *Sorghum bicolor* root tips as a function of NaCl treatments. The dashed line denotes the change in the peak position of the cytoplasmic inorganic phosphate (Pi) signal from *S. bicolor* root tips as a result of treatment with increasing concentrations of NaCl (in mM). Peaks are labeled as 1, glucose-6-phosphate; 2, cytoplasmic Pi; 3, vacuolar Pi; 4 and 5, terminal and α-phosphate of nucleoside tri- and diphosphates (γ nucleoside 5′-triphosphate (NTP) + βNDP, αNTP + αNDP), respectively; 6, NADPH + UDPH; 7, UDPG; and 8, β-phosphate of nucleoside triphosphate (βNTP). (Reprinted from Colmer et al., © 1994,[112] with permission of Oxford University Press.)

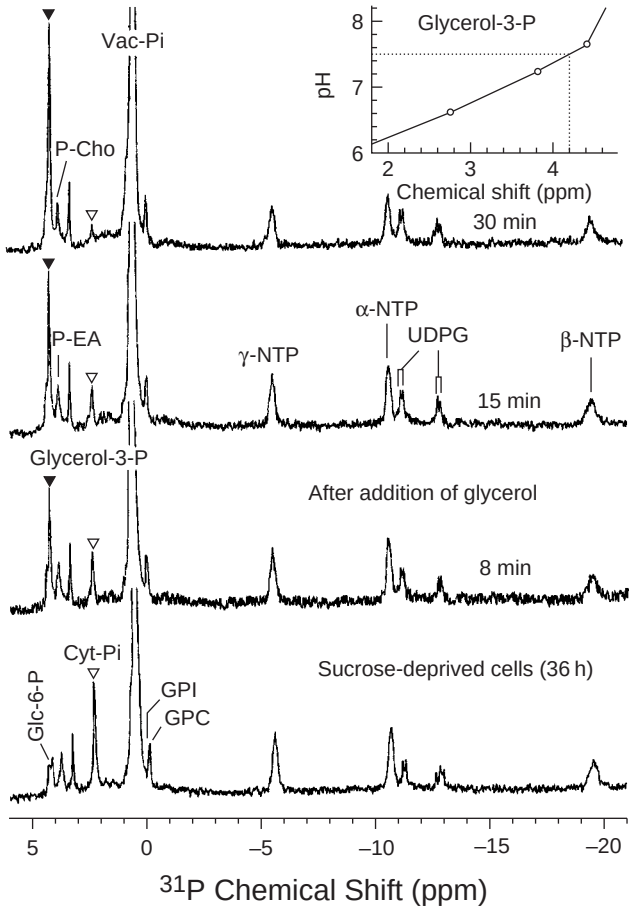

Figure 3 Time course changes in in vivo ^{31}P-NMR spectra of sucrose-deprived sycamore cells after the addition of glycerol. Inset: chemical shift of *sn*-glycerol 3-phosphate versus pH. UDPG, uridine 5′-diphosphate-α-D glucose, P-EA, phosphoethanolamine. (Reprinted from Aubert et al., © 1996,[32] with permission of The American Society of Plant Physiologists.)

application for in vivo NMR.[3] This technique measures the exchange process (i.e. magnetization exchange) between ATP and inorganic phosphate as catalyzed by the mitochondrial ATPase. By following the transfer of magnetization from the γ resonance of ATP to the inorganic phosphate resonance or vice versa, the unidirectional rates of ATP synthesis or hydrolysis can be deduced by taking into account the spin–lattice relaxation and other competing reactions (see Figure 5 for an example). Another common application of in vivo NMR is to follow the changes in metabolic processes with time. This can be performed with or without stable isotope tracers. For example, the time courses of changes in energy metabolism and intracellular pH in plant tissues under hypoxic stress were followed in vivo by ^{31}P-NMR,[49] while complementary changes in fermentation metabolism were revealed by ^{1}H-NMR.[25,30] In vivo NMR analyses that unveil novel metabolic changes under other plant stress conditions (e.g. pH, salinity, sucrose starvation) have also been reported, largely employing ^{31}P-NMR (see reviews by Ratcliffe).[29,40] The ^{1}H- and ^{31}P-NMR applications were performed without labeled tracers since the natural abundance of these two nuclei is virtually 100%.

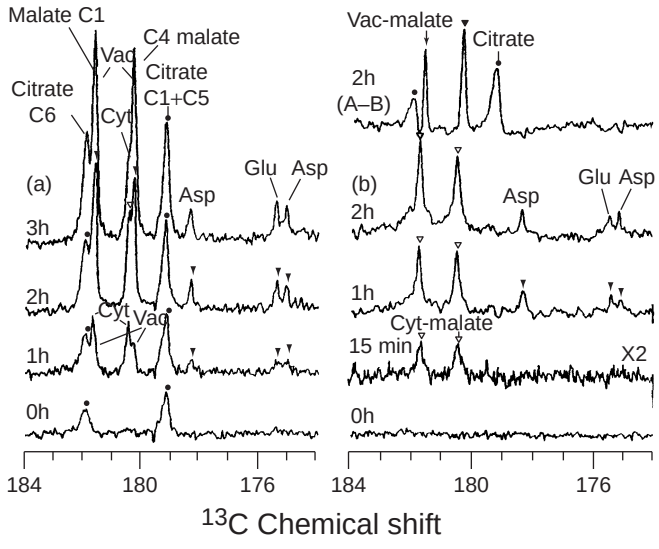

Figure 4 In vivo ^{31}P-NMR measurement of malate and citrate compartmentation in sycamore cells treated with 5 mM $H^{13}CO_3^-$. The two series of spectra were acquired from perfused sycamore cells in the absence (a) and presence (b) of 0.5 mM Mn^{2+}. In (a), cytoplasmic and vacuolar malate signals were observable but severely overlapped, while those for citrate were not discerned. With the addition of Mn^{2+} as in (b), the vacuolar signals for malate and citrate were suppressed, thus allowing their cytoplasmic signals to be observed. The difference spectrum of (a) and (b) (a – b) revealed the vacuolar signals of malate and citrate. (Reprinted from Gout et al., © 1993,[48] with permission of The American Society for Biochemistry & Molecular Biology.)

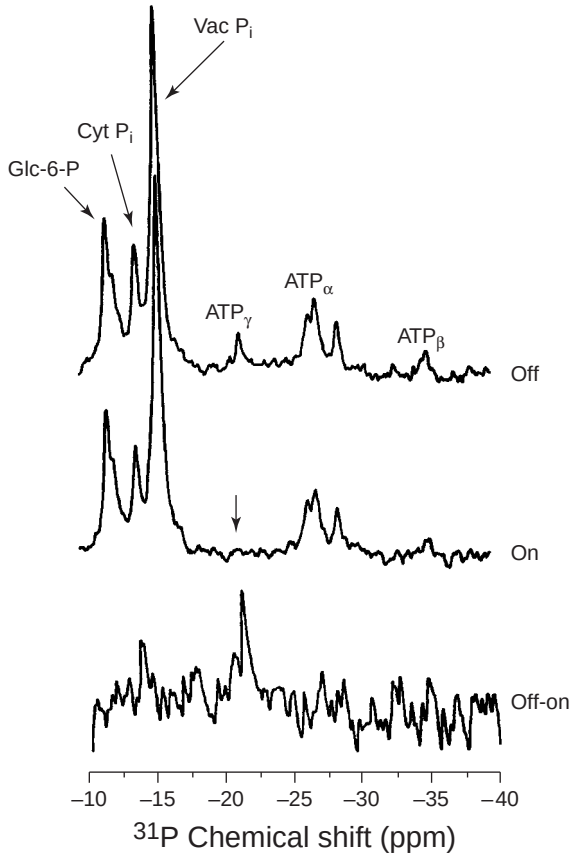

Figure 5 In vivo ^{31}P-NMR saturation transfer measurement of the rate of ATP synthesis in hypoxic rice coleoptiles. The γ-phosphate resonance of ATP (ATP_γ) was saturated in the ''on'' spectrum while no irradiation was applied in the ''off'' spectrum. The difference spectrum (off – on) shows the effect of saturation transfer from the ATP_γ to the cytoplasmic phosphate resonances. (Reprinted from Fan et al., © 1993,[33] with permission of Academic Press, Inc.)

In contrast, ^{13}C- or ^{15}N-enriched tracers are invariably used to measure metabolic changes in vivo via ^{13}C- or ^{15}N-NMR, due to the low natural abundance of these two nuclei. Such labeled tracer studies can also provide metabolic flux and pathway information that is difficult to obtain otherwise. Moreover, isotopic enrichment improves detectability such that in vivo 2-D NMR measurements can be performed for a more reliable assignment of resonances observed in vivo.[44] Examples of in vivo ^{13}C- and ^{15}N-NMR applications are shown in Figures 6 and 7, respectively.[50,51] In Figure 6, the synthesis of trehalose and glycogen from external glucose was demonstrated in mycorihizae-colonized leek roots, and the mycorrhizal symbiosis with roots was shown to be essential to the synthesis. Figure 7 illustrates the synthesis of γ-aminobutyrate (GAB) from external NH_4^+ in carrot cells via the amino acid Glu and possibly Ala. This, together with the intracellular pH data obtained from ^{31}P-NMR, revealed a significant role of GAB in pH homeostasis in plant cells.[51]

2.1.1.3 Limitations Sensitivity is always a limitation in NMR. This limitation accounts in part for the popularity of high-natural-abundance ^{31}P-NMR applications and the need for isotopic enrichment for ^{13}C- and ^{15}N-NMR applications. In addition, tissue samples, unlike homogeneous solution, disrupt magnetic homogeneity, which degrades sensitivity due to increased line width. This problem particularly affects ^{1}H-NMR applications since the ^{1}H spectral range is narrow. Moreover, the cytoplasm (where metabolites of interest often reside) of plant cells often occupies only a small fraction of the cell volume, which further adds to the difficulty of detection. On the other hand, a much larger cytoplasmic volume is present in the fast growing regions (root and shoot tips), which explains their popularity for in vivo NMR applications.

There are also constraints on sample size and types depending on the magnet configuration. In general, maximizing sample density is highly desired, which complicates

Figure 6 In vivo ^{13}C-NMR time-course spectra of mycorrhizae-colonized (a) and uncolonized (b) leek roots superfused with 0.6 mM ^{13}C1-glucose. Peak assignment is as follows: a, sucrose C2′, fructose C2; b, glycogen C1; c, β-glucose C1; d, trehalose C1, 1′; e, α-glucose C1, sucrose C1. (Reprinted from Shachar-Hill et al., © 1995,[50] with permission of The American Society of Plant Physiologists.)

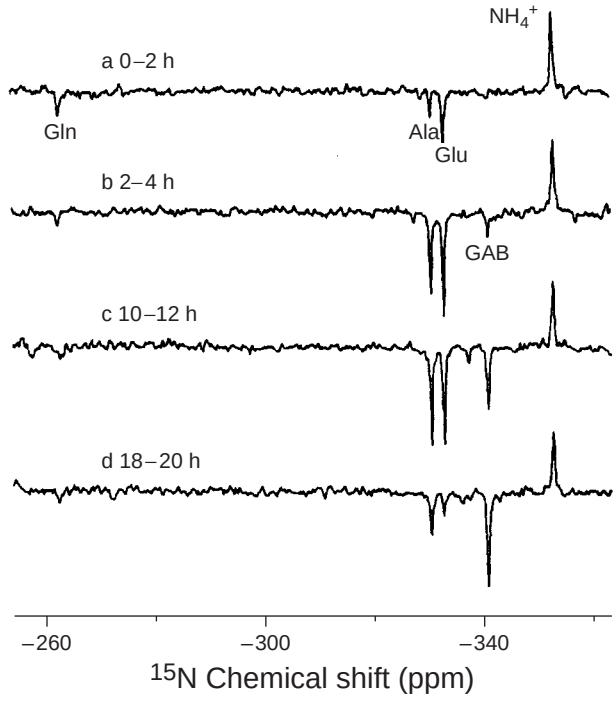

Figure 7 In vivo ^{13}N-NMR evidence for the synthesis of γ-aminobutyrate (GAB) from ^{15}N-ammonium chloride in aerated carrot cells. The ^{1}H-decoupled ^{13}N-NMR spectra were obtained in 2 h. (Reprinted from Carroll et al., © 1994,[51] with permission of The American Society of Plant Physiologists.)

the maintenance of tissues in a physiologically stable condition, particularly for whole plants. Another difficulty with whole plant work is localizing signals from different anatomical regions, e.g. root tips versus more mature root regions. For leaf tissues, the air spaces often cause such strong line broadening that vacuum infiltration is needed to obtain usable spectra.

Once usable spectra have been acquired, there are the problems of signal assignment, verification, and quantification, particularly for those metabolites whose signals are sensitive to cellular conditions such as pH, ionic strength, divalent and paramagnetic ions, and viscosity. The assignment problem can be straightforward for ^{31}P-NMR since the number of metabolites involved is small and their signals are generally sufficiently resolved. The assignment difficulty is perhaps most severe for ^{1}H-NMR for the following reasons. Due to the narrow spectral spread and the higher number of observable metabolites, in vivo ^{1}H-NMR spectra are generally more crowded in appearance and assignment based on chemical shifts alone is grossly inadequate. In addition, the water signal is orders of magnitude stronger than metabolite signals, which makes it difficult to suppress, particularly when the signal is broad. However, the new pulsed-field gradient method for solvent suppression may greatly alleviate the latter problem.[22] To overcome the assignment problem, a comprehensive analysis of NMR-observable metabolites in tissue extracts needs to be obtained to assist the in vivo NMR assignment.[7,8,15] Despite these shortcomings, ^{1}H-NMR remains attractive for its superior sensitivity and high information content.

It should be emphasized that assignment of metabolite signals that arise from different compartments should be independently validated by other analysis. Caution should also be taken in utilizing in vivo NMR signal intensities for absolute quantification, particularly in crowded spectra and when intensities are sensitive to metabolite mobility via binding to membranes or large proteins and to exchange with ligands such as paramagnetic and divalent cations. The latter processes could attenuate or even eliminate metabolite signals.

Signal attenuation by immobilization is substantial for sodium ions while that by exchange with cations is most relevant to phosphate metabolites and organic acids.

2.2 Plant–Soil Interactions

2.2.1 Plant Extract Characterization

In addition to in vivo applications, NMR is eminently suited for profiling metabolites in a wide variety of biological extracts including those of plants, algae, and soil microbes. The NMR analysis has the following unique advantages:

- molecular structure information can be obtained directly in crude extracts;
- extensive sample preparation and associated artifacts are eliminated;
- a comprehensive profile of metabolites can be acquired simultaneously with unequivocal structure confirmation;
- unexpected and/or unknown metabolites can be detected and identified without the need for authentic standards;
- it is vital to a reliable assignment of in vivo NMR resonances.[7,8]

These advantages of in vitro NMR analysis of tissue extracts have been demonstrated in a number of plant metabolic studies.[15,25,30,32,50,52–57] For example, not only were a number of amino and organic acids determined in the extract of a euryhaline alga isolated from an agricultural drainage evaporation pond, but two unexpected metabolites (trehalose and dimethylsulfonium propionate) were also revealed by 2-D ^{1}H-TOCSY analysis (Figure 8).[55] The latter two compounds are likely to be involved in the osmotic adjustment in this alga. In another study, the identity of dimethylselenonium propionate and other organoselenium metabolites was unequivocally confirmed in synthetic mixtures without purification by a combination of 2-D ^{1}H-TOCSY, ^{77}Se–^{1}H-heteroCOSY, and ^{1}H-DQF/COSY (double quantum filtered correlation spectroscopy) analysis. This information was then utilized to determine the identity of selenium metabolites in algal extracts.[56,57] Moreover, 2-D ^{1}H-TOCSY, ^{15}N–^{1}H-HSQC, and ^{15}N–^{1}H-HSQC/TOCSY (Figure 9) experiments were used to determine ^{15}N-labeled metabolites and to estimate the % ^{15}N enrichment in these metabolites in rice coleoptiles in response to a combination of hypoxia, ^{15}N-nitrate, and ^{15}N-ammonium treatments.[54] Together with the position of the ^{15}N label in each metabolite, different assimilation pathways of nitrate and ammonium ion were uncovered.

2.2.2 Plant Exudate Characterization

The utility of in vitro NMR for tissue metabolite profiling is equally valuable for analyzing exudates excreted by plants and/or microbes. This emerging application holds great promise in advancing our knowledge in plant–soil interactions. This is because exudation accounts for a large organic input to the soil and is intimately involved in a number of important processes such as trace nutrient acquisition, plant–plant interaction, microbe–plant interaction, microbial community, and humification. For example, exudation of phytosiderophores and bacteriosiderophores has been shown to play a vital role in Fe acquisition by plants and soil microbes.[58] However, a mechanistic understanding on the role of plant and microbe exudates in these processes has been hampered by a general lack of information on the exudate composition. This gap of knowledge has been a difficult challenge to conventional analytical methods but application of NMR to this problem will greatly help to circumvent the difficulty.

This is illustrated in a recent study on a comprehensive analysis of plant root exudates using the multinuclear and 2-D NMR approach.[59] Three derivatives of mugineic acid (phytosiderophores) were identified and quantified directly in crude barley exudates without fractionation, along with a number of amino and organic acids. The mugineic acids were found to dominate the total organic ligand equivalents in barley exudates only under severe Fe deficiency, while other ligands contributed comparably under less Fe-deficient conditions. This finding may have important implications in agricultural soils where mild Fe deficiency may be most relevant.

2.2.3 Natural Organic Matter Characterization

NOMs represent one of the Earth's most abundant carbon reserves.[60] They are composed of compounds with a wide chemical diversity, particularly the humic and fulvic acids (Hayes et al.,[61] and references therein). Humic/fulvic acids or humates/fulvates (the more likely ionization state in natural soils) are quasi biomacromolecules since they are derived from decomposed plant/algae/animal debris and microbes through complex and yet largely undefined processes[62,63] which may also involve the mineral matrix.[64,65] Consequently, humic/fulvic structures are notoriously heterologous and largely non-polymeric[66] unlike biopolymers such as proteins and nucleic acids. The heterogeneous nature of humates and fulvates has presented a unique challenge to their structural characterization and thus to a mechanistic understanding of their function(s) in the environment. Soil and sediment humates and fulvates are active participants in a number of important processes involving plant

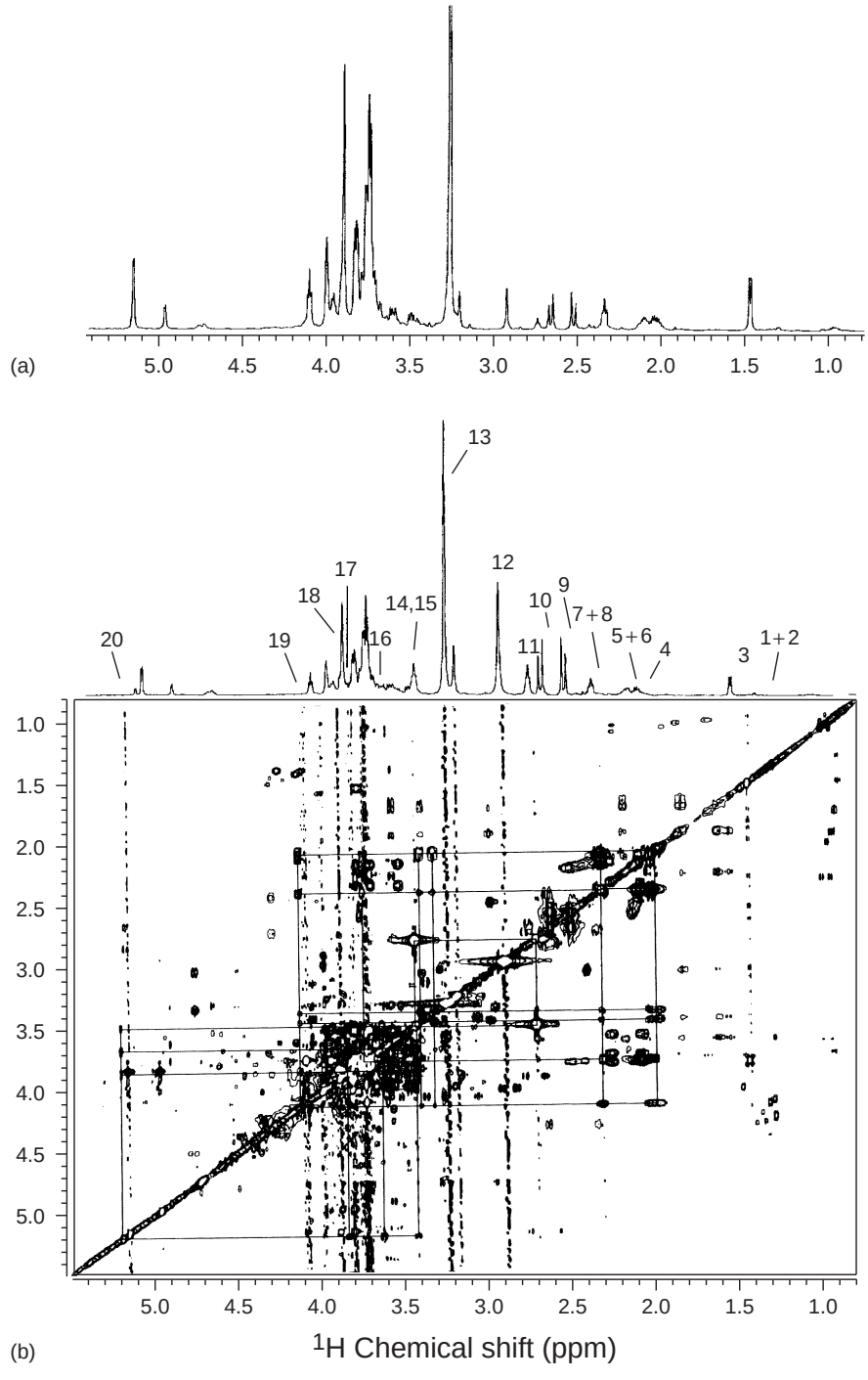

Figure 8 1-D High-resolution and 2-D ^{1}H-TOCSY/NMR spectra of *Chlorella* algae extracts from control and Se treatments. The 1-D ^{1}H-NMR spectra were obtained from the 100 ppm Se (a) and control (b) treatments while the 2-D ^{1}H-TOCSY spectrum (b, displayed as a contour plot below the corresponding 1-D spectrum) was acquired from the control treatment. Rectangular boxes connect the diagonal to the off-diagonal cross-peaks, which illustrate the ^{1}H scalar coupling network of dimethylsulfonium propionate (DMSP), trehalose, and proline. Peak assignment is as follows: 1, βCH_3 of lactate; 2, γCH_3 of Thr; 3, βCH_3 of Ala; 4, 5, 7, 19, γCH_2, $\beta'CH$, βCH, and αCH of Pro; 6, 8, β and γCH_2 of Glu, 9, 10, α, γ and α', γ' CH of citrate; 11, 12, 14, αCH_2, S-CH_3, and βCH_2 of DMSP; 13, 18, N-CH_3 and αCH_2 of glycinebetaine; 15, 16, 17, 20, C4H, C2H, C6H, C3, 6H, and C1H of trehalose, respectively. (Reprinted from Fan et al.,[55] with permission of American Chemical Society © 1997.)

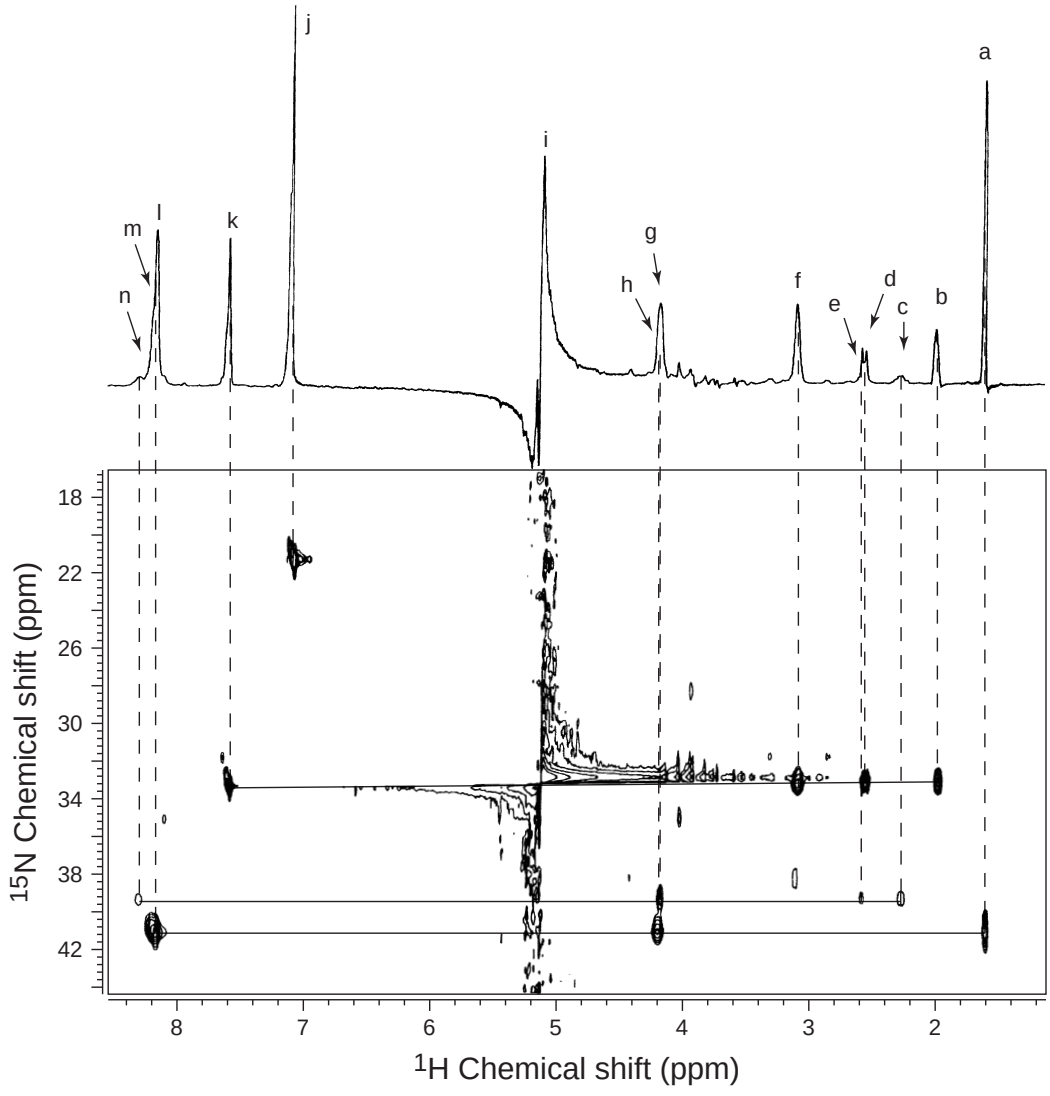

Figure 9 2-D ^{15}N–^{1}H-HSQC/TOCSY spectrum of a rice coleoptile extract after ^{15}N-ammonium treatment. The 2-D HSQC/TOCSY spectrum is displayed as contour plots along with the 1-D high-resolution spectrum. Peak assignment is as follows: a, h, and l, βCH_3, αCH, and αNH_3^+ of Ala; b, d, f, and k, βCH_2, αCH_2, γCH_2, and γNH_3^+ of GAB; c, e, g, and n, βCH_2, γCH_2, αCH, and αNH_3^+ of Gln; m, αNH_3^+ of Glu; j, NH_4^+; i, residual H_2O, respectively. (Reprinted from Fan et al., © 1997,[54] with permission of Oxford University Press.)

mineral nutrition, microbial activities and community, pollutant transport and bioavailability.[67,68]

NMR has been applied to NOM characterization for over three decades.[69] The lack of sensitivity and resolution of earlier NMR instrumentation prevented its routine use in NOM research. However, with the dramatic improvements in NMR technology in the last decade, NMR has emerged as the method of choice for obtaining structure information on soil humates and fulvates. A number of solution- and solid-state NMR methods involving ^{1}H, ^{13}C, ^{31}P, and ^{15}N nuclei have been applied to humate and fulvate characterization as a function of soil sources, soil decomposition/development, humification,[70] cultivation, burning,[71,72] as well as interactions with xenobiotics[73] and metals.[60,65,69,74–76,78] In addition, a few studies have utilized other NMR-active nuclei including ^{27}Al,[79] ^{29}Si,[80–82] ^{31}P,[83] ^{51}V,[84] and ^{113}Cd.[85–87]

2.2.3.1 ^{1}H-Nuclear Magnetic Resonance Despite the high sensitivity of the ^{1}H nucleus, few ^{1}H-NMR methods have been applied to humic and fulvic characterization compared with ^{13}C-NMR applications. This is largely because resonances in 1-D ^{1}H-NMR spectra of humates are severely overlapped, which limits the structural information obtainable. This spectral overlap results from a

combination of the large number of chemical groups, chemical shift heterogeneity (due to variability in chemical environments of similar groups), the macromolecular size, and paramagnetic broadening. The presence of multivalent paramagnetic ions also limits the solubility of humates in D_2O, which necessitates the use of reactive solvents such as d_6-dimethylsulfoxide (d_6-DMSO) and NaOD.[88]

To date, most of the ^{1}H-NMR studies have been performed in the liquid state.[89–93] This is because in the solid state the strong dipolar coupling between protons leads to very broad lines (ca. 50 kHz) that are difficult to narrow by MAS or other standard techniques.[5] This yields poor quality ^{1}H-NMR spectra.[88] However, ^{1}H–^{1}H dipolar interactions were exploited in 2-D solid-state ^{1}H-NMR spectroscopy, where dipolar dephasing was varied in the second dimension to obtain dipolar interactions as a function of chemical shift (i.e. different chemical groups).[88]

The information obtained from ^{1}H-NMR studies of humates includes types of substructures and abundance of a given substructure. A 1-D ^{1}H-NMR spectrum of a sediment humate is shown in Figure 10,[90] where the spectral resolution is superior to that commonly obtained for soil humate. Substructures were assigned

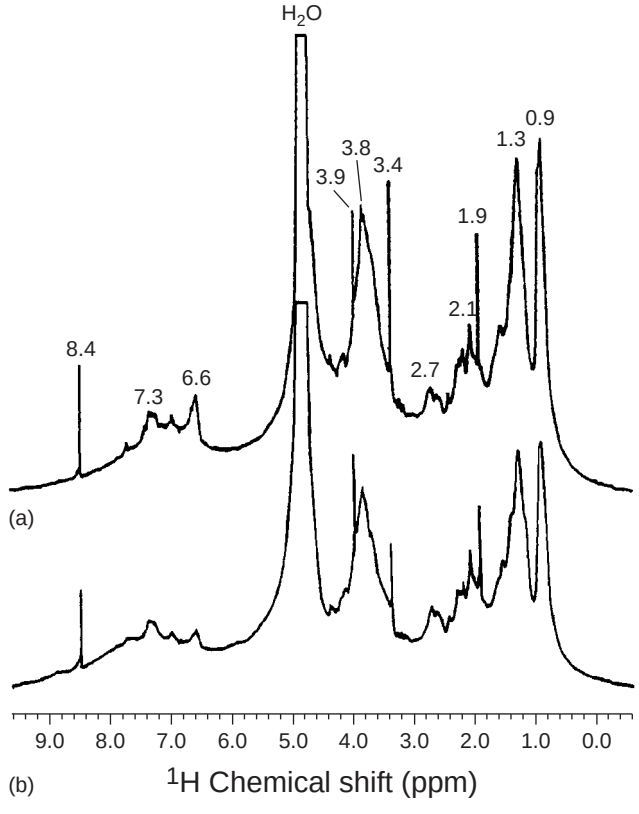

Figure 10 1-D Solution-state ^{1}H-NMR spectra of (a) Lake Zlatari and (b) Vasići humates. (Reprinted from Jokić et al.,[90] with kind permission of Kluwer Academic Publishers © 1995.)

Table 1 ^{1}H Chemical shift ranges assigned to various structural groups in aquatic and terrestrial humates and fulvates

Assigned groups	^{1}H Chemical shift range (ppm)
Aliphatic (R–CH_3 and –CH_2–R)[94]	0.0–1.6
Alkyl protons[82]	0.5–3.0
Terminal methyl of methylene chain[90]	0.91
Aliphatic protons[90]	0.8–1.4
Paraffinic methyl[92]	0.5–1.65
Methyl protons of highly branched aliphatic[90]	1.26
Methylene of acyclic compounds, protons of aliphatic carbon β or more removed from aromatic rings or polar functional groups[90]	1.53
Acetate methyl[90]	1.91
Methylene, methine α to carbonyl/carboxyl/aromatic[94]	1.6–3.0
Methyl/methylene attached to electronegative groups[90]	1.91–2.68
Hydrogen from O–C–H[82]	3.0–5.5
Methyl, methylene, methine bonded to oxygen/nitrogen[94]	3.3–3.5
Methanol methyl[90]	3.35
Alcohol (–OH)	
O-,N-attached protons or protons of carbohydrate, methoxyl, amines[90]	3.35–4.10
Olefinic and aromatic protons[82]	5.5–8.5
Protons attached to unsaturated carbons[94]	5.5–9.0
Polyhydroxyphenols[90]	6.55
Unsubstituted and substituted aromatic protons[94]	6.55–7.7
Sterically unhindered aromatic protons[88]	6.55–7.67
Formate proton[94]	8.40

primarily by ^{1}H chemical shifts and existing knowledge of humate composition. Spin–spin coupling information is typically lost in humate spectra because the peaks are too broad to resolve the spin multiplets. Table 1 lists some ^{1}H chemical shifts and their substructure assignments for different humate preparations. These assignments, although valuable, should be considered tentative since the chemical shift range of various proton groups (e.g. methyl, methylene, methine) depends strongly on their electronic environment and is not sufficiently distinct to allow unequivocal assignment. For example, the chemical shift for methyl protons (<1 ppm) is normally lower than those of methylene and methine protons. However, when the methyl protons are attached to oxygen (i.e. methoxyl), their shifts can reach as high as 3–4 ppm or the methine proton range (Table 1). These ambiguities in chemical shift assignments need to be resolved by other means.

2-D ^{1}H-NMR spectroscopy has a tremendous potential in resolving many of the assignment ambiguities and

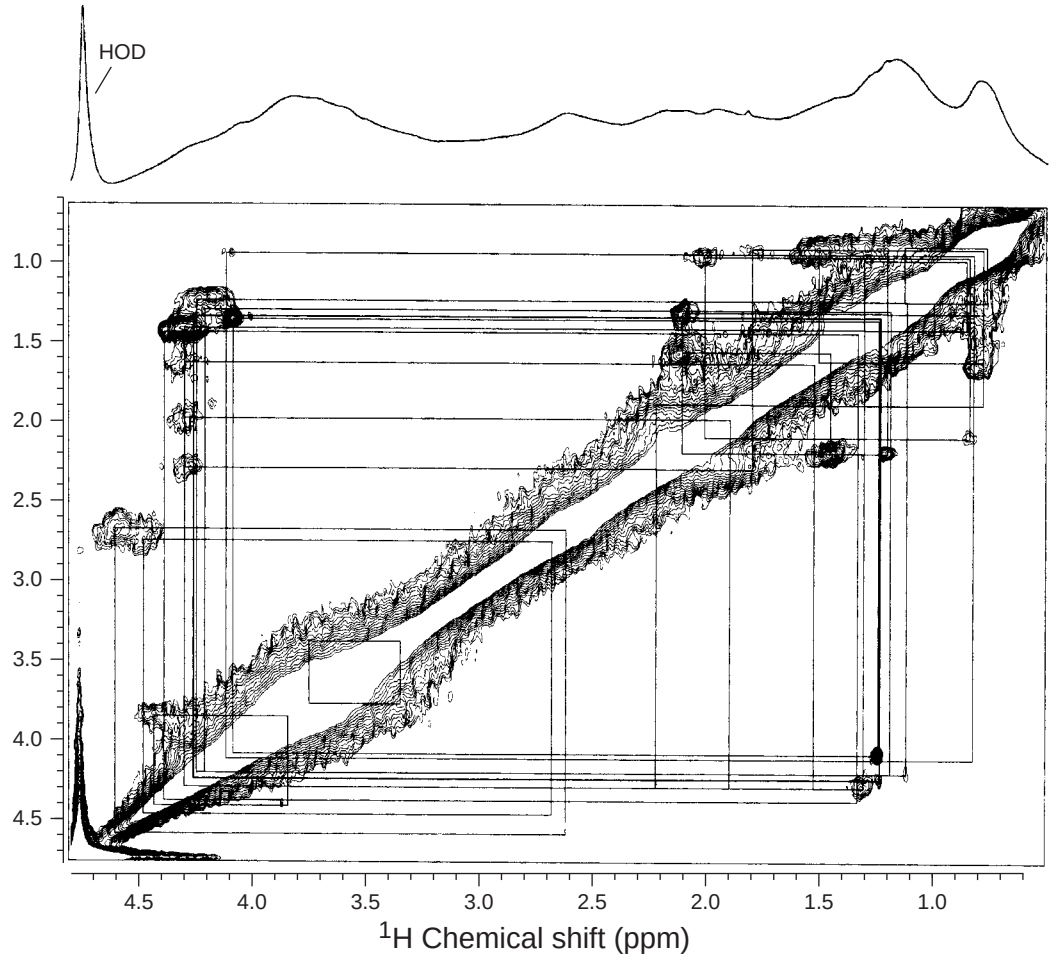

Figure 11 2-D ^{1}H-TOCSY spectrum of a forest soil (Forbes) humate. Forbes humate was extracted with 0.25 M NaOH[12] and dissolved in D_2O for NMR analysis. The TOCSY spectrum was acquired at 11.75 Tesla and displayed as contour plots. The rectangular boxes connect the diagonal peaks to the off-diagonal cross-peaks, the latter of which indicate covalent linkages between the diagonal peaks (Fan, Higashi, and Lane, unpublished results).

in obtaining conformational information, as shown by the numerous 2-D applications in structure/conformation determination of proteins and nucleic acids.[6] To date, only a few 2-D ^{1}H-NMR analyses of soil humates and fulvates have been reported.[12,92,94–96] Particularly interesting was the conformational analysis of humate by 2-D ^{1}H-NOESY that provided evidence for the presence of two distinct domains: a flexible aliphatic chain and a more condensed aromatic domain.[94] The following 2-D ^{1}H-NMR measurements further illustrate the values of 2-D NMR analysis.

Figure 11 shows a total correlation 2-D ^{1}H-NMR spectrum as a contour map of a humate preparation extracted from a Sierra Nevada range forest soil (Forbes). As described in section 1.3.1, TOCSY provides information on an intramolecular covalent bond network which is traced by the off-diagonal cross-peaks (noted by rectangular boxes in Figure 11). Consequently, substructure assignments are derived from multiple pieces of information rather than from chemical shift alone. For example, the resonances at 1.38, 1.6, 1.96, 2.28, and 4.31 ppm are correlated through their cross-peaks, which indicates a substructure of $Z{-}CH_2{-}CH_2{-}CH_2{-}CH_2{-}CH{-}XY$, where X, Y, and Z could be O-, N-containing functional groups or aromatic rings. Also evident in Figure 11 is the presence of similar substructures in slightly different chemical environments. For example, two groups of cross-peaks in the 1.2–1.4 and 4.1–4.4 ppm regions indicate the presence of the $CH_3{-}CH{-}XY$ structure. The small variation of the chemical shifts in these two regions suggests the presence of different X and/or Y functional groups in the substructure, which is consistent with the heterologous nature of humates.

Other than conformational analysis, 2-D ^{1}H-NOESY is useful in probing intermolecular interactions between humates and other ligands since NOE reveals interactions

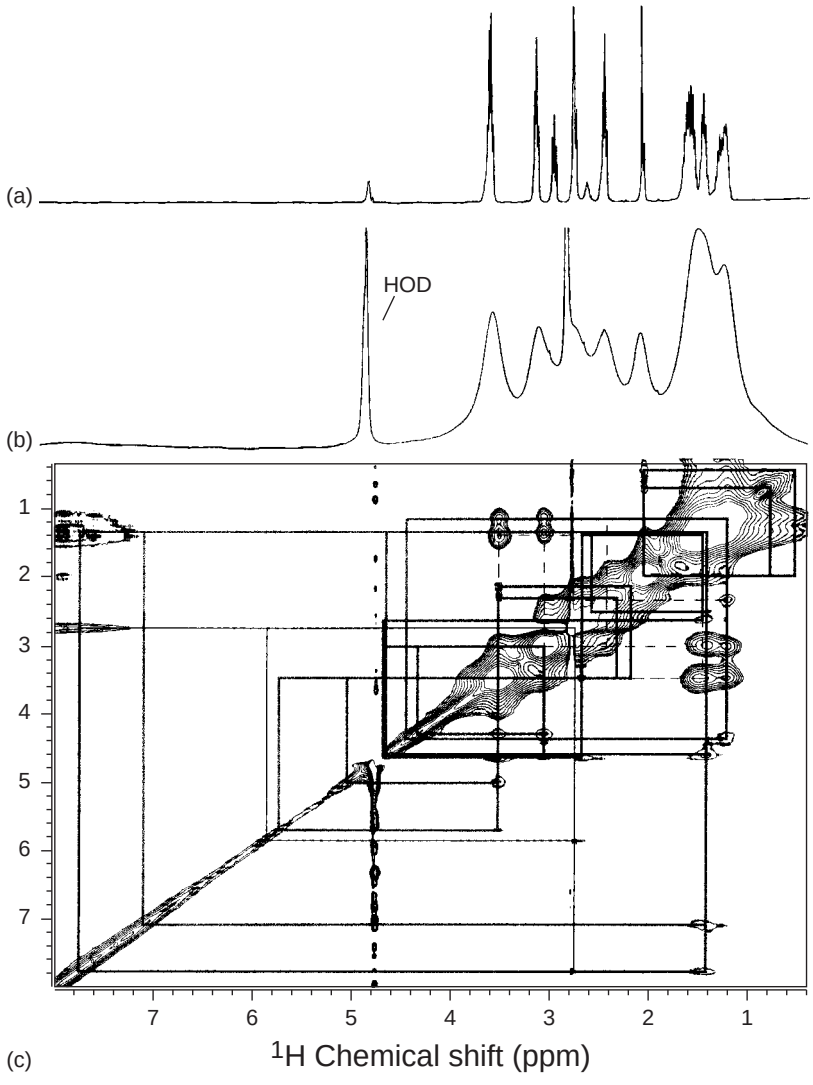

Figure 12 NMR evidence for the interaction between desferrioxamine B (DFOB) and Forbes humate. (a) 1-D ^{1}H-NMR spectrum of DFOB, a common and powerful organic ligand for transition metals in soils. Numbered spectral assignments were based on Borgias et al. (1989). (b) 1-D ^{1}H-NMR spectrum of Forbes humate in the presence of excess DFOB. (c) Contour plot of 2-D ^{1}H-NOESY of the sample from (b). The corresponding off-diagonal cross-peak pattern for free DFOB, HS, and DFOB + HS is represented by thin dashed lines, thin solid lines, and thick solid lines, respectively. These cross-peaks indicate regions of intra- and intermolecular associations via exchange and dipolar interactions. (Reprinted from Higashi et al., © 1998,[12] with permission of the Royal Society of Chemistry.)

between protons that are within 5 Å apart and/or in chemical exchange (see section 1.3.1). Figure 12 shows a NOESY spectrum as a contour map of Forbes humate (as in Figure 11) in the presence of a powerful bacteriosiderophore DFOB.[12] Also shown are the 1-D spectra of DFOB and DFOB + Forbes humate. Here, the off-diagonal cross-peaks represent dipolar couplings while the diagonal peaks correspond to those in the normal 1-D spectrum. From the NOESY spectrum, dipolar couplings originating from DFOB (dashed line) and humate (solid line) can be traced, most notably from 2.75 to 7.76 and 5.87 ppm for the humate. This proton connectivity is absent in the TOCSY spectrum (data not shown), indicating that the interactions are through space, not via scalar coupling. Thus, NOESY and TOCSY analyses are complementary and together they provide information on primary and higher-order structures of humates. In addition, NOESY cross-peaks reveal complexation of DFOB with humate, as traced by thick solid lines. These cross-peaks are absent in either DFOB or humate spectrum, thus providing unequivocal evidence for the binding of DFOB to Forbes humate. Moreover, the molecular detail of the binding sites can be deduced from the NOESY spectrum, which involves

the interaction between the diaminopentyl and succinyl protons of DFOB and the aromatic/phenolic and/or saccharidic groups of the humate. The complexation between siderophores (Fe chelators) or other metal ligands and humates is previously unknown and will have significance in metal transport and bioavailability in the soil environment.

2.2.3.2 ^{13}C-Nuclear Magnetic Resonance ^{13}C-NMR and solid-state analysis, in particular, is by far the most prevalent NMR technique employed for NOM characterization. Despite its low sensitivity, ^{13}C-NMR has been preferred over ^{1}H-NMR because of the greater chemical shift dispersion and the ability to observe important functional groups such as carbonyl, carboxyl, and quaternary carbons. Solid-state measurement is attractive from the standpoint of in situ soil analysis, better sensitivity,[88] and detection of more rigid or immobilized substructures. Humate and fulvate analysis in situ without extraction can eliminate preparation artifacts and avoids the use of reactive solvents such as NaOD and d_6-DMSO, as mentioned above. These, together with a shorter interpulse delay practical for solid-state CP/MAS acquisition,[5,97] can help improve detection sensitivity. Figure 13 compares the solid- and liquid-state ^{13}C-NMR spectra of the same soil fulvate.[98] The solid-state spectrum was acquired with MAS and ^{1}H to ^{13}C CP, which reduced the line-broadening effect of chemical shift anisotropy and the acquisition time to improve detection sensitivity. Both qualitative and quantitative differences are evident, including different sensitivity, spectral resolution, and relative peak heights. For example, a distinct peak at 211 ppm (often assigned to ketone or aldehydic carbons) in the solid-state spectrum is absent in the liquid-state spectrum. In addition, the relative intensity of the 0–80 ppm region (aliphatic carbons) is lower in the liquid-state than in the solid-state spectrum. These differences may result from the inability of liquid-state measurements to detect more rigid structures, which has been postulated for aliphatic groups in humates. Moreover, the relatively higher intensity of the aromatic region in the liquid-state spectrum may be attributed to a high aromatic substitution which would reduce the efficiency of CP transfer and therefore the ^{13}C signal intensity in the solid-state spectrum.[98] Quantitative differences between liquid- and solid-state ^{13}C spectra were also observed in another study, although the two spectra were qualitatively similar.[99]

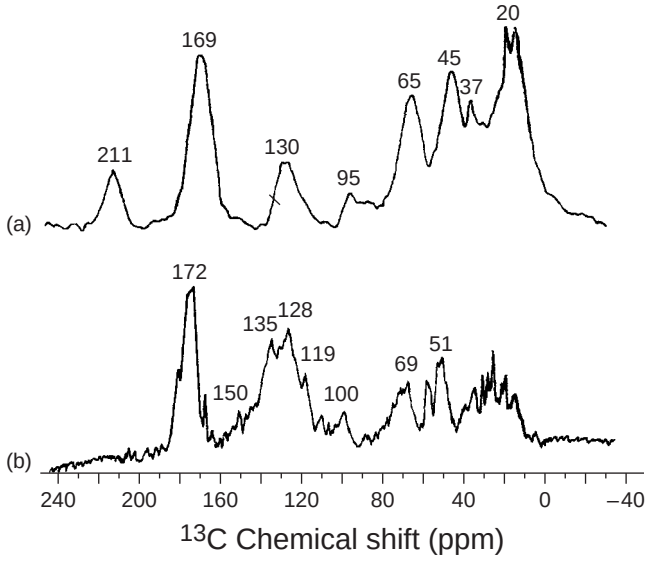

Figure 13 Comparison of liquid- (a) and solid-state (b) ^{13}C-NMR spectra of fulvate extracted from an Italian loam soil. (Reprinted from Conte et al., © 1997,[98] with permission from Elsevier Science.)

On the other hand, solid-state analysis is not without disadvantages. The analysis may be complicated by the presence of other organic compounds and factors such as the presence of excess paramagnetic ions[79,100] and degree of protonation in humates.[101,102] There are also complications associated with the CP/MAS technique.[69,88] MAS can lead to spinning side bands[65] while CP transfer may result in inconsistent peak quantification among different samples or among various chemical groups of the same sample. This is because peak intensities acquired under the CP transfer mode depend on the optimal contact time, which is known to vary with different humates and their chemical groups.[103] These problems need to be carefully addressed for quantitative analysis of humates by solid-state ^{13}C-NMR.

Despite the above limitations, CP/MAS ^{13}C-NMR is presently one of the most popular techniques for determining humate chemical groups and their abundance. Table 2 lists some ^{13}C assignments of chemical groups and their abundances in NOM. As is the case for ^{1}H-NMR, assignment ambiguities exist for ^{13}C-NMR when based on chemical shift alone. Different approaches have been employed to reduce these ambiguities, including the use of INEPT[81,104] and dipolar dephasing ^{13}C-NMR spectroscopy.[60,105] Verification of chemical group assignments by other spectroscopic methods such as infrared spectroscopy and wet chemical analysis[105] has also been utilized.

2.2.3.3 ^{15}N-Nuclear Magnetic Resonance Applications of ^{15}N-NMR in NOM analysis have been limited by experimental difficulties including the low intrinsic sensitivity and natural abundance of the ^{15}N nucleus, negative magnetogyric ratios that result in unfavorable NOEs, and long T_1 values. Inverse-gated decoupling and ^{1}H to ^{15}N polarization transfer (e.g. INEPT), and acoustic methods have been employed to eliminate the problems

Table 2 ^{13}C Chemical shift assignments and abundance in aquatic and terrestrial humates and fulvates

Assigned groups	^{13}C Chemical shift range (ppm)	Abundance (%)
Aliphatic carbon[98]	0–40	12.6–45.2[98]
Alkyl carbon[78]	0–46	15.9–48.2[78]
Aliphatic carbon (sp^3 hybridized)[107]	0–60	13–27[107]
Alkyl carbon[114]	10–48	26.7[114]
	0–50	19–39[43]
Unsubstituted aliphatic carbon[61]	38	
Aliphatic methoxyl carbon[61]	56	
	50–62	6–10[43]
Aliphatic carbons bonded to oxygen (ethers, alcohols, carbohydrates)[107]	60–90	3–15[107]
C–O or singly substituted oxygenated carbon[61]	61–93 (75)	15–19
O-alkyl carbon	62–95	15–30[43]
Peptidic and oxidized carbon[98]	40–110	21.4–46.8[98]
O-alkyl carbon[78]	46–110	22.4–49.9[78]
O-alkyl carbon[114]	48–110	26
Anomeric carbon[61]	106	
Aromatic and phenolic carbon[98]	110–160	0–46.2[98]
		7–21[65]
Unsubstituted aromatic carbon[78]	110–140	10.4–36.9[78]
Aromatic carbon[107]	90–160	31–62[107]
Protonated aromatic carbon[107]	90–140	
Di-O-alkyl carbon and aromatic carbon[43]	95–140	17–33
Olefinic carbon[107]	110–150	
Acetal and hemiacetal carbon[107]	90–110	
Aromatic or ring carbon[61]	128	
O,N-substituted aromatic carbon[78]	140–160	4.3–8.9[78]
Phenol and aromatic ether carbon[61]	153	
Aryl carbon[114]	110–165	37.5
Phenolic carbon[43]	140–160	6–13
Carboxylic, ester, amide, and lactone carbon[107]	160–180	14–22[107]
Carboxylic carbon[98]	160–190	10.1–21.3[98]
		10–18[65]
		4–8[43]
Carboxyl and ester carbon[61]	162–187 (172)	15–17
Carboxyl carbon[114]	165–190	9.8
Carbonyl carbon[78]	160–200	6.8–13.4[78]
Quinone and ketone carbon[107]	180–220	6–7[107]
Aldehydic carbon[107]	192–202	1.91
Ketonic carbon[61]	187–220 (195)	4–6
Carbonyl[43]	190–220	1–2
Methanol methyl[78]	3.35	
O-,N-attached protons or protons of carbohydrate, methoxyl, amines[78]	3.35–4.10	
Olefinic and aromatic protons[98]	5.5–8.5	
Polyhydroxyphenols[78]	6.55	
Unsubstituted and substituted aromatic protons[78]	6.5–7.7	
Sterically unhindered aromatic protons	6.55–7.67	
Formate proton[78]	8.40	

of unfavorable NOEs, to enhance the signal intensities of protonated nitrogens, and to overcome spectral baseline distortion, respectively.[76,106] The INEPT pulse sequence also allowed distinction between protonated and non-protonated nitrogens, but nitrogens attached to rapidly exchanging protons will be lost. Example spectra acquired under these conditions are shown in Figure 14.[76]

To overcome the low abundance problem, ^{15}N enrichment by derivatizing NOM with ^{15}N-labeled precursors

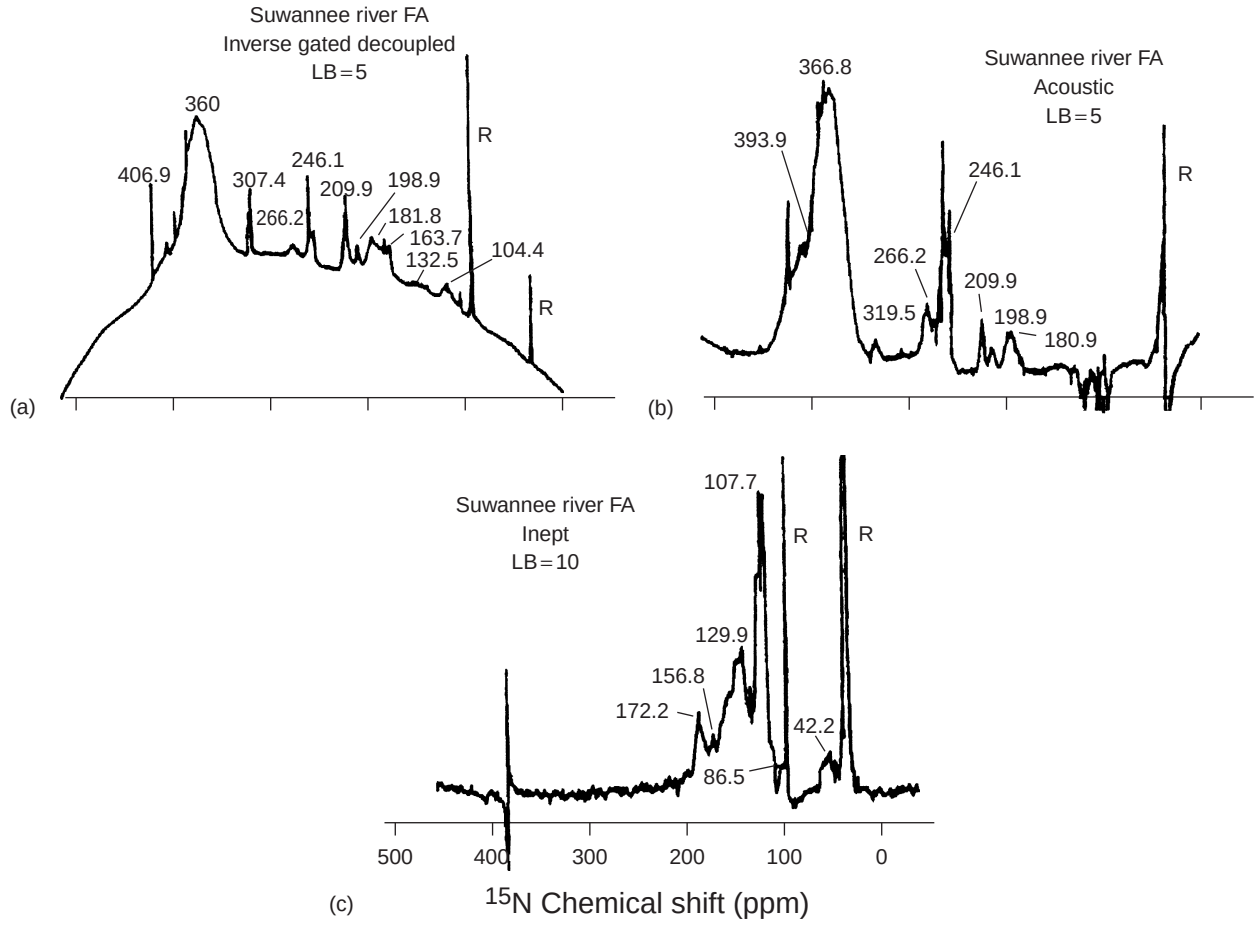

Figure 14 Liquid-state ^{15}N-NMR spectra of Suwannee River fulvate derivatized with ^{15}N-labeled hydroxylamine. (Reprinted from Thorn et al., © 1992,[76] with permission of American Chemical Society.)

Table 3 Assigned ^{15}N-containing functional groups in humates/fulvates derived from ^{15}N-precursors[106]

Added ^{15}N precursors	Functional groups identified in humates/fulvates					
	Oxime	Nitroso phenol	Heterocycle	Amide	Aniline derivatives	Other
NH_4OH[76]			M	S	S	Aminoquinone
NH_2OH[77]						Oxazole, nitrile, hydroxamic acid
NH_2Cl[110]			S	M	T	Aminoquinone
Urea[4]						Lactams, peptides, carbamates
Glycine[115]				T		NH_4^+
Aniline[107]			M		S	Anilinoquinone
NO_2^-[77]			M	S		
NO_3^-[108,109]				S		Peptides, NH_4^+, indole, aminosugar, chitin, nucleic acids

M = major; S = significant; T = trace.

(e.g. ^{15}N-ammonium compounds, ^{15}N-nitrate, and ^{15}N-amino acids, and ^{15}N-herbicides) has been used. Such manipulation not only allowed usable ^{15}N-NMR spectra to be obtained but more importantly it revealed the molecular mechanism(s) of NOM interaction with herbicides,[107] and NOM's role in nitrogen fertilizer,[108,109] nitrification/denitrification processes,[76,77] and transport of aquatic pollutants.[110] A number of N-containing

functional groups have been assigned in NOM that were derived from ^{15}N-labeled fertilizers, amino acids, and herbicides (Table 3, adapted from Steelink[106]). Again, these assignments were mainly based on chemical shifts, which would need verification since ^{15}N chemical shifts are highly sensitive to solvent polarity, pH, presence of paramagnetic ions, and ionic strength.

2.2.3.4 ^{31}P-Nuclear Magnetic Resonance The application of ^{31}P-NMR to NOM analysis was initiated in the early 1980s,[74] and subsequently several studies on soils,[111] compost, and sewage sludge[83,97] have been reported. Example ^{31}P-NMR spectra of humates are shown in Figure 15, where phosphonates, inorganic orthophosphate, inositol hexaphosphate, phosphate monoester, aromatic phosphate diester and nucleic acid, and phosphate diester were assigned[111] Also evident in Figure 15 is the dependence of the ^{31}P profile on the molecular size fraction of humates. Variable ^{31}P profiles were also observed with different cultivation and fertilization history. These studies demonstrate some of the potential values of ^{31}P-NMR in probing NOM structures.

However, there has been relatively little method development in ^{31}P-NMR for NOM analysis. Two major issues need to be resolved, i.e. validation of chemical-shift-based resonance assignment (since ^{31}P chemical shifts are highly dependent on pH, ionic strength, and paramagnetic ions) and quantitative reliability. The latter issue would require systematic studies on extraction recovery of phosphate functional groups (particularly under alkaline conditions), ^{31}P relaxation properties, and NMR ''visibility'' of different ^{31}P signals. Since phosphates have a high affinity towards multivalent cations, the presence of these ions (particularly paramagnetic ions) can significantly affect the line widths of different phosphate signals, and therefore their peak intensities. For quantitative analysis it is necessary to adopt procedures that remove these ions.

2.2.3.5 Other Nuclei Although potentially powerful, there have been few applications of other NMR nuclei to NOM analysis, most likely due to the low concentration and a general lack of knowledge of these nuclei in NOM. In two related studies, ^{27}Al- and ^{29}Si-NMR was employed to examine the aluminum and silicon–oxygen coordination chemistry in soils and clay fractions.[79,80] ^{29}Si-NMR was also applied to analyze for carboxylic, phenolic, and alcoholic groups in silylated humates.[82] In a ^{51}V-NMR study, vanadate reduction and subsequent speciation in the presence of aquatic humates was reported.[84] ^{113}Cd-NMR was employed to probe complexation of Cd^{2+} with NOM and provided valuable evidence for Cd^{2+} coordination with the oxygen and nitrogen donors of NOM.[85] With the dramatic improvement in NMR technology and increasing knowledge of NOM structures, it is foreseeable that nuclei other than ^{13}C, ^{1}H, ^{15}N, and ^{31}P may receive increasing attention.

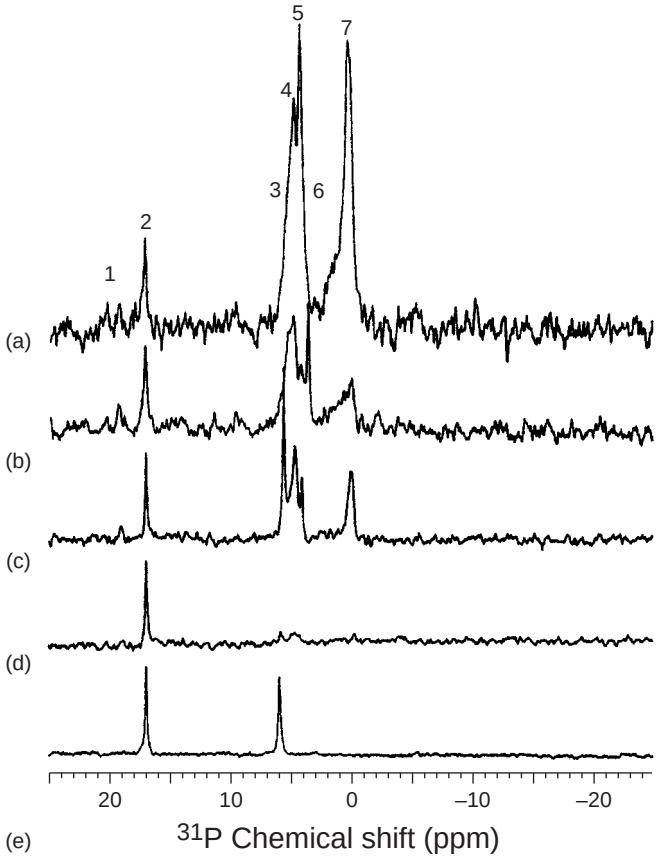

Figure 15 ^{31}P-NMR spectra of an example humate and its various molecular size fractions: (a) humate; (b) >300 kDa; (c) 30–300 kDa; (d) 1–30 kDa; (e) <1 kDa. Resonances assigned: 1, phosphonate (18.9 ppm); 2, internal standard (methylene diphosphonate, 16.8 ppm); 3, inorganic orthophosphate (5.9 ppm); 4, inositol hexaphosphate (4.35 ppm); 5, phosphate monoester (3.9–5.0 ppm); 6, aromatic phosphate diester and nucleic acid (0.3–3.0 ppm); 7, phosphate diester (−0.3 ppm). (Reprinted from Bedrock et al., © 1994,[111] with permission from Elsevier Science.)

3 FUTURE PROSPECTS

3.1 Resolution

For aqueous biological extracts, a large number of compounds can be resolved and identified simultaneously by 2-D NMR such as TOCSY (see above). However, in the sugar region (3.2–4.7 ppm), only a few of the common mono and disaccharides are discernible because of the severe spectral overlap in this region. Resolution can be improved with increasing magnetic field strength, e.g. from low hundred MHz to 800 MHz or 900 MHz, which

is to become available in the near future. However, these ultrahigh field instruments are access limited and the improvement in resolution is modest. A more effective means for improving resolution is to add extra frequency dimensions to NMR experiments. It is now feasible to conduct homonuclear 3-D NMR experiments, such as TOCSY/TOCSY or TOCSY/ROESY, although they are more acquisition time-consuming (e.g. 3 days) and the spectral analysis is more complicated. An alternative to homonuclear 3-D NMR is the heteronuclear approach, e.g. combining TOCSY with HSQC. Here the low natural abundance of ^{13}C and ^{15}N can be a limitation. However, they should be extremely valuable for the analysis of isotopically enriched metabolites.

3.2 Detection Limits

The NMR sensitivity is a function of the instrument hardware and the sample, with the two often interactive. The hardware is continuously improving, e.g. higher magnetic field strength, higher signal-to-noise amplifiers, and improved probe design. Higher sensitivity can also be attained by increasing sample size and/or concentration in the detection region. The standard NMR probe accepts 5-mm diameter tubes, which readily gives a good line shape and good matching and tuning for optimal sensitivity. Larger diameter probes often yield poorer sensitivity due to nonoptimal line shapes and tuning. With improved probe design, the sample size can be increased (e.g. to 8 mm diameter) without compromising line shape and tunability, such that the expected gain (e.g. 2.5-fold compared with 5-mm diameter size) is obtained. When the amount of sample is limited, higher sample concentrations (e.g. twofold) can be obtained by restricting the active volume with doped glasses that match the magnetic susceptibility of the solvent (e.g. Shigemi tubes). Moreover, the advent of pulsed field gradient methods affords much better solvent suppression capability and flexibility in multidimensional NMR experiments, both contributing to better detection limits.

3.3 New Applications

The majority of NMR applications in agriculture have been in the area of chemical identification. As NMR is also a powerful technique for probing molecular dynamics and conformation, applications along these areas should prove to be very fruitful. From the few studies available in the literature,[12,85] the use of NMR analysis for metal speciation and organometallic chemistry, exchange dynamics between ligands and NOM, and NOM conformation and binding site characterization is expected to break exciting new ground in the understanding of plant–soil interactions.

ACKNOWLEDGMENTS

We wish to thank Dr R.M. Higashi for helpful suggestions on the manuscript. We also acknowledge the financial support from US Department of Energy grants DE-FG07-96ER20255 and DE-FG03-97ER62349, as well as US Environmental Protection Agency grant R825960010. Although the information in this document has been funded partly by the US Environmental Protection Agency, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

ABBREVIATIONS AND ACRONYMS

ATP	Adenosine 5′-Triphosphate
COSY	Correlation Spectroscopy
CP	Cross-polarization
d_6-DMSO	d_6-Dimethylsulfoxide
DFOB	Desferrioxamine B
DQF/COSY	Double Quantum Filtered Correlation Spectroscopy
GAB	γ-Aminobutyrate
HETCOR	Heteronuclear Correlation
HMBC	Heteronuclear Multiple Bond Correlation
HOHAHA	Homonuclear Hartmann–Hahn
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclear Enhancement by Polarization Transfer
MAS	Magic Angle Spinning
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NOM	Natural Organic Matter
NTP	Nucleoside 5′-Triphosphate
ROESY	Rotating Frame Overhauser Effect Spectroscopy
TOCSY	Total Correlation Spectroscopy
UDPGlc	Uridine 5′-Diphosphate Glucose
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Nuclear Magnetic Resonance of Biomolecules

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring • Soil Instrumental Methods

Food **(Volume 5)**
Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Steel and Related Materials **(Volume 10)**
Nuclear Magnetic Resonance in Metals Analysis • X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Nuclear Magnetic Resonance of Geological Materials and Glasses • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. J. Cavanagh, W.J. Fairbrother, A.G. Palmer, A.G.N.J. Skelton, *Protein NMR Spectroscopy Principles and Practice*, Academic Press, San Diego, 1996.
2. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1990.
3. D.G. Gadian, *Nuclear Magnetic Resonance and Its Application to Living Systems*, Clarendon Press, Oxford, 1982.
4. K.K.M. Sanders, B.K. Hunter, *Modern NMR Spectroscopy. A Guide for Chemists*, Oxford University Press, Oxford, 1987.
5. E.O. Stejskal, J.D. Memory, *High Resolution NMR in the Solid State. Fundamentals of CP/MAS*, Oxford University Press, Oxford, 1994.
6. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
7. T.W.-M. Fan, 'Recent Advances in Profiling Plant Metabolites by Multi-nuclear and Multi-dimensional NMR', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Shachar-Hill, P.E. Pfeffer, The American Society of Plant Physiologists, Rockville, Maryland, 181–254, 1996.
8. T.W.-M. Fan, 'Metabolite Profiling by One- and Two-dimensional NMR Analysis of Complex Mixtures', *Prog. NMR Spectrosc.*, **28**, 161–219 (1996).
9. G.A. Morris, R. Freeman, 'Enhancement of Nuclear Magnetic Resonance Signals by Polarisation Transfer', *J. Am. Chem. Soc.*, **101**, 760–762 (1979).
10. L. Braunschweiler, R.R. Ernst, 'Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy', *J. Magn. Reson.*, **53**, 521–528 (1983).
11. A. Bax, D.G. Davis, 'MLEV-17 based Two-dimensional Homonuclear Magnetization Transfer Spectroscopy', *J. Magn. Reson.*, **65**, 355–360 (1985).
12. R.M. Higashi, T.W.-M. Fan, A.N. Lane, 'Association of Desferrioxamine with Humic Substances and their Interaction with Cadmium(II) as Studied by Pyrolysis Gas Chromatography Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy', *Analyst*, **123**(5), 911–918 (1998).
13. A.A. Bothner-By, R.L. Stephens, J.M. Lee, C.D. Warren, R.W.J. Jeanloz, 'Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame', *J. Am. Chem. Soc.*, **106**, 811–813 (1984).
14. A. Bax, D.G. Davis, 'Practical Aspects of Two-dimensional Transverse NOE Spectroscopy', *J. Magn. Reson.*, **63**, 207–213 (1985).
15. T.W.-M. Fan, R.M. Higashi, A.N. Lane, O. Jardetzky, 'Combined Use of ^{1}H NMR and GC-MS for Monitoring

Metabolites and *In Vivo* ^{1}H NMR', *Biochim. Biophys. Acta*, **882**, 154–167 (1986).
16. A. Bax, G.A. Morris, 'An Improved Method for Heteronuclear Chemical Shift Correlation by Two-dimensional NMR', *J. Magn. Reson.*, **42**, 501–505 (1981).
17. G. Bodenhausen, D.J. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185–189 (1980).
18. D.H. Live, D.G. Davis, W.C. Agosta, D. Cowburn, 'Observation of 1000-fold Enhancement of ^{15}N NMR via Proton-detected Multiquantum Coherences: Studies of Large Peptides', *J. Am. Chem. Soc.*, **106**, 6104–6105 (1984).
19. G.W. Kellogg, 'Proton-detected HeteroTOCSY Experiments with Applications to Nucleic Acids', *J. Magn. Reson.*, **98**, 176–182 (1992).
20. M.J. Gradwell, T.W.-M. Fan, A.N. Lane, 'Analysis of Phosphorylated Metabolites by ^{31}P{^{1}H} NMR Heteronuclear Total Correlation Spectroscopy (HeteroTOCSY)', *Anal. Biochem.*, **263**, 139–149 (1998).
21. G.M. Clore, A. Bax, P. Wingfield, A.M. Gronenborn, 'Long Range ^{15}N–^{1}H Correlation as an Aid to Sequential Proton Resonance Assignment of Proteins', *FEBS Lett.*, **238**, 17–21 (1988).
22. M. Piotto, V. Saudek, V. Sklenar, 'Gradient-tailored Excitation for Single-quantum NMR Spectroscopy of Aqueous Solutions', *J. Biomol. Struct.*, **2**, 661–665 (1992).
23. J.K.M. Roberts, M.P. Testa, '^{31}P NMR Spectroscopy of Roots of Intact Corn Seedlings', *Plant Physiol.*, **86**, 1127–1130 (1988).
24. R.B. Lee, R.G. Ratcliffe, 'Development of an Aeration System for use in Plant Tissue NMR Experiments', *J. Exp. Bot.*, **34**, 1213–1221 (1983).
25. T.W.-M. Fan, R.M. Higashi, A.N. Lane, 'Monitoring of Hypoxic Metabolism in Superfused Plant Tissues by *In Vivo* ^{1}H NMR', *Arch. Biochem. Biophys.*, **251**(2), 674–687 (1986).
26. G.G. Fox, R.G. Ratcliffe, T.E. Southon, '^{31}P NMR Observations on the Effect of the External pH on the Intracellular pH Values in Plant Cell Suspension Cultures', *Plant Physiol.*, **93**, 512–521 (1990).
27. C. Roby, J.-B. Martin, R. Bligny, R. Douce, 'Biochemical Changes During Sucrose Deprivation in Higher Plant Cells', *J. Biol. Chem.*, **262**, 5000–5007 (1987).
28. R. Callies, R. Altenburger, A. Mayer, L.H. Grimme, D. Leibfritz, 'A New Illumination System for *In Vivo* NMR Spectroscopy', *J. Magn. Reson.*, **90**, 561–566 (1990).
29. R.G. Ratcliffe, '*In Vivo* NMR Spectroscopy: Biochemical and Physiological Applications to Plants', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Shachar-Hill, P.E. Pfeffer, The American Society of Plant Physiologists, Rockville, Maryland, 1–32, 1996.
30. T.W.-M. Fan, R.M. Higashi, A.N. Lane, 'An *In Vivo* ^{1}H and ^{31}P NMR Investigation of the Effect of Nitrate on Hypoxic Metabolism in Maize Roots', *Arch. Biochem. Biophys.*, **266**, 592–606 (1988).
31. K.J. Chang, J.K.M. Roberts, 'Observation of Cytoplasmic and Vacuolar Malate in Maize Root Tips by ^{13}C-NMR Spectroscopy', *Plant Physiol.*, **89**, 197–203 (1992).
32. S. Aubert, R. Bligny, R. Douce, 'NMR Studies of Metabolism in Cell Suspensions and Tissue Cultures', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Shachar-Hill, P.E. Pfeffer, The American Society of Plant Physiologists, Rockville, Maryland, 109–154, 1996.
33. T.W.-M. Fan, A.N. Lane, R.M. Higashi, 'Hypoxia does Not Affect Rate of ATP Synthesis and Energy Metabolism in Rice Shoot Tips as Measured by ^{31}P NMR *In Vivo*', *Arch. Biochem. Biophys.*, **294**, 314–318 (1992).
34. T. Matsunaga, T. Nagata, '*In Vivo* B-11 NMR Observation of Plant Tissues', *Anal. Sci.*, **11**, 889–892 (1995).
35. P.E. Pfeffer, Y. Shachar-Hill, 'Plant/microbe Symbiosis', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Shachar-Hill, P.E. Pfeffer, The American Society of Plant Physiologists, Rockville, Maryland, 77–107, 1996.
36. R.B. Lee, J.V. Purves, R.G. Ratcliffe, L.R. Saker, 'Nitrogen Assimilation and the Control of Ammonium and Nitrate Absorption by Maize Roots', *J. Exp. Bot.*, **43**, 1385–1396 (1992).
37. T.W.-M. Fan, R.M. Higashi, J.D. Norlyn, E. Epstein, '*In Vivo* ^{23}Na and ^{31}P NMR Monitoring of a Tonoplast Na^{+}/H^{+} Exchange Process and its Characteristics in Two Barley Cultivars', *Proc. Natl. Acad. Sci. USA*, **86**, 9856–9860 (1989).
38. L.O. Sillerud, J.W. Heyser, 'Use of ^{23}Na-Nuclear Magnetic Resonance to Follow Sodium Uptake and Efflux in NaCl-adapted and Nonadapted Millet (*Panicum miliaceum*) Suspensions', *Plant Physiol.*, **75**, 269–272 (1984).
39. F. Martin, P. Rubini, R. Cote, I. Kottke, 'Aluminum Polyphosphate Complexes in Mycorrhizal Basidiomycete *Laccaria bicolor*: a ^{27}Al-Nuclear Magnetic Resonance Study', *Planta*, **194**, 241–246 (1994).
40. R.G. Ratcliffe, '*In Vivo* NMR Studies of Higher Plants and Algae', *Adv. Bot. Res.*, **20**, 43–123 (1994).
41. J.K.M. Roberts, 'Study of Plant Metabolism *In Vivo* Using NMR Spectroscopy', *Ann. Rev. Plant Physiol.*, **35**, 375–386 (1984).
42. T. Ogino, J.A. den Hollander, R.G. Shulman, '^{39}K, ^{23}Na, ^{31}P NMR Studies of Ion Transport in *Saccharomyces cerevisiae*', *Proc. Natl. Acad. Sci. USA*, **80**, 5185–5189 (1983).

43. Y.Y. Ford, G.G. Fox, R.G. Ratcliffe, R.J. Robins, '*In Vivo* ^{15}N NMR Studies of Secondary Metabolism in Transformed Root Cultures of *Datura stramonium* and *Nicotiana tabacum*', *Phytochemistry*, **36**, 333–339 (1994).
44. Y. Shachar-Hill, P.E. Pfeffer, M.W. Germann, 'Following Plant Metabolism *In Vivo* and in Extracts with Heteronuclear Two-dimensional Nuclear Magnetic Resonance Spectroscopy', *Anal. Biochem.*, **243**(1), 110–118 (1996).
45. J.K.M. Roberts, N. Wade-Jardetzky, O. Jardetzky, 'Intracellular pH Measurements by ^{31}P Nuclear Magnetic Resonance Influence of Factors other than pH on ^{31}P Chemical Shifts', *Biochemistry*, **20**, 5389–5394 (1981).
46. E. Gout, R. Bligny, R. Douce, 'Regulation of Intracellular pH Values in Higher Plant Cells. Carbon-13 and Phosphorus-31 Nuclear Magnetic Resonance Studies', *J. Biol. Chem.*, **267**, 13 903–13 909 (1992).
47. J.S. Taylor, C.J. Deutsch, 'Fluorinated α-Methylamino Acids as ^{19}F NMR Indicators of Intracellular pH', *Biophys. J.*, **43**, 261–267 (1983).
48. E. Gout, R. Bligny, N. Pascal, R. Douce, '^{13}C Nuclear Magnetic Resonance Studies of Malate and Citrate Synthesis and Compartmentation in Higher Plant Cells', *J. Biol. Chem.*, **268**, 3986–3992 (1993).
49. J.K.M. Roberts, J.-H. Xia, 'NMR Contribution to Understanding of Plant Responses to Low Oxygen Stress', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Shachar-Hill, P.E. Pfeffer, The American Society of Plant Physiologists, Rockville, Maryland, 155–180, 1996.
50. Y. Shachar-Hill, P.E. Pfeffer, D. Douds, S.F. Osman, L.W. Doner, R.G. Ratcliffe, 'Partitioning of Intermediary Carbon Metabolism in Vesicular–Arbuscular Mycorrhizal Leek', *Plant Physiol.*, **108**, 7–15 (1995).
51. A.D. Carroll, G.G. Fox, S. Laurie, R. Phillips, R.G. Ratcliffe, G.R. Stewart, 'Ammonium Assimilation and the Role of γ-Aminobutyric Acid in pH Homeostasis in Carrot Cell Suspensions', *Plant Physiol.*, **106**, 513–520 (1994).
52. T.W.-M. Fan, R.M. Higashi, A.N. Lane, 'Energy and Fermentation Metabolism in Hypoxic Rice Coleoptiles – a Multinuclear NMR Approach', NATO Workshop, Wye College, 1993.
53. T.D. Colmer, T.W.-M. Fan, A. Läuchli, R.M. Higashi, 'Interactive Effects of Salinity, Nitrogen and Sulphur on the Organic Solutes in *Spartina alterniflora* Leaf Blades', *J. Exp. Bot.*, **47**(296), 369–375 (1996b).
54. T.W.-M. Fan, R.M. Higashi, T.A. Frenkiel, A.N. Lane, 'Anaerobic Nitrate and Ammonium Metabolism in Flood-tolerant Rice Coleoptiles', *J. Exp. Bot.*, **48**(314), 1655–1666 (1997).
55. T.W.-M. Fan, A.N. Lane, R.M. Higashi, 'Selenium Biotransformations by a Euryhaline Microalga Isolated from a Saline Evaporation Pond', *Environ. Sci. Technol.*, **31**, 569–576 (1997).
56. T.W.-M. Fan, R.M. Higashi, A.N. Lane, 'Biotransformations of Selenium Oxyanion by Filamentous Cyanophyte Cultured from Agricultural Drainage Waters', *Environ. Sci. Technol.*, **32**, 3185–3193 (1998).
57. T.W.-M. Fan, A.N. Lane, D. Martens, R.M. Higashi, 'Synthesis and Structure Characterization of Selenium Metabolites', *Analyst*, **123**(5), 875–884 (1998).
58. J.A. Manthey, D.E. Crowley, D.G. Luster, *Biochemistry of Metal Micronutrients in the Rhizosphere*, Lewis Publishers, Boca Raton, 1994.
59. T.W.-M. Fan, A.N. Lane, J. Pedler, D. Crowley, R.M. Higashi, 'Comprehensive Analysis of Organic Ligands in Whole Root Exudates Using Nuclear Magnetic Resonance and Gas Chromatography–Mass Spectrometry', *Anal. Biochem.*, **251**, 57–68 (1997).
60. I. Kögel-Knabner, P.G. Hatcher, E.W. Tegelaar, J.W. de Leeuw, 'Aliphatic Components of Forest Soil Organic Matter as Determined by Solid-state ^{13}C NMR and Analytical Pyrolysis', *Sci. Tot. Environ.*, **113**, 89–106 (1992).
61. M.H.B. Hayes, P. MacCarthy, R. Malcolm, R.S. Swift (eds.), Humic Substances II. *In Search of Structure*, Wiley, Chichester, 1989.
62. P.G. Hatcher, R. Rowan, M.A. Mattingly, '^{1}H and ^{13}C NMR of Marine Humic Acids', *Org. Geochem.*, **2**, 77–85 (1980).
63. J.I. Hedges, 'Polymerization of Humic Substances in Natural Environments', in *Humic Substances and Their Role in the Environment*, eds. F.H. Frimmel, R.F. Christman, Wiley, Chichester, 45–58, 1988.
64. B.K.G. Theng, G.J. Churchman, R.H. Newman, 'The Occurrence of Interlayer Clay–Organic Complexes in Two New Zealand Soils', *Soil Sci.*, **142**, 262–266 (1986).
65. R.H. Newman, K.R. Tate, '^{13}C NMR Characterization of Humic Acids from Soils of a Development Series', *J. Soil Sci.*, **42**, 39–46 (1991).
66. M.H.B. Hayes, R.S. Swift, 'Genesis, Isolation, Composition and Structures of Soil Humic Substances', in *ASI Series. Series B: Physics*, eds. M.F. de Boodt, M.H.B. Hayes, A. Herbillon, Plenum Press, New York, 245–305, Vol. 215, 1990.
67. U. Müller-Wegener, 'Interaction of Humic Substances with Biota', in *Humic Substances and Their Role in the Environment*, eds. F.H. Frimmel, R.F. Christman, Wiley, Chichester, 179–192, 1988.
68. D. Chapman, V. Kimstach, 'The Selection of Water Quality Variables', in *Water Quality Assessments*, ed. D. Chapman, Chapman and Hall, London, 81, 1992.
69. C.M. Preston, 'Applications of NMR to Soil Organic Matter Analysis: History and Prospects', *Soil Sci.*, **161**, 144–166 (1996).
70. A. Wais, P. Burauel, A.A. de Graaf, K. Haider, F. Führ, 'Using ^{13}C-NMR Spectroscopy to Evaluate the Binding Mechanism of Bound Pesticide Residues in Soils. 2. Investigations of Model Humic Acids', *J. Environ. Sci. Health*, **B31**(1), 1–24 (1996).

71. G. Almendros, F.J. González-Vila, F. Martin, R. Fründ, H.-D. Lüdemann, 'Solid State NMR Studies of Fire-induced Changes in the Structure of Humic Substances', *Sci. Tot. Environ.*, **117/118**, 63–74 (1992).
72. A. Golchin, J.A. Baldock, P. Clarke, T. Higashi, J.M. Oades, 'The Effects of Vegetation and Burning on the Chemical Composition of Soil Organic Matter of a Volcanic Ash Soil as Shown by ^{13}C NMR Spectroscopy. II. Density of Fractions', *Geoderma*, **76**, 175–192 (1997).
73. P.G. Hatcher, J.M. Bortlatynski, R.D. Minard, J. Dec, J.-M. Bollag, 'Use of High-resolution ^{13}C NMR to Examine the Enzymatic Covalent Binding of ^{13}C-labeled 2,4-Dichlorophenol to Humic Substances', *Environ. Sci. Technol.*, **27**, 2098–2103 (1993).
74. R.H. Newman, K.R. Tate, 'Soil Phosphorus Characterization by ^{31}P Nuclear Magnetic Resonance', *Commun. Soil Sci. Plant Anal.*, **11**, 835–842 (1980).
75. L.M. Condron, K.M. Goh, R.H. Newman, 'Nature and Distribution of Soil Phosphorus as Revealed by ^{31}P Nuclear Magnetic Resonance Analysis', *J. Soil Sci.*, **36**, 199–207 (1985).
76. K.A. Thorn, J.B. Arterburn, M.A. Mikita, '^{15}N and ^{13}C NMR Investigation of Hydroxylamine-derivatized Humic Substances', *Environ. Sci. Technol.*, **26**, 107–116 (1992).
77. K.A. Thorn, M.A. Mikita, 'Ammonia Fixation by Humic Substances: a Nitrogen-15 and Carbon-13 NMR Study', *Sci. Tot. Environ.*, **113**, 67–87 (1992).
78. G. Almendros, M.E. Guadalix, F.J. Gonzalez-Vila, F. Martin, 'Distribution of Structural Units in Humic Substances as Revealed by Multi-step Selective Degradation and ^{13}C-NMR of Successive Residues', *Soil Biol. Biochem.*, **30**(6), 755–765 (1998).
79. M.A. Wilson, *NMR Techniques and Application in Geochemistry and Soil Chemistry*, 1st edition, Pergamon Press, London, 1987.
80. P.F. Barron, M.A. Wilson, A.S. Campbell, R.I. Frost, 'Detection of Imogolite in Soils Using Solid State ^{29}Si NMR', *Nature*, **299**, 616–618 (1982).
81. K.A. Thorn, D.W. Folan, J.B. Arterburn, M.A. Mikita, P. Maccarthy, 'Application of INEPT Nitrogen-15 and Silicon-29 Nuclear Magnetic Resonance Spectrometry to Derivatized Fulvic Acids', *Sci. Tot. Environ.*, **81/82**, 209–218 (1989).
82. H. Herzog, P. Burba, J. Buddrus, 'Quantification of Hydroxylic Groups in a River Humic Substance by ^{29}Si-NMR', *Fresenius' J. Anal. Chem.*, **354**, 375–377 (1996).
83. Z.R. Hinedi, A.C. Chang, R.W.K. Lee, 'Mineralization of Phosphorus in Sludge-amended Soils Monitored by Phosphorus-31 Nuclear Magnetic Resonance Spectroscopy', *Soil Sci. Soc. Am. J.*, **52**, 1593–1596 (1988).
84. X. Lu, W.D. Johnson, J. Hook, 'Reaction of Vanadate with Aquatic Humic Substances: an ESR and ^{51}V NMR Study', *Environ. Sci. Technol.*, **32**, 2257–2263 (1998).
85. J. Li, E.M. Perdue, L.T. Gelbaum, 'Using Cadmium-113 NMR Spectroscopy to Study Metal Complexation by Natural Organic Matter', *Environ. Sci. Technol.*, **32**, 483–487 (1998).
86. C.K. Larive, A. Rogers, M. Morton, W.R. Carper, '^{113}Cd NMR Binding Studies of Cd–Fulvic Acid Complexes: Evidence of Fast Exchange', *Environ. Sci. Technol.*, **30**, 2828–2831 (1996).
87. K.H. Chung, S.W. Rhee, H.S. Shin, C.H. Moon, 'Probe of Cadmium(II) Binding on Soil Fulvic Acid Investigated by Cd-113 NMR Spectroscopy', *Can. J. Chem.*, **74**, 1360–1365 (1996).
88. M.A. Wilson, 'Solid-state Nuclear Magnetic Resonance Spectroscopy of Humic Substances: Basic Concepts and Techniques', in *Humic Substances II*, eds. M.H.B. Hayes, P. MacCarthy, R.L. Malcolm, R.S. Swift, Wiley, New York, 310–335, 1989.
89. K. Yonebayashi, T. Hattori, 'Chemical and Biological Studies on Environmental Humic Acids II. ^{1}H-NMR and IR Spectra of Humic Acids', *Soil Sci. Plant Nutr.*, **35**(3), 383–392 (1989).
90. A. Jokić, R. Srejić, P.A. Pfendt, J. Zakrzewska, 'Characterization of Lake Sediment Humic Acids from the Ćelije Lake System Near Kruševac (Central Serbia) by ^{13}C and ^{1}H Solution NMR and ^{13}C CPMAS NMR', *Water Air Soil Pollut.*, **84**, 159–173 (1995).
91. M. Kawahigashi, N. Fujitake, T. Takahashi, 'Structural Information Obtained from Spectral Analysis (UV/Vis, IR, ^{1}H NMR) of Particle Size Fractions in Two Humic Acids', *Soil Sci. Plant Nutr.*, **42**(2), 355–360 (1996).
92. J. Lambert, J. Buddrus, 'Quantification of Isolated Methyl Groups in Aquatic Humic Substances by Means of ^{1}H and ^{13}C NMR Spectroscopy', *Magn. Reson. Chem.*, **34**, 276–282 (1996).
93. H. Herzog, S. Haiber, P. Burba, J. Buddrus, 'Quantification of Partial Structures of Aquatic Humic Substances by ^{1}H-NMR under Water Conditions', *Fresenius' J. Anal. Chem.*, **359**, 167–170 (1997).
94. Y.-Y. Chien, W.F. Bleam, 'Two-dimensional NOESY Nuclear Magnetic Resonance Study of pH-dependent Changes in Humic Acid Conformation in Aqueous Solution', *Environ. Sci. Technol.*, **32**(23), 3653–3658 (1998).
95. J. Buddrus, P. Burba, H. Herzog, J. Lambert, 'Quantitation of Partial Structures of Aquatic Humic Substances by One-dimensional and Two-dimensional Solution C-13 Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **61**, 628–631 (1989).
96. M.A. Wilson, J.V. Hanna, K.B. Anderson, R.E. Botto, 'H-1 Cramps NMR Derived Hydrogen Distributions in Various Coal Macerals', *Org. Geochem.*, **20**, 985–999 (1993).
97. C.M. Preston, 'Review of Solution NMR of Humic Substances', in *NMR in Humic Substances and Coal*, eds. R.L. Wershaw, M.A. Mikita, Lewis Publishers, Chelsea, MI, 3–32, 1987.
98. P. Conte, A. Piccolo, B. Van Lagen, P. Buurman, P.A. de Jager, 'Quantitative Differences in Evaluating Soil

Humic Substances by Liquid and Solid State ^{13}C-NMR Spectroscopy', *Geoderma*, **80**, 339–352 (1997).
99. R. Fründ, H.-D. Lüdemann, 'The Quantitative Analysis of Solution and CPMAS-C-13 NMR Spectra of Humic Material', *Sci. Tot. Environ.*, **81/82**, 157–168 (1989).
100. P.E. Pfeffer, W.V. Gerasimowicz, E.G. Piotrowski, 'Effect of Paramagnetic Iron on Quantitation in Carbon-13 Cross-polarization Magic Angle Spinning Nuclear Magnetic Resonance Spectroscopy of Heterogeneous Environmental Matrices', *Anal. Chem.*, **56**, 734–741 (1984).
101. L.B. Alemany, D.M. Grant, R.J. Pugmuire, T.D. Alger, K.W. Zilm, 'Cross-polarization and Magic Angle Sample Spinning NMR Spectra of Model Organic Compounds, I. Highly Protonated Molecules', *J. Am. Chem. Soc.*, **105**, 2133–2141 (1983).
102. L.B. Alemany, D.M. Grant, R.J. Pugmuire, T.D. Alger, K.W. Zilm, 'Cross-polarization and Magic Angle Sample Spinning NMR Spectra of Model Organic Compounds, II. Molecules of Low Remote Protonation', *J. Am. Chem. Soc.*, **105**, 2142–2147 (1983).
103. P. Conte, A. Piccolo, B. Van Lagen, P. Buurman, P.A. de Jager, 'Quantitative Aspects of Solid-state ^{13}C-NMR Spectra of Humic Substances from Soils of Volcanic Systems', *Geoderma*, **80**, 327–338 (1997).
104. G. Ricca, F. Severini, 'Structural Investigations of Humic Substances by IR-FT, ^{13}C-NMR Spectroscopy and Comparison with a Maleic Oligomer of Known Structure', *Geoderma*, **58**, 233–244 (1993).
105. A. Ramunni, V. Pignalosa, C. Amalfitano, 'The Lignin Input in the Structure of the Humic Acids from a Farmyard Manured Soil as Detected by FT-IR, UV-visible, and ^{13}C NMR CP-MAS Spectroscopy', *Agrochimica*, **36**(3), 268–281 (1992).
106. C. Steelink, 'Application of N-15 NMR Spectroscopy to the Study of Organic Nitrogen and Humic Substances in the Soil', in *Humic Substances in the Global Environment and Implications on Human Health*, eds. N. Senesi, T.M. Miano, Elsevier Science B.V., London, 405–426, 1994.
107. E.J. Weber, K.A. Thorn, D.L. Spidel, 'Kinetic and N-15 NMR Spectroscopic Studies of the Covalent Binding of Aniline to Humic Substances, Book of Abstracts', *203rd National ACS Meeting*, San Francisco, CA, April 5–10, 1992.
108. L. Benzing-Purdie, M.V. Cheshire, B.L. Williams, G.P. Sparling, C.I. Ratcliffe, J.A. Ripmeester, 'Fate of N-15 Glycine in Peat as Determined by C-13 and N-15 CPMAS NMR Spectroscopy', *J. Agric. Food. Chem.*, **34**, 170–176 (1986).
109. G. Almendros, R. Freund, F.J. Gonzalez-Villa, K.M. Haider, H. Knicker, H. Lüdemann, 'Analysis of C-13 and N-15 CPMAS NMR Spectra of Soil Organic Matter and Composts', *FEBS Lett.*, **282**(1), 119–121 (1991).
110. A.S. Ginwalla, M.A. Mikita, 'Reaction of Suwannee River Fulvic Acid with Chloramine Characterization of Products via N-15 NMR', *Environ. Sci. Technol.*, **26**, 1148–1150 (1992).
111. C.N. Bedrock, M.V. Cheshire, J.A. Chudek, B.A. Goodman, C.A. Shand, '^{31}P NMR Studies of Humic Acid from a Blanket Peat', in *Humic Substances in the Global Environment and Implications on Human Health*, eds. N. Senesi, T.M. Miano, Elsevier Science B.V., London, 227–232, 1994.
112. T.D. Colmer, T.W.-M. Fan, R.M. Higashi, A. Läuchli, 'Interactions of Ca^{2+} and NaCl Stress on the Ion Relations and Intracellular pH of *Sorghum bicolor* Root Tips: an *In Vivo* ^{31}P-NMR Study', *J. Exp. Bot.*, **45**(277), 1037–1044 (1994).
113. B. Borgias, A.D. Hugi, K.N. Raymond, 'Isomerization and Solution Structures of Desferrioxamine B Complexes of Al^{3+} and Ga^{3+}', *Inorg. Chem.*, **28**, 3538–3545 (1989).
114. B. Dehorter, C.Y. Kontchou, R. Blondeau, '^{13}C-NMR Spectroscopy Analysis of Soil Humic Acids Recovered after Incubation with Some White Rot Fungi and Actinomycetes', *Soil Biol. Biochem.*, **24**(7), 667–673 (1992).
115. L. Benzing-Purdie, M.V. Cheshire, B.L. Williams, C.I. Ratcliffe, J.A. Ripmeester, B.A. Goodman, 'Interactions between Peat and Sodium Acetate, Ammonium Sulfate, Urea or Wheat Straw during Incubation Studied by C-13 and N-15 NMR Spectroscopy', *J. Soil Sci.*, **43**(1), 113–125 (1992).

Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Jeffrey S. de Ropp and Michael J. McCarthy
University of California at Davis, Davis, USA

1 Introduction **1**
2 History **1**
3 Instrumentation **2**
3.1 Low-resolution Analysis 2
3.2 High-resolution Analysis 2
3.3 Magnetic Resonance Imaging Analysis 3
4 Foodstuffs **3**
4.1 Laboratory Quality Assurance 3
4.2 Soluble Components 4
4.3 Water 5
4.4 Adulteration and Identification 5
4.5 Structure and Texture 6
4.6 Physical and Transport Properties 7
4.7 Functional Imaging 7
5 Plant Materials **7**
5.1 Introduction 7
5.2 Magnetic Resonance Imaging of Plants 8
5.3 Nuclear Magnetic Resonance Spectroscopy of Plants 12
5.4 Nuclear Magnetic Resonance Spectroscopy of Plant Extracts 14
Abbreviations and Acronyms **17**
Related Articles **17**
References **18**

Nuclear magnetic resonance (NMR) is an experimental technique based on the interaction between an applied magnetic field and nuclear magnetic moments in a sample. NMR provides information on material composition, molecular structure, physical structure, physical properties and dynamics of phase changes. NMR is a noncontacting and noninvasive measurement. NMR has been used for the routine analysis of foodstuffs and plant materials. NMR is used in the analysis of foods to detect moisture content, solid/liquid ratio in fats, adulteration of juices and characterization of texture. Applications of NMR to plant materials include observation and characterization of metabolites and ions, determination of pH and following changes of these over time. NMR is also used to determine structures of natural products isolated from plants. Magnetic resonance imaging (MRI) is used in both foodstuffs and plant materials to study morphology and also water transport, diffusion and variations of these under different conditions. The response of NMR is linear from the detection limits of approximately 10 ppb to 100%. The accuracy of the technique depends upon the signal-to-noise ratio (S/N), which is a function of the sample size, applied magnetic field strength, the radiofrequency (RF) coil, the spectrometer hardware and experimental parameters. Typical accuracy is 0.5% for composition and 1–5% for viscosity.

1 INTRODUCTION

NMR analytical applications in food and plant materials are extensive. The applications range from simple determinations of water content to the determination of molecular structure of compounds and the internal physical structure of the material. The focus within this article is on both standard techniques and emerging techniques. Low-resolution nuclear magnetic resonance (LRNMR) has been used extensively in quality assurance laboratories to measure average sample properties such as fat content and water droplet size distributions. More recently high-resolution and multidimensional experiments have been applied to characterize composition, authenticate the material, measure the internal physical structure and characterize component interactions and functions.

NMR has several advantages as an analytical technique; these include minimal sample preparation, rapid analysis and the potential to run multiple tests on a single sample. To date, primarily LRNMR and NMR spectroscopy have been used in analytical laboratories; however, recent advances in automating experiment set-up and operation coupled with the development of new methods indicate expansion of the use of NMR spectroscopy and the beginning of the application of imaging systems in analytical settings.

2 HISTORY

The initial successful NMR experiments were performed in the laboratories of F. Bloch[1] and E. Purcell.[2] The discovery of the chemical shift[3] permitted the expansion to chemical applications. The use of NMR in the analysis of chemical and biochemical molecules is an essential tool in research and more recently in analytical laboratories.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Advances in fields other than food technology have provided NMR spectroscopists with the ability to improve dramatically both experiments and data analysis. In the area of magnet design, advances enabling very high magnetic field systems have allowed NMR to be used to study the fine detail of molecular structure. Advances in computer technology coupled with digital fast Fourier transforms have permitted the development of multidimensional NMR methods.[4–6] Occurring simultaneously with the development of multidimensional NMR was the technique of generating images of the internal structure of materials.[7,8] Called NMR imaging or MRI, this technique has been primarily developed as a standard radiological diagnostic method.

Use of NMR for routine analytical testing has primarily focused on the use of low-resolution systems to measure sample composition (e.g. moisture content, fat content or polymer content) and/or physical state of the sample (e.g. solid-fat to liquid-fat ratio). These applications are prevalent in the chemical, agricultural, food and medical industries. One of the oldest applications, the use of LRNMR to measure the solid-to-liquid ratio in cocoa butter in chocolate manufacture, was implemented during the 1960s.

3 INSTRUMENTATION

NMR instrumentation used for analytical purposes is based on similar components. Each spectrometer will consist of a magnet, an RF network, amplifiers, a low-noise receiver, an RF coil, a pulse programmer and a computer system for data analysis and system control. These spectrometers can be classified broadly by placing them in categories of low-resolution, high-resolution and MRI analysis. The differentiation between these categories is not distinct and often an NMR spectrometer may be classified in two of the categories, by its use and by its operational characteristics. This distinction between the spectrometers is both in terms of the experiments performed on the system and the magnetic field strength and RF network.

3.1 Low-resolution Analysis

LRNMR spectrometers utilize a low-field permanent magnet to generate the magnetic field. The range of magnetic field strength is 0.1–0.65 T (4.3–27.7 MHz proton frequency). The analysis of the NMR signal is performed in the time domain. The characteristics of the signal that are used for analysis are as follows:

1. The initial amplitude of the signal is directly proportional to the total number of protons in the sample.
2. The proton signal is a superposition of the signals from all proton-containing components.
3. Separation of the signals from different components is based on differences in relaxation rates between different components.

Low-resolution system costs are relatively inexpensive compared with high-resolution systems/MRI spectrometers. The range of cost is typically between $20 000 and $150 000, the difference being the accessories ordered with the basic system. Available accessories include pulsed-field-gradient units and temperature-controlled probes. The instruments have been designed specifically for routine analysis and to achieve a high degree of accuracy and reproducibility.

3.2 High-resolution Analysis

The major differences between low-resolution and high-resolution nuclear magnetic resonance (HRNMR) spectrometers are the properties of the magnet and the use of frequency domain analysis. For high-resolution spectrometers, the main magnetic field is significantly stronger than for low-resolution systems. High-resolution spectrometers utilize a magnet with field strengths normally ranging from 7 to 14 T. Costs of complete systems in this field strength range are typically $300 000–$800 000.

A high-resolution experiment differentiates between nuclei in one molecular environment and another based on differences in nuclear resonance frequency. Differences in the frequency result from variation in structure and the local chemical and electronic environment for each nucleus. The frequency of excitation (ν) for each nucleus is proportional to the magnetic field at the site of that nucleus. The differences in frequency are described by Equation (1):

$$\nu = \frac{\gamma B_0(1-\sigma)}{2\pi} \tag{1}$$

where ν is the nuclear resonance or Larmor frequency in megahertz, B_0 is the magnetic field strength, σ is the shielding constant and γ is the gyromagnetic ratio. Values of the shielding constant range between 10^{-3} and 10^{-6} for most nuclei.[9] Equation (1) demonstrates the importance of increasing the field strength, since the actual frequency separation is proportional to the static magnetic field strength.

For many complex mixtures and systems with small variations in shielding constants, multidimensional (N-dimensional, where N generally is 2 but also sometimes 3 or even 4) experiments are used. Multidimensional experiments generate spectral data that are a function of two or more independent variables. Common two-dimensional (2-D) experiments include such independent variable pairs as chemical shift vs scalar couplings and chemical shift vs dipolar couplings. Data in N-dimensional experiments are collected with N time domains and subjected to

N Fourier transforms to produce *N* axes in the frequency domain.

3.3 Magnetic Resonance Imaging Analysis

MRI experiments can be considered a subset of multidimensional HRNMR experiments. In MRI the independent variable pairs most often relate directly to spatial locations and hence a spatial map of signal intensity is recorded. The additional hardware to permit this spatial localization is the use of pulsed linear magnetic field gradients. The effect of the linear magnetic field gradients is to make the frequency of the spin proportional to its spatial location. The frequency variation developed by the pulsed linear magnetic field gradient must be greater than other sources of frequency variation in the sample (e.g. chemical shift and line broadening).

4 FOODSTUFFS

4.1 Laboratory Quality Assurance

LRNMR has become a standard instrument for use in laboratory quality assurance of fats and oils and for determining the composition of oil seeds.[10,11] Methods employing NMR have been standardized and approved by the American Oil Chemists' Society (AOCS) (AOCSd 16b-93, AOCS Ak 4-95), the International Union of Pure and Applied Chemistry (IUPAC) (solid fat content, IUPAC Norm 2.150) and the International Standards Organization (ISO) (oil seeds, ISO Dis/10565, ISO CD 10632). Additional quality assurance applications of LRNMR are listed in Table 1. The detection limits for LRNMR are a few hundred parts per million for hydrogen and fluorine nuclei. Most low-resolution experiments are based on using a ratio of the signal intensity from two different points in time during the free induction decay (FID) or on the multipoint analysis of the time decay of the signal.

Solid fat content measurements are one of the most successful applications of LRNMR. This measurement can be implemented in a variety of ways with LRNMR. The most common is the direct method that relies on the ratio of NMR signal amplitude at two times during the FID. Figure 1 shows the time evolution of the signal in a direct determination of the solid fat content. The sample of fat is tempered prior to being placed in the NMR spectrometer. After placement in the magnet the sample is permitted to achieve equilibrium which commonly takes between 2 and 6 s. A single RF pulse is applied to the sample and the signal is recorded. The intensity of the signal immediately after the RF pulse is a measure of the total number of protons in the sample and is proportional to the solid plus the liquid protons. The protons in the solid phase of the fat decay much faster than the protons from the liquid phase. This permits the signal from the two phases to be separated by the difference in rates of signal decay.[12] The signal intensity measured after the decay of the solid signal is proportional to the liquid-phase protons only. The intensity of the liquid-phase protons recorded at 70 μs is unaffected by the slight time delay. For example, the T_2 of liquid oils is approximately 100 ms, so the signal recorded at 70 μs is reduced in intensity by less than 0.1% from the initial value.

In practice, the signal immediately after the RF pulse is not recorded for 7–10 μs, because the receiver network of the NMR spectrometer requires time for recovery from the intense RF radiation during the excitation pulse. This recovery time is referred to as the dead time. This dead time is a significant part of the decay of the proton signal from the solid fat fraction. The influence of the dead time on the solid-phase determination is corrected for by using

Table 1 Quality assurance-type applications of NMR in the food industry

Moisture content	Oil content, solid fat content or solid/liquid ratio
Cornstarch	Chocolate
Wheat flour	Edible oils and fats
Rice flour	Marzipan
Corn flour	Rice
Lentils	Olive husk
Dried milk	Milk powders
Skim milk	Baby foods
Milk powders	Milk-based drinks
Cereal grains	Almonds
Corn meal	Mustard
Biscuits and cookies	Biscuits and cookies
Peanuts	Peanuts
Baby foods	Cheese
Oilseeds	Oilseeds

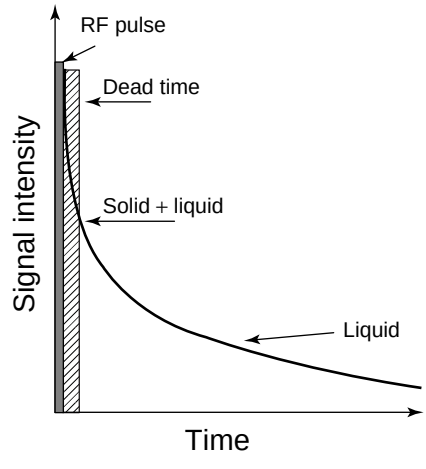

Figure 1 Signal decay from an LRNMR spectrometer.

a factor commonly called the "f factor". It is assumed that the solid signal immediately after the RF pulse is directly proportional to the f factor times the solid signal following the dead time[12] (Equation 2):

$$S_s(0\,\mu s) = fS_s(10\,\mu s) \quad (2)$$

The fraction of solid phase is then calculated by Equation (3):

$$\text{solid phase fraction} = \frac{fS_s(10\,\mu s)}{[S_1(70\,\mu s) + fS_s(10\,\mu s)]} \quad (3)$$

The numerical value of f will depend upon the duration of the dead time and the decay rate of the solid-phase protons. The correction factor is determined using standard oil samples. Ideally the correction factor should be close to unity.

Critical to the success of measuring the solid fat content is the temperature history of the sample prior to measurement.[12] The temperature control of the tempering block used to hold the sample prior to measurement should be at least 0.1 °C. It should be noted that the measurement temperature in most LRNMR systems is 40 °C and hence the oil sample should not be kept in the magnet for more than 10 s. After 10 s in the LRNMR spectrometer the solid phase will begin to melt.

The method described above is commonly referred to as the direct method. This is considered the fastest, simplest and most convenient technique.[12] The reproducibility of the direct method is 0.4%. Other methods employ either a variation of the direct method or curve deconvolution of a complete magnetization decay curve.[12] There are many other LRNMR applications for quality assurance and some common examples are listed in Table 1. Detailed descriptions of these techniques can be found in Bruker Application Notes.[13] An introduction to the most common methods was given by Barker and Stronks.[10] The most recent developments in LRNMR applications are the extension of these techniques to include particle size measurements and measurements of ^{19}F content.

4.2 Soluble Components

High-field studies (500 MHz or greater proton resonance frequency) have proven to be very important in the analysis of chemical and biochemical structures in solution.[4,14,15] Indeed, high-field spectrometers dedicated to molecular solution structure of species from small organic molecules to biological macromolecules are a standard part of the landscape of universities and major private companies. The advantage of high-field spectrometers is the good dispersion and very high S/N.[16] An additional advantage is a very high dynamic range that permits the characterization of molecules with both strong and weak signals at the same time. Applications in food analyses include improving the understanding of ripening in fruits, characterizing the role of fungal and microbial contamination in fruits and juices and authentication of fruit juices. Shown in Figure 2 are one-dimensional (1-D) proton spectra from different apple varieties.

Even with high-field spectrometers the 1-D spectra of plant extracts will contain many overlapping lines. There are several solutions to this problem. One is to utilize multidimensional techniques to resolve separate sets of resonances for each species present in solution. Another is to use a combination of liquid chromatography (LC) and NMR (LC/NMR) to separate the species chemically before spectra are recorded. Useful experiments for detection of specific compounds by NMR are proton homonuclear correlation "through-bond" pulse sequences such as correlation spectroscopy (COSY)[17] and total correlation spectroscopy (TOCSY).[18] Also important are one bond inverse proton–carbon correlation methods such as heteronuclear multiple quantum coherence (HMQC) spectroscopy[19,20]

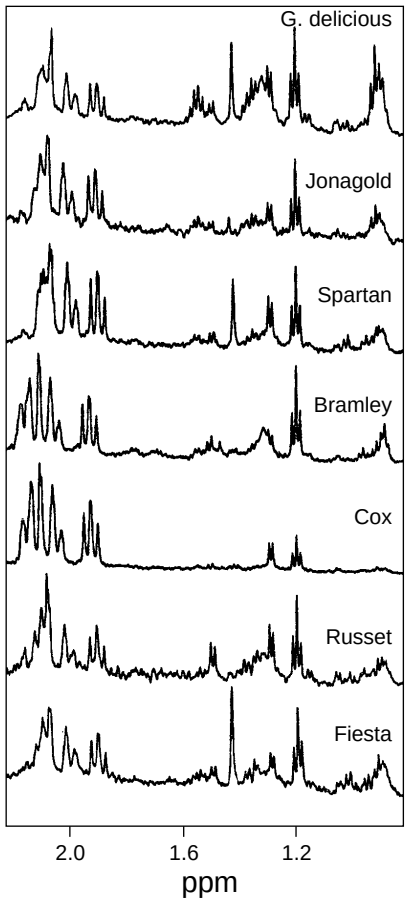

Figure 2 Partial spectra of seven different varieties of apple juices. (Reproduced by permission of Aspen Publishers from P.S. Belton, I. Delgadillo, A.M. Gil, *Semin. Food Anal.*, **3**, 223–234 (1998).)

and heteronuclear single quantum coherence (HSQC) spectroscopy.[21]

LC/NMR is a new and developing method of analysis in which NMR is used as the detector following LC separation. This technique has been used to detect and identify flavanoids in fruit juices and to characterize sugars in wine.[22] The technique is capable of detecting quantities in the lower nanogram range using specially designed flow probes and multiple solvent suppression routines.

Another emerging method is the use of high-resolution magic-angle spinning (HRMAS) to reduce or eliminate the dipolar broadening. The samples are spun at an angle of 54° relative to the magnetic axis at speeds between 3 and 15 kHz.[23,24] This spinning reduces or eliminates the broadening and permits the acquisition of liquid-like spectra in one or two dimensions. These techniques have been used to study wheat proteins and to identify taste-determining compounds in cheese.[22]

4.3 Water

The state of water in foods has been a matter of controversy for many years. The goal has been to derive correlations between water mobility and biological activity and/or food quality. Characterizing the properties of water in a complex and heterogeneous environment such as foods has proven to be very difficult. An approach that has been attempted many times is to measure the relaxation properties (spin–lattice relaxation time, T_1, and spin–spin relaxation time, T_2) of water and use these to characterize the mobility of water. In general, these approaches have not met with significant success due to the complexity of the relaxation mechanisms.

There are many factors which influence the relaxation times of water protons in foods, including intermolecular interactions between protons on different water molecules, intramolecular interactions between two protons on one water molecule, interactions between a water proton and a proton on a macromolecule and exchange between a water proton and a proton bound to a macromolecule.[25] A simple model for a two-state macromolecular hydration exchange is shown in Figure 3. The model assumes that one layer of ''bound'' water with a more rapid relaxation rate is in fast exchange with the remaining water which has properties like those of ''free'' water. This appears to be a reasonable model for protein solutions with more than 50% water. If the water content is reduced below 50% then there are more than two types of water and the range of relaxation times can be vary broad. Recent work has shown that it is most often sufficient to model water behavior with three states in biopolymer systems and biological tissues.[26]

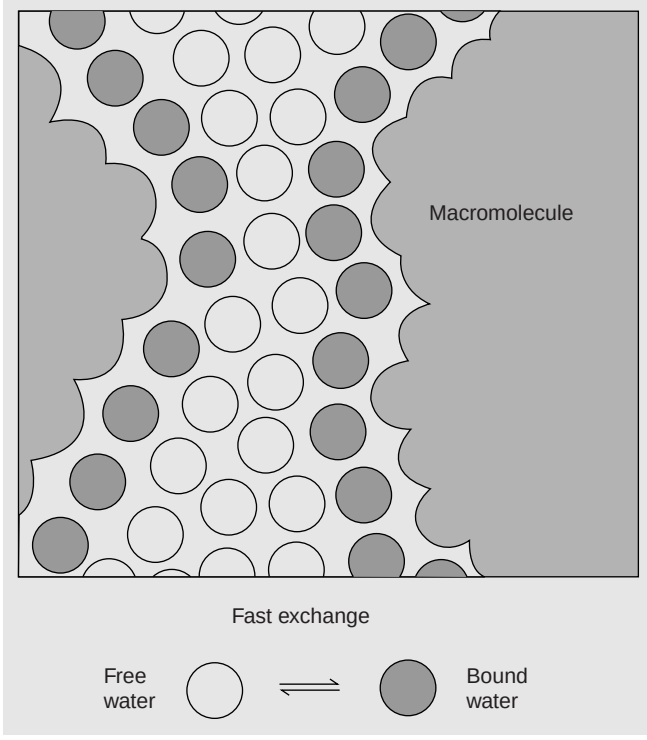

Figure 3 The two-state macromolecular hydration exchange model assumes one layer of ''bound'' water with modified properties in fast exchange with ''free'' water. The bound water is 2–3 times slower than free water and amounts to ~50 g per 100 g of macromolecule. (Reproduced from H.J.C. Berendsen, 'Rationale for Using NMR to Study Water Relations in Foods and Biological Tissues', *Trends Food Sci. Technol.*, **3**, 202–205 (1992), Copyright 1992, with permission from Elsevier Science.)

4.4 Adulteration and Identification

During the 1980s and 1990s, new analytical methods based on NMR and mass spectrometry (MS) were developed to detect food adulteration.[27–30] Results from MS give average isotope ratios for the entire sample, e.g. $^{13}C/^{12}C$ or $^{2}H/^{1}H$ ratios. These ratios can indicate food adulteration; however, this method can be circumvented by isotopic adjustment through addition of a synthetic to the food. The utilization of NMR isotopic analysis has the advantage of a multiisotopic characterization that is more difficult to simulate than an overall isotope ratio. This type of NMR analysis is called SNIF-NMR® (site-specific natural isotope fractionation studied by nuclear magnetic resonance).[30] SNIF-NMR® has been applied primarily to determine whether foods such as fruit juices, flavors, syrup, honey, wines or liquor have been adulterated or inappropriately labeled.

Important atoms with stable isotopes for analyses of food materials are hydrogen, oxygen and carbon. These isotopes occur in specific proportions, e.g. in the case of hydrogen ^{1}H is 99.985% and ^{2}H is 0.015% abundant (^{3}H is essentially nonexistent in nature). The isotopic content of natural molecules will vary based on the photosynthetic metabolism of the plant and on geographical and climatic conditions. The average value of deuterium in water is 150 ppm, yet specific concentrations depend on location. The concentration of deuterium in water at the earth's poles, for example, is approximately 70 ppm.[30]

Water is an important reactant in many metabolic processes in plants. The plant will absorb water from the environment and utilize the water to form proteins, sugars and other metabolites. Each plant species will have specific reaction pathways to produce molecules, and different plants will produce the same natural chemicals from different pathways. Hence the combination of different sources of water and different biochemical pathways results in different distributions of deuterium in the molecule that each plant produces. Consider orange juice and beet juice. Each plant makes sugar yet the SNIF-NMR® hydrogen value for sugar from oranges is different than that from beets. Hence the adulteration of orange juice using beet sugar can be detected.[30]

4.5 Structure and Texture

The size of structural components and their spatial distribution in a processed food are important quality attributes that can be measured using NMR and MRI. The range of length scales accessible with NMR is from the molecular level to the size of the food product itself. Standard MRI protocols can be used to measure the distribution of fat, sugar and water.[31] An MRI image of a rice pudding is shown in Figure 4. The change in moisture content and incompletely hydrated rice grains are clearly distinguished. Structure on the scale of an emulsion droplet can be measured using pulsed-field-gradient nuclear magnetic resonance (PFGNMR) spectroscopy, an experiment designed to measure the self-diffusion of molecules. HRNMR or relaxation time studies can be used to characterize the structure on a molecular level. These features allow the use of NMR techniques to characterize structure in almost all types of food products.[31]

Studies using MRI to characterize the sample on a visual length scale include the determination of defects in cheeses. The size, distribution and shape of the eyes in Swiss-style cheese have been measured using MRI proton images analyzed using standard image processing techniques. All physical features of the cheese were graded according to United States Department of Agriculture (USDA) standards. However, quality factors associated with visual perception such as dullness or color could not be graded using MRI.[32]

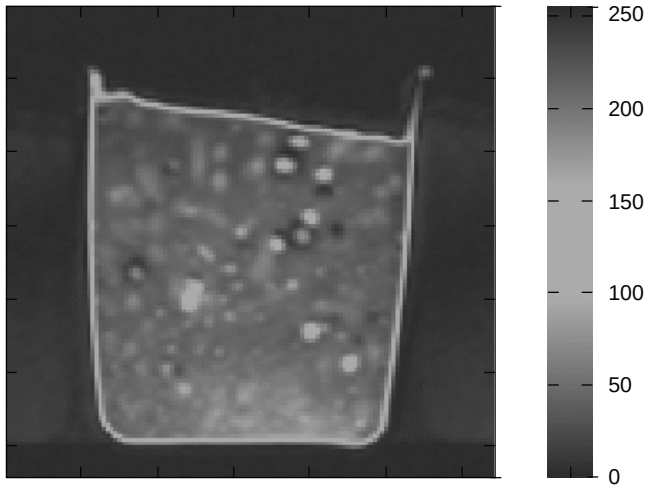

Figure 4 Proton MRI of the central plane through a rice pudding. The image intensity is weighted by spin–spin relaxation effects.

PFGNMR is used to measure the droplet size distribution in food emulsions such as margarine.[33] The procedure is suitable for use both in the laboratory and at a process line.[34] When a system has a heterogeneous structure, the translational motion of diffusing molecules is strongly influenced by the internal structure of the sample. In a confined geometry, the maximum extent of diffusional displacement is limited by the size of the confinement. For example, the diffusion of oil molecules within an emulsion droplet is restricted by the geometry and the dimensions of the internal droplet boundaries. This effect is referred to as ''restricted diffusion''. Analysis of the signal decay in a PFGNMR experiment can provide a measure of the characteristic dimensions of the confinement geometry. One of the most common uses of PFGNMR is to measure the size distribution of water droplets in margarine.[35,36] When this technique is applied to liquid emulsions the temperature of the emulsion may need to be controlled to eliminate the effects of molecules diffusing between droplets.[34,35]

Table 2 Applications of MRI

Processing studies	Food structure/quality
Drying	Meat
Freeze drying	Cheese
Rheology	Fruit
Extrusion	Crystallization
Freezing	Confectionary
Heating	Baked products
Cooling	Vegetables

4.6 Physical and Transport Properties

MRI data can be used to investigate the dynamics of a food process or to calculate physical properties. A variety of food processing operations have been investigated using MRI, including freezing of potatoes,[37] rehydration of pasta,[38] extrusion processing[39] and baking of cookies[40] (Table 2). A comprehensive review and discussion on the use of MRI applied to food processing can be found in books by Hills[41] and McCarthy.[31]

The characterization of physical and transport properties using MRI differs in a fundamental manner from traditional analysis techniques. Most traditional methods employed to measure physical or transport properties rely on an integral approach. For example, classical characterizations of moisture transport in food systems have employed integral techniques such as sorption–desorption or gravimetric methods. MRI provides measurements of component saturations and material properties on a spatially resolved basis. Therefore, instead of an average moisture content for an entire sample, MRI provides the moisture content as a function of spatial coordinates within the object. This permits the calculation of the transport coefficients and the spatial variation of the transport coefficients and also the simultaneous determination of material properties such as membrane permeability.[42]

MRI has provided novel insights into the interaction between products and process equipment. In rotating geometries modeling a couette cell, a scraped surface heat exchanger and a single screw extruder, the velocity and concentration profiles of solid suspensions were measured using MRI.[39] Particle migration was observed to be a function of the geometry of the equipment and the particle migration was sufficient enough to significantly change the local viscosity of the suspension. Corbett et al.[39] concluded that particle migration was best controlled by changes in the geometry of the processing equipment and not by changes in the operating conditions.

4.7 Functional Imaging

Functional magnetic resonance imaging (fMRI) is an experimental technique used to relate sensory perception and/or cognitive activity of the brain to specific locations in the brain.[43] Brain functions associated with visual, auditory and motional activities are actively being investigated using fMRI. These studies have been successful in isolating regions of the human brain that are activated by certain visual and auditory stimuli or tasks.[44] Recently, studies of human taste perception have been initiated.[45] Locations in the brain activated by six different taste stimuli were identified surrounding and buried in the Sylvian fissure: the upper portion of the insula, the frontal operenlum and the pre- and postcentral gyri.

Locations of brain activity are mapped by measuring the associated signal intensity changes in a subject's MRI images. The MRI images are taken very rapidly, of the order of 2–3 s for the entire brain. One of the rapid MRI techniques such as echo planar imaging[41] or spiral imaging is applied. The change in signal intensity between images, several seconds apart, is a result of a rapid momentary increase in blood flow. For example, when a subject moves the index finger there is a rapid momentary increase in the cerebral blood flow in the specific region of the brain controlling the movement of the finger. The increase in blood flow results in a stronger MRI signal intensity at the specific location of the increase in blood flow. The tissue has a net increase in intravascular oxyhemoglobin and a decrease in deoxyhemoglobin. Deoxyhemoglobin is paramagnetic and hence tissues with a greater deoxyhemoglobin content have a faster decaying MRI signal. The change in signal intensity is between 1 and 5% of the total signal on a 1.5-T medical MRI scanner. Statistical techniques are employed to analyze the changes in signal intensity in fMRI data.

fMRI studies will prove valuable in improving the understanding of the sensory perception of foods. This approach should improve the development of new foods and help in the understanding of sensory methods for the analysis of foods. fMRI has recently been extended to include monitoring and mapping the dynamics of metabolism.[46] Understanding of metabolism and the influence of specific foods on metabolism should play an important role in the understanding of many foods, including neutraceuticals. The development of instrumental methods to replace human sensory studies should benefit from the knowledge gained in these investigations.

5 PLANT MATERIALS

5.1 Introduction

NMR investigation of the vast area of plants can largely be divided into those studies using NMR spectroscopy and those utilizing NMR imaging, MRI. Hence we shall commence with an overview of both MRI and spectroscopy applied to plants, and then consider what information each method provides. There is also a short summary of the analysis of plant natural product extracts. While the literature on plant NMR is not as large as that on medical MRI or liquid spectroscopy of biomolecules, it is nonetheless extensive. We make no attempts to review the field (excellent reviews will cited), but in keeping with the focus of this work instead will give an overview of the field with representative examples to indicate to the interested experimentalist the information

to be gained by such a study, its advantages and its drawbacks.

A rough distinction between a plant and a plant-derived food (treated in the previous section) is that the former is obviously a living organism, and much of the interest in NMR of plants centers around the noninvasive nature of the technique for studying living plants. With NMR, repeated studies of (for example) a developing fruit or seedling are possible over an extensive course of time. NMR also offers the possibility of defining both molecular metabolic processes via in vivo spectroscopy and ''supramolecular'' structure or morphology via MRI. To date, most NMR studies have utilized either MRI or spectroscopy, although a few ''image-guided'' localized spectroscopy studies akin to work in the biomedical field have been reported (see below). The distinction between molecular (spectroscopy) and morphology (MRI) is somewhat blurred since MRI can be used to follow ''molecular'' processes such as diffusion and nuclear relaxation, which are critically dependent on molecular motion. However, in general, spectroscopy of plants is applied to the observation of metabolites and ions, determination of intracellular pH and the possible variation of these over time under differing experimental conditions. MRI of plants in general concerns itself with morphology, as well as water transport, diffusion and relaxation, including possible variations of these over time under differing conditions.

5.2 Magnetic Resonance Imaging of Plants

5.2.1 Sample Preparation

In practice, MRI studies of plants are almost invariably concerned with a single component of the plant. Sample preparation varies according to the material studied, with fruits, stems, seeds, leaves and nuts generally being easily investigated by simply placing the sample in the magnet. Light and cooling/warming air can easily be introduced into the magnet; generally, complex perfusion systems are the province of plant NMR spectroscopy (see below). With roots in situ both the roots and soil must be placed in the magnet.

5.2.2 Magnetic Resonance Imaging of Plants. What is Observed?

The first point to consider in an examination of (for example) a fruit by MRI is that what is seen is critically dependent on the instrumental conditions, both the hardware used (field strength, shielded vs unshielded gradients, etc.) and the parameters used in the imaging pulse sequence utilized. This point is illustrated in Figure 5(a–d), which shows a single slice through a potato using different imaging sequences.[47] MRI of plants has almost exclusively been devoted to studies using ^{1}H, hence the proton images in Figure 5(a–d) constitute a map of mobile water protons (the dominant source of proton signal) weighted by their nuclear relaxation times and diffusivity. Use of gradient-echo sequences, choice of T_2 and/or T_1 weighting and use of diffusion weighting are all examples of the choices available to the experimentalist to emphasize different structures within the plant and alter image contrast. The oft-cited simplistic example of an MRI image as a ''picture'' taken of an object must be viewed with caution in the light of these factors. Nonetheless, the studies cited below indicate that first and foremost MRI can give excellent images of the morphology of a plant noninvasively, i.e. without the excising, fixing and staining necessary for optical images. The resolution with a conventional gradient set strength of 1–10 G cm^{-1} is roughly about 100 μm for a 1-mm thick slice for an acquisition time of a few minutes on a small object (e.g. plant stem, grape), or about 200 μm for a 2-mm slice of a larger object (e.g. potato, apple). With microimaging gradient sets (gradient strength 50–100 G cm^{-1}) the resolution can reach <10 μm in-plane with a 250-μm slice thickness. The small number of nuclei generating observable signals in such a small volume leads to increased acquisition times, of the order of 1 h. Rapid acquisition methods such as echo-planar imaging (EPI), gradient-recalled echo in the steady state (GRESS) imaging and the related fast low-angle shot (FLASH) techniques can dramatically reduce total imaging times by increasing the repetition rate of the pulse sequence, but are limited in their utility in many plants by large susceptibility differences between cells and intracellular air spaces,[47] a key distinction for MRI of plants compared with other biological systems. These widespread local susceptibility differences cause a spread in Larmor frequencies leading to loss of resolution and image intensity, especially for sequences (GRESS, FLASH, EPI) that do not utilize spin-echoes. Susceptibility effects increase with B_0 and reduce the sensitivity advantage gained by going to higher fields.[48]

In addition to the inherent noninvasive nature of MRI, an additional advantage is that a single image slice can be taken in any direction or (at the cost of increased acquisition time) an entire three-dimensional (3-D) image of an object can be obtained and any desired slice (including user-defined oblique) can be reconstructed from the data analysis. Other advantages of the NMR method are the possibility of choosing chemical shift imaging (CSI) to select sugar or oil protons rather than water, thus mapping the location of those components in the plant. Finally, diffusion-weighted sequences allow the probing of molecular self-diffusion or flow in the plant. Examples of all these studies follow.

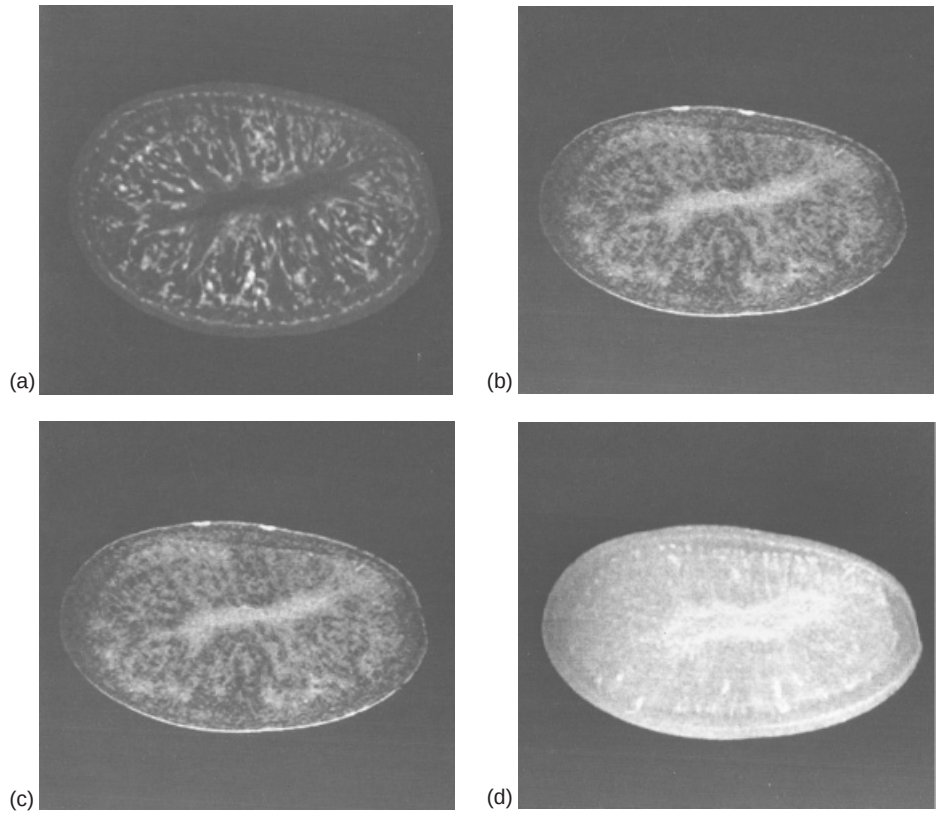

Figure 5 Spin-echo images from a potato showing the variation of contrast and information available from simple variation of pulse sequence parameters. All slices are of the same potato and same position. (a) Image with short recycle time and spin-echo time. Regions of shorter T_1 are brighter. (b, c) Images with long recycle time and two different values of spin-echo time. Regions with short T_2 appear darker. (d) Image with long recycle time and short spin-echo time. Image brightness here is much less dependent on relaxation times and thus reflects water distribution in a straightforward fashion. (Reproduced from MacFall and As[47] by permission of the American Society of Plant Physiologists.)

5.2.3 Disadvantages of Magnetic Resonance Imaging

Having indicated the utility of plant MRI, what are the disadvantages? First, the resolution generally does not compare with optical or electron microscopy.[49] The cost of an MRI scanner dwarfs that of an optical microscope, although steps to reduce system cost while retaining utility for plant studies are under way.[50] Studies on development of growing plants are not optimally conducted on the prevalent horizontal bore imagers owing to the sensitivity of most plants to the direction of gravity.[51] Vertical bore spectrometers are not optimally configured for plant study, being of narrower bore and usually having difficult access through the bottom. Finally, to date, plant imaging has a far from comprehensive database to guide new studies, and questions have arisen of the biological relevance of the NMR data.[49,52]

5.2.4 Fruits and Vegetables

We have divided MRI of plants into subsections of fruits and vegetables, stems and roots, along with consideration of single-cell imaging. MRI studies of fruits and vegetables have been relatively extensive. Primary areas of study include quality control, developmental studies and visualization of disease and infestation.

5.2.4.1 Quality Control; Maturation In this area are studies to examine noninvasively internal structure to determine fruit quality in response to storage, ripening, freezing, bruising and disease.[50] An early study used MRI to detect core breakdown in pear.[53] Chen et al.[54] used MRI to examine bruising defects in apple, peach, pear and onion, and also the presence of voids and seeds. The same report showed that MRI could be used to follow ripening in pineapple, tomato and avocado. Core breakdown in pears after storage and then ripening could be determined by spin-echo MRI.[53] Bruising leading to large changes in T_2 could be detected in strawberries.[55,56] Metabolic or physiological disorders detected by MRI in stored fruit include woolly breakdown in nectarines[57] and dry sacks in oranges.[54] Browning disorder has recently been detected using MRI by Clark and Burmeister.[58]

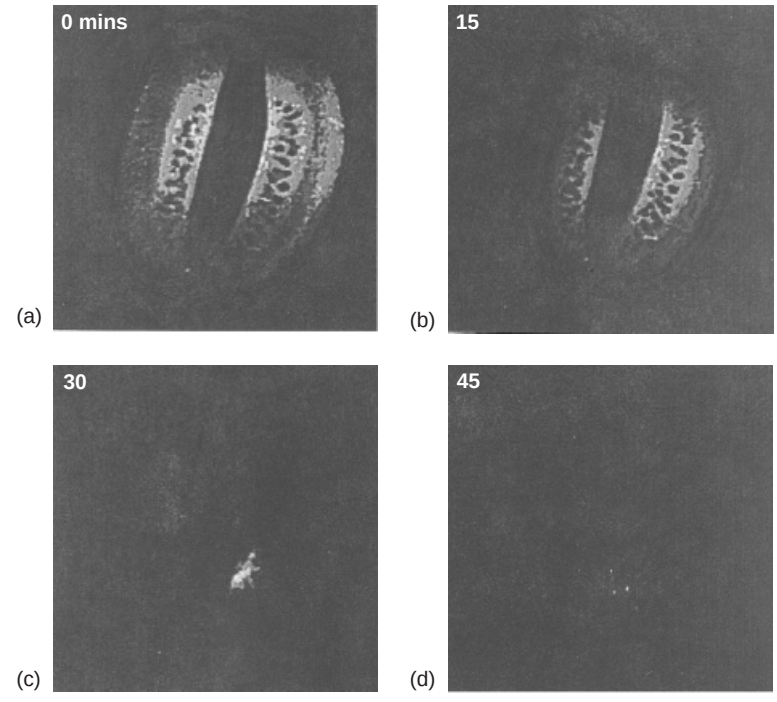

Figure 6 Progression of freezing in a median longitudinal slice of kiwifruit monitored by spin-echo imaging (resolution 390 μm, recycle time = 1000 ms, echo time = 25 ms) at the indicated time points. Freezing was induced by suspending fruit in flowing air at −40 °C in a chamber prior to placing samples in the magnet. (Reproduced from W.L. Kerr, C.J. Clark, M.J. McCarthy, J. de Ropp, 'Freezing Effects in Fruit Tissue of Kiwifruit Observed by Magnetic Resonance Imaging', *Sci. Hort.*, **69**, 169–179 (1997), Copyright 1997, with permission from Elsevier Science.)

Fruit ripening has been successfully followed in tomatoes[59–62] and in peaches, nectarines and plums.[63] Strawberries of varying maturity show clean differentiation in inversion–recovery spin-echo magnetic resonance images.[64] Some investigators have noted T_1 and T_2 variations of water with ripening but the results did not show a clear trend among all species. Freezing can cause severe damage to fruits. Freezing usually releases bound water from damaged cells, resulting in increased T_2 and brighter image intensity in affected areas. Freezing and chill injury changes have been noted for kiwifruit (Figure 6a–d),[37] blueberries[65] and zucchini.[66,67]

5.2.4.2 Disease and Infestation An obvious use of MRI would be in the detection of disease and (insect) infestation in fruit, and indeed many such studies have been conducted. Prominent among these are studies of the progression of fungal infections in fruit. A variety of fruits and infectious fungi have been studied. Raspberry and blackcurrant infection by *Botrytis cinerea* was elucidated by Williamson et al.[68] using GRESS methodology. The spread of infection in raspberries produced a clean boundary in the image of the fruit. Strawberries infected with different pathogens produced different images, suggesting that differentiation of infectious agents may be possible by MRI.[69] Similar results were obtained for two different pathogens of potatoes.[70] Mazucco et al.[50,71] have shown images clearly identifying fruit fly larvae in mangoes.

5.2.5 Stems

The cylindrical symmetry of stems makes them ideal candidates for transverse imaging over a relatively thick slice by MRI (Figure 7a and b). MRI studies of plant items have been used for a variety of purposes; one of the most notable is the investigation of water flow within the stem. The rate and start of rehydration after water stress in *Blechmum unilaterale* stems have been examined;[72] both proton density and T_1 values altered with dehydration/rehydration. Johnson et al.[73] observed changes in water content and T_1 with time during transpiration. Water movement in the pulvinus of *Mimosa* plants was visualized by Tamiva et al.[74] by sequential images. Information on water transport can be obtained from data other than simply proton density. For example, several studies have utilized the variance in T_1 relaxation times to elucidate paths of transport in tissue.[73,75,76] A sophisticated tool for measuring water diffusion in plants is available in the form of diffusion-weighted imaging. This has been applied to water and xylem flow in the plant

Figure 7 Spin-echo images of rose stems of the ''Cara Mia'' cultivar. (a) Close to the bottom of the stem; (b) near the flower. Each image is a 2-mm slice with 40-μm in-plane resolution. Acquisition times were 10 min and the spin-echo time was 18 ms.

stem.[77–79] Drought-stressed plants were compared with well-hydrated plants and the distribution of water was imaged.[80]

5.2.6 *Roots*

MRI provides, under controlled conditions, a unique tool for studying roots in soil. Several early studies showed the feasibility of detecting roots in soil.[81–83] Rogers and Bottomley[83] characterized natural and artificial soils with respect to their suitability for NMR studies. Loamy sand soils are among the best for MRI, yielding a minimal background signal even with a high water content and providing an excellent medium for root MRI.[83,84] Care must be taken when using either natural or artificial soils that the concentration of paramagnetic species be kept within acceptable limits. Water uptake in roots from surrounding soil has been measured in Loblolly pine seedlings.[85,86] Root growth has been followed in the Loblolly pine over several months using 3-D imaging

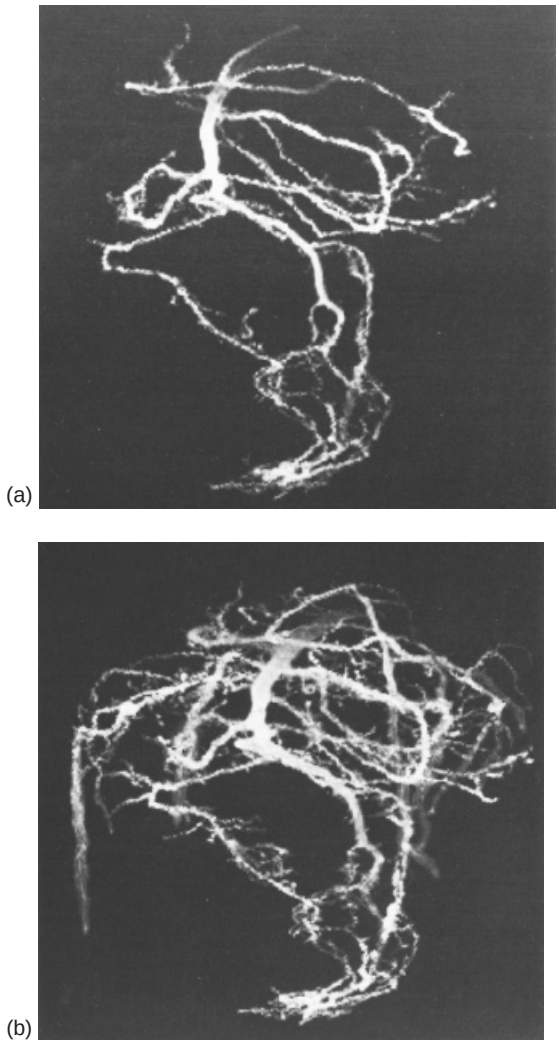

Figure 8 Growth of a root system in soil revealed by MRI. (a) Pine seedling roots at 3 months of age; (b) root system of the same plant at 6 months of age. (Reproduced from MacFall and As[47] by permission of the American Society of Plant Physiologists.)

techniques[47] (Figure 8a and b). In this way, root length, volume and growth pattern could be monitored noninvasively under physiologically relevant conditions. Changes of T_1 of water in regions of nodulated soyabean roots were studied by MacFall et al.[87] under differing perfusion (O_2, air, N_2) conditions. Matyac et al.[88] were able to image over several weeks the growth of nematode-induced galls on tomato roots in soil. Overall MRI has clearly established itself as a technique to visualize roots in situ.

5.2.7 *Single-cell Imaging*

The advances in gradient and probe technology have permitted MRI resolution and sensitivity to reach the microscopic scale (resolution $< 40\,\mu m$), which leads to

the possibility of using MRI to visualize single cells.[89] Since 1986, several studies of (relatively large) single cells have appeared;[90,91] a few of these have reported the resolution of single plant cells. Harrison et al.[92] and McFarland and Mortara[93] both reported images of large single-cell algae. Imaging of higher plant cells has been accomplished for onions,[94,95] although the resolution did not permit the detection of cell organelles. Further advances in methodology are needed for the wider use of MRI for microscopy of plant cells.

5.3 Nuclear Magnetic Resonance Spectroscopy of Plants

NMR spectroscopy of plants provides information on low molecular weight metabolites and ions in plant cells. The spectra of plants resemble liquid-phase spectra and are generally acquired on liquid-phase (''high-resolution'') spectrometers; solid-phase NMR of plant cell wells is not covered here. NMR of purified plant extracts is treated separately in section 5.4. NMR spectroscopy of plants has been reviewed by leaders in the field;[49,52,96,97] treated here are its salient features.

In contrast to the preponderance of proton data for the MRI of plants, spectroscopy has been effectively applied to ^{1}H, ^{13}C, ^{15}N, ^{23}Na and ^{31}P nuclei; problems with overlap of complex spectra over the narrow 0–10 ppm range of proton chemical shifts have helped to accentuate the importance of heteronuclear studies for plant NMR. For metabolic studies ^{31}P and ^{13}C are the most useful nuclei (Figures 9 and 10a–c); the low γ and 1.1% natural abundance of the latter nucleus lead to sensitivity problems that can be alleviated by the incorporation of ^{13}C labels, which in turn provide a fruitful avenue for tracing metabolic pathways. The ^{15}N nucleus has even lower sensitivity than ^{13}C and is likewise isotopically rare; enrichment for use in studying nitrogen metabolism has proven attractive. Some studies have been performed probing ion concentrations and flux using ^{23}Na- and ^{39}K-NMR. In contrast to MRI, where studies have been spread over a range of plant structures, spectroscopic studies have favored perfused seedlings and root tips, along with cell cultures, which can be inserted in relatively high density into a liquid NMR probe of 5–20 mm diameter with an appropriate perfusion system. An effective perfusion system is critical to NMR spectroscopy of plant systems,[96] and in addition to providing nutrients, oxygen and ions, the perfusion medium should have a minimal concentration of whatever nucleus is under study (to minimize the background signal) and a minimal concentration of paramagnetic species (to limit the broadening of lines).

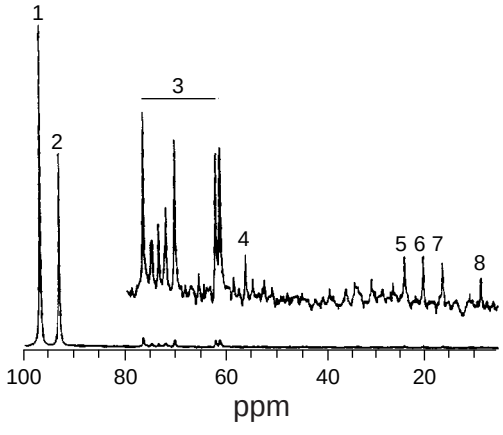

Figure 9 In vivo ^{13}C-NMR spectrum of maize seedling root tips, perfused with nitrogen-saturated 50 mM [1-^{13}C]glucose–0.1 mM $CaSO_4$. The dataset was acquired in 1 h. Roots were prelabeled with oxygenated 50 mM [1-^{13}C]glucose for 8 h prior to data acquisition. Peak assignments: 1, βGlc-C1; 2, αGlc-C1; 3, other sugar peaks; 4, 2-aminobutyrate-C2 (external reference); 5, 2-aminobutyrate-C3; 6, lactate-C3; 7, alanine-C3; 8, 2-aminobutyrate-C4. (Reproduced from Roberts and Xia[96] by permission of Academic Press.)

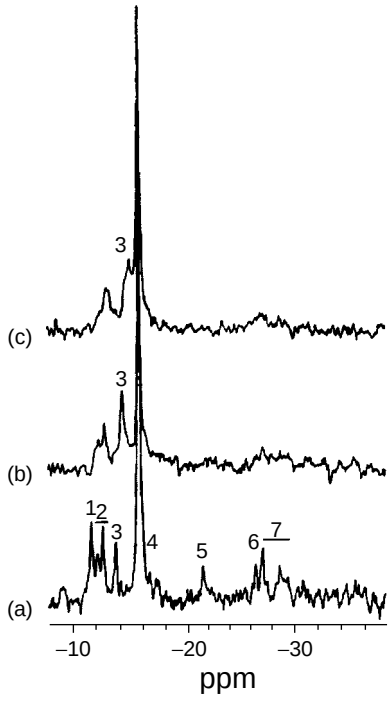

Figure 10 In vivo ^{31}P-NMR spectra of maize seedling root tips under (a) oxygenated conditions or (b, c) anoxia. Peak assignments: 1, Glc-6-P; 2, other monophosphate esters; 3, cytoplasmic P_i; 4, vacuolar P_i; 5, γ-nucleoside triphosphates; 6, α-nucleoside triphosphates; 7, uridinediphosphoglucose and nicotinamide adenine nucleotides. Note the increase in the width of peak 3 from spectrum (a) through spectrum (c), indicating increasing heterogeneity of cytoplasmic pH. Spectrum (c) was acquired over the same 1-h time period as the ^{13}C-NMR spectrum in Figure 9. (Reproduced from Roberts and Xia[96] by permission of Academic Press.)

5.3.1 pH Measurement

The capability of NMR to measure intracellular pH via the pH-dependent chemical shift of low molecular weight metabolites was both an early and yet a continuingly useful application of in vivo plant NMR. The ^{31}P shift of inorganic phosphate has proved particularly useful owing to its ubiquity and easy detection in all plant tissues.[97–99] Also used have been the pH-dependent chemical shifts of organic acids such as malate and citrate, which are especially useful for characterizing vacuolar pH.[100]

5.3.2 Metabolic Studies

The use of ^{13}C labels provides an excellent, non-invasive way of tracing metabolic pathways in plants in vivo. Chang and Roberts[101–103] used ^{13}C-NMR to study inorganic CO_2 fixation in maize roots. They also characterized carbonic anhydrase activity by saturation transfer ^{13}C-NMR. Carbon metabolism, specifically in the form of malate synthesis and degradation, has been followed in maize root tips[101–103] and cultured sycamore cells.[104] The results show that metabolic pathways can be followed in vivo in the fashion of classical radioisotope studies with ^{14}C, but with the advantage of real-time noninvasive monitoring. Monitoring nitrogen metabolism via ^{15}N is another application of in vivo plant NMR spectroscopy. Examples include the study of the biosynthesis of secondary metabolites[105,106] and the study of the release of ammonium from glutamate in carbon-starved carrot cells[107,108] (Figure 11a–d). ^{15}N Labels have been used to characterize the assimilation and metabolism of ammonium and nitrate.[107,109] Uptake of sodium ion and its compartmentalization (intracellular vs extracellular) using shift reagents have been studied by ^{23}Na-NMR.[110–112] ^{31}P-NMR can be used to characterize cytoplasmic vs vacuolar P_i.[102] Inorganic phosphate metabolism has been a frequent topic of ^{31}P-NMR studies. Lee et al.[113] showed that plant cells maintain level P_i concentrations under a variety of nutritional conditions, and Lee and Ratcliffe[114] showed that 1 μM external P_i was sufficient to maintain constant P_i in maize seedlings. Measurement of rates of synthesis of ATP from P_i using ^{31}P saturation transfer in maize root tips has been carried out by Roberts et al.[115,116]

5.3.3 Physiological Response to Environmental Change

Response of plants to hypoxia (low oxygen conditions) has been an area of high interest. Response of metabolism in plant tissues to hypoxia has been successfully characterized by both ^{13}C- and ^{31}P-NMR.[116–119] As has been pointed out by scientists and historians alike, irrigated agriculture carries the seeds of its own destruction in the deposition of salt in irrigated soil, and probing the response of plant tissues to salt stress is an active area of plant NMR. Roberts et al.[115] used ^{31}P-NMR to show enhanced P_i uptake in maize root tips under high-salt conditions. Spickett et al.[112,120] also used ^{31}P-NMR of maize root tips to characterize the effects of salt stress on intracellular pH, P_i and phosphorylcholine. Effects of extremes in pH, which can occur in soil or from influences such as acid rain, have been probed by both ^{13}C- and ^{31}P-NMR.[104,121] Gout et al.[104] showed that intracellular and vacuolar pH in sycamore cells are independent of external pH in the range 4.5–7.5, but outside this range become more aligned with the external pH value.

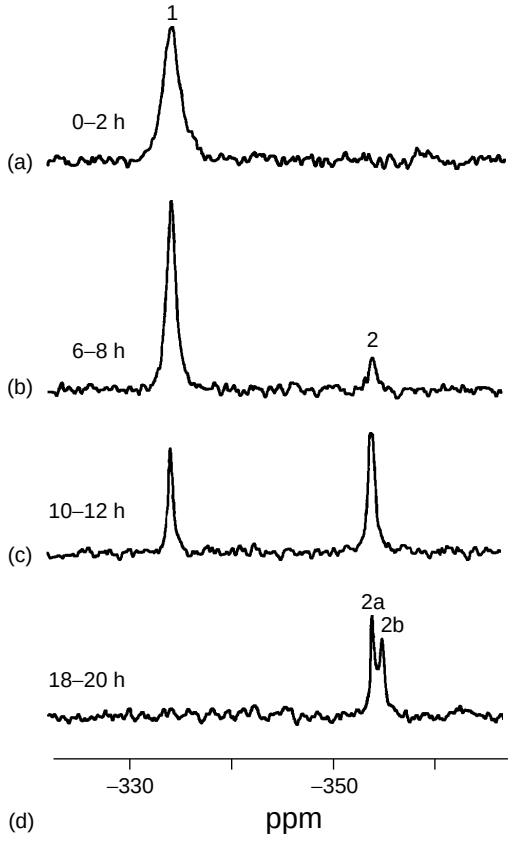

Figure 11 ^{1}H-Coupled ^{15}N-NMR spectra of 14-day-old carrot cells, preincubated for 3 h with 3 mM methionine sulfoximine and 3 mM transaminase inhibitor aminooxyacetate, following the addition of 20 mM [^{15}N]glutamate. Each spectrum was recorded over a 2-h interval. The peak assignments are as follows: 1, glutamate; 2, ammonium; 2a, cytoplasmic ammonium; 2b, external ammonium. (Reproduced from Fox et al.[108] by permission of NRC Research Press.)

5.3.4 Combined Spectroscopy and Imaging

The combination of imaging and localized spectroscopy is an intriguing area that has been the subject of much interest in the medical field (e.g. Gillies[122]). Imaging and spectroscopy can be thought of as interfacing in two

ways. The first is the use of CSI to generate images of the distribution of specific substances in plants based on their chemical shift differences. Generally in CSI a selective pulse excites the chemical shift of interest, followed by application of an imaging sequence to provide the spatial distribution of the species chosen by the selective pulse. CSI is aided by high fields, since isolation of particular metabolites in the image is dependent upon chemical shift differences. The second interface of imaging and spectroscopy is the use of the field gradients employed for imaging to achieve localized spectroscopy of a specific area; in its most elegant form, an image for morphology is obtained first, followed by spectra localized to a region of interest.

5.3.4.1 Chemical Shift Imaging of Plants Oil distribution is usually easy to image selectively owing to the prominent lipid methylene resonance, which is well separated from that of water. Examples of images of oil distribution include those for olives[123] and coconuts.[124] The spatial distribution of sucrose and other metabolites have been imaged in *Ricinus communis* hypocotyls using CSI at high spatial resolution.[125,126] Sugar distribution in grapes was imaged by Pope et al.[127] Goodman et al.[128] used CSI to follow the distribution of sugars in ripening grapes, observing that initial sugar gradients in the immature grape changed to an even distribution throughout the mesocarp at maturity. Tse et al.[129] have imaged the sugar distribution in pea.

5.3.4.2 Localized Spectroscopy of Plants Fewer studies of localized spectroscopy have been published, but interesting work has appeared. Jagannathan et al.[124] used image-guided localized proton NMR to observe changes in fluids in coconut during maturation (Figure 12a–c). ^{1}H Localized spectroscopy and ^{13}C-NMR have also been used to probe the distribution and identity of sugars in germinating barley seeds.[130]

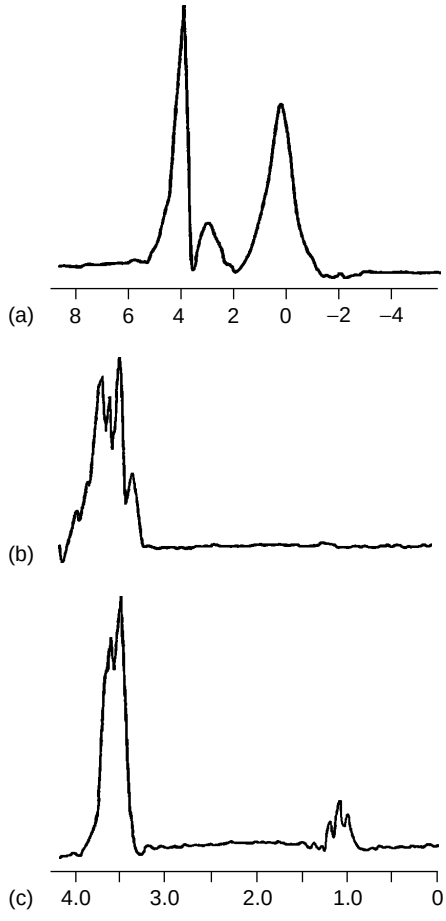

Figure 12 In vivo ^{1}H-NMR spectra at 1.5 T of 8 mL of voxel localized in the water portion of (b) water tender coconut and (c) mature coconut. (a) Spectrum from the pulp region of a mature coconut. (Reproduced from N.R. Jagannathan, V. Govindaraju, P. Raghunathan, 'In Vivo Magnetic Resonance Study of the Histochemistry of Coconut (*Cocos nucifera*)', *Magn. Reson. Imaging*, **13**, 885–892 (1995), Copyright 1995, with permission from Elsevier Science.)

5.4 Nuclear Magnetic Resonance Spectroscopy of Plant Extracts

In contrast to the more recent development of MRI and in vivo spectroscopy for the study of plants, high-resolution liquid-phase spectroscopy of natural product plant extracts dates back over several decades. The literature covering the determination of plant natural product structures by NMR is vast and no attempt will be made to cover it. Here only salient features of this field will be briefly treated.

The usual purpose of the study of isolated pure plant compounds by NMR is to establish scalar and dipolar connectivities leading to the identification of molecular structures. Recent developments in 2-D NMR, inverse detection, gradient methodology and microprobes have all contributed to a major revolution in the power of NMR methods applied to plant extracts and the ease with which total structures are elucidated. We shall look briefly at each of these aspects.

5.4.1 Two-dimensional Nuclear Magnetic Resonance Methods

Early studies of plant extracts, like all NMR, were limited to 1-D studies, principally of ^{1}H and ^{13}C. The development of COSY[17,131] and then TOCSY[132] 2-D scalar correlation methods enabled vicinal ($^{3}J_{HH}$) and geminal proton ($^{2}J_{HH}$) couplings to be elucidated throughout a molecular backbone (Figure 13a–c). Other 2-D pulse sequences also proved to be powerful tools for determination of molecular structure (see below).

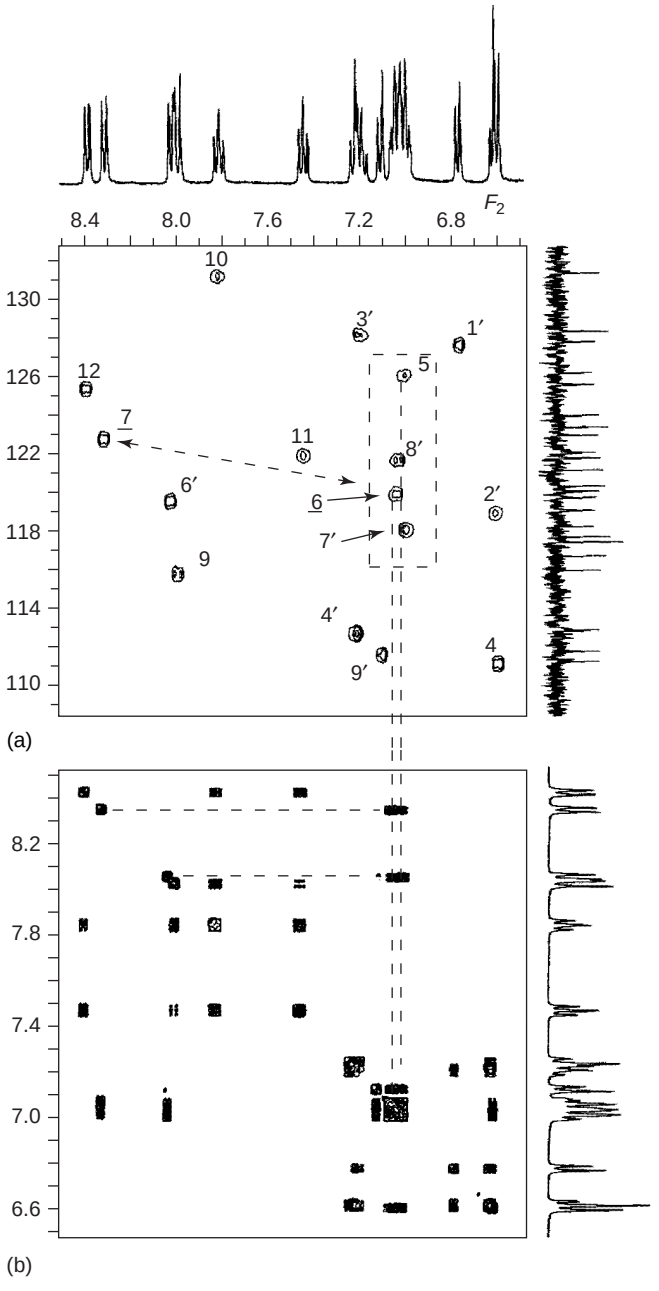

Figure 13 Segments of the COSY (b) and HMQC (a) spectra of cryptospirolepine **(1)** acquired at an ^{1}H frequency of 400 MHz. Trace (b) shows $^3J_{HH}$ scalar correlations and trace (a) shows single-bond carbon–proton correlations. Peak labels correspond to the numbering of structure **(1)**. (Reproduced from G.E. Martin, R.C. Crouch, in *Two-dimensional NMR Spectroscopy*, eds. W.R. Croasmun, R.M.K. Carlson, VCH, New York, 873–914, 1994, © 1994 VCH Publishers, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

5.4.2 Inverse Detection

It is well known that the chemical shift dispersion of ^{13}C (200 ppm, about 20-fold that of ^{1}H) provides an attractive route to well-resolved spectra, but at the cost of insensitive detection of the rare ^{13}C spin. The development of inverse-detected 2-D methods,[20,21,133] where correlations are established between X nuclei and ^{1}H *with ^{1}H detection*, permitted assignments of both ^{13}C and ^{1}H spectra and elucidation of ^{13}C–^{1}H *J*-coupled partners,[134] without the need for isolating the large amounts of material routinely required for traditional ^{13}C studies. The one-bond ^{13}C–^{1}H correlations are obtained via HSQC[135] or HMQC[19,20,136] type experiments as shown for crytospirolepine **(1)** in Figure 13(a) and (b). Frequently in natural products, saturated carbons, carbonyls or heteroatoms disrupt proton connectivity networks and necessitate the use of NMR techniques that can ''navigate through'' these structural features. The HMBC (heteronuclear *multiple* bond correlation with ^{1}H detection)[137] experiment and more recent variants[138,139] allow the determination of $^2J_{CH}$ correlations (C–C–H) and also $^3J_{CH}$ and $^4J_{CH}$ correlations, and are invaluable techniques for assignment and structure determination in ''proton-poor'' natural products (Figure 14).

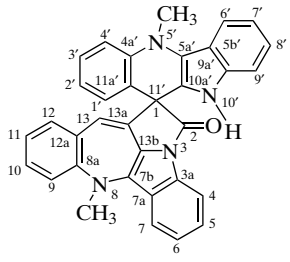

(1)

5.4.3 Gradient Methodology

Adding to the power of the inverse detected methods, and also other NMR techniques, has been the development of pulsed-field gradients for liquid spectroscopy.[140,141] The gradients serve two primary purposes: coherence selection and solvent suppression. Gradient selection of coherence replaces, partially or entirely, methods based on phase cycling and enables some 2-D data to be collected with only a single scan per t_1 increment when sufficient material (with microprobes < 1 mg is sufficient) is available as no minimum phase cycle is required. When sample limitations require signal averaging, gradient methods are still preferable, especially in inverse-detected experiments, for the production of artifact-free spectra. Normally plant natural product spectra are acquired in deuterated solvents but with small sample amounts even the residual proton of a deuterated solvent can be troublesome, and gradient methods provide improved solvent suppression over traditional presaturation methods.

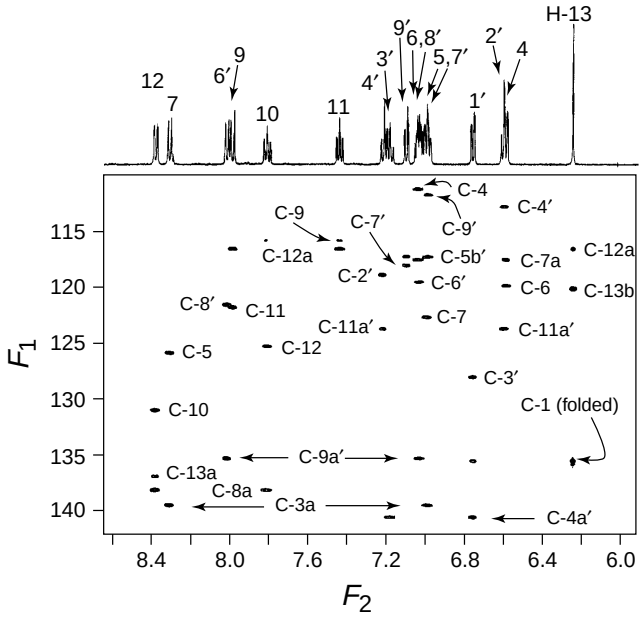

Figure 14 HMBC spectrum acquired in 16 h with a 250-μg sample of structure **(1)** in a microinverse detection probe at 11.75 T. Peak labels correspond to the numbering of structure **(1)**. Cross peaks represent mostly $^{2}J_{CH}$ correlations along with some $^{3}J_{CH}$ correlations. (Reproduced from G.E. Martin, R.C. Crouch, in *Two-dimensional NMR Spectroscopy*, eds. W.R. Croasmun, R.M.K. Carlson, VCH, New York, 873–914, 1994, © 1994 VCH Publishers, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

5.4.4 Microprobes

Commercial manufacturers over the last few years have developed probes with 2.5–3-mm diameters with outstanding performance, typically achieving 40% more sensitivity than comparable 5-mm probes. These tubes use <150 μL of solution, and with "plugs" at either end of the sample constructed of materials matched to the magnetic susceptibility of the NMR solvent, volumes <80 μL can be utilized, with concurrent reduction of sample size to 10–100 μg. It has therefore become possible to determine total structures of natural products using amounts of material an order of magnitude less than possible just a decade ago.[142] Recently, probe diameters and sample volumes have been further reduced with the advent of 1.7-mm "submicro" probes that utilize just 30 μL of solution.[143–145] These probes produce an impressive 2.3-fold improvement in sensitivity over the 3-mm probe[146] and further lower the minimum amount of sample needed for structural studies.

5.4.5 Total Structure Determination: the Strategy

The NMR suite of techniques utilized for structure determination will necessarily vary with the material studied, with a greater number of, and more complex, techniques needed for molecules that do not yield to the basic experiments.[15,134] Generally a COSY or TOCSY experiment provides $^{3}J_{HH}$ correlations. HSQC and HMQC yield $^{1}J_{CH}$ correlations and HMBC the $^{2}J_{CH}$ and $^{3}J_{CH}$ correlations. In many cases this is sufficient to determine the complete structure. Where resolution is poor HMQC–TOCSY[147] or HSQC–TOCSY[148] 2-D spectra provide proton correlation maps "edited" by the chemical shift dispersion of ^{13}C. For nitrogen-containing compounds a $^{15}N-^{1}H$ HMQC can be used to locate and assign the chemical shifts of nitrogens. The utility of low-level, long-range $^{1}H-^{15}N$ correlations has also been reported recently.[149–151] Additional, more elaborate techniques can be utilized, but those listed here are frequently sufficient to elucidate total structure, i.e. the relative configuration of all chemically bonded atoms in the molecule. Tertiary structure, used in the sense of protein NMR, is less of a concern in natural products. Where spatial (dipolar), as opposed to bonding (scalar), interactions need to be determined, then nuclear Overhauser spectroscopy (NOESY) experiments or the related rotating frame nuclear Overhauser enhancement spectroscopy (ROESY) can be performed (Figure 15). The ROESY experiment is the one of choice in many cases since for mid-sized natural products at high field the nuclear Overhauser effect (NOE) may be negligible,

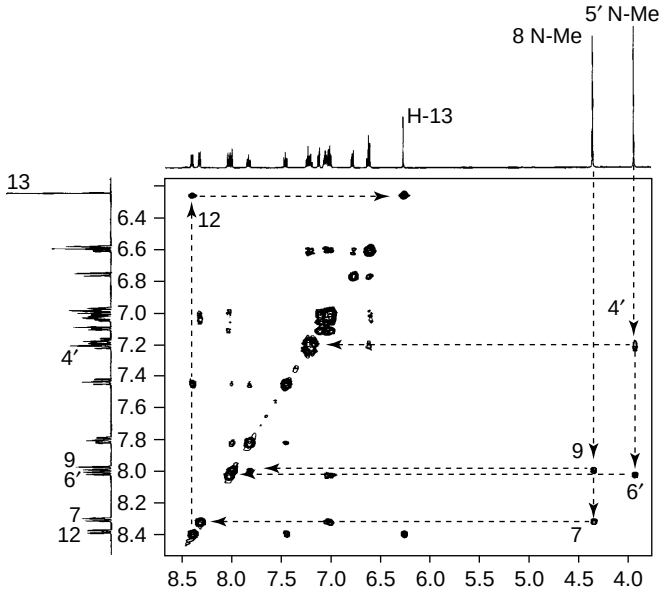

Figure 15 ROESY spectrum of structure **(1)**. Cross peaks represent close dipolar contacts between protons. The cross relaxation between the *N*-Me groups and the various ring systems is shown on the figure. (Reproduced from G.E. Martin, R.C. Crouch, in *Two-dimensional NMR Spectroscopy*, eds. W.R. Croasmun, R.M.K. Carlson, VCH, New York, 873–914, 1994 © 1994 VCH Publishers, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

while the rotating frame nuclear Overhauser effect (ROE) suffers no such limitation.[152]

5.4.6 On the Horizon

The impact of 2-D methods, inverse detection, field gradients and high-sensitivity microprobes and submicroprobes for natural product analysis has been immense. Further technical advances in liquid NMR are continually appearing. The recent use of 1.7-mm diameter probes has allowed a further reduction of the sample amounts needed for analysis (see section 5.4.4) and more routine use of such probes is likely to occur in the near future. Combination LC/NMR spectrometers[153] allow the separation of natural products and ''on-line'' NMR analysis to be conducted simultaneously and may well prove to have a great impact on plant natural product characterization. Superconducting probes offer greatly enhanced sensitivity compared with standard room-temperature probes. The use of superconducting probe technology is only just appearing in the literature[154] and the high cost of these devices currently is a significant barrier to their widespread use, but the anticipated pay-off in sensitivity, thereby further lowering the minimum sample amounts required, will be a powerful driving force for their utilization. Finally, continuing the trend towards smaller probes with higher sensitivity, ''microcoils'' for use in liquid NMR spectroscopy have recently been fabricated.[155–157] These coils have diameters considerably less than 1 mm for sample volumes far less than 1 μL. Although applications to date are limited, this technology has the potential yet again to reduce significantly the sample amount needed for analysis by virtue of its dramatically enhanced mass sensitivity.

ABBREVIATIONS AND ACRONYMS

AOCS	American Oil Chemists' Society
COSY	Correlation Spectroscopy
CSI	Chemical Shift Imaging
EPI	Echo-planar Imaging
FID	Free Induction Decay
FLASH	Fast Low-angle Shot
fMRI	Functional Magnetic Resonance Imaging
GRESS	Gradient-recalled Echo in the Steady State
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Coherence
HRMAS	High-resolution Magic-angle Spinning
HRNMR	High-resolution Nuclear Magnetic Resonance
HSQC	Heteronuclear Single Quantum Coherence
ISO	International Standards Organization
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LRNMR	Low-resolution Nuclear Magnetic Resonance
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Spectroscopy
PFGNMR	Pulsed-field-gradient Nuclear Magnetic Resonance
RF	Radiofrequency
ROE	Rotating Frame Nuclear Overhauser Effect
ROESY	Rotating Frame Nuclear Overhauser Enhancement Spectroscopy
S/N	Signal-to-noise Ratio
SNIF-NMR®	Site-specific Natural Isotope Fractionation Studied by Nuclear Magnetic Resonance
TOCSY	Total Correlation Spectroscopy
USDA	United States Department of Agriculture
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Imaging, Functional • Magnetic Resonance in Medicine, High Resolution Ex Vivo • Magnetic Resonance, General Medical • Multinuclear Magnetic Resonance Spectroscopic Imaging

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Nuclear Magnetic Resonance of Biomolecules

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Viscosity of Food: Measurement and Application

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Imaging of Polymers

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. F. Bloch, W.W. Hansen, M. Packard, 'The Nuclear Induction Experiment', *Phys. Rev.*, **70**, 474–485 (1946).
2. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
3. W.G. Proctor, F.C. Yu, 'The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound', *Phys. Rev.*, **77**, 717 (1950).
4. W.R. Croasmun, R.M.K. Carlson, *Two-dimensional NMR Spectroscopy*, 2nd edition, VCH, New York, 1994.
5. R. Freeman, *Spin Choreography: Basics Steps in High Resolution NMR*, Oxford University Press, New York, 1998.
6. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
7. P.C. Lauterbur, 'Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance', *Nature (London)*, **242**, 190–191 (1973).
8. P. Mansfield, P.K. Grannell, 'NMR ''diffraction'' in Solids?', *J. Phys. C: Solid State Phys.*, **6**, L422–L426 (1973).
9. D.G. Gadian, *Nuclear Magnetic Resonance and Its Applications to Living Systems*, Oxford University Press, New York, 1982.
10. P.J. Barker, H.J. Stronks, 'Application of the Low Resolution Pulsed NMR ''MINISPEC'' to Analytical Problems in the Food and Agriculture Industries', in *NMR Applications in Biopolymers*, eds. S.J.W. Finley, A.S. Serianni, Plenum Press, New York, 481–498, 1990.
11. P.N. Gambhir, 'Applications of Low-resolution Pulsed NMR to the Determination of Oil and Moisture in Oilseeds', *Trends Food Sci. Technol.*, **3**, 191–196 (1992).
12. M.C.M. Gribnau, 'Determination of Solid/Liquid Ratios of Fats and Oils by Low-resolution Pulsed NMR', *Trends Food Sci. Technol.*, **3**, 186–190 (1992).
13. *An Introduction to Analytical Applications of Low Resolution NMR*, Minispec Application Note, Bruker Analytische Messtechnik, Karlsruhe, Vol. 30, 1990.
14. J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, *Protein NMR Spectroscopy*, Academic Press, New York, 1996.
15. G.E. Martin, R.C. Crouch, 'Inverse-detected 2D-NMR Applications in Alkaloid Chemistry', *Mod. Methods Plant Anal.*, **15**, 25–89 (1994).
16. P.S. Belton, I. Delgadillo, A.M. Gil, 'High-field Solution State Proton NMR for Food Analysis', *Semin. Food Anal.*, **3**, 223–234 (1998).
17. W.P. Aue, E. Bartholdi, R.R. Ernst, 'Two-dimensional Spectroscopy. Application to Nuclear Magnetic Resonance', *J. Chem. Phys.*, **64**, 2229 (1976).
18. L. Braunschweiler, R.R. Ernst, 'Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy', *J. Magn. Reson.*, **53**, 521–528 (1983).
19. A. Bax, R.H. Griffey, B.L. Hawkins, 'Correlation of Proton and Nitrogen-15 Chemical Shifts by Multiple Quantum NMR', *J. Magn. Reson.*, **55**, 301–315 (1983).
20. L. Muller, 'Sensitivity Enhanced Detection of Weak Nuclei Using Heteronuclear Multiple Quantum Coherence', *J. Am. Chem. Soc.*, **101**, 4481–4484 (1979).
21. G. Bodenhausen, D.J. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185–189 (1980).
22. E. Humpfer, M. Spraul, 'The Use of High Resolution NMR in Food Analysis', *Semin. Food Anal.*, **3**, 287–302 (1998).
23. P.A. Keifer, 'New Methods for Obtaining High-resolution NMR Spectra of Solid-phase-synthesis Resins, Natural Products, and Solution-state Combinatorial Chemistry Libraries', *Drugs Future*, **23**, 301–317 (1998).
24. J.N. Shoolery, 'The Development of Experimental and Analytical High Resolution NMR', *Prog. Nucl. Magn. Reson. Spectrosc.*, **28**, 37–52 (1995).
25. H.J.C. Berendsen, 'Rationale for Using NMR to Study Water Relations in Foods and Biological Tissues', *Trends Food Sci. Technol.*, **3**, 202–205 (1992).

26. B.P. Hills, C.E. Manning, J. Godward, 'A Multistate Theory of Water Relations in Biopolymer Systems', in *Advances in Magnetic Resonance in Food Science*, eds. P.S. Belton, B.P. Hills, G.A. Webb, Royal Society of Chemistry, Cambridge, 45–62, 1999.
27. C. Guillou, G. Remaud, G.J. Martin, 'Applications of NMR to the Characterization and Authentication of Foods and Beverages', *Trends Food Sci. Technol.*, **3**, 197–201 (1992).
28. G.G. Martin, 'Application of SNIF-NMR® and Other Isotopic Methods for Testing the Authenticity of Food Products', in *Magnetic Resonance in Food Science*, eds. P.S. Belton, I. Delgadillo, A.M. Gil, G.A. Webb, Royal Society of Chemistry, London, 120–126, 1995.
29. G.J. Martin, 'Natural Abundance Deuterium NMR as a Tool for Studying Mechanistic and Environmental Effects', *Semin. Food Anal.*, **3**, 251–266 (1998).
30. www.eurofins.com.
31. M.J. McCarthy, *Magnetic Resonance Imaging in Foods*, Chapman & Hall, New York, 1994.
32. S.-M. Kim, M.J. McCarthy, P. Chen, 'Magnetic Resonance Image Analysis of the Structure and Texture of Swiss-type Cheeses', *J. Magn. Reson. Anal.*, **2**, 281–289 (1996).
33. K.J. Packer, C. Rees, 'Pulsed NMR Studies of Restricted Diffusion. I. Droplet Size Distributions in Emulsions', *J. Colloid Interface Sci.*, **40**, 206–218 (1972).
34. J.C. Van Den Enden, D. Waddington, H.V. Aalst, C.G.V. Kralingen, K.J. Packer, 'Rapid Determination of Water Droplet Size Distributions by PFG-NMR', *J. Colloid Interface Sci.*, **140**, 105–113 (1990).
35. I. Fourel, J.P. Guillement, D.L. Botlan, 'Determination of Water Droplet Size Distributions by Low Resolution NMR, I. Liquid Emulsions', *J. Colloid Interface Sci.*, **164**, 48–53 (1994).
36. I. Fourel, J.P. Guillement, D.L. Botlan, 'Determination of Water Droplet Size Distributions by Low Resolution NMR, II. Solid Emulsions', *J. Colloid Interface Sci.*, **169**, 119–124 (1995).
37. W.L. Kerr, C.J. Clark, M.J. McCarthy, J. de Ropp, 'Freezing Effects in Fruit Tissue of Kiwifruit Observed by Magnetic Resonance Imaging', *Sci. Hort.*, **69**, 169–179 (1997).
38. B.P. Hills, F. Babonneau, V.M. Quantin, P.S. Belton, 'Radial NMR Microimaging Studies of Rehydration of Pasta', *J. Food Eng.*, **27**, 71–86 (1996).
39. A.M. Corbett, R.J. Phillips, R.J. Kauten, K.L. McCarthy, 'Magnetic Resonance Imaging of Concentration and Velocity Profiles of Pure Fluids and Solid Suspensions in Rotating Geometries', *J. Rheol.*, **39**, 907–924 (1995).
40. S.-W. Hong, Z.-Y. Yan, M.S. Otterburn, M.J. McCarthy, 'Magnetic Resonance Imaging (MRI) of a Cookie in Comparison with Time-lapse Photographic Analysis (TLPA) During Baking Process', *Magn. Reson. Imaging*, **14**, 923–927 (1996).
41. B. Hills, *Magnetic Resonance Imaging in Food Science*, John Wiley & Sons, New York, 1998.
42. M.J. McCarthy, D. Lasseux, J.E. Maneval, 'NMR Imaging in the Study of Diffusion of Water in Foods', *J. Food Eng.*, **22**, 211–224 (1994).
43. J. Ellermann, M. Garwood, K. Hendrich, R. Hinke, X. Hu, S.-G. Kim, R. Menon, H. Merkle, S. Ogawa, K. Ugurbil, 'Functional Imaging of the Brain by Nuclear Magnetic Resonance', in *NMR in Physiology and Biomedicine*, ed. R.J. Gillies, Academic Press, San Diego, 137–150, 1994.
44. M.I. Posner, M.E. Raichle, 'The Neuroimaging of Human Brain Function', *Proc. Natl. Acad. Sci. USA*, **95**, 763–764 (1998).
45. A. Faurion, B. Cerf, D.L. Bihan, A.-M. Pillias, 'fMRI Study of Taste Cortical Areas in Humans', presented at International Symposium of Olfaction and Taste XII and Achems XIX, San Diego, CA, 1997.
46. T.L. Davis, K.K. Kwong, R.M. Weisskoff, B.R. Rosen, 'Calibrated Functional MRI: Mapping the Dynamics of Oxidative Metabolism', *Proc. Natl. Acad. Sci. USA*, **95**, 1834–1839 (1998).
47. J.S. MacFall, H.V. As, 'Magnetic Resonance Imaging of Plants', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Schachar-Hill, P.E. Pfeffer, American Society of Plant Physiologists, Rockville, MD, 33–76, 1996.
48. P.T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Clarendon Press, Oxford, 1991.
49. R.G. Ratcliffe, 'In Vivo NMR Studies of Higher Plants and Algae', *Adv. Bot. Res.*, **20**, 44–123 (1994).
50. C.J. Clark, P.D. Hockings, D.C. Joyce, R.A. Mazucco, 'Application of Magnetic Resonance Imaging to Pre- and Post-harvest Studies of Fruits and Vegetables', *Postharvest Biol. Technol.*, **11**, 1–21 (1997).
51. J.A. Chudek, G. Hunter, 'Magnetic Resonance Imaging of Plants', *Prog. Nucl. Magn. Reson. Spectrosc.*, **31**, 43–62 (1997).
52. R.G. Ratcliffe, 'In Vivo NMR Spectroscopy: Biochemical and Physiological Applications to Plants', in *Nuclear Magnetic Resonance in Plant Biology*, eds. Y. Shachar-Hill, P.E. Pfeffer, American Society of Plant Physiologists, Rockville, MD, 1–32, 1996.
53. C.Y. Wang, P.C. Wang, 'Nondestructive Detection of Core Breakdown in "Bartlett" Pears with Nuclear Magnetic Resonance Imaging', *HortScience*, **24**, 106–109 (1989).
54. P. Chen, M.J. McCarthy, R. Kauten, 'NMR for Internal Quality Evaluation of Fruits and Vegetables', *Trans. Am. Soc. Agric. Eng.*, **32**, 1747–1753 (1989).
55. J.L. Maas, M.J. Line, M.M. Millard, G.J. Galletta, 'Nuclear Magnetic Resonance Imaging: a Different View of Strawberry Fruit', *Adv. Strawberry Res.*, **11**, 64–66 (1992).
56. J.L. Maas, M.M. Millard, 'Nuclear Magnetic Resonance Microimaging of Strawberry Fruit', *Acta Hort.*, **348**, 375–377 (1993).

57. L. Sonego, R. Ben-Arie, J. Raynal, J.C. Pech, 'Biochemical and Physical Evaluation of Textural Characteristics of Nectarines Exhibiting Woolly Breakdown: NMR Imaging, X-ray Computed Tomography and Pectin Composition', *Postharvest Biol. Technol.*, **5**, 187–198 (1995).
58. C.J. Clark, D.M. Burmeister, 'Magnetic Resonance Imaging of Braeburn Browning Disorder Development During CA Storage', *HortScience*, **34**, 915–919 (1999).
59. M.E. Saltveit, 'Determining Tomato Fruit Maturity with Nondestructive In Vivo Nuclear Magnetic Resonance Imaging', *Postharvest Biol. Technol.*, **1**, 153–159 (1991).
60. J.C. Pech, A. Latche, M.H. Andrieu, J. Raynal, J. Pradere, 'In Vivo Monitoring of Fruit Ripening by Proton NMR Imaging', presented at XXIII International Horticultural Congress, Florence, 1990.
61. N. Ishida, T. Kobayashi, M. Koizumi, H. Kano, '^{1}H-NMR Imaging of Tomato Fruits', *Agric. Biol. Chem.*, **53**, 2363–2367 (1989).
62. N. Ishida, M. Koizumi, H. Kano, 'Ontogenetic Changes in Water in Cherry Tomato Fruits Measured by Nuclear Magnetic Resonance Imaging', *Sci. Hort.*, **57**, 335–346 (1994).
63. J.R. Heil, P. Chen, F.G. Mitchell, B. Zion, M.J. McCarthy, 'Use of Magnetic Resonance for Testing Peach, Nectarine, and Plum Maturity and Ripening Characteristics', presented at California Fruit Testing Agreement, Sacramento, 1991.
64. M. Faust, P.C. Wang, J. Maas, 'The Use of Magnetic Resonance Imaging in Plant Science', *Hort. Rev.*, **20**, 225–265 (1997).
65. G.R. Gamble, 'Non-invasive Determination of Freezing Effects in Blueberry Fruit Tissue by Magnetic Resonance Imaging', *J. Food Sci.*, **59**, 571–573 (1994).
66. S.L. Duce, T.A. Carpenter, L.D. Hall, 'Nuclear Magnetic Resonance Imaging of Fresh and Frozen Courgettes', *J. Food Eng.*, **16**, 165–172 (1992).
67. C.Y. Wang, P.C. Wang, 'Differences in Nuclear Magnetic Resonance Images between Chilled and Non-chilled Zucchini Squash', *Environ. Exp. Bot.*, **32**, 213–219 (1992).
68. B. Williamson, B.A. Goodman, J.A. Chudek, D.J. Johnston, 'Nuclear Magnetic Resonance (NMR) Microimaging of Soft Fruits Infected by *Botrytis cinerea*', in *Recent Advances in Botrytis Research*, eds. K. Verhoeff, N.E. Malathrakis, B. Williamson, Proceedings of X International *Botrytis* Symposium, Heraklion, Greece, 140–144, 1992.
69. J.L. Maas, M.J. Line, 'Nuclear Magnetic Resonance (NMR) Inversion Recovery Spin Echo Micro-imaging of Fungus Infections in Strawberry Fruit', *Acta Hort.*, **398**, 241–247 (1995).
70. A.J. Snijder, S.M. Glidewell, B.A. Goodman, 'NMR Microimaging of Healthy and Diseased Potato Tubers', presented at III International Conference on Magnetic Resonance Microscopy, Wurzburg, 1995.
71. R.A. Mazucco, D.C. Joyce, P.D. Hockings, 'Magnetic Resonance Imaging Applied to Harvested Mango Fruit', presented at Australasian Postharvest Conference, Gatton, Australia, 1993.
72. J.S. Veres, G.A. Johnson, P.J. Kramer, 'In Vivo Magnetic Resonance Imaging of *Blechnum fems*: Changes in T1 and N(1H) During Dehydration and Rehydration', *Am. J. Bot.*, **78**, 80–88 (1991).
73. G.A. Johnson, J. Brown, P.J. Kramer, 'Magnetic Resonance Microscopy of Changes in Water Content in Stems of Transpiring Plants', *Proc. Natl. Acad. Sci. USA*, **84**, 2752–2755 (1987).
74. T. Tamiva, T. Miyuazaki, H. Ishikawa, N. Iriguchi, T. Maki, J.J. Matsumoto, T. Tsuchiya, 'Movement of Water in Conjunction with Plant Movement Visualized by NMR Imaging', *J. Biochem.*, **104**, 5–8 (1988).
75. J.M. Brown, J.F. Thomas, C.P. Cofer, G.A. Johnson, 'Magnetic Resonance Microscopy of Stem Tissues of *Pelargonium hortorum*', *Bot. Gaz.*, **149**, 253–259 (1988).
76. J.S. MacFall, P. Spaine, R. Doudrick, G.A. Johnson, 'Alteration in Growth and Water-transport Process in Fusiform Rust Gall of Pine, Determined by Magnetic Resonance Microscopy', *Phytopathology*, **84**, 288–293 (1994).
77. J.E.A. Reinders, H.V. As, T.J. Schaafsma, P.A.d. Jager, D.W. Sherrif, 'Water Balance in Cucumis Plants, Measured by NMR', *J. Exp. Bot.*, **39**, 1199–1210 (1988).
78. Y. Xia, P.T. Callaghan, 'One-shot Velocity Microscopy–NMR Imaging of Motion Using a Single Phase-encoding Step', *Magn. Reson. Med.*, **23**, 138–153 (1992).
79. W. Kockenberger, J.M. Pope, Y. Xia, K.R. Jeffrey, E. Komor, P.T. Callaghan, 'Non-invasive Measurement of Phloem and Xylem Flow in Castor Bean Seedlings by Nuclear Magnetic Resonance Microimaging', *Planta*, **201**, 53–63 (1997).
80. U. Zimmermann, F.C. Meinzer, R. Benkert, J.J. Zhu, H. Schneider, G. Goldstein, E. Kuchenbrod, A. Haase, 'Xylem Water Transport: Is the Available Evidence Consistent with the Cohesion Theory?', *Plant Cell Environ.*, **17**, 1169–1181 (1994).
81. K. Omasa, M. Onoe, H. Yamada, 'NMR Imaging for Measuring Root System and Soil Water Content', *Environ. Control Biol.*, **23**, 99–102 (1985).
82. P.A. Bottomley, H.H. Rogers, T.H. Foster, 'NMR Imaging Shows Water Distribution and Transport in Plant Root Systems In Situ', *Proc. Natl. Acad. Sci. USA*, **83**, 87–89 (1986).
83. H.H. Rogers, P.A. Bottomley, 'In Situ Nuclear Magnetic Resonance Imaging of Roots: Influence of Soil Type, Ferromagnetic Particle Content, and Soil Water', *Agron. J.*, **79**, 957–965 (1987).
84. P.A. Bottomley, H.H. Rogers, S.A. Prior, 'NMR Imaging of Root Water Distribution in Intact *Vicia faba* L. Plants in Elevated Atmospheric CO_2', *Plant Cell Environ.*, **16**, 335–338 (1993).

85. J.S. MacFall, G.A. Johnson, P.J. Kramer, 'Observation of a Water-depletion Region Surrounding Loblolly Pine Roots by Magnetic Resonance Imaging', *Proc. Natl. Acad. Sci. USA*, **87**, 1203–1207 (1990).
86. J.S. MacFall, G.A. Johnson, P.J. Kramer, 'Comparative Water Uptake by Roots of Different Ages in Seedlings of Loblolly Pine (*Pinus taeda* L.)', *New Phytol.*, **119**, 551–560 (1991).
87. J.S. MacFall, P.E. Pfeffer, D.B. Rolin, J.R. MacFall, G.A. Johnson, 'Observation of the Oxygen Diffusion Barrier in Soyabean (*Glycine max*) Nodules with Magnetic Resonance Microscopy', *Plant Physiol.*, **100**, 1691–1697 (1992).
88. C.A. Matyac, G.P. Cofer, J.E. Bailey, G.A. Johnson, 'In Situ Observations of Rootgall Formation Using NMR Imaging', *J. Nematol.*, **21**, 131–134 (1989).
89. N.R. Aiken, E.W. Hsu, S.J. Blackband, 'A Review of NMR Microimaging Studies of Single Cells', *J. Magn. Reson. Anal.*, **1**, 41–48 (1995).
90. J.B. Aguayo, S.J. Blackband, J. Schoeniger, M. Mattingly, M. Hintermann, 'Nuclear Magnetic Resonance Imaging of a Single Cell', *Nature (London)*, **322**, 190–191 (1986).
91. J.S. Schoeniger, N. Aiken, E. Hsu, S.J. Blackband, 'Relaxation Time and Diffusion NMR Microscopy of Single Neurons', *J. Magn. Reson. B*, **103**, 261–273 (1994).
92. L.G. Harrison, S.D. Luck, B.D. Munasinghe, L.D. Hall, 'Magnetic Resonance Imaging Approaching Microscopic Scale: Maturation Stages of *Acetabularia mediterranea* Reproductive Caps', *J. Cell Sci.*, **91**, 379–388 (1988).
93. E.W. McFarland, A. Mortara, 'Three-dimensional NMR Microscopy: Improving SNR with Temperature and Microcoils', *Magn. Reson. Imaging*, **10**, 279–288 (1992).
94. P.M. Glover, R.W. Bowtell, G.D. Brown, P. Mansfield, 'A Microscope Slide Probe for High Resolution Imaging at 11.7 Tesla', *Magn. Reson. Med.*, **31**, 423–428 (1994).
95. R. Bowtell, P. Mansfield, J.C. Sharp, G.D. Brown, M. McJury, P.M. Glover, 'NMR Microscopy at 500 MHz. Cellular Resolution in Biosystems', in *Magnetic Resonance Microscopy*, eds. B. Blunich, W. Kuhn, VCH, New York, 427–439, 1992.
96. J.K.M. Roberts, J.H. Xia, 'High-resolution NMR Methods for Study of Higher Plants', *Methods Cell Biol.*, **49**, 245–258 (1995).
97. J.K.M. Roberts, 'Plant Physiology', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, New York, 3631–3633, 1996.
98. G.G. Fox, R.G. Ratcliffe, '^{31}P NMR Observations on the Effect of the External pH on the Intracellular pH Values in Plant Cell Suspension Cultures', *Plant Physiol.*, **93**, 512–521 (1990).
99. E. Gout, R. Bligny, N. Pascal, R. Douce, '^{13}C Nuclear Magnetic Resonance Studies of Malate and Citrate Synthesis and Compartmentation in Higher Plant Cells', *J. Biol. Chem.*, **268**, 3986–3992 (1993).
100. J.K.M. Roberts, M.K.L. Pang, 'Estimation of Ammonium Ion Distribution between Cytoplasm and Vacuole Using Nuclear Magnetic Resonance Spectroscopy', *Plant Physiol.*, **100**, 1571–1574 (1992).
101. K. Chang, J.K.M. Roberts, 'Observation of Cytoplasmic and Vacuolar Malate in Maize Root Tips by ^{13}C-NMR Spectroscopy', *Plant Physiol.*, **89**, 197–203 (1989).
102. K. Chang, J.K.M. Roberts, 'Cytoplasmic Malate Levels in Maize Root Tips during K^+ Ion Uptake Determined by ^{13}C-NMR Spectroscopy', *Biochim. Biophys. Acta*, **1092**, 29–37 (1991).
103. K. Chang, J.K.M. Roberts, 'Quantitation of Rates of Transport, Metabolic Fluxes, and Cytoplasmic Levels of Inorganic Carbon in Maize Root Tips during K^+ Ion Uptake', *Plant Physiol.*, **99**, 291–297 (1992).
104. E. Gout, R. Bligny, R. Douce, 'Regulation of Intracellular pH Values in Higher Plant Cells. Carbon-13 and Phosphorus-31 Nuclear Magnetic Resonance Studies', *J. Biol. Chem.*, **267**, 13 903–13 909 (1992).
105. Y.Y. Ford, G.G. Fox, R.G. Ratcliffe, R.J. Robins, 'In Vivo ^{15}N-NMR Studies of Secondary Metabolism in Transformed Root Cultures of *Datura stramonium* and *Nicotiana tabacum*', *Phytochemistry*, **36**, 333–339 (1994).
106. Y.Y. Ford, R.G. Ratcliffe, R.J. Robins, 'Phytohormone-induced GABA Production in Transformed Root Cultures of *Datura stramonium*: an In Vivo ^{15}N-NMR Study', *J. Exp. Bot.*, **47**, 811–818 (1996).
107. S.A. Robinson, A.P. Slade, G.G. Fox, R. Phillips, R.G. Ratcliffe, G.R. Stewart, 'The Role of Glutamate Dehydrogenase in Plant Nitrogen Metabolism', *Plant Physiol.*, **95**, 509–516 (1991).
108. G.G. Fox, R.G. Ratcliffe, S.A. Robinson, G.R. Stewart, 'Evidence for Deamination by Glutamate Dehydrogenase in Higher Plants: Commentary', *Can. J. Bot.*, **73**, 1112–1115 (1995).
109. T.A. Thorpe, K. Bagh, A.J. Cutler, D.I. Dunstan, D.D. McIntyre, H.J. Vogel, 'A ^{14}N and ^{15}N Nuclear Magnetic Resonance Study of Nitrogen Metablolism in Shoot-forming Cultures of White Spruce (*Picea glauca*) Buds', *Plant Physiol.*, **91**, 193–202 (1989).
110. L.O. Sillerud, J.W. Heyser, 'Use of ^{23}Na-nuclear Magnetic Resonance to Follow Sodium Uptake and Efflux in NaCl-adapted and Nonadapted Millet (*Panicum miliaceum*) Suspensions', *Plant Physiol.*, **75**, 269–272 (1984).
111. W.V. Gerasimowicz, S.-I. Tu, P.E. Pfeffer, 'Energy Facilitated Na^+ Uptake in Excised Corn Roots via ^{31}P and ^{23}Na NMR', *Plant Physiol.*, **81**, 925–928 (1986).
112. C.M. Spickett, N. Smirnoff, R.G. Ratcliffe, 'An In Vivo Nuclear Magnetic Resonance Investigation of Ion Transport in Maize (*Zea mays*) and *Spartina anglica* Roots During Exposure to High Salt Concentrations', *Plant Physiol.*, **102**, 629–638 (1993).

113. R.B. Lee, R.G. Ratcliffe, T.E. Southon, 'P-31 NMR Measurements of the Cytoplasmic and Vacuolar P_i Content of Mature Maize Roots – Relationships with Phosphorus Status and Phosphate Fluxes', *J. Exp. Bot.*, **41**, 1063–1078 (1990).
114. R.B. Lee, R.G. Ratcliffe, 'Subcellular Distribution of Inorganic Phosphate, and Levels of Nucleoside Triphosphate, in Mature Maize Roots at Low External Phosphate Concentrations: Measurements with ^{31}P NMR', *J. Exp. Bot.*, **44**, 587–598 (1993).
115. J.K.M. Roberts, D. Wemmer, O. Jardetzky, 'Measurement of Mitochondrial ATPase Activity in Maize Root Tips by Saturation Transfer ^{31}P Nuclear Magnetic Resonance', *Plant Physiol.*, **74**, 632–639 (1984).
116. J.K.M. Roberts, A.N. Lane, R.A. Clarke, R.H. Nieman, 'Relationships Between the Rate of Synthesis of ATP and the Concentrations of Reactants and Products of ATP Hydrolysis in Maize Root Tips, Determined by ^{31}P Nuclear Magnetic Resonance', *Arch. Biochem. Biophys.*, **240**, 712–722 (1985).
117. J.H. Xia, J.K.M. Roberts, 'Improved Cytoplasmic pH Regulation, Increased Lactate Efflux, and Reduced Cytoplasmic Lactate Levels are Biochemical Traits Expressed in Root Tips of Whole Maize Seedlings Acclimated to a Low-oxygen Environment', *Plant Physiol.*, **105**, 651–657 (1994).
118. J.K.M. Roberts, M.A. Hooks, A.P. Miaullis, S. Edwards, C. Webster, 'Contribution of Malate and Amino Acid Metabolism to Cytoplasmic pH Regulation in Hypoxic Maize Root Tips Studied Using Nuclear Magnetic Resonance Spectroscopy', *Plant Physiol.*, **98**, 480–487 (1992).
119. F. Menegus, L. Cattaruzza, M. Mattana, N. Beffagna, E. Ragg, 'Response to Anoxia in Rice and Wheat Seedlings', *Plant Physiol.*, **95**, 760–767 (1991).
120. C.M. Spickett, N. Smirnoff, R.G. Ratcliffe, 'Metabolic Response of Maize Roots to Hyperosmotic Shock. An In Vivo ^{31}P Nuclear Magnetic Resonance Study', *Plant Physiol.*, **99**, 856–863 (1992).
121. J. Guern, Y. Mathieu, M. Pean, C. Pasquier, J.-C. Beloeil, J.Y. Lallemand, 'Cytoplasmic pH Regulation in *Acer pseudoplatonus* Cells', *Plant Physiol.*, **82**, 840–845 (1986).
122. R.J. Gillies, *NMR in Physiology and Biomedicine*, Academic Press, New York, 1994.
123. M. Gussoni, F. Greco, R. Consonni, H. Molinari, G. Zannoni, G. Bianchi, L. Zetta, 'Application of NMR Microscopy to the Histochemistry Study of Olives (*Olea europaea* L.)', *Magn. Reson. Imaging*, **11**, 259–268 (1993).
124. N.R. Jagannathan, V. Govindaraju, P. Raghunathan, 'In Vivo Magnetic Resonance Study of the Histochemistry of Coconut (*Cocos nucifera*)', *Magn. Reson. Imaging*, **13**, 885–892 (1995).
125. A. Metzler, W. Kockenberger, M.v. Kienlin, E. Komor, A. Haase, 'Quantitative Measurement of Sucrose Distribution in *Ricinus communis* Seedlings by Chemical Shift Microscopy', *J. Magn. Reson., Ser. B*, **105**, 249–252 (1994).
126. A. Metzler, M. Izquierdo, A. Ziegler, W. Kockenberger, E. Komor, M.v. Kiemkin, A. Haase, M. Descorps, 'Plant Histochemistry by Correlation Peak Imaging', *Proc. Natl. Acad. Sci. USA*, **92**, 11 912–11 915 (1995).
127. J.M. Pope, H. Rumpel, W. Kuhn, R. Walker, D. Leach, V. Sarafis, 'Applications of Chemical-shift-selective NMR Microscopy to the Non-invasive Histochemistry of Plant Materials', *Magn. Reson. Imaging*, **9**, 357–363 (1991).
128. B.A. Goodman, B. Williamson, J.A. Chudek, 'Identification of the Distribution of Sugars in Grapes Using Chemical Shift Selective NMR Microscopy', *Magn. Reson. Imaging*, **11**, 1039–1041 (1993).
129. T.Y. Tse, R.M. Spanswick, L.W. Jelinski, 'Quantitative Evaluation of NMR and MRI Methods to Measure Sucrose Concentrations in Plants', *Protoplasma*, **94**, 54–62 (1996).
130. N. Ishida, M. Koizumi, H. Kano, 'Location of Sugars in Barley Seeds During Germination by NMR Microscopy', *Plant Cell Environ.*, **19**, 1415–1422 (1996).
131. A. Bax, R. Freeman, 'Investigation of Complex Networks of Spin–Spin Coupling by Two-dimensional NMR', *J. Magn. Reson.*, **44**, 542–561 (1981).
132. L. Braunschweiler, R.R. Ernst, 'Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy', *J. Magn. Reson.*, **53**, 521–528 (1983).
133. C. Griesinger, H. Schwalbe, J. Schleucher, M. Sattler, 'Proton-detected Heteronuclear and Multidimensional NMR', in *Two-dimensional NMR Spectroscopy*, eds. W.R. Croasmun, R.M.K. Carlson, VCH, New York, 457–580, 1994.
134. G.E. Martin, R.C. Crouch, 'Two-dimensional NMR in Natural Products and Pharmaceutical Chemistry', in *Two-dimensional NMR Spectroscopy*, eds. W.R. Croasmun, R.M.K. Carlson, VCH, New York, 873–914, 1994.
135. G. Bodenhausen, D. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185–189 (1980).
136. A. Bax, S. Subramanian, 'Sensitivity-enhanced Two-dimensional Heteronuclear Shift Correlation NMR Spectroscopy', *J. Magn. Reson.*, **67**, 565–569 (1986).
137. A. Bax, M.F. Summers, '^{1}H and ^{13}C Assignments from Sensitivity Enhanced Detection of Heteronuclear Multiple-bond Connectivity by 2D Multiple Quantum NMR', *J. Am. Chem. Soc.*, **108**, 2093–2094 (1986).
138. C.E. Hadden, G.E. Martin, V.V. Krishnamurthy, 'Improved Performance Accordion Heteronuclear Multiple-bond Correlation Spectroscopy – IMPEACH – MBC', *J. Magn. Reson.*, **140**, 274–280 (1999).
139. G.E. Martin, C.E. Hadden, R.C. Crouch, V.V. Krishnamurthy, 'ACCORD-HMBC: Advantages and Disadvantages of Static Versus Accordion Excitation', *Magn. Reson. Chem.*, **37**, 517–528 (1999).

140. J. Keeler, R.T. Clowes, A.L. Davis, E.D. Laue, 'Pulsed-field Gradients: Theory and Practice', *Methods Enzymol.*, **239**, 145–207 (1994).
141. T. Parella, 'Pulsed Field Gradients: A New Tool for Routine NMR', *Magn. Reson. Chem.*, **36**, 467–495 (1998).
142. R.C. Crouch, G.E. Martin, 'Micro Inverse-detection – A Powerful Technique for Natural Product Structure Elucidation', *J. Nat. Prod.*, **55**, 1343–1347 (1992).
143. G.E. Martin, C.E. Hadden, A.N. Tackie, M.H.M. Sharaf, P.L. Schiff, Jr, 'Rapid Acquisition of $^{1}H-^{13}C$ and $^{1}H-^{15}N$ Heteronuclear Chemical Shift Correlation Data at the Submilligram Level Using SMIDG (Submicro Inverse-detection Gradient) NMR Probe Technology', *Magn. Reson. Chem.*, **37**, 529–537 (1999).
144. G.E. Martin, J.E. Guido, R.H. Robins, M.H.M. Sharaf, P.L. Schiff, Jr, A.N. Tackie, 'Submicro Inverse-detection Gradient NMR: a Powerful New Way of Conducting Structure Euclidation Studies with <0.05 mmol Samples', *J. Nat. Prod.*, **61**, 555–559 (1998).
145. G.E. Martin, R.C. Crouch, A.P. Zens, 'Gradient Submicro Inverse Detection: Rapid Acquisition of Inverse-detected Heteronuclear Chemical Shift Correlation Data on Submicromole Quantities of Material', *Magn. Reson. Chem.*, **36**, 551–557 (1998).
146. G.E. Martin, C.H. Hadden, 'Comparison of 1.7 mm Submicro and 3 mm Micro Gradient NMR Probes for the Acquisition of $^{1}H-^{13}C$ and $^{1}H-^{15}N$ Heteronclear Shift Correlation Data', *Magn. Reson. Chem.*, **37**, 721–729 (1999).
147. D.G. Davis, 'Enhanced Correlations in ^{1}H Detected Heteronuclear Two-dimensional Spectroscopy via Spin-locked ^{1}H Magnetization Transfer', *J. Magn. Reson.*, **84**, 417–424 (1989).
148. T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, I.D. Campbell, 'Comparison of Techniques for ^{1}H Detected Heteronuclear $^{1}H-^{15}N$ Spectroscopy', *J. Magn. Reson.*, **87**, 488–501 (1990).
149. C.E. Hadden, G.E. Martin, 'Low-level Long-range $^{1}H-^{15}N$ Heteronuclear Shift Correlation at Natural Abundance Using Submicro NMR Techniques', *J. Nat. Prod.*, **61**, 969–972 (1998).
150. C.E. Hadden, G.E. Martin, 'Low-level Long-range $^{1}H-^{15}N$ Heteronuclear Shift Correlation at Natural Abundance Using a Combination of 2D and Selective 1D NMR Methods With Submicro NMR Probe Technology', *Magn. Reson. Chem.*, **37**, 385–388 (1999).
151. C.E. Hadden, M.H.M. Sharaf, J.E. Guido, R.H. Robins, A.N. Tackie, C.H. Phoebe, Jr, P.L. Schiff, Jr, G.E. Martin, '11-Isopropylcryptolepine: a Novel Alkaloid Isolated from *Cryptolepis sanguinolenta* Characterized Using Submicro NMR Techniques', *J. Nat. Prod.*, **62**, 238–240 (1999).
152. A.A. Bothner-By, R.L. Stephens, J. Lee, C.D. Warren, R.W. Jeanloz, 'Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame', *J. Am. Chem. Soc.*, **106**, 811–813 (1984).
153. B.C. Sweatman, R.D. Farrant, P.N. Sanderson, I. Philippe, S.R. Salman, J.K. Nicholson, J.C. Lindon, 'Evaluation of the Detection Limits of Directly Coupled 600 MHz ^{1}H and $^{1}H-^{13}C$ HPLC–NMR Spectroscopy', *J. Magn. Reson. Anal.*, **1**, 9–12 (1995).
154. T.M. Logan, N. Murali, G. Wang, C. Jolivet, 'Application of a High-resolution Superconducting NMR Probe in Natural Product Structure Determination', *Magn. Reson. Chem.*, **37**, 762–765 (1999).
155. A.G. Webb, S.C. Grant, 'Signal-to-noise and Magnetic Susceptibility Trade-offs in Solenoid Microcoils for NMR', *J. Magn. Reson., Ser. B*, **113**, 83–87 (1996).
156. R. Subramanian, M.M. Lam, A.G. Webb, 'RF Microcoil Design for Practical NMR of Mass-limited Samples', *J. Magn. Reson., Ser. B*, **133**, 227–231 (1998).
157. D.L. Olson, T.L. Peck, A.G. Webb, R.L. Magin, J.V. Sweedler, 'High-resolution Microcoil ^{1}H-NMR for Mass-limited, Nanoliter Volume Samples', *Science*, **270**, 1967–1970 (1995).

Particle Size Analysis in Food

Margaret M. Robins
Institute of Food Research, Norwich, UK

1 Introduction: The Importance of Particle Size 1
2 Direct Techniques 4
2.1 Microscopy 4
2.2 Sieving 5
3 Optical Methods 6
3.1 Light Transmission and Back-scattering 6
3.2 Variable-angle Light Scattering 8
3.3 Dynamic Light Scattering 10
4 Electric Sensing Zone 12
4.1 Description of Technique 12
4.2 Advantages and Limitations 12
5 Sedimentation 13
5.1 Gravitation 13
5.2 Centrifugation 14
5.3 Field-flow Fractionation 15
6 Acoustic Techniques 16
6.1 Ultrasonic Spectroscopy 16
6.2 Electroacoustics 17
7 Nuclear Magnetic Resonance 18
7.1 Description of the Technique 18
7.2 Advantages and Limitations 19
7.3 Magnetic Resonance Imaging 19
8 Evaluation of the Available Techniques for Differing Particle Types 19
8.1 Powders 19
8.2 Emulsions and Suspensions 19
8.3 Foams 20
8.4 Biological Systems 20
8.5 Complex Dispersions 20
Acknowledgments 21
Abbreviations and Acronyms 21
Related Articles 21
References 21

Techniques are described for determining particle size in food dispersions involving food particles ranging in size from 0.1 to 1000 μm, where the presence of the particles confers special properties such as texture, appearance or flavor release. Those foods of plant origin that are cellular in nature are outside the scope of this article, as their structure is not under the control of the food manufacturer. The techniques described are used to a greater or lesser extent on foods, and they are all methods in use in other industrial sectors. The techniques are classified according to their underlying physical principle. Summary tables are provided to act as quick reference guides for a potential user of the technique. After the detailed descriptions, the final section groups those techniques relevant to a specific type of dispersion.

The two most direct methods, microscopy and sieving, are discussed initially followed by three classes of optical technique, comparatively simple forward or back scattering, angular-dependent light scattering and dynamic light scattering. The next technique to be addressed is the electrical sensing zone originally developed by Coulter. Sedimentation techniques follow: gravitational sedimentation, centrifugal sedimentation and field-flow fractionation (FFF). All depend on the behavior of the particles in a force field that is balanced by viscous drag forces. The next two techniques are both acoustic in operation, either direct ultrasonic attenuation or electroacoustics. Finally, the novel applications based on nuclear magnetic resonance (NMR) are presented.

1 INTRODUCTION: THE IMPORTANCE OF PARTICLE SIZE

It is difficult to formulate a meal without including particulate foods. The particles include gas bubbles in bread, oil droplets in dairy products and sugar crystals in chocolate. The presence of the particles gives the food its character, texture and flavor. The long-term stability is also critically affected by the type and size of the particles in the food.[1] Table 1 gives examples of the particles occurring in a range of foods, their size range and the properties affected by the particle size.

Particles may be gas, liquid, solid or biological. Other particles such as the filled cells in plant tissue are outside the scope of this article, which is confined to systems where the particle size is controllable and an important factor in the properties of the food. The particles of interest in foods are generally in the range 0.1 μm to 1 mm because particles in this size range have the most dramatic effect on the quality of foods.

There is a scarcity of texts devoted solely to particles in foods, but much of the colloidal literature is relevant, and this article draws heavily on particle sizing techniques developed originally for the chemical or mineral industries. The best sources of information on colloidal systems are those by Dickinson and Stainsby[1] (food), Becher[2] (emulsions) and Allen[3] (general particle sizing).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Particles in foods

Food	Particle type	Size range (μm)	Properties affected by the particle size
Milk and cream	Oil droplets	0.1–10	Stability to creaming, texture
Butter and spreads	Water droplets	1–20	Texture, microbiological stability, flavor
Sauces, dressings	Oil droplets	0.2–5	Stability to separation, texture, flavor
Mustard	Solid mustard particles	1–20	Flavor, stability to separation, grittiness
Chocolate	Sugar crystals	5–100	Grittiness
Ice cream	Fat particles	0.2–3	Flavor, texture
	Sugar crystals	0.1–50	Grittiness
	Ice crystals	20–80	Texture
	Air bubbles	10–100	Texture, bulk density, appearance
Bread	Air bubbles	20–10 000	Loaf volume, texture, appearance
Beer	Gas bubbles	1–1000	Stability, texture, appearance, flavor
Beer and wine	Yeast	3–5 (aggregates 10–1000)	Speed of separation after fermentation, clarity

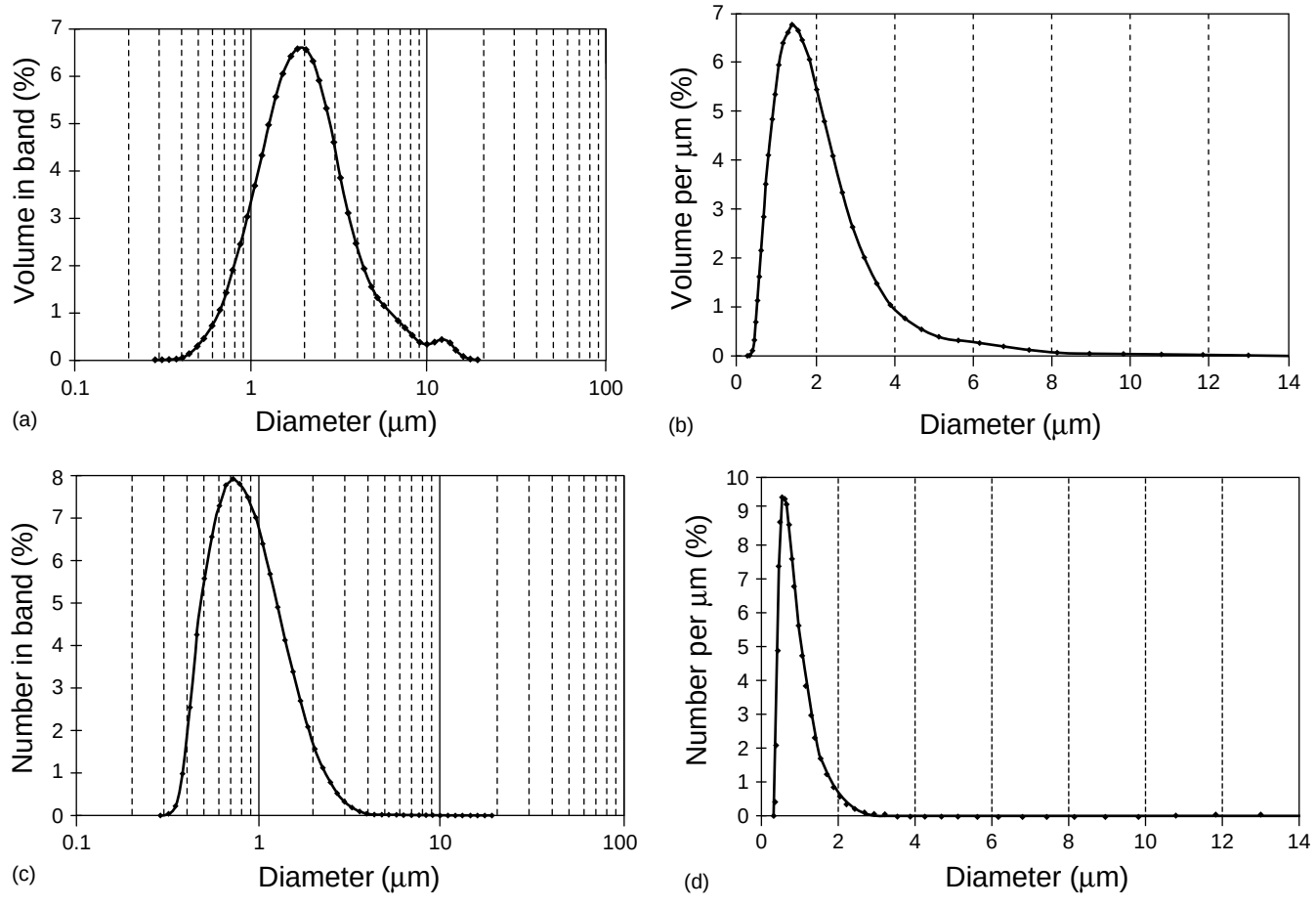

Figure 1 A typical log-normal size distribution of an oil-in-water emulsion measured using a variable-angle light scattering sizer. (a) Volume distribution using log(diameter) axis. (b) Volume distribution using linear diameter axis. (c) Number distribution using log(diameter) axis. (d) Number distribution using linear diameter axis.

Techniques are classified by the physical principle underlying the measurement. The emphasis is on instruments that are available commercially and are most commonly used for food materials. Special emphasis has also been given to the new techniques, such as acoustics and NMR, where there is potential for increased applications in the future. Each section is headed by a summary table for quick reference, and the last entry presents an overview of the techniques suitable for each type of application. The critical factors in selecting the

best method for a particular application include accuracy, susceptibility to perturbing factors, ease and speed of use, breadth of distribution, intrusiveness, preparation procedure (e.g. dilution), commercial availability and cost. The article attempts to address the majority of these factors; the issues of commercial availability and cost are omitted, but Allen[3] gives an excellent list of the instruments currently available.

Particles in food are rarely all the same size (monodisperse), and for many purposes it is important to measure the whole distribution of sizes present. A common distribution produced by a milling or homogenization process is the log-normal distribution of diameters.[2] Figure 1(a) shows the droplet size distribution in an emulsion, expressed as the volume in each size band, where the size bands are plotted on a logarithmic scale. This is the form of data produced by instruments with a wide particle size range, such as the variable-angle light scattering technique described in section 3.2 below. The general Gaussian shape of the distribution, when plotted against log(diameter), indicates the log-normal nature of the distributions. The volume-mean diameter, commonly denoted by D_{43}, is calculated to be 2.56 μm, close to the major peak of the log-normal curve. The data are not exactly log-normally distributed, as a small second peak is observed at approximately 12 μm. However, if the same data are plotted using linear diameter bins, as in Figure 1(b), the skewed shape is evident, and the apparent bimodality disappears. The same set of particle size data can be shown as a number distribution, which is the primary output of techniques involving individual particle counting, such as the electric sensing zone method described in section 4 below. Figure 1(c) shows the previous data converted to a number distribution, where the results are expressed as the number (as a percentage of the total number) of particles in each diameter bin. Comparison with Figure 1(a) shows the dependence of the shape of the distribution on the analysis protocol. In practice it is best to express data in the form (volume or number) that is closest to the primary data collected by the instrument, as the conversion process gives emphasis to regimes of diameter that may not have been warranted by the original measurement. For example, if data are initially collected by volume, which emphasizes the larger diameter fractions, then conversion to a number distribution can give large errors in the number of small particles where the instrument measured very low volumes. Counting methods are often best when the distribution of particle sizes is not very wide; in a highly polydispersed sample, a very large number of particles needs to be detected and classified in order to accumulate enough counts in the largest size categories. The number distributions can also be expressed in linear or logarithmic size bands, and Figure 1(d) shows the effect on the shape of the distribution of plotting on a linear scale.

Even where it is convenient to use a mean value of diameter, the distribution determines the relationship between the various types of mean obtained for a sample. Table 2 shows the definitions of the most common mean values used to describe a particle size and its distribution and the calculated means for the distribution in Figure 1. The data show, for example, that the mean diameter calculated in order to obtain a total surface area of the particles (D_{32} in Table 2) is invariably smaller than that which, when cubed, gives the average particle volume (D_{43}).

It is worth remembering in the discussion of particle size that no two techniques ever agree exactly, and there is always a debate as to the value of the ''true'' size. In practice, the test is whether the measurement gives results that are transparent in their analysis and are proved to be useful in the design and control of foods. Polydispersity and irregular shape are two recurring features, and the optimum technique for a specific product depends on the end-use property that is important. For example, if stability to separation is important, the hydrodynamic diameter is the appropriate size parameter.

Table 2 Definitions and examples of mean particle size parameters for polydisperse size distributions

Name	Symbol	Definition (where $N(D)$ is the number of particles of diameter D)	Value for distribution in Figure 1 (μm)
Arithmetic (number mean)	D_{10}	$\frac{\sum DN(D)}{\sum N(D)}$	1.00
Number-surface mean	D_{20}	$\frac{\sum D^2N(D)}{\sum N(D)}$	1.14
Number-volume mean	D_{30}	$\frac{\sum D^3N(D)}{\sum N(D)}$	1.31
Length-surface mean	D_{21}	$\frac{\sum D^2N(D)}{\sum DN(D)}$	1.29
Length-volume mean	D_{31}	$\sqrt{\frac{\sum D^3N(D)}{\sum DN(D)}}$	1.50
Surface-volume mean (Sauter mean)	D_{32}	$\frac{\sum D^3N(D)}{\sum D^2N(D)}$	1.75
Volume or weight mean	D_{43}	$\frac{\sum D^4N(D)}{\sum D^3N(D)}$	2.56

2 DIRECT TECHNIQUES

2.1 Microscopy

Microscopic techniques are summarized in Table 3.

2.1.1 Description of the Technique

Microscopy is the forerunner of all other sizing techniques. It is still seen by some as the ''gold standard'' as it provides reassuringly direct visual images of the particles. The traditional instrument is the light microscope, capable of imaging and photographing particles in the range 3–150 μm with accuracy, provided that optical contrast can be achieved. Larger particles may be examined using a small magnifier or with the naked eye. Particles smaller than 3 μm may be observed using visible light, but diffraction effects cause overestimation of the diameters by up to 10%.[3] Dry powders can be viewed, but most applications require the sample to be presented as a thin film of particles in an appropriate nonswelling liquid[4] sandwiched between two glass surfaces. Biological systems may need to be stained to achieve the required contrast, and bacteria are at the resolution limit of optical microscopy. Fluorescence techniques use a dye to provide emission in the ultraviolet spectrum, with improvement in resolution proportional to the wavelength. Differential interference techniques are also valuable, to improve contrast between biological cells or oil droplets and an aqueous suspending medium without staining. If the particle is nonspherical, its projected area depends on the orientation on the slide.

Most food systems scatter light strongly, and samples require dilution to resolve nonoverlapping particles. The use of laser scanning confocal microscopy has allowed optical observation of more concentrated dispersions, to a depth of several particle diameters if the particles have been made strongly contrasting using fluorescence techniques.[5] The instrument uses a pin-hole to restrict the image to light scattered from a thin plane at a chosen depth within the sample. The light scattered by the particles in other planes is eliminated by the pinhole; consequently, an image is obtained only of the particles at the required depth. Confocal microscopy has been applied to on-line imaging.

Particles approaching the resolution of optical microscopy are usually subject to visible Brownian motion and image capture must be rapid. It is important to obtain thin specimens without deforming the particles or causing fractionation when excess dispersion flows outside the field of view.

Visual observation or measurement from photographs is very time-consuming if the size distribution has significant width, and care is needed to ensure that representative particles are selected for estimation of their size.[6] As a guide, if particles are counted randomly until the largest size category has collected at least 25 particles, the volume mean will have an error of typically 2%.[3] Each particle usually requires refocusing, as the depth of field is rarely sufficient to observe the range of particles at a single working distance. In recent years, automatic image analysis techniques have enabled more rapid quantification of particle size distributions, but they

Table 3 Summary of microscopic techniques used for particle sizing

Characteristics	Comments
Operating principle	Direct imaging of a projection on a plane, using light or electrons
Particle type	Solid, liquid, gas and biological particles
Property measured	Projected diameter, shape, number distributions of size
Size range	3–1000 μm (optical), 0.02–1000 μm (SEM), 0.002–1 μm (TEM)
Concentration range	Variable
Calibration required	Not in principle, but magnification verified using standard graticules or grids
Model/parameters required	Geometric models to account for shape of particles and obtain surface area and volume from projected areas
Speed of operation	Minutes to hours
Applicability to on-line use	Scanning laser microscopy available for on-line use
Advantages	Direct visualization of individual particles, shape information available, surface characteristics of particles discernible, detects aggregation, automatic image analysis available, chemical information available
Limitations	Slow, sample preparation important, combination of techniques needed for wide distribution, large number of particles required to obtain good statistics, artifacts may perturb automatic analysis, intrusive, limited depth of focus restricts field of view to few particles

SEM, scanning electron microscopy; TEM, transmission electron microscopy.

still depend on the preparation technique, the choice of particles to analyze, depth of focus, the contrast that defines the edge of the particle and shape effects.

Image analysis is the preferred method of characterizing food foams. Solid foams, as in bread and cake, are straightforward to prepare, as a cut edge enables both cell sizes and the cell wall thickness to be measured.[7] Liquid foams are more difficult to handle and are often very unstable. One protocol involves preparing the foam by sparging gas through a known volume of surfactant solution in a glass-walled container. Images of the bubble size viewed through the wall are captured photographically or with a video recorder, and processed using image analysis software.[8] On-line versions of the technique have been developed to monitor in real-time the foam structures produced in processes such as ice-cream manufacture.[9]

Electron microscopy is used to examine and measure the size of particles below the optical range. Scanning electron microscopy requires the sample to be mounted with the particles projecting from the surface, usually achieved by freeze-fracturing procedures, and coating with a conducting layer to present a surface of uniform reflectivity for the electron beam. The scattering intensity as a function of angle is then simply related to the relief of the surface. Particles in the range 0.02–1000 μm are readily examined and measured using this technique; however, since the samples are prepared by fracturing the particles, the estimation of size requires correction factors to account for nondiametric fracture planes. For spherical particles, there are simple algorithms to convert the apparent diameters to the overall size distribution.[10] Scanning electron microscopy is routinely used to examine biological particles such as bacteria, which are marginally too small for optical methods.[11] It is also used for characterizing the small bubbles and particles in complex food dispersions such as ice cream.[10]

Very small particles may be viewed using transmission electron microscopy, which has a resolution of less than a nanometer. Samples are prepared by replication, which requires the particles to be initially projecting from their substrate. It is rarely used for analysis of food particles.

2.1.2 Advantages and Limitations

There is an argument that every sizing experiment should start with examination of the sample in a microscope, to determine the state of dispersion.[3] Microscopy remains one of the fundamental characterization techniques. However, it is highly intrusive, usually requiring dilution or elaborate preparation protocols, and is time-consuming as a primary method for quantitative particle size analysis.

There is always a risk that the preparation or examination procedure changes the particle size or shape, whether by dissolving, flattening, overheating or damage by an electron beam. Since comparatively few particles are viewed in any one slide, the sampling procedure needs to be rigorous. Foams present special problems, as the viewed surface may not be representative of the bulk. Microscopic techniques are superior to other methods for characterizing the shape of particles, provided that the particles are randomly oriented with respect to the viewing direction. Images of irregular particles can be analyzed to give fractal dimensionality from measures of the projected perimeter lengths and particle area.[12]

2.2 Sieving

Techniques to assess particles by sieving are summarized in Table 4.

2.2.1 Description of the Technique

Sieving is one of the oldest particle classification methods; it is very heavily used to characterize flours and starches.[3]

Table 4 Summary of sieving techniques used for particle sizing

Characteristics	Comments
Operating principle	Classification of particles from a fluidized powder by passage through well-defined apertures
Particle type	Solid particles in air or in liquid dispersion
Property measured	Upper limit of the two smaller dimensions
Size range	40 μm to a few millimeters (dry), 5–100 μm (wet)
Concentration range	Dry powder, or <1% (v/v) in wet dispersions
Calibration required	Standards used to monitor size and uniformity of sieve apertures
Model/parameters required	Statistical models and correction factors for irregular particles
Speed of operation	Minutes
Applicability to on-line use	Limited
Advantages	Direct classification of powders, removes aggregates, cheap and simple equipment, easy to use, established standards
Limitations	Slow, moisture content critical, clogging, multistage process to obtain distribution of sizes

The principle is that a particle will pass through the holes in a sieve if two of the dimensions are both smaller than the mesh size. The majority of milling and dry-grinding operations in the food industry rely on sieving to grade the particles and determine their size distribution. The fractions may be separated for different purposes, or the data used to control the milling conditions. Sieves are available in a wide range of mesh sizes, from macroscopic dimensions down to about 5 μm. The apertures are usually square, prepared by weaving wires or synthetic fibers, or circular drill-holes. Banks of sieves are used, the coarsest at the top, with successive mesh sizes reducing by $\sqrt{2}$ in mesh diameter. The sample is loaded onto the top sieve and vibrated until typically 98% of the powder has passed through at least the first mesh. For very small particles, less than 20 μm, wet sieving is used, where the particles are dispersed in a liquid and forced by pressure through the apertures. Ultrasonic waves are used as aids in wet sieving to generate standing waves that suspend the particles according to their volume, which fractionates the suspension prior to the sieving (filtration) process.[3]

2.2.2 *Advantages and Limitations*

The use of sieves is very simple and appealing as a direct method. Sieves allow powders and wet particles to be classified as well as sized, and the technique is universally recognized, with associated industry standards. However, the method is very time-consuming and does not give precise size distributions. Fine sieves clog very easily, and the cleaning process can cause damage to the delicate materials. If the particles are irregular in shape, they may get caught on different sieves depending on the agitation and their orientation on the mesh. Many food powders, such as flour, sugar and starches, are very sensitive to humidity in the atmosphere, and the sieving process is hampered by particle aggregates. Violent agitation to disrupt the aggregates can also damage the sieve mesh.

A number of factors have been found to affect the results of sieving operations.[3] These include the duration of sieving, the variation of sieve apertures from the quoted size, wear, errors of sampling and the effects of different equipment and operation. The use of small volumes in each operation provides more rapid separation but may lead to inaccuracies in recovering and weighing small amounts. In wet sieving, the particles have to be dried before weighing, which introduces a further error in handling.

When a particle can only pass through the sieve if it is presented to the mesh in a particular orientation, it is termed a ''near-mesh'' particle. The fate of such a particle varies according to the probability of its approach in a certain orientation, and its shape. There are mathematical procedures to estimate the shape of particles using well-defined square mesh sieves.[13]

3 OPTICAL METHODS

3.1 Light Transmission and Back-scattering

Table 5 summarizes the characteristics of light transmission and back-scattering methods.

3.1.1 *Turbidity*

One of the simplest measurements to make, although not to analyze, is the turbidity of a suspension or emulsion. The sample is placed in a thin glass optical cell and inserted into the path of a laser beam or other monochromatic light source. The intensity, I, of the transmitted beam is measured using a photodetector and compared with the intensity I_0 of the transmitted light when the cell contains the continuous phase alone. The turbidity τ of the dispersion is defined by the Beer–Lambert law (Equation 1):

$$I = I_0 \exp(-\tau L) \tag{1}$$

where L is the path length of the light in the sample (cell width). The turbidity depends on the concentration of the particles and their scattering cross-section, which in turn is related to the particle cross-sectional area.[14] In principle, if the concentration, refractive index contrast and absorbance of the particles are known, the mean particle area can be derived from a single turbidity measurement, but in practice, several wavelengths of light are used and calibration data obtained for the dispersion under investigation. If the particles are very well characterized in terms of their optical properties, the use of several wavelengths and Mie theory enables the distribution of particle diameters to be inferred.[15]

The technique is routinely used to determine the droplet size of raw and homogenized milk, and after the acidification step in cheese-making. It is also used to determine the emulsifying index of emulsifiers, where the total droplet area is the property required.[16]

3.1.2 *Optical Sensing Zone*

The optical sensing zone technique[17] is similar to the electrical sensing method in section 4 below. Particles are detected as they obscure a beam of light. Their passage produces a change of transmitted light intensity that is related, via the refractive index contrast, to the size of the particle. Calibration is required on the same type of particle because of the refractive index dependence. It is suitable for particles between 0.5 and 1000 μm in diameter.

A dynamic version of the optical sensing zone technique has been developed to monitor aggregation in water treatment processes or mineral processing.[18] In this

Table 5 Summary of light transmission and back-scattering techniques used for particle size analysis

Characteristics	Comments
Operating principle	Absorption of light by the particles in a dispersion or foam, measured in the straight-through or back direction
Particle type	Powders, suspensions, emulsions, foams, bacteria, aggregates
Property measured	Particle cross-sectional area, particle number (dynamic techniques), bubble number and size
Size range	0.08–1.5 μm (turbidity), 1–9000 μm (static or dynamic optical sensing zone), 100–10 000 μm (bubbles)
Concentration range	Very dilute for transmission, only a small percentage value for back-scattered versions
Calibration required	In practice, calibration required for absolute size data of liquid or solid particles, none needed for foam technique
Model/parameters required	Liquid or solid particles: optical models, refractive indices, absorbance values, concentration, but can be used empirically for comparisons; foams: no parameters or models
Speed of operation	Seconds
Applicability to on-line use	In principle, with sampling and dilution procedures for turbidity measurement. Dynamic techniques used for in-line aggregation monitoring. In principle, foam technique adaptable for on-line use
Advantages	Cheap, simple apparatus required. No sample preparation required except dilution; combination of techniques suitable for wide particle size range. Back-scattering versions applicable to intermediate concentration range
Limitations	Turbidity gives mean particle size, single-particle methods provide some information of distribution but long sampling times needed to collect many data, not suitable for very broad size distributions, foam technique is intrusive

method, the fluctuations of light intensity resulting from particles passing through a beam are recorded, and the ratio of the root-mean-square value to the time-averaged mean gives the number and scattering cross-section of the particles. It is generally used to observe changes in particle size and number caused by aggregation but could, in principle, be calibrated to give absolute values.

3.1.3 Related Techniques Using Back-scattered Light

A variant of the optical sensing zone measures the light scattered back from a particle as it flows through a focused laser interrogation zone.[19] The use of a very fine beam and reflected light means that individual particles are sampled even in moderately concentrated dispersions (a small percentage volume per volume). The individual particles produce pulses in light intensity the size of which is related to the diameter of a chord of the particle, so number and chord length are separately recorded. Provided that the shape of the particle is known, the number distribution of diameters may be obtained for a range of 1.0–1000 μm. Although not common in the food industry, there are on-line versions available that are in use for suspensions of irregular solids, such as silica, and for flocs or aggregates.

Another related technique has been developed recently to characterize the bubble size distribution in liquid foams.[20] The foam is prepared under controlled conditions within a tall column. A fiber optic probe (20 μm diameter) is inserted into the foam and moved up or down at a constant speed, typically 10 cm s^{-1}. The fiber and associated detectors and circuitry detect the reflection coefficient of a beam of light projected ahead of the moving fiber tip. The instantaneous value of the reflected beam is dependent on whether the fiber tip is in the middle of an internal film or within a gas bubble in the foam. Sampling the reflected signal at a rate of 1 MHz provides a data point every 0.1 μm in the foam, to obtain a detailed distribution of chord lengths from bubbles in the interior of the foam. Analysis of the chord lengths assuming spherical bubbles gives a bubble size distribution covering the range 50 μm to several millimeters. The technique does not need calibration as it measures directly the distance across each bubble it encounters, but the data can be checked by comparing the measured gas/liquid ratio with the overall density of the foam.

3.1.4 Related Technique: Flow Cytometry

The use of flow cytometry to obtain particle size employs a forward light-scattering technique. It is used principally on biological dispersions, and bacteria are the main food-based application.[21] Individual particles (cells) are enclosed in a sheathing fluid and passed singly through an optical illumination and detector system. The intensity of forward-scattered light from the laser beam that impinges on the cell is related directly to its cross-sectional area. For rod-like bacteria, the flow cytometer can provide an

estimate of both characteristic dimensions of the cell. By detecting the fluorescence of the cell at a range of wavelengths, the flow cytometer can also distinguish cells of different species, either naturally fluorescent or after staining. The use of viable stains distinguishes live and dead cells, and the detector signals can be used to control electrodes to divert the particle stream into different collection pots according to size or viability.

3.1.5 Advantages and Limitations

The majority of the light transmission or reflection methods are nonintrusive and rapid to use, with comparatively simple apparatus, comprising a light source and detector electronics. Measurements are made only in the forward or backward direction; consequently, the light source and detector system may be mounted permanently in a fixed position. The samples need to be very dilute for turbidity or forward scattering using the optical sensing zone method. Dispersions of higher concentrations may be characterized by back-scattering (reflection), but as the methods require interrogation of single particles or single aggregates there is still a practical concentration limit of typically 5% (v/v). The single-particle methods may need to accumulate data over a timescale of at least minutes. Although ideal for comparative measurements, the data are complex to analyze from first principles, and calibration is usually needed to obtain absolute particle size data.

The dynamic versions are particularly well suited to monitoring processes in-line and are in common use in environmental applications. More complex devices include flow cytometry, designed to characterize multiple features of biological particles. Flow cytometry is unsurpassed in its ability to distinguish different populations of cells, by size or by their uptake of a specific fluorescent stain. However, it is inconvenient and slow for routine use on nonbiological particles.

The technique to characterize bubble size in foams has some particular features. This method measures directly the internal chord lengths of each foam bubble sampled, but it is limited by the diameter of the fiber optic probe (currently 20 μm). The probe is intrusive, which may affect the stability of the foam, and the tip may become coated with liquid which perturbs the data from the gas phase. However, it has been shown to give total gas volumes that agree well with the overall density of the foam. The technique is suitable for food and beverage foams of varying stability, which are very difficult to characterize by other techniques.

3.2 Variable-angle Light Scattering

Characteristics of variable-angle light-scattering techniques are summarized in Table 6.

3.2.1 Description of the Technique

Variable-angle light-scattering techniques measure the angular distribution of light scattered by particles in air or liquid. Both the intensity and the angular distribution of scattered light depend on the particle size.[14] For large particles (approximately 10 μm or more), Fraunhofer diffraction theory is applicable, which treats each particle as a black disc that diffracts a laser light beam in a characteristic angular pattern, the first maximum being at an angle $\theta = \lambda/d = 3\pi/2$ where λ is the wavelength of the light and d the disc diameter. Clearly, a polydisperse size distribution will have a set of superposed diffraction patterns, corresponding to each population. Smaller

Table 6 Summary of variable-angle light-scattering techniques used for particle size analysis

Characteristics	Comments
Operating principle	Angular dependence of coherent light scattered from particles of different sizes
Particle type	Aerosols, powders, emulsions, suspensions, biological dispersions
Property measured	Particle volume, expressed as diameter of equivalent spheres
Size range	0.05–3500 μm
Concentration range	Typically <0.01% (depends on size)
Calibration required	Not in principle provided that physical constants known, is needed in practice
Model/parameters required	Fraunhofer and Mie optical diffraction models, refractive index and absorbance of dispersed and continuous phases
Speed of operation	Seconds
Applicability to on-line use	Automatic sampling and dilution adaptations available, in use commercially
Advantages	Suitable for wide range of particles, exceptional range of sizes measurable in a single determination, can be used in sterile environment, rapid, easy to use, highly automated
Limitations	Small particles give high errors, results very sensitive to optical model and appropriate parameters, calibration required for accurate work, some artificial broadening of distributions, samples must be robust to dilution

particles require a much more detailed theory of light scattering, developed by Gustave Mie. The full treatment requires the refractive index of both media, and the particle absorbance. In principle, the angular distribution of light is only dependent on those factors, and the particle size, so the technique may be considered absolute. The commercial instruments combine the material constants into a single parameter, the presentation value, from which the scattering pattern for a specific distribution of particle sizes can be predicted. Comparison of the predicted pattern with the measured intensity as a function of angle provides a basis for iteration of the predictions to obtain the best estimate of the size distribution.

The commercial instruments may be used to size solid, liquid or biological particles. Some models have a facility for spraying a powder or an aerosol into the light path, but most sample chambers are designed for liquids. The cells contain a stirred dispersion of the particles, at low concentration to avoid multiple scattering effects. An axially mounted photodiode monitors the intensity of the unscattered beam to ensure that the turbidity (obscuration) remains in the range 5–40%. For particles of 1 μm, this corresponds to approximately 0.004–0.01% (v/v). The optimum concentration varies with particle size and the presentation value; for example, particles of 100 μm are measured in the range 0.1–1% (v/v). The overall obscuration value also serves as a test on the scattering model used to analyze the data. The scattered light is analyzed as a function of angle using a special lens system (Fourier transform lens) that collects the scattered light from all the particles illuminated by the beam (up to 20 mm diameter) to get a stationary scattering pattern on a set of semicircular detectors. The number of detectors determines the resolution in measured scattering angle; modern instruments typically have 256 detector channels. Figure 2 shows the principle of operation of a commercial sizer.

3.2.2 Advantages and Limitations

The advantage of variable-angle laser light scattering to measure particle size is its applicability to a very wide range of materials and size ranges. Most instruments have lenses of different focal lengths, enabling them to assess solid or liquid particles in the range 1 μm to 1 mm. Obtaining accurate data for particle sizes < 1 μm is more complex, as the light is scattered to very wide angles, which are difficult to obtain practically. Although some information can be gained from the scattering pattern at lower angle, fitting the shape of the rising edge of the primary peak, it is not reliable for real, polydisperse systems with a significant fraction of small particles. Some modern instruments utilize an additional feature to characterize particles in the range 0.1–1 μm. The technique is based on the difference in scattering from vertically and horizontally polarized beams of light, which is dependent on frequency and particle size. This improves the resolution of small particles and extends the range of the diffraction-based instruments.

The methods are suitable for highly polydisperse systems with each lens system capable of spanning typically three decades of particle diameters. Since a volume distribution is obtained, the larger particles tend to dominate and it is unwise to convert the data to number or area distributions when there is a high level of polydispersity. The method relies on established optical models, but the analysis procedure involves many steps, and experimental uncertainty in the measured data can result in erroneous predictions of the particle size distribution. It is also not always straightforward to obtain the parameters for the predictions (the presentation value), as the refractive index and absorbance of the particles may not be known or may vary within the sample. It is also possible that the presence of surfactant or other adsorbed species on the surfaces of particles may perturb the optical properties. In practice, it may be necessary

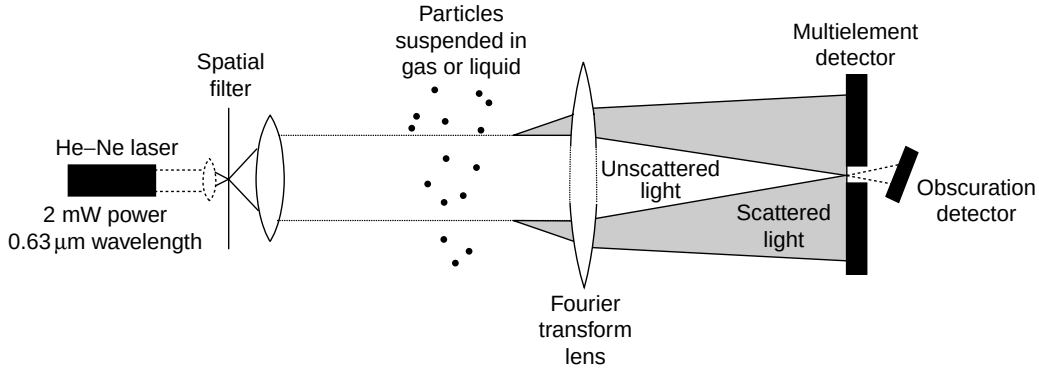

Figure 2 Principle of the variable-angle light-scattering particle sizer. (Reproduced by kind permission of Malvern Instruments Ltd. © Malvern Instruments 1999.)

to conduct preliminary experiments to determine the appropriate presentation value. These may be calibration runs with particles of known size, or a protocol where the obscuration values and fitted size distributions are examined for consistency with varying concentrations of the same sample.

The technique is applicable to dry powders, provided that appropriate sample-handling systems are used: powders dispersed in a nonswelling liquid, oil-in-water or water-in-oil emulsions, bubbles in liquid (provided they are comparatively stable) and biological systems such as yeast[22] and bacteria.[11] The samples must be stable to dilution; as a result, this may not be the system of choice for slightly soluble particles or reversible aggregates. The sample cells can usually be sterilized if necessary, and there are adaptations of the instruments to sample from a process line to obtain real-time particle monitoring.[23]

Recently, there have been modifications to the technique to enable measurements to be made on concentrated colloidal dispersions, such as undiluted milk and parenteral emulsions.[24] The instrument is based on the same principles as the commercial sizers, but the samples are enclosed in a flat cell the thickness of which is variable between 10 μm and 5 mm. The use of very thin samples reduces the likelihood of multiple scattering. The technique also works at higher turbidity, up to 70%, by incorporating the effects of multiple scattering in the algorithm that converts the scattering data into the particle size distribution. Another extension of the technique, currently under development, is to characterize the fractal dimensionality of aggregates of polydisperse particles.[25]

3.3 Dynamic Light Scattering

Particle analysis by dynamic light scattering is summarized in Table 7.

3.3.1 Description of the Technique

Dynamic light-scattering techniques[1,26,27] are based on the Stokes–Einstein relation between the unhindered diffusion coefficient D of a colloidal particle in a liquid and its diameter d (Equation 2):

$$D = \frac{kT}{3\pi\eta d} \tag{2}$$

where k is the Boltzmann constant, T the absolute temperature and η the shear viscosity of the suspending phase. For particles $< 1\,\mu m$, the diffusion is rapid enough to be readily measurable, and the diffusion coefficient provides a measure of the particle diameter. A number of particle-sizing instruments are based on this principle, using photon correlation spectroscopy to measure D. The principle is illustrated in Figure 3.

Photon correlation spectroscopy, or dynamic light scattering, is concerned with the fluctuating signals arising

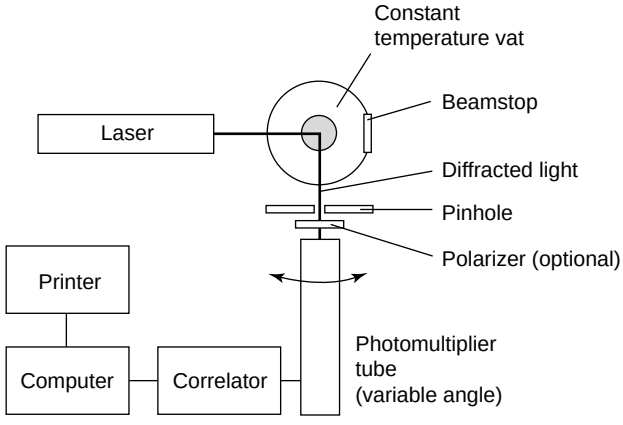

Figure 3 Principle of operation of dynamic light scattering to measure particle size. (Reproduced by kind permission of Malvern Instruments Ltd. © Malvern Instruments 1999.)

Table 7 Summary of the use of dynamic light scattering for particle size analysis

Characteristics	Comments
Operating principle	The measurement of the self-diffusion coefficient of individual particles, using photon correlation spectroscopy
Particle type	Solid or liquid particles dispersed in liquid
Property measured	Particle diameter (inversely related to self-diffusion coefficient)
Size range	3 nm to 3 μm
Concentration range	10^{18} particles m^{-3}, for example, for 0.1 μm < 0.01% (v/v)
Calibration required	Not in principle
Model/parameters required	Stokes–Einstein diffusion relation, viscosity of background liquid, temperature, particle shape
Speed of operation	Seconds (monodisperse particles) to hours (highly polydisperse samples)
Applicability to on-line use	Impractical owing to high dilution, cleanliness required and sensitivity to vibration
Advantages	Very sensitive for small particles, including molecules and micelles, absolute sizes obtained, can measure adsorbed layers of substances such as proteins
Limitations	Very sensitive to contamination and vibration, high dilution required, polydispersity difficult to quantify, results strongly influenced by largest particles present, complex and expensive

from the light scattered by single particles or molecules in a colloidal dispersion or polymer solution. Static light scattering averages the scattered light, often as a function of angle or wavelength, and obtains particle information from the overall intensity of light. In dynamic light scattering, the fluctuations (speckle pattern) in the scattered beam at a given angle provide information on the rate of particle movement owing to diffusion. There are several ways of looking at the phenomenon. One approach is to consider the fluctuations to reflect the instantaneous position of particles in the illuminated volume. As the particles move within the volume element, their position causes changes in the optical path length and, thus, in the phase of scattered light. The changes in phase are directly related to the distance moved in the sampling time. The time between successive changes in phase is a consequence of a characteristic particle diffusion rate, provided that the sampling time is much shorter than the average time between changes of direction of the particle.

Commercial instruments analyze the scattered light by means of an autocorrelation function. This function compares the intensity $I(t)$ at time t with the intensity $I(t+\tau)$ after a time delay τ. If the time delay τ is less than the characteristic diffusion time for the particle, then the two values will be correlated, i.e. both will be high or both will be low. The sum of the product of the two values of intensity for the total measurement time t provides the autocorrelation function $g(\tau)$ (Equation 3):

$$g(\tau) = \langle I(t) \times I(t+\tau) \rangle \tag{3}$$

The autocorrelation function is related to the diffusion coefficient of the particles by Equation (4):

$$g(\tau) = A + B\exp(-\Gamma\tau) \tag{4}$$

where A and B are instrument factors ($B > A$) and Γ is the decay rate, linearly related to the translational diffusion coefficient D.

Therefore, the diffusion coefficient is derived from the decay rate Γ of the autocorrelation function. If the particles are monodisperse, the analysis shows a log-linear dependence of $g(\tau)$ on τ. Polydispersity is detected by curvature in the log correlation function versus correlation time. Then a set of exponential functions must be fitted to the autocorrelation function $g(\tau)$ versus τ.

3.3.2 Advantages and Limitations

Dynamic light scattering is currently the accepted technique for small particles and polymer molecules, which are difficult to characterize by other methods. The precision is high for the estimation of the mean size and polydispersity index of narrow distributions and the measurements are comparatively rapid. However, longer correlation times are needed if there is considerable polydispersity, as the fitting of multiple exponential functions to a single measured curve requires a large number of data points. In wide distributions, another major problem is that the scattering intensity rises with d^6; consequently, large particles dominate over small ones.

Many instruments measure at a single angle – usually 90° to the incident light beam – but improved measurement time and resolution can be gained by measuring at a range of angles.[28] The optimum range, even if the intensities owing to different particles are equalized, is when the largest particles are no more than twice the diameter of the smallest ones. However the resolution can be improved by combining the dynamic measurements with angle-dependent static light scattering.[29]

Table 8 Summary of the electric sensing zone technique to measure particle size

Characteristics	Comments
Operating principle	Impedance change across aperture during passage of particle suspended in electrolyte solution
Particle type	Nonconducting, nonporous, solid, liquid or biological particles suspended in electrolyte solution. Irregular shapes
Property measured	Volume of equivalent sphere
Size range	0.3–1200 μm (using a range of apertures)
Concentration range	Highly dilute
Calibration required	One-off calibration of each aperture, transferable to any sample
Model/parameters required	None
Speed of operation	Seconds
Applicability to on-line use	Feasible in principle if dilution bypass installed and clean environment, free from electrical interference
Advantages	Direct count of individual particles engenders confidence, model-independent results, high precision and accuracy, applicable to wide range of particles
Limitations	Not suitable for highly polydisperse samples, small apertures block easily, must be diluted in electrolyte

4 ELECTRIC SENSING ZONE

Table 8 summarizes the features of the electric sensing zone technique.

4.1 Description of Technique

The electric sensing zone technique is based on the difference in dielectric properties of particles and a surrounding liquid (Figure 4).[3] It was originally developed for blood cells but is now used for a variety of solid and liquid dispersions. The particles are suspended in an electrolytic solution and passed through a small aperture. Electrodes on either side of the aperture measure the change in impedance as the particle passes through the gap. The change in impedance is directly related to the volume of electrolyte displaced by the particle, and since each particle passes individually through the aperture, each one is counted. The number of counts recorded for each size of pulse (divided into up to 256 analysis bins) gives the number distribution of particle volume. This is readily converted into the equivalent spherical diameters, volume distributions, etc. as required.

The range of the instrument is dependent on the aperture size. A commercial instrument, the Coulter Multisizer, uses apertures from 15 to 2000 μm. Particles can be sized if they fall in the range 2–60% of the aperture diameter; therefore, in principle, the instrument has a total range of 0.3–1200 μm. Calibration is required for each aperture to obtain the geometric and electrical constants that convert the pulse height into the required particle volume. Polystyrene latex size standards are used to obtain the calibration values. Provided that the sample particles are also nonconducting and spherical, the calibration is valid for any type of (nonporous) particle.

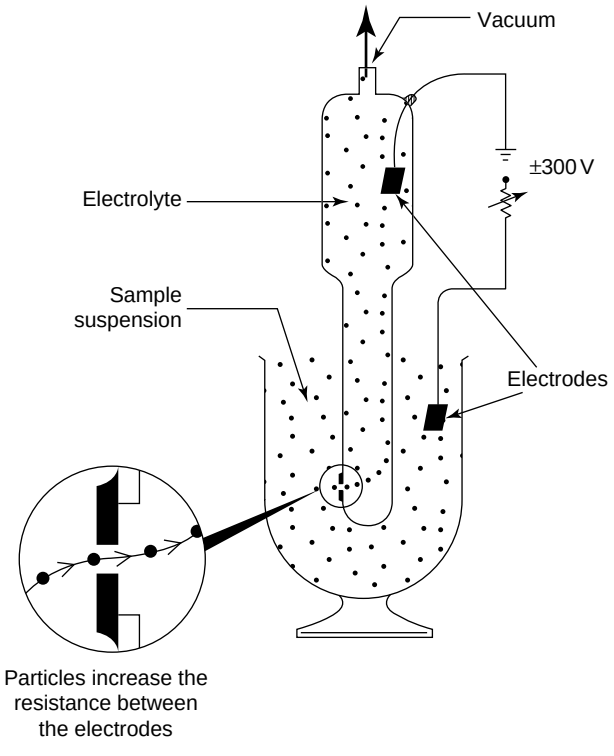

Figure 4 Principle of operation of the electric sensing zone particle-sizing technique. (Reproduced by kind permission of Beckman Coulter UK Ltd. © Beckman Coulter 1999.)

4.2 Advantages and Limitations

The method has the attraction of counting individual particles, with minimal calibration. It is suitable for a wide range of particle type, including biological systems such as blood cells and bacteria. It is excellent for particles in a comparatively narrow size range, accessible by a single aperture size. Where there is a high level of polydispersity, a full size distribution may be built up from measurements made using several apertures, but it is not always straightforward to link the individual distributions into a single data set.

Errors may occur as a result of more than one particle passing through the aperture at any one time (coincidence), but the modern instruments detect this by recording an anomalous pulse shape and discarding the data point. The dilution of the suspension should be adjusted to reduce coincidence to less than 10% of the total count. High dilution means an increase in measurement time, as typically counts are made of approximately 30 000 particles, but this is rarely a limitation in practice, with most characterizations completed within 10 s. An anomalous pulse shape also occurs for spherical particles that pass off-center through the aperture, and these data are also automatically discarded. The volume of nonspherical particles may be measured, but there is an error associated with the irregular pulse shape caused by the orientation of the particles in the aperture, and it may preclude the detection of coincident particles, thus broadening artificially the size distribution. Clearly, the measured diameters relate to the sphere of equivalent volume.

The instrument requires dispersal of the particles in an electrolyte, usually an aqueous salt solution, and this may change the size or stability of certain samples. The electrolyte must be very clean if the small apertures are used because these are prone to blocking by small aggregates of particles or contaminants. The sample must also be robust to a high level of dilution; therefore, the particles must be highly insoluble in the liquid. The technique is not suitable for porous particles or gas bubbles.

5 SEDIMENTATION

5.1 Gravitation

Gravitational sedimentation techniques are summarized in Table 9.

5.1.1 Description of the Technique

Gravitational sedimentation techniques for particle sizing are based on the application of Stokes' law to the individual particles in a (dilute) settling or creaming dispersion. If the particles are denser than the surrounding liquid, they have a tendency to sediment; less dense and the process is called creaming. The term creaming derives from the behavior of oil-in-water emulsions such as milk. The basis of Stokes' law is that particles of different density to the surrounding fluid move under gravity and against a fluid resistance until a constant, terminal velocity is attained. The gravitational force is related to the volume of the particle, and the resistance is determined by its surface area and velocity. When the two forces are equal and opposite the particle is moving at a constant (terminal) velocity, V_s, given by Equation (5):

$$V_s = \Delta\rho g \frac{d^2}{18\eta} \tag{5}$$

where $\Delta\rho$ is the difference in density between the particle and liquid, g is the acceleration from gravity, d is the particle diameter and η the viscosity of the continuous phase.

The practical issue is to measure the terminal velocity of the particles. This is normally done by assaying a dispersion for particle concentration, where changes with time are analyzed in terms of particle flux and, therefore, velocity. A large number of instruments are available to obtain this information.[3] The simplest are intrusive, where a pipette (e.g. the Andreason pipette) is used to sample a dispersion as a function of time, and sometimes vertical position, to conduct off-line concentration assays on the portion removed. A hydrometer may be suspended in a dispersion to monitor the density of the dispersion with time, from which the particle concentration may be inferred. Other devices employ a scale pan suspended in the dispersion to weigh the particles landing on it. Analysis of the signal from an optical concentration scanner is complex (see section 3.1 above) as the turbidity, usually measured at a single frequency, is a function of the concentration and particle size distribution of the particles in the beam; however, this analysis is suitable for very low concentrations. For concentrated dispersions, the concentration–height information may be collected using X-ray attenuation,[30] capacitance measurements[31] or ultrasonic density profiling.[32] In the last (Figure 5), the time-of-flight of a pulse of ultrasound is measured as it passes through the sample. This value is converted to speed of ultrasound, which is usually a simple function of dispersion concentration. Measurement at a series of heights allows detailed characterization of the sedimentation or creaming process, as shown schematically in Figure 5. The concentration profiles are converted into particle velocities and then diameters, using Equation (5).

5.1.2 Advantages and Limitations

The use of gravitational sedimentation techniques is confined to particles that have a Stokes' velocity that is significantly larger than their movement owing to diffusion. In practice, this restricts the use of the techniques to particles of the order of 1 μm and larger, up to a maximum determined by the need to maintain

Table 9 Summary of gravitational sedimentation methods for particle sizing

Characteristics	Comments
Operating principle	Rate of vertical movement of particles
Particle type	Solid or liquid particles in a liquid dispersion
Property measured	Hydrodynamic diameter distribution
Size range	1–200 μm (depending on density difference)
Concentration range	Ideally < 0.2% (v/v), but higher concentrations can be allowed for
Calibration required	Not in principle
Model/parameters required	Stokes' law, density difference, viscosity of liquid, particle shape; at higher concentrations, interaction factors between particles
Speed of operation	Minutes to days
Applicability to on-line use	Not suitable
Advantages	Nonintrusive, comparatively cheap, simple theory, can be used, with correction factors, on concentrated dispersions, obtain full hydrodynamic size distribution, detects flocculation
Limitations	Perturbed by convection, not suitable for porous particles, shape effects difficult to quantify, can be slow, not suitable for submicron particles or where densities closely matched, need Newtonian suspending liquid

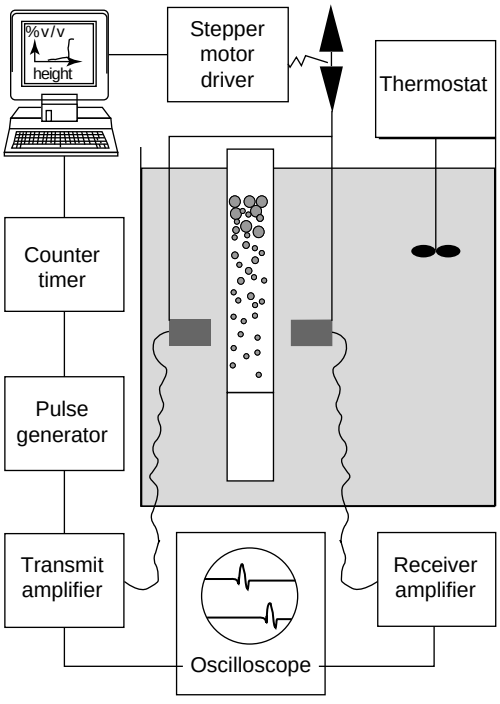

Figure 5 Ultrasonic detection of density profiles for gravitational sedimentation sizing.

laminar flow (Reynolds number < 0.25), which typically gives an upper limit in the region of 200 μm (depending on density difference and viscosity). The dispersion concentration should be low, <0.2% typically, to ensure that the particles are subject to individual movement unperturbed by each other. However, there are correction factors for dispersions of higher concentration that allow for backflow and the hindrance caused by particle interactions.[33,34] The X-ray, capacitance and ultrasonic techniques are particularly suited for the sedimentation of concentrated dispersions, which is important where the particle size may be affected by dilution or where the aim is to characterize weak aggregates. The values of particle size obtained from sedimentation techniques are always based on hydrodynamic diameters, requiring complex analysis if the particles are not uniform in composition and shape, as in the case for porous particles or flocs.

The simplicity of many of the sedimentation techniques makes them an attractive and cheap option for particle sizing in liquid dispersions. They are commonly used for powders dispersed in nonswelling liquids, and the ultrasonic method has been applied to food emulsions. In many cases, the distribution of hydrodynamic diameters is the property required, in order to predict or control separation under gravity, but care must be taken in comparing the data with other measures of diameter, particularly for irregularly shaped particles or aggregates.

5.2 Centrifugation

Centrifugal sedimentation is summarized in Table 10.

5.2.1 Description of the Technique

The principles in centrifugal particle sizing are identical to those in gravitational methods. The particles are made to move in a centrifugal field, which fractionates them according to their hydrodynamic diameter. The samples are enclosed in a disc that is spun; the concentration of particles passing a fixed diameter on the disc is monitored. The detection device is usually an optical photocell, and analysis is complicated by the factors described in section 3.1, the turbidity of the dispersion being a complex function of particle concentration and size distribution. Measuring the particle concentration using X-rays is also available in a commercial instrument,[30] which provides

Table 10 Summary of centrifugal sedimentation methods for particle sizing

Characteristics	Comments
Operating principle	Movement of particles in a centrifugal field, owing to differences in density
Particle type	Solid or liquid dispersed in a liquid
Property measured	Hydrodynamic diameter
Size range	0.05–100 μm (depending on density and viscosity)
Concentration range	<0.2% (v/v)
Calibration required	Not in principle
Model/parameters required	Stokes' law, density difference, centrifugal field, liquid viscosity, Mie scattering theory and optical properties (for optical detection)
Speed of operation	Minutes to hours
Applicability to on-line use	Not applicable
Advantages	Nonintrusive, full hydrodynamic size distribution obtained, comparatively rapid, suitable for broad range of sizes
Limitations	Low concentration needed for optical sensing, uncertainty in density of particles if adsorbed materials present (e.g. surfactant), underestimation of diameters if particles are deformable, shape factors difficult to quantify

unambiguous concentration data at the selected position on the disc.

5.2.2 Advantages and Limitations

The disc centrifuge technique is well established in the food industry, being in routine use for characterizing emulsions. It is suitable for much smaller particles than gravitational sedimentation techniques, and for larger particles it is the more rapid technique. It provides a full hydrodynamic size distribution, and the complexities of analyzing the concentration using optical scattering are avoided in the X-ray versions of the instrument. For small emulsion droplets, however, if there is a substantial amount of emulsifier on their surfaces, the overall density of the particles may not be well known, and in some cases a thick protein layer can make oil droplets denser than a surrounding aqueous phase. If the particles deform under the centrifugal forces, inaccurate results may be obtained. The dispersion must be stable to dilution, to ensure individual particle movement according to Stokes' law.

5.3 Field-flow Fractionation

The technique using FFF to analyse particle size is summarized in Table 11.

5.3.1 Description of the Technique

FFF is a variation on the sedimentation techniques.[35] The method uses a gravitational or centrifugal field to fractionate the particles into size zones, which are washed out (eluted) in turn by a fresh fluid. Solid or liquid particles may be characterized, and the method provides hydrodynamic diameters of nonspherical particles.

Centrifugal fields are most commonly used to determine the size of particles in the range 0.005–1 μm. Essentially the apparatus consists of a flattened toroid tube containing the sample. Input and exit channels allow filling and emptying. The sample cell is set spinning with the continuous phase of the dispersion, to set up a centrifugal field across the chamber. The sample [containing particles at less than 0.2% (v/v)] is injected into the field and the particles separate into regions of varying size. The centrifugal forces are balanced by the back diffusion of the particles against the concentration gradient generated by the centrifugal field. Large particles are flung to the outside of the tube, and smaller ones remain suspended. The field is controlled to achieve maximum separation of the particles in the sample. When the particles have reached their equilibrium positions in the field, a fresh sample of continuous phase is passed through the tube. Using laminar flow conditions to obtain a parabolic velocity profile, the particles at the center of the tube are collected first, followed by the larger size fractions. The concentration with elution time is monitored using optical detectors to obtain the concentration of material that was suspended at each position in the tube. This information allows a full size distribution to be accumulated, usually over a timescale of up to an hour.

5.3.2 Advantages and Limitations

The wide range of FFF instruments available cover a correspondingly wide size range. Within any one sample, the conditions can be set to fractionate to the required resolution, or to emphasize certain parts of the size distribution. Controlling the spin speed to slow down exponentially during the collection process allows the particles to emerge according to a log-linear relationship between mass and retention time.[3] The process can be controlled to be gentle if required, to characterize aggregates or unstable particles.

Table 11 Summary of field-flow fractionation technique

Characteristics	Comments
Operating principle	Classification by centrifugal or gravitational field into size bands, followed by washing out of each size in turn
Particle type	Solid or fluid particles in liquid suspension
Property measured	Hydrodynamic diameter
Size range	0.005–500 μm
Concentration range	0.1% (v/v) typically
Calibration required	Optical detection system
Model/parameters required	Stokes' law, density difference, viscosity, Mie scattering theory, refractive index, absorbance
Speed of operation	Minutes to hours
Applicability to on-line use	Not suitable
Advantages	Wide size distribution, high resolution, can collect fractions, good for aggregates, aqueous or nonaqueous dispersions
Limitations	Resolution dependent on experimental parameters, multistage analysis, fairly dilute

The analysis of the turbidity/time data to obtain the particle size distribution is multistaged, involving knowledge of the optical properties of the dispersions (and their variation with particle size), the flow profile of the eluting fluid, the centrifugal field at each point in the tube under the conditions chosen, and the sedimentation properties of the particles in the dispersing liquid. The fractionation system is never perfect, and a monodisperse sample will still result in a distribution of elution times. This zone broadening generates apparent polydispersity in the derived size distribution. Recent developments[36] are overcoming this problem by using refined detection procedures, involving multiangle light scattering to quantify the polydispersity of each fraction during elution. In this mode, the FFF instrument is operating primarily as a fractionator to prepare near-monodisperse samples for the light-scattering analysis to give the size distribution of each fraction. The advantage over merely using the light-scattering instrument is the resolution obtainable when analyzing sequential samples of limited polydispersity.

6 ACOUSTIC TECHNIQUES

6.1 Ultrasonic Spectroscopy

The principles of ultrasonic spectroscopy for analysis of particle size are described in Table 12.

6.1.1 Description of the Technique

Ultrasonic spectroscopy is based on the size-specific interaction between high-frequency sound (ultrasound) and colloidal particles in suspension.[37] In some ways the technique is analogous to the scattering of light by small particles; it is distinct in that ultrasound is a longitudinal wave and not subject to polarization, and most applications of the technique involve particles smaller than the wavelength of the ultrasound. There are a number of different mechanisms that contribute to the scattering or absorption of ultrasound by particles, which is altered by the contrast in the physical and thermal properties of the particles and the surrounding liquid.[38] Since the contrast in these properties between a liquid and a gas bubble is very high, a small amount of entrained air in a dispersion renders it effectively opaque to ultrasonic characterization.

Ultrasonic sizing measurements are usually made in the straight-through direction using a pair of transducers facing each other across the dispersion. The frequency of the ultrasound is varied, and the overall attenuation of the signal by the dispersion is measured. The attenuation of most dispersions increases strongly with frequency, usually as $(\text{frequency})^2$, so measurements at high frequency are made using a reduced path length. Either a pulse of ultrasound is used to irradiate the sample with a wide range of frequencies in a single measurement, which are analyzed separately, or the frequency is changed for each determination. In both cases, a spectrum of ultrasonic attenuation as a function of frequency is obtained, which is inverted to obtain an inferred particle size distribution. Commercial instruments operate in the frequency range 1–200 MHz and can characterize particles between 0.01 and 1000 μm. Unlike light, it is possible to propagate ultrasound through concentrated dispersions, and particle volume concentrations up to 50% are accessible, although the analysis is considerably simplified if single scattering can be assumed, which requires concentrations of a small percentage (volume per volume).[39]

6.1.2 Advantages and Limitations

The advantage of ultrasonic attenuation for particle sizing is its ability to characterize concentrated dispersions. The

Table 12 Summary of ultrasonic spectroscopy as a technique to measure particle size

Characteristics	Comments
Operating principle	The attenuation of ultrasound as a function of frequency
Particle type	Solid or liquid particles dispersed in liquid
Property measured	Volume distribution of diameters
Size range	0.01–1000 μm
Concentration range	0.5–50% (v/v)
Calibration required	Not in principle
Model/parameters required	Ultrasonic scattering model. Densities, ultrasonic properties, viscosity/modulus and thermal properties of particles and dispersing liquid
Speed of operation	Minutes, but analysis may take longer
Applicability to on-line use	Suitable for on-line applications
Advantages	Wide size range, high concentrations, nonintrusive
Limitations	Very complex analysis, perturbed by imprecision in parameters of particles and liquid, reduced precision at high particle concentrations, very sensitive to small amounts of air

ultrasonic waves propagate readily through dispersions of up to 50% (v/v), although it is not straightforward to remove the effects of multiple scattering from the analysis of the particle size. With an appropriate choice of frequency, ultrasonic instruments are applicable to solid or liquid particles in a very wide range of diameters.

In recent years, several commercial ultrasonic particle sizers have become available. Some have a requirement for a large volume of sample (several liters), which may be inconvenient for certain applications. The other major limitation to the current applications lies in the difficulty of using the complex ultrasonic scattering theory to invert data to yield a polydisperse size distribution. Even for monodispersions, full implementation of the scattering theory requires 14 parameters (particle and liquid density, ultrasonic attenuation and velocity, viscosity or modulus, thermal expansivity, specific heat capacity and thermal conductivity), many of which are not known to a high degree of precision. For mineral dispersions (e.g. silica in water), the physical properties dominate and the thermal ones are less important. For these systems, the inversion of the ultrasonic spectra to obtain particle size distribution is relatively straightforward. However, all the parameters contribute for emulsions, and the results obtained are sensitive to the values of the parameters that are assumed. Although, in principle, the technique is absolute, involving independently derived parameters, in practice it is usually necessary to conduct calibration experiments with the dispersions of interest to obtain data sets for comparison with the spectra of unknown samples. There is a detailed protocol required for robust fitting of the ultrasonic attenuation spectra to particle size distribution and concentration.[40]

The ultrasonic technique is suitable for installation in a processing line, to give real-time changes in particle size.[41,42] The ultrasonic properties are sensitive to aggregation[43,44] and currently the theories are being extended to characterize the degree of flocculation, floc size and structure using ultrasonic spectroscopy.[45,46] There is no doubt that ultrasonic techniques will be developed further and become routine for a number of particle sizing applications. The application to biological systems is problematic, as they are typically of low volume concentration and there is little physical contrast between the cells and the suspending liquid.

6.2 Electroacoustics

Electroacoustic techniques are summarized in Table 13.

6.2.1 Description of the Technique

The use of electroacoustics involves a complex series of measurements. First, an electric field is applied to a sample in a cell, pulsed at a frequency in the megahertz range. A piezoelectric transducer at one end of the cell detects the ultrasonic wave generated by the action of the pulsating electric field on the (charged) particles in suspension. The amplitude of the acoustic waves is related to the electrophoretic mobility of the particles, which in turn is related to their charge and hydrodynamic properties in the suspending liquid. The second step is the determination of the acoustic impedance of the dispersions by measuring the reflection coefficient of an ultrasonic wave propagated from the transducer into the suspension. The theory assumes a constant charge density on the surfaces of the particles and fits a monomodal size distribution (e.g. log-normal) to the electrophoretic mobility spectrum collected at a range of applied frequencies. Calibration experiments are conducted to characterize the electrical constants in the instrument and to provide a background value to allow

Table 13 Summary of application of electroacoustic techniques to particle sizing

Characteristics	Comments
Operating principle	Measurement of the amplitude of acoustic waves generated when an oscillating electric field is applied, related to the particle drag and its charge
Particle type	Solid or liquid particles in aqueous or nonaqueous liquid
Property measured	Hydrodynamic diameter, as volume distribution
Size range	0.01–10 μm
Concentration range	1–50% (v/v)
Calibration required	Yes
Model/parameters required	Electroacoustic theory, particle and liquid densities, liquid viscosity, particle volume fraction
Speed of operation	Seconds to minutes
Applicability to on-line use	Yes
Advantages	Comparatively concentrated dispersions, can be used in nonaqueous systems, nonintrusive
Limitations	Theory only developed for dilute systems, multistage analysis, unsuitable for mixed dispersions as assumes all particles have same surface charge density

for the effect of the ions in the suspending liquid. The technique is routinely used to determine the zeta potential (charge density) of particles, such as the protein-coated emulsion droplets in milk.[47] Various manufacturers of instruments claim they can cover different ranges of particle size, from casein micelles[47] of the order of 0.01 μm, to silica suspensions containing particles 10 μm in diameter.[46] The frequency range depends on the size required and the density difference between the particles and the suspending liquid; values quoted are from 0.3 to 100 MHz.

The electroacoustic theory was originally proposed for noninteracting particles, which restricted the application to low concentrations of a small percentage volume per volume, but recently extensions have been developed to model the electroacoustic properties of dispersions up to 50% (v/v).[48] An on-line application is also reported. The method is not restricted to aqueous dispersions, and it has been used to monitor the size of the water droplets in food water-in-oil emulsions such as spreads.[49]

6.2.2 Advantages and Limitations

The technique of electroacoustics is still in its infancy, and the coming years are likely to yield new developments and modifications to extend the applications of the method. It has the advantage of being nonintrusive and capable of characterizing many dispersions without dilution. In principle, it will have value in characterizing reversible aggregates, although their theoretical analysis will be challenging. The technique is most accurate when applied to particles of narrow size distribution, with a well-defined charge density and in dilute concentrations [<5% (v/v)]. The application of the theory to concentrated dispersions up to 60% (v/v) is currently restricted to particles that have a density close to that of the suspending liquid.[48]

In 1996, the use of combined acoustic attenuation spectroscopy and electroacoustic measurements has been reported,[50] This combination helps to overcome the problems of fitting electroacoustic theory to concentrated dispersions and offers potential for the increased use of acoustic techniques in colloid particle sizing.

7 NUCLEAR MAGNETIC RESONANCE

NMR can also be used to analyse particle size (Table 14).

7.1 Description of the Technique

The sample is subjected to a pulsed magnetic field at radiofrequencies. The magnitude of the field varies with position in the sample, creating a gradient. At a characteristic frequency and field strength, the nuclei in the molecules resonate and cause a small net loss in the energy of the magnetic pulse, detected by measuring the strength of an echo in the absence and presence of the molecules of interest. The position of the molecules in the magnetic field gradient affects the amplitude of the echo, and diffusion of the molecules during the echo

Table 14 Summary of the use of NMR to measure particle size

Characteristics	Comments
Operating principle	Measurement of the restricted diffusion of molecules inside fluid droplets to give relative amounts of ''bulk'' and ''surface'' molecules
Particle type	Liquid particles in solid or liquid continuous phase, biological particles, foam bubbles
Property measured	Specific surface area distribution (expressed as volume distribution of droplet diameters)
Size range	0.5–20 μm (diffusion), >10 μm (imaging)
Concentration range	1–70% (v/v)
Calibration required	Yes
Model/parameters required	Model of restricted diffusion in spherical cavities, self-diffusion coefficient of molecules in droplets, NMR parameters
Speed of operation	Minutes
Applicability to on-line use	In principle, but not implemented
Advantages	Specific measurement on droplets of interest, high concentrations, suitable for solid emulsions (spreads, etc.), aqueous or nonaqueous systems, can be used on mixed systems
Limitations	Expensive, complex model to obtain particle size, assumes mono-modal distribution (e.g. log-normal)

time causes a detectable phase shift of the echoed pulse. As a result, the distribution of diffusion rates of the molecules may be measured as a distribution of echo magnitudes and phases. If the molecules are constrained within a spherical cavity such as a droplet, those near the walls of the cavity suffer restricted diffusion compared with those in the center of the cavity. The size of the cavity determines the proportion of molecules within it that are subjected to restricted diffusion over the timescale of the experiment. The droplet size is estimated from the proportion of molecules exhibiting restricted diffusion.[51]

7.2 Advantages and Limitations

NMR apparatus is highly complex and comparatively expensive, although now benchtop instruments are available. The main advantage of the technique for droplet sizing is its ability to characterize the molecules of interest (those inside the droplets) in the presence of other components, in a liquid or a solid continuous phase and in very concentrated dispersions.[52] The method has been used to examine food emulsions, particularly water-in-oil systems such as spreads,[53] which are difficult to characterize by other methods. Similarly, since the diffusion of water inside biological cells is restricted by the cell wall, in principle NMR methods can be used to determine the size of bacteria and yeast.

The analysis of the NMR results in terms of particle size distribution requires the application of several models: those associated with the response of the nuclei to the field, the effect of diffusion in the gradient, the assumption of spherical geometry and extraction of effective droplet area from the diffusion data, and the fitting of a simple size distribution. It would be difficult to use the methods for a very skewed or multimodal distribution of sizes or to apply them to droplets that might vary in composition. The methods can be used for droplet sizes in the range 0.1–10 μm, which is convenient for many food emulsions unless severe coalescence has occurred. However, to access the full range of sizes, a number of different experimental conditions must be used, which hinders the speed of measurement and may make it less useful for rapidly changing sizes. The technique provides information on the primary droplet size and is not sensitive to flocculation. However, in principle, if the molecules of the continuous phase are subjected to restricted diffusion because the droplets increase the tortuosity of the liquid channels between them, examination of the diffusion rates of the molecules in the external phase may provide microstructural information. The technique has also been applied to examine the voids in a solid foam,[54] and is one of the few methods to be applicable to nonintrusive evaluation of the microstructure of cheese.[55]

7.3 Magnetic Resonance Imaging

In magnetic resonance imaging, the internal structure of a food is probed nonintrusively to provide a chemical map.[56] In principle, any spatial change in composition resulting in contrast in the NMR signal may be processed to form an image of the cross-section under examination. The resolution of current imaging techniques is of the order of 10 μm; consequently, the technique has considerable potential for measuring food particles, including foam bubbles.[57]

8 EVALUATION OF THE AVAILABLE TECHNIQUES FOR DIFFERING PARTICLE TYPES

8.1 Powders

It is important to know the particle size of powders to optimize processing operations such as baking, to control the bulk density when powders are packed or in storage hoppers and to prevent dust in the atmosphere. In foods requiring control of the size of solids in the finished product (e.g. sugar crystal size in chocolate), it is convenient to monitor the size before combining with other ingredients. Powders are generally in the larger size range of food particles, over 10 μm, and they are amenable to size analysis using microscopy, sieving, variable-angle light scattering and the optical obscuration technique. Of these methods, the most highly automated is the variable-angle optical method, and the size range is accessible using comparatively simple diffraction theory. A number of commercial sizers have dry powder feeders to present the sample to the measuring system. Clearly, if there is a suitable carrier liquid, the full range of liquid sizers can be used.

If the powder is of irregular shape, and this is of interest, there is little alternative to the use of microscopy and image analysis. Modern software offers a range of options for characterizing the shape parameters of irregular solids.[12]

8.2 Emulsions and Suspensions

Food emulsions and suspensions contain a very wide range of particle sizes, and the optimum instrument depends heavily on the specific sizes required. For comparatively narrow distributions, containing particles above 0.3 μm that are robust to immersion in electrolyte, there is no substitute for the electric sensing zone methods, providing direct counting of particles classified by volume. If the distribution is broader, then the variable-angle light-scattering techniques are useful; it is claimed that the commercial instruments have a

lower limit of 0.05 μm. However, the small particle fractions are difficult to characterize accurately in a broad distribution. Similarly, the centrifugal and FFF techniques are capable of characterizing broad size distributions, in the 0.1–10 μm range.

The techniques listed so far have all required the use of very dilute samples. If the dispersion is changed by dilution, the particles can be characterized using: confocal or electron microscopy; gravitational sedimentation using X-ray, capacitance or ultrasonic detection; ultrasonic attenuation; electroacoustics, or NMR techniques. NMR is particularly useful where the continuous phase is nonaqueous and/or solid, as in water-in-oil spreads. Development in the area of acoustic attenuation may provide an alternative in the future.

A comparison of different techniques[58] to characterize suspensions with a fairly narrow range of sizes (the largest being typically three times the smallest) concludes that there is good agreement between electron microscopy, variable-angle light scattering and dynamic light scattering, provided that the dynamic method uses an appropriate correction for polydispersity. Both light-scattering techniques were shown to be very sensitive to small amounts of aggregation. Comparison between variable-angle and dynamic light scattering of food emulsions of high polydispersity[59] showed good agreement in the 0.1–0.6 μm range, but, as expected, the dynamic method detected a smaller fraction of droplets < 0.1 μm in diameter. To improve the resolution of dynamic light scattering, the use of nonlinear channel time was proposed.[59] For small particles, the combined technique using dynamic light scattering to analyze samples after separation by FFF is recommended.[60]

8.3 Foams

Foams are very difficult to characterize reliably. If the matrix is solid, as in bread or cake, microscopy and image analysis offer the best means of determining distributions of bubble size and wall thickness. In the case of liquid foams, microscopy through the wall of a container is frequently used, but there is always a concern that the bubbles at the wall may not be representative of the bulk structure. Although intrusive, the fiber optic method described above[20] is an interesting development and may be used routinely in the future. Many researchers on food foams rely on indirect methods (for example overall volume) to characterize foam stability, instead of measuring bubble size directly. In the future, NMR techniques may be developed further in this area.

8.4 Biological Systems

Microscopy is routinely used to measure the size of individual bacteria and yeast, generally electron microscopy to obtain good resolution. The electric sensing zone method was developed initially for blood cells and has a loyal following among biologists. However, increasingly researchers are adapting optical techniques to obtain rapid measures of cell size and viability.[21] Variable-angle light scattering has been compared[11] with transmission electron microscopy and found to be very suitable for spore suspensions Flow cytometry[21] is now routine as a sensitive multipurpose technique to measure cell size and viability.

The main food application of biological particle sizing is that of yeast flocculation after fermentation processes. A number of techniques are suitable for this application, including the dynamic light obscuration method[18] and detecting back-scattered light on-line during industrial processing.[23]

8.5 Complex Dispersions

The term complex dispersions is used here to describe systems containing particles of irregular shape, aggregates or mixtures of different composition. If such a system is suspected, it should be confirmed by a preliminary microscopic examination if possible, before any attempt at quantitative analysis.

Nonspherical particles present problems in their detailed characterization. The only reliable method is optical microscopy, provided that the particles are larger than approximately 1 μm, robust to dilution and randomly oriented on the slide. Electron microscopy may be used for smaller particles; however, if the sample is prepared by freeze-fracturing, the correction factors for nondiametric viewing planes become very complex. Similarly, sieving provides information on the upper bound of the two smaller dimensions of a particle, and there are ways to estimate particle shape using square-mesh sieves.[13] All the indirect methods will provide some estimate of size for nonspherical particles, usually based on a sphere of equivalent volume, area or hydrodynamic diameter, but the usefulness of the results will depend on the aspect ratio, regularity, surface roughness, etc.

Aggregated particles also present problems, either as an unwanted perturbation in the determination of primary particle size or as the population of interest. A sizing determination resulting in a bimodal size distribution should always be investigated for particle aggregation. Optical techniques are particularly sensitive to a small amount of aggregation; to determine only primary particle size, it is important to ensure that the particles are well dispersed before proceeding with a measurement. If aggregates are present during sizing using the electric sensing zone method, they will appear as a large fraction in the measured distribution or simply block the aperture. The best method to characterize the

aggregates themselves depends on their robustness. The technique based on dynamic light obscuration[18] was developed principally to detect and characterize strong aggregates, such as yeast at the end of a fermentation process, but in many other cases the aggregates are disrupted by dilution. These systems need techniques where higher concentrations can be used, such as sedimentation using capacitance, X-ray or ultrasonic detection, or the emerging applications of ultrasonic and electroacoustic spectroscopy. Optical methods are restricted to the back-scattered techniques developed for on-line use. NMR methods are insensitive to aggregation, in general, as the primary size will still emerge from the analysis of restricted diffusion within the droplets. However, if the aggregates are nonmobile and larger than approximately 10 μm, magnetic resonance imaging has potential for characterizing their structure.

Microscopy and image analysis may also be used to obtain the individual size distributions of each component in a mixture, the precision depending on the ease of identification and the number of particles of each type that are analyzed. Mixed dispersions are not readily sized using optical techniques, as the optical properties are dependent on particle composition. Similarly, sedimentation methods are problematic because, even if each population of particles is characterized by a distinct range of Stokes' velocity, the technique used to determine concentration (optical, capacitance, X-ray or ultrasound) depends on the particle properties and may not be able to provide data that can be analyzed using the appropriate parameters for each subpopulation. The overall size distribution can be measured using the electric sensing zone technique, provided that identification of the particles is not required and they are nonporous, nonconducting and within a fairly narrow size range so they are all detectable using a single aperture. The NMR techniques can measure droplet size in an emulsion/suspension mixture, as the signals from the molecules in the droplets are collected independently of other components.

ACKNOWLEDGMENTS

The author thanks David Higgs of Malvern Instruments and James Beattie of Colloidal Dynamics for information regarding their instruments. She is also indebted to colleagues at the Institute of Food Research, especially David Hibberd, Peter Wilde and Brian Hills for providing assistance in specialist techniques. The UK Biotechnology and Biological Sciences Research Council supports the Institute's research on colloidal foods.

ABBREVIATIONS AND ACRONYMS

FFF	Field-flow Fractionation
NMR	Nuclear Magnetic Resonance

RELATED ARTICLES

Food **(Volume 5)**
Viscosity of Food: Measurement and Application

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Filtration in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Sieving in Particle Size Analysis • Surface Area and Pore Size Distributions • Turbidimetry in Particle Size Analysis • Ultrasonic Measurements in Particle Size Analysis • Velocimetry in Particle Size Analysis

REFERENCES

1. E. Dickinson, G. Stainsby, *Colloids in Food*, Applied Science, London, 1982.
2. P. Becher (ed.), *Encyclopedia of Emulsion Technology*, Dekker, New York, 1996.
3. T. Allen, *Particle Size Measurement*, Chapman & Hall, London, 1997.
4. O. Flint, *Food Microscopy*, Bios Scientific, London, 1994.
5. J.C.G. Blonk, H. Vanaalst, 'Confocal Scanning Light-microscopy in Food Research', *Food Res. Int.*, **26**, 297–311 (1993).
6. H.A. Apling, 'The Optical Microscope in Food Analysis', in *Developments in Food Analysis Techniques 2*, ed. R.D. King, Applied Science, London, 1980.
7. H.D. Sapirstein, 'The Imaging and Measurement of Bubbles in Bread', in *Bubbles in Food*, International Conference, UMIST, Manchester, 1998.
8. C. Guillerme, W. Loisel, D. Bertrand, Y. Poponeau, 'Study of Foam Stability by Video Image Analysis', *J. Texture Stud.*, **24**, 287–302 (1993).
9. H. Rohenkohl, R. Kohlus, 'Foaming of Ice Cream and the Time Stability of its Bubble Size Distribution' *Bubbles in Food*, International Conference, UMIST, Manchester, 1998.
10. K.G. Berger, G.W. White, 'Ice Cream', in *Food Microscopy*, ed. J.G. Vaughan, Academic Press, London, 1979.

11. R.G.K. Leuschner, A.C. Weaver, P.J. Lillford, 'Rapid Particle Size Distribution Analysis of *Bacillus* Spore Suspensions', *Colloids Surf. B*, **13**, 47–57 (1999).
12. M.S. Rahman, 'Physical Meaning and Interpretation of Fractal Dimensions of Fine Particles Measured by Different Methods', *J. Food Eng.*, **32**, 447–456 (1997).
13. M. Whiteman, K. Ridgway, 'A Proposed Method for Sorting Particles According to their Length, Width and Thickness', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, London, 1987.
14. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, Academic Press, London, 1969.
15. L.P. Bayvel, P. Eisenklam, M. Rymons, R. Wood, 'Particle Size Measurement in the Sub-micron Size Range by Light Extinction', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, London, 1987.
16. D.R. Cameron, M.E. Weber, E.S. Idziak, R.J. Neufeld, D.G. Cooper, 'Determination of Interfacial Areas in Emulsions using Turbidometric and Droplet Size Data: Correction of the Formula for Emulsifying Activity Index', *J. Agric. Food Chem.*, **39**, 655–659 (1991).
17. R. Davies, H.E. Turner, 'Recent Progress in Rapid Response and On-line Methods for Particle Size Analysis', in *Particle Size Analysis 1977*, ed. M.J. Groves, Heydon, London, 1978.
18. J. Gregory, 'Turbidity Fluctuations in Flowing Suspensions', *J. Colloid Interface Sci.*, **105**, 357–371 (1985).
19. R.A. Williams, S.J. Peng, 'On-line Measurement and Control of Particle Production and Handling Processes', in *Controlled Particle, Droplet and Bubble Formation*, ed. D.J. Wedlock, Butterworth, Oxford, 1994.
20. C.G.J. Bisperink, J.C. Akkerman, A. Prins, A.D. Ronteltap, 'A Moving Optical Fiber Technique for Structure Analysis of Heterogeneous Products: Application to the Determination of Bubble-Size Distribution in Liquid Foams', *Food Struct.*, **11**, 101–108 (1992).
21. A.C. Pinder, S. Gatley, 'Fluorescence Cytometry for the Rapid Analysis of Food Microorganisms', in *Food Process Monitoring Systems*, eds. A.C. Pinder, G. Godfrey, Blackie, Glasgow, 1993.
22. W.R. Bowen, T.J. Ventham, 'Aspects of Yeast Flocculation – Size Distribution and Zeta-Potential', *J. Inst. Brewing*, **100**, 167–172 (1994).
23. E.K. Hobbel, R. Davies, F.W. Rennie, T. Allen, L.E. Butler, E.R. Waters, J.T. Smith, R. Sylvester, 'Modern Methods of On-line Size Analysis for Particulate Process Streams', *Part. Part. Syst. Charact.*, **8**, 29–34 (1991).
24. D. Lehner, G. Kellner, H. Schnablegger, O. Glatter, 'Static Light Scattering on Dense Colloidal Systems: New Instrumentation and Experimental Results', *J. Colloid Interface Sci.*, **201**, 34–47 (1998).
25. G. Bushell, R. Amal, J. Raper, 'The Effect of Polydispersity in Primary Particle Size on Measurement of the Fractal Dimension of Aggregates', *Part. Part. Syst. Charact.*, **15**, 3–8 (1998).
26. N. Ostrowsky, 'Particle Characterisation by Photon Correlation Spectroscopy', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, London, 1987.
27. B.B. Weiner, 'Particle Sizing using Photon Correlation Spectroscopy', in *Modern Methods of Particle Size Analysis*, ed. H.G. Barth, Wiley, New York, 1984.
28. D.S. Horne, 'Determination of the Size Distribution of Bovine Casein Micelles using Photon Correlation Spectroscopy', *J. Colloid Interface Sci.*, **98**, 537–548 (1984).
29. S.E. Bott, 'Enhanced Resolution Particle Size Distributions by Multiple Angle Photon Correlation Spectroscopy', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, London, 1987.
30. T. Allen, L. Svarovsky, 'A New X-ray Sedimentometer', *J. Phys. E.*, **3**, 458–460 (1970).
31. T.M. Shi, S.J.R. Simons, F.J. Dickin, R.A. Williams, 'Electrical Sensing of Dispersion Behaviour', *Colloids Surf. A*, **77**, 9–27 (1993).
32. A.M. Howe, A.R. Mackie, M.M. Robins, 'Technique to Measure Emulsion Creaming by Velocity of Ultrasound', *J. Disp. Sci. Tech.*, **7**, 231–243 (1986).
33. J.F. Richardson, W.N. Zaki, 'Sedimentation and Fluidization I', *Trans. Inst. Chem. Eng.*, **32**, 35 (1954).
34. D.J. Hibberd, 'Use of Ultrasound to Characterise Polymer Induced Flocculation', PhD Thesis, University of East Anglia, Norwich, 1997.
35. K.D. Caldwell, 'Field-flow Fractionation of Particles', in *Modern Methods of Particle Size Analysis*, ed. H.G. Barth, Wiley, New York, 1984.
36. M.R. Schure, 'Advances in the Theory of Particle Size Distributions by Field-flow Fractionation: Outlet and Apparent Polydispersity at Constant Field', *J. Chromatogr. A*, **831**, 89–104 (1999).
37. M.J.W. Povey, *Ultrasonic Techniques for Fluids Characterization*, Academic Press, London, 1997.
38. J.R. Allegra, S.A. Hawley, 'Attenuation of Sound in Suspensions and Emulsions: Theory and Experiment', *J. Acoust. Soc. Am.*, **51**, 1545–1564 (1972).
39. A.K. Hipp, G. Storti, M. Morbidelli, 'On Multiple-particle Effects in the Acoustic Characterization of Colloidal Dispersions', *J. Phys., D: Appl. Phys.*, **32**, 568–576 (1999).
40. A.K. Hipp, G. Storti, M. Morbidelli, 'Particle Sizing in Colloidal Dispersions by Ultrasound. Model Calibration and Sensitivity Analysis', *Langmuir*, **15**, 2338–2345 (1999).
41. U. Riebel, F. Loffler, 'On-line Measurement of Particle Size Distribution and Particle Concentration in Suspensions by Ultrasonic Spectroscopy', *Chem. Eng. Technol.*, **12**, 433–438 (1989).
42. H.P. Pendse, A. Sharma, 'Particle Size Distribution Analysis of Industrial Colloidal Slurries using Ultrasonic Spectroscopy', *Part. Part. Syst. Charact.*, **10**, 229–233 (1993).

43. D.J. Hibberd, A.K. Holmes, M.J. Garrood, A. Fillery-Travis, M.M. Robins, R.E. Challis, 'Ultrasonic Monitoring of Oil-in-water Emulsions Undergoing Depletion Floccuation', *J. Colloid Interface Sci.*, **193**, 77–87 (1997).
44. P. Kippax, J.D. Sherwood, D.J. McClements, 'Ultrasonic Spectroscopy Study of Globule Aggregation in Parental Fat Emulsions Containing Calcium Chloride', *Langmuir*, **15**, 1673–1678 (1999).
45. Y. Hemar, N. Herrmann, P. Lemarchal, R. Hocquart, F. Lequeux, 'Effective Medium Model for the Ultrasonic Attenuation due to the Thermo-elastic Effect in Concentrated Emulsions', *J. Phys.*, **7**, 637–647 (1997).
46. A.S. Dukhin, P.J. Goetz, 'Characterization of Aggregation Phenomena by means of Acoustic and Electroacoustic Spectroscopy', *Colloid Surf., A*, **144**, 49–58 (1998).
47. T. Wade, J.K. Beattie, 'Electroacoustic Determination of Size and Zeta Potential of Fat Globules in Milk and Cream Emulsions', *Colloid, Surf., B.*, **10**, 73–85 (1997).
48. R.W. O'Brien, T.A. Wade, M.L. Carasso, R.J. Hunter, W.N. Rowlands, J.K. Beattie, 'Electroacoustic Determination of Droplet Size and Zeta Potential in Concentrated Emulsions', in *Particle Size Distribution III: Assessment and Characterization*, ed. T. Provider, Chemical Society Symposium Series 693, American Chemical Society, Washington, DC, 311–319, 1998.
49. R.J. Hunter, 'The Electroacoustic Characterization of Colloidal Suspensions', in *Handbook on Ultrasonic and Dielectric Characterization Techniques for Suspended Particulates*, eds. V.A. Hackley, J. Texter, American Ceramic Society, Westerville, OH, 1998.
50. A.S. Dukhin, P.J. Goetz, 'Acoustic and Electroacoustic Spectroscopy', *Langmuir*, **12**, 4334–4344 (1996).
51. K.J. Packer, C. Rees, 'Pulsed NMR Studies of Restricted Diffusion: 1. Droplet Size Distribution in Emulsions', *J. Colloid Interface Sci.*, **10**, 206–218 (1972).
52. I. Lönnqvist, A. Khan, O. Soderman, 'Characterization of Emulsions by NMR Methods', *J. Colloid Interface Sci.*, **144**, 401–411 (1991).
53. I. Fourel, J.P. Guillement, D. le Botlan, 'Determination of Water Droplet Size Distributions by Low Resolution PFG–NMR', *J. Colloid Interface Sci.*, **169**, 119–124 (1995).
54. P.C. Griffiths, C.N. Tipples, P. Stilbs, B.H. McCormick, 'Restricted Diffusion of a ''Gas'' in a Solid Foam', *Langmuir*, **14**, 6603–6605 (1998).
55. P.T. Callaghan, K.W. Jolley, R.S. Humphrey, 'Diffusion of Fat and Water in Cheese as Studed by Pulsed Field Gradient NMR', *J. Colloid Interface Sci.*, **93**, 521–529 (1983).
56. B.P. Hills, *Magnetic Resonance Imaging in Food Science*, Wiley, New York, 1998.
57. J.B. German, M.J. McCarthy, 'Stability of Aqueous Foams: Analysis using MRI', *J. Agric. Food Chem.*, **37**, 1321, 1324 (1989).
58. A. Delgado, E. Matijevic, 'Particle Size Distribution of Inorganic Colloidal Dispersions: a Comparison of Different Techniques', *Part. Part. Syst. Charact.*, **8**, 128–135 (1991).
59. D.G. Dalgleish, S.J. West, F.R. Hallet, 'The Characterization of Small Emulsion Droplets made from Milk Proteins and Triglyceride Oil', *Colloid Surf., A*, 123–124, 145–153 (1997).
60. B. Scarlett, H. Merkus, Y. Mori, 'An Evaluation of the Sedimentation Field Flow Fractionation Technique', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, London, 1987.

Pesticides, Mycotoxins and Residues Analysis in Food

M. Volante
Presidio Multizonale di Igiene e Prevenzione, Unità Operativa Chimica, Azienda Sanitaria Locale Como, Italy

P. Branca
Agenzia Regionale Protezione Ambientale Piemonte, Dipartimento di Torino, Area Fitofarmaci, Italy

1 Introduction **1**
1.1 Sampling 1
2 Pesticides **2**
2.1 Problem of Pesticide Residues in Food 2
2.2 Sample Homogenization 2
2.3 Multiresidue Methods 2
2.4 Special Methods for Specific Classes 13
3 Mycotoxins **14**
3.1 Problem of Mycotoxins in Food 14
3.2 Different Types of Mycotoxin 14
3.3 Methods of Analysis 17
4 Quality Control **21**
4.1 Validation of Method and Internal Laboratory Quality Control 21
4.2 Interlaboratory Quality Control 22
Abbreviations and Acronyms **22**
Related Articles **23**
References **24**

This article will deal with methods for analyzing residues of pesticides and mycotoxins from molds in food. A general description of each topic is given directing the reader towards some key reviews, books or manuals published in the field with specific references added for more recent papers or topics of particular interest; techniques such as chromatography, spectrophotometry and immunochemical methods have been reviewed. In the first section sampling procedures are discussed, in the second section an overall description of the analytical methods used for pesticide residues in food is given with particular attention to multiresidue methods, in general those used most for assessing residue levels in crops at the international level, with consideration given to those less used methods dedicated to specific compound classes not covered elsewhere. The third section deals with mycotoxins with a brief description of the various mycotoxin types and some properties, followed by some discussion of analytical methods. The description of analytical techniques has been divided into extraction, cleanup and instrumental determination techniques for both pesticides and mycotoxins. Some tables of experimental data are also given.

The last section (four) deals with quality assurance aspects outlining the necessary steps, included interlaboratory trials, needed to validate a method and to assure reliability of the data.

1 INTRODUCTION

Since the discovery of the insecticidal properties of DDT (1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane) in 1939 by P. Muller more and more molecules have been studied, synthesized and widely used to protect crops from pests. In the 1960s an american biologist, R. Carson highlighted in her book ''*The Silent Spring*'' the dangers to human health and environment related to the use of pesticides.[1] The highly persistent organochlorine (OC) pesticides were found to accumulate in the fatty tissues of humans and animals, therefore in the 1970s they were banned in USA and Europe. Many alternative molecules (more selective and labile) were studied and produced so that thousands of pesticides are now in use and available.

Some of these OC compounds continue however to be used in other countries for pest control in sanitary use with a rise in the kind of problems[2,3] experienced in USA and Europe in the past.

On the other hand avoidance of all pesticide use may not always lead to a high quality agricultural product, e.g. ''wormed'' olives may have a noticeable content of oxidized oil with the undesirable products of lipid peroxidation and molds growing on stored cereals or other commodities produce toxins that could survive both food processing or an animal's metabolic activity and could be present in everyday food like beer, milk, coffee.

1.1 Sampling

In the analysis of pesticide residues and mycotoxins, the heterogeneous residue distribution among several single vegetables or fruit composing a given lot may be due, e.g. to the fact that some are more exposed and some are less exposed during spraying in the field of pesticides, or for mycotoxins, to variations in conditions affecting mold growth in different parts of the stored commodity. Usually the size of a sample collected for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

pesticide analysis is of the order of a few kilograms of vegetables. Obviously this problem may be of less importance for those commodities that are already in a homogeneous form (e.g. wine, juices, milk). The Codex Alimentarius[4] gives detailed guidelines for sampling: the number of primary samples must be related to the size of the lot and the minimum sample sizes recommended vary from 2 kg for large vegetables, to 1 kg for vegetables and to 0.5 kg (generally) for fatty products. In particular a minimum of 3 (for lots > 50 kg) to 15 primary samples (for lots over 2000 kg) and between 6 and 30 primary samples for lots suspected of having history of contamination is recommended.

For mycotoxins other problems arise from the very low levels of the residues investigated[5] and from the very heterogeneous distribution. The variations between pesticide residues is that a single piece in a lot may be strongly contaminated with values up to a tenth of a microgram per gram, caused by molds developed there. For these reasons correct sampling must encompass large amounts of each commodity in order to have at least one contaminated piece and a representative sample from the lot; furthermore each sample must be well homogenized to achieve representative subsamples for laboratory analysis. Literature reports sample sizes of 5–30 kg for the analysis of mycotoxins in different size lots.[6,7]

A review of suitable methods for correct sampling in food analysis was published by Horwitz.[8] Studies[9] also point out the increase in the error when decreasing the amount of sample collected.

2 PESTICIDES

2.1 Problem of Pesticide Residues in Food

All pesticides to a greater or lesser extent are toxic molecules, some exhibiting very high toxic activities for humans as well as for insects, therefore after they have performed their function and protected the plant against pests, they should have as little residue as possible. For the reasons mentioned in the introduction, several monitoring programs were started and are still fully active in countries worldwide as a technical tool for assessing residue levels. Apart from these, countries have also implemented policies to enable the best compromise to be found between protection of crops (of particular relevance in developing countries) and protection of the environment and human health from pollution by harmful substances.

Many studies and analytical methods for pesticide residues analysis are present in the literature, but the main problem for chemists working in this area is how to determine reliable data from hundreds of molecules[10] that exhibit completely different properties and behaviors. The residues are in fact present at trace levels (from 0.001 to nearly 10 mg kg^{-1}) in matrices that make analysis particularly difficult with the need for complex preparative procedures and sophisticated instrumental analysis used in conjunction with different selective detectors.

Some general reviews have been published[11–14] concerning methods for determining residues in plants and vegetables. Other references in this field are manuals like the FDA (Food and Drug Administration, USA) *Pesticide Analytical Manual*,[15] the *Analytical Methods for Residues of Pesticides in Foodstuffs*,[16] and other manuals[17–22] or series of books such as those by Zweigg[23] and Sherma and Cairns.[24,25] Detailed reviews,[26–30] covering only limited topics, have been published.

2.2 Sample Homogenization

Sample homogenization is the first preparative step: all the vegetables composing each single sample must be homogenized by a mechanical blender. In this operation all those parts that cannot be homogenized (e.g. core of the fruits) or are of minimal relevance to the presence of residues (e.g. nonedible parts) are normally excluded. Pesticide manuals also give details of this aspect of sample preparation. Care must be taken when isolating a portion for dithiocarbamates analysis (by cutting away some pieces for a representative subsample) before homogenization since some dithiocarbamates can degrade in the mechanical milling process. About 0.5–1 kg of homogenized sample is enough[31] to ensure representative data.

2.3 Multiresidue Methods

A considerable problem in pesticide residues analysis is that the analyst is engaged to investigate the 'potential' presence of several hundreds of active molecules in each commodity. To do this, the analyst needs methods that can simultaneously detect and quantify the largest number of active molecules possible. This goal is, in theory, very difficult (if not impossible) to achieve but methods are available in pesticide analytical manuals and elsewhere in the literature that can determinate up to 100–200 or more pesticides at the same time. These are the so-called *multiresidue methods* where the preparative steps are the same for all the molecules and the instrumental determination is made on different types of instrument; usually most molecules are determined by gas chromatography (GC) with appropriate detectors and a minimum number of pesticides determined by liquid

chromatography. Methods describing the development of single multiresidue instrumental detection by GC/mass spectrometry (MS) systems have been published.[60]

Two volumes of the FDA *Pesticide Analytical Manual*[15] are dedicated to multiresidue methods; a review of official multiresidue methods was published by Motohashi et al.[14]

Multiresidue methods cannot analyze all the active molecules at the same time; some classes (e.g. phenoxy acid herbicides, basic benzimidazole fungicides or urea herbicides) require their own methods. Some advances have been reported in efforts to include part of these classes in a general multiresidue method using specially tailored sample processing or new extraction techniques.[32,33]

2.3.1 Sample Preparation

Sample preparation normally consists of an extraction step and a cleanup step; several different techniques may be combined together to provide a much larger number of methods as can be seen from the abundant literature in this field. The first main division can be approximately drawn between the methods for nonfatty and fatty foods. Some authors[15,34] classified matrices into groups that can undergo the same extraction technique, e.g. high/low water content, high/low fat content, high/low sugar content, dry commodities, high/low acidity fruits. Acetone or a similar water-soluble solvent is suggested for high-water content samples whereas dichloromethane is suggested for dry commodities.

2.3.1.1 Techniques for Extraction The homogenized sample is normally extracted by contact with an organic solvent. In this step, because many pesticides are also present in the inner tissues of the vegetable (e.g. systemic compounds) full contact between the solvent and the sample is important for a quantitative extraction.

For aqueous matrices or vegetables with high-water content, solvents like acetone, acetonitrile and methanol are widely used, alone or in mixtures. Extraction is often aided by ultrasound or mechanical stirring with blenders or high-speed laboratory homogenizers to enable the solvent to penetrate the vegetable tissues deeply. Addition of a small quantity of diatomaceous earth helps the successive filtration. The extract obtained contains water and needs to be purified. In some methods, some water must be added to dry matrices like cereals before the solvent extraction.

The aqueous–organic extract then is extracted by partition with a nonmiscible solvent like the petroleum ethers, dichloromethane or chloroform or is diluted with water for cleanup by reverse phase solid-phase extraction (SPE). Aqueous samples like wine and juices are also extracted and purified directly onto reverse phase SPE C_{18} (octadecyl) cartridges.[35–38] SPE limits the use of solvents and can also be automated, but exhibit difficulties in recovering most polar compounds; these latter, however, can be better recovered using high-surface area C_{18} SPE substrates.[39]

Association of Official Analytical Chemists (AOAC) methods[15] use acetonitrile for the extraction with some variants in relation to the sugar content of the matrix. Honey is usually extracted by water dissolution and direct extraction both with LLP (liquid–liquid partition) using e.g. dichloromethane,[40] and hexane/isopropanol,[41] and with SPE onto C_{18} functionalized silica[42] columns.

For fatty matrices many methods[15] extract the fat and the residue values are given in milligrams per kilogram of fat; several variations may exist with respect to all types of fatty commodities, milk, meat, fish, eggs, cheese, and so on. Commodities with low fat content (less than 10%) may also be processed with methods for nonfatty matrices. AOAC[15] give a method for fatty matrices extracting fat and residues by partition in petroleum ether–acetonitrile–water.

Another technique of extraction was developed[43] using solid-phase matrix dispersion (SPMD) onto a solid phase (Florisil®, diatomaceous earth, sodium sulfate) which provides some elimination of the water and a large exchange interface for extraction with a water nonsoluble solvent. The solvents with the best extracting power are chloroform, dichloromethane, dichloromethane–acetone 9 : 1 and ethyl acetate, the former also being characterized by low boiling temperatures for the successive concentration step. Recently similar methods were developed with less harmful solvents[44,45] such as ethyl acetate, but with some problems due to its higher boiling point.

Some classes of pesticides, because of their acid or base properties require appropriate methods of extraction. Among these the most important are benzimidazolic fungicides, phenolic and urea herbicides.

Basic benzimidazole (carbendazim, thiophanate methyl, benomyl and thiabendazole) compounds are extracted by partition with an acidified aqueous or aqueous/methanolic phase and an organic phase (normally ethyl acetate or dichloromethane); in a second step the acid aqueous phase is alkalinized by means of sodium hydroxide and reextracted with ethyl acetate or dichloromethane. This extract can undergo some further cleanup.[15] Some extraction methods[46] operate with ethyl acetate or dichloromethane in the presence of a basic medium (or other catalysts, e.g. copper acetate) which catalyzes transformation of benomyl and thiophanate methyl into carbendazim, before the acid/base partition. A study[47] of extraction of benzimidazole compound with hot pressurized water has been published.

Acid herbicides (chlorophenoxy acids and chlorophenols) are extracted in methanol acidified with sulfuric acid and with successive partition with methylene chloride, petroleum ether or other suitable solvents.[15]

Urea herbicides[15] are extracted with methanol whereas *N*-methylcarbamates[15] are extracted with methanol and partitioned with acetonitrile, then in both cases partitioning is with petroleum ether (removal of nonpolar coextractives) and dichloromethane. In some multiresidue methods it has been shown[43] that urea compounds and *N*-methylcarbamates are also extracted.

Developments in some methods allow harmful solvents like acetonitrile or methanol and dichloromethane to be substituted with acetone and ethyl acetate, respectively.

In a single more sophisticated experiment some new methods achieve the extraction of phenolic and the urea compounds together with the other molecules normally detected by classical multiresidue methods[32] by modifying the pH value during the extraction.

Supercritical fluid extraction (SFE) is a new technique[48] that operates by extracting a sample with liquid supercritical carbon dioxide instead of using a solvent. This technique, however, requires sophisticated apparatus and, at present, is not yet widely used in laboratories.

General reviews of applications of SFE in extracting contaminants from food[49] or environmental samples[50] are available, together with some studies for the optimization of the experimental conditions for extraction of pesticides.[51,52] SFE can be performed either statically or dynamically (with carbon dioxide flowing through the sample), with several parameters having an influence on the recoveries. The operating pressure of the carbon dioxide must be above a given critical pressure where the analyte tends to increase its solubility in the supercritical extractant until it reaches a maximum, and the temperature must be optimized because an increase in temperature may or may not enhance the solubility of the analyte. These two parameters depend upon the properties of the molecules investigated. Often a modifier is added to a supercritical fluid in order to increase the polarity of the medium and achieve the extraction of polar pesticides. The water content of the sample also affects the extraction. In fatty matrices good selectivity for pesticides with respect to fat may be obtained by selecting appropriate conditions of temperature and pressure of the supercritical fluid; too high a pressure would start extraction of substances from the matrix as well and lead to a too complex chromatogram. Conditions for extracting pesticides are, for instance, pressures ranging from 5000–10 000 psi (35–70 MPa) and temperatures between 40–80 °C. A limitation of SFE apparatus is the relatively small sample extraction chamber (10–15 mL) which does not allow extraction of more than 5 g of sample dispersed in a solid phase. Another step needing control in SFE is the final trapping of the analytes onto a sorbent substrate and the consequent elution with a solvent. Lehotay and Valverde Garcia[53] studied a series of trap–solvent combinations, after extraction from fruits and vegetables carried out dynamically at 50 °C and 350 bar; the best combination found was an octadecyl–silica trap with acetone as eluting solvent.

Accelerated solvent extraction (ASE)[45,54] extracts with solvent at high pressure and temperature leading to recoveries mostly falling in the 70–110% range with standard deviations between 5 and 11%. It represents an alternative to Soxhlet apparatus.

2.3.1.2 Techniques for Cleanup of the Extract A review[55] of cleanup methods was published by Tekel and Hatrik, however similar information is often included in other reviews.

Many extraction procedures already operate as a first clean up as the analytes are preferentially extracted more than many other substances contained in the matrix (matrix coextractive). This could be the case for traditional LLP or the more recent reversed phase octadecyl silica SPE and SPMD that were discussed earlier.

A traditional technique widely used to clean up the extracts is preparative chromatography onto silica, alumina, Florisil®, or Celite® columns which are devoted to retaining coextracted substances from the matrix. Analytes are usually eluted from the column with pure solvents or solvent mixtures effecting a compromise between elution of a wide range of compounds, recovering the most polar active molecules and limited coelution of the interferents from the matrix. Solvents normally used in mixtures are hexane, dichloromethane, petroleum ether, or acetonitrile. Florisil® is used most and is particularly good in retaining fats; dedicated columns are needed for special cases, e.g. silver added to alumina to detect sulfur compounds present in cabbage and celery.

These cleanup procedures were developed using 6–3 mL/1–0.5-g mini columns which like those used for SPE are readily available for commercial use. Some studies have been published concerning stability of adsorbed compounds on SPE columns.[37]

SPE columns are also used with other stationary phases: graphitized carbon black as a completely nonpolar substrate and the polar aminopropyl column. DeKok[56] and others[57] used amino columns for cleaning up carbamates prior to HPLC (high-performance liquid chromatography) analysis; for amino columns polar solvents such as acetonitrile or methanol–dichloromethane are necessary for eluting polar molecules, such as captan, acephate and vamidothion. Bicchi et al.[58] and Di Muccio et al.[59] achieved cleanup of acetonic or methanolic

extracts on Extrelut® prepacked columns and eluting with dichloromethane. Other methods were also reported using SPE columns in succession[60–62] with different sorbent phases.

A method[63] developed for fatty matrices combines a first cleanup by centrifuging at −10 °C in the extraction step and a successive cleanup on Florisil® cartridges.

Among the instrumental techniques for purifying the samples are gel permeation chromatography (GPC) and sweep codistillation (SCOD).

In GPC[64–66] separation of the analytes onto a styrene–divinylbenzene cross-linked packed column is achieved on the basis of molecular size. Eluents used are normally dichloromethane–cyclohexane 1 : 1 and cyclohexane–ethyl acetate 1 : 1; fatty matrices are purified well with a 85 : 15 cyclohexane–dichloromethane solvent mixture. GPC does not always provide a satisfactory purification with respect to the costs of instrument and solvents; while fatty matrices, especially those with fat content approaching to 80–100%, are purified satisfactorily, in vegetable extracts chlorophyll is not completely eliminated. Columns are normally about 30 cm long with 2-cm diameter but some methods use a smaller diameter (1 cm) columns for a minor solvent consumption.

SCOD is a completely different technique[16,17] based upon the relatively high volatility of pesticides with respect to matrix components. The sample is put into a heating device filled with silanized glass wool where an inert gas is flowing through, the analytes are distilled and transported towards a cold trap where they condense and can be recovered for successive instrumental detection.

2.3.1.3 Concentration of Sample Extract The sample extracted (usually 100–200 mL) must be concentrated either for cleanup procedures or for instrumental analysis. Usually, the final purified extract is concentrated to a few milliliters to obtain 10 g of extracted vegetable in 1 mL with a ten-fold concentration factor. This concentration factor is much less than in water analysis where up to 1000-fold concentrations are usually achieved, although in the case of foods the limiting factors of concentration are coextracted substances from the matrix. It must also be considered that the injection of too concentrated extracts shortens the life of the chromatographic columns.

In the Rotavapor®, the extract is rotated in a flask at temperatures between 30–50 °C under reduced pressure, and the solvent is distilled. Extract must be only reduced to a small volume and the solvent evaporation must be completed under a gentle stream of nitrogen in order to avoid losses of the more volatile pesticides (e.g. dichlorvos, methamidophos, biphenyl, diphenylamine, dimethylaniline metabolites, acephate).

In the Kuderna–Danish apparatus the solvent evaporated from the extract by heating flows first through an expansion chamber and then through a Snyder column where the heavier fraction (which may include most volatile pesticides) condenses back; the solvent may be recovered by condensation into an apposite vessel. Some Kuderna–Danish apparatus devices achieve solvent evaporation by inert gas (nitrogen) flowing instead of heating.

2.3.1.4 Solid-phase Microextraction Solid-phase microextraction (SPME) is based upon direct adsorption of the analytes from aqueous solutions onto the organic phase (polidimethylsiloxane, carbowax, divinylbenzene, methacrylate are used) coating a glass fiber. The adsorbed analytes are then directly desorbed in the injector of a gas chromatograph or in the loop (by means of a special device) valve of a HPLC.

This technique is particularly interesting because it avoids the use of solvent, simplifies the preparative step and good results were obtained in analyzing pollutant in water,[67,68] as reviewed by Eisert and Levsen.[67] This advantage is partially counterbalanced by the fact that each gas chromatographic run needs its extraction step and the calibration requires extraction for each point (in this case the concept of ''recovery'' is meaningless).

The first experimental applications of SPME to pesticide residue analysis on food concerned mainly wine and juices,[69,70] and vegetables.[71] The technique seems applicable for the determination of many active molecules. Honey[71] is a particularly interesting water-soluble matrix for such an application but the high content of sugar present interferes by dirtying the fiber and causing consequent losses in sensitivity.

2.3.2 Instrumental Determination

2.3.2.1 Gas Chromatographic Techniques GC is the main technique devoted to the pesticide residue analysis due to its high power of resolution of complex mixtures and its high sensitivity.

Splitless or cold on-column injection are the two most used injection techniques in this analysis because pesticide residues are present in the sample at trace levels. Splitless injection preserves the clean column but introduces the sample directly into a high temperature chamber whereas cold on-column injection is undoubtedly preferred in the case of labile compounds (e.g. *N*-methylcarbamates), although introducing the whole extract into the column leads to a shortening of the life of the column.

There is no requirement for a specific column for analyzing pesticide residues; usually general purpose columns with nonpolar or slightly medium polar stationary phases

are used (100% methylsilicone, or methylsilicone containing 5–50% phenylsilicone or 6–14% cyanopropylphenylsilicone). Carbowax columns are limited because they cannot reach temperatures above 220 °C without bleeding, while several pesticides and several matrix substances elute at higher temperatures. Special columns also exist devoted to optimizing the separation for determinate classes of compounds (e.g. triazine herbicides).

Despite the fact that many different compounds with different vapor pressures, boiling points and polarity properties have to be analyzed and then separated, programmed oven temperature runs from 50–100 °C to 250–280 °C are usually used. Solvent focusing allows good separation of many compounds with temperature programs increasing from 40–50 °C up to 150 °C at 10–20 °C min^{-1} then slowly at 250–280 °C at 2–6 °C min^{-1} (data are indicative); a further increase in oven temperature cleans the column from the heavier coextractives (e.g. chlorophyll, fats).

Detection must be both selective and sensitive. Flame ionization detector (FID) is rarely used. Electron capture detectors (ECDs) provide good detection (down to near 0.01 μg mL^{-1} of sample extract) for OC residues and many other molecules. Herbicides and fungicides that possess halogen atoms on the molecules or have high electron affinity centers like double bonds, aromatic rings, carbonyl or nitro groups can be detected by ECD with a limit of detection (LOD) from 0.05–1 μg mL^{-1} of sample extract depending upon the molecule. Pyrethroid insecticides and elemental sulfur are detectable at the ECD. Nitrogen phosphorus detectors (NPDs) assure good detection for all the organophosphates pesticides (LOD = 0.01–0.05 μg mL^{-1} of sample extract) and organonitrogen herbicides like triazines. Carbamate insecticides and herbicides are detectable at the NPD only at higher concentrations (0.5–1 μg mL^{-1} of sample extract). Flame photometric detector (FPD), although less sensitive than NPD gives good selectivities for organophosphorous and organosulfur compounds; sulfur compounds may interfere when used in the phosphorus mode.

Some molecules do not possess a heteroatom and so are not detectable by the detectors previously described. Thus the analysis must be done by a FID without any selectivity or directly by a more sophisticated system coupling a gas chromatograph with a mass spectrometer (GC/MS) used as detector. For this purpose quadrupole mass spectrometers are more suitable since they scan spectra faster where many spectra have to be scanned in the short time corresponding to the elution of a gas chromatographic peak. If the whole mass spectrum is acquired, GC/MS gives a positive identification of the compound investigated but does not allow any great sensitivity (0.1–1 μg mL^{-1} of sample extract). In order to achieve the necessary sensitivity and selectivity, three or four characteristic mass fragments must be selected from the spectrum of the substances and the acquisition of such ions must be programmed in a time window across the retention time of the substance. Since the complete investigation may require at least two or three chromatographic runs on different instruments with different detectors, some methods were proposed that detected up to 100 pesticides in one or two GC/MS runs by programmed acquisition[60] of given fragments in given time intervals. In Table 1 are listed relative retention times with respect to internal standard fenchlorphos recorded using different columns for most common pesticides and their main mass fragments.

Some molecules require derivatization into more volatile (and/or more detectable) molecules prior to being analyzed by GC. This is the case for phenoxy acid herbicides that are converted by methylation into the corresponding methyl esters. Some other derivatizing agents act by inserting some organohalogen groups into the molecule so that it becomes detectable by the highly sensitive ECD. Methyl iodide, trifluoroacetic acid (TFA) and bistrimethylsilyltrifluoroacetamide are common derivatizing agents.

2.3.2.2 High-performance Liquid Chromatography Techniques (Labile Molecules and High-molecular-weight Molecules) HPLC is used less compared with GC, but it is a necessary tool for all those molecules that are either very labile at high temperature or require temperatures too high to be eluted from the column. Compounds that are usually determined by HPLC are *N*-methylcarbamates insecticides, benzimidazole fungicides, urea compounds and carbamates herbicides.

Injection in the column (10–50 μL) is generally effected from a loop valve. Separations are carried out by reversed phase chromatography onto C_{18} (or C_8) bonded silica columns with acetonitrile, methanol and water as eluents with gradient elution. In order to compensate for changes in light absorption from the eluent during gradient elution and to avoid any baseline drifting, water can be replaced by dilute ammonium acetate[72] (about 0.1 mmol L^{-1}) solutions.

Fixed wavelength ultraviolet (UV) simple detection must be done by choosing an appropriate wavelength where the analyte absorption is maximized and the absorption of interfering substances is minimized. This is quite difficult in some cases because the UV spectra of some pesticides do not exhibit a characteristic peak of absorption. Diode array detectors (DADs), which acquire the UV spectrum in real time, provide special aid in identifying the eluted compound by the UV spectrum pattern. This is of noticeable importance for those molecules that cannot be analyzed at all by

Table 1 Relative retention times with respect to fenchlorphos and detection data of GC-amenable pesticides

Compound name	Molecular formula	CAS-RN	Column[a]			Column[b]		Detector[c]				*m/z* Fragments			
			DB-5	DB-1	Optima Delta-3	PTE-5	OV-1	ECD	NPD	FPD	MS	1	2	3	4
Acephate	$C_4H_{10}NO_3PS$	30560-19-1	0.47	0.39	0.48	0.46	0.46		xx	xx	xx	94	136	95	125
Alachlor	$C_{14}H_{20}ClNO_2$	15972-60-8	0.97		0.97	0.98		xx			xx	160	188	146	161
Aldrin	$C_{12}H_8Cl_8$	309-00-2	1.08	1.10	1.05	1.06		xx			xx	263	265	261	66
Ametryn	$C_9H_{17}N_3S$	834-12-8	1.01	0.97	1.01	0.98	0.97		x		xx	207	212	185	170
Amitraz	$C_{19}H_{23}N_3$	33089-61-11				1.88		x	x		xx	132	147	162	
Anilazine	$C_9H_5Cl_3N_4$	101-05-3						xx			xx	239	241	178	180
Anthraquinone	$C_{14}H_8O_2$	84-65-1	1.12		1.15						xx	208	180	152	76
Atrazine	$C_8H_{14}ClN_5$	1912-24-9	0.78		0.81	0.77	0.77		x		xx	200	202	215	173
Atrazine desethyl	$C_6H_{10}ClN_5$					0.67	0.67		x		xx	172	174	187	
Azinphos ethyl	$C_{12}H_{16}N_3O_3PS_2$	2642-71-9	1.99	1.94	2.02	1.92	1.78	x	xx	xx	xx	132	160	77	105
Azinphos methyl	$C_{10}H_{12}N_3O_3PS_2$	86-50-0	1.85	1.81	1.92	1.82	1.69	x	xx	xx	xx	132	160	77	105
Barban	$C_{11}H_9Cl_2NO_2$	101-27-9		1.37	1.33			x	x		xx	153	155	125	126
Benalaxyl	$C_{20}H_{23}NO_3$	71626-11-4	1.54	1.58	1.46	1.55		x			xx	148	176	206	234
Benfluralin	$C_{13}H_{16}F_3N_3O_4$	1861-40-1	0.67		0.63	0.70		xx			xx	292	264	318	335
Benzoximate	$C_{18}H_{18}ClNO_5$	29104-30-1									xx	105	77	199	202
Bifenox	$C_{14}H_9Cl_2NO_5$	42576-02-3				1.77		xx			xx	173	189	341	343
Biphenyl	$C_{12}H_{10}$	92-52-4	0.36	0.36	0.41	0.43					xx	154	153	152	155
Bitertanol I	$C_{20}H_{23}N_3O_2$	55179-31-2	2.15	2.07	2.04	1.99		x			xx	170	168	152	141
Bitertanol II	$C_{20}H_{23}N_3O_2$		2.18	2.09	2.07	2.00		x			xx	170	168	152	141
Bromophos ethyl	$C_{10}H_{12}BrCl_2O_3PS$	4824-78-6	1.27	1.30	1.19	1.27	1.25	xx	xx	xx	xx	359	357	303	301
Bromophos methil	$C_8H_8BrCl_2O_3PS$	2104-96-3	1.15	1.16	1.12			xx	xx	xx	xx	331	329	333	125
Bromopropylate	$C_{17}H_{16}Br_2O_3$	18181-80-1	1.72	1.74	1.67	1.72		xx			xx	341	339	343	185
Bromoxynil methyl ester	$C_8H_5Br_2O$					0.65		x			xx	276	278	291	293
Bupirimate	$C_{13}H_{24}N_4O_3S$	41483-43-6	1.39	1.42	1.29	1.40		xx	x		xx	208	273	166	316
Butylate	$C_{11}H_{23}NOS$	2008-41-5	0.42	0.42	0.41				x		xx	146	156	174	57
Captafol	$C_{10}H_9Cl_4NO_2S$	2425-06-1	1.65	1.61	1.65	1.62		xx			xx	79	92	151	
Captan	$C_9H_8Cl_3NO_2S$	133-06-2	1.24	1.18	1.22	1.19		xx			xx	79	149	117	264
Carbaryl	$C_{12}H_{11}NO_2$	63-25-2	1.01	0.94	1.03	0.97			x		xx	144	115	117	201
Carbofuran	$C_{12}H_{15}NO_3$	1563-66-2	0.76	0.73	0.82	0.76		x	x		xx	164	149	131	221
Carbophenothion	$C_{11}H_{16}ClO_2PS_3$	786-19-6	1.54	1.56	1.47			xx	xx	xx	xx	157	121	97	342
Chinomethionat	$C_{10}H_6N_2OS_2$	2439-01-2	1.30	1.25	1.26			xx	xx	xx	xx	206	234	116	174
Chlorbicyclène	$C_9H_6Cl_8$	255075-6	1.20	1.22	1.16			xx			xx	272	274	227	229
Chlorbufam	$C_{11}H_{10}ClNO_2$	1967-16-4	0.80	0.75	0.81			x	x		xx	153	155	127	90
Chlordane *cis*	$C_{10}H_6Cl_8$	5103-71-9				1.29		xx			xx	371	373	375	
Chlordane *trans*	$C_{10}H_6Cl_8$	5103-74-2				1.24		xx			xx	371	373	375	
Chlorfenson	$C_{12}H_8Cl_2O_3S$	80-33-1	1.36	1.32	1.30			xx			xx	175	177	111	302
Chlorfenvinphos I	$C_{12}H_{14}Cl_3O_4P$	470-90-6	1.19	1.20	1.13	1.20	1.17	xx	xx		xx	267	269	323	325
Chlorfenvinphos II	$C_{12}H_{14}Cl_3O_4P$		1.22	1.23	1.17	1.21	1.17	xx	xx		xx	267	269	323	325
Chlormephos	$C_5H_{12}ClO_2PS_2$	24923-91-6	0.42	0.41	0.22			xx	xx	xx	xx	121	154	234	236
Chloroneb	$C_8H_8Cl_2O_2$	2675-77-6	0.51	0.50	0.52			xx			xx	161	163	206	208
Chloropropylate	$C_{17}H_{16}Cl_2O_3$	5836-10-2	1.45	1.49	1.34			xx			xx	139	141	251	253
Chlorothalonil	$C_8Cl_4N_2$	1897-45-6	0.84	0.81	1.02	0.86		x	x		xx	266	264	268	230
Chlorpropham	$C_{10}H_{12}ClNO_2$	101-21-3	0.67	0.66	0.66	0.66	0.67	x	x		xx	127	153	171	213
Chlorpyriphos ethyl	$C_9H_{11}Cl_3NO_3PS$	2921-88-2	1.09	1.11	1.07	1.09	1.08	xx	xx	xx	xx	197	199	314	316
Chlorpyriphos methyl	$C_7H_7Cl_3NO_3PS$	5598-13-0	0.95	0.94	0.98	0.96	0.95	xx	xx	xx	xx	286	238	225	209
Chlothal dimethyl	$C_{10}H_6Cl_4O_4$	1861-32-1	1.09	1.13	1.08	1.11		xx			xx	301	303	299	332
Chlozolinate	$C_{13}H_{11}Cl_2NO_5$	84332-86-5	1.22	1.23	1.15	1.20		xx	x		xx	188	259	261	331
Climbazole	$C_{15}H_{17}ClN_2O_2$	38083-17-9	1.36		1.25			x	x		xx	207	209	128	109
Coumaphos	$C_{14}H_{16}ClO_5PS$	56-72-4				2.04			x		xx	226	334	362	364
Cyanazine	$C_9H_{13}ClN_6$	21725-46-2	1.14	1.06	1.13			x	x		xx	212	225	240	198
Cyanofenphos	$C_{15}H_{14}NO_2PS$	13067-93-1	1.55	1.56	1.51			xx	xx	xx	xx	157	169	141	303
Cyfluthrin I	$C_{22}H_{18}Cl_2FNO_3$	68359-37-5	2.32	2.24	2.18	2.12		xx			xx	163	165	199	227
Cyfluthrin II	$C_{22}H_{18}Cl_2FNO_3$		2.35	2.27	2.22	2.14		xx			xx	163	165	199	227
Cyfluthrin III	$C_{22}H_{18}Cl_2FNO_3$		2.36	2.30	2.23	2.15		xx			xx	163	165	199	227
Cyfluthrin IV	$C_{22}H_{18}Cl_2FNO_3$		2.37	2.31	2.25	2.16		xx			xx	163	165	199	227
Cyhalothrin-lambda	$C_{23}H_{19}ClF_3NO_3$	91465-08-6				1.92		xx			xx	141	181	197	208
Cymiazole	$C_{12}H_{14}N_2S$	61676-87-7				0.97					xx	144	185	218	
Cypermethrin I	$C_{22}H_{19}Cl_2NO_3$	52315-07-8	2.41	2.35	2.28	2.15		xx			xx	163	165	281	209
Cypermethrin II	$C_{22}H_{19}Cl_2NO_3$		2.45	2.38	2.32	2.17		xx			xx	163	165	281	209
Cypermethrin III	$C_{22}H_{19}Cl_2NO_3$		2.58	2.40	2.33	2.18		xx			xx	163	165	281	209
Cypermethrin IV	$C_{22}H_{19}Cl_2NO_3$		2.61	2.42	2.35	2.19		xx			xx	163	165	281	209

(*continued overleaf*)

Table 1 (*continued*)

Compound name	Molecular formula	CAS-RN	Column[a]			Column[b]		Detector[c]				*m/z* Fragments			
			DB-5	DB-1	Optima Delta-3	PTE-5	OV-1	ECD	NPD	FPD	MS	1	2	3	4
Cyproconazole I	$C_{15}H_{18}ClN_3$	113096-99-4	1.45	1.44	1.36	1.41		x	x		xx	222	224	139	148
Cyproconazole II	$C_{15}H_{18}ClN_3$		1.46		1.37	1.42		x	x		xx	222	224	139	148
2,4-D methyl ester	$C_{11}H_{12}Cl_2O_3$	1928-38-7	0.65		0.65	0.65		xx			xx	199	201	234	236
2,4-DB methyl ester	$C_9H_8Cl_2O_4$	25333-2-15				0.91		x			xx	101	133	162	164
DDD – *op*	$C_{14}H_{10}Cl_4$	53-19-0	1.38	1.40	1.31	1.37		xx			xx	235	237	165	176
DDD – *pp*	$C_{14}H_{10}Cl_4$	72-54-8	1.48	1.49	1.42	1.47		xx			xx	235	237	165	176
DDE – *op*	$C_{14}H_8Cl_4$	3424-82-6				1.26		xx			xx	246	248	316	318
DDE – *pp*	$C_{14}H_8Cl_4$	72-55-9	1.36	1.39	1.28	1.35		xx			xx	246	248	316	318
DDT – *op*	$C_{14}H_9Cl_5$	784-02-6	1.47	1.51	1.31	1.48		xx			xx	235	237	165	176
DDT – *pp*	$C_{14}H_9Cl_5$	50-29-3	1.56	1.60	1.52	1.58		xx			xx	235	237	165	176
Deltamethrin	$C_{22}H_{19}Br_2NO_3$	52918-63-5	2.92		3.05	2.46		xx			xx	181	251	253	255
Dialiphos	$C_{14}H_{17}ClNO_4PS_2$	10311-84-9	2.00	1.92	1.99			xx	xx	xx	xx	208	210	173	357
Diazinon	$C_{12}H_{21}N_2O_3PS$	333-41-5	0.81	0.83	0.81	0.84	0.85	x	xx	xx	xx	179	199	137	304
Dicamba methyl ester	$C_9H_8Cl_2O_3$	6597-78-0				0.54		xx			xx	203	205	234	236
Dichlofluanid	$C_9H_{11}Cl_2FN_2O_2S_2$	1085-98-9	1.06	1.05	1.06	1.05		xx	x		xx	123	167	200	92
Dichlorvos	$C_4H_7Cl_2O_4P$	62-73-7	0.22	0.19	0.28	0.35	0.35	xx	xx	xx	xx	109	185	187	220
Dicrotophos	$C_8H_{16}NO_5P$	3735-78-3				0.68			xx		xx	109	127	193	237
Dicofol	$C_{14}H_9Cl_5O$	115-32-2	1.63	1.75	1.36			xx			xx	139	141	251	253
Dieldrin	$C_{12}H_8Cl_6O$	60-57-1	1.38	1.38	1.30	1.35		xx			xx	79	263	265	277
Diethofencarb	$C_{14}H_{21}NO_4$	87130-20-9	1.11		1.07				x		xx	151	150	124	267
Dimefox	$C_4H_{12}FN_2OP$	115-26-4	0.01		0.10			x	x	x	xx	110	154	92	
2,4-Dimethylaniline	$C_8H_{11}N$	95-68-1				0.31			x		xx	106	120	121	
Dimethoate	$C_5H_{12}NO_3PS_2$	60-51-5	0.75	0.71	0.82	0.60	0.74	x	xx	xx	xx	87	93	125	229
Dimetilan	$C_{10}H_{16}N_4O_3$	644-64-4	0.96	0.95	1.02				xx		xx	72	73	170	
Dioxathion	$C_{12}H_{26}O_6P_2S_4$	78-34-2	0.78	0.77	0.83				xx	xx	xx	125	153	197	270
Diphenylamine	$C_{12}H_{11}N$	122-39-4	0.63	0.61	0.64	0.63		x	x		xx	169	168	170	
Disulfoton	$C_8H_{19}O_2PS_3$	298-04-4				0.85	0.86		xx		xx	142	186	274	
Ditalimfos	$C_{12}H_{14}NO_4PS$	5131-24-8	1.32		1.27			xx	xx	xx	xx	130	148	243	299
Endosulfan alfa	$C_9H_6Cl_6O_3S$	959-98-7	1.31	1.32	1.24	1.28		xx			xx	195	237	235	241
Endosulfan beta	$C_9H_6Cl_6O_3S$	33213-65-3	1.47	1.44	1.43	1.43		xx			xx	195	237	235	241
Endosulfan sulfate	$C_9H_6Cl_6O_4S$	1031-07-8	1.57	1.55	1.57	1.56		xx			xx	272	274	387	389
Endrin	$C_{12}H_8Cl_6O$	72-20-8	1.43	1.43	1.36	1.41		xx			xx	263	261	265	245
Esaconazole	$C_{14}H_{17}Cl_2N_3O$	79983-71-4	1.37	1.36	1.27			xx	x		xx	214	216	231	233
Esfenvalerate I	$C_{25}H_{22}ClNO_3$	66230-04-4	2.61		2.64			x			xx	125	167	225	449
Esfenvalerate II	$C_{25}H_{22}ClNO_3$		2.67		2.72			x			xx	125	167	225	449
Ethalfluralin	$C_{33}H_{14}F_3N_3O_4$	55283-68-6	0.64	0.68	0.61			xx			xx	276	292	316	269
Ethiofencarb	$C_{11}H_{15}NO_2S$	29973-13-5	0.92	0.87	0.96	0.90					xx	107	168	108	78
Ethion	$C_9H_{22}O_4P_2S_4$	563-12-2	1.48	1.53	1.37	1.49	1.42	xx	xx	xx	xx	231	125	153	384
Ethofumesate	$C_{13}H_{18}O_5S$	26225-79-6				1.04		x			xx	137	161	207	286
Ethoprophos	$C_8H_{19}O_2PS_2$	13194-48-4	0.63	0.63	0.63			x	xx	xx	xx	158	200	139	126
Etrimfos	$C_{10}H_{17}N_2O_4PS$	38260-54-7	0.86	0.87	0.87		0.89	xx	xx	xx	xx	181	168	292	153
Famphur	$C_{10}H_{16}NO_5PS_2$	52-85-7	1.54	1.54	1.50			xx	xx	xx	xx	218	125	93	
Fenamiphos	$C_{13}H_{22}NO_3PS$	22224-92-6	1.34	1.34	1.26				xx	xx	xx	154	217	260	303
Fenarimol	$C_{17}H_{12}Cl_2N_2O$	60168-88-9	1.98	1.92	1.95	1.90		xx	xx		xx	139	219	251	253
Fenazaflor	$C_{15}H_7Cl_2F_3N_2O_2$	14255-88-0	1.51	1.57	1.41			xx	x		xx	234	236	374	376
Fenazaquin	$C_{20}H_{22}N_2O$	120928-09-8	1.78		1.75				x		xx	145	160	146	117
Fenclorphos	$C_8H_8Cl_3O_3PS$	299-84-3	1.00	1.00	1.00	1.00	1.00	xx	xx	xx	xx	285	287	125	270
Fenfuram	$C_{12}H_{11}NO_2$	24691-80-3	0.88	0.83	0.91				x		xx	109	201	110	202
Fenitrothion	$C_9H_{12}NO_5PS$	122-14-5	1.05	1.02	1.06	1.03	1.01	xx	xx	xx	xx	125	109	277	260
Fenoxycarb	$C_{17}H_{19}NO_4$	79127-80-3	1.93		1.73	1.73			x		xx	116	186	88	255
Fenpropathrin	$C_{22}H_{22}NO_3$	64257-84-7	1.78	1.77	1.70			xx	x		xx	181	209	265	97
Fenson	$C_{12}H_9ClO_3S$	80-38-6	1.16	1.11	1.15			xx		xx	xx	268	270	141	77
Fenthion	$C_{10}H_{15}O_3PS_2$	55-38-9	1.11	1.09	1.09	1.09	1.07		xx	xx	xx	278	169	125	109
Fenvalerate I	$C_{25}H_{22}ClNO_5$	51630-58-1	2.61	2.58	2.64	2.35		x			xx	125	127	167	169
Fenvalerate II	$C_{25}H_{22}ClNO_5$		2.67	2.64	2.72	2.46		x			xx	125	127	167	169
Fluazifop butyl	$C_{19}H_{20}F_3NO_4$	79241-46-6	1.45	1.49	1.30			xx	x		xx	282	383	254	227
Fluorodifen	$C_{13}H_7F_3N_2O_5$	15457-05-3	1.35	1.33	1.29			xx	xx		xx	190	160	162	126
Flusilazole	$C_{16}H_{15}F_2NSi$	85509-19-9	1.41	1.41	1.31			xx	x		xx	233	206	315	234
Folpet	$C_9H_4Cl_3NO_2S$	133-07-3	1.26	1.21	1.24	1.21		xx	x		xx	75	76	106	260
Fonofos	$C_{10}H_{15}OPS_2$	944-22-9	0.80	0.80	0.84	0.82	0.83	xx	xx	xx	xx	109	137	246	174
Formothion	$C_6H_{12}NO_4PS_2$	2540-82-1	0.92		0.98			xx	xx	xx	xx	93	125	126	170
Furalaxyl	$C_{17}H_{19}NO_4$	57646-30-7				1.23			x		xx	95	242	301	
Furathiocarb	$C_{18}H_{26}N_2O_5S$	65907-30-4				1.83					xx	135	163	194	325

Table 1 (*continued*)

Compound name	Molecular formula	CAS-RN	Column[a]			Column[b]		Detector[c]				*m/z* Fragments			
			DB-5	DB-1	Optima Delta-3	PTE-5	OV-1	ECD	NPD	FPD	MS	1	2	3	4
Heptachlor	$C_{10}H_5Cl_7$	76-44-8	0.98	0.99	0.98	0.98		xx			xx	272	274	100	66
Heptachlor epoxide	$C_{10}H_5Cl_7O$	1024-57-3	1.21	1.22	1.17	1.19		xx			xx	183	185	217	237
Heptenofos	$C_9H_{12}ClO_4P$	23560-59-0	0.57	0.56	0.57	0.59		xx	xx	xx	xx	124	126	250	109
Hexachlorobenzene	C_6Cl_6	118-74-1	0.70	0.73	0.77	0.74		xx			xx	284	282	280	288
Hexythiazox	$C_{17}H_{21}ClN_2O_2S$	78587-05-0				1.24					xx	156	184	227	
HCH alpha	$C_6H_6Cl_6$	319-84-6				0.72		xx			xx	181	183	217	219
HCH beta	$C_6H_6Cl_6$	319-85-7				0.78		xx			xx	181	183	217	219
HCH delta	$C_6H_6Cl_6$	319-86-8				0.85		xx			xx	181	183	217	219
Imazalil	$C_{14}H_{14}Cl_2N_2O$	35554-44-0	1.43	1.36	1.28	1.34		xx	x		xx	215	217	173	175
Iprodione	$C_{13}H_{13}Cl_2N_3O_3$	36734-19-7	1.72	1.72	1.71	1.70		xx	x		xx	314	316	187	189
Isofenphos	$C_{15}H_{24}NO_4PS$	25311-71-1	1.21	1.25	1.14	1.21	1.19	xx	xx	xx	xx	58	121	255	213
Jodfenfos	$C_8H_8Cl_2IO_3PS$	18181-70-9	1.35	1.35	1.27			xx	xx	xx	xx	377	379	125	
Lindane (HCH gamma)	$C_6H_6Cl_6$	58-89-9	0.77	0.76	0.85	0.79		xx			xx	181	183	217	219
Malathion	$C_{10}H_{19}O_6PS_2$	121-75-5	1.08	1.07	1.05	1.07	1.05	xx	xx	xx	xx	173	125	127	158
Mecarbam	$C_{10}H_{20}NO_5PS_2$	2595-54-2	1.24	1.23	1.17	1.22	1.19	xx	xx	xx	xx	131	159	160	125
Mecoprop methyl ester	$C_{11}H_{13}ClO_3$	2786-19-8				1.80		xx			xx	142	169	228	230
Metalaxyl	$C_{15}H_{21}NO_4$	57837-19-1	1.00	0.99	1.01	0.99			x		xx	206	160	146	249
Metamidophos	$C_2H_8NO_2PS$	10265-92-6	0.27	0.16	0.30	0.34	0.33		xx	xx	xx	94	95	141	
Metamitron	$C_{10}H_{10}N_4O$	41394-05-2				1.40		xx			xx	104	173	174	202
Methacrifos	$C_7H_{13}O_5PS$	30864-28-9	0.49		0.50		0.52		xx	xx	xx	125	180	208	240
Methidathion	$C_6H_{11}N_2O_4PS_3$	10265-92-6	1.27	1.26	1.24	1.25	1.20	xx	xx	xx	xx	145	85	93	125
Methiocarb	$C_{11}H_{15}NO_2S$	2032-65-7	1.06	1.02	1.06	1.03			x		xx	168	153	225	109
Metolachlor	$C_{15}H_{22}ClNO_2$	51218-45-2	1.09		1.06	1.07		xx	x		xx	162	238	163	146
Metribuzin	$C_8H_{14}N_4OS$	21087-64-9	0.96	0.90	1.01				xx		xx	198	199	214	171
Mevinphos	$C_7H_{13}O_6P$	7786-34-7		0.40	0.44				xx	xx	xx	127	109	192	
Molinate	$C_9H_{17}NOS$	2212-67-1	0.53			0.55	0.56		x		xx	126	187	99	84
Monocrotophos	$C_7H_{14}NO_5P$	6923-22-4	0.70		0.75	0.70	0.67		xx	xx	xx	127	97	192	109
Myclobutanil	$C_{15}H_{17}ClN_4$	88671-89-0	1.40	1.38	1.34	1.38		xx	x		xx	179	181	150	152
Naled	$C_4H_7Br_2Cl_2O_4P$	300-76-5	0.69	0.65	0.68			xx	xx	xx	xx	109	185	187	220
Nicotine	$C_{10}H_{14}N_2$	54-11-5	0.34	0.32	0.37				x		xx	84	133	162	
Nuarimol	$C_{17}H_{12}ClFN_2O$	63284-71-9	1.62	1.63	1.57	1.62		xx	x		xx	235	314	203	139
Omethoate	$C_5H_{12}NO_4PS$	1113-02-6		0.55		0.60	0.60	x	x	x	x	110	156	79	80
Orthophenylphenol	$C_{12}H_{10}O$	90-43-7	0.53	0.50	0.53	0.53					xx	170	149	141	115
Oxadiazon	$C_{15}H_{18}Cl_2N_2O_5$	19666-30-9	1.38	1.42	1.26			xx	x		xx	175	177	258	260
Oxadixyl	$C_{14}H_{18}N_2O_4$	77732-09-3	1.49	1.48	1.46				x		xx	132	163	105	118
Oxyfluorfen	$C_{15}H_{11}ClF_3NO_4$	42874-03-3				1.39		xx			xx	252	300	361	
Parathion ethyl	$C_{10}H_{14}NO_5PS$	56-38-2	1.12	1.10	1.10	1.09	1.07	xx	xx	xx	xx	109	97	291	139
Parathion methyl	$C_8H_{10}NO_5PS$	298-00-0	0.98	0.93	1.01	0.96	0.94	xx	xx	xx	xx	109	125	263	93
Penconazole	$C_{13}H_{15}Cl_2N_3$	66246-88-6	1.21	1.20	1.17	1.19		xx	x		xx	159	161	248	250
Pendimethalin	$C_{13}H_{19}N_3O_4$	40487-42-1	1.20		1.15			xx	xx		xx	252	281	191	162
Pentachloroaniline	$C_6H_2Cl_5N$	527-20-8	0.91	0.88	0.97	0.91		xx	x		xx	265	263	267	269
Pentachloroanisole	$C_7H_3Cl_5O$	1825-21-4	0.72	0.74	0.77			xx			xx	280	282	237	265
Pertane	$C_{18}H_{20}Cl_2$	72-56-0				1.43		x			xx	165	223	224	
Phenthoate	$C_{12}H_{17}O_4PS_2$	2597-03-7	1.23	1.23	1.18			xx	xx	xx	xx	274	121	125	246
Phorate	$C_7H_{17}O_2PS_3$	298-02-2	0.69	0.70	0.71	0.71	0.72	xx	xx	xx	xx	75	121	231	260
Phosalone	$C_{12}H_{15}ClNO_4PS_2$	2310-37-0	1.84	1.81	1.87	1.82	1.71	xx	xx	xx	xx	182	184	367	369
Phosmet	$C_{11}H_{12}NO_4PS_2$	732-11-6	1.71	1.70	1.78	1.71		xx	xx	xx	xx	160	161	133	104
Phosphamidon	$C_{10}H_{19}ClNO_5P$	13171-21-6	0.93	0.91	0.96		0.92	xx	xx	xx	xx	127	138	264	72
Piperonyl butoxide	$C_{19}H_{30}O_5$	51-03-6	1.65	1.69	1.53	1.66					xx	176	178	149	
Pirimicarb	$C_{11}H_{18}N_4O_2$	23103-98-2	0.89	0.88	0.92	0.91			x		xx	166	238	72	
Pirimiphos methil	$C_{11}H_{20}N_3O_3PS$	29232-93-7	1.04	1.05	1.01	1.04	1.03	xx	xx	xx	xx	290	125	233	305
Prochloraz	$C_{13}H_{11}Cl_2NO_2$	32809-16-8				2.04		xx			xx	180	266	308	310
Procymidone	$C_{13}H_{11}Cl_2NO_2$	32809-16-8	1.27	1.26	1.19	1.23		xx	x		xx	96	67	283	285
Profenofos	$C_{11}H_{15}BrClO_3PS$	41198-08-7	1.36	1.38	1.28	1.34	1.30	xx	xx	xx	xx	208	139	337	339
Promecarb	$C_{12}H_{17}NO_2$	2631-37-0	0.70	0.69	0.71				x		xx	135	150	92	108

(*continued overleaf*)

Table 1 (*continued*)

Compound name	Molecular formula	CAS-RN	Column[a]			Column[b]		Detector[c]				*m/z* Fragments			
			DB-5	DB-1	Optima Delta-3	PTE-5	OV-1	ECD	NPD	FPD	MS	1	2	3	4
Propanil	$C_9H_9Cl_2NO$	709-98-8	0.99	0.90	0.99			xx			xx	161	163	217	219
Propargite I	$C_{19}H_{26}O_4S$	2312-35-8	1.61	1.66	1.52	1.63					xx	135	173	150	350
Propargite II	$C_{19}H_{26}O_4S$		1.62		1.53	1.64					xx	135	173	150	350
Prometryn	$C_{10}H_{19}N_5O$	7287-19-6				1.00			x		xx	226	241		
Propazine	$C_9H_{16}ClN_5$	139-40-2				0.79			x		xx	172	214	229	
Propham	$C_{10}H_{13}NO_2$	122-42-9	0.45	0.45	0.23	0.48			x		xx	93	137	179	120
Propiconazole I	$C_{15}H_{17}Cl_2N_3O_2$	60207-90-1	1.58	1.60	1.49	1.56		xx	xx		xx	173	175	259	261
Propiconazole II	$C_{15}H_{17}Cl_2N_3O_2$		1.59	1.62	1.51	1.57		xx	xx		xx	173	175	259	261
Propoxur	$C_{11}H_{15}NO_3$	114-26-1	0.60	0.60	0.62	0.62			x		xx	110	152	92	
Propyzamide	$C_{12}H_{11}Cl_2NO$	23950-58-5	0.82	0.81	0.84	0.82		xx			xx	173	175	145	147
Prothiofos	$C_{11}H_{15}Cl_2O_2PS_2$	34643-46-4	1.34		1.25			xx	xx	xx	xx	162	164	267	309
Pyracarbolide	$C_{13}H_{15}NO_2$	24691-76-7	1.18	1.13	1.15				x		xx	125	126	217	218
Pyrazophos	$C_{14}H_{20}N_3O_5PS$	13457-18-6	1.97	1.98	1.94			xx	xx	xx	xx	221	232	265	373
Pyridaphenthion	$C_{14}H_{17}N_2O_4PS$	119-12-0	1.70	1.71	1.69	1.71	1.64	xx	xx	xx	xx	340	199	188	204
Pyrimethanil	$C_{12}H_{13}N_3$	53112-28-0	0.83	0.81	0.87	0.83			xx		xx	198	199	200	
Quinalfos	$C_{12}H_{15}N_2O_5PS$	13593-03-8	1.24	1.24	1.18	1.21	1.18	xx	xx	xx	xx	146	157	156	298
Quintozene	$C_6Cl_5NO_2$	82-68-8	0.77	0.78	0.86	0.81		xx			xx	293	295	297	237
Secbumeton	$C_{10}H_{19}N_5O$	26259-45-0	0.87	0.84	0.88				xx		xx	196	210	225	169
Simazine	$C_8H_{15}N_5S$	122-34-8	0.77		0.81	0.76	0.75		x		xx	201	203	186	158
Simetryn	$C_7H_{12}ClN_5$	1014-70-6				0.97			x		xx	155	170	213	
Tebuconazole	$C_{16}H_{22}ClN_3O$	107534-96-3	1.63		1.58	1.62		x	x		xx	125	127	250	252
Tebufenpyrad	$C_{18}H_{24}ClN_3O$	119168-77-3	1.78		1.72			x	x		xx	171	276	318	333
TEPP	$C_8H_{20}O_7P_2$	104-49-3				0.57	0.57	xx			xx	161	179	235	263
Terbumeton	$C_{10}H_{19}N_5O$	33693-04-8	0.78	0.77	0.81				xx		xx	210	225	169	141
Terbuthylazine	$C_9H_{16}ClN_5$	5915-41-3	0.81	0.84	0.84	0.81	0.81		x		xx	214	216	173	229
Terbutryn	$C_{10}H_{19}N_6S$	886-50-0				1.03			x		xx	185	226	241	
Tetrachlorvinphos	$C_{10}H_9Cl_4O_4P$	22248-79-9	1.30	1.30	1.23			xx	xx	xx	xx	329	331	333	109
Tetraconazole	$C_{13}H_{11}Cl_2F_4N_3O$	11228-77-3				1.12		xx			xx	159	171	336	338
Tetradifon	$C_{12}H_6Cl_4O_2S$	116-29-0	1.83	1.79	1.85	1.80		xx			xx	159	161	356	229
Thiabendazole	$C_{10}H_7N_3S$	148-79-8	1.30		1.24				x		xx	201	174	202	
Tiocarbazil I	$C_{16}H_{25}NOS$	36756-79-3	1.13	1.14	1.07				x		xx	156	100	91	57
Tiocarbazil II	$C_{16}H_{25}NOS$		1.14	1.15	1.08				x		xx	156	100	91	57
Tolclophos methyl	$C_9H_{11}Cl_2O_3PS$	57018-04-9	0.97	0.95	1.00	0.97	0.96	xx	xx	xx	xx	265	267	125	93
Tolifluanide	$C_{10}H_{13}Cl_2OF_2S_2$	731-27-1	1.20		1.17			xx			xx	137	238	240	181
Triadimefon	$C_9H_{16}ClN_3O_2$	43121-43-3	1.14	1.12	1.09	1.10		xx	x		xx	208	210	181	57
Triadimenol I	$C_{14}H_{18}ClN_3O_2$	55219-65-3	1.27	1.25	1.19	1.22		xx	x		xx	112	128	168	70
Triadimenol II	$C_{14}H_{18}ClN_3O_2$		1.29	1.27	1.20	1.24		xx	x		xx	112	128	168	70
Triazophos	$C_{12}H_{16}N_3O_3PS$	24017-47-8				1.53	1.45	x	xx	xx	xx	161	162	172	256
Trifluralin	$C_{13}H_{16}F_3N_3O_4$	1582-09-8	0.66	0.70	0.62			xx	xx		xx	264	306	276	290
Vamidothion	$C_9H_{18}NO_4PS_2$	2275-23-2	1.31		1.27	1.27	1.21	x	xx	xx	xx	87	145	109	119
Vinclozolin[d]	$C_{12}H_9Cl_2NO_3$	50471-44-8	0.97	0.95	0.97	0.96		xx	x		xx	285	287	198	212

[a] Data from ARPA Dip. di Torino using 30-m columns (0.25-mm internal diameter and 0.25-μm film thickness) and the following temperature program: 1 min at 100 °C, from 100 °C to 190 °C at 7 °C min^{-1}, 5 min at 190 °C, from 190 °C to 250 °C at 5 °C min^{-1}, 10 min at 250 °C, then from 250 °C to 270 °C at 5 °C min^{-1}.

[b] Data from PMIP – ASL Como using a 30-m column (PTE-5 and OV1: 0.25 and 0.32-mm internal diameter, 0.25 and 0.4-μm film thickness, respectively) and the following temperature program: 1 min at 60 °C, from 60 °C to 150 °C at 25 °C min^{-1} and to 270 °C at 3 °C min^{-1}; 15 min at 270 °C.

[c] (xx) = good detectability; (x) = medium to scarce detectability.

[d] Use also ions with $m/z = 178$ and 187 in the case of the presence and coelution of chlorpyriphos methyl.

DDD, 1,1-dichloro-2,2-bis[*p*-chlorophenyl]ethane; DDE, 1,1-dichloro-2,2-bis[*p*-chlorophenyl]ethylene; TEPP, tetraethyl pyrophosphate; HCH, hexachlorocyclohexane.

GC/MS (at least for qualitative confirmation) and when the laboratory does not possess MS interfaced HPLC. Fluorimetric detectors are much more sensitive than UV detectors whereas electrochemical detectors are as selective (by selecting the characteristic potential of the analyte's redox couple), although expensive and delicate.

Derivatization with dansyl- or *o*-phthalaldehyde (OPA) is often carried out in order to achieve more selectivity and sensitivity. Either group with good fluorescence properties or a halogen atom that makes the analyte itself fluorescent can be inserted in the molecule, but because these groups increase the molecular dimensions,

thus minimizing the differences between the analytes that have to be separated, it is necessary to perform the derivatization after each molecule has been eluted from the column and has been separated from the others. Precolumn derivatization is also used, but although postcolumn[73] derivatization involves the use of additional devices it gives better selectivity.

HPLC has been interfaced with MS spectrometers (liquid chromatography/mass spectrometry (LC/MS)) allowing the application of such detectors to heavy molecular weight or labile molecules. A review of the application of LC/MS to the analysis of pesticide residues, drugs and mycotoxin in food was published by Careri et al.[74]

2.3.2.3 Qualitative Identification and Confirmation of Residues All the chromatographic methods generally identify the respective analytes by the time it takes for elution of each compound from the column within the definition of absolute and relative retention time. However, because of potential interferences from the matrix and the possibilities of coelution of different molecules, whatever column is used, some confirmation by an alternative instrumental technique, detector or method is needed. The confirmation becomes a ''must'' when a residue is found to be over the active limits stated by regulation.

The most widely used and selective technique used to confirm results is a MS detector coupled to a gas chromatograph. The whole mass spectrum is a unique fingerprint of the compound detected which provides the best confirmation but unfortunately it is not always possible to acquire the whole mass spectrum down to levels of $0.1-0.05\,\mu g\,mL^{-1}$ of sample extract. Therefore selected ion monitoring (SIM) choosing the three–four most specific fragments (see also Table 1) must be used.

For confirmation by GC/MS the chosen fragment must have a m/z value (when possible) over 100. Further confirmation can be obtained when the residue found exceeds the active limit (or maximum residue limit (MRL)), by comparing abundance ratios of the acquired ionic fragments, which must fit those of the standard within a tolerance of $\pm 20\%$.

In other cases useful confirmation can also be obtained with the same detector and a different polarity column. Confirmation can also be obtained for some compounds that respond to more than one detector by comparing chromatographic runs with NPD/FPD and ECD detectors. HPLC with UV or fluorimetry detection can also be used to this purpose. DADs which acquire the whole UV spectrum, also provide a fingerprint of the eluted compounds.

The FDA *Pesticide Analytical Manual*,[15] Vol. 2, parts 1–5 provides several alternative analytical methods for residues confirmation.

More and more applications have been published concerning the use of tandem GC tandem mass spectrometry (MS/MS) where an ion selected from the spectrum produced by the first mass spectrometer is fragmented into a second mass spectrometer with a greatly enhanced selectivity. The availability of relatively low cost instruments based on the ion trap will make probably MS/MS the main technique[75] for residues confirmation in the future.

2.3.2.4 Quantitative Determination Quantitative determinations, as in all chromatographic methods, are made by comparison of the analyte's peak with a peak from standard solution of known concentration. Owing to the usually very low volume injected in GC ($1-2\,\mu L$), errors in measuring this volume can occur, and the use of internal standard can automatically correct these errors. On the other hand, in HPLC, since loop valves always inject the same standardized amount of sample ($10-100\,\mu L$), the column does not need an internal standard. The use of an internal standard also corrects for cases where variations in the absolute response given from detector are caused by uncontrolled phenomena. Instrumental errors in the 5–15% range are common for external standard quantitations, and use of an internal standard can decrease this error down to 5–2%.

Calibration can also be performed on the matrix extract or by addition to the sample in order to take account of the potential effect of the matrix (see below).

In quantitation performed with GC/MS instruments it is recommended that as large m/z values as possible for the fragments be chosen so that less inteference can occur from the matrix. Another method used in GC/MS quantitation is isotopic dilution,[76] where an isotope-enriched standard for the compound being investigated is added to the sample giving rise to a change in the isotopic ratio from which the concentration of the compound can be calculated.

Since the cleanup step normally does not remove all the matrix, the organic matter in the extract that is present in concentrations greater by orders of magnitude than those of the pesticides investigated, may interfere heavily with the instrumental determination of some molecules. This happens especially in GC multiresidue techniques where the matrix progressively dirties the column during use. Some pesticides are quite labile (e.g. captan, captafol, dicofol, pyrethroids, *N*-methylcarbamates, *o*,*p*-DDT) and undergo partial degradation during the gas chromatographic run, whereas other molecules (i.e. acephate, metamidofos, vamidothion, methomyl and other *N*-methylcarbamates) strongly interact with active sites created in dirty used columns with peak broadening and decrease in sensitivity. Thus when a standard pesticide mixture in pure solvent is injected, peaks corresponding to the active molecules cited above are

broadened or are not detectable especially at low concentrations 0.1–0.5 μg/mL. The so-called ''matrix-effect'' works so that, at the same concentration, in the presence of the matrix, the coextracts saturate the active sites in the column or in the liner and the pesticide can be observed with its whole peak intensity. On the other hand, in the absence of the matrix, the pesticide interacts with the column (or liner) active sites bringing about peak intensity reduction and consequent overestimation of the data. These effects are sometimes not in evidence because the overestimation produced by the matrix effect goes to balance a recovery under 100%, whereas in some cases recovery values over 200% are produced; some papers[77–79] report studies characterizing the matrix effect. To take into account this phenomenon, which is particularly severe in the case of cold-on-column injections it is necessary: (1) to have a guard column of 2–5-m length that can be periodically changed, (2) to change the liner insert in the injector if present, and (3) to perform the calibration in the same (or closest) matrix extract as the sample, even for those active molecules for which this effect produces recoveries of >120%. Kocourek et al.[80] studied the stability of standard solutions in extracts and they found that at 20 °C, over a two-month period, degradation of several pesticides was observed, so that storage of sample and calibration extracts in a freezer is recommended.

2.3.2.5 Determination of Metabolites In many cases the active molecule originally sprayed on the plant may not be the only residue to look for because degradation on the plant and metabolism in animal tissues originates new products. In some cases these metabolites exhibit similar or more toxicity with respect to the parent compound (e.g. DDE and DDD from DDT). Table 2 lists the main metabolites for some active molecules. The metabolites in most cases are extracted together with the active parent molecules by the multiresidue methods according to the capability of the method to extract large spectra of different organic compounds and classes. Information about pesticide metabolites, their fates and relative studies are also available in book series such as *Reviews of Environmental Contamination and Toxicology*.[81] Other information about methods for metabolites can also be found in the report[82] series *Pesticide Residues in Food* by Joint Meeting for Pesticide Residues (JMPR) which reports the trials performed worldwide in order to assess the residue levels for a given active molecule in each commodity and for each country following good agricultural practice (GAP).

A particularly well-studied metabolite is ethylene-thiourea (ETU) from the degradation of ethylenebis- and propylenebis-dithiocarbamate fungicides either on plants or during food processing; this is a case when the metabolite is more toxic than the parent compound. ETU must be derivatized in order to make it detectable well by GC. Dubey et al.[83] recently proposed a double step

Table 2 Main metabolites of some pesticides

Active molecule	Metabolites
Acephate	Methamidophos
Aldicarb	Aldicarb sulfoxide, aldicarb sulfone, aldicarb oxime, aldicarb nitrile
Aldrin	Dieldrin
Benfuracarb	Carbofuran, 3-hydroxycarbofuran, 3-ketocarbofuran
Captan	Captan epoxide, 1,2,3,6-THI, 3-OH-THI, 5-OH-THI, 4,5-di-OH-THI, THI-epoxide
Carbaryl	2-Naphtol
Chlordane	Oxychlordane, nonachlor
Chlorpropham	3-Chloroaniline
Chlorpyriphos methyl	Chlorpyriphos methyl-oxon, 3-chloropyridinol
DDT	DDD, DDE
Diazinon	OH-Diazinon, diazo-oxon
Dicofol	*p*,*p*′-DCBF, dichlorodicofol, *p*,*p*′-dichlorobenzidrole
Dimethoate	Omethoate
Diuron	1-(3,4-Dichlorophenyl)urea, 3,4-dichloroaniline
Dithiocarbamates (ethylene-bis-, propylene-bis-)	ETU – PTU
Endosulfan	Endosulfan sulfate
Ethefon	Ethylene
Ethion	Ethion-monoxon, ethion-dioxon, diethyldihydrogenphosphate
Folpet	Phthalimyide
Gliphosinate ammonium	3-OH-(Methyl)-phosphinoyl-propionic acid
Gliphosate	Aminomethylphosphonic acid
Heptaclor	Heptachlor epoxide
Iprodione	*N*,(3,5-Dichlorophenyl)-3-isopropyl-2,4-dioxyimidazolydine-1-carbox-amide; 3-(3,5-dichlorophenyl)-2,4-dioxyimidazolidine-1-carboxamide; 3,5 dichloroaniline; 3,5 dichloro-phenylurea
Malathion	Malaoxon, malathion monocarboxylic acid
Parathion	Paroxon, 4-nitrophenol
Parathion methyl	Paraoxon methyl, 4-nitrophenol,
Phosalone	Phosaloneoxon, phosalone sulfide, phosalone thiol, phosalone glycoside, phosalone aglycone
Tebuconazole	Terbutyl derivative,
Tecnazene	Tetrachloroaniline, tetrachlorophenol, tetrachlorobenzene
Tolclophos methyl	2,6-Dichloro-4-methylphenyldimethyl-phosphate, 2,6-dichloro-4-methyl-phenol
Vinclozolin	3,5-Dichloroaniline

THI, tetrahydrophthalimide; DCBF, dichlorobenzophenone; PTU, propylenethiourea.

derivatization method with benzyl chloride, followed by LLP, and then with TFA and double detection by ECD and NPD.

2.4 Special Methods for Specific Classes

2.4.1 Gas Chromatography Headspace Methods

2.4.1.1 Fumigants Although official methods for fumigants in grains still use extraction with solvents, some methods[84,85] have been developed using headspace techniques, relying mostly on the high volatility of the compound investigated. OC fumigants, e.g. are not different from the volatile organic compounds (VOCs) analyzed in water samples by headspace methods. Headspace methods require less preparation of the sample and have less risk of analyte losses.

2.4.1.2 Dithiocarbamates Dithiocarbamates are usually detected by headspace methods as total carbon sulfide, after reduction with tin(II) chloride in acid medium. General purpose columns for VOCs are used with ECDs or FPDs (sulfur mode).[86] These methods are fast and allow analysis of much more sample at the same time than with colorimetric methods (see below).

2.4.2 Ultraviolet/Visible Colorimetric Methods (Dithiocarbamates, Phosphine)

Ultraviolet/visible (UV/VIS) colorimetric methods are based upon selective reaction with a chromogenic reagent and a single pesticide (or class of pesticides) to give a colored adduct that is spectrophotometrically measurable.

Obviously colored substances from the vegetable and coextracted substances that may react with chromogenic reagent may cause interference. Some methods that provide isolation of the analyte by distillation or degradation into other products are more or less widely used.

2.4.2.1 Methods for Dithiocarbamates A dated method that is still widely used is that by Mestres[87] where dithiocarbamates are reduced to carbon sulfide by heating the sample in the presence of tin(II) chloride and potassium iodide, in an acidic medium. The carbon sulfide developed is fixed by chromogenic reagent (copper acetate, diethanolamine in ethanol) as copper dithiocarbamate and measured spectrophotometrically at 435 nm. Two traps (sodium hydroxide for acidic substances and zinc acetate for bases) must be inserted before chromogenic reaction to avoid interference. Naturally producing carbon sulfide substances can give noticeable interference or false positives (rocket, capers, cabbages).

Methods that distinguish between single dithiocarbamates compounds are not widely developed. Some methods are proposed for thiram and ziram in the FDA *Pesticide Analytical Manual* Vol. 2[15] and for thiram by Ekroth et al.[88]

2.4.2.2 Methods for Phosphine Phosphine can be distilled from treated commodities in the presence of concentrated sulfuric acid, transformed into orthophosphate anion by oxidation with bromine[89] or potassium permanganate[90] and detected spectrophotometrically by methods used for phosphates. Alternative simple methods measure phosphine by direct reaction[91] with silver nitrate giving a yellow complex in methanol.

2.4.2.3 Methods for o-*Phenilphenol and Biphenyl* An old method[92] measures these compounds by refluxing an acidified sample with apposite apparatus (Clavenger) and the distilled pesticide residues are collected in 5–10 mL of hexane floating on the condensed water. Pesticides extracted can be determined either by UV/VIS spectrophotometric methods or by chromatographic methods.

2.4.3 Atomic Absorption Methods (Inorganic and Organometallic Compounds)

Atomic absorption methods do not differ from those used in assessing metal trace levels in food (see **Atomic Spectroscopy in Food Analysis**). A limiting factor may be the natural content of a given metal in a crop, which may originate from a pollution history in its own soil, so that the pesticide levels to be detected are lower.

Modern inductively coupled plasma atomic emission spectrometry (ICPAES) can determine elements other than metals such as phosphorus. An example for phosphine/orthophosphate is given by Mortensen et al.[90]

2.4.4 Immunoenzymatic Methods and Other Methods

Immunoenzymatic techniques are less used methods and are mainly used in screening. Many of them detect the function characteristics of given groups (e.g. triazines, OC, organophosphates) without discriminating between the single compounds in the class and are therefore of less interest because residue limits are fixed for each compound by regulatory acts. More selective methods known as enzyme-linked immunosorbent assay[28,93] (ELISA) based on the reaction of a pesticide-horseradish peroxidase (HRP) conjugate with a monoclonal or polyclonal antibody have been developed. The ELISA method allows determination of specific compounds at parts per trillion levels. The main advantage of these methods are that they are less time-consuming, analysis is in situ, and they have high sensitivity and specificity. More information is available elsewhere.[94–98]

Another technique that must be taken into account is thin-layer chromatography (TLC), which despite

the many advances that have been brought about by more sophisticated techniques like GC and liquid chromatography has remained a useful technique for screening and quantitative analysis in the field of pesticide residues. A comprehensive and synthetic review has been published by Rathore and Begum[99] where stationary phases, mobile phases, detectors and other details are tabulated. Silica gel is the stationary phase used most, while hexane, chloroform, acetone, and other solvents in many different combinations are used as eluents. Detection is by means of reflectance densitometers or characteristic reaction of a single or a class of compounds (e.g. cholinesterase for organophosphates). Rathore and Begum also published an interesting study[100] where different stationary and mobile phases and the relative effects on the separation of a class of compounds (carbamate) were tested.

The advantages of TLC are that it is a simple technique, not requiring sophisticated instruments, so that every laboratory can use it.

3 MYCOTOXINS

3.1 Problem of Mycotoxins in Food

Apart from the problem caused by the widespread use of pesticides and their residues, in commodities (mainly cereals and fruits) infected by a particular species of fungus, the metabolic activity of the mold gives rise to substances known as ''mycotoxins''. Mycotoxins are characterized by their acute toxicity and carcinogenicity which is on the same level as most toxic pesticides. Mycotoxins that are present in commodities stored in moist conditions can injure animals given contaminated feed (e.g. ''turkey's X disease'' in Great Britain or ''porcine nephropathy'' in Denmark) or people (e.g. ''Balkan endemic disease'' in Southern Yugoslavia).

3.2 Different Types of Mycotoxin

The mycotoxins are metabolites originating from particular species of fungi and molds. Their structures are mostly based on condensed heterocycles bearing mainly oxygen as heteroatom and $-OH$, $>C=O$ and esterified functions. In this section the most relevant molecules that are usually sought in food will be briefly described.

3.2.1 Aflatoxins

The aflatoxins are among the most toxic and carcinogenic mycotoxins and are produced by two species of *Aspergillus* (*A. flavus* and *A. parasiticus*). These molecules are derived from heterocycles and their structures are shown here; several species of aflatoxin named B_1 [CAS 1162-65-8] **(1)**, B_2 [CAS 7220-81-7] **(2)**, G_1 [CAS 1165-39-5] **(3)**, G_2 [CAS 7241-98-7] **(4)**, M_1 [CAS 6795-23-9] **(5)** and M_2 [CAS 6885-57-0] **(6)** exist. Other secondary metabolites have also been identified like aflatoxicol, dihydroaflatoxicol and aflatoxins P and Q.

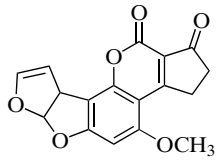

(1) Aflatoxin B_1

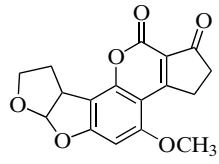

(2) Aflatoxin B_2

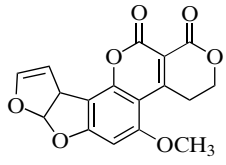

(3) Aflatoxin G_1

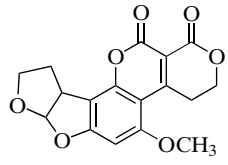

(4) Aflatoxin G_2

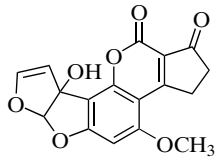

(5) Aflatoxin M_1

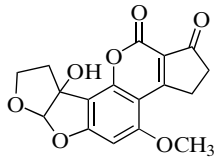

(6) Aflatoxin M_2

Aflatoxins are characterized by fluorescence ranging from green to violet. Aflatoxins B and G are usually found on contaminated moldy cereals, cotton, and peanuts used either as food or as animal feed. Aflatoxins M1 and M2 are metabolites found normally in the milk of animals who have ingested contaminated feed. Species B and G are very toxic with LD_{50} values ranging from less than 0.5 to near 20–30 mg kg^{-1} body weight. These molecules are particularly dangerous for the liver and can also damage the immune and reproductive systems; other effects observed in animals were a reduction in consumption of feed and in production of milk. Carcinogenic effects were observed in experimental animals given contaminated feed at levels ranging from a few to tens of micrograms per kilogram and with induction times from one to two years. Mutagenic activity seems to be in relation to the ability of the molecule to interfere with DNA synthesis.

3.2.2 Ochratoxins

Ochratoxins are toxins produced from *Aspergillus ochraceus* and other species of *Aspergilli* and *Penicilli*; they originate from biosynthesis from phenylalanine and dihydrocoumarin moieties.

There are three types of ochratoxin: ochratoxin A [CAS 303-47-9] **(7)**, ochratoxin B [CAS 4825-86-9] **(8)**

(differing from the species A as not being chlorinated) and ochratoxin C [CAS 4865-85-4] **(9)** (the ethyl ester of species A); all these molecules exhibit a blue-greenish fluorescence.

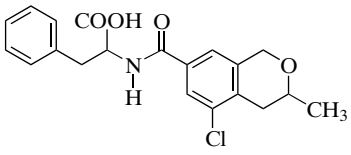

(7) Ochratoxin A

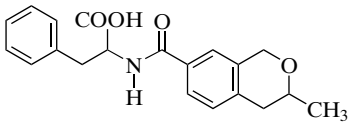

(8) Ochratoxin B

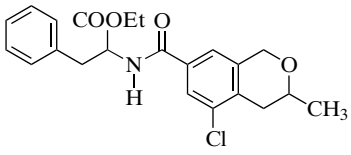

(9) Ochratoxin C

The ochratoxin A **(7)** is well known as a very toxic molecule with LD_{50} values being very close to those characteristic of the aflatoxins; it has been found on coffee grains and moldy cereals. Ochratoxins A **(7)** and B **(8)** are characterized respectively by green and blue fluorescence.

3.2.3 Patulin

Patulin [CAS 149-29-1] **(10)** is produced from *Aspergillus*, *Penicilium* and *Byssochlamis* and affects mainly fruits, where inadequate collecting (tumble shocks) and storage (moisture) conditions favors rotting and consequent mold growth. Apple is particularly attacked by molds producing this toxin and residues of patulin are also found on fruit juices and ciders. The chemical formula of patulin **(10)** is shown.

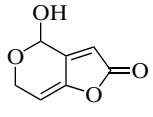

(10) Patulin

At toxic levels patulin can originate gastrointestinal diseases and can also damage the spleen, kidneys and lungs as demonstrated by experimental animals. There is also some evidence of teratogenicity and possible carcinogenicity. Typical LD_{50} values (e.g. for rat) range from 10 to 50 mg kg^{-1} body weight. Its carcinogenicity may originate from the conjugated double bond system present in the molecular structure.

3.2.4 Trichothecenes, Fumonisins and Zearalenone

These molecules are all produced by several species of *Fusarium* and other fungi (*Tricohthecium*, *Stachybotrys* etc.). Trichothecene toxins cover a wide group of molecules (over 40) based on the structure of another mycotoxin, the trichothecin [CAS 6379-69-7]. Trichothecenes may be divided between two main groups: the nonmacrocyclic trichothecenes and the more labile macrocyclic ones that can be considered to be polyesters of simple trichothecenes. Among the best known trichothecenes are nivalenol [CAS 23282-20-4] **(11)**, deoxynivalenol (DON, named also "vomitoxin") [CAS 51841-10-8] **(12)** and 3-acetyldeoxynivalenol **(13)**, neosolaniol [CAS 36519-25-2] **(14)**, T_2 [CAS 21259-20-1] **(15)** and HT_2 [CAS 26934-87-2] **(16)** toxins, verrucarol, scirpentriol **(17)**, diacetoxyscirpentriol (DAS) **(18)**, 15-monoacetoxyscirpentriol **(19)** and fusarenone [CAS 3255-69-8] **(20)**. Among macrocyclic trichotecenes there are the H **(21)** and G **(22)** satratoxins, roridin A **(23)** and verrucarin A [CAS 3148-09-2] **(24)**. Schemes 1 and 2 show the chemical formulas of nonmacrocyclic trichothecenes, whereas Scheme 3 shows chemical formulas of some macrocyclic trichothecenes.

Trichothecenes and zearalenone [CAS 17924-92-4] **(25)** are found on moldy cereals. Generally the trichothecenes affect the immune system so that targeted animals or humans become more subject to infections. Symptoms of severe intoxication in man are vomiting, fever, headache, skin irritations, respiratory disorders and diarrhea.

The nivalenol **(11)** and DON **(12)** vomitoxins are known to cause refusal of food, vomiting and subsequent weight loss together with diarrhea and other gastroenteric diseases in animals given infected feed; in the most severe intoxication damage to kidney, stomach and heart can also occur both in animals and in humans. T_2 **(15)** toxin is very toxic (LD_{50} orders range from near 1 to 5 mg kg^{-1} body weight) and injures the digestive apparatus and reproductive system.

Zearalenone **(25)** is a toxin produced over a quite wide range of temperatures (about 10–30 °C). Its activity is based on estrogenic action causing diseases to male and female genital apparatus. The chemical formula is shown by structure **(25)** and that of zearalenol by structure **(26)**.

Fumonisins (B_1, **(27)**; B_2, **(28)**; B_3, **(29)**) are recently discovered mycotoxins produced from *Fusarium moniliforme*. Fumonisins are now known to cause leukoencephalomalachia in horses or pulmonary edema in swine and is suspected to cause esophageal cancer in humans.

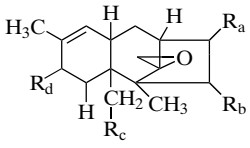

	R_a	R_b	R_c	R_d
Neosolaniol **(14)**	OH	$OCOCH_3$	$OCOCH_3$	OH
Scirpentriol **(17)**	OH	OH	OH	H
Monoacetoxyscirpentriol **(19)**	OH	OH	$OCOCH_3$	H
Diacetoxyscirpentriol **(18)**	OH	$OCOCH_3$	$OCOCH_3$	H
HT_2 Toxin **(16)**	OH	OH	$OCOCH_3$	$OCOCH_2CH(CH_3)_2$
T_2 Toxin **(15)**	OH	$OCOCH_3$	$OCOCH_3$	$OCOCH_2CH(CH_3)_2$
T_2 Triol	OH	OH	OH	$OCOCH_2CH(CH_3)_2$
T_2 Tetraol	OH	OH	OH	OH
Verrucarol **(24)**	H	OH	OH	H

Scheme 1 Nonmacrocyclic trichothecenes.

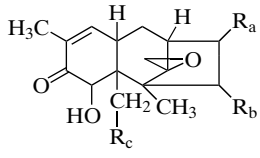

	R_a	R_b	R_c
Fusarenone **(20)**	OH	$OCOCH_3$	OH
Nivalenol **(11)**	OH	OH	OH
Deoxynivalenol **(12)**	OH	H	OH
15-Acetyldeoxynivalenol	OH	H	$OCOCH_3$
3-Acetyldeoxynivalenol **(13)**	$OCOCH_3$	H	OH

Scheme 2 Nonmacrocyclic trichothecenes.

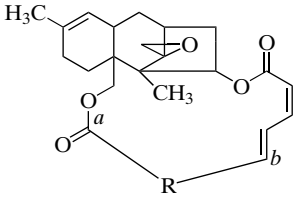

Satratoxin G **(22)**

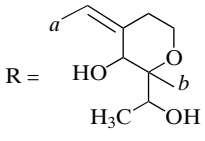

Satratoxin H **(21)**

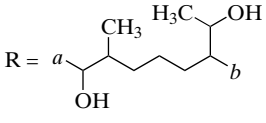

Roridin A **(23)**

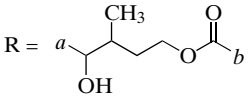

Verrucarin A **(24)**

Scheme 3 Macrocyclic trichothecenes.

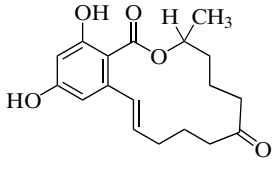

Zearalenone **(25)**

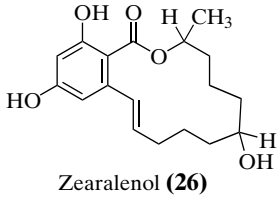

Zearalenol **(26)**

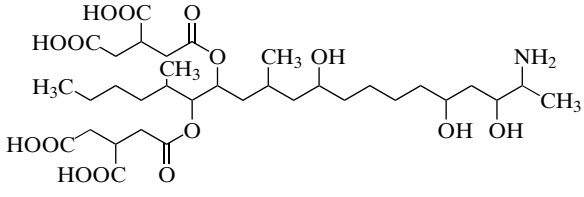

Fumonisin B_1 **(27)**

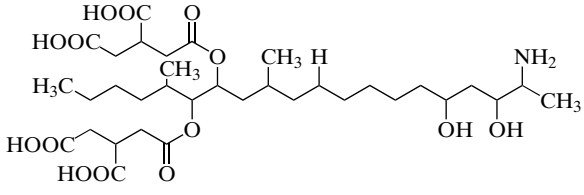

Fumonisin B_2 **(28)**

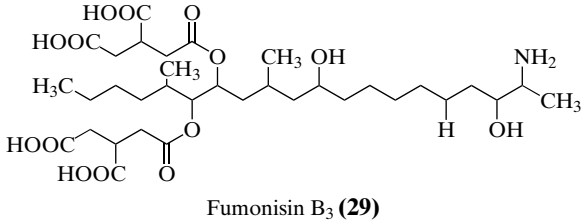

Fumonisin B_3 **(29)**

3.2.5 Toxin from Alternaria *and Others*

Alternaria fungi, in particular *Alternaria alternata* are known to produce, as a result of their metabolism, some

molecules (alternariol [CAS 641-38-3] **(30)**, alternariol methyl ester [CAS 23452-05-3] **(31)**, altenuene **(32)**, tenuazonic acid [CAS 610-88-8] **(33)**) that have shown some evidence of mutagenic and carcinogenic activity towards animals. Molecular structures of these molecules and altertoxin I **(34)** are shown. Mycotoxins from *Alternaria* have been found in degraded vegetables, seeds and cereals.

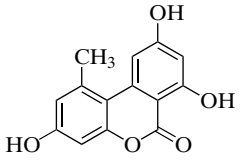

Alternariol **(30)**

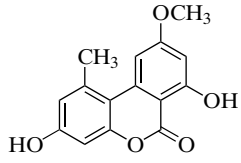

Alternariol methyl ester **(31)**

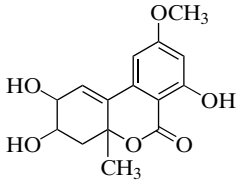

Altenuene **(32)**

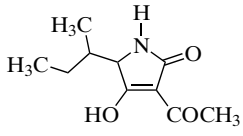

Tenuazonic acid **(33)**

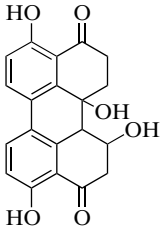

Altertoxin I **(34)**

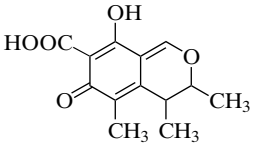

Citrinin **(35)**

Among the other mycotoxins, of noticeable importance are sterigmatocystine and citrinin [CAS 518-75-2] **(35)**. Sterigmatocystine is produced from different species of *Aspergillus* and is believed to be a precursor of the aflatoxins; the molecular structure is similar to that of the aflatoxins and this molecule also exhibits hepatotoxicity, kidney toxicity and carcinogenic activity. Citrinin **(35)** is often found together with ochratoxins in moldy cereals and like ochratoxins it exhibits toxicity towards the kidney.

3.3 Methods of Analysis

Like pesticides, a large number of articles have been published on mycotoxins covering many different experimental conditions both in the preparative step and in instrumental detection and this is reflected in the different values of sensitivity (LODs) reported in different papers.

Official methods are generally available for mycotoxins,[101] furthermore, several reviews or papers covering general[102] or particular topics in mycotoxins residue analysis in food have been published. Reviews concern aflatoxins,[103] analysis of residues in milk,[104] mycotoxins[105] with COOH groups, chromatographic methods[106] for ochratoxin A, fumonisins[107] or others,[108] use of immunoaffinity[109] columns, high-performance TLC[110] for aflatoxins and TLC[111] for mycotoxins, applications with HPLC/MS[74] systems, methods for nonmacrocyclic[112] trichothecenes, and sample preparation.[113,114]

Owing to the extreme toxicity of some mycotoxins, pure standards for aflatoxins and ochratoxins, for example, are available only on the milligram scale with unknown purity; therefore concentration of these standard solutions must be determined by spectrophotometry using extinction coefficients and their purity assessed by TLC, see also AOAC[101] methods 970.44 and 971.22.

3.3.1 Sample Preparation

3.3.1.1 Sample Extraction Owing to the polar characteristics of the molecules, extraction of the mycotoxin from the matrices is usually carried out with polar solvents including water, methanol, acetonitrile, acetone, chloroform and their mixtures. The weighed amounts for extraction vary from 10 g to near 100 g with the obvious effect on the final detection limits. In some cases amounts of the magnitude order of kilograms are extracted, taking into consideration that the toxin distribution in the commodity is very heterogeneous.

In some cases, e.g. dry cereals, the sample is blended with water prior to extraction in order to homogenize it and to make the analyte more extractable.

Typical extraction mixtures for aflatoxins, ochratoxins, fumonisins and trichothecenes from cereals and seed are water–acetonitrile with ratios from 2:3 to 1:9 and water–methanol with ratios from 1:1 to nearly 9:1; acetone with a low percentage water content and water nonmiscible solvents like dichloromethane, chloroform or the less harmful ethyl acetate are also used.[108] For trichothecenes the extractant most used[112] is acetonitrile–water in the ratio 86:14.

Extraction with an aqueous/organic phase needs to be filtered in order to eliminate water-insoluble substances. Several papers report the use of solutions of acetic acid, sodium chloride, and zinc or lead acetate to aid in defatting the extract by precipitating fatty acids as metal soaps. Defatting of the extract by extraction with nonpolar solvents like hexane, heptane or petroleum ether, where polar mycotoxins are practically insoluble, has also been achieved.

In some cases aqueous solutions of sodium hydrogencarbonate (≈1–3%) are used as extractants.[115]

Extraction of aflatoxins from milk[116,117] samples often requires a less complicated preparative step and can be defatted simply by centrifugation, aided by addition of some sodium hydroxide or sodium chloride, and the remaining aqueous sample extracted either by partition with chloroform or by adsorption with SPE or onto an immunodiffusion cartridge.

Methods for ochratoxin[101] often report use of partition with chloroform or methylene chloride and aqueous phosphate buffer.

Applications of direct ochratoxin extraction from beer by passing through a C_{18} SPE cartridge[118] are also reported. A German[119] method extracted ochratoxins from wine using toluene with added magnesium chloride at pH 2.

Patulin from ciders and juices is generally extracted, owing to aqueous nature of the sample, with ethyl acetate[120] which is nonmiscible with water, followed sometimes by partition with sodium hydrogencarbonate aqueous solution.

SFE, as a technique recently introduced and requiring sophisticated and expensive apparatus, is used less. Applications[121,122] are reported for extraction of DON **(12)** and some others trichothecenes operating at 40–60 °C and 300–500 bar with carbon dioxide containing 3–5% methanol on 1–5 g of sample. Extraction can be achieved either statically or dynamically, in less time (e.g. 15 min instead of 30 min). A review concerning uses of SFE in extracting flavors, lipids fragrances and mycotoxins from natural products was published by Modey et al.[123]

3.3.1.2 Sample Cleanup Cleanup onto solid substrates is normally achieved on silica or other substrates like Celite®, Florisil®, or alumina. The sample extract applied onto a silica column is washed with apolar solvents like hexane or diethylether, whereas the elution of the mycotoxin is carried out with chloroform (or dichloromethane) containing some 1–10% of acetone or methanol to increase the polarity of the medium. Widely used multifunctional columns (Mycosep®) consist of different stationary phases with alumina, Celite®, charcoal and ion exchangers which retain substances from the matrix when the water–acetonitrile extract is simply passed through.[103,112,124]

The use of SPE columns involves silica, C_{18} octadecylsilica or phenylsilica functionalized stationary phases, where the applied sample is eluted with chloroform (or dichloromethane) with about 10% acetone; ion-exchange stationary phases (e.g. Amberlite® XAD, strong anion exchange (SAX)) are also used in SPE cleanup.

Immunoaffinity columns (IAC) represent a technique of purification[109] and/or separation and preconcentration which is very selective, based on the interaction between the mycotoxin and a specific toxin antibody linked to a solid substrate. Usually before the cartridge is loaded with the sample (the extract concentrated in methanol, acetone or acetonitrile must be diluted), it is washed with a buffered saline solution (e.g. phosphate buffered saline (PBS) containing a surfactant, such as tween-20). The sample is then adsorbed and washed with PBS buffer, water or aqueous 10% methanol. The mycotoxins are finally eluted with methanol or aqueous methanol (e.g. 80%) solutions or acetonitrile. Some commercial specific cartridges are now available, for example for aflatoxin, ochratoxins, fumonisins and zearalenone. The advantages of this type of cleanup come from its selectivity whereas disadvantages are the high cost of a single column and the nonavailability in commerce of ''multitoxin'' columns performing SPE in the multiresidue methods.

Comparisons between different cleanup and extraction techniques have also been published: e.g. multifunctional columns, IAC and SFE techniques[125] or IAC versus solvent extraction techniques[126] where the latter (with ethyl acetate) showed less capacity for extraction.

Finally, in the same way as already seen for pesticides, the organic extract can also be purified by GPC.[127]

3.3.2 Instrumental Methods

3.3.2.1 Gas Chromatography and Gas Chromatography/Mass Spectrometry Techniques GC is not extensively used in analysis of mycotoxins due to the characteristics of the molecules (relatively high-molecular weight, high-melting point and the presence of strong intermolecular interaction due to polar function like −OH, C=O and −COOH groups) that make them unsuitable for such instrumental analysis. Derivatization techniques must be used in order to make the molecule more volatile so that the temperature and time needed to elute the compounds from the column are not too great.

Most applications found in the literature are for trichothecenes and zearalenone but applications for patulin,[128] alternaria toxins[129] and aflatoxins[130] with ≈6–10-m short columns have also been reported. Heptafluorobutyric acid (HFBA), TFA and heptafluorobutyrylimidazole are typical derivatizing agents that insert fluoro-substituted groups into the molecules by esterification of their OH functions. Silylating reagents like bistrimethylsilylacetamide (BSTFA), trimethylsilylimidazole and trimethylchlorosilane are other derivatizing agents. The latter two in particular are used to ensure complete silylation because of the difficulty of silylating some functions arising from double peaks (two silylated products) in the chromatogram.

A general oven temperature program runs from 80–100 °C up to 140–210 °C at 10–20 °C min^{-1} and then

up to 280–310 °C at various rates from 5–10 °C min^{-1}. Capillary columns up to 50-m long are used but mostly columns 15–30-m long are used. The columns are coated with 100% methylsilicone or containing 5% phenylsilicone, 6–14% cyanopropylphenylsilicone or other phases.

Detectors used are FID, ECD and MS in GC/MS (electron impact (EI), chemical ionization and negative ion). In GC/MS systems it is necessary to operate SIM to achieve the required sensitivity. In the case of mycotoxin, ions of TMS (trimethylsilyl), TFA or HFBA derivatives are usually collected. Derivatives are also characterized by fragments with greater values of m/z so that fewer undergo interference from the matrix. In Table 3 some characteristic ions of toxins or their derivatives are shown.

The FID is not very sensitive and sensitivities of the order of 0.1–1 mg kg^{-1} are reported even with the use of derivatizing agents. More selective and sensitive detectors like ECD or MS allow detection limits in the range 0.01–0.05 mg kg^{-1} to be attained.

3.3.2.2 High-performance Liquid Chromatography and Liquid Chromatography/Mass Spectrometry Techniques HPLC analysis of mycotoxins is usually performed by reversed-phase liquid chromatography; the columns most used are 25-cm long (4.6-mm diameter) and have C_{18} octadecylsilane functionalized silica (ODS) as stationary phase. Columns used to a lesser extent are C_8 columns and 10, 12.5 and 15 cm-long columns. Methanol–water (40–80% methanol) or methanol–acetonitrile–water in various ratios are used as mobile phases also with aqueous phase containing 1–5% acetic acid, phosphate buffers or other buffers depending on the method.

Detection is mostly based on the fluorescence properties of the molecule (as is the case for example for aflatoxins, ochratoxins and fumonisins) by fluorimetric detectors working in the range 330–380 nm for excitation wavelength and in the range of 420–460 nm for emission wavelength; in other cases UV detection is used with DADs. Derivatization is often performed in order to enhance the detection, either by precolumn derivatization with TFA, boron trifluoride, phenilendiamine, OPA (for fumonisins) or by postcolumn reaction with I_2 or Br_2. Some authors studied the effect of detection enhancement originated by cyclodextrins.[131] Limits of detection are strongly related to the preparative route in addition to the instrumental conditions used, in particular the amounts of sample extracted and the final volume of the extract. Limits of detection ranging from a few nanograms per kilogram (for aflatoxins) to micrograms per kilogram for less detectable fumonisins, trichothecenes and zearalenone. For patulin LODs of tenths of micrograms per kilogram are normally reported.

Chromatograph from HPLC have been interfaced with mass spectrometers by means of apposite devices that eliminate a large amount of interference due to the solute being in the mobile phase (e.g. buffers, salts). Electrospray and thermospray devices are usually used as interfaces, electrospray being preferred as they give more abundant molecular ions than do thermospray giving more fragmentation with APCI or EI. Ionization in the negative ion mode gives better sensitivities than ionization with positive ions.[113] As in GC/MS systems, operating in SIM allows greater sensitivity; in Table 3 some examples of ions that characteristically have toxins as parent molecules or as derivatives in different LC/MS or LC/MS/MS systems are reported.

3.3.2.3 Capillary Electrophoresis Techniques Relatively few papers report on this topic. Separation of mycotoxins is generally performed on 50–75-cm long capillary silica columns with diameters ranging from 50 to 75 μm. The potentials applied vary from 10 to 20 kV with 50–70-cm distances from anode to detector. Detection methods used are UV/VIS/DAD, conventional and laser-induced[141] fluorescence. Derivatization and hydrodynamic system of injection at high pressures are used for more sensitive determinations.

3.3.2.4 Other Techniques

Thin-layer Chromatography. A multiresidue method[142] determining 22 mycotoxins simultaneously by TLC has been published. TLC is a quite simple method that does not require sophisticated instruments but, owing to the complexity of the sample and the low levels investigated, still requires the preparative step of cleanup of the sample like those described above. TLC separations are mostly performed with silica as stationary phase whereas a very great number of solvent mixtures and reagents are used respectively as eluents and as detecting agents.[111] Quantitation is made by spotting different amounts of sample and standards on the plate and matching the intensity of the respective spots to assess which spots exhibit the same intensity of response.

The review by Ling et al.[111] also compares TLC with other methods. TLC is a very simple technique having sensitivity and precision similar to the HPLC method whereas ELISA is less reproducible.

3.3.3 Immunochemical Methods

Radioimmunoassay (RIA) is a technique not widely adopted in mycotoxin analysis,[143] while ELISA tends to substitute for RIA because the former is a simpler and safer method (handling of harmful radioactive substances is avoided). ELISA measures mycotoxin levels by competition of the analyte and its analogs conjugatedwith an

Table 3 Examples of selected ions for mycotoxins in MS, GC/MS and LC/MS determinations

Toxin	Technique	Derivatives	Selected ions[a] (m/z)	Ref.
Aflatoxin B_1	Direct probe MS-injection	–	297–311	101
Aflatoxin B_1	Direct probe		227–256–284–312	132[b]
Aflatoxin G_1	MS-injection		328	
Aflatoxin G_2	GC/MS, EI		287–330	
Aflatoxin M_1			271–299–328	
Aflatoxin M_2			246–273–330	
Ochratoxin A			228–239–255	
Ochratoxin A		methyl ester	221–239–255	
Ochratoxin B		methyl ester	187–205–221–383	
Patulin		acetate	136–137–154–196	
Alternariol		monomethyl ether	199–201–243–272	
Diacetylnivalenol			179–247–336	
DON			248–249	
DON		acetate	173–223–290–338	
Zearalenone–monomethoxy-zearalenone		TMS	247–275–293–386–404	
		dimethyl derivative	189–217–346	
Ochratoxin A	HPLC/MS/MS Electrospray positive	TFA	239–404–538	133
Patulin	GC/MS	hexafluorobutyrate	97–110–125–136–153	128
Patulin	HPLC/MS, EI	–	110–125–126–136–154	134
Fumonisins B_1	HPLC/MS/APCI Electrospray	–	723	135
Fumonisin B_1	HPLC/MS/MS		334–353–704–(precursor: 722)	136
Fumonisin B_2	Ion spray-API		318–337–355–(precursor: 706)	
DON	GC/MS, EI	–	135–248	137
Fusarenone-x			179–306	
3-Acetyl-DON			223–290	
DAS			278–306	
T_2-Toxin			180–364	
Zearalenone			188–312	
DON	GC/MS, EI	TMS	103–235–512	138[c]
7-Acetyl-DON			103–289–392–482	
Fusarenone-x			103–480–570	
3-Acetyl-DON			117–392–482	
15-Acetyl-DON			193–235–392–482	
Nivalenol			191–289–600	
3,15-Diacetyl-DON			320–540	
4,15-Diacetyl-DON			117–408–540	
T_2-Toxin	HPLC/GC/MS/MS	–	245–327–387–489	139[d]
T_2-Toxin-triol	Electrospray		303–334–405	
T_2-Toxin-tetraol	Ion trap		260–263–291–321	
HT_2-toxin			285–345–447	
Verrucarol			245–259–274–276–289	
Verrucarin A			245–327–387–489	
DAS			247–329–389	
Satratoxin G			492–493–523–567	
Satratoxin H			451–468–523	
Roridin A			279–325–527–555	
Reserpine			397–448–609	
Zearalenone	HPLC/MS/APCI Negative ion and positive ion	–	279–301–319–391 (p.ion) 317 (neg.ion)	140

Table 3 (*continued*)

Toxin	Technique	Derivatives	Selected ions[a] (m/z)	Ref.
Alternariol	GC/MS	bis/tris TMS	402 (bis)–459 (tris)	129[e]
		bis/tris HFB	650 (bis)–845 (tris)	
Alternariol methyl ester		mono/bis TMS	344 (mono)–401 (bis)	
		mono/bis HFB	468 (mono)–664 (bis)	
Altenuene		bis/tris TMS	421 (bis)–493 (tris)	
		bis/tris HFB	682 (bis)–880 (tris)	
Altertoxin I		tetra HFB	1100	
		bis/tris/tetra TMS	496 (bis)–568 (tris)–625–640 (tetra)	
Tenuazonic acid		bis/tris TMS	326–341 (bis)–398–413 (tris)	
(CuII-salt)		bis/tris HFB	589 (bis)–785 (tris)	

[a] Selected ions indicated by the authors of the cited papers or by the authors of the present article.
[b] Adapted from Wiley 130 K Mass Spectral Database.[132]
[c] Adapted from Mirocha et al.[138]
[d] Adapted from Tuomi et al.[139]
[e] Adapted from Scott et al.[129]
APCI, atmospheric pressure chemical ionization; API, atmospheric pressure interface; HFB, hexafluorobutyl.

enzyme molecule, for the site on an antibody immobilized on a solid support (membrane, test tube, microtiter well plate and so on). The enzymatic activity is related to the toxin–enzyme conjugate present on the antibody, whose presence is also related to the free mycotoxin (the analyte) present. The enzymatic activity is measured simply, after washing the solid support with a saline buffered solution (also added with a surfactant, e.g. Tween), by adding the enzyme substrate and a chromogenic solution. A typical enzyme is HRP with a chromogen like *o*-phenylendiamine or tetramethylbenzidine; the absorbance of the solution is read on a spectrophotometer at 440–490 nm. These methods are now also available in commercial prepared kits and are as sensitive as RIAs but simpler and cheaper and allow the operator to avoid handling radioactive isotope solutions. Sensitivities for ELISA range from less than 0.1 to 50 $\mu g\,L^{-1}$.

4 QUALITY CONTROL

4.1 Validation of Method and Internal Laboratory Quality Control

Analytical methods can be regarded as transformation processes where the end products are the data. The quality of data and thus reliability of the methods must be assured by a series of experimental steps in what is called ''validation''. Some aspects must be characterized and some parameters must be defined experimentally as stated in the EN45001[144–147] regulation.

The linearity of the analytical response of the method must be checked by increasing the analyte concentration. In doing this, the fact that the different detectors may not have a linear response over the usual range covered by the residues (more or less than 0.001–10 $mg\,kg^{-1}$) must be taken into account. The first step is to study the instrumental response, followed by a more severe approach which includes the preparative step in the linearity study so that the overall measured linearity is accurate with a precision calculated at different (usually from three to five) concentration levels of the analyte. All such work takes a great deal of time and effort especially in the case of a multiresidue method, and thus must be adequately planned.

The accuracy is assessed either by determining recoveries on added matrices in the laboratory or less frequently by analyzing reference materials, whereas precision is assessed by repeating the analysis on the same sample from five to eleven times. The preparative route for this type of analysis greatly affects the data in terms of accuracy and precision. Precision[148] depends on the potential of many random errors at every step; accuracy depends on sample losses from the several preparative (negative) steps or from matrix effects (positive). Usually accuracy and precision are determined at the same time by repeated analysis on fortified matrices. In the field of pesticide residues reference materials are rare and expensive due to complexity of preparation and instability, on the other hand, in the field of mycotoxins, reference materials, although expensive, are very important.

A reference material for residues analysis must always be stored at low temperature (−25 °C). Other methods of checking accuracy and precision are by participating in proficiency tests and collaborative/cooperative trials. Recently some organizations that have performed proficiency tests make test materials available after each set have been performed and the average consensus value for added analytes (normally from 3 to 6) has been assessed.

The identification of a compound must always be confirmed by another detection technique or another method. GC/MS and LC/MS or at least HPLC/DAD (with acquisition of the whole UV spectrum) are the

instruments used most. If SIM/GC/MS (or SIM/LC/MS) is used, the acquisition of the whole mass spectrum or at least three characteristic ions with m/z values >100 (when possible) is recommended when residues found exceed the active limits.

Quantitation must be performed with appropriate calibration with at least two calibrating points ''bracketing'' the concentration found in the sample; if single point is used, it should not exceed ±25% of the value of the sample. Calibration must also take into account matrix effects described above. Limits of detection must be checked with respect to recovery from the matrix and not only with respect to instrumentation.

Blanks of the entire procedure are recommended because the quality of materials (solvents, salts) may not be constant at the time and fine cleaning of the glassware with pure solvents must be assured before each analysis.

Instrument calibration and performance must be checked at regular intervals (e.g. once for each group of samples analyzed) following manuals and good laboratory practice principles; instruments used in preparation must be also controlled, e.g. GPC cleanup systems must be checked for recoveries and cleanup power by eluting special sample mixtures or by simply testing recovery of some pesticides (chosen between those with higher and lower retention volumes) and their purification from the matrix.

It is recommended that both the laboratory and the instruments used in preparation be dedicated to this kind of analysis; the original samples and their extracts must be kept separate from the pure standards and their concentrated and diluted solutions. All items must be stored at a low temperature in a refrigerator or freezer. Studies concerning standard solutions stability with respect to solvent and storage conditions used have been published.[149]

Control charts are not easy to prepare for such a complex analysis, however the performance of the method can be checked by inserting into each sample batch of 20–30 analyzed samples a blank sample and a fortified sample to check for interferences and analysis recoveries. For common pesticides it is recommended to check the recovery in every sample batch frequently, whereas for rare pesticides a yearly recovery test is adequate. It may be the case that for pesticides of the same class and which exhibit the same behavior, a recovery test performed on selected molecules may be indicative of all the pesticides in the class.

4.2 Interlaboratory Quality Control

Quality control programs performed by a laboratory undoubtedly greatly enhance the reliability of the data produced from that laboratory; however interlaboratory tests present a more powerful system for assessing the ability of one or more laboratories to produce true data. Such trials acquire importance particularly with respect to official programs for monitoring residues to establish knowledge about the contamination status of food and the undesirable persistence of some active molecules.

Among interlaboratory tests are intercalibrations, collaborative/cooperative studies and proficiency tests.

Intercalibrations may even be simple exercises where matrices which have known substances added at unknown level (or even standard solutions in the simplest trials) are distributed among several laboratories. On the basis of the data reported, the accuracy and the precision of the testing group of laboratories can be established. Several examples have been reported.[150,151]

More sophisticated studies are collaborative (several laboratories using the same methods) and cooperative (several laboratories using their own methods) performed by each laboratory to give a series of samples at different fortification levels, in duplicate. Each trial may be repeated a number of times during a specific period (e.g. six months, one year). In this case too, known compounds but at unknown levels are quantified, see cited examples.[120,124,152–155]

On the other hand proficiency tests are spot exercises where the ability both to identify given compounds to quantify them is evaluated. In these exercises a fortified matrix with between four and six unknown compounds at unknown levels, chosen from a list of some 20–50 known compounds, is distributed to several laboratories. These have to discover which pesticides are present and in what relative amounts. Results presented by each laboratory are compared with the medium values or the robust consensus mean value obtained from all the laboratories and the difference from it (z-score) is expressed in terms of Horwitz's standard deviation[156] at that concentration level. Normally performances are considered acceptable where difference from the consensus means not exceeding two (or also three) times the Horwitz's standard deviation. Several examples have been reported.[157,158]

A survey carried out on the results of many interlaboratory trials showed that although the averaged interlaboratory reproducibility is located between 15% and 20%, the distribution of single results recorded in each exercise can range[159,160] from less than 5% up to 50%. Proficiency testing is recommended in European Community guidelines[161] for pesticide residual analysis.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Interface

ASE	Accelerated Solvent Extraction
BSTFA	Bistrimethylsilylacetamide
DAD	Diode Array Detector
DAS	Diacetoxyscirpentriol
DCBF	Dichlorobenzophenone
DDD	1,1-Dichloro-2,2-bis[*p*-chlorophenyl]-ethane
DDE	1,1-Dichloro-2,2-bis[*p*-chlorophenyl]-ethylene
DDT	1,1,1-Trichloro-2,2-bis(*p*-chloro-phenyl)ethane
DON	Deoxynivalenol
ECD	Electron Capture Detector
EI	Electron Impact
ELISA	Enzyme-linked Immunosorbent Assay
ETU	Ethylenethiourea
FAO	Food Agriculture Organization
FDA	Food and Drug Administration
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
GAP	Good Agricultural Practice
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
HCH	Hexachlorocyclohexane
HFBA	Heptafluorobutyric Acid
HPLC	High-performance Liquid Chromatography
HRP	Horseradish Peroxidase
IAC	Immunoaffinity Columns
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
JMPR	Joint Meeting for Pesticide Residues
LC/MS	Liquid Chromatography/ Mass Spectrometry
LLP	Liquid–Liquid Partition
LOD	Limit of Detection
MRL	Maximum Residue Limit
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NPD	Nitrogen Phosphorus Detector
OC	Organochlorine
OPA	*o*-Phthalaldehyde
PBS	Phosphate Buffered Saline
PTU	Propylenethiourea
RIA	Radioimmunoassay
SAX	Strong Anion Exchange
SCOD	Sweep Codistillation
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SPMD	Solid-phase Matrix Dispersion
SPME	Solid-phase Microextraction
TEPP	Tetraethyl Pyrophosphate
TFA	Trifluoroacetic Acid
THI	Tetrahydrophthalimide
TLC	Thin-layer Chromatography
TMS	Trimethylsilyl
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VOCs	Volatile Organic Compounds
WHO	World Health Organization

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Dioxin-like Compounds, Screening Assays • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Polychlorinated Biphenyls Analysis in Environmental Samples • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicide Residues in Biota, Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and

Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Gradient Elution Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. G. Tyler Miller, Jr, 'Pesticides and Pest Control', in *Living in the Environment*, 7th edition, Wadsworth Publishing, Belmont, CA, USA, 633–650, 1992.
2. I. Mukherjee, M. Gopal, 'Organochlorine Residues in Dairy Milk in and Around Delhi', *J. AOAC Int.*, **76**(2), 283–286 (1993).
3. S.M. Waliszewski, V.T. Pardio, K.N. Waliszewski, J.N. Chantiri, A.A. Aguirre, R.M. Infanzon, J. Rivera, 'Organochlorine Pesticide Residues in Cows Milk and Butter in Mexico', *Sci. Total Environ.*, **208**(1–2), 127–132 (1997).
4. Codex Alimentarius Commission, Joint FAO, 'Codex Alimentarius, Vol 2: Pesticide Residues in Food', WHO Food Standard Programme, Rome, Italy, 1993.
5. J. Gilbert, 'Sampling and Analysis of Ochratoxin A in Food', *Food Addit. Contam.*, **13**, 1317–1318 (1996).
6. T.B. Whitaker, J. Springer, P.R. Defize, W.J. DeKoe, R. Coker, 'Evaluation of Sampling Plans used in the United States, United Kingdom and The Netherlands to Test Raw Shelled Peanuts for Aflatoxins', *J. AOAC Int.*, **78**(4), 1010–1018 (1995).
7. M. Hussain, F. Vojir, 'Sampling Plan for the Control of Aflatoxin B1 in Imported Dried Figs', *Dtsch-Lebensm-Rundsch*, **89**(12), 379–383 (1993).
8. W. Horwitz, 'Sampling and Preparation of Sample for Chemical Examination', *J. AOAC Int.*, **71**(2), 241–245 (1988).
9. A. Ambrus, 'Estimation of Uncertainty of Sampling for Analysis of Pesticide Residues', *J. Environ. Sci. Health, B*, **31**(3), 435–442 (1996).
10. C.D.S. Tomlin, *The Pesticide Manual*, 11th edition, The British Crop Protection Council, 1997.
11. J. Sherma, 'Pesticides', *Anal. Chem.*, **67**(12), 1R–20R (1995).
12. K. Fodor-Csorba, 'Chromatographic Methods for Determination of Pesticides in Food', *J. Chromatogr.*, **624**, 353–367 (1992).
13. C.M. Torres, Y. Picó, J. Mañes, 'Determination of Pesticide Residues in Fruit and Vegetables', *J. Chromatogr., A*, **754**, 301–331 (1996).
14. N. Motohashi, H. Nagashima, C. Pàrkànyi, B. Subrahmanyam, G. Zhang, 'Official Multiresidue Methods of Pesticide Analysis in Vegetables Fruits and Soils', *J. Chromatogr., A*, **754**, 333–346 (1996).
15. 'US Food Drug Administration', *FDA Pesticide Analytical Manual*, Volume 1, Base Manual part 1 and 2 – October 1991 and Volume 2, Base Manual part 1 to 5 – October, 1991, Washington, DC, 1992.
16. P.A. Greve (ed.), *Analytical Methods for Residues of Pesticides in Foodstuffs*, 5th edition, Government Publishing Office, The Hague, Netherlands, 1988.
17. L.D. Sawyer, B. McMahon, W.H. Newsome, 'Chapter 10: Pesticides and Industrial Chemical Residues', in *Official Methods of Analysis of AOAC International*, 16th edition, ed. P. Cunnif, Arlington, Virginia, Vol. I, 1995.
18. Istituto Superiore di Sanità, Gruppo di Lavoro per i Residui di Antiparassitari della Commissione Perm. Coord. Interregion. Controllo Ufficiale Prodotti Alimentari, 'Metodi Multiresiduo per l'analisi di Residui di Antiparassitari in Prodotti Vegetali', eds. A. DiMuccio, T. Generali, D. Attard-Barbini, Rapporti ISTISAN 97/23, Rome, 1997.
19. UNICHIM, 'Manuale N.118 residui dei pesticidi', Vol. 1, 1981, Vol. 2, 1983 and Vol. 3, 1989.
20. 'Methodensammlung zur Ruckstandsanalytik von Pflanzenschutzmittel', Lieferung Verlag Chemie Gmbh, Weinheim, Germany, 1–11, 1991.
21. 'Pesticide Analytical Methods in Sweden, Part 1', National Food Administration Report 17/98, Uppsala, Sweden, 1998.

22. Health and Welfare – Health Protection Branch, 'Manual on Analytical Methods for Pesticide Residues in Foods', Ottawa, Canada, 1985.
23. G. Zweigg, J. Sherma, 'Analytical Methods for Pesticides, Plant Growth Regulators and Food Additives', Academic Press, NY, Vol. VII to Vol. XVII, 1974–1992.
24. T. Cairns, J. Sherma, *Emerging Strategies for Pesticide Analysis*, CRC Press, Boca Raton, 1992.
25. J. Sherma, T. Cairns, *Comprehensive Analytical Profiles of Important Pesticides*, CRC Press, Boca Raton, 1992.
26. J. Tekel, J. Kovačičovà, 'Chromatographic Methods in the Determination of Herbicide Residues in Crops, Food and Environmental Samples', *J. Chromatogr. A*, **643**, 291–303 (1993).
27. H.B. Wan, M.K. Wong, 'Minimization of Solvent Consumption in Pesticide Residue Analysis', *J. Chromatogr. A*, **754**, 43–47 (1996).
28. S.S. Yang, A.I. Goldsmith, I. Smetena, 'Recent Advances in the Residue Analysis of N-methylcarbamate Pesticides', *J. Chromatogr. A*, **754**, 3–16 (1996).
29. I. Mukherjee, M. Gopal, 'Chromatographic Techniques in the Analysis of Organochlorine Pesticide Residues', *J. Chromatogr. A*, **754**, 33–42 (1996).
30. Z.M. Chen, Y.H. Wang, 'Chromatographic Methods for the Determination of Pyrethrin and Pyrethroid Pesticide Residues in Crops, Foods and Environmental Samples', *J. Chromatography A*, **754**, 367–395 (1996).
31. A. Ambrus, E.M. Solymosne, I. Korsos, 'Estimation of Uncertainty of Sample Preparation for the Analysis of Pesticide Residues', *J. Environ. Sci. Health, Part B*, **31**(3), 443–450 (1996).
32. P. Johnson, D.A. Rimmer, R.H. Brown, 'Adaptation and Application of a Multiresidue Method for the Determination of a Range of Pesticides, Including Phenoxy Acid Herbicides in Vegetation, Based on High-resolution Gel Permeation Chromatographic Cleanup and GC Analysis with MS Selective Detection', *J. Chromatogr. A*, **765**, 3–11 (1997).
33. M. Anastassiades, W. Schwack, 'Analysis of Carbendazim, Benomyl, Thiophanate Methyl and 2,4-dichlorophenoxyacetic Acid in Fruits and Vegetables after Supercitical Fluid Extraction', *J. Chromatogr. A*, **825**, 45–54 (1998).
34. A. Ambrus, J. Lantos, E. Visi, C. Csatlos, L. Sarvari, 'General Method for Determination of Pesticide in Samples of Plant Origin, Soil and Water. I. Extraction and Cleanup', *J. Assoc. Off. Anal. Chem.*, **64**, 733–742 (1981).
35. A. Kaufmann, 'Fully Automated Determination of Pesticides in Wine', *J. AOAC Int.*, **80**(6), 1302–1307 (1997).
36. P.T. Holland, D.E. McNaughton, C.P. Malcolm, 'Multiresidue Analysis of Pesticides in Wines by Solid-phase Extraction', *J. AOAC Int.*, **77**(1), 79–86 (1994).
37. I. Liška, K. Bìlìkovà, 'Stability of Polar Pesticides on Disposable Solid-phase Extraction Precolumns', *J. Chromatogr. A*, **795**, 61–69 (1998).
38. P. Branca, P. Quaglino, 'Determinazione Rapida di Pesticidi Organofosforati e Diserbanti Triazinici in Acque e Alimenti', *Boll. Chim. Igien. Parte Sci.*, **40**, 71–78 (1989).
39. P. Branca, G. Sacchero, 'Solid Phase Extraction (SPE) of Pesticide Residues in Fruits Followed by GC/MS Determination', *Ind. Alimentari*, **36**(358), 482–488 (1997).
40. C. Franchi, A. Severi, 'Comportamento dell'Amitraz e dei Suoi Prodotti di Degradazione nel Miele', in A.A.M.DelRe, Proceedings of 7° Simposio Chimica degli Antiparassitari, Piacenza, 8–9 June, 1989, 89–96.
41. M. Barbina Taccheo, M. DePaoli, C. Spessotto, 'Determination of Total Amitraz Residues in Honey by Electron Capture Capillary Gas Chromatography – a Simplified Method', *Pestic. Sci.*, **23**, 59–64 (1988).
42. M. Barbina-Taccheo, M. DePaoli, C. Spessotto, 'The Determination of Coumaphos Residues in Honey by HPTLC with in situ Fluorimetry', *Pestic. Sci.*, **25**, 11–15 (1989).
43. L. Kadenczky, Z. Arpad, A. Ambrus, L. Gyorfi, G. Reese, W. Ebing, 'Column Extraction of Residues of Several Pesticides from Fruits and Vegetables: A Simple Multiresidue Analysis Method', *J. AOAC Int.*, **75**, 53–61 (1992).
44. F. Sicbaldi, A. Sarra, G.L. Copeta, 'Diatomaceous Earth-assisted Extraction for the Multiresidue Determination of Pesticides', *J. Chromatogr. A*, **765**, 23–30 (1997).
45. A. Franchi, L. Bensi, S. Viti, F. Abballe, 'Applicazione della Tecnica di Estrazione Accelerata con Solvente (ASE) nell'Analisi dei Residui di Antiparassitari in Alimenti di Origine Vegetale', *Boll. Chim. Igien.*, **49**, 119–125 (1998).
46. M. Baldi, L. Bazzani, I. Scaroni, S. Coppi, M.L. Davi, S. Benedetti, A. Boloventa, L. Penazzi, M.P. Previati, 'Considerazioni sui Risultati delle Analisi Effettuate nel Triennio 1987–1989 per la Determinazione dei Residui di Fitofarmaci Utilizzati in Post-Raccolta su Mele e Pere', *Boll. Chim. Igien.*, **42**, 1–13 (1991).
47. T.M. Pawlowsky, C.F. Poole, 'Extraction of Thiabendazole and Carbendazim from Foods Using Pressurized Hot (Subcritical) Water for Extraction: A Feasibility Study', *J. Agric. Food Chem.*, **46**, 3124–3132 (1998).
48. L.T. Taylor, 'Supercritical Fluid Extraction', J. Wiley & Sons, New York, 1997.
49. S.J. Lehotay, 'Supercritical Fluid Extraction of Pesticides in Food', *J. Chromatogr. A*, **785**, 289–312 (1997).
50. V. Camel, A. Tambutè, M. Caude, 'Analytical Scale Supercritical Fluid Extraction: A Promising Technique for the Determination of Pollutants in Environmental Matrices', *J. Chromatogr.*, **642**, 263–281 (1993).
51. J.W. King, M.L. Hopper, R.G. Lutchefeld, S.L. Taylor, W.L. Orton, 'Optimization of Experimental Conditions for the Supercritical Carbon Dioxide Extraction of Pesticide Residues from Grains', *J. AOAC Int.*, **76**(4), 857–864 (1993).

52. R.K. Juhler, 'Supercritical Fluid Extraction of Pesticides from Meat: A Systematic Approach for Optimization', *Analyst*, **123**, 1551–1556 (1998).
53. S.J. Lehotay, A. Valverde Garcia, 'Evaluation of Different Solid Phase Traps for Automated Collection and Cleanup in the Analysis of Multiple Pesticides in Fruits and Vegetables after Supercritical Fluid Extraction', *J. Chromatogr. A*, **765**, 69–84 (1997).
54. US EPA Method 3454, Federal Register, Vol. 60, No. 142, 37974, 1995.
55. J. Tekel, Š. Hatrìk, 'Pesticide Residue Analyses in Plant Material by Chromatographic Methods: Cleanup Procedures and Selective Detectors', *J. Chromatogr. A*, **754**, 397–410 (1996).
56. A. DeKok, M. Hiemstra, 'Optimization, Automation and Validation of the SPE Cleanup and On-line Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides in Fruits and Vegetables', *J. AOAC Int.*, **75**(6), 1063–1072 (1992).
57. M.J. Page, M. French, 'Determination of *N*-methyl Carbamate Insecticides in Vegetables Fruits and Feeds Using SPE Cleanup in the Normal Phase', *J. AOAC Int.*, **75**(6), 1073–1083 (1992).
58. C. Bicchi, F. Belliardo, L. Cantamessa, G. Gasparini, M. Icardi, E. Sesia, 'Simultaneous Determination of Benzimidazole Fungicides by HPLC on Apples, Pears and their Pulps', *Pestic. Sci.*, **25**, 355–360 (1989).
59. A. DiMuccio, R. Dommarco, D. Attard-Barbini, A. Santilio, S. Girolimetti, A. Ausili, M. Ventriglia, T. Generali, L. Vergori, 'Application of Solid Phase Partition Cartridges in the Determination of Fungicide Residues in Vegetable Samples', *J. Chromatogr. A*, **643**, 363–368 (1993).
60. J. Filion, R. Hindle, M. Lacroix, J. Selwin, 'Multiresidue Determination of Pesticide in Fruit and Vegetables by Gas Chromatography–Mass-selective Detection and Liquid Chromatography with Fluorescence Detection', *J. AOAC Int.*, **78**(5), 1252–1266 (1995).
61. SUPELCO, 'Solid Phase Extraction of Pesticides from Fruits and Vegetables for Analysis by GC or HPLC', *Supelco Bulletin 900*, 1–6 (1996).
62. J. Frank, R. Wagner, M.K. Hennessy, J.L. Obkrasinski, Jr, 'Screening of Organochlorine Pesticide and Polychlorinated Biphenil Residues in Nonfatty Seafood Products by Tandem Solid Phase Extraction Cleanup', *J. AOAC Int.*, **77**(1), 102–106 (1994).
63. F. Bordet, D. Inthavong, J. Mallet, L. Maurice, 'Analyse des Trace de Pesticides Organo-chlorés et de Polychloro-biphényles dans les Aliments D'origine Animale. Methode Rapide Multiresidus. Étude de la Fidélité et Amelioration de la Purification', *Analusis*, **24**, 328–333 (1996).
64. W. Specht, M. Tillkes, 'Gas-chromatographische Bestimmung von Rueckstanden an Pflanzenbehandlungsmitteln nach Cleanup uber Gel-chromatographie und Mini-kieselgel-saulen-chromatographie', *Fresenius' J. Anal. Chem.*, **322**, 443–455 (1985).
65. J.A. Ault, C.M. Schofield, L.D. Johnson, R.H. Waltz, 'Automated Gel-permeation Chromatographic Preparation of Vegetables, Fruits and Crops for Organophosphate Residues Determination', *J. Agric. Food Chem.*, **27**, 4 (1979).
66. M. Baldi, S. Coppi, M.C. Cristofori, M. Davì, S. Benedetti, A. Boloventa, L. Penazzi, M.P. Previati, 'Applicazione della Cromatografia a Permeazione di gel nella Purificazione di Estratti Organici per la Determinazione di Contaminanti a Livello Residuale', *Boll. Chim. Igien.*, **41**(S4), 295–324 (1990).
67. R. Eisert, K. Levsen, 'Solid Phase Micro-extraction Coupled to Gas Chromatography: A New Method for the Analysis of Organics in Water', *J. Chromatogr. A*, **795**, 143–157 (1996).
68. A. Boyd-Boland, S. Magdic, J. Pawliszyn, 'Simultaneous Determination of 60 Pesticides in Water Using SPME and GC/MS', *Analyst*, **121**, 929–938 (1996).
69. M. Guidotti, 'Ricerca di Residui di Fitofarmaci in Succhi di Mela Mediante SPME e GC/MS', *Boll. Chim. Igien.*, **48**, 43–54 (1997).
70. M. Montury, 'La SPME pour les Résidus de Pesticides – le Taux de Recouvrement', Communication, at 'International Seminar 1999 – Pesticides in Food in Mediterranean Countries', Cagliari, April 29–30, 1999.
71. M. Volante, M. Cattaneo, M. Bianchi, G. Zoccola, 'Some Applications of Solid Phase Micro-extraction in the Analysis of Pesticide Residues in Food', *J. Environ. Sci. Health*, **B33**, 179–292 (1998).
72. R. Reupert, E. Plöger, G. Brausen, 'HPLC Determination of 29 Controlled Herbicides in Water Supplies – Application Note', Hewlett Packard Seminar, 'Analisi di Inquinanti Nelle Acque', Cernusco sul Naviglio (Milan), March 18, 1994.
73. R. Schuster, A. Gratzfeld-Husgen, 'A Comparison of Pre- and Post-column Sample Treatment for the Analysis of Glyphosate', Hewlett Packard Seminar, 'Analisi di inquinanti nelle acque', Cernusco sul Naviglio (Milan), March 18, 1994.
74. M. Careri, A. Mangia, M. Musci, 'Applications of Liquid Chromatography–Mass Spectrometry Interfacing Systems in Food Analysis: Pesticide Drugs and Toxic Substance Residues', *J. Chromatogr. A*, **727**(2), 153–184 (1996).
75. S. Schachterle, C. Feigel, 'Pesticide Residue Analysis in Fresh Produce by Gas Chromatography–Tandem Mass Spectrometry', *J. Chromatogr. A*, **754**, 411–422 (1996).
76. J.F. Sabot, 'Concepts et équivoques en Diluition Isotopique Stable. Considérations Mathématiques Fondamentales', *Analusis*, **22**, 381–391 (1994).
77. D.R. Erney, T.M. Pawlowsky, 'Matrix Induced Peak Enhancement of Pesticides in Gas Chromatography: Is There a Solution?', *J. High Resolut. Chromatogr.*, **20**, 375–378 (1997).

78. M. Volante, R. Guida, 'L'effetto Matrice nell'analisi GC', *Laboratorio Duemila*, **10**(1), 60–63 (1996).
79. J. Haišlovà, K. Holadovà, V. Kocourek, J. Poustka, M. Godula, P. Cuhra, M. Kempnỳ, 'Matrix-induced Effect – a Critical Point in the Gas Chromatographic Analysis of Pesticide Residues', *J. Chromatogr.*, **800**(2), 283–295 (1998).
80. V. Kocourek, J. Hajšlovà, K. Holadovà, J. Poutska, *J. Chromatogr. A*, **800**(2), 297–304 (1998).
81. *Reviews of Environmental Contamination and Toxicology*, Springer Verlag, New York.
82. Joint Meeting for Pesticide Residues (FAO/WHO), 'Pesticide Residues in Food 1994 – Part 1: Residues', Rome, Vol 1 & 2, 1995 and other volumes in the series.
83. J.K. Dubey, T. Heberer, H.J. Stan, 'Determination of Ethylenethiourea in Food Commodities by a Two-step Derivatization Method and Gas Chromatography with Electron-capture and Nitrogen–Phosphorus Detector', *J. Chromatogr. A*, **765**, 31–38 (1997).
84. J. Gan, S. Papiernik, S.R. Yates, 'Static Headspace and Gas Chromatographic Analysis of Fumigant Residues in Soil and Water', *J. Agric. Food Chem.*, **46**, 986–990 (1998).
85. K.N.T. Norman, 'A Rapid Method for the Determination of Liquid Fumigants Residues in Food Commodities Using Automated Headspace Analysis', *Pestic. Sci.*, **33**, 23–34 (1991).
86. MAFF (UK), Committee for Analytical Methods for Residues for Pesticides and Vetinary Products in Foodstuffs, 'Determinations of Residues of Dithiocarbamate Pesticides in Foodstuffs by a Headspace Method', *Analyst*, **106**, 782–787 (1981).
87. R. Mestres, 'Pesticides Analytical Methods', Montpellier, 11–14, 1977.
88. S.B. Ekroth, B. Ohlin, B.G. Osterdahl, 'Rapid and Simple Method for Determination of Thiram in Fruits and Vegetables with High Performance Liquid Chromatography with Ultraviolet Detection', *J. Agric. Food Chem.*, **46**, 5302–5304 (1998).
89. S.G. Calzolari, 'Determinazione di Residui di Fosfina in Cereali Sottoposti a Fumigazione', *Boll. Chim. Unione Ital. Lab. Prov. Parte Sci. (*now *Boll. Chim. Igien.)*, **34**(S1), 29–41 (1983).
90. G. Mortensen, K. Sondergaard, B. Pedersen, B. Rietz, 'ICP-AES and Colorimetry as Analytical Tools for the Determination of Phosphorus-containing Compounds, Including Phosphine', *Anal. Lett.*, **22**(7), 1791–1806 (1989).
91. J.R. Rangasawami, 'Spectrophotometric Method for Determination of Phosphine Residues in Coffee Beans', *J. Food Sci. Technol.*, **27**(1), 33–37 (1990).
92. J.E. Farrow, A. Hoodless, M. Sargent, J.A. Sidwell, 'Determination of Residues of Post-harvest Fungicides on Citrus Fruit by HPLC', *Analyst*, **102**, 752–758 (1977).
93. R.J. Bushway, K. Larkin, B. Perkins, 'Determination of Thiabendazole in Potatoes by ELISA', *Food Agric. Immunol.*, **9**(4), 249–255 (1997).
94. B.B. Dzhantiev, A.V. Zherdev, O.G. Romanenko, J.N. Trubaceva, 'Development of Various Enzyme Immunotechniques for Pesticides Detection', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS, Washington, Vol. 657, 87–96, 1996.
95. C.D. Ergovich, 'Analysis of Pesticide Residues', ACS, Washington, 62–77, 1977.
96. S.J. Gee, B.D. Hammock, J.M. Van Emon, 'A User's Guide to Environmental Immunochemical Analysis', EPA/540/R/94/509, 1994.
97. E.P. Meulenberg, W.H. Mulder, P.G. Stocks, 'Immunoassays for Pesticides', *Environ. Sci. Technol.*, **29**, 553–561 (1995).
98. G.S. Nunes, I.A.S. Toscano, D. Barcelò, 'Analysis of Pesticides in Food and Environmental Samples by Enzyme-linked Immunosorbent Assays', *Trends. Anal. Chem.*, **17**, 79–87 (1998).
99. H.S. Rathore, T. Begum, 'Thin Layer Chromatographic Methods for Use in Pesticide Residue Analysis', *J. Chromatogr.*, **643**, 271–290 (1993).
100. H.S. Rathore, T. Begum, 'Thin-layer Chromatographic Behaviour of Carbamate Pesticides and Related Compounds', *J. Chromatogr.*, **643**, 321–329 (1993).
101. P.M. Scott, 'Chapter 49: Natural Toxins' in *Official Methods of Analysis of AOAC International*, 16th edition, ed. P. Cunnif, Arlington Virginia, Vol. II, 1995.
102. P.M. Scott, 'Recent Developments in Methods of Analysis for Mycotoxins in Foodstuffs', *Trends. Anal. Chem.*, **12**(9), 373–381 (1993).
103. M. Holcomb, D.M. Wilson, M.W. Trucksess, H.C. Thompson, Jr, 'Determination of Aflatoxins in Food Products by Chromatography', *J. Chromatogr.*, **624**, 341–352 (1992).
104. P.S. Diaz, M.A. Moreno, J. Prieta, L. Dominguez, 'Aflatoxins in Milk and Milk Products: A Review of Analytical Techniques, including Dialysis Methods', *Alimentaria (Madrid)*, **258**, 19–23 (1994).
105. J.G. Wilkes, J.B. Sutherland, 'Sample Preparation and High-resolution Separation of Mycotoxins Possessing Carboxyl Groups', *J. Chromatogr., B: Biomed. Appl.*, **717**(1–2), 135–156 (1998).
106. H. Valenta, 'Chromatographic Methods for the Determination of Ochratoxin A in Animal and Human Tissues and Fluids', *J. Chromatogr. A*, **815**(1), 75–92 (1998).
107. G.S. Shephard, 'Chromatographic Determination for Fumonisin Determination', *J. Chromatogr. A*, **815**(1), 31–39 (1998).
108. O.G. Larionov, O. Yu.Pisareva, 'Methods for Sample Preparation in Chromatographic Analysis of Some Mycotoxins', *Khim.-Farm. Zh.*, **31**(7), 49–53 (1997).

109. P.M. Scott, M.W. Trucksess, 'Application of Immunoaffinity Columns to Mycotoxin Analysis', *J. AOAC Int.*, **80**(5), 941–949 (1997).
110. S. Nawaz, R.D. Coker, S.J. Haswell, 'HPTLC – a Valuable Chromatographic Tool for the Analysis of Aflatoxins', *J. Planar Chromatogr.-Mod. TLC*, **8**(1), 4–9 (1995).
111. L.M. Ling, J. Zhang, P. Wang, Y.S. Wang, C.P. Chen, 'Thin Layer Chromatography of Mycotoxins and Comparison with Other Chromatographic Methods', *J. Chromatogr. A*, **815**(1), 3–20 (1998).
112. W. Langseth, T. Rundberget, 'Instrumental Methods for Determination of Non-macrocyclic Trichothecenes in Cereals, Foodstuffs and Cultures', *J. Chromatogr. A*, **815**(1), 103–121 (1998).
113. R. Krska, 'Performance of Modern Sample Preparation Techniques in the Analysis of Fusarium Mycotoxins in Cereals', *J. Chromatogr. A*, **815**, 49–57 (1998).
114. M. Soffrizzo, G. Avvantaggiato, A. Visconti, 'Use of Various Cleanup Procedures for the Analysis of Ochratoxin A in Cereals', *J. Chromatogr. A*, **815**, 67–73 (1998).
115. M. Nakajima, H. Tsubouchi, M. Miyabe, Y. Ueno, 'Survey of Aflatoxin B1 and Ochratoxin A in commercial Green Coffee Beans by HPLC Linked with Immunoaffinity Chromatography', *Food Agric. Immunol.*, **9**(2), 77–83 (1997).
116. E. Ioannou-Kakouri, M. Christodoulidou, E. Christou, E. Constandinidou, 'Immunoaffinity Column-HPLC. Determination of Aflatoxin M1 in Milk', *Food Agric. Immun.*, **7**(2), 131–137 (1995).
117. M. Gelosa, I. Buzzetti, 'Tecniche Cromatografiche per la Determinazione di Aflatossine nel Latte', *Ind. Aliment.*, **3**, 315–317 (1994).
118. P.M. Scott, S.R. Kanhere, 'Determination of Ochratoxin A in Beer', *Food Addit. Contam.*, **12**(4), 591–598 (1995).
119. P. Majerus, H. Ottender, 'Detection and Occurrence of Ochratoxin A in Wine and Grape Juice', *Dtsch. Lebensm. Rundsch.*, **92**(12), 338–390 (1996).
120. A.R. Brause, M.W. Trucksess, F.S. Thomas, S.W. Page, 'Determination of Patulin in Apple Juice by Liquid Chromatography: Collaborative Study', *J. AOAC Int.*, **79**(2), 451–455 (1996).
121. R.D. Josephs, R. Krska, R. Grasserbauer, J.A.C. Broekaert, 'Determination of Trichothecene Mycotoxin in Wheat by Use of SFE and HPLC with DAD or GC with ECD', *J. Chromatogr. A*, **795**(2), 297–304 (1998).
122. R.P. Huopalahti, J. Ebel, Jr, J.D. Henion, 'Supercritical Fluid Extraction of Mycotoxins from Feed with Analysis by LC-UV and LC-MS', *J. Liq. Chromatogr. Relat. Technol.*, **20**(4), 537–551 (1997).
123. W.K. Modey, D.A. Mullholland, M.W. Raynor, 'Analytical Supercritical-Fluid Extraction of Natural Products', *Phytochem. Anal.*, **7**(1), 1–15 (1996).
124. M.W. Trucksess, M.E. Stack, S. Nesheim, R.H. Albert, T.R. Romer, 'Multifunctional Column Coupled with Liquid Chromatography for Determination of Aflatoxins B1, B2, G1 and G2 in Corn, Almond, Brazil Nuts, Peanuts and Pistachio Nuts: Collaborative Study', *J. AOAC Int.*, **77**(6), 1512–1521 (1994).
125. R. Krska, 'Performance of Modern Sample Preparation Techniques in the Analysis of Fusarium Mycotoxins in Cereals', *J. Chromatogr. A*, **815**(1), 49–57 (1998).
126. R. Schuhmacher, R. Krska, M. Grasserbauer, W. Edinger, H. Lew, 'Immunoaffinity Columns Versus Conventional Cleanup: A Method-Comparison Study for Determination of Zearalenone in Corn', *Fresenius' J. Anal. Chem.*, **360**(2), 241–245 (1998).
127. K.A. Scudamore, S. Nawaz, M.T. Hetmanski, S.C. Rainbird, 'Mycotoxins in Ingredients of Animal Feeding Stuffs. III. Determination of Mycotoxins in Rice Bran', *Food Addit. Contam.*, **15**(2), 185–194 (1998).
128. E.J. Tarter, P.M. Scott, 'Determination of Patulin by Capillary Gas–Chromatography of the Heptafluorobutyrate Derivative', *J. Chromatogr. A*, **538**(2), 441–446 (1991).
129. P.M. Scott, D. Weber, S.R. Kanhere, 'Gas Chromatography–Mass Spectrometry of Alternaria Mycotoxins', *J. Chromatogr. A*, **765**, 255–263 (1997).
130. T. Goto, 'Determination of Aflatoxins by Capillary Column Gas Chromatography', *J. Chromatogr.*, **447**, 410–414 (1988).
131. A. Cepeda, C.M. Franco, C.A. Fente, B.I. Vasquez, J.L. Rodriguez, P. Prognon, G. Mahuzier, 'Postcolumn Excitation of Aflatoxins Using Cyclodextrins in Liquid Chromatography for Food Analysis', *J. Chromatogr. A*, **721**(1), 69–74 (1996).
132. 'Wiley 130K Mass Spectral Database' for HP Vectra and IBM AT Compatible DOS Systems, Licensed to Hewlett Packard (c) John Wiley & Sons, Inc., 1986.
133. M. Becker, P. Dengelmann, M. Herderich, P. Schreier, H.U. Humpf, 'Column Liquid Chromatography – Electrospray Ionization Tandem Mass Spectrometry for the Analysis of Ochratoxin', *J. Chromatogr. A*, **818**(2), 260–264 (1998).
134. M. Rychlik, P. Schieberle, 'Synthesis of ^{13}C-Labeled Patulin to be Used as Internal Standard in a Stable Isotope Dilution Assay', *J. Agric. Food Chem.*, **46**, 5163–5169 (1998).
135. H.U. Humpf, 'Fumonisins in Food: Sensitive and Selective Detection by HPLC-MS-MS', *GIT Labor-Fachz.*, **42**(2), 84–85 (1998).
136. A. Carniel, B. Casetta, S. Ceccone, 'Caratterizzazione e Dosaggio delle Tossine Della Famiglia delle Fumonisine per LC-MSMS', *Boll. Chim. Igien.*, **47**, 289–299 (1996).
137. Y. Onji, Y. Aoki, N. Tani, K. Umebayashi, Y. Kitada, Y. Dohi, 'Direct Analysis of Several Fusarium Mycotoxins in Cereals by Capillary Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **815**, 59–65 (1998).
138. C.J. Mirocha, E. Kolakowski, W. Xie, H. Yu, H. Jelen, 'Analysis of Deoxynivalenol and Its Derivatives (Batch and Single Kernel) Using Gas Chromatography/Mass

Spectrometry', *J. Agric. Food Chem.*, **46**, 1414–1418 (1998).
139. T. Tuomi, L. Saarinen, K. Reijula, 'Detection of Polar and Macrocyclic Trichothecene Mycotoxins from Indoor Environment', *Analyst*, **123**, 1835–1841 (1998).
140. E. Rosemberg, R. Krska, R. Wissiack, V. Kmetow, R. Joseph, E. Razzazi, M. Grasserbauer, 'HPLC – APCI/MS as a New Tool for the Determination of the Mycotoxin Zearalenone in Food and Feed', *J. Chromatogr. A*, **819**, 277–288 (1998).
141. S. Corneli, C.M. Maragos, 'Capillary Electrophoresis with Laser-induced Fluorescence: Method for the Mycotoxin Ochratoxin A', *J. Agric. Food Chem.*, **46**, 3162–3165 (1998).
142. P. Nowotny, W. Baltes, W. Kroenert, R. Weber, 'Thin Layer Chromatographic Method for the Determination of Twenty-two Mycotoxins in Mouldy Food', *Chem. Mikrobiol. Technol. Lebensm.*, **8**(1), 24–28 (1983).
143. G.S. Qian, P. Yasei, G.C. Yang, 'Rapid Extraction and Detection of Aflatoxin M1 in Cow's Milk by High-performance Liquid Chromatography and Radioimmunoassay', *Anal. Chem.*, **56**(12), 2079–2080 (1984).
144. CEN/CENELEC, The Joint European Standard Institution, 'EUROPEAN STANDARD EN45001 – General Criteria for the Operation of Testing Laboratories', Bruxelles, 1989.
145. A.R.C. Hill, '*Quality Control in Pesticide Residues Analysis*', Ministry of Agriculture Food and Fisheries, UK, 1994.
146. Istituto Superiore di Sanità – Gruppo di lavoro per i residui di antiparassitari della commissione permanente di coordinamento interregionale per i problemi relativi al controllo ufficiale dei prodotti alimentari, 'Linee Guida per l'applicazione della Buone Pratiche di Laboratorio e l'assicurazione ed il Controllo della Qualità nell'analisi dei Residui di Prodotti Fitosanitari', eds. A. Di-Muccio, T. Generali, D. Attard-Barbini, Rapporti ISTISAN 97/24 – Rome, 1997.
147. S.A. Thorpe, S.L. Reynolds, 'Accreditation in the UK for Pesticide Residue Analysis in Foods', *J. Chromatogr. A*, **737**, 3–7 (1996).
148. International Standard Organization, 'Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 1:', 'General Principles and Definitions', Document 5725–1, 1994, and Part 3: Intermediate Measures of the Precision of a Standard Measurement Method', Document 5725–3, 1994.
149. J.L. Bernal, M.J. Nozal, J.J. Jiménez, 'Influence of Solvent and Storage Conditions on the Stability of Acaricide Standard Stock Solutions', *J. Chromatogr. A*, **765**, 109–114 (1997).
150. F.J. Schenck, 'Screening of Nonfatty Fish for Organochlorine Pesticide Residues by Solid-phase Extraction Cleanup – Interlaboratory Study', *J. AOAC Int.*, **79**(5), 1215–1219 (1996).
151. W. Gilsbach, 'Interlaboratory Studies of the German Working Group Pesticides Residues of the GDCH for Validating Analysis of Dithiocarbamate and Thiuramdisulfide Fungicides. 1. Communication–Validation of the Method DFG S 15', *Dtsch. Lebensm.-Rund.*, **92**(11), 351–353 (1996).
152. K.A. Scudamore, S.J. McDonald, 'A Collaborative Study of an HPLC Method for Determination of Ochratoxin A in Wheat Using Immunoaffinity Column Cleanup', *Food Addit. Contam.*, **15**(4), 401–410 (1998).
153. K. Larsson, T. Moeller, 'Liquid Chromatographic Determination of Ochratoxin A in Barley, Wheat Bran and Rye by the AOAC/IUPAC/NMKL Method: NMKL Collaborative Study', *J. AOAC Int.*, **79**(5), 1102–1105 (1996).
154. G.A. Bennett, T.C. Nelsen, B.M. Miller, 'ELISA for Detection of Zearalenone in Corn, Wheat and Pig Feed: Collaborative Study', *J. AOAC Int.*, **77**(6), 1500–1511 (1994).
155. Regione Lombardia, 'Studio Cooperativo Interlaboratoriale – Residui di Antiparassitari in Acque Potabili–Biennio 1997–1998', International Center for Pesticide Safety, Busto Garolfo (Milan), 1999.
156. W. Horwitz, 'Evaluation of Analytical Methods used for Regulation of Food and Drugs', *Anal. Chem.*, **54**(1), 67A–76A (1982).
157. P. Key, A.L. Patey, S. Rowling, A. Wilbour, F. Worner, 'International Proficiency Testing of Analytical Laboratories for Foods and Feeds from 1990–1996: The Experience of the UK Food Analysis Performance Assessment Scheme', *J. AOAC Int.*, **80**(4), 895–899 (1997).
158. P. Weigert, J. Gilbert, A. Patey, P. Key, R. Wood, N. Barylko-Pikielna, 'Analytical Quality Assurance for the WHO GEMS/Food EURO Programme-results of 1993/94 Laboratory Proficiency Testing', *Food Addit Contam.*, **14**(4), 399–410 (1997).
159. N.A. Smart, 'Interlaboratory Comparability of Accuracy in Residue Analysis', *Anal. Proc.*, **30**(2), 75–77 (1993).
160. N.A. Smart, 'Examination of Performance in Collaborative Studies of Recommended Methods for Pesticide Residues Analysis', *Analyst*, **109**(6), 781–786 (1984).
161. European Community Commission, 'Recommendation 199/333/CE March 3-1999', Official Gazette of the European Community, L128, 25–55 (May 21, 1999).

Proteins, Peptides, and Amino Acids Analysis in Food

Stephan Dierckx, Karen Boeve, John Van Camp, and André Huyghebaert
University of Ghent, Ghent, Belgium

1 Introduction 1
2 Quantification of Total Nitrogen and Protein 1
2.1 Kjeldahl Method 3
2.2 Infrared Spectroscopy 4
2.3 Dye Binding 4
3 Determination of Specific Nitrogen Compounds 5
3.1 Gas Chromatography 5
3.2 High-performance Liquid Chromatography 6
3.3 Electrophoresis 10
3.4 Immunological Detection Methods for Food Proteins 12
4 Certified Reference Materials 14
5 Conclusions 14
Abbreviations and Acronyms 15
Related Articles 15
References 16

Proteins are an important fraction in most foodstuffs, ranging from 1.8% in roots and tubers up to about 25% in legumes.[1] However, in foods they are present in a complex multicomponent matrix also containing water, lipids, carbohydrates, and some minor components, all of which might interfere during the analysis of proteins in foods. A whole variety of analytical tools for protein analysis are available to the analyst[2] but not all of them are suitable for routine analysis procedures. This article will deal with some key analytical techniques in order to give an overview of possibilities and shortcomings of current techniques and some promising future trends. Well established methods are the Kjeldahl method, spectroscopy in the ultraviolet (UV), visible (VIS), and infrared (IR) ranges of the spectrum, chromatography, and conventional electrophoresis, while recent developments are to be found in the field of capillary electrophoresis (CE) and immunoassay. Since validation of methods and results are important features of today's analysis, some attention is paid to the availability and use of reference materials.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Proteins are defined as complex polymers of amino acids, which are incorporated in a fixed sequence given by the information encoded in the DNA. They can vary in structure depending on their sequence of amino acids and chain lengths, thereby making possible a diversity of structures and functions. Proteins constitute the primary material of living organisms[3–5] and actually every living organism can be considered as a source of protein.

The use of different protein sources can vary widely from one country to another depending on habits and traditions. Available protein sources can roughly be divided into two major groups: plant and animal proteins. A major portion of the world protein supply is derived from plant proteins such as cereals, legumes, and oilseed proteins.[6] From an economic and quantitative point of view, the most important animal protein sources are muscle tissue, eggs, and milk.[5] Cereal grains are of critical importance for humanity and are major foods in every country, either directly or indirectly as animal feeds. Analysis of cereal proteins is especially challenging to scientists owing to their complexity and heterogeneity. Proteins are important from a nutritional and technological point of view, they affect every property that characterizes a living organism, and they play different roles in the human body.[4,6] Owing to their importance and because each protein or protein source has its own functional and nutritional characteristics, both protein analysis and amino acid analysis are often a prerequisite for the characterization of a food product.

Various techniques are available to quantify the total N and protein content in food products. In section 2, three classical procedures are discussed: Kjeldahl, IR spectroscopy, and dye binding. The Kjeldahl method is an important procedure, at the least as the most frequently used reference method. Also, many official protein determination methods include the Kjeldahl technique. IR spectroscopy and dye binding are used as common routine methods in the quantification of N compounds and protein. For the quantification of specific N and protein compounds, three different methods are discussed: electrophoresis, chromatography, and immunology. These techniques are very specific and find their application in, e.g. authenticity research, determination of nutritional values, and quality control. Several N and protein determination methods have been described in review articles.[2,7–11]

2 QUANTIFICATION OF TOTAL NITROGEN AND PROTEIN

Protein determination methods are classified in two groups: some methods determine the protein content

directly, while others determine total organic N which is taken to be equivalent to crude protein. Of the latter procedures, the most important technique is the determination of N by the Kjeldahl method. Direct protein determination methods include procedures such as IR spectroscopy, turbidimetry, fluorimetry, refractometry and polarography.[2]

In indirect protein determination methods, the factors used to convert total organic N into protein are based on the average N content of proteins in particular foods. An important condition for the conversion of N into protein is that the food product may not include other N-containing matter besides protein. A number of conversion factors for different cereals and other products have been adopted by the Food and Agriculture Organization and the World Health Organization.[12] Some other scientists criticized those factors and have published factors for a wide range of foods based on detailed amino acid analysis.[13] Another list of conversion factors was given by Gaspar.[7] If there is no specified conversion factor, the N content is multiplied by 6.25, a factor corresponding to an average protein N content of 16%. When quoting data on protein content derived from indirect protein quantification methods, it is required to report also the conversion factor employed.

In protein analysis, a frequent operation is the determination of the nonprotein nitrogen (NPN) content, or its separation from the ''true protein'' nitrogen. Not taking into account the NPN fraction can lead to an overestimation of the true protein content, e.g. in the case of milk by about 5–6%. The NPN generally includes peptides with molecular weight (MW) <10 kDa, free amino acids, ammonia, urea, and trimethylamine oxide. The methods most often employed to separate NPN are dialysis, ultrafiltration and precipitation.[7] The precipitation methods most frequently employed use heat, alcohol, HCl–acetone, trichloroacetic acid (TCA), perchloric acid (PCA) and sulfosalicylic acid. Since various concentrations of certain reagents precipitate different percentages of protein, specification of the type and concentration of the precipitating agents used is required.[14] The nature of the NPN fraction in cows' milk and methods for determining the NPN constituents were described by Wolfschoon-Pombo et al.[15,16]

Possible interfering compounds in protein determination methods are carbohydrates, salts, lipids, and detergents. Also, the amino acid composition of a particular protein can influence the response of proteins, e.g. color development in a dye-binding assay. Therefore, the calibration of the assay using an appropriate protein mixture is recommended in order to obtain accurate measurements. In the following section, specific attention will be given to three methods: Kjeldahl, IR spectroscopy and dye binding. Advantages and disadvantages of these three methods are presented in Table 1.

Table 1 Advantages and disadvantages of the Kjeldahl method, IR spectroscopy and dye binding. [Adapted from Van Camp and Huyghebaert[2]]

Method	Advantages	Disadvantages
Kjeldahl	Precise and versatile Quantitate nitrogen from either soluble or insoluble samples Included in methods approved by international organizations	Time consuming Use of expensive, toxic materials Interference by the NPN fraction Not the most sensitive technique Indirect method: choice of conversion factor
IR spectroscopy	Reproducible Rapid analysis time Excellent precision Nondestructive analytical method Minimal sample preparation No need for skilled technical labor (minimal technical effort) No need for expensive and corrosive chemicals Applicable to solid materials Multicomponent analysis	Requires calibration against a reference method. The calibration is difficult, but, once determined, universal constants allow stable measurements to be achieved over long periods without the need of regular checking Interference by starch and lipids Errors from factors such as particle size differences, moisture losses during grinding, and sample homogeneity
Dye binding	Very sensitive Rapid Included in methods approved by international organizations	Dependent on amino acid composition Interferences from laboratory chemicals Adsorption of the dye on glassware and cuvettes Insolubility of proteins in acidic system Destructive Requires calibration Variations of binding capacity for different batches of commercial-grade dyes

2.1 Kjeldahl Method

Although it has been subject to modification over the years, the basic Kjeldahl procedure is still the most universal technique available for the determination of organic N, and it is therefore internationally recognized for official and statutory purposes. A great advantage of the Kjeldahl procedure over other protein determination methods is the possibility of determining insoluble proteins. The Kjeldahl method is applied as a reference method to calibrate direct methods of protein determination.

The Kjeldahl procedure involves two steps: digestion of the sample by heating with concentrated sulfuric acid and the determination of the ammonia released. Concentrated sulfuric acid is an effective dehydrating agent. At a given critical temperature, the elements of water from oxygen-containing organic matter are extracted and carbon is liberated. At the decomposition temperature sulfuric acid also acts as an oxidant and oxidizes the liberated carbon to carbon dioxide. This causes conversion of organic N into ammonia which is retained in solution as ammonium sulfate. In the first stage, the transformation of N into ammonia is rapid, but it subsequently becomes increasingly slower.[7] To guarantee the complete decomposition of more resistant intermediates, an additional period of boiling (after-boil) after the digest is recommended.

The decomposition step can be accelerated and completed by adding salts, oxidants, or reductants and using catalysts. When using these accelerating agents, a process time of 10–45 min for the total digestion step has been reported.[7] The addition of salts increases the boiling-point and shortens the decomposition time for organic compounds. The salt generally used is K_2SO_4. The addition of oxidants significantly accelerates digestion and decreases foaming.[12] The oxidants mainly applied are hydrogen peroxide, PCA, persulfates, and chromic acid. To determine the total N content in plant materials containing nitrate, the addition of reductants is necessary for the conversion of the nitrate N into ammonium ions. Reduction is also necessary in the determination of the N of azo, nitro, oxime, isoxazole, hydrazine, and hydrazone groups. Metallic zinc or copper, titanous chloride, sodium dithionite, benzoic acid, or phenol can be used. Salicylic acid and metallic zinc are recommended for enhancement of the reduction process.[7] The digestion time also depends on the catalyst used. Mercury, as mercuric oxide, is generally agreed to be the most effective, with selenium almost as effective, but both have toxic hazards and waste disposal problems. Another disadvantage of mercury is the formation of ammonia complexes in the digest, which requires sodium thiosulfate to break the complex and to release the ammonia into solution.[12] Copper is the most widely used catalyst nowadays. Additional information on the use of salts or oxidants in the digestion process was given by Gaspar[7] and Lakin[17] reported some guidelines for the selection of a suitable catalyst.

The ammonia formed in the digest can be determined by titrimetry, spectroscopy, or potentiometry. In the titrimetric method, the digest is distilled or steam distilled in the presence of alkali to release the ammonia. By direct distillation with moderate heating, the ammonia can be quantitatively distilled off in <20 min, depending on the size of the apparatus. A steam-distillation apparatus needs a maximum of 10 min for the quantitative distillation of the ammonia.[7] In a frequently used procedure, the liberated ammonia is adsorbed in an excess of boric acid solution and can then be titrated with standard acid (HCl, H_2SO_4) in the presence of an indicator. Spectroscopic methods are mainly used for the direct determination of the ammonia in the decomposition solution, making unnecessary the time-consuming distillation and titration steps. In addition, they are more advantageous than titration for the determination of microquantities ($0.1–1.5\,\mu g\,mL^{-1}$).[7] Color reactions of ammonia with Nessler reagent (HgI_4^{2-}), indophenol, and ninhydrin can be used for spectrometric determination. However, the catalysts used in the Kjeldahl method can disturb the color reactions of ammonia.

In the potentiometric determination, the distillation–titration steps are replaced by the application of an ammonia-specific electrode (ammonia gas-sensing probe), which shortens the digest analysis time to 5–10 min. Ammonia gas-sensing probes measure the partial pressure of gases in solution.[18] Copper and mercury in digestion catalysts may interfere by complexing ammonia. This may be prevented by the addition of the disodium salt of ethylenediaminetetraacetic acid (EDTA) or sodium iodide, respectively.[18,19]

On a macroscale, modern types of Kjeldahl apparatus allow the analysis of N in the range 0.2–160 mg per sample.[9] The application of the Kjeldahl method can be manual, or semi- or fully automated. Fully unattended operation allows the analysis of 60 samples in up to 4 h. The Berthelot procedure is particularly applicable to micro-Kjeldahl analysis. This procedure involves a wet digestion similar to that in the Kjeldahl method, after which the ammonia formed is determined by colorimetry. Nitrogen concentrations in the range $0.3–1.3\,\mu g\,N\,mL^{-1}$ for the sample digest could be determined.[20] Automation of the Berthelot reaction allowed the determination of ammonia at the rate of 40 samples per hour.[9] Different procedures based on the Kjeldahl method have been studied and optimized for meat,[21,22] milk,[23–25] and cereals.[20]

2.2 Infrared Spectroscopy

Most organic products have an absorption in the IR region. Absorption of radiation at IR frequencies by molecules is the result of transitions between vibrational energy levels. Atoms vibrate in complicated patterns affected by atomic masses and strength of atomic bonds. Functional groups, such as O–H, N–H, and C–H bonds, will absorb various IR frequencies. The pattern of the various IR frequencies absorbed provides information about the sample composition. The degree of IR absorbance by functional groups at specific wavelengths can be correlated with the concentration of constituent compounds of the sample. The near-infrared (NIR) spectral region is the part of the electromagnetic spectrum between 800 and 2500 nm. It is the intermediate region between the VIS part of the electromagnetic spectrum (400–750 nm) and the IR spectrum (2500–30 000 nm). The higher energy content of NIR compared with IR radiation is suitable to induce transitions to higher vibrational levels and to cover absorption bands related to combined interactive vibrations. NIR has important advantages over IR spectroscopy for quantitative analysis. Since the absorptivities are much lower for overtone and combination bands compared with the fundamental bands, moderately concentrated samples and pathlengths comparable to those usually applied in the VIS or UV region may be used, thus permitting the analysis of intact samples such as protein fibers.[9]

In the quality control of milk, IR spectroscopy is routinely used to determine the protein content by measuring the absorption band at a wavelength of 1548 cm^{-1}, which corresponds to the absorption maximum of the peptide bond. The quantification of lactose (1042 cm^{-1}) and fat (1745 cm^{-1}) can be applied simultaneously. The two main technological problems are the high water content of milk and the heterogenity of the fat globules.[8] Water absorbs a large quantity of energy between 1000 and 5000 cm^{-1} which disturbs the protein absorption bands in the IR spectrum. To neutralize this problem, a double-beam spectrophotometer with water in the reference cell can be used. Fat globules in milk display a similar diameter to the wavelength of the light beam, which causes light diffusion and diminuation of the absorption peak. To reduce these phenomena, fat globules have to be homogenized (to $<2\,\mu m$) and absorption bands have to be selected in the region of longer wavelengths.

In NIR spectroscopy, a spectrum may be collected from a sample in two ways: by transmission through the sample and by reflection from the sample. Transmission techniques are suitable for a wider range of samples, from clear films and liquids through highly scattering solutions, pastes, slurries, and suspensions. The reflectance mode is best suited to opaque powders and to solids moving in a dense phase. NIR technology offers the facility for rapid (the analysis time is <1 min) and accurate protein testing and the technique is used on a routine basis. In addition to the fact that NIR reflectance makes it possible to determine insoluble proteins and analyze almost intact food samples, another advantage over other methods of protein determination is the simultaneous detection of other constituents such as moisture and lipids.[9,26]

NIR reflectance analysis had a considerable impact on the cereal industry. To avoid interferences due to starch, oil, and water, the optimum wavelength for the quantification of the protein constituents is 2180 nm, which is the position of a band assigned to a combination of twice amide I plus amide III.[9] Variations in particle size between different samples and variations in the levels of starch, oil, and water cause larger changes in NIR reflectance at 2180 nm than changes in protein content. A correction for these effects can be made by using an adapted calibration procedure where specific absorption bands for protein, starch, oil, and water are taken into account. For all chemical methods of protein determination, the sample of grain must be ground prior to analysis to render the grain to a meal having a similar particle size distribution to that of flour.[9] The standardization of the grinding method for NIR reflectance analysis is more important than that for chemical methods because the reflectance characteristics are strongly influenced by particle size. For ground wheat, some errors in predicting protein in flour with NIR reflectance result from such factors as particle size differences, moisture losses during grinding, and sample homogeneity. A study[27] claimed that under normal practical conditions, the NIR analysis of whole wheat samples may prove to be the more accurate procedure because these errors with ground wheat are known to be even larger under practical field conditions than under careful research conditions.

On-line NIR systems to evaluate the quality of meat are already on the market to assay water, proteins, lipids, and collagen. A difficulty is found with products such as ham and other fine meat products; while the absorption bands in the NIR method are mainly influenced by the protein fraction, the Kjeldahl procedure, as a reference method, determines the total N content and thus also the nitrates.[28,29] A further improvement of the NIR technique is obtained by using the whole spectrum or large parts of it, rather than one or a few wavelengths. Quantification is then achieved by chemometrics.[30]

2.3 Dye Binding

Under appropriate conditions (pH and ionic strength of the medium), the acidic and basic groups of proteins

interact with dissociated groups of organic dyes (e.g. the acid radicals of sulfonic acids) to form colored precipitates. Examples of colorants used are Amido Black 10-B, Acid Orange 12 and Coomassie Brilliant Blue (CBB).

In indirect protein determination methods utilizing dye binding, the excess of dye was found to be indirectly proportional to the amount of protein present. The principle of those methods was to remove the unsoluble protein–dye complex by filtration or centrifugation and to measure by spectroscopy the excess of dye in solution. An indirect dye-binding method was employed[7] to determine the basic and acidic groups, and their ratios, in the proteins of wheat flour fractions. Dye binding became popular with the development by Bradford[31] of a direct protein determination method using CBB G_{250}. The binding of the dye to protein causes a shift in the absorption maximum of the dye from 465 nm (red form) to 595 nm (blue form), which is quantified by monitoring the increase in absorption at 595 nm. The Bradford procedure is very sensitive, working in the range 0.2–1.4 mg protein mL^{-1} for the standard assay and 5–100 μg protein mL^{-1} for the microassay procedure.[32] The method is a rapid, simple, one-step procedure and shows color stability for 1 h. There is little or no interference from cations such as sodium or potassium or from carbohydrates such as sucrose. However, relatively large amounts of detergents such as sodium dodecyl sulfate (SDS), Triton X-100™, and commercial glassware detergents are found to give excessive interfering color in the assay.[31] A small amount of color is developed in the presence of strongly alkaline buffering agents. Interferences by nonprotein compounds may be eliminated by the use of proper controls. The main problems in the Bradford procedure are technical: the dye adsorbs on glassware and cuvettes, which can be solved by the use of disposable cuvettes. Also, the adsorbed dye can be quickly removed by rinsing in 0.1 M HCl or by washing with concentrated detergent followed by water and acetone.[32,33]

In the dye-binding methods, the response of proteins may vary with the amino acid composition of the particular protein, e.g. primarily basic (especially arginine) and aromatic amino acids residues bind to CBB G_{250} in the Bradford method.[32] Whey proteins adsorb 1.27 times more colorants as caseins, whereas the NPN fraction does not interact with the dyes.[33] Also, nonprotein groups, such as starch and bran in cereals, can adsorb significant amounts of dye. Hence accurate determinations require standardization with an appropriate protein mixture. To allow for an accurate protein quantification, each type of sample should have its own calibration which must be rechecked frequently owing to variations in the binding capacities of different batches of commercial grade dyes.[9] Since protein dye binding is a colorimetric procedure, there are interferences by colored food substances.[8,34] Colors having absorption bands nearest to those of the dyes interfere the most. Increasing color concentrations cause greater depressions of dye binding, and hence protein percentage. Also, Maillard browning reactions, e.g. in milk powder, involve the lysine residues, rendering them partially or completely unavailable to react with the dye.[35] Instruments for manual and semi- and automatic determination of protein by a dye-binding method are on the market for cereals[36] and dairy products.[37,38]

3 DETERMINATION OF SPECIFIC NITROGEN COMPOUNDS

Various instrumental techniques are available to determine specific N compounds in food such as amino acids, peptides, and proteins. Separation of the individual N compounds is based on specific properties, mainly the charge-to-mass ratio and structural properties. These specific characteristics are exploited in techniques such as gas chromatography (GC), high-performance liquid chromatography (HPLC), electrophoresis, CE, and immunological determinations. In the following sections each of these techniques will be briefly discussed with emphasis on their applicability for the separation and/or determination of food proteins, peptides, and amino acids. Examples of recent applications and some general guidelines are given. The separation and analyses of individual N compounds is for quantification purposes in addition to identification of particular proteins or protein fractions. For instance, detection of the adulteration of high-quality products by addition of components with lower quality or which are cheaper than the original product is becoming more and more important.[39]

3.1 Gas Chromatography

GC is based on differences in the affinities of the compounds to be separated for the mobile phase, usually a carrier gas such as He or N_2, and the stationary phase. A prerequisite for the application of GC for separations of compounds in general is that these compounds must be volatile or can be converted into volatile derivatives. Taking into account the relatively high MW of peptides and proteins and their characteristic structure and conformation, GC is only useful for the separation of amino acids. A serious drawback of GC separations of amino acids, compared with other analytical methods, is the complicated and not automatable derivatization.[40] Analysis of amino acids by GC after acid hydrolysis of the proteins has been applied to canned foods. Amino

acids were determined as heptafluorobutyryl isobutyl ester derivatives after separation on a packed column with methylsilyl or an open column with dimethylsilyl as mobile phase. Detection was carried out with a thermionic nitrogen–phosphorus detector.[41] A rapid screening method for free amino acids consists of picric acid treatment and solvent washing of a water extract of the sample for the isolation of free amino acids. The amino acids are derivatized into zwitterionic *N*(*O*,*S*)-isobutyloxycarbonylcarboxylic acids and subsequently silylated to stable *tert*-butyldimethylsilyl derivatives. Detection was carried out with mass spectrometry.[41] The GC analysis of amino acids is a very sensitive method. However, the intrinsic disadvantages have led to GC being dominated by HPLC as the prefered method for amino acid analysis.

3.2 High-performance Liquid Chromatography

The principle of HPLC is very similar to that of GC but in the case of HPLC the mobile phase is a solvent or a mixture of solvents. Based on the separation mechanism and stationary phase, different HPLC modes can be distinguished. The different HPLC modes and some properties are listed in Table 2. A review of HPLC techniques for the separation of food proteins and applications was given by Dierckx and Huyghebaert.[42] HPLC is an excellent technique for the separation of proteins and also peptides and amino acids. However, depending on the sample type, HPLC mode, and analytes to be detected, thorough sample preparation and cleanup procedure may be required. This is especially the case for amino acid analysis, which is one of the most difficult and complicated analyses to be carried out in food analysis procedures. Sample preparation can include removal of interfering components such as lipids and nucleic acids and solubilization of the analytes. Separation of the different classes of proteins in a sample can be achieved by selective precipitation. This includes the use of salts, organic solvents, or heat. Some examples of sample pretreatments are given in Table 4.

Separation conditions and equipment requirements depend very much on the choice of the HPLC mode. From this point of view, SEHPLC can be considered as the simplest method to perform since no gradient elution, in which the composition of the mobile phase changes during elution, is used. A typical mobile phase for SEHPLC consists of a buffer at near neutral pH, e.g. phosphate or Tris buffer, with a low concentration of salt in order to supress ionic interactions of proteins with the stationary phase. The other modes of HPLC are more complicated to perform since elution of analytes is based on the competition of analyte molecules and the (changing) mobile phase for the stationary phase. Typical mobile phases for RPHPLC of proteins, peptides, and amino acids are aqueous mixtures of acetonitrile or propanol in the presence of small amounts (~0.1%) of, e.g. TFA or heptafluorobutyric acid. Increasing the hydrophobicity of the mobile phase by increasing the ratio of organic solvent to water in the mobile phase causes elution of the proteins, based on their hydrophobicity. In the case of HIHPLC, elution is obtained by changes in the ionic strength of the mobile phase. This principle is also used in IEHPLC, in addition to changes in the pH of the elution buffer. Since good control of the gradient conditions is necessary to obtain good reproducibility, perfect control of the composition of the mobile phase is required. This is achieved by using two solutions which are mixed in a certain, changing ratio throughout the separation and which requires high-quality equipment.[42]

The method of preference for the determination of amino acids is HPLC. However, the analytical procedure is very much dependent on the purpose of the analysis: determination of free amino acids or the total amino acid profile of a food product, and quantification of amino acids or identification of the presence of any particular amino acid. Determination of free amino acids is far simpler than the determination of the total amino acid profile since the latter requires hydrolysis of the protein molecules into their constituent monomers, preferably without destruction or transformation of any particular amino acid and with the total hydrolysis and release of the hydrolyzed amino acids. Free amino acid analysis usually includes two steps: rendering the material extractable in the case of solid foods, e.g. by homogenization or grinding, and solubilization of the free amino acids. This is achieved by stirring with a suitable solvent, e.g. dilute HCl. Precautions have to be taken because some amino acids might be tightly bound to proteins or may be labile to acidic conditions. Total hydrolysis of proteins for determination of the total amino acid profile is generally performed by acid hydrolysis of the peptide bond (6 M HCl, 105–120 °C for 18–24 h under N_2). However, this method has some drawbacks, especially when critical parameters such as purity of acid, acid-to-protein ratio, temperature, time, and O_2 are not under control. Possible problems are the degradation of Trp, oxidation of Cys, dehydration of Ser and Thr, and conversion of Gln and Asn into their acid forms. This has led to the development of other methods such as the use of other acids (methanesulfonic acid, performic acid),[43] alkaline (4.2 M NaOH, 110 °C for 22 h)[44] and enzymatic hydrolysis,[45] or the incorporation of protecting agents [thioglycolic acid, tryptamine, or β-mercaptoethanol (ME)].[46] A further complication is the fact that most amino acids do not have useful UV or

Table 2 Modes of HPLC used for the analysis of proteins, peptides, and amino acids in foods. [Adapted from Van Camp and Huyghebaert[2] and Dierckx and Huyghebaert[42]]

HPLC method	Separation mechanism	Typical stationary phase	Typical mobile phase	Advantages	Disadvantages
SEHPLC	Differences in size of protein molecules	Silica gel Synthetic polymers	Phosphate–NaCl buffers, pH 7.0 Tris–HCl–SDS buffers, pH 7.0	Simple one-step analysis Mild separation conditions, no protein denaturation	Peak broadening at large sample volumes Decrease in resolution at high flow rates Interactions between analytes and stationary phase interfere Elution behavior dependent on molecular conformation
RPHPLC	Differences in polarity of the amino acid side chains	Linear hydrocarbons of 4–18 residues in length Phenyl group	Acetonitrile–TFA mixture	High resolution power Analysis possible at low ionic strengths	Risk of protein denaturation Use of toxic solvents Interference by hydrophobic contaminants High system requirements
HIHPLC	Differences in the polarity of amino acid side chains	Phenyl group	Salt solutions of high ionic strength	High resolution power Large number of potential elution conditions Less risk of protein denaturation Analysis at high ionic strengths possible	Corrosive effects of salts Interference by hydrophobic contaminants
IEHPLC	Differences in the pK_a values of the functional groups of the side chains	Anion exchange: diethylaminoethyl, amines Cation exchange: sulfopropyl, sulfonic acid	Salt buffers (varying ionic strength or pH)	Mild separation process with limited protein denaturation High resolution power Wide range of variables, allowing fine tuning of separations	Laborious optimization procedure Corrosive effects of salts Interactions of analytes with stationary phase might interfere

HIHPLC, hydrophobic interaction high-performance liquid chromatography; IEHPLC, ion-exchange high-performance liquid chromatography; RPHPLC, reversed-phase high-performance liquid chromatography; SEHPLC, size-exclusion high-performance liquid chromatography; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane.

Table 3 Precolumn derivatization reagents used for the analysis of amino acids by HPLC

Reagent	Detection	Sensitivity	Advantages	Disadvantages
o-Phthalaldehyde	Fluorescence: $\lambda_{ex} = 330–340$ nm, $\lambda_{em} = 430–450$ nm	100 fmol	Strong fluorescence Rapid reaction Aqueous medium	No reaction with secondary amino acids Poor sensitivity for Cys and Lys Unstable derivatives
Phenyl isothiocyanate	UV: $\lambda = 254$ nm	10 pmol	Reactions with secondary amino acids Stable derivatives	Poor reproducibility for Cys Rapid deterioration of column
9-Fluorenylmethyl chloroformate	UV/fluorescence: $\lambda_{ex} = 263$ nm, $\lambda_{em} = 313$ nm	100 fmol	Reaction with secondary amino acids Stable derivatives Good reproducibility	Need to eliminate excess of derivatizing agent
Dansyl chloride	UV: $\lambda = 250–254$ nm Fluorescence: $\lambda_{ex} = 360$ nm, $\lambda_{em} = 470$ nm	200 fmol	Determination of Cys Stable derivatives	Formation of two derivatives possible Poor reproducibility for His
Dabsyl chloride	UV/VIS: $\lambda = 436$ nm	1 pmol	Stable derivatives	Interference by salt, urea, SDS, phosphate and NH_4HCO_3

UV/VIS, ultraviolet/visible.

fluorescence absorbances,[45] requiring derivatization of amino acids to permit detection by conventional means (UV or fluorescence detection). Derivative formation can be done pre- or postcolumn, primarily depending on the HPLC mode used for separation. Precolumn derivatization is commonly used with RPHPLC since it renders the amino acids more hydrophobic. A major advantage of this procedure is an unlimited reaction time, allowing the tailoring of the analysis to a particular matrix. An overview of possible precolumn derivatization techniques with some properties is given in Table 3. Postcolumn derivatizations for the analysis of amino acids are almost exclusively used in combination with IEHPLC (cation exchange) with ninhydrin (λ 570 nm) or *o*-phthaldialdehyde in the presence of a thiol as derivatizing agents. Various problems are connected with this method: the postseparation volume is increased and a postcolumn reaction vessel is required, which results in loss of resolution. Underivatized amino acid separation limits flexibility and the derivatization reaction needs to be very fast.[45] Both, IEHPLC and RPHPLC are currently in use for amino acid analysis[47,48] and a comparison between the two methods was made by Bütikofer et al.[49,50] Various automated HPLC systems for the analysis of amino acids are available, e.g. the Pico-Tag system from Waters (Millipore, Bedford, MA, USA).[51]

Peptides can be analyzed by RPHPLC in much the same way as proteins.[52–54] However, some problems may occur, e.g. with UV detection, where other food components may interfere at the low wavelengths used for peptides, during SEHPLC, where peptides all elute in the void fraction, or separation of small hydrophilic peptides (di- and tripeptides) which are normally not resolved during a classical RPHPLC separation.[55] Selective and sensitive methods have been developed for the detection of peptides, usually based on the same principles as those used for the detection of amino acids including both precolumn[55] and postcolumn[56] derivatizations. A good overview of the possibilities of RPHPLC for peptide separations and evaluation of various detection techniques was given by Herraiz.[57]

Analysis of proteins by HPLC has become a widely applied method in many laboratories since the introduction of columns with suitable packings for protein analysis by Waters (Millipore) in the late 1970s. Nevertheless, routine applications are mainly limited to milk and cereal proteins, as can be derived from the number of publications in this field.[42] Analysis of milk proteins includes quantification of caseins and whey proteins,[58,59] identification of milk from different species,[60–65] phenotyping of milk proteins,[66,67] detection of addition of bovine milk to soybean milk,[68,69] influence of technological parameters on milk proteins,[70,71] and the study of cheese ripening.[54,72] Applications to wheat proteins are found in the differentiation between various cultivars,[73–75] adulteration of durum wheat with common wheat,[76,77] and relating protein composition with dough-making properties and bread quality.[78–80] Furthermore, HPLC has proven to be useful for the analysis of egg,[81,82] and meat[83,84] proteins, fish species identification,[85,86] soybean hydrolysates,[87] peanut proteins,[88] the analysis of mixed protein systems,[89,90] and species identification in meat samples.[91] In order to give an overview of the possibilities of HPLC for the analysis of proteins in foods, a summary of some procedures is outlined in Table 4.

Table 4 Some examples of the use of HPLC, electrophoresis, and immunological techniques for the analysis of food proteins

Matrix	Purpose	Method	Detection	Sample preparation	Ref.
Milk and milk products	Phenotyping of bovine milk proteins	RPHPLC (C_{18}), 30 °C, Acetonitrile–TFA	UV, 220 nm	Dissolution of freeze-dried milk samples in 0.1 M Bis–Tris, 4 M urea, 0.3% ME buffer at pH 7.0 Dilution of samples with running buffer containing 6 M urea	67
	Rapid detection of genetic variants of milk proteins	IEF/PAGE	CBB	Separation of caseins and whey proteins by acid precipitation at pH 4.6 and centrifugation (3000 g for 10 min.)	174
	Analysis of genetic variants of milk proteins from different species	FSCE, 45 °C, 25 kV HCFS[a]–citrate, urea, MHEC[b] buffer, pH 3.0	UV, 214 nm	Incubation of milk samples with 167 mM Tris, 42 mM 3-(*N*-morpholino)-2-hydroxypropane, 67 mM EDTA, 17 mM DL-DTT, 6 M urea, 0.05% MHEC[b], pH 8.6, without further treatment	122
	Detection of cows' milk and caseinate in goats' and ewes' cheese	Indirect competitive ELISA	Alkaline phosphatase	Bovine casein is bound to microtiter plates Primary antibodies were raised against γ_3-casein	162
	Evaluation of residual antigenicity of hydrolyzed whey proteins	Indirect competitive ELISA	Peroxidase	α-Lactalbumin, β-lactoglobulin, BSA, or Ig were coated on to wells Primary antiserum derived from persons allergic to cows' milk	170
Meat and meat products	Species identification in cooked meat samples	SDS/PAGE	CBB and silver staining	Extraction with 1% (w/v) SDS, 10 mM EDTA, 1 mM DTT Centrifugation (80 000 g for 30 min)	128
	Species identification in unheated samples	RPHPLC, Acetonitrile–TFA	UV, 280 nm	Mix chopped samples with distilled water Filtration over paper filter and addition of NaN_3	91
Wheat	Classification of wheat varieties	RPHPLC (C_4), 60 °C, Acetonitrile–TFA	UV, 210 nm	Extraction of defatted flour with 0.05 M NaCl followed by centrifugation (15 000 g) Extraction of residue with 70% ethanol and centrifugation (15 000 g for 12 min)	74
	Classification of wheat varieties	IEF	CBB	Extraction of crushed and sieved grains with ethylene glycol Centrifugation (2000 g for 10 min) and used without further treatment	117
	Wheat varietal identification	FSCE, 40 °C, 22 kV UFS[c]–phosphate, HPMC[d] buffer, pH 2.5	UV, 200 nm	Extraction of gliadins with aqueous ethanol Filtration of extracts	175

BSA, bovine serum albumin; DTT, dithiothreitol; ELISA, enzyme linked immunosorbent assay; Ig; immunoglobulin; FSCE, free-solution capillary electrophoresis; IEF, isoelectric focusing; SDS/PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis.

[a] Hydrophilically coated fused silica.

[b] Methylhydroxyethylcellulose.

[c] Uncoated fused silica.

[d] Hydroxypropylmethylcellulose.

3.3 Electrophoresis

Electrophoresis refers to the movement of charged particles in an external electrical field. The mobility of a charged molecule is dependent on its charge-to-size ratio, the size being determined by MW, the three-dimensional structure, and the degree of solvation.[10] Nowadays, conventional electrophoresis is almost exclusively performed in slab gel systems. The gel medium stabilizes the buffer medium against convectional flow and improves the separations by its sieving action.[92] The sample mixture is brought on to the gel and a voltage difference is applied over the gel. Proteins and peptides, having a specific charge at a certain pH value, will migrate towards the electrode of the opposite sign. The net charge, however, is conformation dependent, so protein electrophoresis is strongly influenced by pH, temperature, and ionic strength. In addition to charge, the electrophoretic mobility is also influenced by macromolecular size and shape of the proteins and peptides, and thus the pore-size distribution of the gel matrix.

A frequently used gel medium is polyacrylamide in a method known as polyacrylamide gel electrophoresis (PAGE). Several samples can be analyzed at the same time and therefore under the same conditions of electrophoresis. In most commercially available systems, the gel is mounted on a cooling plate, so the heat produced during the electrophoresis is removed, resulting in a superior resolving power and greater flexibility.[93] Gels can be either homogeneous (fixed pore size) or gradient (increasing pore size) type. Different electrophoretic modes can be distinguished:

- SDS/PAGE: SDS is an anionic detergent. When proteins are dissolved in SDS solution they lose their individual charges and all proteins acquire the same negative surface charge density, giving a separation based on differences in molecular size. The MW of proteins can be determined by measuring their electrophoretic mobility in SDS-containing polyacrylamide gels.[94]
- Urea or native PAGE: proteins are separated based on their charge-to-mass ratio.
- IEF: a stable pH gradient is created between the anode and the cathode by the use of ampholytes. The proteins migrate towards the electrode of opposite charge and focus at those positions on the pH gradient where they have no net charge, also called the isoelectric point (IEP).

Classical slab electrophoresis is still performed mainly manually, although semiautomated separation, staining, and scanning units exist, and requires highly skilled personnel in the gel pouring, sampling, separation, and staining/destaining procedures. A wide variety of premade gels, with various degrees of cross-linking and with fixed gradients, are commercially available. Using these gels has the advantage of a higher reproducibility and a shorter analysis time. Furthermore, it reduces the risk of exposure of researchers to the neurotoxic acrylamide. A more comprehensive description of gel electrophoresis was given by Creighton.[94]

After separation, proteins and peptides have to be fixed on the gel and have to be visualized. Fixation agents such as 20% (w/v) TCA or methanolic solutions of acetic acid can be used. After the separation process, proteins and peptides are precipitated and immobilized by these agents. Peptides, however, need to have a certain size, i.e. ~10 kDa, in order to be separated and fixed on the gel. Interfering nonprotein components are removed prior to staining. Since proteins and peptides lack sufficiently high UV absorption or fluorescence, the use of specific dyes is necessary. Protein staining is usually achieved by the use of CBB G_{250} or R_{250}, which benefits from its speed of operation and relatively low background staining but which is rather insensitive (0.5 μg protein cm^{-2} gel).[2] Furthermore, CBB does not obey Beer's law at high concentrations, which can interfere during the spectrophotometric scanning of the gels and can give erroneous results in quantification procedures based on the density profiles.[95] Silver staining techniques may increase the sensitivity by a factor of 20–200 but they are more expensive and time-consuming than CBB staining.[2] Since the number of samples per gel is limited and differences in the degree of staining between different protein components exist, electrophoresis can only be used as a qualitative or semiquantitative analytical tool for N compounds in foods.

Sample preparation in the case of conventional electrophoresis is usually very simple and often consists of mixing or grinding the product to be analyzed with a suitable buffer, containing urea (6–8 M) and ME (0.1–5%) or DTT to dissociate protein complexes. A prefractionation of proteins can be achieved by selective precipitation at a certain pH.[96,97] In the case of fat-containing products, e.g. cheeses or milk, the fat can be removed by centrifugation, with[96] or without[98] the addition of organic solvents. For SDS/PAGE, boiling proteins at slightly alkaline pH (~8) in the presence of SDS (~0.1%) and ME is usually included in the sample preparation. Disulfide linkages are reduced and the dissociated chains bind SDS and adopt a random coil configuration. The denaturating capacity of SDS is sometimes unfavorable since it changes the naturally occurring state of proteins. If this native stage is required, e.g. for detection of electrophoretically separated proteins by immunological

techniques (immunoblotting), the cationic detergent *N*-cetyl-*N*,*N*,*N*-trimethylammonium bromide (CTAB) can be used instead.[99]

Electrophoresis can also be performed in free solution if the heat produced can be dissipated efficiently. Tubes with small inner diameter have a high surface-to-volume ratio, which facilitates heat dissipation. In the early 1980s, the use of capillaries for electrophoretic separations of analytes was demonstrated[100] and ever since CE has developed into a powerful separation technique. It is ideal for handling small sample amounts and is similar to HPLC in its ease of operation, speed, on-line detection, and full automation capabilities.[101] The high efficiency of CE is due to its characteristic flat flow profile, which allows for narrower peaks and therefore a higher resolution than conventional HPLC. Furthermore, CE requires no staining and destaining steps, and shows real time separation. However, the use of small sample amounts often means a high concentration limit of detection and considerable effort is devoted to improving the concentration sensitivity of the technique. Different detection methods are being used with CE: UV/VIS absorbance, fluorescence, derivatization, mass spectrometry, and indirect, laser-based, or electrochemical detection. However, only UV/VIS and to a lesser extent derivatization are of practical importance for proteins.[102] CE is almost exclusively performed in fused-silica tubing which exhibits good thermal properties, is transparent to UV/VIS light and can be made with an inner diameter of less then 100 μm.[10] Typical for fused-silica capillaries is that, for a wide range of pH values, the inner wall is negatively charged owing to ionization of the silanol groups. Consequently, it attracts cationic counterions from the bulk solution with the formation of an electric double layer at the silica–solution interface. When an electric field is applied, the hydrated cationic species will migrate towards the cathode causing bulk solutes to move in the same direction. This phenomenon is called electroosmotic flow (EOF) and is often greater than the electrophoretic mobility of any molecule present in solution.[10] Different modes of operation[10] and detection modes[103] have been developed.

For FSCE, the capillary is filled with electrolyte and the ends are immersed in vials containing electrolyte or sample solution. Because of differences in the intrinsic electrophoretic mobilities and the phenomenon of EOF, both cations and anions can be separated in a single run. Neutral molecules, however, will not be resolved since they will travel in one zone with a velocity equal to that of the EOF. Although proteins can be separated by simple FSCE, precautions have to be taken since proteins tend to adsorb to the capillary wall through a variety of mechanisms,[104] causing broadening and/or tailing of peaks and sometimes poor quantification. This can be (partially) overcome by using a high- or low-pH buffer, a high-ionic-strength buffer, coated capillaries, or buffer additives. Since protein adsorption can influence the migration time reproducibility, rinsing the capillary between injections, e.g. with water, NaOH, air, N_2, and buffer for a short period of time, is often necessary. Which approach is used depends very much on the proteins to be separated.[105] Buffer additives are exploited to separate neutral molecules and to improve the separation of proteins and peptides by the inclusion of micelle-forming detergents or surfactants in the electrophoretic medium; this is called micellar electrokinetic chromatography (MEKC).[10,104,106] Surfactants exhibit both hydrophobic and hydrophilic character and, in an aqueous environment, self-aggregate into micelles if the surfactant concentration exceeds the critical micelle concentration (cmc). During an MEKC separation, proteins associate with micelles through hydrophobic, hydrophilic, and electrostatic mechanisms. Many successful CE separations of proteins have been facilitated by the presence of both charged, e.g. SDS, and neutral or zwitterionic surfactants.[104] In addition to protein adsorption on the inner wall, the introduction of a well-defined sample volume can be a problem. Samples can be introduced into the column hydrodynamically or by electrokinetic introduction. The former can be achieved by applying a positive pressure to the inlet of the capillary or negative pressure at the outlet, for a specified period of time. Electrokinetic sample introduction is based on the transport of an aliquot of the sample in the column by the EOF. Parameters to be controlled are time, temperature, and liquid levels at inlet and outlet reservoirs. It seems that hydrodynamic injection is more reproducible than electrokinetic injection;[107] however, the use of internal standards is often required for either introduction technique. This approach permits corrections for the variations in introduced volume and for losses during sample cleanup. A further complication is that analytes do not pass the detector all at the same speed since differences in migration velocities are the basis for separation. Zones traveling past the detector at high velocity will produce sharper peaks with smaller area than to slow-traveling zones, even though they might have the same concentration and extinction coefficients.[108,109] Since this is a systematic bias, peaks can be corrected for differential mobilities by normalization with their retention time.[102] Although sharp peaks can be obtained which facilitate integration, the small sample volume introduced together with the short optical pathlength often result in high concentration limits for detection. Different attempts are being made to overcome this major drawback, including modifying capillary geometries,[110] preconcentration of analytes, e.g. by stacking,[111] indirect detection methods,[112] and the development of alternative detection modes.[103,113]

Since conventional electrophoresis is generally accepted as being a qualitative rather than a quantitative analytical tool, the quantification of proteins and peptides is mostly limited to CE. CE has the advantage over HPLC that it has high theoretical plate numbers (200 000–500 000),[114] although results are still less repeatable for complex mixtures.[10]

A good example of the resolving power of electrophoresis is the separation of wheat proteins from different cultivars where electrophoretic techniques permit differentiation between closely related wheat cultivars.[115,116] A main problem in cultivar identification by using conventional electrophoresis arises from experimental variations such as partial deformations on the whole gel and partial distortion of single bands.[117] This led to a lot of effort being put into the development of HPLC methods but not all separations were performed as well as with electrophoresis.[118] A good separation of individual milk proteins either in milk or dairy products can be achieved by conventional electrophoresis methods such as SDS/PAGE and IEF.[92,119,120] Separation of genetic variants is usually carried out by IEF or urea–PAGE since the genetic variants have comparable MWs but slightly different IEPs.[121] Since its introduction, CE has become a popular technique to analyze milk proteins.[71,105,114,122] Separation of the genetic variants of milk proteins is strongly dependent on buffer pH and ionic strength, and analysis times comparable to those in HPLC are often necessary for a complete resolution of caseins and whey proteins in a single run. Further optimizations have been done and variants differing in only a single amino acid could be separated as long as this substitution involved a change in the net charge at pH 3.0.[123] Quantification of the major whey proteins has been continuously improved during the past 5 years[124,125] and current research on CE of milk proteins is mainly directed towards the optimization of the separation conditions, e.g. the inclusion of surfactants in the separation buffer,[126] in order to reduce the analysis time further[122] and to have a good quantification of all milk proteins in a single run.[126] The possibilities of conventional electrophoresis for identification purposes were very well realized during the early days of electrophoresis.[127] Since the early research in the 1970s and 1980s, much effort has been directed to refining the procedures, to lowering the detection limits, and to reducing the time required for an analysis. Detection of adulterations by electrophoretic techniques have been reported for both raw and cooked meat and meat products.[99,128–131] Detection of the incorporation of low-quality meat such as mechanically recovered meat (MRM) in all types of meat products, including cooked meat products, has also been reported, with detection limits as low as 5–10%.[132] The identity of raw[133] and cooked[134] fish is determined with differences between species being much greater with IEF than with SDS/PAGE.[135] Semiquantitative results with SDS/PAGE can be obtained with the inclusion of an internal standard, e.g. BSA for the determination of beef myofibrillar proteins[136] or the peptide Lys-Trp-Lys.[137] Adulteration of high-quality products with lower quality or lower priced ingredients is also of importance in the dairy industry.[138] Sheep, goat and buffalo milk are all used to make cheese and detection of the use of mixtures of milks from different species is of great economic importance. Differences in the electrophoretic pattern of the milk of the various species can be found in the casein fraction[98,127,139–142] and whey proteins.[97,143] Changes in the protein fraction during processing, e.g. protein hydrolysis during the ripening of cheese[140] or aging of dry cured ham,[144] are succesfully studied by both conventional electrophoresis[96,145–149] and CE.[52,150] The importance of different factors during processing has also been evaluated by electrophoresis.[149,151–153] Protein denaturation can be evaluated by electrophoresis, e.g. protein denaturation by heat[154] or high pressure.[155] Applications of electrophoresis to proteins other than milk, meat, and cereal proteins are much less numerous but have been reported for, e.g. garlic,[156] kiwifruit,[157] peanut,[158] and fermented beverages.[159–161] Examples of analysis conditions for electrophoresis are given in Table 4.

3.4 Immunological Detection Methods for Food Proteins

Several protein detection methods based on immunology have been described in review articles.[2,11] ELISA seems particularly suitable for the routine analysis of proteins in food matrices. These immunoassays are based on the specific interaction between an antigen (in this case the food protein) and an antibody which reacts specifically with the antigen. The antigen–antibody interaction can be detected by the use of enzymes, e.g. alkaline phosphatase[162–164] and peroxidase,[165–168] which catalyze the transformation of a colorless substrate into a colored product to a degree which is related to the amount of antigen present. The method is particularly suitable for application in multiwell plates, is relatively fast in operation, and can be applied by non-specialized laboratory personnel. A significant amount of experimental research has now been performed to evaluate the suitability of ELISA in the detection and quantification of protein compounds in food matrices. Applications are found in the detection of allergenic compounds in foods, e.g. peanut proteins,[165] gliadin,[169] and rice bran protein,[166] and in tracing the adulteration of foods, e.g. detection of milk proteins in goats' and ewes'

milk and cheese,[162,163] and detection of soy proteins in milk[170] and meat products.[169,171]

Different variants of ELISA techniques have been applied to determine proteins in food products. Common are *heterogeneous* assays where the enzyme activity is not affected by the antigen–antibody interaction and where a separation between bound and unbound materials occurs.[172] Several authors have used *indirect competitive ELISA* in food protein analysis.[162–164,167,173] The antigen is immobilized to a solid support, after which the primary antibody and the sample (containing the antigen) is added to the solid phase. The primary antibody binds to the antigen present in the sample which prevents the antibody from binding to the immobilized antigen. After removal of unbound material, the antibody attached to the solid phase is determined by addition of an enzyme-labeled secondary antibody, which specifically interacts with the remaining primary antibody, and an enzyme substrate.[11,172] Methods based on this principle were used to detect the presence of cows' milk and caseinate in goats' and ewes' cheese,[162] and to evaluate the residual antigenicity of whey proteins after hydrolysis with endo- and exopeptidases[167] (Table 4).

Proteins in foods have also been determined by *double antibody sandwich ELISA*.[165] Antibodies raised against the protein are coated on a solid support, after which the sample containing the antigen is added. Unbound material is removed by washing and bound antigen detected by a second antibody labeled with an enzyme.[172] The method was used by Hefle et al.[165] to quantify peanut proteins in extracts from different foods with a concentration range from 40 μg peanut protein mL^{-1} up to 2.0 mg mL^{-1}. The secondary antibody was labeled with biotin, which was coupled in a separate step to peroxidase by streptavidin.

A critical step in the development of immunoassays is the formation of antibodies which are highly specific for the component of interest. Both monoclonal[165,176] and polyclonal[11,163,165,166,173] antibodies have been developed for analysis of food proteins. Cross-reactivity with closely related antigens may reduce the specificity of the assay, e.g. indirect competitive ELISA developed by Richter et al.[162] for detection of bovine casein in ewes' and goats' milk cheese shows cross-reactivity with closely related buffalo γ-caseins. Nevertheless, by using immunoabsorption chromatography, they were able to remove all antibodies cross-reacting with ewes' and goats' milk protein. The specificity of the antibody can also be influenced by the three-dimensional structure of the protein–antigen, e.g. the immunoreactivity of food proteins can be altered due to denaturation during manufacturing. Beer et al.[163] succeeded in developing an ELISA to detect both native and heat-denaturated bovine β-lactoglobulin in ewes' and goats' milk cheese. Since the degree of denaturation of bovine whey proteins used in adulterated cheese is not exactly known, it was chosen to use an extra standardized heat treatment (30 min at 90 °C) during the sample preparation to reduce the influence of individual thermal history of the samples. The method was found suitable to detect adulteration of ewes' and goats' milk cheese with ultrahigh temperature (UHT) and pasteurized cows' milk. In the case of soy proteins, an antiserum was produced by using a standardized denaturation and renaturation procedure.[171] Immunoreactivity was found against glycinin, which was denatured by boiling for 30 min in 10 M urea. Nevertheless, nonspecific interactions occurred with urea-denatured casein. Meisel[170] developed antibodies with high specificity but still suitable for detection of glycinin A subjected to different food processing conditions, by denaturating the proteins at 100 °C for 5 min in the presence of SDS and DTT. At a concentration of 0.05%, the SDS was able to reduce nonspecific interactions of the antibody with faba bean and pea proteins. The effect may be due to blocking by SDS of binding sides involved in nonspecific antigen–antibody interactions. Hydrolysis of a whey protein concentrate (WPC) with corolase 7092 (a technical mixture of endo- and exopeptidases derived from *Aspergillus* sp.) resulted in a decrease in the antigenicity of BSA and Ig up to a degree of hydrolysis (DH) of 25%, after which a significant increase in antigenicity was found. It was suggested that prolonged enzyme action let to the exposure of more antigenic sides due to protein unfolding.[167] The binding ability of proteins for antibodies can also be altered by the degree of lipophilization. Akita and Nakai[164] found that during indirect competitive ELISA the interaction with immunoglobulin G (IgG) increased at low and medium levels of fatty acid incorporation (0.3 mol of stearic acid per mole of lysine), while the binding activity decreased when more fatty acids were attached to the protein (>13 mol stearic acid per mole of lysine). The result was attributed to differences in the degree of unfolding in function of lipophilization, which can alter the exposure of antigenic sides during ELISA.[164] In addition to the development of highly specific antibodies, the degree of nonspecific binding in ELISA can also be reduced by treatment of the solid support with a suitable blocking agent.[171] After coating the plates with antigens or antibodies, unoccupied sides remain on the support where nonspecific binding can take place. Blocking agents frequently used are gelatin,[11,163,167,170] BSA,[165,171] and ovalbumin.[168]

Heddleson et al.[177] indicated that peptides with MW below 2500 Da are not immunogenic without modification, whereas proteins with MW below 10 kDa show only a weak immunogenic response. Rats injected with caseinphosphopeptides derived from tryptic hydrolysis of β-casein did not produce IgG antibodies that

cross-reacted with β-casein or skim-milk proteins.[177] Ena et al.[167] demonstrated that permeates of WPC hydrolyzates obtained by ultrafiltration and showing the same DH or a similar MW distribution gave differences in antigenicity of BSA and IgG compared with α-lactalbumin and β-lactoglobulin. This led to the conclusion that the residual antigenicity of whey proteins after hydrolysis is determined by the specificity of the enzymes used rather than by the DH or the MW distribution of the protein hydrolyzates. Also for chickpea proteins a dominating influence of enzyme activity was found. The antigenicity of native chickpea proteins during indirect competitive ELISA was reduced by 58% after hydrolysis with α-chymotrypsin compared with 45% for bromelain.[178]

4 CERTIFIED REFERENCE MATERIALS

Certified reference materials (CRMs) are an important tool in the validation of methods used to analyze N-containing compounds in foods. These materials are accompanied by a certificate, giving certified values for the amount of N present in the food matrix with an indication of the degree of uncertainty at a stated level of confidence (usually 95% confidence interval).[179] Several CRMs to analyze N-containing components in foods are available at the Institute for Reference Materials and Measurements (IRMM) of the European Commission. Kjeldahl N has been determined in rye flour (CRM 381), wheat flour (CRM 382 and CRM 563), beans (CRM 383), and pork muscle (CRM 384).

CRMs are prepared in the form of dry and homogeneous powders with good stability characteristics. Skim-milk powder (CRM 063R) has been prepared by homogenizing spray-dried skim-milk powder under dry argon in a mixing drum for a period of 14 days.[180] Lyophilized pork muscle (CRM 384) is derived from ca. 200 kg of muscle separated from connective tissue and visible fat in lean pigs. The muscle is minced and blended with subcutaneous pork fat and antioxidant (α-tocopherol), frozen at −20 °C, and lyophilized.[181] Powders are packaged under dry N_2 and stored at 4 °C. Samples of the powdered reference materials are analyzed by different laboratories with a high level of expertise, where each laboratory is permitted to use its own well-established method of analysis. Although for the determination of Kjeldahl-N in pork muscle a wide range of catalysts ($CuSO_4$, HgO, Se, TiO_2), reaction conditions (application of different digestion times and amounts of H_2SO_4, presence or absence of H_2O_2), and equipment (automated versus manual titration) have been used, low levels of uncertainty were obtained, i.e. 13.7 ± 0.2 g Kjeldahl-N per 100 g dry matter for pork muscle.[181] Results are expressed on a dry matter basis, since moisture was not always homogeneously distributed between the sachets, and because of the risk of gaining or losing moisture during storage or handling of the samples. CRMs based on Kjeldahl-N can be used to check the performance of Kjeldahl-N determination methods (section 2.1), and also for the calibration of N (protein) determination methods which require Kjeldahl as a reference method (e.g. NIR reflectance, section 2.2). For natural skim-milk powder (CRM 063R) and single-cell protein (CRM 273), only indicative values could be given for Kjeldahl-N, in view of the large spread of results obtained during interlaboratory tests. However, a certified value for total N was obtained for methods based on combustion and/or chemical conversion of N compounds followed by separation and determination of the N_2 formed, e.g. by catharometry.[180] Two freeze-dried curd reference materials derived from a mixture of ewes' and goats' milk and suitable for the detection of adulteration of cheeses made from ewes' and goats' milk with cows' milk are available at IRMM (CRM 599). One material is not adulterated (0%) and the second one contains 1% of cows' milk. The CRMs can be used in the validation of detection methods which trace cows' milk proteins in ewes' and goats' milk curds, e.g. IEF (section 3.3).

5 CONCLUSIONS

Analysis of N compounds in food is not always a simple task to perform and is often determined by the objective of the analysis. Conventional methods such as Kjeldahl, spectroscopy, and electrophoresis, together with more recent developments such as HPLC, CE, and immunology, offer the analyst a wide choice of techniques. However, all these methods somewhat lack sufficient flexibility to adapt to the chemical complexity in the composition of foods and possible changes of food proteins during manufacturing and processing. This "handicap" is especially prominent in the detection of adulteration of foods, where very low detection limits together with a high degree of certainty is often required. A possible solution is the use of more than one separation technique (two-dimensional analysis). This has been an approach to refine further the identification of wheat varieties.[181–183] A recent approach is the use of multivariate statistics such as multiple regression, principal component analysis, discriminant analysis, or neural networks, and the use of chemometrics. Applications can be found in the recognition of electrophoretic patterns of wheat proteins[117,184] and HPLC analysis of

mixed foods.[90] As shown, immunological detection is a promising technique, but further research in this area is necessary. For the detection of peptides with a low immunogenicity, formation of antibodies can be increased by conjugation through glutaraldehyde cross-linking to a larger carrier, e.g. hemocyanin[172] or ovalbumin.[168] Antibodies raised against E8E, an octapeptide in fragment 141–148 with glutamic acid (E) at the C-terminus (specific for bovine α_{s1}-casein) and E8Q, which differs from E8E by replacement of glutamic acid by glutamine (Q) at the C-terminus (specific for ovine, water buffalo, and caprine α_{s1}-casein), did not cross-react with the heptapeptide E7P in fragment 141–147 lacking glutamic acid or glutamine at the C-terminus (common in the four species of α_{s1}-casein). Addition of extra amino acids to the N-terminus of E8E gave an immune response for anti-E8E but not for anti-E8Q. Care has to be taken when the antigenicity of peptides is used to detect whole protein fractions in foods. Bovine casein reacted with anti-E8Q although the bovine α_{s1}-casein fraction does not contain the E8Q-peptide. This may be due to the presence of interfering compounds in whole casein, which further have to be identified by additional techniques, e.g. immunoblotting.[134]

ABBREVIATIONS AND ACRONYMS

BSA	Bovine Serum Albumin
CBB	Coomassie Brilliant Blue
CE	Capillary Electrophoresis
cmc	Critical Micelle Concentration
CRM	Certified Reference Material
CTAB	*N*-Cetyl-*N*,*N*,*N*-trimethylammonium bromide
DH	Degree of Hydrolysis
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme Linked Immunosorbent Assay
EOF	Electroosmotic Flow
FSCE	Free-solution Capillary Electrophoresis
GC	Gas Chromatography
HIHPLC	Hydrophobic Interaction High-performance Liquid Chromatography
HPLC	High-performance Liquid Chromatography
IEF	Isoelectric Focusing
IEHPLC	Ion-exchange High-performance Liquid Chromatography
IEP	Isoelectric Point
Ig	Immunoglobulin
IgG	Immunoglobulin G
IR	Infrared
IRMM	Institute for Reference Materials and Measurements
ME	β-Mercaptoethanol
MEKC	Micellar Electrokinetic Chromatography
MRM	Mechanically Recovered Meat
MW	Molecular Weight
NIR	Near-infrared
NPN	Nonprotein Nitrogen
PAGE	Polyacrylamide Gel Electrophoresis
PCA	Perchloric Acid
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEHPLC	Size-exclusion High-performance Liquid Chromatography
TCA	Trichloroacetic Acid
TFA	Trifluoroacetic Acid
Tris	Tris(hydroxymethyl)aminomethane
UHT	Ultrahigh Temperature
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible
WPC	Whey Protein Concentrate

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Infrared Spectroscopy of Biological Applications

Food **(Volume 5)**
Electrophoresis and Isoelectric Focusing in Food Analysis • Enzyme Analysis and Bioassays in Food Analysis • Fluorescence Spectroscopy in Food Analysis • Near-infrared Spectroscopy in Food Analysis

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Gel Electrophoresis in Protein and Peptide Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

Liquid Chromatography **(Volume 13)**
Biopolymer Chromatography

REFERENCES

1. J. Cheftel, J. Cuq, D. Lorient, 'Amino Acids, Peptides, and Proteins', in *Food Chemistry*, 2nd edition, ed. O. Fennema, Marcel Dekker, New York, 241–371, 1985.
2. J. Van Camp, A. Huyghebaert, 'Proteins', in *Handbook of Food Analysis: Physical and Nutrient Analysis*, ed. L. Nollet, Marcel Dekker, New York, 282–290, Vol. 1, 1996.
3. D. Boulter, E. Derbyshire, 'The General Properties, Classification and Distribution of Plant Proteins', in *Plant Proteins*, ed. G. Norton, Butterworths, London, 3–21, 1978.
4. T. Creighton, *Proteins: Structure and Molecular Properties*, 2nd edition, W.H. Freeman, New York, 1996.
5. R. Ludescher, 'Physical and Chemical Properties of Amino Acids and Proteins', in *Food Proteins: Properties and Characterization*, ed. S. Nakai, VCH Publishers, New York, London, Cambridge, 23–70, 1996.
6. S. Damodaran, 'Functional Properties', in *Food Proteins: Properties and Characterization*, ed. S. Nakai, VCH Publishers, New York, London, Cambridge, 67–224, 1996.
7. L. Gaspar, 'General Laboratory Methods', in *Methods of Protein Analysis*, ed. I. Kerese, Ellis Horwood, Chichester, 30–86, 1980.
8. H. Guillou, J.P. Pelissier, R. Grappin, 'Méthodes de dosage des Protéines du Lait de vache', *Lait*, **66**, 143–175 (1986).
9. B.G. Osborne, 'Determination of Protein in Cereals', in *Developments in Food Proteins*, ed. B.J.F. Hudson, Elsevier Applied Science, London, 247–290, Vol. 4, 1986.
10. J. Lindeberg, 'Capillary Electrophoresis in Food Analysis', *Food Chem.*, **55**, 73–94 (1996).
11. C.J. Smith, 'Evolution of the Immunoassay', in *Development and Application of Immunoassay for Food Analysis*, ed. J.H. Rittenburg, Elsevier Applied Science, London, 3–27, 1990.
12. H. Egan, R.S. Kirk, R. Sawyer, 'General Chemical Methods', in *Pearson's Chemical Analysis of Foods*, eds. H. Egan, R.S. Kirk, R. Sawyer, Longman, London, 7–37, 1981.
13. N.D. Heidelbaugh, C.S. Huber, J.F. Bednarczyk, M.C. Smith, P.C. Rambaut, H.O. Wheeler, 'Comparison of Three Methods for Calculating Protein Content of Foods', *J. Agric. Food Chem.*, **23**, 611–613 (1975).
14. J.M. Regenstein, C.E. Regenstein, 'Protein Quantitation', in *Food Protein Chemistry*, eds. J.M. Regenstein, C.E. Regenstein, Academic Press, London, 90–108, 1984.
15. A. Wolfschoon-Pombo, H. Klostermeyer, 'Nonprotein N Fraction in Cows' Milk. I. Amount and Composition', *Milchwissenschaft*, **36**, 598–600 (1981).
16. A. Wolfschoon-Pombo, H. Klostermeyer, G. Weiss, 'Nonprotein N Fraction in Cows' Milk. V. Analytical Methods', *Milchwissenschaft*, **37**, 80–85 (1982).
17. A.L. Lakin, 'Determination of Nitrogen and Estimation of Protein in Food', in *Developments in Food Analysis Techniques*, ed. R.D. King, Elsevier Applied Science, London, 43–74, Vol. 1, 1978.
18. L.V. Coveney, 'Application of Ion Selective Electrodes and Gas Sensing Probes to Food Analysis – a Survey', *Sci. Tech. Surv.*, **118**, 46–47 (1980).
19. P.M. Todd, 'The Use of the Ammonia Selective Electrode for the Determination of Nitrogen in Meat Products', *J. Sci. Food Agric.*, **24**, 488–489 (1973).
20. C. Nkonge, G.M. Ballance, 'A Sensitive Colorimetric Procedure for Nitrogen Determination in Micro-Kjeldahl Digests', *J. Agric. Food Chem.*, **30**, 416–420 (1982).
21. F. Kulcsar, L. Gaspar, M. Csont, 'A New Method for Detecting Ammonia in Kjeldahl Protein Analysis', *Proc. Eur. Meet. Meat Res. Workers*, **25**, 521–523 (1979).
22. W. Arneth, B. Herold, 'Studies on the Hach System for Rapid Analysis of Nitrogen in Meat and Meat Products', *Mitt. Bundesanstalt Fleischforsch., Kulmbach*, **110**, 410–415 (1990).
23. H. Konrad, 'The Semi-micro Kjeldahl Determination of Milk Protein as a Reference Method', *Nahrung*, **13**, 143–149 (1969).
24. B.G. Venter, H.P. Scheepers, J. Floor, J.H. Snyman, 'Semi-micro Kjeldahl Procedures for Protein Determination of Dairy Products', *S. Afr. J. Dairy Technol.*, **17**, 107–111 (1985).
25. M. Collomb, M. Spahni, G. Steiger, 'Kjeldahl Determination of Nitrogen and Nitrogen Fractions in Dairy Products Using an Automatic System', *Mitt. Geb. Lebensm. Hyg.*, **81**, 499–509 (1990).
26. P.C. Williams, B.N. Thompson, D. Wetzel, G.W. McLay, D. Loewen, 'Near-infrared Instruments in Flour Mill Quality Control', *Cereal Foods World*, **26**, 234–237 (1981).
27. R. Tkachuk, 'Protein Analysis of Whole Wheat Kernels by Near Infrared Reflectance', *Cereal Foods World*, **26**, 584–587 (1981).
28. M. Mitsumoto, S. Maeda, T. Mitsuhashi, S. Ozawa, 'Near-infrared Spectroscopy Determination of Physical and Chemical Characteristics in Beef Cuts', *J. Food Sci.*, **56**, 1493–1496 (1991).
29. A. Ripoche, 'La Spectrométrie Infrarouge Pour Évaluer la Qualité de la Viande', *Viandes Produits Carnés*, **18**, 222–226 (1997).
30. M. Defernez, R.H. Wilson, 'Mid-infrared Spectroscopy and Chemometrics for Determining the Type of Fruit Used in Jam', *J. Sci. Food Agric.*, **67**, 461–467 (1995).
31. M.M. Bradford, 'A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein–Dye Binding', *Anal. Biochem.*, **72**, 24–254 (1976).

32. M.J. Dunn, 'Determination of Total Protein Concentration', in *Protein Purification Methods: A Practical Approach*, eds. E.L.V. Harris, S. Angal, Oxford University Press, New York, 10–16, 1989.
33. R.K. Scopes, *Purification Principles and Practice*, 2nd edition, Springler-Verlag, New York, 1987.
34. M. Kroger, E.E. Katz, J.C. Weaver, 'Determining Protein Content of Ice Cream and Frozen Desserts', *J. Dairy Sci.*, **61**, 274–277 (1978).
35. N. Rawson, R.R. Mahony, 'A Modified Method for Determining of Reactive Lysine in Milk Powder Using Remazol Brilliant Blue R', *Lebensm. Wiss. Technol.*, **16**, 1–4 (1983).
36. C.T. Greenwood, D.J. Stevens, 'Determination of Protein in Wheat and Flour with the 'Research Association Protein Indicator Device' (Rapid-Meter)', *Veroeff. Arbeitsgemeinschaft Getreideforsch.*, **150**, 75–80 (1973).
37. D.L. Park, R.L. King, 'Evaluation of Automated Dye Binding Determination of Protein in Milk', *J. Assoc. Off. Anal. Chem.*, **57**, 42–47 (1974).
38. J.A. O'Connell, 'Evaluation and Modification of the Pro-Milk Tester Mk II for Protein Estimation in Milk', *Lab. Pract.*, **19**, 1119–1120 (1970).
39. M.J. Dennis, P.R. Ashurst, 'An Introduction to Food Authentication', in *Food Authentication*, eds. P.R. Ashurst, M.J. Dennis, Chapman & Hall, London, 1–14, 1997.
40. W. Büser, H.F. Erbersdobler, 'Gaschromatographische Aminosäurenbestimmung Mit Stickstoffselektiver Detektion', *Z. Lebensm.-Unters. Forsch.*, **186**, 509–513 (1988).
41. C.-H. Oh, J.-H. Kim, K.-R. Kim, T.J. Mabry, 'Rapid Gas Chromatographic Screening of Edible Seeds, Nuts and Beans for Nonprotein and Protein Amino Acids', *J. Chromatogr. A*, **708**, 131–141 (1995).
42. S. Dierckx, A. Huyghebaert, 'HPLC Analysis of Food Proteins', in *HPLC Analysis of Food*, ed. L. Nollet, Marcel Dekker, New York, Chapter 4, 2000.
43. Z. Deyl, J. Hyanek, M. Horakova, 'Profiling of Amino Acids in Body Fluids and Tissues by Means of Liquid Chromatography', *J. Chromatogr.*, **379**, 177–250 (1986).
44. H. Nielsen, R. Hurrell, 'Tryptophan Determination of Foods Proteins by HPLC After Alkaline Hydrolysis', *J. Sci. Food Agric.*, **36**, 893–907 (1985).
45. J.H. Baxter, 'Amino Acids', in *Handbook of Food Analysis. Vol. 1. Physical Characterization and Nutrient Analysis*, ed. L.M.L. Nollet, Marcel Dekker, New York, 197–228, 1996.
46. L. Ng, A. Pascaud, M. Pascaud, 'Hydrochloric Acid Hydrolysis of Proteins and Determination of Tryptophan by Reversed-phase High-performance Liquid Chromatography', *Anal. Biochem.*, **173**, 47–52 (1987).
47. V. Lebet, H. Schneider, E. Arrigoni, R. Amadó, 'Kritische Beurteilung der Proteinbestimmung nach der Methode von Kjeldahl Aufgerund der Aminosäureanalyze', *Mitt. Geb. Lebensm. Hyg.*, **85**, 46–58 (1994).
48. J.J. Córdoba, T.A. Rojas, C.G. González, J.V. Barroso, C.L. Bote, M.A. Asensio, 'Evolution of Amino Acids and Amines During Ripening of Iberian Cured Ham', *J. Agric. Food Chem.*, **42**, 2296–2301 (1994).
49. U. Bütikofer, D. Fuchs, J.O. Bosset, W. Gmür, 'Automated HPLC Amino Acid Determination of Protein Hydrolysates by Precolumn Derivatization with OPA and FMOC and Comparison with Classical Ion exchange Chromatography', *Chromatographia*, **31**, 441–447 (1991).
50. U. Bütikhofer, D. Fuchs, J.O. Bosset, 'Quantitative Bestimmung von Aminosäuren aus Hydrolysaten: Auswertung eines Ringversuches Mit klassischen Ionenaustausch-Aminosäureanalysatoren und HPLC-systemen Mit OPA/FMOC-vorsäulenderivatisierung', *Mitt. Geb. Lebensm. Hyg.*, **83**, 457–466 (1992).
51. I. Molnár-Perl, M. Khalifa, 'Analysis of Foodstuff Amino Acids Using Vapour-phase Hydrolysis', *LC/GC Int.*, **7**, 395–398 (1994).
52. E. Molina, M. Ramos, A. Cifuentes, R. López-Fandiño, 'Characterization of Cheese Proteolysis by Capillary Electrophoresis and Reverse-phase HPLC Analysis of Peptides', *Z. Lebensm.-Unters. Forsch. A*, **206**, 259–263 (1998).
53. E. Rodríguez-Nuñez, M.-C. Aristoy, F. Toldrá, 'Peptide Generation in the Processing of Dry-cured Ham', *Food Chem.*, **53**, 187–190 (1995).
54. K.D. Lee, J.J. Warthesen, 'Mobile Phases in Reverse-phase HPLC for the Determination of Bitter Peptides in Cheese', *J. Food Sci.*, **61**, 291–294 (1996).
55. J.M. Roturier, D. Le Bars, J.C. Gripon, 'Separation and Identification of Hydrophilic Peptides in Dairy Products Using FMOC Derivatization', *J. Chromatogr. A*, **696**, 209–217 (1995).
56. T. Herraiz, V. Casal, M.C. Polo, 'Reversed-phase HPLC Analysis of Peptides in Standard and Dairy Samples Using On-line Absorbance and Post-column OPA-Fluorescence Detection', *Z. Lebensm.-Unters. Forsch.*, **199**, 265–269 (1994).
57. T. Herraiz, 'Sample Preparation and Reversed-phase High-performance Liquid Chromatography Analysis of Food-derived Peptides', *Anal. Chim. Acta*, **352**, 119–139 (1997).
58. A.F. Groen, R. van der Vegt, M.A.J.S. van Boekel, O.L.A.M. de Rouw, H. Vos, 'Case Study on Individual Variation in Milk Protein Composition as Estimated by High-pressure Liquid Chromatography', *Neth. Milk Dairy J.*, **48**, 201–212 (1994).
59. C.M. Hollar, A.J.R. Law, D.G. Dagleish, R.J. Brown, 'Separation of Major Casein Fractions Using Cation-exchange Fast Protein Liquid Chromatography', *J. Dairy Sci.*, **74**, 2403–2409 (1991).
60. M. de Frutos, A. Cifuentes, J.C. Diez-Masa, L. Amigo, M. Ramos, 'Application of HPLC for the Detection of Proteins in Whey Mixtures from Different Animal

Species', *J. High Resolut. Chromatogr.*, **14**, 289–291 (1991).
61. A.J.R. Law, C.M. Papoff, D.G. Dalgleish, R.L. Campus, 'Quantitative Fractionation of Ovine Casein by Cation-exchange FPLC', *Milchwissenschaft*, **47**, 279–282 (1992).
62. C. Romero, O. Perez-Andújar, A. Olmedo, S. JiméNez, 'Detection of Cow's Milk in Ewe's or Goat's Milk by HPLC', *Chromatographia*, **42**, 181–184 (1996).
63. I. De Noni, A. Tirelli, F. Masotti, 'Determinazione del Latte di Vacca nei Formaggi di Specie Minori Mediante HPLC Delle Sieroprotiene: Applicazione ai Formaggi di Capra', *Sci. Tec. Latt.-Cas.*, **47**, 7–17 (1996).
64. H.K. Mayer, D. Heidler, C. Rockenbauer, 'Determination of the Presence of the Percentages of Cows', Ewes' and Goats' Milk in Cheese by Isoelectric Focusing and Cation-exchange HPLC of γ- and Para-κ-caseins', *Int. Dairy J.*, **7**, 619–628 (1997).
65. R. López-Fandiño, M.I. Acedo, M. Ramos, 'Comparative Study by HPLC of Caseinomacropeptides from Cows', Ewes' and Goats' Milk', *J. Dairy Res.*, **60**, 117–121 (1998).
66. S. Visser, C.J. Slangen, H.S. Rollema, 'Phenotyping of Bovine Milk Proteins by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **548**, 361–370 (1991).
67. S. Visser, C.J. Slangen, H.S. Rollema, 'Phenotyping of Bovine Milk Proteins by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **548**, 361–370 (1991).
68. C. García, L. Marina, M. Torre, 'Simultaneous Separation of Soya Bean and Animal Whey Proteins by Reversed-phase High-performance Liquid Chromatography. Quantitative Analysis in Edible Samples', *Anal. Chem.*, **69**, 2217–2220 (1997).
69. M.C. García, M.L. Marina, M. Torre, 'Simultaneous Separation of Soya Bean and Animal Whey Proteins by Reversed-phase High-performance Liquid Chromatography. Quantitative Analysis in Edible Samples', *Anal. Chem.*, **69**, 2217–2220 (1997).
70. A.J.R. Law, D.S. Horne, J.M. Banks, J. Leaver, 'Heat-induced Changes in the Whey Proteins and Caseins', *Milchwissenschaft*, **50**, 483–488 (1995).
71. I. Recio, M. de Frutos, A. Olano, M. Ramos, 'Protein Changes in Stored Ultra-high Temperature-treated Milks Studied by Capillary Electrophoresis and High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **44**, 3955–3959 (1996).
72. M. McGoldrick, P.F. Fox, 'Intervarietal Comparison of Proteolysis in Commercial Cheese', *Z. Lebensm.-Unters. Forsch. A*, **208**, 90–99 (1999).
73. G.L. Lookhart, R. Graybosch, J. Peterson, A. Lukaszewski, 'Identification of Wheat Lines Containing the 1BL/1RS Translocation by High-performance Liquid Chromatography', *Cereal Chem.*, **68**, 312–316 (1991).
74. F.R. Huebner, J.A. Bietz, 'Improved Chromatographic Separation and Characterization of Ethanol Soluble Wheat Proteins', *Cereal Chem.*, **70**, 506–511 (1993).
75. M.G. Scanlon, H.D. Sapirstein, W. Bushuk, 'Evaluation of the Precision of High-performance Liquid Chromatography for Wheat Cultivar Identification', *Cereal Chem.*, **66**, 112–116 (1989).
76. I. de Noni, G. de Bernardini, L. Pellegrino, 'Detection of Common-wheat (*Triticum aestivum*) Flour in Durum-wheat (*Triticum Durum*) Semolina by Reverse-phase High-performance Liquid Chromatography (RP-HPLC) of Specific Albumins', *Food Chem.*, **51**, 325–329 (1994).
77. P.K. McCarthy, B.F. Scanlon, I.C. Lumley, M. Griffin, 'Detection and Quantification of Adulteration of Durum Wheat Flour by Flour from Common Wheat Using Reverse-phase HPLC', *J. Sci. Food Agric.*, **50**(2), 211–226 (1989).
78. R.B. Gupta, I.L. Batey, F. MacRitchie, 'Relationships Between Protein Composition and Functional Properties of Wheat Flours', *Cereal Chem.*, **69**, 125–131 (1992).
79. M.G. Scanlon, P.K.W. Ng, D.E. Lawless, W. Bushuk, 'Suitability of Reversed-phase High-performance Liquid Chromatographic Separation of Wheat Proteins for Long-term Statistical Assessment of Breadmaking Quality', *Cereal Chem.*, **67**, 395–399 (1990).
80. N.K. Singh, R. Donovan, F. MacRitchie, 'Use of Sonication and Size-exclusion High-performance Liquid Chromatography in the Study of Wheat Flour Proteins. II. Relative Quantity of Glutenin as a Measure of Breadmaking Quality', *Cereal Chem.*, **67**, 161–170 (1990).
81. J. Sajdok, D. Požárková, P. Rauch, J. Káš, 'Thermal Denaturation of Hen Egg White Studied by Chromatographic and Immunological Techniques', *J. Food Sci.*, **54**, 906–908 (1989).
82. S. Takeuchi, T. Saito, T. Itoh, 'Rapid Analysis of Chicken Egg White Proteins Via High-performance Liquid Chromatography', *Anim. Sci. Technol. (Jpn.)*, **63**, 598–600 (1992).
83. S.H. Ashoor, M.A. Osman, 'Liquid Chromatographic Quantitation of Chicken and Turkey in Unheated Chicken–Turkey Mixtures', *J. Assoc. Off. Anal. Chem.*, **71**, 403–405 (1988).
84. I.M. Oellingrath, A. Iversen, G. Skrede, 'Quantitative Determination of Myoglobin and Haemoglobin in Beef by High-performance Liquid Chromatography', *Meat Sci.*, **28**, 313–320 (1990).
85. S.G. Armstrong, D.N. Leach, S.G. Wyllie, 'The Use of HPLC Protein Profiles in Fish Species Identification', *Food Chem.*, **44**, 147–155 (1992).
86. C. Piñeiro, C.G. Sotelo, I. Medina, J.M. Gallardo, R.I. Pérez-Martin, 'Reversed-phase HPLC as a Method for the Identification of Gadoid Fish Species', *Z. Lebensm.-Unters. Forsch. A*, **204**, 411–416 (1997).
87. I.L. Kukman, M. Zelenik-Blatnik, V. Abram, 'Isolation of Low-molecular Mass Hydrophobic Bitter Peptides

in Soybean Protein Hydrolysates by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. A*, **704**, 113–120 (1995).
88. S.M. Basha, 'Resolution of Peanut Seed Proteins by High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **36**, 778–781 (1988).
89. S.H. Ashoor, P.G. Stiles, 'Determination of Soy Protein, Whey Protein, and Casein in Unheated Meats by High-performance Liquid Chromatography', *J. Chromatogr.*, **393**, 321–328 (1987).
90. Z.L. Ma, Y.P. Wang, C.X. Wang, F.-Z. Miao, 'HPLC Determination of Muscle, Collagen, Wheat, Shrimp, and Soy Proteins in Mixed Food with the Aid of Chemometrics', *Am. Lab.*, **29**, 27–29 (1997).
91. S.H. Ashoor, W.C. Monte, P.G. Stiles, 'Liquid Chromatographic Identification of Meats', *J. Assoc. Off. Anal. Chem.*, **71**, 397–402 (1988).
92. L.K. Creamer, 'Electrophoresis of Cheese', *Bull. Int. Dairy Fed.*, **261**, 14–28 (1991).
93. I.M. Mackie, 'Authenticity of Fish', in *Food Authentication*, eds. P.R. Ashurst, M.J. Dennis, Chapman & Hall, London, 141–170, 1997.
94. T.E. Creighton, *Protein Structure. A Practical Approach*, 2nd edition, Oxford University Press, Oxford, 1997.
95. E.R. Maurer, *Disc Electrophoresis and Related Techniques of Polyacrylamide Gel Electrophoresis*, Walter de Gruyter, Berlin, 1971.
96. K. Dewettinck, S. Dierckx, P. EichwalDer, A. Huyghebaert, 'Comparison of SDS–PAGE Profiles of Four Belgian Cheeses by Multivariate Statistics', *Lait*, **77**, 77–89 (1997).
97. L. Amigo, I. Ibanez, C. Fernandez, G. Santa-Maria, M. Ramos, 'Comparison of an Electrophoretic and an Immunological Method for the Determination of Goat and Cow Milk in Cheese', *Milchwissenschaft*, **44**, 215–218 (1989).
98. F. Addeo, L. Moio, L. Chianese, C. Stingo, A. Di Luccia, 'Improved Procedure for Detecting Bovine and Ovine Milk Mixtures in Cheese by Isoelectric Focusing of Para-κ-casein', *Milchwissenschaft*, **45**, 221–224 (1990).
99. M. Körs, H. Steinhart, 'CTAB Electrophoresis and Immunoblotting: a New Method for the Determination of Soy Protein in Meat Products', *Z. Lebensm.-Unters. Forsch. A*, **205**, 224–226 (1997).
100. J.W. Jorgenson, K.D. Lukacs, 'Capillary Zone Electrophoresis', *Science*, **222**(4621), 266–272 (1983).
101. A. Guttman, 'Capillary Sodium Dodecyl Sulfate–Gel Electrophoresis of Proteins', *Electrophoresis*, **17**, 1333–1341 (1996).
102. W.G. Kuhr, 'Capillary Electrophoresis', *Anal. Chem.*, **62**, 403R–414R (1990).
103. M. Albin, P.D. Grossman, S.E. Moring, 'Sensitivity Enhancement for Capillary Electrophoresis', *Anal. Chem.*, **65**(10), 489A–497A (1993).
104. M.A. Strege, A.L. Lagu, 'Micellar Electrokinetic Chromatography of Proteins', *J. Chromatogr. A*, **780**, 285–296 (1997).
105. N. de Jong, S. Visser, C. Olieman, 'Determination of Milk Proteins by Capillary Electrophoresis', *J. Chromatogr. A*, **652**, 207–213 (1993).
106. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open-tubular Capillaries', *Anal. Chem.*, **56**, 111–113 (1984).
107. H. Wätzig, C. Dette, 'Capillary Electrophoresis (CE) – a Review. Strategies for Method Development and Applications Related to Pharmaceutical and Biological Sciences', *Pharmazie*, **49**, 83–96 (1994).
108. S. Hjertén, K. Elenbring, F. Kilár, J.-L. Liao, A.J.C. Chen, C.J. Siebert, M.-D. Zhu, 'Carrier-free Zone Electrophoresis, Displacement Electrophoresis and Isoelectric Focusing in a High-performance Electrophoresis Apparatus', *J. Chromatogr.*, **403**, 317–322 (1987).
109. X. Huang, W.F. Coleman, R.N. Zare, 'Analysis of Factors Causing Peak Broadening in Capillary Zone Electrophoresis', *J. Chromatogr.*, **480**, 95–110 (1989).
110. D.N. Heiger, P. Kaltenbach, H.-J. Sievert, 'Diode-array Detection in Capillary Electrophoresis', *Electrophoresis*, **15**, 1234–1237 (1994).
111. R.-L. Chien, D.S. Burgi, 'On-column Sample Concentration Using Field Amplification in CZE', *Anal. Chem.*, **64**, 489A–496A (1992).
112. E.S. Yeung, W.G. Kuhr, 'Indirect Detection Methods for Capillary Separations', *Anal. Chem.*, **63**, 275A–282A (1991).
113. 'Online Coupling of Capillary Electrophoresis to Electrospray Ionization Mass Spectrometry', Technical Note No. 12-5965–5985^{E}, Hewlett-Packard, Avondale, PA, 1997.
114. M. Kanning, M. Casella, C. Olieman, 'Milk and Soy Protein Analysis Using Capillary Zone Electrophoresis', *LC-GC Int.*, **6**, 701–704, 706 (1993).
115. B.L. Jones, G.L. Lookhart, S.B. Hall, K.F. Finney, 'Identification of Wheat Cultivars by Gliadin Electrophoresis: Electrophoregrams of the 88 Wheat Cultivars Most Commonly Grown in the United States in 1979', *Cereal Chem.*, **59**, 181–188 (1982).
116. L. Capelli, F. Forlani, F. Perini, N. Guerrieri, P. Cerletti, P.G. Righetti, 'Wheat Cultivar Discrimination by Capillary Electrophoresis of Gliadins in Isoelectric Buffers', *Electrophoresis*, **19**, 311–318 (1998).
117. I. Søndergaard, K. Jensen, B.N. Krath, 'Classification of Wheat Varieties by Isoelectric Focusing Patterns of Gliadins and Neural Networks', *Electrophoresis*, **15**, 584–588 (1994).
118. S.R. Bean, J.A. Bietz, G.L. Lookhart, 'High-performance Capillary Electrophoresis of Cereal Proteins', *J. Chromatogr. A*, **814**, 25–41 (1998).
119. D.L. Van Hekken, M.P. Thompson, 'Application of Phast System® to the Resolution of Bovine Milk Proteins

on Urea–Polyacrylamide Gel Electrophoresis', *J. Dairy Sci.*, **75**, 1204–1210 (1992).
120. H. Bovenhuis, A.J.M. Verstege, 'Improved Method for Phenotyping Milk Protein Variants by Isoelectric Focusing Using Phast System', *Neth. Milk Dairy J.*, **43**, 447–451 (1989).
121. W.N. Eigel, J.E. Butler, C.A. Ernstrom, H.M. Farrell Jr, V.R. Harwalkar, R. Jennes, R. McL Whitney, 'Nomenclature of Proteins of Cow's Milk: Fifth Revision', *J. Dairy Sci.*, **67**, 1599–1631 (1984).
122. B. Vallejo-Cordoba, 'Rapid Separation and Quantification of Major Caseins and Whey Proteins of Bovine Milk by Capillary Electrophoresis', *J. Capillary Electrophoresis*, **4**, 219–224 (1997).
123. I. Recio, M.-L. Pérez-Rodríguez, M. Ramos, L. Amigo, 'Capillary Electrophoretic Analysis of Genetic Variants of Milk Proteins From Different Species', *J. Chromatogr. A*, **768**, 47–56 (1997).
124. N.M. Kinghorn, G.R. Paterson, D.E. Otter, 'Quantification of the Major Bovine Whey Proteins Using Capillary Zone Electrophoresis', *J. Chromatogr. A*, **723**, 371–379 (1996).
125. J.A.H.J. Otte, K.R. Kristiansen, M. Zakora, K.B. Qvist, 'Separation of Individual Whey Proteins and Measurement of α-Lactalbumin and β-Lactoglobulin by Capillary Zone Electrophoresis', *Neth. Milk Dairy J.*, **48**, 81–97 (1994).
126. J.-F. Fairise, P. Cayot, 'New Ultrarapid Method for the Separation of Milk Proteins by Capillary Electrophoresis', *J. Agric. Food Chem.*, **46**, 2628–2633 (1998).
127. K.-P. Kaiser, I. Krause, 'Analytik von Proteinen in Lebensmitteln mit Elektrophoretischen un Chromatographischen Verfahren', *Z. Lebensm.-Unters. Forsch.*, **180**, 181–201 (1985).
128. A. Zerifi, C. Labie, G. Benard, 'SDS–PAGE Technique for the Species Identification of Cooked Meat', *Fleischwirtschaft*, **71**, 1060–1062 (1991).
129. B. Schröder, D. Horn, W. Darge, 'Tierartdifferenzierung Mittels Isoelektrischer Fokussierung bei erhitztem Fleisch', *Fleischwirtschaft*, **70**, 102–107 (1990).
130. A. Skrökki, O. Hormi, 'Composition of Minced Meat. Part A: Methods', *Meat Sci.*, **38**, 497–501 (1994).
131. A. Skrökki, O. Hormi, 'Composition of Minced Meat. Part B: a Survey of Commercial Ground Meat', *Meat Sci.*, **38**, 503–509 (1994).
132. H. Brown, L. Day, 'Detection of Mechanically Recovered Chicken Meat Using Capillary Gel Electrophoresis – A feasibility Study', R&D Report No. 22, Campden and Chorleywood Food Research Association, Chipping Campden, Gloucestershire, 1995.
133. I.M. Mackie, 'Identifying Species of Fish', *Anal. Proc.*, **27**, 89–92 (1990).
134. A.E. Scobbie, I.M. Mackie, 'The Use of Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis in Fish Species Identification – a Procedure Suitable for Cooked and Raw Fish', *J. Sci. Food Agric.*, **44**, 343–351 (1988).
135. A. Craig, A.H. Ritchie, I.M. Mackie, 'Determining the Authenticity of Raw Reformed Breaded Scampi (*Nephrops norvegicus*) by Electrophoretic Techniques', *Food Chem.*, **52**, 451–454 (1995).
136. E. Claeys, L. Uytterhaegen, B. Buts, D. Demeyer, 'Quantification of Beef Myofibrillar Proteins by SDS–PAGE', *Meat Sci.*, **39**, 177–193 (1995).
137. I. Recio, E. Molina, M. Ramos, M. de Frutos, 'Quantitative Analysis of Major Whey Proteins by Capillary Electrophoresis Using Uncoated Capillaries', *Electrophoresis*, **16**, 654–658 (1995).
138. E.L. Malin, J.J. Basch, J.J. Shieh, B.C. Sullivan, VH. Holsinger, 'Detection of Adulteration of Buttermilk Powder by Gel Electrophoresis', *J. Dairy Sci.*, **77**, 2199–2206 (1994).
139. M.P. Cattaneo, F. Nigro, G.F. Greppi, 'Analysis of Cow, Goat and Ewe Milk Mixtures by Capillary Zone Electrophoresis (CZE): Preliminary Approach', *Milchwissenschaft*, **51**, 616–619 (1996).
140. P.F. Fox, 'Proteolysis During Cheese Manufacture and Ripening', *J. Dairy Sci.*, **72**, 1379–1400 (1989).
141. R.M. Whitney, 'Proteins of Milk', in *Fundamentals of Dairy Chemistry*, 3rd edition, ed. N.P. Wong, Van Nostrand Reinhold, New York, 81–169, 1988.
142. *Off. J. Eur. Commun.*, L 142, 15/06/1996, 0015–0025 (1996).
143. L. Amigo, M. Ramos, P.J. Martin-Alvarez, M. Barbosa, 'Effect of Technological Parameters on Electrophoretic Detection of Cow's Milk in Ewe's Milk Cheeses', *J. Dairy Sci.*, **74**, 1482–1490 (1991).
144. S. Buscailhon, G. Monin, 'Time Related Changes in Nitrogen Fractions and Free Amino Acids of Lean Tissue of French Dry-cured Ham', *Meat Sci.*, **37**, 449–456 (1994).
145. T.M.I.E. Christensen, A.-M. Bech, H. Werner, 'Methods for Crude Fractionation (Extraction and Precipitation) of N Components in Cheese', *Bull. Int. Dairy Fed.*, **261**, 4–9 (1991).
146. F. Addeo, G. Garro, N. Intorcia, L. Pellegrino, P. Resmini, L. Chianese, 'Gel Electrophoresis and Immunoblotting for the Detection of Casein Proteolysis in Cheese', *J. Dairy Res.*, **62**, 297–309 (1995).
147. C. Carretero, A.J. Trujillo, M. Mor-Mur, R. Pla, B. Guamis, 'Electrophoretic Study of Casein Breakdown During Ripening of Goat's Milk Cheese', *J. Agric. Food Chem.*, **42**, 1546–1550 (1994).
148. T. Marshall, K.M. Williams, 'High Resolution Two-dimensional Electrophoresis and Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis of Cheese Proteins after Rapid Solubilization', *Electrophoresis*, **9**, 143–147 (1988).
149. J. Centeno, J.L. Rodrigez-Otero, A. Cepeda, 'Changes in the Protein Profile of Arzua Cheese (NW Spain) During Ripening', *Milchwissenschaft*, **49**, 319–322 (1994).
150. M. Strickland, B.C. Weimer, J.R. Broadbent, 'Capillary Electrophoresis of Cheddar Cheese', *J. Chromatogr. A*, **731**, 305–313 (1996).

151. Y.K. Jin, Y.W. Park, 'SDS–PAGE of Proteins in Goat Milk Cheeses Ripened Under Different Conditions', *J. Food Sci.*, **61**, 490–494 (1996).
152. P. Ollikainen, T. Kivelä, 'The Importance of Plasmin in Swiss-type Cheese Ripening', *Milchwissenschaft*, **44**, 204–206 (1989).
153. I. Recio, L. Amigo, M. Ramos, R. Lopez-Fandino, 'Application of Capillary Electrophoresis to Study the Proteolysis of Caseins', *J. Dairy Res.*, **64**, 221–230 (1997).
154. I. Recio, C. Olieman, 'Determination of Denatured Serum Proteins in the Casein Fraction of Heat-treated Milk by Capillary Zone Electrophoresis', *Electrophoresis*, **17**, 1228–1233 (1996).
155. J. Van Camp, W. Messens, J. Clement, A. Huyghebaert, 'Influence of pH and Calcium Chloride on the High-pressure Induced Aggregation of a Whey Protein Concentrate', *J. Agric. Food Chem.*, **45**, 1600–1607 (1997).
156. E. Mochizuki, T. Yamamoto, S. Suzuki, H. Nakazawa, 'Electrophoretic Identification of Garlic and Garlic Products', *J. AOAC Int.*, **79**, 1466–1470 (1996).
157. M. Möller, M. Kayma, H. Steinhart, A. Paschke, 'Isolation and Characterization of a Major Allergen in Kiwi Fruit', *Z. Lebensm.-Unters. Forsch. A*, **205**, 364–369 (1997).
158. S.M. Basha, 'Separation of Peanut Proteins by Capillary Electrophoresis', *J. Agric. Food Chem.*, **45**, 400–402 (1997).
159. K.M. Williams, T. Marshall, 'Protein Composition of Beer as Revealed by High-resolution Two-dimensional Electrophoresis', *Anal. Proc. Incl. Anal. Commun.*, **32**, 25–28 (1995).
160. E.J. Waters, W. Wallace, M.E. Tate, P.J. Williams, 'Isolation and Partial Characterization of a Natural Haze Protective Factor from Wine', *J. Agric. Food Chem.*, **41**, 724–730 (1993).
161. C. Luguera, V. Moreno-Arribas, E. Pueyo, M.C. Polo, 'Capillary Electrophoretic Analysis of Wine Proteins. Modifications During the Manufacture of Sparkling Wines', *J. Agric. Food Chem.*, **45**, 3766–3770 (1997).
162. W. Richter, I. Krause, C. Graf, I. Sperrer, C. Schwarzer, H. Klostermeyer, 'An Indirect Competitive ELISA for the Detection of Cows' Milk and Caseinate in Goats' and Ewes' Milk and Cheese Using Polyclonal Antibodies Against γ-Caseins', *Z. Lebensm.-Unters. Forsch. A*, **204**, 21–26 (1997).
163. M. Beer, I. Krause, M. Stapf, C. Schwarzer, H. Klostermeyer, 'Indirect Competitive Enzyme-linked Immunosorbent Assay for the Detection of Native and Heat-denaturated Bovine β-Lactoglobulin in Ewes' and Goats' Milk Cheese', *Z. Lebensm.-Unters. Forsch.*, **203**, 21–26 (1996).
164. E.M. Akita, S. Nakai, 'Lipophilization of β-Lactoglobulin: Effect on Allergenicity and Digestibility', *J. Food Sci.*, **55**, 718–723 (1990).
165. S.L. Hefle, R.K. Bush, J.W. Yunginger, F. Sun Chu, 'A Sandwich Enzyme-linked Immunosorbent Assay (ELISA) for the Quantitation of Selected Peanut Proteins in Foods', *J. Food Prot.*, **57**, 419–423 (1994).
166. P.S. Landers, B.R. Hamaker, 'Antigenic Properties of Albumin, Globulin, and Protein Concentrate Fractions from Rice Bran', *Cereal Chem.*, **71**, 409–411 (1994).
167. J.M. Ena, E.C.H. Van Beresteijn, A.J.P.M. Robben, D.G. Schmidt, 'Whey Protein Antigenicity Reduction by Fungal Proteinases and a Pepsin/Pancreatin Combination', *J. Food Sci.*, **60**, 104–110, 116 (1995).
168. N. Horn, M.I. Martinez, J.M. Martinez, P.E. Hernandez, M.J. Gasson, J.M. Rodriguez, H.M. Dodd, 'Production of Pediocin Pa-1 by *Lactococcus lactis* Using the Lactococcin A Secretory Apparatus', *Appl. Environ. Microbiol.*, **64**(3), 818–823 (1998).
169. J.H. Skerritt, 'Immunoassays of Non-meat Protein Additives in Foods', in *Development and Application of Immunoassay for Food Analysis*, ed. J.H. Rittenburg, Elsevier Applied Science, London, 81–126, 1990.
170. H. Meisel, 'Enzyme-linked Immunosorbent Assay and Immunoblotting Using IgY Antibodies Against Soybean Glycinin A', *Int. Dairy J.*, **3**, 149–161 (1993).
171. J.C. Allen, 'The Value of Immunoassays to Food Analysis', in *Development and Application of Immunoassay for Food Analysis*, ed. J.H. Rittenburg, Elsevier Applied Science, London, 29–57, 1990.
172. J.H. Rittenburg, 'Fundamentals of Immunoassay', in *Development and Application of Immunoassay for Food Analysis*, ed. J.H. Rittenburg, Elsevier Applied Science, London, 29–57, 1990.
173. R. Pizzano, M. Adalgisa Nicolai, F. Addeo, 'Antigenicity of the 139–149 α_{s1}-Casein Region in Different Species Revealed by ELISA and Immunoblotting Using Antipeptide Antibodies', *J. Agric. Food Chem.*, **45**, 2807–2813 (1997).
174. G.E. Vegarud, T.S. Molland, M.J. Brovold, T.G. Devold, P. Alestrøm, T. Steine, S. Rogne, T. Langsrud, 'Rapid Separation of Genetic Variants of Caseins and Whey Proteins Using Urea-modified Gels and Fast Electrophoresis', *Milchwissenschaft*, **44**, 689–691 (1989).
175. J.A. Bietz, G.L. Lookhart, 'Wheat Varietal Identification by Capillary Electrophoresis: an Inter-laboratory Comparison of Methods', *Lebensm. Wiss. Technol.*, **30**, 210–213 (1997).
176. A. Venien, D. Levieux, C. Astier, L. Briand, J.-M. Chobert, T. Haertle, 'Production and Epitopic Characterization of Monoclonal Antibodies Against Bovine β-Lactoglobulin', *J. Dairy Sci.*, **80**, 1077–1987 (1997).
177. R.A. Heddleson, O. Park, J.C. Allen, 'Immunogenicity of Casein Phosphopeptides Derived from Tryptic Hydrolysis of β-Casein', *J. Dairy Sci.*, **80**, 1971–1976 (1997).
178. F. Hindi-Tamelikecht, C. Dauphin, M. Hamon, J.P. Grangaud, D. Pradeau, 'Analytic and Immunologic Characterization of Chickpea (*Cicer arietinum*) Protein

Hydrolysates Obtained by Bromelain and α-Chymotrypsin', *J. Agric. Food Chem.*, **45**, 4758–4762 (1997).
179. *Terms and Definitions Used in Connection with Reference Materials*, ISO Guide 30, International Standards Organization, Geneva, 1992.
180. B. Griepink, H. Gonska, H. Muntau, 'The Certification of the Contents (Mass Fractions) of Nitrogen, Phosphorus, Chloride, Sodium, Potassium, Magnesium and Calcium and of the Kjeldahl-nitrogen Content in a Skim Milk Powder', Report EUR 9138 EN, Commission of the European Communities, Community Bureau of Reference, Brussels, 1983.
181. P. Hollman, A. Boenke, P. Finglas, P.J. Wagstaffe, 'The Certification of the Mass Fractions of Major Components and Essential Elements in a Whole Milk Powder and a Lyophilized Pork Muscle Reference Material', Report EUR 14554 EN, Commission of the European Communities, Community Bureau of Reference, Brussels, 1993.
182. S.R. Bean, G.L. Lookhart, 'Separation of Wheat Proteins by Two-dimensional Reversed-phase High-performance Liquid Chromatography Plus Free Zone Capillary Electrophoresis', *Cereal Chem.*, **74**, 758–765 (1997).
183. P. Picard, M. Bourgoin-Grenèche, M. Zivy, 'Potential of Two-dimensional Electrophoresis in Routine Identification of Closely Related Durum Wheat Lines', *Electrophoresis*, **18**, 174–181 (1997).
184. K. Jensen, C. Keşmir, I. Søndergaard, 'From Image Processing to Classification: IV. Classification of Electrophoretic Patterns by Neural Networks and Statistical Methods Enable Quality Assessment of Wheat Varieties for Breadmaking', *Electrophoresis*, **17**, 964–698 (1996).

Proximate Assays in Food Analysis

S. Suzanne Nielsen
Purdue University, West Lafayette, USA

1 Introduction **1**
1.1 Reasons for Measuring Food Components 1
1.2 Choice of Methods 1
1.3 Proximate Composition 1
1.4 Sampling and Sample Preparation 2
2 Moisture Analysis **2**
2.1 Oven Drying 2
2.2 Distillation 2
2.3 Karl Fischer 2
2.4 Physical Methods 3
3 Ash Analysis **3**
3.1 Dry Ashing 4
3.2 Wet Ashing 4
3.3 Microwave Ashing 4
4 Lipid Analysis **4**
4.1 Solvent Extraction 4
4.2 Wet Extraction Without Organic Solvent 5
4.3 Instrumental Methods 5
5 Protein Analysis **5**
5.1 Nitrogen Content Methods 5
5.2 Infrared Spectroscopy 5
5.3 Other Methods 5
6 Carbohydrate Analysis **6**
6.1 Total Carbohydrate 6
6.2 Mono- and Oligosaccharides 6
6.3 Starch 7
6.4 Dietary Fiber 7
Acknowledgments **7**
Abbreviations and Acronyms **7**
Related Articles **7**
References **7**

The proximate composition of foods includes moisture, ash, lipid, protein, and carbohydrate contents. These food components may be of interest in the food industry for product development, quality control (QC), or regulatory purposes. Analyses used may be rapid methods for QC or more accurate but time-consuming official methods. Sample collection and preparation must be considered carefully to ensure analysis of a homogeneous and representative sample, and to obtain accurate results. Specific methods of analysis for moisture, ash, lipid, protein, and carbohydrate are described in this article, focusing on the principle of each method, application, and sources of error.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Reasons for Measuring Food Components

Food products produced in the food industry require analysis as part of quality management, product development, or research programs, in efforts to monitor food composition and to ensure the quality and safety of the food.[1–3] Food components often are measured for the following reasons: (1) formulating and developing new products, (2) evaluating new processes for making food products, (3) identifying unacceptable products, (4) developing or checking a nutrition label, (5) checking the quality of raw ingredients, (6) checking the composition during processing, (7) determining the problem in consumer complaint samples, and (8) checking the composition of competitors' samples.

1.2 Choice of Methods

The choice of analytical method is dictated commonly by the nature of the sample and the specific reason for the analysis. For example, official methods, which are published by numerous organizations,[4–7] are used for nutrition labeling purposes or other legal requirements, and generally are more accurate than methods used for rapid QC or in-line processing measurements. Key factors in the choice of method often are speed, precision, accuracy, and durability. Also, it is necessary to validate any method for the specific type of food sample being analyzed.[1]

1.3 Proximate Composition

The proximate composition of foods includes the contents of the macrocomponents, specifically moisture, ash, lipid, protein, and carbohydrate.[8] Other food components, such as vitamins, pigments, flavors, non-nutritive sweeteners, and preservatives, constitute a much smaller proportion of the food. This article covers only the analyses for proximate composition, and not for other food components.

1.4 Sampling and Sample Preparation

All results obtained in analyzing foods depend on obtaining a representative sample and properly converting the sample to a form that can be analyzed.[2,9,10] In general, a sample is collected and analyzed to estimate the quality characteristics of the entire population. The sampling technique must be selected carefully to obtain accurate and precise estimates of the quality of the population. Selection of the sampling procedure is determined by the purpose of the analysis and the nature of the food product, test method, and population being investigated. Once selected for analysis, the sample often must be reduced in bulk or particle size by a method such as grinding. In grinding, care must be taken to avoid moisture loss or gain and to avoid contamination by metals from the grinder. Some samples must be protected prior to analysis from enzymatic degradation, lipid oxidation, and microbial growth and contamination.

2 MOISTURE ANALYSIS

The moisture content of foods is important for reasons such as food quality, cost and convenience of shipping, legal standards, and uniformity for expressing other analytical determinations (i.e. dry weight basis).[11,12] Total solids is the term that refers to the dry matter in food that remains after moisture removal. Moisture analysis may seem simplistic, but obtaining accurate and precise data can be very difficult. Numerous sources describe official, rapid, and in-line methods for moisture analysis including the principles, procedures, applications, and sources of error.[4–7,11–13]

2.1 Oven Drying

Oven drying methods for moisture analysis rely on heating the sample under specified conditions to evaporate water, with weight loss equal to the moisture content. Obtaining accurate and precise results depends on careful temperature control, preventing decomposition of other food constituents, preventing splattering or spillage, and complete removal of the moisture. Forced draft ovens, while being an official method for many foods and easy for handling a large number of samples, often require a time and temperature (e.g. 100–130 °C for 1–24 h) that may make the method too slow for QC applications and cause decomposition of other food constituents. Vacuum ovens allow for drying under reduced pressure, to reduce the drying time and decomposition. In microwave oven drying, the sample is placed between two fiberglass pads, weighed on a balance in the microwave oven, then heated using microwave energy under selected conditions of time and power. This method is considered accurate enough for routine moisture determinations, and has the advantage of obtaining results within minutes. Using another fast method, rapid moisture analyzers, the sample on an aluminum pan or filter paper is placed on a digital balance, then heated by an overhead heating element to evaporate moisture.

Microwave drying ovens and rapid moisture analyzers are considerably more expensive than forced draft or vacuum ovens, but the latter two methods require hours compared with minutes for analysis. Results from microwave drying ovens and rapid moisture analyzers are commonly checked against the reference methods using forced draft or vacuum ovens. Errors in moisture analysis (i.e. over- or underestimation) by oven drying methods occur if food particles are too large, volatile compounds are lost, lipids are oxidized, samples are very hygroscopic, food carbohydrates are altered, a crust is formed on the surface of the food, or the sample splatters.

2.2 Distillation

Distillation procedures involve codistilling the moisture in a food sample with a high-boiling-point solvent that is immiscible with water. The mixture that distills off is collected and measured by volume. Reflux distillation with toluene is the most commonly used distillation technique for moisture analysis. The food sample is heated in toluene, the toluene and water are codistilled, the mixture is collected, and the volume of water removed from the food is measured. Distillation methods prevent exposure of the sample to oxygen and cause less thermal decomposition than do oven drying methods, but distillation techniques are not adaptable to the routine testing that is possible with oven drying methods. Distillation methods have the greatest advantage with samples that are high in volatile compounds (e.g. spices) and lipids (e.g. nuts, oils).

2.3 Karl Fischer

The Karl Fischer method is based on a chemical reaction between the Karl Fischer reagent (KFR) and water in the food sample. The reaction involves the reduction of iodine by sulfur dioxide in the presence of water, to cause a change that can be detected visually or with a potentiometer in a volumetric titration of the water-containing sample with the KFR. A coulometric titration is preferable for products with very low levels of moisture. The volume of the KFR used in the titration, along with a KFR water equivalence determined using a primary standard, allow for the calculation of moisture content.

The Karl Fischer method is the preferred method for many low-moisture foods (e.g. dried fruits, candies,

chocolate, and roasted coffee). The method uses no heat and is rapid and sensitive, but is subject to error from interfering food constituents (e.g. ascorbic acid, carbonyl compounds, unsaturated fatty acids). Other sources of error include incomplete water extraction, atmospheric moisture, and moisture adhering to the wall of the Karl Fischer unit.

2.4 Physical Methods

Physical methods used commonly for a rapid determination of moisture content in foods include hydrometry, refractometry, and infrared (IR) spectroscopy. These methods do not involve the removal of water as is used in oven drying or distillation methods, or the chemical reaction of water as is used in the Karl Fischer method. In contrast, physical methods relate certain physical properties of the food to the moisture content. Some physical methods can be applied directly in the food processing line, for immediate control of the moisture/solids content.

2.4.1 Hydrometry

Hydrometry, the science of measuring specific gravity or density, is used commonly for rapid routine testing of moisture/solids content of foods such as beverages, salt brines, and sugar solutions. While a pycnometer or Westphal balance are used sometimes to measure the specific gravity of foods, various types of hydrometers are used most commonly. These include a lactometer for milk, Brix hydrometer for sugar solutions, Baumé hydrometer for salt solutions, alcoholometer for alcohol content of beverages, and Twaddell hydrometer for liquids heavier than water. A hydrometer is based on Archimedes' principle, which states that a solid suspended in a liquid will be buoyed by a force equal to the weight of the liquid displaced. The volume of test liquid displaced by an object of standard weight (i.e. hydrometer) is measured to determine the weight per unit volume of the liquid. The hydrometer placed in the liquid displaces a weight of liquid equal to its own weight. The relative density (i.e. specific gravity) of the sample is compared with that of water at the same temperature. The spindle of the hydrometer is calibrated to read specific gravity directly at 15.5 or 20 °C. This method requires temperature control and the use of a clean hydrometer.

2.4.2 Refractometry

Refractometry is based on the bending of light (i.e. refraction) as it passes from one medium (e.g. air) to another (e.g. food product). The bending of the light beam is a function of the media and the sines of the angles of incidence and refraction at any given temperature and pressure, and is thus a dimensionless constant that can be used to describe the nature of the media. This constant, the refractive index (RI), is the ratio of the sines of the angle of incidence and the angle of refraction. All chemical compounds have a characteristic RI. The RI varies with concentration of the compound, temperature, and wavelength of light. Therefore, by holding the temperature and wavelength of light constant, one can use the RI to determine the concentration of the compound of interest. Refractometers are commonly used as a rapid method to determine the moisture (total solids) content of liquid sugar products and condensed milk and the soluble solids content of fruits and fruit products. Abbe refractometers are popular benchtop units for laboratory use, but hand-held units are convenient for rapid testing at the processing line, and refractometers can be installed directly in a liquid processing line. Because of their application in determining solids contents of carbohydrate-based foods, refractometers are calibrated in °Brix (g of sucrose/100 g of sample).

2.4.3 Infrared Spectroscopy

IR spectroscopy can be used to determine not only moisture content of foods, but also lipid, protein, and carbohydrate contents. This method measures the absorption of radiation (near- or mid-IR) by molecules in foods. Different functional groups characteristic of the molecules in food absorb different frequencies of IR radiation. A sample is irradiated with a wavelength of IR light specific for the constituent of interest. The energy that is reflected or transmitted by the sample, which is inversely proportional to the energy absorbed, is measured as an indicator of the concentration of the food constituent. For accurate results, the IR spectrometer must be calibrated for each analyte to be measured, and the analyte of interest must be uniformly distributed in the sample. IR spectroscopy has a wide range of food applications, and has proven successful in the laboratory, at the processing line, and in-line.

3 ASH ANALYSIS

Ash in a food product represents the total mineral content, and refers to the inorganic residue remaining after either ignition or complete oxidation of organic matter. Ash determination is part of the proximate analysis for nutritional evaluation, and is the first step in the preparation of a food sample for the analysis of specific minerals. Ash content of foods is determined most commonly by dry, wet, or microwave ashing methods.[14,15] Many food products must be dried and/or ground before ash analysis. Sources of error for these

methods include use of an unrepresentative sample, loss of samples in a drying step before ashing, contamination with minerals, volatilization of elements, and incomplete combustion of organic matter.

3.1 Dry Ashing

In dry ashing, the sample is placed in a crucible and heated in a muffle furnace to a high temperature (500–600 °C). This incinerates all the organic matter, leaving inorganic material to be quantitated gravimetrically. The procedure is time-consuming (12–18 h) and loss of volatile elements is a risk, but the procedure is simple and a large number of samples can be handled at one time.

3.2 Wet Ashing

In wet ashing, the organic matter in a sample is oxidized using acids and oxidizing agents, leaving only inorganic matter. This is a liquid treatment that uses low temperatures, compared with the high temperature and drying conditions of dry ashing. Wet ashing is used primarily to prepare a sample for analysis of specific minerals because the minerals usually stay in solution and there is little or no loss from volatilization. However, wet ashing requires almost constant operator attention, a limited number of samples can be handled at one time, corrosive reagents are necessary, and an appropriate fume hood is required. If perchloric acid is used in the procedure, special precautions and facilities are required.

3.3 Microwave Ashing

Microwave systems are available for both wet and dry ashing. Programmed microwave wet digesters and muffle furnaces can dehydrate then ash the sample and exhaust the system. The time for analysis is reduced from hours with conventional dry ashing to minutes, but a limited number of samples can be processed at one time.

4 LIPID ANALYSIS

The term lipids refers to a wide range of compounds soluble in organic solvents but only sparingly soluble in water. The terms lipids, fats, and oils sometimes are used interchangeably, but fats and oils are distinguished generally by whether they are solid or liquid at room temperature. The lipid content of foods is determined commonly by solvent extraction methods, but wet extraction methods without an organic solvent are applicable to liquid samples.[16,17] As alternatives, rapid instrumental methods based on various principles have been developed for lipid analysis.

4.1 Solvent Extraction

Prior to using solvent extraction methods, samples may need to be subjected to drying, particle size reduction, and acid hydrolysis. The last treatment is used because a significant portion of lipids in some foods is bound covalently or ionically to proteins and carbohydrates, which makes direct extraction with solvents inefficient. In solvent extraction methods, the fat is extracted from the food sample with a solvent (e.g. diethyl ether, petroleum ether). Fat content is measured by weight loss of the sample or by the weight of fat removed.

4.1.1 Traditional Methods

The traditional solvent extraction techniques can be classified as continuous, semicontinuous, or discontinuous. The Goldfish method is continuous, with the solvent from a boiling flask continuously flowing over the sample held in a ceramic thimble. The Soxhlet method is semicontinuous, with solvent from the boiling flask being volatilized and condensed to build up in the extraction chamber that holds the sample. The sample becomes completely soaked with solvent before the solvent siphons back to the boiling flask. The Mojonnier method (similar to the Roese–Gottlieb method) is discontinuous, uses a combination of solvents, and does not require prior removal of moisture from the sample. In this method, the sample is treated in series with ammonia solution, 95% ethanol, diethyl ether, and petroleum ether. The solvent-soluble material is poured from the Mojonnier flask into a drying dish. After a second and third series of extractions are performed, the combined solvent-soluble material is dried to evaporate the solvent, leaving the fat to be determined by weight.

4.1.2 Rapid Methods

Lipid solvent extraction methods utilizing microwave energy, high pressure, or high temperature have been developed to reduce the time required for analysis and reduce or eliminate solvent usage. With these methods, the initial investment in instrumentation is considerable compared with traditional methods, but there are great savings in time and solvent usage.

In microwave-assisted solvent extraction, the sample is dried in a microwave oven, extracted with mechanical agitation (e.g. 6 min), and redried in the microwave oven to evaporate the solvent. In supercritical fluid extraction (SFE), a fluid (usually CO_2) is brought to a specific pressure–temperature combination. This allows it to attain supercritical solvent properties to extract fat selectively from the sample. The fat dissolved in the supercritical fluid is precipitated by decreasing the pressure of the solution, so the precipitated fat can

be dried and weighed. In accelerated solvent extraction, the fat is extracted when the sample is exposed to a nonpolar solvent in a static (i.e. no outflow of solvent) or dynamic (i.e. fresh solvent continuously flows through the sample) mode, under elevated temperature and pressure conditions. The solvent is then evaporated from the extracted fat, and the fat is dried and weighed.

4.2 Wet Extraction Without Organic Solvent

The Babcock and Gerber methods were developed for lipid analysis of dairy products, and utilize acid treatment rather than the use of an organic solvent. In the Babcock method, the sulfuric acid added to the sample in a Babcock bottle digests protein, generates heat, and releases the fat. With centrifugation and hot water addition, fat rises to the graduated portion of the test bottle so it can be quantitated by volume. The Gerber method is similar, but uses both sulfuric acid and amyl alcohol.

4.3 Instrumental Methods

As described above, IR spectroscopy can be used to predict the lipid content of foods. Mid-IR spectroscopy is used in IR milk analyzers to determine milk fat content. Near-IR spectroscopy has been used to measure the fat content of meats, cereals, and oilseeds, and is being adapted for in-line measurements. Other rapid instrumental methods for fat analysis are based on acoustic properties of the fat (i.e. ultrasonic method), X-ray absorption of fat, the dielectric constant of the food as affected by lipid content, specific gravity, and light scattering caused by fat globules.

5 PROTEIN ANALYSIS

Protein analysis is important for nutrition labeling, assessing nutritive value, and during isolation and purification of a protein. Protein content methods based on nitrogen content are typical official methods of analysis, but wet chemistry methods, based on other principles, and rapid instrumental methods are used for other applications.[18–20]

5.1 Nitrogen Content Methods

The Kjeldahl and Dumas methods are both based on determining the nitrogen content of the sample and using that value to calculate protein content. In the Kjeldahl method, proteins and other organic food components are digested with sulfuric acid in the presence of catalysts, with organic nitrogen being converted to ammonium sulfate. The digested sample is neutralized with alkali and distilled into a boric acid solution. The boric anions formed are titrated with standardized acid to a specific pH or color. The volume of standardized acid used in the titration is used to calculate the nitrogen content of the sample, and that value is converted to crude protein content. In the Dumas, or nitrogen combustion, method, the sample is combusted at high temperatures (700–800 °C). The nitrogen released is measured by gas chromatography using a thermal conductivity detector. The nitrogen value is converted to protein content. For both the Kjeldahl and Dumas methods, nitrogen to protein conversion factors are appropriate for certain specific foods, but the general factor used is 6.25 ($100/16 = 6.25$, since most proteins contain 16% nitrogen).

Both the Kjeldahl and Dumas methods can be applied to all types of foods; total organic nitrogen is measured and not just protein nitrogen. The Kjeldahl method has been an official method for crude protein content for many years, and the Dumas method is increasingly being approved as such. Unlike the Dumas method, the Kjeldahl method uses corrosive reagents and is time-consuming. The Dumas method utilizes an expensive instrument, but results are available within minutes and automated instruments can analyze over 100 samples without attention.

5.2 Infrared Spectroscopy

Whereas mid-IR spectroscopy is used widely to determine the protein content of milk, near-IR spectroscopy is applicable to a wide range of food products. Wavelengths of IR radiation characteristic of the peptide bond are used to estimate the protein content, by the procedure for IR spectroscopy described above. Although such instruments are expensive and must be calibrated properly, samples can be analyzed rapidly.

5.3 Other Methods

Numerous wet chemistry methods to measure protein content rely on various principles: peptide bond, aromatic amino acids, ultraviolet absorptivity of certain amino acids, free amino groups, light scattering properties, and dye-binding capacity. Many of these methods are most applicable to the research laboratory. For example, measuring absorbance at 280 nm, which is based on the presence of tyrosine and tryptophan, is a qualitative, nondestructive method applied to protein samples as they elute from a chromatographic column. The bicinchoninic acid (BCA) method is a quantitative, sensitive, destructive method based on the fact that proteins reduce cupric ions to cuprous ions under alkaline conditions.

The cuprous ion complexes with the BCA reagent to form a purplish color that is proportional to the protein concentration.

6 CARBOHYDRATE ANALYSIS

Total carbohydrate content is of interest for nutrition labeling purposes and for samples in a research laboratory. However, of more use in developing food products and ensuring their quality may be the content of specific carbohydrates or groups of carbohydrates. The content in foods of specific mono- and oligosaccharides may be determined by enzymatic assays or high-performance liquid chromatography (HPLC). Starch and total dietary fiber contents of food are determined by methods that utilize enzymes.[21–25]

6.1 Total Carbohydrate

Total carbohydrate content for nutrition labeling purposes is calculated by difference after determining all other components of proximate composition (Equation 1):

$$\text{total carbohydrate}(\%) = 100\% - (\%\ \text{moisture} + \%\ \text{ash} + \%\ \text{lipid} + \%\ \text{protein}) \quad (1)$$

However, a phenol–sulfuric acid method is used commonly in research laboratories to determine the total carbohydrate content of samples. In this method, carbohydrates are heated in the presence of concentrated sulfuric acid to produce furan derivatives. These derivatives condense with phenol to form compounds that can be quantitated spectrophotometrically. Absorbance values are related to a calibration curve developed using a single sugar.

6.2 Mono- and Oligosaccharides

The food content of carbohydrates such as glucose, fructose, sucrose, lactose, maltose, raffinose, or stachyose may be of interest. Samples may require extraction before analysis of total reducing sugars or individual sugars by HPLC or enzymatic methods.

6.2.1 Extraction

Extraction is used to remove interfering substances from samples prior to analysis for mono- and oligosaccharides. The most common procedure used is extraction with hot 80% ethanol (final concentration), in which mono- and oligosaccharides are soluble but polysaccharides and proteins are insoluble. Calcium carbonate is added to neutralize any acidity, so sucrose is not hydrolyzed to glucose and fructose. The 80% ethanol extract will contain components other than carbohydrates, which can be removed by ion-exchange techniques if necessary. The aqueous alcohol of the ethanol extract can be removed under reduced pressure at low temperatures using a rotary evaporator.

6.2.2 Reducing Sugars

Reducing sugars are those mono- and oligosaccharides that contain a free carbonyl (aldehyde or ketone) group, so can act as reducing agents. These sugars are of interest because they can react with proteins to cause Maillard browning in foods. The Somogyi–Nelson, Munson–Walker, and Lane–Eynon methods all can be used to measure reducing sugars, and are based on the reduction of cupric ions in alkaline solution to cuprous ions by reducing sugars. In the Somogyi–Nelson method, which is the most often used method, the amount of cuprous ions generated in the reaction with reducing sugars is quantitated in a second oxidation–reduction reaction that produces a colored complex that can be measured spectrophotometrically. Glucose or another specific reducing sugar is used to generate a calibration curve for comparison.

6.2.3 High-performance Liquid Chromatography

HPLC is the most commonly used chromatographic method to quantitate specific mono- and oligosaccharides. Stationary phases most often used for HPLC are anion-exchange, cation-exchange, normal-phase, and reversed-phase materials. Detectors most commonly used for such analyses are RI, electrochemical (pulsed amperometric), and postcolumn derivatization with absorbance or fluorescence detection.

6.2.4 Enzymatic Assays

Enzymatic methods have been developed to quantitate specific monosaccharides (e.g. fructose, glucose, galactose), oligosaccharides (e.g. lactose, maltose, sucrose), and polysaccharides (amylose, amylopectin, β-glucan, inulin, starch). For many of these, commercial kits are available that contain appropriate enzymes, other required reagents, buffers, and detailed instructions. In these enzymatic assays, enzyme concentration, substrate concentration, concentration of other required reagents, pH, and temperature must all be controlled because they will affect the reaction rates and results. Enzymatic methods usually are very specific for the substance being measured and have low detection limits. However, care must be taken to remove interfering compounds present. The enzymatic methods are useful to quantitate individual specific carbohydrates, but not for measuring total carbohydrates, a class of carbohydrates, or all carbohydrates individually.

6.2.5 Physical Methods

As described above for moisture/solids analysis, specific gravity, refractometry, and near-IR spectroscopy can be used to measure carbohydrate content of some types of foods. Polarimetry is another physical method applicable to carbohydrates. Carbohydrates, like most other compounds that contain one or more chiral carbon atoms, will rotate the plane of polarization of polarized light. The extent to which a compound in solution will rotate the plane of polarized light can be measured with a polarimeter. When the nature of the compound, temperature, and wavelength of light are kept constant, the concentration of the compound can be determined by measuring optical rotation. For example, the process of hydrolyzing sucrose into D-glucose and D-fructose can be monitored by determining the specific optical rotation before and after sucrose hydrolysis. The methods of polarimetry, refractometry, and specific gravity are applicable only for liquids, and are accurate only for solutions of pure sucrose or other sugars.

6.3 Starch

The starch content of a food is best determined by a combination of enzymes in sequential enzyme-catalyzed reactions. Starch is converted totally to D-glucose by enzymes specific for starch (amylases), then D-glucose is quantitated enzymatically (e.g. combination of glucose oxidase and peroxidase). Two cautions are that the amylases must be pure, and the assay may not be quantitative for high-amylose starch or samples high in starch resistant to digestion.

6.4 Dietary Fiber

Dietary fiber constituents are cellulose, lignin, hemicellulose, pectins, and hydrocolloids. Some of these constituents are soluble in water and some are insoluble. These constituents in total are measured by the official method for total dietary fiber, which is an enzymatic–gravimetric method. A dried, defatted sample is enzymatically digested with α-amylase, amyloglucosidase, and protease to remove the starch and protein. The soluble fiber components are precipitated by adding ethanol. Soluble and insoluble fiber components are collected as the residue upon filtration. The weight of the dried residue, corrected for ash and protein contents (determined by dry ashing and the Kjeldahl method, respectively), is used to calculate the dietary fiber content.

ACKNOWLEDGMENTS

The author thanks the authors of the chapters on sampling, and on moisture, ash, lipid, protein, carbohydrate, and fiber analysis from the book she edited. These chapters were invaluable in writing this article.

ABBREVIATIONS AND ACRONYMS

BCA	Bicinchoninic Acid
HPLC	High-performance Liquid Chromatography
IR	Infrared
KFR	Karl Fischer Reagent
QC	Quality Control
RI	Refractive Index
SFE	Supercritical Fluid Extraction

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Dietary Fiber Analysis as Non-starch Polysaccharides • Enzyme Analysis and Bioassays in Food Analysis • Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Lipid Analyses in Food • Liquid Chromatography in Food Analysis • Near-infrared Spectroscopy in Food Analysis • Proteins, Peptides, and Amino Acids Analysis in Food • Sample Preparation for Food Analysis, General • Starch Analysis in Food • Water Determination in Food

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

REFERENCES

1. S.S. Nielsen, 'Introduction to Food Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 1, 3–13, 1998.
2. Y. Pomeranz, C.E. Meloan, 'Sampling', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 2, 16–25, 1994.
3. J.-L. Multon, W.J. Stadleman, B.A. Watkins, *Analysis and Control Methods for Foods and Agricultural Products,*

Vol. 4: Analysis of Food Constituents, John Wiley & Sons, New York, 1997.
4. AOAC International, *Official Methods of Analysis*, 16th edition, AOAC International, Gaithersburg, MD, 1995.
5. AACC, *Approved Methods of Analysis*, 9th edition, American Association of Cereal Chemists, St. Paul, MN, 1995.
6. AOCS, *Official Methods and Recommended Practices*, 4th edition, American Oil Chemists' Society, Champaign, IL, 1996.
7. R.T. Marshall (ed.), *Standard Methods for the Examination of Dairy Products*, 16th edition, American Public Health Association, Washington, DC, 1992.
8. Y. Pomeranz, C.E. Meloan, 'General Remarks', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 33, 567–574, 1994.
9. A. Proctor, J.-F. Meullenet, 'Sampling and Sample Preparation', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 5, 71–82, 1998.
10. Y. Pomeranz, C.E. Meloan, 'Preparation of Samples', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 3, 26–36, 1994.
11. R.L. Bradley, Jr, 'Moisture and Total Solids Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 8, 119–139, 1998.
12. Y. Pomeranz, C.E. Meloan, 'Determination of Moisture', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 34, 575–601, 1994.
13. J. Giese, 'In-line Sensors for Food Processing', *Food Technol.*, **47**(5), 87–95 (1993).
14. L.H. Harbers, 'Ash Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 9, 141–149, 1998.
15. Y. Pomeranz, C.E. Meloan, 'Ash and Minerals', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 35, 602–624, 1994.
16. D.B. Min, D.F. Steenson, 'Crude Fat Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 13, 201–215, 1998.
17. Y. Pomeranz, C.E. Meloan, 'Lipids', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 37, 678–732, 1994.
18. S.K.C. Chang, 'Protein Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 15, 201–215, 1998.
19. R.Y. Yada, R.L. Jackman, J.L. Smith, A.G. Marangoni, 'Analysis: Quantitation and Physical Characterization', in *Food Proteins. Properties and Characterization*, eds. S. Nakai, H.W. Modler, VCH, New York, Ch. 7, 333–403, 1996.
20. Y. Pomeranz, C.E. Meloan, 'Nitrogenous Compounds', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 38, 733–757, 1994.
21. J.N. BeMiller, N.H. Low, 'Carbohydrate Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 11, 167–187, 1998.
22. R.L. Whistler, J.N. BeMiller, *Carbohydrate Chemistry for Food Scientists*, Eagen Press, St. Paul, MN, 1997.
23. M.R. Bennink, 'Fiber Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, Ch. 12, 189–199, 1998.
24. S.C. Lee, L. Prosky, 'International Survey on Dietary Fiber: Definition, Analysis, and Reference Materials', *J. AOAC Int.*, **78**, 22–36 (1995).
25. Y. Pomeranz, C.E. Meloan, 'Carbohydrates', in *Food Analysis Theory and Practice*, 3rd edition, Chapman & Hall, New York, Ch. 36, 625–677, 1994.

Sample Preparation Analytical Techniques for Food

Michael Rothaupt
Givaudan Roure Research Ltd, Dübendorf, Switzerland

1 Introduction 1
2 Hyphenation to Gas Chromatography 1
2.1 Thermal Desorption 1
2.2 Thermal Extraction 3
2.3 Supercritical Fluid Extraction 3
2.4 Supercritical Fluid Extraction with a Modifier 3
2.5 Solid-phase Microextraction/Headspace Gas Chromatography 3
2.6 Accelerated Solvent Extraction 4
3 Conclusions 4
Abbreviations and Acronyms 4
Related Articles 4
References 4

Sample preparation is one of the important steps in flavor and food analysis. Thermal extraction (TE) of volatile substances simplifies sample preparation by separating volatile compounds from nonvolatile ones. A temperature vaporizing program allows additional prefractionation of the analytes. Solvent venting (SV) supports chromatography by removing a large amount of the solvent. SV also helps eliminate interfering and problematic ingredients from the sample. Sample preparation by supercritical fluid extraction (SFE) is compared to TE, with specific advantages and disadvantages being discussed.

1 INTRODUCTION

Sample preparation is one of the important steps in flavor and food analysis. Solvent extraction is still considered to be a reliable and well-understood method of sample preparation. New techniques are being developed for specialized needs, which exhibit superior performance for specific cases. However, as general approaches to extraction problems, few of these techniques can be considered equal or superior to the classical methods.

Thermal desorption/extraction (TDS/TE) of volatile substances simplifies sample preparation by separating volatile from nonvolatile compounds. A specific temperature vaporizing program allows additional prefractionation of the analytes. SV supports chromatography by removing a large amount of solvent. In this way, interfering and problematic ingredients in the sample can also be removed, while the analytes of interest are absorbed onto a specific stationary phase placed in a cryotrap. Sample preparation by SFE is discussed with respect to the specific advantages and disadvantages of several other techniques. Solid-phase micro-extraction (SPME) and accelerated solvent extraction (ASE) are also covered. Unambiguous control of the sample determination is carried out by mass spectrometry (MS). The coupling of high-performance liquid chromatography (HPLC) with MS is an established standard technique. Capillary electrophoresis (CE) coupled with MS is a relatively new development, and CE has yet to establish itself as a routine analytical method. However, current developments in this technique have demonstrated its potential use in the near future.

2 HYPHENATION TO GAS CHROMATOGRAPHY

2.1 Thermal Desorption

Thermal desorption (TDS) is the extraction of compounds by their vapor pressure from a nonvolatile, complex matrix. The sample is placed into a specifically designed glass tube which is inserted into a system with temperature control from $-150\,^{\circ}C$ to $+250\,^{\circ}C$. The carrier gas flow over the sample can be controlled over the same temperature range. The cooled GC liner can also be filled with adsorbent material in order to trap the analytes quantitatively. This additional focusing mechanism is known as the cooled injection system (CIS). For TDS, the influence of temperature and carrier gas flow are the most important parameters. The flow can be set to such high rates that compounds of low volatility, i.e. having very low vapour pressures, can be transferred to the injector. Higher temperatures can be used when thermally sensitive substances are present. The temperature can be set so that thermally sensitive compounds are desorbed at low temperatures. Then the TDS process can be set to a higher temperature so that desorption occurs of molecules with lower vapour pressures. Desorption of polar compounds with high boiling points is possible. For example the desorption of glycerol (having a boiling point of $290\,^{\circ}C$), takes place with a high carrier gas flow (Figure 1).

An additional SV feature reduces chromatographic interference by the solvent; the analytes are trapped in the CIS. The split valve is open in SV mode. The thermally extracted volatiles enter the liner and the analytes are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

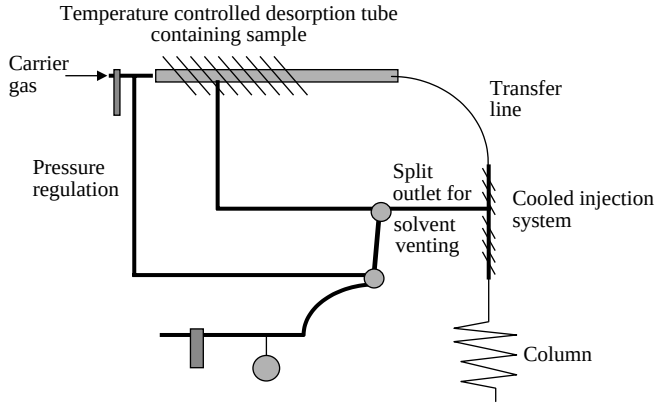

Figure 1 Principle of TDS.

trapped in the CIS at a very low temperature, while the solvent is vented through the split outlet. The venting of the solvent is dependent on the temperature and the carrier gas flow. Too high flow rates will result in loss of analytes. If the trapping temperature cannot be set low enough, it is possible to trap the analytes by placing an adsorbent into the liner (e.g. Tenax®, Porapak®). This option also enables the use of higher (e.g. ambient) temperatures.

When the desorption process is complete, the injection is carried out by heating up the liner to transfer the sample to the analytical column. Mild desorption conditions can be achieved by heating in a suitable temperature gradient. The desorption is also very mild when other compounds are present that can form azeotropes with the analytes. At less steep temperature gradients, peak broadening has to be taken into consideration when choosing the cryotrap parameters.

The analysis of fragrance in an air freshener stick which contained a wet soapy matrix was analyzed without any sample preparation by TDS. If carried out by classical extraction, this analysis would require extensive sample clean-up. Sample preparation consists of weighing the sample and placing it in the TDS tube. The weighing should take place quickly because recovery rates are influenced by the exposure time to air. This has a stronger influence on the recoveries of highly volatile compounds than loss in the CIS. The recovery rate has to be determined for each compound. Quantitation is best achieved by adulteration of the sample with a quantity of the analyte comparable to the amount present in the sample. External standard quantitations give reliable results only when the standard is applied to the same matrix. This is because sample recovery is strongly dependent on the polarity of the analyte and the matrix. Figure 2 shows the result of the analysis of an air freshener.

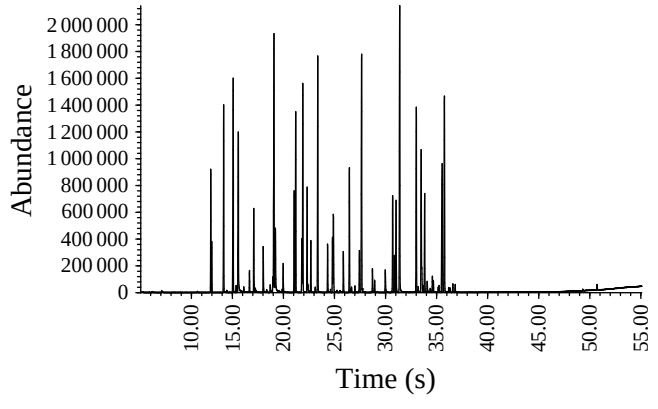

Figure 2 TDS of an air-freshener stick which contained a wet, soapy matrix.

The fragrance of a shampoo was injected to determine the recovery of all fragrance compounds from the matrix. It was observed that most compounds were extracted quantitatively, but a few compounds were hardly recovered. This may have several reasons: too long venting time, too high trap temperature, or incompatibility between the adsorbent material and these molecules. Incompatibility here means that the compounds are not retained in the trap on the adsorbent and are lost through the solvent split, or that they are not released from the adsorbent because a strong binding interaction takes place between the compounds and the trapping material. These observations make it necessary to carry out a thorough quantitation, as described above.

Analysis of autoxidation products in moistening creams with a considerable amount of water can be carried out by placing the sample into the TDS tube and thermally desorbing using SV to reduce the problem of interfering compounds, such as water. A temperature of 80 °C is low enough not to also desorb the fat, and only the analytes are transferred to the injection system. The low content of these analytes can be overcome by using a large volume of oil in order to obtain enough analyte to be determined.

A difficult analysis using TDS alone is when the sample contains a considerable amount of water. In this case it is necessary to use an off-line TE instrument.

The analysis of fresh ginger roots by TDS is an example of the moderate conditions that are applied during sample extraction by TDS. Conventional sample preparation for the analysis of essential oils is carried out by steam distillation. This exposes the analytes to high temperatures for quite a long time in the distillation vessel, with the potential risk of degradation. TE can be used to remove most of the water content in a short period of time. This is necessary because the analysis is carried out on a polar (wax) column. The desorption is so mild that thermally sensitive compounds can be detected in much higher concentrations than expected from the literature. Figure 3 shows the analysis of fresh ginger roots.

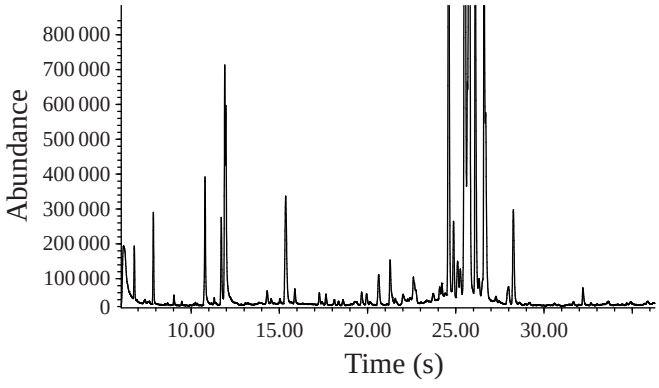

Figure 3 TDS of fresh ginger roots.

TDS does not require knowledge of the physicochemical properties of the analyte, and in most cases leads to good results more quickly. For individual compounds the extraction efficiency and recovery rate of TDS has to be determined by spiking the sample with the compound of interest. The amount of spiking material must be in the same range as in the sample.

TDS is a new and powerful sample preparation technique. The appropriate approach to sample preparation depends on the compounds that need to be quantified, and on the sample matrix. Used correctly, a wide range of compounds can be determined in a reliable and reproducible way.

2.2 Thermal Extraction

TE is an off-line TDS with a TDS tube as focusing device. The focusing is carried out by an appropriate absorption material. This device is used when the original sample contains considerable amounts of water. Otherwise, the TDS procedure can result in severe losses of analytes when the temperature is not low enough. Also, when the temperature of the cryotrap is below 0 °C, the water in the sample would freeze immediately and block the system. In principle it is possible to extract pesticides from an aqueous sample by this technique (Figure 4).

2.3 Supercritical Fluid Extraction

SFE as an alternative extraction method without solvent consumption is comparable to TDS. This technique is particularly useful in the extraction of hydrophobic compounds. Even low-volatile compounds are extractable under very mild conditions. The recoveries are comparable to the results obtained with TDS for nonpolar compounds. For polar compounds an additional modifier is added to the supercritical fluid. The extraction efficiency of SFE and TDS is comparable for polar compounds. Usually the same solvent is used as applied for the rinse step from the trap. Extracts obtained by SFE are compatible with several separation techniques, such as gas chromatography, HPLC, or capillary electrochromatography systems. However, TE is designed specifically for use with gas chromatography.

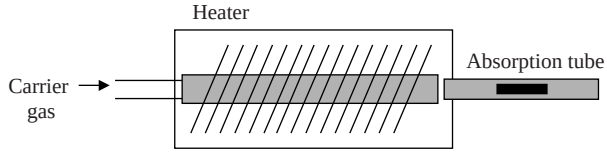

Figure 4 Principle of TE.

The extraction of hydrophobic compounds from complex matrices containing sugar, proteins, and fat can be achieved almost quantitatively, but polar molecules give poor recovery rates. The recovery of these compounds can be improved significantly by the addition of a modifying solvent.

Changes in the density of supercritical fluids influence the extraction of nonpolar compounds of different molecular weight. These compounds can be fractionated in a short period of time by the correct choice of density.

2.4 Supercritical Fluid Extraction with a Modifier

The addition of a modifying eluent such as acetonitrile or methanol enables other compounds to be extracted by SFE. Thus selection by polarity provides an additional fractionation parameter.

There are two possibilities when adding a modifier: static addition of modifier directly into the extraction thimble, and addition of the solvent by a modifier pump. The pump delivers a constant flow rate and improves extraction efficiency and reproducibility. Modification can also mean addition of an adsorbent to the sample to retain specific substances that are problematic to the analysis.

For example, at low densities, the extraction of flavors out of a fatty matrix can be achieved by adding cellulose powder to the sample. Supercritical carbon dioxide then extracts the flavoring compounds without any of the problematic sample matrix compounds, such as fatty acids or triacylglycerols. After SFE extraction, the analysis can be carried out by gas chromatography, HPLC, or CE. The recovery of terpenes from a hydrophobic matrix by SFE can be as high as 80%. For more polar compounds the addition of a modifier is essential for SFE.

2.5 Solid-phase Microextraction/Headspace Gas Chromatography

SPME represents a special technique for the absorption for volatile compounds. The technique is very specific and has to comply with the physical and chemical properties of the analytes. If the properties of the analytes are

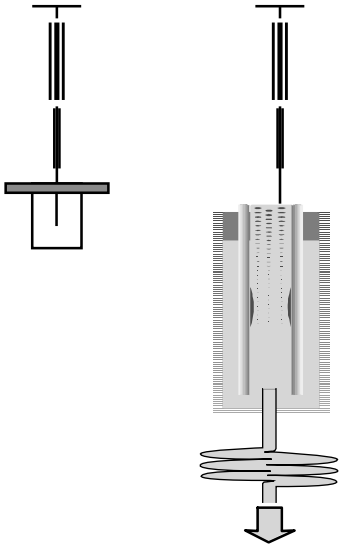

Figure 5 Principle of SPME.

known the absorption material can be chosen correctly and the analysis shows both high selectivity and high sensitivity. The principle of the technique is shown in Figure 5. The absorbent material may be located in the headspace above the sample or can be immersed in the solvent containing the analyte. The latter variation is more critical because the absorption material can become saturated by the solvent, thus changing the material's properties.

2.6 Accelerated Solvent Extraction

ASE is similar to SFE. The difference is that organic solvents are used instead of supercritical fluids. The use of temperature and pressure as variables, together with a choice of organic solvent, extends this extraction method to include polar materials, which are usually covered by conventional extraction methods with polar solvents. Compared to liquid–liquid extraction, this technique shows a higher efficiency (i.e. recovery rate) in a shorter time. With an organic solvent such as acetonitrile, the extraction of polar metal ion complexes from soap was almost as efficient as extraction by SFE with a modifier. Compared to Soxhlet extraction or continuous liquid–liquid extraction, the time required was reduced 10-fold. This technique is most effective when the physicochemical properties of the molecules and the matrix are known, thus requiring a detailed understanding of the underlying chromatography.

3 CONCLUSIONS

The extraction methods described in this article extend the range of applicability to fast and specific extraction efficiency for known compounds. These new approaches deal with very specific conditions in which the matrix and the properties of the compounds influence extractability and recovery. Data evaluation can only be considered reliable when a quantitation of the analytes has been carried out according to accepted guidelines.

ABBREVIATIONS AND ACRONYMS

ASE	Accelerated Solvent Extraction
CE	Capillary Electrophoresis
CIS	Cooled Injection System
HPLC	High-performance Liquid Chromatography
MS	Mass Spectrometry
SFE	Supercritical Fluid Extraction
SPME	Solid-phase Micro-extraction
SV	Solvent Venting
TDS	Thermal Desorption
TDS/TE	Thermal Desorption/Extraction
TE	Thermal Extraction

RELATED ARTICLES

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

REFERENCES

1. G. Antinescu, C. Doneanu, V. Radulescu, *Flavour Fragrance J.*, **12**, 173 (1997).
2. A.J. Borgerding, C.W. Wilkerson, *Anal. Chem.*, **68**, 2874 (1996).
3. W.M. Coleman, B.M. Lawrence, *Flavour Fragrance J.*, **12**, 1 (1997).
4. K. Grob, *Anal. Chem.*, **66**(20), 1009–1018 (1994).
5. H.G.-J. Mol, M. Althuizen, H.-G. Janssen, C.A. Cramers, *J. High Resol. Chromatogr.*, **19**, 69 (1996).
6. E. Reverchon, G. Della Porta, D. Gorgoglione, *Flavour Fragrance J.*, **12**, 37 (1997).
7. H. Soerensen, S. Soerensen, C. Bjergegard, S. Michaelsen, *RSC Food Analysis Monographs*, Royal Society of Chemistry, 69–79, 1999.

Sample Preparation for Food Analysis, General

Michael J. Lichon
University of Tasmania, Hobart, Australia

1 Introduction 1
2 Sampling for Analysis 2
2.1 Aims of Sampling 2
2.2 Nature of Matrix and Analyte 3
3 Homogenization of Samples 5
3.1 Reduction of Particle Size 6
3.2 Mixing of Powders 7
3.3 Evaluation of Homogenization 8
4 Sample Preparation 9
4.1 Separation of Portions 9
4.2 Preparing the Samples 10
4.3 Storage of Samples 10
4.4 Weighing and Dilution 10
4.5 Digestion, Extraction, Cleanup Separation and Derivatization 10
4.6 Deterioration and Contamination 11
Abbreviations and Acronyms 12
Related Articles 12
References 13

Sample preparation is the obtaining and manipulation of representative portions of a whole prior to subjecting it to any particular analysis. This includes all steps from design and implementation of sampling from the whole, bulk matrix, or population, in such a manner as to represent the whole (or some determinable aspect of the whole), reduction of particle size, mixing, drying, weighing, dilution, and any number of physical or chemical manipulations prior to analysis. These may include apportionment, storage, digestion, extraction, cleanup separation, and derivatization. While the purpose of these essential steps is to enhance precision and accuracy in determining an analyte concentration relating to the whole, each step adds inherent errors to the result. Furthermore, they may contribute to the contamination and deterioration of the sample relative to the original whole. The consistent use of observation, documenting, replication of test portions, recoveries and reference materials should highlight problem areas. These may be reduced, or at least quantified, using the skill of the professional analyst.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

With few exceptions, matrices containing analytes are complex, inhomogeneous mixtures of a large range of chemical substances which may include intermediate and side reaction products. The isolation and measurement of the individual chemical compounds of interest usually represent a difficult challenge, particularly with modern requirements for environmental management and food standards. Even with powerful modern techniques of separation and identification, such as high-performance liquid chromatography (HPLC), rarely is it possible to load a syringe directly with an intact matrix and inject to obtain a sensible result. Perhaps surprisingly, it is not unusual to find analytical methods published with precision data reflecting repeated direct assays of standard solutions. This tells the reader little about the practicality of the methods with regard to real-world samples. Analytical errors usually cost money: wrong decisions based on erroneous results, lost business, false security, litigation, etc. The identification and minimization of sources of error, of which sampling and sample preparation usually feature prominently, need special attention and resources. Procedures for preparation of the sample should be developed, evaluated and published as an integral part of any analytical method and application.

Three basic steps are involved in sample preparation for chemical analysis: (1) sampling, obtaining a sample for the laboratory; (2) homogenization of the laboratory sample to permit the taking of test portions; and (3) sample preparation, physical and chemical manipulation of the test portion prior to analytical measurement. It should be appreciated that elements of these steps may occasionally occur in the reverse order or as combined operations. The fourth and final step of the analysis is the actual determinative assay procedure. Paradoxically, although the purpose of each of the three steps is to increase the accuracy and precision of the analysis so the test portion reflects the composition of the bulk, each step also introduces inherent errors. The error contributions of these steps for a typical analysis scheme are shown in Figure 1. Analyte concentration is limited at one end by detection limit and at the other by overloading of preparation stages or the measurement instrument by either analyte or matrix. The significance of the contributions of these steps to the total error for the analysis are mathematically described by the relative standard deviation (RSD) relation (Equation 1):

$$\mathrm{RSD}_{\mathrm{total}} = (\mathrm{RSD}^2_{\mathrm{sampling}} + \mathrm{RSD}^2_{\mathrm{homogenization}} + \mathrm{RSD}^2_{\mathrm{sample\ preparation}} + \mathrm{RSD}^2_{\mathrm{analytical}})^{1/2} \quad (1)$$

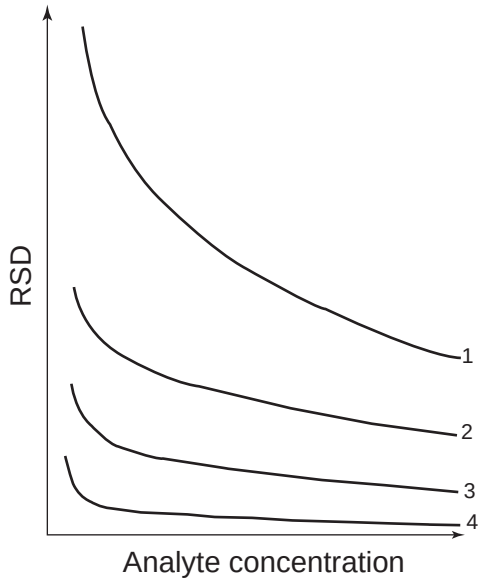

Figure 1 Typical RSD (relative standard deviation) of error components for an inhomogeneous food matrix. 1 = sampling; 2 = homogenization; 3 = sample preparation; 4 = assay procedure. (Reproduced from Lichon[1] by courtesy of Marcel Dekker, Inc.)

This equation shows the errors to be additive. Either a greater number of preparative steps or larger errors in a smaller number of steps may create a worse overall error. Furthermore, the errors of each step are not only cumulative, but also usually strictly irreversible. Any error generated, by sampling (for example), can usually not be compensated for by any subsequent treatment.

From Equation (1), it is clear that if any one of the contributing factors is significantly greater than the others, it is futile to attempt to reduce any of the other contributors, as the total error will be disproportionately dictated by the most dominant factor being squared. A good example is aflatoxin analysis of bulk peanuts, where as long as the sampling error contributes 90% of the error there is little incentive to improve the analytical precision[2] beyond academic interest.

It may often be clear to an experienced analyst what the approximate proportions of the contributing errors are of the total for a familiar analysis. If this is not so, then these may be defined by rigorous assays of replicates and recoveries testing the effects of each successive step of the analysis. This review of preparing samples for chemical analysis is divided into three sections to consider problems encountered in each of these respective steps. Examples are drawn from food and environmental analysis from the author's experience of these arguably more challenging fields of analytical chemistry.

2 SAMPLING FOR ANALYSIS

There has been great concern amongst analysts over the validity of analytical methods. Attempts to define rigorously the precision and accuracy of methods include measures such as international collaborative trials. For all the benefits of implementing such expensive measures there is one important oversight: the issue of sampling is not examined. Experience demonstrates that sampling can often be the greatest source of error in chemical analysis, particularly for environmental[3,4] and food[5] matrices (Figure 1). Frequently lack of attention to sampling is the hidden factor behind ''unexpected'' analysis results.

> The classic example of incorrect sampling procedure and its ridiculous consequences is given by the fable of the blind men and the elephant. The consequences are sometimes no less ridiculous for incorrect chemical sampling
>
> W.J. Blaedel and V.W. Meloche[6]

Undergraduate texts often form the basis of analysts' future attitudes. The spectrum of emphasis on sampling ranges from serious but brief mentions[7–9] through a good treatment, but at chapters at the end of the book,[6] to an integral treatment from the beginning.[10] This choice resides with course designers and coordinators.

Unfortunately, in the real world, the benefits of correct sampling come at a cost. However, if one considers the assay of a sample that does not represent the bulk, or ''population'', of interest as near useless, then to obtain meaningful analytical data one must be prepared to pay for it. If the samples, individually and collectively, cannot provide the required information, they are seldom worth the time and expense of analysis. Correct sampling by competent personnel may require an investment of anything up to, or even more than, the cost of the laboratory assays.[11] Planning for informative sampling must be an integral part of any study. Subsequent occasions of analysis of the same situation may cost somewhat less on account of experience.

2.1 Aims of Sampling

The degree of justifiable investment often depends on the purpose of the assay: for assay leading to the determination of a market price of a shipment, it may well be acceptable to spend 1% of the market value of the commodity on reliable sampling. Although this may at first seem a high cost, it may be low when regarded as an insurance policy against larger losses. If a pesticide content of an export shipment of beef is suspected to approach the acceptable threshold, the sampling could

justify a far higher expense than, for example, when routinely examining for other contaminants known to be far below regulatory limits.

The best method of sampling in a given situation will depend on such issues as what information is sought, resources available, accessibility of the target population, what the required turnaround time is, perishability of the matrix and the analyte, what the criteria are for acceptability and sampling correctness, whether samples are to be pooled or replicated, whether analysis should be performed separately on different portions of the sample, whether the surface is to be included in the bulk and the monitoring which should occur to prevent contamination and abuse.

The compilation of food composition data presents specific sampling problems. Typically cited nutrient contents frequently reflect limited localized sampling.[12] The task of nationwide representative sampling is fraught with bias, difficult to organize and resource intensive.

Environmental monitoring may frequently best involve indirect sampling, e.g. of animals, such as trout or molluscs which accumulate heavy metals from the environment. Hair or blood of mine workers may be analyzed to estimate exposure to ore dust as less intrusive than biopsy sampling of vital organs or bone accumulation. Breath or urine samples are commonly used to estimate blood concentrations of drugs. Accumulator plants have been used as indicators for mineral prospecting. Foods such as milk, through which children may be particularly exposed to some pollutants, are often the subject of monitoring programs.[3]

Production sampling may seek to determine a market price based on positive analytical content (such as Cu and Ag in a bulk mineral concentrate) or to price-set or outright select against contaminant content (mycotoxins, pesticides, etc., in foods).

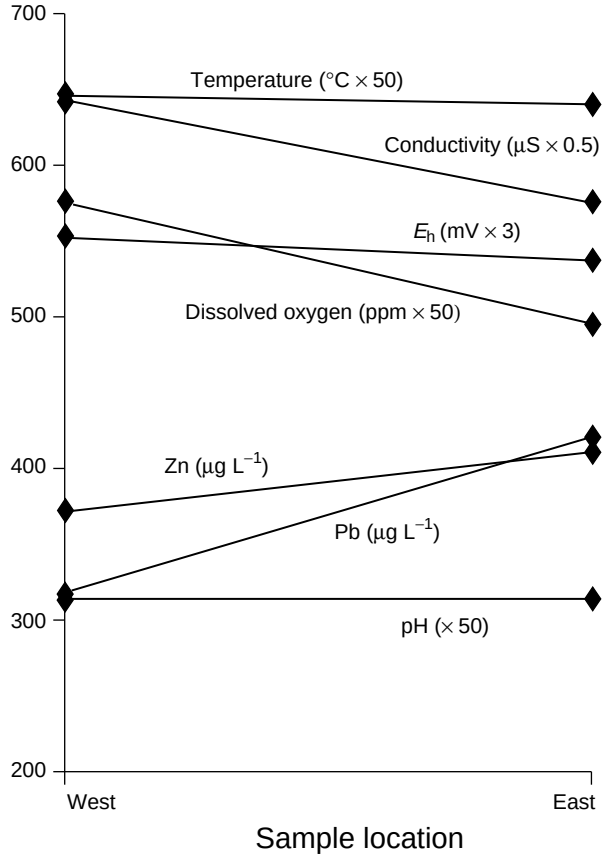

Figure 2 The distribution of analytes across a 20-cm wide outlet channel draining an apparently homogeneous settling pond with a single inlet source. (Reproduced from Lichon[13] by permission of the Australian Water and Wastewater Association.)

2.2 Nature of Matrix and Analyte

Is the matrix/population really heterogeneous? A poignant example of an apparently homogeneous flowing liquid showing unexpected variation is shown in Figure 2. Is spatial variation general, localized or stratified? Should the population be sampled critically or representatively, randomly or systematically? Table 1 offers some consideration of this question. The low bias of random sampling has obvious attractions, and may be desirable to minimize sampler's liability. Systematic grid sampling is particularly useful in measuring changes over time;[13] an example being monitoring soil fertility changes. Targeted sampling may be robust if the background research shows that the analyte of interest is associated with a readily identifiable feature of the matrix and only spot measurements of high or low analyte provide the required data. Is there temporal variation, and over what scale of

Table 1 Styles of sampling: main advantages and drawbacks

Style	Bias	No. of samples required	Sample sites	Associated costs
Random	Low	Large	Random allocation	Analyze large sample numbers
Systematic	Moderate	Moderate	Grid pattern	Establish consistent sampling stations
Targeted	High	Least	Critical judgement based on understanding of analyte distribution	Research work to form basis of understanding

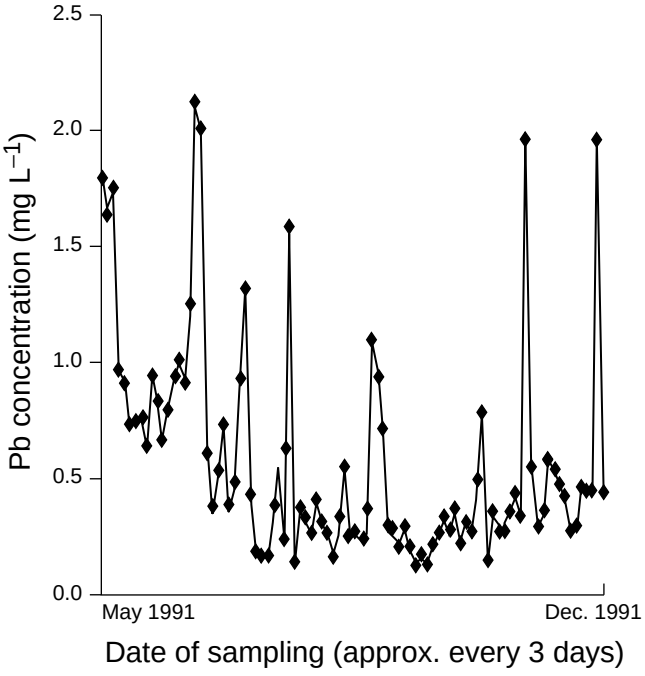

Figure 3 The temporal variation of Pb in a dam outlet with an assumed residence half-life of 3 days. The fluctuations of Pb indicate more rapid transit of channeled flows. (Reproduced from Lichon[13] by permission of the Australian Water and Wastewater Association.)

time? What factors influence the changes? The example in Figure 3 demonstrates unexpected analyte fluctuations in dam overflow samples due to overestimating the optimum sampling regime timescale – some fluctuations may well be missed by the data.

The difficulties of obtaining good representative sampling depend on the nature of the ''population''. Discrete populations are relatively easy to sample. In the case of a shipment of pallets of canned food, the units are discrete and may be mathematically modeled and sampled; the only inconvenience then is the necessity for dismantling some of the pallets to extract the desired target cans. This correct procedure is often bypassed, and cans are taken from the top corner of the most accessible pallets to give a highly biased sampling.

Sampling of continuous populations is by far more common. Most difficult sampling problems usually involve a 3D bulk population: simple sampling is rarely possible.[11] A more reliable approach is to convert the bulk into a one-dimensional flowing stream. The stream is then divided systematically, randomly, or by a combined formula of random delineation of systematic segments, into target segments in time, such that any part of the bulk has an equal probability of scrutiny.[11] Entire cross-section segments (strictly delineated by parallel planes) of the stream are sampled at the target times, then taken either separately or pooled for analysis. In a factory situation, the best sampling points are often conveyor belts, from which segments may be easily extracted. Another method of treatment of a three-dimensional bulk is to convert it into a flattened two-dimensional configuration. The two-dimensional bulk may then be sampled by cylindrical corer, at random or systematic points on the surface, ensuring that the entire thickness of the bulk is recovered. In practice, this method often fails to recover a strictly delineated cylindrical core, or adequately to expose the bottom of the bulk to scrutiny by virtue of the device being unable to retrieve the lowest part of the cylinder. The two-dimensional bulk may be fashioned into a long narrow pile, and cross-sampled in space as a one-dimensional bulk, with randomly or systematically located segments delineated by parallel planes. The classical method of coning and quartering can rarely be done correctly (fundamental problems of delimitation and by unmixing, both spontaneously and when being moved; see section 3.2), and operator bias further fails the process.[11]

It may on some occasions be desirable to analyze unrepresentative target, or extreme portions of a population, as well as a representative sample (Table 1). It may be of interest to find extreme values for vitamins, to find the worst case for environmental contamination, or to link distribution with a causative factor. It may be desirable to select specimens with a view to breeding selection.

To consider some additional variables in real sampling problems, take the example of analysis of peas: sample variables include size distributions, position of individual peas in their pods, height of pods up the individual vine, individual plant genetics, cultivar, time of planting, efficiency of pollination, watering history, soil type and underlying geology, previous crop history, soil fertilization, crop maturity, disease and pestilence attack, length of time and handling and storage conditions since harvest. Some further questions follow: will either the sample deteriorate or the data expire before the sampling, analysis, and reporting are complete?; are the peas to be cooked?; if so, how and for how long?; is the analysis to represent data for a single plot, farm, locality, or national database?; does the sampling need to be certified, for legal or economic reasons?

The importance of sampling is highlighted by problems encountered in analysis of aflatoxins in peanuts. Several workers have specifically addressed this difficult sampling situation.[2,14–17] The highly inhomogeneous distribution of analyte in the matrix leads to the proportion of total analytical error attributable to sampling commonly exceeding 90%. Similar distribution inhomogeneity may present problems in sampling individual chocolate chip cookies, cans of sausages and beans, and croutons in a soup packet. The best strategy may often be to take a very large initial sampling quantity prior to

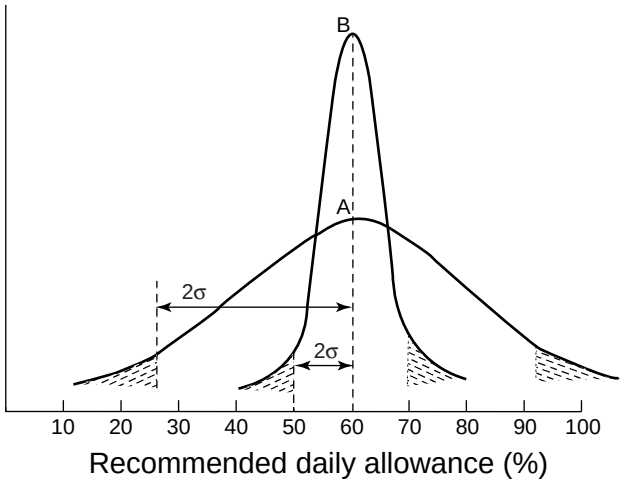

Figure 4 Distribution curves for vitamin A content of canned food with carrot chunks. (A) Single can samples; (B) 12-can pooled samples. (Reprinted from NBS Special Publication 519, 'The Role of Sample Preparation in Nutritional Labeling Analysis', by H.G. Lento,[18] courtesy of the National Institute of Standards and Technology. Not copyrightable in the United States.)

preliminary homogenization. Another approach is to take many replicate samples (20–200) for analysis to show individual unit variability. The various approaches yield different information that may meet different requirements. Pooling of several individual units may give better definition of the true mean of assay results (Figure 4). However, individual unit variation may better reflect single-serving variability for the consumer. The example of β-carotene content in canned stew in Figure 4 is a reflection of statistical sampling of units containing relatively few, large particles of carrots in a small discrete sampling bulk, from a large population. By pooling 12 cans together for assay, the manufacturer is able to meet specifications for vitamin A content using less of the ingredient, simply by the reduction of the RSD.

Often the analyst has little influence over the taking of the sample. Ideally the analyst should appraise the problem and take the samples personally. Failing that, the analyst should brief the sampler as to appropriate methods.[13] An instance of such difficulty concerned an untrained sampler assigned to study traditional indigenous foods of Australian Aboriginals. He was briefed to collect specimens for nutritional analysis, but was apparently little influenced by the analysts. Constraints apart from the lack of training were hot climate, lack of refrigeration, desirability of taking pocket-size samples, and remoteness from the laboratory, some 3000 km and several days freight away. All of these factors contributed to degradation of sample integrity. Many of the samples analyzed consisted, for example, of three individual thawed fruits totalling 15 g. Despite obvious limitations of nutritional assays on such poor samples, the data are nevertheless at least indicative and still useful in this context,[19,20] where previously nothing was known. Investment in an expert sampler can clearly boost the value of analytical data.

3 HOMOGENIZATION OF SAMPLES

The classical wet chemistry methods of analysis held sway for decades. These rarely used test portions less than 1 g, and the measurement precision was variable. Hence the conventional methods of homogenization usually provided adequate test portion sampling. More recently, with the development of analytical instrumentation, detection limits and test portions decreased dramatically, while sensitivity to contaminants and artifacts increased. Simultaneously, as matrices became more diverse, so did the diversity of analytes of interest. The need arose to review existing methods for homogenization and develop new techniques to meet the new demands adequately.[21]

Several workers have recently reconsidered homogenization methods, conventional and otherwise, applied to foods.[18,21–25] For assays using test portions of 1 g or more, several conventional methods prove satisfactory with compatible matrices.[21] Many method–sample category combinations prove to be incompatible, some unexpectedly so. Conventional techniques usually operate by the mechanisms on cutting and shearing for samples of wet consistency, and shearing and impact for dry samples. Weaknesses of these techniques include the reliance on the inertia of the supporting fluid, or surrounding particles, as the ''anvil'' for the grinding action. Examples of failure of these methods include the persistence of corn kernel skins and long meat fibers. For techniques such as hammer milling[21] the serrated body of the grinder serves this purpose. The CO_2 milling technique[21] uses dry ice as inert mass particles, which also serve to embrittle the food particle. The liquid N_2 milling technique[21] similarly uses cryogenic embrittlement, and uses a metal–metal impacting action. For assays using smaller test portions or requiring stringent homogeneity of very heterogeneous foods, the cryogenic treatments prove well worth the extra effort after a conventional pretreatment and freeze-drying. This dual treatment reduces particle sizes to below 60 μm (97% below 10 μm) for one of the most difficult matrices (muesli bars) with sufficient mixing to take reproducible test portions of 1 mg. The average RSD of microcombustion protein assays for a range of foods was 1.33%, performed on test portions of 2–5 mg.[21] The number of particles included in such a 2-mg test portion following N_2 milling is approximately 1×10^5.

The required size of the test portion and the sample's characteristics will dictate the degree and type of

homogenization required. If several different assays are to be performed on a sample, then whichever has the most stringent requirements will often dictate the homogenization requirements. Experience shows that it is often prudent to use two or more homogenization techniques sequentially. It may be desirable to split the sample after an initial wet-basis homogenization treatment, analyzing the first part for labile vitamins directly, and further rigorous homogenizing after freeze-drying before subjecting the second part to other analysis. A generalized example scheme for homogenization of samples for nutritional analysis is shown in Figure 5. This scheme includes approximations for quantities and particle sizes at each stage, and what types of assays are amenable to the products of each stage. However, it must be reiterated that each sample will have different characteristics which will require different homogenization treatments. Table 2 provides a summary of example assays and the consequences of using three example homogenization treatments. The main conclusion that should be drawn is that each sample should be homogenized by methods that have proven effectiveness with the particular matrix, either from experience, from the literature, or by experiment, to a degree that meets the test portion requirements and other facets of the assay procedures. This must be confirmed by the precision of replicate assays, blanks, and recoveries.

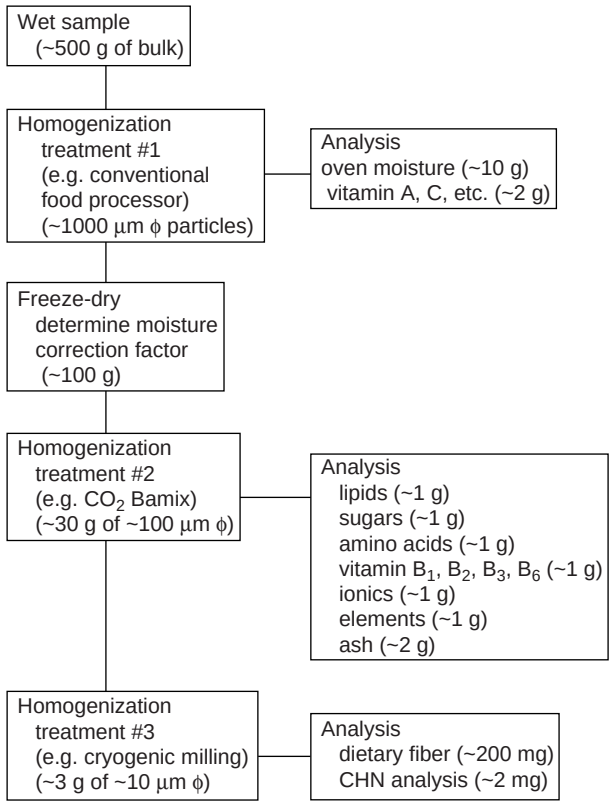

Figure 5 A sample homogenization scheme for nutritional analysis of food. ϕ = diameter. (Reprinted from M.J. Lichon,[26] 'Sample Preparation Methods for Chromatographic Analysis in Food Chemistry', *J. Chromatogr.*, **624**, 3–9 (1992), Copyright 1992, with permission from Elsevier Science.)

3.1 Reduction of Particle Size

There are two functions of homogenization: reduction of particle size and mixing. Reduction of particle size involves cutting, shattering, and shearing. The various homogenization devices achieve these in different ways to differing degrees. This necessitates a judicious choice of homogenization methods that have demonstrated applicability for use on particular matrices. The efficacy of a method may be observed by microscopic examination

Table 2 Consequences of using various homogenization treatments prior to undertaking a selection of food assays

Assay and test portion[a]	Type of homogenization treatment, approx. particle size		
	Conventional 1–2 mm	CO_2 mill 100–200 μm	N_2 mill 10–20 μm
Fat 1–5 g	Poor extraction, OK sampling	Better	Best extraction but possible for emulsion problems
Fiber – AOAC 1 g	Poor sampling, OK digestion	Best	Losses through 90 μm pore filter
Fiber – Englyst 200 mg	Sampling too poor	OK	Best
Moisture 2–10 g	Best	NA – process changes moisture	
Vitamins 1–2 g	Best for labile vitamins	Better sampling (if vitamins stable after grinding treatment)	
Bomb calorimeter 0.5 g	Poor	OK	Preferred
GFAAS direct 1–10 mg	Useless	Poor sampling	Preferred
CHN 2–5 mg	Useless	Poor sampling	Preferred

[a] GFAAS = graphite furnace atomic absorption spectroscopy; CHN = semimicro combustion (carbon, hydrogen, nitrogen) elemental analysis. (Reproduced from Lichon[1] by courtesy of Marcel Dekker, Inc.)
AOAC, Association of Official Analytical Chemists.

or sieving of the product. The importance of particle size reduction is intuitive; the smaller the particles, the greater is the number of particles that will be included in a given test portion, hence the greater is the probability of sampling of more parts of the original bulk. Quantifying this statistical notion is more difficult. A simplified treatment with graphed relationships[27] is recommended reading. The standard deviation (SD) plot in Figure 6 clearly shows the improvement in precision that results from particle size reduction. For reasonable precision in assaying a complex food, the sample should be ground sufficiently such that the smallest test portion should contain at least 500 particles of the least common ingredient.

Apart from the aspect of test portion sampling, particle size may influence other facets of the analytical procedure, where surface area and/or particle radius are critical to reagent exposure and penetration. Hence the degree of particle size reduction may affect both the accuracy and the precision of the analysis (Table 2), in ways that may not be readily revealed by recovery data. A typical case is the accessibility of the food matrix to enzymatic digestion. The efficiency of enzymatic digestion is proportional to the surface area (or degree of homogenization) of the food substrate, that is,

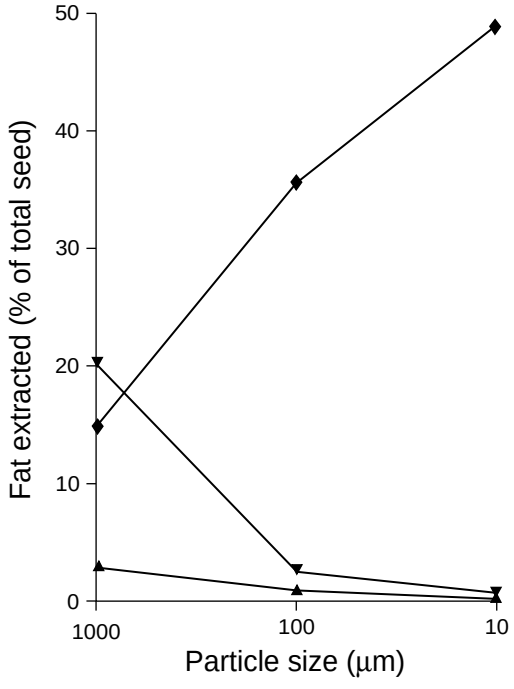

Figure 6 Fat extraction from pumpkin seeds homogenized to three particle sizes by conventional treatment, CO_2 milling, and N_2 milling. Total extractable fat and error (both absolute and RSD) improve dramatically with reduction of particle size. , Fat(%); , Error (SD) (%); , RSD (%). (Reproduced from Lichon[1] by courtesy of Marcel Dekker, Inc.)

inversely proportional to particle size (radius). Large particle size may inhibit enzyme access to the whole of the food, or specific portions may be encapsulated and not be available for subsequent extraction; such problems have been encountered in thiamine,[28] lipid,[29] moisture, and dietary fiber[30] analysis. Efficient thiamine extractions were only possible after digestion of finely homogenized foods. In the second example, digestion of a well homogenized substrate was required for complete release of lipids. Efficient solvent extraction was found to be possible only with small particle radius (Figure 6). However, if some foods were too finely ground (with a very large reactive surface area), problems with emulsion formation hampered the efficiency of the subsequent liquid–liquid extraction. Moisture determinations are similarly affected by large particle size.[11] Large-radius particles often fail to yield water quantitatively during the procedure, as the innermost molecules are trapped by the depth of surrounding material. Moisture–time curves for such samples are shallow and poorly defined, and do not reach a recognizable plateau. In contrast, Englyst dietary fiber determinations[30] on poorly homogenized samples give spuriously high results owing to the incorporation of undigested starch in the fiber fractions. Analysis of total dietary fiber by the AOAC method[31] has different problems: if the sample is too finely ground then there is the risk of low results through losing fiber through the 90-μm porosity filter, even with the use of filter aids; if the sample is too coarse, then there will be high results from insufficient enzymatic digestion of other components, and a loss of precision from poorer test portion sampling of larger particles (Table 2).

3.2 Mixing of Powders

Particle size reduction helps to improve sampling at the test portion by increasing the number of particles sampled per test portion, but to achieve the full gain in accuracy and precision the sample particles must also be thoroughly mixed so the test portion includes a random selection of the different particles. Mixing is not easy – ask anyone who has attempted to blend home-made muesli thoroughly enough to prevent children from segregating the mix to maximize their favourite ingredient content!

Most sampling and homogenization processes for foods lead to powdered samples. Simple observation of powders leads to the definition of two broad categories: free flowing and cohesive.[32] The distinct properties of each lead to important consequences for the analysis of test portions.

Free-flowing powders, such as rice or coarse sands, flow smoothly, have a non-dusty appearance, and tend not to adhere to container walls. In broad terms, particle size for free flow needs to be above 50 μm, given

average surface properties and sphericity of the particles. This cut-off size increases with particle elongation or stickiness. On the face of it, flowable powder samples are attractive for homogenization, as they are easy to mix and handle. However, the flowable properties lead to mixtures unmixing as easily as mixing. As individual particles are free to move, they will move and settle differentially according to properties such as mass, density, shape, angularity, angle of repose, size and surface nature. Therefore, any induced motion will lead to classification and segregation. The muesli example clearly falls into this category. Segregation mechanisms include projection, percolation, and elutriation. Coarse particles, by virtue of momentum, are projected further than finer particles. Smaller, denser particles have a greater ability and mobility to percolate through a loose mass of particles. Percolation may take place when moving down an incline, rolling and tumbling, pouring into a pile, or when subject to vibration. Elutriation is the loss of fines as dust. This commonly occurs when filling a container, where the displacement air blows dust out of the mixture, or when stirring or mixing in an open vessel. The use of jaw crushers and open milling techniques risk differential losses of fines. This loss of fines may be critical to analysis, as the composition of the dusts often differs markedly from that of the coarser particles. For example, dust in peanuts may be disproportionately rich in mycotoxins. The counters to these drawbacks of flowable powders are to reduce particle size by homogenization into a cohesive powder, or to ensure that any mixing is active, not passive. Mixing should involve active displacement of particles rather than allow them to choose their own paths. A rolling, tumbling mixer gives individual particles freedom to unmix and is undesirable, while a paddle mixer pushes and relocates groups of particles and is preferred.

Cohesive powders, such as flour, have an erratic stick-slip flow, are often dusty, and stick to container walls. A smaller particle size ($<50\,\mu m$) by nature enhances the potential for mixing to meet test portion precision requirements. Furthermore, the cohesive structure inhibits segregation. Cohesive powders lack mobility as light individual particles are held in a structure and cannot easily move independently. Hence mixing tends to behave irreversibly, the structure preventing classification. Thorough mixing of a cohesive powder is more difficult owing to poor mobility, is slower and is a more intensive process. Many of the mixing properties of cohesive powders are shared with thick fluids and muds, typical of many wet-basis food samples.

Mixing needs attention in two areas: adequate shearing and the inclusion of all the sample. Shearing is necessary to overcome cohesive forces, to break up and disperse aggregates into the mixture: it is necessary repeatedly to break individual particles free from their structuring neighbors. For some powders gentle rolling and breaking of a tumbling mixer may cause sufficient restructuring. For more cohesive aggregates, it may be necessary to put in more energy in the form of high-speed impactors to create a high shear zone in the mixer. However, for sensitive matrices or analytes, care is required to avoid overheating the mixture either locally or as a whole. The second area of attention is to ensure the inclusion of the whole sample in the mixing process. The cohesion of finer powders lends them to be held up in dead spaces in the mixer and be isolated from the mixing action. Efficient scraping and sweeping of the entire mixing volume are essential.

There are several means that may prevent adequate mixing of foods: classification, agglomeration, and phase separation. Causes include differentiated particle sizes and shapes, surface adhesivity, electrostatic charging, disruption of stable structures maintaining surface tension, destruction of encapsulating structures, and various hydrophobic–hydrophilic interactions. Typical practical food examples include oil separation in finely ground nuts and classification of whole grain flour. Samples of brans and flours circulated for dietary fiber collaboration were found impossible to mix;[30] the only means of foiling the classification of these flowable powders was to grind the samples further into cohesive powders. One common means of inducing classification error is the use of a vibrating spatula. This device should only be used for pure compounds. Mixing may be thwarted by the presence of a few large particles, especially if they contain the analyte of interest. Such a mixture needs to be reground to eliminate the large particles.

One may avoid such problems by selecting appropriate homogenization methods through experience and learning, but just as often by intelligent trial and error experimentation. The CO_2 Bamix homogenization method,[21] for example, usually produces a powder with particle size around $100\,\mu m$, yet is usually cohesive in mixing properties. The shearing action of this technique tends to produce irregular elongated particles, and biological particles are often adherent, giving cohesive properties. The cohesion of the sample–CO_2 mixture requires systematic shaking and occasional scraping of the assembly during grinding to ensure inclusion of ''dead'' pockets. Mechanical mortar and pestle[21] grinding similarly requires regular scraping, additional to that provided by the mechanical scraper. This method is very successful on some samples, but is poor for fatty foods where phase separation becomes a problem.

3.3 Evaluation of Homogenization

Mixing may be more difficult to examine than particle size reduction. Experience suggests that visual inspection

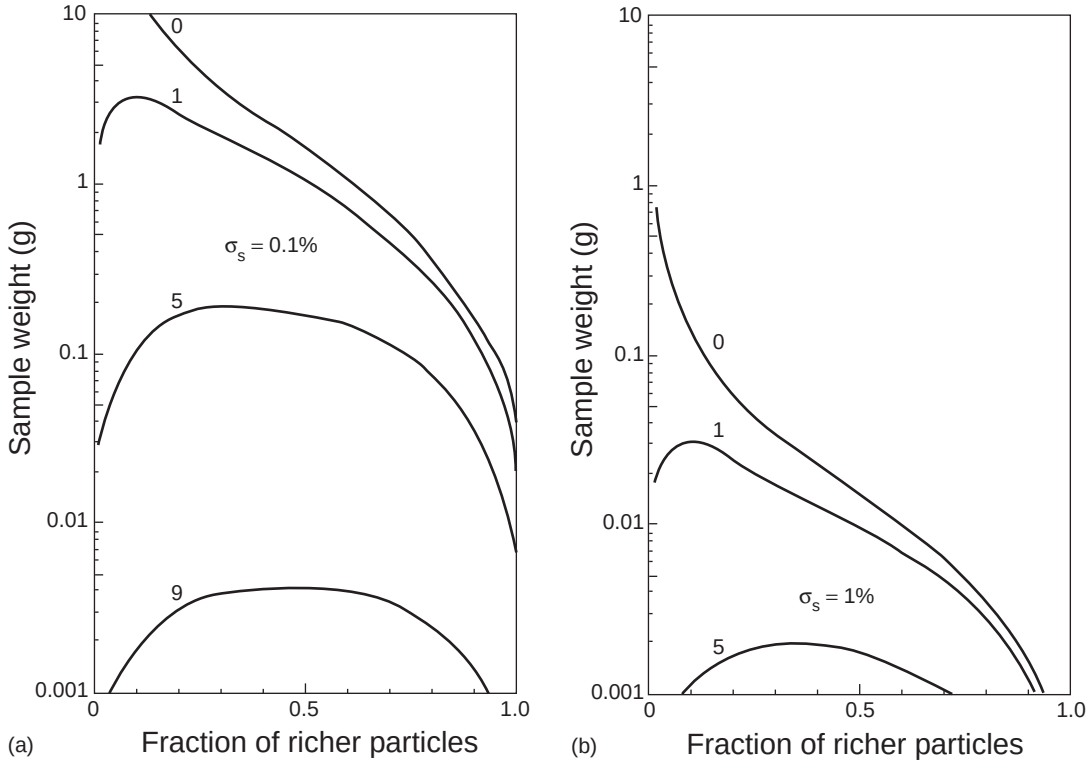

Figure 7 Relation between minimum test portion and fraction of the richer particles in a mixture of two types of spherical particles (diameter 0.1 mm and density 3) for sampling SD of (a) 0.1% and (b) 1%. Richer particles contain 10% of analyte, and leaner ones contain 0, 1, 5 or 9%. (Reproduced from Harris and Kratochvil[6] by permission of W.E. Harris.)

of color and texture is very useful, but not necessarily rigorous, especially in the case of a sample consisting of components of similar appearance. The combined effect of reduction and mixing may be examined by performing assays on replicate test portions. For this examination it may be prudent in some circumstances to run simple, cheap assays rather than to use the actual target assay. The use of an assay more sensitive to variation than the target assay may be warranted. However, for a complex mixture, the confirmation of adequate randomization of one ingredient does not necessarily ensure that other components are mixed. It may often be safer to carry out separate statistical analyses for at least the key different types of analytes associated with different ingredients.

Another method of potential valve is statistical image analysis of the homogenized test portions under a microscope. The determination of particle size distribution is relatively straightforward. It may be necessary to color-label some components prior to treatment to aid differentiation of the constituent particles, and/or observe under various wavelengths. A simple example is the addition of iodine to differentiate starch particles in a light-colored mixture.

The nature of the components, in addition to homogenization, of a sample can be critical to the precision of a given test portion. Figure 7(a) and (b) show huge variations in theoretical curves for simple mixtures; the case of aflatoxins in peanuts has already been mentioned. Fortunately, most recommended homogenization techniques are found simultaneously to reduce particle size and mix samples both with at least reasonable efficiency.

4 SAMPLE PREPARATION

Sample preparation includes any operation performed to the test portion prior to the analytical measurement: storage, preservation, weighing, dilution, cleanup, extraction, digestion, purification, separation, derivatization, etc. Descriptions of these are usually included in publications, although they frequently lack background information, finer practical detail, or rigorous theoretical or practical error analysis.

4.1 Separation of Portions

Careful thought and often some background research are required to decide what parts of the sample are to be analyzed. For example, ought the surface scum and

sediment from a river sample be included in a digest for heavy metals – will the water piped by users include these fractions? Is fresh produce to be washed prior to pesticide analysis? If so, how? What constitutes the ''edible portion'' for a food sample? Examples of such separations include removal of outer leaves, peels, and pips from fresh fruits and vegetables, removal of bones and trimming of excess fat from meat, exclusion of brines from canned vegetables, but inclusion of liquid from canned fruit. Many such choices are subject to debate, such as the inclusion or otherwise of seeds in a sample of blackberry jam. It may be desirable to analyze both portions. In any case, it is standard practice to weigh the separate portions, analogous to the determination of moisture when drying a sample.

4.2 Preparing the Samples

Several workers have emphasized that for nutritional evaluation of food it is desirable to prepare the sample in the same way as it is commonly consumed[18,33] prior to analysis. What are the customary methods and times of cooking of meals? What constitutes complete preparation of powdered soups and hot beverages? An example of inappropriate preparation is the analysis of such beverages for vitamins B_1 and C, where the assay of the beverage powder was assumed to be the measure of dietary intake.[34] In reality, these vitamins degrade significantly when the powders are stirred into boiling water, especially if the water has been sterilized by chlorine or iodine agents, as is usual practice in this context. Vitamins in cooked foods may be reduced by as much as 80% when compared with the raw ingredient vegetables and meats.[35,36] Food held for serving in bains-marie suffers further losses. When bread is baked, 30% of thiamine is lost; if alkaline baking powder is used, the loss can increase to 80%.

4.3 Storage of Samples

Storage and identification of samples require meticulous attention and control. Analyzing the wrong samples is worse than any error generated by sampling, homogenization, preparation, or assay technique! Labeling of samples and all tubes, glassware, and containers used in their analysis must be meticulous, consistent, and unambiguous. The sample trail ought to be auditable at any time. The identification markings should endure any treatment (for example by solvent, microwave digestion, steam bath, freezing, or abrasion) encountered in the handling and analysis process. Storage conditions will be determined by the perishability of the sample/analyte, the delay before assay, and the duration of sample archival. Foods are generally frozen in sealed plastic after preliminary homogenization, or freeze-dried before and after further homogenization treatments and stored in desiccators. Tundra and ice core samples must be kept frozen and tightly sealed to slow deterioration and sublimation freezer burning. Geological drill core samples are best kept in as large pieces as possible in order to slow the degradation of unstable minerals from weathering processes. Change of moisture content, loss of dust fractions, pest attack, and oxidation are all such factors to minimize.

4.4 Weighing and Dilution

A practically universal step in the manipulation of the test portion is the weighing step. Fortunately, the precision of modern balances is commonly six significant figures or better. A judicious choice of balance can maintain this precision for a large range of test portion size, typically from 10 g down to 1 mg for most analyses. However, the analyst must ensure the operation is performed accurately: moisture variation of samples requires particular attention. Electrostatic charging may cause problems with dry samples, and may be eliminated through the use of ionizing tools, such as ''antistatic guns'' sold in audiophile stores for discharging vinyl records. These same errors must be considered when drying samples and determining moisture correction factors. Once dry, samples should be stored in a desiccator. Weighing of dry powders should occur with desiccant present in the weighing chamber of the balance; boats of silica gel are adequate for general use, phosphorus pentoxide being recommended for operations with more hygroscopic materials. Care is needed to avoid moisture adsorption problems with weigh containers and tools: it is prudent to store these in a desiccator, after drying in an oven.

Volumetric dilution is a virtually universal operation contributing to the sample manipulation error. Volumetric errors are inherently orders of magnitude greater than those for mass measurements. The practice of good basic laboratory techniques helps minimize the volume errors, as does to some degree the minimization of the number of volumetric steps. Volumetric errors include the delivery volume of gas chromatography (GC) syringes and the like. For vapor sampling, the size, pressure, equilibration time, and temperature of the vapor space are all critical. It may on occasion be prudent to weigh liquids to reduce the error contribution.

4.5 Digestion, Extraction, Cleanup Separation and Derivatization

Digestion, ashing, filtration, extraction, and derivatization should all be quantitative. Digestion efficiency may be enhanced, for example, by application of microwaves for HPLC,[37] Kjeldahl nitrogen,[38] and metal[39–43] analysis, but only if sample integrity is maintained. Minor elements

may be determined after bomb calorimeter analysis in preference to open-oven dry ashing and digestion.[44] Confirmatory tests should be used in any doubtful cases. A typical simple test is the testing for residual starch with drops of iodine to confirm completion of amylase digestion.[30]

The efficiency of extraction should be shown to be quantitative. Re-extraction of residues should be blank. Recovery of added spikes should be 100%, although this does not conclusively mean that the (bound) sample analyte is necessarily fully extracted. Better assurance may be achieved by the use of two independent methods or the use of standard reference materials. As intimated earlier, the degree of homogenization (particle size) may influence the efficiency of extraction to a greater (Figure 6) or lesser[45] degree. Recent developments in chromatographic analysis include the use of supercritical fluid extraction (SFE).[46–48] The use of CO_2 reduces the potential for contamination and eliminates the problems associated with toxic solvents. Another novel cleanup method is the use of dialysis to remove all but small molecules prior to chromatography.[49] Large molecules may also be digested by enzyme treatment, e.g. prior to amino acid[50] or nitrate analysis.[51]

The methods of ensuring good recoveries while using solid-phase absorption columns prior to liquid chromatography are straightforward. The use of preconcentration methods in GC[52] can be fraught with complications of volatility differences, reactivity, and adsorption. These issues are generally adequately discussed in methodology papers alongside chromatography details.

4.6 Deterioration and Contamination

Degradation of the sample and analyte integrity may take place at any stage, from the taking, transport, and storage of the sample, drying, homogenization, and sample preparation to the analysis of the test solution. Addition of contaminants, exposure of samples to heat, warmth (microbiological activity), moisture, oxygen, visible and ultraviolet light, and reagent fumes can all compromise accuracy. These problems are considerable in vitamin analysis, e.g. riboflavin is sensitive to ultraviolet light. Vitamins A, B_6, D, and E and folic acid are light sensitive. Laboratory manipulations are usually performed using tinted glassware and preferably in the dark.[28,53,54] Plant tocopherols are sensitive to oxidation. Losses during sample preparation are minimized by pyrogallol addition.[55]

Ascorbic acid is particularly sensitive to oxidation, promoted by atmospheric exposure, heat, or high pH. Analysis schemes aim to reduce manipulation and turnaround time to an absolute minimum, making use of various stabilizing agents. A number of methods use metaphosphoric acid addition at the homogenization stage. The 10-fold variation in ascorbic acid found in *Terminalia ferdinandiana*, a native Australian plum rich in this vitamin,[19] is partly due to degradation during sample transit. Furthermore, contamination by traces of copper catalyzes oxidation.

Contamination of samples can occur at any stage of analysis from contact with unsuitable container surfaces, lids or seams, reagents, ingress to open containers, carryover from past samples, diluents, and preservatives. Polytetrafluoroethylene (PTFE), polypropylene, glass, and stainless steel are most commonly used for laboratory ware, but contact with other materials may be unavoidable. However, even glass or stainless steel may leach boron, silica, sodium, chromium, iron, nickel, and molybdenum under certain conditions.

Metal contamination, particularly by trace elements, is a common laboratory problem. Trace metals may be of interest from a nutrient point of view or as environmental contaminants. Many laboratories redistil acids on-site, in addition to using elaborate water purifiers. Recent work quantified the metal increment of using laboratory blenders.[56,57] This increment is able to be reduced by chemical treatment of the blender surfaces.[57] Another approach is to design[58] or seek out homogenization devices with inert working surfaces. For laboratories dedicated to trace metal work, a clean room is a good investment to minimize laboratory environmental contamination.[59] High purity, smooth surfaces, filtered ventilation, and positive pressure airlocks are typical features. Precautions may include the use of overclothing and hair confinement and strict ''hygiene'' practices analogous to those used in infectious bacteriology. Care is taken to avoid transfers via all contact surfaces, hands and implements, air, and dust. Such practices are not confined to metal analysis, but extend to any trace and environmental analysis where contamination of the samples or personnel is a potential hazard. Introduction of metal contamination may be serious beyond the simple raising of metal content. The presence of metal may promote degradation reactions, compromising sample integrity, as mentioned above, but may also interfere with extraction, cleanup, and enzymatic digestion steps of sample preparation procedures.

Contaminants may be introduced by reagents. Method development for enzymatic digestive release of lipids[29] revealed that several commercial enzyme preparations contain unacceptably significant amounts of ether-extractable contaminants, adding artifact to the total determination of fat.

Contamination and the potential complications in sample preparation procedures are highlighted by the GC of alditol acetates in the Englyst dietary fiber determination. Plasticizers may contaminate food

samples at literally any point from the farm to the chromatograph. Pure samples of ubiquitous plasticizer contaminants chromatograph at similar times to some of the analytes,[60] but are be resolvable from analyte peaks by use of capillary columns. However, exposing plasticizers to the derivatization procedure used to form alditol acetates yields multiple and broad peaks that interfere with the analytes (Figure 8a and b). This is an additional artifact caused by the sample preparation technique fundamental to the analysis. Thus one eliminates leachable plastics [such as poly(vinyl chloride)] from the analytical procedure. Sample preparation artifacts have similarly been observed in fatty acid methyl ester analysis,[61] where compounds synthesized by the derivatization interfere with the GC peaks of interest. A novel approach to derivatization of fatty acid mixtures is to react the sample with reagent in the GC injection port,[62] thus eliminating bench chemistry and attendant problems of sample contamination, errors, and limited control of reaction conditions. A monograph[63] describes such analytical artifacts, with considerable attention to problems with gas chromatography/mass spectrometry (GC/MS), a technique widely regarded as definitive with respect to analyte specificity.

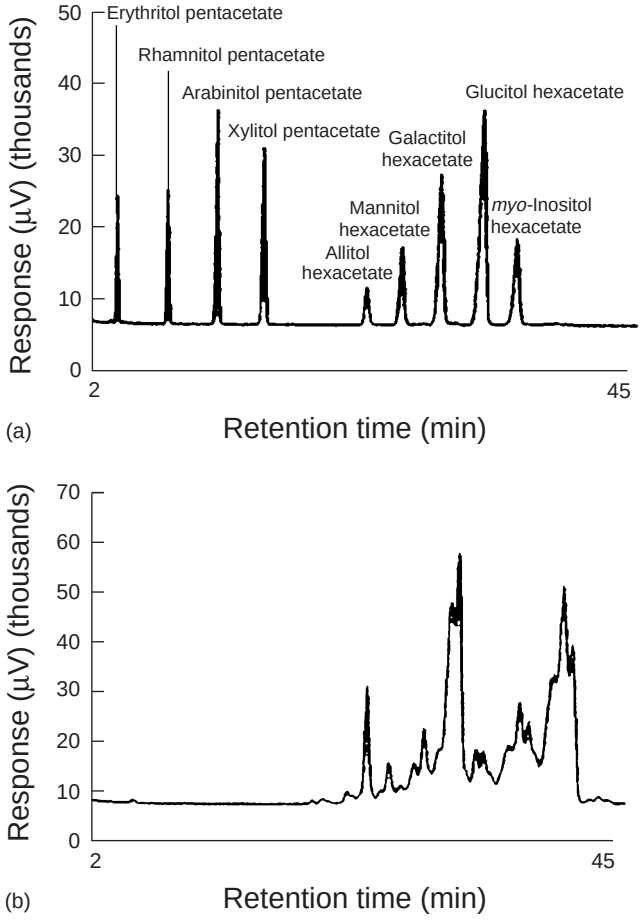

Figure 8 Artifact peaks from "derivatized" contaminant plasticizers (b) can interfere with measurement of alditol acetate analyte peaks (a) for dietary fiber fractionation analysis. (Reproduced from Lichon[1] by courtesy of Marcel Dekker, Inc.)

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
PTFE	Polytetrafluoroethylene
RSD	Relative Standard Deviation
SD	Standard Deviation
SFE	Supercritical Fluid Extraction

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses • Soil Sampling for the Characterization of Hazardous Waste Sites • Soxhlet and Ultrasonic Extraction of Organics in Solids

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Sample Preparation, Headspace Techniques

Industrial Hygiene **(Volume 5)**
Industrial Hygiene: Introduction

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry • Spot Test Analysis • Traceability in Analytical Chemistry

REFERENCES

1. M.J. Lichon, 'Sample Preparation', in *Handbook of Food Analysis*, ed. L.M.L. Nollet, Marcel Dekker, New York, 1–19, 1996.
2. W. Horwitz, J.W. Howard, 'Sampling Methods for Trace Organic Analysis in Foods', *NBS Spec. Publ.*, **519**, 231–242 (1979).
3. L.H. Keith, *Environmental Sampling and Analysis: a Practical Guide*, Lewis, Chelsea, MI, 1991.
4. American Chemical Society, 'Principles of Environmental Analysis', *Anal. Chem.*, **55**, 2210–2218 (1983).
5. Y. Mälkki, 'Collaborative Testing of Methods for Food Analysis', *J. Assoc. Off. Anal. Chem.*, **69**(3), 403–404 (1986).
6. W.E. Harris, B. Kratochvil, *An Introduction to Chemical Analysis*, Saunders, Philadelphia, 1981.
7. J.S. Fritz, G.H. Schenk, *Quantitative Analytical Chemistry*, Allyn & Bacon, Boston, 1978.
8. W.F. Pickering, *Modern Analytical Chemistry*, Marcel Dekker, New York, 1971.
9. A.I. Vogel, *A Textbook of Quantitative Inorganic Analysis*, Longman, London, 1978.
10. B.W. Woodget, D. Cooper, *Samples and Standards*, Wiley, Chichester, 1987.
11. F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, CRC Press, Boca Raton, FL, 1989.
12. L.F. Russell, L.W. Douglass, J.T. Vanderslice, 'Preparation of Sampling Plans for B_2 Vitamins in Fast Food Hamburgers', *J. Food Compos. Anal.*, **5**(3), 224–235 (1992).
13. M.J. Lichon, 'Environmental Monitoring: Problems with Stream Sampling', *Water*, **23**(1), 30–32 (1996).
14. A.D. Campbell, T.B. Whitaker, A.E. Pohland, J.W. Dickens, D.L. Park, 'Sampling, Sample Preparation, and Sampling Plans for Foodstuffs for Mycotoxin Analysis', *Pure Appl. Chem.*, **58**(2), 305–314 (1986).
15. D.L. Park, A.E. Pohland, 'Sampling and Sample Preparation for Detection and Quantitation of Natural Toxicants in Food and Feed', *J. Assoc. Off. Anal. Chem.*, **72**(3), 399–404 (1989).
16. N. Apro, S. Resnik, C. Ferro Fontan, 'Obtaining Representative Samples for Mycotoxin Analysis', *An. Asoc. Quim. Argent.*, **75**(5), 501–510 (1987).
17. S. Hisai, 'Sampling Plan for Inhomogeneously Distributed Contaminants', *Kogai to Taisaku*, **15**(3), 417–423 (1979).
18. H.G. Lento, 'The Role of Sample Preparation in Nutritional Labeling Analysis', *NBS Spec. Publ.*, **519**, 243–248 (1979).
19. K.W. James, A.T. Hancock, M.J. Lichon, L. Robertson, 'Nutritional Assessment of Bush Tucker', in *Chemical International Food Forums Proceedings, August 1989, Brisbane*, 247–290, 1989.
20. J. Brand Miller, K.W. James, P.M.A. Maggiore, *Tables of Composition of Australian Aboriginal Foods*, Aboriginal Studies Press, Canberra, 1993.
21. M.J. Lichon, K.W. James, 'Homogenisation Methods for Analysis of Foodstuffs', *J. Assoc. Off. Anal. Chem.*, **73**, 820–825 (1990).
22. R.B.H. Wills, N. Balmer, H. Greenfield, 'Composition of Australian Foods 2. Methods of Analysis', *Food Technol. Aust.*, **32**, 198–204 (1980).
23. R.A. Beebe, E. Lay, S. Eisenberg, 'Homogeneity of Meats Prepared for Analysis with a Commercial Food Processor: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **72**(5), 777–783 (1989).
24. J.D. Pettinati, S.A. Ackerman, R.K. Jenkins, M.L. Happich, J.G. Phillips, 'Comparative Analysis of Meat Samples Prepared with Food Chopper and Bowl Cutter', *J. Assoc. Off. Anal. Chem.*, **66**(3), 759–765 (1983).
25. J.W. Dorner, R.J. Cole, 'Variability Among Peanut Subsamples Prepared for Aflatoxin Analysis with Four Mills', *J. Assoc. Off. Anal. Chem.*, **76**, 983–987 (1993).
26. M.J. Lichon, 'Sample Preparation Methods for Chromatographic Analysis in Food Chemistry', *J. Chromatogr.*, **624**, 3–9 (1992).
27. W.E. Harris, B. Kratochvil, 'Sampling Variance in Analysis for Trace Components in Solids', *Anal. Chem.*, **46**, 313–315 (1974).
28. K.W. James, A.T. Hancock, 'Simultaneous Determination of Thiamine and Riboflavin by HPLC with Fluorometric Detection', in *Chemical International Food Forums Proceedings, August 1989, Brisbane*, 343–360, 1989.
29. M.J. Lichon, P.J. Tattersall, K.W. James, in *Proceedings of 'Determination of Fat in Foodstuffs', 9th Australian Symposium on Analytical Chemistry, April 1987, Sydney*, Vol. 1, 282–285, 1987.
30. M.J. Lichon, K.W. James, 'Determination of Dietary Fiber and Carbohydrate Composition of Foods Using a Modified Version of the Englyst Method', *J. AOAC Int.*, **79**(1), 54–61 (1996).
31. *Official Methods of Analysis of the Association of Official Analytical Chemists*, AOAC, Arlington, VA, sect. 985.29, 1990.

32. N. Harnby, M.F. Edwards, A.W. Neinow, *Mixing in the Process Industries*, Butterworth-Heinemann, Oxford, 1992.
33. F.M. Garfield, 'Sampling in the Analytical Scheme', *J. Assoc. Off. Anal. Chem.*, **72**, 405–411 (1989).
34. K.W. James, G.F. Thomson, A.T. Hancock, G.J. Walker, R.A. Coad, M.J. Lichon, 'Laboratory Evaluation of Australian Ration Packs', MRL Report, MRL-TR-92-30, Defence Science and Technology Organization, Australia, Melbourne, 1993.
35. C.A. Kwiatkowska, P.M. Finglas, R.M. Faulks, 'The Vitamin Content of Retail Vegetables in the UK', *J. Hum. Nutr. Dietet.*, **2**, 159–172 (1989).
36. R.A. McCance, E.M. Widdowson, *The Composition of Foods*, HMSO, London, 1978.
37. K. Ganzler, A. Salgó, K. Valcó, 'Microwave Extraction: a Novel Sample Preparation Method for Chromatography', *J. Chromatogr.*, **371**, 299–306 (1986).
38. L.W. Colins, S.J. Chalk, H.M. Kingston, 'Atmospheric Pressure Microwave Sample Preparation Procedure for the Combined Analysis of Total Phosphorus and Kjeldahl Nitrogen', *Anal. Chem.*, **68**, 2610–2614 (1996).
39. H.M. Kuss, 'Applications of Microwave Digestion Technique for Elemental Analysis', *Fresenius' J. Anal. Chem.*, **343**(9–10), 788–793 (1992).
40. C.L. Suard, M.H. Feinberg, J. Ireland-Ripert, R.M. Mourel, 'Interlaboratory Validation of Analysis of Minerals by Microwave Digestion', *Analysis*, **21**(6), 287–291 (1993).
41. H.M. Kingston, L.B. Jassie, 'Microwave Energy for Acid Decomposition at Elevated Temperatures and Pressures Using Biological and Botanical Samples', *Anal. Chem.*, **58**, 2534–2541 (1986).
42. M.D. Mingorance, M.L. Perez-Vazquez, M. Lachica, 'Microwave Digestion Methods for the Atomic-spectrometric Determination of Some Elements in Biological Samples', *J. Anal. At. Spectrom.*, **8**, 853–858 (1993).
43. A. Krushevska, R.M. Barnes, C. Amarasiriwaradena, 'Decomposition of Biological Samples for Inductively Coupled Plasma Atomic Emission Spectrometry Using an Open Focused Microwave Digestion System', *Analyst*, **118**(9), 1175–1181 (1993).
44. G. Schwedt, 'Combined Calorimetry and Ion Chromatography for the Determination of Food Energy and Minerals', *Dtsch. Lebensm.-Rundsch.*, **87**(7), 223–224 (1991).
45. M. Takeda, H. Sekita, K. Otsuki, H. Tanabe, Y. Kawai, Y. Mishima, K. Kogo, Y. Katayama, 'Analysis of Pesticide Residues in Foods. XI. Influence of the Degree of Powdering on the Extraction of Organochlorine Pesticides from Cereals and Beans', *Shokuhin Eiseigaku Zasshi*, **14**(6), 569–573 (1973).
46. L.J.D. Myer, J.H. Damian, P.B. Lieschеski, J. Tehrani, D. Conseil, 'Comparison of Supercritical Fluid Extraction and the Soxhlet Method', *Spectra-2000 (Deux-Mille)*, **165**, 57–63 (1992).
47. F. David, 'Supercritical Fluid Extraction as a Sample Preparation Technique', *Chem. Mag. (Ghent)*, **17**(4), 12, 14, 15 (1991).
48. F. Verillon, 'Supercritical Fluids as an Effective, Rapid, Economical and Non-polluting Solution for Industrial Analysis', *Analusis*, **24**(3), M17–M19 (1996).
49. F. Verillon, F. Qian, 'Coupling of Dialysis with Liquid Chromatography for Automated Analysis of Complex Samples', *Analusis*, **19**(9), 271–277 (1991).
50. M. Hauck, 'Enzymic Hydrolysis of Proteins for Amino Acid Determination and Sample Preparation', *Dtsch. Lebensm.-Rundsch.*, **86**(1), 12–15 (1990).
51. R. Gromes, T. Siegl, 'Enzymic Sample Preparation for Food Analysis', *GIT Fachz. Lab.*, **35**(6), 623–626 (1991).
52. W.G. Jennings, M. Filsoof, 'Comparison of Sample Preparation Techniques for Gas Chromatographic Analysis', *J. Agric. Food Chem.*, **25**, 440–445 (1977).
53. D.J. Aulik, 'Sample Preparation for Nutrient Analysis', *J. Assoc. Off. Anal. Chem.*, **57**(5), 1190–1192 (1974).
54. M.H. Bui, 'Sample Preparation and Liquid Chromatographic Determination of Vitamin D in Food Products', *J. Assoc. Off. Anal. Chem.*, **70**(5), 802–805 (1987).
55. T. Ujiie, A. Tanaka, A. Kondo, R. Hiroe, T. Tetsushige, 'Effect of Pyrogallol Addition in Sample Preparation on the Determination of Tocopherols in Foods', *Vitamin*, **66**(2), 101–107 (1992).
56. C. Schlage, B. Wortberg, 'Influence of Homogenization Processes on the Iron Content of Food', *Z. Lebensm.-Unters.-Forsch.*, **145**(2), 97–99 (1971).
57. I.B. Razagui, P.J. Barlow, 'A Chemical Cleanup Procedure to Reduce Trace Metal Contamination from Laboratory Blenders', *Food Chem.*, **44**, 309–312 (1992).
58. V.W. Bunker, H.T. Delves, R.F. Fautley, 'A System to Minimize Trace Metal Contamination of Biological Material During Homogenization', *Ann. Clin. Biochem.*, **19**(6), 444–445 (1982).
59. G.E. Batley, *Trace Element Speciation: Analytical Methods and Problems*, CRC Press, Boca Raton, FL, 1989.
60. R.J. Henry, P.J. Harris, A.B. Blakeney, B.A. Stone, 'Separation of Alditol Acetates from Plasticizers and Other Contaminants by Capillary Gas Chromatography', *J. Chromatogr.*, **262**, 249–256 (1983).
61. C.F. Moffat, A.S. McGill, R.S. Anderson, 'Production of Artefacts During Preparation of Fatty Acid Methyl Esters from Fish Oils, Food Products and Pathological Samples', *J. High Resolut. Chromatogr.*, **14**(5), 322–326 (1991).
62. E.L. Nimz, S.L. Morgan, 'Online Derivatization for Complex Fatty Acids Mixtures by Capillary Gas Chromatography–Mass Spectroscopy', *J. Chromatogr. Sci.*, **31**(4), 145–149 (1993).
63. B.S. Middleditch, *Analytical Artifacts*, Elsevier, New York, 1986.

Sample Preparation, Headspace Techniques

Alain Chaintreau*
Nestlé Research Centre, Lausanne, Switzerland

1 Introduction 4230
2 Hyphenation to Gas Chromatography 4230
2.1 Static Headspace 4230
2.2 Purge and Trap Headspace 4232
2.3 Static and Trapped Headspace 4234
2.4 Solid-phase Microextraction Headspace 4235
2.5 Vacuum Headspace 4236
2.6 Membrane Techniques 4236
2.7 In Situ Headspace Gas Chromatography 4236
2.8 General Remarks 4237
3 Direct Hyphenation to a Detector 4237
3.1 Sensor Array 4237
3.2 Mass Spectrometer 4238
4 Applications 4238
4.1 Foods, Flavors, Fragrances, Essential Oils 4238
4.2 Environment 4239
4.3 Health and Biosciences 4239
4.4 Other Applications 4240
5 Comparison of Recovery Efficiencies of Headspace Methods 4240
6 Determination of Thermodynamic Constants 4240
6.1 Air-to-liquid Partition Coefficients 4240
6.2 Liquid-to-liquid Partition Coefficients 4241
6.3 Activity Coefficients 4241
6.4 Other Thermodynamic Constants 4241
7 Conclusion 4241
Acknowledgments 4242
Abbreviations and Acronyms 4242
Related Articles 4242
References 4242

Among sample preparation techniques used prior to gas chromatography (GC), headspace sampling (HS) offers many advantages as it only collects vapors that are directly compatible with GC. The main variants are: (1) collection of a fraction of the HS (static headspace, SHS), (2) gas stripping of volatiles from a solution and subsequent trapping in an adsorbent (purge and trap headspace, P&THS), (3) collection and trapping of volatiles from a whole HS in equilibrium with a liquid (static and trapped headspace, S&THS), (4) partition of volatiles between a sample, its HS, and an adsorbent fiber (headspace solid-phase microextraction (HSSPME)). For each one, a design and some guidelines for qualitative and quantitative use are given, as well as the equation giving the theoretical recovery. Performances are compared between HS techniques and others such as simultaneous distillation extraction (SDE).

A special section is devoted to the timesaving hyphenation of HS to detectors without chromatographic separation like HS-sensor array and headspace sampling mass spectrometry (HSMS).

Main HS applications in various areas (food, environment, health and biosciences) since the late 1980s are illustrated. In addition, HS appears to be a unique tool to measure thermodynamical constants of volatiles characterizing their partitioning between phases and their interactions with a nonvolatile matrix.

1 INTRODUCTION

Discussing HS techniques in a sample preparation article is something of a paradox as this methodology needs little sample preparation, if any at all! As GC is only suitable for vaporizable components, these latter must normally be isolated from nonvolatile compounds of the matrix by an appropriate technique prior to injection (extraction, distillation, etc.). In contrast HS only samples vapors which are directly compatible with the GC process.

HS using a gas syringe is one of the oldest injection techniques in GC (see section 2.1.1). This article mostly deals with the headspace sampling gas chromatography (HSGC) hyphenation (section 2) because it is the subject of the majority of published papers. However, there is increasing interest in direct coupling of HS to detectors enabling the specific monitoring of a given compound in a mixture without preseparation of its constituents. This timesaving approach is also considered hereafter (section 3). At the end of this article, the use of HSGC as a unique technique to measure some thermodynamic properties of volatiles that are difficult to obtain by other means is considered (section 6).

* Present address: Firmenich SA, Corporate Research, 1 route des Jeunes, CP 239, Geneva 8, Switzerland

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 HYPHENATION TO GAS CHROMATOGRAPHY

2.1 Static Headspace

Readers interested in an extensive review of SHS should refer to Kolb and Ettre's book.[1]

2.1.1 Design and Principle

2.1.1.1 Manual Injection The most simple SHS injection mode only requires a gas-tight syringe and a closed vial containing the liquid or solid sample. Both are kept in a thermostated oven to equilibrate their temperature prior to withdrawing the headspace. The usual split/splitless injector may be used. Detailed procedures exist for volatiles in polymers and flavors.[2]

2.1.1.2 Autosampler To overcome reproducibility problems that are inherent in manual injection (thermostatization, control of the pressure within the vial and the syringe, injection duration, etc.), SHS autosamplers were developed early on. The vapor phase may be transferred from the vial to the column in two different ways. In the first,[3] a vacuum is applied to a gas loop connected to the vial by a needle. After filling the loop, its content is pushed by the carrier gas onto the column by switching a multiport valve (Figure 1). In the second,[1] the vial is pressurized by the carrier gas, then the carrier gas flow is stopped to release the vial pressure directly onto the GC column (Figure 2).

In both cases, there is a modification of the HS pressure and only a fraction of the gas phase is collected. In commercial autosamplers, the vial as well as the transfer line are thermostated to avoid any condensation. The pressure and the times of equilibration and sampling are electronically controlled to ensure good reproducibility.

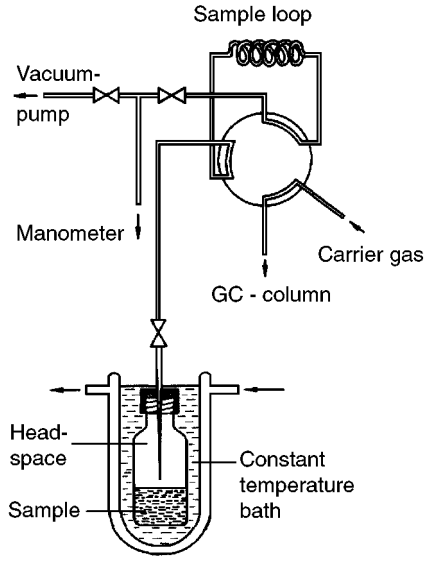

Figure 1 SHS injector based on a gas loop. (Reproduced with permission from B.R. Rao, 'Determination of the Maximum Emission from Storage Tanks for Heavy Fuel Oil', in *Applied Headspace Gas Chromatography*, ed. B. Kolb, Heyden, London, 61–68. Copyright John Wiley & Sons Limited (1980).[3])

2.1.2 Theoretical Recovery

The gas-to-liquid partition coefficient is defined as the ratio of the concentration in air C_g to that in the liquid C_l, Equation (1):

$$k_{gl} = \frac{C_g}{C_l} \tag{1}$$

If a fraction f of the HS volume is transferred onto the GC column, the recovery ρ_i of a volatile is given by

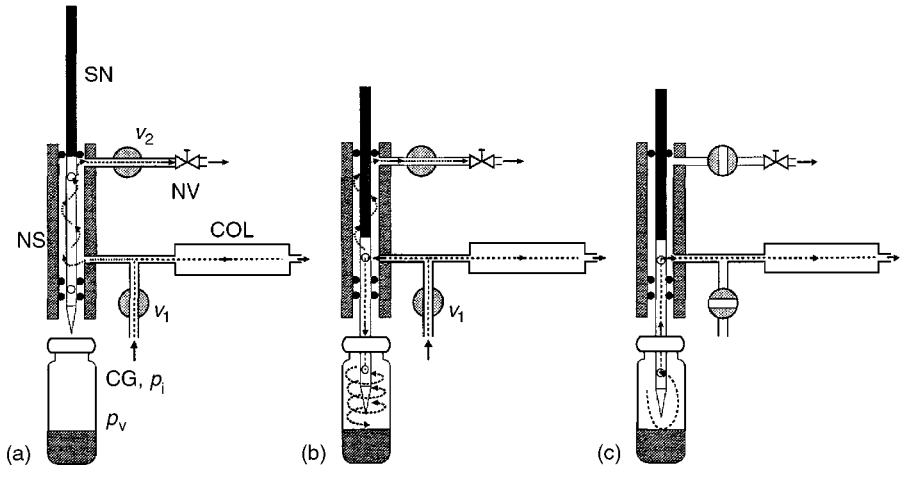

Figure 2 SHS injector based on vial pressurization. (a) equilibration, (b) pressurization, (c) sample transfer. (Reproduced with permission of Wiley-VCH, Copyright 1997.[1])

Equation (2):[1]

$$\rho_i = \frac{f k_{gl} V_g}{V_l + k_{gl} V_g} \quad (2)$$

where V_g is volume of HS and V_l is volume of liquid.

From Equation (2) the injected amount strongly depends on the partition coefficient k_{gl}, the fraction of the collected gas volume f, and the ratio of phase volumes. Maximizing V_g/V_l maximizes ρ_i, while f is fixed by the hardware to less than 1. Alternatively, a higher temperature improves the sensitivity as k_{gl} increases with this parameter. However, for aqueous samples this possibility is limited by the enhancement of the moisture level which disturbs the chromatography.

2.1.3 *Use for Quantitation*

Usual GC quantitation methods, *internal normalization* and *internal standardization*, are theoretically applicable to SHS/GC. The former requires reconstitution of a mixture containing all constituents of the sample (all volatiles diluted *in the same matrix*). In practice only fully gaseous samples that have a relatively simple composition can be analyzed in SHS/GC by internal normalization.[1]

Internal standardization applied to SHS also suffers from some limitations. The added internal standard should have the same air-to-matrix partition coefficient as the target compounds, otherwise partition coefficients of the internal standard and target volatiles must be accurately known. In addition, the matrix used to dilute the standards must be identical to that of the sample, without containing target compounds or the internal standard. The same requirements for the matrix and partition coefficients remain valid for a third method, *external standardization*. The removal of volatiles from the liquid sample has been proposed so that the "deodorized" liquid can be used as a solvent for the preparation of standard solutions.[4]

Because of the above mentioned limitations, two other techniques are generally preferred when a SHS sampling is performed: standard addition, and multiple headspace extraction (MHE). Standard addition overcomes the problem of the matrix choice because known amounts of the target compound are added to the sample itself. Details of an example of quantitation by standard addition are given in section 2.3.3.

MHE eliminates the matrix problem, as a total extraction of the analyte is simulated. Therefore it has often been applied to very complex matrices such as those that are found in the environment (e.g. soils). A series of pressurizations with the carrier gas, and GC injections of the HS are performed using the same sample contained in

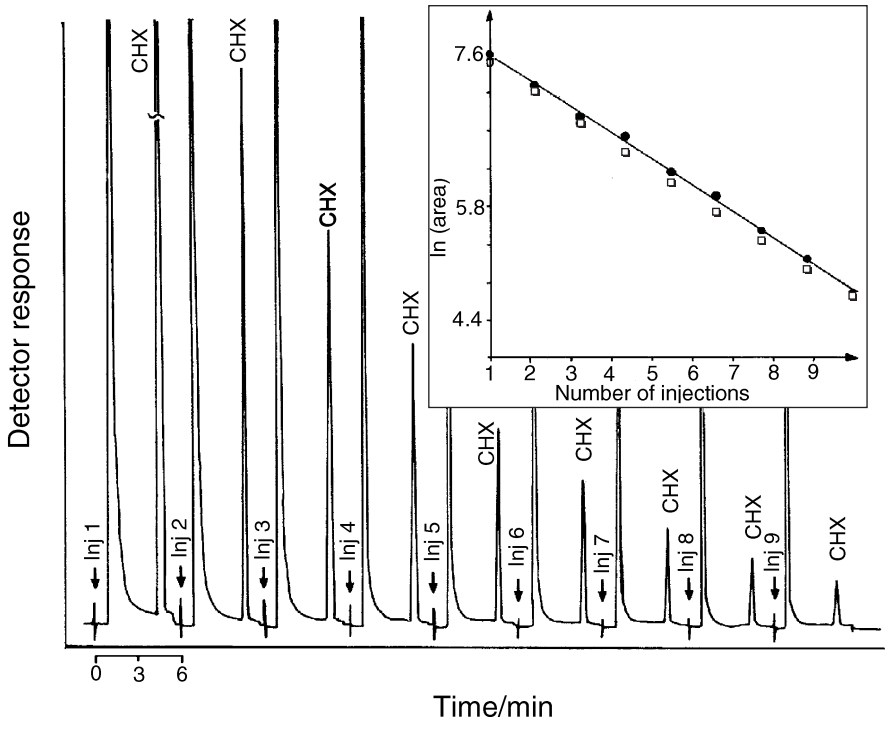

Figure 3 Chromatogram from a nine-step MHE determination of cyclohexanone (CHX) in a contaminated soil sample, and regression lines using a standard (❑) and the sample (●). (Reproduced from Milana et al., *J. Chromatogr.*, **552**, 205–211. Copyright 1991 with permission of Elsevier Science.[8])

a closed vial. After each injection, volatiles migrate from the matrix to the gas phase to re-establish a new partition between the phases. Amounts of volatile extracted into the HS decrease according to Equation (3):[5]

$$\ln(A_i) = -k(n-1) + \ln(A_1) \quad (3)$$

where A_i is the area of the analyte in the ith HS injection, n is the number of extractions and A_1 is the area of the analyte in the first injection.

Consequently, the area A_T corresponding to the total amount originally present in the sample is the sum of areas of all extracted amounts, Equation (4):[5]

$$A_T = \sum(A_i) = \frac{A_1}{1 - e^{-k}} \quad (4)$$

In practice a few extraction steps are sufficient to obtain k from Equation (3), by applying a regression calculation (Figure 3). The calibration may be performed by external standardization using the same volatile in a different matrix, or by internal standardization with another volatile, submitted to the same MHE process.[6] In both cases, the difference of volatility inherent to the different matrices or compounds does not influence the A_T value because this area represents the whole quantity present in the matrix. Recoveries in the 90–110% range are usually observed with relative standard deviations often less than 15%.[5–7]

2.2 Purge and Trap Headspace

2.2.1 Design and Principle

To collect a large volume of HS, and subsequently a greater amount of its volatiles than those present at equilibrium, a liquid sample may be continuously stripped by a flow of pure and inert gas. Volatiles must then be trapped from the gas flow.[9] This can be achieved in three ways:

1. They can be dissolved in a solvent reflux and GC injected.[10]
2. They may be cryofocused in a cold loop,[11] or on the beginning of the GC-column itself.[12]
3. Many automated systems trap them in an adsorbent (graphite, Tenax®, Porapak®, etc.), from which they are either thermally desorbed,[9] or eluted with a solvent.[13] Figure 4 shows the principle of a usual automated P&THS system. As a variant for solid materials, a piece of the sample can be put directly into a thermal desorber under a flow of gas. Volatiles released by the controlled heating are trapped and then GC injected.[14] To analyze plants, only a small amount (1–40 mg) of the raw material is required.

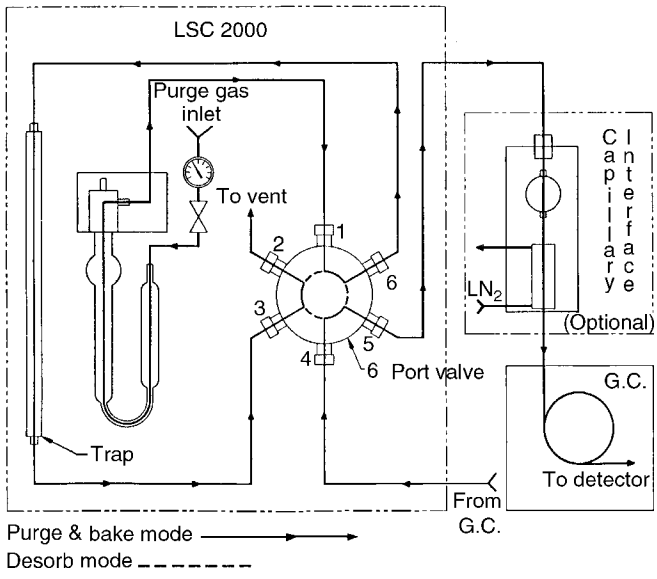

Figure 4 Scheme of an automated P&THS injector. (Reproduced with permission of Tekmar-Dohrmann, USA.)

Desorption from a trap may be achieved in two ways. The adsorbant can be rinsed off with a small amount of solvent (diethyl ether, CS_2, etc.) that is then GC injected.[15–17] As the solvent choice sometimes gives rise to selective desorption, recoveries must be checked.[18] The second way is heating of the trap. This alternative is the most frequently used as numerous automated thermal desorbers are now commercially available. Thermolabile components such as terpenes are advantageously desorbed with a solvent to avoid artifacts.[17] However, adsorbent pretreatment with an antioxidant and a buffer has been said to prevent degradation of these fragile compounds during thermal desorption.[19]

Most P&THS applications have been developed for aqueous samples because the solvent must be nonvolatile or it must not be trapped on the adsorbent. Water fits the second requirement. However, any other solvent answering these conditions could be used, as well as solid matrices.

In a variant, the purging and trapping systems are connected in a closed loop assembly owing to a pump circulating the stripping gas in the loop.[20–22] This method is claimed to perform significant extractions of furanol and maltol, scarcely strippable aroma compounds.[21]

2.2.2 Theoretical Recovery

The recovery of a given volatile stripped from water by the gas flow is given by:[23]

$$\rho_i = 1 - e^{\frac{-k_{gl}Ft}{k_{gl}V_g + V_l}} \quad (5)$$

where F is the flow of stripping gas, t is the duration of the stripping, V_g is the void volume of the sampling cell, V_l is the volume of the solution, and k_{gl} is the gas-to-liquid partition coefficient.

If $k_{gl}V_g$ is negligible (small void volume of the cell and/or small k_{gl} values), Equation (5) can be simplified:[23]

$$\rho_i = 1 - e^{\frac{-k_{gl}Ft}{V_l}} \quad (6)$$

Equation (6) shows that, for a given k_{gl} value, the sensitivity is better than when using SHS because the total gas volume (Ft) may be much greater. Therefore the GC profile obtained after P&THS is not representative of the HS composition at equilibrium. For an infinite volume V_g of stripping gas, the recovery of all volatile solutes should be quantitative. However V_g is limited by the trapping step, as described in the next paragraph. Practically speaking, hydrophobic compounds are often fully recovered, whereas hydrophilic ones remain in the aqueous solution.[18] Therefore recoveries strongly depend on the air-to-liquid partition coefficients. As these latter increase with temperature, sensitivity may also be increased by heating the stripped sample.[18] The influence of various experimental parameters has been investigated.[24] Equation (5) is also valid for closed loop sampling.[23]

2.2.3 Adsorbent Properties

Several types of traps are used:

- porous polymers (Amberlite®, Chromosorb®, Porapak®, Tenax®, etc.) that are generally hydrophobic[23]
- active charcoal (Ambersorb®, Carbosieve®, Carbopack®, etc.)[23,25]
- liquid chromatographic phases (e.g. polysiloxane) either ground,[26] or as multichannel tubes,[27] or bonded to the inner wall of an open tubular capillary column.[28]

The first two materials utilize adsorption, while the third utilizes dissolution in the liquid phase. Traps made from several layers which provide increasing trapping efficiency may be helpful for trace analysis down to the parts per trillion level.[29,30] Hydrophobic adsorbents retain organic compounds advantageously without trapping water, which could disturb the chromatographic process when present in large amounts.

It must be recalled that a trapping step is only chromatography through a very retentive phase. Care must be taken to limit stripping gas volumes to less than the breakthrough volume of the target compound in a given adsorbent, otherwise losses must be expected.[20,31] Breakthrough volumes of many volatiles in various adsorbents have been published.[20,31] Very efficient adsorbents such as active charcoal may not quantitatively release trapped components.[32] The desorption step may give rise to artifact formation. For instance microwave heating of graphite traps seems to yield benzene derivatives systematically.[33] Therefore, blank experiments must be run, using the same system and, possibly, the same matrix (at least pure water for aqueous samples).

2.2.4 Use for Quantitation

In many cases, quantitation using P&THS is not simple because several parameters influence the experiment. Recoveries of volatiles vary according to temperature and gas flow fluctuations, stripping duration and different air-to-matrix partition coefficients of components. This can be partially overcome by electronically controlled purge and trap autosamplers (time, flow and temperature regulation). However, the variability of k_{gl} values between solutes of a sample is sometimes forgotten by analysts.

2.2.4.1 External Standardization Some authors calibrate by injecting known standard amounts either into the trap using a syringe,[34] or into splitless injectors without changing other chromatographic conditions.[17] Such a procedure does not take into account the stripping efficiencies of the analyte, which must be evaluated separately to correct quantities found using the calibration curve.[34] The calibration curve is also often established by submitting model solutions containing standards to the same purge and trap conditions as the target sample.[35] This requires either that recoveries are quantitative or that identical efficiency is obtained for the quantitated sample and standard solutions. This last condition is only fulfilled if standard solutions are prepared in the same matrix as the sample. This matrix must obviously not contain any of the compounds to be quantified.

2.2.4.2 Internal Standardization The sample is spiked with a standard prior to stripping. It assumes that standard and target compounds are fully recovered or that they possess identical air-to-sample partition coefficients. Although it is rarely checked, internal standardization is often used, with the risks inherent in this approximation.[36,37] The correction of results with recoveries estimated in a different matrix from the sample does not seem satisfactory.[16]

MHE has also been used, with some modifications of the static MHE model to take into account the purge process.[38]

2.2.4.3 Standard Addition According to Equations (5) or (6) recoveries are independent of the concentration. Therefore, added or pre-existent amounts of

the target compound are recovered in identical ratios and a calibration curve may be drawn.[35] The procedure is similar to that used for SHS and S&THS.

2.2.4.4 Labeled Standards The use of an isotopomer of the target compound allows its accurate quantitation when it is represented by a very small peak included in a complex chromatogram.[29] The problem of purge and trap recovery efficiency does not intervene because both labeled and unlabeled compounds are assumed to behave similarly.

2.3 Static and Trapped Headspace

2.3.1 Design and Principle

The injection of large gas volumes exceeds capacities of classical injectors used in SHS and the measurement of the total gas volume is not accurate in P&THS. S&THS combines advantages of both techniques (Figure 5).[39] Volatiles are equilibrated between the sample receptor and the HS chamber. When the piston is pressed, volatiles are evacuated through the transfer tube and trapped by the adsorbent. This latter is then transferred to a thermal desorber for injection onto the GC column.

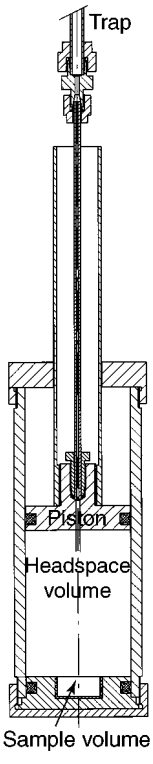

Figure 5 Cell for S&THS sampling. (Reproduced with permission from Chaintreau et al., *Anal. Chem.*, **34**, 3300–3304. Copyright 1995 American Chemical Society.[39])

2.3.2 Theoretical Recovery

The formula representing the recovery is similar to that of SHS, but the gas volume is greater and fully recovered, Equation (7):[39]

$$\rho_i = \frac{k_{gl} V_g}{V_l + k_{gl} V_g} \tag{7}$$

where V_g is the volume of headspace and V_l is the volume of liquid.

For a given volatility, the sensitivity mainly depends on the gas volume V_g which is only limited by the size of the cell. As for SHS, a temperature enhancement also increases k_{gl} and then ρ_i.

2.3.3 Use for Quantitation

Two S&THS methods have been proposed for quantitation. The first one[39] requires determination of air-to-liquid partition coefficients, using the same HS cell (see section 6.1). Then HS concentrations may be calculated from the initial concentration of aqueous calibration samples. Using the resulting curve, HS peak areas of the sample to be analyzed can be transformed into quantities. The coefficient of variation for the method was estimated to be 3.4% for 2-butanone.[39] The MHE procedure is suitable to determine the concentration in the solution. No application of this first method has yet been published.

The second method uses the standard addition procedure.[40] Known quantities of the target compound are added to the analyzed solution. The HS of the spiked samples is then collected by S&THS and injected. By determining the straight line representing peak areas as a function of the added standard concentrations, including the area of the compound without addition, the initial amount of the target compound is given by extrapolating the curve to zero area. The intercept with the abscissa represents its concentration (Figure 6). Applied to yogurt, recoveries were 87–114% and standard deviations less than 5.5%.[40]

2.4 Solid-phase Microextraction Headspace

2.4.1 Design and Principle

Solid-phase microextraction (SPME) was proposed in 1990 as a rapid extraction technique from a liquid sample. In the original concept, a fiber coated with a bonded liquid phase is dipped into the liquid and then desorbed in the GC injector. It was observed that volatiles could also be extracted from the vapor phase by exposing the fiber to the HS of a sample contained in a closed vial (Figure 7). It has been shown that recoveries from the HS at equilibrium are often close to those from the liquid phase (see section 2.4.2). In addition HSSPME avoids polluting the fiber with nonvolatiles in the liquid. The

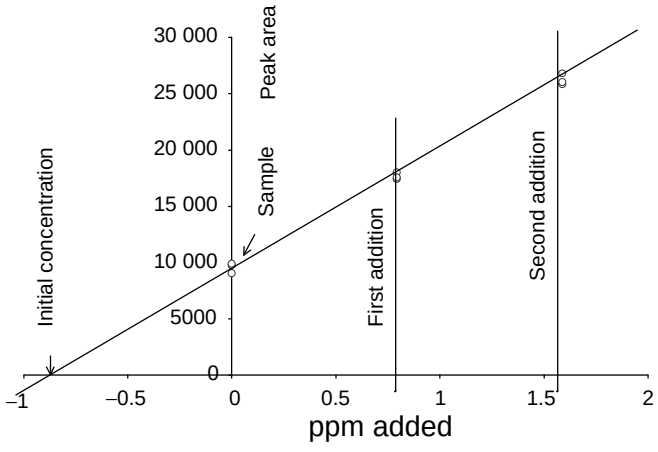

Figure 6 Quantification by standard additions in yogurt using 3-penten-2-one as a test compound and the S&THS cell.[40]

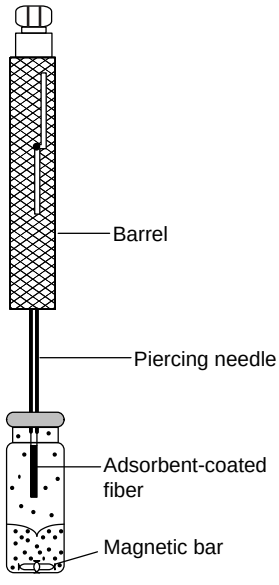

Figure 7 Principle of HSSPME.

technique has been reviewed by its inventor.[41] Fibers of various polarities (Carbowax, polyacrylate, carboxen, polydimethylsiloxane (PDMS) alone or as mixture with polydivinylbenzene) and phase thickness (7–100 μm) are available. Fiber performances rapidly change as a function of the number of extractions and the fiber must be regularly renewed.[42]

2.4.2 Theoretical Recovery

In the closed vial, the recovery accounts for a three-phase equilibrium: sample/HS/fiber, Equation (8):[43]

$$\rho_i = \frac{V_l k_{fl}}{V_f k_{fl} + V_g k_{gl} + V_l} \quad (8)$$

where k_{fl} is the fiber-to-liquid partition coefficient.

When the term $k_{gl}V_g$ is small (e.g. pyridine: $k_{gl} = 5.71 \times 10^{-4}$ at 25 °C, $V_g = V_l = 1$ mL), Equation (8) can be simplified as Equation (9):[43]

$$\rho_i = \frac{V_l k_{fl}}{V_f k_{fl} + V_l} \quad (9)$$

which represents the recovery from a liquid sample. Therefore it must be pointed out that, in spite of its name, *HSSPME is not representative of the headspace composition*, but rather that it represents the direct partition between the fiber and the liquid.

If the sampling temperature rises, the extraction is faster, but the effect on the HS concentration at equilibrium depends on the compound. If k_{fl} is greater than 1 (exothermic partitioning into the fiber), increasing the temperature decreases k_{fl} and vice versa.[41]

Equation (8) also shows that the sensitivity is improved by maximizing V_f/V_l or minimizing V_g/V_l.

2.4.3 Use for Quantitation

SPME has been shown to be reproducible when experimental conditions are kept constant (stirring rate, temperature, sampling time).[44] The time to reach the air/matrix equilibrium varies greatly according to the solute, the matrix and the fiber coating. For instance HS-fiber exposures of about 25 and 150 min were necessary for Dichlorvos and Napromide in water, respectively.[45] This means that this time to equilibrium should always be determined prior to developing a quantitation method. However quantitation under nonequilibrium conditions has been proven to be feasible.[46]

External standardization is the most common SPME quantitation method.[47,48] The calibration solvent must obviously be identical to the sample one, without containing target components. Internal standardization is not easy because of recoveries varying with varying partitions of the internal standard and the target compounds. Its application to tobacco alkaloids was not found to be fully satisfactory.[49] Standard addition overcomes matrix problems better by using the sample as the solvent.[50] Best results are obtained with labeled standard and MS (mass spectrometry) detection.[51]

2.5 Vacuum Headspace

The literature sometimes mentions as ''vacuum headspace'' sampling[52] techniques involving the distillation of volatiles under high vacuum. These methods will not be mentioned in this article because, in the author's opinion, they relate to distillation or steam-distillation, rather than to HS.

2.6 Membrane Techniques

2.6.1 Design and Principle

Pervaporation membranes may help HS and can be used in two ways:

- One side of the membrane is contacted with the liquid sample, while the second side is swept by the GC carrier gas flow (Figure 8).[53,54] Volatile solutes permeate through the membrane and are conveyed to the GC column. They can be concentrated by interfacing the membrane device and the chromatograph with a microtrap. Aqueous benzene solutions may be quantitated in the 10–150 ppb range by external standardization.
- Membrane extraction with sorbent interface (MESI). Instead of contacting one of its sides with the liquid sample, the membrane separates the sample HS from the carrier gas flow which extracts volatiles passing across the hollow-fiber membrane.[55] Quantitation by standard addition has been used.

2.7 In Situ Headspace Gas Chromatography

2.7.1 Design and Principle

In situ HS refers to the monitoring of volatiles emitted by a source as a function of time. A typical example is the release of odorants by a living plant over the day.[56] This requires a noninvasive sampling technique such as trapping of volatiles from air by a trap, subsequent desorption and GC analysis. The emitter (e.g. a living flower) is enclosed in a HS chamber made of glass, Plexiglas™, or in a polyvinyl fluoride bag, connected to a pure air inlet (Figure 9). The gas exiting the chamber passes through a trap such as those described for P&THS. Recent systems use an array of several traps in parallel, into which the exiting air is successively admitted. Time programming of traps and other parameters (temperature, pressure, air humidity, etc.) are controlled by computer.[57,58]

2.7.2 Quantitative Use

Obtaining real quantitative results by in situ HS may be hard because leaks are difficult to prevent (e.g. because of the difficulty of sealing a fragile stem passing through

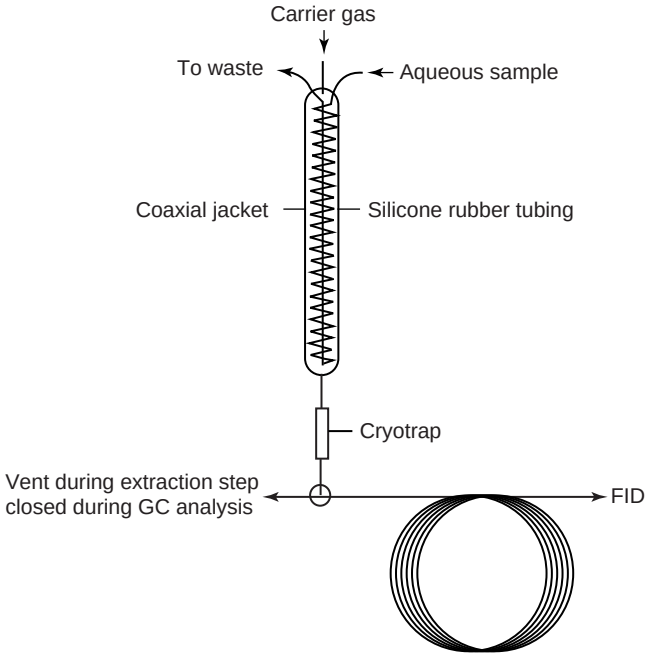

Figure 8 HS by permeation through a hollow-fiber membrane in conjunction with carrier gas stripping. FID, flame ionization detector. (Reproduced with permission of Wiley-VCH from *J. High. Resolut. Chromatogr.* Copyright 1996.[53])

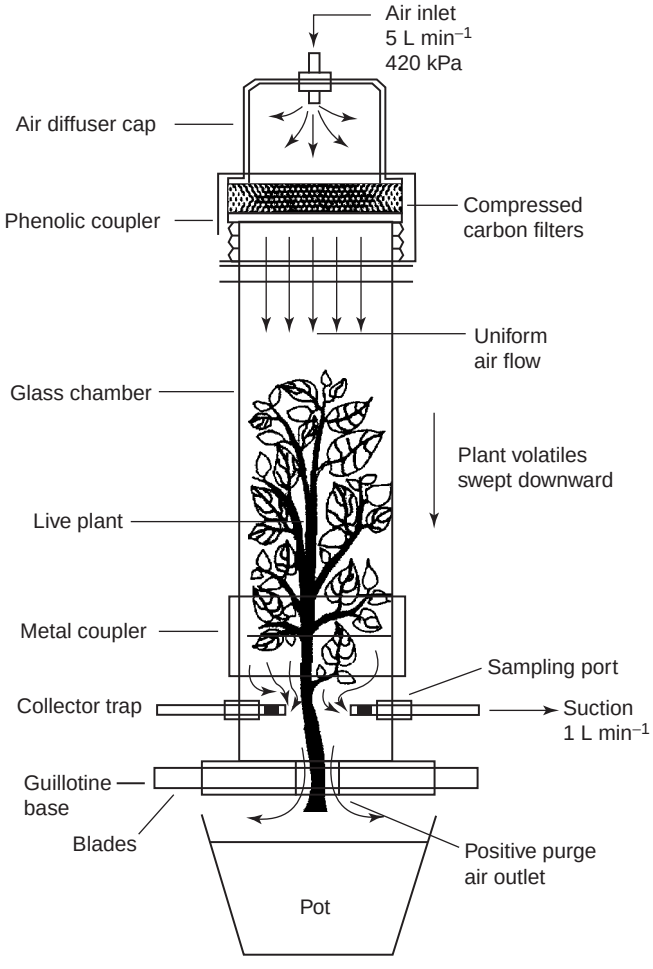

Figure 9 Collection assembly for in situ HS analysis of volatiles emitted by a plant. (Reproduced with permission of Plenum Publishing Corp. from *J. Chem. Ecol.* Copyright 1994.[57])

the chamber wall). In addition the incorporation of an internal standard to the emitter is not possible. Addition of the internal standard after solvent elution of the adsorbent has been mentioned.[57]

2.8 General Remarks

2.8.1 Sample Storage

When analyses cannot be performed close to the HS collection area, samples may need to be stored and transported to the laboratory. Several means have been proposed. Teflon® or Tedlar® bags are often used, but permeation of volatiles through the polymeric wall occurs and they are not recommended for long-term storage.[59] The air to be analyzed can be admitted into metallic canisters or glass bulbs previously evacuated to a low pressure.[30] To avoid degradation of sulfur-containing constituents, the stainless steel wall can be coated with a bonded silica layer.[60] For large volumes, the gas can be percolated through an adsorbent which only retains organic compounds (see section 2.2.3). Airtight tubes containing trapped volatiles may be stored up to 2 years.[61]

2.8.2 Salting Out

A usual means of increasing the volatility of organics in aqueous samples consists in the addition of salt. Recoveries of volatiles are normally greatly improved.[42,62–64]

2.8.3 Labeled Standards

With any HSGC technique the best way of calibrating for quantitation remains the use of an isotopomer of the target compound itself as the internal standard. This requires detection by MS.

3 DIRECT HYPHENATION TO A DETECTOR

Hyphenation of HS to GC exhibits a major drawback for some applications: it does not provide the analyst with real-time monitoring of the volatiles. Even automated systems described for in situ HS require an off-line trap desorption and a time-consuming GC elution. Therefore the direct hyphenation of HS to a detector has been studied. When the objective is to monitor one or several compounds in a mixture, selective detectors must be chosen.

3.1 Sensor Array

This detection means has become very popular and is often improperly called an ''electronic nose'' because it mimics odor detection by the array of receptors in the nose. However its sensitivity and selectivity cannot be compared to those of the mammalian organ which is able to detect a component having an extremely low odor threshold in a complex mixture of less intense odorants (e.g. 18 ppb of 2,4,6-trichloroanisole in cork-tainted wine[65]).

Semiconductor metal oxides, conductive polymers or quartz resonators have been used to build sensors. When volatile organic compounds (VOCs) are present in the sample's HS, the sensor's conductivity or resonance frequency changes according to the nature of the sensor's material and the nature of the organic compounds.[66] Individual signals produced by the elements of the sensor array generate a response profile more or less characteristic of the detected odor. Although sensors exhibiting different selectivities may be used in the same array, this selectivity remains low. However, in spite of their limited ability to provide information about HS composition, sensor arrays can easily detect composition differences in the vapor phase, especially with the aid of chemometrics to treat raw data, such as principal component analysis (PCA).[67] Therefore sensors are most often used in quality assurance to classify global profiles of volatiles present in the gas phase, without giving information on its composition.[68]

3.2 Mass Spectrometer

3.2.1 Chemical Sensor

Routine MS detectors have been directly hyphenated with HS autosamplers without any GC separation in between. These systems use an electronic impact ionization (70 eV) yielding a strong fragmentation. The resulting complex spectra are used as ''fingerprints''. Therefore data treatment also requires chemometrics to extract relevant information.[25,69]

3.2.2 Proton-transfer Reaction Mass Spectrometry

This technique[70] uses chemical ionization by H_3O^+ ions which form MH^+ quasimolecular ions with little fragmentation. In addition, H_3O^+ does not react with the natural constituents of air. Therefore, since spectra of mixtures mainly contain quasimolecular ions of their constituents, following one of them becomes possible, as long as no isobaric ion interferes with that of the compound of interest. Consequently, monitoring the HS composition as a function of (real) time, or of any parameter that can be varied as a function of time, is feasible, with a quantitative signal. It is an alternative to in situ HSGC (section 2.7) with a lower selectivity and a quicker response.

Resonance-enhanced multiphoton ionization, hyphenated to MS, also permits direct monitoring of the HS without previous separation. Laser excitation of the volatiles leads to specific ionization of a chemical class of compounds.[71]

4 APPLICATIONS

Between 1987 and 1998, the Chemical Abstracts' compilation lists about 1070 papers related to the ''headspace'' keyword. This does not take into account publications using sensor arrays. Three main application areas of HS may be distinguished: (1) food, flavors and fragrances (34%); (2) environment (26%); (3) health and biosciences (21%). The remaining papers deal with method development, phase equilibrium, plastics and inorganic analysis. Only examples or review papers will be cited to illustrate each topic because references are too numerous to be exhaustive within the limits of this article.

4.1 Foods, Flavors, Fragrances, Essential Oils

4.1.1 Identification and Quantification of Flavor and Fragrance Constituents

Many articles and reviews have already been dedicated to the use of all HS techniques for the identification and quantification of flavors and fragrances.[36,72,73] Because S&THS collects odorants that are in equilibrium with a given food, for example in its packaging, this sampling method was recently hyphenated to GC-olfactometry to determine aroma impact compounds.[40,74] With the help of statistical tools (e.g. PCA) to treat HSGC results, some sample classification can be achieved according, for instance, to botanical varieties such as arabica/robusta coffees[75] or grape vines.[42]

4.1.2 Off-flavors

HS is a suitable technique for investigating off-flavors such as trichloroanisole which confers the cork taint on wine.[65] It can be associated with GC-olfactometry to determine undesirable odorants by comparison with the aromagram of the nonaltered sample.[76]

4.1.3 Food Process Monitoring

Maturing processes may need an objective evaluation in addition to sensory analysis. This is especially valuable for fermented foods such as dry ham[37] or olives in brine.[77] The thermal generation of an aroma, like that of roast coffee[78] or process flavors,[79] can also be monitored by HS.

4.1.4 Food Preservation

The quality and safety of industrial food products requires their shelf life and storage conditions to be determined.[80] In this respect, lipid oxidation is a major cause of deterioration that may be detected by the formation of volatile degradation products.[34,81,82]

Applications of sensor arrays to food have been reviewed.[83]

4.2 Environment

4.2.1 Chemicals in Air

HS is obviously the technique of choice for air pollution. This topic starts with indoor pollutants imparting, for example, an off-odor in a smoker's office.[30] The composition of outside air reflects the great variety of human activity. It can be controlled at this level, or emitting sources like incinerators may be monitored on-line.[78] Diurnal variations in the production of natural organics (e.g. isoprene) released by green plants in the atmosphere can be followed.[70] An uncommon hyphenation with microwave-induced plasma detection HSGC/(MIP) has been used to investigate the effect of methylmercury on wildlife.[84]

4.2.2 Water Pollutants

This has been the area of predilection for the development of P&THS.[11,20] Membrane techniques[53] have appeared, but HSSPME has gained a great popularity because of its simplicity.[63]

4.2.3 Soil Contamination

In addition to the usual methods (extraction, column partitioning), SHS and P&THS are useful. However, preparing representative calibration standards is difficult because of the heterogeneous nature of the matrix. This problem can be overcome by MHE which does not require any spiking of the matrix for the calibration.[7,8]

4.2.4 Pollutants in Food

Undesirable volatile compounds contaminating foods and beverages may be monitored by HS techniques. They can originate from agricultural treatments, from the animal feeding chain, or from monomers coming from the packaging polymer, and so on.[6,29,85]

4.3 Health and Biosciences

This section includes pharmaceutical and medical applications, forensic science, and fermentative processes.

4.3.1 Metabolites from Fermentative Processes

Although HS analyses are unable to provide direct information about bacteria, these latter may be recognized from to their metabolites.[86] Using precursors labeled with stable isotopes, biological pathways leading to the production of volatiles in yogurt have been investigated.[87,88] The production of human growth hormone by *Escherichia coli* was monitored with a sensor array.[67] HSSPME allows a quick analysis of aroma compounds generated by staphylococci.[48]

4.3.2 Health Diagnostic and Forensic Science

HS is very useful for the detection of volatiles in biological fluids, such as alcohol in blood,[89,90] odorants in urine,[22] and toxic substances in cadavers.[91,92] HS methods, mainly static ones, applied to biological samples to analyze carbon monoxide, alcohols and their metabolites, solvents, gases, hydrogen cyanide and endogenous volatile metabolites have been reviewed.[62]

Direct HSMS hyphenation has permitted the real-time monitoring of volatiles in breath. As an example, smokers were shown to breath out benzene and acetonitrile as long as 1 week after stopping smoking (Figure 10).[70,93]

4.3.3 Pharmacy

HSGC analysis is used to analyze pharmaceutical products.[94] Main applications relate to the determination of residual solvents or other volatile impurities,[86] the evaluation of residual ethylene oxide used as a sterilizing agent[86] and active ingredients such as medicinal plants and flowers.[73]

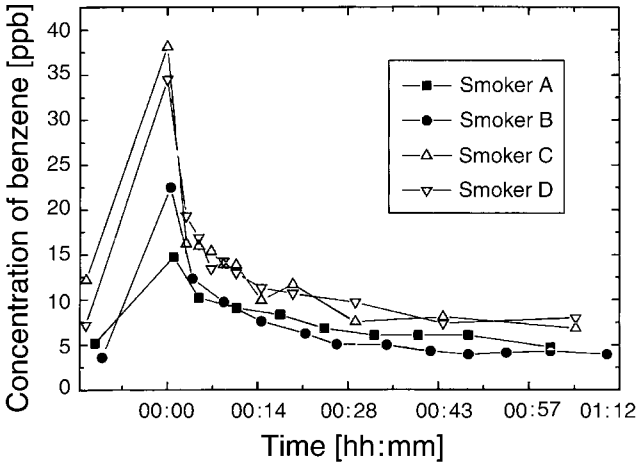

Figure 10 Concentration of benzene in the breath of four test persons smoking a cigarette before $t = 0$. (Reproduced from W. Lindinger, A. Hansel, A. Jordan, *Int. J. Mass Spectrom. Ion Processes*, **173**, 191–241. Copyright 1998 with permission from Elsevier Science.[70])

4.4 Other Applications

References dealing with method development and phase equilibria are not mentioned in this section as details are given in sections 2, 3 and 6.

4.4.1 Agriculture

Pests are sometimes attracted by volatiles emitted by colonized plants. HS analyses enable scientists to determine possible attractants.[95] In the same way fly pheromones were identified.[96]

4.4.2 Polymers

Polymers often contain residual solvents, monomers and antioxidants that are analyzed after desorption and HS.[38] Gas permeation through polymers used as packaging material can be investigated by HS.[80]

5 COMPARISON OF RECOVERY EFFICIENCIES OF HEADSPACE METHODS

Equations given in the ''theoretical recovery'' sections of this article (sections 2.1.2, 2.2.2, 2.3.2, 2.4.2), allow expected results from various techniques to be compared (Table 1). SHS is not included, as its yield is a fraction of that of S&THS. Partition coefficients and recoveries vary in the same direction, for SHS, S&THS and P&THS. HSSPME does not exhibit the same trend, owing to the previously mentioned fact (section 2.4.2) that recoveries were close to those of a direct contact of the fiber with the liquid. HSSPME for 1 and 100 mL HS and the same sample amount (1 mL) illustrate the lower yields with increasing gas volumes. Highest theoretical recoveries are obviously observed for P&THS, which is itself less efficient than SDE,[72] a non-HS technique which is, however, based on volatilities of solutes.

Some experimental method comparisons have been published. HSSPME gave lower detection limits (DLs) than SHS, and both results were highly correlated when applied to the quantitation of alcohols and esters in beer.[47] Similar DLs were observed for substituted benzenes in water using HSSPME and P&THS and quantitative results with both techniques were also strongly correlated.[63]

More generally, DLs were in the parts per million range for SHS, in the parts per billion range for SPME, solid-phase extraction, liquid extraction, and in the parts per billion range for P&THS.[97] Experimental recovery comparisons with other, HS and non-HS, sample preparation techniques (extraction, SDE) have been

Table 1 Calculated recoveries of selected compounds under standard operating conditions according to published theoretical models

	Air/water partition coefficient (25 °C)	S&THS yield (%)	HSSPME (PDMS) for V_g =		P&THS yield (%)
			1 mL yield (%)	100 mL yield (%)	
Vanillin	6.49×10^{-9}	9.7×10^{-7}	2.5×10^{-5}	2.5×10^{-5}	2.6×10^{-5}
2-Phenylethanol	1.57×10^{-4}	2.3×10^{-2}	4.3×10^{-2}	4.3×10^{-2}	6.3×10^{-1}
Pyridine	5.70×10^{-4}	7.9×10^{-2}	7.8×10^{-1}	7.4×10^{-1}	2.3
2-Propanone	1.47×10^{-3}	1.8×10^{-1}	1.1×10^{-3}	9.6×10^{-4}	5.7
Benzene	2.50×10^{-1}	9.7×10^{-1}	6.1×10^{-1}	2.9×10^{-2}	100

S&THS $V_l = 2$ mL, $V_g = 300$ mL. P&THS $V_l = 5$ mL; $F \times t = 200$ mL ($F \times t$ is the total volume of stripping gas).
SPME $V_l = 1$ mL; $V_g = 1$ or 100 mL; $V_f = 1.32 \times 10^{-4}$ mL for a film thickness of 7 μm.
Additional numerical values for HSSPME and P&THS have been given elsewhere.[72]

published.[98] Direct thermal desorption of volatiles from a solid mentioned in section 2.2.1 was found to be more reproducible and to yield less artifacts than SDE under atmospheric pressure.[14]

6 DETERMINATION OF THERMODYNAMIC CONSTANTS

HSGC is not only a unique tool in chemical analysis, it can also be used to measure thermodynamic parameters characterizing equilibria between phases, interactions between compounds, and so on. Only three examples are given here: air-to-liquid and liquid-to-liquid partition coefficients and activity coefficients.

6.1 Air-to-liquid Partition Coefficients

The gas-to-liquid partition coefficient may be determined by several methods. For instance, using SHS and the ''vapor phase calibration'' method,[1] an aliquot of the HS is injected after fully vaporizing a known amount of volatile in a closed empty vial of volume V_v. The same amount is introduced into another closed vial, of the same volume, containing a volume V_s of solvent. Two peak areas A_c and A_g, respectively, are obtained. The coefficient is then calculated as, Equation (10):

$$k_{gl} = \frac{A_g V_s}{A_c V_v - A_g V_g} \tag{10}$$

As many methods involve conditions that may alter sampled quantities (full vaporization calibration, pressure change in the vial during sampling, adsorptions and leaks when using glass syringes, and so on), a more sophisticated method has been proposed using the S&THS cell shown in Figure 5.[39] GC peak areas versus initial concentrations in the liquid phase form a straight line of slope p. Determining this slope using two different volumes of the liquid sample, V_{l1} and V_{l2} gives two values p_1 and p_2. Using the ratio $\rho = p_1/p_2$, the air-to-liquid partition coefficient is given by Equation (11):

$$k_{gl} = \frac{(1-\rho)V_{l1}V_{l2}}{(\rho V_{l2} - V_{l1})V_g} \tag{11}$$

6.2 Liquid-to-liquid Partition Coefficients

For a volatile solute, these can be indirectly determined from separate measurements of the two air-to-liquid partition coefficients of the volatile in each solvent, Equations (12) and (13)

$$k'_{gl} = \frac{C'_g}{C'_l} \tag{12}$$

$$k''_{gl} = \frac{C''_g}{C''_l} \tag{13}$$

where k'_{gl} and k''_{gl} are the air-to-liquid partition coefficients in both systems.

As both gas phase concentrations are equal at equilibrium, Equation (14):

$$k_{l'l''} = \frac{C'_l}{C''_l} = \frac{k''_{gl}}{k'_{gl}} \tag{14}$$

where $k_{l'l''}$ is the liquid-to-liquid partition coefficient.

This has been applied to calculate water-to-fat partition coefficients because extracting organic solutes from a fat to determine their concentration is not easy.[99]

6.3 Activity Coefficients[100]

For ideal solutions (no interaction between components of the mixture), the partial pressure p_i of a given

compound in the vapor phase obeys Raoult's law, Equation (15):

$$p_i = p_i^0 x_i \tag{15}$$

where p_i is the vapor pressure of compound i and x_i is the molar fraction of i in the mixture.

However real solutions often deviate from this ideality, then Henry's law applies, Equation (16):

$$p_i = p_i^0 \gamma_i x_i \tag{16}$$

where γ_i is the activity coefficient of i.

As the GC peak area of a volatile is proportional to its partial pressure in the gas phase, Henry's law for a binary mixture can be rewritten as, Equation (17):

$$\frac{A_i}{\lambda} = \frac{A_i^0}{\lambda^0}\gamma_i x_i \tag{17}$$

where λ, λ^0 are GC response factors.

Considering GC response factors to be independent of the concentration, $\lambda = \lambda^0$, then, measuring GC peak areas from the HS of pure i and in the mixture gives the activity coefficient, Equation (18),

$$\gamma_i = \frac{A_i}{A_i^0 x_i} \tag{18}$$

By stripping a solution with an inert gas and HSGC analysis using thermal conductivity and a FID in series, the determination of activity coefficients of volatile solutes and water activity were realized in a single operation.[101]

6.4 Other Thermodynamic Constants

Other properties of volatile substances may be characterized by HSGC. The reader is invited to consult references cited for vapor pressures, partial free molar energy of mixing,[1] fugacity[102] and interactions with nonvolatile solutes (e.g. proteins): number of binding sites, intrinsic affinity constant, enthalpy, entropy and free energy of binding.[103,104]

7 CONCLUSION

Innovation in the HS area since the late 1980s has been characterized by three main topics:

- Automation of injectors allowing a better reproducibility. Consequently the quantitative performances have been improved as well.
- Invention of SPME, and its application to HS. Its simplicity has been responsible for its rapid popularity.
- Direct hyphenation of HS and detectors. The greatest development was dedicated to sensor arrays, but additional efforts are required to improve their sensitivity and to mimic the human nose better. New alternatives like HSMS are only in their infancy, but seem very promising because they give a more informative response.

ACKNOWLEDGMENTS

We gratefully acknowledge Professors C. Bicchi, A. Voilley, Dr Prior and Mr Pollien who reviewed this paper.

ABBREVIATIONS AND ACRONYMS

DL	Detection Limit
FID	Flame Ionization Detector
GC	Gas Chromatography
HS	Headspace Sampling
HSGC	Headspace Sampling Gas Chromatography
HSMS	Headspace Sampling Mass Spectrometry
HSSPME	Headspace Solid-phase Microextraction
MESI	Membrane Extraction with Sorbent Interface
MHE	Multiple Headspace Extraction
MIP	Microwave-induced Plasma
MS	Mass Spectrometry
PCA	Principal Component Analysis
PDMS	Polydimethylsiloxane
P&THS	Purge and Trap Headspace
SDE	Simultaneous Distillation Extraction
SHS	Static Headspace
SPME	Solid-phase Microextraction
S&THS	Static and Trapped Headspace
VOC	Volatile Organic Compound

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Food **(Volume 5)**
Flavor Analysis in Food

REFERENCES

1. B. Kolb, L. Ettre (eds.), 'Static Headspace/Gas Chromatography', Wiley-VCH, New York, 1997.
2. American Society for Testing and Materials (ASTM), 'Standard Practice for Determination of Volatiles in Polymers by Static Headspace Gas Chromatography', in *Annual Book of ASTM Standards*, American Society for Testing and Materials (ASTM), 1–3, 1997 and Working Group on Methods of Analysis, 'Recommended Methods', International Organization of the Flavor Industry (IOFI), Geneva (CH), IOFI, Methods 1 and 22, 1973.
3. B.R. Rao, 'Determination of the Maximum Emission from Storage Tanks for Heavy Fuel Oil', in *Applied Headspace Gas Chromatography*, ed. B. Kolb, Heyden, London, 61–68, 1980.
4. E.A. Baldwin, M.O. Nisperos-Carriedo, R. Baker, J.W. Scott, 'Quantitative Analysis of Flavor Parameters in Six Florida Tomato Cultivars *(Lycopersicon esculentum Mill)*', *J. Agric. Food Chem.*, **39**, 1135–1140 (1991).
5. A.D. Uhler, L.J. Miller, 'Multiple Headspace Extraction Gas Chromatography for the Determination of Volatile Halocarbon Compounds in Butter', *J. Agric. Food Chem.*, **36**, 772–775 (1994).
6. B. Kolb, 'Multiple Headspace Extraction. A Procedure for Eliminating the Influence of the Sample Matrix in Quantitative Headspace Gas Chromatography', *Chromatographia*, **15**, 587–594 (1982).
7. A. Maggio, M.R. Milana, M. Denaro, R. Feliciani, L. Gramiccioni, 'Multiple Headspace Extraction/Capillary-gas Chromatography (MHE/CGC) for the Quantitative Determination of Volatiles in Contaminated Soils', *J. High Resolut. Chromatogr.*, **14**, 618–620 (1991).
8. M.R. Milana, A. Maggio, M. Denaro, R. Feliciani, L. Gramiccioni, 'Modern Approach to the Quantitative Determination of Volatiles in Solid Samples. Multiple Headspace Extraction Gas Chromatography for the Determination of Cyclohexanone Residues in Soil', *J. Chromatogr.*, **552**, 205–211 (1991).
9. S.M. Abeel, A.K. Vickers, D. Decker, 'Trends in Purge and Trap', *J. Chromatogr. Sci.*, **32**, 328–338 (1994).
10. A. Rapp, W. Knippser, 'Eine Neue Methode zur Enreicherung von Dampfkomponenten. Dargestellt am Beispiel des Weines', *Chromatographia*, **13**, 698–702 (1980).
11. P.G. Simmonds, 'Analysis of Trace Halocarbons in Natural Waters by Simplified Purge and Cryotrap Method', *J. Chromatogr.*, **289**, 117–127 (1984).
12. J.F. Pankow, M.E. Rosen, 'Determination of Volatile Compounds in Water by Purging Directly to a Capillary Column with Whole Column Cryotrapping', *Environ. Sci. Technol.*, **22**, 398–405 (1988).
13. R. Buttery, R. Teranishi, L.C. Ling, 'Fresh Tomato Volatiles: A Quantitative Study', *J. Agric. Food Chem.*, **35**, 540–544 (1987).
14. J.L. Esteban, I. Martinez-Castro, R. Morales, B. Frabrellas, J. Sanz, 'Rapid Identification of Volatile Compounds in Aromatic Plants by Automatic Thermal Desorption – GC/MS', *Chromatographia*, **43**, 63–72 (1996).
15. R. Buttery, R. Teranishi, R.A. Flath, L.C. Ling, 'Fresh Tomato Volatiles. Composition and Sensory Studies', in *Flavor Chemistry: Trends and Developments*, American Chemical Society, 213–222, 1989.
16. R. Buttery, R. Teranishi, L.C. Ling, 'Fresh Tomato Aroma Volatiles: A Quantitative Study', *J. Agric. Food Chem.*, **35**, 540–544 (1987).
17. B.V. Burger, Z. Munro, 'Liquid Desorption of Headspace Volatiles Trapped on Activated Carbon Open Tubular Traps', *J. Chromatogr.*, **402**, 95–103 (1987).
18. J. Curvers, T. Noy, C. Cramers, J. Rijks, 'Possibilities and Limitations of Dynamic Headspace Sampling as a Pre-concentration Technique for Trace Analysis of Organics by Capillary Gas Chromatography', *J. Chromatogr.*, **289**, 171–182 (1984).
19. A.-M. Strömvall, G. Petersson, 'Protection of Terpenes Against Oxidative and Acid Decomposition on Adsorbent Cartridges', *J. Chromatogr.*, **589**, 385–389 (1992).
20. A. Bianchi, M.S. Varney, J. Phillips, 'Modified Analytical Technique for the Determination of Trace Organics in Water Using Dynamic Headspace and Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **467**, 111–128 (1989).
21. R. Buttery, L.C. Ling, 'Methods for Isolating Food and Plant Volatiles', in *Biotechnology for Improved Foods and Flavors*, eds. G. Takeoka, R. Teranishi, P.J. Williams, A. Kobayashi, ACS, Washington, DC, 241–248, 1996.
22. H.J. Bestmann, K. Haberkorn, O. Vostrowsky, R. Ferstl, F. Eggert, 'GC Profile of Volatile Constituents from Human Urine Obtained by Closed Loop Stripping, Purge and Trap Technique and Simultaneous Steam Distillation-extraction', *Z. Naturforsch.*, **51**, 849–852 (1996).
23. A.J. Nuñez, L.F. Gonzalez, J. Janak, 'Pre-concentration of Headspace Volatiles for Trace Organic Analysis by Gas Chromatography', *J. Chromatogr.*, **300**, 127–162 (1984).
24. R. Kostiainen, 'Effect of Operating Parameters in Purge-and-trap GC/MS of Polar and Nonpolar Organic Compounds', *Chromatographia*, **38**, 709–714 (1994).
25. C.K. Meng, P.L. Wylie, C. Cai, C. Sharrar, R. Firor, 'A Chemical Sensor for Fast Chemical Integrity Monitoring', *Am. Lab.*, News Edition, **31**(13), 28–29 (1999).
26. E. Baltussen, F. David, P. Sandra, H.G.M. Janssen, C.A. Cramers, 'Sorption Tubes Packed with Polydimethysiloxane: A New and Promising Technique for the Preconcentration of Volatiles and Semivolatiles from

Air and Gaseous Samples', *J. High Resolut. Chromatogr.*, **21**, 332–340 (1998).
27. E.K. Ortner, E.R. Rohwer, 'Trace Analysis of Semivolatile Organic Air Pollutants Using Thick Film Silicone Rubber Traps with Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **19**, 339–344 (1996).
28. C. Bicchi, A. D'Amato, F. David, P. Sandra, 'Capturing of Volatiles Emitted by Living Plants by Means of Thick Film Open Tubular Traps', *J. High Resolut. Chromatogr.*, **12**, 316–321 (1989).
29. C. Nerin, C. Rubio, J. Cacho, J. Salafranca, 'Parts-per-trillion Determination of Styrene in Yogurt by Purge-and-trap Gas Chromatography with Mass Spectrometry Detection', *Food Addit. Contam.*, **15**, 346–354 (1998).
30. C.W. Bayer, 'Advances in Trapping Procedures for Organic Indoor Pollutants', *J. Chromatogr. Sci.*, **32**, 312–316 (1994).
31. I. Maier, M. Fieber, 'Retention Characteristics of Volatile Compounds on Tenax® TA', *J. High Resolut. Chromatogr.*, **11**, 566–576 (1988).
32. J.A. Rojas, A. Voilley, C. Thomas, 'Study of Adsorption of 1-Octen-3-ol on Porous Hydrophobic Polymers', *Gas Sep. Pur.*, **5**, 187–194 (1991).
33. B. Toulemonde, D. Beauverd, 'Headspace Analysis: Trap Desorption by Microwave Energy. Application to the Volatile Components of some Tropical Fruits', in *Progress in Flavor Research 1984*, ed. J. Adda, Elsevier Science B.V., Amsterdam, 533–548, 1984.
34. P.S.W. Park, R.E. Goins, 'Determination of Volatile Lipid Oxidation Products by Dynamic Headspace-capillary Gas Chromatography Analysis with Application to Milk-based Nutritional Products', *J. Agric. Food Chem.*, **40**, 1581–1585 (1992).
35. O.L. Lum, M.K. Wong, C.K. Lee, 'A Simple Headspace-gas Chromatography Method for the Quantitative Determination of Organic Volatiles of Fresh Orange Juice', *Food Chem.*, **37**, 313–317 (1990).
36. T.G. Hartman, J. Lech, K. Karmas, J. Salinas, R.T. Rosen, C.T. Ho, 'Flavor Characterization Using Adsorbent Trapping/Thermal Desorption or Direct Thermal Desorption/Gas Chromatography and Gas Chromatography/Mass Spectrometry', in *Flavor Measurement*, eds. C.T. Ho, C.H. Manley, M. Dekker, New York, 37–60, 1993.
37. M. Flores, C.C. Grimm, F. Toldra, A.M. Spanier, 'Correlation of Sensory and Volatile Compounds of Spanish ''Serrano'' Dry-cured Ham as a Function of Two Processing Times', *J. Agric. Food Chem.*, **45**, 2178–2186 (1997).
38. A. Hagman, S. Jacobsson, 'Quantitative Determination of Volatiles in Polyolefins by Dynamic Headspace/Capillary Gas Chromatography/Mass Spectrometry', *J. High Resolut. Chromatogr.*, **11**, 830–836 (1988).
39. A. Chaintreau, A. Grade, R. Munoz-Box, 'Determination of Partition Coefficients and Quantitation of Headspace Volatile Compounds', *Anal. Chem.*, **34**, 3300–3304 (1995).
40. A. Ott, M. Baumgartner, A. Chaintreau, 'Aroma Comparisons of Traditional and Mild Yogurt: Headspace-GC Quantitation of Volatiles and Origin of α-diketones', *J. Agric. Food Chem.*, **47**, 2379–2385 (1999).
41. J. Pawliszyn (ed.), *Solid Phase Microextraction*, Wiley-VCH, New York, 1997.
42. D. De la Calle Garcia, M. Reichenbächer, K. Danzer, C. Hurlbeck, C. Bartzsch, K.H. Feller, 'Analysis of Wine Bouquet Components Using Headspace Solid-phase Microextraction-capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **21**, 373–377 (1998).
43. X. Yang, T. Peppard, 'Solid Phase Microextraction for Flavor Analysis', *J. Agric. Food Chem.*, **42**, 1925–1930 (1994).
44. P.A. Martos, J. Pawliszyn, 'Calibration of Solid Phase Microextraction for Air Analyzes Based on Physical Chemical Properties of the Coating', *Anal. Chem.*, **69**, 206–215 (1997).
45. J. Dugay, C. Miege, M.C. Hennion, 'Effect of Various Parameters Governing Solid-phase Microextraction for the Trace-determination of Pesticides in Water', *J. Chromatogr.*, **795**, 27–42 (1998).
46. J. Ai, 'Solid Phase Microextraction for Quantitative Analysis in Nonequilibrium Situations', *Anal. Chem.*, 1230–1238 (1997).
47. H.H. Jelen, K. Wlazly, R. Wasowicz, E. Kaminski, 'Solid-phase Microextraction for the Analysis of Some Alcohols and Esters in Beer – Comparison with Static Headspace Method', *J. Agric. Food Chem.*, **46**, 1469–1473 (1998).
48. L. Vergnais, F. Masson, M.C. Montel, J.L. Berdague, R. Talon, 'Evaluation of Solid-phase Microextraction for Analysis of Volatile Metabolites Produced by Staphylococci', *J. Agric. Food Chem.*, **46**, 228–234 (1998).
49. S.S. Yang, I. Smetana, 'Determination of Tobacco Alkaloids Using Solid Phase Microextraction and GC/NPD', *Chromatographia*, **47**, 443–448 (1998).
50. A. Steffen, J. Pawliszyn, 'Analysis of Flavor Volatiles Using Headspace Solid-phase Microextraction', *J. Agric. Food Chem.*, **44**, 2187–2193 (1996).
51. Z. Zhang, J. Pawliszyn, 'Headspace Solid-phase Microextraction', *Anal. Chem.*, **65**, 1843–1852 (1993).
52. P. Werkhoff, M. Güntert, G. Krammer, H. Sommer, J. Kaulen, 'Vacuum Headspace Method in Aroma Research: Flavor Chemistry of Yellow Passion Fruits', *J. Agric. Food Chem.*, **46**, 1076–1093 (1998).
53. B.V. Burger, W.J.G. Burger, I. Burger, 'Trace Determination of Volatile Organic Compounds in Water Using Permeation through a Hollow Fiber Membrane and Carrier Gas Stripping', *J. High Resolut. Chromatogr.*, **19**, 571–576 (1996).
54. E.C. Goosens, D. de Jong, G.J. de Jong, U.A.T. Brinkman, 'On-line Sample Treatment/Capillary Gas Chromatography', *Chromatographia*, **47**, 313–345 (1998).

55. M.J. Yang, J. Pawliszyn, 'Membrane Extraction with a Sorbent Interface', *LC/GC Int.*, May, 283–296 (1996).
56. R. Kaiser, 'Trapping, Investigation and Reconstitution of Flower Scents', in *Perfumes*, eds. P.M. Müller, D. Lamparsky, Elsevier Applied Science, London, 213–250, 1991.
57. R.R. Heath, A. Manukian, 'An Automated System for Use in Collecting Volatiles Released from Plants', *J. Chem. Ecol.*, **20**, 593–608 (1994).
58. H.B. Jakobsen, 'The Preisolation Phase of In Situ Headspace Analysis: Method and Perspectives', in *Plant Volatile Analysis*, eds. H.F. Linskens, J.F. Jackson, Springer, Berlin, 1–22, 1997.
59. R.R. Arnts, E.L. Singsaas, T.D. Sharkey, 'Gas Chromatographic Techniques for the Measurement of Isoprene in Air', in *Modern Methods of Plant Analysis*, eds. H.F. Linskens, J.F. Jackson, Springer, Berlin, 47–71, 1997.
60. D. Shelow, 'SilcoCan – The Ideal Canister for Sulfur Compound Storage', *Restek – The Advantage*, 6–7 (1997).
61. 'ATD 400 User's manual', Perkin Elmer, Beaconsfield, UK, 1991.
62. Y. Seto, 'Determination of Volatile Substances in Biological Samples by Headspace Chromatography', *J. Chromatogr.*, **674**, 25–62 (1994).
63. B. McGillivray, J. Pawliszyn, 'Headspace Solid-phase Microextraction Versus Purge and Trap for the Determination of Substituted Benzene Compounds in Water', *J. Chromatogr. Sci.*, **32**, 317–322 (1994).
64. S.E. Ebeler, R.M. Pangborn, W.G. Jennings, 'Influence of Dispersion Medium on Aroma Intensity and Headspace Concentration of Menthone and Isoamyl Acetate', *J. Agric. Food Chem.*, **36**, 791–796 (1988).
65. C. Fischer, U. Fischer, 'Analysis of Cork Taint in Wine and Cork Material at Olfactory Subthreshold Levels by Solid Phase Microextraction', *J. Agric. Food Chem.*, **45**, 1995–1997 (1997).
66. P.N. Bartlett, J.M. Elliott, J.W. Gardner, 'Electronic Noses and their Application in the Food Industry', *Food Technol.*, **51**, 44–48 (1997).
67. C.F. Mandenius, A. Hagman, F. Dunas, H. Sundgren, I. Lundstrom, 'A Multisensor Array for Visualizing Continuous State Transitions in Biopharmaceutical Processes Using Principal Component Analysis', *Biosensors Bioelectronics*, **13**, 193–199 (1998).
68. T. Talou, J.M. Sanchez, B. Bourrounet, 'New Trends in Electronic Noses for Flavour Chemistry', in *Flavour Science*, eds. A.J. Taylor, D.S. Mottram, Royal Society of Chemistry, Cambridge, 277–282, 1996.
69. B. Dittman, S. Nitz, G. Horner, 'A New Chemical Sensor on a Mass Spectrometric Basis', *Adv. Food Sci.*, **20**, 115–121 (1998).
70. W. Lindinger, A. Hansel, A. Jordan, 'On-line Monitoring of Volatile Organic Compounds at pptv Levels by Means of Proton-transfer-reaction Mass Spectrometry (PTR/MS). Medical Applications, Food Control and Environmental Research', *Int. J. Mass Spectrom. Ion Processes*, **173**, 191–241 (1998).
71. R. Zimmermann, H.J. Heger, C. Yeretzian, H. Nagel, U. Boesl, 'Application of Laser Ionisation Mass Spectrometry for On-line Monitoring of Volatiles in the Headspace of Food Products: Roasting and Brewing of Coffee', *Rapid Commun. Mass Spectrom.*, **10**, 1975–1979 (1996).
72. A. Chaintreau, 'Analysis Technology', in *Current Topics in Flavors and Fragrances, Towards a New Millenium of Discovery?*, ed. K.A.D. Swift, Kluwer Academic, Dordrecht, The Netherlands, 1999.
73. C. Bicchi, D. Joulain, 'Headspace-gas Chromatographic Analysis of Medicinal and Aromatic Plants and Flowers', *Flavor and Fragrance J.*, **5**, 131–145 (1990).
74. P. Pollien, A. Ott, F. Montigon, M. Baumgartner, R. Munoz-Box, A. Chaintreau, 'Hyphenated Headspace-Gas Chromatography-sniffing: Screening of Impact Odorants and Quantitative Aromagram Comparisons', *J. Agric. Food Chem.*, **45**, 2630–2637 (1997).
75. C. Bicchi, O. Panero, G.M. Pellegrino, A.C. Vanni, 'Characterization of Roasted Coffee and Coffee Beverages by Solid Phase Microextraction/Gas Chromatography and Principal Component Analysis', *J. Agric. Food Chem.*, **45**, 4680–4686 (1997).
76. J.H. Linssen, L. Rijnen, A. Leggerhuiysman, J.P. Roozen, 'Combined GC and Sniffing Port Analysis of Volatile Compounds in Rubber Rings Mounted on Beer Bottles', *Food Addit. Contam.*, **15**, 79–83 (1998).
77. A. Montano, A.H. Sanchez, L. Rejano, 'Rapid Quantitative Analysis of Headspace Components of Green Olive Brine', *J. Chromatogr.*, **521**, 153–157 (1990).
78. R. Zimmermann, U. Boesl, H.J. Heger, E. Rohwer, E.R. Ortner, E.W. Schlag, 'Hyphenation of Gas Chromatography and Resonance-enhanced Laser Mass Spectrometry (REMPI/TOFMS): a Multidimensional Analytical Technique', *J. High Resolut. Chromatogr.*, **20**, 461–470 (1997).
79. E. Ibañez, R.A. Bernhard, 'Solid-phase Microextraction (SPME) of Pyrazines in Model Reaction Systems', *J. Sci. Food Agr.*, **72**, 91–96 (1996).
80. B. Johnson, 'Headspace Analysis and Shelf Life', *Cereal Food World*, **42**, 752–754 (1997).
81. K. Andersson, H. Lingnert, 'Influence of Oxygen Concentration on the Flavor and Chemical Stability of Cream Powder', *Lebensm.-Wiss. Technol.*, **31**, 245–251 (1998).
82. M. Lennersten, H. Lingnert, 'Influence of Different Packaging Materials on Lipid Oxidation in Potato Crisps Exposed to Fluorescent Light', *Lebensm.-Wiss. Technol.*, **31**, 162–168 (1998).
83. E. Schaller, J.O. Bosset, F. Escher, ' ''Electronic Noses'' and their Application to Food', *Lebensm. -Wiss. Technol.*, **31**, 305–316 (1998).

84. M.F. Wolfe, S. Schwarzbach, R.A. Sulaiman, 'Effect of Mercury on Wildlife – a Comprehensive Review', *Environ. Toxicol Chem.*, **17**, 146–160 (1998).
85. T.P. McNeal, H.C. Hollifield, 'Quantitative Multiresidue Analyses for Volatile Organics in Water and Milk, Using a Fused Silica Open-tubular Wide-bore Capillary Column and Automated Headspace Gas Chromatography', *J. Assoc. Off. Anal. Chem.*, **73**, 328–331 (1990).
86. R.L. Barnes, 'Headspace Analysis of Pharmaceuticals', *Chromatogr. Sci.*, **90**, 149–165 (1990).
87. A. Ott, J.E. Germond, A. Chaintreau, 'Origin of Acetaldehyde during Milk Fermentation Using ^{13}C Labeled Precursors', *J. Agric. Food Chem.*, in press (2000).
88. A. Ott, J.E. Germond, A. Chaintreau, 'Vicinal Diketones in Yogurt: Precursors and Effect of Branched Chain Amino Acids', *J. Agric. Food Chem.*, in press (2000).
89. X.P. Lee, T. Kumazawa, K. Sato, H. Seno, A. Ishii, O. Suzuki, 'Improved Extraction of Ethanol from Human Body Fluids by Headspace Solid-phase Microextraction with a Carboxen–Polydimethylsiloxane-coated Fiber', *Chromatographia*, **47**, 593–595 (1998).
90. T. Kumazawa, H. Seno, X.P. Lee, A. Ishii, O. Suzuki, K. Sato, 'Detection of Ethanol in Human Body Fluids by Headspace Solid-phase Microextraction (SPME)/Capillary Gas Chromatography', *Chromatographia*, **43**, 393–397 (1996).
91. K.D. Pohl, E. Keller, 'Application of Gas Chromatographic Headspace Analysis for Forensic Science', in *Applied Headspace Gas Chromatography*, ed. B. Kolb, Heyden, London, 165–170, 1980.
92. J. Schuberth, 'Joint Use of Retention Index and Mass Spectrum in Post-mortem Tests for Volatile Organics by Headspace Capillary Gas Chromatography with Ion-trap Detection', *J. Chromatogr.*, **674**, 63–71 (1994).
93. R.F. Service, 'Breathanalyzer Device Sniffs for Disease', *Science*, **281**, 1431 (1998).
94. K. Bartle, M. Cikalo, T. Clifford, A. Lewis, P. Myers, M. Robson, H. Vandenburg, 'New Sample Preparation Methods used in GC Pharmaceutical Analysis', *LC/GC Int.*, May, Suppl. S, 32–36 (1998).
95. F. Mathieu, C. Malosse, B. Frerot, 'Identification of the Volatile Components Released by Fresh Coffee Berries at Different Stages of Ripeness', *J. Agric. Food Chem.*, **46**, 1106–1110 (1998).
96. J. Auger, S. Rousset, E. Thibout, B. Jaillais, 'Solid-phase Microextraction–Gas Chromatography–Direct Deposition Infrared Spectrometry as a Convenient Method for the Determination of Volatile Compounds from Living Organisms', *J. Chromatogr.*, **819**, 45–50 (1998).
97. Anonymous, 'Microextraction en phase solide – une méthode de préparation d'échantillons sans solvant pour les composés organiques dans l'eau', Le Supelco Reporter, XII, 3–5, 1993.
98. B.H. Heath, G. Reineccius, 'Flavor and its Study', in *Flavor Chemistry and Technology*, eds. B.H. Heath, G. Reineccius, Macmillan, Basingstoke, 3–42, 1986.
99. P.E. Nelson, J.E. Hoff, 'Food Volatiles: Gas Chromatographic Determination of Partition Coefficients in Water–Lipid Systems', *J. Food Sci.*, **33**, 479–482 (1968).
100. B. Kolb, 'Physicochemical Applications for Headspace Gas Chromatography', in *Applied Headspace Gas Chromatography*, ed. B. Kolb, Heyden, London, 1–11, 1998.
101. A. Voilley, 'Activity Coefficients of Aroma Compounds and Water Activity in Model Systems', *J. Process Eng.*, **8**, 159–170 (1986).
102. C. Neill, K.M. Johnson, E. Lewis, D.W.R. Wallace, 'Accurate Headspace Analysis of f_{CO_2} in Discrete Water Samples Using Batch Equilibration', *Limnol. Oceanogr.*, **42**, 1774–1783 (1997).
103. J.E. Kinsella, 'Flavor Perception and Binding to Food Components', in *Flavor Chemistry of Lipid Foods*, eds. D.B. Min, T.H. Smouse, American Oil Chemists' Society, Champaign, IL, 376–403, 1989.
104. P. Landy, C. Druaux, A. Voilley, 'Retention of Aroma Compounds by Proteins in Aqueous Solution', *Food Chem.*, **54**, 387–392 (1995).

Starch Analysis in Food

Klaus N. Englyst, Geoffrey J. Hudson, and Hans N. Englyst
Englyst Carbohydrate Services, Eastleigh, UK

1 Introduction 1
2 History 4
3 Applications of Starch Measurements 4
3.1 Food Tables and Food Labeling 4
3.2 Public Health 5
4 Methods 5
4.1 Reagents 5
4.2 Apparatus 7
4.3 Sample Preparation 8
4.4 The Colorimetry Version (Figure 1) 8
4.5 The High-performance Liquid Chromatography Version (Figure 2) 10
4.6 Calculations 11
4.7 Measurement of Resistant Starch 1, 2 and 3 (Using Colorimetry or High-performance Liquid Chromatography) 11
5 Quality Control and Troubleshooting 12
5.1 Quality Control 12
5.2 Troubleshooting 13
6 Method Development and Validation 13
6.1 Methodological Considerations 13
6.2 Validation of Rapidly Digestible Starch and Slowly Digestible Starch Values 14
6.3 Validation of Resistant Starch Values 14
7 Comparison with Other Methods 15
7.1 Total Starch 15
7.2 Rate and Extent of Starch Digestion 15
Abbreviations and Acronyms 15
Related Articles 15
References 15

Starch is quantitatively an important component of the human diet, being present in grains, tubers and legumes. Starch has for a long time been considered by many as being slowly but completely digested in the small intestine, resulting in modest glycemic responses and with no physiological role other than as an energy source. It is now understood that in fact the metabolic fate and physiological properties of starch can vary considerably, and both the botanical source and the effects of food processing are major determinants of starch digestibility. In addition to the nature of the starch itself, the site, rate and extent of digestion of starch in the human small intestine are influenced by a number of host factors. The rate at which starch is digested in the human small intestine results in a wide range of glycemic responses, and this physiological measurement has been used to rank foods by their glycemic index. In vitro studies have indicated that glycemic response and the rate of starch digestion are closely correlated. Rapidly digestible starch (RDS) and slowly digestible starch (SDS) fractions together represent the starch that is likely to be digested completely in the human small intestine, with any remaining starch defined as the resistant starch (RS) fraction that is available for fermentation in the large bowel. Measurements of RDS, SDS and RS can be obtained by one simple procedure. Values for the different starch fractions obtained by the in vitro method described here represent reproducible measurements that can be used to classify dietary starch according to its potential digestibility. In addition to these starch fractions, two terms, rapidly available glucose (RAG) and slowly available glucose (SAG), are introduced to reflect the rate at which glucose (from both sugars and starch) is likely to be absorbed in the small intestine.

The proportions of RAG, SAG, RDS, SDS and RS in foods can be controlled by food processing. The implications of altering the rate and extent of starch digestion are potentially of great importance to public health. A full understanding of the links between dietary carbohydrates and health and the underlying mechanisms will come only from the specific measurement of individual types of dietary carbohydrates.

1 INTRODUCTION

Advances in knowledge of the relations between diet and disease have implicated the importance of dietary carbohydrates. A full understanding of the links between dietary carbohydrates and health and the underlying mechanisms will come only from the specific measurement of individual types of dietary carbohydrates. Table 1 shows the overall nutritional classification of plant carbohydrates in the human diet. This classification is based primarily on carbohydrate chemistry, with groupings and sub-divisions related to nutritional and physiological properties.

Starch is composed of D-glucose units and exists as a mixture of two polymers, amylose and amylopectin. Amylose is an essentially linear compound with an average degree of polymerization (DP) of 500 with mainly α-D-(1–4)-glucosidic bonds. It has a helical structure

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Classification of the carbohydrates in plant foods

Class/components	Comments
Sugars:	
Mono- and disaccharides and their alcohols	Physiological response depends on identity and rate of release. Free glucose + glucose from sucrose = FSG
Short-chain carbohydrates:	
Maltodextrins	Measured as RDS
Resistant short-chain carbohydrates (non-digestible oligosaccharides)	Fermented in the large bowel and may stimulate growth of bifidobacteria
Starch:	
RDS	RDS + rapidly released FSG = RAG
SDS	SDS + slowly released FSG = SAG
RS	Escapes digestion in the small intestine
NSP:	
Plant cell-wall NSP	Encapsulate and slow absorption of other nutrients. Marker for naturally high-fiber diets for which health benefits have been shown. Fermented in the large bowel to different extents
Other NSP	Food additives. Minor components of the human diet. Fermented in the large bowel to different extents

NSP, nonstarch polysaccharides; FSG, free-sugar glucose.

with a hydrophobic cavity allowing complexation with hydrophobic molecules such as lipids and iodine (giving the characteristic blue–black color). Amylopectin has a DP in the thousands with α-D-(1–4)-linked chains extensively branched through α-D-(1–6)-glucosidic linkages (about 5% of the total linkages). The ratio of amylose to amylopectin varies considerably between species and has been manipulated by selective breeding to produce both high-amylose and high-amylopectin (waxy) varieties of several cereals.

Starch is the storage polysaccharide of many higher plants, especially cereals, legumes and most tubers. In the plant, starch is stored within the cells as partially crystalline granules. An extensive review of the physical characteristics of starch granules has been given by Gallant et al.[1] The size and shape of the granules and the crystal structure within them are characteristic of the botanical source. Starch granules have been characterized as type A, B or C according to their X-ray diffraction pattern.[2,3] In general, cereal starches have the type A pattern, starches from tubers such as potatoes tend to have the type B pattern, and C-type starch, found in some legumes, is a combination of the A and B patterns. In general, starch granules with B or C type patterns are more resistant to digestion by pancreatic amylase, although the degree of resistance varies with botanical origin.[4] This resistance to hydrolysis affects the digestibility of starchy foods that are normally eaten raw, such as bananas, and of processed foods such as biscuits in which the starch is incompletely gelatinized.

Heating starch granules in a moist environment leads to loss of the crystalline structure, swelling and leaching of amylose molecules from the granules. This process, known as gelatinization, usually occurs in the temperature range 50–70 °C, or at higher temperatures in low-moisture conditions. Cooling of gelatinized starch results in gelation and retrogradation, which is the recrystallization of the starch molecules, particularly amylose, into a form with the type B X-ray pattern.

Starch enclosed within particulate matter or plant cell walls and starch in the form of intact granules is digested slowly, because the physical access of amylolytic enzymes is restricted. Disruption of the particulate matter and removal of plant cell walls during food processing, especially milling, increase the surface area and can damage the structure of starch granules, increasing the susceptibility to enzymatic degradation. Gelatinization tends to increase and retrogradation to reduce starch digestibility. Colonna et al. have provided a detailed account of the influence of food processing on starch granules and their digestion.[5]

The rate at which starch is digested in the human small intestine results in a wide range of glycemic responses, and this physiological measurement has been used to rank foods by their glycemic index.[6] In vitro studies have indicated that glycemic response and the rate of starch digestion are closely correlated.[7–15]

Starch that escapes digestion in the small intestine is termed RS, which, by definition, reaches the large intestine and is a potential substrate for microbial fermentation.[16] The main end-products of this

fermentation are the gases hydrogen and carbon dioxide, and methane in about 50% of people, and short-chain fatty acids, which may be absorbed and utilized.[17]

It is clear from this brief discussion that the metabolic fate and physiological properties of starch can vary considerably, and that the botanical source and the effects of food processing are major determinants of starch digestibility. In addition to the nature of the starch itself, the site, rate and extent of digestion of starch in the human small intestine are influenced by a number of other factors. These include the extent of chewing, the rate of gastric emptying, the transit time of food along the small intestine, the concentration of pancreatic amylase available for breakdown of the starch, the amount of starch consumed and the influence of other food components that may inhibit enzymatic hydrolysis.[18] The rate and extent of digestion of starch in vivo is variable both within and between individuals, which is recognized in the definition of RS as ''the starch (and starch degradation products) that, on average, reaches the human large intestine''.[16] The values for the different starch fractions obtained by the method described here represent reproducible in vitro measurements that can be used to classify dietary starch according to its potential digestibility.

Measurements of RDS, SDS and RS can be obtained by one simple procedure.[16,18,19] RDS and SDS fractions together represent the starch that is likely to be digested completely in the human small intestine, with any remaining starch defined as the RS fraction. Three types of RS have been categorized: physically inaccessible starch (RS1), resistant starch granules (RS2) and retrograded starch (RS3).

In addition to these starch fractions, two terms, RAG and SAG, reflect the rate at which glucose is likely to be absorbed in the small intestine. RAG and SAG may include glucose derived from RDS and SDS, respectively, but the glucose may be derived from sugars. FSG is usually included in the RAG fraction, but may form part of the SAG fraction if trapped within particulate matter or plant cell walls. This ambiguity as to what proportions of RAG and SAG are derived from starch or from FSG complicates the determination of RDS and SDS fractions in foods containing appreciable amounts of FSG. The use of RDS and SDS should therefore be restricted to starchy foods that contain only small amounts of FSG, otherwise RAG and SAG are more appropriate.

Table 2 shows the proportions of the various fractions of starch and the RAG values for a range of foods. These fractions are all expressed in units of grams per 100 g of food ''as eaten''. The legumes have the lowest RAG values, which are associated with low levels of free sugar and a low proportion of starch measuring as RDS. The cause of the slow and incomplete digestion of legume starch is probably a combination of starch granules being encapsulated by cell walls (dietary fiber) and not being fully gelatinized. Spaghetti, macaroni and pearled barley are examples of foods that have a moderate RAG value as the result of a dense structure. For digestive biscuits, a considerable proportion of the starch is measured as SDS, reflecting incomplete gelatinization, but the high dry matter content results in similar RAG values for digestive biscuits and white bread. Cooked potato has a low RDS value, reflecting the low dry matter content, but RDS represents a high proportion of the total starch (TS) content. In general, highly processed breakfast cereals have high RAG values.

Table 2 Starch fractions and RAG content of some starchy foods (grams per 100 g as eaten)

Food	TS	RDS	SDS	RS	RAG
Cereals:					
Pearled barley	17.1	8.0	7.0	2.1	9
Sweetcorn	17.1	15.4	1.4	0.3	18
Bread:					
Rye crispbread	59.8	48.8	6.7	4.3	55
Rye wholemeal	33.8	23.2	7.4	3.2	27
Wheat white	41.7	37.4	3.7	0.6	42
Wheat wholemeal	35.0	32.1	1.4	1.5	36
Biscuits:					
Digestive	46.5	32.0	12.6	1.9	44
Oatmeal	55.9	48.8	6.2	0.9	55
Breakfast cereals:					
All Bran	22.2	20.6	0.5	1.1	35
Oat bran	45.8	31.2	13.6	1.0	36
Porridge Oats	13.0	9.9	3.1	0.1	11
Puffed Wheat	68.7	62.5	0.0	6.2	70
Rice Krispies	69.8	65.6	1.7	2.5	80
Shredded Wheat	62.2	48.7	11.9	1.6	55
Weetabix	57.0	56.8	1.0	0.0	65
Rice:					
Brown, long grain	23.8	14.6	9.2	0.0	16
Parboiled	27.8	16.6	10.0	1.2	19
White, long grain	23.0	17.4	5.6	0.0	19
Pasta:					
Macaroni	26.2	13.4	12.0	0.8	15
White spaghetti	23.5	13.5	9.0	1.0	15
Legumes:					
Butter beans	11.4	9.4	0.8	1.2	11
Chickpea	16.4	5.1	8.8	2.5	6
Frozen peas	7.2	4.1	1.0	2.1	6
Haricot beans	18.2	4.1	5.8	8.3	5
Kidney beans	17.0	4.7	9.8	2.5	6
Pinto beans	16.1	9.3	5.0	1.8	11
Red lentils	15.8	7.3	6.1	2.4	8
Tubers:					
Instant potato	12.7	10.9	1.1	0.8	12
Potato	16.0	15.2	0.7	0.1	17
Potato crisps	50.0	42.7	2.8	4.5	48
Sweet potato	9.3	7.5	0.8	1.1	11
Yam	16.8	14.3	0.4	2.1	18

The proportions of RAG, SAG, RDS, SDS and RS in foods, and thus the expected rate and extent of digestion in the human small intestine, can be controlled by food processing. The implications of altering the rate and extent of starch digestion are potentially of great importance to public health. The identification of specific health benefits related to the ingestion of various types of starch will be possible only when separate measurements of these are available. It is therefore important to be able to measure nutritionally relevant starch fractions in vitro, and such methods are the subject of this article.

2 HISTORY

The classification and measurement of dietary carbohydrates have developed during the 20th century as knowledge of the nutritional importance of this fraction of the diet has accumulated. The perception of dietary carbohydrates as simply an energy source has been echoed in the traditional calculation of ''carbohydrate by difference'', as the portion remaining once moisture, protein, fat and ash are accounted for.

In 1929, McCance and Lawrence highlighted the general lack of available data for the carbohydrate content of foods and the enormous variation that existed between some of the published values.[20] McCance and Lawrence were well aware that the carbohydrate composition of raw and cooked foods can be very different and they identified the limitation of much of the available data, which were nearly all for the composition of raw foods. McCance and Lawrence were primarily concerned with the dietary management of diabetes, and they made the distinction between ''available'' and ''unavailable carbohydrates''. They identified available carbohydrates as starch and the soluble sugars, sucrose, glucose and fructose.

McCance and Lawrence compiled a table of the available carbohydrate content of a range of plant foods, which were analyzed in the form in which they were normally eaten, i.e. cooked or raw as appropriate.[20] They used hydrolysis with hot 7.5% (v/v) hydrochloric acid under a reflux condenser, followed by neutralization with sodium hydroxide, and they measured both reducing sugars and pentoses.

McCance and Lawrence introduced two important principles into the measurement of dietary carbohydrates for nutritional purposes: (1) the need to analyze foods in the form in which they are eaten; and (2) the need to make specific chemical measurements of dietary carbohydrates. The analysis and measurement of dietary carbohydrates in later years has been greatly facilitated by the advent of highly purified enzyme preparations and advances in techniques for the separation and specific measurements of individual sugars.

Increased interest in dietary fiber (plant cell walls) as a potentially protective agent against Western diseases of affluence led to the development of the NSP-type measurement techniques for the dietary carbohydrates that McCance and Lawrence referred to as unavailable. This type of procedure was pioneered by Southgate and later developed by Englyst, who identified a fraction of RS3 that could not be hydrolyzed without prior dispersion with dimethyl sulfoxide (DMSO) or KOH.[21–26] This starch fraction, subsequently identified as RS3, was termed RS. Later, in conjunction with ileostomy studies, two other forms of RS were identified; RS1 and RS2. The definition of RS includes all three fractions, and no starch is included in the measurement of NSP.

There is growing interest in the rate and extent of starch digestion with respect to the magnitude of the glycemic response. Jenkins et al.[9] have pioneered the glycemic index, which may be used to rank foods by their likely effect on the glycemic response. In a classical study, Haber et al.[27] demonstrated the effect of particle size, i.e. degree of food processing, on the rate of absorption of sugars from apples, and several in vitro studies have shown that the rate of starch digestion varies according to its botanical source and degree of processing.[5,10,12,27] Measurement of RAG, SAG and starch fractions as described here represent tools for studying the mechanisms that underlie the links between dietary carbohydrates and health.

3 APPLICATIONS OF STARCH MEASUREMENTS

3.1 Food Tables and Food Labeling

Food tables should supply meaningful, reliable data that are suitable for use by dietitians, epidemiologists and the food industry. The Food and Agriculture Organization (FAO) of the United Nations and the World Health Organization (WHO) have recommended that compositional data for dietary carbohydrates for use in food tables and food labeling should be based on chemically identified components, which may be further subdivided by nutritional properties.[28] The classification of dietary carbohydrates presented in Table 1 conforms with these criteria. The inclusion in food tables of the RDS, SDS and RS subdivisions of starch provides the tools for investigating the health implications of altering the rate and extent of starch digestion. The RAG and SAG terms provide information on the rate at which glucose from starch and sugars is likely to become available for absorption in the human small intestine,

and may be used as a guide to the likely magnitude of the glycemic response.[19] Such values do not become obsolete and can be used in different combinations for different purposes. The use of these values for food labeling will depend on confirmation of their importance for public health.

3.2 Public Health

A review of current knowledge suggests that a diet rich in slowly digested carbohydrates has an overall protective effect against the non-communicable diseases of westernization.[28,29] The results of epidemiological studies have suggested that diets with a high content of rapidly digested carbohydrates from refined foods increase the relative risk of developing diabetes mellitus and coronary heart disease.[29–32] There are indications from studies with low glycemic index diets that the rate of starch digestion is of importance in the glycemic control of people with diabetes.[33]

The impact of RS on health remains uncertain. It has been suggested that the high proportion of butyrate produced during the fermentation of RS in the colon may benefit the gut epithelium and act as a protective agent against colon cancer. However, caution should be applied until the health implications of RS fractions have been assessed, and safe levels of RS intakes have been determined.

4 METHODS

The main analytical procedure described here quantifies the various fractions of starch through enzymatic hydrolysis and measurement of the glucose released. Samples are analyzed ''as eaten'' and are treated with protease to disrupt any starch–protein interaction. This is followed by incubation with amylolytic enzymes under conditions controlled for temperature, pH, viscosity and rate of mechanical mixing. Subsamples are taken at 20 min [the glucose measured in the portion removed from the analysis at 20 min after the addition of enzymes (G20)] and 120 min [the glucose measured in the portion removed from the analysis at 120 min after the addition of enzymes (G120)] as measures of the rate and extent of starch digestion. Any remaining starch is dispersed before enzymatic hydrolysis to give a value for total glucose (TG). A value for RS is calculated as the difference between G120 and TG. Separate values may be obtained for RS1, RS2 and RS3.

Two versions of the main procedure are described: one is based on the colorimetric determination of glucose and the other on the measurement of glucose by high-performance liquid chromatography (HPLC), which includes the use of an internal standard. Figures 1 and 2 summarize the colorimetry and HPLC procedures.

The subsidiary procedure for the measurement of FSG is described, which is required for calculation of RDS and TS, and this procedure may be used to obtain a rapid measurement of TS.

Regarding analytical throughput, a shaking water-bath capable of holding 18 tubes allows the analysis of five test samples in triplicate. Both the main and the subsidiary procedures may be completed within a working day.

4.1 Reagents

4.1.1 Common to the Colorimetry and High-performance Liquid Chromatography Procedures

Distilled water or water of equivalent purity should be used throughout the method. All sugars should be dried to constant weight under reduced pressure with phosphorus pentoxide before use.

Amyloglucosidase EC 3.2.1.3 (Englyst Carbohydrate Services Ltd, Cat. No. 61-002).

Amyloglucosidase Solution Amyloglucosidase diluted 1 : 7 (v/v) with water.

Benzoic Acid, Saturated Prepare a saturated solution of benzoic acid at room temperature.

Enzyme Mixture For 18 samples/standards, into each of six centrifuge tubes weigh 3.0 g of pancreatin and suspend in 20 mL of water using a vortex-mixer. Add a magnetic stirring bar and mix for 10 min. Centrifuge at 1500 *g* for 10 min, then remove 15 mL of the cloudy supernatant from each tube and combine (90 mL total). Add 4 mL of amyloglucosidase and 6 mL of invertase. Mix well. The enzyme mixture should be prepared immediately before use.

Guar Gum Powder (Sigma, Cat No. G-4129).

Heat-stable Amylase EC 3.2.1.1 (Termamyl, Englyst Carbohydrate Services Ltd, Cat. No. 61-005).

Hydrochloric Acid 0.05 mol L^{-1}.

Invertase EC 3.2.1.26 (Merck, Cat. No. 39020).

Pancreatin (Sigma, 8 × USP, Cat. No. P-7545).

Pepsin Powder EC 3.4.23.1 (Sigma, Cat. No. P-7000).

Pepsin–Guar Gum Solution For 18 samples/standards, add 1 g of pepsin powder to 200 mL of 0.05 mol L^{-1} hydrochloric acid and mix with a magnetic stirring

Figure 1 Flow diagram of the steps in the colorimetry version of the procedure.

bar. Just before use, add 1 g of guar gum and mix well. The pepsin–guar gum solution should be prepared immediately before use.

Potassium Hydroxide 7 mol L^{-1}.

Sodium Acetate Buffer 0.1 mol L^{-1}, pH 5.2 Dissolve 13.6 g of sodium acetate trihydrate in water, add 250 mL of saturated benzoic acid solution and make up to 1 L with water. Adjust to pH 5.2 with 0.1 mol L^{-1} acetic acid. To stabilize and activate enzymes, add 4 mL of 1 mol L^{-1} calcium chloride to 1 L of buffer.

4.1.2 For Colorimetry Only

Acetic Acid 0.5 mol L^{-1}. Plus 4 mL L^{-1} of 1 mol L^{-1} calcium chloride.

Ethanol 66% (v/v).

Glucose Oxidase Diagnostic Kit [Boehringer glucose test combination, Cat. No. 166391].

Glucose Standard Weigh 50 g of glucose to the nearest 0.1 mg. Make up to 200 mL with sodium acetate buffer to give a 25 mg mL^{-1} solution.

Sodium Acetate 0.25 mol L^{-1}. Weigh 34 g of sodium acetate trihydrate and make up to 1 L with water.

4.1.3 For the High-performance Liquid Chromatography Procedure Only

Acetic Acid 1 mol L^{-1}. Plus 4 mL L^{-1} of 1 mol L^{-1} calcium chloride.

Ethanol Absolute.

Figure 2 Flow diagram of the steps in the HPLC version of the procedure.

Internal Standard Dissolve 40 g of arabinose in water, add 500 mL of saturated benzoic acid and make up to 1 L with water.

Stock Sugar Mixture Dissolve 50 g of glucose and 25 g of fructose in water, add 500 mL of saturated benzoic acid and make up to 1 L with water.

Sodium Acetate 0.5 mol L^{-1}. Weigh 68 g of sodium acetate trihydrate and make up to 1 L with water.

Sodium Hydroxide 50% (w/v).

4.2 Apparatus

4.2.1 Common to the Colorimetry and High-performance Liquid Chromatography Procedures

Balance Accurate to 0.1 mg.

Centrifuge Capable of exerting 1500 *g*.

Centrifuge Tubes Polypropylene (Copolymer) centrifuge tubes of length 11.5 cm and diameter 3 cm (50-mL capacity), with screw caps (Alpha Laboratories, Cat. No. LW 1110).

Glass Balls (Marbles) Approximately 1.5 cm in diameter, sufficient to provide five balls per sample.

Magnetic Stirrer

Mincer Hand-driven, with a plate with 0.9-cm diameter holes.

Test Tubes Glass or plastic, of 15- and 30-mL capacity, preferably with lids, capable of withstanding low-speed centrifugation.

Vortex-mixer

Water-bath Capable of maintaining 100 °C.

Water-baths Capable of maintaining temperatures in the range 35–70 °C. This or another similar bath must have a linear shaking capacity of not less than 160 strokes min^{-1} and a stroke length of approximately 35 mm. The shaking bath must be fitted with the means to hold all the centrifuge tubes exactly horizontally under the water, with the long axis of each tube exactly parallel with the direction of movement. Each bath should be of sufficient capacity that there is no significant change in temperature when a rack containing all the tubes is placed in it.

4.2.2 For the Colorimetry Procedure Only

Spectrophotometer

4.2.3 For the High-performance Liquid Chromatography Procedure Only

Autoinjector

Ion-exchange Column with Guard Column For separation of sugars.

Ion-exchange Column For trapping amino acids and peptides.

Column-switching Mechanism

Gradient Pump

Electrochemical Detector

Computerized Data-handling System

4.3 Sample Preparation

Starch digestibility is greatly influenced by food processing and samples must be analyzed as eaten. Foods normally eaten dry are analyzed dry, and foods normally eaten hot and wet are cooked and maintained at 70–80 °C immediately before analysis. Foods with a recognizable structure that would normally require chewing (e.g. pasta, rice, maize) are passed through the mincer. For determination of FSG and the rapid measurement of TS, samples may be milled or homogenized. Determine the moisture content of samples as weight loss after overnight incubation at 104 °C.

Sample weights should be chosen to contain 500–600 mg of starch and sugars, which can be estimated from food tables. Examples of suitable sample weights are given in Table 3.

The procedure should now continue, using either the colorimetry or the HPLC version.

4.4 The Colorimetry Version (Figure 1)

4.4.1 Measurement of Rapidly Available Glucose, Slowly Digestible Starch, Resistant Starch and Total Starch

Preparation of the Standard and Blank Into each of three centrifuge tubes weigh 50 mg of guar gum powder and add five glass balls. Pipet 20 mL of glucose standard into each of two of these and 20 mL of acetate buffer into the other (blank). Cap and shake well to disperse the gum. Treat these tubes exactly as the samples after the addition of sodium acetate.

Sample Weight Weigh an appropriate amount of sample, to the nearest 0.1 mg, into 50-mL tubes. If the sample is hot, cap the tube and place in a water-bath at 70–80 °C until ready to analyze. Samples should be analyzed at least in duplicate. Reference samples should be included in every batch (see section 5).

Incubation with Pepsin Put 10 mL of the freshly prepared pepsin–guar gum solution into each sample tube, vortex-mix and place the tubes in a water-bath at 37 °C. After 30 min of incubation, remove the samples from the water-bath and add five glass balls and 10 mL of 0.25 mol L^{-1} sodium acetate to each tube. Shake the tubes gently to disperse the contents and replace the sample tubes, standards and blanks in the water-bath at 37 °C to equilibrate.

Enzymatic Hydrolysis of Starch Remove one sample tube from the 37 °C water-bath and add 5 mL of enzyme mixture. Immediately cap the tube and mix the contents gently by inversion before securing the tube horizontally in the 37 °C shaking water-bath. Start the shaking action of the water-bath; this is time zero for the incubation

Table 3 Examples of suitable sample weights

Dry matter (%)	Examples	Weight (g)
75–100	Starch	0.6
75–100	Flours, breakfast cereals	0.8
55–75	Bread, cakes	1.0–2
35–55	Beans, pasta, rice	1.5–3
15–35	Canned foods, sauces	3.0–4

and the shaking action is not interrupted until all the G120 portions have been collected (see below). Repeat the addition of enzyme mixture for the rest of the sample tubes, at 1-min intervals to aid timing of the procedure, and place them into the shaking water-bath. Remove each tube from the bath at exactly 20 min after addition of the enzyme mixture, transfer 0.5 mL into 20 mL of 66% ethanol and vortex-mix to stop the hydrolysis; this is the G20 portion. Return the sample tube to the shaking water-bath immediately after the sample has been taken. After a further 100 min (a total of 120 min of incubation), transfer 0.5 mL from the sample tube into 20 mL of 66% ethanol and vortex-mix; this is the G120 portion.

Dispersion of Resistant Starch Having removed all the tubes from the shaking water-bath, vortex-mix vigorously to break up any large particles, then place the rack of tubes into a boiling water-bath for 30 min. Remove the tubes from the bath, vortex-mix and place the tubes in ice–water until thoroughly chilled. Add 10 mL of 7 mol L^{-1} KOH, cap the tube and mix the contents by inversion. Immerse the tubes horizontally in a shaking water-bath containing ice–water for 30 min.

Hydrolysis of Starch to Glucose Remove the samples from the waterbath and immediately transfer 1 mL of the contents into a 50-mL tube containing 10 mL of 0.5-mol L^{-1} acetic acid and mix well. Add 0.2 mL of amyloglucosidase solution to each sample. Mix and place in a water-bath at 70 °C. After 30 min, remove the rack of tubes to a boiling water-bath for 10 min. Cool the tubes to room temperature, add 40 mL of water, cap and mix; this is the TG portion to be measured in step 4.4.3.

4.4.2 Measurement of Free-sugar Glucose; Optional Rapid Measurement of Total Starch

Preparation of the Standard and Blank Into each of two 50-mL tubes, pipet 25 mL of glucose standard and into a third add 25 mL of sodium acetate buffer (blank). Treat these tubes exactly as the samples.

Pretreatment of the Samples Samples may be analyzed in the form in which they are eaten, as in the main procedure. Alternatively, samples may be finely divided by milling or homogenization before analysis, in which case the glass balls may be replaced by a magnetic stirring bar and mixing of the sample achieved with a magnetic stirrer instead of a shaking water-bath.

Measurement of Free-sugar Glucose Weigh an appropriate amount of sample, to the nearest 0.1 mg, into 50-mL tubes and add five glass balls to each. Pipet 25 mL of 0.1-mol L^{-1} sodium acetate buffer into each tube, cap and vortex-mix vigorously to begin breaking up large particles. Place the rack of tubes into a boiling water-bath. After 30 min remove the rack of tubes from the boiling water-bath and vortex-mix vigorously to break up any remaining particles of sample. Cool to 37 °C, then add 0.2 mL of invertase to each tube. Cap the tubes, immerse horizontally in the shaking water-bath at 37 °C and incubate for 30 min.

Vortex-mix vigorously before transferring 1 mL of each sample or standard solution into a test tube containing 2 mL of absolute ethanol and vortex-mix. Centrifuge at 500 *g* for 5 min, then transfer 1 mL of the supernatant into 5 mL of water and mix well by inversion. (For the standard, transfer 1 mL of supernatant into 20 mL of water.) This is the FSG portion to be measured in step 4.4.3.

If TS is to be determined by the main method, the procedure is finished; otherwise, the procedure can be continued for a rapid measurement of TS.

Measurement of Total Glucose To the remainder of the sample solution add 0.1 mL of heat-stable amylase, cap the tube, vortex-mix vigorously, then place in a boiling water-bath for 15 min. Remove the tubes from the bath, vortex-mix and then place the rack of tubes in ice–water and leave them to chill thoroughly.

Add 10 mL of 7-mol L^{-1} KOH, cap the tube and mix the contents by inversion. Immerse the tubes horizontally in a shaking water-bath containing ice–water for 30 min.

Remove the tubes from the water-bath and immediately transfer 1 mL into a 50-mL tube containing 10 mL of 0.5-mol L^{-1} acetic acid and mix well. Add 0.2 mL of amyloglucosidase solution to each sample. Mix and place in a water-bath at 70 °C. After 30 min remove the rack of tubes to a boiling water-bath for 10 min. Cool the tubes to room temperature, add 40 mL of water, cap and mix; this is the TG portion.

4.4.3 Determination of Glucose Using Glucose Oxidase

Centrifuge the samples at 500 *g* for 5 min to pellet any precipitate before determination of glucose.

Pipet 100 µL of water, enzyme blank, samples and standards in duplicate into test tubes containing 2 mL of the glucose oxidase reagent (prepared according to the manufacturer's instructions) and vortex-mix. Place the tubes in a water-bath at 37 °C for 20 min. Measure the absorbance of the standards and samples in the spectrophotometer at 510 nm against the reagent blank. The glucose concentration (grams per 100 g of sample) is

Table 4 Values of V and C for the different analytes

Analyte	V (mL)[a]	C (mg mL^{-1})
G20	25	20
G120	25	20
TG	35.4	14.12
FSG	25	7.1
TG (rapid)	35.7	17.5

[a] +1 mL per 1 g of sample.

given by Equation (1):

$$\text{Glucose}(\%) = \frac{A(\text{t})VC}{A(\text{s})W} \quad (1)$$

where A(t) is the absorbance of the test solution, V is the total volume of the test solution (milliliters), C is the concentration (milligrams per milliliter) of the standard used, A(s) is the absorbance of the standard used and W is the weight (milligrams) of sample taken for analysis (this figure may be corrected for moisture), as illustrated in Table 4.

4.5 The High-performance Liquid Chromatography Version (Figure 2)

4.5.1 Measurement of Rapidly Available Glucose, Slowly Digestible Starch, Resistant Starch and Total Starch

Sample Weight and Internal Standard and Blanks Weigh an appropriate amount of sample, to the nearest 0.1 mg, into 50-mL tubes. If the sample is hot, cap the tube and place in a water-bath at 70–80 °C until ready to analyze. Samples should be analyzed at least in duplicate. Reference samples should be included in every batch (see section 5). Two tubes, one with 4 mL of stock sugar mixture added, are included as enzyme blanks. Add 5 mL of internal standard to the sample and blank tubes and treat exactly as the samples.

Incubation with Pepsin Add 10 mL of the freshly prepared pepsin–guar gum solution to each tube, vortex-mix and place the tubes in a water-bath at 37 °C for 30 min. Remove the samples from the water-bath and add five glass balls and 5 mL of 0.5-mol L^{-1} sodium acetate to each tube. Shake the tubes gently to disperse the contents and replace them in the water-bath at 37 °C to equilibrate.

Enzymatic Hydrolysis of Starch Remove one sample tube from the 37 °C water-bath and add 5 mL of enzyme mixture. Immediately cap the tube and mix the contents gently by inversion before securing the tube horizontally in the 37 °C shaking water-bath. Start the shaking action of the water-bath; this is time zero for the incubation and the shaking action is not interrupted until all the G120 portions have been collected (see below). Repeat the addition of enzyme mixture for the rest of the sample tubes, at 1-min intervals to aid timing of the procedure, and place them into the shaking water-bath. Remove each tube from the bath at exactly 20 min after addition of the enzyme mixture, transfer 0.2 mL into 4 mL of absolute ethanol and vortex-mix to stop the hydrolysis; this is the G20 portion to be measured in step 4.5.3. Return the sample tube to the shaking water-bath immediately after the sample has been taken. After a further 100 min (a total of 120 min of incubation), transfer 0.2 mL from the sample tube into 4 mL of absolute ethanol and vortex-mix; this is the G120 portion to be measured in step 4.5.3.

Dispersion of Resistant Starch Having removed all the tubes from the shaking water-bath, vortex-mix vigorously to break up any large particles, then place the rack of tubes in a boiling water-bath for 30 min. Remove the tubes from the bath, vortex-mix and place the tubes in ice–water until thoroughly chilled. Add 10 mL of 7-mol L^{-1} KOH, cap the tube and mix the contents by inversion. Immerse the tubes horizontally in a shaking water-bath containing ice–water for 30 min.

Hydrolysis of Starch to Glucose Remove the tubes singly from the ice–water and transfer 0.2 mL of the contents into a tube containing 1 mL of 1-mol L^{-1} acetic acid. To these tubes add 40 μL of amyloglucosidase solution. Mix and place the tubes in a 70 °C water-bath for 30 min followed by 10 min in a boiling water-bath. Cool the tubes to room temperature before adding 12 mL of absolute ethanol; this is the TG portion to be measured in step 4.5.3.

4.5.2 Measurement of Free-sugar Glucose; Optional Rapid Measurement of Total Starch

Preparation of Enzyme Blanks No enzyme blank is required for the measurement of FSG. For the rapid measurement of TS, two tubes, one with 4 mL of the stock sugar mixture added, are included as enzyme blanks.

Pretreatment of the Samples Samples may be analyzed in the form in which they are eaten, as in the main procedure. Alternatively, samples may be finely divided by milling or homogenization before analysis, in which case the glass balls may be replaced by a magnetic stirring bar and mixing of the sample achieved with a magnetic stirrer instead of a shaking water-bath.

Measurement of Free-sugar Glucose Weigh an appropriate amount of sample, to the nearest 0.1 mg, into 50-mL tubes and add five glass balls to each. Add 5 mL of internal standard to each sample tube. Pipet 20 mL of 0.1-mol L^{-1}

acetate buffer into each tube, cap and vortex-mix vigorously in order to begin breaking up large particles. Place the rack of tubes in a boiling water-bath. After 30 min, remove the rack of tubes from the boiling water-bath and vortex-mix vigorously to break up any remaining particles of sample. Cool to 37 °C, then add 0.2 mL of invertase to each tube. Cap the tubes and immerse horizontally in the shaking water-bath at 37 °C for 30 min.

Vortex-mix vigorously before removing 0.2 mL of each sample or standard solution into a test tube containing 4 mL of absolute ethanol; this is the FSG portion to be measured in step 4.5.3.

If TS is to be determined by the main method, the procedure is finished here; otherwise, the procedure can be continued for a rapid measurement of TS.

Dispersion and Hydrolysis of Total Starch To the remainder of the sample solution add 0.1 mL of heat-stable amylase, cap the tube, vortex-mix vigorously, then place in a boiling water-bath for 15 min. Remove the tubes from the bath, vortex-mix and place the tubes in ice–water until thoroughly chilled. Add 10 mL of 7-mol L^{-1} KOH, cap the tube and mix the contents by inversion. Immerse horizontally in a shaking water-bath containing ice–water for 30 min.

Remove the samples from the shaking water-bath and immediately transfer 0.2 mL of the contents into a tube containing 1 mL of 1-mol L^{-1} acetic acid and mix well. Add 40 μL of amyloglucosidase solution. Mix and place the tubes in a 70 °C water-bath for 30 min followed by 10 min in a boiling water-bath. Cool the tubes to room temperature before adding 12 mL of absolute ethanol; this is the TG portion to be measured in step 4.5.3.

4.5.3 High-performance Liquid Chromatography Analysis of Sugars

Prepare two sugar standards in 50-mL tubes: standard 1, 1 mL of stock sugar mixture, 19 mL of water and 5 mL of internal standard; standard 2, 10 mL of stock sugar mixture, 10 mL of water and 5 mL of internal standard. Mix well and transfer 0.2 mL from each into a tube containing 4 mL of absolute ethanol.

Before HPLC analysis, centrifuge all the ethanolic fractions for 5 min at 500 *g*. The amount to be taken for analysis varies according to sugar content: typically 70 μL for the sugar standards and the G20 and G120 portions, 200 μL for the TG portions and 70–120 μL for the FSG portions. Place the samples in HPLC vials, add 1 mL of deionized water and vortex-mix.

A Dionex AS3500 autoinjector may be used for injection of 20 μL of the diluted ethanolic fractions. Sugar separation may be achieved with a Dionex Carbopac PA100 column using a Dionex GP40 gradient pump and a Dionex Carbopac PA10 guard column. Column switching and a Dionex Aminotrap may be used to prevent pancreatic peptides from reaching the analytical column. The eluents, high-purity water and 200 mol L^{-1} NaOH [16 mL L^{-1} of 50% (w/v) NaOH solution in high-purity degassed water] should be degassed. The flow-rate is 0.8 mL min^{-1} and the elution conditions are shown in Table 5. Monosaccharide detection may be achieved with a Dionex ED40 electrochemical detector with the following pulse potentials, E, and durations, t: E_1 0.05 V, t_1 400 ms; E_2 0.75 V, t_2 200 ms; and E_3 0.15 V, t_3 400 ms. A response time of 1 s is used and the output on the detector is set at 300 nA.

Table 5 Sequence of elution conditions for the HPLC measurement of sugars[a]

Switch position	[NaOH] (mmol L^{-1})	Time (min)
A	10	0–3.5
B	70	3.6–14.0
B	200	14.1–15.0
B	10	15.1–20.0

[a] In switch position A, the flow is from Aminotrap to guard column to separation column. In switch position B, the flow is from guard column to separation column to Aminotrap. Sample injection is at time 0.1 min.

4.6 Calculations

Values for RAG, SAG, RDS, SDS, RS and TS are calculated from the measured FSG, G20, G120 and TG values according to Equations (2–7). Values for the starch fractions are expressed as polysaccharides, using a factor of 0.9 to convert the measured glucose value to polysaccharide.

$$\text{RAG} = \text{G20} \tag{2}$$

$$\text{SAG} = \text{G120} - \text{G20} \tag{3}$$

$$\text{RDS} = 0.9(\text{G20} - \text{FSG}) \tag{4}$$

$$\text{SDS} = 0.9(\text{G120} - \text{G20}) \tag{5}$$

$$\text{TS} = 0.9(\text{TG} - \text{FSG}) \tag{6}$$

$$\text{RS} = 0.9(\text{TG} - \text{G120}) \tag{7}$$

4.7 Measurement of Resistant Starch 1, 2 and 3 (Using Colorimetry or High-performance Liquid Chromatography)

4.7.1 Resistant Starch 1, Physically Inaccessible Starch

There are three forms of RS: RS1 is present in foods having a dense or rigid structure, e.g. boiled rice, pasta,

whole-grain bread, maize and legumes; RS2 is present in raw foods, e.g. bananas, and raw cereal flours and grains, in foods cooked with very little water, e.g. some dry-baked biscuits, and in legumes; RS3 is present in foods that have been cooked and then cooled, e.g. bread, breakfast cereals and cold potatoes. (Note: for a few high-amylose products, e.g. Hylon VII, the boiling step in the procedure will not gelatinize all of the starch. Any starch that remains ungelatinized after this step will therefore be included as RS3.)

A value for total RS is obtained by the main procedure. Values for the individual RS1, RS2 and RS3 fractions may be obtained by the parallel analysis of three samples of the test food, which are prepared and analyzed in different ways.

Sample A is prepared in the form in which the food would normally be eaten, and is then passed through a hand-operated mincer. The analytical sample is weighed into 50-mL tubes and treated as described for the main procedure. Only the G120(A) value is required for calculation of the RS1 fraction but it is recommended that G20(A) and TG(A) are measured during the procedure and the values used for calculation of RAG, SAG and total RS.

Sample B is homogenized for moist foods or milled for dry foods to disrupt the food matrix before analysis but ball milling should not be used, as this may damage starch granules. The analytical sample is weighed into 50-mL tubes and treated as described for the main procedure. Only the G120(B) value is required for calculation of the RS2 fraction.

Sample C is homogenized or milled as described for sample B. The analytical sample is weighed into 50-mL tubes and treated as described for the main procedure up to and including the addition of sodium acetate. At this stage, the tubes are placed in boiling water for 30 min to gelatinize any native starch granules that would normally be gelatinized during cooking. The tubes are cooled to 37 °C and the remainder of the main procedure is followed to obtain the G120(C) value that is required for calculation of the RS2 and RS3 fractions. The analysis is continued to obtain a TG(C) value, which is required for calculation of the RS3 fraction. The step in the boiling water-bath in the main procedure after the G120(C) subsample has been withdrawn may be omitted, instead continuing with the dispersion of RS3 by treatment with KOH.

Calculations are performed using Equations (8–10).

$$\mathrm{RS1} = 0.9[\mathrm{G120(B)} - \mathrm{G120(A)}] \quad (8)$$

$$\mathrm{RS2} = 0.9[\mathrm{G120(C)} - \mathrm{G120(B)}] \quad (9)$$

$$\mathrm{RS3} = 0.9[\mathrm{TG(C)} - \mathrm{G120(C)}] \quad (10)$$

5 QUALITY CONTROL AND TROUBLESHOOTING

5.1 Quality Control

For reproducible results to be obtained within and between laboratories, it is essential that stringent quality control is applied. The procedures described here are based on the measurement of the rate and extent of starch digestion in vitro and require standardized enzyme preparations and incubation conditions. A kit is available (Englyst Carbohydrate Services Ltd, Cat. No. 61-000), which includes the necessary enzymes and three reference samples for calibration purposes. It is suggested that reference samples 1 and 2 are included in every batch of samples analyzed. The reference samples have been selected for their specific starch digestion profiles.

Reference sample 1 (white wheat flour; Homepride, UK), with its high SDS content, is used to check the efficiency of starch hydrolysis during the amylolytic incubation. Values for G120 that are too low (see target values below) may indicate that the activity of the amylolytic enzymes has decreased. Reference sample 2 (raw potato starch; Kartoffelmel Centralen, Denmark), with its high RS content, is used to establish the optimum stroke speed of the shaking water-bath during the amylolytic incubation. If the G120 value for the potato starch is too high (see target values in Table 6), decrease the stroke speed and vice versa, e.g. we observe a linear increase in the G120 value for reference sample 2 from 12.4 to 36.4 g per 100 g between 120 and 170 strokes min^{-1}.

Reference sample 3 (corn flakes; Kellogg's, UK), with its high RDS content, is used to check the efficiency of starch and maltose hydrolysis during the first stage of amylolytic incubation. Values for G20 that are too low (see target values below) may indicate that the amyloglucosidase activity has decreased.

Target values for the three reference samples are given in Table 6.

The method yields coefficients of variation of 3.7% intraassay and 6.6% interassay for the measurement of RAG, 2.2% and 3.9% for G120 (RAG + SAG) and 1.6% and 2.5% for TG for the three reference samples.

Table 6 Target values (grams per 100 g) for the reference samples

Reference sample	G20	G120	TG
Potato starch	4 (±0.5)	26 (±1)	89 (±1)
Wheat flour	35 (±1)	77 (±1)	78 (±1)
Corn flakes	79 (±1)	81 (±1)	85 (±1)

5.2 Troubleshooting

5.2.1 Troubleshooting for the Common Hydrolysis Steps

Accurate timing during subsampling for G20 and G120 measurements is critical. Pitted glass balls may increase the mechanical disruption of the samples, and should be replaced. Failure to ensure that tubes are held exactly horizontal in the shaking water-bath and exactly parallel with the direction of movement may result in differences in mechanical disruption of samples.

Variation in Reference Sample Values Between Batches Use the reference samples to check enzyme activity levels. Consistently low results may indicate incorrect pH during hydrolysis: check the buffer system. Variation in G120 values between batches for reference sample 2 may be due to changes in the shaking speed or the temperature of the water-bath. It is recommended that the shaking speed and temperature controls are physically fixed once optimized to the target values.

Variation in Replicate Analysis The sample heterogeneity of many foods analyzed ''as eaten'' can make it difficult to take fully representative subsamples, leading to greater variation in results than for homogenized or milled samples. It is recommended that an increased number of replicates are analyzed for such problem samples.

5.2.2 Troubleshooting for the Colorimetric Procedure

Variation in Replicate Analysis This may be due to inaccurate pipeting when taking subsamples for measurement of G20, G120 and TG, due to the presence of large food particles in the sample and/or the high viscosity of the suspension. The reproducibility of pipeting may be improved by cutting off the end of the pipet tip to produce a larger orifice and by the repeated washing of the pipet tip (after wiping the outside) into the ethanol fraction to ensure complete transfer of the subsample.

Variation in Absorbance Measurements Accurate pipeting is essential, and may be improved by repeatedly washing the tip (after wiping the outside) into the color reagent to ensure complete transfer of the subsample. Ensure that each set of samples is analyzed with suitable standards and blanks, and that the interval between the addition of sample to the color reagent and the measurement of the absorbance is identical for all the samples, standards and blanks.

5.2.3 Troubleshooting for High-performance Liquid Chromatography Procedure

Variation Between Replicates This may be caused by inaccurate pipeting during the addition of internal standard to the samples.

Spurious Peaks on the Chromatogram Peaks that significantly interfere with the integration of sugar peaks may be due to amino acids or peptides in the sample that have eluted on to the separator column from the guard column intended for their retention. Clean the guard column or alter the timing of the column-switching mechanism.

6 METHOD DEVELOPMENT AND VALIDATION

The measurements of starch fractions obtained by the procedures described here characterize starchy foods in terms that reflect the average rate and extent of starch digestion in the human gut. Owing to the variation in both the rate and extent of starch digestion that is seen both within and between individuals in human studies, the measurement of RS is based on the average of measurements observed in vivo. The measurement of RAG, which includes RDS but is not limited to starch, has been shown to be correlated with in vivo glycemic response measurements.[19]

6.1 Methodological Considerations

This section provides a brief discussion of the development of methodology for the measurement of nutritionally important starch fractions. A full description has been given in the papers cited below.

For foods with a well-defined particulate structure, the rate and extent of starch digestion are critically dependent on the way the food sample is divided and the method chosen should reflect the average disruption of food structure achieved by chewing. Comparison of the starch digestibility in food samples that were chewed before analysis and those minced as described here showed that, although both gave similar mean values, the values for the minced samples had a smaller standard deviation.[18]

The addition of glass balls to the incubation tubes provides further mechanical disruption of particulate matter during the hydrolysis with pancreatic amylase. This allowed the determination of in vitro RS values for polished rice commensurate with the starch recovered in ileostomy studies with similar products. To prevent excessive grinding of starch granules by the glass balls, the viscosity of the sample mixture was increased by the addition of guar

gum, which keeps the sample material in suspension. The amounts of the enzymes used in this procedure are calculated to be present in excess so that they are not rate-limiting in the hydrolysis of starch to glucose.

Three approaches are used to disperse the starch fraction that resists hydrolysis during the incubation with pancreatic α-amylase. Physical disruption of particulate material (RS1) by vigorous vortex-mixing at several stages is aided by the presence of the glass balls. A 30 min boiling step is used to gelatinize starch granules (RS2). Retrograded amylose (RS3) is dispersed by 30 min treatment with 2-M KOH, at 0 °C to avoid the destruction of sugars that can occur at higher temperatures. These treatments in combination have been shown to totally disperse starch in all but the hardest whole grains, which may require milling to obtain a reliable value for TS.

6.2 Validation of Rapidly Digestible Starch and Slowly Digestible Starch Values

In the procedure described here, the analytical time points of 20 and 120 min for the measurement of RDS and SDS are not intended to represent that time of passage through the human gut. The choice of 20 min of incubation for the measurement of RDS is based on observed rates of hydrolysis of starch in vitro. For foods in which starch is rapidly digested, such as bread and corn flakes, the majority of starch is digested by 20 min under the conditions described here, whereas a considerable amount of starch in foods such as spaghetti and pearled barley is digested between 20 and 120 min.

A recent study has shown a good correlation between glycemic index values from the literature and in vitro RAG values, which include RDS, and RAG is proposed as an important food-related determinant of the glycemic response.[34]

More direct evidence for this was obtained when the glycemic response to four starchy foods with different proportions of RDS and SDS was investigated. Two of the foods, pearled barley and spaghetti, have a low proportion of starch as RDS, whereas corn flakes and white bread contain a high proportion of starch as RDS. The four foods were each fed in two portion sizes, providing 25 and 50 g of glucose (from RDS, SDS and FSG) likely to be available for absorption in the small intestine. The eight test meals represented a range of intake from 11 to 49 g of RAG. There was a very strong correlation ($P < 0.001$) between RAG intake and glycemic response, with RAG explaining 70% of the variance in glycemic response after subject variation was accounted for (Figure 3).[19]

The results of this study and the good agreement between RAG and glycemic index strongly suggest that

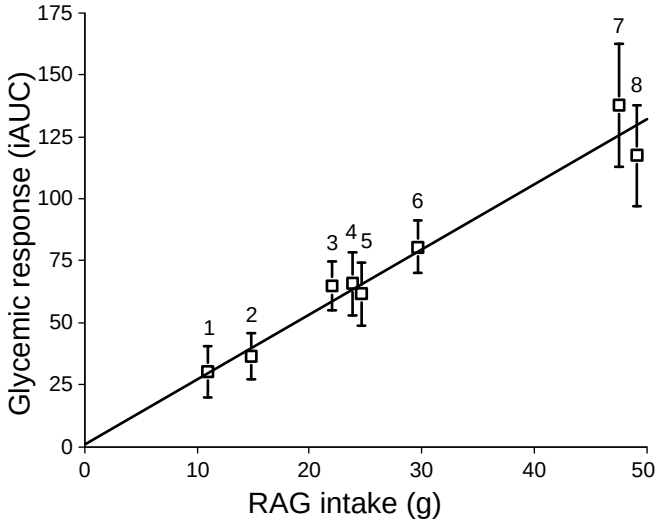

Figure 3 Correlation of glycemic response and RAG content of test meals. The line is fit by least-squares analysis to show the correlation between the mean glycemic response, measured as the incremental area under the curve (iAUC; error bars indicate ± standard error of the mean), and the eight intakes of RAG over all subjects ($n = 8$). The numerals indicate small and large portions, respectively, of the different test meals: 1 and 3, pearled barley; 2 and 6, spaghetti; 4 and 7, corn flakes; 5 and 8, white bread.

measurements of RAG and SAG, and the corresponding starch fractions, RDS and SDS, have physiological significance.[19]

6.3 Validation of Resistant Starch Values

Quantifying the amount of starch entering the human colon presents a considerable technical problem. Hydrogen excreted in breath has been used as an indicator of fermentation in the colon, but it is considered to lack the sensitivity required for quantification. Intubation has been used to sample digesta from the ileum, but these studies were restricted to liquidized meals. Several studies have used human ileostomy subjects as a model to investigate the digestive physiology of the small intestine. Carbohydrate analysis of the ileostomy effluent allows determination of the amount of starch and starch digestion products escaping digestion in the small intestine.[35–37] The amount of starch recovered in the ileostomy effluent varied between individuals by ±20% around the mean. It is from these studies that the currently accepted definition of RS is derived "the sum of starch and starch degradation products that, on average, reach the human large intestine".[16] The in vitro methodology has been tuned to yield values for RS that match the mean proportion of starch recovered in ileostomy studies for both single foods and mixed meals.[14,35–38]

7 COMPARISON WITH OTHER METHODS

7.1 Total Starch

A discussion of the principles of starch analysis has been provided by Southgate.[39] Most methods for the measurement of TS are dependent on acid or enzymatic hydrolysis of starch and measurement of the released glucose. Generally, enzymatic approaches are more specific and do not result in destruction of sugars. Samples are usually finely divided and may be boiled to gelatinize starch before enzymatic hydrolysis with α-amylase and/or amyloglucosidase. However, retrograded amylose is not accessible to enzymatic hydrolysis before it has been dispersed with DMSO or 2-M KOH.[22] Procedures that do not contain an efficient dispersal step will underestimate the starch content of some processed foods containing RS3. The values obtained by such methods will therefore be lower than those obtained by the method described here.

7.2 Rate and Extent of Starch Digestion

The increased knowledge of the different metabolic fates of dietary starch has resulted in the development of several methods for assessing the rate of starch digestion.[8–13,15,18,22,40–43]

The basis of the methods investigating the rate of starch digestion is the measurement of starch hydrolysis after a specified period of incubation with amylolytic enzymes. Meaningful comparison between methods is difficult, owing to differences in sample preparation, enzymes and hydrolysis conditions. However, the ranking between foods is often similar between methods, and most have shown strong correlations with glycemic response data. The RDS and SDS terms described here are at present the only measures of rate of starch digestion expressed in grams per 100 g of food.

The basis of methods for the determination of RS is the measurement of starch remaining unhydrolyzed after a defined period of incubation with amylolytic enzymes. Measurements of total RS must include RS1, RS2 and RS3. However, owing to inappropriate sample preparation and gelatinization of starch during the procedure, many methods measure only RS3 or RS2 and RS3 fractions. In addition, few of the methods have been developed and validated in conjunction with in vivo studies.

Methods designed to investigate the rate and/or the extent of starch digestion in the human gut should incorporate the following principles:

- analysis of foods prepared ''as eaten'';
- reproducible disruption of the physical structure of food;
- standardized conditions of amylolytic hydrolysis;
- development and validation of methods in conjunction with in vivo studies.

The method described here is based on these principles, and incorporates in vitro measurements that reflect the likely rate and extent of starch digestion in the human gut.

ABBREVIATIONS AND ACRONYMS

DMSO	Dimethyl Sulfoxide
FAO	Food and Agriculture Organization
FSG	Free-sugar Glucose
G20	The Glucose Measured in the Portion Removed from the Analysis at 20 min After the Addition of Enzymes
G120	The Glucose Measured in the Portion Removed from the Analysis at 120 min After the Addition of Enzymes
HPLC	High-performance Liquid Chromatography
NSP	Nonstarch Polysaccharides
RAG	Rapidly Available Glucose
RDS	Rapidly Digestible Starch
RS	Resistant Starch
RS1	Physically Inaccessible Starch
RS2	Resistant Starch Granules
RS3	Retrograded Starch
SAG	Slowly Available Glucose
SDS	Slowly Digestible Starch
TG	Total Glucose
TS	Total Starch
WHO	World Health Organization

RELATED ARTICLES

Food **(Volume 5)**
Dietary Fiber Analysis as Non-starch Polysaccharides

REFERENCES

1. D.J. Gallant, B. Bouchet, A. Buleon, S. Perez, 'Physical Characteristics of Starch Granules and Susceptibility to Enzymatic Degradation', *Eur. J. Clin. Nutr.*, **46**, Suppl. 2, S3–S16 (1992).
2. J.R. Katz, 'The Amorphous Part of Starch in Fresh Bread, and in Fresh Pastes and Solutions of Starch', *Recl. Trav. Chim. Pays Bas*, **18**, 55–59 (1937).
3. M.J. Gidley, 'Factors Affecting the Crystalline Type (A–C) of Native Starches and Model Compounds:

A Rationalisation of Observed Effects in Terms of Polymorphic Structures', *Carbohydr. Res.*, **161**, 301–304 (1987).
4. H. Fuwa, T. Takaya, Y. Sugimoto, 'Degradation of Various Starch Granules by Amylases', in *Mechanisms of Saccharide Polymerization/Depolymerization*, ed. J.J. Marshall, Academic Press, New York, 73–100, 1980.
5. P. Colonna, V. Leloup, A. Buleon, 'Limiting Factors of Starch Hydrolysis', *Eur. J. Clin. Nutr.*, **46**, Suppl. 2, S17–S32 (1992).
6. D.J.A. Jenkins, T.M.S. Wolever, R.H. Taylor, H. Barker, H. Fielen, J.M. Baldwin, A.C. Bowling, H.C. Newman, A.L. Jenkins, D.V. Goff, 'Glycemic Index of Foods: A Physiological Basis for Carbohydrate Exchange', *Am. J. Clin. Nutr.*, **34**, 362–366 (1981).
7. K. O'Dea, P.J. Nestel, L. Antonoff, 'Physical Factors Influencing Postprandial Glucose and Insulin Responses to Starch', *Am. J. Clin. Nutr.*, **33**, 760–765 (1980).
8. K. O'Dea, P. Snow, P. Nestel, 'Rate of Starch Hydrolysis in Vitro as a Predictor of Metabolic Responses to Complex Carbohydrate in Vivo', *Am. J. Clin. Nutr.*, **34**, 1991–1993 (1981).
9. D.J.A. Jenkins, T.M.S. Wolever, R.H. Taylor, H. Ghafari, A.L. Jenkins, H. Barker, M.J.A. Jenkins, 'Rate of Digestion of Foods and Postprandial Glycaemia in Normal and Diabetic Subjects', *Br. Med. J.*, **2**, 14–17 (1980).
10. K.W. Heaton, S.N. Marcus, P.M. Emmett, C.H. Bolton, 'Particle Size of Wheat, Maize and Oat Test Meals: Effects on Plasma Glucose and Insulin Responses and on the Rate of Starch Digestion in Vitro', *Am. J. Clin. Nutr.*, **47**, 675–682 (1988).
11. J. Holm, I. Lundquist, I. Bjork, A.C. Eliasson, N.-G. Asp, 'Degree of Starch Gelatinization, Digestion Rate of Starch in Vitro, and Metabolic Response in Rats', *Am. J. Clin. Nutr.*, **47**, 1010–1016 (1988).
12. F. Bornet, A. Fontvieille, S. Rizkalla, P. Colonna, A. Blayo, C. Mercier, G. Slama, 'Insulin and Glycemic Responses in Healthy Humans to Native Starches Processed in Different Ways, Correlation with in Vitro α-Amylase Hydrolysis', *Am. J. Clin. Nutr.*, **50**, 315–323 (1989).
13. Y. Granfeldt, I. Bjork, A. Drews, J. Tovar, 'An in Vitro Procedure Based on Chewing to Predict Metabolic Response to Starch in Cereal and Legume Products', *Eur. J. Clin. Nutr.*, **46**, 649–660 (1992).
14. H.N. Englyst, S.M. Kingman, G.J. Hudson, J.H. Cummings, 'Measurement of Resistant Starch in Vitro and in Vivo', *Br. J. Nutr.*, **75**, 749–755 (1996).
15. I. Goni, A. Garcia-Alonso, F. Saura-Calixto, 'A Starch Hydrolysis Procedure to Estimate Glycemic Index', *Nutr. Res.*, **17**, 427–437 (1997).
16. H.N. Englyst, S.M. Kingman, 'Dietary Fiber and Resistant Starch. A Nutritional Classification of Plant Polysaccharides', in *Dietary Fiber*, eds. D. Kritchevsky, C. Bonfield, J.W. Anderson, Plenum Press, New York, 49–65, 1990.
17. J.H. Cummings, H.N. Englyst, 'Fermentation in the Human Large Intestine and the Available Substrates', *Am. J. Clin. Nutr.*, **45**, 1243–1255 (1987).
18. H.N. Englyst, S.M. Kingman, J.H. Cummings, 'Classification and Measurement of Nutritionally Important Starch Fractions', *Eur. J. Clin. Nutr.*, **46**, S33–S50 (1992).
19. K.N. Englyst, H.N. Englyst, G.J. Hudson, T.J. Cole, J.H. Cummings, 'Rapidly Available Glucose in Foods. A Measurement in Vitro that Reflects the Glycemic Response', *Am. J. Clin. Nutr.*, **69**, 448–454 (1999).
20. R.A. McCance, R.D. Lawrence, *The Carbohydrate Content of Foods*, MRC Special Report, HMSO, London, 1929.
21. D.A.T. Southgate, 'Determination of Carbohydrates in Foods. II. Unavailable Carbohydrates', *J. Sci. Food Agric.*, **20**, 331–335 (1969).
22. H.N. Englyst, H.S. Wiggins, J.H. Cummings, 'Determination of the Non-starch Polysaccharides in Plant Foods by Gas–Liquid Chromatography of Constituent Sugars as Alditol Acetates', *Analyst*, **107**, 307–318 (1982).
23. H.N. Englyst, M.E. Quigley, G.J. Hudson, J.H. Cummings, 'Determination of Dietary Fiber as Non-starch Polysaccharides by Gas–Liquid Chromatography', *Analyst*, **117**, 1707–1714 (1992).
24. M.E. Quigley, H.N. Englyst, 'Determination of Neutral Sugars and Hexosamines by High-performance Liquid Chromatography with Pulsed Amperometric Detection', *Analyst*, **117**, 1715–1718 (1992).
25. H.N. Englyst, M.E. Quigley, G.J. Hudson, 'Determination of Dietary Fiber as Non-starch Polysaccharides with Gas–Liquid Chromatographic, High-performance Liquid Chromatographic or Spectrophotometric Measurement of Constituent Sugars', *Analyst*, **119**, 1497–1509 (1994).
26. M.E. Quigley, H.N. Englyst, 'Determination of the Uronic Acid Constituents of Non-starch Polysaccharides by High-performance Liquid Chromatography with Pulsed Amperometric Detection', *Analyst*, **119**, 1511–1518 (1994).
27. G.B. Haber, K.W. Heaton, D. Murphy, L.F. Burroughs, 'Depletion and Disruption of Dietary Fiber. Effects on Satiety, Plasma Glucose and Serum Insulin', *Lancet*, **i**, 679–682 (1977).
28. FAO/WHO, 'Carbohydrates in Human Nutrition. Report of a Joint FAO/WHO Expert Consultation, Rome, 14–18 April 1997', FAO Food and Nutrition Paper 66, FAO, Rome, 1998.
29. W.C. Willett, 'The Dietary Pyramid: Does the Foundation Need Repair', *Am. J. Clin. Nutr.*, **68**, 218–219 (1998).
30. J. Salméron, A. Ascherio, E.B. Rimm, G.A. Colditz, D. Spiegelman, D.J. Jenkins, M.J. Stampfer, A.L. Wing, W.C. Willett, 'Dietary Fiber, Glycemic Load, and Risk of NIDDM in Men', *Diabetes Care*, **20**, 545–550 (1997).
31. J. Salméron, J.E. Manson, M.J. Stampfer, G.A. Colditz, A.L. Wing, W.C. Willett, 'Dietary Fiber, Glycemic Load, and the Risk of NIDDM in Women', *J. Am. Med. Assoc.*, **277**, 472–477 (1997).

32. D.R. Jacobs, K.A. Meyer, L.H. Kushi, A.R. Folsom, 'Whole Grain Intake May Reduce the Risk of Ischemic Heart Disease Death in Postmenopausal Women: The Iowa Women's Health Study', *Am. J. Clin. Nutr.*, **68**, 248–257 (1998).
33. J. Brand-Miller, 'Importance of Glycemic Index in Diabetes', *Am. J. Clin. Nutr.*, **59**, Suppl, S747–S752 (1994).
34. H.N. Englyst, J. Veenstra, G.J. Hudson, 'Measurement of Rapidly Available Glucose (RAG) in Plant Foods: A Potential in Vitro Predictor of the Glycaemic Response', *Br. J. Nutr.*, **75**, 327–337 (1996).
35. H.N. Englyst, J.H. Cummings, 'Digestion of the Polysaccharides of Some Cereal Foods in the Human Small Intestine', *Am. J. Clin. Nutr.*, **42**, 778–787 (1985).
36. H.N. Englyst, J.H. Cummings, 'Digestion of the Carbohydrates of Banana (*Musa paradisiaca sapientum*) in the Human Small Intestine', *Am. J. Clin. Nutr.*, **44**, 42–50 (1986).
37. H.N. Englyst, J.H. Cummings, 'Digestion of the Polysaccharides of Potato in the Small Intestine of Man', *Am. J. Clin. Nutr.*, **45**, 423–431 (1987).
38. K.R. Silvester, H.N. Englyst, J.H. Cummings, 'Ileal Recovery of Starch from Whole Diets Containing Resistant Starch Measured in Vitro and Fermentation of Ileal Effluent', *Am. J. Clin. Nutr.*, **62**, 403–411 (1995).
39. D.A.T. Southgate, *Determination of Food Carbohydrates*, 2nd edition, Elsevier Applied Science, London, 1991.
40. I. Bjork, M. Nyman, B. Pedersen, M. Siljestrom, N.-G. Asp, B.O. Eggum, 'On the Digestibility of Starch in Wheat Bread – Studies in Vitro and in Vivo', *J. Cereal Sci.*, **4**, 1–11 (1986).
41. C.S. Berry, 'Resistant Starch: Formation and Measurement of Starch that Survives Exhaustive Digestion with Amylolytic Enzymes During the Determination of Dietary Fiber', *J. Cereal Sci.*, **4**, 301–314 (1986).
42. J.G. Muir, K. O'Dea, 'Measurement of Resistant Starch: Factors Affecting the Amount of Starch Escaping Digestion in Vitro', *Am. J. Clin. Nutr.*, **56**, 123–127 (1992).
43. N. Faisant, D.J. Gallant, B. Bouchet, M. Champ, 'Banana Starch Breakdown in the Human Small Intestine Studied by Electron Microscopy', *Eur. J. Clin. Nutr.*, **49**, 98–104 (1995).

Viscosity of Food: Measurement and Application

M.A. Rao
Cornell University, Geneva, NY, USA

1 Introduction **1**
1.1 Newtonian and non-Newtonian Foods 2
1.2 Description of Time-independent Flow Behavior 2
1.3 Time-dependent Behavior 4
1.4 Effect of Soluble and Insoluble Solids Concentration on Apparent Viscosity of Foods 4
1.5 Effect of Biopolymer Concentration on Zero-shear Viscosity 5
1.6 Food Dispersions 6
1.7 Effect of Temperature on Viscosity 7
2 Measurement of Viscosity and Rheological Behavior **8**
2.1 Viscometer Geometries 8
2.2 Measurement of Viscoelastic Behavior of Fluid Foods 9
2.3 Relationships Between Steady Shear and Oscillatory Shear Parameters 11
3 Stimuli for Evaluation of Viscosity **12**
4 Velocity Profiles in Tube Flow **12**
5 Energy Requirements for Pumping **13**
5.1 Mechanical Energy Balance Equation 13
5.2 Friction Losses in Pipes 14
List of Symbols **14**
Greek Letters 15
Dimensionless Numbers 15
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

The measurement of viscosity and viscoelastic behavior of fluid foods are reviewed. The role of factors such as rate of shear, temperature, and concentration of soluble and insoluble solids on the viscosity of food is covered. Two important areas in food technology, sensory assessment of viscosity and role of viscosity in pipe flow, are discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The study of viscosity is but one aspect of rheology, which by definition is the study of deformation and flow of matter. The science of rheology grew considerably due to concurrent developments in polymer science, that in turn were necessary due to the many industrial uses of synthetic polymers. The rheological properties of foods are based on flow and deformation responses when subjected to stress. The word ''texture'' is generally associated with solid foods and is based on empirical tests using puncture, compression, and extrusion that are still specific to each food commodity. In most, if not all, instances the test geometries used in texture measurement are too complex for reliable mathematical analysis of the forces and strains involved.

Fluid foods containing relatively large amounts of dissolved low-molecular-weight compounds (e.g. sugars) and no significant amount of polymers or insoluble solids can be expected to exhibit Newtonian behavior:

- Newtonian foods – water, filtered depectinized juices, refined vegetable oils, sugar syrup, and wines;
- non-Newtonian foods – solutions of biopolymers (also called gums and hydrocolloids), chocolate, starch dispersions, baby foods and other puréed fruit and vegetable products, orange juice, tomato juice, paste and concentrate, mustard, mayonnaise, salad dressing, concentrated milk, and yoghurt.

A small amount (ca. 1%) of a dissolved polymer can substantially increase the viscosity of water and also alter the flow characteristics from Newtonian of water to non-Newtonian of the aqueous dispersion; although the rheological properties are altered substantially, the magnitudes of the thermal properties (e.g. density and thermal conductivity) of the dispersion remain relatively close to those of water. The techniques for measuring rheological properties, mathematical models for describing rheological behavior, and application to sensory assessment and handling/heating of foods have been reviewed by several authors. Here, they are covered in brief.

Shear rate, denoted by the symbol $\dot{\gamma}$, is the velocity gradient established in a fluid as a result of a shear stress acting on it. It is expressed in units of reciprocal seconds. Shear stress is the stress component applied tangentially; it is equal to the force vector (a vector has both magnitude and direction) divided by the area of application and is expressed in units of force per unit area, pascals. The nomenclature committee of the Society of Rheology recommended that the symbol σ be used to denote shear stress. However, the symbol τ that was used

to denote shear stress for a long time is often encountered in the literature.

Viscosity, is the internal friction of a fluid or its tendency to resist flow. It is denoted by the symbol η for Newtonian fluids, whose viscosity does not depend on the shear rate, and for non-Newtonian fluids to indicate shear rate dependence by η_a. Depending on the flow system and choice of shear rate and shear stress, there are several equations to calculate the magnitude of η_a. Here, it is defined by Equation (1):

$$\eta_a = \frac{\text{shear stress}}{\text{shear rate}} = \frac{\sigma}{\dot{\gamma}} \quad (1)$$

The preferred units of viscosity are Pa s or mPa s, but other units such as the centipoise (1 cP = 1 mPa s) can be found in the literature.

1.1 Newtonian and non-Newtonian Foods

1.1.1 Newtonian

A plot of shear rate versus shear stress can be used to classify flow behavior of foods (Figure 1). With Newtonian fluids, the shear rate is directly proportional to the shear stress and the plot begins at the origin. Typical Newtonian foods are those containing only low-molecular-weight compounds (e.g. sugars) and that do not contain large concentrations of either dissolved polymers (e.g. pectins, proteins, starches) or insoluble solids. For example, water, sugar syrups, most honeys, most carbonated beverages, edible oils, filtered juices, and milk are Newtonian fluids.

All other types of fluid foods are non-Newtonian which means that either the shear stress–shear rate plot is not linear and/or the plot does not begin at the origin, or the material exhibits time-dependent rheological behavior as a result of structural changes. Flow behavior may depend only on shear rate and not on the duration of shear (time-independent) or may depend also on the duration of shear (time-dependent). Several types of time-independent flow behavior of foods have been encountered.

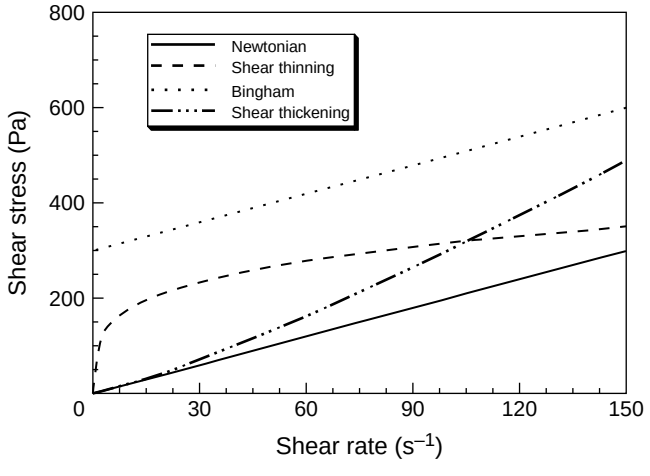

Figure 1 Shear rate versus shear stress characteristics of Newtonian, Bingham, shear-thinning (pseudo plastic), and shear-thickening (dilatant) fluids.

1.1.2 Shear Thinning

With shear-thinning fluids, the curve begins at the origin of the shear stress–shear rate plot but is concave upwards, i.e. an increasing shear rate gives a less than proportional increase in shear stress. Shear-thinning fluids are popularly called pseudoplastic. The term ''shear thinning'' is preferred because it is an accurate description of the shear rate–shear stress curve. Shear thinning may be thought of being due to the breakdown of structural units in a food due to the hydrodynamic forces generated during shear. Most non-Newtonian foods exhibit shear-thinning behavior, including many mayonnaises, salad dressings, and some concentrated fruit juices as well as puréed fruits and vegetables.

1.1.3 Yield Stress

The flow of some materials may not commence until a threshold value of stress, the yield stress (σ_0), is exceeded. Shear thinning with yield stress behavior is exhibited by foods, such as tomato concentrates, tomato ketchup, mustard, and mayonnaise. In the event, the shear rate–shear stress data follow a straight line with a yield stress, the food is said to follow the Bingham plastic model.

1.1.4 Shear Thickening

In shear-thickening behavior also, the curve begins at the origin of the shear stress–shear rate plot and is concave downwards, i.e. an increasing shear stress gives a less than proportional increase in shear rate. This type of flow has been encountered in partially gelatinized starch dispersions. The term ''dilatant'' is popularly and incorrectly used to describe shear thickening. Because dilatancy implies an increase in the volume of the sample during the test, it is incorrect to use it to describe shear-thickening rheological behavior. Strictly speaking, shear-thickening should be due to increase in the size of the structural units as a result of shear.

1.2 Description of Time-independent Flow Behavior

A flow model may be considered to be a mathematical equation that can describe rheological data, such as shear rate–shear stress data in a basic shear diagram; it provides a convenient and concise manner of describing the data.

Occasionally, such as for the viscosity versus temperature data during starch gelatinization, more than one equation may be necessary to describe the rheological data.

1.2.1 Newtonian Model

The model for a Newtonian fluid is described by Equation (2):

$$\sigma = \eta\dot{\gamma} \tag{2}$$

From the definition of a Newtonian fluid, the shear stress and the shear rate are proportional to each other, and a single parameter η characterizes the data. For a Bingham plastic fluid that exhibits a yield stress (σ_0), the model is

$$\sigma - \sigma_0 = \eta'\dot{\gamma} \tag{3}$$

where in Equation (3) η' is called the Bingham plastic viscosity.

As shown in Figure 1, the Newtonian model and the Bingham plastic model can be illustrated by straight lines in terms of shear rate and shear stress, and the former can be described by one parameter η and the latter by the two parameters η' and σ_0. However, the shear rate–shear stress data of shear-thinning and shear-thickening fluids are curves that require more than one parameter to describe their data. Because the equation of a straight line is simple in that it can be described by just two parameters (the slope and the intercept) shear rate–shear stress data are often transformed in to such lines.

1.2.2 Power Law Model

Shear stress–shear rate plots of many fluids become linear when plotted on double logarithmic coordinates and the power law model describes the data of shear-thinning and shear thickening fluids (Equation 4):

$$\sigma = K\dot{\gamma}^n \tag{4}$$

where K, the consistency index (Pa s^n), is the shear stress at a shear rate of 1.0 s^{-1} and the exponent n, the flow behavior index, is dimensionless that also reflects the closeness to Newtonian flow. For the special case of a Newtonian fluid ($n = 1$), the consistency index K is identically equal to the viscosity of the fluid. When the magnitude of $n < 1$ the fluid is shear thinning and when $n > 1$ the fluid is shear thickening in nature. Taking logarithms of both sides of Equation (4),

$$\log\sigma = \log K + n\log\dot{\gamma} \tag{5}$$

In Equation (5) the parameters K and n are determined from a plot of $\log\sigma$ versus $\log\dot{\gamma}$, and the resulting straight-line intercept is $\log K$ and the slope is n. If a large number (e.g. >15) of σ, $\dot{\gamma}$ data points are available (it is easy to obtain a large number of points with automated viscometers), linear regression of $\log\dot{\gamma}$ versus $\log\sigma$ will provide statistically best values of K and n. Nevertheless, a plot of experimental and predicted values of $\log\dot{\gamma}$ and $\log\sigma$ is useful for observing trends in data and ability of the model to follow the data. Linear regression techniques can also be used for determination of the parameters of the Herschel–Bulkley (when the magnitude of the yield stress is known) and the Casson models, discussed later.

Because it contains only the two parameters K and n, the power law model has been used extensively to characterize fluid foods, and in studies on handling and heating/cooling of foods. Extensive compilations of the magnitudes of power law parameters can be found in Holdsworth[1] and Rao.[2] Because it is convenient to group foods into commodities, Rao[2] compiled the magnitudes of power law parameters of food commodities. In addition, the influence of temperature in quantitative terms of activation energies, and the effect of concentration of soluble and insoluble solids on the consistency index are given.

Although the power law model is popular and useful, its empirical nature should be noted. Another reason for its popularity results from its applicability over the shear rate range 10^1–10^4 s^{-1} that can be obtained in many commercial viscometers. Often, the magnitudes of the consistency and the flow behavior indexes of a food sample depend on the specific shear rate range being used so that when comparing the properties of different samples an attempt should be made to determine K and n over a specific range of shear rates.

One drawback of the power law model is that it does not describe the low-shear and high-shear rate constant-viscosity data of shear-thinning foods where limiting zero-shear (η_0) and infinite-shear viscosities (η_∞) are reached, respectively. The models of Cross and Carreau were capable of describing apparent viscosity versus shear rate data over the entire ranges of data obtained on food polymer dispersions.[2] Very low shear rates (ca. 10^{-3} s^{-1}) are necessary to obtain reliable values of η_0.

1.2.3 Herschel–Bulkley Model

When yield stress of a food is measurable, it can be included in the power law model and the model is known as the Herschel–Bulkley model (Equation 6):

$$\sigma - \sigma_0 = K_H(\dot{\gamma})^{n_H} \tag{6}$$

where $\dot{\gamma}$ is the shear rate (s^{-1}), σ is shear stress (Pa), n_H is the flow behavior index, K_H is the consistency index, and σ_0 is the yield stress. The concept of yield stress has been challenged by Barnes and Walters[3] because a fluid may deform minutely at stress values lower than the yield stress. Nevertheless, yield stress may be considered to

be an engineering reality and plays an important role in many food products.

If the yield stress of a sample is known from an independent experiment, K_H and n_H can be determined from linear regression of $\log(\sigma - \sigma_0)$ versus $\log(\dot{\gamma})$ as the intercept and slope, respectively. Alternatively, nonlinear regression techniques can be used to estimate σ_0, K_H, and n_H of tomato concentrates.[4] However, estimated values of yield stress and other rheological parameters should be used only when experimentally determined values are not available.

1.2.4 Casson Model

The Casson[5] model (Equation 7) is a structure-based model that, although originally developed for characterizing printing inks, has been used for a number of food dispersions:

$$\sigma^{0.5} = K_{0c} + K_c(\dot{\gamma})^{0.5} \tag{7}$$

For a food following the Casson model, a straight line results when the square root of shear rate, $(\dot{\gamma})^{0.5}$, is plotted against the square root of shear stress, $(\sigma)^{0.5}$ with slope K_c and intercept K_{0c}. The Casson yield stress is calculated as the square of the intercept, $\sigma_{0c} = (K_{0c})^2$ and the Casson plastic viscosity as the square of the slope, $\eta_{Ca} = (K_c)^2$. The International Office of Cocoa and Chocolate has adopted the Casson model as the official method for interpretation of flow data on chocolate samples. The Casson plastic viscosity can be used as being equivalent to the infinite shear viscosity (η_∞) of dispersions by considering the limiting viscosity at infinite shear rate.[2]

1.3 Time-dependent Behavior

Foods that exhibit time-dependent shear-thinning behavior are said to exhibit thixotropic flow behavior. Most of the foods that exhibit thixotropic behavior are heterogeneous systems containing a dispersed phase containing small particles. At rest, the particles or molecules in the food are linked together by weak forces. When the hydrodynamic forces during shear are sufficiently high, the interparticle linkages are broken, resulting in reduction in the size of the structural units that in turn offer lower resistance to flow during shear. This type of behavior is common to foods such as salad dressings and soft cheeses where the structural adjustments take place in the food due to shear until an equilibrium is reached.

Time-dependent shear-thickening behavior, formerly called rheopectic behavior, is called antithixotropic behavior. This type of behavior, although rare, is being detected in many foods due to the development of sensitive automated rheometers. When shear rate versus shear stress data are obtained first in ascending order of shear rate and immediately afterwards in descending order, the two curves will not coincide and values of the latter will be lower than the former for thixotropic foods. Repetition of the experiments will result in an equilibrium hysteresis loop.

In antithixotropic behavior, the shear stress values in descending order of shear rates are higher than those in ascending order. The primary difficulty in obtaining reliable thixotropic or antithixotropic data is that during loading of the test sample in to a measurement geometry structural changes occur that cannot be either controlled or expressed quantitatively. Therefore, considerable caution should be exercised in using the area of hysteresis loop as a measure of thixotropic behavior and a characteristic of the food.

1.4 Effect of Soluble and Insoluble Solids Concentration on Apparent Viscosity of Foods

1.4.1 Intrinsic Viscosity

The effect of concentration on the Newtonian or zero-shear viscosity of biopolymer dispersions can be expressed in terms of the coil overlap parameter $c[\eta]$, where c is the concentration of the polymer and is the intrinsic viscosity of the polymer in a specific solvent.

Many foods contain high-molecular-weight polymers, such as proteins, pectins, and others. Often, they contribute significantly to the structure and viscosity of foods. In dilute solutions, the polymer chains are separate and the intrinsic viscosity, denoted as $[\eta]$, of a polymer in solution depends only on the dimensions of the polymer chain. Because $[\eta]$ indicates the hydrodynamic volume of the polymer molecule and is related to the molecular weight and to the radius of gyration, it reflects important molecular characteristics of a biopolymer. The concentrations of polymers used should be such that the relative viscosities (η/η_s) of the dispersions are from about 1.2 to 2.0 to ensure good accuracy and linearity of extrapolation to zero concentration.[6,7] Intrinsic viscosity can be determined from dilute solution viscosity data as the zero concentration limit of specific viscosity η_{sp} divided by concentration c (Equation 8):

$$[\eta] = \lim_{c \to 0} \frac{\eta_{sp}}{c} \tag{8}$$

where $\eta_{sp} = [(\eta - \eta_s)/\eta_s]$, and η and η_s are the viscosities of the solution and the solvent, respectively. When a dilute solution exhibits shear-thinning behavior, its zero shear viscosity, η_0, at very low shear rates (vanishing shear rates) may be used in place of the Newtonian viscosity η. There are several ways of determining the intrinsic viscosity $[\eta]$ from experimental dilute solution viscosity data.[8] The two equations commonly employed

for determining $[\eta]$ of food gums are those of Huggins (Equation 9) and Kraemer (Equation 10):

$$\frac{\eta_{sp}}{c} = [\eta] + k_1[\eta]^2 c \tag{9}$$

$$\frac{\ln(\eta_r)}{c} = [\eta] + k_2[\eta]^2 c \tag{10}$$

In Equation (10), η_r is the relative viscosity (η/η_s). Equations (9) and (10) imply that plots of η_{sp}/c versus c and $\ln(\eta_r)/c$ versus c would result in straight lines, respectively. To obtain reliable values of intrinsic viscosity, the extrapolations are usually done for relative viscosity values between 1.2 and 2.0 when the corresponding specific viscosities are between about 0.2 and 1.0. The Huggins constant k_1 is considered to be an index of the polymer–polymer interaction with a large number of the reported values being between 0.3 in good solvents and 1.0 in theta solvents; the higher values are attributed to the existence of association between the macromolecules. The Huggins and Kraemer constants (k_1 and k_2) are theoretically related by Equation (11):

$$k_1 = k_2 + 0.5 \tag{11}$$

The Huggins equation has been used more extensively than the Kraemer equation. For polyelectrolytes (charged polymers), Tam and Tiu[9] utilized the equation of Fuoss–Strauss (Equation 12):

$$\frac{\eta_{sp}}{c} = \frac{[\eta]}{(1 + Bc^{1/2})} \tag{12}$$

When (c/η_{sp}) is plotted against $c^{1/2}$, a straight line is obtained with an intercept of $1/[\eta]$ and a slope of $B/[\eta]$. The Mark–Houwink empirical equation relates intrinsic viscosity and the average molecular weight M (Equation 13):

$$[\eta] = K(M)^a \tag{13}$$

where K and a are constants at a specific temperature for a given polymer–solvent system. The magnitude of the constant a is generally in the range 0.5–0.8 when conformation of the polymer is random coil.[10] For polymers that assume compact conformations, the magnitudes of the exponent a are lower than 0.5. Magnitudes of the constant a higher than 0.8 are encountered when hydrodynamic interactions are absent (free draining random coil), i.e. when the solvent can drain freely between the macromolecules, as is the case when the conformations are elongated and rigid (rod-like polymers). A large number of models have been used to deduce intrinsic viscosity–molecular weight relationships,[10] including the equivalent-sphere model (Flory–Fox equation), the random-flight model for flexible chain molecules, the worm-like chain model, and rigid elongated molecules. Rigid elongated conformations have been associated with large values of the Mark–Houwink exponent, with values of 1.7–1.8 being reported for some polyelectrolytes at low ionic strength and high degrees of disassociation. The magnitudes of intrinsic viscosity and of molecular weight of several biopolymers are available;[2] significant differences in the magnitudes are encountered depending on the structural characteristics, such as molecular mass (length of the chain) and different degrees of substitution. For polymers with charges, both the magnitude of $[\eta]$ and the rheological behavior are also affected by the pH and the ionic strength of the solution; therefore, their values should be controlled in studies of $[\eta]$ and their rheological behavior.

1.5 Effect of Biopolymer Concentration on Zero-shear Viscosity

In earlier studies on solutions of synthetic polymer melts,[11] the zero-shear viscosity was found to be related to the molecular weight of the polymers. Plots of $\log \eta$ versus $\log M$ often resulted in two straight lines with the lower-M section having a slope of about 1.0 and the upper-M section having a slope of about 3.4. Because the apparent viscosity also increases with concentration of a specific polymer, the roles of both molecular size and concentration of polymer need to be understood. In polymer dispersions of moderate concentration, the viscosity is controlled primarily by the extent to which the polymer chains interpenetrate that is characterized by the coil overlap parameter $c[\eta]$.[12] The product $c[\eta]$ is dimensionless and indicates the volume occupied by the polymer molecule in the solution.

A plot of $\log c[\eta]$ against $\log[(\eta_0 - \eta_s)/\eta_s]$ of several food gum solutions at 25 °C adapted from Morris et al.[13] is shown in Figure 2. In Figure 2, there are two regions: a region of dilute dispersions where the viscosity dependence on concentration follows a 1.4 power, and a region of concentrated dispersions where the viscosity dependence on concentration follows a 3.3 power. The transition from the dilute to the concentrated region occurred at a value of $c[\eta] = 4$. However, data on a few biopolymers (e.g. guar gum and locust bean gum) were found to deviate from the above observations. First, the region of concentrated solution behavior began at lower values of the coil overlap parameter $c[\eta] = 2.5$. Second, the viscosity showed a higher dependence on concentration with a slope of 5.1 instead of about 3.3. These deviations were attributed to specific intermolecular associations (hyperentanglements) between regular and rigid chain sequences in addition to the simple process of interpretation. Launay et al.[10] suggested that there could be two transitions, instead of one transition shown in Figure 2,

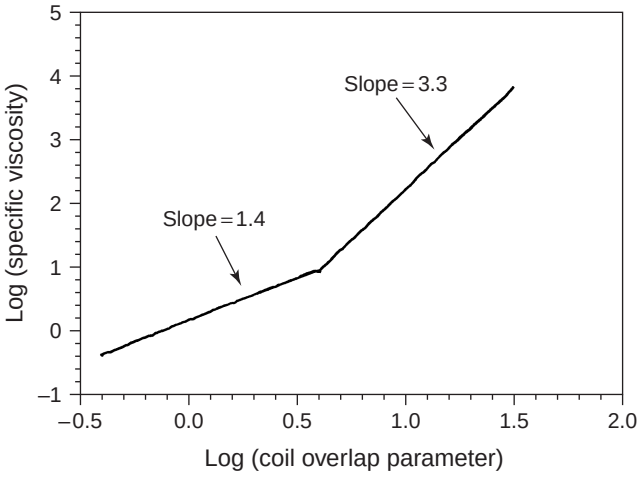

Figure 2 Plot of log coil overlap parameter ($c[\eta]$) versus $\log[(\eta_0 - \eta_s)/\eta_s]$ of several biopolymer dispersions; η_0 is zero shear viscosity and η_s is viscosity of the solvent at 25 °C. (Reproduced with permission from M.A. Rao, 'Engineering Properties of Foods: Current Status', in *Food Engineering 2000*, eds. P. Fito, E. Oretga Rodriguez, G. Barbosa-Cánovas, Aspen Publishers Inc., Gaithersburg MD, 39–54.)

before the onset of high concentration–viscosity behavior. Because intrinsic viscosity of a biopolymer can be determined with relative ease, Figure 2 can be used to estimate the polymer's zero-shear viscosity of any food polymer at a specific polymer concentration at 25 °C.

Unlike biopolymer dispersions where the intrinsic viscosity is known and the concentration can be chosen a priori, often for fluid foods the concentration of soluble (e.g. pectins in fruit juices) and insoluble solids can be determined only posteriori, and the determination of their zero-shear viscosities is also difficult due to instrument limitation and the existence of yield stress. In most foods, often it is possible to identify the component(s), called key component(s) that play an important role in the rheological properties.[2] The effect of concentration c of soluble or insoluble solids on either apparent viscosity η_a or the consistency index of the power law model K can be described by either exponential or power law relationships (Equations 14–17):

$$\eta_a \propto \exp(ac) \tag{14}$$

$$\eta_a \propto c^b \tag{15}$$

$$K \propto \exp(a'c) \tag{16}$$

$$K \propto c^{b'} \tag{17}$$

where a, a', b and b' are constants to be determined from experimental data. For example, the effect of concentration c of soluble solids (°Brix) and insoluble solids (pulp) on either apparent viscosity or the consistency index of the power law model of concentrated orange juice (COJ) can be described by exponential relationships.[14,15] In the case of COJ, it should be noted that insoluble solids are expressed in terms of pulp content determined on a 12 °Brix sample by centrifugation for 10 min at $360 \times g$ (Equations 18 and 19):

$$K = K^c \exp(B_K^c\,^\circ\text{Brix}) \tag{18}$$

$$K = K^P \exp(B_K^P\ \text{pulp}) \tag{19}$$

where K^c, K^P, B_K^c, and B_K^P are constants. The role of insoluble solids can be also studied in terms of the relative viscosity, η_r = (apparent viscosity of COJ/apparent viscosity of serum), and pulp fraction and, as expected, such a plot has the limiting value of 1.0 at zero pulp fraction (serum).[2]

1.6 Food Dispersions

When dispersions contain insoluble particles of nearly uniform size, their viscosity can be expressed in terms of the Peclet number.[16,17] Yoo and Rao[18] showed that the influence of two different sizes of tomato pulp particles (TP4 and TP6) at a pulp weight fraction of 17% could be described in terms of Peclet number and relative viscosity (Figure 3). Magnitudes of the Peclet numbers (Pe) were calculated using Equation (20):

$$\text{Pe} = \frac{\eta_0 r_0^3 \dot{\gamma}}{kT} \tag{20}$$

where η_0 is the viscosity of the suspending liquid (serum), r_0 is the particle radius, k is the Boltzmann constant

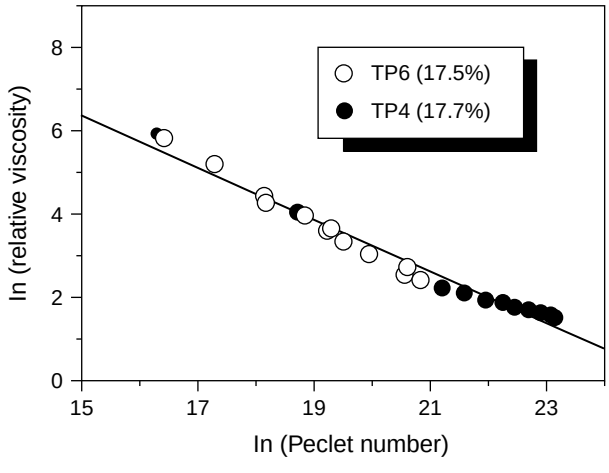

Figure 3 Peclet number versus relative viscosity of tomato pulp particles. At equal values of pulp weight fraction, a TP6 sample with small diameter particles was more viscous than a TP4 sample with large diameter particles. (Reproduced by permission of Food and Nutrition Press, Inc., from Yoo and Rao.[18])

(1.38×10^{-23} N m K^{-1}), and T is the absolute temperature. The Peclet number can be used to compare the effect of imposed shear with the effect of diffusion of the particles. For example, when Pe $\gg 1$, hydrodynamic effects dominate and a dispersion of spherical particles exhibits shear-thinning behavior; in contrast, when Pe $\ll 1$, the distribution of particles is only slightly altered by the flow.[19]

Starch is the major energy reserve of seeds, roots, and tubers, and is produced in small packets called granules. Irreversible changes occur, often referred to as gelatinization, when a starch–water mixture is heated above a critical temperature that depends on the source of starch. These changes include loss of birefringence and, more importantly, crystallinity that allows the granules to swell under excess water conditions. A gelatinized starch dispersion may be considered to be a composite material consisting of swollen granules and granular fragments dispersed in a continuous biopolymer matrix, so that the properties of the dispersed phase, the continuous phase, and interactions between the components are three important considerations for insights into the rheology of gelatinized starch suspensions.[2]

The viscosity–temperature profile of a starch dispersion exhibits an initial phase during which the viscosity increases, reaches a peak value, and at high temperatures decreases exponentially to a nearly equilibrium value. Starch gelatinization up to peak viscosity studied by various techniques such as viscometry and particle size measurement by laser diffraction, is essentially a first-order kinetic process whose extent is determined principally by pasting temperature and time.[2] At low starch concentration, the simple power law model (Equation 4) describes the shear responses over a wide range of shear rates. However, gelatinized starch suspensions often exhibit yield stresses at low shear rates.

The mass fraction of granules plays an important role in the magnitudes of rheological properties of starch dispersions.[20] Because of their deformable nature, it is difficult to determine volume fraction of starch dispersions accurately and it is preferable to work with starch granule mass fraction, denoted as cQ, where c is the dry starch concentration, w/w, and Q is the mass of hydrated starch granules per unit weight of dry starch. The quantity cQ takes into consideration both the solids concentration and volume of starch granules after water absorption, and plays an important role in consistency and yield stress,[21] and viscoelastic properties.[2]

1.7 Effect of Temperature on Viscosity

A wide range of temperatures is encountered during processing and storage of fluid foods, so that the effect of temperature on rheological properties needs to be described quantitatively. The effect of temperature on the viscosity of Newtonian fluids, or on either the apparent viscosity at a specified shear rate or the consistency index K of the power law model of a fluid can be described often by the Arrhenius relationship. For the apparent viscosity, the Arrhenius relationship is written as Equation (21):

$$\eta_a = \eta_{\infty A} \exp\left(\frac{E_a}{RT}\right) \tag{21}$$

where η_a is the apparent viscosity at a specific shear rate, $\eta_{\infty A}$ is the frequency factor, E_a is the activation energy (J mol^{-1}), R is the gas constant (J mol^{-1} K^{-1}), and T is the temperature (K).

The quantity E_a is the energy barrier that must be overcome before the elementary flow process can occur and E_a is the activation energy for viscous flow. From the slope of a plot of $\ln \eta_a$ (ordinate) versus $(1/T)$ (abscissa), E_a can be calculated as (slope $\times R$); we note that η_∞ is the exponential of the intercept. The Arrhenius equation for the consistency index is (Equation 22)

$$K = K_\infty \exp\left(\frac{E_{ak}}{RT}\right) \tag{22}$$

where K_∞ is the frequency factor, E_{aK} is the activation energy (J mol^{-1}), R is the gas constant, and T is the temperature (K). A plot of $\ln K$ (ordinate) versus $(1/T)$ (abscissa) results in a straight line, and E_{aK} = (slope $\times R$), and K_∞ is the exponential of the intercept. Although the activation energy should be expressed in joules, it is often expressed in calories (1 cal = 4.1868 J). The applicability of the Arrhenius model to the apparent viscosity versus

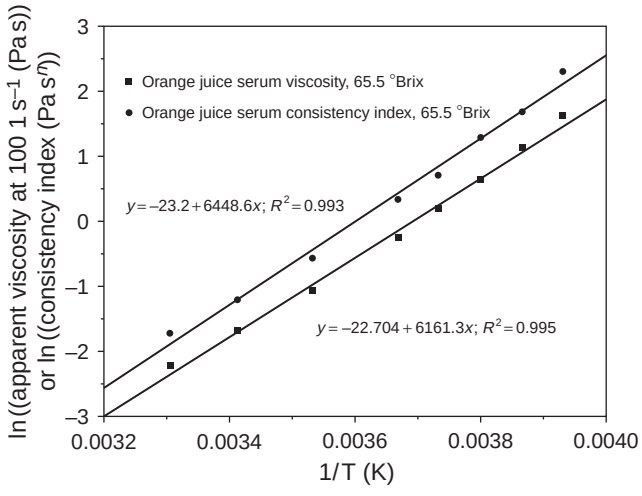

Figure 4 Arrhenius plot for apparent viscosity and power law consistency index of orange juice.[2] (Reprinted with permission from M.A. Rao, *Rheology of Fluid and Semifluid Foods: Principles and Applications*, © 1999, Aspen Publishers Inc., Gaithersburg, MD, 501.)

temperature data on a COJ serum sample[14,15] is shown in Figure 4.

The Arrhenius equation did not describe very well the influence of temperature on viscosity data of concentrated apple and grape juices in the range 60–68 °Brix.[22,23] It was determined that the empirical Fulcher equation (see Ferry,[11] p. 289, and Soesanto and Williams[24]) described the viscosity versus temperature data on those juice samples better than the Arrhenius model:[23]

$$\log \eta = A' + \frac{B'}{T - T_\infty} \tag{23}$$

where in Equation (23) A' and B' are constants.

2 MEASUREMENT OF VISCOSITY AND RHEOLOGICAL BEHAVIOR

A well-designed viscometer must be capable of providing readings that can be converted to shear rate and shear stress. In addition, it should allow for the recording of the readings so that time-dependent flow behavior can be studied. The shear stresses that need to be measured in the case of low-viscosity foods are low in magnitudes so that instruments that minimize friction by the use of gas bearings are very useful. The flow in the selected geometry should be steady, laminar and fully developed, and the temperature of the test fluid should be maintained uniform. For obtaining data under isothermal conditions the temperature of the sample should be maintained constant at the desired level.

2.1 Viscometer Geometries

For foods that exhibit Newtonian behavior, viscometers that operate at a single shear rate (e.g. glass capillary) are acceptable. A glass capillary viscometer is shown in Figure 5. A test sample is loaded to mark E by first inverting the viscometer, immersing tube A in the sample, and applying suction to tube G. After the sample had attained the desired temperature in a bath, suction is applied to tube A so that the test liquid is drawn above mark C. The efflux time t for the liquid to pass from mark C to mark E is noted and used to calculate the viscosity using calibration constants provided by the manufacturer. Alternatively, the elapsed times for a test liquid t and a standard liquid t_{st} of the same density to flow between two etched lines are determined. Because the magnitude of viscosity is directly proportional to the time of flow, the viscosity of a fluid can be calculated as Equation (24):

$$\eta = \left(\frac{t}{t_{st}} \right. \quad \eta_{st} \tag{24}$$

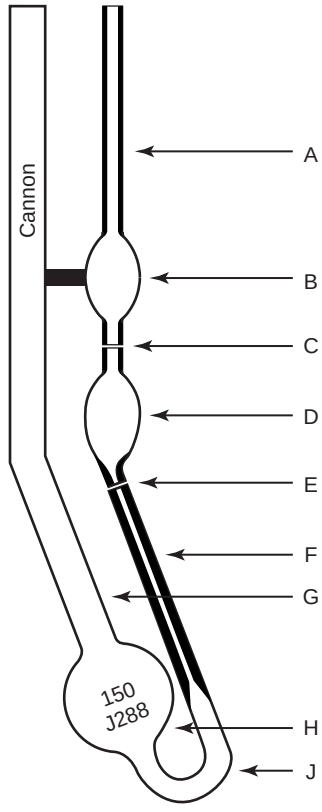

Figure 5 Glass capillary viscometer. (Reproduced by courtesy of Cannon Instrument Co.)

The magnitudes of t and t_{st} must be determined with care at a specific temperature, and the magnitude of viscosity of the standard fluid η_{st} must be reliable. Glass capillary viscometers are not suitable for liquids that deviate substantially from Newtonian flow or contain large-size particles or a high concentration of suspended solids.

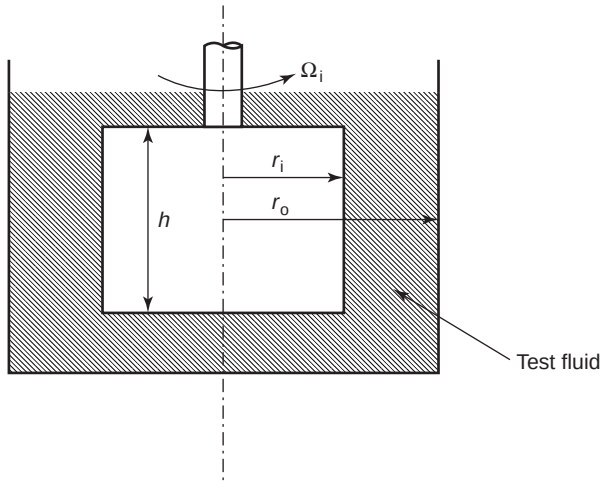

Figure 6 Diagram of a concentric cylinder geometry.

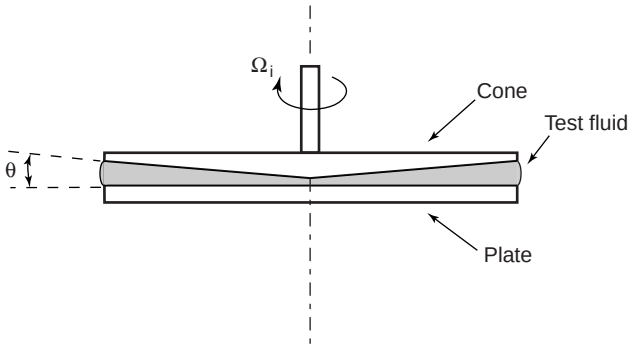

Figure 7 Diagram of a cone-plate geometry.

For foods that exhibit non-Newtonian behavior, data should be obtained at several shear rates and the commonly used viscometric flow geometries in rheological studies on foods include: (a) concentric cylinder (Figure 6), (b) plate and cone (cone-plate; Figure 7), (c) parallel disk (also called parallel plate), (d) capillary/tube/pipe, and (e) slit flow. The relationships for shear stress for flow in capillary/tube and slit geometries can be deduced relatively easily from the pressure drop over a fixed length after fully developed flow has been achieved. Likewise, in concentric cylinder, cone-plate, parallel disk geometries, the shear stress can be calculated from the measured torque and the dimensions of the test geometry being used. In contrast, the derivation of expressions for the shear rate requires solution of the continuity and momentum equations with the applicable boundary conditions can be found in several texts, such as Whorlow[25] and Rao,[2] and will not be covered here. Further, modern commercial viscometers/rheometers are automated so that the shear rate and shear stress are calculated by a computer. It is noted that in a properly designed cone-plate geometry the shear rate depends only on the rotational speed and not on the geometrical characteristics. In all other flow geometries, the dimensions of the measuring geometry (capillary, concentric cylinder, parallel disks) play important roles.[2] Because the concentric cylinder and the plate-cone geometries are encountered commonly, the equations for shear stress and Newtonian shear rate are given.

2.1.1 Concentric Cylinder Viscometer

In a concentric cylinder geometry, the shear stress can be determined from the total torque M (Equation 25):

$$\sigma = \frac{M}{2\pi r_i^2 h} \tag{25}$$

where r_i is the radius of the rotating bob. The Newtonian shear rate in a concentric cylinder geometry can be calculated exactly from Equation (26):

$$\dot{\gamma}_N = \frac{2\Omega r_o^2}{[1/r_i^2 - 1/r_o^2]} = \frac{2\Omega}{[1 - (r_i/r_o)^2]} \tag{26}$$

where Ω is the angular velocity of the rotating bob, r_i is the radius of the bob, and r_o is the radius of the cup. Recalling that viscosity is obtained by dividing the magnitude of shear stress by the corresponding shear rate, the viscosity of a Newtonian fluid from concentric cylinder flow data is given by the Margules Equation (27):

$$\eta = \left(\frac{M}{4\pi h\Omega} \right. \quad \left(\frac{1}{r_i^2} - \frac{1}{r_o^2} \right. \tag{27}$$

2.1.2 Cone-plate Viscometer

The shear rate and shear stress are given by Equations (28) and (29):

$$\text{shear stress, } \sigma = \frac{3T_c}{D} \tag{28}$$

$$\text{shear rate, } \dot{\gamma} = \frac{\Omega}{\theta} \tag{29}$$

where T_c is the torque per unit area, D is the diameter of the rotating cone or plate, Ω is the angular velocity, and θ is the cone angle in radians. In general, θ is usually quite small (2–4°).

2.2 Measurement of Viscoelastic Behavior of Fluid Foods

Viscometric functions are generally used to relate stress with shear rate in liquid systems, whereas elastic functions are related to the appropriate stress function to strain in solids. Viscoelastic properties cover the combination where a material exhibits both viscous and elastic properties. The relationships between stress and strain, and the influence of time on them, are generally described by constitutive equations or rheological equations of state.[11] When the strains are relatively small, i.e. in the linear range, the constitutive equations are relatively simple. In the linear viscoelastic range, one can express a large linear deformation as the sum of all the small linear deformations.

Viscoelastic behavior of many foods has been studied by means of dynamic shear, creep compliance, and stress relaxation techniques. The paradigm shift from emphasis on normal stress measurements to dynamic rheological technique took place gradually, beginning in the late 1950s. It is customary to employ different symbols for the various rheological parameters in different types of deformation – shear, bulk, or simple – and the symbols

Table 1 Viscoelastic parameters from shear, simple extension, and bulk compression

Parameter	Shear	Simple extension	Bulk compression
Stress relaxation modulus	$G(t)$	$E(t)$	$K(t)$
Creep compliance	$J(t)$	$D(t)$	$B(t)$
Storage modulus	$G'(w)$	$E'(w)$	$K'(w)$
Loss modulus	$G''(w)$	$E''(w)$	$K''(w)$
Complex modulus	$G^*(w)$	$E^*(w)$	$K^*(w)$
Dynamic viscosity	$\eta'(w)$	$\eta'_e(w)$	$\eta'_v(w)$
Complex viscosity	$\eta^*(w)$	$\eta^*_e(w)$	$\eta^*_v(w)$

employed for the three types of deformation are given in Table 1.[11]

2.2.1 *Oscillatory Shear Flow*

Small amplitude oscillatory shear (SAOS), also called dynamic rheological experiment, can be used to determine viscoelastic properties of foods. The experiments can be conducted with the Couette, plate and cone, and parallel plate geometries. As noted earlier, uniform shear is achieved in the gap of a cone-plate geometry. However, the narrow gap may not be suitable for studying foods containing solids with dimensions greater than about 100 μm. In the SAOS tests, a food sample is subjected to a small sinusoidally oscillating strain or deformation $\gamma(t)$ at time t according to Equation (25) and the phase difference between the oscillating stress and strain as well as the amplitude ratio are measured. The information obtained should be equivalent to data from a transient experiment at time $t = w^{-1}$ (Equation 30)

$$\gamma(t) = \gamma_0 \sin(wt) \tag{30}$$

where γ_0 is the strain amplitude and w the angular frequency. The applied strain generates two stress components in the viscoelastic material: an elastic component in line with the strain and a 90° out of phase viscous component. Equations (31–33), that define viscoelastic behavior, can be derived:

$$G' = \frac{\sigma_0}{\gamma_0} \cos\delta \tag{31}$$

$$G'' = \frac{\sigma_0}{\gamma_0} \sin\delta \tag{32}$$

$$\tan\delta = \frac{G''}{G'} \tag{33}$$

where G' (Pa) is the storage modulus, G'' (Pa) is the loss modulus, and $\tan\delta$ is the loss tangent.

The storage modulus G' expresses the magnitude of the energy that is stored in the material or recoverable per cycle of deformation; G'' is a measure of the energy that is lost as viscous dissipation per cycle of deformation. Therefore, for a perfectly elastic solid, all the energy is stored, i.e. G'' is zero and the stress and the strain will be in phase. In contrast, for a liquid with no elastic properties all the energy is dissipated as heat, i.e. G' is zero and the stress and the strain will be out of phase by 90° (Figure 8).

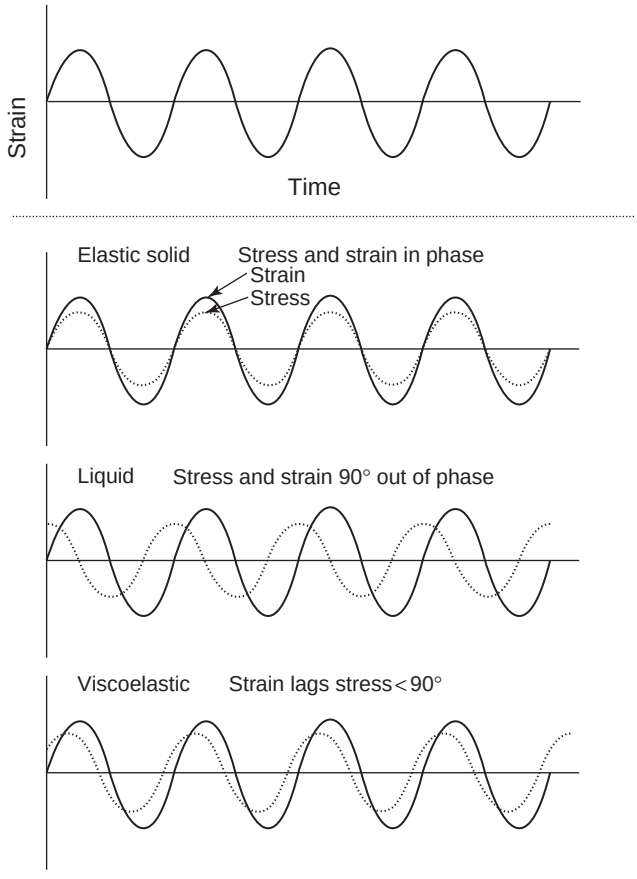

Figure 8 Illustration of strain and stress profiles in dynamic tests. When the strain is the stimulus, the stress is the response and both are recorded.

For a specific food, magnitudes of G' and G'' are influenced by frequency, temperature, and strain. For strain values within the linear range of deformation, G' and G'' are independent of strain. The loss tangent, is the ratio of the energy dissipated to that stored per cycle of deformation. These viscoelastic functions have been found to play important roles in the rheology of structured polysaccharides. One can also employ notation using complex variables and define a complex modulus $G^*(w)$ (Equation 34):

$$|G^*| = \sqrt{(G')^2 + (G'')^2} \tag{34}$$

We note that the dynamic viscosity and the dynamic rigidity are components of the complex dynamic

viscosity η^*:

$$\eta^* = \left(\frac{G^*}{w} \quad = \eta' - i\left(\frac{G'}{w}\right.\right. \tag{35}$$

where in Equation (35) w is the frequency of oscillation; the real part of the complex viscosity η' is equal to (G''/w) and the imaginary part η'' is equal to (G'/w). One can also determine the loss modulus G'', from oscillatory shear data using Equation (36):

$$G'' = w\eta' \tag{36}$$

It should be noted that if G' is much greater than G'', the material will behave more like a solid, i.e. the deformations will be essentially elastic or recoverable. However, if G'' is much greater than G' the energy used to deform the material is dissipated viscously and the materials behavior is liquid like. The viscoelastic moduli determined over a range of frequencies will indicate transition zones corresponding to relaxation processes that depend on the material's structure.[11]

2.2.2 Types of Dynamic Rheological Tests

Dynamic rheological experiments provide suitable means for monitoring the gelation process of many biopolymers and for obtaining insight into gel/food structure because they satisfy several conditions:[26,27]

- they are nondestructive and do not interfere with either gel formation or softening of a structure;
- the time involved in the measurements is short relative to the characteristic times of the gelation and softening processes;
- the results are expressible in fundamental terms so that they can be related to the structure of the network.

For the same reasons, they were used to follow the changes induced on potato cells by cellulase,[28] α-amylase on wheat starch,[27] and can be used in studying other phase transitions in foods.

A strain sweep in which the strain is varied over a range of values is an essential test to determine the linear viscoelastic range. With controlled stress rheometers, a stress sweep can be conducted. The limit of linearity can be detected when dynamic rheological properties (e.g. G' and G'') change rapidly from their almost constant values. Three other types of dynamic rheological tests can be conducted to obtain useful properties of viscoelastic foods, such as gels, and of gelation and melting:

- Frequency sweep studies in which G' and G'' are determined as a function of frequency w at a fixed temperature.
- Temperature sweep studies in which G' and G'' are determined as a function of temperature at fixed w. This test is well suited for studying gel formation during cooling of a heated polymer (e.g. pectin, gelatin) dispersion,[29] gelatinization of a starch dispersion during heating,[30] and gel formation of proteins.[31]
- Time sweep study in which G' and G'' are determined as a function of time at fixed w and temperature. This type of test, often called a gel cure experiment, is well suited for studying structure development in physical gels.

2.3 Relationships Between Steady Shear and Oscillatory Shear Parameters

The relationship between the rheological parameters determined from oscillatory tests such as η^* and η' on one hand and the apparent viscosity (from shear stress and shear rate) on the other is of interest. The superimposition of the shear rate dependence of steady shear viscosity η_a and of the frequency dependence of the complex viscosity η^*, at equal values of frequency and shear rate, was first reported by Cox and Merz[32] for polystyrene samples, and is known as the Cox–Merz rule (Equation 37):

$$\eta^*(w) = \eta_a(w)|_{w=\dot{\gamma}} \tag{37}$$

This empirical correlation has been confirmed experimentally for several synthetic polymers,[11] and for several solutions of random-coil polysaccharides.[33] However, many exceptions to the Cox–Merz rule have also been found for synthetic polymers[34] as well as for biopolymer systems and foods.[2] Relationships between the parameters were determined for equal magnitudes of the shear rate $\dot{\gamma}$ and the frequency w.[35] Based on data on 10 fluid and semisolid foods, power-type relationships were found to be satisfactory between the parameters (Equations 38 and 39):

$$\eta^*(w) = C[\eta(\dot{\gamma})]^{\alpha}|_{\dot{\gamma}=w} \tag{38}$$

$$\eta'(w) = C''[\eta(\dot{\gamma})]^{\alpha'}|_{\dot{\gamma}=w} \tag{39}$$

Similarly, a power-type equation related G'/w^2 and the first normal stress coefficient ψ_1 (Equation 40):

$$\frac{G'}{w^2(w)} = C^*[\Psi_1(\dot{\gamma})]^{\alpha'} \tag{40}$$

Deviation from the Cox–Merz rule appears to be an indication of structural heterogeneities in a food. For example, significant deviations from the Cox–Merz rule were found in dispersed systems, such as tomato

concentrates[36] and cross-linked waxy maize starch dispersions.[37] In contrast, the Cox–Merz rule was found to be applicable to fluids with homogeneous structures, such as dispersions of guar gum and locust bean gum. When departures from the Cox–Merz rule are attributed to structure decay in the case of steady shear, the complex viscosity is usually larger than the steady viscosity.[2] Nevertheless, the relationship between magnitudes of η_a and η^* can be dependent on the strain amplitude used.[38]

3 STIMULI FOR EVALUATION OF VISCOSITY

To describe the mouthfeel of beverages, 136 terms were generated from 5350 responses.[39] Of the 136 terms, 30.7% were viscosity-related terms. In addition, it was suggested that the term ''mouthfeel'' was better than ''viscosity'' that was applicable to liquid foods, whereas the former has the same meaning as ''texture''. Wood[40] suggested that the stimulus associated with oral assessment was the apparent viscosity determined at $50\,s^{-1}$. Shama and Sherman[41] made an important contribution in identifying the stimuli associated with sensory assessment of foods by oral methods. They showed that the stimulus depends on whether one is dealing with a low-viscosity or a high-viscosity food. As illustrated in Figure 9 for low-viscosity foods, the stimulus is the shear rate developed at an almost constant shear stress of 10 Pa. In contrast, for high-viscosity foods, the stimulus is the shear stress developed at an approximately constant shear rate of $10\,s^{-1}$. Other studies on the stimuli associated with nonoral assessment of viscosity and perceived thickness of liquid foods have been reviewed.[2]

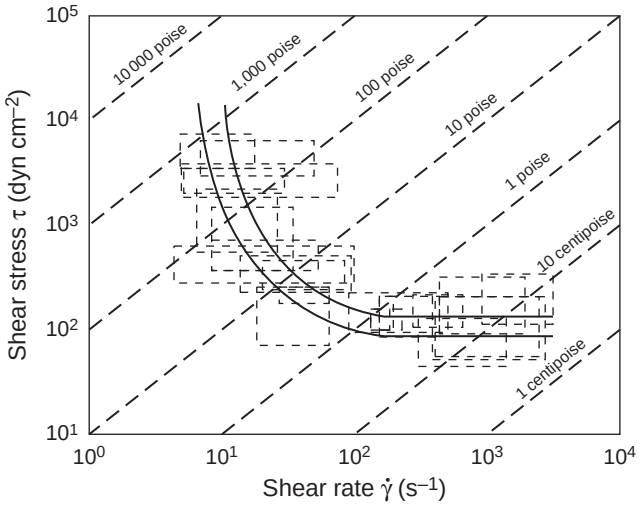

Figure 9 The curved lines represent borders of the stimuli associated with sensory assessment of foods by oral methods. Constant-viscosity lines are also shown (1 P = 0.1 Pa s). (Reproduced by permission of Food and Nutrition Press, Inc., from Sharma and Sherman.[41])

4 VELOCITY PROFILES IN TUBE FLOW

Equations describing velocity profiles can be used, among other applications, to study the effect of different rheological models on the distribution of velocities and to understand the concept of residence time distribution across the cross-section of a pipe or a channel. The velocity profile of a fluid flowing in a tube can be derived from Equation (41),

$$v_z = \int_r^{r_o} \left(\frac{dv_z}{dr} \right) dr \tag{41}$$

where v_z is the velocity in the axial direction, r is the radial coordinate, and r_o is the radius of the tube (Figure 10). A simple force balance for tube flow yields Equation (42),

$$\sigma_{rz} = \sigma_w \frac{r}{r_o} = -\frac{r}{2} \frac{dp}{dz} \tag{42}$$

where σ_{rz} is the shear stress at any radius r, σ_w is the magnitude of the shear stress at the wall ($r = r_o$), p is the pressure, and z is the axial coordinate. Utilizing Equations (36) and (37), and noting that in the rheological equations $\sigma = \sigma_{rz}$ and $\dot{\gamma} = (dv_z/dr)$, one can derive equations describing the velocity profiles for laminar flow in a tube.

The power law model (Equation 4) has been used extensively in handling applications. The relationship between the maximum velocity (v_{zm}) and the average velocity ($\bar{v}$) for the design of the length of a holding tube of a pasteurizing system is of special interest (Equation 43):

$$\frac{v_{zm}}{\bar{v}_z} = \frac{3n+1}{n+1} \tag{43}$$

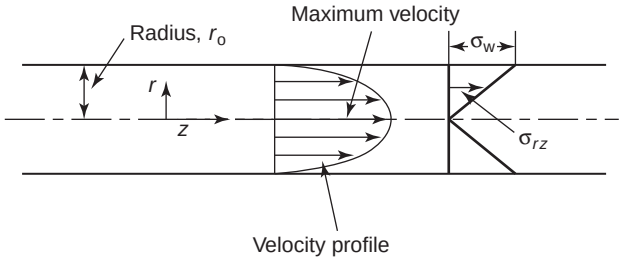

Figure 10 Diagram of the coordinate system for velocity profiles in tube flow.

For Newtonian and pseudoplastic fluids in laminar flow, one can deduce from Equation (38), the popular relationship that maximum velocity, at most, is equal to twice the average velocity. In contrast, in the case of shear thickening (dilatant) fluids, the maximum velocity would be more than twice the average velocity.

The volumetric flow rate Q is given by Equation (44)

$$Q = \int_0^{r_o} 2\pi r v_z \, dr = \pi \int_0^{r_o} v_z(r) \, d(r^2) \qquad (44)$$

Integrating by parts the second part of Equation (39), and using the boundary condition $v = 0$ at $r = r_o$, one can obtain the general equation for the volumetric flow of a fluid in a tube (Equation 45):

$$Q = \left(\frac{\pi r_o^3}{\sigma_w^3}\right) \int_0^{\sigma_w} \sigma_{rz}^2 \left(\frac{dv_z}{dr}\right) d\sigma_{rz} \qquad (45)$$

Table 2 Velocity profile[a] and volumetric flow rate equations for power law, Herschel–Bulkley, and Bingham plastic fluids

Power law model: $\sigma = K\dot{\gamma}^n$

Velocity profile[b]

$$v_z = \left(\frac{n}{n+1}\right)\left(\frac{\Delta p}{2KL}\right)^{(1/n)} r_o^{(n/n+1)} - r^{(n/n+1)}$$

Volumetric flow rate

$$\frac{Q}{\pi r_o^3} = \left(\frac{n}{n+1}\right)\left(\frac{\sigma_w}{K}\right)^{(1/n)}$$

Herschel–Bulkley model: $\sigma - \sigma_0 = K_H(\dot{\gamma})^{n_H}$

Velocity profile[c]

$$v_z = \frac{2L}{\Delta p(m+1)K_H^m}(\sigma_w - \sigma_0)^{(m+1)} - \left(\frac{r\Delta p}{2L} - \sigma_0\right)^{(m+1)}$$

when $r_0 \le r \le r_o$.

Volumetric flow rate

$$\frac{Q}{\pi r_o^3} = \frac{(\sigma_w - \sigma_0)^{(m+1)}}{\sigma_w^3 K_H^m}\left[\frac{(\sigma_w - \sigma_0)^2}{m+3} + \frac{2\sigma_0(\sigma_w - \sigma_0)}{m+2} + \frac{\sigma_0^2}{m+1}\right]$$

Note that in the equations for the Herschel–Bulkley model $m = (1/n_H)$.

Bingham plastic model: $\sigma - \sigma_0 = \eta'\dot{\gamma}$

Velocity profile

$$v_z = \frac{1}{\eta'}\left[\frac{\Delta p}{L}(r_o^2 - r^2) - \sigma_0(r_o - r)\right]$$

when $r_0 \le r \le r_o$.

Volumetric flow rate

$$\frac{4Q}{\pi r_o^3} = \frac{\sigma_w}{\eta'}\left[1 - \frac{4}{3}\left(\frac{\sigma_0}{\sigma_w}\right) + \frac{1}{3}\left(\frac{\sigma_0}{\sigma_w}\right)^4\right]$$

[a] The average velocity can be obtained by dividing the equation for volumetric flow rate by the area of cross-section of the pipe.
[b] Maximum velocity occurs at the center line, $r = 0$.
[c] Velocity profiles are valid for $r_o \le r \le R$, where the radius of plug $r_o = (2\sigma_0 L/\Delta p)$; the maximum velocity occurs when $0 \le r \le r_o$ and is obtained by substituting r_o for r.

Substituting for the shear rate appropriate expressions from different rheological models, one can derive equations relating Q and pressure drop Δp.[2] Table 2 contains expressions for the velocity profiles and the volumetric flow rates of the three rheological models: the power law, Herschel–Bulkley, and the Bingham plastic models.

In the case of fluids obeying the power-law model, the pressure drop per unit length $\Delta p/L$ is related to Q and r_o by the relationship:

$$\frac{\Delta p}{L} \propto \frac{Q^n}{r_o^{3n+1}} \qquad (46)$$

From Equation (46) it can be seen that for Newtonian foods ($n = 1$) the pressure gradient is proportional to the $(r_o)^{-4}$ power. Therefore, a small increase in the radius of the tube will result in a major reduction in the magnitude of the pressure gradient. In contrast, for a highly pseudoplastic fluid (e.g. $n = 0.2$), increasing the pipe radius does not have such a profound effect on the pressure gradient.

5 ENERGY REQUIREMENTS FOR PUMPING

5.1 Mechanical Energy Balance Equation

The energy required to pump a liquid food through a pipeline can be calculated from the mechanical energy balance (MEB) equation. The MEB equation can be used to analyze pipe flow systems. For the steady-state flow of an incompressible fluid, the MEB can be written as Equation (47):[2]

$$gZ_1 + \frac{P_1}{\rho} + \frac{v_{z1}^2}{\alpha} - W = gZ_2 + \frac{P_2}{\rho} + \frac{v_{z2}^2}{\alpha} + E_f \qquad (47)$$

where g is the acceleration due to gravity, Z is the height above a reference point, P is the pressure, v is the fluid velocity, W is the work output per unit mass, E_f is the energy loss per unit mass, α is the kinetic energy correction factor, and the subscripts 1 and 2 refer to two points in the pipe system. In order to accurately estimate the energy required for pumping of a food in a specific piping and equipment system, the term $(-W)$ $(J\,kg^{-1})$ has to be estimated from Equation (47) after the other terms have been evaluated.[2,42]

The velocities at the entrance and exit of the system can be calculated from the volumetric flow rate of the food, and the respective diameters of the tanks or pipes. The energy loss term E_f consists of losses due to friction in pipe and that due to friction in valves and fittings:

$$E_f = \frac{2fv_z^2 L}{D} + \sum_1^b \frac{k_f v_z^2}{2} \qquad (48)$$

where f is the pipe friction factor, v_z is the velocity, L is the length of straight pipe of diameter D, k_f is the friction coefficient for a fitting, and b is the number of valves or fittings. It is emphasized that k_f is unique to a particular fitting and that different values of u, k_f, and f may be required when the system contains pipes of different diameters. Further, losses due to special equipment, such as heat exchangers, must be added to E_f.[42]

5.2 Friction Losses in Pipes

Because many fluid foods are non-Newtonian in nature, estimation of friction losses (pressure drop) for these fluids in straight pipes is of interest. One can estimate the friction losses in straight pipes and tubes can be estimated from the magnitude of the fanning friction factor f (Equation 49):

$$\frac{\Delta P_f}{\rho} = \frac{2fLv_z^2}{D} \tag{49}$$

For fluids that can be described by the power law (Equation 4), the generalized Reynolds number (GRe) is calculated from Equation (50):

$$\mathrm{GRe} = \frac{D^n v_z^{2-n}\rho}{8^{(n-1)}K}\left(\frac{4n}{3n+1}\right)^n \tag{50}$$

where D is the pipe diameter, v_z is the average velocity, ρ is the density, and K and n are the power law parameters of the food. When $n = 1$, Equation (50) reduces to the Reynolds number for Newtonian foods:

$$\mathrm{Re} = \frac{Dv_z\rho}{\eta} \tag{51}$$

In laminar tube flow, the fanning friction factor can be calculated from Equation (52):

$$f = \frac{16}{\mathrm{GRe}} \tag{52}$$

This relationship can also be used for Newtonian fluids in laminar flow with the Reynolds number being used in place of GRe. In fact, the Newtonian f versus Re relationship was established much earlier than extension to non-Newtonian fluids. Once the magnitude of the friction factor is known, the pressure drop in a pipe can be estimated from Equation (49).

5.2.1 Friction Loss Coefficients for Fittings

Steffe et al.[43] determined magnitudes of the coefficient of a fully open plug valve, a tee with flow from line to branch, and a 90° short elbow as a function of GRe using apple sauce as the test fluid. They found that, as for Newtonian fluids, k_f increases with decreasing values of GRe. The regression equations for the three fittings are given by Equations (53–55):

$$\text{three-way plug valve:}\quad k_f = 30.3\,\mathrm{GRe}^{-0.492} \tag{53}$$

$$\text{tee:}\quad k_f = 29.4\,\mathrm{GRe}^{-0.504} \tag{54}$$

$$\text{elbow:}\quad k_f = 191.0\,\mathrm{GRe}^{-0.896} \tag{55}$$

In many instances, the practice is to employ values determined for Newtonian fluids, such as those in the *Chemical Engineers' Handbook*.[44]

LIST OF SYMBOLS

c	polymer or solute concentration (w w^{-1})
c^*	coil overlap concentration (w w^{-1})
cQ	starch granule mass fraction
D	starch granule and pipe diameter
E_a	activation energy of flow (J mol^{-1})
E_f	energy loss per unit mass (J kg^{-1})
E_g	activation energy of gelatinization (J mol^{-1})
E_i	elastic modulus (Pa)
E_{ak}, $E_{a\eta}$	activation energy for viscous flow
f	Fanning friction factor, dimensionless
G'	storage modulus (Pa)
G''	loss modulus (Pa)
$G^* = \sqrt{(G')^2 + (G'')^2}$	complex modulus (Pa)
K, K_H	consistency index (Pa s^n)
k_f	is friction coefficient of a fitting
L	length of tube/pipe (m)
n, n_H	flow behavior index, dimensionless
p	pressure (Pa)
r_i	radius of inner cylinder (m)
r_o	radius of outer cylinder; particle radius (m)
R	gas constant (8.314 J mol^{-1} K^{-1})
T	temperature of fluid (°C, K)
v_z	velocity (m s^{-1})
W	work per unit mass (J kg^{-1})
Z_1 and Z_2	heights above a reference point (m)

Greek Letters

δ	phase angle (rad)
$\gamma(t)$	oscillatory strain or deformation
$\dot{\gamma}$	shear rate (s^{-1})
γ_0	strain amplitude
η_a	apparent viscosity (Pa s)
$\eta_{Ca} = (K_c)^2$	Casson viscosity (Pa s)
η_o	zero shear viscosity (Pa s)
$[\eta]$	intrinsic viscosity ($dL\,g^{-1}$)
$\eta^* = \left(\frac{G^*}{w}\right.$	complex viscosity (Pa s)
η_∞	infinite shear viscosity (Pa s)
ϕ	volume fraction, $\phi = cQ$ for starch dispersions
ϕ_m	maximum volume fraction
ρ	density ($kg\,m^{-3}$)
σ	shear stress (Pa)
$\sigma_{0c} = (K_{0c})^2$	Casson yield stress (Pa)

Dimensionless Numbers

Re	Reynolds number, $\frac{Dv_z\rho}{\eta}$
GRe	pipe flow generalized Reynolds number $\mathrm{GRe} = \frac{D^n v_z^{2-n}\rho}{8^{(n-1)}K}\left(\frac{4n}{3n+1}\right.^n$

ABBREVIATIONS AND ACRONYMS

COJ	Concentrated Orange Juice
MEB	Mechanical Energy Balance
SAOS	Small Amplitude Oscillatory Shear

RELATED ARTICLES

Coatings **(Volume 2)**
Rheology in Coatings, Principles and Methods

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Mechanical Properties of Polymers and Rubbers

REFERENCES

1. S.D. Holdsworth, 'Rheological Models Used for the Prediction of the Flow Properties of Food Products: A Literature Review', *Trans. Instit. Chem. Eng. C*, **71**, 139–179 (1993).
2. M.A. Rao, *Rheology of Fluid and Semisolid Foods: Principles and Applications*, Aspen Publishers, Gaithersburg, MD, 1999.
3. H.A. Barnes, K. Walters, 'The Yield Stress Myth?', *Rheol. Acta*, **24**, 323–326 (1985).
4. M.A. Rao, H.J. Cooley, 'Applicability of Flow Models with Yield for Tomato Concentrates', *J. Food Process Eng.*, **6**, 159–173 (1983).
5. N. Casson, 'A Flow Equation for Pigment-oil Suspensions of the Printing Ink Type', in *Rheology of Disperse Systems*, ed. C.C. Mill, Pergamon Press, New York, 82–104, 1959.
6. E.R. Morris, S.B. Ross-Murphy, 'Chain Flexibility of Polysaccharides and Glicoproteins from Viscosity Measurements', in *Techniques in Carbohydrate Metabolism*, ed. D.H. Northcote, Elsevier, Amsterdam, B310, 1–46, 1981.
7. J.A.L. Lopes da Silva, M.A. Rao, 'Viscoelastic Properties of Food Gum Dispersions', in *Viscoelastic Properties of Food*, eds. M.A. Rao, J.F. Steffe, Elsevier Applied Science Publishers, London, 285–316, 1992.
8. T. Tanglertpaibul, M.A. Rao, 'Intrinsic Viscosity of Tomato Serum as Affected by Methods of Determination and Methods of Processing Concentrates', *J. Food Sci.*, **52**, 1642–1645, 1688 (1987).
9. K.C. Tam, C. Tiu, 'Improved Correlation for Shear-dependent Viscosity of Polyelectrolyte Solutions', *J. Non-Newtonian Fluid Mech.*, **46**, 275–288 (1993).
10. B. Launay, J.L. Doublier, G. Cuvelier, 'Flow Properties of Aqueous Solutions and Dispersions of Polysaccharides', in *Functional Properties of Food Macromolecules*, eds. J.R. Mitchell, D.A. Ledward, Elsevier Applied Science Publishers, London, 1–78, 1986.
11. J.D. Ferry, *Viscoelastic Properties of Polymers*, John Wiley, New York, 1980.
12. W.W. Graessley, 'Polymer Chain Dimensions and the Dependence of Viscoelastic Properties on Concentration, Molecular Weight and Solvent Power', *Polymer*, **21**, 258–315 (1980).
13. E.R. Morris, A.N. Cutler, S.B. Ross-Murphy, D.A. Rees, 'Concentration and Shear Rate Dependence of Viscosity in Random Coil Polysaccharide Solutions', *Carbohydrate Polymers*, **1**, 5–21 (1981).
14. A.A. Vitali, M.A. Rao, 'Flow Properties of Low-pulp Concentrated Orange Juice: Effect of Temperature and Concentration', *J. Food Sci.*, **49**, 882–888 (1984).
15. A.A. Vitali, M.A. Rao, 'Flow Properties of Low-pulp Concentrated Orange Juice: Serum Viscosity and Effect of Pulp Content', *J. Food Sci.*, **49**, 876–881 (1984).
16. I.J. Krieger, 'Rheology of Polymer Colloids', in *Polymer Colloids*, eds. R. Buscall, T. Corner, J.F. Stageman, Elsevier Applied Science, New York, 219–246, 1985.
17. S.C. Tsai, K. Zammouri, 'Role of Interparticular van der Waals Force in Rheology of Concentrated Suspensions', *J. Rheol.*, **32**, 737–750 (1988).

18. B. Yoo, M.A. Rao, 'Effect of Unimodal Particle Size and Pulp Content on Rheological Properties of Tomato Puree', *J. Texture Stud.*, **25**, 421–436 (1994).
19. P.C. Hiemenz, R. Rjagopalan, *Principles of Colloid and Surface Chemistry*, 3rd edition, Marcel Dekker, New York, 1997.
20. I.D. Evans, D.R. Haisman, 'Rheology of Gelatinized Starch Suspensions', *J. Texture Stud.*, **10**, 347–370 (1979).
21. E.B. Bagley, D.D. Christianson, 'Swelling Capacity of Starch and its Relationship to Suspension Viscosity: Effect of Cooking Time, Temperature and Concentration', *J. Texture Stud.*, **13**, 115–126 (1982).
22. M.A. Rao, H.J. Cooley, A.A. Vitali, 'Flow Properties of Concentrated Fruit Juices at Low Temperatures', *Food Technol.*, **38**(3), 113–119 (1984).
23. M.A. Rao, R.S. Shallenberger, H.J. Cooley, 'Effect of Temperature on Viscosity of Fluid Foods with High Sugar Content', in *Engineering and Food*, eds. M. LeMaguer, P. Jelen, Elsevier Applied Science Publishers, New York, 23–31, Vol. 1, 1986.
24. T. Soesanto, M.C. Williams, 'Volumetric Interpretation of Viscosity for Concentrated and Dilute Sugar Solutions', *J. Phys. Chem.*, **85**, 3338–3341 (1981).
25. R.W. Whorlow, *Rheological Techniques*, Halsted Press, New York, 1980.
26. J.A. Lopes da Silva, M.A. Rao, J.-T. Fu, 'Rheology of Structure Development and Loss During Gelation and Melting', in *Phase/State Transitions in Foods: Chemical, Rheological and Structural Changes*, eds. M.A. Rao, R.W. Hartel, Marcel Dekker, NY, 111–156, 1998.
27. Y.C. Champenois, M.A. Rao, L.P. Walker, 'Influence of α-Amylase on the Viscoelastic Properties of Starch–Gluten Pastes and Gels', *J. Sci. Food Agric.*, **78**, 127–133 (1998).
28. I. Shomer, M.A. Rao, M.C. Bourne, D. Levy, 'Rheological Behaviour of Potato Tuber Cell Suspensions During Temperature Fluctuations and Cellulase Treatments', *J. Sci. Food. Agric.*, **63**, 245–250 (1993).
29. M.A. Rao, H.J. Cooley, 'Dynamic Rheological Measurement of Structure Development in High-methoxyl Pectin/Fructose Gels', *J. Food Sci.*, **58**, 876–879 (1993).
30. W.H. Yang, M.A. Rao, 'Complex Viscosity–Temperature Master Curve of Cornstarch Dispersion During Gelatinization', *J. Food Proc. Eng.*, **21**, 191–207 (1998).
31. S.R. Owen, M.A. Tung, A.T. Paulson, 'Thermorheological Studies of Food Polymer Dispersions', *J. Food Eng.*, **16**, 39–53 (1992).
32. W.P. Cox, E.H. Merz, 'Correlation of Dynamic and Steady Flow Viscosities', *J. Polym. Sci.*, **28**(118), 619–622 (1958).
33. E.R. Morris, 'Rheology of Hydrocolloids', in *Gums and Stabilisers for the Food Industry*, 2nd edition, eds. G.O. Philips, D.J. Wedlock, P.A. Williams, Pergamon Press, Oxford, UK, 57–78, 1981.
34. W.M. Kulicke, R.S. Porter, 'Relation Between Steady Shear Flow and Dynamic Rheology', *Rheologica Acta*, **19**, 601–605 (1980).
35. K.L. Bistany, J.L. Kokini, 'Dynamic Viscoelastic Properties of Foods in Texture Control', *J. Rheol.*, **27**, 605–620 (1983).
36. M.A. Rao, H.J. Cooley, 'Rheology of Tomato Pastes in Steady and Dynamic Shear', *J. Texture Stud.*, **23**, 415–425 (1992).
37. P.M.S. Da Silva, J.C. Oliveira, M.A. Rao, 'The Effect of Granule Size Distribution on the Rheological Behavior of Heated Modified and Unmodified Maize Starch Dispersions', *J. Texture Stud.*, **28**, 123–138 (1997).
38. J.A. Lopes da Silva, M.P. Gonçalves, M.A. Rao, 'Viscoelastic Behavior of Mixtures of Locust Bean Gum and Pectin Dispersions', *J. Food Eng.*, **18**, 211–228 (1993).
39. A.S. Szczesniak, 'Classification of Mouthfeel Characteristics of Beverages', in *Food Texture and Rheology*, ed. P. Sherman, Academic Press, New York, 1979.
40. F.W. Wood, 'Psychophysical Studies on the Consistency of Liquid Foods', in *Rheology and Texture of Foodstuffs*, Society of Chemical Industry, London, 40–49, 1968.
41. F. Shama, P. Sherman, 'Identification of Stimuli Controlling the Sensory Evaluation of Viscosity. II. Oral Methods', *J. Texture Stud.*, **4**, 111–118 (1973).
42. J.F. Steffe, R.G. Morgan, 'Pipeline Design and Pump Selection for Non-Newtonian Fluid Foods', *Food Technol.*, **40**(12), 78–85 (1986).
43. J.F. Steffe, I.O. Mohamed, E.W. Ford, 'Pressure Drop Across Valves and Fittings for Pseudoplastic Fluids in Laminar Flow', *Trans. Am. Soc. Agric. Eng.*, **27**, 616–619 (1984).
44. R.H. Perry, C.H. Chilton, 'Chemical Engineer's Handbook', McGraw-Hill Book Company, New York, 1973.

Vitamins Analysis in Food

Lin Ye and Ronald R. Eitenmiller
University of Georgia,
Athens, USA

1 Introduction **1**
2 Chemistry **1**
2.1 Vitamin A and Carotenoids 1
2.2 Vitamin D 2
2.3 Vitamin E 2
2.4 Vitamin K 5
2.5 Vitamin C 5
2.6 Thiamin (Vitamin B_1) 5
2.7 Riboflavin (Vitamin B_2) 5
2.8 Niacin 6
2.9 Vitamin B_6 6
2.10 Folate 6
2.11 Vitamin B_{12} 10
2.12 Biotin 10
2.13 Pantothenic Acid 10
3 Analysis of Water-soluble Vitamins **10**
3.1 Microbiological Methods 10
3.2 Chromatographic Methods 22
4 Analysis of Fat-soluble Vitamins **22**
4.1 Chromatographic Methods 22
4.2 Multianalyte Methods 24
Abbreviations and Acronyms **33**
Related Articles **33**
References **33**

Vitamin assay of biological materials represents a large and rapidly expanding area. Methods now rely heavily on high-performance liquid chromatography (HPLC). However, microbiological assay of the water-soluble vitamins is still routinely used for food analysis owing to the sensitivity and specificity of the methods. An extensive table on physical properties of the vitamins is provided. Such information is dispersed widely in the literature and is often not at the analyst's fingertips when needed. This article concentrates on providing the reader with current HPLC approaches to vitamin analysis of food. Tables are provided that give method summaries referenced to the original work. Because of the scope of the area, recent text references are provided for readers requiring more in-depth knowledge.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Significant advances have been made in vitamin assay methodology since the 1970s. Older, established methods include chemical methods such as the thiochrome method for thiamin and the various chemical procedures available for ascorbic acid (AA). Such methods, together with microbiological methods still comprise the bulk of accepted methods of vitamin analysis available in methodology sources, including *Official Methods of Analysis*, 16th edition [Association of Official Analytical Chemists (now known as AOAC International) (AOAC) International, Arlington, VA, 1995]. Most advances in vitamin quantification have come through use of HPLC. At present, HPLC can be used for quantification of each of the vitamins depending upon the matrix and concentration of the vitamin. For each of the water-soluble vitamins except AA, microbiological assay is routinely used for food analysis. Since the field is so large, we have attempted in this article to present an overview of useful approaches to vitamin assay. Realizing the scope of the subject, our concentration is on HPLC assays. Reference is made to several recent texts for readers needing more complete information.

2 CHEMISTRY

2.1 Vitamin A and Carotenoids

2.1.1 Vitamin A

The structure of all-*trans*-retinol (vitamin A) is given in Figure 1. Vitamin A refers to all isoprenoid compounds from animal products that possess the biological activity of all-*trans*-retinol which could prevent night blindness and xerophthalmia. The parent vitamin A compound, retinol, contains a β-ionone (cyclohexenyl) ring attached at the C-6 position to a side chain composed of two isoprene units. The four double bonds in the polyene side chain give rise to *cis–trans* (*Z–E*) isomerism. Retinoids include all substances with vitamin A activity, some of which differ structurally from all-*trans*-retinol. Retinyl acetate and retinyl palmitate are used commercially in synthetic form by the pharmaceutical and food industry. Esterification greatly stabilizes the vitamin toward oxidation.

Vitamin A has strong ultraviolet (UV) absorption properties owing to the conjugated double bond system. Absorption maxima vary depending on the solvent and the presence of *cis*-isomers (*Z*-). Isomeration to the *cis* form lowers the absorption maximum and molar absorptivity ($A^{1\%}_{1\,cm}$) values relative to all-*trans*-retinol.[1] Retinol and retinyl esters exhibit strong native fluorescence with excitation and emission maxima at wavelengths of 325–330 and 470–490 nm, respectively.[2]

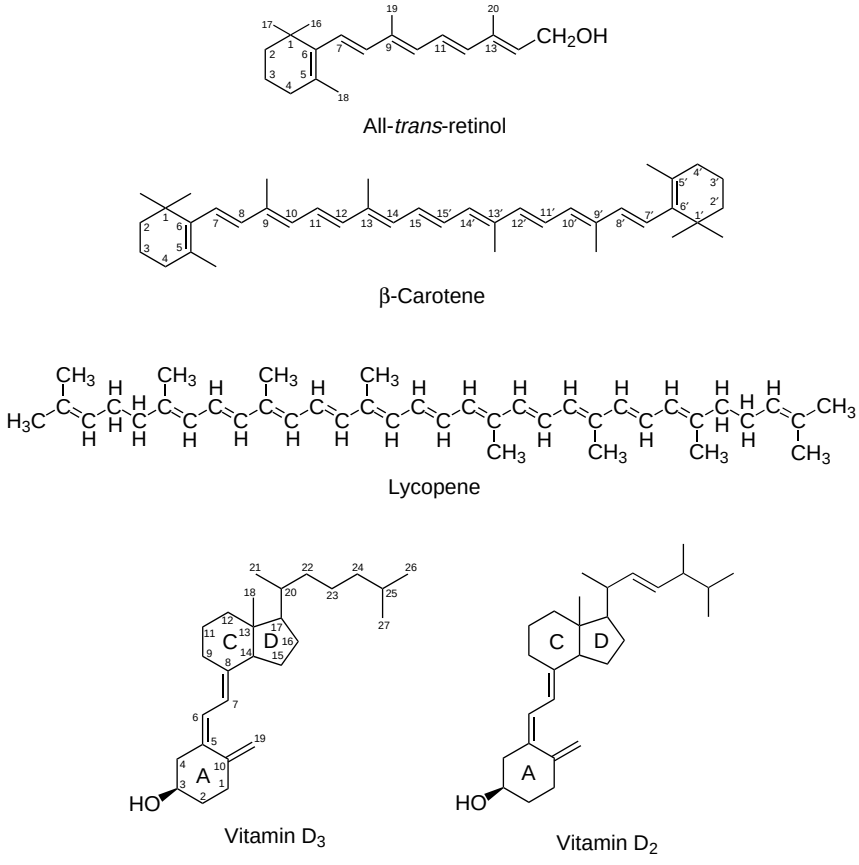

Figure 1 Selected structures of vitamin A, carotenoids and vitamin D.

Physical properties of all-*trans*-retinol, all-*trans*-retinyl palmitate and all-*trans*-retinal are given in Table 1.

2.1.2 Carotenoids

Carotenoids are classified chemically as carotenes, which are hydrocarbons and xanthophylls. The structures are formed by the head-to-tail linkages of isoprene units to provide a C_{40} skeleton. Lycopene (Figure 1) shows the acyclic hydrocarbon backbone chain. Structural modifications of lycopene lead to the diverse nature of the carotenoids present in the plant kingdom. The nutritionally most important carotenoid, β-carotene, is composed of two molecules of retinol joined tail to tail (Figure 1); thus, the compound possesses maximum provitamin A activity. Addition of oxygenated functions to the molecule yields the xanthophylls. These oxygen-containing groups include hydroxyl, carbonyl, carboxylic acid, ester, epoxide, glycoside and ether.[3]

Physical properties of β-carotene are given in Table 1.

2.2 Vitamin D

Vitamin D is the inclusive term for steroids that are antirachitic. There are only two D vitamins of significance nutritionally and biologically, vitamin D_2 and D_3, which are also known as ergocalciferol and cholecalciferol, respectively. The structures of vitamin D_2 and D_3 are given in Figure 1. Vitamin D_2 and vitamin D_3 differ structurally by only a double bond and an additional methyl group in the side chain of vitamin D_2.

All vitamin D forms show a characteristic broad UV spectrum with a maximum absorption near 264 nm and a minimum near 228 nm.[4] Vitamin D does not fluoresce.[5]

Physical properties of vitamin D are given in Table 1.

2.3 Vitamin E

Vitamin E includes Ts and T3s (Figure 2) that possess the biological activity of α-T. They occur in nature as a group of substances including eight chemically similar compounds which are derivatives of 6-chromanol. The Ts have a phytol side chain and the T3s have a similar structure with double bonds at the 3′, 7′ and 11′ positions of the side chain.[6,7] The four homologs of Ts and T3s differ in the number and location of the methyl groups on the chroman ring. Because of the presence of three asymmetric carbon atoms at position 2 of the chroman ring and at the 4′ and 8′ carbons of the side chain, the Ts

Table 1 Physical properties of fat-soluble vitamins[a]

Substance	Formula	Molar mass	Solubility	Melting point (°C)	Crystal form	UV absorption			Fluorescence[b]	
						λ_{max} (nm)	$A^{1\%}_{1\,cm}$	Solvent	λ_{ex} (nm)	λ_{em} (nm)
Vitamin A										
All-*trans* retinol (CAS 68-26-8)	$C_{20}H_{30}O$	286.46	Soluble in absolute alcohol, methanol, chloroform, ether, fats, oils	62–64	Yellow prisms	325 325	1845 1810	Ethanol Hex	**325**	**470**
All-*trans* retinyl palmitate (CAS 79-81-2)	$C_{36}H_{60}O_2$	524.88		28–29	Amorphous or crystalline	325	940	Ethanol	**325**	**470**
All-*trans* retinal (CAS 116-31-4)	$C_{20}H_{28}O$	284.44	Soluble in ethanol, chloroform, cyclohexane, PE, oils	61–64	Orange crystals	383 368	1510 1690	Ethanol Hex		
Carotenonids										
β-carotene (CAS 7488-99-5)	$C_{40}H_{56}$	536.88	Soluble in CS_2, benzene, chloroform	183	Red rhombic square leaflets	425 453 479 452	– 2592 – 2505	PE PE PE Cyclohexane		
Vitamin D										
Vitamin D_3 (CAS 67-97-0)	$C_{27}H_{44}O$	384.65	Soluble in most organic solvents; insoluble in water	84–85	Fine needles yellow to white	264	485	Ethanol		
Vitamin D_2 (CAS 50-14-6)	$C_{28}H_{44}O$	396.66	Soluble in most organic solvents; insoluble in water	115–118	Prisms yellow to white	264	462	Ethanol		
Vitamin E										
α-T (CAS 59-02-9)	$C_{29}H_{50}O_2$	430.71	Freely soluble in oils, fats, acetone, chloroform, ether, alcohol, other fat solvents; insoluble in water			292	75.8	Ethanol	295	320
β-T (CAS 148-03-8)	$C_{28}H_{48}O_2$	416.69				296	89.4	Ethanol	297	322

(*continued overleaf*)

Table 1 (*continued*)

Substance	Formula	Molar mass	Solubility	Melting point (°C)	Crystal form	UV absorption			Fluorescence[b]	
						λ_{max} (nm)	$A^{1\%}_{1\,cm}$	Solvent	λ_{ex} (nm)	λ_{em} (nm)
γ-T (CAS 7616-22-0)	$C_{28}H_{48}O_2$	416.69	Freely soluble in oils, fats, acetone, chloroform, ether, alcohol, other fat solvents; insoluble in water			297	91.0	Ethanol	297	322
δ-T (CAS 119-13-1)	$C_{27}H_{40}O_2$	402.66				297	87.3	Ethanol	297	322
α-T3 (CAS 2265-13-4)	$C_{29}H_{44}O_2$	424.67				292	86.0	Ethanol	290	323
β-T3 (CAS 49-23-3)	$C_{28}H_{42}O_2$	410.64				296	86.2	Ethanol	290	323
γ-T3 (CAS 14101-61-2)	$C_{28}H_{42}O_2$	410.64				297	91.0	Ethanol	290	324
δ-T3 (CAS 26512-59-3)	$C_{27}H_{40}O_2$	396.61				297	85.8	Ethanol	292	324
Vitamin K										
Phylloquinone (K_1) (CAS 84-80-0)	$C_{31}H_{46}O_2$	450.71	Sparingly soluble in methanol; soluble in ethanol, acetone, benzene, PE, Hex, $CHCl_3$, ether		None; yellow oil	242 248 260 269 325	396 419 383 387 68	PE PE PE PE PE		
MK-4 (Vitamin K_2)	$C_{31}H_{40}O_2$	444.70		35	Yellow crystals	248	439	PE		
Menadione (Vitamin K_3) (CAS 58-27-5)	$C_{11}H_8O_2$	172.18	Insoluble in water; moderately soluble in $CHCl_3$	105–107	Bright yellow crystals					

Source: references.[1,4,21–26]

[a] Table adapted from Eitenmiller and Landen.[5]

[b] In hexane (Hex); bold values in 2-propanol (IPA).

T3, tocotrienol; T, tocopherol; PE, petroleum ether; MK-4, menaquinone-4.

can exist in eight diastereomeric forms. T3s with double bonds in the side chain can form *cis* and *trans* isomers.

Physical properties of vitamin E homologs are given in Table 1.

2.4 Vitamin K

Vitamin K is used as a generic descriptor for 2-methyl-1,4-naphthoquinone and all its derivatives providing the antihemorrhagic activity of vitamin K_1, phylloquinone.[8] Figure 2 shows the structures of vitamin K_1, vitamin K_2 and vitamin K_3. Vitamin K_3, known as menadione, is the synthetic and simplest form of vitamin K. Vitamin K_1 is synthesized by green plants in the chloroplasts and vitamin K_2 is synthesized by bacteria.[9] Alkylation at position 3 of the 2-methyl-1,4-naphthoquinone ring with five-carbon isoprenoid units produces the vitamin K series found in nature, $K_{1(20)}$, the most common being phylloquinone. Vitamin K_2 refers to a family of structural analogs called the menaquinone-*n* series (MK-*n*) (Figure 2), the predominant members of which have side chains composed of 6–10 unsaturated isoprene units.[9]

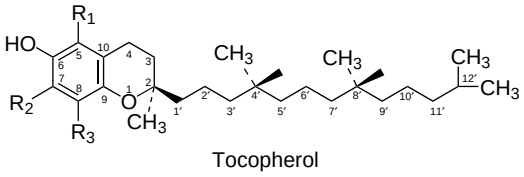

Tocopherol or tocotrienol	R_1	R_2	R_3
α-5,7,8-Trimethyl	CH_3	CH_3	CH_3
β-5,8-Dimethyl	CH_3	H	CH_3
γ-7,8-Dimethyl	H	CH_3	CH_3
δ-8-Methyl	H	H	CH_3

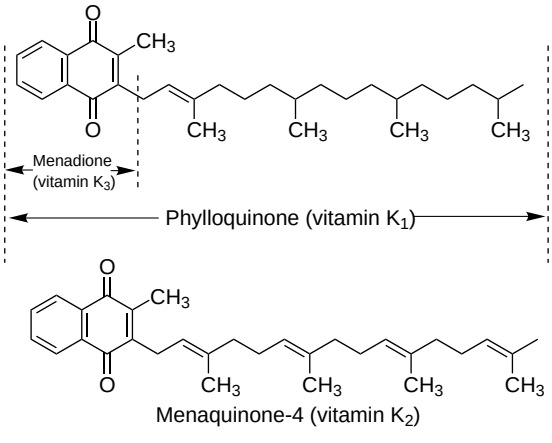

Figure 2 Selected structures of vitamin E and vitamin K.

Physical properties of phylloquinone, MK-4 and menadione are given in Table 1.

2.5 Vitamin C

Vitamin C refers to compounds exhibiting full or partial biological activity of L-AA that prevent and cure scurvy. These include esters of AA such as ascorbyl palmitate, 100% relative activity, synthetic forms (6-deoxy-L-AA, 33% relative activity) and the primary oxidized form of L-AA, dehydroascorbic acid (DHAA), with 80% relative activity.[10] The structures of L-AA and related compounds are given in Figure 3. L-AA and DHAA are the primary dietary sources of vitamin C. Since animals efficiently reduce DHAA to L-AA, both forms must be measured to quantitate accurately "total" vitamin C activity of food or other biological samples.[11] L-AA is a widely used food additive with many functional roles, many of which are based upon its oxidation–reduction properties.

The UV absorption of L-AA is strongly pH dependent: at pH 2 the absorption maximum is at 245 nm and at pH 6.4 it is at 265 nm.[12] L-AA does not fluoresce, but its derivatization with *o*-phenylenediamine (OPD) gives a highly fluorescent product which is used advantageously in chemical and HPLC methods. Physical properties of L-AA are given in Table 2.

2.6 Thiamin (Vitamin B_1)

Thiamin is characterized by a pyrimidine linked by a methylene bridge to a substituted thiazole (Figure 3). The naturally occurring thiamin exists as thiamin pyrophosphate (TPP) (Figure 3), whereas thiamin hydrochloride and thiamin mononitrate are the commercially available forms used in pharmaceuticals and for food fortification. Physical properties of the thiamin salts are given in Table 2. The most distinguishing difference between the hydrochloride salt and the mononitrate is water solubility. The hydrochloride is soluble in water ($1\,g\,mL^{-1}$) and the mononitrate is only slightly water-soluble ($0.027\,g\,mL^{-1}$).[5]

The absorption spectrum of thiamin hydrochloride is pH dependent. At pH 2.9, a single maximum is present at 246 nm. At pH levels above 5 but below neutrality, two absorption maxima exist at 234 and 264 nm, representing the pyrimidine and thiazole rings, respectively.[13]

2.7 Riboflavin (Vitamin B_2)

Riboflavin, 7,8-dimethyl-10-(1′-D-ribityl)isoalloxazine, is the generic term for the group of compounds that exhibit the biological activity of riboflavin (Figure 3). Phosphorylation of the 5′-position of the ribityl side

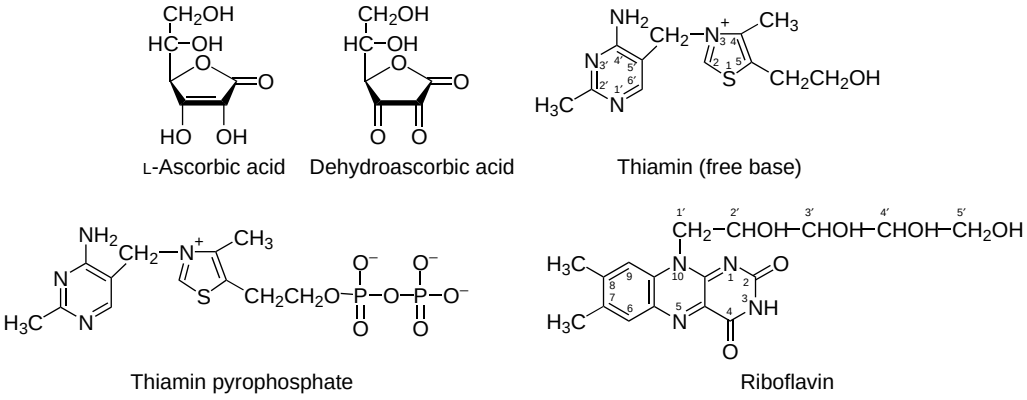

Figure 3 Selected structures of AA, thiamin and riboflavin.

chain yields flavin mononucleotide (FMN) and addition of 5′-adenosyl monophosphate moiety yields FAD. The carbon–nitrogen bond linking the ribityl side chain to the isoalloxazine withstands hydrolysis, whereas FMN and FAD are easily converted to riboflavin below pH 5.0.[14] Therefore, analytical procedures often incorporate acid hydrolysis as the first step, with subsequent quantitation of total riboflavin. In contrast, acid hydrolysis must be avoided in the quantitation of the coenzyme forms. Riboflavin is a yellow-green, light-sensitive compound and is widely distributed in animal and plant cells.[15]

The characteristic spectral properties of riboflavin, FMN and FAD provide the basis for analytical methods for their determination. Most chemical and HPLC methods capitalize on the strong native fluorescence of the flavins with excitation wavelength from 440 to 500 nm and emission wavelength from 520 to 530 nm. In aqueous solutions, riboflavin shows absorption maxima of 223, 266, 373 and 445 nm.[9] Physical properties of riboflavin, FMN and FAD are given in Table 2.

2.8 Niacin

Niacin is the generic descriptor for nicotinic acid and nicotinamide, which have equal biological activities. The coenzyme forms of niacin are nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). Nicotinic acid has been designated as pyridine-3-carboxylic acid and nicotinamide is known as pyridine-3-carboxamide (Figure 4). Both compounds are white, needle-shaped crystals. Nicotinamide is more water-soluble (100 g per 100 mL) than nicotinic acid (1.67 g per 100 mL).[16]

Nicotinic acid and nicotinamide show similar absorption properties with an absorption maximum near 260 nm. The absorption intensity is strongly affected by pH. The absorptivity is higher in acidic than in alkaline solution.[13] Niacin does not fluoresce; however, the coenzymes fluoresce at 470 nm when excited at either 260 or 340 nm.[4] Physical properties of nicotinic acid and nicotinamide are listed in Table 2.

2.9 Vitamin B_6

Vitamin B_6 is a generic term for the group of 2-methyl-3-hydroxy-5-hydroxymethylpyridine compounds that possess the biological activity of PN in rats.[5] There are six known B_6 vitamers in nature: PN or pyridoxol, pyridoxal (PL), PM, with alcohol, aldehyde and amine groups in the 4-position, respectively. The phosphorylation at the 5′-hydroxymethyl group of PN, PL and PM yield pyridoxine-5′-phosphate (PNP), pyridoxal-5′-phosphate (PLP) and pyridoxamine-5′-phosphate (PMP), respectively. PN, PL and PM are metabolically interconvertible and considered to be biologically active equivalents.[5] The structure of PN (vitamin B_6) is given in Figure 4.

The UV absorption spectra of PL, PN, PM and their 5′-phosphate esters are similar (in 0.1 M HCl) and the absorption maxima are at about 290 nm.[17] At pH 7.0, PN, PM, PNP and PMP have absorption maxima at 253 and 325 nm, whereas the other vitamers absorb maximally at 390 nm.[18] Physical properties of vitamin B_6 are given in Table 2.

2.10 Folate

Folate is the general term inclusive of folic acid (pteroyl glutamate) and poly-γ-glutamyl conjugates that exhibit the biological activity of folic acid. The structure of folic acid is given in Figure 4. Folate refers to the large group of heterocyclic compounds based on the pteroic acid structure conjugated with one or more L-glutamates linked through the γ-carboxyl group of the amino acid. Folic acid is not found in nature; rather, it is in its reduced form, either dihydrofolate (H_2 folate) or tetrahydrofolate (H_4 folate). Folic acid is yellow and slightly soluble in water in the acid form but fairly soluble in the salt form.

Table 2 Physical properties of water-soluble vitamins[a]

Substance	Formula	Molar mass	Solubility	Melting point (°C)	Crystal form	UV absorption[b]			Fluorescence[c]	
						λ_{max} (nm)	$A^{1\%}_{1\,cm}$	Solvent	λ_{ex} (nm)	λ_{em} (nm)
AA (CAS 50-81-7)	$C_6H_8O_6$	176.13	Soluble in water 30 g/100 mL	190–192 (dec.)	Monoclinic platelets and needles, white or yellow	245	695	H_2O, pH 2.0		
						265	940	H_2O, pH 6.4		
			Soluble in alcohol; insoluble in ether, $CHCl_3$, benzene, PE, oils and fats							
Thiamin (CAS 67-03-5)	$C_{12}H_{17}ClN_4OS{\cdot}HCl$	337.26	Water 1.0 g mL^{-1}; ethanol (95%) 1.0 g/100 mL; ethanol(100%) 0.3 g/100 mL; glycerol 5 g/100 mL; insoluble in ether, benzene, Hex, chloroform	246–250 (dec.)	White crystalline powder, colorless, monoclinic needles	246	425	0.1 M phosphate buffer, pH 2.9		
						234	345	0.1 M phosphate buffer, pH 5.5		
						264	255			
Thiamin mononitrate (CAS 532-43-4)	$C_{12}H_{17}N_5O_4S$	327.37	Water 2.7 g/100 mL	196–200 (dec.)	White to yellow, white crystals					
TPP (CAS 157-87-0)	$C_{12}H_{18}N_4O_7P_2S$	424.31		(Tetrahydrate form) 220–222 (dec.)						
Riboflavin (Vitamin B_2) (CAS 83-88-5)	$C_{17}H_{20}N_4O_6$	376.37	Soluble but unstable in dilute alkali; slightly soluble in water, 0.10–0.13 g L^{-1}; slightly soluble in alcohol, phenol; insoluble in $CHCl_3$, acetone, benzene, ether	278–282	Fine yellow-orange needles	260	**736**	0.1 M phosphate buffer	360, 465 pH 3.5–7.5	521
						375	**282**			
						450	**324**			
Riboflavin-5′-phosphate (FMN) (CAS 130-40-5)	$C_{17}H_{21}N_4O_9P$	456.35	Soluble in water 30 g L^{-1} (Na salt); insoluble in acetone, benzene, ether	280–290 (dec.)	Fine yellow-orange crystalline powder	260	**594**	0.1 M phosphate buffer	440–500 (3.5–7.5)	530
						375	**228**			
						450	**267**			

(*continued overleaf*)

Table 2 (*continued*)

Substance	Formula	Molar mass	Solubility	Melting point (°C)	Crystal form	UV absorption[b]			Fluorescence[c]	
						λ_{max} (nm)	$A^{1\%}_{1\,cm}$	Solvent	λ_{ex} (nm)	λ_{em} (nm)
FAD (CAS 146-14-5)	$C_{27}H_{33}N_9O_{15}P_2$	785.56	Soluble in water; insoluble in $CHCl_3$, acetone, benzene, ether			260 375 450	**471** **118** **144**	0.1 M phosphate buffer	440–500 (2.7–3.1)	530
Niacin										
Nicotinic acid (CAS 58-67-6)	$C_6H_5NO_2$	123.11	Water, 1.67 g/100 mL; ethanol, 0.73 g/100 mL; soluble in alkali metal hydroxides and carbonates, propylene glycol; insoluble in ether	236.6	Needles	260	**227**	50 mM potassium phosphate buffer, pH 7.0		
Nicotinamide (CAS 68-92-0)	$C_6H_6N_2O$	122.13	Water, 100 g/100 mL; ethanol, 66.6 g/100 mL; soluble in glycerol	128–131	Needles	261	478	0.1 N H_2SO_4		
PN (Vitamin B_6) (CAS 65-23-6)	$C_8H_{11}NO_3$	169.18	Soluble in water; weakly soluble in alcohol, acetone; insoluble in ether, $CHCl_3$	160 (dec.)		324 254	**428** **23**	pH 6.8 pH 6.8	332 (6.5–7.5) 320 (12.0–14.0)	400 (6.5–7.5) 380 (12.0–14.0)
PM dihydrochloride (Vitamin B_6) (CAS 524-36-7)	$C_8H_{12}N_2O_2{\cdot}2HCl$	241.12	Soluble in water, 95% alcohol	226–227 (dec.)		328 253	322 **190**	Water Water	337 (4.0–5.5) 320 (pH 14)	400 (4.0–5.5) 370 (pH 14)

Folate										
Folic acid (CAS 59-20-3)	$C_{19}H_{19}N_7O_6$	441.40				282 350 254 282 363	**612** **159** **566** **532** **197**	Phosphate buffer, pH 7.0 0.1 M NaOH, pH 13	363 (pH 9)	450–460 (pH 9)
H_4 folate	$C_{19}H_{23}N_7O_6$	445.44				297	**606**	Phosphate buffer, pH 7.0	305–310 (pH 3)	360 (pH 3)
Cyanocobalamin (Vitamin B_{12}) (CAS 68-19-9)	$C_{63}H_{88}CoN_{14}O_{14}P$	1355.38	Soluble in water (1 g/80 mL)		Dark red Hygroscopic Darkness at 210–220 °C	278 361 551	115 204 64	Water Water Water		
Biotin (CAS 58-85-5)	$C_{10}H_{16}N_2O_3S$	244.31	Soluble in dilute alkalis; sparingly soluble in water, alcohol; insoluble in most organic solvents	232–233	Colorless, fine long needles	204 Very weak				
Pantothenic acid (CAS 79-83-4)	$C_9H_{17}NO_5$	219.24	Freely soluble in water, ethyl acetate, dioxane; moderately soluble in ether, amyl alcohol; insoluble in $CHCl_3$			204 Very weak				

Source: references.[4,13,19,21,31–33]
[a] Table adapted from Eitenmiller and Landen.[5]
[b] Values in bold are calculated from corresponding ε values.
[c] Values in parentheses are pH range.
FAD, flavin adenine dinucleotide; PN, pyridoxine; PM, pyridoxamine; TPP, thiamin pyrophosphate.

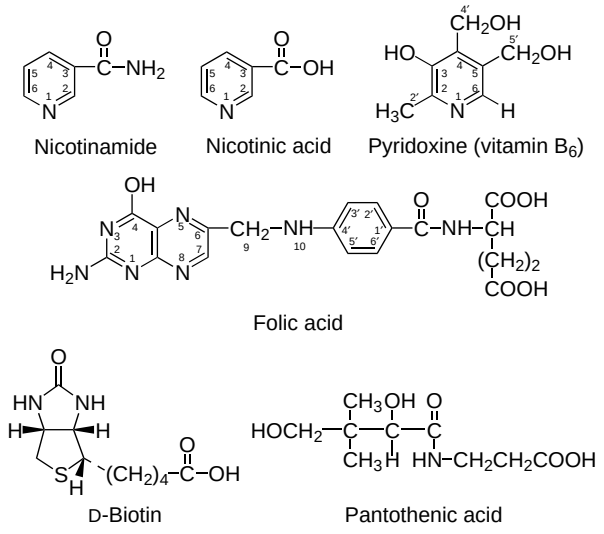

Figure 4 Selected structures of niacin, vitamin B_6, folate, biotin and pantothenic acid.

Folic acid and derivatives of tetrahydrofolate (except for 5,10-methenyltetrahydrofolate with a maximum at 352 nm) exhibit absorption spectra with maxima around 290 nm in a buffered medium at pH 7.[19] The fluorescence properties of naturally occurring folate and related compounds have been reported by Uyeda and Rabinowitz.[20] The fluorescence intensity is strongly influenced by pH and buffer composition. For reduced folates, the fluorescence intensity increases as the acidity of the solvent increases. Physical properties of folates are given in Table 2.

2.11 Vitamin B_{12}

Vitamin B_{12} is the collective name for cobalt-containing corrinoids with the biological activity of cyanocobalamin. Cyanocobalamin is the permissive or trivial name for vitamin B_{12}. Nutritional references to vitamin B_{12} usually refer to all cobalamins biologically active in humans. Aqueous solutions of cyanocobalamin have absorption maxima at 278, 361 and 551 nm.[2,4] Physical properties of cyanocobalamin are given in Table 2.

2.12 Biotin

Biotin is *cis*-hexahydro-2-oxo-1*H*-thieno[3,4-*d*]imidazole-4-pentanoic acid.[27] The molecule contains three asymmetric carbon atoms and therefore eight stereoisomers are possible. D-(+)-Biotin is the only biologically active form and other isomers are not found in nature.[28] The structure of biotin is shown in Figure 4. In nature, biotin occurs free or bound through ε-amino linkages with lysine as biocytin or to carrier proteins and peptides. Physical properties of biotin are given in Table 2. Biotin shows weak absorbance at 200–220 nm with an absorption maximum at 204 nm from the carbonyl group.[13] HPLC methods are only useful for the assay of highly concentrated pharmaceuticals or vitamin premixes unless the UV absorbance or fluorescence is enhanced through derivatization.

2.13 Pantothenic Acid

Pantothenic acid (Figure 4) is systematically named D-(+)-*N*-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine.[27] The trivial name is D-(+)-α,γ-dihydroxy-β,β-dimethylbutyl-β-alanine. The molecule is optically active but only the D-(+)-enantomers are biologically active and present in nature. Synthetic pantothenic acid is a racemic mixture. Pantothenic acid does not contain a chromophore. The carbonyl group absorbs weakly below 210 nm. The physical properties of pantothenic acid are given in Table 2.

3 ANALYSIS OF WATER-SOLUBLE VITAMINS

3.1 Microbiological Methods

Microbiological assays are available for each water-soluble vitamin except AA. These assays were concurrently developed with, or shortly after, development of bioassays for the different vitamins earlier in the twentieth century. Therefore, the method development activities and the availability of analytical procedures that provide similar responses to longer duration bioassays played a historically significant role in understanding not only microbial metabolism but also human and animal metabolism. Interestingly, except in a few cases, methodology developed during the 1940s has remained essentially unchanged. Microbiological assay of the water-soluble vitamins continues to play a major role in current vitamin analysis programs. Although HPLC techniques are available, microbiological techniques have not been replaced by the chromatographic assays. Microbiological assays offer specificity and sensitivity in a cost-effective manner often not obtainable with HPLC methods.[11]

Microbiological vitamin assays utilize dose–response curves of microorganisms that possess a highly specific requirement for the vitamin. Such requirements need to be absolute in that the nutritional requirement for the vitamin can not be altered by other naturally occurring compounds. When this requirement is met by the microorganism, a basal medium can be formulated that provides all of the growth requirements of the organism except for the vitamin being assayed.[29] When a standard or an extract of the material to be quantified is added to the clear growth medium, the microorganism

reproduces in proportion to the amount of added vitamin. Over a defined concentration range, the response will be directly proportional to the amount of vitamin present. Within this range, sample and standard solutions can be compared accurately.[30] Growth is measured by turbidity, production of metabolite (acid), gravimetrically or by respiration.

Requirements of bacteria, yeast, molds, and protozoans have been well defined. Microorganisms chosen for water-soluble vitamin assays must not be able to synthesize vitamin, utilize metabolites not active for humans or animals or utilize structurally related compounds. The microorganism, therefore, must possess the following characteristics: (1) specifically, require vitamin forms that are biologically active in higher animals; (2) be genetically stable; (3) possess a rapid growth cycle; (4) provide a growth response that is not easily modified by other substances that are present in the extract (stimulators or inhibitors); and (5) be nonpathogenic.[30]

Eitenmiller[30] summarized the advantages and disadvantages of microbiological assay as follows. Advantages: (1) sensitive; (2) capable of assaying naturally occurring levels from complex matrices (food samples); and (3) minimal laboratory requirements of space, materials and instrumentation. Disadvantages: (1) labor intensive; (2) poor precision; (3) difficult to streamline or automate, with limited sample output; (4) varying responses of the assay organism to different vitamin forms (bound forms or multiple chemical forms); (5) simultaneous vitamin assay not possible; (6) drift problems; and (7) provides little information on bioavailability.

Compared with bioassays using animals, microbiological assay requires less space, labor, material and time. Start-up costs can be minimal.

Because of extensive use worldwide, the assay techniques have been reviewed extensively and readers requiring in-depth knowledge are referred to the literature.[5,9,11,13,16,34]

As an overall reference to currently accepted procedures, AOAC International[34] presents detailed procedural guides for the analysis of many different food matrices. Since it is beyond the scope of this brief article to cover assays for each water-soluble vitamin, we have provided an overview in Table 3.

Table 3 Microbiological assay of water-soluble vitamins

Analyte	Assay organism	AOAC International method[a]	Extraction
Thiamin	*L. viridescens* ATCC 12706	–	0.1 N HCl hydrolysis and phosphatase digestion (follows AOAC International thiochrome digestion method 942.23)
Riboflavin	*L. rhamnosus* (casei) ATCC 7469	960.46 Chapter 45.2.06	0.1 N HCl hydrolysis and adjust pH to 4.5 (deproteinization), final pH 6.8
Niacin	*L. plantarum* ATCC 8014	944.13 Chapter 45.2.04 985.34 Chapter 50.1.19	1.0 N H_2SO_4 hydrolysis and adjust pH to 4.5 (deproteinization). Note: cereals contain bound niacin. Alkaline hydrolysis will free bound forms
Vitamin B_6	*S. uvarum* ATCC 9080	961.15 Chapter 45.2.08 985.32 Chapter 50.1.18	Plant tissue, 0.44 N HCl hydrolysis; animal product, 0.055 N HCl hydrolysis
Folate	*E. hirae* ATCC 8043 *L. rhamnosus* ATCC 7469	944.12 Chapter 45.2.03 992.05 Chapter 50.1.21	Total folate digest with conjugase or trienzyme procedure using conjugase, α-amylase and protease
Vitamin B_{12}	*L. delbrueckii* ATCC 7830	952.20 Chapter 45.2.02 986.23 Chapter 50.1.20	Extract with phosphate buffer in the presence of reducing agent
Biotin	*L. plantarum* ATCC 8014	–	Remove fat from high-fat product using Hex prior to extraction: plant products, 2 N H_2SO_4 hydrolysis; animal products, 6 N H_2SO_4 hydrolysis
Pantothenic acid	*L. plantarum* ATCC 8014	945.74 Chapter 45.2.05 992.07 Chapter 50.1.22	Natural pantothenate ''double enzyme'' treatment with alkaline phosphatase and pantetheinase

[a] *AOAC Official Methods of Analysis*, 16th edition, 1995.[34]

Table 4 Selected HPLC methods for the analysis of water-soluble vitamins in various foods[a,b]

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
AA						
AA, IAA, uric acid	Cured meats	Add 15 mL 5% MPA containing 0.1 mg mL^{-1} to 4–10 g sample, homogenize with and without the addition of AA (IS), centrifuge, dilute 10-fold with mobile phase	Altex Ultrasphere ODS 5 µm 250 × 4.6 mm	Isocratic, 0.04 M NaOAc, 0.005 M TBAP, 0.2 mg mL^{-1} EDTA, adjust pH to 5.25 with HOAc, 0.4 or 0.8 mL min^{-1}	EC, glassy carbon electrode, +0.6 V vs Ag/AgCl	36
AA, IAA	Foods, beverages, fruits, vegetables	Blend with 0.3 N TCA containing 0.5 mL octanol per 60 mL, add NaOAc buffer and ascorbate oxidase, incubate, 37 °C, 30 min	Spherisorb ODS2 5 µm 125 × 4.6 mm	Isocratic, 0.08 M KH_2PO_4–MeOH (80 : 20), pH 7.8, 1 mL min^{-1}	Fluorescence, λ_{ex} 365 nm, λ_{em} 418 nm	37–39
AA, DHAA, IAA, DHIAA	Various foods	Blend with 0.1 M citric acid containing 5 mM EDTA plus equal volume of Hex, centrifuge, dilute 1 : 1 with mobile phase	Altex ODS, 3 columns in series 5 µm 250 × 4.6 mm	Isocratic, 0.1 M NaH_2PO_4 containing 5 mM EDTA and 5 mM TBAP, pH 5.0, 0.5 mL min^{-1}	Postcolumn oxidation with OPD, fluorescence, λ_{ex} 350 nm, λ_{em} 430 nm	40–42
AA	Fruit juice	Dilute with mobile phase, filter	PLRP-S, 100 Å 5 µm 250 × 4.6 mm	Isocratic, 0.2 M NaH_2PO_4, pH 2.14, 0.5 mL min^{-1}	220 or 244 nm	43
AA, DHAA	Various foods	Solid foods, homogenize with 10–50-fold excess 20 mM H_2SO_4; juice, homogenize, centrifuge; DHAA, reduce with DTT	Sulfonated polystyene–divinylbenzene 100 × 4.6 mm	20 mM H_2SO_4, 0.6 mL min^{-1}	EC, Pt electrode +0.6–0.8 V vs Ag/AgCl	44
AA	Diets	Mix with 6% MPA, filter	Inertsil ODS-2 5 µm 250 × 4.6 mm	Isocratic, 100 mM KH_2PO_4 containing 1 mM EDTA. Na, pH 3.0, 0.6 mL min^{-1}	EC, 70 mV vs Ag/AgCl	45
AA, DHAA, IAA, DHAA	Various foods, animal tissues	Homogenize in MPA–HOAc (30 g + 80 mL diluted to 1 L), vortex with Hex, centrifuge, filter	PLRP-S, 2 columns in series 5 µm 150 × 4.6 and 250 × 4.6 mm	Isocratic, 0.2 M NaH_2PO_4, pH 2.14, 4 °C, 0.5 mL min^{-1}	Postcolumn oxidation with $HgCl_2$ and OPD, fluorescence, λ_{ex} 350 nm, λ_{em} 430 nm	46

AA, DHAA	Various foods	Homogenize in 17% MPA (0.85% final MPA concentration), measure before and after conversion of DHAA acid to AA by homocysteine	C_{18} 5 µm 250 × 4.6 mm	Isocratic, 80 mM NaOAc, pH 4.8, containing 1 mM octylamine–MeOH–MPA (85 : 15 : 0.00015), pH 4.6, 0.9 mL min^{-1}	EC, glassy carbon electrode, +0.7 V vs Ag/AgCl, 50 mA	47
AA	Various foods	Homogenize in MPA–HOAc (3% MPA in 8% HOAc)	µBondapak C_{18} 10 µm 300 × 3.9 mm	MeOH–water (55 : 45), 1 mL min^{-1}	Precolumn oxidation with Norit, OPD derivatization, fluorescence, λ_{ex} 350 nm, λ_{em} 430 nm	48
AA, DHAA, IAA, DHIAA	Various foods	Homogenize in 17% MPA, centrifuge, filter, reduce with homocysteine to determine AA and IAA	Supelcosil LC-18 DB, 3 columns in series 5 µm 250 × 4.6 mm	Isocratic, 0.08 M NaOAc, pH 5.4, with 5 mM TBAHS and 0.15% MPA	EC, glassy carbon electrode, +0.6 V vs Ag/AgCl	49, 50
AA, DHAA, IAA, DHIAA	Various foods	Homogenize in MPA–HOAc–EDTA (30 g MPA + 5 g EDTA + 80 mL HOAc diluted to 1 L); for processed foods containing IAA, AA can be used as IS, if the vitamin is not present in the product	PLRP-S, 100 Å, 2 columns in series 5 µm 150 × 4.6 and 250 × 4.6 mm	Isocratic, 0.2 M NaH_2PO_4, pH 2.14, 0.5 mL min^{-1}	Postcolumn derivatization of dehydroforms with OPD, fluorescence, λ_{ex} 350 nm, λ_{em} 430 nm	51
AA, DHAA	Infant formula	Homogenize and dilute 5 g to 100 mL with 1% MPA, centrifuge, filter	Superspher 100 RP-18 5 µm	Isocratic, 0.1 M KH_2PO_4, pH 3.5, 1 mL min^{-1}	254 nm	52
AA, sulfites	Beer, beverages	Add 1.5% MPA (20 mL) to 2 mL sample, flush flask with He, dilute to volume with 1.5% MPA, inject	Ion-exchange, Fast Acid 100 × 7.8 mm or Aminex HPX 87H 300 × 7.8 mm	Isocratic, 0.005 M H_2SO_4 containing 0.001 M chloride, 1 mL min^{-1}	EC, glassy carbon electrode, +0.6 V vs Ag/AgCl	53
AA	Fruits, vegetables	Grind with quartz sand, mix, macerate with 2% MPA (12 g + 48 mL), transfer and shake for 10 min, filter	Spherisorb ODS-2 10 µm 250 × 4.6 mm	Isocratic, 0.01 M KH_2PO_4–MeOH–20% TBAHS (970 : 30 : 1), adjust pH to 2.75 with 85% H_3PO_4	PDA, 190–340 nm	54

(*continued overleaf*)

Table 4 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
AA	Beverages, apple juice	Filter, neutralize with 0.1 M NaOH, dilute	5 C_{18} AR 250 × 4.6 mm	Isocratic, 15 mM NaH_2PO_4–KH_2PO_4, pH 6.5, 0.3 mL min^{-1}	Electrochemi-luminescence, 3 electrode system, glassy carbon, stainless steel, +1.5 V vs Ag/AgCl, λ_{em} 610 nm, $Ru(bpy)_3^{2+}$	55
AA	Green beans, Padrón peppers	Homogenize 20 g sample with 60 mL 4–5% MPA, 15 min, filter	ODS-2 C_{18} 5 µm 250 × 4 mm	Water, adjust pH to 2.2 with MPA, 0.5 mL min^{-1}	245 nm	56
Thiamin						
Thiamin TMP, TPP, TTP	Various foods	Add amprolium (IS), homogenize with 5% sulfosalicylic acid, centrifuge, filter water layer	Perkin-Elmer C_{18} 3 µm 300 × 3 mm	Gradient, 0.1 M Na_3PO_4, pH 5.5, 6 min, 0.1 M Na_3PO_4, pH 2.6, 19 min	Thiochrome, postcolumn fluorescence, λ_{ex} 339 nm, λ_{em} 432 nm	57
Thiamin, nonphos-phorylated	Infant formula, various foods	Adjust pH to 1.7–2.0 with 6 N HCl. Adjust pH to 4.0–4.2 with NaOH, filter, inject	µBondapak C_{18} 10 µm 300 × 4.6 mm	Isocratic, EDTA (2 g) + sodium Hex sulfonate (3 g) + acetic acid (15 mL) + MeOH (400 mL) dilute to 2 L with water, 50 °C, 2.5 mL min^{-1}	248 nm	58
Thiamin	Various foods	Hydrolyze with 0.1 M HCl (30 mL) and 6 M HCl (0.1 mL), 121 °C for 1 min, adjust pH to 4.0–4.5, digest with takadiastase®, 48 °C, 3 h	µBondapak C_{18} 10 µm 300 × 3.9 mm	Isocratic, MeOH–HOAc–sodium Hex sulfonate, various, 1.0 mL min^{-1}	254 nm	59
Thiamin	Various foods	Add 0.25 N H_2SO_4 (10 mL) to 2 g sample, autoclave, 30 min, adjust pH to 4.6 with NaOH–HOAc, digest with takadiastase®, 40–50 °C, 25 min, digest with papain, 40–45 °C, 2 h, add TCA, heat at 50–60 °C, 5 min, centrifuge	Mercksorb Si60 10 µm 250 × 4.5 mm	Isocratic, phosphate buffer pH 5.6–EtOH (100 : 12), 1.0 mL min^{-1}	Thiochrome, postcolumn fluorescence, λ_{ex} 366 nm, λ_{em} 464 nm	60

Thiamin	Various foods	Add HCl, digest, 100 °C for 30 min, adjust pH to 4–4.5, digest with takadiastase®, 47 °C, 3 h	Lichrospher 100 RP-18 5 µm 125 × 4 mm	Isocratic, H_3PO_4–KH_2PO_4, 10^{-2} M, pH 2.8–MeOH (85 : 15) with 5 µM Hex sulfonic acid and 0.1% triethylamine	254 nm	61
Thiamin	Dried yeast	Add 1 mL 10% HCl and 80 mL water to 1 g sample, incubate at 80–85 °C, 30 min, add NaOAc buffer, digest with takadiastase®, 45–50 °C, 3 h	Nucleosil C_{18} 5 µm or Capcell-pak C_{18} 5 µm 150 × 4.6 mm	Isocratic, 0.02 M phosphate buffer, pH 3.5, containing 0.2% sodium 1-octanesulfonate–MeCN (4 : 1), 40 °C, 1 mL min^{-1}	254 nm	62
Thiamin	Cooked sausages	10 g ground sample, homogenize with 60 mL 0.1 N HCl, autoclave, 20 min, cool, adjust pH to 4.0–4.5 with 2.5 M NaOAc, add 5 mL 6% claradiastase, incubate, 3 h, add 2 mL 50% TCA, 90 °C, 15 min, filter	Spherisorb C_{18} 5 µm 250 × 4 mm	Phosphate buffer (5 mM, pH 7.0)–MeCN (70 : 30), 35 °C, 0.650 mL min^{-1}	Fluorescence, λ_{ex} 360 nm, λ_{em} 430 nm	63
Riboflavin						
Riboflavin	Cheese	Homogenize in water–MeOH (2 : 1), acidify with HOAc, centrifuge	LiChrosorb RP_{18} 5 µm 250 × 4 mm	Isocratic, MeCN–water (20 : 80), 1 mL min^{-1}	446 nm	64
Riboflavin	Milk, dairy products	Milk: pass through C_{18} Sep-Pak, elute riboflavin with 0.02 M acetate buffer pH 4.0–MeOH (1 : 1). Dairy products: blend with 0.02 M acetate buffer pH 4.0, pass through C_{18} Sep-Pak	Bio-Sil ODS-5S C_{18} 250 × 4 mm	Isocratic, water–MeOH–HOAc (65 : 35 : 0.1), 1 mL min^{-1}	270 nm	65
Riboflavin	Milk	Mix with 10% TCA, dilute and filter	Spherisorb ODS 5 µm 250 × 4.5 mm	Isocratic, MeCN–water (20 : 80), 1.5 mL min^{-1}	Fluorescence, λ_{ex} 453 nm, λ_{em} 580 nm	66
Riboflavin	Various foods	Add 0.1 N HCl, autoclave, 30 min, adjust pH to 4.5, dilute, filter	LiChrosorb C_{18} 10 µm 250 × 4 mm	Isocratic, 0.005 M 1-Hex sulfonic acid–MeOH (60 : 40), 1.0 mL min^{-1}	Fluorescence, λ_{ex} 440 nm, λ_{em} 565 nm	67

(*continued overleaf*)

Table 4 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Riboflavin	Various foods	Homogenize, add 0.2 M H_2SO_4, autoclave 20 min, adjust pH to 4.5 with 2.5 M acetate buffer, digest with claradiastase or takadiastase®, 45 °C, dilute, filter	Spherisorb C_{18} 10 µm 250 × 4.6 mm	Isocratic, MeOH–water (35 : 65), 1 mL min^{-1}	Fluorescence, λ_{ex} 445 nm, λ_{em} 525 nm	68
Riboflavin, FMN, FAD	Dairy products	Add formic acid–urea (6% formic acid containing 2 M urea), homogenize or mix, centrifuge to remove fat, add sorboflavin (IS)	Supelco LC-18 3 µm 75 × 4.6 mm	Isocratic, MeCN–0.1 M KH_2PO_4 (14 : 86), pH 2.9, 1 mL min^{-1}	Fluorescence, λ_{ex} 450 nm, λ_{em} 530 nm	69
Riboflavin	Milk, nondairy imitation milk	Add 10% (w/w) lead acetate solution, pH 3.2, filter	Spherisorb ODS 5 µm 150 × 3.9 mm	Isocratic, water–acetic acid–MeOH (70 : 30), add 15 mL HOAc to 1 L water, 0.6 mL min^{-1}	270 nm	70
Riboflavin, FMN, FAD, total riboflavin	Various foods	Add MeOH (9 mL) and CH_2Cl_2 (10 mL) per 0.5–4.0 g sample, add 7-ethyl-8-methylriboflavin (IS), homogenize, add 0.1 M citrate–phosphate buffer, pH 5.5, containing NaN_3, homogenize, centrifuge, filter	2 PLRP-S 100 Å in series 5 µm 250 × 4.6 and 150 × 4.6 mm	MeCN–0.1% sodium azide in 0.01 M citrate–phosphate buffer, pH 5.5, in multistep gradient elution program	Fluorescence, λ_{ex} 450 nm, λ_{em} 522 nm	71
Riboflavin	Cooked sausages	Homogenize 10 g ground sample with 60 mL 0.1 N HCl, autoclave, 20 min, cool, adjust pH to 4–4.5 with 2.5 M NaOAc, add 6% claradiastase, incubate, 50 °C, 3 h, add 2 mL 50% TCA, 90 °C, 15 min, filter	Spherisorb ODS-2 5 µm 250 × 4 mm	5 mM HEP sulfonic acid (pH 2.7)–MeCN (75 : 25), 35 °C, 0.6 mL min^{-1}	Fluorescence, λ_{ex} 277 nm, λ_{em} 520 nm	72
Niacin						
Nicotinic acid, nicotinamide	Meats	Add 30 mL water to 5 g sample, homogenize, Polytron®, boil, 10 min, cool, dilute to 50 mL, filter, 0.45 µm	Partisil SCX cation-exchange 10 µm 250 × 4.6 mm	Isocratic, 50 mM phosphate buffer, pH 3.0, 1 mL min^{-1}, 25 °C	260 nm	73

Nicotinic acid, nicotinamide	Beef, pork	Blend, add 30 mL water to 10 g sample, homogenize (Biotron), dilute to 100 mL with water, centrifuge, transfer 20 mL to 25 mL volumetric flask, add 1 mL saturated $ZnSO_4$ and 0.5 L 1 N NaOH (deproteinization), dilute to volume, allow to stand at ambient temperature 30 min, filter	μBondapak C_{18} 300 × 3.9 mm or Radial-Pak 2 module	Isocratic. Nicotinic acid: MeOH–water (1 : 9) with 5 mm PIC-A, pH 3.5, 1 mL min^{-1}. Nicotinamide: 2 PIC-B7 vials in 1000 mL water, 1 mL min^{-1}	263 nm	74
Nicotinic acid	Vegetables, meat	Add 30 mL 0.1 N HCl to 1–10 g sample plus 1 mL 6 N HCl, autoclave, adjust pH to 4.0–4.5 with 2 N NaOAc, add 5 mL of 6% takadiastase®, digest at 48 °C, 3 h, filter, dilute to 100 mL with water	μBondapak C_{18} 10 μm 300 × 3.9 mm or Spherisorb ODS-2 10 μm 300 × 3.9 mm	Isocratic, 0.005 M TBAHS in MeOH–0.01 M NaOAc (1 : 9), pH 4.72, 1.5 mL min^{-1}	254 nm	75
Nicotinic acid	Fortified foods	Weigh sample containing 100 mg niacin into 500 mL Phillips beaker, add 50 mL water, add 6 mL H_2SO_4 (1 : 1), mix, autoclave, 45 min, cool, adjust pH to 6.0 with 7.5 N NaOH, dilute to 100 mL with water	PRP-X100 250 × 4.1 mm	Isocratic, 20 mL HOAc diluted to 1 L with water	254 nm	76
Niacin	Infant formula wheat flour	Weigh sample containing 100–200 μg niacin, add 8 mL DMW, 2 mL (1 + 1) H_2SO_4, mix, autoclave, 45 min, cool, adjust pH to 6.5 with 7.5 N NaOH, adjust pH to ≤1 with H_2SO_4, filter	PRP-X100 anion-exchange 250 × 4.1 mm	0.1 M NaOAc–HOAc (pH 4–4.2), 1.5 mL min^{-1}	260 nm	77
Niacin	Various foods	1–5 g ground sample, add 30 mL 0.1 M HCl, heat at 100 °C, 1 h, cool, dilute, filter, autoclave, 1 h, cool, adjust pH to 4.5, dilute, filter	Lichrospher 100 RP-18 5 μm 250 × 5 mm	0.07 M KH_2PO_4, 0.075 M hydrogen peroxide, 5.10×10^{-6} M copper(II) sulfate solution, 1 mL min^{-1}	Fluorescence, λ_{ex} 322 nm, λ_{em} 380 nm	78

(*continued overleaf*)

Table 4 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Vitamin B_6						
PL, PM, PN	Human milk	Add 4′-deoxypyridoxine (IS), digest with potato acid phosphatase, 1 h, 37 °C, add TCA, add CH_2Cl_2, centrifuge, adjust supernatant to pH 5.2, filter	μBondapak C_{18} 10 μm 300 × 3.9 mm	Gradient: (A) MeOH–water (85 : 15), (B) 0.005 M HEP sulfonic acid in 1% HOAc; 0–17 min, 2–40% A; 17–20 min to 100% B; 1.5 mL min^{-1}	Fluorescence, λ_{ex} 300 nm, λ_{em} 375 nm	79
PL, PM, PLP, PMP	Chicken, raw, cooked	Extract with 1 M MPA and water, Polytron®, filter	Biosil ODS-5S 250 × 4 mm	Isocratic 0.066 M KH_2PO_4, pH 3.0, 1 mL min^{-1}	Fluorescence, λ_{ex} 330 nm, λ_{em} 400 nm	80
PL, PM, PN, PLP, PMP, PNP, PN-glucoside	Various foods	Extract with 1 M cold perchloric acid, centrifuge, adjust pH to 3.5 with 50% KOH, precipitate overnight, divide into two parts, inject one and treat other with KCN to 5 mM to convert PLP to 4-PA-P	TSK Gel ODS 120 Å 250 × 4.6 mm	Isocratic, 1% MeCN in 0.1 M sodium perchlorate and 0.1 M KH_2PO_4, pH 3.5, 0.5 mL min^{-1}	Fluorescence, λ_{ex} 330 nm, λ_{em} 400 nm. PLP determined as 4-PA-P or postcolumn derivatization with sodium hydrogensulfite	81, 82
PL, PM, PN, PMP, PN-glucoside	Wheat	Add 4′-deoxypyridoxine (IS), homogenize in water, deproteinize with MPA, centrifuge, filter	Ultramex C_{18} 3 μm 150 × 4.6 mm	(A) 0.033 M H_3PO_4 and 8 mM 1-octanesulfonic acid, pH 2.2; (B) 0.033 M H_3PO_4 and 10% MeCN, pH 2.2; gradient program	Fluorescence, λ_{ex} 311 nm, λ_{em} 360 nm	83
PL, PM, PN, PLP, PMP, PNP, 4-PA	Egg, milk, baker's yeast extract, yeast cell-free culture media	Baker's yeast extract: 50 mg sample, add 50 μL 1 mM iso-PL, 3 mL 1 M $HClO_4$. Egg yolk: 2 g sample, add 100 μL 1 mM iso-PL, 6 mL 1 M $HClO_4$. Milk: 2 g sample, add 10 μL 1 mM iso-PL, 1 mL 1 M $HClO_4$; mix, centrifuge, decant supernatant, adjust pH to 3–4 with 10 M KOH, refrigerate, centrifuge, filter	Phenospher ODS-2 5 μm 250 × 4.6	0.15 M NaH_2PO_4, adjust pH to 2.5 with 70% $HClO_4$, 1 mL min^{-1}	Postcolumn derivatization, sodium hydrogensulfite, fluorescence, λ_{ex} 290 nm, λ_{em} 389 nm	84

PM, PL, PN	Pork meat, pork meat products	5 g sample, add 60 mL 0.1 M HCl and 3 mL 100 μg mL^{-1} 4-deoxypyridoxine (IS), water-bath, 100 °C, 30 min, cool, adjust pH to 4–4.5 with 2 M NaOAc, add 5 mL 10% takadiastase®, incubate, 45 °C, 3 h, cool, add 2 mL 50% TCA, heat, 100 °C, 5 min, filter	Spherisorb ODS C_{18} 5 μm 250 × 4 mm	0.01 M H_2SO_4, 1 mL min^{-1}	Fluorescence, λ_{ex} 290 nm, λ_{em} 395 nm, 30 °C	85
Folate						
Polyglutamyl folates 1–7, folic acid 1–7, H_2 folate 1–7, H_4 folate 1–7, 5-CH_3-H_4 folate 1–7, 10-CHO-H_4 folate 1–7	Various foods	Suspend or dilute in solution of 2% AA, 10 mM 2-MCE and 100 mM bis-tris, pH 7.8 in autoclave, 30 min, cool, homogenize, centrifuge, mix with [^{3}H]folic acid	Econospher 5 μm 100 × 4.6 mm	(A) 5 mM TBAP, 0.5 mM DET in 25 mM phosphate–Tris buffer, pH 7.4, or 25 mM NaCl in water; (B) 5 mM TBAP, 0.5 mM DET in 25 mM–Tris buffer, pH 7.4, in MeCN–EtOH–water (64 : 9 : 27); gradient program	PDA, 350 nm; 10 CHO-H_4 folate, 258 nm	86, 87
Folic acid, 5-CH_3-H_4 folate, 5-CHO-H_4 folate, 10-CHO-H_2 folate, 10-CHO-folate	Cereal foods	Add 10 volumes HEPES–CHES buffer, pH 7.85 (50 mM HEPES, 50 mM CHES) containing 2% AA and 10 mM 2-MCE, vortex, boiling water-bath, 10 min, cool, homogenize, digest rat plasma conjugase and α-amylase (37 °C, 4 h) followed by protease (37 °C, 1 h), boiling water-bath, 5 min, cool, centrifuge, resuspend residue in extraction buffer, centrifuge filter (Whatman No.1, flush N_2)	Ultramex C_{18} 5 μm 250 × 4.6 mm	Gradient, (A) MeCN, (B) 0.033 M H_3PO_4, pH 2.3: 0–8 min isocratic 5% A, 8–33 min linear gradient to 17.5% A, 1 mL min^{-1}	PDA, UV, 280 nm	88

(*continued overleaf*)

Table 4 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Folic acid, H_2 folate, 5-CHO-H_4 folate, 10-CHO-H_4 folate	Milk, dairy products	Adjust pH to 4.5 with HOAc, homogenize, centrifuge, decant supernatant, add phosphate buffer, pH 4.5, containing 10% AA and 1 M 2-MCE to concentration of 0.1% AA and 0.1 M 2-MCE, digest with HK conjugase, 37 °C, 2 h, centrifuge, filter, 0.45 µm	C_{18} Microsorb 3 µm 100 × 4.6 mm	Isocratic, phosphate buffer, pH 6.8–MeOH (50 : 50) containing 50 mL L^{-1} of 1.0 M TBAP, 1 mL min^{-1}	Fluorescence, λ_{ex} 238 nm, λ_{em} 340 nm; postcolumn pH adjustment with 4.25% H_3PO_4 or enhancement hypochlorite oxidation	89
5-CH_3-H_4 folate	Citrus juice	Centrifuge diluted juice, adjust pH to 5.0 (1.0 M NaOH), digest with HK conjugase, 37 °C, 1.5 h, cool, centrifuge, filter (0.45 µm)	Zorbax ODS C_{18} 5 µm 250 × 4.6 mm	(A) MeOH–phosphate/acetate buffer pH 5.0 (10 : 90) containing 5 mM TBAP, (B) similar to A but containing 30% MeOH, for backflushing and analysis	EC, glassy carbon electrode, 0.2 V vs Ag/AgCl	90, 91
Folates	Vegetables, pig liver, milk powder, wholemeal flour	Milk powder: mix with water (1 : 1); weigh individual sample (0.5–4 g), add CHES–HEPES buffer, homogenize, place in boiling water, 10 min, cool to 0 °C, adjust pH to 7 with HCl, weigh each extract into centrifuge tube (folate ≤ 500 ng), digest with trienzymes (conjugase, amylase and protease), 37 °C, 4 h, place tubes in boiling water-bath, 5 min, cool in ice, centrifuge, dilute supernatant to 2 mL with CHES–HEPES buffer	Vydac 201 TP 54 5 µm 250 × 4.6 mm 20 °C	(A) phosphate buffer (0.033 M, pH 2.1), (B) MeCN, gradient program	Two fluorescences, λ_{ex} 280 nm, λ_{em} 360 nm and λ_{ex} 359 nm, λ_{em} 460 nm; PDA, 280 nm	92

Folic acid	Fortified cereal products	Homogenize 2.0 g ground sample with 50 mL 0.1 M K_2HPO_4 (pH 8–9) containing 0.05% ascorbate, 1 h, adjust pH to 6.9 with H_3PO_4, α-amylase digestion, 65 °C, 1 h, inactivate enzyme at 90 °C, cool, centrifuge, filter	Microsorb-MV C_{18} 3 µm 100 × 4.6 mm	Isocratic, MeOH–K_2HPO_4 (pH 6.8) (25 : 75) containing 5 mM tetrabutylammmonium dihydrogenphosphate, 1 mL min^{-1}	280 nm	93
Biotin						
D-Biotin D-Biocytin	Various foods	Ground sample (5–10 g) [for high-starch sample, add 100 mg takadiastase® (60.5 µ mg^{-1})], add reduced glutathione (300 µL), EDTA (300 µL), citrate buffer (30 mL) and papain (3 mL), shake and incubate, 37 °C, overnight, cool, filter	Lichrospher 100 RP-18 5 µm 250 × 5 mm	0.1 M phosphate buffer (pH 6)–MeOH (81 : 19), 0.4 mL min^{-1}	Postcolumn derivatization with avidin–fluorescein isothiocyanate; fluorescences, λ_{ex} 490 nm, λ_{em} 520 nm	94

[a] Table adapted from Eitenmiller and Landen.[5]

[b] Abbreviations (where not already defined elsewhere): 4-PA, 4-pyridoxic acid; 4-PA-P, 4-pyridoxic acid 5′-phosphate; CHES, 2-(*N*-cyclohexylamino)ethanesulfonic acid; DET, dithioerythritol; DHIAA, dehydroisoascorbic acid; DMW, demineralized water; DTT, dithiothreitol; EC, electrochemical; EDTA, ethylenediaminetetraacetic acid; HEP, heptane; HEPES, *N*-(2-hydroxyethyl)piperazine-N′-ethanesulfonic acid; IAA, isoascorbic acid; IS, internal standard; MCE, mercaptoethanol; MPA, metaphosphoric acid; ODS, octadecylsilane; TBAHS, tetrabutylammonium hydrogensulfate; TBAP, tetrabutylammonium phosphate; TCA, trichloroacetic acid; TMP, thiamin monophosphate; TPP, thiamin pyrophosphate; TTP, thiamin triphosphate.

3.2 Chromatographic Methods

HPLC is the most commonly used chromatographic method to assay water-soluble vitamins. HPLC can provide selectivity and accuracy with fast results and the ability to assay different forms of the vitamin in biological matrices. HPLC techniques have been widely used for the simultaneous determination of fat-soluble vitamins in food; however, their application to water-soluble vitamin analysis is limited owing to the availability of sufficiently sensitive and specific detection systems that are capable of quantifying several vitamers from a complex mixture of compounds. Examples of selected HPLC methods for water-soluble vitamins are given in Table 4; Ball[9,35] and Eitenmiller and Landen[5] provide in-depth reviews on HPLC procedures for water-soluble vitamin analysis.

Only reversed-phase high-performance liquid chromatography (RPHPLC) is used in water-soluble analysis owing to the chemical properties of this group. Columns used vary depending on the analytes; they include ion-exchange columns for vitamin C[53] and niacin,[73] PLRP-S for riboflavin[71] and vitamin C[43,46,51] and C_{18} columns for most analytes. The mobile phase consists of methanol or acetonitrile together with water, acetic acid or buffers. The extraction procedures usually involve acid hydrolysis for vitamin C, thiamin, riboflavin, niacin and vitamin B, and enzyme hydrolysis for thiamin, riboflavin, niacin, vitamin B_6 and folate. UV, EC and fluorescence detectors have often been used in the methods. Pre- or postderivatization is used for those analytes that lack native fluorescence.[40–42,46,48,51,57,60,81,82,84,94]

Some multianalyte methods are available for the determination of riboflavin and thiamin and, occasionally, other water-soluble vitamins. A simultaneous determination of thiamin and riboflavin in selected foods was described by Sims and Shoemaker.[95] After autoclave extraction, samples are derivatized to form thiochrome (a highly fluorescent oxidation product of thiamin). Riboflavin has strong native fluorescence. The interferences were removed on a C_{18} cartridge. Analytes were separated on a reversed-phase system with a nonaqueous mobile phase. Detection limits were 0.05 ng and relative standard deviations were <2% for both analytes.

Albalá-Hurtado et al.[96] reported a rapid HPLC method for the determination of nicotinamide, thiamin, riboflavin, PN, PL, PM, cyanocobalamine and folic acid in liquid and powdered infant milk. Sample preparation involved acidification to precipitate proteins, centrifugation, then filtration. Ion-pair chromatography with a reversed-phase C_{18} column was used for analysis of the six vitamins. Methanol–water (15 : 85) containing 5 mM octanesulfonic acid with 0.5% triethylamine at pH 3.6 was used as the mobile phase. The method is simple and reliable.

4 ANALYSIS OF FAT-SOLUBLE VITAMINS

4.1 Chromatographic Methods

Chromatographic methods are the most versatile and useful methods for the determination of the fat-soluble vitamins. These methods include column chromatography, paper chromatography, thin-layer chromatography (TLC), gas chromatography (GC) and HPLC.

Paper chromatography has been used by several workers for the determination of Ts.[97–99] One- and two-dimensional paper chromatography were used successfully by Green et al. in structural studies on Ts.[98,99] However, paper chromatography is tedious, time-consuming and inaccurate for the quantification of individual Ts.[100]

TLC, compared with paper chromatography, has several advantages, which include a shorter development time, superior separation of the individual Ts, larger loading capacity and better quantitative elution of compounds from the absorbent.[100] TLC has traditionally been a method of choice for the analysis of lipophilic compounds and for confirming the purity of concentrated solutions of retinoids and carotenoids, where impurities that do not absorb markedly at the wavelength of detection or that elute slowly from the HPLC column might not be detected.[1] For quantitative analysis, HPLC or GC have surpassed TLC, as quantitative TLC assays lack precision, are labor intensive and are difficult to automate. Moreover, the risk of oxidative losses by exposing Ts on large surface areas to light and air is enhanced. However, TLC, has value as a straightforward and inexpensive qualitative or semiquantitative assay method.[7]

GC is a useful technique for the simultaneous separation and quantitative detection of free Ts, T esters and their derivatives. However, it has limited value in retinoid analysis owing to the instability of conjugated unsaturated compounds.[101,102] With respect to vitamin E analysis, GC is fairly sensitive and microgram quantities of different Ts can be readily detected.[24,100] Sample preparation for GC analysis of vitamin E is more critical than with other chromatographic techniques. Interferences of lipids present in the unsaponifiable fraction of biological materials are more likely to occur. As a result, sample pretreatment prior to GC often becomes tedious and time-consuming.[103]

The two major, distinct methods of HPLC separation are referred to as normal-phase (adsorption) chromatography and reversed-phase (partition) chromatography. In normal-phase chromatography, the sample is preferentially adsorbed on a polar stationary phase, usually silica (sometimes alumina), and displaced from its adsorption sites by mobile phase, typically Hex containing small amounts of polar modifiers. Hence less polar compounds elute before more polar compounds. RPHPLC

is characterized by partitioning of the sample between a hydrophobic stationary phase (typically C_2, C_8, C_{18} or phenyl chains covalently bonded to silica particles) and a more polar mobile phase, typically methanol or acetonitrile with small amounts of water added. In contrast to adsorption chromatography, polar components are eluted before less polar components in RPHPLC. RPHPLC has many advantages over the normal-phase mode, such as (1) less sensitivity to changes in retention time due to the presence of water; (2) more easily cleaned of contaminants; (3) more stable to small changes in mobile phase composition; (4) more quickly equilibrated to mobile phase composition changes, permitting use of gradients; and (5) capable of resolving compounds with a wide range of polarities.[1,102,104–106] Adsorption HPLC, however, is capable of resolving compounds (e.g. geometric isomers of retinoids and the β- and γ-positional isomers of vitamin E) that are difficult or impossible to separate by RPHPLC. HPLC for the determination of fat-soluble vitamins can be utilized in both normal-phase and reversed-phase modes. However, β- and γ-isomers of the Ts and T3s cannot be separated by RPHPLC.[107–112] Moreover, the nonpolar column does not tolerate heavy loads of triglycerides or other nonpolar compounds.[112,113] Normal-phase columns can tolerate lipid up to 2 mg per injection.[113]

Efficient extraction of vitamins from the sample matrix before HPLC analysis is as important as the determinative chromatography. Extraction can be very time-consuming. The extraction methods used for fat-soluble vitamins include alkaline hydrolysis (saponification), enzymatic hydrolysis, direct solvent extraction, Soxhlet extraction and supercritical fluid extraction (SFE) of the total lipid component.

Saponification is the most widely used and generally accepted method for the determination of vitamins A, D and E and carotenoids. It is conventionally carried out by refluxing an appropriate amount of suitably prepared sample with ethanolic KOH solution, in the presence of an antioxidant, such as pyrogallol,[113–115] or AA,[116–118] with heat treatment[115,117,119–121] or without heat treatment.[122,123] The unsaponifiable matter containing the fat-soluble vitamins, sterols, carotenoids etc. is extracted with Hex[116,123] or a mixture of ethyl acetate and Hex[124,125] or PE[121], leaving the fatty acid salts and other water-soluble components behind. The liberation of the unstable retinol from the relatively stable esters during saponification demands protective measures against light and oxygen during the analysis.

Several factors affect the efficiency of the extraction. In 1987, Ueda and Igarashi[126] reported that the ethanol concentration in the extraction medium must be kept below 30% for the quantitative extraction of Ts and increasing levels of co-existing fat decreased the extraction recovery of Ts, except α-T. Hewavitharana and van Brakel[117] found that the use of 3% (m/v) AA in the saponification mixture along with the addition of butylated hydroxytoluene (BHT) to the extracting solution gave higher recoveries of retinol, α-T and β-carotene compared with the use of ethanolic pyrogallol or 0.5% (m/v) ethanolic AA alone. Saponification could reduce the recovery of all Ts and in particular, δ-T[114,126–129] and α-T.[128] The reduction in recovery was dependent on the concentration of KOH,[127] the temperature of saponification [room temperature (15, 20 °C) vs 70 °C][114,126,128] and triacylglycerol content.[126] Saponification is a cheap and easy method, but time-consuming, and the process often facilitates the isomerization of α- and β-carotene and destruction of α-T.[130] Owing to the instability of vitamin K under alkaline conditions, saponification cannot be used for extraction.

Enzymatic hydrolysis using lipase has been employed as an effective alternative to saponification to reduce the risk of vitamin degradation. For the simultaneous determination of vitamins A, D, E and K in fortified milk- and soy-based infant formulas and dairy products,[131] an amount of sample containing approximately 3.5–4.0 g of fat was digested for 1 h with lipase at 37 °C and pH 7.7. The hydrolyzate was made alkaline in order to precipitate the fatty acids as soaps. The solution was then diluted with ethanol and extracted with pentane. A final water wash yielded an organic phase containing, primarily, the fat-soluble vitamins and cholesterol. This treatment effectively hydrolyzed the glycerides, but only partially converted retinyl palmitate and α-tocopheryl acetate to their alcohol forms.[35] Lietz and Henry[130] used nonspecific *Candida cylindracea* lipase to achieve complete hydrolysis of red palm oil after 4 h with no significant loss of carotenoids and Ts.

Soxhlet extraction is often the simplest and most efficient method for vitamin E extraction if the sample is dry.[5] Håkansson et al.[109] used Soxhlet extraction for the determination of vitamin E in wheat products. The sample (5–15 g) was weighed and placed in a Soxhlet extraction thimble. Glass wool was placed on the top of the sample and the thimble was placed in a 100-mL Soxhlet extractor. The extractor was wrapped in aluminum foil in order to protect vitamin E from light and to minimize cooling. The sample was extracted with Hex containing BHT (1 mg per 125 mL) in the Soxhlet extractor placed in a water-bath (90 °C) for 4 h. Similar procedures have been employed in HPLC methods for the determination of Ts in pecans[132] and fruit.[133]

SFE is an extraction technique which exploits the solvent properties of fluids above their critical point. Carbon dioxide is often used instead of the organic solvent that is normally employed in conventional extraction methods. The extraction conditions are related to the relative

solvent strength which, in a given fluid, is primarily dependent upon its density. This method can eliminate harmful organic solvents in the extraction of lipophilic plant components and avoid environmental pollution. Marsili and Callahan[134] compared a liquid–solvent extraction technique with SFE for the determination of α- and β-carotene in vegetables. They concluded that SFE provides recoveries of α- and β-carotene from vegetables that are equal to or greater than those obtained with traditional solvent extraction techniques.

The direct solvent extraction of the fat-soluble vitamins from the sample matrix with a suitable solvent system provides a milder environment than alkaline hydrolysis and maintains esterified vitamins in their original form.[35] Solvent systems must be capable of penetrating tissues and breaking lipoprotein bonds to free the analytes.[2] Hex is commonly used as the extraction solvent. Other useful solvents include diethyl ether, ethyl acetate and various solvent mixtures. Thompson and Hatina[113] presented a solvent extraction method for Ts and T3s in foods. Samples (10 g) of foods and tissues were homogenized in IPA and acetone. Low-polarity lipids were obtained in the epiphase after the addition of Hex and water to the filtered extract. After separation, the hypophase was extracted twice with Hex and the combined epiphases were evaporated. The efficiency of the procedure was $\geq$97% depending on the matrix being extracted. This method has been used by many investigators for vitamin E extraction from a wide array of natural products.[5]

Ultraviolet/visible (UV/VIS) and fluorescence detectors have been used extensively in food analysis to quantify fat-soluble vitamins. EC detection has been used mainly in procedures using RPHPLC. The major reason for this includes the poor solubility of most electrolytes which are used for conductivity in organic solvents and the incompatibility of organic solvents with some commonly used working electrodes.[135]

Absorbance UV/VIS measurement represents the most widely used detection method employed in HPLC, and it can be performed using either a fixed-wavelength photometer or a continuously variable wavelength [photodiode-array (PDA)] detector. Although the detector PDA shares many elements with the UV/VIS detector, the essential difference is that it can record the entire spectral range (190–800 nm) during analysis, thus monitoring the chromatogram at selected wavelengths while recording the spectra of eluates simultaneously. Therefore, it eliminates the need to generate multiple chromatograms at different wavelengths and/or collecting fractions to record spectra later. This ability significantly reduces the time taken for analysis.

Fluorescence detection (FLD) is considerably more sensitive than UV/VIS detection.[136–139] The sensitivity of FLD depends, in part, on the composition of the mobile phase. For example, α-T and retinol display a 5–6-fold decrease in intensity on changing from Hex to acetonitrile–water (50 : 50);[2] hence normal-phase HPLC provides greater sensitivity for FLD than reversed-phase systems.[140] Caution has to be taken to avoid quenching (reduced intensity of fluorescence) when FLD is used. The factors leading to quenching are (1) dissociation of molecule by the excitation wavelength, (2) absorbance of emitted light by other molecules or the sample itself and (3) collisions of molecules. FLD is also more selective than UV/VIS detection because two wavelengths are required in the measurement and the structural features necessary for a molecule to fluoresce are more limited. Most lipids, including glycerides and sterols, do not fluoresce.[35] Therefore, injection of fat extracts can be made directly on silica column, for the analysis of vitamin E.[5]

Ueda and Igarashi[141] studied the application of EC detectors for detection of Ts in feeds by HPLC. They found that EC detection had a 20-fold higher sensitivity than FLD; however, β- and α-T cannot be separated owing to the required use of reversed-phase chromatography, because the EC detector needs a conducting eluent. Other examples of the use of EC detection include work by Schneiderman et al.[142] on the determination of retinyl palmitate in cereal products, by Dionisi et al.[143] on the assay of olive oil for Ts and T3s and by Huang et al.[135] on the analysis of vitamin A and vitamin E in plasma. Some selected HPLC methods for determination of fat-soluble vitamins are given in Table 5.

4.2 Multianalyte Methods

Development of vitamin assays that can simultaneously quantify two or more analytes has been an important part of method development in food and biological chemistry. Multianalyte methods save time and resources. Simultaneous vitamin assay method development faces the following areas of difficulty, as summarized by Eitenmiller and Landen:[5]

1. often fat-soluble vitamins are present in nature as multiple, structurally related compounds;
2. the presence of synthetic forms or metabolites can complicate quantitation;
3. each fat-soluble vitamin has its own spectral characteristics and therefore multiple detectors or programmable detectors are required;
4. stability differences among fat-soluble vitamins often complicate extractions;
5. concentrations of the various vitamins differ considerably in biological materials, pharmaceuticals, foods, feeds and fortified products;

Table 5 Selected HPLC methods for fat-soluble vitamins in foods[a,b]

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Vitamin A						
Retinol	Dried skimmed milk	Saponify, reflux 30 min, extract with PE, dilute with PE after addition of BHT, evaporate aliquot to dryness, dissolve residue in MeOH	C_8 or C_{18} 10 µm 250 × 4.6 mm	Isocratic, MeOH–water (90 : 10), 2 mL min^{-1}	325 nm	144
Retinol	Cheese	Homogenize in water–EtOH–Hex (8 : 20 : 20) containing 2% BHT, add water, centrifuge, remove Hex layer, dilute with Hex	LiChrosorb Si 60 250 × 4.6 mm	Isocratic, Hex–IPA (99.8 : 0.2), 2 mL min^{-1}	325 nm	145
Retinol	Milk	Saponify, 70 °C, 20 min extract with HEP–DIPE (3 : 1), re-extract, inject extract	HS-5 silica 125 × 4.0 mm	Isocratic, HEP–IPA (60 : 1), 1 mL min^{-1}	Fluorescence, λ_{ex} 344 nm, λ_{em} 472 nm	146
Retinol, 7-*cis*-, 9-*cis*-, 9,13-di-*cis*-, 13-*cis*-, 11,13-di-*cis*-, 11-*cis*	Liver, liver products, infant foods	Low-fat samples: homogenize liver and infant foods, digest with pancreatin at pH 9.0, saponify, ambient, 16 h. High fat samples: homogenize liver sausage with water–EtOH (1 : 1), saponify, ambient, 16 h, dilute digest with water and EtOH to give 1 : 1 ratio, pass aliquot through Kieselguhr column, elute with PE, evaporate, dissolve residue in iOCT	Spherisorb SW 3 µm 100 × 2.0 mm	Isocratic, iOCT–IPA (98.75 : 1.25), 0.4 mL min^{-1}	Retinol 325 nm, 13-*cis*-retinol 329 nm	147, 148
Retinol	Milk powder, flour	Mix 1.25 g sample with 5 mL aqueous NaOH (50%), heat 30 °C, 3 min, add 25 mL EtOH and 0.5 mL hydroquinone (20% in EtOH), heat, 80 °C, 30 min, cool, add water, extract with Et_2O–PE (1 : 1), re-extract, filter, evaporate and re-dissolve in MeOH	LiChrosorb RP-18 5 µm 125 × 4.5 mm 30 °C	Isocratic, MeCN–water (80 : 20), 0.8 mL min^{-1}	325 nm	149

(*continued overleaf*)

Table 5 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Carotenoids						
α-, β-, γ-carotene, lutein, zeaxanthin, β-cryptoxantin, phytoene, phytofluene	Vegetables	Add retinyl palmitate (IS), blend with Na_2SO_4, $MgCO_3$ and THF, concentrate samples, partition THF extract with PE, evaporate to dryness; system I, dissolve residue in THF; system II, dissolve residue in THF–EtOH (1 : 3)	Spherisorb-S RP-18 or Spherisorb-ODS 220 × 4.6 mm	System I: isocratic MeCN–CH_2Cl_2–MeOH (70 : 20 : 10), 1.8 mL min^{-1} System II: MeCN–MeOH (85 : 15), flow-rate gradient	Carotenoids 450 nm, phytoene 286 nm, phytofluene 370 nm	150
Carotenoids, 30 esters	Fruits, vegetables	Add Na_2CO_3, extract with acetone, evaporate, 40 °C, add NaCl-saturated water, extract with benzene–Et_2O (1 : 1), dry over Na_2SO_4, evaporate, dissolve in $CHCl_3$ and mobile phase	LiChrosorb C_{18} 10 µm 250 × 4.6 mm	Isocratic, MeCN–IPA–water (39 : 57 : 4) or MeCN–IPA–MeOH–water (39 : 52 : 6 : 4), 0.9–1.5 mL min^{-1}	PDA, multiple λ	151
α-, β-, γ-, δ-carotene, lutein, lycopene-5,6-diol, lycopene, β-crypto-xanthin, neurosporene, phytoene, phytofluene	Tomato juice	Add β-apo-8′-carotenal (IS), blend with $MgCO_3$, Celite and THF, filter, re-extract until colorless, reduce volume, partition with CH_2Cl_2 and NaCl-saturated water, dry over Na_2SO_4, reduce volume to 10 mL, filter, dilute with CH_2Cl_2	Microsorb MV C_{18} 5 µm 250 × 4.6 mm	Isocratic, MeCN–MeOH–CH_2Cl_2–Hex (40 : 20 : 20 : 20)	PDA, 450 nm except lycopene 470 nm, β-cryptoxanthin 445 nm, γ-carotene 400 nm, phytofluene 350 nm, phytoene 290 nm	152
α-, β-carotene, 13-*cis*-β-carotene, 9-*cis*-β-carotene	Green vegetables	Blend with water containing 0.5% AA, extract with acetone–PE (3 : 2) containing 0.5% BHT, saponify extract, ambient, 15 min, wash 3× with 10% NaCl solution, dry over Na_2SO_4, evaporate, dissolve in MeOH–CH_2Cl_2 (9 : 1)	Vydac TP-201 5 µm 250 × 4.6 mm	Isocratic. Prepared isomers: MeOH–CH_2Cl_2–water (79 : 15 : 6) Sample: MeOH–CH_2Cl_2–water (80 : 12.5 : 4.8), 1 mL min^{-1}	450 nm	153
β-carotene, all-*trans*, *cis*-zeaxanthin, β-cryptoxantin, luteoxanthin isomers, violaxanthin, all-*trans*, *cis*-neoxanthin	Mango	Extract with cold acetone, partition with Et_2O–PE (1 : 1), saponify, ambient, overnight, wash organic layer free of base, concentrate, add Sudan I (IS) to aliquot, evaporate, dissolve in Hex	Spherisorb Nitrile 5 µm 150 × 4.6 mm	Multilinear gradient acetone in Hex: acetone, 0–15% in 10 min, to 20% in 20 min, to 30% in 10 min, to 40% in 2 min, 1 mL min^{-1}	PDA, 400–500 nm; identification by mass spectrometry, molecular ion, fragmentation pattern	154

Neoxanthin, violaxanthin, zeaxanthin, cryptoxanthin, α-, β-carotene, β-zeacarotene, lycopene	Vegetables and fruits	5 g cut sample, freeze-dried, 24 h, weigh again for dry weight%, grind sample, extract sample (0.5–1 g) with 5 mL THF–MeOH (1 : 1) containing β-apo-8′-carotenal or echinenone (IS), re-extract twice, centrifuge, mix supernatant with 10 mL Hex, 2 mL 10% NaCl, evaporate Hex layer, dissolve in 200 µL CH_2Cl_2, inject	Vydac 201 TP54 C_{18} 5 µm 250 × 4.6 mm 30 °C	Isocratic, MeOH–MeCN (9 : 1), 1.0 mL min^{-1}, 30 °C	450 nm	155
Lutein, zeaxanthin, β-cryptoxanthin, α-, β-carotene	Orange juice	Extract orange juice (5 mL) 3× with EtOAc containing 0.004% BHT, transfer extract through anhydrous Na_2SO_4, add 50 mL MeOH (0.004% BHT) and 100 mL 1 M NaCl, extract with 75 and 25 mL EtOAc (0.004% BHT), evaporate, 40 °C, add 1.5 mL mobile phase and 100 µL Sudan I (50 mg L^{-1}) (IS), dilute to 10 mL with mobile phase	Vydac 201 TP54 C_{18} 5 µm 250 × 4.6 mm	Isocratic, MeCN–MeOH–1,2-dichloroethanol (60 : 35 : 5) containing 0.1% BHT, 0.1% triethylamine with 0.05 M ammonium acetate (in MeOH) added, 1 mL min^{-1}	450 nm	156
Vitamin D						
D_2, D_3	Milk, infant formula	Saponification, 75 °C, 30 min, extract with Et_2O, evaporate, redissolve in Hex, clean up	Apex silica 3 µm 150 × 4.5 mm	Isocratic, Hex–amyl alcohol (100 : 0.15), 3 mL min^{-1}	254 nm	157, 158
D_2, D_3	Infant formula, enteral nutritionals	Add D_2 (IS), saponification, 60 °C, 30 min, extract with Hex, evaporate, clean up	Vydac 201 TP54 C_{18} 5 µm 250 × 4.6 mm	Gradient, MeCN–MeOH–EtOAc, 0.7–2.5 mL min^{-1}	265 nm	159, 160
D_3, 25(OH)-D_3	Fish, fish products	Add D_2 and 25(OH)-D_2 (IS), saponification, extract with PE–Et_2O (1 : 1), evaporate, redissolve in Hex, clean up	Vydac 201 TP54 C_{18} 250 × 4.6 mm	Isocratic. For 25(OH)-D_2 and D_3: MeOH–water (83 : 17) For D_2 and D_3: MeOH–water (93 : 7)	264 nm	161
D_2, D_3	Medical nutritionals	Saponification, extract with Et_2O, add D_2 (IS), evaporate, redissolve in Hex, clean up	Hitachi Gel 3056 5 µm 250 × 4 mm	Isocratic, MeCN–MeOH–50% $HClO_4$ (970 : 30 : 1.2) containing 0.057 M $NaClO_4$, 1.2 mL min^{-1}	Dual-electrode EC, redox mode, detector 1 0.65 V, detector 2 0.20 V, H_2/H reference	162

(*continued overleaf*)

Table 5 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
D_2, D_3	Egg yolk	Add D_2 (IS), saponification, ambient, overnight, extract with Hex, evaporate, redissolve in Hex, clean up	Vydac 201 TP54 C_{18} 5 µm 250 × 4.6 mm	Isocratic, MeOH–water (94 : 6), 1 mL min^{-1}	264 nm	163
D_2, D_3	Meat, animal fat	Saponification, extract with Hex–CH_2Cl_2 (85 : 15), evaporate, redissolve in Hex, no IS, clean up	Apex silica 3 µm 150 × 4.5 mm	Isocratic, MeOH, 100%, 1.5 mL min^{-1}	254 nm	164
D_2, 25(OH)-D_2, pre-D_2	Mushrooms	Add D_3 and 25(OH)-D_3(IS), saponification, extract with PE–Et_2O (1 : 1), evaporate, clean up	Vydac 201 TP54 C_{18} 250 × 4.6 mm	Isocratic. For D_2, D_3, pre-D_2, pre-D_3: MeOH–water (93 : 7), 1 mL min^{-1} For 25(OH)-D_2, 25(OH)-D_3: Hex–IPA (98 : 2), 1 mL min^{-1}	264 nm	165, 166
Vitamin E						
α-, β-, γ-, δ-T, α-, β-,γ-, δ-T3	Vegetable oils, cod liver oil, margarine, butter, dairy spread	Dilute with Hex, direct injection	LiChrosorb Si60 5 µm 250 × 4.0 mm 30 °C	Gradient, 8–17% DIPE in Hex	Fluorescence, λ_{ex}290 nm, λ_{em}325 nm	167
α-, β-, γ-, δ-T, α-, β-, γ-, δ-T3	Olive oil	(a) Normal-phase HPLC: dilute with Hex, direct injection (b) RPHPLC: dilute with THF, dilute with MeOH, direct injection	(a) LiChrosorb Si60 5 µm 250 × 4 mm (b) Spherisorb ODS 5 µm 250 × 4.6 mm	Isocratic. (a) Hex–IPA (99.7 : 0.3), 1.7 mL min^{-1} (b) 0.05 M $NaClO_4$–MeOH (10 : 90), 2.0 mL min^{-1}	(a) fluorescence, λ_{ex}290 nm, λ_{em}330 nm; PDA, 280 nm (b) Amperometric, 0.6 V	168
α-T, α-tocopheryl acetate	Infant formula	Extract fat by the Ross–Gottlieb procedure, saponify lipid fraction, extract with Hex, evaporate, redissolve in IPA–EtOH–Hex (1 : 0.5 : 98.5)	LiChrosorb Si60 120 × 4.6 mm	Isocratic, IPA–EtOH–Hex (1 : 0.5 : 98.5), 1 mL min^{-1}	Fluorescence, λ_{ex}292 nm, λ_{em}320 nm	143
α-, β-, γ-, δ-T	Infant formula, milk, various foods	Saponify, extract with light petroleum–DIPE (2 : 1), centrifuge, inject 10 µL of upper layer	(a) Rad-Pak silica cartridge 5 µm, RCM-100 (b) Rad-Pad C_{18} cartridge 5 µm, RCM-100	Isocratic. (a) Hex–IPA (99 : 1) (b) 100% MeOH, 1 mL min^{-1}	Fluorescence, λ_{ex}295 nm, λ_{em}330 nm	114

α-, β-, γ-, δ-T, α-, β-, γ-, δ-T3	Vegetables, oils, wheat flour, barley milk, frozen dinners, beef, spinach, infant formula	Extract with boiling IPA, filter, extract with acetone, add water, Hex, collect Hex, layer	LiChrosorb Si60 5 µm 250 × 3.2 mm	Isocratic, moist Hex–Et_2O (95 : 5), 2 mL min^{-1}	Fluorescence, λ_{ex}290 nm, λ_{em}330 nm	113
α-, β-, γ-, δ-T, α-, β-, γ-, δ-T3	Finnish foods	(a) Butter, margarine: dissolve in Hex, direct injection (b) Other foods : saponification, overnight, ambient	LiChrosorb Si60 5 µm 250 × 4 mm	Isocratic, Hex–DIPE (93 : 7), 2.1–2.5 mL min^{-1}	Fluorescence, λ_{ex}290 nm, λ_{em}325 nm	122, 123, 127, 169–172
α-, δ-T	Chicken muscle	Saponification, overnight, ambient, followed by 2 h at 50 °C, extract with Hex, evaporate	BioSil ODS-5S 250 × 4 mm	Isocratic, MeOH, 100%, 1 mL min^{-1}	Fluorescence, λ_{ex}296 nm, λ_{em}330 nm	173
α-, β-, γ-, δ-T	Peanuts, pecans	Soxhlet, Hex containing 0.01% BHT, 90 °C, 6 h, evaporate, redissolve in Hex containing 0.01% BHT	LiChrosorb Si60 5 µm 250 × 4.0 mm	Isocratic, Hex–IPA (99 : 1), 1 mL min^{-1}	Fluorescence, λ_{ex}290 nm, λ_{em}330 nm	132, 174, 175
α-, β-, γ-, δ-T, α-, β-, γ-, δ-T3	Grain amaranths	Extract with MeOH, evaporate, extract with Hex	Waters silica 300 × 4 mm 10 µm	Isocratic, Hex–IPA (99.8 : 0.2), 1 mL min^{-1}	Fluorescence, λ_{ex}295 nm, λ_{em}330 nm	176
α-T, α-T3	Fruits	SFE	(a) Merck Si60 250 × 4 mm (b) RP-18 5 µm 250 × 4 mm	Isocratic. (a) Hex–IPA (99.5 : 0.5), 1.5 mL min^{-1} (b) MeOH–1,4-dioxane–water (86 : 10 : 4) containing 75 mM $NaClO_4$, 1 mL min^{-1}	(a) Fluorescence, λ_{ex}295 nm, λ_{em}333 nm (b) EC, 950 mV	133
α-, β-, γ-, δ-T	Fruits	Saponification, overnight, ambient, followed by 2 h at 50 °C, extract with Hex, evaporate	Phenomenex Prodigy ODS-2 5 µm 250 × 2 mm	Isocratic, MeCN–MeOH (85 : 15), 1 mL min^{-1}	294 nm	177
α-, γ-, δ-T	Margarine, reduced-fat products	Dissolve sample (5 g) in Hex containing 0.1% BHT, add $MgSO_4$ to remove water, filter	LiChrosorb Si60 5 µm 250 × 4.6 mm	Isocratic, Hex–IPA (99.1 : 0.9), 1.0 mL min^{-1}	Fluorescence, λ_{ex}290 nm, λ_{em}330 nm	178
Vitamin K						
K_1	Infant formula	Lipase hydrolysis, add aqueous NH_3, EtOH, extract with pentane	µBondapak C_{18} 5 µm 250 × 4.6 mm	Isocratic, MeOH–MeCN–THF–water (39 : 39 : 16 : 6), 1.5 mL min^{-1}	254 nm	179

(*continued overleaf*)

Table 5 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
K_1	Infant formula	Add aqueous NH_3, MeOH, CH_2Cl_2–iOCT (2 : 1), collect organic layer, evaporate, clean up	Apex I, silica 5 µm 250 × 4.6 mm	Isocratic, CH_2Cl_2–iOCT (30 : 70) containing 0.02% IPA, 1 mL min^{-1}	254 nm	158
K_1	Various foods	Add dihydro-K_1 (IS), homogenize, extract with Hex, evaporate, redissolve in Hex, add equal volume of MeOH–water (9 : 1) to Hex extract, mix, centrifuge, evaporate upper Hex layer to dryness, dissolve residue in mobile phase	Hypersil ODS 5 µm 250 × 4.6 mm	Isocratic, MeOH–CH_2Cl_2 (90 : 10) containing 5 mL methanolic solution containing 1.37 g $ZnCl_2$, 0.41 g NaOAc and 0.3 g HOAc, 1 mL min^{-1}, 40 °C	Postcolumn reduction. Fluorescence, λ_{ex}243 nm, λ_{em}430 nm	180
K_1	Vegetables	1–5 g sample, sonicate with 10 mL MeOH, 15 min, centrifuge, mix 2 mL extract with 4 mL carbonate, heat at 80 °C, 1 h, partition with 4 mL Hex (3×), evaporate Hex layer, redissolve in MeOH	LiChrosorb RP-8 10 µm 250 × 4.6 mm	Isocratic, MeOH, 100%, 0.6 mL min^{-1}	247 nm and particle beam mass spectrometry	181
K_1, MK-4	Cereal products	2.5 g sample, add MK-4 (40–365 ng), 15 mL IPA, digest in boiling water, 10 min, cool, add 10 mL Hex twice, homogenize 2 min each time, add 10 mL water and shake, evaporate Hex layer, redissolve in Hex, clean up	Vydac 201 TP54 C_{18} 5 µm 250 × 4.6 mm	Isocratic, MeOH (96%)–0.05 M NaOAc, pH 3, 1 mL min^{-1}	EC, redox mode, upstream 1.1 V, downstream 0 V	182
K_1	Oil, margarine, butter	Oil: add MK-4 (IS) to 0.5–1 g sample, dilute to 10 mL with Hex Butter, margarine: extract with 3 methods: (a) Hex; (b) IPA–Hex (1 : 1); (c) add 2 mL water and 2 mL 25% ammonia, shake, add 5 mL EtOH, extract with Et_2O and PE; wash with water, evaporate, dissolve in Hex, clean up	Vydac 205 TP54 5 µm 250 × 4.6 mm	Isocratic, MeOH (95%)–0.05 M NaOAc, pH 3.0, 1 mL min^{-1}	EC, redox mode, upstream 1.1 V, downstream 0 V	183

[a] Table adapted from Eitenmiller and Landen.[5b] Abbreviations (where not already defined elsewhere): DIPE, diisopropyl ether; THF, tetrahydrofuran. iOCT, isooctane.

Table 6 Selected multianalyte assays for fat-soluble vitamins in foods[a]

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Retinyl palmitate, β-carotene	Margarine, milk	Dissolve in Hex, shake with 60% EtOH, centrifuge, inject directly	LiChrosorb Si60 5μm 250 × 3.2 mm	Isocratic, wet Hex–dry Hex–Et_2O (49:49:2), 2 mL min^{-1}	Retinyl palmitate 325 nm β-carotene 453, 436 nm	184
Vitamins K_1, D_2, D_3, other fat-soluble vitamins	Infant formula, dairy products	Add cholesterol phenyl acetate (IS), lipase hydrolysis, add aqueous NH_3, EtOH, extract with pentane	Two Zorbax ODS in series 250 × 4.6 mm	Gradient, MeOH–EtOAC (86:14), against 100% MeCN, variable flow rate	265 nm	131
All-*rac*-α-tocopheryl acetate, retinyl palmitate, β-carotene, vitamin D_2 or D_3, K_1	Margarine, oils, infant formula, milk, cereals	Homogenization with mixture IPA–CH_2Cl_2 with $MgSO_4$ added to remove water, clean up	Zorbax ODS 6 μm 250 × 4.6 mm	Isocratic, CH_2Cl_2–MeCN–MeOH (300:700:2), 1 mL min^{-1}	Retinyl palmitate 325 nm; β-carotene 436 nm; vitamin D_2 or D_3 280 nm; vitamin E 280 nm; vitamin K_1 280 nm	185–188
α-T, γ-T, retinyl esters	Milk, human	Add α-tocopheryl acetate (IS), dilute with EtOH, extract with Hex, evaporate, dissolve in Hex containing 0.1% BHT	Rad-Pak silica cartridge 10 μm 5 mm i.d.	Isocratic, Hex–DIPE (95:5), 2.5 mL min^{-1}	280 nm	189
All-*trans*-retinol, α-carotene, β-carotene, lycopene	Forty foods of animal origin	Saponification, heating mantle, 30 min, extract with Hex (4×), dry, Na_2SO_4, evaporate, dilute with Hex	μBondapack C_{18} 10 μm 300 × 3.9 mm	Isocratic, MeCN–MeOH–EtOAc (88:10:2), 2 mL min^{-1}	Carotenoids 436 nm, retinol 313 nm	190
All-*trans*-retinol, α-, β-, γ-, δ-T, β-carotene	Italian cheese	Saponification, 70 °C, 30 min, extract with Hex–EtOAc (90:10), evaporate, dissolve in mobile phase	Ultrasphere Si 5 μm 250 × 4.6 mm	Isocratic: Hex–IPA (99:1), 1.5 mL min^{-1} Gradient: multilinear pump A, Hex–IPA (99:1), pump B, Hex (100%), 1.5 mL min^{-1}	Ts, fluorescence, λ_{ex}280 nm, λ_{em}325 nm; retinol, fluorescence, λ_{ex}325 nm, λ_{em}475 nm; β-carotene, 450 nm	124

(*continued overlead*)

Table 6 (*continued*)

Analyte	Sample type	Sample preparation	Column	Mobile phase	Detection	Refs.
Vitamin D_3, all-*trans*-retinol, α-T	Milk, milk powder	Saponification (on-line), neutralization (on-line) Sep-Pak Plus C_{18} cartridge (on-line concentration), clean up	Brownlee OD-224 RP-18 5 µm 220 × 4.6 mm	Isocratic, HOAc–NaOAc (25 mM) in MeOH–water (99 : 1), 1 mL min^{-1}	(a) 280 nm (b) EC + 1300 mV	191
Vitamin K_1, α-T, all-*trans* retinol	Milk, milk powder	(a) α-T, retinol: saponification, overnight, ambient, extract with Hex, evaporate, dissolve in MeOH (b) vitamin K_1: lipase hydrolysis, add alcoholic NaOH and immediately extract with Hex, clean up	Brownlee OD-224 RP-18 5 µm 220 × 4.6 mm	Isocratic, MeOH–water (99 : 1) containing 2.5 mM HOAc–NaOAc, 1.25 mL min^{-1}	EC, dual-EC glassy carbon, −1100 mV, +700 mV vs Ag/AgCl	192–193
α-T, retinol	Infant formula	Saponification, overnight, ambient, extract (5×) with Hex, wash extract with water, dry over Na_2SO_4, filter, evaporate, 40 °C, dissolve in MeOH, filter	Spherisorb ODS-2 C_{18} 5 µm 250 × 4.6 mm	Isocratic, water–MeCN–MeOH (4 : 1 : 95)	α-T 292 nm, retinol 323 nm	118
α-T, retinol acetate, vitamin D_3, provitamin D_2	Animal feed	Extract 1 g ground sample with 4 mL Hex, centrifuge, evaporate 1 mL supernatant, redissolve in 0.3 mL 1-butanol	Novapak C_{18} 150 × 3.9 mm	Isocratic, MeOH, 1.5 mL min^{-1}	290 nm	194
All-*trans*-retinol, 13-*cis*-retinol, α-, β-T, β-carotene, retinyl palmitate	Milk	Saponification, 70 °C, 7 min, extract with Hex–DIPE (75 : 25) containing 0.04% BHT, add 30 mL Milli-Q water, gently invert the tube, centrifuge, evaporate supernatant, dissolve in 2 mL Hex containing 0.4% BHT	Alltech Econosphere silica 3 µm 150 × 4.6 mm	Isocratic, 0.01% acetic acid–0.5% IPA–0.02 mg L^{-1} α-T in Hex, 1.5 mL min^{-1}	Vitamin A, fluorescence, λ_{ex} 330 nm, λ_{em} 470 nm; vitamin E, fluorescence, λ_{ex} 295 nm, λ_{em} 330 nm; β-carotene 450 nm	117

[a] Table adapted from Eitenmiller and Landen.[5]

6. co-eluting interferences often lead to the need for specialized sample cleanup approaches.

Therefore, there is no single method that can be used to analyze all the fat-soluble vitamins, their synthetic forms and metabolites. However, many good multivitamin methods have been developed with the improvement of modern instrumentation. They are summarized briefly in Table 6.

ABBREVIATIONS AND ACRONYMS

AA	Ascorbic Acid
AOAC	Association of Official Analytical Chemists (now known as AOAC International)
BHT	Butylated Hydroxytoluene
CHES	2-(*N*-Cyclohexylamino)ethanesulfonic Acid
DET	Dithioerythritol
DHAA	Dehydroascorbic Acid
DHIAA	Dehydroisoascorbic Acid
DIPE	Diisopropyl Ether
DMW	Demineralized Water
DTT	Dithiothreitol
EC	Electrochemical
EDTA	Ethylenediaminetetraacetic Acid
FAD	Flavin Adenine Dinucleotide
FLD	Fluorescence Detection
FMN	Flavin Mononucleotide
GC	Gas Chromatography
HEP	Heptane
HEPES	*N*-(2-Hydroxyethyl)piperazine-N′-ethanesulfonic Acid
Hex	Hexane
HPLC	High-performance Liquid Chromatography
IAA	Isoascorbic Acid
iOCT	Isooctane
IPA	2-Propanol
IS	Internal Standard
MCE	Mercaptoethanol
MK-4	Menaquinone-4
MPA	Metaphosphoric Acid
NAD	Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
ODS	Octadecylsilane
OPD	*o*-Phenylenediamine
PDA	Photodiode-array
PE	Petroleum Ether
PL	Pyridoxal
PLP	Pyridoxal-5′-phosphate
PM	Pyridoxamine
PMP	Pyridoxamine-5′-phosphate
PN	Pyridoxine
PNP	Pyridoxine-5′-phosphate
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SFE	Supercritical Fluid Extraction
T	Tocopherol
TBAHS	Tetrabutylammonium Hydrogensulfate
TBAP	Tetrabutylammonium Phosphate
TCA	Trichloroacetic Acid
THF	Tetrahydrofuran
TLC	Thin-layer Chromatography
TMP	Thiamin Monophosphate
TPP	Thiamin Pyrophosphate
TTP	Thiamin Triphosphate
T3	Tocotrienol
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
4-PA	4-Pyridoxic Acid
4-PA-P	4-Pyridoxic Acid 5′-phosphate

RELATED ARTICLES

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Fluorescence Spectroscopy in Food Analysis • Liquid Chromatography in Food Analysis • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. H.C. Furr, A.B. Barua, J.A. Olson, 'Retinoids and Carotenoids', in *Modern Chromatographic Analysis of*

the Vitamins, eds. A.P. De Leenheer, W.E. Lambert, H.J. Nelis, Marcel Dekker, New York, Chapter 1, 1992.
2. G.F.M. Ball, *Fat-soluble Vitamin Assays in Food Analysis; a Comprehensive Review*, Elsevier Applied Science, Barking, 1988.
3. K.L. Simpson, C.O. Chichester, 'Metabolism and Nutritional Significance of Carotenoids', *Annu. Rev. Nutr.*, **1**, 351 (1981).
4. W.F. Friedrich, *Vitamins*, Walter de Gruyter, Berlin, 1988.
5. R.R. Eitenmiller, W.O. Landen Jr, *Vitamin Analysis for the Health and Food Sciences*, CRC Press, New York, 1999.
6. L.J. Machlin, 'Vitamin E', in *Handbook of Vitamins: Nutritional, Biochemical, and Clinical Aspects*, ed. L.J. Machlin, Marcel Dekker, New York, 99–145, 1984.
7. J.K. Lang, M. Schillaci, B. Irvin, 'Vitamin E', in *Modern Chromatographic Analysis of Vitamins*, 2nd edition, eds. A.P. Deleenbeer, W.E. Lamberty, H.J. Nelis, Marcel Dekker, New York, 153–195, 1992.
8. J.W. Suttie, 'Vitamin K', in *The Vitamins*, ed. F.C. Gerald, Academic Press, New York, Chapter 8, 1992.
9. G.F.M. Ball, *Bioavailability and Analysis of Vitamins in Foods*, Champman & Hall, New York, 1998.
10. G.R. Combs Jr, *The Vitamins, Fundamental Aspects in Nutrition and Health*, Academic Press, New York, Chapter 9, 1992.
11. R.R. Eitenmiller, W.O. Landen Jr, 'Vitamins', in *Analyzing Food for Nutrition Labeling and Hazardous Contaminants*, eds. I.J. Jeon, W.G. Ikins, Marcel Dekker, New York, Chapter 9, 1995.
12. M.T. Parviainen, K. Nyyssönen, 'Ascorbic Acid', in *Modern Chromatographic Analysis of Vitamins*, 2nd edition, eds. A.P. Deleenbeer, W.E. Lamberty, H.J. Nelis, Marcel Dekker, New York, 235–360, 1992.
13. G.F.M. Ball, *Water-soluble Vitamin Assays in Human Nutrition*, Chapman and Hall, New York, 1994.
14. L.F. Russell, J.T. Vanderslice, 'Non-degradative Extraction and Simultaneous Quantitation of Riboflavin, Flavin Mononucleotide and Flavin Adenine Dinucleotide in Foods by HPLC', *Food Chem.*, **43**, 151–162 (1992).
15. P. Nielsen, 'Flavins', in *Modern Chromatographic Analysis of Vitamins*, eds. A.P. De Leenheer, W.E. Lambert, H.J. Nelis, Marcel Dekker, New York, Chapter 9, 1992.
16. R.R. Eitenmiller, S. DeSouza, 'Niacin', in *Methods of Vitamin Assay*, 4th edition, eds. J. Augustin, B.P. Klein, D.A. Becker, P.B. Venugopal, John Wiley & Sons, New York, Chapter 15, 1985.
17. J.W. Bridges, D.S. Davies, R.T. Wlilliams, 'Fluorescence Studies on Some Hydroxypyridines Including Compounds of the Vitamin B_6 Group', *Biochem. J.*, **98**, 451–468 (1966).
18. J.B. Ubbink, 'Vitamin B_6', in *Modern Chromatographic Analysis of Vitamins*, 2nd edition, eds. A.P. De Leenheer, W.E. Lambert, H.J. Nelis, Marcel Dekker, New York, Chapter 10, 1992.
19. C. Temple Jr, J.A. Montgomery, 'Chemical and Physical Properties of Folic Acid and Reduced Derivatives', in *Folates and Pterins*, eds. R.L. Blakely, S.J. Benkovic, John Wiley & Sons, New York, Chapter 2, Vol. 1, 1984.
20. K. Uyeda, J.C. Rabinowitz, 'Fluorescence Properties of Tetrahydrofolate and Related Compounds', *Anal. Biochem.*, **6**, 100–106 (1963).
21. S. Budavari, *The Merck Index*, 12th edition, Merck, Whitehouse Station, NJ, 1996.
22. H.C. Furr, A.B. Barua, J.A. Olson, 'Analytical Methods', in *The Retinoids, Biology, Chemistry and Medicine*, 2nd edition, eds. M.B. Sporn, A.B. Roberts, D.S. Goodman, Raven Press, New York, Chapter 3, 1994.
23. M. Mino, H. Nakamura, A.T. Diplock, H.J. Kayden, *Vitamins*, Japan Scientific Societies Press, Tokyo, 1993.
24. L.J. Machlin, *Handbook of Vitamins*, 2nd edition, Marcel Dekker, New York, 1991.
25. H.F. DeLuca, 'Vitamin D', in *Handbook of Lipid Research 2, The Fat-soluble Vitamins*, ed. H.F. Deluca, Plenum Press, New York, 69–132, 1978.
26. J.C. Bauernfeind, *Carotenoids as Colorants and Vitamin A Precursors, Technological and Nutritional Applications*, Academic Press, New York, 1981.
27. Anon., 'Nomenclature Policy: Generic Descriptors and Trivial Names for Vitamins and Related Compounds', *J. Nutr.*, **120**, 12–19 (1990).
28. M. Gaudry, O. Ploux, 'Biotin', in *Modern Chromatographic Analysis of Vitamins*, 2nd edition, eds. A.P. De Leenheer, W.E. Lambert, H.J. Nelis, Marcel Dekker, New York, Chapter 11, 1992.
29. M.N. Voigt, R.R. Eitenmiller, 'Microbiological Assays', in *Methods of Vitamin Assay*, eds. J. Augustin, B.P. Klein, D.A. Becker, P.B. Venugopal, John Wiley & Sons, New York, 49–63, 1985.
30. R.R. Eitenmiller, 'Strengths and Weaknesses of Assessing Vitamin Content of Foods', *J. Food Qual.*, **13**, 7–20 (1990).
31. Committee on Food Chemicals Codex, *Food Chemical Codex*, 4th edition, National Academy of Sciences, Washington, DC, 1996.
32. J. Augustin, P.K. Brbara, D.A. Becker, P.B. Benugopal, *Methods of Vitamin Assay*, John Wiley & Sons, New York, 1985.
33. A.P. De Leenheer, W.E. Lambert, H.J. Nelis, *Modern Chromatographic Analysis of Vitamins*, Marcel Dekker, New York, 1992.
34. *Official Methods of Analysis of AOAC International*, 16th edition, AOAC International, Arlington, VA, 1995.
35. G.F.M. Ball, 'The Fat-soluble Vitamins', in *Food Analysis by HPLC*, ed. L.M.L. Nollet, Marcel Dekker, New York, 275–340, 1992.
36. M.A. Kutnink, S.T. Omaye, 'Determination of Ascorbic Acid, Erythorbic Acid, and Uric Acid in Cured Meats by High Performance Liquid Chromatography', *J. Food Sci.*, **52**, 53–56 (1987).

37. A.J. Speek, J. Schrijver, W.H.P. Schreurs, 'Fluorometric Determination of Total Vitamin C and Total Isovitamin C in Foodstuffs and Beverages by High-performance Liquid Chromatography with Precolumn Derivatization', *J. Agric. Food Chem.*, **32**, 352–355 (1984).
38. M. Hägg, S. Ylikoski, J. Kumpulainen, 'Vitamin C and α- and β-carotene Contents in Vegetables Consumed in Finland During 1988–1989 and 1992–1993', *J. Food Compos. Anal.*, **7**, 252–259 (1994).
39. M. Hägg, S. Ylikoski, J. Kumpulainen, 'Vitamin C Content in Fruits and Berries Consumed in Finland', *J. Food Compos. Anal.*, **8**, 12–20 (1995).
40. J.T. Vanderslice, D.J. Higgs, 'HPLC Analysis with Fluorometric Detection of Vitamin C in Food Samples', *J. Chromatogr. Sci.*, **22**, 485–489 (1984).
41. J.T. Vanderslice, D.J. Higgs, 'Robotic Extraction of Vitamin C from Food Samples', *J. Micronutr. Anal.*, **1**, 143–154 (1985).
42. J.T. Vanderslice, D.J. Higgs, 'Chromatographic Separation of Ascorbic Acid, Isoascorbic Acid, Dehydroascorbic Acid and Dehydroisoacorbic Acid and Their Quantitation in Food Products', *J. Micronutr. Anal.*, **4**, 109–118 (1988).
43. L.L. Lloyd, F.P. Warner, 'Ion Suppression Reversed-phase High-performance Liquid Chromatography Method for the Separation of L-Ascorbic Acid in Fresh Fruit Juice', *J. Chromatogr.*, **437**, 447–452 (1988).
44. H.J. Kim, 'Determination of Total Vitamin C by Ion Exclusion Chromatography with Electrochemical Detection', *J. Assoc. Off. Anal. Chem.*, **72**, 681–686 (1989).
45. H. Iwase, 'Determination of Ascorbic Acid in Elemental Diet by High-performance Liquid Chromatography with Electrochemical Detection', *J. Chromatogr.*, **606**, 277–280 (1992).
46. J.T. Vanderslice, D.J. Higgs, 'Separation of Ascorbic Acid, Isoascorbic Acid, Dehydroascorbic Acid and Dehydroisoascorbic Acid in Food and Animal Tissue', *J. Micronutr. Anal.*, **7**, 67–70 (1990).
47. W.A. Behrens, R. Madère, 'Ascorbic and Dehydroascorbic Acid Contents of Canned Food and Frozen Concentrated Orange Juice', *J. Food Compos. Anal.*, **3**, 3–8 (1990).
48. K.Y. Dodson, E.R. Young, A.M. Soliman, 'Determination of Total Vitamin C in Various Food Matrixes by Liquid Chromatography and Fluorescence Detection', *J. AOAC Int.*, **75**, 887–891 (1992).
49. W.A. Behrens, R. Madère, 'Quantitative Analysis of Ascorbic Acid and Isoascorbic Acid in Foods by High-performance Liquid Chromatography with Electrochemical Detection', *J. Liq. Chromatogr.*, **15**, 753–765 (1992).
50. W.A. Behrens, R. Madère, 'Ascorbic Acid, Isoascorbic Acid, Dehydroascorbic Acid, and Dehydroisoascorbic Acid in Selected Food Products', *J. Food Compos. Anal.*, **7**, 158–170 (1994).
51. J.T. Vanderslice, D.J. Higgs, 'Quantitative Determination of Ascorbic Acid, Dehydroascorbic, Isoascorbic, and Dehydroisoascorbic Acids by HPLC in Foods and Other Matrices', *J. Nutr. Biochem.*, **4**, 184–190 (1993).
52. M.J. Esteve, R. Farré, A. Frigola, J.C. López, J.M. Romera, M. Ramirez, A. Gil, 'Comparison of Voltammetric and High Performance Liquid Chromatographic Methods for Ascorbic Acid Determination in Infant Formulas', *Food Chem.*, **52**, 99–102 (1995).
53. R. Leubolt, H. Klein, 'Determination of Sulphite and Ascorbic Acid by High-performance Liquid Chromatography with Electrochemical Detection', *J. Chromatogr.*, **640**, 271–277 (1993).
54. H.G. Daood, P.A. Biacs, M.A. Dakar, F. Hajdu, 'Ion-pair Chromatography and Photodiode-array Detection of Vitamin C and Organic Acids', *J. Chromatogr. Sci.*, **32**, 481–487 (1994).
55. X. Chen, M. Sato, 'High-performance Liquid Chromatographic Determination of Ascorbic Acid in Soft Drinks and Apple Juice Using Tris(2,2′-bipyridine)ruthenium(II) Electrochemiluminescence', *Anal. Sci.*, **11**, 749–754 (1995).
56. M.J. Oruña-Concha, M.J. Gonzalez-Castro, J. Lopez-Hernandez, 'Monitoring of the Vitamin C Content of Frozen Green Beans and Padròn Peppers by HPLC', *J. Sci. Food Agric.*, **76**, 477–480 (1998).
57. J.T. Vanderslice, M.A. Huang, 'Liquid Chromatographic Analysis of Thiamin and Its Phosphates in Food Products Using Amprolium as an Internal Standard', *J. Micronutr. Anal.*, **2**, 189–199 (1986).
58. E.C. Nicolas, K.A. Pfender, 'Fast and Simple Liquid Chromatographic Determination of Nonphosphorylated Thiamine in Infant Formula, Milk, and Other Foods', *J. Assoc. Off. Anal. Chem.*, **73**, 792–798 (1990).
59. C. Vidal-Valverde, A. Reche, 'An Improved High Performance Liquid Chromatographic Method for Thiamin Analysis in Foods', *Z. Lebensm.-Unters. Forsch.*, **191**, 313–318 (1990).
60. Z.M. Abdel-Kader, 'Comparison of AOAC and High-performance Liquid Chromatographic Methods for Thiamin Determination in Foods', *Food Chem.*, **43**, 393–397 (1992).
61. D. Blanco, M.B. Llaneza, M.D. Gutierrez, 'A Paired Ion Liquid Chromatographic Method for Thiamine Determination in Selected Foods', *J. Liq. Chromatogr. Relat. Technol.*, **19**, 2155–2164 (1996).
62. K. Yamanaka, M. Matsuoka, K. Banno, 'Determination of Thiamine in Dried Yeast by High-performance Liquid Chromatography Using a Clean-up Column of CM-cellulose', *J. Chromatogr. A*, **726**, 237–240 (1996).
63. F. Valls, M.A. Checa, M.A. Fernández-Muiño, M.T. Sancho, 'Determination of Thiamin in Cooked Sausages', *J. Agric. Food Chem.*, **47**, 170–173 (1999).
64. B. Stancher, F. Zonta, 'High Performance Liquid Chromatographic Analysis of Riboflavin (Vitamin B_2) with

Visible Absorbance Detection in Italian Cheeses', *J. Food Sci.*, **51**, 857–858 (1986).
65. S.H. Ashoor, M.J. Knox, J.R. Olsen, D.A. Deger, 'Improved Liquid Chromatographic Determination of Riboflavin in Milk and Dairy Products', *J. Assoc. Off. Anal. Chem.*, **68**, 693–696 (1985).
66. F. Ribarova, S. Shishkov, N. Obretenova, L. Metchkueva, 'Comparative Determination of Riboflavin in Milk by HPLC and Lumiflavin Methods', *Nahrung*, **31**, 77–79 (1987).
67. E.P. Reyes, K.M. Norris, C. Taylor, D. Potts, 'Comparison of Paired-ion Liquid Chromatographic Method with AOAC Fluorometric and Microbiological Methods for Riboflavin Determination in Selected Foods', *J. Assoc. Off. Anal. Chem.*, **71**, 16–19 (1988).
68. V. Ollilainen, P. Mattila, P. Varo, P. Koivistoinen, J. Huttunen, 'The HPLC Determination of Total Riboflavin in Foods', *J. Micronutr. Anal.*, **8**, 199–207 (1990).
69. N. Bilic, R. Sieber, 'Determination of Flavins in Dairy Products by High-performance Liquid Chromatography Using Sorboflavin as Internal Standard', *J. Chromatogr.*, **511**, 359–366 (1990).
70. A. Muñoz, R. Ortiz, M.A. Murcia, 'Determination by HPLC of Changes in Riboflavin Levels in Milk and Nondairy Imitation Milk During Refrigerated Storage', *Food Chem.*, **49**, 203–206 (1994).
71. L.F. Russell, L. Brooks, K.B. McRae, 'Development of a Robotic-HPLC Determination of Riboflavin Vitamers in Foods', *Food Chem.*, **63**, 125–131 (1998).
72. F. Valls, M.T. Sancho, M.A. Fernández-Muiño, M.A. Checa, 'Determination of Total Riboflavin in Cooked Sausages', *J. Agric. Food Chem.*, **47**, 1067–1070 (1999).
73. T. Hamano, Y. Mitsuhashi, N. Aoki, S. Yamamoto, 'Simultaneous Determination of Niacin and Niacinamide in Meats by High-performance Liquid Chromatography', *J. Chromatogr.*, **457**, 403–480 (1988).
74. K. Takatsuki, S. Suzuki, M. Sato, K. Sakai, I. Ushizawa, 'Liquid Chromatographic Determination of Free and Added Niacin and Niacinamide in Beef', *J. Assoc. Off. Anal. Chem.*, **70**, 698–702 (1987).
75. C. Vidal-Valverde, A. Reche, 'Determination of Available Niacin in Legumes and Meat by High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **39**, 116–121 (1991).
76. G.W. Chase Jr, W.O. Landen Jr, A.M. Soliman, R.R. Eitenmiller, 'Liquid Chromatographic Analysis of Niacin in Fortified Food Products', *J. AOAC Int.*, **76**, 390–393 (1993).
77. D.E. Lacroix, W.R. Wolf, J.T. Vanderslice, 'Determination of Niacin in Infant Formula and Wheat Flour by Anion-exchange Liquid Chromatography with Solid-phase Extraction Cleanup', *J. AOAC Int.*, **82**, 128–133 (1999).
78. S. Lahély, M. Bergaentzlé, C. Hasselmann, 'Fluorimetric Determination of Niacin in Foods by High-performance Liquid Chromatography with Post-column Derivatization', *Food Chem.*, **65**, 129–133 (1999).
79. L.A. Morrison, J.A. Driskell, 'Quantities of B_6 Vitamers in Human Milk by High-performance Liquid Chromatography', *J. Chromatogr.*, **337**, 249–258 (1985).
80. C.Y.W. Ang, M. Cenciarelli, R.R. Eitenmiller, 'A Simple Liquid Chromatographic Method for Determination of B_6 Vitamers in Raw and Cooked Chicken', *J. Food Sci.*, **53**, 371–375 (1988).
81. T. Toukaairin-Oda, E. Sakamoto, N. Hirose, M. Mori, T. Itoh, H. Tsuge, 'Determination of Vitamin B_6 Derivatives in Foods and Biological Materials by Reversed-phase HPLC', *J. Nutr. Sci. Vitaminol.*, **35**, 171–180 (1989).
82. K. Tadera, U. Naka, 'Isocratic Paired-ion High-performance Liquid Chromatographic Method to Determine B_6 Vitamers and Pyridoxine Glucoside in Foods', *Agric. Biol. Chem.*, **52**, 563–564 (1991).
83. D.A. Sampson, L.A. Eoff, X.L. Yan, K. Lorenz, 'Analysis of Free and Glycosylated Vitamin B_6 in Wheat by High-performance Liquid Chromatography', *Cereal Chem.*, **72**, 217–221 (1995).
84. C.J. Argoudelis, 'Simple High-performance Liquid Chromatographic Method for the Determination of all Seven Vitamin B_6-related Compounds', *J. Chromatogr. A*, **790**, 83–91 (1997).
85. M.J. Esteve, R. Frré, A. Frígola, J.M. García-Cantabella, 'Determination of Vitamin B_6 (Pyridoxamine, Pyridoxal and Pyridoxine) in Pork Meat and Pork Meat Products by Liquid Chromatography', *J. Chromatogr. A*, **795**, 383–387 (1998).
86. J. Selhub, 'Determination of Tissue Folate Composition by Affinity Chromatography Followed by High-pressure Ion Pair Liquid Chromatography', *Anal. Biochem.*, **182**, 84–93 (1989).
87. E. Seyoum, J. Selhub, 'Combined Affinity and Ion Pair Column Chromatographies for the Analysis of Food Folate', *J. Nutr. Biochem.*, **4**, 488–494 (1993).
88. C.M. Pfeiffer, L.M. Rogers, J.F. Gregory, 'Determination of Folate in Cereal-grain Food Products Using Trienzyme Extraction and Combined Affinity and Reversed-phase Liquid Chromatography', *J. Agric. Food Chem.*, **45**, 407–413 (1997).
89. D.L. Holt, R.L. Wehling, M.G. Zeece, 'Determination of Native Folates in Milk and Other Dairy Products by High-performance Liquid Chromatography', *J. Chromatogr.*, **449**, 271–279 (1988).
90. D.R. White Jr, 'Determination of 5-Methyltetrahydrofolate in Citrus Juice by Reversed-phase High-performance Liquid Chromatography with Electrochemical Detection', *J. Agric. Food Chem.*, **38**, 1515–1518 (1990).
91. D.R. White Jr, H.S. Lee, R.E. Krüger, 'Reversed-phase HPLC/EC Determination of Folate in Citrus Juice by Direct Injection with Column Switching', *J. Agric. Food Chem.*, **39**, 714–717 (1991).

92. E.J.M. Konings, 'A Validated Liquid Chromatographic Method for Determining Folates in Vegetables, Milk Powder, Liver, and Flour', *J. AOAC Int.*, **82**, 119–127 (1999).
93. E.S. Osseyi, R.L. Whling, J.A. Albrecht, 'Liquid Chromatographic Method for Determining Added Folic Acid in Fortified Cereal Products', *J. Chromatogr. A*, **826**, 235–240 (1998).
94. S. Lahély, S. Ndaw, F. Arella, C. Hasselmann, 'Determination of Biotin in Foods by High-performance Liquid Chromatography with Post-column Derivatization and Fluorimetric Detection', *Food Chem.*, **65**, 253–258 (1999).
95. A. Sims, D. Shoemaker, 'Simultaneous Liquid Chromatographic Determination of Thiamine and Riboflavin in Selected Foods', *J. AOAC Int.*, **76**, 1156–1160 (1993).
96. S. Albalá-Hurtado, M.T. Veciana-Nogués, M. Izquierdo-Pulido, 'Determination of Water-soluble Vitamins in Infant Milk by High-performance Liquid Chromatography', *J. Chromatogr. A*, **778**, 247–253 (1997).
97. F. Brown, 'The Estimation of Vitamin E. 2. Quantitative Analysis of Tocopherol Mixtures by Paper Chromatography', *Biochem. J.*, **52**, 523–526 (1952).
98. J. Green, D. McHale, S. Marcinkiewicz, P. Mamalis, P.R. Watt, 'Tocopherols, Part V. Structural Studies on epsilon- and delta-Tocopherol', *J. Chem. Soc.*, 3362–3373 (1959).
99. J. Green, P. Mamalis, S. Marcinkiewicz, D. McHale, 'Structure of Epsilon-Tocopherol', *Chem. Ind. (London)*, 73–74 (1960).
100. I.D. Desai, 'Assay Methods', in *Vitamin E, a Comprehensive Treaties*, ed. L.J. Machlin, Marcel Dekker, New York, 67–98, 1980.
101. A.P. De Leenheer, H.J. Nelis, W.E. Lambert, R.M. Bauwens, 'Chromatography of Fat-soluble Vitamins in Clinical Chemistry', *J. Chromatogr.*, **429**, 3–58 (1988).
102. R. Wyss, 'Chromatographic and Electrophoretic Analysis of Biomedically Important Retinoids', *J. Chromatogr. B*, **671**, 381–425 (1995).
103. H.J. Nelis, V.O.R.C. De Bevere, A.P. De Leenheer, 'Vitamin E: Tocopherols and Tocotrienols', in *Modern Chromatographic Analysis of the Vitamins*, eds. A.P. De Leenheer, W.E. Lambert, M.G.M. De Ruyter, Marcel Dekker, New York, 129, 1985.
104. A.P. De Leenheer, W.E. Lambert, E. Meyer, *Progress in Research and Clinical Applications*, Marcel Dekker, New York, 1993.
105. R. Wyss, 'Chromatography of Retinoids', *J. Chromatogr.*, **531**, 481–508 (1990).
106. W.E. Lambert, H.J. Nelis, M.G.M. De Ruyter, A.P. De Leenheer, *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1985.
107. G.T. Vatassery, V.R. Maynard, D.F. Hagen, 'High-performance Liquid Chromatography of Various Tocopherols', *J. Chromatogr.*, **161**, 299–302 (1978).
108. E. Westerberg, M. Friber, B. Alesson, 'Assay of Brain Tocopherols Using High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **4**, 109–121 (1981).
109. B. Håkansson, M. Jägerstad, R. Oste, 'Determination of Vitamin E in Wheat Products by HPLC', *J. Micronutr. Anal.*, **3**, 307–318 (1987).
110. B. Tan, L. Brzuskiewicz, 'Separation of Tocopherol and Tocotrienol Isomers Using Normal- and Reversed-phase Liquid Chromatography', *Anal. Biochem.*, **180**, 368–373 (1989).
111. Y. Satomura, M. Kimura, Y. Itokawa, 'Simultaneous Determination of Retinol and Tocopherols by High-performance Liquid Chromatography', *J. Chromatogr.*, **625**, 372–376 (1992).
112. M.K. Balz, E. Schulte, H.P. Thier, 'Simultaneous determination of α-Tocopheryl Acetate, Tocopherols and Tocotrienols by HPLC with Fluorescence Detection in Foods', *Fat Sci. Technol.*, **95**, 215–220 (1993).
113. J.N. Thompson, G. Hatina, 'Determination of Tocopherols and Tocotrienols in Foods and Tissues by HPLC', *J. Liq. Chromatogr.*, **2**, 327–344 (1979).
114. H.E. Indyk, 'Simplified Saponification Procedure for the Routine Determination of Total Vitamin E in Dairy Products, Foods and Tissues by High-performance Liquid Chromatography', *Analyst*, **113**, 1217–1221 (1988).
115. C. Maraschiello, J.A.G. Regueiro, 'Procedure for the Determination of Retinol and α-Tocopherol in Poultry Tissues Using Capillary Gas Chromatography with Solvent Venting Injection', *J. Chromatogr. A*, **818**, 109–121 (1998).
116. J.K.G. Kramer, L. Blais, R.C. Fouchard, 'A Rapid Method for the Determination of Vitamin E Forms in Tissues and Diet by High-performance Liquid Chromatography Using a Normal-phase Diol Column', *Lipids*, **32**, 323–330 (1997).
117. A.K. Hewavitharana, A.S. van Brakel, 'A Rapid Saponification Procedure for the Simultaneous HPLC Determination of the Natural Levels of Vitamins A, E and β-carotene in Milk', *Milchwissenschaft*, **53**, 11–15 (1998).
118. S. Albalà-Hurtado, S. Novella-Rodriguez, M.T. Veciana-Nogués, 'Determination of Vitamins A and E in Infant Milk Formulae by High-performance Liquid Chromatography', *J. Chromatogr. A*, **778**, 243–246 (1997).
119. T. Ueda, O. Igarashi, 'New Solvent System for Extraction of Tocopherols from Biological Specimens for HPLC Determination and the Evaluation of 2,2,5,7,8-Pentamethyl-6-chromanol as an Internal Standard', *J. Micronutr. Anal.*, **3**, 185–198 (1987).
120. D.B. Dennison, J.R. Kirk, 'Quantitative Analysis of Vitamin A in Cereal Products by High Speed Liquid Chromatography', *J. Food Sci.*, **42**, 1376–1379 (1977).
121. J.I. Rader, C.M. Weaver, L. Patrascu, H. Ali, G. Gangyal, 'α-Tocopherol, Total Vitamin A and Total Fat in Margarines and Margarine-like Products', *Food Chem.*, **58**, 373–379 (1997).

122. V. Piironen, E.L. Syvaoja, P. Varo, K. Salminen, P. Koivistoinen, 'Tocopherols and Tocotrienols in Finnish Foods: Vegetables, Fruits and Berries', *J. Agric. Food Chem.*, **34**, 742–746 (1986).
123. E.E. Syvaoja, V. Piironen, P. Varo, P. Korvistoinen, K. Salminen, 'Tocopherols and Tocotrienols in Finnish Foods: Human Milk and Infant Formulas', *Int. J. Vitam. Nutr. Res.*, **55**, 159–166 (1985).
124. G. Panfili, P. Manzi, L. Pezzoferrato, 'High-performance Liquid Chromatographic Method for the Simultaneous Determination of Tocopherols, Carotenes, and Retinol and its Geometric Isomers in Italian Cheese', *Analyst*, **119**, 1161–1165 (1994).
125. E.J.M. Konings, H.H.S. Roomans, R.R. Beljaars, 'Liquid Chromatographic Determination of Tocopherols and Tocotrienols in Margarine, Infant Foods, and Vegetables', *J. AOAC Int.*, **79**, 902–906 (1996).
126. T. Ueda, O. Igarashi, 'Effect of Coexisting Fat on the Extraction of Tocopherols from Tissues after Saponification as a Pretreatment of HPLC Determination', *J. Micronutr. Anal.*, **3**, 15–25 (1987).
127. V. Piironen, P. Varo, E.L. Syvaoja, K. Salminen, P. Koivistoinen, 'High-performance Liquid Chromatography Determination of Tocopherols and Tocotrienols and its Application to Diets and Plasma of Finnish Men', *Int. J. Vitam. Nutr. Res.*, **53**, 35–40 (1984).
128. C.K. Chow, H.H. Draper, A.S. Csallany, 'Method for the Assay of Free and Esterified Tocopherols', *Anal. Biochem.*, **32**, 81–90 (1969).
129. C.J. Hogarty, C. Ang, R.R. Eitenmiller, 'Tocopherol Content of Selected Foods by HPLC/Fluorescence Quantitation', *J. Food Compos. Anal.*, **2**, 200–209 (1989).
130. G. Lietz, C.J.K. Henry, 'A Modified Method to Minimize Losses of Carotenoids and Tocopherols During HPLC Analysis of Red Palm Oil', *Food Chem.*, **60**, 109–117 (1997).
131. S.A. Barnett, L.W. Frick, H.M. Baine, 'Simultaneous Determination of Vitamins A, D_2 or D_3, E and K_1 in Infant Formulas and Dairy Products by Reversed-phase Liquid Chromatography', *Anal. Chem.*, **52**, 610–614 (1980).
132. F. Yao, G. Dull, R.R. Eitenmiller, 'Tocopherol Quantification by HPLC in Pecans and Relationship to Kernel Quality During Storage', *J. Food Sci.*, **57**, 1194–1197 (1992).
133. M.L. Colombo, A. Corsini, A. Mossa, L. Sala, M. Stanca, 'Supercritical Carbon Dioxide Extraction, Fluorometric and Electrochemical High Performance Liquid Chromatographic Detection of Vitamin E from *Hordeum vulgare*', *Phytochem. Anal.*, **9**, 192–195 (1998).
134. R. Marsili, D. Callahan, 'Comparison of a Liquid Solvent Extraction Technique and Supercritical Fluid Extraction for the Determination of α- and β-carotene in Vegetables', *J. Chromatogr. Sci.*, **31**, 422–428 (1993).
135. M.L. Huang, G.J. Burckart, R. Venkataramanan, 'Sensitive High-performance Liquid Chromatographic Analysis of Plasma Vitamin E and Vitamin A Using Amperometric and Ultraviolet Detection', *J. Chromatogr.*, **380**, 331–338 (1986).
136. W.M. Cort, T.S. Vicente, E.H. Waysek, B.D. Williams, 'Vitamin E Content of Feedstuffs Determined by High-performance Liquid Chromatographic Fluorescence', *J. Agric. Food Chem.*, **31**, 1330–1333 (1983).
137. K. Seta, H. Nakamura, T. Okuyama, 'Determination of α-Tocopherol, Free Cholesterol, Esterified Cholesterols and Triacylglycerols in Human Lipoproteins by High-performance Liquid Chromatography', *J. Chromatogr.*, **515**, 585–590 (1990).
138. E. Teissier, E. Walters-Laporte, C. Duhem, G. Luc, J.C. Fruchar, P. Duriez, 'Rapid Quantification of α-Tocopherol Plasma and Low- and High-density Lipoproteins', *Clin. Chem.*, **42**, 430–435 (1996).
139. D. Hoehler, A.A. Frohlich, R.R. Marquardt, H. Stelsovsky, 'Extraction of α-Tocopherol from Serum Prior to Reversed-phase Liquid Chromatography', *J. Agric. Food Chem.*, **46**, 973–978 (1998).
140. A.T. Rhys Williams, 'Simultaneous Determination of Serum Vitamin A and E by Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **341**, 198–201 (1985).
141. T. Ueda, O. Igarashi, 'Evaluation of the Electrochemical Detector for the Determination of Tocopherols in Feeds by High-performance Liquid Chromatography', *J. Micronutr. Anal.*, **1**, 31–38 (1985).
142. M.A. Schneiderman, A.K. Sharma, D.C. Locke, 'Determination of Vitamin A Palmitate in Cereal Products Using Supercritical Fluid Extraction and Liquid Chromatography with Electrochemical Detection', *J. Chromatogr. A*, **765**, 215–220 (1997).
143. F. Dionisi, J. Prodolliet, E. Tagliaferri, 'Assessment of Olive Oil Adulteration by Reversed-phase High-performance Liquid Chromatography/Amperometric Detection of Tocopherols and Tocotrienols', *J. Am. Oil Chem. Soc.*, **72**, 1505–1511 (1995).
144. E.J. de Vries, C.J. Olling, U. Manz, E. Tagliaferri, 'Dried Skimmed Milk – Determination of Vitamin A – Colorimetric and Liquid Chromatographic Methods', *Bulletin of the International Dairy Federation*, **285**, 53–64 (1993).
145. R. Marsh, P. Kajda, J. Ryley, 'The Effect of Light on the Vitamin A Content of Cheess', *Nahrung*, **38**, 527–532 (1994).
146. S.K. Jensen, 'Retinol Determination in Milk by HPLC and Fluorescence Detection', *J. Dairy Res.*, **61**, 233–240 (1994).
147. E. Brinkmann, I. Mehlitz, H.B. Oei, R. Tiebach, W. Baltes, 'Determination of Vitamin A in Liver and Liver-containing Products Using Narrow-bore Normal Phase HPLC', *Z. Lebensm.-Unters. Forsch.*, **199**, 206–209 (1994).

148. E. Brinkmann, L. Dehne, H.B. Oei, R. Tiebach, W. Baltes, 'Separation of Geometrical Retinol Isomers in Food Samples by Using Narrow-bore High-performance Liquid Chromatography', *J. Chromatogr. A*, **693**, 271–279 (1995).
149. M. Aké, A.M. Contento, B. Mandrou, M.D. Blanchin, H. Fabre, 'Experimental and Statistical Approach for Validating an HPLC Assay for the Determination of Vitamin A in Flour and Milk Powder for Children', *Chromatographia*, **47**, 716–720 (1998).
150. F. Granado, B. Olmedilla, I. Blanco, E. Rojas-Hidalgo, 'Carotenoid Composition in Raw and Cooked Spanish Vegetables', *J. Agric. Food Chem.*, **40**, 2135–2140 (1992).
151. P.A. Biacs, H.G. Daood, 'High-performance Liquid Chromatography with Photodiode-array Detection of Carotenoids and Carotenoid Esters in Fruits and Vegetables', *J. Plant Physiol.*, **143**, 520–525 (1994).
152. L.H. Tonucci, J.M. Holden, G.R. Beecher, F. Khachik, C.S. Davis, G. Mulokozi, 'Carotenoid Content of Thermally Processed Tomato-based Food Products', *J. Agric. Food Chem.*, **43**, 579–586 (1995).
153. H. Nyambaka, J. Ryley, 'An Isocratic Reversed-phase HPLC Separation of the Stereoisomers of the Provitamin A Carotenoids (α- and β-carotene) in Dark Green Vegetables', *Food Chem.*, **55**, 63–72 (1996).
154. A.Z. Mercadante, D.B. Rodriguez-Amaya, G. Britton, 'HPLC and Mass Spectrometric Analysis of Carotenoids from Mango', *J. Agric. Food Chem.*, **45**, 120–123 (1997).
155. A. Ben-Amotz, R. Fishler, 'Analysis of Carotenoids with Emphasis on 9-*cis*-β-carotene in Vegetables and Fruits Commonly Consumed in Israel', *Food Chem.*, **62**, 515–520 (1998).
156. A.M. Pupin, M.J. Dennis, M.C.F. Toledo, 'HPLC Analysis of Carotenoids in Orange Juice', *Food Chem.*, **64**, 269–275 (1999).
157. D.C. Sertl, B.E. Molitor, 'Liquid Chromatographic Determination of Vitamin D in Milk and Infant Formula', *J. Assoc. Off. Anal. Chem.*, **68**, 177–182 (1985).
158. J.T. Tanner, S.A. Barnett, M.K. Mountford, 'Analysis of Milk-based Infant Formula, Phase IV. Iodide Linoleic Acid, and Vitamins D and K: US Food and Drug Administration–Infant Formula Council: Collaborative Study', *J. AOAC Int.*, **76**, 1042–1056 (1993).
159. M.G. Sliva, A.E. Green, J.K. Sanders, J.R. Euber, J.R. Saucerman, 'Reversed-phase Liquid Chromatographic Determination of Vitamin D in Infant Formulas and Enteral Nutritionals', *J. AOAC Int.*, **75**, 566–571 (1992).
160. M.G. Sliva, J.K. Sanders, 'Vitamin D in Infant Formula and Enteral Products by Liquid Chromatography: Collaborative Study', *J. AOAC Int.*, **79**, 73–80 (1996).
161. P. Mattila, V. Piironen, E. Uusi-Rauva, P. Koivistoinen, 'Cholecalciferol and 25-Hydroxycholecalciferol Contents in Fish and Fish Products', *J. Food Compos. Anal.*, **8**, 232–243 (1995).
162. H. Hasegawa, 'Vitamin D Determination Using High-performance Liquid Chromatography with Internal Standard–redox Mode Electrochemical Detection and Its Application to Medical Nutritional Products', *J. Chromatogr.*, **605**, 215–220 (1992).
163. P. Mattila, V. Piironen, C. Bäckman, A. Asunmaa, E. Uusi-Rauva, P. Koivistoinen, 'Determination of Vitamin D_3 in Egg Yolk by High-performance Liquid Chromatography with Diode Array Detection', *J. Food Compos. Anal.*, **5**, 281–290 (1992).
164. J.N. Thompson, L. Plouffe, 'Determination of Cholecalciferol in Meat and Fat from Livestock Fed Normal and Excessive Quantities of Vitamin D', *Food Chem.*, **46**, 313–318 (1993).
165. P.H. Mattila, V.I. Piironen, E.J. Uusi-Rauva, P.E. Koivistoinen, 'Vitamin D Contents in Edible Mushrooms', *J. Agric. Food. Chem.*, **42**, 2449–2453 (1994).
166. P.H. Mattila, V.I. Piironen, E.J. Uusi-Rauva, P.E. Koivistoinen, 'New Analytical Aspects of Vitamin D in Foods', *Food Chem.*, **57**, 95–99 (1996).
167. E.L. Syväoja, V. Piironen, P. Varo, P. Koivistoinen, K. Salminen, 'Tocopherols and Tocotrienols in Finnish Foods: Oils and Fats', *J. Am. Oil Chem. Soc.*, **63**, 328–329 (1986).
168. E.J. Rogers, S.M. Rice, R.J. Nicolosi, D.R. Carpenter, C.A. McClelland, L.J. Romanczyk Jr, 'Identification and Quantitation of γ-Oryzanol Components and Simultaneous Assessment of Tocols in Rice Bran Oil', *J. Am. Oil Chem. Soc.*, **70**, 301–307 (1993).
169. V. Piironen, E.L. Syväoja, P. Varo, K. Salminen, P. Koivistoinen, 'Tocopherols and Tocotrienols in Finnish Foods: Meat and Meat Products', *J. Agric. Food Chem.*, **33**, 1215–1218 (1985).
170. E.L. Syväoja, K. Salminen, V. Piironen, P. Varo, D. Kerojoki, P. Koivistoinen, 'Tocopherols and Tocotrienols in Finnish Foods: Fish and Fish Products', *J. Am. Oil Chem. Soc.*, **62**, 1245–1248 (1985).
171. V. Piironen, E.L. Syväoja, P. Varo, K. Salminen, P. Koivistoinen, 'Tocopherols and Tocotrienols in Cereal Products from Finland', *Cereal Chem.*, **63**, 78–81 (1986).
172. V. Piironen, P. Varo, P. Koivistoinen, 'Stability of Tocopherols and Tocotrienols in Food Preparation Procedures', *J. Food Compos. Anal.*, **1**, 53–58 (1987).
173. C.Y.W. Ang, G.K. Searcy, R.R. Eitenmiller, 'Tocopherols in Chicken Breast and Leg Muscles Determination by Reverse Phase Liquid Chromatography', *J. Food Sci.*, **55**, 1536–1539 (1990).
174. I.B. Hashim, P.E. Koehler, R.R. Eitenmiller, 'Tocopherols in Runner and Virginia Peanut Cultivars at Various Maturity Stages', *J. Am. Oil Chem. Soc.*, **70**, 633–635 (1993).
175. I.B. Hashim, P.E. Koehler, R.R. Eitenmiller, C.K. Kvien, 'Fatty Acid Composition and Tocopherol Content of Drought Stressed Florunner Peanuts', *Peanut Sci.*, **20**, 21–24 (1993).

176. J.W. Lehmann, D.H. Putnam, A.A. Qureshi, 'Vitamin E Isomers in Grain Amaranths (*Amaranthus* spp.)', *Lipids*, **29**, 177–181 (1994).
177. J.A. Osuna-Garcia, M.M. Wall, C.A. Waddell, 'Endogenous Levels of Tocopherols and Ascorbic Acid During Fruit Ripening of New Mexican-type Chile (*Capsicum annuum* L.) Cultivars', *J. Agric. Food Chem.*, **46**, 5093–5096 (1998).
178. L. Ye, W.O. Landen Jr, J. Lee, R.R. Eitenmiller, 'Vitamin E Content of Margarine and Reduced Fat Products Using a Simplified Extraction Procedure and HPLC Determination', *J. Liq. Chromatogr. Relat. Technol.*, **21**, 1227–1238 (1998).
179. M.P. Bueno, M.C. Villalobos, 'Reverse Phase High Pressure Liquid Chromatographic Determination of Vitamin K_1 in Infant Formula', *J. Assoc. Off. Anal. Chem.*, **66**, 1063–1066 (1983).
180. E. Jakob, I. Elmadfa, 'Application of a Simplified HPLC Assay for the Determination of Phylloquinone (Vitamin K_1) in Animal and Plant Food Items', *Food Chem.*, **56**, 87–91 (1996).
181. M. Careri, A. Mangia, P. Manini, N. Taboni, 'Determination of Phylloquinone (Vitamin K_1) by High Performance Liquid Chromatography with UV Detection and with Particle Beam–Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **355**, 48–56 (1996).
182. T. Koivu, V. Piironen, P. Mattila, 'Phylloquinone (Vitamin K_1) in Cereal Products', *Cereal Chem.*, **75**, 113–116 (1998).
183. V. Piironen, T. Koivu, O. Tammisalo, P. Mattila, 'Determination of Phylloquinone in Oils, Margarines and Butter by High-performance Liquid Chromatography with Electrochemical Detection', *Food Chem.*, **59**, 473–480 (1997).
184. J.N. Thompson, G. Hatina, W.B. Maxwell, 'High Performance Liquid Chromatographic Determination of Vitamin A in Margarine, Milk, Partially Skimmed Milk, and Skimmed Milk', *J. Assoc. Off. Anal. Chem.*, **63**, 894–898 (1980).
185. W.O. Landen Jr, R.R. Eitenmiller, 'Application of Gel Permeation Chromatography and Non-aqueous Reverse Phase Chromatography to High Pressure Liquid Chromatographic Determination of Retinyl Palmitate and β-carotene in Oil and Margarine', *J. Assoc. Off. Anal. Chem.*, **62**, 283–289 (1979).
186. W.O. Landen Jr, 'Application of Gel Permeation Chromatography and Non-aqueous Reverse Phase Chromatography to High Performance Liquid Chromatographic Determination of Retinyl Palmitate and α-Tocopheryl Acetate in Infant Formulas', *J. Assoc. Off. Anal. Chem.*, **65**, 810–816 (1982).
187. W.O. Landen, D.M. Hines, T.W. Hamill, J.I. Martin, E.R. Young, R.R. Eitenmiller, A.M. Soliman, 'Vitamin A and Vitamin E Content of Infant Formulas Produced in the United States', *J. Assoc. Off. Anal. Chem.*, **68**, 509–511 (1985).
188. W.O. Landen, R.R. Eitenmiller, A.M. Soliman, 'Vitamin D_3 and Vitamin K_1 Levels in Infant Formula Produced in the United States', *J. Food Compos. Anal.*, **2**, 140–148 (1989).
189. J.E. Chappell, T. Francis, M.T. Clandinin, 'Simultaneous High Performance Liquid Chromatography Analysis of Retinol Ester and Tocopherol Isomers in Human Milk', *Nutr. Res.*, **6**, 849–852 (1986).
190. E.S. Tee, C.L. Lim, 'Reanalysis of Vitamin A Values of Selected Malaysian Foods of Animal Origin by the AOAC and HPLC Methods', *Food Chem.*, **45**, 289–296 (1992).
191. M.M.D. Zamarreño, A.S. Perez, M.C.G. Pereze, J.H. Mendez, 'Directly Coupled Sample Treatment–High-performance Liquid Chromatography for On-line Automatic Determination of Liposoluble Vitamins in Milk', *J. Chromatogr. A*, **694**, 399–406 (1995).
192. M.M.D. Zamarreño, A.S. Perez, M.C.G. Pereze, J.H. Mendez, 'High Performance Liquid Chromatography with Electrochemical Detection for the Simultaneous Determination of Vitamin A, D_3, and E in Milk', *J. Chromatogr.*, **623**, 69–74 (1992).
193. M.M.D. Zamarreño, A.S. Perez, M.C.G. Pereze, M.A.F. Moro, J.H. Mendez, 'Determination of Vitamins A, E, and K_1 in Milk by High-performance Liquid Chromatography with Dual Amperometric Detection', *Analyst*, **120**, 2489–2492 (1995).
194. H. Qian, M. Sheng, 'Simultaneous Determination of Fat-soluble Vitamins A, D and E and Pro-vitamin D_2 in Animal Feeds by One-step Extraction and High-performance Liquid Chromatography Analysis', *J. Chromatogr. A*, **825**, 127–133 (1998).

Water Determination in Food

David Reid
Department of Food Science and Technology, University of California, Davis, USA

1 Introduction **2**
1.1 Special Precautions Required for Meaningful Measurement 3
2 Gravimetric Water Content **3**
2.1 Methods of Sample Handling 3
2.2 Oven Drying Methods 4
2.3 Distillation Methods 5
2.4 Chemical Methods 6
2.5 Indirect Methods of Water Content Determination 7
2.6 Thermogravimetric Analysis 8
2.7 Solvent Extraction 8
3 Water Activity (or Equilibrium Water Vapor Pressure) **9**
3.1 Vapor Pressure Manometer 9
3.2 Dew Point Determination 9
3.3 Water Activity Sensors 11
3.4 Measuring Related Thermodynamic Quantities 12
3.5 Isopiestic Methods 12
3.6 Sorption Isotherms 13
4 Categories of Water **13**
4.1 "Bound" Water 14
4.2 Water Mobility 14
Abbreviations and Acronyms **14**
Related Articles **14**
References **14**

Water determination in foods is a deceptively simple theme. Defining the quantity to be measured identifies the inherent complexity. Three separate types of measure may be appropriate: a gravimetric measure; a measure related to vapor pressure; and a measure of classes of water. The ubiquitous nature of water in our environment provides additional complexity in the challenge of preventing transfer of water between sample and environment. The earliest measures of amount of water were all gravimetric, determining the weight fraction of water in the food. These methods range from simple direct weighing using a difference technique, to more complex methods where the amount of water is determined by spectroscopic methods or by chemical assay. A wide range of methods have been developed and are in daily use because gravimetric water content is important for formulation and for labeling purposes. This measure, however, is of little value for the prediction of the stability of a food, even though water plays a critical role in determining the stability characteristics of foods. For a measure of the amount of water relevant to stability concerns, the vapor pressure or its related thermodynamic parameters is more relevant. Determination of vapor pressure uses methods developed from thermodynamic roots, although if the product is not at true equilibrium the measured quantity is not a thermodynamic descriptor of the product but it is still a measure of a product characteristic. Rather than describing necessarily the thermodynamic state of the water in the food, these measures all determine the vapor pressure of the atmosphere in contact with the food. Whether this is the equilibrium vapor pressure, and therefore the thermodynamic description, has to be established on a case by case basis. The vapor pressure per se is an important parameter. Vapor pressure can be measured directly in a manometric system or can be inferred from the dew point temperature (at which the relative humidity just reaches 100%). It can also be inferred from the electrical characteristics of sensors that have electrical conductance or capacitance that is a function of the relative humidity of their immediate environment, once sensor and immediate environment are equilibrated.

A special set of applications measure both the gravimetric water content and the vapor pressure that corresponds to this water content. The result is a sorption isotherm. In this case, as an alternative to measuring the vapor pressure, a stable, known vapor pressure can be established and the product allowed to equilibrate to constant weight. The establishment of an atmosphere of known vapor pressure/relative humidity is the essential component of isopiestic techniques. On the assumption of equilibrium, analysis of sorption isotherms according to various models can lead to categorizing various classes of water and identifying the amount of water in each class. Spectroscopic measurements can also lead, more directly, to the identification of the amount of water in different classes. Many different labels, such as bound water, unfreezable water and restricted water, are attached to the categories of water, depending upon the method used to distinguish between the classes. For direct and indirect estimation of vapor pressures it is critically important that the temperature of the sample be stable and accurately known, and also that the temperature in the region where the measurement that will translate into vapor pressure is to be made is accurately known. This is because the driver for many properties that depend upon water content is the relative vapor pressure, and the vapor pressure of pure water has a steep dependence

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

upon temperature. It is also important that sufficient time be taken to allow for the establishment of a uniform and steady vapor pressure throughout the vapor spaces of the system, otherwise moisture transfer will occur, driven by the pressure gradients, changing the characteristics of the materials.

1 INTRODUCTION

The theme of ''water determination in foods'' is deceptively simple. It is necessary as a first step to define the quantity to be measured. This immediately leads us to the inherent complexity of the topic. The amount of water can be identified through the use of three separate types of measure: a measure of the gravimetric quantity of water; a measure of the thermodynamic impact of the water; and a measure of the amounts of water that fall into different classifications. Because water is a ubiquitous component of our environment, there are significant challenges in preventing the transfer of water to or from the sample under study. This results in additional complexity in many of the measurement systems, in an attempt to get a clear and repeatable definition of the sample and of the conditions of measurement.

Historically, the earliest measures of amount of water were gravimetric measurements: the weight fraction of water in the food. Although this is a useful quantity to determine, its value in understanding such factors as the stability of the food and the susceptibility to microbial attack is very limited. Pioneering publications by Scott[1,2] underlined the importance of a thermodynamic characterization of the ''effective amount of water'', otherwise termed water activity. This is essentially a description of the amount of water by a measure of the equilibrium vapor pressure of water in the atmosphere in contact with the food, in contrast to a measure of the weight of water in the food. The information that we seek is often the water activity at a known gravimetric water content, hence there is a need to determine both water activity and moisture content. Descriptions of systems designed for this purpose are given by Gal[3,4] in a discussion of methods for the determination of sorption isotherms. A schematic of a typical system is shown in Figure 1, and a typical shape for a sorption isotherm is illustrated in Figure 2. In recent years, attempts to understand the dependence of the properties of foods in terms of their water contents, whether defined as gravimetric water content or as water activity, have led to models that categorize the fractions of water that appear to have different properties. This led to such measures as unfreezable water, bound water, etc. Sorption isotherms, for example, have been analyzed, using models, to arrive at estimates of the amount of water in different categories as indicated in Figure 2, which

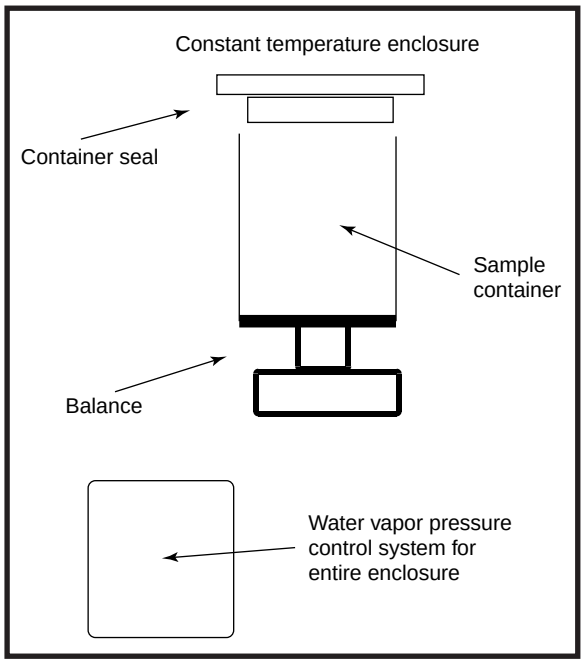

Figure 1 Schematic diagram of an apparatus to determine sorption isotherms. The sample is introduced into the enclosure in a sealed container, and the container seal is opened to initiate the sorption process. Sample weight is monitored. The entire enclosure is maintained as close to isothermal as possible, and a known partial vapor pressure of water is maintained.

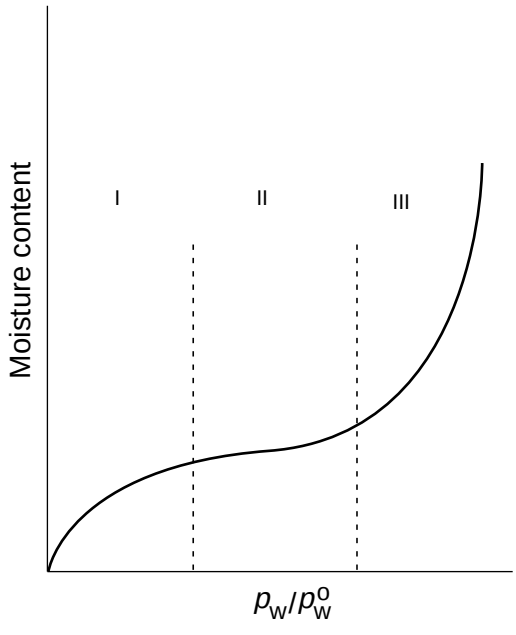

Figure 2 Schematic representation of a sorption isotherm. Zones I, II and III represent low, intermediate and high moisture.

identifies ranges for different moisture categories.[5–7] Allocation of boundaries is qualitative at best. In more general terms, identification of different categories can be seen as identifying the amounts of water that have molecular mobilities in certain ranges. Clearly solid and liquid (i.e. unfrozen) is a simple division, but how liquid is liquid? More complex divisions can be made. Modern advances in spectroscopic techniques allow for the estimation of molecular mobilities and for the quantification of the amounts of water in various mobility classifications. In recent years, categorization of water mobilities has become an important descriptor of the state and amount of water in foods. It must be realized, however, that the range of molecular mobilities is a continuum, and any separation into categories is in some way arbitrary.

This article will discuss the means of estimating gravimetric water content and also water activity. It will conclude with a brief discussion of water mobility, etc.

1.1 Special Precautions Required for Meaningful Measurement

Water is a ubiquitous component of our environment. Exchange of water between materials in contact is common. Atmospheric moisture may be picked up readily by a material if the conditions are favorable. Similarly, moisture may be lost readily to the atmosphere if the conditions are favorable. Because moisture is so readily exchanged, it is important to define the conditions of measurement and to define the state of the sample if a meaningful measure of water content is to be made and reported. It is appropriate to discuss methods of equilibration, of sample collection and the handling required to maintain a sample in a defined condition. The precautions required are dependent upon the particular measure of moisture content that is to be made. For gravimetric determination, it is important that no moisture be gained or lost from the sample, once it has been collected for assay. For vapor pressure measurement, it is important that the amount of moisture that transfers to establish the vapor pressure in the measuring space does not cause a significant change in the equilibrium vapor pressure. It is also important for the temperature to remain constant, because the saturation vapor pressure of water is very temperature dependent. Because of the temperature dependence of vapor pressure, temperature control is an important aspect of moisture measurement. Although absolute vapor pressure is very much temperature dependent, equilibrium relative vapor pressure, is much less dependent upon temperature. Thus, for measurement of relative vapor pressure, the uniform temperature is necessary, whereas for measurement of absolute vapor pressure a precise uniform temperature should be established. The significance of this is that if a sample is in contact with an atmosphere of constant relative humidity, then maintenance of a uniform temperature is sufficient for a defined equilibration, whereas for a sample in contact with an atmosphere of maintained vapor pressure it is necessary to establish a precisely known temperature.

2 GRAVIMETRIC WATER CONTENT

Gravimetric water content describes a measure of amount of water in terms of the weight of water in a given weight of sample. It may be determined directly or indirectly. Care has to be taken to prevent loss or gain of moisture once a sample has been taken. Both the Association of Official Analytical Chemists (AOAC) and the American Association of Cereal Chemists (AACC) publish recommended methods for moisture determination.[8,9] Bradley[10] discusses methods for the analysis of moisture and total solids in some detail.

2.1 Methods of Sample Handling

The general procedures for sampling and sample handling prior to gravimetric estimation of water content are similar to those for most analytical protocols. Sampling procedures are required that give a statistically valid sample. After the sample is taken, it must be protected from contamination and from change. Also, the sampling procedure itself, namely any methodology to mix, subdivide or reduce particle size, must have minimal impact upon the sample composition. General sampling procedures are discussed in most textbooks of food analysis, e.g. that of Pomeranz and Meloan.[11] As indicated above, special precautions are required to minimize inadvertent loss or gain of moisture during handling. Exposure to an open atmosphere must be minimized. The rate of moisture exchange is related to the relative humidity of the atmosphere. Only exposure to controlled atmospheres for the purposes of controlled moisture exchange in a well-developed protocol is acceptable. Excessive heating during grinding must be minimized because this increases the risk of moisture loss. The headspace of sample storage containers must also be minimal, because moisture will equilibrate between the sample and the headspace. The preferred storage container should be airtight, with a hermetic seal, to minimize the possibility of moisture exchange with the environment.

2.2 Oven Drying Methods

In order to determine moisture content, a sample of known initial weight may be heated in specified conditions and the loss of weight determined. The results are dependent upon the type of oven used, the time and temperature of drying, and the characteristics of the product. Such methods are simple and direct, and can allow for simultaneous analysis of a large number of samples. Care is required to ensure that the samples cannot pick up significant moisture between drying and weighing. The rate and extent of moisture removal increase as temperature increases. Pure water boils at 100 °C at sea level. Dissolved solute raises the boiling point, so that as moisture is removed the boiling point increases. This makes it more difficult to evaporate additional moisture under constant temperature conditions. The level of drying that can be achieved is then dependent upon the drying temperature. At higher temperatures, more moisture can be removed. However, as the drying temperature increases, loss of other volatile components also increases, somewhat negating the advantage in terms of moisture removal. Also, at higher temperatures the extent of thermal degradation increases. Weight loss due to thermal degradation and loss of volatile materials is a source of error that should be minimized, so the choice of drying temperature is dependent upon sample type. Frequently 130 °C is recommended for air ovens and 100 °C for vacuum ovens. For more sensitive materials, such as fruits, the corresponding temperatures are 100 °C and 70 °C. The recommended temperature for drying will often be recommended in a specified procedure for the commodity. Because the boiling point of water is pressure dependent, reducing the environmental pressure by using a vacuum oven makes removal of moisture easier. Volatile substances are also removed more easily, but the extent of thermal degradation is reduced. Drying ovens are available from a wide range of manufacturers, and frequently come with detailed instructions for their use and clear indications of their suitability for different product types. All drying ovens must be protected from moisture ingress, hence any air streams entering the oven must be passed through an air dryer.

2.2.1 Convection/Conduction Ovens

As a consequence of the different characteristics of different materials, the oven drying methods recommended by, for example, the AOAC and the AACC specify both drying temperatures and times. These require careful control. The uniformity of temperature in the oven must also be considered. The three most common oven types are convection ovens, forced draft ovens, and vacuum ovens. The convection oven exhibits the greatest internal temperature variations and a 10 °C variation is not uncommon. This influences the attainable accuracy, because the location of a sample in the oven will influence drying rate and drying efficiency. This type of oven is not suited to precise measurement. Forced draft ovens maintain temperature differentials of less than 1 °C. These are suited to much higher precision of measurement of moisture loss. Vacuum ovens are intermediate in temperature uniformity but have the advantage of allowing more water removal at much lower temperatures, thereby minimizing thermal degradation. Heat transfer in a vacuum oven cannot be through convection or forced draft of an atmosphere, because the atmosphere density is much reduced. The heat transfer must be through the structural framework. This must be taken into account when choosing materials for the sample containers.

The design of the pan that holds the sample influences the ease and precision of measurement. Pans may be reusable or disposable. The costs of cleaning reusable pans can be significant. Covers should be utilized, because their use helps to prevent loss of sample through spattering. However, it is important that the cover design and positioning should not hinder evaporation. For most work, it is recommended that disposable aluminum pans be used, with fiberglass covers. A metal pan enhances thermal conduction to the sample, which is especially important in a vacuum oven. Fiberglass is preferred over metal for the cover, because metal covers can hinder evaporation. A fiberglass cover allows for evaporation, while preventing loss of solids content through spattering. If metal covers are used, their position must be adjusted prior to drying to allow for moisture removal. Both pans and covers should be thoroughly predried before use, and should be handled with tongs to prevent inadvertent contamination. Samples should be transferred rapidly into the pans and weighed immediately. After drying, pans must be handled in a very dry atmosphere prior to reweighing, because dried samples will rapidly pick up moisture. The appearance of the sample should be noted. If there is evidence of crusting or lumping, which leads to erratic drying, it may be necessary to use the sand pan technique, where the sample is mixed with a preweighed amount of dried sand in the pan to help prevent clumping and crusting and lead to more uniform and complete drying.

Drying may be performed either for a preset time or until two successive weighings of the drying sample show no significant weight change. For convenience, most standard methods define a time that will have been determined by experiment to be sufficient for effective drying. As indicated, both the AOAC and AACC methods give precise details as to the protocols for different types of materials and also for different types of oven. For vacuum ovens, in addition to specifying

temperature and pressure, a controlled rate of dry air purge is required in defined methods.

2.2.2 Infrared Drying

The use of radiant (infrared) heating can shorten drying times compared to convection or conduction heating in the conventional drying oven. However, it results in significant temperature gradients and a greater probability of thermal degradation. Surface temperatures can reach over 200 °C. This method is most suited to rapid qualitative process monitoring. If used in a vacuum oven system, infrared heating enables more rapid evaporation. The lower pressures and hence lower temperatures also help to reduce the amount of thermal degradation compared with conduction drying. The method should be confirmed by use of calibration standards on a regular basis. Equipment is available from many sources.

2.2.3 Microwave Drying

An alternative method to conductive, convective or radiant heating for energy transfer to the sample employs microwave energy. Specially designed microwave oven systems are widely available for moisture content assays. These have the advantage of ease of use and often ease of automation. The input power can be controlled readily but the energy pattern is very much dependent upon oven design. The sample, too, can influence the heating pattern. Standard methods exist for different materials that specify the particular oven to be used. The usual procedure is for the internal balance to be tared with fiberglass pads, the sample to be placed rapidly between the pads, and power applied to dry the sample. Sample weighing is continuous. Care is needed in sample placement to ensure uniform heating and to prevent localized charring. The method is suited for high throughput but does not lead to high precision. Each design of microwave oven has a unique heating profile, and, as a consequence, standard methods (e.g. AOAC Method 977.11 for cheese) specify the model of oven employed. Calibration against a vacuum drying procedure is required (e.g. AOAC Method 926.08 for cheese).

As already indicated, commercial drying ovens from a range of manufacturers are available for all of the above methods. These ovens come with detailed instructions for use. Most recently, microwave ovens and radiant heating ovens seem to have captured the largest market share due to their ease of use and rapid sample throughput, but the need to confirm their reliability by use of standardization procedures must be borne in mind.

2.3 Distillation Methods

Distillation methods also fairly directly determine water content and are defined as standard methods. For example

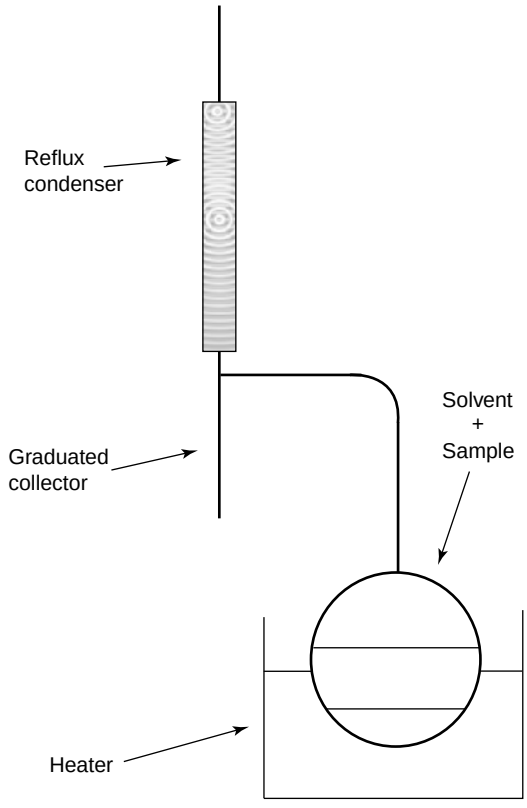

Figure 3 Schematic diagram of a distillation assembly.

AOAC Method 969.19 defines a distillation method to determine the moisture content of cheese. These methods involve the co-distillation of water in a food sample with a high-boiling solvent immiscible with water. Because the water that co-distills is immiscible with the solvent, its volume can be determined easily. Direct and reflux distillation methods may be used. In direct distillation, the sample is heated in a high-boiling solvent at a temperature well above the boiling point of water, and the resulting vapors are condensed and collected. In reflux methods the solvent is re-circulated so that extraction can be more efficient. Lower temperatures can be effective. Special traps have been designed to prevent contamination and moisture pick-up. A typical distillation assembly is shown schematically in Figure 3. There are several sources of error associated with this type of method.

2.3.1 Errors Associated with Distillation Methods

Because these methods are predicated upon immiscibility of water and solvent, any event that prevents the water and the solvent from separating can lead to errors in estimation of water volume. Typical problems include water droplets sticking to surfaces in the apparatus and failing to merge with the main volume of water. Also, the formation of emulsions leads to entrainment of water

in the solvent, unless the emulsions can somehow be broken. Another source of error is chemical reactions that produce water through degradation of components of the sample. These can be minimized by reducing the temperatures required for co-distillation.

2.4 Chemical Methods

A potential disadvantage of drying or distillation methods is that the water content is not determined directly as such, but rather it is inferred from a weight change or from a volumetric measurement. Loss of volatile substances and degradation can lead to weight change. Soluble contaminants and partitioning of partially soluble components between phases can influence volume change. As an alternative to these methods, chemical methods are available to determine the amount of water by evaluating the extent of a chemical reaction. As such, the sources of error and the limitations are different. The most widely applicable chemical method is the Karl Fischer titration.

2.4.1 Karl Fischer Titration

This method does not require sample heating and so does not lead to thermal degradation. It is particularly effective for low-moisture foods. It is based on the reduction of iodine by sulfur dioxide in the presence of water, according to Equation (1):

$$2H_2O + SO_2 + I_2 \longrightarrow H_2SO_4 + 2HI \qquad (1)$$

In order to dissolve the iodine and sulfur dioxide, methanol and pyridine are used as solvents. The resultant reaction scheme is shown by Equations (2) and (3):

$$C_5H_5N \cdot I_2 + C_5H_5NSO_2 + C_5H_5N + H_2O \longrightarrow 2C_5H_5N \cdot HI + C_5H_5N \cdot SO_3 \qquad (2)$$

$$C_5H_5N \cdot SO_3 + CH_3OH \longrightarrow C_5H_5N(H)SO_4 \cdot CH_3 \qquad (3)$$

For each mole of water, 1 mol of iodine, 1 mol of sulfur dioxide, 3 mol of pyridine and 1 mol of methanol are used. Two detection methods can be used. The volumetric titration adds the reagents to the sample in a closed container, isolated from atmospheric moisture. Excess, unreacted iodine can be detected visually or potentiometrically. This procedure is effective for samples with moisture content in excess of 0.03%. At lower moisture levels, a coulometric titration is preferred. Here, the reagent iodine is produced electrolytically, to titrate the water. The amount of iodine required to titrate the water is determined from the current needed to generate it.

The volumetric titration may be direct, adding the Karl Fischer reagent as a titrant directly to the sample, or it may be indirect, extracting water from the sample into a solvent such as methanol and then titrating the methanol extract.

Standardization of the reagent is necessary before the amount of water can be determined. Because the reagent is not stable, this must be performed before each use. Standardization requires a standard of known water content. This can be pure water, a solution of water in methanol, or sodium tartrate dihydrate (or some other characterizable hydrate). Pure water is not an easy standard to employ, because of the difficulty in precisely measuring small amounts. A standard solution of water in methanol can slowly change with time, due to absorbing atmospheric moisture. Sodium tartrate dihydrate is quite stable and is the standard of choice.

2.4.1.1 Sources of Error in the Karl Fischer Titration Although the Karl Fischer titration is an effective method for the measurement of water content, there are several sources of error that must be taken into account. Incomplete extraction of water from the sample can lead to error. To enhance water extraction, the fineness of grinding can be important. Other errors involve contamination by water from sources other than the sample. Because the method is a very sensitive measure of moisture, external air must be prevented from access to the reaction chamber, or else atmospheric moisture will be absorbed. All glassware and tools must be dried rigorously before use in order to prevent inadvertent addition of moisture. In addition to ensuring carefully that all the water in the sample and only the water in the sample is determined, it must be realized that substances other than water may react with the assay reagent. Thus, certain interferences must be recognized. The most important of these are from ascorbic acid, which is oxidized by Karl Fischer reagent, from carbonyl compounds, which react with methanol to form acetals and water, and from unsaturated fatty acids, which react with iodine.

Given the unstable nature of the reagent, the difficulties that are experienced in its preparation, and the critical requirement for preventing access to atmospheric moisture, it is best to employ one of the many systems that are available from instrument manufacturers. Although the principles of operation are common, the details of operation differ from one supplier to another.

2.4.2 Alternative Chemical Assays for Water

Although there are other chemical methods for determining water, such as the use of calcium carbide, which releases acetylene on reacting with water through the reaction depicted by Equation (4):

$$CaC_2 + 2H_2O \longrightarrow Ca(OH)_2 + C_2H_2 \qquad (4)$$

these are of limited application and difficult to use with any precision. The Karl Fischer reaction is the only chemical method with wide application.

2.5 Indirect Methods of Water Content Determination

So far, only fairly direct methods for estimating the gravimetric water content have been described. These all tend to be demanding of technique, and are slow. Indirect methods exist for estimating water content. These rely on careful calibration for particular materials against the direct methods described above. Common indirect methods require the measurement of some easily determined property of the food material and then application of a calibration chart to estimate the appropriate water content. For an indirect method to be successful, the relationship between water content and the property must be repeatable, and the change in property with water content should be such that the method is sensitive to small changes in water content.

2.5.1 Infrared Spectroscopy

Water can be determined by measuring absorption at infrared wavelengths characteristic of the absorption frequencies of water. Typical wavelengths are 3.0 and 6.1 μm (fundamental vibrational mode), 1.93 μm (combination) and 1.45 μm (first overtone OH stretch). Samples for evaluation must be thin, for transmission measurements. An alternative to using a thin sample is to disperse a finely divided sample in a suitable carrier fluid before performing a transmission measurement. Reflectance measurements characterize the surface region more than the entire sample. Generally, infrared absorption is calibrated by using samples of known moisture content, and wavelengths for measurement are chosen to minimize interferences. Fourier transform spectrometers collect data for all wavelengths by interferometry, followed by mathematical transform. This is a more rapid procedure than the wavelength scanning of a dispersive instrument. It is necessary to correlate the infrared spectrum with water content data obtained by more direct measurement techniques. Modern computers make this a simple task, but care should be taken to confirm that the wavelengths identified by the correlation routine are significant in terms of the spectroscopic absorbance of the sample.[12]

2.5.1.1 Near-infrared Spectroscopy Near-infrared reflectance measurements can be applied to the estimation of water content.[12] Once again, this is a method in which data sets are correlated for samples of differing water contents, and the resulting correlation equations are then used to estimate the water content of suitable samples. Although early instruments utilized only two wavelengths, this was unduly influenced by factors such as particle size. More recent work uses at least three wavelengths: two reference wavelengths and a measuring wavelength. For a measurement to be reliable, it is important that the sample is equilibrated. For many samples, the extent of absorption is a linear function of moisture content. Further improvement in accuracy comes from measuring near-infrared reflectance at a wider range of wavelengths. Calibrations employ mathematical analysis of the entire data set by multi-variate statistics. As computer power continues to increase, this method becomes more and more effective and is of increasing popularity. It is essential that it be calibrated against a standard gravimetric procedure before being applied as routine.

2.5.2 Electrical Properties

The electrical properties of many systems vary in a repeatable fashion with water content, making such properties suitable as indirect probes of water content.

2.5.2.1 Electrical Conductivity Measurement of electrical conductivity is simple and rapid. Samples are prepared with specific geometries between measuring electrodes. This can be through pressing a standard mass of sample into a cup with preconfigured electrodes. Alternatively, it can be the sectioning of a sample to fit between electrodes, or the insertion of an electrode jig into a sample of controlled density. In all cases the conductivity or resistance is measured using standard electrical methods. The moisture content is obtained from a calibration curve generated from samples of the materials that have had their moisture content determined by conventional measurement. The AOAC methods that exist for dried fruits utilize electrical conductivity.

2.5.2.2 Electrical Capacitance Sample preparation for electrical capacitance measurement is very similar to that for electrical resistance measurement. The methods differ in that the measurement performed is of the capacitance between the electrodes. These measurements can be made at a range of frequencies. Measurement of capacitance is based on the fact that the dielectric permittivity of water is much higher than that of most other fluids or solvents. The variation in capacitance with frequency can lead to information about the environments experienced by water in the sample. Once again, for estimation of moisture content, calibration curves must be established for the particular material under investigation.

For both electrical methods, it must be realized that the solid composition of the sample will influence the calibration curve. In particular, the ratios of salts to nonelectrolytes will have a significant influence on the correlations between water content and electrical

conductivity and electrical capacitance. The value of these methods is very much linked to the effectiveness of the calibration procedures that are followed.

2.5.3 Estimation of Soluble Solids

Certain properties of aqueous solutions are dependent upon the solute concentration. In appropriate cases it is possible to estimate the moisture content of a food by estimating the concentration of soluble solids. This requires that the ratio of the content of soluble to insoluble solids is known or can be estimated. Given the concentration of soluble solids, the amount of water can be calculated by difference. Easily measured properties that can be used to estimate the concentration of soluble solids include refractive index, solution density and freezing point. Refractive indices are often utilized to estimate the content of soluble solids in many fruits, fruit products and sugar solutions. The use of a simple refractometer, such as an Abbe refractometer, allows for the determination of the refractive index of the solution. Rather than reading the refractive index, the instrument scale is calibrated to read the Brix of a sugar solution that has this refractive index. A range of AOAC methods exist, e.g. AOAC Method 970.59, which utilize refractometry to determine soluble solids and hence, by inference, the moisture content. Measurements of density, utilizing pycnometers, hydrometers or specific gravity balances, are often used for milk products, beverages, pickling brines and sugar solutions. Once again, the density correlates directly to the solute concentration of the solution under study, as long as the principal solutes are known in advance. Freezing point measurements are often used to check for adulteration of milk.

2.6 Thermogravimetric Analysis

One further gravimetric method should be mentioned: the method of thermogravimetric analysis.[13] A small sample attached to a weighing sensor is heated through a preset temperature program, and the changes in weight are followed. Loss of moisture and other volatile materials is readily seen, as is thermal degradation of the sample. A typical thermogravimetric analyzer scan is shown in Figure 4. This technique can help to identify the appropriate range of temperatures for drying procedures. The temperature at which thermal degradation sets in can often be seen clearly on the thermogravimetric trace.

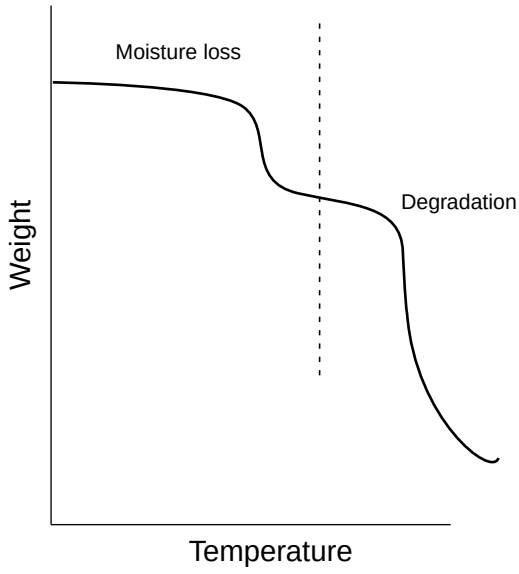

Figure 4 A typical thermogravimetric analysis curve. The first zone represents the loss of sorption moisture, up to around 110 °C. Weight loss in the second zone reflects loss of water of hydration and also chemical degradation.

2.7 Solvent Extraction

In describing the Karl Fischer method, passing reference was made to using an extraction solvent for removal of water. This provides yet another route for the determination of water content. Provided that extraction of water by the solvent is quantitative and complete, then the task becomes determination of the water content of the extraction solvent. With a suitable solvent, this can be achieved through the use of gas chromatography. This requires a liquid or gaseous sample, which can be introduced into the column. Determination of water will be effective only if extraction into the solvent is quantitative and if there are no substances extracted that will provide interference. After extraction, the gas chromatography system is used to determine the moisture content of the extraction solvent, from which the moisture content of the original food can be determined. A typical protocol is to take 15 g of sample and blend with 100 mL of solvent (e.g. 95% absolute methanol and 5% secondary butanol). After the mixture has settled for 20 s, a 2-μL aliquot of the clear supernatant is analyzed using a gas chromatograph. Calibration samples are solvent with known amounts of water added. The method is applicable to a range of foods with moisture contents varying from 8 to 60%.

The advantage of indirect methods, and many correlation procedures, is that, unlike the direct determination of moisture content, the indirect method may be rapid and simple, allowing for the monitoring of a large number of samples. The methods are also often conducive to automation. However, the value of any procedure is dependent on the robustness of the correlation of the property determined to the moisture content, and also on the quality of the original data used to establish the correlation equations. Care must be taken to establish a calibration that is reliable.

3 WATER ACTIVITY (OR EQUILIBRIUM WATER VAPOR PRESSURE)

Although gravimetric water content is an important parameter, a thermodynamic measure of ''amount of water'' is more appropriate for describing the potential of water to participate in chemical and physical processes. Activity – a thermodynamic measure of ''quantity of substance'' – was defined by Lewis and Randall.[14] ''Activity is, at a given temperature, the ratio of the fugacity, f, of a substance in some given state and its fugacity, f^{o}, in some state which for convenience has been chosen as a standard state''. When a vapor behaves in a manner close to ideality, the fugacity is closely approximated by the partial vapor pressure. A common standard state is the pure liquid. For most practical purposes, water activity is defined as the ratio of the equilibrium partial vapor pressure of water above the sample to the equilibrium vapor pressure above pure water *at the same temperature*. The equilibrium partial vapor pressure of water over the sample, p_w, may be measured directly or determined indirectly. The saturation partial vapor pressure of water at the same temperature, p_w^o, can be obtained from tables if the temperature is known. Great care must be taken to ensure that equilibrium conditions have been established within the sample, and between the sample and its atmosphere, because a measure of a transient partial vapor pressure is of little value. True equilibrium is seldom seen in foods, but a measurement made at steady-state conditions can convey important information. It is important either that the whole measuring system be at a uniform temperature or that the temperature of the sample be precisely known. In some measuring systems, the sensor that determines the vapor pressure may not be at the same temperature as the sample. In this situation the temperatures of the sample environment and the sensor environment must be known precisely. Leung[15] and Chirife[16] provide useful discussions of the applications of relative vapor pressure/water activity information. Levine and Slade[17] organized a meeting at which many of the relevant issues have been discussed.

3.1 Vapor Pressure Manometer

This entails the direct measurement of the partial vapor pressure above a sample of known water content. Temperature control of the sample is critically important, because the saturation vapor pressure of water and the equilibrium partial pressure above an aqueous solution change by about 5% for every 1° rise in temperature around 25 °C (see Table 1). A typical vapor pressure manometer system is illustrated in Figure 5. Although a vapor pressure manometer is the most direct method of determination of partial vapor pressure, it is a very demanding technique if good accuracy is to be achieved. Given the dimensions of the system, temperature control is difficult. The presence of other volatiles leads to an erroneous estimate. Exclusion of air is also difficult.

3.2 Dew Point Determination

As with direct measurement of the vapor pressure, this method entails establishing a constant vapor pressure above the sample, with careful temperature control of the sample and its immediate environment. In this case, because the determination of the partial pressure is by

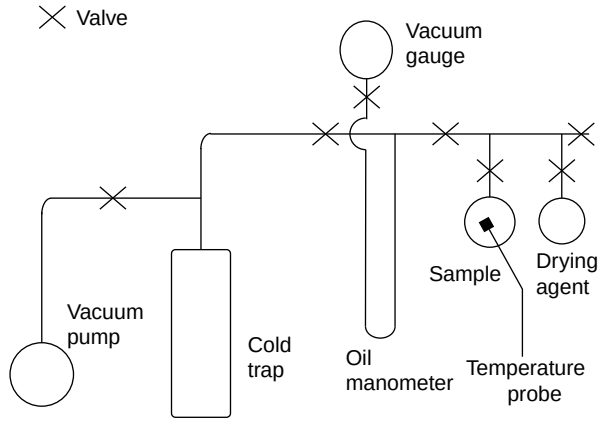

Figure 5 Schematic diagram of a vapor pressure manometer. Prior to a measurement, the system must be evacuated and dried. A complex arrangement of valves is necessary to protect the sample while this process is in hand.

Table 1 Temperature dependence of the saturation vapor pressure of water (p_w^o) and water activity (a_w) for selected sulfuric acid concentrations. (Adapted from Rizvi.[21])

Temperature	p_w^o	a_w for H_2SO_4 concentration (%)						
		5	10	15	20	25	30	35
0	4.58							
10	9.19	0.9804	0.9555	0.9230	0.8779	0.8183	0.7429	0.6514
20	17.53	0.9806	0.9558	0.9237	0.8796	0.8218	0.7491	0.6607
30	31.82	0.9808	0.9562	0.9245	0.8814	0.8252	0.7549	0.6693
40	55.33	0.9811	0.9565	0.9253	0.8831	0.8285	0.7604	0.6773
50	97.48	0.9814	0.9570	0.9261	0.8848	0.9317	0.7655	0.6846

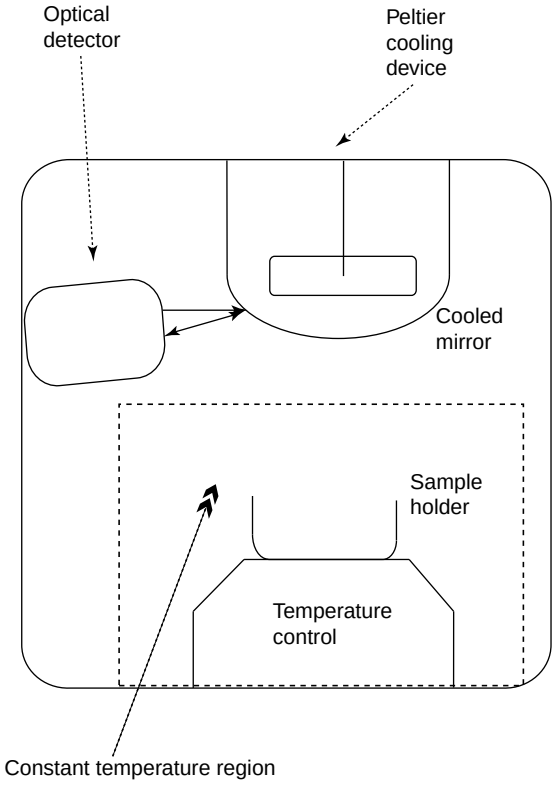

Figure 6 Schematic diagram of a dew point cell. The sample is placed in a thermostatically controlled holder. The partial vapor pressure is equal to the saturation pressure at the dew point, measured by the initiation of condensation on the cooled mirror. Condensation is detected by an optical sensor, and the temperature of the mirror, cooled by a Peltier device, is recorded. The vapor spaces are kept to the minimum necessary. Except for the sensor, other electrical devices for measuring a_w are of similar layout, as described in the text.

a linked property rather than direct, the influence of air and other vapors on the measurement may be less critical. The partial pressure of water is deduced in the following manner. Inserted in the vapor space is a probe whose temperature can be controlled independently. This probe is sufficiently far removed from the sample so as not to influence its temperature. The temperature at which dew forms on the probe is determined. Dew formation is detected optically (e.g. through the change in reflectivity of a mirrored surface). This temperature of the probe when initial dew formation is detected is the dew point of the atmosphere above the sample. At the dew point, the partial vapor pressure of water is the saturation pressure. Thus, if the dew point temperature is known, the saturation water vapor pressure at this temperature can be obtained from tables. This is equal to the partial pressure established above the sample. Because the saturation pressure is very temperature dependent, the temperature of the dew point probe must be measured precisely. The design of a dew point system must be such that the dew point probe does not thermally influence the immediate environment of the sample, because if the sample temperature changes then the partial vapor pressure changes. A typical configuration is illustrated schematically in Figure 6. Note that two features are particularly important. First, the vapor space must be small enough so that the loss of moisture from the sample in establishing the atmosphere is insignificant. Second, the temperatures in the sample region must be uniform and known. The dew point method has become quite popular. Peltier cooling of the mirror allows for sensitive control of the mirror temperature, and several devices are available that operate on this principle.

Determination of the dew point temperature by use of a cooled mirror surface is a fairly direct method for determination of the partial vapor pressure of the water. Alternative methods to estimate partial water vapor pressure exist. In the schematic of Figure 6 the dew point sensor could be replaced by alternative sensors that respond in some predictable fashion to the water vapor pressure of their surroundings. The generic cell design can remain the same as illustrated in Figure 6, except that now the region of the sensor also must be maintained at a uniform, known, temperature rather than be adjusted to the dew point. Recalling the discussion of sorption isotherms, and also the indirect methods for arriving at moisture content, it is clear that the physical characteristics of many systems are dependent on moisture content and the associated water activity. If a material has a repeatable sorption isotherm, then establishing the dependence of some property on moisture content/water activity allows for its use as a vapor pressure sensor. If the temperature of the sensor is known and sufficient time is allowed for the sensor to equilibrate with the atmosphere around it, then measurement of the property in question allows for estimation of the water activity a_w of the sensor and hence the p_w of the atmosphere adjacent to the sensor. In a well-designed system, this p_w will also be the partial vapor pressure of water close to the sample. Given that the sample is at uniform temperature and that equilibration has been achieved, this allows for estimation of the equilibrium partial vapor pressure of water above the sample at the temperature of the sample. Because the sample temperature is known, p_w^o can be obtained from tables. If the whole system is at a uniform temperature, then a_w of the sensor is a_w of the sample. It is not necessary to establish a uniform temperature throughout the chamber: it is only necessary to establish uniform, known, temperatures in the region of the sample and in the region of the sensor, provided that the sample is the only significant source of water vapor.

3.3 Water Activity Sensors

Further to the preceding discussion, by ''water activity sensors'' we mean those sensors that, when equilibrated with an atmosphere of known equilibrium relative humidity (or water activity), establish a stable water content. Stated differently, these are materials with highly repeatable sorption isotherms. Measurement of a property that depends upon water content then allows for identification of the equilibrium relative humidity of the atmosphere in contact with the sensor and, because the temperature at the sensor is known, the actual vapor pressure of the atmosphere.

3.3.1 Electric Hygrometers

As indicated earlier, the electrical properties of materials are dependent upon their water contents. This can be utilized to provide sensors that change in properties according to the relative humidity of the atmosphere to which they are exposed. For an effective sensor, the sorption isotherm (water content versus water activity) will be reproducible, and the dependence of measured property to water content will be a monotonous function. It has been found that electrical conductance and electrical capacitance for many substances are monotonous functions of water content at a defined temperature. Because these are measured readily, they provide the basis for a range of sensors. Many commercial designs are on the market, all based on the same principles.

3.3.2 Conductance

Sensors based on this principle measure the conductance of an immobilized film containing a salt such as lithium chloride. The sensor equilibrates, at a known temperature, with the atmosphere above the sample. The relative humidity of this atmosphere determines the moisture content and hence the conductivity of the cell. Using standard samples, which establish atmospheres of known relative humidity, a calibration curve of sensor conductance against relative humidity is established. Several commercial instruments utilize this type of sensor. Early designs were prone to contamination, but modern sensors seem to have good reliability.

3.3.3 Capacitance

The electrical capacitance of a film such as aluminum oxide varies with moisture content. Capacitance is measured readily. A sensor based on aluminum oxide can therefore be calibrated to allow for estimation of the relative humidity of any atmosphere in which it is equilibrated. Once again, there are several commercial instruments that utilize this type of sensor. In some capacitance sensors, a polymer film is utilized rather than an aluminum oxide film. These sensors have proved quite reliable in use.

It should be noted that the sensors described in section 3.3 should all be checked on a regular basis. Solutions of known equilibrium relative humidity should be used to test the operation. Table 2 lists some appropriate saturated solutions. Another list is given as AOAC Table 978.18. To achieve equilibrium relative humidity values other than those of saturated solutions, solutions of measured concentration can be used. Care must be taken to prevent significant change in concentration, because equilibrium relative humidity is sensitive to concentration. Note that, unlike the situation for saturated solutions, for sodium chloride solutions of fixed concentration the water activity is not very dependent upon temperature in the range around 25 °C (see Table 3). As long as sample temperature is known precisely, this makes sodium chloride solution a good calibration standard.

Table 2 Water activity above saturated salt slurries. (From Greenspan.[22])

Salt	a_w		
	20 °C	25 °C	30 °C
LiBr	0.0661	0.0637	0.0616
LiCl	0.1131	0.1130	0.1128
CH_3COOK	0.2311	0.2251	0.2161
$MgCl_2$	0.3307	0.3278	0.3244
K_2CO_3	0.4316	0.4316	0.4317
NaBr	0.5914	0.5757	0.5603
KI	0.6990	0.6886	0.6789
NaCl	0.7547	0.7529	0.7509
$(NH_4)_2SO_4$	0.8134	0.8099	0.8063
KCl	0.8511	0.8434	0.8362
KNO_3	0.9462	0.9358	0.9231

Table 3 Water activities of selected NaCl concentrations, valid for range 20–40 °C. (Adapted from Chirife and Resnik.[23])

a_w	NaCl concentration (wt. fraction)
0.98	0.035
0.96	0.065
0.94	0.093
0.92	0.118
0.90	0.142
0.88	0.163
0.86	0.182
0.84	0.200
0.82	0.217
0.80	0.232
0.78	0.246
0.76	0.260

3.4 Measuring Related Thermodynamic Quantities

In addition to measurement (directly or indirectly) of the vapor pressure established in the atmosphere above a sample, other thermodynamic properties that are related to the water activity can be determined. Properties that are influenced by concentration are often termed colligative properties. Two common colligative properties that are related to water activity are osmotic pressure and freezing point depression. Measurement of osmotic pressure is impractical for many food systems, so in general only freezing point depression has been used to estimate water activity.

3.4.1 Freezing Point Depression

The freezing point depression of water, ΔT_f, is related to water activity through the expression shown in Equation (5):

$$-\log a_w = 0.004207\Delta T_f + 2.1 \times 10^{-6}\Delta T_f^2 \qquad (5)$$

Determination of freezing point depression requires a sensitive thermometer and also requires that steps be taken to eliminate supercooling. This method is used most frequently to check for adulteration of milk.

3.5 Isopiestic Methods

The methods of determining water activity described so far are suitable for application to samples of defined gravimetric water content. In all cases, such samples are allowed to equilibrate only within the measuring cell. There must be little loss of water from the sample to the measuring cell or from the measuring cell to its environment for valid measurements.

An alternative approach, often used in the generation of sorption isotherms, is to establish in some suitable apparatus, or section of apparatus, an atmosphere with a controlled, known partial pressure of water at a known temperature and to equilibrate a sample with it. After equilibration, the gravimetric water content of the sample is determined by weighing. Methods of this type are termed isopiestic methods. A typical apparatus for isopiestic equilibration is shown in Figure 7. There are alternative strategies for the control of the partial pressure of water vapor within the isopiestic system.

3.5.1 Solution Equilibration

At constant temperature, solutions of known composition establish above them equilibrium vapor pressures that can be obtained from tables. This vapor pressure is, as indicated earlier (Table 1), dependent upon temperature. The equilibration solutions can be of fixed concentration, establishing an atmosphere of stable equilibrium relative humidity at constant temperature. Provided that uniform temperature is maintained throughout the system, the equilibrium relative humidity (equilibrium relative vapor pressure) does not change much for the small variations in the constant temperature that is maintained. Hence, although equilibration requires elimination of temperature gradients and precise control of temperature, this need not be accurate at the target temperature for valid measurement. It is necessary to have a sufficient volume of the equilibration solution to ensure that the concentration does not change significantly as moisture is taken up by the samples under equilibration. An alternative approach, which maintains constant concentration regardless of water vapor transfer, is to use saturated solutions for equilibration. Here, however, temperature must be known accurately as well as be maintained precisely constant. Not only is the equilibrium vapor pressure above the saturated solution temperature dependent, but now also the equilibrium relative vapor pressure is temperature dependent (Table 2), because saturation solubility is temperature dependent. The advantage of this approach, which may justify dealing with the need for more stringent temperature control, is that as long as excess solute is present, the vapor pressure will remain constant, unaffected by the uptake of water by the samples. The

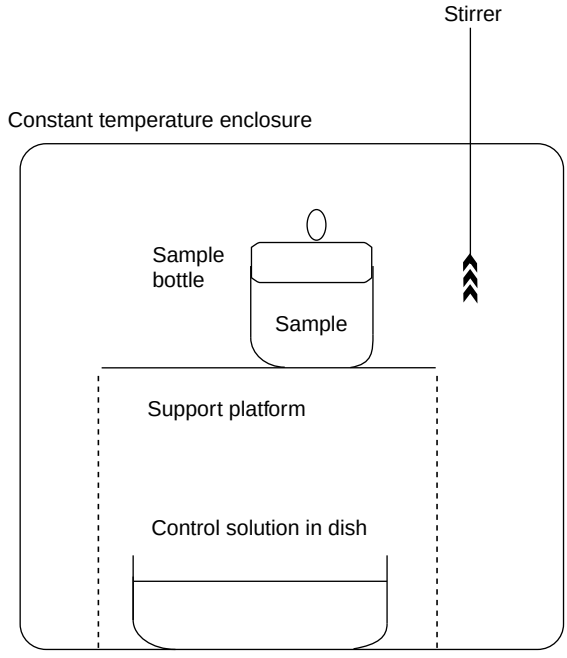

Figure 7 Schematic diagram of isopiestic equilibration by solution system. The atmosphere, above a suitable solution (see Tables 1–3) is maintained at constant temperature throughout the enclosure. It may be evacuated prior to equilibration. The sample is exposed to the atmosphere for a sufficient time to reach a steady state. The sample container is readily sealed or opened.

challenge in isopiestic measurements is first to equilibrate the samples as rapidly as possible and then to determine the sample weight accurately. Because large numbers of samples may be equilibrated together, some design of sample containers is required that allows for easy access of the equilibration atmosphere, while at the same time allowing for the sealing off of the sample once equilibration is complete, to allow for transfer to the weighing system without too much moisture exchange.

3.5.2 Bithermal Equilibration

An alternative approach to the use of equilibration solutions, which also allows for maintenance of an atmosphere of controlled relative humidity in the equilibration zone, employs the same principle as is used in the dew point method for determining vapor pressure. In this case, the temperature of a reservoir containing pure water is controlled. This establishes a vapor pressure equal to the saturation vapor pressure at that temperature. If a vapor path is provided to an equilibration cell at some higher temperature, the saturation pressure at T1 is established as the pressure at T2. Hence, the relative humidity at T2 is maintained at $(p_w^o)_{T1}/(p_w^o)_{T2}$.

This type of approach is particularly suited to the establishment of sorption isotherms, because the sample can be maintained at constant temperature, attached to some form of balance, and sample weight (and hence water content) can be monitored continuously. By stepwise increase of the temperature of the water reservoir, the water content at increasing relative humidity can be followed. If the reservoir temperature is lowered in steps, the isotherm during desorption can be obtained. Care must be taken to allow for equilibration before moving to the next step. The two chambers must also be able to maintain constant, uniform temperatures within the volume of the chamber. A typical configuration is shown in Figure 8.

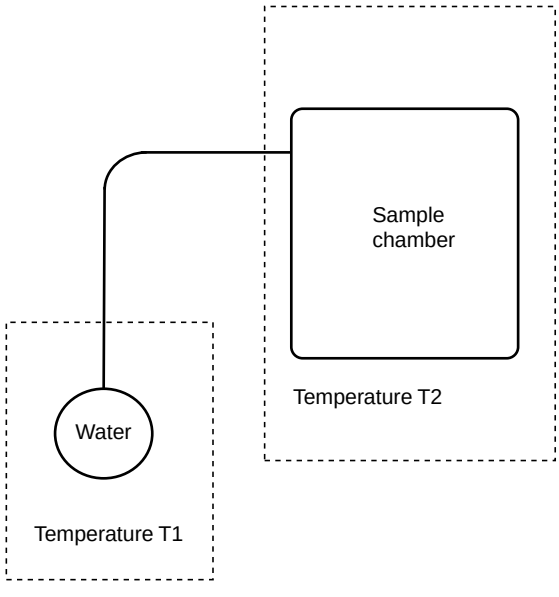

Figure 8 Schematic diagram of isopiestic equilibration by bithermal system. The partial pressure in the sample chamber (which is maintained at temperature T2) is established by the saturation vapor pressure of the water in the reservoir maintained at temperature T1. Sorption is therefore taking place in an atmosphere of relative humidity $(p_w^o)_{T1}/(p_w^o)_{T2}$.

3.6 Sorption Isotherms

A variety of systems exist for the measurement of sorption isotherms. Spiess and Wolf[17] describe a system recommended by the participants in COST90. This is the best-documented system available. Note, however, that unless the samples studied are very carefully standardized, it is difficult to replicate results from one center to another. The establishment of internal equilibrium is also difficult to achieve. Equipment has become available that allows for the rapid determination of water sorption and desorption on small samples. Sensitive balances can determine small changes in mass. For well-characterized materials, this can lead to valuable information, and indeed the rapid sorption systems are in widespread use in the pharmaceutical industry. In food systems, the ability to identify rapidly a sorption steady state can be useful, and helps to characterize the surface of the material, but the slow process of moisture equilibration in the bulk of a food, and the nonequilibrium nature of many foods, make it difficult to compare results between laboratories.

4 CATEGORIES OF WATER

Both gravimetric water content and water vapor pressure are single points of information. Today, often more detail is requested. Although for many years the measurement of water content, water activity, and their interrelation through moisture sorption isotherms has sufficed to describe water in foods, it is no longer sufficient. In past research, the shapes of sorption isotherms and the equations and models that attempt to fit the sorption isotherm data have been interpreted in terms of classes of water. Fitting of isotherms to equations such as the BET equation led to allocation of amounts of water to various categories. A variety of methods have evolved that estimate water in different environments. There is much discussion as to the applicability of these classifications. Although useful, they tend to obscure the reality that water in foods exists as a continuum, with a gradation of properties of the individual molecules that depends upon individual molecular environments and also reflects the ability of water molecules to exchange rapidly between environments. The many classification

models must therefore be seen as ''lumping'' models, or arbitrary classifications. Also, it has to be realized that the rapid exchange means that as water content changes, the environments and exchange patterns of all water molecules change. There are no individual water molecules that remain in a particular environment for an extended time, although there may always be water molecules in that environment.

4.1 "Bound" Water

The concept of bound water is rather imprecise. It suggests that water may exist as free (solution) water, or as water in some way restricted by its environment. Variously described, the estimates of ''bound'' water differ for the same system. Bound water has been described as the water that remains unfrozen at a specified temperature, determined by thermal analysis techniques. It can also be determined by spectroscopic techniques, the most common being nuclear magnetic resonance, where the liquid signal at subfreezing temperatures is interpreted as a measure of the unfrozen water. The topic is covered in great detail in the proceedings of a 1989 meeting discussing water in foods,[18] and in the publications of the ISOPOW symposia.[19,20]

4.2 Water Mobility

Spectroscopic measurements are influenced by the properties of the solvent water. A variety of models exist that take the time- or frequency-dependent signal from nuclear magnetic resonance, dielectric or other spectroscopic techniques, and derive estimates for the number of water molecules that exist in a variety of categories. The results of these studies may be published in great detail, and many useful correlations have emerged between classifications of water and system properties. It must be realized, however, that all of these results are model dependent, and that, often, different authors have interpreted essentially the same spectroscopic data in widely different ways. The spectroscopic measurement itself is an experimental fact. The interpretation of the data in terms of classes or fractions of water is a hypothesis and requires independent validation by other methods before it can be said to identify real situations.

ABBREVIATIONS AND ACRONYMS

AACC	American Association of Cereal Chemists
AOAC	Association of Official Analytical Chemists

RELATED ARTICLES

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Near-infrared Spectroscopy in Food Analysis • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General • Sample Preparation, Headspace Techniques

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction

REFERENCES

1. W.J. Scott, 'Water Relations of Food Spoilage Microorganisms', *Adv. Food Res.*, **7**, 83–127 (1957).
2. W.J. Scott, 'Water Relations of *Staphylococcus aureus* at 30 °C', *Aust. J. Biol. Sci.*, **6**, 549–564 (1953).
3. S. Gal, 'Recent Advances in the Determination of Sorption Isotherms', in *Water Relations of Foods*, ed. R.B. Duckworth, Academic Press, London, 139–150, 1975.
4. S. Gal, 'Practical Applications of Sorption Data', in *Physical Properties of Foods*, eds. R. Jowitt, F. Escher, B. Hallstrom, H.F.T. Meffert, W.E.L. Spiess, G. Vos, Applied Science Publishers, London, 12–15, 1983.
5. L.B. Rockland, L.R. Beuchat, *Water Activity, Theory and Application to Food*, Marcel Dekker, New York, 1987.
6. W. Wolf, W.E.L. Spiess, G. Jung, *Sorption Isotherms and Water Activity of Food Materials, a Bibliography*, Science & Technology Publishers, London, 1985.
7. O.R. Fennema, 'Water and Ice', in *Food Chemistry*, 3rd edition, ed. O.R. Fennema, Marcel Dekker, New York, 19–94, 1996.
8. Association of Official Analytical Chemists, *Official Methods of Analysis*, 16th edition, Association of Official Analytical Chemists, Gaithersburg, MD, 1994.
9. American Association of Cereal Chemists, *Approved Methods of Analysis*, 9th edition, American Association of Cereal Chemists, St Paul, MN, 1995.
10. R.L. Bradley, 'Moisture and Total Solids Analysis', in *Food Analysis*, 2nd edition, ed. S.S. Nielsen, Aspen Publishers, Gaithersburg, MD, 119–139, 1998.
11. Y. Pomeranz, C.E. Meloan, *Food Analysis, Theory and Practice*, 3rd edition, Chapman & Hall, New York, 1994.

12. I.B. Benson, 'Compositional Analysis Using Near Infrared Absorption Spectroscopy', in *Instrumentation and Sensors for the Food Industry*, ed. E. Kress-Rogers, Butterworth, London, 121–166, 1993.
13. V. Harwalkar, C.-Y. Ma, *Thermal Analysis of Foods*, Elsevier Applied Science, New York, 1990.
14. G.N. Lewis, M. Randall, *Thermodynamics*, 2nd edition, McGraw Hill, New York, 1961.
15. H. Leung, 'Water Activity and Other Colligative Properties of Foods', in *Physical and Chemical Properties of Foods*, ed. M.R. Okos, ASAE, Michigan, 138–185, 1986.
16. J. Chirife, 'An Update on Water Activity Measurements and Prediction', in *Food Preservation by Moisture Control*, eds. G.V. Barbosa-Canovas, J. Welti-Chanos, Technomic, Lancaster, PA, 169–190, 1995.
17. W.E.L. Spiess, W. Wolf, 'Critical Evaluation of Methods to Determine Moisture Sorption Isotherms', in *Water Activity, Theory and Application to Food*, Marcel Dekker, New York, 1987.
18. H. Levine, L. Slade, *Water Relationships in Foods*, Plenum Press, New York, 215–238, 1991.
19. D.S. Reid, *The Properties of Water in Foods: ISOPOW 6*, Blackie, New York, 1996.
20. Y.H. Roos, R.B. Leslie, P.J. Lillford, *Water Management in the Design and Distribution of Quality Foods: ISOPOW 7*, Technomic, Lancaster, PA, 1999.
21. S.S.H. Rizvi, 'Thermodynamic Properties of Foods in Dehydration', in *Engineering Properties of Foods*, eds. M.A. Rao, S.S.H. Rizvi, Marcel Dekker, New York, 133–214, 1986.
22. L. Greenspan, 'Humidity Fixed Points of Binary Saturated Aqueous Solutions', *J. Res. Nat. Bur. Stand. Sect.*, **81A**, 89–96 (1977).
23. J. Chirife, S.L. Resnik, 'Unsaturated Solutions of Sodium Chloride as Reference Sources of Water Activity at Various Temperatures', *J. Food Sci.*, **49**, 1486–1488 (1984).

Forensic Science: Introduction

T.A. Brettell
New Jersey State Police, West Trenton, USA

Analytical chemistry as applied to forensic science involves the analysis of drugs and poisons, forensic DNA analysis and the analysis of trace evidence.

Drugs and poisons include ethanol and volatiles, cannabinoids, morphine and related narcotics, cocaine, amphetamines, barbiturates and benzodiazepines. Forensic DNA analysis involves the following techniques: DNA extraction and quantification, restriction fragment length polymorphisms, amplified fragment length polymorphisms (Amp FLP), short tandem repeats (STR), gender identification, AmpliType PM and HLA-DQA1 and mitochondrial DNA typing. Trace evidence includes the analysis of paints, fibers, glass, gunpowder and primer residue detection, petroleum products and fire debris, explosives, and the comparison of impression evidence including fingerprints, footwear impressions, and tiremarks.

This wide range of samples, which must be investigated, requires the use of chemical, chromatographic, spectroscopic, spectrometric, microscopic and immunological techniques which are in common with the techniques presented in other sections of this Encyclopedia. These other sections, which present techniques that are utilized or can be utilized in forensic analysis include: *Clinical Chemistry*, for example see **Drugs of Abuse, Analysis of** and also **Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry**; *Coatings*, for example see **Microscopy of Coatings** and **Atomic Spectroscopy in Coatings Analysis**; *Environment: Water and Waste*, for example see **Biological Samples in Environmental Analysis: Preparation and Cleanup; Explosives Analysis in the Environment; Nitroaromatics, Environmental Analysis of**; *Industrial Hygiene*, for example see **Metals in Blood and Urine: Biological Monitoring for Worker Exposure**; *Nucleic Acids Structure and Mapping*, for example see **Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes; Polycyclic Aromatic Compounds Mapping; Restriction Landmark Genomic and cDNA Scanning; Sequencing Strategies and Tactics in DNA and RNA Analysis**; *Petroleum and Liquid Fossil Fuels Analysis*, for example see **Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels; Nuclear Magnetic Resonance Characterization of Petroleum**; *Pharmaceuticals and Drugs*, for example see **Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis; Mass Spectrometry in Pharmaceutical Analysis; Planar Chromatography in Pharmaceutical Analysis**; *Polymers and Rubbers*, for example see **Gas Chromatography in Analysis of Polymers and Rubbers; X-ray Scattering in Analysis of Polymers**; *Pulp and Paper*, for example see **Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry; X-ray Photoelectron Spectroscopy, Paper Surface Analysis by**; and *Mass Spectrometry*, for example see **Liquid Chromatography/Mass Spectrometry**.

There are 14 forensic science techniques presented in the Forensic Analysis section that follows this Introduction. These articles, together with the related articles from other sections cited above, provide a core resource of forensics chemical analysis.

Atomic Spectroscopy for Forensic Applications. Atomic absorption (AA) spectroscopy is utilized in forensic examinations involving suspected heavy metal poisoning determinations in tissue samples, other biological materials as well as for gunshot residue analyses. Forensic applications covered include: (i) analysis of foods, plants, snow, and tobacco (minor and trace elements in); (ii) trace element patterns in artificial fibers by inductively coupled plasma atomic emission spectroscopy (ICPAES); (iii) use of inductively coupled plasma mass spectrometry (ICPMS) to determine major elements in drugs as well as ICPAES for the determination of contamination elements that could serve as elemental pattern markers in sourcing batch production, manufacturing sites, or other discriminators that might lead to identifying clandestine laboratories; (iv) determination of modes of suspected poisonings often involving the measurement of mercury, arsenic, and so on, in foodstuffs and drugs, as well as the now common use of human hair as substrate; (v) determination and identification of ammunition manufacturers through the measurement of trace elements in bullets and cartridges utilizing ICPMS and ICPAES; (vi) gunshot residue analysis to determine if an individual or an object was in the vicinity of a firearm at the time of its discharge; (vii) metals analysis with ICPAES or ICPMS to discriminate between various steel or aluminum objects; (viii) discrimination of precious metals and glass manufacturers according to the trace element composition.

Capillary Ion Electrophoresis in Forensic Science. This covers the analyses of highly mobile, low molecular weight ions for the determination of potential antemortem disease states or time intervals between time of death and time of autopsy. This technique is also utilized for the detection and quantification of potassium and lithium ions in biological fluids and tissues collected in clinical and forensic toxicology.

Chiroptical Spectroscopy in Drug Analysis. This technique using polarimetry for characterization of enantiomerically resolved drug compounds is useful in identifying drugs in forensic analysis. This article describes

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

circular birefringence, optical rotary dispersion and circular dichromism as used in identifying pharmaceutical diastereomers including alkaloids, terpenes, barbiturates, benzodiazepines, indoalkaloids, nucleic acids and nucleotides as well as amino acids and peptides and proteins and sugars. Metabolites of these compounds can also be identified.

DNA Extraction Methods in Forensic Analysis. The extraction of DNA from scarce or badly degraded substances remains a significant problem and the ability to overcome it remains crucial since the success of all DNA analyses depends on successful initial extraction. Forensic substrates vary from fresh materials with abundant DNA, through putrefied materials also with abundant DNA, to clean materials and dirty materials with scarce DNA such as sperm, saliva, blood stains, nails and feces. Extraction procedures from cell lysis, protein digestion and solubilization of DNA as well as nonaffinity extraction, affinity and commercial methods are all covered.

Fluorescence in Forensic Science. This covers the uses of photoluminescence in criminalistics for ultrasensitive latent fingerprint detection, detection of bodily fluids, visualization of fibers and ink discrimination. While the focus of the article is on visualization and imaging, a discussion of basic molecular photoluminescence phenomena and a description of some useful spectroscopic measurements are also presented.

Ion Mobility Spectrometry in Forensic Science. The ion mobility spectrometer is a device that is used primarily for the detection, identification and monitoring of trace amounts of gases and vapors. The instruments used are typically small, portable, handheld or even pocket sized and respond within seconds to parts per billion or even parts per trillion level concentrations. Principles of operation, instrument design and sampling as well as sample introduction are presented. Forensic applications for the detection and determination of explosives, illicit drugs and drug abuse, tear gas, arson, chemical warfare, chemical warfare agents and biological materials are discussed.

Immunoassays in Forensic Toxicology. This covers enzyme immunoassay (EIA), radioimmunoassay (RIA), fluorescent polarization immunoassay (FPIA), kinetic interaction of microparticles in solution (KIMS), enzyme-linked immunosorbent assay (ELISA) as well as on-site drug testing immunoassays. Sample preparation and matrix factors are described as well as key issues for forensic applications such as detection limits, cut-offs in ranges, cross-reactivities, interfering substances and interpretation of results. Government regulations, quality control, methods development and comparison with other screening procedures are also covered.

Mass Spectrometry for Forensic Applications. This covers mass spectrometry for DNA fragment sizing as well as a number of forensic applications including DNA fingerprinting and the determination of STR and single nucleotide polymorphisms as well as mass spectrometry for analysis of DNA samples for contaminants from the environment. The techniques include electrospray and electrospray time-of-flight mass spectrometry.

Microspectrophotometry in Forensic Science. This covers the use of infrared, Raman and ultraviolet (UV) microspectrophotometry, as a combination of optical microscopy and spectrometry, to supply specific information about the composition and structure of forensic materials. Typically, a Fourier transform infrared spectrometer or ultraviolet/visible (UV/VIS) spectrometer is connected online with an optical microscope. In forensic analysis this is used for the identification of very small amounts of fibers, paints, traces of ink or ballpoint ink on a forged document, and so on. Advantages outlined include the ability to analyze picograms or smaller samples, unambiguous identification and photographing of measured areas of samples, the archiving of measured areas and the elimination of the burdensome process of preparing samples prior to analysis. It is also applicable to forensic samples such as single fibers, inks, hair and oils from the clothes of victims. Additional techniques described include fluorescent microspectrometry.

Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs. This covers sample preparation, nuclear magnetic resonance experiments and applications in the confirmation of known compounds, identification of unknown compounds, quantification, impurity profiling, optical purity and analysis of body fluids. The determination of a number of drugs is covered, including amphetamines, phencyclidines, fentanyls, opiate alkaloids, cocaine, cannabinoids, ergot, alkaloids, quinazolinones and barbiturates.

Polymerase Chain Reaction in the Forensic Analysis of DNA. This covers the polymerase chain reaction (PCR) technologies for amplification of a particular sequence of DNA in forensic science. A description of the technique is provided, followed by coverage of the analysis of PCR product by Amp FLP and STR methods as well as the use of fluorescent PCR primers. Then, the various techniques used in forensic PCR testing are described as a function of chromosome number and gene name (if named).

Pyrolysis Gas Chromatography in Forensic Science. This covers the pyrolytic breakdown of organic molecules by the application of heat in the absence of oxygen and the separation and detection of the pyrolysis products by gas chromatography. This technique is used most often in forensic science to help answer the question, ''Could samples A and B have originated from a common source?'' Pyrolysis methods are described, followed by the experimental parameters and quality control,

with a description of forensic sample types, including, ink, fibers, rubber and plastic, photocopy toners and adhesives.

Scanning Electron Microscopy in Forensic Science. This presents a description of the scanning electron microscope technique including sample preparation and instrumental parameters. Applications covered include forensic testing of paint chips, glass, metallic fragments, explosives and arson residues, hairs and fibers.

X-ray Fluorescence in Forensic Science. This includes a description of elemental analysis, X-ray fluorescence in forensic protocols, conventional X-ray fluorescence spectrometry, electron probe microanalysis, X-ray fluorescence microanalysis and total reflection X-ray fluorescence spectrometry. Applications are described for the forensic analysis of metals, gunshot residues, paint, glass and soil, fibers, plastics and general polymers.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
Amp FLP	Amplified Fragment Length Polymorphisms
EIA	Enzyme Immunoassay
ELISA	Enzyme-linked Immunosorbent Assay
FPIA	Fluorescent Polarization Immunoassay
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
KIMS	Kinetic Interaction of Microparticles in Solution
PCR	Polymerase Chain Reaction
RIA	Radioimmunoassay
STR	Short Tandem Repeats
UV/VIS	Ultraviolet/Visible

Atomic Spectroscopy for Forensic Applications

Diana M. Grant and Charles A. Peters
FBI Laboratory, Washington, USA

1 Introduction 1
2 Development of Atomic Spectroscopy 2
3 Applications 3
3.1 Natural Products 4
3.2 Drugs of Abuse 5
3.3 Suspected Poisonings 6
3.4 Lead Applications 8
3.5 Gunshot Residue Analysis 10
4 Metals Analysis 11
4.1 General Applications 11
4.2 Precious Metals 12
5 Glass 13
6 Laser Ablation 15
Acknowledgments 17
Abbreviations and Acronyms 17
Related Articles 17
References 18

Forensic science is a discipline that evolved from the interaction between law enforcement and the scientific community. The typical operating budget of laboratory facilities created to support law enforcement is often far below what is necessary for practical modernization. Hence, routine analyses that require low maintenance and can be readily performed with practical experience were quickly instituted as standard procedures in many such laboratories. As an example, atomic absorption (AA) spectroscopy has served the forensic community for over 40 years and continues to work effectively for such diverse applications as gunshot powder residue analysis and toxicological examinations in suspected heavy metal poisoning cases.

These and other traditional forensic applications have also utilized other forms of spectrochemical analysis such as neutron activation. The advantage of the various forms of atomic spectroscopy over these other methods lies in the practical considerations of accessibility and cost as much as in sampling requirements and detection limits. From well established AA methods to the increasing number of forensic applications being developed for laser ablation/inductively coupled plasma mass spectrometry (LA/ICPMS), modern atomic spectroscopy has proven itself to be a compelling and evolving tool in the investigation of forensic evidence.

1 INTRODUCTION

Forensic science may be defined as the application of scientific methods to problems that are addressed in a courtroom setting. Given that the results of an examination have the potential to affect a verdict at trial, the law provides that physical evidence may be examined by both prosecution and defense experts. The size of a forensic specimen can thus present a sampling challenge for analysts, particularly when a requested examination is destructive. Preservation of collected evidence is always a consideration prior to examination. The analyst must ask the question: ''Has the evidence been altered since or as a result of its collection?'' Another issue that arises is the order in which examinations are performed so as not to preclude subsequent analyses.

Evidentiary samples are generally analyzed for relational value to a suspect, victim, or crime scene. Specimens may readily associate a subject with a crime scene or victim, as with DNA or fingerprints.[1,2] More often, however, evidence from a victim or crime scene may be used as a circumstantial link to known materials found on a suspect or in his environment.

For this type of specimen, the associative link will be comparative. That is, the relationship between a known and the questioned evidence will be assessed as to the degree of similarity. Whether the association is formed on the basis of color, texture, diameter, physical impressions, or compositional similarities at the elemental level, the forensic examiner's written opinion will assess the degree of comparability between the submitted known and questioned specimens.[2]

Forensic scientists are often asked to determine which examinations are best suited to the collected evidence in a given case. Dialog between the scientist and the evidence contributor is then necessary to establish what questions are being asked and if answers are reasonably achievable given the limitations of sample size and the laboratory's analytical capabilities. After all known or available analytical tools are employed to address the question(s) posed to the laboratory, a solid scientific foundation should be established upon which a legal opinion may be rendered. An ideal scenario would allow that any necessary instrumentation would be readily accessible, specimen size would be plentiful, and the condition of the evidence would be pristine.

External contamination and minimal sample size are just two examples of the problems often encountered

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

with forensic evidence. As with any object in limited supply, preservation of materials deemed to be evidential in nature is often a fundamental consideration in requests for analysis. Thus, an arguably nondestructive method such as instrumental neutron activation analysis (INAA) is quite useful for a wide range of applications in forensic science assays.[3–7] However, INAA has not been readily available for use by most forensic laboratories. The lack of convenient accessibility, required licensing for use of such facilities, and regulatory waste disposal issues has led to sustained interest in alternative methods of analysis such as atomic spectroscopy.

AA spectroscopy has been well documented in forensic examinations involving suspected heavy metal poisoning determinations in tissue samples, other biological materials, and for gunshot residue (GSR) analyses.[8–19] While INAA has the advantage of multielement capability for solid samples, which preserves the condition of a specimen, AA is more readily accessible, and ultimately cheaper to use and maintain. The merits of both methods were compared in early forensic application papers.[20,21]

Atomic spectroscopy most routinely involves dissolution of a solid matrix before analysis. Traditional AA utilizes a flame or furnace system to atomize aqueous samples. Later systems that allow for solid sample atomization have improved sensitivity for a variety of matrices.[22] Most recently, plasma technology has grown in popularity due to its efficiency in ionizing an aqueous sample.[23] Modes of detection have also evolved to improve sensitivity for a variety of matrices with these systems. They range from photomultiplier tubes (PMTs) or electrodeless discharge lamps (EDLs) used in conventional AA instruments to the charge transfer devices (CTDs) found in inductively coupled plasma atomic emission spectrometry (ICPAES) instrumentation. Some ICPAES instruments still rely on PMT technology for wavelength recording and enhancement. These systems are still quite useful for some applications and can reach comparable detection levels in certain well-defined matrices. Further evolution has produced inductively coupled plasma mass spectrometry (ICPMS) instrumentation, which utilizes either a quadrupole or magnetic sector detection system for mass-to-electric charge ratio (m/z) determinations.

Limits of detection for these various atomic methods can range from microgram per gram ($\mu g\,g^{-1}$) to femtogram per gram ($fg\,g^{-1}$) concentrations depending on the analyte of interest and the sample matrix.[23] Many good general reference books have been written about AA spectroscopy, ICPAES, and ICPMS respectively. However, for a practical introduction to each, Beaty and Kerber,[22] Boss and Fredeen,[23] and Montaser[24] are useful for both beginner and refresher usage.

2 DEVELOPMENT OF ATOMIC SPECTROSCOPY

The notion of coupling a plasma excitation source with a multielement detection system in the mid-1960s ushered in a new generation of workhorse instrumentation for the modern analytical laboratory. Since then, industrial, environmental, and academic laboratories have demonstrated a wide applicability of plasma techniques to trace element studies.[25–32] Forensic laboratories have also published studies of atomic spectrometry applications.[33–36] However, as a discipline, forensic science seeks to apply analytical techniques to evidence in a criminal matter for the purpose of weighing the probative value of that evidence. Therefore, only those techniques that have been tested and validated by the scientific community are suitable for presentation in a courtroom. Such methods are thus no longer considered research endeavors as much as they are new applications of established protocols, already discussed extensively in relevant journals.

Moreover, while all of the conventional atomic spectrometry techniques mentioned thus far have become well established in a variety of analytical problem solving arenas, AA spectroscopy is still the most popular of these methods for routine forensic examinations.[37] Examples range from GSR analysis to toxicological assays for the presence of heavy metal poisoning in hair.[38,39] Therefore, a discussion of atomic spectroscopy techniques in forensic applications should begin with classical AA protocols.

Slavin[40] credits a 1930 publication by Muller and Pringsheim as the initial discussion of AA instrumentation. This paper described the use of AA spectroscopy in the determination of mercury, an element that remains important in current literature.[41] The first articles that addressed the use of AA spectroscopy in forensic applications, however, did not appear until the early 1970s.[10,12,13,42] These articles, which pertained to GSR analysis, represent early examples of the use of modern atomic spectrometers in forensic laboratories. Before this series of articles, GSR literature concentrated almost exclusively on wet chemistry techniques.[43]

Forensic assays have recently focused on the organic nature of evidence. The most current forensic science review article by Brettell and Saferstein[44] illustrates the number of articles that characterize materials by their organic components. This review of forensic science publications also contains a section devoted to trace evidence studies. It is interesting to note that there are no atomic spectrometry articles listed for the surveyed period that encompasses January 1994 to December 1996. However, more than 300 references related to assorted atomic spectrometry applications appear in an April 1994

review published in the *Journal of Analytical Atomic Spectrometry*.[45] Moreover, many of these topics could be readily interpreted as forensic applications.

The publication of this type of research in a more mainstream spectroscopic journal lends strength to the argument that forensic science is simply applied analytical chemistry. That it is performed as a service to the criminal justice system is secondary to its nature as scientific analysis. For this reason, it is vital that forensic scientists also maintain familiarity with mainstream scientific literature, which often contains articles regarding forensic topics.

Several papers that have concentrated on the use of atomic spectroscopy in forensic applications were published in 1985 by Carpenter,[46] in 1993 by Koons,[47] and by Lalchev et al.[48] in 1997. Each article focused on the use of ICPAES for trace element patterns in materials such as glass, brass, bullet lead, and elevated levels of toxic elements in animal tissues. These articles demonstrate the range of materials that can be of forensic interest, a discipline rich in compositional diversity.

In this diverse environment, developing a preparation method best suited for a given sample size and matrix is a prudent approach to any research study. Several journals publish annual reviews to highlight articles that have been published in a given specialty. These sections are generally grouped by application, with separate editions dedicated to general instrumental advances in atomic emission,[49–53] AA and fluorescence,[54] and atomic mass spectrometry,[55,56] respectively. Perusal of these types of review issues is a useful and inexpensive means of staying current with new instrumentation. Moreover, journal sections that describe recent applications in environmental, industrial, clinical and biological materials all contain novel sampling discussions that can be useful to forensic analysts regarding these types of materials.[57–69]

As discussed earlier, there have been numerous forensic applications throughout the development of atomic spectroscopy. The following sections will attempt to discuss these topics in terms of the various analytical methods that have been applied to a given sample type. Some areas will show the evolution of the spectroscopic techniques available for use. Others will demonstrate that ''new'' is not always improved, depending on a given sample matrix and analyte of interest.

3 APPLICATIONS

A 1993 article in the *Fresenius Journal of Analytical Chemistry* by Paudyn and Smith[70] describes a procedure for the analysis of lead-based paint scrapings by ICPAES after microwave assisted digestion. Using aqueous standards and neutron activation analysis (NAA) as a reference technique, results were reported for both major and minor trace elements in several types of paint samples, including automotive spray paint. A low pressure nitric acid–hydrofluoric acid digestion mixture followed by ICPAES yields comparable results to neutron activation data with equivalent relative standard deviations (RSDs) of approximately 5%. For studies that serve to associate the trace elemental compositions of two or more specimens, this level of uncertainty is generally considered to be acceptable. Furthermore, in comparative examinations, which is the basis of most forensic work, it is more critical to develop associative information about each of the known and questioned specimens than to attempt to fully characterize the trace element profile of any one specimen. Therefore, precision errors approximating 5% between replicate trials of a given specimen have been deemed reasonable for this type of analysis.

The article also mentions earlier work using AA and direct current plasma (DCP) spectrometry, which is beneficial to readers who have access to this type of instrumentation. Analytical instrumentation purchasing decisions in forensic laboratories are often based on practical budgeting rather than ideal conditions. As a result AA, particularly flame AA, remains a popular technique. Part of its ubiquity is due to the plethora of experienced technicians who can operate, maintain, and, if necessary, repair such instrumentation.

Virtually everyone with a physical science degree has had exposure to AA in some college course, internship, or employment opportunity. Moreover, procedures for its use with many forensic sample matrices as well as its limitations and interferences are thoroughly documented in textbook literature. Thus, it is widely accepted in the forensic community.

Another reality of forensic work is that specimens are frequently received in less than ideal sample quantities or conditions. While the article by Paudyn and Smith[70] suggests an attractive forensic application, most evidentiary paint specimens are chips rather than scrapings. Moreover, comparison of such a complex matrix would most assuredly require a multielement examination technique such as inductively coupled plasma (ICP) or DCP, rather than the time-consuming single-element capability of the traditional AA equipment found in most forensic laboratories.

It is also interesting to note that forensic applications which appear in mainstream atomic spectrometry journals often cite the lack of available background information for the reported research. A 1998 paper by Zieba-Palus[71] published in *Forensic Science International* describes the study of used motor oils by flame AA as a potential forensic examination. The paper also notes that similar work has been published by two separate groups using electrothermal AA spectroscopy and ICPAES. Both of

the latter techniques would be more desirable alternatives to flame AA. Graphite furnace atomic absorption spectroscopy (GFAAS) has detection limits about two orders of magnitude better than either flame AA or ICPAES. For its part, ICP is much faster at detecting emission signals simultaneously or in rapid sequence. Moreover, the higher temperature of a plasma derived excitation source allows for analysis of refractory metals.

Therefore, since this application could be developed for use with any of the aforementioned techniques, the underlying considerations of adequate sample size and predetermination of what conclusions could be drawn from the resultant data must still be assessed for each situation. For example, one must ask what probative value an oil spot on a suspected carjacker's clothing would have if he were an automechanic. However, if the victim was a mechanic and the suspect was an office worker who had motor oil on his shirt, then the compositional similarities between the victim's work environment and the suspect's stained clothing might prove to be significant corroboration of contact between these two individuals. Often, the circumstances of how transference of trace evidence can occur need to be assessed prior to initiation of a potentially meaningless examination.

3.1 Natural Products

Natural consumer products manufactured and grown for human consumption or use is another popular area of study in atomic spectrometry. The distinction between what constitutes a clinical, environmental, or regulatory application versus one that is forensic often depends on the objectives of a given study. While none of the following publications was written to address forensic issues per se, it is not unreasonable to view these articles as a model for that future use.

An article appearing recently in *Today's Chemist at Work*, a journal published by the American Chemical Society, described the role of the United States Customs Service in the analysis of imported foods.[72] Customs, a branch of the US Treasury Department, oversees enforcement of trade and tariff regulations. For several years, this agency has utilized ICPMS for the analysis of the trace element compositions found in various food products. Since these levels will vary depending upon where the product was grown, such data are a good indication of the country of origin.

This information aids Customs in several important commerce and food safety issues. First, the amount of a particular product from a given country is often regulated in accordance with trade agreements. Some countries are further banned from trade with the USA as a result of sanctions. Likewise, tariff rates can vary between countries that import the same commodity. Also, certain foods are not accepted from countries whose products could damage domestic crops or place consumers at risk, due to inherent disease, insect infestation, or pesticides used during their growing. This topic, as a novel example of the diversity of forensic work, was also covered by the *Washington Post*[73] and was presented by Schwartz,[74] the research chemist who developed the protocol, at the 1998 Winter Conference on Plasma Spectrometry held in Scottsdale, Arizona.

Originally, this work was performed with ICPAES, but has been furthered using ICPMS. Decreased instrument sampling time and improved sensitivity over a mass range comprised of 70 elements has allowed for more specimens to be analyzed in a timely fashion. As a result of the additional elements of interest and increased specimen throughput, better differentiation has been achieved between a larger number of samples. Thus, discrimination by country of origin is possible based upon a statistical modeling of the observed concentration differences between products from various global locales. To date, this work has been applied to imported raw peanuts, garlic, coffee beans, and orange juice.

Other articles have focused on the use of atomic spectrometric techniques for analyses of such diverse matrices as olive oil,[75] plants,[76–78] wine,[79–81] milk,[82] snow,[83] and tobacco.[84] Given the current litigious climate surrounding tobacco and its by-products, the ability to differentiate plant origins by trace element concentration levels would seem to be of interest to the legal community.

Similarly, an article in the *Indian Journal of Forensic Science* makes an argument that trace element composition comparisons between ashed tobacco specimens can be made using copper, cobalt, chromium, and zinc.[85] The article does not mention what type of standards were used. However, use of commercially available plant and leaf reference materials in addition to aqueous solutions would seem to provide adequate calibration and check standards.

Another study that looks promising utilized ICPMS to analyze minor and trace elements in cigarettes and raw tobacco.[86] Aqueous stock standard solutions, traceable to the National Institute of Standards and Technology (NIST, Gaithersburg, MD) were used along with standard reference material (SRM) citrus leaves as a check standard. A total of 30 elements were analyzed, including some of the lanthanides and uranium. Rhodium (^{103}Rh) was used as the internal standard. No mention is made regarding selected isotopes or interference corrections, both of which are quite important in interpreting the significance of the findings. Regardless, tobacco crops appear to be a promising area for continued forensic research endeavors.

Trace element patterns in man-made fibers have also been explored as a potential forensic application of ICPAES.[87] This study focused on inorganic additives in polyester fibers. An earlier article also discussed analysis of trace element distributions in several types of commercially manufactured fibers using activation analysis.[88] Comparing the two techniques as described in these articles, NAA would appear to be a better alternative for this application given its sensitivity for intact 2-mg fiber samples. However, uncertainties of ±20% are reported for major elements of interest, such as Ti, Sb, Mg, Cu, V, Al, Cl, and Na.

In the former study, it is difficult to ascertain the success of the ICPAES methodology. Little attention is paid to the use and development of the technique itself. Rather, the focus of the article is the statistical manipulation that may be applied to the volume of data that was collected using ICPAES instrumentation of unknown manufacture and configuration. This article therefore serves to highlight an unfortunate gap between research and practical applications; that is to say, the ability to combine critical procedural details with useful applicative conclusions often remains an unrealized goal in forensic literature. The information gap between the research scientist and the caseworking technician must be lessened if productive studies are to be implemented in real time laboratory settings. This disparity is not exclusive to the forensic community. However, it is a scenario that demands attention within this discipline as it continues to arise in evidence admissibility hearings.

New technology is increasingly being introduced into courtrooms, where it is placed under intense scrutiny. In these *Daubert* hearings (the result of the Supreme Court's opinion in Daubert vs Merrell Dow Pharmaceuticals 509 US ___, 113 S. Ct 2786), the weight of a technology's evidentiary merits is evaluated. In this regard, it is often easier to define the contribution of developmental research to the advancement of the larger scientific community than to the smaller populations it attempts to serve.

3.2 Drugs of Abuse

Atomic spectrometry is clearly impacting upon one high profile forensic application: the prosecutorial war on illicit drug manufacture and trade. Five separate articles were found in nonforensic journals that detail the interest and capabilities of atomic spectrometric techniques in the analysis of drugs of abuse and their derivatives. As early as 1988, when ICPMS had only been commercially available for approximately 5 years, an article in the *Journal of Chromatography* described its comparative use to ion chromatography for detection of inorganic impurities in organically pure methamphetamine.[89] This application featured analysis of major elements in the alkali and alkali earth series as well as iodine, barium, lead, and several light transition metals. Conclusions resulting from the work emphasized the inhomogeneity of the methamphetamine crystals studied, which therefore went on to require discrete multiple aliquot sampling per specimen. The article also notes that ICPMS was attractive for its speed, and sensitivity. The inaccessibility of neutron activation beyond government laboratories in the authors' native Japan was also cited as a reason for exploration of an alternative analytical method. A separate research interest was explored in the same year regarding the use of ICPMS for detection of trace elements in methamphetamine hydrochloride.[90] Many of the same elements were analyzed with NAA used as a reference method.

More detailed papers using ICPAES and ICPMS separately or in concert for analysis of trace element signature impurities in drugs of abuse appeared in the literature in 1995. In a paper authored by members of the National Forensic Chemistry Center, the United States Food and Drug Administration Laboratory in Cincinnati, Ohio, γ-hydroxybutyrate (GHB) was analyzed by ICPAES for the presence of Ba, Ca, Cd, Fe, K, Mg, Ni, P, Pb, Si, Sr, and Zn.[91] The results were promising as a starting point for identification of ''contamination'' elements that could serve as elemental pattern markers in sourcing batch production, manufacturing sites, or other discriminators that might lead to identifying clandestine laboratories. To date, however, it is unknown to the authors if any of this research has been introduced in court. Recall that the Daubert standard, with its four prongs to determine admissibility – testing, validation, error rate, and general acceptance in the scientific community – now applies to all scientific evidence presented in US federal courts.

Also described in this paper was the analysis of ephedrine hydrochloride via ICPMS. The presence of 14 elements including Na, Al, Mn, Rb, Sr, Ba, and Pt was found to serve as a good discriminator between batches of ephedrine samples. Sodium later showed poor day-to-day reproducibility, although it had originally proved encouraging in same-day production trials. The authors theorized that this discrepancy was attributable to cross-calibration problems between the pulse counting and analog detector modes in the ICPMS instrumentation. Given their respective concentrations, detection of the 14 analytes would likely split evenly between the two detector modes. Thus, this problem is nontrivial to atomic spectroscopists and yet not obvious to inexperienced users of the technique.

Another ICPMS paper from 1995 detailed the analysis of trace elements in heroin.[92] The approach used for sample comparison resembled the paper described earlier by Schwartz of the US Customs Service, in that the heroin

was analyzed for country of origin comparisons between major producers such as India, Turkey, Iran, Pakistan, Mexico, and Burma. Not surprisingly, Mexican data were quite different from the other countries in the study, which share a common geographical locale. Further, Iranian and Pakistani data were indistinguishable on the dendrogram of all analyzed samples. While this paper is a nice representation of the power of ICPMS for use in multivariate statistical compilations, it is more useful for data processing than method development. Conversely, studies such as Schwartz's work can be an essential component in comparative analytical research due to its description of the analytical method, equipment, and data interpretation steps.

Further controlled substance research utilized electrothermal vaporization AA spectrometry for the analysis of Ag, Al, Cd, and Mn in cocaine and heroin samples.[93] The charring step was omitted from this research in favor of decreasing the sample pretreatment time. Higher drying temperatures were studied, as were lower temperatures in the presence of chemical matrix modifiers. The beauty of this work from the viewpoint of a bench chemist was its use of instrumentation that may already be standard equipment in most forensic laboratories. However, the challenges of adapting a procedure toward a given set of practical conditions including available instrumentation and software options, sample design, operator experience, and time and budget constraints can represent a potentially lengthy technical and logistical research effort.

A more recent application of ICPMS with regard to analyses of drugs of abuse was introduced at the 1998 Winter Conference on Plasma Spectrometry.[94] It was later described in the September 1998 issue of the *Journal of Analytical Atomic Spectrometry*.[95] Researchers in Australia have attempted to source the provenance of cannabis crops using LA/ICPMS. Elemental patterns of crops known to be grown in different parts of western Australia were studied in an attempt to characterize plant trace element compositions as a product of the soil environment from which they were harvested. Plant specimens were freeze dried, pressed into disks and ablated for sample introduction into a conventional ICP mass spectrometer. Optimization of the instrumental operating parameters was achieved using glass SRMs commercially available from NIST. It is the author's conclusion that these data would be quite favorably received by a jury of nonscientists due to the obvious concentration differences that could be observed in plants that were geographically isolated from one another. However, the subtle differences that can occur in plants from similar regions, such as that noted in the data of heroin specimens from Iran and Pakistan previously cited, would require a more thorough interpretation in order for the probative weight of the results to be assessed.

3.3 Suspected Poisonings

The distinction between environmental, clinical, and forensic applications is not easily delineated when discussing cases where poisoning is suspected. The methylmercury exposure of hundreds of Iraqi citizens in the early 1970s was attributed to ingestion of grain that had been treated with fungicides that contained mercury.[96] While the circumstances were extremely unfortunate, the intent was not criminal. Similarly, the chronic and acute arsenic poisoning of almost 75 consumers in Singapore of traditional Chinese medicines in 1975 was characterized by researchers as overuse of the product and poor quality assurance testing by the manufacturers.[97] However, the presence of cyanide in Tylenol capsules in the USA in 1982 was attributable to an individual's willful intent to endanger public safety. Therefore, the results of an environmental offense may often have similar consequences as those of crimes against either specific or nonspecific victims. Chain of custody issues and potentially limited sample availability, then, strengthen the commonality in the analyses used for the respective investigations.

Atomic spectroscopy has long been utilized in biological applications that embody a general forensic description. In one study, silicon, in the form of organosilicon oxide polymers, was analyzed by AA with a hollow graphite tube assemblage.[8] This research was conducted as a result of the use of silicone in breast implants. Similar analyses regarding the chronic effects and environmental impact of heavy metals on populations have also been conducted using emission spectroscopy.[98,99] Mercury determinations by a variety of atomic techniques have also continued to span a wide series of forensic interests.[100–104] Moreover, the development of ICPMS has generated tremendous interest in the levels of trace elements in biological matrices.[105–113] A 1997 paper even compared the potential of ICPMS versus NAA in the multielement analysis of earthworms, as indicators of ecotoxicological assessments and the bioavailability of metals in soil.[114]

Arguably, the greatest contributions that ICPMS has made to spectrometric studies have been in the areas of elemental speciation and isotopic ratios. Specifically, arsenic and mercury speciation studies are of interest to forensic analysts, and these topics have been well documented in recent literature.[114–123] Lead isotope ratios are of particular interest due to the use of lead in water pipes, paint, gasoline, and pottery. Isotopic ratios of lead are often studied in an attempt to trace the lead back to a specific source for both environmental and

archaeological purposes. If the lead is known to have been mined as a primary resource, rather than recycled from a previous use, these ratios can provide useful information as to the isotopic composition of this lead source. Compiling data from various lead mining sites provides a compositional map that can be compared to the lead specimen under investigation in a given study.[124]

Lead isotope ratios in modern-day goods such as ammunition would be of interest to the forensic community, and has been explored by several researchers. However, bullet lead is not culled from a primary source. Rather, it is mostly harvested from recycled automotive batteries, and therefore is not directly traceable back to its origins. Greater emphasis will be placed on this application in a later section.

One of the most popular biological matrices that has been studied in suspected heavy metal poisoning cases is human head hair.[125–127] This matrix is easier to work with than most biologicals, in terms of sample preparation and analytically detectable element profiles, when analyzed in bulk. Elemental concentrations in hair can be as high as 300 times that of blood specimens, with microgram per gram quantities generally considered as an average baseline concentration.[128] Each cited reference describes what a trace element examination of human hair can reveal regarding the subject's recent environment. They also offer suggestions for sampling.

Much of the early analytical literature, which detailed potential differences between rate of hair growth in males versus females, elemental deposition along the length of a single hair, and, suspected poisonings, were studied using NAA.[129–134] As with many types of specimens, the ability to analyze several milligrams of a solid for up to 30 elements simultaneously is an attractive feature of NAA, assuming that adequate facilities are available for use.

Analyses of hair have also been routinely performed using AA methods, with detailed studies reporting arsenic levels in hair as indicators of environmental exposure or chronic poisoning.[135–145] Such work was also conducted in the early years of suspected arsenic poisoning cases that were analyzed by the FBI Laboratory. However, the advent of the graphite furnace gave way to the development of a protocol that was instituted in 1994. Using this procedure, individual solid head hair samples were placed directly onto a graphite platform situated inside a hollow graphite tube. A chemical matrix modifier composed of palladium and magnesium nitrate was then instrumentally pipetted onto the hair within the graphite tube.[146] NAA was used as a reference method. This work was of great importance to the laboratory for several reasons. Hair specimens, 5 mm in length, were oriented root-to-tip and segmentally analyzed in order to develop a theoretical modeling of a time line during which the poisoning was thought to have occurred. Bulk analysis of hair specimens can prove to be troublesome when multiple strands are analyzed simultaneously because hair can remain attached to the scalp through several stages of maturation. However, this process allows for a representative sampling of the submitted specimen. Moreover, the solid sampling technique eliminated timely dissolution procedures and inherent contamination risks associated with this sample preparation step. While this method is quite useful for single element determinations, the sensitivity and multielement capabilities of a technique such as ICPMS make it a more attractive alternative when the constituents in a suspected inorganic poisoning case are unknown.

ICPAES has also been used for bulk analyses of metals in cut head hair. However, the cited study determined concentrations for major (Na, Mg, Ca, and K) and first row transition elements using a sample size of 800 mg. Specimens of this magnitude are a rarity for forensic case submissions.[147] As with other applications, ICPMS has created new and exciting possibilities for the examination of trace elements in hair. At least two papers describe recent exploration into environmental pollution effects on cut, bulk head hair samples using ICPMS. A paper from 1994 gave detailed sample preparation criteria and instrumental parameters.[148] That work featured indium as an internal standard, with emphasis placed on cadmium and lead determinations. Levels of these elements and several others, including arsenic, were monitored in hair as a bioindicator of exposure levels in smokers and nonsmokers. A separate 1998 publication described the method used to monitor levels of metals (As, Cd, Cr, Hg, Mn, Ni, Pb, and Sn) in the hair of children in Spain.[113] An internal standard was not used in the latter study, in which arsenic and mercury analyses were conducted using hydride generation rather than a conventional sample introduction system for ICPMS.

Single strands of hair were analyzed for mercury and thallium using ICPMS in a 1993 article by members of the National Institute for Environmental Studies (NIES) in Japan.[149] Flow injection which requires a small sample volume, typically 50 μL, was used for specimen introduction to the plasma. Each sample was segmented in 3–5 mm increments from proximal to distal ends of a single hair. A human hair certified reference material (NIES CRM No. 5) commercially available from the NIES was similarly studied. Mercury was analyzed in the hair of a healthy male whose baseline mercury hair content was observed to fluctuate over a period of 6 months while he lived in Japan. The authors attributed these data to an increase in fish intake in a subject who prior to this timeframe had rarely consumed fish as part of his daily diet. Hair from a patient who was known to have been poisoned with thallium was also

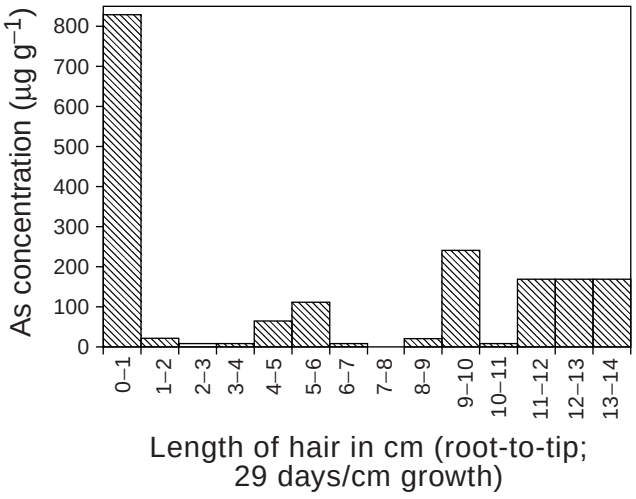

Figure 1 Chronic poisoning occurring over the course of 1 year.

analyzed. Specimens were collected 3 weeks after he was hospitalized. Levels of $35\,\mu g\,g^{-1}$ were detected in the peak section of the hair sample, with the lowest levels ($\sim 3\,\mu g\,g^{-1}$) also exceeding predicted normal thallium levels ($0.016\,\mu g\,g^{-1}$). It was theorized that the patient had been exposed to thallium from as early as 8 months prior to this sampling study, which was conducted 3 weeks before his demise.

In the Spring of 1998, the FBI Laboratory received two suspected arsenic poisoning cases within weeks of one another. The decision was made to develop a protocol for ICPMS in anticipation of eventually using this technique for multielement, heavy metal screening purposes. This course of action proved to be particularly timely when the hair of a cohabitating suspect in one of these cases was later submitted in order to rule out environmental factors, after the victim's hair was found to contain elevated levels of arsenic. Figure 1 represents data from a typical chronic poisoning event that was calculated to have occurred over the course of 1 year.

The two original submissions were first analyzed by the single solid hair segment protocol developed for GFAAS noted earlier. These analyses did not yield the expected results, based on details submitted by the contributors in each of the two cases. Therefore, ICPMS was attempted on bulk hair segments, with solution AA also performed on the digested samples as a reference method. Aqueous arsenic standard concentrations ranged from expected normal arsenic levels of roughly $0.04\,\mu g\,g^{-1}$ up to $64\,\mu g\,g^{-1}$. Indium (^{115}In) was used as an internal standard.

Two isotopes of chlorine (^{35}Cl, ^{37}Cl) were monitored for comparison to the expected argon chloride ($^{40}Ar^{35}Cl$) interference on the lone naturally occurring arsenic isotope at mass 75. After repeated analyses with this method and discussions with Wallace[150] of Perkin-Elmer, it is now thought that concerns regarding the introduction of chlorine from the hair matrix are unfounded given that the hair is thoroughly washed and rinsed in alcohol prior to analysis. However, if continued monitoring for chlorine is desired, a better gauge of its presence would be observation of the selenium isotope at mass 77, where $^{40}Ar^{37}Cl$ would be observed as an interferent. Similarly, selenium at mass 82 should be monitored for comparison to Se-77.

Table 1 GFAAS versus ICPMS values in a suspected arsenic poisoning

Hair length (cm)[a]	Sample weight (mg)	GFAAS (ppm)	ICPMS (ppm)
0–1 (root)	1.655	Not detected	Not detected
1–2	1.500	(3)	$7.8 \pm 0.6\%$
2–3	1.638	$72 \pm 3\%$	$64 \pm 0.6\%$
3–4	1.631	(5)	$4.2 \pm 1\%$
4–5	1.536	$81 \pm 2\%$	$74 \pm 0.4\%$
5–6	1.544	$490 \pm 0.4\%$	$458 \pm 0.6\%$
6–7	1.365	(1)	$2.7 \pm 3\%$
7–8	1.245	Not detected	$9.6 \pm 2\%$
8–9	1.103	$38 \pm 15\%$	$41 \pm 1\%$
9–10	0.850	Not detected	$3.6 \pm 3\%$
10–11 (tip)	0.506	$110 \pm 14\%$	$124 \pm 0.6\%$

[a] Approximately 25–50 head hairs in each segment.

Agreement between the two atomic methods was not as good as expected. Concentrations differed by as much as 10–30% between the GFAAS and ICPMS data. Results derived from the conventional AA procedure were therefore reported, given that this protocol was better defined and validated. Table 1 lists the values obtained for both methods in a representative suspected poisoning case.

Subsequent cases were analyzed as bulk specimens due to a lack of sample. Both the aqueous AA and ICPMS protocols were utilized for these cases as well. The AA data were used as the reported method while development work continues on the ICPMS procedure.

3.4 Lead Applications

Using conventional mass spectrometry equipment, lead isotope ratios have been discussed in analytical literature since the mid-1970s.[151,152] This type of research was first published as applied work in specific fields of study that sought to trace lead based objects back to an original source. An article from 1976 appearing in *Archaeometry* discussed the lead isotope ratios in artists' lead white pigment.[153] This material is ubiquitous in paint media, but its isotopic ratio information has the potential to aid archaeologists in the characterization and source identification of specific production lots. An earlier

paper from 1975 introduced the concept of analyzing lead ratios in bullets for the same purpose.[154] Similar studies have been carried out more recently.[155–157] The original paper by Haney and Gallagher[154] noted that previous work had been done in this area for archaeological and environmental purposes. However, archaeologists traditionally analyze lead pigment from a time when trade and travel was conducted along specific routes that can be usefully mapped. Also, the production of this material was limited. Neither scenario applies to modern-day manufacturing processes.

Ammunition manufacturers estimate that current production rates yield approximately 9 billion bullets per year. This number does not include lead used for shotshell pellets. Another complication is that the four leading ammunition manufacturers in the USA routinely obtain lead ingots from common distribution sources. Moreover, this lead is not from an original source either. Bullet lead is recycled from sources, such as automotive batteries, which may be doped with manufacturer specified levels of antimony at a smelting facility and then shipped to the respective ammunition producers. Trace element concentrations in the recycled lead are generally not a concern for bullet manufacturers, so long as the majority of the product is properly formed and passes routine quality assurance test firing.

With the exception of increasing automation, bullet manufacturing has not changed significantly since mass production became popular during World War I. Lead ingots are purchased from a recycling facility that smelts the lead into a manufacturer's specified alloy that may contain up to 3% antimony for hardness. Each ingot, referred to as a ''pig'', weighs about 60 pounds. Pigs are melted in a batch process that allows for a continuous supply of the molten product to be solidified into billets, weighing about 125 pounds each. The billets are then forced through a die of specified diameter to form wire, from which the individual bullets are cut and shaped. The uniqueness of each batch of lead is a result of the continuous addition of ingots and recycled scrap lead back into the primary cauldron or melting pot. A batch is thus determined by the size of a given manufacturer's final melt.[158]

The formed bullets are only one component of the finished product known as a cartridge. Once the bullet is placed in a cartridge case that houses a primer cup and gunpowder, the assembled cartridges are packaged in sets that range from 20 to 100 per box. Lot numbers are assigned at the time of cartridge assembly and packaging. Therefore, hundreds of boxes may have the same lot number. Due to the mixing of formed bullets prior to assembly and packaging, a single box has been observed to contain up to 12 unique trace elemental composition groups depending on the manufacturer's protocol for loading cartridges into lots for distribution.[158]

Several forensic laboratories have investigated compositional bullet lead analysis in research efforts using ICPMS.[159–161] However, ICPAES remains the preferred method for this type of examination in the FBI Laboratory's Materials and Devices Unit, where hundreds of these analyses are performed routinely each year.[162] The FBI's standard protocol utilizes seven trace elements to compositionally differentiate between lots of bullet lead: Sb, As, Cu, Bi, Ag, Sn, and Cd. Triplicate samples of each specimen are digested in 20% HNO_3 in an 80 °C oven for 4–5 h. After removal from the oven, most specimens are already a homogeneous solution; however, as a final aid to the digestion process, 10 μL of reagent grade HF (49%) is added prior to analysis. Specimens are run against a calibration curve that includes a 99.99995% lead blank standard.

Three SRMs commercially available from NIST are run in triplicate as check standards against this curve. The described method has been employed for over 10 years in the FBI Laboratory. During that time, recorded values for the stated SRMs have varied within an acceptable margin of relative error from those concentrations listed on the certificates issued by NIST. A reference database is utilized to store this information regarding repeated analyses of the SRMs. It is comprised of compositions that have been recorded using NAA, and three different ICPAES instruments, with quantitative information for Bi, Sn, Ag, and Cd resulting from ICPAES analyses only. Figures 2–4 illustrate the average precision obtained for arsenic, antimony, and copper, respectively, in SRM 2416: Bullet Lead (NIST, Gaithersburg, MD) during 30

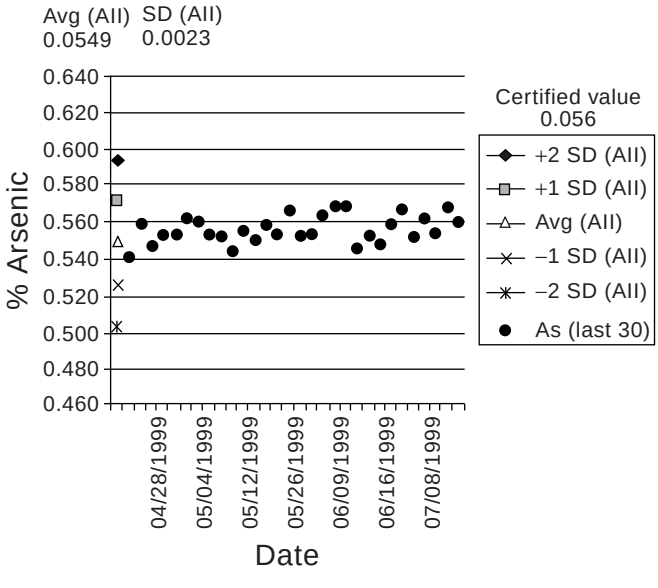

Figure 2 Average precision obtained for arsenic.

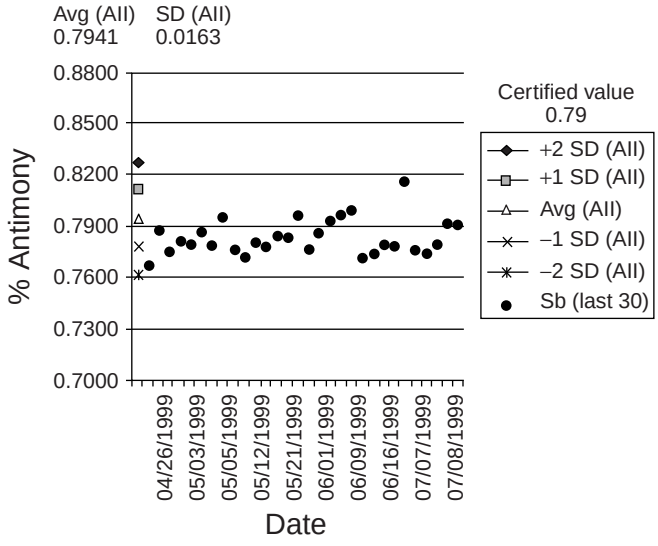

Figure 3 Average precision obtained for antimony.

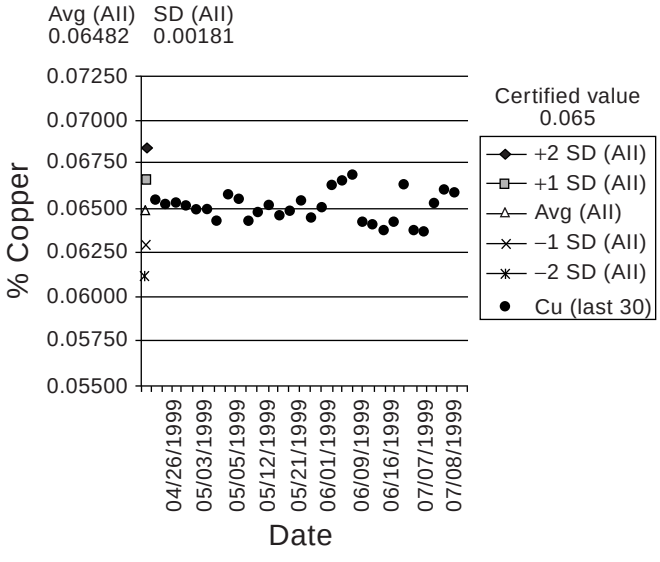

Figure 4 Average precision obtained for copper.

separate analyses of this SRM in the FBI Laboratory. Each graph depicts the mean and first and second standard deviations obtained for the respective elements as a function of the date on which the analysis was performed. These control charts serve as instrumental response monitors for the three main elements of interest in the comparative bullet lead methodology.[163]

3.5 Gunshot Residue Analysis

A much more common forensic examination involving fired ammunition is primer GSR analysis. GSR for the purposes of this treatise is described as the inorganic residues left behind from a cartridge's primer components. Organic residues derived from unburned or partially burned gunpowder are referred to as gunpowder GSR. Both species are associated with the discharge of a firearm. Therefore, GSR analyses are performed in order to determine if an individual or an object was in the vicinity of a firearm at the time of its discharge. Currently, the most common techniques used for this examination are AA spectrometry for bulk analysis of swabbed areas and scanning electron microscopy (SEM) examination of dabbed surfaces for particle analysis.[38]

The presence of GSR indicates that a weapon was fired. GSR examinations are used to determine if the firearm left behind detectable amounts of inorganic residue on surfaces in the vicinity of the discharge or surfaces with which the firearm was in direct contact sometime after its discharge. The results cannot determine the caliber of the weapon used, nor does it identify the shooter.

Subsequent handling of a discharged weapon would be an example of GSR deposition through surface contact. Assuming normal activity, GSR is not likely to remain in detectable concentrations on a surface 6–8 h after a weapon's discharge. Common sampling surfaces include objects at a crime scene and the hands or clothing of a person who may have been at the scene at the time of the weapon's discharge. However, if GSR is not detected, no conclusions may be drawn from this test.[164]

Research is currently in progress to bring GSR testing back on-line for a limited number of approved cases within the FBI Laboratory. The techniques that will be employed are SEM and ICPMS. The latter technique will serve as a screening process for the proper ratio of three elements that define the presence of residue from a discharged weapon: antimony, barium, and lead. If the ICPMS results are conclusive, particle analysis using SEM will be performed to detect the presence of primer particles.

One of the first GSR test methodologies to be published was the paper by Harrison and Gilroy[43] in 1959. There are still a handful of new applications in modern day literature despite the controversy surrounding what conclusions may be drawn from the examination.[165,166] For example, there is little scientific need to perform a GSR study on a victim who has been shot. Similarly, some 0.22 caliber ammunition does not contain antimony in the primer, which is a reality that the forensic scientist must be aware of in order to properly interpret the data from this test.[10]

While early publications detailed appropriate wet chemical methods for GSR detection, many subsequent articles were published beginning in the 1970s, which focused on the promise of new or combined techniques for this analysis.[167–176] Drawbacks to these methods include

NAA's inability to detect lead and flame AA's lack of sufficient sensitivity for barium and antimony. However, flameless AA applications were quite promising for the advancement of GSR studies involving all three elements of interest. In 1987, three papers were published that demonstrated the popularity of GSR research by a variety of techniques.[168–170]

The first paper described a procedure whereby gunpowder stabilizers were analyzed by high-performance liquid chromatography for their organic content followed by GFAAS for the characteristic presence of barium, antimony, and lead. In a typical forensic laboratory, it would seem as if this dual approach could be time-consuming and costly. Moreover, the results were not necessarily more compelling than AA measurements alone, given the extra time involved as weighted against information gained from the double analysis.[168]

The second paper investigated the use of Auger electron spectroscopy as a substitution technique for energy dispersive X-ray analysis (EDX).[169] The authors described this alternative as a means of depth profiling the layers of particles of interest. Better lateral resolution and peak differentiation were also cited as advantages over EDX measurements. However, this argument does not seem compelling in light of the number of scanning electron microscopy/energy dispersive X-ray analysis (SEM/EDX) papers that continue to appear in the literature. Comparatively speaking, Auger electron spectroscopy does not decrease the time constraints of the examination, instrumentation costs are higher than for SEM/EDX, and a multitude of problems were quickly associated with the carbon planchet collection stubs when compared to those used for EDX.

The third article that was published in 1987 made an argument for the use of SEM/EDX exclusively for GSR analyses.[170] The opening paragraph acknowledged that the current method of choice for GSR was a bulk analytical approach in which NAA (for barium and antimony) and AA (for lead) were used in concert. In fact, this procedure was employed by the FBI Laboratory at the time of the publication. The paper argued the merits of defining the morphology of a particle as being consistent with GSR as opposed to using AA to quantitate elements that were readily found in nature. Indeed this rationale has been adopted by many laboratories so that the majority of modern GSR analyses are performed almost exclusively by SEM/EDX.

Extensive studies have been performed in this area by the FBI Laboratory as well. As one of the premiere forensic laboratories in the world, there has never been a lack of funding or research talent to promote innovative scientific advances within the forensic community. Work by the FBI Laboratory in the area of bulk GSR examinations included a published hand-blank study from 1990,[174] in addition to previous research in 1989 that described the use of flameless AA spectroscopy and ICPAES using an updated swab extraction procedure.[172] Most recently, a 1998 publication details the protocol under investigation by the FBI for the use of ICPMS as a bulk screening technique prior to the use of SEM/EDX for particle identification.[166] With proper sample collection, including consideration of the timeframe for sample collection and requisite environmental conditions necessary for a valid analysis, ICPMS screening for the presence of GSR will offer significant time savings in the examination process. If ICPMS results produce cause for further study, SEM/EDX can be employed for verification of the presence of specific GSR particles.

The FBI Laboratory's ICPMS research protocol for GSR was developed on the basis of years of experiments using flameless AA sample collection, storage and analysis. Early models included publications by Triplett,[176] Gonzales,[177] and a study by Vandecasteele et al.[178] The latter article described elemental concentrations that were found to be present in air samples collected in an indoor shooting range. Such studies provide essential background information as to the levels of GSR that can be exhibited by persons in close proximity to a discharged firearm, even if the individual has not actively engaged in the discharge of a weapon. The 1998 paper by Koons[166] in the *Journal of Forensic Science* lists the solutions, method and instrumental conditions employed in this pilot study. GFAAS and ICPAES were used as control methods of analysis for this work with comparable results being realized from all three analytical techniques.

A further interest in the development of this protocol is currently two-fold. There is a desire to characterize other elements associated with specific primers or ammunitions. For example, future work may include the study of strontium in nontoxic primers.

Another publication of merit in the analysis of GSR by a variety of spectrochemical methods was published in 1997 by Meng and Caddy.[165] As a review article, it is a comprehensive analysis of the history of this examination, with information on the composition of primers and the myriad of techniques available for study of the residues produced. While the discussion leans toward an argument in favor of analysis of the organic components in GSR, it is nonetheless a comprehensive treatment of the many considerations that play a role in this examination.

4 METALS ANALYSIS

4.1 General Applications

Another area of analytical endeavor that has embraced the development and increasing availability of ICPAES

is the metals industry.[179] The ability to monitor and thus control elemental impurities in industrial alloys has made it possible to discriminate between these objects for forensic purposes as well.[180,181] Nonhomogeneous leaded brass (copper–zinc alloys) studies were published in 1985.[180] Sample sizes of 10–100 μg were found to be adequate for ICPAES analyses.[181] However, specimens of 20 mg in triplicate were used for a similar study involving common household aluminum foils. This material is of interest to forensic scientists due to its prevalence at crime scenes involving narcotics production or the assemblage of improvised explosive devices. The elements of interest in this initial study were Cu, Fe, Ga, Mg, Mn, Ni, Si, Ti, V, and Zn. Analytical precision was determined to be about 1–2% RSD for most of the studied elements.

Further work on aluminum alloy trace element patterns has been reported using ICPMS and glow discharge mass spectrometry by Feng and Horlick.[182] The ion source for the glow discharge measurements was a substitution of the plasma sample introduction system found on conventional ICPMS instruments. While this approach is still more research oriented than applied, using ICPMS for this examination is noteworthy. Moreover, the article provides good background for scientists interested in pursuing laser ablation studies of solid metal samples. The advantages and disadvantages of sample dissolution ICPMS measurements are described in relation to laser ablation analyses.

Since laser ablation is likely to remain uncommon in forensic laboratories for some time, a 1998 paper by Coedo et al.[183] is helpful for its discussion of boric acid as an aid in fluoride detection in steels via ICPMS. The dissolution involves microwave digestion of the specimens for rare earth analysis using flow injection for sample introduction into the plasma. As has been demonstrated for years by archaeologists using NAA and more recently, ICPAES and ICPMS for trace analysis, the rare earth elements offer exceptionally consistent concentration patterns. For this reason, they have served historically as unblemished geographic markers for sourcing production sites.

A feature of flow injection analysis that may increase interest in the forensic community is the small sample volume needed for analysis. Load volumes are typically of the order of microliter quantities. Moreover, the limited exposure of the sampler and skimmer cones, in the ICPMS instrument, to the highly concentrated salt solutions that typify such matrices improves precision measurements because a steady state signal is readily achieved. As a routine application, the technique developed in this study is relegated to the distant future. However, knowledge of its existence is of benefit to front-line forensic laboratories where new challenges are often encountered.

In a 1993 article, Wynn[184] discussed a dissolution procedure for tin–lead solders using a combination of nitric and hydrochloric acids for detection of both trace and major elements using inductively coupled plasma optical emission spectroscopy. The procedure utilized a 1-g sample size, which was digested by the acid matrix on an open hotplate in about 15 min. Calibration standards were comprised of purified solid tin and lead standards, digested in the acid media for matrix matching purposes. Similarly treated SRMs from NIST were used to verify the standard solution concentrations.

Differences between the obtained ICP results and the NIST certified values averaged about 10%. This work is of particular interest in this forensic laboratory because of the frequently encountered domestic terrorism bombing cases in which solder features in the construction of improvised explosive devices. If signature trace element patterns could be developed to source solder back to a unique origin, this information would serve as a powerful investigative tool.

Scientists at the University of Pretoria published a 1996 article in the *Journal of Forensic Science* describing the use of ICPAES for the study of trace element compositional differences in mild steel.[185] This material is an iron-based alloy containing manganese and carbon in quantities that are difficult to detect and quantify using other methods. The authors explain that mild steel is quite prevalent as a construction material for storage objects such as safes. A common mode of entry into these structures is to cut through the container with an oxygen–acetylene torch. While X-ray diffraction can distinguish the use of a torch from an abrasive cutting tool via oxidation effects, it is not as sensitive to trace levels of impurities in the steel.

This paper further describes the process whereby ICPAES can be used to relate metallic debris found on a suspect to the cut object at a crime scene. First row transition metals were analyzed in two rods of mild steel as a test case. Four specimen aliquots were assayed from a total drilled sample weight of 180 mg. Specimens were digested in a hydrochloric acid–deionized water (40 : 30 v/v) matrix. The signal from each analyte of interest was ratioed to the intensity of manganese, as it is the most abundant alloying metal present in mild steel. It was concluded that silicon and aluminum could not be reliably distinguished using this method of analysis; however, chromium, nickel, cadmium, zinc, and copper in cutting debris could be sourced back to the parent sample when expressed relative to the intensity of manganese in the specimen.

4.2 Precious Metals

The use of ICPMS to study natural silver artifacts was reported by Longerich et al.[186] in 1987. An argument

is made for the use of ICPMS as a sensitive analytical tool that yields useful trace element compositional information on small, often nonhomogeneous specimens. Twelve elements were studied via ICPMS with comparison data obtained for cobalt, nickel, copper, and zinc using flame AA spectroscopy. While rare earth elements were not targeted in this study, lanthanum contamination was observed as a result of its ubiquity in the given laboratory environment. Moreover, two silver–argon species ($^{107}Ag^{40}Ar^{+}$ and $^{109}Ag^{40}Ar^{+}$) were identified at masses 147 and 149, respectively. These two isotopes would be of interest if samarium were to be included in the study. The absence of other rare earth elements such as cerium, neodymium, and europium supported the discussed conclusions regarding the observed interfering intensities at masses associated with lanthanum and samarium.

Memory effects were also discussed in this paper. High acid content (0.8 M nitric acid versus 0.2 M) and a high dissolved solids ratio were cited as contributors to this problem. One resolution would be frequent cleaning or replacement of the sampler and skimmer cones that separate the air/vacuum interface between the ionization source and the detector. An alternative for stabilization of the signal prior to analysis would be to introduce a slightly more concentrated solution of a given matrix in order to achieve a steady state layer of solid material on the cone surfaces. This method would apply more specifically to a dedicated instrument or set of sampling and skimmer cones, where such luxuries are available.

Related articles that further develop the concept of ICPMS as a tool for the analysis of precious metals have since been published.[187] High purity metals (cadmium, copper, gallium, and zinc) were analyzed for 30 trace elements using high resolution ICPMS.[188] High resolution ICPMS utilizes a magnetic sector mass spectrometer such as that employed in thermal ionization and spark source mass spectrometry instruments. Conventional ICPMS instrumentation is equipped with a quadrupole mass analyzer. In high resolution ICPMS, a quadrupole is utilized as a mass filter prior to sample introduction to the magnetic sector spectrometer. This configuration can achieve mass resolution of the order of 0.1 of an atomic mass unit for certain elements. Mass discrimination in the presence of overlapping isotopes at this level of detection would serve the forensic community in applications such as compositionally sourcing wires found in lamp cords used to bind a victim, or in improvised detonation devices.

Precious metals such as platinum group elements and gold in geological materials have also been analyzed by ICPMS.[187] As an alternative to conventional fire assay methods, ICPMS offers a cleaner and safer sample digestion environment in addition to greater sensitivity in analytical determinations.[189] Detection of phosphorus in high purity iron (steel) specimens has also been reported via electrothermal vaporization ICPMS.[190]

The most recent alternative that has been developed for solution nebulization ICPMS where ultra trace analyses are required is the collision cell.[191] This reaction chamber may be either a quadrupole or hexapole that is placed before the conventional quadrupole detector in an ICPMS instrument. Its purpose is to allow for an inlet whereby a previously chosen gas will be introduced to the transported ions in order to form competing reactions with interferents of the analyte of interest. In this way, determinations such as that for calcium at mass 40 are possible when the potential argon interference is eliminated by its preferential bonding with the reaction gas.

Similarly, the mass filter portion of this chamber sets a window through which only the analyte of interest is able to pass into the conventional quadrupole mass separator under the specified conditions. The obvious first-generation users of such advanced technology would be the semiconductor industry in which high purity acids are required for production use.[192] Environmental and biological uses are also facilitated by the collision cell's ability to characterize, for example, chlorine, selenium, and lead isotope ratios.[193,194]

As stated earlier with regard to arsenic in hair analyses, chloride levels are not thought to be a significant analyte interferent for prewashed hair specimens. Therefore, such technology as the collision cell chamber would not be warranted for this examination on a routine basis. Moreover, the instrument design is such that only highly pure solutions may be analyzed in this manner, a requirement that perhaps disregards forensic specimens by definition. However, it is an advancement of note for potential future applications.[195]

5 GLASS

As with metals, glass is a material that has been studied from a multidisciplinary perspective. Its trace element profile is of interest in manufacturing, both from an historical and an industrial perspective.[196] Given the types of glass used to manufacture window panes in homes, offices, and automobiles, as well as, its use for headlamps, flashlights, and costume jewelry, compositional analysis is also of interest in the forensic community. As the manufacturing process is refined and tighter quality control limits are set, refractive index differences continue to diminish between batch lots of container and float glasses. In cases where only small fragments of glass are available for comparative analysis,

trace element patterns become critical for a relational assessment to be made between known and questioned specimens.[197]

Early work in this area was carried out in 1970 using NAA to distinguish between glass specimens from vehicles involved in a two-car collision. Based on the NAA results, glass transfer was found to have occurred between the victim's shattered Jeep windscreen and the carriage of the truck that hit it. The truck fled the scene and the Jeep's driver was killed in the collision. The truck's driver denied both knowledge of the incident and his truck's involvement in the accident. However, glass debris found on his vehicle compositionally matched glass specimens collected from the Jeep's windscreen, indicating that these two vehicles were in contact at some point.[198]

Glass analysis was subsequently discussed from the perspective of atomic spectrometry by Hughes et al.[199] in 1976. Both flame and furnace AA were used in their research for the determination of manganese and magnesium. These elements were selected as useful discriminators between the following selected glass types: sheet, container, tableware, and headlamp glass specimens. Given the technology that exists today, this procedure would serve as the most basic of elemental screening tests. However, the article does provide useful background information regarding the process of sample discrimination, digestion, and early sample homogeneity results for common glass types.[199]

This article was cited and expanded upon by Catterick and Wall[200] from the Metropolitan Police Forensic Science Laboratory (MPFSL) in the UK in a 1978 publication that included iron as an element of interest.[200] The described digestion was unique in that it substituted the heating of small fragments with a 30-min ultrasonic treatment. A combination of these two practices along with the use of both hydrochloric and hydrofluoric acids is now a generally accepted digestion procedure.

Research into trace element characterization of glass types using ICPAES continued at the MPFSL. As the method was refined and more elements were successfully added to the list of possible discriminators, interest shifted to what conclusions might be based upon trace element patterns in a given class of these specimens. Another scientist at the MPFSL, Hickman,[201] discussed the merits of glass classification in a 1983 article.

Hickman worked with archaeologists and long-time neutron activation researchers, Harbottle and Sayre, at Brookhaven National Laboratory in New York, to ascribe meaning to the voluminous data that had been collected to that point. Using ICPAES for Mn, Fe, Mg, Al, Sr, and Ba, NAA for Na, Cs, Sc, La, Ce, Eu, Th, Cr, Se, U, Ca, and Zn, flame emission for K, Rb, and Li, and AA spectroscopy for As, 15 colorless sheet glass samples were analyzed.

ICPAES was further utilized for the study of an additional 232 samples using Mn, Fe, Mg, Al, Sr, and Ba. Sample selection was represented by 10 narrow ranges of refractive index within the range of 1.5150–1.5210. Each of these 10 groups of sheet glass specimens were assessed for the overall discrimination capabilities of this six-element composition assay. The author concluded that the six-element compositional analysis coupled with refractive index measurements was a better discriminator than refractive index alone. However, better analytical precision of the ICPAES technique was needed to further improve discriminating ability between certain regions of refractive index.[201]

A 1987 article by the same author furthered this research into the use of multivariate statistical classification for ICPAES glass data.[202] An internal standard was added to this study in which both discriminant analysis and cluster analysis were used. Other models such as Bayesian statistics were also discussed as alternative classification treatments for this type of sample matrix and analytical method.

Also published in 1987, Koons et al.[203] described similar work that was carried out in the FBI Laboratory using pattern recognition with ICPAES for classification and discrimination of sheet and container glasses. A 1991 article further developed the chemical discriminating capabilities available for glass samples with the inclusion of energy dispersive X-ray fluorescence to research involving refractive index and ICPAES.[204] These studies have led to the exploration of ICPMS as an analytical tool for glass samples.[205–207]

The first article of note in this area was published in 1990.[205] It described the dramatic compositional differences that exist between glass specimens from the USA and Australia. Discrimination was defined between two samples as a measure of overlap for any element within a mean confidence of two or three standard deviations. If no such overlap occurred for the elements of interest, the samples were considered to be distinguishable or separate in composition. Such conclusions were based on mean concentrations obtained from five independent samples for both control and suspect glass specimens. Sample sizes as small as 500 μg were utilized, with the mean sample size ranging between 800 μg and 1200 μg. The concentrations of 48 elements were quantified in this study with RSD values of 4% or less.

Fifteen trace elements were identified as discriminators of float glass specimens, such that origin and further subset classification was possible for individual samples. It was concluded that Australian glasses are better characterized by ICPMS than ICPAES when a single technique is preferable, due to the presence of trace elements at nanogram per gram levels in the samples. However,

ICPAES can accomplish this same task in glasses manufactured in the USA, where trace elements are typically three orders of magnitude more concentrated in the matrix (microgram per gram quantities).[205]

Parouchais et al.[206] supported this conclusion in a 1996 paper in the *Journal of Forensic Sciences*. It was further stated that samples sizes of 100 μg were necessary to adequately analyze headlamp specimens, while double that amount was needed for container and sheet glass. Using conventional solution nebulization technology, the authors concluded that the minimum sample volume necessary for current ICPMS instrumentation would be of the order of 2 mL. However, a precision of 1–5% RSD would require a minimum sample volume of 5 mL for a 1 mg mL^{-1} glass solution.[206] It can often be a difficult task to achieve this sample size in forensic cases. Thus, solid sampling is of great interest, both as a microanalytical technique and for its preservation of specimen integrity. Specifically, the coupling of laser ablation sampling with the sensitivity of ICPMS detection has generated widespread attention in the pursuit of forensic applications.[207]

6 LASER ABLATION

Ten years after the commercial development and widespread availability of ICPMS instrumentation, its early proponents were actively seeking the means to expand its versatility. Dissolution requirements of conventional nebulization technology greatly detracted from the appeal of ICPMS. Therefore, nondestructive solid sampling methodologies were independently pursued in both academia and industry for semiquantitative and quantitative analysis of subpicogram per gram levels of trace elements in inorganic matrices of limited sample size.[208–216] From this work evolved the concept of directly introducing solid material to a plasma ionization source.

Lack of dissolution provides better sensitivity due to the absence of polyatomic species generated by the aqueous medium necessary for conventional nebulization. Specimen size requirements are significantly reduced with solid sampling; therefore, the sampling process can be virtually nondestructive to rare or evidential materials.

Basic explanations of the theory and evolution of laser ablation sampling as related to ICPMS may be found in these literature references. However, the popularity of the technique has resulted in its discussion in a recent textbook as well.[24] Most commercial systems employ an Nd : YAG laser at a wavelength setting of 1064 nm, which emits a pulsed signal of 500 mJ for sampling purposes. The laser may be rastered across a surface or used for depth penetration studies depending upon the user's intent.

Sample pretreatment usually involves the removal of surface contamination with the laser system prior to collection of quantitative samples. Determinations of the amount of material ablated with each laser pulse have also been extensively discussed in the literature. Differential weighing is most convenient with adequate weighing apparatus as long as sample positioning is not critical. However, transport efficiency is not factored into this equation and therefore must be assessed using some other criteria, such as internal standardization.[217]

An example of industrial interest in laser ablation was demonstrated when semiconductor companies sought to quantitate the level of actinides in aluminum used to create integrated circuitry. As a quality control issue, this research was given a high priority, so that LA/ICPMS was aggressively studied and found to be ideally suited for this task.[218–221] For forensic purposes, this research serves as a basis for LA/ICPMS studies of a variety of metal objects used in the commission of a crime.[222]

Simultaneously, academic researchers became interested in laser ablation as a geochemical sampling tool for the study of minerals. Geological studies by Jarvis and Williams and others serve as good examples of early precious metal applications.[223–225] Subsequent studies of LA/ICPMS by Watling et al.[226] have yielded trace element characterization of diamonds and their indicator minerals, namely garnets and chromites. From this work and similar research, forensic interest was generated in the prospect of analyzing evidential glass fragments a priori.[227–233]

Early studies discussed the merits of ablating a solid material that had been added to a flux mixture and fused to a glass bead.[217] Particular attention was paid to the use of a matrix element as an internal standard, which is a convenient calibration strategy in homogeneous specimens. The application of this method to the study of rare earth elements in geological materials was further discussed, with detection limits reported in the range of 0.02–3 μg g^{-1}. The significant reduction of polyatomic interferences as compared to solution nebulization was also noted.[217]

Further interest in the use of LA/ICPMS for forensic casework has since been generated by the continuation of promising work in this area from researchers at the Chemistry Centre of Western Australia.[234] This group first reported on the use of LA/ICPMS for trace element determinations in cannabis plants as detailed in an earlier section. That work developed from promising results for similar studies using steel and glass specimens found at crime scenes. These analyses were reported on in the *Journal of Analytical Atomic Spectrometry* in early

1997.[234] The laser system used in this work was focused to a spot size of 20 μm with sample sizes of the order of 50 μm in diameter. Sixty-two glass specimens consisting of 31 float glasses, 4 sheet glasses, and 27 container glasses were analyzed. Steel samples were obtained from 69 sources such as safes, firearm barrels, tools, angle iron, rods, and crowbars.

With respect to the glass samples, the authors' argument against dissolution noted the necessity of hydrofluoric acid in order to digest the silicate matrix. Formation of insoluble rare earth fluorides can greatly affect quantitation of these important fingerprint elements. Thus, the digestion process can be quite lengthy in order to assure removal of excess hydrofluoric acid after the initial matrix breakdown.

According to these authors, the LA/ICPMS process for complex forensic matrices is not yet completely quantitative. However, it is still ideal for these types of specimens because virtually all forensic work is comparative, not absolute in its mission. The question most often asked is: ''Do the known and suspect samples match?'', rather than: ''What is it?'', or ''How much is there?'' Thus, for the comparison of small, solid matrices, LA/ICPMS is fast becoming the survey technique of choice.

In the described paper, sample preparation resembled that used for SEM or EDX work, in which the specimen (average size equal to 15 mm in diameter) was embedded in a plastic block of 25 mm diameter and 10 mm thickness. The surface of the block was then polished to expose the specimen. Alternatively, cyanoacrylate was used to embed smaller samples onto the surface of the plastic blocks.[234]

The authors further explain that the laser signal optimization was achieved for both sample types using the SRM 610 for Trace Elements in Glass (NIST, Gaithersburg, MD). This particular standard contains a high concentration of the mid-mass rare earth element lanthanum (^{139}La), which makes it ideal for optimizing both the laser and the ICPMS systems. Further, this glass reference material was used as a standard for both specimens for its production of a clean steady state signal. Continuous ablation with a metal standard could coat the laser cell and contaminate the system over time. It should also be noted that concerns regarding the uniformity of ablation of differing materials which could result in variation of the plasma load function is not at issue in this work because of the comparative nature of the study. Associations between samples of the same type will be based on their relative concentrations, not absolute quantitation of each specimen with respect to a standard or reference material.[234]

Lastly, in the Apparatus section of the article by Watling et al.,[234] it is stated that the laser can be operated in either the Q-fixed (free running) mode or the Q-switched mode. Steel samples were ablated using the laser in the Q-fixed pulse mode. This setting works well for thermally conducting samples as the energy of the laser radiation is absorbed by the surface, thereby creating deep narrow craters. One to five shots were fired on each site in quick succession rather than over a timed interval.[234]

Nonconducting specimens, such as glass, were analyzed using the Q-switched pulse mode, whereby the laser energy generates an ionizing plasma above the focal point on the sample. The authors further explain that sampling is achieved in this mode as the laser-induced plasma couples with the surface of the specimen and cuts into it. Research demonstrated to the authors that analysis of the varying glass types in this study could be achieved with a laser energy of 750 V at a rate of 15 Hz over a 60-s interval.[234]

Results of this extensive study demonstrated that reproducibility was readily achieved on a day-to-day basis. Similarly, differentiation was possible on five colorless glass samples with relative refractive index differences of less than 0.1%. Most promising for all aspects of lot production comparisons was the work that demonstrated that brown glass bottles manufactured at the same plant on different days could also be distinguished on the basis of ternary plots of the replicate data. Results for replicate analyses of steel samples showed similar results. Over time, the precision of the analytical technique showed slight variations that in no way compared with the intersample variability demonstrated by repeated sampling of five different specimens of the same type.[234]

The development of methodologies such as those demonstrated in this work would obviously require expertise, time, and large specimen banks to begin the evolution of a database for historical comparability. However, once the initial protocols were developed and the software capabilities in place for data storage and access, it is conceivable that training of less skilled or inexperienced technicians for routine use of this technology would be easily achieved. That prospect makes the potential of LA/ICPMS highly promising for routinely analyzed forensic specimens such as bullet lead, glass, steel, copper wire, soils, paints, polymers, and fibers.

The description of each of the techniques mentioned in this article is by no means complete. With regard to forensic applications, the methodologies are not routine in some instances; nor has the need for further exploration of useful methods been exhausted. There will continue to be a need for better detection limits for smaller specimens with less alteration to the integrity of the evidence. As atomic spectrometric techniques become less intrusive through the advent of sample introduction capabilities such as laser ablation, the ability to justify the need

for such equipment in modern forensic laboratories will become easier. These methods have the advantage of preserving the sample and minimizing the preparation time, both of which are essential criteria for meeting the needs of the criminal justice system.

An inexpensive and continually evolving mechanism for forensic laboratories to stay current with modern trends in atomic spectrometry is through the use of the internet. The July/August 1998 issue of *Today's Chemist at Work* featured a list of common internet search engine sites that can be used to search for information regarding all aspects of scientific problem solving.[235] Common spectroscopy and spectrometry newsgroups such as *sci.techniques.spectroscopy* and *sci.techniques.mass-spec* were listed, along with World Wide Web sites (Deja News) that do not require individual subscriptions for access. These features, as well as on-line journals and the ease of international communication via e-mail, have increased the forensic community's ability to stay abreast of research pursuits on a scale that was not previously possible. As a unifying instrument for international communications networking between cooperating forensic and scientific organizations, the internet is developing into one of the most utilized and cost efficient sources of information exchange available to the global forensic community.

ACKNOWLEDGMENTS

The authors wish to thank their colleagues in the Elemental Analysis Operations Group of the Materials Analysis Unit for their helpful comments and suggestions. Also, the staff of the Forensic Science Information and Resource Center and the Automation Unit provided immense support, always in a friendly and timely manner.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
CTD	Charge Transfer Device
DCP	Direct Current Plasma
EDL	Electrodeless Discharge Lamp
EDX	Energy Dispersive X-ray Analysis
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
GHB	γ-Hydroxybutyrate
GSR	Gunshot Residue
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
INAA	Instrumental Neutron Activation Analysis
LA/ICPMS	Laser Ablation/Inductively Coupled Plasma Mass Spectrometry
MPFSL	Metropolitan Police Forensic Science Laboratory
NAA	Neutron Activation Analysis
NIES	National Institute for Environmental Studies
NIST	National Institute of Standards and Technology
PMT	Photomultiplier Tube
RSD	Relative Standard Deviation
SEM	Scanning Electron Microscopy
SEM/EDX	Scanning Electron Microscopy/Energy Dispersive X-ray Analysis
SRM	Standard Reference Material

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Drugs of Abuse, Analysis of

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Scanning Electron Microscopy in Forensic Science

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction • Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis • Metal Analysis, Sampling and Sample Preparation in

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry •

Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Ablation in Atomic Spectroscopy • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Nuclear Methods **(Volume 14)**
Elemental Analysis by Isotope Dilution • Instrumental Neutron Activation Analysis

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Archaeological Chemical Analysis • Literature Searching Methodology • Microwave Techniques • Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration • Traceability in Analytical Chemistry

REFERENCES

1. F.S. Baechtel, 'The Identification and Individualization of Semen Stains', in *Forensic Science Handbook*, ed. R. Saferstein, Prentice Hall, New Jersey, Vol. II, 1988.
2. B. Robertson, G. Vignaux (eds.), 'Transfer Evidence', in *Interpreting Evidence: Evaluating Forensic Evidence in the Courtroom*, Wiley, New York, 135–154, 1995.
3. V.P. Guinn, 'Applications of Neutron Activation Analysis in Scientific Crime Detection', GA-8013, Division of Technical Information, US Atomic Energy Commission, 1967.
4. J.B. Weinstein, 'Enhancing the Relationship of Science and the Courts', *J. Forensic Sci.*, **43**, 242–245 (1998).
5. P.M. Grant, K.J. Moody, I.D. Hutcheon, D.L. Phinney, J.S. Haas, A.M. Volpe, J.J. Oldani, R.E. Whipple, N. Stoyer, A. Alcaraz, J.E. Andrews, R.E. Russo, G.L. Klunder, B.D. Andersen, S. Cantlin, 'Forensic Analyses of Suspect Illicit Nuclear Material', *J. Forensic Sci.*, **43**, 680–688 (1998).
6. R.C. McFarland, M.E. McLain, 'Rapid Neutron Activation Analysis for Gunshot Residue Determination', *J. Forensic Sci.*, **18**, 226–231 (1973).
7. K.K. Pillay, W.A. Jester, H.A. Fox, 'New Method for the Collection and Analysis of Gunshot Residues as Forensic Evidence', *J. Forensic Sci.*, **19**, 768–783 (1974).
8. S. Felby, 'Determination of Organosilicon Oxide Polymers in Tissue by Atomic Absorption Spectroscopy Using HGA Graphite Furnace, *Forensic Sci. Int.*, **32**, 61–65 (1986).
9. E.C. Blacklock, P.A. Sadler, 'A Rapid Screening Method for Heavy Metals in Biological Materials by Emission Spectroscopy', *Clin. Chim. Acta*, **113**, 87–94 (1981).
10. A.L. Green, J.P. Sauve, 'The Analysis of Gunshot Residue by Atomic Absorption Spectrophotometry', *At. Absorpt. Newsl.*, **11**, 93–95 (1972).
11. S.S. Krishnan, 'Firing Distance Determination by Atomic Absorption Spectrophotometry', *J. Forensic Sci.*, **18**, 351–356 (1973).
12. G.D. Renshaw, C.A. Pounds, E.F. Pearson, 'The Quantitative Estimation of Lead, Antimony, and Barium in Gunshot Residues by Non-flame Atomic Absorption Spectrophotometry', *At. Absorpt. Newsl.*, **12**, 55–56 (1973).
13. R.D. Cone, 'Detection of Barium, Antimony, and Lead in Gunshot Residue by Flameless Atomic Absorption Spectrophotometry', *Police Weapons Center Bull.*, **6**, 4–6 (1973).
14. J.A. Goleb, C.R. Midkiff, 'Barium Contamination in Gunshot Residue Collection Kits', *Appl. Spectrosc.*, **28**, 382–383 (1974).
15. I.C. Stone, C.S. Petty, 'Examination of Gunshot Residues', *J. Forensic Sci.*, **19**, 784–788 (1974).
16. J.A. Goleb, C.R. Midkiff, 'The Determination of Barium and Antimony in Gunshot Residue by Flameless Atomic Absorption Spectroscopy Using a Tantalum Strip Atomizer', *Appl. Spectrosc.*, **29**, 44–48 (1975).
17. M.L. Newbury, 'The Analysis of Gunshot Residue for Antimony and Barium by Flameless Atomic Absorption Spectrometry', *Can. Soc. Forensic Sci. J.*, **13**, 19–26 (1980).
18. J.T. Newton, 'Rapid Determination of Antimony, Barium, and Lead in Gunshot Residue Via Automated Atomic Absorption Spectrophotometry', *J. Forensic Sci.*, **26**, 302–312 (1981).
19. L.C. Portis, R.K. Tilley, 'A Filtration Technique for Gunshot Residue Analysis by SEM and/or FAAS', *Assoc. Firearms Toolmarks Examiners*, **13**, 93–99 (1981).
20. S.S. Krishnan, K.A. Gillespie, E.J. Anderson, 'Rapid Detection of Firearm Discharge Residues by Atomic Absorption and Neutron Activation Analysis', *J. Forensic Sci.*, **16**, 144–151 (1971).
21. W.D. Kinard, D.R. Lundy, 'A Comparison of Neutron Activation Analysis and Atomic Absorption Spectroscopy on Gunshot Residue', *ACS Symp. Ser.*, **13**, 97–107 (1975).
22. R.D. Beaty, J.D. Kerber, 'Concepts, Instrumentation, and Techniques', in *Atomic Absorption Spectrophotometry*, The Perkin-Elmer Corporation, Norwalk, 1993.
23. C.B. Boss, K.J. Fredeen, 'Concepts, Instrumentation, and Techniques', in *Inductively Coupled Plasma Optical Emission Spectrometry*, The Perkin-Elmer Corporation, Norwalk, 1997.
24. A. Montaser, *Inductively Coupled Plasma Mass Spectrometry*, Wiley-VCH, New York, 1998.
25. J.D. Kerber, 'The HGA Graphite Furnace Increases the Scope of Atomic Absorption Spectroscopy', *Instrum. News*, **24**, 2–4 (1974).
26. J. Chen, H. Berndt, G. Tolg, 'A Contribution to the Trace Analysis of High-purity Iron by Atomic Spectrometry (Flame-AAS, Furnace-AAS, ICP/OES)', *Fresenius' J. Anal. Chem.*, **344**, 526–534 (1992).

27. T.J. Gluodenis, Jr, J.F. Tyson, 'Flow Injection Systems for Directly Coupling On-line Digestions with Analytical Atomic Spectrometry. Part 1: Dissolution of Cocoa under Stopped-flow, High-pressure Conditions', *J. Anal. At. Spectrom.*, **7**, 301–306 (1992).
28. M.S. Cresser, 'Atomic Spectrometry in Environmental Analysis: Jam Today or Jam Tomorrow?', *J. Anal. At. Spectrom.*, **8**, 269–272 (1993).
29. D.J. Halls, 'Analytical Minimalism Applied to the Determination of Trace Elements by Atomic Spectrometry', *J. Anal. At. Spectrom.*, **10**, 169–175 (1995).
30. H. van der Jagt, P.J. Stuyfzand, 'Methods for Trace Element Analysis in Surface Water, Atomic Spectrometry in Particular', *Fresenius' J. Anal. Chem.*, **354**, 32–40 (1996).
31. M.E. Moreno, C. Perez-Conde, C. Camara, 'Sensitization of Stibine Generation: Antimony Determination in Environmental Samples by Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1181–1187 (1998).
32. E. Ivanova, W. Van Mol, F. Adams, 'Electrothermal Atomic Absorption Spectrometric Determination of Cadmium and Lead in Blood Using Flow Injection On-line Sorption Preconcentration in a Knotted Reactor', *Spectrochim. Acta, Part B*, **53**, 1041–1048 (1998).
33. G. Norheim, P.E. Paus, 'Arsenic Determination in Autopsy Material Using Atomic Absorption Spectroscopy', *Z. Rechtsmed.*, **75**, 127–130 (1974).
34. S.P. Singhal, 'Atomic Absorption Spectroscopy in Forensic Chemistry', *J. Can. Soc. Forensic Sci.*, **7**, 7–17 (1974).
35. G.W. Walker, 'Atomic Absorption Spectroscopy', *Chem. Br.*, **11**, 440–441 (1975).
36. M. Ravreby, 'Analysis of Long-range Bullet Entrance Holes by Atomic Absorption Spectrophotometry and Scanning Electron Microscopy', *Israel J. Forensic Sci.*, **27**, 92–112 (1982).
37. W. Slavin, 'A Comparison of Atomic Spectroscopic Analytical Techniques', *Spectroscopy*, **6**, 16–21 (1991).
38. D.M. Grant (ed.), 'Gunshot Residue Testing Laboratories in the United States, Unpublished List', revised by the FBI Laboratory, June, 1998.
39. C.A. Peters, Forensic Examiner, personal communication, 1998.
40. W. Slavin, 'Atomic Absorption Spectroscopy: Why Has It Become Successful?', *Anal. Chem.*, **63**, 1033A–1038A (1991).
41. R. Knight, S.J. Haswell, S.W. Lindow, J. Batty, 'Determination of Mercury in Hair by Coupled CVAA/ICPMS', *J. Anal. At. Spectrom.*, **14**, 127–129 (1999).
42. J.H. Sherfinski, 'A Graphite Tube Degradation Study of Barium at Gunshot Residue Levels', *Atom. Abs. Newsletter*, **14**, 26–28 (1975).
43. H.C. Harrison, R. Gilroy, 'Firearms Discharge Residue', *J. Forensic Sci.*, **4**, 184–199 (1959).
44. T.A. Brettell, R. Saferstein, 'Forensic Science', *Anal. Chem.*, **69**, 123R–143R (1997).
45. 'Atomic Spectrometry Updated References', *J. Anal. At. Spectrom.*, **9**, 135R–146R (1994).
46. R.C. Carpenter, 'The Analysis of Some Evidential Materials by Inductively Coupled Plasma Optical Emission Spectrometry', *Forensic Sci. Int.*, **27**, 157–163 (1985).
47. R. Koons, 'ICP Atomic Spectrometry in the Forensic Laboratory', *Spectroscopy*, **8**, 16–21 (1993).
48. M. Lalchev, I. Ionov, N. Daskalova, 'Application of Inductively Coupled Plasma Atomic Emission Spectrometry in Forensic Science', *J. Anal. At. Spectrom.*, **12**, 21–24 (1997).
49. J.B. Dawson, R.D. Snook, W.J. Price, 'Background and Background Correction in Analytical Atomic Spectrometry Part 1: Emission Spectrometry. A Tutorial Review', *J. Anal. At. Spectrom.*, **8**, 517–537 (1993).
50. B.L. Sharp, S. Chenery, R. Jowitt, S.T. Sparkes, A. Fisher, 'Atomic Spectrometry Update – Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **9**, 171R–188R (1994).
51. J. Marshall, A. Fisher, S. Chenery, S.T. Sparkes, 'Atomic Spectrometry Update – Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **11**, 213R–238R (1996).
52. J. Marshall, E.H. Evans, A. Fisher, S. Chenery, 'Atomic Spectrometry Update – Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **12**, 263R–290R (1997).
53. K.W. Jackson, S. Lu, 'Atomic Absorption, Atomic Emission, and Flame Emission Spectrometry', *Anal. Chem.*, **70**, 363R–383R (1998).
54. S.J. Hill, J.B. Dawson, W.J. Price, I.L. Shuttler, C.M.M. Smith, J.F. Tyson, 'Atomic Spectrometry Update – Advances in Atomic Absorption and Fluorescence Spectrometry and Related Techniques', *J. Anal. At. Spectrom.*, **12**, 327R–379R (1997).
55. J.R. Bacon, J.S. Crain, A.W. McMahon, J.G. Williams, 'Atomic Spectrometry Update – Atomic Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 407R–448R (1997).
56. J.R. Bacon, J.S. Crain, L. Van Vaeck, J.G. Williams, 'Atomic Spectrometry Update – Atomic Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 171R–208R (1998).
57. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **9**, 87R–134R (1994).
58. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **10**, 61R–112R (1995).
59. A. Taylor, S. Branch, H.M. Crews, D.J. Halls, L.M.W. Owen, M. White, 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **12**, 119R–221R (1997).
60. A. Taylor, S. Branch, D.J. Halls, L.M.W. Owen, M. White, 'Atomic Spectrometry Update – Clinical and

Biological Materials, Food and Beverages', *J. Anal. At. Spectrom.*, **13**, 57R–106R (1998).
61. M.S. Cresser, J. Armstrong, J. Cook, J.R. Dean, P. Watkins, M. Cave, 'Atomic Spectrometry Update – Environmental Analysis', *J. Anal. At. Spectrom.*, **8**, 1R–78R (1993).
62. J.R. Dean, L.M. Garden, J. Armstrong, M.S. Cresser, M. Cave, P. Watkins, 'Atomic Spectrometry Update – Environmental Analysis', *J. Anal. At. Spectrom.*, **12**, 19R–87R (1997).
63. J.R. Dean, O. Butler, A. Fisher, L.M. Garden, M.S. Cresser, P. Watkins, M. Cave, 'Atomic Spectrometry Update – Environmental Analysis', *J. Anal. At. Spectrom.*, **13**, 1R–56R (1998).
64. R.E. Sturgeon, 'Future of Atomic Spectrometry for Environmental Analysis', *J. Anal. At. Spectrom.*, **13**, 351–361 (1998).
65. J. Marshall, J. Carroll, J.S. Crighton, 'Atomic Spectrometry Update – Industrial Analysis: Metals, Chemicals, and Advanced Materials', *J. Anal. At. Spectrom.*, **6**, 283R–339R (1991).
66. J. Marshall, J. Carroll, J.S. Crighton, C.L.R. Barnard, 'Atomic Spectrometry Update – Industrial Analysis: Metals, Chemicals, and Advanced Materials', *J. Anal. At. Spectrom.*, **7**, 349R–411R (1992).
67. J. Marshall, J. Carroll, J.S. Crighton, C.L.R. Barnard, 'Atomic Spectrometry Update – Industrial Analysis: Metals, Chemicals, and Advanced Materials', *J. Anal. At. Spectrom.*, **9**, 319R–356R (1994).
68. J.S. Crighton, J. Carroll, B. Fairman, J. Haines, M. Hinds, 'Atomic Spectrometry Update – Industrial Analysis: Metals, Chemicals, and Advanced Materials', *J. Anal. At. Spectrom.*, **11**, 461R–508R (1996).
69. J.S. Crighton, B. Fairman, J. Haines, M. Hinds, S.M. Nelms, D.M. Penny, 'Atomic Spectrometry Update – Industrial Analysis: Metals, Chemicals, and Advanced Materials', *J. Anal. At. Spectrom.*, **12**, 509R–542R (1997).
70. A.M. Paudyn, R.G. Smith, 'Determination of Elements in Paints and Paint Scrapings by Inductively Coupled Plasma Atomic Emission Spectrometry Using Microwave Assisted Digestion', *Fresenius' J. Anal. Chem.*, **345**, 695–700 (1993).
71. J. Zieba-Palus, 'Examination of Used Motor Oils by Flame AAS for Criminalistic Purposes: A Diagnostic Study', *Forensic Sci. Int.*, **91**, 171–179 (1998).
72. M.A. Ryan, 'Finding Food Imposters: It's Not Just Peanuts', *Today's Chemist at Work*, **7**, 63–66 (1998).
73. G. Gugliotta, 'High-tech Trade Enforcement Tracks Peanuts Across Borders', *Washington Post*, Federal Page (4 December, 1997).
74. R. Schwartz, 'ICPMS Determination of Trace Element Composition of Agricultural Products for Geographic Origin Determination', 1998 Winter Conference on Plasma Spectrometry, Scottsdale, AZ, 5–10 January, Poster FP01, 1998.
75. I. Karadjova, G. Zachariadis, G. Boskou, J. Stratis, Electrothermal Atomic Absorption Spectrometric Determination of Aluminum, Cadmium, Chromium, Copper, Iron, Manganese, Nickel, and Lead in Olive Oil', *J. Anal. At. Spectrom.*, **13**, 201–204 (1998).
76. F. Dolinsek, J. Stupar, V. Vrscaj, 'Direct Determination of Cadmium and Lead in Geological and Plant Materials by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**, 653–660 (1991).
77. T. Ozaki, S. Enomoto, Y. Minai, S. Ambe, F. Ambe, T. Tominaga, 'Determination of Lanthanides and Other Trace Elements in Ferns by Instrumental Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **217**, 117–124 (1997).
78. H. Zhou, J. Liu, 'The Determination of Rare Earth Elements in Plant Foods by ICPMS', *At. Spectrosc.*, **18**, 192–194 (1997).
79. J.D. Greenough, H.P. Longerich, S.E. Jackson, 'Element Fingerprinting of Okanagan Valley Wines Using ICPMS: Relationships Between Wine Composition, Vineyard, and Wine Color', *Aust. J. Grape Wine Res.*, **3**, 75–83 (1997).
80. M.Y. Perez-Jordan, J. Soldevila, A. Salvador, A. Pastor, M. de la Guardia, 'Inductively Coupled Plasma Mass Spectrometry Analysis of Wines', *J. Anal. At. Spectrom.*, **13**, 33–39 (1998).
81. M. Segura, Y. Madrid, C. Camara, 'Evaluation of Atomic Fluorescence and Atomic Absorption Spectrometric Techniques for the Determination of Arsenic in Wine and Beer by Direct Hydride Generation Sample Introduction', *J. Anal. At. Spectrom.*, **14**, 131–135 (1999).
82. J.A. Nobrega, Y. Gelinas, A. Krushevska, R.A. Barnes, 'Direct Determination of Major and Trace Elements in Milk by Inductively Coupled Plasma Atomic Emission and Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 1243–1246 (1997).
83. C. Barbante, T. Bellomi, G. Mezzadri, P. Cescon, G. Scarponi, C. Morel, S. Jay, K. Van de Velde, C. Ferrari, C.F. Boutron, 'Direct Determination of Heavy Metals at Picogram per Gram Levels in Greenland and Antarctic Snow by Double Focusing Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**, 925–931 (1997).
84. G. Yaprak, B. Uysal, 'Determination of Low Level ^{210}Pb in Tobacco', *J. Radioanal. Nucl. Chem.*, **229**, 153–155 (1998).
85. R.Y.P. Bhatia, M. Prabhakar, T.S. Rao, K. Raghunath, 'Determination of Chromium, Cobalt, Copper, and Zinc Contents in Tobacco Ash of Different Cultivars of Tobacco by Inductively Coupled Plasma Atomic Emission Spectrometry', *Indian J. Forensic Sci.*, **2**, 145–156 (1988).
86. M.O.A. Oladipo, O.O. Ajayi, S.B. Elegba, S.O. Alonge, S.A. Adeleye, 'The Determination of Minor and Trace Elements in Some Nigerian Cigarettes and Raw Tobacco Using Inductively Coupled Plasma Mass Spectrometry

(ICPMS)', *J. Environ. Sci. Health A*, **28**, 839–857 (1993).
87. C. Roux, 'The Analysis of Inorganic Additives in Polyester Fibers by Inductively Coupled Plasma Atomic Emission Spectroscopy', in *Proc. Meet. Int. Assoc. Forensic Sci.*, 216–224, Vol. 4, 1995.
88. J.R. Vogt, D.M. Hashii, 'Trace Element Distributions in Man-made Fibers', *J. Radioanal. Nucl. Chem.*, **15**, 355–366 (1973).
89. S. Suzuki, H. Tsuchihashi, K. Nakajima, A. Matsushita, T. Nagao, 'Analyses of Impurities in Methamphetamine by Inductively Coupled Plasma Mass Spectrometry and Ion Chromatography', *J. Chromatogr.*, **437**, 322–327 (1988).
90. T. Kishi, 'Analysis of Trace Elements in Methamphetamine Hydrochloride by Inductively Coupled Plasma Mass Spectrometry', *J. Res. Natl. Bur. Stand.*, **93**, 469–471 (1988).
91. K.A. Wolnik, D.T. Heitkemper, J.B. Crowe, B.S. Barnes, T.W. Brueggemeyer, 'Application of Inductively Coupled Plasma Atomic Emission and Mass Spectrometry to Forensic Analysis of Sodium Gamma Hydroxy Butyrate and Ephedrine Hydrochloride', *J. Anal. At. Spectrom.*, **10**, 177–181 (1995).
92. R.J. Wells, S.V. Skopec, R. Iavetz, J. Robertson, 'Trace Element Analysis of Heroin by ICPMS', *Chem. Australia*, **62**, 14–25 (1995).
93. P. Bermejo-Barrera, A. Moreda-Pineiro, J. Moreda-Pineiro, A. Bermejo-Barrera, 'Application of Rapid Electrothermal Atomic Absorption Spectrometric Methods to the Determination of Ag, Al, Cd, and Mn in Cocaine and Heroin Samples', *Fresenius' J. Anal. Chem.*, **358**, 844–847 (1997).
94. R.J. Watling, 'Sourcing the Provenance of Cannabis Crops Using Inter-element Association Patterns ''Fingerprinting'' and Laser Ablation ICPMS', 1998 Winter Conference on Plasma Spectrometry, Scottsdale, AZ, 5–10 January, W2, 1998.
95. R.J. Watling, 'Sourcing the Provenance of Cannabis Crops Using Inter-element Association Patterns ''Fingerprinting'' and Laser Ablation ICPMS', *J. Anal. At. Spectrom.*, **13**, 917–926 (1998).
96. H. Al-Shahristani, I.K. Al-Haddad, 'Mercury Content of Hair from Normal and Poisoned Persons', *J. Radioanal. Nucl. Chem.*, **15**, 59–70 (1973).
97. E.O. Espinoza, M. Mann, B. Bleasdell, S. DeKorte, M. Cox, 'Toxic Metals in Selected Traditional Chinese Medicinals', *J. Forensic Sci.*, **41**, 453–456 (1996).
98. M. Piette, B. Desmet, R. Dams, 'Determination of Strontium in Human Whole Blood by ICPAES', *Sci. Total Environ.*, **141**, 269–273 (1994).
99. P. Capota, M. Ropota, M. Constantin, V. Stoenescu, 'Determination of Microelements in Human Teeth Using the ICPAES Method', *Chem. Anal.*, **42**, 721–724 (1997).
100. N.K. Mottet, R.L. Body, 'Mercury Burden of Human Autopsy Organs and Tissues', *Arch. Environ. Health*, **29**, 18–24 (1974).
101. M.W. Hinds, 'Determination of Mercury in Gold Bullion by Flame and Graphite Furnace Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **53**, 1063–1068 (1998).
102. Z. De-Qiang, N. Zhe-Ming, S. Han-Wen, 'Stabilization of Organic and Inorganic Mercury in the Graphite Furnace with $(NH_4)_2PdCl_6$-$(NH_4)_3$ $RhCl_6$ as a Mixed Chemical Modifier', *Spectrochim. Acta, Part B*, **53**, 1049–1055 (1998).
103. P. Canada Rudner, J.M. Cano Pavon, F. Sanchez Rojas, A. Garcia de Torres, 'Use of Flow Injection Cold Vapor Generation and Preconcentration on Silica Functionalized with Methylthiosalicylate for the Determination of Mercury in Biological Samples and Seawater by Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1167–1171 (1998).
104. G.R. Boaventura, A.C. Barbosa, G.A. East, 'Multivessel System for Cold-vapor Mercury Generation: Determination of Mercury in Hair and Fish', *Biol. Trace Elem. Res.*, **60**, 153–161 (1997).
105. J.A. Moreton, H.T. Delves, 'Simple Direct Method for the Determination of Total Mercury Levels in Blood and Urine and Nitric Acid Digests of Fish by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 659–665 (1998).
106. J. Yoshinaga, N. Matsuo, H. Imai, M. Nakazawa, T. Suzuki, 'Application of Inductively Coupled Plasma Mass Spectrometry (ICPMS) to Multi-element Analysis of Human Organs', *Int. J. Environ. Anal. Chem.*, **41**, 27–38 (1990).
107. T.D.B. Lyon, G.S. Fell, K. McKay, R.D. Scott, 'Accuracy of Multi-element Analysis of Human Tissue Obtained at Autopsy Using Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **6**, 559–564 (1991).
108. S.F. Durant, 'Inductively Coupled Plasma Mass Spectrometry for Biological Analysis', *Trends Anal. Chem.*, **11**, 68–73 (1992).
109. T. Tanaka, K. Hara, A. Tanimoto, K. Kasai, T. Kira, N. Tanaka, T. Takayasu, 'Determination of Arsenic in Blood and Stomach Contents by Inductively Coupled Plasma/Mass Spectrometry (ICP/MS)', *Forensic Sci. Int.*, **81**, 43–50 (1996).
110. L. Moens, 'Applications of Mass Spectrometry in the Trace Element Analysis of Biological Materials', *Fresenius' J. Anal. Chem.*, **359**, 309–316 (1997).
111. C.J. Amarasiriwardena, N. Lupoli, V. Potula, S. Korrick, H. Hu, 'Determination of the Total Arsenic Concentration in Human Urine by Inductively Coupled Plasma Mass Spectrometry: A Comparison of the Accuracy of Three Analytical Methods', *Analyst*, **123**, 441–445 (1998).
112. C. Latkoczy, T. Prohaska, G. Stingeder, M. Teschler-Nicola, 'Strontium Isotope Ratio Measurements in

Prehistoric Human Bone Samples by Means of High-resolution Inductively Coupled Plasma Mass Spectrometry (HR/ICPMS)', *J. Anal. At. Spectrom.*, **13**, 561–566 (1998).
113. S. Granero, J.M. Llobet, M. Schuhmacher, J. Corbella, J.L. Domingo, 'Biological Monitoring of Environmental Pollution and Human Exposure to Metals in Tarragona, Spain. Part I. Levels in Hair of School Children', *Toxic Elem. Electrol.*, **15**, 39–43 (1998).
114. J. Zheng, W. Goessler, A. Geiszinger, W. Kosmus, B. Chen, G. Zhuang, K. Xu, G. Sui, 'Multi-element Determination in Earthworms with Instrumental Neutron Activation Analysis and Inductively Coupled Plasma Mass Spectrometry: A Comparison', *J. Radioanal. Nucl. Chem.*, **223**, 149–155 (1997).
115. B.S. Sheppard, J.A. Caruso, D.T. Heitkemper, K.A. Wolnik, 'Arsenic Speciation by Ion Chromatography with Inductively Coupled Plasma Mass Spectrometric Detection', *Analyst*, **117**, 971–975 (1992).
116. I.B.-A. Razagui, S.J. Haswell, 'The Determination of Mercury and Selenium in Maternal and Neonatal Scalp Hair by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Toxicol.*, **21**, 149–153 (1997).
117. J. Yoshinaga, M. Morita, 'Determination of Mercury in Biological and Environmental Samples by Inductively Coupled Plasma Mass Spectrometry with the Isotope Dilution Technique', *J. Anal. At. Spectrom.*, **12**, 417–420 (1997).
118. H.P. Chen, D.C. Paschal, D.T. Miller, J.C. Morrow, 'Determination of Total and Inorganic Mercury in Whole Blood by On-line Digestion with Flow Injection', *At. Spectrosc.*, **19**, 176–179 (1998).
119. S. Saverwyns, X. Zhang, F. Vanhaecke, R. Cornelis, L. Moens, R. Dams, 'Speciation of Six Arsenic Compounds Using High Performance Liquid Chromatography/Inductively Coupled Plasma Mass Spectrometry with Sample Introduction by Thermospray Nebulization', *J. Anal. At. Spectrom.*, **12**, 1047–1052 (1997).
120. N. Miekeley, M.O. Amato, 'Fast Hg Determination in Biological Samples by ICPMS Using Minitube Furnace Catalytic Combustion', *At. Spectrosc.*, **18**, 186–192 (1997).
121. J. Lintschinger, P. Schramel, A. Hatalak-Rauscher, I. Wendler, B. Michalke, 'A New Method for the Analysis of Arsenic Species in Urine by Using HPLC/ICPMS', *Fresenius' J. Anal. Chem.*, **362**, 313–318 (1998).
122. D. Karunasagar, J. Arunachalam, S. Gangadharan, 'Development of a ''Collect and Punch'' Cold Vapor Inductively Coupled Plasma Mass Spectrometric Method for the Direct Determination of Mercury at Nanograms per Liter Levels', *J. Anal. At. Spectrom.*, **13**, 679–682 (1998).
123. K. Lee, S. Jiang, H. Liu, 'Determination of Mercury in Urine by Electrothermal Vaporization Isotope Dilution Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1227–1231 (1998).
124. M.C. Webb, D.C. Paschal, W.C. Elliott, H.P. Hopkins, A.M. Ghazi, B.C. Ting, I. Romieu, 'ICPMS Determination of Lead Isotope Ratios in Whole Blood, Pottery, and Leaded Gasoline: Lead Sources in Mexico City', *At. Spectrosc.*, **19**, 156–163 (1998).
125. T.Y. Toribara, A.G. Muhs, 'Hair: A Keeper of History', *Arctic Anthropol.*, **21**, 99–108 (1984).
126. K. Bencze, 'What Contribution Can Be Made to Biological Monitoring by Hair Analysis? Part 1', *Fresenius' J. Anal. Chem.*, **337**, 867–876 (1990).
127. K. Bencze, 'What Contribution Can Be Made to Biological Monitoring by Hair Analysis? Part 2', *Fresenius' J. Anal. Chem.*, **338**, 58–61 (1990).
128. V. Valkovic, *Human Hair, Trace-element Levels*, CRC Press, Boca Raton, Vol. II, 1988.
129. G.C. Verghese, R. Kishore, V.P. Guinn, 'Differences in Trace-element Concentrations in Hair Between Males and Females', *J. Radioanal. Chem.*, **15**, 329–335 (1973).
130. V.P. Guinn, M. Gavrilas-Guinn, R. Demiralp, 'Measurement of Arsenic in Sectioned Hair Samples by Instrumental Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **179**, 365–368 (1994).
131. V.P. Guinn, R. Demiralp, 'Arsenic in Hair by INAA – A Major Recent Murder Case', *J. Radioanal. Nucl. Chem.*, **168**, 249–256 (1993).
132. A. Pazirandeh, A.H. Brati, M. Ghannadi Marageh, 'Determination of Arsenic in Hair Using Neutron Activation', *Appl. Radiat. Isot.*, **49**, 753–759 (1998).
133. M. Saiki, M.B.A. Vasconcellos, L.J. de Arauz, R. Fulfaro, 'Determination of Trace Elements in Human Head Hair by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **236**, 25–28 (1998).
134. S.Y. Cho, S.G. Jang, Y.S. Chung, 'Human Hair Identification by Instrumental Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **229**, 143–147 (1998).
135. H. Hagedorn-Gotz, G. Kuppers, M. Stoeppler, 'On Nickel Contents in Urine and Hair in a Case of Exposure to Nickel Carbonyl', *Arch. Toxicol.*, **38**, 275–285 (1977).
136. F. Diaz-Barriga, M.A. Santos, J. de Jesus Mejia, L. Batres, L. Yanez, L. Carrizales, E. Vera, L.M. del Razo, M.E. Cebrian, 'Arsenic and Cadmium Exposure in Children Living Near a Smelter Complex in San Luis Potosi, Mexico', *Environ. Res.*, **62**, 242–250 (1993).
137. M. Keynes, 'Did Napoleon Die from Arsenical Poisoning?', *Lancet*, **344**, 276 (1994).
138. C.T. Wang, W.T. Chang, C.W. Huang, S.S. Chou, C.T. Lin, S.J. Liau, R.T. Wang, 'Studies on the Concentrations of Arsenic, Selenium, Copper, Zinc, and Iron in the Hair of Blackfoot Disease Patients in Different Clinical Stages', *Eur. J. Clin. Chem. Clin. Biochem.*, **32**, 107–111 (1994).
139. W. Ashraf, M. Jaffar, D. Mohammad, 'Trace Metal Contamination Study of Scalp Hair of Occupationally Exposed Workers', *Bull. Environ. Contam. Toxicol.*, **53**, 516–523 (1994).

140. W. Ashraf, M. Jaffar, D. Mohammad, 'Levels of Selected Trace Metals in Hair of Urban and Rural Adult Male Population of Pakistan', *Bull. Environ. Contam. Toxicol.*, **54**, 207–213 (1995).
141. A. de Peyster, J.A. Silvers, 'Arsenic Levels in Hair of Workers in a Semiconductor Fabrication Facility', *Am. Ind. Hyg. Assoc. J.*, **56**, 377–383 (1995).
142. V. Bencko, 'Use of Human Hair as a Biomaker in the Assessment of Exposure to Pollutants in Occupational and Environmental Settings', *Toxicology*, **101**, 29–39 (1995).
143. M.A. Armienta, R. Rodriguez, O. Cruz, 'Arsenic Content in Hair of People Exposed to Natural Arsenic Polluted Groundwater at Zimapan, Mexico', *Bull. Environ. Contam. Toxicol.*, **59**, 583–589 (1997).
144. T. Lin, Y. Huang, M. Wang, 'Arsenic Species in Drinking Water, Hair, Fingernails, and Urine of Patients with Blackfoot Disease', *J. Toxicol Environ. Health, A*, **53**, 85–93 (1998).
145. J. Buchancova, G. Klimentova, M. Knizkova, D. Mesko, E. Galikova, J. Kubik, E. Fabianova, M. Jakubis, 'Health Status of Workers of a Thermal Power Station Exposed for Prolonged Periods to Arsenic and Other Elements from Fuel', *Centr. Eur. J. Publ. Health*, **6**, 29–36 (1998).
146. R.D. Koons, C.A. Peters, 'Axial Distribution of Arsenic in Individual Human Hairs by Solid Sampling Graphite Furnace AAS', *J. Anal. Toxicol.*, **18**, 36–40 (1994).
147. W. Galas, J. Trzcionka, 'ICPAES Method in the Determination of the Metals in Human Hair', *Chem. Anal.*, **42**, 697–702 (1997).
148. M. Wolfsperger, G. Hauser, W. Gossler, C. Schlagenhaufen, 'Heavy Metals in Human Hair Samples from Austria and Italy: Influence of Sex and Smoking Habits', *Sci. Tot. Environ.*, **156**, 235–242 (1994).
149. J. Yoshinaga, Y. Shibata, M. Morita, 'Trace Elements Determined Along Single Strands of Hair by Inductively Coupled Plasma Mass Spectrometry', *Clin. Chem.*, **39**, 1650–1655 (1993).
150. G. Wallace, personal communication, 1999.
151. G.W. Stupian, 'Lead Isotope Ratio Measurements: A Potential Method for Bullet Identification', *J. Forensic Sci. Soc.*, **15**, 161–164 (1975).
152. R.M. Farquhar, J.A. Walthall, R.G.V. Hancock, '18th Century Lead Smelting in Central North America: Evidence from Lead Isotope and INAA Measurements', *J. Arch. Sci.*, **22**, 639–648 (1995).
153. B. Keisch, R.C. Callahan, 'Lead Isotope Ratios in Artists' Lead White: A Progress Report', *Archaeometry*, **18**, 181–193 (1976).
154. M.A. Haney, J.F. Gallagher, 'Elemental Analysis of Bullet Lead by Spark Source Mass Spectrometry', *Anal. Chem.*, **47**, 62–65 (1975).
155. J. Andrasko, I. Kopp, A. Abrink, T. Skiold, 'Lead Isotope Ratios in Lead Smears and Bullet Fragments and Application in Firearm Investigations', *J. Forensic Sci.*, **38**, 1161–1171 (1993).
156. C.R. Midkiff, 'SIRA – A New Forensic Tool, Part I: Bullet Lead Examinations', *Mid Atlant. Assoc. Forensic Sci. Newsl.*, **26**(3), 9–11 (1998).
157. C.R. Midkiff, 'SIRA – A New Forensic Tool, Part II: Bullet Lead and More', *Mid Atlant. Assoc. Forensic Sci. Newsl.*, **26**(4), 7–9 (1998).
158. D.G. Havekost, Report of the Visits of SA Donald G. Havekost and Dr Max A. Haney to Winchester-Western and Smith and Wesson Ammunition Plants, East Alton, and Alton, Illinois, unpublished report, 1974.
159. Y. Suzuki, Y. Marumo, 'Determination of Trace Impurities in Lead Shotgun Pellets by ICPMS', *Anal. Sci.*, **12**, 129–132 (1996).
160. T. Dufosse, P. Touron, 'Comparison of Bullet Alloys by Chemical Analysis: Use of the ICPMS Method', *Forensic Sci. Int.*, **91**, 197–206 (1998).
161. R.O. Keto, 'Analysis and Comparison of Bullet Leads by Inductively Coupled Plasma Mass Spectrometry', *J. Forensic Sci.*, **44**, 1020–1026 (1999).
162. C.A. Peters, D.G. Havekost, R.D. Koons, 'Multielement Analysis of Bullet Lead by Inductively Coupled Plasma/Atomic Emission Spectrometry', *Crime Lab. Dig.*, **15**, 33–38 (1988).
163. Unpublished Database of Compositional Bullet Lead Analyses by the Elemental Analysis Operations Group, Materials and Devices Unit, Scientific Analysis Section, FBI Laboratory, Washington, DC, 1999.
164. R.S. Maloney, J.I. Thornton, Color Tests for Diphenylamine Stabilizer and Related Compounds in Smokeless Gunpowder, 3rd Annual Meeting of the American Academy of Forensic Science, Los Angeles, CA, 1981.
165. H. Meng, B. Caddy, 'Gunshot Residue Analysis – A Review', *J. Forensic Sci.*, **42**, 553–570 (1997).
166. R.D. Koons, 'Analysis of Gunshot Primer Residue Collection Swabs by Inductively Coupled Plasma/Mass Spectrometry', *J. Forensic Sci.*, **43**, 748–754 (1998).
167. M.H. Mach, A. Pallos, P.F. Jones, 'Feasibility of Gunshot Residue Detection Via Its Organic Constituents. Part I: Analysis of Smokeless Powders by Combined Gas Chromatography/Chemical Ionization Mass Spectrometry', *J. Forensic Sci.*, **23**, 433–445 (1978).
168. D.B. Dahl, P.F. Lott, 'Gunshot Residue Determination by Means of Gunpowder Stabilizers Using High-performance Liquid Chromatography with Electrochemical Detection and Analysis of Metallic Residues by Graphite Furnace Atomic Absorption Spectrophotometry', *Microchem. J.*, **35**, 347–359 (1987).
169. G. Hellmiss, W. Lichtenberg, M. Weiss, 'Investigation of Gunshot Residues by Means of Auger Electron Spectroscopy', *J. Forensic Sci.*, **32**, 747–760 (1987).
170. R.S. White, A.D. Owens, 'Automation of Gunshot Residue Detection and Analysis by Scanning Electron

Microscopy/Energy Dispersive X-ray Analysis (SEM/EDX)', *J. Forensic Sci.*, **32**, 1595–1603 (1987).
171. R.D. Koons, D.G. Havekost, C.A. Peters, 'Determination of Barium in Gunshot Residue Collection Swabs Using Inductively Coupled Plasma/Atomic Emission Spectrometry', *J. Forensic Sci.*, **33**, 35–41 (1988).
172. R.D. Koons, D.G. Havekost, C.A. Peters, 'Analysis of Gunshot Primer Residue Collection Swabs Using Flameless Atomic Absorption Spectrophotometry and Inductively Coupled Plasma/Atomic Emission Spectrophotometry: Effects of a Modified Extraction Procedure and Storage of Standards', *J. Forensic Sci.*, **34**, 218–221 (1989).
173. L. Varetto, 'The Use of Plasma Ashing on Samples for Detection of Gunshot Residues with Scanning Electron Microscopy and Energy-dispersive X-ray Analysis (SEM/EDXA)', *J. Forensic Sci.*, **35**, 964–970 (1990).
174. D.G. Havekost, C.A. Peters, R.D. Koons, 'Barium and Antimony Distributions on the Hands of Nonshooters', *J. Forensic. Sci.*, **35**, 1096–1114 (1990).
175. W.A. MacCrehan, K.D. Smith, W.F. Rowe, 'Sampling Protocols for the Detection of Smokeless Powder Residues Using Capillary Electrophoresis', *J. Forensic Sci.*, **43**, 119–124 (1998).
176. G.M. Triplett, 'Gunshot Residue: A Review of Analytical Techniques', *Forensic Sci. Dig.*, **5**, 194–210 (1979).
177. G. Gonzales, A Simple Method for Concentrating Gunshot Residue Particles, FBI Laboratory Archived Research Materials and Notes, 1979.
178. C. Vandecasteele, G. Vermeir, R. Dams, 'Element Concentration in the Air of an Indoor Shooting Range', *Environ. Tech. Lett.*, **9**, 1287–1294 (1988).
179. J. Borszeki, E. Gegus, P. Halmos, E. Mako, 'Quality Control of Industrial Alloys Using ICPAES and Pattern Recognition', *Can. J. Appl. Spectrosc.*, **37**, 1–5 (1992).
180. R.C. Carpenter, 'The Analysis of Casework Sized Alloy Fragments by Inductively Coupled Plasma/Optical Emission Spectrometry (ICP/OES) Using Discrete Nebulization', *Forensic Sci. Int.*, **27**, 165–170 (1985).
181. R.D. Koons, C.A. Peters, R.A. Merrill, 'Forensic Comparison of Household Aluminum Foils Using Elemental Composition by Inductively Coupled Plasma/Atomic Emission Spectrometry', *J. Forensic Sci.*, **38**, 302–315 (1993).
182. X. Feng, G. Horlick, 'Analysis of Aluminum Alloys Using Inductively Coupled Plasma and Glow Discharge Mass Spectrometry', *J. Anal. At. Spectrom.*, **9**, 823–831 (1994).
183. A.G. Coedo, M.T. Dorado, I. Padilla, F.J. Alguacil, 'Use of Boric Acid to Improve the Microwave-assisted Dissolution Process to Determine Fluoride Forming Elements in Steels by Flow Injection Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1193–1197 (1998).
184. D.A. Wynn, 'A Complete Dissolution Procedure for Sn–Pb Solders Using Nitric and Hydrochloric Acids with Simultaneous Determination of Major and Trace Elements by ICP/OES', *Talanta*, **40**, 1207–1211 (1993).
185. D.G. Poolman, P.C. Pistorius, 'The Possibility of Using Elemental Analysis to Identify Debris from the Cutting of Mild Steel', *J. Forensic. Sci.*, **41**, 998–1004 (1996).
186. H.P. Longerich, B.J. Fryer, D.F. Strong, 'Trace Analysis of Natural Alloys by Inductively Coupled Plasma/Mass Spectrometry (ICP/MS): Application to Archaeological Native Silver Artifacts', *Spectrochim. Acta, Part B*, **42**, 101–109 (1987).
187. R.R. Barefoot, 'Determination of the Precious Metals in Geological Materials by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1077–1084 (1998).
188. R. Matschat, M. Czerwensky, M. Hamester, S. Pattberg, 'Investigations Concerning the Analysis of High-purity Metals (Cd, Cu, Ga, and Zn) by High-resolution Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **359**, 418–423 (1997).
189. J.C. Van Loon, R.R. Barefoot, 'Rocks, Water, and Biogeochemical Materials', in *Determination of the Precious Metals: Selected Instrumental Methods*, Wiley, Chichester, 175–220, 1991.
190. D.M. Hughes, D.C. Gregoire, H. Naka, C.L. Chakrabarti, 'Determination of Trace Amounts of Phosphorus in High Purity Iron by Electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta, Part B*, **53**, 1079–1085 (1998).
191. S.D. Tanner, V.I. Baranov, 'Theory, Design, and Operation of a Dynamic Reaction Cell for ICPMS', *At. Spectrosc.*, **20**, 45–52 (1999).
192. D.S. Bollinger, A.J. Schleisman, 'Analysis of High Purity Acids Using a Dynamic Reaction Cell ICPMS', *At. Spectrosc.*, **20**, 60–63 (1999).
193. K. Neubauer, U. Vollkopf, 'The Benefits of a Dynamic Reaction Cell to Remove Carbon- and Chloride-based Spectral Interferences by ICPMS', *At. Spectrosc.*, **20**, 64–68 (1999).
194. D.R. Bandura, S.D. Tanner, 'Effect of Collisional Damping in the Dynamic Reaction Cell on the Precision of Isotope Ratio Measurements', *At. Spectrosc.*, **20**, 69–72 (1999).
195. E. Debrah, G. Legere, 'Improving Sensitivity and Detection Limits in ICPMS with a Novel High-efficiency Sample Introduction System', *At. Spectrosc.*, **20**, 73–77 (1999).
196. M.P. Heyworth, J.R. Hunter, S.E. Warren, The Role of Inductively Coupled Plasma Spectrometry in Glass Provenance Studies, contribution prepared for a meeting on The Chemical Analysis of Roman Pottery, King's College, May, 1984.
197. R.F. Coleman, G.C. Goode, 'Comparison of Glass Fragments by Neutron Activation Analysis', *J. Radioanal. Chem.*, **15**, 367–388 (1973).

198. L.T. Atalla, F.W. Lima, 'An Actual Case of Glass Debris Identification by NAA in an Automobile Accident', *Radiochem. Radioanal. Lett.*, **3**, 13–22 (1970).
199. J.C. Hughes, T. Catterick, G. Southeard, 'The Quantitative Analysis of Glass by Atomic Absorption Spectroscopy', *Forensic Sci.*, **8**, 217–227 (1976).
200. T. Catterick, C.D. Wall, 'Rapid Analysis of Small Glass Fragments by Atomic-absorption Spectroscopy', *Talanta*, **25**, 573–577 (1978).
201. D.A. Hickman, 'Elemental Analysis and the Discrimination of Sheet Glass Samples', *Forensic Sci. Int.*, **23**, 213–223 (1983).
202. D.A. Hickman, 'Glass Types Identified by Chemical Analysis', *Forensic Sci. Int.*, **33**, 23–46 (1987).
203. R.D. Koons, C. Fiedler, R.C. Rawalt, 'Classification and Discrimination of Sheet and Container Glasses by Inductively Coupled Plasma/Atomic Emission Spectrometry and Pattern Recognition', *J. Forensic Sci.*, **33**, 49–67 (1988).
204. R.D. Koons, C.A. Peters, P.S. Rebbert, 'Comparison of Refractive Index, Energy Dispersive X-ray Fluorescence and Inductively Coupled Plasma Atomic Emission Spectrometry for Forensic Characterization of Sheet Glass Fragments', *J. Anal. At. Spectrom.*, **6**, 451–456 (1991).
205. A. Zurhaar, L. Mullings, 'Characterization of Forensic Glass Samples Using Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **5**, 611–617 (1990).
206. T. Parouchais, I.M. Warner, L.T. Palmer, H. Kobus, 'The Analysis of Small Glass Fragments Using Inductively Coupled Plasma Mass Spectrometry', *J. Forensic Sci.*, **41**, 351–360 (1996).
207. W.T. Perkins, N.J.G. Pearce, T.E. Jeffries, 'Laser Ablation Inductively Coupled Plasma Mass Spectrometry: A New Technique for the Determination of Trace and Ultra-trace Elements in Silicates', *Geochim. Cosmochim. Acta*, **57**, 475–482 (1993).
208. C. Ludke, E. Hoffman, J. Skole, 'Comparative Studies on Metal Determination in Airborne Particulates by LA/ICPMS and Furnace Atomization Non-thermal Excitation Spectrometry', *Fresenius' J. Anal. Chem.*, **350**, 272–276 (1994).
209. K. Ulens, L. Moens, R. Dams, S. Van Winckel, L. Vandevelde, 'Study of Element Distributions in Weathered Marble Crusts Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **9**, 1243–1248 (1994).
210. D.J. Kontak, S. Jackson, 'Laser-ablation ICP/MS Microanalysis of Calcite Cement from a Mississippi-valley-type Zn–Pb Deposit, Nova Scotia: Dramatic Variability in REE Content on Macro- and Micro-scales', *Can. Miner.*, **33**, 445–467 (1995).
211. F.E. Lichte, 'Determination of Elemental Content of Rocks by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**, 2479–2485 (1995).
212. E.F. Cromwell, P. Arrowsmith, 'Semiquantitative Analysis with Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **67**, 131–138 (1995).
213. A. Raith, R.C. Hutton, I.D. Abell, J. Crighton, 'Nondestructive Sampling Method of Metals and Alloys for Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **10**, 591–594 (1995).
214. T.E. Jeffries, N.J.G. Pearce, W.T. Perkins, A. Raith, 'Chemical Fractionation During Infrared and Ultraviolet Laser Ablation Inductively Coupled Plasma Mass Spectrometry – Implications for Mineral Microanalysis', *Anal. Commun.*, **33**, 35–39 (1996).
215. D. Gunther, R. Frischknecht, C.A. Heinrich, H.-J. Kahlert, 'Capabilities of an Argon Fluoride 193 nm Excimer Laser for Laser Ablation Inductively Coupled Plasma Mass Spectrometry Microanalysis of Geological Materials', *J. Anal. At. Spectrom.*, **12**, 939–944 (1997).
216. R.E. Russo, X. Mao, O.V. Borisov, 'Laser Ablation Sampling', *Trends Anal. Chem.*, **17**, 461–469 (1998).
217. A.A. van Heuzen, 'Analysis of Solids by Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry', *Spectrochim. Acta, Part B*, **46**, 1803–1817 (1991).
218. P. van de Weijer, P.J.M.G. Vullings, W.L.M. Baeten, W.J.M. de Last, 'Determination of Uranium and Thorium in Aluminum with Flow Injection and Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **6**, 609–614 (1991).
219. H. Yasuhara, T. Okano, Y. Matsumura, 'Determination of Trace Elements in Steel by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Analyst*, **117**, 395–399 (1992).
220. E.F. Cromwell, P. Arrowsmith, 'Fractionation Effects in Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **49**, 1652–1660 (1995).
221. Y. Ishibashi, 'Rapid Analysis of Steel by Inductively Coupled Plasma/Atomic Emission Spectrometry and Mass Spectrometry with Laser Ablation Solid Sampling', *Iron Steel Inst. Jpn. Int.*, **37**, 885–891 (1997).
222. A.I. Yuzefovsky, D.E. Miser, 'Analysis of Discolored Aluminum Foil by Laser Ablation Inductively Coupled Plasma Mass Spectrometry and Scanning Electron Microscopy', *Appl. Spectrosc.*, **52**, 629–637 (1998).
223. K.E. Jarvis, J.G. Williams, 'Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA/ICPMS): A Rapid Technique for the Direct, Quantitative Determination of Major, Trace and Rare-earth Elements in Geological Samples', *Chem. Geol.*, **106**, 251–262 (1993).
224. K.E. Jarvis, J.G. Williams, S.J. Parry, E. Bertalan, 'Quantitative Determination of the Platinum-group Elements and Gold Using NiS Fire Assay with Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry', *Chem. Geol.*, **124**, 37–46 (1995).
225. S. Chenery, J.M. Cook, M. Styles, E.M. Cameron, 'Determination of the Three-dimensional Distributions of Precious Metals in Sulfide Minerals by Laser Ablation

Microprobe/Inductively Coupled Plasma/Mass Spectrometry (LAMP/ICPMS)', *Chem. Geol.*, **124**, 55–65 (1995).
226. R.J. Watling, H.K. Herbert, I.S. Barrow, A.G. Thomas, 'Analysis of Diamonds and Indicator Minerals for Diamond Exploration by Laser Ablation/Inductively Coupled Plasma Mass Spectrometry', *Analyst*, **120**, 1357–1364 (1995).
227. P.M. Outridge, 'Potential Applications of Laser Ablation ICPMS in Forensic Biology and Exploration Geochemistry', *Spectroscopy*, **11**, 21–26 (1996).
228. J.L. Imbert, P. Telouk, 'Application of Laser Ablation ICPMS to Elemental Analysis of Glasses', *Mikrochim. Acta*, **110**, 151–160 (1993).
229. J.A. Westgate, W.T. Perkins, R. Fuge, N.J.G. Pearce, A.G. Wintle, 'Trace-element Analysis of Volcanic Glass Shards by Laser Ablation Inductively Coupled Plasma Mass Spectrometry: Application to Tephrochronological Studies', *Appl. Geochem.*, **9**, 323–335 (1994).
230. J. Stix, G. Gauthier, J.N. Ludden, 'A Critical Look at Quantitative Laser Ablation ICPMS Analysis of Natural and Synthetic Glasses', *Can. Miner.*, **33**, 435–444 (1995).
231. D. Figg, M.S. Kahr, 'Elemental Fractionation of Glass Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **51**, 1185–1192 (1997).
232. V. Kanicky, V. Otruba, J.-M. Mermet, 'Use of Internal Standardization to Compensate for a Wide Range of Absorbance in the Analysis of Glasses by UV Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry', *Appl. Spectrosc.*, **52**, 638–642 (1998).
233. M.D. Norman, W.L. Griffin, N.J. Pearson, M.O. Garcia, S.Y. O'Reilly, 'Quantitative Analysis of Trace Element Abundances in Glasses and Minerals: A Comparison of Laser Ablation Inductively Coupled Plasma Mass Spectrometry, Solution Inductively Coupled Plasma Mass Spectrometry, Proton Microprobe and Electron Microprobe Data', *J. Anal. At. Spectrom.*, **13**, 477–482 (1998).
234. R.J. Watling, B.F. Lynch, D. Herring, 'Use of Laser Ablation Inductively Coupled Plasma Mass Spectrometry for Fingerprinting Scene of Crime Evidence', *J. Anal. At. Spectrom.*, **12**, 195–203 (1997).
235. T.A. Rooney, 'Rev Up Your Search Engines', *Today's Chemist at Work*, **7**, 17–20 (1998).

Capillary Ion Electrophoresis in Forensic Science

Kenneth E. Ferslew and Andrea N. Hagardorn
East Tennessee State University, Johnson City, USA

1 Introduction 1
1.1 History 1
1.2 Theory 2
2 Specimen Handling 2
2.1 Blood 2
2.2 Vitreous Humor 2
2.3 Other Biological Tissues 3
3 Materials and Methods 3
3.1 Reagents 3
3.2 Specimen Preparation 3
3.3 Instrumentation 3
3.4 Calculations 4
3.5 Linearity and Reproducibility 4
4 Application Results 4
4.1 Potassium 4
4.2 Lithium 5
5 Discussion and Conclusions 6
5.1 Advantages 6
5.2 Disadvantages 8
5.3 Potential 8
Acknowledgments 9
Abbreviations and Acronyms 9
Related Articles 9
References 9

Electrolyte concentrations can be determined in biological specimens by a number of analytical techniques including flame emission spectrophotometry, atomic absorption spectrometry, high-performance liquid chromatography (HPLC), and potentiometry with an ion-specific electrode. Capillary ion electrophoresis (CIE) is an adaptation of free zone capillary electrophoresis to a particular electrophoretic analysis of highly mobile low-molecular-weight ions.

Vitreous humor electrolyte and nitrogenous compound concentrations are often used to determine potential antemortem disease states or the time interval between time of death and time of autopsy (postmortem interval). Potassium ion concentration in the vitreous humor increases after death, and determination of its concentration in vitreous humor is often used to estimate the postmortem interval.

Lithium ion is considered valuable in the treatment of certain types of mania and major endogenous depression and is even used in the treatment of cluster headaches. The optimum dosage is best determined by monitoring serum lithium concentrations. Lithium toxicity can occur with elevated serum lithium concentrations and repeated measurement every 3 h of serum lithium concentrations is recommended for intoxicated patients to differentiate acute poisoning from chronic toxicity and to determine the best course of therapy.

The development of the CIE methodology reported herein offers a different qualitative and quantitative analytical technique to detect these ions in biological fluids and tissues. Application of CIE analysis for detection and quantitation of potassium and lithium ions in biological fluids and tissues commonly collected and analyzed in clinical and forensic toxicology is described.

1 INTRODUCTION

1.1 History

Electrolyte concentrations can be determined in biological specimens by a number of analytical techniques including flame emission spectrophotometry, atomic absorption spectrometry, HPLC, and potentiometry with an ion-specific electrode.[1] The development of the CIE methodology reported herein offers a different qualitative and quantitative analytical technique with which to detect ions in biological specimens.

CIE is an adaptation of capillary electrophoresis to a particular electrophoretic analysis of highly mobile low-molecular-weight ions. The use of capillary electrophoresis for ion analysis was first reported by Hjerten[2] for the separation of bismuth and copper cations. Since then, several investigators have used capillary electrophoresis to separate cations[3–5] and anions.[6] More recently, the addition of a carrier electrolyte component (a co-ion) with matching polarity and ionic mobility to the desired analytes has provided rapid analysis time of less than 5 min, improved separation, more symmetrical peak shapes and enhanced detection limits,[7–10] and allowed application of this technique to the analysis of ion concentrations in biological specimens.[6,11,12] Improvements in CIE techniques have made it a potentially desirable technique for the analysis of very low ion concentrations in small biological specimens.

Vitreous humor refers to the fluid that fills the vitreous body (or posterior chamber) of the eye. Analysis of endogenous constituents of vitreous humor is often of forensic significance in determining cause of death

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

or postmortem interval.[13,14] Comparison of the results of vitreous humor analysis to those of other biological specimens such as blood, bile, and urine can be beneficial in determining drug absorption or elimination states, exogenous or endogenous sources of ethanol, pharmacokinetic drug disposition, and diagnosis of diabetes in the deceased.[14] Vitreous humor electrolyte (sodium, chloride, potassium, calcium, and magnesium) and nitrogenous compound (urea, ammonia, creatinine, amino acids, and xanthine) concentrations are often used to determine potential antemortem disease states or time interval between time of death and time of autopsy (the postmortem interval).[13]

Potassium ion concentration in the vitreous humor increases after death. This finding was initially reported by Jaffe[15] and subsequently corroborated by others.[16–23] The increase in vitreous potassium concentration has been found to correlate to postmortem interval but with greater variation with increased interval. Variation in potassium concentration/interval correlations due to cause of death[24–26] as well as variability in potassium concentration measurement due to method of analysis has made reliability in vitreous potassium determined postmortem intervals questionable.[27,28]

Lithium ion is considered valuable in the treatment of certain types of mania and major endogenous depression. It is commonly prescribed by psychiatrists for the treatment of manic-depressive illness[29] as well as aggressive and self-mutilating behavior.[30] It is even used in the treatment of cluster headaches.[31] The mechanisms underlying how lithium works remain largely unknown, although there is strong evidence to suggest that it affects the phosphoinositide second messenger system in the brain.[32] Lithium is administered orally as a carbonate or a citrate in daily doses of 600–2400 mg. The optimum dosage is best determined by monitoring serum concentrations.[33] The optimum serum lithium concentrations have been found to range from 0.5 to 1.3 mmol L^{-1} for most patients.[34] Signs of lithium toxicity include nausea, vomiting, diarrhea, weakness, ataxia, blurred vision, confusion, stupor, seizures, and coma. Lithium toxicity is often seen if serum concentrations exceed 2 mmol L^{-1}. Repeated measurement every 3 h of serum lithium concentrations is recommended for intoxicated patients to differentiate acute poisoning from chronic toxicity and to determine the best course of therapy.[35,36] Recent reports suggest that rapid removal of pharmacologically excessive lithium concentrations from lithium intoxicated patients, as might occur with hemodialysis, can lead to serious toxic effects.[37,38] The deleterious effects of rapid lithium removal may in part be similar to what is observed in rats when they are rapidly corrected to normal sodium levels after being hypernatremic.[39]

The methods described herein[11,40] were developed to determine the applicability of CIE analysis for detection and quantitation of significant ions in biological fluids and tissues commonly collected and analyzed in clinical and forensic toxicology.

1.2 Theory

CIE is a free zone capillary electrophoresis separation technique for the analysis of low-molecular-weight inorganic and organic ions in aqueous solutions. An aqueous solution of electrolytes is drawn into a hollow silica capillary by hydrostatic sampling and a voltage is applied. Ions migrate toward the electrodes of opposite charge causing the fluid within the capillary to move, defined as electroosmotic flow. An ion's mobility or migration toward an electrode in an electrolyte solution is directly proportional to the ion's equivalent ionic conductance, which is related to the ratio of the ion's net charge to ionic size in solution. Thus the greater the mobility value the faster the ion moves toward the attracting electrode. Differences in ionic mobility results in separation of the ions as they migrate through the capillary. Individual ion mobility and selectivity can be altered in the electrolyte by addition of osmotic flow modifiers (e.g. surfactants and metal–organic complexation) for faster and more reproducible separations.

Indirect ultraviolet (UV) absorption is used to detect the ions as they migrate within the capillary and through the detector. A UV-absorbing electrolyte co-ion is added to the electrolyte solution. The analyte ion will proportionally displace the UV-absorbing electrolyte co-ion producing a decrease in background absorbance which is detected as a peak in the electropherogram. Determination of each ion's relative migration time is used for identification, and interpolation of the relative peak size for each ion is used for quantitation.

2 SPECIMEN HANDLING

2.1 Blood

Blood was collected from fish via cardiac puncture into microhematocrit tubes (Fisher Scientific Co., Fairlawn, NJ). Plasma was separated by centrifugation in a Model 235C microcentrifuge (Fisher Scientific Co., Fairtown, NJ) at 15 000 *g* and stored at 2–4 °C until analysis.

2.2 Vitreous Humor

Human vitreous humor was collected by needle puncture of the posterior chamber of both eyes and stored in a sterile Vacutainer® tube. Specimens were stored at 2–4 °C until analysis.

2.3 Other Biological Tissues

Brain and liver specimens from fish were collected at autopsy and stored in plastic sealable bags (Fisher Scientific Co., Fairlawn, NJ). Tissue specimens were stored at −70 °C until analysis.

3 MATERIALS AND METHODS

3.1 Reagents

3.1.1 Reagents for Potassium Analysis

Run electrolyte was prepared by placing 100 mL of water (18 MΩ) in a 250 mL flask on a magnetic stirrer. Then 67.7 mg of hydroxyisobutyric acid (HIBA), 64 µL of UV-CAT-1 reagent (containing 4-methylbenzylamine, Waters Corporation, Milford, MA) and 52.8 mg of 18-crown-6-ether (Fisher Scientific Co., Fairlawn, NJ) were added. The solution was stirred to dissolve, and degassed under vacuum. Run electrolyte was prepared fresh daily. An internal standard (40 ppm barium) was prepared by diluting 40 mL of barium (1000 ppm, Fisher Scientific Co., Fairlawn, NJ) to 1 L with water (18 MΩ). Standard solutions of the cations were prepared from 1000 ppm standards of each ion (Fisher Scientific Co., Fairlawn, NJ) as follows: 1000 ppm, 100 µL of each ion quanta satis (qs) 10 mL with 40 ppm barium; 500 ppm; 50 µL of each ion + 200 µL water qs 10 mL with 40 ppm barium; 250 ppm, 25 µL of each ion + 300 µL water qs 10 mL with 40 ppm barium; 100 ppm, 10 µL of each ion + 360 µL water qs 10 mL with 40 ppm barium.

3.1.2 Reagents for Lithium Analysis

Run electrolyte was prepared by placing 100 mL of water (18 MΩ) in a 250 mL flask on a magnetic stirrer. Then 67.7 mg of HIBA, 64 µL of UV-CAT-1 reagent (containing 4-methylbenzylamine, Waters Corporation, Milford, MA) and 52.8 mg of 18-crown-6-ether (Fisher Scientific Co., Fairlawn, NJ) were added. The solution was stirred to dissolve and degassed under vacuum. Run electrolyte was prepared fresh daily.

An internal standard (40 ppm is equivalent to 0.29 mM barium) was prepared by diluting 40 mL of barium (1000 ppm, Fisher Scientific Co., Fairlawn, NJ) to 1 L with water (18 MΩ). A standard 500 ppm solution of the cations (potassium, calcium, sodium, barium, and lithium) was prepared by dilution of 1000 ppm standards of each ion (Fisher Scientific Co., Fairlawn, NJ) with water (18 MΩ).

3.2 Specimen Preparation

3.2.1 Vitreous Potassium Analysis

Vitreous humor specimens were prepared as follows: 1 : 100 dilution, 100 µL vitreous fluid + 300 µL water qs to 10 mL with 40 ppm barium; 1 : 500 dilution, 1 mL of 1 : 100 dilution + 200 µL water qs to 5 mL with 40 ppm barium. Dual dilution is needed to cover the vast range of cation concentration variation in the vitreous humor (sodium is much greater than the other cations). Specimens, standards and controls were all analyzed at 1 : 100 dilutions in 40 ppm barium. Specimens outside of the range of linearity were diluted 1 : 500 for quantitation. Aliquots (100 µL) of specimens and controls were transferred to 0.5 mL polypropylene sample vials (Waters Corporation, Milford, MA) and placed in the sample carousel of the Quanta 4000 for analysis.

3.2.2 Lithium Analysis

Ten-µL aliquots of plasma, 10 µL of 0.29 mmol barium in 18 MΩ water and 10 µL of lithium were added to 30 µL of 18 MΩ deionized water and centrifuged in a microcentrifuge (Model 235C, Fisher Scientific Co., Fairlawn, NJ) at 15 000 *g* in 10 000 nominal molecular weight limit (NMWL) Ultrafree®-MC Centrifugal Filter Unit tubes (Millipore Corporation, Bedford, MA) to remove high-molecular-weight components that might clog the electrophoretic capillary tube. The lithium concentrations were 0.1, 0.25, 0.5, 1.0, 1.5 and 2.0 mmol for development of standard curves or 0.375 and 0.75 mmol to assess within and between-day precision. Final sample assay volumes were 30 µL.

Liver and brain specimens (300–500 mg) were homogenized 1 : 4 w/v with 18 MΩ water. Whole tissue homogenates were spun in a microcentrifuge (Model 235C, Fisher Scientific Co., Fairlawn, NJ) at 15 000 *g* and supernates were collected. Ten-µL aliquots of supernate, 10 µL of 0.29 mM barium in 18 MΩ water and 10 µL of lithium were added to 30 µL of 18 MΩ deionized water and were centrifuged at 15 000 *g* in the Millipore Ultrafree® units. Final sample assay volumes were 30 µL.

3.3 Instrumentation

Samples were analyzed on a Waters Quanta 4000 Capillary Electrophoresis System with a Waters 745 Data Module (Waters Corporation, Milford, MA). Analytical conditions: hydrostatic sampling time, 5 s for plasma and 10 s for vitreous, liver and brain; run time, 7 min at 25 °C; power supply, positive 20 kV; autopurge between samples, 1 min with buffer (the capillary is flushed with clean buffer to remove any contaminants); current, approximately 5 µA; detector polarity, negative; UV

detection, 214 nm; absorbance range, 0.02 absorbance units full scale; time constant, 0.3 s. The 745 Data Module settings were as follows: attenuation, 4 mV to full scale deflection; chart speed, 2 cm min^{-1}; peak width, 3; time function now (TFN), peak marker (PM) on; peak threshold (PT) evaluation, run after electrolyte was allowed to stabilize in the capillary for 10 min (value should be between 50 and 200). Dialog for integration was as follows: time, 0.1 min, PT 90 (only if PT evaluation is greater than 90); time 0.2 min, auto zero (AZ) on; time 2.0 min, AZ on; time 7 min, end run (ER) on.

A Waters Accusep Capillary (75 μm × 60 cm) was used for these analyses. Carousel set-up was performed by filtering 17 mL of capillary ion electrolyte through a 0.2 μm Anotop™ 25 inorganic membrane filter (Whatman International, Ltd, Maidstone, England) using a 20 mL syringe (Fisher Scientific Co., Fairlawn, NJ) into each electrolyte reservoir to be used. One reservoir was needed for each group of five samples. The maximum number of samples per carousel was 20. No preconditioning was performed on the capillary, and the run electrolyte was allowed to equilibrate within the capillary for 10 min (as noted above) before analyses were performed. Stabilization produced a steadier baseline and better electropherograms with this procedure and integrator.

3.4 Calculations

3.4.1 Vitreous Potassium Calculations

Cation concentrations were used in ppm to make handling of the solutions easier. Concentrations of each ion were converted into millimoles per liter based on their molecular weight, to be relevant to the forensic and clinical literature. Linear regressions for the standard curves of concentrations and peak area ratios (peak area of analyte: peak area of internal standard) as well as means, standard deviations, and errors were calculated using a model 1860 statistical programmable printing calculator (Monroe, The Calculator Co., Orange, NJ). Specimen concentrations were determined by interpolation of their peak area ratios from those of the standard curves. Coefficients of variation (CV) were determined by dividing the standard deviations by the means and multiplying by 100.

3.4.2 Lithium Calculations

The lithium concentrations used were in millimoles, to make handling of the solutions easier. Concentrations for individual tissue samples were converted into nanograms of Li per milligram of tissue or nanograms of Li per microliter of plasma. Linear regressions for the standard curves of concentrations and lithium: barium peak area ratios as well as means, standard deviations and errors, and CV were determined using Stat View Student Software (Abacus Concepts, Inc., Berkeley, CA) for a Power Macintosh computer (Apple Computer, Cupertino, CA). Specimen concentrations were determined by interpolation of their lithium: barium peak area ratios from those of the standard curves.

3.5 Linearity and Reproducibility

3.5.1 Vitreous Potassium

Standards of each analyte were prepared in deionized water over a concentration range of 100–1000 ppm for determination of linearity. Aliquoted samples of each cation were analyzed to determine within-day ($n = 5$) and between-day ($n = 5$) CV.

3.5.2 Lithium

Standards of lithium were prepared in plasma and brain or liver homogenate over a lithium concentration range of 0.1, 0.25, 0.5, 1.0, 1.5, and 2.0 mM for determination of linearity. Control specimens were analyzed at 0.375 and 0.75 mM lithium concentrations. Aliquoted samples of each concentration were analyzed to determine within-day ($n = 5$) and between-day ($n = 5$) CV.

4 APPLICATION RESULTS

4.1 Potassium

A typical electropherogram for a 1 : 100 dilution of a mixture containing 250 ppm standards of all the cations (and barium, internal standard 40 ppm) is illustrated in Figure 1. Migration times were: potassium, 3.24 min; calcium, 3.84 min; sodium, 3.98 min; barium (internal standard), 4.68 min; lithium, 4.79 min. A typical vitreous humor electropherogram is illustrated in Figure 2. As many of the vitreous humor specimens are collected from severely decayed bodies, it is possible for them to contain significant concentrations of ammonia. An electropherogram of an actual specimen illustrating the separation of ammonia from the other cations is presented in Figure 3.

Correlation coefficients (r) between peak area ratios and concentration ranges of 100–1000 ppm for each cation are given in Table 1 (each value is for a between-day analysis). Correlation coefficients ranged from 0.9855 to 0.9999. Both within-day and between-day CV are given in Table 2. CVs ranged from 1.45% to 13.8% between-days and from 1.38% to 9.43% within-day.

Application of this methodology to 25 vitreous humor specimens from forensic autopsies was compared to analysis by ion-specific electrode for potassium concentration. Comparison of CIE to ion-specific electrode analysis

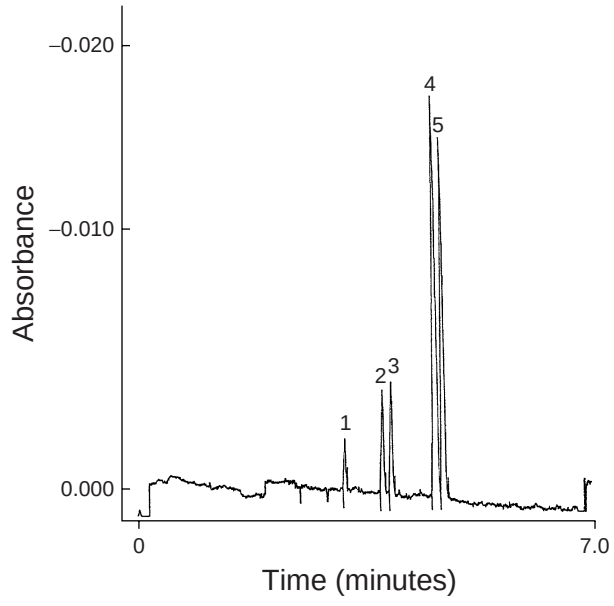

Figure 1 Electropherogram of standard cations (1 : 100 dilution of 250 ppm each ion in 40 ppm barium): 1 = potassium, 2 = calcium, 3 = sodium, 4 = barium internal standard, 5 = lithium. (Reproduced by permission of Wiley-VCH, STM from K.E. Ferslew, A.N. Hagardorn, M.T. Harrison, W.F. McCormick, *Electrophoresis*, **19**, 6–10 (1998).)

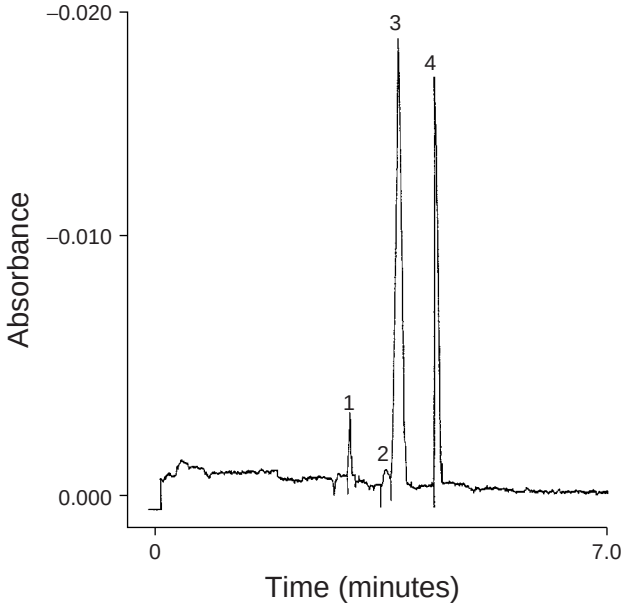

Figure 2 Typical vitreous humor electropherogram: 1 = potassium, 2 = calcium, 3 = sodium, 4 = barium internal standard. (Reproduced by permission of Wiley-VCH, STM from K.E. Ferslew, A.N. Hagardorn, M.T. Harrison, W.F. McCormick, *Electrophoresis*, **19**, 6–10 (1998).)

of vitreous humor potassium concentrations (Figure 4) revealed a correlation coefficient of 0.9642.

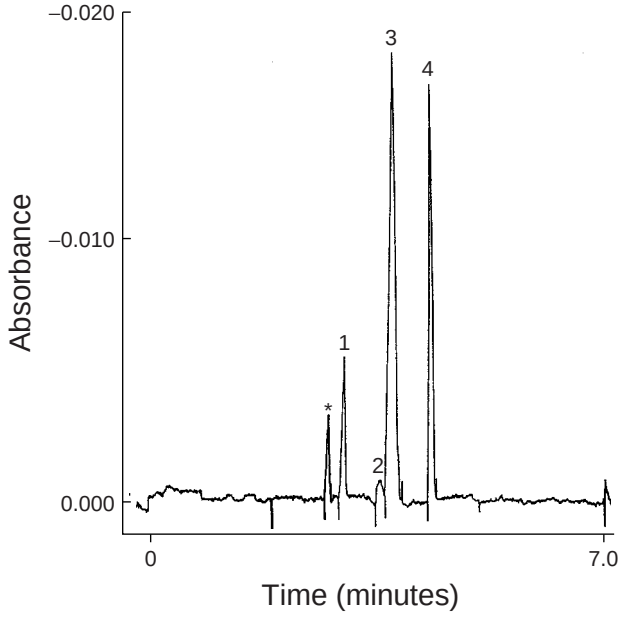

Figure 3 Electropherogram of vitreous humor containing ammonia: ∗ = ammonia, 1 = potassium, 2 = calcium, 3 = sodium, 4 = barium internal standard. (Reproduced by permission of Wiley-VCH, STM from K.E. Ferslew, A.N. Hagardorn, M.T. Harrison, W.F. McCormick, *Electrophoresis*, **19**, 6–10 (1998).)

Table 1 Standard curve correlation coefficients (r)

	Cation			
	Potassium	Calcium	Sodium	Lithium
	0.9989	0.9998	0.9981	0.9996
	0.9984	0.9989	0.9982	0.9999
	0.9950	0.9992	0.9972	0.9994
	0.9994	0.9994	0.9976	0.9996
	0.9855	0.9995	0.9981	0.9995
Mean	0.9954	0.9994	0.9978	0.9996
SE[a]	0.0023	0.0001	0.0002	0.0001

[a] SE = standard error.

4.2 Lithium

A typical electropherogram for a mixture containing 500 ppm standards of cations is illustrated in Figure 5. The migration times were potassium 2.98 min, calcium 3.48 min, sodium 3.60 min, barium 4.15 min, and lithium 4.26 min. Lithium and barium migration times were stable and reproducible (mean relative migration time for lithium/barium from all standards was 1.0184 ± 0.0011, $n = 78$). A representative electropherogram of fish liver extract is illustrated in Figure 6. Electropherograms for plasma and brain extract (not shown) were similar. Standard curves of lithium/barium peak area ratios versus lithium concentrations for plasma and for liver and brain

Table 2 Within-day and between-day statistics

Analyte	Concentration[a]	Within-day		Between-day	
		Mean ± SE	CV	Mean ± SE	CV
Potassium (K)	2.56 (100)	2.82 ± 0.12	9.43	2.42 ± 0.15	13.8
	6.39 (250)	6.49 ± 0.17	5.7	6.69 ± 0.36	11.92
	12.79 (500)	11.48 ± 0.45	8.79	12.96 ± 0.73	12.58
Calcium (Ca)	2.50 (100)	2.71 ± 0.05	4.54	2.71 ± 0.04	3.45
	6.24 (250)	6.45 ± 0.04	1.47	6.51 ± 0.06	1.96
	12.48 (500)	11.94 ± 0.17	3.13	12.36 ± 0.19	3.44
Sodium (Na)	4.35 (100)	5.78 ± 0.09	3.57	6.03 ± 0.04	1.45
	10.87 (250)	10.82 ± 0.15	3.09	10.89 ± 0.16	3.36
	21.75 (500)	20.90 ± 0.42	4.47	20.89 ± 0.28	3.03
Lithium (Li)	14.41 (100)	14.71 ± 0.11	1.74	14.61 ± 0.12	1.78
	36.02 (250)	36.19 ± 0.24	1.46	35.70 ± 0.29	1.82
	72.04 (500)	68.35 ± 0.42	1.38	70.06 ± 0.63	2.00

[a] Concentrations are given in mmol L^{-1} (ppm).

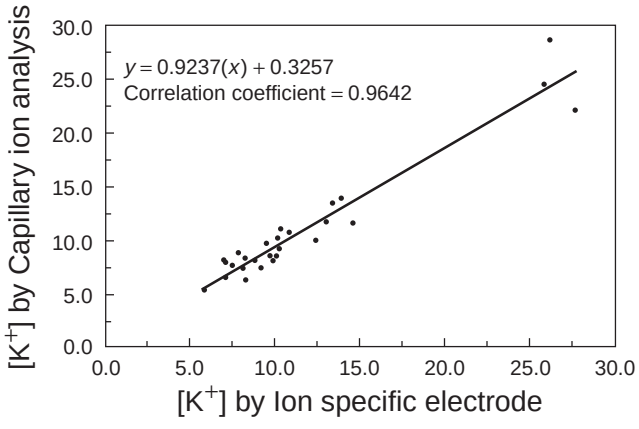

Figure 4 Correlation of vitreous humor potassium concentrations (mmol L^{-1}) by CIE versus ion-specific electrode. (Reproduced by permission of Wiley-VCH, STM from K.E. Ferslew, A.N. Hagardorn, M.T. Harrison, W.F. McCormick, *Electrophoresis*, **19**, 6–10 (1998).)

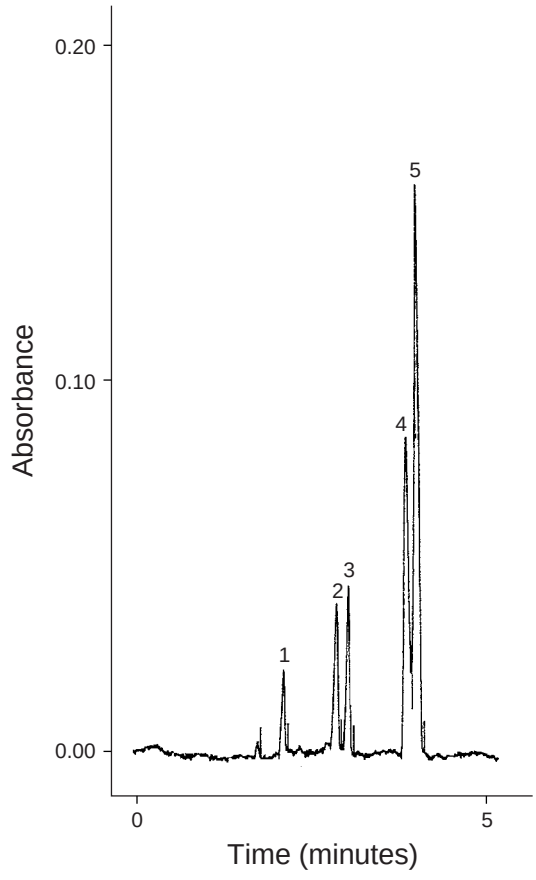

Figure 5 Electropherogram of standard cations (500 ppm of each ion): 1 = potassium, 2 = calcium, 3 = sodium, 4 = barium internal standard, 5 = lithium. (Reproduced by permission of Wiley-VCH, STM from T.K. Creson, P.J. Monaco, E.M. Rasch, A.N. Hagardorn, K.E. Ferslew, *Electrophoresis*, **19**, 3018–3021 (1998).)

tissue extracts are shown in Figure 7. Equations for linear regression and variances between peak area ratios and concentrations are also shown in Figure 7. Correlation coefficients ranged from 0.976 to 0.996. Both within-day and between-day CVs are given in Table 3. CV values for lithium concentrations ranged from 4.07% to 15.71% between-days and from 4.38% to 7.76% within-day.

Application of CIE lithium methodology to the determination of lithium from plasma and tissues of fish consistently dosed for 23 days is given in Table 4.

5 DISCUSSION AND CONCLUSIONS

5.1 Advantages

The CIE cation separation was obtained with a short, small-bore capillary column with minimal run electrolyte solution (<2 mL per analysis). Barium can be used as an internal standard to improve accuracy and precision

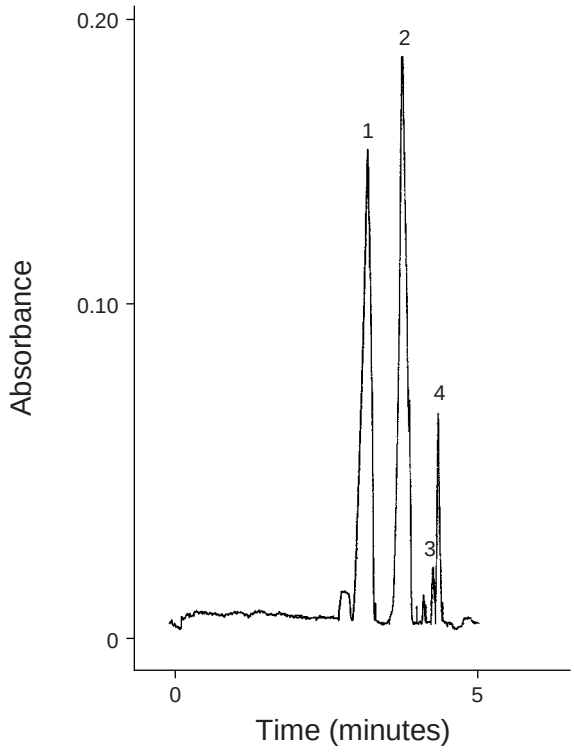

Figure 6 Typical fish liver electropherogram: 1 = potassium and calcium, 2 = sodium, 3 = barium internal standard, 4 = lithium. (Reproduced by permission of Wiley-VCH, STM from T.K. Creson, P.J. Monaco, E.M. Rasch, A.N. Hagardorn, K.E. Ferslew, *Electrophoresis*, **19**, 3018–3021 (1998).)

as well as quality assurance with each analysis. Baseline separation was obtained for the cations of interest in under 5 min. Direct analysis of the vitreous humor (after appropriate dilution to obtain concentrations within the instrument's sensitivity and analytical range) is a distinct advantage. The aqueous nature of the vitreous humor allows direct analysis without pretreatment or filtration. Analysis of biological specimens with large concentrations of proteins, lipids or other endogenous constituents (i.e. plasma, urine, or tissue homogenates) can be accomplished with easy removal of interfering substances by centrifugal ultrafiltration to allow for determination of electrolyte concentrations in the resulting filtrate solution.

Ion electropherograms are more similar to capillary gas chromotography (CGC) chromatograms than HPLC chromatograms in peak shape and size characteristics. Separation efficiencies are equal or greater than HPLC. The use of short (60 cm) noncoated capillaries which are relatively inexpensive is a distinct advantage versus typical columns used in HPLC. Bulk capillary tubing can be easily cut to the desired length and the optical detection section made by burning the polyamide coating off the fused silica capillary, or standard commercial capillaries

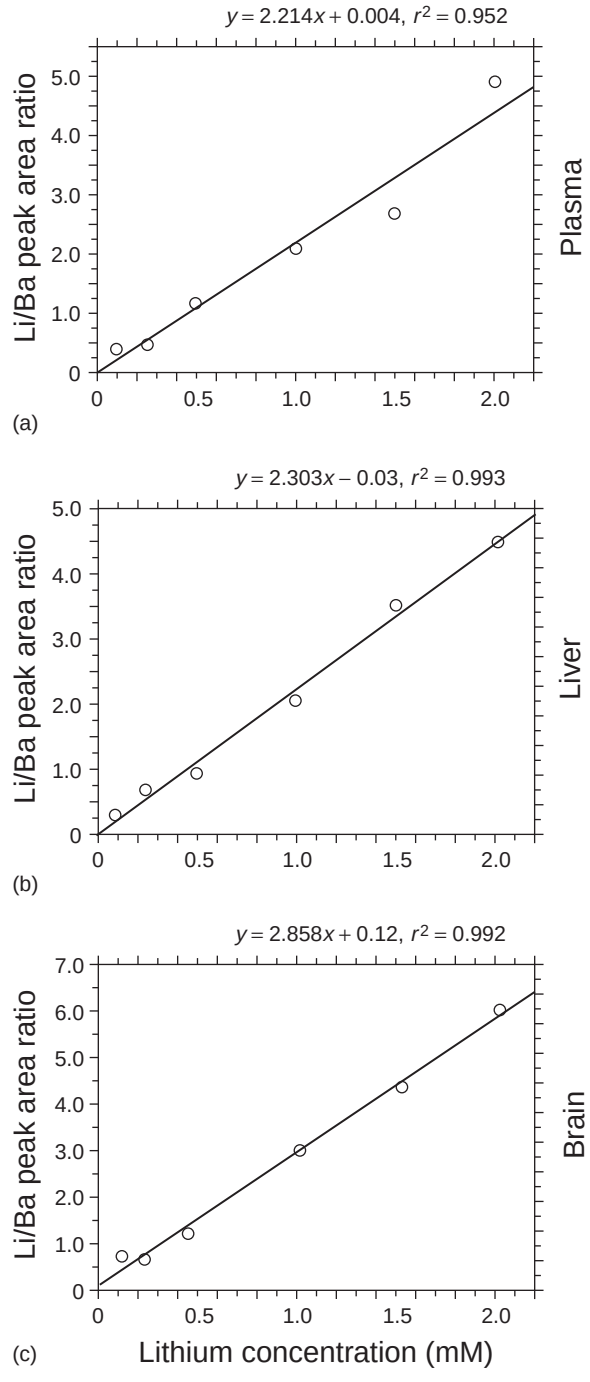

Figure 7 Correlation of lithium/barium peak area ratios to lithium concentrations for (a) plasma and for extracts of (b) liver and (c) brain. Regression equations and variances are shown for each graph. (Reproduced by permission of Wiley-VCH, STM from T.K. Creson, P.J. Monaco, E.M. Rasch, A.N. Hagardorn, K.E. Ferslew, *Electrophoresis*, **19**, 3018–3021 (1998).)

can be used. Capillaries of various bore sizes (10–100 μm) can be used which allows for detection of a wide range of analyte concentrations. The CIE methodology described

Table 3 Within-day and between-day statistics

Sample lithium concentration	Within-day		Between-day	
	Mean (SE)[a]	CV	Mean (SE)[a]	CV
Plasma 0.375 mM	0.383 (0.01)	6.84	0.361 (0.03)	15.71
Plasma 0.75 mM	0.590 (0.02)	7.76	0.588 (0.01)	5.54
Liver 0.375 mM	0.356 (0.01)	4.38	0.369 (0.02)	13.81
Liver 0.75 mM	0.747 (0.03)	7.68	0.646 (0.03)	10.6
Brain 0.375 mM	0.330 (0.01)	5.48	0.334 (0.01)	4.07
Brain 0.75 mM	0.754 (0.02)	5.49	0.718 (0.02)	5.71

[a] Mean and SE values are given in mmol L^{-1}.

Table 4 Lithium concentrations detected in fish after 23 days of LiCl administration

Tissue sample	Lithium concentration
Plasma	8.593 (0.616)
Liver	4.903 (0.991)
Brain	1.354 (0.127)

Data represent the mean (and SE) from five fish. Plasma lithium concentrations are in ng μL^{-1}, brain and liver lithium concentrations are in ng per mg of tissue. Fish were sacrificed on day 23, immediately after LiCl administration.

is automated and as easily operated as any automated HPLC systems encountered by the authors.

The excellent separation capabilities of CIE are demonstrated by the baseline separation of potassium, sodium, barium, lithium, calcium, and ammonia, and yield identification and quantitation of multiple ions on each specimen in a single analysis. The hydrostatic sampling (versus electromigratory) and CIE separation produced accurate and precise quantitation. The UV detection at 214 nm (versus 254 nm) offered an adequate minimum detectable concentration and a linear range of detection. As the optical detection cell is within the capillary, the bore size has a directly proportional effect on the limit of detection and the separability of the analytes. The internal diameters most useful are 50–75 μm, as narrower bores clog easily and larger bores result in increased heat production. The limit of detection with biological specimens is not a problem and allows for determination of forensically and clinically relevant cation concentrations.

Recent research developments in run electrolyte make improved analysis attainable. HIBA is used in the run electrolyte as a chelating agent (i.e. electroosmotic flow modifier) to control the flow of buffer through the capillary, improving elution and peak shape. Addition of the 18-crown-6-ether provides separation of ammonia from potassium so that comigration of these cations will not occur causing false quantitation of potassium concentrations, which is of particular importance in analyzing postmortem specimens. Addition of 4-methylbenzylamine to the run electrolyte produces a high background absorbance. Detection of the separated ions is by indirect negative UV absorbance produced by displacement of the background co-ion. Decreased absorbance is directly proportional to the concentration of the ion eluting from the capillary. This addition improves sensitivity of detection, especially at 214 nm.

5.2 Disadvantages

CIE is a more complex methodology than potentiometry with ion-specific electrodes. The analysis time per specimen is distinctly longer due to the serial elution of numerous endogenous analytes in any given biological specimen. Elution of multiple analytes of varying concentrations (and often very high concentrations in biological fluids, i.e. sodium and potassium) can cause capillary overloading resulting in misshaped peaks and peak masking. This may necessitate serial dilution and multiple analysis to identify and quantitate multiple ions in a given specimen. Significant preparation of run electrolyte and instrument set-up are needed for a single analysis or multiple analyses with CIE which would not be necessary for potentiometry with ion-specific electrodes.

The operation of the CIE methodology is more similar to HPLC than it is to CGC. Fresh daily prepared run electrolytes must be made precisely to achieve good reproducible electropherograms, and be filtered and degassed in order to avoid gas bubble formation within the capillary as it heats up under large potentials. Bubbles can produce noisy baselines and false peaks that interfere in the analyses, especially with low-wavelength UV detection, and can even short out the field and abort the run. Run electrolyte depletion can occur with CIE, which does not occur with HPLC because the mobile phase is homogeneous and is under continuous flow through the column. Electrophoretic flow of ions occurs from one electrode reservoir to the other through the capillary. Repeated analyses with the same reservoirs can therefore result in run electrolyte depletion, which can effect both injection accuracy and analyte separation. Limiting the number of analyses per buffer aliquot to five or less circumvents this problem.

5.3 Potential

The CIE methodology has great potential for the analysis of a number of significant chemicals in forensic

toxicology. Although this article focuses on the analysis of potassium and lithium cations in biological fluids and tissues, CIE with a negative power supply can detect anions such as bromide, chloride, fluoride, nitrite, nitrate, and oxalate as well. Toxicologically significant concentrations of these anions in biological fluids and tissues are easily within the detection capabilities of CIE. Quantitation of toxic concentrations of these anions in forensic specimens to determine toxicity or lethality in clinical and forensic cases, as well as the detection of adulterants (e.g. nitrites, nitrates, and hypochlorite), in forensic urine specimens is the focus of future research with CIE.

Other biological fluids (such as perilymph and endolymph of the ear, cerebrospinal fluid, amniotic fluid, and urine) which are primarily aqueous in nature, like vitreous humor, also have the potential for direct analysis by CIE. CIE analysis of ions in these specimens, especially when the amount of fluid specimen attainable for analysis is minimal (i.e. a few microliters), can be beneficial in both forensic toxicology practice and research. In specimens where large concentrations of proteins, lipids, or other endogenous constituents are present and may interfere with analysis or potentially clog the capillary, such as plasma and tissue homogenates, and particularly with the products of putrefaction in forensic specimens, the combination of centrifugal ultrafiltration and CIE analysis will allow for determination of electrolyte concentrations in the resulting filtrate solution. Further application to the analysis of ions from tissue digests or ashed specimens is possible by CIE. The only limitation is that the chemicals to be analyzed must be soluble in aqueous solution.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance and support of Dennis Beaudry of Waters Corporation in conducting these investigations. Travis Harrison's assistance with original potassium research on vitreous specimens was supported by a Howard Hughes Medical Grant through King College, Bristol, VA. The authors wish to express their appreciation for the collaboration of Ellen Rasch, Paul Monaco, Thomas Creson, Travis Harrison, and William McCormick in conducting the original research.

ABBREVIATIONS AND ACRONYMS

AZ	Auto Zero
CGC	Capillary Gas Chromotography
CIE	Capillary Ion Electrophoresis
CV	Coefficients of Variation
ER	End Run
HIBA	Hydroxyisobutyric Acid
HPLC	High-performance Liquid Chromatography
NMWL	Nominal Molecular Weight Limit
PM	Peak Marker
PT	Peak Threshold
qs	Quanta Satis
SE	Standard Error
TFN	Time Function Now
UV	Ultraviolet

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Ion Chromatography in Environmental Analysis

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Ion Chromatography • Micellar Electrokinetic Chromatography

REFERENCES

1. N.W. Tietz, E.L. Pruden, O. Siggard-Andersen, 'Electrolytes, Blood, Gases, and Acid–Base Balance, Section One, Electrolytes', in ed. N.W. Tietz, *Fundamentals of Clinical Chemistry*, 3rd edition, W.B. Saunders Company, PA, 614–624, 1997.
2. S. Hjerten, 'Free Zone Electrophoresis', *Chromatogr. Rev.*, **9**, 122–219 (1967).

3. T. Tsuda, K. Nomura, G.J. Nakagawa, 'Separation of Organic and Metal Ions by High-voltage Electrophoresis', *J. Chromatogr.*, **264**, 385–392 (1983).
4. X. Huang, T.K.J. Pang, M.J. Gordon, R.N. Zare, 'On-column Conductivity Detection or Capillary Zone Electrophoresis', *Anal. Chem.*, **59**, 2747–2749 (1987).
5. J.L. Beckers, T.P.E.M. Verheggen, F.M. Everaerts, 'Use of a Double Detector System for the Measurement of Mobilities in Zone Electrophoresis', *J. Chromatogr.*, **452**, 591–600 (1988).
6. A. Meulemans, F. Delsenne, 'Measurement of Nitrite and Nitrate Levels in Biological Samples by Capillary Electrophoresis', *J. Chromatogr. B*, **660**, 401–404 (1994).
7. P. Jandik, W.R. Jones, A. Weston, P.R. Brown, 'Electrophoretic Capillary Ion Analysis: Origins, Principles, and Applications', *LC/GC*, **9**, 634–645 (1991).
8. F. Foret, S. Fanali, A. Nardi, P. Bocek, 'Capillary Zone Electrophoresis of Rare Earth Metals with Indirect UV Absorbance Detection', *Electrophoresis*, **11**, 780–783 (1990).
9. W.R. Jones, P. Jandik, A. Weston, Abstract No. LTG, presented at the Third International Symposium on High-performance Capillary Electrophoresis, San Diego, CA, 3–6 February, 1981.
10. A. Weston, W.R. Jones, P. Jandik, Abstract No 158, presented at the 42nd Pittsburgh Conference and Exposition, Chicago, IL, 4–8 March, 1991.
11. K.E. Ferslew, A.N. Hagardorn, M.T. Harrison, W.F. McCormick, 'Capillary Ion Analysis of Potassium Concentrations in Human Vitreous Humor', *Electrophoresis*, **19**, 6–10 (1998).
12. S. McClean, E. O Kane, D.J.M. Coulter, S. McLean, W.F. Smyth, 'Capillary Electrophoretic Determination of Trace Metals in Hair Samples and its Comparison with High-performance Liquid Chromatography and Atomic Absorption Spectrometry Techniques', *Electrophoresis*, **19**(1), 11–18 (1998).
13. J.I. Coe, 'Postmortem Chemistry Update. Emphasis on Forensic Application', *Am. J. For. Med. & Path.*, **14**, 91–117 (1993).
14. R.O. Bost, 'Analytical Toxicology of Vitreous Humor' in *Analytical Toxicology of Vitreous Humor. Handbook of Analytical Therapeutic Drug Monitoring and Toxicology*, eds. S.H.Y. Wong, I. Sunshine, CRC Press, Boca Raton, FL, 281–302, 1997.
15. F. Jaffe, 'Chemical Postmortem changes in the Intraocular Fluid', *J. Forensic Sci.*, **7**, 231–237 (1962).
16. J.I. Coe, 'Postmortem Chemistries on Human Vitreous Humor', *Am. J. Clin. Path.*, **51**, 741–750 (1969).
17. L. Adelson, I. Sunshine, N.B. Rushforth, M. Mankoff, 'Vitreous Potassium Concentration as an Indicator of the Postmortem Interval', *J. Forensic Sci.*, **8**, 503–514 (1963).
18. L. Hanson, V. Votila, R. Lindfors, K. Laiho, 'Potassium Content of the Vitreous Body as an Aid in Determining the Time of Death', *J. Forensic Sci.*, **11**, 390–394 (1966).
19. W. Hughes, 'Levels of Potassium in the Vitreous Humor after Death', *Med. Sci. Law*, **5**, 150–156 (1965).
20. J.T. Lie, 'Changes of Potassium Concentration in the Vitreous Humor after Death', *Am. J. Med. Sci.*, **254**, 136–142 (1967).
21. W.Q. Sturner, 'The Vitreous Humor: Postmortem Potassium Changes', *Lancet*, **1**, 807–808 (1963).
22. W.Q. Sturner, G.E. Ganter, 'The Postmortem Interval: A Study of Potassium in the Vitreous Humor', *Am. J. Clin. Path.*, **42**, 137–144 (1964).
23. G. Adjutantis, A. Coutselinis, 'Estimation of the Time of Death by Potassium Levels in the Vitreous Humor', *Forensic Sci.*, **1**, 55–60 (1972).
24. L. Rammer, B. Gerdin, 'Dilution of Blood in Fresh Water Drowning. Post-mortem Determination of Osmolarity and Electrolytes in Blood, Cerebrospinal Fluid and Vitreous Humor', *Forensic Sci.*, **8**, 229–234 (1976).
25. W.Q. Sturner, A. Balko, J. Sullivan, 'Magnesium and Other Electrolytes in Bovine Eyeballs Immersed in Seawater and other Fluids', *Forensic Sci.*, **8**, 139–150 (1976).
26. E.D. Baker, K.E. Ferslew, R.O. Orcutt, W.F. McCormick, 'Vitreous Potassium Concentration and Postmortem Intervals in relation to cause of Death', presented at the American Academy of Forensic Sciences, Boston, MA, February, 1993.
27. A.W.R. Forrest, Personal Communication, (1995) in R.O. Bost, 'Analytical Toxicology of Vitreous Humor', in *Analytical Toxicology of Vitreous Humor. Handbook of Analytical Therapeutic Drug Monitoring and Toxicology*, eds. S.H.Y. Wong, I. Sunshine, CRC Press, Boca Raton, FL, 291, 1997.
28. J. Lauridson, Personal Communication, (1995) in R.O. Bost, 'Analytical Toxicology of Vitreous Humor', in *Analytical Toxicology of Vitreous Humor. Handbook of Analytical Therapeutic Drug Monitoring and Toxicology*, eds. S.H.Y. Wong, I. Sunshine, CRC Press, Boca Raton, FL, 291, 1997.
29. N.E. Rosenthal, F.K. Goodwin, 'The Role of the Lithium Ion in Medicine', *Annu. Rev. Med.*, **33**, 555–568 (1982).
30. E.A. Wickham, J.V. Reed, 'Lithium for the Control of Aggressive and Self-mutilating Behavior', *Int. Clin. Psychopharmacol.*, **2**, 181–190 (1987).
31. K. Ekbom, 'Lithium for Cluster Headache: Review of the Literature and Preliminary Results of Long-term Treatment', *Headache*, **21**, 132–139 (1981).
32. R.S. Jope, M.B. Williams, 'Lithium and Brain Signal Transduction Systems', *Biochem. Pharmacol.*, **47**(3), 429–441 (1994).
33. R.C. Baselt, R.H. Cravey, *Disposition of Toxic Drugs and Chemicals in Man*, 4th edition, Chemical Toxicology Institute, Foster City, CA, 429–432, 1995.
34. A. Amdisen, 'Monitoring of Lithium Treatment through Determination of Lithium Concentration', *Dan. Med. Bull.*, **22**, 277–291 (1977).

35. A. Amdisen, C.G. Gottsfries, L. Jacobsson, B. Winbald, 'Grave Lithium Intoxication with Fatal Outcome', *Acta Psych. Scand. Supp.*, **4**, 25–33 (1974).
36. A. Amdisen, 'Clinical and Serum Level Monitoring in Lithium Therapy and Lithium Intoxication', *J. Anal. Tox.*, **2**, 193–202 (1978).
37. C.M. Swartz, P. Jones, 'Hyperlithemia Correction and Persistent Delirium', *J. Clin. Pharmacol.*, **34**, 865–870 (1994).
38. C. Swartz, 'Commentary on Encephalopathy Associated with Rapid Decrease of High Levels of Lithium', *Ann. Clin. Psych.*, **8**(2), 111–112 (1996).
39. J.C. Ayus, R.K. Krothapalli, A.I. Arieff, 'Treatment of Symptomatic Hyponatremia and its Relation to Brain Damage. A Prospective Study', *N. Engl. J. Med.*, **317**, 1190–1195 (1987).
40. T.K. Creson, P.J. Monaco, E.M. Rasch, A.N. Hagardorn, K.E. Ferslew, 'Capillary Ion Analysis of Lithium Concentration in Biological Fluids and Tissues of *Poecilia* (Teleost)', *Electrophoresis*, **19**, 3018–3021 (1998).

Chiroptical Spectroscopy in Drug Analysis

Harry G. Brittain
Center for Pharmaceutical Physics, Milford, USA

1 Introduction **1**
2 Circular Birefringence (Polarimetry) **2**
2.1 Background 2
2.2 Instrumentation 2
2.3 Application: Enantiomeric Identity and Enantiomeric Purity 2
3 Optical Rotatory Dispersion **3**
3.1 Background 3
3.2 Instrumentation 4
3.3 Application: the Octant Rule and Ketone Chirality 5
4 Circular Dichroism **5**
4.1 Background 5
4.2 Instrumentation 6
4.3 Application: Studies of Compounds of Pharmaceutical Interest 6
Abbreviations and Acronyms **8**
Related Articles **8**
References **8**

Many pharmaceutical compounds contain one or more centers of dissymmetry, thus presenting a unique series of regulatory and compendial requirements. The regulatory authorities are well acquainted with the fact that one given enantiomer or diastereomer of a drug entity may possess more efficacy (or fewer side effects) than its mirror image. Consequently, unless a compelling argument exists for the marketing of a racemate, it is more appropriate to develop enantiomerically pure therapeutic agents. Although most often characterized using chiral chromatography, these molecules can be effectively studied using the various techniques of chiroptical spectroscopy. Techniques that have been found to be very useful for such work include polarimetry (i.e. circular birefringence), optical rotatory dispersion (ORD), and circular dichroism (CD). At present, regulatory authorities remain fixated on the use of polarimetry to establish enantiomeric identity and purity, but as their awareness grows so will an acceptance of the more sophisticated chiroptical techniques. In this article, the principles underlying each chiroptical effect are briefly outlined and the application to problems of pharmaceutical interest is illustrated through the inclusion of appropriate examples.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Optical spectroscopy is generally concerned with measurements of the interaction of electromagnetic radiation (whose energy lies between 10 and 50 000 cm^{-1}) with isotropic (materials incapable of influencing the polarization state of light) or anisotropic (materials having the ability to affect the polarization properties of transmitted light) compounds. The various effects which are sensitive to the polarization state of radiant energy are ignored during the conduct of ordinary spectroscopy, but can be used to study a wide variety of phenomena for molecules lacking certain types of molecular symmetry.

Molecules for which the mirror images cannot be superimposed are denoted as being dissymmetric, or chiral, and these enantiomer structures are capable of being physically separated from each other. The fundamental requirement for the existence of molecular dissymmetry is that the molecule cannot possess an improper axis of rotation. This rule is often trivialized to state that the molecule in question cannot contain either a center of inversion or a reflection plane. The term dissymmetric should be used rather than asymmetric, since it is possible for a molecule to contain proper axes of rotation and still be capable of existing as nonsuperimposable mirror images.

Chiral molecules interact with electromagnetic radiation in exactly the same fashion as do achiral molecules in that they will exhibit optical absorption, have a characteristic refractive index and can scatter oncoming photons. Dissymmetric compounds will exhibit additional interactions with light whose electric vectors are circularly polarized. For example, the plane of linearly polarized light will undergo an apparent rotation (associated with the circular birefringence of the medium) as it passes through a chiral medium (polarimetry), and the magnitude of this rotation will depend on the wavelength of the light used ORD. Similarly, left- or right-circularly polarized light will be preferentially absorbed within the electronic transitions of a dissymmetric compound CD. General and readable introductions to molecular optical activity have been published by Charney[1] and Nakanishi et al.,[2] and a number of monographs exist which deal with chiroptical issues ranging from the theoretical to the practical.[3–11]

The pharmaceutical industry has made considerable use of polarimetry as a fundamental technique for

the characterization of enantiomerically resolved compounds, primarily owing to the specific requirements of regulatory authorities. The use of ORD and CD spectroscopy has been more along the lines of research studies, but these applications can be very important. CD spectroscopy, for example, is the technique of choice to establish the secondary structure of proteins and peptides.

2 CIRCULAR BIREFRINGENCE (POLARIMETRY)

2.1 Background

The study of molecular optical activity can be considered as beginning with the work of Biot, who demonstrated that the plane of linearly polarized light would be rotated upon passage through an optically active medium. Mitscherlich improved the method of measurement through the use of calcite prisms, and the double-field method of detection was introduced. Since these early developments, much advancement in polarimetry has been made, and a large number of detection schemes are now possible.

The velocity of light (v) passing through a medium is determined by the refractive index (n) of that medium (Equation 1):

$$v = \frac{c}{n} \tag{1}$$

where c is the velocity of light in a vacuum. For an achiral medium, the refractive index will not exhibit a dependence on the sense of the polarization state of the light. When the medium is chiral, however, the refractive index associated with left-circularly polarized light will not equal the refractive index associated with right-circularly polarized light, and the velocities of left- and right-circularly polarized light will differ on passage through a chiral medium. Since linearly polarized light is merely the vector resultant of two in-phase, oppositely signed, circularly polarized components, then the differing velocities of the components produce a phase difference as they pass through the chiral medium. Upon leaving the chiral medium, the components are recombined, and linearly polarized light is obtained whose plane is rotated (relative to the original plane) by an angle equal to half the phase angle difference of the circular components. Charney[1] has shown that this phase angle difference is given by Equation (2):

$$\theta = \frac{2\pi b'}{\lambda_0}(n_L - n_R) \tag{2}$$

where θ is the phase difference, b' is the medium pathlength (in centimeters), λ_0 is the vacuum wavelength of the light used and n_L and n_R are the refractive indices for left- and right-circularly polarized light, respectively. The quantity $n_L - n_R$ defines the circular birefringence of the chiral medium, and this quantity is the origin of what is commonly referred to as optical rotation.

2.2 Instrumentation

The experimental measurement of optical rotation is extremely simple. The incident light is collimated and plane-polarized, and then passed through the medium under study. Since the polarization plane of the incident light is fixed, the angle of rotation is defined with respect to this original plane. This is carried out by first determining the orientation of the polarizer and analyzer for which no light can be transmitted (the null position), and then finding the new null position after the medium containing the optically active material is introduced between the prisms. The observed angle of rotation is taken as the difference between the two null angles, and a variety of measurement schemes have been developed to obtain the best experimental results. Those more interested in the experimental measurement of circular birefringence should consult the extensive summary of methodology that has been provided by Heller.[12]

2.3 Application: Enantiomeric Identity and Enantiomeric Purity

The use of polarimetry data for the definitive probing of chemical problems has been effectively superseded by the use of the spectroscopic techniques yet to be discussed. However, this trend has been resisted somewhat by the pharmaceutical industry, where regulatory agencies normally require measurements of specific rotation as one proof of chiral identity. In fact, the measurement of optical rotation (general test ⟨781⟩) is the only chiroptical test quoted in the United States Pharmacopeia,[13] which ensures a continued interest in polarimetry. For example, Chafetz has provided a compilation of the specific rotation values obtained for a very extensive list of steroids.[14]

In principle, the determination of enantiomeric identity by polarimetry is extremely simple. Knowing the sign of the optical rotation for a given enantiomer of a particular compound enables the enantiomers of that compound to be readily differentiated. For instance, (*R*)-glycerylaldehyde will exhibit positive rotation, whereas (*S*)-glycerylaldehyde exhibits negative optical rotation. The enantiomer present to the largest extent can therefore be readily determined in unknown samples of glycerylaldehyde.

Unfortunately, the assignment of enantiomeric identity by polarimetry is filled with difficulties and ambiguities unless the conditions of measurement are exactly specified. For instance, L-asparagine will exhibit negative optical rotation when dissolved in water, but positive rotation in dilute hydrochloric acid.[15] When compounds contain more than one dissymmetric atom, the sign of the observed optical rotation may not enable one to specify fully the relative configurations of all chiral atoms present. Consequently, unless the measurements are performed on highly defined situations, the determination of enantiomeric identity is best accomplished using a spectroscopic or chromatographic technique.

Polarimetry can be extremely useful, however, in the determination of the enantiomeric purity of a resolved or partially resolved compound. In fact, the commonly misused term optical purity properly relates only to the determination of enantiomeric purity by polarimetry. The term enantiomeric purity itself has no clear definition, and its use differs from investigator to investigator. One term used to define the enantiomer composition is the enantiomer ratio (Equation 3):

$$ER = \frac{\%[R]}{\%[S]} \tag{3}$$

where $\%[R]$ and $\%[S]$ denote the relative percentages of the two enantiomers. A much more useful term in the enantiomeric excess, defined according to Equation (4):

$$EE = \frac{[R] - [S]}{[R] + [S]} \tag{4}$$

where $[R]$ and $[S]$ represent the concentrations of the two enantiomers. The enantiomeric excess is also given by Equation (5):

$$EE = \%[R] - \%[S] \tag{5}$$

3 OPTICAL ROTATORY DISPERSION

3.1 Background

In his fundamental studies on tartaric acid, Biot found that the magnitude of observed optical rotation varied with the wavelength of light used for the measurement.[16] The wavelength dependence was found to be influenced by the nature of the solvent used and by the presence of certain cosolutes, and Biot recognized that such investigations could be used to study the interaction phenomena which existed in fluid solution. Until this time, optical rotatory power had been viewed as an inherent fundamental property of a dissymmetric molecule, whose exact sign and magnitude set that compound apart from all others. These findings naturally sparked intense studies.

Biot's discovery of the wavelength dependence of optical rotation, and his explanation for the phenomena that he observed, followed earlier observations that the optical rotation exhibited by a quartz plate also varied with wavelength. As a result, the measurement of optical rotation has essentially been a study of ORD since the very beginning.

3.1.1 Ordinary Dispersion

Biot's earliest quantitative measurements showed that the optical rotation (α) of quartz varied inversely as the square of the wavelength (λ) used (Equation 6):

$$\alpha = \frac{A}{\lambda^2} \tag{6}$$

where the constant of proportionality (A) is a fundamental property of the dissymmetric molecule. Biot subsequently divided all optically active substances into two classes, one of which obeyed his inverse square law and the other containing those that violated the law. Improvements were later suggested by von Lang,[17] who added an additional molecular constant (B) to improve the fit with experimental data (Equation 7):

$$\alpha = B + \frac{A}{\lambda^2} \tag{7}$$

Others continued to deduce modifications to the law of inverse squares, but none of these proved to be completely satisfactory.

Investigators quickly became aware that the compounds which violated the inverse square law were those which exhibited light absorption at the wavelength of investigation. The relationship between optical rotatory power and wavelength outside of regions of absorption was expressed by Drude,[18] who introduced the equation that bears his name:

$$[\phi] = \frac{K}{\lambda^2 - \lambda_0^2} \tag{8}$$

where K is a molecular constant, $[\phi]$ is the total molecular rotation at the given wavelength λ and λ_0 is the wavelength of maximum absorbance for the chromophore. The Drude equation works fairly well for many organic compounds over the range of 450–700 nm, since their absorption bands are generally located in the ultraviolet (UV) region of the spectrum.

Lowry has provided an extensive summary of the variation of optical rotation with wavelength, temperature, solvent and concentration for numerous organic liquids.[19] As an example of ordinary dispersion, data reported for the methyl and ethyl ethers of *d*-benzylmethylcarbinol

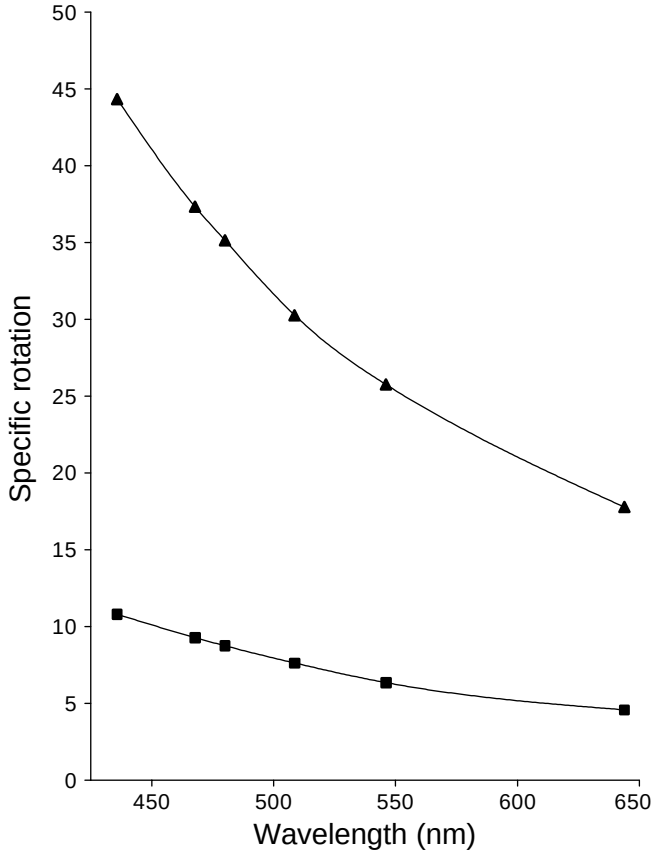

Figure 1 Ordinary dispersion curves obtained for the methyl (■) and ethyl ( ) ethers of *d*-benzylmethylcarbinol. (The data were plotted from the table given in Lowry.[19])

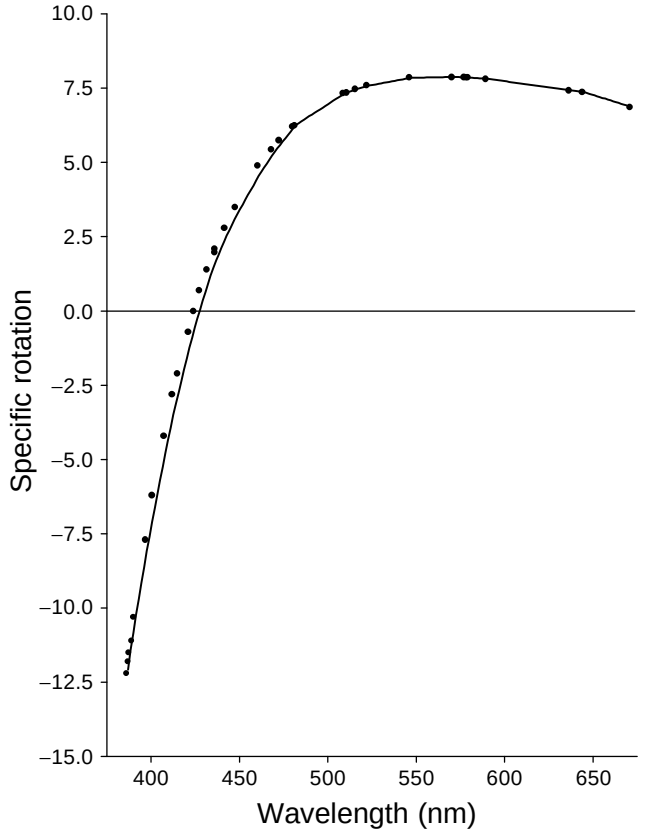

Figure 2 ORD spectrum obtained for ethyl tartrate. (The data were plotted from a table given in Lowry and Dickson.[20])

are plotted in Figure 1. The optical rotations of both compounds exhibit the typical dispersion which is characteristic of circular birefringence outside of absorption bands.

3.1.2 Anomalous Dispersion

As just discussed, the magnitude of circular birefringence increases with decreasing wavelength. This behavior persists until the light is capable of being absorbed by the chiral substance, whereupon the refractive index exhibits anomalous behavior. The anomalous dispersion observed in ORD spectra arises since the refractive index of a material is the sum of a real and imaginary part (Equation 9):

$$n = n_0 + \mathrm{i}k \tag{9}$$

where n is the observed refractive index at some wavelength, n_0 is the refractive index at infinite wavelength and k is the absorption coefficient of the substance. It has been amply demonstrated that if $n_L - n_R$ does not equal zero, then $k_L - k_R$ will not equal zero either.

The optical rotation of ethyl tartrate was found to exhibit anomalous dispersion, and attracted much study.[20] The original data of Lowry and Dickson have been plotted in Figure 2, where one may note the existence of ordinary dispersion between 550 and 675 nm. As the wavelength is decreased further, the magnitude of optical rotation rapidly decreases and changes sign below 425 nm. The data could be fitted reasonably well by modifying the Drude equation to contain a second term whose effect opposed that of the first (Equation 10):

$$\alpha = \frac{K_1}{\lambda^2 - 0.026} - \frac{K_2}{\lambda^2 - 0.061} \tag{10}$$

where K_1 and K_2 are molecular constants and α is the observed specific rotation at the given wavelength λ.

3.2 Instrumentation

An ORD spectrometer in its simplest form would consist of a manual polarimeter through which light of various wavelengths is passed. One measures the optical rotation at each incident wavelength, and obtains the ORD spectrum by plotting the observed values. In an

early improvement, the incident polarizer was rocked through a small angle by an electric motor at low frequency, the passed light measured by a photoelectric cell and the analyzer rotated until a null position was again achieved. Recording ORD spectrometers have been developed which either use the Faraday effect to obtain both modulation and balance or which use servo-driven analyzers to determine the null position. The calibration of ORD spectrometers is verified using the same standards as used for the calibration of ordinary polarimeters. Those seeking more detailed information regarding ORD instrumentation should consult the book by Djerassi.[3]

3.3 Application: the Octant Rule and Ketone Chirality

The earliest chiroptical work involving chiral organic molecules was entirely based on ORD methods, since little else was available at the time. One of the largest data sets collected to date concerns the chirality of ketone and aldehyde groups,[21] which eventually resulted in the deduction of the octant rule.[22] The octant rule was an attempt to relate the absolute stereochemistry within the immediate environment of the chromophore with the sign and intensity of the ORD Cotton effects.

To apply the octant rule, the ORD within the n $\rightarrow \pi^*$ transition around 300 nm is obtained, and its sign and intensity noted. The rule developed by Djerassi and co-workers states that the three nodal planes of the n- and π^*-orbitals of the carbonyl group divide the molecular environment into four front octants and four back octants. A group or atom situated in the upper-left or lower-right rear octant (relative to an observer looking at the molecule parallel to the C=O axis) induces a positive Cotton effect in the n $\rightarrow \pi^*$ band. A negative Cotton effect would be produced by substitution within the upper-right or lower-left back octant. The use of the octant rule is illustrated in Figure 3(a) and (b) for the particular example of a generic 3-hydroxy-3-alkylcyclohexanone.

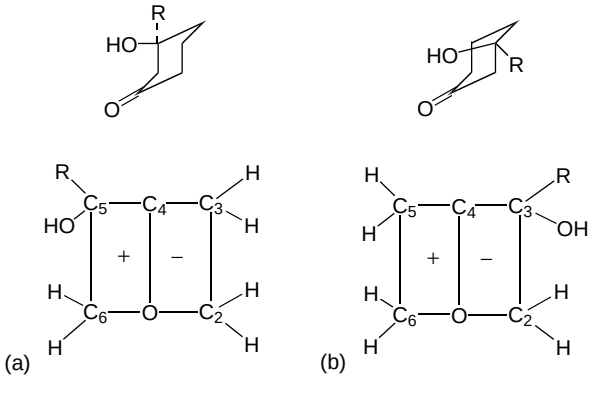

Figure 3 Application of the octant rule, applied to the generic enantiomers of 3-hydroxy-3-alkylcyclohexanone.[20] (a) Positive and (b) negative Cotton effect.

Although exceptions to the octant rule have been shown, its wide applicability has remained established. The ability to deduce molecular conformations in solution on the basis of ORD spectra data has proven to be extremely valuable to synthetic and physical organic chemists, and enabled investigators of the time to develop their work without requiring the use of more heroic methods.

As an example, one may consider the case of *trans*-10-methyl-2-decalone, which has been reported to exhibit a positive Cotton effect.[23] Placing the molecule in its all-chair conformation and then viewing the molecule along the direction of its C=O bond enable one to deduce the position of the molecular groups in the various quadrants. The substituents of C-10 (the angular methyl group and C-5) lie in a vertical plane, and are ignored. There are no axial substituents on C-1 and C-3, placing no contributions in the lower-left or lower-right octants. C-4 bears no substituents, hence there is no contribution from the upper-right octant. However, there are three carbon atoms (C-6, C-7 and C-8) in the upper-left octant (which would make a positive contribution to the ORD), leading one to conclude that the overall sign of the Cotton effect would be positive.

4 CIRCULAR DICHROISM

4.1 Background

As just discussed, circular birefringence and ORD are easily interpretable outside regions of electronic absorption, but exhibit anomalous properties within absorption bands. This effect arises since the refractive index also contains a contribution related to molecular absorptivity, as discussed earlier. Not only will the phase angle between the projections of the two circularly polarized components be altered by passage through the chiral medium, but also the amplitudes of these components will be modified by the degree of absorption experienced by each component. This differential absorption of left- and right-circularly polarized light is termed CD, and is given by $k_L - k_R$.

The effect of one circularly polarized component being more strongly absorbed than the other is that when the projections are recombined after leaving the chiral medium, they no longer produce plane polarized light. Instead, the resulting components describe an ellipse, the major axis of which lies along the angle of rotation. The measure of the eccentricity of the ellipse that results from the differential absorption is termed the ellipticity, ψ. It

is not difficult to show that[1] (Equation 11)

$$\psi = \frac{\pi z}{\lambda}(k_{\mathrm{L}} - k_{\mathrm{R}}) \qquad (11)$$

where z is the pathlength in centimeters and λ is the wavelength of the light.

4.2 Instrumentation

Most instrumentation suitable for measurement of CD is based on the design of Velluz et al.[5] Linearly polarized light is passed through a dynamic quarter-wave plate, which modulates the beam alternately into left- and right-circularly polarized light. The quarter-wave plate is a piece of isotropic material, which is rendered anisotropic through the external application of stress. The device can be a Pockels cell (in which stress is created in a crystal of ammonium dideuterium phosphate through the application of ac high voltage), or a photoelastic modulator (in which the stress is induced by the piezoelectric effect). The light leaving the cell is detected by a photomultiplier tube, whose current output is converted into voltage and split. One signal consists of an ac signal proportional to the CD, and is due to the differential absorption of one component over the other. This signal is amplified by means of phase-sensitive detection. The other signal is averaged, and is related to the mean light absorption. The ratio of these signals varies linearly as a function of the CD amplitude, and is the recorded signal of interest.

The calibration of CD spectrometers has been approached most often by the use of reference standard materials, with various salts of *d*-10-camphorsulfonic acid having received the greatest attention. Highly accurate CD spectra of the free acid,[24,25] the *n*-propylammonium and *n*-butylammonium salts,[26] and the tris(hydroxymethyl)aminomethane salt[27] have been published, and the relative virtues of each material discussed in detail.

4.3 Application: Studies of Compounds of Pharmaceutical Interest

It would be impossible to summarize the utility of CD spectroscopy for the study of molecular stereochemistry, but the cited monographs are sufficient for this purpose.[1–11] Being a variant of absorption spectroscopy, the technique can be used in exactly the same manner, with the differential absorption following its own analog of Beer's law. In its simplest application, the CD spectrum of a given compound can be used for diagnostic purposes, as was demonstrated in the identification of morphine, codeine, thebaine and noscapine.[28] Being proportional to concentration, CD intensities can be used to evaluate the concentration of drug substances in pharmaceutical dosage forms.[29]

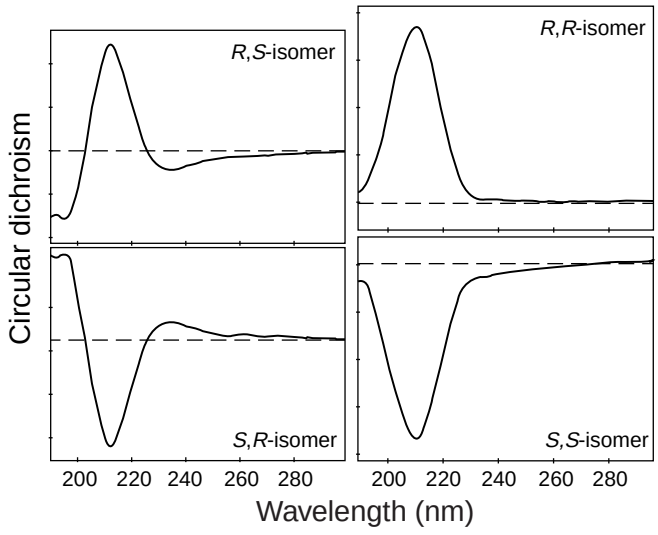

Figure 4 CD spectra obtained on the four diastereomers of captopril, 1-[(2*S*)-3-mercapto-2-methylpropionyl]-(*S*)-proline.[30] The diastereomers are identified first by the absolute configuration of the side chain, followed by the absolute configuration of the proline group.

As one example of the use of CD spectroscopy for the characterization of compounds of pharmaceutical interest, the CD of highly purified captopril {1-[(2*S*)-3-mercapto-2-methylpropionyl]-(*S*)-proline}, its component fragments and its three other diastereomers have been reported.[30] The UV absorption spectrum of captopril consists of a single band maximum at 200 nm, while the CD spectrum consists of a single negative peak located at 210 nm. The CD spectra obtained for captopril and its three diastereomers are illustrated in Figure 4, and can be explained largely in terms of a summation of the chirality of the individual components, (*S*)-proline and (2*S*)-3-mercapto-2-methylpropionic acid. However, this summation was found to be only approximate since the individual chiralities were found to be enhanced in the diastereomer compounds.

According to Purdie,[31] natural products that are amenable to CD analysis would include alicyclic compounds (e.g. alkaloids and terpenes), heterocyclic compounds (e.g. barbiturates, diazepams, indole alkaloids, quinolines, nucleic acids and nucleotides), amino acids and peptides, oligopeptides, and proteins and sugars (saccharides and polysaccharides). In addition, the technology may also be used to characterize the many molecules that are the metabolites and/or the naturally occurring condensation products of carbohydrates with the other categories of compounds, such as glucuronides and glycosides of these.

One early application of CD spectroscopy for the characterization of pharmaceutically important compounds concerned studies of aqueous solutions of various morphine alkaloids.[32] Although the most intense optical activity was observed at wavelengths shorter than 230 nm, the CD bands most useful for discrimination among the structural analogs occurred between 240 and 300 nm. Spectra obtained for some analogous compounds containing phenolic groups were found to invert depending on whether the phenol was ionized or not, a feature that was exploited for the analysis of mixtures. Heroin could be identified in real samples, since the CD enabled one to discriminate the chiral substance from its matrix.[33]

CD spectroscopy has been useful in the study of various antibiotics, since many of these contain one or more centers of dissymmetry. In one particular example, the strong optical activity exhibited in the visible region by members of the tetracycline family has been reported.[34] One may readily differentiate between the parent tetracycline and its degradant 4-epitetracycline, and this selectivity has been applied to the analysis of commercial products.[35] Penicillin-V and cephalothin were simultaneously determined in buffered solutions of prepared laboratory mixtures, and penicillin-V was easily measured in various media.[36]

Discrimination between penicillins and cephalosporins is nearly impossible unless the compounds are derivatized prior to the CD analysis.

CD spectroscopy was used to evaluate the ability of a large number of different solvents to influence the aggregation state of amphotericin B.[37] Using the relative donor/acceptor tendencies known for each solvent system, the factors leading to self-association were deduced. In the absence of strong electron-pair–solvent interaction, amphotericin B prefers to self-associate into oligomeric species. As illustrated in Figure 5, the CD spectra were extremely sensitive toward the degree of aggregation, and the equilibrium could be shifted to any desired position depending on the exact composition of the solvent medium.

CD spectra for the diastereomeric pairs quinine–quinidine and cinchonine–cinchonidine were found to be the necessary mirror images of each other, and the compositions of mixtures of these were determined.[38] Spectra for the pilocarpine–isopilocarpine pair were of such low quality that they could be used only for qualitative distinction. In the same study, CD detection combined with UV detection was used to measure enantiomeric excesses in mixtures of L-hyoscyamine and atropine (i.e. racemic hyoscyamine).

Without question, the largest body of chiroptical work on compounds of pharmaceutical interest has been performed on the steroids.[39] Although these compounds contain multiple centers of dissymmetry and

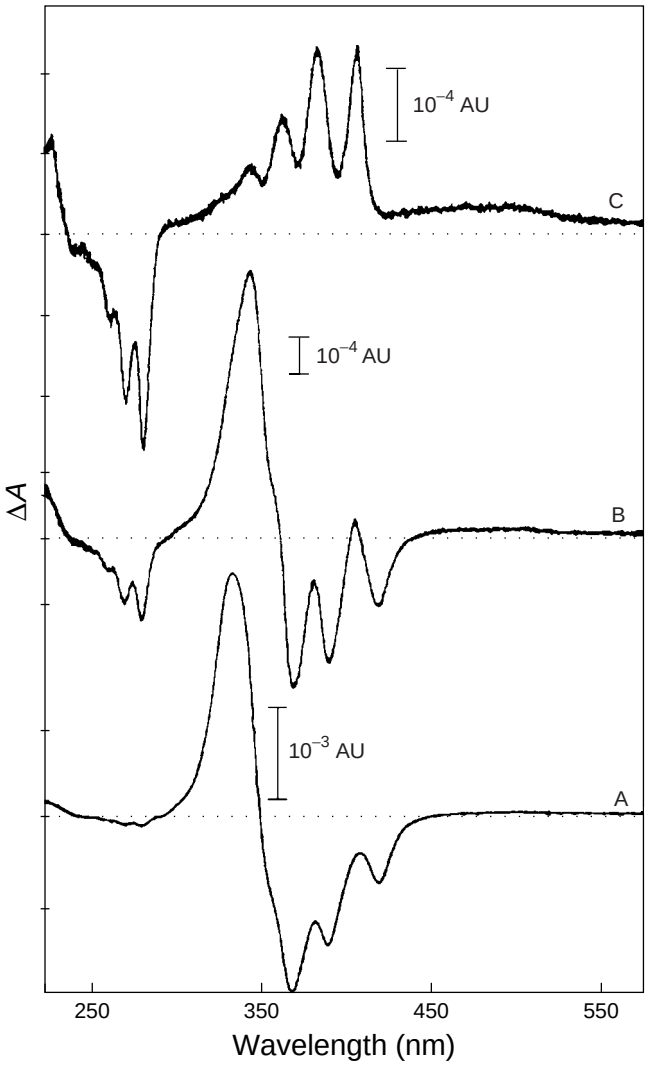

Figure 5 CD spectra obtained of amphotericin B in mixed solvents consisting of (A) 90:10, (B) 80:20 and (C) 70:30 acetonitrile–methanol.[37]

theoretically are amenable to polarimetric study, many present a difficult detection problem owing to the small magnitude of the observed effects. This problem is severe for compounds lacking accessible chromophores. For instance, the only unsaturation in cholesterol is the Δ^4-double bond, which exhibits an absorption maximum around 205 nm.

Of course, the presence of a chromophore ordinarily yields strong and useful CD, as was previously discussed for ketone groups that permitted the development of the octant rule. CD spectroscopy has been use to characterize and determine a large number of steroids, such as the Δ^4-3-keto steroids and 17-keto and 17-ethynyl compounds.[29] Members of the latter series are normally prepared from their corresponding keto compounds, which then presents an impurity species to detect and

quantitate. CD spectroscopy has proven to be an excellent technique for such work.

One very useful approach to the analysis of mixtures is to enhance the selectivity of the chiroptical method further by measuring spectra of the analyte in two different solvent systems, and to use the spectral differences for quantitative analysis. This technique has been applied to the determination of corticosteroids and Δ^4-3-keto steroids in pharmaceutical products.[29] Simple derivatization reactions will also affect the spectra in many ways, often resulting in extra selectivity through shifts in the absorption wavelengths, sign reversals and the appearance or disappearance of Cotton effects. One often obtains sufficient selectivity to permit the analysis of binary mixtures using measurements made at only two wavelengths.

In addition to the examples just discussed, a variety of other pharmaceutically active compounds have been characterized by means of CD spectroscopy, including penicillin,[40] tetracycline,[41] anhydrotetracycline,[42] nicotine,[43] various cannabinoids,[44] various corticosteroids,[45] reserpine,[46] ampicillin,[47] naproxen[48] and various cephalosporins.[49]

ABBREVIATIONS AND ACRONYMS

CD	Circular Dichroism
ORD	Optical Rotatory Dispersion
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Circular Dichroism in Analysis of Biomolecules • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Circular Dichroism and Linear Dichroism

REFERENCES

1. E. Charney, *The Molecular Basis of Optical Activity*, John Wiley & Sons, New York, 1979.
2. K. Nakanishi, N. Berova, R.W. Woody, *Circular Dichroism: Principles and Applications*, VCH Publishers, New York, 1994.
3. C. Djerassi, *Optical Rotatory Dispersion: Applications to Organic Chemistry*, McGraw-Hill, New York, 1960.
4. P. Crabbe, *ORD and CD in Chemistry and Biochemistry*, Academic Press, New York, 1972.
5. L. Velluz, M. Legrand, M. Grosjean, *Optical Circular Dichroism: Principles, Measurements, and Applications*, Verlag Chemie, Weinheim, 1965.
6. G. Snatzke (ed.), *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*, Heyden, London, 1967.
7. K. Nakanishi, *Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry*, University Science Books, Mill Valley, CA, 1983.
8. D.J. Caldwell, H. Eyring, *The Theory of Optical Activity*, Wiley-Interscience, New York, 1971.
9. L. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
10. S. Mason, *Molecular Optical Activity and the Chiral Discrimination*, Cambridge University Press, Cambridge, 1982.
11. N. Purdie, H.G. Brittain (eds.), *Analytical Applications of Circular Dichroism*, Elsevier, Amsterdam, 1994.
12. W. Heller, 'Optical Rotation – Experimental Techniques and Physical Optics', in *Physical Methods of Chemistry*, eds. A. Weissberger, B.W. Rossiter, John Wiley & Sons, New York, 51–181, Chapter 2, Vol. I, 1972.
13. 'Optical Rotation', General Test ⟨781⟩ in *United States Pharmacopeia*, 23rd edition, United States Pharmacopeial Convention, Rockville, MD, 1812–1813, 1995.
14. L. Chafetz, 'Implementing Changes in Polarimetry – Steroid Drugs', *Pharmacop. Forum*, **19**, 6159–6162 (1993).
15. H. Landolt, *The Optical Rotating Power of Organic Substances and its Practical Applications*, Chemical Pub. Co., Easton, PA, 541–542, 1902.
16. J.B. Biot, 'Mémoire sur la Polarisation Circulaire et sur ses Applications à la Chimie Organique', *Mem. Acad. Sci.*, **13**, 39–175 (1835); 'Methodes Mathématiques et Expérimentales pour Discerner les Mélanges et les Combinaisons Chimiques Défines or Non-Défines qui Agissent sur la Lumière Polarisée, Suivies d'Applications aux Combinaisons de l'Acide Tartrique avec l'Eau, l'Alcool, et l'Esprit de Bois', **15**, 93–279 (1838); 'Mémoire sur Plusiers Points Fondamentaux de Méchanique Chimique', **16**, 229–396 (1838).
17. T.M. Lowry, 'Complex and Anomalous Rotatory Dispersion', in *Optical Rotatory Power*, Longmans, Green, and Co., London, 136–148, Chapter 10, 1935.
18. P. Drude, *Lehrbuch der Optik*, 2nd edition, S. Hirsel Verlag, Leipzig, 1906.
19. T.M. Lowry, *The Optical Rotation of Liquids: Its Variation with Wavelength, Temperature, Solvent, and Concentration*, US Government Printing Office, Washington, DC, 7, 1932.

20. T.M. Lowry, T.W. Dickson, 'Rotatory Power of Organic Compounds. VI. Complex Rotatory Dispersion of Ethyl Tartrate', *J. Chem. Soc.*, 1173–1187 (1915).
21. D.N. Kirk, 'The Chiroptical Properties of Carbonyl Compounds', *Tetrahedron*, **42**, 777 (1986).
22. W. Moffitt, R.B. Woodward, A. Moscowitz, W. Klyne, C. Djerassi, 'Structure and the Optical Rotatory Dispersion of Saturated Ketones', *J. Am. Chem. Soc.*, **83**, 4013–4018 (1961).
23. C. Djerassi, R. Riniker, B. Riniker, 'Optical Rotatory Dispersion Studies. VII. Application to Problems of Absolute Steroid Configurations', *J. Am. Chem. Soc.*, **78**, 6362–6377 (1956).
24. D.F. DeTar, 'Suggested Preliminary Standards for Calibration of Optical Rotatory Dispersion and Circular Dichroism Instruments', *Anal. Chem.*, **41**, 1406–1408 (1969).
25. G.C. Chen, J.T. Yang, 'Two-point Calibration of Circular Dichrometer with *d*-10-Camphorsulfonic Acid', *Anal. Lett.*, **10**, 1195–1207 (1977).
26. M.F. Gillen, R.E. Williams, 'The Use of the *n*-Propylammonium and *n*-Butylammonium Salts of *d*-10-Camphorsulfonic Acid for the Calibration of Spectropolarimeters: Circular Dichrometer Calibration', *Can. J. Chem.*, **53**, 2351–2353 (1975).
27. K.H. Pearson, V.C. Zadnik, J.L. Scott, 'Calibration of Circular Dichrometers', *Anal. Lett.*, **12**, 1049–1061 (1979).
28. S.M. Han, N. Purdie, 'Simultaneous Determination of Opiates by Circular Dichroism', *Anal. Chem.*, **58**, 113–116 (1986).
29. A. Gergely, 'A Review of the Application of Chiroptical Methods to Analytical Chemistry', *J. Pharm. Biomed. Anal.*, **7**, 523–541 (1989).
30. H.G. Brittain, H. Kadin, 'Ultraviolet Absorption and Circular Dichroism Spectra of Captopril', *Pharm. Res.*, **7**, 1082–1085 (1990).
31. N. Purdie, 'Analytical Applications of CD to Forensic, Pharmaceutical, Clinical, and Food Sciences', in *Analytical Applications of Circular Dichroism*, eds. N. Purdie, H.G. Brittain, Elsevier, Amsterdam, 241–278, Chapter 8, 1994.
32. T.A. Crone, N. Purdie, 'Circular Dichroism Spectra of Opium Alkaloids in Aqueous Media', *Anal. Chem.*, **53**, 17–21 (1981).
33. J.M. Bowen, T.A. Crone, R.K. Kennedy, N. Purdie, 'Determination of Heroin by Circular Dichroism', *Anal. Chem.*, **54**, 66–68 (1982).
34. L.A. Mitscher, *Chemistry of the Tetracycline Antibiotics*, Marcel Dekker, New York, Vol. 13, 1978.
35. R.F. Miller, T.D. Sokoloski, L.A. Mitscher, A.C. Bonacci, B.A. Howner, 'Use of Circular Dichroism in Analysis of Mixtures of Tetracycline and 4-Epitetracycline and its Application to Assay of Commercial Products', *J. Pharm. Sci.*, **62**, 1143–1147 (1973).
36. N. Purdie, K.A. Swallows, 'Direct Determination of β-Lactam Antibiotics by Circular Dichroism', *Anal. Chem.*, **59**, 1349–1351 (1987).
37. H.G. Brittain, 'Circular Dichroism Studies of the Self-association of Amphotericin B', *Chirality*, **6**, 665–669 (1994).
38. S.M. Han, N. Purdie, 'Determination of Stereoisomers by Circular Dichroism', *Anal. Chem.*, **58**, 455–458 (1986).
39. A. Gergely, 'Application of Circular Dichroism to the Determination of Steroids', in *Analytical Applications of Circular Dichroism*, eds. N. Purdie, H.G. Brittain, Elsevier, Amsterdam, 293–305, Chapter 10, 1994.
40. C.E. Rasmussen, T. Higuchi, 'Spectropolarimetric and Circular Dichroism Methods for Determining the Activity of Penicillin', *J. Pharm. Sci.*, **60**, 1608–1616 (1971).
41. E.C. Newman, C.W. Frank, 'Circular Dichroism Spectra of Tetracycline Complexes with Mg(II) and Ca(II)', *J. Pharm. Sci.*, **65**, 1728–1732 (1976).
42. L.J. Stoel, E.C. Newman, G.L. Asleson, C.W. Frank, 'Potentiometric and Spectral Investigations of Anhydrotetracycline and Its Metal-ion Complexes', *J. Pharm. Sci.*, **65**, 1794–1799 (1976).
43. W.M. Atkinson, S.M. Han, N. Purdie, 'Determination of Nicotine in Tobacco by Circular Dichroism Spectropolarimetry', *Anal. Chem.*, **56**, 1947–1950 (1984).
44. S.M. Han, N. Purdie, 'Determination of Cannabinoids by Circular Dichroism', *Anal. Chem.*, **57**, 2068–2071 (1985).
45. A. Gergely, Gy. Szasz, J. Soos, 'New Method for the Selective Determination of Corticosteroids by Means of Difference CD Spectroscopy', *Fresenius' J. Anal. Chem.*, **323**, 157–161 (1986).
46. K.A. Swallows, N. Purdie, 'Determination of Reserpine by Circular Dichroism', *Pharm. Res.*, **6**, 521–524 (1989).
47. S.-J. Chen, H. Aki, M. Yamamoto, 'Characteristics of Circular Dichroism Spectrum and Degradation Mechanism of Ampicillin in Various Buffer Solutions', *Chin. Pharm. J.*, **42**, 423–433 (1990).
48. S. Wenzel, V. Buss, 'Circular Dichroism and Electronic Structure Calculations on Naproxen', *J. Phys. Org. Chem.*, **5**, 748–754 (1992).
49. P. Gortazar, J.T. Vaquez, 'Discrimination and Direct Determination of Cephalosporins by Circular Dichroism', *J. Pharm. Sci.*, **83**, 1204–1208 (1994).

DNA Extraction Methods in Forensic Analysis

Cristina Cattaneo
Instituto di Medicina Legale, Universita' degli Studi di Milano, Milan, Italy

K. Gelsthorpe and R.J. Sokol
National Blood Service, Sheffield, UK

1 Introduction: Problems Concerning the Extraction of DNA **2**
2 Possible Forensic Substrates for DNA Analysis, Their Peculiarities and Preparation for Extraction **4**
2.1 Fresh Materials with Abundant DNA 4
2.2 Putrefied Materials with Abundant DNA 4
2.3 Clean Material with Scarce DNA 5
2.4 Dirty Material with Scarce DNA 6
2.5 Formalin-fixed and Paraffin-embedded Tissues, and Stained Smears on Slides 6
2.6 Calcified Tissues (Bone and Teeth) 7
3 DNA Extraction **7**
3.1 Cell Lysis, Protein Digestion and Solubilization of DNA 8
3.2 Nonaffinity Extraction Methods 9
3.3 Affinity Methods 15
3.4 Commercial Methods 18
4 Conclusions **18**
Acknowledgments **18**
Abbreviations and Acronyms **19**
Related Articles **19**
References **19**

Advances in molecular biology can be applied to forensic practice with spectacular success. However, the extraction of DNA from scarce or badly degraded substances remains a significant problem and an ability to overcome it remains crucial because the success of all DNA analyses depend on a successful initial extraction. Extraction methods must solve problems associated with low quantities of DNA, DNA degradation, polymerase chain reaction (PCR) inhibitors and failure to solubilize or separate DNA from the substrate. These problems are compounded by the relative lack of knowledge concerning DNA degradation in different forensic environments.

The extraction method must be suitable for the starting material. Fresh material with abundant DNA, such as blood, soft tissue, or saliva, presents no difficulties and DNA suitable for PCR analysis can be extracted easily. In mixed substrates, for instance vaginal swabs from rape victims, male and female DNA can be separated by differential lysis procedures. Putrefied material with abundant DNA, such as blood, soft tissues, and entomological and botanical material, has plenty of DNA, but PCR inhibitors may be present and extraction methods must be chosen that select DNA but not the inhibitors. Clean material with scarce DNA, such as stains (blood, saliva, sperm), swabs (vaginal and from bitemarks, fingernails and stains on bodies), fingernails, hair, desquamated cells (skin, dandruff), and certain body liquids (urine, plasma), has a paucity of cellular material and hence low quantities of DNA. The prerequisite here is to reduce the loss of DNA, for example by introducing more specific methods of extraction and reducing the number of steps in the procedure. In the case of dirty material with scarce DNA, such as sperm, saliva and blood stains, nails and feces, not only are there low quantities of DNA, but there are also large amounts of organic and inorganic contamination. The problems faced by any extraction method are to recuperate small quantities of DNA as well as eliminating inhibitors of PCR. Formalin-fixed and paraffin-embedded tissues and stained smears on slides present particular problems associated with damage to DNA due to the processing of the materials. Nevertheless, DNA can be successfully extracted from these substrates. The greatest challenge in forensic DNA analysis is extraction from calcified tissues (bone and teeth). These substrates have obviously undergone considerable degradation and are often contaminated with soil. In addition, DNA undergoes binding interactions with the hydroxyapatite matrix which chemically stabilizes DNA; this aids its survival, but makes it difficult to extract. An essential part of the extraction process is therefore to release DNA from this binding.

Before DNA can be isolated from a substrate, several important steps must be taken. These involve lysing the cells to release the DNA and digesting protein so that as much DNA as possible is solubilized. Many (if not most) techniques – and there are many variations – involve the use of proteinase K (PK) to digest protein and a detergent (often sodium dodecyl sulfate (SDS)) to lyse the cell membranes. The extraction procedures themselves can be divided into two categories: nonaffinity methods where all extraneous material is removed leaving DNA in solution, and affinity methods whereby the DNA is targeted directly and removed from the solution. Each category has many variations and on occasions they can be combined. The

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

main nonaffinity techniques are phenol/chloroform, salt precipitation and chelex. The phenol/chloroform method is the best known and has an excellent forensic record; it is based on removing protein, thus purifying the nucleic acids which are extracted in aqueous solution; the disadvantages are that it does not remove nonprotein contaminants, several steps are involved in the procedure, and the reagents are toxic. The salt precipitation method is based on salting-out by dehydration and precipitation with a saturated salt solution. It is an excellent method when abundant clean samples are available and is quicker, cheaper and less toxic than the phenol/chloroform procedure. The disadvantages are that it does not eliminate all PCR inhibitors and considerable amounts of protein may be retained. With chelation (e.g. chelex), the active reagent binds to inorganic substances and clears the extract of certain PCR inhibitors, particularly metal ions; an alkaline pH disrupts cell membranes leaving DNA in solution. The method is simple, rapid and does not involve multiple tube transfers, though it does not have a particularly selective action on protein. The novelty of affinity methods is that they actively select DNA by either specific or nonspecific reactions. The main nonspecific affinity method is the glass–milk or silica gel procedure. Very small glass beads are combined with isothiocyanate which itself is bound to guanidium. The positively charged amine groups on the guanidium link with the negatively charged phosphate groups on DNA. Once bound to the glass beads, the DNA can be repeatedly washed and cleansed of all other material before being eluted and subjected to PCR analysis. This is currently the method of choice for difficult substrates; the disadvantage is that when a great deal of contaminant is present, it might sterically hinder the adsorption process of DNA on to the glass substrate. The only specific affinity method yet published uses a murine anti-dsDNA monoclonal antibody of IgG2A subclass bound to paramagnetic beads via anti-murine IgG. The antibody-coated beads react specifically with DNA and can easily be washed free of contaminants. The method is extremely specific and sensitive, positive PCR results being obtainable from as little as 0.005 ng per microliter of DNA. In addition, the bead–anti-DNA–DNA complex can be put directly into the PCR. As yet this method has not been fully tested on all forensic substrates, but initial results are very promising. Commercial DNA extraction kits, for example Dynal DNA direct™, QIAamp®, Prep-A-Gene® and Scotlab Geneclean®, are becoming increasingly available and some provide protocols for forensic substrates. Most of these kits are based on an affinity mechanism. In practice they seem to give reliable results.

In conclusion, DNA extraction methods in forensic analysis seem to be going two ways. For substrates with fresh abundant DNA, extraction presents no problems and any method would suffice; it would seem logical to choose one that uses nontoxic reagents and is inexpensive. The most interesting evolution in extraction techniques is with those dealing with difficult substrates where DNA is scarce or tightly bound and non-DNA contamination is abundant. Recent work suggests that the future in these situations lies with affinity methods and the development of immunoaffinity seems to point the way for future research. Such studies should be combined with investigations into the interaction between DNA and forensic environments so that extraction procedures can be devised for particular circumstances, for only with proper extraction procedures will subsequent DNA analysis be successful.

1 INTRODUCTION: PROBLEMS CONCERNING THE EXTRACTION OF DNA

In spite of recent advances in molecular biology, particularly in the field of forensic sciences (which include the advent of PCR, sequencing, and the plethora of microsatellites and other length polymorphisms used for identification purposes), the extraction of DNA, especially from substrates which may be scarce or badly degraded (for example, dry bone and teeth, old blood stains and hair shafts), still remains a problem.

The possibilities offered by modern-day technology would seem highly applicable to the forensic world, as with specific PCR techniques such as nested PCR,[1] it is possible to isolate typable DNA from a single cell. A single cell contains approximately 10 pg of DNA (i.e. 10 000–100 000-fold less than required for restriction fragment length polymorphism (RFLP) analysis and 100–1000-fold less than for standard PCR techniques), but over 40 cycles of nested PCR will produce a signal visible by ethidium bromide staining of polyacrylamide gel.[1] However, success in practice is greatly different from theory, and methods that allow extraction of adequate amounts of relatively pure DNA from qualitatively or quantitatively scarce material are crucial for the success of all subsequent analyses.

It is frequently reported in the forensic literature that DNA is a fairly robust molecule and that relatively short sequences can survive under appropriate conditions. However, common experience teaches us that with particularly difficult substrates, DNA extraction and/or amplification is often unsuccessful. The main reasons for failure are DNA degradation, PCR inhibition, and failure to solubilize and separate DNA from the substrate of origin. Unfortunately, as often happens among differing scientific disciplines, forensic scientists rarely share or acquire information that originates from other branches of molecular biology. For this reason, valuable

information on diagenesis (the study of the biochemical alterations that biomolecules go through after cell or organism death) which come from the fields of chemistry and geochemistry, may be lost. It therefore seems mandatory to approach briefly the problems behind DNA extraction – or rather the risks of failure to extract informative DNA – before addressing the actual methods themselves. For this reason, it is necessary to understand what little is known about DNA's behavior, its degradation, and its interaction with other molecules.

Various theories and opinions, more or less corroborated by valid practical results, exist on the survival and behavior of DNA.[2,3] DNA has been reported as surviving and having been successfully extracted even from historical and archaeological material. The remains of the Romanovs who died in 1918 were identified by DNA analysis.[4] Several authors have identified sex[5] and other informative genetic material from archaeological bone.[6,7] None the less, it must be stressed that these findings are exceptional, and other experts feel that nucleic acids are amongst the most unstable molecules,[8] and that DNA is likely to survive only in protected settings. This may be the result of its interaction with protective substances surrounding it, for example after having been conglomerated with resistant macromolecules, being bound to other molecules, or encased by crystals. An example is hydroxyapatite crystals in the case of bone. DNA is chemically unstable because of various phenomena such as hydrolysis, oxidation, and nonenzymatic methylation.[2] The cellular destruction that occurs soon after cell death exposes DNA molecules to enzymes and other reagents that cause breakage and further chemical alterations.[3] In addition, environmental agents such as heat, radiation, and microbial attack take their toll. Recent diagenesis studies seem to show that hydrolytic scission events may be crucial to DNA loss, even though the molecule may survive autolytic and biological degradation. Even when it survives, DNA may not be in a PCR-amplifiable form.[9] The lability of DNA and its susceptibility to harsh exogenous and endogenous agents should therefore caution against extreme optimism for the survival of informative DNA in old material. These factors also suggest adopting extraction methods that are not harsh, that avoid further loss of whatever useful DNA may remain, and are chemically suited to the substrate.

Another problem in obtaining a good yield of DNA suitable for PCR analysis is the presence of non-DNA contaminants that may inhibit the PCR. The main contaminant is usually protein and many methods of extraction use phenol because this targets proteins, eliminating them from the DNA solution. However, contaminants of a different nature exist, such as hemoglobin and minerals, which may also act as inhibitors of the PCR. Specific methods of extraction must therefore be used in these cases if inhibition is to be avoided. For example, a copurified hemoglobin derivative (hematin) is postulated to be a major contaminant, and its inhibitory effect can be countered by adding bovine serum albumin to the extract, or by using ultrafiltration, chelating resin treatment, gel filtration, or anion exchange column chromatography.[10] The presence of inhibitors should always be suspected and tested for when a PCR has given an unexpected negative result, by simply adding the extract in different dilutions to known positive controls and verifying whether inhibition occurs.

Perhaps the most difficult problem to solve is the actual detachment of DNA from the substrate. Some materials, for example bone, provide excellent protection of DNA as a result of their crystalline structure on to which DNA seems to adsorb; however, for the same reason it is equally difficult to solubilize DNA from the bone crystals. Any extraction methods therefore need to be adequate to overcome the chemical bonds between DNA and bone without destroying the nucleic acids.

The difficulty in devising such specific methods is that very little is known about the chemical behavior of DNA, even in vitro, although it seems to have been the most studied molecule since the late 1950s. It appears that a variety of structurally diverse low-molecular-weight ligands react with DNA, mainly by noncovalent associations with the minor groove of DNA, intercalation between DNA base pairs, and by alkylation of the component nucleotides.[11] In spite of the innumerable methods used for studying DNA interactions (such as X-ray crystallography and computational modeling methods), its mode of interacting with small molecules or even with macromolecular complexes (such as proteins)[12] remains unknown. Certain molecules, such as water, seem to affect interactions of DNA with other molecules, but the exact modalities, even of this simple reaction, remain unknown.[13] For these reasons, it is all the more difficult to devise adequate extraction techniques for forensic scenarios.

DNA contamination should be briefly mentioned. The extracted DNA may be contaminated by exogenous DNA, which may belong to the person who collected the sample, or extracted it, or may derive from pipets or other laboratory hardware that have been contaminated by other DNA samples, or may be present in aerosol from the laboratory atmosphere. DNA contamination presents one of the greatest risks of PCR analysis because it may lead to dangerous exclusions or false conclusions. To overcome this problem, numerous cautionary measures are recommended and taken by most laboratories working with DNA. These include appropriate washing and sterilization of crushing devices and other tools (with anti-DNA reagents and/or ultraviolet (UV) light), the use of positive displacement pipets, and the testing of

positive and negative control samples with each series of extractions. Also, mock extractions should be carried out in parallel with the specific extraction method being used. For example, nonhuman substrate should be tested (when working with human material), as well as complete blank controls, i.e. distilled water and reagents on their own as in the case of affinity methods, silica powder, or magnetic beads. Contamination risks come not only from the laboratory but also from the forensic site itself, such as the room or area in general where a crime has been committed, where a body has been found, or even the autopsy room. Interestingly, a recent study performed in the New York City Medical Examiner facilities showed that there were moderate quantities of typable DNA contamination, not only in the autopsy room, but also in adjacent room hoods and air vents. However, DNA contamination from airborne sources was not evident. These authors concluded that there should be no contamination of specimens so long as samples were collected into clean containers and not placed directly in contact with contaminated surfaces.[14]

In conclusion, provided that DNA has survived in a suitable form for PCR analysis in substrates such as old blood stains, dry bone, and hair shafts, then one must make sure the extraction technique chosen will avoid exogenous contamination, eliminate possible PCR inhibitors and release DNA from the substrate. Different substrates react with DNA in different manners and the extraction method must be chemically and physically tailored to the specific kind of substrate. It should be remembered that this is a rapidly growing field, and as more understanding of interrelationships between DNA and other molecules develop, the techniques for extraction will change. This article therefore gives guidance to those methods the authors feel to be most popular and recommendable at the present moment.

2 POSSIBLE FORENSIC SUBSTRATES FOR DNA ANALYSIS, THEIR PECULIARITIES AND PREPARATION FOR EXTRACTION

The extraction method must be suitable for the starting material or substrate on which the analysis is to be performed, different methods being appropriate to different situations. The following briefly lists the major forensic substrates and their peculiarities for DNA extraction. Not all samples need to be treated in the same manner and many different ways of collecting, treating, or preparing the sample prior to extraction have been recommended. This section includes general considerations of these methods.

2.1 Fresh Materials with Abundant DNA

These substrates, which include blood, soft tissue, and saliva, usually give no problems and contain abundant cellular material from which DNA can be extracted by standard methods. They are the substrates on which most of the methods described in section 3 were devised and tested. Fresh blood, soft tissue, and saliva contain a sufficient quantity of well-preserved blood, parenchymal cells and epithelial cells, respectively, for successful DNA extraction. Some soft tissues are preferable, favorites being cervical lymph nodes and muscle.

Fresh blood is perhaps the easiest substrate to deal with. Depending on the method employed, white cells are usually isolated as a buffy coat and DNA extracted from this or, more simply, a few drops of whole blood are used. The need for fresh blood to be extracted in forensic cases is usually related to samples taken from living persons in paternity cases or from suspected perpetrators of a crime who have left organic residues at the scene or on a body.

Relatively fresh samples from soft tissues excised with a sterile scalpel from fresh cadavers must be homogenized in a mill, taking care to thoroughly sterilize the equipment in between the preparation of different specimens. When sampling from a cadaver, muscle and lymph nodes are optimal substrates. About one cubic centimeter of tissue is usually sufficient for several tests.

Blood, fresh tissues and saliva should be stored at $-20\,^{\circ}C$ and homogenized only when needed.

2.2 Putrefied Materials with Abundant DNA

Typical examples of these materials are blood, soft tissues, and entomological and botanical material. These substrates present with putrefied, lysed cells, fatty putrefaction products and/or fibrous organic tissue (the last two are particularly associated with maggots and plants) which tend to inhibit PCR analysis, though not much is known about the mechanism of such inhibition. Although large amounts of DNA are still available, great care must be taken to eliminate PCR inhibitors.

Putrefied blood and substrates containing putrefied blood, in particular, quickly decay, producing hematin and porphyrins which are severe inhibitors of the PCR. In the case of cadavers, blood tends to clot soon after death, entrapping cellular elements and leaving behind few cells actually available for DNA extraction in the fluid component.[15] Extraction priorities with such abundant, but difficult, substrates must therefore be to discard debris (proteins, lipids, metals) in order to quantitatively and qualitatively eliminate inhibitors and then actively select the DNA. Putrefied soft tissues are extracted for DNA largely in cases of identification of decomposed human remains.

DNA typing of maggots found on cadavers is required to aid in determining time and place of death. Different insect species proliferate at different temperatures, with different velocities and in different environments. At times there may be difficulties in diagnosing the species of a maggot by morphological means alone, so genetic typing must be performed. In the case of plants, toxicological diagnoses may require genetic analysis, as in circumstances where there is a need for diagnosing the species with accuracy, for example *Cannabis sativa* from a pack of drugs. In all cases of entomological and botanical material the sample is simply crushed to a pulp or to a fine powder and extracted. Fibrous material needs to be eliminated in botanical samples. Maggots have a particularly high fat content, so great care must be taken to remove this component, for example with solvents.

2.3 Clean Material with Scarce DNA

These materials include stains (blood, saliva, sperm), swabs (vaginal, from bitemarks, from fingernails, from stains on bodies or other surfaces), fingernails, hair, desquamated cells (skin, dandruff), and other body liquids (urine, plasma). The main problem with these substrates is the paucity of typable DNA (if indeed there is any at all), because of degradation related to environmental phenomena and time and/or because of the small number of cells involved. In many instances, these substrates are relatively clean from inorganic debris. An essential prerequisite for the method of extraction is therefore to reduce loss of DNA to a minimum, for example by introducing more specific methods aimed at the small quantities of DNA present and reducing the number of steps in the process.

Blood stains can be found on many types of substrates (cotton, denim, carpet, wall board, envelopes, cigarette butts) and DNA is routinely and successfully extracted from all of them. Special care should be taken in these cases to eliminate the hematin and porphyrin components that inhibit PCR. Similar problems arise with fingernails of a victim, which may contain blood belonging to an aggressor.

Most authorities believe that shredding (or cutting into small pieces) fabric or fingernails, prior to elution is the method of choice. Others prefer to elute any residues off the nails, stains or other substrates by simply incubating the entire sample in extracting solution. Time and temperature of incubation may also be important, especially for blood stains because blood seems to be tightly bound to some substrates. In this case, incubation prior to extraction in a buffer at 56 °C for long periods (e.g. overnight) may be useful.

In the case of stains (from semen, blood, saliva, sweat, nasal secretions) on fabric, the stain is cut with sterile scissors. The same is valid for fingernails, which can be cut from the cadaver or from a live victim of an attack where there was a fight during which the victim possibly scratched the assailant. The main aim in extracting DNA from any kind of substrate is to expose the maximum amount of surface area of the stain to the extracting agents (most preparations of tissues and other substrates for these purposes involve reduction to small pieces or pulverization).

Nails can also be boiled for 5 min in sterile water prior to chopping as this apparently helps digestion.[15] Nail scrapings or debris should be recovered by swabbing all nails of a hand with a single cotton swab moistened with sterile water.

Similarly, if blood or other body liquids are on the skin, they can be transferred on to a damp sterile cotton swab by swabbing the area in a circular motion for about 30 s. The swab is then dried in the open air for 6 h or immediately frozen at −20 °C. The same is valid for vaginal, anal, and buccal swabs for dead or living presumed victims of rape or sexual violence. These are collected by repeatedly swabbing the vagina (in some cases different swabs are taken for external genitalia, vaginal wall and cervix), anus and mouth.

Vaginal swabs in victims of rape are a special case where the presence of genetic material in mixed substrates originating from more than one individual is a frequent problem. The vaginal swabs may contain female epithelial cells and equal amounts of sperm cells. In these instances, where the substrate is relatively abundant, physical separation of the genotypes by differential lysis methods is usually feasible. This involves separating the female epithelial cell fraction from the male sperm cells by using different buffers[16] (section 3). Problems arise with such mixed substrates when the sperm cell fraction is small or when material from victim and aggressor are of the same cell type, for example when the perpetrator's skin/blood is under the victim's fingernails. In such cases, allele-specific amplification systems have been devised that allow the detection of individual alleles in a sample with mixed genotype.[17]

Saliva stains can also be useful for analysis, and genomic DNA has been isolated from saliva found on a variety of substrates, such as chewing gum, cigarette butts, and postage stamps.[18] A greater problem arises with saliva left on human skin (transferred by biting, kissing, sucking, licking). In these cases, saliva is usually collected by a different swabbing technique from that described previously for other stains, i.e. the double-swab technique. The affected skin area is swabbed with a dry swab first, then with one previously moistened in sterile filtered distilled water.

Desquamated cells (e.g. from fingerprints or dandruff) are another possible substrate for DNA typing. Contact of

a suspect's hand, for example fingertips, with household objects, body parts as well as lesions on body parts such as bruises, can leave desquamated epithelial cells which can be collected and typed.[19] The print only needs to be swabbed, as described above, after having adequately searched and tested the body for fingerprints. For similar reasons, certain objects that have been worn may be proper substrates, such as condoms even when no ejaculation has taken place. Dandruff, a clinical alteration of skin with ''orthokeratotic clumps with minute parakeratotic foci'' found in dermatological pathologies,[20] is also a suitable substrate for DNA extraction if detected at a scene of crime.

Body fluids of forensic interest (other than blood, saliva, or sperm) which contain scarce amounts of DNA usually include urine and plasma. When the material is dry or semidry in the form of a stain or smear on fabric, it can be treated as blood stains or collected by swabbing if on a body. Though less frequent, it does happen that urine or plasma/serum (without the cellular components of blood) must be tested to verify whether the sample has been attributed to the correct person. An example may be urine testing for steroids in the sports industry, where positive results usually provoke contestation of whether the sample actually belongs to that person. Protocols for DNA extraction from liquids such as urine[21] and plasma[22] have been published.

Hair and body hair are other sources of DNA, although these substrates are controversial as most literature sources suggest that extraction of DNA is only possible from the bulb. However, there is at least one reference to extraction of DNA from hair shafts, but this refers to mitochondrial DNA.[18] Recent studies have shown that appropriately examining the microscopic structure of hair, as well as diagnosing its growth phase, is crucial. Telogen hairs are most commonly found at the scene of crime, as the average human will shed 50–100 hairs during daily activities. Anagen/catagen hair bulbs and translucent sheath tissue are excellent candidates for nuclear DNA examination, whereas telogen hair clubs should not be submitted initially for such analysis, but should be saved for other tests (for example morphological comparative studies), before being subjected to grinding to a powder for DNA analysis.[23]

Scalp hair is usually found on-site, on clothes and on bodies, and is easily collected using sterile forceps.

Pubic hair from the perpetrator of rape is usually found in the genital areas and on underwear of the victim and is collected and tested in the same manner. Pubic combing of suspected rape victims should always be performed in order to isolate suspicious hair that does not belong to the victim.

In murder cases, evidence such as hairs and fibers can be obtained from the body with strips of adhesive tape. Such tape fragments must then be carefully scanned for possible sources of the perpetrator's DNA.

2.4 Dirty Material with Scarce DNA

These substrates include sperm, saliva and blood stains, nails and feces. With these substrates, not only are there small amounts of DNA, but there are also large quantities of organic and inorganic contamination from endogenous and exogenous sources, and from other substances. For example, nail scrapings, blood stains or swabs may contain soil, soot or other inorganic contaminants which can easily be present at a forensic scene. The method for extracting DNA must therefore prioritize on recuperating small quantities of material, as well as on eliminating the inorganic components that may inhibit PCR.

Feces are peculiar in that they comprise a complex mixture of diverse microorganisms, digested food residues, mucus, and enzymes. All these must therefore be removed prior to PCR, especially bilirubin and bile salts which are strong PCR inhibitors.[24]

2.5 Formalin-fixed and Paraffin-embedded Tissues, and Stained Smears on Slides

Tissues in formalin and paraffin also constitute a problem. Paraffin must be removed prior to DNA testing and formalin has deleterious effects on the DNA while fixation is taking place, as it alters its chemical structure.

Formaldehyde oxidizes to formic acid which acts as a reducing agent. At a concentration of 4% (usually the minimum at which it is used) formaldehyde breaks the purine–pyrimidine hydrogen bonds resulting in rapid and irreversible denaturation of DNA by hydrolysis of the *N*-glycosidic bonds which release the purine residue, followed by gradual hydrolysis of the phosphodiester bonds between nucleotides; cross-links between protein and DNA also occur.[25] Paraffin seems to do less damage chemically, but tissue sections must be deparaffinized by dissolving, usually in octane, and then the DNA extracted with standard techniques, or the tissue simply digested.[26] Specific methods involving phenol/chloroform also exist for dealing with formalin-fixed tissues. Extraction from these kinds of tissues is more frequent in scenarios pertaining to natural history and archaeology, although its applications are becoming increasingly frequent in paternity cases and malpractice suits, where a fixed piece of tissue may be the only reasonably well-preserved substrate left of a putative father. Similarly, histopathology sections and fixed tissue samples may be crucial in verifying whether there has been a specimen mix-up and therefore a wrong diagnosis.

At times, and for similar reasons, it may be necessary to extract DNA from stained cytological smears. Studies

have shown, at least with Papanicolau, May Grunwald Giemsa, hematoxylin and eosin and Baecchi stains,[27] that this can be successful.

2.6 Calcified Tissues (Bone and Teeth)

The presence of a mineral matrix on to which DNA adsorbs requires the extraction process to overcome the bonds between minerals and DNA. These materials currently present one of the greatest challenges in forensic (and archaeological) DNA analysis.

Once remains have reached the skeletal stage, it is obvious that considerable degradation and usually soil contamination have occurred. Dry bone (and to a lesser extent tooth) exhibit large amounts of high-molecular-weight microbial DNA due to bone-dwelling organisms. There are also contaminants of humic acid and tannins derived from the soil. All these factors inhibit PCR and must be eliminated.[28] For skeletal material, various groups in the field of microbial ecology and paleogenetic biology have implicated a plethora of inhibitors, namely Maillard reaction products or chemically altered carbohydrates, residues of porphyrins, damaged DNA, bacterial (nontarget) DNA, and soil components such as fulvic and humic acid, tannins, or complexing ions like iron (Fe^{++}). Although the common trend is to think of inhibitors as being ''outside the substrate'', it is now being realized that they may be endogenous to the sample. Collagen, particularly human collagen type I, has been proved to be a potent PCR inhibitor.[29]

However, the real challenge with calcified tissues (dry bone and teeth) lies in the binding interactions of the hydroxyapatite matrix that chemically stabilize DNA, thereby aiding its survival, but making it difficult to extract. The mechanisms of such interaction must be known in order to devise proper extraction methods for DNA from calcified tissues.

Bones and teeth are usually crushed and powdered by using liquid nitrogen in specific mills, though one could envisage that violent crushing may actually damage DNA. Decalcifying the bone or tooth does not seem to increase the yield of DNA; in fact, some studies have shown that it actually reduces the yield.[30]

In the case of dry bone, it is still not known whether cortical or trabecular bone is the better substrate for extracting DNA. Cortical bone contains only bone cells (osteocytes, osteoblasts, and osteoclasts) and cells from vessels and nerves running through the Haversian canals. Trabecular bone additionally contains possible bone marrow residues in between the trabeculae, which may actually increase the amount of extractable cellular material, although it may also bring along the problem of greater quantities of cell debris that need to be purified.

Whichever type of bone is preferred, pulverisation seems to be the chosen form for extraction, as it maximizes exposure of the bone to extracting agents, a very important step considering the strong bond between DNA and mineral.

For teeth, the pulp should be the main source of DNA. If one considers the morphology of a tooth, the pulp chamber is the only dental tissue with nucleated cells. Neurovascular tissue here should be the best source of DNA and some authors therefore recommend extracting the pulp. However pulp morphology goes through a series of postmortem alterations which may make it impossible to distinguish the pulp from the pulp chamber walls and neatly pull it off (it usually has a film-like appearance and elastic or spongy consistency that, in fresh samples, makes it easy to tease out with clean forceps). In cases when it has disappeared or turned into a crystalline or sandy substance adhering to the walls of the pulp chamber, it is best to crush the entire tooth. It is possible in theory that DNA may still be present in old cementoblasts around the root or in extensions of odontoblasts in dentine, but the quantities are minimal. Teeth with endodontic treatment are particularly difficult, and it has been shown that the quantity of DNA recovered is 100-fold less than in a normal tooth.[31] Therefore treated teeth, as well as immature permanent teeth with open roots and deciduous teeth with open or reabsorbing roots, should not be used.

3 DNA EXTRACTION

In order to clarify the various stages and methods of DNA extraction, this section comprises two basic parts, one dealing mainly with nonaffinity extraction methods and the other with affinity-based procedures. It must be kept in mind that the protocols given are mainly designed for PCR analysis although, depending on the substrate, some may also be used for the extraction of abundant, high-molecular-weight DNA useful for RFLP.

In general, extraction methods come under headings that bear the name of the type of chemical element involved, for example organic or inorganic. However, for this article, the simplest and most appropriate way of classifying DNA extraction techniques is by their essential mode of action. Methods directed towards all material but DNA (e.g. protein precipitation or the chelation of inorganic substances) are placed in the nonaffinity category, whereas those that specifically react with DNA and selectively extract it discarding everything else are considered in the affinity category. In the chemical sense, the term affinity relates to the property or attraction by which differing chemical elements or groups of

elements unite to form a new compound. In the case of affinity DNA extraction methods, the bonds that form between the DNA and the extracting reagent are reversible, and therefore allow subsequent release of the DNA.

Nonaffinity methods include:

- the well-known phenol/chloroform technique, frequently referred to as the main organic method;
- inorganic methods, based on the use of saturated salts;
- chelating agents, such as Chelex or ethylenediaminetetraacetic acid (EDTA);
- miscellaneous methods that include the use of other substances such as NaOH (alkaline extraction) and detergents.

All have in common the fact that they are designed for ''extracting'' non-DNA substances from the solution, leaving DNA behind to be concentrated and tested.

The affinity methods include techniques using silica particles, monoclonal antibodies and several commercial kits. All allow the user to work with material on to which DNA has been selectively bound.

Before specific extraction procedures are presented (meaning the purification and isolation of DNA), some important preliminary steps are considered. These are mainly the stages involving lysis of any surviving whole cells or other organic material and solubilization of DNA. In practice, this stage is already part of the DNA extraction procedure, and is included in a preliminary section because the rationale behind such a step is similar (as are the reagents) for most protocols. Differential lysis will also be dealt with.

Finally, this section describes the main nonaffinity and affinity methods of extracting DNA, based on the different principles by which they function. After a brief general consideration of the rationale, advantages, and disadvantages, two protocols for each method will be given: one for an easy substrate (e.g. fresh blood) and one for a difficult substrate (e.g. dry bone), when there is a distinction in the application of the method. Many combinations of extraction procedures exist. The protocols given in detail are those considered either to be the most widely tested for that particular category, or those which personal experience has suggested as being the optimal method at the present time. Tables of references are given for variations in protocols and other specific substrates. The previous comments regarding rapid advances in this field should be constantly borne in mind and the literature should be searched continually for more recent work on particular problems/substrates.

3.1 Cell Lysis, Protein Digestion and Solubilization of DNA

Although the main extraction step consists of the isolation of DNA prior to PCR, the preliminary phases, particularly that of lysis, are important.

These consist of dissociating the nucleic acids from their surrounding proteins and/or tissues, and the digestion of most material present in order to solubilize as much DNA as possible. Most techniques are based on incubation of the substrate in an extraction buffer (most commonly containing tris(hydroxymethyl)aminomethane (Tris)–HCl and EDTA), PK and a detergent (usually SDS) followed by what is usually considered the actual extraction phase or DNA isolation phase (e.g. phenol/chloroform extraction).

The role of the PK is to digest the protein, whereas the SDS is added to lyse cell membranes thus releasing the DNA. Other detergents that can be used include Nonidet and Triton X 100. Basically, many protocols in this sense are similar as they employ similar reagents and temperatures with varying incubation times (ranging from 30 min to 48 h), according to substrate and protocol.

Some protocols divide the lysis phase into two steps when fresh blood is the substrate: leukocyte separation by centrifuging the sample and/or by using chemical methods, including a red cell lysis buffer containing dextrose and a lysis agent; and pelleting the leukocytes. RNAses have also been used in the past to eliminate any RNA present. The pH of these solutions and buffers is also important: at a pH of 5–6, the extraction of RNA is favored; at a pH of 8–9 both RNA and DNA are liberated.[32] Sodium perchlorate has also been used in these preliminary steps – it is another reagent useful both for lysing and denaturing cellular proteins and, according to some authors, can replace the organic extraction phase. This is not a method of choice, but it has been successfully used on fresh[33] and ancient material.[34,35]

Finally, guanidine–HCl is sometimes used for cell lysis as it is highly denaturing.[36] It has also been employed in conjuction with glass beads and sodium perchlorate.

Protocols for the lysis and digestion steps are given within the detailed extraction procedures identified in sections 3.2 and 3.3.

3.1.1 Differential Lysis in Cases of Sperm-stained Material

In the special case of vaginal swabs and sperm stains where an abundant quantity of both sperm and female epithelial cells may be present, a differential lysis procedure (section 2) is necessary, prior to extraction. A suitable protocol is now briefly given.[16]

When a vaginal swab from a rape case, for example, or a piece of fabric containing mixed male and

female secretions are the substrate, these are soaked in phosphate-buffered saline (PBS) in an appropriate tube according to the size, and incubated overnight at 4 °C. The following steps should be followed:

1. Vortex briefly and remove the swab or fabric; spin down the cells at 2500 *g* for 10 min at 4 °C; save supernatant in a separate tube.
2. Wash the swab or fabric in 1.5 mL PBS/S (phosphate-buffered saline + 2% sarcosyl) in a microcentrifuge tube. Punch a small hole through the bottom of the closed tube, insert the tube into a second open tube and punch a hole in the lid of the tube containing the sample. Spin both tubes for 10 min at 1500 *g*; the upper tube should contain the swab or fabric and the lower tube the buffer and cell pellet. Combine the cell pellet with the pellet from Step 1, saving the supernatant.
3. Wash the pellet twice in 1.5 mL PBS/S in a microfuge tube and spin it down for 10 min at 1500 g. Resuspend the pellet in 50 µL PBS and remove 3–5 µL for microscopic analysis.
4. Add 500 µL PBS/PKS (proteinase K + sodium dodecyl sulfate) and incubate for 2 h at 50 °C with mild agitation for lysis of vaginal epithelial cells. Recover sperm heads by centrifugation at 1500 *g* for 10 min at 4 °C.
5. Save the supernatant (i.e. the female fraction) for further analysis of female DNA. Resuspend the pellet containing the sperm heads in 30 µL PBS and remove 2–3 µL for microscopic analysis.
6. Lyse the sperm heads (i.e. the male fraction) by adding 500 µL PBS/Lys and incubate for 3 h at 50 °C with mild agitation.
7. Separately purify DNA from female and male fractions by the phenol/chloroform method described in section 3.2.1.
8. Transfer both supernatants to two separate Centricon-30 microconcentrator tubes and purify DNA from salt, SDS, and contaminants with three washes with 2 mL distilled water each. Concentrate samples to a final volume of 100 µL.

3.2 Nonaffinity Extraction Methods

Nonaffinity methods are based on the indirect extraction of DNA by the selective precipitation of protein and other contaminants, or by chelation of inorganic substances (Figure 1). The main methods belonging to this category, which have excellent credentials for dealing with forensic material, are the phenol/chloroform method, Chelex resin and saturated salt procedures.

3.2.1 Phenol/Chloroform

Since the beginnings of DNA analysis, the organic method of extraction par excellence has been phenol/chloroform;[37] it is still the most popular method among the forensic community.

3.2.1.1 Method Rationale, Advantages and Disadvantages

Rationale. The main mode of functioning is to remove the protein component thus purifying the nucleic acids; this is usually carried out by simply extracting aqueous solutions of the nucleic acids with phenol and/or phenol/chloroform. When dealing with complex mixtures of cell lysates and debris – and this is usually the case with forensic substrates – it is sometimes necessary to employ additional measures. In these cases, more protein can be eliminated by digestion with proteolytic enzymes (such as PK) which are active against a broad spectrum of native proteins (section 3.1), before the extraction with organic solvents. The standard way to remove proteins is to extract once with phenol, once with a 1 : 1 mixture of phenol–chloroform, and once with

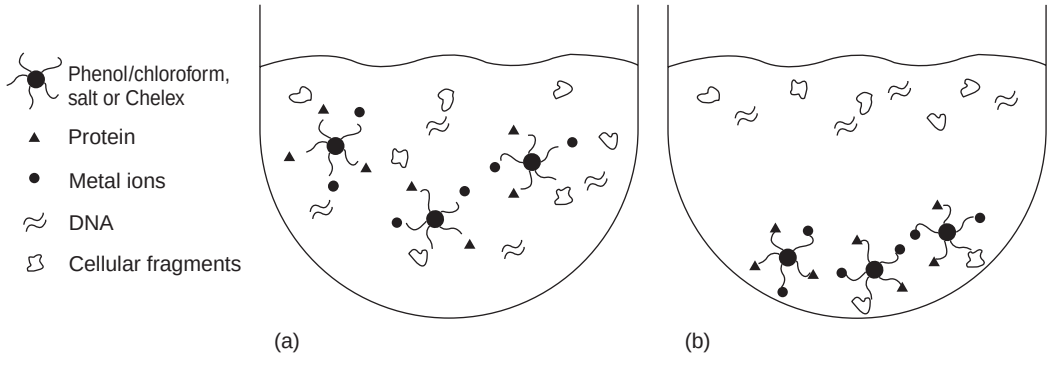

Figure 1 Mechanisms of nonaffinity methods where (a) the extracting agent (phenol/chloroform, salt or Chelex resin) pull away protein, metal ions, and components of cell walls, leaving (b) DNA in solution.

chloroform. Deproteinization is more efficient when two different organic solvents are used instead of one.[37]

Advantages. This is the most tested method in forensics with protocols (including alternatives) for myriad substrates.

Disadvantages. The method does not eliminate other nonprotein contaminants as efficiently and may lower DNA yield because of the number of steps involved; when organic material is separated from the double-layer solution, some of the aqueous DNA-containing layer is lost with it, although this loss may be diminished by piercing a hole through the tube and allowing the phenol to drip out or by using a pipet where the narrow tip has been cut off.[16] Phenol and chloroform can be harmful both for the substrate and for the operator.[32] It is possible that DNA may be damaged as a result of steric shearing during the multiple tube transfers necessary in this method.[38] Furthermore, phenol is a well-known carcinogen and chloroform can cause liver damage.

3.2.1.2 Selected Protocol for Easy Substrate: Fresh Blood[16]

Materials.

1. 50 mM KCl (for hypotonic lysis of erythrocytes).
2. Lysis buffer: 25 mM EDTA, 75 mM NaCl, 10 mM Tris–HCl pH 7.5. Add 200 mg per milliliter of PK immediately before use.
3. 10% (w/v) SDS.
4. Sodium acetate, sodium chloride, ethylenediaminetetraacetic acid (ANE) buffer: 10 mM sodium acetate, 100 mM NaCl, 1 mM EDTA.
5. Buffered phenol: use only crystallized phenol, dissolve in a water bath at 65 °C and add approximately 0.1–0.2 volume of ANE buffer until an aqueous phase forms above the phenol; add 0.01% 8-hydroxyquinoline (w/v solid crystals).
6. Chloroform/isoamyl alcohol: to 24 parts of chloroform, add one part of isoamyl alcohol.
7. 6 M NaCl.
8. 3 M sodium acetate, pH 5.2: dissolve 40.8 g sodium acetate in 80 mL distilled water, bring to pH 5.2 with concentrated acetic acid, and adjust final volume with water to 100 mL.
9. TE (tris(hydroxymethyl)aminomethane–ethylenediaminetetraacetic acid) buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 7.5.
10. Twentyfold sodium citrate (SSC): 175.3 g NaCl, 88.2 g sodium citrate per liter, adjust to pH 7.0 with NaOH.
11. 0.2 M sodium acetate, pH 7.0.
12. 10 mg per milliliter PK stock solution: dissolve 10 mg PK in 1 mL of 10 mM Tris–HCl, pH 8.0 and store frozen in small aliquots.
13. PBS: 50 mM phosphate buffer, pH 7.4, 0.9% NaCl.
14. PBS/S: PBS + 2% sarcosyl.
15. PBS/PKS: PBS + 100 μg per milliliter PK + 1% SDS (=10 μL PK stock solution + 100 μL 10% SDS per milliliter PBS).
16. PBS/Lys: PBS + 100 μg per milliliter PK + 2% sarcosyl + 10 mM dithiotreitol (DTT) + 25 mM EDTA (=10 μL PK stock + 10 μL 1 M DTT stock 50 μL + 0.5 M EDTA stock per milliliter PBS/S).

Method.

(a) Purification of white blood cells

1. Place 10–15 mL EDTA blood into a 50-mL polypropylene tube. In addition, transfer two 0.5-mL aliquots of blood into 1.5-mL microfuge tubes as reference samples and then store them frozen at −20 °C.
2. Add 50 mM KCl up to 50 mL, mix well, and incubate in water bath at 37 °C for 10 min.
3. Spin for 10 min at 500 *g* in a clinical centrifuge at room temperature.
4. Remove the supernatant with a Pasteur pipet connected to a water jet or vacuum pump; leave the pellet intact.
5. Repeat Steps 2–4 once or twice until the pellet is free from red cells.
6. Add 15 mL lysis buffer and shake vigorously to resuspend the cell pellet.
7. Add 1.5 mL 10% SDS and mix carefully. The sample should now become very viscous because of cell lysis.
8. Incubate overnight at 37 °C or at 55 °C for 4–5 h.

(b) Organic phase of DNA extraction

1. Add 1 volume buffered phenol to the PK-digested cell extract and mix the aqueous and organic phases carefully to achieve a homogeneous suspension; avoid vigorous shaking.
2. Spin for 10 min at 1500 *g* in clinical centrifuge at room temperature.
3. The aqueous phase containing the DNA is on top and the phenolic phase is below. Transfer the aqueous phase to a fresh 50-mL polypropylene tube using a widebore glass pipet. The interphase containing proteins and protein DNA complexes may also be transferred at this step.

4. Add 1 volume of a 1:1 mixture of phenol and chloroform–isoamyl alcohol (24:1) and extract as described in Steps 1–3.
5. Add 1 volume of choroform/isoamyl alcohol and repeat the extraction as described in Steps 1–3. At this step, avoid transferring any residual protein debris from the interphase.

(c) DNA precipitation

1. After organic extraction add 0.1 volume of 3 M sodium acetate and 2 volumes of ice-cold absolute ethanol.
2. Mix carefully without vigorous shaking. The DNA should precipitate by forming viscous strings and finally a compact pellet which may float on top of the solution.
3. Use a glass hook to recover the floating DNA (authors' variation).
4. Rinse the DNA pellet attached to the glass hook twice in 70% ethanol to remove excess salt.
5. Dry the pellet briefly in the air and resuspend the DNA in an appropriate volume (300–500 µL depending on the size) of TE (1 in 10 in distilled water).
6. Incubate the sample for 1 h at 65 °C or overnight at 37 °C in a water bath to dissolve the DNA. If the sample is still very viscous, add more TE (1 in 10 in distilled water) and incubate again at 65 °C until a homogeneous solution is obtained.

3.2.1.3 Selected Protocol for a Difficult Substrate: Dry Bone This protocol is modified from Cattaneo et al.[39,40]

Materials.

1. Extraction buffer: 10 mM Tris–HCl pH 8.
2. 0.1 M EDTA pH 8.0.
3. 20 µg mL^{-1} RNAse.
4. PK 20 µg mL^{-1}.
5. 0.5% SDS.
6. 10 M ammonium acetate.
7. Phenol.
8. Chloroform.
9. Isoamyl alcohol.

Method.

1. Add powdered bone tissue slowly to approximately 10 volumes of extraction buffer in a 50-mL centrifuge tube. Add RNAse to a final concentration of 20 µg mL^{-1}.
2. Incubate at 37 °C for 1 h.
3. Add PK to a final concentration of 100 µg mL^{-1}; incubate at 50 °C in a water bath for 3 h. Swirl the viscous solution periodically.
4. Cool the solution to room temperature. Add an equal volume of phenol equilibrated with Tris–HCl (pH 8) and gently mix the two phases by slowly swirling the tube end over end for 10 min. Separate the two phases by centrifugation at 2500 rpm for 15 min at room temperature.
5. With a widebore pipet (0.3 cm) slowly draw the DNA into the pipet to avoid disturbing the interface and transfer the viscous aqueous phase to a clean centrifuge tube and repeat the extraction with phenol/chloroform and then with chloroform/isoamyl alcohol.
6. After the third extraction transfer the aqueous phase to a fresh tube and add 0.2 volumes of 10 M ammonium acetate and two volumes of ethanol at room temperature and swirl the tube until the solution is thoroughly mixed.
7. Centrifuge the DNA down to a pellet.
8. Allow the pellet to air dry, then dissolve in sterile filtered distilled water.

3.2.1.4 Variations in the Protocol and References for Other Substrates Although the phenol/chloroform method has disadvantages, it has been reported as giving useful DNA for PCR from most substrates.

Few studies have been performed comparing phenol/chloroform with other extraction methods. However, one study by the present authors showed that, as far as DNA yield and PCR success were concerned, inorganic methods (section 3.2.2) were equal if not better than organic methods with difficult substrates such as dry bone.[39]

One other point that should be mentioned is that some authors have added extra postextraction concentration or purification steps. After extraction of DNA, various methods can be used to concentrate the end solution. Some workers prefer to use miniconcentrators, others to pellet the DNA down by centrifugation and resuspend it in the desired quantity of sterile filtered distilled water or deionized distilled water. Other purification steps mainly involve isolation of the molecular weight region of interest by electrophoresis of the PCR product, for example on agarose gel, cutting it out, and re-extracting it by boiling.[39] These extra steps may be included at the end of most types of DNA extractions.

Table 1 lists, in chronological order, some major references in the forensic and archaeological fields, where phenol/chloroform has been used successfully to extract

Table 1 Phenol/chloroform extraction methods with different substrates

Substrate	Year	Reference	Comments/peculiarities
Forensic: decomposed bodies and skeletal remains; vaginal swabs	1998	15	Differential lysis performed for vaginal swabs
Forensic: bodies, blood stains	1998	16	
Forensic: hair shafts	1998	18	Sample is first decontaminated – use only for hair shafts, not bulbs
Forensic: maggots	1998	41	
Forensic: saliva-stained material (cigarette butts, postage stamps, envelope sealing flaps)	1998	42	Declares better results than Chelex
Forensic: dandruff	1998	20	Comparative study with Chelex – both methods yield DNA but phenol/chloroform in greater amounts
Forensic: fingerprints on objects	1997	19	
Forensic: general forensic specimens	1996	43	Adds H_2O_2 to the digestion phase in order to purify from heme
Archaeological: ancient and old formalin-fixed tissues	1996	25	Recommends final DNA precipitation with NaCl, not sodium acetate
Archaeological: ancient teeth and bone	1996	44	
Forensic: human bodies in mass disasters, soft tissue, bone marrow	1995	45	
Archaeological: ancient mummies	1995	46	
Forensic: putrefied and burnt liver and brain	1994	10	
Archaeological: ancient bone and soft tissue	1994	47	
Forensic: blood, blood stains, semen stains, vaginal swabs, hair bulbs	1993	48	Differential lysis performed for vaginal swabs
Forensic: maggots	1994	49	
Archaeological: historical bone	1993	30	Showed that decalcifying does not improve DNA yield
Forensic: Mengele's skeletal remains	1992	50	
Forensic/experimental: teeth exposed to different environmental conditions up to 20 years	1992	51	
Forensic: urine sediment	1992	21	Used phenoldichloromethane
Forensic: decomposed remains, hair, skin, bone	1992	52	Decalcified bone
Archaeological: ancient bone	1992	28	
Forensic: eight-year old skeletal remains	1991	53	
Forensic: blood stains, vaginal swabs	1991	54	Differential lysis for vaginal stains
Forensic: bodies of fire victims	1991	55	
Archaeological: human 5000-year-old bone	1991	35	
Archaeological: ancient human bone	1991	5	
Archaeological: ancient bone and teeth	1990	7	
Archaeological: mummified tissue and bone	1990	56	
Archaeological: ancient human bone	1989	57	
Forensic: single hairs with bulbs	1988	58	

DNA from various substrates; protocols may vary from the ones mentioned above. Points of interest, when present, are listed on the right.

3.2.2 Precipitation with Salts

Methods involving the use of salts are widely employed in blood transfusion and clinical scenarios where abundant clean samples are available. Although rarely used in forensic laboratories, recent literature has shown them to be as valuable with some forensic substrates as phenol/chloroform and give DNA yields that could be adequate for PCR and RFLP.

3.2.2.1 Method Rationale, Advantages and Disadvantages

Rationale. The rationale is that certain inorganic reagents, mainly salts, will precipitate proteins, and therefore act very similarly to phenol/chloroform. The cellular proteins are salted-out by dehydration and precipitation with a saturated salt solution.

Advantages. Compared to organic methods, the techniques are quicker and cheaper[59] and much less toxic.

Disadvantages. They still do not eliminate all PCR inhibitors and considerable amounts of protein may be retained.

3.2.2.2 Selected Protocol using Saturated NaCl for Easy Substrate: Fresh Blood[60]

Materials.

1. Lysis buffer: (10 mM Tris–HCl, pH 7.6; 2 mM Na_2EDTA, 400 mM NaCl).
2. SDS at 10%.
3. PK at 20 mg mL^{-1}.
4. Saturated NaCl 6 M.
5. Absolute ethanol.
6. TE buffer (10 mM Tris–HCl, 0.2 mM Na_2EDTA, pH 7.5).

Method.

1. Buffy coats of nucleated cells obtained from anticoagulated blood are resuspended in 15 mL polypropylene centrifugation tubes with 3 mL nucleic lysis buffer.
2. Cell lysates are digested overnight at 37 °C with 0.2 mL SDS and 50 µL PK.
3. After digestion is complete, 1 mL of saturated NaCl is added to each tube and shaken vigorously for 15 s, followed by centrifugation at 2500 rpm for 15 min.
4. The precipitated protein pellet is left at the bottom of the tube and the supernatant containing DNA transferred to another 15-mL polypropylene tube.
5. Two volumes of absolute ethanol at room temperature are added and the tubes rotated until DNA is precipitated (forming a medusa).
6. The precipitated DNA strands are removed with a plastic spatula or pipet and transferred to a 1.5-mL microcentrifuge tube containing 100–200 µL TE buffer.
7. The DNA is then allowed to dissolve for 2 h at 37 °C before being stored at −20 °C.

3.2.2.3 Selected Protocol using Saturated Sodium Acetate for Difficult Substrate: Dry Bone[39,40]

Materials.

1. Buffer (10 mM Tris–HCl pH 7, 10 mM Na_2EDTA, 50 mM NaCl).
2. PK (20 µg mL^{-1}).
3. SDS at 10%.
4. 100% isopropanol.
5. 70% ethanol.
6. Saturated sodium acetate.

Method.

1. 1.5 g pulverized bone are incubated in a polypropylene tube containing 3 mL white cell lysis buffer with the addition of 25 µL PK and 50 µL SDS in a water bath overnight at 42 °C.
2. 1 mL saturated sodium acetate is then added and the tube shaken manually for 30 s and then centrifuged at 4000 *g* for 10 min.
3. The supernatant containing the DNA is then transferred to a new polypropylene tube; 4 mL 100% isopropanol are added and mixed for at least 10 min to precipitate the DNA before centrifugation at 4000 *g* for 10 min.
4. The supernatant is discarded; the remaining whitish-yellow pellet of DNA suspended in 250 µL 70% ethanol, is transferred to a 1.5-mL microcentrifuge tube and centrifuged at 13 000 *g* for 10 min.
5. The pellet is dried in an incubator at 37 °C, reconstituted in 250 µL sterile filtered distilled water and the resulting DNA solution divided into 20-µL aliquots which can be stored at −20 °C until required for testing.

3.2.2.4 Variations in Protocol and References for Other Substrates The literature does not provide as wide a variety of protocols for salt extraction in forensic scenarios as for the other methods. However, a previously mentioned comparative study between phenol/chloroform and saturated salt precipitation was performed for the extraction of DNA from putrefied cadavers.[39] The latter seemed the better extraction method, although a post-PCR purification step still had to be added in order to get typable DNA.

3.2.3 Chelating Methods

3.2.3.1 Method Rationale, Advantages and Disadvantages

Rationale. The rationale is that the active reagent will bind to inorganic substances and clear the extract of PCR inhibitors in this category. Although simple chelating methods exist which employ the properties of substances such as EDTA to neutralize certain ions, one of the main chelating methods involves the use of Chelex. Chelex is a resin composed of styrene–divinylbenzene copolymers containing paired iminodiacetate ions that act as chelating groups in binding polyvalent metal ions. It has a high selectivity for divalent ions and

differs from ordinary ion exchangers because of higher bond strength.[61] It has been postulated that boiling a sample in the presence of Chelex prevents the degradation of DNA.[62] Metal ions which act as catalysts in DNA breakdown at high temperatures and at low ionic strength are chelated and inhibited from this action.

Advantages and Disadvantages. The technique is simple and rapid and does not involve multiple transfer tubes. The Chelex resin removes impurities and the alkaline pH disrupts the cell membranes resulting in the release of DNA. It does not have a particularly selective activity on protein and can be unreliable unless relatively large amounts of DNA are present.

3.2.3.2 Selected Single Protocol Using Chelex for an Easy Substrate: Fresh Blood[61]

Materials.

1. Chelex 100.
2. Sterile deionized water.

Methods on Whole Blood and Blood Stains.

1. Pipette 1 mL sterilized deionized water into a 1.7-mL microcentrifuge tube.
2. Add 3 µL whole blood or a piece of blood-stained material (3 mm square) and incubate at room temperature for 15–30 min.
3. Prepare a 5% Chelex solution in sterile deionized water.
4. Centrifuge samples for 3 min at 10 000 *g* in a microcentrifuge to pellet the red blood cells.
5. Carefully remove all but 20–30 µL of the supernatant and discard. Leave the substrate and pelleted material in the tube.
6. Resuspend the pellet in 5% Chelex 100 to a final volume of 200 µL.
7. Incubate at 56 °C for 30 min.
8. Vortex at high speed for 5 s.
9. Incubate in a boiling water bath for 8 min.
10. Vortex at high speed for 5 s.
11. Centrifuge samples for 3 min at 10 000 *g* to pellet the Chelex 100 resin, substrate, and remaining tissue.
12. The extracts are now ready. For short-term usage, store at 2–8 °C on the Chelex 100 resin. For long-term storage (over 1 month), the extracts should be centrifuged and the supernatant removed from the Chelex 100 resin and stored in a new tube at −20 °C.

3.2.3.3 Selected Protocol using Chelex for a Difficult Substrate: Bone[61]

Materials. As described above.

Method.

1. Prepare a 5% Chelex 100 suspension in sterile deionized water.
2. Add 200 µL of 5% Chelex 100 suspension to a 1.7-mL microcentrifuge tube.
3. Place 0.5 g crushed bone tissue directly into the Chelex 100 suspension.
4. Incubate at 56 °C for 30 min.
5. Continue with the protocol for whole blood (Steps 8–12).

3.2.3.4 Variations in Protocol and References for Other Substrates It has been reported that Chelex can also be used to extract DNA from hair, seminal stains, saliva, and postcoital samples.[63] Its advantage with saliva stains, as compared to phenol/chloroform, has been verified,[64] although in this case a modified version of the Chelex method was adopted, which involved pre-extraction with PK and incubation at 56 °C and 100 °C plus microconcentration of the solution.

References to other chelation protocols on different substrates are given in Table 2.

Table 2 Chelex extraction methods with different substrates

Substrate	Year	Reference	Comments/peculiarities
Forensic: blood, blood stains, semen, hair, buccal swabs, postcoital swabs	1998	61	
Forensic: hair	1998	23	
Forensic: stained smears on slide	1997	27	Vaginal and postcoital smear slides
Forensic: saliva stains on human skin	1996	64	Modified Chelex method with pre-extraction in PK
Forensic: urine stains	1996	65	
Forensic: teeth in various conditions up to 36 months	1996	66	
Forensic: blood, blood stains, swabs, nail scrapings	1995	17	
Archival tissue: paraffin wax-embedded tissue	1994	67	
Archaeological: insect (weevil) in amber	1993	68	

3.2.4 Miscellaneous Methods

Other less popular nonaffinity methods exist, some of which rely on the use of detergents. The CTAB (cetyltrimethylammonium bromide) and DTAB (dodecyltrimethylammonium bromide) method[69] can be considered a valuable, relatively recent technique for DNA extraction. After incubation in DTAB, a cationic detergent which lyses the cell membrane and denatures protein, chloroform is added to further denature and deproteinize the extract, which is then centrifuged; the DNA-containing aqueous layer is kept and CTAB is added to it. The purpose of CTAB, another detergent, is to precipitate DNA. These reagents are highly toxic and have featured in few reports in the field of forensics.

Alkali extraction is another alternative and is among one of the simplest strategies – whereas the primary structure of DNA is relatively stable in alkaline solution, other proteins are efficiently solubilized due to the ionization of aspartic, glutamic, cysteic, and tyrosine residues. Proponents of the method state that efficient extraction of DNA from whole blood can be performed with alkaline treatment at room temperature in a time frame as short as 1 min. Alkaline extraction of buccal epithelial cells, blood and semen stains can be accomplished in 5–6 min at 75 °C, whereas liquid semen requires 30 min at 75 °C. The technique involves incubating aliquots of whole blood (5 μL) with 20 μL of NaOH at temperatures ranging from ambient to 90 °C for 60 min. The reaction is stopped by adding Tris–HCl, pH 7.5.[70] Potassium hydroxide can be used as a substitute for sodium hydroxide.

Some authors have even suggested the use of anhydrous hydrogen fluoride to extract DNA from yeast and fungi,[71] whereas others have reported extracting DNA from archaeological material by simply incubating bone powder in water[72] or by incubating it in an extraction buffer.[73]

3.3 Affinity Methods

The novelty of affinity methods is that they rely on the process of actively selecting, to varying degrees, the DNA by either a nonspecific or specific reaction, thus allowing the examiner to wash out all other substances.

3.3.1 Nonspecific Affinity Methods

The main nonspecific affinity methods employ (a) glass (silica) beads in combination with isothiocyanate bound to guanidium[74] or (b) magnetic beads from Dynal[75] coated with a substance(s) specific for DNA but which is not revealed in the manufacturer's data sheets (dealt with briefly in section 3.3.3).

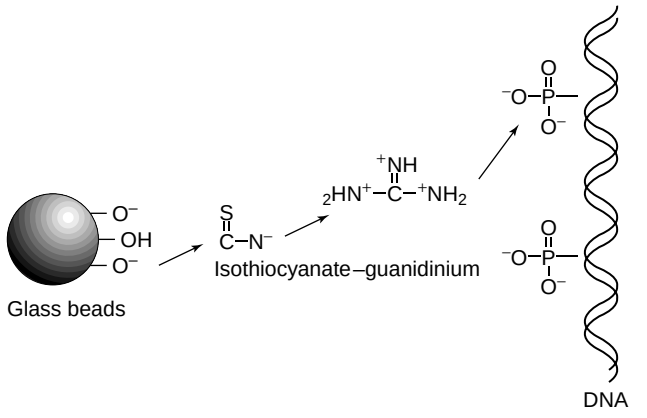

Figure 2 Chemical interaction between the silica bead and DNA.

3.3.1.1 Method Rationale, Advantages and Disadvantages of the Silica Beads Procedure

Rationale. The silica glass particle method (sometimes called the glass milk[76] or silica gel method) is based on glass particles actively attaching to isothiocyanate which is itself bound to guanidium. It has been hypothesized that the positively charged amine groups on the guanidium act as the link to the negatively charged phosphate groups on DNA. Once DNA is bound to the glass particles, it can be repeatedly washed and spun, as the glass particles will retain the attached DNA and allow elimination of other material and inhibitors (Figure 2). However other mechanisms of DNA binding to silica have been suggested. These may involve chaotropic salt disruption of the water structure around negatively charged silica, allowing a cation bridge to form between it and the negatively charged phosphate backbone of DNA; this reaction is reversible in low salt concentrations.[77]

Advantages and Disadvantages. The main advantage is that, contrary to nonaffinity methods, DNA is selectively bound (even if only temporarily) to a substrate (glass) that can be repeatedly washed. Once it has been satisfactorily cleansed, the DNA can be eluted. This offers a much better chance of purifying DNA from PCR inhibitors.

The disadvantages lie in the fact that when a great deal of contaminant is present, it may sterically hinder the adsorption process of DNA on to the glass substrate. In addition, the silica method preferentially recovers high-molecular-weight DNA and so may fail to recover any DNA from samples in which molecules are present but in a highly fragmented form (less than 100 bp).

3.3.1.2 Selected Protocol Using Silica Particles for any Substrate This protocol is modified from Höss and Pääbo[78].[40]

Table 3 Application of silica-based extraction methods with different substrates

Substrate	Year	Reference	Comments/peculiarities
Forensic/experimental: human bone and blood stains up to 43 years	1997	40	Comparative study with sodium acetate and DNA direct™ (Dynal) showed greater efficiency of silica
Serum and plasma	1996	22	Bio-Rad Prep-A-Gene® DNA purification kit
Archaeological: ancient skeleton	1996	79	
Archaeological: ancient human skeletons	1996	80	
Archaeological: botanical and entomological samples from herbaria and amber-embedded insects	1993	81	Remark that Chelex extracts will give similar results but degrade if frozen for 2 weeks
Archaeological: Pleistocene bone	1993	77	
Zoological: modern animal droppings	1992	82	

Materials.

1. Extraction buffer (4 M guanidine thiocyanate, 0.05 M Tris–HCl pH 7.6, 0.01 M EDTA, 1% Triton X-100).
2. The silica suspension: 12 g Sigma S-5631 is mixed with 100 mL sterile filtered distilled water and left to sediment for 24 h; the top 86 mL of liquid (water) are removed and discarded. The volume is made up with distilled water to 100 mL and the silica resuspended and left to sediment for 5 h. Finally, the top 88 mL are removed and discarded. 120 µL of 10 M HCl are added to the sediment, resuspended and aliquoted in sterile eppendorf tubes and kept at 4 °C.
3. Wash buffer (extraction buffer with 150 µL sterile filtered distilled water added).

Methods for any Substrate.

1. The pulverised bone (1 g) or shredded blood stain or 1 mL whole blood is incubated in 1 mL extraction buffer at 60 °C for 1 h.
2. Centrifuge at 4000 *g* for 5 min.
3. 500 µL of supernatant are added to 500 µL extraction buffer and 40 µL silica suspension.
4. Rotate the mixture for 10 min at room temperature.
5. Centrifuge at 4000 *g* for 15 min.
6. Wash twice with wash buffer.
7. Wash twice with 70% ethanol.
8. Centrifuge at 4000 *g* for 5 min.
9. Dry the resulting silica pellet for 10 min at 56 °C, then resuspend in 65 µL sterile filtered distilled water and leave at 56 °C with occasional mixing.
10. Centrifuge the suspension at 4000 *g* for 3 min and transfer the supernatant to a clean tube.
11. The remaining silica pellet is resuspended, centrifuged for a second time and the supernatant removed and pooled with the previous one.
12. The DNA solution (pooled supernatants) is finally centrifuged at 6000 *g* for 5 min and the supernatant removed and stored at −20 °C.

3.3.1.3 Variations in Protocol and References for Other Substrates Glass milk has been used in clinical settings for the recovery of DNA from minute amounts of serum and plasma (50 µL)[22] and also after PCR in order to further purify DNA.[79] The silica glass milk method is currently the optimal one for archaeological material and is very likely to be useful in the forensic field when dry bone is being extracted.

In a comparative study between Dynal DNA-direct™, sodium acetate precipitation and glass milk extraction methods, on dry bone and blood stains over 40 years old, the present authors found that amplifiable DNA was obtained more often with the glass milk technique, although the quantity of DNA was less.[40]

Table 3 presents a series of references from authors who have used different silica affinity protocols on various substrates.

3.3.2 Specific Affinity Methods: Monoclonal Antibodies

This method employs anti-dsDNA IgG-coated beads.

3.3.2.1 Method Rationale, Advantages and Disadvantages

Rationale. Immunoaffinity methods are those based on the use of antibodies devised specifically for detecting and attaching to DNA. They are scarce and, up to the time of writing, only one protocol has been published,[84] although a full description of the development of the method is available.[32] Briefly, the technique consists of the production of a monoclonal antibody to double-stranded DNA; it then uses magnetic beads coated with the antibody to specifically select DNA from a solution (Figure 3). Although contamination of the monoclonal anti-DNA with murine and bovine DNA might occur during preparation, neither reduced sensitivity due to blocked antibody sites nor false positive PCR reactions from species cross-reactivity have been observed. Contamination with human DNA, always a potential problem with such a sensitive method, can be avoided by meticulous technique and by including antibody-coated

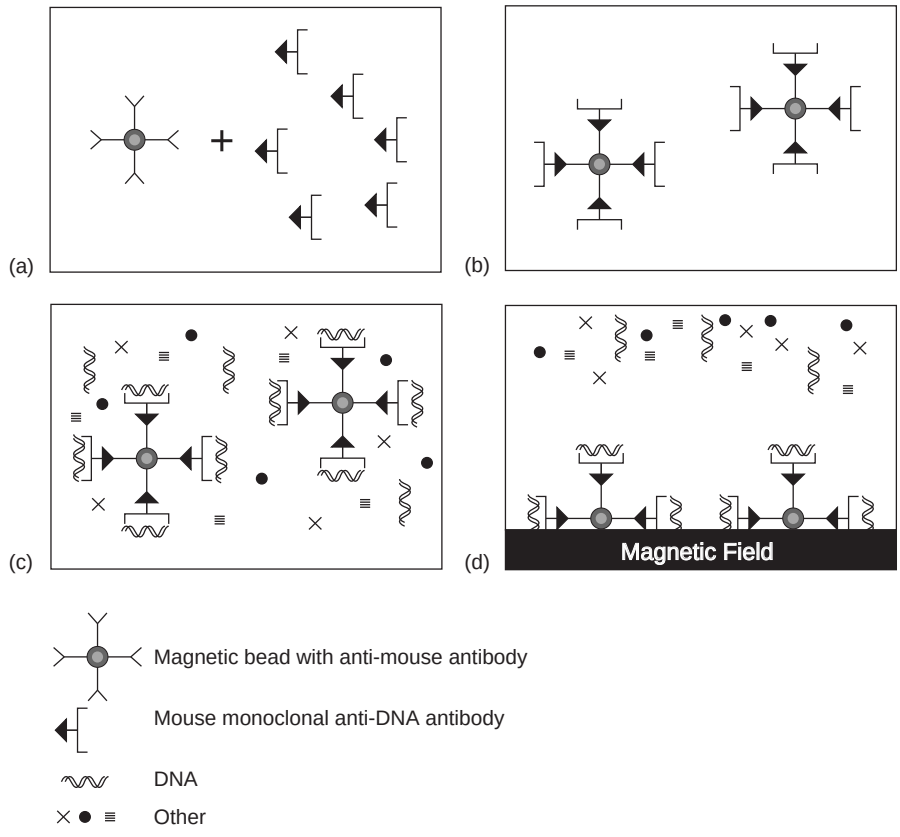

Figure 3 Diagram of the specific immunoaffinity method. (a) Mouse monoclonal anti-dsDNA IgG. (b) Binds to the magnetic beads coated with anti-mouse immunoglobulins. (c) When incubated with a solution containing DNA and other contaminants, the antibody-coated beads selectively attract DNA. (d) The beads bearing the anti-DNA–DNA complex are then separated from the remaining solution by a magnetic field.

beads on their own in every experiment as a negative control.

Advantages and Disadvantages. The great advantage of this method is that the DNA is selectively bound by a monocolonal antibody. This represents the maximum specificity one can achieve for selecting DNA from a solution and positive results can be obtained with as little as 0.005 ng per microliter of DNA. Another advantage is that the beads–anti-DNA–DNA complex can be put directly into the PCR solution.[84]

The disadvantage lies in the fact that the method has not yet been tested sufficiently on forensic material, and, at the moment, is quite expensive. Also, although the method is very specific, it does not yield large volumes of DNA.

3.3.2.2 Protocol Using Anti-DNA Antibody-coated Magnetic Beads on a Single Substrate[84]

Materials.

1. Murine monoclonal antibody (IgG2A class) to dsDNA are bound on to paramagnetic beads coated with anti-mouse IgG by incubating at 20 °C for 2 h. The beads are then washed and resuspended to a concentration of 4×10^8 per milliliter.
2. Digestion buffer (0.5% Na_2EDTA, 0.1 M Tris base, pH 7.8, 0.1 M NaCl, 1% Nonidet P-40).
3. PK (20 mg per milliliter).
4. Fetal calf serum.
5. Isotonic saline.
6. Distilled water.

Methods.

1. Test samples (e.g. 0.05 mL fresh blood, 1 cm^2 blood-stained gauze, 0.5 g powdered bone) are incubated at 56 °C in 3 mL digestion buffer and 100 µL PK for up to 36 h for bone (30 min for fresh blood).
2. Supernatants are collected and approximately 20×10^5 antibody-coated beads are added and incubated at 20 °C with constant mixing for varying periods (2 h for blood, 24 h for bone).
3. The beads are recovered magnetically and washed three times in 3% fetal calf serum in

isotonic saline and once in sterile filtered distilled water.

4. The beads are then suspended in either sterile filtered distilled water or with the reaction mixture and immediately subjected to PCR (approximately 16×10^5 beads per reaction).

To our knowledge no other protocol has been devised based on immunoaffinity. Preliminary studies of this method on dry bone and blood stains over 40 years old have given positive results for HLA DR loci,[32] suggesting that it has a promising future in forensic DNA analysis.

3.4 Commercial Methods

Commercial kits are being increasingly used and several provide protocols for forensic substrates. They are usually based on an affinity mechanism. Although one cannot comment on their specific mode of action, they seem reliable on scarce forensic substrates in some cases. Detailed protocols for their use are supplied with the kits. The ones usually referenced in the forensic literature are DNA direct™ from Dynal,[75] QIAamp® (QIAGEN),[85] Prep-A-Gene®,[86] and GeneClean®[77] DNA extraction kits.

Dynabeads DNA Direct™ are uniform superparamagnetic polystyrene beads designed for the simple and rapid isolation of DNA from crude materials. The process relies upon the rapid adsorption of the released DNA on to the surface of the Dynabeads. The DNA–bead complex is then separated magnetically and washed.[75] One study compared the DNA yield from feces between phenol/chloroform and the Dynal DNA direct™ kit. The latter gave by far the better results, being faster and more reliable.[24]

QIAamp® DNA kits are designed for the rapid purification of DNA from various substrates. DNA is adsorbed on to the QIAamp® silica gel membrane during a brief centrifugation or vacuum step, then washed, eluted and concentrated.[85] When compared to phenol/chloroform and Chelex methods in the extraction of DNA from blood stains and mixed sperm/epithelial substrates, it proved more efficient.[87]

regGeneClean® DNA Extraction kit[77] and the Bio-Rad Prep-A-Gene® DNA purification kit[86] are also based on the use of a silica gel matrix and have proved useful in extracting DNA from substrates such as plasma and serum.[22]

Some commercial kits have also been used in the study of botanical specimens of toxicological interest.[88]

4 CONCLUSIONS

The aim of this article is two-fold. First, it evaluates the difficulties in DNA extraction for PCR technology in forensic scenarios, considers the versatility of DNA survival, and stresses the need for the extraction method to relate to the substrate. Second, it provides at least two protocols (one for easy substrates and one for difficult ones) selected from the existing forensic literature for each basic extraction method. It is difficult to provide a method of choice for some situations, and techniques are evolving and improving all the time. Nevertheless some general conclusions can be drawn from our own experience and from examining the existing literature.

Methods involving phenol/chloroform have been used since the beginning of RFLP and PCR technology, and therefore have a better tested record with the widest variety of substrates. Although they have been very successful, they are dangerous, toxic, time-consuming and involve many steps. Furthermore, they are not directly selective for DNA. At present, DNA extraction methods in forensic analysis seem to be going two ways:

- For clinical material, with fresh, abundant sources of DNA, extraction presents no problems – salting out methods are just as successful as methods using organic solvents and are far less toxic. Commercial extraction kits provide a suitable alternative.
- The most interesting evolution in extraction methods is taking place with difficult substrates, where DNA is scarce and contamination abundant. In these cases, recent work has suggested that the future lies in affinity methods – this would seem a more logical approach and is supported from a physical and chemical point of view. The extracted DNA is purer and present in relatively greater amounts (with respect to non-DNA contaminants).

Affinity methods, such as the use of glass particles or (even better) monoclonal antibodies, which select out DNA, seem to be the techniques of choice for difficult substrates. If the substrate is not extremely scarce, mixed procedures could be used (e.g. phenol/chloroform and glass beads). The only drawback of affinity methods is that they have not yet been routinely tested on most forensic substrates and so do not have a proven record. It is for forensic research to remedy this and, together with data from other disciplines concerning the physical and chemical interaction of DNA with various reagents, to produce tailor-made extraction methods for the plethora of scenarios encountered daily in forensic practice. Only with proper extraction procedures will subsequent DNA analysis be successful.

ACKNOWLEDGMENTS

We thank Prof. M. Grandi and Prof. A. Farneti of the Institute of Legal Medicine in Milan for granting

Dr C. Cattaneo study leave to prepare this article and Mrs. C.A. Mitchell for secretarial assistance.

ABBREVIATIONS AND ACRONYMS

ANE	Sodium Acetate, Sodium Chloride, Ethylenediaminetetraacetic Acid
CTAB	Cetyltrimethylammonium Bromide
DTAB	Dodecyltrimethylammonium Bromide
DTT	Dithiotreitol
EDTA	Ethylenediaminetetraacetic Acid
PBS	Phosphate-buffered Saline
PBS/S	Phosphate-buffered Saline + 2% Sarcosyl
PCR	Polymerase Chain Reaction
PK	Proteinase K
PKS	Proteinase K + Sodium Dodecyl Sulfate
RFLP	Restriction Fragment Length Polymorphism
SDS	Sodium Dodecyl Sulfate
TE	Tris(hydroxymethyl)aminomethane–Ethylenediaminetetraacetic Acid
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

RELATED ARTICLES

Forensic Science **(Volume 5)**
Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Molecules, Properties and Detection of Single • Polymerase Chain Reaction and Other Amplification Systems • Sequencing Strategies and Tactics in DNA and RNA Analysis

REFERENCES

1. C.M. Strom, S. Rechitsky, 'Use of Nested PCR to Identify Charred Human Remains and Minute Amounts of Blood', *J. Forensic Sci.*, **43**(3), 696–700 (1998).
2. T. Lindahl, 'Instability and Decay of the Primary Structure of DNA', *Nature*, **362**, 709–715 (1993).
3. T.A. Brown, K.A. Brown, 'Ancient DNA and the Archaeologist', *Antiquity*, **66**, 10–23 (1992).
4. P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully et al., 'Identification of the Remains of the Romanov Family by DNA Analysis', *Nat. Genet*, **6**, 130–135 (1994).
5. S. Hummel, B. Herrmann, 'Y-chromosome-specific DNA Amplified in Ancient Human Bone', *Naturwissenschaften*, **78**, 266–267 (1991).
6. E. Hagelberg, L.S. Bell, T. Allen, A. Boyde, S.J. Jones, J.B. Clegg, 'Analysis of Ancient Bone DNA: Techniques and Applications', *Phil. Trans. R. Soc. London, B*, **333**, 399–407 (1991).
7. C. Hänni, V. Laudet, M. Sakka, A. Bègue, D. Stéchelin, 'Amplification de Fragments d'ADN Mitochondrial à Patir de Dents et d'os Humains Anciens', *C R Acad. Sci.*, **310**, 365–370 (1990).
8. D.E.G. Briggs, 'Fossil Biomolecules', *The Biochemist*, Aug/Sept, 8–12 (1993).
9. E.R. Waite, A.M. Child, O.E. Craig, M.J. Collins, K. Gelsthorpe, T.A. Brown, 'A Preliminary Investigation of DNA Stability in Bone During Artificial Diagenesis', *Bull. Soc. Géol. France*, **168**(5), 547–554 (1997).
10. A. Akane, K. Matsubara, H. Nakamura, S. Takahashi, K. Kinura, 'Purification of Highly Degraded DNA by Gel Filtration for PCR', *BioTechniques*, **16**, 235–238 (1994).
11. J.A. Mountzouris, L.H. Hurley, 'Small Molecules-DNA Interaction', in *Bioorganic Chemistry*, ed. S.M. Hecht, Oxford University Press, Oxford, 288–323, 1996.
12. C.J. Larson, G.L. Veroline, 'The Chemistry of Protein-DNA Interactions', in *Bioorganic Chemistry*, ed. S.M. Hecht, Oxford University Press, Oxford, 324–346, 1996.
13. H.M. Berman, B. Schneider, 'Nucleic Acid Hydration', in *Nucleic Acid Structure*, ed. S. Neidle, Oxford Science Publications, Oxford, 295–310, 1996.
14. T. Toledano, L. Quarino, S. Leving, P. Buffolino, H. Baum, R. Shaler, 'An Assessment of DNA Contamination Risks in New York City Medical Examiner Facilities', *J. Forensic. Sci.*, **42**(4), 721–724 (1997).
15. M.N. Hochmeister, 'PCR Analysis of DNA from Fresh and Decomposed Bodies and Skeletal Remains in Medicolegal Death Investigations', *Methods Mol. Biol.*, **98**, 19–26 (1998).
16. P.M. Schneider, 'Recovery of High-molecular Weight DNA from Blood and Forensic Specimens', *Methods Mol. Biol.*, **98**, 1–7 (1998).
17. M. Allen, T. Saldeen, U. Gyllensten, 'Allele-specific HLA-DRB1 Amplification of Forensic Evidence Samples with Mixed Genotypes, *BioTechniques*, **19**, 454–463 (1995).
18. E. Jahaes, A. Gilissen, J.-J. Cassiman, R. Decorte, 'Evaluation of a Decontamination Protocol for Hair Shafts before mtDNA Sequencing', *Forensic Sci. Int.*, **94**, 65–71 (1998).
19. R.A.H. van Oorschot, M.K. Jones, 'DNA Fingerprints from Fingerprints', *Nature*, **387**, 767 (1997).
20. M. Lorente, C. Entrala, J.A. Lorente, J.C. Alvarez, E. Villanneva, B. Budowle, 'Dandruff as a Potential Source of DNA Forensic Casework', *J. Forensic. Sci.*, **43**(4), 901–902 (1998).
21. B. Brinkmann, S. Rand, T. Bajanowski, 'Forensic Identification of Urine Samples', *Int. J. Legal Med.*, **105**, 59–61 (1992).

22. K.R. Fowke, F.A. Plummer, J.N. Simonsen, 'Genetic Analysis of Human DNA Recovered from Minute Amounts of Serum or Plasma', *J. Immunol. Methods*, **180**, 45–51 (1994).
23. C.A. Linch, S.C. Smith, J.A. Prahlow, 'Evaluation of the Human Hair Root for DNA Typing Subsequent to Microscopic Comparison', *J. Forensic Sci.*, **43**(2), 305–314 (1998).
24. A.J. Hopgood, A. Mannucci, K.M. Sullivan, 'DNA Typing from Human Faeces', *Int. J. Legal Med.*, **108**, 237–243 (1996).
25. A.-M. Vachot, M. Monnerot, 'Extraction, Amplification and Sequencing of DNA from Formaldehyde-fixed Specimens', *Ancient Biomol.*, **1**(1), 3–16 (1996).
26. D.N. Fredricks, D.A. Redman, 'Paraffin Removal from Tissue sections for Digestion and PCR Analysis', *BioTechniques*, **26**, 198–200 (1999).
27. N. Dimo-Simonin, F. Grange, C. Brandt-Casaderall, 'PCR-based Forensic Testing on DNA from Stained Cytological Smears', *J. Forensic Sci.*, **42**(3), 506–509 (1997).
28. S. Hummel, G. Nordsiek, B. Herrmann, 'Improved Efficiency in Amplification of Ancient DNA and its Sequence Analysis', *Naturwissenschaften*, **79**, 359–360 (1992).
29. M. Scholz, I. Giddings, C.M. Pusch, 'A Polymerase Chain Reaction Inhibitor of Ancient Hard and Soft Tissue DNA Extracts is Determined as Human Collagen Type 1', *Anal. Biochem.*, **259**, 283–286 (1998).
30. D.L. Fisher, M.M. Holland, L. Mitchell, P.S. Sledzik, A.W. Wilcox, M. Wadhams, V.W. Weedn, 'Extraction, Evaluation and Amplification of DNA from Decalcified and Undecalcified United States Civil War bone', *J. Forensic Sci.*, **38**(1), 60–68 (1993).
31. D. Sweet, D. Hildebrand, 'Recovery of DNA from Human Teeth by Cryogenic Grinding', *J. Forensic Sci.*, **43**(6), 1199–1202 (1998).
32. A.R. Gelsthorpe, 'Development of an Affinity Method for the Extraction of DNA', PhD thesis, University of Sheffield, Sheffield, UK, 1996.
33. M.B. John, J.E. Paulus-Thomas, 'Purification of Human Genomic DNA from Whole Blood Using Sodium Perchlorate in Place of Phenol', *Anal. Biochem.*, **180**, 276–278 (1989).
34. S. Pääbo, 'Molecular Cloning of Ancient Egyptian Mummy DNA', *Nature*, **314**, 644–645 (1985).
35. E. Hagelberg, J.B. Clegg, 'Isolation and Characterization of DNA from Archaeological Bone', *Proc. R. Soc. London, B*, **244**, 45–50 (1991).
36. P. Chomcyznski, 'A Reagent for the Single-step Simultaneous Isolation of RNA, DNA and Proteins from Cell and Tissue Samples', *BioTechniques*, **15**, 532–536 (1993).
37. T. Maniatis, E.F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York, 468, 1982.
38. L. Tilzer, S. Thomas, R.F. Moreno, 'Use of Silica Gel Polymer for DNA Extraction with Organic Solvents', *Anal. Biochem.*, **183**, 13–15 (1989).
39. C. Cattaneo, D.M. Smillie, K. Gelsthorpe, A. Piccinini, A.R. Gelsthorpe, R.J. Sokol, 'A Simple Method for Extracting DNA from Old Skeletal Material', *Forensic Sci. Int.*, **74**, 167–174 (1995).
40. C. Cattaneo, O.E. Craig, N.T. James, R.J. Sokol, 'Comparison of Three DNA Extraction Methods on Bone and Blood Stains up to 43 Years Old and Amplification of Three Different Gene Sequences', *J. Forensic Sci.*, **42**(6), 1126–1135 (1997).
41. M. Benecke, 'Random Amplified DNA (RAPD) Typing of Necrophageous Insects (diptera, coleoptera) in Criminal Forensic Studies: Validation and Use in Practice', *Forensic Sci. Int.*, **98**, 157–168 (1998).
42. M.N. Hochmeister, O. Rudin, E. Ambach, 'PCR Analysis from Cigaret Butts, Postage Stamps, Envelope Sealing Flaps and Other Saliva Stained Material', *Methods Mol. Biol.*, **98**, 27–32 (1998).
43. A. Akane, 'Hydrogen Peroxide Decomposes the Heme Compound in Forensic Specimens and Improves the Efficiency of PCR', *BioTechniques*, **21**(3), 392–394 (1996).
44. C.L. Fox, 'Analysis of Ancient Mitochondrial DNA from Extinct Aborigines from Tierra del Fuego-Patagonia', *Ancient Biomol.*, **1**(1), 43–54 (1996).
45. T.M. Clayton, J.P. Whitaker, C.M. Maguire, 'Identification of Bodies from the Scene of a Mass Disaster Using DNA Amplification of Short Tandem Repeat (STR) Loci', *Forensic Sci. Int.*, **76**, 7–15 (1995).
46. Z. Lin, T. Kondo, T. Minamino, M. Ohtsuji, J. Nishigami, T. Takasayuh, R. Sun, T. Ohshima, 'Sex Determination by Polymerase Chain Reaction on Mummies Discovered at Taklamakan Desert in 1912', *Forensic Sci. Int.*, **75**, 197–205 (1995).
47. C. Lassen, S. Hummell, B. Herrmann, 'Comparison of DNA Extraction and Amplification from Ancient Human Bone and Mummified Soft Tissue', *Int. J. Legal Med.*, **107**, 152–155 (1994).
48. H. Pfitzinger, B. Ludes, P. Mangin, 'Sex Determination of Forensic Samples: Co-amplification and Simultaneous Detection of a Y-specific and an X-specific DNA Sequence', *Int. J. Legal Med.*, **105**, 213–216 (1993).
49. F.A.H. Sperling, G.S. Anderson, D.A. Hickey, 'A DNA-based Approach to the Identification of Insect Species used for Postmortem Interval Estimation', *J. Forensic Sci.*, **39**(2), 418–427 (1994).
50. A.J. Jeffreys, M. Allen, E. Hagelberg, A. Sonnberg, 'Identification of the Skeletal Remains of Josef Mengele by DNA Analysis', *Forensic Sci. Int.*, **56**, 65–76 (1992).
51. C. Ginther, L. Issel-Tarver, M.-C. King, 'Identifying Individuals by Sequencing Mitochondrial DNA from Teeth', *Nat. Genet*, **2**, 135–138 (1992).
52. K.M. Sullivan, R. Hopwood, P. Gill, 'Identification of Human Remains by Amplification and Automated

Sequencing of Mitochondrial DNA', *Int. J. Legal Med.*, **105**, 83–86 (1992).
53. E. Hagelberg, I.C. Gray, A.J. Jeffreys, 'Identification of the Skeletal Remains of a Murder Victim by DNA Analysis', *Nature*, **352**, 427–429 (1991).
54. O. Pascal, D. Aubert, E. Gilbert, J.P. Moison, 'Sexing of Forensic Samples Using PCR', *Int. J. Legal Med.*, **104**, 205–207 (1991).
55. A. Sajantila, M. Ström, B. Budwole, P.J. Karhunen, L. Peltonen, 'The Polymerase Chain Reaction and Post-mortem Forensic Identity Testing: Application of Amplified D1S80 and HLA-DQ∝ loci to the Identification of Fire Victims', *Forensic Sci. Int.*, **51**, 23–34 (1991).
56. I. Thuesen, J. Engberg, 'Recovery and Analysis of Human Genetic Material from Mummified Tissue and Bone', *J. Archaeol Sci.*, **17**, 679–689 (1990).
57. S. Horai, K. Hayosaka, K. Murayama, N. Wate, H. Koike, N. Nakai, 'DNA Amplification from Ancient Human Skeletal Remains and their Sequence Analysis', *Proc. Jpn. Acad.*, **65**(B), 229–233 (1989).
58. R. Higuchi, C.H. von Berholdingen, G.F. Sensabaugh, H.A. Erlich, 'DNA Typing from Single Hairs', *Nature*, **332**, 543–546 (1988).
59. J. Laitinen, J. Samarut, E. Hölttä, 'A Nontoxic and Versatile Protein Salting-out Method for Isolation of DNA', *BioTechniques*, **17**(2), 316–322 (1994).
60. S.A. Miller, D.D. Dykes, H.F. Polesky, 'A Simple Salting Out Procedure for Extracting DNA from Human Nucleated Cells', *Nucleic Acids Res.*, **16**(3), 1215 (1988).
61. J.M. Willard, D.A. Lee, M.M. Holland, 'Recovery of DNA for PCR Amplification from Blood and Forensic Samples Using a Chelating Resin', *Methods Mol. Biol.*, **98**, 9–18 (1998).
62. J. Singer-Sam, R.L. Tanguay, A.D. Riggs, 'Use of Chelex to Improve the PCR Signal from a Small Number of Cells', *Amplifications: A Forum for PCR users*, **3**, 11 (1989).
63. P.S. Walsh, D.A. Metzger, R. Higuchi, 'Chelex 100 as a Medium for Simple Extraction of DNA for PCR-based Typing from Forensic Material', *BioTechniques*, **10**, 507–513 (1991).
64. D. Sweet, M. Lorente, A. Valenzuela, J.A. Lorente, J.C. Alvarez, 'Increasing DNA Extraction Yield from Saliva Stains with a Modified Chelex Method', *Forensic Sci. Int.*, **83**, 167–177 (1996).
65. L. Casarino, A. Mannucci, C.P. Kimpton, S. Presciuttini, G. Bruni, M.G. Costa, F. De Stefano, 'Forensic Evaluation of HUMCD4: An Italian Database', *Int. J. Legal Med.*, **109**, 49–51 (1996).
66. A. Alvarez-Garcia, I. Munoz, C. Pestoni, M.V. Lareu, M.S. Rodriguez-Calvo, A. Carracedo, 'Effect of Environmental Factors on PCR–DNA Analysis from Dental Pulp', *Int. J. Legal Med.*, **109**, 125–129 (1996).
67. R. Sepp, I. Szabó, H. Uda, H. Sakamoto, 'Rapid Techniques for DNA Extraction from Routinely Processed Archival Tissue for use in PCR', *J. Clin. Pathol.*, **47**, 318–323 (1994).
68. R.J. Cano, H.N. Poinar, N.J. Pieniazek, A. Acra, G.O. Poinar, 'Amplification and Sequencing of DNA from a 120–135-million-year-old Weevil', *Nature*, **363**, 536–538 (1993).
69. S. Gustincich, G. Manfioletti, G. Del Sal, C. Schneider, 'A Fast Method for High-quality Genomic DNA Extraction from Whole Human Blood', *BioTechniques*, **3**, 298–301 (1991).
70. L. Rudbeck, J. Dissing, 'Rapid, Simple Alkaline Extraction of Human Genomic DNA from Whole Blood, Buccal Epithelial Cells, Semen and Forensic Stains for PCR', *BioTechniques*, **25**(4), 588–592 (1998).
71. S. Oshiro, N. Katsura, K. Kutida, N. Gunge, 'A Simple Method for Extraction of DNA from Fungi and Yeasts with Anhydrous Hydrogen Fluoride', *FEBS Lett.*, **220**, 283–286 (1987).
72. V.N. Petrishchev, A.B. Kutueva, Yu.G. Rychkov, 'A Simple and Efficient Method for DNA Extraction from Ancient Bones for the use in the Polymerase Chain Reaction', *Genetika*, **29**, 690–693 (1993).
73. H. Meijer, W.R.K. Perizonius, J.P.M. Geraedts, 'Recovery and Identification of DNA Sequences Harboured in Preserved Ancient Human Bones', *Biochem. Biophys. Res. Commun.*, **183**(2), 367–374 (1992).
74. R. Boom, C.J. Sol, M.M. Salimans, C.L. Jansen, P.M. Wertheim-van Dillen, J. van der Noordaa, 'Rapid and Simple Method for Purification of Nucleic Acids', *J. Clin. Microbiol.*, **28**(3), 495–503 (1990).
75. Dynabeads DNA Direct™, *Dynabeads for the Isolation of PCR-ready DNA*, Dynal (UK) Ltd, Bromsborough, UK, 1999 (http://www.dynal.no).
76. H. Poinar, 'Glass Milk, a Method of Extracting DNA from Fossil Material', *Ancient DNA Newsletter*, **2**(1), 12–13 (1994).
77. GeneClean®, *Guide to Protocols and Procedures*, Bio101, 1999 (http://www.bio101.com).
78. M. Höss, S. Pääbo, 'DNA Extraction from Pleistocene Bones by a Silica-based Purification Method', *Nucleic Acids Res.*, **21**(16), 3913–3914 (1993).
79. L.S. Smith, T.L. Lewis, S.M. Matsui, 'Increased Yield of Small DNA Fragments Purified by Silica Binding', *BioTechniques*, **18**, 970–975 (1995).
80. A.C. Stone, G.R. Milner, S. Pääbo, M. Stoneking, 'Sex Determination of Ancient Human Skeletons Using DNA', *Am. J. Phys. Anthropol.*, **99**, 231–238 (1996).
81. A.C. Stone, M. Stoneking, 'Genetic Analysis of an 8000 Year-old Native American Skeleton', *Ancient Biomol.*, **1**(1), 83–87 (1996).
82. R.J. Cano, H.N. Poinar, 'Rapid Isolation of DNA from Fossil and Museum Specimens Suitable for PCR', *BioTechniques*, **15**, 432–436 (1993).
83. M. Höss, M. Kohn, S. Pääbo, F. Knauer, W. Schröder, 'Excrement Analysis by PCR', *Nature*, **359**, 199 (1992).

84. A.R. Gelsthorpe, K. Gelsthorpe, R.J. Sokol, 'Extraction of DNA Using Monoclonal Anti-DNA and Magnetic Beads', *BioTechniques*, **22**(6), 1082–1083 (1997).
85. QIAamp®, *DNA Mini Kit and QIAamp® DNA Blood Mini Kit Handbook*, QIAGEN, 1999 (http://www.qiagen.com).
86. Prep-A-Gene®, *DNA Purification Kit*, 1999 (http://www.biorad.com).
87. S.A. Greenspoon, M.A. Scarpetta, M.C. Drayton, S.A. Turek, 'QIamp Spin Columns as a Method of DNA Isolation for Forensic Casework', *J. Forensic Sci.*, **43**(5), 1024–1030 (1998).
88. A. Linacre, J. Thorpe, 'Detection and Identification of Cannabis by DNA', *Forensic Sci. Int.*, **91**, 71–76 (1998).

Fluorescence in Forensic Science

E. Roland Menzel
Center for Forensic Studies, Texas Tech University, Lubbock, USA

1 Introduction 1
2 Spectra 2
2.1 Excitation Spectrum 2
2.2 Emission Spectrum 3
3 Lifetimes 3
3.1 Time-gated Measurement 3
3.2 Phase-resolved Measurement 3
4 Photoluminescence in Fingerprint Detection 3
4.1 Light Sources 4
4.2 Inherent Fingerprint Fluorescence 4
5 Physical Fingerprint Treatments 4
5.1 Dusting with Fluorescent Powder 5
5.2 Staining with Fluorescent Dye 5
6 Chemical Fingerprint Treatments 5
6.1 Ninhydrin and Ninhydrin Analogs 5
6.2 1,8-Diazafluoren-9-one 6
6.3 Miscellaneous Treatments 7
7 Fingerprint Treatments of Future Potential 7
7.1 Lanthanide-based Fingerprint Detection 7
7.2 Photoluminescent Semiconductor Nanoparticles 8
8 Time-resolved Imaging 8
9 Fluorescence in Trace Evidence Analysis 9
9.1 Fiber Analysis 9
9.2 Document Examination 9
9.3 Miscellaneous Trace Evidence Detection 10
Abbreviations and Acronyms 10
Related Articles 10
References 11

Emission spectroscopy, which in the forensic science context historically has meant (carbon) arc spectroscopy, is occasionally still employed in paint examination and some other trace evidence areas, but has largely been replaced by atomic absorption (AA), Fourier transform infrared (FTIR), gas chromatography/mass spectrometry (GC/MS) and other analytical techniques. Notwithstanding its intrinsically very high sensitivity, molecular fluorescence spectroscopy is not utilized as a general technique of forensic analysis either, because spectra tend to be broad and featureless. Fluorescence spectral examination may occasionally be used in areas such as document examination, specifically for ink differentiation in instances of document alteration, but fluorescence spectra as such tend to pertain primarily to the arena of research. This is not to say, however, that the photoluminescence phenomenon is not exploited in criminalistics, the examination of articles of physical evidence recovered from the crime scene. Indeed, it is widely and effectively employed in tagging or labeling modalities and via the inherent fluorescence itself of trace evidence. Fluorescence-based approaches to authentication of currency and passports, identification of valuables, origin specification of gun powders and explosives, surveillance, and so on, are less directly pertinent to forensic science. These areas of fluorescence tagging will thus not be dealt with in detail. In criminalistics, photoluminescence is today the mode par excellence for ultrasensitive latent fingerprint detection. It is also employed in detection of body fluids, visualization of fibers, ink discrimination, and so on. Thus, the focus in this article will be on visualization and imaging. We begin, however, with discussion of the basic molecular photoluminescence phenomenon and description of some useful spectroscopic measurements, namely luminescence and excitation spectra, and luminescence lifetime determination by time-gated and phase-resolved techniques. These set the stage for the latest advances in photoluminescence visualization in physical evidence processing, with emphasis on fingerprints, the most prominent photoluminescence application in forensic science.

1 INTRODUCTION

The emission of light by substances can have various causes, among them heat (thermoluminescence or incandescence), electrical charge (electroluminescence, cathodoluminescence) and chemical reaction (chemiluminescence), as in the reaction of luminol with blood. However, for our purposes the most important origin of the luminescence is the prior absorption of light, giving rise to photoluminescence. The phenomenon was studied long before its quantum mechanical underpinnings were understood. Early on, two categories of photoluminescence were distinguished on the basis of lifetime, namely fluorescence, associated with short lifetime, and phosphorescence, with long lifetime. The slowness concept of distinction between the two light emissions evolved

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

over time, with a trend toward shorter decay times. At first, phosphorescence involved emission with a lifetime longer than roughly a few seconds. Today, emissions called phosphorescence may involve lifetimes shorter than microseconds. The distinction between fluorescence and phosphorescence is placed on a less arbitrary footing as follows. The photoluminescent (typically organic) molecules of interest to good approximation can be thought of as being two-electron systems. When the molecule is in its ground state, the two spectroscopically active electrons fill the highest occupied molecular orbital (HOMO). Their spins must be antiparallel to satisfy the exclusion principle and the corresponding state is thus a singlet state. When one of the two electrons resides in the lowest unoccupied molecular orbital (LUMO), the molecule is in the excited state that gives rise to photoluminescence. If no spin flip has taken place in the process of electron promotion to LUMO, the excited state is a singlet state, and the decay to the (singlet) ground state, accompanied by the emission of a photon, is termed fluorescence. If a spin flip has taken place (which is quantum mechanically legal because the two electrons no longer occupy the same spatial orbital), the excited state is a triplet state and the decay back to the ground state via emission of a photon is termed phosphorescence. We generalize the distinction between fluorescence and phosphorescence on the basis of the spins of the two states between which the light-emitting transition takes place. The electric dipole operator which describes the photon emission does not operate on spin. When states of the same spin multiplicity are involved, we thus have fluorescence. Otherwise, we have phosphorescence, with violation of the basic $\Delta S = 0$ spin selection rule, hence the longer lifetime. This definition is not universally followed. For instance, one often hears of fluorescence of lanthanides, even though the involved states differ in spin multiplicities $(2S + 1)$, often even by spin quantum number change greater than 1. Recombination and trap luminescences in semiconductors are a different matter still. Thus, the term photoluminescence will be utilized in general, with the term fluorescence reserved for the transition between molecular singlet states.

In photoluminescence visualizations associated with examination of articles of evidence, one is concerned with the optimization of the wavelength of the incident light to maximize the intensity of the photoluminescence of interest, and with the wavelength of the photoluminescence itself, for optimization of contrast with background fluorescences that almost invariably are present. Excitation spectral measurement aids the first of the above two optimizations and emission spectra pertain to the second one. In some instances, background fluorescence is extremely intense and of the same color as the photoluminescence of interest. In such instances, time-resolved visualization may be employed for the suppression of the background. Photoluminescence lifetime measurements define the optimization of this background removal.

2 SPECTRA[1,2]

The optimization of the excitation that produces a photoluminescence typically involves absorption spectroscopy in solution. If the sample of interest does not scatter light appreciably and is reasonably transparent, absorption spectroscopy of solid samples may be performed also. However, in the forensic science context one often has to contend with opaque samples and samples that contain a number of absorbing species. Thus, excitation spectroscopy is often performed to obtain the same spectral information one would ordinarily obtain from an absorption measurement.

2.1 Excitation Spectrum

Figure 1 shows the block diagram of a generic set-up for excitation spectroscopy. The light source depends on the spectral range to be covered, but might typically be a xenon arc lamp. Monochromator 1, the excitation monochromator, is wavelength-scanned while monochromator 2, the emission monochromator, is set at a fixed wavelength, at or near the photoluminescence maximum of the sample. The photoluminescence intensity variation versus excitation wavelength provides the same spectral information as absorption spectroscopy, excepting absolute optical density. This does not matter since the excitation optimization only calls for relative absorbance monitoring. The light source/excitation monochromator combination is on occasion replaced by a tunable laser.

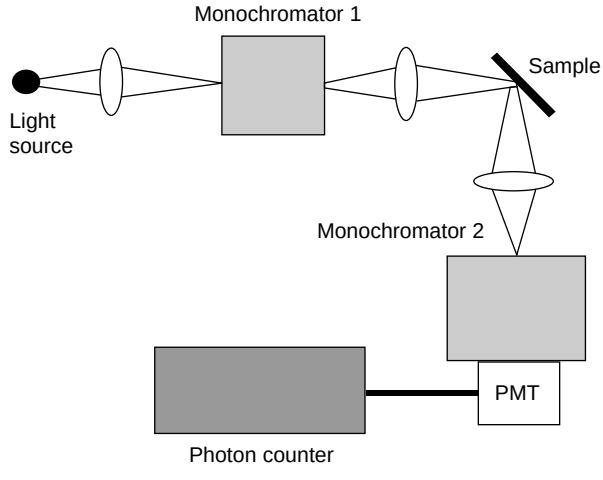

Figure 1 Excitation/emission spectroscopy instrumentation.

The emission monochromator is sometimes replaced by a band-pass filter.

2.2 Emission Spectrum

The set-up shown in Figure 1 is also applicable to measurement of emission spectra. In this case, however, monochromator 1 is set at a fixed wavelength (at which the photoluminescent material absorbs well) while monochromator 2 is scanned. The photomultiplier tube (PMT) shown in the figure and associated signal processing electronics, such as a photon counter perhaps, may be replaced by an optical multichannel analyzer, which basically is a charge-coupled device (CCD), in the form of a linear array of the photosensitive elements. With this replacement, scanning of the emission monochromator is no longer required because the array covers the spectral range of interest. Again, the light source/monochromator combination may be replaced by a laser. The emission monochromator may be replaced by a wedge filter,[2] namely a variable band-pass filter.

3 LIFETIMES

Photoluminescence lifetimes of interest to us are typically measured either by time-gated or phase-resolved techniques. We are not here concerned with picosecond or femtosecond techniques. For our purposes, time-gated techniques are best suited to lifetimes longer than microseconds and phase-resolved techniques are used for shorter lifetimes, down to the subnanosecond regime.

3.1 Time-gated Measurement

Figure 2 depicts the general approach to time-gated photoluminescence lifetime measurement and Figure 3 shows a representative experimental arrangement. The gate, of adjustable width, is scanned through the dark period of the chopped excitation, shown in Figure 3 as employing a laser. The output of the boxcar averager or similar signal processing instrument is then an exponentially decaying signal.

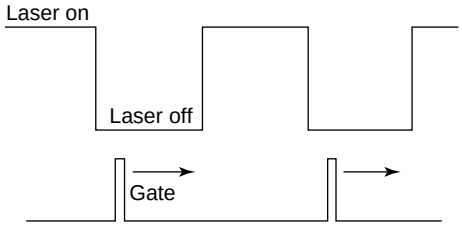

Figure 2 General scheme of time-gated luminescence lifetime measurement.

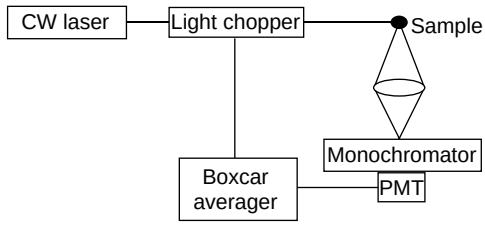

Figure 3 Instrumentation for time-gated luminescence lifetime measurement.

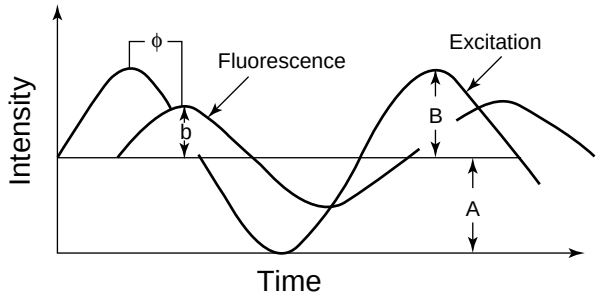

Figure 4 General scheme of phase resolved luminescence lifetime measurement. $\tan\phi = w\tau$, $m = b/B = (1 + w^2\tau^2)^{-(1/2)}$. See text.

3.2 Phase-resolved Measurement

The essence of the phase-resolved technique for photoluminescence lifetime measurement is shown in Figure 4. ϕ is the phase shift between excitation and emission, w is the modulation angular frequency, $2\pi f$, where f is the modulation frequency, τ is the luminescence lifetime, B/A is the modulation depth of the excitation light and b/A that of the luminescence (with average intensities normalized to be equal as drawn in the figure), and m is the demodulation factor. The experimental set-up is similar to that of Figure 3. However, the excitation light is now not chopped on-off, but is modulated sinusoidally in intensity. When the sample of interest displays multiple lifetimes, the phase-resolved measurement is made over a range of modulation frequencies. Lifetimes are extracted by computer analysis of the phase shifts or demodulation factors. The boxcar averager shown in Figure 3 is replaced with a phase and frequency sensitive instrument, essentially a lock-in amplifier in function.

4 PHOTOLUMINESCENCE IN FINGERPRINT DETECTION[2,3]

When a finger is pressed against an article, about a tenth of a milligram of material is typically deposited, forming a latent fingerprint. About 99% of this is water, which soon evaporates to leave behind about a

microgram of material. Roughly half of this is made up of inorganic salts, chlorides and nitrates, and the rest is a complex organic mixture including amino acids, lipids, urea, vitamins such as riboflavin, and so on.[4–6] When a fingerprint is detected, one of these organic ingredients (excepting silver nitrate fingerprint development, which targets chloride anions), or a class of them, is generally physically or chemically attacked to form either a colored or a photoluminescent product visible to the naked eye. In the absorption/reflection-based traditional colorimetric fingerprint detection procedures, such as dusting, one generally deals with a small difference between relatively large signals, namely the small difference between light reflected from a (weak) fingerprint ridge and from between ridges. This is an inherently insensitive detection mode. In the photoluminescence approach one deals with photons emitted from fingerprint ridges but not from between ridges (in absence of background fluorescence), that is with small signal detection rather than small difference detection, with the former quite generally more sensitive than the latter. Today, we are approaching single-photon sensitivity. Photoluminescence approaches are thus intrinsically ultrasensitive and widely employed in many analytical fields. The target material of a latent fingerprint is usually present only in quantities of picomole order, and yet has to luminesce strongly, to be detectable visually, on a very wide range of articles, often under dirty conditions.

4.1 Light Sources

Not only the presence but also the spatial distribution of minute quantities of material is probed in fingerprint detection. Given that photoluminescence intensity is proportional to excitation intensity (barring complications that are not operative in fingerprint work), photoluminescence detection is thus often performed using high power lasers, such as argon-ion lasers that deliver 5–20 W in the blue-green region and CW (continuous wave) frequency-doubled Nd : YAG (yttrium aluminum garnet) lasers with output powers of several watts at 532 nm. Albeit at the sacrifice of sensitivity, band-pass-filtered arc lamps are often utilized as well. Their virtues are price and portability (although portable lasers of comparable power, several hundred milliwatts, are available also). For evidence examinations not concerned with fingerprints, lamp systems offer much flexibility in terms of light color tuning. For fingerprint work, the blue-green region is mostly of interest. Near ultraviolet (UV), and at times even deep UV, are useful in special instances. Ar-lasers can operate in these spectral ranges but ordinarily operate in the blue-green range. Lasers are mandatory in time-resolved imaging approaches. Current routine detection involves illuminating the article under scrutiny with the light from the laser, dispersed with a lens or a fiber-optic cable to form an illumination spot a few centimeters in diameter, or with the filtered lamp. The article is visually examined (in a darkened room) through goggles that block the scattered or reflected illumination light but transmit the fingerprint luminescence (long-wavelength pass filter). Once a fingerprint is observed, it is photographed through the same type of filter as in the observation goggles. Band-pass filters are additionally employed occasionally for background fluorescence suppression.

4.2 Inherent Fingerprint Fluorescence

Fingerprint residue absorbs primarily in the deep UV (wavelengths shorter than 300 nm), not surprisingly given the abundance of small organic molecules in the residue. This is not a generally useful wavelength domain because the resulting photoluminescence would be expected mostly to occur in the near-UV, not visible to the eye. Moreover, there is a dearth of intense and user-friendly light sources for the deep UV, lasers in particular. Fortunately, however, there are fluorescent components (riboflavin being one of them) in the residue that absorb in the blue-green, a convenient range because powerful and user-friendly lasers operating in this range were already in existence when photoluminescence detection of fingerprints as a general technique began to be explored in 1976.[7] Inherent fingerprint fluorescence detection is a useful procedure, more so because it is nondestructive, such that all other fingerprint processing approaches can subsequently still be employed. Inherent fingerprint fluorescence detection has been used successfully in numerous cases,[8,9] but is applicable only when background fluorescence is minimal because the inherent fluorescence of fingerprints is weak even under excitation with powerful Ar-lasers. The detection applies to old as well as fresh prints. In casework, the oldest print was developed 41 years after the fact.[8] In the author's laboratory, a print dating to the early 1860s was detected. When there is prohibitive background luminescence, procedures such as dusting with luminescent powder, staining with luminescent dye and chemical treatments that lead to photoluminescent products are employed. Truly difficult situations call for time-resolved techniques.

5 PHYSICAL FINGERPRINT TREATMENTS

A number of fingerprint treatments for photoluminescence have been developed since the advent of inherent fingerprint fluorescence detection. They have mostly been designed to respond to the blue-green excitation suitable for inherent fingerprint fluorescence excitation, such that one light source suffices for most purposes.

5.1 Dusting with Fluorescent Powder

Numerous fluorescent powders are available commercially. They can also be made at home rather easily by blending fluorescent dye with iron filings, for example, to produce fluorescent magnetic powder.[2] Successful casework involving dusting for fluorescence detection of fingerprints includes the famous Polly Klaas (abduction/murder, 1993) case. As with dusting generally, the approach is applicable only to relatively fresh fingerprints because the powder adherence to fingerprint residue suffers when fingerprints dry.

5.2 Staining with Fluorescent Dye

In much the same way in which dye staining can improve visualization in the cell studies of biology, fluorescent dyes can attach preferentially to fingerprint residue. This mode of fingerprint detection was recognized early on[7] but suffered from the problem of fingerprints being washed away when the article was sprayed with the dye solution or when it was immersed in the solution. With the advent of (methyl or ethyl) cyanoacrylate ester fuming in the early 1980s, however, staining with fluorescent dye, notably rhodamine 6G,[10] became the most successful bread-and-butter fingerprint detection methodology. The cyanoacrylate fuming process was first explored in Japan and was brought to the USA by personnel of the US Army Criminal Investigation Laboratory.[11] The volatilization of the cyanoacrylate ester, the active ingredient in Superglue® and similar adhesives, is achieved by heating or placement of drops of the glue onto cotton pads that are soaked with NaOH and left to dry prior to the glue application. Commercial pouches (Hard Evidence, Loctite Corp., Newington, CT) that deliver the fumes on opening are especially useful for field work. The fuming is done in a humid atmosphere. A white polymer then forms on fingerprints and stabilizes them such that the subsequent dye staining does not wash them away. In the absence of humidity, the polymer formed is clear, as it is when the fuming utilizes a vacuum chamber, an approach that is gaining use. In the latter two situations, staining, preferential to the polymer, and fluorescence visualization are mandatory for fingerprint detection. In the former approach, visibly developed white prints may develop. From the perspective of the dye staining, however, it is not necessary that the Superglue® polymerization be that extensive. The dye staining, by dipping or spraying, typically utilizes a methanol solution. If the staining of the article is excessive, a neat methanol rinse can be applied to wash off excess dye. However, this potentially risky procedure is not usually necessary if the dye concentration is kept low, about 10^{-5} M. Coumarin dyes are sometimes employed, as are analogs of crystal violet for fingerprints on the sticky side of adhesive tapes. Lanthanide-based formulations are explicitly taken up below. A number of commercial dye formulations are in use. Rhodamine 6G, however, remains an unsurpassed option because it is a laser dye (thus endowed with very high fluorescence quantum yield) that is superbly matched to Ar-laser excitation and that is very good in its fingerprint selectivity. Old as well as fresh fingerprints are amenable to the dye staining approach. The Nightstalker case[12] is an example of successful fingerprint detection by rhodamine 6G staining/laser examination. Vapor dye staining procedures are reported in the literature for processing difficult surfaces (such as cloth and skin) but are not currently employed routinely.

6 CHEMICAL FINGERPRINT TREATMENTS

Dusting and staining approaches are applicable only to reasonably smooth surfaces since the processing of porous items, such as paper for instance, will not be fingerprint selective but will cover the whole article. Examination of articles of this kind calls for chemical procedures that involve reaction with a fingerprint constituent leading to a photoluminescent product.

6.1 Ninhydrin and Ninhydrin Analogs

Ninhydrin **(1)** was introduced to the fingerprint field in 1954.[13] The compound was first prepared by Ruhemann in 1910[14] and was recognized by him to be an amino acid assay reagent, a use it has been put to widely since then in biochemistry. Since the early 1960s, ninhydrin has become the workhorse of chemical fingerprint processing. The reagent usually is first dissolved in methanol or ethanol and the solution is then diluted with a volatile carrier solvent in order to minimize bleeding of fingerprint detail. A few grams of ninhydrin per liter of solvent system (about 90% volatile carrier) is a typical concentration. For many years, 1,1,2-trichlorotrifluoroethane

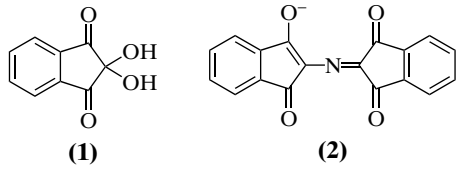

(1) **(2)**

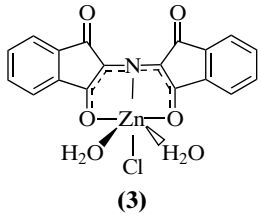

(3)

(also known as fluorisol, freon TF, freon 113, CFC 113) was the preferred volatile carrier. However, since it is ozone depleting, it is now replaced by solvents such as petroleum ether, heptane, HFE 7100® (the hydrofluoroether 1-methoxynonafluorobutane, 3M product), and so on. Acetic acid (a few milliliters) is sometimes added to counteract the alkalinity of some papers. Ethyl acetate may be added as well to counteract the esterification reaction between the alcohol and acetic acid that is accompanied by the formation of water as an undesirable by-product. The reagent concentrations are not at all critical. The reaction requires ambient humidity and is slow at room temperature, taking several hours. It can be expedited in an incubator. The ninhydrin–amino acid product **(2)**, referred to as Ruhemann's Purple, is unfortunately not fluorescent, even under optimized orange dye laser excitation. This can be remedied by a subsequent treatment with zinc chloride.[15] This salt is dissolved to saturation in methanol and diluted with a volatile carrier, much like the ninhydrin. The reaction with Ruhemann's Purple requires ambient humidity and is nearly instantaneous. The product formed **(3)** is orange in color, and thus from the absorption perspective well suited to blue-green excitation. The product displays an intense, broad fluorescence peaked in the yellow. The fluorescence increases substantially at low (near 77 K) temperature. The low-temperature fingerprint development (including photography) is relatively straight forward.[2] An example of the sensitivity gain of the (room temperature) fluorescence detection with ninhydrin/zinc chloride as compared to the corresponding traditional ninhydrin development of the same sample is shown in Figure 5. The sensitivity gain shown translates to the casework scenario.[16] Ninhydrin/zinc chloride has become routine in casework since the mid-1980s. It is applicable to old as well as fresh prints.

Benzo(*f*)ninhydrin **(4)** functions much like ninhydrin but yields a green product instead of a purple one when reacting with amino acid. When this is followed by zinc chloride and fluorescence examination, the excitation maximum shifts to about 530 nm, well matched with the frequency-doubled Nd : YAG laser.[17] Moreover, the (orange) fluorescence intensity is substantially higher at room temperature than that obtained (under comparably intense 490-nm excitation) with ninhydrin/zinc chloride. It is of high quantum yield such that little is gained by the additional complication of liquid nitrogen work. In terms of excitation and fluorescence color, 5-methoxyninhydrin **(5)** resembles ninhydrin, but in its fluorescence quantum efficiency in concert with zinc chloride it has the superior characteristics of benzo(*f*)ninhydrin.[18] The benzo and methoxy analogs of ninhydrin are commercially available (Aldrich) but are comparatively expensive. In the author's laboratory, they have nonetheless supplanted ninhydrin in casework because of sensitivity. Other ninhydrin analogs of excellent characteristics and 1,2 indanedione analogs of ninhydrin have been reported, but are not available commercially.

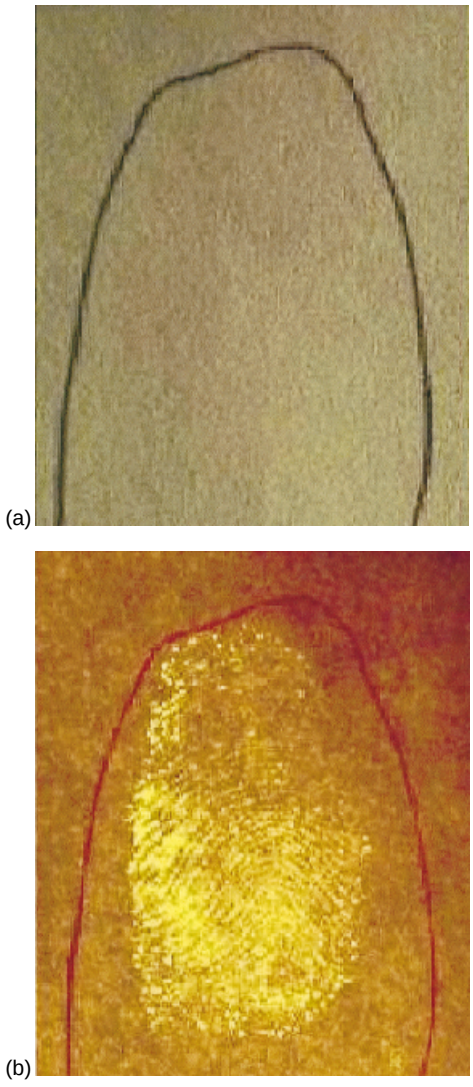

Figure 5 Fingerprint on paper developed by ninhydrin only (a) and by subsequent application of zinc chloride followed by laser examination (b).

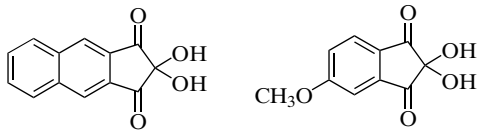

(4) **(5)**

6.2 1,8-Diazafluoren-9-one

The compound 1,8-diazafluoren-9-one **(6)**, referred to as DFO,[19] is employed much like ninhydrin. However,

its reaction with fingerprint amino acid requires elevated temperature (100 °C). The product responds to green excitation and yields an orange fluorescence. In the author's experience, it is (at room temperature) intermediate in sensitivity between ninhydrin/zinc chloride and the two ninhydrin analogs/zinc chloride mentioned above. DFO has been in routine use as a fluorogenic reagent since the early 1990s in a number of law enforcement agencies.

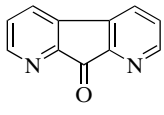

(6)

6.3 Miscellaneous Treatments

Dimethylaminocinnamaldehyde or *para*-dimethylaminocinnamaldehyde (DMAC) or (PDMAC) was investigated in the late 1970s for fluorescence detection of fingerprints[9] because it was thought to react with urea, a major component of eccrine secretions. The reagent is well known to forensic biologists for use in urine stain detection. While its fluorescence use for fingerprint work was successful, it was supplanted by ninhydrin/zinc chloride and later DFO. However, it may be making a comeback because it offers the prospect of dry fingerprint development by vapor or contact procedures,[20,21] and thus has prospects for processing fingerprints on items, such as cloth for instance, that do not lend themselves to the application of solutions. Several heme-catalyzed procedures, such as merbromin plus hydrogen peroxide, exist for fluorescence detection of blood prints. Fluorescein, and even ninhydrin/zinc chloride and DFO can be effective in blood print detection as well. No truly effective procedures for fingerprint detection on skin exist at present.

7 FINGERPRINT TREATMENTS OF FUTURE POTENTIAL

We concern ourselves here with approaches to latent fingerprint detection that have the potential of becoming universal strategies, that is applicable to fingerprints on all kinds of evidence, rather than with incremental modifications of fingerprint procedures such as those discussed above.

7.1 Lanthanide-based Fingerprint Detection[2]

A number of the lanthanide ions (which typically exist in the +3 valence state in compounds), notoriously europium and terbium, emit with high quantum yields and also with long lifetimes, in the millisecond domain. Unfortunately, these ions absorb poorly, such that emission intensities tend to be rather low. In the biochemical arena and also in the fingerprint field, the remedy for this problem has been to chelate the lanthanide ion with a ligand that absorbs well and that transfers its excitation energy to the lanthanide ion, basically by the Forster energy transfer process.[22] The basic scheme is depicted in Figure 6 for the case of europium.

The sensitizing ligand may result from a reaction with fingerprint residue, in which case for example, the counterpart to ninhydrin/zinc chloride has been obtained. The complex may be prepared in solution to lend itself to application in the dye staining mode, akin to rhodamine 6G. The complex may be blended with dusting powder. All three modes have been demonstrated in fingerprint work, with a variety of sensitizing ligands, and with amino acid and lipid sensitive chemistries. However, pervasive problems persist in connection with fingerprints that are not fresh. The most salient virtue of the lanthanide strategy is its utility in concert with time-resolved imaging to suppress the obnoxious background fluorescences which all too often hamper fingerprint detection. Europium complexes, where the emission is

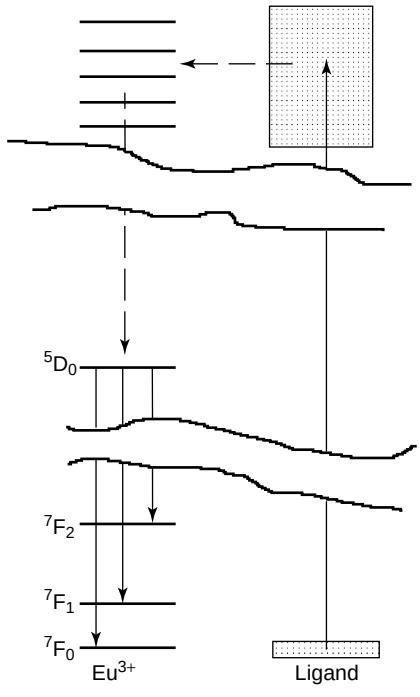

Figure 6 Energy transfer scheme of photoluminescence in europium complexes. Dashed arrows denote radiationless transitions and solid arrows denote photon absorption or emission. The Eu states are labeled with $^{2S+1}L_J$ quantum number designations.

in the red and typically is excited by UV light absorbed by the sensitizing ligand, often lend themselves nicely to background suppression by simple optical filtering because of the large absorption–emission wavelength discrepancy. The current most effective application of lanthanides to fingerprint detection involves the staining approach using europium with thenoyltrifluoroacetone and/or *ortho*-phenanthroline as sensitizing ligands. Preparation of lanthanide complexes may be as simple as mixing a europium salt in solution, such as $EuCl_3 \cdot 6H_2O$, with an excess of ligand, although a stoichiometric proportion of the bidentate ligands is often chosen, keeping in mind that the full coordination of the Eu^{3+} ion is nine-fold.

7.2 Photoluminescent Semiconductor Nanoparticles

The application of photoluminescent semiconductor nanoparticles (also referred to as nanocrystals, quantum dots, or nanocomposites) to latent fingerprint detection is likely to be the next milestone in the evolution of the field, for a number of reasons:[2,23]

1. Photoluminescent semiconductor nanocrystals, such as CdS and CdSe, emit with efficiencies comparable to those of laser dyes.
2. The luminescence lifetimes are on the order of 10^{-7} s. This is a range very well suited to phase-resolved imaging for background fluorescence suppression.
3. The excitation and emission wavelengths can be tailored by adjusting the size of the nanocrystal.
4. CdSe nanocrystals can be encapsulated with ZnS or silica,[24] and can be made soluble in both polar and nonpolar media. The encapsulation may permit the attachment of ligands designed to bind to constituents of fingerprints.[25]
5. The nanocrystal surfaces may be functionalized with ligands designed for selective binding to fingerprints. Nanocrystals may be prepared within voids of dendrimers, which, in turn, may be functionalized for selective attachment to fingerprints.[26]

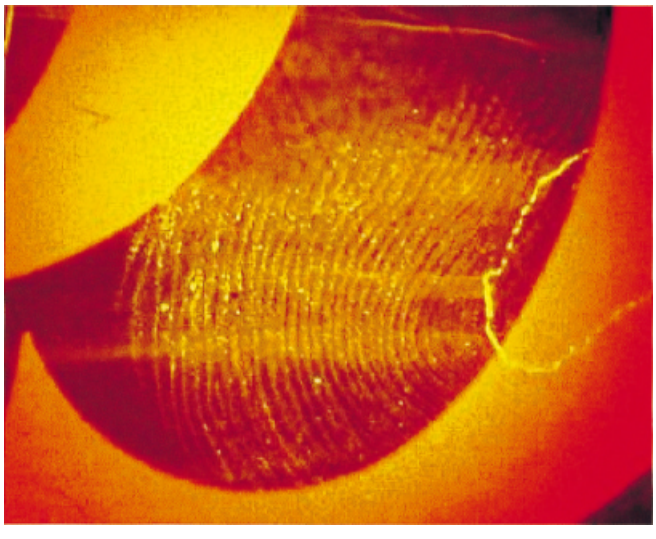

Figure 7 Fingerprint on soft drink can developed by cyanoacrylate fuming followed by staining with CdS nanocrystal solution.

Indeed, recent preliminary work[27,28] suggests that selective fingerprint tagging with photoluminescent semiconductor nanocrystals and dendrimers is feasible. Figure 7 shows the photoluminescence detection (under blue-green Ar-laser excitation) of a fingerprint fumed with cyanoacrylate ester and then stained (in heptane solvent) with CdS nanocrystal functionalized with dioctyl sulfosuccinate.

8 TIME-RESOLVED IMAGING[2]

The suppression of the background fluorescence that often inhibits fingerprint detection typically uses a CCD camera equipped with a microchannel plate image intensifier which can be gated on-off or modulated sinusoidally. In the former instance, the luminescence excitation source is chopped on-off and the camera is synchronized to be on in the light-off portion of the chopping cycle, with a delay past laser cut-off such that the background has already decayed by the time the camera turns on. The camera turns off before the next laser-on cycle portion. This is the imaging counterpart to the lifetime measuring scheme of Figure 2, but with a gate that is not scanned and that has a width that spans most of the chopper dark period. When luminescence lifetimes shorter than about 10^{-6} s are involved, the light chopping becomes expensive and difficult because full extinction in the light off portion of the cycle becomes difficult at the requisite high chopping frequencies. For that domain, phase-resolved imaging becomes of interest because it is not necessary for the light modulation depth to be perfect when phase shifts are probed. Now the imaging involves a phase delay between the sinusoidal light modulation and the sinusoidal image intensifier gain modulation. This is the imaging counterpart to the scheme of Figure 4. There are variants in which the light modulation and camera gain modulation are at slightly different frequencies such that the image information is extracted at the low difference-frequency in a manner reminiscent of the frequency mixing used in radios, for instance.

9 FLUORESCENCE IN TRACE EVIDENCE ANALYSIS

Curiously, fluorescence spectroscopy as such is not widely exploited in criminalistics. However, fluorescence visualization is gaining frequent use in various areas, such as document examination, where infrared fluorescence has been utilized for some time already, fiber analysis, and trace evidence more generally. Typically, the article under scrutiny is simply illuminated and visually inspected for fluorescence. Photography and visual comparisons may be made. We do not consider here highly specialized areas such X-ray fluorescence or chemiluminescence.

9.1 Fiber Analysis

The scrutiny of fibers encountered at the crime scene proceeds primarily by comparison optical microscopy, to examine for color, shape and texture. Polarization microscopy may also be done to determine index of refraction and birefringence. Additional modes of examination include visible and infrared absorption microspectrometry and chromatography of fiber dye content.[29–31] Currently, the utilization of photoluminescence is confined to finding of fibers at the scene via their fluorescence. Otherwise elusive fibers can be found in this way. In principle, excitation and emission spectroscopy can be done on single fibers without much difficulty. However, this is presently not a tool used routinely in fiber analysis. Because fibers come in a wide range of colors, a correspondingly wide range of excitation wavelengths is called for to produce fluorescence adequately. White light source/band-pass filter or monochromator combinations are thus more commonly employed than lasers which operate at a single wavelength or a set of discrete wavelengths. An Ar-laser/dye laser combination, with the Ar-laser operable in the near-UV and blue-green, is very effective, however. As a rough rule of thumb, yellow compounds call for blue excitation and display green to yellow fluorescence, red or orange materials call for green excitation and show yellow, orange or red fluorescence, and blue or green substances require orange to red excitation to display red or near-infrared emission. Colorless or white materials call for UV excitation and typically emit in the near-UV or blue.

9.2 Document Examination

In the area of document alteration, where the interest is to distinguish between writings made with different inks that are of the same color and thus difficult to distinguish by ordinary inspection, plugs of writing are sometimes removed from the document and subjected to thin-layer chromatography (TLC) or the writing is subjected to infrared luminescence examination under visible excitation, most often in the green, especially in examination of black inks. We inquire into the reason for this excitation/emission situation.[2] Generally, absorption and emission do not occur at the same color, even though the same molecular electronic states (ground, first excited singlet) are involved in both phenomena. Instead, the emission occurs at a longer wavelength than the absorption. This is referred to as Stokes shift, and arises from a physical distortion of the molecule after absorption and is related also to vibrational relaxation/wave function overlap considerations. The general scheme is shown in Figure 8(a). Instead of emitting light, a (donor) molecule may transfer its excitation energy to a nearby (acceptor) molecule, as depicted in Figure 8(b). This process, referred to as Forster energy transfer, is not the same as emission of a photon by the donor molecule followed by subsequent absorption of that photon by the acceptor molecule, a process sometimes referred to as self-absorption. Both processes are active in particles of colloidal dimensions, comprised of many molecules in close vicinity to each other, such as ink particles. The energy transfer or self-absorption processes may involve identical molecules or different species. The latter is of most interest in inks made up of dyes of several colors.

The excitation may be most effective at a wavelength where a large number of the ink molecules absorb, i.e. in the green on average for black ink (ideally composed of yellow, magenta and cyan molecules). The emission, on the other hand, comes only from the molecules with the lowest energy gap between the ground and first excited singlet states. The emission is thus typically in the near-infrared. Because of self-absorption/energy transfer, most of the molecules in the ink particle do not emit. Since the emission process generally has quantum efficiency less than one, the net luminescence is also far weaker than it would be if the molecules in the ink particle were far apart from each other. The separation of molecules is in document examination achieved by TLC. Once the components of an ink are thus separated,

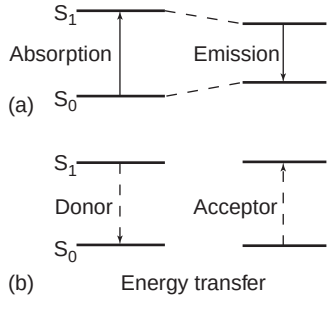

Figure 8 General scheme of Stokes shift (a) and Forster energy transfer (b). S_0 = ground state, S_1 = first excited singlet state.

visibly fluorescent components that otherwise would lose their energy via self-absorption/energy transfer may be detected on the TLC plate. TLC sensitivity can be substantially enhanced in this way. If there is not enough altered writing to permit TLC, in situ fluorescence examination can be much improved by painting the writing with a clear and rapidly drying lacquer that serves to separate ink components slightly to suppress the quenching via self-absorption/energy transfer.[32] This mode of ink examination is currently not routinely performed, however. In situ excitation and luminescence spectroscopies are not generally done either, even though they are not difficult to do in principle. Direct visual examination (with visible and infrared photography) is the most widely utilized fluorescence application in document work. It is often effective in recovering erased or obliterated writing. In the latter case, examination from the back side of the page may be effective.

9.3 Miscellaneous Trace Evidence Detection

Numerous cases of contamination of probative value on articles of evidence, detected by fluorescence, have been reported,[33] including transferred printing, glass fragments, paint, and so on. Often, the detection is serendipitous, but fluorescence crime scene examination for general purposes is becoming more common. For example, deliberate fluorescence detection is at times attempted to find biological fluids, semen stains in particular. The visualization utilizes blue-green or UV excitation. Once such stains are found, they are processed by the customary biological techniques of forensic science. The application of fluorescence to DNA profiling is currently routine only in quality assurance of digestion and electrophoresis, by staining with fluorescent dye, most notably ethidium bromide. The profiling as such in the RFLP (restricted fragment length polymorphism) approach utilizes radioactive tagging, and colorimetric spot development is typical in PCR (polymerase chain reaction) processing. However, the fluorescence approaches common in (nonforensic) DNA sequencing[34] are rapidly gaining use in criminalistics. Glass fragments and paint chips are sometimes examined by fluorescence comparisons but the fluorescence detection is usually followed by more standard techniques of analysis. Laser examination has been performed on cyanide-laced Tylenol capsules. PDMAC is occasionally applied to some drug examinations. However, the analysis of poisons or drugs does not usually employ photoluminescence techniques, generally relying instead on absorption spectroscopy, chromatography and mass spectrometry. For drugs, color and microcrystalline tests are also used, as preliminary examinations. In explosives and firearms examination, techniques such as electron microscopy (with X-ray elemental analysis) and color tests for metal or explosive residue on the skin of a suspect are typical, rather than fluorescence methodology.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
CCD	Charge-coupled Device
CW	Continuous Wave
DFO	1,8-Diazafluoren-9-one
DMAC	Dimethylaminocinnamaldehyde
FTIR	Fourier Transform Infrared
GC/MS	Gas Chromatography/Mass Spectrometry
HFE	Hydrofluoroether
HOMO	Highest Occupied Molecular Orbital
LUMO	Lowest Unoccupied Molecular Orbital
PCR	Polymerase Chain Reaction
PDMAC	*para*-Dimethylaminocinnamaldehyde
PMT	Photomultiplier Tube
RFLP	Restricted Fragment Length Polymorphism
TLC	Thin-layer Chromatography
UV	Ultraviolet
YAG	Yttrium Aluminum Garnet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Laser- and Optical-based Techniques for the Detection of Explosives

Environment: Water and Waste **(Volume 4)**
Luminescence in Environmental Analysis

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Microspectrophotometry in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
Fluorescence In Situ Hybridization

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Surfaces **(Volume 10)**
Photoluminescence in Analysis of Surfaces and Interfaces

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Absorption and Luminescence Probes • Circular Dichroism and Linear Dichroism • Detectors, Absorption and Luminescence • Fluorescence Imaging Microscopy • Fluorescence in Organized Assemblies • Fluorescence Lifetime Measurements, Applications of • Indirect Detection Methods in Capillary Electrophoresis • Near-infrared Absorption/Luminescence Measurements • Phosphorescence Measurements, Applications of • Surface Measurements using Absorption/Luminescence • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

Infrared Spectroscopy **(Volume 12)**
Emission Spectroscopy, Infrared

REFERENCES

1. E.R. Menzel, *Laser Spectroscopy, Techniques and Applications*, Marcel Dekker, New York, 1995.
2. E.R. Menzel, *Fingerprint Detection with Lasers*, 2nd edition, Marcel Dekker, New York, 1999.
3. H.C. Lee, R.E. Gaensslen (eds.), *Advances in Fingerprint Technology*, Elsevier, New York, 1991.
4. R.D. Olsen, Sr, 'The Chemical Composition of Palmar Sweat', *Fingerprint and Identification Mag.*, **53**(April), 3–23 (1972).
5. A.M. Knowles, 'Aspects of Physicochemical Methods for the Detection of Latent Fingerprints', *J. Phys. E: Sci. Instrum.*, **11**, 713–721 (1978).
6. M.V. Buchanan, K. Asano, A. Bohanon, 'Chemical Characterization of Fingerprints from Adults and Children', in *Forensic Analysis and Crime Scene Investigation*, eds. J. Hicks, P.R. DeForest, V.M. Baylor, Proc. SPIE, **2941**, 89–95, 1997.
7. B.E. Dalrymple, J.M. Duff, E.R. Menzel, 'Inherent Fingerprint Luminescence–Detection by Laser', *J. Forensic Sci.*, **22**, 106–115 (1977).
8. N.S. Stames, 'Interesting Case', *Identification News*, **34**(10), 13 (1984).
9. E.R. Menzel, *Fingerprint Detection with Lasers*, Marcel Dekker, New York, 1980.
10. E.R. Menzel, J.A. Burt, T.W. Sinor, W.B. Tubach-Ley, K.J. Jordan, 'Laser Detection of Latent Fingerprints: Treatment with Glue Containing Cyanoacrylate Ester', *J. Forensic Sci.*, **28**, 307–317 (1983).
11. F.G. Kendall, 'Superglue Fuming for the Development of Latent Fingerprints', *Identification News*, **32**(5), 13–14 (1982).
12. 'LA Times Reports on Laser's Role in Arrest of Murder Suspect', *Laser Topics*, September, 20 (1985).
13. S. Oden, B. Von Hofsten, 'Detection of Fingerprints by the Ninhydrin Reaction', *Nature*, **173**, 449–450 (1954).
14. S. Ruhemann, 'Triketohydrindene Hydrate', *J. Chem. Soc.*, **97**, 2025–2031 (1910).
15. D.W. Herod, E.R. Menzel, 'Laser Detection of Latent Fingerprints: Ninhydrin followed by Zinc Chloride', *J. Forensic Sci.*, **27**, 513–518 (1982).
16. E.R. Menzel, 'Laser Detection of Latent Fingerprints', *Law Enforcement Comm.*, March/April, 4–5 (1983).
17. E.R. Menzel, J. Almog, 'Latent Fingerprint Development by Frequency-doubled Neodymium : Yttrium Aluminum Garnet (Nd : YAG) Laser: Benzo(f)ninhydrin', *J. Forensic Sci.*, **30**, 371–382 (1985).
18. J. Almog, A. Hirshfeld, '5-Methoxyninhydrin: A Reagent for the Chemical Development of Latent Fingerprints that is Compatible with the Copper-vapor Laser', *J. Forensic Sci.*, **33**, 1027–1030 (1988).
19. C.A. Pounds, 'Developments in Fingerprint Visualization', in *Forensic Science Progress*, eds. A. Maehly, R.L. Williams, Springer, Berlin, 91–119, Vol. 3, 1988.
20. J.S. Brennan, 'The Development of Fingerprints by Fuming with Dimethylaminocinnamaldehyde (DMAC)', *Proceedings International Symposium on Fingerprint Detection and Identification*, Ne'urim, Israel, 1995, eds. J. Almog, E. Springer, Israel Natl. Police, Jerusalem, 85–90, 1996.
21. S. Ramotowski, 'Fluorescence Visualization of Latent Fingerprints on Paper Using *p*-Dimethylaminocinnamaldehyde (PDMAC)', *Proceedings International Symposium on Fingerprint Detection and Identification*, Ne'urim, Israel, 1995, eds. J. Almog, E. Springer, Israel Natl. Police, Jerusalem, 91–94, 1996.
22. T. Forster, 'Zwischenmolekulare Energiewanderung und Fluoreszenz', *Ann. Phys.*, **2**, 55–75 (1948).
23. E.R. Menzel, 'Fluorescent Semiconductor Nanocrystals for Latent Fingerprint Development', Patent Applied for.
24. M. Bruchez Jr, M. Moronne, P. Gin, S. Weiss, A.P. Alivisatos, 'Semiconductor Nanocrystals as Fluorescent Biological Labels', *Science*, **281**, 2013–2016 (1998).
25. W.C.W. Chan, S. Nie, 'Quantum Dot Bioconjugates for Ultrasensitive Nonisotopic Detection', *Science*, **281**, 2016–2018 (1998).
26. K. Sooklal, H.L. Hanus, H.J. Ploehn, C.J. Murphy, 'A Blue-emitting CdS/Dendrimer Nanocomposite', *Adv. Mater.*, **10**, 1083–1087 (1998).
27. E.R. Menzel, S.M. Savoy, S.J. Ulvick, K.H. Cheng, R.H. Murdock, M.R. Sudduth, 'Photoluminescent Semiconductor Nanocrystals for Fingerprint Detection', *J. Forensic Sci.*, in press.

28. E.R. Menzel, M. Takatsu, R.H. Murdock, K.K. Bouldin, K.H. Cheng, 'Photoluminescent CdS/Dendrimer Nanocomposites for Fingerprint Detection', *J. Forensic Sci.*, in press.
29. S. Saferstein (ed.), *Forensic Science Handbook*, Prentice-Hall, Englewood Cliffs, Vol. 1, 1981.
30. S. Saferstein (ed.), *Forensic Science Handbook*, Prentice-Hall, Englewood Cliffs, Vol. 2, 1988.
31. S. Saferstein (ed.), *Forensic Science Handbook*, Prentice-Hall, Englewood Cliffs, Vol. 3, 1993.
32. T.W. Sinor, J.P. Wilde, K.E. Everse, E.R. Menzel, 'Lasers and Optical Spectroscopy in Questioned Document Examination', *J. Forensic Sci.*, **31**, 825–839 (1986).
33. K.E. Creer, 'The Detection and Photography of Fluorescent Marks', in *Fluorescence Detection*, ed. E.R. Menzel, Proc. SPIE, **743**, 175–179, 1987.
34. J.A. Luckey, H. Drossman, A.J. Kostichka, D.A. Mead, J. D'Cunha, T.B. Norris, L.M. Smith, 'High Speed DNA Sequencing by Capillary Electrophoresis', *Nucleic Acids Res.*, **18**, 4417–4421 (1990).

Immunoassays in Forensic Toxicology

David E. Moody
University of Utah, Salt Lake City, USA

1 Introduction **2**
1.1 Forensic Toxicology 2
1.2 Principles of Immunoassay 3
1.3 Introduction of Immunoassays into Forensic Toxicology 3
2 Immunoassays Used in Forensic Toxicology **3**
2.1 Homogeneous Immunoassays 4
2.2 Heterogeneous Immunoassays 8
2.3 On-site Drug Testing Immunoassays 10
3 Sample Preparation and Matrix Factors **12**
3.1 Introduction 12
3.2 Urine 12
3.3 Blood 12
3.4 Hair 14
3.5 Other Matrices 16
4 Forensic Applications **17**
4.1 Detection Limits and Cutoffs 17
4.2 Cross-reactivities 19
4.3 Interfering Substances 22
4.4 Interpretation of Results (Need for Confirmation and Quantitative Versus Qualitative) 22
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **25**
References **25**

Forensic toxicology encompasses the determination of the presence and concentration of drugs, other xenobiotics and their metabolites in physiological fluids and organs and the interpretation of these findings as they may impact on legal issues. These include medical examiner investigations, driving under the influence and other transportation accident investigations, workplace pre-employment, random and for-cause drug testing and judicial monitoring of arrestees and parolees. Similar techniques are employed in emergency room clinical toxicology and to monitor the efficacy of substance abuse treatment. The introduction of immunoassays into forensic toxicology in the early 1970s has had a major impact on the speed and efficiency that samples can be screened for the presence of certain drug classes. For the most part, forensic toxicologists use commercial immunoassays directed primarily towards abused drugs. Commercial immunoassays developed for therapeutic monitoring of other drugs, veterinary drugs and pesticides, as well as immunoassays developed in research laboratories for specialized studies, may find a role in the forensic toxicology laboratory for specialized cases.

Most immunoassays used for forensic toxicology are competitive. An antigen structurally similar to the target compound is conjugated to a signalling molecule and competes with target drug in a sample for antibody binding. Immunoassays are also classified as homogeneous and heterogeneous. Homogeneous assays do not separate the original sample from the final detection sample. They must use a signal that changes when antibody is bound. Homogeneous immunoassays include enzyme immunoassay (EIA) (enzyme activity decreases when bound), fluorescent polarization immunoassay (FPIA) (emission in a polar field increases when bound) and kinetic interaction of microparticles in solution (KIMS) immunoassay (lattice formation inhibited when bound). Heterogenous immunoassays include radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA) where unbound radiolabeled antigen and enzyme conjugated antigen, respectively, are removed from the sample before measurement. In general, the homogeneous immunoassays are more ameniable to full automation, and thereby quicker throughput. The heterogeneous immunoassays are less susceptible to matrix interference, and thereby more versatile with nonurine matrices.

While most commercial immunoassays have been developed for a urine matrix, they have been applied by forensic toxicologists to other matrices, including blood, hair, saliva, sweat, tissue homogenates, blood stains and most other physiological samples that may be of value in the investigation. The use of nonurine matrices must contend with two factors. With the exception of parenchymal tissues, the concentration of the target compound is often lower and the sensitivity of the immunoassay may be limiting. In addition, the nonurine matrix usually is much more complex in its composition. Sample pretreatments that range from simple deproteinations to multistep extractions to remove matrix components and/or concentrate the sample are often required. The heterogenous RIAs and ELISAs usually require less rigorous, if any, pretreatments.

Interpretation of immunoassay results must take into consideration the limits of detection of the assay, the cross-reactivity of the antibody(ies) and the potential for interference. Appropriate controls should be included to demonstrate adequate signal separation from blanks (drug-free sample in the same matrix). The concentration of the low control is often determined by the sensitivity

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of the assay, or in workplace testing programs by administratively determined cutoffs that draw the line between a negative and presumptive positive. The ability of the antibody to detect compounds other than the target compound (its cross-reactivity) can be a useful characteristic or, when the drugs that can be confirmed are limited, a nuisance. The amphetamines, barbiturates and benzodiazepines, in particular, all have a number of licit (and for amphetamines illicit) analogs that could be present in a sample. Cross-reactivities are antibody-source (i.e. manufacturer) dependent. Further testing to determine the immunoreactive compound often requires rigorous methodologies. In some instances, the potency of the drug and its poor cross-reactivity make detection by immunoassay difficult (e.g. nitrosobenzodiazepines). Interference in immunoassays may arise from compounds that appear in the matrix during disease states, or those that are intentionally added to a sample in the hope of negating a positive test. These act through many different mechanisms and may decrease or increase the immunoassay test result.

Immunoassays have added an extremely useful tool to the forensic toxicology investigation. They can be used to screen rapidly a large number of samples for the potential presence of a drug group. With rare exceptions (emergencies, limited sample volume), their use without a confirmation assay (e.g. gas or liquid chromatography/mass spectrometry (LC/MS)) is unwarranted, as it leads to a risk of improper test result interpretation.

1 INTRODUCTION

1.1 Forensic Toxicology

Forensic toxicology encompasses the determination of the presence and concentration of drugs, other xenobiotics and their metabolites in physiological fluids and organs and the interpretation of these findings as they may impact on legal issues. Forensic toxicology has grown in scope from investigations conducted by the medical examiner's office to include driving under the influence and other transportation accident investigations, workplace pre-employment, random and for-cause drug testing and judicial monitoring of arrestees and parolees. Similar techniques are employed in emergency room clinical toxicology and to monitor the efficacy of substance abuse treatment. The capabilities of the forensic toxicologist to detect and quantify drugs, other xenobiotics and their metabolites have grown along with the field of analytical chemistry. The introduction of the RIA by Yalow and Berson in 1959[1,2] was followed in the early 1970s by the introduction of immunoassays into forensic toxicology. This form of analysis has had a major impact on the speed and efficiency that samples can be screened for the presence of certain drug classes.

For the most part, forensic toxicologists use commercial immunoassays directed primarily towards abused drugs (Table 1). These immunoassays will be the primary subject of this article. Commercial immunoassays have also been developed for therapeutic monitoring of other drugs. Some of these assays can also be of great use to the forensic toxicologist investigating causes of death or toxic injury (Table 2). The routine use of these will vary from laboratory to laboratory, as some laboratories will depend upon other screening methods such as thin-layer chromatography (TLC) or gas chromatography/mass spectrometry (GC/MS) screens to extend their qualitative detection. Some therapeutic drug assays may not be used routinely, but may be called upon if the medical history dictates (e.g. epileptic deaths and anticonvulsants). Other immunoassays have been developed for the detection of drugs specific to veterinary toxicology and the detection of pesticides for environmental monitoring. In addition, a number of immunoassays with limited distribution have been developed in research laboratories for specialized studies. All of these assays may find a role in the forensic toxicology laboratory for specialized cases, but they will not be discussed further in this article.

This article has been prepared so that a reader unfamiliar with immunoassays and/or their application to forensic toxicology can come away with a basic understanding

Table 1 Drug groups targeted by immunoassays commonly used in forensic toxicology

No.	[a]Drug group	Most common antigen
1	Amphetamines	Amphetamine or methamphetamine
2	Cannabinoids	THCA
3	Cocaine	Benzoylecgonine
4	Opiates	Morphine
5	PCP	Phencyclidine
6	Barbiturates	Secobarbital
7	Benzodiazepines	Oxazepam or nordiazepam
8	LSD	Lysergic acid diethylamide
9	Methadone	Methadone

[a] The number assigned to the drug group will be used in subsequent tables to identify which drug groups were studied. PCP, phencyclidine; LSD, lysergic acid diethylamide; THCA, 11-nor-Δ^9-tetrahydrocannabinol-9-carboxylic acid.

Table 2 Therapeutic drug immunoassays that are often employed in forensic toxicology investigations

Propoxyphene	Acetaminophen
Salicylates	Cardiac glycosides (digoxin)
Phenytoin	Valproate
Carbamazepine	

of how immunoassays work, their limitations and how they have been applied for studies in forensic toxicology. Much of the tabular material is, however, practically complete, with studies listed in chronological order. It is hoped that the inclusion of this material will make this article also of value to the experienced forensic toxicologist.

1.2 Principles of Immunoassay

Immunoassays require an antibody that is selective for the drug or drug group of interest (antigen), a label that will be part of the antibody antigen complex that emits a detectable signal and a means of detecting the signal. Immunoassays may be competitive (Figure 1), where the analyte competes with labeled analyte for the antibody, or noncompetitive, where the antibody, or a secondary antibody is labeled. The commercial immunoassays in use to detect drugs of interest to forensic toxicologists are competitive immunoassays.

Immunoassays can also be distinguished as heterogeneous versus homogeneous. Heterogeneous assays require the separation of bound from unbound antigen, while homogeneous assays do not require separation. In general, the separation step of a heterogeneous assay will remove the initial specimen matrix, and this in turn will reduce the number of potential interfering substances. Therefore, heterogeneous assays are considered more specific and sensitive. On the other hand, the additional steps required to perform a heterogeneous assay make total automation of the assay more difficult. Most homogeneous immunoassays are easily automated. The heterogeneous immunoassays in common use in the forensic toxicology laboratory are the RIA and the ELISA and the homogeneous immunoassays are the EIA, FPIA and KIMS immunoassay.

1.3 Introduction of Immunoassays into Forensic Toxicology

Urine drug testing had long been a tradition in the forensic toxicology laboratory for medical examiner studies and the extension of driving under the influence investigations since the increased drug use in the 1960s. Many techniques were employed. The introduction of TLC in the 1960s had a major impact. Two driving factors led to the introduction of immunoassays into the forensic (and clinical) laboratory. The first was the need for methadone treatment clinics to monitor drug use of the patients. The second, which was to have an overwhelming impact, was the use of drugs in the Vietnam war. Both dictated the need for rapid analysis of urine samples for morphine.

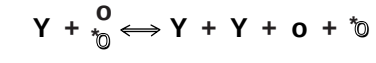

Figure 1 Schematic of a competitive immunoassay.

In 1970, Spector, at the Roche Institute of Molecular Biology, developed an RIA for morphine.[3,4] A refinement of this RIA was then developed at Hoffman-La Roche that could be routinely used to measure morphine in urine and serum.[5] The use of radioactive material in this assay, however, limited the sites where it could be performed. Leute et al. at Syva (formerly Synavar) developed an immunoassay that utilized antibodies directed towards morphine and electron spin resonance to detect morphine conjugated to a spin label, nitroxide radical.[6,7] The spin of the unpaired electron in the nitroxide radical can be detected by electron spin resonance. The signal produced by a rapidly rotating (unbound) radical is sharp and distinct; it diminishes as the conjugated antigen is bound by antibody. This was named the free radical assay technique (FRAT). Although initially developed for use by methadone clinics, this technique was seized upon by the US Army, and beginning in 1971 was used to screen military personnel in Vietnam. FRAT instrumentation was expensive. During the same time period, Adler et al. developed a hemagglutination-inhibition (HI) assay for morphine that was both rapid and simple.[8,9] Syva then developed an EIA that had antigen conjugated to the enzyme lysozyme for detection of morphine in urine [enzyme-multiplied immunoassay technique (EMIT)].[10,11] This assay could be performed on a number of commonly available laboratory instruments and thereby became readily usable by numerous forensic and clinical laboratories.

Although both the FRAT and HI methods were used and evaluated for a time,[12–14] they did not gain sufficient favor to have a long-term impact. Both the RIA and EIA methods, however, became popular and were expanded to encompass the detection of a number of abused and therapeutic drugs.

2 IMMUNOASSAYS USED IN FORENSIC TOXICOLOGY

The commercial production of immunoassays that are used in forensic toxicology laboratories is a dynamic market, and some vendors advertise their products to a greater extent than others. The following sections are based on vendors that have come to the attention of the author (Table 3). The omission of any vendors does not imply any negative attributes towards their products, it simply demonstrates the author's ignorance of their existence.

Table 3 Instrument-dependent immunoassays marketed for use in forensic toxicology

Immunoassay	Company
Homogeneous immunoassays:	
EIA	DRI
	Microgenics CEDIA DAU
	Syva EMIT
FPIA	Abbott Diagnostics, TDx
KIMS	Roche Diagnostics, OnLine
Heterogeneous immunoassays:	
RIA	DPC
	Immunalysis Corp.
ELISA	Diagnostix Ltd.
	STC Diagnostics

DRI, Diagnostic Reagents Inc.; DPC, Diagnostic Products Corp.; CEDIA, cloned enzyme donor immunoassay.

2.1 Homogeneous Immunoassays

2.1.1 Enzyme Immunoassay

EIAs are homogeneous immunoassays that utilize an antigen conjugated to an enzyme (Figure 2a) or enzyme fragment (Figure 2b). When the conjugated antigen is bound to antibody, the enzyme is less active or inactive. As the amount of target drug in a sample increases, it displaces the conjugated antigen, thereby activating the enzyme. The enzyme will then form product(s) from the substrate(s). Product formation is monitored with the detector at an absorbance selective for the product. The rate of product formation increases with the amount of target drug in the sample (Figure 2c).

Syva's EMIT assays were the initial immunoassays available for EIA detection of drugs. The EMIT initially used lysozyme as the enzyme conjugate in urine assays.[10,11] For serum therapeutic drug monitoring assays this was not practical because of endogenous lysozyme activity; glucose-6-phosphate dehydrogenase was used for these assays. Eventually, glucose-6-phosphate dehydrogenase was incorporated into the urine assays as lysozyme substrate was found to degrade with aging, some individuals (~2–4% of the population) were found to excrete endogenous lysozyme into their urine and a number of inhibitors of lysozyme were being encountered in tested urines. Malate dehydrogenase was also used for a time in the urine assays,[15] but it also suffered from excretion of endogenous enzyme. The glucose-6-phosphate dehydrogenase reaction requires glucose-6-phosphate and nicotinamide adenine dinucleotide (NAD) as substrates. The formation of nicotinamide adenine dinucleotide (reduced form) (NADH) is monitored spectrophotometrically.

When the patent on EIA for drug detection lapsed, other companies entered this market. DRI, (a subsidiary of Sybron International Corp.) markets a number of

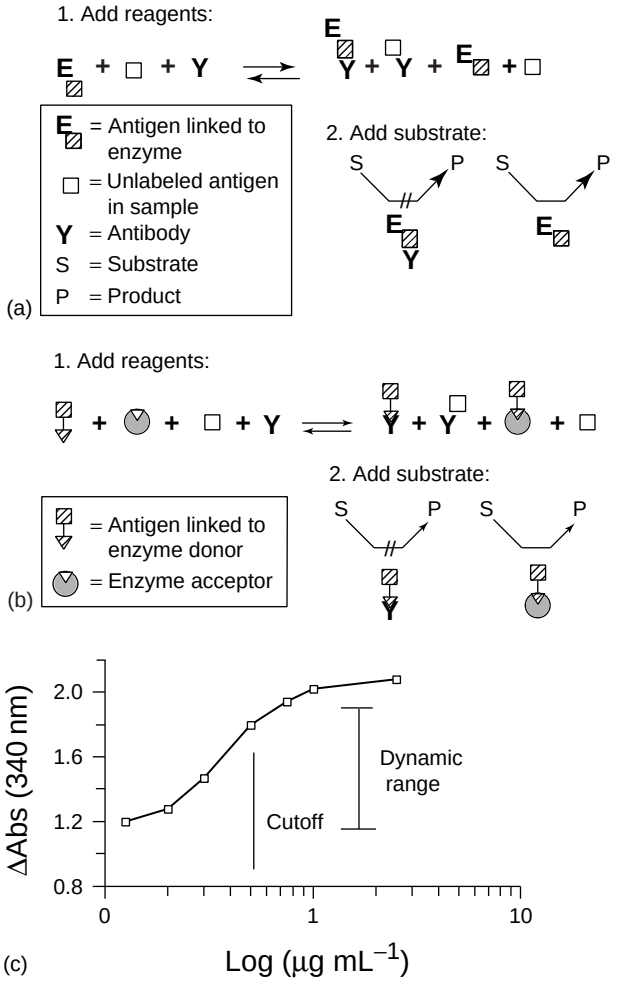

Figure 2 Schematic of an EIA showing (a) an EIA where the antigen is conjugated to an intact enzyme (e.g. EMIT), (b) an EIA where a donor portion of the enzyme is conjugated to the antigen and the acceptor portion of the enzyme is free in solution (e.g. CEDIA) and (c) a representative plot of signal versus drug concentration.

EIAs that also use antigen conjugated to glucose-6-phosphate dehydrogenase. A different approach was taken by Microgenics (now also a subsidiary of Sybron International Corp.). They invented the CEDIA that uses recombinant DNA expressed fragments of β-galactosidase as the enzyme conjugate.[16] Two enzyme fragments exist that are inactive by themselves; the smaller one is conjugated to the drug–antigen. In the unbound state, the fragments will freely associate and form active enzyme that forms product from the substrate, chlorophenol red β-D-galactopyranoside (Figure 2b). The rate of product formation increases with the amount of target drug in the sample.

The EIAs utilized in forensic toxicology laboratories are amenable to many different automated analyzers. The manufacturers should be consulted to see which

instruments have proven useful in their experiments. Use of other instruments, or modifications of the procedures are best accompanied by data that validate the modifications. A number of studies have been reported on the EMIT (Table 4), CEDIA and DRI assays (Table 5).

2.1.2 *Fluorescent Polarization Immunoassay*

The FPIAs for drug analysis are homogeneous immunoassays marketed by Abbott Diagnostics.[77] The antigen is conjugated to fluorescein. The fluorescein is excited with light in a single plane (polarized light). When conjugated antigen is bound to the antibody, the rate of rotation of the complex is relatively slower, and more light is emitted. The unbound conjugated antigen rotates faster and emits less light (Figure 3a). As drug in the sample increases and more binds to the antibody, displacing the conjugated antigen, the amount of light emitted decreases (Figure 3b). Abbott's FPIAs require dedicated instruments such as the TDx or ADx. These will be provided by Abbott as part of a reagent agreement. Representative studies of the TDx and ADx reagents for detection of abused drugs are shown in Table 6.

2.1.3 *Kinetic Interaction of Microparticles in Solution*

Roche Diagnostics markets a number of KIMS immunoassays for the detection of drugs. The antigen is

Table 4 Representative studies on the EMIT assays for drugs of abuse in urine

Drug groups studied[a]	Reagent	Instrument	Methods compared	Refs.
4	–	Gilford 300-N	None	11
1, 3, 4, 6, 9	–	ND[b]	TLC	14
4	–	Beckman 25	RIA	17
4	–	Shimadzu UV-200	GC/MS[b]	18
3	–	ND	RIA, TLC, GC/FID[b]	19
2	–	Stasar III	None	20
4, 6, 7, 9	–	Beckman 25	GC/MS	21
2	d.a.u.	Syva 5000	RIA, GC/MS	22
2	d.a.u.	Emit System	RIA, GC/MS, HPLC[b], GC/ECD	23
2	d.a.u.	Syva 6000	RIA, TLC[b], GC/MS	24
2	d.a.u.	Emit System	RIA, GC/MS	25
2	d.a.u.	ND	GC/MS	26
1–9	d.a.u.	Multistat III	None	27
3	d.a.u.	Stasar III	FPIA, GC/MS	28
2	d.a.u.	Cobas Bio	FPIA, HPLC	29
3	d.a.u.	ND	FPIA, RIA, TLC, GC/MS	30, 31
5, 6	d.a.u.	Multistat III	FPIA, GC/MS	32
2	d.a.u.	*aca* discrete	FPIA, RIA	33
2	d.a.u.	Stasar III	FPIA, TLC	34
5	d.a.u.	ETS®	None	35
2	d.a.u.	EPOS 5060	TLC, GC/MS	36
3	d.a.u.	Cobas-Bio	FPIA	37
1–5	d.a.u.	Monarch 2000	None	38
2	d.a.u.	Technicon RA-500	RIA, GC/MS	39
2	d.a.u.	Monarch 2000	FPIA, RIA, GC/MS	40
7	d.a.u.	Stasar III	FPIA, GC/MS	41, 42
1	d.a.u.	Monarch 2000	FPIA, GC/MS	43
7	d.a.u.	Monarch 2000	FPIA	44
1–7	d.a.u.	ND	Triage, GC/MS	45
5	d.a.u.	Cobas Mira	GC/MS	46
4	d.a.u.	Hitachi 717	FPIA, RIA, GC/MS	47, 48
5	d.a.u.	ETS®	GC/MS	49
2, 5	II/d.a.u.	Hitachi 717/ETS®	GC/MS	50
1	II/d.a.u.	Hitachi 704	GC/MS	51
2–4, 6	II	Hitachi 717	FPIA, KIMS, RIA	52
2–6	II	Hitachi 717	KIMS, RIA, GC/MS	53
1–6	d.a.u.	ETS®/Hitachi 717	Triage, GC/MS	54
2–7	II	Chem 1	EMIT d.a.u.	55
7	d.a.u./II	Cobas Fara II	FPIA, KIMS, RIA, Triage, GC/MS	56
2–6	II/700	Hitachi 717	RIA	57
7	d.a.u.	ND	RRA[b], HPLC	58

Table 4 (*continued*)

Drug groups studied[a]	Reagent	Instrument	Methods compared	Refs.
7	d.a.u.	ETS®	Triage, GC/MS	59
7	II	Hitachi 717	FPIA, GC/MS	60
1–6	II	Hitachi 717	CEDIA, FPIA, KIMS, RIA, GC/MS	61
2	d.a.u.	Cobas Fara II	KIMS	62
2	d.a.u.	ETS®	FPIA, KIMS, GC/MS	63
1	d.a.u.	Hitachi 737	KIMS, GC/MS	64
1–7	II	Hitachi 717	CEDIA, GC/MS	65
7	II/d.a.u.	Hitachi 717/ETS®	CEDIA, FPIA, Triage, GC/MS	66
3, 4	d.a.u.	Solaris	KIMS, GC/MS	67
7	II	ND	CEDIA, GC/MS	68
1–4	d.a.u.	ETS®	RapiTest, GC/MS	69
7	II/d.a.u.	Hitachi 717	FPIA, KIMS, GC/MS	70
2	d.a.u.	*aca* discrete	KIMS, GC/MS	71
2–4	d.a.u.	ND	Frontline, FPIA, GC/MS	72
1–4, 7	d.a.u., ETS®/II,	Cobas Mira/II, Hitachi 711	3 Onsite kits, GC/MS	73
8	II	Hitachi 917	CEDIA, RIA, GC/MS	74

[a] See Table 1 for number assignment to drug groups.
[b] Abbreviations not previously defined: HPLC, high-performance liquid chromatography; GC/ECD, gas chromatography/electron capture detection; GC/FID, gas chromatography/flame ionization detection; ND, not determined; RRA, radioreceptor assay.

Table 5 Representative studies on the CEDIA and DRI assays for drugs of abuse in urine

Drug groups studied[a]	Reagent	Instrument	Methods compared	Ref.
1–6	CEDIA	Hitachi 717	EMIT, FPIA, KIMS, RIA, GC/MS	61
1–7	CEDIA	Hitachi 717	EMIT, GC/MS	65
7	CEDIA	Corning Express 550	EMIT, FPIA, Triage, GC/MS	66
8	CEDIA	Hitachi 911	RIA, GC/MS	75
7	CEDIA	ND	EMIT, GC/MS	68
8	CEDIA	Hitachi 917	EMIT, RIA, GC/MS	74
1–4, 6, 7, 9	DRI	Hitachi 911	EMIT	76

[a] See Table 1 for number assignment to drug groups.

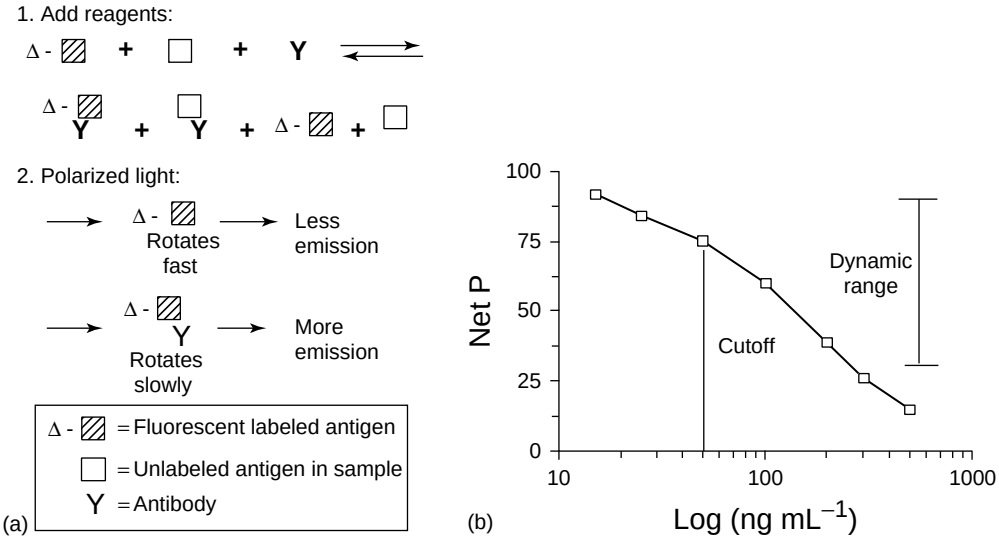

Figure 3 (a) Schematic of an FPIA and (b) a representative plot of signal versus drug concentration.

Table 6 Representative studies on the TDx FPIAs for drugs of abuse in urine

Drug groups studied[a]	Instrument	Methods compared	Refs.
1	TDx	GC/NPD[b]	78
3	TDx	EMIT, GC/MS	28
2	TDx	EMIT, HPLC	29
2	TDx	RIA, GC/MS	79
2	TDx	EMIT, RIA	33
2	TDx	EMIT, TLC	34
2	TDx	GC/MS	80
5, 6	TDx, ADx	EMIT, GC/MS	32
3	TDx	EMIT, RIA, TLC, GC/MS	30, 31
3	TDx	EMIT	37
2	TDx	EMIT, RIA, GC/MS	40
7	TDx, ADx	EMIT, GC/MS	41, 42, 81
1	TDx, ADx	EMIT, GC/MS	43
7	ADx	EMIT	44
4	TDx	EMIT, RIA, GC/MS	47, 48
2–4, 6	TDx	EMIT, KIMS, RIA	52
7	TDx	EMIT, RIA, Triage, GC/MS	56
3	TDx	None	82
7	TDx	EMIT, GC/MS	60
1, 2, 7	ADx	Triage, GC/MS	83
1–3	TDx	CEDIA, EMIT, KIMS, RIA, GC/MS	61
2–4	ADx	EZ-Screen, GC/MS	84
2	ADx	EMIT, KIMS, GC/MS	63
7	TDx	CEDIA, EMIT, Triage, GC/MS	66
7	ADx	EMIT, KIMS, GC/MS	70
2–4	TDx	Frontline, EMIT, GC/MS	72
1–4	ADx	Accusign, GC/MS	85

[a] See Table 1 for number assignment to drug groups.
[b] Abbreviation not previously defined: GC/NPD, gas chromatography/nitrogen–phosphorus detection.

conjugated to microparticles. When antibody binds the conjugated antigen, a lattice of microparticles will form. The lattice decreases as conjugated antigens are unbound owing to competition for antibody binding sites with target drug in the sample (Figure 4a). The greater the lattice formation, the higher is the absorbance, or the absorbance decreases as the target drug in the sample increases (Figure 4b). The KIMS immunoassays utilized in forensic toxicology laboratories are amenable to many different automated analyzers. The manufacturers should

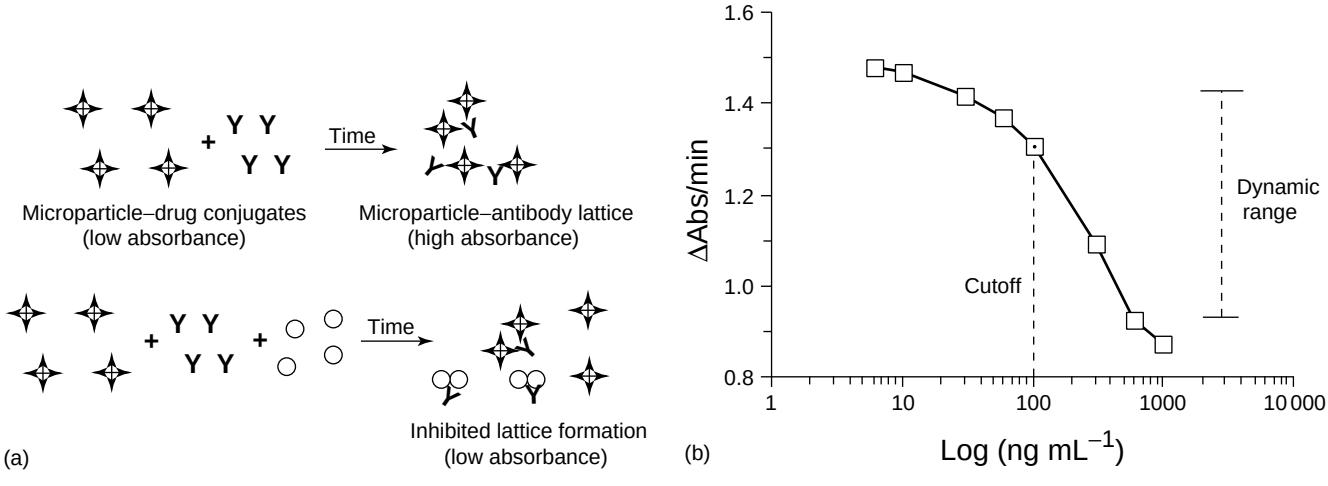

Figure 4 (a) Schematic of a KIMS immunoassay and (b) a representative plot of signal versus drug concentration.

Table 7 Representative studies on the OnLine KIMS immunoassays for drugs of abuse in urine

Drug groups studied[a]	Instrument	Methods compound	Ref.
2–4, 6	Hitachi 717	EMIT, FPIA, RIA	52
2–6	Hitachi 717	EMIT, RIA, GC/MS	53
7	Cobas Fara II	EMIT, FPIA, RIA, Triage, GC/MS	56
1–3	Hitachi 717	CEDIA, EMIT, FPIA, RIA, GC/MS	61
2	Cobas Fara II	EMIT	62
2	Cobas Mira	EMIT, FPIA, GC/MS	63
1	Hitachi 737	EMIT, GC/MS	64
8	Hitachi 717	RIA, GC/MS	86
3, 4	Cobas Mira	EMIT, GC/MS	67
7	Cobas Integra	OnTrak, GC/MS	87
7	Cobas Mira	EMIT, FPIA, GC/MS	70
2	Dimension	EMIT, GC/MS	71
2–4	Hitachi 717	Ontrack, Testcup, GC/MS	88

[a] See Table 1 for number assignment to drug groups.

be consulted to see which instruments have proven useful in their experiments. Use of other instruments or modifications of the procedures are best accompanied by data that validate the modifications. A summary of the use of KIMS assays for detection of drugs of abuse in urine is presented in Table 7.

2.2 Heterogeneous Immunoassays

2.2.1 Radioimmunoassay

RIAs incorporate a radiolabeled antigen that competes for binding to the antibody with unlabeled drug in the sample. Most of the RIAs used in forensic laboratories use ^{125}I-labeled antigens. ^{125}I can be detected with gamma counters or liquid scintillation counters. Some specialty RIAs (e.g. fentanyl assays prepared by Janssen Laboratories) are available for drug analysis that use ^{3}H-labeled antigens. These require a liquid scintillation counter. RIAs are heterogeneous assays and separation of the bound from unbound labeled antigen is required. For the common ^{125}I-dependent assays, two separation methods have been used. The double antibody method is used for the RIAs that used to be marketed by Roche Diagnostics (Abuscreen), many of the RIAs marketed by DPC and all of the RIAs marketed by Immunalysis Corp. (the Immunanalysis drugs of abuse panel incorporates many of the reagents previously supplied by Roche Diagnostics as their Abscreen line of RIAs). Following the initial antigen–primary antibody complex formation, a second antibody specific for the primary antibody is added. This allows the formation of a precipitable lattice, that is pelleted by centrifugation (Figure 5a). DPC also markets RIAs for some drug groups with antibody attached to a plastic tube (Coat-A-Count RIAs). Here the antigen–antibody complex remains attached to the tube. Centrifugation is not required; the unbound material is simply decanted out of the tube (Figure 5b).

As the amount of drug in the sample increases, more of the radiolabeled antigen will remain in the unbound state; the amount of radioactivity in the pellet, or attached to the tube, decreases (Figure 5c). As the short half-life of the ^{125}I-labeled antigens results in rapid loss of counts per minute (cpm) from one day to the next, comparison of RIA data usually depends upon use of the ratio of bound cpm to the cpm in a sample where the concentration of unknown drug is zero (B/B_0) (Figure 5c).

Because they use radioactive materials, RIAs require special laboratory conditions for the safe storage, handling and disposal of the radioactive material. A related aspect of RIAs is the above-mentioned half-life of the radiolabel. ^{125}I has a 60 day half-life that is essentially the shelf-life of the reagent. Although this limits the time over which the reagents can be used, it also makes storage of waste material simpler. After a number of half-lives the material can be more readily disposed of. ^{3}H with a half-life of 12.4 years can be stored almost indefinitely, but will always require special disposal procedures. A summary of the use of RIAs to detect drugs of abuse in urine is presented in Table 8.

2.2.2 Enzyme-linked Immunosorbent Assay

ELISAs have been used in immunochemistry for a number of years. The absorbance, or binding, of an antibody or antigen to the wells of 96-well plates (or strips of plates) makes these assays amenable to rapid handling of up to 96 samples at a time – a number that can be readily multiplied, as the handling of more than one plate at a time is not problematic. ELISAs have been formulated in many different styles that can

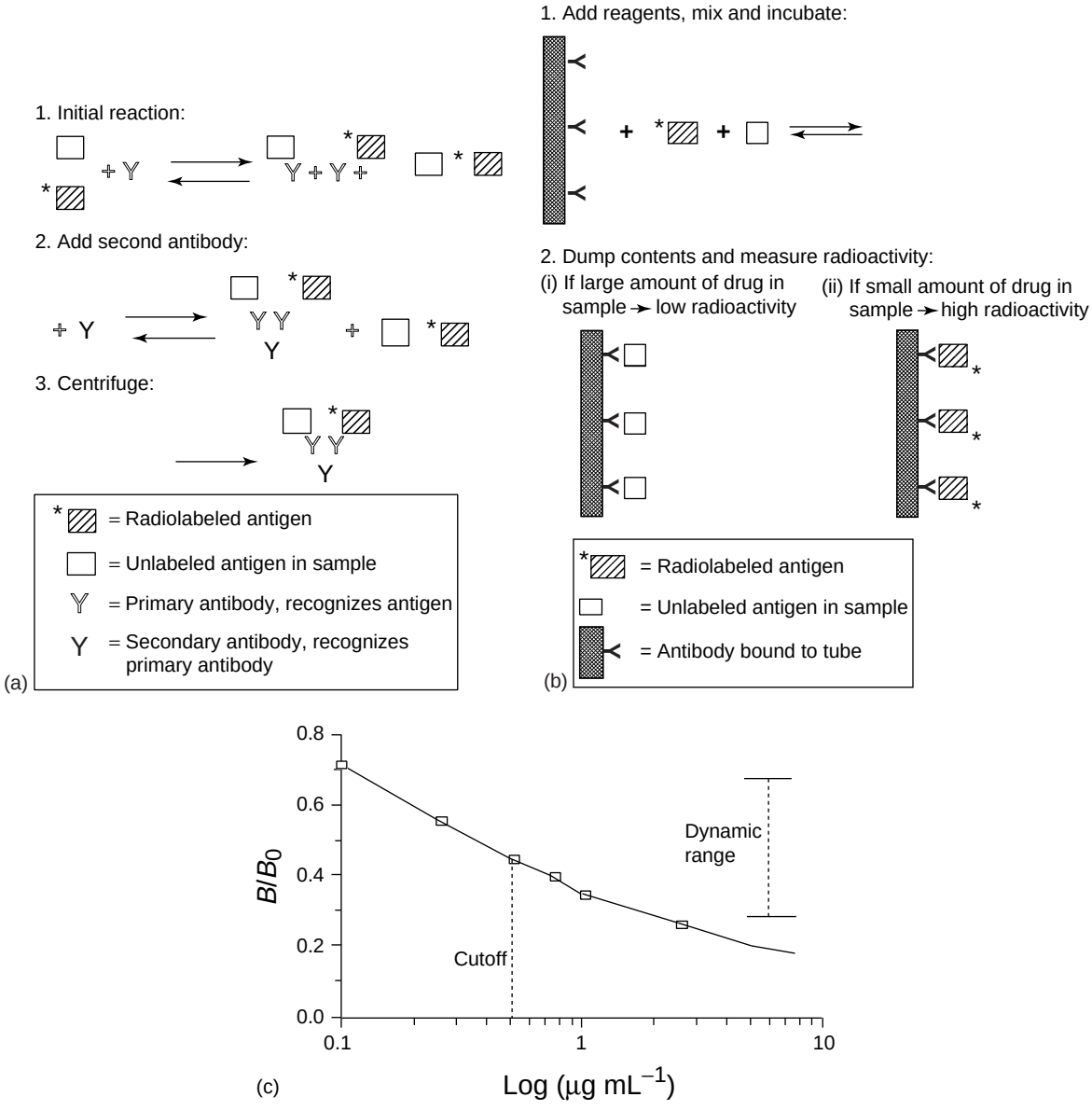

Figure 5 A schematic of RIAs showing (a) an RIA that uses double antibody method to separate bound from unbound antigen, (b) an RIA that uses antibody-coated tubes to allow separation of bound from unbound antigen (e.g. DPC Coat-a-Count), and (c) a representative plot of signal versus drug concentration.

use secondary, and even tertiary, antibodies or other variations to amplify the reaction. The two types of ELISAs marketed for detection of drugs use a simple system. The antibody is bound to the well. Sample and horseradish peroxidase conjugated antigen compete for the binding sites. These are then rinsed out, and the bound enzyme is measured with the substrate 3,3′,5,5′-tetramethylbenzidine (Figure 6). The enzymatic reaction forms a blue product that is measured at 630–650 nm. As recommended for the STC Diagnostics products, and offered as an option for the Diagnostix products, the reaction can be stopped by addition of 2 N sulfuric acid, which will convert the blue product to a yellow product monitored at 450 nm. As the amount of target drug in the sample increase, enzyme-conjugated antigen will be displaced and the resultant absorbance will decrease, as seen with the EIAs (Figure 2c). The assay requires a plate reader capable of monitoring the appropriate wavelengths. The assay can be facilitated with the use of a plate washer. The only published study of these ELISAs for urine drug testing is an evaluation of the STC Diagnostics method for the detection of LSD in urine.[93] Further discussion of these methods will be presented in section 3.5 on other matrices.

Table 8 Representative studies on RIAs for drugs of abuse in urine

Drug group studied[a]	Reagent	Methods compared	Refs.
4	Abuscreen	HI, FRAT, TLC	5, 12
4	Abuscreen	FRAT	13
4	Abuscreen	EMIT, FRAT, HI, TLC	14
1, 4, 6	Abuscreen	EMIT, TLC	89
4	Abuscreen	EMIT	17
3	Abuscreen	EMIT, TLC, GC/FID	19
2	Abuscreen	GC/MS	90
2	Abuscreen	EMIT, GC/MS	22
2	Abuscreen	EMIT, GC/MS, HPLC, GC/ECD	23
2	Abuscreen	EMIT, GC/MS	25
2	Abuscreen	EMIT, TLC, GC/MS	24
2	Abuscreen	FPIA, GC/MS	79
2	Abuscreen, DPC, Immunanalysis	EMIT, FPIA	33
1–3, 5	DPC	GC/MS	91
3	Abuscreen, DPC	EMIT, FPIA, TLC, GC/MS	30, 31
8	DPC	In-house RIA	92
2	DPC	EMIT, GC/MS	39
2	Abuscreen	EMIT, FPIA, GC/MS	40
4	Abuscreen, DPC	EMIT, FPIA, GC/MS	47, 48
2–4, 6	Abuscreen	EMIT, FPIA, KIMS	52
2–6	Abuscreen	EMIT, KIMS, GC/MS	53
7	DPC	EMIT, FPIA, KIMS, Triage, GC/MS	56
2–6	Abuscreen	EMIT	57
1–6	Abuscreen	EMIT, CEDIA, FPIA, KIMS, GC/MS	61
8	Abuscreen	KIMS, GC/MS	86
8	DPC	ELISA	93
8	DPC	LC/MS	94
8	DPC	CEDIA, GC/MS	75
8	DPC	CEDIA, EMIT, GC/MS	74

[a] See Table 1 for number assignment to drug groups.

2.3 On-site Drug Testing Immunoassays

Over the past few years, a number of companies have marketed ''on-site'' drug testing devices. Most of these kits utilize immunochromatographic techniques that allow the detection of a complex of membrane-adherent protein-conjugated antigen with antibody conjugated to a colored particle (e.g. colloidal gold, blue-dyed microparticles). With application of urine to a test strip, the target drug and conjugated antibody move up the strip by capillary motion (Figure 7a). Unbound antibody will bind to antigen at the viewing membrane. When no target drug is present, sufficient antibody binds at the viewing membrane to form a colored line (negative result, Figure 7b). When sufficient target drug is present, all the antibody will be bound and will not bind to the viewing window with no color formation (positive result, Figure 7c). Each strip also contains a control antibody that is specific for a control (nondrug) antigen adhered to a different section of the viewing membrane. This control reaction tests for appropriate functioning of the kit, and must form a colored line for acceptable results (Figure 7d).

Sample application to the kit may be done by dipping a strip(s) directly into the urine collection device, pipetting a portion of the urine into a test well or tilting a collection cup so that a portion of the urine flows into attached testing wells. Kits are configured to test from one to 10 drugs in a single device. Some kits also include a human urine test, where the same principle is used to detect human immunoglobulin G (IgG) in the sample (anti-human IgG is conjugated to the test antibody, and human IgG is adhered to the viewing membrane).

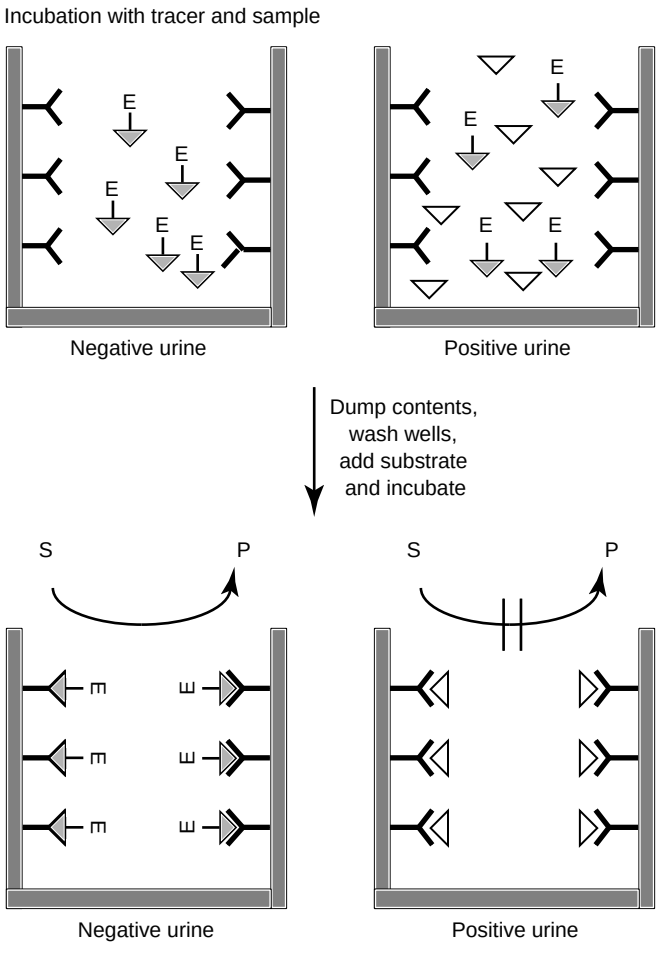

Figure 6 A schematic of a simple ELISA (see Figure 2c for a representative plot of signal versus drug concentration).

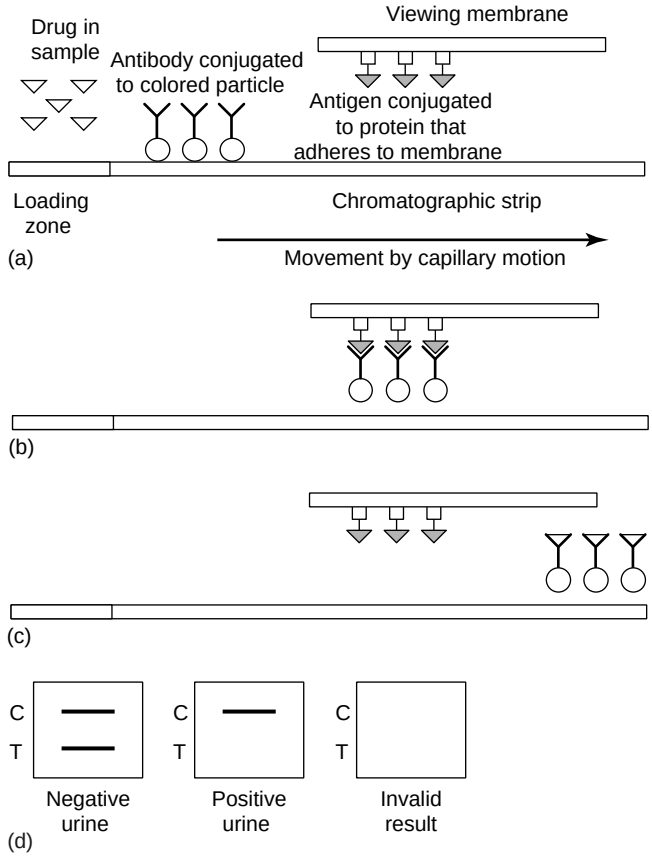

Figure 7 A schematic showing the principle of many on-site drug testing devices that shows (a) components of the device, (b) the end result with a negative urine, (c) the end result with a positive urine and (d) the results seen in the viewing window including both that for the target compound and that for the assay control.

A number of the kits have now undergone rigorous evaluations comparing their results with those of laboratory-dependent immunoassays and GC/MS confirmation with fairly favorable results (Table 9). Their primary limitation may arise from the subjective nature of detecting a minor color difference between a sample that contains drug at concentrations just below, or just above, the cutoff. In part, this will be determined by

Table 9 Evaluated on-site immunoassays for detection of drugs of abuse

Manufacturer	Kit Name	Refs.
Roche Diagnostics	ONTRAK	73, 87, 88, 95–99
	ONTRAK TESTCUP	88, 99–101
Biosite Diagnostics Inc.	Triage	45, 54, 56, 66, 73, 83, 99, 102, 103
Abbott Laboratories	Advisor Reaction Disc	104
Hycore Biomedical Inc.	accuPINCH	105
Editek Inc.	EZ-Screen	73, 84, 106
Morwell Diagnostics GmbH	RapiTest	69
Princeton BioMedi-tech	Accusign	85, 99, 101
Boehringer Mannheim GmbH	FRONTLINE	72
Pharmchem Laboratories	PharmScreen	101
Life-Sign (previously Orion Diagnostics)	Status DS	101
American Bio Medica	Rapid Drug Screen	101

the rigor employed at any one testing site to establish concomitant testing of control urines.

''On-site'' kits offer a means for rapid sampling of specimens soon after their collection, precluding the need for packaging and shipping of specimens to a laboratory. For this purpose, they may find favor for nonregulated workplace drug testing, emergency room toxicology, testing at drug treatment centers and roadside drug testing. Their use in crime scene investigations is another possibility.

3 SAMPLE PREPARATION AND MATRIX FACTORS

3.1 Introduction

Most of the immunoassays used to detect drugs of abuse in forensic toxicology laboratories are formulated for use with urine samples. The analysis of physiological fluids, besides urine, and tissues for drugs and drug metabolites, however, is an essential component of forensic toxicology. Many immunoassays have therefore been evaluated for their use with these nonurine matrices. Three main factors must be considered in the use of urine-formulated immunoassays in analysis of other fluids and tissues. First, with the exception of most parenchymal organs with their rich blood supply, other fluids and tissues will have lower concentrations of the target drug than found in urine. Second, with the exception of other nonviscous fluids (e.g. vitreous humor, cerebrospinal fluid), the matrix has to be converted in to a more soluble form for analysis. Third, the analyte targeted by the urine-formulated immunoassay may not be the metabolite (or parent drug) that is present in that particular matrix at the highest concentration. Many strategies have been devised to deal with these issues. Those used for analysis of blood and hair will be addressed in some detail. Issues pertaining to the immunoanalysis of other tissues and fluids, including urine, will also be addressed.

3.2 Urine

Urine is the most common matrix used for forensic drug testing. While some issues of privacy occur with urine collection, it is still a noninvasive mode of sampling and relatively large volumes can be readily collected. Furthermore, urinary excretion of most drugs, or their metabolites, leads to an accumulation of the analyte in the bladder; concentrations of drugs and/or drug metabolites in urine are relatively high. Aliquots of urine are commonly used without any prior sample preparation. A common exception is the need to centrifuge particularly turbid specimens prior to analysis by assays that use spectrometric detection methods. Another key exception to this rule is immunoassay screening for benzodiazepines. As first demonstrated by Beck et al.[44,107] and subsequently substantiated by others,[60,108,109] enzymatic hydrolysis of urine samples will increase the immunoassay detection of many benzodiazepines.

3.3 Blood

Blood, or the blood fractions serum and plasma, has traditionally been the favored sample for drug testing that is conducted to assess potential impairment or intoxication by the drug. Much more is known about correlation between blood concentrations of the drug or active metabolite with pharmacodynamic functions than with any other matrix. Two main issues have to be considered when transferring technology formulated for urine drug testing to blood drug testing. The first, is that most drugs and drug metabolites will be present in blood at much lower concentrations than are found in urine. Second, bloods contains cellular material (either whole or fragmented in hemolyzed blood) at much higher concentrations than found in urine. Attempts to use blood in urine immunoassays have had to develop procedures to enhance the sensitivity of the assay (or accept the limitations of sensitivity of the assay in regards to the matrix), and to transform the matrix to an acceptable form for analysis (Table 10).

The RIA has been applied to blood analysis with the fewest modifications. With the exception of analysis for cannabinoids, blood can be directly analyzed by RIA.[92,116] Changes in sample and reagent volumes have been added to enhance sensitivity.[91,113,128] Extractions of blood for RIA analysis have also been proposed when the sample was being analyzed using a robotic system[124] and to enhance the sensitivity for benzodiazepines.[128] Cannabinoid detection by RIA has required a simple methanolic deproteinization;[90] blank whole blood itself has too much variation in signal to allow proper calibration.[123]

Like RIAs, ELISAs are heterogeneous immunoassays. Theoretically, matrix differences will also have less impact on ELISAs used for drug testing. These assays are new to the drug testing field, however, and only a few studies have been published on their use. Those now available show that ELISA has promise for testing of whole blood with little or no modification. Methanolic deproteinization was used by Perrigo and Joynt to study a number of drugs,[127] Cassells et al.[93] were able to use whole blood directly to detect LSD in LSD-fortified blood samples and Spiehler et al.[131] diluted blood samples 1 : 2 with buffer to facilitate ELISA detection of a number of drugs.

Analysis of blood with EMIT, CEDIA and TDx homogeneous immunoassays has always required, at the

Table 10 Use of immunoassays for detection of drugs of abuse in whole or hemolyzed blood

Drug groups studied[a]	Immunoassay	Modification	Refs.
2–4, 6, 7	EMIT	Drug group-specific organic extractions with reconstitution in aqueous media	110
2	EMIT	Mix blood with methanol (1 : 2), centrifuge, store in freezer overnight and use dilution of supernate in buffer (1 : 4). Subsequent modifications reported in 1989	111, 112
2	Abuscreen RIA	Extract blood with methanol (1 : 2), centrifuge and use 50 µL of supernate.	90
3–6	Abuscreen RIA	Increased sample size to 100 µL	113
1–7, 9	EMIT	Extract blood with methanol (1 : 2), centrifuge, filter and use 50 µL of supernate. Update of instrument settings in 1992 to enhance sensitivity and reagent efficiency	114, 115
4	DPC RIA	Use whole blood directly	116
1–4, 6, 7	EMIT	Add blood to acetone (2 : 5), vortex, centrifuge, remove extract and dry. Reconstitute in EMIT buffer–methanol (1 : 1), centrifuge and use supernate	117
4	TDx FPIA	Mix blood with equal volume of saline	118
1, 3–5	DPC RIA	Whole blood used with stated modifications in sample and tracer volume	91
2–4, 7	EMIT	Mix blood with *N*,*N*-dimethylformamide (1 : 2), centrifuge and use supernate	119
8	DPC RIA	Use whole or hemolyzed blood directly	92
1, 2, 4, 7	EMIT	Mix blood with methanol (1 : 2), place at −20 °C for at least 20 h, thaw, centrifuge, filter and use supernate with stated Monarch settings	120, 121
1	TDx FPIA	Mix 0.2 mL blood with 3.5% zinc sulfate in 50% methanol (1 : 1), centrifuge and use supernate. (Good evaluation of other precipitation methods)	122
2	Abuscreen RIA	Mix 0.2–1.0 mL blood with methanol (1 : 2), centrifuge and use 0.1 mL of supernate. Comparison with unextracted blood	123
1, 3–7	DPC RIA	Mix 1.0 mL blood with 3% zinc sulfate, 0.5% sulfosalicyclic acid in 50% methanol (1 : 1), centrifuge and use supernate in automated RIA system	124
2	ADx FPIA	Mix 0.5 mL blood with 0.5 mL phosphate buffer (pH 6)–methanol (80 : 20) and centrifuge in Ultrafree® tubes; use filtrate	125
1–7	EMIT	Multi-step extraction of 1.0 mL of blood that results in two concentrated fractions for analysis	126
1–5	STC ELISA	Mix blood with methanol (1 : 2), centrifuge, store in freezer overnight and use dilution of supernate in buffer (1 : 4)	127
7	EMIT, FPIA, RIA	Add 7.0 mL butyl chloride to 1.0 mL blood, mix, centrifuge, dry organic phase and reconstitute with 0.5 mL TDx buffer. RIAs can use whole blood	128, 129
8	STC ELISA	Used 50 µL whole blood directly	93
2	CEDIA	Add 1 mL blood to 2.5 mL acetone while mixing, centrifuge and filter supernate through SPE frit. Add 0.05 mL 1% HCl in methanol, dry and reconstitute in 0.5 mL EMIT buffer–methanol (1 : 1). Centrifuge and use supernate	130
2	AxSym FPIA	Add 0.5 mL blood to 1.0 mL acetonitrile and 0.2 mL saturated sodium sulfate while mixing, centrifuge and collect top layer for analysis.	130
1, 3, 4, 6	STC ELISA	Use whole blood that is diluted 1 : 2 with phosphate buffer in microplate	131
1–4, 7, 9	CEDIA	Add 0.5 mL blood to 1.0 mL acetone, mix and centrifuge. Dry supernate and reconstitute with 0.5 mL ADx buffer	132

[a] See Table 1 for number assignment to drug groups.
[b] Abbreviation not previously defined: SPE, solid-phase extraction.

least, deproteinization or more thorough extractions with reconstitution (Table 10). Simple deproteinizations that can be followed by centrifugation and direct analysis of the supernatant (sometimes after frozen storage) have been reported using methanol,[111,114,120] *N*,*N*-dimethylformamide,[119] zinc sulfate in methanol[122] and sodium sulfate in acetonitrile.[130] Yee et al.[125] reported a simple modification of these simple deproteinizations where blood was mixed with phosphate buffer–methanol (80 : 20) and centrifuged in a filtering tube, so the filtrate could be used. Extractions with reconstitutions have ranged from simply adding acetone[117,132] or butyl chloride,[128] drying the organic phase and reconstituting, to more elaborate extraction schemes.[110,126] The more elaborate extractions will recover and concentrate more of the drug, but they can be much more time-consuming.

Serum has less protein and other cellular material than whole blood, and many automated homogeneous immunoassays have been developed for the direct measurement of therapeutic drugs in serum. As serum is the most common matrix collected in emergency room settings, there is often a need to use serum for drugs of abuse testing. Applying immunoassays developed for urine drug testing to serum samples, however, still has to deal with the lower concentrations of drug found in this matrix. Examples of assay of serum samples follow.

Poklis et al.[133] found that they could measure serum directly for benzoylecgonine with EMIT d.a.u. reagents on an ETS® using calibrators modified to a lower cutoff of 50 ng mL^{-1}. At this cutoff, positive versus negative results in serum matched the results of urine samples screened at a 300 ng mL^{-1} cutoff. GC/MS of the 45 positive sera with a cutoff of 50 ng mL^{-1} was only positive for 67%, but benzoylecgonine was detected below the cutoff in the remaining samples.[133] Moody and Medina[134] compared unextracted serum with serum extracted with chloroform–2-propanol (9 : 1) and reconstitution in half the volume of 0.1 M potassium phosphate (pH 7.4) to detect cannabinoids, benzoylecgonine and morphine using OnLine KIMS reagents on a Cobas Mira. They found that extraction was necessary to detect the target drugs at concentrations anticipated in serum. Furthermore, extraction provided greater precision of analysis.[134] Iwersen-Bergmann and Schmoldt[132] used the same acetone deproteinization–extraction with reconstitution in ADx buffer as applied to whole blood to detect cannabinoids, benzoylecgonine, opiates, benzodiazepines and methadone successfully using CEDIA d.a.u. reagents on a Hitachi 911 instrument with the same settings as used for urine. It was found that further modifications would be required to detect amphetamines.[132]

3.4 Hair

Hair is an important and accessible matrix for the detection of exposure to drugs. One of the more unique features of hair testing is the preservation of the evidence of exposure for periods of time approximately proportional to the length of the hair.[135–139] Ongoing controversies about the use of hair as a matrix for workplace drug testing exist. These include the potential for external contamination and amassing evidence that basic drugs accumulate in darker hair to a greater extent because of drug binding to melanin.[135,139–142] Hair is, however, a well-studied matrix, that with proper allowances for extenuating circumstances can be used for forensic drug testing.

Hair cannot be measured for drugs of abuse directly by immunoassay. A number of thorough reviews of techniques to prepare hair for both immunoassay and chromatographic assays have been published.[136,143–146] Representative studies in which hair extracts were analyzed by immunoassay are presented in Table 11. The hair sampling should avoid inclusion of the hair follicle, as this will include some blood and other tissue. Whether hair is further segmented by length to assess approximate time intervals is optional, but should be noted. The hair sample usually goes through a rinsing/washing step followed by an extraction/digestion of the hair. A number of different washing techniques have been employed that included dilute detergents, surfactants and weak solvents (Table 11). Methanolic extraction was used in initial studies[147] and has continued to be used in some laboratories. Most laboratories, however, have employed extraction methods that essentially digest the hair protein matrix. These have included digestions in strong acid (HCl) or base (NaOH) and proteolytic digestions (Table 11). The first two methods require neutralization of the digest before it can be used with an immunoassay. Although proteolytic digestion was initially introduced in 1989 by Offidani et al.,[155] it has only recently become more widely used. It is recommended when unstable compounds are to be detected. Supercritical fluid extraction of rinsed hair has also been used to prepare extracts for immunoassay.[172,173]

RIA has been used almost exclusively for the detection of abused drugs in hair (Table 11). RIAs are generally more sensitive, which permits the detection of the low concentrations of drug or drug metabolites that are present in hair. In addition, because RIAs are heterogeneous immunoassays, they are probably less susceptible to interference from residues of the hair processing methods. Franceschin et al.,[153] Kintz et al.[159] and Lewis et al.[175] have reported the use of TDx FPIAs to detect drugs in hair. Whereas Franceschin et al.[153] and Lewis et al.[175] required a more extensive extraction technique, Kintz et al. [159] reported that a number of different drugs could

Table 11 Representative use of immunoassays for detection of drugs of abuse in hair

Drug groups studied[a]	Immunoassay	Modification	Refs.
4	Abuscreen RIA	Wash 10 mg hair in 1 mL detergent (3×) and 1 mL H_2O (3×), dry and pulverize. Heat in 5 mL of methanol for 2 h, centrifuge, collect methanol and dry. Reconstitute in 1 mL phosphate buffer (pH 7.4)	147
3, 4	Abuscreen RIA	Incubate 50 mg hair with 0.1 M HCl at 45 °C overnight, neutralize with 0.1 mL 1 M NaOH and 0.9 mL 1 M phosphate buffer (pH 7.5). Use 100 µL for analysis. (Presents results on a number of different methods)	148
6	Abuscreen RIA	Wash 0.4 mg hair in 20 mL H_2O (10×), cut and add to 0.2 mL 0.1% SDS–saline.[b] After 24 h, remove 0.1 mL for analysis	149
3	Abuscreen RIA	Shake 10–20 mg hair in 2% Prell shampoo in H_2O for 10 min. Rinse with 10 mL H_2O (10×) and dry. Reflux in ethanol for 4 h, collect extract, dry and reconstitute in 0.2 mL ethanol. Use 20 µL for analysis	150
3	Abuscreen RIA	Wash 58 mg with 100 mL 0.05% SDS (6×), cut and pulverize. Reflux in 10 mL ethanol for 2 h, filter, dry and reconstitute with 0.2 mL negative urine	151, 152
4	TDx FPIA	Incubate 50–200 mg overnight with 2 mL 0.1 M HCl, neutralize with 0.2 mL 1 M NaOH, collect liquid, wash residue with 2 mL saline, combine, extract, dry and reconstitute with 0.2 mL TDx buffer	153
3	Abuscreen RIA	Wash 100 mg hair with 10 mL H_2O (3×) and 10 mL ethanol (3×), crush 50 mg hair in 1.0 mL 0.1 M HCl and incubate overnight at 45 °C, neutralize with 0.1 mL 1 M NaOH, add 0.9 mL phosphate buffer (pH 7.4) and use mixture	154
3, 4	DPC RIA	Wash 20–100 ng hair with methylene chloride (2×), Treat with 1 mL of 1 mg mL^{-1} pronase in 0.05 M TRIS (pH 8.25) for 24 h at 39 °C. Terminate with exposure to 100 °C and add 0.2 mL 1 M phosphate buffer (pH 7). Compared with 1 M NaOH. See Strano-Rossi et al.[171] for update on method	155
4	DPC RIA	Wash 50 mg hair in acetone, H_2O and acetone; dry. Boil hair in 1 mL 1 M NaOH, cool and neutralize with 1 mL 1 M HCl. Use mixture for analysis	156, 157
3	DPC RIA	Wash 100 mg hair with 2 mL ethanol at 37 °C for 15 min, wash 2× with 3 mL 0.1 M phosphate buffer (pH 7) at 37 °C for 15 min (additional wash if wash is not drug-free). Remove buffer and incubate with 1 mL 0.1 M HCl overnight at 45 °C. Neutralize 25 µL with 25 µL saturated NH_4Cl–NH_4HCO_3 (1 : 1), add 25 µL phosphate buffer and analyze	158
1–4, 6, 7	TDx FPIA	Wash 50 mg hair in 5 mL ethanol for 15 min at 37 °C. Dry and incubate in 3 mL 1 M NaOH for 1 h at 100 °C. Neutralize with 3 M HCl, centrifuge and dilute supernate (1 : 1) with ADx buffer	159
4, 9	DPC RIA	Evaluated a number of different washing procedures. After wash, incubate overnight in 0.1 M HCl at 55 °C, neutralize 0.5 mL with 0.5 mL 1 M NaOH and 0.5 mL 0.1 M phosphate buffer (pH 7). Use 50 µL of mix	160–163
3	Abuscreen RIA	Wash 2 mg hair in 2 mL ethanol of 37 °C for 30 min (4×) and for 60 min (2×). Sonicate 2–5 mg hair in 1 mL methanol for 30 min and incubate in same overnight. Collect methanol, rinse with 1 mL methanol and combine. Dry and reconstitute in 0.1 mL PBS[b]	164, 165
Fentanyl	DPC RIA	Wash 10 mg hair with 1 mL methanol (3×), Incubate with 1 mL methanol at 40 °C overnight, collect, rinse with 1 mL methanol, combine, dry and reconstitute with 0.1 M citrate buffer (pH 6)	166
3	DPC RIA	Wash hair with 10% tincture green soap in H_2O, rinse 4× with H_2O and dry. Incubate 100 mg with 1 mL 0.1 M HCl at 37 °C for 18 h, neutralize 0.1 mL 1 M NaOH and 0.9 mL 0.15 M PBS (pH 7.4), use 50 µL	167
3, 4	DPC RIA	Rinse 50 mg hair with 5 mL 0.1% Tween 80 (2×), 5 mL H_2O (2×) and 1 mL acetone. Dry and digest overnight with 1 mL pronase (1 mg mL^{-1}) and dithiothreitol (6 mg mL^{-1}) in TRIS buffer (pH 7.1). Collect filtrate and use 25 µL for analysis	168–170

Table 11 (*continued*)

Drug groups studied[a]	Immunoassay	Modification	Refs.
3, 4	DPC RIA	Wash 30 mg hair with 10 mL 0.1% Tween 80 (2×), with H_2O (2×) and then 1 mL acetone. Dry and incubate overnight in 1 mL pronase (1 mg mL^{-1}) and dithiothreitol (6 mg mL^{-1}) in TRIS buffer (pH 7.1). Centrifuge and use 25 μL of supernatant for analysis	171
3	DPC RIA	Wash 100 mg hair with 2 mL ethanol at 37 °C for 15 min, wash 2× with 3 mL 0.1 M phosphate buffer (pH 7) at 37 °C for 15 min. Perform supercritical fluid extraction with triethylamine under a CO_2 flow. Dry off solvent and reconstitute in methanol	172, 173
4	DPC RIA	Rinse hair in 10 mL acetone, 10 mL H_2O and 10 mL acetone and dry. Incubate with 1 M NaOH for 18 h at 55 °C, neutralize with 0.1 mL HCl and then 0.1 M phosphate buffer (pH 7) to pH 6.5–7.5. Use 50 μL for analysis. (Note study compares a number of NaOH digestion methods)	174
1–3	TDx FPIA	Wash 50 mg hair with H_2O, 3 mL methanol and 3 mL acetone. Incubate with 3 mL 0.1 M HCl overnight at 55 °C, collect acid and incubate in mixture of 0.2 mL ethanol, 2 mL H_2O and 0.2 mL 12 M H_2O_2 at 80 °C for 30 min. Cool and add 1 mL glacial acetic acid, centrifuge, add supernatant to collected acid fraction and add 9 mL H_2O. Then extract using SPE that provides neutral and basic drug fractions	175
3, 4	DPC RIA	Wash 50–200 mg hair with 20 mL 0.3% Tween 20 (2×), and then H_2O. Dry, cut and incubate in 0.25 M HCl overnight at 45 °C. Neutralize with 0.5 mL 1 M NaOH and extract (2×) with Toxi-tubes A. Pool the organic layers, dry and reconstitute in 0.5 mL 0.05 M phosphate buffer (pH 5)	176

Note: A number of studies in the literature had hair drug analysis conducted at Pychomedics Corp. These studies were not included in the table as the wash and extraction procedures were described as ''proprietary''. Although the current Psychomedics procedure has now been detailed,[136] it was not possible to ascertain which specific procedure was employed at the time of any one study.

[a] See Table 1 for number assignment to drug groups.

[b] Abbreviations not previously defined: PBS, phosphate-buffered saline; SDS, sodium dodecyl sulfate; TRIS, tris (hydroxymethyl)aminomethane.

be detected following a more routine extraction of the hair. At this time no other laboratories have reported the use of this assay for hair analysis. The new STC Diagnostics ELISAs for cocaine appear to provide sufficient sensitivity for hair screening; a current study in our Center has used this approach with GC/MS confirmation (D.G. Wilkins, personal communication). While RIAs will continue to play a major role in the immunoassay detection of drugs of abuse in hair, further research will probably find roles for other immunoassays.

While some of the larger commercial laboratories used immunoassay screening of hair on a routine basis,[136,158,175] a great deal of research on the preparation of hair for drug testing has relied solely on chromatographic techniques for drug measurement. As this review has focused on the application of immunoassays, these other studies have not been cited, but should be consulted for more information on hair preparation techniques. Many are referenced in the reviews on hair preparation.[136,143–146]

3.5 Other Matrices

A number of other tissues and physiological fluids have been studied as matrices for the detection of drugs of abuse (Table 12). Three that have been examined extensively are saliva, meconium and sweat. Saliva can be collected in a noninvasive manner, and therefore has been advocated as an alternative to blood for drug testing. Concentrations of some drugs are present in saliva at concentrations close to those in blood.[206,207] Many factors affect the secretion of drug and drug metabolites into saliva, pK_a and plasma protein binding being two of the most important.[208] As salivary pH is variable, it usually differs from that of blood, and this difference appears to have the greatest impact on blood to saliva ratios of drugs with pK_a values <6 or >8.[208] These variations can have an impact on interpretation of saliva drug concentrations, when most of the pharmacodynamic studies are based upon correlates to blood concentrations. A number of immunoassay approaches to detecting drugs in saliva have been reported (Table 12).

Meconium has been widely used in studies on the exposure of the fetus to drugs from drug use during pregnancy.[209,210] As the inital fecal excretion of a newborn, meconium can be collected in the diaper(s) over 24–36 h after birth, and provides an accumulative record of drugs that circulated in the amniotic fluid during the latter stages of gestation. As an accumulative

Table 12 Representative use of immunoassays for detection of drugs of abuse in other matrices

Drug groups studied[a]	Immunoassay	Physiological fluid or tissue	Refs.
4	FRAT	Saliva	7
4	Abuscreen RIA	Saliva	177
2–4, 6, 7	EMIT	Liver, brain and kidney extracts	110
7	Abuscreen RIA	Saliva	178
6	Abuscreen RIA	Bloodstains, semen, seminal stains, perspiration stains	149
2	EMIT	Saliva	179
3	Abuscreen RIA	Perspiration stain, menstrual bloodstain	151
2–4	Abuscreen RIA	Meconium	180–186
4	TDx FPIA	Liver homogenate, bile	118
3	Abuscreen RIA	Saliva	187
8	DPC RIA	Stomach contents	92
4	DPC RIA	Saliva	187
9	DPC RIA	Sweat	188
1, 3–6	DPC RIA	Liver and spleen extracts	124
3	TDx FPIA	Meconium	189, 190
3	DPC RIA	Saliva	191
3	Abuscreen RIA	Dried blood spot	192
2–4	EMIT d.a.u.	Meconium	193
3, 4	EMIT	Breast milk extract	194
1, 3–5	EMIT d.a.u.	Meconium	195
2–4, 9	EMIT d.a.u.	Meconium	196
3	DPC RIA	Sweat	197
1–4	TDx FPIA	Meconium	198–200
3	STC ELISA	Sweat	201
1	STC ELISA	Sweat	202
1–7	TDx FPIA	Synovial fluid, vitreous humor	203
2	EMIT	Meconium	204
2	DPC RIA	Fingernails	205

[a] See Table 1 for number assignment to drug groups.

record, meconium measurements do not pinpoint times of exposure. The analytes present in meconium are a product of both maternal and fetal metabolism. Some metabolites rarely seen are found in meconium, probably because of unique fetal metabolic pathways. A number of studies have addressed the use of immunoassays to detect drugs of abuse in meconium homogenates or extracts of meconium homongenates.

Sweat is another physiological fluid that is gaining popularity as a matrix for drug testing. The presence of drugs in sweat has been known for many years.[211] The use of this fluid has lately become more popular as collection patches have been developed that permit the evaporation of water and other small molecules. This alleviates the irritation that was common with older methods of collection[212] and precluded sampling for more than a day or two. Patches can now be applied for 1–2 weeks, and are therefore advocated for testing programs where long-term monitoring is important, such as for patients in drug treatment programs and parolees. Immunoassays, particularly RIA and more recently ELISA, have been used to detect drugs of abuse in sweat (Table 12).

Several other matrices have been reported as useful for the detection of abused drugs and other drugs of interest to the forensic toxicologist (Table 12). The use of a particular matrix depends upon the situation at hand (e.g. stains or dried blood spots), limitations of other matrix availability and the specific population to be studied. If the technology applied is sensitive enough, a drug or drug metabolite may be found in the matrix. Immunoassays occasionally provide the sensitivity for this purpose (Table 12).

4 FORENSIC APPLICATIONS

4.1 Detection Limits and Cutoffs

The sensitivity of an immunoassay is commonly defined from the signal generated by a blank matrix. The usual experiment would analyze a number (e.g. 20) of replicate

blank matrix samples and determine the mean and standard deviation (SD) of the blanks. The concentration that is equivalent to the blank signal ±3SD would be established as the limit of detection (LOD).[213] The lower limit of quantitation (LLOQ) and upper limit of quantitation (ULOQ) should be established if the immunoassay is to be used in a quantitative manner. The LLOQ can be defined in a manner similar to the LOD, in this case using the blank signal ±10SD.[213] It is becoming common practice, however, to define both upper and lower limits from actual replicate measurements of test calibrators, and accepting the limits when precision and accuracy are within certain limits (e.g. 15–20%).

Cutoffs are used in many testing procedures to establish strict guidelines as to whether the result of an immunoassay screen will result in further confirmational testing, or in screening-only programs a positive test result. In those programs, cutoffs are also used for confirmations to dictate a concentration that will trigger a positive report. Cutoffs do not always reflect the LOD of a particular immunoassay. For example, 300 ng mL^{-1} is a commonly used cutoff for methadone detection. Stated LODs of immunoassays for methadone range from 0.02 to 20 ng mL^{-1} (Table 13). EIA reagent manufacturers rarely state a sensitivity for their assay, in part because they are formulated for many different instruments and sensitivity can be instrument dependent (Table 13).

Testing programs, such as medical examiner's offices, that are interested in the detection of any drug present will often establish cutoffs based upon analytical sensitivity. Not only must the LOD of the immunoassay be considered, but also the LOD or LLOQ of the confirmational assay must be considered. It is worthless to establish a screening cutoff that is lower than the confirmational LOD or LLOQ. Legal challenge is another factor to be considered in establishing the screening cutoff. Two well-known examples involve cannabinoids and opiates.

Most immunoassay methods were initally designed and had sufficient sensitivity to screen for cannabinoids at 20 ng mL^{-1} with few false-positive screens. However, passive inhalation of marijuana smoke (under potentially extreme conditions) was found to result in some urines that screened positive at the 20 ng mL^{-1} cutoff, but few or none at the 100 ng mL^{-1} cutoff.[214–217] Therefore, when cutoffs for federal urine drug testing programs were initially established, the potential for a ''passive inhalation'' defense played a role in setting the cannabinoid screen cutoff at 100 ng mL^{-1}. With experience, it became evident that many cannabinoid users were being missed with the 100 ng mL^{-1} cutoff and, along with other factors, that passive exposure would not surpass a 50 ng mL^{-1} cutoff. The cannabinoid screen cutoff was lowered to 50 ng mL^{-1}.

Table 13 Comparative LODs for methadone immunoassays where the common cutoff is 300 ng mL^{-1}

Immunoassay	LOD (ng mL^{-1})
Syva d.a.u. EIA	Not stated
CEDIA d.a.u. EIA	≤225 (instrument specific)
DRI EIA	10
TDx FPIA	100
OnLine KIMS	<20
DPC RIA	0.2
Diagnostix ELISA	0.02
STC ELISA	<1

Poppy seeds are legally used in many food products. They contain opiates, including morphine and codeine, or compounds metabolized to morphine and codeine. It has long been known that ingestion of foodstuffs containing sufficient amounts of poppy seeds can result in a confirmed positive screen for opiates, particularly at the commonly employed cutoff of 300 ng mL^{-1}.[218–220] The ''poppy seed'' defense for an opiate positive urine drug test often overturned a positive finding. Other factors were also having an impact on the use of a 300 ng mL^{-1} cutoff for opiates. Prescription use of codeine was causing a number of positives. While the licit nature of these positives would eventually be unraveled, it required a great deal of energy from the medical review officers and laboratory directors. To counter these and other factors, the screening cutoff for opiates has been increased to 2000 ng mL^{-1} in many drug testing programs. In addition, to focus on positive illicit opiate use, confirmations must also include the detection of 6-monoacetylmorphine, a metabolite unique to heroin.

Other licit foodstuffs and drugs may cause similar problems if either an incomplete history is available or if the immunoassay test result is the sole finding. Examples of foodstuffs include hemp seed oil and coca tea consumption. Selegiline is metabolized to *l*-amphetamine and *l*-methamphetamine; some inhaler cold remedies contain *l*-amphetamine. Use of either of these in sufficient doses can result in urine concentrations that trigger a positive immunoassay test in assays that do not have specificity for the *d*-isomer. A chiral chromatographic confirmation is then required to discriminate these from *d*-amphetamine and *d*-methamphetamine. Such a confirmation is now used by many laboratories.

The establishment of the cutoff can be mandated by law, or determined by negotiation between the laboratory and its client. Examples of some federally mandated cutoffs for pre-employment and random urine drug testing are given in Table 14. Railroad employees will be routinely tested for drugs in urine under DOT guidelines. The Federal Railroad Administration, a component of the DOT, also operates a mandatory postaccident drug testing program.[221] This program will screen for drugs

Table 14 Screening (immunoassay) and confirmation cutoffs ($ng\,mL^{-1}$) applied in representative urine drug testing programs

Drug group for screen/ Analyte for confirm	SAMHSA[a]/DOT[a]		DOD[a]		NRC[a]	
	Screen	Confirm	Screen	Confirm	Screen	Confirm
Amphetamines	1000		500		1000	
Amphetamine		500		500[b]		500
Methamphetamine		500		500[b]		500
MDA[a]				500		
MDMA[a]				500		
MDEA[a]				500		
Cannabinoids	50		50		50/100[c]	
THCA		15		15		15
Cocaine metabolite	300		150		300	
Benzoylecgonine		150		100		150
Opiates	2000		2000		300	
Morphine		2000		4000		300
Codeine		2000		2000		300
Phencyclidine	25		25		25	
PCP		25		25		25
Barbiturates				200		
Butalbital					200	
Phenobarbital					200	
LSD				0.05		
LSD					0.015	

[a] Abbreviations not previously defined: SAMHSA, Substance Abuse and Mental Health Service Administration; DOT, Department of Transportation; DOD, Department of Defense; NRC, Nuclear Regulatory Administration; MDA, methylenedioxyamphetamine; MDMA, methylenedioxymethamphetamine; MDEA, methylenedioxyethylamphetamine.

[b] Confirmation of *d*- and *l*-isomers required.

[c] While the current mandated cannabinoid screening cutoff is $100\,ng\,mL^{-1}$, an interim cutoff of $50\,ng\,mL^{-1}$ has been implemented until it is officially approved. Both cutoffs are used, the $50\,ng\,mL^{-1}$ to determine which samples go on to confirmation and the $100\,ng\,mL^{-1}$ for summary reports (D. Kuntz, NWT Inc., personal communication).

in urine with confirmation in urine and blood. The cutoffs employed by this program (Table 15) illustrate the focus of this program in determining drug involvement in the accident, rather than just prior use of drugs.

4.2 Cross-reactivities

Antibody detection of an analyte is fairly selective, but not perfect. Immunoassays designed for a specific drug (e.g. methadone) will detect other compounds. The extent to which they do that is called the cross-reactivity of the antibody. Cross-reactivity is usually found for analogs and metabolites of the primary analyte. Amphetamines, opiates, barbiturates and benzodiazepines each have a number of over-the-counter (amphetamines) and prescription analogs within their drug class. Many of these will be detected by the respective immunoassay. Depending upon the purpose of the immunoassay, this can be beneficial or a hindrance. In the medical examiner laboratory, this is beneficial. In the workplace setting, where only restricted drugs can be confirmed positive, it is often a hindrance.

With amphetamines, the drugs *d*-amphetamine and *d*-methamphetamine, and in some drug-testing programs, the illicit designer amphetamines MDA, MDMA and MDEA, are confirmed after positive immunoassay screens. When positive screens are caused by over-the-counter amphetamines such as phenylpropanolamine, this is a hindrance to the workplace drug-testing process, and accounts for some of the false-positive screens. A similar situation exists for the opiates. Many drug-testing programs (e.g. SAMHSA) are primarily interested in the detection of illicit heroin use. For this purpose, confirmations are now for morphine and 6-monoacetylmorphine (metabolites of heroin), and codeine (to rule out another source of morphine). There are, however, many other prescription opiates, such as hydrocodone, hydromorphone, oxycodone and oxymorphone, that will cross-react to a certain extent with many immunoassays. The detection of most of these drugs, including the over-the-counter amphetamines when at toxic doses, is important in most other toxicology laboratories.

For benzodiazepines and barbiturates, the cross-reactivity is usually within the drug group for the large number of analogs available on the prescribed drug market. When these drugs are part of the drug screen, the

Table 15 Screening (immunoassay) and confirmation cutoffs (ng mL^{-1}) applied in the Federal Railroad Administration Post-accident Testing Program for Urine and Blood (Revised 2 February 1999).

Drug group for screen/ Analyte for confirm	Urine		Blood	
	Screen[a]	Confirm	Screen[a]	Confirm
Amphetamines	300		50	
Amphetamine		100[b]		20
Methamphetamine		100[b]		20
Cannabinoids	20		10	
THCA		15		5
THC		–		1
Cocaine metabolite	300		20	
Benzoylecgonine		150		10
Cocaine		50		10
Opiates	300		50	
Morphine (total urine/free blood)		300		200
Codeine (total urine/free blood)		300		200
Phencyclidine	25		2.5	
PCP		25		2.5
Barbiturates	200		100	
Amobarbital		200		100
Butalbital		200		100
Pentobarbital		200		100
Phenobarbital		1000		1000
Secobarbital		200		100
Benzodiazepines	100		50	
Nordiazepam		LLOQ[c]		20
Oxazepam		LLOQ[c]		20
Temazepam		LLOQ[c]		20
N-Desalkylflurazepam		LLOQ[c]		20
α-Hydroxyalprazolam		LLOQ[c]		–
α-Hydroxytriazolam		LLOQ[c]		–
Diazepam		–		20
Flurazepam		–		20
Chlordiazepam		–		20
Alprazolam		–		20
Triazolam		–		20
Ethanol[d]	0.01 g/100 mL		0.01 g/100 mL	
Ethanol		0.01 g/100 mL		0.01 g/100 mL

[a] Urine is normally the only specimen screened (KIMS); blood is only screened (RIA) if urine is not available; both blood and urine are confirmed if urine screens positive.
[b] A confirmed urine positive for amphetamine or methamphetamine will result in a *d*- and *l*-isomer analysis and is reported as the percentage of each isomer present.
[c] Urine benzodiazepine concentrations are reported if above the LLOQ and only if the blood concentrations (of same drug or parent drug) are above the cutoff. If a blood specimen is not received and the urine concentration is above the LLOQ, then the urine specimen may be reported.
[d] The blood specimen (or urine if blood is not available) is analyzed for ethanol by gas chromatography (GC). If the blood is positive, the analysis is repeated using a separate portion of the specimen and urine is also analyzed by GC.

difficulty is in the extensive number of drugs that must be considered during confirmation assays. Additional problems occurs for benzodiazepines. The antibody used in most immunoassays is directed towards oxazepam or nordiazepam that are themselves first-generation benzodiazepines, or are metabolites of many other first-generation benzodiazepines.[129] Some of the newer benzodiazepines differ in structure sufficiently such that there is only poor cross-reactivity. This is particularly true for the nitrosobenzodiazepines such as clonazepam and flunitrazepam, and structurally diverse benzodiazepines such as mexazolam and haloxazolam[129] In addition, the newer generations are more potent and have resultant lower urine and blood concentrations than the first-generation benzodiazepines.[222] This compounds the problem of detecting the benzodiazepines with poor cross-reactivity in urine. In blood, the low concentrations can even impede detection of benzodiazepines with relatively good cross-reactivity such as alprazolam and triazolam. Immunoassay detection of some benzodiazepines remains a challenge.

Occasionally, an antibody will be cross-reactive with a drug from a different drug class. These include

Table 16 Representative studies on the cross-reactivities of commercial immunoassays

Drug groups studied[a]	Immunoassay	Class studied	Refs.
1	TDx FPIA, EMIT d.a.u.	Analogs	75
	EMIT d.a.u., TDx FPIA, Abuscreen RIA	Analogs	226
	EMIT d.a.u., TDx FPIA, Abuscreen RIA, DPC RIA	Metabolites	227
	EMIT d.a.u., TDx FPIA, Abuscreen RIA, DPC RIA	Analogs	228
	TDx FPIA	Analogs	229
2	EMIT	Analogs/metabolites	230
	Abuscreen RIA	Analogs/metabolites	23
	Abuscreen RIA	Metabolites	23
	Abuscreen RIA	Metabolites	40, 231
	OnLine KIMS	Metabolites	71
	EMIT	Metabolites	204
3	EMIT d.a.u., TDx FPIA, Abuscreen RIA, DPC RIA	Isomers/metabolites	31
	EMIT II, TDx FPIA	Metabolites	232
4	EMIT d.a.u., TDx FPIA, Abuscreen RIA, DPC RIA	Analogs	47, 233
	OnLine KIMS	Rifampicin	234, 235
6	TDx FPIA	Phenytoin/metabolite	225
7	EMIT d.a.u.	Analogs/metabolites	41, 236
	TDx FPIA	Analogs/metabolites	41, 42, 81, 237
	EMIT d.a.u., TDx FPIA, Abuscreen RIA, DPC RIA	Analogs/metabolites	128, 129
	EMIT d.a.u., Triage	Analogs/metabolites	59
	EMIT d.a.u.	Analogs/metabolites	58
	CEDIA d.a.u., EMIT d.a.u., TDx FPIA	Oxaprozin	223, 224
8	DPC RIA	Analogs/metabolites	92
	OnLine KIMS	Analogs/metabolites	86
	STC ELISA	Analogs/metabolites	93
9	EMIT d.a.u., DRI EIA, TDx FPIA, OnLine KIMS DPC RIA, STC ELISA, Diagnostix ELISA	Analogs/metabolites	238
Fentanyl	Janssen ELISA (same antibody as RIA)	Analogs	239

[a] See Table 1 for number assignment to drug groups.

oxaprozin with benzodiazepine immunoassays[223,224] and phenytoin and its metabolite with the TDx FPIA for barbiturates.[225] As these drugs will rarely be on the respective confirmation list, those false-positive screens are rarely explained.

Many assays will also detect closely related metabolites. As the metabolites will be present in the same urine sample, sometimes in higher concentrations than the target drug, cross-reactivity for the metabolite will contribute to the signal generated by the immunoassay. This is often the case for cannabinoids and is the main reason that the immunoassay cutoff is usually much higher than the GC/MS cutoff.

A representative list of studies on the cross-reactivity of commercial immunoassay for analogs and metabolites is given in Table 16. These studies are best used to see how structurally related compounds can be detected by immunoassays and for methodologies of conducting cross-reactivity studies. The antibodies used in commercial kits will change over time. Therefore, the most up-to-date, albeit often limited, cross-reactivity data is that present in the package insert.

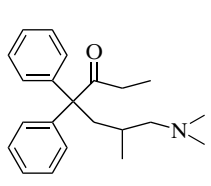

(1)

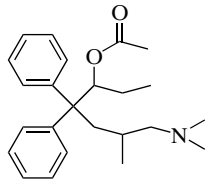

(2)

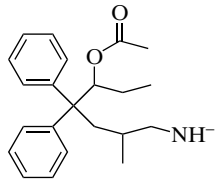

(3)

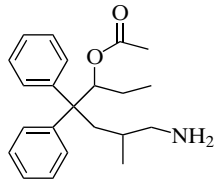

(4)

An example of the effect of analog and metabolite cross-reactivity can be seen with the use of methadone **(1)** immunoassay kits for the detection of *l*-α-acetylmethadol (LAAM) **(2)**. LAAM is a longer acting analog of methadone recently approved as an alternative for maintenance therapy of opioid-dependent subjects.[240] LAAM is successively N-demethylated to norLAAM **(3)**

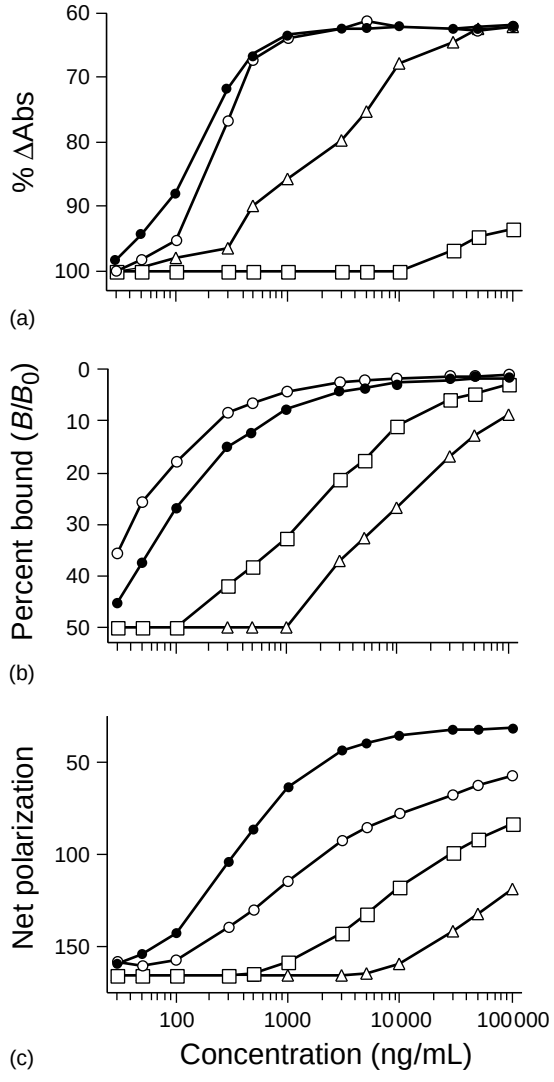

Figure 8 The concentration dependence of detection of methadone, LAAM, norLAAM and dinorLAAM by (a) the OnLine KIMS, (b) the DPC RIA and (c) the TDx FPIA. • – Methadone; ○ – LAAM; – norLAAM; Δ – dinorLAAM.

and dinorLAAM **(4)** and all three compounds are pharmacodynamically active. The structural similarity of methadone, LAAM, norLAAM and dinorLAAM can be seen from the formulas shown. To determine the cross-reactivity of LAAM, norLAAM and dinorLAAM, the compounds were prepared separately in drug-free urine at concentrations ranging from 30 to 100 000 ng mL^{-1} and analyzed using the OnLine KIMS immunoassay, the DPC RIA and the TDx FPIA (Figure 8a–c). Cross-reactivity for LAAM was RIA > KIMS > FPIA. Cross-reactivity for the LAAM metabolites decreased with each N-demethylation (Figure 8a–c). These data suggest that these kits will detect LAAM in urine samples and that methadone immunoassays will not discriminate people who are taking methadone from those taking LAAM.

4.3 Interfering Substances

Particularly in workplace drug testing, an ongoing battle has been occurring to try to interfere with positive urine drug tests. A whole host of common household items or commercially available interferents have been used to alter submitted urines. As seen for cannabinoid immunoassays, many of these interferents will cause positive urines to test negative. Some, however, will cause negative urines to screen positive (Table 17). Many of these substances will be detected with the use of common adulterant tests that include pH, specific gravity, smell and appearance.[241,243,244,246] A notable exception is Visine eyedrops.

Many other substances have been reported to interfere with immunoassays for drugs of abuse. Nonsteroidal anti-inflammatory drugs have been found to interfere with some EMIT and TDx FPIA assays, but to varying degrees.[251–253] Addition of excess antibody reagent to the EMIT cocaine metabolite assay resulted in false-negative screens.[254] False-positive EMIT results have been seen in postmortem urine from subjects with both proteinuria and lactic aciduria,[255] as the presence of both the lactate dehydrogenase and its substrate leads to conversion of NAD to NADH. Denture-cleaning tablets were found to reduce the results of TDx FPIA analysis for amphetamines, benzoylecgonine and THCA.[256]

Although nitrites have not been studied much for their effect on immunoassays, they should also be mentioned as they are a common adulterant that appear to act by degrading THCA. Urines containing high concentrations will screen positive, but in the short time required to go to confirmation the THCA will have decreased sufficiently to not be confirmed.[257]

Concentrations of nitrites that exceed those associated with disease states or therapy have been established[258] and methods to determine nitrite concentrations have been published.[258–260] The test-strip method also tests for glutaraldehyde, which is the other most common adulterant used at this time. As the battle continues, we can be assured that new adulterants will enter the picture.

4.4 Interpretation of Results (Need for Confirmation and Quantitative Versus Qualitative)

All commercial immunoassays have a statement in their package insert that says the assay is only an initial test and all positive results should be confirmed with a second test, GC/MS being currently recommended. The preceding portions of this section should offer the rationale for this statement. An immunoassay screen is quick, but confirmations are not. This often leads to situations, such as the emergency room or the admittance of workers to dangerous workplace settings, where action is taken based upon the result of an immunoassay itself.

Table 17 Representative studies on the effect of adulterants on cannabinoid immunoassays

Adulterant	Type of urines studied							Studies with adulterant on other drug groups
	Positive only[241]	Positive and negative[242]			Positive and negative[243,244]		Positive only[65]	
	EMIT	EMIT	RIA	FPIA	RIA	FPIA	CEDIA	
Ammonia (NH_3)	–	–	–	–	NE[a]	NE	–	243–245
Ascorbic acid	–	–	–	–	FN[a]	FN	–	243, 244, 246
Baking soda ($NaHCO_3$)	–	NE	NE	NE	–	–	NE	65, 242, 246, 247
Bleach ($HClO_4$)	FN	FN	NE	FN	FN	FN	FN	65, 241–247
Blood	–	–	–	–	NE	FN	–	243, 244, 246
Dish soap	–	FN	FP[a]	FP	FP	FP	FN	65, 242–245
Drano ($NaOH–HClO_4$)	FN	–	–	–	NE	NE	FN	65, 241, 243, 244
Ethanol	–	NE	NE	NE	–	–	–	242
Ethylene glycol	–	NE	NE	NE	–	–	–	242
Glutaraldehyde	–	–	–	–	–	–	FN	65, 246, 248
Golden seal tea	FN	–	–	–	FN	FN	FN	65, 241, 243, 244, 246
Hand soap (liquid)	FN	–	–	–	FP	FP	–	241, 243, 244, 246
Hydrogen peroxide (H_2O_2)	–	NE	NE	NE	–	–	–	242
2-Propanol	–	NE	NE	NE	–	–	–	242, 245
Lemon juice	NE	–	–	–	NE	NE	FN	241, 243, 244, 246
Lime-A-Way	–	–	–	–	FP	NE	–	243, 244, 246
Methanol	–	–	–	–	NE	NE	–	243, 244, 246
Salt (NaCl)	FN	FN	NE	NE	FN	FN	NE	65, 241–247, 249
Sodium phosphate, tribasic	–	–	–	–	FP	NE	–	243, 244
Vanish	–	–	–	–	FN	NE	–	243, 244
Vinegar (acetic acid)	FN	–	–	–	NE	FN	FN	65, 241, 243, 244, 246, 247
Visine	FN	–	–	–	FN	FN	FN	65, 241, 243, 244, 246, 250

[a] NE, no effect; FN, false negative, caused urine fortified above cutoff to test below the cutoff; FP, false positive, caused urine with no drug present to test above the cutoff. Consult the specific reference to find concentrations of adulterants used that gave indicated result. In general, liquid adulterants were not studied at concentrations in excess of 10% (v/v).

This is considered an acceptable practice. For further consequences (e.g. termination of employment), resting upon the immunoassay result alone is not an acceptable practice. The value of the immunoassay in the forensic laboratory is as a screen. It provides a yes/no answer to the question of whether further testing is required. When a confirmed positive result arises, it provides evidence for qualitative detection by a second unique methodology.

Immunoassays provide concentration-dependent signals. What is the quantitative value of the immunoassay? As discussed under cross-reactivity, however, these signals can result from the detection of more than one analyte. An example of the selectivity of the immunoassay is given in Figure 9(a) and (b). Marijuana contains a number of cannabinoids that will appear in the urine after smoking or oral ingestion. While commercial immunoassays are designed to detect THCA, many of these cannabinoids are recognized by immunoassays. When a number of urine samples were measured by Abuscreen immunoassay in a quantitative mode and for THCA by GC/MS, a very poor correlation was found between the two assays (Figure 9a). Cocaine is metabolized predominantly to benzoylecgonine and ecgonine methyl ester. As ecgonine methyl ester results from hydrolysis of a bulky benzoyl group, it loses much of its structural identity to benzoylecgonine and is general poorly recognized by immunoassays designed to detect benzoylecgonine. The consequence of this is that cocaine metabolite immunoassays designed to detect benzoylecgonine are fairly specific (Figure 9b).

Two approaches to the use of immunoassay results have been taken. One, more commonly used in the forensic laboratory setting, is the determination of the need for confirmation on a diluted sample and setting the order of samples in a GC/MS batch to diminish carryover. In this manner, a sample with a high immunoassay result may be prepared using a full aliquot and a diluted (e.g. 1 : 10) aliquot, and nondiluted aliquots with high readings being grouped together in a GC/MS batch. This can save time and the need to repeat samples.

Quantitative immunoassays have been used in a number of pharmacokinetic studies. These must be interpreted carefully and with knowledge of the cross-reactivity of the antibody for metabolites of the primary analyte. In substance abuse treatment programs, the use of quantitative measurements to monitor increases and decreases in drug use has been proposed as a better measure of the efficacy of treatment.[261,262] This requires assays with relatively good selectivity for the analyte in question, as can be achieved with mass spectrometric

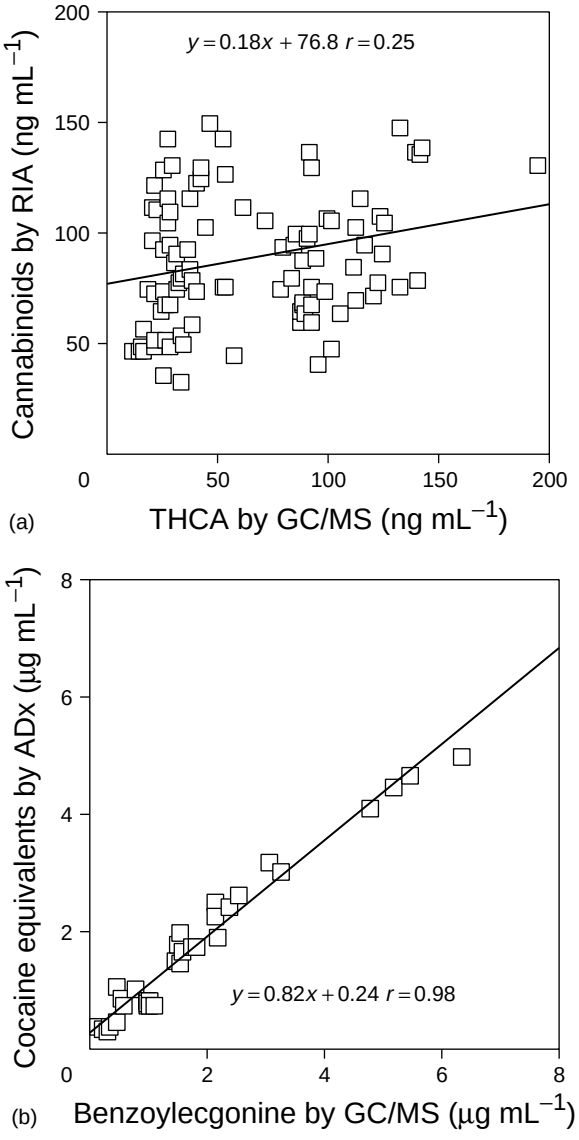

Figure 9 Correlation between quantitative immunoassay and GC/MS analysis between (a) the Abuscreen RIA for cannabinoids and GC/MS for THCA and (b) the ADx FPIA for cocaine metabolite and GC/MS for benzoylecgonine.

methods of analysis for most drugs or drug metabolites. This is also the case for immunoassays designed for the detection of cocaine metabolite in urine. Indeed, the accuracy and precision of immunoassays for the semiquantitative detection of benzoylecgonine (BE) have been demonstrated.[263–265]

Immunoassays have provided a powerful tool for the forensic toxicologist to screen large numbers of specimens for the presence of certain drug classes. When combined with a selective confirmational method such as GC/MS or LC/MS, they have become indispensible in the current laboratory setting. Immunoassays do, however, have limitations. They detect other analytes than the primary analyte, and they are subject to certain interferences. Furthermore, some drugs in the drug-class of an immunoassay may not have sufficient cross-reactivity at the concentration present in the sample to be detected (e.g. the potent benzodiazepines, some opiates). Interpretation of the immunoassay result must consider these limitations, particularly if the immunoassay result is used by itself or in a quantitative manner.

ACKNOWLEDGMENTS

Much of my education in the use of immunoassays must be attributed to my fellow directors at the Center for Human Toxicology, Douglas Rollins, Rodger Foltz, Dennis Crouch, Diana Wilkins and David Andrenyak. I must also acknowledge the input I have received from several colleagues at meetings that include the Society of Forensic Toxicologists, the California Association of Toxicologists and the International Association of Forensic Toxicologists. For any knowledge I impart with this article I thank them. Mistakes and misinterpretations are solely my responsibility. Supported in part by USPHS grant RO1DA 10100 and contract NO1DA-7-8074.

ABBREVIATIONS AND ACRONYMS

CEDIA	Cloned Enzyme Donor Immunoassay
cpm	counts per minute
DOD	Department of Defense
DOT	Department of Transportation
DPC	Diagnostic Products Corp.
DRI	Diagnostic Reagents Inc.
EIA	Enzyme Immunoassay
ELISA	Enzyme-linked Immunosorbent Assay
EMIT	Enzyme-multiplied Immunoassay Technique
FPIA	Fluorescent Polarization Immunoassay
FRAT	Free Radical Assay Technique
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/FID	Gas Chromatography/Flame Ionization Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GC/NPD	Gas Chromatography/Nitrogen–Phosphorus Detection
HI	Hemagglutination-inhibition
HPLC	High-performance Liquid Chromatography
IgG	Immunoglobulin G
KIMS	Kinetic Interaction of Microparticles in Solution

LAAM	*l*-α-Acetylmethadol
LC/MS	Liquid Chromatography/Mass Spectrometry
LLOQ	Lower Limit of Quantitation
LOD	Limit of Detection
LSD	Lysergic Acid Diethylamide
MDA	Methylenedioxyamphetamine
MDEA	Methylenedioxyethylamphetamine
MDMA	Methylenedioxymethamphetamine
NAD	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide (reduced form)
NRC	Nuclear Regulatory Administration
PBS	Phosphate-buffered Saline
PCP	Phencyclidine
RIA	Radioimmunoassay
RRA	Radioreceptor Assay
SAMHSA	Substance Abuse and Mental Health Service Administration
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SPE	Solid-phase Extraction
THCA	11-nor-Δ^9-Tetrahydrocannabinol-9-carboxylic Acid
TLC	Thin-layer Chromatography
TRIS	Tris (hydroxymethyl)aminomethane
ULOQ	Upper Limit of Quantitation

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of • Immunochemistry

Environment: Water and Waste **(Volume 3)**
Immunoassay Techniques in Environmental Analyses

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Mass Spectrometry for Forensic Applications

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis

REFERENCES

1. R.S. Yalow, S.A. Berson, 'Assay of Plasma Insulin in Human Subjects by Immunological Methods', *Nature (London)*, **184**, 1648–1649 (1959).
2. R.S. Yalow, S.A. Berson, 'Immunoassay of Endogenous Plasma Insulin in Man', *J. Clin. Invest.*, **60**, 1157–1175 (1960).
3. S. Spector, C.W. Parker, 'Morphine: Radioimmunoassay', *Science*, **168**, 1347–1348 (1970).
4. S. Spector, 'Quantitative Determination of Morphine in Serum by Radioimmunoassay', *J. Pharmacol. Exp. Ther.*, **178**, 253–258 (1971).
5. D. Catlin, R. Cleeland, E. Grunberg, 'A Sensitive, Rapid-Radioimmunoassay for Morphine and Immunologically Related Substances in Urine and Serum', *Clin. Chem.*, **19**, 216–220 (1973).
6. R.K. Leute, E.F. Ullman, A. Goldstein, L.A. Herzenberg, 'Spin Immunoassay Technique for Determination of Morphine', *Nature New Biol.*, **236**, 93–94 (1972).
7. R. Leute, E.F. Ullman, A. Goldstein, 'Spin Immunoassay of Opiate Narcotics in Urine and Saliva', *J. Am. Med. Assoc.*, **221**, 1231–1234 (1972).
8. F.L. Adler, C.-T. Liu, 'Detection of Morphine by Hemagglutination-Inhibition', *J. Immunol.*, **106**, 1684–1685 (1971).
9. F.L. Adler, C.-T. Liu, D.H. Catlin, 'Immunological Studies on Heroin Addiction I. Methodology and Application of a Hemagglutination Test for Detection of Morphine', *Clin. Immunol. Immunopathol.*, **1**, 53–68 (1972).
10. K.E. Rubenstein, R.S. Schneider, E.F. Ullman, ' ''Homogeneous'' Enzyme Immunoassay. A New Immunochemical Technique', *Biochem. Biophys. Res. Commun.*, **47**, 846–851 (1972).
11. R.S. Schneider, P. Lindquist, E.T.-I. Wong, K.E. Rubenstein, E.F. Ullman, 'Homogenous Enzyme Immunoassay for Opiates in Urine', *Clin. Chem.*, **19**, 821–825 (1973).
12. D.H. Catlin, 'Urine Testing: a Comparison of Five Current Methods for Detecting Morphine', *Am. J. Clin. Pathol.*, **60**, 719–728 (1973).
13. C.W. Gorodetzky, C.R. Angel, D.J. Beach, D.H. Catlin, S.-Y. Yeh, 'Validity of Screening Methods for Drugs of Abuse in Biological Fluids I. Heroin in Urine', *Clin. Pharmacol. Ther.*, **15**, 461–472 (1973).
14. S.J. Mule, M.L. Bastos, D. Jukofsky, 'Evaluation of Immunoassay Methods for Detection, in Urine, of Drugs Subject to Abuse', *Clin. Chem.*, **20**, 243–248 (1974).
15. G.L. Rowley, K.E. Rubenstein, J. Huisjen, E.F. Ullman, 'Mechanism by Which Antibodies Inhibit Hapten–Malate Dehydrogenase Conjugates. An Enzyme Immunoassay for Morphine', *J. Biol. Chem.*, **250**, 3759–3766 (1975).
16. D.R. Henderson, S.B. Friedman, J.D. Harris, W.B. Manning, M.A. Zoccoll, 'CEDIA™, a New Homogeneous Immunoassay System', *Clin. Chem.*, **32**, 1637–1641 (1986).
17. V.R. Spiehler, D. Reed, R.H. Cravey, W.P. Wilcox, R.F. Shaw, S. Holland, 'Comparison of Results for Quantitative Determination of Morphine by Radioimmunoassay, Enzyme Immunoassay and Spectrofluorometry', *J. Forensic Sci.*, **20**, 647–655 (1975).
18. E.P.J. van der Slooten, H.J. van der Helm, 'Comparison of the EMIT (Enzyme Multiplied Immunoassay Technique) Opiate Assay and a Gas-chromatographic–

Mass-spectrometric Determination of Morphine and Codeine in Urine', *Clin. Chem.*, **22**, 1110–1111 (1976).
19. S.J. Mule, D. Jukofsky, M. Kogan, A. De Pace, K. Verebey, 'Evaluation of the Radioimmunoassay for Benzoylecgonine (a Cocaine Metabolite) in Human Urine', *Clin. Chem.*, **23**, 796–801 (1977).
20. R. Rodgers, C.P. Crowl, W.E. Eimstad, M.W. Hu, J.K. Kam, R.C. Ronald, G.L. Rowley, E.F. Ullman, 'Homogeneous Enzyme Immunoassay for Cannabinoids in Urine', *Clin. Chem.*, **24**, 95–100 (1978).
21. J. Fenton, M. Schaffer, N.W. Chen, E.W. Bermes, Jr, 'A Comparison of Enzyme Immunoassay and Gas Chromatography/Mass Spectrometry in Forensic Toxicology', *J. Forensic Sci.*, **25**, 314–319 (1980).
22. J. Irving, B. Leeb, R.L. Foltz, C.E. Cook, J.T. Bursey, R.E. Willette, 'Evaluation of Immunoassays for Cannabinoids in Urine', *J. Anal. Toxicol.*, **8**, 192–196 (1984).
23. A.B. Jones, H.H. ElSohly, E.S. Arafat, M.A. ElSohly, 'Analysis of the Major Metabolite of Δ^9-Tetrahydrocannabinol in Urine. IV. A Comparison of Five Methods', *J. Anal. Toxicol.*, **8**, 249–251 (1984).
24. M.J. Kogan, J. Al Razi, D.J. Pierson, N.J. Willson, 'Confirmation of Syva Enzyme Multiple Immunoassay Technique (EMIT®) d.a.u. and Roche Abuscreen® Radioimmunoassay (RIA) (^{125}I) Urine Cannabinoid Immunoassays by Gas Chromatographic/Mass Spectrophotometric (GC/MS) and Bonded-phase Adsorption/Thin-layer Chromatographic (BPA-TLC) Methods', *J. Forensic Sci.*, **31**, 494–500 (1986).
25. M.L. Abercrombie, J.S. Jewell, 'Evaluation of EMIT® and RIA High Volume Test Procedures for THC Metabolites in Urine Utilizing GC/MS Confirmation', *J. Anal. Toxicol.*, **10**, 178–180 (1986).
26. L.J. McBurney, B.A. Bobbie, L.A. Sepp, 'GC/MS and EMIT Analysis for Δ^9-Tetrahydrocannabinol Metabolites in Plasma and Urine of Human Subjects', *J. Anal. Toxicol.*, **10**, 56–64 (1986).
27. D.L. Black, B.A. Goldberger, Y.H. Caplan, 'Enzyme Immunoassay Method for Comprehensive Drug Screening in Micro-samples of Urine', *Clin. Chem.*, **33**, 367–371 (1987).
28. A. Poklis, 'Evaluation of TDx Cocaine Metabolite Assay', *J. Anal. Toxicol.*, **11**, 228–230 (1987).
29. L. Karlsson, M. Strom, 'Laboratory Evaluation of the TDx Assay for Detection of Cannabinoids in Urine from Prison Inmates', *J. Anal. Toxicol.*, **12**, 319–321 (1988).
30. E.J. Cone, S.L. Menchen, B.D. Paul, L.D. Mell, J. Mitchell, 'Validity Testing of Commercial Urine Cocaine Metabolite Assays: I. Assay Detection Times, Individual Excretion Patterns, and Kinetics After Cocaine Administration to Humans', *J. Forensic Sci.*, **34**, 15–31 (1989).
31. E.J. Cone, J. Mitchell, 'Validity Testing of Commercial Urine Cocaine Metabolite Assays: II. Sensitivity, Specificity, Accuracy, and Confirmation by Gas Chromatography/Mass Spectrometry', *J. Forensic Sci.*, **34**, 32–45 (1989).
32. Y.H. Caplan, B. Levine, 'Abbott Phencyclidine and Barbiturates Abused Drug Assays: Evaluation and Comparison of ADx FPIA, TDx FPIA, EMIT, and GC/MS Methods', *J. Anal. Toxicol.*, **13**, 289–292 (1989).
33. D.J. Wells, M.T. Barnhill, Jr, 'Comparative Results with Five Cannabinoid Immunoassay Systems at the Screening Threshold of 100 µg/L', *Clin. Chem.*, **35**, 2241–2243 (1989).
34. P. Dextraze, W.C. Griffiths, P. Camara, L. Audette, M. Rosner, 'Comparison of Fluorescent Polarization Immunoassay, Enzyme Immunoassay, and Thin-layer Chromatography for Urine Cannabinoid Screening. Effects of Analyte Adsorption and Vigorous Mixing of Specimen on Detectability', *Ann. Clin. Lab. Sci.*, **19**, 133–138 (1989).
35. M.A. ElSohly, D.F. Stanford, 'Cutoff of 25 ng/mL for the EMIT d.a.u. Phencyclidine Assay', *J. Anal. Toxicol.*, **14**, 192–193 (1990).
36. R. Foltz, I. Sunshine, 'Comparison of a TLC Method with EMIT® and GC/MS for Detection of Cannabinoids in Urine', *J. Anal. Toxicol.*, **14**, 375–378 (1990).
37. F.M.L. Moore, D. Simpson, 'Detection of Benzoylecgonine (Cocaine Metabolite) in Urine: a Cost-effective Low Risk Immunoassay Procedure', *Med. Lab. Sci.*, **47**, 85–89 (1990).
38. J.C. Standefer, R.C. Backer, 'Drug Screening with EMIT Reagents: a Quantitative Approach to Quality Control', *Clin. Chem.*, **37**, 733–738 (1991).
39. D. Altunkaya, A.J. Clatworthy, R.N. Smith, I.J. Start, 'Urinary Cannabinoid Analysis: Comparison of Four Immunoassays with Gas Chromatography/Mass Spectrometry', *Forensic Sci. Int.*, **50**, 15–22 (1991).
40. M.L. Weaver, B.K. Gan, E. Allen, L.D. Baugh, F.-Y. Liao, R.H. Liu, J.G. Langner, A.S. Walia, L.F. Cook, 'Correlations of Radioimmunoassay, Fluorescence Polarization Immunoassay, and Enzyme Immunoassay of *Cannabis* Metabolites with Gas Chromatography/Mass Spectrometry Analysis of 11-nor-Δ^9-Tetrahydrocannabinol-9-carboxylic Acid in Urine Specimens', *Forensic Sci. Int.*, **49**, 43–56 (1991).
41. A.D. Fraser, W. Bryan, A.F. Isner, 'Urinary Screening for Midazolam with the Abbott ADx and TDx analyzers and the EMIT d.a.u. Benzodiazepine Assay with Confirmation by GC/MS'. *J. Anal. Toxicol.*, **15**, 8–12 (1991).
42. A.D. Fraser, W. Bryan, A.F. Isner, 'Urinary Screening for Alprazolam and its Major Metabolites by the Abbott ADx and TDx Analyzers with Confirmation by GC/MS', *J. Anal. Toxicol.*, **15**, 25–29 (1991).
43. M.A. Przekop, J.E. Manno, G.W. Kunsman, K.R. Cockerham, B.R. Manno, 'Evaluation of the Abbott ADx Amphetamine/Methamphetamine-II Abused Drug Assay: Comparison to TDx, EMIT, and GC/MS Methods', *J. Anal. Toxicol.*, **15**, 323–326 (1991).
44. O. Beck, P. Lafolie, P. Hjemdahl, S. Borg, G. Odelius, P. Wirbing, 'Detection of Benzodiazepine Intake in

Therapeutic Doses by Immunoanalysis of Urine: Two Techniques Evaluated and Modified for Improved Performance', *Clin. Chem.*, **38**, 271–275 (1992).
45. K.F. Buechler, S. Moi, B. Noar, D. McGrath, J. Villela, M. Clancy, A. Shenhav, A. Colleymore, G. Valkirs, T. Lee, J.F. Bruni, M. Walsh, R. Hoffman, F. Ahmuty, M. Nowakowski, J. Buechler, M. Mitchell, D. Boyd, N. Stiso, R. Anderson, 'Simultaneous Detection of Seven Drugs of Abuse By the Triage™ Panel for Drugs of Abuse', *Clin. Chem.*, **38**, 1678–1684 (1992).
46. P.L. Cary, C.A. Johnson, T.M. Folsom, W.R. Bales, 'Immunoassay Method Validation for a Modified EMIT Phencyclidine Assay', *J. Anal. Toxicol.*, **16**, 48–51 (1992).
47. E.J. Cone, S. Dickerson, B.D. Paul, J.M. Mitchell, 'Forensic Drug Testing for Opiates. IV. Analytical Sensitivity, Specificity, and Accuracy of Commercial Urine Opiate Immunoassays', *J. Anal. Toxicol.*, **16**, 72–78 (1992).
48. E.J. Cone, S. Dickerson, B.D. Paul, J.M. Mitchell, 'Forensic Drug Testing for Opiates. V. Urine Testing for Heroin, Morphine, and Codeine with Commercial Opiate Immunoassays', *J. Anal. Toxicol.*, **17**, 156–164 (1993).
49. T.C. Sneath, N.C. Jain, 'Evaluation of Phencyclidine by EMIT® d.a.u.™ Utilizing the ETS® Analyzer and a 25-ng/mL Cutoff', *J. Anal. Toxicol.*, **16**, 107–108 (1992).
50. G. Tso, A.H.B. Wu, D. Kuntz, S.S. Wong, 'Screening for Drugs of Abuse by Syva EMIT® II Reagents on the BM/Hitachi 717 Analyzer', *Res. Commun. Subst. Abuse*, **13**, 256–268 (1992).
51. A. Dasgupta, S. Saldana, G. Kinnaman, M. Smith, K. Johansen, 'Analytical Performance Evaluation of EMIT® II Monoclonal Amphetamine/Methamphetamine Assay: More Specificity than EMIT® d.a.u.™ Monoclonal Amphetamine/Methamphetamine Assay', *Clin. Chem.*, **39**, 104–108 (1993).
52. D.A. Armbruster, R.H. Schwarzhoff, E.C. Hubster, M.K. Liserio, 'Enzyme Immunoassay, Kinetic Microparticle Immunoassay, Radioimmunoassay, and Fluorescence Polarization Immunoassay Compared for Drugs-of-abuse Screening', *Clin. Chem.*, **39**, 2137–2146 (1993).
53. D.A. Armbruster, R.H. Schwarzhoff, B.L. Pierce, E.C. Hubster, 'Method Comparison of EMIT II and OnLine with RIA for Drug Screening', *J. Forensic Sci.*, **38**, 1326–1341 (1993).
54. A.H.B. Wu, S.S. Wong, K.G. Johnson, J. Callies, D.X. Shu, W.E. Dunn, S.H.Y. Wong, 'Evaluation of the Triage System for Emergency Drugs-of-abuse Testing in Urine', *J. Anal. Toxicol.*, **17**, 241–245 (1993).
55. R.C. Sreenivasam, T.C. Sneath, N.C. Jain, 'Evaluation of Emit® II Reagents on the Chem 1', *J. Anal. Toxicol.*, **17**, 370–373 (1993).
56. R.L. Fitzgerald, D.A. Rexin, D.A. Herold, 'Detecting Benzodiazepines: Immunoassay Compared with Negative Chemical Ionization Gas Chromatography/Mass Spectrometry', *Clin. Chem.*, **40**, 373–380 (1994).
57. D.A. Armbruster, R.H. Schwarzhoff, B.L. Pierce, E.C. Hubster, 'Method Comparison of EMIT 700 and EMIT II with RIA for Drug Screening', *J. Anal. Toxicol.*, **18**, 110–117 (1994).
58. B. Borggard, I. Joergensen, 'Urinary Screening for Benzodiazepines with Radioreceptor Assay: Comparison with EMIT® d.a.u. Benzodiazepine Assay and High-performance Liquid Chromatography', *J. Anal. Toxicol.*, **18**, 243–246 (1994).
59. L.E. Edinboro, A. Poklis, 'Detection of benzodiazepines and Tribenzazolams by TRIAGE™: Confirmation by Solid-phase Extraction Utilizing SPEC®•3ML•MP3 Microcolumns and GC/MS', *J. Anal. Toxicol.*, **18**, 312–316 (1994).
60. R. Meatherall, 'Benzodiazepine Screening Using EMIT II® and TDx®: Urine Hydrolysis Pretreatment Required', *J. Anal. Toxicol.*, **18**, 385–390 (1994).
61. D.A. Armbruster, E.C. Hubster, M.S. Kaufman, M.K. Ramon, 'Cloned Enzyme Donor Immunoassay (CEDIA) for Drugs-of-abuse Screening', *Clin. Chem.*, **41**, 92–98 (1995).
62. M. Hailer, Y. Glienke, I.-M. Schwab, L. von Meyer, 'Modification and Evaluation of Abuscreen OnLine Assays for Drug Metabolites in Urine Performed on COBAS FARA II in Comparison with EMIT d.a.u. Cannabinoid 20', *J. Anal. Toxicol.*, **19**, 99–103 (1995).
63. P. Kintz, D. Machart, C. Jamey, P. Mangin, 'Comparison Between GC/MS and the EMIT II, Abbott ADx, and Roche OnLine Immunoassays for the Determination of THCCOOH', *J. Anal. Toxicol.*, **19**, 304–306 (1995).
64. D.P. Baker, M.S. Murphy, P.F. Shepp, V.R. Royo, M.E. Caldarone, B. Escoto, S.J. Salamone, 'Evaluation of the Abuscreen OnLine Assay for Amphetamines on the Hitachi 737: Comparison with EMIT and GC/MS Methods', *J. Forensic Sci.*, **40**, 108–112 (1995).
65. A.H.B. Wu, E. Forte, G. Casella, K. Sun, G. Hemphill, R. Foery, H. Schanzenbach, 'CEDIA for Screening Drugs of Abuse in Urine and the Effect of Adulterants', *J. Forensic Sci.*, **40**, 614–618 (1995).
66. A.D. Fraser, R. Meatherall, 'Comparative Evaluation of Five Immunoassays for the Analysis of Alprazolam and Triazolam Metabolites in Urine: Effect of Lowering the Screening and GC/MS Cut-off Values', *J. Anal. Toxicol.*, **20**, 217–223 (1996).
67. F.P. Smith, C. Lora-Tamayo, R. Carvajal, B. Caddy, F. Tagliaro, 'Assessment of an Automated Immunoassay Based on Kinetic Interaction of Microparticles in Solution for Determination of Opiates and Cocaine Metabolite in Urine', *Ann. Clin. Biochem.*, **34**, 81–84 (1997).
68. R.L. Fitzgerald, D.A. Herold, 'Improved CEDIA® Benzodiazepine Assay Eliminates Sertraline Crossreactivity', *J. Anal. Toxicol.*, **21**, 32–35 (1997).
69. T. Korte, J. Pykalainen, P. Lillsunde, T. Seppala, 'Comparison of RapiTest® with Emit® d.a.u. and GC/MS for

the Analysis of Drugs in Urine', *J. Anal. Toxicol.*, **21**, 49–53 (1997).
70. O. Beck, Z. Lin, K. Brodin, S. Borg, P. Hjemdal, 'The OnLine Screening Technique for Urinary Benzodiazepines: Comparison with EMIT, FPIA, and GC/MS', *J. Anal. Toxicol.*, **21**, 554–557 (1997).
71. D.M. Obzansky, E.G. Gorman, S.P. Kramer, I.S. Masulli, E.A. Nuzzaci, W.F. Skogen, 'Development and Analytical Performance of an Automated Screening Method for Cannabinoids on the Dimension® Clinical Chemistry System', *J. Autom. Chem.*, **19**, 169–173 (1997).
72. R. Wennig, M.R. Moeller, J.M. Haguenoer, A. Marocchi, F. Zoppi, B.L. Smith, R. de la Torre, C.A. Carstensen, A. Goerlach-Graw, J. Schaeffler, R. Leinberger, 'Development and Evaluation of Immunochromatographic Rapid Tests for Screening of Cannabinoids, Cocaine, and Opiates in Urine', *J. Anal. Toxicol.*, **22**, 148–155 (1998).
73. D. Crouch, J. Frank, L. Farrell, H. Karsch, J. Klaunig, 'A Multi-site Laboratory Evaluation of Three On-site Urinalysis Drug-testing Devices', *J. Anal. Toxicol.*, **22**, 493–502 (1998).
74. A.G. Verstraete, S. Steyaert, 'Evaluation of the Diagnostic Performance of the Boehringer Mannheim CEDIA® LSD Assay', *J. Anal. Toxicol.*, **22**, 601–604 (1998).
75. A.H.B. Wu, Y.-J. Feng, A. Pajor, T.G. Gornet, S.S. Wong, E. Forte, J. Brown, 'Detection and Interpretation of Lysergic Acid Diethylamide Results by Immunoassay Screening of Urine in Various Testing Groups', *J. Anal. Toxicol.*, **21**, 181–184 (1997).
76. V. Haver, D. Black, J. Garbin, 'Evaluation of DRI Drug-of-abuse Reagents on the Hitachi 911', *J. Anal. Toxicol.*, **22**, 403–404 (1998).
77. S.R. Popelka, D.M. Miller, J.T. Holen, D.M. Kelso, 'Fluorescence Polarization Immunoassay II. Analyzer for Rapid, Precise Measurement of Fluorescence Polarization with Use of Disposable Cuvettes', *Clin. Chem.*, **27**, 1198–1201 (1981).
78. Y.H. Caplan, B. Levine, B. Goldberger, 'Fluorescence Polarization Immunoassay Evaluated for Screening for Amphetamine and Methamphetamine in Urine', *Clin. Chem.*, **33**, 1200–1202 (1987).
79. D.E. Moody, L.A. Reinersman, K.M. Monti, D.S. Barberio, A.M. Cairns, T.A. Jennison, 'A New TDx™ Procedure for Urine Cannabinoid Detection', in *Proceedings of the 24th Annual Scientific Meeting of the International Association of Forensic Toxicologists*, eds. G.R. Jones, P.P. Singer, Banff, Alberta, Canada, 151–159, 1988.
80. R. Meatherall, 'An Evaluation of the Abbott TDx® Urinary Cannabinoids Assay', *Can. Soc. Forensic Sci. J.*, **22**, 89–95 (1989).
81. A.D. Fraser, W. Bryan, A.F. Isner, 'Urinary Screening for α-OH Triazolam by FPIA and EIA with Confirmation by GC/MS', *J. Anal. Toxicol.*, **15**, 347–350 (1992).
82. E. Hanseler, H. Keller, 'Sensitivity and Specificity of Cocaine Metabolite Screening in View of the Analytical Performance of a Fluorescence Polarisation Immunoassay', *Eur. J. Clin. Chem. Clin. Biochem.*, **32**, 865–871 (1994).
83. J. Rohrich, K. Schmidt, H. Bratzke, 'Application of the Novel Immunoassay Triage™ to a Rapid Detection of Antemortem Drug Abuse', *J. Anal. Toxicol.*, **18**, 407–414 (1994).
84. H.R. Kransler, J. Stone, L. McLaughlin, 'Evaluation of a Point-of-care Testing Product for Drugs of Abuse: Testing Site is a Key Variable', *Drug Alcohol Depend.*, **40**, 55–62 (1995).
85. J.J.W. Ros, M.G. Pelders, A.C.C. Egberts, 'Performance of Abusign™ Drugs-of-abuse Slide Tests with Particular Emphasis on Concentrations Near the Cutoff: Comparison with FPIA–ADx and Confirmation of Results with GC/MS', *J. Anal. Toxicol.*, **22**, 40–44 (1998).
86. A.J. McNally, K. Goc-Szkutnicka, Z. Li, I. Pilcher, S. Polakowski, S.J. Salamone, 'An OnLine Immunoassay for LSD: Comparison with GC/MS and the Abuscreen® RIA', *J. Anal. Toxicol.*, **20**, 404–408 (1996).
87. S.J. Salamone, S. Honasoge, C. Brenner, A.J. McNally, J. Passarelli, K. Goc-Szkutnicka, R. Brenneison, M.A. ElSohly, S. Feng, 'Flunitrazepam Excretion Patterns Using the Abuscreen OnTrak and OnLine Immunoassays: Comparison with GC/MS', *J. Anal. Toxicol.*, **21**, 341–345 (1997).
88. D.J. Crouch, M.L. Cheever, D.M. Andrenyak, D.J. Kuntz, D. Loughmiller, 'A Comparison of ONTRAK TESTCUP™, Abuscreen ONTRAK™, Abuscreen ONLINE® and GC/MS Urinalysis Test Results', *J. Forensic Sci.*, **43**, 35–40 (1998).
89. S.J. Mule, E. Whitlock, D. Jukofsky, 'Radioimmunoassay of Drugs Subject to Abuse: Critical Evaluation of Urinary Morphine–Barbiturate, Morphine, Barbiturate and Amphetamine Assays', *Clin. Chem.*, **21**, 81–86 (1975).
90. P.S. Childs, H.H. McCurdy, 'Evaluation of a Radioimmunoassay (^{125}I) Kit for Cannabinoid Metabolites in Urine and Whole Blood', *J. Anal. Toxicol.*, **8**, 220–223 (1984).
91. T. Appel, N. Wade, 'Screening of Blood and Urine for Drugs of Abuse Utilizing Diagnostic Products Corporation's Coat-A-Count® Radioimmunoassay Kits', *J. Anal. Toxicol.*, **13**, 274–276 (1989).
92. D. Altunkaya, R.N. Smith, 'Evaluation of a Commercial Radioimmunoassay Kit for the Detection of Lysergide (LSD) in Serum, Whole Blood, Urine and Stomach Contents', *Forensic Sci. Int.*, **47**, 113–121 (1990).
93. N.P. Cassells, D.H. Craston, C.W. Hand, D. Baldwin, 'Development and Validation of a Nonisotopic Immunoassay for the Detection of LSD in Human Urine', *J. Anal. Toxicol.*, **20**, 409–415 (1996).
94. K.S. Webb, P.B. Baker, N.P. Cassells, J.M. Francis, D.E. Johnston, S.L. Lancaster, P.S. Minty, G.D. Reed, S.A.

White, 'The Analysis of Lysergide (LSD): the Development of Novel Enzyme Immunoassay and Immunoaffinity Extraction Procedures Together with an HPLC–MS Confirmation Procedure', *J. Forensic Sci.*, **41**, 938–946 (1996).
95. R.H. Schwartz, S. Bogema, M.M. Thorne, 'Evaluation of a Rapid Latex-particle Agglutination–Inhibition Screening Assay for Cocaine in Urine', *J. Pediatr.*, **117**, 670–672 (1990).
96. D.A. Armbruster, J.M. Krolak, 'Screening for Drugs of Abuse with Roche ONTRAK Assays', *J. Anal. Toxicol.*, **16**, 172–175 (1992).
97. E. Welch, L.E. Fleming, I. Peyser, W. Greenfield, B.W. Steele, E.S. Bandstra, 'Rapid Cocaine Screening of Urine in a Newborn Nursery', *J. Pediatr.*, **123**, 468–470 (1993).
98. D.J. Birnbach, D.J. Stein, A. Grunebaum, B.I. Danzer, D.M. Thys, 'Cocaine Screening of Parturients Without Prenatal Care: an Evaluation of a Rapid Screening Assay', *Anesth. Analg.*, **84**, 76–79 (1997).
99. B.J. Buchan, J.M. Walsh, P.E. Leaverton, 'Evaluation of Accuracy of On-site Multi-analyte Drug Testing Devices in the Determination of the Prevalence of Illicit Drugs in Drivers', *J. Forensic Sci.*, **43**, 395–399 (1998).
100. J. Towt, S.-C.J. Tsai, M.R. Hernandez, A.D. Klimov, C.V. Kravec, S.L. Rouse, H.S. Subuhi, B. Twarowska, S.J. Salamone, 'ONTRAK TESTCUP: a Novel, On-site, Multi-analyte Screen for the Detection of Abused Drugs', *J. Anal. Toxicol.*, **19**, 504–510 (1995).
101. E. Taylor, E. Oertli, J. Wolfgang, E. Mueller, 'Accuracy of Five On-site Immunoassay Drugs-of-abuse Testing Devices', *J. Anal. Toxicol.*, **23**, 119–124 (1999).
102. T.R. Koch, R.L. Raglin, S. Kirk, J.F. Bruni, 'Improved Screening for Benzodiazepine Metabolites in Urine Using the Triage™ Panel for Drugs of Abuse', *J. Anal. Toxicol.*, **18**, 168–172 (1994).
103. R. de la Torre, L. San, M. Torrens, 'Clinical Evaluation of the Triage® Analytical Device for Drugs-of-abuse Testing', *Clin. Chem.*, **42**, 1433–1438 (1996).
104. R.G. Parsons, R. Kowal, D. LeBlond, V.T. Yue, L. Neargarder, L. Bond, D. Garcia, D. Slater, P. Rogers, 'Multianalyte Assay System Developed for Drugs of Abuse', *Clin. Chem.*, **39**, 1899–1903 (1993).
105. A.J. Jenkins, W.D. Darwin, M.A. Huestis, E.J. Cone, J.M. Mitchell, 'Validity Testing of the AccuPINCH™ THC Test', *J. Ana. Toxicol.*, **19**, 5–12 (1995).
106. A.J. Jenkins, L.C. Mills, W.D. Darwin, M.A. Huestis, E.J. Cone, J.M. Mitchell, 'Validity Testing of EZ-SCREEN Cannabinoid Test', *J. Anal. Toxicol.*, **17**, 292–298 (1993).
107. O. Beck, P. Lafolie, G. Odelius, L.O. Boreus, 'Immunological Screening of Benzodiazepines in Urine: Improved Detection of Oxazepam Intake', *Toxicol. Lett.*, **52**, 7–14 (1990).
108. P. Simonsson, A. Liden, S. Lindberg, 'Effect of β-Glucuronidase on Urinary Benzodiazepine Concentrations Determined by Fluorescence Polarization Immunoassay', *Clin. Chem.*, **41**, 920–923 (1995).
109. R.C. Meatherall, A.D. Fraser, 'CEDIA dau Benzodiazepine Screening Assay: a Reformulation', *J. Anal. Toxicol.*, **22**, 270–273 (1998).
110. E.L. Slightom, 'The Analysis of Drugs in Blood, Bile, and Tissue with an Indirect Homogeneous Enzyme Immunoassay', *J. Forensic Sci.*, **23**, 292–303 (1978).
111. H.W. Peel, B.J. Perrigo, 'Detection of Cannabinoids in Blood Using EMIT', *J. Anal. Toxicol.*, **5**, 165–167 (1981).
112. B.J. Perrigo, B.P. Joynt, 'Optimization of the EMIT Immunoassay Procedure for the Analysis of Cannabinoids in Methanolic Blood Extracts', *J. Anal. Toxicol.*, **13**, 235–237 (1989).
113. V.R. Spiehler, P. Sedgwick, 'Radioimmunoassay Screening and GC/MS Confirmation of Whole Blood Samples for Drugs of Abuse', *J. Anal. Toxicol.*, **9**, 63–66 (1985).
114. W.M. Asselin, J.M. Leslie, B. McKinley, 'Direct Detection of Drugs of Abuse in Whole Hemolyzed Blood Using the EMIT d.a.u. Urine Assays', *J. Anal. Toxicol.*, **12**, 207–215 (1988).
115. W.M. Asselin, J.M. Leslie, 'Modification of Emit Assay Reagents for Improved Sensitivity and Cost Effectiveness in the Analysis of Hemolyzed Whole Blood', *J. Anal. Toxicol.*, **16**, 381–388 (1992).
116. C.W. Hand, R.A. Moore, J.W. Spear, 'Comparison of Whole Blood and Plasma Morphine', *J. Anal. Toxicol.*, **12**, 234–235 (1988).
117. L.J. Lewellen, H.H. McCurdy, 'A Novel Procedure for the Analysis of Drugs in Whole Blood by Homogeneous Enzyme Immunoassay (EMIT®)', *J. Anal. Toxicol.*, **12**, 260–264 (1988).
118. C.-W. Lee, H.-M. Lee, 'Evaluation of the Abbott TDx Analyzer', *J. Anal. Toxicol.*, **13**, 50–56 (1989).
119. L.M. Blum, R.A. Klinger, F. Rieders, 'Direct Automated EMIT® d.a.u. Analysis of *N*,*N*-Dimethylformamide-modified Serum, Plasma, and Postmortem Blood for Benzodiazepines, Benzoylecgonine, Cannabinoids, and Opiates', *J. Anal. Toxicol.*, **13**, 285–288 (1989).
120. H. Gjerde, A.S. Christophersen, B. Skuterud, K. Klemetsen, J. Morland, 'Screening for Drugs in Forensic Blood Samples Using EMIT® Urine Assays', *Forensic Sci. Int.*, **44**, 179–185 (1990).
121. H. Gjerde, 'Screening for Cannabinoids in Blood Using EMIT: Concentrations of Δ-9-Tetrahydrocannabinol in Relation to EMIT Results', *Forensic Sci. Int.*, **50**, 121–124 (1991).
122. T.F. Simonick, V.W. Watts, 'Preliminary Evaluation of the Abbott TDx for Screening of *d*-Methamphetamine in Whole Blood Specimens', *J. Anal. Toxicol.*, **16**, 115–118 (1992).
123. D.E. Moody, L.F. Rittenhouse, K.M. Monti, 'Analysis of Forensic Specimens for Cannabinoids. I. Comparison of RIA and GC/MS Analysis of Blood', *J. Anal. Toxicol.*, **16**, 297–301 (1992).

124. C. Collins, J. Muto, V. Spiehler, 'Whole Blood Deproteinization for Drug Screening Using Automatic Pipettors', *J. Anal. Toxicol.*, **16**, 340–342 (1992).
125. H.Y. Yee, J.D. Nelson, V.M. Papa, 'Measurement of Benzoylecgonine in Whole Blood Using the Abbott ADx® Analyzer', *J. Anal. Toxicol.*, **17**, 84–86 (1993).
126. D.T. Diosi, D.C. Harvey, 'Analysis of Whole Blood for Drugs of Abuse Using EMIT d.a.u. Reagents and a Monarch 1000 Chemistry Analyzer', *J. Anal. Toxicol.*, **17**, 133–137 (1993).
127. B.J. Perrigo, B.P. Joynt, 'Use of ELISA for the Detection of Common Drugs of Abuse in Forensic Whole Blood Samples', *Can. Soc. Forensic Sci. J.*, **28**, 261–269 (1995).
128. W. Huang, D.E. Moody, D.A. Andrenyak, D.E. Rollins, 'Immunoassay Detection of Nordiazepam, Triazolam, Lorazepam, and Alprazolam in Blood', *J. Anal. Toxicol.*, **17**, 365–369 (1993).
129. W. Huang, D.E. Moody, 'Immunoassay Detection of Benzodiazepines and Benzodiazepine Metabolites in Blood', *J. Anal. Toxicol.*, **19**, 333–342 (1995).
130. J.C. Cagle, H.H. McCurdy, Y.M. Pan, K.J. Ayton, W.H. Wall, E.T. Solomons, 'Evaluation of the CEDIA™ DAU Assays and the AxSym™ Systems for the Analysis of Cannabinoids in Whole Blood', *J. Anal. Toxicol.*, **21**, 213–217 (1997).
131. V.R. Spiehler, I.B. Collison, P.R. Sedgwick, S.L. Perez, S.D. Le, D.A. Farnin, 'Validation of an Automated Microplate Enzyme Immunoassay for Screening of Postmortem Blood for Drugs of Abuse', *J. Anal. Toxicol.*, **22**, 573–579 (1998).
132. S. Iwersen-Bergmann, A. Schmoldt, 'Direct Semiquantitative Screening of Drugs of Abuse in Serum and Whole Blood by Means of CEDIA® DAU Urine Immunoassays', *J. Anal. Toxicol.*, **23**, 247–256 (1999).
133. A. Poklis, S. Jortani, L.E. Edinboro, J.J. Saady, 'Direct Determination of Benzoylecgonine in Serum by EMIT® d.a.u. Cocaine Metabolite Immunoassay', *J. Anal. Toxicol.*, **18**, 419–422 (1994).
134. D.E. Moody, A.M. Medina, 'OnLine® Kinetic Microparticle Immunoassay of Cannabinoids, Morphine and Benzoylecgonine in Serum', *Clin. Chem.*, **41**, 1664–1665 (1995).
135. M.R. Harkey, G.L. Henderson, 'Hair Analysis for Drugs of Abuse', *Adv. Anal. Toxicol.*, **2**, 298–329 (1989).
136. W.A. Baumgartner, V. Hill, 'Hair Analysis for Organic Analytes: Methodology, Reliability Issues, and Field Studies', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 223–267, 1996.
137. P. Kintz, 'Clinical Applications of Hair Analysis', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 267–278, 1996.
138. H. Sachs, 'Forensic Applications of Hair Analysis', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 211–222, 1996.
139. D.E. Rollins, D.G. Wilkins, S.P. Gygi, M.H. Slawson, P.R. Nagasawa, 'Testing for Drugs of Abuse in Hair – Experimental Observations and Indications for Future Research', *Forensic Sci. Rev.*, **9**, 24–35 (1997).
140. E.J. Cone, R. Joseph, Jr, 'The Potential for Bias in Hair Testing for Drugs of Abuse', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 69–94, 1996.
141. M.A. Huestis, 'Technical and Legal Aspects of Drugs of Abuse Testing for Hair', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 5–16, 1996.
142. D.A. Kidwell, D.L. Blank, 'Environmental Exposure – the Stumbling Block of Hair Testing', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 17–68, 1996.
143. C.M. Selavka, F. Rieders, 'The Determination of Cocaine in Hair: a Review', *Forensic Sci. Int.*, **70**, 155–164 (1995).
144. C. Staub, 'Analytical Procedures for Determination of Opiates in Hair: a Review', *Forensic Sci. Int.*, **70**, 111–123 (1995).
145. M.R. Moeller, H. Eser, 'The Analytical Tools for Hair Testing', in *Drug Testing in Hair*, ed. P. Kintz, CRC Press, Boca Raton, FL, 95–120, 1996.
146. M. Chiarotti, S. Strano-Rossi, 'Preparation of Hair Samples for Drug Analysis', *Forensic Sci. Rev.*, **8**, 111–127 (1996).
147. A.M. Baumgartner, P.F. Jones, W.A. Baumgartner, C.T. Black, 'Radioimmunoassay of Hair for Determining Opiate-abuse Histories', *J. Nucl. Med.*, **20**, 748–752 (1979).
148. D. Valente, M. Cassini, M. Pigliapochi, G. Vansetti, 'Hair as the Sample in Assessing Morphine and Cocaine Addiction', *Clin. Chem.*, **27**, 1952–1953 (1981).
149. F.P. Smith, D.A. Pomposini, 'Detection of Phenobarbital in Bloodstains, Semen, Seminal Stains, Saliva, Saliva Stains, Perspiration Stains, and Hair', *J. Forensic Sci.*, **29**, 582–586 (1981).
150. W.A. Baumgartner, C.T. Black, P.F. Jones, W.H. Blahd, 'Radioimmunoassay of Cocaine in Hair: Concise Communication', *J. Nucl. Med.*, **23**, 790–792 (1982).
151. F.P. Smith, R.H. Liu, 'Detection of Cocaine Metabolite in Perspiration Stain, Menstrual Bloodstain and Hair', *J. Forensic Sci.*, **31**, 1269–1273 (1986).
152. S.A. Reuschel, F.P. Smith, 'Benzoylecgonine (Cocaine Metabolite) Detection in Hair Samples of Jail Detainees Using Radioimmunoassay and Gas Chromatography/Mass Spectrometry (GC/MS)', *J. Forensic Sci.*, **36**, 1179–1185 (1991).
153. A. Franceschin, L. Morosini, L. Dell'Anna, 'Detection of Morphine in Hair with the Abbott TDx', *Clin. Chem.*, **33**, 2125 (1987).
154. S. Balabanova, H. Brunner, R. Nowak, 'Radioimmunological Determination of Cocaine in Human Hair', *Z. Rechtsmed.*, **98**, 229–234 (1987).
155. C. Offidani, A. Carnevale, M. Chiarotti, 'Drugs in Hair: a New Extraction Procedure', *Forensic Sci. Int.*, **41**, 35–39 (1989).

156. H. Sachs, W. Arnold, 'Results of Comparative Determination of Morphine in Human Hair Using RIA and GC/MS', *J. Clin. Chem. Clin. Biochem.*, **27**, 873–877 (1989).
157. K. Puschel, P. Thomasch, W. Arnold, 'Opiate Levels in Hair', *Forensic Sci. Int.*, **21**, 181–186 (1993).
158. D. Fritch, Y. Groce, F. Rieders, 'Cocaine and Some of Its Products in Hair by RIA and GC/MS', *J. Anal. Toxicol.*, **16**, 112–114 (1992).
159. P. Kintz, B. Ludes, P. Mangin, 'Detection of Drugs in Human Hair Using Abbott ADx, with Confirmation by Gas Chromatography/Mass Spectrometry (GC/MS)', *J. Forensic Sci.*, **37**, 328–331 (1992).
160. A. Marsh, M.E. Carruthers, N. Desouza, M.B. Evans, 'An Investigation of the Effect of Washing Upon the Morphine Content of Hair Measured by a Radioimmunoassay Technique', *J. Pharm. Biomed. Anal.*, **10**, 89–93 (1992).
161. A. Marsh, M.B. Evans, 'Challenging Declarations of Abstinence by the Determination of Morphine in Hair by Radioimmunoassay', *J. Pharm. Biomed. Anal.*, **11**, 693–698 (1993).
162. A. Marsh, M.B. Evans, 'Radioimmunoassay of Drugs of Abuse in Hair. 1. Methadone in Human Hair, Method Adaptation and the Evaluation of Decontamination Procedures', *J. Pharm. Biomed. Anal.*, **12**, 1123–1130 (1994).
163. A. Marsh, M.B. Evans, J. Strang, 'Radioimmunoassay of Drugs of Abuse in Hair. 2. The Determination of Methadone in the Hair of Known Drug Users', *J. Pharm. Biomed. Anal.*, **13**, 829–839 (1995).
164. G. Koren, J. Klein, R. Forman, K. Graham, 'Hair Analysis of Cocaine: Differentiation Between Systemic Exposure and External Contamination', *J. Clin. Pharmacol.*, **32**, 671–675 (1992).
165. F. Ursitti, J. Klein, G. Koren, 'Clinical Utilization of the Neonatal Hair Test for Cocaine: a Four-year Experience in Toronto', *Biol. Neonate*, **72**, 345–351 (1997).
166. W.-L. Wang, E.J. Cone, J. Zacny, 'Immunoassay Evidence for Fentanyl in Hair of Surgery Patients', *Forensic Sci. Int.*, **61**, 65–72 (1993).
167. F. Martinez, T.S. Poet, R. Pillai, J. Erickson, A.L. Estrada, R.R. Watson, 'Cocaine Metabolite (Benzoylecgonine) in Hair and Urine of Drug Users', *J. Anal. Toxicol.*, **17**, 138–142 (1993).
168. C. Offidani, S. Strano-Rossi, M. Chiarotti, 'Improved Enzymatic Hydrolysis of Hair', *Forensic Sci. Int.*, **63**, 171–174 (1993).
169. C. Offidani, S. Strano Rossi, M. Chiarotti, 'Drug Distribution in the Head, Axillary and Pubic Hair of Chronic Addicts', *Forensic Sci. Int.*, **63**, 105–108 (1993).
170. M. Chiarotti, S. Strano-Rossi, C. Offidani, A. Fiori, 'Evaluation of Cocaine Use During Pregnancy Through Toxicological Analysis of Hair', *J. Anal. Toxicol.*, **20**, 555–558 (1996).
171. S. Strano-Rossi, A. Bermejo-Barrera, M. Chiarotti, 'Segmental Hair Analysis for Cocaine and Heroin Abuse Determination', *Forensic Sci. Int.*, **70**, 211–216 (1995).
172. J.F. Morrison, S.N. Chesler, J.L. Reins, 'Supercritical Fluid Extraction–Immunoassay for the Rapid Screening of Cocaine in Hair', *J. Microcol. Sep.*, **8**, 37–45 (1996).
173. J.F. Morrison, S.N. Chesler, W.J. Yoo, C.M. Selavka, 'Matrix and Modifier Effects in the Supercritical Fluid Extraction of Cocaine and Benzoylecgonine from Human Hair', *Anal. Chem.*, **70**, 163–172 (1998).
174. S. George, R.A. Braithwaite, 'The Measurement of Morphine in the Hair of Heroin Abusers', *Ann. Clin. Biochem.*, **34**, 375–383 (1997).
175. D. Lewis, C. Moore, P. Morrissey, J. Leikin, 'Determination of Drug Exposure Using Hair: Application to Child Protective Cases', *Forensic Sci. Int.*, **84**, 123–128 (1997).
176. F. Tagliaro, Z. De Battisti, G. Lubli, C. Neri, G. Manetto, M. Margio, 'Integrated Use of Hair Analysis to Investigate the Physical Fitness to Obtain the Driving License: a Casework Study', *Forensic Sci. Int.*, **84**, 129–135 (1997).
177. C.W. Gorodetzky, M.P. Kullberg, 'Validity of Screening for Drugs in Biological Fluids. II. Heroin in Plasma and Saliva', *Clin. Pharmacol. Ther.*, **15**, 579–587 (1974).
178. R. Lucek, R. Dixon, 'Chlordiazepoxide Concentrations in Saliva and Plasma Measured by Radioimmunoassay', *Res. Commun. Chem. Pathol. Pharmacol.*, **27**, 397–400 (1980).
179. H.W. Peel, B.J. Perrigo, N.Z. Mikhael, 'Detection of Drugs in Saliva of Impaired Drivers', *J. Forensic Sci.*, **29**, 185–189 (1984).
180. J. Ostrea, P.M. Parks, M.J. Brady, 'Rapid Isolation and Detection of Drugs in Meconium of Infants of Drug-dependent Mothers', *Clin. Chem.*, **34**, 2372–2373 (1988).
181. E.M. Ostrea, M.J. Brady, P.M. Parks, D.C. Asenio, A. Naluz, 'Drug Screening of Meconium in Infants of Drug-dependent Mothers: an Alternative to Urine Testing', *J. Pediatr.*, **115**, 474–477 (1989).
182. E.M. Ostrea, M. Brady, S. Gause, A.L. Raymundo, M. Stevens, 'Drug Screening of New-borns by Meconium Analysis: a Large-scale, Prospective, Epidemiological Study', *Pediatrics*, **89**, 107–113 (1992).
183. R. Needlman, B. Zuckerman, G.M. Anderson, M. Mirochnick, D.J. Cohen, 'Cerebrospinal Fluid Monoamine Precursors and Metabolites in Human Neonates Following in Utero Cocaine Exposure: a Preliminary Study', *Pediatrics*, **92**, 55–60 (1993).
184. E.M. Ostrea, A. Romero, D.K. Knapp, A.R. Ostrea, J.E. Lucena, R.B. Utarnachitt, 'Postmortem Drug Analysis of Meconium in Early-gestation Human Fetuses Exposed to Cocaine: Clinical Implications', *J. Pediatr.*, **124**, 477–479 (1994).

185. P. Nair, S. Rothblum, R. Hebel, 'Neonatal Outcome in Infants with Evidence of Exposure to Opiates, Cocaine, and Cannabinoids', *Clin. Pediatr.*, **33**, 280–285 (1994).
186. V. Delaney-Black, C. Covington, E. Ostrea, A. Romero, D. Baker, M. Tagle, B. Nordstrom-Klee, M.A. Silvestre, M.L. Angelilli, C. Hack, J. Long, 'Prenatal Cocaine and Neonatal Outcome: Evaluation of Dose–Response Relationship', *Pediatrics*, **98**, 735–739 (1996).
187. E.J. Cone, W.W. Weddington, Jr, 'Prolonged Occurrence of Cocaine in Human Saliva and Urine After Chronic Use', *J. Anal. Toxicol.*, **13**, 65–68 (1989).
188. S. Balabanova, E. Schneider, G. Buhler, H. Krause, 'The Presence of Methadone in Human Sweat', *Z. Rechtsmed.*, **103**, 323–327 (1990).
189. G.D. Clark, I.B. Rosenzweig, V.A. Raisys, C.M. Callahan, T.M. Grant, A.P. Streissguth, 'The Analysis of Cocaine and Benzoylecgonine in Meconium', *J. Anal. Toxicol.*, **16**, 261–263 (1992).
190. C.M. Callahan, T.M. Grant, P. Phipps, G. Clark, A.H. Novack, A.P. Streissguth, V.A. Raisys, 'Measurement of Gestational Cocaine Exposure: Sensitivity of Infants Hair, Meconium, and Urine', *J. Pediatr.*, **120**, 763–768 (1992).
191. W. Schramm, P.A. Craig, R.H. Smith, G.E. Berger, 'Cocaine and Benzoylecgonine in Saliva, Serum, and Urine', *Clin. Chem.*, **39**, 481–487 (1993).
192. L.O. Henderson, M.K. Powell, W.H. Hannon, B.B. Miller, M.L. Martin, R.L. Hanzlick, D. Vroon, W.R. Sexson, 'Radioimmunoassay Screening of Dried Blood Spot Materials for Benzoylecgonine', *J. Anal. Toxicol.*, **17**, 42–47 (1993).
193. E.M. Ostrea, A. Romero, H. Yee, 'Adaptation of Meconium Drug Test for Mass Screening', *J. Pediatr.*, **122**, 152–154 (1993).
194. P.H. Dickson, A. Lind, P. Studts, H.C. Nipper, M. Makoid, D. Therkildsen, 'The Routine Analysis of Breast Milk for Drugs of Abuse in a Clinical Toxicology Laboratory', *J. Forensic Sci.*, **39**, 207–214 (1994).
195. F. Moriya, K.M. Chan, T.T. Noguchi, P.Y.K. Wu, 'Testing for Drugs of Abuse in Meconium of Newborn Infants', *J. Anal. Toxicol.*, **18**, 41–45 (1994).
196. W.E. Wingert, M.S. Feldman, M.H. Kim, L. Noble, I. Hand, J.J. Yoon, 'A Comparison of Meconium, Maternal Urine and Neonatal Urine for Detection of Maternal Drug Use During Pregnancy', *J. Forensic Sci.*, **39**, 150–158 (1994).
197. M. Burns, R.C. Baselt, 'Monitoring Drug Use with a Sweat Patch: an Experiment with Cocaine', *J. Anal. Toxicol.*, **19**, 41–48 (1995).
198. D.E. Lewis, C.M. Moore, J.B. Leikin, A. Koller, 'Meconium Analysis for Cocaine: a Validation Study and Comparison with Paired Urine Analysis', *J. Anal. Toxicol.*, **19**, 148–150 (1995).
199. C. Moore, D. Lewis, J. Leikin, 'False-positive and False-negative Rates of Meconium Drug Testing', *Clin. Chem.*, **41**, 1614–1616 (1995).
200. C. Moore, D. Lewis, J. Becker, J. Leikin, 'The Determination of 11-nor-Δ^9-Tetrahydrocannabinol-9-carboxylic acid (THCCOOH) in Meconium', *J. Anal. Toxicol.*, **20**, 50–51 (1996).
201. V. Spiehler, J. Fay, R. Fogerson, D. Schoendorfer, R.S. Niedbala, 'Enzyme Immunoassay Validation for Qualitative Detection of Cocaine in Sweat', *Clin. Chem.*, **42**, 34–38 (1996).
202. J. Fay, R. Fogerson, D. Schoendorfer, R.S. Niedbala, V. Spiehler, 'Detection of Methamphetamine in Sweat by EIA and GC/MS', *J. Anal. Toxicol.*, **20**, 398–403 (1996).
203. D. Felscher, G. Gastmeier, J. Dressler, 'Screening of Pharmaceuticals and Drugs in Synovial Fluid of the Knee Joint and Vitreous Humor by Fluorescence Polarization Immunoassay (FPIA)', *J. Forensic Sci.*, **43**, 619–621 (1998).
204. M.A. ElSohly, S. Feng, 'Δ^9-THC Metabolites in Meconium: Identification of 11-OH-Δ^9-THC, 8β,11-diOH-Δ^9-THC, and 11-nor-Δ^9-THC-9-COOH as Major Metabolites of Δ^9-THC', *J. Anal. Toxicol.*, **22**, 329–335 (1998).
205. N.P. Lemos, R.A. Anderson, J.R. Robertson, 'Nail Analysis for Drugs of Abuse: Extraction and Determination of Cannabis in Fingernails by RIA and GC/MS', *J. Anal. Toxicol.*, **23**, 147–152 (1999).
206. W. Schramm, R.H. Smith, P.A. Craig, 'Drugs of Abuse in Saliva: a Review', *J. Anal. Toxicol.*, **16**, 1–9 (1992).
207. E.J. Cone, 'Saliva Testing for Drugs of Abuse', *Ann. N.Y. Acad. Sci.*, **694**, 91–127 (1993).
208. W.J. Jusko, R.L. Milsap, 'Pharmacokinetic Principles of Drug Distribution in Saliva', *Ann. N.Y. Acad. Sci.*, **694**, 36–47 (1993).
209. C. Moore, A. Negrusz, D. Lewis, 'Determination of Drugs of Abuse in Meconium', *J. Chromatogr. B*, **713**, 137–146 (1998).
210. S. Strano-Rossi, 'Methods Used to Detect Drug Abuse in Pregnancy: a Brief Review', *Drug Alcohol Depend.*, **53**, 257–271 (1999).
211. H.L. Johnson, H.I. Maibach, 'Drug Excretion in Human Sweat', *J. Invest. Dermatol.*, **56**, 182–188 (1971).
212. R. Aly, C. Shirley, B. Cunio, H.I. Maibach, 'Effect of Prolonged Occlusion on the Microbial Flora, pH, Carbon Dioxide and Transepidermal Water Loss on Human Skin', *J. Invest. Dermatol*, **71**, 378–381 (1978).
213. J.K. Taylor, *Quality Assurance of Chemical Measurements*, Lewis, Chelsea, MI, 1988.
214. M. Perez-Reyes, S. Di Guiseppi, A.P. Mason, K.H. Davis, 'Passive Inhalation of Marijuana Smoke and Urinary Excretion of Cannabinoids', *Clin. Pharmacol. Ther.*, **34**, 36–41 (1983).
215. J. Morland, A. Bugge, B. Skuterud, A. Steen, G.H. Wethe, T. Kjeldsen, 'Cannabinoids in Blood and Urine After Passive Inhalation of *Cannabis* Smoke', *J. Forensic Sci.*, **30**, 997–1002 (1985).
216. E.J. Cone, R.E. Johnson, 'Contact Highs and Urinary Cannabinoid Excretion After Passive Exposure to

Marijuana Smoke', *Clin. Pharmacol. Ther.*, **40**, 247–256 (1986).
217. S.J. Mule, P. Lomax, S.J. Gross, 'Active and Realistic Passive Marijuana Exposure Tested by Three Immunoassays and GC/MS in Urine', *J. Anal. Toxicol.*, **12**, 113–116 (1988).
218. L.W. Hayes, W.G. Krasselt, P.A. Mueggler, 'Concentrations of Morphine and Codeine in Serum and Urine After Ingestion of Poppy Seeds', *Clin. Chem.*, **33**, 806–808 (1987).
219. H.N. ElSohly, D.F. Stanford, A.B. Jones, M.A. ElSohly, H. Snyder, C. Pedersen, 'Gas Chromatographic/Mass Spectrometric Analysis of Morphine and Codeine in Human Urine of Poppy Seed Eaters', *J. Forensic Sci.*, **33**, 347–356 (1988).
220. K.D. Meneely, 'Poppy Seed Ingestion: the Oregon Perspective', *J. Forensic Sci.*, **37**, 1158–1162 (1992).
221. D.E. Moody, D.J. Crouch, R.P. Smith, C.W. Cresalia, P. Francom, D.G. Wilkins, D.E. Rollins, 'Drug and Alcohol Involvement in Railroad Accidents', *J. Forensic Sci.*, **36**, 1474–1484 (1991).
222. G.R. Jones, P.P. Singer, 'The Newer Benzodiazepines', in *Analytical Toxicology*, ed. R.C. Baselt, Year Book Medical Publishers, Chicago, IL, 1–69, Vol. 2, 1989.
223. A.D. Fraser, P. Howell, 'Oxaprozin Cross-reactivity in Three Commercial Immunoassays for Benzodiazepines in Urine', *J. Anal. Toxicol.*, **22**, 50–54 (1998).
224. T. Nishikawa, Y. Kamijo, H. Ohtani, A.D. Fraser, 'Oxaprozin Interference with Urinary Benzodiazepine Immunoassays and Noninterference with Receptor Assay', *J. Anal. Toxicol.*, **23**, 125–126 (1999).
225. K.S. Siff, A.E. Finkler, 'False-positive Barbiturate Test in Urine Owing to Phenytoin and 5-(*p*-Hydroxyphenyl)-5-phenyhydantoin', *Clin. Chem.*, **34**, 1359–1360 (1988).
226. W. Ruangyuttikarn, D.E. Moody, 'Comparison of Three Commercial Amphetamine Immunoassays for Detection of Methamphetamine, Methylenedioxyamphetamine, Methylenedioxymethamphetamine, and Methylenedioxyethylamphetamine', *J. Anal. Toxicol.*, **12**, 229–233 (1988).
227. D.E. Moody, W. Ruangyuttikarn, M.Y. Law, 'Quinidine Inhibits in Vivo Metabolism of Amphetamine in Rats: Impact upon Correlation Between GC/MS and Immunoassay Findings in Rat Urine', *J. Anal. Toxicol.*, **14**, 311–317 (1990).
228. J. D'Nicuola, R. Jones, B. Levine, M.L. Smith, 'Evaluation of 6 Commercial Amphetamine and Methamphetamine Immunoassays for Cross-reactivity to Phenylpropanolamine and Ephedrine in Urine', *J. Anal. Toxicol.*, **16**, 211–213 (1992).
229. J.T. Cody, R. Schwarzhoff, 'Fluorescence Polarization Immunoassay Detection of Amphetamine, Methamphetamine, and Illicit Amphetamine Analogues', *J. Anal. Toxicol.*, **17**, 26–30 (1993).
230. M.J. DeLaurentis, K. McNeil, A.J. Mann, S. Clark, H.M. Greenwood, 'An EMIT Assay for Cannabinoid Metabolites in Urine', *NIDA Res. Monogr.*, **42**, 69–85 (1982).
231. R.H. Liu, C. Edwards, L.D. Baugh, J.-L. Weng, M. Fyfe, A.S. Walia, 'Selection of Initial Test Cutoff Concentration for Workplace Drug Analysis – *Cannabis* Example', *J. Anal. Toxicol.*, **18** (1994).
232. J. Oyler, W.D. Darwin, K.L. Preston, P. Suess, E.J. Cone, 'Cocaine Disposition in Meconium from Newborns of Cocaine-abusing Mothers and Urine of Adult Drug Users', *J. Anal. Toxicol.*, **20**, 453–462 (1996).
233. M.L. Smith, R.O. Hughes, B. Levine, S. Dickerson, W.D. Darwin, E.J. Cone, 'Forensic Drug Testing for Opiates. VI. Urine Testing of Hydromorphone, Hydrocodone, Oxymorphone, and Oxycodone with Commercial Immunoassays and Gas Chromatography/Mass Spectrometry', *J. Anal. Toxicol.*, **19**, 18–26 (1995).
234. P.H. Trevilla, E.O. Jimenez, T. Tena, C.L. Tamayo, 'Presence of Rifampicin in Urine Causes Cross-reactivity with Opiates Using the KIMS Method', *J. Anal. Toxicol.*, **19**, 200 (1995).
235. H. van As, L.M.L. Stolk, 'Rifampicin Cross-reacts with Opiate Immunoassay', *J. Anal. Toxicol.*, **23**, 71 (1999).
236. A.D. Fraser, 'Urinary Screening for Alprazolam, Triazolam and Their Metabolites with the EMIT d.a.u. Benzodiazepine Metabolite Assay', *J. Anal. Toxicol.*, **11**, 263–266 (1987).
237. A.D. Fraser, W. Bryan, 'Evaluation of the Abbott TDx® Serum Benzodiazepine Immunoassay for the Analysis of Lorazepam, Adinazolam, and *N*-Desmethyladinazolam', *J. Anal. Toxicol.*, **19**, 281–284 (1995).
238. M. Cheever, G. Armendariz, D. Moody, 'Detection of Methadone, LAAM and Their Metabolites by Methadone Immunoassays', *J. Anal. Toxicol.*, **23**, 500–505 (1999).
239. W. Ruangyuttikarn, M.Y. Law, D.E. Rollins, D.E. Moody, 'Detection of Fentanyl and Its Analogs by Enzyme-linked Immunosorbent Assay', *J. Anal. Toxicol.*, **14**, 160–164 (1990).
240. The Medical Letter, 'LAAM – a Long-acting Methadone for Treatment of Heroin Addiction', *Med. Lett.*, **36**, 52 (1994).
241. S.L. Mikkelsen, K.O. Ash, 'Adulterants Causing False Negatives in Illicit Drug Testing', *Clin. Chem.*, **34**, 2333–2336 (1988).
242. A. Warner, 'Interference of Common Household Chemicals in Immunoassay Methods for Drugs of Abuse', *Clin. Chem.*, **35**, 648–651 (1989).
243. J.T. Cody, R.H. Schearzhoff, 'Impact of Adulterants on RIA Analysis of Urine for Drugs of Abuse', *J. Anal. Toxicol.*, **13**, 277–284 (1989).
244. R. Schwarhoff, J.T. Cody, 'The Effects of Adulterating Agents on FPIA Analysis of Urine for Drugs of Abuse', *J. Anal. Toxicol.*, **17**, 14–17 (1993).
245. W. Bronner, P. Nyman, D. von Minden, 'Detectability of Phencyclidine and 11-nor-Δ^9-Tetrahydrocannabinol-9-carboxylic Acid in Adulterated Urine by

Radioimmunoassay and Fluorescence Polarization Immunoassay', *J. Anal. Toxicol.*, **14**, 368–371 (1990).
246. N.P. Cassells, D.H. Craston, 'The Effects of Commonly Used Adulterants on the Detection of Spiked LSD by an Enzyme Immunoassay', *Sci. Justice*, **38**, 109–117 (1998).
247. F.P. Smith, S.A. Reuschel, K.C. Jenkins, 'OnLine™ Opiate Immunoassay Evaluation: Precision, Accuracy and Adulterants', *Sci. Justice*, **35**, 65–67 (1995).
248. S. George, R.A. Braithwaite, 'The Effect of Glutaraldehyde Adulteration of Urine Specimens on Syva EMIT II Drugs-of-abuse Assays', *J. Anal. Toxicol.*, **20**, 195–196 (1996).
249. O. Andersen, B. Eriksen, 'Specific Conductivity of Urine and Sensitivity of Enzyme Immunoassay Methods of Analysis for Drugs of Abuse', *Clin. Chem.*, **23**, 751–753 (1977).
250. S.D. Pearson, K.O. Ash, F.M. Urry, 'Mechanism of False-negative Urine Cannabinoid Immunoassay Screens by Visine™ Eyedrops', *Clin. Chem.*, **35**, 636–638 (1989).
251. J. Larsen, R. Fogerson, 'Nonsteroidal Anti-inflammatory Drug Interference in TDx® Assays for Abused Drugs', *Clin. Chem.*, **34**, 987–988 (1988).
252. D.E. Rollins, T.A. Jennison, G. Jones, 'Investigation of Interference by Nonsteroidal Anti-inflammatory Drugs in Urine Tests for Abused Drugs', *Clin. Chem.*, **36**, 602–606 (1990).
253. R. Joseph, S. Dickerson, R. Willis, D. Frankenfield, E.J. Cone, D.R. Smith, 'Interference by Nonsteroidal Anti-inflammatory Drugs in EMIT® and TDx® Assays for Drugs of Abuse', *J. Anal. Toxicol.*, **19**, 13–17 (1995).
254. G.C. Critchfield, D.G. Wilkins, D.L. Loughmiller, B.W. Davis, D.E. Rollins, 'Antibody-mediated Interference of a Homogeneous Immunoassay', *J. Anal. Toxicol.*, **17**, 69–72 (1993).
255. G. Sloop, M. Hall, G.T. Simmons, C.A. Robinson, 'False-positive Postmortem EMIT Drugs-of-abuse Assay Due to Lactate Dehydrogenase and Lactate in Urine', *J. Anal. Toxicol.*, **19**, 554–556 (1995).
256. L.M. Stolk, J.L. Scheijen, 'Urine Adulteration with Denture-cleaning Tablets', *J. Anal. Toxicol.*, **21**, 403 (1997).
257. M.A. ElSohly, S. Feng, W.J. Kopychi, T.P. Murphy, A.B. Jones, A. Davis, D. Carr, 'A Procedure to Overcome Interferences Caused by the Adulterant ''Klear'' in the GC/MS Analysis of 11-nor-Δ^9-THC-9-COOH', *J. Anal. Toxicol.*, **21**, 240–242 (1997).
258. F.M. Urry, G. Komaromy-Hiller, B. Staley, D.K. Crockett, M. Kushnir, G. Nelson, R.E. Struempler, 'Nitrite Adulteration of Workplace Urine Drug-testing Specimens I. Sources and Associated Concentrations of Nitrite in Urine and Distinction Between Natural Sources and Adulteration', *J. Anal. Toxicol.*, **22**, 89–95 (1998).
259. J. Singh, J.A. Elberling, D.G. Hemphill, J. Holmstrom, 'The Measurement of Nitrite in Adulterated Urine Samples by High-performance Ion Chromatography', *J. Anal. Toxicol.*, **23**, 137–140 (1999).
260. E.J. King, 'Performance of AdultaCheck 4 Test Strips for the Detection of Adulteration at the Point of Collection of Urine Specimens Used for Drugs-of-abuse Testing', *J. Anal. Toxicol.*, **23**, 72 (1999).
261. S.L. Batki, L.B. Manfredi, P. Jacob, R.T. Jones, 'Fluoxetine for Cocaine Dependence in Methadone Maintenance: Quantitative Plasma and Urine Cocaine/Benzoylecgonine Concentrations', *J. Clin. Psychopharmacol.*, **13**, 243–250 (1993).
262. S.-H. Li, C.N. Chiang, B.C. Tai, C.K. Marschke, R.L. Hawks, 'Quantitative Versus Qualitative Urinalysis for Benzoylecgonine in Clinical Trials for the Assessment of Cocaine Use', *Psychopharmacol. Bull.*, **31**, 671–679 (1995).
263. E.J. Cone, D. Yousefnejad, S.L. Dickerson, 'Validity Testing of Commercial Urine Cocaine Metabolite Assays: IV. Evaluation of the EMIT® d.a.u.™ Cocaine Metabolite Assay in a Quantitative Mode for Detection of Cocaine Metabolite', *J. Forensic Sci.*, **35**, 786–791 (1990).
264. R.D. Crosby, J.A. Halikas, L.D. Bowers, 'Quantitative Cocaine Urine Testing in Clinical Research Settings', *Ann. Clin. Physchiatry*, **3**, 243–249 (1991).
265. R.L. Foltz, C. Botelho, S.A. Reuschel, D.J. Kuntz, D.E. Moody, G.M. Bristow, 'Comparison of Immunoassays for Semiquantitative Measurement of Benzoylecgonine in Urine', *NIDA Res. Monogr.*, **175**, 287–302 (1997).

Ion Mobility Spectrometry in Forensic Science

Zeev Karpas
Nuclear Research Center, Negev, Israel

1 Introduction and History 1
1.1 Principle of Operation 2
1.2 Mobility and Reduced Mobility 2
1.3 History 3
2 Instrument Design 3
2.1 Sampling and Sample Introduction 4
2.2 Ionization: Sources and Chemistry 5
2.3 The Drift Tube 5
2.4 Signal Acquisition and Processing 5
3 Forensic Applications 6
3.1 Explosives 6
3.2 Illicit Drugs and Drug Abuse 9
3.3 Other Applications: Tear Gas, Arson, and Chemical Warfare Agents and Biological Materials 11
4 Ion Mobility Spectrometers in the Field and in the Laboratory 13
4.1 Operational Considerations 13
4.2 Sampling for Detection of Hidden Contraband Materials 14
4.3 Sample Preparation 14
4.4 Sensitivity and Limit of Detection 14
5 Related and Hyphenated Ion Mobility Spectrometry Techniques 15
5.1 Field Ion Spectrometry 15
5.2 Gas Chromatography Ion Mobility Spectrometer 16
5.3 Ion Mobility Spectrometer Mass Spectrometry 17
6 Merits, Limitations, and Future Trends 17
6.1 Merits and Limitations in Forensic Science 17
6.2 Future Developments 18
Acknowledgment 18
Abbreviations and Acronyms 19
Related Articles 19
References 19

The ion mobility spectrometer (IMS) is a device used primarily for detection, identification, and monitoring of trace amounts of gases or vapors. Generally, sample introduction, ionization, ion separation, and detection all take place at ambient pressure, so that elaborate vacuum systems are not required. IMSs are typically small (portable, hand-held or even pocket-sized) and respond within seconds to part per billion (ppb), or even part per trillion (ppt) concentration levels. The IMS is not a universal detector and is not suitable for all compounds. It has a limited dynamic range due to its finite reservoir of charge, and may suffer from matrix effects which could interfere with the ionization or ion separation processes and cause spectral interferences. These drawbacks have limited the application of IMS technology to a few fields that require high sensitivity and rapid response, such as the detection of chemical warfare agents (CWA), explosives, and illicit drugs, and the identification and monitoring of toxic gases or vapors.

The array of field and laboratory instruments available today have become useful tools in the prevention of illicit use of contraband substances and in the rapid identification of these and other toxic substances. This article describes the principle of operation of IMSs and the basic gas-phase ion chemistry underlying the technology. The proven experience, both in the field and in the laboratory, of IMSs for forensic science applications is also described, and the merits and shortcomings of present-day IMS technology, and future trends, are discussed.

1 INTRODUCTION AND HISTORY

The IMS device is used primarily for detection, identification and monitoring of trace amounts of gases or vapors. Its high sensitivity arises from the fact that the atmospheric pressure ionization (API) processes, that occur in the source region, efficiently transfer charge to the analyte molecules present in trace amounts in the sample and produce ions that are typical of the analyte. IMSs are particularly suitable for detection of compounds that have a high proton affinity (PA) and form stable positive ions, or for compounds that have a high electronegativity (EN) and readily form stable negative ions.

Almost since its advent in the early 1970s, the potential of plasma chromatography, as ion mobility spectrometry was then called, for forensic applications was realized by its inventors. Indeed, the first report on the technique deals with detection of personnel[1] and other early reports explored detection of explosives as well as narcotics.[2–4] However, as with other promising techniques, time was needed for the technology to mature, and for the array of field and laboratory instruments now available to become useful tools in the prevention of illicit use of contraband substances and in the rapid identification of these and other toxic substances. The

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

modern IMS contains the following basic features: a sampling and sample introduction system, an ionization source, a device to separate ions according to their mobility and to detect them and, finally, a signal-processing capability to derive the desired information, as shown in Figure 1.

A rather loose definition of forensic sciences is used throughout this article. In addition to the classical areas of detection of concealed explosives and illicit drugs, IMSs have been used to identify residues of explosives on suspects or in post-detonation debris, or illicit drugs and abuse of prescription drugs in clinical samples and on objects, IMSs have also found applications in identification of arson materials and tear-gas residues. Nerve agents, which after being used in a terrorist attack in a train station in Japan, and other highly toxic chemicals and biological materials, may now be part of a terrorist's arsenal and are no longer confined to the military battlefield.

To date, only two monographs devoted solely to IMSs have been published.[5,6] Several review articles have been published in English,[7–12] but only one was dedicated to forensic applications of the technique,[12] and recently a review of explosive detection by IMSs appeared as a chapter in a book on explosives.[11] A two-part comprehensive review on forensic applications of IMSs, authored by forensic experts, appeared recently in German.[13,14] Since the last review paper in English appeared progress has been made in this particular field, and the most rapidly growing commercial applications of IMS technology are in areas which may be categorized as forensic sciences.

The fact that most of the work reviewed here was published by instrument manufacturers and by law-enforcement agents that describe actual field experience with IMSs is indicative of the maturity of the technology.

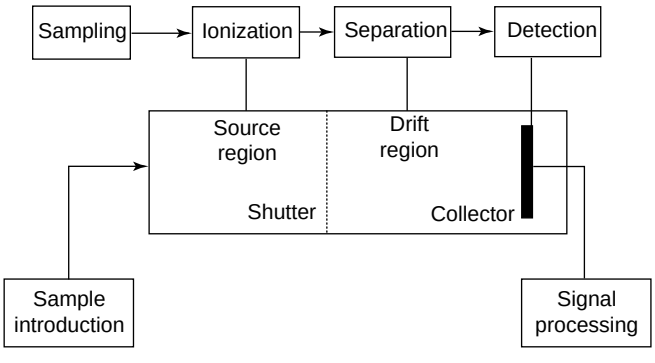

Figure 1 Block diagram of a generic IMS. Note the two major sections of the drift tube: the ionization (or source) region where ions are formed and the drift region where they are separated according to their mobilities.

1.1 Principle of Operation

The IMS utilizes the fact that a mixture of different ions may be separated into individual components as they travel at different velocities (mobilities) under the influence of an electric field through a tube containing a drift gas. The IMS may also be, somewhat naively, viewed as a variation on some more familiar techniques. On the one hand, it can be compared to the time-of-flight (TOF) mass spectrometer operating at an elevated pressure (usually ambient) with a very short flight (or drift) tube. On the other hand, the IMS may be compared to a gas chromatography (GC) apparatus with a virtual column (drift gas) as the stationary phase and may even be regarded as gas-phase electrophoresis device where the drift gas replaces the liquid.

The principles that govern the ionization and ion formation processes in the IMS are quite similar to those that dominate API and chemical ionization (CI) mass spectrometry (MS). The more specific gas-phase ion chemistry and ion–molecule reactions that pertain to explosives, drugs, and the other substances discussed in this article are reviewed in the relevant sections.

1.2 Mobility and Reduced Mobility

A comprehensive and detailed discussion of mobility theory, as it relates to ion mobility spectrometry, i.e. under conditions of ''low-field'', may be found in the fundamental paper by Revercomb and Mason[15] and in the two monographs mentioned above.[5,6] This section presents only a brief and basic explanation of mobility and reduced mobility that pertain to the IMS in practice.

When an ion moves under the influence of an electric field, through a bath gas, its drift velocity v_d is proportional to the strength of the field E (Equation 1):

$$v_d = KE \tag{1}$$

where the proportionality constant K represents the mobility of the ion. If the drift velocity v_d is substituted with the quotient d/t, where d is the distance the ion traverses in time t and Equation (1) is rearranged, the expression for the mobility becomes (Equation 2)

$$K = \frac{d}{Et} \tag{2}$$

When d is given in centimeters, E in volts per centimeter and t in seconds, the units of the mobility constant are $cm^2\,V^{-1}\,s^{-1}$. However, as the ambient conditions of temperature T and pressure P that directly affect the ion velocity may vary from one location to the next and between instruments, it is common practice to normalize these values to standard temperature (273 K) and pressure (760 torr) conditions and use the ''reduced

mobility'' K_o. This is a universal characterization of the ion, according to Equation (3):

$$K_o = \frac{d}{Et}\frac{273}{T}\frac{P}{760} \quad (3)$$

where T is expressed in degrees Kelvin and P in torr. The units of the reduced mobility are also $cm^2\,V^{-1}\,s^{-1}$.

When two ions, i and j, with different mobilities are formed and they drift along the same IMS tube, i.e. the same electric field, temperature and pressure affect both ions as they traverse the same distance, only the time it takes them to travel this length, t_i and t_j, would differ. As seen from Equation (3), the ratio between the drift times of these ions, t_i/t_j, would be inversely proportional to the reduced mobility ratio:

$$\frac{K_{o_j}}{K_{o_i}} = \frac{t_i}{t_j} \quad (4)$$

Equation (4) actually represents a functional approximation of the reduced mobility, which is widely used by practitioners of IMS technology. Usually, i would be an ion from a reference compound with a known reduced mobility and j the ion (or ions) derived from the analyte. It should be noted that none of the above equations include any information on the ion itself: its polarity, charge, mass, shape, internal charge distribution or structure. Nor is there any reference to the type of gas through which the ion drifts. The term for the mobility (or reduced mobility) should contain this information and reflect how all these parameters affect the motion of the ion. According to low-field mobility theory, the mobility of an ion may be expressed in Equation (5) as[15]

$$K = \frac{3e}{16N}\left(\frac{2\pi}{\mu k T_{\text{eff}}}\right)^{1/2}\frac{1+\alpha}{\Omega_D(T_{\text{eff}})} \quad (5)$$

where e is the ion charge, N is the density of the drift gas molecules, μ is the reduced mass of the ion–neutral collision pair $[mM/(m+M)]$ (where m is the ion mass and M is the mass of the drift gas molecules), k is the Boltzmann constant and T_{eff} the effective temperature of the ions. The parameter α is a correction factor (usually below 0.02 for $m > M$) and $\Omega_D(T_{\text{eff}})$ is the collision cross-section which depends on the effective ion temperature. This latter term represents the interactions between the ion and drift gas molecules, which depend on the ion size, shape, charge distribution, and the nature of the drift gas.

1.3 History

A detailed history of the evolution of modern ion mobility spectrometry is available.[6] In brief, ideas that originated at the turn of the twentieth century, and further developed half a century later, reached maturity and were presented in the form of a practical instrument in the late 1960s.[16] A considerable boost to the new technique was given by Karasek and co-workers who published an impressive array of articles that outlined the potential of the technology for solving practical problems, particularly for explosives[17] and narcotics[18] detection. It is noteworthy that the first publication involving an IMS dealt with a forensic application: the marking and detection of personnel.[1] The potential for forensic applications, mainly detection and identification of explosives and drugs, is still one of the main driving forces of on-going IMS development as reflected by the numerous papers presented at the annual IMS workshops,[19–24] and by the number of commercial manufacturers of IMSs.

2 INSTRUMENT DESIGN

A block diagram of the main components of the IMS is shown in Figure 1, and Figure 2 is a diagram of a commercial IMS (the Ionscan™, by Barringer). As can be seen in the photographs of IMSs in the following sections, the actual design may vary considerably from one instrument to the other. However, all IMSs must have a means of introducing the sample and of creating ions from vapors that emanate from the sample. The drift tube is the key component of the IMS, as this is where ions are separated according to their mobilities. Finally, a detector measures the ion current, the signal is processed, and the results are displayed according to the requirements of the particular application.

This basic scheme may be modified, as necessary. In the following sections the common designs for these components are described. It should be borne in mind that for most forensic applications the instrument should be able to detect and identify the substances of interest and provide the operator with ''hit'' results if contraband substances are found. This requires, first and foremost, a means of bringing the target analyte molecules to the IMS and introducing sufficient vapor into the drift tube to enable sensitive detection and true identification. Furthermore, identification of the target compound may be based on an algorithm for signal processing that uses a reference compound for calibration of the mobility scale; thus the results may only be displayed as a tabulated list, or the whole spectrum may be shown so that the operator has to make a judgment as to whether a target analyte was detected.

Most ionization sources produce both positive and negative ions. By selecting the direction of the electric field gradient in the source and along the drift tube, the polarity of ions that are detected may be chosen. Some IMSs are capable of measuring both polarities simultaneously (with two drift tubes) or sequentially (by periodically switching

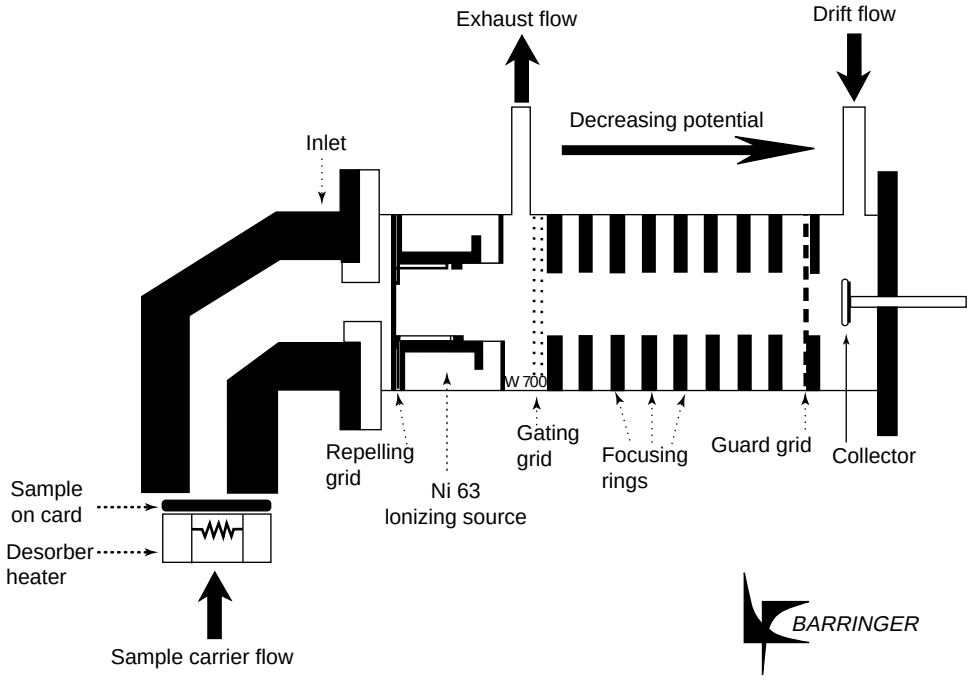

Figure 2 Diagram of the Ionscan™ (Barringer) IMS. From left to right: the desorber heater unit for introduction of vapors emanating from swipe samples through the inlet unit into the ionization region with the ^{63}Ni source, and then through the gating grid and along the drift region with its focusing rings and electric field gradient to the collector plate. (Reproduced with permission of Barringer.)

between the two polarities). The polarity of choice would depend on the application – generally negative polarity for explosive detection and positive polarity for drugs.

2.1 Sampling and Sample Introduction

For several applications, such as monitoring the concentration of vapors in air or for identification of a substance absorbed on a sample holder, a simple system for drawing ambient air into the instrument is quite suitable. However, detection of hidden contraband material requires special sampling techniques, in order to gather enough material from which to produce ions that will allow the unequivocal detection and identification of the target analytes. Some of these substances have exceptionally low vapor pressures, so that straightforward vapor-sniffing techniques would not be sensitive enough to detect them. Thus, in addition to direct vapor measurement, two sampling strategies have been adopted by instrument designers. When contraband detection is based on air sampling without direct contact with the suspect object or person, then a means of drawing a large volume of air and trapping particulate matter and vapors is of vital importance. Modified handheld or commercial vacuum cleaners, equipped with a removable filter for collecting particles, have been popular with some manufacturers and are widely applied in the field. Novel air-sampling systems with cyclones to separate particles according to their size have also been proposed.[25] Another approach relies on swabbing (or swiping) and therefore requires actual physical contact between the suspect material and the sampling device, and is recommended when there are no time constraints for taking a sample and access to the suspect object or person is possible. The US Coast Guard[26] has devoted time and effort to select the most suitable sample holder materials, and other law-enforcement agencies, such as the Drug Enforcement Administration (DEA)[27–30] and the Federal Bureau of Investigation (FBI),[31] have developed sampling procedures to ensure successful collection of particles and to minimize cross-contamination and interferences.

Both sampling strategies – vapor sniffing and particle collection – have been successfully applied, as demonstrated by the cases described in the following sections. Many IMSs also include an oven attachment, or heated inlet port, for vaporizing solid samples or for injection of liquid or gaseous samples, much like the injection port in GC.

Some IMSs have a membrane inlet that separates the ion source from the sampled ambient air, generally equipped with a hydrophobic membrane that rejects most of the moisture and allows the analyte molecules to permeate through it. This is advantageous when high levels of moisture are present, and also protects the IMS

from dust, lint, and other contaminants. The membrane inlet is especially utilized in IMSs that operate at near-ambient temperatures as it reduces the clustering of water vapors with the reactant and analyte ions. The disadvantages of the membrane inlet are a reduction in sensitivity, longer response times and sometimes more severe memory effects.

A special type of inlet system, combining GC in the front-end of the IMS, is discussed in section 5.2, and is meant to overcome one of the major limitations of the IMS that arise from competitive ionization in mixtures, where compounds with a high PA or EN will interfere with the formation of ions in compounds with a lower PA or EN.

2.2 Ionization: Sources and Chemistry

In the original IMSs the ion source contained a radionuclide, usually a β particle emitter such as ^{63}Ni or ^{3}H (tritium) or an α particle source like ^{241}Am. However, the regulatory obligations involved with using radioactive materials, and the costs inferred by adhering to them, created a driving force toward alternative nonradioactive sources. These include electrospray,[32] photoionization (either from a laser or an ultraviolet (UV) lamp),[33] corona discharge,[34–36] surface ionization,[37] and high-energy electron[38] sources. Each has its own merits and limitations, as discussed in more detail elsewhere.[6] To date, ^{63}Ni is still the most commonly deployed ionization source, due to its long life, low maintenance and the fact that it has no power requirements. A gating system made of two wire grids (or two interwoven wires) is required to introduce a discrete packet of ions into the drift section (analogous to syringe used for injection of a sample in GC or high-performance liquid chromatography (HPLC)) and allows for the separation of ions of different mobilities.

The chemical processes that first lead to the formation of reactant ions and subsequently to production of analyte ions are rather complex. They depend to some extent on the type of ionization source used, but mainly on the gases and vapors present in the source region. For example, in the IMS source, trinitrotoluene (TNT) vapor in a pure nitrogen atmosphere will simply capture an electron to form TNT^- ions (mass 227 Da), whereas in air, where the reactant ion is O_2^-, a proton will be abstracted to form $(TNT-H)^-$ ions (mass 226 Da).[39] The ion chemistry in the source may be, and often is, modified by doping the source with a substance (reagent) that changes the reactant ion. For example, addition of organochlorine compounds (such as CH_2Cl_2) would produce chloride ions, which could readily attach to analyte vapors (ethylene glycol dinitrate (EGDN) for example), and form $EGDN-Cl^-$ ions.[40] Doping of the ion source is advisable when the analyte of interest may be present in a complex matrix, but has a high PA or EN.

2.3 The Drift Tube

This is the central part of the IMS and the quality of the instrument's performance is determined mainly by this component. The chief function of this device is to separate ions of different types according to their mobilities and transport them from the gated ionization source to the detector, with as little loss and with as little peak broadening as possible. The conventional design relies on a direct current (DC) voltage gradient that creates an electric field along the drift tube and a purified gas stream (drift gas), usually at atmospheric pressure, that flows counter to the ion motion. The original design included a segmented drift tube design, where alternate conductor and insulator rings were joined together to form a pneumatically sealed tube and a voltage divider provides each conductor ring with a suitable voltage so that an electric field gradient is obtained. Simpler (and much cheaper) designs involve a series of thin conductor rings placed inside or outside a sealed tube or even, as proposed in the past, an insulator material (glass or ceramic) coated with a resistant material to form the field gradient.[41] Some of these designs are so simple that they can almost be made at home. A novel design, called field ion spectrometry (FIS) by its developers and discussed in section 5.1, operates with an alternating current (AC) voltage to create a nonuniform electric field.[42,43]

The drift tube in IMSs may be heated, which is usually the case for forensic applications, retained at room temperature (as with several hand-held models), or even cooled to subambient temperatures (mainly for research). Operating the drift tube at an elevated temperature is advantageous as it reduces clustering which may alter the drift times of reactant and analyte ions, thus causing difficulties in spectra processing; heating also shortens the clearance times after a sample was taken (which reduces memory effects). The electronic shutter (gate) is located at one end of the drift tube and the detector, usually a plate that serves as a Faraday cup, is at the other end (see Figure 1). The drift time of an ion is defined as the time it takes, under given operating conditions, to traverse the length of the drift tube and is, of course, instrument dependent. Therefore, it is customary to convert the measured drift time into more universal units, such as the mobility or reduced mobility, as explained above.

2.4 Signal Acquisition and Processing

A plot of the ion current at the detector as a function of the drift time (beginning when the gate opens) is called a mobility spectrum or, in the older terminology, a

plasmagram. A typical example, showing that the vapors emitted from a wipe sample contained heroin (the peak at 10.35 ms) as well as the reactant ion (at 8.4 ms) is shown in Figure 3.

As seen from the *x*-axis, a single spectrum is acquired typically in less than 25 ms. However, the spectrum obtained from a single scan can be quite noisy due to electronic interferences and noise caused by the shutter grid. Simply by averaging several spectra, a matter of a few seconds, the signal-to-noise ratio may be improved considerably. As the shutter gate is typically open for 0.2 ms and spectrum acquisition takes 20 ms or more, only a small fraction (about 1%) of the ions produced in the source leave the shutter grid and are actually measured. A different approach, based on Fourier transformation has been proposed,[44] its main advantage being that a larger fraction of the ions can be utilized. However, this promising approach requires a second shutter gate near the collector plate together with more complex electronics and signal processing, and therefore has not been adopted by IMS manufacturers. In the novel FIS instrument, this problem is also overcome, as described in section 5.1.

Modern IMSs deploy microprocessors or computers for advanced signal processing. Their output may be in the form of a list of detected compounds (as on the right-hand side of Figure 3) with, or without, weighted probabilities. They may also be accompanied by sounding of an alarm or flashing a light when a hit is scored. Other software packages may store all the spectra for off-line processing by a trained operator or for use as evidence in court. It is generally advisable to have at least one reference compound for calibration of the mobility scale, as well as means for quality assurance. Commercial IMSs usually define a hit after testing whether an ion peak is within a preset drift-time window (i.e. a time position with a defined error), whether the peak width and amplitude meet the design criteria, and whether other peaks also appear in the proper position (when more than one peak is expected).

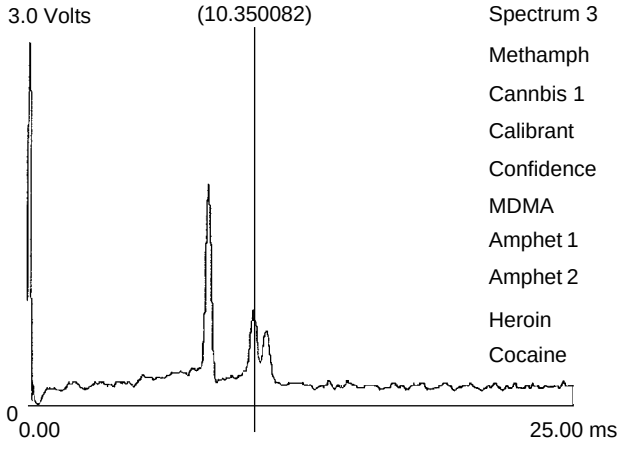

Figure 3 A mobility spectrum generated by a Narcotec™ (Graseby) IMS, from a wipe sample taken from an aluminum foil touched by a proxy in a simulated drug transport scenario. Note the clear presence of heroin (peak at 10.35 ms). Also note the list of compounds on the right. (Reproduced with permission of Graseby.)

3 FORENSIC APPLICATIONS

For the purpose of the present article, the forensic applications of IMSs have been divided into three categories. The two main applications concern the detection of explosives and of drugs (illicit and prescription), whereas all other forensic applications are discussed in the third subsection.

3.1 Explosives

IMS technology has been successfully applied to detect hidden explosive charges, for identification of explosives in cases when a suspect substance had to be identified, or when postdetonation traces and debris were collected in order to obtain preliminary identification in the field of the type of explosive used or for selection of evidence for laboratory tests. The IMS is also considered for detection of taggants, which are substances with relatively high vapor pressures added to explosives by some manufacturers in order to increase their detectability.[45,46]

Generally, when it is used for detection of explosives, the IMS is operated in the negative ion mode. In some scenarios the positive ion mode may be preferred, or a more definite identification may be achieved when both modes are used simultaneously or sequentially. Most common explosives contain nitro (i.e. NO_2) functional groups which render the molecule somewhat electronegative, and therefore negatively charged ions are readily formed in collisions with electrons or less stable negative ions. The sensitivity of the IMS for the detection of explosives is because a large part of the charge generated by the ion source may end up as negative ions derived from the explosive molecules that were introduced into the ion source. Some types of explosives that do not contain such electronegative groups present the forensic chemist with a special challenge, as they may not be detected by the procedures used for common explosives. This is especially relevant for instruments that base their detection solely on nitro groups, but IMS operators may also be required to take special measures for these substances.

There are three major mechanisms of negative ion formation from most explosive compounds, and which serve as the basis for detection and identification of explosives by IMS (Equation 6):

$$\text{charge transfer} \quad M + X^-(\text{or } e^-) \longrightarrow M^- + X \qquad (6a)$$

proton transfer $M + X^- \longrightarrow (M-H)^- + XH$ (6b)

association $M + X^- \longrightarrow MX^-$ (6c)

dimerization $MX^- + M \longrightarrow M_2X^-$ (6d)

where M is a molecule of the explosive and X^- a reactant anion. Charge transfer occurs when a negative reactant ion encounters an electronegative molecule, and an electron is transferred from this ion to the neutral molecule. Such a reaction channel will be dominant when the electron affinity of M is higher than that of X and other reaction mechanisms are less exothermic. Proton transfer, or proton abstraction, will take place when the analyte molecule has a labile proton and the negative ion is a conjugate base which could readily accommodate the proton, i.e. when the PA of X^- is higher than that of MH^-. The third mechanism, association or clustering, may occur when the explosive molecule has an electrophilic site for binding the negative ion. Most explosives, particularly aromatic compounds that contain electronegative substituents, are good candidates for such reaction, although other compounds may also readily associate with small ions. One of the most common explosives, TNT, undergoes electron attachment (which can be viewed as a type of charge transfer) when there is no oxygen (only an inert gas or nitrogen) in the ion source, proton abstraction if O_2^- ions from air are present, or association in the presence of chloride or bromide ions. Under certain conditions, several types of these ions may be formed simultaneously.

The product ions in Equation (6a–d) are all quasi-molecular ions, as they retain the parent molecule's structure almost intact. However, fragmentation processes that lead to creation of sub-molecular ions may also occur, especially if the product quasimolecular ion has an energy excess due to its exothermic formation process or due to the fact that the molecule or quasi-molecular ion is thermally unstable. Formation of NO_2^- or NO_3^- is quite common for explosives such as pentaerythritol tetranitrate (PETN). A comprehensive study of the complex fragmentation processes that occur in TNT, PETN, and hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) after zero-energy electron attachment emphasizes this point.[47] Although the experimental conditions in that study were not similar to those prevalent in the IMS, it should be noted that between five and 13 different ions were observed in these compounds. In the IMS, the fragment ions from the thermally unstable compounds tend to form cluster ions, including at times clustered dimer ions as shown in Equation (6d), such as RDX_2Cl^-, with concentration-dependent relative intensities.[11]

The improvement in sensitivity that can be obtained by doping the reactant ion has been quantitatively demonstrated.[48] The signal intensity from EGDN increased by factors of 7 and 4.7 with bromide and chloride reactant ions, respectively, and the taggant DMDNB (2,3-dimethyl-2,3-dinitrotoluene), which gave no signal in the undoped IMS, gave a detectable signal with both these dopant ions.

Naturally, the more types of ions that are formed from a given analyte, the more positive is the identification of the explosive. However, if the charge is distributed between several ions, the intensity of each may be too small to be detectable.

Some examples that illustrate applications of IMSs for explosives detection and identification are given below. First is the use of IMSs for the identification of explosives residues in post-detonation debris.[49] A Barringer Ionscan™ instrument was used to confirm and enhance results obtained by HPLC for samples collected at real detonation scenes. Debris with explosive residues were treated with acetone to perform solvent extraction: small samples in a Soxhlet system, medium-sized samples by sonification, and large samples by swabbing. The extracts were then filtered, reduced to a small volume, and dried under a stream of nitrogen. The dried samples were dissolved in acetonitrile and solid-phase extraction on a Florisil® column was performed, followed by HPLC analysis. The HPLC fractions were collected, dried and redissolved, and placed on a Teflon® filter and the filter was dried under a stream of dry air. Analysis by IMS was carried out according to the standard procedure of the Ionscan™ instrument and detection limits of a few hundred picograms were obtained. The authors concluded that the IMS results for several types of explosives compared well with those obtained by HPLC with a UV detector and that in some cases the IMS confirmed the presence of explosives at levels that were below the detection limit of HPLC.[49]

Application of IMS technology to the detection of concealed explosives on personnel was recently demonstrated.[50] The subject stands in a mock walk-through portal, while an overhead fan blows air over the subject's body (at approximately $2\,m\,s^{-1}$) so that explosives vapors and dislodged particles are carried through two vents ($160\,L\,s^{-1}$) located close to the feet and collected by a preconcentrator (a filter screen that traps heavy organic molecules). After the sampling cycle is finished, the screen is heated in a stream of air ($4\,L\,s^{-1}$) so that explosive vapors are desorbed into an IMS detector (phenylcyclidine (PCP)-110). The whole process takes about 12 s, but installation of two preconcentrators and two IMSs should shorten the time per person. The system is claimed to be sensitive enough to detect subfingerprint quantities of RDX deposited on cloth patches as well as

TNT vapors (and possibly more volatile explosives and taggants).[50]

Not all explosives contain nitro groups, a fact that may preclude their detection by techniques based on fluorescence from excited NO species. However, an undoped IMS (Graseby, chemical agent monitor (CAM)) can be used for some peroxide-based explosives, such as TATP (triacetone triperoxide) and its degradation products (namely DADP (diacetone diperoxide)), which have been involved in recent terrorist attacks.[51]

An interesting test of the capability of an IMS to detect explosives was carried out at Toronto International Airport.[52] Over 10 000 items, six of which were spiked with the explosive PETN, were sampled at a rate of 80–120 items per hour. Sampling was by direct vacuuming, swiping and swabbing. The false alarm rate was low (0.05%) and success in detection of all six spiked items sampled by swabbing was reported.[52] Figure 4 depicts a typical mobility spectrum and table of identified compounds obtained from the sample.

Table 1 presents a summary of the measured reduced mobilities and limits of detection (LOD) (where available) when using an IMS of some of the common explosives. In several cases, the operational conditions of the measurement are concisely given, such as the drift tube temperature, type of carrier gas and method

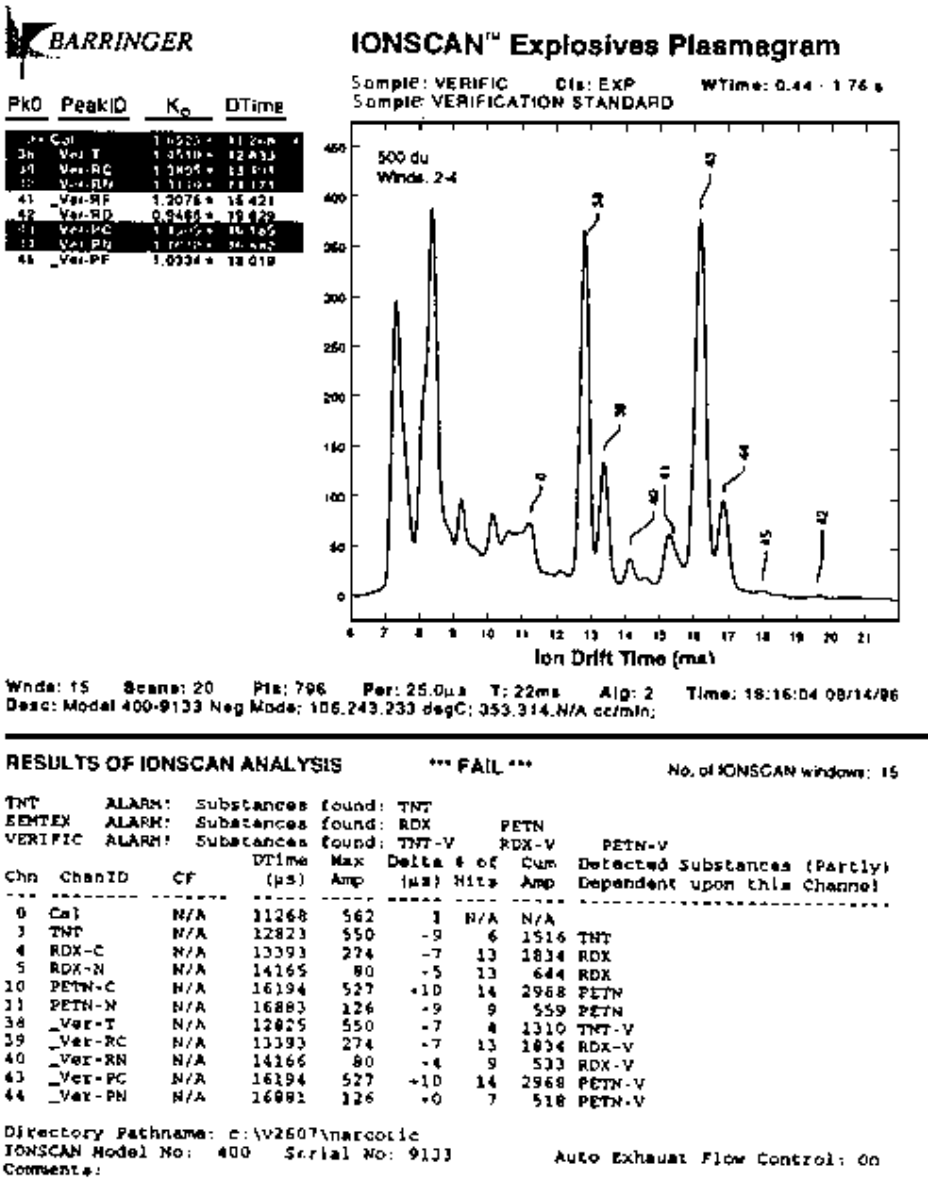

Figure 4 A mobility spectrum and listing of the identified compounds (explosives), generated by an Ionscan™ (Barringer). Note that several of the compounds have more than one ion species (peak). (Reproduced with permission of Barringer.)

Table 1 The reduced mobility and LOD of common explosives

Name	CAS number	LOD[a] (pg)	K_o (−) ($cm^2\ V^{-1}\ s^{-1}$)	Notes[b]
MNT	99-08-1		1.74, 1.81, 2.40	Air, 166, dilu.
2,4-DNT	121-14-2		1.68, 2.10	Air, 200, mem.
2,6-DNT	121-14-2		1.67	Air, 250, wire
TNT	118-96-7	1.45, 200	1.49, 1.54, 1.59	Air, 166, dilu.
Dynamite			2.10, 2.48	Air, 200, mem.
NG	55-63-0	1.34 (Cl^-), 50 1.28 (NO_3^-), 200	1.32	Air, 150, sniff
EGDN[c]	628-96-6		2.46	Air, 150, sniff
HMX	2691-41-0		1.30, 1.25	Air, 250, wire
RDX	121-82-4	1.39 (Cl^-), 200 1.31 (NO_3^-), 800 0.95 (RDX_2Cl^-), 1000	1.48	Air, 250, wire
PETN	78-11-5	1.21 (−H), 80 000 1.145 (Cl^-), 200 1.10 (NO_3^-), 1000	1.48	Air, 166
Tetryl	479-45-8		1.45, 1.62	Air, 250, wire
Comp B			1.57, 1.70, 1.81	Air, 200, mem.

[a] (−H) denotes the $(M-H)^-$ ion; (NO_3^-) denotes the $M \cdot NO_3^-$ association product ion; (Cl^-) denotes the $M \cdot Cl^-$ association.

[b] Air was used to carry the sample, for the ionization source and drift gas flow; 150, 166, 200 or 250 is the drift tube temperature in degrees centigrade; wire = insertion of a metal wire with absorbed vapors or solution; mem. = introduction of vapors through a membrane; dilu. = sample vapors introduced from a solid reservoir with a flow of dilution gas; sniff = means sampling ambient air by pumping the air directly into the instrument's inlet.

[c] Detection limit for EGDN vapors with Graseby Dynamic GVD6 with dibromomethane as dopant-0.3 ppb. (A. Brittain, personal communication).

MNT, mononitrotoluene; DNT, dinitrotoluene; NG, nitroglycerin; Comp B is a mixture of RDX and TNT; HMX, octahydro-1,3,5,7-tetranitro-1,3,5,7-tetraazine.

of sample introduction. The reduced mobility values shown in Table 1 are those that are judged by the author as being the best available values. These may differ somewhat between IMSs, according to the operational parameters, so that a calibration with a suitable reference material under the very same conditions is recommended. The tabulated LOD are based on reported values.

IMS technology for the detection of explosives is already in wide use for screening passengers and their luggage in airports, visitors and their bags in popular tourist sites (such as the Tower of London), and at restricted installations or security zones. However, screening large amounts of luggage and people is still time-consuming, and the slow procedures still cause queues to be formed during crowded periods. In addition, although the sensitivity of the IMS may be marginally satisfactory with detection limits in the subnanogram range, false readings (positive and negative) that arise from artifacts (mainly matrix effects that interfere with the detection or identification) demonstrate that this technology is not foolproof. The IMS must therefore be regarded as one of the powerful tools in the war against terrorism, but by no means the only tool.

3.2 Illicit Drugs and Drug Abuse

In this article, the term ''illicit drugs'' refers to substances that are outlawed and their possession is considered an offense (the quantity will determine how severe the offense is and the exact amount varies from country to country and state to state). The most common of these are the narcotics marijuana, cocaine and heroin, and the hallucinogens (psychedelic drugs) lysergic acid diethylamide (LSD) and psilocybin. As an offender will usually try to conceal the drugs the problem facing the forensic chemist is twofold: detection of hidden drugs, and obtaining identification once a suspect material is found. Drug abuse refers to prescription drugs, mainly barbiturates and amphetamines, that are used by people for nonmedical purposes (mood alteration).

Most of these drugs contain amine or amide groups, resulting in a relatively high PA, and therefore tend to efficiently produce positive ions, usually a protonated quasimolecular ion (Equation 7).

$$\text{proton transfer} \quad M + BH^+ \longrightarrow MH^+ + B \tag{7}$$

where M is the drug molecule and BH^+ a reactant ion with a labile proton. This proton transfer reaction will occur with high efficiency when the PA of M is higher

than that of B. Exothermic protonation may also lead to fragmentation of some of these quasimolecular ions and formation of smaller ions. Another process, frequently observed when the concentration of the analyte molecules is high, is the formation of protonated dimer ions, which can also be viewed as association of an analyte molecule with a protonated quasimolecular ion:

$$\text{association (dimerization)} \quad MH^+ + M \longrightarrow M_2H^+ \qquad (8)$$

Equation (11) occurs because a symmetric proton-bridged species is formed. In principle, association between two different molecules may also occur, such as MHB^+, but these ions are generally unstable because the proton would favor the species with the higher PA.

Some cases in which the IMS has been used to detect and identify drugs are discussed below. A scenario in which drugs are smuggled in a vehicle was simulated in order to obtain an estimate of typical levels that can be detected.[53] It was found that all hard drugs left detectable traces on surfaces in the vehicle and its driver, as well as on secondary surfaces touched by the driver (Figure 3). A Graseby Narcotec™ IMS (Figure 5), was used for detection of vapors that were generated by thermal desorption with a stream of hot air from wipe samples.

Typically, hard drugs (such as heroin, cocaine, ecstasy and amphetamines) were present in microgram quantities on primary contact surfaces, and therefore easily detected (above 10 ng per wipe). Cannabis detection depended on the nature of the material, due particularly to the weak transfer of THC (tetrahydrocannabinol, the active ingredient in hashish) to the wipe material.[53]

Another interesting scenario, using a Barringer Ionscan™, where it was necessary to detect drugs on board a ship has been described.[27–30] The effect of different aqueous media (tap water, deionized water, seawater, and sweat) on the detectability of cocaine was tested, with particular attention given to the temporal behavior of the desorption of drug vapors. It was found that the presence of water delays the release of cocaine vapors which may lead to a false negative response. This can be overcome by extending the desorption time to 15 s or by performing a second desorption cycle. The presence of large amounts of methanol had little effect on the maximum amplitude of the cocaine signal. In a later work, Su et al. tested the effect of the presence of oil on drug detection whose desorption was delayed. Once again, the oily sample was heated in a programmed temperature ramp, so that the initial release of the oil vapors did not interfere with the detection of the drug.[54]

A particularly innovative idea was to apply an IMS to the detection of drug residues in hair samples.[55–57] Alkaline digestion (5 M NaOH in a 4 : 1 methanol–water medium) was used to unbind methamphetamine, MDMA (the abusive drug ''Adam'' or ''Ecstasy'') and MDEA (the abusive drug ''Eve'') prior to vaporization and IMS analysis (both MDMA and MDEA are derivatives of MDA (3,4-methylenedioxyamphetamine)). The use of an internal standard made it possible to obtain quantitative results using an IMS, with good agreement with the laborious conventional gas chromatography/mass spectrometry (GCMS) method.[55,56] Analysis of 20 mm long sections of hair taken near the scalp and further from it can yield information on the drug abuse habits of the subject.[57] Washing the hair sample to remove external traces prior to analysis was necessary to obtain results that truly reflect the abuse habits. The reported detection limit for methamphetamine (MA) was 0.5 ng per milligram of hair.

In this article, the term ''drugs of abuse'' refers to substances that may be legally used, but either require prescriptions or are used in excess of the recommended amount. The IMS may be used to identify such substances

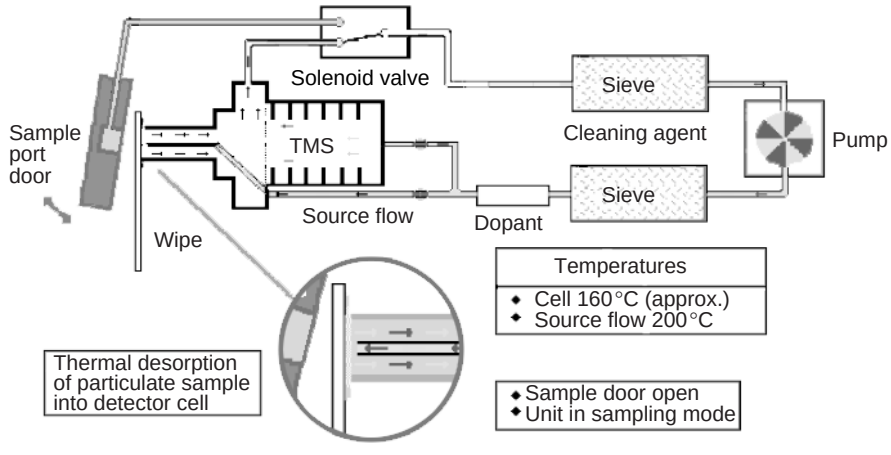

Figure 5 Diagram of the Narcotec™ (Graseby) IMS (the Plastec is quite similar). Note the thermal desorption unit for swipe samples and the option for introducing a dopant. (Reproduced with permission of Graseby.)

and help investigators and medical personnel determine what treatment is required. In these scenarios the emphasis would be on rapid positive identification of suspect substances rather than on detection of hidden contraband materials. For example, skin surface sampling and sniffing, in combination with IMS detection, were used to screen patients with suspected drug abuse and overdoses in an emergency room scenario.[58–60] The response of the drug psilocybin, used for mind alteration, has been recently characterized by the IMS[61] with good detection limits.

Table 2 presents a compilation of useful data required for identification of common illicit drugs and popular drugs of abuse, including the reduced mobility and LOD where available. Most of the reduced mobility values and LOD were taken from the reviews by Keller et al.;[13,14] other data were taken from the Barringer Ionscan™ tables and from the review by Karpas.[12]

The LOD for most drugs is on the order of 1 ng, which represents the minimum amount of substance on the sampler (swab or filter) required for obtaining a positive response under normal operating conditions. Depending on the laws of the country, the IMS could be used for screening only, and subsequent positive identification by other techniques (usually GCMS) may be required for judicial procedures.

3.3 Other Applications: Tear Gas, Arson, and Chemical Warfare Agents and Biological Materials

These forensic applications are grouped together either because the IMS is not yet as widely recognized as suitable for their detection or identification (arson and tear gas), or because they are not generally considered as classical forensic applications (chemical and biological warfare agents (BW)). In some cases, the realm of forensic science may not be clearly defined; for example, detecting spoiled fish may be considered as a pure health hygiene problem or may be regarded as prevention of a criminal act.

3.4.1 Tear Gas

''Tear gas'' is the generic name given to a group of lachrymatory chemical compounds. They are used as a nonlethal riot-control agent by law-enforcement authorities worldwide as well as for personal protection by individuals (although fatalities from severe exposure are known). Tear gas is increasingly being used in criminal acts as an offensive weapon. The IMS has been used to characterize the common tear gas sprays and a commercial Riot Agent Monitor™[62] is available which detects CN, CS, and BBC (bromobenzyl cyanide). A detailed study of CN, CS, and their isomers, and also the residues and degradation products found after their use, has been reported.[63] Positive mode and negative mode spectra were obtained. The detection limits for CN were 0.1 μg and 0.5 μg for the two modes, respectively, whereas CS was only identified in the negative mode at levels of 10 mg and 16 mg at ambient temperature and 50 °C, respectively. At higher temperatures (even 75 °C) analyte ion peaks were not observed in either mode. Thus, it was pointed out that the IMS can be used for screening of suspect spray cans and to identify the use of tear gas from residues on clothes or on swipe samples.[63]

3.4.2 Arson Cases

The IMS has been used to detect the presence of accelerants at the scene of arson crimes, and to help investigators locate the site where the fire was started and gather evidence for further, more definitive, laboratory testing.[64] This is just one example of where the portability of the IMS can be of help on the scene for screening debris and collecting meaningful samples which, after more analysis, may be presented in court.

3.4.3 Detection of Chemical Warfare Agents and Biological Materials

Although CWA detection is not usually considered as part of the forensic sciences, CWA use in terrorist attacks is already on record.[65] Another reason for the inclusion of CWA detection in this section is that the most widespread application of IMS technology is in this field – these instruments are the mainstay in several countries for military and civilian defense purposes. Hand-held IMSs such as that shown in Figure 6 are used to alert troops to dangerous levels of CWA and to monitor personnel, equipment and areas suspected of being contaminated after an attack. Pocket-size IMSs may be used by personnel in the battlefield for similar purposes (Figure 7).

Portable instruments are used by emergency response teams and other stationary IMS systems have been developed to monitor the air intake in armored vehicles, as well as in ships or in military and civilian installations and in populated areas.[66,67] This success of IMS technology results from its sensitivity toward these highly toxic chemicals (Table 3) and that specific IMS instruments are relatively inexpensive and easy to operate and maintain.

Unlike the detection of concealed drugs and explosives, where the objective is to try and detect a substance that is hidden while the perpetrators try to prevent vapors from being emanated, when monitoring for CWA it is necessary to detect only those substances that are present (substances are harmless if not present above a certain level). In other words, searching for explosives

Table 2 The reduced mobility and LOD of illicit drugs and drugs of abuse

Name	CAS number	LOD[a] (ng)	$K_o(+)$[b] ($cm^2\ V^{-1}\ s^{-1}$)	Notes[c]
Acetaminophen	103-90-2		1.70, 1.76, 1.97	Air, 220, skin/cart
Acetylcodeine	6703-27-1		1.09, 1.21	Air, 220, sol/wire
Alprazolan	28981-97-7		1.15	Air, 220, sol/wire
Amitryptiline	50-48-6		1.19	Air, 220, skin/cart
Amobarbital	57-43-2		1.36, 1.53	N_2, 230, sol/GC
Amphetamine	300-62-9	0.3	1.66, 1.65	Air, 220, sol/wire
Antipyrene	60-80-0	1.0		
Aprobarbital	77-02-1		1.39, 1.56, 1.75	N_2, 230, sol/GC
Barbital	57-44-3		0.99, 1.50	N_2, 230, sol/GC
Benzocain	94-09-7	1.0		
Benzoylecgonin	519-09-5	0.5		
Bromazepam	1812-30-2		1.24	Air, 220, sol
Butabarbital	125-40-6		1.28, 1.35	N_2, 230, sol/GC
Cannabinol	521-35-7		1.06	Air, 220, sol/wire
CDA	93-71-0		1.18	Air, 220, sol/wire
Cocaine	50-36-2	0.5	1.16, 1.14	Air, 220, sol/wire
Chlorazepam	57109-90-7	0.5		
Codeine	76-57-3	3.0	1.18, 1.21	Air, 220, sol/wire
Diazepam	439-14-5	0.5	1.21	Air, 220, sol/wire
EC		0.5		
Ephedrin	299-42-3	3.0		
Flurazepam	17617-23-1		1.03	Air, 220, sol/wire
Heroin	561-27-3	3	1.03, 1.04, 1.14	Air, 220, sol/wire
LSD	50-37-3	5.0	1.085, 1.065	
Lidocaine	137-58-6	1.0		
Lorazepam	846-49-1		1.19, 1.22	Air, 220, sol/cart
MDA		0.3	1.49, 1.48	Air, 220, sol/wire
MDEA	105-59-9	0.3	1.40	
MDMA	42542-10-9	0.3	1.47, 1.45	
Mescalin	54-04-6	5.0		
Methamphetamine	537-46-2	0.3	1.62, 1.63	Air, 220, sol/wire
Methaqualon	72-44-6	1.0		
Methadon	76-99-3	1.0	1.08	
Mephobarbital	115-38-8		1.63, 1.81	N_2, 230, sol/GC
Methyprylan	125-64-4		1.52	Air, 220, sol/wire
Morphine	57-27-2	3.0	1.22, 1.26	Air, 220, sol/wire
Nicotine	54-11-5		1.57	
Nitrazepam	146-22-5		1.22	Air, 220, sol/wire
OMAM			1.13, 1.26	Air, 220, sol/wire
Opium		5.0	1.55	Air, 250, wire
Oxazepam	604-75-1		1.23, 1.28	Air, 220, sol/cart
Pentobarbitol	76-74-4	3.0	1.38	N_2, 230, sol/GC
Phenobarbital	50-06-6		1.44	N_2, 230, sol/GC
PCP	77-10-1	1.0	1.27, 1.63, 2.01, 2.23	Air, 220, sol/wire
Procaine	59-46-1	1.0	1.31, 1.29	
Psilocybin	520-52-5	10.0	1.42	
Secobarbital	76-73-3	1.0	1.31, 1.48	N_2, 230, sol/GC
THC	33086-25-8	5.0	1.03, 1.05	Air, 220, sol/wire
Triazolam	28911-01-5		1.13	Air, 220, sol/wire

[a] Based on Keller et al.[13,14]

[b] Reduced mobility values based on Karpas review,[12] Barringer database, or Keller et al.[13,14] Some ion identifications were confirmed by Lawrence using IMS/MS. Most values are consitent within $0.01-0.02\ cm^2\ V^{-1}\ s^{-1}$, unless other ion species were observed.

[c] Air or N_2 = air or nitrogen were used to carry the sample, for the ionization source and drift gas flow; 150, 166, 200 or 250 is the drift tube temperature in degrees centigrade; wire = insertion of a metal wire with absorbed vapors or solution; cart = sample was introduced from skin onto a cartridge and then thermally desorbed; skin = sampling subject by swabbing skin surface; sol = introducing sample by evaporating a dried solution droplet. EC, ecgonine methylester.

Figure 6 A photograph of the CAM, the handheld IMS instrument widely used for detection of CWA (Graseby). Several modifications including a nonradioactive source and other applications are also on the market. (Reproduced with permission of Graseby.)

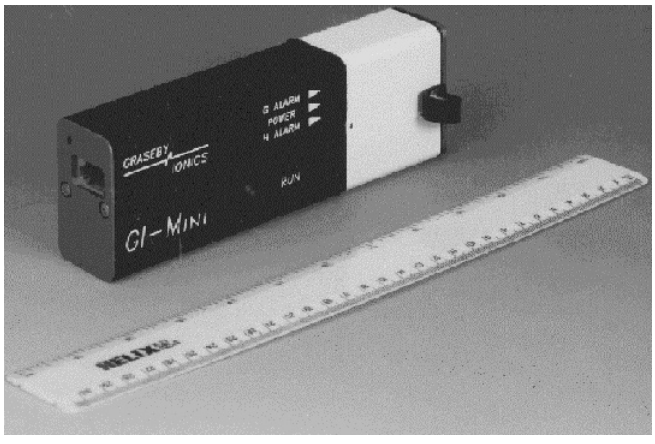

Figure 7 The GI-Mini™, a pocket-size IMS for detection of CWA. (Reproduced with permission of Graseby.)

is looking for something that supposedly ''isn't there'', whereas when looking for CWA the concern is about what ''is there''.

Detection of biological substances in general, and BW in particular, is a relatively new field for IMS technology.[68,69] Once again, use of BW or other toxins in forensic investigations is not very common, although their potential use in certain crime scenarios cannot be ignored. The response of the IMS to the presence of biological materials is sometimes an indirect one in that the action of the active biological materials releases a chemical substance, and that substance is detected by the IMS. More direct methods have been developed which detect biomarkers using a pyrolysis–GC/IMS (gas chromatograph ion mobility spectrometer) configuration.[70,71]

4 ION MOBILITY SPECTROMETERS IN THE FIELD AND IN THE LABORATORY

4.1 Operational Considerations

The type of IMS chosen would depend on the application and on operational considerations, with different types used in different scenarios. IMSs may be used as laboratory instruments for detection and identification of substances that were sampled elsewhere, either without pretreatment or after some chemical procedures have been carried out to preconcentrate the analyte and remove interferences. If a rapid response is required, as in the examination of luggage or the screening passengers entering an airport terminal, an IMS may be installed on-site. A number of fixed-point IMSs may be connected in a network in a railway station, for example, to monitor the presence of a chemical agent to counter terrorist attacks. If custom authorities or coast guard units need to examine cargo in a container, the IMS must be brought on-site and therefore hand-held devices are advantageous. Finally, if personnel need to be alerted to the presence of hazardous vapors or gases, pocket-size instruments may be used that

Table 3 Reduced mobilities and LOD of CWA

Name	CAS number	LOD ($mg\,L^{-1}$)	K_o (−)	K_o (+)	Reference
GA (Tabun)	77-81-6	5×10^{-6}	1.52, 1.76, 2.44	1.51, 1.06	Adler[67]
GB (Sarin)	107-44-8	5×10^{-6}		1.62, 1.22	Adler[67]
GD (Soman)	96-64-0	5×10^{-6}		1.51, 1.01	Adler[67]
VX	50782-69-9	5×10^{-6}		1.23	Adler[67]
HD (mustard)	505-60-2	2×10^{-5}	1.55, 2.40, 2.73		Adler[67]
HN	538-07-8	2×10^{-5}	1.33	1.47	Adler[67]
CG (phosgene)	75-44-5	2×10^{-3}	1.76, 1.90, 2.68	1.84	Adler[67]
HCN (AC)			2.44, 2.98		Adler[67]
ClCN (CK)			2.12, 2.44, 2.58		Adler[67]
G, V agents		$5\,\mu g\,m^{-3}$			RAID-16
H, L agents		$20\,\mu g\,m^{-3}$			RAID-16
HCN (AC)		$200\,\mu g\,m^{-3}$			RAID-16

RAID-16 is information taken from a brochure produced by Bruker Saxonia Analytik, Leipzig, Germany.

continually sample the ambient air and sound an alarm if a threshold is exceeded.

4.2 Sampling for Detection of Hidden Contraband Materials

A simplified statement of the main problem involved with detection of hidden contraband materials is that of ''bringing enough analyte molecules to the IMS''. Once the concentration of analyte vapors is high enough, positive identification may be achieved and detection obtained. The amount required for most common explosives or drugs is on the order of 10^{-8}–10^{-11} g, depending on the analyte, the matrix in which it is present, and the instrument.

The sampling system must be capable of collecting vapors that emanate from contraband materials that, on the one hand, have relatively high vapor pressures (like dynamite) and, on the other hand, should trap and concentrate microparticles from those materials that have low vapor pressures (as do most drugs and plastic explosives). These two demands are somewhat contradictory as sampling large volumes of air would bring more of these particles to the IMS but may also dilute the analyte vapors present near the suspect material with room air from other areas, thus reducing the vapor concentration to undetectable levels. In practical terms, an efficient sampling system must perform the following steps: detach analyte microparticles from the surface of the contraband packaging material or person, carry the particles and vapors to the preconcentrating device (a filter, for example) that allows the air to pass through it while trapping the analyte molecules, and then the trapped analyte particles must be easily vaporized and introduced into the IMS ionization region. Sampling may be performed with a modified vacuum cleaner equipped with a filter for trapping the microparticles and special portals for screening personnel or luggage have been developed for this purpose.

Several search methods, based on swabbing or swiping, rely on physical contact between the suspect material and the sampling device. A suitable sampling medium (cotton, cotton-wool, paper, etc.), that can be either dry or wetted with a solvent, is rubbed on the suspect package or hand and then placed in a heated sample introduction port (Figure 8). The particles taken up by the swabbing action are vaporized in a stream of carrier gas (usually air) that transports them into the IMS, where they are ionized and detected.

4.3 Sample Preparation

Detection of gases or vapors by the IMS can be done by direct sampling of ambient air, if the concentration is in the ppb to part per million (ppm) range, and in some cases even ppt concentration may yield a clear signal. In such cases, no sample preparation is required. However, when particulate matter is trapped on a filter, or swab and swipe samples are collected, the sampling medium has to be heated so that vapors enter the IMS. Thus, in several IMS models, particularly those used for forensic application, a small oven is coupled to the IMS inlet, and the sampling medium is inserted into this oven and sealed, so that a stream of purified air (or another gas) carries the vapors emanating from the sample into the ion source (Figure 5).

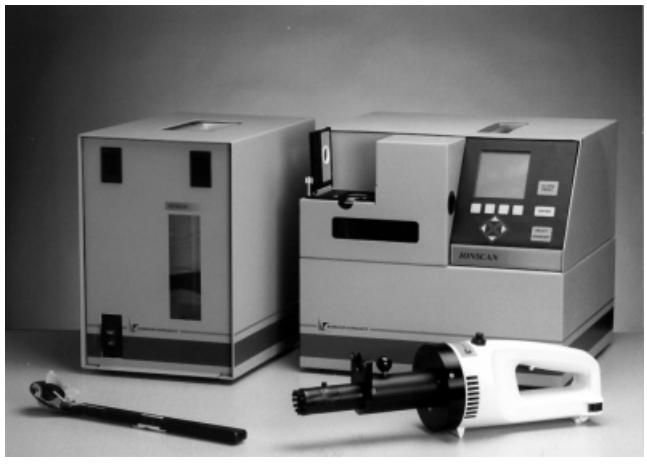

Figure 8 A photograph of the Ionscan™ 400 (Barringer) IMS and two sampling devices: a tool for collecting wipe samples (left foreground) and a handheld vacuum cleaner for collecting particles on a filter (right foreground). (Reproduced with permission of Barringer.)

If, for example, detection of drugs in hair samples or explosive residues is the application, then more elaborate procedures need to be carried out, as described in the previous sections. In such cases, the sample is usually treated with a suitable solvent, which may then be evaporated to a minimal volume and injected into the heated inlet in a fashion similar to GC samples. Very complex procedures are generally not performed for routine analysis by the IMS.

4.4 Sensitivity and Limit of Detection

The LOD of an IMS, like any other gas sensor, may be expressed in terms of the minimum detectable concentration of the analyte vapors in air, and is in the ppt to ppm range for most explosive and drug vapors. A more practical definition is that microparticles from nonvolatile substances are trapped and collected on a sampling medium, which is then introduced into the inlet oven. In this case the LOD is expressed in terms of the minimum detectable mass of analyte on the sample, which is in the picogram to nanogram range for most explosives and drugs. Tables 1–3 summarize selected

Table 4 Vapor concentration of some common drugs[12,13] and vapor pressure of explosives [Dionne et al. *J. Energ. Mater.*, **4**, 447–472 (1986)] at 20 °C

Substance	Vapor concentration	Vapor concentration
Cocaine	1.1×10^5 ppt	1.4 ng L^{-1}
Heroin	0.35 ppt	5.5 pg L^{-1}
Amphetamine	3.2×10^9 ppm	13.6 μg L^{-1}
TNT	1.3×10^{-3} Pa	
PETN	2.7×10^{-6} Pa	
RDX	8.5×10^{-7} Pa	

LOD, as reported by independent researchers and by some of the major manufacturers of IMSs in the forensic market. These summaries are based on published works and on the claims made in the most recent brochures and promotional literature available at the time of writing. Table 4 summarizes some of the values that represent the vapor pressure or equilibrium concentrations of common explosives and drugs.

5 RELATED AND HYPHENATED ION MOBILITY SPECTROMETRY TECHNIQUES

5.1 Field Ion Spectrometry

FIS is a technology in which mobilities of ions in an AC electric field are determined by scanning a voltage, instead of by measurement of drift times in a static DC electric field as in classical ion mobility spectrometry. Nevertheless, the two techniques are closely related as both depend on API processes for ion formation and on the fact that ions may be separated according to their different mobilities. FIS was initiated in the former USSR and was first presented in a paper as recently as 1993.[72] Commercial interest and support has given the technique a considerable boost. Although not yet in wide use, FIS is being evaluated by government and law enforcement agencies in several countries and, as forensic applications appear to be most promising, the technique warrants some discussion.

The principle of FIS is reminiscent of a quadrupole MS, i.e. the device operates by filtering ions as they travel from the ion source to the detector through parallel electrodes across which an oscillating electric field is applied. Figure 9 is a diagram of an improved FIS design, with separate sample and carrier streams to overcome the detrimental effect of uncontrolled moisture levels in the sample on the ion mobility. Ions are injected into the carrier stream in the space between the inner and outer electrodes, and travel along the ion filter until they reach the ion collector. The path along which an ion travels depends on its mass, size, shape and charge. By sweeping a second voltage, known as the compensating (or traverse field) voltage, different ions are allowed to reach the detector. The plot of ion current as function of the compensating voltage, called an ionogram, is equivalent to the mobility spectrum in conventional IMSs, as shown in Figure 10 for some explosives.

The enhanced sensitivity of FIS instruments arises mainly because its duty cycle is close to 100%, whereas the need for a shutter-gate in a conventional IMS allows only about 1% of the ions formed in the source to be detected.

The reported LOD for several chemical substances of relevance to forensic science (Table 5) are superior to those attainable by a conventional IMS. Furthermore, the controlability of the high-field conditions of FIS can be advantageous in reducing some of the interferences that hinder the conventional IMS, such as overlapping peaks

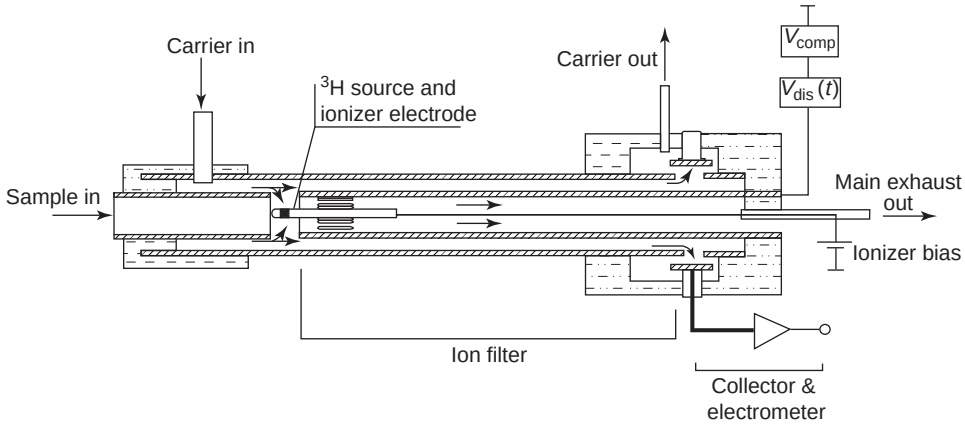

Figure 9 Diagram of the FIS, a novel mobility spectrometer. Note the fact that the ions have to traverse a path from the source to the collector between two concentric electrodes in an AC electric field, rather than the conventional DC field. (Reproduced with permission of MSA.)

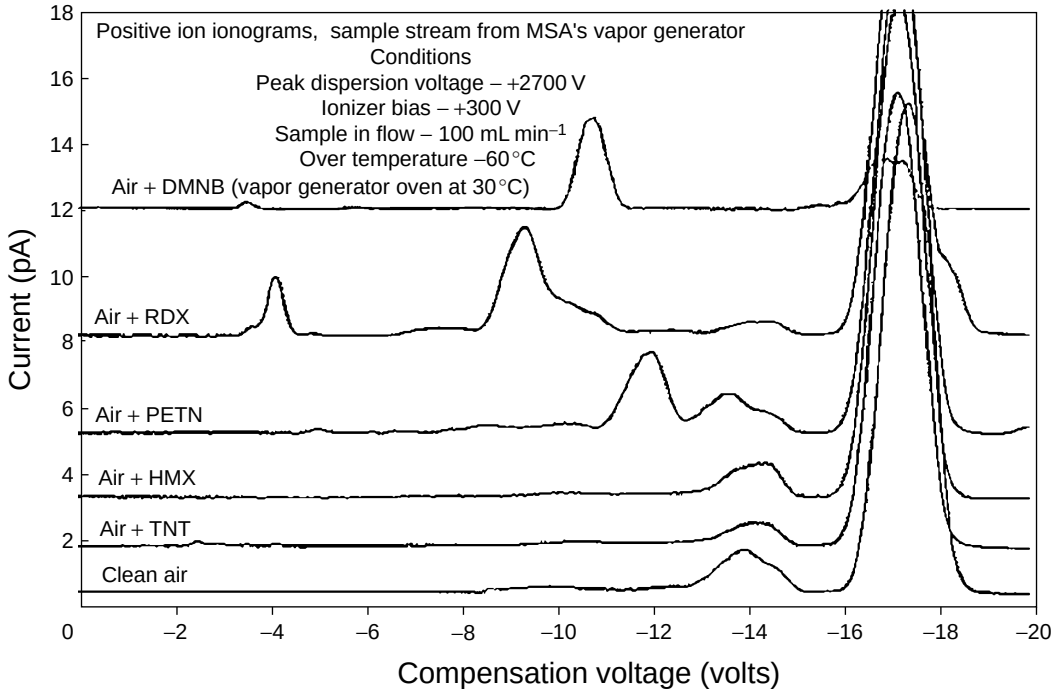

Figure 10 The positive ion mobility spectrum (ionogram) of DMNB, RDX, PETN, HMX and TNT, and the clean air spectrum, obtained using an FIS instrument. (Reproduced with permission of MSA.)

Table 5 The LOD for several compounds, as reported for an FIS instrument (manufactured by MSA, Pittsburgh)

Compound	CAS number	LOD (ppt by volume)	LOD (pg by weight)
Cocaine base	50-36-2	10	0.2
Heroin base	561-27-3	20	0.5
TNT	118-96-7	<80	<1
NG	55-63-0	20	–
RDX	628-96-6	25	0.4
PETN	78-11-5	1	0.02
TATP		60	–
DMNB		70	1
DMMP		10	0.1
GB (Sarin)	107-44-8	8	0.08
HD (Mustard)	505-60-2	620	7
HCN (AC)		$0.006\,mg\,m^{-3}$	
ClCN (CK)		$0.3\,mg\,m^{-3}$	
Cl_2CO (CG)	65-44-5	$4\,ppt; 2 \times 10^{-5}\,mg\,m^{-3}$	

DMMP, dimethyl methyl phosphonate.

resulting from ions with similar mobilities. However, skilful operation of FIS under optimal operating conditions, requires more training than the simple IMS, and it is more difficult to predict the mobilities of new chemicals, or to estimate the molecular weights of peaks that appear in the mobility spectrum of FIS. The technology relies on the high-field effect on the mobility of the ions, K, which may be expressed by: $K = K(0) \times \alpha \times E^2 + \cdots$, where α is another unkown and can be related to +ve, −ve, or a more complicated expression. Calibration of the high-field effect is required for each ion.

5.2 Gas Chromatography Ion Mobility Spectrometer

As mentioned above, one of the main limitations of the IMS arises from competitive ionization processes in

mixtures, which may hinder the formation of ions from some analytes, whereas the measured concentration of others may be distorted. In order to overcome this, a GC column may be used to separate the sample into discrete components before they enter the ionization source. As both the IMS and GC operate at ambient pressure, their coupling need not involve elaborate pumping systems. However, attention to some technical details is required so as to reduce memory effects. This mainly concerns proper positioning of the column-end in the IMS and use of a make-up gas stream to transport the analytes as they elute from the capillary GC column to the ionization source.

It is interesting to note that this idea first appeared when the IMS was introduced[73] as a laboratory GC/IMS/MS instrument. In addition to simplifying the mobility spectra, the GC/IMS affords two-dimensional information for positive identification of components in mixtures: first a retention time is derived from the GC data (in some instruments the IMS serves as a detector for GC by monitoring the decrease in the reactant ion peak, as an indication that a chemical has entered the ion source), and then mobility spectra may be examined and the peaks compared with a library of mobility spectra (just like GC/MS). The main drawback of this technology arises from the relatively long time required for separation on a GC column, therefore making it less suitable for field applications. Evolution of IMS and GC technologies, and developments in microelectronics have led to the introduction of hand-held fast GC/IMS instruments,[74] first for the military and later for civilian purposes. An example of the latter is the EVM™ (Environmental Vapor Monitor). A recent innovation combines a novel pneumatic sampling system that prevents saturation with the separation capability and speed of a high speed capillary GC column and the sensitivity of the IMS. This is known as the EVM-II or by the acronym AVS/TLGC/IMS (automatic vapor sampling/transfer line gas chromatography/ion mobility spectrometer).[75]

5.3 Ion Mobility Spectrometer Mass Spectrometry

The combination of IMS and MS involves differential pumping to overcome the pressure gradient between the ambient pressure IMS and the 10^{-5} torr operating pressure of MS, as well as a system of ion lenses to transport the ion cloud from the IMS detector region into the MS device. The only manufacturer of custom made IMS/MS systems is still the original inventor of modern IMS technology (PCP, Inc.), although some API/MS systems may be quite easily converted into IMS/MS. The IMS/MS is a typical laboratory system, with large power requirements and heavy equipment although, judging by recent trends in MS, portable IMS/MS is a realistic expectation. The main advantage of the IMS/MS system is that it affords positive identification of the ions formed in the IMS, and is vital for establishing a correlation between mobilities (ion drift times) and masses (required in some court cases). If forensic applications of the IMS are viewed mainly as a screening technique, then the IMS/MS data is the linchpin on which a court case may rest.

6 MERITS, LIMITATIONS, AND FUTURE TRENDS

6.1 Merits and Limitations in Forensic Science

The main advantages of the IMS for forensic applications arise from its sensitivity, portability, simple operation even by nonspecialists, and its price, all of which make IMS a powerful tool for screening. Its major merit is derived from the sensitivity of the method which requires as little as a nanogram of contraband substances for detection and identification purposes in a short time, with relatively few false-alarms. The fact that IMSs can easily be carried to the scene or can directly examine suspect personnel or objects, and that elaborate laboratory pretreatment is not usually required is another advantage for field applications. Little training is required for operators and the results may be presented in an unambiguous manner, which enables law-enforcement officers to obtain relevant evidence for further laboratory processing which may be used in court. Finally, the price of hand-held instruments, and even of the more sophisticated portable systems, and the minimal maintenance requirements are low enough for widespread use.

All these advantages are valid only when suitable applications are selected. Fortunately, detection of most explosives is relatively straightforward, due to the efficient production of negative ions from the analytes, and detection of the common drugs relies on the high PA and the tendency to form stable positive ions. However, the detection of some explosives and drugs requires special operational procedures, as they do not respond in the conventional mode of operation.

IMSs have some limitations which affect their applicability in some scenarios. For example, if quantitative information is necessary, the IMS with its limited dynamic range and problematic response to mixtures may not be able to provide the required data. Further limitations may be imposed by unfavorable operational conditions, such as high humidity or the presence of interfering substances at the sampling point. Naturally, some substances are not suitable for IMS detection owing to their chemical properties.

6.2 Future Developments

The future trends of IMS technology in general, and its application in forensic sciences, are already reflected in presentations and publications. Some of these improvements may require a sacrifice, or a trade-off, of some of the features that make the technology so appealing, namely price, size, simplicity of operation, and portability.

6.2.1 Nonradioactive Sources

The emergence of alternative ionization sources that do not require radioactive materials is evident from the number of papers presented in the most recent IMS workshops. Some of these sources may even be superior to the traditional ^{63}Ni in view of their higher ion production capability (an extended dynamic range and diminished saturation problems) as well as their flexibility in operational conditions which may lead to better selectivity and specificity. Furthermore, ionization of less-volatile compounds, particularly those of higher molecular weight, may be possible with devices that spray a solution with a dissolved sample, rather than vapors, into the IMS, like electrospray and ion spray sources. Other developments may include flame or pulsed-flame ionization sources that will be element specific, rather than compound specific.

6.2.2 Increasing Sensitivity

IMS technology is rapidly approaching picogram sensitivities (of the order of a billion molecules), which is typical of the most advanced explosive detectors of the present generation. Further improvements in sensitivity could arise from more efficient ionization and ion introduction processes (most gated IMSs use only 1% of the ions), more efficient drift tubes for better ion transmission, collection and detection, improved ion chemistry, and better signal acquisition and processing. For example, sophisticated strategies of ion collection (bunching), in analogy to the ion trap mass spectrometer, could yield great benefits. The idea is to thermally desorb the vapors from the sample, or to introduce vapors, for a few seconds, while collecting the ions of interest (rejecting other ions to avoid large space charges), before sending them in a short pulse, as a preconcentrated ion pack, down the drift tube. This works well under high-vacuum conditions; it may encounter difficulties, and possibly insurmountable practical problems, at atmospheric pressure. Improved detectors, similar to electron multipliers, that operate at near-ambient pressures could replace the traditional collector electrode. A noteworthy development is the use of novel drift tubes that deploy a dynamic AC electric field (rather than the traditional static DC field) and a reduced operating pressure (subambient), such as with the FIS instrument.

6.2.3 Improved Sampling Methods

The two key factors in detection of hidden contraband materials are collection of enough molecules (vapors or particles) and a sensitive detector. Direct sampling of vapors has a built-in paradox: high sampling rates are required to carry over enough vapor from across the room (or wherever the contraband material is) which will lead to dilution of the vapors to perhaps undetectable concentrations, whereas low sampling rates may not be sufficiently effective in bringing the vapor to the detector. Particle collection by swipe sampling may not always be possible, either because of time restrictions or because direct contact is not feasible. Therefore, better sample collection and preconcentration is the most promising avenue for enhanced detection capability. This could be done by using an approach similar to that deployed in the analysis of particulate matter in air, with an additional device for vapor collection, probably based on a liquid medium for dissolving the vapor and amassing the particles.[25]

6.2.4 Smaller Detectors

One evident trend is toward smaller, even personal or lapel-worn, IMS-based detectors and monitors. These could be carried around unobtrusively by security personnel in airports, customs, sensitive installations, and public areas to continuously sample the environment for presence of explosives, drugs and potentially hazardous chemicals. These detectors would have to be cheap and simple, which could be achieved with mass production techniques.

6.2.5 Explosives Not Containing Nitro Groups

Preliminary work indicates that the IMS can be used to identify explosives such as TATP, which some other techniques are incapable of doing. Further development is required, both to overcome the ever-changing chemistry resulting from the instability of such compounds, and to enhance the detection capability.

ACKNOWLEDGMENT

The assistance and information given by colleagues in the IMS field and by commercial companies that allowed their photographs to be used are greatfully acknowledged. Special thanks are due to Dr Alan Brittain for critically reviewing the manuscript and for his suggestions and comments.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
API	Atmospheric Pressure Ionization
AVS/TLGC/IMS	Automatic Vapor Sampling/ Transfer Line Gas Chromatography/Ion Mobility Spectrometer
BBC	Bromobenzyl Cyanide
BW	Biological Warfare Agents
CAM	Chemical Agent Monitor
CI	Chemical Ionization
CWA	Chemical Warfare Agents
DADP	Diacetone Diperoxide
DC	Direct Current
DEA	Drug Enforcement Administration
DMDNB	2,3-Dimethyl-2,3-dinitrotoluene
DMMP	Dimethyl Methyl Phosphonate
DNT	Dinitrotoluene
EC	Ecgonine Methylester
EGDN	Ethylene Glycol Dinitrate
EN	Electronegativity
FBI	Federal Bureau of Investigation
FIS	Field Ion Spectrometry
GC	Gas Chromatography
GC/IMS	Gas Chromatograph Ion Mobility Spectrometer
GCMS	Gas Chromatography/Mass Spectrometry
HMX	Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetraazine
HPLC	High-performance Liquid Chromatography
IMS	Ion Mobility Spectrometer
LOD	Limits of Detection
LSD	Lysergic Acid Diethylamide
MA	Methamphetamine
MDA	3,4-Methylenedioxyamphetamine
MDEA	The Abusive Drug ''Eve''
MDMA	The Abusive Drug ''Adam'' or ''Ecstasy''
MNT	Mononitrotoluene
MS	Mass Spectrometry
NG	Nitroglycerin
PA	Proton Affinity
PCP	Phenylcyclidine
PETN	Pentaerythritol Tetranitrate
ppb	part per billion
ppm	part per million
ppt	part per trillion
RDX	Hexahydro-1,3,5-trinitro-1,3,5-triazine
TATP	Triacetone Triperoxide
THC	Tetrahydrocannabinol
TNT	Trinitrotoluene
TOF	Time-of-flight
UV	Ultraviolet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Mass Spectrometry for Forensic Applications

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction

REFERENCES

1. D.I. Carroll, 'Personnel Marking and Detection Based upon the Plasma Chromatography(Trade Name) Concept', Final Report 510-F; AFATL-TR-68-79, Franklin GNO Corporation, West Palm Beach, FL, July 1968.
2. W.D. Kilpatrick, 'Plasma Chromatography and Dynamite Vapor Detection', Final Report FAA-RD-71-7, Contract DOT-FA71WA-2491, Federal Aviation Administration, Washington, DC, January 1971.
3. J.E. Cline, J.R. Hobbs, A.E. Barrington, 'Laboratory Evaluation of Detectors of Explosives' Effluents', Final Report, DOT/TSC, Cambridge, MA, November 1972.
4. W.A. Wall, H.M. Gage, 'Determination of the Sensitivity and Specificity of Vapor Detection Systems for Explosives, Narcotics, and Related Compounds,' Technical Report, USALWL, Aberdeen Proving Ground, Maryland, 1973.
5. T.W. Carr (ed.), *Plasma Chromatography*, Plenum Press, New York, 1984.
6. G.A. Eiceman, Z. Karpas, *Ion Mobility Spectrometry*, CRC Press, Boca Raton, FL, 1993.
7. R.H. St. Louis, H.H. Hill, Jr, 'Ion Mobility Spectrometry in Analytical Chemistry', *CRC Crit. Rev. Anal. Chem.*, **21**, 321–355 (1990).
8. G.A. Eiceman, 'Advances in Ion Mobility Spectrometry: 1980–1990', *Crit. Rev. Anal. Chem.*, **22**, 17–36 (1991).
9. J.A. Reategui, 'Ion Mobility Spectrometry: Theory and Applications', *Sensors*, **11**(4), 14–20, 59 (1994).

10. H.H. Hill, G. Simpson, 'Capabilities and Limitations of Ion Mobility Spectrometry for Field Screening Applications', *Field Anal. Chem. Technol.*, **1**, 119–134 (1997).
11. R. Wilson, A.H. Brittain, 'Explosive Detection by Ion Mobility Spectrometry', in *Explosives in the Service of Man, the Nobel Heritage*, eds. J.E. Dolan, S.S. Langer, The Royal Society of Chemistry, Cambridge, 1997.
12. Z. Karpas, 'Forensic Applications of Ion Mobility Spectrometry', *Forens. Sci. Rev.*, **1**, 104–119 (1989).
13. T. Keller, R. Binz, P. Regenscheit, W. Bernhard, R. Dirnhofer, A. Chlewinski, 'Ionenmobilitats-spektrometrie In der Forensischen Analytik und Kriminalistik', *Kriminalistik Schweiz*, (1), 67–70 (1996).
14. T. Keller, R. Binz, P. Regenscheit, W. Bernhard, R. Dirnhofer, A. Chlewinski, 'Ionenmobilitats-spektrometrie In der Forensischen Analytik und Kriminalistik', (Second part of ref. 13), *Kriminalistik Schweiz*, (1), 137–141 (1996).
15. H.E. Revercomb, E.A. Mason, 'Theory of Plasma Chromatography/Gaseous Electrophoresis – A Review', *Anal. Chem.*, **47**, 970–983 (1975).
16. M.J. Cohen, F.W. Karasek, 'Plasma Chromatography – A new dimension for Gas Chromatography and Mass Spectrometry', *J. Chromatogr. Sci.*, **8**, 330–337 (1970).
17. F.W. Karasek, D.W. Denney, 'Detection of 2,4,6-Trinitrotoluene Vapours in Air by Plasma Chromatography', *J. Chromatogr.*, **93**, 141–147 (1974).
18. F.W. Karasek, H.H. Hill, Jr, S.H. Kim, 'Plasma Chromatography of Heroin and Cocaine with Mass Identified Mobility Spectra', *J. Chromatogr.*, **117**, 327–(1976).
19. *The 1992 Workshop on Ion Mobility Spectrometry*, ed. G.A. Eiceman, US Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, USA, 1992.
20. *Second International Workshop on Ion Mobility Spectrometry*, as part of the 39th Canadian Spectroscopy Conference, eds. A. Lawrence, P. Pilon, Quebec, Canada, 1993.
21. *Third International Workshop on Ion Mobility Spectrometry*, ed. J.H. Cross, NASA Conference Publication 3301, 1994.
22. *Fourth International Workshop on Ion Mobility Spectrometry*, ed. A. Brittain, Cambridge, UK, 1995.
23. *Fifth International Workshop on Ion Mobility Spectrometry*, eds. J. Baumbach, J. Stach, Bastei, Germany, 1996.
24. *Sixth International Workshop on Ion Mobility Spectrometry*, 'Recent Developments in Ion Mobility Spectrometry', eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 1997.
25. S. Zaromb, J. Alcaraz, D. Lawson, C.S. Woo, 'Detection of Airborne Cocaine and Heroin by High-throughput Liquid-absorption Preconcentration and Liquid Chromatography–Electrochemical Detection', *J. Chromatogr.*, **643**, 107–115 (1993).
26. C.W. Su, J.R. Jadamec, L. Norwood, S.M. Rigdon, 'The Effect of Sample Holder Material on Ion Mobility Spectrometry Reproducibility', in *Second International Workshop on Ion Mobility Spectrometry*, eds. A. Lawrence, P. Pilon, Quebec, Canada, 221–228, 1993.
27. C.W. Su, J.R. Jadamec, L. Norwood, S.M. Rigdon, 'A Novel Approach to Increasing Cocaine Detection Confidence Utilizing Ion Mobility Spectrometry', in *Second International Workshop on Ion Mobility Spectrometry*, eds. A. Lawrence, P. Pilon, Quebec, Canada, 229–234, 1993.
28. P.A. Brown, J.H. Comparin, 'Application of the Ionscan™ for Detection of Methamphetamine and Ephedrine in Abandoned Clandestine Laboratories', in *Second International Workshop on Ion Mobility Spectrometry*, eds. A. Lawrence, P. Pilon, Quebec, Canada, 245–252, 1993.
29. P.A. Brown, H.F. Skinner, 'The Application of Ion Mobility Spectrometry for the Detection of Phenylcyclidine and Intermediate, 1-Piperidinocyclohexanecarbonitrile', in *Fifth International Workshop on Ion Mobility Spectrometry*, 'Recent Developments in Ion Mobility Spectrometry', eds. J. Baumbach, J. Stach, Bastei, Germany, 235–241, 1996.
30. A.M. DeTulleo-Smith, 'Barringer Ion Mobility Spectrometer Filter Disk Cleaning and Packaging Method', in *Fifth International Workshop on Ion Mobility Spectrometry*, 'Recent Developments in Ion Mobility Spectrometry', eds. J. Baumbach, J. Stach, Bastei, Germany, 243–247 1996.
31. D.D. Fetterolf, T.D. Clark, 'Detection of Trace Explosive Evidence by Ion Mobility Spectrometry', *J. Forens. Sci.*, **38**, 28–39 (1993).
32. C. Wu, H.H. Hill, Jr, 'Electrospray High Resolution Ion Mobility Spectrometry Mass Spectrometry', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 1997.
33. M.A. Baim, R.L. Eatherton, H.H. Hill, Jr, 'Ion Mobility Detector for Gas Chromatography with a Direct Photoionization Source', *Anal. Chem.*, **55**, 1761–1766 (1983).
34. A. Clark, P.D. Arnold, A.H. Brittain, R.C. Wilson, 'The Use of Pulsed Corona Discharge Ionisation in High Temperature IMS Systems', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 92–108, 1997.
35. J. Adler, G. Arnold, H.R. Doring, V. Starrock, E. Wulfing, 'First Results with the Bruker Saxonia Corona Discharge IMS', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 109–119, 1997.
36. O. Soppart, S. Sielemann, J.I. Baumbach, D. Klokow, 'Overcoming the Nickel Partial Discharges as Ionisation Source for Ion Mobility Spectrometry', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 312, 1997.

37. U.Kh. Rasulev, E.G. Nazarov, V.V. Palitsin, 'Identification of Trace Amounts of Hard-volatile Nitrous Bases Using Thermoflash, Surface Ionization and IMS Methods', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 1997.
38. G. Arnold, V. Budovich, A. Mikhailov, 'A High-energy Electron Source for an IMS', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 1997.
39. G.E. Spangler, P.A. Lawless, 'Ionization of Nitrotoluene Compounds in Negative Ion Plasma Chromatography', *Anal. Chem.*, **50**, 884 (1978).
40. C.J. Proctor, J.F.J. Todd, 'Alternative Reagent Ions for Plasma Chromatography', *Anal. Chem.*, **56**, 1794–1797 (1984).
41. J.P. Carrico, D.W. Sickenberger, G.E. Spangler, K.N. Vora, 'Simple Electrode Design for Ion Mobility Spectrometry', *J. Phys. E: Sci. Instrum.*, **16**, 1058–1062 (1983).
42. B. Carnahan, S. Day, V. Kouznetsov, A. Tarassov, 'Development and Applications of a Transverse Field Compensation Ion Mobility Spectrometry', in *Fourth International Workshop on Ion Mobility Spectrometry*, ed. A. Brittain, Cambridge, UK, 1995.
43. B. Carnahan, S. Day, V. Kouznetsov, A. Tarassov, 'Field Ion Spectrometry Applied to the Detection of Illicit Drugs and Explosives', in *Fifth International Workshop on Ion Mobility Spectrometry*, eds. J. Baumbach, J. Stach, Bastei, Germany, 249–274, 1996.
44. F.J. Knorr, R.L. Eatherton, W.F. Siems, H.H. Hill, Jr, 'Fourier Transform Ion Mobility Spectrometry', *Anal. Chem.*, **57**, 402–406 (1985).
45. A.H. Lawrence, P. Neudorfl, J.A. Stone, 'The Dependence of Cluster Ion Formation in IMS on Molecular Geometry with Particular Reference to Dinitroalkanes and Chloride Ions', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 320, 1997.
46. W.A. Munro, C.L.P. Thomas, M.L. Langford, 'Characterisation of the IMS Response of DMNB Under Water, Ammonia and Dichloromethane Reactant Ion Chemistries', in *Sixth International Workshop on Ion Mobility Spectrometry*, eds. H.H. Hill, Jr, D. Atkinson, Jackson, Wyoming, 120–158, 1997.
47. S. Boumsellek, S.H. Alajajian, A. Chutjian, 'Negative-ion Formation in the Explosives RDX, PETN and TNT by Using the Reversal Electron Attachment Detection Technique', *J. Am. Soc. Mass Spectrom.*, **3**, 243–247 (1992).
48. A.H. Brittain, J.L. Brokenshire, 'Quantitative Investigation of the IMS Detection of Nitro-compounds Over a Wide Temperature Range and The Effect of Dopants', in *Fifth International Workshop on Ion Mobility Spectrometry*, eds. J. Baumbach, J. Stach, Bastei, Germany, 67–86, 1996.
49. F. Garofolo, V. Migliozzi, B. Roio, 'Application of Ion Mobility Spectrometry to the Identification of Trace Levels of Explosives in the Presence of Complex Matrices', *Rapid Commun. Mass Spectrom.*, **8**, 527–532 (1994).
50. C. Rhykerd, K. Linker, D. Hannum, J. Parmeter, 'Walk-through Explosive Detection Portal', International Nuclear Material Management 38th Conference, 1997.
51. R. Wilson, A.H. Brittain, personal communication, 1998.
52. L.L. Danylewych-May, L. Kim, R. Jackson, 'Improvements to the Field Portable Ion Mobility Spectrometer', in '*Fifth International Workshop on Ion Mobility Spectrometry*, eds. J. Baumbach, J. Stach, Bastei, Germany, 335–352, 1996.
53. R. Wilson, A. Brittain, 'A Study to Investigate the Trace Levels of Contamination on Surfaces when Narcotic Contraband is Concealed in a Vehicle', *SPIE*, **2932**, 27–40 (1997).
54. C.W. Su, S. Rigdon, K. Babcock, 'The Detection of Illicit Drugs on Wet Samples Utilizing Ion Mobility Spectrometry', in *Fifth International Workshop on Ion Mobility Spectrometry*, eds. J. Baumbach, J. Stach, Bastei, Germany, 199–214, 1996.
55. A. Miki, T. Keller, P. Regenscheit, W. Bernhard, M. Tatsuno, M. Katagi, M. Nishikawa, L. Kim, S. Hatano, H. Tsuchihashi, 'Detection of Internal and External Methamphetamine in Human Hair by Ion Mobility Spectrometry', *Jpn. J. Toxicol. Environm. Health*, **43**, 15–24 (1997).
56. T. Keller, A. Miki, M. Tatsuno, M. Katagi, M. Nishikawa, H. Tsuchihashi, P. Regenscheit, R. Dirnhoer, 'Detection of Amphetamine, DMDA and MDEA in Human Hair by means of Ion Mobility Spectrometry', *IMS Journal*, (in press).
57. A. Miki, T. Keller, P. Regenscheit, R. Dirnhoer, M. Tatsuno, M. Katagi, M. Nishikawa, H. Tsuchihashi, 'Application of Ion Mobility Spectrometry to the Rapid Screening of Methamphetamine Incorporated in Hair', *J. Chromatogr. B (Biomed. Appl.)*, **692**, 319–328 (1997).
58. A.A. Nanji, A.H. Lawrence, N.Z. Mikhael, 'Use of Skin Surface Sampling and Ion Mobility Spectrometry as a Preliminary Method for Drug Detection in an Emergency Room', *Clin. Toxicol.*, **25**, 501–515 (1987).
59. A.H. Lawrence, A.A. Nanji, J. Taverner, 'Skin-sniffing Ion Mobility Spectrometric Analysis: A Potential Screening Method in Clinical Toxicology', *J. Clin. Lab. Anal.*, **2**, 101–107 (1988).
60. A.H. Lawrence, 'Detection of Drug Residues on the Hands of Subjects by Surface Sampling and Ion Mobility Spectrometry', *Forens. Sci. Int.*, **34**, 73–83 (1987).
61. T. Keller, A. Schneider, P. Regenscheit, R. Dirnhoer, T. Rucker, J. Jaspers, W. Kisser, 'Analysis of Psilocybin and Psilocin in *Psilocybe subcubensis Guzman* by Ion Mobility Spectrometry and Gas Chromatography/Mass Spectrometry', *Forens. Sci. Int.*, (in press).

62. 'Riot Agent Monitor™' (RAM™), Graseby Dynamics Limited, Watford, UK.
63. G. Allinson, C.W. McLeod, 'Characterization of Tear Gas Residues by Ion Mobility Spectrometry', *Appl. Spectrosc.*, **51**, 1880–1889 (1997).
64. K. Maekawa, 'The Sarin Poisoning Incident in Tokyo Subways', in Proceedings of the Fifth International Symposium on Protection against Chemical and Biological Warfare Agents, Stockholm, supplement 31–37, June 1995.
65. J. Stotz, P. de B. Harrington, 'Accelerent Detection for Arson Investigations Using Ion Mobility Spectrometry', Paper 884, Pittsburgh Conference, New Orleans, LA, 1998.
66. Detection Systems for Toxic Gases in Subway Stations, Brochure, Bruker Saxonia Analytik GMBH.
67. J. Adler, H.R. Doring, V. Starrock, E. Wulfing, 'A New Interference-resistant Ion Mobility Spectrometer', in Proceedings of 4th International Symposium on Protection Against Chemical Warfare Agents, Stockholm, Sweden, 175–182, 1992.
68. I.D. Ogden, N.J.C. Strachan, 'Enumeration of *Escherichia coli* in Cooked and Raw Meats by Ion Mobility Spectrometry', *J. Appl. Bacteriol.*, **74**, 402–405 (1993).
69. A.P. Snyder, M. Miller, D.B. Shoff, G.A. Eiceman, D.A. Blyth, J.A. Parsons, 'Enzyme-substrate Assay for the Qualitative Detection of Microorganisms by Ion Mobility Spectrometry', *J. Microbiol. Methods*, **14**, 21–32 (1991).
70. A.P. Snyder, S.N. Thornton, J.P. Dworzanski, H.L.C. Meuzelaar, 'Detection of the Picolinic Acid Biomarker in Bacillus Spores Using a Field-portable Pyrolysis–Gas Chromatography/Ion Mobility Spectrometry System', *Field Anal. Chem. Technol.*, **1**(1), 49–58 (1996).
71. J.P. Dworzanski, W.H. McClennen, W.H. Cole, S.N. Thornton, H.L.C. Meuzelaar, A.P. Snyder, 'Field-portable, Automated Pyrolysis–GC/IMS System for Rapid Biomarker Detection in Aerosols: A Feasibility Study', *Field Anal. Chem. Technol.*, **1**(5), 295–305 (1997).
72. I.A. Buryakov, E.V. Krylov, E.G. Nazarov, U.K. Rasulev, 'A New Method of Separation of Multiatomic Ions by Mobility at Atmospheric Pressure Using High-frequency Amplitude Asymmetric Strong Electric Field', *Int. J. Mass Spectrom. Ion Processes*, **128**, 143–148 (1993).
73. M.J. Cohen, F.W. Karasek, 'Plasma Chromatography – A New Dimension for Gas Chromatography and Mass Spectrometry', *J. Chromatogr. Sci.*, **8**, 330–337 (1970).
74. A.P. Snyder, C.S. Harden, A.H. Brittain, M.G. Kim, N.S. Arnold, H.L.C. Meuzelaar, 'Portable, Hand-held Gas Chromatography/Ion Mobility Spectrometry', *Int. Lab.*, 30–33, March 1993.
75. N.S. Arnold, M.G. Kim, W.H. McClennen, J.P. Dworzanski, H.L.C. Meuzelaar, 'Optimization of Transfer Line Gas Chromatography for Direct Atmospheric Vapor Sampling Ion Mobility Spectrometry Applications', in *The 1992 Workshop on Ion Mobility Spectrometry*, ed. G.A. Eiceman, US Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, USA, 11–17, 1992.

Mass Spectrometry for Forensic Applications

Chung-Hsuan Chen
Oak Ridge National Laboratory, Oak Ridge, USA

1 Introduction **2**
1.1 Characteristics of DNA 2
1.2 Brief Historical Review of DNA Typing for Forensic Applications 2
1.3 Forensic DNA Sample Preparation 2
1.4 Population Genetics for Reliability Estimation 3
1.5 Gel Electrophoresis for DNA Analysis 3
1.6 Matrix-assisted Laser Desorption/Ionization for Biomolecule Detection 4

2 Mass Spectrometry DNA Sizing for Forensic Applications **5**
2.1 Restriction Fragment Length Polymorphism and Variable Numbers of Tandem Repeats Analysis 5
2.2 Short Tandem Repeats 6
2.3 Single-nucleotide Polymorphism 7
2.4 Gender Determination 9

3 DNA Sequencing by Mass Spectrometry for Forensic Applications **10**
3.1 DNA Sequencing by Mass Spectrometry with DNA Ladders 11
3.2 Direct Sequencing 12

4 Conclusion **13**

Acknowledgments **13**

Abbreviations and Acronyms **13**

Related Articles **14**

References **14**

Recent advances in DNA technology for the detection of variation in specific DNA sequences have proved that DNA fingerprinting analysis can become a very valuable technology for forensic applications. By analyzing a sufficient number of regions of DNA that show variability between nonrelated individuals, one can reduce the probability of two individuals matching by chance, to practically zero. Thus, it can be demonstrated that an individual's genetic fingerprint is as unique as their skin fingerprints, and often easier to obtain. However, the requirements for DNA fingerprinting for forensic application involve not only the technology but also the cost, time and reliability of the analysis. DNA fingerprinting technology has tremendous power of exclusion and inclusion and is used routinely as legal evidence. Due to the high reliability of DNA fingerprinting, courts often require DNA analysis as evidence in the most serious criminal cases. However, present gel electrophoresis analysis of DNA in forensic applications is considered to be slow and expensive. Consequently, a backlog of DNA analysis for forensics is common. Mass spectrometry DNA analysis can help to resolve this difficulty, since it can be faster, cheaper and more reliable for forensic sample analysis.

DNA fingerprinting is a technology that identifies individuals based on patterns of DNA markers detected in the genomic DNA. It can be used in forensic analysis to identify suspects or victims. It is particularly valuable at scenes of violent crimes where a body may not be available, or in instances where decomposition or dismemberment excludes the use of standard forensic techniques. There are several different approaches available to obtain DNA fingerprints for forensic application. These include restriction fragment length polymorphism (RFLP), short tandem repeats (STRs), single-nucleotide polymorphism (SNP) and DNA sequencing. Currently, the bulk of RFLP, STR, SNP analysis and DNA sequencing utilize gel electrophoresis, which is time-consuming and labor intensive. With the newly developed mass spectrometry technology for DNA analysis, the analysis time can possibly be reduced by more than 90% and automation of sample analysis is much easier. Furthermore, mass spectrometric analysis does not require the use of chemical dyes or radioactive materials, thus increasing safety and reducing cost in sample preparation and handling. In addition, the cost and concern of disposing of hazardous waste are eliminated.

The major difficulties preventing the use of mass spectrometry for DNA analysis are the very low vapor pressure of DNA and the fragmentation of these large DNA molecules by most ionization methods. In 1987, Professors Hillenkamp, Karas and their colleagues successfully developed matrix-assisted laser desorption/ionization (MALDI) for protein analysis. In this approach, a protein sample was mixed with a much larger quantity of small organic compounds which serve as a matrix. These matrixes, such as sinapinic acid, functioned to prevent the protein molecules from becoming entangled with one another. In MALDI, the DNA sample is irradiated with a laser beam, which causes desorption. The wavelength of the laser is often chosen so as to have strong absorption by the small matrix molecules, but not the protein. After the absorption of the laser photons by the matrix molecules, the matrix molecules are vaporized and carry the much

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

larger protein molecules into space. Some ions and electrons are also produced due to this laser ablation process. Some protein ions can be produced by protonation or deprotonation processes caused by the interaction of desorbed protein molecules and desorbed matrix molecules or ions.

During the past few years, MALDI has also been successfully applied to DNA detection through the discovery of new matrixes and improvements in instrumentation. The main advantage of MALDI analysis is its very fast speed. However, the size of DNA fragments which can be detected is still limited to less than 3000 base pairs (bp). Mass spectra for DNA fragments larger than 100 nucleotides usually have poor mass resolution ($M/\Delta M < 200$). Nevertheless, there are many forensic applications using mass spectrometry for DNA analysis which are not limited by modest mass resolution and detection efficiency of large DNA fragments.

1 INTRODUCTION

1.1 Characteristics of DNA

DNA molecules are present in all living cells. In general, DNA contains a repeated structure of four different molecules: adenine (A), guanine (G), cytosine (C) and thymine (T). The chemical structures and molecular weights of these four molecules are shown in **(1)**. The geometric structure of DNA was found by Watson and Crick to be a double helix with antiparallel chains of deoxyribonucleotides. The strands are joined by

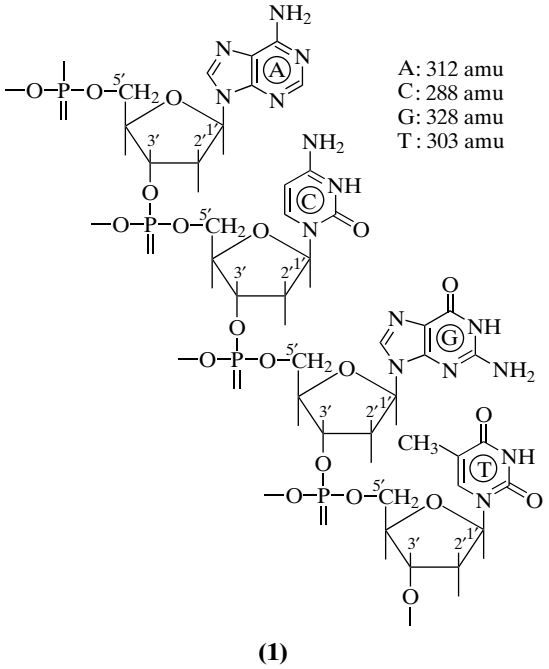

(1)

complementary base pairing consisting of hydrogen bonding between A and T, and G and C. Owing to the strong A–T and G–C coupling, the length of a double-stranded DNA (dsDNA) molecule is often referred to as the number of bp. In a human cell, there are 23 pairs of chromosomes, which together contain the estimated 3.6 billion bp of the human genome. DNA identification analysis in general refers to the specific features of an individual genome. Every person has a unique genotype (genetic make-up) which subsequently leads to different phenotypes (physical manifestation of genotype). Although all humans have all the same genetic make-up, there is a great deal of difference in nucleotide sequence between different individuals. The DNA sequence of any individual is identical even when it is extracted from different cells. The individual's genetic uniqueness and identical DNA sequence within all tissues of the same person provide a strong basis for using DNA typing for individual identification through DNA fingerprinting.

1.2 Brief Historical Review of DNA Typing for Forensic Applications

The foundation of DNA typing for person identification was established by the observation of RFLP in which polymorphic DNA loci are determined from restriction enzyme digestion.[1] In 1985, Jeffreys et al.[2] discovered 33 and 16 bp repeated structures vary in number between different individuals, referred to as minisatellites, that can be used for fingerprinting. In 1987, Nakamura et al.[3] reported the isolation and characterization of variable numbers of tandem repeats (VNTRs) for single locus DNA profiling. More recently, emphasis has been shifted to the typing of STRs[4] because of the high reliability resulting from the large number of STRs in the human genome. In the last few years, it was discovered that SNPs occur once every 300–1000 bp in DNA fragments. Thus, SNP typing is expected to become a very reliable tool for person identification in the future. The ultimate accuracy of DNA fingerprinting for personal identification will be achieved once the entire human genome is sequenced. At present, the time and cost factors prohibit the use of sequencing the entire genome for forensic purposes. However, the rapid progress in sequencing technology as a result of the Human Genome Project[5] suggests the possibility of using a sequencing approach for forensic applications.

1.3 Forensic DNA Sample Preparation

In the process of using DNA typing for forensic application, it is critically important to have very reliable and stringent protocols established. The major steps for DNA typing often include:

1. Sample collection: blood, buccal cells and hair follicles are the general sources for fresh DNA. For dried stains, samples should be collected and maintained in a dry, cool state. For autopsy tissue, it can be more reliable to have samples from more than one organ. All samples must be labeled with the source, tissue type and the time and date of sample collection.
2. Storage and transport: DNA samples can be stored and shipped as nonextracted tissue, a lysate with tissue in an appropriate EDTA (ethylenediamine tetraacetic acid) solution or fully extracted pure DNA. For nonextracted tissues, sample refrigeration during shipping is often recommended.
3. DNA extraction: standard phenol–chloroform extraction and alcohol precipitation processes are often used to extract DNA from biological samples. An automated DNA extractor such as the nucleic acid extractor (Applied Biosystems, Perkin Elmer) can achieve high-quality extraction. It is desirable to complete DNA extraction as soon as possible.
4. DNA amplification: DNA amplification is often necessary to increase the quantity of sample available for analysis. The polymerase chain reaction (PCR)[6] is the standard technique for the amplification of a selected DNA segment. Since an amplification of 10^6 or more can be achieved by PCR, any contamination by other DNAs can lead to a false conclusion. Thus, great care must be taken to prevent cross-contamination of sample material at the collection site and at any time during the analysis.
5. DNA size or sequence determination. This is the ultimate step leading to the conclusion of person identification. Up to now, gel electrophoresis has been used for both DNA size determination and sequencing. In this article, analysis by mass spectrometry will be emphasized.

1.4 Population Genetics for Reliability Estimation

The primary goal of DNA profiling for forensic application is to get the assurance of a match (inclusion) or mismatch (exclusion) of the evidence DNA sample with the suspect or victim's sample. In general, DNA profiling can be used easily to rule out the suspect when DNA typing between suspect and evidence shows clear differences. Thus, DNA fingerprinting is a very powerful tool for proving that an individual does not match the DNA collected at a crime scene. On the other hand, if the DNA profile from the evidence sample matches with the suspect, there is still a possibility that the evidence DNA comes from another person. It might be a coincidental match between two persons having similar DNA profiles. Thus, population genetics plays an important role in suspect identification. The simplest approach is to assume mating is random. The random mating of persons produces the same genotypes as random combinations of sperm and egg. Assume at a locus X, that 10% of the alleles are X_1 and 12% are X_2. For an X_1 homozygote (X_1X_1), the frequency is $10\% \times 10\% = 1\%$. For a heterozygote (X_1X_2), the frequency is $10\% \times 12\% \times 2 = 2.4\%$. Randomly mating populations are described as being in Hardy–Weinberg (HW) equilibrium.[7] Although local populations (especially small populations) are not usually strictly randomly mating, the HW calculation provides a good approximation when no special genetic linkage is involved. For subpopulation groups such as people in an isolated village, the random mating approach is not generally a reliable assumption. However, it can probably be used as a guide for the minimum number of loci to be analyzed for the confirmation of matching suspect and evidence.

Since the confirmation of matching the suspect with the evidence DNA often requires the analysis of several loci, the time needed for DNA profiling can be long and the cost high. At present, all forensic DNA samples for court evidence are analyzed by gel electrophoresis. Visualization of DNA in this procedure requires the use of either radioactive material or dye tagging. VNTR electrophoretic analysis, especially with radioactive probes, can take several weeks to finish the analysis of different loci[8] for forensic confirmation. The human factor introduced by lengthy handling increases the risk of error or contamination and reduces reliability. It is obvious that a faster, less expensive and more reliable method is required for processing the large number of forensic samples.

1.5 Gel Electrophoresis for DNA Analysis

The principle of gel electrophoresis is to separate different size DNA fragments by their different migration speeds in gel when an electric field is applied. Electrophoresis through an agarose or polyacrylamide gel matrix is the standard method used to separate, identify, and purify DNA segments. Agarose and polyacrylamide gels can be poured in a variety of shapes, sizes and porosities with different configurations. The selection of these parameters primarily depend on the sizes of DNA fragments to be separated. Polyacrylamide gel can have single base resolution for DNA fragments smaller than 500 bp. Due to their high resolution, polyacrylamide gels are often used for DNA sequencing. However, handling acrylamide gels is often tricky and requires skill and experience for reliable and reproducible results. Polyacrylamide gels also take a long time to run, and staining requires additional lengthy steps. Agarose gels have a lower resolution power than polyacrylamide gels

but have a greater range of separation and generally run faster and are easier to stain. DNAs up to 50 kbp can be separated. Larger DNA fragments up to 10 Mbp can be pursued by pulsed gel electrophoresis.[9] The detection of DNA with gel electrophoresis needs dye tagging such as ethidium bromide, or radioactive material tagging such as ^{32}P or ^{35}S. However, the migration speed depends not only on the molecular sizes of DNA but also the conformation of DNAs, gel concentration, electrophoresis buffer, presence of intercalating dyes, etc. For example, circular DNA and linear DNA with the same molecular weight can have different migration speeds. Ethidium bromide can reduce the electrophoretic mobility by more than 10%. With a mass spectrometric approach to DNA analysis, all parameters which can change the ion mobility, except the size of DNA fragments, are no longer a concern.

1.6 Matrix-assisted Laser Desorption/Ionization for Biomolecule Detection

It is naturally logical to consider a mass spectrometry approach for DNA sequencing, since the development of mass spectrometry for large bio-organic molecules detection has already aroused major interest for decades.[10] In particular, the principle of a time-of-flight (TOF) mass spectrometer is somewhat similar to gel electrophoresis. However, most mass spectrometric approaches failed to keep the large biomolecules intact during the vaporization and ionization processes. In order to overcome this difficulty, mass spectrometry coupled with electrophoresis for DNA analysis was pursued.[11] In 1987 Karas et al.[12] discovered that large protein molecular ions can be produced by laser desorption without much fragmentation if these biomolecules are mixed with small organic compounds that serve as matrixes for strong absorption of a laser beam. This process is now called MALDI. The typical preparation technique for MALDI is to dissolve biomolecular samples in solution, then prepare a matrix that contains small organic compounds such as 3-hydroxypicolinic acid. These sample and matrix solutions are subsequently mixed and a small amount of the mix is placed on a metal sample plate to dry. After the organic material crystallizes, the sample plate is placed in the mass spectrometer for analysis. During the MALDI process, matrix materials strongly absorb the laser energy and quickly become vaporized. Large biomolecules are carried along during the fast vaporization process. By this method large biomolecules can be delivered into space without breakup, which is probably due to the minimal direct absorption of laser energy; thus, ''soft'' desorption can be achieved. Furthermore, it has also been found that protein ions can be produced during the laser desorption process, in addition to the expected neutral molecules. Ionization is thought to involve a proton transfer process.

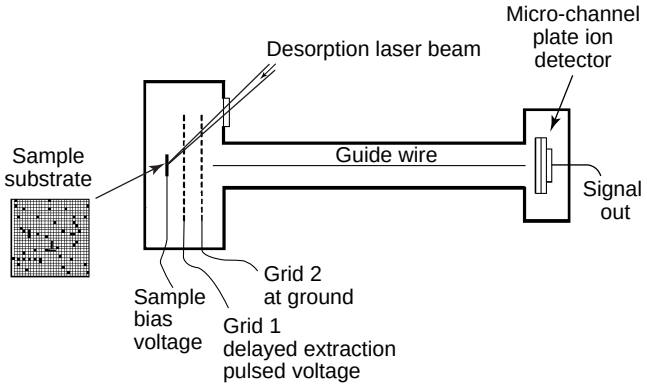

Figure 1 The experimental schematic of a typical MALDI/TOF mass spectrometer. Multiple samples can be placed in the mass spectrometer for analysis. Most MALDI DNA analysis has been performed in a linear TOF mass spectrometer. Both linear and reflectron TOF mass spectrometers have been used for protein and other organic polymers analysis. Recently, delayed pulsed ion extraction has been used to enhance mass resolution. In addition to TOF, MALDI has also been used in Fourier transform mass spectrometers. The desorption/ionization step was accomplished either by a pulsed ultraviolet (UV) or infrared (IR) laser. Ion detection was mostly based on secondary electron emission.

Most MALDI experiments have been carried out by using a TOF mass spectrometer for mass determination, since TOF mass spectrometers have the advantages of simplicity, high sensitivity, fast speed, probing the whole mass region, and nearly unlimited mass range. The schematic of a MALDI device is shown in Figure 1. A pulsed UV or IR beam is used to achieve matrix-assisted desorption and ionization. The target plate with biological samples is usually biased to 20–50 kV for ion acceleration. The field free drift region for a linear TOF mass spectrometer is for the separation of different sizes of biomolecules. A reflectron TOF mass spectrometer is often used to improve the mass resolution for protein samples. However, spectra of large DNA molecules by a reflectron TOF often fails to achieve the same degree of improvement on mass resolution because of the limited lifetimes of DNA ions produced by laser desorption.[13] The detection of DNA ions is typically by an electron multiplier or a channeltron or microchannel plates. When the front surface of the detector is impacted by biomolecular ions, secondary electrons can be ejected and then subsequently be amplified by the detector, a fast preamplifier and amplifier. The output is fed into an on-line computer for data processing. Several commercial instruments have the capability for automatic sample preparation and data processing for peak identification and resolution determination. In addition, it is plausible that portable MALDI/TOF mass spectrometers can be developed in field use by forensic investigators, thus decreasing the potential of crime scene/sample contamination.

Since the discovery of MALDI, many research groups have succeeded in measuring various proteins and large organic compounds by MALDI.[14] MALDI has also been applied to DNA segments. Initially, success was limited to small DNA detection[15] and gaining some understanding of the mechanism of the MALDI process.[16] Then, Wu et al.[17] discovered 3-hydroxypicolinic acid was a good matrix for mixed-base oligonucleotide and succeeded in detecting oligonucleotides of 67 bases. With the development of an instrument to give high ion energy and the use of new matrices, Tang et al.[18] succeeded in using MALDI to detect single-stranded DNA (ssDNA) of 500 nucleotides and dsDNA of 500 bp. Berkenkamp et al.[19] also reported the detection of 2000 bp DNAs with IR/MALDI. This clearly indicates that MALDI can be used for large DNA analysis. Recently, MALDI was also used for mutation detection for disease diagnosis caused by base deletion[20] and point mutation[21] for cystic fibrosis. Successful sequencing short DNAs by mass spectrometry has also reported.[22–27] The success of mass spectrometry for DNA sizing and sequencing paved the way to use mass spectrometry for DNA typing for forensic applications. With mass spectrometry DNA profiling, there is a potential that the analysis time for each sample can be a few seconds instead of the hours needed by standard gel electrophoresis. Furthermore, no chemical or radioactive tagging or gel preparation is required.

2 MASS SPECTROMETRY DNA SIZING FOR FORENSIC APPLICATIONS

Each DNA typing involves (1) measurements of DNA fragments associated with genetic markers, (2) match determination for various genetic markers for different samples, and (3) statistical analysis for the type match to determine the possibility of an accidental match. As described above, all sequencing or size determination of DNA fragments for forensic applications has been performed with gel electrophoresis. Various approaches for DNA applications in the forensic community include the analysis of VNTR, STR, SNP and DNA sequencing. Laboratory protocol and standard procedures for data analysis were set up for forensic sample analysis. With the recent progress in development of mass spectrometry for DNA analysis, there is a good potential that mass spectrometry can be used for all the above applications. Since the use of mass spectrometry for DNA analysis is still in the early developmental stages, standard protocols for its use in forensic analysis have not been established. The potential applications of using mass spectrometry in DNA fingerprinting for forensic applications will be presented and discussed below.

2.1 Restriction Fragment Length Polymorphism and Variable Numbers of Tandem Repeats Analysis

DNA fingerprinting relies on the use of DNA markers. Among the most widely used DNA markers are RFLPs. RFLP analysis relies on differences that affect the position of restriction endonuclease recognition sites. Restriction endonucleases are enzymes which recognize and cut dsDNA at sequence specific recognition sites. Homologous DNA fragments from different individuals will contain different cut sites based on random mutations within the stretch of nucleotides. Consequently, fragments differing in length and number will be generated upon restriction digestion. These constitute RFLPs and, because they differ in size (length), the RFLP fragments can be readily separated by routine agarose electrophoresis.

When an individual's entire genomic DNA is digested by a restriction endonuclease, hundreds to thousands of restriction fragments are produced. When digested genomic DNA is subjected to gel electrophoresis, individual bands are not discernible. In such a case, analysis is generally facilitated by the Southern blotting technique. In Southern blotting, digested DNA is separated by electrophoresis. The digested fragments are transferred to a nitrocellulose or nylon membrane through capillary action. The DNA binds to the membrane by electrostatic attraction. Labeled probes specific for the RFLP of interest are then hybridized to the DNA-containing membrane. A banding pattern can be visualized using autoradiography or chemical detection techniques. Each individual will exhibit a characteristic pattern. Several RFLPs must be used to statistically ensure an individual's pattern is unique. The fact that RFLPs can occur anywhere in the genome and that there are potentially millions of RFLPs detectable with different probes makes RFLP analysis a tremendously powerful tool for forensic use.

A special class of RFLPs is based on DNA sequences that occur in tandem repeats. Tandem repeats arise when slippage mutation occurs during DNA replication. It has been observed that this type of mutation occurs frequently enough to generate significant variation over many generations. Polymorphism for these markers is so high in humans that only identical twins will have the same patterns. This factor makes the use of tandem repeats analysis extremely valuable to forensic scientists.

VNTR analysis has been used extensively in forensic applications. VNTRs consist of approximately 10 to 100 bp repeated several times in tandem. For example, a tandem repeat of the trinucleotide CTG would appear in the genome as CTGCTGCTGCTGCTGCTGCTG. In this particular example this sequence has seven repeats. VNTRs are generally bordered by unique sequence DNA,

a characteristic that can be exploited by current technology. In VNTR analysis, genomic DNA is digested with a restriction endonuclease which does not contain a cut site within the tandem repeat sequence. Since the restriction enzyme cuts outside of the tandem repeat region, the tandem repeat is left intact. Tandem repeat fragments are separated by gel electrophoresis and analyzed with Southern blotting. The length of the tandem repeat will vary between individuals. Individuals can be homozygous or heterozygous for a particular VNTR. The large number of VNTR-containing loci creates millions of possible pattern combinations. If there are n loci, there should be n homozygotes and $n(n-1)/2$ heterozygotes. If m loci are analyzed, the number of possible genotypes is $[n(n+1)/2]^m$. If $n = 20$ and $m = 4$, there are more than 1 billion genotypes. As a result, VNTR provides the forensic analyst with a powerful tool for inclusion and/or exclusion.

The need for Southern blotting can be eliminated if the VNTR region is selectively amplified with the PCR. PCR primers complementary to sequences flanking the tandem repeat region can be designed and used to amplify the tandem repeat region. The length of an individual's tandem repeat can then be obtained by gel electrophoresis. It is PCR amplification of VNTRs that shows a lot of promise for MALDI detection.

By using MALDI for RFLP detection, gel electrophoresis, Southern blotting and radioactive material or dye tagging can be eliminated.[29,30] Figure 2 shows the mass spectra for restriction enzyme digested DNA samples. Hybridization probes can also be detected by MALDI.[31] An example is shown in Figure 3. At present, detection of DNA of a size much larger than 500 bp is still not routine. It is preferred to use MALDI in RFLP for detection of small DNA fragments. A PCR process can be used to replicate a selected part of genomic DNAs. The PCR products can then be subjected to enzyme digestion to produce short DNA fragments for MALDI detection.

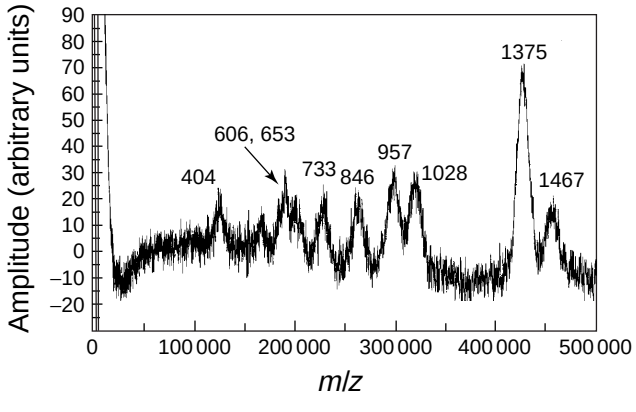

Figure 2 A UV/MALDI negative ion mass spectrum from an enzyme digestion of a microbial DNA sample. DNA fragments of more than 1000 bp in size were detected.

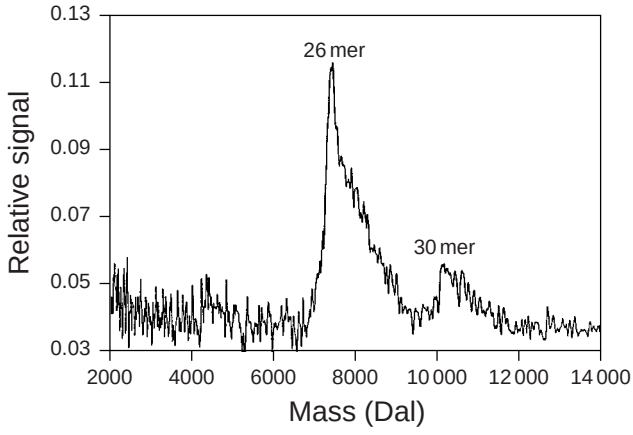

Figure 3 Negative ion mass spectrum showing the use of MALDI for the detection of probes for hybridization. Two probes with different lengths were used for hybridization at a single spot. The result indicates that high multiplexing is feasible when mass resolution is improved.

Since mass spectrometry can be used to measure RFLP with fewer steps than gel electrophoresis and it can have better resolution, MALDI is expected to play an important role in RFLP detection in the future.

2.2 Short Tandem Repeats

For VNTR analysis by measuring RFLP without the use of PCR, one major disadvantage is that the quantity of the DNA sample is so small that a long analysis time is required. One promising alternative is to amplify and measure loci containing STRs. There is a very large number of STRs distributed throughout the human genome. STR analysis can be used for as many loci as needed for reliable identification. PCR is nearly always used in STR analysis. Since PCR can amplify the selected DNA segments by more than six orders of magnitude, only a very small quantity of DNA sample is required. However, the disadvantage is that any procedure using PCR amplification is highly susceptible to contamination. Any trace foreign DNA contamination can be amplified and detected to potentially lead to a false conclusion. However, careful handling of DNA samples or applying quantitative PCR techniques can more or less resolve the concerns about DNA contamination. Since the analysis of STR is by measurement of the sizes of DNA products from the PCR process, the size of DNA fragments is often less than 300 bp and the quantity is often in the range of a few picomoles, which can be readily detected by MALDI.

A typical mass spectrum of a DNA sample amplified by PCR from a part of the vWFII gene is shown in Figure 4. Peaks for PCR products are clearly visible; a difference of 12 bases with three 4-base tandem repeats can clearly be observed. Both primary and doubly charged

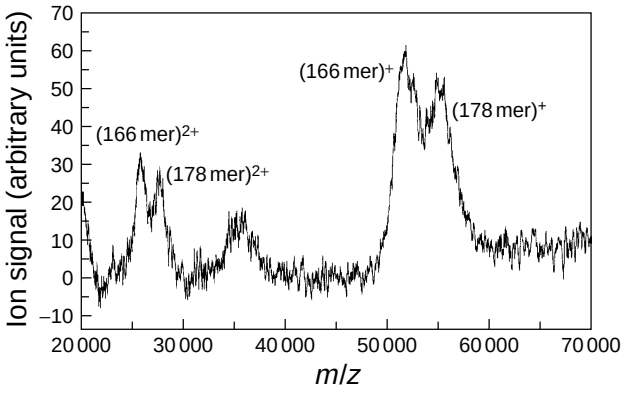

Figure 4 Negative ion mass spectrum of PCR product: 166 and 178 bp in the intron 40 of human WFII gene.

ion peaks are observed. The spectrum indicates that one chromosome in this region is 12 bp longer than the same region in the other chromosomes. Thus, the result indicates that the sample is from a person with heterozygous vWFII. Several different genes were used for mass spectrometric detection; their repeat sequences and the range of their allele sizes are shown in Table 1.[32] Take the F13B locus as an example. The maximum difference for the number of 4-base repeats is 5. There are 15 different genotypes for this locus. Since different people have different numbers of repeats in each of these genes, measurements of the number of repeats for several loci can be used for person identification. Examples of various mass spectra for different people for different loci are shown in Figure 5. The results were compared with results from gel electrophoresis, with good agreement. There are many tetranucleotide repeats which can be used for forensic identification. HUMTH01 is one that is often used for forensic application.[33] Eight alleles were distinguished for identification purposes.

In addition to tetranucleotide repeats, trinucleotide repeats are common in many well-characterized genes which include Huntington's disease, DRPLA (dentatorubral–pallidoluysian atrophy), Fragile X, etc. The numbers of the repeats for these genes are shown in Table 2. The use of MALDI for trinucleotide repeat detection in the DRPLA gene is shown in Figure 6.[34] The number of repeats and the confirmation of homozygosity or heterozygosity can be readily obtained. With the combination of these loci, the number of possible genotypes can reach billions. Consequently, this technique proves to be useful for reliable identification. However, the use of genes associated with disease requires special caution regarding the use of random mating approximations for forensic application, since disease genes may have special distributions among different ethnic groups.

2.3 Single-nucleotide Polymorphism

There are several types of DNA sequence variation, which include deletion, insertion, difference in number of repeats such as STR, and single bp differences. The single base difference is the most common. It was estimated that two random human homologous chromosomes differ at a frequency of about 1 in 300–1000 bases. This single bp difference is often referred to as a SNP. There are several million SNPs in the human genome. The predominance of SNPs and the much lower mutation rate makes SNP analysis extremely valuable for both forensic and medical applications.

Allele-specific hybridization is often used for SNP detection. It often requires incubation, washing and careful control of temperature and salt concentration. However, discrimination often can be difficult due to imperfect complement hybridization. Another approach is to use allele-specific polymerase chain reaction (ASPCR). Mass spectrometry can be used for the analysis of PCR products. ASPCR/MALDI has been pursued for a simulated fragment of the p53 gene.[21] The sequence of the template and the primers used are:

Primer 1 (23 nucleotides): 5′-CAGCTGTGGGTTGATTCCACACC-3′

Primer 2 (16 nucleotides): 3′-GGGGCGGGCCGTGGGC-5′

Template (60 bp): 5′-CAGCTGTGGGTTGATTCCACAC<u>C</u>CCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTAC-3′

Table 1 Sequence polymorphism and frequency of alleles

Sample	Locus (Gene)	Chromosome location	Repeat sequence	Allele size range (bp)
GZ	LPL (Human lipoprotein lipase)	8 p22	AAAT	105–133
F13B	F13B (Blood coagulation factor X111 B)	1 q31-q32.1	AAAT	169–189
vWFII[a]	vWF (Human vWF gene)	12 p12-pter	AGAT	131–191
vWA[a]	vWF (Human vWF gene)	12 p12-pter	AGAT (or TCTA)	102–154

[a] Sequence located in intron 40 with different primers to generate different sizes of repeat-containing fragments.

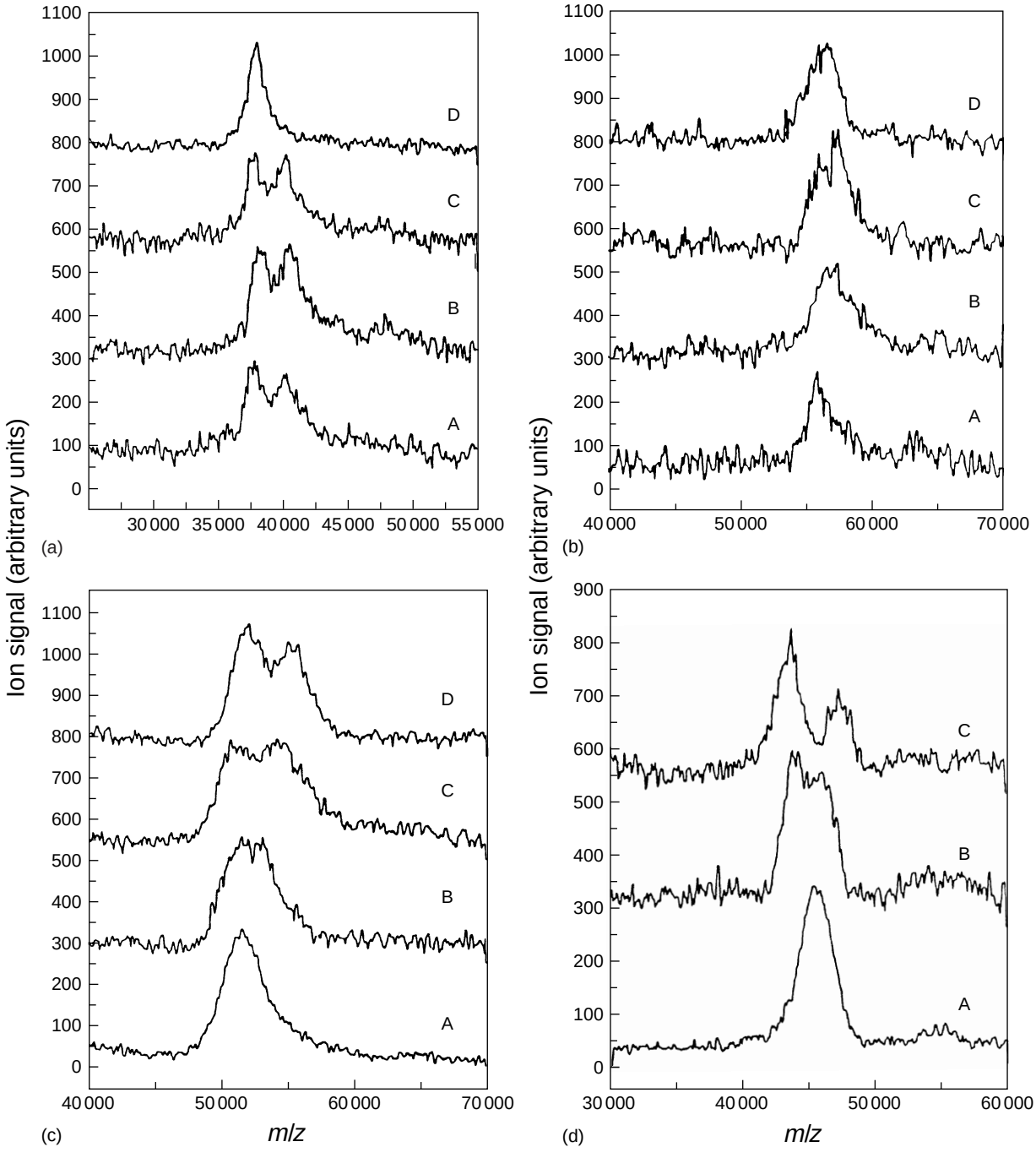

Figure 5 Negative ion mass spectra of PCR products of (a) LPL, (b) F13B, (c) vWFII and (d) vWA genes from different people. Homozygosity and heterozygosity as well as the number of repeats can be determined.

Upstream and downstream primers are overlapped at the base, which may be the site for base difference. If the template is normal, PCR will proceed. The size of amplified DNA product is 38 bp. The experimental result is shown in Figure 7. For a template with the base difference, no PCR products are expected. Results for this case are shown in Figure 8. This approach has also been applied to G551 in the cystic fibrosis transmembrane conductance regulator (CFTR) gene with human samples.[35] Since the lengths of primers can be varied, multiplexing can be applied to monitor several SNP in a single sample.

Another approach to the use of MALDI for SNP detection is to hybridize a primer adjacent to the

Table 2 Various genes with trinucleotide repeats

Disease	Repeated sequence	Normal number of copies
Fragile X syndrome	CCG	6–50
Spinobulbar muscular atrophy (Kennedy disease)	AGC	11–31
Myotonic dystrophy	AGC	5–35
Huntington's disease	AGC	9–37
Spinocerebellar ataxia type 1	AGC	25–36
Fragile X	CCG	6–25
Dentatorubral and pallidoluysian atrophy	AGC	7–23

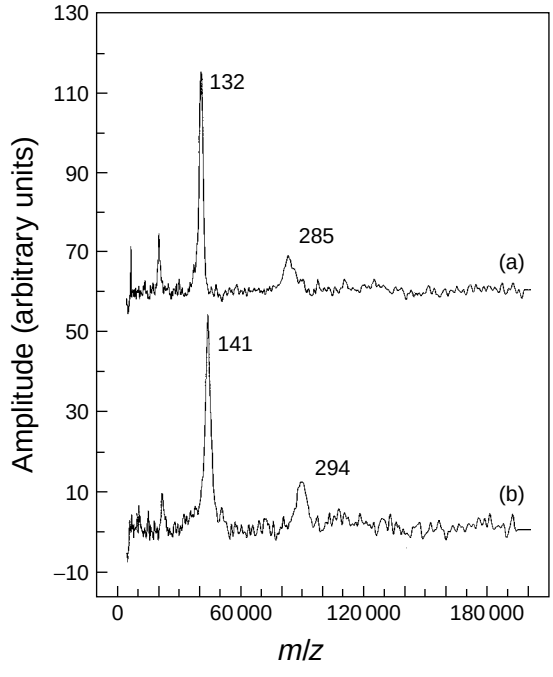

Figure 6 Mass spectra of samples from two patients, showing expanded CAG repeats. (a) Mass spectrum of PCR product of 132 and 285 bp with the number of CAG repeats being 12 and 63. (b) Mass spectrum of PCR product of 141 and 294 bp with the number of repeats being 15 and 66 for DRPLA locus.

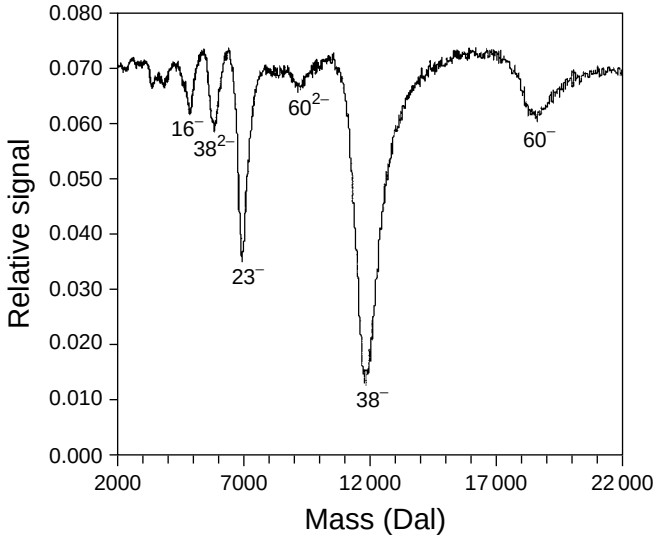

Figure 7 Negative ion mass spectrum of 38 bp PCR product with two primers overlapping at one base. The lengths of the primers are 16 and 23 nucleotides.

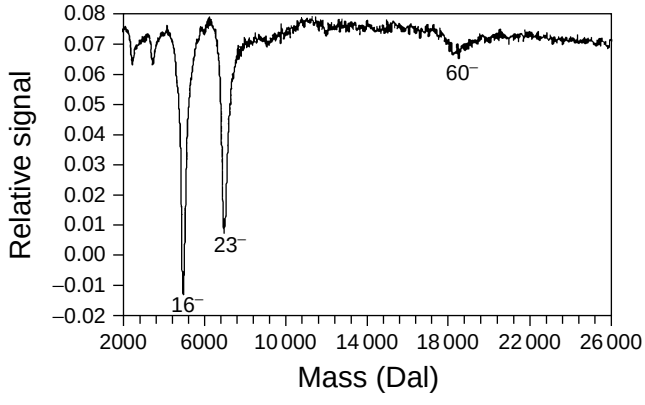

Figure 8 Negative ion mass spectrum with primers not complementary to the template. No PCR products were detected.

polymorphic locus and extend it by a single base by dideoxynucleotide addition. Mass spectrometry can be used to accurately measure the mass of the extension product in order to identify the base.[36] For homozygotes, one peak will be observed. However, two peaks should be observed for a heterozygote. Since different lengths of primers can be used for extension, it is possible to achieve a high throughput.[37]

2.4 Gender Determination

Sex determination is often important for identifying suspects and victims. A reliable gender test is PCR amplification of a segment of the X–Y homologous gene amelogenin.[38] A single pair of primers spanning part of the first intron generates 106 bp and 112 bp PCR products from the X and Y chromosomes respectively. MALDI can be subsequently used for the measurements of PCR products. Recently, it was found that the deletion of the Y-encoded gene occurs in a small percentage of Y chromosomes. Thus, it is more reliable to test both amelogenin and the male sex determining gene SRY for gender determination. Co-amplifications of genomic templates are listed in Table 3. Male DNA generates three products of 93 bp (SRY), 106 bp (amelogenin in the X chromosome) and 112 bp (amelogenin in the Y chromosome). Co-amplification of female DNA

Table 3 Sex-typing markers for human genomic DNA

Gene	Primer names	Primer sequence	Product size (bp)
Amelogenin	Amel-A	5′-CCCTGGGCTCTGTAAAGAATAGTG-3′	X = 106
	Amel-B	5′-ATCAGAGCTTAAACTGGGAAGCTG-3′	Y = 112
SRY	F11	5′-ATAAGTATCGACCTCGTCGGAA-3′	
	R7	5′-GCACTTCGCTGCAGAGTACCGA-3′	Y = 93

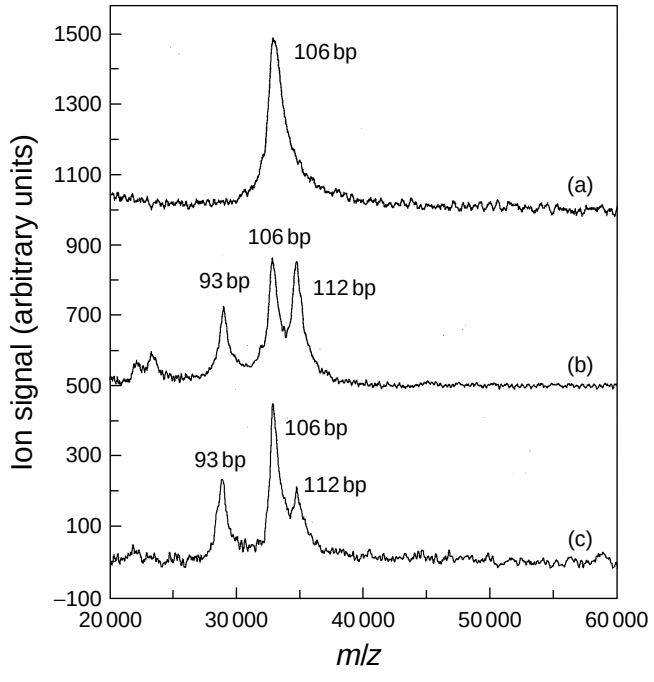

Figure 9 Negative ion mass spectra of sex-specific co-amplification (amelogenin X/Y plus SRY) of loci for (a) female, (b) male and (c) female plus male (sample ratio: 1 to 1), respectively. Each sample contained 10 ng of DNA sample. The laser wavelength was 337 nm, with a laser fluence of 65 mJ cm^{-2}.

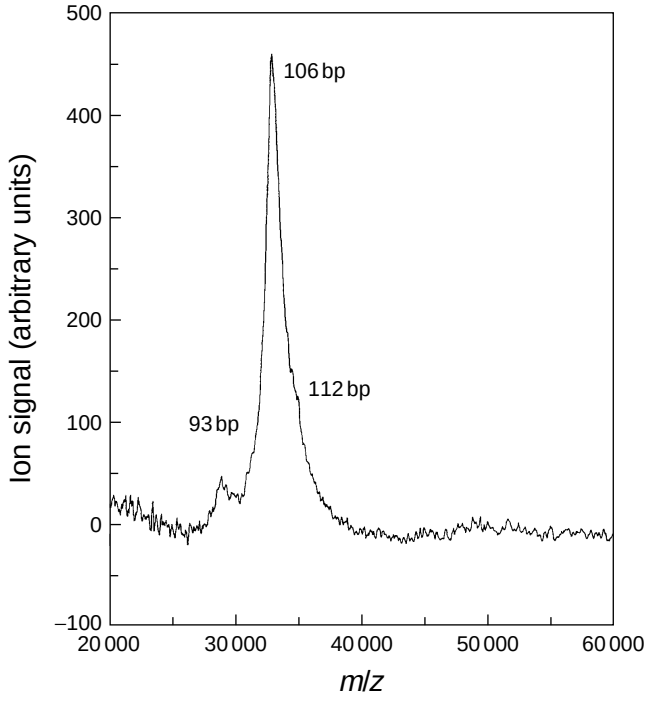

Figure 10 Negative ion mass spectrum of PCR amplified product of female (10 ng) DNA with 0.01 ng of male DNA.

generates only the 106 bp. As shown in Figure 9, the distinction of genotyping from male and female is quite clear. Figure 9 shows examples of mass spectra of PCR products for female gender, male gender, and a mixture of male and female gender. Of special interest in forensic analysis, especially for sexual assault, is the detection of a trace amount of male-specific DNA in a female sample. Such detection can be performed by demonstrating the presence of male-derived products of 93 bp and 112 bp. The MALDI spectra of composite sample containing mostly female species (10 ng) and only a trace amount of male sample (0.01 ng) can be detected (see Figure 10). It indicates that as little as 0.1% of male DNA in a female DNA sample can be detected. It also indicates that the detection sensitivity can reach 0.01 ng or less, which is about one order of magnitude better than gel electrophoresis.[39]

3 DNA SEQUENCING BY MASS SPECTROMETRY FOR FORENSIC APPLICATIONS

It is obvious that a complete sequencing of the human genome for each individual should reach the highest reliability for person identification. The sequencing of a selected DNA fragment should also be more reliable than just measuring STR or SNP. However, sequencing with gel electrophoresis is much more time-consuming and labor-intensive than just measuring the size of DNA. Furthermore, the sequencing process often needs the use of polyacrylamide gel instead of agarose as the gel matrix. Acrylamide requires more careful handling and generally takes longer to run. DNA sequencing by mass spectrometry is orders of magnitude faster than gels and can be done directly, without generating DNA ladders.[40] Thus, direct sequencing with mass spectrometry can be

just as fast as measuring the size of DNA fragments alone. At the present time, the low mass resolution of MALDI for DNA limits the size of DNA fragments which can be sequenced to approximately 200 bp. However, it is not always necessary to sequence a long piece of DNA for forensic application because a polymorphic profile is already known. Thus, DNA sequencing by mass spectrometry can become a valuable tool for forensic application in the future.

3.1 DNA Sequencing by Mass Spectrometry with DNA Ladders

DNA sequencing has been broadly used for biomedical research and clinical applications over the past two decades. Rapid and reliable DNA sequencing can also be very valuable for forensic applications. With the conventional sequencing approach, different sizes of DNA ladders, which are produced by either Sanger's enzymatic method[41] or Maxam–Gilbert's chemical cleavage method,[42] are separated by gel electrophoresis to achieve sequencing. With MALDI for sequencing, the speed can be significantly faster than gel electrophoresis. In addition, MALDI sequencing does not require labeling for identification, which saves both time and cost.

During the past few years, mass spectrometry for sequencing short DNA fragments has been pursued. In 1995, Shaler et al.[22] reported the first sequencing of a 45-mer ssDNA using MALDI/TOF with enzymatic preparation of DNA ladders. Recently, Roskey et al.[25] reported the results of sequencing a synthetic ssDNA template of 50 nucleotides. Koster et al.[24] also succeeded in sequencing synthetic ssDNA of 39 nucleotides using streptavidin-coated magnetic beads for purification of DNA ladders. Mouradian et al.[26] reported MALDI analysis of sequencing reactions from bacteriophage M13 and succeeded in determining the sequence for samples up to 35 bases in length. A complete sequencing effort for a ssDNA with 50 bases was demonstrated with no false stops, no serious fragmentation, and no unidentified peaks by Taranenko et al.[23] Recently, sequencing ssDNA higher than 100 nucleotides by using cycle sequencing to produce DNA ladders was reported.[27] Figure 11 shows the negative-ion mass spectra of DNA ladders from A, C, G and T reactions using dsDNA 130 bp as the template with reverse primer. Forward primers were also used for sequencing, with results similar to the data shown in Figure 11. Since sequencing by using both forward and backward primers can reach 120 mer, complete sequencing of 200 bp DNA can be achieved. Figure 12 illustrates the idea of combining the spectra from forward and backward primers to sequence dsDNA with a 200 bp template. Owing to the measurement of mass to identify the size of DNAs, there is no concern about a missing band, which can occur in the gel electrophoresis method. Thus, redundant sequencing can possibly be eliminated with MALDI DNA sequencing.

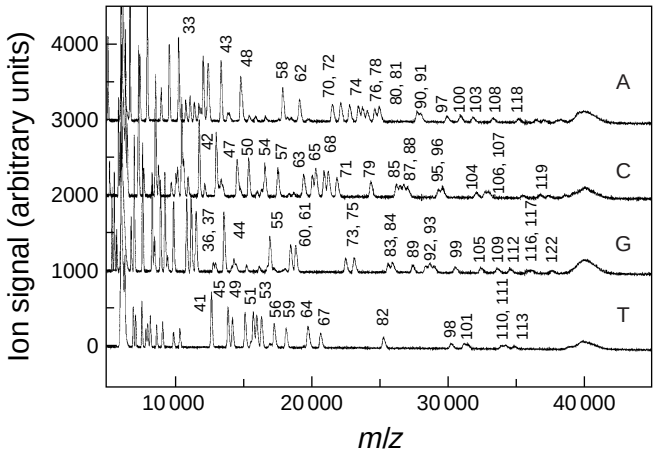

Figure 11 Negative ion mass spectra of Sanger sequencing ladders produced by dideoxynucleotide chain termination of the PCR product of a double-stranded 130 bp template with a 20 nucleotide reverse primer.

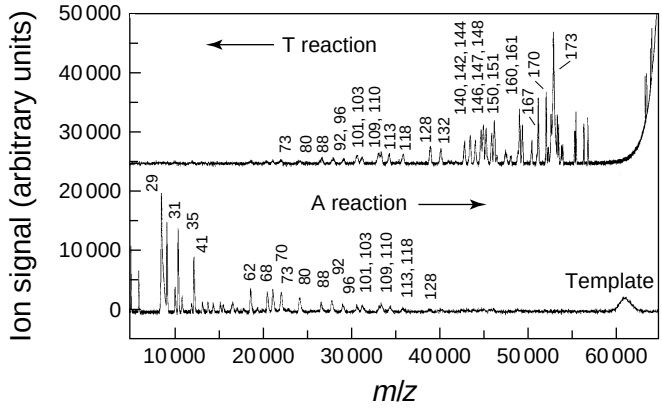

Figure 12 Negative ion mass spectra of Sanger sequencing ladders produced by dideoxynucleotide chain termination of the PCR product of a double-stranded 200 bp template. The results were from the A reaction with the reverse primer and the T reaction with the forward primer.

At present, automatic gel sequences can read up to 1000 bp, but most routine sequencing ranges from 300 to 500 bases. With MALDI now able to sequence 200 bp, mass spectrometry is emerging as a useful tool for DNA sequencing for forensic analysis, since most STR and SNP analysis does not need sequencing longer than 200 bp. Mass spectrometry sequence detection utilizes inexpensive unlabeled primers and also has the additional advantages of higher sequencing speed and the ability to obtain sequence information close to the primer.

3.2 Direct Sequencing

When using mass spectrometry for sequencing DNA with ladders, the time needed for a TOF mass spectrometer to separate and detect various ladders only takes a few seconds. However, the time needed to prepare DNA ladders needs many minutes to hours. Thus, the ideal sequencing method would be if mass spectrometry could be used to do direct DNA sequencing without the need to produce DNA ladders with chemical or enzymatic methods. This also implies that DNA ladders need to be produced during the laser ablation process.

Direct DNA sequencing can be easily understood by using Figures 13 and 14. Assuming the selective cleavage of the P–O bond at 3′-linkage for an oligonucleotide of 5′-CTGTGA-3′, the primary fragmentation of such bond breaking will produce two series of fragments. These are labeled as 5′- and 3′-termini (see Figure 13). Each series has six members ranging from 1- to 6-mers. The vertical dotted line in Figure 13 represents the position

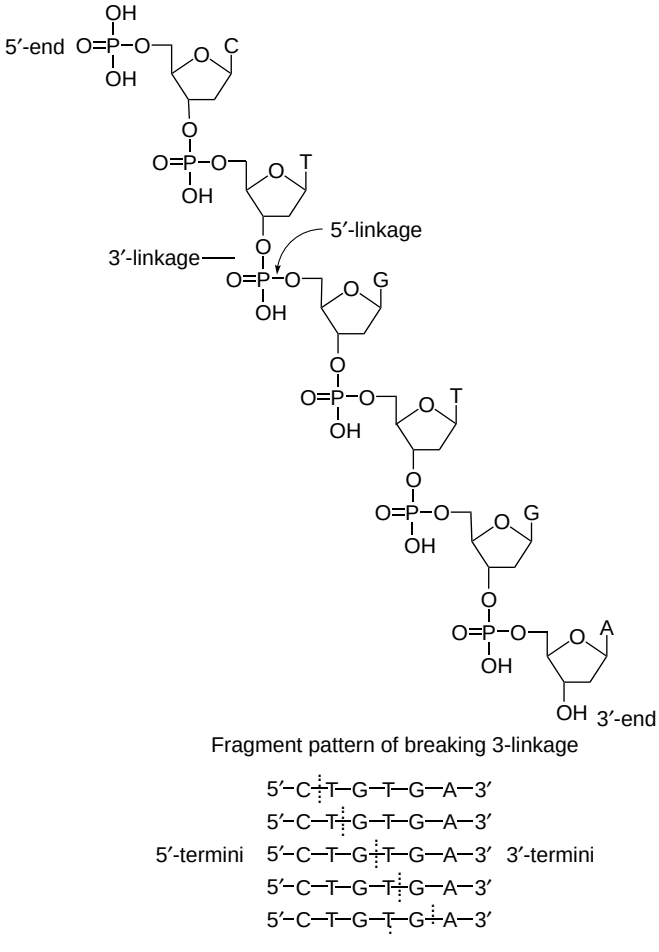

Figure 13 Illustration of the selective cleavage of the P–O bond and the fragmentation pattern for a 5′-CTGTGA-3′ oligonucleotide.

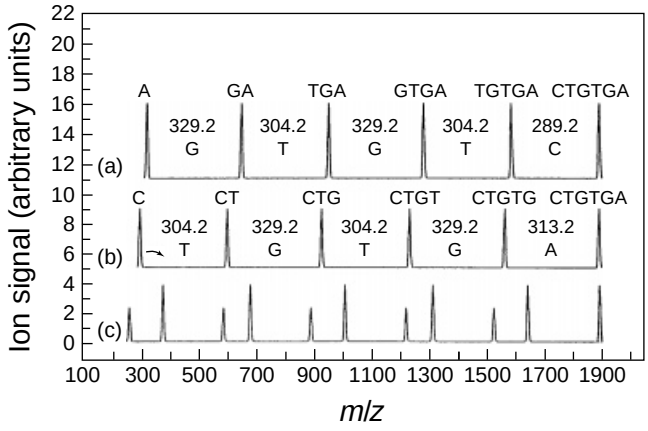

Figure 14 Simulated spectra of the selective cleavage of the P–O 3′-linkage and 5′-linkage for 5′-CTGTGA-3′ oligonucleotide; (a) simulated spectrum of fragment series with 3′-termini by the 3′-cleavage; (b) series with 5′-termini by 3′-cleavage; and (c) the combination of (a) and (b).

of cleavage. The fragments in both series can be ionized and resolved in MALDI spectra. Figure 14 gives the simulated mass spectra for these two series. Figure 14(a) shows the spectrum of the series with 3′-termini. The mass difference between two adjacent peaks in the series provides the information for each extra base. For example, the first member of this series is A, and the second member of this series is GA. Due to the selective cleavage of 3′-linkage, the mass difference between these two peaks will be exactly 329.2 Da, which represents the mass of dGMP. The sequence information can be obtained by analyzing all values of mass difference in this series, together with the total mass of 5′-CTGTGA-3′. The same information can also be obtained from the series with 5′-termini (Figure 14b). If a complete series (either 5′- or 3′-termini) of such fragmentation for a DNA sample can be experimentally resolved, the full sequence information will be obtained. Figure 14(c) shows a mass spectrum of two series from the cleavage of a 5′-linkage P–O bond. This sequencing method is very reliable, since the sequence result from one series can be reconfirmed by the other series. Recently, UV/MALDI was tried to achieve direct DNA sequencing by carefully adjusting the experimental conditions, which include the use of newly discovered two-component matrixes and adequate laser fluence.[43] Figure 15 shows the direct sequencing of DNA probes up to 35 nucleotides. Similar results by IR/MALDI,[44] and by electrospray high-resolution Fourier transform mass spectrometry[45] for short DNAs, have also been reported. At present, direct sequencing is still limited to short oligonucleotides, owing to the limited mass resolution, since the mass difference between different bases needs to be identified. If a 60-mer oligonucleotide is to be sequenced, the accuracy of the

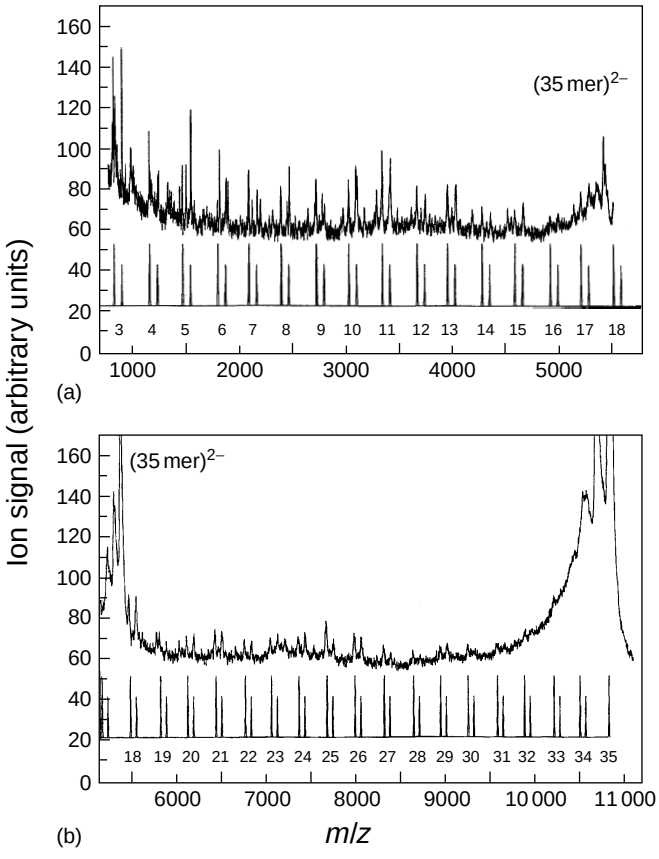

Figure 15 The mass spectrum of MALDI, with selective fragmentation for direct sequencing of a 35 mer oligonucleotide, 5′-GCGTGATGGAATCGATGACGTGCGATGTCGTGT-TT-3′. Simulated spectra are shown as the lower traces in (a) and (b). The high-intensity series in the simulated spectra is the 5′-termini and 3′-cleavage series and the lower-intensity series is the 3′-termini and 3′-cleavage series. A mixture of 2,4,6-trihydroxyacetophenone (1.5×10^{-5} mol) was used as a matrix. The quantity of 35 nucleotide DNA was 100 pmol. (a) and (b) are the low- and high-mass regions, respectively.

ion peaks needs to be better than 1 in 2000, since the mass difference between A and T is only 9 Da. When a DNA segment is longer than 100-mer, the effect of the abundance of the ^{13}C isotope makes the distinction between A and T extremely difficult. However, direct sequencing can still be applied to certain SNP samples when the sequence and the positions of possible variants are known and the segments to be sequenced can be short.

4 CONCLUSION

Mass spectrometry for DNA fingerprinting by VNTR, STR, SNP and sequence analysis is emerging as a new and powerful tool for forensic applications. Its special advantages include its much faster speed than conventional gel methods and its ability for detection without radioactive or dye tagging. In the near future, special effort should be put into improving mass resolution so that longer DNA samples can be used for STR, VNTR and SNP analysis. Better mass resolution can also extend SNP analysis and DNA sequencing to longer DNA segments for broader forensic applications. Robotic automation for sample preparation and analysis also needs to be developed to achieve high throughput analysis to obtain a database for future searches for suspects. Since the transport of DNA samples can sometimes lead to the contamination or deterioration of samples, it is desirable to have a transportable device for DNA analysis. Mass spectrometry for DNA analysis can be more convenient than gel electrophoresis to adapt for in situ and real time analysis as well as for automation for high throughput analysis.

ACKNOWLEDGMENTS

This work was primarily supported by the National Institute of Justice, grant number 97-LB-VX-A047, and supported in part by the Office of Biological and Environmental Research, US Department of Energy, under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

ABBREVIATIONS AND ACRONYMS

ASPCR	Allele-specific Polymerase Chain Reaction
bp	base pairs
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
DRPLA	Dentatorubral–Pallidoluysian Atrophy
dsDNA	Double-stranded DNA
EDTA	Ethylenediamine Tetraacetic Acid
HW	Hardy–Weinberg
IR	Infrared
MALDI	Matrix-assisted Laser Desorption/Ionization
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
SNP	Single-nucleotide Polymorphism
ssDNA	Single-stranded DNA
STR	Short Tandem Repeat
TOF	Time-of-flight
UV	Ultraviolet
VNTR	Variable Numbers of Tandem Repeat

RELATED ARTICLES

Forensic Science **(Volume 5)**
Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

REFERENCES

1. D. Botstein, R.L. White, M. Skolnick, R.W. Davis, 'Construction of a Genetic Linkage Map in Man using Restriction Fragment Length Polymorphisms', *Am. J. Human Genet.*, **32**(3), 314–331 (1980).
2. A.J. Jeffreys, V. Wilson, S.L. Thein, 'Hypervariable Minisatellite Regions in Human DNA', *Nature*, **314**, 67–73 (1985).
3. Y. Nakamura, M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, R. White, 'Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping', *Science*, **235**, 1616–1622 (1987).
4. E.A. Fischer, L.R. Paulson, M.L. Straf, J.R. Tucker, P.A. Adams, M. Grossblatt, P. Gilman, J.F. Crow, M.A. Berger, S.S. Diamond, D.H. Kaye, H.H. Kazazian, A.G. Motulsky, T.A. Nagylaki, M. Nei, G.F. Sensabaugh, D.O. Siegmund, S.M. Stigler, T.D. Pollard, P.J. Bickel, M.T. Clegg, B. Singer, *The Evaluation of Forensic DNA Evidence*, National Research Council, National Academy Press, Washington, DC, 23, 1996.
5. L. Rowen, G. Mahairas, L. Hood, 'Sequencing the Human Genome', *Science*, **278**, 605–607 (1997).
6. R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, 'Enzymatic Amplification of Beta Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickel Cell Anemia', *Science*, **230**, 1350–1354 (1985).
7. L.D. Mueller, 'Population Genetics of Hypervariable Human DNA', in *Forensic DNA Technology*, eds. M.A. Farley, J.J. Harrington, Lewis Publishers, Inc., Chelsea, MI, 51–62, 1991.
8. N. Ip, I. van de Stadt, Z. Loewy, S. Leary, K. Grzechik, I. Balazs, 'Identification and Characterization of a Hypervariable Region[D18S27] on Chromosome 18', *Nucleic Acids Res.*, **17**, 4427 (1989).
9. D.C. Schwartz, C.R. Cantor, 'Separation of Yeast Chromosome-size DNA by Pulsed Field Gradient Gel Electrophoresis', *Cell*, **37**, 67–75 (1994).
10. A.L. Burlingame, T.A. Baillie, D.H. Russell, 'Mass Spectrometry', *Anal. Chem.*, **64**, 467R–502R (1992).
11. K.B. Jacobson, H.F. Arlinghaus, 'Development of Resonance Ionization Spectroscopy for DNA Sequencing and Gene Mapping', *Anal. Chem.*, **64**, 315A–328A (1992).
12. M. Karas, D. Beckmann, U. Bahr, F. Hillenkamp, 'Matrix-assisted Ultraviolet Laser Desorption of Non-volatile Compounds', *Int. J. Mass Spectrom. Ion Process.*, **78**, 53–68 (1987).
13. R.B. Jones, S.L. Allman, K. Tang, W.R. Garrett, C.H. Chen, 'Neutralization of Negatively Charged Oligonucleotide', *Chem. Phys. Lett.*, **212**, 451–457 (1993).
14. M.M. Siegel, I.J. Hollander, P.R. Hammann, J.P. James, L. Hinman, B.J. Smith, A.P.H. Farnsworth, A. Phipps, D.J. King, M. Karas, A. Ingendoh, F. Hillenkamp, 'Matrix-assisted UV–Laser Desorption/Ionization Mass Spectrometric Analysis of Monoclonal Antibodies for the Determination of Carbohydrate, Conjugated Chelator and Conjugated Drug Content', *Anal. Chem.*, **63**, 2470–2481 (1991).
15. D.M. Schieltz, C. Chou, C. Luo, R.M. Thomas, P. Williams, 'Mass Spectrometry of DNA Mixtures by Laser Ablation from Frozen Aqueous Solution', *Rapid Commun. Mass Spectrom.*, **6**, 631–636 (1992).
16. L. Zhu, G.R. Parr, M.C. Fitzgerald, C.M. Nelson, L.M. Smith, 'Oligonucleotide Fragmentation in MALDI/TOF Mass Spectrometry Using 355-nm Radiation', *J. Am. Chem. Soc.*, **117**, 6048–6056 (1995).
17. K.J. Wu, A. Steding, C.H. Becker, 'Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry of Oligonucleotide using 3-Hydroxypicolinic Acid as an Ultraviolet-sensitive Matrix', *Rapid Commun. Mass Spectrom.*, **7**, 142–146 (1993).
18. K. Tang, N.I. Taranenko, S.L. Allman, L.Y. Chang, C.H. Chen, 'Detection of 500-nucleotide DNA by Laser Desorption Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 727–730 (1994).
19. S. Berkenkamp, F. Kirpekar, F. Hillenkamp, 'Infrared MALDI Mass Spectrometry of Large Nucleic Acid', *Science*, **281**, 266–268 (1998).
20. L.Y. Chang, K. Tang, M. Schell, C. Ringelberg, K.J. Matteson, S.L. Allman, C.H. Chen, 'Detection of ΔF 508 Mutation of the Cystic Fibrosis Gene by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 772–774 (1995).
21. N.I. Taranenko, K.J. Matteson, C.N. Chung, Y.F. Zhu, L.Y. Chang, S.L. Allman, L. Haff, S.A. Martin, C.H. Chen, 'Laser Desorption Mass Spectrometry for Point Mutation Detection', *Genetic Analysis: Biomol. Eng.*, **13**, 87–94 (1996).
22. T.A. Shaler, Y. Tan, J.N. Wickham, K.J. Wu, C.H. Becker, 'Analysis of Enzymatic DNA Sequencing Reactions by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 537–540 (1995).

23. N.I. Taranenko, C.N. Chung, Y.F. Zhu, S.L. Allman, V.V. Golovlev, N.R. Isola, S.A. Martin, L.A. Haff, C.H. Chen, 'Matrix-assisted Laser Desorption/Ionization for Sequencing Single-stranded and Double-stranded DNA', *Rapid Commun. Mass Spectrom.*, **11**, 386–392 (1997).
24. H. Koster, K. Tang, D. Fu, A. Braun, D. van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, 'A Strategy for Rapid and Efficient DNA Sequencing by Mass Spectrometry', *Nature/Biotechnology*, **14**, 1123–1128 (1996).
25. M.T. Roskey, P. Jukasz, J.P. Smirnov, E.J. Takach, S.A. Martin, L.A. Haff, 'DNA Sequencing by Delayed Extraction-matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **93**, 4724–4729 (1996).
26. S.D. Mouradian, R. Rank, L.M. Smith, 'Analyzing Sequencing Reactions from Bacteriophage M13 by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass. Spectrom.*, **10**, 1475–1478 (1996).
27. N.I. Taranenko, S.L. Allman, V.V. Golovlev, N.V. Taranenko, N.R. Isola, C.H. Chen, 'Sequencing DNA Using Mass Spectrometry for Ladder Detection', *Nucleic Acids Res.*, **26**, 2488–2490 (1998).
28. Committee on DNA Technology in Forensic Science, National Research Council, *DNA Technology in Forensic Science*, National Academy Press, Washington, DC, 1992.
29. Y.H. Liu, J. Bai, Y. Zhoe, Xiao Liang, D. Siemieniak, P. Venta, D.M. Lubman, 'Rapid Screening of Genetic Polymorphisms using Buccal Cell DNA with Detection by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 735–743 (1995).
30. Y. Wada, M. Yamamoto, 'Detection of Single-nucleotide Mutations Including Substitutions and Deletions by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 1657–1660 (1997).
31. D.P. Little, T.J. Cornish, M.J. O'Donnell, A. Braun, R.J. Cotter, H. Koster, 'MALDI on a Chip: Analysis of Arrays of Low-femtomole to Subfemtomole Quantities of Synthetic Oligonucleotides and DNA Diagnostic Products Dispensed by a Piezoelectric Pipet', *Anal. Chem.*, **69**, 4540–4546 (1997).
32. N.I. Taranenko, V.V. Golovlev, S.L. Allman, N.V. Taranenko, C.H. Chen, J. Hong, L.Y. Chang, 'Matrix-assisted Laser Desorption/Ionization for Short Tandem Repeat Loci', *Rapid Commun. Mass Spectrom.*, **12**, 413–418 (1998).
33. J.M. Butler, B.R. McCord, J.M. Jung, R.O. Allen, 'Rapid Analysis of Short Tandem Repeat HUMTH01 by Capillary Electrophoresis', *BioTechniques*, **17**, 4–7 (1994).
34. N.I. Taranenko, N.T. Potter, S.L. Allman, V.V. Golovlev, C.H. Chen, 'Detection of Trinucleotide Expansion in Neurodegenerative Disease by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Genetic Analysis: Biomol. Eng.*, **15**, 25–31 (1999).
35. J. Riordan, J.M. Rommens, B. Kerem, M.L. Drumm, N. Alon, R. Rozmachel, Z. Grzelczrk, J. Zielenski, S. Lok, N. Plavsic, J.L. Chou, M. Drumm, M.C. Iannuzzi, F.S. Collins, L.-C. Tsui, 'Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complimentary DNA', *Science*, **245**, 1066–1073 (1989).
36. L.A. Haff, I.P. Smirnov, 'Single-nucleotide Polymorphism Identification Assays Using a Thermostable DNA Polymerase and Delayed Extraction MALDI/TOF Mass Spectrometry', *Genome Res.*, **7**, 378–388 (1997).
37. G.M. Church, S. Kieffer-Higgins, 'Multiplex DNA Sequencing', *Science*, **240**, 185–188 (1988).
38. F.R. Santos, A. Pandya, C. Tyler-Smith, 'Reliability of DNA-based Sex', *Nature Genetics*, **18**, 103 (1998).
39. N.I. Taranenko, N.T. Potter, S.L. Allman, V.V. Golovlev, C.H. Chen, 'Gender Identification by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **71**(18), 3974–3976 (1999).
40. Y.F. Zhu, N.I. Taranenko, S.L. Allman, N.V. Taranenko, S.A. Martin, L.A. Haff, C.H. Chen, 'Oligonucleotide Sequencing by Fragmentation in Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass. Spectrom.*, **11**, 897–903 (1997).
41. F. Sanger, S. Nicklen, A.R. Coulson, 'DNA Sequencing with Chain-terminating Inhibitor', *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467 (1977).
42. A.M. Maxam, W. Gilbert, 'A New Method for Sequencing DNA', *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 560–564 (1977).
43. Y.F. Zhu, N.I. Taranenko, S.L. Allman, S.M. Martin, L. Haff, C.H. Chen, 'The Effect of Ammonium Salt and Matrix in the Detection of DNA by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 1591–1596 (1996).
44. E. Nordhoff, M. Karas, R. Cramer, S. Hahner, F. Hillenkamp, F. Kirpekar, A. Lezins, J. Muh, C. Meier, J.W. Engels, 'Direct Mass Spectrometric Sequencing of Low Picomole Amounts of Oligodeoxynucleotides with up to 21 Bases by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *J. Mass Spectrom.*, **30**, 99–112 (1995).
45. D.P. Little, R.A. Chorush, J.P. Speir, M.W. Senko, N.L. Kelleher, F.W. McLafferty, 'Rapid Sequencing of Oligonucleotides by High-resolution Mass Spectrometry', *J. Am. Chem. Soc.*, **116**, 4893–4897 (1994).

Microspectrophotometry in Forensic Science

Janina Zięba-Palus
Institute of Forensic Research, Cracow, Poland

1 Introduction **1**
2 Analysis of Chemical Contents by Infrared Microspectrometry **2**
2.1 History 2
2.2 Technique Evaluation 2
2.3 Microspectrometry in Forensic Work 4
2.4 Raman Microspectroscopy 8
2.5 Databases and Libraries 9
3 Color Analysis by Microspectrophotometry **9**
3.1 History 9
3.2 Spectral Measurements and Analysis 10
3.3 Application in Examinations 10
3.4 Spectral Analysis Using Ultraviolet Radiation and Fluorescence 11
Acknowledgments **11**
Abbreviations and Acronyms **11**
Related Articles **12**
References **12**

Infrared microspectrophotometry is a combination of two techniques – optical microscopy and infrared (IR) spectrometry – which supply specific information about the composition and structure of a given material. The equipment used is a Fourier transform infrared (FTIR) spectrometer connected on line with an optical microscope.

The microscope allows observation of a sample in white light at relatively high magnification; this allows its morphology and microstructure to be established, and facilitates selection of the area which will subsequently be subjected to IR spectrometric analysis. The microscope also permits use of polarized light in investigations.

The spectrometer enables analysis of a sample with a chosen measuring technique – transmission or reflection – with the aim of determining the chemical composition of the sample. It also supplies information about its microstructure and orientation.

This method is particularly useful in the analysis of trace amounts of various substances secured as material evidence in court cases, and in analysis of the homogeneity of the sample, identification of inclusions and contaminants on the surface, and detection of structural defects.

A specific advantage of this method is its ability to unambiguously photograph and record marked and measured areas of the sample and minimize the process of preparing the sample for analysis.

Furthermore, IR microspectrophotometry gives the rare opportunity of studying small crystals or areas in the oriented materials by using polarized IR radiation, both by the transmission and the reflection technique.

Its fundamental drawback is that the physical nature of the microsample may influence the precision of the photometric measurements and cause distortion of the spectra (artefacts) obtained. Furthermore, because only a small area is studied the heterogeneity of the sample and the amount of contaminants may significantly influence the results of spectrometric measurements.

Microspectrophotometry in the ultraviolet/visible (UV/VIS) range is a combination of photometric measuring techniques and optical microscopy. It allows comparison of the color of very small samples of various materials, such as single fibers, tiny amounts of paint, traces of ink or ball-point pen ink on a forged document in an objective way, independently of the sharpness and quality of the observers's vision. It yields immediate information about spectral differences existing between two samples of similar color, which are indistinguishable by use of the optical microscope. Additionally, using appropriate software to analyze the results of microspectrometric measurements, it enables precise measurement of a color, and – applying the theory of colors – gives a color a defined numerical value; this facilitates communication between experts working with this method. The equipment consists of an optical microscope with a spectrometer for analysis in the visible and ultraviolet range, connected to a microcomputer via an analog-to-digital converter.

1 INTRODUCTION

Microspectrometry is a new method applied in chemical microanalysis which has simplified the procedure for preparing and studying samples in a fundamental way.

Microspectrometry is a combination of optical microscopy and spectrometry. Microscopy deals with creating, recording, and interpreting magnified images, and spectrometry makes use of emission, absorption, and reflection of radiation energy through materials to determine their structure, properties, and composition. Both methods have been developed from optics and the action of radiation energy on matter. Microspectrometry is divided into IR and UV/VIS, according to the type of energy used.

IR microspectrometry enables determination and comparison of composition of examined samples, whereas

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

UV/VIS microspectrometry serves to establish and compare colors of analyzed samples in an objective way, independent of the observer and the quality and sharpness of human vision and perceived color.

The fundamental advantages of microspectrometric techniques are the possibility of studying relatively small amounts of sample, eliminating the burdensome process of preparing the sample for study and the ability to repeat measurements without destroying the sample. These advantages have led to these methods being used today in most forensic laboratories in the study of traces revealed at the crime scene.

2 ANALYSIS OF CHEMICAL CONTENTS BY INFRARED MICROSPECTROMETRY

2.1 History

IR microscopes for dispersive spectrometers were developed in several academic laboratories in the late 1940s and early 1950s but were used by only a handful a researchers.[1] Even when Perkin-Elmer introduced a commercial unit in 1953, few people applied it in research work. In the late 1970s, FTIR spectrometers were developed. The virtues of the new type of instruments, such as higher resolution, lower noise and better frequency accuracy than that of the ubiquitous dispersive instruments, allow the application of the microscope attachment.

However, the IR microscope attached to the FTIR spectrometer tends to restrict throughput (Jacquinot). The optical throughput of an optical system is described by Equation (1),[1]

$$\theta_x \theta_s = A\Omega_s \qquad (\mathrm{cm}^2\,\mathrm{sr}) \tag{1}$$

where A is the area and Ω_s is the solid angle subtended by the limiting aperture of the system (in the FTIR microscope, this would be the area-defining aperture). Because a microscope is by definition used for small samples, the sample area contribution to the throughput calculation is reduced. So, one tries to increase the solid angle as much as possible, in order to improve signal and spatial resolution. One normally still ends up in a region where more optical throughput is attained than by a dispersive spectrometer.

By 1982, the sensitivity of IR spectroscopy had been developed to a point where low-energy and microscopic techniques were again being tried. Today, FTIR microscopy is the chosen technique for a wide range of sampling problems. The first microscopes only enabled transmission spectra to be obtained. Today one can also register reflection spectra. The type of sample studied usually determines the study technique chosen.

2.2 Technique Evaluation

It would be a gross oversimplification to claim that all you need to do to perform IR microspectrometry is to place the studied sample on the microscope stage in a beam of IR radiation and carry out the measurements.[2] Because all studied samples are unique, especially those which are the subject of forensic investigations, the optimum method of analysis must be used, encompassing the following: choice of the appropriate microscopy technique, preparation of the sample to be studied by that technique, selection of instrument, study of the microstructure and choice of area of sample for analysis, measurement of the spectrum in chosen area, interpretation of results of analysis and correlation of these results with the microstructure of the sample.

To gain a good quality IR spectrum two measuring conditions must be fulfilled:

- IR radiation energy reaching the detector must be very high (high throughput);
- the area of the sample analyzed spectrometrically must be strictly defined.

Nonfulfillment of these conditions leads to lowering of the signal-to-noise ratio (S/N) which leads to such phenomena as radiance, diffraction, and aberration, causing optical noise. A similar thing happens in the case of irregular lighting of the field of vision and measurement. It is easy to define the area of a sample which interests the analyzer, using a beam of white light. The area which will be analyzed by IR is larger as a consequence of diffraction of the IR radiation. Consequently, with a smaller sample, an exact delineation of the area for spectrometric research becomes critical in order to achieve a good spectrum. A badly defined area of the measured sample results in a bad spectrum. In such a situation apertures that limit the field of vision are helpful, being located between the source of radiation and the sample, and also between the sample and the detector.

Initially, IR spectrometry was applied to the analysis of very small (nanogram) samples. Today the method is frequently used in the analysis of small areas in large samples, because they can be isolated selectively using apertures, without the need to prepare additional specimens. Measurements can be carried out by reflection and transmission techniques.

Measurement of reflection spectra in IR microspectrometry is the best method with which to analyze surfaces, thin films, and strongly reflecting or dispersed materials. It does not require preparation of the sample for study and does not cause destructive changes to the substance studied. The spectra obtained by reflection do, however, generally differ from transmission spectra in

shape, outline, and intensity of bands. It is therefore necessary to correct the spectra using the Kramers–Kronig transformation. Depending on the type of sample studied the following reflection techniques can be applied:

- specular reflection
- diffuse reflection
- attenuated total reflection (ATR)
- reflection–absorption.

The transmission technique is characterized by excellent photometric accuracy and introduces the least spectral interference during measurements, so it has been applied in quantitative analysis. However, its fundamental drawback is the necessity to prepare samples for study. The aim of preparation is to separate the studied sample from the surface on which it is found and to obtain the appropriate thickness. Many traditional techniques applied in the preparation of samples for microscopic investigations can be used unchanged, or with only slight modification in IR microspectrometry.

A thin sample is obtained for transmission investigations by cutting the specimen, for example by using a microtome, abrasion and polishing, crushing, flattening through rolling out, or the use of a diamond cell anvil and grinding and pressing tablets with potassium bromide (if the morphology of the sample is not interesting). The ideal sample for transmission investigations is transparent, 1–10 μm thick, and positioned on a KBr plate to eliminate potential sources of error. In some cases transmission of IR radiation is accompanied by specular and/or internal reflectance. This effect can be eliminated by arranging the sample between two KBr windows or immersing it in immersion oil. In the case of samples suspended in KBr, if the size of the sample grains is comparable with the length of the IR radiation wave, then scattering radiation occurs.

The quality of the spectrum gained by IR microspectrometry may be influenced by the thickness of the sample, the averaging of the signal and the optical resolution.[3–5]

The S/N may be improved by changing the thickness of the sample, or by performing measurements for a large area of sample or for a thick sample, so that the absorption bands have an intensity of about 0.8. The level of noise can be reduced by carrying out many measurements and averaging them as the S/N is proportional to time spent conducting measurements. In routine studies a resolution of 8 cm^{-1} is most commonly applied.

Sometimes, good quality spectra can be obtained by the microscope technique for nanogram amounts of samples, i.e. amounts which cannot be analyzed by other IR spectrometric techniques. However, one must be aware that the physical nature of microsamples may cause distortion of the absorption bands or influence the photometric accuracy. Thus, minute pieces that are analyzed microtransmissionally, such as individual bits of paint or fabric, have their own geometric shapes that will influence the spectral data collected. The effect of dispersed light is visible on these spectra, especially clearly in the case of strong bands. It is then necessary to flatten the sample, reduce the size of the aperture or analyze by microreflection.

As a result of radiation dispersion, thick-grained powder samples give spectra with a sloped baseline, especially in the case of shorter waves. It is then advisable to correct of baseline. However, as a result of the Christiansen effect in the fingerprint area, bands appear with sides going down below the baseline. Then it is essential to prepare tableted samples with KBr. A problem that occurs when optically analyzing nontransparent tablets is stray light effects. In such situations the thickness of the sample must be reduced.

Every feature of a spectrum – the size, shape, and position of peaks, and even the nature of the baseline – contains information concerning the sample. Not every feature, however, is related to the chemical structure of the sample, as some arise from optical phenomena. Therefore spectra should be obtained and examined with a view to identifying these spectral artefacts (interference patterns, stray radiation) so that in comparative work the possibility of a false negative conclusion being drawn is minimized.

The recently introduced ATR microscope objective has the significant advantage of allowing the analyst to view the specimen directly where the measurements are taken.[4] This method is nondestructive and generally allows the analysis of specimens in situ, on evidence such as garments or documents.

Another advantage of analysis by IR microspectrometry is the possibility of examining microscopic contaminants in chemical systems and observing the homogeneity of the composition of the studied sample both on the surface (by way of mapping) and deep within the sample (by so-called depth profiling).[6] For these techniques to be performed it is essential for the microscope to possess a motorized stage, allowing automatic moving of the sample by a defined distance over several micrometers. Measurements can then be performed for a chosen defined area of the sample at set intervals, such as every 10 μm, effectively providing a series of IR spectra. These spectra can be presented on a three-dimensional system of coordinates, namely the wavelength, absorption, and position of the sample. This gives a three-dimensional array of spectra, illustrating the variability of composition of the studied area of the sample. On this basis a compositional contour map can be prepared, which ensures that

a more exact picture of the distribution of a selected component in the sample is gained. This procedure is often used in studies of heterogeneity or defects of structure of polymer laminates.

2.3 Microspectrometry in Forensic Work

The basic advantages of IR microspectrometry are:

- use of picogram or smaller samples;
- unambiguous identification and photographing of measured areas of sample;
- archivization of measured areas;
- elimination of the burdensome process of preparing the sample for research;
- repetition of measurement (nondestructive analysis).

These make the technique especially useful to forensic studies, such as in the identification of microtraces, namely paint chips, fragments of polymers, rubber, glue, and single fibers. This type of material is secured for study from the scene of the event, most frequently a road accident, break-in, rape, or homicide. This technique is also used in the analysis of chemical composition of inks and toners, used in the forging of documents.

2.3.1 Examination of Paint Fragments and Laminates

The authoritative reviews are given by Messerschmidt[7] and Humecki.[8] Ryland[9] and Suzuki[10] have published comprehensive bibliographies of applications, including paint examination.

In routine comparative examinations of paints, transmission IR techniques are mainly used[11–16] to provide information about binders, pigments and additives used in various types of coatings. Because usually only a small amount of sample is available for study, up to now little bits of paint have been prepared for study using microtableting methods with KBr or a diamond cell anvil. The IR beam is focused on the microtablets or the diamond cell using a beam condenser. Nowadays, a microscope attachment is used.[17–28] IR microspectrometry is one of the most valuable tools of the forensic paint examiner. It allows highly reproducible nondestructive individual layer comparisons with a very limited time investment.

Microfragments of car paint found on a road or revealed on the clothes of a victim of a traffic accident are usually multilayer fragments of paint. A sample of an individual layer may be prepared manually using a scalpel, blades, or needles, and then placed on a salt plate for analysis. Samples also can be obtained for spectrometric investigations by cross-sectioning the multilayer coating. A fragment of paint may be cut into

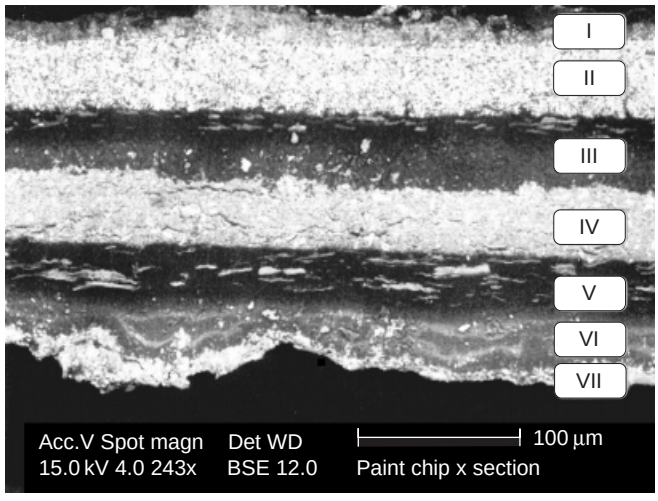

Figure 1 Cross-section of a paint chip (magnification 300×). (Reproduced by permission of the Institute of Forensic Research in Cracow from J. Zięba-Palus, 'Identification of Paint Fragments', *Probl. Forens. Sci.*, **XXXII**, 114–118 (1995).)

slices 2–5 µm thick, using a microtome perpendicular to the surface. The specimen gained in this way is then placed on the microscope stage and its morphology is observed, i.e. the number, thickness, color, and grain of the layers visible on a cross-section of the fragment. Then, limiting the field of vision with the microscope aperture, the fragment's IR spectrum is determined for only one layer using the transmission technique (Figures 1 and 2). Comparison of the spectrum with library spectra provides information about the type of paint in the layer, i.e. the type of resins occurring in the binder and the basic pigments. This same specimen may be used in further studies, such as with a scanning microscope linked to an X-ray microspectrometer for determining the elemental composition of particular layers, and for identifying all inorganic components (pigments, fillers). Instead of preparing the specimen using the microtome, fragments of paint are flattened in the diamond cell anvil, and then positioned on the stage of the microscope in the path of the IR radiation. Samples of polymer laminate are prepared for research and analyzed in a similar way.

Paint and polymer microtraces transferred from a car to the clothes of victims of traffic accidents are studied using the IR microspectrophotometry method. They are visible macroscopically as colored streaks. However in the microscopic image, bits of material can be observed which have been ''rubbed in'' or melted in-between fibers of the fabric. For identification of the transferred material and possibly comparison with material taken from the car of the person suspected of participating in the accident, it is necessary to separate microflakes from the surface. Microflakes that have been separated out are positioned directly on the KBr plate lying on

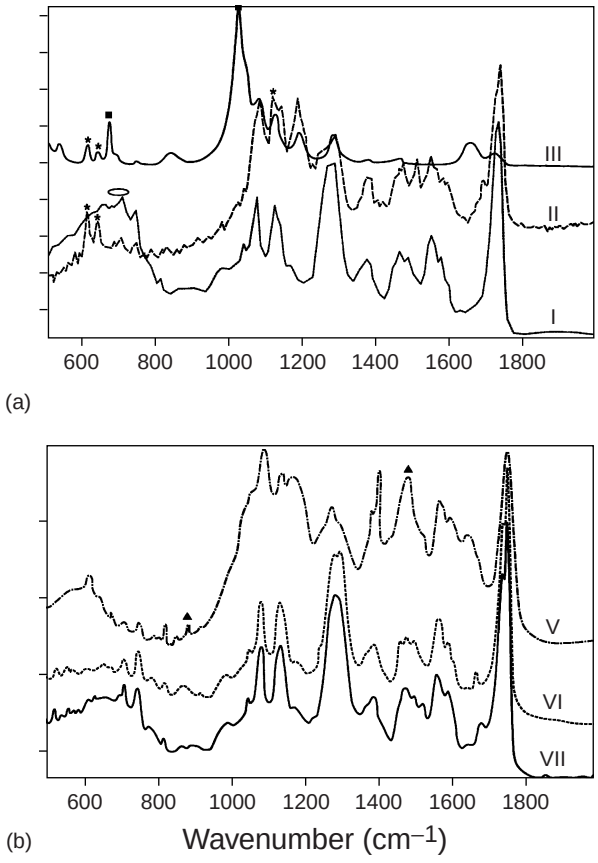

Figure 2 IR spectra of particular layers, visible on the cross-section of paint chip. Absorption bands of pigments: sulfate, ■ talcum, ⬭ titanium dioxide, carbonate. (Reproduced by permission of the Institute of Forensic Research in Cracow from J. Zięba-Palus, 'Identification of Paint Fragments', *Probl. Forens. Sci.*, **XXXII**, 114–118 (1995).)

the microscope stage and a spectrum is collected in transmission mode. Study of the microtraces by the reflection technique directly on the fabric frequently does not give satisfactory results, because the absorption bands of the fabric may mask the absorption bands of the material being investigated.

Transfers of coatings resulting in smears on other substrata may be sampled in situ using an ATR objective on an IR microscope. As a control, the substratum should also be analyzed to verify that its absorption bands are not contributing to the spectrum of the smear.

In comparative studies of paint samples, the reproducibility of the method of study is of fundamental significance.[9] If it is weak then differences observed between spectra of samples of the same paint coat (resulting from its natural variability) can be interpreted as significant differences, wrongly indicating that the analyzed samples were different and originating from various objects. Conversely, if the differences between spectra of samples of different paint coats (variability between samples) are of the same magnitude as the differences in spectra obtained for the same sample, then one might wrongly conclude that the samples studied were identical and so originating from one source.

This technique is in most cases repeatable but, because it is in principal a point analysis, it is necessary to carry out several measurements for several micro-areas of the same sample. It is worth adding that the reproducibility of the measurements is influenced by the fact that the paint is heterogeneous. This has no significance in analysis of macrosamples, but if measurements are carried out for microfragments when using an aperture smaller than $100\,\mu m^2$ it becomes critical. Reproducibility depends on the technique of preparing the sample for measurement. The most popular method of flattening a sample, by pressing and squashing it, introduces orientation in polymer molecules and changes their crystallinity. There is no one single universal technique for preparing a sample. It depends on the type of sample. For example, problems encountered when studying paint transfers can be due to difficulties in separating the sample from the surface on which it was found. Subtraction of spectra usually does not provide good results.

Using microspectrophotometry, both the binder and pigment portions of the paint are represented in the data obtained, allowing quick eliminations without the use of multiple instrumental techniques. Microspectrophotometry is quite discriminating for most binder types, especially those employing copolymer modifications. This technique is also sensitive to changes in organic pigment composition. However, IR microspectrometry has its limitations and should not be the sole instrumental technique used for a full comparison of the evidence and a standard sample. IR microspectrometry does not provide sensitive detection of minor components. Constituents present in concentrations less than 5% will typically go unnoticed. The problem of poor sensitivity to minor components is not limited to organic compounds. Thus broad band nonorganic pigments, which are present in low concentrations, will be masked by the major inorganic extenders or organic binders, especially in the region $700–2000\,cm^{-1}$. Furthermore, there are some types of binders whose IR spectra change very little from one to another. The presence of high concentrations of extender pigments can cause problems in binder comparisons and classifications. Additionally, not all extender pigments have similar compositions, yet different crystalline forms can be differentiated by IR spectroscopy.

In summary, IR microspectrometry is a fast, nondestructive technique allowing individual layer comparisons on casework-sized samples. It can provide information about both the binder and the pigment portions of a cured paint specimen, as well as their relationship to one

another. It possesses the ability to discriminate between a great number of paints having similar colors and similar binders.

2.3.2 *Analysis of Single Fibers*

Microspectrometry is used in analysis of microfibers conveyed from fabric to fabric or other surface as a result of contact between the two. For a single fiber of length 1 mm one can obtain an interpretable IR spectrum by the transmission technique, on condition that the fiber is sufficiently thin and transparent for IR radiation.[29–37]

The way in which a fiber is prepared influences the fine detail of the spectra obtained.[38] Squashing of the fiber is essential to ensure that sufficient energy is transmitted through the fiber substance. To get reproducible results the technique of squashing needs to be reproducible. Thus, in any one laboratory it will be necessary to build up a library of spectra from authenticated samples against which spectra from case materials can be empirically compared.

There are many methods of preparing and mounting fiber samples for analysis by IR microspectrometry. The choice of method depends on the individual analyst, the particular requirements of the laboratory, and whether or not subsequent analyses may be required.

Unflattened fibers may be mounted across an aperture or placed on a KBr window and analyzed directly. The aperture may be a hole in a metal disk with the fiber mounted on double-sided adhesive tape. This method has two major disadvantages. First, fibers greater than 50 μm in diameter are generally too thick for obtaining good-quality spectral data. Fibers with circular cross-sections have a variable path length and act like a lens because the beam passes through the sides and the center. Under these conditions the ratios of less intense bands to stronger bands are greater than those obtained with a uniform path length. Second, circular and irregular cross-sectional shapes tend to diffract and scatter incident IR radiation. In general, the use of unflattened fibers is good for obtaining initial spectra of weaker crystalline bands. The fiber should subsequently be flattened by some means to obtain a good overall spectrum. Note that crystallinity changes may occur during fiber flattening and the spectra may be slightly distorted.

Flattening the fibers prior to analysis offers several advantages. Deviations from Beer's law are reduced because samples have more uniform thickness. The reduction in path length is also important for fibers possessing high absorptivity. Additionally, flattening increases the surface area of the sample available for analysis, thereby enhancing S/N while reducing diffraction effects at the fiber edges.

Fibers may be flattened with a roller knife or between plates in a laboratory press. Fibers may be placed between two KBr disks and compressed with a compression cell or a diamond cell anvil may be used, which involves sandwiching a sample between diamond windows.

However, flattening destroys the physical shape of the fiber. In addition, the application of pressure can change the crystalline composition of the polymer. Although no major spectral differences have been observed as a result of applying pressure to fiber samples, minor changes in peak frequencies, intensities, and shapes do occur for certain fibers.

Flattening fibers on a hard surface also effectively produces a thin film and may result in interference fringes in the IR spectra. These fringes do not interfere with qualitative interpretation of the spectra, but they can affect quantitative analysis and comparisons made with large databases.

The polymer composition of the fiber is identified on the basis of analysis of IR spectra, and then classified (Figure 3). When results of other investigations, i.e. color and morphology, are also taken into account, the fiber can be identified. Contaminants occurring on fibers can also be identified by microspectrometry.

In certain instances, the presence of a contaminant can increase the evidential value of a fiber match when the contaminant is shown to be present on both the questioned and known fibers. The technique is useful in the identification of man-made fibers.

To sum up, IR microspectrometry is quick, nondestructive, highly sensitive, and provides a wealth of information from a short piece of single fiber. Every feature of a spectrum contains information concerning the sample. Not every feature, however, is related to the chemical structure of the sample. Certain features, such as interference patterns and stray radiation, may arise from optical phenomena.

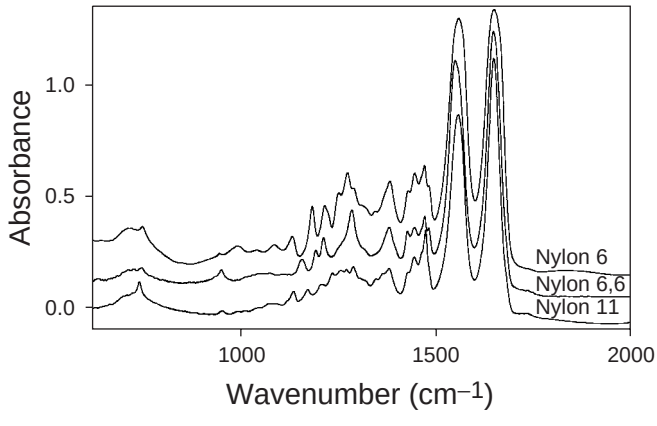

Figure 3 IR spectra of different kinds of single Nylon® fibers.

2.3.3 Examination of Inks

Writing materials, i.e. toners and various types of ink (in the form of text or writing on paper), are also studied by IR microspectrometry.[39–42] In cases relating to forgeries and suspicious documents, it is often necessary to know whether a questioned fragment of text or signature was created with the same type of writing/printing material as the remaining part of the text; sometimes it is necessary to establish whether an entry was made by the writing instrument sent in for study (Figure 4). In order not to destroy a studied document, the reflection technique can be applied, placing the document under the microscope and limiting the field of vision and measurement to the line of the questioned fragment of writing. However, one rarely gains good results by this method. Experience teaches that a spectrum obtained in this way mainly contains absorption bands of cellulose, i.e. paper, which masks the absorption bands originating from the ink or toner. It is more advantageous to separate out a fragment of writing agent with a scalpel and, after conveying it to a KBr plate, to obtain a spectrum by the transmission technique. That is the procedure for ball-point pen inks and toners. The loss of this agent in the line of writing is then minimal, invisible to the naked eye. In the case of some types of agents (e.g. inks for fiber pens) it is necessary to cut out a several millimeter long line of writing from the studied fragment of text and to carry out microextraction of the sample of ink or paste from the paper, using a mixture of solvents. The extract is then studied by the transmission method after being placed on the KBr window lying on the microscope stage. Independently, thin-layer chromatography may be used.

The reflection technique is successfully applied in the study of banknotes, tickets and other special forms where the layer of printing ink deposited on a significant area of the paper is sufficiently thick. Then absorption bands originating from the material deposited on the surface and not from the paper (the base of the document) are visible in the IR spectrum. Then it is especially useful to use a microscope with an ATR objective. Spectra can be obtained in several minutes without the need to remove the agent from the document.

The recently introduced ATR microscope objective has the significant advantage of enabling the analyst to view the specimen directly where the measurements are taken. The method is nondestructive and generally allows the analysis of specimens in situ on the evidence, i.e. on the document.

2.3.4 Others

2.3.4.1 Hair Hair is often found at the scene of the crime. The IR spectra of hairs are not readily distinguishable between individuals. However, the presence of hair spray or other coatings on hair (Figure 5) can significantly increase the value of evidence.[4]

2.3.4.2 Oils Another type of trace is so-called greasy stains, revealed on the clothes of victims of road accidents as a result of contact with a vehicle's chassis bearing oil or lubricant, or on the clothes of an arsonist. Stains visible on the fabric of the clothes may be of varying size; however, barely a drop (of oil or other lubricant) can be

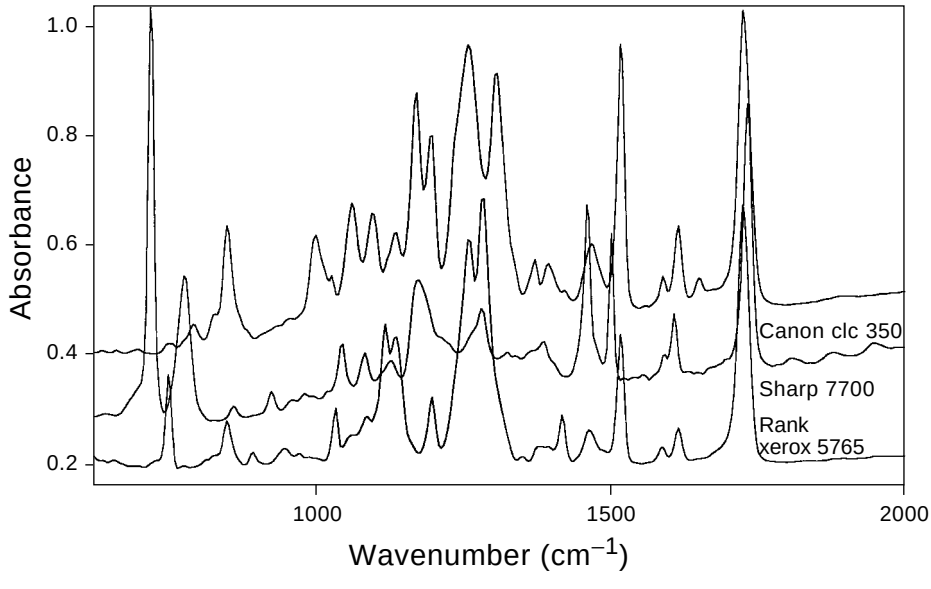

Figure 4 IR spectra of different kinds of black toners.

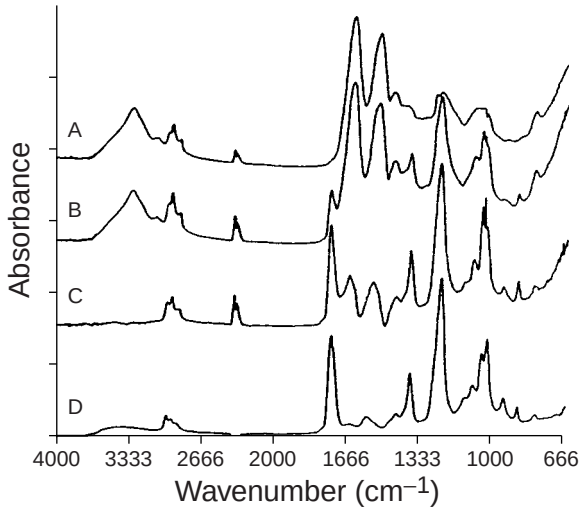

Figure 5 ATR spectra of hair (curve A) and hair with hairspray (curve B). Curve C: difference spectrum obtained by subtracting spectrum A from spectrum B. Curve D: hairspray reference spectrum. (Reprinted from E.G. Bartick, M.V. Tungal, J.A. Reffner, 'A New Approach to Forensic Analysis with Infrared Microscopy Internal Reflection Spectroscopy', *Anal. Chim. Acta*, **288**, 35–42, (1994). Copyright with permission from Elsevier Science.)

obtained by extraction with solvents. This small amount of substance can only be analyzed by IR spectrometry if the microscope attachment is used.[43,44] In order to take a measurement using the transmission method the sample is deposited on a KBr plate; for measurement by reflection it is placed on a metal plate. Comparison of the spectrum with the catalog of spectra of oil products sometimes enables identification of the lubricant. More frequently the spectrum obtained serves to establish the degree of similarity of the material forming the stain and the material secured from the vehicle of the suspect (Figure 6).

2.4 Raman Microspectroscopy

Raman and IR microspectroscopy are two complementary techniques. Whereas IR spectra are the result of absorption of radiation, Raman spectra arise from an inelastic scattering effect. The selection rules differ for IR and Raman. The bands that appear weak in the IR spectrum will be strong in the Raman spectrum, and vice versa.

There are only a few data in the literature on the subject of using this technique in the study of fragments of paint coatings.[45,46] It is interesting that some paint components may have vibrational modes that produce no IR absorption bands, but may produce Raman bands.

Raman spectroscopy is used increasingly in the study of fibers.[47] The Raman technique is sensitive enough to analyze single fibers readily with microscopes coupled to spectrometers. In addition to the complementary vibrational information obtained, Raman spectroscopy offers several other advantages. First, sample preparation is eliminated because scattering and not transmission is measured. Second, diffraction effects are greatly reduced because the exciting radiation is of a lower wavelength than IR. Third, spatial resolution is enhanced. Finally, narrow-range IR detectors generally have a low-frequency range near 700 cm^{-1}, whereas Raman experiments are generally performed down to 200 cm^{-1},

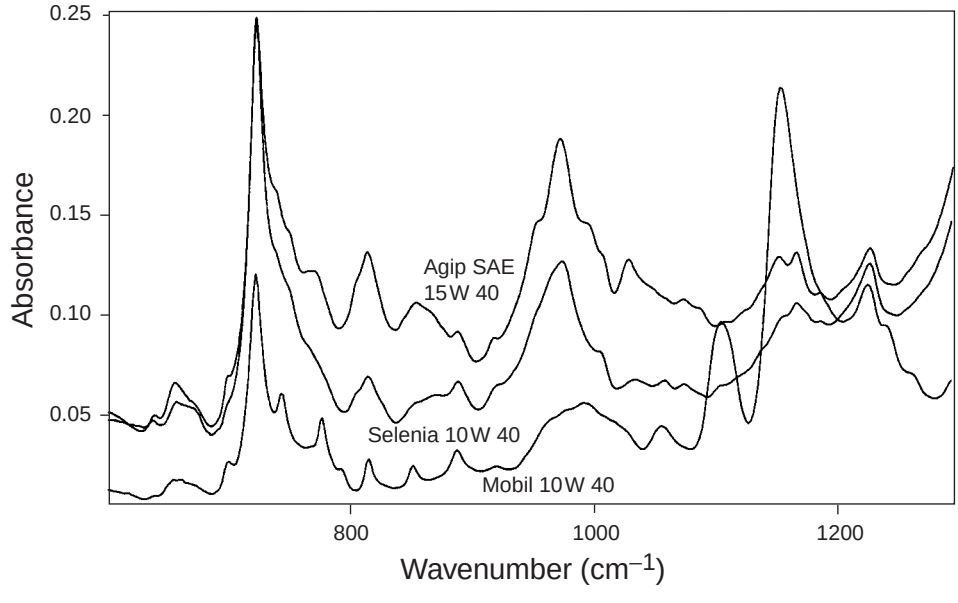

Figure 6 IR spectra of different kinds of motor oils.

and new holographic notch filters are pushing this limit down to $50\,cm^{-1}$. In this far-IR region one can find information relating primarily to the inorganic components of fibers, such as delusterants. Because Raman spectroscopy is not a transmission technique, fibers do not have to be pretreated before analysis, but there is potential for the surface composition of the fibers to be over-represented in the spectrum.

2.5 Databases and Libraries

The high repeatability which characterises the IR microspectrometry method enables the creation of libraries of spectra of samples of various materials as well as computer databases containing (in addition to catalogs of IR spectra) data on the subject of morphology, structure, physical and optical properties of materials and also technological data. They help to identify and classify samples quickly.[48–50] Libraries of spectra formed in one's own laboratory are best suited to these aims, which reflect the influence of local factors on the spectrum, such as the equipment, sample preparation method, the measuring technique, and the measuring conditions. Specialized firms offer libraries of IR spectra for sale. These are most frequently catalogs of the spectra of polymers, pigments, and other pure ingredients used in production of paints, fibers, writing agents, and other materials (Sadtler, Hummel, Aldrich). They are also available on diskette or CD (e.g. Nicolet, Spectra-Tech.).

It is more difficult to obtain databases of ready-made products, as opposed to the individual component materials. Catalogs of automobile paints (such as the Collaborative Testing Services (CTS) automotive paint collection) or fibers (such as the CTS fiber collection) do exist, but they only cover a few thousand samples of materials produced in a defined period of time (10–20 years).

Databases are being created by forensic laboratories in some countries. These encompass the products encountered or produced in a given country. The creation of a full and current catalog of products produced around the world is almost impossible, as this would require permanent cooperation of all manufacturers worldwide with research laboratories.

To sum up, IR microspectrometry has made it possible to identify and detect picogram quantities of some chemical substances, eliminating the need for previously applied techniques of sample preparation, which required a lot of time and effort. Together with optical and scanning electron microscopy, the method has become an indispensable tool in forensic research for the identification of trace amounts of various materials. When interpreting the IR spectra gained by this method it should be borne in mind that IR spectra obtained for nanogram quantities of unknown samples may not be the spectra of just the samples under investigation, but those of ubiquitous contaminants, such as fragments of epidermis, dust, or wood dust, which may be identified, respectively, as a fiber of Nylon®, silicon, or cellulose. By analyzing microregions of heterogeneous samples in a similar way, different IR spectra can be obtained at various places on the same sample. Repeated measurements for a given sample are therefore recommended.

3 COLOR ANALYSIS BY MICROSPECTROPHOTOMETRY

3.1 History

The first microspectrometers were built in the late 1950s by individual laboratories.[38,51] They were made for the purpose of measuring absorption spectra of colored fibers, for the types of traces which were frequently secured in connection with crimes at the time. These instruments consisted of a monochromator or filtering device as the source of light, a microscope, and a sensitive photometer. They were built because microscopic comparisons of two fibers could only reveal that ''the fibers are similar'', or ''the colors match'', which was insufficient for courts. Furthermore, in the case of analysis of very short, millimeter-long fibers it was not possible to identify the dye by thin-layer chromatography.

Today most laboratories use microspectrophotometers which are produced by specialized firms. A microspectrophotometer system has as its component parts a microscope, a spectrophotometer, a control unit for the photometer (a spectral data processor), and a chart recorder or plotter. Many users also incorporate as part of their system an analog-to-digital converter to interface with the microcomputer. The photometer is mounted on a microscope and consists of a grating monochromator and photomultiplier detector.

Analysis of the color of materials for forensic purposes was introduced in Europe at the beginning of the 1970s, especially in Switzerland where it was routinely used in the analysis of fibers. Only at the end of the 1970s was this technique tested and its discriminating power assessed. In 1979 Macrae[52] analyzed blue woollen fibers with the help of a comparison microscope and the microspectrophotometry method. He stated that in studying fibers using a comparison microscope the discriminating power was 68%, whereas for measurements made by microspectrometry it was 99%. It also turned out that in some cases it was possible to distinguish fibers by thin-layer chromatography, although when using microspectrophotometry in the visible range they were indistinguishable

and vice versa. Both techniques are currently considered complementary in the analysis of fiber color.

The effectiveness of discrimination by microspectrophotometry was also checked in the case of cotton fibers.[53] For red cotton fibers, 320 out of 1035 pairs of fibers could not be distinguished by microscopy, but only 20 could not be distinguished by microspectrophotometry. The fact that microspectrophotometry is highly discriminating and nondestructive, and that when using it spectra can be gained, even for very small microfibers or little bits of other materials, means that UV/VIS microspectrophotometry has begun to be universally applied in forensic laboratories.

3.2 Spectral Measurements and Analysis

Color is an essential feature of materials which is taken into account in comparative forensic studies.[38,51] A visual description of color, especially a comparison of the color of two objects, is difficult because the human eye only sees within a narrow range of electromagnetic radiation (the so-called visual range), and even in this range it is not uniformly sensitive to the energies of all wavelengths. The color of an object also changes depending on the lighting and background. So lighting, background, observation conditions and a good observer are prerequisites for a good observation of color, essential for an accurate comparison of the color of two samples. In order for a color to be measured objectively, it was necessary to find a method of description of color which would correlate with its visual perception.

Color is linked with absorption of incident radiation. According to Beer's law absorbance of a material is linearly related to its concentration. Hence colorimetric methods have been used for years to define it. The new generation of spectrophotometers for the visual range, which measures transmitted or reflected light by a sample, is universally used in the measurement and comparison of the color of two objects.

The color of an object depends on three factors: the light source, the observed object, and the color response of the eye. Visual descriptions of color are very subjective, and simply recording absorption maxima alone will afford only limited information. A number of color order systems have been developed, which are of great value for storing and communicating color information. The method developed by Peterson and Cook[54] was based on a widely accepted color ordering system originally proposed by the Commission International de l'Eclairage (CIE) of 1931.[38,55]

Each shade of color can be obtained by mixing the three basic colors in appropriate proportions, namely red, yellow and blue. Every color is therefore a function of three variables. This can be expressed mathematically as a system of linear equations. These allow the calculation of so-called chromaticity coordinates from the tristimulus values. Appropriate coefficients of proportionality are worked out by experiment.[51] A digital code is ascribed to each color, so that instead of a verbal description of the color an unambiguous digital description is introduced. Color catalogs have also been developed that are used by experts and facilitate understanding between them.

A color comparison is carried out on the basis of results of spectrometric studies, taking into account the position of maximum absorption on the spectrum and the course of the graph over the whole spectral range. The wavelength corresponding to maximum absorption and the sum of squares of the differences between points on spectra corresponding to each other may be differentiating parameters.

3.3 Application in Examinations

There is a constant need in forensic science for comparison of colors to establish identity between two parts and, as the difference between identity and nonidentity may be very small, great care must be used to convince a court. Usually the samples are small – such as a hair, a fiber, or a paint chip – so special apparatus is needed. For straightforward comparison a comparison microscope with two fields is used, but when the color, as seen by the instrument, must be recorded, a specialized colorimeter or spectrometer is used.

Many forensic laboratories worldwide have successfully used microscope-mounted spectrophotometers to compare fibers and other materials such as inks, paints, and plastics.

Microspectrophotometry, which uses a microscope and photometric measuring techniques, has proved to be an extremely valuable, if not indispensable, tool in comparing the colors of fibers.[53,55–58] With its ability to provide color information from very small areas of single fibers, microspectrophotometry can confirm suspected color matches while removing the disadvantages of subjectivity which are associated with human perception of color. Equally important is the ability of microspectrophotometry to immediately inform the user of spectral differences between two colored fibers, in spite of a convincing visual match on the comparison microscope – the phenomenon known as metamerism.

The examiner of a questioned document is faced with the problem of identification and comparison of inks which are apparently similar in color and shade. Inks are characterized by various physical, chemical, and chemicophysical techniques. Nondestructive optical methods do not always reveal the identity of individual ink components and may not be entirely conclusive.

Visual examination with or without the aid of a microscope is operator dependent and lacks the sensitivity and precision of spectrophotometric instrumentation. Application of microspectrophotometry, however, enables examination and comparison of different kinds of inks. Nowadays color measurement in document examination has to be performed on ink traces and occasionally on small fibers such as those found in banknotes. This requires a microspectrophotometer with capabilities in transmission, reflectance, and sometimes fluorescence modes.

An object to be examined, such as an ink line, is viewed through a microscope and the light reflected from it is measured over the visible spectrum. The reflected energy is then compared with a standard stored in the system's microprocessor, and a spectral curve is provided which is characteristic for the color of the object examined.[59]

Visible spectra have been measured from minimal quantities of commonly encountered ball-point and fiber-tip pen inks. Comparison of these spectra provided a high degree of discrimination between similarly colored inks.[60]

More recent applications in document examination should find increasing application in the authentication of passports, visas, green cards, credit cards, money orders, international banknotes, and many other valuable or critical documents.

This procedure has been applied to the detection of forgeries in documents by showing up the difference in color of ink in different parts of a document, whether handwritten or typed.[59–64] Even if an addition or alteration has been made with the same ink at a later time, this will show and be recorded, because color changes at a regular pace as a result of atmospheric oxidation.

Another application is in the confirmation that a particular paint chip came from a particular car, or even to identify the make of car from which a chip originated.[65–68] Microspectrophotometry may be required to provide objective color data for paint comparison due to the typically small size of samples. The technique can be applied to the outer surfaces of paint films by diffuse reflectance measurements with visible spectrum illumination. Comparison of paint layers by transmission microspectroscopy of thin edge sections offers a more definitive form of color analysis for these samples compared to reflectance techniques. Individual makers have a standardized procedure of a number of undercoats and topcoats. A microscopic examination of a cross-section of the chip with a measurement of their colors can identify the make, and, furthermore, pin down an individual car, if an authentic sample is available for comparative examination.

The application of this technique in forensic research is a relatively new development. In the case of very small samples of paint it is irreplaceable in defining and comparing colors. Although up to now no research has been published indicating an improvement in the differentiation of samples using UV/VIS microspectrophotometry, this technique shows promise for the compilation of databases targeted at vehicle make/model finish identifications.

3.4 Spectral Analysis Using Ultraviolet Radiation and Fluorescence

Interest is growing in microspectrometry in the UV region.[55,57] Some spectrometers enable the measurement of spectra in the 240–760 nm range, and even 190–900 nm. They can be adapted to the recording of fluorescent spectra. These techniques are particularly valuable in comparison studies of fibers. Fibers which cannot be differentiated in the visible region can sometimes be distinguished in the UV range, especially when the differences relate to the presence or absence of optical brighteners or application of different brightening agents.

Fluorescence microscopy of thin cross-sections, as an aid in differentiating samples or various layers within intact paint fragments, is discussed by Stoecklein and Tuente.[69] When using an excitation wavelength of 365 nm, the technique may be sensitive to differences in organic pigments, additives, and film-forming components.

ACKNOWLEDGMENTS

I wish to gratefully acknowledge the assistance of Zuzanna Brożek-Mucha who proofread various versions of this manuscript and offered valuable suggestions.

I would like to thank Małgorzata Pomianowska for help in preparing the figures for this manuscript.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflection
CIE	Commission International de l'Eclairage
CTS	Collaborative Testing Services
FTIR	Fourier Transform Infrared
IR	Infrared
S/N	Signal-to-noise Ratio
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymers and Rubbers

Infrared Spectroscopy **(Volume 12)**
Microspectroscopy

Raman Spectroscopy **(Volume 15)**
Fourier Transform Raman Instrumentation

REFERENCES

1. R.G. Messerschmidt, 'Minimizing Optical Nonlinearities in Infrared Microspectroscopy', in *Practical Spectroscopy Series*, ed. H.J. Humecki, Marcel Dekker, New York, 1–40, Vol. 19, 1995.
2. J.A. Reffner, P.A. Martoglio, 'Uniting Microscopy and Spectroscopy', in *Practical Spectroscopy Series*, ed. H.J. Humecki, Marcel Dekker, New York, 41–84, Vol. 19, 1995.
3. S. Compton, J. Powell, *Infrared Microscopy Using the UMA 150 Accessory: Some Forensic Applications*, FTS 7/IR Notes No.2, Bio-Rad, Cambridge, MA, 1991.
4. E.G. Bartick, M.V. Tungol, J.A. Reffner, 'A New Approach to Forensic Analysis With Infrared Microscopy: Internal Reflection Spectroscopy', *Anal. Chim. Acta*, **288**, 35–42 (1994).
5. J. Katon, G. Pacey, J.O. Keefe, 'Vibrational Molecular Microspectroscopy', *Anal. Chem.*, **58**(3), 465A–474A (1986).
6. R.W. Duerst, W.L. Stebbings, G.J. Lillquist, J.W. Westberg, W.E. Breneman, C.K. Spicer, R.M. Dittmar, M.D. Duerst, J.A. Reffner, 'Depth Profiling and Defect Analysis of Films and Laminates: an Industrial Approach', in *Practical Spectroscopy Series*, ed. H.J. Humecki, Marcel Dekker, New York, 137–162, Vol. 19, 1995.
7. R.G. Messerschmidt, 'Infrared Microspectroscopy. Theory and Application', *Practical Spectroscopy Series*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, New York, Vol. 6, 1988.
8. H.J. Humecki (ed.), *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York, 1995.
9. S.G. Ryland, 'Infrared Microspectroscopy of Forensic Paint Evidence', in *Practical Spectroscopy Series*, ed. H.J. Humecki, Marcel Dekker, New York, 163–243, Vol. 19, 1995.
10. E.M. Suzuki, 'Forensic Applications of Infrared Spectroscopy', in *Forensic Science Handbook*, ed. R. Saferstein, Regents Prentice Hall, Englewood Cliffs, NJ, 71–195, Vol. 3, 1993.
11. J.I. Thornton, 'Forensic Paint Examination', in *Forensic Science Handbook*, ed. R. Safestein, Regents Prentice Hall, Englewood Cliffs, NJ, 529–571, Vol. 1, 1982.
12. P.G. Rodgers, 'The Classification of Automotive Paint by Diamond Window Infrared Spectrophotometry', Part II: Automotive Topcoats and Undercoats, *Can. Soc. Forens. Sci. J.*, **9**, 49–68 (1976).
13. G. Dabdoub, P. Severin, 'The Identification of Domestic and Foreign Automobile Manufacturers Through Body Primer Characterization', *J. Forens. Sci.*, **34**(6), 1395–1404 (1989).
14. S.G. Ryland, R.J. Kopec, 'The Evidential Value of Automobile Paint Chips', *J. Forens. Sci.*, **24**(1), 140–147 (1979).
15. E.M. Suzuki, 'Infrared Spectra of US Automobile Original Topcoats (1974–1989)', *J. Forens. Sci.*, **41**(3), 376–406 (1996).
16. R.F.E. Percy, R.J. Audette, 'Automotive Repaints: Just a New Look?', *J. Forens. Sci.*, **25**(1), 189–239 (1980).
17. J.M. Wilkinson, J. Locke, D.K. Laing, 'The Examination of Paint as Thin Sections Using Visible Microspectrophotometry and Fourier Transform Infrared Microscopy', *Forens. Sci. Int.*, **38**, 43–52 (1988).
18. D.R. Cousins, 'The Use of Microspectrophotometry in The Examination of Paints', *Forens. Sci. Rev.*, **1**(20), 141–162 (1989).
19. D.R. Cousins, C.R. Platoni, L.W. Russell, 'The Variation in The Color of Paint on Individual Vehicles', *Forens. Sci. Int.*, **24**, 197–208 (1984).
20. D.J. McEwen, G.D. Cheever, 'Infrared Microscoping Analysis of Multiple Layers of Automotive Paints', *J. Coatings Technol.*, **65**(819), 35–41 (1993).
21. D.R. Cousins, C.R. Platoni, L.W. Russell, 'The Use of Microspectrophotometry for the Identification of Pigments in Small Paint Samples', *Forens. Sci. Int.*, **24**, 183–196 (1984).
22. T.J. Allen, 'Paint Sample Presentation for Fourier Transform Infrared Microscopy', *Vib. Spectrosc.*, **3**, 217–237 (1992).
23. E.G. Bartick, M.W. Tungol, 'Infrared Microscopy and its Forensic Applications', in *Forensic Science Handbook*, ed. R. Saferstein, Regents Prentice Hall, Englewood Cliffs, NJ, 196–252, Vol. 3, 1993.
24. J. Zięba-Palus, 'The Use of Micro Fourier Transform Infrared Spectroscopy and Scanning Electron Microscopy With X-Ray Microanalysis for the Identification of Automobile Paint Chips', *Microchim. Acta*, **14**, 357–359 (1997).
25. J. Zięba-Palus, 'Selected Cases of Forensic Paint Analysis', *Sci. Justice*, **39**(2), 123–127 (1999).
26. J. Zięba-Palus, 'Application of Micro Fourier Transform Infrared Spectroscopy to the Examination of Paint Samples', *J. Mol. Struct.*, **511/512**, 327–335 (1999).

27. J. Zięba-Palus, 'Identification of Paint Fragments', *Prob. Forens. Sci.*, **XXXII**, 114–118 (1995).
28. D.W. Schiering, 'A Beam Condenser/Miniature Diamond Anvil Cell Accessory for the Infrared Microspectrometry of Paint Chips', *Appl. Spectrosc.*, **42**, 903 (1988).
29. M.W. Tungol, E.G. Bartick, A. Montaser, 'The Development of a Spectral Database for the Identification of Fibers by Infrared Microscopy', *Appl. Spectrosc.*, **44**, 543 (1990).
30. D. Lin-Vien, B.J. Bland, V.J. Spence, 'An Improved Method of Using the Diamond Anvil Cell for Infrared Microprobe Analysis', *Appl.Spectrosc.*, **44**, 1227 (1990).
31. M.C. Grieve, 'Another Look at the Classification of Acrylic Fibers, Using FTIR Microscopy', in *Proceedings of the 13th Meeting of the International Association of Forensic Sciences*, Dusseldorf, Germany, 1993.
32. M.W. Tungol, E.G. Bartick, A. Montaser, 'Analysis of Single Polymer Fibers by Fourier Transform Infrared Microscopy: the Results of Case Studies', *J. Forens. Sci.*, **36**, 1027 (1991).
33. M.W. Tungol, E.G. Bartick, A. Montaser, 'Forensic Analysis of Acrylic Copolymer Fibers by Infrared Microscopy', *Appl. Spectrosc.*, **47**, 1655 (1993).
34. M.W. Tungol, E.G. Bartick, A. Montaser, 'Polarized Infrared Microscopical Analysis of Single Poly(Ethylene Terephtalate) Fibers', *Proceedings of the 13th Meeting of the International Association of Forensic Sciences*, Dusseldorf, Germany, 1993.
35. M.C. Grieve, J. Dunlop, T.M. Kotowski, 'Bicomponent Acrylic Fibers: their Characterization in the Forensic Science Laboratory', *J. Forens. Sci. Soc.*, **28**, 25 (1988).
36. G.C. Pandey, 'Fourier Transform Infrared Microscopy for the Determination of the Composition of Copolymer Fibers: Acrylic Fibers', *Analyst*, **114**, 231 (1989).
37. M.W. Tungol, E.G. Bartick, A. Montaser, 'Forensic Examination of Synthetic Textile Fibers by Microscopic Infrared Spectrometry', in *Practical Spectroscopy Series*, ed. H.J. Humecki, Marcel Dekker, New York, 245–286, Vol. 19, 1995.
38. K.P. Kirkbride, 'The Application of Infrared Microspectrometry to the Analysis of Single Fibers', in *Forensic Examination of Fibers*, ed. J. Robertson, Ellis Horwood, New York, 181–218, 1992.
39. R.A. Merill, E.G. Bartick, W.D. Mazzella, 'Studies of Techniques for Analysis of Photocopy Toners by IR', *J. Forens. Sci.*, **41**, 2 (1996).
40. J. Andrasko, 'A Simple Method for Sampling Photocopy Toners for Examination by Microreflectance Fourier Transform Infrared Spectrophotometry', *J. Forens. Sci.*, **39**, 1 (1994).
41. J. Andrasko, 'Microreflectance FTIR Techniques Applied to Materials Encountered in Forensic Examination of Documents', *J. Forens. Sci.*, **41**, 5 (1996).
42. B.M. Trzcińska, Z. Brożek-Mucha, 'The Possibilities of Identifying Photocopy Toners by Means of Infrared Spectroscopy (FTIR) and Scanning Microscopy (SEM/EDX)', *Mikrochim. Acta*, **14**, 235–237 (1997).
43. J. Zięba, 'Differentiation of Lubricating Oil Stains on Textile by Infrared Spectroscopy', *Forens. Sci. Int.*, **30**, 45–51 (1985).
44. J. Zięba-Palus, P. Kościelniak, 'Differentiation of Motor Oils by Infrared Spectroscopy and Elemental Analysis for Criminalistic Purposes', *J. Mol. Struct.*, **482–483**, 533–538 (1999).
45. A.H. Kuptsov, 'Application of Fourier Transform Raman Spectroscopy in Forensic Science', *J. Forens. Sci.*, **39**(2), 305–318 (1994).
46. M. Claybourn, J.K. Agbenyega, P.J. Hendra, G. Ellis, 'Fourier Transform Raman Spectroscopy in the Study of Paints', *Adv. Chem. Ser.*, **236**, 443–482 (1993).
47. R.G. Messerschmidt, D.B. Chase, 'FT–Raman Microscopy: Discussion and Preliminary Results', *Appl. Spectrosc.*, **43**, 11 (1989).
48. D.K. Laing, R. Cook, A.W. Hartshorne, G. Robinson, 'A Fiber Data Collection for Forensic Scientists Collection and Examination Methods', *J. Forens. Sci.*, **32**, 364–360 (1987).
49. G.R. Carroll, W.C. Lalonde, B.D. Gaudette, S.L. Hawley, R.S. Hubert, 'A Computerized Database for Forensic Textile Fibers', *Can. Soc. Forens. Sci. J.*, **21**, 1–10 (1989).
50. J.L. Buckle, D.A. Macdougall, R.R. Grant, 'PDQ-Paint Data Queries: The History and Technology Behind the Development of the Royal Canadian Mounted Police Forensic Laboratory Services Automotive Database', *Can. Soc. Forens. Sci. J.*, **30**(4), 199–212 (1997).
51. G.J. Chamberlin, D.G. Chamberlin, *Color – its Measurement, Computation and Application*, Heyden, London, 1980.
52. R. Macrae, R.J. Dudley, K.W. Smalldon, 'The Characterization of Dyestuffs on Wool Fibers with Special Reference to Microspectrophotometry', *J. Forens. Sci.*, **24**, 117–129 (1979).
53. M.C. Grieve, J. Dunlop, P. Haddock, 'An Assessment of the Value of Blue, Red and Black Cotton Fibers as Target Fibers in Forensic Science Investigations', *J. Forens. Sci.*, **33**, 1332–1334 (1988).
54. M.D. Paterson, R. Cook, 'The Production of Color Coordinates from Microgram Quantities of Textile Fibers. Part 1', *Forens. Sci. Int.*, **15**, 249–258 (1980).
55. J. Dunlop, 'Color Analysis by Microspectrophotometry', in *Forensic Examination of Fibers*, ed. J. Robertson, Ellis Horwood, New York, 127–140, 1992.
56. A.W. Hartshorne, D.K. Laing, 'The Definition of Color for Single Textile Fibers by Microspectrophotometry', *Forens. Sci. Int.*, **34**, 107–129 (1987).
57. A.W. Hartshorne, D.K. Laing, 'Color Matching Within a Fiber Data Collection', *J. Forens. Sci.*, **33**, 1345–1354 (1988).

58. D.K. Laing, A.W. Hartshorne, R.J. Harwood, 'Color Measurements on Single Textile Fibers', *Forens. Sci. Int.*, **30**, 65–77 (1986).
59. L.A. Olson, 'Color Comparison in Questioned Document Examination Using Microspectrophotometry', *J. Forens. Sci.*, **31**, 1330–1340 (1986).
60. R.N. Totty, M.R. Ordridge, 'A Comparison of the Use of Visible Microspectrophotometry and High Performance Thin Layer Chromatography for the Discrimination of Aqueous Inks Used in Porous Tip and Roller Pens', *Forens. Sci. Int.*, **28**, 137–144 (1985).
61. D.K. Laing, M.D.J. Isaacs, 'The Comparison of Nanogram Quantities of Ink Using Visible Microspectrometry', *J. Forens. Sci. Soc.*, **23**, 147–154 (1983).
62. A. Zeichner, B. Glattstein, 'Some Special Observations Regarding Visible Transmission Spectra of Inks and an Improved Method for their Discrimination by Microspectrophotometry', *J. Forens. Sci.*, **37**, 738–749 (1992).
63. P.W. Pfefferli, 'Application of Microspectrophotometry in Document Examination', *Forens. Sci. Int.*, **23**, 129–136 (1983).
64. A. Zeichner, N. Levin, A. Klein, Y. Novoselsky, 'Transmission and Reflectance Microspectrophotometry of Inks', *J. Forens. Sci.*, **33**, 1171–1184 (1988).
65. D.K. Laing, R.J. Dudley, M.D.J. Isaacs, 'Colorimetric Measurements on Small Paint Fragments Using Microspectrometry', *Forens. Sci. Int.*, **16**, 159–171 (1980).
66. J. Nowicki, R. Patten, 'Examination of US Automotive Paints: Make and Model Determination of Hit-and-run Vehicles by Reflectance Microspectrophotometry', *J. Forens. Sci.*, **32**(2), 464–470 (1986).
67. G.P. Voskertchian, 'Quantitative Analysis of Organic Pigments in Forensic Paint Examination', *J. Forens. Sci.*, **40**(5), 823–825 (1995) .
68. D.R. Cousins, 'The use of Microspectrophotometry in the Examination of Paints', *Forens. Sci. Rev.*, **1**(2), 141–161 (1989).
69. W. Stoecklein, J. Tuente, 'Using the Light Microscope for Analytical Procedures – Aids for Solving Cases Involving Hit-and-run Offences', *Zeiss Inf. Jena Rev.*, **3**(4), 19–22, 669–678 (1994).

Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Brian Dawson
Bureau of Biologics and Radiopharmaceuticals, Therapeutic Products Programme, Ottawa, Canada

1 Introduction 1
2 Methodology 2
2.1 Sample Preparation 2
2.2 Nuclear Magnetic Resonance Experiments 3
3 Applications 7
3.1 Confirmation of Known Compounds 7
3.2 Identification of Unknown Compounds 7
3.3 Quantification 7
3.4 Impurity Profiling 8
3.5 Optical Purity 8
3.6 Analysis of Body Fluids 8
4 Comparison with Other Forensic Techniques 8
4.1 Advantages 8
4.2 Disadvantages 9
5 Survey of the Literature 10
5.1 Amphetamines and Related Substances 10
5.2 Phencyclidine and Related Substances 11
5.3 Fentanyls 12
5.4 Opiate Alkaloids 12
5.5 Cocaine and Related Substances 13
5.6 Cannabinoids 13
5.7 Ergot Alkaloids 13
5.8 Quinazolinones 14
5.9 Barbiturates 14
5.10 Benzodiazepines 14
5.11 Other Abused Drugs 14
6 Conclusions 14
Abbreviations and Acronyms 15
Related Articles 15
References 15

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Nuclear magnetic resonance (NMR) spectroscopy provides the forensic analyst with an extremely powerful tool for the detection and quantification of abused drugs. A whole range of one-dimensional (1-D) and two-dimensional (2-D) NMR techniques is available for performing the required analyses. These NMR methods may be used for routine purposes, such as to confirm the identity of a drug or quantify the amount of illicit substance present in a police exhibit. However, the area where NMR stands out as an analytical tool is in the identification of unknown compounds, such as ''designer drugs''. NMR is also used in police intelligence work, as it can provide clues to the synthetic route used to prepare the drug. This is done by impurity profiling or by determining the drug's optical purity. Although NMR has been used for many years to analyze abused drugs, even the most modern spectrometers lack the sensitivity obtainable by other techniques such as mass spectrometry (MS) or high-performance liquid chromatography (HPLC). However, NMR is a nondestructive technique which provides essential structural information which cannot be obtained from these other methods. NMR also has the distinct advantage of not requiring reference standards for the identification of unknowns.

1 INTRODUCTION

The nuclei of atoms with an odd number of protons, neutrons, or both, have an intrinsic nuclear spin. The spinning of charged particles, the circulation of charges, produces a magnetic moment along the axis of spin. The simplest nucleus possessing a spin is the proton, the nucleus of the hydrogen atom. If hydrogen atoms are placed in an external magnetic field, their magnetic moments, according to quantum mechanics, will become aligned in two ways, with and against the external field. For hydrogen, the more stable alignment is with the field and energy must be absorbed to ''flip'' the proton over to the less stable alignment against the field. The amount of energy needed to flip the proton depends on the strength of the external field: the stronger the field, the greater the energy required to produce the flip. The energy required corresponds to electromagnetic radiation in the radiofrequency (RF) range. The observation frequency is given by Equation (1):

$$\nu = \frac{\gamma H_0}{2\pi} \tag{1}$$

where ν = frequency, in cycles per second, γ = a nuclear constant (the gyromagnetic ratio), and H_0 = strength of the static magnetic field, in Tesla.

When the nucleus is irradiated at an appropriate frequency, absorption occurs, followed by the nuclei re-emitting the RF radiation to return to the lower energy

state. This process gives rise to a signal and the spectrum arising from this signal is called an NMR spectrum. The frequency at which a nucleus absorbs depends on the magnetic field which it ''feels'' and this effective field strength is not exactly the same as the applied field strength. The effective field strength at each nucleus depends on its environment, e.g. the electron density, closeness of other nuclei, and hence will require a slightly different applied frequency in order for absorption to take place. Thus each different nucleus will give rise to its own signal.

There are two basic NMR spectrometer designs, the continuous-wave and the pulsed or Fourier transform nuclear magnetic resonance (FTNMR). The continuous-wave spectrometers had low-field, water-cooled electromagnets which consumed large amounts of electricity. The newer FTNMR spectrometers use superconducting magnets cooled with liquid helium which consume very little electricity, but require relatively expensive cryogens (liquid nitrogen and liquid helium) to keep them operational. Although continuous-wave spectrometers are perfectly suitable for many routine applications, they have largely been replaced by the higher-field pulsed FTNMR instruments.

Because the energy difference between the nuclei aligned with and against the field is small, the numbers of nuclei in each spin state are almost equal. There is, however, a slight excess of spins in the more stable alignment (of the order of 1 in 10^5 for protons at 200 MHz). It is the signal from this excess which gives rise to the NMR signal, thus explaining the relative insensitivity of this technique. However, over the last 20 years, the field strength (and thus sensitivity) has grown enormously. In forensic science laboratories, modern spectrometers can easily obtain good quality spectra on amounts of abused drugs that are much less than a single dose, even for drugs such as lysergic acid diethylamide (LSD) where the dosage is about 50 μg.

Since the results obtained in forensic science laboratories are used in the criminal justice system, the methods used to obtain them must leave no room for doubt. Thus, the analytical techniques used must meet a number of strict criteria. Besides being sufficiently sensitive, they must be highly selective, reproducible, and suitable to give qualitative and, in many cases, quantitative analyses. Other desirable features of analytical techniques within forensic science include that they require the minimum amount of sample pretreatment and that they can handle mixtures of substances without prior separation. As will be seen from the following discussion, NMR spectroscopy meets all these requirements and is becoming more and more important for the identification of unknown substances, as well as for routine analyses, in many forensic science laboratories worldwide.

Reports in the open literature of uses of NMR in areas of forensic interest began with a few papers in the late 1960s. Starting in the early 1970s, reports on NMR results for illicit drugs and some closely related analogs began to appear more frequently. Most of these studies provided forensic scientists with valuable data useful in the identification and differentiation of series of abused substances with varying substitution patterns. These types of papers have continued to appear, often utilizing state-of-the-art NMR techniques and equipment to solve forensic problems such as structures of ''designer'' drugs. The use of NMR spectrometers in forensic laboratories has become very common since the early 1980s.

2 METHODOLOGY

2.1 Sample Preparation

There is a large variety in the types of samples submitted by law enforcement officers which are amenable to forensic NMR spectroscopic analysis. These include abused drugs, explosives, fire accelerants, hydrocarbon fuels, body fluids and tissues, lachrymators, poisons, and fingerprint reagents. The focus of this article will be on abused drugs.

Illicit drug seizures usually consist of solids (normally powders), plant materials (or their extracts) or dosage forms (tablets, capsules, or elixirs). In most cases, it is only necessary to identify the major component in the sample. However, it is sometimes necessary to quantify the amount of abused substance present. Also, identification of ''impurities'' in the exhibits (e.g. synthetic by-products, coextracts, decomposition products) may provide police with valuable intelligence information as to the source of the material (e.g. synthetic route used, possible country of origin of the plant).

The methods used to prepare the samples for NMR analyses depend on the source of the sample and the information required. Procedures range from simple dissolution to complex extraction and derivatization reactions which must be done under very stringent conditions.

For the simplest case of ''pure'' drug, a suitable quantity (usually a few milligrams) of the sample is dissolved in a deuterated NMR solvent. The choice of solvent depends on the solubility of the abused drug. Commonly used deuterated solvents include chloroform ($CDCl_3$), water (D_2O), dimethyl sulfoxide (DMSO-d_6) or a mixture of chloroform and methanol ($CDCl_3/CD_3OD$). The amount of sample used depends on the quantity of sample available, the field strength of the spectrometer, and the type of NMR experiments to be performed.

In cases where the exhibit is a mixture, such as in tablets or powders where the abused drug is ''cut'' with sugar or starch, sample preparation methods depend on the nature of the components. It is often possible to choose a solvent which will dissolve only the drug, leaving the other excipients behind, or to choose two immiscible solvents where the drug will go into one solvent while the excipients go into the other. For example, deuterochloroform will extract many drugs from formulations where the excipients are sugars, since sugars will not dissolve in this solvent but may be taken up in water. In many situations this simple procedure will not work and one must use the chemical characteristics of the drug to separate it from its excipients. Drugs may conveniently be divided into four general categories, acidic, basic, amphoteric, and neutral. For each of these, the partition coefficients between aqueous and immiscible organic solvents differ. The hydrogen ion concentration, which varies with pH in the aqueous phase, affects the ionic state or polarity of the drug, depending on the polar substituents present. It is often possible to separate drugs from different classes from each other or from other organic compounds by proper adjustment of the pH. For example, if the aqueous solution is made acidic, organic solvents may be used to remove acidic and neutral components. On the other hand, if the aqueous phase is made basic, basic and neutral components may be extracted with organic solvents. For amphoteric abused drugs such as morphine, the pH must be adjusted to its isoelectric point for it to be extracted. At this pH, the drug exists in a nonpolar or noncharged form, making it more soluble in appropriate organic solvents than in water. It is possible to isolate or remove neutral drugs or other organic components from acidic and basic drugs by extracting a basic aqueous solution with an organic solvent, and then washing this solvent with aqueous acid. Only the neutral materials should remain in the organic solvent.

When quantification is required, sample preparation procedures are more demanding. First, the amount of sample used must be accurately known. If an extraction procedure is necessary, one must be sure that the drug is quantitatively removed. Next, an appropriate internal standard must be chosen. This standard must meet a number of criteria. It must be pure (or at least be of known purity) and be soluble in the solvent used for analysis. It must be stable and not reactive to any compounds present (including the solvent). For convenience, it is usually easier to work with a solid standard, but if a liquid is chosen, it should be nonvolatile in order to avoid losses during the sample preparation and analysis. Lastly, the standard must have an appropriate NMR spectrum such that both the drug and standard have well separated resonances whose integrals or peak heights may be used in the quantification. It is usually best to choose a standard which has either a suitable singlet or narrow multiplet to minimize the chances of overlap of signal resonances.

Other procedures, such as extraction and derivatization, may be required for different types of plant materials. Special precautions may be necessary when the drugs in these materials are not stable under certain conditions, e.g. in the presence of acid or base. Because of the complexity of the extracts from most plants, NMR is usually not the method of choice for analyzing these materials. However, there have been a few cases where NMR has been used to advantage.

For drugs where legislation refers to only one optical isomer, procedures involving the addition of chiral shift reagents or chiral solvating agent (CSA), or reaction with chiral derivatization agents, are required.

2.2 Nuclear Magnetic Resonance Experiments

Since the mid 1980s, advances in NMR techniques have grown dramatically. As the number of experiments increases, most NMR spectroscopists opt to choose a set of these experimental techniques which work well for them in obtaining the kind of information they require. What follows is a discussion of the general types of experiments which have most often been used in the analyses of abused drugs.

2.2.1 One-dimensional Techniques

2.2.1.1 High-resolution Spectra When most analysts with a passing knowledge of NMR think about this technique, they recall the traditional 1-D high-resolution proton (^{1}H) and carbon (^{13}C) spectra. These spectra are highly informative in the hands of a trained NMR spectroscopist. For forensic analysts who may not be experts in the field, these spectra may be used as ''fingerprints'' for comparison with those for authentic materials. For this type of work, the use of ^{13}C-NMR spectra is most suitable. ^{1}H-NMR spectra are not as convenient for a number of reasons. The main reason is that the appearance of the proton spectra differ significantly when obtained at different field strengths – the higher the field, the more first-order (and thus simpler) the spectra become. At lower fields, the ^{1}H-NMR spectra for many abused drugs consist of broad and uninterpretable envelopes of signals, whereas at higher fields the corresponding spectra contain large numbers of sharp resonances which may be readily assigned using modern techniques. On the other hand, ^{13}C spectra look very similar, regardless of the field strength used. The only difference would be for those signals which may only be resolved from each other at higher fields.

^{1}H-NMR spectra can give detailed information on the nature of unknown substances. Each different proton in the molecule gives rise to a peak or group of peaks. When there is no overlap in the ^{1}H-NMR spectra, integration may be used to get the relative number of each type of proton. The chemical shift of the protons indicate the environment in which they exist. The majority of protons fall in a range of about 10 ppm. The chemical shifts are usually referenced to tetramethylsilane (TMS), which is assigned a value of 0.00 ppm. Chemical shifts provide evidence as to whether the protons are aliphatic (0.8–3.0 ppm), olefinic (4.5–6.5 ppm), or aromatic (6.5–8.5 ppm). Protons attached to carbon atoms that are bonded to nitrogen resonate at about 2.0–3.5 ppm, while if the carbon is bound to an oxygen, the protons resonate at about 3.3–4.5 ppm. Aldehyde protons are found around 8.5–10.0 ppm. Complete tables of chemical shifts may be found in most introductory text books.

High-resolution ^{13}C-NMR spectra also provide a great deal of information. The chemical shift range for carbon is more than 200 ppm. As with protons, the chemical shift ranges are indicative of the kind of carbon. Carbon shifts follow basically the same order as protons, with aliphatic carbons being at highest fields, followed by olefinic and then aromatic carbons. Acid and ester carbonyls are at lower fields than aromatic carbons, followed by aldehydes and isolated (nonconjugated) carbonyls at lowest fields. As with protons, tables of typical chemical shifts may be found in most introductory text books.

The number of carbon signals (excluding those from the solvent) usually gives a direct count of the number of chemically distinct carbons in the molecule studied. Exceptions to this occur when two or more different carbons have chemical shifts that are too close to be resolved under the conditions of the experiment, or when one or more carbons are coupled to another magnetic nucleus that is not decoupled (^{19}F or ^{31}P, for example). In a routine ^{13}C-NMR spectrum, the signals obtained for quaternary carbons are smaller than those for protonated carbons. It is possible, by using longer delays during acquisition and taking care to eliminate the nuclear Overhauser effect (NOE), to obtain carbon spectra such that each carbon gives rise to the same size of peak. When this type of quantitative experiment is performed, one can easily determine if there are unresolved carbon shifts or if any carbons are coupled to any other nuclei, thus enabling the analyst to get an accurate count of the number of carbons in the molecule.

One class of 1-D experiment that is very important is that which gives a count of the number of protons attached to each carbon. The best known of these is called distortionless enhancement by polarization transfer (DEPT). By suitable choice of experimental parameters, the ^{13}C spectra obtained have positive peaks for CH and CH_3 groups, negative peaks for CH_2 groups, and no signal for quaternary carbons. Varying the parameters allows spectra showing only CH or CH_2 or CH_3 carbons to be obtained. Thus the number of protons attached to every carbon in the compound being studied can be readily determined.

Although it is more common to obtain high-resolution spectra for protons and carbons, spectra may be obtained for many other nuclei. Those which are most often studied include ^{15}N, ^{19}F, and ^{31}P. Many abused drugs contain nitrogen. However, the low natural abundance and relative insensitivity of the ^{15}N nucleus make it a very difficult atom to study. Comparatively large amounts of material are required in order to obtain spectra in a reasonable period of time. Despite this, a number of reports have been published containing ^{15}N chemical shift data for abused drugs. The other two nuclei (^{19}F and ^{31}P) are essentially 100% abundant and are NMR sensitive but, unfortunately, do not occur in many abused drugs.

2.2.1.2 Nuclear Overhauser Effect When magnetic nuclei are close to each other, there is a dipolar interaction. If the chemical shift of one of these nuclei is irradiated in order to saturate its spin, any nuclei that are close to it may experience a small change in the size of the signal, referred to as the NOE. For small molecules, such as abused drugs, this effect is usually a signal enhancement. The closer the nuclei are to each other, the greater the increase in signal. The maximum amount of enhancement for small molecules is only a few percent of the original size of the signal. Because the strength of the dipolar interaction falls off at a rate which is proportional to the inverse of the distance raised to the sixth power, nuclei must normally be within a distance of about 5 Å in order to see any observable enhancement. In order to observe the small signal enhancements, a difference spectrum must be obtained by subtracting a reference spectrum obtained under ''identical'' conditions, with the only difference being that the reference spectrum is obtained while irradiating at a blank part of the spectrum instead of at the frequency for the chemical shift for one of the nuclei. When the difference spectrum is plotted, a large negative peak for the irradiated nucleus and small positive peaks for nuclei experiencing an observable NOE are obtained. Thus, this procedure often allows the determination of which protons are close to each other in space, often allowing the analyst to determine the relative stereochemistry in unknown molecules.

2.2.1.3 Lanthanide Shift Reagent Studies The magnetic moments of the unpaired electrons of paramagnetic reagents are capable of producing relatively large chemical shift changes in molecules which are in close proximity.

Most paramagnetic reagents are detrimental to the appearance of NMR spectra because they cause severe line broadening of all the signals. However, a few complexes of the rare-earth elements (lanthanides) do cause substantial chemical shift changes without substantial line broadening on lower field-strength instruments. The most common of these, referred to as lanthanide shift reagents (LSRs), are composed of complexes of europium, praseodymium, or ytterbium with a number of different diketones. The most popular LSRs are complexes with 6,6,7,7,8,8,8-heptafluoro-2,2-dimethyloctane-3,5-dione [$Eu(fod)_3$, $Pr(fod)_3$, and $Yt(fod)_3$] or 2,2,6,6-tetramethylheptane-3,5-dione [$Eu(dmp)_3$, $Pr(dmp)_3$, and $Yt(dmp)_3$].

When a shift reagent is added to a solution of the abused drug, it forms a weak bond to the most basic site in the drug, usually a nitrogen or oxygen. The amount of change in the chemical shift is inversely proportional to the distance between the lanthanide atom and the proton (or carbon). The experimental procedure followed in this type of study is to obtain a spectrum of the drug alone and in the presence of varying amounts of the shift reagent. By following the shifting of the different nuclei, one can determine their relative distances from the reagent, thus providing help in assigning the shifts to the appropriate nuclei. In the case of proton studies, the spreading out of the resonance signals may cause the spectrum to be more first-order and thus easier to interpret.

The amount of line broadening experienced by nuclei with a given concentration of LSR is directly proportional to the square of the magnetic field strength. Thus, these reagents are not suitable for use on the more modern high-field spectrometers (>300 MHz). However, on lower-field instruments, they have been used for a wide variety of abused drugs for many years.

2.2.1.4 Optical Purity Methods Another important class of 1-D experiments is that used to determine the optical isomer content of drugs. There are a number of abused drugs for which only one isomer is illicit in many countries. There are three basic types of NMR experiments which may be used to solve this problem. The first method to be developed was the use of chiral lanthanide shift reagents (CLSRs). A related method developed later was the use of CSAs. The final method is the use of chiral derivatizing agents (CDAs).

The use of LSRs was discussed in section 2.2.1.3. There are a number of shift reagents which have optically active ketone groups. Some of the common CLSRs include complexes with the ligands 3-(trifluoroacetyl)-D-camphor [$Eu(tfc)_3$], 3-(heptafluorobutyryl)-D-camphor [$Eu(hfc)_3$] and D,D-dicampholylmethane [$Eu(dcm)_3$]. These materials work in basically the same manner as LSRs, but when these compounds are used, they may induce different chemical shift changes for nuclei of optical antipodes. The CLSRs have been used to determine the optical purity of several abused drugs.

An alternative to CLSRs is to use one of a series of compounds referred to as chiral solvents or CSAs. These are optically active compounds possessing an aromatic group and a polar functional group such as an amine, alcohol, or acid. When a suitable CSA is added to a solution of an abused drug which also has a polar group, the CSA forms a weak hydrogen bond with the drug. Because the CSA is optically pure, it forms different complexes with each optical antipode of the drug. Thus, the aromatic-ring current of the CSA is located at different distances from the corresponding nuclei in the drug isomers. This causes the nuclei of the drug to give rise to separate signals for each antipode, thus enabling the determination of the optical purity of the drug. Unlike the CLSRs, there is no line broadening when one uses CSAs, and therefore the higher-field modern spectrometers are the instruments of choice for this type of experiment.

The third method for determining the optical isomer content of abused drugs is the use of CDAs. As with CSAs, the CDAs usually contain an optical center and an aromatic group. They also contain a functional group (such as an acyl chloride) which is capable of reacting and forming a covalent bond with the abused drug. As with the CSAs, the aromatic group is in different proximity to the nuclei of the drug's antipodes, causing differential shifts for some of the corresponding resonances. In many cases, this difference is sufficient to allow accurate quantification of each isomer. Once again, there is no line broadening like that with CLSRs and the higher-field spectrometers give better separation than the lower-field instruments.

2.2.2 Two-dimensional Techniques

2-D techniques in NMR spectroscopy all involve multiple-pulse experiments containing three basic stages, preparation, evolution, and detection. The spectrum that is finally obtained depends on the nature of the preparation and the length of the evolution time. The effects of evolution are not observed directly, but are detected by carrying out a series of experiments in which the evolution time is systematically varied. By following this procedure, one obtains two distinct sets of spectroscopic features, those that influence the magnetization during the evolution time and those that influence the magnetization during the acquisition time. Whereas in 1-D NMR, Fourier transformation converts the time dependence of the magnetization evolution during the acquisition time into frequency data, 2-D NMR uses a second Fourier transformation to convert the time dependence of the evolution time into a second frequency.

2.2.2.1 $^{1}H-^{1}H$ Correlation Spectroscopy Experiments There is a class of 2-D experiments which may be performed on any modern high-resolution NMR spectrometer which is referred to as correlation spectroscopy (COSY). One of the most common 2-D experiments is the proton–proton ($^{1}H-^{1}H$) COSY. The results for this type of experiment are usually displayed as a 2-D contour plot with the 1-D proton spectrum plotted along both the *x*- and *y*-axes. The contours in this plot indicate the chemical shifts of the protons that are coupled to each other, as well as for the protons themselves (i.e. diagonal peaks). Figure 1 shows the aliphatic portion of a typical COSY spectrum obtained for 4-propoxy-2,5-dimethoxyamphetamine. Looking under the peak at about 1.85 ppm, it can be seen that there is coupling to protons at 1.05 and 4.10 ppm. These are the signals for the propoxy chain. Similarly, looking under the signal at about 3.70 ppm, coupling to signals at 1.35 and 2.90 ppm can be seen, showing the grouping for the amphetamine alkyl chain. Note that there are no contours under the resonance at 3.90 ppm, other than the one on the diagonal, since there are no protons coupled to the methoxy protons. Thus, to determine which protons are coupled to a particular proton resonance, one simply scans down a column beneath its peak plotted on the *x*-axis. Contours will be found on rows corresponding to peaks plotted on the *y*-axis for protons to which it has detectable coupling. By varying the experimental parameters, one can look for either larger short-range or smaller long-range coupling constants. This experiment allows the analyst to determine which protons are separated from each other by two, three, or even four bonds.

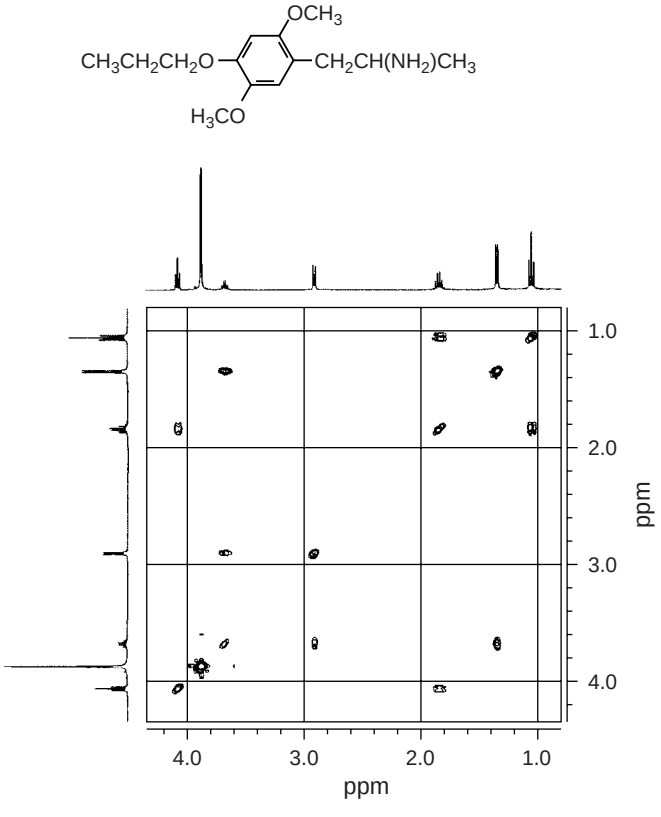

Figure 1 Partial COSY spectrum showing the correlations in the aliphatic region for 4-propoxy-2,5-dimethoxyamphetamine.

2.2.2.2 $^{13}C-^{13}C$ Correlation Spectroscopy Experiments The other homonuclear shift correlation experiment which may be used by the forensic analyst is the $^{13}C-^{13}C$ COSY. This technique is often referred to in the literature by its acronym INADEQUATE (incredible natural abundance double quantum transfer experiment). This is a very powerful experiment which allows the determination of all the one-bond carbon–carbon connections in the molecule, enabling the complete skeleton of the compound to be determined, except for those parts joined by heteroatoms. However, because of the extremely low sensitivity of this experiment, and the length of time required to accumulate the data, it is seldom used. Instead, analysts use a combination of other techniques to obtain the same information much more quickly.

2.2.2.3 $^{13}C-^{1}H$ Correlation Spectroscopy Experiments Because of the importance of determining the connections between hydrogen atoms and carbon atoms, a large number of $^{13}C-^{1}H$ COSY experimental techniques have been developed. In the 2-D contour plots for this type of experiment, the carbon spectrum appears along one axis and the proton spectrum along the other. Parameters may be chosen to emphasize large coupling constants in order to observe one-bond couplings, or to emphasize smaller couplings so that two- and three-bond couplings may be observed as well. In a related experiment, called relayed coherence transfer (RCT), parameters are used which give contours for each carbon atom corresponding to the directly attached protons as well as all the protons coupled to them (i.e. protons on nearby carbons).

2.2.2.4 Nuclear Overhauser Effect Spectroscopy The 1-D experiment for the NOE was discussed in section 2.2.1.2. There is a corresponding 2-D version of this experiment which goes by the acronym NOESY (nuclear Overhauser effect spectroscopy). Although the $^{1}H-^{1}H$ version of this experiment is more popular with forensic analysts, the $^{13}C-^{1}H$ version is also available. In the 2-D plot, contours are found for each pair of nuclei which would have produced a medium to strong NOE in the 1-D counterpart. In the 2-D experiment, some weak interactions that are detectable in the 1-D method are sometimes missed. However, one can get most of the results in a single 2-D experiment that would be obtained in many 1-D NOE experiments.

During recent years, advances in NMR technology have led to techniques which provide increased sensitivity over these 2-D methods just discussed. These advances include the introduction of ''inverse'' or ''reverse'' detection probes, gradient probes and decouplers which are capable of producing shaped pulses. These exciting advances mean that the forensic analyst can determine the structures of unknown compounds by using ever decreasing amounts of material.

3 APPLICATIONS

3.1 Confirmation of Known Compounds

Although forensic scientists have used NMR for tackling many types of problems, it is most often used in the study of abused drugs. While it is true that the majority of illicit drugs may be conveniently analyzed by other methods such as infrared (IR) and mass spectra, NMR is still used in a number of forensic laboratories. Some drugs that possess hygroscopic or polymorphic properties are difficult to analyze by IR. Also, many drugs fragment in such a manner that their mass spectra are ambiguous and difficult to interpret, especially in cases where there may be a number of closely related isomers. NMR provides an attractive alternative approach for these types of drugs since it is not affected by such problems and is a very powerful method for distinguishing between isomers.

3.2 Identification of Unknown Compounds

NMR spectroscopy provides the forensic analyst with one of the most powerful techniques for the identification of unknown compounds. For laboratories which have modern spectrometers, the full range of experiments described in section 2.2 are available. While IR and MS are useful for identification of known compounds, especially by means of their library search capabilities, NMR is usually the key spectroscopic technique for identifying true unknowns. For new abused substances such as ''designer drugs'', NMR proves to be invaluable in the determination of their structures. The term ''designer drug'' refers to compounds, related to abused drugs, which have been synthesized to circumvent legislation in the country of sale. These materials are usually a modified illicit drug in which a substituent has been added or changed (substituting an ethyl for a methyl, for example). Examples of these were reported in Canada and the USA during the late 1980s and early 1990s for fentanyls, amphetamines, synthetic opiates, and phencyclidines.

The exact series of NMR experiments used to solve unknown structures may vary somewhat, depending on the preferences of the particular spectroscopist. However, the general procedure will be the same. The analyst will carry out the following series of experiments until enough information has been obtained to identify the unknown, or until enough information is obtained for the investigator to be able to determine which specialized experiment is required to complete the study.

First, the abused drug sample is prepared for study as described in section 2.1. Then the spectroscopist obtains a high-resolution ^{1}H-NMR spectrum using an appropriate deuterated NMR solvent. Careful inspection of this spectrum gives a great deal of structural information. Chemical shifts and integration reveal the relative number of protons in each area (aliphatic, olefinic, aromatic) and the coupling patterns reveal information on the proximity of many protons. Next, ^{13}C-NMR and DEPT spectra are obtained to determine the number of different carbons and the numbers of protons attached to each. Chemical shifts of the carbons give important information about each carbon's environment. Often, the combination of proton and carbon data is enough to solve the structure of the unknown. If it is not, 2-D spectra and NOE experiments are used to establish connectivity and proximity of the nuclei in the molecule. The information obtained is sufficient to establish the skeleton of the molecule. Sometimes, especially when there are NMR-''silent'' nuclei (such as chlorine, bromine, oxygen, nitrogen) present, it may be necessary to obtain MS and/or IR data in order to complete the structure. While it is true that spectra can be obtained for most of the ''silent'' nuclei, it is usually much quicker and easier to use these other spectroscopic techniques to establish their presence.

3.3 Quantification

After the abused drug in a police exhibit has been identified, it is sometimes necessary for the analyst to provide quantitative results. In many seizures, the drug is found to be ''cut'' with other compounds ranging from sugars or starch to other drugs. Cocaine is one drug that is frequently cut with cheaper, more readily available drugs such as lidocaine. NMR provides a quick, accurate, and reliable method for quantifying abused drugs in most seizures. Depending on the complexity of the mixture, the analyst may use either proton or carbon spectral analyses. If absolute quantification is required, an accurately known amount of a suitable standard (see section 2.1) must be added to the sample before the analysis. If only relative quantities of substances present are required, there is no need for an internal standard. For relatively simple mixtures, proton NMR, because of the shorter relaxation times and higher sensitivity when compared with ^{13}C-NMR, is often the method of choice. However, for more

complex mixtures, ^{13}C-NMR may prove to be the better option. Although NMR is well suited for quantification of abused drugs, relatively few publications have appeared in this area. In part this is due to the fact that this method is so routine as to not warrant publishing the results.

3.4 Impurity Profiling

When a ^{1}H-NMR spectrum of an exhibit is recorded, signals are obtained for the protons of the impurities as well as for the abused drug itself. Careful examination of these ''extra'' peaks will often allow the identification of one or more impurities. Integration, or comparison of relative peak heights, will usually give a very good estimate of the percentage of each impurity present. In cases where the impurity(ies) cannot be identified in the presence of the drug, it is often possible to separate it (them) from the drug by methods discussed in section 2.1. Once the drug resonances no longer mask the impurity peaks, it is much easier to perform the required NMR techniques for the identification of this material.

Identification and quantification of the impurities in drugs is often helpful in police intelligence operations. Knowledge of the impurities in exhibits (including any solvent residues) can often give the analyst information concerning the synthetic route used to prepare the drug and solvents used in their cleanup/purification. Comparison of results for different seizures may then be used to determine if they are likely to be from the same source.

3.5 Optical Purity

There are a number of drugs for which only certain optical isomers are illegal. This differs from country to country. For example, methamphetamine is an illicit substance in Canada, whereas in the USA only the (+)-isomer is controlled. Levomethorphan is an internationally recognized narcotic, but its optical isomer, dextromethorphan, is a common ingredient in over-the-counter cough medications. Dextropropoxyphene is internationally recognized as a narcotic, while levopropoxyphene, which is used as an antitussive, is not. By using the methods described in section 2.2.1.4 NMR may quickly allow the determination of the optical purity of these types of drugs.

The study of optical purity in abused drugs is not limited to those drugs for which there is legislation against specific isomers. There are a number of abused drugs whose precursors are available in either racemic or optically pure form. Determining the optical content of the abused drug will sometimes allow the forensic analyst to determine the starting material and/or give evidence as to the synthetic route.

3.6 Analysis of Body Fluids

Modern forensic laboratories must often investigate the involvement of abused drugs in traffic accidents, murder, sexual assault, and other crimes. In many of these cases, the analysts will analyze samples of body fluids or tissues in attempts to measure levels of abused drugs. The concentration of drugs in the blood varies over a wide range, from tens of milligrams per milliliter for ethanol, to nanograms per milliliter for most drugs and their metabolites. However, in some extreme cases of overdoses, drug levels of 100 micrograms per milliliter are observed. The lower range of these drug and metabolite levels is beyond the sensitivity of NMR, but the higher end values are not. Detection limits for ^{1}H-NMR are in the low microgram per milliliter range on newer high-field spectrometers. It has proven difficult to obtain any lower limits due to the background signals from endogenous compounds in the blood and urine samples. The fluorine signals of fluorinated steroids are detectable at submicrogram per milliliter levels because there is no background. For nonfluorinated abused drugs, analysts have the option of performing sample work-up procedures such as solid-phase extraction or preparative HPLC cleanup before analysis. This also allows the analyst the opportunity to concentrate the drug or metabolite from several milliliters, down to the volume ($<$0.5 ml) required for NMR analysis, thus bringing many more samples into the range of levels detectable by NMR.

4 COMPARISON WITH OTHER FORENSIC TECHNIQUES

4.1 Advantages

NMR will detect all soluble organic compounds present in the sample. Other techniques, such as HPLC or MS, will often not detect impurities. For example, these techniques will not observe any compounds that either elute in the solvent front or do not come off the column at all under the conditions used. Careful examination of the impurity peaks in the NMR will often give significant clues as to their identity and allow quantitative estimates as to the amounts present. HPLC provides little information about the identity of unknown impurities and is not reliable for quantifying them since it does not give equimolar responses for different materials. In fact, compounds with low or no ultraviolet (UV) absorbance may not be detected, even when present at significant levels.

NMR is a nondestructive technique. As long as the sample material is stable in the solvent chosen for the study, the entire sample may be recovered and then

used for other studies. Hence, because a sample is not recoverable when techniques such as HPLC and IR are used, NMR studies should be performed before these other techniques whenever there is a limited amount of sample.

NMR spectroscopy provides the most powerful structure elucidation technique available to the forensic analyst for unknown materials. One area where NMR outshines other spectroscopic techniques is in the determination of isomers. Its true value comes from its ability to establish the proximity of nuclei in a molecule by using coupling constants (through one, two, or three bonds) or with NOE measurements (to within approximately 5 Å for nuclei). The NMR chemical shifts and coupling constants easily allow the discrimination of geometric or positional isomers, even when there are no authentic materials for comparison purposes. It would be much more difficult, if not impossible, to obtain this information using spectroscopic techniques such as IR or MS (without reference standards).

The most common method by which analysts determine the optical purity of drugs has been by either specific optical rotation or HPLC, using a chiral column or a chiral derivatization procedure. However, NMR has a number of advantages over optical rotation and HPLC in studies aimed at determining enantiomeric composition of any organic compound, including abused drugs. This is especially true if methods are required which are capable of determining optical antipodes at levels of less than a few percent. This can be important for police intelligence work where the starting materials for the abused drug may contain small amounts of its enantiomer. Establishing that the abused drug has the same optical content as the potential starting material gives good evidence of the synthetic route used. Using modern spectrometers and the techniques described above, it is usually possible to develop methods that can quantify the abused drug's enantiomer at or below levels of 0.1%. Some of the advantages of NMR would include the following.

- NMR results are accurate and precise, even when the substance of interest is present only at lower percentages. Optical rotation cannot be used to determine low levels of impurity, especially in drugs with a low specific optical rotation.
- The time required for analysis is usually less for NMR than for HPLC. This is especially true if there is no existing method, since it is usually quicker to develop an NMR method than an HPLC method. If a method already exists it is still much faster to obtain the results by NMR. For NMR, the procedure would normally involve simply dissolving the drug in an appropriate solvent, adding the calculated amount of CSA, acquiring the proton spectrum, and measuring the appropriate peaks in the spectrum. This may normally be done within half an hour. For HPLC, one must set up the LC system (put in the most suitable column for the analysis, prepare the mobile phase), check the performance of the column by running a standard, and then run the sample. All this requires much more than half an hour.
- NMR methods are more "robust" than HPLC methods. Once a method has been developed for NMR, it should work on any spectrometer of equal field strength (or higher field strength for CSAs and CDAs). Because of variability of columns, HPLC methods have to be checked and sometimes modified when changing columns.
- NMR provides "bonus" information. Since NMR detects all signals for the nucleus under study, it will detect not only the abused drug (and its enantiomer), but also any impurities present.

4.2 Disadvantages

One of the main criticisms of NMR as an analytical technique is its relative insensitivity when compared to other methods such as MS or HPLC. While these other instruments may require only nanograms or less of the sample, NMR usually needs micrograms or more. While this limitation is of little concern for drug seizures of one or more doses of the abused substances, it does limit its use in analyses of body fluids for drugs and their metabolites.

Another common complaint about NMR is the cost of the instrumentation. While it is true that the highest-field-strength spectrometers are very expensive (millions of dollars each), the models that have been around for several years (currently field strengths of 200 to 600 MHz) tend to be comparable in cost to mass spectrometers. In general, NMR spectrometers have proven to be very reliable instruments, with relatively low repair and maintenance costs. Thus, if one considers the cost per year, NMR spectrometers may not be as bad as is generally thought.

Another concern often brought up when discussing the use of NMR is the need for highly specialized expertise in the area. While it is true that considerable expertise is required in order to take full advantage of the full capabilities of modern spectrometers, very little knowledge is needed to use NMR results for more routine purposes. Thus, for positive identification of known compounds, simple comparison of spectra with those of authentic materials is all that is required. This is very similar to the use of IR "fingerprinting" to identify compounds. It is also possible to automate the identification by use of library searches of NMR databases, comparable to those for IR or MS.

5 SURVEY OF THE LITERATURE

Abused drugs may be described as belonging to either of two categories, (1) fully synthetic (e.g. amphetamines, Phencyclidine; 1-(1-phenylcyclohexyl)piperidine (PCPs)); and (2) natural source or semisynthetic modifications (e.g. cocaine, heroin). NMR spectra and spectral data have been reported for most abused drugs. There are a number of books which contain spectral data for drugs and related compounds. One set of books of particular interest to forensic analysts is the excellent collection of instrumental data, including 300 MHz proton NMR, MS, IR and UV spectra, found in Mills and Roberson.[1] There have also been two recent reviews[2,3] of the uses of NMR in forensic science.

5.1 Amphetamines and Related Substances

Amphetamine **(1)** belongs to a class of compounds referred to as phenethylamines. Abuse of these compounds began to increase progressively in the 1960s. Their effects range from aggression and psychosis to hallucination and distorted perception. Amphetamine and its analog, methamphetamine **(2)**, have medical uses and are legitimately manufactured. Clandestine laboratories produce these materials, as well as many substituted analogs, using a variety of synthetic routes. Although many possible synthetic routes for the production of amphetamines are described in the literature, only four are generally used in clandestine laboratories. These are the Leuckart synthesis, reductive amination, or reduction of the appropriate oxime or nitrostyrene. An examination of the impurities and optical state of the resulting materials is often useful in police intelligence work.

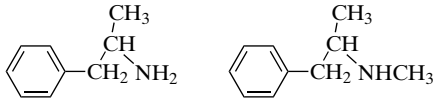

(1) **(2)**

The large number of substituted amphetamines and methamphetamines has required the development of analytical techniques capable of identifying these closely related compounds. Consider, for example, the simple case of ethoxydimethoxyamphetamines (analogs of trimethoxyamphetamines, with one methoxy being replaced by an ethoxy). There are 16 possible isomers, which may be split into three groups, those with the two aromatic protons ortho to each other (six isomers, Figure 2a), those with the aromatic protons meta to each other (seven isomers, Figure 2b) and those with the aromatic protons para to each other (three isomers, Figure 2c). The proton NMR spectra for each of these

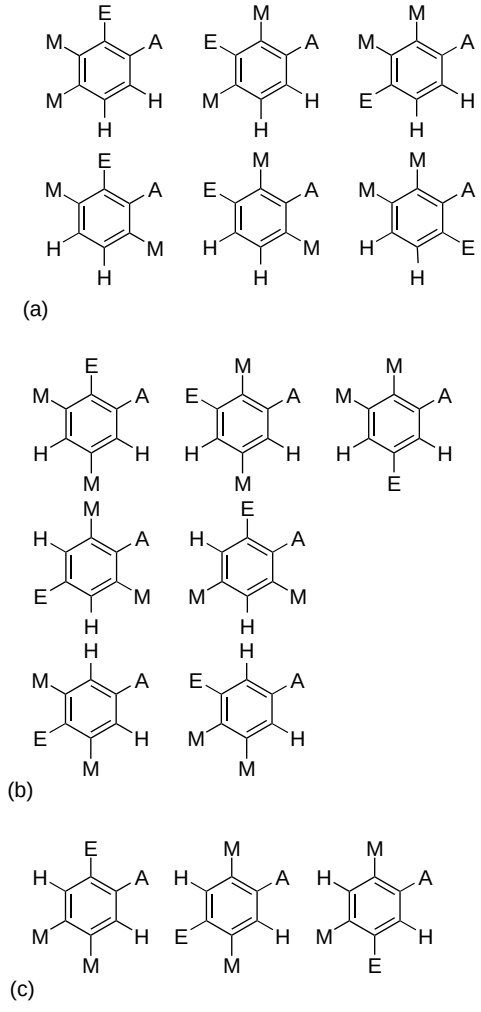

Figure 2 The 16 possible isomers for ethoxydimethoxyamphetamine divided into three groups: (a) those with the aromatic protons ortho to each other; (b) those with the aromatic protons meta to each other; (c) those with the aromatic protons para to each other. (A = amphetamine side chain ($CH_2CH(NH_2)CH_3$); E = ethoxy (OCH_2CH_3); M = methoxy (OCH_3)).

compounds would be expected to be distinct. Even without standards, it is not difficult to determine which isomer is present if NMR is used. First, an examination of the coupling constants for the signals for the aromatic protons reveals immediately to which of the three groups of isomers the compound belongs, since ortho couplings are ~8 Hz, meta couplings are ~2 Hz and para couplings are usually not resolvable (~0 Hz). Performing a few NOE experiments will then allow the determination of which substituents are adjacent to each other, thus completing the structure determination.

Figure 3 illustrates how the proton spectrum and the results of some NOE experiments may be used to determine the structure for one isomer. The resonances in this spectrum (a) are readily assignable: the two aromatic

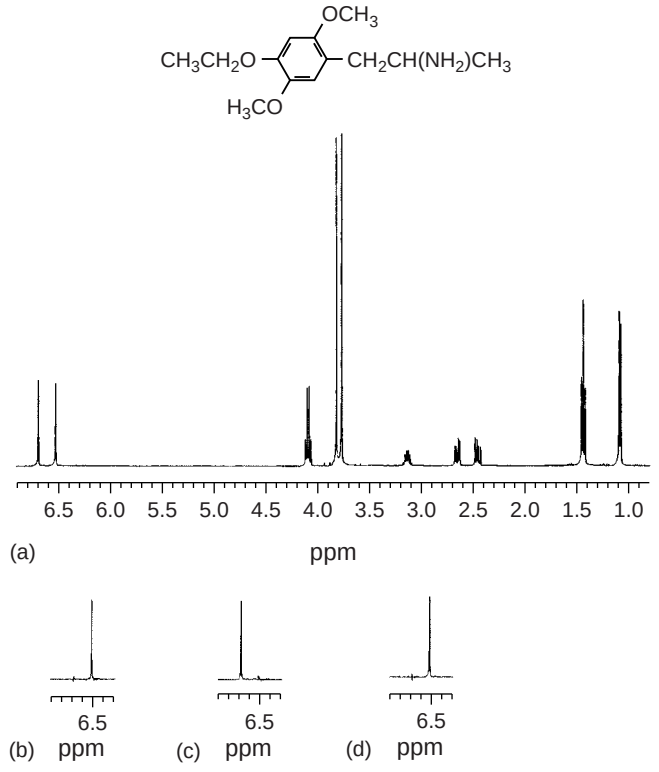

Figure 3 Spectrum of one isomer of ethoxydimethoxyamphetamine (a) and results in the aromatic region for the NOE experiments for the irradiation of the methoxy groups at 3.78 ppm (b) and 3.82 (c) and the ethoxy methylene protons at 4.10 ppm (d).

protons resonate at 6.52 and 6.70 ppm, the methoxy groups at about 3.78 and 3.82 ppm, the ethoxy shows methylene protons as a quartet at 4.10 ppm and a methyl triplet at 1.43 ppm, and the amphetamine alkyl chain gives a multiplet at 3.14 for the methine, two multiplets between 2.4 and 2.7 for the methylene, and a doublet for the methyl at 1.1 ppm. Since the aromatic protons appear as singlets, the two protons are para to each other. The results in the aromatic region for the NOE experiments, irradiating the methoxy and ethoxy ethylene protons, are shown in Figure 3. Since enhancement for the aromatic proton at 6.52 ppm is observed for the NOE experiments with the high-field methoxy (Figure 3b) and the ethoxy methylene (Figure 3d), this proton must be assigned to the 3 position, with the 6.70 ppm then being for the proton in the 6 position. Since the ethoxy NOE experiment enhances the proton in the 3 position, the ethoxy group must be in the 2 or 4 position. Irradiation of the amphetamine methylene protons gives an NOE to the high-field methoxy signal (results not shown), meaning that it is in the 2 position and thus the ethoxy is in the 4 position, completing the assignment. Similar procedures could be used for the other isomers.

Most of the abused drugs in this class are synthetic in origin. There are a few exceptions. The hallucinogen mescaline, from the peyotyl cactus, *Lophophora williamsii* or *Anhalonium lewinii*, was popular in the 1960s. More recently, another natural product, khat, *Catha edulis*, has shown up in a number of countries. Khat is a small tree or shrub that is native to east Africa and southern Arabia. Its leaves are used for chewing to get an amphetamine-like stimulation. The principal active components in khat are cathinone **(3)** and cathine **(4)**. Study of the alkaloid content, including optical isomers, by NMR methods can be used to provide police with information about the origin of the material.

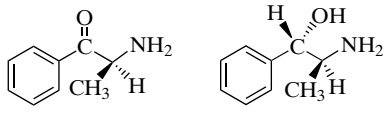

(3) **(4)**

5.2 Phencyclidine and Related Substances

PCP **(5)** was first marketed as an anesthetic in 1958, but was withdrawn in 1967 because of adverse effects. Despite its neurotoxic effects, it became a drug of abuse because of its hallucinogenic and stimulant effects. A large number of analogs have also appeared as "street drugs" or "designer drugs". The PCPs produced in clandestine laboratories are known to often be impure, sometimes containing the toxic intermediate 1-piperidinocyclohexanecarbonitrile (PCC) **(6)**. As with the amphetamine designer drugs, unknown new PCP analogs are usually most easily identified by means of NMR spectroscopy. Figure 4 shows the spectra of two PCP analogs and illustrates the ease

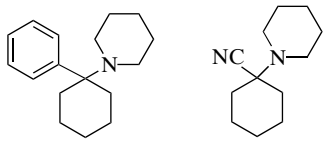

(5) **(6)**

with which they may be identified from simple proton NMR spectra. MS may be used to determine that the molecular weight of these compounds is 14 higher than that of PCP, indicating a possible methylation product, but without authentic standards the isomers may not be easily determined. Examination of the spectrum in Figure 4(a) reveals a singlet at about 2.4 ppm which integrates to three protons. This indicates that the compound probably contains a methyl on the aromatic ring. The appearance of the aromatic proton resonances indicates para substitution, showing that the methyl is in the 4 position. The spectrum of the other isomer shown

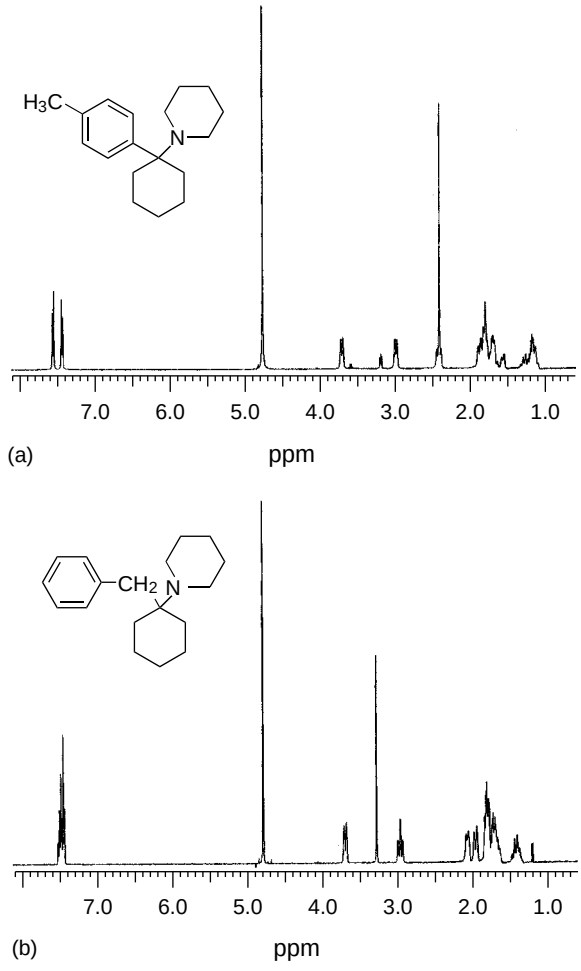

Figure 4 Proton spectra of two PCP analogs: (a) the 4-methyl derivative; (b) the benzyl derivative.

in Figure 4(b) contains a singlet at about 3.2 ppm which integrates to two protons. The aromatic ring resonances integrate to five protons. Combining this information reveals that this isomer is the benzyl derivative, rather than another aromatic methylated isomer.

5.3 Fentanyls

The abuse of fentanyl (*N*-(1-phenethyl-4-piperidyl)propionanilide) **(7)** and its related compounds began in the USA in 1979. Fentanyl is a narcotic analgesic widely used as a surgical anesthetic. This drug and many related compounds have activities which make them more than 100 times more potent than morphine. Because of this, more than 100 deaths have been attributed to overdose from these compounds. One of the first analogs to appear on the streets was referred to as ''China White'' or ''synthetic heroin''. Its identity was determined as α-methylfentanyl **(8)** by using a combination of MS and ^{1}H-NMR. A large number of designer drugs of this class appeared in the early 1980s and clandestine laboratories continue to be found, especially in the USA.

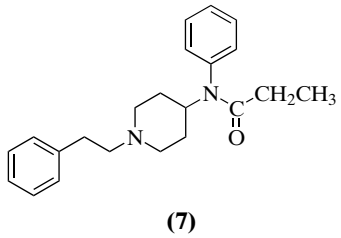

(7)

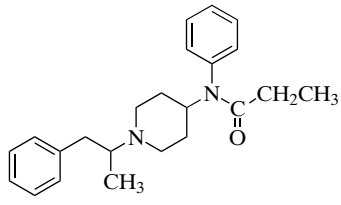

(8)

5.4 Opiate Alkaloids

Opium is derived from the opium poppy (*Papaver somniferum*) or other poppy species which may be grown in many countries. Untreated opium contains a mixture of about 30 alkaloids. Morphine **(9)** is the principal opium alkaloid, with some of the other common alkaloids being codeine **(10)**, thebaine **(11)**, papaverine **(12)**, and noscapine **(13)**. Heroin **(14)**, the main opiate-derived abused drug, is produced by the simple diacetylation

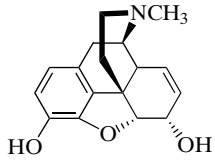

(9)

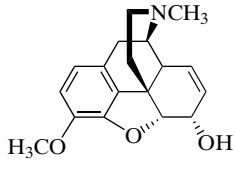

(10)

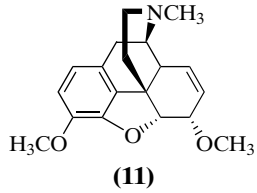

(11)

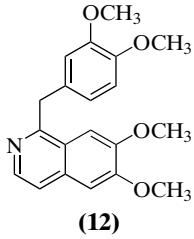

(12)

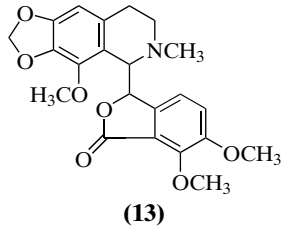

(13)

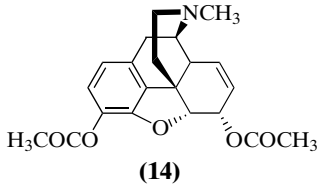

(14)

of morphine. All these related compounds may be readily identified by their distinctive NMR spectra. Also, the impurity profiles for these compounds may be very useful in police intelligence work, as discussed above. Illicit heroin samples are usually cut with a range of other substances, such as caffeine, diazepam, and methaqualone. It has been found that the street samples of heroin generally average a purity of less than 40%. Identification of the other materials may allow law enforcement agencies to link seizures from different sources.

5.5 Cocaine and Related Substances

Cocaine **(15)**, the first local anesthetic to be discovered, is a strong central nervous system stimulant whose abuse has grown greatly in the last two decades in many parts of the world. Cocaine is a tropane alkaloid, as are atropine and scopolamine, both of which have valuable medical uses. Although cocaine has been synthesized, and there have been reports of a few clandestine laboratories producing it, cocaine is obtained primarily from the leaves of the shrubs *Erythroxylum coca* or *Erythroxylum novogranatense*, which are cultivated in South America. The leaves are extracted in clandestine laboratories close to the growing regions. Once again the impurities (including the other minor alkaloids) give valuable intelligence information to the law enforcement agencies. As with the other drugs discussed, NMR provides a quick and accurate method for easily distinguishing between cocaine and all the other closely related alkaloid compounds. Since cocaine is rapidly metabolized, only metabolites are seen in body fluids. These include benzoyl ecgonine and ecgonine methyl ester. If alcohol drinking is combined with cocaine abuse, benzoyl ecgonine ethyl ester is formed in the body. All these compounds have been studied and fully characterized by NMR spectroscopy.

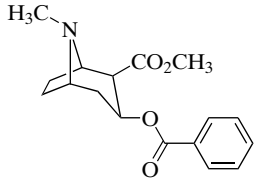

(15)

5.6 Cannabinoids

Cannabis is the name given to material prepared from the *Cannabis sativa* plant. The preparation occurs as herbal cannabis (marijuana, the flowering or fruiting part of the plant) or cannabis resin (an amorphous semi-solid exuded by the flowering tops). Cannabis is a mixture of more than 60 compounds, the cannabinoids. It is a mild hallucinogen, which induces light-headed euphoric intoxication. The primary active substance in cannabis is Δ^9-tetrahydrocannabinol **(16)**, which was identified by use of ^{1}H-NMR. Since the mid-1960s, more than 100 other cannabis substances have been identified, with ^{1}H-NMR playing a major role in the elucidation of their structures. NMR has also been used in the identification of a number of related semisynthetic drugs. Although NMR is not appropriate for the routine analyses of cannabis products, it is used for the authentication of pure substances to be used by other methods and for the identification of new related unknown compounds.

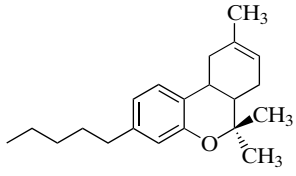

(16)

5.7 Ergot Alkaloids

Ergot is a dried material from the fungus *Claviceps purpurea*, which grows on rye and other grains. Ergot alkaloids have also been found in many plant species, such as the *Convolvulaceae* (morning glories). Ergot material yields four main classes of alkaloids: lysergic acids, lysergic amides, clavine, and ergot peptides. The most common drug of abuse derived from ergot is LSD (or lysergide) **(17)**, which is produced by the amination of lysergic acid. The abuse of LSD began in the 1960s. Currently, LSD is encountered as either microdots (small tablets), or, more commonly, as impregnated paper or card squares in one of more than 200 printed designs reported so far.

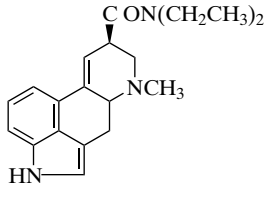

(17)

LSD is a very potent hallucinogen, with street samples normally containing between 25 and 100 μg per dose. Since LSD is unstable in the presence of light or high temperatures and moisture, the forensic analyses present some difficulty. Gas chromatography (GC) analysis requires conditions that are highly specific for this drug. Thus either HPLC or ^{1}H-NMR is more useful. Even with the lowest amount which will give a perceptible effect (~10 μg), it is feasible to use modern high-field

spectrometers (≥400 MHz) to obtain proton spectra in a reasonable amount of time.

As discussed above, it is very important to know the identity of a substance when working in the judicial system. It is crucial to be able to distinguish between isomers of compounds, especially if one of them may be an illicit material while the other may be a legal substance. In the early 1980s in the USA, it was realized that existing chromatographic methods for LSD were not adequate for distinguishing this abused drug from the noncontrolled isomer, lysergic acid methylpropylamide (LAMPA). It is no problem to distinguish LSD and LAMPA (or the methylisopropyl isomer) by simple ^{1}H-NMR. Chromatographic and MS methods have also been developed for differentiating LSD and LAMPA.

5.8 Quinazolinones

There are a number of abused drugs which belong to the quinazolinone category. The two most often encountered in forensic analysis are methaqualone **(18)** and mecloqualone **(19)**. Both are hypnotics, manufactured as proprietary pharmaceuticals. Illicit samples may arise either by diversion of prescription tablets or from clandestine laboratories. Methaqualone is one of the substances which is sometimes used as a diluent for heroin samples. NMR spectra for these compounds are readily available in the open literature.

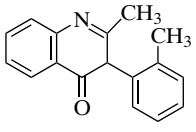

(18)

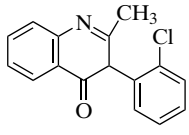

(19)

There has been one report[4] of a ''designer'' methaqualone. Spectroscopic evidence indicated that it was a methylated analog. The problem was to determine the position of the methyl group. It was known to be one of the four isomers with the methyl on the phenyl ring. The coupling pattern indicated that the methyl was in either the 4 or 5 position. NOESY correlations were then used to show definitively that the compound was 2-methyl-3-(2,4-dimethylphenyl)-4(3*H*)-quinazolinone **(20)** rather than the 2,5-dimethyl isomer. The structure was confirmed by synthesis. The authors discussed the improved efficiency of using NMR as a means of eliminating the need for synthesis of all the other possible isomers.

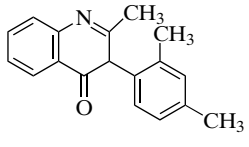

(20)

5.9 Barbiturates

Some of the earliest forensic NMR work was done on barbiturates. These studies showed that NMR provided an excellent method for identifying these compounds, even in mixtures which were difficult to analyze by other methods, without the need for prior separation. There are a number of papers, including one by Wilson et al.[5] (which contains spectra of all barbiturates and their salts that are under international control), which may be used to provide the forensic analyst with all the spectral data necessary to identify any abused drugs in this class.

5.10 Benzodiazepines

In most countries, benzodiazepines have, to a large extent, replaced barbiturates as the sedative or tranquilizer of choice by drug users. There have been reports of the use of these medications as ''club'' or ''date rape'' drugs. Wilson and Avdovich[6] have shown that ^{1}H-NMR may be used for the identification, differentiation, and analysis of these drugs and presented a complete collection of 80 MHz spectra for all 33 benzodiazepines under international control.

5.11 Other Abused Drugs

NMR spectra and/or spectral data have been reported for most abused drugs that do not fall into one of the classes discussed above. In most cases, although the data were not reported in the context of forensic analysis, they may still be used for such purposes.

Extensive amounts of ^{1}H- and ^{13}C-NMR data have been published for steroids, including anabolic steroids which have been a problem in sports regulation for many years. Currently, GC/MS is the most important technique for the forensic detection of steroids and their metabolites. There have been very few reports of the use of NMR in forensic analysis of steroids. However, NMR should not be overlooked as a quick method of identification of (purified) steroids when GC/MS is not appropriate. An important advantage of NMR over other techniques is its inherent ability to identify ''designer'' modifications in drugs such as steroids.

6 CONCLUSIONS

NMR has played an important role in the forensic analysis of drugs of abuse for many years. Although it has not been used as the method of choice for routine analyses, the advent of NMR spectrometer accessories for automated data collection and analysis may change this situation in the near future. There is no doubt that NMR will continue

to play a vital role in the identification of unknown compounds such as ''designer drugs''.

ABBREVIATIONS AND ACRONYMS

CDA	Chiral Derivatizing Agent
CLSR	Chiral Lanthanide Shift Reagent
COSY	Correlation Spectroscopy
CSA	Chiral Solvating Agent
DEPT	Distortionless Enhancement by Polarization Transfer
FTNMR	Fourier Transform Nuclear Magnetic Resonance
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
IR	Infrared
LAMPA	Lysergic Acid Methylpropylamide
LSD	Lysergic Acid Diethylamide
LSR	Lanthanide Shift Reagent
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PCC	1-Piperidinocyclohexane-carbonitrile
PCP	Phencyclidine; 1-(1-Phenylcyclohexyl)piperidine
RCT	Relayed Coherence Transfer
RF	Radiofrequency
TMS	Tetramethylsilane
UV	Ultraviolet
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Two-dimensional Nuclear Magnetic Resonance of Small Molecules

REFERENCES

1. T. Mills, III, J.C. Roberson, *Instrumental Data for Drug Analysis*, Elsevier, New York, Vols. 1–4, 1984.
2. B.A. Dawson, 'The Use of Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantitation of Abused Drugs', in *The Analysis of Drugs of Abuse*, ed. T.A. Gough, John Wiley & Sons Ltd, New York, 283–309, 1991.
3. C.J. Groombridge, 'NMR Spectroscopy in Forensic Science', *Annu. Rep. NMR Spectrosc.*, **32**, 215–297 (1996).
4. S.A. Angelos, D.C. Lankin, J.A. Meyers, J.K. Raney, 'The Structural Identification of a Methyl Analog of Methaqualone via 2-dimensional NMR Techniques', *J. Forens. Sci.*, **38**, 455–465 (1993).
5. W.L. Wilson, B.A. Dawson, H.W. Avdovich, 'The Identification and Analysis of Barbiturates Under International Control. I. ^{1}H Nuclear Magnetic Resonance Spectroscopy', United Nations Scientific and Technical Notes, SCITEC/7, United Nations, New York, 1989.
6. W.L. Wilson, H.W. Avdovich, 'The Identification and Analysis of Benzodiazepines Under International Control. II. Nuclear Magnetic Resonance Spectroscopy', United Nations Scientific and Technical Notes, SCITEC/4, United Nations, New York, 1988.

Polymerase Chain Reaction in the Forensic Analysis of DNA

Charles M. Strom
Illinois Masonic Medical Center, Chicago, USA

1 Introduction 1
1.1 The Human Genome 1
1.2 Hypervariable Regions 1
1.3 History of DNA Forensic Analysis 1
1.4 Comparison of Polymerase Chain Reaction to Other Techniques 2
2 Description of Technique 2
2.1 Description of Various Forms of Polymerase Chain Reaction 2
2.2 Polymerase Chain Reaction Primers 2
2.3 DNA Preparation 3
2.4 Preparation of Polymerase Chain Reaction Mixture 3
2.5 Model Polymerase Chain Reaction Cycle 3
2.6 Complexity of Polymerase Chain Reactions 3
2.7 Exponentiality of Reaction 5
3 Analysis of Polymerase Chain Reaction Products 5
3.1 Manual Techniques 5
3.2 Automated Techniques 5
4 Loci used in Forensic Polymerase Chain Reaction Testing 5
4.1 Short Tandem Repeats and Variable Number of Tandem Repeats 5
4.2 Kits 5
4.3 Mitochondrial DNA Testing 5
Abbreviations and Acronyms 5
Related Articles 6
References 6

This article summarizes the theory behind the uses of DNA analysis for identification purposes. Various techniques are summarized and their strengths and weaknesses are contrasted. The theoretical basis of the polymerase chain reaction (PCR) and various types of PCR reactions are introduced. A table is provided with loci that can be used for the forensic analysis of DNA. The availability of kits is summarized. Finally, a discussion of mitochondrial DNA analysis is performed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 The Human Genome

The human genome consists of approximately 3×10^9 basepairs (bp) of DNA arranged on 46 chromosomes. An individual gene or location on a chromosome is called a locus. There are 44 autosomes (numbered 1–22) and two sex chromosomes (X or Y) in each individual. There are two homologous autosomes, one inherited from the mother and one inherited from the father. Females inherit an X chromosome from both parents and males inherit an X from their mother and a Y from the father. Thus, every individual has two copies of every autosomal gene. Each copy of a gene is called an allele. When the paternal and maternal alleles are identical, an individual is homozygous at that locus, and when the paternal and maternal alleles are different, an individual is heterozygous at that locus.

1.2 Hypervariable Regions

In the human genome there are noncoding regions that consist of stretches of repeating sequences of DNA. The sequences can vary in size from 2 bp (dinucleotide repeats) to more than 50 bp repeated for several iterations. The number of these repeated sequences present on any individual's chromosomes differs from chromosome to chromosome and from individual to individual. Repeated fragments of 350–1000 bp are generally referred to as amplified fragment length polymorphisms (AmpFLPs) and shorter fragment sizes in the range 100–300 bp are called short tandem repeats (STRs).[1] There are other areas of variability, the most important of which are the human lymphocyte A (HLA) histocompatability loci. Analysis of the protein polymorphisms in the HLA loci has long been a cornerstone of serological paternity testing. PCR-based HLA typing is also possible and is the basis of commercially available kits.[2]

1.3 History of DNA Forensic Analysis

Initially DNA forensic testing relied on one of two seminal techniques: Southern blot analysis of regions with a variable number of tandem repeats (VNTR) (single locus gene testing);[3] and a technique developed by Dr Alec Jeffries, which used probes that recognize a number of loci in the genome, called multilocus DNA testing (popularly known as DNA fingerprinting).[4] It is important to note that DNA fingerprinting is a patented technology of Cellmark Diagnostics and should not be used to describe any other type of DNA test. Both of these techniques have largely been supplanted by PCR technologies. PCR is a technique

that amplifies a particular sequence of DNA and allows the analysis of any known DNA sequence by amplifying the sequence of interest approximately a million times. Once the PCR product is obtained, analysis can be performed by a wide variety of techniques that include size fractionation, hybridization, DNA sequencing and various microchip technologies (see below).

1.4 Comparison of Polymerase Chain Reaction to Other Techniques

The advantages of PCR over prior technologies are that PCR requires much less DNA (even a single human cell can be genotyped reliably) and is successful on partially degraded or damaged DNA, technical aspects are simple and allow automation, analysis does not require radioactivity, and the analysis can be accomplished in a matter of hours rather than days for Southern blot technologies. In addition, the fragment sizes are discrete, allowing for easy comparison among laboratories and standardization for the purposes of offender databases. The disadvantages of PCR are that the polymorphic loci tend to be less variable, so a larger number of loci are required to achieve equivalent statistical significance, and because PCR is capable of amplifying minute amounts of DNA, contamination by extraneous DNA sequences at crime scenes or forensic laboratories is a much more significant problem in PCR than in standard DNA techniques.

2 DESCRIPTION OF TECHNIQUE

2.1 Description of Various Forms of Polymerase Chain Reaction

PCR uses a thermally stable DNA polymerase and two or more synthetic oligonucleotides that are usually 16–25 bp in length. For forensic analysis the primers are designed to frame a hypervariable region of DNA. In simple PCR two primers are used: an upstream primer on the 5′ end of the locus of interest and a downstream primer on the 3′ end. The two primers must be complementary to opposite strands of the DNA (see Figures 1 and 2). In nested PCR,[5] two sequential PCR reactions are performed. The first reaction uses two outside primers (see Figure 2) and 20–30 cycles of PCR are performed. Following this first round of PCR, a small aliquot of the initial PCR mixture is added to a new tube with PCR buffer and two inside primers. Then a second round of PCR is performed. Nested PCR was developed to genotype reliably single cells and is sometimes able to amplify minute quantities of DNA when standard PCR fails. In hemi-nested PCR one of the primers added to the second round is the same as one of the primers used in the first round of PCR.

2.2 Polymerase Chain Reaction Primers

Primers are usually 16–30 bp in length and can be synthesized on commercially available oligonucleotide synthesizers or ordered from a number of commercial laboratories. Oligonucleotide synthesizers use phosphoramidite chemical synthesis, with the first nucleotide linked to a

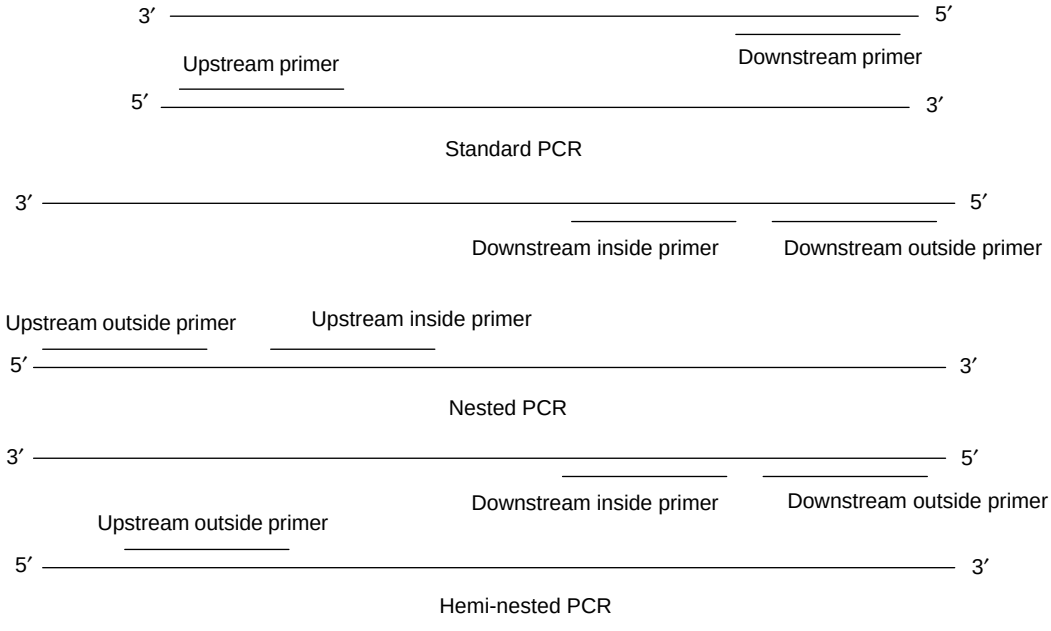

Figure 1 Various schemes of PCR reactions.

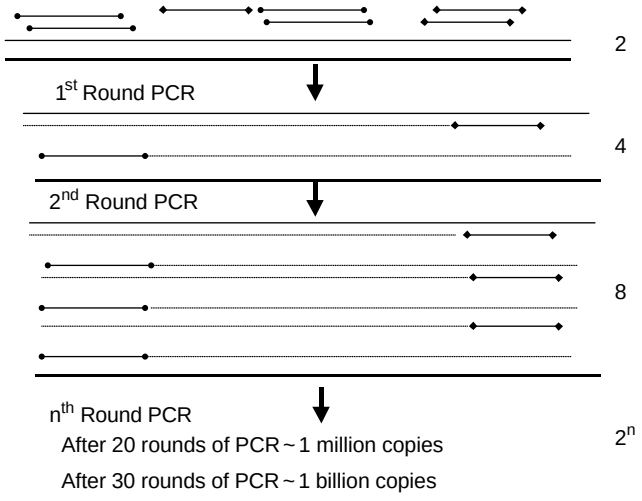

Figure 2 Schematic of PCR reaction.

solid support. Then successive nucleotides are added in a 3′–5′ direction by automated phosphoramidite synthetic steps. Following the synthesis, the oligo is removed from the column. The final nucleotide is protected with a trityl group. Synthesizers can be programmed to leave the final trityl group ''on'' (trityl-on synthesis) or to cleave the final group (trityl-off synthesis). In trityl-on synthesis all oliognucleotides containing less than the full-size product will not have a trityl group at the 5′ end, because the first step in the subsequent round of synthesis de-protects the last nucleotide. Therefore, a purification step that separates all trityl-containing nucleotides from nontrityl-containing nucleotides will assure that only full-length oligonucleotides are present. However, for most forensic purposes, primers synthesized trityl-off are equally efficacious.

2.3 DNA Preparation

There are several methods to prepare DNA for PCR reactions. These include proteinase K digestion with or without organic extraction, organic extraction only, chelex, alkaline lysis buffer, and simply dissolving the specimen in PCR buffer. DNA preparation from vaginal or anal swabs in rape cases can be subjected to differential lysis procedures. In this procedure, a gentle lysis is performed using detergent and proteinase K to lyse vaginal epitheleal cells, leaving the sperm intact. Then the two fractions are separated and the sperm are lysed by the addition of ethylenediaminetetraacetic acid (EDTA) or dithiothreitol (DTT) to the lysis mixture. Successful PCR analysis has been made from licked stamps, saliva on skin, cigarette butts, fingerprints, hair roots, tooth pulp, paraffin-imbedded tissue sections, and histology slides, in addition to the standard tissues of blood and cheek swabs.[6–40]

2.4 Preparation of Polymerase Chain Reaction Mixture

Once the DNA has been prepared, the PCR mixture is prepared. In simple PCR, one set of two primers is added, along with sufficient quantities of the four deoxyribonucleotide triphosphates (dNTPs), a thermally stable DNA polymerase, in the appropriate salt and pH for the particular polymerase and primer set. The reaction mixture is then placed in a thermocycler (PCR machine). Thermocyclers are programmable instruments that perform repeated cycles of heating and cooling of samples for PCR.[41–43]

2.5 Model Polymerase Chain Reaction Cycle

For most applications, ''Hot Start'' PCR is the preferred method. If all the reagents are present in the PCR reaction when it is placed into the thermocycler, nonspecific primer binding to DNA can occur as the temperature begins to increase from room temperature to the initial melting temperature. As soon as the priming occurs, the polymerase can begin chain elongation, thus creating a nonspecific PCR product that participates in subsequent cycles and can interfere with interpretation of the PCR reaction. In Hot Start PCR, one or more critical reagents are left out of the PCR mixture (dNTPs or polymerase) and not added until after the initial melting step while the reaction mixture is held at the elongation temperature. A typical PCR cycle would be as follows:

95 °C for 10 min (melting of DNA performed only once)
72 °C until final reagent is added (for Hot Start)
Begin performing the following cycle
65 °C for 2 min (reannealing)
72 °C for 2 min (elongation)
95 °C for 2 min (melting)
Cycle for 18–40 cycles
72 °C for 10–20 min (final chain elongation, performed only once)
4 °C indefinitely

2.6 Complexity of Polymerase Chain Reactions

PCR reactions can be separated according to the complexity of analysis, as follows:

Standard PCR – one set of two primers, one PCR reaction, one locus analyzed
Fully nested PCR – two sets of primers, two PCR reactions, one locus analyzed
Hemi-nested PCR – two sets of primers, but one outside primer is shared by both reactions
Multiplex PCR – two or more sets of primers, one PCR reaction, two or more loci analyzed
Nested multiplex PCR – several reactions, $2n$ sets of primers (n = number of cycles)

Table 1 Loci that can be used for forensic analysis of DNA

Locus	Chromosome	Gene name	Type
D1S80	1	pMCT118	~10-bp STR
D1S103	1	None	STR
D1S1656	1	None	4-bp STR
Apo-B	2	Apolipoprotein B	AmpFLP
TPOX	2	Thyroid peroxidase	4-bp STR
D2S436	2	None	4-bp STR
D3S38	3	None	STR
D3S1358	3	None	STR
D3S1359	3	None	STR
FIBRA	4	FGA	STR
D4S43	4	None	AmpFLP
GABRB15	4	None	4-bp STR
CSF1PO	5	cfms oncogene	4-bp STR
D5S818	5	None	STR
F13A101	6	Coagulation factor XIII	4-bp STR
ACTBP2	6	β-Actin pseudogene	4-bp STR
FOLP23	6	None	4-bp STR
HLA-DR	6	HLA-DR	Polymorphism
D7S820	7	None	STR
D7S809	7	None	4-bp STR
D7S22	7	None	VNTR
D8S1179	8	None	4-bp STR
Lewis antigen	9	Lewis blood type	Polymorphism
ABO	9	ABO blood typing	Polymorphism
THO1	11	Tyrosine hydroxylase	4-bp STR
D11S554	11	None	4-bp STR
D12S391	12	None	STR
CD4	12	CD4	5-bp STR
DRPLA	12	Dentatorubral-pallidoluysian atrophy	STR
VWF31A	12	von Willebrand factor	4-bp STR
D13S317	13	None	STR
D13S358	13	None	STR
CYAR04	15	None	4-bp STR
D16S539	16	None	STR
FES/FPS	15	cfes/fps oncogene	4-bp STR
D17S5	17	YNZ22	AmpFLP
D18S51	18	None	4-bp STR
D18S35	18	None	4-bp STR
D20S470	20	None	4-bp STR
D20S85	20	None	STR
D21S11	21	None	4-bp STR
GPT		Glutamate pyruvate transaminase	PCR/RFLP
DXS52	X	None	AmpFLP
HPRT	X	Hypoxanthine phosphoribosyltransferase	
Amelogenin	X and Y	Amelogenin	Polymoprhism
ZFX/ZFY	X and Y	Zinc finger protein	RFLP
Alphoid repeats	X and Y	None	Polymorphism
DXYS156	X and Y	None	STR
DYS389	Y	None	STR
SRY	Y	Testis determining factor	Polymorphism
DYS18	Y	None	4-bp STR
DYS385	Y	None	STR
DYS19	Y	None	STR
DYS390	Y	None	STR
DYS391	Y	None	STR
DYS393	Y	None	STR
DYS389-II	Y	None	STR
DYS392	Y	None	STR

RFLP, restriction fragment length polymorphism.

2.7 Exponentiality of Reaction

The mechanism of the PCR reaction is shown in Figure 2. Theoretically, the reaction is exponential, with $2n$ copies made.

3 ANALYSIS OF POLYMERASE CHAIN REACTION PRODUCTS

3.1 Manual Techniques

Once the PCR reaction has been completed, the products can be analyzed in a number of ways. The most common AmpFLP and STR analytical method is gel electrophoresis with ethidium bromide staining. Ethidium bromide is a dye that intercollates into the DNA helix between bases. When exposed to ultraviolet light, the ethidium-bromide-stained DNA fluoresces, allowing a photograph to be taken. Silver stain has also been used to visualize PCR products.[44–53]

3.2 Automated Techniques

Newer methodologies utilize fluorescent PCR primers. This allows detection of PCR fragments on automated DNA sequencers. In the sequencer detection, the PCR products are separated on either denaturing acrylamide gels or by capillary electrophoresis. The fragments are detected by a laser fluorescence detector (or detectors). Some instruments have the capability of analyzing four different colors. If one designs the primer systems correctly so that ranges of the fragment lengths differ for each system without overlap, it is possible to analyze several different systems using the same color. Then, combining such a system with the capability of using three different colors allows a simultaneous assay of more than 20 systems.[44–53]

4 LOCI USED IN FORENSIC POLYMERASE CHAIN REACTION TESTING

4.1 Short Tandem Repeats and Variable Number of Tandem Repeats

Table 1 is a summary of published loci that have been described to discriminate among individuals.[54–145]

4.2 Kits

Some commercially available kits are available for forensic PCR analysis. Some of these kits are listed in Table 2.[146–162]

Table 2 Kits available commercially for forensic polymerase chain reaction analysis

Kit	Manufacturer
AmpliType – LDLR,GYPA, HBGG,D7S8,Gc	Roche Molecular Systems/ Perkin-Elmer
PM + HLA DQ-alpha	Perkin-Elmer
AmpFISTR-Blue	PE – Applied Biosystems
FK06 (CD4, FES,FGA)	Serak (Germany)

4.3 Mitochondrial DNA Testing

PCR can also be used to analyze mitochondrial DNA. Mitochondrial DNA analysis has two advantages over standard PCR analysis. Because mitochondrial DNA is inherited completely from the mother, mitochondrial DNA analysis allows easier identification of unknown individuals from mass graves and mass disasters. In standard STR or AmpFLP analysis, reverse paternity testing must be performed. In order to accomplish this, samples from both parents are required for analysis. In cases of genocide and war crimes, this is rarely possible. In addition, any relative related through the mother can be used as a source of DNA for mitochondrial DNA identification. Another advantage is that mitochondrial DNA is present in hundreds to thousands of copies per cell, in contrast to genomic DNA that is present in only one copy per cell. Therefore, some samples that have insufficient DNA for standard PCR or nested PCR analysis may yield results with mitochondrial DNA amplification. However, mitochondrial DNA analysis has drawbacks compared to standard analysis for identity testing. In general, mitochondrial DNA is much less polymorphic than the STR or AmpFLP loci, so that in the event of a nonexclusion, the statistical power of the conclusions is not as significant.[163–169]

ABBREVIATIONS AND ACRONYMS

AmpFLP	Amplified Fragment Length Polymorphism
dNTP	Deoxyribonucleotide Triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
HLA	Human Lymphocyte A
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
STR	Short Tandem Repeat
VNTR	Variable Number of Tandem Repeats

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry • DNA Arrays: Preparation and Application • Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry

Forensic Science **(Volume 5)**
Forensic Science: Introduction • DNA Extraction Methods in Forensic Analysis

Nucleic Acids Structure and Mapping **(Volume 6)**
Polymerase Chain Reaction and Other Amplification Systems • RNA Tertiary Structure • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

Liquid Chromatography **(Volume 13)**
Ion Chromatography

REFERENCES

1. B. Brinkman, 'Overview of PCR Based Systems in Identity Testing', in *Forensic DNA Profiling Protocols*, eds. P.J. Lincoln, J. Thomson, Humana Press, Totowa, NJ, 105–120, 1998.
2. M. Baird, 'Use of the Amplitype PM + HLA DQA1 PCR Amplification and Typing Kits for Identity Testing', in *Forensic DNA Profiling Protocols*, eds. P.J. Lincoln, J. Thomson, Humana Press, Totowa, NJ, 261–278, 1998.
3. J. Thompson, 'Southern Blotting of Genomic DNA for DNA Profiling', in *Forensic DNA Profiling Protocols*, eds. P.J. Lincoln, J. Thomson, Humana Press, Totowa, NJ, 49–56, 1998.
4. A.J. Jeffries, V. Wilson, S.L. Thein, 'Individual-specific ''Fingerprints'' of Human DNA', *Nature*, **316**, 76–79 (1985).
5. C.M. Strom, L. Rechitsky, 'Use of Nested PCR to Identify Charred Remains and Minute Amounts of Blood', *J. Forensic Sci.*, **43**(3), 696–700 (1998).
6. D. Sweet, D. Hildebrand, 'Recovery of DNA from Human Teeth by Cryogenic Grinding', *J. Forensic Sci.*, **43**(6), 1199–1202 (1998).
7. M.N. Hochmeister, 'PCR Analysis of DNA from Fresh and Decomposed Bodies and Skeletal Remains in Medicolegal Death Investigations', *Methods Mol. Biol.*, **98**, 19–26 (1998).
8. D.R. Linfert, A.H. Wu, G.J. Tsongalis, 'The Effect of Pathologic Substances and Adulterants on the DNA Typing of Urine', *J. Forensic Sci.*, **43**(5), 1041–1045 (1998).
9. M. Lorente, C. Entrala, J.A. Lorente, J.C. Alvarez, E. Villanueva, B. Budowle, 'Dandruff as a Potential Source of DNA in Forensic Casework', *J. Forensic Sci.*, **43**(4), 901–902 (1998).
10. B. Herber, K. Herold, 'DNA Typing of Human Dandruff', *J. Forensic Sci.*, **43**(3), 648–656 (1998).
11. R.L. Romero, A.C. Juston, J. Ballantyne, B.E. Henry, 'The Applicability of Formalin-fixed and Formalin-fixed Paraffin-embedded Tissues in Forensic DNA Analysis', *J. Forensic Sci.*, **42**(4), 708–714 (1997).
12. L. Potsch, U. Meyer, S. Rothschild, P.M. Schneider, C. Rittner, 'Application of DNA Techniques for Idenfication using Human Dental Pulp as a Source of DNA', *Int. J. Legal Med.*, **105**(3), 139–143 (1992).
13. F. Fridez, R. Coquoz, 'PCR DNA Typing of Stamps: Evaluation of the DNA Extraction', *Forensic Sci. Int.*, **78**(2), 103–110 (1996).
14. S.A. Del Rio, M.A. Marino, P. Belgrader, 'PCR-based Human Leukocyte Antigen (HLA) DQ Alpha Typing of Blood Stained Light and Dark Blue Denim Fabric', *J. Forensic Sci.*, **41**(3), 490–492 (1996).
15. S. Yazawa, H. Oh-kawara, T. Nakajima, O. Hosomi, S. Akamatsu, K. Kishi, 'Histo-blood Group Lewis Genotyping from Human Hairs and Blood', *Jpn. J. Hum. Genet.*, **41**(1), 177–188 (1996).
16. L. Roewer, J.T. Epplen, 'Rapid and Sensitive Typing of Forensic Stains by PCR Amplification of Polymorphic Simple Repeat Sequences in Case Work', *Forensic Sci. Int.*, **53**(2), 163–171 (1992).
17. M.N. Hochmeister, 'A Discussion of ''PCR DNA Typing of Stamps: Evaluation of the DNA Extraction'' [letter; comment]', *Forensic Sci. Int.*, **83**(1), 75–79 (1996).
18. C. Stein, S.H. Kyeck, C. Henssge, 'DNA Typing of Fingerprint Reagent Treated Biological Stains', *J. Forensic Sci.*, **41**(6), 1012 (1996).
19. G.J. Tsongalis, D.E. Anamani, A.H. Wu, 'Identification of Urine Specimen Donors by the PM + DQA1 Amplification and Typing Kit', *J. Forensic Sci.*, **41**(6), 1031–1034 (1996).
20. D. Sweet, J.A. Lorente, A. Valenzuela, M. Lorente, E. Villanueva, 'PCR-based DNA Typing of Saliva Stains Recovered from Human Skin', *J. Forensic Sci.*, **42**(3), 447–451 (1997).
21. N. Dimo-Simonin, F. Grange, C. Brandt-Casadevall, 'PCR-based Forensic Testing of DNA from Stained Cytological Smears', *J. Forensic Sci.*, **42**(3), 506–509 (1997).
22. R.A. van Oorschot, M.K. Jones, 'DNA Fingerprints from Fingerprints [letter; comment]', *Nature*, **387**, 767 (1997).
23. G.J. Tsongalis, M.M. Berman, 'Application of Forensic Identity Testing in a Clinical Setting Specimen Identification', *Diagn. Mol. Pathol.*, **6**(2), 111–114 (1997).
24. B. Karger, E. Meyer, A. DuChesne, 'STR Analysis on Perforating FMJ Bullets and a New VWA Variant Allele', *Int. J. Legal Med.*, **110**(2), 101–103 (1997).
25. C.A. Linch, S.L. Smith, J.A. Prahlow, 'Evaluation of the Human Hair Root for DNA Typing Subsequent

to Microscopic Comparison', *J. Forensic Sci.*, **43**(2), 305–314 (1998).
26. D.M. Skinker, S.C. Warnecke, J.K. Morrow, J.C. Hunsaker, '3rd DNA Typing of Azoospermic Semen at the D1S80 Locus', *J. Forensic Sci.*, **42**(4), 718–720 (1997).
27. I. Findlay, P. Quirke, 'Fluorescent Polymerase Chain Reaction: Part I. A New Method Allowing Genetic Diagnosis and DNA Fingerprinting of Single Cells', *Hum. Reprod. Update*, **2**(2), 137–152 (1996).
28. W.C. Burns, Y.S. Liu, C. Dow, R.J. Thomas, W.A. Phillips, 'Direct PCR from Paraffin-embedded Tissue', *Biotechniques*, **22**(4), 638–640 (1997).
29. M.P. Evison, D.M. Smillie, A.T. Chamberlain, 'Extraction of Single-copy Nuclear DNA from Forensic Specimens with a Variety of Postmortem Histories', *J. Forensic Sci.*, **42**(6), 1032–1038 (1997).
30. C. Cattaneo, O.E. Craig, N.T. James, R.J. Sokol, 'Comparison of Three DNA Extraction Methods on Bone and Blood Stains up to 43 Years Old and Amplification of Three Different Gene Sequences', *J. Forensic Sci.*, **42**(6), 1126–1135 (1997).
31. U. Reischl, B. Haber, W. Vandezande, T. De Baere, M. Vaneechoutte, 'Long-term Storage of Unamplified Complete PCR Mixtures', *Biotechniques*, **23**(4), 580–582, 584 (1997).
32. M. Takahashi, Y. Kato, H. Mukoyama, H. Kanaya, S. Kamiyama, 'Evaluation of Five Polymorphic Microsatellite Markers for Typing DNA from Decomposed Human Tissue – Correlation Between the Size of the Alleles and that of the Template DNA', *Forensic Sci. Int.*, **90**(1/2), 1–9 (1997).
33. M. Sasaki, H. Shiono, T. Fukushima, K. Shimizu, 'Human Identification by Genotyping of Personal Articles', *Forensic Sci. Int.*, **90**(1/2), 65–75 (1997).
34. R. Palmirotta, F. Verginelli, A. Cama, R. Mariani-Costanni, L. Fra, P. Batsta, 'Origin and Gender Determination of Dried Blood on a Statue of the Virgin Mary', *J. Forensic Sci.*, **43**(2), 431–434 (1998).
35. M. Allen, A.S. Engstrom, S. Meyers, O. Handt, T. Saldeen, A. von Haeseler, S. Paabo, U. Gyllensten, 'Mitochondrial DNA Sequencing of Shed Hairs and Saliva on Robbery Caps: Sensitivity and Matching Probabilities', *J. Forensic Sci.*, **43**(3), 453–464 (1998).
36. C. Schyma, W. Huckenbeck, W. Bonte, 'DNA-PCR Analysis of Bloodstains Sampled by the Polyvinylalcohol Method', *J. Forensic Sci.*, **44**(1), 95–99 (1999).
37. J.M. Willard, D.A. Lee, M.M. Holland, 'Recovery of DNA for PCR Amplification from Blood and Forensic Samples using a Chelating Resin', *Methods Mol. Biol.*, **98**, 9–18 (1998).
38. M.N. Hochmeister, O. Rudin, E. Ambach, 'PCR Analysis from Cigaret Butts, Postage Stamps, Envelope Sealing Flaps, and Other Saliva-stained Material', *Methods Mol. Biol.*, **98**, 27–32 (1998).
39. L. Rudbeck, J. Dissing, 'Rapid, Simple Alkaline Extraction of Human Genomic DNA from Whole Blood, Buccal Epithelial Cells, Semen and Forensic Stains for PCR', *Biotechniques*, **25**(4), 588–590, 592 (1998).
40. N. Dimo-Simonin, F. Grange, C. Brandt-Casadevall, 'PCR-based Forensic Testing of DNA from Stained Cytological Smears', *J. Forensic Sci.*, **42**(3), 506–509 (1997).
41. N. Morling, 'Amplification of Short Tandem Repeat Loci Using PCR', in *Forensic DNA Profiling Protocols*, eds. P.J. Lincoln, J. Thomson, Humana Press, Totowa, NJ, 173–180, 1998.
42. G. Caetano-Anolles, 'DAF Optimization using Taguchi Methods and the Effect of Thermal Cycling Parameters on DNA Amplification', *Biotechniques*, **25**(3), 472–476, 478–480 (1998).
43. J.M. Robertson, J. Walsh-Weller, 'An Introduction to PCR Primer Design and Optimization of Amplification', *Methods Mol. Biol.*, **98**, 121–154 (1998).
44. C.P. Phillips, A. Carracedo, M.V. Lareu, 'Manual Electrophoretic Methods for Genotyping Amplified STR Loci', *Methods Mol. Biol.*, **98**, 181–192 (1998).
45. J.J. Lin, J. Ma, J. Kuo, 'Chemiluminescent Detection of AFLP Markers', *Biotechniques*, **26**(2), 344–348 (1999).
46. R.R. Frazier, E.S. Millican, S.K. Watson, N.J. Oldroyd, R.L. Sparkles, K.M. Taylor, S. Panchal, L. Bark, C.P. Kimpton, P.D. Gill, 'Validation of the Applied Biosystems Prism 377 Automated Sequencer for the Forensic Short Tandem Repeat Analysis', *Electrophoresis*, **17**(10), 1550–1552 (1996).
47. S. Chiba, S. Oshida, K. Okano, H. Kambara, M. Esumi, 'Paternity Testing by the Detection of D1S80 VNTR Using a Fluorescence Image Analyzer (Dualcolor System)', *Forensic Sci. Int.*, **83**(2), 87–94 (1996).
48. A.C. Syvanen, 'Solid-phase Minisequencing as a Tool to Detect DNA Polymorphism', *Methods Mol. Biol.*, **98**, 291–298 (1998).
49. J.M. Butler, 'The Use of Capillary Electrophoresis in Genotyping STR Loci', *Methods Mol. Biol.*, **98**, 279–289 (1998).
50. D. Watts, 'Genotyping STR Loci using an Automated DNA Sequencer', *Methods Mol. Biol.*, **98**, 193–208 (1998).
51. F. Rousselet, H. Pfitzinger, P. Mangin, 'A Pentaplex Automated Fluorescent Typing System for Forensic Identification and French Caucasian Population Data', *J. Forensic Sci.*, **42**(3), 500–503 (1997).
52. R. Roy, D.L. Steffens, B. Gartside, G.Y. Jang, J.A. Brumbaugh, 'Producing STR Locus Patterns from Bloodstains and Other Forensic Samples using an Infrared Fluorescent Automated DNA Sequencer', *J. Forensic Sci.*, **41**(3), 418–424 (1996).
53. D.L. Steffens, R. Roy, J.A. Brumbaugh, 'Multiplex Amplification of STR Loci with Gender Alleles using Infrared Fluorescence Detection', *Forensic Sci. Int.*, **85**(3), 225–232 (1997).

54. B. Brinkmann, 'Overview of PCR-based Systems in Identity Testing', *Methods Mol. Biol.*, **98**, 105–119 (1998).
55. Z. Tun, K. Honda, M. Nakatome, M.N. Islam, H. Bai, Y. Ogura, H. Kuroki, M. Yamazaki, M. Terada, C. Wakasugi, 'Rapid and Clear Detection of ABO Genotypes by Simultaneous PCR–RFLP Method', *J. Forensic Sci.*, **41**(6), 1027–1030 (1996).
56. B. Budowle, R.C. Allen, 'Analysis of Amplified Fragment-length Polymorphisms (VNTR/STR Loci) for Human Identity', *Methods Mol. Biol.*, **98**, 155–171 (1998).
57. N. Morling, 'Amplification of Short Tandem Repeat Loci using PCR', *Methods Mol. Biol.*, **98**, 173–180 (1998).
58. S.J. Park, W.G. Lee, S.W. Lee, S.H. Kim, B.S. Koo, B. Budowle, H.M. Rho, 'Genetic Variations at Four Tetrameric Tandem Repeat Loci in Korean Population', *J. Forensic Sci.*, **42**(1), 125–129 (1997).
59. S. Rand, C. Puers, K. Skowasch, P. Wiegand, B. Budowle, B. Brinkmann, 'Population Genetics and Forensic Efficiency Data of 4 AMPFLP's', *Int. J. Legal Med.*, **104**(6), 329–333 (1992).
60. A. Sajanla, B. Budowle, M. Strom, V. Johansson, M. Lukka, L. Peltonen, C. Ehnholm, 'PCR Amplification of Alleles at the DIS80 Locus: Comparison of a Finnish and a North American Caucasian Population Sample, and Forensic Casework Evaluation', *Am. J. Hum. Genet.*, **50**(4), 816–825 (1992).
61. A. Sajanla, S. Puomilah, V. Johnsson, C. Ehnholm, 'Amplification of Reproducible Allele Markers for Amplified Fragment Length Polymorphism Analysis', *Biotechniques*, **12**(1), 16, 18, 20–22 (1992).
62. F. Pinheiro, L. Pontes, M. Gene, E. Huguet, J.P. da Costa, P. Moreno, 'Population Study of the HUMTH01, HUMVWA31A, HUMF13A1, and HUMFES/FPS STR Polymorphisms in the North of Portugal', *J. Forensic Sci.*, **42**(1), 121–124 (1997).
63. M. Lucas, C. Munoz, E. Pintado, F. Solano, 'Highly Informative Single-stranded Conformation Polymorphism (SSCP) of Short Tandem Repeats in DNA Identification', *J. Forensic Sci.*, **42**(1), 118–120 (1997).
64. I. Zupanic, J. Balazic, R. Komel, 'Analysis of Nine Short Tandem Repeat (STR) Loci in the Slovenian Population', *Int. J. Legal Med.*, **111**(5), 248–250 (1998).
65. L. Garofano, M. Pizzamiglio, C. Vecchio, G. Lago, T. Floris, G. D'Errico, G. Brembilla, A. Romano, B. Budowle, 'Italian Population Data on Thirteen Short Tandem Repeat Loci: HUMTH01, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358', *Forensic Sci. Int.*, **97**(1), 53–60 (1998).
66. B. Budowle, F.S. Baechtel, R. Fejeran, 'Polymarker, HLA-DQA1, and D1S80 Allele Frequency Data in Chamorro and Filipino Populations from Guam', *J. Forensic Sci.*, **43**(6), 1195–1198 (1998).
67. M.M. de Pancorbo, A. Castro, I. Fernandez-Fernandez, A. Garcia-Orad, 'Population Genetics and Forensic Applications using Multiplex PCR (CSF1PO, TPOX, and TH01) Loci in the Basque Country', *J. Forensic Sci.*, **43**(6), 1181–1187 (1998).
68. R.A. van Oorschot, S.J. Gutowski, S.L. Robinson, J.A. Hedley, I.R. Andrew, 'HUMTH01 Validation Studies: Effect of Substrate, Environment, and Mixtures', *J. Forensic Sci.*, **41**(1), 142–145 (1996).
69. K.M. Keys, B. Budowle, S. Andelinovic, M. Definis-Gojanovic, I. Drmic, M. Mladen, D. Primorac, 'Northern and Southern Croatian Population Data on Seven PCR-based Loci', *Forensic Sci. Int.*, **81**(2/3), 191–199 (1996).
70. M. Ciesielka, P. Koziol, A. Krajka, 'Allele Frequency Distributions of D1S80 in the Polish Population', *Forensic Sci. Int.*, **81**(2/3), 141–147 (1996).
71. K. Tamaki, X.L. Huang, H. Nozawa, T. Yamamoto, R. Uchihi, Y. Katsumata, J.A. Armour, 'Evaluation of Tetranucleotide Repeat Locus D7S809 (wg1g9) in the Japanese Population', *Forensic Sci. Int.*, **81**(2/3), 133–140 (1996).
72. B. Budowle, K.L. Monson, R. Chakraborty, 'Estimating Minimum Allele Frequencies for DNA Profile Frequency Estimates for PCR-based Loci', *Int. J. Legal Med.*, **108**(4), 173–176 (1996).
73. K.E. al-Nassar, J. Mathew, N. Thomas, H.R. Fatania, 'Analysis of the D1S80 (pMCT118) VNTR Locus Polymorphism in a Native Kuwai Population by the Polymerase Chain Reaction', *Forensic Sci. Int.*, **78**(2), 131–138 (1996).
74. M. Klintschar, R. Crevenna, 'Validation of the STR System FXIIIB for Forensic Purposes in an Austrian Population Sample', *Forensic Sci. Int.*, **81**(1), 35–42 (1996).
75. V.W. Weedn, 'Forensic DNA Tests', *Clin. Lab. Med.*, **16**(1), 187–196 (1996).
76. W.M. Yun, S.G. Yun, 'Analysis of the VNTR Locus DXS52 by the Amp-FLP Technique', *J. Forensic Sci.*, **41**(5), 859–861 (1996).
77. K. Yoshida, K. Sekiguchi, K. Kasai, H. Sato, S. Seta, G.F. Sensabaugh, 'Evaluation of New Primers for CSF1PO', *Int. J. Legal Med.*, **110**(1), 36–38 (1997).
78. H.G. Zhou, K. Sato, Y. Nishimaki, L. Fang, H. Hasekura, 'The HumD21S11 System of Short Tandem Repeat DNA Polymorphisms in Japanese and Chinese', *Forensic Sci. Int.*, **86**(1/2), 109–118 (1997).
79. A.M. Lins, K.A. Micka, C.J. Sprecher, J.A. Taylor, J.W. Bacher, D.R. Rabbach, R.A. Bever, S.D. Creacy, J.W. Schumm, 'Development and Population Study of an Eight-locus Short Tandem Repeat (STR) Multiplex System', *J. Forensic Sci.*, **43**(6), 1168–1180 (1998).
80. B. Budowle, T.R. Moret, K.M. Keys, B.W. Koons, J.B. Smerick, 'Validation Studies of the CTT STR Multiplex System', *J. Forensic Sci.*, **42**(4), 701–707 (1997).

81. F. Neuhuber, M. Radacher, 'A Genetic Study of the Short Tandem Repeat Systems VWA and TH01 in an Austrian Population', *Forensic Sci. Int.*, **87**(3), 211–217 (1997).
82. F. De Stefano, L. Casarino, M.G. Costa, G. Bruni, A. Mannucci, M. Unseld, R. Hiesel, M. Canale, 'Analysis of a Short Tandem Repeat Locus on Chromosome 19 (D19S253)', *Int. J. Legal Med.*, **108**(5), 256–258 (1996).
83. L. Busque, D. Desmarais, S. Provost, J.W. Schumm, Y. Zhong, R. Chakraborty, 'Analysis of Allele Distribution for Six Short Tandem Repeat Loci in the French Canadian Population of Quebec', *J. Forensic Sci.*, **42**(6), 1147–1153 (1997).
84. A.M. Gross, G. Carmody, R.A. Guerrieri, 'Validation Studies for the Genetic Typing of the D1S80 Locus for Implementation into Forensic Casework', *J. Forensic Sci.*, **42**(6), 1140–1146 (1997).
85. S. Furedi, B. Budowle, J. Woller, Z. Padar, 'Hungarian Population Data on Six STR Loci – HUMVWFA31, HUMTH01, HUMCSF1PO, HUMFES/FPS, HUMTPOX, and HUMHPRTB – Derived using Multiplex PCR Amplification and Manual Typing', *Int. J. Legal Med.*, **109**(2), 100–101 (1996).
86. F. Neuhuber, R. Lamprecht, M. Radacher, 'Population Genetic Study of the AmpFLP System APO B in an Austrian Population Sample', *Int. J. Legal Med.*, **109**(1), 45–46 (1996).
87. J. Andersen, P. Marn, A. Carracedo, M. Dobosz, B. Eriksen, V. Johnsson, C. Kimpton, A. Kloosterman, C. Konialis, A. Kratzer, P. Phillips, B. Mevag, H. Pfitzinger, S. Rand, B. Rosen, H. Schmitter, P. Schneider, M. Vide, 'Report on the Third EDNAP Collaborative STR Exercise. European DNA Profiling Group', *Forensic Sci. Int.*, **78**(2), 83–93 (1996).
88. R. Sparkes, C. Kimpton, S. Watson, N. Oldroyd, T. Clayton, L. Barnett, J. Arnold, C. Thompson, R. Hale, J. Chapman, A. Urquhart, P. Gill, 'The Validation of a 7-locus Multiplex STR Test for Use in Forensic Casework. (I). Mixtures, Ageing, Degradation and Species Studies', *Int. J. Legal Med.*, **109**(4), 186–194 (1996).
89. R. Andreassen, B. Olaisen, 'De Novo Mutations and Allelic Diversity at Minisatellite Locus D7S22 Investigated by Allele-specific Four-state MVR-PCR Analysis', *Hum. Mol. Genet.*, **7**(13), 2113–2120 (1998).
90. M. Sasaki, H. Shiono, 'ABO Genotyping of Suspects from Sperm DNA Isolated from Postcoital Samples in Sex Crimes', *J. Forensic Sci.*, **41**(2), 275–278 (1996).
91. J.F. Andersen, M.J. Greenhalgh, H.R. Butler, S.R. Kilpatrick, R.C. Piercy, K.A. Way, H.S. Myhill, J.C. Wright, R. Hallet, B.H. Parkin, 'Further Validation of a Multiplex STR System for Use in Routine Forensic Identity Testing', *Forensic Sci. Int.*, **78**(1), 47–64 (1996).
92. Y. Liu, Y. Koda, M. Soejima, N. Uchida, H. Kimura, 'PCR Analysis of Lewis-negative Gene Mutations and the Distribution of Lewis Alleles in a Japanese Population', *J. Forensic Sci.*, **41**(6), 1018–1021 (1996).
93. C.H. Tzeng, J.Y. Lyou, Y.R. Chen, H.Y. Hu, J.S. Lin, C.H. Yung, K.B. Choo, S.Y. Wang, J.C. Lee, 'Polymorphisms of Twelve Short Tandem Repeat Loci in a Taiwanese Population and their Application in Parentage Testing', *J. Formos. Med. Assoc.*, **97**(11), 738–744 (1998).
94. B. Budowle, L.T. Nhari, T.R. Moret, S.B. Kanoyangwa, E. Masuka, D.A. Defenburgh, J.B. Smerick, 'Zimbabwe Black Population Data on the Six Short Tandem Repeat Loci – CSF1PO, TPOX, THO1, D3S1358, VWA and FGA', *Forensic Sci. Int.*, **90**(3), 215–221 (1997).
95. P. Brner, M. Shpitzen, M. Freund, N. Manny, 'The Effects of Blood Transfusions on PCR DNA Typing at the CSF1P0, TP0X, TH01, D1S80, HLA-DQA1, LDLR, GYPA, HBGG, D7S8 and GC Loci', *J. Forensic Sci.*, **42**(6), 1154–1156 (1997).
96. S. Liech-Galla, D. Neeser, 'Efficient and Reliable PCR-based Detection of the ABO Blood Group Alleles: Genotyping on Stamps and Other Biological Evidence Samples', *J. Forensic Sci.*, **41**(4), 653–657 (1996).
97. M. Crespillo, J.A. Luque, R. Fernandez, E. Ramirez, P. Garcia, J.L. Valverde, 'Allele Frequency Distributions of 13 PCR-based Systems in a Population from North-east Spain', *Int. J. Legal Med.*, **110**(4), 223–225 (1997).
98. G. Mertens, N. Mommers, H. Heylen, M. Gielis, L. Muylle, A. Vandenberghe, 'Allele Frequencies of Nine STR Systems in the Flemish Population and Application in Parentage Testing', *Int. J. Legal Med.*, **110**(4), 177–180 (1997).
99. M.M. de Pancorbo, J. Rodriguez-Alarcon, A. Castro, I. Fernandez-Fernandez, J.C. Melchor, A. Linares, A. Garcia-Orad, L. Fernandez-Llebrez del Rey, G. Aranguren, L. Sanllana, 'Newborn Genetic Identification: A Protocol using Microsatellite DNA as an Alternative to Foot Printing', *Clin. Chim. Acta*, **263**(1), 33–42 (1997).
100. B. Glock, E.M. Dber, D.W. Schwartz, W.R. Mayr, 'Additional Variability at the D12S391 STR Locus in an Austrian Population Sample: Sequencing Data and Allele Distribution', *Forensic Sci. Int.*, **90**(3), 197–203 (1997).
101. J.C. Lee, J.G. Chang, 'ABO Genotyping by Polymerase Chain Reaction', *J. Forensic Sci.*, **37**(5), 1269–1275 (1992).
102. H. Jorquera, B. Budowle, 'Chilean Population Data on Ten PCR-based Loci', *J. Forensic Sci.*, **43**(1), 171–173 (1998).
103. M.L. Baird, 'Use of DNA Identification for Forensic and Paternity Analysis', *J. Clin. Lab. Anal.*, **10**(6), 350–358 (1996).
104. R.A. Bayoumi, L.I. al-Gazali, U. Jaffer, M.S. Nur-E-Kamal, A. Dawodu, A. Bener, V. Eapen, B. Budowle, 'United Arab Emirates Population Data on Three DNA

Tetrameric Short Tandem Repeat Loci – HUMTHO1, TPOX and CSF1PO – Derived Using Multiplex Polymerase Chain Reaction and Manual Typing', *Electrophoresis*, **18**(9), 1637–1640 (1997).
105. M. Klintschar, N. al-Hammadi, T. Lux, B. Reichenpfader, 'Genetic Variation at the Short Tandem Repeat Loci HumvWA, HumFXIIIB, and HumFES/FPS in the Egyptian and Yemenian Populations', *J. Forensic Sci.*, **43**(4), 850–853 (1998).
106. A. Tagliabracci, L. Buscemi, F. Bianchi, C. Sassaroli, U. Ricci, T.M. Neri, D. Rodriguez, 'Polymorphism and Sequence Variations of the HumCD4 Pentameric Microsatellite in an Italian Population Sample', *J. Forensic Sci.*, **43**(4), 841–844 (1998).
107. L. Garofano, G. Lago, C. Vecchio, M. Pizzamiglio, C. Zanon, A. Virgili, L. Albonici, V. Manzari, B. Budowle, 'Italian Population Data on the Polymarker System and on the Five Short Tandem Repeat Loci CSF1PO, TPOX, TH01, F13B, and vWA', *J. Forensic Sci.*, **43**(4), 837–840 (1998).
108. S. Ameno, K. Ameno, H. Kinoshita, N. Tanaka, I. Ijiri, 'Lewis Genotyping by the PCR–RFLP Method in a Japanese Population and its Evaluation in Forensic Analysis', *Int. J. Legal Med.*, **110**(4), 232–234 (1997).
109. Y. Hanaoka, K. Minaguch, 'D4S43 Locus DNA Typing in the Japanese Population and Application to Teeth with Degraded DNA', *J. Forensic Sci.*, **43**(2), 406–409 (1998).
110. M. Gene, M. Fuentes, E. Huguet, E. Pique, F. Bert, A. Corella, A. Perez-Perez, J. Corbella, P. Moreno, 'Quechua Amerindian Population Characterized by HLA-DQ Alpha, YNZ22, 3′APO B, HUMTH01, and HUMVWA31A Polymorphisms', *J. Forensic Sci.*, **43**(2), 403–405 (1998).
111. E. Arroyo, F. Garcia-Sanchez, L. Prieto, J.M. Ruiz de la Cuesta, J.L. Vicario, 'Polymorphism Analysis of the VNTR Locus D17S5 in Central Spain', *Int. J. Legal Med.*, **109**(2), 98–99 (1996).
112. H. Nishimukai, Y. Fukumori, T. Okiura, I. Yuasa, T. Shinomiya, S. Ohnoki, H. Shibata, U. Vogt, 'Genotyping of the ABO Blood Group System: Analysis of Nucleotide Position 802 by PCR–RFLP and the Distribution of ABO Genotypes in a German Population', *Int. J. Legal Med.*, **109**(2), 90–93 (1996).
113. M.D. Barber, B.H. Parkin, 'Sequence Analysis and Allelic Designation of the Two Short Tandem Repeat Loci D18S51 and D8S1179', *Int. J. Legal Med.*, **109**(2), 62–65 (1996).
114. G. Watanabe, K. Umetsu, T. Suzuki, 'Determination of the HUMTH01 Alleles by the APLP Method', *Int. J. Legal Med.*, **112**(2), 134–135 (1999).
115. M.V. Lareu, M.C. Pestoni, F. Barros, A. Salas, A. Carracedo, 'Sequence Variation of a Hypervariable Short Tandem Repeat at the D12S391 Locus', *Gene*, **182**(1/2), 151–153 (1996).
116. I. Rostedt, K. Lalu, M. Lukka, A. Sajanla, 'Genotyping of Five Short Tandem Repeat Loci via Triplex and Duplex PCR', *Forensic Sci. Int.*, **82**(3), 217–226 (1996).
117. Z. Lin, T. Kondo, T. Minamino, E. Sun, G. Liu, T. Ohshima, 'Genotyping of ABO Blood Group System by PCR and RFLP on Mummies Discovered at Taklamakan Desert in 1912', *Nippon Hoigaku Zasshi*, **50**(5), 336–342 (1996).
118. P. Gill, E. d'Aloja, B. Dupuy, B. Eriksen, M. Jangblad, V. Johnsson, A.D. Kloosterman, A. Kratzer, M.V. Lareu, B. Mevag, N. Morling, C. Phillips, H. Pfitzinger, S. Rand, M. Sabaer, R. Scheither, H. Schmitter, P. Schneider, I. Skitsa, M.C. Vide, 'Report of the European DNA Profiling Group (EDNAP) – an Investigation of the Hypervariable STR Loci ACTBP2, APOAI1 and D11S554 and the Compound Loci D12S391 and D1S1656', *Forensic Sci. Int.*, **98**(3), 193–200 (1998).
119. L. Garofano, M. Pizzamiglio, C. Vecchio, G. Lago, T. Floris, G. D'Errico, G. Brembilla, A. Romano, B. Budowle, 'Italian Population Data on Thirteen Short Tandem Repeat Loci: HUMTH01, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358', *Forensic Sci. Int.*, **97**(1), 53–60 (1998).
120. I.M. Sebetan, H.A. Hajar, 'Analysis of the Short Tandem Repeat (STR) Locus HumVWA in a Qatari Population', *Forensic Sci. Int.*, **95**(2), 169–171 (1998).
121. F.X. Xiao, A. Gilissen, J.J. Cassiman, R. Decorte, 'Quadruplex Fluorescent STR Typing System (HUMVWA, HUMTH01, D21S11 and HPRT) with Sequence-defined Allelic Ladders Identification of a New Allele at D21S11', *Forensic Sci. Int.*, **94**(1/2), 39–46 (1998).
122. R. Szibor, S. Ltsch, I. Plate, K. Bender, D. Krse, 'Population Genetic Data of the STR HumD3S1358 in Two Regions of Germany', *Int. J. Legal Med.*, **111**(3), 160–161 (1998).
123. R. Poltl, C. Luckenbach, H. Ritter, 'The Short Tandem Repeat Locus D3S1359', *Forensic Sci. Int.*, **95**(2), 163–168 (1998).
124. M. Klintschar, R. Crevenna, 'HumCD4 – Validation of a STR System for Forensic Purposes in an Austrian Caucasian Population Sample', *J. Forensic Sci.*, **42**(5), 907–910 (1997).
125. E. Ambach, W. Parson, H. Niederstatter, B. Budowle, 'Austrian Caucasian Population Data for the Quadruplex Plus Amelogenin: Refined Mutation Rate for HumvWFA31/A', *J. Forensic Sci.*, **42**(6), 1136–1139 (1997).
126. W. Wang, K. Okazaki, T. Kishida, M. Fukuda, Y. Tamaki, 'Subtyping of D20S85 STR Alleles by Single-strand Conformation Polymorphism (SSCP) Analysis', *Forensic Sci. Int.*, **86**(3), 187–192 (1997).
127. M. Lorente, J.A. Lorente, M.R. Wilson, B. Budowle, E. Villanueva, 'Spanish Population Data on Seven Loci: D1S80, D17S5, HUMTH01, HUMVWA, ACTBP2,

D21S11 and HLA-DQA1', *Forensic Sci. Int.*, **86**(3), 163–171 (1997).
128. M.A. Cariolou, P. Manoli, M. Christophorou, E. Bashiardes, A. Karagrigoriou, B. Budowle, 'Greek Cypriot Allele and Genotype Frequencies for Amplitype PM-DQA1 and D1S80 Loci', *J. Forensic Sci.*, **43**(3), 661–664 (1998).
129. B. Marnez-Jarreta, B. Budowle, E. Abecia, B. Bell, Y. Casalod, M. Castellano, 'PM and D1S80 Loci Gene Frequencies in the Zaragoza Population of Northern Spain', *J. Forensic Sci.*, **43**(5), 1094–1096 (1998).
130. M. Klintschar, Z. Kozma, N. al Hammadi, M. Abdull Fatah, C. Nohammer, 'A Study on the Short Tandem Repeat Systems HumCD4, HumTH01 and HumFIBRA in Population Samples from Yemen and Egypt', *Int. J. Legal Med.*, **111**(2), 107–109 (1998).
131. A. Bjerre, D.S. Court, P. Lincoln, N. Morling, 'A Report of the 1995 and 1996 Paternity Testing Workshops of the English Speaking Working Group of the International Society for Forensic Haemogenetics', *Forensic Sci. Int.*, **90**(1/2), 41–55 (1997).
132. B. Budowle, F.S. Baechtel, R. Fejeran, 'Polymarker, HLA-DQA1, and D1S80 Allele Frequency Data in Chamorro and Filipino Populations from Guam', *J. Forensic Sci.*, **43**(6), 1195–1198 (1998).
133. A. Alessio, C. Cldio, 'The Distribution of D1S80 (pMCT118) Alleles in a Southern Italian Population Sample', *J. Forensic Sci.*, **43**(4), 911–912 (1998).
134. M.M. de Pancorbo, A. Castro, I. Fernandez-Fernandez, A. Garcia-Orad, 'Population Genetics and Forensic Applications using Multiplex PCR (CSF1PO, TPOX, and TH01) Loci in the Basque Country', *J. Forensic Sci.*, **43**(6), 1181–1187 (1998).
135. A.M. Lins, K.A. Micka, C.J. Sprecher, J.A. Taylor, J.W. Bacher, D.R. Rabbach, R.A. Bever, S.D. Creacy, J.W. Schumm, 'Development and Population Study of an Eight-locus Short Tandem Repeat (STR) Multiplex System', *J. Forensic Sci.*, **43**(6), 1168–1180 (1998).
136. C. Entrala, M. Lorente, J.A. Lorente, J.C. Alvarez, T. Moret, B. Budowle, E. Villanueva, 'Fluorescent Multiplex Analysis of Nine STR Loci: Spanish Population Data', *Forensic Sci. Int.*, **98**(3), 179–183 (1998).
137. E. Rossi, B. Rolf, M. Schurenkamp, B. Brinkmann, 'Y-Chromosome STR Haplotypes in an Italian Population Sample', *Int. J. Legal Med.*, **112**(1), 78–81 (1999).
138. A. Caglia, M. Dobosz, I. Boschi, E. d'Aloja, V.L. Pascali, 'Increased Forensic Efficiency of a STR-based Y-specific Haplotype by Addition of the Highly Polymorphic DYS385 Locus', *Int. J. Legal Med.*, **111**(3), 142–146 (1998).
139. R. Zehner, U. Bohrer, 'DYS19 and Amelogenin in Artificial Blood Stains with Defined Amounts of Male and Female Cells', *Int. J. Legal Med.*, **111**(6), 340–342 (1998).
140. A. Akane, 'Sex Determination by PCR Analysis of the X–Y Amelogenin Gene', *Methods Mol. Biol.*, **98**, 245–249 (1998).
141. B. Stacks, M.M. Witte, 'Sex Determination of Dried Blood Stains using the Polymerase Chain Reaction (PCR) with Homologous X–Y Primers of the Zinc Finger Protein Gene', *J. Forensic Sci.*, **41**(2), 287–290 (1996).
142. R. Reynolds, J. Varlaro, 'Gender Determination of Forensic Samples using PCR Amplification of ZFX/ZFY Gene Sequences', *J. Forensic Sci.*, **41**(2), 279–286 (1996).
143. Y. Hanaoka, K. Minaguchi, 'Sex Determination from Blood and Teeth by PCR Amplification of the Alphoid Satellite Family', *J. Forensic Sci.*, **41**(5), 855–858 (1996).
144. J.L. Finch, R.M. Hope, A. van Daal, 'Human Sex Determination using Multiplex Polymerase Chain Reaction (PCR)', *Sci. Justice*, **36**(2), 93–95 (1996).
145. A. Akane, S. Seki, H. Shiono, H. Nakamura, M. Hasegawa, M. Kagawa, K. Matsubara, Y. Nakahori, S. Nagafuchi, Y. Nakagome, 'Sex Determination of Forensic Samples by Dual PCR Amplification of an X–Y Homologous Gene', *Forensic Sci. Int.*, **52**(2), 143–148 (1992).
146. S. Inoue, Y. Yamamoto, O. Okamoto, H. Murakami, S. Miyaishi, H. Ishizu, 'Improvement of Sensitivity in HLA-DRB1 Typing by Semi-nested PCR–RFLP', *Acta Med. Okayama*, **52**(6), 289–296 (1998).
147. L.B. Jankowski, B. Budowle, N.T. Swec, J.A. Pino, S. Freck-Tootell, H.W. Corey, R. Schwartz, E.J. LaRue, W.L. Rochin, C.J. Kearney, M.L. Tarver, 'New Jersey Caucasian, African American, and Hispanic Population Data on the PCR-based Loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc', *J. Forensic Sci.*, **43**(5), 1037–1040 (1998).
148. M. Ota, Y. Katsuyama, C.Y. Liu, A. Arakura, H. Fukushima, 'Validation of HLA-DR Locus Typing in Forensic Specimens by Combining PCR–SSP with PCR–RFLP', *J. Forensic Sci.*, **42**(5), 929–934 (1997).
149. I.M. Sebetan, H.A. Hajar, 'HLA DQ Alpha Genotype and Allele Frequencies in Qatari Population', *Forensic Sci. Int.*, **90**(1/2), 11–15 (1997).
150. B. Budowle, B.W. Koons, T.R. Moret, 'Subtyping of the HLA-DQA1 Locus and Independence Testing with PM and STR/VNTR Loci' [published erratum appears in *J. Forensic Sci.*, **43**(5), 1105 (1998).], *J. Forensic Sci.*, **43**(3), 657–660 (1998).
151. Y. Watanabe, S. Yamada, A. Nagai, T. Takayama, K. Hirata, Y. Bunai, I. Ohya, 'Japanese Population DNA Typing Data for the Loci LDLR, GYPA, HBGG, D7S8, and GC', *J. Forensic Sci.*, **42**(5), 911–913 (1997).
152. J.M. Wallin, M.R. Buoncrisani, K.D. Lazaruk, N. Fildes, C.L. Holt, P.S. Walsh, 'TWGDAM Validation of the AmpFISTR Blue PCR Amplification Kit for Forensic Casework Analysis', *J. Forensic Sci.*, **43**(4), 854–870 (1998).
153. R. Susukida, A. Kido, M. Oya, 'GPT Genotyping by Polymerase Chain Reaction and Restriction Fragment

Length Polymorphism Analysis', *Forensic Sci. Int.*, **98**(3), 185–191 (1998).
154. M.S. Taylor, A. Challed-Spong, E.A. Johnson, 'Co-amplification of the Amelogenin and HLA DQ Alpha Genes: Optimization and Validation', *J. Forensic Sci.*, **42**(1), 130–136 (1997).
155. R. Zehner, D. Mebs, H. Bratzke, 'Population Genetic Study of the Simultaneously Amplified Loci HLA DQA1, LDLR, GYPA, HBGG, D7S8, and GC in a German Population Sample', *J. Forensic Sci.*, **43**(4), 913–914 (1998).
156. M.L. Baird, 'Use of the AmpliType PM + HLA DQA1 PCR Amplification and Typing Kits for Identity Testing', *Methods Mol. Biol.*, **98**, 261–277 (1998).
157. E. Blake, J. Mihalovich, R. Higuchi, P.S. Walsh, H. Erlich, 'Polymerase Chain Reaction (PCR) Amplification and Human Leukocyte Antigen (HLA)-DQ Alpha Oligonucleotide Typing on Biological Evidence Samples: Casework Experience', *J. Forensic Sci.*, **37**(3), 700–726 (1992).
158. C.J. Word, T.M. Sawosik, D.H. Bing, 'Summary of Validation Studies from Twenty-six Forensic Laboratories in the United States and Canada on the Use of the AmpliType PM PCR Amplification and Typing Kit', *J. Forensic Sci.*, **42**(1), 39–48 (1997).
159. A.M. Gross, R.A. Guerrieri, 'HLA DQA1 and Polymarker Validations for Forensic Casework: Standard Specimens, Reproducibility, and Mixed Specimens', *J. Forensic Sci.*, **41**(6), 1022–1026 (1996).
160. G. Marnelli, E. Trabet, P. Farabegoli, M. Buzzi, A. Zaccaria, N. Testoni, M. Amabile, A. Casartelli, A. de Vivo, P.F. Pignat, S. Tura, ' ''Fingerprinting'' of HLA-DQA by Polymerase Chain Reaction and Heteroduplex Analysis', *Mol. Cell Probes*, **10**(2), 123–127 (1996).
161. F. De Stefano, L. Casarino, A. Mannucci, L. Delfino, M. Canale, G.B. Ferrara, 'HLA-DQA1 Allele and Genotype Frequencies in a Northern Italian Population', *Forensic Sci. Int.*, **55**(1), 59–66 (1992).
162. M.S. Rodriguez-Calvo, S. Bellas, L. Souto, C. Vide, E. Valverde, A. Carracedo, 'Population Data on the Loci LDLR, GYPA, HBGG, D7S8, and GC in Three Southwest European Populations', *J. Forensic Sci.*, **41**(2), 291–296 (1996).
163. R.J. Steighner, M. Holland, 'Amplification and Sequencing of Mitochondrial DNA in Forensic Casework', *Methods Mol. Biol.*, **98**, 213–223 (1998).
164. M. Bataille, K. Crainic, M. Leterreux, M. Durigon, P. de Mazancourt, 'Multiplex Amplification of Mitochondrial DNA for Human and Species Identification in Forensic Evaluation', *Forensic Sci. Int.*, **99**(3), 165–170 (1999).
165. K.M. Sullivan, R. Hopgood, P. Gill, 'Identification of Human Remains by Amplification and Automated Sequencing of Mitochondrial DNA', *Int. J. Legal Med.*, **105**(2), 83–86 (1992).
166. C. Ginther, L. Issel-Tarver, M.C. King, 'Identifying Individuals by Sequencing Mitochondrial DNA from Teeth', *Nat. Genet.*, **2**(2), 135–138 (1992).
167. A. Baasner, C. Schafer, A. Junge, B. Madea, 'Polymorphic Sites in Human Mitochondrial DNA Control Region Sequences: Population Data and Maternal Inheritance', *Forensic Sci. Int.*, **98**(3), 169–178 (1998).
168. G. Barreto, A.R. Vago, C. Ginther, A.J. Simpson, S.D. Pena, 'Mitochondrial D-loop ''Signatures'' Produced by Low-stringency Single Specific Primer PCR Constitute a Simple Comparative Human Identity Test', *Am. J. Hum. Genet.*, **58**(3), 609–616 (1996).
169. H. Druid, P. Holmgren, B. Carlsson, J. Ahlner, 'Cytochrome P450 2D6 (CYP2D6) Genotyping on Postmortem Blood as a Supplementary Tool for Interpretation of Forensic Toxicological Results', *Forensic Sci. Int.*, **99**(1), 25–34 (1999).

Pyrolysis Gas Chromatography in Forensic Science

Robert D. Blackledge
Navy Criminal Investigative Service, Regional Forensic Laboratory, San Diego, USA

1 Introduction 1
2 Pyrolysis Methods 1
2.1 Resistive Heating 1
2.2 Curie Point (Inductive Heating) 2
2.3 Microfurnace 2
3 Experimental Parameters 2
4 Quality Control 3
5 Forensic Sample Types 4
5.1 Paint 4
5.2 Fibers 4
5.3 Rubber and Plastic 4
5.4 Photocopy Toners 6
5.5 Adhesives and Tapes with Adhesive Backings 6
5.6 Other 6
6 Simultaneous Pyrolysis Alkylation 6
7 Mass Spectrometry Detection 7
8 Data Interpretation 9
9 Analysis Plan Considerations 10
Abbreviations and Acronyms 11
Related Articles 12
References 12

Pyrolysis (as the term is used by analytical chemists) is the breakdown of organic molecules by the application of heat in the absence of oxygen. Because oxygen is absent, combustion (which would produce primarily carbon dioxide and water vapor) does not occur, and the pyrolysis products are characteristic of their parent molecules. Pyrolysis is most often applied to characterize polymeric materials or other samples that cannot normally be vaporized without undergoing decomposition. In pyrolysis gas chromatography (PY/GC) the carrier gas (usually helium or nitrogen) sweeps the pyrolysis products onto a gas chromatography (GC) column.

1 INTRODUCTION

PY/GC is the breakdown of organic molecules by the application of heat in the absence of oxygen and the separation and detection of the pyrolysis products by GC. PY/GC is most often used in forensic science to help answer the question: ''Could samples A and B have originated from a common source?'' A ''yes'' answer may help to provide an association between a suspect and the crime. PY/GC is a destructive technique and forensic science samples are often quite limited. However, a sample size of a few micrograms down to even tenths of a microgram is all that is required. If a flame ionization detector (or other standard GC detector) is used to identify the elution of the pyrolysis products, the comparison of samples A and B is based on pattern matching (retention times and relative peak heights). If the GC is equipped with a mass selective detector (MSD), a positive comparison is strengthened since it may be possible to identify the various peaks based on their mass spectra.

2 PYROLYSIS METHODS

To be of forensic science value the pyrolysis must be performed in a reproducible manner. PY/GC will then produce pyrolysis chromatograms (known as pyrograms) that serve as identifying ''fingerprints''. Resistive heating, Curie point (inductive heating) and microfurnace are the three most common pyrolysis methods, and commercial instrumentation is available for each.

2.1 Resistive Heating

In the most popular pyrolysis system based on resistive heating a self-sensing resistively heated platinum wire coil or ribbon is used to heat the sample rapidly. In the coil probe the sample may be placed either in a quartz boat or inside a quartz tube, which is then inserted into the middle of the coil. In the ribbon probe (Figure 1) the sample (which is usually a high boiling liquid or a solid dissolved in a volatile solvent) is applied directly to the ribbon. The liquid should flow along the ribbon leaving a relatively uniform thin coating. The sample-containing resistive heating unit (probe) may be directly coupled to the GC injection port (Figure 2),[1] or it may be in a separate unit having a heated transfer line interfaced to the GC injection port. In either case, upon pyrolysis the carrier gas sweeps the vapors into the GC column. An electrical cord extends from the probe to a control unit which controls the heating modes (pulse or programmed) and the pyrolysis temperature and duration. Pyrolysis temperatures may be selected in 1 °C increments up to 1400 °C. Figure 3 shows the results from the successive pyrolysis of a polystyrene sample and is indicative of the potential reproducibility of the method.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

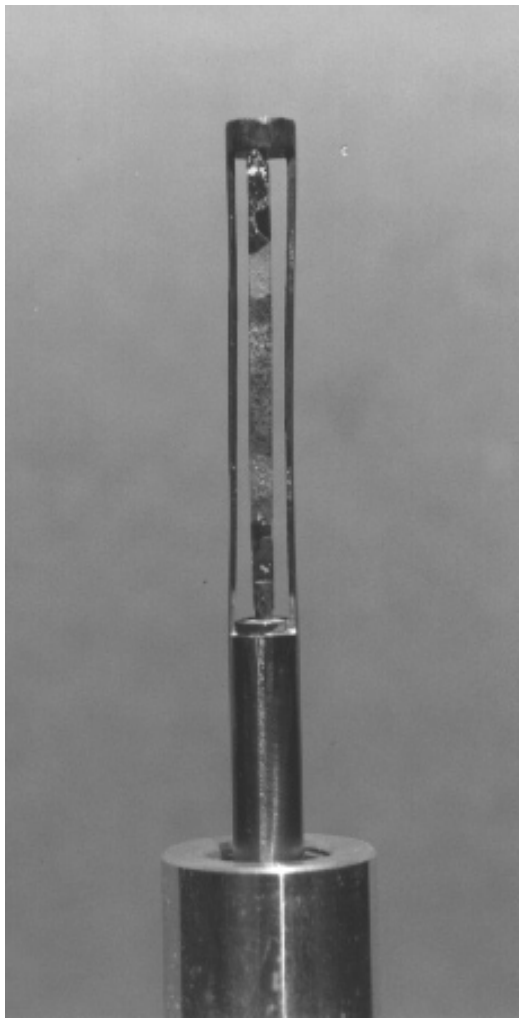

Figure 1 Tip of ribbon probe version of Pyroprobe® (CDS Analytical Inc., Oxford, PA, USA).

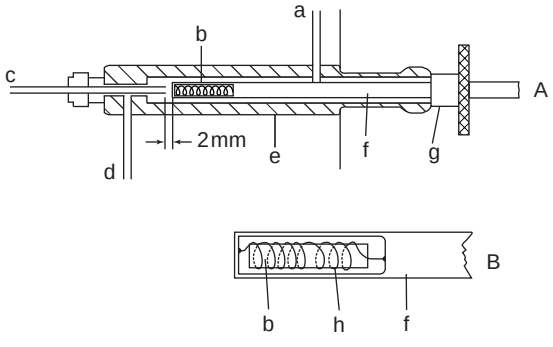

Figure 2 Schematic of a resistive heating pyrolysis unit (Pyroprobe®, CDS Analytical, Oxford, PA) interfaced to the injection port of a capillary column GC. a, carrier gas; b, quartz sample tube plus sample; c, capillary column; d, split vent; e, injector port; f, Pyroprobe®; g, septum retainer; h, platinum wire heater coil. (Reproduced by permission of Elsevier Science B.V. from S.A. Liebman, T.P. Wampler, E.J. Levy, *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **7**, 172–184 (1984).[1])

2.2 Curie Point (Inductive Heating)

In Curie point pyrolysis the sample is on or contained within a ferromagnetic wire or foil. The foil is within a radiofrequency field (RFF). By induction heating the RFF heats the foil, whose temperature rises until it reaches its Curie point, the temperature at which it changes from ferromagnetic to paramagnetic. At the Curie point the temperature remains constant. Thus Curie point pyrolysis affords a very reproducible temperature rise time and final temperature (Figure 4). Because the Curie point varies according to the ratio of ferromagnetic metals (iron, nickel, and chromium) in the foil alloy, various Curie point temperatures ranging from below 200 °C to over 1000 °C may be utilized by selection of the appropriate commercially-available foil. Figure 5 is a schematic of a commercial system in which the sample is on a ferromagnetic alloy filament which is centered in a stainless steel injector needle which pierces the GC septum. The photo, Figure 6, illustrates inserting the sample-containing ferromagnetic filament into the stainless steel injector needle. In order to be suitable for the sample type and analysis requirements, ferromagnetic filaments are available as wires, spirals, or tubes. The table in Figure 7 illustrates the reproducibility of three different peaks from five successive pyrograms of the same white automotive paint sample.

2.3 Microfurnace

In microfurnace pyrolysis systems the sample drops or is injected into the center of the reactor zone of a resistively heated, continuous mode microfurnace. Carrier gas sweeps the pyrolysis vapors into the column of the gas chromatograph. A diagram of the furnace assembly of a commercial system is illustrated in Figure 8. Injectors are available for both liquid and solid samples. The four pyrograms in Figure 9 illustrate how the pyrolysis temperature influences the distribution of pyrolysis products.

3 EXPERIMENTAL PARAMETERS

PY/GC must have run-to-run reproducibility to have any analytical value. Only those factors unique to pyrolysis will be considered here, but one should bear in mind that variations may also result from changes in the GC conditions (for example, injection port liner cleanliness and type, injector split/splitless conditions, transfer line and injection port temperatures, carrier gas (usually helium or nitrogen), carrier gas flow rate, column (type, length, diameter, cleanliness and bleed), temperature program, and the type of detector and the conditions for its operation). If a mass spectrometer is used as the detector for the gas chromatograph, it too may be a factor in a lack of

Apparatus	Model	Conditions
Pyrolyzer	PYROLA Pyrolab Sweden	Temp. rise time (t_1): 8 ms Pyrolysis temp.: 700 °C Pyrolysis time (t_2): 2 s Chamber temp. (T_c): 150 °C
GC	3400 Varian USA	Detector: FID (Flame ionization detector) (Attn. 1×10^{-11}) Det. temp.: 325 °C Inj. temp.: 250 °C Carrier gas: He, 20 mL min^{-1} Split: 20:1 Column: DB-1 (J&W), 0.25 µm, 30 m, 0.25 mm i.d.
Integrator	SP 4270 Spectra Physics USA	Col.temp.: 50 °C → (15 °C min^{-1}) → 300 °C

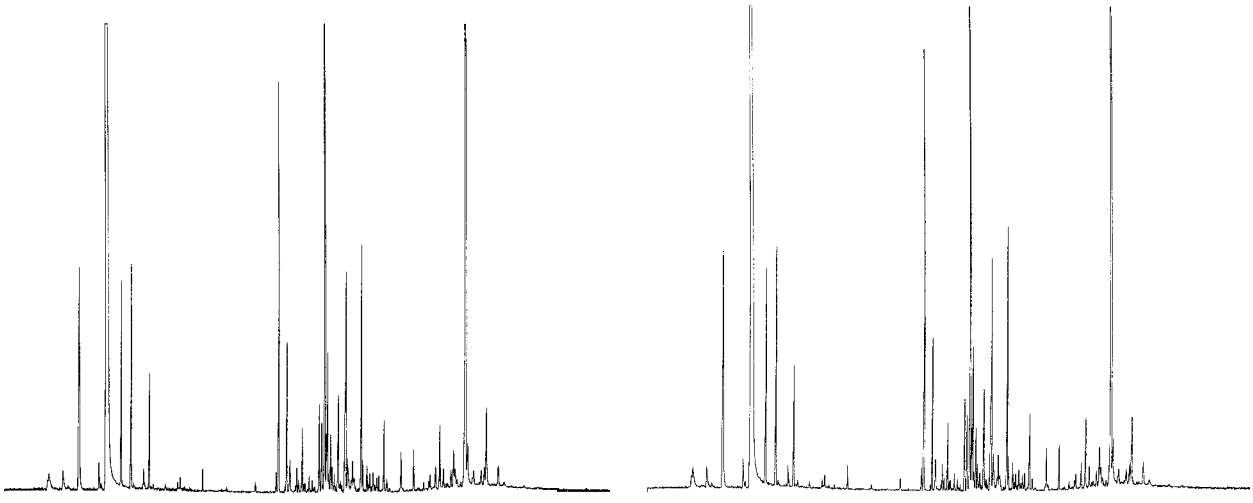

Figure 3 Successive pyrograms of a polystyrene sample by the resistive heating method. (Reproduced by permission of Pyrolab, Lund, Sweden.)

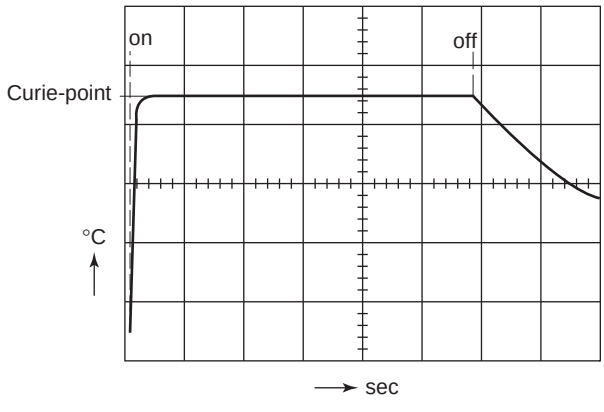

Figure 4 Temperature versus time profile for Curie point pyrolysis for 2.8 s with 760 °C ferromagnetic filaments. (Reproduced by permission of Fischer GmbH, Meckenheim, Germany.)

reproducibility and it is important that factors such as the tune program used, scan range and so on are the same.

Various pyrolysis parameters have been studied by Ericsson[2] (sample type, temperature–time profile, temperature rise time, ambient temperature, carrier gas flow rate, and sample size). Tsuge et al.[3] considered transfer line and injection port heating effects, carrier gas effects (helium was better than nitrogen) and the effects of the inertness of capillary column materials and the buildup of tarry degradation products. Wampler and Levy[4,5] investigated the effects of various injection methods including cryogenic trapping.

4 QUALITY CONTROL

Wampler and Levy[4] have extensively investigated the problem of reproducibility in PY/GC. Just as it is routine practice in GC to inject a test mix of hydrocarbon standards daily, in a similar fashion, Wampler and Levy[6] recommend the use of a pyrolysis standard. They obtained

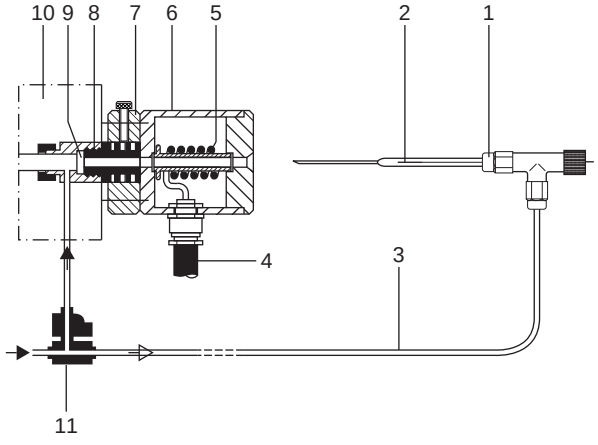

Figure 5 Schematic of the interface of a Curie point pyrolysis system with a capillary column GC: 1, pyrolysis injector made of glass with stainless steel injection needle as well as axial ferromagnetic filament 2; 3, Teflon® carrier gas hose; 4, impulse cable; 5, induction coil; 6, aluminum housing; 7, adapter for fastening; 8, GC inlet; 9, GC septum; 10, GC; 11, carrier gas changeover valve. (Reproduced by permission of Fischer GmbH, Meckenheim, Germany.)

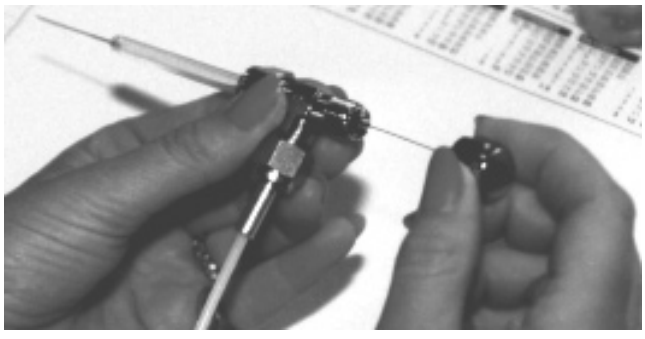

Figure 6 Photo showing insertion of a sample-containing ferromagnetic wire into the Curie point pyrolysis injector. (Reproduced by permission of Fischer GmbH, Meckenheim, Germany.)

good results with Kraton® 1107, a styrene–isoprene copolymer.

5 FORENSIC SAMPLE TYPES

The application of PY/GC to forensic science samples has been reviewed by Wheals[7] (1981), Challinor[8] (1983), Saferstein[9] (1985), Blackledge[10] (1992), and Challinor[11] in 1995.

5.1 Paint

Paint is by far the most common type of evidence sample compared by PY/GC. Paint traces may be exchanged in hit-and-run vehicle accidents and there is frequently an exchange of paint traces when tools such as screwdrivers, crowbars, and hammers are used in break-ins. If the paint traces have only a single layer, and assuming they match in terms of color, texture, and layer thickness, then the analyst need only make sure that the questioned and known samples for pyrolysis approximately match in terms of weight and dimensions (i.e. one sample should not be thin, long and narrow and the other thick and square). If the pyrolysis samples are placed in a quartz tube or boat which is then inserted into a coil of platinum wire, it is important that both questioned and known samples be placed in the center of the tube/boat, which is then placed into the coil so that the samples are at the midpoint of the coil's length. The above admonisions are true for all solid samples. If the paint traces consist of two or more layers the analyst can either try to separate the individual layers and pyrolyze each separately, or they can pyrolyze the intact chips as long as questioned and known samples both have the same layers and roughly the same dimensions and layer thicknesses. This latter approach is frequently easier since it may be difficult to separate the layers without introducing contaminants.

5.2 Fibers

Although synthetic fibers may be identified/compared by methods such as polarized light microscopy and ultraviolet/visible/infrared microspectrophotometry, pyrolysis/capillary column GC in some cases may be able to show differences between different brands, especially if they are polymer blends.

For example, Almer[12] used pyrolysis capillary GC to examine 63 acrylic and 22 modacrylic fibers from various manufacturers. Differences were seen because ''acrylics can contain up to almost 15% of non-acrylonitrile copolymers, and modacrylics can have about 65% non-acrylonitrile copolymers''.[12] On the basis of variations in peak proportions and extra peaks, she was able to subclassify the acrylics into nine groups (not shown) and the modacrylics into six groups (Figure 10).

5.3 Rubber and Plastic

Rubber and plastic bumper guard and trim material may be exchanged in vehicle accidents;[13,14] acts such as kicking in a door may also leave traces of rubber/plastic from shoe soles;[15,16] and rubber traces may be left on knives used in tire slashings.[17] Hu[18] reported a two-stage method for compounded rubber products that first produces a chromatogram of any volatile materials (such as residual solvent traces from the compounding process) and then produces a pyrogram. Hu's method has greater reproducibility and also produces more

t_R	8.4 min	10.2 min	22.2 min
Pyrolysis/GC		Content (%)	
I	26.48	9.89	7.47
II	25.74	11.06	8.74
III	25.15	11.46	9.87
IV	26.28	10.84	8.36
V	24.18	10.42	8.90

t_R	8.4 min	10.2 min	22.2 min
Pyrolysis/GC		Content (%)	
X	25.57	10.73	8.67
S_x	0.833	0.539	0.779
$S_{x.r}$ %	3.26	5.03	8.98

Analytical conditions

(1) *Fischer Curie Point Pyrolysis Model 0316*

Pyrolysis conditions:

Pyrolysis temperature 700 °C
Pyrolysis time 9.9 s
Pyrolysis reactor for manual operation
Ferromagnetic tube for 700 °C
Sample quantity approx > 2–5 μg

(2) *Hewlett Packard GC 5890II with HP 3396A Integrator*

GC conditions:

Fused silica capillary column 30 m × 0.32 mm
Stationary phase SP-2100, 0.25 μm layer thickness
Carrier gas helium 2 cm^3 min^{-1}
Injector split/splitless 290 °C, split outlet 30 cm^3 He min^{-1}
FID 310 °C
Oven temperature programmed from 40 °C after 3 min up to 300 °C at a rate of 5 °C min^{-1}

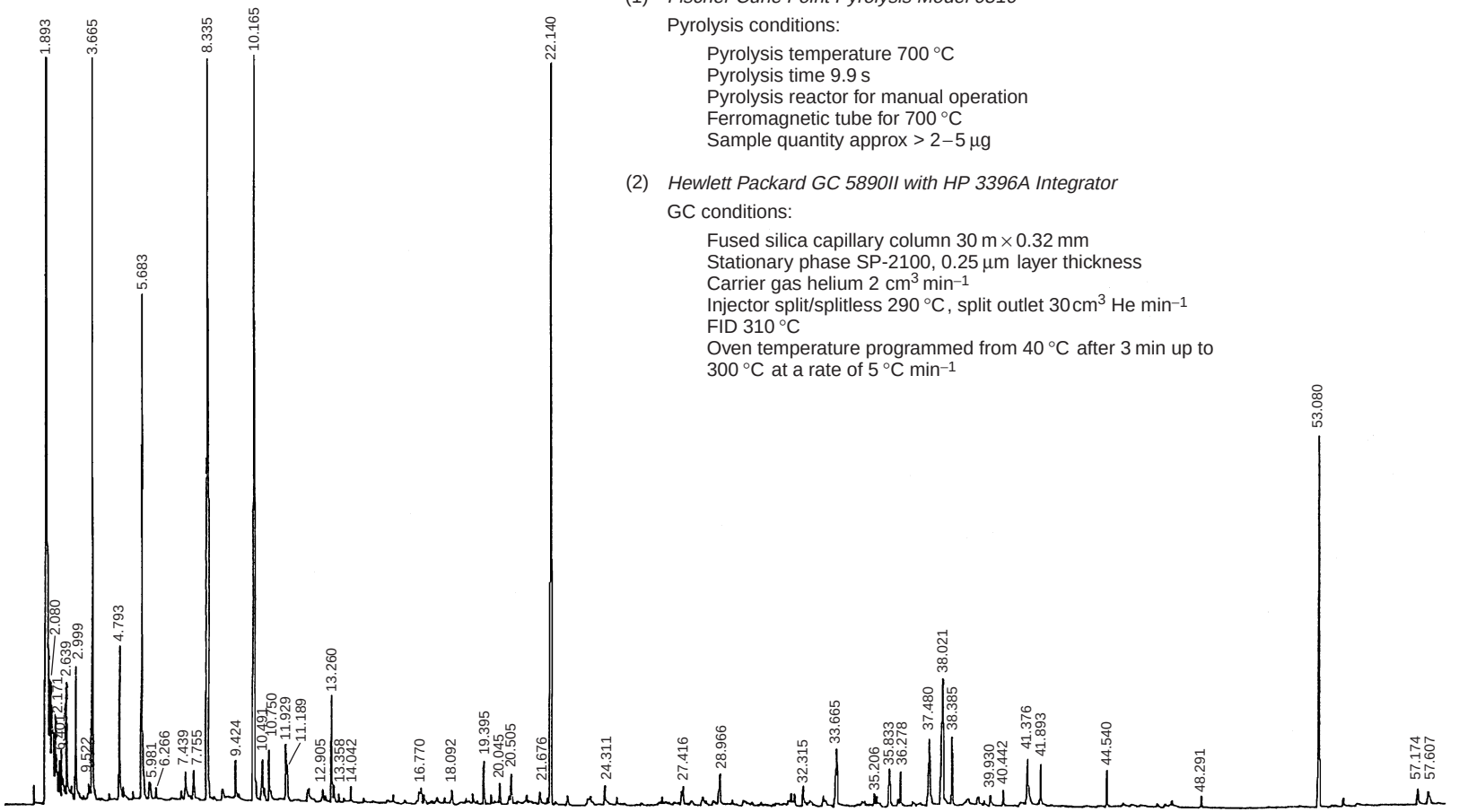

Figure 7 Pyrogram of a white automotive paint sample by Curie point pyrolysis. The table in the upper right indicates the reproducibility of three peaks from five successive programs. (Reproduced by permission of Fischer GmbH, Meckenheim, Germany.)

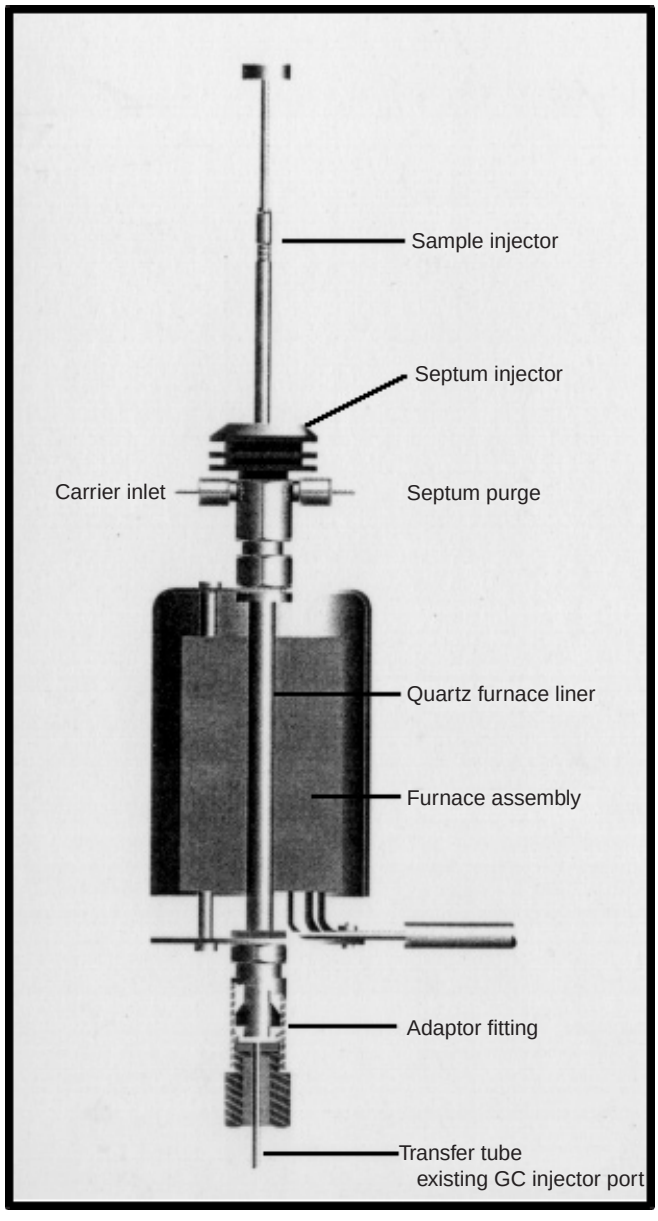

Figure 8 Schematic of a furnace-type pyrolyzer (PYROJECTOR®) interfaced to a capillary column GC. (Reproduced by permission of Scientific Glass Engineering, Incorporated, Austin, TX, USA.)

information for forensic comparisons. Figure 11 shows the results of this method used on questioned and known rubber bumper guard samples from a hit-and-run vehicle investigation.[13]

5.4 Photocopy Toners

Although different brands of toner on photocopy pages may be distinguished by Fourier transform infrared spectroscopy, excellent discrimination may also be achieved with pyrolysis capillary column GC.[19,20]

5.5 Adhesives and Tapes with Adhesive Backings

Adhesive traces and tapes such as duct tape, electrical tape, and masking tape may be important evidence. Subsequent to nondestructive examination methods such as Fourier transform infrared microscopy, excellent reproducibility and discrimination may be provided by pyrolysis capillary column GC.[19–24]

5.6 Other

Other less commonly encountered forensic sample types that have been examined by pyrolysis methods include different types of wood,[25] chewing gum,[26] natural gums, waxes, and resins,[27] and adult and fetal bloodstains.[28] Although hair,[29,30] and different brands of smokeless powder[31] have been examined, little discrimination was provided by PY/GC.

6 SIMULTANEOUS PYROLYSIS ALKYLATION

Challinor[32–34] has extensively studied a PY/GC technique where simultaneous with pyrolysis the pyrolysis products react with a derivatizing agent. In general the method is used with oxygen-containing macromolecular material which is susceptible to hydrolysis with subsequent alkylation. For suitable samples the method provides more structural information about polar components and greater sensitivity. Figure 12 illustrates the pyrogram of an epoxy resin by conventional pyrolysis (top), and then when the same material is subjected to simultaneous pyrolysis methylation using tetramethylammonium hydroxide (TMAH) as a derivatizing agent.

Challinor refers to this simultaneous pyrolysis derivatization reaction as ''thermally assisted hydrolysis and methylation'' (THM). Challinor[35] described four criminal cases in which his THM procedure produced important findings:

- ''The THM reaction is particularly suited to the identification of trace quantities of lipid-containing material such as cosmetic products. A multiple murder, where a fingerprint in a cosmetic body lotion was left by a suspect at the scene, was solved by using this application. The suspect was a known associate of the family victims and, therefore, [the print was] not of great significance. However, body swabs from the victims, who had been sexually assaulted, were examined and found to contain traces of the same body lotion as found in the fingerprint. The suspect confessed to the crime when confronted with this evidence''.

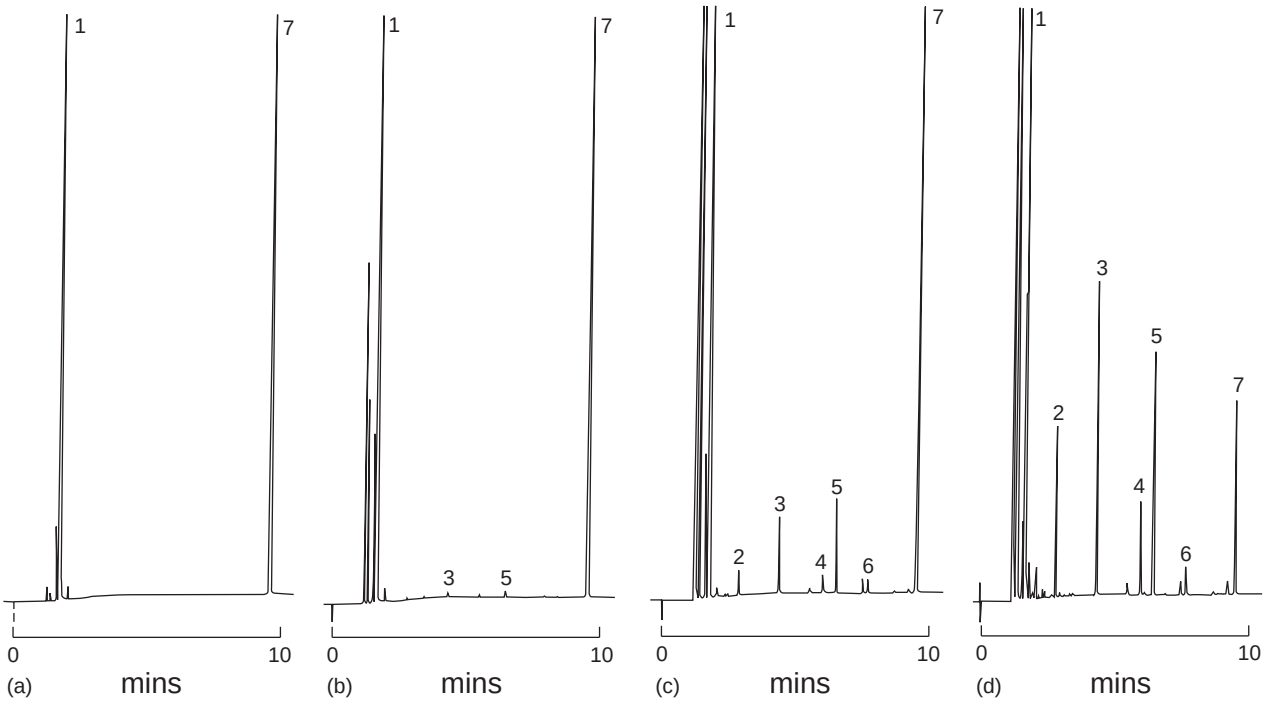

Instrumentation and conditions

Pyrolysis temperature 600 °C–700 °C, Column 25QC3/BP1/0.5, Length 25 m, Inside diameter 0.32 mm, Stationary phase cross-linked methyl silicone, BP1, Film thickness 0.5 μm.

Initial temperature 40 °C for 1 min

Program rate 10 °C min, Final temperature 130 °C for 1 min, Detector flame ionization, Carrier gas hydrogen, Linear carrier velocity 35cm s^{-1} at 40 °C, Split ratio 100 : 1, Injection mode pyrolysis split, Sensitivity 256×10^{-12} AFS, Sample *n*–butylbenzene in pentane (2.5% v/v), Injection volume 0.4 μL, Pressure differential 2 psi.

Figure 9 Pyrolysis in a furnace-type pyrolyzer (PYROJECTOR®) of an *n*-butylbenzene sample in pentane (2.5% v/v) at four different pyrolysis temperatures, pyrogram (a) 600 °C, (b) 700 °C (c) 800 °C, (d) 900 °C. Components 1 pentane, 2 benzene, 3 toluene, 4 ethylbenzene, 5 styrene, 6 *n*-propylbenzene, 7 *n*-butylbenzene. (Reproduced by permission of Scientific Glass Engineering Incorporated, Austin, TX, USA.)

- ''The female proprietor of a gift shop was brutally murdered with a blunt metal instrument. Minute traces of a blue paint were found in the head wounds of the victim. The paint binder was identified as an alkyd enamel by THM/GC/MS (mass spectrometry) using selected ion monitoring for dimethyl phthalate (*m*/*z* 163). Two peaks, corresponding to the *ortho* and *iso*-isomers were detected indicating that the alkyd enamel was a mixed polybasic acid type. This type of alkyd is often used in industrial coatings and could have been found on a tool. The alleged offender admitted that he had hit the victim with a blue painted adjustable spanner [crescent wrench]''.
- ''Fragments of light brown vegetable matter were found in the head wounds of a young person attacked by youths with a cricket bat. The fragment of vegetable matter was identified by THM/GC/MS and pyrolysis of a 1% aqueous TMAH extract as wood belonging to the genus Salix (willow) and was consistent with originating from the cricket bat''.
- ''The driver of a car, which had stopped due to mechanical failure, was badly injured by another vehicle while he was standing by his vehicle. The suspect drove away without stopping but was eventually apprehended. Black smears, identified by THM/GC/MS as polyethylene terephthalate and matching the victim's polyester trousers, were found on the paintwork of the suspect vehicle. As a result, the suspect pleaded guilty in Court''.

7 MASS SPECTROMETRY DETECTION

Poor inter- and intralab reproducibility and an inability to create computer-searchable databases are deficiencies

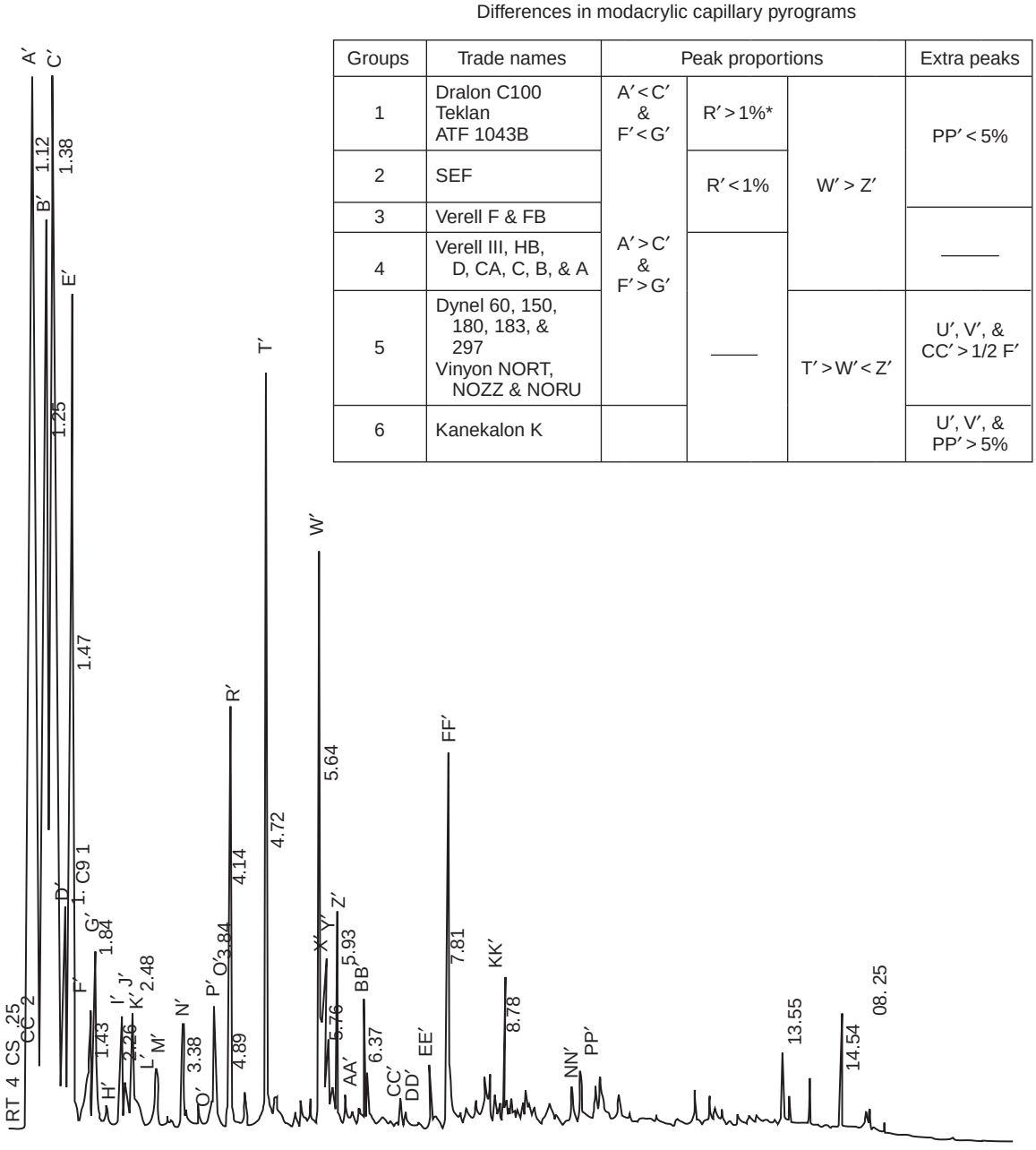

Figure 10 Pyrogram of one of 22 different types of modacrylic fibers and table showing that on the basis of capillary PY/GC they formed six groups. (Reproduced (and adapted) by permission of the Canadian Society of Forensic Science from J. Almer, *Can. Soc. Forens. Sci. J.*, **24**, 51–64 (1991).[12])

frequently sited for PY/GC. Such deficiences are hard to overcome when GC detectors such as flame ionization are used, and sample comparisons are largely based on pattern matching. However, a basis for a reliable computer-searchable database exists if good separation of the pyrolysis products is achieved by capillary column GC and then the essentially pure peaks are identified by their mass spectra.[36] For example, Figure 13 shows the total ion chromatograms obtained from the pyrolysis of four 1995 black automotive paints.[37] As shown in Table 1, most of the pyrogram peaks could be identified from their mass spectra. In some cases it is not necessary to pass the pyrolysis vapors through a gas chromatograph first. Direct pyrolysis/MS eliminates the variables introduced by the GC.

PY/GC/MS (pyrolysis gas chromatography/mass spectrometry) was one of several analytical techniques employed by Sandia National Laboratories (SNL) to

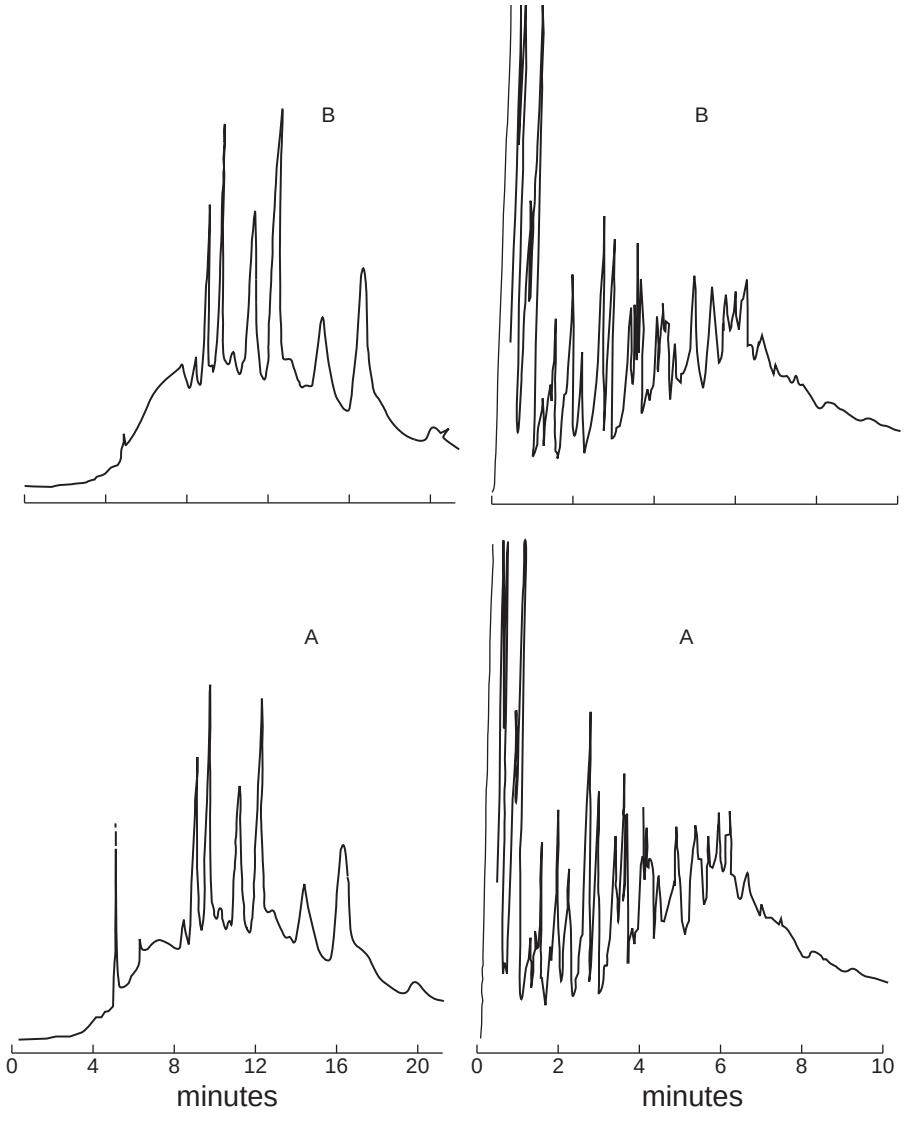

Figure 11 Chromatograms (left) of the volatile components and pyrograms (right) from the nonvolatile polymeric component of a piece of rubber bumper guard material found at accident scene (A) and sample taken from suspect's vehicle (B). (Reproduced by permission of American Society for Testing and Materials from R.D. Blackledge, *J. Forens. Sci.*, **26**, 557–559 (1981).[13])

refute the Navy's contention that the explosion in 1989 in a 16-inch gun turret of the battleship *U.S.S. Iowa*, which killed 47 crew members, was caused by a chemical ignition device that a suicidal guncrew member had placed between powder bags about to be rammed into the breech.[38,39]

Along with other traces, the Navy had found glycols and a single fragment of polyethylene–polyethylene terephthalate (PE-PET). The Navy reasoned that the glycols came from brake fluid or a similar material and that the PE-PET fragment came from a plastic food bag. They postulated an ignition device containing calcium hypochlorite, brake fluid, and steel wool all contained in a plastic bag.

However, SNL pointed out that glycols are present in Break-Free CLP®, a liquid that was regularly used for maintenance of the 16-inch guns. SNL found that Dacron® fibers covered with Break-Free CLP® produced a PY/GC/MS spectrum indistinguishable from that of the PE-PET fragment, and that Dacron® fibers were in a brush used to clean the guns.

8 DATA INTERPRETATION

Although for complex pyrograms capillary column GC does a much better job of separating the individual pyrolysis products than packed columns, the length and

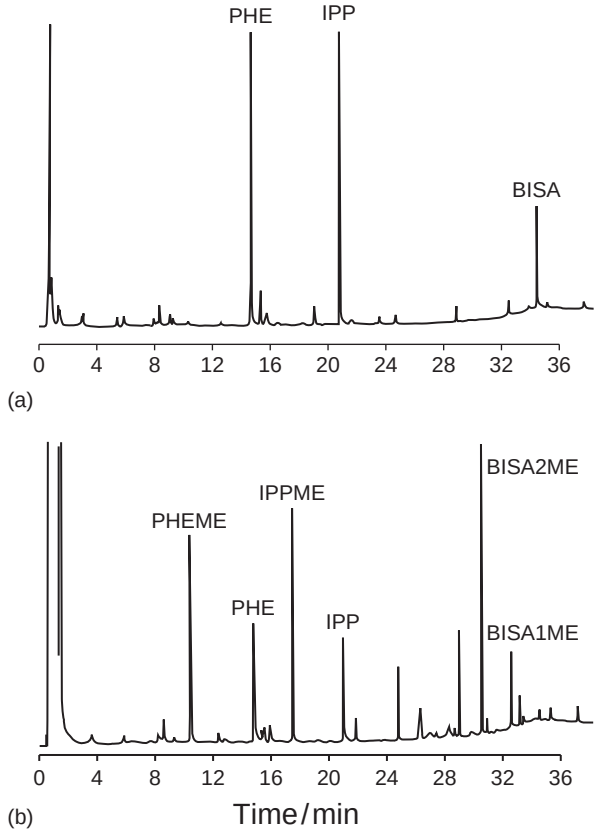

Figure 12 Pyrogram of an epoxy resin by conventional Curie point capillary column PY/GC (a), and by simultaneous pyrolysis methylation (b) using TMAH as a derivatizing agent. PHE, phenol; IPP, isopropenyl phenol; BISA, bisphenol A; PHEME, phenol methyl ether; IPPME, isopropylphenol methyl ether; BISA2ME, bisphenol A monomethyl ether. (Reproduced by permission of Elsevier Science B.V. from J.M. Challinor, *J. Anal. Appl. Pyrolysis*, **16**, 323–333 (1989).[32])

complexity of the patterns can make difficult the visual comparison of questioned and known samples. The use of discriminant analysis methods such as principal component analysis and cluster analysis may aid in data analysis and the discrimination of samples by pyrolysis capillary column GC or PY/GC/MS.[40,41]

9 ANALYSIS PLAN CONSIDERATIONS

Samples for forensic science analysis and comparison are frequently quite limited in amount and possibly contaminated by extraneous materials. These factors must be taken into consideration when developing an analysis plan. Pyrolysis methods will destroy that portion of the sample that was examined. The original sample must be photographed first. If the amount of sample is quite limited it is important to consider first

Table 1 MS peak retention times (RT) and identities for the four black automotive paint PY/GC peaks in Figure 13

Peak RT (min)	Name	C	D	F	G
2.55	Cyclohexane	0	0	X	0
2.81	MMA	X	0	X	0
3.71	Trimethylcyclopentane	0	0	X	0
4.86	Xylene	0	0	X	0
5.17	Styrene	X	X	0	X
5.72	Hydroxy EA	X	0	0	0
6.41	*n*-BMA	X	X	0	0
6.49	Methyl styrene	0	0	0	X
7.24	Hydroxypropyl MA	0	X	X	0
9.13	Naphthalene	0	X	0	X
9.29	Cyclohexyl MA	0	0	X	0
9.54	*iso*-Octyl MA	0	0	X	0
10.40	*n*-Octyl MA	0	0	X	0
11.04	HDI	X	0	X	0
13.32	IDI	0	0	X	0
17.38	MDI	0	X	0	0
19.05	Benzylbutyl-phthalate	X	0	0	0

MMA, methyl methacrylate; hydroxy EA, hydroxyethyl acrylate; BMA, butyl methacrylate; HDI, 1,6-diisocyanatohexane; IDI, 5-isocyanato-1-(isocyanatomethyl)-1,3,3-trimethylcyclohexane; MDI, 1,1-methylenebis-4-isocyanatobenzene. (Adapted from T.P. Wampler, G.A. Bishea, W.J. Simonsick, *J. Anal. Appl. Pyrolysis*, **40–41**, 79–89 (1997).[37])

analytical tests that are nondestructive. These could include examination with stereobinocular microscope, polarized light microscope, and tests for fluorescence when irradiated at various wavelengths. Sometimes a sample that was tested by other methods may be recovered and examined by pyrolysis methods just as long as the questioned and known samples were treated in the same way. For example, samples that were carbon- or gold-coated for examination by scanning electron microscopy/energy dispersive spectroscopy (elemental analysis), might still be recovered and examined by PY/GC. To obtain their infrared spectra, paint samples may be ground up with KBr and pressed into a pellet, but afterwards the intact pellets may be recovered and directly placed in quartz tubes for examination by PY/GC.[42] Or paint, rubber/plastic, or fiber samples could be squashed between diamonds for examination by Fourier transform infrared microscopy[14,43] and afterwards the squashed sample can be recovered. Rubber samples may be extracted with organic solvents and the extracts examined by methods of fluorescence spectroscopy.[44] The solid extracted material can still be examined by pyrolysis methods just as long as questioned and known samples are both extracted in the same manner. Alternatively, adhesive samples may be dissolved in an organic solvent and examined first by methods of fluorescence spectroscopy.[45] Then a portion of the solution can placed on a suitable substrate (KBr or

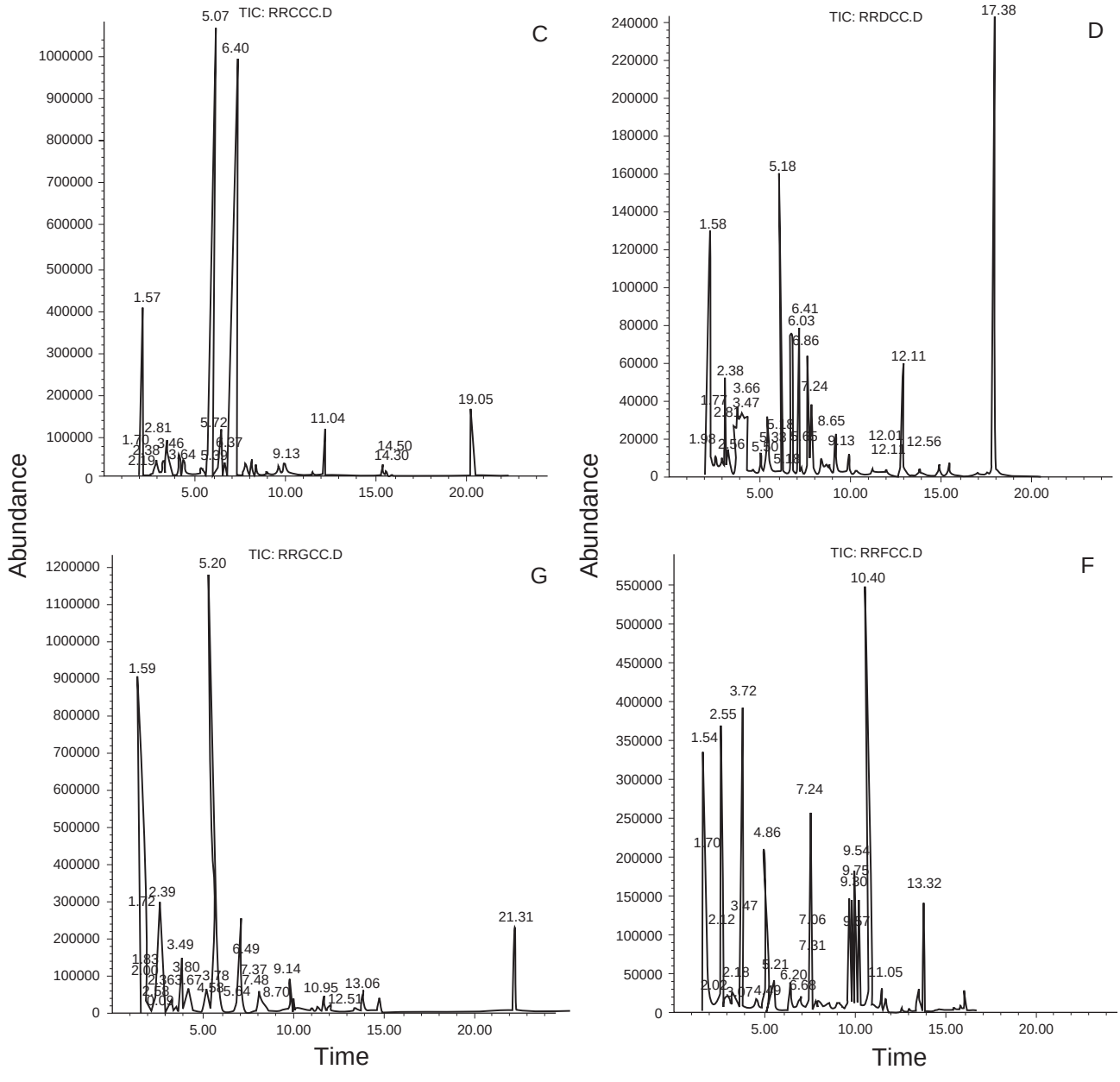

Figure 13 Total ion chromatograms from PY/GC/MS of four 1995 black automotive paints: C, Dupont; D, BASF; G, PPG; F, BASF. (Reproduced by permission of Elsevier Science B.V. from T.P. Wampler, G.A. Bishea, W.J. Simonsick, *J. Anal. Appl. Pyrolysis*, **40–41**, 79–89 (1997).[37])

NaCl disc, 3M IR Card®, ground KBr or NaCl, or gold-coated metal microscope slide), the solvent allowed to evaporate off and the adhesive residue examined in infrared transmission, diffuse reflectance infrared Fourier transform spectroscopy, or Fourier transform infrared reflectance microscopy. The remaining solution could then either be injected in some pyrolysis systems or, in others, dripped onto a pyrolysis ribbon probe with the solvent evaporating and leaving a uniform adhesive coating on the ribbon.

ABBREVIATIONS AND ACRONYMS

GC	Gas Chromatography
MS	Mass Spectrometry
MSD	Mass Selective Detector
PE-PET	Polyethylene–Polyethylene Terephthalate
PY/GC	Pyrolysis Gas Chromatography
PY/GC/MS	Pyrolysis Gas Chromatography/Mass Spectrometry

RFF	Radiofrequency Field
SNL	Sandia National Laboratories
THM	Thermally Assisted Hydrolysis and Methylation
TMAH	Tetramethylammonium Hydroxide

RELATED ARTICLES

Forensic Science **(Volume 5)**
Forensic Science: Introduction

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers • Pyrolysis Techniques in the Analysis of Polymers and Rubbers

Chemometrics **(Volume 11)**
Chemometrics • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Soft Modeling of Analytical Data

REFERENCES

1. S.A. Liebman, T.P. Wampler, E.J. Levy, 'Developments in Pyrolysis Capillary GC', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **7**, 172–184 (1984).
2. I. Ericsson, 'Influence of Pyrolysis Parameters on Results in Pyrolysis/Gas Chromatography', *J. Anal. Appl. Pyrolysis*, **8**, 73–86 (1985).
3. S. Tsuge, H. Ohtani, H. Matsubara, M. Ohsawa, 'Some Empirical Considerations on the Pyrolysis/Gas Chromatography Conditions Required to Obtain Characteristic and Reliable High-resolution Pyrograms for Polymer Samples', *J. Anal. Appl. Pyrolysis*, **11**, 181–194 (1987).
4. T.P. Wampler, E.J. Levy, 'Reproducibility in Pyrolysis: Recent Developments', *J. Anal. Appl. Pyrolysis*, **12**, 75–82 (1987).
5. T.P. Wampler, E.J. Levy, 'Cryogenic Focusing of Pyrolysis Products for Direct (Splitless) Capillary Gas Chromatography', *J. Anal. Appl. Pyrolysis*, **8**, 65–71 (1985).
6. T.P. Wampler, E.J. Levy, 'The Modern Molecular Thermometer: a Standardization Method for Pyrolysis Gas Chromatography', *J. Chromatogr. Sci.*, **22**, 49–55 (1984).
7. B.B. Wheals, 'Review – Analytical Pyrolysis Techniques in Forensic Science', *J. Anal. Appl. Pyrolysis*, **2**, 277–292 (1981).
8. J.M. Challinor, 'Forensic Applications of Pyrolysis Gas Chromatography', *Forens. Sci. Int.*, **21**, 269–285 (1983).
9. R. Saferstein, 'Forensic Aspects of Analytical Pyrolysis', in *Pyrolysis and GC in Polymer Analysis*, eds. S.A. Liebman, E.J. Levy, Marcel Dekker, New York, Chapter 7, 339–371, 1985.
10. R.D. Blackledge, 'Application of Pyrolysis Gas Chromatography in Forensic Science', *Forens. Sci. Rev.*, **4**(1), 1–16 (1992).
11. J.M. Challinor, 'Examination of Forensic Evidence', in *Applied Pyrolysis Handbook*, ed. T.P. Wampler, Marcel Dekker, New York, 207–217, 1995.
12. J. Almer, 'Subclassification of Polyacrylamide Fibers by Pyrolysis Capillary Gas Chromatography', *Can. Soc. Forens. Sci. J.*, **24**, 51–64 (1991).
13. R.D. Blackledge, 'Pyrolysis Gas–Chromatography of Automobile Rubber Bumper Guard Samples', *J. Forens. Sci.*, **26**, 557–559 (1981).
14. A.E. Parybyk, H.J. Kobus, 'The Characterization of Plastic Automobile Bumper Bars Using Fourier Transform Infrared Spectroscopy (FTIR), Pyrolysis Gas Chromatography (PGC), and Energy Dispersive X-ray Microanalysis Combined with a Scanning Electron Microscope (SEM-EDX)', *J. Forens. Sci.*, **35**, 281–292 (1990).
15. C.J. Lennard, P.A. Margot, 'Analysis of Synthetic Shoe Soles by FTIR Microspectrometry and Pyrolysis GC: A Case Study', *J. Forens. Ident.*, **39**, 239–243 (1989).
16. J-K. Ding, H.S. Liu, 'A Study of Identification of Trace Rubber Residues in Marks from Rubber-soled Shoes and Tyres by PY/GC', *Forens. Sci. Int.*, **43**, 45–50 (1989).
17. J.J. Payne, 'Application of Chemistry to Criminalistics: A. Pyrolysis Gas Chromatography of Tire Rubber on Implements Involved in Tire Slashings', Ph.D. dissertation, University of Michigan Press, Ann Arbor, MI, 1987.
18. J.C. Hu, 'Pyrolysis Gas Chromatography Analysis of Rubbers and Other High Polymers', *Anal. Chem.*, **49**, 537–540 (1977).
19. C.J. Lennard, W.D. Mazella, 'A Simple Combined Technique for the Analysis of Toners and Adhesives', *J. Forens. Sci. Soc.*, **31**, 365–371 (1991).
20. R.N. Totty, 'Analysis and Differentiation of Photocopy Toners', *Forens. Sci. Rev.*, **2**, 1–23 (1990).
21. R.D. Blackledge, 'Tapes with Adhesive Backings: Their Characterization in the Forensic Laboratory', in *Applied Polymer Analysis and Characterization*, ed. J. Mitchell, Hanser Publishers, Munich, Germany, Chapter III-F, 413–421, 1987.
22. E.R. Williams, T.O. Munson, 'The Comparison of Black Polyvinylchloride (PVC) Tapes by Pyrolysis Gas Chromatography', *J. Forens. Sci.*, **33**, 1163–1170 (1988).
23. N.L. Bakowski, E.C. Bender, T.O. Munson, 'Comparison and Identification of Adhesives Used in Improvised Devices by Pyrolysis/Capillary Column Gas Chromatography/Mass Spectrometry', *J. Anal. Appl. Pyrolysis*, **24**, 483–492 (1985).
24. C.J. Curry, 'Pyrolysis/Mass Spectrometry Studies of Adhesives', *J. Anal. Appl. Pyrolysis*, **11**, 213–225 (1987).
25. Anon, 'Cedar, Cypress, or Pine?', PEAK, Hewlett-Packard Company, Palo Alto, CA, 2–3, Summer, 1988.
26. F.H. Cassidy, 'Chewing Gum Analysis by Pyrolysis Gas Chromatography', TIELINE, California, Department of

Justice, Bureau of Forensic Services, Sacramento, CA, 59–60, **5**(4), 1979.
27. M.M. Wright, B.B. Wheals, 'Pyrolysis/Mass Spectrometry of Natural Gums, Resins and Waxes and its use for Detecting such Materials in Ancient Egyptian Mummy Cases (Cartonnages)', *J. Anal. Appl. Pyrolysis*, **11**, 195–211 (1987).
28. P.K. Clausen, W.F. Rowe, 'Differentiation of Fetal and Adult Bloodstains by Pyrolysis/Gas–Liquid Chromatography', *J. Forens. Sci.*, **25**, 765–778 (1980).
29. T.O. Munson, 'The Status of Hair Comparison by Pyrolysis/Gas Chromatography', *Crime Lab. Digest*, **14**, 153–154 (1987).
30. F. Ishizawa, S. Misawa, 'Capillary Column Pyrolysis/Gas Chromatography of Hair; a Short Study in Personal Identification', *J. Forens. Sci. Soc.*, **30**, 201–209 (1990).
31. R.O. Keto, 'Comparison of Smokeless Powders by Pyrolysis Capillary Gas Chromatography and Pattern Recognition', *J. Forens. Sci.*, **34**, 74–82 (1989).
32. J.M. Challinor, 'A Pyrolysis Derivatization Gas Chromatography Technique for the Structural Elucidation of Some Synthetic Polymers', *J. Anal. Appl. Pyrolysis*, **16**, 323–333 (1989).
33. J.M. Challinor, 'Structure Determinations of Alkyd Resins by Simultaneous Pyrolysis Methylation', *J. Anal. Appl. Pyrolysis*, **18**, 233–244 (1991).
34. J.M. Challinor, 'The Scope of Pyrolysis Methylation Reactions', *J. Anal. Appl. Pyrolysis*, **20**, 15–24 (1991).
35. J.M. Challinor, 'Pyrolysis Gas Chromatography – Derivatization Techniques for Problem Solving in Forensic Science', Presented at the 12th International Symposium on Analytical and Applied Pyrolysis, Venice, Italy, October, 1996.
36. P. Burke, C.J. Curry, L.M. Davies, D.R. Cousins, 'A Comparison of Pyrolysis Mass Spectrometry, Pyrolysis Gas Chromatography and Infra-red Spectroscopy for the Analysis of Paint Resins', *Forens. Sci. Int.*, **28**, 201–219 (1985).
37. T.P. Wampler, G.A. Bishea, W.J. Simonsick, 'Recent Changes in Automotive Paint Formulation Using Pyrolysis/Gas Chromatography/Mass Spectrometry for Identification', *J. Anal. Appl. Pyrolysis*, **40–41**, 79–89 (1997).
38. M. Heylin, 'Sabotage Refuted as Cause of Battleship Explosion', *C&EN*, page 14, Sept 23, 1991.
39. H. Trewhitt, P. Cary, 'The Sleuths who Cracked the Iowa Case', US News & World Report, page 24, June 18, 1990.
40. R.S. Sahota, S.L. Morgan, 'Vector Representation, Feature Selection, and Fingerprinting: An Application of Pattern Recognition to Pyrolysis/Gas Chromatography/Mass Spectrometry of Nucleosides', *Anal. Chem.*, **65**, 70–77 (1993).
41. D.W. Wright, K.O. Mahler, L.B. Ballard, E. Dawes, 'The Application of Multidimensional Techniques to the Rapid Pyrolysis/GC Profiling of Synthetic Polymers', *J. Chromatogr. Sci.*, **24**, 13–17 (1986).
42. R.J. Audette, R.F.E. Percy, 'A Novel Pyrolysis Technique for Micro Paint Analysis', *J. Forens. Sci.*, **23**, 672–678 (1978).
43. R.D. Blackledge, 'Examination of Automobile Rubber Bumper Guards by Attenuated Total Reflectance Spectroscopy Using a Fourier Transform Infrared Spectrometer', *J. Forens. Sci.*, **26**, 554–556 (1981).
44. R.D. Blackledge, 'Examination of Automobile Rubber Bumper Guards by Synchronous Excitation Spectrofluorometry', *J. Forens. Sci.*, **25**, 583–588 (1980).
45. R.D. Blackledge, 'Comparison of Masking Tapes by Fluorescence Spectroscopy', *Proceedings of the International Symposium on the Analysis and Identification of Polymers*, July 1984, US Government Printing Office, stock number 027-000-01265-5.

Scanning Electron Microscopy in Forensic Science

Samarendra Basu
Albany State University, Albany, USA

1 Introduction **2**
2 History **2**
3 Sample Preparation and Instrumental Parameters **2**
3.1 Trace Evidence 3
3.2 Gunshot Residue 3
3.3 Paint Chips, Glass, and Metallic Fragments 5
3.4 Explosive and Arson Residues 6
3.5 Hairs and Fibers 6
3.6 Mounting, Embedding and Sectioning of Gunshot Residue Particulates 6
3.7 Specimen Coating, Energy-dispersive X-ray Microanalysis and Scanning Electron Microscope Parameters 7
4 Scanning Electron Microscope Signals and Their Forensic Applications **8**
4.1 Secondary Electrons and Surface Topography 8
4.2 Backscattered Electrons and Atomic Number Contrast 9
4.3 Characteristic X-rays and Elemental Analysis 10
4.4 Cathodoluminescence 14
5 Scanning Electron Microscope Signal Processing and Applications **14**
5.1 Dot Maps and Line Scans of Energy-dispersive X-rays 14
5.2 Deflection Modulation Imaging with Energy-dispersive X-rays 17
5.3 Deflection Modulation Imaging with Secondary Electrons and Backscattered Electrons 17
6 Forensic and Environmental Regulations, Methods, and Laws **17**
6.1 History of the Scanning Electron Microscope and Energy-dispersive X-ray Microanalysis in Court 17
6.2 Daubert (1993) v. Frye (1923) Rules 19
6.3 Federal Rule of Evidence 702 and Expert Witness 19
6.4 Environmental Monitoring and Scope 19
6.5 Monitoring of Radiation Hazards 20
6.6 Survey of Methods 20
7 Quality Control, Standards and Sources of Error or Troubleshooting **20**
7.1 Data Quality Objectives 20
7.2 American Society for Testing Materials E-30/National Institute of Standards and Technology Standards 21
7.3 Monitoring of Vibration and Stray Magnetic Fields 21
7.4 Control or Elimination of Charging Effects 21
7.5 Control or Elimination of the Sources of Spurious X-rays 21
7.6 Control of Instrument Performance and Detector Geometry 21
8 Comparison with Other Microbeam Analytical Methods **21**
8.1 Comparison of Technology 21
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **24**

Scanning electron microscopy is providing increasingly definitive solutions to criminal problems since its commercial debut in 1965. This has been due to the ability of the scanning electron microscope (SEM) to simultaneously produce several electron probe-induced signals from the specimen, which generate readily interpretable images of surface topography and material composition. The successful applications of these signals are determined by sample preparation and instrumental parameters. No other microbeam technology combines high resolution (2–5 nm) of the topographic (secondary) electrons with large depth of field for three-dimensional viewing. The SEM is indeed ideal for stereomicroscopy. The versatility of the SEM stems from its additional capability to process each specimen signal by various contrast enhancement methods, such as line scanning, deflection modulation (DM), area mapping, etc. These methods allow an intuitive, stylistic, and synthetic analysis of the image and are ideal for quality control analysis. Digital SEMs have pioneered in automated image processing and unattended search and analysis of particulates. The combined SEM and energy-dispersive X-ray microanalysis (EDX) is the most definitive technique in testing for gunshot residue (GSR) particles, collected by the glue-lift technique. In the analyses of other trace evidence, such as hair and fibers, in physical matching, and in nondestructive elemental analysis of physical evidence, the SEM/EDX is the most efficient of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

all microbeam technologies. From firearms, bullet wounds and human bones, to plants, pollen and fungi, the list of criminal evidence examined by SEM/EDX is endless. However, the SEM/EDX is not ideal for quantitative analysis of elements present as traces (<1% w/w). The SEM lacks the three-dimensional sectioning abilities of scanning beam confocal microscopes. The imaging capabilities of the SEM surpass these limitations. The SEM is, therefore, a major tool in forensic research and investigation.

1 INTRODUCTION

SEMs use a finely focused primary beam of electrons (energy 5–30 kV) to make sweeps as a raster does across the specimen surface in synchronization with a cathode ray tube (CRT).[1–5] As the primary electrons impinge on the specimen, a multitude of electron-beam-induced signals are generated within the specimen and from the specimen surfaces. Of these, the most important signals are secondary electrons (SEs), backscattered electrons (BSEs), characteristic X-ray photons (i.e. X-rays), and cathodoluminescence (CL). Each signal is detected and amplified to modulate the brightness of the CRT i.e. the SEM video monitor. The resultant information in each signal is obtained as an image of the specimen.[1–6] Several of these signals may be recorded concurrently. These images can be obtained separately or in superimposition. The major advantage of the SEM is the high resolution obtainable with the SE, or topographic electrons. Another important advantage of the SEM is its large depth of field. The large depth of field and the shadow relief effect of the SE and BSE electron contrasts make the SEM images appear three-dimensional. Therefore, the SEM is also ideal for stereomicroscopy. The combined SEM/EDX technique is indispensable for fast, nondestructive, physicochemical analysis of trace evidence and solid objects of any size. Environmental SEMs are now able to study corrosion, catalysis and condensation problems at high (atmospheric) pressure. Digital SEMs allow automated image processing and unattended search and analysis of particulates. The intrinsic drawback of the SEM is that it lacks the three-dimensional sectioning abilities of scanning beam confocal microscopes. The SEM/EDX is not ideal for quantitative analysis of elements present as traces. Energy-dispersive X-rays are subject to the matrix effects of Z (the atomic number), absorption, and fluorescence, and a high Bremsstrahlung continuum for low-atomic-number elements. These effects reduce the ability of the SEM/EDX to detect an element below a minimum detection limit (MDL) (0.1–0.3% w/w).

2 HISTORY

The history of the SEM started long before Professor C.W. Oatley introduced the concept of an SEM in 1948 to the department of Electrical Engineering in the University of Cambridge in England. Three German physicists who laid down theories and designs of electron optics were H. Stintzing, M. Knoll, and M. Von Ardenne. Stintzing first applied for a patent on the designs of electron optics on May 13, 1927 and his application (German patent number 485,155) was granted in October of 1929. In 1935 Knoll first proposed the principle of the SEM. In 1938 Von Ardenne first built a scanning transmission electron microscope (STEM) by adding scan coils to a transmission electron microscope (TEM). The outpouring of interest in Germany in the development of the TEM and the outbreak of war put an end to the work of these physicists. The SEM had to be developed in the laboratory of Professor Oatley in the hands of his able doctoral students, D. McMullan, K.C.A. Smith, T.E. Everhart, O.C. Wells, R.F.W. Pease and others. Oatley and McMullan built their first SEM in 1952. This instrument was able to produce a resolution of 50 nm (500 Å). By 1963 R.F.W. Pease built a system (SEMV – the fifth SEM) which became the prototype of the Cambridge Scientific Instruments Mark I Stereoscan®. A.D.G. Stewart, chief engineer at Cambridge Instruments Ltd (England), who was once a student of Professor Oatley, commercialized this model of SEM in 1965. Since this commercial debut of the SEM, many technological advances have been made to improve the instrumentation and performance. Of these advances the notable ones are the developments of the lanthanum hexaboride (LaB_6) cathode, the field emission gun, and associated high-vacuum system (all in 1969) and the digital imaging, image processing, and automation of the SEM in 1976. The years 1965–1968 also marked the advents of wavelength-dispersive spectrometry (WDS) and energy-dispersive spectrometry (EDS) of characteristic X-rays, and their coupling to the SEM and the electron probe microanalyzer (EPMA), which is frequently referred to as the electron microprobe.[7] Historically, the SEM and the EPMA evolved as separate instruments although these two instruments are similar. They differ mainly in the manners they are used, i.e. one for scanning and the other for fixed beam analysis.

3 SAMPLE PREPARATION AND INSTRUMENTAL PARAMETERS

Sample preparation for the SEM and instrumental parameters are keys to successful applications of the signals arising from SE, BSE, X-rays, and CL. Each

signal seems to have a preference for certain sample characteristics and substrate or mounting materials. With the SE electrons, the imaging of highly irregular and convoluted objects is as much of an art as a science. With characteristic X-rays in microanalysis, the sample surface must be flat so that matrix effects due to Z, absorption, and fluorescence are minimal.[5] This is also true for the BSE signal, giving Z contrast with normal incidence and high take-off angles, because the BSE signals producing amplitude contrast (i.e. topography) are minimal with a flat surface. At oblique incidence and low take-off angles the Z contrast vanishes and the amplitude contrast remains.[8,9] Most adhesive tapes and conducting paints used to mount SEM samples for SE and BSE imaging are cathodoluminescent. One must consider noncathodoluminescent substrate and mounting materials for CL imaging.[10]

A variety of procedures are generally used to prepare inorganic materials for SEM.[5,11] Metallic specimens are prepared according to the type of information to be obtained from the specimen. Biological, organic, polymeric, and hydrated materials must be preserved by fixation, dehydration, drying, and coating.[12] Critical-point drying (CPD) is suitable for most biological objects[13,14] except membranous organelles like mitochondria and specimens containing a polysaccharide surface coat (e.g. yeast and other fungi).[15] Freeze drying is also suitable for all types of biological specimens.[12] An authoritative article on general procedures for specimen mounting for the SEM is that by Murphy.[16] The following section outlines only certain procedures used by this author to prepare forensic materials for SEM/EDX analysis. Details of these and specific procedures of several forensic, occupational, and environmental heath practitioners have been compiled elsewhere.[17]

3.1 Trace Evidence

Polished carbon planchets (diameter 13 mm or 0.5 in; thickness 3 mm or 0.125 in; Ernest F. Fullam, Inc.®) are perhaps the most suitable substrates for the SEM/EDX analysis of trace evidence such as GSR particles.[18] These polished carbon planchets are more expensive than their precursors – the highly coarse graphite planchets. The graphite planchets are smoothened by grinding with silicon carbide papers. These smooth planchets are further polished by using some polishing compounds. A typical grinding and polishing sequence might be 320, 400, 600 grit silicon carbide papers followed by 3 μm, 1 μm and 0.25 μm alumina or gold or diamond dusts.[5] This standard metallographic work may be done manually or by using an automated rotating belt on a wheel. The planchets are cleaned with soap and water in an ultrasonic cleaner. A further sonication with petroleum ether or methanol removes the remaining carbon dusts from these polished carbon planchets.

3.2 Gunshot Residue

For GSR collection by the modified glue-lift technique,[19–23] these polished carbon planchets are mounted with adhesive tapes on the 13-mm diameter sanded top of hat-like plastic thimbles (Caplug®). These plastic thimbles are inexpensive. They also allow ease of operation because the collector can be lifted by inserting a finger into the thimble. A dilute (1 : 3 v/v) solution of Rubber Cement® (Stafford-Reeves, Inc.®; Carter's Ink Co.®) in toluene is applied as one or two droplets to the polished surface of the planchet and the glue solution is allowed to spread uniformly. The excess liquid is soaked off the edges of the planchet with a torn filter paper. The thimble supporting this glue-lift disk is then covered from the top with an unsanded thimble. This collector set is allowed to stand inside a transparent cylindrical plastic box and is held in place by the snap-in (tight-fitting) cover of the box (height 2.54 cm). Four such boxes are stored in a square plastic box (height 3.2 cm). The GSR kit consists of this box, the collection procedure, and a pair of polythene gloves, all packaged into a rectangular paper box as shown in Figure 1. The New York State Police have used these kits in the field during the period 1980–1989.[20,21,23] With a finger inserted into a collector thimble, the attached planchet was gently touched five times to a specific area: either the palm or the back of a particular hand, as marked on the side of each thimble. Thus, two collections were

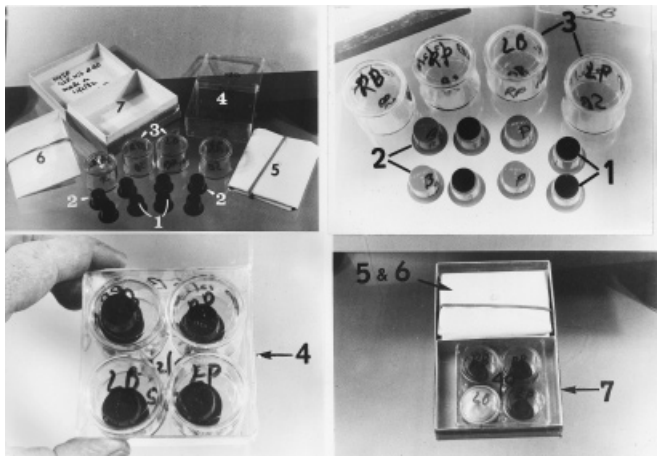

Figure 1 New York State Police GSR kit based on the glue-lift collection technique: 1, carbon planchets (four); 2, thimbles; 3, transparent plastic cylindrical boxes (four); 4, square plastic box (one); 5, GSR instruction and form; 6, polyethylene gloves (one pair); 7, paper box (one). (Reprinted with permission from Basu.[21])

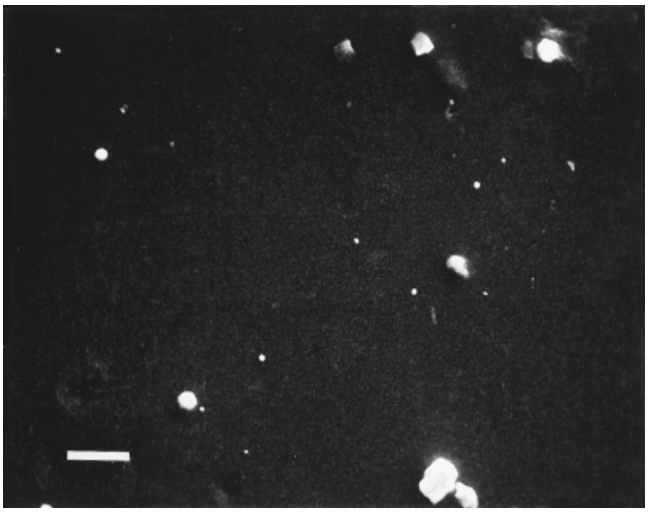

Figure 2 SE image of GSR distribution by glue-lift. Collection by only three touches from the web area (between thumb and forefinger) of the test shooter's hand after one firing. The shooter's hands and the gun were both precleaned prior to firing.[18–20] The field contains two GSR spheroids (PbSbBa), three binary spheroids (2BaSb + 1PbSb) and six monomers (Pb). The field represents only 6% of a 1.5-mm diameter marked circle (search area). In the glue-lift technique four such circles are usually examined per collector disk to render an opinion. Gun, 0.38 caliber revolver; ammunition, Remington 38 spl 200 grain lead bullet; magnification, 550×; bar = 20 μm; 20 kV.

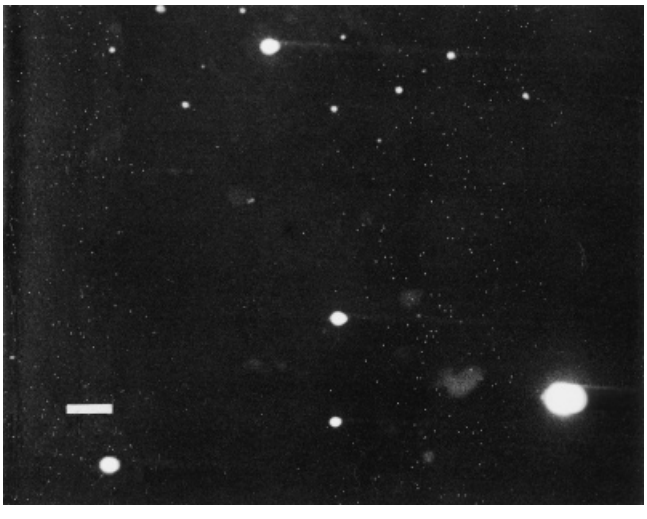

Figure 3 BSE image (Z contrast) of GSR distribution by glue-lift. Collection by five touches from the back of a subject's (victim) right hand. The field shows one GSR and 13 monomers (Pb) of diameters in the range 0.1–0.8 μm. The back of the hand had more than 125 particles (i.e. >8 PbSbBa + 27 PbSb + 89 Pb/Ba in four circles; diameters were 0.1–4.0 μm). These were mainly smooth regular spheroids – a sign of breech GSR! See the examples in Figure 4 and elsewhere.[23] Case, a homicide; gun, 0.32 caliber 5 shot Harrington and Richardson revolver; ammunition, 0.32 caliber lead bullet; magnification, 7930×; bar, 1 μm; 20 kV.

made from each hand, with the total collections being four per subject or per test firing. Optimal tackiness, thickness, and smoothness of the glue layer results in selective rejection of interfering skin cells and sufficient collection of GSR spheroids having condensate or molten morphologies.[18,19] Figure 2 shows GSR distribution and Figures 3–6 the morphology and elemental make-up of GSR.

The shelf-life (14 months or more) of glue-lift disks (GSR kits) improves remarkably when toluene is used instead of 1,1,1-trichloroethane as the solvent to dilute the stock glue. This particular glue contains traces of silicon.[18] Scotch® brand 465 transfer tape also

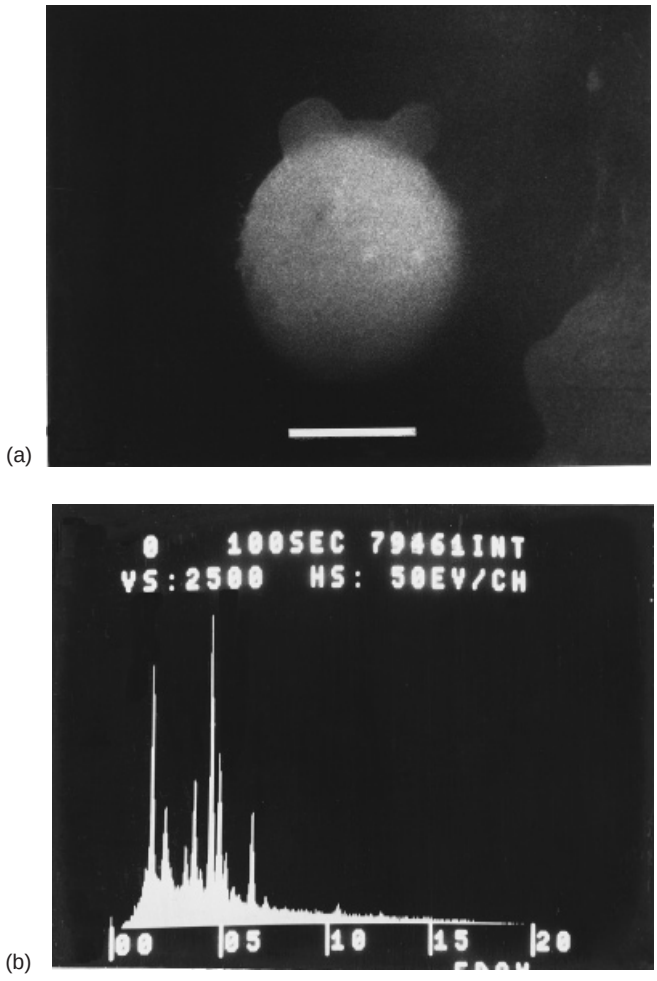

Figure 4 (a) SE image of a regular GSR spheroid of smooth surface topography with two nodules on its head and (b) its EDX spectrum. GSR diameter, 1.9 μm; composition, PbBaSb < CaSiFe. This and other smooth GSRs were detected on the right hand back of the subject in Figure 3. These fluid GSR droplets were forced to escape through the breeches by the backward thrust of the propellant gas mixture. Case, gun and ammunition as for Figure 3; magnification, 24 000×; bar, 1 μm; 20 kV.

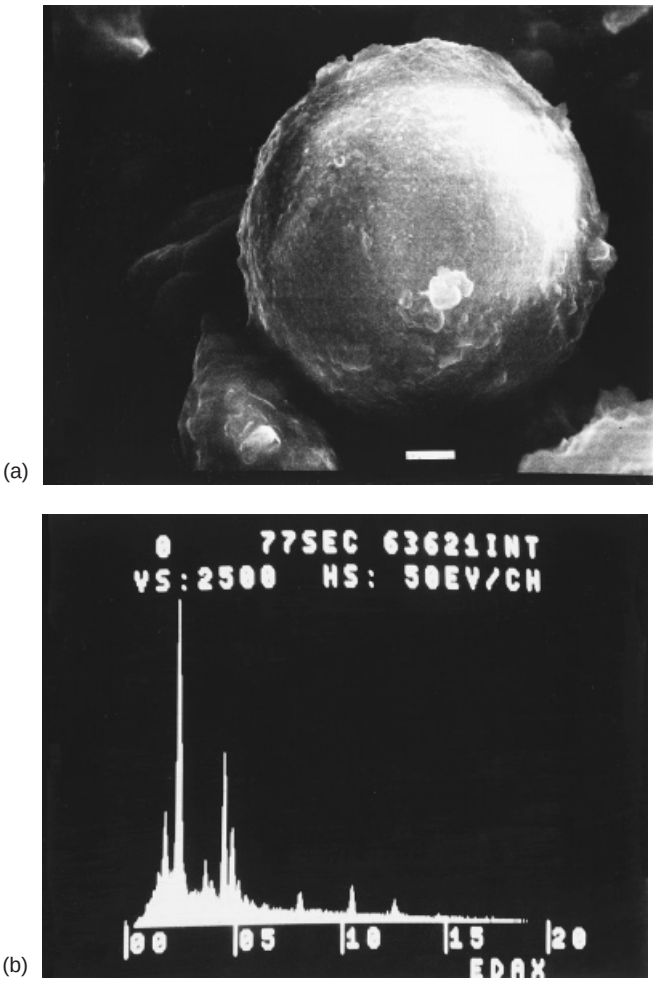

Figure 5 (a) SE image of a Peeled-orange GSR of irregular surface topography and (b) its EDX spectrum. GSR diameter 15.5 μm. [Peeled-orange GSR has a Ba–Sb core and an outer layer of Pb (see Basu and Ferriss[18] and Basu[19]).] This and other GSRs of irregular surface topography were detected along with bullet fragments and Pb residues on both hands (back of each hand) of the subject (victim). These are the characteristics of rebound muzzle-blast residues. The semisolid muzzle-blast GSR captures idle residues and contaminants of the barrel on their surfaces. Case, a homicide; gun, 12 gauge Amnican Eagle model 60 G shotgun; ammunition, 12 gauge Remington rifled slug (Pb); magnification, 4500×; bar, 2 μm; 20 kV.

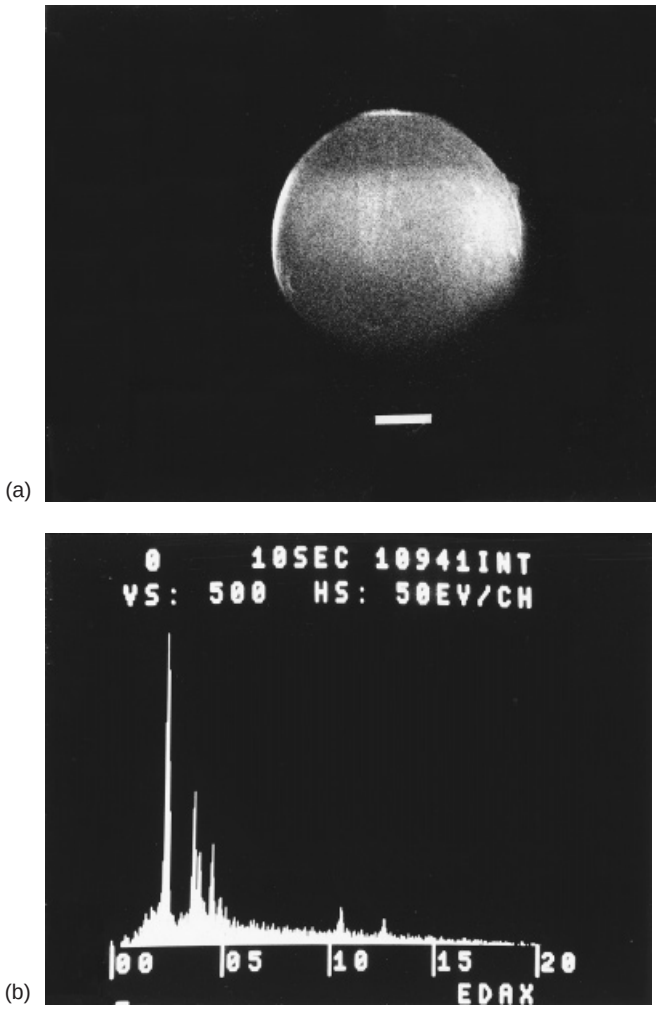

Figure 6 (a) SE image of a regular GSR spheroid of smooth surface topography and (b) its EDX spectrum. GSR diameter 4.5 μm. A total of 21 such smooth GSRs were detected on the right hand back and four on the left back of the subject (suspect). Case, a homicide; gun, 0.38 caliber Rossi revolver; ammunition, Winchester 0.38 caliber certerfire cartridge; magnification, 10 000×; bar, 1 μm; 20 kV.

contains a negligible amount of barium and silicon. This tape is used to mount specimens greater than 1 mm in width. Large objects such as bullets and cartridge cases are mounted on special types of specimen holders.[16]

3.3 Paint Chips, Glass, and Metallic Fragments

Paint chips as small as 1 × 1 mm are held in droplets of glue solution (dilution 1/5, solvent 1,1,1-trichloroethane) until the solvent evaporates and the glue holds the paint chips on edge, i.e. perpendicular to the surface of the polished carbon planchet. This method has been particularly suitable for X-ray dot mapping of elements in the successive layers of paint chips.[24] Comparative elemental and topographic analyses of trace evidence, such as fragments of glass, are performed by mounting the evidence and the control substance on two halves of the same glue-lift disk separated by a diameter. This diameter and the item numbers of the evidence and the control are scribed on the glue layer with the sharp end of a tweezer (Figure 7). The glue-lift disks are also used to lift a monolayer of particulates, granules, metallic dusts and/or scrapings.

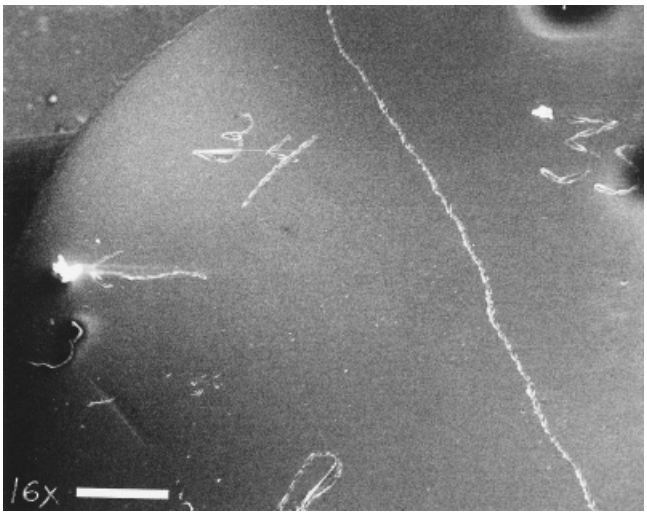

Figure 7 SE image of a glue-lift disk containing evidence (#34) and control (#35). Magnification, 16×; bar, 1 mm; 20 kV.

3.4 Explosive and Arson Residues

The interelement effect[4–7,25,26] of X-ray absorption is minimized when the specimen is well-spread as a thin layer on a smooth, spectroscopically pure substrate, the element of which (e.g. carbon) cannot be detected. Therefore, if the sample is in a suspension, a small volume of it is withdrawn with a Pasteur pipet and settled as a droplet on a polished carbon planchet. The planchet is air-dried (Figure 8). This method is used to detect lead and bromine containing gasoline, arson residues, black powders, pyrotechnique residues, and inorganic explosive residues.

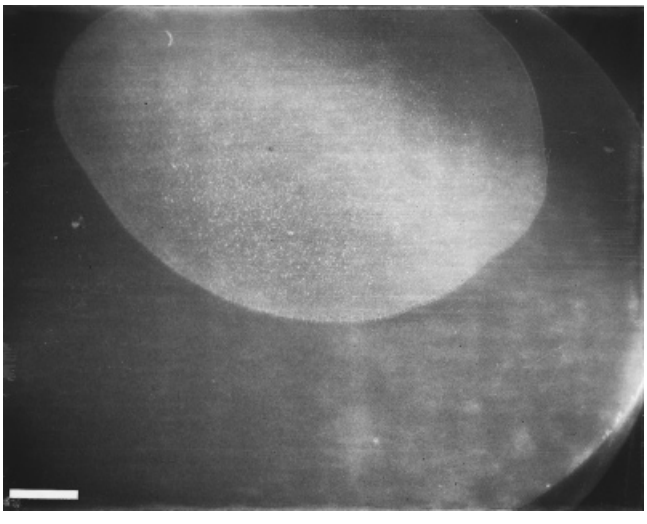

Figure 8 SE image of a polished carbon planchet containing a dried droplet of a flammable liquid (gasoline) recovered from an arson scene. Magnification, 12×; bar, 1 mm; 20 kV.

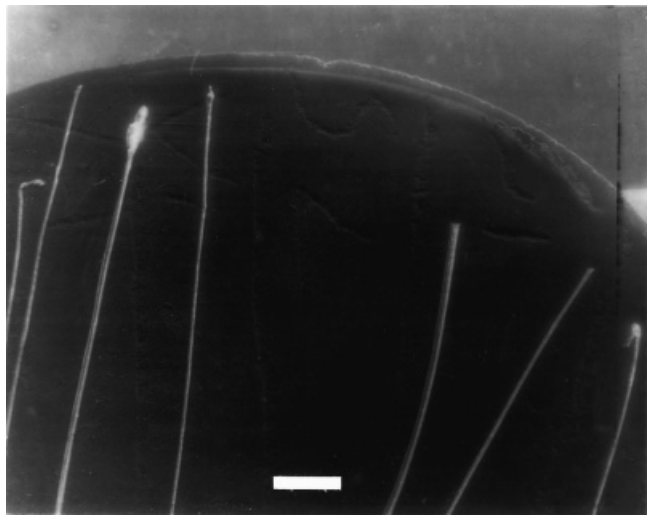

Figure 9 SE image of a glue-lift disk containing hair and fragments of hair (total of seven) in a blind matching test based on comparison of elements and sulfur indices.[27] Magnification, 14×; bar, 1 mm; 20 kV. (Reprinted with permission from Basu.[27])

3.5 Hairs and Fibers

Cut segments of hairs and fibers are arranged in an array on a smooth glassine paper. These are lifted by one touch of the glue-lift disk (Figure 9). Carbon rods or cylinders are also used to wrap hairs around them. When the glue layer is not desirable (as in CL studies) the cut segments of hair are attached to a polished carbon planchet with a clipper.[10] The specimen planchets are usually affixed to an aluminum pin (13 mm or 0.5 in diameter) for carbon coating and further examination in the SEM. When necessary, the disk is separated from the aluminum pin with a scalpel. A convenient alternative to this is the use of pin-type polished carbon mounts.

3.6 Mounting, Embedding and Sectioning of Gunshot Residue Particulates

The distribution of elements within the interiors of particulates is directly related to their origin and/or the process of their formation and growth by the condensation of vapors. These problems can be solved by sectioning particulates and performing X-ray dot mappings of the elements on the cross-sections of the particulates. The following method of mounting, embedding and sectioning was used to examine the formation of GSR particles[19] and can be used for other particulates or residues. The embedding compounds are Epoxy-812, two hardeners, namely nadic methyl anhydride and dodecenyl succinic anhydride, and DMP-30, the accelerating compound (Ernest F. Fullam®). The first three compounds are mixed and applied on several 13 × 102-mm cut pieces

Figure 10 Cross-sectioning of GSR particles: (a) microtome with a steel blade (Will Corp.®); (b) sectioned GSR. Magnification, 600×; bar, 20 µm; 20 kV. (Reprinted with permission from Basu.[19] Copyright ASTM.)

of aluminum foil (Reynolds Wrap), or 8 × 100-mm pieces of Scotch® copper tape. These are fixed in arrays on two cardboards. One is held in close proximity (5 mm) to the breech of a firearm, while the other one is held horizontally about 100 mm below the muzzle as the gun is discharged three to five times. The emanated GSR is captured by the target aluminum foils or copper tapes. Large quantities of primer GSR are obtained by shaking primer cups of expended bullet casings. These are sieved through a 300-µm aperture sieve and spread as a monolayer on the adhesive layer of Scotch® copper tape. The accelerating compound (DMP-30) is added to the epoxy mixture and this embedding medium is applied on the particles of the foils and tapes. Each foil or tape is tightly rolled over a piece of plastic tubing, cut on both sides to make spools. The aluminum or copper spool is inserted into a beam capsule filled with the embedding mixture. The capsule is covered and epoxy is allowed to harden at 65 °C for 6 h and then for another 5 days at room temperature. The capsules are sectioned in a microtome (Will Corp.®) with a steel blade at an angle of 75° (Figure 10a).[19] Figure 10(b) shows a sectioned GSR particle.

3.7 Specimen Coating, Energy-dispersive X-ray Microanalysis and Scanning Electron Microscope Parameters

Whether the mounted specimen should be vacuum coated with carbon (thickness 20 nm) or with a heavy metal (Au, Pt–C, etc.; thickness 1–3 nm), or whether the coating should be avoided, all depends on the purpose of the experiment, the conductivity of the specimen, and on the charging effects. The specimen must always remain grounded. The sample disk must be turned towards the X-ray detector (tilt 30–45°) and be in line with it to produce optimal X-ray counts (about 500–1500 counts per second). The maximal integral (pile-up) count, for example, with the EDAX® (Model 707A) memory is 263 744 (i.e. 2^{18}). This number should not be exceeded. The conditions (SEM/EDX) required for X-ray dot mapping of the elements are described elsewhere.[19] These SEM/EDX studies must be preceded by a thorough search for the sources of spurious metallic emissions in the SEM chamber. The brass surfaces of solid-state (''quad'') backscatter detectors around the aperture of the final lens are the main sources of X-rays due to copper and zinc. The major source of aluminum X-rays is the collimator window of the X-ray detector. These spurious X-ray emissions are caused by high-energy (>10 kV) BSEs reaching those surfaces from the specimen mount. Therefore these surfaces must be heavily coated with evaporated carbon (thickness > 10 µm).[27] This precaution allows the specimens to be examined in close proximity (3 cm) of the collimator window of the X-ray detector. This increases the take-off angles and collection of X-rays. All specimens are examined with 20-kV electrons at an emission current of 75–90 µA. The corresponding beam currents are 0.1–10 nA for a 100–200 µm objective aperture. Accurate measures of the electron beam current are obtained at the specimen level by using a Faraday cup.[5,11] The magnification standards used to calibrate the SEM are 1 µm diameter latex spheres (Figure 11) and a waffle grating.[12] Evaporated gold particles on a carbon film and evaporated gold film on a magnetic recording tape are used to obtain the SE images for any given conditions of the SEM. These images are examined to obtain the resolution of the SEM. This resolution test is crucial during installation of an SEM, and afterwards as the observed resolution may change due to stray electric and magnetic fields and vibrations.

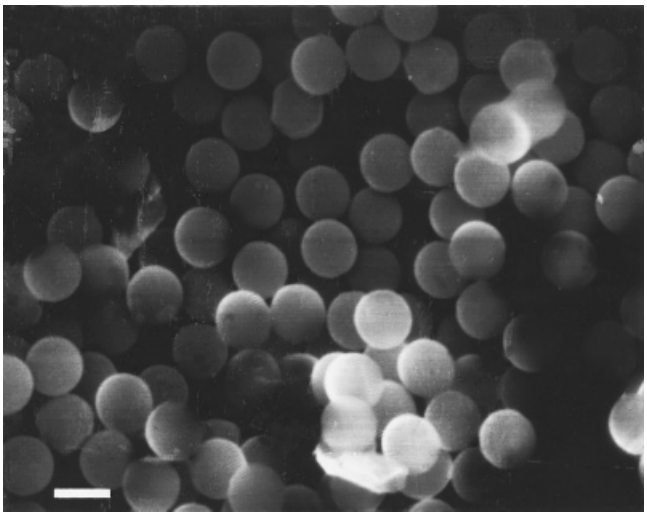

Figure 11 SE image of 1-μm diameter latex spheres used as magnification standard to calibrate an SEM. Magnification, 10 000×; bar, 1 μm; 20 kV.

4 SCANNING ELECTRON MICROSCOPE SIGNALS AND THEIR FORENSIC APPLICATIONS

An authoritative review of SEM in forensic science, though dated, is that by Taylor.[28] Taylor and Noguchi[29] and Pfister[30] have published comprehensive bibliographic updates of forensic applications of SEM up to 1979 and 1982, respectively. Forensic science has advanced since then hand in hand with the technology and understanding of the SEM. Forensic applications of the signals arising from SE, BSE, characteristic X-rays, and CL are identified.

4.1 Secondary Electrons and Surface Topography

As their energies are low (0–50 eV), SEs escape from only a very thin layer of the specimen surface. Because many of them are reabsorbed by the specimen itself, the effective area of SE emission is close to the diameter of the incident electron probe.[31,32] No matter how irregular the specimen is these electrons are collected from every detail of the specimen surfaces within the scanned width of the image.

4.1.1 Detection Limits and Ranges

The result of this superior collection of SEs is the combination of superb depth of field (300–500 times that of a light microscope) and high resolution (7 nm or better) at wide ranges of continuously variable magnifications (20–500 000×).[31–34] Resolutions of the order of 2–5 nm are now usually quoted for commercial instruments.

4.1.2 Applications

This SE imaging mode has been used for the examination of toolmarks, such as firing pin impressions on cartridge cases,[35–39] bullet markings,[35,40,41] inscriptions on coins,[42] staple impressions,[43] toolmarks on cut ends of wires,[39] and other surface marks.[44] The physical matching of two broken surfaces (e.g. glass, ceramics, and vehicle filaments),[45–51] comparison between chemically modified hairs[52] and fibers,[40] and characterization of hairs[53–55] based upon surface texture, dimensional width and mass distribution in cross-sections have been possible. Pelton[56] has used this imaging mode to determine the cause of fiber damage using models based on single fiber fracture, such as lateral compression, clean-cut and cap, etc. Microfibers can also be identified using SEs.[57] Surface analysis using the SE mode has identified a musket shot dated between 1645 and 1655, found in historic St Mary's City in Maryland.[58] Morgan[59] has used SEs to visualize the restored serial number of a firearm. Using the SE micrographs of magnetically separated particles formed by abrasive cutting versus oxygen cutting of the same safe, Zeichner[60] has shown that debris of the two cutting processes have distinguishable characteristics. In a recent study by McDermott[61] topographic analysis of metal particles generated by a grinder fitted with an abrasive disk has linked some suspects with three criminal cases as per Locard's[62] exchange principle. Discrimination of ceramics based on microstructures is now possible using SEs with computerized image analysis to determine grain size distribution, and X-ray diffraction patterns.[63] Topographic analysis in the SEM shows that suspected marijuana leaves, hashish, heroin crystals and mixtures of narcotics and sands, all recovered from crime scenes, have identifiable surface textures and/or characteristics.[41,64,65] It should be noted that surface textures and characteristics are enhanced by a heavy metal (e.g. gold) coating because the coefficient of SE emission for a heavy metal is greater than that of carbon, the heavy metal being a higher-atomic-number element.[66] Unless specified otherwise, these organic and bioorganic physical evidence specimens are usually examined in the SEM after gold or palladium coating with a sputter coater. Quantitative support to these dimensional studies can be given by obtaining stereopairs of the specimen image.[67] In fact, this technique has been successfully applied to the examination of line crossings.[68]

Recent years have observed breakthrough applications of SEM imaging to visible[24] and latent[69–71] fingerprints. The development of latent fingerprints requires the use of magnetic flakes on paper[69] and the deposition of palladium or gold on cartridges.[70,71] The SE imaging has found numerous applications in various branches of forensic medicine. Bohm[72] first utilized the SEM to examine skin injuries and bullet wounds. These studies

were extended to routine autopsy examinations by his group.[73] The SE imaging has solved many problems associated with thermal denaturation of bone collagens[74] and changes of color, morphology, crystal structure, and shrinkage of burnt bones and teeth.[75] Experimentally-controlled heat-treated human bones have shown that ultrastructural features are quite specific to the associated temperature transitions within 200 °C and the age of the deceased.[76] The SEM image patterns of burned or carbonized human remains also correlate with the furnace temperatures between 150 and 1150 °C.[77] The SEM images have been used for estimating age in nonburned archaeological samples from vertebrae and normal ribs and other skeletal remains.[78–80]

Botany had been an underutilized resource in forensic investigation. The SE imaging has made new advances in this area. Plants, pollen, fungi, strawberry, and olive seeds are now being used as both weapons against, and evidence of, criminal acts.[81,82] The trichromes owe their presence to postcrash, not precrash, events.[83] The SEM has contributed to the development of a simple and rapid method for processing organ and water samples for identification of diatoms and for diagnosis of drowning.[84] Recent applications of the SEM in forensic entomology include the detection of chironomid larvae in a homicide victim submersed in water.[85] Other entomological evidence has also been identified using the SE imaging mode.[86] The SEM is now being used in forensic forestry to determine the origin and cause of forest fire by an examination of urban airborne particulate matter.[87]

The SEM has made an unusual contribution to DNA identification by the polymerase chain reaction (PCR). The adhesive tape-lift stubs are normally used for collection and detection of GSR with SEM/EDX to solve shooting crimes. Because the adhesive tape lifts profuse amounts of skin cells, which bury the GSR particles, should the attribution to a specific suspect become uncertain, the epidermal cells on the stubs may be used to extract the DNA, which is then amplified by PCR and typed.[88] The Drugfire Program of the Federal Bureau of Investigation (FBI) can strongly benefit from innovative applications of the SEM potential in computerized imaging.[89] Commercial SEMs (e.g. CamScan®) with a large specimen chamber allow comparative studies of bullets and cartridge cases. These systems can surpass the light microscopic limitations of the Integrated Ballistic Identification System® (IBIS) which are currently in use by the Bureau of Alcohol, Tobacco and Firearms and many state crime laboratories and their networks in the United States. David[90] has applied the SEM to bite-mark analysis. He has shown that the SEM can readily demonstrate individual characteristics in carbon-coated bite marks. The high topographic contrast achievable with a heavy metal (e.g. chromium) coating and the high resolution and superb depth of field obtainable in the SEM, make possible the effective use of stereo techniques in these diverse forensic studies.

4.2 Backscattered Electrons and Atomic Number Contrast

BSEs are high-energy ($\gg$50 eV) elastically scattered electrons which emerge from a larger volume (greater depth) of a specimen.[32] The images from these electrons are inferior in resolution to SE images.[31,32] However, BSEs provide additional information because of their greater sensitivity to atomic number variations with the specimen.[91–95] Currently this is referred to as compositional contrast.[5,11,12]

4.2.1 Detection Limits and Ranges

The threshold equation based on the Rose criterion and the brightness of the electron gun forms the basis for estimating the detection limits and ranges of resolution in SEM images.[5] One must first consider the type of contrast mechanism and the numerical value of contrast to be expected for a specimen. For a flat, polished sample consisting of alternating layers of aluminum and iron of various thickness viewed edge-on with BSE, we can calculate the Al–Fe contrast as $C = 0.45$ (45%). Assuming an average detection efficiency of 0.20, the threshold beam current needed to image this contrast in a 100-s frame is 1×10^{-12} A, or 1 pA. The corresponding probe size can be calculated from the brightness equation. The brightness of a tungsten filament is 5×10^4 A cm^{-2} sr^{-1} at 20 keV and its divergence for a 200-μm aperture and 10-mm working distance is 10^{-2} radians. The corresponding probe size is 2.8 nm. The actual probe size will enlarge to about 5–10 nm due to aberrations. The directionality of BSE causes penetration into the specimen and they undergo plural scattering before they are emitted from a larger volume. Therefore the limiting resolution with BSE is 20 nm or more under the conditions specified. For another specimen containing layers of platinum and gold, the atomic number contrast is only 0.0041 (0.41%). The threshold beam current for this contrast is 9.5×10^{-9} A (9.5 nA). The probe diameter required to contain this current increases to 280 nm (0.28 μm). The corresponding resolution with BSE would be a few micrometers. The chemical composition, mass thickness variation and the presence of edges thinner than the sampling depth of BSE and/or the range of incident beam electrons, are all taken into account to develop strategies for obtaining optimal BSE images.[5]

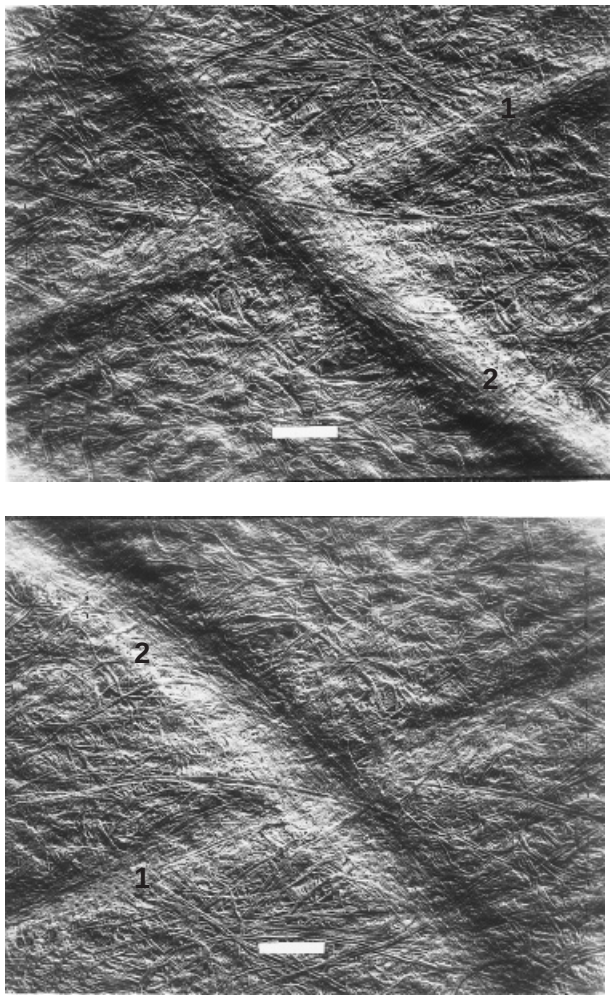

Figure 12 Amplitude contrast BSE images of a cross-over of two pencil lines 1 and 2: (a) recorded first; (b) recorded after the cross-over was rotated clockwise through 180°. Magnification, 60×; bar, 0.2 mm; 20 kV. (Reprinted with permission from Basu.[24])

4.2.2 Applications

Matricardi[96] has used BSE imaging to examine glass fractures. Both SE and BSE imaging have been used to determine the sequence of intersecting lines.[24,97,98] A cross-over of pencil lines is shown in Figures 12(a) and (b). The BSE image in Figure 12(a) was recorded first. The image in Figure 12(b) was recorded after the cross-over was rotated through 180°. Both these images show that line 2 has cut through line 1, which was drawn first.[24]

4.3 Characteristic X-rays and Elemental Analysis

Chemical elements present in a specimen are characterized by their X-ray emissions. Nondestructive analysis of inorganic elements with $Z > 11$ (sodium) is performed either by EDS or by WDS. The chemical information can be obtained as a spectral read-out of the X-ray peaks of the detected elements. EDS is much faster than WDS and, therefore, EDS is mainly used by crime laboratories. However, the EDX spectra generated by EDS are less suitable for quantitative analysis than the wavelength-dispersive X-ray (WDX) spectra produced by WDS. WDS offers 10 times more resolution than does EDS and, therefore, WDS is mainly used with EPMA for research purposes.

4.3.1 Detection Limits and Ranges

The elements detected by SEM/EDX may occupy only a few cubic micrometers of sample volume.[99] However, this capability should not be confused with the analytical ability of this technique. The MDL of an element in a matrix by SEM/EDX is about 0.1–0.3%. The corresponding values at the two-sigma confidence level (twice the background) is 0.2–0.6%. Flat-polished (not etched) elemental standards of the National Institute of Standards and Technology (NIST) are usually preferred for determination of MDLs.[11] In the absence of these standards, seven to eight dissolved oxides of the Spectra-Tech® elements are spread as a monolayer on precleaned polished carbon planchets (Figure 8). The selected elements are those that frequently occur in forensic or physical evidence, such as GSR (PbBaSb), bullets (PbCu) and paints (AlSCa, PbCr), etc. Four of these elements are studied at one time using a multiple pin holder. This allows the operating conditions of the SEM/EDX to remain the same for each element. The electron beam energy (voltage) is chosen to obtain an overvoltage (beam energy/critical ionization energy of the shell) of 2.0. A decrease in overvoltage below 2.0 reduces the amount of the incident electron beam energy made available for the production of characteristic X-rays and the energy range of the emitted X-rays. A beam energy of 20 keV is used for all pure elements and specimens as this beam energy is a good compromise between the need for adequate overvoltage and the need to minimize X-ray absorption in the specimen. The measured peak intensities are adjusted for background radiation by linear interpolation of the lower and higher energy ends of each Gaussian elemental peak, and by background extrapolation of the linear interpolations (i.e. within peaks) for a mixture of elements. The plots of background-corrected intensities ($\log_{10}$ of total counts per second) versus concentrations ($\log_{10}$ of w/w(%)) of the element are fitted to straight lines by the least-squares approach. These plots show intercepts on the axis of concentration when the concentration is below 1% (w/w). These intercepts are the MDLs of the element at the X-ray lines being studied. These intercepts are smaller for Kα lines but larger for Kβ lines. For example, the

MDLs of the Ga Kα and Kβ lines are about 0.012% and 0.04%, respectively. The existence of these MDLs (zero X-ray counts) is clear evidence that SEM/EDX or EDS must be avoided for trace (i.e. less than 1%) elemental analysis when the elements are in a matrix. The described method offers only crude estimates of MDLs and should not substitute the practice of using polished elemental standards. Unfortunately the lithium-drifted silicon EDS with a beryllium window cannot be applied to low-atomic-number elements (Ne, F, O, N, C, B, and Be). Windowless or thin-window EDS detectors allow access to X-ray lines below 1 keV and these systems are necessary for the analysis of these light elements.

The accuracy with which the X-ray peak intensities are converted to elemental concentrations depends on the chemical make-up and configuration (e.g. topography) of the sample, and on the access to a computer program that corrects for the interelement effects due to atomic number, absorption, and fluorescence (ZAF).[4–7,11,25,26] Furthermore, the ZAF method is suitable only for homogeneous samples and when the specimen surfaces are polished enough (not etched) to minimize the erroneous X-ray absorption (e.g. topography or take-off angle effect). Also, application of the ZAF method requires that a value for the pure element intensity is obtained for every element in the sample. The ZAF method and other empirical approaches are more suitable for thin films and thin sections than for bulk samples. The SEM/EDX technique is limited to surface analysis, due to lack of penetration of low-energy electrons (5–30 keV) into the specimen. Therefore, because the homogeneity aspect of physical evidence is often unknown, this analysis is only qualitative. With irregular and inhomogeneous specimens, several examinations are made and the X-ray intensities are corrected for the background X-ray radiation. Such examinations reveal basically the major (>10%) and the minor (>1%) elements detected in the sample, and whether or not two or more submitted samples (evidence and control) have the same general composition. Specific guidelines for automatic EDS analysis and WDS qualitative analysis are given elsewhere.[5,11]

4.3.2 *Applications of Simultaneous Secondary Electrons, Backscattered Electrons, and Energy-dispersive X-rays or Wavelength-dispersive X-rays*

An important application of the signals described up to this point (i.e. BSEs, X-rays, and SEs) is the search and identification of GSR particles, as these particles are characterized by their constituent primer elements (lead, antimony, and barium),[18–23,100–114] their specific spherical shape and condensed (molten) morphologies, both on the surfaces[18,19,101–105,107] and in the interiors,[19] and by their wide range of diameters – from submicrometer to 55 μm or more.[18,19] These GSR particles are distinct from many environmental and occupational particles.[104,108] The physical processes involved in GSR formation,[19] in the transfer of elements from bullet to GSR,[111] in the escape of GSR from the breeches and the muzzle of a gun, and their forced deposition on the firing hand(s) or any hand in close proximity,[21,22] make GSR infallible physical evidence. Consequently, the density distribution[20–22] of all deposited particles, including GSR (PbSbBa), and perhaps also their spatial distribution[113] on the back of the shooting hand(s) have been an aid to reconstructing shooting crimes.[20,21,24] The success rate of the SEM/EDX detection of GSR is incredibly high when precautions are taken against losses and transfers of GSR from the shooting hand(s).[23] The yield of observable GSR on a specimen disk depends on the efficiency of the collection technique, such as glue-lift[18–23] or tape-lift.[101–104,107,115] Two methods of concentrating GSR from tape-lift disks are now available.[106,114,116]

The tape-lift technique is essentially a skin (i.e. DNA) collection technique.[88] The adhesive used in this technique is Scotch® transfer tape (3M brand 465). This tape is 40–60 μm thick and very sticky.[18] The aluminum stub containing this tape is usually dabbed on the shooter's hand many times (>50), or until the stickiness of the adhesive is over. As a result the GSR particles collected in earlier touches are buried under epidermal cells and extraneous materials that pile up in subsequent touches. Very few of these GSR are made available for topographic identification using SE imaging.[18,103] Only the high-energy BSEs and characteristic X-rays which escape the thin regions of these epidermal cells may register some of the particles, but not their real morphologies. This is being practised in automated GSR research and analysis.[112,116–122] By applying plasma-ashing to tape-lifts of GSR, Sild and Pausak[47] have shown that, as these epidermal cells are removed by ashing or burning, the GSR appear on the sample disks. This suggests that the GSR found on the tape-lift disks are mainly those particles lifted in the last few dabbings. The moist endodermal sides of the accumulated epidermal cells may have this lifting ability. However, the retention of these particles requires double carbon coating because the adhesive tape also cracks, makes blisters and pockets, and melts under the electron beam in the SEM.[18,47,103,115]

As a result of these drawbacks tape-lift has the potential to generate many false positive, false negative, and inconclusive results. Obviously these cases require further DNA analysis using the collected epidermis.[88] A recent report on GSR cases using the tape-lift technique shows that GSR positive samples containing one GSR, from two to five GSRs, and more than five GSRs were, respectively, 3.9%, 4.4%, and 1.7% of all submitted samples, with an

overall success rate of 10%. Of the GSR-positive suspects, 31% had only one GSR on their hands.[117] In this and other casework assessments of automated GSR analysis using the tape-lift collection technique, the threshold level of GSR-positive cases happens to be one GSR.[112,118–122] This threshold, consisting of either one ''Unique'' GSR or one ''Consistent'' GSR, originated from the research of the Aerospace Corporation.[100,101,104,107–109] Stronger evidence is forthcoming to suggest that this threshold is highly ambiguous. The potential of secondary GSR transfer from police officers on to subjects has been tested.[123] Forty-three officers were sampled with tape-lifts which were subsequently concentrated[114,116] and analyzed by SEM/EDX. Three of these 43 officers had on their hands ''Unique'' GSR, each containing lead, barium and antimony. [''Unique'' GSR are spheroids of four different compositions (see Wolten et al.[104,108,109]). These are (a) Pb, Sb and Ba; (b) Ba, Ca, Si with a trace of S; (c) Ba, Ca, Si with a trace of Pb if Cu and Zn are absent; (d) Sb and Ba. ''Consistent'' GSR, also caused ''characteristic GSR'', especially if spheroidal but are not ''unique'' to it, are of the following five different compositions (see Wolten et al.[104,108,109]): (a) Pb, Sb; (b) Pb, Ba; (c) Pb; (d) Ba, only if S is absent or only a trace; (e) Sb (rare). ''Full'' GSR (see Basu and Ferriss[18]) are full complements of GSR spheroids each containing Sb, Ba and Pb.] Fifteen other officers had particles containing at least one (Pb/Ba/Sb) or two (PbSb/SbBa/PbBa) of the three elements. Twenty-five officers had no GSR on their hands.[123] This statistic suggests that one-GSR cases must be discarded. Alternatively, the hands of the arresting officers must be sampled. The current users of the tape-lift technique ignore the spherical shape, the condensation (e.g. homogeneity) characteristics and the molten features of GSR in their automated GSR analysis. The irony is that these GSR characteristics were introduced by the developers of the tape-lift technique.[100,101,104,107–109] Furthermore, the discharge residues from cartridge-operated industrial tools (e.g. studguns),[124] cap guns,[104,108] and residues of some paints and fireworks may also contain one or more of the GSR elements (PbSbBa).

The yield of GSR by only five touches of the glue-lift disk is usually sufficient for GSR detection and analysis by SEM/EDX.[18–23] Figure 2 is an example of such collections from a test shooter's hand at a low magnification. The field of view shown contains two GSR spheroids each containing PbSbBa, three binary spheroids containing BaSb (two) and PbSb (one), and six monomer spheroids containing only Pb. Figure 3 represents the collection from a subject's (victim) right hand (back) at 7930× in a case submitted as a homicide. This glue-lift collector had hundreds of GSR in only one 1.5-mm diameter area of the disk. This case turned out to be a suicide based only on the GSR analysis. The glue-lift technique has been tested by officers and investigators of the New York State Police in several such cases. This technique has provided them with a wealth of information about their cases. Figure 3 also highlights that it is only the glue-lift that is able to collect the smallest of the smaller GSR particles as well as larger particles in a wide range of diameters (namely 0.1–55 μm). The tape-lift technique has never been able to produce such data of both quantity and quality.

The glue-lift technique has been applied to 298 cases (total number of kits 380) of the New York State Police with an overall success rate of 70.4%.[23] This success rate does not include an additional 6.4% of the cases where in each case more than 25 binaries (PbSb/SbBa/BaPb) were detected along with bullet fragments and bullet residues on hands. None of these residues were ''Full''[18] or ''Unique''[104] GSR (PbSbBa) spheroids. They had the characteristics of muzzle-blast residues which usually contaminate a hand touching a discharged firearm at the muzzle. Similarly the additional 20.8% of the cases had in each case less than 25 binaries (PbSb/SbBa/BaPb) plus monomers (Pb/Sb/Ba). No opinion has been given in these cases because the detected particles were all contaminants. The success rates with suicides, suspected suicides, and joint homicides and suicides were 81%, 86%, and 90%, respectively. With homicides, assault, reckless endangerment, and prohibited use of firearms, the success rates were 54%, 57%, and 50% respectively. Definitive opinions, including specific hand positions in self-inflicted deaths, were given in many of the successful cases (66.4%). None of these opinions has been challenged or disproved. Figures 4–6 represent typical GSR detected in three submitted cases of homicides. Two of these cases turned out to be suicides based on the analysis of GSR morphologies and their distribution on the back of hands (Figures 4 and 5). These and other cases were guided by the established finding that regular GSR spheroids of smooth surface topography are infallible evidence of fresh breech GSR,[20–23] namely the hand being on the trigger or in its close proximity at the instant of firing. These fresh GSR (PbSbBa) spheroids bounce and roll off the surfaces of the gun due to the pressure of the trigger blast. Conversely, these semisolid regular GSR become flatter and irregular as they strike the exposed surfaces of the gun.[125] Only irregular GSR occur on the surfaces of the fired gun and these could be transferred to a nonshooter's hand due to physical contact with the most contaminated areas of the fired weapon.[21,22] The glue-lift threshold used (N) in these successful cases was 10, where $N = 2\,\mathrm{GSR} + 8$ binaries plus monomers (GSR = PbSbBa; binaries = PbSb/SbBa/BaPb; monomers = Pb/Sb/Ba). This threshold is much higher than the threshold used by the

users of tape-lift (i.e. one ''Unique'' or ''Consistent'' GSR).[109,117–123] If one applies this threshold of tape-lift to casework using glue-lift, then the success rate would be 97.6%. The objective of the glue-lift development in SEM/EDX was to turn this cutting-edge technology into an investigative tool. The users of tape-lift cannot even imagine this fundamental potential of glue-lift even though they are searching for GSR in every case.

The threshold question avoids the ambiguity raised earlier with the tape-lift collection technique. This threshold is also consistent with the physical process of GSR condensation from its three major elemental vapors at a dynamic equilibrium[19] and it is usually representative of the deposits found in test firings where no loss of these deposits is allowed (Figure 2). It is senseless to imagine that physical activity of the suspect shooter selectively removes the binaries (PbSb/SbBa/BaPb) and monomers (Pb/Sb/Ba) but retains only the ''Full''[18] or ''Unique''[104] GSR (PbSbBa) on hands, and vice versa. A further distinction between glue-lift and tape-lift is that glue-lift is able to show consistency in deposition per firing with precleaned guns, also known as the density distribution criterion and is, therefore, able to reconstruct shootings.[20–23] This is beyond the reach of tape-lift because the collection efficiency of tape-lift is only 33%, whereas the collection efficiency of glue-lift is at least 84%. DeGaetano et al.[126] used the glue-lift technique inappropriately by substituting the needed polished carbon planchets with the coarser smooth planchets because the polished carbon planchets cost four times more than the smooth planchets. They failed in their effort to make a good comparison between the two GSR collection techniques. If tape-lift is better than glue-lift, then the users of tape-lift should document thousands of GSRs per sample disk, not one or two.[126] The surfaces of adhesive tapes are much coarser than the surfaces of glue-lift disks and, therefore, make fewer contacts with the GSR deposits on hands. This is another reason why the tape-lift technique has failed to find the smallest GSRs (0.1 μm diameter or less) and to document these at high magnifications (Figure 3). The glue-lift has shown more discrete morphologies of GSR spheroids than all other attempts combined since GSR was first discerned both chemically[127] and physically.[128] Due to larger yields of GSR the glue-lift allows the search of GSR in a few marked circles (total area 7.1 mm^2). With the tape-lift, the entire 1-in or 0.5-in disk has to be searched for GSR. Statistical considerations do not provide sufficient justification for any significant reduction in the sample area of tape-lift to be searched.[129] In summary, the advantages of glue-lift over tape-lift are smaller search time due to greater collection or yield of GSR; minimal pick-up of skin cells and other debris and hence cleaner collection; ability to lift GSR of a wide range of diameters (0.1–55 μm or above) and diversified condensed GSR morphologies; ability to avoid imbedment of GSR in the thin glue layer (a few micrometers thick); no need for coating the sample disks unless these have to be re-examined several months later; and a smaller search area (6%) of each disk to obtain adequate numbers of GSRs.

The SEM/EDX detection of GSR is also applicable to the determination of target distances.[110] These studies involve examination of larger substrate areas than that of a specimen stub and so require automated or semiautomatic search techniques.[110,112,113] The SEM/EDX analysis of GSR has confronted a new issue in recent years. The build-up of residual lead or GSR particles inside the barrel, the chambers, and the trigger housing of an unclean gun have the potential for carry-over of the deposited metals (Pb, Sb, Ba) to residues of lead-free[22,130–132] and antimony-free[133] primers. Environmental health problems associated with airborne lead particulates[134] and acid-rain precursors (sulfates)[135] may have generated a new awakening and have possibly tightened the production of lead- and antimony-containing primers. Therefore the development of new sampling procedures for GSR and explosive residues must continue.[136] An important application of SEM/EDX for the medical examiner's office is the detection and quantitation of GSR, bullet residues and bullet fragments in tissue scrapings collected from a bullet wound.[137] The SEM/EDX has also determined that bone particles and bone-plus-bullet particles are a feature of backscatter from close-range shots to heads.[138] The SEM/EDX has been used in combination with neutron activation analysis (NAA) to identify ammunitions used in a lethal robbery.[139]

A routine undertaking of SEM/EDX is the inorganic elemental analysis of vehicular paint chips,[24,41,140–145] glasses,[45] ceramics,[63] and bullets.[40,41,45,58,146] Chemical markers obtained from the scenes of arson[61,147] and theft,[148] and dental alginate material[149] have been identified with this technique. Traces of toxic elements (e.g. lead in regular gasoline, arsenic and mercury in contaminated food and ingested materials) are concentrated and the technique applied. The technique has produced evidence of a sulfur pump at the root of scalp hair and has been used for hair comparison based upon increased sulfur intensity along the length of a hair.[27,150] The analysis of other biological inorganic elements such as sodium, phosphorus, potassium, and calcium in cross-sections of scalp hair has been used to compare hair.[151–153] SEM/EDX has been applied to the examination of absorbed copper ions in human head hair.[27,154] This technique is well-suited to examining papers[155] and to identifying dynamite wrapping papers,[156] to solve forensic problems involving handwritten and printed documents,[157,158] and inks.[158,159] SEM/EDX has also been used to study the surface morphology and elemental

composition of the toner deposits on the copying papers and to identify the origins of questioned photocopied documents.[160–163] Minute lipstick smears are sometimes found as evidence on clothing, cigarette butts, and other crime scene objects. SEM/EDX is able to differentiate these and to identify the manufacturer or brand name of lipsticks.[164] The intrinsic ability of SEM/EDX to differentiate particulates based on their sizes and elemental make-ups has been used successfully to identify the origin of soil samples involved in an airplane crash.[165]

4.4 Cathodoluminescence

CL is the emission of visible light under electron impact; it is akin to scintillation emission. The cascade theory suggests that several successive primary and secondary events follow after electron irradiation in organic compounds, which lead to primary π-electron excitation. The latter contributes to CL. Only a small proportion (10% or less) of the energy of incident electrons is expended in the excitation of π-electronic singlet states of unsubstituted aromatic hydrocarbons. The larger proportions of electron energy transfer is expended in ionizations and the σ transitions. De-excitation resulting from the primary excitations (π and σ) create secondary luminescence excited by SEs, X-ray photons, ultraviolet and visible photons. All of these contribute to the final CL emission. The cross-section of CL is determined by the energy of the incident electrons and the target molecules, the generalized oscillator strength of which plays the central role in the processes of energy transfer.[166]

4.4.1 Detection Limits and Ranges

Two types of CL detection systems have been used to determine the detection limits and ranges of CL emission. These are the surface emissive detector and the transmitted (STEM) detector.[10] The volume of sample involved in CL emission is much greater than that of SEs and BSEs. This volume generally approximates the total volume of primary electron diffusion. The STEM system is able to resolve structures up to 50 nm.[166,167] The surface emissive detector offers a much poorer resolution (100 nm).[10]

4.4.2 Applications

The CL studies of fluorochromes show that certain dye compounds such as fluorescein, thioflavine T, brilliant yellow 6G, and ethidium bromide are fairly electron resistant. They exhibit durable CL despite limited molecular damage by the incident electrons. Compounds such as acridine orange, 9-acridanone, and rhodamine-B are sensitive to electrons, and their CL emission is short-lived.[166,167] Giemsa stain is a mixture of eosin-y, methylene blue, and azurs. This stain is not a fluorochrome but it shows intense CL up to 45 min after electron irradiation.[166,167] The potent hallucinogenic drug LSD (lysergic acid diethylamide) is highly cathodoluminescent. The stimulant drug cocaine-HCl is less luminescent than LSD. Time-dependent studies of LSD incorporation into scalp hair show that CL studies of hair could possibly be utilized to determine the history of this drug abuse.[10]

5 SCANNING ELECTRON MICROSCOPE SIGNAL PROCESSING AND APPLICATIONS

The imaging capabilities of SEM/EDX allow both compositional and topographical displays that can be readily interpreted. Illustrative evidence of this type is valuable in forensic comparisons. The ability of SEM/EDX to process each signal by various contrast enhancement methods allows an intuitive, stylistic, and synthetic analysis of the image. Forensic applications of three such methods are stressed in this section. These are X-ray dot mapping, line scanning, and DM or Y-modulation.

5.1 Dot Maps and Line Scans of Energy-dispersive X-rays

Conventional X-ray area scanning or dot mapping is similar to other kinds of SEM imaging – as the electron beam scans the specimen in raster mode, in synchronization with the CRT, the X-ray signal is detected by an EDS or WDS detector. The X-rays produce photoionizations in a liquid-nitrogen-cooled Si(Li) detector. A high voltage bias to the Si(Li) detector generates output pulses of collected electrons proportional to the X-ray energy. These pulses are integrated by a preamplifier, which is a field effect transistor (FET) and generates step signals. These signals are again amplified and given shapes of gaussian pulses by a linear amplifier. These pulses are processed sequentially by an analog-to-digital converter, a multichannel analyzer (MCA) for spectrum accumulation and storage, and a computer for data processing and display of the X-ray spectrum. The output of the analyzer is used to modulate the brightness of the CRT until dots appear on its display screen to specify the location of these X-ray emissions. By selecting an elemental X-ray peak and by adjusting an energy window of the MCA to two-thirds of half-band width of the peak, a dot map of that element is obtained.

Compared to BSE and SE signals, the X-ray signals are weaker by a factor of 10^4–10^6; thus X-ray detection

takes much longer (a millisecond or more) than the detection of electrons (a nanosecond or less). A good practice is to subtract background X-rays and noise and to simultaneously enhance the image signal by using a digital filter before the X-ray signal is fed to the CRT, and then to generate an acceptable dot map by successive exposures at a scan rate of 2 min per frame.[19] The operating conditions of the SEM are critically judged to avoid any significant drift in the dot map.[19] Unless the X-ray background is subtracted, detection limits in the analog dot mapping mode with EDS are 2–5% w/w, whereas for WDS, which has a greater peak to background ratio, the limit is about 0.5–1% w/w.[5] Analog dot mapping is useful but it has artifacts as well as limitations. The long pulse-processing time of EDS leads to dead time effects, and these must be avoided.[19] The dot maps are used for qualitative purposes, not for quantitative information. A large part of this problem results from the way the net scanning time is divided between x and y raster (i.e. scan) speeds. These scan speeds are determined by the x and y fixed resistors of the analog scan generator. In dot mapping in the raster mode, both speeds are equivalent. This often makes the dwelling time of the beam much shorter than in the detection of X-rays. The dots therefore appear sparsely and they are all of the same intensity. Therefore the area density of dots becomes dependent both on the density of the element in question and the geometry of the scanned area in a very complex manner that is not truly dependent on the specimen characteristics.

The chemical make-up of forensic substances can be deciphered by obtaining X-ray maps of the detected elements. The elemental maps are generally complementary to Z-contrast images of high-energy BSEs. Whether the spatial distributions of the detected elements are in successive layers (e.g. paint chips[24]) or are highly discontinuous (i.e. heterogeneous),[142–144] they can be determined by dot mapping. Despite the time involved, this technique is well-suited for the examination of documents[155–159] and fingerprints of metallic inks.[24] The dot mapping technique has been applied to cross-sections of GSR to determine the formation processes of these particles during firearms discharge.[19]

The dot maps are usually examined together with SE and BSE images of the physical evidence to obtain specific information of the constituents. An example is shown in Figure 13 using a hair formula (Grecian Formula®) as evidence. The hair formula was settled as a droplet on a polished carbon planchet as described earlier in Figure 8. The SE image (Figure 13a) showed that the dried sediment contained a mixture of amorphous particles of various sizes, and spongy, thin, disk-like crystals in stacks. The EDX spectrum (Figure 13d) showed that the major elements (>10% w/w) of the mixture were sulfur, chlorine, and lead, with titanium as the minor (>1% w/w) element. The only heavy and toxic element detected was lead and so more information was necessary about this element. The Z-contrast BSE image in Figure 13(b) indicated that these could be the spongy, thin, disk-shaped crystals in the SE image (Figure 13a). Confirmation of this was achieved by dot mapping at the Lα peak of lead (Figure 13c) and/or by recording the X-ray spectrum of these crystals in the spot mode (Figure 13f). A lower magnification of the Z-contrast BSE image in Figure 13(e) showed that these Pb crystals are abundant in the hair formula. The X-ray peaks due to S(Kα), Pb(M), and Cl(Kα) are fairly close to each other and they affect each other by their own continuum. Therefore the Pb(M) line was not considered for the dot mapping of lead. These three elemental peaks are overblown in Figure 13(d) due to adjustment of the vertical scale of X-ray intensity (count rate). This adjustment was necessary in order to show that the Pb(Lα) peak was at or slightly above the two-sigma level (i.e. twice the background). Despite the fact that a digital filter was used to subtract the background, several dots due to the continuum are evident in Figure 13(c). Yet this dot map was preferably used for interpretation.

In analog line scanning the selected window signal is applied to an R–C (resistance–capacitance) ratemeter (e.g. EDAX®-352 module). The ratemeter output is applied to the vertical deflection (y) of the electron beam of the CRT. The x and y resistors of the analog scan generator are selected such that most of the scan time is spent on the x speed and very little on the y speed. The result is a single-line scan along x. Because the y scan is much faster than the x scan, the fast change in count rate produced by the y scan creates R–C distortions. These are due to the lack of correspondence between the time constant of the R–C circuit of the ratemeter, count rate, and scan speed. The major problem is tailing or the masking of changes when the time constant for the R–C circuit is too high. The other problem is due to the high count rate (the dead time effect) when the signal reaches an immediate saturation instead of rising or decreasing in proportion with the count rate. These and other problems of analog line scans are minimized by proper combinations of count rate, slower scan speed, and ratemeter settings (scale factors for integration).[26,27,150] Furthermore, the ratemeter allows smoothing of the line-scanning signal, which removes noise due to fluctuation in background counts.[27,150]

The line scanning technique has the inherent ability to produce useful parameters for comparative hair identification. As sulfur (S) is the major (6% w/w) inorganic element of hair keratin, both perpendicular and longitudinal line scans of S (Kα) serve as a reference for determining the modes of distribution of other elements in hair.[27,154] An SE image of the sample area is

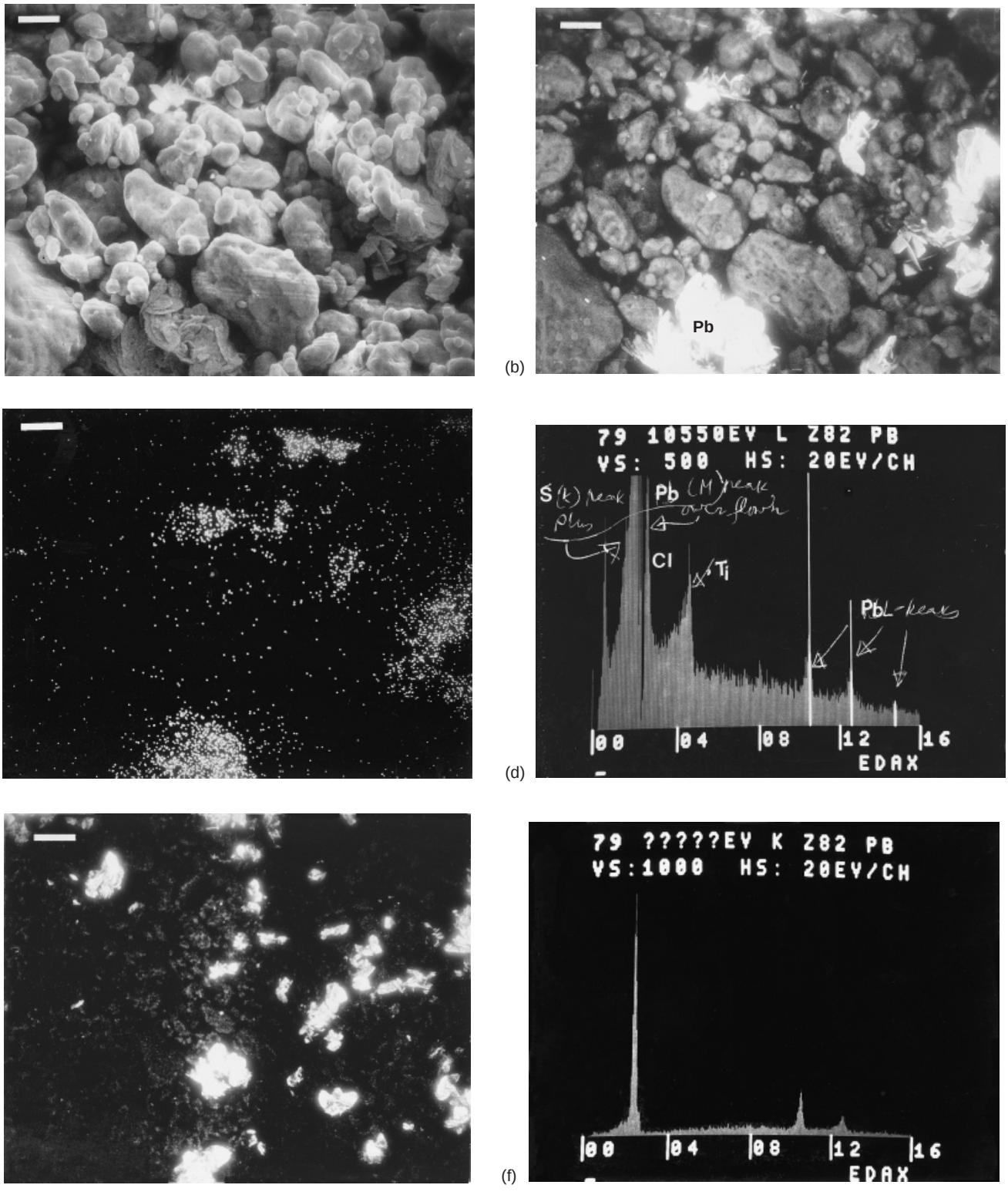

Figure 13 A hair liquid (Grecian Formula®) settled as a droplet on a polished carbon planchet: (a) SE image of the dried sediment, magnification 500×; bar 20 µm; (b) *Z*-contrast BSE image, magnification 500×; bar 20 µm; (c) dot map at Lα peak of detected lead, magnification 500×; bar 20 µm; (d) EDX spectrum showing major (>10% w/w) elements, namely sulfur, chlorine, and lead, and the minor (>1%) element titanium; (e) lower magnification (200×) of the *Z*-contrast BSE image in (a), bar 50 µm; (f) X-ray spectrum of the disk-shaped crystals in (a), 20 kV.

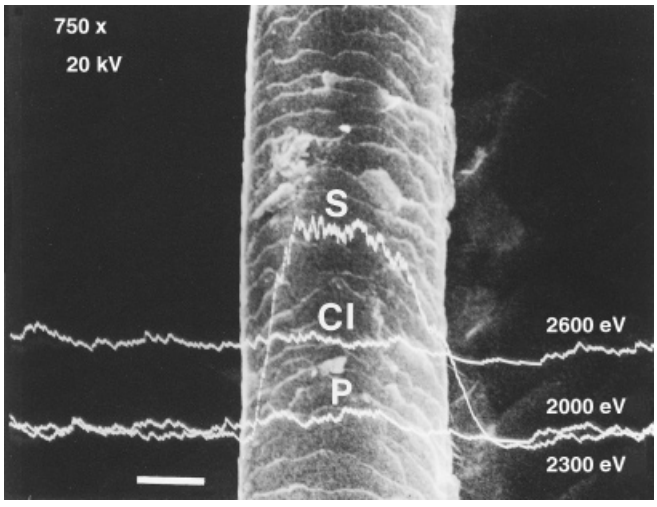

Figure 14 Superimposed SE image and elemental line scans of a human head hair. The dome-shaped S(Kα) line scan shows that concentration-dependent change in intensity of sulfur is modulated by topography of hair. This line scan is marginally affected by the continuum if other elements such as phosphorus and chlorine are absent from the area of examination. Note that the line scans obtained by setting energy windows at 2000 eV (P) and 2600 eV (Cl) are flatter. These are on either side of the S(Kα) line at 2300 eV. (Magnification 750×; bar 20 µm; 20 kV; counts per second full scale 0.5×10^3; scan rate 45 s.)

superimposed over the elemental line scans (Figure 14). These line scans are marginally affected by the continuum if other elements (phosphorus, chlorine, etc.) are absent from the area of examination. An important parameter of hair identification is offered by a sudden rise in the uptake of sulfur within the endoderm portion of the root of hair, which is followed by a gradual increase in sulfur concentration along the hair shaft. The rate of incorporation of sulfur at the root of a hair (the sulfur index) varies remarkably from one individual to another.[27,150] This information has been confirmed by measuring the count rates in the spot mode at the roots of hairs, and then along the length of both anagen and telogen hairs of many individuals (Figure 15).[150]

5.2 Deflection Modulation Imaging with Energy-dispersive X-rays

A novel technique for precisional identification of trace evidence using SEM/EDX is by DM of elemental X-ray signals to generate multiple line scans.[168] This technique requires a potentiometer to vary the *x* speed while the *y* speed (i.e. its resistor) remains fixed. This generates numerous line scans per frame (e.g. 48–100, 110–375, etc.). The X-ray continuum and noise in the image are removed by a threshold setting to the count rate. The threshold setting is achieved by a negative voltage shift of the brightness potentiometer of the CRT on the SEM panel. The recording time as determined mostly by the *y* speed is 3 min 40 s for each elemental image.[168]

This image processing technique accentuates both concentration-dependent and topography-dependent intensity variations. The surface topography selects one or the other. The concentration dependence of the X-ray intensity predominates on smooth flat surfaces, whereas with irregular objects or surfaces, the intensity variations are topography-dependent. These pseudotopographic images of elements are like fingerprints of element distribution in three dimensions in the object. Figure 16 is the lead image of a GSR particle. These particles form by condensation from a state of vapors. The technique offers a nondestructive test for such cases.[168] The concentration-dependent images are able to identify scraping and scratching in violent attacks (e.g. on a watch, watch band, rings, etc.). This technique has generated illustrative evidence of the sulfur index which is characterized by a sudden rise in the intensity of sulfur a short distance away from the root of head hair.[169] This technique is undoubtedly superior to dot mapping[169] and has the potential to solve difficult cases of toolmarks where these are faint or have been obliterated (e.g. firing pin impressions of shotguns and rifles).

5.3 Deflection Modulation Imaging with Secondary Electrons and Backscattered Electrons

The image processing in the SEM by DM allows augmentation of depths and topographic features. DM also recovers some of the flaws which are suppressed in the original SEM signal. A critical evaluation of this method, in comparison with the SE and BSE imaging methods, has been made elsewhere.[24] This technique is suitable for sequencing of intersecting lines of pencils and ball-point pens which produce relatively deep and equal impressions on the paper.[24] The technique has been used to identify individual characteristics of the firing pin impressions of shotguns.[24]

6 FORENSIC AND ENVIRONMENTAL REGULATIONS, METHODS, AND LAWS

6.1 History of the Scanning Electron Microscope and Energy-dispersive X-ray Microanalysis in Court

In admitting GSR evidence examined by SEM/EDX in People v. Palmer (1978),[170] the court in California for the first time noted that the SEM/EDX technique had been uniformly accepted by the scientific community. The accused's objection as to the reliability of the technique was rejected. With the recognition of the potential of the SEM/EDX technique, efforts to overcome its limitations and to improve its credibility have continued. In 1980, in

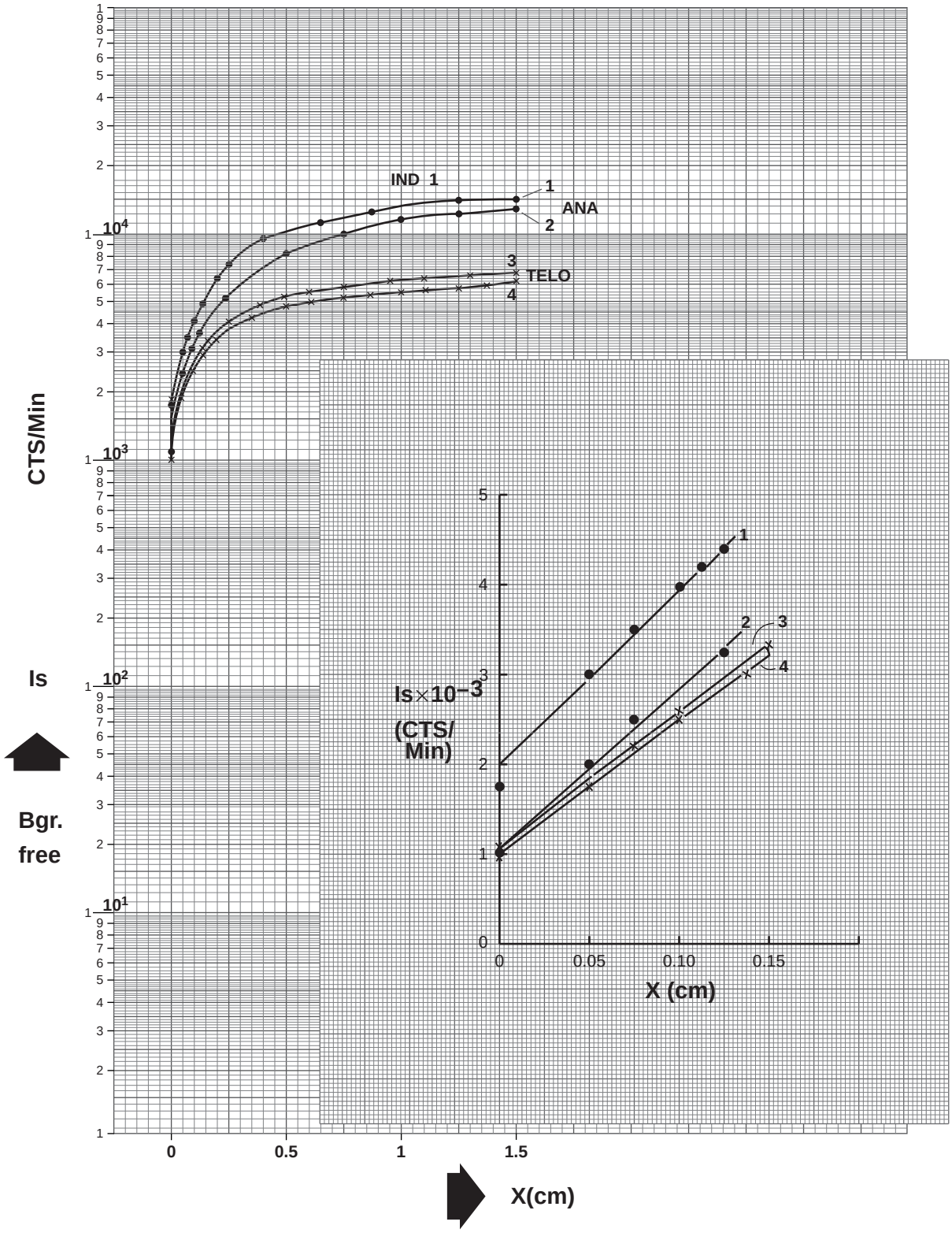

Figure 15 Plots of background-free X-ray intensity (Is) of sulfur (S) versus distance measured from the proximal (root) end of an individual's hair. The count rate was measured in the spot mode at the root of hair and then along length of both anagen (ANA) and telogen (TELO) hair. The rate of incorporation of sulfur at the root of hair (the sulfur index) is almost invariable for different hair of a particular individual (see insert on left) but varies significantly from one individual to another.[150]

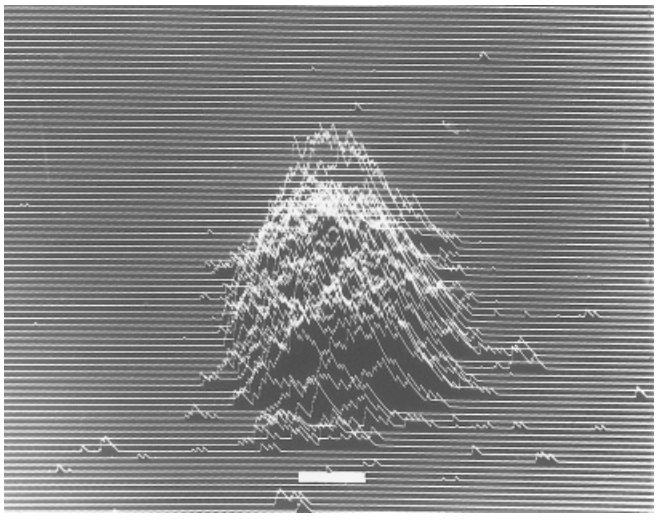

Figure 16 DM image of lead in a GSR particle. Multiple line scans are due to Pb(M) X-rays.[168] (Magnification 10 000×; bar 1 μm; 20 kV.)

People v. Jean Harris, the court of West Chester County, New York, admitted the elemental analysis of bullets and bullet fragments using the SEM/EDX. GSR analysis with the glue-lift has been admitted into evidence by various courts in the state of New York in many cases since 1981, including People v. Patricia Thrush, People v. D.K. Ventiquattro, and D. & T. Wronka's v. Trooper R. Galarza.

6.2 Daubert (1993) v. Frye (1923) Rules

The general acceptance principle, which serves as a criterion for the judicial admissibility of scientific evidence, was set forth on 3 December 1923, by the Court of Appeals of the District of Columbia in the case of Frye v. United States.[171] The trial court had sustained the prosecutor's objection to the admission of lie detector (polygraph) evidence. The appellate court set forth the general acceptance standards by saying that it is difficult to define just when a scientific principle or discovery has crossed the line between the experimental and demonstrable stages. ''Somewhere in this twilight zone the evidential force of the principle must be recognized, and while the courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.'' This Frye rule has been debated in recent years for three reasons. The rule has no safeguard to justify the quality of expert witnesses who may be using the same principle and technique but a variant of methodologies of their own. The rule does not oversee if the scientists openly discuss the pros and cons of accepted practices and the veracity of claims which are unnoticed by the court. The rule enforces a clinging of scientists to accepted principles, barring the entry of new and novel principles which as a matter of fact do not gain widespread support within the scientific community.

In 1993, in a landmark ruling in Daubert v. Merrell Dow Pharmaceutical, Inc.,[172] the US Supreme Court took a different view about the Frye rule by asserting that this rule is not an absolute prerequisite to the admissibility of scientific evidence under the Federal Rules of Evidence.[173] The Court observed that according to Federal Rule 702 (Testimony by Experts) trial judges hold the ultimate responsibility for the admissibility and validity of scientific evidence presented in their courts.

6.3 Federal Rule of Evidence 702 and Expert Witness

Rule 702 defines expert testimony. Under this rule a witness may be an expert by virtue of ''knowledge, skill, experience, training or education''. If they are thus an expert, they may testify to ''scientific, technical, or other specialized knowledge'' in their area, but only if it will ''assist the trier of fact to understand the evidence or determine a fact in issue''.[173] *Daubert*[172] has put the science and technology where they should be. Using the Federal Rule 702 as a perspective this rule has added credibility to reliability standard of *Frye*.[171] In fact the court in *Daubert*[172] has given several guidelines for a judge to qualify an expert witness and the scientific evidence, e.g. whether the technique or theory has been tested and can be tested, whether these have been peer reviewed and published, what the limitations are and the potential rate of error of the technique, whether standards exist and were maintained to control the technique's operation, and whether the theory or method has received widespread acceptance within a particular field in which it belongs. Forensic experts must notice that these legal guidelines set the stage for the subsequent sections of this review.

6.4 Environmental Monitoring and Scope

The users of the SEM must be aware of environmental and occupational hazards associated with the equipment and reagents in the SEM laboratory. They must know the laws and follow proper procedures in all hazardous situations to dispose of dangerous wastes. The environmental laws of using various chemicals and the equipment must be discussed in a training course so that students and trainees can make informed choices of their lines of research. The objective is to protect the environmental health and the associated quality of living. Routine environmental monitoring of the SEM laboratory must cover (1) personal safety and first aid kits, (2) chemical safety as

per OSHA (Occupational Safety and Health Administration (US)) standards, (3) fire safety, (4) electrical safety, (5) physical and mechanical hazards associated with equipment and accessories and consumables, including compressed gas and liquid cylinders, (6) training and orientation programs, and (7) hot lines.[12] Some of the hot lines (telephone numbers) to be kept in the laboratory are Environmental Protection Agency (wastes 800-424-8802; toxins 800-424-9065), National Institute of Occupational Health and Safety (800-356-4674), Consumer Product Safety Commission (800-638-2772), American Chemical Society – Health and Safety Referral Service (202-872-4511), Chemical Transportation Emergency Center (800-424-9300), and Chemical Manufacturers Association (800-262-8200).

On 31 January 1990 the OSHA published a set of rules in the Federal Register for dealing with dangerous chemicals in the workplace. This standard is entitled *Occupational Exposures to Hazardous Chemicals in Laboratories*, and it must be complied with by all science laboratories in the USA. This standard enforces a *Chemical Hygiene Plan* for each laboratory. According to this plan the laboratory must develop a set of *Standard Operating Procedures* that cover general safety precautions while using chemical exposure control measures (fume hoods, respirators, gloves, goggles, etc.), spill control measures, accident responses, and disposal methods to follow upon completion of the procedure. A safety officer must be designated and the employer is required to provide training on the *Chemical Hygiene Plan*. The OSHA standard also applies to gases, vapors, and aerosols created in the laboratory. As the SEM has to run all year in many laboratories, its mechanical pumps must have exhausts through the laboratory into the open air.

6.5 Monitoring of Radiation Hazards

Commercial SEMs are usually well shielded against X-ray leakage, but a new SEM must always be checked for leakage with a Geiger counter for beta and gamma radiation. Leakage greater than $0.5\,\mathrm{mR\,h^{-1}}$ measured at 5 cm from the SEM column is considered significantly hazardous by the Electron Microscopy Society of America (currently Microscopy Society of America) Radiation Committee of 1973. The SEM should be monitored periodically for leakage of X-rays by operating it at the highest voltage with apertures slightly out of alignment and the beam expanded. A well-shielded SEM will register zero counts above the background (25 counts per second) due to cosmic rays. Obviously each operator of the SEM must bear a radiation badge (X-ray film), and these badges must be developed and monitored periodically by an external radiation detection/monitoring service provider.

6.6 Survey of Methods

Just as a regular safety inspection is necessary in order to assure compliance with established rules and to stop negligence before an accident occurs, the surveying of methods used in the SEM laboratory is important as it can save reputations and the cost of litigation. An example of this is the CPD method. The CPD apparatus is often called a bomb because extreme pressures build up inside these units during the drying of biological evidence. The apparatus must be well shielded. Accidents happened with some commercial CPD units in the late 1970s. There are several alternative protocols of CPD that produce the same results, and the overwhelming concensus has shown that the best method is using liquid CO_2.

7 QUALITY CONTROL, STANDARDS AND SOURCES OF ERROR OR TROUBLESHOOTING

7.1 Data Quality Objectives

It is fair to assume that most SEM laboratories have their own priorities or goals. The priority of forensic science laboratories is to generate casework, whereas the priority of a manufacturer's laboratory is to make quality products. These priorities are different due to different strategies. The priorities require implementation of objectives as these can be achieved, measured and evaluated with respect to standards. Fortunately, the SEM is a versatile instrument and is able to meet the demands of quantity and quality. An example of this is elemental analysis. The SEM analyst of a crime laboratory can generate casework by analyzing each item of evidence in the spot mode. This method is considered as reliable as, and much faster than, the image processing methods using dot mapping, line scanning, DM, etc. The SEM analyst in the manufacturer's laboratory, however, will try these contrast enhancement methods to show the quality of a product. The thing that is expected of these two SEM analysts is that each has used a set of controlled operating procedures to secure data, and that such data are evaluated on a routine basis using standard materials. In the words of *Daubert*[172] these data can, therefore, be tested. These objectives which assure quality of data are called data quality objectives. These specific and measurable objectives of the quality of the product or service by an analytical laboratory are the foundation on which the International Organization for Standardization (ISO) 9000 program is built.[174]

7.2 American Society for Testing Materials E-30/National Institute of Standards and Technology Standards

Founded in 1898 and currently headquartered at West Conshohocken, Pennsylvania, the American Society for Testing and Materials (ASTM) has for a century been in the forefront of developing consensus standards.[175] The ASTM standards began with A-1 on steel, stainless steel, and related alloys in 1898 under the initiative of Dr Charles Dudley. The ASTM standard E-30 on forensic science was developed in 1970. The ASTM committee E-30 began actively formulating standards for laboratory procedures and standard practices after a reorganization meeting. This meeting took place in Las Vegas in 1989 in association with the Annual Meeting of the American Academy of Forensic Sciences. Twenty voluntary consensus standards were adopted,[176] including E1588-95 *Standard Guide for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy-Dispersive Spectroscopy*. Developed by ASTM subcommittee E30.01,[177] this guide covers the analysis of GSR by SEM/EDS. The analysis may be performed manually, with the operator manipulating the microscope stage controls and the EDS system software, or in an automated fashion, where some amount of the analysis is controlled by preset software functions requiring little or no operator intervention. As software and hardware formats vary among commercial systems, guidelines were offered in general terms. Three NIST/NIH (National Institutes of Health) softwares for image analysis and microanalysis can serve both as practical tools for solving problems in the SEM/EDX laboratory as well as standards for microanalysts at all levels of experience. These softwares are NIH Image,[178] MacLispix,[179] and NIST Micro MC.[180] These softwares, as well as Desktop Spectrum Analyzer (DTSA) of Chuck Fiori, Carol Swyt-Thomas and Bob Myklebust, are freeware available from a source (CW11) of NIST (http://micro.nist.gov/dtsa/dtsa.html).

7.3 Monitoring of Vibration and Stray Magnetic Fields

Two environmental barriers to high-resolution imaging are mechanical vibration and alternating magnetic field interference. These must be minimal at the time of installation of the SEM and are monitored periodically. The probe size increases drastically with these barriers. Low-frequency vibrations (2–10 Hz) are due to natural vibrations of the building. Motors are sources of a large range of frequencies. Commercial SEM equipment usually comes with an antivibration table under the microscope column. Power supplies (60 Hz or higher) are major sources of stray magnetic fields. The electron beam at low voltages is particularly sensitive to oscillating stray magnetic fields.

7.4 Control or Elimination of Charging Effects

When an electron beam impinges on the specimen, a large proportion of electrons remains in the specimen as they lose all of their initial energy. If the specimen is a conductor and the specimen stage is grounded, this charge flows to ground. If the ground path is disconnected, even the conducting specimen quickly accumulates charge and can deflect the scanning beam off the specimen. Charging is frequently encountered with insulators or specimens containing insulating materials. Charging manifests itself in images in a variety of ways.[5] Nonconducting specimens invariably need coating and conductivity techniques before they can be examined.[5]

7.5 Control or Elimination of the Sources of Spurious X-rays

The brass surfaces of solid-state backscatter detector systems are the main sources of spurious X-rays due to copper and zinc. After removal of this system from the SEM, the solar windows and their leads to a chain of amplifiers are carefully covered with small circular pieces of masking tape. The system is then coated with carbon (thickness $> 10\,\mu m$) by several cycles of carbon evaporation in a vacuum evaporator. The masking tape is removed and the system reinstalled. The collimator window of the X-ray detector (EDS) is also coated with thick evaporated carbon. This avoids spurious X-rays due to aluminum. Commercial-grade carbon paints and inks contain ferromagnetic elements (iron and cobalt). These are not suitable for carbon coating. BSEs reaching those surfaces from the specimen mount are completely absorbed by the thick carbon coating and are not able to produce the spurious X-rays.[27]

7.6 Control of Instrument Performance and Detector Geometry

The control of SEM operating conditions and their optimizations are discussed in sections 4.1.1, 4.2.1, and 4.3.1. The SEM images of cavities, grooves, lines, line-crossings, and faceted objects suffer from shadow effects depending upon the shape and orientation, toward or away from the SE detector. These and other effects of image contrast due to SE detector geometry have been studied elsewhere.[24]

8 COMPARISON WITH OTHER MICROBEAM ANALYTICAL METHODS

8.1 Comparison of Technology

The SEM's ability to generate easily interpretable images of surface topography, crystallography, and material

composition, and the ability to process each specimen signal by various contrast enhancement methods that allow an intuitive, stylistic, and synthetic analysis of the image, has made the SEM one of the most versatile instruments available for microstructural examination and analysis of physical evidence.

Today, it is estimated that well over half of all SEMs are now equipped with X-ray analytical capabilities.[5] In trace evidence analysis involving the examination of GSR, hair, fiber, etc., physical matching and non-destructive elemental analysis of physical evidence, the SEM/EDX is the most efficient of all microbeam technologies. No other microbeam technology combines high resolution with large depth of field for three-dimensional viewing and then allows image processing. The SEM is indeed a research tool in industry and academia.

The digital SEM, or the SEM coupled to a computer to provide a digital signal, now has the capability to process images stored on a magnetic disk and to generate all the desirable effects achievable by the older analog SEM. These systems have allowed automated and unattended search and analysis of forensic, occupational and environmental particulates.[17,110,112,117–121,134] The current imaging trend is to record all data digitally. The advantages of digital acquisition are instant viewing and recording, immediate interpretation and quantification, fast distribution, and storage for further analysis. Digital image processing has affected all microbeam technologies including the SEM. In this regard the SEM is both competitive with, and complementary to, the capabilities of other microbeam technologies.

In the analysis of elemental composition, the SEM's counterparts are the EPMA, Auger electron spectroscopy (AES), and X-ray fluorescence (XRF). In the EPMA the primary radiation of interest is the characteristic X-rays produced by electron bombardment. The microprobes in combination with WDS allow both qualitative and quantitative compositional information from areas of a specimen as small as a micrometer in diameter. Both EPMA and AES have the benefit of superior lateral resolution of elements over EDX. Also, with AES, elements of low mass number (carbon, nitrogen, oxygen, etc.) and elements whose X-ray peaks overlap with the EDX can be quantified. When AES is used in combination with sputtering equipment, depth profiles of the elements can be obtained.[181] XRF is used in forensic laboratories to detect the presence of a wide range of elements present at concentrations of about $10\,\mu g\,g^{-1}$ or greater in trace evidence. The concentrations of most elements needed for characterization of source of trace evidence by SEM/EDX are too low to be detectable.[182] The SEM/EDX is not ideal for quantitative analysis of elements present as traces (<1% w/w).

The SEM is more suitable for examining irregular objects than the scanning tunneling microscope (STM), or the atomic force microscope (AFM). In the STM and AFM the probe interactions become difficult to control and understand when irregular or multiple surfaces are involved. The capabilities of STM and AFM are most utilized with flat surface-like membranes.[183] Some of the best modern SEMs now have spatial resolutions in the nanometer range. These are fairly close to resolutions achieved with a TEM. The SEM lacks the three-dimensional sectioning abilities of scanning beam confocal microscopes. However, the analytical capabilities of the SEM cannot be matched by light microscopes because the resolution is far superior. The principles of image formation and image processing in the SEM are well understood. The SEM is, therefore, a major tool in forensic research and investigation.

ACKNOWLEDGMENTS

Thanks are due to New York State Police Crime Laboratory for implementing the NYSP GSR Program using the glue-lift technique during 1980–89. The participation of many officers from all troops of the New York State Police in the GSR program is thankfully acknowledged. This review began in 1985 at the author's previous address at the New York State Police Crime Laboratory in Albany, New York, and has been completed recently under the auspices of the HBCU Strengthening Program of the United States Department of Education, Contract #PO31B20023, awarded to Albany State University during 1992–98.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscope
ASTM	American Society for Testing and Materials
BSE	Backscattered Electron
CL	Cathodoluminescence
CPD	Critical-point Drying
CRT	Cathode Ray Tube
DM	Deflection Modulation
DTSA	Desktop Spectrum Analyzer
EDS	Energy-dispersive Spectrometry
EDX	Energy-dispersive X-ray Microanalysis
EPMA	Electron Probe Microanalyzer
FBI	Federal Bureau of Investigation
FET	Field Effect Transistor

GSR	Gunshot Residue
IBIS	Integrated Ballistic Identification System®
ISO	International Organization for Standardization
LSD	Lysergic Acid Diethylamide
MCA	Multichannel Analyzer
MDL	Minimum Detection Limit
NAA	Neutron Activation Analysis
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
OSHA	Occupational Safety and Health Administration (US)
PCR	Polymerase Chain Reaction
R–C	Resistance–Capacitance
SE	Secondary Electron
SEM	Scanning Electron Microscope
STEM	Scanning Transmission Electron Microscope
STM	Scanning Tunneling Microscope
TEM	Transmission Electron Microscope
WDS	Wavelength-dispersive Spectrometry
WDX	Wavelength-dispersive X-ray
XRF	X-ray Fluorescence
ZAF	Computer Program that Corrects for the Interelement Effects due to Atomic Number, Absorption, and Fluorescence

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Coatings **(Volume 2)**
Microscopy of Coatings

Environment: Water and Waste **(Volume 3)**
Asbestos Analysis • Detection and Quantification of Environmental Pollutants • Explosives Analysis in the Environment • Heavy Metals Analysis in Seawater and Brines • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring

Food **(Volume 5)**
Particle Size Analysis in Food

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Polymerase Chain Reaction in the Forensic Analysis of DNA

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air • Direct Reading Instruments for the Determination of Aerosols and Particulates

Industrial Hygiene cont'd **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Surface Area and Pore Size Distributions

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction • Pulp and Paper Matrices

Pulp and Paper cont'd **(Volume 10)**
X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in

Surfaces **(Volume 10)**
Surfaces: Introduction • Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • Photoluminescence in Analysis of Surfaces and Interfaces • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Probe Microscopy,

Industrial Applications of • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Absorption and Luminescence Probes • Fluorescence Imaging Microscopy • Fluorescence in Organized Assemblies • Phosphorescence Measurements, Applications of • Surface Measurements using Absorption/Luminescence

Mass Spectrometry **(Volume 13)**
Inorganic Substances, Mass Spectrometric in the Analysis of

Nuclear Methods **(Volume 14)**
Rutherford Backscattering Spectroscopy

REFERENCES

1. P.R. Thornton, *Scanning Electron Microscopy*, Chapman and Hall, London, 1968.
2. O.C. Wells, *Scanning Electron Microscopy*, McGraw-Hill Company, New York, 1974.
3. M.A. Hayat, 'Principles and Techniques of Scanning Electron Microscopy', *Biological Applications*, Van Nostrand Reinhold, New York, 4, 1975.
4. J.I. Goldstein, J.J. Hren, D.C. Joy, *Introduction to Analytical Electron Microscopy*, Plenum Press, New York, 1979.
5. J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, A.D. Romig, Jr, C.E. Lyman, C. Fiori, E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd edition, Plenum Press, New York, 1992.
6. T.R. Allmand, S.N. Jagger, 'Electron Beam X-ray Microanalysis Systems', Cambridge Instruments, England/USA. (Cambridge Instruments, 40 Robert Pitt Drive, Monsey, N.Y. 10952), (Personal Communication).
7. C.A. Anderson, *Microprobe Analysis*, John Wiley and Sons, NY, 1973.
8. L. Reimer, 'Scanning Electron Microscopy–Present State and Trends', *Scanning*, **1**(1), 3–16 (1978).
9. H. Niedrig, 'Physical Background of Electron Backscattering', *Scanning*, **1**(1), 17–34 (1978).
10. S. Basu, 'Cathodoluminescence Studies: From Chromatin and Chromosomes to Human Hair', in *Advances in Microscopy*, eds. R.R. Cowden, F.W. Harrison, Alan R. Liss, Inc., New York, **196**, 265–288, 1985.
11. C.E. Lyman, D.E. Newbury, J.I. Goldstein, D.B. Williams, A.D. Romig, Jr, J.T. Armstrong, P. Echlin, C.E. Fiori, D.C. Joy, E. Lifshin, K.R. Peters, 'Microanalysis and Analytical Electron Microscopy-A Laboratory Workbook, 1st edition, Plenum Press, New York, 1990.
12. J.J. Bozzola, L.D. Russell, *Electron Microscopy: Principles and Techniques for Biologists*, 2nd edition, Jones and BarHett Publishers, Sudbury, MA, 49–71, 203–239, 617–640, 1998.
13. T.F. Anderson, 'Techniques for Preservation of Three Dimensional Structure in Preparing Specimens for the Electron Microscope', *Trans. NY Acad. Sci.*, **13**(1), 130–134 (1951).
14. A.L. Cohen, 'A Critical Look at Critical Point Drying-Theory, Practice and Artifacts', SEM/II, IIT Research Institute, Chicago, IL, 60616, 525–536, 1977.
15. S. Basu, 'Segmentation of Wet-Elongated Mitochondria into Spheres', in *Advances in Microscopy*, eds. R.R. Cowden, F.W. Harrison, Alan R. Liss, Inc., New York, 313–325, **196**, 1985.
16. J.A. Murphy, 'Considerations, Materials and Procedures for Specimen Mounting Prior to Scanning Electron Microscopic Examination', Scanning Electron Microscopy, Part II, SEM Inc., AMF O' Hare, Chicago, IL, 60666, 657–696, 1982.
17. S. Basu, J.R. Millette, *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, 1st edition, Plenum Press, New York, 1986.
18. S. Basu, S. Ferriss, 'A Refined Collection Technique for Rapid Search of Gunshot Residue Particles in the SEM', Scanning Electron Microscopy, Part I, SEM Inc., AMF O'Hare, Chicago, IL, 60666, **392**, 375–384, 1980.
19. S. Basu, 'Formation of Gunshot Residues', *J. Forensic Sci.*, **27**(1), 72–91 (1982).
20. S. Basu, S. Ferriss, R. Horn, 'Suicide Reconstruction by Glue-Lift of Gunshot Residue', *J. Forensic Sci.*, **29**(3), 843–864 (1984).
21. S. Basu, 'The Mechanism of Gunshot Residue Deposition and Its Probing Characteristics to Reconstruct Shootings', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 1–22, 1986.
22. S. Basu, C.E. Boone, D.J. Denio, Jr, R.A. Miazga, 'Fundamental Studies of Gunshot Residue Deposition by Glue-lift', *J. Forensic Sci.*, **42**(4), 571–581 (1997).
23. S. Basu, 'Smoking Gun in Hand: The Glue-Lift Technique', Current Topics in Forensic Science, Proc. of IAFS, 14th Meeting of the International Association of Forensic Sciences, Tokyo, Japan, August 26–30, 1996, 4, 187–190, 1997.
24. S. Basu, 'Forensic Uses of Deflection (Y) Modulation and X-Ray Dot Mapping', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 61–84, 1986.
25. N.C. Barbi, 'Electron Probe Microanalysis Using Energy Dispersive X-ray Spectroscopy', published by Princeton

Gamma-Tech., Box 641, Princeton, NJ 08540 (Personal Communication).
26. J.C. Russ, B.J. Panessa, 'Thin Section Microanalysis', Proceedings of a Symposium, Nov. 18, 1972, St. Louis, Missouri, Publ. EDAX® Lab, 4509 Creedmoor Road, Raleigh, North Carolina 27612 (Personal Communication).
27. S. Basu, 'Spatial Variations of Sulfur and Absorbed Copper Ions in Human Hair by X-Ray Line Scanning', in *Electron Microscopy in Forensic, Occupational, and Environmental Health sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 23–41, 1986.
28. M.E. Taylor, 'SEM in Forensic Sciences', *J. Forensic Sci.*, **13**(4), 296–280 (1973).
29. R.L. Taylor, T.T. Noguchi, 'Applications of SEM and X-ray Microanalysis in Criminalistics: An Indexed Bibliography', Scanning Electron Microscopy, Part II, SEM Inc., AMF O' Hare, Chicago, IL, 60666, 209–216, 1979.
30. R. Pfister, 'The Use of Scanning Electron Microscope and Associated Techniques in Forensic Sciences', Scanning Electron Microscopy, Part III, SEM Inc., AMF O' Hare, Chicago, IL, 60666, 1037–1042, 1982.
31. T.E. Everhart, O.C. Wells, C.W. Oatley, 'Factors affecting Contrast and Resolution in the Scanning Electron Microscope', *J. Electronics and Control*, **7**(1), 97–111 (1959).
32. S. Kimoto, H. Hashimoto, 'On the Contrast and Resolution of the Scanning Electron Microscope', Scanning Electron Microscopy, Proceedings of Symposium, IIT Research Institute, Chicago, IL, 63–78, 1968.
33. V.N.E. Robinson, 'The Origins of the Secondary Electron Signal in Scanning Electron Microscope', *J. Phys. D: Appl. Phys.*, **7**, 2169–2173 (1974).
34. D.A. Moncrieff, P.R. Barker, 'Secondary Electron Emission in the Scanning Electron Microscope', *Scanning*, **1**(3), 195–197 (1978).
35. H.L. MacDonell, L.H. Pruden, 'Applications of the Scanning Electron Microscope to the Examination of Firearms Markings', Scanning Electron Microscopy, Proceedings of a Workshop on Forensic Applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 569–576, 1971.
36. C.A. Grove, G. Judd, R. Horn, 'Examination of Firing Pin Impressions by Scanning Electron Microscopy', *J. Forensic Sci.*, **17**(4), 645–658 (1972).
37. G. Judd, R. Wilson, H. Weiss, 'A Topographical Comparison Imaging System for SEM Applications', Scanning Electron Microscopy, Part I, IIT Research Institute, Chicago, IL, 60616, 167–172, 1973.
38. C.A. Grove, G. Judd, R. Horn, 'Evaluation of SEM Potential in the Examination of Shotgun and Rifle Firing Pin Impressions', *J. Forensic Sci.*, **19**(3), 441–447 (1974).
39. R.P. Singh, 'Examination of Tool Marks in the Scanning Electron Microscope', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 109–115, 1986.
40. M.E. Taylor, 'Forensic Applications of Scanning Electron Microscopy', Proceedings of the Workshop on Forensic Applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 537–544, 1971.
41. J.R. Devanay, L.W. Bradford, 'Applications of Scanning Electron Microscopy to Forensic Science at Jet Propulsion Laboratory', Scanning Electron Microscopy, Proceedings of the Workshop on Forensic Applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 561–568, 1971.
42. P.G. Rodgers, J.R.G. Jacob, P. Blais, D.C. Harris, 'The Scanning Electron Microscopy of Counterfeit Coins', Scanning Electron Microscopy, Proceedings of a Workshop on Forensic Applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 585–586, 1971.
43. S. Basu, 'Evaluation of Staple Impressions in the Scanning Electron Microscope', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 43–60, 1986.
44. H. Katherine, 'Modern Approaches for the Examination of Tool Marks and other Surface Marks', *Forensic Sci. Rev.*, **8**(1), 45–71 (1996).
45. R.L. William, 'An Evaluation of the SEM with X-Ray Microanalyzer Accessory for Forensic Work', Scanning Electron Microscopy, Proceedings of a Workshop on Forensic Applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 537–544, 1971.
46. V.R. Matricardi, M.S. Clark, F.S. DeRonja, 'The Comparison of Broken Surfaces: A Scanning Electron Microscopic Study', *J. Forensic Sci.*, **20**(3), 507–523 (1975).
47. E.H. Sild, S. Pausak, 'Forensic Applications of SEM/EDX', Scanning Electron Microscopy, Part II, SEM Inc., AMF O' Hare, IL, 185–192, 1979.
48. J.I. Thornton, G.T. Mitosinka, 'Comparison of Tungsten Filaments by means of the Scanning Electron Microscope', *J. Forensic Sci. Soc.*, **11**(3), 197–200 (1971).
49. G.L.F. Powell, 'Interpretation of Vehicle Globe Failures: The Unit Condition', *J. Forensic Sci.*, **22**(3), 628–635 (1977).
50. K.A. Thorsen, 'Examination of Bulb Filaments by the Scanning Electron Microscope', *Canadian Soc. Forensic Sci. J.*, **14**(2), 55–69 (1981).
51. K.A. Thorsen, 'Forensic Examination of Bulb Filaments by SEM', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 117–147, 1986.
52. M.Y. Choudhry, C.R. Kingston, L. Kobilinsky, P.R. DeForest, 'Individual Characteristics of Chemically Modified Hairs Revealed by Scanning Electron Microscopy', *J. Forensic Sci.*, **28**(2), 293–306 (1983).

53. J.L. Clement, R. Hagege, A. Le Pareux, J. Connet, G. Gastaldi, 'New Concepts About Hair Identification Revealed by Electron Microscopic Studies', *J. Forensic Sci.*, **26**(3), 447–458 (1981).
54. J.L. Clement, A. Le Pareux, P.F. Ceccaldi, 'The Specifity of the Ultrastructure of Human hair Medulla', *J. Forensic Sci. Soc.*, **22**(4), 396–398 (1982).
55. J.M. Riggott, E.H. Wyatt, 'Mensuration of Scanning Electron Micrographs. A Possible means of Hair Identification', *J. Forensic Sci. Soc.*, **2**, 155–160 (1983).
56. W.R. Pelton, 'Distinguishing the Cause of Textile Fiber Damage Using the Scanning Electron Microscope', *J. Forensic Sci.*, **40**(5), 874–882 (1995).
57. N.J. Clayson, K.G. Wiggins, 'Microfibers – A Forensic Perspective', *J. Forensic Sci.*, **42**(5), 842–845 (1997).
58. C.C. Stacy, H.M. Miller, R.C. Froede, 'Surface Analysis of a Musket Shot Dated Between 1645 and 1655 Found in Historic St. Mary's City, Maryland', *J. Forensic Sci.*, **35**(3), 753–761 (1990).
59. A.L. Morgan, 'Visualization of a Restored Serial Number Using Scanning Electron Microscopy', *J. Forensic Sci.*, **41**(6), 1074–1076 (1996).
60. A. Zeichner, 'Abrasive Cutting of a Safe – A Case Study', *J. Forensic Sci.*, **38**(6), 1516–1522 (1993).
61. S.D. McDermott, 'Metal Particles As Evidence in Criminal Cases', *J. Forensic Sci.*, **39**(6), 1552–1559 (1994).
62. E. Locard, 'Dust and Its Analysis', *The Police Journal*, **1**(1), 177–192 (1928).
63. T. Hashimoto, S. Deki, Y. Kanaji, 'Discrimination of Ceramics–Study of the Microstructures of Ceramics', *J. Forensic Sci.*, **39**(3), 824–838 (1994).
64. K.A. Siegesmund, G.M. Hunter, 'Scanning Electron Microscopy of Selected Crime Laboratory Specimens', Scanning Electron Microscopy, Proc. of a Workshop on Forensic applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 557–584, 1971.
65. P. Pilon, M. Hupe, M. Chauhan, A. Lawrence, 'Development of Test Material for Narcotics Detection Equipment: Sand/ Drug Mixture', *J. Forensic Sci.*, **4**(3), 371–375 (1996).
66. D.A. Moncrieff, P.R. Barker, 'Secondary Electron Emission in the Scanning Electron Microscope', *Scanning*, **1**(3), 195–197 (1978).
67. A. Boyde, 'Cutting Teeth in the SEM', *Scanning*, **1**(3), 157–165 (1978).
68. A. Blueschke, A. Lacis, 'Examination of Line Crossings by Low KV Scanning Electron Microscopy Using Photographic Stereoscopic pairs', *J. Forensic Sci.*, **41**(1), 80 (1996).
69. B. Wilshire, N. Hurley, 'Development of Latent Fingerprints on Paper using Magnetic Flakes', *J. Forensic Sci.*, **40**(5), 838–842 (1995).
70. Y. Migron, D. Mandler, 'Development of Latent Fingerprints on Unfired Cartridges by Palladium Deposition: A Surface Study', *J. Forensic Sci.*, **42**(6), 986–992 (1997).
71. Y. Miogron, G. Hocherman, E. Springer, J. Almog, D. Mandler, 'Visualization of Sebaceous Fingerprints on Fired Cartridge Cases: A Laboratory study', *J. Forensic Sci.*, **43**(3), 543–548 (1998).
72. E. Bohm, 'Application of the SEM in Forensic Medicine', Scanning Electron Microscopy, Proceedings to Workshop on Forensic Applications of the SEM, Part II, IIT Research Institute, Chicago, IL, 60616, 553–560, 1971.
73. E. Bohm, K.H. Hochkirchen, 'Sur Ultrastruktur Vitaler, Postmortaler und Autolysierter Gerinnsel', *Forensic Sci. Int.*, **21**(2), 117–127 (1983).
74. L.C. Boner, M.J. Glimcher, 'Thermal Denaturation of Mineralized Bone Collagens', *J. Ultrastruct Res.*, **32**, 454–461 (1970).
75. P. Shipman, G. Foster, M. Schoeninger, 'Burnt Bones and Teeth and Experimental Study of Color, Morphology, Crystal Structure and Shrinkage', *J. Archael Sci.*, **11**, 307–325 (1084).
76. J.L. Holden, P.P. Phakey, J.G. Clement, 'Scanning Electron Microscope Observations of Heat Treated Human Bone', *Forensic Sci. Int.*, **74**(1–2), 29–45 (1995).
77. G. Quatrehomme, M. Bolla, M. Muller, Jean Paul Rocca, G. Grevin, P. Bailet, A. Ollier, 'Experimental Single Controlled Study of Burned Bones: Contribution of Scanning Electron Microscopy', *J. Forensic Sci.*, **43**(2), 417–422 (1998).
78. L.D.M. Nokes, M. Green, B. Knight, 'The Use of Scanning Electron Microscopy in the Dating of Human Skeletal Remains', *J. Forensic Sci. Soc.*, **27**, 413–416 (1987).
79. J. Wakely, K. Manchester, C. Roberts, 'Scanning Electron Microscope Study of Normal Vertebrae and Ribs from Early Medieval Human Skeleton', *J. Archaeol Sci.*, **16**, 627–642 (1989).
80. M. Yoshino, T. Kimijima, S. Miyasaka, H. Sato, S. Seta, 'Microscopical study on Estimation of Time Since Death in Skeletal Remains', *Forensic Sci. Int.*, **49**, 143–158 (1991).
81. J.H. Bock, D.O. Norris, 'Forensic Botany: An Under-Utilized Resource', *J. Forensic Sci.*, **42**(3), 364–367 (1997).
82. H. Demmelmeyer, 'Foliar Morphology–A Key for the Identification of Small Leaf Fragments', in *Current Topics in Forensic Science*, eds. T. Takatori, A. Takasu, Shunderman Communications, Ottawa, Canada Vol. 4, 56–59, 1997.
83. D.M. Bates, G.J. Anderson, R.D. Lee, 'Forensic Botany: Trichome Evidence', *J. Forensic Sci.*, **42**(3), 380–386 (1997).
84. J.V. Pachar, J.M. Cameron, 'Scanning Electron Microscopy: Application in the Identification of Diatoms in Cases of Drowning', *J. Forensic Sci.*, **37**(3), 860–866 (1992).
85. D.A. Hawley, N.H. Haskell, D.J. McShaffrey, R.E. Williams, J.E. Pless, 'Identification of a Red fiber: Chironomid Larvae', *J. Forensic Sci.*, **34**(3), 617–621 (1989).

86. J.G. Rozen, Jr, G.C. Eickwort, 'The Entomological Evidence', *J. Forensic Sci.*, **42**(3), 394–397 (1997).
87. K.A. Berube, T.P.Jones, B.J. Williamson, 'Electron Microscopy of Urban Airborne Particulate Matter', *Microscopy and Analysis*, USA Edition, **26**, 11–13 (1997).
88. C. Torre, S. Gino, 'Epidermal Cells on Stubs Used for Detection of GSR with SEM-EDX: Analysis of DNA Polymorphisms', *J. Forensic Sci.*, **41**(4), 658–659 (1996).
89. 'FBI Drugfire Program: A Research and Development–Potential Sources Sought', *Commerce Business Daily*, Issue # PSA 0268, Mon. 28 Jan 1991.
90. T.J. David, 'The Use of Scanning Electron Microscopy in Bitemark Analysis', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 85–95, 1986.
91. S. Kimoto, H. Hashimoto, 'On a Scanning Electron Microscope', 6th International Congress of Electron Microscopy, Kyoto, 197–198, 1966.
92. V.N.E. Robinson, 'Imaging with Backscattered Electrons in a Scanning Electron Microscope', *Scanning*, **3**(1), 15–26 (1980).
93. O.C. Wells, 'Bibliography on the Backscattered Electron Image in the Scanning Electron Microscope', Scanning Electron Microscopy, Part I, Proc. of a Symposium, IIT Research Institute, Chicago, IL, 60616, 747–771, 1977.
94. L. Reimer, C. Tollkamp, 'Measuring the Backscattered Coefficient and Secondary Electron Yield inside a Scanning Electron Microscope', *Scanning*, **3**(1), 35–39 (1980).
95. V.N.E. Robinson, 'The Construction and Uses of an Efficient Backscattered Electron Detector for Scanning Electron Microscopy', *J. Phys. E: Sci. Instrum.*, **7**, 650–652 (1974).
96. V.R. Matricardi, 'The Application of Backscattered Electron Image to Forensic Science', Scanning Electron Microscopy, Part II, SEM Inc., AMF O' Hare, IL, 159–165, 1979.
97. M. Oron, V. Tamir, 'Development of SEM Methods for Solving Forensic Problems Encountered in Handwritten and Printed Documents', Scanning Electron Microscopy, Proceedings of a Symposium, Part I, IIT Research Institute, Chicago, IL, 60616, 207–214, 1974.
98. R.D. Koons, 'Sequencing of Intersecting Lines by Combined Lifting Process and Scanning Electron Microscopy', *Forensic Sci. Int.*, **27**(2), 261–268 (1985).
99. D.E. Newbury, 'Microanalysis in the Scanning Electron Microscope: Progress and Prospects', Scanning Electron Microscopy, Part II, SEM Inc., AMF O' Hare, IL, 1–20, 1979.
100. J.E. Wessel, P.F. Jones, W.Y. Kwan, R.S. Nesbitt, E.J. Rattin, 'Gunshot Residue Detection: Survey and Assessment and Identification of Alternative Concepts', Report ATR-75 (7915)-1, The Aerospace Corporation, El Segundo, CA, 161–176, 1976.
101. R.S. Nesbitt, J.E. Wessel, P.F. Jones, 'Detection of Gunshot Residue by use of the Scanning Electron Microscope', *J. Forensic Sci.*, **21**(3), 595–610 (1976).
102. J. Andrasco, A.C. Maehly, 'Detection of Gunshot Residues on Hands by Scanning Electron Microscopy', *J. Forensic Sci.*, **22**(2), 279–287 (1977).
103. V.R. Matricardi, J.W. Kilty, 'Detection of Gunshot Residue Particles from the Hands of a Shooter', *J. Forensic Sci.*, **22**(4), 725–738 (1977).
104. G.M. Wolten, R.S. Nesbitt, A.R. Calloway, G.L. Loper, P.F. Jones, 'Final report on Particle Analysis for Gunshot Residue Detection', Report ATR-77 (7915)-3, The Aerospace Corporation, El Segundo, CA, 1–196, 1977.
105. M. Tassa, N. Zeldes, 'Characterization of Gunshot Residue Particles by Localizations of their Chemical Constituents', Scanning Electron Microscopy, Part II, SEM, Inc., AMF O' Hare, IL, 175–178, 1979.
106. J.S. Wallace, R.H. Keeley, 'A Method for Preparing Firearms Residue Samples for Scanning Electron Microscopy', Scanning Electron Microscopy, Part II, SEM Inc., AMF O' Hare, IL, 179–184, 1979.
107. G.M. Wolten, R.S. Nesbitt, A.R. Calloway, G.L. Loper, P.F. Jones, 'Particle Analysis for the Detection of Gunshot Residue. I: Scanning Electron Microscopy/ Energy Dispersive X-Ray Characterization of Hand Deposits from Firing', *J. Forensic Sci.*, **24**(2), 409–422 (1979).
108. G.M. Wolten, R.S. Nesbitt, A.R. Calloway, G.L. Loper, 'Particle Analysis for the Detection of Gunshot Residue. II. Occupational and Environmental Particles', *J. Forensic Sci.*, **24**(2), 423–430 (1979).
109. G.M. Wolten, R.S. Nesbitt, A.R. Calloway, 'Particle Analysis for the Detection of Gunshot Residue. III: The Case Record', *J. Forensic Sci.*, **24**(4), 864–869 (1979).
110. M. Ueyama, R.L. Taylor, T.T. Noguchi, 'SEM/EDS Analysis of Muzzle Deposits at Different Target Distances', Scanning Electron Microscopy, Part I, SEM, Inc., AMF O' Hare, IL, 367–374, 1980.
111. G.M. Wolten, R.S. Nesbitt, 'On the Mechanism of Gunshot Residue Particle Formation', *J. Forensic Sci.*, **25**(3), 533–545 (1980).
112. R.H. Keeley, W.B. Drayton, 'Automatic Particle Analysis with the SEM', Abstract, 9th Meeting of the IAFS, 22–26 June, 1981, Bergen, Norway, *J. Forensic Sci. Soc.*, **21**(2), 115–116 (1981).
113. H. Gansau, U. Becker, 'Semi- Automatic Detection of Gunshot Residue by SEM- EDX', Scanning Electron Microscopy, Part I, SEM Inc., AMF O' Hare, IL, 107–114, 1982.
114. D.E. Ward, 'Gunshot Residue Collection for Scanning Electron Microscopy', Scanning Electron Microscopy, Part III, SEM Inc., AMF O' Hare, IL, 60161, 1031–1036, 1982.
115. H.A. Wrobel, J.J. Millar, Marek Kijek, 'Comparison of Properties of Adhesive Tapes, Tabs and Liquids used for the Collection of Gunshot Residue and other Trace

Materials for SEM Analysis', *J. Forensic Sci.*, **43**(1), 178–181 (1998).
116. L.A. Sugerman, 'The Concentration and Isolation of Gunshot Residues for Particle Analysis', *J. Forensic Sci. Soc.*, **27**, 135 (1987).
117. A. Zeichner, N. Levin, 'Casework Experience of GSR Detection in Israel, on Samples from Hands, Hair, and Clothing Using an Autosearch SEM/EDX System', *J. Forensic Sci.*, **40**(6), 1082–1085 (1995).
118. T.G. Kee, C. Beck, 'Casework assessment of an Automated Detection and Analysis by Scanning Electron Microscope/Microanalysis System for the Detection of Firearms Discharge Particles', *J. Forensic Sci. Soc.*, **27**, 321–330 (1987).
119. J. Tillman, 'Automated Gunshot Residue Particle Search and Characterization', *J. Forensic Sci.*, **32**(1), 62–71 (1987).
120. R.S. White, A.D. Owens, 'Automation of Gunshot Residue Detection and Analysis by Scanning Electron Microscopy/Energy Dispersive X-ray Analysis', *J. Forensic Sci.*, **32**(6), 1595–1603 (1987).
121. M.S. Germani, 'Evaluation of Instrumental Parameters for Automated Scanning Electron Microscopy/Gunshot Residue Particle Analysis', *J. Forensic Sci.*, **36**(2), 331–342 (1991).
122. A. Zeichner, N. Levin, 'Collection Efficiency of Gunshot Residue Particles from Hair and Hands using Double Side Adhesive Tape', *J. Forensic Sci.*, **38**(3), 571–584 (1993).
123. D.M. Gialamas, E.F. Rhodes, L.A. Sugarman, 'Officers, Their Weapons and Their Hands: An Emperical Study of GSR on the Hands of Non-shooting Police Officers', *J. Forensic Sci.*, **40**(6), 1086–1089 (1995).
124. J.S. Wallace, J. McQuillan, 'Discharge Residues from Cartridge - Operated Industrial Tools', *J. Forensic Sci.*, **24**, 495–508 (1984).
125. B. Burnett, 'The Form of Gunshot Residue is Modified by Target Impact', *J. Forensic Sci.*, **34**(4), 808–822 (1989).
126. D. DeGaetano, J.A. Siegal, K.L. Klomparens, 'A Comparison of Three Techniques Developed for Sampling and Analysis of Gunshot Residue by Scanning Electron Microscopy/Energy Dispersive X-Ray Analysis', *J. Forensic Sci.*, **37**(1), 281–300 (1992).
127. H.C. Harrison, R. Gilroy, 'Firearms Discharge Residues', *J. Forensic Sci.*, **4**, 184–190 (1959).
128. G. Price, 'Firearm discharge Residues on Hands', *J. Forensic Sci. Soc.*, **5**, 199–200 (1965).
129. A.D. Owens, 'A Reevaluation of the Aerospace Corporation Final Report on Particle Analysis–When to Stop Searching for Gunshot Residue?', *J. Forensic Sci.*, **35**(3), 698–705 (1990).
130. L. Gunaratnam, K. Himberg, 'The Identification of Gunshot Residue Particles from Lead-Free Sintox Ammunition', *J. Forensic Sci.*, **39**(2), 532–536 (1994).
131. A. Harris, 'Analysis of Primer Residue from CCI Blazer Lead Free Ammunition by Scanning Electron Microscopy/Energy Dispersive X-ray', *J. Forensic Sci.*, **40**(1), 27–30 (1995).
132. A. Zeichner, N. Levin, E. Springer, 'Gunshot Residue Particles Formed by using Different Types of Ammunition in the Same Firearm', *J. Forensic Sci.*, **36**(4), 1020–1026, 1991.
133. A. Zeichner, B. Schecter, R. Brener, 'Antimony Enrichment on the Bullet's Surfaces and the Possibility of Finding it in Gunshot Residue of the Ammunition Having Antimony-Free Primers', *J. Forensic Sci.*, **43**(3), 493–501 (1998).
134. F.E. Doern, D.L. Wotton, 'Microanalysis of Airborne Lead Particles in an Urban Industrial Environment', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 237–252, 1986.
135. J.S. Webber, 'Using Microparticle Characterization to Identify Sources of Acid-Rain Precursors Reaching Whiteface Mountain, New York', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Millette, Plenum Press, New York, 261–273, 1986.
136. J.S. Wallace, W.J. McKeown, 'Sampling Procedures for Firearms and/or Explosive Residues', *J. Forensic Sci.*, **33**, 107–116 (1993).
137. S. Basu, 'Gunshot Residue Analysis in Bullet Wounds with the SEM-EDX', in *Microbeam analysis–1986*, eds. A.D. Romig, Jr, W.F. Chambers, San Francisco Press, CA, 461–462, 1986.
138. B.R. Burnett, 'Detection of Bone and Bone Plus Bullet Particles in Backspatter from Close-Range Shots to Heads', *J. Forensic Sci.*, **36**(6), 1745–1752 (1991).
139. A. Mauro, G. Falso, 'Identification of Ammunition used in a Lethal Robbery, Comparison between Scanning Electron Microscopy/Energy Dispersive X-ray Analysis and Instrumental Neutron Activation Analysis (INAA) Measurements', *J. Forensic Sci.*, **38**, 1237–1242 (1993).
140. E.M. Butler, 'The Examination of the Paint with Electron Microscope', *J. Criminal Law, Criminol. and Police Sci.*, **58**(4), 596–602 (1967).
141. H.R. MacQueen, G. Judd, S. Ferriss, 'The Application of Scanning Electron Microscopy to the Forensic Evaluation of Vehicular Paint Samples', *J. Forensic Sci.*, **17**(4), 659–667 (1972).
142. R. Wilson, G. Judd, S. Ferriss, 'Characterization of Paint Fragments by Combined Topographical and Chemical Electron Optics Techniques', *J. Forensic Sci.*, **19**(2), 363–371 (1974).
143. B. Collins, 'Micro-Techniques on the Examination of Problems in Surface Coatings', *Australian Oil Color Chemists Assoc. Proc. News*, **13**(1–2), 6–16 (1976).
144. R.H. Meinhold, R.M. Sharp, 'The Application of a Multistyles Recorder to the Energy Dispersive Analysis

of Paint Flakes in the Scanning Electron Microscope', *J. Forensic Sci.*, **22**(2), 274–282 (1977).
145. M.E. Anderson, 'Identification of Pigments of Artistic, Forensic and Industrial Importance by the Raman Microprobe and Scanning Electron Microscopy with Energy Dispersive X- Ray Analysis', in *Microbeam Analysis*, ed. K.F.J. Heinrich, San Francisco Press, CA, 197–201, 1982.
146. R.L. Taylor, M.S. Taylor, T.T. Noguchi, 'Firearm Identification by Examination of Bullet Fragments: An SEM/EDS Study', Scanning Electron Microscopy, Part II, SEM, Inc., AMF O'Hare, IL, 167–174, 1979.
147. E.L. Covey, 'Application of Energy Dispersive X-ray Spectroscopy in Fire Investigation', *J. Forensic Sci.*, **21**, 325–328 (1976).
148. R. Goebel, 'Identification of Chemical Markers by Application of the Scanning Electron Microscope in Cases of Theft', Scanning Electron Microscopy, Part I, SEM, Inc., AMF O' Hare, IL, 60666, 141–146, 1982.
149. L. Gentleman, R.F. Carr, E.E. Herschaft, 'Identification of a Foreign Body using Energy Dispersive X-ray Analysis', *J. Forensic Sci.*, **26**(1), 58–63 (1981).
150. S. Basu, 'Hair Comparison in the SEM-EDX Based upon Increased Sulfur Intensity along the Length of Hair: Evidence of a Sulfur Pump', in *Microbeam analysis–1986*, eds. A.D. Romig, Jr, W.F. Chambers, San Francisco Press, CA, 457–460, 1986.
151. S. Seta, H. Sato, M. Yoshino, 'Quantitative Investigation of Sulfur and Chlorine in Human Head Hairs by Energy Dispersive X-ray Microanalysis', Scanning Electron Microscopy, Part II, SEM, Inc., AMF O'Hare, IL, 193–202, 1979.
152. S. Seta, H. Sato, M. Yoshino, S. Miyasaki, 'SEM/EDX Analysis of Inorganic Scalp Hairs with Special Reference to the Variation with Different Locations on the Head', Scanning Electron Microscopy, Part I, SEM, Inc., AMF O' Hare, IL, 60666, 127–140, 1982.
153. S. Seta, 'Analysis of Biological Elements in Scalp Hairs with SEM-EDX and TEM and its Application to the Hair Comparison', in *Electron Microscopy in Forensic, Occupational, and Environmental Health Sciences*, eds. S. Basu, J.R. Milette, Plenum Press, New York, 97–108, 1986.
154. S. Basu, S. Alexander, 'The Capability of the SEM-EDX to Detect Absorbed Cu-ions into Human Hair: A New Application of Line Scanning', Ninth Annual Meeting of Northeastern Association of Forensic Scientists, Sheraton Heights, Hasbrouck, NJ, Oct., 7–8, 1983.
155. R.A. Parham, 'X-ray/SEM Analysis of Materials: Principles and Applications to Paper', *Paperi Ja Puu*, **12**, 1–9 (1973).
156. M. Ueyama, I. Suzuki, R.L. Taylor, T.T. Noguchi, 'The Characterization of Dynamite Wrapping Paper in SEM', Scanning Electron Microscopy, Part II, SEM, Inc., AMF O' Hare, IL, 203–208, 1979.
157. M. Oron, V. Tamir, 'Development of SEM Methods for Solving Forensic Problems Encountered in Handwritten and Printed Documents', Scanning Electron Microscopy, Proceedings of a Symposium , Part I, IIT Research Institute, Chicago, IL, 60616, 207–214, 1974.
158. P.J. Nolan, M. England, C. Davis, 'The Examination of Documents by Scanning Electron Microscopy and X-ray Spectrometry', Scanning Electron Microscopy, Part II, SEM, Inc., AMF O'Hare, Chicago, IL, 60666, 599–610, 1982.
159. M.F. Cassidy, D.E. Allen, 'Application of the Microdroplet Method of X-ray Analysis to the Characterization of Inks', *J. Forensic Sci.*, **38**(1), 40–47 (1993).
160. R.N. Totty, B. Baxendale, 'Defect Marks and Identification of Photocopy Toners', *J. Forensic Sci.*, **21**(23), 23–30 (1981).
161. G.P. Warman, C.J. Curry, P. Burke, 'Evaluation of SEM with Energy Dispersive Spectrometry for Examining Plain Paper Photocopies', 9th International Association of Forensic Science Meeting, Oxford, UK, Sept., 1984.
162. D.A. Crown, 'The Differential of Electrostatic Photocopy Machines', *J. Forensic Sci.*, **34**(1), 142–155 (1989).
163. R.N. Totty, 'Analysis and Differentiation of Photocopy Toners', *Forensic Sci. Rev.*, **2**(1), 1–23 (1990).
164. M.Y. Choudhry, 'Comparison of Minute Smears of Lipstick by Microspectrophotometry and Scanning Electron Microscopy/Energy Dispersive Spectroscopy', *J. Forensic Sci.*, **36**(2), 366–375 (1991).
165. L.A. Daugherty, 'Soil Science Contribution to an Airplane Crash Investigation, Ruidodo, New Mexico', *J. Forensic Sci.*, **42**(3), 401–405 (1997).
166. S. Basu, 'Theoretical Estimates of Cathodoluminescence of Giemsa Stained Chromosomes', *J. Theor. Biol.*, **85**, 125–151 (1980).
167. S. Basu, 'Detection of Cathodoluminescence of Giemsa Stain and its Applications', *Rev. Sci. Instrum.*, **51**(4), 435–439 (1980).
168. S. Basu, 'Deflection Modulation Imaging of Elements in the Scanning Electron Microscope with Energy Dispersive X-rays', *J. Comput. Assisted Microscopy*, **2**(1), 25–45 (1990).
169. S. Basu, 'Elemental X-ray Imaging of Forensics by a Multiple Line Scanning Technique in the SEM-EDX', Current Topics in Forensic Science, Proc. of IAFS, 14th Meeting of the International Association of Forensic Sciences, Tokyo, Japan, Aug. 26–30, 1996, Vol. 14, 43–49, 1997.
170. People vs. Palmer (1978, 1st. Dist.) 80 cal App 3d 239; 145 Cal Rptr 466, 1 ALR 4th 1056.
171. Frye vs. United States 293 Fed. 1013, 1923.
172. Daubert vs. Merrell Dow Pharmaceutical, Inc., 113 S. Ct. 2786, 1993.
173. Federal Rules of Evidence: Rules of Evidence for the United States Courts and Magistrates, Practice

Comments by Paul F. Rothstein, Clark Boardman Company, Ltd., New York, Rules, 702–705, 281–294, 1973.
174. L.A. Heimbrook, 'Implementation of ISO 9000 Standards in an Analytical Laboratory', *Microbeam Anal.*, **4**(4), 209–214 (1995).
175. ASTM, 'A Century of Progress/ASTM: 1898-1998', *Standardization News*, **26**(1), 25–57 (1998).
176. J.J. Lentini, 'ASTM standards for Forensic Sciences', *J. Forensic Sci.*, **40**(1), 146–149 (1995).
177. E1588-95 Standard Guide for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy-Dispersive Spectrometry; http://www.astm.org.
178. W.S. Rasband, D.S. Bright, 'NIH Image: A Public Domain Image Processing Program for the Macintosh', *Microbeam Anal.*, **4**(4), 137–149 (1995).
179. D.S. Bright, 'MacLispix: A Special Purpose Public Domain Image Analysis Program for the Macintosh', *Microbeam Anal.*, **4**(4), 151–163 (1995).
180. D.E. Newbury, R.L. Myklebust, 'NIST Micro MC: A User's Guide to the NIST Microanalysis Monte Carlo Electron Trajectory Simulation Program', *Microbeam Anal.*, **4**(4), 165–175 (1995).
181. G. Hellmiss, W. Lichtenberg, M. Weiss, 'Investigation of Gunshot Residues by Means of Auger Electron Spectroscopy', *J. Forensic Sci.*, **32**(3), 747–760 (1987).
182. R.D. Koons, 'Comparison of Individual Carpet Fibers Using Energy Dispersive X-ray Fluorescence', *J. Forensic Sci.*, **41**(2), 199–205 (1996).
183. W.R. Bowen, 'Atomic Force Microscopy as a Tool for the Membrane Technologist', *The Americas Microscopy and Analysis*, November (33), 7–9 (1998).

X-ray Fluorescence in Forensic Science

Claude Roux
University of Technology, Sydney, Australia

Chris Lennard
Australian Federal Police, Canberra, Australia

1 Introduction 1
2 Elemental Analysis and X-ray Fluorescence in Forensic Protocols 1
3 Technology 2
3.1 Conventional X-ray Fluorescence Spectrometry 2
3.2 Electron Probe Microanalysis 2
3.3 X-ray Fluorescence Microanalysis 3
3.4 Total Reflection X-ray Fluorescence Spectrometry 3
4 Forensic Applications 3
4.1 Metals 3
4.2 Gunshot Residues 4
4.3 Paint 5
4.4 Glass and Soil 5
4.5 Fibers 6
4.6 Plastics and General Polymers 6
4.7 Miscellaneous 8
5 Discussion and Conclusions 8
Acknowledgments 9
Abbreviations and Acronyms 9
Related Articles 9
References 9

This article describes the applications of X-ray fluorescence (XRF) as an analytical tool in forensic science, including limitations and practical aspects. The use of XRF is specifically discussed for the forensic analysis of materials of forensic interest such as metals, gunshot residues (GSR), paint, glass, soil, fibers, plastic and general polymers, and miscellaneous types of evidence. Recent developments such as micro XRF and total XRF are also described. These specialized techniques have highly desirable characteristics for the elemental profiling of a wide range of forensic samples, and hence it can be anticipated that they will find increasing use in forensic laboratories in the future.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

XRF techniques are based upon the emission of characteristic X-rays from the elements in the sample due to their excitation by absorption of a primary beam of X-rays. These techniques can perform rapid and nondestructive qualitative multielemental analyses of barely visible samples, which constitutes a major advantage in a forensic context. Modern XRF instruments can also produce quantitative analyses of complex materials. However, most XRF techniques are not as sensitive as other methods, such as inductively coupled plasma mass spectrometry (ICP/MS), for example.

This article discusses the application of XRF techniques in forensic science, including limitations and practical aspects. For completeness, electron probe microanalysis (EPMA) or scanning electron microscopy/energy-dispersive X-ray spectrometry (SEM/EDX) is also considered, although in this case the X-ray spectrum is induced by an absorption of electrons instead of a primary beam of X-rays.

2 ELEMENTAL ANALYSIS AND X-RAY FLUORESCENCE IN FORENSIC PROTOCOLS

Most forensic examinations deal with the identification of a sample, the comparison of samples coming from different sources, and the interpretation of the results obtained in the context of the events. Forensic science is largely based on a comparison process. The aim of this process is to attain the highest degree of discrimination between very similar samples. In a particular case, choosing appropriate techniques from the wide range available will depend on several factors including the nature (primarily organic versus inorganic matrix) and the form of the sample, the quantity of material available, the exact circumstances of the case under investigation, and the instrumental techniques available to the laboratory. It is good forensic practice to follow a stepwise progression from simple, general, rapid, widely applicable, and nondestructive screening methods to more discriminating, specific and sometimes destructive methods. The comparison process considers morphological, physical and chemical characteristics. In modern forensic laboratories these characteristics are largely determined by microscopic examinations, color/dye characterization using microspectrophotometry and/or chromatography (thin-layer chromatography or high-performance liquid chromatography), and other instrumental techniques. In this context, it becomes obvious that techniques yielding elemental information are generally only used during

the later stages of a forensic examination. In addition, in some cases there might be no need for an elemental analysis because enough discriminative information has been obtained by simpler tests.

Nevertheless, in many cases it may be important to characterize the sample not only in terms of the organic matrix, but also in terms of the elemental profile. The elemental analysis is obvious for samples which are in essence inorganic, such as metals and glass, or for samples with a significant inorganic content, such as GSR, soil, or paint containing inorganic pigments. This type of analysis can also be useful to compare samples which are in essence organic, such as plastics/general polymers or fibers on the basis of their trace element profile. In some cases, this information can lead to samples being differentiated which would have remained otherwise indistinguishable.

XRF techniques can be used whenever there is a need for elemental profile characterization. The main advantage of these methods over other elemental techniques such as atomic absorption spectrometry (AAS) or ICP/MS is their nondestructive nature, which is a highly desirable feature in forensic science. The ability of these methods to analyze small solid samples with accuracy, precision and speed constitutes another advantage in the forensic context. The spectra are also relatively simple to interpret. Disadvantages include a relatively mediocre sensitivity (in the ppm range in the best cases), inconveniences associated with quantitative analyses, and difficulties in detecting and measuring light elements.

3 TECHNOLOGY

The common XRF techniques used in forensic science include:

- conventional XRF spectrometry;
- EPMA (also known as SEM/EDX or scanning electron microscope (SEM)/EDS);
- X-ray fluorescence microanalysis (XRFM or micro-XRF);
- total reflection X-ray fluorescence spectrometry (TXRF).

Key factors to consider when determining which technique will best serve the analytical needs of the forensic scientist include:

- What is the sample size and/or feature size to be analyzed?
- What elemental sensitivity and/or compositional variations need to be resolved?
- What degree of sample preparation is acceptable or required?
- What analytical environment is acceptable: air, helium, or vacuum?
- What are typical sample sizes and manipulation requirements?
- Is the sample physically and chemically sensitive to beam damage and/or contamination?
- What is the sample analysis turnaround time?
- Is automation important?

Due to the size of typical forensic samples, and the need for nondestructive analyses, the forensic scientist will generally favor techniques with a good spatial resolution, and those which minimize sample preparation. In some cases, automation and/or elemental sensitivity can also be key features. In summary, an ideal XRF technique for forensic applications should combine good spatial resolution, elemental sensitivity, easy sample preparation, and automation capabilities.

3.1 Conventional X-ray Fluorescence Spectrometry

Conventional XRF has good elemental sensitivity and is widely used for the qualitative analysis of elements with an atomic number greater than oxygen (i.e. >8). It can also be used for semiquantitative or quantitative analyses. Two types of spectrometers exist, wavelength-dispersive and energy-dispersive instruments. In the first case, the separation of the polychromatic XRF is achieved by dispersion through a rotating crystal, whereas in the latter the different energies composing the spectrum are separated by a multichannel pulse-height analyzer. It is beyond the scope of this article to discuss in detail the difference between the two systems. Suffice to say that energy-dispersive spectrometers are best suited to forensic applications because of their increased sensitivity and their improved signal-to-noise ratio.

The sample can be analyzed in various forms including solids, powders, evaporated films, and solutions. This renders conventional XRF suitable for analyzing a range of forensic samples (e.g. pieces of metal, large glass fragments, soil samples, etc.). However, conventional XRF is a macroscopic technique, and does not have good spatial resolution. It is therefore not applicable to the analysis of minute samples often encountered in forensic casework, such as fibers or glass fragments. As a result, forensic scientists generally prefer to use one of the specialized techniques described below.

3.2 Electron Probe Microanalysis

EPMA differs from the other techniques described in this article in terms of the type of interaction inducing the X-ray spectrum. In EPMA, the X-rays characteristic of the specimen are induced by a beam of electrons instead of primary X-rays. In the forensic context, the

primary advantage of EPMA is its strong link with the imaging of the SEM. Coupled with the SEM, and thanks to its excellent spatial resolution (of the order of 1 μm), this technique can characterize minute particles by observing the morphology and size of the particle in addition to carrying out the elemental analysis on a specific area of the specimen. For example, it can be used for the analysis of a microscopic glass bead deposited on a headlight filament, or a single gunshot residue particle. Compositional mapping showing the elemental distribution in the area of interest is also possible, which can have some application in the analysis of automotive paints, for example.

However, the relatively poor sensitivity of EPMA due to the high X-ray continuum background (Brehmstrahlung) precludes its use for trace element characterization, such as the minor elements contained in a single textile fiber or in general polymers. In addition, nonconducting samples require coating. The analysis also requires high vacuum and therefore is not well suited for liquids, volatile or biological samples. Despite these limitations, EPMA is often encountered in forensic laboratories because it constitutes the standard technique for GSR analysis. It is useful for the analysis of any forensic sample where the need for good spatial resolution outweighs the need for high elemental sensitivity.

3.3 X-ray Fluorescence Microanalysis

This combines the excellent elemental sensitivity of conventional XRF with spatial resolution in the EPMA range. A typical XRFM system is equipped with a collimated microfocus X-ray tube, a rhodium anode, an optical sample viewing/positioning device, and a precision computer controlled *xyz* stage.[1] With XRFM it is possible to scan, locate and analyze particles of 10 μm and larger (up to some 20 cm in diameter). Multiple-point microfluorescence analyses are also possible and can provide an average composition and information on elemental heterogeneity. Elemental maps can indicate the distribution of elements of interest across the sample, which can show the distribution of GSR around a bullet entry hole or the distribution of inorganic pigments on a surface coating. Multielemental analyses are possible from sodium to uranium in a rapid and nondestructive manner. The sensitivity is 100–1000 times better than that of EPMA, and the sample preparation is much simpler (no coating is required). Heavier elements can be analyzed at atmospheric pressure, whereas lighter elements can be analyzed under helium or at reduced pressure. Samples such as pastes, solutions, and volatile liquids that cannot be analyzed by EPMA (due to the requirement to work in a vacuum) can be readily analyzed by XRFM, making it a technique of choice for the elemental analysis of a wide range of forensic samples thanks to the features indicated above.

Unfortunately, most XRFM systems are not able to analyze particles less than 10 μm because the performance is restricted by the size of the collimator (minimum 50 μm) and the power of the X-ray tube (maximum 100 W). The smaller the collimator, the less energy hits the sample, and the less signal is detected.

A new type of XRFM has been recently developed with an emphasis on forensic applications. A novel approach has been employed, using capillary optics to guide and compress X-rays to very intense beams with a small angular dispersion. The two key features of good spatial resolution and high elemental sensitivity are obtained by this approach. This technology is claimed to be able to analyze particles as small as 5 μm in diameter. In some cases this improved spatial resolution may be significant. Otherwise, the capabilities of such a system are not significantly different from those of a more traditional XRFM system.

3.4 Total Reflection X-ray Fluorescence Spectrometry

This techniques differs from conventional dispersive energy XRF in that the incident X-rays arrive near the critical angle of reflection without penetration into the sample substrate. The sensitivity is greatly improved (in the picogram range) thanks to a better signal-to-noise ratio. Another advantage of the technique is its ability to perform quantitation using a single element internal standard. The TXRF method appears to combine the much needed requirements of high sensitivity and the ability to analyze minute solid samples. The current price for such a system is rather prohibitive for routine analyses (in the order of US$250 000). Only a few studies have been reported on forensic applications (see below); however, TXRF is already seen as an excellent method for characterizing the trace element profiles of minute samples, including specimens that are essentially organic in nature (such as plastic wraps, synthetic fibers, etc.).

4 FORENSIC APPLICATIONS

4.1 Metals

In forensic science it is sometimes important to analyze metals such as debris from projectiles, metal filings, or other metallic exhibits relevant to a case under investigation. The choice of the technique is dependant on the size of the sample and on the resources available to the forensic laboratory.

If the sample is macroscopic, e.g. a metal plate, it can be analyzed by conventional XRF with no major

problem. If the sample is microscopic in size, EPMA or XRFM becomes the method of choice. The first is chosen for single-particle analyses when elemental sensitivity is not an issue. The latter is more appropriate when good sensitivity and/or elemental mapping is required. For example, with XRFM, it is possible to identify the position and composition of microscopic metal particles originating from a bullet on cloth, skin, etc., without removing the particles and with limited sample preparation.[2,3]

4.2 Gunshot Residues

Most of the techniques available for detecting GSR on a suspect's hands or clothing are based on the inorganic primer components of the ammunition. In the past, GSR analysis has been carried out using a number of different bulk elemental analysis techniques including neutron activation analysis (NAA), AAS, and inductively coupled plasma (ICP) techniques. Currently, the standard technique is EPMA.[4,5] An adhesive SEM stub is repeatedly pressed over the area of interest (for example, the hands of a suspect). The stub is then analyzed by EPMA. The advantage of EPMA over the bulk elemental analysis techniques is that it is a nondestructive method that characterizes individual GSR particles by morphology and size, in addition to carrying out the elemental analysis. Generally, GSR particles are spherical and between 0.5 and 5.0 μm in diameter, although particles of over 100 μm have been observed.[6] Also, whereas the bulk elemental analysis techniques only allow a total quantity of each of the elements to be given, with EPMA it is possible to determine whether each of the elements observed are present in a single particle. As each of these requirements are filled, it becomes more likely that the particle in question is in fact GSR, and not an environmental contaminant.

The main disadvantage of EPMA is that it is time-consuming and particles may be overlooked. This problem has been somewhat overcome by the introduction of automated instrumentation.[7–9]

The potential of XRFM as an alternative to EPMA for gunshot residue analysis has been recently reported.[3,10] Although this technique is ineffective at analyzing individual particles that are less than 10 μm in diameter, it shows interesting features. The elemental mapping capabilities can be used for the on-target analysis of the dispersion of GSR around bullet holes. The analysis of residues on target substrates can potentially indicate the type of primer employed (lead based or lead free) and the nature of the projectile (for example, fully jacketed or exposed lead). An indication of the muzzle-to-target distance (for example, less than 30 cm, or more than 30 cm) can also be obtained from this analysis. As this procedure is nondestructive, it does not preclude the application of more conventional techniques. An example XRFM

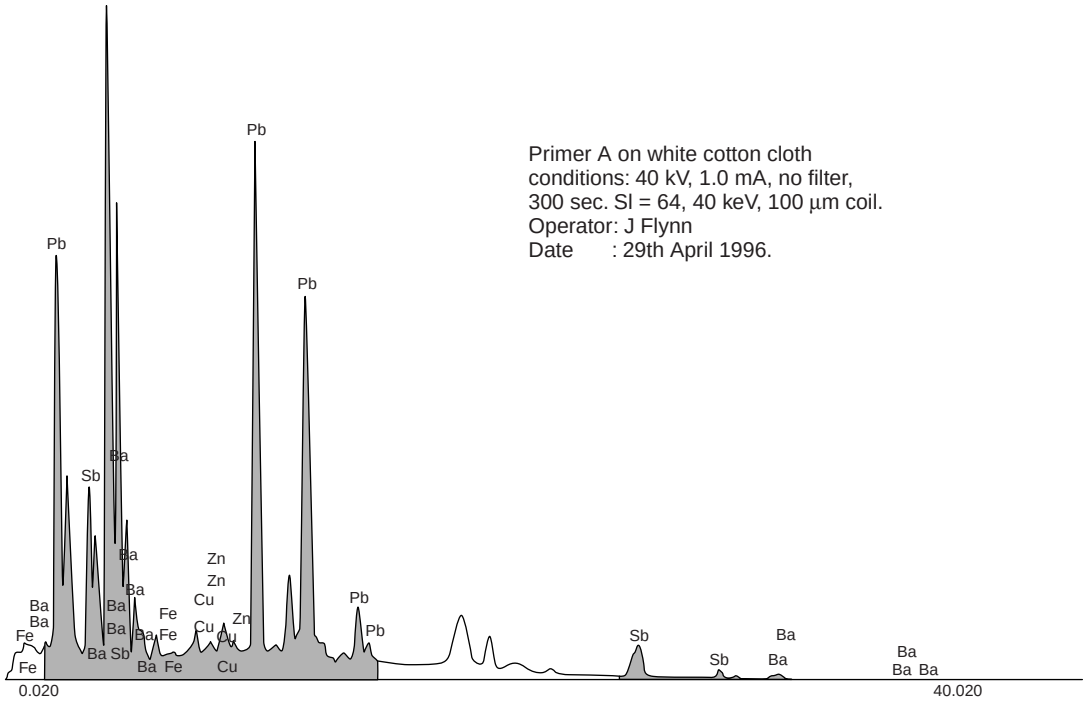

Figure 1 Example XRFM spectrum of GSR from Winchester 9 mm NATO ammunition. (Courtesy ref. 3; Copyright Elsevier Science – reprinted with permission.)

spectrum of GSR is shown in Figure 1. An example of on-target elemental mapping by XRFM for the analysis of GSR dispersion is shown in Figure 2.

It should be pointed out that a number of manufacturers have recently developed new types of ammunition which are lead or heavy metal free.[11] The main inorganic constituents of GSR derived from this ammunition type are not specific to gunshot residue, and may be present as an environmental contamination. This means that the applicability of EPMA or XRFM, like other inorganic techniques, to the detection of these types of GSR is limited.[12] As a result, these techniques may be less useful in the future, and the forensic examination of GSR may swing towards the organic analysis of propellant residues.

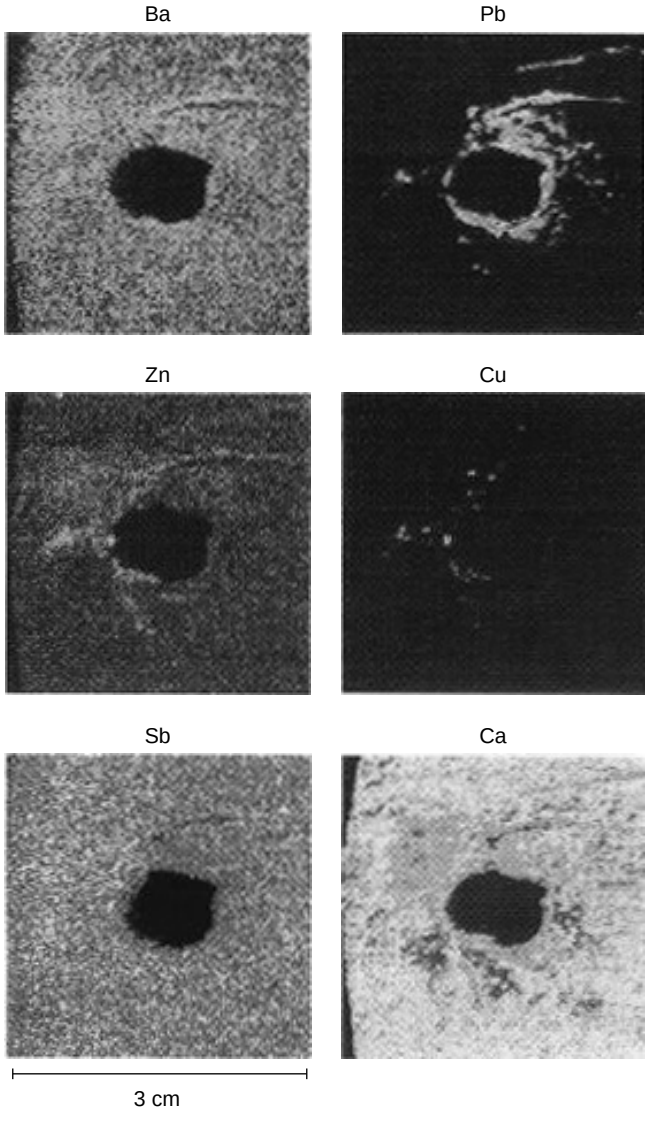

Figure 2 Elemental maps from the on-target XRFM analysis of a section of parietal bone excised from a suicide victim. (Courtesy ref. 3; Copyright Elsevier Science – reprinted with permission.)

4.3 Paint

Paint is a typical example of trace evidence, and is routinely analyzed by forensic laboratories around the world. The analysis of paint smears or paint flakes can provide crucial information in linking two items, or an item with a person. Common examples include motor vehicle accidents and burglaries, e.g. contact between a painted tool and a window frame. Forensic paint examination protocols include elemental profile characterization, in addition to optical examination and organic analysis. Elemental analysis targets inorganic pigments and extenders present in the paint sample.

The application of XRF techniques to forensic paint analysis is well established and has been reported by many authors.[13–17] It has been shown that automotive paint samples coming from different makes and models are easily differentiated using XRF. However, samples coming from different locations from the same vehicle may also show some differences. When dealing with automotive paints it is therefore important to collect appropriate samples in terms of number of samples and sampling locations on the car.

Recent studies have shown that XRFM is particularly useful with such small solid samples and provides a better discrimination than X-ray diffraction measurements.[18] The mapping capabilities of XRFM are of particular interest as well.

4.4 Glass and Soil

When glass is examined in forensic science, the most common analysis performed is the refractive index (RI) determination. Nowadays this can be quickly and easily achieved even with microscopic fragments using a glass refractive index measurement (GRIM) automated system (Foster and Freeman, UK). However, in some cases it may be important to characterize the elemental composition of the glass. The elemental composition has the potential to further discriminate similar glass fragments which remain indistinguishable on the basis of their RI.[19–23] The elemental composition can also indicate the end-use category of the glass being examined (i.e. container glass, headlamp, window, etc.).[21,23] An understanding of the distribution of the elemental composition within and between glass sources undoubtedly enhances the interpretation of this type of evidence. The elements of interest include Al, Fe, Mg, Mn, Sr, Zr, Ca, Ba, Ti, Rb, Sb, Pb, Na, and K.[21,23–26]

Several techniques can be applied to obtain elemental composition information on forensic glass samples.

However, due to the nature of the material, solid-state techniques such as XRF are often preferred. In addition, as glass is often encountered as submillimeter particles, micro techniques are best suited (i.e. EPMA or XRFM). The EPMA method can be used for the in situ characterization of very small samples such as glass beads on headlight filament, and has a better sensitivity than XRFM for lighter elements. The method of choice is probably XRFM because it requires little sample preparation and has a better sensitivity than EPMA for elements heavier than Ca. For the analysis, a submillimeter sample can be stuck onto a piece of clear sticky tape, or glued on a thread of cotton.

In forensic science, soil can be useful for linking the suspect or a victim to a scene, or an object to a scene. A typical protocol for forensic soil analysis includes pH and color determination, particle size distribution, particle density distribution, and some kind of instrumental analysis, possibly including XRF. For example, Hiraoka[27] could classify 110 soil samples into nine categories using their element profiles based on Si, K, Ca, Ti, Fe, Rb, and Sr. These categories showed good agreement with known geological features. Although the potential of XRF for soil analysis has been demonstrated, it is fair to say that it does not constitute a routine technique for soil analysis.

4.5 Fibers

Fibers are generally examined using a protocol with a large emphasis on microscopic techniques. In addition, fibers are commonly analyzed using visible microspectrophotometry and Fourier transform infrared microspectroscopy. Elemental analysis is rarely carried out, but can supplement the existing pool of techniques used for forensic fiber examination.[28]

Most textile fibers are organic polymers but contain inorganic constituents (residues from the manufacturing process, additives/finishing agents, or environmental contaminants). Inorganic constituents are only present in minute quantities in single fibers. Therefore, extremely sensitive methods are required for their detection. In this context, XRF techniques have proved to be useful, although they are rarely employed in routine casework. The EPMA method is generally considered relatively insensitive for such an application, whereas XRF and XRFM have been applied, respectively, to carpet fibers,[29] and to colorless acrylic and polyester fibers.[30] The TXRF technique has been investigated for the analysis of colorless samples (single fibers and weighable amounts) of wool, viscose and polyester,[31] and individual automotive carpet fibers.[32] These studies have shown that various fibers made of the same material (i.e. same polymer or same generic class) can be distinguished by their trace element content. Elements of interest are P, Ti, Mn, Sb, Cr, Co, Cu, Zn, Fe, Ca, and Ti. Both XRFM and TXRF appear to be promising methods for the elemental analysis of single fibers. These methods fulfil the requirements of elemental sensitivity and the ability to deal with minute solid samples, including single fibers as small as 5 mm and 2 mm in length, respectively. It should be pointed out that, when dealing with these techniques, it is very important to use appropriate plastic microtools to minimize external contamination. Environmental contamination due to washing or wearing of garments is generally unavoidable.[33,34] As a result, extra care must be taken when interpreting localized variations and/or when the known sample (e.g. suspect's garment) has been obtained a significant period of time after the transfer of the unknown sample (e.g. fibers found on the victim).

These techniques can be applied to dyed and colorless fibers, especially in cases where limitations are imposed by similarities in morphological features. It is anticipated that, in the future, with the advances in this field, the elemental analysis of fibers will be more widely used and accepted.

Examples of XRF spectra of carpet fibers are shown in Figures 3 and 4.

4.6 Plastics and General Polymers

Plastics in general, and plastic bags and polymer films in particular, are in common use in a wide range of daily activities. As a result, they are commonly employed in relation to the commission of crime, the most obvious example being packaging material in drug trafficking cases. Other examples include plastic fragments found amongst debris from a bomb scene, pieces of adhesive tapes, traces of plastic bumper bars, etc. The ability to compare and ultimately identify plastics from crime scenes with similar items in the possession of suspects, or

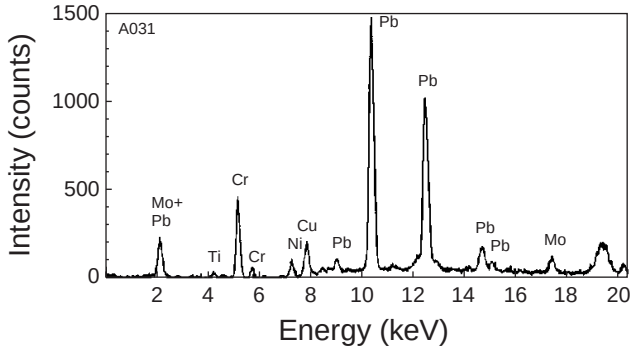

Figure 3 The XRF spectrum of a polypropylene automotive carpet fiber, with sample holder blank subtracted, acquisition time 789 s. (Reprinted with permission from Koons,[29] Copyright ASTM.)

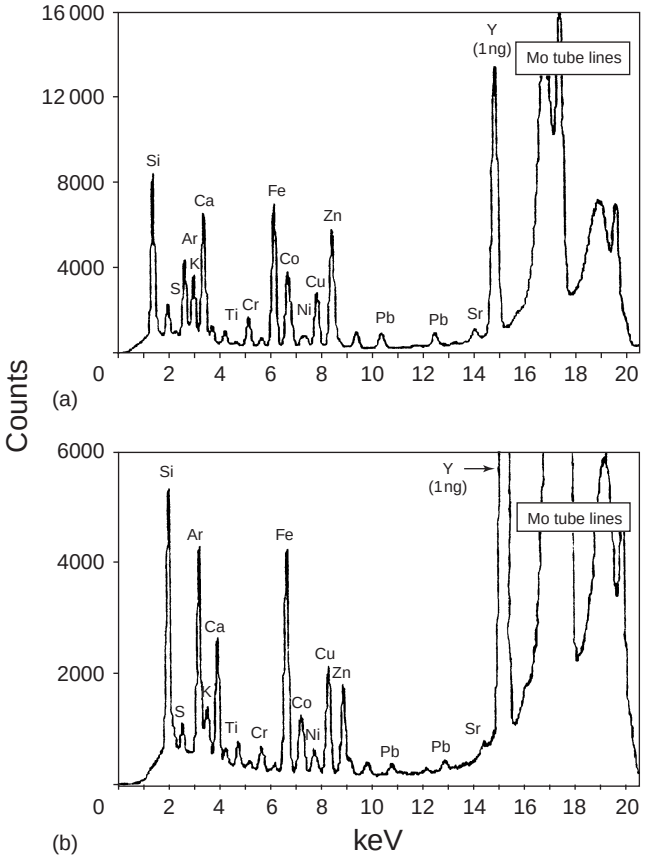

Figure 4 The similar but differentiable TXRF spectra of grey nylon carpet fibers from two different automobiles of the same year, make and model. These spectra were acquired after plasma ashing for 2000 s with Mo excitation (50 kV, 38 mA) using a thin filter and 1 ng Y as internal standard. (Courtesy J.A. Buscaglia, Forensic Science Research and Training Center, FBI Academy, Quantico, VA – ref. 28, Copyright Taylor & Francis, reprinted with permission.)

found in serial cases, is of prime importance. In general, this examination relies on optical, physical, and chemical techniques. A wide variety of chemical methods can be applied to the examination of plastic packaging materials. These methods characterize the make-up of the plastics, including the polymer base, along with the dyes and pigments. These methods, including XRF techniques, can differentiate samples which appear visually similar in the first instance.

The XRF techniques target inorganic additives, and are well suited for the nondestructive elemental characterization of such samples. Almost any XRF technique can be applied, depending on the size of the sample and the level of trace elements targeted. Conventional XRF is well suited if the sample is not too small. If high sensitivity and the ability to analyze small samples are required, XRFM and TXRF are the methods of choice.

For example, elemental profiles of colored polyethylene bags found in illicit drug cases as determined by XRF were found to be very reliable for discriminating between a piece cut from a bag and a batch of bags.[35] Similarly, the value of the elemental profile of green polyethylene garbage bags was investigated using TXRF.[36] The classification model based on the linear discriminant analysis (LDA) of data matrices obtained for Ti, Ca, Pb, Zn, Cu, Cr, and Fe could discriminate samples within a specific package and also different packages. The TXRF method appeared to be more discriminating than differential scanning calorimetry, infrared spectroscopy or optical examination. The technique was found to be particularly useful for the comparison of small amounts of very similar plastics. The usefulness of TXRF for the forensic analysis of polymeric materials, including adhesive tapes, plastic bags and wraps, wire insulation and synthetic fibers, has been recently confirmed.[37,38]

The EPMA technique was found to be the most discriminating in the sequence FTIR, PGC, and EPMA for the analysis of plastic bumper bars.[39] In this study, 28 bumper bar samples could be classified into 17 categories on the basis of the element profiles. However, the authors pointed out that the nonideal nature of casework samples may reduce the degree of discrimination unless care is taken to present them as polished samples.

Elemental profiles as determined by XRF techniques are obviously very useful and should be incorporated into the forensic protocol for the analysis of plastics and general polymers. However, it should be pointed out that extreme care must be paid to the issue of contamination. In this respect the selection of an appropriate sample preparation protocol is crucial. In addition, the interpretation of the data is not always straightforward.

An example XRF spectrum of a green polyethylene bag is shown in Figure 5.

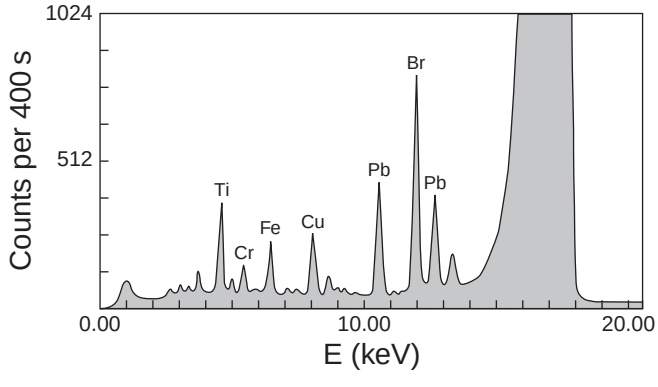

Figure 5 The XRF spectrum of a green polyethylene bag. (Reprinted with permission from Nir-El,[35] Copyright ASTM.)

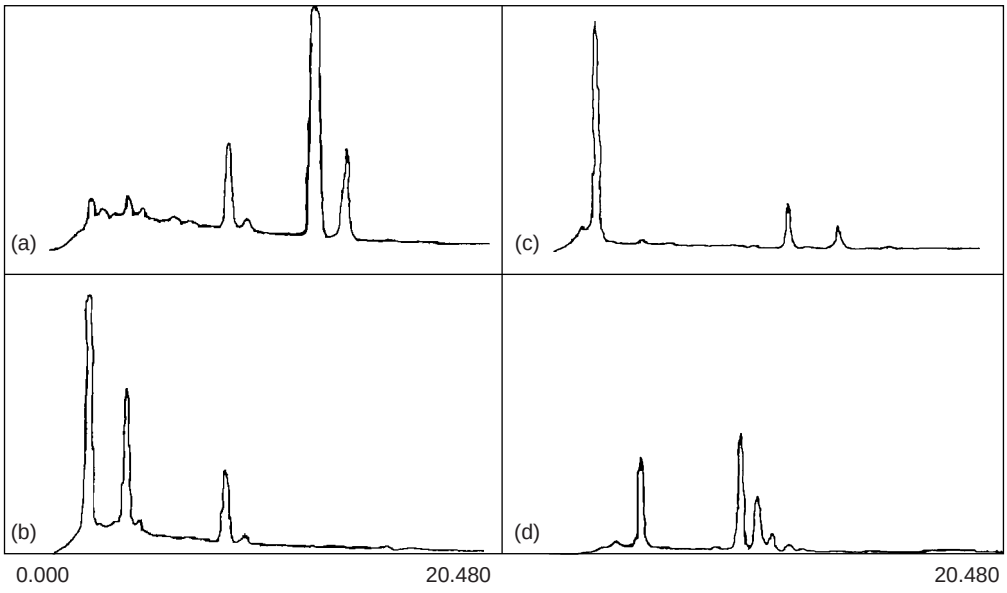

Figure 6 The XRF analyses of four lipstick samples from the same manufacturer (Clinique). (Reprinted with permission from Choudhry,[41] Copyright ASTM.)

4.7 Miscellaneous

In addition to the examples described above, a wide variety of materials can be encountered in forensic casework. Many of these materials may require an elemental analysis and therefore can be analyzed using XRF. Several miscellaneous applications are discussed below.

- *Cosmetics* Cosmetic traces, and especially lipstick smears, are sometimes found as evidence on clothing, cigarette butts, crime scene surfaces, etc. They can be successfully analyzed using XRF techniques. The elemental profile may be useful in identifying samples where different pigment formulations are used for similar colors and shades of lipstick made by different manufacturers. Applications of EPMA have been reported by Andrasko[40] and by Choudhry.[41] Example XRF spectra of cosmetics are shown in Figure 6.
- *Unreacted Explosive Mixtures* Unreacted explosive mixtures, including match head residues, can be found in postexplosion debris or at fire scenes. These residues can be submitted to an elemental analysis including XRF techniques.[42–44] The elemental profiles obtained (principally based on Cl, K, S, Si, and other minor elements such as Ca, Ti, Mn, Fe, and Zn) can bring useful information to identify residues and link different cases, or to compare samples found at different locations.
- *Oil and Lubricants* Oil and lubricants can be present as evidence in a variety of forensic cases. The organic characterization of oil and lubricant samples is usually determined by Fourier transform infrared spectroscopy, gas chromatography/mass spectroscopy, fluorescence spectroscopy, and/or high-performance liquid chromatography. Trace element profiling can be carried out and can be useful for discriminating this type of sample. In addition, it appears that engine wear and specific environmental conditions such as soil contamination can contribute to the individualization of a sample.[45]
- *Questioned Documents* Materials such as paper, ink, correction fluid, toner, etc. are often analyzed in association with the forensic examination of documents. Common protocols rarely include elemental profile characterization, but it has been shown that elemental analysis, including XRF techniques, can provide still greater discriminating power and reduce the chance of an accidental match.[46–49]

5 DISCUSSION AND CONCLUSIONS

A major part of forensic science is based on the chemical characterization of samples of interest. Elemental profile characterization is a common procedure in analytical chemistry, and is therefore often carried out in forensic casework.

Typical forensic samples are generally very small and complex matrices. In addition, due to the legal implications of any forensic procedure, the samples should be kept intact as far as possible. For these reasons,

sensitive solid-state and nondestructive techniques such as XRF are ideal for the forensic scientist. XRF analysis is applicable to a wide range of samples of forensic interest. In addition, specialized XRF techniques such as XRFM and TXRF are extremely interesting from a forensic perspective. These methods fulfil the requirements of good elemental sensitivity and the ability to deal with minute and complex solid samples.

However, it should be pointed out that XRF, as any analytical technique, has some limitations. The major limitation is due to the inhomogeneity of most forensic samples. To discriminate samples of similar make-up, it is often necessary to consider minor elements, because there is little variation in the concentration of major elements present. The intrasample variation can often be close to the intersample variation. It is therefore sometimes difficult to ascertain whether differences are due to the inhomogeneity of the samples or to the fact that the samples come from different sources. In addition, when dealing with highly sensitive techniques such as TXRF, it is very important to pay extra care with sample preparation and handling to minimize external contamination. For example, it is advisable to use plastic microtools rather than metallic implements. It becomes obvious that the comparison and interpretation of elemental profiles requires experience and careful checks to arrive at a meaningful conclusion. Similarly, the data available are often not extensive enough to state that the manufacturing source can be identified on the basis of the elemental profile to the exclusion of all other sources. However, matching profiles are good indicators of a common manufacturing source. When added to the other analytical data available, and in conjunction with industrial inquiries, this information may lead to a positive identification.

The EPMA technique has long been a mainstay in forensic laboratories, particularly for its application to the analysis of GSR. However, with the advent of heavy metal free ammunition, this technique may be of reduced usefulness in the future. More recent developments such as XRFM and TXRF have highly desirable characteristics for the elemental profiling of a wide range of forensic samples. It can be anticipated that these techniques will find increasing use in forensic laboratories in the future.

ACKNOWLEDGMENTS

The authors would like to thank JoAnn Buscaglia, Forensic Science Research and Training Center – FBI Academy, for her assistance in providing examples of TXRF spectra.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
EPMA	Electron Probe Microanalysis
GRIM	Glass Refractive Index Measurement
GSR	Gunshot Residues
ICP	Inductively Coupled Plasma
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
LDA	Linear Discriminant Analysis
NAA	Neutron Activation Analysis
RI	Refractive Index
SEM	Scanning Electron Microscope
SEM/EDX	Scanning Electron Microscopy/Energy-dispersive X-ray Spectrometry
TXRF	Total Reflection X-ray Fluorescence Spectrometry
XRF	X-ray Fluorescence
XRFM	X-ray Fluorescence Microanalysis

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Proton-induced X-ray Emission in Environmental Analysis • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
Forensic Science: Introduction

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. D.C. Wherry, B.J. Cross, T.H. Briggs, 'An Automated X-ray Microfluorescence Materials Analysis System', *Adv. X-ray Anal.*, **31**, 93 (1988).

2. G. Nilsson, B. Stocklassa, 'Analysis of Particles Around Bullet Holes with X-ray Fluorescence and Determination of the Composition of Bullet Jackets', presented at the *13th Meeting of the International Association of Forensic Sciences*, Düsseldorf, 1993.
3. J. Flynn, M. Stoilovic, C. Lennard, I. Prior, H. Kobus, 'Evaluation of X-ray Microfluorescence Spectrometry for the Elemental Analysis of Firearm Discharge Residues', *Forens. Sci. Int.*, **97**, 21–36 (1998).
4. J. Andrasko, M.A. Maehly, 'Detection of Gunshot Residues on Hands by SEM', *J. Forens. Sci.*, **22**, 279–287 (1977).
5. ASTM Committee E-30, 'Standard Guide for Gunshot Residue Analysis by Scanning Electron Microscopy/Energy-Dispersive Spectroscopy', ASTM Standard E 1588-94, 960–962, 1994.
6. R.S. Nesbitt, J.E. Wessel, G.M. Wolten, P.F. Jones, 'Evaluation of a Photoluminescence Technique for the Detection of Gunshot Residue', *J. Forens. Sci.*, **22**, 288–303 (1977).
7. T.G. Kee, C. Beck, 'Casework Assessment of an Automated Scanning Electron Microscope/Microanalysis System for the Detection of Firearms Discharge Particles', *J. Forens. Sci. Soc.*, **27**, 321–330 (1987).
8. R.S. White, A.D. Owens, 'Automation of Gunshot Residue Detection and Analysis by Scanning Electron Microscopy/Energy Dispersive X-ray Analysis (SEM/EDX)', *J. Forens. Sci.*, **32**, 1595–1603 (1987).
9. M.S. Germani, 'Evaluation of Instrumental Parameters for Automated Scanning Electron Microscopy/Gunshot Residue Particle Analysis', *J. Forens. Sci.*, **36**, 331–342 (1991).
10. J. Brazeau, R.K. Wong, 'Analysis of Gunshot Residues on Human Tissues and Clothing by X-ray Microfluorescence', *J. Forens. Sci.*, **42**, 424–428 (1997).
11. G.M. Lawrence, 'Lead-Free or Clean-Fire?', Presented at the *Southwestern Association of Forensic Scientists (SWAFS) Spring 1993 meeting*, South Padre Island, Texas, 1–7, 1993.
12. H. Meng, B. Caddy, 'Gunshot Residue Analysis – A Review', *J. Forens. Sci.*, **42**, 553–570 (1997).
13. D. Smale, 'The Examination of Paint Flakes, Glass and Soils for Forensic Purposes, with Special Reference to Electron Probe Microanalysis', *J. Forens. Sci. Soc.*, **13**, 5–15 (1973).
14. L. Haag, 'Element Profiles of Automotive Paint Chips by X-ray Fluorescence Spectrometry', *J. Forens. Sci. Soc.*, **16**, 225–263 (1976).
15. R. Wilson, G. Judd, S. Ferris, 'Characterization of Paint Fragments by Combined Topographical and Chemical Electron Optics Techniques', *J. Forens. Sci.*, **19**, 363–371 (1974).
16. D. Cousins, D. Rendle, P. Nolan, 'The Application of Microspectrophotometry, X-ray Diffraction and Electronprobe Micro-analysis to the Analysis of Pigments in Paint', *J. Forens. Sci. Soc.*, **24**, 348 (1984).
17. T. Beam, V. William, 'Analysis Protocol for Discrimination of Automotive Paints by SEM/EDX Using Beam Alignment by Current Centering', *J. Forens. Sci.*, **35**, 1055–1063 (1990).
18. G. Massonnet, 'Comparison of X-ray Fluorescence and X-ray Diffraction for the Forensic Analysis of Automobile Paint', *Adv. Forens. Sci.*, **5**, 321–326 (1995).
19. R. Koons, C. Fiedler, R. Rawalt, 'Classification and Discrimination of Sheet and Container Glasses by Inductively Coupled Plasma/Atomic Emission Spectrometry and Pattern Recognition', *J. Forens. Sci.*, **33**, 49–67 (1988).
20. R. Koons, C. Peters, P. Rebbert, 'Comparison of Refractive Index, Energy Dispersive X-ray Fluorescence and Inductively Coupled Plasma Atomic Emission Spectrometry for Forensic Characterization of Sheet Glass Fragments', *J. Anal. At. Spectrom.*, **6**, 451–456 (1991).
21. B. Stocklassa, G. Nilsson, C. Karlsson, R. Miguel, L. Halberg, N. Paulsson, 'Analysis of Sub-mm Fragments of Glass Using X-ray Fluorescence', presented at the *13th Meeting of the International Association of Forensic Sciences*, Düsseldorf, 1993.
22. T. Goldman, T. Hicks, P. Margot, 'The Analysis of Glass Fragments Using Energy Dispersive X-ray Microfluorescence', in *Proceedings of the 14th Meeting of the International Association of Forensic Sciences*, Tokyo, **4**, 74–78, 1996.
23. J.R. Almirall, M. Cole, K. Furton, G. Gettinby, 'Classification and Discrimination of Forensic Glass Samples Using the Statistical Analysis of their Elemental Composition and Refractive Index Data', in *Proceedings of the 14th Meeting of the International Association of Forensic Sciences*, Tokyo, **4**, 79–89, 1996.
24. C.R. Howden, R.S. Dudley, K.W. Smalldon, 'The Analysis of Small Glass Fragments Using Energy Dispersive X-ray Fluorescence Spectrometry', *J. Forens. Sci. Soc.*, **18**, 99–112 (1978).
25. J. Locke, D. Boase, K.W. Smalldon, 'The Use of Spark Source Mass Spectroscopy for the Analysis of Small Glass Fragments', *J. Forens. Sci. Soc.*, **18**, 123–131 (1978).
26. D.A. Hickman, G. Harbottle, E.V. Sayre, 'The Selection of the Best Elemental Variables for the Classification of Glass Samples', *Forens. Sci. Int.*, **23**, 189–212 (1983).
27. Y. Hiraoka, 'A Possible Approach to Soil Discrimination Using X-ray Fluorescence Analysis', *J. Forens. Sci.*, **39**, 1381–1392 (1994).
28. C. Roux, 'Scanning Electron Microscopy and Elemental Analysis', in *Forensic Examination of Fibers*, eds. J. Robertson, M. Grieve, Taylor & Francis, 239–250, 1999.
29. R.D. Koons, 'Comparison of Individual Carpet Fibers Using Energy X-ray Fluorescence', *J. Forens. Sci.*, **41**, 199–205 (1996).
30. J. Cartier, C. Roux, 'A Study to Investigate the Feasibility of Using X-ray Fluorescence Microanalysis to Improve

Discrimination between Colorless Synthetic Fibers', *J. Forens. Sci.*, **42**, 1019–1026 (1997).
31. A. Prange, U. Reus, H. Böddeker, R. Fischer, F.P. Adolf, 'Microanalysis in Forensic Science: Characterization of Single Textile Fibers by Total Reflection X-ray Fluorescence', *Anal. Sci.*, **11**, 483–487 (1995).
32. J.A. Buscaglia, R.D. Koons, 'The Characterization of Automotive Carpet Fibers by Total Reflection X-ray Fluorescence Spectrometry', presented at the *50th Meeting of the American Academy of Forensic Sciences*, San Francisco, 1998.
33. G. Prior, *Röntgenanalytische Untersuchungen zur Differenzierbarkeit von Jeansfasern mittels charakteristischer Elementgehalte*, unpublished dissertation, Wilhelms University, Münster, 1991.
34. R. Jallard, *Application de la Microfluorescence X lors de la Discrimination de Fibers de Blue Jeans*, unpublished seminar thesis, University of Lausanne, 1995.
35. Y. Nir-El, 'Forensic Characteristics of Colored Polyethylene Bags', *J. Forens. Sci.*, **39**, 758–768 (1994).
36. F. Düwel, R. Fisher, T. Schönberger, U. Simmross, D. Weis, 'Quantitative Elemental Analysis of Thermoplastic Remains Using Total Reflexion X-ray Fluorescence (TXRF) Spectrometry', in *Proceedings of the 14th Meeting of the International Association of Forensic Sciences*, Tokyo, **4**, 50–53, 1996.
37. J.A. Buscaglia, R.D. Koons, 'The Application of Total-Reflection X-ray Fluorescence Spectrometry to the Elemental Analysis of Trace Evidentiary Materials', presented at the *50th Meeting of the American Academy of Forensic Sciences*, San Francisco, 1998.
38. J.A. Buscaglia, R.D. Koons, 'Elemental Analysis for the Comparison of Black Polyethylene Trash Bags', presented at the *15th the International Association of Forensic Sciences*, Los Angeles, 1999.
39. A.E. Parybyk, H.J. Kobus, 'The Charaterization of Plastic Automobile Bumper Using Fourier Transform Infrared Spectroscopy (FTIR), Pyrolysis Gas Chromatography (PGC), and Energy Dispersive X-ray Microanalysis Combined with a Scanning Electron Microscope (SEM/EDX)', *J. Forens. Sci.*, **35**, 281–292 (1990).
40. J. Andrasko, 'Forensic Analysis of Lipstick', *Forens. Sci. Int.*, **17**, 235–251 (1981).
41. M.Y. Choudhry, 'Comparison of Minute Smears of Lipstick by Microspectrophotometry and Scanning Electron Microscopy/Energy-Dispersive Spectroscopy', *J. Forens. Sci.*, **36**, 366–375 (1991).
42. A.D. Beveridge, S.F. Payton, R.J. Audette, A.J. Lambertus, R.C. Shaddick, 'Systematic Analysis of Explosive Residues', *J. Forens. Sci.*, **20**, 431–454 (1975).
43. J. Andrasko, 'Identification of Burnt Matches by Scanning Electron Microscopy', *J. Forens. Sci.*, **23**, 637–642 (1978).
44. B. Glattstein, E. Landau, A. Zeichner, 'Identification of Match Head Residues in Post-explosion Debris', *J. Forens. Sci.*, **36**, 1360–1367 (1991).
45. E.O. Espinoza, M.J. Mann, L. Leonardo, A. Copeland, 'Differentiation of Automotive and Locomotive Lubricant Oil', *J. Forens. Sci.*, **39**, 839–843 (1994).
46. D.E. Polk, A.E. Attard, B.C. Giessen, 'Forensic Characterization of Papers. II: Determination of Batch Differences by Scanning Electron Microscopic Elemental Analysis of the Inorganic Components', *J. Forens. Sci.*, **22**, 524–533 (1977).
47. H. Harada, 'A Rapid Identification of Black Color Materials with Specific References to Ballpoint and India Ink', *J. Forens. Sci. Soc.*, **28**, 167–177 (1987).
48. M.F. Cassidy, D.E. Allen, 'Application of the Microdroplet Method of X-ray Analysis to the Characterization of Inks', *J. Forens. Sci.*, **38**, 40–47 (1993).
49. B. Stocklassa, G. Nilsson, 'Elemental Analysis of Ball-Point Pen Ink on Paper', presented at the *13th Meeting of the International Association of Forensic Sciences*, Düsseldorf, 1993.

Industrial Hygiene: Introduction

P.R. Michael
Monsanto Company, St. Louis, USA

C.R. Glowacki
CERP, McClellan AFB, USA

M. Harper
University of Alabama at Birmingham, USA

Industrial hygiene is classically defined as the effort to protect worker health by the recognition, evaluation, and control of stressors in the workplace. Stressors can be in the form of excess energy, for example noise, heat, radiation, or physical stress (ergonomics) or they can be in the form of excessive exposure to a particular chemical element or compound or mixture of compounds. Analytical chemistry plays a fundamental role in identifying and quantifying potentially harmful chemicals in the environment. The determination of chemicals in the workplace is key to all three aspects of industrial hygiene. Recognizing chemicals present in the workplace is sometimes as simple as taking an inventory of the chemicals on hand. Frequently, however, new compounds are formed during chemical reactions and these reactions may not be anticipated in the process design. Evaluating the amounts of chemicals to which workers are actually exposed, for example by the determination of ambient air concentrations, requires the collection of a representative sample typically followed by instrumental analysis to determine the quantities of the compounds or elements in the sample. Once employee exposure has been characterized then possible health consequences can be anticipated and, if needed, appropriate action taken to reduce the exposure. Reliable data are of utmost importance. Not identifying a health hazard can have disastrous consequences ranging from immediate injury or death to more insidious chronic health problems that become manifest years after the exposure has ceased. Conversely, reacting to a hazard that does not exist can result in costly, unneeded engineering modifications or complacency on the part of employees. Follow-up sampling is normally required to judge the efficacy of control measures.

Judgment of ''safe'' working conditions is usually made by comparing personal exposure concentrations to concentrations believed to be tolerated by the majority of average healthy workers without health consequences. Acceptable exposure concentrations can be established as law by government organizations such as the Occupational Safety and Health Administration in the USA and the Health and Safety Executive in the UK, as voluntary standards such as the Threshold Limit Values set by the American Conference of Governmental Industrial Hygienists, Inc., or in some cases as advisory limits set by the manufacturer of a specific chemical. These concentrations provide a frame of reference for analytical chemists involved with industrial hygiene. However, advances in toxicology and occupational medicine usually result in the identification of biological effects at lower and lower dose levels, challenging chemists to measure smaller and smaller quantities of chemicals in the workplace.

Chemicals can cause local irritation or enter the human body via inhalation, ingestion, or through the skin or mucus membranes. All routes of exposure are relevant to airborne materials and so the measurement of air concentrations is the most common activity in which analytical chemists are involved. Determining materials in the air (which can enter the body through the lungs or which can be deposited in the upper respiratory tract, washed into the gut and ultimately ingested) is a key part of assessing the safety of the workplace. However, in recent years, recognition of significant potential for adsorption through the skin has necessitated the development of techniques to assess surface contamination, the amount of skin adsorption, and permeation through protective clothing.

Once within the body, chemicals can cause acute or chronic systemic effects depending on the chemical species, its dose, and the time interval over which the dose is received, as well as on individual susceptibility. Sampling and analysis of the environment only defines exposure and not actual dose or biological effect. Biological monitoring is a holistic approach for the determination of dose, or, in some cases, the biological effect (e.g. monitoring enzyme efficiency). However, due to a number of complications, it is employed much less frequently than air monitoring. To produce meaningful results biological monitoring must be done in the context of a complete understanding of a compound's metabolism in the human body, including metabolites, rates of metabolism, and routes of elimination. Biological matrices are much more complex than air samples and analysis methods are usually much more costly to develop and to implement. Also, biological monitoring is confounded by possible off-the-job exposure, interferences related to medications and personal habits, the personal nature and invasiveness of sample collection, coupled with worker concerns about right to privacy. Methods for biological monitoring of organic compounds in the human body are highly dependent on the compound of interest and are not covered in detail here. There are a number of references that provide more background on biological monitoring

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

in occupational situations.[1–5] A search of the current scientific literature should be performed for information on a specific compound(s).

Applications of analytical chemistry in industrial hygiene are closely related to other fields, especially monitoring of the ambient environment, chemical warfare detection, clinical chemistry, and the development and characterization of field portable instruments. In many cases, sample analysis is very similar and it is prudent to keep abreast of advances in those fields. However, the emphasis in industrial hygiene in determining human exposure dictates sampling or sensing with portable, unobtrusive devices and, in the case of biological monitoring, defines the sample matrix as human in origin (typically fluid, tissue, or exhaled breath). In many cases, sample collection, especially when analytes need to be preserved due to instability, becomes an integral part of the analysis method.

Advances in analytical chemistry, such as a greater ability to measure specific compounds at lower and lower concentrations, will benefit industrial hygiene by allowing a better characterization of the work environment. Increasingly, a main interest in industrial hygiene method development is the ability to have rapid, on-site evaluations. Transporting samples back to a laboratory for analysis will undoubtedly continue to play a large role in industrial hygiene, but real time or near real time data allow rapid identification of hazardous conditions and rapid feedback on corrective measures. Improvements in miniature analyzers, remote sensors, and prepackaged chemistries such as enzyme-linked immunosorbent assay techniques will receive a great deal of attention if they can be made sufficiently rugged to tolerate the work environment.

Because chemicals will always play a vital role in industry and since all chemicals are potentially harmful, depending on the dose, the risk of harmful chemical exposure will always exist. In addition, the efficiency of containment measures is not always as good as anticipated and may deteriorate with time. Therefore hazard assessment is required at all stages of industrial hygiene, and exposure monitoring is a vital part of this process.

This section acquaints the reader with applications of analytical chemistry in industrial hygiene. Articles such as **Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air** and **Surface and Dermal Monitoring** cover techniques used to obtain samples from the workplace. Laboratory analysis and characterization of those samples is addressed in the articles **Chromatographic Techniques in Industrial Hygiene; Spectroscopic Techniques in Industrial Hygiene** and **Dust, Measurement of Trace Elements in**. Devices capable of providing real time or near real time data on site are covered in the articles **Direct Reading Instruments for the Determination of Aerosols and Particulates** and **Sensors in the Measurement of Toxic Gases in the Air**. Both laboratory and on-site measurements are discussed in the articles **Aerosols and Particulates Analysis: Indoor Air** and **Carcinogens, Monitoring of Indoor Air**. Selected approaches to the analysis of biological monitoring samples are covered by the articles **Metals in Blood and Urine: Biological Monitoring for Worker Exposure** and **Parent and Progeny Compounds in Exhaled Breath, Determination of**.

REFERENCES

1. R.C. Baselt, *Biological Methods for Industrial Chemicals*, 2nd edition, Year Book Medical Publishers, Littleton, MA, 1988.
2. M.H. Ho, H.K. Dillon (eds.), *Biological Monitoring of Exposure to Chemicals – Organic Compounds*, John Wiley & Sons, New York, 1987.
3. T.J. Kniep, J.V. Crable (eds.), *Methods for Biological Monitoring: A Manual for Assessing Human Exposure to Hazardous Substances*, American Public Health Association, Washington, DC, 1988.
4. R.R. Lauwerys, P. Hoet, *Industrial Chemical Exposure: Guidelines for Biological Monitoring*, 2nd edition, Lewis Publishers, Boca Raton, FL, 1993.
5. S.S. Que Hee (ed.), *Biological Monitoring: an Introduction*, Van Nostrand Reinhold, New York, 1993.

Activation Analysis of Large Samples

Peter Bode
Delft University of Technology, Delft, The Netherlands

1 Introduction **1**
2 Large Sample Activation Analysis **2**
2.1 Large Sample Neutron Activation Analysis 2
2.2 Large Sample Prompt γ-Activation Analysis 3
2.3 Large Sample Photon Activation Analysis 4
3 Measurement Equation of Large Sample Neutron Activation Analysis **4**
4 Instrumentation **5**
4.1 Neutron Sources for Large Sample Activation Analysis 5
4.2 Irradiation Facility 6
4.3 Sample Containers for Irradiation 7
4.4 Counting Facility 8
5 Calibration **10**
5.1 Absolute Method 10
5.2 The Comparator Method 11
5.3 Single Comparator Method 11
5.4 k_0-Based Method for Standardization 11
5.5 Internal Monostandard Method 12
5.6 Neutron Fluence Rate Monitoring 12
5.7 γ-Ray Self-Attenuation 13
5.8 Extreme Inhomogeneities 13
6 Quality Control **14**
6.1 Quality Control in Large Sample Analysis 15
7 Sensitivity **17**
7.1 Natural Background 18
8 Applications **18**
8.1 Materials Difficult to Homogenize: Geological Samples, Ores, and Waste 18
8.2 Materials That May be Contaminated During Homogenization: High-Purity Materials 19
8.3 Materials Difficult to Subsample: Nutritional Studies 19
8.4 Valuable Material of Irregular Shape 19
8.5 Other Applications 19
References **19**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

There is a trend in many analytical techniques towards the use of smaller sizes of the test portion, and sometimes submilligram amounts are being used. The limitation to the size of the test portion can pose the analyst with problems when the amount of material collected is large. Subsampling and preparation of a representative test portion may be difficult if homogenization is impossible or extremely expensive, or if homogenization introduces contamination. An alternative approach has been introduced in the 1990s by the development of large sample neutron activation analysis (LSNAA), later followed by large sample photon activation analysis. These techniques are capable of direct analysis of samples with masses of hundreds of grams to several kilograms. Though the principles and physics of large sample activation analysis are thoroughly understood, the method is still not as versatile or applicable as, for example, normal small sample neutron activation analysis. In this article, the physics of LSNAA is described, including methods of calibration. Examples are given of irradiation and counting facilities and the special aspects of quality control are discussed. Several examples are given of applications of large sample analysis, e.g., for studies related to electronic waste, complete archaeological and cultural artifacts, high-purity materials, and materials of irregular shape.

1 INTRODUCTION

All the routine multielemental analysis techniques (inductively coupled plasma atomic emission spectrometry (ICP-AES), inductively coupled plasma mass spectrometry (ICP-MS), and instrumental neutron activation analysis (INAA), etc.) employ rather small test portions of material, varying from a few milligrams to a few grams of solids or in the range of a few milliliters in the case of liquids[1] (see Table 1). There is even a tendency to go for smaller test portions, such as in solid-state atomic absorption spectrometry (AAS), laser-ablation ICP, and total reflection X-ray fluorescence (XRF) spectrometry. In XRF, the use of quantities larger than required to prepare the target is anyhow meaningless, as the derived information is from the surface layers, representing a few milligrams only.

The limitation to the size of the test portion can pose problems to the analyst when the amount of material collected is large. This is often the case since soils, rocks, plant material, etc. can be more easily and

Table 1 Sizes of the samples and analytical portions handled in several multielement analysis techniques[1]

Analysis technique		Solid material mass used or prepared to test portion	Volume used as test portion
Atomic absorption spectroscopy (AAS)	Gas furnace	Typically 1–2 g dissolved; maximum approximately 10 g	10–20 μL
AAS	Flame		1–2 mL
Inductively coupled plasma spectroscopy (ICP)		Typically 1–2 g dissolved; maximum approximately 10 g	Approximately 500 μL
X-ray fluorescence spectroscopy (XRF)		10 g	
Instrumental neutron activation analysis (INAA)		Typically approximately up to 500 mg; in some cases up to 30 g	1–50 mL

representatively sampled at quantities in the order of hundreds of grams to kilograms than at quantities less than 1 g. A sample is denoted as ''representative'' when ''it can be expected to exhibit the average properties of the material, environment or population it was taken from''.[2] Representativeness is a priori preserved when (i) the sampling is performed according to specific, certified norms or when (ii) a truly homogeneous material is sampled. Homogeneity is defined as ''the degree to which a property or substance is randomly distributed throughout the material''.[2]

Homogeneity depends on the size of the units under consideration. A mixture of minerals may be inhomogeneous at the molecular or atomic level but homogenous at the particle level. In chemical analysis this unit is obviously correlated with its effect in the conduct of the analysis (e.g. differences in solubility) or in its interpretation. Thus, both from practical and sampling considerations often more material is collected and presented for analysis than can be handled.

Irrespective of the analysis technique selected, attention has to be paid to representative subsampling to obtain a relevant final analytical portion from the originally collected material. When restricting the discussion to the analysis of solid materials, this subsampling may imply sample size reduction techniques and other processing such as sieving, crushing, milling, or blending. Problems of the following types arise:

1. Homogenization is impossible, or extremely expensive, because of material properties. Examples are electronic circuits on printed boards, household waste, and scrap from recycled electronics, automobiles, and plastics. A solution to this problem is to sort the material and to perform individual homogenizations and, subsequently, analyze, thereby increasing the total project costs.
2. The homogenization step results in contamination of the sample. Often the contamination due to crushing or milling is not controllable on every sample type. When processing a large series of samples, careful interim cleaning may get less attention since it is time consuming and therefore expensive.

Testing the degree of homogeneity is a common practice in the preparation of reference materials, but for routine operations, such a procedure, the requirement of analysis and statistical evaluation of at least five or more test portions of each sample, would raise the cost of the analysis considerably.

Considerations from the above indicate that direct analysis of the voluminous solid sample as collected might have advantages, both analytical and economical.

Detection limits in trace-element studies are based on the signal-to-noise ratio. An additional feature of analysis of large test portions is that the detection limit for trace elements may be decreased considerably in case of high-purity materials. This was demonstrated, for example, by Verheijke[3] in the assessment of impurities in (5 in. diameter) silicon wafers to be used in the electronic industry.

2 LARGE SAMPLE ACTIVATION ANALYSIS

2.1 Large Sample Neutron Activation Analysis

In analytical terms, a ''large sample'' can be anything exceeding the regular size of a test portion in the process to determine the components of the material. The regular mass of a test portion in neutron activation analysis (NAA) varies from a few milligrams to 1 g. As already indicated, instrumental neutron activation analysis (INAA) has all the potentials to analyze, with adequate accuracy, test portions[4] in the kilogram range.[5–8] Both the incoming radiation for activation (neutrons) and the outgoing radiation to be measured (γ-rays) have sufficiently high penetrating power to facilitate NAA of portions of samples weighing kilograms. A ''large sample'' in NAA is defined as a test portion in which neutron and γ-ray self-attenuation cannot be neglected in view of the required degree of accuracy.

A few phenomena need more attention in large sample neutron activation analysis (LSNAA) than in normal NAA (handling test portions varying from micrograms to a maximum of 0.5 g), where these phenomena usually have only insignificant impact to the degree of accuracy of the results.[9] In large test portions, e.g., of kilogram size, neutron absorption and scattering result in substantial self-shielding, causing depression of the neutron flux at the center of the sample compared to the periphery. Neutron self-thermalization may cause substantial changes in the neutron spectrum throughout the sample if the sample material also contains, for example, hydrogen.

Similarly, the γ-radiation of the activation products deep inside in the sample will be more strongly absorbed and scattered before leaving the sample than the radiation resulting from, e.g., the surface of the sample; moreover the absorption and scattering increase rapidly at lower γ-ray energies. This effect is denoted as *γ-ray self-attenuation*. In addition, a sample of, say, 1 kg cannot be considered anymore as a more-or-less ''point source'' during counting at normal sample–detector distances of, e.g., 10–30 cm, resulting in a corresponding different response of the detector for the γ-radiation. In contrast to conventional INAA with small samples, analysis of test portions larger than a few hundred milligrams requires correction for these neutron self-shielding and γ-ray self-absorption effects, either via calibration or by modeling.

Trace-element determinations in large test portions have been carried out for decades in areas such as well logging,[10] on-line conveyor belt industrial analyzers,[11] and in vivo studies of, e.g., Ca in bones and Cd in kidneys.[12] In all these studies, the NAA technique is applied since the strong penetration of the neutrons as well as the resulting emitted γ-radiation (either prompt during the nuclear reaction or delayed from the activation products) allows bulk analysis rather than surface analysis, the ''limiting'' factor in XRF spectrometry. For cases such as coal and ore analysis, these bulk analyses have been applied in industrial settings and for field-mineral explorations using 14 MeV neutron generators[13] and isotopic neutron sources.[14] The analyses are mainly focused on raw material analysis and for product control of one or a few major constituents. These procedures are, through calibration, customized for the problems they have been developed for and cannot be translated into a routinely applicable method for the analysis of a large variety of sample types. The same applies to the use of NAA principles in well-logging devices, in which the entire surrounding rock may be considered as a large sample.

The advantage of reactor-based INAA in the analysis of large samples lies with the higher available neutron fluxes and activation cross sections, all leading to better sensitivities for trace elements than can be obtained with 14 MeV neutron generators and isotopic neutron sources and providing the opportunity for full multielement analysis.

LSNAA has evolved over the years[15–20] toward a capability for analysis of various samples types which otherwise would have been difficult to analyze. Both off-line (LSNAA) and on-line (LS-prompt γ-NAA) activation methods are used. The emphasis in this article is on (off-line) LSNAA; prompt γ-LSNAA is discussed briefly below.

2.2 Large Sample Prompt γ-Activation Analysis

Large sample prompt γ-neutron activation analysis (PGNAA) is being applied for many years in well logging, industrial (conveyor) belt analyzers, etc. using isotopic neutron sources such as ^{252}Cf or ^{241}Am(Be). The advantage of isotopic source-based PGNAA compared to normal NAA lies in the fact that the test portion may be analyzed locally rather than having to be taken to the laboratory and on-line information is obtained. Since the output of the sources is rather low, large samples are needed anyhow to obtain a measurable signal, usually from the main components in the material of interest. Industrial analyzers are commercially available for the on-line analysis of cement,[21] the determination of the sulfur content on coal,[22] for the detection of explosives in airline cargo,[23] etc.

Reactor-based large sample PGNAA, i.e. using an external neutron beam, was demonstrated by Sueki et al.[24] for a pottery sample of 15 cm diameter, 10 cm width, and 0.5 cm wall thicknesses. The neutron beam dimensions were approximately 2 cm $\times$ 2 cm.

Similar to ''normal'' LSNAA, in large sample PGNAA also the problems of neutron attenuation and γ-ray self-shielding have to be solved. In the example quoted above, the internal monostandard was used (see Section 5.3). Also, other intact archaeological objects were analyzed by this method, such as bronzes.[25] Moreover, neutron beams from reactors are relatively limited in dimensions (on the order of 5 cm $\times$ 3 cm), which sets also a limit to the size of the object activated. This limitation can be overcome by repositioning the sample in the beam.

An advantage of large sample PGNAA over normal LSNAA is that no special facilities have to be constructed in the reactor, and that the sample contains hardly any induced radioactivity, which is of importance when dealing with, for example, archaeological or cultural artifacts. The PGNAA setup can be standard,[24] but care has to be taken that the large object does not ''transform'' into a very intense source of prompt γ-radiation with associated radiation dose hazards for the researchers.

Other methods for standardization have been proposed too, and mostly are based on a priori available information on the (gross) composition of the object; e.g. using Monte Carlo simulations[26] or neutron transport codes[27] (''fixed point iteration method''). Degenaar developed a method in which no a priori information is used and the neutron self-shielding is estimated on basis of the attenuation and scattering of the neutron beam, measured outside the sample.[28]

2.3 Large Sample Photon Activation Analysis

Photon activation analysis has the potential to analyze very large samples for reasons similar to NAA: large penetration power of the incident bremsstrahlung photons (typically in the order of several tens of million electronvolts), and similar to NAA, large penetration of the γ-radiation from the induced radioactivity. There is also some similarity to prompt γ-NAA with respect to the size of the object that can be exposed at a time; here also, the sample can be ''moved'' through the beam to attain a homogeneous activation, or the beam can be scanned over the sample. It introduces an additional complication if the integral sample is counted after exposure: the different activated parts have different decay times but their signals are registered simultaneously. Alternatively, one may choose to limit the sample size.

At the Bundesanstalt fur Materialprufung (BAM) in Berlin, Germany, large sample photon activation analysis has been applied[29] using the 30 MeV linear accelerator for studies involving samples with sizes in the order of 8 cm height and approximately 2 cm thickness with masses of up to 100–200 g. The measurements were done using a twin detector set up; i.e. the sample was ''sandwiched'' between two side-looking semiconductor detectors.

One of the advantages of photon activation analysis over NAA is that the corrections for self-attenuation of the incoming bremsstrahlung photons are relatively easy to be applied on the basis of fluence rate monitors positioned before and after the sample.[29] Moreover, given the high energy of the photons, this attenuation is mostly relevant for thick targets with high average atomic number.

Large sample photon activation analysis has many interesting aspects and advantages compared to LSNAA, including its capability to detect elements such as C, N, and O, as well as Tl, Bi, and Pb. However, the number of photon activation analysis laboratories worldwide is very small and most of the large sample activation analysis studies are done with neutrons. For these reasons, this type of large sample analysis is not further elaborated upon in this article, and the reader is directed to the available literature.[29]

3 MEASUREMENT EQUATION OF LARGE SAMPLE NEUTRON ACTIVATION ANALYSIS

The basic measurement equations of NAA by which the mass of the unknown element is calculated directly demonstrates the fact that the technique does not set a priori constraints to the mass of the sample analyzed:

$$m_{\mathrm{unk}} = m_{\mathrm{std}} \frac{(A_{0,\mathrm{x}})}{(A_{\mathrm{s}})} R_\theta R_\phi R_{\mathrm{En}} R_\sigma R_{\mathrm{nss}} R_\varepsilon R_{\gamma\mathrm{ss}} R_{\mathrm{inh}} \quad (1)$$

$$A_0 = \Phi_{\mathrm{th}} \sigma_{\mathrm{eff}} \frac{N_{\mathrm{Av}} \theta m}{M} (1 - \mathrm{e}^{-\lambda t_{\mathrm{ir}}}) \mathrm{e}^{-\lambda t_{\mathrm{d}}} \frac{(1 - \mathrm{e}^{-\lambda t_{\mathrm{m}}})}{\lambda} \gamma \varepsilon \quad (2)$$

in which the subscripts ''unk'' and ''std'' refer to unknown and standard, respectively, and

A_0 = the area of the relevant peak in the γ-ray spectrum, corrected for differences in decay and measurement time between the unknown (x) and the standard (s);
R_θ = ratio of isotopic abundance of the element of interest in test portion and standard (often = 1);
R_ϕ = ratio of thermal neutron fluence rates in test portion and standard;
R_{En} = ratio of neutron energy distribution in test portion and standard;
R_σ = ratio of effective activation cross sections for the test portion and standard at the different neutron energy spectra;
R_{nss} = ratio of the neutron self-shielding in test portion and standard;
R_ε = ratio of the photopeak efficiency for the test portion and standard;
$R_{\gamma\mathrm{ss}}$ = ratio of the γ-ray self-attenuation in test portion and standard; and
R_{inh} = ratio of the effect of extreme inhomogeneities in test portion and standard.

Also,
Φ_{th} is the thermal neutron fluence rate ($\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$),
σ_{eff} is the effective absorption cross section (cm^2),
N_{Av} is the Avogadro's number (mol^{-1}),
θ is the isotopic abundance,
m is the mass of the irradiated element (g),
M is the atomic mass number ($\mathrm{g\,mol}^{-1}$),
λ is the decay constant of the radioisotope formed (s^{-1}),
t_{ir} is the irradiation duration (s),
t_{d} is the decay time (s),
t_{m} is the (live time) measuring time (s),
γ is the abundance in the nuclear decay of the γ-ray measured, and
ε is the full energy photopeak efficiency of the detector for the energy of the γ-ray measured.

Many of the correction terms, R_i, can often be neglected in normal sample analysis but some of them like R_{nss}, $R_{\gamma ss}$, R_{ε}, R_{En}, and R_{inh} become significant in large sample analysis. As such, algorithms in large sample NAA differ from normal NAA by the calculation/estimation of

- the neutron self-shielding and/or neutron fluence rate profile inside the test portion,
- the γ-ray self-attenuation,
- the volumetric photopeak source efficiency of the detector, and
- the impact of extreme inhomogeneity effects.

There are many approaches for these calculations, varying from pure theoretical modeling,[9] Monte Carlo modeling,[20] and modeling using a priori available information about the test portion composition[7] to pragmatic empirical estimations of the correction factors.[17] Modeling may even be avoided when, e.g., for routine applications a representative well-characterized (large sample) standard – even a reference material – is available.[18] These standardization methods are further discussed below.

4 INSTRUMENTATION

4.1 Neutron Sources for Large Sample Activation Analysis

The type and strength of the neutron source and energy characteristics play an important role in any type of NAA including LSNAA, as the radioactivity produced is directly proportional to the neutron flux (φ) and energy-dependent neutron absorption cross section (σ). The neutron source should provide a sufficiently high neutron fluence rate so as to keep the product of neutron fluence rate and large test portion mass almost equal to that in small test portion NAA. This criterion indicates that for test portions with masses in the order of 2 kg a neutron fluence rate of approximately $5 \times 10^{12} \times 0.2/2000 = 5 \times 10^{8}\,cm^{-2}\,s^{-1}$ would result in an adequate induced radioactivity during the irradiation time, similar to that applied in conventional NAA in which a 200 mg test portion is processed. Fluence rates on the order of 10^{8}–$10^{10}\,cm^{-2}\,s^{-1}$ are found at an extended distance from the core of small and medium-sized reactors,[15] in beam tubes, and in thermal columns (TCs).[5,19,20] However, low fluence rates can also be realized – or even may be preferred – by lowering the reactor power because of fuel economy considerations.[30] Table 2 also provides an indication of typical neutron sources available to provide the required neutron fluence rate.

The advantage of reactor TCs above, e.g., poolside facilities, is that the longitudinal neutron flux gradient (i.e. horizontally away from the reactor core) over the sample is much less steep because of the multiple neutron scattering in the graphite inside the TC, as can also be derived from the differences in thermal neutron diffusion length in carbon and water, viz., 64.2 and 2.76 cm, respectively. Rotating the sample along its vertical axis compensates partly for these gradients. However, in some materials the neutron self-attenuation combined with the neutron flux gradient may result in situations in which the center of the test portion is hardly activated compared to the periphery. In such cases, the measured γ-rays originate mainly from the periphery and the analysis result will apply merely to the outermost layers of the sample rather than reflecting the bulk composition. In such facilities, the approach is to set limits to the dimensions of the test portion. Moreover, neutron self-moderation will occur owing to the less thermalized neutron spectrum in poolside facilities, resulting in difficulties in the calculation of the element amounts.

Table 2 Typical neutron fluence rates for activation analysis of test portions of different mass along with neutron sources

Sample mass (g)	n-flux ($cm^{-2}\,s^{-1}$)	Neutron source
1–10	2.10^{12}–2.10^{11}	Reactor
10–100	2.10^{11}–2.10^{10}	Reactor
100–1000	2.10^{10}–2.10^{9}	Reactor
>1 kg	$<2.10^{9}$	Reactor, Isotopic n-source n-generator

(Mass × flux) = constant = 200 mg × $10^{13}\,cm^{-2}\,s^{-1}$

Neutron fluence rates $<10^{10}\,cm^{-2}\,s^{-1}$ may also be attainable with isotopic neutron sources and high-intensity neutron generators, and in (reactor-based) external neutron beams. ^{252}Cf is probably the most attractive isotopic neutron source from the point of view of the neutron spectrum shape and easy thermalization aspects. However, the short half-life (2.64 years) may be seen as an economical disadvantage. Other isotopic neutron sources have a relatively hard neutron spectrum, resulting in relatively low thermal neutron fluence rate equivalents. The applicability of such neutron sources may, therefore, be limited to the determination of the major components in a sample.

Neutron generators (3–14 MeV) have their own scope of applications.[31,32] One of the problems in using neutron generators and isotopic neutron sources is that the neutron fluence is anisotropic and therefore the neutron flux seen by a large sample is not the same in all parts of the sample. This can be overcome to some extent by rotating the test portion during irradiation. Given the fact that the thermal neutron fluence rate in, e.g., a D-D

(deuterium ions on a deuterium target) generator is of the order of 10^4 lower than in research reactors, an increase in sample mass of at least 10^4 (i.e. from 200 mg to 2 kg) would be needed to compensate for this low thermal fluence rate.

An advantage of large (kilogram size) sample NAA with D-D generators over reactors is that the sample may be quickly removed from its irradiation position upon shutting down the accelerator, facilitating the measurement of radionuclides with short half-lives. For samples in the order of tens of grams, pneumatic facilities may be designed. Larger sample masses may also be considered for delayed neutron counting procedures to reach lower minimum detectable amounts, although this implies that a larger delayed neutron counter is needed too.

However, several technological obstacles exist both with isotopic neutron sources as with neuron generators, such as the large void needed within the moderator of the device. Moreover, since these moderators are often based on hydrogenous materials such as polyethylene or paraffin, steep flux gradients may occur over the sample, similar to those outlined earlier for poolside reactor facilities. The large sample approach may also be considered for irradiation with the sub-fast neutrons, although here an increase in sample mass from, e.g., 200 mg to 200 g might be sufficient for reaching the sensitivity required. However, an additional problem is that large sample masses will increase the effect of neutron self-moderation.

Most of the external neutron beams from nuclear reactor, with neutron fluence rates of 10^6–$10^8\,cm^{-2}\,s^{-1}$, are suitable for PGNAA. An advantage associated with PGNAA is the flexibility in choosing the mass and shape of the test portion. However, one should be careful in increasing the test portion mass, as it might adversely affect the measurements because the background is sample dependent in PGNAA.[33] It becomes severe particularly in the cases where the hydrogen or boron fraction in the large test portion is high, as it results in an extremely intense source of prompt γ-radiation, which will affect the results.[34] External neutron beams of isotopic neutron sources have usually neutron fluence rates of $<10^7\,cm^{-2}\,s^{-1}$ and can in principle be used for PGNAA, though the energy definition becomes a tedious problem. Steep neutron gradients over the sample occur in neutron beam activation analysis too.

4.2 Irradiation Facility

Transferring the test portions to the irradiation position of the neutron source is usually done with pneumatic/hydraulic transfer facilities and/or manually or automatic loading facilities. Most of the pneumatic facilities are designed to handle volumes up to 5–50 mL, which are placed in a sample carrier known as ''rabbit''. Use of rabbit systems places constraints on the sample shape so as to maintain the defined geometry. In principle, it is possible to transfer test portions up to 1 kg or more through such facilities – such big systems already exist for transferring documents in offices and banks; however, it is yet to be explored and examined whether large rabbits can be obtained with the required specifications (quality of the rabbit materials, purity, and radiation/mechanical resistance) for application in reactors.

In some reactors TCs are available for accommodating a large sample irradiation facility (see Figures 1 and 2). Samples are placed in the irradiation position of the TC by the mechanical movement of a tray that houses the sample in a defined position.[5,19,20] In Table 3, suggestions for facilities for irradiation of test portions of different masses are indicated.

There are various design aspects to be taken into account for irradiation facilities:

1. A large-volume facility near the core of a nuclear reactor creates a void in the reactor's reflector, whereas loading and unloading may cause unwanted fluctuations in the core's reactivity. Moreover, a high amount of[41] Ar will be produced from activation of the air in the container.
2. The thermal neutron fluence rate gradient in the water reflector of a light-water-moderated reactor is quite steep, typically by a factor of 3 per each 3 cm, which is due to the neutron diffusion length (2.84 cm) in water. Such a strong gradient would also create an unwanted strong flux variation over the large test portion to be activated. This may be corrected for by rotating the sample during counting,[5] by mixing the sample after irradiation,[17] or by the use of *in situ* relative efficiency method.[19] Mixing, however, eliminates information about inhomogeneities. The problem of heterogeneity may also be addressed by dividing the large test portion into many smaller fractions, to be processed individually followed later on by combining of the results.
3. Large hydrogen mass fractions may result in neutron spectrum changes over the test portion volume due to self-thermalization. This phenomenon is difficult to correct for mathematically, and may be an additional reason to consider an irradiation facility with well-thermalized neutron spectrum, for instance, to be realized in a TC.
4. Large sample activation facilities at isotopic neutron sources must be designed in such a way that adequate shielding is ensured against the prompt γ-rays, which will be several orders of magnitude higher than with normal small samples.

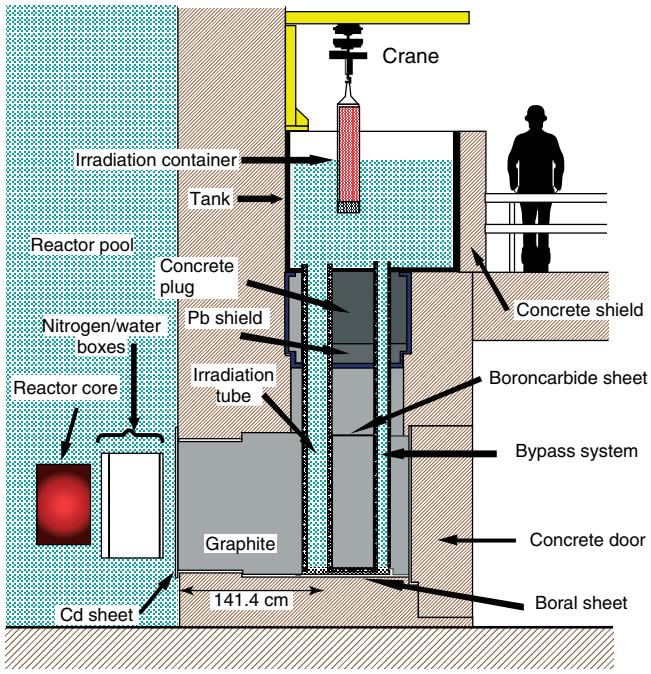

(a)

(b)

Figure 1 Schematic drawing (a) and picture of model (b) of the large sample neutron activation analysis facility in the thermal column of the Hoger Onderwijs Reactor in Delft.

A few of the large sample NAA reactor facilities that are currently operational are given in Table 4.[35] Large sample PGNAA facilities have been realized in Hungary and Japan. Isotope neutron source-based large sample PGNAA facilities are in use in some places.[36]

4.3 Sample Containers for Irradiation

The large sample container itself may be of any shape and type. A wide-neck bottle is easy to fill when coarse material has to be analyzed (Figure 3).[5] A container of inexpensive plastic may be preferred, as the impurities in the plastic itself (blank contribution) may be neglected at a given sample size (see Table 5). If the contribution from the sample holder is substantial, the irradiated large samples may also be easily transferred after irradiation into nonirradiated containers and possible small losses during transfer can be neglected in view of the large mass of the test portion. As such, a Marinelli beaker geometry or multisample container (Figure 4)[37] may

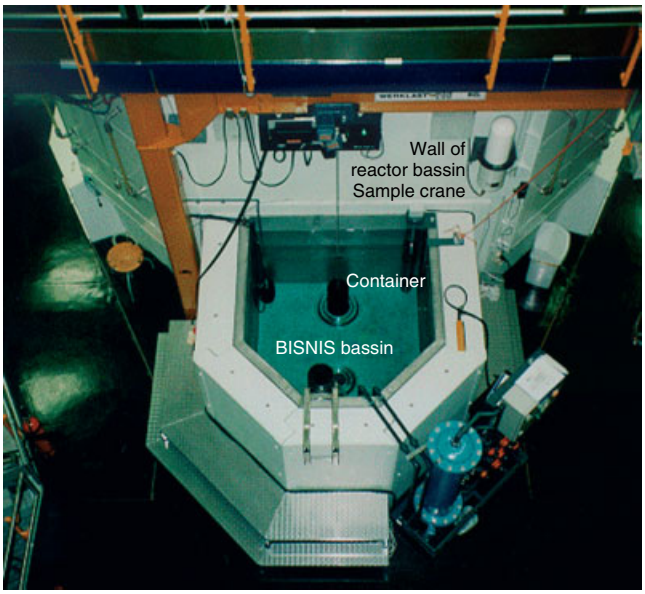

Figure 2 Top view of the large sample activation analysis facility in Delft.

have certain advantages, depending on the application (see below).

4.4 Counting Facility

Very large Ge detectors are available (crystal sizes up 400–800 cm^3, comparable to ''relative efficiencies'' of 100–200%). Such big detectors are an additional tool to maintain adequate sensitivity in NAA. Side-looking detectors (''horizontal dipstick'') have the advantage that cylindrical samples, positioned perpendicular to the detector axis, can easily be rotated around the sample axis to reduce geometrical effects. Vertical dipstick detectors have the advantage to measure large samples in the Marinelli beaker geometry. Well-type Ge detectors can handle test portion volumes of up to approximately 8 mL, and thus have their own niche in large sample NAA, especially as an addition to enhance sensitivity for test portions in the grams range.

In general, it is preferable to count the large sample by placing it at a certain distance from the detector end cap to minimize complications in the efficiency calculations, in particular in the coincidence summing corrections. The distance between the test portion and detector is guided by the sample activity; the higher the activity, the farther the sample to be placed from the detector. Automatic sample changing can be realized irrespective of the sample size. Sample changers are already commercially available for containers with volumes up to 1 L, although they have been designed for the Marinelli beaker measurement.

Table 3 Irradiation facilities needed for test portions of different masses

	Test portion mass (g)	Type of facility
1	1–10	Existing pneumatic facility
2	10–100	Existing/special pneumatic facility
3	100–1000	Existing manual loaded or special facility such as the thermal column of a reactor
4	>1000	Often a new special facility

Table 4 Details of some of the irradiation facilities used for LSNAA[37]

Institute	Nation	Reactor type	Test portion mass	Facility (TC = thermal column)	Neutron fluence rate ($cm^{-2}\,s^{-1}$)
Dalhousie University	Canada	SLOWPOKE	30 g	Rabbit system	2.5×10^{11}
International Centre for Environmental and Nuclear Sciences University of the West Indies	Jamaica	SLOWPOKE	30 g	Rabbit system	2.5×10^{11}
Atominstitut	Wien, Austria	TRIGA	5 g	Fast and normal Rabbit system	2×10^{12}
FRG-II	Munich, Germany	TRIGA	1 kg	Manual loading	6×10^{9}
Delft University of Technology, Reactor Institute Delft	Netherlands	Swimming pool reactor	50 kg	Manual loading, TC	3×10^{8}
BARC, Mumbai	India	Swimming pool reactor, Apsara	1–4 kg	Manual loading, TC	2×10^{8}
Demokritos	Greece	Swimming pool reactor	2 kg	Manual loading, TC	5×10^{6}
Institute of Nuclear Physics	Kazakhstan	Pool type reactor	10 × 100 mL	Manual loading, core	Low power operation

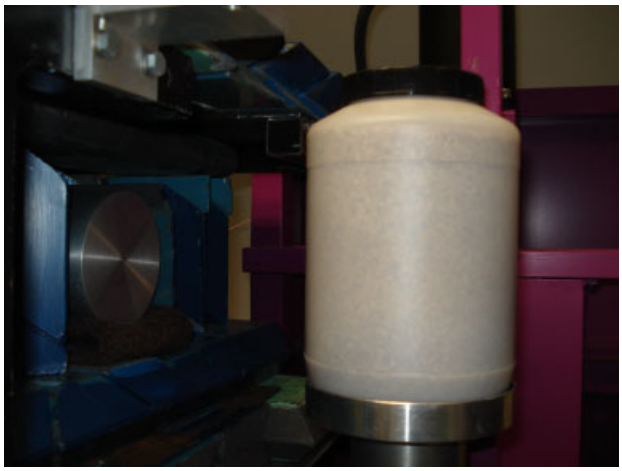

Figure 3 Large sample container (diameter 12 cm, height 20 cm) in front of side-looking Ge detector (cryostat diameter 9 cm).

Table 5 Typical masses of test portions and required bottles, indicating the reduction of the ratio of masses in the case of large sample NAA

Bottle volume (mL)	Bottle mass	Sample mass	Sample mass/ bottle mass
0.5	0.2 g	0.2–0.4 g	1–2
50	10	30–60	3–6
100	15	60–120	4–8
250	25	150–300	6–12
500	37	300–500	8–15
1000	100	600–1000	6–10

Adequate shielding of the stored samples remains, of course, a prerequisite.

Different detector calibration approaches are needed – taking into account voluminous photo peak efficiency, γ-ray self-attenuation, and coincidence summing correction if Marinelli beaker geometries or multisample containers are applied. In the multisample container setup, as the geometry for each container is the same, one can start with one container and measure up to 20 samples according to their decay.

The spectrometer may have to be equipped with a separate device to allow the determination of the effective γ-ray linear attenuation coefficients using a multi-γ-ray pencil beam of an external source (e.g. ^{152}Eu). An example of such a setup is given in Figure 5.

It may be necessary to collimate the detector in case the localization of the inhomogeneity is the subject of interest in large sample analysis (see Figure 6). Such a setup will allow, in principle, for emission tomography of the activated sample.[38]

More advanced spectrometer systems may be designed, in which two or more Ge detectors surround the large sample to create nearly 4π geometry of the detectors.[29]

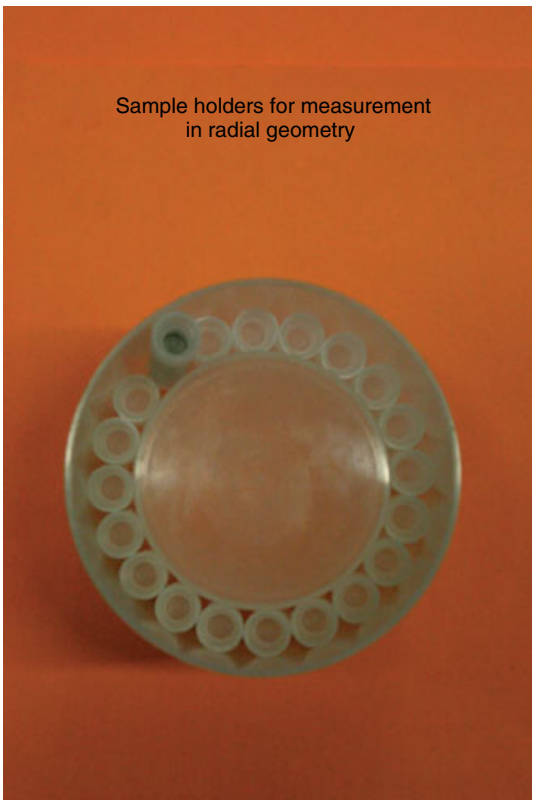

Figure 4 A multisample container.[37]

The individual spectra can be added later to create a composite spectrum with better statistics.

The γ-ray spectrometer should be equipped with dedicated high-count-rate electronics if the induced radioactivity to be measured would result in a count rate $>20\,000\,s^{-1}$. The prerequisites are a transit–reset preamplifier and pulse processing electronics (analog or digital), allowing for on-line dead-time compensation, e.g., based on the loss-free counting principle.[39] The latter is relevant only if count rates vary significantly during the counting time, as may be the case in counting radionuclides with very short half-lives (in the range of seconds to several minutes).

It is interesting to take note that there are a few applications where counting could be carried out using well-type scintillation detectors. In the cases where major components only have to be analyzed and the neutron sources are isotope based, then the resulting activity could be low and the γ-ray spectra might be relatively simple, which could be measured using high-efficiency scintillation detectors. Large sample NAA with isotopic sources and the use of scintillation detectors may be considered as an additional opportunity to further enhance the sensitivity. One may even construct a simple 4π detector by putting two well-type detectors against each other.

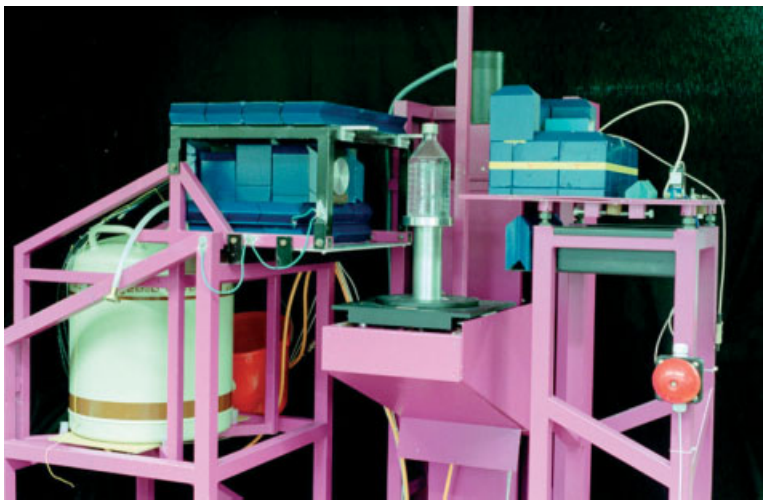

Figure 5 Large sample γ-ray spectrometer with (from left to right) the shielded side-looking Ge detector, the sample on the rotating turntable, and the lead shield with the ^{152}Eu source for γ-ray transmission. The source can be pneumatically moved in front of a point-source collimator.[1]

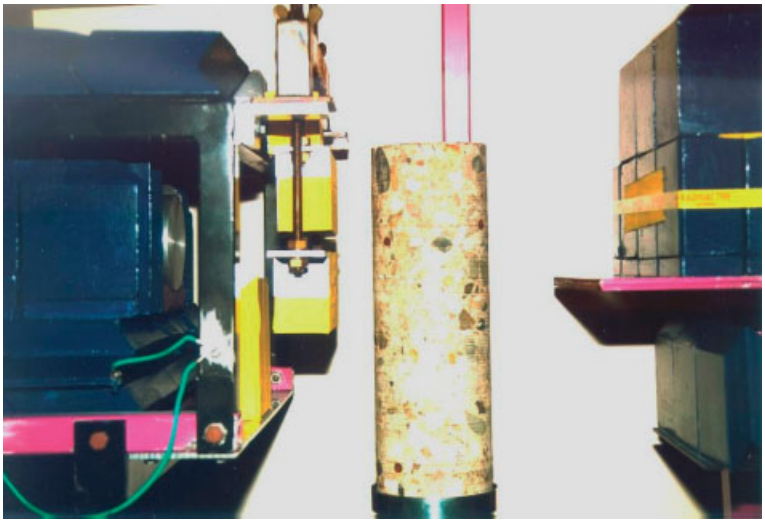

Figure 6 Close-up view of the γ-ray spectrometer showing (from left to right) the shielded Ge detector, a slit collimator, a (simulated) large sample, and the lead shield for the transmission source.[1]

5 CALIBRATION

Determination of the elemental masses in large sample NAA may be done as in small sample NAA, via

- the absolute method,
- the comparator method, or
- the internal standard method.

5.1 Absolute Method

The absolute method for standardization in NAA is based on using known values for neutron fluence rate and activation cross sections, derived either from previously performed measurements or from reactor physics estimations (neutron fluence rate) and from the literature data (cross sections); the same is the case with θ, N_{Av}, M, σ_{eff}, γ, and λ. For many (n,γ) reactions and radionuclides, the parameters σ_{eff}, γ, and λ are not precisely known, while in some cases θ also is not known accurately. Since the various parameters are often achieved via independent methods, their individual imprecisions will add up in the calculation of the elemental amounts, leading to large systematic errors. This method is best applicable if the composition of the sample matrix is well established in advanced, as is the case, e.g., when dealing with pure materials. It is well known that even in normal NAA this approach may not yield highly accurate data but still the results may be adequate

for the intended purpose. Still, additional estimates are needed for correcting the neutron fluence rate gradient and the γ-ray self-absorption. The first may be done using simplified models or using neutron transport codes, and the second after simple transmission measurements or using tabulated linear attenuation coefficients in case the sample is well defined with respect to its composition.

5.2 The Comparator Method

The test portion is irradiated together with a calibration sample containing a known amount of the element(s) of interest. The calibration sample is measured under (preferably) the same conditions as the sample (sample-to-detector distance, sample size, and if possible composition). From a comparison of the net peak areas in the two measured spectra, the mass(es) of the element(s) of interest can be calculated (see above, Equations (1) and (2)):

$$mass_{\mathrm{unk}} = mass_{\mathrm{std}} \frac{(A_{0,\mathrm{x}})}{(A_{\mathrm{s}})} R_\theta R_\phi R_{\mathrm{En}} R_\sigma R_{\mathrm{nss}} R_\varepsilon R_{\gamma\mathrm{ss}} R_{\mathrm{inh}} \qquad (3)$$

The relative standardization based on element standards is not immediately suitable for laboratories aiming at the full multielement powers of INAA. It is virtually impossible to produce a multielement standard containing known amounts of all 70 detectable elements with sufficient accuracy in a volume closely matching the size and the shape of the samples. For this reason, some laboratories prefer to use (certified) reference materials as a multielement standard. However, if dealing with large samples (gram to kilogram size), the use of (certified) reference materials is not practical (major differences may exist in neutron spectrum characteristics in the real sample and in the standard), economical (reference materials are expensive and large sample analysis would imply a high consumption rate), or ethical (certified reference materials are produced for method validation purposes, not necessarily for calibration).

Multielement INAA based on the relative standardization method is feasible when performed according to the principles of the single comparator method. Assuming that all relevant experimental conditions are stable over time, standards for all elements are co-irradiated, each in turn with the chosen single comparator element. Once the sensitivity for all elements relative to the comparator element has been determined (expressed as the so-called k-factor, see Section 5.3), only the comparator element has to be used in routine measurements instead of individual standards for each element.

5.3 Single Comparator Method

Originally, the single comparator method for multielement INAA was based on the ratio of proportionality factors of the element of interest and of the comparator element after correcting for saturation, decay, counting, and sample weights. Girardi et al.[40] defined the k-factor as

$$k = \frac{M_{\mathrm{t}} \gamma_{\mathrm{c}} \varepsilon_{\mathrm{c}} \theta_{\mathrm{c}} \sigma_{\mathrm{eff,c}}}{M_{\mathrm{c}} \gamma_{\mathrm{t}} \varepsilon_{\mathrm{t}} \theta_{\mathrm{t}} \sigma_{\mathrm{eff,t}}} \qquad (4)$$

in which the subscripts "t" and "c" refer to the element of interest in the sample and comparator, respectively. Mass fractions can then be calculated from these k-factors; for an element determined via a directly produced radionuclide, the mass fraction ρ follows from

$$\rho = \frac{(\mathrm{A/SDCw})_{\mathrm{t}}}{(\mathrm{A/SDCw})_{\mathrm{c}}} \cdot k \qquad (5)$$

where $S = (1 - \mathrm{e}^{-\lambda . t\mathrm{ir}})$

These experimentally determined k-factors are often more accurate than those calculated on basis of literature data as in the absolute standardization method. However, the k-factors are valid only for a specific detector, a specific counting geometry, and the irradiation facility and remain valid only as long as the neutron flux parameters of the irradiation facility remain stable.

The main problem of the single comparator method in LSNAA is that differences in neutron exposure, γ-ray attenuation, and volumetric counting efficiency between comparators and samples all have to be accounted for. In normal NAA, most of these differences can be neglected, also because the comparators are co-irradiated with the samples. But in LSNAA this may be practically impossible; the irradiation facility may not be spacious enough, and substantial difference may exist in the neutron exposure and flux gradients. Only if the neutron flux spectrum is well known – as in TCs – and the neutron fluence rate gradient can be established in each individual test portion, the comparator method provides an opportunity for applicability. In that case, even the k_0-based method for standardization may be applied.

The comparator method is, however, very well usable for large sample analysis of, e.g., large liquid samples such as water or oil since standard samples with element spikes into a similar matrix can easily be prepared.

5.4 k_0-Based Method for Standardization

At the Institute for Nuclear Sciences in Ghent, Belgium, an attempt has been made to define k-factors that should be independent of neutron flux parameters as well as spectrometer characteristics.

The expression for the activation reaction rate can written as

$$R = \Phi_{th}\sigma_0 + \Phi_{epi} I_0(\alpha) \tag{6}$$

The ratio f of the thermal neutron flux and the epithermal neutron flux is $f = \Phi_{th}/\Phi_{epi}$ and the ratio of the resonance integral and the thermal activation cross section can be expressed as $Q_0(\alpha) = I_0(\alpha)/\sigma_0$; thus the effective cross section is

$$\sigma_{eff} = \sigma_0 \left(1 + \frac{Q_0(\alpha)}{f}\right) \tag{7}$$

The k_0-factor is now defined as

$$k_0 = \frac{1}{k}\frac{1 + (Q_{0,c}(\alpha)/f)}{1 + (Q_{0,t}(\alpha)/f)}\frac{\varepsilon_c}{\varepsilon_t} = \frac{M_c\theta_t\sigma_{0,t}\gamma_t}{M_t\theta_c\sigma_{0,c}\gamma_c} \tag{8}$$

and the mass fraction, again for an element determined via a directly produced radionuclide, is found from

$$\rho = \frac{1 + (Q_{0,c}(\alpha)/f)}{1 + (Q_{0,t}(\alpha)/f)}\frac{\varepsilon_c}{\varepsilon_t}\frac{(\text{A/SDCw})_t}{(\text{A/SDCw})_c}\frac{1}{k_0} \tag{9}$$

The k_0-factor has thus become a purely nuclear parameter for the thermal neutron spectrum. In the k_0 convention, Au is proposed as the comparator element. The neutron flux parameters f and α no longer cancel out in concentration calculations and must be measured in each irradiation facility, preferably even for each irradiation and sample.[41] The k_0-factors are used in Delft for the analysis of very large samples.

5.5 Internal Monostandard Method

In the internal monostandard method, one of the radionuclides produced during activation of the test portion is used as monostandard. The rationale behind this is that the effect of neutron spectrum perturbation is the same for this parent element of this radionuclide as well as for all other elements in the sample; as such there is an implicit assumption that the test portion is ''macroscopically'' homogeneous.

In the case of internal monostandard method using TC irradiations followed by γ-spectrometric method, the ratio of mass (m) of an element (x) in the test portion (t) to mass of the internal comparator element (c) in the sample is given by Equation (10):

$$\frac{m_t}{m_c} = \frac{\left((\text{S.D.C}) \cdot (f + Q_o(\alpha))\right)_c}{\left((\text{S.D.C}) \cdot (f + Q_o(\alpha))\right)_t} \cdot \frac{A_{At}}{A_{Ac}} \cdot \frac{\left(\varepsilon_\gamma\right)_t}{\left(\varepsilon_\gamma\right)_c} \cdot \frac{1}{k_{0,c}(x)} \tag{10}$$

Here, $k_{0,c}(x)$ represents the relative sensitivity of element x with respect to y and is calculated from the $k_{0,Au}$ factors in the literature.

The internal monostandard method either results in elemental mass ratios (element of interest vs. monostandard element) and thus may serve for comparative studies, or, in the case of, e.g., materials of high purity and known stoichiometry, directly into mass fractions of the elements of interest if (one of) the major component(s) is used as the monostandard element.

5.6 Neutron Fluence Rate Monitoring

Neutron fluence rate monitoring is needed for the k_0 method of standardization and may be done noninvasively with the flux monitors outside the sample, as well as by inserting flux monitors inside the sample. The first approach may use, e.g., the neutron depression outside the large sample to estimate the neutron flux distribution;[42] the second approach provides direct information on the flux distribution.[43,44] The first approach is applied in Delft for large sample irradiations in the reactor's TC.[5] During the irradiation, the sample is surrounded by four flux monitors at any desired height around the sample. Since the unperturbed neutron flux gradient in the TC can be derived from the irradiation of a pure graphite sample, the neutron flux depression outside the sample can be estimated (Figure 7). This forms the basis for estimating the effective neutron diffusion length and neutron diffusion coefficient. Finally, the overall correction factor is calculated – which reflects how the large sample compares to a small sample – using the volume efficiency of the Ge

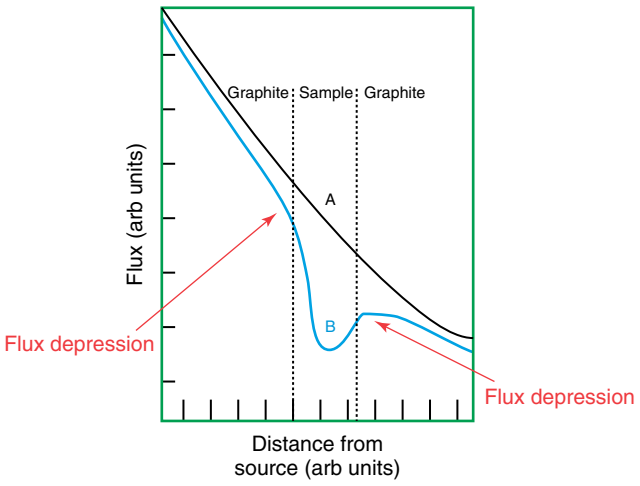

Figure 7 Neutron flux depression outside a large sample placed inside the graphite-filled thermal column. The curve A indicates the unperturbed flux gradient; curve B schematically shows the flux gradient if a sample absorbing and scattering thermal neutrons is placed in the irradiation position.

detector, the neutron diffusion length and coefficient, and the effective linear attenuation coefficients.

Changes in the neutron spectrum due to self-thermalization are much more difficult to deal with, as these are not easy to monitor. The extent of this effect depends, of course, on the neutron spectrum shape and the fraction of epithermal and fast neutrons compared to the thermal neutrons. In TC facilities, the ratio of thermal over nonthermal neutrons may be much larger than a factor of 1000, eliminating the significance of neutron self-thermalization. The user of the irradiation facility should be familiar with this phenomenon, and a priori information must always be collected about the sample composition so as to estimate the extent of these effects and to decide if empirical correction factors can be applied or if additional in situ monitoring is needed (i.e. invasive, by inserting suitable monitors inside the sample).

5.7 γ-Ray Self-Attenuation

The γ-ray self attenuation correction is relatively easy to establish once the effective linear attenuation coefficients are available, either by measurement or by calculation from the approximate (or well-known) elemental composition. A multi-γ-ray emitting source, with γ-ray energies distributed over the entire range of interest (such as ^{152}Eu, ^{182}Ta) can be used for this. A nearly pencil beam geometry can be created by locating this source behind a pinhole collimator, and the γ-ray transmission can be measured at several heights along the sample. This forms the basis for the estimation of the effective linear γ-ray attenuation coefficients.

The volumetric photopeak efficiency can be determined by Monte Carlo modeling, but this requires precise information about the inner geometry of the cryostat and detector configuration (including the dead-layer thickness). Empirical curves may be determined using standard sources in water, as the self-attenuation can simply be subtracted from the measured efficiency. In situ relative detection efficiency in a voluminous sample was determined using the multi-γ-emitters produced in the sample, and was adequate to calculate mass ratios with respect to the comparator using Equation (5).[45]

Pragmatic approaches have been suggested in which the large sample, after activation, is repacked into many small-diameter containers that are placed in a cylindrical holder surrounding the detector. If the detector crystal is perfectly symmetrically mounted inside the cryostat, the detection efficiency for each of the positions around the detector is the same, which simplifies the calculations. Besides, the dimensions of the small containers can be chosen such that γ-ray attenuation effects may be neglected.

It should be noted that the γ-ray spectrum due to the natural radioactivity of the sample material has also to be measured in large sample analysis, prior to the activation. These "sample background" peaks in the γ-ray spectrum should be separately treated later on in the neutron and γ-ray self-attenuation corrections.

5.8 Extreme Inhomogeneities

Combination of correction algorithms for neutron and γ-ray self-attenuation as well as for the volumetric photopeak efficiency yields an "overall correction factor", which reflects the difference between the actual detector response for a given γ-ray energy and the theoretical detector response if the sample were a massless point source located in the large sample's center, without any neutron and γ-ray attenuation.[9]

In these corrections, it is assumed implicitly that both trace elements and major (matrix) elements in the sample are homogeneously distributed on a macroscopic scale. If this condition is not met, there is a high probability that owing to the neglect of inhomogeneities the concentrations determined are not correct. How large these deviations may be as a result of such neglect has been studied via computer simulations. Inhomogeneities may influence the results of the irradiation as well as of the measurement; therefore, both have been treated separately. Inhomogeneous matrix composition has been modeled by composing a sample of cylinders with strongly differing neutron or γ-ray attenuation properties. Inhomogeneity for trace elements has been simulated by modeling extreme distributions of a trace element with either neutron or γ-ray attenuation properties, strongly differing from those of the main composition of the sample.[20,46] Both inhomogeneities in matrix composition (e.g., layered structures) and trace-element inhomogeneities (e.g., "hot spots") were taken into account.

Obviously, the results of these simulations demonstrated that false concentrations may be obtained if inhomogeneities are not accounted for in the interpretation step of large sample INAA. The smallest errors may occur for matrix inhomogeneities; the most pronounced effects can be expected when the trace element of interest is distributed either on the outside or on the axis of the cylindrical sample.

In these simulations, materials or elements were selected with neutron and γ-ray attenuation properties that were strongly different from the rest of the sample; as such they may be considered as "extremities", and the consequences on the inaccuracy of the results rather indicate "worst case" conditions.

5.8.1 *Determination of Inhomogeneities*

The presence of extreme inhomogeneities in large samples may be considered a nuisance. On the other hand, large sample analysis is a unique tool for determining these inhomogeneities without destroying the test portion. To this end, sample scanning using a collimated detector has been introduced, and it is assumed that the sample consists of volume elements that individually are considered homogeneous.

The set of spectra constituting one scan is statistically evaluated to determine whether fluctuations over the scan of the count rates of γ-rays of a given energy are only due to Poisson counting statistics or are also due to inhomogeneities in the sample. If inhomogeneities have only a layered structure in the direction of the cylinder axis of the sample, the sample can be analyzed layer by layer, and for the most extreme cases the analysis can be performed for each voxel separately. An example of a collimated detector system for large sample scanning is shown in Figure 6.

Baas et al.[38,47] developed a method for the detection of local inhomogeneities. The count rate in each channel of each segmented measurement can therefore be compared with the average count rate in each channel after summing all individual measurements. Such a comparison, taking into account uncertainties, is made analogous to the ζ-score principle. Values of $|\zeta| > 2$ or > 3 (depending on the analyst's fitness-for-purpose criterion) indicate a local inhomogeneity at the respective γ-ray energy in a particular voxel. This approach is visualized in Figure 8. A bottle of approximately 25 cm length and 8 cm diameter filled with soil was irradiated in the large sample facility in Delft. The induced radioactivity was measured with a 96% Ge detector, collimated with a 10-cm-thick Pb collimator with a 2-cm split opening. The figure shows the z' scores for each channel number (γ-ray energy) along the height of the sample. From the zoomed details it can clearly be seen that strongly deviating ζ-scores occur at, e.g., 439 and 1115 keV, indicating an inhomogeneity for zinc. In addition, the histogram of all ζ-scores in this figure provides also an insight into the presence of this inhomogeneity.[47]

6 QUALITY CONTROL

The high degree of accuracy in normal activation analysis results from decades of experience in the development of certified reference materials. Many sources of error and the quantification of their impact are known.[48,49] Methods commonly referred to as *quality control practices* have been developed to inspect the occurrence of errors during the analysis, whereas implementation of quality assurance contributes to minimizing and even avoiding the occurrence of errors. The known sources of error in normal activation analysis may occur in large sample analysis too. Some of them – such as γ-ray self-attenuation and neutron/photon fluence rate, or neutron spectrum gradients – have much larger effects. Extreme inhomogeneities are an additional phenomenon in large sample analysis,[46] with an impact on the degree of accuracy.

Quality control in normal activation analysis includes the simultaneous analysis of well-characterized quality

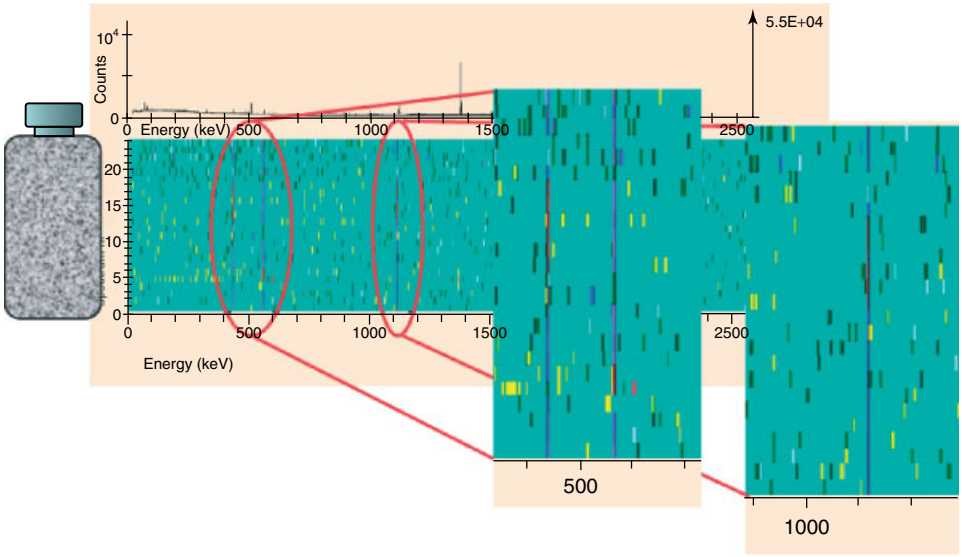

Figure 8 Energy- and position-correlated ζ-scores (see text) of measured radioactivity, indicating location of inhomogeneities in a sample of 20 cm height and 10 cm diameter.

Table 6 Opportunities of quality control measures traditionally applied in normal activation analysis for samples of larger sizes

	Quality control samples	Blanks	Duplicates	γ-Ray intensity ratios and multiple radio nuclides
1 g	Y	Y	Y	Y
10 g	N	Y/l.r.	Y/n.r.	Y
100 g	N	l.r.	n.r.	Y
1 kg	N	l.r.	n.r.	Y

Y = yes, application possible; N = no, not possible; l.r. = less relevant (see text); n.r. = not relevant.

control samples, blanks, and sometimes duplicates. In addition, inspection of the intensity ratios of γ-ray peaks of a given nuclide and/or the quantified results based on different radionuclides formed from a given element also provide a unique opportunity to inspect for errors. The applicability of these quality control approaches for samples of increasing mass is given in Table 6.

It is clear that basic problems emerge when extending the traditional approaches to samples with weights of more than a few grams. Firstly, well-characterized control samples of the size of large samples (several grams to kilograms) are either very expensive to use or not available at all. Secondly, large sample analysis may be required because of the heterogeneity of the object, which cannot be simulated by a control sample even if it were available. Thirdly, duplicates – assuming identical composition in mass fraction and in degree of homogeneity – may probably not be available when larger sample masses are needed. The problem related to the blank – impurities in the sample container and/or contamination – has, on the contrary, a smaller impact on the final result since the increase in the ratio of sample mass to container mass may result in negligible contribution of the blank (Table 5).

New strategies have to be developed to control the analytical quality in large sample analysis. One of the opportunities is to continue with the use of performance indicators, derived from the actual sample analyzed. In fact, this is not different from most quality control procedures in, e.g., manufacturing and production, in which the quality of a final product depends on predefined specifications being met, such as dimensions, tolerances, mass, color, or operation characteristics. The inspection of γ-ray intensity ratios and the use of different radionuclides of one element are already examples of such a form of quality control in activation analysis. This approach can further be extended to other sample/material characteristics on the basis of physical sample properties such as γ-ray self-attenuation and neutron attenuation parameters, as well as via the degree of inhomogeneity (Table 3).

6.1 Quality Control in Large Sample Analysis

Some materials may be difficult to be processed to such homogeneity that representative subsamples can be taken at the <1 g level. For such materials, it may be advantageous to analyze much larger quantities without homogenization and to assume that the inhomogeneities are randomly distributed throughout the sample, so that the entire quantity can be considered as homogeneous. However, this assumption has some limitations. Overwater and Bode demonstrated the impact of extreme inhomogeneities on the correction mechanisms for the attenuation of γ-ray attenuation and neutron self-shielding.[46] Inhomogeneities with strong γ-ray absorbing properties have stronger effects on the degree of accuracy than those with strong neutron absorbing properties. It is therefore relevant to inspect for the presence of such extreme inhomogeneities in order to decide on the value of the finally calculated mass fractions. Two opportunities to inspect such inhomogeneities are given here.

The effective linear γ-ray attenuation coefficient is usually determined by measuring the transmission of γ-rays of different energies emitted by an external source with known emission rates.[9,45] The values of the effective linear attenuation coefficients may be estimated using the tabulated values for the elements. Typically, for example, biological and geological material values of $\sim 0.15\,\mathrm{cm}^{-1} < \mu < \sim 0.60\,\mathrm{cm}^{-1}$ at ~ 100 keV; $\sim 0.12\,\mathrm{cm}^{-1} < \mu < \sim 0.25\,\mathrm{cm}^{-1}$ at ~ 300 keV; and $\sim 0.05\,\mathrm{cm}^{-1} < \mu < \sim 0.15\,\mathrm{cm}^{-1}$ at ~ 1000 keV can be found. A ''bandwidth'' of the linear attenuation coefficient can thus be determined at different γ-ray energies for different types of materials (e.g. environmental, geological) (Figure 9). This can assist in inspecting whether the experimentally determined attenuation coefficients have realistic values. Moreover, if scanned measurements are carried out, an indication of local (layer-type) inhomogeneities with strong γ-ray absorbing properties may already be obtained. An example of this approach is given in Figure 10. The transmission of the γ-rays of a ^{152}Eu source was measured at different heights along a $\sim 1-$m long, $\sim 15-$cm diameter soil drill core sample prior to neutron activation. The effective linear attenuation coefficients for each of the γ-ray energies were fitted with a polynomial so as to estimate the effective linear γ-ray attenuation at other γ-ray energies. The attenuation coefficients at 100, 300, and 1000 keV all fall within the expected bandwidths. There are no indications in this example for layer inhomogeneities with strongly differing γ-ray absorbing properties.

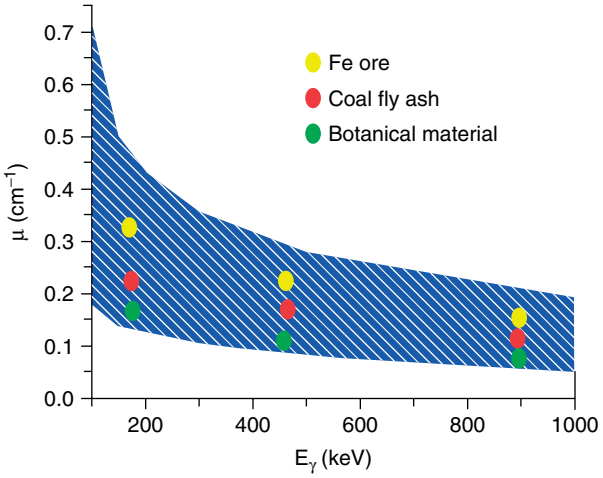

Figure 9 Range of linear γ-ray attenuation coefficients for typical materials.[50]

Additionally, the methods developed by Baas et al. (described in the preceding text) can be used for the detection of local inhomogeneities. It can now be decided on a case-to-case basis whether such inhomogeneities have any unwanted impact on the final analysis result.

The correction for neutron self-shielding in LSNAA may be made using information derived from the neutron fluence rate depression at positions in the irradiation facility just outside the sample. Overwater and Hoogenboom[42] developed this approach to estimate the thermal neutron diffusion length L_s and the thermal neutron diffusion coefficient D_s, which subsequently were used to reconstruct the neutron fluence rate profile inside the large sample. Both L_s and D_s are physical element properties and, similar to the effective linear γ-ray attenuation coefficient, boundaries can be estimated for the values of these two parameters in real materials. Though strongly correlated ($L_s^2 = D_s/\Sigma_a$, with Σ_a the macroscopic absorption cross section (cm^{-1})), typical values are $\sim 1\,cm < L_s < \sim 20\,cm$ and $\sim 0.2\,cm < D_s < \sim 3\,cm$. An example is given in Figure 4. A ~1-m long, ~15-cm diameter water basin sediment drill core was analyzed in the frame of a pollution research project. Zinc foils were used as neutron flux monitors. The monitors were positioned just outside the sample container to monitor neutron fluence rate depression by comparison with the neutron fluence rates as monitored in a separate irradiation with a solid graphite sample, thus simulating the unperturbed neutron flux. The calculated values of L_s and D_s at different heights are plotted within the bandwidths for these values (Figure 11). Also, the average neutron fluence rates as a function of sample height is given. The fluence rates can be fitted with a cosine function, reflecting the flux distribution within the reactor's TC.

Assuring the quality of the results requires insight, monitoring, and control of the sources of error. Quality control procedures as traditionally applied in chemical analysis are not fully applicable in large sample analysis.[50] One of the advantages of activation analysis is that some of the measured sample parameters dealing with γ-ray and neutron attenuation can only vary in ranges set by well-known values of elemental constants. These parameters can be much more easily determined in large sample analysis than with samples in the milligram range, thereby offering an outlook for direct verification of the quality of the related correction algorithms.

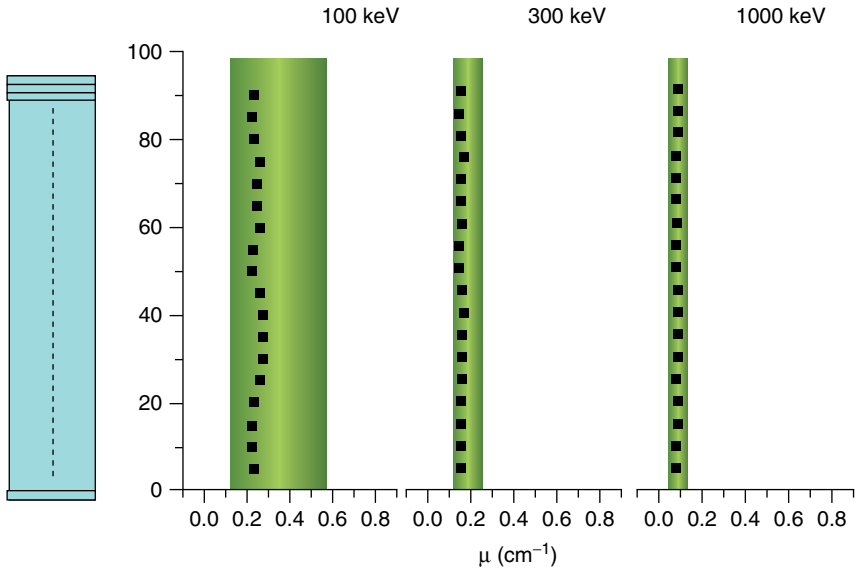

Figure 10 Measured linear attenuation coefficients at 100, 300 and 1000 keV of a 15 cm thick soil drill core.

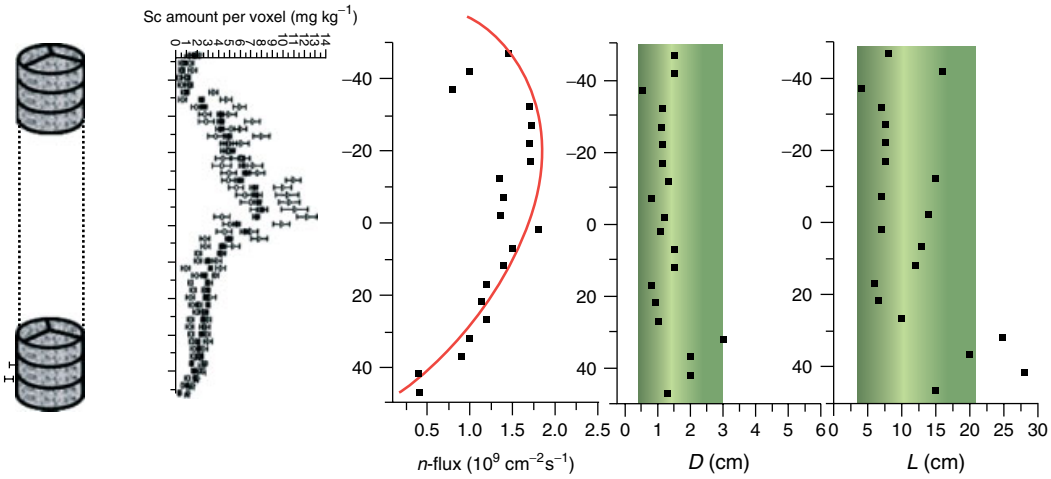

Figure 11 Results of collimated large sample analysis of a stack of soil samples, total of 1 m length, 12 cm diameter, showing (left to right) Sc amount per 120° voxel, neutron flux gradient variations in neutron diffusion coefficient and neutron diffusion length; the band widths indicate the theoretical boundaries of these parameters.[50]

7 SENSITIVITY

Sensitivity is defined as the gradient of the response curve: i.e., the change in instrument response that corresponds to a change in analyte concentration. This definition translates in INAA into the net peak area as a function of the analyte mass. Larger peak areas at a given sample mass can be obtained by the following:

- higher neutron fluxes and longer irradiation times
- more efficient detectors
- larger sample masses.

Higher neutron fluxes and longer irradiation times are often not easily attainable, as the first is limited by reactor design, whereas longer irradiation times only have a positive effect on the sensitivity for radionuclides with very long half-lives.

Absolute photopeak efficiencies of detectors of different sizes and for different geometries are given in Table 7, and a comparison of the absolute efficiencies of a regular 17% coaxial and one of the largest well-type detectors reported in the literature[51,52] is given in Figure 12. It should be noted that the use of Compton suppression shields does not result in an increase in sensitivity – as often erroneously suggested; after all, the signal resulting from the induced radioactivity does not increase and sometimes even decreases due to summing-out effects or larger sample-detector distances. Compton suppression systems find their advantage in a decrease in uncertainty of measurement due to decrease of the background under a peak.

Large sample masses can compensate for low neutron fluxes. As the limiting factor in NAA is merely the maximum acceptable induced radioactivity upon

Table 7 Absolute photopeak efficiencies at some photon energies for 3 detector types and different geometries[51,52]

Absolute photopeak efficiency (%)					
		1332 keV	750 keV	500 keV	250 keV
20%	0 cm	1.5	2.5	3.2	5
	5 cm	0.22	0.35	0.5	1
100%	0 cm	7.5	10	12	16
	5 cm	1	1.4	1.7	2.4
125 cm^3	Well	5	9	13	25

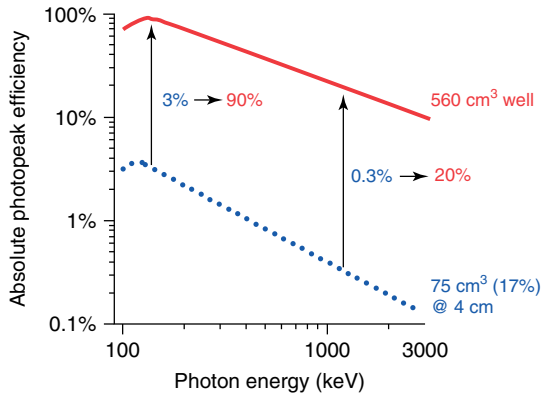

Figure 12 Schematic absolute full energy photopeak efficiency curves for a regular coaxial Ge detector (17% relative efficiency) and a very large well-type detector.

counting, a comparison has been made with normal INAA in which a hypothetical 200-mg sample is irradiated for a given time at a neutron flux of 10^{13} cm^{-2} s^{-1}. Assuming a certain neutron fluence rate and a cylindrical sample of similar length and diameter, and an average density of 0.5 g cm^{-3}, a first estimate can be made of the minimum

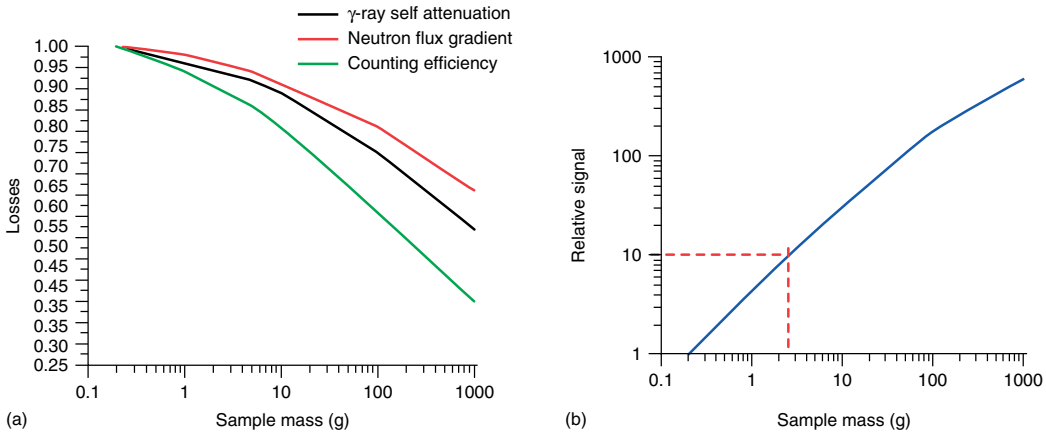

Figure 13 (a) Losses due to neutron and γ-ray self-shielding and enlarged source-to-detector distance, compared to an analysis of a 200-mg powdered siliceous sample and (b) net gain in signal at increased sample masses.

Table 8 Indication of minimum sample mass at a given neutron fluence rate to attain similar induced radioactivity as a 200-mg sample, irradiated at 10^{13} cm^{-2} s^{-1}

Neutron fluence rate (cm^{-2} s^{-1})	Sample size (L, D) (cm)	Sample mass (g)
2.10^{12}–2.10^{11}	1–2	1–10
2.10^{11}–2.10^{10}	2–4	10–100
2.10^{10}–2.10^{9}	4–7	100–1000
$<2.10^{9}$	>7	>1 kg

mass needed to reach the same sensitivity as for the small sample at a high neutron fluence rate (Table 8).[35]

However, these results must be corrected for the losses due to neutron self-shielding and γ-ray self-attenuation, and the fact that the center of voluminous samples, simply for physical reasons, is always positioned further away from the detector than in case of a small sample. In Figure 13 these effects have been combined, and it can be derived that at a given neutron fluence rate a net gain in signal of a factor of 10 can be obtained by increasing the sample mass by a factor of 15 (e.g., from 200 mg to 3 g).

7.1 Natural Background

A large sample NAA procedure should start with measurement of the natural radioactivity of the sample, as the corresponding peaks in the γ-ray spectrum should be separately treated later on in the neutron and γ-ray self-attenuation corrections.

8 APPLICATIONS

LSNAA has evolved over the years towards a capability for various samples types that otherwise would have been difficult to analyze. Accordingly, the developments faced many challenges, each one different from the other, and in this process its horizon of application to various fields has increased: e.g., nutritional studies, geology, biology, archaeology, waste analysis, high-purity material characterization, precious samples, and liquid samples, from all walks of life. A few of them are given in the following text.

8.1 Materials Difficult to Homogenize: Geological Samples, Ores, and Waste

LSNAA is highly suitable for the analysis of heterogeneous geological material such as rocks, coal (determination of quality), ores, and mineral concentrates. Conveyor belt monitoring of elemental concentrations by PGNAA in coal and cement raw materials has lead to an increased efficiency of coal-fired power plants and cement factories. It appears that this technique has the required potential to trigger industrial processes and gives reliable results.

Waste material in many cases is considered to contain hazardous substances whose behavior could result in their entry into biosphere through the atmosphere or ground water. Therefore, appropriate classification of the waste material is required in order to ensure safe disposal or further treatment and recycling. Construction material, domestic and electronic waste, as well as contaminated sediment and compost material are considered to be highly heterogeneous and therefore elaborate sampling procedures are required if representative sampling of these materials for analysis is needed. LSNAA has been effectively used to analyze large samples of soil,[53] electronic waste,[54,55] and other materials. These are the materials in which both subsampling and homogenization, steps not required in LSNAA, are very difficult and cost intensive. A typical example is the elemental composition analysis of waste from an incineration plant. This type of

sample cannot be easily homogenized and analyzed by other analytical techniques. However, a 1-kg portion of this waste was analyzed by using the LSNAA without homogenization.

8.2 Materials That May be Contaminated During Homogenization: High-Purity Materials

LSNAA is an extremely useful technique to analyze the metals and alloys for impurities. It is being used to analyze various finished products of alloys (Zircaloy 2, Zircaloy 4, SS-316M (stainless steel), and 1S aluminum) that are used in reactor technology[56] and for impurities in high technological materials such as silicon and superalloys.[3] The biggest advantage is that information is obtained on the entire specimen, and because of the absence of subsampling, contamination can be minimized.

8.3 Materials Difficult to Subsample: Nutritional Studies

LSNAA is extremely useful in the determination of major, minor, and trace elements in foodstuff, as large samples can be analyzed without resorting to subsampling. In fact, comparison of the results obtained from subsamples of varying mass gave an indication that a sample of 1 kg of wheat is more representative than small samples in the range of 40–1000 mg.[57] It is also feasible to determine trace elements in liquid diets, e.g., juices and milk, by LSNAA.

8.4 Valuable Material of Irregular Shape

Subsampling of archaeological and cultural heritage objects is generally prohibited, as these objects have to be preserved intact. LSNAA has the capability for nondestructive bulk analysis of the whole object. In comparison, other established nondestructive analytical methods, such as XRF or analytical techniques based on charged particle irradiation (PIXE, particle induced X-ray emission, IBA, ion beam analysis), can only analyze superficial layers of the sample and provide limited information over the whole volume of the object of interest.[58,59]

The particular advantage of INAA being noninvasive and a true multielemental technique is combined in LSNAA with the ability to analyze bulky objects as a whole, without any visual damage to the valuable cultural heritage objects. Art historians, conservators, and museum staff do not generally allow damaging such valuable objects by removing a portion for analytical purposes. However, sometimes only the elemental composition can decisively distinguish whether an object is different from what it appears to be from visual inspection.

8.5 Other Applications

Direct analysis of large samples provides a unique opportunity for (the validation of) (sub)sampling studies.

REFERENCES

1. P. Bode, 'Instrumental and Organizational Aspects of a Neutron Activation Analysis Laboratory', Ph.D. thesis, Delft University of Technology, 1996, ISBN 90-73861-42-X.
2. B. Kratochvil, J.K. Taylor, 'Sampling for Chemical Analysis', *Anal. Chem.*, **54**, 924A (1981).
3. M.I. Verheijke, H.J.J. Jaspers, J.M.G. Hanssen, 'Neutron Activation of Very Pure Silicon Wafers', *J. Radioanal. Nucl. Chem.*, **131**, 197 (1989).
4. IUPAC, 'Terminology in Soil Sampling', *Pure Appl. Chem.*, **77**, 825 (2005).
5. P. Bode, R.M.W. Overwater, 'Trace Element Determination in Very Large Samples: A New Challenge for Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **167**, 169 (1993).
6. R.M.W. Overwater, P. Bode, J.J.M. de Goeij, 'Large-sample Neutron Activation Analysis: Present Status and Prospects', *J. Radioanal. Nucl. Chem.*, **216**, 5 (1997).
7. R. Gwozdz, F. Grass, 'Activation Analysis of Large Samples', *J. Radioanal. Nucl. Chem.*, **244**, 523 (2000).
8. X. Lin, R. Henkelmann, 'Instrumental Neutron Activation Analysis of Large Samples: A Pilot Experiment', *J. Radioanal. Nucl. Chem.*, **251**, 197 (2002).
9. R.M.W. Overwater, 'The Physics of Big Sample Instrumental Neutron Activation Analysis', PhD thesis, Delft University of Technology, 1994.
10. C.G. Clayton, A.M. Hassan, M.R. Wormald, 'Multi-element Analysis of Coal During Borehole Logging by Measurement of Prompt γ-Rays from Thermal Neutron Capture', *Int. J. Appl. Radiat. Isot.*, **34**, 83 (1983).
11. G. Vourvopoulos, 'Industrial on Line Bulk Analysis Using Nuclear Techniques', *Nucl. Instrum. Methods*, **B56**, 917 (1991).
12. D.M. Franklin, R. Amstrong, D.R. Chettle, M.C. Scott, 'An Improved in vivo Neutron Activation System for Measuring Kidney Cadmium', *Phys. Med. Biol.*, **35**, 1397 (1990).
13. C.S. Lim, B.D. Sowerby, 'On Line Bulk Elemental Analysis in the Resource Industries Using Neutron-gamma Techniques', *J. Radioanal. Nucl. Chem.*, **264**, 15 (2005).
14. R. Doczi, M.A. Ali, M. Fayez-Hassan, J. Csikai, 'Determination of Hydrogen Content in Bulk Samples Using Neutron Activation Analysis', *Appl. Radiat. Isot.*, **63**, 137 (2005).

15. P.A. Beeley, R.G. Garrett, 'Neutron Activation Analysis of Multiple Large Geological Samples by in-pool Irradiation Using a Slowpoke-2 Reactor', *J. Radioanal. Nucl. Chem.*, **167**, 177 (1993).

16. P. Bode, R.M.W. Overwater, J.J.M. de Goeij, 'Large Sample Neutron Activation Analysis: Status and Prospects', *J. Radioanal. Nucl. Chem.*, **216**, 5 (1997).

17. R. Gwozdz, H.J. Hansen, K.L. Rasmussen, H. Kunzendorf, 'Instrumental Neutron Activation Analysis of Samples with Masses from Micrograms to Hectograms', *J. Radioanal. Nucl. Chem.*, **167**, 161 (1993).

18. J.I. Kim, H. Staerk, J. Fiedler, 'A Method of Long-time Irradiation of a Voluminous Liquid Sample in a Reactor Neutron Flux for Activation Analysis of Water', *Nucl. Instrum. Methods*, **177**, 557 (1980).

19. A.G.C. Nair, R. Acharya, K. Sudarshan, S. Gangotra, A.V.R. Reddy, S.B. Manohar, A. Goswami, 'Development of an Internal Monostandard INAA Method Based on in-situ Detection Efficiency for Analysis of Large and Non-standard Geometry Samples', *Anal. Chem.*, **75**, 4868 (2003).

20. F. Tzika, I. Stamatelatos, J. Kalef-Ezra, 'Neutron Activation Analysis of Large Volume Samples: the Influence of Inhomogeneity', *J. Radioanal. Nucl. Chem.*, **271**, 233 (2007).

21. M. Borsaru, R.J. Holmes, P.J. Mathew, 'Bulk Analysis Using Nuclear Techniques', *Int. J. Appl. Radiat. Isot.*, **34**, 397 (1983).

22. M.R. Wormald, C.G. Clayton, 'In-situ Analysis of Coal by Measurement of Neutron-induced Prompt Gamma-rays', *Int. J. Appl. Radiat. Isot.*, **34**, 71–82 (1983).

23. A.T. Farsoni, S.A. Mireshgi, 'Design and Evaluation of a TNA Explosive-detection System to Screen Carry-on luggage', *J. Radioanal. Nucl. Chem.*, **248**, 695 (2001).

24. K. Sueiki, K. Kobayashi, W. Sato, H. Nakahara, T. Tomizawa, 'Non Destructive Determination of Major Elements in Large Sample by Prompt Gamma Ray Neutron Activation Analysis', *Anal. Chem.*, **68**, 2203 (1996).

25. K. Sueki, Y. Oura, W. Sato, H. Nakahara, T. Tomizawa, 'Analysis of Archaelogical Samples by the Internal Monostandard Method of PGAA', *J. Radioanal. Nucl. Chem.*, **234**, 27 (1998).

26. I.H. Degenaar, M. Blaauw, 'The Neutron Energy Distribution to Use in Monte Carlo Modeling of Neutron Capture in Thermal Neutron Beams', *Nucl. Instrum. Methods*, **B 207**, 131 (2003).

27. J.H. Holloway, H. Akkurt, 'The Fixed Point Formulation for Large Sample PGNAA', *Nucl. Instrum. Methods*, **A522**, 529 (2004).

28. I.H. Degenaar, 'Towards a Methodology for Large Sample Prompt Gamma Neutron Activation Analysis', Ph.D. thesis, Delft University of Technology, 2004, ISBN 90-407-2509-8.

29. W. Gorner, S. Alber, A. Berger, O. Haase, G. Manse, C. Segebade, 'BEAMGAA A Chance for High Precision Analysis of Big Samples', *J. Radioanal. Nucl. Chem.*, **263**, 791 (2005).

30. R. Gwodz, 'Instrumental Neutron Activation Analysis of Samples with Volumes from 2 to 350 mL', *J. Radioanal. Nucl. Chem.*, **271**, 751 (2007).

31. S.S. Nargolwalla, E.P. Przybylowicz, *Activation Analysis with Neutron Generators*, John Wiley & Sons, New York, 1973.

32. K.G. Broadhead, D.E. Shanks, 'The Application of 2.8 MeV (D,D) Neutrons to Activation Analysis', *Int. J. Appl. Radiat. Isot.*, **18**, 279 (1967).

33. C.J. Evans, S.J.S. Ryde, D.A. Hancock, F. Al-Agel, 'Monte Carlo Simulation of Prompt Gamma Neutron Activation Analysis Using MCNP Code', *Int. J. Appl. Radiat. Isot.*, **49**, 541 (1998).

34. S.A. Natto, D.G. Lewis, S.J.S. Ryde, 'Benchmarking the MCNP Code for Monte Carlo Modeling of an in vivo Neutron Activation Analysis System', *Int. J. Appl. Radiat. Isot.*, **49**, 545 (1998).

35. 'Large Sample NAA Using Low Flux Irradiation Facilities', Report of a Technical Meeting, IAEA, Vienna, Austria, 7–11 November 2005.

36. R. Khelifi, Z. Idiri, P. Bode, A. Amokrane, 'Flux Calculation in LSNAA Using an Am-Be Source', *J. Radioanal. Nucl. Chem.*, **274**, 639 (2007).

37. R. Gwozdz, F. Grass, 'Application of a Radial Geometry for Counting Samples of Mass 10g, to 300g with a Ge-detector', *Environmental Radiochemical Analysis*, Special Publication No. 234, (G.W.A. Newton, Editor), Royal Society of Chemistry, Cambridge, UK, 289 1999.

38. H.W. Baas, M. Blaauw, P. Bode, J.J.M. de Goeij, 'Collimated Scanning Toward 3D-INAA of Inhomogeneous Large Samples', *Fresenius J. Anal. Chem.*, **363**, 753 (1999).

39. G.P. Westphal, 'On the Performance of Loss-Free Counting – A Method for Real-time Compensation of Dead-Time and Pileup Losses in Nuclear Pulse Spectroscopy', *Nucl. Instrum. Methods*, **163**, 189 (1979).

40. F. Girardi, G. Guzzi, J. Pauly, 'Reactor Neutron Activation Analysis by the Single Comparator Method', *Anal. Chem.*, **37**, 1085 (1965).

41. A. Simonits, F. de Corte, J. Hoste, 'Single Comparator Methods in Reactor Neutron Activation Analysis', *J. Radioanal. Chem.*, **24**, 31 (1975).

42. R.M.W. Overwater, J.E. Hoogenboom, 'Accounting for the Thermal Neutron Flux Depression in Voluminous Samples for Instrumental Neutron Activation Analysis', *Nucl. Sci. Eng.*, **117**, 141 (1994).

43. N.S. Shakir, R.E. Jervis, 'Correction Factors Required for Quantitative Large Volume Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **248**, 61 (2001).

44. K. Sudarshan, R. Tripathi, A.G.C. Nair, R. Acharya, A.V.R. Reddy, A. Goswami, 'A Simple Method for Correcting the Neutron Self-shielding Effect of Matrix and Improving the Analytical Response in PGNAA', *Anal. Chim. Acta*, **549**, 205 (2005).

45. R.M.W. Overwater, P. Bode, J.J.M. de Goeij, 'Gamma-ray Spectroscopy of Voluminous Sources: Corrections for Source Geometry and Self-attenuation', *Nucl. Instrum. Methods*, **A324**, 209 (1993).

46. R.M.W. Overwater, P. Bode, 'Computer Simulations of the Effects of Inhomogeneities on the Accuracy of Large Sample INAA', *Int. J. Appl. Radiat. Isot.*, **49**, 967 (1998).

47. H.W. Baas, 'Neutron Activation Analysis of Inhomogeneous Large Samples – An Explorative Study', Ph.D. Dissertation, Delft University of Technology, 2004.

48. P. Bode, E.A. De Nadai Fernandes, R.R. Greenberg, 'Metrology for Chemical Measurements: Purism, Pragmatism and the Position of INAA', *J. Radioanal. Nucl. Chem.*, **245**, 109 (2000).

49. R.M. Lindstrom, R. Zeisler, R.R. Greenberg, 'Accuracy and Uncertainty in Radioactivity Measurement for INAA', *J. Radioanal. Nucl. Chem.*, **271**, 311 (2007).

50. P. Bode, 'Quality Control in Large Sample Analysis', *J. Radioanal. Nucl. Chem.*, **271**, 333 (2007).

51. P. Bode, 'Detectors and Detection Limits in INAA. I. General Theoretical Relationships Between Detector Specifications and Detection Limits', *J. Radioanal. Nucl. Chem.*, **222**, 117 (1997).

52. P. Bode, 'Detectors and Detection Limits in INAA. II. Calculated and Observed Improvements in Detection Limits with Large Ge-detectors, Well-type Ge-detectors and Anti-Compton Spectrometers', *J. Radioanal. Nucl. Chem.*, **222**, 127 (1997).

53. M. Blaauw, H.W. Baas, M. Donze, 'Height-resolved Large-sample INAA of a 1 m long, 13 cm diameter Ditch-bottom Sample', *Nucl. Instrum. Methods*, **A505**, 512 (2003).

54. R.M.W. Overwater, P. Bode, J.J.M. de Goeij, J.E. Hoogenboom, 'Feasibility of Elemental Analysis of Kilogram-Size Samples by Instrumental Neutron Activation Analysis', *Anal. Chem.*, **68**, 341 (1996).

55. C. Segebade, P. Bode, W. Gorner, 'The Problem of Large Samples: An Activation Analysis Study of Electronic Waste Material', *J. Radioanal. Nucl. Chem.*, **271**, 261 (2007).

56. R. Acharya, A.G.C. Nair, A.V.R. Reddy, A. Goswami, 'Standard-less Analysis of Zircaloy Clad Samples by an Instrumental Neutron Activation Method', *J. Nucl. Mater.*, **326**, 80 (2004).

57. R. Acharya, A.G.C. Nair, K. Sudarshan, A.V.R. Reddy, A. Goswami, 'Development and Applications of the k_0-based Internal Mono Standard INAA Method', *Appl. Radiat. Isot.*, **65**, 164 (2007).

58. K. Janssens, R. van Grieken (eds.), *Comprehensive Analytical Chemistry XLII: Non-destructive microanalysis of cultural heritage materials*, Elsevier, 2005.

59. IAEA STI/DOC/010/416, 'Nuclear Analytical Techniques in Archaeological Investigations', Vienna, 2003.

Activity Determination and Localization of Radioactive Point Source in a Large Medium

Zeev B. Alfassi

Ben Gurion University, Beer Sheva, Israel Università degli Studi and INFN of Milano, Segrate

1 Introduction **1**
2 γ-Emitting Source **1**
2.1 Two Collinear Detectors 1
2.2 Four Detectors 5
3 Neutron Sources **10**
3.1 Two Neutron Detectors 11
3.2 More than Two Detectors 11
References **14**

The activity and localization of a small-sized radioactive source in a large medium was studied for various cases by the simultaneous use of several detectors. The γ-emitting source and the neutron-emitting source were both studied. This also led to an α-emitting source, which cannot be detected from outside of the medium unless the neutrons from their (α, n) reaction with low Z elements are detected. A special actual case of commercial lung counter with four semiplanar HPGe detectors is described in detail. For large boxes, NaI(Tl) detectors were used for γ measurements measuring only the full-energy peaks. For neutrons, only thermal neutrons were measured with ^{3}He detectors. To increase the fraction of the thermal neutrons and to protect the measurer, the sample was surrounded by a fence made of 5-cm polypropylene.

1 INTRODUCTION

Many measurements of radioactivity involve counting of γ rays or neutrons emitted by bulky samples, which are referred to as *a box*. While it is simple experimentally in the case of a point source in a known position, to transform the measurements of count-rate to activity (disintegration per unit time) and hence to the number of radioactive atoms, the problem can be the inaccuracy in the case of a small source in an unknown position in a bulky medium. The inaccuracy is because of two factors: (i) the self-absorption and scattering of the photons in the sample and (ii) the unknown solid angle in which the detector sees the point source. Thus, in the case of a radioactive small source in an unknown position in a large box, we need to find its position in order to transform the count-rate to activity (*see* **Instrumental Neutron Activation Analysis**).

The measurements are different if the source is emitting γ rays or neutrons. In the case of the γ-emitting source, measurements can be done only for the photons that do not interact with the media surrounding the source, except for the media of the detector, by measuring the full-energy peak. In the case of neutron emission, the measurement of the spectrum of the emission is more difficult. In addition, most of the emitted neutrons are fast neutrons, while most of the detectors are more sensitive to thermal neutrons; hence, for better statistics, the neutrons are measured after their interaction with the surrounding material.

The localization of the source can be done only by measurement with more than one detector. The simple case is when it is known that the source in on a given line. In this case, there are only two unknowns: the activity and the location on the line; and two detectors on parallel faces of the box, collinear with the source, are sufficient. However, in the general case, there are four unknowns: activity and the three coordinates in space; then, at least four detectors are needed.

2 γ-EMITTING SOURCE

2.1 Two Collinear Detectors

2.1.1 Nonabsorbing Media

Two factors cause the decrease in efficiency with the increase in distance from the detector: (i) the isotropic propagation of γ rays, which leads to a smaller solid angle in which the detector sees the source with the increasing distance, i.e. the 1/distance^2 law and (ii) the absorption of γ rays inside the sample, i.e. the $\exp(-\mu \cdot \text{distance})$ law. In order to separate these factors and dealing with only the first factor, a relative nonabsorbing media should be used. This can be done either with a low dense sample or, in the case of a low value of μ, the linear absorption coefficient, with high-energy γ rays.[1]

The mathematical treatment uses the assumption of the ''virtual point detector'' model for the voluminous γ detectors.[2–9] The bulky detector is treated as a virtual point detector on the symmetry axis of the detector at a distance d from the cap. This means that the count-rate in the absence of an absorbing/scattering medium due to a radioactive source at a distance y from the detector cap $C(y)$ is given by the equation $C(y) = \frac{C(0) \cdot d^2}{(y+d)^2}$, where d is

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

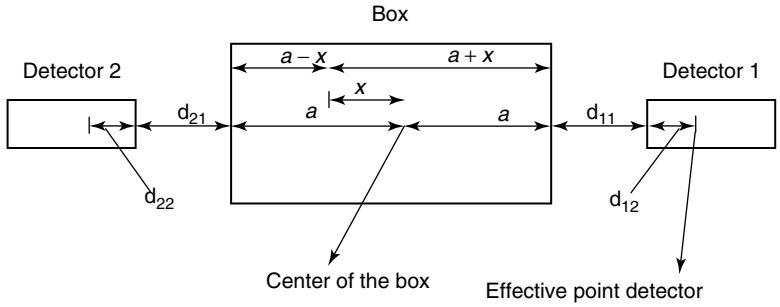

Figure 1 The dual detectors measuring system.

the distance from the detector cap to the virtual detector point, y is the distance from the source to the detector cap, and $C(0)$ is the count-rate measured on the detector cap.

The localization of the radioactive point source is done by two detectors on parallel faces of the box, collinear with the source, as shown in Figure 1. In Figure 1, d_{j1} is the distance between the inner face of the box and the detector j cap and d_{j2} is the distance between the detector cap and its effective virtual point detector; thus, if $d_j = d_{j1} + d_{j2}$, we get the equation for the count-rate because of a source at a distance y from the inner face of the box.

$$C_j(y) = \frac{C_j(0) \cdot {d_j}^2}{(y + d_j)^2} \quad j = 1, 2 \tag{1}$$

For a point source at the center of a box with a linear dimension $2a$ (on the line of centers), this equation leads to

$$C_j(\text{center}) = \frac{C_j(0) \cdot {d_j}^2}{(a + d_j)^2} \quad j = 1, 2 \tag{2}$$

For a point source located at a distance x from the center, the distances from the box faces are $a + x$ and $a - x$, and hence Equation (2) yields

$$C_1(y) = \frac{C_1(0) \cdot {d_1}^2}{(a + x + d_1)^2}; \quad C_2(y) = \frac{C_2(0) \cdot {d_2}^2}{(a - x + d_2)^2} \tag{3}$$

The normalized count-rate (normalized to the count-rate at the center of the box) $N_1(x)$ and $N_2(x)$ are

$$N_1(x) = \frac{(a + d_1)^2}{(a + x + d_1)^2}; \quad N_2(x) = \frac{(a + d_2)^2}{(a - x + d_2)^2} \tag{4}$$

Since both detectors are quite similar, we can accurately approximate $d_1 = d_2$. Equation (4) shows that, if we take the square roots of the reciprocal count-rate, the denominators are the same, while in the numerators, x appears once with a plus sign and once with a minus sign, which means that adding them cancels the terms with x:

$$\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} = 2 \tag{5}$$

This sum has the units $s^{0.5} \cdot \text{count}^{-0.5}$ and hence the square of its reciprocal value has the units counts per second, which are the units of count-rate. The definition of a new mean according to this equation, the harmonic mean of the square roots, leads to a mean that is independent of the position of the source.

$$M_{\text{SH}} = \left(\frac{2}{\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}}} \right)^2 \Rightarrow M_{\text{SH}} = \frac{4}{\left(\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} \right)^2}$$

$$M_{\text{SH}} = \frac{4N_1N_2}{(\sqrt{N_1} + \sqrt{N_2})^2} \tag{6}$$

Equation (5) indicates that M_{SH}, defined for the normalized counts, should be equal to 1, independent of x. It was found experimentally[1] that the deviation of M_{SH} from 1 is always below 10% and, in most cases, not more than 2–3%.

This constancy of M_{SH} indicates the method of measuring the activity of a point source accurately within a box without having to know the position of the source. As we do not know the counts with the source at the center of the box, M_{SH} is not calculated with the normalized count-rates, but with the experimental ones. In this way, M_{SH} is independent of the position of the source, but dependent proportionally on the activity of the source. The box is calibrated for its efficiency by a standard point source with known activity, which is denoted by Act(Std), positioned anywhere on the line of the centers. Let us assign its mean square harmonic M_{SH} by M_{SH}(Std). The activity of any unknown source Act(u) of the same energy, positioned

somewhere (unknown) on the line of centers, is given by

$$\mathrm{Act}(u) = \frac{M_{\mathrm{SH}}(u)}{M_{\mathrm{SH}}(\mathrm{Std})} \cdot \mathrm{Act}(\mathrm{Std}) \tag{7}$$

M_{SH} is calculated from the count-rates of the two detectors (Equation 6). As the unknown source cannot be removed for the measurement of the standard source, the counts due to the standard source for each detector are calculated by the subtraction of two counts; one with the unknown source alone and the other with both of the two, the unknown source and the standard one.

If the exact place of the source in the box is required, the position of the source can be found from the ratio of the count-rates of the two detectors. From Equation (5):

$$R = \sqrt{\frac{N_1}{N_2}} = \frac{[(d+a)+x]}{[(d+a)-x]} \tag{8}$$

and hence

$$x = (d+a) \cdot \frac{1-R}{1+R} \tag{9}$$

where x is the distance from the center of the box toward detector 2. Care should be taken to include d, besides the distance of the virtual point detector from the detector cap (which depends on the γ-ray energy) and also the thickness of the box and the distance of the outer face of the box from the detector cap.

2.1.1.1 Calculated Error for Longitudinal Sources The calculation till now assumes that the source is a very small one, close to a point. In the following, we calculate the error induced by a longitudinal source of length $2b$ lying on the line of centers with its midpoint coinciding with the midpoint of the line of centers (Figure 2).

The activity of an infinitesimal piece of the source with length dx is $A \cdot dx/2b$, where A is the total activity of the source. The normalized count-rates of the longitudinal source–detector (normalized to the whole activity being concentrated in the center) is given by the following integral

$$N = \frac{\dfrac{A}{2b}\displaystyle\int_0^{2b} \mathrm{d}x/(d+a-b+x)^2}{A/d+a}$$

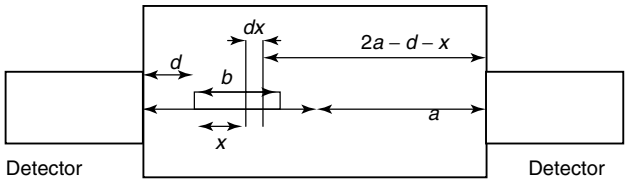

Figure 2 Longitudinal measuring geometry.

By integration, the normalized count-rates are

$$N = \frac{(d+a)^2}{(d+a)^2 - b^2} = \frac{1}{1-[b/(d+a)]^2} \tag{10}$$

If $b/(d+a) \ll 1$, N can be approximated as $N = 2b/(d+a)$; this is the fractional error because the source is not a point source. As long as $b/(d+a) < 0.3$, the error due to a longitudinal source (Equation 10) is less than 10%.

2.1.2 *Absorbing Media – Parallel γ Lines*

2.1.2.1 Activity Calculation The use of collimated detectors to measure only parallel photons, as, for example, in a medical double-headed γ camera, eliminates the decrease in the solid angle in which the detector sees the source with the increase in the source–detector distance and leaves the effect of the source–detector distance to the sole effect of absorption/scattering.[10] In this case, the geometric mean of the count-rate of the two detectors can be used to calculate the activity of the unknown source:

$$C_1 = C_{10} \cdot \mathrm{e}^{-\mu(a-x)}; \quad C_2 = C_{20} \cdot \mathrm{e}^{-\mu(a+x)} \tag{11}$$

From Equation (11), the geometric mean of the two count-rates measured by the two detectors is given by

$$\begin{aligned} M_{\mathrm{G}}(x) &= \sqrt{C_1(x) \cdot C_2(x)} = \sqrt{C_{10} \cdot C_{20} \cdot \mathrm{e}^{-2\mu \cdot a}} \\ &= \mathrm{e}^{-\mu \cdot a}\sqrt{C_{10} \cdot C_{20}} \end{aligned} \tag{12}$$

Equation (12) shows that the geometric mean of the two count-rates is independent of the position of the point source but depends linearly on the activity of the source. Thus, the geometric mean can be used to measure the activity of an unknown source using a calibrated source positioned on one of the faces of the box between the detector and the box:

$$\mathrm{Act}(u) = \frac{M_{\mathrm{G}}(u)}{M_{\mathrm{G}}(\mathrm{Std})} \cdot \mathrm{Act}(\mathrm{Std}) \tag{13}$$

2.1.2.2 The Error due to a Longitudinal Source The mathematical approach described previously assumed that the active source is very thin. In the following, the error is induced if the source has a dimension in the plane of the detectors. Let us assume that the whole activity is spread over a length b (Figure 2). Then, the count-rates due to infinitesimal length dx, assuming there is no absorption by the medium, are $\frac{C_{01}}{b}\mathrm{d}x$ and $\frac{C_{02}}{b}\mathrm{d}x$. The

count-rates due to the absorption effects are (Figure 3):

$$C_1 = \frac{C_{01}}{b}\int_0^b \mathrm{e}^{-\mu(x+d)}\,\mathrm{d}x = \frac{C_{01}}{\mu b}\mathrm{e}^{-\mu d}\cdot(1-\mathrm{e}^{-\mu b})$$

$$C_2 = \frac{C_{02}}{b}\int_0^b \mathrm{e}^{-\mu[2a-(x+d)]}\,\mathrm{d}x = \frac{C_{02}}{\mu b}\cdot\mathrm{e}^{-\mu 2a}\cdot\mathrm{e}^{\mu d}(\mathrm{e}^{\mu b}-1)$$

where $2a$ is thickness of the box, d is the distance of one side of the source from one of the detectors, and x is the distance of the infinitesimal length $\mathrm{d}x$ from the same side of the source. The geometric mean is

$$M_{\mathrm{G}} = \sqrt{C_1\cdot C_2} = \mathrm{e}^{-\mu a}\sqrt{C_{01}\cdot C_{02}}\frac{\sqrt{(1-\mathrm{e}^{-\mu b})(\mathrm{e}^{\mu b}-1)}}{\mu b} \tag{14}$$

The correction factor K due to the dimension b perpendicular to the detector's plane is given by

$$K = \frac{\sqrt{(1-\mathrm{e}^{-\mu b})(\mathrm{e}^{\mu b}-1)}}{\mu b} \Rightarrow K = \frac{\mathrm{e}^{\mu b/2}(1-\mathrm{e}^{-\mu b})}{\mu b} \tag{15}$$

As long as $\mu b < 1.5$, the error is <10% ($K < 1.10$). For example, for $^{99\mathrm{m}}$Tc in water $\mu = 0.152\,\mathrm{cm}^{-1}$, then for an object with size <10 cm, the error is <10%.

2.1.2.3 Source Localization The previous mathematical model yields the activity of the unknown source in the box without knowing where the activity is located. However, the location of the source can also be calculated using Equation (11). Dividing the two, Equation (11) yields

$$R(x) = \frac{C_1(x)}{C_2(x)} = \left[\frac{C_{10}}{C_{20}}\right]\cdot\mathrm{e}^{\mu\cdot 2x} \tag{16}$$

$$x = \frac{\ln[R(x)/(C_{10}/C_{20})]}{2\mu} \tag{17}$$

If the two detectors are equal, or are made to give equal responses by adding a factor to one of them, then $C_{10}/C_{20} = 1$.

If the material in the box is known, Equation (15) can be used for calculating x from the known μ. However, in case the content of the box is unknown, μ can also be calculated from the measurement of the external calibrated standard on the face of the box by the two detectors, as the only decrease in the counts is due to the absorption.

$$\mu = \frac{\ln[R(a)]}{2a} \tag{18}$$

Here, $R(a)$ is the ratio of the two count-rates due to the external source alone.

2.1.3 The General Case

The general case is where both, the solid angle in which the detector sees the source and the absorption/scattering, are operational.

2.1.3.1 Activity Calculation In the previous sections, it was shown that, when only the solid angle is important, the activity can be calculated from the square root harmonic average, while when only absorption is operational the geometric average is the right method to calculate the activity. Since these two effects are independent, it can be expected that, in the general case, the right answer is a weighted product of these two averages. The weighing must ensure that the units of the new average are the same as count-rate[10]:

$$M_{\mathrm{GSH}} = {M_{\mathrm{SH}}}^{\alpha}\cdot {M_{\mathrm{G}}}^{1-\alpha} \tag{19}$$

It was found that α depends on the contents of the box[10] and α can be found from the measurement of the count-rates at several points of the box. However, for an unknown box, this means opening the box or reproducing a similar box for calculating α. The method to measure the radioactivity in the box without the need to open it is by previous localization of the radioactive point source, as discussed in the next paragraph.

2.1.3.2 Localization of the Point Source It was shown[11] that the count-rate of a detector positioned on the face of a box as a function of the distance of a point source, positioned inside the box filled with polypropylene, from the detector is close to exponential dependence, but is not very accurate. The exponential dependence on the distance from the box's face is improved if, instead of taking the count-rate of one detector, the ratio of the count-rates of two detectors positioned on parallel faces is taken. Another possibility to improve the exponential dependence is to position the detector at a distance from the face of the box. However, increasing the distance between the detector and the box leads to worse statistics because of a smaller count-rate. In addition, the use of the ratio of the count-rates of two detectors is preferable as the equation of the count-rate $C(x)$ versus the distance of the source from the box face $-x$,

$$C(x) = C(0)\cdot\mathrm{e}^{-\mu\cdot x} \tag{20}$$

has two parameters, $C(0)$ and μ, which require at least two calibration experiments. Performing two calibration measurements is difficult if we do not want to open the box. On the other hand, if the two detectors are equal, or made to be equal, by adding a factor to one of them,

the two parameters of the equation of the ratio of the count-rates

$$R(x) = R(0) \cdot e^{-\mu \cdot x} \tag{21}$$

can be shown to be correlated, leading to only one parameter.

If the length of the box is a, then $R(0)$ and $R(a)$ are reciprocal numbers, as changing from 0 to a means replacing the detectors. Substituting $R(a) = 1/R(0)$ yields

$$R(0) = e^{\mu \cdot a/2} \tag{22}$$

Hence

$$x = -\frac{a}{2} \cdot \left(1 - \frac{\ln[R(x)]}{\ln[R(0)]}\right) \tag{23}$$

The activity of the source cannot be calculated from the ratio of the count-rates as $R(x)$ is independent of the source activity. For the determination of the source activity, the measurement of the count-rate of one detector must be used, together with the calculated x from Equation (23). Since in the case of the detectors close to the box, the dependence of the count-rate on the distance x is quite complex, it is preferable to use measurements at a distance between the detectors and the box. As mentioned before, there are two unknowns, $C(0)$ and μ, but μ for one detector measurement can also be deduced from the measurements of the ratio of the count-rates at the same distance detector box. It was shown earlier that μ for the expression $R(x)$ can be found from $R(0)$ and the box length. The μ for one detector count-rate was proven to be one-half of the μ for the ratio of the count-rates.[11]

2.2 Four Detectors

Old lung counters were made of two large-sized phoswich detectors. In order to increase the resolution of modern lung counters, semi-planar Ge detectors were used. Owing to the small size of the available Ge detectors, 3 or 4 detectors were typically used to cover the area of the two lungs previously covered by two phoswich detectors.[12,13] A commercially available set of four detectors is the Lung-Counter composed of four semiplanar HPGe detectors that were positioned two on each lung on the chest.

2.2.1 Lung Detector

The basic assumption generally made when calibrating a lung counter is that the deposition of the radioactive aerosols in lungs is homogenous; however, it has been reported that the distribution is a function of the aerosol size and the breathing rate and changes with time.[14] In some cases, the contamination can be present in the form of a single ''hot'' particle. The count-rate of an external detector is strongly dependent on the position of the ''hot particle''; the possible errors in activity determination due to deviations from homogeneity were evaluated to be factors of 26, 3, and 1.8 for photon energies of 17, 60, and 1000 keV respectively,[15] and factors of up to 20 were measured for natural uranium.[16] The extra information due to more than one detector allows the localization of a point source in the lungs.

2.2.1.1 The Mathematical Method The mathematical method used is applicable not only for localization of radioactive sources but also for any technique using several measurements to identify an unknown.[17] The most well known analysis using this mathematical method is the identification of an analyte from its electron-impact mass spectrum (EIMS),[18] but the method can be also used for source identification.[19–21] It is based on a calibration with a radioactive point source positioned consecutively in various locations in a phantom, which is the actual bulky material. For every positioning, the four detectors measure the pulse height spectra, and the count-rate for each γ full-energy peak is then calculated. The position of an unknown radioactive point source can be found by comparing its count-rates measured by the four detectors with similar data obtained for each of the various locations during the calibration. The mathematical problem is to find which of the sets of the standards (which, in our case, contain the results of the four detectors) best fits the four count-rates of the unknown.

The values of the count-rates of the four detectors are set as the components of a vector and all the vectors are normalized to unity length.[22] Each normalized vector can be seen as a single point on a hypersphere with a unity radius in an n-dimensional hyperspace, where n is the number of the vector components (the number of the detectors). If two vectors (points on the n-dimensional hypersphere) are identical in all the components, there will be a perfect ''match'' that represents the same point in the hyperspace. However, because of instrumental instability and the statistical nature of the measurements, the point of an unknown coincides exactly with one of the points in the library of standards only rarely. Actually, even repeating the same measurement at the same point rarely gives exactly the same count-rates. The problem of finding the best match for the vector of an unknown in the library of the vectors of standards can be solved in principle by two methods (two matching factors). The first method can be described as the ''nearest-neighbor'' technique, while the second can be defined as the ''minimal contact angle'', although, in many cases, it is called the *dot product method* after the mathematical operation used to calculate the angle.

The first method calculates the distances between the head of the vector of the unknown and the heads of each of the vectors of the standards (the library), looking for the minimum distance, i.e. the minimum of

$$\Delta_j = \sum_{i=1}^{n} (u_i - s_{i,j})^2 \tag{24}$$

where u_i is the ith component (the detector i) of the vector of the count-rates of the unknown and $s_{i,j}$ represents the ith component of the jth vector in the library of standards. The summation is performed over all the components of the vector (all the detectors). It should be emphasized that all the vectors must be normalized (to be of unity length) in order to compensate for different activities of the standard and the unknown:

$$u_i = \frac{\mathbf{U}_i}{\sum_{k=1}^{n} \mathbf{U}_k}; \quad s_{i,j} = \frac{\mathbf{S}_{i,j}}{\sum_{k=1}^{n} \mathbf{S}_{k,j}} \tag{25}$$

where **U** and **S** are the original un-normalized vectors.

In the second method, the angles between the unknown vector and each of the standard vectors are calculated. The cosine of the angle between two vectors can be calculated from the values of their components by using the scalar ''dot product'':

$$\cos\theta_j = \frac{\underline{s}_j \cdot \underline{u}}{|\underline{s}_j| \cdot |\underline{u}|} = \frac{\sum_{i=1}^{n} s_{i,j} \cdot u_i}{\sqrt{\sum_{i=1}^{n} s_{i,j}^2 \cdot \sum_{i=1}^{n} u_i^2}} \tag{26}$$

The summation is done over all the components of the two vectors (which are of the same dimension). Actually, the calculation of the angle does not require the normalization of the vectors and can be performed on the original data:

$$\cos\theta_j = \frac{\underline{\mathbf{S}}_j \cdot \underline{\mathbf{U}}}{|\underline{\mathbf{S}}_j| \cdot |\underline{\mathbf{U}}|} = \frac{\sum_{i=1}^{n} \mathbf{S}_{i,j} \cdot \mathbf{U}_i}{\sqrt{\sum_{i=1}^{n} \mathbf{S}_{i,j}^2 \cdot \sum_{i=1}^{n} \mathbf{U}_i^2}} \tag{27}$$

However, it is preferable to use the normalized vectors at least for all the standards. Even if the vectors are not normalized explicitly, normalization is still performed in the process of calculating the lengths of the vectors in the denominator of Equation (27), which includes division by the lengths of the vectors. It is better to calculate the lengths only once in the normalization process rather than to calculate them for each unknown.

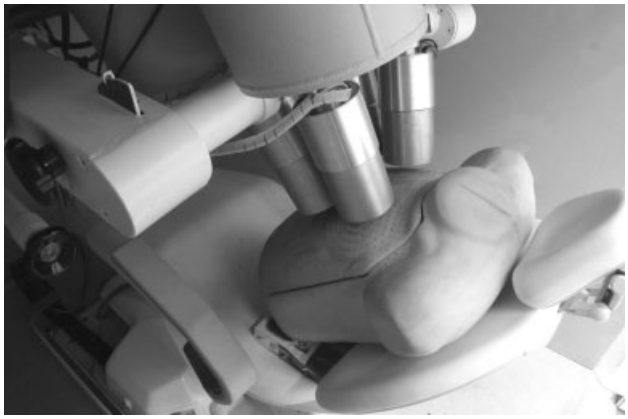

Figure 3 The realistic phantom used in the experiments.

Lengths of normalized vectors are unities, which simplifies Equation (26):

$$\cos\theta_j = \frac{\underline{s}_j \cdot \underline{u}}{|\underline{s}_j| \cdot |\underline{u}|} = \sum_{i=1}^{n} s_{i,j} \cdot u_i \tag{28}$$

The Experimental System The measurements were performed by the NRC-Negev lung counter system, which consists of four semiplanar HPGe detectors measuring a uranium source inside a phantom, as can be seen in Figure 3. The information from each detector was analyzed separately by a multichannel analyzer. The uranium source was a sealed cylinder with a radius of 1.5 mm and length of 5 mm, which contained natural uranium (99.27% ^{238}U and 0.72% ^{235}U) with an activity of 1.85 kBq. The main γ lines were 92 keV because of the ^{238}U chain (due to ^{234}Th) and 186 keV mainly because of ^{235}U. The quantitative information from each measurement was the count-rates of the 186 and 92 keV peaks (the areas under the peaks divided by the counting time). Other γ peaks were also observed; however, their counts were smaller, implying larger statistical errors, and they were not used in the analysis.

The detectors were calibrated using a realistic phantom designed by Lawrence Livermore National Laboratory. The phantom lungs used were made of tissue-equivalent plastic material and had 28 cylindrical holes, where point sources of natural uranium could be placed, as can be seen in Figure 4. All holes were normally filled with cylinders made of the same tissue-equivalent plastic as the lungs, and a small portion of the cylindrical filling was replaced by the source only at the checked position, so that the absorbing properties of the lungs remained practically unchanged and fully simulated the human lungs. The 56 points for placing the sources at different locations (two points in each hole) were distributed all over the lungs. Twenty-eight points were located on the upper surface of the lungs and 28 on the lower surface.

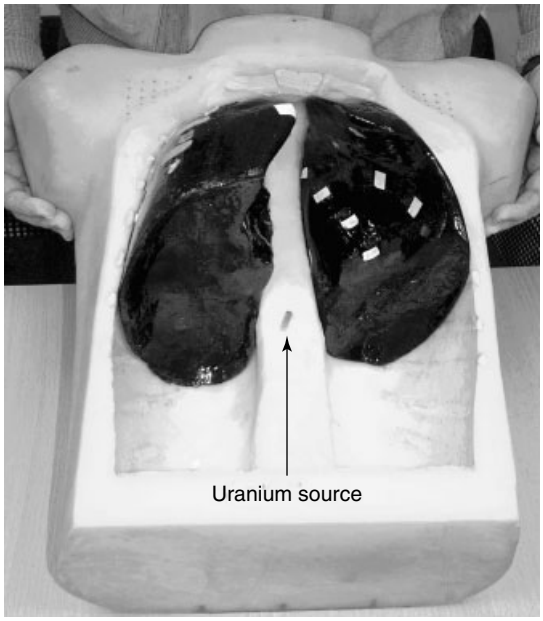

Figure 4 The four lung counters placed over the phantom.

The matching factors (Δ_j and $\cos\theta_j$) of the counts for the source placed at an unknown position with each of the 56 different points of the library data base were calculated. On the basis of the minimum value of Δ_j, the matching factor, or a maximum of $\cos\theta_j$, the other matching factor, a ''predicted point'' was determined and compared with the actual point location. If there was a match between them, it was called a *success* or just a *hit*.

The difference between the calculated and the actual activity (the error in calculation of the activity) was calculated as follows. After determination of the ''guessed'' point location, the guessed (calculated) source activity was calculated from the count-rates of the source at the unknown position, and the efficiency factor for the guessed point was determined. The difference between this calculated activity for each single measurement and the known source activity was the error for the specific measurement.

There was a one-to-one correlation between $\cos\theta$ (θ stands for the angle between the two vectors) and the distance between the points (heads of the vectors) in the hyperspace, and it was found, in the case of localization of a radioactive point source in the lung phantom, that Equations (3) and (5) actually gave the same results.[19]

The Use of Weighted Results Stein[23] found that the identification in EIMS gets better if, instead of raw ion intensities, the values (ion intensity · ion mass)$^{1/2}$ are taken as the components of the vectors. This choice of components of the vectors increases the weights of low-intensity peaks produced by ions of big masses.

In our case of the lung phantom, it was also found that the number of ''hits'' increased slightly if raw values of the count-rates were multiplied by weighing factors. The various weighing factors studied are given in Table 1. For calculation of the least distance, it was found that the best results were obtained for weights taken as the reciprocal of the mean of the normalized count-rates of the unknown and the library vectors. The raw vectors are first normalized, and the weights are calculated from the normalized vectors. In a mathematical notation, this means that, instead of searching for the library vector $\mathbf{L}$ that leads to the minimal values of $\sqrt{\sum_{i=1}^{n}(\mathbf{L}_i - \mathbf{U}_i)^2}$ for the various vectors of the library (where $\mathbf{U}$ is the vector of the unknown), we are looking for the vector $\mathbf{L}$ that yields a minimal value of $\sum_{i=1}^{n}(\mathbf{U}_i - \mathbf{L}_i)^2/(\mathbf{U}_i + \mathbf{L}_i)$. Because we are looking for the minimal distance, this assigns bigger weights to the measurements of larger counts with smaller statistical errors. When these weights are used, the fraction of the ''hits'' (where ''hit'' means that the method detects the correct location of the source) increases from 78 to 82%.[19] Although there were only 82% of ''hits'', the mean error in the calculated activity was only 10.4% because, in the ''nonhit'' cases, the resulted position was

Table 1 The various test criteria

1	2	3	4	5
$\cos^2\theta = \dfrac{\left(\sum \mathbf{L}_i \cdot \mathbf{U}_i\right)^2}{\sum \mathbf{L}_i^2 \cdot \sum \mathbf{U}_i^2}$	$\lvert r\rvert = \dfrac{\sum(\mathbf{U}_i - \overline{\mathbf{U}})\,(\mathbf{L}_i - \overline{\mathbf{L}})}{\sqrt{\sum(\mathbf{U}_i - \overline{\mathbf{U}})^2 \sum(\mathbf{L}_i - \overline{\mathbf{L}})^2}}$	$\sum_{i=1}^{n} \mathbf{L}_i \cdot \lvert\mathbf{U}_i - \mathbf{L}_i\rvert$	$\sum_{i=1}^{n} \mathbf{L}_i \cdot (\mathbf{U}_i - \mathbf{L}_i)^2$	$\sum_{i=1}^{n} \left\lvert \dfrac{\mathbf{U}_i}{\mathbf{L}_i} - 1 \right\rvert$
6	7	8	9	10
$\sum_{i=1}^{n} \left(\dfrac{\mathbf{U}_i}{\mathbf{L}_i} - 1\right)^2$	$\sum_{i=1}^{n} \dfrac{\lvert\mathbf{U}_i - \mathbf{L}_i\rvert}{\mathbf{U}_i}$	$\sum_{i=1}^{n} \dfrac{\lvert\mathbf{U}_i - \mathbf{L}_i\rvert}{\sqrt{\mathbf{U}_i \cdot \mathbf{L}_i}}$	$\sqrt{\sum_{i=1}^{n}(\mathbf{L}_i - \mathbf{U}_i)^2}$	$\sum_{i=1}^{n} \lvert\mathbf{L}_i - \mathbf{U}_i\rvert$
11	12	13	14	15
$\sum_{i=1}^{n} \dfrac{(\mathbf{U}_i - \mathbf{L}_i)^2}{(\mathbf{U}_i + \mathbf{L}_i)}$	$\sum_{i=1}^{n} \left[\dfrac{\mathbf{U}_i - \mathbf{L}_i}{\mathbf{U}_i + \mathbf{L}_i}\right]^2$	$\sum_{i=1}^{n} \dfrac{(\mathbf{U}_i - \mathbf{L}_i)^2}{\mathbf{U}_i \cdot \mathbf{L}_i}$	$\sum_{i=1}^{n} \dfrac{(\mathbf{U}_i - \mathbf{L}_i)^2}{\sqrt{\mathbf{U}_i \cdot \mathbf{L}_i}}$	$\sum_{i=1}^{n} \dfrac{\lvert\mathbf{U}_i - \mathbf{L}_i\rvert}{\mathbf{U}_i + \mathbf{L}_i}$

Table 2 The results obtained when either the 186 keV or the 92 keV peaks were used separately and when both peaks were used simultaneously, using the library of 224 measurements

Criterion number	92 keV		186 keV		92 + 186 keV	
	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)
1	78	17.6	81	10.9	90	5.5
2	67	23.8	71	16.8	73	9.5
3	64	19.2	72	20.4	74	13.0
4	70	21.0	77	16.1	79	10.7
5	79	14.6	81	11.4	88	6.0
6	81	14.0	80	11.0	89	5.5
7	82	11.3	81	11.4	89	5.5
8	82	11.3	80	12.3	90	5.4
9	78	17.6	81	10.9	90	5.5
10	80	14.5	82	10.4	88	6.3
11	82	11.2	84	9.8	94	4.1
12	82	13.9	81	10.9	90	5.7
13	82	13.9	81	10.9	90	5.7
14	82	11.3	82	9.5	94	3.8
15	83	11.3	80	12.3	90	5.4

quite close to the real one, which led to a relatively small error in the calculated activity.

Homogenous versus Point-wise Distribution Application of this method to a homogenously distributed source in the lungs of the phantom results in the best "fit" for one of the points in the center. The error in the activity is quite large in this case, almost 40%. It seems that this method can work only when applied to inhomogenously distributed radioactive contaminants. However, this problem can be solved by adding another vector to the set of the standards.[24] This vector is the response of the four detectors to a phantom with homogenously distributed radionuclides, which can be defined as a "virtual point". Another possibility for a vector representing the homogenously distributed lungs is to take the arithmetic mean of the 56 calibrating points for each one of the four detectors. With this extended set of standards, the "guessed point" for homogenously distributed lungs is always found in this virtual vector, and the error in the calculated activity is below 5%. Thus, one can actually distinguish between a homogenous distribution and a point location.

Simultaneous use of two γ lines In the case of uranium contamination, the count-rates of two relatively strong γ lines were measured by the four detectors, namely, the lines at 92 and 186 keV. The 92-keV line is due to a daughter in the ^{238}U chain, while the 186 keV line originates mainly from ^{235}U. These γ lines, when used separately, provide approximately the same fraction of "hits" and the same mean error in the activity, but a closer observation shows that the cases of "nonhits" are not the same for both the lines. This is reasonable because the main cause for a "miss" is the statistical nature of the count-rates. It was found[20] that analyzing the two energy peaks together could provide better results. Eight-component vectors were composed. Out of these, four components were the 92 keV count-rates measured by the four detectors, while the other four components were count-rates measured by the same detectors for the 186 keV line. The normalization of the vectors involved all the eight components. It can be seen in Table 2 that the fraction of "hits" increased from 82%, when each of the energies was analyzed separately, to 94%, when the two energies were analyzed together using an eight-dimensional vector. The mean error in the activity decreased from 10.4 to 3.8%.

Table 2 gives the percentage of "hits" and the error in the calculated activity for four detectors, both for using only the count-rates at either the 186 or the 92 keV peak (vectors of four dimensions) and the simultaneous count-rates at the 92 and the 186 keV peaks (eight-dimensional vectors). In the case of Table 2, the library was built from all the 224 experiments, and each standard vector was the average of four repeated experiments. The unknown vector was one of the 224 vectors, leading to the same vectors being used as part of the library, although in an average with three other experiments and as the unknown. In Table 3, the same results are given when the average of only two repeating experiments (112 measurements) was used to calculate the library, while

Table 3 Average results obtained when either the 186 or the 92 keV peaks were used separately and when both peaks where used simultaneously, using all six library combinations of 112 measurements

Criterion number	92 keV		186 keV		92 + 186 keV	
	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)
1	64 ± 2	28.1	69 ± 2	17.8	74 ± 2	12.4
2	54 ± 4	51.6	58 ± 4	34.5	68 ± 3	16.3
3	51 ± 4	43.4	59 ± 5	30.6	60 ± 4	19.8
4	57 ± 4	36.4	63 ± 3	26.8	64 ± 3	18.2
5	64 ± 3	26.9	69 ± 2	16.4	76 ± 2	12.5
6	64 ± 2	28.9	68 ± 2	17.7	74 ± 2	13.3
7	65 ± 5	25.7	69 ± 2	16.6	75 ± 5	12.6
8	65 ± 4	27.2	69 ± 2	16.3	77 ± 2	11.4
9	64 ± 2	28.1	69 ± 2	17.8	74 ± 2	12.4
10	63 ± 1	28.4	68 ± 2	21.5	76 ± 3	11.3
11	68 ± 2	23.4	71 ± 2	16.4	79 ± 3	10.5
12	66 ± 3	26.6	69 ± 2	16.9	76 ± 2	12.0
13	66 ± 2	26.6	68 ± 2	17.3	75 ± 2	12.2
14	68 ± 2	23.2	71 ± 2	16.0	79 ± 3	10.7
15	65 ± 4	27.3	69 ± 2	16.3	77 ± 2	11.6

the other 112 measurements were treated as unknowns. There are six possibilities (C_4^2) to choose the two repeating experiments used for the library; the results in Table 3 are the average of these six possibilities.

However, using the two γ lines presents a problem when the isotopic composition of the uranium in the calibration standards is different from that of the unknown, because the two lines are produced by different isotopes. It was shown[21] that, if the isotopic compositions were different, errors in localizing a source and determining its activity were larger with the eight-dimensional vector than with the four-dimensional one. This study[21] shows that, in such a case, the method of simultaneous analysis of two γ energy peaks leading to eight-dimensional vectors can still be used to improve the accuracy, but the eight-dimensional vectors should be normalized in an unusual way, which can be called *double half-normalization*. In this method, the four count-rates of each γ energy are normalized separately. The components of the normalized vector c_j are

$$c_j = \frac{C_{92,j}}{\sum_{k=1}^{4} C_{92,k}}; \quad \text{for } j = 1\text{–}4$$

$$c_j = \frac{C_{186,j-4}}{\sum_{k=1}^{4} C_{185,k}}; \quad \text{for } j = 5\text{–}8$$

In the expressions above, the measured count-rates are $C_{92,j}$ and $C_{186,j}$, where $j = 1$–4 represent the four different detectors. Pelled[21] found that this complex normalization (the lengths of the vectors are actually 2) decreases the fraction of "hits" in the case of the same isotopic composition only from 94 to 93%, which is still much higher than the success rate when a single γ line in used. When the isotopic compositions are different, this larger dimensionality increases the fraction of "hits" almost to the same degree as in the case of the same isotopic composition. However, when the isotopic compositions are extremely different, the statistics of one of the energy lines is expected to be poor; therefore, it is better to use only the most abundant γ line.

Detectors at the Back Another way to increase the dimensionality of the vectors is to use more detectors. In practice, the four detectors cover most of the area of the lungs, and, physically, there is no place for additional ones. However, they can be mounted on the back of the phantom. Counts by detectors mounted on the back are usually lower due to stronger absorption of the photons. Hence, longer counting times are required to get good statistics. Count-rates for the normal counting time with the detectors positioned at the back of the phantom were studied,[24] thus simulating an array of eight detectors. The results can be represented by eight-dimensional vectors or 16-dimensional vectors if both the γ energies (92 and 186 keV) are used. The percentage of "hits" exceeded 99%, suggesting that placing lung detectors behind the phantom, by mounting them in a chair support, is worthwhile. There is actually no need for eight detectors as only two additional detectors at the back increase the fraction of "hits" to about 99%.

Table 4 The percentage of "hits" and the error in the activity calculation with four detectors for 56 points (224 measurements)

Combination number	Number of detectors	Detectors	92 + 186 keV		186 keV	
			Number of "hits" (%)	Error in activity (%)	Number of "hits" (%)	Error in activity (%)
1	**4**	**A, B, C, D**	**94**	**3.8**	**83**	**9.4**
2	4	BA, BB, BC, BD	93	4.7	82	10.1
3	4	A, B, BA, BB	98	1.9	95	2.4
4	4	A, D, BA, BD	98	2.1	92	5.9
5	4	B, C, BB, BC	99	1.8	95	2.9
6	4	A, B, BC, BD	98	1.3	97	1.4
7	4	C, D, BA, BB	96	2.0	95	2.2
8	4	C, D, BC, BD	96	1.5	96	1.6
9	4	A, B, D, BA	99	1.6	96	2.2
10	4	A, B, D, BB	99	1.6	97	1.8
11	4	A, B, D, BC	99	1.4	98	1.6
12	4	A, B, D, BD	99	1.7	95	2.1
13	4	A, B, C, BA	98	1.8	92	3.2
14	4	A, B, C, BB	97	2.1	96	2.1
15	4	A, B, C, BC	99	1.6	97	1.7
16	4	A, B, C, BD	99	1.5	95	1.8
17	4	A, C, D, BA	98	2.0	93	3.4
18	4	A, C, D, BB	99	1.7	96	2.3
19	4	A, C, D, BC	99	1.5	97	1.8
20	4	A, C, D, BD	98	1.9	94	2.5
21	4	B, C, D, BA	98	1.7	94	2.4
22	4	B, C, D, BB	98	1.7	95	2.1
23	4	B, C, D, BC	100	1.5	97	1.9
24	4	B, C, D, BD	98	1.7	93	2.3

Possibly more interesting results of this study are that, even when only four detectors are used, it is considerably better to locate one or two of them at the back.[24] Tables 4 and 5 yield the percentage of "hits" and the error of the calculated activity both for using the 186 keV peak alone and for the simultaneous use of the 92 and 186 keV peaks. Each line gives the results for another arrangement of the four detectors. The detectors in the front are assigned by A, B, C, and D, while the detectors at the back are denoted by BA, BB, BC, and BD. Detector A is located opposite the upper part of the right lung. Detector B is located opposite the upper part of the left lung. Detector C is located opposite the lower part of the left lung, and detector D is located opposite the lower part of the right lung. When measuring very low energy γ photons, as, for example, when measuring contamination with ^{239}Pu, the detectors at the back hardly see anything. However, Tables 4 and 5, in which each line giving a different arrangement of the four detectors shows that, even for natural uranium, with not very high energy photons, positioning of one or two, out of the four, detectors at the back leads to considerably more accurate results. This is clearer in the case of only one photon energy or if the library used were counted for a relatively short time (6 h in Table 5 compared to 12 h in Table 4).

3 NEUTRON SOURCES

The main difference between neutron sources and γ sources is that, while γ sources emit photons of discrete energies, mainly one or two major lines, the radionuclidic neutron source, both fission sources or (α, n) sources, emits a spectrum of neutrons, mainly fast ones with an average of more than 1 MeV. Another difference is that it is quite easy to measure the spectrum of the γ photons emitted from the box, and the resolution of the energy spectrum is quite high, while, on the other hand, measurement of the spectrum of the neutrons is quite a task, and the resolution of the energy spectrum is quite low. Taking these two differences and the much larger sensitivity of usual neutron detectors to thermal neutrons than to more energetic neutrons, it is almost impossible to measure only neutrons that reach the detectors without interaction with the box and the surrounding materials, as was done for γ photons. Not only the interaction with the box but also the interaction with the shield of the experimental system should be considered, as the detectors must be shielded from outside radiation and the researcher must be protected from the neutron irradiation. This shield, if made from hydrogen-containing material, not only shields the detector from outside

Table 5 The average percentage of "hits" and the average error in the activity calculation with four detectors for 56 points (112 measurements, average of six combinations)

Table combination number	Number of detectors	Detectors	92 + 186 keV		186 keV	
			Number of "hits" (%)	Error in activity (%)	Number of "hits" (%)	Error in activity (%)
1	**4**	**A, B, C, D**	**79.8 ± 3.0**	**10**.2	**71.8 ± 1.9**	**16.0**
2	4	BA, BB, BC, BD	77.0 ± 4.8	13.1	66.7 ± 3.4	18.3
3	4	A, B, BA, BB	91.8 ± 1.2	3.6	86.5 ± 3.4	4.6
4	4	A, D BA, BD	92.7 ± 1.4	4.4	83.5 ± 2.3	14.8
5	4	B, C, BB, BC	95.2 ± 1.0	3.0	88.0 ± 2.8	6.4
6	4	A, B, BC, BD	95.0 ± 1.8	2.3	93.7 ± 2.9	2.4
7	4	C, D, BA, BB	91.3 ± 1.5	3.7	90.7 ± 2.9	3.8
8	4	C, D, BC, BD	93.7 ± 1.4	2.4	89.8 ± 3.5	3.4
9	4	A, B, D, BA	93.7 ± 0.5	3.0	86.0 ± 2.8	5.7
10	4	A, B, D, BB	96.3 ± 1.2	2.6	92.7 ± 2.5	3.4
11	4	A, B, D. BC	97.8 ± 1.8	2.1	95.7 ± 1.8	2.6
12	4	A, B, D, BD	94.7 ± 1.4	2.5	89.7 ± 3.7	3.5
13	4	A, B, C, BA	92.7 ± 1.2	3.2	85.3 ± 3.4	5.3
14	4	A, B, C, BB	92.3 ± 1.0	3.4	87.8 ± 3.5	4.8
15	4	A, B, C, BC	92.7 ± 3.3	3.3	92.2 ± 2.4	3.3
16	4	A, B, C, BD	95.5 ± 1.9	2.3	93.0 ± 2.8	2.6
17	4	A, C, D, BA	93.0 ± 0.9	3.8	86.8 ± 2.5	5.5
18	4	A, C, D, BB	95.2 ± 1.6	3.3	89.7 ± 1.2	4.6
19	4	A, C, D, BC	96.3 ± 1.4	2.6	93.3 ± 1.9	3.0
20	4	A, C, D, BD	92.5 ± 1.4	2.9	88.7 ± 2.9	3.7
21	4	B, C, D, BA	92.7 ± 1.4	3.2	86.0 ± 1.9	4.3
22	4	B, C, D, BB	91.5 ± 1.2	3.3	88.5 ± 2.5	4.0
23	4	B, C, D, BC	95.8 ± 1.8	2.5	92.5 ± 1.4	3.2
24	4	B, C, D, BD	91.3 ± 2.1	2.8	86.7 ± 2.0	3.6

radiation but also increases the count-rate at the detector due to reflection of thermal neutrons from the shielding material.

3.1 Two Neutron Detectors

It was found[25] that, when two ^{3}He detectors were positioned on two parallel faces of a box with a small neutron source in the box on the line connecting the two detectors, the ratio of the two count-rates was an exponential function of the distance between the source and one of these faces. The exponential dependence was found to be independent of the medium in the box, from air to highly hydrogenated material, although the coefficient of the exponentiality depended on the medium. This is the same as that described earlier for two γ detectors in the general case and hence the same equation should apply:

$$x = -\frac{a}{2} \cdot \left(1 - \frac{\ln[R(x)]}{\ln[R(0)]}\right) \tag{29}$$

where x is the distance from the box's face, a is the length of the box (the distance between the two faces on which the detectors are positioned), and R is the measured ratio of the two count-rates. $R(x)$ is the ratio for the unknown source and $R(0)$ is the ratio for an external source positioned on the face of the box ($x = 0$). It was found that Equation (29) yields quite accurate results, as can be seen in Table 6. In this table, $\Delta x/a$ is the fractional error in the location. As can be seen in this table, the calculated location is quite accurate except for the cases where the source is positioned very close to one of the two faces of the box. This is probably due to thermal neutrons coming to the close detector only by scattering from the shielding, as the neutrons coming directly from the source are too energetic to react with the ^{3}He in the detector. In order to get more accurate results, 5-cm polypropylene slabs were positioned between the detectors and the box.[26,27]

3.2 More than Two Detectors

When the unknown neutron source is not on the line (actually a plane as the detectors are quite long) connecting the two detectors, it is found that Equation (29) gives the correct result only if $R(0)$ is measured for the external source positioned on the face of the box in a plane parallel to the detectors' plane but passes through the unknown source. Therefore, now the problem is to find this plane. Two solutions are found to this problem. In the first solution, three pairs of ^{3}He

Table 6 The measured (x_{meas}) and calculated (x_{cald}) source-to-detector distance (cm) for an AmBe source

Wood			Paraffin			Air		
x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$
0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000
3.3	2.8	0.012	4.0	2.4	0.040	4.0	4.8	0.020
7.0	5.9	0.026	8.0	6.0	0.050	8.0	8.8	0.020
10.3	9.2	0.027	12.0	10.5	0.037	12.0	12.4	0.009
13.8	13.0	0.019	16.0	14.3	0.044	16.0	16.2	0.004
17.5	17.0	0.013	20.0	19.1	0.023	20.0	19.7	0.007
20.7	20.6	0.002	24.0	24.0	0.000	24.0	23.3	0.018
24.0	21.7	0.058	28.0	28.7	0.018	28.0	26.8	0.030
27.8	28.2	0.011	32.0	33.4	0.034	32.0	30.2	0.044
31.2	32.0	0.019	36.0	37.0	0.025	36.0	33.8	0.054
35.0	35.4	0.010	40.0	39.3	0.018	39.0	38.0	0.024
40.6	39.8	0.020						
Foamplast			**Paper**			**Concrete tiles**		
x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$
0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000
3.0	4.1	0.028	2.4	1.5	0.022	4.8	4.8	0.001
5.0	6.7	0.041	5.2	3.3	0.047	10.0	9.8	0.006
9.0	10.3	0.031	7.8	6.0	0.046	12.1	12.2	0.001
13.0	14.0	0.025	12.0	10.6	0.036	16.8	17.4	0.015
17.0	17.6	0.015	16.0	15.5	0.012	21.7	22.5	0.019
21.0	20.8	0.004	21.5	20.9	0.016	26.6	27.4	0.021
25.0	24.2	0.020	25.0	25.0	0.000	29.0	29.8	0.019
27.8	27.0	0.020	29.0	29.3	0.007	33.6	34.3	0.017
30.6	29.9	0.018	31.5	32.4	0.022	38.5	38.3	0.005
33.4	33.1	0.007	34.0	36.4	0.061			
36.2	34.6	0.041	36.5	38.0	0.039			
39.0	38.3	0.016	39.0	39.9	0.023			

detectors are positioned on two parallel faces of the box.[27,28] It is found that the pair that gives the highest geometric mean of the count-rates is the closest to the unknown source and hence the external source should be positioned in this plane. However, this still gives an error of one-sixth of the width of the box and, consequently, for larger boxes it will require more detectors. A better solution that requires only four detectors involves the moving of the external source on different faces by an iterative method.[29]

The iterative method was applied to localize the unknown neutron source by finding successive approximations to the solution, starting from an initial estimate regarding the unknown source plane (y coordinate). First, the x coordinate of the unknown source was calculated, using Equation (29), from the results of the two detectors on the y axis (detectors 1 and 2, which are on parallel faces of the box, as can be seen in Figure 5). Both $R(x)$ and $R(0)$ were measured by detectors 1 and 2. Measurements were made of the unknown source alone and together with a known (calibrating) source placed outside the box; $R(0)$ was the difference between the two measurements. In this first step, $R(0)$ was measured with the known

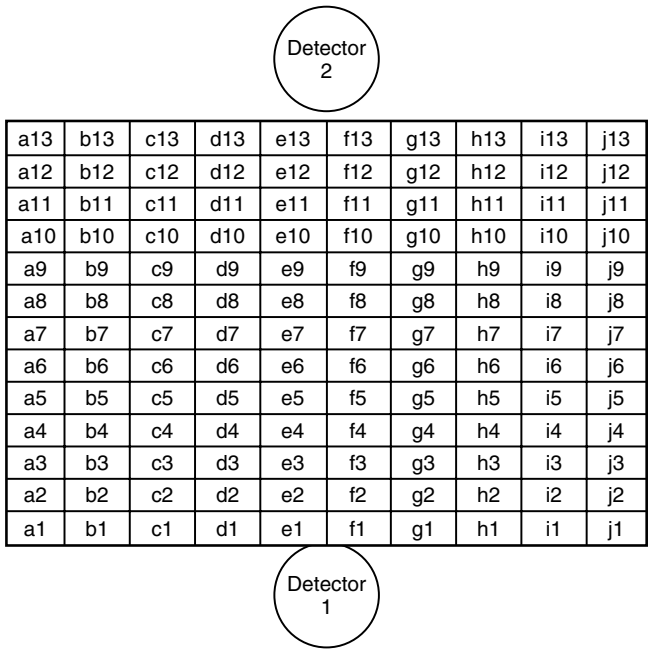

Figure 5 Schematic top view of the source positions measured in the 40×40 cm^2 sample cells. The sample was divided to 10 planes parallel to the detectors' plane.

Table 7 Iteration calculations of the source-to-detector distance (cm) for row b of the experimental results

	y_{real}	x_{real}	Before iterations	Initial estimate	First iteration		Second iteration		Third iteration		After iterations	
			y	x	y	x	y	x	y	x	y	x
b01	4	0	10.95	3.79	6.04	1.54	**6.04**	–	–	–	6.04	1.54
b02	4	3	9.27	5.64	5.23	2.29	3.18	**2.29**	–	–	3.18	2.29
b03	4	6	7.97	7.58	5.31	4.68	3.33	**4.68**	–	–	3.33	4.68
b04	4	9	6.25	10.16	3.14	7.86	**3.14**	–	–	–	3.14	7.86
b05	4	12	4.85	12.41	3.22	10.63	**3.22**	–	–	–	3.22	10.63
b06	4	15	3.93	15.20	3.22	14.07	**3.22**	–	–	–	3.22	14.07
b07	4	20	3.29	19.61	3.29	19.52	**3.29**	–	–	–	3.29	19.52
b08	4	25	3.98	23.92	3.31	24.84	**3.31**	–	–	–	3.31	24.84
b09	4	28	4.77	26.40	4.13	27.90	3.24	**27.90**	–	–	3.24	27.90
b10	4	31	6.27	29.16	4.92	31.29	3.48	**31.29**	–	–	3.48	31.29
b11	4	34	7.75	31.70	5.31	34.44	3.50	**34.44**	–	–	3.50	34.44
b12	4	37	9.11	33.71	5.42	36.92	3.31	**36.92**	–	–	3.31	36.92
b13	4	40	10.42	35.46	7.28	37.61	5.48	39.07	3.52	**39.07**	3.52	39.07
Average absolute error (cm)			2.95	1.91	1.19	0.84	–	–	–	–	0.79	0.73

Table 8 Pre- and postiteration distance calculations between the real source position and the calculated one

Row, y_{real} →		a, $y = 0$		b, $y = 4$		c, $y = 8$		d, $y = 12$		e, $y = 16$	
Line	x_{real}	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations
1	0	10.36	0.00	7.92	2.56	5.60	0.07	3.39	0.27	1.74	1.56
2	3	8.06	3.74	5.89	1.08	4.08	1.06	2.63	1.21	1.60	1.24
3	6	6.07	3.37	4.27	1.48	2.74	1.45	2.03	1.47	1.65	1.41
4	9	3.98	3.37	2.53	1.43	1.41	1.39	1.09	0.95	1.27	1.07
5	12	2.08	2.85	0.94	1.58	0.33	1.29	0.33	0.96	0.83	0.75
6	15	0.88	1.39	0.21	1.21	0.34	1.12	0.19	0.63	0.69	0.63
7	20	0.59	0.78	0.81	0.85	0.90	0.93	0.37	0.39	0.55	0.55
8	25	1.36	0.19	1.08	0.71	0.80	0.89	0.44	0.59	0.30	0.23
9	28	2.50	0.38	1.77	0.76	1.18	0.89	0.74	0.33	0.68	0.54
10	31	4.09	0.68	2.92	0.60	1.90	0.70	1.05	0.35	0.78	0.65
11	34	5.99	0.78	4.40	0.67	2.87	0.79	1.45	0.67	1.13	0.92
12	37	8.11	0.35	6.07	0.69	4.14	0.54	2.41	0.42	1.10	0.73
13	40	10.24	3.30	7.86	1.05	5.68	0.86	3.41	0.55	1.63	1.36
Average		4.95	1.63	3.59	1.13	2.46	0.92	1.50	0.68	1.07	0.90

Row, y_{real} →		f, $y = 20$		g, $y = 24$		h, $y = 28$		i, $y = 32$		j, $y = 36$	
Line	x_{real}	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations
1	0	1.38	0.73	3.20	0.30	5.70	0.24	8.09	0.65	10.47	2.87
2	3	1.22	0.92	2.16	1.03	3.90	1.31	5.89	1.50	8.19	1.30
3	6	1.67	1.51	2.00	1.28	3.04	1.13	4.34	1.56	6.33	1.61
4	9	1.10	0.92	1.35	0.88	1.99	0.89	3.08	1.08	4.61	1.61
5	12	1.20	1.13	0.83	0.59	1.04	0.70	1.76	0.90	3.00	1.50
6	15	1.07	1.04	0.32	0.25	0.24	0.75	0.68	0.75	1.49	1.00
7	20	1.32	1.32	0.47	0.47	0.14	0.16	0.14	0.16	0.77	0.78
8	25	1.32	1.28	0.79	0.54	0.75	0.27	1.11	0.39	2.17	1.15
9	28	1.10	0.99	1.00	0.34	1.43	0.21	2.00	0.41	3.28	1.33
10	31	1.28	1.12	1.35	0.50	2.05	0.62	3.01	1.02	4.69	1.54
11	34	1.50	1.32	1.95	0.93	3.04	1.01	4.62	1.10	6.35	2.94
12	37	1.40	1.07	2.51	0.86	4.08	0.87	5.99	1.17	8.16	1.20
13	40	1.57	1.15	3.31	0.42	5.46	0.28	7.78	0.42	10.39	0.84
Average		1.32	1.12	1.63	0.64	2.53	0.65	3.73	0.86	5.38	1.51

(calibrating) source positioned at the center of the y axis (on the central plane (initial guess)), connecting the two detectors 1 and 2 (at $x = 0$, 5 cm from detector $2^{(2)}$). This $R(0)$ was assigned $R(0)_x$. Then, the calibrating source was moved to the x axis (right side of the box, $y = 0$) at the calculated x (not to the exact value, but to one of the 13 values of the cells; to the cell closest to the calculated value) and $R(0)$ was measured again using the two detectors on the y axis (detectors 3 and 4, on the other two parallel faces of the box). This $R(0)$ was assigned $R(0)_y$ and was used to calculate the y coordinate, according to Equation 29, using $R(x)$, also from detectors 3 and 4. Then, the calibrating source was moved to the y axis at the calculated y (as close as possible to the calculated y cell). $R(0)_x$ was measured again and a new value of x was calculated. This process was repeated until either x or y was the same as in the previous iteration.

Table 7 shows the various calculations of x and y done by iterations. Table 7 presents, as an example, the iteration calculations of row b (Figure 5) of the experimental results. For most points, two iterations were performed after the initially estimated calculation of x, until x or y recurred. It was found that one iteration (two consecutive calculations of y and x) after the initial estimate already yielded a major improvement to the accuracy.

Table 8 shows the distance between the real source position and the calculated one, $\sqrt{(y_{\text{cald}} - y_{\text{real}})^2 + (x_{\text{cald}} - x_{\text{real}})^2}$ both for x and y calculated with central $R(0)$ for both coordinates and the distances calculated after completing all the iterations. One to four iterations are required, depending on the source's position. Iteration calculations reduce the average calculated distance between the real source position and the calculated one from 2.46 to 1 cm and reduce the maximal distance from above 10 to <4 cm.

REFERENCES

1. O. Presler, O. Pelled, U. German, Y. Leichter, Z.B. Alfassi, 'Determination of a Source in a Box with Two Detectors. I Non-absorbing Media', *Nucl. Instrum. Methods*, **A 491**, 314 (2002).
2. A. Notea, 'The Ge(Li) Spectrometer as a Point Detector', *Nucl. Instrum. Methods*, **91**, 513 (1971).
3. K. Debertin, R.G. Helmer, *γ - and X-Ray Spectrometry with Semiconductor Detectors*, Elsevier, New York, 250–255, 1988.
4. O. Presler, U. German, O. Pelled, Z.B. Alfassi, 'The Validity of the Virtual Point Detector Concept for Absorbing Media', *Appl. Radiat. Isot.*, **60**, 213 (2004).
5. O. Presler, O. Pelled, U. German, Y. Leichter, Z.B. Alfassi, 'Off-center Efficiency of HPGe Detectors', *Nucl. Instrum. Methods Phys. Res.*, **484A**, 462 (2002).
6. Z.B. Alfassi, O. Pelled, U. German, 'The Virtual Point Detector Concept for HPGe Planar and Semi-planar Detectors', *Appl. Radiat. Isot.*, **64**, 574 (2006).
7. S. Mahling, I. Orion, Z.B. Alfassi, 'The Dependence of the Virtual Point Detector on the HPGe Detector Dimensions', *Nucl. Instrum. Methods Phys. Res.*, **A557**, 544 (2006).
8. O. Presler, U. German, V. Pushkarski, Z.B. Alfassi, 'Virtual Point Detector: on the Interpolation and Extrapolation of Scintillation Detectors Counting Efficiencies', *Nucl. Instrum. Methods Phys. Res.*, **A565**, 704 (2006).
9. Z.B. Alfassi, N. Lavi, O. Presler, V. Pushkarski, 'HPGe Virtual Point Detector for Disk Radioactive Sources', *Appl. Radiat. Isot.*, **65**, 253 (2007).
10. O. Presler, U. German, H. Golan, Z.B. Alfassi, 'Determination of a Source in a Box with Two Detectors. The General Case', *Nucl. Instrum. Methods Phys. Res.*, **A527**, 632 (2004).
11. J. Feldman, R. Presler, M. Segal, Y. Sabov, O. Presler, W. Weinstein, U. German, Z.B. Alfassi, 'The Localization of Gamma Emitting Point Source in a Long Rectangular Medium Using Two Detectors on the Parallel Faces', *to be published*.
12. M.A. Lopez, T.N. Bravo, 'A Low Energy Germanium Detector System for Lung Counting at the WBC Facility of Ciemat', *Radiat. Prot. Dosim.*, **89**, 221 (2000).
13. J.L. Webb, M. Gadd, F. Bronsen, O. Tench, 'An Evaluation of Recent Lung Counting Technology', *Radiat. Prot. Dosim.*, **89**, 325 (2000).
14. K. Rhoads, J.A. Mahaffey, C.L. Sanders, 'Distribution of Inhaled PuO_2 in Rat and Hamster Lung', *Health Phys.*, **42**, 645 (1982).
15. G.H. Kramer, L.C. Burns, S. Yiu, 'Lung Counting: Evaluation of Uncertainties in Lung Burden Estimation Arising from a Heterogeneous Lung Deposition Using Monte Carlo Code Simulations', *Radiat. Prot. Dosim.*, **74**, 173 (1997).
16. O. Pelled, U. German, G. Pollak, Z.B. Alfassi, 'Estimation of Errors Due to Inhomogeneous Distribution of Radionuclides in Lungs, *Nucl. Instrum. Methods Phys. Res.*, **A564**, 491 (2006).
17. Z.B. Alfassi, 'The Vector Analysis of Multi Measurements Identification', *J. Radioanal. Nucl. Chem.*, **266**, 245 (2005).
18. S.E. Stein, D.R. Scott, 'Optimization and Testing of Mass Spectra Library Search Algorithms for Compound Identification', *J. Am. Soc. Mass Spectrom.*, **5**, p.859 (1994).
19. O. Pelled, S. Tzroya, U. German, G. Haquin, Z.B. Alfassi, 'Locating a ''Hot Spot'' in the Lungs when Using an Array

of Four HPGe Detectors', *Appl. Radiat. Isot.*, **61**, 107 (2004).

20. O. Pelled, S. Tzroya, U. German, A. Abraham, Z.B. Alfassi, 'Improved Localization of a ''Hot Spot'' in the Lungs for an Array of Four HPGe Detectors – the Simultaneous Use of Two Gamma Energies', *Nucl. Instrum. Methods Phys. Res.*, **A551**, 553 (2005).
21. O. Pelled, A. Abraham, U. German, Z.B. Alfassi, 'Localization of a ''Hot Spot'' of Uranium in the Lungs by an Array of Four HPGe Detectors. The Effect of Variation in the Isotopic Composition', *Nucl. Instrum. Methods Phys. Res. A*, **A584**, 406 (2008).
22. Z.B. Alfassi, 'On the Normalization of a Mass Spectrum for Comparison of Two Spectra', *J. Am. Soc. Mass Spectrom.*, **15**, 335 (2004).
23. S.E. Stein, 'An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data', *J. Am. Soc. Mass Spectrom.*, **10**, 770 (1999).
24. O. Pelled, Ph.D. thesis, presented to Ben Gurion University, Beer Sheva, Israel, 2006 approved 2006.
25. S. Dubinski, A. Talmor, O. Presler, A. Tshuva, I. Yaar, I. Orion, Z.B. Alfassi, 'The Determination of a Neutron Source Position in an Unknown Homogeneous Medium; the Planar Case', *Nucl. Instrum. Methods Phys. Res.*, **A548**, 555 (2005).
26. S. Dubinski, O. Presler, Z.B. Alfassi, 'The Determination of a Neutron Source Position in an Unknown Homogeneous Medium; the 2D Case', *Radiat. Prot. Dosim.*, (2007) DOI:10.1093/rpd/ncm048.
27. S. Dubinski, O. Presler, Z.B. Alfassi, 'The Localization of a Neutron Source in an Unknown Homogeneous Medium; the 3D Case', *Nucl. Instrum. Methods Phys. Res.*, **574A**, 315 (2007).
28. O. Presler, S. Dubinski, Z.B. Alfassi, 'The Determination of a Neutron Source Position in an Unknown Homogenous Medium, the 3D Case', *Radiat. Prot. Dosim.*, (2007) DOI:10.1093/rpd/ncm050.
29. S. Dubinski, O. Presler, Z.B. Alfassi, 'An Iterative Method for the Localization of a Neutron Source in a Large Box (container)', *Nucl. Instrum. Methods Phys. Res.*, **583A**, 439 (2007).

Aerosols and Particulates Analysis: Indoor Air

Radu Branisteanu
Institute of Public Health, Iasi, Romania

1 Introduction 1
2 From Respiratory Toxicology to Industrial Hygiene 2
2.1 Deposition of Particulates in Human Airways 2
2.2 The Behavior of the Liquid Particle Aerosols 2
2.3 Characteristics of Interest 2
2.4 Aerosol Classification 3
3 Sources and Types of Aerosols in Indoor Air 4
3.1 Industrial Environment 4
3.2 Home 9
3.3 Nonindustrial Workplaces 10
3.4 Transportation 10
4 Factors Influencing Sampling, Storage and Analysis of Aerosols 11
4.1 Intrinsic Factors 11
4.2 Factors Related to Sampling Strategy 12
5 Industrial Hygiene Applications 12
5.1 Sampling Methodology 12
5.2 Identification and Quantitation of Workplace Pollutants 13
5.3 Assessment of Human Exposure 17
6 Nonindustrial Applications 18
6.1 Objectives and Procedures 18
6.2 Analysis of Bioaerosols 19
7 Regulations and Air Standards 19
8 Comparison Between Industrial and Nonindustrial Indoor Environments 20
Abbreviations and Acronyms 21
Related Articles 21
References 22

This article is intended to provide information and directions needed in the assessment of indoor aerosols, both in industrial and nonindustrial environments. Section 2 of this material is dedicated to basic explanations needed to understand the scope and directions of the analysis of aerosols in indoor air. It comprises brief considerations about the deposition of particulates in human airways, the characteristics of the particles and aerosol classification. Section 3 describes some widespread types of aerosols found in occupational and domestic environments. The characterization comprises considerations upon chemical composition, size, shape and the influence upon human health. Methods of analysis are also mentioned. Section 4 describes the factors that influence the sampling and storage of aerosols. Section 5 deals with applications of aerosol analysis in industrial hygiene. Sampling, sample preparation and analyse are exemplified on three selected compounds: metallic species, polycyclic aromatic hydrocarbons (PAHs) and silica. Alternative methods for specific applications are presented. Section 6 refers to home and office environments and approaches strategies used for indoor air quality (IAQ) assessment and the analysis of bioaerosols. Section 7 addresses the air standards and their influence upon analytical methodology and instrumentation. Section 8 comprises a brief discussion regarding differences encountered in aerosol analysis in industrial and nonindustrial environments.

1 INTRODUCTION

There is no doubt that exposure to indoor aerosols represents a cause of concern. Humans breathe the air no matter whether they work, eat, sleep or relax. The levels of contaminants in buildings depend on the type of activity and the number of inhabitants, but also on general economic development.

Traditionally, the studies dedicated to human exposure to airborne toxins contained within buildings focus separately on occupational and non-occupational environments. However, the enclosed spaces with human activity comprise either industrial or nonindustrial environments. Apart from the well-known differences, these two types of environment have in common a very important element: the human being. Therefore, IAQ represents in fact an integrated issue, as the two faces of the same coin.

The assessment of human exposure to indoor aerosols is a multidisciplinary action that includes the identification and quantitation of airborne toxins. In order to obtain representative and reliable analytical results, the comprehensive understanding of the physical, chemical and toxicological behavior of the aerosols is essential. The chemist should have all that information in order to choose the analytical method that gives the best cost/performance ratio in a particular application.

This article is intended to provide information and directions needed in the assessment of indoor aerosols, both in industrial and nonindustrial environments. It approaches different aspects, such as general toxicological

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

knowledge, aerosol characteristics, sampling methodology and laboratory analysis. Regulations and conventions are also included.

2 FROM RESPIRATORY TOXICOLOGY TO INDUSTRIAL HYGIENE

2.1 Deposition of Particulates in Human Airways

Upon inhalation, particles are arrested and deposited differentially according to the location in the respiratory tract (RT). There are four main mechanisms that contribute to particle deposition in the human airways: interception, inertial impaction, gravitational sedimentation and radial diffusion. In the case of the first three mechanisms, the deposition efficiency increases with the increase of the particle diameter. With respect to radial diffusion, the deposition efficiency increases with the decrease of the particle diameter. The electrostatic attraction/repulsion is a mechanism that contributes to a lesser extent to the particle deposition in the human RT. Detailed description of these processes can be found in Lippmann,[1] ICRP,[2] and Rando.[3]

The particle size represents the key parameter which influences deposition into the RT. The size of the particles that deposit due to the impaction and sedimentation mechanisms is described in terms of aerodynamic diameter, d_{ae}. The aerodynamic diameter is defined as the diameter of a unit density sphere having the same terminal settling velocity as the particle in question. For very small particles which deposit primarily by diffusion the size is described in terms of thermodynamic diameter, d_{th}. The d_{th} is equivalent to the particle volume equivalent diameter. The transition from d_{ae} to d_{th} occurs in the region of 0.5–1 μm.

Figure 1 shows a typical pattern of particle deposition in human RT obtained by Freijer[4] using computer simulation.

For the purpose of analysis, the RT is divided into five compartments, each representing a region. The stacked graph expresses the cumulative fraction deposited in the five regions of RT as a function of d_{ae}. The total fraction deposited is high (90–99%) for particles with d_{ae} of 0.001, 0.005 and 5 μm, intermediate (27–75%) for 0.02 and 1 μm, whereas for particles in the range of 0.1–0.5 μm the deposition is low (12–17%). It should also be noted that the particles < 0.003 μm and particles > 0.5 μm deposit predominantly in the extrathoracic regions, while particles between >0.003 and <0.5 μm deposit mainly in bronchial and alveolar regions.

The particles trapped in the nasal region are carried out by the mucus towards the pharynx and swallowed. The particles deposited in the bronchial and bronchiolar regions are continuously evacuated in the oral cavity by the ciliated ephithelium. Upon arrival, they are swallowed or expectorated. The particles arrested in the alveolar region are eliminated more slowly and spend more time in the lung. These clearance mechanisms remove the particles deposited in the airways and transfer the burden to the digestive tract. Therefore, the toxins contained in the particles may be absorbed in the organism by dissolution and absorption either from the RT or from the digestive tract.

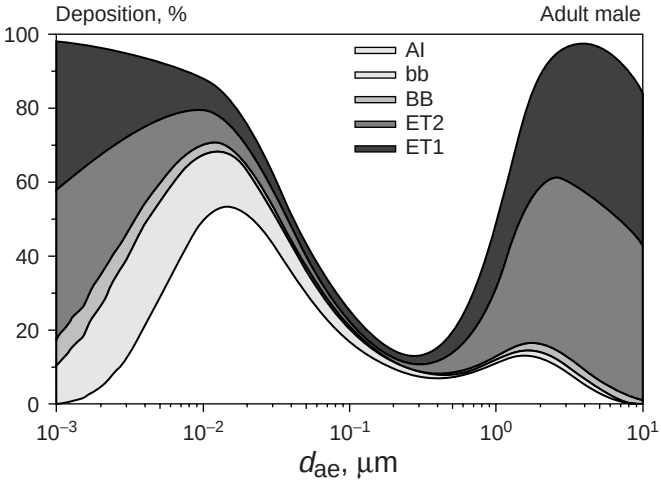

Figure 1 Deposition fraction in different regions of the RT as a function of aerodynamic diameter. Key: AI = alveolar-interstitial region, bb = bronchiolar region, ET1 = extrathoracic region with the anterior nasal passages, ET2 = extrathoracic region with the posterior nasal passages, the naso- and oropharynx and the larynx. [Reproduced by permission of National Institute of Public Health and the Environment (RIVM) from J.I. Freijer, F.R. Cassee, van Bree, The Netherlands, 47 (1997).]

2.2 The Behavior of the Liquid Particle Aerosols

Aerosols having the disperse phase composed of liquid droplets exhibit a different behavior inside the human RT. After inhalation, liquid particles, which are very small, may reach all parts of the lung, from where they may be readily absorbed. The volatility of the material may also lead to concentrations that rise to values far above those possible from partial pressures of the gaseous phase at room temperature. This is the case of some solvents, for which the vapors are relatively harmless at ambient temperature but are hazardous when they exist as a mist.

2.3 Characteristics of Interest

Human exposure to indoor aerosols and particulates depends on a number of determinants pertaining to the nature of the contaminant, the characteristics of the confined space (e.g. microclimate, ventilation, etc.) and

the individual traits of the exposed subject. The physical and chemical properties of the aerosols influence their emission and behavior in the atmosphere, but also the deposition and uptake, once inhaled by humans. Some of these characteristics may be measured.

1. *Chemical composition*. The chemical composition of aerosols and particulate matter is the main determinant of toxicological interest. It reflects the action of toxins upon the living organisms and gives information about other properties of interest such as vapor pressure, solubility, and reactivity. Chemical composition influences the strategy of collection and analysis of aerosols.
2. *Mass*. The action of the toxins upon living organisms depends on both the quality and the quantity of the compounds. The amount of toxin in the air is quantified in terms of concentration, which, in most cases, represent the mass for a given volume of air (e.g. $mg\,m^{-3}$). In the case of aerosols, the concentration may refer to the entire quantity of particles present in the air or to a particular toxin previously quantified by chemical analysis.
3. *Number*. The number of particles is also a quantitative measurement. In the assessment of exposure to fibers the airborne concentration of particles is expressed in terms of number for a given volume of air (number m^{-3}). Sloff and Blokzijl[5] mention that the number of fibers for a given quantity of dust (e.g. number ng^{-1}) may be sometimes encountered.
4. *Size*. Particle dimensions influence the behavior of a certain aerosol both in the surrounding atmosphere and inside the human RT. Smaller particles spend more time in the gaseous medium before they settle but penetrate deeper into the lung compartments.
5. *Shape*. The shape of the particulates influences their aerodynamic behavior. The ratio length: diameter constitutes a criterion for the counting rules employed in fiber exposure assessment.
6. *Surface properties*. In some cases, the level of biological activity is strongly influenced by the surface properties of particulates (e.g. adsorption capacity, hydrophilicity, surface radicals, defects, surface impurities). These characteristics cannot be determined by chemical analysis of bulk samples nor by conventional microscopy, and surface analysis must be performed.
7. *Vapor pressure*. Apparently, vapor pressure is a characteristic with low implication in aerosol assessment. However, it determines the saturated vapor concentration in air. At concentrations above this value, a less volatile material will partition between the vapor phase and the aerosol (condensed) phase. In addition, the adsorption of very volatile pulmonary irritants onto particulates may increase their injury potential for the deeper parts of RT since the cleanup action of the upper airways is bypassed. A thorough explanation of these processes is that by Rando[3] (see also section 2.2).

2.4 Aerosol Classification

The term aerosol refers to ''a dispersion, in a gaseous medium, of solid or liquid particles which have a negligible settling velocity'' – World Health Organisation (WHO).[6] The solid particles form ''solid particles aerosols'' (and not ''solid aerosols''), the liquid particles form ''liquid particles aerosol'' (and not ''liquid aerosols''). Particulate pollutants are usually referred to as airborne dispersions of solids that have essentially no vapor phase present (e.g. silica particles, coal dust).

From a practical viewpoint it is impossible to classify an aerosol according to a single criterion. There are several forms of classification based mainly on their physical state, behavior, and action upon living organisms. Some common forms of classification follow.

2.4.1 Physical Classification (Common Terms Used in Describing Aerosols)

Dust is a loose term used for solid particle aerosols. It usually refers to particulates predominantly larger than colloidal dimensions and capable of temporary suspension in air or other gases. Dust is commonly obtained from larger masses through the application of physical force.

Fume refers to solid particles generated by condensation from gaseous state, usually after volatilization from melted substances. Fume formation often implies a chemical reaction such as oxidation. Popular usage includes any type of contaminant.

Smoke refers to particles resulting from incomplete combustion of organic matter and consisting predominantly of carbon and other combustible materials.

Mist is a term applied to liquid particle aerosols composed mainly of particles large enough to be seen by the naked eye.

Fog is a loose term applied to liquid particle aerosols of natural provenance. Formation by condensation is implied.

Smog represents a combination of smoke and fog implying natural and man-made sources.

Bioaerosols is a loose term used to describe airborne particles of biological origin. According to American Conference of Governmental Industrial Hygienists (ACGIH)[7] bioaerosols include microorganisms and

fragments, toxins and particulate waste products from all varieties of living things.

The usual types of aerosols encountered in indoor environment include dust, fume, smoke, mist, and bioaerosols.

2.4.2 Classification According to Dimensions and Behavior

Settlable Particulates refers to particles large enough to settle out of the air fairly fast. The term implies particles usually larger than 30 μm, but in still air some particles of 10 μm or smaller may settle. The settled particles are sometimes collected for bulk analysis.

Suspended Particulates refers to aerosol particles that tend to remain suspended in the atmosphere for long periods of time. The term commonly addresses particles of 10 μm and smaller. Alternative names are total suspended particulate (TSP) or suspended particulate matter (SPM).

Condensation Nuclei – sometimes called Aitken nuclei – refers to particle in atmospheric suspension having a diameter smaller than 10^{-1} μm but greater than 10^{-3} μm.[6] Due to their small dimensions, the Aitken particles do not act as normal condensation sites except as for the important super-saturated vapors in the atmosphere.

Agglomerates refers to particles composed of several smaller particles, attracted to a large particle or to each other, that travel together in the atmosphere as a single particle.

There is no sharp cut-off between particle size in any of these definitions and the dimensions of the particles in each class overlaps its neighboring class. Several factors such as shape (which affects the aerodynamic properties), density and velocity, and electrical charge affect the proper collection and classification of the particles.

2.4.3 Classification According to Industrial Hygiene Conventions and Regulations

Inhalable fraction refers to particles that may deposit anywhere in the RT.

Thoracic fraction refers to particles that, upon inhalation, deposit mainly in the thoracic region of the RT.

Respirable fraction refers to particles that, upon inhalation, deposit mainly in the alveolar region. For complete definitions of inhalable, thoracic and respirable fractions, see also section 7.

PM 10 (Particulate Matter 10) refers to particles equal to or smaller than 10 μm.

PM 2.5 (Particulate Matter 2.5) refers to particles equal to or smaller than 2.5 μm.

Ultrafine particles usually refer to particles having a diameter smaller than 0.1 μm. The toxicological significance of the ultrafine particles is still under research. Therefore, this term is encountered in the exposure regulatory acts only in few countries. Such an example is The German List of Maximum Allowable Concentrations in Workplaces.[8]

2.4.4 Classification According to Health Effects

Irritants are materials which possess a corrosive or vesicant action. They may affect the human RT but also the eyes and the skin. Typical examples are the acid mists and alkaline dusts. *Fibrosis-producing dusts* represent solid particle aerosols that cause lung fibrosis (e.g. silica and asbestos). *Systemic poisons* are those aerosols that may cause organic injury to several tissues or organs (e.g. toxic metals such as lead, mercury, cadmium, etc.). *Allergens* apply especially to dusts causing allergic reactions (e.g. home dust mites). *Carcinogens* represent aerosols that contain substances with carcinogenic potential (e.g. nitrosamines, asbestos etc.).

It should be noted that classification according to the health effect does not allow a sharp distinction between various types of aerosols, because it is frequently impossible to place a material in a single class. For example, asbestos may be regarded both as a fibrosis-producing and carcinogen toxin.

3 SOURCES AND TYPES OF AEROSOLS IN INDOOR AIR

3.1 Industrial Environment

The industrial environment is characterized by the presence of various types of fumes, mists, dusts and smoke, according to the type of industry and the technological processes involved.

3.1.1 Metallic Fumes

Metallic fumes are widespread contaminants in industry. Among these, welding fumes are of particular importance since they may be encountered in any industrial environment. According to WHO,[9] welders constitute between 0.5 and 2% of the total working populations with elevated fractions for certain industrial sectors.

The target organs in exposure to welding fumes are the eyes, skin, respiratory system, and the central nervous system. Symptoms include metal fume fever, dyspnea, cough, and muscle pain. Ozdemir[10] pointed out that welders have increased risk of chronic bronchitis and impairment of pulmonary function. Antonini[11–13] revealed that freshly generated stainless steel welding fumes were more biologically active than the ''aged'' ones. In 1990 the International Agency for Research on Cancer (IARC) declared welding fumes as possibly carcinogenic

to humans (Group 2B) with lung as the cancer site. In addition, Gustavsson[14] emphasized an association between welding fume exposure and an increased risk of pharyngeal and laryngeal cancer. However, there are studies that do not agree with significant excess of lung cancer, such as that of Danielsen.[15] The complexity and diversity of the biological effect is strongly related to the welding technology.

Roughly, four techniques account for 60–80% of all welding. Manual metal arc welding (MMA) and tungsten inert gas (TIG) are performed with hand-held electrodes with or without an inert gas shielding. The corresponding semi-automatic processes are generically entitled metal active gas (MAG) and metal inert gas (MIG). Other techniques encountered in practice are oxyacetylene (autogen), submerged arc, laser beam and electron-beam welding. The materials commonly welded are mild steel, stainless steel and aluminum. Inert gas techniques produce smaller quantities of fumes. In fumes generated by MMA the relative concentration of the elements may be very different from those in the consumables. Some elements with lower boiling points (F, Mn, Zn, Pb, As, Ca, Si) are significantly enriched in the fume over their sometimes trace concentration in the consumable or the parent metal. For the MMA/mild steel technique, the primary constituents of the fumes are Fe, Mn, Si, Na, Ca, F, while stainless steel welding introduces Cr and Ni as additional compounds. Fumes generated by MIG welding are less complex because of the absence of flux coating. If mild steel is welded, the primary constituents are Fe, Mn, Si and their oxides. If stainless steel is welded, Cr and Ni are also found. The hexavalent chromium content is dependent on the welding technique. MMA/stainless steel fumes contain approximately 3–4% Cr(VI) while MIG/stainless steel fumes contain 14–18%. In the case of the MMA, the crystalline fraction in the fumes may fall in a very wide range, from 20 to 90% of the composition. The predominant crystalline substance is Fe_3O_4 but X-ray diffraction (XRD) also revealed other compounds such as NaF, CaF, $KCaF_3$, MgO, K_2CO_3, Na_2CO_3, $MnFe_2O_4$ and KCF_3.

MIG and MMA produce fumes with a mass median d_{ae} in the range of 0.1–0.5 μm which may deposit predominantly in the lower RT. Larger particles are also generated due to the spray from the arc. Hewett[16,17] emphasized that the particle size distribution and physical properties of the fumes depend more on the welding technique than on the welded material. Inert gas welding techniques generate smaller particles regardless of whether mild or stainless steel is welded.

Due to the above mentioned factors, welding fumes must frequently be tested for individual constituents (see also section 5.2.1). If no toxic elements are present in the welding rod, or in the welded metal or metal coating, conclusions based on total concentration (gravimetric assay) are adequate.

Other branches of industry, such as metallurgy and metal surface treatment shops, are also assessed for potential exposure to metallic as well as nonmetallic fumes. These fumes may contain metallic species already mentioned earlier, but also cobalt fume, Co [CAS 7440-48-4], copper fume, CuO [CAS 1317-38-0], magnesium fume, MgO [CAS 1309-48-4], rhodium fume, Rh [CAS 7440-16-6], vanadium fume, V_2O_5 [CAS 1314-62-1] and zinc chloride fume, $ZnCl_2$ [CAS 7646-85-7]. The primary target for the metallic fumes is the respiratory system and the impact may vary from symptoms like ''fume fever'' and muscle ache to diseases such as occupational asthma. Of a distinctive toxicological significance is the cadmium fume. It contains mostly CdO [CAS 1306-19-0] and is a potent occupational carcinogen (Group 1 in the IARC classification).

The ammonium chloride fume, NH_4Cl [CAS 12125-02-9] is a widely encountered nonmetallic aerosol, mainly in metallic surface treatment shops. This type of fume is primarily an irritant due to the emission of HCl [CAS 7647-01-0] which takes place in contact with moist surfaces (such as the inner surface of RT).

3.1.2 Mists

Mists, particularly acid mist and oil mist, are widely encountered in the industrial environment.

Acid mist is a loose term describing a liquid particle aerosol composed of strong inorganic acids. The acid mists are primarily irritants but in 1992 IARC[18] classified occupational exposure to strong inorganic acid mists containing sulfuric acid as being carcinogenic to humans (Group 1). Further studies such as of Steenland[19] confirmed this classification.

Strong inorganic acids may be encountered in the workplace atmosphere as mist, vapor and gas. The inorganic acids widely encountered as mist are: hydrochloric acid (mostly gaseous but also as mist), nitric acid [CAS 7697-37-2], sulfuric acid [CAS 7664-93-9] and phosphoric acid [CAS 7664-38-2]. Hydrogen bromide [CAS 10035-10-6] is also encountered but to a lesser extent. These acids are usually sampled and analyzed together. For example, the Occupational Safety and Health Administration (OSHA) analytical method ID-165SG quantifies HBr, HNO_3, H_3PO_4 and H_2SO_4 using the ion chromatography (IC) technique. The industries in which there is potential exposure to strong inorganic acids are listed in Table 1.

Another mist of particular toxicological concern is the chromic acid mist. It contains CrO_3 [CAS 1333-82-0] and chromates. Due to the content of chromium compounds, IARC (1990) classified chromic acid mist as human carcinogen (Group 1). Recent studies such as of

Table 1 Industries in which there is potential exposure to strong inorganic acids

Industry	Sulfuric acid	Hydrochloric acid	Nitric acid	Phosphoric acid
Aerospace		+	+	
Building and construction	+	+	+	+
Chemical manufacture	+	+	+	+
• isopropanol	+			
• sulfuric acid	+			
• synthetic ethanol	+			
• vinyl chloride		+		
Detergents	+			
Electric and electronic equipment	+	+	+	+
Fertilizers	+		+	+
Food products	+	+	+	+
Health services	+	+	+	+
Instruments	+	+	+	+
Lead-acid batteries	+			
Leather	+	+	+	+
Metal plating	+	+	+	+
Metal cleaning and pickling	+	+	+	+
Metal extraction and ore processing	+	+	+	+
Oil and gas extraction	+	+		+
Petroleum and coal products	+	+	+	+
Photography shops	+	+	+	+
Printing and publishing	+	+	+	+
Paper and allied products	+	+	+	+
Rubber and plastic products	+	+		+
Semiconductors		+	+	
Soap	+			
Steel	+	+	+	
Textile products	+	+	+	+
Zinc galvanizing		+		

Reproduced by permission of IARC from *IARC Monographs on the Evaluation of Carcinogenic Risk to Humans*, Vol. 54, 9, 1992.

Sorahan et al.[20] are consistent with this classification. The exposure to chromic acid mist occurs mainly in the metal plating industry.

The size distribution of acid mists is practically unknown. Routine analytical methods, such as of OSHA method ID-215 for the determination of hexavalent chromium, do not imply size-selective collection of this type of aerosol.

Oil mists are in general referred to aerosols composed of white mineral oil. Alternative names of some water-insoluble petroleum-based cutting oils are sometimes used: cable oil, cutting oil, drawing oil, engine oil, heat-treating oils, hydraulic oils, machine oil and transformer oil. A more detailed definition refers to mist composed of heavy mineral oil [CAS 8012-95-1] having the general formula C_nH_{2n+2} ($n \geq 16$).

Mineral oil mists may be encountered in industries which involve processing of metallic components such as machine-tool construction, ball bearing manufacturing, automotive and aircraft construction, but also in small workshops in which metal cutting is performed. Industries involving the use of heavy machinery (such as turbines or rolling mills) are also potential oil mist environments.

The chemical composition is variable. Fresh oil is supposed to contain only aliphatic hydrocarbons. The burned oils resulting from the lubrication of the mechanisms that operate at high temperatures also contain aromatic hydrocarbons such as PAHs.

The influence upon human health depends strongly on the nature of the oil. In the case of mineral oil [CAS 8012-95-1], the common symptoms are irritation of the eyes, skin, and RT. In the case of mineral oil [CAS 8002-05-9] (Chemical Abstract Name: Petroleum, synonym "Petroleum distillate") IARC decided that untreated and mildly-treated oils are carcinogenic to humans (Group 1, 1987).

As in the case of the acid mists, not very much is known about the size distribution of the droplets. The routine assays, such as that described in method number 5026 issued by the National Institute of Occupational Safety and Health (USA) (NIOSH), involve collection of the inhalable fraction on filters. However, Kira[21] emphasized the importance of the respirable fraction and suggested that measurements of oil mists have to be size-selective.

3.1.3 Dusts

Dusts represent widely spread contaminants in the workplace that may contain metallic or nonmetallic species.

Metallic dusts contain elements that are found also in metallic fumes. Of particular interest are dusts that contain Cr, Cd, and Pb, because of their carcinogenic potential. Lead dust may be encountered also in nonindustrial environments (see also section 3.2).

Silica dust is one of the most studied chemical hazards. Its fibrogenic action has been known for a long time and the occupational exposure is regulated in most countries. Silica has many crystallographic forms but only crystalline forms are known to impose a serious threat to the human health. Two natural forms, α, β-quartz [CAS 14808-60-7] and α, β-cristobalite [CAS 14464-46-1] are classified as being carcinogens to humans (Group 1). An authoritative review of the silica is that by IARC.[22]

The fibrogenic and carcinogenic potential of silica dusts depend on two factors: the crystalline silica content of the dust and the size of particulates. Recent data revealed that the biological response to crystalline silica may differ for samples identical in their bulk analysis and the discrepancies are to be attributed to surface properties such as hydrophilicity, surface radicals or the presence of some surface impurities.

Table 2 Main activities in which workers may be exposed to crystalline silica

Industry/activity	Specific operation/task	Source material
Agriculture	Ploughing, harvesting, use of machinery	Soil
Mining and related milling operations	Most occupations (underground, surface, mill) and mines (metal and nonmetal, coal)	Ores and associated rock
Quarrying and related milling operations	Crushing stone, sand and gravel processing, monumental stone cutting and abrasive blasting, slate work, diatomite calcination	Sandstone, granite, flint, sand, gravel, slate, diatomaceous earth
Construction	Abrasive blasting of structures, buildings	Sand, concrete
	Highway and tunnel construction	Rock
	Excavation and earth moving	Soil and rock
	Masonry, concrete work, demolition	Concrete, mortar, plaster
Glass, including fiberglass	Raw material processing	Sand, crushed quartz
	Refractory installation and repair	Refractory materials
Cement	Raw materials processing	Clay, sand, limestone, diatomaceous earth
Abrasives	Silicon carbide production	Sand
	Abrasive products fabrication	Tripoli, sandstone
Ceramics, including bricks, tiles, sanitary ware, porcelain, pottery, refractories, vitreous enamels	Mixing, molding, glaze or enamel spraying, finishing	Clay, shale, flint, sand, quartzite, diatomaceous earth
Iron and steel mills	Refractory preparation and furnace repair	Refractory material
Silicon and ferro-silicon	Raw materials handling	Sand
Foundries (ferrous and nonferrous)	Casting, shaking out	Sand
	Abrasive blasting, fettling	
	Furnace installation and repair	
Metal products including structural metal, machinery, transportation equipment	Abrasive blasting	Sand
Shipbuilding and repair	Abrasive blasting	Refractory material
Rubber and plastics	Raw material handling	Sand
Paint	Raw material handling	Fillers (tripoli, diatomaceous earth, silica flour)
Soaps and cosmetics	Abrasive soaps, scouring powders	Silica flour
Asphalt and roofing felt	Filling and granule application	Sand and aggregate, diatomaceous earth
Agricultural chemicals	Raw material crushing, handling	Phosphate ores and rock
Jewellery	Cutting, grinding, polishing, buffing	Semi-precious gems or stones, abrasives
Dental material	Sand blasting, polishing	Sand, abrasives
Automobile repair	Abrasive blasting	Sand
Boiler scaling	Coal-fired boilers	Ash and concretions

Reproduced by permission of IARC from *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, Vol. 68, 62, 1997.

The main activities in which occupational exposure may occur are listed in Table 2.

Fiber dusts (or simply fibers) refer to aerosols constituted mainly of fibrous materials.

Asbestos [CAS 1332-21-4] is a generic name used for a naturally occurring number of silicates divided in two main groups: the serpentine and the amphiboles. Chrysotile [CAS 12007-29-5], a serpentine, is widely used for the reinforcement of various materials (e.g. cement) due to its high resistance to alkalis and high tensile strength. The amphiboles, anthophyllite [CAS 17068-78-9], amosite [CAS 121172-73-5], actinolite [CAS 121172-67-7], tremolite [CAS 14567-73-8] and crocidolite [CAS 12001-28-4] are also used but to a lesser extent. Asbestos is very persistent in indoor environments. Once released in the atmosphere, it spends a long time before it settles and the depositions may be further re-dispersed due to various indoor activities. Chronic inhalatory exposure to asbestos leads to asbestosis, a historically known lung fibrosis, but the great concern arises from its carcinogenic potential. Asbestos is a potent carcinogen classified by IARC in Group 1. All the crystallographic varieties have been linked to excess of lung cancer and mesotheliomas of the pleura and peritoneum. Advised monographs concerning asbestos problems are that by IARC (Vol. 14, Suppl. 7; 1987) or Sloff[5] and Montizaan.[23]

Man made mineral fibers (MMMFs) is a term used to denote the amorphous glassy fibers made from molten

slag, rock or glass. Three are three main types of MMMFs: ''mineral wool'' (made from slag or rock), ''fibrous glass'' (glass wool and continuous filament glass fibers) and ''fine glass fibers'' (diameter 1–3 μm). The latter category include also ''very fine glass fibers'' (diameter < 1 μm). Sometimes the term of MMMFs is replaced by man made vitreous fibers (MMMVs) which covers also ceramic fibers (amorphous) and potassium titanate, silica, and zirconia fibers (crystalline). An informal term encountered in practice is mineral wool fiber.

The hazardous potential to human health implies irritation to RT, eyes and skin. Glasswool, rockwool, slagwool, and ceramic fibers are classified by IARC as possibly carcinogenic to humans (Group 2B, 1988).

The exposure assessment to asbestos and MMMFs relies on the methodology historically developed for asbestos. The concentration in the air is expressed in terms of number of fibers per a given volume of air and is determined by fiber counting. Only the respirable fibers are counted. Respirable fibers are defined as those fibers that have a length (l) $\geq 5\,\mu m$, a diameter (d) $< 3\,\mu m$ and a ratio $l:d \geq 3:1$. The most popular method of fiber counting is that performed by means of phase contrast optical microscopy (PCOM). An example is the method NIOSH 7400. Other microscopic techniques imply the usage of scanning electronic microscopy (SEM) or transmission electronic microscopy (TEM). Notably is that SEM and TEM may be coupled with X-ray microanalysis (such as in NIOSH 7402), thus enhancing the possibility for distinguishing asbestos from other fibers present in the sample. A reference method for the microscopic determination of MMMFs is that by WHO.[24] The fiber concentration in the atmosphere can be measured continuously by means of portable monitors. These devices operate on light scattering principle. The identification and quantitation of asbestos in various products require the analysis of bulk samples. The most frequently encountered techniques are the XRD (e.g. method NIOSH 9000), infrared spectroscopy (IR), polarized light microscopy (e.g. method NIOSH 9002), differential thermogravimetry, light scattering by magnetically orientated fibers and selective staining of asbestos fibers.

Asbestos and MMMFs are used as reinforcing agents for composite materials and also as insulating materials. Therefore, exposure is encountered in a wide variety of industries such as construction materials, automotive and aircraft industry, metallurgy and chemical industry.

Organic fibers is a generic name for fibers produced by chemical synthesis. A review of health data by the International Programme on Chemical Safety[25] is recommended. The organic fibers extensively used in practice are carbon graphite fibers, aramid fibers (e.g. Kevlar®) and polyolefin fibers. The aramid fibers are produced by reaction of aromatic diamines and aromatic diacid chlorides. The polyolefin fibers are synthesized by polymerization of different olefin units such as ethylene or propylene.

Little is known about their hazardous potential excepting the fact that para-aramid fibers may release very fine fibrils by peeling of the relatively large fibers. The para-aramid fibers were further re-evaluated in 1997 by IARC[22] and explicitly classified in Group 3 (unclassifiable as to carcinogenicity to humans). The exposure assessment methodology derives from that of asbestos and MMMFs. Fibers are collected by air filtration and further evaluated by PCOM. Carbon and aramid fibers are used in aerospace, military and other industries for the improvement of composite materials. Polyolefin fibers are used in textile industry.

Natural fibers such as cotton, wool, and hemp are common materials used in the textile industry. Natural fibers exert their hazard potential mostly upon RT. Symptoms include cough and discomfort. Chronic exposure to cotton dust leads to bronchitis and a specific disease called byssinosis. Unlike the mineral fibers, the hazardous potential arises mainly from the very nature of the fiber rather than the size. Consequently, concentration in the air is expressed in terms of quantity per a given volume of air (e.g. $mg\,m^{-3}$) and is determined by gravimetric assay.

Of particular importance is the *wood dust*. Apart from its irritative potential, in 1995 IARC classified wood dust as being carcinogenic to humans (Group 1). The cancer site is the nose sinuses. The hazardous potential seems to arise more from the very nature of the particulates and less from their dimensions. The hard essence woods are more biologically active than the soft ones. The concentration in the air is determined by gravimetric assay.

Alkaline dusts is a loose term referring to particulate matter containing strong alkalis such as NaOH [CAS 1310-73-2], KOH [CAS 1310-58-3], LiOH [CAS 1310-65-2], and basic salts. Alkaline dusts are basic, hygroscopic, caustic aerosols encountered in various branches of chemical industry (petroleum refining, fertilizers and polymer syntheses, detergents and cosmetics manufacturing, etc). The target organs are the respiratory system, eyes and skin. The primary symptom is irritation. Over long time exposures pneumonitis and temporary loss of hair may develop. The alkaline dusts are sampled as inhalable fraction and the alkali content is quantified by titration as in method NIOSH 7401.

3.1.4 Tar and Soot

Workplace exposure to tar and soot occurs mainly in the coke production and coal based power plants. Coke furnaces emit tar through slits around the doors and charging holes. IARC (1987) classified coke production as an exposure circumstance carcinogenic to humans (Group 1). As

in many combustion products, the carcinogenic potential arises from the presence of PAHs. For sampling and analysis of PAHs see section 5.2.2.

3.2 Home

Home air contaminants may originate within the building or be drawn in from outdoors. There are hundreds of types and sources of particles present in the home environment depending on the individual habits, tradition, and the general economic development. Home aerosols comprise fumes, mists, dusts, and particles of biological origin.

3.2.1 Fumes and Mists

Fumes and mists are generated by heating and cooking. These processes involve a large number of ingredients, materials and techniques which strongly depend on the lifestyle of the inhabitants. Smith cited by Gold[26] estimated that about one half of the world's population is exposed to concentrations of PAHs that are orders of magnitude greater than those normally found in areas using more advanced combustion systems. The investigations undertaken in the Chinese province of Xuan Wei, in which nonsmoker women had higher cancer incidence rate due to exposure of fumes originating from smoky coal combustion, revealed a higher PAH airborne burden. Table 3 presents a list of compounds.

Chiang[27,28] and more recently Wu[29] demonstrated that the usage of lard, soybean oil, and peanut oil in Taiwanese kitchens generates fat and oil mist that exhibit mutagenic and genotoxic activity. Chemical analysis of the oily mists revealed extremely high concentrations of dibenz[*a*,*h*]anthracene, benz[*a*]anthracene, and benzo[*a*]pyrene in the microgram range (for CAS numbers see Table 6). These concentrations are of entirely comparable range with workplace concentrations previously reported for various branches of industry in Western Europe by Montizaan.[30]

By contrast, in Western countries aerosols generated by heating and cooking result in much lower concentrations. A thorough discussion of indoor carcinogens is that by Gold.[26]

Environmental tobacco smoke (ETS) is a combination of sidestream smoke released into the air from tobacco burning and mainstream smoke exhaled by the active smoker. ETS is the most widely spread source of indoor pollutants in residences, offices and public places. Tobacco smoke contains several hundreds of toxic compounds both in gaseous form (carbon monoxide, nitrogen dioxide, ammonia, etc.) and particulate-bound form (PAHs, nitrosamines, cadmium, etc.). The size of the aerosol is in the respirable range (usually between 0.2 and 0.5 μm) and varies due to the evaporation or coagulation of the individual components. The composition is also variable according to the tobacco brand and the smoking habit (e.g. cigarette or pipe). The marker compounds used for ETS evaluation are: particulates (assessed both as TSP or respirable fraction), nicotine, carbon monoxide and nitrogen oxides. Lebret et al. cited by Gold[26] suggested cadmium as a possible marker.

Acute effects of exposure to indoor combustion fumes (heating, cooking, ETS) are irritation of RT and eyes, coughing, wheezing, headache and related sinus problems. The carcinogenic potential arises from the content in chemicals such as benzene [CAS 71-43-2], PAHs, and nitrosamines.

3.2.2 Dust

Dust is commonly referred to particles of various origin present in indoor environments. Apart from being annoying, the home dust may impose a serious threat

Table 3 PAH levels in organic extracts of indoor air particles from burning different fuels during cooking in Xuan Wei homes

Hydrocarbon	Smoky coal		Smokeless coal		Wood	
	μg g^{-1} organics	μg m^{-3} air	μg g^{-1} organics	μg m^{-3} air	μg g^{-1} organics	μg m^{-3} air
Benz[*a*]anthracene	1450	32.83	2060	1.01	264	4.20
5-Methylchrysene	729	16.51	464	0.23	2	0.03
Benzofluoranthenes	1370	31.02	3780	1.86	194	3.08
Benzol[*a*]pyrene	850	19.25	1160	0.57	204	3.24
Indeno[1,2,3-*cd*]pyrene	529	11.98	915	0.45	127	2.02
Dibenzo[*a*,*e*]pyrene	771	17.46	1310	0.64	BDL[a]	BDL
Dibenz[*a*,*j*]acridine	31	0.70	14	0.007	BDL	BDL

[a] Below detection limit.

[Reproduced by permission of IARC from H.J. Seifert, H.J. van de Wiel, B. Dodet, I.K. O'Neill (eds.) *Environmental Carcinogens – Methods of Analysis and Exposure Measurement*, Vol. 12, Indoor Air, IARC, 1993.]

to the human health, depending on the contained compounds. Some common toxins present in the home dust are fibers, lead, pesticides, and particles of biological origin.

Fibers may be either natural (hair, wool, cotton, etc.) or MMMFs. Asbestos, if present in home environment, represents a source of concern. The main sources are the insulation products and building materials (e.g. floor tiles, dry wall compounds, reinforced plaster, insulation of pipes, etc.), the use of diverse home appliances (hair dryers, air treatment devices) and the outdoor air. High concentrations of asbestos develop indoor during maintenance operations. Extensively used in the past, asbestos is now severely limited in many countries.

The main sources of lead are paints, pipes and outdoor particulates. Lead paints may impose a risk when the painted surfaces are not properly maintained. During renovation, sanding and scraping of the surfaces may release large amounts of lead dust. Outdoor SPM, especially in the vicinity of heavy traffic areas, contain significant amounts of lead. The quantities are much lower in countries where lead-free gasoline is used. Contaminated soil is carried in from outside on occupant's clothes and shoes and may accumulate in carpets. Re-suspension may subsequently develop elevated concentrations in air. Inorganic lead and compounds are possibly carcinogenic to humans (IARC 1987, group 2B).

Pesticides, such as insecticides, rodenticides, termicides, germicides, fungicides, and herbicides, are widely used in residences or in close proximity. Powder conditioned pesticides (especially termicides) and spraying application may lead to airborne concentrations in the microgram range. The threat imposed by indoor pesticides remain somewhat controversial. Lemus et al.[31] detected concentrations up to $2.13\,\mu m^{-3}$ of chlorpyrifos but could not emphasize clear acute or long time effects. More recently, Davis et al.[32] declared that spraying of chlorpyrifos constituted a great risk for the children, due to accumulation on toys or other surfaces currently touched by children.

Particles of biological origin include bacteria, viruses, fungi, algae, house dust mite particles, dander, and pollen grains. The generic term is bioaerosols. Bioaerosols arise both from outdoor and indoor sources. Indoor sources include air handling system condensate, water damaged materials, high humidity indoor areas, damp organic material and porous wet surfaces, humidifiers, hot water systems, plants, animals and insects, food, and food products. Indoor levels of bacteria, fungi, and viruses may rise to higher values than outdoors, depending on the ventilation system and degree of activity. Acute health effects include allergic reactions (hypersensitivity pneumonitis, humidifier fever, allergic rhinitis, etc.) and infections (such as legionellosis).

3.3 Nonindustrial Workplaces

Sick building syndrome (SBS) is a general term defining a condition associated with complaints of discomfort (headache, nausea, dizziness, eye, nose, throat, and respiratory irritation, etc.). The specific causes are often not known but sometimes are attributed to the conjugated effects of indoor pollution and individual susceptibility. The symptoms associated with periods of occupancy often disappear after the worker leaves the building. By contrast, the term ''building related illness'' (BRI) refers to symptoms of diagnosable illness that can be attributed directly to airborne building contaminants. The causes of SBS and BRI are inadequate ventilation and contamination with chemical compounds and bioaerosols from indoor and outdoor sources.

Traditionally, indoor air in health care facilities was investigated mainly for the presence of pathogen agents. However, especially in the pharmacy areas of the hospitals, aerosols containing hazardous drugs are frequently encountered. Antineoplastic drugs, such as vincristine, dacarbazine, mitomycin, cytosine or arabinoside, impose a serious threat during the preparation of injectable solutions or disposal of the equipment. Exposure to methyl methacrylate, an acrylic cement-like substance used in prosthetics, may occur during preparation and mixing. Aerosol treatment sessions are another exposure circumstance. Drugs administrated as aerosols (such as ribavirin) form liquid particle aerosols of about 1 μm that may be also inhaled by the person who conducts treatment and not only by the patient. The substances mentioned above impose a serious threat due to their acute and chronic effects. Most of them exhibit carcinogenic potential. Dental care facilities represent occupational environments in which chemical aerosols may be generated. Harrel[33] revealed that dentistry equipment, such as ultrasonic scalers, generate aerosols and splatter regardless of the type, the power level or the insert. Barnes[34] showed that these aerosols may contain also blood droplets.

3.4 Transportation

Not much is known about composition and size distribution of aerosols in transportation. In aviation, attention was traditionally paid to flight deck crew, cabin crew, and ground support personnel, and less to cabin air quality. However, Dechow[35] measured the concentrations of several contaminants in the passenger cabin and reported that particulate matter contained tobacco smoke and bacteria. The particulates size was identified in respirable range but the air purification systems maintain low levels.

Svendsen[36] investigated marine engineers that are potentially exposed to aerosols arising from engine

functioning. They reported concentrations of oil mists in range of 0.12 to 0.74 mg m^{-3}.

4 FACTORS INFLUENCING SAMPLING, STORAGE AND ANALYSIS OF AEROSOLS

It is well documented that the quality of the sample submitted to analysis is crucial in obtaining a reliable result. No laboratory analysis, regardless of the quality of the equipment or the skills of the analyst, can overcome the defects of sampling and storage. The determinants that influence the sampling and storage of the aerosols may be roughly divided in factors pertaining to the very nature of the aerosol (intrinsic) and factors related to the sampling strategy.

4.1 Intrinsic Factors

4.1.1 The Momentum of a Particle

Momentum of a particle is the product of its mass and its velocity. The particulates are much larger than the molecules of the gas in which they are dispersed. If the flow direction of the gas changes, larger particles tend to continue their original trajectory. Consequently, larger particles are dispersed to the periphery of the stream and may deposit, thus affecting the original distribution and concentration. The situation is encountered in stack gases and ventilation ducts or if a large object is placed in the main stream. In order to obtain a representative sample, all parts of a gas stream must be sampled.

A greater speed of a gas stream implies a greater momentum for the particles which tend to be lost from the surface of the sampler. Consequently, when collecting a sample this is to be done in such conditions that there is no change in the momentum. The condition is referred to as *isokinetic sampling* and is accomplished by means of a thin-walled probe aligned with the stream flow and drawing sample at the same (linear) velocity as of the stream flow at that point. Even so, a theoretically 100% efficiency of the sampling is not possible because the probe always has a finite wall thickness that disturbs the flow. Laminar flow of the gas stream will allow truly isokinetic sampling whereas turbulent flow will impede the collection of the larger particles. The collection efficiency may also be affected by the deposition of particles on the walls of the probe. Sample velocity of the air and alignment of the probe are both important conditions of the isokinetic sampling. Fortunately, isokinetic sampling is not necessary for all the applications. Sampling in calm ambient air is not influenced by velocity because slow moving particles have a negligible momentum. Also, particles under about 3 μm do not require isokinetic sampling because their small mass minimizes inertial effects.

4.1.2 Gravitational Effects

Gravitational effects impose a small effect upon sampling in the case of relatively quiescent air masses such as indoor environment. A thorough discussion about the influence of both the momentum and the gravitational effect may be found in Katz.[37]

4.1.3 Size

Size influences the sampling of both solid particle and liquid particle aerosols. A larger size implies a greater momentum. In the case of oil mist, Wilsey[38] found that the amount of aerosol collected by means of an Institute of Occupational Medicine (IOM) sampler was greater than that collected with an open face sampler.

4.1.4 Evaporation and Sublimation

Evaporation and sublimation of aerosol samples during storage is a common phenomenon. Particulate samples are less affected because solids generally have low vapor pressure and the evaporative loss is small. By contrast, liquid samples are susceptible to evaporation even during sampling. Leith[39] revealed a significant loss of sample in the case of filter collection when compared with electrostatic precipitation sampling of mineral oil mist. Although the mineral oil used as a machining fluid is not normally considered volatile, when dispersed as mist its surface becomes high. The mist droplets deposited on the surface of the filter are in permanent contact with the air passing through the filter and evaporation takes place. In the case of electrostatic precipitators, droplets are separated from the air flow and coalesce on the precipitator wall to form a film with smaller surface. Differences were emphasized even between different types of filters. Mixed cellulose esters (MCE) retain more oil mist than polyvinyl chloride (PVC) filters.

4.1.5 Electrostatic Charge

Electrostatic charge may contribute to sampling errors. Particles may be lost by electrostatic attraction to various components of the sampler. This is of particular importance in asbestos collection. In area sampling, the conductive cowl of the sampler is connected to the ground (e.g. cold water pipe) in order to eliminate the electrostatic charge.

4.2 Factors Related to Sampling Strategy

4.2.1 Location

Inside large industrial buildings the aerosol concentrations may vary by several orders of magnitude within a relatively short radius. Obstacles (such as machinery or elements of infrastructure) interposed into the air stream cause deposition. If the source is small, as in manual welding, high concentration of contaminants are encountered only in the proximate vicinity. The variations of concentration impose that exposure assessments must be achieved using exclusively data based on personal sampling. Area sampling is used only in general surveys of the workplace.

4.2.2 Temperature

Temperature influences the mass and number of the particles. If the temperature is too high, liquid aerosols may vaporize, and if it is too low, water or other vapors may form mists. Also, at the point of emission, some substances may have a fraction in the gaseous phase which in standard collection conditions (21 °C and 700 torr) may turn into liquid or solid particles. This potential fraction is called condensible particulate matter.

4.2.3 Humidity

In environments with high relative humidity (RH), both particulates and the filters used in their collection may absorb water vapors. Liquid particulates (such as of acid mists) may become larger due to absorption. Water absorption is somehow negligible up to about 50% RH but may increase considerably in environments where the steam is used.

4.2.4 Sampling Rate

Sampling rate influences the collection efficiency of filters. A higher sampling rate means a higher air velocity at the surface of the filter. Also, the size-selective sampling devices (e.g. cyclones and cascade impactors) require a certain flow rate in order to operate properly.

4.2.5 Time Interval

The time interval for a proper collection results from the total volume of air which must be drawn and the flow rate used for collection. However, if the concentration of airborne particles is too low, and the total amount of collected material falls under the detection limit of the method, the duration of sampling must be increased. The increase of the flow rate is not recommended because it affects the collection efficiency.

5 INDUSTRIAL HYGIENE APPLICATIONS

There are two main objectives for the aerosol sampling and analyses in industrial indoor environments: the workplace air quality characterization and the personal exposure assessment. In both cases, the accuracy of the evaluation requires a thorough approach to the type of aerosol, the sampling strategy and the method of analysis. Since the situations encountered in practice differ very much it is the analyst's choice what methodology to apply. Some guidelines and examples follow.

5.1 Sampling Methodology

There are several criteria used in the classification of sampling methodology: the time duration, the size of the particles and the location of the sampling.

According to the time duration, the sampling methods may be classified as:

1. full-period, continuous single sampling; refers to sampling over the entire sampling period with only one sample (e.g. a full-shift sample). This is the preferred method for the collection of aerosols.
2. full-period, consecutive sampling; refers to sampling using consecutive samples of smaller time duration which combined equal the total duration of the sample period. This method is used in case that continuous single sampling leads to filter overloading due to high concentrations of SPM.
3. Grab sampling is defined as the collection of several short-term samples that combined provide an estimate of exposure over the entire period of observation. Grab sampling is applied in aerosol sampling for the assessment of high concentrations that arise during some technological operations with short duration.

According to the size of the particles, the sampling methods are:

1. Sampling for the collection of the entire amount of airborne SPM burden. This operation is encountered in the collection of aerosols for which it is assumed that the hazard potential is not dependent on a certain dimension of the particles.
2. Size-selective sampling implies the collection of particles according to certain dimensional intervals. The sampling uses devices such as stage impactors that are capable of collecting particles of certain dimensions onto separate collection media. Of particular importance is the so-called respirable fraction (see also sections 2 and 7). Substances such as silica have

the hazard potential associated with this mass fraction. The standard device for the respirable fraction collection is the cyclone.

As a function as the location the sampling may be:

1. Area sampling (static) that implies the collection of SPM in a certain fixed location.
2. Personal sampling which means that the samples are collected in the breathing zone of the worker (e.g. within a radius of 30 cm from the nose). This sampling procedure aims to identify the concentrations of the compounds as close as possible to those present in the air that enters RT. The sampler is attached to the worker's collar and connected to a battery operated pump. The sampling is not performed in a particular area but follows the worker's motion.

Depending on the objective of the investigation, the sampling strategy combines two or more sampling methods. For instance, in area sampling either the entire amount of SPM or certain fractions of interest, such as PM 10, PM 2.5, and respirable fraction, may be collected. Also, personal sampling may be conducted for the assessment of the inhalable fraction or the respirable fraction. In either case, sampling may be continuous sampling, consecutive sampling, or grab sampling.

5.2 Identification and Quantitation of Workplace Pollutants

The nature of workplace contaminants is generally predictive because they originate from a well-defined and controlled technological process. However, the concentration of individual compounds vary according to the nature and location of the operations. The SPM burden in large industrial buildings comprise substances generated in different workplaces. Therefore, the first step in establishing the background profile is to screen for the expected contaminants in order to gain data regarding their presence and individual concentrations. Screening may involve the evaluation of the number and size distribution of SPM, the quantity of SPM and the presence of the individual compounds. These goals imply the usage of portable monitors, gravimetric sampling and chemical analysis.

The assessment of individual contaminants in the screening procedure requires methods of analysis capable of identifying a number of compounds simultaneously. These methods may also quantify some compounds with reliable accuracy. In the case of other compounds, screening methods give only qualitative information. Therefore, identification is followed by further sampling and quantitation using specific methods.

5.2.1 Screening and Quantification of Metals in Fumes and Dusts

The evaluation of metallic species in fumes and dusts is a typical example of simultaneous analysis. It implies gravimetric assay and chemical analysis. Characterization of different size fractions may be sometimes required. Metals that can be determined by simultaneous or specific analysis are: silver (Ag) [CAS 7440-22-4], aluminum (Al) [CAS 7429-90-5], arsenic (As) [CAS 7440-38-2], beryllium (Be) [CAS 7440-41-7], calcium (Ca), cadmium (Cd) [CAS 7440-43-9], cobalt (Co) [CAS 7440-48-4], chromium (II) (Cr) [CAS 22541-79-3], chromium (III) (Cr) [CAS 16065-83-1], chromium (VI) (Cr) [CAS 18540-29-9], copper (Cu) [CAS 7440-50-8], iron (Fe) [CAS 1309-37-1], lithium (Li), magnesium (Mg) [CAS 1309-48-4], manganese (Mn) [CAS 7439-96-5], molybdenum (Mo) [CAS 7439-98-7], nickel (Ni) [CAS 7440-02-0], lead (Pb) [CAS 7439-92-1], platinum (Pt) [CAS 7440-06-4], selenium (Se) [CAS 7782-49-2], tellurium (Te) [CAS 13494-80-9], titanium (Ti) [CAS 7440-32-6], thallium (Tl) [CAS 7440-28-0], vanadium (V) [CAS 7440-62-2], yttrium (Y), zinc (Zn) [CAS 1314-13-2] and zirconium (Zr) [CAS 7440-67-7].

The content of metals in dusts and fumes is commonly assessed by means of inductively coupled argon plasma atomic emission spectroscopy (ICPAES), flame atomic absorption spectroscopy (FAAS), and graphite furnace atomic absorption spectroscopy (GFAAS). Table 4 presents a summary of three routine methods used in industrial hygiene.

Sampling is performed by collecting the inhalable fraction of dusts and fumes onto MCE filters. Area sampling and personal sampling may be used. In welding fume sampling for individual assessment, the sampler is placed inside the helmet of the welder. Sample preparation involves hot plate acid digestion. However, this procedure does not destroy all matrices. Microwave digestion is a powerful alternative that uses acid treatment at elevated temperature and pressure.

Detection is performed for each element using specific wavelengths. ICPAES based procedure offers simultaneous analysis of all elements but FAAS technique requires changing the lamp for each metallic specie. In addition, FAAS requires a typical 1–2 ml of the final solution for each element. Diluting the solution in order to screen for more elements may result in falling above the detection limit. These limitations of FAAS affects the versatility of the procedure when used for screening purposes. However, if ICPAES equipment is not available, FAAS represents an alternative because it has detection limits suitable for industrial hygiene applications.

Some elements require specific analysis methods. Table 5 presents a list of routine specific methods for different metallic species. These procedures use either

Table 4 Routine methods for simultaneous analysis of metal content in fumes and dusts

Method number	NIOSH 7300	OSHA ID-125G	P & CAM 173
Technique	ICPAES	ICPAES	FAAS
Matrix	Air	Air, Wipe, Bulk	Air
Elements	Ag, Al[a], As, Be[a], Ca, Cd, Co[a], Cr[a], Cu, Fe, Li[a], Mg, Mn[a], Mo[a], Na, Ni, P, Pb[a], Pt[a], Se, Te, Ti, Tl, V, Y, Zn, Zr[a]	Ag[b], Al[b], As[b], Be, Ca[b], Cd, Co, Cr, Cu, Fe, Mg[b], Mn, Mo, Ni, Pb, Sb, Se[b], Si[b], Sn[b], V, Zn. (Te may replace Si)	Ag, Al[a], Ba, Be[a], Bi, Ca, Cd, Co[a], Cr[a], Cu, Fe, In, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Sr, Tl, V[a], Zn
Limits of detection	Instrument dependent	Instrument dependent	Instrument dependent
Sample preparation	Hotplate acid digestion; 4 : 1, HNO_3 : $HClO_4$	Wet and hotplate acid digestion 1 : 1, H_2SO_4 : deionized water oxidizing agent H_2O_2 30%	Hotplate acid digestion HNO_3 conc.

[a] Some compounds of these elements require special sample treatment. Further specific methods will be used for exact quantitation.
[b] Only semiquantitative analysis. For exact quantitation specific methods are required.

Table 5 Specific routine methods for metal quantitation

Element	Method no.	Sample preparation	Technique	Interference
Al	NIOSH 7013	Hotplate acid digestion	FAAS; nitrous oxide-acetylene, reducing	Possible Fe and V
As	NIOSH 7900	Hotplate acid digestion	FAAS hydrogen-argon	None, background absorption
Be	NIOSH 7102	Hotplate acid digestion	GFAAS	Possible Ca, Na, K, Al
Cr (metal, Cr II, Cr III)	NIOSH 7024	Hotplate acid digestion	FAAS; nitrous oxide-acetylene, reducing	Possible Fe and Ni
Cr VI (hexavalent)	NIOSH 7600	Acid treatment; treatment with diphenylcarbazide solution	VIS; 540 nm wavelength	Possible Fe, Cu, Ni, V
Co	NIOSH 7027	Aqua regia, hotplate digestion	FAAS; air-acetylene, oxidizing	None, background correction
Cu (dust and fumes)	NIOSH 7029	Hotplate acid digestion	FAAS	None
Pb	NIOSH 7105	Hotplate acid digestion; alternative, microwave digestion	GFAAS	Possible Ca
V (oxides, dust and fumes)	NIOSH 7504	Ultrasonication in tetrahydrofuran; redeposit on Ag filter	XRD	α-Quartz

VIS, visible spectrophotometry.

enhanced sample procedure steps or different analytical techniques.

There are some alternative methods available. Hexavalent chromium may be analyzed with NIOSH 7604 method based on IC conductivity detection; cations of metals interfering with the colorimetric method, do not interfere with this method.

NIOSH 7082 is a laboratory method based on FAAS technique for the identification of lead and lead compounds. It has a higher working range than NIOSH 7105 but lower detection limit. Lead sulfide, a compound with particular toxicological interest, is evaluated in the respirable fraction (NIOSH 7505) using XRD.

In field conditions, lead may be detected using chemical spot test (rhodizonate-based)-NIOSH 7700. This is a qualitative colorimetric method for identification of lead in the particulates sampled on filters. On-site filter analysis may be performed by means of portable anodic stripping voltammeters (e.g. method NIOSH 7701) but samples must be first ultrasonicated in HNO_3. Method NIOSH 7702 avoids sample preparation in field conditions because it uses portable instruments based on X-ray fluorescence (XRF), L-shell excitation technique (cadmium-109 source).

Arsenic compounds may be analyzed by GFAAS. The method NIOSH 7901 is a procedure for the determination of As_2O_3 [CAS 1327-53-3] and NIOSH 6001 for the analysis of AsH_3 [CAS 7784-42-1].

5.2.2 Sampling and Analysis of Polycyclic Aromatic Hydrocarbons

In practice, the PAH mixtures are too complex and not all individual compounds may be assessed. Therefore, PAHs are evaluated (a) as a total amount of compounds (i.e. not distinguishing between the individual compounds) and (b) using the quantitation of a number of compounds

selected as markers of the presence of PAHs in the mixture.

Some problems are usually encountered in PAH analysis. One drawback is the complexity of the incorporating matrices. The presence of interfering substances that have physical and chemical behavior similar to that of PAHs require complex sample preparation by repeated extraction and purification steps. On the other hand, PAHs are present as complex mixtures containing isomeric structures and derivatives that vary greatly in relative concentration of the individual compounds. Therefore, the analysis focuses only on certain markers.

The analysis of PAHs requires the use of methods with high resolution, response selectivity and constancy, and acceptable cost/performance ratios. It is often impossible to achieve some of the goals simultaneously and the analyst should apply the technique that offers the best result in the given application.

Sampling of PAHs usually requires the collection of large amounts of SPM. The smaller the amount collected, the greater the sensitivity required in identification. Particle-bound PAHs are collected onto glass fiber filters. For the collection of both gaseous and particle-bound PAHs the sampling train comprises filter plus sorbent tube.

Sample preparation implies extraction with solvents such as cyclohexane, acetone, benzene, toluene, and methylene chloride. Furton et al.[40] underlined that supercritical fluid extraction (SFE) is a powerful alternative to conventional liquid extraction. SFE uses mainly carbon dioxide which is a nonpolar compound and limits the recovery of higher molecular PAHs.

The extracts may be further cleaned up by adsorption and elution on silica gel, Florisil or alumina columns or by means of solid phase extraction (SPE). SPE uses cartridges filled with various materials (e.g. octadecylsilane C_{18} – ammonia (NH_2), silica–cyano) and require much smaller amounts of solvents than purification on columns.

Separation and detection: Reversed-phase high-performance liquid chromatography (RPHPLC) using octadecylsilane (C_{18}) columns is one of the most popular methods for the separation of PAHs. However, not all C_{18} stationary phases provide the selective separation for all PAHs. A quantitative measure of phase selectivity is given by the selectivity factor in Equation (1):

$$\alpha_{\mathrm{TBN/BaP}} = \frac{k'_{\mathrm{TBN}}}{k'_{\mathrm{BaP}}} \qquad (1)$$

The k'_{TBN} represents the retention factor of 1,2:3,4:5,6:7,8-tetrabenzonaphthalene (alternate name dibenzo[*g*,*p*]chrysene) and k'_{BaP} the retention factor of benzo[*a*]pyrene. The selectivity factor $\alpha_{\mathrm{TBN/BaP}}$ correlates with the retention behavior of PAHs and the type of the bonded phase. A value of $\alpha_{\mathrm{TBN/BaP}} \leq 1$ indicates polymeric C_{18} phases while $\alpha_{\mathrm{TBN/BaP}} \geq 1.7$ indicates monomeric C_{18} phases. For values of $1 < \alpha_{\mathrm{TBN/BaP}} < 1.7$ the bonded phase synthesis is somehow less certain and may indicate a densely loaded monomeric phase or light polymerization with di- or trifunctional reagents (Wise et al.[41]). The separation of EPA 16 priority pollutant PAHs can be properly achieved only with columns which have $\alpha_{\mathrm{TBN/BaP}}$ between ca. 0.6 and 0.9. Widely used columns are Supelcosil LC-PAH, Vydac 201 TP, Spherisorb PAH, Erbasil C_{18} H.

Detection of PAHs is commonly based on ultraviolet (UV) absorbance measurements. The usage of diode-array detectors allow the generation of UV spectra. Further comparison with library spectra leads to enhanced identification. Another common procedure is to use a fluorescence detector in tandem with the UV detector. The identification relies on the fact that many PAHs are good fluorophores. However, some compounds such as acenaphthylene, benzo[*ghi*]perylene, and indeno[1,2,3-*cd*]pyrene exhibit low or no fluorescence.

The 254 nm wavelength is used for simple UV scan. For the enhanced identification of a particular compound there are three alternative wavelengths:

(a) 236 nm for fluoranthene, and pyrene;
(b) 254 nm for naphthalene, phenanthrene, anthracene, benz[*a*]anthracene, and chrysene;
(c) 299 nm for benzo[*b*]fluoranthene, benzo[*k*]fluoranthene, benzo[*a*]pyrene, dibenz[*a*,*h*]anthracene, benzo[*ghi*]perylene, and indeno[1,2,3-*cd*]pyrene.

The general analytical steps presented above vary according to the application's requirements. Four routine methods OSHA 58, NIOSH 5506, NIOSH 5800 and Method 102A & D[42] will be briefly presented as an example. Table 6 presents the list of compounds that may be identified by means of these methods.

NIOSH 5506 is a method directed towards personal sampling. It allows sampling of both gaseous and particle-bound PAHs. The absence of a cleanup step reduces the loss of PAHs but may impose a negative influence upon chromatographic separation. The characteristic peak of a compound may overlap the peak of the neighboring compound. A thorough adjustment of the solvent gradient may be necessary for good separation. The application area includes evaluation of fumes generated in the aluminum industry, coal liquefaction plants, coal gasification plants, creosote treatment facilities, graphite electrode manufacturing, and also the analysis of asphalt fume, coal tar pitch and coke oven emissions, diesel exhaust, petroleum pitch, and fumes that are generated during roofing.

Table 6 PAH compounds currently analyzed

Compound	CAS numbers	OSHA 58	NIOSH 5506	Method 102A & D
Acenaphthene	CAS 83-32-9	no	yes	yes
Acenaphthylene	CAS 208-96-8	no	yes	yes
Anthracene	CAS 120-12-7	yes	yes	yes
Benz[*a*]anthracene	CAS 56-55-3	no	yes	yes
Benzo[*b*]fluoranthene	CAS 205-99-2	no	yes	yes
Benzo[*k*]fluoranthene	CAS 207-08-9	no	yes	yes
Benzo[*ghi*]perylene	CAS 191-24-2	no	yes	yes
Benzo[*a*]pyrene	CAS 50-32-8	yes	yes	yes
Benzo[*e*]pyrene	CAS 192-97-2	no	yes	no
Chrysene	CAS 218-01-9	yes	yes	yes
Dibenz[*a*,*h*]anthracene	CAS 53-70-3	no	yes	yes
Fluoranthene	CAS 206-44-0	no	yes	yes
Fluorene	CAS 86-73-7	no	yes	yes
Indeno[1,2,3-*cd*]pyrene	CAS 193-39-5	no	yes	yes
Naphthalene	CAS 91-20-3	no	yes	yes
Phenanthrene	CAS 85-01-8	yes	yes	yes
Pyrene	CAS 129-00-0	yes	yes	yes
Total (gravimetric)		yes	no	no

Table 7 Comparative parameters in some routine methods for PAHs analysis

Analytical step	OSHA 58 (HPLC)	NIOSH 5506 (HPLC)	NIOSH 5800 (fluorescence)	Method 102A & D (HPLC)
Sampling	Filter	Filter + sorbent tube	Filter + sorbent tube	Filter
Extraction	Benzene, shake	Ultrasonication or Soxhlet. Solvents: benzene, cyclohexane, methylene chloride[a]	Hexane, rotated for at least 12 hours	Soxhlet. Solvent: cyclohexane
Cleanup	–	–	SPE (cyano)	Silicagel column; eluent, pentane–methylene chloride
Separation column; conditions	DuPont Zorbax ODS; gradient	Vydac 201TP or equivalent; gradient	–	Vydac 201TP or equivalent; gradient
Mobile phase	Acetonitrile–water	Acetonitrile–water	Acetonitrile	Acetonitrile–water
Detection	UV/fluorescence	UV @ 254 nm; Fluorescence: excitation @ 240 nm, emission @ 425 nm	Fluorescence. Detector 1: excitation @ 254 nm, emission @ 370 nm. Detector 2: excitation @ 254 nm, emission @ 400 nm	UV @ 254 nm. Fluorescence (optional): excitation @ 240 nm, emission @ 425 nm

[a] The solvent choice is based on comparative trials.

NIOSH 5800 is not based on high-performance liquid chromatography (HPLC); it uses flow injection into a fluorescence detector and identifies the total polycyclic aromatic compounds (PACs) fraction in aerosols (defined as all compounds producing detectable fluorescence at the chosen wavelengths). This method is used for the comparison of environments using the same fume source. The total PACs burden is assessed by comparison with standard solutions of PACs. Therefore, only PACs present in the standards will be identified.

OSHA 58 represent in fact two methods in one: (a) the quantitation of total PAHs by gravimetric assay of the benzene extractable fraction, and (b) the identification of some individual compounds. The gravimetric measurement is very useful since regulation of PAHs in many countries rely on it. The identification of individual compounds is somehow limited. Also, the method is limited to the particle-bound PAHs.

Method 102A & D is tedious but very flexible. It may be used for area sampling, personal sampling and size selective sampling. The extraction solvent (cyclohexane) is selective for nonsubstituted PAHs but Soxhlet extraction assumes long periods (a minimum 8 hours is required). Cleanup on silicagel is versatile and ensures better conditions for the separation and identification even when simple UV detection is involved. However, this step may induce sample loss and is time-consuming. This method is suitable for both industrial and

nonindustrial applications and is dedicated to particle-bound PAHs.

Table 7 presents a summary of these four methods.

HPLC has limitations in terms of detection limits. Increasing the sampled volume might not represent always the best solution (see section 8). Consequently, more sensitive techniques are sometimes required. A common alternative is gas chromatography/mass spectrometry (GC/MS). The use of capillary columns provides good separation for PAHs and the mass spectrometry detectors are suitable for identification of compounds in mixtures. Another alternative is to upgrade the HPLC analysis by using it in conjunction with a more powerful method of identification such as Shpol'skii spectroscopy (SS). SS is especially suited for the qualitative analysis of PAHs at trace levels. An industrial hygiene application is reported by Ariese.[43]

5.3 Assessment of Human Exposure

The assessment of human exposure to aerosols implies measurements of the airborne concentrations of those toxins to which a person comes into direct contact during a certain time interval. The measurements rely exclusively on personal sampling. The samples are collected as inhalable dust or as respirable fraction according to the toxicological behavior of the compound that will be analyzed.

The assessment of human exposure to silica dust requires the identification and quantitation of the crystalline silica in the respirable fraction (see also section 3.1). Routine analyses rely on three techniques: XRD, IR and VIS. Some common methods are: NIOSH 7500, NIOSH 7501, NIOSH 7601, NIOSH 7602, NIOSH 7603, and OSHA ID-142. A brief characterization is presented in Table 8.

5.3.1 Sampling

Sampling implies personal sampling using cyclones. The flow rate is adjusted in order to achieve a median cut-off point of 4 μm (see also section 7). The samples are collected on filters that must be weighted before and after the sampling. Sample weights of 0.5 to 3.0 mg are preferred. Some substances such as aluminum phosphate, feldspars, graphite, iron carbide, lead sulfate, micas, montmorillonite, silver chloride, and talc may interfere during the analysis. Therefore, the field operations must include the acquisition of information regarding the possible presence of interfering substances and also bulk samples.

Sample preparation is specific to the analytical method. General procedures for XRD and IR techniques involve the destruction of the filter used in collection by one of the following techiques: muffle furnace, low-temperature (radio-frequency) plasma ashing or filter dissolution in organic solvents (e.g. PVC filters in tetrahydrofuran). Additional steps of ultrasonication or acid treatment are included in order to break up the agglomerates and to overcome the influence of interfering substances. For XRD analysis, the recovered material is deposited onto a silver membrane filter. In case of the IR analysis, the ashen material is usually incorporated into KBr pellets, as in method NIOSH 7602. An alternative is that of scanning directly on thin layer deposits (method NIOSH 7603). In this case, the ashed material is suspended in 2-propanol and then redeposited by filtration on a 47-mm diameter, 0.45-μm pore size, vinyl chloride–acrylonitrile copolymer membrane filter.

The VIS technique involves hot plate acid digestion of the collected filters. The noncrystalline fraction is then treated with concentrated H_3PO_4 (85%) and the phosphates are dissolved in water and filtered. The crystalline silica deposited on filters is treated with HF 48% and the acid excess is neutralized with boric acid solution. The first step in the color development is the treatment with molybdate reagent prepared from ammonium molybdate tetrahydrate [CAS 12054-85-2], in deionized water and concentrated sulfuric acid. The silicomolybdate complex is yellowish. If color does not develop, a second treatment with 1-amino-2-naphthol-4-sulfonic acid is performed in order to obtain the molybdenum blue complex (blue color). The silicomolybdate complex will cover a range of 0.1 to 2.5 mg of SiO_2 and the molybdenum blue complex a range of 0.02 to 0.15 mg of SiO_2 in the sample.

Table 8 Routine methods for the analysis of crystalline silica

Method	Technique	Polymorphs	Working range
NIOSH 7500	XRD	Quartz, cristobalite, tridimite	0.025 to 2.5 mg m^{-3} for an 800 l air sample
OSHA ID-142	XRD	Quartz, cristobalite	Validation for quartz, 50 to 160 μg per sample; 816 l sample recommended volume
NIOSH 7602	Spectrophotometry, IR	No, quartz	0.025 to 0.4 mg m^{-3} for a 400 l air sample
NIOSH 7603	Spectrophotometry, IR, specific to quartz in coal mine dust	No, quartz	0.03 to 2 mg m^{-3} for a 1000 l sample
NIOSH 7601	Spectrophotometry, VIS	No, all polymorphs evaluated as total	0.04 to 5 mg m^{-3} for a 500 l air sample

5.3.2 Analysis

Equation 2 gives the Bragg relation on which the X-ray technique is based:

$$n\lambda = 2\,d\sin\theta \qquad (2)$$

in which:

$n =$ order of diffracted beam
$\lambda =$ wavelength of X-ray beam (in angstroms)
$d =$ distance between diffracting planes (in angstroms)
$\theta =$ angle between incident X-rays and the diffracting planes (in degrees); 2θ for most spectrometers

Quartz, cristobalite and tridimite can be identified and quantified on three or more different angles of diffraction. Informational values are listed in Table 9. Values correspond to 2θ degrees and should be regarded rather as informative than absolute since peak locations depend also on the instrument.

In IR, scanning is performed usually from $1000\,cm^{-1}$ to $600\,cm^{-1}$. The peak of interest for quartz is located at $800\,cm^{-1}$. If the samples were ashed at low temperature and kaolinite is present, a correction must be applied. Five different concentrations of purely kaolinite samples are scanned at $800\,cm^{-1}$ and at $915\,cm^{-1}$ (the maximum absorption bands for kaolinite). Then a curve is prepared and used for correction of the absorbance value at $800\,cm^{-1}$.

VIS involves a classic color measurement. The silicomolybdate is measured at 420 nm. The molybdenum blue complex is measured at 820 nm.

The methods based on XRD and IR are highly specific and require relatively simple sample preparation. The VIS method is rather tedious but offers a fairly good precision and require lower skilled personnel. A comparison between the performances of these techniques applied in different laboratories over a long period of time may be found in Shulman.[44]

As stated earlier in this article, surface analysis is required for complete investigation of toxicological properties of silica. Low-voltage scanning electron microscopy may be employed for the detection of clay occlusion of quartz particles. The surface composition may be determined by laser microprobe mass analysis. This technique uses a laser beam to vaporize a small volume of material from the surface of a single quartz particle and the ions generated are identified with a time-of-flight mass spectrometer. Detailed surface analysis may be also performed by means of scanning electron microscopy with energy dispersion X-ray analysis or X-ray photoelectron spectroscopy.

Table 9 Diffraction angles for silica identification by XRD

Mineral	Primary	Secondary	Tertiary
Quartz	26.66	20.85	50.16
Cristobalite	21.93	36.11	31.46
Tridymite	21.62	20.50	23.28
Silver	38.12	44.28	77.47

6 NONINDUSTRIAL APPLICATIONS

6.1 Objectives and Procedures

Common objectives for the analysis of indoor SPM are the investigation of the IAQ profile, the investigation of the sources, and the assessment of exposure. Wilson[45] emphasized that the comparison with outdoor concentration (the calculation of the indoor/outdoor SPM ratio) is a relevant parameter for epidemiological studies. It may indicate the prevalent source of a particular contaminant. The operations involved in the assessment of indoor SPM comprise chemical and biological analysis, counting, and sizing.

The measurements of indoor SPM are conducted as to assess the average concentrations as well as the maximum concentrations. Some directions for maximum concentration measurements are:

1. ventilation must be diminished to facilitate accumulation (e.g. doors and windows shut or air conditioning systems kept to a minimum recirculation rate);
2. the samples must be taken at 1–2 m height in the area closest to suspected source and during the maximum source activity;
3. the occupants must carry out full normal activity including those that are supposed to create resuspension of the settled particulate matter;
4. samples must be taken both in winter and in summer;
5. the duration of sampling depends mostly on activity of the source of SPM contamination; direct instruments must be used if the emissions are short timed so that the sampling would not provide enough material for the analysis;
6. the number of samples is roughly double in air conditioning environments compared to natural ventilated buildings.

An analysis of sampling strategies for indoor sampling is given by van de Wiel.[46]

The SPM is collected on filters. The investigation often require the usage of size-selective sampling devices. Chemical analysis employs techniques usually derived from industrial applications.

Heating, ventilation and air conditioning (HVAC) systems that are used in many buildings, both in office and residential environments, are subjected to verifications during IAQ assessment. One characteristic of interest for these devices is the filtration efficiency, which represents their ability to clean the air by retaining the airborne SPM. The size and number of particles are determined before and after passing the HVAC. Then the filtration efficiency is expressed as the percent of particles of a certain size fraction captured by the filtration system (e.g. HVAC retains particles of $\leq 1\,\mu m$ with 98% filtration efficiency). Counting and sizing for the investigation of HVACs is performed by direct reading instruments based usually on light scattering. The laser diffraction particle counters represent an alternative because of their enhanced precision and lower detection limits. Detection limit refers to the smallest particle that can be detected and counted.

In the case of ventilation duct measurements (e.g. inlets for fresh air supply) the sensors are equipped with isokinetic probes (see also section 4). It is also advisable to use dual particle counters in order to minimize the errors that occur in field particle counting.

Detailed investigations of HVAC systems may include biological contamination because humidifiers are often a source of microorganisms.

6.2 Analysis of Bioaerosols

Indoor biological pollutants often constitute a cause of concern in the indoor nonindustrial environment. The assessment of bioaerosols is a complex operation which requires knowledge of many specialists as industrial hygienists, biologists, chemists, medical professionals, epidemiologists, statisticians, ventilation engineers, etc.

Sampling of aerosols includes almost every principle applied for separating particles from air, from sedimentation, filtration, impingement to electrical and thermal precipitation. Use of open plates of solid nutrient media is inefficient. Cascade impactors and all-glass impingers proved to have collection efficiency good enough to be considered as standard sampling instruments. The collection efficiency of a bioarerosol sampler refers to: (a) the physical collection efficiency, which implies the capacity of the sampler to separate the biological particles from the air; and (b) the biological collection efficiency, which implies the capacity of the sampler to preserve the integrity and biological activity of the microorganisms during sampling.

A general procedure for sampling culturable microorganisms such as bacteria, fungi, and thermophilic actinomycetes is the method NIOSH 0800. The sampling involves drawing the air through an Andersen cascade impactor fitted with culture media plates. Culture media include agar derivatives such as malt extract agar and trypticase soy agar, but other media may be used according to the microorganisms intended for identification.

Once the particles have been collected they can be analyzed by one of the following techniques:

1. Culturing of viable cells is one of the most frequently used methods. The samples are inoculated onto any type of culture media and incubated under whatever conditions of temperature, humidity, and lighting as are necessary to grow the collected microorganisms. The microorganisms are then identified by their distinctive individual colonies. The culture of sample isolates is suitable for the identification of infectious agents as Legionella, the measles virus (rubeola), etc.
2. Direct examination of particles is useful in identification of allergic agents as pollen grains and spores of the fungi. The samples are stained and then examined under a microscope. The staining methods include either direct use of Gram stain or the immunoassay technique using fluorescent antibodies. In the latter case, a specific antibody that is conjugated with a fluorescent dye combines with the microbial agent located onto the surface of the cell forming a complex. The fluorescent complex may be observed by means of a fluorescent microscope.
3. Instrumental analysis employs the use of an analytical technique such as chromatography. NIOSH 0801 method provides a procedure for the identification of airborne bacteria by means of gas chromatography/flame ionization detector (GC/FID). The samples are cultured onto nutrient media and than harvested. The fatty acids C_9-C_{20} are first separated from the harvested material by saponification and further methylated to obtain fatty acid methyl esters (FAME). The material is subjected to GC/FID analysis in order to obtain a FAME profile characteristic to each bacteria. Identification is based on comparison with a profile library of FAME.

In the field, identification of biological particulates may be performed by means of direct reading instruments. These portable devices use the fluorescent properties exhibited by biological particulates when exposed to UV radiation.

7 REGULATIONS AND AIR STANDARDS

In the working environment, the control of airborne toxins relies on the use of the threshold limit values (TLVs). TLVs are numbers that refer to airborne concentrations of substances and represent conditions

under which it is believed that workers may be repeatedly exposed without adverse health effects. However, due to variations in individual susceptibility, some workers may experience discomfort or even serious disturbance at concentrations at or below TLV. Therefore, TLVs do not represent perfectly safe limits but only recommendations and guidance for practice. There are three ways of expressing TLVs:

1. Threshold limit value/time weighted average (TLV/TWA, or simply TWA); ACGIH[7] defines TWA as being "the time-weighted average concentration for a conventional 8-hour workday and a 40-hour workweek, to which it is believed that nearly all workers may be repeatedly exposed, day after day, without adverse effect";
2. Threshold limit value/short term exposure limit (TLV/STEL, or simply STEL), defined as a 15-minute TWA exposure not to be exceeded during the workday even if TWA is in limits; STEL supplements TWA in cases where there are acute effects from a substance whose toxic effects are primarily of chronic nature;
3. Threshold limit value/ceiling (TLV/C) refers to the concentration that should not be exceeded during any part of the working exposure.

Exposure to particulate matter is normally expressed as TWA in terms of mg m^{-3}. However, substances that may exert immediate effects have also STELs and TLV/C (e.g. acid mists containing sulfuric acid, sodium hydroxide, etc.).

Some substances have TLVs expressed size-selectively. The TLV for crystalline silica is expressed as the mass of the particles which belong to the fraction of respirable size. The TLV for asbestos is defined as the number of fibers belonging to a certain dimensional interval which is also considered a respirable fraction. Many other chemicals are suspected of having a potential hazard associated with the mass concentration of a certain size fraction. Consequently, more compounds are expected to have mass size-selective TLVs in the near future. The mass fraction for a dimensional interval (e.g. the mass of the respirable particles) refer to those particles that may be captured by a sampler in certain boundary conditions, as follows:

1. Inhalable particulates mass comprises particles that may be captured with 50% collection efficiency for a corresponding dimension of 100 μm; for particles with dimensions below 100 μm the collection efficiency increases with the decrease of the d_{ae}; in the case of particles greater than 100 μm, the collection efficiency decreases with the increase of d_{ae}; inhalable particulate mass replaces the old term "Total" particulate mass that referred the total amount of airborne particulates;
2. Thoracic particulates mass, comprises particles that may be captured with 50% collection efficiency for a corresponding dimension of 10 μm; for particles with dimensions below 10 μm the collection efficiency increases with the decrease of the d_{ae}; in the case of particles greater than 10 μm, the collection efficiency decreases with the increase of d_{ae};
3. Respirable particulates mass comprises particles that may be captured with 50% collection efficiency for a corresponding dimension of 4 μm; for particles with dimensions below 4 μm the collection efficiency increases with the decrease of the d_{ae}; in the case of particles greater than 4 μm, the collection efficiency decreases with the increase of d_{ae}; Equations followed by examples that illustrate these descriptions in quantitative terms may be found in ACGIH.[7]

The explanations presented above rely on international conventions provided by the International Organization for Standardization/European Standardization Committee protocol. However, not all particulate measurements are made in accordance with these boundary conditions for samplers. In some countries, the median cut point for respirable particulate matter samplers is of 5 μm. Also, some older measurements performed in USA implied the use of samplers with a median cut point of 3.5 μm. Therefore, when comparing concentration measurements performed in different areas or at different intervals of time, it is recommended that additional data are obtained regarding the conditions in which the sampling was performed.

All exposure limits refer to personal exposure. Consequently, the analysis of toxins for compliance with TLVs must rely on personal sampling.

8 COMPARISON BETWEEN INDUSTRIAL AND NONINDUSTRIAL INDOOR ENVIRONMENTS

The size, shape, and chemical composition of particulates in the industrial environment strongly relate to the technological process involved. If there is no change of the operational parameters, the pollutants follow a reproducible pattern of concentration and distribution. The quality and quantity of home emissions depend on the inhabitants' lifestyle and the general economic development. Therefore, it is difficult to establish a concentration profile.

In the industrial environment it is relatively easy to identify the sources that emit particulate matter.

By contrast, in the home and office environment the sources of emissions are small and diffuse. In general, the airborne concentrations of SPM in the home and office environments are lower than in the industrial workplaces.

For most toxins, the methods of chemical analysis for aerosols present in the two types of environments share the same techniques. Because the nonindustrial environment has lower levels of concentrations, the common way to adapt a method from industrial hygiene is to sample larger volumes of air. This is achieved either by increasing the sampling rate or the sampling period. The increase of the sampling rate by means of high volume samplers may be inadequate in some circumstances because they disturb the occupants of the building. On the other hand, longer sampling periods may result in loss of information regarding the fluctuating high concentrations. Some operations (e.g. cooking) release relatively large quantities of particles which may be underestimated because of the limited period of time in which they may be sampled. For such circumstances, it is necessary to use more sensitive, and thus more expensive, analytical techniques.

As stated before, the control of airborne particulates in the industrial environment rely on personal exposure. In nonindustrial environments it is difficult to perform personal exposure assessments. The house inhabitants are generally resistant to the requirement of carrying uncomfortable sampling equipment. The requisite for more expensive analytical methods may represent another drawback. Consequently, many exposure assessments in indoor environment use area sampling for the purpose of concentration measurements.

ABBREVIATIONS AND ACRONYMS

ACGIH	American Conference of Governmental Industrial Hygienists
BRI	Building Related Illness
ETS	Environmental Tobacco Smoke
FAAS	Flame Atomic Absorption Spectroscopy
FAME	Fatty Acid Methyl Esters
GC/FID	Gas Chromatography/Flame Ionization Detector
GC/MS	Gas Chromatography/Mass Spectrometry
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
HPLC	High-performance Liquid Chromatography
HVAC	Heating, Ventilation and Air Conditioning
IAQ	Indoor Air Quality
IARC	International Agency for Research on Cancer
IC	Ion Chromatography
ICPAES	Inductively Coupled Argon Plasma Atomic Emission Spectroscopy
IOM	Institute of Occupational Medicine
IR	Infrared Spectroscopy
MAG	Metal Active Gas
MCE	Mixed Cellulose Esters
MIG	Metal Inert Gas
MMA	Manual Metal Arc Welding
MMMF	Man Made Mineral Fiber
MMMV	Man Made Vitreous Fiber
NIOSH	National Institute of Occupational Safety and Health (USA)
OSHA	Occupational Safety and Health Administration
PACs	Polycyclic Aromatic Compounds
PAH	Polycyclic Aromatic Hydrocarbon
PCOM	Phase Contrast Optical Microscopy
PVC	Polyvinyl Chloride
RH	Relative Humidity
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RT	Respiratory Tract
SBS	Sick Building Syndrome
SEM	Scanning Electronic Microscopy
SFE	Supercritical Fluid Extraction
SPE	Solid Phase Extraction
SPM	Suspended Particulate Matter
SS	Shpol'skii Spectroscopy
TEM	Transmission Electronic Microscopy
TIG	Tungsten Inert Gas
TLV	Threshold Limit Value
TLV/C	Threshold Limit Value/Ceiling
TLV/STEL	Threshold Limit Value/Short Term Exposure Limit
TLV/TWA	Threshold Limit Value/Time Weighted Average
TSP	Total Suspended Particulate
UV	Ultraviolet
VIS	Visible Spectrophotometry
WHO	World Health Organisation
XRD	X-ray Diffraction
XRF	X-ray Fluorescence

RELATED ARTICLES

Industrial Hygiene **(Volume 5)**
Carcinogens, Monitoring of Indoor Air • Chromatographic Techniques in Industrial Hygiene • Direct Reading Instruments for the Determination of Aerosols

and Particulates • Dust, Measurement of Trace Elements in

Industrial Hygiene cont'd **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

REFERENCES

1. M. Lippmann, D.B. Yeates, R.E. Albert, 'Deposition, Retention and Clearance of Inhaled Particles', *Br. J. Industr. Med.*, **37**, 337–362 (1980).
2. International Commission for Radiological Protection, 'Human Respiratory Tract Model for Radiological Protection', ICRP publication 66, Elsevier, UK, 1994.
3. R.J. Rando, 'Occupational Exposure to Respiratory Toxins', in *Occupational Lung Disorders*, ed. C.E. Mapp, European Respiratory Society Journals Ltd., Sheffield, UK, 1999.
4. J.L. Freijer, F.R. Cassee, L. van Bree, 'Modeling of Particulate Matter Deposition in the Human Airways', National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands, 1997.
5. W. Sloff, P.J. Blokzijl, 'Integrated Criteria Document Asbestos', National Institute of Public Health and Environmental Protection (RIVM), Bilthoven, The Netherlands, 1989.
6. O.M.S. Publications régionales, 'Glossaire de la Pollution de l'air' (in French), Série européenne, No. 9, Bureau régional de l"Europe, Copenhague, 1987.
7. ACGIH 'Documentation of the Threshold Limit Values and Biological Exposure Indices', ACGIH, Cincinnati, USA, 1999.
8. Deutsche Forschungsgemeinschaft 'MAK -und BAT – Werte-Liste 1998' (in German), Senatkomission zur Prüfung gesundheitsschädlicher Arbeitsstoffe, Mitteilung 34. Wiley-VCH, 1998.
9. WHO 'Health Hazard and Biological Effects of Welding Fumes and Gases, Report on an International Conference', WHO Regional Office for Europe, Copenhagen, 1986.
10. O. Ozdemir, N. Numanoglu, U. Gönüllü, I. Savas, D. Alper, H. Gürses, 'Chronic Effects of Welding Exposure on Pulmonary Function Tests and Respiratory Symptoms', *Occup. Environ. Med.*, **52**(12), 800–3 (1995).
11. J.M. Antonini, G.G. Krishna Murthy, R.A. Rogers, R. Albert, G.D. Ulrich, J.D. Brain, 'Pneumotoxicity and Pulmonary Clearance of Different Welding Fumes After Intratracheal Instillation in the Rat', *Toxicol. Appl. Pharmacol.*, **140**(1), 188–99 (1996).
12. J.M. Antonini, G.G. Krishna Murthy, J.D. Brain, 'Responses to Welding Fumes: Lung Injury, Inflammation, and the Release of Tumor Necrosis Factor-alpha and Interleukin-1 Beta', *Exp. Lung. Res.*, **23**(3), 205–27 (1997).
13. J.M. Antonini, R.W. Clarke, G.G. Krishna Murthy, P. Sreekanthan, N. Jenkins, T.W. Eagar, J.D. Brain, 'Freshly Generated Stainless Steel Welding Fume Induces Greater Lung Inflammation in Rats as Compared to Aged Fume'. *Toxicol. Lett.*, **98**(1–2), 77–86 (1998).
14. P. Gustavsson, R. Jakobsson, H. Johansson, F. Lewin, S. Norell, L.E. Rutkvist, 'Occupational Exposures and Squamous Cell Carcinoma of the Oral Cavity, Pharynx, Larynx, and Oesophagus: A Case-control Study in Sweden'. *Occup. Environ. Med.*, **55**(6), 393–400 (1998).
15. T.E. Danielsen, S. Langard, A. Andersen, 'Incidence of Lung Cancer Among Shipyard Welders Investigated for Siderosis', *Int. J. Occup. Environ. Health.*, **4**(2), 85–8 (1998).
16. P. Hewett, 'Estimation of Regional Pulmonary Deposition and Exposure for Fumes from SMAW and GMAW Mild and Stainless Steel Consumables', *Am. Ind. Hyg. Assoc. J.*, **56**(2), 136–42 (1995).
17. P. Hewett, 'The Particle Size Distribution, Density, and Specific Surface Area of Welding Fumes from SMAW and GMAW Mild and Stainless Steel Consumables', *Am. Ind. Hyg. Assoc. J.*, **56**(2), 128–35 (1995).
18. IARC 'Monographs on The Evaluation of Carcinogenic Risk to Humans, Occupational Exposure to Mists and Vapors from Strong Inorganic Acids; and Other Industrial Chemicals', Volume 54, Lyon, 1992.
19. K. Steenland, 'Laryngeal Cancer Incidence Among Workers Exposed to Acid Mists (United States)', *Cancer Causes Control*, **8**(1), 34–8 (1997).
20. T. Sorahan, D.C. Burges, L. Hamilton, J.M. Harrington, 'Lung Cancer Mortality in Nickel/Chromium Platers, 1946–95', *Occup. Environ. Med.*, **55**(4), 236–42 (1998).
21. S. Kira, Y. Nogami, K. Taketa, M. Ogata, 'Size-selective Sampling of Oil Mist in Air and Subjective Symptoms Among Machine Workers', *Ind. Health.*, **35**(3), 394–8 (1997).
22. IARC 'Monographs on the Evaluation of Carcinogenic Risks to Humans, Silica, Some Silicates, Coal Dust and Para-aramid Fibrils, Vol. 68', Lyon, France, 1997.
23. G.K. Montizaan, C.A. van der Heijen, 'Integrated Criteria Document Asbestos Effects', Report No. 758473013, National Institute of Public Health and Environmental Protection (RIVM), Bilthoven, The Netherlands, (2) 1989.
24. WHO 'Reference Method for Measuring Airborne Man-Made Mineral Fibers', World Health Organization, Regional Office for Europe, Copenhagen, 1992.
25. I.P.C.S. Environmental Health Criteria 151 'Selected Synthetic Organic Fibers', WHO, Geneva, 1993.
26. K.W. Gold, 'Indoor Concentration of Environmental Carcinogens', in *Environmental Carcinogens – Methods of Analysis and Exposure Measurement*, eds. H.J. Seifert, H.J. van de Wiel, B. Dodet, I.K. O'Neill, Volume 12, 'Indoor Air', IARC, Lyon, France, 41–72, 1993.
27. T.A. Chiang, P.F. Wu, Y.C. Ko, 'Prevention of Exposure to Mutagenic Fumes Produced by Hot Cooking Oil in

Taiwanese Kitchens', *Environ. Mol. Mutagen.*, **31**(1), 92–6 (1998).
28. T.A. Chiang, P.F. Wu, L.F. Wang, H. Lee, C.H. Lee, Y.C. Ko, 'Mutagenicity and Polycyclic Aromatic Hydrocarbon Content of Fumes from Heated Cooking Oils Produced in Taiwan', *Mutat. Res.*, **381**(2), 157–61 (1997).
29. P.F. Wu, T.A. Chiang, Y.C. Ko, H. Lee, 'Genotoxicity of Fumes from Heated Cooking Oils Produced in Taiwan', *Environ. Res.*, **80**(2) Pt 1, 122–6 (1999).
30. G.K. Montizaan, P.G.N. Kramers, J.A. Janus, R. Posthumus, 'Integrated Criteria Document, Polynuclear Aromatic Hydrocarbons (PAH): Effects of 10 Selected Compounds', National Institute of Public Health and Environmental Protection (RIVM), Bilthoven, The Netherlands, 1989.
31. R. Lemus, A.A. Abdelghani, T.G. Akers, W.E. Horner, 'Potential Health Risks from Indoor Exposure to Chlorpyrifos (O,O-diethyl O-[3,5,6-trichloro-2-pyridyl] Phosphorothioate)', *Rev. Environ. Health.*, **12**(2), 91–7 (1997).
32. D.L. Davis, A.K. Ahmed, 'Exposures from Indoor Spraying of Chlorpyrifos Pose Greater Health Risks to Children than Currently Estimated', *Environ. Health. Perspect.*, **106**(6), 299–301 (1998).
33. S.K. Harrel, J.B. Barnes, F. Rivera Hidalgo, 'Aerosol and Splatter Contamination from the Operative Site During Ultrasonic Scaling', *J. Am. Dent. Assoc.*, **129**(9), 1241–9 (1998).
34. J.B. Barnes, S.K. Harrel, F. Rivera Hidalgo, 'Blood Contamination of the Aerosols Produced by In Vivo Use of Ultrasonic Scalers', *J. Periodontol.*, **69**(4), 434–8 (1998).
35. M. Dechow, H. Sohn, J. Steinhanses, 'Concentrations of Selected Contaminants in Cabin Air of Airbus Aircrafts', *Chemosphere*, **35**(1–2), 21–31 (1997).
36. K. Svendsen, B. Hilt, 'Exposure to Mineral Oil Mist and Respiratory Symptoms in Marine Engineers', *Am. J. Ind. Med.*, **32**(1), 84–9 (1997).
37. M. Katz (ed.), *Methods of Air Sampling and Analysis*, 2nd edition, APHA Intersociety Committee ACS, ACGIH, AIHA, AIChE, APCA, APWA, ASCE, ASME, AOAC, HPS, ISA, SAE, American Public Health Association, Washington DC, 27–37, 1977.
38. P.W. Wilsey, J.H. Vincent, M.J. Bishop, L.M. Brosseau, I.A. Greaves, 'Exposures to Inhalable and 'Total' Oil Mist Aerosol by Metal Machining Shop Workers', *Am. Ind. Hyg. Assoc. J.*, **57**(12), 1149–53 (1996).
39. D. Leith, F.A. Leith, M.G. Boundy, 'Laboratory Measurements of Oil Mist Concentrations Using Filters and an Electrostatic Precipitator', *Am. Ind. Hyg. Assoc. J.*, **57**(12), 1137–41 (1996).
40. K.J. Furton, E. Jolly, G. Pentzke, 'Recent Advances in the Analysis of Polycyclic Aromatic Hydrocarbons and Fullerenes', *J. Chromatogr.*, **642**, 35–47 (1993).
41. S.A. Wise, L.C. Sander, W.E. May, 'Determination of Polycyclic Aromatic Hydrocarbons by Liquid Chromatography', *J. Chromatogr.*, **642**, 329–349 (1993).
42. Method 102A & D. In *Methods of Air Sampling and Analysis*, 3rd edition, ed. J.P. Lodge, APHA Intersociety Committee APCA, ACS, AICHE, APWA, ASME, AOAC, HPS, ISA, American Public Health Association, Washington DC, 249–61, 1989.
43. F. Ariese, 'Shpol'skii Spectroscopy and Synchronous Fluorescence Spectroscopy: (Bio)Monitoring of Polycyclic Aromatic Hydrocarbons and Their Metabolites', Thesis, Vrije Universiteit Amsterdam, 1993.
44. S.A. Shulman, J.H. Groff, M.T. Abell, 'Performance of Laboratories Measuring Silica in the Proficiency Analytical Testing Program', *Am. Ind. Hyg. Assoc. J.*, **53**(1), 49–56 (1992).
45. W.E. Wilson, H.H. Suh, 'Fine Particles and Coarse Particles: Concentration Relationships Relevant to Epidemiologic Studies', *J. Air. Waste. Manag. Assoc.*, **47**(12), 1238–49 (1997).
46. H.J. van de Wiel, L. Mølhave, 'Sampling Strategies for Indoor Air Analysis', in *Environmental Carcinogens – Methods of Analysis and Exposure Measurement*, eds. H.J. Seifert, H.J. van de Wiel, B. Dodet, I.K. O'Neill, Volume 12, ''Indoor Air'', IARC, Lyon, France, 96–118, 1993.

Carcinogens, Monitoring of Indoor Air

Max R. Peterson
Research Triangle Institute, USA

1 Introduction **1**
1.1 Carcinogens in Indoor Air 2
1.2 Sampling and Analysis 2
1.3 Basic Monitoring Approaches 3
1.4 Overview 4
2 Radon **4**
2.1 Overview 4
2.2 Monitoring Methods 5
3 Organic Compounds **7**
3.1 Overview 7
3.2 General Monitoring Methods for Organic Compounds 7
3.3 Organic Carcinogens of Particular Concern 9
4 Inorganic Species **10**
4.1 Overview 10
4.2 General Monitoring Methods for Inorganic Species 10
4.3 Inorganic Carcinogens of Particular Concern 11
5 Asbestos **13**
5.1 Overview 13
5.2 Sampling Airborne Fibers 13
5.3 Laboratory Analysis Methods 13
6 Particles **14**
6.1 Overview 14
6.2 Sampling Airborne Particles 14
6.3 Analyzing Particles and Particle-bound Chemical Species 15
6.4 Particles of Particular Concern 15
7 Remarks **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

Carcinogens in indoor air include radioactive radon, a variety of organic compounds, several inorganic species, asbestos fibers, and a large number of chemical species associated with suspended particulate matter (SPM). Within these categories, carcinogens may be grouped by either chemical class and physical properties or by the methods used to measure them. Generally, measurement methods within a category are at least similar if not identical.

A typical measurement method includes procedures for sampling a known volume of air in a way that quantitatively collects the species of interest, transferring collected samples to a laboratory, and measuring species of interest in the samples using an appropriate validated analysis method. The sampling location may be in the breathing zone of a person (to measure personal exposure) or at a specific location in a building or room (to measure potential for exposure at that location). A sample may be collected very quickly (to obtain a snapshot of carcinogen concentration at a specific time) or very slowly over a period of hours, days, or weeks (to obtain an integrated or average carcinogen concentration).

The analysis method is essentially independent of the sampling location and sampling time, but collection of an adequate quantity of the species of interest is crucial. In general, radon is measured as radiation emitted from the sample; organic compounds are measured using gas or liquid chromatography with an appropriate detector; carcinogenic inorganic species are measured using atomic absorption spectroscopy (AAS) or atomic emission spectroscopy (AES); asbestos fibers are visually counted at high magnification; and particles are extracted with a solvent to quantitatively remove the species of interest, which are then measured using one of the organic or inorganic methods described above. Concentrations of the species of interest can then be calculated by dividing the quantity of each species measured in the collected sample by the volume of air sampled.

1 INTRODUCTION

Indoor air contains a complex mixture of chemical species in the vapor state and associated with airborne particulate matter. The number of chemical species present at one time can range into the thousands, especially if environmental tobacco smoke (ETS) is present. Concentrations of chemical species and particles, while usually quite low from an analytical perspective, may be significant, and the concentrations vary with location and with time within a given indoor air space. Some of the pollutants within this complex mixture appear to be harmless; but some are irritants, some are toxic, and some are known or suspected human carcinogens. Carcinogens are the focus of this article, although many of the monitoring approaches described here are also used for other species. (This article is an updated version of a paper by Peterson and Jayanty.[1])

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1.1 Carcinogens in Indoor Air

Pollutants, including known or suspected carcinogens, found in residential or nonoccupational indoor air come from a variety of sources. Some pollutants come from building materials, carpet, fabrics, adhesives, sealants, paints, stains, and other materials used to construct or maintain a home or other building; some pollutants are brought in or tracked in from outdoors; and some pollutants are generated by human activities such as smoking tobacco products, cooking, or using cleaning and personal care products. Even seemingly innocuous activities like sweeping or vacuuming floors and furnishing can fill the air in an indoor environment with respirable particles containing heavy metals tracked in from outdoors and carcinogenic species formed during heating and cooking activities.

1.1.1 Categories

For monitoring purposes, chemical species, including carcinogens and other pollutants, are often grouped on the basis of chemical class, chemical properties, or physical state. Species that can be measured by a single method are often treated as a group, but particularly hazardous members of a group may also be treated separately in more tightly focused monitoring programs. The categories of carcinogens in the following sections have been arranged in the general order of simplest to most complex, although there is considerable overlap. Major categories include radon, organic compounds, inorganic species, asbestos, and particles. Radon and asbestos could have been included under inorganic species and particles, respectively, but methods to measure these two carcinogens are quite different from methods for measuring other carcinogens.

1.1.2 Physical States

Carcinogens in indoor air may exist in the gaseous, liquid, or solid state; they may be dissolved in liquid or adsorbed on to solid material associated with airborne particulate matter; or they may exist in any combination of these states. Some carcinogens, like radon and formaldehyde, exist primarily in the gaseous state, while others, like asbestos and some of the heavy metals, exist only in the solid state and occur in air as suspended particles or fibers. A third group of carcinogens, which includes low-volatility compounds like some pesticides and some components of ETS, may exist simultaneously in any combination of physical states.

1.1.3 Complex Mixtures

Indoor air is a complex mixture of gaseous species, suspended liquid and solid particles (SPM), and airborne fibers. Chemical species, including carcinogens, can be found in any or all of those states. Some individual chemical species may exist in the gaseous phase, be dissolved in suspended liquid particles, and be adsorbed on to solid particles at the same time, with molecules constantly moving from one state to another. The distribution of a particular species among the various physical states is determined by its vapor pressure (or volatility), its solubility in whatever liquid aerosol materials are present, and its affinity for active sites on the surface of whatever solid particles are present; and its concentrations in the gaseous, dissolved, and particle-bound states are affected by the temperature of the air. In addition, the sources of a species and the so-called ''sinks'' that remove the species from air are not evenly distributed, nor are source contribution rates and sink removal rates constant and perfectly balanced. Thus, the concentration of a species also varies with time and with location in a given indoor environment.

1.2 Sampling and Analysis

Measurement methods generally involve sampling indoor air to collect the species of interest and analyzing the collected sample to measure those species. The two parts, sampling and analysis, may be done on-site and essentially simultaneously to obtain more or less real-time measurements. Alternatively, collected samples may be transported to a laboratory for analysis.

Carcinogens are typically present in indoor air at extremely low concentrations.[2] A sufficient volume of air must be sampled to yield an amount of the species of interest that is measurable by the chosen analysis method. Both the practical quantitation limit (PQL) of the analysis and an approximate or a maximum allowable (e.g. threshold limit value) concentration of the species of interest at the sampling location must be known before the volume of air that must be sampled can be determined.

1.2.1 Fundamental Objectives of Sampling

The fundamental objective of sampling is to obtain samples that contain representative concentrations of the species of interest for the purposes of the measurement. Samples are generally collected in a fashion that either mimics human exposure or provides data that can be related to human exposure. Whole-air samples can be collected in suitable containers, or the species of interest may be removed by some physical or chemical process from the air sample during collection.

Particles and fibers are typically collected on a filter, while gaseous species may be collected in a stainless steel canister or a Tedlar® or aluminized bag, bubbled through a liquid-filled impinger, or concentrated on a

sorbent. Filters and sorbents remove and concentrate species of interest from a known volume of sampled air. Impingers are often used for collecting reactive or unstable species and generally are filled with a solution containing a derivatizing agent, which converts reactive species to stable derivatives. Species existing in both gaseous and particle-bound states are typically sampled using both a filter and a sorbent. Thus, collected samples may be whole-air samples collected in canisters or bags, particles or fibers collected on filters, or chemical species collected on sorbents or in a liquid-filled impinger.

The volume of air sampled (or sample volume) must be large enough to provide a measurable amount of the species of interest. On the other hand, the volume of air sampled is limited by the capacity of the sampling device or medium. Canisters, for example, generally have volumes of 6 L or less. Thus, using negative pressure to drive sampling into an evacuated canister limits the volume of air that can be collected to something less than the volume of the canister.

The volume of air that can be sampled through a filter is limited by increasing pressure drop across the filter as filter pores become clogged with collected particles or fibers. Samples are invalid if the pressure drop across the filter exceeds the maximum drop specified in the chosen sampling method.

The volume of air that can be sampled through a sorbent is limited by the weakness of the attachment of the species of interest to the sorbent. A sorbent is generally selected on the basis of its ability both to retain a species during sampling and to release the species for analysis. Thus, molecules of the target species that are trapped from the air moving over the sorbent subsequently move back into the gas phase, flow with the sampled air for a short distance, and then reattach to the sorbent. This process is repeated over and over again so that molecules of the target species migrate further and further into the sorbent bed. Breakthrough is said to have occurred if a measurable amount of the target species actually passes beyond the downstream end of the sorbent bed. The total volume of air sampled when breakthrough occurs is called the breakthrough volume. The volume of air sampled through a sorbent must be less than the breakthrough volume of the target species for that sorbent bed.

Whatever the method of sampling, the temperature, pressure, and volume of air sampled must be known for later calculations of species concentrations. In some cases, the humidity of the sampled air is also a factor.

1.2.2 Fundamental Objectives of Analysis

The fundamental objective of analysis is to provide reliable, accurate, and precise measurements of the species of interest. The analysis must provide an unambiguous measurement of the species without loss (e.g. through decomposition) or enhancement (e.g. through contamination) of the species. Often the species of interest must be separated from other species present in collected samples prior to measurement. In this situation, the analytical system may be used to both isolate and measure species of interest.

An appropriate analytical system is one that responds in a characteristic and quantifiable way to species of interest. In modern instruments, the response of an analytical system to a species is often converted into an electrical signal with a magnitude that is directly related to the amount of the species present in the detection device of the instrument. In such systems, the response of the instrument to the species in a collected sample is compared to the response of the instrument to standards containing known amounts of the species. Analysis of standards containing different amounts of a species and the subsequent plotting of instrument response versus amount of species analyzed is referred to as calibration. The range of the calibration (or range of amounts of species in the standards analyzed for the calibration) must bracket the amounts of species measured in collected samples.

1.3 Basic Monitoring Approaches

A monitoring approach must be chosen based on the goals of the desired measurement, the sampling and analysis methods available for the species of interest, and the concentrations of the species of interest in the indoor environment to be monitored. Some of the choices to be considered in designing an appropriate monitoring plan are described below.[3]

1.3.1 Selection of Monitoring Location

Because concentrations of species vary with location and with time within a given indoor environment and because people often move in and out among a variety of environments, monitoring or sampling locations must be carefully chosen to ensure that measurements are meaningful. The selection of a monitoring or sampling location is based on the purpose of the measurement. To measure the exposure of a particular person to a particular species of interest, monitoring or sampling is done near the breathing zone of the person during his or her usual daily activities. To measure the concentration of a particular species in a particular indoor environment, monitoring or sampling is done at fixed points within that environment. In survey studies, many people and many indoor environments are monitored or sampled according to a specific protocol, which details the sampling and analysis parameters.

1.3.2 Real-time Measurements versus Laboratory Analysis of Samples

So-called ''real-time'' measurements require that both sampling and analysis be done on-site. The term ''real-time'' is somewhat misleading because there is often a lag time between sampling and analysis, particularly if the target species must be concentrated prior to measurement. ''Real-time'' usually means that measurement results are available a short time (a few minutes to an hour or so) after the sample is taken. Real-time measurements generally require a dedicated on-site sampling and analysis system or an automated monitor.

Analyzing collected samples in a laboratory means that measurement results will be available within a few days or weeks. The advantages of laboratory analysis include a controlled analytical environment, the availability of more sophisticated equipment, and the ability to analyze samples from many locations on a single instrument.

1.3.3 Grab Sampling, Integrated Sampling, and Continuous Sampling

Samples of indoor air may be collected quickly over a very short time or at a slower sampling rate over an extended period of time. Samples taken quickly are referred to as grab samples, and samples that are taken slowly over an extended period of time are referred to as integrated samples. Most continuous monitoring devices cycle repeatedly between sampling and analysis, but more sophisticated instruments allow simultaneous analysis (of the previous sample) and collection (of the next or subsequent sample).

A grab sample, which may be collected by simply opening a valve on an evacuated canister for a few seconds, represents a snapshot of the air at the sampling location at the specific time the sample was taken. A series of grab samples taken at a single location at specific time intervals can be used to determine variations in concentrations of species over time. An integrated sample, which is collected at a fixed sample flow rate over an extended time period, can be used to determine average concentrations of species over time. Integrated samples are usually used to measure individual human exposure. Continuous sampling devices can provide near-real-time measurements, but such devices require a dedicated analytical system at the sampling site.

1.3.4 Active Sampling versus Passive Sampling

In active sampling, air is drawn through or into a sample collection device by means of a pump or an evacuated canister. The volume of air sampled can be calculated by multiplying the flow rate through the sampling pump by the sampling time or from the pressure increase inside and the volume and temperature of the canister. Active sampling can be used for essentially all of the carcinogens that occur in indoor air, although samples collected in this way may not provide the best available estimates of human exposure.

In passive sampling, gaseous species diffuse through a permeable membrane and interact with, or are trapped by, some medium (e.g. a sorbent). Passive sampling does not require a pump and represents a fairly inexpensive way to obtain integrated samples. Passive samplers are also small and unobtrusive.

1.4 Overview

Monitoring approaches used to measure various categories of carcinogens in indoor air are presented below. Some of the methods were developed to measure species in workplace air or in ambient (outdoor) air but should, in principle, also be applicable to measurement of the same species in indoor air. The categories addressed include radon, organic compounds, inorganic species, asbestos, and particles. Methods sometimes span several categories; for example, radon decay products, SVOCs and nonvolatile organic compounds, and inorganic species are associated with particles.

2 RADON

Radon is considered a carcinogen because of its radioactivity. The actual harm to human health comes from the radioactivity of the decay products of radon. Radon is a gas and, if inhaled, is not retained well by the lungs. When radon atoms undergo radioactive decay, the decay products do not remain in the gas phase but become attached to airborne particles, which, if inhaled, can be retained in the lungs indefinitely. It is the particle-bound radon decay products that are actually the cause of the increased risk of lung cancer attributed to radon.

2.1 Overview

Concentrations of radon and radon decay products (or progeny) in indoor air vary with location and with time. The primary isotope of interest, radon-222 or $^{222}_{86}\mathrm{Rn}$, and its decay products are part of the radioactive decay series leading from a naturally occurring, long-lived radioactive isotope (uranium-238) to a stable isotope (lead-206). A portion of this series, showing the formation of radon-222 and its decay products of interest (polonium-218, lead-214, bismuth-214, and polonium-214), is shown in Figure 1. The primary decay mode of an isotope is denoted by the type of radiation associated with fission of that isotope. An α-particle is a helium-4 nucleus ($^{4}_{2}\mathrm{He}^{2+}$)

Isotope	$^{226}_{88}Ra$	→ $^{222}_{86}Rn$	→ $^{218}_{84}Po$	→ $^{214}_{82}Pb$	→ $^{214}_{83}Bi$	→ $^{214}_{84}Po$
Decay mode		α	α	α	β	β	
Half-life	1600 years	3.82 days	3.04 min	26.9 min	19.7 min	164 μs

Figure 1 Formation of radon-222 and its decay products.

ejected from the nucleus of an isotope during radioactive decay, and a β-particle is a high-energy electron ejected in a similar fashion. Loss of an α-particle reduces the mass number (number of protons plus number of neutrons) of an isotope by four and reduces the atomic number (number of protons) by two. A β-particle is formed when a neutron in the nucleus of an atom is converted to a proton, which remains in the nucleus, and a high-energy electron (the β-particle), which is ejected from the nucleus. Loss of a β-particle does not affect the mass number of the atom but increases the atomic number by one. In both types of radioactive decay, an atom of one element is changed into an atom of another element because the atomic number changes. Emission of γ-rays, which are high-energy photons, during radioactive decay does not affect the mass number or the atomic number of an isotope. The half-life of an isotope is the time required for half of the atoms of that isotope in a sample to undergo radioactive decay.

A second radon isotope, radon-220 (sometimes called thoron) or $^{220}_{86}Rn$, and its decay products are part of the radioactive decay series leading from another naturally occurring, long-lived radioactive isotope (thorium-232) to a different stable isotope of lead (lead-208). A portion of this series, showing the formation of radon-220 and its decay products of interest (polonium-216, lead-212, bismuth-212, and polonium-212), is shown in Figure 2. Radon-220 and its decay products, while just as deadly as isotopes associated with radon-222, pose a lower risk to human health because of the short (55.6 s) half-life of radon-220, which limits its time in the gas phase and therefore its migration distance.

2.1.1 Radon

Of all of the isotopes in the two decay series, only the radon isotopes are gases at ambient conditions. Existence as a gas, which allows radon to migrate from soil or rock into indoor air in the first place, provides a distinguishing characteristic that can be used to separate radon from other radioactive species.

Isotope	$^{224}_{88}Ra$	→ $^{220}_{86}Rn$	→ $^{216}_{84}Po$	→ $^{212}_{82}Pb$	→ $^{212}_{83}Bi$	→ $^{212}_{84}Po$
Decay mode		α	α	α	β	β	
Half-life	3.66 days	55.6 s	0.15 s	10.6 h	60.6 min	298 ns

Figure 2 Formation of radon-220 and its decay products.

2.1.2 Radon Decay Products

When atoms of radon gas undergo radioactive decay, the newly formed polonium atoms (and any subsequently formed polonium decay products) immediately condense or attach on to any surface with which they come into contact. In indoor air, that surface is usually associated with airborne particulate matter.

2.2 Monitoring Methods

Monitoring of radon and radon decay products in indoor air is complicated by the nuclear instability of the target species. With half-lives ranging from almost 4 days for radon-222 to a fraction of a second for polonium-216, measurements based on emitted radiation must be made shortly after collection.

Indoor air may be sampled for radon or radon progeny measurements by grab sampling, integrated sampling, or continuous sampling. Separation of radon from progeny may be accomplished by allowing the radon to diffuse through a passive barrier (e.g. foam rubber) that does not allow the passage of particulate material containing the radon progeny.[4]

Methods for measuring radon and radon progeny actually measure emitted radiation. In the absence of other radioactive species, emitted radiation is directly proportional to the concentration of radon and radon progeny in the sample. The emitted radiation may be in the form of α- or β-particles or γ-rays, all of which can be measured with appropriate instrumentation.[5] Table 1 provides an overview of applications and sensitivities of devices used to measure radon and its decay products.

2.2.1 Direct Measurement

Direct measurements are made with either a scintillation cell or an ionization chamber. Both approaches can be used in continuous monitoring or to analyze grab samples. The inside walls of a scintillation cell are coated with a phosphorescent material, which emits light when struck by an α-particle from decaying radon. The light from the phosphor is measured with the aid of a photomultiplier tube. The measurement is used to count the scintillations due to radioactive decay within the cell. The cell can be used to measure radon in grab samples or in a flow-through mode for continuous measurements. Progeny can be excluded (as described above) to allow only radon into the cell. In this mode, radon decaying within the cell forms progeny that attach to the walls

Table 1 Applications and sensitivities of some radon measurement devices[a]

Instrument type	Sampling	Sensitivity[c] ($Bq\ m^{-3}$)	Purpose[d]
Direct measurement			
Scintillation cell	Grab or continuous	3.7	Screening, diagnostic
Ionization chamber	Grab or continuous	3.7	Screening, diagnostic
Passive barrier method[b]			
Scintillator	Continuous	3.7	Screening, diagnostic
Thermoluminescent detector chip	Integrating	0.08–8.1	Screening, large-scale survey
Two-filter method	Grab or continuous	3.7	Diagnostic
Passive sampling devices			
Activated charcoal	Integrating	7.4 (100-h exposure)	Screening, large-scale survey
α-Track	Integrating	18.5 (30-days exposure)	Screening, large-scale survey

[a] Adapted from Cothern and Smith, Jr.[5]
[b] With radon progeny collected on or close to a scintillator or thermoluminescent detector chip.
[c] A sensitivity less than the value shown is generally achievable, depending on the specific instrument used ($1\ pCi\ L^{-1} = 37\ Bq\ m^{-3}$).
[d] Three typical purposes for the measurement include:
- screening – to detect rapidly and inexpensively the occurrence of high radon concentrations;
- diagnostic – to measure such specific parameters as short-term spatial and temporal concentration variations, relationship with other factors (e.g. ventilation rate), equilibrium fraction of each of the radon progeny, or the effect of remedial actions;
- large-scale survey – to obtain a large number of time-averaged measurements aimed at evaluating the exposure of the public.

of the cell and subsequently further decay. This decay of progeny within the cell also contributes to the scintillation measurement and must be taken into account in flow-through mode.[6]

In an ionization chamber, the electric field between two electrodes is altered when the decay of radon or progeny ionizes some of the air molecules in the chamber. The current flow between the two electrodes is measured with an electrometer, and the measurement is converted into concentration of radon or radon decay products.

2.2.2 *Passive Barrier Method*

In the passive barrier method, radon decay products are collected on or in close proximity to a scintillator or a thermoluminescent dosimeter (TLD) chip. The scintillator allows continuous measurement of decay products, and the TLD chip, which is usually made of lithium fluoride or calcium fluoride, provides an integrated measurement.

2.2.3 *Two-filter Method*

In the two-filter method, sampled air is drawn through a small tube that is equipped with filters on both ends. The upstream (or entrance) filter removes radon decay products from the sample allowing only radon to pass into the tube. Some of the radon undergoes radioactive decay while it is transversing the tube, and the resulting polonium-218 and any subsequent decay products formed in the tube are collected on the downstream (or exit) filter. After sampling, the downstream filter is removed and the collected decay products counted by measurement of α-decays. This method can be used in a continuous or integrative mode by placing a TLD chip at the downstream filter during sampling.

2.2.4 *Passive Sampling Devices*

Common passive sampling devices include charcoal canisters and α-track detectors. In a charcoal canister, radon is collected by adsorption on to activated charcoal. Radon collected on charcoal can be measured in several ways. In one approach, radon is de-emanated from charcoal into a scintillation cell and measured by counting α-decays. In another approach, the charcoal is heated to desorb collected radon, which is then measured by counting γ-emissions. In a third approach, the charcoal is placed in a liquid scintillation fluid and measured by counting scintillation events.

An α-track detector contains a small slab of material that sustains microscopic damage when struck by an α-particle. The microscopic gouge marks or tracks made by the α-particles are enlarged by chemical etching and visually counted with the aid of a microscope. An α-track

detector can be exposed for up to a year before it is etched and counted.

3 ORGANIC COMPOUNDS

3.1 Overview

Organic compounds are simply compounds containing the element carbon. The vast majority of pollutants that may be present in indoor air are organic compounds.

3.1.1 Classification of Organic Pollutants

In addition to the usual classification of organic compounds by functional group (e.g. alcohols, ketones, aromatic compounds, halogenated compounds, etc.), organic pollutants in indoor air are often classified on the basis of boiling points. The boiling point of a compound is indicative of the compound's volatility, which to a large extent determines the distribution of the compound between gaseous and condensed phases. A general classification scheme based on boiling points is shown in Table 2. The boiling point ranges assigned to various classes are somewhat arbitrary, and compounds with boiling points outside the range for a particular class may be included with that class in a particular study or for analysis by a particular analytical method.[7]

VVOCs exist almost completely in the gaseous state in indoor air. Some VVOCs actually have boiling points below room temperature. VOCs can exist in the gaseous state in air at relatively high concentrations. SVOCs are much less volatile and tend to have low gas-phase concentrations, but SVOCs can exist at high concentrations in liquid aerosol droplets and on solid particles. Some organic compounds, often called particulate organic matter (POM), have such low volatilities that they exist in air almost exclusively as or on particles.

3.1.2 Complexity of Organic Mixtures

Measuring concentrations of specific organic compounds in indoor air can be quite a formidable challenge because of the complexity and variability of the indoor air matrix. Literally thousands of organic compounds are present in such complex mixtures as ETS. If suspended liquid- and solid-phase particulate matter is present, the VOCs and SVOCs exist in a complex, constantly changing distribution among gaseous, condensed, dissolved, and adsorbed states.

3.2 General Monitoring Methods for Organic Compounds

Monitoring of organic compounds is generally accomplished in three steps: collection, separation, and measurement.

3.2.1 Collection

VOCs in indoor air are collected either as whole-air samples in evacuated stainless steel canisters or by passing a measured volume of sampled air through a solid sorbent, which removes compounds of interest from the sampled air.[8] While both approaches to collection work well for time-averaged or integrated samples (up to 24 h), evacuated canisters can also be used for grab sampling. Because the sample is collected over a very short time interval (typically a few seconds to a minute), grab sampling can be used to obtain a snapshot of pollutant concentrations. A series of grab samples can be used to measure concentration variability over time. Whole-air samples collected in a canister must be concentrated, usually by passing a known amount of air from the canister through a cryogenic or sorbent-based preconcentration unit, prior to analysis.

Sorbents are granular solids and are packed in glass tubes in sections held in place by plugs of glass wool.

Table 2 Classification of organic pollutants by boiling point

Description	Abbreviation	Approximate boiling point range (°C)[a]		Sampling
		Lower end	Higher end	
Very volatile organic compounds	VVOCs	<0	50–100	Batch sampling; adsorption on to activated charcoal
Volatile organic compounds	VOCs	50–100	240–260	Adsorption on to Tenax, carbon molecular sieves, of activated charcoal
Semivolatile organic compounds	SVOCs	240–260	380–400	Adsorption on polyurethane foam or XAD-2
Organic compounds associated with particulate matter	POM	>380		Collection on a filter

[a] Polar compounds generally have higher boiling points than nonpolar compounds of similar molecular weight.

Sorbents trap or collect compounds of interest from sampled air as it is drawn through the tube. Sorbents and the stationary phases used in chromatography (discussed in the next section) operate according to the same basic mechanisms. In some applications, several different sorbents may be packed in separate sections within the same glass tube.[9] These multisorbent tubes are often used if the target compounds span a wide range of polarities or molecular sizes.

The ideal sorbent for a given set of target compounds will completely trap those compounds from a large volume of sampled air without loss (or breakthrough) and will completely release those compounds when either heated (thermal desorption) or treated with a solvent (solvent desorption). Real sorbents, of course, generally have collection efficiencies and desorption efficiencies that are less than 100% for most organic compounds. Some of the more common sorbents used for collecting organic compounds in indoor air are described in Table 3.[10]

With thermal desorption, the entire sorbent tube is quickly heated to a high temperature (300 °C or so) and desorbed compounds are backflushed off the sorbent in the tube with a flow of helium (or other carrier gas). The entire sample is then used in a single analysis. Although only a single measurement for each compound is obtained per tube, the method limit of quantitation (MLOQs), which take into account the fraction of a collected sample that is actually analyzed) are much lower because all of each target compound in the sample is delivered to the analysis system at one time.

With solvent desorption, the sorbent sections from each tube are desorbed separately with one or more suitable solvents, and the resulting solutions analyzed separately. Concentration comparisons can then be made to see if any of the target compounds passed through the front section of sorbent and were collected on some downstream section. Solvent desorption offers the advantage of multiple analyses of each desorption solution but the MLOQs are typically higher because only a small fraction of the collected sample is actually analyzed.

Reactive or polar (heteroatom-containing) VOCs are usually collected in a glass impinger filled with a solvent in which the VOCs are very soluble and perhaps containing, if one or more of the target VOCs is reactive, a derivatizing agent. The derivatizing agent converts reactive VOCs into stable derivatives. Derivatives are often solvent-extracted from the impinger solution prior to analysis.

Table 3 Common sorbents for collecting organic compounds in air

Tenax GC resin	Porous polymer of 2,6-diphenyl-*p*-phenylene oxide; hydrophobic; not suitable for very light organics; lower capacity than XAD-2 but may be thermally desorbed
XAD-2 resin	Polystyrene–divinylbenzene porous polymer; hydrophobic; not suitable for very light organics; higher capacity than Tenax GC but requires solvent desorption
Activated charcoal	Typically a coconut charcoal; relatively high retention of water; organics are very strongly adsorbed; requires solvent desorption
Graphitized carbon black	Prepared by heating thermal carbon blacks at 3000 °C under an inert gas atmosphere; nonselective; low retention of water and light gases; may be thermally desorbed
Carbon molecular sieves	Pyrolyzed porous beads of poly(vinylidene chloride); low capacity for water; organics are very strongly adsorbed; higher capacity for organics than graphitized carbon blacks; better than Tenax GC for very volatile compounds; very light, volatile compounds may be thermally desorbed
Polyurethane foam	Usually used to collect SVOCs, including those volatilized from particulate matter on a filter; must be cleaned prior to use

3.2.2 Separation

Either gas chromatography (GC) or high-performance liquid chromatography (HPLC) is used to separate organic compounds. As a general rule, thermally stable gaseous, volatile, or semivolatile compounds are separated by GC, and reactive or nonvolatile compounds are separated by HPLC. Both involve differential migration of organic compounds moving in a mobile phase across a stationary phase. The time required for a particular compound to pass through a chromatography column is referred to as the retention time for that compound. The goal of chromatographic separation is to isolate each compound of interest from all other species present so that the pure compounds can be measured one by one as they exit (or elute from) the chromatography column.

In GC, the mobile phase is a gas and the stationary phase is a solid or high-boiling liquid. In modern GC columns, the stationary phase is usually chemically bonded to a solid support or to the inside of a long (30–60 m) capillary column. Temperature programming of the column is used to speed the analysis without jeopardizing the complete separation of the compounds of interest from other species present.

In HPLC, the mobile phase is a liquid or a solution. Unlike GC, the composition of the mobile phase is crucial

to HPLC separations. Solvent programming, or changing the ionic strength of the mobile phase during analysis, is used in a fashion analogous in many ways to the use of temperature programming in GC.

3.2.3 Measurement

Measurement of organic compounds exiting a GC or HPLC column is accomplished by means of a detector, which generates an electrical response when a compound passes through it. The magnitude of the electrical response is proportional to the amount of the compound present in the detector. From a measurement perspective, concentrations of individual organic compounds in indoor air are usually quite low. The analytical limit of quantitation (ALOQ) for the measurement of a particular compound is usually expressed as the minimum mass of that compound that results in a measurable and reproducible electrical response on the detector used for the measurement. The ALOQ is usually estimated as the mass of compound required to give a detector response that is 10 times the standard deviation of detector responses to at least seven replicate measurements for a mass of the compound that is slightly above the ALOQ. Often, organic compounds in indoor air must be concentrated to provide sufficient mass of each target compound to give detector responses above the individual ALOQs in a given analysis. Concentrating of compounds is usually accomplished by passing sampled air through a collection medium, usually a solid sorbent, and subsequently analyzing the collection medium.

Table 4 lists some common GC detectors and approximate detection limits and applications for each.[10] Some of the detectors are universal (respond to all compounds), and some respond only to compounds that meet certain criteria. All of the detectors rely on retention time for compound identification, and, for all but one of the detectors, retention time is the only evidence available for compound identification.

Only the mass spectrometric detector (MSD) allows unambiguous identification of compounds. In an MSD, a portion of the column effluent passes through an ionization chamber, which ionizes molecules in the effluent with sufficient energy to cause some of them to break into fragments. The fragments are separated on the basis of their mass-to-charge (m/z) ratios, and the relative abundances of the fragments are measured. In scan mode, relative abundances are monitored for a wide range of m/z values for the entire chromatogram. This provides a three-dimensional array of data (time, m/z value, and relative abundance). A mass spectrum, which is a plot of m/z value versus relative abundance, can be displayed for any given time in the chromatogram. Identification of a compound is accomplished by electronically matching the mass spectrum measured during analysis at the known retention time for the compound to mass spectra stored in an on-line library. In single-ion monitoring (SIM) mode, only selected m/z values are measured. Increased sensitivity is obtained in SIM mode because the instrument's data collection resources are focused on monitoring only one or, at most, a few m/z values instead of the several hundred values monitored in scan mode.

Common HPLC detectors include refractometers and ultraviolet (UV) detectors, including the photodiode array (PDA) detector. The PDA detector monitors UV light absorption over a range of wavelengths to give a three-dimensional data set that can be used for compound identification.

3.3 Organic Carcinogens of Particular Concern

Organic compounds that are classified as known or suspected human carcinogens and that may be present in indoor air include benzene (CAS 71-43-2), *p*-dichlorobenzene (CAS 541-73-1), carbon tetrachloride (CAS 56-23-5), chloroform (CAS 67-66-3), methylene chloride (CAS 75-09-2), formaldehyde (CAS 50-00-0), polycyclic aromatic hydrocarbons (PAHs), and some pesticides. Except for the last three, all of the compounds in the list can be measured using methods described above.

Table 4 Some common GC detectors

Detector	Detection limit	Applications
Mass spectrometric	Sub ng	Responds to all compounds
Flame ionization	Sub ppm, sub ng	Responds to all combustible substances
Thermal conductivity	<100 ppm	Responds to all compounds
Electron capture	pg	Responds to all electron-capturing species, especially halogenated compounds
Nitrogen/phosphorus	10 pg	Responds to compounds containing nitrogen or phosphorus
Flame photometric	≥Sub ppb	Responds to compounds containing sulfur or phosphorus
Photoionization	10 pg, ≥ppb	Responds to compounds with an ionization potential < 11.7 eV
Argon ionization	Sub ppm to ppb	Responds to compounds with an ionization potential < 11.8 eV
Helium ionization	Sub ppm to ppb	Responds to compounds with an ionization potential < 19.8 eV

3.3.1 Formaldehyde

Formaldehyde is a fairly reactive VVOC that is very difficult to measure directly, and devising an appropriate sampling strategy is not always straightforward.[11] Formaldehyde is usually converted to a stable derivative (often by reaction with pararosaniline or 2,4-dinitrophenylhydrazine) by bubbling the sampled air through an aqueous solution of the derivatizing agent or by pulling the sampled air through a cartridge containing the derivatizing agent coated on some solid support, and the derivative is measured by some calorimetric, HPLC, or GC method.[8]

Passive badge monitors are also used to sample formaldehyde in air. The US Occupational Safety and Health Administration (USOSHA) Method ID-205 provides for sampling using a passive badge containing bisulfite-impregnated paper. Samples are desorbed with deionized water, the desorption solution is acidified and treated with chromotrophic acid, and the color complex is analyzed using a UV spectrophotometer at a wavelength setting of 580 nm. Method ID-205 has a quantitative detection limit of 0.11 parts per million of formaldehyde in air with a 4-h badge exposure time.[12]

3.3.2 Polycyclic Aromatic Hydrocarbons

PAHs are SVOCs containing only carbon and hydrogen with molecular structures composed of multiple, fused 5- and 6-carbon aromatic rings. Carcinogenic PAHs that may be present in indoor air include benzo[*a*]pyrene (CAS 50-32-8) and benzo[*b*]fluoranthene (CAS 205-99-2). As a group, PAHs have high melting and boiling points and very low vapor pressures. Consequently, PAHs exist primarily as or on particles and are typically collected on filters. Concentrations of PAHs in air are generally extremely low (1–100 μg per 1000 m^3 of air) so that very large volumes of air must be sampled to obtain sufficient mass of each PAH for measurement. Although a variety of methods have been used to successfully measure PAHs,[13] the approach of choice is most often extraction from the filter using an appropriate solvent, separation by GC, and measurement with a flame ionization detector (FID) or with an MSD. The latter approach allows confirmation of the identity of the PAHs in the sample.[8]

3.3.3 Pesticides

Organic pesticides are SVOCs that are classified on the basis of use and not molecular structure. Pesticides that are known or suspected carcinogens and that may be found in indoor air include heptachlor (CAS 76-44-8), heptachlor epoxide (CAS 1024-57-3), and hexachlorobenzene (CAS 118-74-1). Pesticides in indoor air are found condensed or adsorbed on to dust particles and in the vapor state. Pesticides in the vapor state are collected on a sorbent, desorbed with a solvent, and analyzed by GC/MSD. Pesticides in airborne dust are collected on a filter and extracted with a solvent. The solvent solution is then concentrated and the resulting concentrate analyzed by GC/MSD.[14] In some applications, an electron-capture detector can provide greater analytical sensitivity for pesticides containing chlorine or fluorine.[8]

4 INORGANIC SPECIES

4.1 Overview

Inorganic carcinogens include some of the heavy metals and their compounds. The heavy metals of primary concern are arsenic (CAS 7440-38-2) (salts, arsenates, and arsenites), beryllium (CAS 7440-41-7), cadmium [oxide (CAS 1306-19-0) fume, bromide (CAS 7789-42-6), and chloride (CAS 10108-64-2)], chromium (hexavalent) (CAS 18540-29-9), nickel [carbonyl (CAS 13463-39-3) and subsulfide (CAS 12035-72-2)], and selenium [sulfide (CAS 7446-34-6)].

4.1.1 Behavior of Ionic Species

Most, but not all, of the inorganic carcinogens listed above exist as salts. Because of the omnidirectional nature of ionic chemical bonds, salts tend to exist as high-melting solids. Such salts are found in air only as a component of SPM.

4.1.2 Oxidation States

The oxidation state of an element in a compound is assigned by a set of rules that allot valence (or outermost) electrons, which are involved in chemical bonding, to the more electronegative (or electron-attracting) element and leave the less electronegative element electron deficient. Thus, the more electronegative element is assigned a negative oxidation state (more electrons assigned than protons in the nucleus), and the less electronegative element is assigned a positive oxidation state (fewer electrons assigned than protons in the nucleus). Metal atoms in compounds are usually assigned positive oxidation states. Oxidation state is a factor in the carcinogenicity of a metal. Hexavalent chromium has an oxidation number of +6 and is classified as a carcinogen, while chromium in other oxidation states (e.g. +3) is not.

4.2 General Monitoring Methods for Inorganic Species

Standard methods have been developed for measuring individual metals in workplace atmospheres. Those

methods should also be appropriate for indoor air. In some cases, more accurate methods of collection and analysis have been reported, but they have not yet been validated or accepted as standard methods.

4.2.1 Collection

Inorganic salts, which exist in air as particulate matter, are collected on filters. Metals are removed from filters by acid digestion and the resulting solution analyzed. Volatile inorganic compounds are usually collected in liquid-filled impingers.

4.2.2 Measurement

Measurement of metals in nonvolatile species is usually accomplished by some type of AAS or AES. More sophisticated instruments, including inductively coupled plasma atomic emission spectroscopy (ICPAES) or inductively coupled plasma mass spectrometry (ICPMS), can measure many different metals in a single analysis. This type of measurement can often be carried out on an aliquot of the sample with no separation of species. Unfortunately, because of chemical changes caused by the rather severe analytical conditions, such approaches measure only the total amount of a metal present and cannot distinguish between different oxidation states of the same metal.

Measurement of a metal in a particular oxidation state, for example, chromium (VI), requires that the species be separated from the matrix by ion chromatography (IC) or some other means, or the species must exhibit some chemical property that is unique among the components of the matrix.

4.3 Inorganic Carcinogens of Particular Concern

The species listed in Table 5 represent inorganic carcinogens that may be present in air and that are of particular concern. Table 5 gives an overview of some published methods for collecting and measuring carcinogenic metals in air and the analyte actually measured.[12,15,16] Brief descriptions of several general methods for metals in air are given below.

Table 5 Measurement of heavy metals in air

Analyte	Method(s)[a]	Collection	Analysis[b]
Arsenic	NIOSH 7900	Cellulose ester membrane filter	AAFA
	NIOSH 7300/IARC 1	Cellulose ester membrane filter	ICPAES
	NIOSH 7901/IARC 2	Sodium carbonate-impregnated cellulose ester filter	AAS
	USOSHA ID-105	Mixed-cellulose ester filter	AAS/HGA
	USOSHA ID-125G	Mixed-cellulose ester membrane filter	ICPAES
Beryllium	NIOSH 7102/IARC 8	Mixed-cellulose ester membrane filter	GFAAS
	NIOSH 7300/IARC 1	Cellulose ester membrane filter	ICPAES
	USOSHA ID-125G	Mixed-cellulose ester membrane filter	ICPAES
Cadmium	NIOSH 7048	Cellulose ester membrane filter	FAAS
	NIOSH 7300/IARC 1	Cellulose ester membrane filter	ICPAES
	USOSHA ID-189	Mixed-cellulose ester membrane filter	AAS/HGA
	USOSHA ID-121	Mixed-cellulose ester filter	FAAS or AES
	USOSHA ID-125G	Mixed-cellulose ester membrane filter	ICPAES
Chromium	NIOSH 7300/IARC 1	Cellulose ester membrane filter	ICPAES
	USOSHA ID-121	Mixed-cellulose ester filter	AAS or AES
	USOSHA ID-125G	Mixed-cellulose ester membrane filter	ICPAES
Chromium (VI)	NIOSH 7600/IARC 4	PVC filter	VIS
	NIOSH 7604	PVC filter	IC
	USOSHA ID-215	PVC filter	IC/UV/VIS
Nickel	NIOSH 7300/IARC 1	Cellulose ester membrane filter	ICPAES
	USOSHA ID-121	Mixed-cellulose ester filter	AAS or AES
	USOSHA ID-125G	Mixed-cellulose ester membrane filter	ICPAES
Nickel carbonyl	NIOSH 6007	Coconut shell charcoal tube (low Ni)	HGAAS
	IARC 5	Alcoholic iodine solution	Calorimetry
	IARC 6	Continuous	Chemiluminescence
Selenium	USOSHA ID-125G	Mixed-cellulose ester membrane filter	ICPAES

[a] See USOSHA,[12] Eller,[15] and O'Neill et al.[16]

[b] AAFA, atomic absorption, flame arsine generation; GFAAS, graphite furnace atomic absorption spectrophotometry; VIS, visible absorption spectrophotometry; NIOSH, National Institute of Occupational Safety and Health; FAAS, flame atomic absorption spectroscopy; HGAAS, graphite furnace atomic absorption spectroscopy; PVC, poly(vinyl chloride); UV/VIS, ultraviolet/visible.

4.3.1 General Method

The US NIOSH has developed a single method (Method 7300) that can be used to measure 29 elements, including all of those addressed here except nickel.[15] Particulates, including the metal salts, are collected on a cellulose ester membrane filter and analyzed by ICPAES.

USOSHA has developed two methods for measuring concentrations of elements, including the carcinogenic metals of concern here, from metal and metalloid particulates in workplace atmospheres. Both USOSHA Methods ID-121 and ID-125G provide for collection of airborne particulates on a mixed-cellulose ester filter. USOSHA Method ID-121, which can be used to measure three of the six carcinogenic metals, requires analysis by AAS; and USOSHA Method ID-125G, which can be used to measure all six carcinogenic metals, requires analysis by AES.[12]

4.3.2 Arsenic

NIOSH Method 7900 measures arsenic and its compounds as total arsenic.[15] The sample is collected on a cellulose ester membrane and analyzed by AAFA. The estimated limit of detection for this method is 0.02 μg per sample.

USOSHA Method ID-105 measures arsenic and its compounds as total arsenic. If volatile inorganic arsenic compounds (arsine, for example) are present, a chemically treated backup pad is added to the usual mixed-cellulose ester filter used for collection of airborne particulates. Arsenic is removed from the collection media by acid-digestion, and the resulting solution analyzed by AAS using a heated graphite atomizer (HGA). The method has a quantitative detection limit of 0.01 $\mu g\,mL^{-1}$, which corresponds to a concentration of 0.0005 $mg\,m^{-3}$ of arsenic in air for a sample volume of 480 L of air and a solution volume of 25 mL.[12]

Total arsenic in air can also be determined by USOSHA Method ID-125G. Airborne particulate matter is collected on a mixed-cellulose ester membrane filter, and arsenic is measured using ICPAES. The nonoptimized quantitative detection limit is 5.7 μg arsenic assuming a solution volume of 50 mL.[12]

4.3.3 Beryllium

Beryllium also occurs in air as a component of SPM, but beryllium compounds are sufficiently volatile to be analyzed by GC, although a pretreatment step offsets some of the advantages of using GC. Beryllium is collected on a membrane filter and can be measured by inductively coupled plasma emission or by atomic absorption–graphite furnace. The limit of detection for the latter, which is used in NIOSH Method 7102, is 0.005 μg per sample.[15]

Beryllium in air can also be measured using USOSHA Method ID-125G. Airborne particulate matter is collected on a mixed-cellulose ester membrane filter, and beryllium is measured using ICPAES. The quantitative detection limit is 0.043 μg beryllium assuming a solution volume of 50 mL.[12]

4.3.4 Cadmium (Oxide, Bromide, and Chloride)

According to NIOSH Method 7048, cadmium compounds in air can be collected on a cellulose ester membrane filter and analyzed by atomic absorption with flame. The estimated limit of detection is 0.05 μg per sample.[15]

Three methods for measuring cadmium in air have been validated by USOSHA. USOSHA Method ID-189 requires collection of SPM on a mixed-cellulose ester membrane filter and analysis by either FAAS or flameless AAS using an HGA. The Method ID-189 quantitative detection limit for cadmium in air is 0.70 $\mu g\,m^{-3}$ for a 200-L sample by FAAS and 0.025 $\mu g\,m^{-3}$ for a 60-L sample by AAS/HGA. USOSHA Method ID-121 requires collection of SPM on a mixed-cellulose ester filter and analysis by AAS or AES. The Method ID-121 analytical detection limit for cadmium in solution is 0.004 $\mu g\,mL^{-1}$. USOSHA Method ID-125G requires collection of SPM on a mixed-cellulose ester membrane filter, and cadmium is measured using ICPAES. The quantitative detection limit is 0.47 μg cadmium assuming a solution volume of 50 mL.[12]

4.3.5 Hexavalent Chromium

Chromium (VI) is usually collected from air on filters of cellulose, polyethylene, polystyrene, PVC, or glass and may be separated from matrix materials by IC. In NIOSH Method 7600, samples are collected on a PVC membrane filter and chromium is measured by VIS. The estimated limit of detection by Method 7600 is 0.05 μg per sample.[15] In USOSHA Method ID-215, SPM is collected on a PVC filter and chromium (VI) is measured by IC with UV/VIS detection. The quantitative detection limit for chromium (VI) by Method ID-215 is 0.0030 $\mu g\,m^{-3}$ in a 960-L air sample.[12]

4.3.6 Nickel (Carbonyl and Subsulfide)

Most compounds of nickel exist in ambient air as a component of SPM and may be collected on a filter. Unfortunately, nickel carbonyl is quite volatile and cannot be collected efficiently on a filter. Volatile compounds of nickel are usually collected in liquid-filled impingers. Analytical methods for nickel include AAS with and without flame, X-ray fluorescence spectrometry, inductively coupled argon plasma spectroscopy, calorimetry, spark

source mass spectrometry, neutron activation analysis, and flame emission spectrophotometry.

USOSHA Methods ID-121 and ID-125G, which are described above, also have been validated to measure total nickel in airborne particulate matter. The analytical detection limit for nickel in solution by Method ID-121 is $0.1\,\mu g\,mL^{-1}$, and the quantitative detection limit by Method ID-125G is 2.0 μg of nickel based on a 50-mL solution volume.[12]

4.3.7 Selenium (Sulfide)

Selenium may be measured by NIOSH Method 7300, which is described above, with an estimated limit of detection of 1 μg per sample.[15] USOSHA Method ID-125G may be used as a screening method for selenium with an estimated quantitative detection limit of 11 μg based on a 50-mL solution volume.[12]

5 ASBESTOS

5.1 Overview

Asbestos (CAS 1332-21-4) is a group of naturally occurring fibrous minerals that are nonflammable and chemically unreactive. Unfortunately, the inertness and chemical stability that made asbestos popular as a component of insulation materials and floor and ceiling tiles are the properties that keep inhaled asbestos fibers intact in lung tissue for decades. Mineral forms of asbestos include actinolite (CAS 13768-00-8), amosite or grunerite (CAS 12172-73-5), anthophylite (CAS 17068-78-9), chrysotile (CAS 12001-29-5), crocidolite (CAS 12001-28-4), and tremolite (CAS 14567-73-8).

5.2 Sampling Airborne Fibers

Fibers are typically collected on a cellulose ester or polycarbonate filter, transferred to a slide, and measured visually under relatively high magnification. The volume of air sampled and the surface area of the filter are used in calculating the concentration of fibers.

5.3 Laboratory Analysis Methods

Asbestos fibers in collected samples are magnified for counting by either phase contrast microscopy (PCM), scanning electron microscopy (SEM), or transmission electron microscopy (TEM). Table 6 gives a comparative summary of the three types of microscopy. The extremely small size of airborne asbestos fibers makes them difficult to distinguish from other natural and man-made fibers. Only TEM offers definitive identification of asbestos fibers.

Table 6 Comparison of methods for measuring asbestos[a]

Method	Specificity for asbestos	Magnification	Sensitivity[b]
PCM	Not specific[c]	400×	0.15 μm (best) 0.25 μm (typical)
SEM	More specific than PCM, but not definitive	1000–2000×	0.05 μm (best) 0.20 μm (typical)
TEM	Definitive (with options)	20 000×	0.0002 μm (best) 0.0025 μm (typical)

[a] See USEPA.[18]
[b] Diameter of thinnest fiber visible.
[c] All fibers with a length > 5 μm and with a length-to-width ratio ≥ 3 : 1 are counted.

5.3.1 Phase Contrast Microscopy

NIOSH Method 7400 describes a procedure for collecting and counting asbestos and other fibers by PCM. Fibers are collected on a 0.45–1.2-μm cellulose ester membrane filter. The volume of air sampled is adjusted to give 100–1300 fibers per square millimeter of surface area on the filter. A section of the filter is placed on a glass microscope slide, and, after some additional sample treatment, positioned under the lens of a positive phase contrast microscope equipped with a Type G-22 Walton-Beckett graticule (100-μm diameter field of view). Magnified fibers in randomly selected grid positions are then manually counted according to rigid counting rules. Fiber concentration (fibers per liter of air) in the sampled air is computed from the average fiber density (fibers per square millimeter of surface area) on the filter, the total effective collection area of the filter (in square millimeters), and the volume of air sampled (in liters).[15] USOSHA Method ID-160 describes a similar procedure.[12]

Although it is used to estimate asbestos concentrations in air, PCM does not allow differentiation between asbestos and other fibers. All fibers and particles meeting the criteria set out in the counting rules (i.e. longer than 5 μm and with a length-to-width ratio greater than or equal to 3 : 1) are counted. Fibers with diameters less than about 0.25 μm are not detected by PCM.

5.3.2 Scanning Electron Microscopy

At present, there is no standard method for the measurement of airborne asbestos using SEM, but procedures for doing so have been developed. Typically, fibers are collected on a 0.4–0.8-μm pore-size polycarbonate (or cellulose ester) filter and carbon-coated directly on the filter. The sample is then transferred to a grid for measurement. The fiber substrate is relatively thick, and the electrons bombarding the specimen during visual analysis

are scattered and reflected rather than transmitted. Only fibers with a diameter of about 0.20 μm or larger can be detected visually by SEM because the scattered and reflected electrons are detected as noise by the microscope. SEM is more specific for asbestos than PCM, but it is not definitive. However, analysis by SEM is much less expensive and more widely available than analysis by TEM.

5.3.3 Transmission Electron Microscopy

NIOSH Method 7402 describes a procedure for measuring asbestos fibers in air by TEM, which is designed to complement results obtained by the PCM method (NIOSH Method 7400) described above.[15] Fibers are collected on the same kind of filter and in the same manner as in the PCM method. Following slide preparation, fibers are identified as asbestos (from fiber morphology, selected area electron diffraction patterns, and energy dispersive X-ray analysis); measured; and counted at a magnification of 20 000×.[17]

Lack of availability, cost, and the time required for analysis are the major disadvantages of TEM analysis. The procedure can be broken down into three levels to reduce the cost and time required for analysis: (1) identification of asbestos for screening purposes; (2) elemental analysis of selected fibers for regulatory action; and (3) quantitative analysis of a few representative fibers for confirmatory analysis.[18]

6 PARTICLES

6.1 Overview

Respirable particles (i.e. particles with aerodynamic diameters of 10 μm or less) represent not only a source of respiratory irritation and allergic response but a pathway for delivery of significant amounts of chemical species, including nonvolatiles, directly to the lungs. Virtually all known or suspected carcinogens can be delivered to the lungs and ultimately to the bloodstream through inhalation of particles containing those species. It is often the chemical species attached to particles, not the particles themselves, that pose the greater threat to human health. Table 7 provides an overview of sampling devices and analytical methods that are used to collect and measure particles in industrial workplaces.[19,20] Only those devices and methods that are most applicable to monitoring human exposure to carcinogens in indoor air are discussed below.

Table 7 Some sampling devices and analysis methods for particles[a]

Device or method	Description
Sampling devices	
Settling chamber	Particles from a trapped sample of air settle by gravity on to a microscope slide
Centrifugal device	Centrifugal force, inside a small cyclone or curved surface trap, is used to separate particles on the basis of size
Impinger	A glass nozzle submerged in a liquid is used to collect particles
Impactor	Particles are collected from an aerosol stream by impaction on to a surface
Filter	Particles are collected by passing an air sample through a filter
Electrostatic precipitator	An ionizing electrode and a collection (or grounded) electrode are used to collect particles
Thermal precipitator	Particles are removed from an aerosol on the basis of response to a temperature gradient
Analysis methods	
Microscopy	Allows previously collected dust particles to be counted, sized, and/or identified
Piezoelectric	Allows real-time monitoring of particulate mass concentration by deposition of particles on to a quartz crystal
Optical	Allows a direct measurement of particles in air on the basis of particle interactions with light
Electrical	Allows direct measurement of particles in air on the basis of the tendency of airborne particles to acquire electrical charge
β-Attenuation	The collected mass of airborne particles is measured by its attenuation of β-radiation passing through it

[a] Based on Lioy and Lioy.[19]

6.2 Sampling Airborne Particles

Particles in air are collected on a filter, usually with a pore size of about 5 μm. Filters may be made from an organic polymer (e.g. Teflon® or PVC) or from glass fibers. The filter is weighed before and after sampling to determine the mass of particles collected. In industrial workplaces, recommended sample volumes range from 1 to more than 1000 L, depending upon the concentration of particles (or the chemical species of interest attached to them) present in the sampled air. Volumes of indoor or residential air that must be sampled should be on the high end of the range, because indoor air tends to be cleaner than industrial workplace air that would be targeted for monitoring.

Human exposure studies generally focus on respirable particles rather than total SPM.[21] A cyclone or a virtual

impactor can be used to limit the size of particles allowed into the sampler. Typical particle-size cuts are 10 μm or smaller and 2.5 μm or smaller.[8]

If the species of interest include VOCs and SVOCs that exist in both gaseous and particle-bound states in air, the sampler may consist of a filter to collect particles followed by a sorbent to collect gaseous species. The filter–sorbent approach works well for measuring total concentrations of such species in air, but it does not allow accurate, separate measurements of gaseous and particle-bound concentrations. During sampling, gaseous semivolatiles from sampled air may be adsorbed by particles already collected on the filter, and particle-bound semivolatiles collected on the filter may be volatilized into the sampled air and swept on to the sorbent. Thus, the particles on the filter may become enriched with some species and depleted of others during sampling.

An approach has been developed for sampling ambient (or outdoor) air that allows separate measurements of gaseous and particle-bound concentrations of semivolatiles. The approach takes advantage of the vast difference in size (actually mass) between molecules and particles. Molecules in air move at speeds that are orders of magnitude faster than even small particles. As a result, molecules diffuse much faster than particles. A diffusion denuder is designed to take advantage of this difference in rates of diffusion to remove organic molecules (but not particles) from sampled air before the air (and particles) reach the filter.[22] A diffusion denuder is an array (usually either a honeycomb or annular configuration) of glass or fused-silica tubes that are coated on interior surfaces with a sorbent. The length of the denuder, the internal cross-sectional area of openings through the denuder, and the flow rate of sampled air determine the residence time of sampled gas within the denuder. The dimensions of the denuder and the sampling flow rate are adjusted to values that allow sufficient residence time within the denuder for gaseous molecules, but not particles, to diffuse to the walls of the denuder, where they are trapped by the sorbent. Particles pass through the denuder and collect on the filter. Target species that volatilize from the collected particles during sampling are collected on a bed of sorbent downstream from the filter. Thus, gaseous material from the sampled air is collected in the denuder, and particle-bound material, on the filter and the sorbent. The fractions are then desorbed and analyzed separately.

6.3 Analyzing Particles and Particle-bound Chemical Species

Typically, total suspended particle concentration is calculated as the mass of particles collected divided by the volume of gas sampled. The same term generally applies to samples collected with or without a particle-size selection device in the sampling system. The particle-size cut points most often used are 10 μm, which allow collection of particles (or particulate matter) with aerodynamic diameters of 10 μm or less (PM_{10}), and 2.5 μm, which allows collection of particles with aerodynamic diameters of 2.5 μm or less ($PM_{2.5}$). The former (PM_{10}) particles are small enough to be inhaled into the lungs, and the latter ($PM_{2.5}$) are small enough to be inhaled into the smallest passageways within the lungs. Obviously, $PM_{2.5}$ is a subset of PM_{10}, but $PM_{2.5}$ also represents the most deeply inhaled particles.

If chemical species associated with collected particulate matter are being measured, the filter, the sorbent, and the denuder, if used, are solvent-desorbed and analyzed according to procedures for organic and inorganic species described elsewhere.

6.4 Particles of Particular Concern

Particles that deliver extremely hazardous or carcinogenic species directly to the lungs over extended periods of time represent a serious threat to human health. Several sources of particles that are found in indoor air and that carry known or suspected carcinogens are discussed below.

6.4.1 Environmental Tobacco Smoke

ETS is a complex mixture of volatile, semivolatile, nonvolatile, and particulate matter that may contain more than 3800 chemical species, including heavy metals. Many of the organic and inorganic species discussed above are present in tobacco smoke. The composition and distribution of the components of ETS change continuously as compounds are adsorbed on to or desorbed from particles and as the more reactive compounds are converted into other species.

Because of the complexity of the mixture and the fact that some of the compounds in ETS also come from other sources, exposure to ETS is quite difficult to measure directly. Typically, respirable suspended particle concentration or the concentration of nicotine (CAS 54-11-5),[8] an organic compound that is specific to ETS, is measured by procedures described in previous sections for particles and organic compounds.

Yet another way to assess exposure to ETS is through analysis of physiological fluids of exposed persons. In this case, exposed persons have become the collection medium. The presence of nicotine or its major metabolite, cotinine, in saliva, blood, or urine is entirely due to exposure to tobacco, tobacco smoke, or ETS. The measurement of nicotine and cotinine in physiological fluids of active smokers or persons exposed to ETS

(sometimes called passive smokers because they are exposed to ETS but do not smoke themselves) is generally accomplished by GC with a nitrogen-sensitive detector. The GC method can be used to measure nicotine concentrations as low as 1 ng mL^{-1} of fluid and concentrations of cotinine as low as 6 ng mL^{-1}.

6.4.2 Particles from Cooking and Heating Activities

Soot particles from poorly vented heaters and oily mist and smoke from cooking can contain PAHs and other carcinogens. As described in the organic compounds section, PAHs are present in both particle-bound and gaseous states. The particles are collected as described above, and the PAHs are solvent-desorbed from the collection media and are typically measured by GC with an appropriate detector.

6.4.3 Other Respirable Particles

Any kind of respirable particle, whatever the source, has the potential to introduce carcinogens or other harmful species directly to the lungs and bloodstream of exposed persons.

7 REMARKS

Monitoring of carcinogens in indoor air is obviously not a simple task. The methods described here have, for the most part, been thoroughly evaluated by many different laboratories. A steady stream of newer, more sensitive techniques and instruments is appearing in the environmental literature and in the marketplace. As an example, a recent journal article reported the measurement, by an automated analysis, of 1 pg of a VOC, which corresponds to less than 1 part per trillion of the compound in a 1-L air sample.[23] Such a feat would have been impossible only a few years ago. Unfortunately, so-called allowable concentrations of some toxic and carcinogenic species are still defined by the sensitivity limitations of the methods available to measure them rather than by the risk those species pose to human health.

ABBREVIATIONS AND ACRONYMS

AAFA	Atomic Absorption, Flame Arsine Generation
AAS	Atomic Absorption Spectroscopy
AES	Atomic Emission Spectroscopy
ALOQ	Analytical Limit of Quantitation
ETS	Environmental Tobacco Smoke
FAAS	Flame Atomic Absorption Spectroscopy
FID	Flame Ionization Detector
GC	Gas Chromatography
GFAAS	Graphite Furnace Atomic Absorption Spectrophotometry
HGA	Heated Graphite Atomizer
HGAAS	Graphite Furnace Atomic Absorption Spectroscopy
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
MLOQ	Method Limit of Quantitation
MSD	Mass Spectrometric Detector
NIOSH	National Institute of Occupational Safety and Health
PAH	Polycyclic Aromatic Hydrocarbon
PCM	Phase Contrast Microscopy
PDA	Photodiode Array
POM	Particulate Organic Matter
PQL	Practical Quantitation Limit
PVC	Poly(vinyl chloride)
SEM	Scanning Electron Microscopy
SIM	Single-ion Monitoring
SPM	Suspended Particulate Matter
SVOC	Semivolatile Organic Compound
TEM	Transmission Electron Microscopy
TLD	Thermoluminescent Dosimeter
USOSHA	US Occupational Safety and Health Administration
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible Absorption Spectrophotometry
VOC	Volatile Organic Compound
VVOC	Very Volatile Organic Compound

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Asbestos Analysis • Formaldehyde, Environmental Analysis of • Gas Chromatography with Selective Detectors for Amines • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Field-portable Instrumentation **(Volume 5)**
Radon, Indoor and Remote Measurement of

Industrial Hygiene **(Volume 5)**
Industrial Hygiene: Introduction • Aerosols and Particulates Analysis: Indoor Air • Chromatographic Techniques in Industrial Hygiene • Direct Reading Instruments for the Determination of Aerosols and Particulates • Dust, Measurement of Trace Elements in

Industrial Hygiene cont'd **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure • Parent and Progeny Compounds in Exhaled Breath, Determination of • Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air • Spectroscopic Techniques in Industrial Hygiene

Particle Size Analysis **(Volume 6)**
Optical Particle Counting

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Surfaces **(Volume 10)**
Scanning Electron Microscopy in Analysis of Surfaces

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Gas Chromatography/Mass Spectrometry

Radiochemical Methods **(Volume 14)**
Speciation of Radionuclides in the Environment

REFERENCES

1. M.R. Peterson, R.K.M. Jayanty, 'Carcinogens, Monitoring Indoor Air', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley & Sons, New York, 929–940, 1998.
2. K.W. Gold, D.F. Naugle, M.A. Berry, 'Indoor Concentrations of Environmental Carcinogens', in *Environmental Carcinogens Methods of Analysis and Exposure Measurement*, eds. B. Seifert, H.J. van de Wiel, B. Dodet, I.K. O'Neill, International Agency for Research on Cancer, Lyon, France, 41–71, Vol. 12, 1993.
3. J.E. Yocom, S.M. McCarthy, *Measuring Indoor Air Quality – A Practical Guide*, John Wiley & Sons, New York, 1991.
4. W.W. Nazaroff, 'Measurement Techniques', in *Radon and Its Decay Products in Indoor Air*, eds. W.W. Nazaroff, A.V. Nero, Jr, John Wiley & Sons, New York, 491–504, 1988.
5. C.R. Cothern, J.E. Smith, Jr (eds.), *Environmental Radon*, Plenum Press, New York, 1987.
6. R.C. Fortmann, 'Measurement Methods and Instrumentation', in *Radon: Prevalence, Measurements, Health Risks and Control*, ASTM Manual Series, ed. N.L. Nagda, ASTM Philadelphia, 53, 1994.
7. M.R. Peterson, D.F. Naugle, M.A. Berry, 'Methods of Analysis for Environmental Carcinogens', in *Environmental Carcinogens Methods of Analysis and Exposure Measurement*, eds. B. Seifert, H.J. van de Wiel, B. Dodet, I.K. O'Neill, International Agency for Research on Cancer, Lyons, France, 126–135, Vol. 12, 1993.
8. W.T. Winberry, Jr, L. Forehand, N.T. Murphy, A. Ceroli, B. Phinney, A. Evans, *Compendium of Methods for the Determination of Air Pollutants in Indoor Air*, US Environmental Protection Agency, Research Triangle Park, 1990. Also published as *Methods for Determination of Indoor Air Pollutants: EPA Methods*, Noyes Data Corporation, Park Ridge, 1992.
9. A.T. Hodgson, J.R. Girman, 'Application of a Multisorbent Sampling Technique for Investigations of Volatile Organic Compounds in Buildings', in *Design and Protocol for Monitoring Indoor Air Quality*, ASTM STP 1002, eds. N.L. Nagda, J.P. Harper, ASTM, Philadelphia, 244–256, 1989.
10. M.R. Peterson, D.F. Naugle, M.A. Berry, 'Suggested Methods of Analysis for Indoor Air Environmental Carcinogens', in *Sick Building Syndrome – Sources, Health Effects, Mitigation*, eds. M.C. Baechler, D.L. Hadley, T.J. Marseille, R.D. Stenner, M.R. Peterson, D.F. Naugle, M.A. Berry, Noyes Data Corporation, Park Ridge, 291–328, 1991.

11. M. Gavin, D.R. Crump, V.M. Brown, 'Appropriate Sampling Strategies for the Measurement of Formaldehyde in Indoor Air', *Environ. Technol.*, **16**(6), 579–586 (1995).
12. OSHA, *OSHA Analytical Methods Manual*, American Conference of Governmental Industrial Hygienists, Cincinnati, 1999. Methods may be downloaded from http://www.osha-slc.gov/ftp/Analytical_Methods (May 1999).
13. H. Egan, M. Castegnaro, P. Bogovski, H. Kunte, E.A. Walker (eds.), *Environmental Carcinogens Selected Methods of Analysis, Volume 3 – Analysis of Polycyclic Aromatic Hydrocarbons in Environmental Samples*, International Agency for Research on Cancer, Lyons, France, 1979.
14. K.S. Roinestad, J.B. Louis, J.D. Rosen, 'Determination of Pesticides in Indoor Air and Dust', *J. AOAC Int.*, **76**(5), 1121–1126 (1993).
15. P.M. Eller (ed.), *NIOSH Manual of Analytical Methods*, 4th Edition, US Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, Division of Physical Sciences and Engineering, Cincinnati, Ohio, 1994. Methods may be downloaded from http://www.cdc.gov/niosh/nmam/nmammenu.html (May 1999).
16. I.K. O'Neill, P. Schuller, L. Fishbein, *Environmental Carcinogens Selected Methods of Analysis, Volume 8 – Some Metals: As, Be, Cd, Cr, Ni, Pb, Se, Zn*, International Agency for Research on Cancer, Lyon, France, 1986.
17. N.E. Clark, D.K. Verma, J.A. Julian, 'Determination of the Variability in Elemental Composition of Asbestos Fibres by Analytical Transmission Electron Microscopy', *Ann. Occup. Hyg.*, **39**(1), 79–88 (1995).
18. USEPA, *Measuring Airborne Asbestos following an Abatement Action*, EPA 600/4-85-049, US Environmental Protection Agency, Research Triangle Park, NC, 1985.
19. P.J. Lioy, M.J.Y. Lioy (eds.), *Air Sampling Instruments for the Evaluation of Atmospheric Contaminants*, 6th edition, American Conference of Governmental Industrial Hygienists, Cincinnati, 1983.
20. R.K.M. Jayanty, M.R. Peterson, D.F. Naugle, M.A. Berry, 'Exposure Assessment: Methods of Analysis for Environmental Carcinogens', *Risk Analysis*, **10**(4), 587–595 (1990).
21. L. Wallace, 'Indoor Particles: A Review', *J. Air Waste Manage. Assoc.*, **46**(2), 98–126 (1996).
22. U. Risse, E. Flammenkamp, A. Kettrup, 'Denuder Sampling System for the Determination of Gaseous and Particle-bound Organic Health Hazardous Substances', *Fresenius' J. Anal. Chem.*, **350**(7/9), 454–460 (1994).
23. E. Woolfenden, 'Monitoring VOCs in Air Using Sorbent Tubes Followed by Thermal Desorption–Capillary GC Analysis: Summary of Data and Practical Guidelines', *J. Air Waste Manage. Assoc.*, **47**(1), 20–36 (1997).

Charged Particle Activation Analysis

Karel Strijckmans
Laboratory Analytical Chemistry, Institute Nuclear Sciences, Gent University, Belgium

1 Introduction **2**
2 Interaction of Charged Particles with Matter **2**
2.1 Stopping Power 2
2.2 Range 2
2.3 Stopping Power and Range Data 3
3 Charged Particle Induced Nuclear Reactions **3**
3.1 *Q*-value and Threshold Energy 3
3.2 Coulomb Barrier 4
3.3 Nuclear Reaction Cross-section 4
3.4 Cross-section Data: Excitation Function and Activation Curve 4
3.5 Radionuclides Formed by Charged Particle Induced Reactions 5
4 Interferences **6**
4.1 Nuclear Interference 6
4.2 Spectral Interference 7
4.3 Matrix Interference 7
5 Standardization **7**
5.1 General 8
5.2 Trace Element Determination in the Bulk of a Sample 9
5.3 Surface Characterization of Thin Layers 10
5.4 Conclusions 10
6 Charged Particle Irradiation **10**
6.1 Accelerator 10
6.2 Heat Release in a Target 11
6.3 Target Holder 12
6.4 Direct Beam Intensity Monitoring 12
6.5 Quantitative Beam Intensity Monitoring 13
7 Activity Measurement **14**
7.1 Radionuclides Formed and Decay Radiation Measured in Charged Particle Activation Analysis 14
7.2 γ-Spectrometry 14
7.3 Positron Counting 15
8 Procedure **16**
8.1 General 16
8.2 Trace Elements in the Bulk of a Sample 17
8.3 Surface Characterization of Thin Layers 18
8.4 Radiochemical Separation 18
9 Data Handling **20**
9.1 γ-Spectrum Analysis 20
9.2 Decay Curve Analysis 20
9.3 Quantitation 21
10 Perspectives **21**
10.1 Charged Particle Activation Analysis 21
10.2 Charged Particle Activation for the Development of Other Analytical Methods 21
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Charged particle activation analysis (CPAA) is an analytical method for elemental analysis. It is based on charged particle (CP) induced nuclear reactions producing radionuclides, that are identified and quantified by their characteristic decay radiation. CPAA allows trace element determination in the bulk of a solid sample, as well as characterization of a thin surface layer, i.e. determination of mass thickness or composition.

Advantages of CPAA are: (1) good precision, (2) low detection limit, (3) outstanding accuracy, (4) absolute, i.e. reference method, (5) independent, i.e. based on a totally different principle than more common methods for elemental analysis (atomic/optical emission and mass spectrometry), and hence not subject to the same systematic errors, (6) traceability, (7) not subject to surface contamination for trace element determination in the bulk, (8) if instrumental analysis is not possible it is not subject to reagent blank errors, and (9) errors due to nonquantitative yield (and even nonreproducible yield) can be corrected for. Disadvantages of CPAA are its inherent complexity and costs, it is not suitable for liquid samples, and heating occurs during irradiation. Speciation, depth profiling, or scanning is not possible.

This article covers (1) the principles on the slowing down of CPs in matter, nuclear reactions, and interferences; (2) experimental details concerning irradiation and activity measurement; (3) conceptual and experimental procedures and data handling, from the problem to the final result; (4) perspectives. References to relevant databases are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

made. A systematic overview of all possible CPAA applications is beyond the scope of this article.

1 INTRODUCTION

CPAA is an analytical method for elemental analysis. Speciation (i.e. discrimination between different oxidation states of the analyte element or between different compounds) is *not* possible. CPAA is based on CP induced nuclear reactions producing radionuclides and their characteristic decay radiation is measured. The radionuclide is identified by measuring its energy and/or half-life, i.e. qualitative analysis. Quantitative analysis is performed by measuring the number of particles or photons emitted, i.e. radioactivity.

CPAA provides a good precision, low detection limits, and accuracy. It can be considered as an absolute (reference) and an independent method for elemental analysis of solid samples. There is no need for standards that are comparable to the samples, or for standards where the analyte content is determined by another analytical method. Traceability conditions are fulfilled, except for the stopping power correction. CPAA is based on a totally different principle than more common methods for elemental analysis (e.g. atomic or optical emission and mass spectrometry), and hence is not subject to the same systematic errors. It is not subjected at all to errors related to dissolution of the sample and chemical separation (e.g. preconcentration). If instrumental analysis is not possible, then radiochemical CPAA can be applied. Errors due to reagent blanks do not exist. Errors due to surface contamination (e.g. determination of bulk oxygen in aluminum) can be avoided easily. Finally, errors due to nonquantitative yield, and even nonreproducible yield, can be corrected for.

CPAA can be applied for trace element determinations as well as for characterization of thin surface layers. It has proven its outstanding capabilities for the determination of trace light elements (boron, carbon, nitrogen, and oxygen) in the bulk of high purity metals and semiconductors. Recently, CPAA has been applied for the characterization of surface layers, such as the determination of mass thickness or composition (stoichiometry). The disadvantages of CPAA are its inherent complexity and cost, it is not suitable for liquid samples, and heating occurs during irradiation. Neither depth profiling nor scanning is possible for surface characterization.

2 INTERACTION OF CHARGED PARTICLES WITH MATTER

Most CPs do not induce nuclear reactions and are slowed down in matter. The energy loss is described by the stopping power, and the penetration depth in matter by the range.

2.1 Stopping Power

The stopping power of a target for a CP is the energy loss of the CP for an amount of target atoms. In CPAA it is most convenient to use mass stopping power, i.e. the energy loss per unit mass thickness, as given in Equation (1).

$$S = \frac{-\mathrm{d}E}{\mathrm{d}x} \tag{1}$$

where S is the mass stopping power, in $\mathrm{MeV\,g^{-1}\,cm^2}$; E is the energy, in MeV; and $x = \rho l$, the mass thickness, in $\mathrm{g\,cm^{-2}}$ (ρ = mass density in $\mathrm{g\,cm^{-3}}$ and l = thickness in cm).

The stopping power is determined by the nature of the CP, energy of the CP, and nature of the target, as given in Equation (2).

$$S_\mathrm{a} = Z_\mathrm{a}^2 S_\mathrm{p} \left(\frac{E_\mathrm{a}}{A_\mathrm{a}} \right. \tag{2}$$

where Z_a is the atomic number and A_a is the mass number. The subscripts ''a'' and ''p'' refer to a CP and proton, respectively. Figure 1(a) and Equation (2) show that the stopping power for 20-MeV deuterons is equal to the one for 10-MeV protons, and the stopping power for 30-MeV helium-3 particles and 40-MeV alpha particles is four times that of 10-MeV protons. The stopping power decreases with the CP energy ($\sim 1/E$) (Figure 1b) and for increasing atomic number of the target (Figure 1c). This decrease is steep for low-Z targets.

Stopping power data for elemental matter can be calculated based on the Bethe formula. For mixtures, the additivity rule of Bragg and Kleeman (Equation 3) is applied.

$$S = \sum_{i=1}^{n} w_i S_i \tag{3}$$

where w_i is the mass fraction of the ith component, S_i is the mass stopping power of the ith component, and n is the number of components. For compounds, deviations from the Bragg rule may occur due to chemical binding effects. For the energy range of interest in CPAA (>1 MeV) these deviations can be neglected.

2.2 Range

As a consequence of the stopping power phenomenon, CPs lose their energy and are stopped at a depth called the ''range''. From Equation (1) the mass range, R, for a

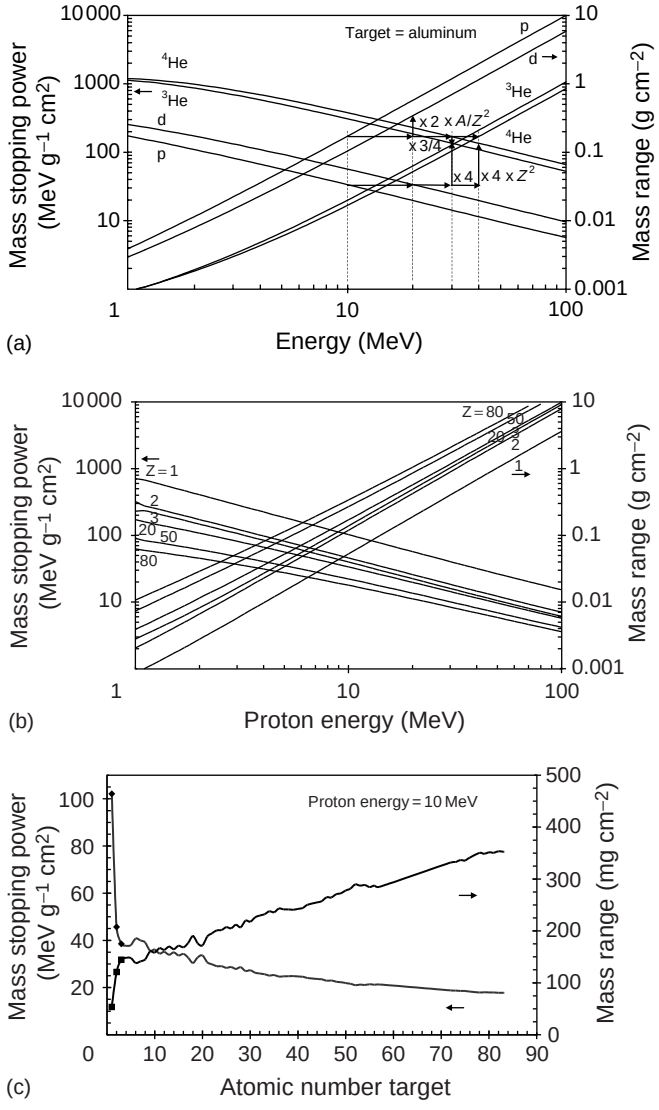

Figure 1 Mass stopping power and mass range for: (a) protons, deuterons, helium-3, and α-particles in an aluminum target as a function of the energy; (b) for elemental targets with atomic numbers $Z = 1, 2, 3, 20, 50$, and 80 as a function of the proton energy; and (c) for 10-MeV protons as a function of the atomic number of an elemental target.

CP with incident energy E_i can be calculated according to Equation (4).

$$R = \int_0^R \mathrm{d}x = \int_{E_i}^0 \frac{\mathrm{d}E}{\mathrm{d}E/\mathrm{d}x} = \int_0^{E_i} \frac{\mathrm{d}E}{S} \tag{4}$$

where R is the mass range, in $\mathrm{g\,cm^{-2}}$; and E_i is the incident CP energy, in MeV. Consequently the range is also determined by the nature of the CP, the energy of the CP, and the nature of the target, as given in Equation (5).

$$R_a = \frac{A_a}{Z_a^2} R_p \left(\frac{E_a}{A_a} \right. \tag{5}$$

Figure 1(a) and Equation (5) show that the range of 20-MeV deuterons is twice that of 10-MeV protons, the range of 30-MeV helium-3 particles is 75% that of 10-MeV protons, and the range of 40-MeV α-particles is equal to that of 10-MeV protons. The range increases with the CP energy ($\sim E^2$) (Figure 1b) and for increasing atomic number of the target (Figure 1c). The increase is steep for low-Z targets.

For CPAA the CP interaction with matter can be described simply: CPs traverse the target undeflected (i.e. nearly straight lines), the beam intensity (the number of CPs per unit time) is almost unchanged, and all CPs are stopped at the same depth in the target (at the range). Consequently, linear versus projected range, and range straggling, should not be considered for CPAA.

2.3 Stopping Power and Range Data

Stopping power and range data can be obtained from the SRIM (Stopping and Range of Ions in Matter) PC software package of Ziegler, IBM Research.[1] The principles are described by Ziegler et al.[2] To calculate ranges expressed in length (as in SRIM) the mass density is required. For mixtures and compounds (not listed in the compound dictionary) the mass density should be introduced, by the user, into the SRIM package. (An updated review of *Nuclear Data for Charged Particle Activation* is available at http://allserv.rug.ac.be/~kstrykmn/nuclear.)

3 CHARGED PARTICLE INDUCED NUCLEAR REACTIONS

CPAA is based on a nuclear reaction $\mathrm{A + a \longrightarrow B + b + } Q$, or A(a,b)B, where A is the stable target nuclide (at rest), "a" is a CP (accelerated), B is a radionuclide, "b" is the particle(s) or photon emitted, and Q is the Q-value. Two important characteristics are the minimum CP energy required to induce such a nuclear reaction, and the probability that this reaction will proceed. As CPs are slowed down when traversing matter, as described in section 2, the latter characteristic should be known as a function of the CP energy.

3.1 *Q*-value and Threshold Energy

The Q-value is the energy released by one single nuclear reaction. It is related to the difference between the resting masses of reactants and products, as given in Equation (6).

$$Q = (m_A + m_a - m_B - m_b)c^2 \tag{6}$$

where m is the nuclidic mass and c is the velocity of light. Nuclidic masses can be replaced by atomic masses,

as electron masses are cancelled, at least if this also is done for ''a'' and ''b'' (e.g. for a (p,α) reaction the atomic masses of ^{1}H and ^{4}He, respectively, are used). A nuclear reaction can be endoergic ($Q < 0$) or exoergic ($Q > 0$). To induce endoergic reactions the CPs should have a minimum energy, i.e. the threshold energy E_t, which is slightly higher than $-Q$. The compound nucleus, formed by collision of the CP ''a'' and the target nucleus A, retains a fraction of the kinetic energy of the CP, i.e. $m_a/(m_A + m_a)$. This fraction is ''lost'' to compensate for the shortfall of mass or energy for an endoergic reaction.

Approximating the nuclidic mass by the mass number A for the recoil energy correction, the threshold energy is given by Equation (7).

$$E_t = -\frac{A_A + A_a}{A_A} Q \qquad \text{for } Q < 0$$

$$E_t = 0 \qquad \text{for } Q \geq 0 \tag{7}$$

where E_t is the threshold energy. For exoergic nuclear reactions the threshold energy is zero by definition.

Q-values and threshold energies can be calculated from the Audi and Wapstra ''adopted masses'' 1993[3] or 1995[4] with the on-line program Q-calc, at the Lund Nuclear Data WWW Service (Lund University, Sweden), at the National Nuclear Data Center (NNDC, Brookhaven National Laboratory, USA), and at the International Atomic Energy Agency (IAEA, Austria),[5] or as a Fortran-77 program[6] for a VAX-VMS system.

3.2 Coulomb Barrier

The Coulomb barrier also determines the minimum CP energy needed to induce a nuclear reaction. The coulombic repulsive force between the target nucleus and the CP dominates at ''large'' distances, and increases when the CP approaches the target nucleus. At some particular distance (i.e. the sum of the radii of CP and target nucleus) the attractive nuclear force balances the Coulomb repulsive force. The decrease in the CP kinetic energy is given by the Coulomb barrier, as in Equation (8).

$$E_C \approx \frac{A_A + A_a}{A_A} \frac{Z_A Z_a}{\sqrt[3]{A_A} + \sqrt[3]{A_a}} \tag{8}$$

where E_C is the Coulomb barrier, in MeV; and $\sqrt[3]{A}$ is proportional to the nucleus radius. The Coulomb barrier for ^{3}He and ^{4}He particles is about twice as high as that for protons or deuterons, and it increases with the atomic number of the target nucleus, roughly $Z_A^{2/3}$, as shown in Figure 2.

The kinetic energy – the CP lost by the Coulomb barrier – is released again when a nuclear reaction occurs. Consequently, the Coulomb barrier influences the energetics of a nuclear reaction only in that a CP must have a kinetic energy higher than the Coulomb barrier before the reaction can occur. Quantum mechanical treatment of the problem explains that the reaction probability (see section 3.3) for CPs with a kinetic energy lower than the Coulomb barrier is not exactly zero, but very low, and rapidly drops as the CP energy decreases.

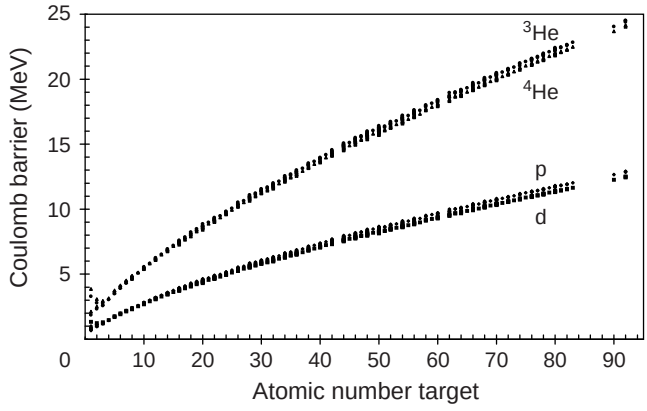

Figure 2 Coulomb barrier of all stable nuclides for proton, deuteron, helium-3, and helium-4 particle induced reactions as a function of the atomic number of the target nucleus.

3.3 Nuclear Reaction Cross-section

The probability of a nuclear reaction is expressed as the (nuclear reaction) cross-section, which has the dimensions of area. This originates from the simple picture that the probability for a reaction between the target nucleus and the incident CP is proportional to the geometric cross-section that the target nucleus presents to a beam of CPs. The average cross-section is somewhere near $10^{-28}\,m^2$. The barn ($1\,b = 10^{-28}\,m^2$) is used as a unit for nuclear reaction cross-section.

The cross-section for a particular reaction (also called the partial reaction cross-section) depends on the energy of the CP. It is zero for CPs below the threshold energy of that reaction. If the Coulomb barrier is higher than the threshold energy (i.e. always the case for exoergic reaction, as $E_t = 0$ for $Q > 0$), then the cross-section is very low in the energy interval between the threshold energy and the Coulomb barrier. For CP energies exceeding the threshold energy and Coulomb barrier, the cross-section increases up to a maximum (typically 1 b), and decreases as more complex reactions become competitive. Two annotated examples are given in Figure 3(a) and (b).

3.4 Cross-section Data: Excitation Function and Activation Curve

As CPs are slowed down in matter, knowledge of the cross-section as a function of the energy is required

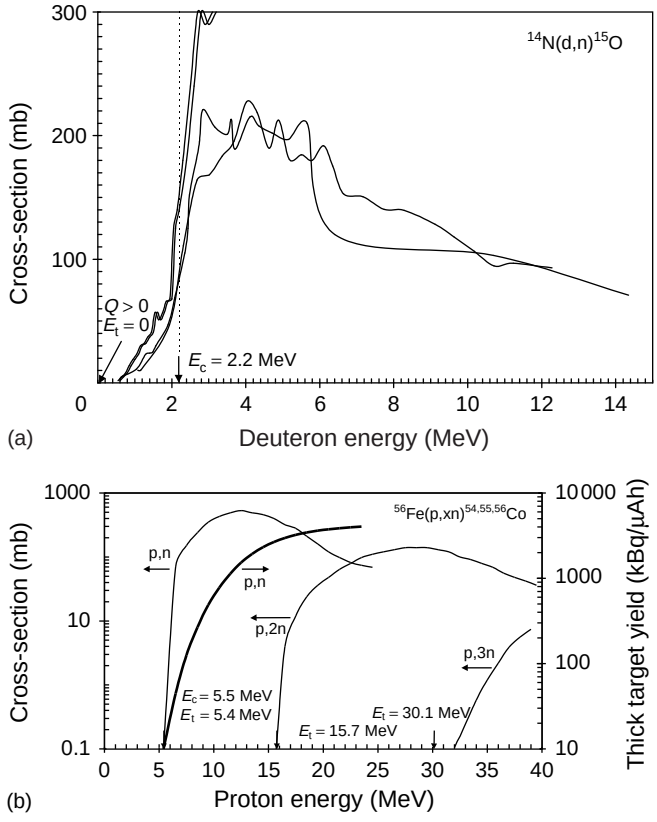

Figure 3 Excitation functions. (a) The $^{14}N(d,n)^{15}O$ reaction is exoergic ($Q > 0$), consequently its threshold energy is zero. Below the Coulomb barrier (2.2 MeV) the cross-section is low, but not zero. It increases steeply around the Coulomb barrier energy. (b) The $^{56}Fe(p,n)^{56}Co$ reaction is endoergic ($Q = -5.3$ MeV), hence its threshold energy equals 5.4 MeV. The Coulomb barrier equals 5.5 MeV. The cross-section is zero below 5.4 MeV, increases to a maximum at 12–13 MeV, and decreases at higher energies, as the $^{56}Fe(p,2n)^{55}Co$ reaction ($E_t = 15.7$ MeV) becomes competitive. For the (p,2n) reaction the reaction cross-section also increases up to a maximum at 27–30 MeV, and decreases at higher energies, as the $^{56}Fe(p,3n)^{54}Co$ reaction ($E_t = 30.1$ MeV) becomes competitive. For thick target yield (TTY), see section 3.5.

for CPAA. However, absolute data are not required (sections 5.2 and 5.3). Therefore the *excitation function* (or curve) indicates the absolute cross-section as a function of the energy, while the *activation curve* indicates a relative data set.

Experimental excitation functions can be obtained from the EXFOR (or CSISRS) database at NNDC mirrored at IAEA and the Nuclear Energy Agency (NEA), France.[7] For the data sets in Figure 3(a), obtained from this database, the agreement is fairly good, but this is *not* representative for the vast majority of CP induced reactions. Actually the $^{14}N(d,n)^{15}O$ reaction has been studied in great detail, because it is extremely important for positron emission tomography (PET), an imaging technique for quantifying the oxygen metabolism, blood flow and blood volume in nuclear medicine.

If no experimental data are available, excitation functions can be estimated using the very simple equations and data from Keller et al.[8]

The experimental procedure needed to obtain an activation curve is quite simple if the target element (A) is available as "thin" foil. "Thin" means much smaller than the range. A stack of "thin" foils (total thickness $\geq$ range) is irradiated, and the foils are measured individually. After decay correction the count rate measured is proportional to the cross-section, as will be shown in Equation (17). Excitation function data require absolute activity measurements, absolute beam intensity monitoring, and absolute energy measurement of the incident CP beam. However, an activation curve (relative cross-section data) is sufficient for CPAA.

In contrast to neutron induced reactions, no experimental data are available for most of the CP induced reactions. Additionally, if different data sets are available, their agreement is often unsatisfactory. Therefore approximate methods for standardization have been developed in CPAA.

3.5 Radionuclides Formed by Charged Particle Induced Reactions

Which radionuclides are formed by irradiation of a particular matrix by a CP with a particular energy, and what is their approximate activity level?

The first question can be answered by observing a chart of nuclides. This represents all nuclides with relevant data for CPAA as chemical symbol, mass number, isotopic abundance (stable nuclides), half-life, decay mode, and decay energies (radionuclides). For proton irradiation, for example, the reaction types given in Figure 4 should be applied to all stable isotopes of all elemental components of the matrix. In the inventory of possible nuclear reactions two categories can be rejected: those leading to stable nuclides, and those with a threshold energy higher than the CP energy. Reactions for which the Coulomb barrier is higher than the CP energy induce only low activity levels, but cannot be excluded. An annotated example is given in Figure 4 for a 10-MeV proton irradiation of nitrogen.

The induced activity is proportional to the number of CPs per time unit, irradiation time (not linearly), number of target nuclides, and reaction cross-section. The number of target nuclides is inversely proportional to the relative atomic mass of the element, and proportional to the mass fraction of the element in the target, the isotopic abundance of the target nuclide, and the activated depth, i.e. the range of the target at the incident energy minus

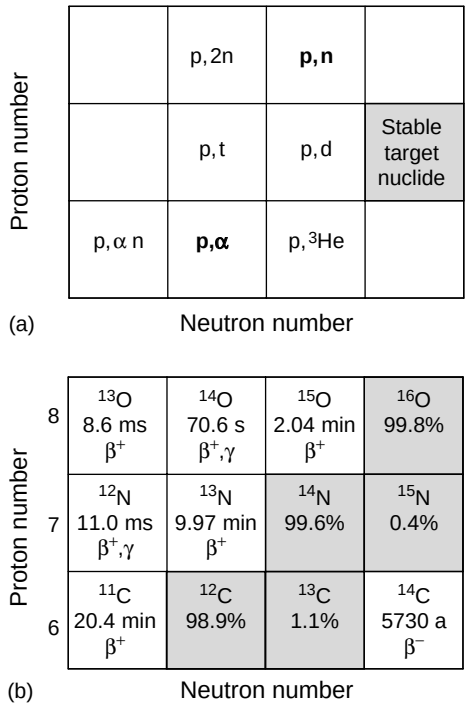

Figure 4 (a) Reaction scheme for proton irradiation to be applied to a chart of nuclides. In general, for a proton energy below 10 MeV only the (p,n) and (p,α) reactions have to be considered; for a proton energy up to 20 MeV the (p,2n), (p,d), (p,t), (p,^{3}He), and (p,αn) reactions also have to be taken into account. (b) Part of a chart of nuclides: irradiation of nitrogen by 10-MeV protons mainly induces ^{14}O and ^{11}C by a (p,n) and (p,α) reaction, respectively, from ^{14}N, i.e. the most abundant stable isotope of nitrogen, and to a negligible extent ^{15}O by a (p,n) reaction from ^{15}N. The (p,α) reaction has not been considered for CPAA as it produces ^{12}C, a stable nuclide.

the range at the threshold energy or the Coulomb barrier, whichever is higher. This will be treated in more detail in section 5.1. A rough estimate of the induced activity, i.e. the thick target yield (TTY), can be calculated from Equation (9).

$$A \approx 380 \frac{1}{Z_a} I \frac{1}{A_r} w\theta (1 - e^{-\lambda t_i}) \overline{\sigma} \left(R_{E_i} - R_{E_{t/C}} \right) e^{-\lambda t_w} \quad (9)$$

where

A = activity, in kBq,
Z_a = atomic number of the CP (1 for p and d; 2 for ^{3}He and ^{4}He (α-particles)),
I = beam intensity, in μA,
A_r = relative atomic mass of the target element,
w = mass fraction of the target element in the matrix, in %,
θ = isotopic abundance, in atom %,
λ = $\ln(2)/t_{1/2}$, decay constant of the radionuclide (where $t_{1/2}$ is the half-life),
t_i = irradiation time,
t_w = waiting time,
$\overline{\sigma}$ = average reaction cross-section, in mb (generally, $100 < \overline{\sigma} < 1000$),
R = mass range, in g cm^{-2} (section 2.2 and Figure 1),
E_i = incident energy,
$E_{t/C}$ = max(E_t, E_C) (sections 3.1 and 3.2).

An example of the TTY as a function of the incident energy is given in Figure 3(b) for the (p,n) reaction.

A chart of nuclides is available as:

- a wall chart or booklet,
  - Kernforshungszentrum Karlsruhe (Germany), (see reference 9)
  - General Electric & KAPL, Knoll Atomic Power Lab. (USA),[10]
- a PC program,
  - NuChart from IAEA (1994), no further updates,[11]
  - Nucleus from AMDC, Atomic Mass Data Center (France),[12]
  - MacNuclide from Center for Nuclear Information Technology (for Mac),[13]
- a PC program in combination with a database on the internet, Isotope Explorer (formerly VuENSDF) from LBNL, Lawrence Berkeley National Laboratory (USA) and Lund University (Sweden),[14]
- on-line databases on the internet,
  - Lund Nuclear Data WWW Service from Lund University (Sweden),[15]
  - jvNuBase from AMDC, Atomic Mass Data Center (France),[16]
  - KAERI Table of the Nuclides from KAERI, Korea Atomic Energy Research Institute (also mirrored).[17]

4 INTERFERENCES

The determination of an analyte element is based on the reaction A(a,b)B, where A is a stable nuclide of that analyte element and B is a radionuclide (section 3). Three different types of interference can be distinguished: nuclear, spectral, and matrix.

4.1 Nuclear Interference

A radionuclide B is also formed from an element other than the analyte element by a nuclear reaction C(c,d)B. C is a stable nuclide of the interfering element.

The CP "c" is mostly identical to "a", but nuclear interferences induced by secondary particles, e.g. fast or thermalized neutrons produced by (a,n) reactions, have to be considered also. Nuclear reactions can only be avoided by a proper choice of the CP energy if the threshold energy E_t of the interfering reaction C(c,d)B exceeds both the threshold energy and the Coulomb barrier of the analyte reaction A(a,b)B. To obtain high sensitivity, the CP energy is chosen at or just below the threshold energy of the interfering reaction.

Interference by the $^{14}N(d,t)^{13}N$ reaction (Figure 4b) could affect the determination of carbon by the $^{12}C(d,n)^{13}N$ reaction. As the threshold energy of the interfering reaction (4.9 MeV) exceeds that of the analyte reaction (0.3 MeV), as well as the Coulomb barrier (2.0 MeV), carbon can be determined sensitively and interference free by irradiation with 4.9-MeV deuterons. It is obvious that the difference in Coulomb barrier is largely insufficient to avoid nuclear interferences, and that the reaction $^{13}C(d,2n)^{13}N$ does not interfere with the determination of carbon.

4.2 Spectral Interference

A radionuclide D is formed by the nuclear reaction C(c,d)D and the measurement does not resolve the activity of B and D. Spectral interferences can be avoided by:

(1) The proper choice of the incident CP energy with respect to the threshold energy and the Coulomb barrier (analogous to nuclear interference).

(2) Selective measurement of the radionuclide D with respect to radionuclide B by spectrometry or decay curve analysis (sections 7 and 9.2). The former is possible if the γ-lines differ by more than the energy resolution of the spectrometer used. The latter is possible if the half-lives of the radionuclides C and D are sufficiently different (about a factor of 2 for comparable activities). If D is short-lived, i.e. its half-life is $\geq$10 times lower as compared to B, then selective measurement (interference $< 0.4\%$) is possible for an irradiation and waiting time equal to one half-life of B (for comparable relative atomic mass, mass fraction, isotopic abundance, cross-section, and range, Equation 9).

(3) Radiochemical separation of B from D, if both are not radioisotopes, i.e. radionuclides of the same element (section 8.4).

Nitrogen can be determined by the $^{14}N(p,n)^{14}O$ as well as by the $^{14}N(p,\alpha)^{11}C$ reaction (Figure 4b). The former reaction leads to ^{14}O, a β^+,γ-emitter. Measuring the 2313-keV γ-ray with a Hewlett-Packard Ge spectrometer, the former reaction only undergoes interference by other γ-emitters within the energy resolution ($\approx$2 keV) of that spectrometer. There could be interference with the latter reaction by, for example, the $^{13}C(p,n)^{13}N$, $^{14}N(p,n)^{14}O$, $^{15}N(p,n)^{15}O$, and $^{16}O(p,\alpha)^{13}N$ reactions, all leading to β^+ (positron) emitters. Measurement of their annihilation radiation (section 7.1) is aspecific and thus could cause spectral interference. As the threshold energies (3–6 MeV) and the Coulomb barriers are not sufficiently different, the interference cannot be avoided by a proper choice of the proton energy. Measurement of the activity as a function of time followed by decay curve analysis (section 9.2) leads to the initial activity of each component if the activity levels are comparable within 1 to 3 orders of magnitude. It is obvious that the other reactions [(p,n), (p,α)] on the analyte element (N) also cause spectral interference (in contrast to nuclear interference). Alternatively, ^{11}C can be radiochemically separated by combustion in oxygen and trapping of the carbon dioxide formed, as the other radionuclides are not radioisotopes of carbon. The latter is called radiochemical analysis (section 8.4) or determination, in contrast to instrumental analysis.

4.3 Matrix Interference

Matrix interference is a special case of spectral interference. As CPAA is primarily a method for trace element determinations or thin-layer characterization, the activity level of the main matrix components could be several orders of magnitude higher than the activity of the analyte element, e.g. 10^6 for a trace element determination at the $\mu g\,g^{-1}$ or ppm level. Then instrumental analysis is almost impossible, even for quite different γ-energies and/or half-lives. Instrumental analysis is only possible for a proper choice of the CP energy, as an alternative for a highly specific radiochemical separation. It is obvious that the determination of light elements (boron, carbon, nitrogen, oxygen) in a high-Z sample is favored by the Coulomb barrier (section 3.2).

5 STANDARDIZATION

Standardization or calibration in CPAA is performed in a relative way, i.e. a sample and a standard are both irradiated and measured. To preserve the advantage of CPAA as an "absolute" or "independent" method, there is no need for a standard similar to the sample and certified for the analyte concentration or mass thickness. The standard is pure elemental matter or a pure compound with exactly known stoichiometry (section 8.1). In a relative method there is no need for absolute data, and errors are cancelled.

For trace element determination in the bulk of a sample, the experimental data (activity) are related to the mass fraction (or concentration) expressed in, for example, mg kg^{-1} (SI), µg g^{-1}, ppm (parts per million) or µg kg^{-1}, ng g^{-1}, ppb (parts per billion).

Surface characterization of thin layers stands for the determination of the partial mass thickness. For an Al_2O_3 layer the partial/total mass thickness is the amount of Al/Al_2O_3 per surface unit, expressed in µg cm^{-2}. From the partial mass thickness the total mass thickness, or thickness or composition (mass fraction or stoichiometry), can be calculated (section 10).

As standardization for trace element determinations in the bulk of a sample and surface characterization have a common and a specific part, the following is divided into three sections. The word ''target'' will be replaced by ''sample'' or ''standard'' in sections 5.2 and 5.3.

5.1 General

The number of radionuclides produced per unit time by irradiation of an infinitesimally thin target is determined by the balance of the increase due to nuclear reactions (proportional to the number of CPs per time unit (i.e. beam intensity, the cross-section and the number of target nuclides per surface unit) and the decrease due to radioactive decay ($-\lambda N$) (Equation 10).

$$\frac{\mathrm{d}N}{\mathrm{d}t} = I\sigma n\,\mathrm{d}l - \lambda N \tag{10}$$

where

N = number of radionuclides,
t = time,
I = number of CPs per unit time or beam intensity,
σ = partial nuclear reaction cross-section,
n = number of target nuclides per unit volume,
l = thickness of the target,
λ = $\ln(2)/t_{1/2}$, decay constant of the radionuclide formed (where $t_{1/2}$ = half-life).

From Equation (10), the number of radionuclides formed and the induced activity after an irradiation time t_i can be calculated. On condition that the beam intensity is constant, this yields Equation (11).

$$A = \lambda N = Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\sigma\,\mathrm{d}x \tag{11}$$

where

A = activity in Bq,
w = mass fraction of the activated element in the target,
M = molar mass of the activated element,
$\mathrm{N_A}$ = Avogadro's constant,
θ = isotopic abundance of the target nuclide,
t_i = irradiation time.

A CP beam is stopped completely in a ''thick'' target, in which the mass thickness D is at least equal to the range R. To calculate the activity for such a ''thick'' target one has to integrate $\sigma\,\mathrm{d}x$ in Equation (11), using the stopping power (Equation 1), as the cross-section has to be expressed as a function of the energy. The calculation is given by Equation (12).

$$\int_0^R \sigma\,\mathrm{d}x = \int_{E_i}^{0} \frac{\sigma\,\mathrm{d}E}{\mathrm{d}E/\mathrm{d}x} = \int_{E_t}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \tag{12}$$

where R is the mass range of the CP in the target, in g cm^{-2}; E_i is the incident CP energy, in MeV; E_t is the threshold energy, in MeV; and S is the mass stopping power, in MeV g^{-1} cm^2. The integration limit $E = 0$ (corresponding to $x = R$) may be replaced by the threshold energy E_t, as the cross-section is zero for $E < E_t$.

If the excitation function is not known, an energy $E_{t/C}$ can be defined as the maximum of threshold energy and Coulomb barrier. The cross-section can be supposed to be zero below $E_{t/C}$, while above $E_{t/C}$ it is constant. Then the following approximation (Equation 13) is obtained from Equations (2) (range definition) and (12).

$$\int_0^R \sigma\,\mathrm{d}x = \int_{E_t}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \approx \overline{\sigma}\left(R_{E_i} - R_{E_{t/C}}\right) \tag{13}$$

if $\sigma(E) = 0$ for $0 \le E \le E_{t/C}$; and $\sigma(E) = \overline{\sigma}$ for $E_{t/C} \le E \le E_i$; where $\overline{\sigma}$ = average cross-section, and $E_{t/C} = \max(E_t, E_C)$.

From Equations (11), (12), and (13) the activity induced in a thick target (i.e. TTY) is given by Equation (14), which approximates to Equation (15).

$$A = Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\int_{E_t}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \tag{14}$$

$$A \approx Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\overline{\sigma}\left(R_{E_i} - R_{E_{t/C}}\right) \tag{15}$$

The target is assumed to be homogeneous, and the decrease in beam intensity (due to nuclear reactions) is negligible. Consequently, neither w nor I is a function of x or E.

For a ''thin'' target, i.e. a target that nearly reduces the CP energy, where the mass thickness D is much lower than the mass range R, the induced activity is given by Equation (16), which approximates to Equation (17).

$$A = Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\int_{E_o}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \tag{16}$$

$$A \approx Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\sigma_0 D \tag{17}$$

where $E_o \approx E_i$, the outgoing energy, in MeV; σ_0 is the cross-section for $E_o < E < E_i$; and $D \ll R$, the mass thickness of the ''thin'' target, in g cm^{-2}.

5.2 Trace Element Determination in the Bulk of a Sample

A relative method is applied, whereby a ''thick'' sample and a ''thick'' standard are irradiated and measured separately. As the nature of sample and standard is different, a correction, F, has to be made owing to the different stopping power (or range) of sample versus standard. Applying Equations (14) and (15) for, respectively, a sample ''x'' and a standard ''s'', and assuming the sample and standard to have equal natural abundance, the mass fraction of the analyte in the sample yields Equation (18), in which F is given by Equation (19), which approximates to Equation (20). (To comply with SI the ''mass fraction'' w is used, while in the CPAA literature often ''concentration'' c is used.)

$$w_x = w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} F \tag{18}$$

where the subscripts x and s refer to a ''thick'' sample, and a ''thick'' standard, respectively.

$$F = \frac{\int_{E_t}^{E_i} \frac{\sigma \, dE}{S_s}}{\int_{E_t}^{E_i} \frac{\sigma \, dE}{S_x}} \tag{19}$$

$$F \approx \frac{\left(R_{E_i} - R_{E_{t/C}}\right)_s}{\left(R_{E_i} - R_{E_{t/C}}\right)_x} \tag{20}$$

The mass fraction (concentration) of the analyte in the sample w_x can be calculated from:

- the mass fraction of the analyte element in the standard w_s,
- the ratio of the activity in sample and standard at the end of irradiation A_x/A_s,
- the ratio of the beam intensity for standard and sample I_s/I_x,
- the decay constant of the radionuclide formed λ,
- the F-factor, i.e. a correction for different stopping power of sample versus standard.

The F-factor can be calculated exactly by Equation (19), requiring:

- relative cross-section data σ from the threshold energy E_t up to the incident energy E_i,
- relative stopping power data for standard and sample S_s, S_x in the same energy interval.

If no cross-section data are available, the F-factor can be approximated by Equation (20). This approximation yields quite good results if one or more of the following conditions (in decreasing order of importance) are fulfilled:

- the atomic number of sample and standard are similar,
- the threshold energy or the Coulomb barrier for the nuclear reaction is high,
- the incident energy is high.

The ''atomic number'' of a compound is, in principle, that of the element with the same stopping power. Practically, the atomic numbers of the elemental components can be averaged: mass fraction weighed. For simplicity Equation (13) was derived supposing a ''step-wise'' excitation function, which does not agree with reality. A detailed study[18] showed that this approximation can also be obtained by assuming that S_x/S_s is a constant as a function of the energy. If the above mentioned conditions are fulfilled, the latter assumption holds better, as can be observed from the parallel lines in Figure 1(b).

Several approximate methods, not requiring cross-section data, have been proposed and evaluated.[18] Equation (19) is known as the first method of Ricci and Hahn.[19] The second method of Ricci and Hahn[20] is fundamentally false. The method of Chaudri et al.[21] yields nearly the same approximation as the first method of Ricci and Hahn. Chaudri's method makes use of stopping power data instead of range data, as shown in Equation (21).

$$F \approx \frac{\left(S_{E_M}\right)_x}{\left(S_{E_M}\right)_s} \tag{21}$$

where $E_M = (E_i + E_{t/C})/2$. The ''average stopping power method'' is an outstanding approximate method with a negligible error that can be calculated.[22,23] However it needs cross-section data, and therefore there is no reason to prefer it to Equation (19). The method has more intellectual than practical merits.

The internal standardization method, not requiring stopping power data for the sample, has more practical applications.[24] The major composition of a sample is often not known, or the composition changes during irradiation. Samples can lose their organic fraction due to heating, and consequently positive errors are made because the stopping power of hydrogen is 3–5 times higher than for the other elements (Figure 1c). Also, approximate methods not requiring cross-section data have been proposed.

The ''two reactions method'' does not require stopping power data.[25] Its main advantage is that accuracy of the

method is not influenced by the accuracy of the stopping power data. Experimentally, the method is quite complex.

Also, the standard addition method does not require stopping power data.[26,27] As CPAA is mainly intended for solid samples, the applicability of this method is limited to powdered samples.

A critical review is given by Strijckmans.[28]

5.3 Surface Characterization of Thin Layers

As for trace element determinations, a relative method is also applied. A ''thin'' surface layer is irradiated and measured together with a ''thin'' or ''thick'' standard. From Equations (16) and (17) the partial mass thickness, using a ''thin'' standard, yields Equation (22), which approximates to Equation (23).

$$w_x D_x = w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} \int_{E_{os}}^{E_i} \frac{\sigma/\sigma_0 \, dE}{S} \qquad (22)$$

$$w_x D_x \approx w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} D_s \qquad (23)$$

where wD is the partial mass thickness; $D \ll R$, the total mass thickness; and the subscripts x and s refer to a ''thin'' surface layer and a ''thin'' standard, respectively.

For a ''thin'' surface layer (order of magnitude μm to nm) the approximation made in Equation (17) holds. For the ''thin'' standard (order of magnitude 10 to 100 μm) one has to check if the cross-section is still constant in the energy interval $E_{o_s} - E_i$. Although Equation (23) does not contain cross-section data, a knowledge of such data, at least relatively in the above mentioned interval, is required in order to evaluate the equation's applicability.

If thin standards are not available, analogous formulae can be obtained from Equations (14), (15), and (17) for a ''thick'' standard. The partial mass thickness using a ''thick'' standard yields Equation (24), which approximates to Equation (25).

$$w_x D_x = w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} \int_{E_t}^{E_i} \frac{\sigma/\sigma_0 \, dE}{S} \qquad (24)$$

$$w_x D_x \approx w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} \frac{\overline{\sigma}}{\sigma_0} \left(R_{E_i} - R_{E_{t/C}}\right)_s \qquad (25)$$

where the subscripts x and s refer to a ''thin'' surface layer and a ''thick'' standard, respectively. Equation (24) requires at least relative cross-section data. From Equation (25) no approximation can be written that does not require cross-section data. The average cross-section in the energy interval $E_{t/c}$ to E_i is never comparable to the cross-section at the incident energy (actually in the very narrow interval E_o to E_i). For trace element analysis in the bulk of a sample such an approximation was possible (section 5.2, and Equations 18 and 20), and it results in accurate analyses observing the conditions mentioned in section 5.2.

Table 1 Standardization in CPAA

Determination	Cross-section data known	No cross-section data known
Trace element determination in the bulk of a sample (mass fraction)	Equations (18) and (19)	Equations (18) and (20) or (21)[a]
Surface characterization of a thin layer (mass thickness)	–	–
Thick standard	Equation (24)	–
Thin standard	Equation (22)	–
Very thin standard[b]	–	Equation (23)[c]

[a] If $Z_x \approx Z_s$ and/or $E_{t/c} \gg 0$, and/or $E_i \gg 0$.
[b] If $\sigma = \sigma_0$ for $E_o < E < E_i$.
[c] Cross-section data only in the energy interval E_o to E_i.

5.4 Conclusions

The standardization formulae to be used are summarized in Table 1. For trace element determinations in the bulk of a sample (mass fraction), accurate results can be obtained without a knowledge of cross-section data if one or more of the conditions are fulfilled. For surface characterization (mass thickness determination) accurate results can also be obtained without a knowledge of cross-section data if a ''very thin'' standard is used.

6 CHARGED PARTICLE IRRADIATION

6.1 Accelerator

CPs have to be accelerated to an energy higher than the Coulomb barrier and lower than the threshold energy of reactions that are more complex than (a,n) and (a,α). The lower limit is to obtain high reaction cross-sections, and consequently high sensitivity; the upper limit is to avoid nuclear interferences by these complex reactions. For protons and deuterons the ideal energy range should be 5–25 MeV, while for helium-3 and α-particles it should be twice as high. A cyclotron is an ideal accelerator for this energy range, although a tandem Van de Graaff also covers the lower energy range.[29]

A tandem Van de Graaff is a DC linear accelerator, the DC source being an electrostatic Van de Graaff generator. The tandem principle is related to the use of the same voltage to accelerate the particles (e.g. protons) twice. First, negative hydrogen ions (H^-) are accelerated to

5 MeV by applying 5 MV from the electrostatic generator; then the ion is stripped from its electrons, and the same reversed voltage is applied again to accelerate to 10 MeV.

The tandem principle is applied a thousand-fold in AC accelerators. A cyclotron is an AC circular accelerator. A high-frequency electric field accelerates the CPs, following a spiral path, due to the magnetic field applied. Ideally, an isochronous cyclotron should be a multi-particle and variable energy accelerator. This implies that the magnetic field and the high-frequency electric field can be tuned, and that the magnetic field as a function of the radius can be adjusted (to compensate for relativistic mass increase during acceleration). Typically, an isochronous cyclotron has radial or spiral sectors in the magnet poles, a prerequisite for beam focusing. Cyclotrons that accelerate helium-3 particles should be equipped with a recuperation system. Typically, the energy can be tuned by a factor of 5–10.

However, the vast majority of cyclotrons in the world is the two-particle and fixed energy type, the so-called ''baby cyclotron''. Their high-frequency electric and magnetic fields are fixed, and the magnetic field increases with the radius by the shape of the magnet poles. Protons or deuterons can be accelerated, the deuteron energy being half the proton energy. Such cyclotrons are negative-ion machines (H^- ions are accelerated, stripped, and the protons formed leave the magnetic field). Extraction of negative ions with a stripping foil is technically very simple compared to extraction of positive ions. These are extracted with a deflector: a strong DC electric field ''extracts'' the accelerated beam out of the magnetic field. The stripping foil (in contrast to a deflector) can be moved into the cyclotron, extracting a lower energy beam. Energy tuning of up to a factor of 2 is possible. Baby cyclotrons are widely used in nuclear medicine departments for PET, i.e. a functional imaging technique using biologically active molecules labeled with cyclotron-produced, short-lived positron emitters.

The beam requirements are not very severe. The energy should be well-defined (within 0.1 MeV is largely sufficient) to control interferences by their threshold energy and/or Coulomb barrier. The beam energy should be reproducible within several days, because this influences slightly the sensitivity. The beam intensity should be tunable from 0.01 to 5 μA, which is largely under the upper limit of these accelerators, and kept constant during irradiation (Equation 11 was calculated assuming that the beam intensity is constant). The unfavorable beam characteristics of a cyclotron versus a Van de Graaff accelerator, such as pulsed beam and energy resolution, are of no importance in CPAA.

6.2 Heat Release in a Target

CP irradiation can cause substantial heating to the sample. Indeed, standards can be irradiated for shorter times and/or at lower beam intensity. The heat release (power) in a ''thick'' target is given by the energy of one single CP multiplied by the number of CPs per unit time (Equation 26).

$$Q = E\frac{I}{Z_a e} \tag{26}$$

where

Q = heat release (power), in W;
E = CP energy, in J;
I = beam intensity, in A;
Z_a = atomic number of the CP;
e $\approx 1.6 \times 10^{-19}$ C (the elementary charge);

or more practically, according to Equation (27),

$$Q = \frac{EI}{Z_a} \tag{27}$$

where Q is the heat release (power), in W; E is the CP energy, in MeV; and I is the beam intensity, in μA. For a ''thin'' target E should be replaced by the energy decrease for the CP traversing the sample, which can be written as the mass thickness of the target and its stopping power (Equation 28).

$$Q = \frac{(E_i - E_o)I}{Z_a} = \frac{SD}{Z_a} \tag{28}$$

where E_i is the incident energy, in MeV; E_o is the outgoing energy, in MeV; S is the mass stopping power, in MeV g^{-1} cm^2; D is the mass thickness of the ''thin'' target, in g cm^{-2}.

Heat is not released uniformly in the sample. It is obvious that the volume is determined by the irradiation surface and the CP range in the target. The depth distribution is given by the stopping power, as shown in Figure 5 for 10-MeV protons in a ''thick'' aluminum target. One can observe the Bragg peak at the very low energy side, i.e. for a depth approaching the range. Two conclusions can be drawn. For ''thick'' targets, heating is more pronounced at the lower energy side, i.e. below the Coulomb barrier, and consequently not in the activated (and thus analyzed) part of the sample. The most efficient back-side cooling is for a sample thickness not exceeding the range. Remember that for trace element determinations in the bulk of a sample one has to always irradiate ''thick'' samples and standards. For surface characterization of thin layers (nm to μm) the heat release is several orders of magnitude lower. It is obvious that also the heat release in the substrate should be considered.

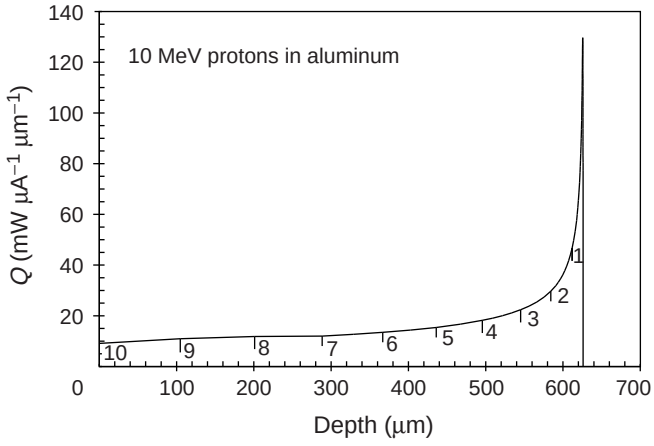

Figure 5 Heat (i.e. power) release as a function of the depth for an aluminum target irradiated with 10-MeV protons. The proton energy decrease is also indicated. The heat release in $mW\,\mu A^{-1}\,\mu m^{-1}$ is equal to the numerical value of the stopping power expressed as $keV\,\mu m^{-1}$, according to Equation (28).

6.3 Target Holder

CPAA has been applied for solid samples (massive or powdered), such as semi-conductors, metals and alloys, ceramics, geological materials, and solid environmental samples (aerosols, soils). It has not been applied for aqueous solutions because optical (atomic) and mass spectrometric methods of analysis are much more suited.

A simple target holder design is a water cooled copper or aluminum base plate, on which a ''thick target'' is mounted, that can be irradiated in the vacuum system of the cyclotron. For surface characterization, the target should be backed (e.g. by aluminum for proton activation) to obtain ''thick'' target conditions, and so avoid activation of the target holder. For powdered samples with poor thermal conductivity, irradiation in a helium atmosphere is recommended.[30] An annotated example is shown in Figure 6.

6.4 Direct Beam Intensity Monitoring

Although the beam intensity I is quantified a posteriori by I-monitors (see below), direct I-monitoring is necessary to keep it constant during irradiation, a prerequisite for obtaining Equation (11). Therefore the target holder should be electrically insulated ($>10^{10}\,\Omega$) from the beam transport system. Moreover, the cooling water should be deionized, as this is the case for the cooling system of the cyclotron. The target holder (i.e. A + B or A + C in Figure 6) is connected to the ground by an ammeter that will measure in the μA range, as shown in Figure 7(a). During irradiation the target tends to charge positively,

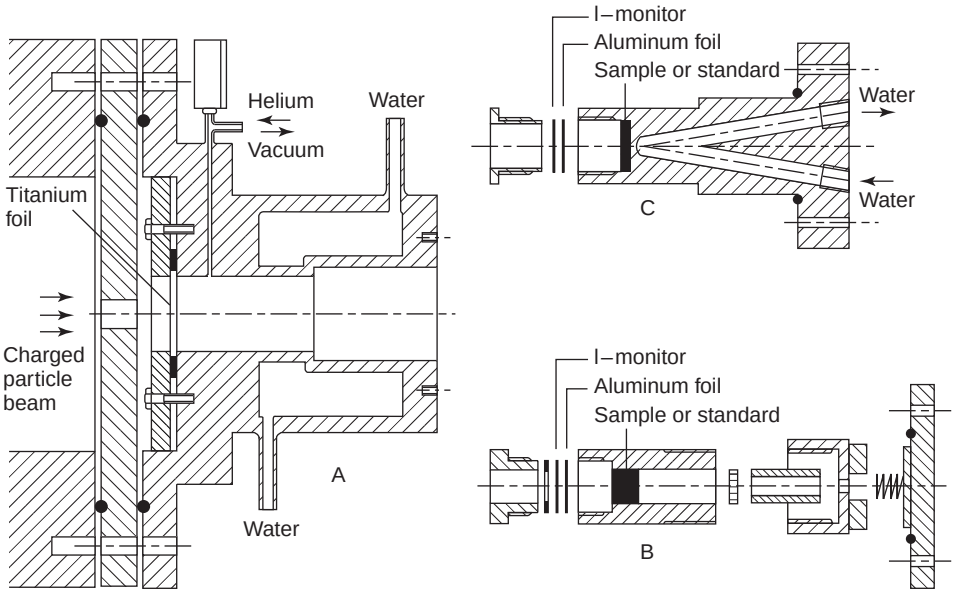

Figure 6 Target system for irradiation in a helium atmosphere.[30] A titanium foil separates the vacuum of the cyclotron from the helium atmosphere. A powdered target is loaded in a dedicated target holder (B), together with a beam intensity monitor foil and an aluminum foil. A disk target is loaded in a dedicated target holder (C), together with a beam intensity foil and an aluminum foil. Powders are loaded from the back-side (B) and, after irradiation, quantitatively unloaded from the front side. Disk targets with good thermal conductivity are also water cooled from the back-side. The target holder (B or C) is brought in a water cooled set-up (A), which is evacuated and filled with helium. (Reproduced by permission of Elsevier Science, from N. De Brucker et al., *Anal. Chim. Acta*, **220**, 93–102 (1989).)

which is balanced by the ''electron'' current measured. By impact of the CPs on the target, secondary electrons are emitted towards grounded parts (e.g. diaphragm) of the beam transport system, which causes positive systematic errors in the *I*-measurement. To avoid emission of secondary electrons (and their positive systematic errors for *I*-measurement) two experimental set-ups are used. A ring-shape electrode at negative potential (−100 V, as the electron energy does not exceed 100 eV) is inserted between the diaphragm and the target, as shown in Figure 7(b). Alternatively, the target holder is electrically connected to a tube, preventing the escape of secondary electrons over a wide spatial angle, as shown in Figure 7(c). A suitable diaphragm prevents the beam from hitting both devices as well as the target holder itself.

6.5 Quantitative Beam Intensity Monitoring

Knowledge of the ratio of the beam intensity for standard and sample (Equations 18, 22–24) is necessary for quantitative analysis in CPAA. There are three reasons for not using direct beam intensity measurement: (1) no absolute beam intensity data are required, (2) to obtain accurate data (no systematic errors due to secondary electrons and/or due to the beam hitting the target holder), and (3) to obtain precise data at low beam intensities (which is mostly the case for standards) compared to the blank. Sample and standard are covered prior to irradiation with a ''thin'' foil (i.e. *I*-monitor) of exactly the same thickness. After irradiation both *I*-monitor foils are measured. From Equation (17) the ratio of beam intensities, I_s/I_x, is given

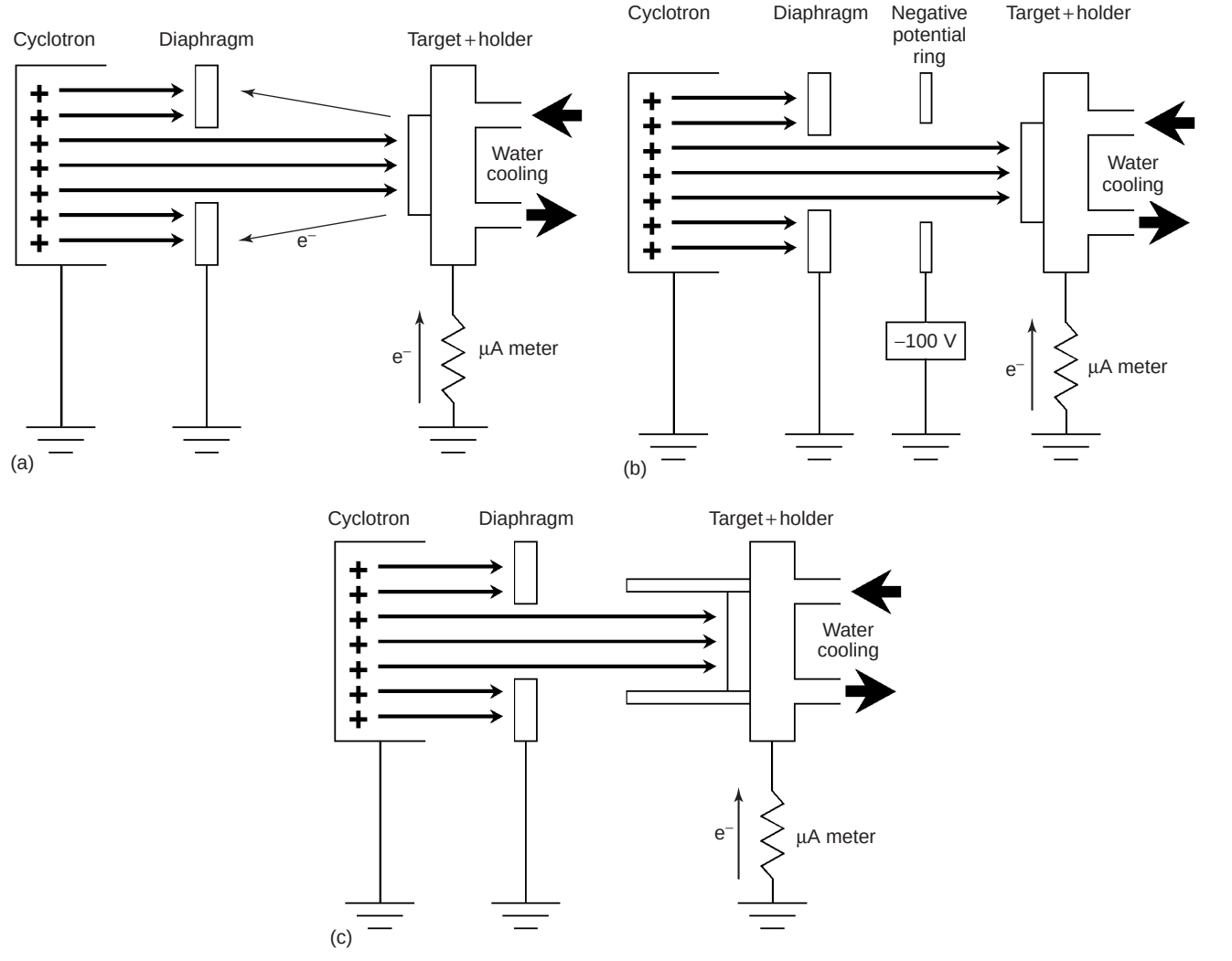

Figure 7 Direct beam intensity monitoring:[28] emission of secondary electrons causes systematic positive errors (a) that can be avoided by a ring shape electrode at negative potential (b) or a tube preventing escape of secondary electrons over a wide spatial angle (c). (Reproduced by permission of Wiley, from K. Strijckmans, 'Charged Particle Activation Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, J. Wiley & Sons, Chichester, 1994.)

by Equation (29).

$$\frac{I_s}{I_x} = \frac{A_s\left(1 - e^{-\lambda t_{is}}\right)}{A_x\left(1 - e^{-\lambda t_{is}}\right)} \quad (29)$$

where I is the beam intensity; A is the activity at the end of irradiation; $\lambda = \ln(2)/t_{1/2}$, the decay constant of the radionuclide (where $t_{1/2}$ = half-life); and t_i is the irradiation time. The subscripts x and s refer to the sample and standard, respectively.

Pure metal foils are the obvious choice as I-monitors because they are: (1) good thermal conductors, (2) monoelemental (the γ-spectra obtained are simple and thus less subject to spectral interference), and (3) available in different thickness.

During irradiation of the I-monitor, recoil nuclides are formed that may leave the I-monitor and enter sample or standard. These recoil nuclides may interfere with the measurement of the sample (or standard) activity, by nuclear interference (if the I-monitor foil is the analyte element, which is the worst possible choice) or by spectral interference. Therefore, a catcher foil is inserted between the I-monitor and the sample or standard that is not activated, e.g. aluminum for proton irradiation.

Sometimes several extra foils are inserted between the I-monitor and sample or standard to reduce the energy to just below the threshold energy or Coulomb barrier of an interfering reaction. This is also the case for a series of standards, when the samples are etched after irradiation (section 8.2).

7 ACTIVITY MEASUREMENT

Section 7.1 explains which radionuclides are formed by CP induced reactions, and gives arguments for measuring γ-spectra, where possible. If this is not possible, measuring the decay curve of annihilation radiation is described. Sections 7.2 and 7.3 give more details about instrumentation.

7.1 Radionuclides Formed and Decay Radiation Measured in Charged Particle Activation Analysis

CP induced reactions produce radionuclides that decay by positron emission and/or electron capture (EC) because they contain one proton in excess compared to stable nuclides. Figure 4(b) shows stable nuclides ^{12}C, ^{14}N, ^{16}O, for which the number of protons (Z) equals the number of neutrons ($A - Z$). In contrast, ^{14}O formed by the $^{14}N(p,n)^{14}O$ reaction is unstable, because of an unbalanced proton/neutron number (8/6). ^{14}O decays by positron emission, i.e. transmutation of a proton into a neutron (that stays in the nucleus), a positron, and a neutrino, both of which are emitted. In this way a stable nuclide ^{14}N is formed again. A positron (β^+) is a positive electron, the anti-matter of an electron. A neutrino has no charge and a negligible mass, and consequently is nearly impossible to detect. The difference in energy level between the unstable radionuclide ^{14}O and the excited state of the nuclide ^{14}N ($E_{\beta_{max}}$) is randomly distributed between the positron and the neutrino. The excited state of ^{14}N further decays to the ground state by photon emission, i.e. γ-rays. Consequently, the energy spectrum of a positron is continuous, from zero to a specific maximum energy ($E_{\beta_{max}}$), while γ-rays are mono-energetic. To identify radionuclides (i.e. qualitative analysis) specifically (i.e. without interferences), γ-spectrometry is preferred.

An alternative to positron emission is EC (or ε). As the decay radiation produced is Auger electrons and/or X-rays, which can easily be absorbed by the sample itself, (in contrast to the γ-rays) they are not used in CPAA.

A limited number of positron emitters decay directly to the ground state without γ-emission. This is the case for the determination of light elements (boron, carbon, nitrogen, oxygen), which is actually the most important application of CPAA. As can be seen from Figure 4(b), the above mentioned reaction is an exception: ^{15}O (determination of N), ^{13}N (determination of C and N), and ^{11}C (determination of B and N) are all pure positron emitters. This is also the case for ^{18}F (determination of O). Selective and sensitive measurement of positron emitters is possible by measuring their annihilation radiation. In positron annihilation a positron loses its kinetic energy, annihilates with an electron, and two annihilation photons (γ-rays) are emitted in opposite direction. According to the mass–energy equivalence law, their energy is 511 keV, corresponding to the mass of an electron at rest. It is obvious that the information about the characteristic maximum energy (and thus the identity of the radionuclide) is lost during the annihilation process. The positron emitter must now be identified by its characteristic half-life. Therefore the activity of a mixture of positron emitter(s) is measured as a function of time, i.e. the decay curve. Decay curve analysis is detailed in section 9.2.

7.2 γ-Spectrometry

γ-Spectrometry can be performed by an NaI scintillation detector and a Ge semiconductor detector. To reduce spectral interferences as much as possible, only a Ge detector is used because of its superior energy resolution characteristics compared to the NaI detector. The Ge detector is a diode, reverse biased. One γ-ray interacts with the detector by Compton scattering or the photoelectric effect. The photo- or Compton electron formed causes ionization in the detector and many electron–positive-hole pairs are formed. The electrons and

positive holes are collected by the electric field applied, and produce a charge pulse. The charge is proportional to the γ-energy (photoelectric effect) or lower (Compton scattering). After linear amplification and analog-to-digital conversion (ADC), the multi-channel analyzer shows a digital spectrum with photopeak(s) corresponding to the γ-energy/energies, and, at the lower energy side, a Compton continuum.

The characteristics of a Ge detector are described by the performance towards the 1333-keV γ-ray of ^{60}Co: (1) energy resolution, i.e. the full width at half-maximum (fwhm), 2 keV or better; (2) peak to Compton ratio, i.e. the ratio of the photopeak height to the Compton continuum at 1 MeV, typically 40–60; and (3) relative detection efficiency, i.e. relative to a $7.6 \times 7.6\,cm^2$ ($3'' \times 3''$) NaI detector at 25 cm source–detector distance, up to 50%. The absolute detection efficiency is of the order of 10^{-3}.

The energy calibration of the γ-spectrometer is carried out by calibration sources covering the whole energy range of interest (0.1–2 MeV), such as ^{241}Am, ^{57}Co, ^{24}Na and ^{60}Co. Activity calibration, i.e. determination of the absolute detection efficiency, is not necessary as absolute activity measurements are not required in CPAA (Equations 18, 22–24). Quantitative analysis requires correction for ''counting losses'' or ''dead-time'', i.e. the spectrometer does not detect (is ''dead'') a γ-ray (a ''count'') because it is still processing the previous one. As radioactive decay is random, the counting-loss probability increases exponentially with the count rate. Therefore, the dynamic range of a γ-spectrometer is limited. Different systems are in common use for the dead-time correction: the ''live-timer'', which is standard in any spectrometer but not applicable for short-lived isotopes, in contrast to the very simple ''pulser method'',[31] the dead time stabilizer (DTS),[32] and the loss-free counting (LFC) module.[33]

Samples and standards should be measured at the same detection efficiency, which is determined by the source-to-detector geometry, and self-absorption as well. For instrumental analysis of solid samples, it is mostly sufficient to measure samples and standards positioned at exactly the same distance from the detector, with the irradiated side towards the detector. Because of the limited range of CPs in matter, the self-absorption of sample and standard (and hence the difference in self-absorption) is very low or negligible, except for very low γ-energies and high-Z targets. This can be roughly estimated, assuming a point source absorbed by half the range, using mass attenuation coefficients compiled by Hubbell.[34,35] For radiochemical analysis (section 8.4) the geometry and self-absorption of the sample (solution or precipitate) can be quite different compared to the standard (a ''thick'' or ''thin'' foil). The relative (i.e. sample vs standard) detection efficiency has to be determined experimentally. A tracer, preferably the mono-elemental standard or, if not, a radionuclide emitting almost the same γ-energy, is irradiated, measured in the standard geometry, brought into the same chemical and physical form as the sample, and measured again. The chemical form should be as similar as possible with respect to its atomic number, Z.

7.3 Positron Counting

Pure positron emitters have to be measured by their 511-keV annihilation photons, emitted in opposite direction (section 7.1). Although this can be measured by a Ge detector, it is more convenient to use an NaI detector because: (1) the detection efficiency is at least equal, even for small NaI detectors, compared to the state-of-the-art Ge detectors; (2) an NaI detector is much cheaper than a Ge detector; (3) a Ge detector should be cooled with liquid nitrogen but an NaI detector does not need cooling; (4) there is no advantage in obtaining the better energy resolution of the Ge detector, as all positron emitters produce annihilation radiation of the same energy. Therefore, there is no need for a multi-channel analyzer. It is sufficient to use a single-channel analyzer that selects all γ-rays measured in the 0.4–0.6 MeV energy interval.

The signal-to-noise ratio can also be improved by the use of a $\gamma-\gamma$ coincidence set-up. Two NaI detectors are used, facing each other, and the β^+ emitter is placed between them. The annihilation photons are emitted in opposite direction, and can be detected simultaneously in both detectors. A coincidence circuit selects all simultaneous events (typically within 40 ns). The system is very insensitive for all γ-radiation other than annihilation photons, even if present in the sample, for positron sources outside the gap between both detectors and for background radiation or random noise. However, the detection efficiency is lower compared to a single detector, because both annihilation photons should produce a photoelectric effect in each detector. The detection efficiency sharply decreases when the source is moved away from the space between the two detectors. Identical geometry for both sample and standard is still more important than for Ge spectrometry. The relative (sample vs standard) detection efficiency can be determined experimentally, as explained in section 7.1, using a pure positron source like ^{68}Ge/^{68}Ga ($t_{1/2} = 270$ days), which is commercially available.

For tuning the set-up (energy range, coincidence time interval) and checking its stability (constant detection efficiency), long-lived positron sources like ^{22}Na ($t_{1/2} =$ 2.6 years, also a γ-emitter) or ^{68}Ge/^{68}Ga ($t_{1/2} = 270$ days) can be used.

As annihilation only occurs after the positron slows down, the positron source should be enveloped by an absorber (of any value of Z). An absorber thickness

equal to the positron range for the characteristic maximum energy of the positron emitter (0.2–0.9 g cm^{-2} for 0.5–2 MeV positrons) is largely sufficient for two reasons: (1) the energy spectrum of the positron emitter is continuous from zero to the characteristic maximum energy; and (2) the projected range (i.e. trajectory projected on the original direction) is much smaller than the linear range (i.e. total path length) for positrons. Indeed, positrons are more deflected when interacting with electrons than with protons. Of course, the annihilation radiation can be absorbed also, therefore a low-Z absorber (plastic, aluminum) is used. The γ absorption of 1-mm aluminum for annihilation photons is 2%.

As for γ-spectrometry, the dynamic range of a γ–γ coincidence set-up is limited by ''counting losses'' or ''dead-time''. A correction is possible, as given in Equation (30), if the dead-time is known, i.e. the time the set-up needs to process one coincident event.

$$R_c = \frac{R_m}{1 - R_m\tau} \tag{30}$$

where R_c is the count rate corrected for dead-time, R_m is the measured count rate, and τ is the dead-time. This equation is only valid for a non-extensible dead-time, i.e. one that is constant as a function of the count rate. Although this is not the case for a γ–γ coincidence set-up, this correction can be applied for counting losses up to 10%, e.g. for $\tau = 10^{-6}$ min per count and $R < 10^5$ counts per min. The dead-time can be determined experimentally by repeated measurements of a short-lived radionuclide that is radioisotopically pure and has a well-known half-life. Irradiation of pure graphite or polyethene with deuterons (<12 MeV) produces the pure positron emitter ^{13}N (Figure 4b) with a half-life of 9.965 ± 0.004 min. Measurements at low count rates, with negligible counting losses, are extrapolated to high count rates (R_c) in Equation (30). The dead-time, τ, can then be calculated from the measured count rate R_m. Alternatively, the best fitted dead-time can be found with a decay curve analysis program (section 9.2) by trial and error.

8 PROCEDURE

The conceptual and experimental procedures to be followed for trace element determinations in the bulk of a sample and for surface characterization are described as a whole, making reference to the other sections if appropriate. The common part is given in the next section. Finally, a radiochemical separation can improve the detection limit and precision.

8.1 General

The first step is to make an appropriate choice for nuclear reaction and CP energy to avoid possible nuclear, spectral, or matrix interference (section 4). For trace element analysis in the bulk of a sample the matrix is the main component of the sample; for surface characterization the main component is the substrate of the surface layer.

The tools are data for threshold energy, Coulomb barrier, cross-section or TTY (if these are not available then Equation 9 may be used), stopping power or range, and a chart of the nuclides (sections 2 and 3). Examples are given in section 4. Activation curves can be obtained experimentally as described in section 3.4.

To preserve the advantage of CPAA as an ''absolute'' or ''independent'' method, no standards may be used that are certified or calibrated by another method. Mono-elemental standards are ideal. Therefore, pure elemental standards (i.e. metals) or a pure compound with exactly known stoichiometry can be used to obtain accurate data for the mass fraction for the analyte element as well as for stopping power. The standard should be stable during irradiation (when heating occurs, as described in section 6.2) in a vacuum. Some organic and hygroscope compounds cannot be used, not only because the mass fraction of the analyte element increases during irradiation, but more importantly because the stopping power decreases as the hydrogen content decreases. Indeed, the stopping power for hydrogen is 3–5 times higher than for the other elements (Figure 1c). Both phenomena lead to positive errors. Not only should the standard itself be stable during irradiation, but so also should the radionuclide formed or any compound of it. Boric acid is a good boron standard for the nuclear reaction ^{10}B(p,α)^{7}Be, but not for ^{10}B(p,n)^{11}C, as the ^{11}C formed volatilizes as carbon dioxide. Boron metal must be used for the latter reaction. To avoid nuclear or spectral interferences the mono-elemental standard should be very pure. Ideally, only activity from the analyte element is found in the γ-spectrum or decay curve.

For irradiation, samples and standards are covered with an I-monitor foil, a catcher foil for recoil nuclides from the I-monitor foil, and, if necessary, extra foils for fine tuning the incident energy on sample or standard (section 6.5). Never use the analyte element as an I-monitor. The thickness of the catcher foil is at least equal to the range of the recoil nuclides. The energy of the recoil nuclides can be calculated exactly, but is always smaller than $E_a + Q$. So the maximum range can be calculated as described in section 2.3. In practice a few μm is sufficient.

The use of an I-monitor, a catcher foil and more extra foils slightly reduces the incident energy on the

sample and the standard. Using the definition of stopping power (Equation 1) this can be calculated by using Equation (31).

$$E_o = E_i - SD \tag{31}$$

where E_o is the outgoing energy, in MeV; E_i is the incident energy, in MeV; S is the mass stopping power, in MeV g^{-1} cm^2; and D is the mass thickness of the "thin" foil, in g cm^{-2}. For each foil with mass thickness D, the outgoing energy E_o of the previous foil is the incident energy E_i of the following foil. For "thin" foils the stopping power is considered constant in the energy interval $E_o - E_i$. The mass thickness of the foils has to be determined experimentally by weighing and measuring its surface, rather than using its nominal or measured thickness multiplied by a tabulated mass density. For an aluminum foil 10×10 cm^2 with a 5-μm nominal thickness, measured with an uncertainty of 0.1 cm and 0.1 mg, the mass thickness yields 1.35 ± 0.03 mg cm^{-2}. The precision on the energy, calculated by Equation (1), yields about 3 keV for α-particles ($S \approx 100$ MeV g^{-1} cm^2), which is largely sufficient. The precision is much better for protons, deuterons, and helium-3 particles, for foils with an atomic number higher than that of aluminum (i.e. foils with a lower stopping power) or for foils with a mass density higher than aluminum or thicker than 5 μm (i.e. foils with a higher mass).

Irradiations are carried out under vacuum or helium, monitoring the beam intensity as described in sections 6.3 and 6.4. For the standards, shorter irradiation times and/or beam intensities are applied, as the amount of analyte element irradiated is orders of magnitude higher than for the sample. To avoid radioactive contamination from the standards, it is good practice to irradiate the samples prior to the standards.

After irradiation, a waiting time, also called the decay or cooling time, is observed for sample handling, chemical etch (section 8.2) or radiochemical separation (section 8.4).

Measurements of samples and standards are done by γ-spectrometry (section 7.2) or positron counting (section 7.3), I-monitors by γ-spectrometry. Time management is important to improve signal-to-noise ratio (and thus the detection limit and repeatability) and to avoid or suppress spectral or matrix interferences. For an irradiation time equal to one half-life one obtains half of the activity for an infinite irradiation time (Equation 9). For a waiting time equal to one half-life one obtains half of the activity for no waiting time (Equation 9). Shorter irradiation or waiting times decrease the analyte activity compared with short-lived interfering activities. Longer irradiation or waiting times decrease the analyte activity compared with long-lived interfering activities. The same holds for the measuring time, simply replacing "activity" by "number of counts" (see Equation 34 in section 9.3).

Data handling is described in section 9.

8.2 Trace Elements in the Bulk of a Sample

To obtain high sensitivity, samples and standards are irradiated at an energy slightly above the maximum cross-section, i.e. when the TTY reaches a plateau. For the determination of iron by the $^{56}Fe(p,n)^{56}Co$ reaction (Figure 3b) the proton energy is about 15–20 MeV. The initial choice is further refined or rejected by considerations of nuclear, spectral and matrix interferences, as explained in section 8.1.

Samples and standards are always "thick" targets (thicker than the range, section 5.2). However, for backside cooling the thickness should not exceed the range too much (sections 6.2 and 6.3). For massive samples (not powders) interference from the surface can be avoided by chemical etch prior to irradiation, at least if further surface contamination can be avoided until the end of irradiation. It is obvious that chemical contamination after irradiation does not interfere any more. For the determination of carbon, nitrogen or oxygen the sample may be contaminated after chemical etch and prior to irradiation. For example, the alumina layer found in air may interfere with trace oxygen determination in the bulk of pure aluminum. The alumina layer is also activated and the oxygen content is found to be too high. Chemical etch *after* irradiation is then the method of choice. The activated surface layer will be removed and replaced by an inactive surface layer that will no longer interfere. The thickness to be removed is not determined by the thickness of the oxide layer, but by the range of the recoil nuclides formed from oxygen in the alumina layer. The energy of the recoil nuclides is less than $E_a + Q$, and its range can be calculated as described in section 2.3. It is good practice to check experimentally the apparent oxygen content compared with the amount removed from the surface layer: a steep decrease followed by a plateau at a few μm. Chemical etch after irradiation reduces the incident energy on the sample, i.e. the sample as it will be measured. Etching the standard is not convenient and may be impossible for powdered standards. A set of two standards is irradiated at two different energies by insertion of additional foil(s) between the recoil foil and the standard. That energy interval should cover the incident energies for all the samples after etch, which is not always very reproducible. Then the standard activity for each individual sample can be obtained by linear interpolation of the set of standards. Energy calculations are done using Equation (31). The mass thickness of the removed surface layer can be obtained by measuring the dimensions of the sample

and weighing prior to irradiation, rather than etch, and after etch. Linear interpolation is justified if the incident energy is chosen in the plateau of the TTY.

8.3 Surface Characterization of Thin Layers

To obtain high sensitivity, samples and standards are irradiated around the energy for maximum cross-section. If Equation (23) is applied, the cross-section should be nearly constant for the energy interval of the ''very thin'' standard (sections 5.1 and 5.3). The initial choice is further refined or rejected by considerations of nuclear, spectral and matrix interferences, as explained in section 8.1.

For surface layer characterization samples are always ''thin''; the standards can be ''thick'', ''thin'', or ''very thin'', each requiring appropriate standardization, as explained in section 5.3. A ''thin'' standard does not mean that the standard should be comparable to the sample thickness (the order of magnitude is μm–nm), or that its thickness or composition should be determined by another method. ''Thin'' standards are typically 10–100 μm thick. Their mass thickness has to be determined experimentally by weighing and measuring the surface, as explained in section 8.2. For a standard that is not available as a foil, a virtual ''thin'' standard can be used. A set of two ''thick'' standards is irradiated at two different energies, one without and the other with an additional foil between the recoil foil and the standard. The difference in standard activities is equal to a ''thin'' standard with a thickness that yields the same energy reduction as the additional foil. The mass thickness of the virtual ''thin'' standard can be calculated from Equation (32).

$$D_s = D_f \frac{S_f}{S_s} \tag{32}$$

where the subscripts s and f refer to the virtual ''thin'' standard and the foil covering the second ''thick'' standard, respectively. As the virtual ''thin'' standard activity results from a small difference in activities, both standards should be measured with good counting statistics. Moreover the virtual ''thin'' standard should be as thick as possible and still comply with the ''nearly constant cross-section'' condition.

Heating of the sample during irradiation is described in section 6.2. To avoid activation of the target holder, samples and ''thin'' standards should be backed by a ''thick'' thermal conductor, e.g. aluminum for proton irradiation.

8.4 Radiochemical Separation

Instrumental analysis is possible if spectral or matrix interference can be avoided by an appropriate choice of the incident energy and/or the measuring conditions (section 4) and/or time management for irradiation and measurement (section 8.1). If not, the radionuclide B formed from the analyte element A has to be separated radiochemically from an interfering radionuclide D formed from an interfering element C. The latter technique is called radiochemical analysis. This section deals with some major differences between a radiochemical separation and a common chemical separation as used in nonnuclear methods of analysis.

For most CP induced reactions, the atomic number of the radionuclide B is different from that of the analyte element A. This is the case for (p,n), (p,α), (d,n), (d,α), (^{3}He,n), (^{3}He,d), (α,n), and (α,d) reactions, but not for (p,d) and (^{3}He,α) reactions (section 3.4 and Figure 4). The chemical separation to be developed for CPAA is thus different from that for all nonnuclear analytical methods and for some other methods based on activation analysis, such as thermal and fast neutron activation analysis using the (n,γ) or (n,2n) reaction, respectively, and photon activation analysis using the (γ,n) reaction. In principle, it is also not necessary to separate the matrix, but rather the radionuclide(s) formed from the matrix element(s). Again, the atomic number of a radionuclide formed from a matrix element is generally different from that of the matrix element itself. Considering only the chemical separation involved, CPAA is an independent analytical method, not subject to the same systematic errors as other analytical methods. Radiochemical separation has three important advantages over conventional chemical separation:

1. An inactive carrier can be added for the elements to be separated (B and D). This avoids the difficulties of a chemical separation at the trace level.
2. Reagent impurities (or blanks) do not influence the detection limit capabilities of the analytical method.
3. Separations need not be quantitative or even reproducible (see section 8.4.2).

For the choice and development of a radiochemical separation the following points have to be considered:

1. *Selectivity of the separation*. As pointed out in section 4, the induced matrix activity (D) can be orders of magnitude higher than the activity to be measured (B). It is not possible to resolve such a low activity in the presence of such a high activity, as the dynamic range of the measuring equipment is limited (sections 7.2 and 7.3). Therefore the decontamination factor (the ratio of D before separation to that after separation) should be up to 10^4.
2. *Quantitative nature of the separation*. Quantitative recovery of B is to be preferred (99% or more).

If the separation is not quantitative and not even reproducible, then determination of the yield for each individual separation is possible. Two approaches can be followed: the addition, before separation, of an accurately known amount of an inactive/active carrier and measurement of the mass/activity of the carrier after separation. It is obvious that the radionuclide added and the analyte radionuclide should be different isotopes of the same element.

3. *Speed of the separation*. To obtain an optimum detection limit, the time needed to perform the whole separation procedure should not exceed a few half-lives of the radionuclide to be measured.
4. *Inactive carrier*. It is not only an advantage, but mostly also a necessity, to add inactive carriers of the elements to be separated. Indeed, if the atomic number of the radionuclide B is different from that of the analyte element A (which is usually the case; see above), this radionuclide B is produced ''carrier free''. As the number of radionuclides is given by

$$N = \frac{At_{1/2}}{\ln(2)} \tag{33}$$

where N is the number of radionuclides, A is the activity (in Bq) and $t_{1/2}$ is the half-life (in s), the amount of substance is 10^{-16} mol for $A = 1$ kBq and $t_{1/2} = 1$ day. It is clear that if there is no accidental addition of inactive element B (e.g. as impurity in the matrix or reagents) it is not possible to separate chemically such a low mass.

5. *Detection efficiency of measurement*. To obtain a high sensitivity, the activity should be measured after separation with the highest possible detection efficiency. This is also determined by the geometry (size and distance from the detector) of the source. Separation procedures ending with volumes of solution are less favorable than those ending with a precipitate.
6. *Repeatability of the measurement*. The repeatability of the measurement is determined by the counting statistics and also by the possibility of reproducing measurements with exactly the same detection efficiency. The former is expressed as the relative standard deviation calculated from counting statistics (i.e. $C^{-1/2}$, where C is the number of counts measured). For a high detection efficiency, one has to use a small source to detector distance, which is less reproducible geometrically. Obtaining a good repeatability and a high detection efficiency are thus contradictory demands.
7. *Self-absorption of the source*. Self-absorption of the source is in principle no problem, as long as the absorption is identical for all samples and standards, which is mostly the case for instrumental analysis (section 7.2). Self-absorption can be important for low-energy γ-rays or for sources containing components of high atomic number.
8. *Separation of the matrix*. Although in principle not necessary, it may be convenient to separate the (inactive) main matrix components and so reduce the mass of material involved in the separation procedure. The following separation process is then faster, leads to smaller volumes of solution or precipitate and allows further evaporation of a solution to a smaller volume if necessary. As a consequence, the detection efficiency is improved and the self-absorption is lowered substantially or becomes negligible.
9. *Health physics*. The activity level required to perform precise measurements (i.e. at the kBq level) provides no health physics risk, as doses are of the order of nSv h^{-1}. The induced matrix activity, however, may be much higher. It is good practice to first separate the main matrix activity, choosing a procedure that limits manual intervention to the strict minimum. Ion exchange chromatography, for instance, is a better choice than manual solvent extraction.

The experimental development of a radiochemical separation is supported by the use of radiotracers. The chemical separation is simulated step by step with the addition of radiotracers (and inactive tracers as well) for the element to be separated (B) and for the interfering element (D), in order to check quantitativeness and selectivity. These radiotracers are preferably produced by thermal neutron activation, if such a facility is available, rather than by CP activation. Indeed, production by, for example, a (p,n) reaction yields a radiotracer that has to be separated from the target material, whereas for production by an (n,γ) reaction the target material can be used as the inactive carrier. The tracer has to be brought to the same chemical form as the radionuclide to be separated. This is often not possible for the dissolution step preceding the actual separation procedure. So, the choice of the dissolution procedure should guarantee quantitative recovery of the radionuclide to be measured.

For each separation step, three measurements can be made: the initial activity, as a reference; and, after separation, the two fractions in which the major and minor activity are collected. From the measurement of the minor fraction the most precise yield can be calculated. However, it is not always possible to perform all measurements with the same detection efficiency. For a precipitation, for example, the reference measurement is made on the solution before precipitation. After precipitation and filtration, a fraction of the filtrate (equal

in volume to the reference) is taken and the yield for the ion to be precipitated can be calculated very precisely. To check coprecipitation (e.g. of matrix activity) the activity of the precipitate is measured. As the detection efficiency is not comparable with that of the reference measurement, the result has to be considered as an estimate. This is largely sufficient for the purpose of evaluating the selectivity of this separation.

There is no need to apply radiochemical separation to mono-elemental standards. However, standards and samples have to be measured at identical detection efficiency. Therefore the standards are brought into exactly the same geometrical form as the samples after separation. If the last step of a separation is ion chromatography, the standards are dissolved and the total volume is brought exactly to the volume of the eluate. Possible differences in self-absorption have to be considered. The eluate may contain an acid other than the one used to dissolve the standard; the eluate may contain only the added inactive carrier (B), whereas the standard contains a larger amount of the analyte element (A).

9 DATA HANDLING

This section follows chronologically from the previous one. Once the irradiation and measurement data (γ-spectrum or decay curve) have been collected, they have to be analyzed and the result (i.e. trace element concentration in the bulk of a sample, mass thickness of a surface layer, composition of a surface layer) has to be calculated.

9.1 γ-Spectrum Analysis

The manufacturers of γ-spectrometers provide suitable programs to resolve the spectrum, i.e. find the significant peaks and calculate their energy, resolution (i.e. fwhm, and comparison to the detector characteristics), net count area, standard deviation (from counting statistics), and detection limit (in number of counts). Once such data have been obtained, it is not sufficient to search only for the most intense γ-ray of the analyte radionuclide. The whole spectrum has be analyzed, starting at the high energy side of the spectrum, using databases with γ-energies and γ-intensities (number of γ-rays emitted per 100 disintegrations). It should be noted that the detection efficiency of a γ-detector is a function of the γ-energy. For high energy γ-rays ($E_\gamma > 1022\,\text{keV}$) single and double escape peaks can be observed at an energy equal to $E_\gamma - 511\,\text{keV}$ and $E_\gamma - 1022\,\text{keV}$. Two intense γ-rays, or γ-rays emitted in cascade, can give rise to sum-peaks. All this can provide useful information concerning the main, minor or trace components in the matrix, which can cause nuclear, spectral and/or matrix interference. Peak broadening is evidence of spectral interference, although no peak broadening does not imply the absence of spectral interference. In case of doubt, decay curve analysis (section 9.2) can be applied to check for spectral interference. All possible reactions on all stable isotopes for an element observed in the spectrum have to be checked, as explained in section 3.4.

Erdtmann and Soyka published a table with γ-energies and γ-intensities, ordered by radionuclide as well as energy.[36] This database is also available at the Lund Nuclear Data WWW Service (Lund University, Sweden).[37] The production mode can be specified as ''CP reaction''. The database is not updated. The NuDat (Nuclear Data) database is retrievable by, and can be sorted by, radionuclide, half-life, γ-energy and/or γ-intensity. The NuDat database at the Lund Nuclear Data WWW Service (Lund University, Sweden) was last updated in January 1996, and will not be updated in the future. In 1999, it is actually the only NuDat base where the production mode can be specified as ''CP reaction''. The NNDC databases are updated three to four times a year, and mirrored at IAEA and NEA.[38] The same database is also obtainable as a PC/DOS program: PCNUDAT.[39] The Table of Isotopes (TOI) database is available as a book and a CD-ROM,[40] and on the Internet at the Lund Nuclear Data WWW Service (Lund University, Sweden).[41] Both are updated yearly.

9.2 Decay Curve Analysis

Positron counting, followed by decay curve analysis, is only applied when the analyte radionuclide is a pure positron emitter. Unfortunately, all CP induced reactions induce positron emitters, and consequently the decay curve may be very complex. In contrast to γ-spectra, which may contain hundreds of γ-rays, the number of components in a decay curve is rather limited (<10). Ideally, a decay curve analysis should answer the questions: (1) how many components are present; (2) what is the half-life of each component; and (3) what is its activity? As most positron emitters are also γ-emitters, their possible presence in the decay curve can be checked by γ-spectrometry and avoided by a proper choice of irradiation and measurement conditions. Moreover, the number of pure positron emitters is rather limited and the number and identity of possible components in the decay curve are mostly known. A decay curve analysis that provides an answer to the second question (qualitative analysis) and the third question (quantitative analysis) is acceptable. By weighed linear regression analysis the third question can be answered, the goodness of fit indicating if the initial assumptions about the number and half-life of

the components are correct. The half-life cannot be fitted because the decay curve is a sum of exponential functions, thus the half-life parameter is not linear. Assuming an initial value for the half-life, the problem can be linearized, and so, by iteration, the best half-life can be fitted. Components that are not present can be rejected because their activity is extremely low or negative. Such an algorithm has been developed by Cumming,[42] and applied successfully. More recently, Prony's improved method has been implemented, tested extensively, and compared to Cumming's method by Schreurs et al.[43] This method also answers the first question, but requires data points equidistant in time.

9.3 Quantitation

Standardization is done in a relative way, the sample and standard being irradiated and measured under, principally, identical conditions. As these identical conditions are not fulfilled in practice, some corrections or normalizations are performed first.

The number of counts measured are corrected for:

1. Counting losses (dead-time), according to sections 7.2 and 7.3, and background (for γ-spectrometry this is mainly the Compton continuum of other γ-rays), using a decay curve analysis program.
2. Differences in measurement (t_w, t_m) conditions (Equation 34), using a decay curve analysis program. The normalized count rate represents the count rate for a decay time zero, i.e. at the end of irradiation. In Equations (22–24) it can replace A for equal detection efficiency of sample and standard.

$$R = \frac{\lambda C\, e^{\lambda t_w}}{1 - e^{-\lambda t_m}} \tag{34}$$

   where
   R = normalized count rate, the number of counts per unit time;
   $\lambda = \ln(2)/t_{1/2}$ decay constant (where $t_{1/2}$ = half-life);
   C = number of counts measured;
   t_w = waiting time;
   t_m = measuring time.
3. Differences in irradiation conditions (I, t_i), corrected for by Equation (18), (22), (23) or (24) (sections 5.2 and 5.3).
4. If applicable, differences in detection efficiency, due to different sample/standard-to-detector geometry and/or self-absorption for sample versus standard (section 7.2).

Finally, difference between the mono-elemental standard and the sample is corrected by Equation (19), (20) or (21) for trace element analysis and Equation (22), (23) or (24) for surface characterization, as explained in sections 5.2 and 5.3. Each individual result can be completed by the standard deviation, calculated from counting statistics, which can be compared to the reproducibility.

For surface characterization, primarily a partial mass thickness is obtained. The total mass thickness can be calculated if the mass fraction (or stoichiometry) is known. The thickness can also be calculated if the mass density is known. Be aware that the mass density can be much lower than the tabulated values, because less dense structures can be deposited, depending on the production mode. The partial mass thickness can also be used to calculate the mass fraction if the total mass thickness is known or if all mass fractions are determined. Also the stoichiometry can be calculated from partial mass thickness data.

10 PERSPECTIVES

10.1 Charged Particle Activation Analysis

During the last two decades CPAA has proved its outstanding capabilities in trace element determinations in solid samples. For the determination of light elements CPAA is still unique. Blondiaux et al. reviewed its applications.[44] Recently CPAA has been proposed for surface characterization[45] and experimentally validated.[46] This method, despite its inherent limitations, is expected to be a reference method, because of its outstanding accuracy, precision and detection limit.

10.2 Charged Particle Activation for the Development of Other Analytical Methods

In addition to CPAA, the production of carrier-free (section 8.4) radionuclides by charged particle activation (CPA) can be a magnificent tool in the development of other analytical methods. A radioactive tracer is a molecule labeled with a radionuclide. A carrier-free radionuclide or tracer is one with a high specific activity, i.e. activity per mass unit. At very low concentrations some analytical methods require preconcentration or chemical separation prior to measurement. These methods can be developed using these tracers, because the extremely low amount of analyte is almost unchanged. The tracer experiments can be performed at the same concentration levels as the final analyses. The activity levels required are very low, and do not present a significant risk. Using rather short-lived radionuclides, there is no waste problem either. An example[47,48] is the determination of platinum by inductively coupled plasma mass spectrometry (ICPMS), using ^{191}Pt.

ACKNOWLEDGMENTS

Grateful acknowledgements are made to the FWO, Fund for Scientific Research, Flanders, Belgium for financial support.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Conversion
CP	Charged Particle
CPA	Charged Particle Activation
CPAA	Charged Particle Activation Analysis
DTS	Dead Time Stabilizer
EC	Electron Capture
fwhm	Full Width at Half-maximum
IAEA	International Atomic Energy Agency
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LFC	Loss-free Counting
NEA	Nuclear Energy Agency
NNDC	National Nuclear Data Center
PET	Positron Emission Tomography
TOI	Table of Isotopes
TTY	Thick Target Yield

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Photon Activation Analysis • Radiochemical Neutron Activation Analysis • Radiotracer Methods

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • γ-Spectrometry, High-resolution, for Radionuclide Determination • Nuclear Detection Methods and Instrumentation

REFERENCES

1. http://www.research.ibm.com/ionbeams/SRIM
2. J.F. Ziegler, J.P. Biersack, U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, New York, Vol. 1, 1985.
3. G. Audi, A.H. Wapstra, 'The 1993 Atomic Mass Evaluation (I) Atomic Mass Table', *Nucl. Phys.*, **A565**, 1–65 (1993).
4. G. Audi, A.H. Wapstra, 'The 1995 Update to the Atomic Mass Evaluation', *Nucl. Phys.*, **A595**, 409–480 (1995).
5. • http://nucleardata.nuclear.lu.se/database/masses *Select:* reaction calculator
   • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* UTIlities – *select:* QCAlc
   • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* UTIlities – *select:* QCAlc
6. • http://www.nea.fr/abs/html/uscd1200.html *Apply for QCALC, i.e. a VAX-VMS program, only for registered users at* http://www.nea.fr/html/signon.html
7. • http://www.nndc.bnl.gov/nndc/exfor/
   • http://www-nds.iaea.or.at/exfor/
   • http://www.nea.fr/html/dbdata/x4/welcome.html *Only for registered users at* http://www.nea.fr/html/signon.html
   • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* DATa bases – *select:* CSISRS
   • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* DATa bases – *select:* CSISRS
   • telnet://db.nea.fr *User:* NEADB – *password:* GUEST – *select:* Data services – *select:* EXFOR
8. K.A. Keller, J. Lange, H. Münzel, *Q-values and Excitation Functions of Nuclear Reactions, Part c: Estimation of Unknown Excitation Functions and Thick Target Yields for p, d, ^{3}He and α Reactions*, ed. H. Schopper, Springer-Verlag, Berlin, 1974.
9. W. Seelmann-Eggebert, G. Pfennig, H. Münzel, H. Klewe-Nebenius, *Karlsruher Nuklidkarte – Chart of the Nuclides*, 5th edition, Kernforschungszentrum Karlsruhe, 1981.
10. J.R. Parrington, H.D. Knox, S.L. Breneman, E.M. Baum, F. Feiner, *Chart of the Nuclides*, 15th edition, General Electric Co. & Knolls Atomic Power Lab., 1996.
11. • ftp://iaeand.iaea.or.at *User:* NDSOPEN – *password:* GUEST – *change directory:* UD4:[SCR.NUCHRT] – *download:* DISKx.ZIP ($x = 1, 2, 3$)
    • http://www.nea.fr/abs/html/iaea1320.html *Apply for* NUCHART; *only for registered users at* http://www.nea.fr/html/signon.html
12. • http://www-csnsm.in2p3.fr/amdc/nubcons_en.html
    • ftp://csnftp.in2p3.fr/pub/AMDC/nucleus/README.NUC
13. http://www.best.com/~sdv/MacNuclide2
14. http://ie.lbl.gov/isoexpl/isoexpl.htm
15. http://nucleardata.nuclear.lu.se/nucleardata/nd.asp?page=Charts
16. http://csnwww.in2p3.fr/amdc/jvnubase
17. • http://hpngp01.kaeri.re.kr/CoN/index.html
    • http://www.dne.bnl.gov/CoN/index.html
    • http://sutekh.nd.rl.ac.uk/CoN/
    • http://www.nes.ruhr-uni-bochum.de/CoN/index.html

18. C. Vandecasteele, K. Strijckmans, 'Standardization in Charged Particle Activation Analysis', *J. Radioanal. Chem.*, **57**, 121–136 (1980).
19. E. Ricci, R.L. Hahn, 'Theory and Experiments in Rapid, Sensitive Helium-3 Activation Analysis', *Anal. Chem.*, **37**, 742–751 (1965).
20. E. Ricci, R.L. Hahn, 'Sensitivities for Activation Analysis of 15 Light Elements with 18-MeV Helium-3 Particles', *Anal. Chem.*, **39**, 794–797 (1967).
21. M.A. Chaudri, G. Burns, E. Reen, J.L. Rouse, B.M. Spicer, 'A Method for Charged Particle Activation Analysis and its Application to Fluorine Determination by The $^{19}F(p,\alpha\gamma)^{16}O$ Reaction', *J. Radioanal. Chem.*, **37**, 243–253 (1977).
22. K. Ishii, M. Valladon, J.L. Debrun, 'The Average Stopping Power Method for Accurate Charged Particle Activation Analysis', *Nucl. Instrum. Methods*, **150**, 213–219 (1978).
23. K. Ishii, M. Valladon, C.S. Sastri, J.L. Debrun, 'Accurate Charged Particle Activation Analysis: Calculation of the Average Energy in the Average Stopping Power Method', *Nucl. Instrum. Methods*, **153**, 503–505 (1978).
24. K. Casteleyn, K. Strijckmans, 'Internal Standardization in Charged Particle Activation Analysis', *Nucl. Instr. Meth., B*, **68**, 161–165 (1992).
25. K. Ishii, C.S. Sastri, M. Valladon, B. Borderie, J.L. Debrun, 'Two Reactions Method for Accurate Analysis by Irradiation with Charged Particles', *Nucl. Instr. Meth.*, **153**, 507–509 (1978).
26. K. Masumoto, M. Yagi; 'Simultaneous Determination of Phosphorus, Chlorine, Potassium and Calcium in Several Control Serums by Alpha-particle Activation Analysis using the Internal Standard Method', *J. Radioanal. Nucl. Chem. Art.*, **109**, 449–458 (1987).
27. M. Yagi, K. Masumoto, 'Instrumental Charged-particle Activation Analysis of Several Selected Elements in Biological Materials using the Internal Standard Method', *J. Radioanal. Nucl. Chem. Art.*, **111**, 359–369 (1987).
28. K. Strijckmans, 'Charged Particle Activation Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, J. Wiley & Sons, Chichester, 215–252, 1994.
29. K. Strijckmans, 'Charged Particle Accelerators', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, J. Wiley & Sons, Chichester, 64–74, 1994.
30. N. De Brucker, J. Dewaele, K. Strijckmans, C. Vandecasteele, 'Determination of Thallium in Zinc by Proton Activation Analysis', *Anal. Chim. Acta*, **220**, 93–102 (1989).
31. K. Strijckmans, C. Vandecasteele, J. Hoste, 'Non-destructive Proton Activation Analysis for Nitrogen in Refractory Metals', *Anal. Chim. Acta*, **89**, 255–266 (1977).
32. J. Bartošek, G. Windels, J. Hoste, 'A New Circuit for Dead-time Compensating Timer', *Nucl. Instrum. Methods*, **103**, 43–44 (1972).
33. G.P. Westphal, 'Real-time Correction of Counting Losses in Nuclear Pulse Spectroscopy', *J. Radioanal. Chem.*, **70**, 387–410 (1982).
34. J.H. Hubbell, 'Photon Mass Attenuation and Energy-absorption Coefficients from 1 keV to 20 MeV', *Int. J. Appl. Radiat. Isot.*, **33**, 1269–1290 (1982).
35. • http://physics.nist.gov/PhysRefData/Xcom/Text/XCOM.html
    • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* DATa bases – *select:* XRAY
    • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* DATa bases – *select:* XRAY
    • http://www.nea.fr/abs/html/dlc-0174.html *Apply for* ZZ-XCOM, *i.e. a PC program, only for registered users at* http://www.nea.fr/html/signon.html
36. G. Erdtmann, W. Soyka, *The Gamma Rays of the Radionuclides*, Verlag Chemie, Weinheim, 1979.
37. http://nucleardata.nuclear.lu.se/database/es
38. • http://www.nndc.bnl.gov/nndc/nudat
    • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* DATa bases – *select:* NUDAT
    • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* DATa bases – *select:* NUDAT
    • telnet://db.nea.fr *User:* NEADB – *password:* GUEST – *select:* DATa bases – *select:* NUDAT
39. http://www.nndc.bnl.gov/nndcscr/pc_prog/ *Only* NDTxx_16.EXE *and* RADTIONS.ZIP *are required*.
40. R.B. Firestone, V.S. Shirley, C.M. Baglin, S.Y.F. Chu, J. Zipkin, *Table of Isotopes*, 8th edition (Book and CD-ROM + yearly update), J. Wiley & Sons, Chichester, 1996.
41. http://nucleardata.nuclear.lu.se/database/toi/
42. J.B. Cumming, 'CLSQ, the Brookhaven Decay Curve Analysis Program', in *Applications of Computers to Nuclear and Radiochemistry*, ed. G.D. O'Kelley, Symp. Proc. Gatlinburg, Tennessee, Office Technical Services, Dept. of Commerce, USA, 25–33, 1962.
43. S. Schreurs, K. Strijckmans, J.-P. François, *Comparison of Prony's Improved and Cumming's Methods for the Analysis of Multicomponent Radioactive Decay Curves*, submitted for publication, (1999).
44. G. Blondiaux, J.L. Debrun, C.J. Maggiore, 'Charged Particle Activation Analysis', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, J.C. Barbour, C.J. Maggiore, J.W. Mayer, Material Research Society, Pittsburgh, 205–230, 1995.
45. K. Strijckmans, 'Charged-particle Activation Analysis', in *Surface Characterisation: a User's Sourcebook*, eds. D. Brune, R. Hellborg, H.J. Whitlow, O. Hunderi, Wiley-VCH, Weinheim, 169–175, 1997.
46. K. De Neve, K. Strijckmans, R. Dams, 'Feasibility Study on the Characterization of Thin Layers by Charged

Particle Activation Analysis (CPAA)', *Anal. Chem.*, **72** (2000).

47. M. Parent, R. Cornelis, F. Alt, K. Strijckmans, R. Dams, 'Use of Pt-191 Radiotracer for the Development of Enrichment Procedures to Detect Natural Levels of Platinum in Biological and Environmental Materials', *Biol. Trace Elem. Res.*, **43–45**, 109–115 (1994).
48. M. Parent, K. Strijckmans, R. Cornelis, J. Dewaele, R. Dams, 'Production of Pt-191 Radiotracer with High Specific Activity for the Development of Preconcentration Procedures', *Nucl. Instrum. Methods B*, **86**, 355–366 (1994).

Chemical Analysis by Nuclear Methods: Introduction

Zeev B. Alfassi
Ben Gurion University, Beer Sheva, Israel

Chemical analysis by nuclear methods is usually an elemental analysis, i.e. it determines the contents of the various elements in the analyzed sample but cannot tell in what chemical form (compounds, valence states) they are. The analysis is based on either a reaction of the analysed element with nuclear projectiles (neutrons or accelerated charged particles, e.g. electrons, protons, heavy ions or γ-photons) (see the following articles within this section: **Instrumental Neutron Activation Analysis; Instrumental Neutron Activation Analysis: Gamma Lines Table; Charged Particle Activation Analysis; Cyclic Activation Analysis; Particle-induced γ-Ray Emission; PIXE (Particle-induced X-ray Emission); Radiochemical Neutron Activation Analysis; Nuclear Reaction Analysis** and **Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons**) or on the scattering of accelerated charged particles. The reaction can be written in the same form as a chemical reaction:

$$\text{Target} + \text{Projectile} \longrightarrow \text{Light product} + \text{Heavy product}$$

or in the more concise form of writings of nuclear physics:

$$\text{Target (Projectile, Light product) Heavy product}$$

Thus, for example, the first production of artificial radionuclide by Joliot and Curie was done by the reaction of α-particles with aluminum metal:

$$^{27}\text{Al} + \alpha \longrightarrow {}^{30}\text{P} + \text{n} \quad (\alpha \text{ is the } {}^4\text{He nucleus})$$

or in the physics notation: $^{27}\text{Al}(\alpha, \text{n})^{30}\text{P}$.

The light product is similar to the projectile and can be a neutron, a small charged particle or a photon. The basis of the nuclear methods for chemical analysis via nuclear reaction (nuclear activation) is the measurement of the amount of light or heavy product formed in a known flux of projectiles for a known length of time (irradiation time). The amount produced is proportional to the number of target atoms, and hence measurement of the amount of the products yields the amount of target atoms. The amount of products formed is too small to be measured chemically (except in very rare cases), and the only way to measure them is by the nuclear physics method of pulse counting. The light product can be measured due to its energy if it is a photon or a charged particle of high kinetic energy, or due to its reaction if it is a neutron. The common denominator of all these processes is that they must be done a very short time after formation of the product, otherwise the light product will lose either its energy or its identity by reaction with the surrounding media. Consequently, measurement of the light product must be done during bombardment of the target with projectiles. This kind of measurement is called prompt activation analysis.

A similar situation exists when the analysis is done by scattering of the projectile, and the measurement is made simultaneously with the irradiation. In this case qualitative analysis is based not on the identity of the products but rather on the kinematics of the scattering, i.e. the energy and direction of the scattered particles. Quantitative measurement is done by counting the number of scattered particles. In the case where the heavy product is radioactive, i.e. it undergoes a spontaneous nuclear transformation, its identity and amount can be measured by its radioactivity. Because the radioactivity can be measured after the end of the interaction between the target and the beam of projectiles, this method of analysis is called delayed activation analysis. Because the heavy product must be radioactive, not all elements can be measured with each projectile. However, by careful choice of the bombarding projectile, almost every element can be determined.

Because many elements are activated simultaneously, the radioactivity measurement must yield both the identity of the radioactive nuclides and their amounts. The term that is sometimes used – radioisotopes – is wrong because when saying isotopes we should say of what element, see **Elemental Analysis by Isotope Dilution**. Nuclide is the nuclear physics term that is parallel to ''atom'' in chemistry. It consists of a nucleus (composed of neutrons and protons) and electrons around it. Because the identity of radionuclides (a short term for radioactive nuclides) in a mixture usually cannot be determined from β^-- or β^+-emission (unless there are only very few in the mixture and they differ strongly in their half-lives or energy of the β-particles), only radionuclides that also emit γ-rays can be used in activation analysis. The radionuclides used for identification are called indicator radionuclides. Only few pure β^--emitting indicator radionuclides, such as ^{32}P, can be used in activation analysis. In this case the phosphorus should be separated from the other radionuclides in the activated sample or the other radionuclides are left to decay before the measurement. The chemical analysis by nuclear activation in which chemical separation precedes the radioactivity measurement is named **Radiochemical Neutron Activation Analysis**, whereas analysis by nuclear activation with direct measurement of the nuclear activity is named Instrumental, see **Instrumental Neutron Activation Analysis** and **Instrumental Neutron Activation Analysis: Gamma**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Lines Table. Owing to the method of measurement of pulse counting, very small amounts of radionuclides can be assessed quantitatively. This and the fact that most natural nuclides are nonradioactive are the basis for the use of radionuclides as tracers in chemical analysis.

Scattering of charged particles is measured both in the forward direction (by **Elastic Recoil Detection Analysis**) and in the backward direction (by **Rutherford Backscattering Spectroscopy**). In the first years of studies mainly the scattering of light ions was studied, but later the scattering of heavier ions became more frequent. The widespread availability of 2–3-MeV accelerators made their use more popular. However, in forward scattering they are restricted to the analyses of H and D. Heavier elements can be analyzed with the beams by backward scattering or by forward scattering of heavier ions. In the recent Ion Beam Analysis Conference at Lisbon (1997) there were two papers on He-induced elastic recoil detection analysis compared to 26 papers on heavy-ion elastic recoil detection analysis. However, for routine analysis the situation is different owing to the availability of the low-energy He accelerators.

Activation by neutrons is carried out mainly in the delayed mode, where very low detection limits can be obtained. Prompt activation is used mainly for in situ and on-line analysis and for elements that cannot be determined by the delayed mode due to the product being nonradioactive or having too short or too long a half-life. An example is the determination of hydrogen and nitrogen atoms by PGNAA (prompt gamma neutron activation analysis), mainly in food samples, see **Prompt γ-Neutron Activation Analysis**.

The slowing down of high-kinetic-energy charged particles passing through matter allows not only the determination of the total concentrations of some elements but also their distribution as a function of distance from the surface, i.e. depth profiles of the concentrations of the various elements. The slowing down can change the energy of the emitted charged particle, e.g. in the reaction $^{10}B(n, \alpha)^{7}Li$. The energy of the α-particle crossing the surface of the sample to reach the detector indicates the distance the α-particle travels in the sample via the energy loss from the original energy of the α-particle.

The slowing down of an impinged high-energy projectile can reduce its energy to the range where a resonance reaction can occur. Thus, by changing the original energy of the projectile, the depth profile of the concentration can be measured.

All nuclear methods, except those using radiotracers, use either nuclear reactors or ion accelerators. These are quite expensive machines that are not found in common analytical laboratories but rather in special laboratories. Many of these methods involve the cooperation of chemists and nuclear physicists.

Chromatographic Techniques in Industrial Hygiene

Rolf M.A. Hahne
University of Washington, Seattle, USA

1 Introduction **1**
2 History **2**
2.1 Origins of Chromatographic Techniques Applied to Industrial Hygiene 2
2.2 Organizations Developing Methods 2
2.3 Overview of Improvements in Chromatographic Techniques 3
2.4 Review Articles 4
3 Sample Collection and Preparation **4**
3.1 Collection of Airborne Samples 4
3.2 Additional Preconcentration of Collected Samples 4
3.3 Collection of Biological Samples 5
4 Gas Chromatography **5**
4.1 Flame Ionization Detection 5
4.2 Electron Capture Detection 5
4.3 Mass Spectrometric Detection 6
4.4 Fourier Transform Infrared Spectroscopic Detection 6
4.5 Other Gas Chromatographic Detectors 6
4.6 High-speed Gas Chromatography 7
4.7 Gas Chromatography Used in Portable, Direct-reading Instruments 7
4.8 Compilation of Gas Chromatographic Analytical Methods 7
4.9 Micro Gas Chromatography 7
5 High-performance Liquid Chromatography **18**
5.1 Ultraviolet/Visible Radiation Detection 18
5.2 Fluorescence Detection 18
5.3 Mass Spectrometric Detection 18
5.4 Atomic Absorption Spectrometry 19
5.5 Inductively Coupled Plasma Spectroscopy 19
5.6 Table of Methods 19
6 Ion Chromatography **19**
7 Other Chromatographic Techniques With Present and Future Industrial Health Applications **22**
7.1 Supercritical Fluid Chromatography 22
7.2 Supercritical Fluid Chromatography and Fourier Transform Infrared 23
7.3 Thin-layer Chromatography 23
8 Quality Assurance and Chromatographic Techniques **24**
8.1 Requirements for Method Validation 24
8.2 Acceptability Criteria 24
8.3 Internal Compared with External Standards 24
9 Alternatives to Chromatographic Techniques **24**
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **25**
References **26**

Chromatographic techniques in industrial hygiene are defined as those separation and analysis techniques which rely on the separation of multiple chemical constituents by taking advantage of the differential adsorption, absorption, mobility, or affinity of the components in a system with a stationary phase (solid, liquid coated on solid, reactive substituent attached to solid surface, ion exchange resin, etc.) and a mobile phase (gas or liquid) which transports the separated constitutents to a detector for immediate identification and possible quantitation or to a collector for later identification and possible quantitation.

By industrial (occupational) hygiene is meant that science and art which is the anticipation, identification, assessment, and control of chemical and physical hazards in the workplace (meant to include offices, places of business, retail establishments, and schools) and home. Included among those methods that make possible the identification and assessment of such hazards are those methods which are used to determine the concentration of airborne and surface chemical contaminants, as well as those which assess levels of exposure to chemical contaminants by the analysis of exhaled breath, blood, urine, or other biological samples.

1 INTRODUCTION

This article will address all the various chromatographic techniques that are employed by industrial (occupational)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

hygienists as part of a strategy for assessing workplace exposure to airborne and surface-borne chemical contaminants. It will also address biological monitoring techniques, which are used by industrial hygienists and occupational health physicians to assess exposure to workplace chemical contaminants, whether through inhalation, dermal, or ingestion pathways. Finally, since industrial hygienists have become involved in assessing more than just traditional workplaces, this article will also include methods that are appropriate for determining relatively low concentrations of chemical contaminants that are found in nonindustrial environments such as schools, public buildings, retail establishments, and homes.

2 HISTORY

2.1 Origins of Chromatographic Techniques Applied to Industrial Hygiene

The technique of gas–liquid chromatography (GLC) was first described by James and Martin[1] in 1952. Two years earlier, Elkins[2] reported on a method for separating halogenated hydrocarbons by adsorption on a sorbent column and then air elution from the column. He also indicated that benzene and toluene could be similarly separated. Although not strictly gas chromatography (GC), it is clearly a chromatographic technique and probably the earliest mention of such an approach for trapping and then separating contaminant mixtures. By 1956, Peterson et al.[3] had reported the use of GC, although not identified as such, applied to monitoring the workplace environment. In 1958 Rushing[4] published an article which predicted some of the areas in which GC would be of value to industrial hygiene. History has proved his visions to be correct. That same year, West et al.[5] published an article related to air pollution monitoring but which is relevant to industrial hygiene as well. The first mentions of GC specifically in the *American Industrial Hygiene Association Journal* (AIHA) were in 1959[6] and in 1960.[7] As an indication of the general level of publication on chromatographic analyses in industrial hygiene, it was only first in 1960 that *Chemical Abstracts* initiated a section of Water, Wastes, and Air Pollutants. By 1962, that section has become Toxicology, Air Pollution, and Industrial Hygiene, and in 1972, a section covering Air Pollution and Industrial Hygiene alone was begun in *Chemical Abstracts*. In 1963, Altschuller[8] wrote a survey article describing GC methods in air analysis, but no compound-specific methods were identified. In 1965, a paper presented at a National American Chemical Society (ACS) Meeting addressed the analysis of a mixture of hydrocarbons in ambient air.[9] The late 1960s saw a dramatic increase in the publication of GC methods applicable to industrial hygiene, and the technique has become the single most prevalent industrial hygiene analytical technique.

Although column chromatography has been in use for most of this century, having been described by Tswett in 1906,[10] high-performance (pressure) liquid chromatography (HPLC) was developed over a period of time beginning in the mid 1960s and grew rapidly in the early 1970s. As late as 1968, an article in *Advances in Chromatography*,[11] reported ''By comparison, GC and thin-layer chromatography (TLC) made the greatest strides in recent years and such methods as column and paper chromatography have received little attention''. However, in a 1972 review article, Zweig, Moore, and Sherma[12] reported that ''Modern high pressure liquid chromatography emerged as a widely used analytical technique which promises to become one of the most popular and powerful of all chromatographic procedures''. By 1974, the review *Advances in Chromatography*, Vol. 12,[13] included an article on the use of HPLC in pharmacology and toxicology.

Similarly, ion chromatography (IC) built on the ion-exchange chromatography which had been used extensively for inorganic separations, but the use of high pressures, small bore columns, and direct measurements of column effluent was reported by Small et al.[14] in 1975. One of the earliest papers applying IC to industrial hygiene was for the measurement of chloracetyl chloride in workplace air.[15]

Not only was Elkins one of the earliest in suggesting chromatography-like methods for workplace air monitoring, but he was also foresighted in his proposal[16] to use biological monitoring as a means of assessing human exposure to various chemicals.

A historical overview of chromatographic methods, but without reference to industrial hygiene is found in a volume edited by Ettre and Zlatkis.[17]

2.2 Organizations Developing Methods

The AIHA developed a series of hygiene guides, starting in the mid-1950s and these guides included analytical methods for the contaminants, which then became AIHA Analytical Guides. In the early 1960s, the AIHA Analytical Guides[18] had no chromatographic techniques included in them, but by 1969, GC methods for ethyl benzene and 2-hexanone were included in this compilation.

In 1974, a recently formed National Institute for Occupational Safety and Health (NIOSH) published the first edition of the National Institute for Occupational Safety and Health Manual of Analytical Methods (NMAM)[19] with the methods designated by the prefix P&CAM (Physical and Chemical Analytical Methods), and including

several methods that used GC. Some of these methods were carried over from the old Bureau of Occupational Safety and Health of the Public Health Service. In 1974, NIOSH sub-contracted with SRI International for the completion of the analytical methods for existing contaminants for which a Permissible Exposure Limit existed and a series of methods were published with the prefix "S" (for Standards Completion Program).[20] Now in the fourth edition, the NMAM[21] includes 467 methods which use GC, HPLC, or IC – over 80% of the total in the manual.

The laboratory of the Occupational Safety and Health Administration (OSHA) in Salt Lake City has also developed a significant number of chromatographic methods which are available to the public.[22] The United States Environmental Protection Agency (USEPA) has developed a series of chromatographic methods directed toward the analysis of low levels of airborne contaminants which one might find in the home or school, or general ambient environment. These are designated as TO-Series (total organic) methods.[23] In 1980, the USEPA also published a compilation, "Analysis of Pesticide Residues in Human and Environmental Samples", which contains solely chromatographic methods of analysis.[24] The American Society for Testing and Materials (ASTM) has included methods for workplace air in its publications since the 1950s (although as late as 1969 there were no chromatographic methods among them) and currently has a large number of GC, HPLC, and IC methods in Volume 11.03 of its Annual Book of Standards.[25] The Intersociety Committee has also published a compilation of (among others) chromatographic methods[26] used in both workplace and ambient air monitoring. A reference which has 21 detailed chromatographic methods (as well as a number of non-chromatographic methods) for the measurement of substances in humans as an indication of exposure has been published by the American Public Health Association.[27]

In the United Kingdom, the Health and Safety Executive (HSE) has developed a series of chromatographic methods (and a number of non-chromatographic methods)[28] which are used for assessing workplace exposure. These methods, part of the Methods for the Determination of Hazardous Substances (MDHS) series, include 34 methods which employ chromatographic techniques out of a total of 81 (including methods and procedures).

In Canada, until 1996, the Ontario Ministry of Labor operated a laboratory which was involved in the development and evaluation of industrial hygiene methods. Notable among their developments was the tryptamine (TRYP) isocyanate method,[29] which has since been incorporated into the NMAM as Method 5522. In British Columbia, the Workmen's Compensation Board of the province has developed and evaluated a series of methods,[30] including a number of chromatographic methods, for use in occupational health hazard evaluations in that province.

2.2.1 Industry Developments

Many large multinational chemical and pharmaceutical firms (e.g. Dow Chemical, Rohm and Haas, Dupont, Abbott Laboratories, American Cyanamid, BASF) have developed or modified existing chromatographic methods intended for workplace monitoring of hazardous substances. Some of these methods have been published in the literature and many more are found in internal compilations. To some extent, individual internal methods may be available to professionals, through the Industrial Hygiene Methods Exchange Network of the AIHA in Fairfax, Virginia.[31] A number of these companies have indicated to the AIHA Methods Network which methods they have developed and these are available, upon request, directly from the company that developed them.

2.3 Overview of Improvements in Chromatographic Techniques

Dramatic improvements in chromatographic techniques for industrial hygiene have taken place since the first methods were applied in the 1950s. In the field of GC, thermal conductivity detectors, isothermal separations, and packed copper columns have been supplanted in the past 40 years by the introduction of temperature programming, high-resolution capillary[32] and megabore columns, a variety of highly sensitive detectors such as flame ionization, electron capture, mass spectrometric, flame photometric, thermionic, photoionization, and Fourier transform infrared (FTIR) spectroscopy. High-speed GC has, in recent years, allowed the use of transportable gas chromatographs for making field measurements (see below).

HPLC has been dramatically improved in its sensitivity and selectivity by going from coulometric and differential refractive index detectors to multiple wavelength ultraviolet/visible (UV/VIS), diode array, mass selective, and fluorescence detectors and enhanced its separation power by having instruments that can change eluent composition among multiple components as a function of time. The development of reversed-phase columns has also increased the separation power of HPLC. An example of this enhanced sensitivity and separation power is the analysis of complex mixtures of polynuclear aromatic hydrocarbons (PAHs) found in the air in coal conversion (gasification and liquefaction) plants. Such a

compound-specific analysis, which would have been difficult, if not impossible with packed columns and thermal conductivity detectors, was reported in the late 1970s.[33] The entire field of HPLC has grown up during the period of maturation of industrial hygiene monitoring in the United States, and made simple many analyses that had been difficult, or insensitive, or both, prior to that.

2.4 Review Articles

There are periodic review articles which generally cover the entire field of analytical chemistry applied to industrial hygiene, but in which methods that use chromatographic techniques are the overwhelming majority. The first identified review article in *Analytical Chemistry* wholly dedicated to industrial hygiene methods was in 1983,[34] although earlier review articles (in odd-numbered years) included some industrial hygiene methods in articles discussing advances in air pollution analysis. The most recent applied review on industrial hygiene in *Analytical Chemistry*, which contains many advances in the area of chromatography applied to industrial hygiene was written by Harper et al.[35] Review articles related to biological monitoring are not only focused on the chemical analysis of biological specimens, but also the rationale for establishing biological exposure guides. In 1975, R. Lauwerys published one of the earliest reviews of biological monitoring.[36] A more recent book, which contains 26 different chromatographic (primarily GC) methods for biological monitoring, is a useful reference for such analyses.[37] The USEPA has also produced a compilation of methods for pesticide analysis in human (and environmental) samples.[38] Recent *Analytical Chemistry* reviews on industrial hygiene methods currently also include a review of recently-developed biological monitoring methods.

3 SAMPLE COLLECTION AND PREPARATION

3.1 Collection of Airborne Samples

Collection of airborne samples for chromatographic analyses takes a number of different approaches. In some cases, the contaminants of interest (gases and vapors) are collected, as found, in such devices as evacuated glass cylinders, plastic (Tedlar®, Teflon®, Mylar®, etc.) bags, electropolished stainless steel containers, or evacuated aerosol cans and then taken back to the laboratory for analysis as is, or with concentration prior to analysis using a solid sorbent, midget impinger, or cryogenic trap.[39]

In many cases, the (gas and vapor) contaminants are concentrated in the field by drawing the air through a solid sorbent such as activated charcoal (Pittsburgh coconut-based charcoal and Columbia JXC charcoal are a couple of examples), silica gel, various synthetic polymers (such as various Chromosorbs®, Anasorbs®, Porapaks®, or XAD-2),[40] as well as such sorbents treated with a derivatizing agent (e.g. 2,4-dinitrophenylhydrazine for aldehydes {NIOSH Method 2539}, HBr for ethylene oxide {NIOSH Method 1614}, or 1,2-pyridyl piperazine for certain isocyanates {NIOSH Method 5521}) and then desorbed with an appropriate solvent prior to chromatographic analysis.

In other cases, the airborne contaminant (gas, vapor, or particulate) is concentrated by trapping on a filter (e.g. glass fiber, polyvinyl chloride, cellulose ester or silver membrane), either through physical capture or else by reaction with a derivatizing agent which is impregnated in the filter. Some examples of this are the capture of MDI (methylene bisphenyl isocyanate, 4,4′-methylene diphenyl isocyanate) and TDI (toluene 2,4-diisocyanate) on 1,2-pyridylpiperazine-treated glass fiber filters[41] and the collection of PAHs in a dual collector sampling train[42] consisting of a silver membrane filter backed up with a solid sorbent tube which traps any subliming particulate matter from the filter.

Airborne contaminants of interest to the industrial hygienist can also be concentrated by trapping by passing the air through a midget impinger or midget bubbler[43] which contain an appropriate solvent with or without a derivatizing agent (e.g. 1,2-pyridylpiperazine for some isocyanates, 2,4-dinitrophenylhydrazine or aqueous $NaHSO_3$ for formaldehyde {NIOSH Methods 2539 and 2541}).

3.2 Additional Preconcentration of Collected Samples

Chromatographic methods for the analysis of such samples has also been enhanced by postcollection concentration of the contaminant(s) of interest. In addition to the cryogenic or sorbent trapping of contaminants from workplace air samples, several other techniques have been used. In headspace analysis, the sample of interest (e.g. a construction material or piece of carpet suspected of emitting a compound which causes indoor air problems, a blood sample containing low levels of an airborne contaminant, or a plastic film which is heated during food packaging) is heated in a closed, inert-gas purged container and the headspace (and any volatile contaminants released from the sample) swept onto a trap which preconcentrates the sample prior to chromatographic analysis. In a related technique, an inert gas is bubbled through the liquid of interest and the gas flow trapped prior to analysis (purge and trap).[44] As part of the sample preparation prior to analysis, certain samples that contain large amounts of interfering materials are purified

by suspending the sample of interest in an appropriate solvent and then passing the entire mixture through a solid phase, which retains the compounds of interest but allows other interferents to pass through or else be collected on a pre-filter (if undissolved). This is solid phase extraction (see below).

3.3 Collection of Biological Samples

In a number of countries, exposure standards are set on the basis of actual measurements of biological samples. Also, actual measurements of exposure, rather than surrogates such as measuring the contaminant's concentration in the ''breathing zone'' of the worker are sometimes preferable, if more difficult. The monitoring techniques relate to such parameters as allowable concentrations of volatile solvent contaminants, contaminant metabolites, or anesthetic gases in exhaled breath, concentration of hexachlorobenzene in blood, or concentrations of organophosphate metabolites or anesthetic gases in the urine of exposed individuals. The collection techniques for each of these analyses – which require chromatographic analysis – are varied, and require specific consideration for maintaining sample integrity and simplicity of preparation for analysis. Sample preservation prior to shipping, minimizing interferences, and other issues of quality assurance (QA) are of paramount interest.

4 GAS CHROMATOGRAPHY

In a period of less than 40 years, GC has gone from being used in a handful of industrial hygiene analytical methods to the dominant method for analysis of workplace contaminants and biological markers. As an example, of the 574 compounds or groups of compounds listed in the NMAM, 67% of them call for GC methods. Over the past 15–20 years, GC methods have evolved from utilizing packed columns to using capillary columns, both standard bore and megabore, in both split and splitless modes. Early on, temperature programming was introduced to enhance separations and speed up analyses, and the use of multiple detectors was important for compound verification until the introduction of the mass spectrometric detector for GCs about 25 years ago. An indication of the early status of the separation and detection combination of gas chromatography/mass spectrometry (GC/MS) was the appearance of a paper which was presented at an international symposium on PAHs in 1973.[45] GC/MS is now frequently used for the identification and quantitation of volatile and semivolatile organic compounds in air. A recent example is related to pesticide usage indoors.[46] Another recent paper using this technique focuses on the determination of microbial contamination in indoor air,[47] a topic of considerable interest in the industrial hygiene community.

4.1 Flame Ionization Detection

The single most common detector used for the analysis of gas chromatographic eluents is the flame ionization detector (FID). Introduced in the late 1950s, FID relies upon the production and subsequent collection (in an electric field) of ions when a GC column eluent enters into an air–hydrogen flame and the molecules present are ionized. The detector is highly sensitive to compounds containing carbon–hydrogen bonds (but is responsive to other bond types as well) and is linear over many orders of magnitude of concentration. FIDs are commonly used for the detection of various hydrocarbons – aliphatic, aromatic, and alicyclic – oxygenated hydrocarbons such as alcohols, ethers, ketones, acids, aldehydes, and esters, and many chlorinated hydrocarbons. The FID is relatively rugged and is simple to maintain and operate. The variables that are used to change sensitivity include the air–hydrogen ratio in the flame and the flow rate through the flame. Collection voltage typically remains constant. NIOSH has a number of GC methods which call for the use of a FID, including those for various aromatic hydrocarbons (NIOSH Method 1501), aliphatic hydrocarbons (NIOSH Method 1500), chlorinated hydrocarbons (NIOSH Method 1003), ketones (NIOSH Method 1300), and alcohols (NIOSH Methods 1400–1403). Although not as sensitive as some detectors, for certain compounds, the linearity of response of FIDs over many orders of magnitude makes them a highly desirable detector for gas chromatographic analyses. An example of the application of gas chromatography/flame ionization detection (GC/FID) is a paper by Wellson, et al.[48] reporting on monitoring for glutaraldehyde, a disinfectant and preservative used as a formaldehyde replacement. Cresols and phenol are common analytes measured in the urine as measures of exposure to toluene, and benzene and phenol, respectively, and a method for their determination using GC/FID was published by Needham et al. in 1984.[49] Expired air measurements using GC/FID was reported by Guillemin and Guberan.[50]

4.2 Electron Capture Detection

Electron capture detection (ECD), which relies on the attenuation by electronegative atoms of a beta-particle beam from a radioactive source (e.g. ^{63}Ni, ^{3}H) within the detector, has been developed and improved within the past 25 years. It is very effective for the detection of compounds containing halogen atoms, which has made

it extremely useful for the detection of halogenated compounds such as dichlorodiphenyltrichloroethane (DDT), 2,4,5-trichlorophenoxyaceticacid (2,4,5-T), Lindane, Chlordane, polychlorinated biphenyls (PCBs), and polybrominated biphenyls (PBBs). Human exposure to these compounds, whether in a farm field, in a residence, or in a workplace can now be determined with considerable sensitivity. For example, the analysis of PBBs in blood serum using gas chromatography/electron capture detection (GC/ECD) was reported in 1980 by Burse et al.[51] Another example of the use of GC/ECD is a recently reported method[52] developed for assessing occupational exposure to two compounds used as taggants in plastic explosives.

Another analytical application in the area of biomonitoring is discussed in a recent paper on the biological monitoring of workers for 4,4′-methylenebis(2-chloroaniline) using GC/ECD.[53]

4.3 Mass Spectrometric Detection

In the mid-1960s, in order to enhance the ability to unambiguously identify compounds eluted from a GC column, mass spectrometers were interfaced with the effluent from GC columns.[54,55] The critical feature of the mass spectrometric detectors used was that they were quadrupole detectors, rather than magnetic sector detectors, which allowed one to do a scan of mass to charge ratios from 28 to 300 within a second, allowing one to look at a whole range of ion fragments within the time duration of a capillary GC peak. The mass spectral fingerprint gave more reliable information than retention time alone, about the compound being eluted. The development of high-speed computers allowed the resulting mass spectra to be matched against large libraries of known mass spectra, using various algorithms, and assigned a ''matching index'' which indicated the degree of correspondence between the unknown peak's mass spectrum and the library spectrum. EPA (Environmental Protection Agency) Methods TO-2 and TO-13[23] are particular approaches for using GC/MS to identify and quantitate a large array of organic compounds that might be found in indoor or ambient air. One can also use the mass selective detector to monitor a single ion (selective ion monitoring, (SIM)), if one knows what compound/ion to expect, considerably enhancing the sensitivity of this detection method and reaching into the picrogram range for limits of detection (LOD). A method for derivatization and GC analysis of formaldehyde using GC/MS using SIM has been reported by Wu and Que Hee.[56]

The GC/MS technique is also a powerful one in the identification of unaltered compounds or metabolites in biological specimens.[57]

4.4 Fourier Transform Infrared Spectroscopic Detection

The development of FTIR detectors (see **Gas Chromatography/Infrared Spectroscopy**), which can perform a scan of infrared absorption over a range of ~8000 to 350 cm^{-1} in a few seconds, has allowed this technique to also do a number of scans during the duration of even a highly resolved capillary GC peak. As with MS detectors, the ability to store vast libraries of infrared spectra in a high-speed computer's memory allows one to match the FTIR spectrum of an unknown against those in the library, with the creation of similar ''matching indices''. A paper by Wieboldt et al.[58] discusses the use of FTIR in the analysis of certain pyrethrin pesticides, compounds included in the NMAM.

4.5 Other Gas Chromatographic Detectors

Flame photometric detection (FPD) has gained widespread use for industrial hygiene analyses because of its sensitivity and selectivity to sulfur and phosphorus in the compounds of interest. These flame emission detectors are used for the detection of low concentrations of organosulfur compounds in workplace air (e.g. hydrogen sulfide, mercaptans, and organo-sulfides and -disulfides) and for the detection of metabolites of organophosphate phosphates, such as dimethyl- and diethyl-phosphates and in urine.[59] Detection levels at the nanogram level allow measurement of low-level exposure to such compounds. Another example, a paper by Kennedy et al.[60] describes a method for the simultaneous determination of 19 organophosphate pesticides in air using GC and a flame photometric detector. FPD has also been used for GC effluent in the separation of organogermanium and organotin compounds.[61]

Hall® (electrolytic) conductivity detectors operate by running the organic eluates from a GC column through a high-temperature reactor, which decomposes them to carbon dioxide and small inorganic molecules (the latter, if the molecules contain N, S, or a halogen atom), and then runs the decomposition products through an electrolytic conductivity cell. Changes in the conductivity are a measure of the concentration of the compound of interest in the effluent. The Hall® conductivity detector is typically used for pesticide and herbicide analyses and thus is useful in assessing workplace exposures to such compounds. A couple of examples of such applications of Hall® detectors with GC are found in Lawrence and McLeod[62] and Stetten et al.[63]

Argon ionization detectors have been used in at least one transportable GC sold to the industrial hygiene community (Sentex Systems, Inc., Ridgefield, NJ.). This detector relies on the use of argon atoms, excited to a neutral, energetic state by a radioactive source, transferring their excitation energy to organic molecules

and, as a result, ionizing them. The ionized molecules are then measured in a manner similar to that for flame ionization. The advantage of an argon ionization detector is that no flame is needed, thus eliminating the need for fuel and oxidant gases, and reducing the weight of the instrument.

Photoionization detectors (PID) rely on the detection of species being eluted from a GC column by their ionization by ultraviolet (UV) light. UV lamps with energies ranging from 8.3 to 11.7 eV are currently available (not all lamp energies are available for all instruments), and gas chromatographs with PIDs can be purchased from several manufacturers of such instruments. An article by Driscoll and Clarici gives a good overview of PIDs and their possible use in industrial hygiene chromatography.[64]

Thermionic ionization detectors, sometimes called nitrogen–phosphorus detectors, are GC detectors which can be particularly sensitive to N and P under the right conditions. They typically cause ionization of the compound(s) of interest through surface ionization processes in the presence of a low-work-function alkali-metal-containing bead, with subsequent ion collection and measurement. Depending on the gaseous boundary around the detector and the geometry of the detector, they can be made highly specific for nitrogen- and phosphorus-containing compounds or for other families of compounds.[65] An example of this detection is found in a paper by Skarping et al.[66] Another application of a thermionic detector with GC is the measurement of low concentrations of phosphine in workplace air, as reported by Vinsjansen and Thrane.[67]

Compounds containing metals atoms have also been chromatographed, using graphite furnace atomic absorption spectrophotometry as the detector. In a paper by Robinson et al.[68] trapped alkyl lead compounds were chromatographed and the column effluent connected directly into the graphite furnace atomizer for lead quantitation.

4.6 High-speed Gas Chromatography

Opportunities for improving the productivity of the GC technique lie in the reduction of the separation time for compounds of interest to industrial hygienists. A number of developments in this area have been made through which the separation time can be reduced significantly. Sandra[69] reviews developments in fast capillary GC relevant to industrial hygiene. As an example, he shows a mixture of ten aliphatic and aromatic hydrocarbons which can be separated (*m*- and *p*-xylene are unresolved) in about 150 s. This is accomplished by reducing film thickness inside the capillary to 0.25 μm, while a chromatogram for the same compounds using a 3 μm film thickness has a retention time about five times longer. Compounds which elute prior to benzene required the thicker film to resolve them. A commercially available system which claims separation times ten times shorter than regular GC[70] uses a patented cryofocusing inlet system, short column lengths, and high carrier-gas velocities to achieve much-shortened retention times. A complete chromatogram – with nearly baseline separation – of 24 volatile organic compounds commonly found in the workplace is generated in about 135 s.

4.7 Gas Chromatography Used in Portable, Direct-reading Instruments

Starting in the early 1980s, manufacturers have begun to build portable, or at least transportable, GCs for use in the field as direct-reading instruments. Initially quite heavy and bulky because of the the necessary weight of batteries, carrier and flame gases, and thermal isolation of the GC column, as well as the size of the column and associated electronics, these instruments are now found in truly hand-held devices. Part of that development has come from batteries with higher energy densities, lighter-weight detectors, and the development of GC columns that are extremely small and have been etched on silicon wafers on the order of 25 cm^2.

A recent addition to the portable GC market employs a surface acoustic wave (SAW) detector,[71] rather than FIDs[72] or PIDs[73] more commonly found in such devices. The SAW detector relies on the measurement of the change in (radio)frequency of mechanical waves on the surface of a piezoelectric substrate. The frequency changes as a function of mass adsorbed on its surface due to the presence of an adsorptive coating on the surface of the detector. These devices are being marketed as nearly real-time measurements of workplace and ambient air, are quite simple, as well as rugged and will certainly see wider use in the next decade, perhaps more often in multisensor arrays.

4.8 Compilation of Gas Chromatographic Analytical Methods

Table 1 gives a summary of GC methods which have been validated by one of six sources (NIOSH, OSHA, USEPA, Intersociety Committee, ASTM, or the HSE) for the measurement of contaminants in airborne, surface wipe, or biological specimens.

4.9 Micro Gas Chromatography

The rapid development of micromachining techniques has had its influence on GC related to occupational hygiene. Capillary GC columns 9 m long, with dimensions of 300 μm wide by 10 μm high, along with similar dimensions for the detectors, have been fabricated by

Table 1 Gas chromatographic methods for airborne and biological contaminants, from six different references

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Acenaphthene	83-32-9	5515		EPA TO-13; MASA 102D
Acenaphthylene	208-96-8	5515		EPA TO-13; MASA 102D
Acetaldehyde	75-07-0	2538, 2539, 3507	68	EPA TO-5
Acetic acid	64-19-7	1603		EPA TO-17
Acetic anhydride	108-24-7	3506	102	
Acetone	67-64-1	1300, 2549	69	EPA TO-5, TO-17; MASA 834;
Acetone cyanohydrin	75-86-5	2506		
Acetonitrile	75-05-8	1606		EPA TO-17
Acetylene	74-86-2			EPA TO-14A, TO-15
Acetylene dichloride	540-59-0	1003		
Acetylene tetrabromide	79-27-6	2003		
Acetylene tetrachloride	79-34-5	1019		
Acrolein	107-02-8	2501, 2539	52	EPA TO-5
Acrylamide	79-06-1		21	
Acrylic acid	79-10-7		28	
Acrylonitrile	107-13-1	1604	37	EPA TO-2, TO-17;MDHS 1(rev), MDHS 2(rev), 55
Alachlor	15972-60-8			ASTM D-4861
Aldicarb®	116-03-1		74	
Aldrin	309-00-2	5502		EPA TO-4A, TO-10A; ASTM D-4861
Allyl alcohol	107-18-6	1402		
Allyl chloride	107-05-1	1000		EPA TO-2
Allyl glycidyl ether	106-92-3	2545		MASA 834
Allyl trichloride	96-18-4	1003		
Amines, aliphatic	–	2010		
Amines, aromatic	–	2002		MASA 827
Aminobenzene (aniline)	62-53-3	2002		
4-Aminodiphenyl	92-67-1		93	
2-Aminoethanol	141-43-5	2007		
2-Aminotoluene (*o*-toluidine)	95-53-4	2002		
n- and *sec*-Amyl alcohol	71-41-0; 6032-29-7			EPA TO-17; ASTM D3686
n-Amyl acetate	628-63-7	1450, 2549		
sec-Amyl acetate	626-38-0	1450, 2549		
Aniline (aminobenzene)	62-53-3	2002		MASA 827
o-,*p*-Anisidine	90-04-0 104-94-9			MASA 827
Anthracene	120-12-7	5515	58	EPA TO-13A; MASA 102D
Aroclor™ 1242	55469-21-9			ASTM D-4861
Aroclor™ 1254	11097-69-1			ASTM D-4861
Aroclor™ 1260	11096-82-5			ASTM D-4861
Atrazine	1912-24-9			ASTM D-4861
Azelaic acid	123-99-9	5019		
Azinphosmethyl	86-50-0	5600		
Benz[*a*]anthracene	56-55-3	5515		EPA TO-13A; MASA 102D
Benzaldehyde	100-52-7	2549		EPA TO-5
Benzene	71-43-2	2549, 1501, 1500	65	EPA TO-1, TO-2, TO-3, TO-14A, TO-15, TO-17: MDHS 17, 22 (rev), 50; MASA 834; ASTM 5466
Benzene hexachloride	4 isomers			EPA TO-10A; ASTM D-4861
Benzidine (in urine)	92-87-5	8306		
Benzidine	92-87-5		65	
Benzo[*a*]pyrene	50-32-8	5515	58	EPA TO-13A; MASA 102D
Benzo[*b*]fluoranthene	205-99-2	5515		EPA TO-13A; MASA 102D
Benzo[*e*]pyrene	192-97-2	5515		
Benzo[*k*]fluoranthene	207-08-9	5515		EPA TO-13A; MASA 102D

Table 1 (*continued*) vfill

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Benzo[*ghi*]perylene	192-24-2	5515		EPA TO-13A; MASA 102D
Benzyl chloride	100-44-7	1003	07	EPA TO-14A, TO-15; ASTM D 5466
Biphenyl (diphenyl)	92-52-4	2530		
Bis(chloromethyl ether)	542-88-1		10	MASA 828, 829
Bromobenzene	108-86-1			EPA TO-14A, TO-15
Bromochloromethane	74-97-5			EPA TO-14A, TO-15
Bromodichloromethane	75-27-4			EPA TO-14A, TO-15
o-,*m*-,*p*-Bromofluorobenzene	1072-85-1 1073-06-9 460-00-4			EPA TO-14A, TO-15
Bromoform	75-25-2	1003	07	EPA TO-1
1,3-Butadiene	106-99-0	1024	56	EPA TO-14A, TO-15; MDHS 53, 63
n-Butamine	109-73-9	2012		
n-Butanal				EPA TO-17
n- and *i*-Butane	106-97-8 75-28-5			EPA TO-17; MASA 101
2-Butanone	78-93-3	2500	84	MASA 834
2-Butoxyethanol	111-76-2	2549, 1403	83	EPA TO-17; MASA 834; MDHS 21, 23
2-Butoxyethylacetate	112-07-2			EPA TO-17; MDHS 21, 23
n-Butyl acetate	123-86-4	1450	07	EPA TO-17
sec-Butyl acetate	105-46-4	1450	07	
tert-Butylacetate	540-88-5	1450	07	EPA TO-17
n-Butyl alcohol	71-36-3	1401	07	EPA TO-17; MASA 834
sec-Butyl alcohol	105-46-4	1401	07	MASA 834
tert-Butyl alcohol	540-88-5	1400	07	MASA 834
1,3-Butylene glycol	107-88-0	5523		
n-Butyl glycidyl ether	2426-08-6	1616	07	MASA 834
n-Butyl mercaptan	109-79-5	2525, 2542		
p-tert-Butyltoluene	98-51-1	1501	07	
Butyraldehyde	123-78-8	2539		
Camphor	76-22-2	1301	07	
Captan	133-06-2			EPA TO-10A; ASTM D-4861
Carbaryl (Sevin™)	63-25-2	5006	63	
Carbon dioxide	124-38-9	6603		MASA 133
Carbon disulfide	75-15-0	1600		MDHS 17, 22(rev)
Carbon monoxide	630-08-0			MASA 133
Carbon tetrachloride	56-23-5	1003	07	EPA TO-1, TO-2, TO-14A, TO-15, TO-17; MASA 834; MDHS 28; ASTM D 5466
3-Carene	13466-78-9	1552		
Chlordane	57-74-9	5510	67	EPA TO-4, TO-10A; ASTM D 4947-94, D 4861
Chlorinated camphene	8001-35-2	5039		
Chlorinated diphenyl oxide	55720-99-5	5025		
Chlorinated hydrocarbons	–			MDHS 28(Rev)
Chlorinated terphenyl	11126-42-4	5014		
Chloroacetaldehyde	107-20-0	2015	76	
o-Chloroaniline	95-51-2			MASA 827
Chlorobenzene	108-90-7	1003	07	EPA TO-1, TO-3, TO-14A, TO-15, TO-17; MASA 834; MDHS 28; ASTM D 5466
Chlorobromomethane	74-97-5	1003		
Chlorodifluoromethane	75-45-6	1018		
Chlorodiphenyl (42% 54%) (Aroclor™ 1242)	53469-21-9	5503		EPA TO-4A; ASTM D 4861

(*continued overleaf*)

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
1-Chloro-2,3-epoxy-propane (epichlorohydrin)	106-89-8			MASA 834
2-Chloroethanol	107-07-3	2513		
Chloroform	67-66-3	1003	05	EPA TO-1, TO-2, TO-3, TO-14A, TO-15; MASA 834; MDHS 28; ASTM D 5466
Chloromethyl methyl ether	107-30-2		10	MASA 829
4-Chloronitrobenzene	100-00-5	2005		
Chloroprene (2-chloro-1,3-butadiene)	126-99-9	1002	07	EPA TO-14A, TO-15
Chlorpyrifos	2921-88-2	5600	62	EPA TO-10A; ASTM D 4861
Chlorothalonil	2136-79-0			EPA TO-10A; ASTM D 4861
Chrysene	218-01-9	5515	58	EPA TO-13A; MASA 102D
Coal tar naphtha	8030-30-6	1550		
Coal tar pitch volatiles	65996-93-2		58	
Coke oven emissions	–		58	
Colophony	8050-09-7			MDHS 83
o-,*m*-,*p*-Cresol	95-48-7; 108-39-4; 106-44-5	2546, 2549	32	MASA 121
p-Cresol (in urine)	106-44-5	8305		
Crotonaldehyde	123-73-9	2539, 3516	81	EPA TO-5
Cumene	98-82-8	1501	07	EPA TO-1
Cyclohexane	110-82-7	1500	07	MASA 834
Cyclohexanol	108-93-0	1402	07	
Cyclohexanone	108-94-1	1300, 2549	01	EPA TO-5, TO-17; MASA 834
Cyclohexene	110-83-8	1500	07	
1,3-Cyclopentadiene	542-92-7	2523		
2,4-D(2,4-Dichloro-phenoxyacetic acid)	94-75-7	5001		
2,4-D Esters	–			EPA TO-10A; ASTM D 4861
1,1-Dichloro-2,2-bis(4-chlorophenyl)ethene	72-55-9			EPA TO-4, TO-10A; ASTM D 4861
Dacthal™ (chlorthal)	2136-79-0			ASTM D 4861
n-Decane	124-18-5	2549		EPA TO-17
Demeton	8065-48-3	5514		
Desflurane (1,2,2,2-tetrafluoroethyl-difluoromethyl ether)	57041-67-5		106	
Diacetone alcohol(4-hydroxy-4-methyl-2-pentanone)	123-42-2	1402	07	
o-Dianisidine	119-90-4		71	
Diazinon	333-41-5	5600	62	ASTM D 4861
Diazomethane	334-88-3	2515		
Dibenz[*a*,*h*]anthracene	53-70-3	5515		MASA 102D
Dibromochloromethane	124-48-1			EPA TO-14A, TO-15
1,2-Dibromoethane	106-93-4			EPA TO-14A, TO-15
Dibromodifluoromethane	75-61-6	1012		
2-Dibutylaminoethanol	102-81-8	2007		
Dibutylphosphate	107-66-4	5017		
Dibutylphthalate	84-74-2	5020		
o-Dichlorobenzene	95-50-1	2549, 1003	07	EPA TO-14A, TO-15; MASA 834; MDHS 28; ASTM D 5466
m-Dichlorobenzene	541-73-1			EPA TO-14A, TO-15; ASTM D 5466
p-Dichlorobenzene	106-46-6	2549, 1003	07	EPA TO-14A, TO-15; MASA 834; ASTM D 5466

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
3,3′-Dichlorobenzidine	91-94-1	5509	65	
Dichlorodifluoromethane	75-71-8	1018	07	ASTM D 5466
Dichlorodiphenyltrichloroethane	50-29-3			EPA: TO-4A, TO-10A
1,1-Dichloroethane (ethylidene chloride)	75-34-3	1003	07	EPA TO-14A, TO-15; MDHS 28; ASTM D 5466
1,2-Dichloroethane	107-06-2	1003	07	EPA TO-1, TO-2, TO-3, TO-14A, TO-15, TO-17; MASA 834; MDHS 28; ASTM D 5466
Dichloroethylether	111-44-4	1004	07	
(*cis*) 1,2-Dichloroethylene	156-59-2	1003	07	EPA TO-14A, TO-15; MDHS 28; ASTM D 5466
(*trans*) 1,2-Dichloroethylene	156-60-5			ASTM D 5466
Dichlorofluoromethane	75-43-4	2516		EPA TO-14A, TO-15
Dichloromethane (methylene chloride)	75-09-2	2549	80	EPA TO-14A, TO-15, TO-17; MASA 834
1,1-Dichloro-1-nitroethane	594-72-9	1601	07	
1,2-Dichloropropane	78-87-5	1013		EPA TO-1, TO-14A, TO-15
1,3-Dichloropropane	142-28-9			EPA TO-1
1,2-Dichloropropene	78-88-6			EPA TO-14A, TO-15; ASTM D 5466
1,3-Dichloropropene	542-75-6			EPA TO-14A, TO-15
1,2-Dichloro-1,1,2,2-tetrafluoroethane	76-14-2	1018		EPA TO-14A, TO-15; ASTM 5466
Dichlorvos	62-73-7		62	EPA TO-10A; ASTM D 4861
Dichrotophos	141-66-2	5600		
Dicloran	99-30-9			ASTM D 4861
Dicofol	115-32-2			EPA TO-10A; ASTM D 4861
Dieldrin	60-57-1			EPA TO-10A; ASTM D 4861
Diethylamine	109-89-7	2010	41	
2-Diethylaminoethanol	100-37-8	2007		
Diethylene glycol	111-46-6	5523		
Diethyl ether	60-29-7			MASA 834
Di(2-ethylhexyl) phthalate	117-81-7	5020		MDHS 32
Diethyldisulfide	110-81-6			MASA 709A
Diethylsulfide	352-93-2			MASA 709A
Difluorodibromomethane	75-61-6	1012, 1018	07	
Difluorodichloromethane	75-71-8	1018		
Diisobutyl ketone	108-83-8	1300	07	
Diisooctyl phthalate	27554-26-3			MDHS 32
N,*N*-Dimethylacetamide	127-19-5	2004		
Dimethylamine	124-40-3	2010	34	
N,*N*-Dimethylaniline	121-69-7	2002	07	MASA 827, 834
N,*N*-Dimethylformamide	68-12-2	2004	66	
2,6-Dimethylpyridine	108-48-5			MASA 834
N,*N*-Dimethyl-*p*-toluidine	99-97-8	2002		
Dimethyl sulfate	77-78-1	2524		
Dinitrotoluene	121-14-2		44	
1,4-Dioxane	123-91-1	1602	07	MASA 834
Diphenyl	92-52-4	2530		
Diphenylamine	122-39-4		44	
Dipropylene glycol methyl ether	34590-94-8		07, 101	
Disulfoton	298-04-4	5600		
Divinylbenzene	1321-74-0		89	
Endrin	72-20-8	5519		EPA TO-10A
Endrin aldehyde	7421-93-4			EPA TO-10A
Enflurane	13838-16-9		29	

(*continued overleaf*)

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Epichlorohydrin	106-89-8	1010	07	
o-Ethyl-*o*-(*p*-nitrophenyl) phenylphosphothioate	2104-64-5	5012		MASA 835
1,2-Epoxypropane	75-56-9	1512		
Ethane	74-84-0			MASA 101
Ethanolamine	141-43-5	2007		
3-Ethenylpyridine (3-vinylpyridine)	100-43-6			ASTM D 5075-96
Ethion	563-12-2	5600		
Ethoprop	13194-48-4	5600		
2-Ethoxyethanol	110-80-5	1403	79	EPA TO-17; MDHS 21, 23
2-Ethoxyethanol acetate	111-15-9	1450	79	EPA TO-17; MDHS 21, 23
Ethyl acetate	141-78-6	1457	07	EPA TO-17; MASA 834
Ethyl acrylate	140-88-5	1450	92	EPA TO-17
Ethyl alcohol (ethanol)	64-17-5	2549, 1400	100	EPA TO-17
Ethylamine	75-04-7		36	
Ethyl amyl ketone (3-octanone)	106-68-3	1301		MASA 834
Ethyl *sec*-amyl ketone (5-methyl-3-heptanone)	541-85-5		07	
Ethyl benzene	100-41-4	1501	07	EPA TO-1, TO-14A, TO-15, TO-17; MASA 834; ASTM D 5466
Ethyl bromide	74-96-4	1011	07	EPA TO-14A, TO-15
Ethyl butyl ketone (3-heptanone)	106-35-4	1301	07	MASA 834
Ethyl chloride	75-00-3	2519		EPA TO-14A, TO-15
Ethyl-2-cyanoacrylate	7085-85-0		55	
Ethylene chlorohydrin	107-07-3	2513	07	
Ethylene diamine	107-15-3		60	
Ethylene dibromide	106-93-4	1008	02	EPA TO-14A, TO-15; MDHS 45; ASTM D 5466
Ethylene dichloride (1,2-dichloroethane)	107-06-2	1003	03	
Ethylene glycol	107-21-1	5523		
Ethylene glycol dinitrate	628-96-6	2507	43	MASA 832
Ethylene oxide	75-21-8	1614, 3702	49, 50	ASTM D 4413-88, D 5578-94; MDHS 26
Ethylene thiourea	96-45-7		95	
Ethyl ether	60-29-7	1610	07	
Ethyl-2-ethoxyacetate	817-95-8			MASA 834
Ethyl formate	109-94-4	1452	07	MASA 834
Bis(2-ethylhexyl)phthalate	117-81-7			MDHS 32
Ethylidene chloride (see 1,1-dichloroethane)	75-34-3			
Ethyl mercaptan	75-08-1	2542		MASA 709A
Ethyl-3-methylbutanoate (ethyl isovalerate)	108-64-5			MASA 834
Ethyl-2-methylpropanoate	1067-08-9			MASA 834
Ethyl parathion	56-38-2			EPA TO-10A; ASTM D 4861
Ethyl pentanoate	539-82-2			MASA 834
Ethyl propanoate	105-37-3			MASA 834
o-,*m*-,*p*-Ethyl toluene	611-14-3			EPA TO-14A, TO-15
	620-14-4			
	622-96-8			
Ethylvinylbenzene	28106-30-1		89	
Fenamiphos	22224-92-6	5600		
Fluoranthene	206-44-0	5515		MASA 102D
9*H*-Fluorene	86-73-7	5515		MASA 102D
Fluorotrichloromethane	75-69-4	1006		

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Folpet	133-07-3			EPA TO-10A; ASTM D 4861
Fonofos	66767-39-3	5600		
Formaldehyde	50-00-0	2541	52	EPA TO-5
Furfural	98-01-1	2539	64	EPA TO-17; MASA 834
Furfuryl alcohol	98-00-0	2508		
Glutaraldehyde	111-30-8	2532	64	
Glycidol (2,3-epoxy-1-propanol)	556-52-5	1608	07	
Glycol ethers	–			MDHS 21, 23
Glycol ether acetates	–			MDHS 21
Halothane	1151-67-7		29	
Heptachlor	76-44-8			EPA TO-10A; ASTM D 4947-94. D 4861
Heptachlor epoxide	1024-57-3			EPA TO-10A; ASTM D 4861
Heptanal	111-71-7	2539		
n-Heptane	142-82-5	1500	07	EPA TO-1; MASA 834
3-Heptanone	106-35-4			
1-Heptene	592-76-7			EPA TO-1
Hexachlorobenzene	118-74-1			ASTM D 4861
Hexachlorobutadiene	87-68-3	2543		EPA TO-14A, TO-15; ASTM D 5466
Hexachloro-1,3-cyclopentadiene	77-47-4	2518		ASTM D 4861
Hexachlorocyclohexane (benzene hexachloride)	4 isomers			EPA TO-10A, 608; ASTM D-4861
Hexachloroethane	67-72-1	1003	07	
Hexafluorobenzene	392-56-3			EPA TO-14A, TO-15
Hexanal	66-25-1	2539, 2549		
n-Hexane	110-54-3	1500, 2549	07	EPA TO-1; MASA 834; MDHS 74
2-Hexanone (MBK)	591-78-6	1300	07	
Hexone (MIBK)	108-10-1	1300, 2549	07	
sec-Hexyl acetate	108-84-9		07	
Hydrocarbons, mixed	–			MDHS 60
Hydrogen sulfide	7783-06-4			MASA 709A
Indeno[1,2,3-*cd*] pyrene	193-39-5	5515		EPA TO-13A; MASA 102D
Isoamyl acetate	123-92-2	1450	07	
Isobutyl acetate	110-19-0	1450	07	EPA TO-17
Isobutyl alcohol	78-83-1	1401	07	EPA TO-17
Isophorone	78-59-1	2506	07	MASA 834
Isopropyl acetate	108-21-4	1454	07	EPA TO-17
Isopropyl alcohol	67-63-0	1454	07, 109	EPA TO-17
N-Isopropylaniline	768-52-5		78	
Isopropyl ether	108-20-3	1618	07	
Isopropyl glycidyl ether	4016-14-2	1620	07	MASA 834
Isovaleraldehyde	590-86-3	2539		
Kepone	143-50-0	5508		
Kerosene	8008-20-6	1550		
Limonene	138-86-3	2549		
Lindane	58-89-9	5502		EPA TO-4A, TO-10A; ASTM D 4861
Malathion™	121-75-5	5600	62	EPA TO-10A; MASA 835; ASTM D 4861
Maleic anhydride	108-31-6			EPA TO-17
Maneb	12427-38-2		107	
Mercaptans	–	2542	26	
Mesityl oxide	141-79-7	1301	07	
Methamidophos	10265-92-6	5600		
Methane	74-82-8			MASA 101

(*continued overleaf*)

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Methoxychlor	72-43-5			EPA TO-10A; ASTM D 4861
2-Methoxyethanol	109-86-4	1403	79	EPA TO-17
2-Methoxyethanol acetate	110-49-6	1451	79	EPA TO-17
1-Methoxy-2-propanol	107-98-2		99	EPA TO-17
2-Methoxy-1-propanol	1589-47-5		99	EPA TO-17
1-Methoxy-2-propyl acetate	108-65-6		99	
2-Methoxy-1-propyl acetate	70657-70-4		99	
Methyl acetate	79-20-9	1458	07	EPA TO-17
Methyl acetylene-propadiene mixture	59355-75-8		07	
Methyl acrylate	96-33-3	1459	92	EPA TO-17
Methylal (dimethoxymethane)	109-87-5	1611	07	
Methyl alcohol (methanol)	67-56-1	2549, 2000	91	EPA TO-17
Methylamine	74-89-5		40	
Methyl-(*n*-amyl) ketone (2-heptanone)	110-43-0	1301		
Methyl bromide	74-83-9	2520		EPA TO-14A, TO-15; ASTM D 5466
Methyl-*t*-butyl ether	1634-04-4	1615		
Methylbutyl ketone (2-hexanone)	591-78-6	1300		MASA 834
Methyl cellosolve	109-86-4	1403		EPA TO-17
Methyl cellosolve acetate	110-49-6	1451		
Methyl chloride	74-87-3	1001		EPA TO-14A, TO-15; ASTM D 5466
Methyl chloroform (1,1,1-Trichloroethane)	71-55-6	2549		EPA TO-14A, TO-15; ASTM D 5466
Methyl cyanide (acetonitrile)	75-05-8	1606		
Methyl-2-cyanoacrylate	137-05-3		55	
Methylcyclohexane	108-87-2	1500	07	
4,4′-Methylenebis(2-chloroaniline) (in urine)	101-14-4	8302		
4,4′-Methylenebis(2-chloroaniline)	101-14-4		71	
Methylenebis(4-phenylisocyanate)	101-68-8		47	
Methylene chloride	75-09-2	2549	80	EPA TO-2; MDHS 28; ASTM D 5466
4,4′-Methylenedianiline	101-77-9		57	
1-Methyl-2-ethylbenzene	611-14-3			EPA TO-17
1-Methyl-3-ethylbenzene	620-14-4			EPA TO-17
1-Methyl-4-ethylbenzene	622-96-8			EPA TO-17
Methyl ethyl ketone (2-butanone)	78-93-3	2500, 2549		EPA TO-17
Methylethylketone peroxide	1338-23-4	3506	77	
5-Methyl-3-heptanone	541-85-5	1301		
Methyl iodide	74-88-4	1014		
Methyl isoamyl acetate	108-84-9	1450		
Methyl isobutyl carbinol	108-11-2	1402	07	
Methyl isobutyl ketone	108-10-1	2549		EPA TO-17; MASA 834
Methyl isocyanate	624-83-9		54	
Methyl mercaptan	74-93-1	2542	26	MASA 709A
Methyl methacrylate	80-62-6	2537	94	EPA TO-17
Methyl-2-methyl-propanoate	547-63-7			MASA 834
Methyl propenoate	96-33-3			MASA 834
α-Methylstyrene	98-83-9	2501	07	MASA 834
4-Methylstyrene	622-97-9			EPA TO-17; MASA 834
Methyl parathion	298-00-0			EPA TO-10A; ASTM D 4861
Methyl-*tert*-butyl ether	1634-04-4	1615		

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Metolachlor	51218-45-2			ASTM D 4861
Mevinphos (phosdrin)	7786-34-7	5600		
Mexacarbate	315-18-4			EPA TO-10A; ASTM D 4861
Mineral spirits	8052-41-3	1550	48	
Mirex	2385-85-5			EPA TO-10A; ASTM D 4861
Monochrotophos	6923-22-4	5600		
Monomethylaniline	100-61-8	3511		
Naphtha (coal tar)	8030-30-6	1550		
Naphthalene	91-20-3	1501, 5515	35	EPA TO-13A
1-Naphthylamine	134-32-7	5518	93	
2-Naphthylamine	91-59-8	5518	93	
Nicotine	54-11-5	2544		ASTM D 5075-96
p-Nitroaniline	100-01-6			MASA 827
Nitrobenzene	98-95-3	2005		EPA TO-17
p-Nitrochlorobenzene	100-00-5	2005		
Nitroethane	79-24-3	2526		
Nitrogen	7727-37-9			MASA 133
Nitroglycerin	55-63-0	2507	43	MASA 832
Nitromethane	75-52-5	2527		
1-Nitropropane	108-03-2		46	
2-Nitropropane	79-46-9	2528	46	
Nitrosamines	–	2522	27, 31, 38, 23	EPA TO-7; MASA 833
trans-Nonachlor	3734-49-4			EPA TO-10A; ASTM D 4861
n-Nonane	111-84-2			EPA TO-17
Octamethylcyclotetrasiloxane	556-67-2	2549		
n-Octane	111-65-9	1500, 2549	07	EPA TO-17; MASA 834
1-Octanethiol	111-88-6	2510		
Octanol	111-87-5			EPA TO-17
Organochlorine pesticides	–			EPA TO-4A
Organophosphorus pesticides	–	5600	62	
Oxychlordane	27304-13-8			EPA TO-10A; ASTM D 4861
Oxygen	7782-44-7			MASA 133
Parathion	56-38-2	5600	62	MASA 835
Pentachlorobenzene	608-93-5	5517		EPA TO-10A; ASTM D 4861
Pentachloroethane	76-01-7	2517		
Pentachlorophenol	87-86-5	5512	39	EPA TO-10A; ASTM D 4861
Pentanal	110-62-3			EPA TO-5
n-Pentane	109-66-0	1500, 2549	07	EPA TO-17; MASA 101, 121, 834
2-Pentanone	107-89-9	1300	07	MASA 834
Perchloroethylene	127-18-4	2549		ASTM D 5466
Perfluorobenzene	392-56-3			EPA TO-14A, TO-15
Perfluorotoluene	434-64-0			EPA TO-14A, TO-15
Petroleum ether/naphtha (VM and P naptha)	8032-32-4	1550		
Petroleum distillate fractions	–		48	
Phenanthrene	85-01-8		58	
Phenol	108-95-2	2546, 2549	32	EPA TO-17; MASA 121
Phenol (in urine)	108-95-2	8305		
o-,*m*-,*p*-Phenylenediamine	108-45-2; 95-54-5; 106-50-3		87	
Phenyl ether (diphenyl oxide)	101-84-0	1617	07	
Phenyl glycidyl ether	122-60-1	1619	07	
N-Phenyl-1-naphthylamine	90-30-2		96	
N-Phenyl-2-naphthylamine	135-88-6		96	
Phorate	298-02-2	5600		ASTM D 4861
Phosdrin (mevinphos)	7786-34-7	5600		
Phosgene	75-44-5		90	

(*continued overleaf*)

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Phosphorus	7723-14-0	7905		
α-Pinene	80-56-8	1552, 2549		MASA 834
β-Pinene	127-91-3			
PAHs	–	5515		
Polychlorobiphenyl	1336-36-3	5503		EPA TO-4A; ASTM D 4861-94a
Propane	74-98-6			MASA 101
Propazine	139-40-2			ASTM D 4861
Propenenitrile	107-13-1			MASA 834
Propionaldehyde	123-38-6	2539		EPA TO-5
Propionitrile	107-12-0			EPA TO-17
n-Propyl acetate	109-60-4	1450	07	EPA TO-17
i-Propyl alcohol	67-63-0		109	EPA TO-17
n-Propyl alcohol	71-23-8	1401	07	EPA TO-17
n-Propylbenzene	103-65-1			EPA TO-17
Propylene dichloride (1,2-dichloropropane)	78-87-5	1013	07	ASTM D 5466
Propylene glycol	57-55-6	5523		
Propylene oxide	75-56-9	1612	88	
n-Propyl nitrate	627-13-4		07	
Propargyl alcohol	107-19-7		97	
Pyrene	129-00-0	5515	70	EPA TO-13; MASA 102D
Pyrethrum	8003-34-7	5008	70	
Pyridine	110-86-1	1613	07	EPA TO-17; ASTM D 4861
Ronnel	299-84-3	5600		EPA TO-10A
Stoddard solvent	8052-41-3	1550	48	
Styrene	100-42-5	1501	89	EPA TO-14A, TO-15, TO-17; MDHS 20, 31(rev), 43, 44; MASA 834
Sulfur dioxide	7446-09-5			MASA 709A
Sulfur hexafluoride	2551-62-4	6602		
Sulprofos	35400-43-2	5600		
Terbufos	13071-79-9	5600		
Terpenes	–	1552		
o-Terphenyl	85-15-1	5021		
1,1,2,2-Tetrabromoethane	79-27-6	2003		
1,2,3,4-Tetrachlorobenzene	634-66-2			ASTM D 4861
1,2,4,5-Tetrachlorobenzene	634-90-2	5617		
1,1,1,2-Tetrachloro-2,2-difluoroethane	76-11-9	1016	07	
1,1,2,2-Tetrachloro-1,2-difluoroethane	76-12-0	1016	07	
1,1,1,2-Tetrachloroethane	630-20-6			EPA TO-17
1,1,2,2-Tetrachloroethane	79-34-5	1019		EPA TO-14A, TO-15, TO-17; MASA 834; MDHS 28; ASTM D 5466
Tetrachloroethylene	127-18-4	1003		EPA TO-1, TO-3, TO-14A, TO-15, TO-17; MDHS 28
2,3,4,6-Tetrachlorophenol	58-90-2		45	
Tetrachlorophthalic anhydride	117-08-8			MDHS 62
Tetraethylene glycol	112-60-7	5523		
Tetraethyl lead	78-00-2	2533		
Tetraethylpyrophosphate	107-49-3	2504		
Tetraethyl tin (tetraethylstannane)	597-64-8		110	
Tetrahydrofuran	109-99-9	1609	07	MASA 834
Tetramethyl lead	75-74-1	2534		
Tetramethylsuccinonitrile	3333-52-6		07	
Tetranitromethane	509-14-8	3513		

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
Thiram	137-26-8	5005		
o-Tolidine	119-93-7		71	
Toluene	108-88-3	1500, 1501, 4000	07	EPA TO-1, TO-2, TO-14A, TO-15, TO-17; MDHS 36(rev), 40(rev), 64; MASA 834; ASTM D 5466
2,4- and 2,6-Toluenediamine	95-80-7; 823-40-5		65	
o-,*m*- and *p*-Toluidine	95-53-4; 108-44-1; 106-49-0	2002	73	MASA 827
Tribromomethane	75-25-2	1003		
Tributyl phosphate	126-73-8	5034		
1,2,3- or 1,2,4-Trichlorobenzene	120-82-1	5517		EPA TO-14A, TO-15; ASTM D 4861, D 5466
1,1,1-Trichloro-2,2-bis-(chlorophenyl)ethane	50-29-3			ASTM D 4861
1,1,1-Trichloroethane (methylchloroform)	71-55-6	1003, 2549	14	EPA TO-1, TO-2, TO-3, TO-14A, TO-15, TO-17; MASA 834; MDHS 28
1,1,2-Trichloroethane (vinyl trichloride)	79-00-5	1003	11	EPA TO-14A, TO-15, TO-17; MASA 834; MDHS 28; ASTM D 5466
Trichloroethylene	79-01-6	1022	07	EPA TO-1, TO-3, TO-14A, TO-15, TO-17; MDHS 28; ASTM D 5466
Trichlorofluoromethane	75-69-4	1006		EPA TO-14A, TO-15; ASTM D 5466
Trichloromethane (chloroform)	67-66-3	1003	05	MASA 834
2,4,5-Trichlorophenol	88-06-2			EPA TO-10A; ASTM D 4861
1,2,3-Trichloropropane	96-18-4	1003	07	
1,1,2-Trichloro-1,2,2-trifluoroethane	76-13-1	1020, 2549		EPA TO-14A, TO-15; MASA 834: ASTM D 5466
Triethyleneglycol	112-27-6	5523, 2540		
Trifluorobromomethane	75-63-8	1017		
Trifluralin	1582-09-8			ASTM D 4861
Trimellitic anhydride	552-30-7	5036	98	
1,2,3-Trimethylbenzene	526-73-8			EPA TO-17
1,2,4-Trimethylbenzene (pseudocumene)	95-63-6			EPA TO-14A, TO-15, TO-17; ASTM D 5466
1,3,5-Trimethyl-benzene(mesitylene)	108-67-8			EPA TO-17; ASTM D 5466
3,5,5-Trimethylcyclohex-2-eneone	78-59-1			EPA TO-17
2,4,6-Trinitrotoluene	118-96-7		60	
tri-*o*-Cresyl phosphate	78-30-3	5037		
tri-*o*-Tolyl phosphate	78-30-8			
Triphenyl phosphate	115-86-6	5038		
Turpentine	8006-64-2	1551		
n-Undecane	1120-21-4			EPA TO-17
Valeraldehyde	110-62-3	2536, 2539	85	
Vinyl acetate	108-05-4	1453	51	
Vinyl benzene (styrene)	199-42-5	1501		ASTM D 5466
Vinyl bromide	593-60-2	1009	08	
Vinyl chloride	75-01-4	1007	75	EPA TO-2, TO-14A, TO-15; ASTM D 4766-88; MDHS 24; ASTM D 5466
Vinylidene chloride	75-35-4	1018	19	EPA TO-2, TO-14A, TO-15; ASTM D 5466
Vinyl toluene	25013-15-4	1501		

(*continued overleaf*)

Table 1 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Other[c]
VM and P naphtha	8032-32-4	1550	48	
Volatile organic compounds	–	2543, 2549		
o-,*m*-,*p*-Xylene	95-47-6; 108-38-3; 106-42-3; 1330-20-7 (mixed)	1501, 2549	07	EPA TO-1, TO-14A, TO-15, TO-17; MASA 834; ASTM D 5466
2,4-Xylidine	1300-73-8	2002		MASA 827
Zineb	12122-67-7		107	

[a] NIOSH Manual of Analytical Methods.[21]
[b] OSHA Analytical Methods.[22]
[c] Other Methods: EPA TO-Methods,[23] MASA: Methods of Air Sampling and Analysis,[26] MDHS: HSE, Methods for the Determination of Hazardous Substances.[28] ASTM: 1997 Annual Book of ASTM Standards; section 11.[25]

micromachining. An excellent review article by Manz, et al.[74] discusses these issues, and an overview article by Marshall[75] gives some idea of what lies ahead with regard to microchromatography, both with regard to columns and detectors. These developments begin to make possible truly hand-held GC, which have become available, and will certainly be common in the not-too-distant future.

5 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

The development of HPLC for the separation of nonvolatile or low-volatility organic compounds led to the rapid application of such methods for industrial hygiene monitoring. One of the earliest uses of this technique was in the separation and analysis of PAHs, which can also be analyzed using gas chromatographic methods. The sensitivity of detection for many of the PAHs by using a fluorescence detector in combination with liquid chromatographic separation has proved to be highly desirable, in view of the carcinogenicity of a number of PAHs. More recent developments which utilize HPLC include the analysis of isocyanate derivatives formed when the airborne isocyanate is derivatized by contact in an impinger or on a glass fiber filter, with 2-pyridyl piperazine or several other common derivatizing agents. The analysis is commonly done using either UV or fluorescence detection. A number of airborne aldehydes are commonly derivatized with 2,4-dinitrophenylhydrazine (typically in a solvent in an impinger, or coated on a solid substrate in a sorbent tube) and the derivatives analyzed using HPLC.

5.1 Ultraviolet/Visible Radiation Detection

UV/VIS light detection of HPLC effluents is a very common technique which has gained widespread use in industrial hygiene applications of this chromatographic technique. A method for the simultaneous analysis of organonitrogen pesticides[76] is a good example. Another common use of the technique and detector is for the analysis of isocyanate derivatives prepared in the course of measuring a variety of isocyanates in workplace air.[77] In the area of biological monitoring Sanderson et al.[78] have reported on an analytical method for the determination of metabolites of the herbicide alachlor in urine using HPLC with UV detection as one of the analytical methods. Another biomonitoring application of HPLC with UV related to isocyanates has been published by Dalene et al.[79]

5.2 Fluorescence Detection

HPLC with a fluorescence detector has been used for the determination of exposure of workers to PAHs during aluminum smelting as reported by Ny et al.[80] The monitoring was based on the analysis of 1-hydroxypyrene as a marker for PAH exposure, which was reported elsewhere.[81] In seeking to identify a single method for workplace air analysis of a variety of isocyanates, Streicher et al.[82] have used both fluorescence and UV detectors to analyze a number of different isocyanate derivatives in order to determine total isocyanates (including nonmonomers) in an air sample. Using four different derivatizing agents, 1-(9-anthracenylmethyl) piperazine (MAP), 1-(2-methoxyphenyl) piperazine (MOPP), 9-(methylaminomethyl)anthracene (MAMA) and TRYP, HPLC separation of the derivatives and quantitation using both methods was done using phenyl isocyanate, 1,6-hexamethylene diisocyanate, TDI MDI and butyl isocyanate. An interesting application for a high speed HPLC analysis of a complex mixture of PAHs using fluorescence detection was reported by Dong et al.[83]

5.3 Mass Spectrometric Detection

For those situations in which traditional detection alone was not satisfactory for the identification and quantitation

of compounds being analyzed in industrial hygiene samples, mass spectral detectors were developed which would enable the analyst to identify the eluted compound by its fragmentation pattern-matched against a library of fragmentation patterns stored in the memory of a computer. Three different methods of introducing the eluted sample into the mass spectrometer have been commercialized: moving belt interface, direct liquid injection, and thermospray. These are discussed elsewhere in detail (see **Liquid Chromatography/Mass Spectrometry**). Liquid chromatography/mass spectrometry (LC/MS) has been used in the analysis of chlorinated herbicides which is useful in the identification of indoor air contamination in residences which had historically or unintentionally been treated with such compounds.[84] It has also been used in biomonitoring for exposure to a series of phenolic compounds.[85] A method that combines HPLC and MS, along with GC for the quantitation of PAHs in air particulates, of considerable interest to industrial hygienists in coal gasification and liquefaction and coking operations, was reported by Wise et al.[86]

5.4 Atomic Absorption Spectrometry

A technique which is used with some frequency combines HPLC with an atomic absorption analysis. The compounds of interest are separated by HPLC and then analyzed, on a continuous basis, by an atomic absorption technique (see **Flame and Vapor Generation Atomic Absorption Spectrometry** and **Graphite Furnace Atomic Absorption Spectrometry**), either flame or flameless. An example of this is the separation of arsenic-containing species in urine using HPLC and analysis using hydride-generation atomic absorption spectrometry (AAS) for analysis.[87]

5.5 Inductively Coupled Plasma Spectroscopy

Another technique reported in the literature combines HPLC with ICP (inductively coupled plasma) spectroscopy. In this case, HPLC is used to separate Cr(III) and Cr(VI), of considerable interest to industrial hygienists, and then ICP is used to quantitate chromium in each of the two oxidation states.[88]

5.6 Table of Methods

Table 2 summarizes HPLC methods found for airborne, surface, or biological specimens found in six different references.

6 ION CHROMATOGRAPHY

The development by Small et al.[14] of IC led to widespread use of this technique in industrial hygiene.

Table 2 High-performance liquid chromatographic methods for airborne and biological contaminants from six different references

Compound	CAS no.	NIOSH[a]	OSHA[b]	Others[c]
Acenaphthene	83-32-9	5506		MASA 102D
Acenaphthalene	208-96-8	5506		MASA 102D
Acetaldehyde	75-07-0	3507		EPA: TO-5; MASA 122
Acrolein	107-02-8			EPA: TO-5; MASA 122
Acrylamide	79-06-1			MDHS 57
3-Aminofluorene	6344-66-7			MASA 836
Aniline	62-53-3			MDHS 75
o-Anisidine	90-04-0	2514		
p-Anisidine	104-94-9	2514		
Anthracene	120-12-7	5506		MASA 836
Aroclor™ 1260 (60% chlorine)	11096-82-5			MASA 836
Aroclor™ 5432 (32% chlorine)	12642-23-8			MASA 836
Aromatic amines	–			MDHS 75
Aspartame	22839-47-0	5031		
Azulene	275-51-4			MASA 836
Bendiocarb	22781-23-3			ASTM D 4861
Benzaldehyde	100-52-7			MASA 836
Benz[*a*]anthracene	56-55-3	5506		MASA 102D, 836
Benzene	71-43-2			MASA 836
Benzidine	92-87-5	5509		
Benzidine dyes	–	5013		
4-*H*-Benzo(*def*)-carbazole	203-65-6			MASA 836
11-*H*-Benzo[*a*]-carbazole	239-01-0			MASA 836
Benzo[*a*]pyrene	50-32-8	5506		MASA 102D, 836

(*continued overleaf*)

Table 2 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Others[c]
Benzo[*b*]fluoranthrene	205-99-2	5506		MASA 102D, 836
11*H*-Benzo[*b*]fluorene	243-17-4			MASA 836
Benzo[*e*]pyrene	192-97-2	5506		MASA 102D, 836
Benzo[*k*]fluoranthrene	207-08-9	5506		MASA 102D, 836
Benzo[*ghi*]perylene	191-24-2	5506		MASA 102D, 836
Benzoyl peroxide	94-36-0	5009		
Bromoxynil	1689-84-5	5010		
Bromoxynil octanoate	1689-99-2	5010		
Butyraldehyde	123-72-8			MASA 122
Capsaicin	404-86-4	5041		
Carbaryl	63-25-2			ASTM D 4861
Carbazole	10605-21-7			MASA 836
Carbofuran	1563-66-2			ASTM D 4861
o-Chloroaniline	95-51-2			MDHS 75
p-Chlorophenol	106-48-9	2014		
Chlorotoluron	15545-48-9			ASTM D 4861
Chrysene	218-01-9	5506		MASA 102D, 836
Coal tar pitch volatiles	–		58	
Coronene	191-07-1			MASA 836
o-,*m*- and *p*-Cresol	95-48-7			EPA: TO-8
	108-39-4			
	106-44-5			
Cyanuric acid	108-80-5	5030		
1,1-Dichloro-2,2-bis(4-chlorophenyl) ethane	72-54-8			MASA 836
1,1-Dichloro-2,2-bis(4-chlorophenyl) ethene	72-55-9			MASA 836
2,4-D (Dichlorophenoxyacetic acid)	94-75-7	5001		
2,3,4,5,6,2′,3′,4′,5′,6′-Decachlorobiphenyl	2051-24-3			MASA 836
o-Dianisidine	119-90-4	5013		
Dibenz[*a*,*h*]anthracene	53-70-3	5506		MASA 102D, 836
7-*H*-Dibenzo[*c*,*g*]-carbazole	194-59-2			MASA 836
Dibenzo[*g*,*p*]chrysene	1919-68-4			MASA 836
Dibenzofuran	132-64-9			MASA 836
Dibenzo[*fg*,*op*]naphthacene	192-51-8			MASA 836
Dibenzothiophene	132-65-0			MASA 836
Dibutyltinbis(isooctylmercaptoacetate)	25168-24-5	5504		
3,3′-Dichlorobenzidine	91-94-1	5509		
Dicrotofos	141-66-2			ASTM D 4861
Diethylenetriamine	111-40-0	2540		
Dihydrocapsaicin	19408-84-5	5041		
7,12-Dimethyl-benz[*a*]anthracene	57-97-6			MASA 836
2,3-Dimethylindole	91-55-4			MASA 836
1,3-Dimethylpyrene	64401-21-4			MASA 836
3,6-Dinitrodibenzoselenophene	10941-23-8			MASA 836
2,5-Dinitrofluorene	15110-74-4			MASA 836
Diuron	330-54-1			ASTM D 4861
Dyes: benzidine-, *o*-toluidine-, *o*-dianisidine-based	–	5013		
N-Ethylcarbazole	86-28-2			MASA 836
Ethylenediamine	107-15-3	2540		
2-Ethylfluorene	1207-20-1			MASA 836
Ethylenimine	151-56-4	3514		
Fenvalerate	51630-58-1			ASTM D 4861
Fluoranthene	206-44-0	5506		MASA 102D
9-Fluorenal	51367-77-2			MASA 836
Fluometuron	2164-17-2			ASTM D 4861

Table 2 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Others[c]
Fluoranthene	206-44-0			MASA 836
9*H*-Fluorene	86-73-7	5506		MASA 102D, 836
Formaldehyde	50-00-0			EPA: TO-5, TO-11; ASTM D 5197-97; MASA 122; MDHS 78
Formaldehyde (on wood or textile dust)	50-00-0	5700		
Glutaraldehyde	111-30-8	2532		EPA: TO-5
Halowax™ 1099	39450-05-0			MASA 836
n-Heptylbenzene	1078-71-3			MASA 836
Hexamethylene diisocyanate	822-06-0	5521,5522		
n-Hexylbenzene	1077-16-3			MASA 836
Hippuric acid (in urine)	495-69-2	8301		
Hydrazine	302-01-2		108	
Hydroquinone	123-31-9	5004		
Indeno[1,2,3-*cd*]pyrene	193-39-5	5506		MASA 102D
Indole	120-72-9			MASA 836
Isocyanates	–	5521,5522		MDHS 25
Linuron	330-55-2			ASTM D 4861
Maleic anhydride	108-31-6	3512		
Maneb	12427-38-2		108	
9-Methylanthracene	2381-16-0			MASA 836
2-Methylcarbazole	3652-91-3			MASA 836
5-Methyl-5,10-dihydroindeno-[1,2-*b*]indole	7428-85-5			MASA 836
4,4′-Methylene-bisphenyl isocyanate	101-68-8	5521,5522		MDHS 75
4,4′-Methylene-dianiline	101-77-9	5029		
o-Methylhippuric acid (in urine)	42013-20-7	8301		
m-Methylhippuric acid (in urine)	27115-49-7	8301		
p-Methylhippuric acid (in urine)	27115-50-0	8301		
4-Methylpyrene	3353-12-6			MASA 836
Monuron	150-68-5			ASTM D 4861
Naphthacene	92-24-0			MASA 836
2-Naphthaldehyde	66-99-9			MASA 836
Naphthalene	91-20-3	5506		MASA 102D, 836
Naphtho[2,1,8,*qrs*]-naphthacene	196-42-9			MASA 836
p-Nitroaniline	100-01-6	5033		
2-Nitrofluorene	607-57-8			MASA 836
1,2,3,4,5,6,7,8-Octa-chlorodibenzofuran	39001-02-0			MASA 836
1,2,3,4,6,7,8,9-Octachlorodibenzo-*p*-dioxin	3268-87-9			MASA 836
Octachloronaphthalene	2234-13-1			MASA 836
Paraquat	4685-14-7	5003		
Pentachlorophenol	87-86-5	5512		
Pentamidine (isethionate)	140-64-7	5032		
cis- and *trans*-Permethrin	52645-53-1			ASTM D 4861
Perylene	195-55-0			MASA 836
Phenanthrene	85-01-8	5506		MASA 102D, 836
Phenol	108-95-2			EPA: TO-8
Phenoxathiin	262-20-4			MASA 836
N-Phenylcarbazole	1150-62-5			MASA 836
o-Phenylphenol	90-43-7			ASTM D 4861
Phosgene	75-44-5			EPA: TO-6
Phthalic anhydride	85-44-9		90	MDHS 62
Picene	213-46-7			MASA 836
PAHs	–	5506		EPA: TO-13; MASA 102D
Propionaldehyde	123-38-6			MASA 122
Propoxur (Baygon®)	114-26-1			ASTM D 4861

(*continued overleaf*)

Table 2 (*continued*)

Compound	CAS no.	NIOSH[a]	OSHA[b]	Others[c]
Pyrene	129-00-0	5506		MASA 102D, 836
Pyrethrum	121-21-1	5008		ASTM D 4861
Resmethrin	10453-86-8			ASTM D 4861
Ribavirin	36791-04-5	5027		
Rotenone	83-79-4	5007		
Simazine	122-34-9			ASTM D 4861
Strychnine	57-24-9	5016		
Tebuthiuron (Preflan)	34014-18-1			ASTM D 4861
Tetradecachloro-*p*-terphenyl	31710-32-4			MASA 836
2,4,5-T	93-76-5	5001		
Tetrabutyl tin	1461-25-2	5504		
Thiram	137-26-8	5005		
Tin, organic compounds	–	5504		
o-Tolidine	119-93-7	5013		
2,4- and 2,6-Toluenediamine	95-80-7 823-40-5	5516		
Toluene 2,4- and 2,6-diisocyanate	584-84-9 91-08-7	2535,5521, 5522		
o-Toluidine	95-53-4			MDHS 75
1,1-Trichloro-2,2-bis(chlorophenyl) ethane	50-29-3			MASA 836
Tributyltin chloride	688-73-3	5034		
Tricyclohexyltin hydroxide (Plictran)	13121-70-5	5504		
Triethylenetetramine	112-24-3	2540		
Trimellitic anhydride	552-30-7			MDHS 62
2,4,7-Trinitrofluoren-9-one	129-79-3	5016		
Triphenylene	217-59-4			MASA 836
Warfarin	81-81-2	5002		
Xanthine	69-89-6			MASA 836
Zineb	12122-67-7		107	

[a] NIOSH Manual of Analytical Methods.[21]
[b] OSHA Analytical Methods.[22]
[c] Other Methods: EPA TO-Methods,[23] MASA,[26] MDHS: HSE, Methods for the Determination of Hazardous Substances.[28] ASTM: 1997 Annual Book of ASTM Standards; section 11.[25]

Detectors for IC column effluents now include those based on the absorption of UV/VIS light, amperometry, fluorescence, flame emission, radioactivity, atomic absorption, and ICP spectroscopy. IC has allowed the individual determination of a mixture of airborne inorganic acids (NIOSH 7903) instead of together, as had been previously done using titrimetric methods. IC has been used, in conjuction with cold vapor AAS to speciate a mixture of arsenic-containing compounds, which are used as herbicides, pesticides, and wood preservatives.[89] Abuku et al.[90] described the analysis of bromide ions in biological fluids using IC with amperometric quantitation. A paper by Lorberau[91] described a method for the determination of both gaseous and particulate fluorides using IC. MASA[26] has an entire section (pp. 230–242) on IC applied to air monitoring. Table 3 shows IC methods found in NMAM, OSHA, and other common references.

7 OTHER CHROMATOGRAPHIC TECHNIQUES WITH PRESENT AND FUTURE INDUSTRIAL HEALTH APPLICATIONS

7.1 Supercritical Fluid Chromatography

The technique of supercritical fluid chromatography (SFC) has been described in detail elsewhere (see **Supercritical Fluid Chromatography**) and has application in industrial hygiene in the assessment of occupational exposure by means of biological monitoring. Although one author commented that SFC bridges the gap between GC and liquid chromatography (LC) and offers a feasible alternative for the analysis of PAHs, there has not been a lot of method development of this technique related to industrial hygiene. There are some early applications of SFC[92,93] which is of interest to industrial hygienists. In contrast to air sample analysis, biological samples are quite complex chemically and require much more

Table 3 Ion chromatographic methods for airborne and biological contaminants from several different references

Substance	CAS No.	NIOSH[a]	Other[b]
Acids, inorganic	–	7903	
2-Aminoethanol	141-43-5	3509	
p-Aminophenylarsonic acid (*p*-arsanilic acid)	98-50-0	5022	
Ammonia	7664-41-7	6016	
Bicarbonate ion	–		MASA 720A
Bromide ion	–		MASA 720A
Bromine	7726-95-6	6011	
Chloride ion	–		ASTM D 5085-95; MASA 720A
Chlorine	7782-50-5	6011	
Chloroacetic acid	79-11-8	2008	
Chromic acid	7738-94-5	7604	
Chromium, hexavalent	18540-29-9	7604	
Diethanolamine	111-42-2	3509	
Dimethylarsinic acid (cacodylic acid) (cacodylic acid)	75-60-5	5022	
Fluoride ion	–	7906	MASA 720A
Formic acid	64-18-6	2011	
Hydrogen bromide	10035-10-6	7903	
Hydrogen chloride	7647-01-0	7903	
Hydrogen fluoride	7664-39-3	7903, 7908	
Hydrogen sulfide	7783-06-4	6013	
Iodine	7553-56-2	6005	
Methylarsonic acid	124-58-3	5022	
Nitrate ion	–		ASTM D 5085-95
Nitric acid	7697-37-2	7903	
Nitrite ion	–		MASA 720A
Phosphate ion	–		MASA 720A
Phosphoric acid	7664-38-2	7903	
Sodium hexafluoroaluminate	13775-53-6	7906	
Sulfate ion	–		ASTM D 5085-95; MASA 720A
Sulfur dioxide	7446-09-5	6004	
Sulfuric acid	7664-93-9	7903	
Sulfuryl fluoride	2699-79-8	6012	
Triethanolamine	102-71-6	3509	

[a] NIOSH Manual of Analytical Methods.[21]
[b] MASA.[26] ASTM: 1997 Annual Book of ASTM Standards; section 11.[25]

sample preparation than air samples, and SFC offers that capability. For example, Kam et al.[94] have reported on a method for the analysis of a PCB in blood using SFC. A review article[95] gives an excellent overview of the use of SFC for industrial hygiene and other related samples.

7.2 Supercritical Fluid Chromatography and Fourier Transform Infrared

Although not frequently used for industrial hygiene monitoring, SFC using FTIR spectrophotometry was identified as a method for pesticide monitoring – relevant in agricultural workplaces. The method involved the separation and quantitation of aldicarb®, methomyl, captan, and phenmedipham.[96] SFC has also been applied to the separation and quantitation of PAHs.[97]

7.3 Thin-layer Chromatography

Although infrequently used for industrial hygiene analyses, there is an entire method devoted to such analyses using this technique. Method 102B of MASA: ''Separation and Microanalysis of Airborne Particulate Matter for Benzo[*a*]pyrene using Thin Layer Chromatography and Spectrofluorimetry''.[26] Chapter 22 (pp. 127–130) of MASA[26] addresses the application of TLC to air monitoring. Another possibility of increasing interest to industrial hygienists is the analysis of mycotoxins in air, related to both indoor and agricultural settings, and a paper by Ripphahn and Halpaap[98] describes the use of high-performance TLC to separate some aflatoxins. TLC has been commonly replaced by either GC or HPLC methods for this compound.

8 QUALITY ASSURANCE AND CHROMATOGRAPHIC TECHNIQUES

8.1 Requirements for Method Validation

As new chromatographic methods are developed for industrial hygiene-related monitoring, they must be validated, i.e. evaluated to assure that they are reproducible, accurate and precise, and that any interferences or confounders are identified. To that end, NIOSH has defined several protocols for such method validations. One such validation is intended for solid sorbent tubes, using a pump to draw air through the tubes.[99] Another one is intended for the use of diffusional (passive) monitors, in which the airborne contaminant (vapor or gas) diffuses to the sampling device as described in Fick's Laws of Diffusion.[99] The parameters that are generally evaluated in validating a method are trapping efficiency, desorption recovery, sampling rate, sampling volume, storage stability (both with regard to time and temperature), and interferences. For passive monitors, an experimental determination of sampling rate, air velocity and monitor orientation effects, as well as sorbent saturation mass, are also determined.

8.2 Acceptability Criteria

NIOSH has developed criteria for acceptability of industrial hygiene analytical methods. The most generally applied is that the accuracy of the method as described in the reference is $\leq 25\%$.[99] In addition, NIOSH does not consider a method to be acceptable if recovery of the contaminant from the trapping medium is $\leq 75\%$, and methods with recoveries of $>10\%$ after 7 days of storage are considered to be suspect.[100]

8.3 Internal Compared with External Standards

By internal standards are meant compounds similar to the analytes sought by the chromatographic analysis, which are added in known amount at the outset of the analysis to the sample being analyzed, and serve as an indicator of the overall recovery of the analytical method. When internal standards are used, then the results are corrected for any diminution of the internal standard, compared to internal standard analyzed directly in the solvent being used, without processing. External standards are standards of the unknown material which are analyzed separately to determine the instrument detector's response, i.e. to create a calibration curve. External standards and separately determined recoveries are used in conjunction to determine the level of analyte in the sample being analyzed.

9 ALTERNATIVES TO CHROMATOGRAPHIC TECHNIQUES

The sheer number of chromatographic methods that have been developed and validated for workplace and biological contaminants suggests that it would be difficult, if not impossible to replace them entirely. The power of chromatographic methods to separate and quantitate mixtures of organic compounds in an airborne or biological specimen is unmatched. In those cases in which a single, known compound is found in such samples, and chromatographic separation is not needed, then a few methods exist for the quantitation of such compounds. Direct MS, without chromatographic separation might be applicable in a few cases. The same might be true when using detectors such as PID, FID, and ECD. Often, such detectors are used, without chromatographic separation, in direct reading instruments which are used in a mono-contaminant environment or as leak detectors or as indicators of qualitative levels of contamination.

In a few cases, non-chromatographic methods have been developed which are compound specific. As an example, NIOSH developed a polarographic method for the quantitation of formaldehyde, which appeared in NMAM but also appeared in the literature as a paper by Septon and Ku.[101]

ACKNOWLEDGMENTS

The author would like to thank those colleagues, past and present, from the University of Iowa Hygienic Laboratory, The Dow Chemical Company Corporate Industrial Hygiene Group, and the University of Washington Department of Environmental Health for helping me mature into an industrial hygiene chemist over the past 24 years.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
ACS	American Chemical Society
AIHA	American Industrial Hygiene Association
ASTM	American Society for Testing and Materials
DDT	Dichlorodiphenyltrichloroethane
ECD	Electron Capture Detection
EPA	Environmental Protection Agency
FID	Flame Ionization Detector
FPD	Flame Photometric Detection

FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/FID	Gas Chromatography/Flame Ionization Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GLC	Gas–Liquid Chromatography
HPLC	High-performance Liquid Chromatography
HSE	Health and Safety Executive
IC	Ion Chromatography
ICP	Inductively Coupled Plasma
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LOD	Limits of Detection
MAMA	9-(Methylaminomethyl)anthracene
MAP	1-(9-Anthracenylmethyl) Piperazine
MASA	Methods of Air Sampling and Analysis
MDHS	Methods for the Determination of Hazardous Substances
MDI	Methylene Bisphenyl Isocyanate, 4,4′-Methylene Diphenyl Isocyanate
MOPP	1-(2-Methoxyphenyl) Piperazine
NIOSH	National Institute for Occupational Safety and Health
NMAM	National Institute for Occupational Safety and Health Manual of Analytical Methods
OSHA	Occupational Safety and Health Administration
PAH	Polynuclear Aromatic Hydrocarbon
PBB	Polybrominated Biphenyl
PCB	Polychlorinated Biphenyl
PID	Photoionization Detectors
QA	Quality Assurance
SAW	Surface Acoustic Wave
SFC	Supercritical Fluid Chromatography
SIM	Selective Ion Monitoring
TDI	Toluene 2,4- and 2,6-diisocyanate
TLC	Thin-layer Chromatography
TRYP	Tryptamine
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
2,4,5-T	2,4,5-Trichlorophenoxyaceticacid

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Supercritical Fluid Chromatography in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Gas Chromatography with Selective Detectors for Amines • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Forensic Science **(Volume 5)**
Capillary Ion Electrophoresis in Forensic Science

Industrial Hygiene **(Volume 6)**
Parent and Progeny Compounds in Exhaled Breath, Determination of

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Ion Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

General Articles **(Volume 15)**
Traceability in Analytical Chemistry

REFERENCES

1. A.T. James, A.J.P. Martin, 'Gas-Liquid Partition Chromatography: A Technique for the Analysis of Volatile Materials', *Analyst*, **77**, 915–932 (1952).
2. H.B. Elkins, *The Chemistry of Industrial Toxicology*, John Wiley & Sons, New York, 67, 1950.
3. J.E. Peterson, H.R. Hoyle, E.J. Schneider, 'The Analysis of Air for Halogenated Hydrocarbon Contaminants by Means of Absorption on Silica Gel', *Am. Ind. Hyg. Assoc. Quart.*, **17**, 429–432 (1956).
4. D.E. Rushing, 'Gas Chromatography in Industrial Hygiene and Air Pollution Problems', *Am. Ind. Hyg. Assoc. J.*, **17**, 238–245 (1958).
5. P.W. West, B. Sen, N.A. Groson, 'Gas Liquid Chromatographic Analysis Applied to Air Pollution', *Anal. Chem.*, **30**, 1390–1397 (1958).
6. R.H. Mansur, R.F. Pero, L.A. Krause, 'Vapor Phase Chromatography in The Quantitative Determination of Air Samples Collected in the Field', *Am. Ind. Hyg. Assoc. J.*, **20**, 175–182 (1959).
7. B. Levadie, J.F. Harwood, 'An Application of GC to the Analysis of Solvent Vapors in Industrial Air', *Am. Ind. Hyg. Assoc. J.*, **21**, 20–24 (1960).
8. A.P. Altschuller, 'Air Pollution', *Anal. Chem.*, **35**, 3–10 (1963).
9. E. Jeung, H.L. Helwig, 'Method for the Analysis of C_1–C_4 Hydrocarbons in Atmospheric Samples by Gas Chromatography', ACS Division of Water and Waste Chemistry, Preprints, 286–288, 1963.
10. K. Sakodynskii, 'The Life and Scientific Works of Michael Tswett', *J. Chromatogr.*, **73**, 303–360 (1972).
11. P.L. Kirk, 'Chromatographic Advances in Toxicology', in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, 79–126, Vol. 5, 1968.
12. G. Zweig, R.B. Moore, J. Sherma, 'Chromatography', *Anal. Chem.*, **44**, 42R–79R (1972).
13. P.R. Brown, 'The Use of High-Pressure Liquid Chromatography in Pharmacology and Toxicology', in *Advances in Chromatography*, eds. J.C. Giddings, E. Grushka, R.A. Keller, J. Cazes, Marcel Dekker, New York, Vol. 12, 1975.
14. H. Small, T.S. Stevens, W.C. Bauman, 'Novel Ion Exchange Chromatographic Method Using Conductimetric Detection', *Anal. Chem.*, **47**, 1801–1809 (1975).
15. P.R. McCullough, J.W. Warley, 'Sampling of Chloracetyl Chloride in Air on Solid Support – Determination by Ion Chromatography', *Anal. Chem.*, **51**, 1120–1125 (1979).
16. H.B. Elkins, 'Analyses of Biological Materials as Indices of Exposure to Organic Solvents', *Arch. Ind. Hyg. Occ. Med.*, **9**, 212–222 (1954).
17. L.S. Ettre, A. Zlatkis (eds.), *Seventy Five Years of Chromatography – A Historical Dialogue*, Elsevier, Amsterdam, 1979.
18. *AIHA Analytical Guides*, Prepared by Analytical Chemistry Committee, American Industrial Hygiene Association, Akron, OH, 1965.
19. *NIOSH Manual of Analytical Methods*, 1st edition, DHEW Publication (NIOSH) No. 75–121, Washington, DC, 1975.
20. 'Documentation of the NIOSH Validation Tests', DHEW Publication (NIOSH) 77–185, Washington, DC, 1977.
21. *NIOSH Manual of Analytical Methods*, 4th edition, ed. P.M. Eller, DHHS Publication (NIOSH), 94–113, 1994.
22. *OSHA Analytical Methods Manual*, 2nd edition, US Department of Labor, Occupational Safety and Health Administration, Salt Lake Analytical Laboratory, Salt Lake City, UT, 1990.
23. *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, 2nd edition, EPA/625/R-96/010(b), Research Triangle Park, NC, USA, 1996.
24. *Manual of Analytical Methods for the Analysis of Pesticides in Humans and Environmental Samples*, USEPA Publication 600/8-80-038, Research Triangle Park, NC, USA, 1980.
25. *1997 Annual Book of ASTM Standards*, section 11, 'Water and Environmental Technology', ASTM, West Conshohocken, PA, Vol. 11.03, 1997.
26. J.P. Lodge (ed.), *Methods of Air Sampling and Analysis*, 3rd edition, Lewis Publishers, Chelsea, MI, 1989.
27. T.J. Kneip, J.V. Crable (eds.), *Methods for Biological Monitoring: A Manual for Assessing Human Exposure to Hazardous Substances*, American Public Health Association, Washington, DC, 1988.
28. Health and Safety Executive, 'Methods for the Determination of Hazardous Substances' (MDHS), Health and Safety Executive/Occupational Safety and Hygiene Laboratory, London, 1987.
29. W.S Wu, R.E. Stoyanoff, R.S. Szklar, V.S. Gaind, M. Rakanovic, 'Application of Tryptamine as a Derivitizing Agent for Airborne Isocyanate Determination, Part 3: Evaluation of Total Isocyanate Analysis by High-performance Liquid Chromatography with Fluorescence and Amperometric Determination', *Analyst*, **115**, 801–807 (1990).
30. *Laboratory Analytical Methods Manual*, Laboratory Services, Workers' Compensation Board of British Columbia, Vancouver, BC, 1989.
31. Personal Communication: American Industrial Hygiene Association.
32. M.J.E. Golay, 'Theory of Chromatography in Open and Coated Tubular Columns with Round and Rectangular

Cross Sections', in *Gas Chromatography*, ed. D.H. Desty, Butterworth, London, 36–55, 1962.
33. H. Kubota, W.H. Griest, M.R. Guerin, 'Determination of Carcinogens in Tobacco Smoke and Coal-Derived-Samples – Trace Polynuclear Aromatic Hydrocarbons', in *Trace Substances in Environmental Health – IX. A Symposium*, ed. D.D. Hemphill, University of Missouri, Columbia, MO, 1975.
34. R.G. Melcher, 'Industrial Hygiene', *Anal. Chem.*, **55**, 40R–56R (1983).
35. M. Harper, C.R. Glowacki, P.R. Michael, 'Industrial Hygiene', *Anal. Chem.*, **69**, 307R–327R (1997).
36. R. Lauwerys, 'Biological Criteria for Selected Industrial Toxic Chemicals: A Review', *Scand. J. Work Env. Health*, **1**, 139–172 (1975).
37. R.C. Baselt, *Biological Monitoring Methods for Industrial Chemicals*, 2nd edition, PSG Publishing, Littleton, MA, 1988.
38. 'Analysis of Pesticide Residues in Human and Environmental Samples', ed. R.R. Watts, Environmental Protection Agency, EPA-600/8-80-038, Research Triangle Park, NC, 1980.
39. Jimmy L. Perkins, *Modern Industrial Hygiene*, Van Nostrand Reinhold, New York, 468–478, 1997.
40. Jimmy L. Perkins, *Modern Industrial Hygiene*, Van Nostrand Reinhold, New York, 426–454, 1997.
41. Personal Communication: Industrial Hygiene Laboratory, The Dow Chemical Company, Midland, MI.
42. D. Rondia, 'Regarding the Volatility of Polycyclic Hydrocarbons' (in French), *Int. J. Air & Water Poll.*, **9**, 113–121 (1965).
43. S. DiNardi (ed.), *The Occupational Environment – Its Evaluation and Control*, American Industrial Hygiene Association, Fairfax, VA, 221–222, 1997.
44. 'Volatile Organic Compounds in Whole Blood: Determination by Heated Dynamic Headspace Purge and Trap Isotope Dilution GC/MS', EPA 560/5-87-008 (1987).
45. R.C. Lao, R.S. Thomas, H. Oja, L. Dubois, 'Application of Gas Chromatograph-Mass Spectrometer Data Processor Combination to the Analysis of the Polycyclic Aromatic Hydrocarbon Content of Airborne Pollutants', *Anal. Chem.*, **45**, 908–920 (1973).
46. Y. Matoba, Y. Takimoto, T. Kato, 'Indoor Behavior and Risk Assessment Following Spraying of d-Resmethrin and d-Tetramethrin', *Am. Ind. Hyg. Assoc. J.*, **59**, 181–190 (1998).
47. S.J. Macnaughton, T.L. Jenkins, S. Alugupali, D.C. White, 'Quantitative Sampling of Indoor Air Biomass by Signature Lipid Biomarker Analysis: Feasibility Studies in a Model System', *Am. Ind. Hyg. Assoc. J.*, **58**, 270–77 (1997).
48. S.L. Wellson, E.G. Trawick, M.F. Stowers, S.L.P. Jordan, T.L. Wass, 'Laboratory and Hospital Evaluation of Four Personal Monitoring Methods for Glutaraldehyde in Air', *Am. Ind. Hyg. Assoc. J.*, **59**, 96–103 (1998).
49. L.L. Needham, S.L. Head, R.E. Cline, 'Determination of Phenols and Cresols in Urine by Gas Chromatography', *Anal. Lett.*, **17**(B14), 1555–1565 (1984).
50. M. Guillemin, E. Guberan, 'Value of Simultaneous Determination of P_{CO_2} in Monitoring 1,1,1-Trichloroethane by Breath Analysis', *Brit. J. Indust. Med.*, **39**, 161–168 (1982).
51. V.W. Burse, L.L. Needham, J.A. Liddle, D.D. Bayse, H.A. Price, 'Interlaboratory Comparison for Results of Analyses for Polybrominated Biphenyls in Human Serum', *J. Anal. Tox.*, **4**, 22–26 (1980).
52. R.M. McKenzie, Y.Y. Kim, R.W. Bishop, R.J. Gaffney, 'The Collection and Analysis of 2,3-Dimethyl-2,3-Dinitrobutane Vapor', *Am. Ind. Hyg. Assoc. J.*, **59**, 388–392 (1998).
53. Y. Ichikawa, M. Yoshida, A. Okayama, I. Hara, K. Morimoto, 'Biological Monitoring for Workers Exposed to 4,4′-methylenebis(2-chloroaniline)', *Am. Ind. Hyg. Assoc. J.*, **51**, 5–7 (1990).
54. D. Henneberg, 'Combination of Gas Chromatography and Mass Spectrometry for the Analysis of Organic Mixtures', *Z. Anal. Chem.*, **183**, 12–23 (1961).
55. J.A. Dorsey, R.H. Hunt, M.J. O'Neal, 'Rapid-Scanning Mass Spectrometry: Continuous Analysis of Fractions from Capillary Gas Chromatography', *Anal. Chem*, **35**, 511–515 (1963).
56. L-J Wu, S.S. Que Hee, 'A Solid Sorbent Personal Air Sampling Method for Aldehydes', *Am. Ind. Hyg. Assoc. J.*, **56**, 362–367 (1995).
57. R.S. Dills, S. Kent, H. Chekoway, D. Kalman, 'Volatile Solvents in Blood by Static Headspace Analysis', *Talanta*, **38**, 367–374 (1991).
58. R.C. Wieboldt, K.D. Kempfert, D.W. Later, E.R. Campbell, 'Analysis for Pyrethrins Using Capillary Supercritical Fluid Chromatography and Capillary Gas Chromatography with Fourier-transform Infrared Detection', *J. High Resolut. Chromatogr.*, **12**, 106–111 (1989).
59. T.F. Moate, C. Lu, R.A. Fenske, R. Hahne, D. Kalman, 'Improved Cleanup and Determination of Dialkyl Phosphates in the Urine of Children Exposed to Organophosphorus Insecticides', *J. Anal. Tox.*, **23**, 230–236 (1999).
60. E.R. Kennedy, M.T. Abell, J. Reynolds, D. Wickman, 'A Sampling and Analytical Method for the Simultaneous Determination of Multiple Organophosphorus Pesticides in Air', *Am. Ind. Hyg. Assoc. J.*, **55**, 1172–1177 (1994).
61. C.G. Flinn, W.A. Aue, 'Gas Chromatographic Determination of Organogermanium Compounds by Surface or Gas-phase Luminescence', *J. Chrom.*, **186**, 299–306 (1979).
62. J.F. Lawrence, H.A. McLeod, 'Evaluation of Electrochemical Conductivity Detection for Gas Chromatographic Screening of some Organo-nitrogen Herbicides and Fungicides in Foods', *J. Assoc. Off. Anal. Chem.*, **60**, 979–981 (1977).

63. J.R. Stetten, J.M. Sedlak, K.F. Blurton, 'Electrochemical Gas Chromatographic Detection of Hydrogen Sulfide at ppm and ppb Levels', *J. Chromatog. Sci.*, **15**, 125–128 (1977).
64. J.N. Driscoll, J.B. Clarici, 'A New Photoionization Detector for Gas Chromatography', *Chromatographia*, **9**, 567–570 (1976).
65. P.L. Patterson, 'Recent Advances in Thermionic Ionization Detection for Gas Chromatography', *J. Chromatog. Sci.*, **24**, 41–52 (1986).
66. G. Skarping, M. Dalene, T. Brorson, J. Sandstroem, C. Sangoe, A. Tjander, 'A Chromatographic Determination of Amines in Biological Fluids with Special Reference to the Biological Monitoring of Isocyanates and Amines. I. Determination of 1,6-hexamethylenediamine Using Glass Capillary Chromatography and Thermionic Specific Detection', *J. Chromatog.*, **479**, 125–133 (1989).
67. A. Vinsjansen, K.E. Thrane, 'Gas-chromatographic Detection of Phosphine in Ambient Air', *Analyst*, **103**, 1195 (1978).
68. J.W. Robinson, E.L. Kiesel, J.P. Goodbread, R. Bliss, R. Marshall, 'The Development of a Gas Chromatography/Furnace Atomic Absorption Combination for the Determination of Organic Lead Compounds', *Anal. Chim. Acta*, **92**, 321–328 (1977).
69. P. Sandra, 'Fast Capillary Gas Chromatography', *LC-GC*, **5**, 236–246 (1987).
70. M.A. Klemp, M.L. Akard, R.D. Sacks, 'Cryofocusing Inlet with Reverse Flow Sample Collection for Gas Chromatography', *Anal. Chem.*, **65**, 2516–20 (1993).
71. H. Wohltjen, D.S. Ballantine, Jr, N.L. Jarvis, 'Vapor Detection with Surface Acoustic Wave Microsensors', in *Chemical Sensors and Microinstrumentation*, eds. R.W. Murray, R.E. Dessy, W.R. Heineman, J. Janata, W.R. Seitz, ACS Symposium Series 403, ACS, Washington, DC, 1989.
72. R.M.A. Hahne, 'Direct Reading Instruments for Gases, Vapors, and Particulates', in *Fundamentals of Industrial Hygiene*, 4th edition, ed. B.A. Plog, National Safety Council, Itasca, IL, 520–521, 1996.
73. 'The Use of Portable Photoionization Gas Chromatographs for Rapid Screening of Samples for Purgeable Organic Compounds in the Field and Laboratory', USEPA-Reg. 01A–5959 (1983).
74. A. Manz, D.J. Harrison, E. Verpoorte, H.M. Widmer, 'Miniaturization of Separations Systems', in *Advances in Chromatography*, eds. P.R. Brown, E. Grushka, Marcel Dekker, New York, Vol. 33, 1993.
75. S. Marshall, 'Microinstrumentation in the Coming Decade', *R&D Magazine*, **39**, March 1997.
76. E.R. Kennedy, J-J. Lin, J.M. Reynolds, J.B. Perkins, 'A Sampling and Analytical Method for the Simultaneous Determination of Multiple Organonitrogen Pesticides in Air', *Am. Ind. Hyg. Assoc. J.*, **58**, 720–725 (1997).
77. S.P. Levine, K.J.D. Hillig, V. Dharmarajan, M.W. Spence, M.D. Baker, 'Critical Review of Methods of Sampling, Analysis and Monitoring for TDI and MDI', *Am. Ind. Hyg. Assoc. J.*, **56**, 581–589 (1995).
78. W.T. Sanderson, R. Biagini, W. Tolos, G. Henningsen, B. MacKenzie, 'Biological Monitoring of Commercial Pesticide Applicators for Urine Metabolites of the Herbicide Alachlor', *Am. Ind. Hyg. Assoc. J.*, **56**, 883–889 (1995).
79. M. Dalene, G. Skarping, P. Ling, 'Worker Exposure to Thermal Degradation Products of TDI and MDI-based Polyurethane: Biomonitoring of 2,4-Toluenediamine, 2,6-Toluenediamine, and 4,4′-Methylenedianiline in Hydrolyzed Urine and Plasma', *Am. Ind. Hyg. Assoc. J.*, **58**, 587–591 (1997).
80. E.T. Ny, D. Heederik, H. Kromhout, F. Jongeneelen, 'The Relationship between Polycyclic Hydrocarbons in Air and Urine of Workers in a Söderberg Potroom', *Am. Ind. Hyg. Assoc. J.*, **54**, 277–284 (1993).
81. F.J. Jongeneelen, R.B.M. Auzion, P.Th. Henderson, 'Determination of Hydroxylated Metabolites of Polynuclear Aromatic Hydrocarbons in Urine', *J. Chromatogr.*, **413**, 227–232 (1987).
82. R.P. Streicher, J.E. Arnold, M.K. Ernst, C.V. Cooper, 'Development of a Novel Derivatization Reagent for the Sampling and Analysis of Total Isocyanate Group in Air and Comparison of its Performance with That of Several Established Reagents', *Am. Ind. Hyg. Assoc. J.*, **57**, 905–913 (1996).
83. M.W. Dong, K. Ogan, J.L. DiCesare, 'Very High Speed Liquid Chromatography for Polynuclear Aromatic Hydrocarbon Analysis: System and Applications', in *Polynuclear Aromatic Hydrocarbons: 6th International Symposium*, eds. M. Cooke, A.J. Dennis, G.L. Fisher, Battelle Press, Columbus, OH, 1982.
84. T.L. Jones, L.D. Betowski, J. Yinon, 'Analysis of Chlorinated Herbicides by High-performance Liquid Chromatography/Mass Spectrometry', in *Liquid Chromatography/Mass Spectrometry: Applications in Agricultural, Pharmaceutical, and Environmental Chemistry*, ed. M.A. Brown, ACS Symposium Series 420, ACS, Washington, DC, 62–74, 1990.
85. W.M. Draper, F.R. Brown, R. Bethem, and M.J. Mille, 'Anion Exchange Thermospray Tandem Mass Spectrometry of Polar Urinary Metabolites and Metabolic Conjugates', in *Liquid Chromatography/Mass Spectrometry: Applications in Agricultural, Pharmaceutical, and Environmental Chemistry*, ed. M.A. Brown, ACS Symposium Series 420, ACS, Washington, DC, 253–269, 1990.
86. S.A. Wise, S.N. Chesler, L.R. Hilpert, W.E. May, R.E. Rebbert, C.R. Vogt, 'Characterization of a Polynuclear Aromatic Hydrocarbon Mixture from Air Particulate Samples Using Liquid Chromatography, Gas Chromatography, and Mass Spectrometry', in *Polynuclear Aromatic Hydrocarbons: Mechanisms, Methods, and*

Metabolism, *8th International Symposium*, ed. M. Cooke, A. Dennis, Battelle Press, Columbus, OH, 1413–1427, 1985.

87. E. Hakala, L. Pyy, 'Selective Determination of Toxicologically Important Arsenic Species in Urine by HPLC/Hydride Generation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **7**, 191–196 (1992).
88. D. Bushee, I.S. Krull, P.R. Demko, S.B. Smith, Jr, 'Trace Analysis and Speciation for Arsenic Anions by HPLC/Hydride Generation/Inductively Coupled Plasma Emission Spectroscopy', Abstracts, Pittsburgh Conference on Analytical Chemistry, Atlantic City, NJ, Paper 397, 1983.
89. R. Dagani, 'Improved Arsenic Analysis May Aid Regulators', *Chem. Eng. News*, **59**, 29–30 (1981).
90. S. Abuku, S. Tanaka, Y. Seki, S. Imamiya, 'Determination of Bromide Ions in Total Blood, Plasma, and Urine by Ion Chromatography with Amperometric Detection', *Nippon Elseigaku Zasshi*, **44**, 945–952 (1989).
91. C. Lorberau, 'Determination of Gaseous and Particulate Fluorides by Ion Chromatographic Analysis', *Appl. Occup. Environ. Hyg.*, **8**, 775–784 (1993).
92. R.E. Jentoft, T.H. Gouw, 'Separation of Halogenated Paraffins', *Anal. Chem.*, **44**, 681 (1972).
93. S.T. Sie, G.W. Rinders, 'Separation of Mono- and Polynuclear Aromatic Hydrocarbons', *Anal. Chim. Acta*, **38**, 31 (1967).
94. K.S. Kam, S. Kapita, A.F. Yanders, R.K. Puri, 'A Multiple Extraction and On-line System for the Analysis of Chlorinated Compounds', *Chemosphere*, **23**, 1109–1116 (1991).
95. L.J. Mulcahey, C.L. Rankin, M.E.P. McNally, 'Environmental Applications of Supercritical Fluid Chromatography', in *Advances in Chromatography*, eds. P.R. Brown, E. Grushka, Marcel Dekker, New York, Vol. 34, 1994.
96. R.C. Wieboldt, J.A. Smith, 'Supercritical Fluid Chromatography with Fourier-transform Infrared Detection', *Supercritical Fluid Extraction and Chromatography: Techniques and Applications*, eds. B.A Charpentier, M.R. Sevenants, ACS Symposium Series 366, ACS, Washington, DC, 229–242, 1988.
97. W.P. Jackson, R.C. Kong, M.L. Lee, 'Capillary Supercritical Fluid Chromatography of Polycyclic Aromatic Compounds', *Polynuclear Aromatic Hydrocarbons: Mechanisms, Methods, and Metabolism*, *8th International Symposium*, ed. M. Cooke, A. Dennis, Battelle Press, Columbus, OH, 1985.
98. J. Ripphahn, H. Halpaap, 'Quantitation in High Performance Micro-thin Layer Chromatography: Separation of Aflatoxins B1,B2, G1 and G2', *J. Chromatog.*, **112**, 81–96 (1975).
99. K. Busch, E. Gunderson, C. Anderson, R. Smith, L. Doemeny, 'Development and Validation of Methods for Sampling and Analysis of Workplace Toxic Substances with a Statistical Appendix', DHHS(NIOSH) Publ. No 80–133, Cincinnati, OH, 1980.
100. E.R. Kennedy, T.J. Fishbach, R. Song, P.M. Eller, S.A. Shulman, R.D. Hull, 'Development and Evaluation of Methods', in NIOSH Manual of Analytical Methods, 4th edition, Section E, 1998.
101. J.C. Septon, J.C. Ku, 'Workplace Air Sampling and Polarographic Determination of Formaldehyde', *Am. Ind. Hyg. Assoc. J.*, **43**, 845–852 (1982).

Cyclic Activation Analysis

Xiaolin Hou

Risø National Laboratory for Sustainable Energy, Technical University of Denmark, Roskilde, Denmark

1 Introduction **1**
2 Theory **2**
2.1 Equation of Cyclic Activation Analysis 2
2.2 Selection of Time Parameters 3
2.3 Estimation or Confirmation of the Half-life of Nuclide 6
2.4 Dead Time and Pileup Correction 6
2.5 Cumulative Activation Analysis 7
3 Classification and Facilities of Cyclic Activation Analysis **7**
4 Application of Cyclic Activation Analysis **8**
4.1 Determination of Se, F, and O in Biological Materials 8
4.2 Determination of Pb, F, Sc, and Ag in Environmental Samples 10
4.3 Determination of O, Sc, Si, Al, Ag, Au, Rh, Hf, Pb, and U in Geological and Industrial Materials 11
5 Conclusion and Perspective **12**
Abbreviations and Acronyms **12**
Related Articles **13**
References **13**

Cyclic activation analysis (CAA) is a method of activation analysis for elemental analysis, in which a sample is irradiated, decayed, counted, and then irradiated again. This process is repeated for a number of cycles, and the spectra from each counting are summed together to give one final total spectrum. By this process, the counts of a short-lived nuclide of interest are considerably increased, and the analytical sensitivity of elements is significantly improved. The most commonly used CAA is the cyclic neutron activation analysis (CNAA) by irradiation with the thermal, epithermal, and fast neutrons produced from nuclear reactor, accelerator, and isotopic neutron source. The nuclear reactor can supply a much high neutron flux and is most often used for this purpose. At least 20 elements produced short-lived nuclides (half-life <100 s) by thermal neutron bombardment, and more than 10 elements produced nuclides with half-life of 100–600 s. These can be determined by thermal and epithermal neutron CAA. This technique has been widely applied in biological, environmental, geological, and industrial studies, and most often measured elements include Se, F, Pb, Hf, Sc, O, Ag, and Rh.

The advantages of CAA, as compared with conventional activation analysis, include significant improvement in the detection limit, analytical precision, and accuracy for the elements by using short-lived nuclides; short experimental time and increased analytical number of samples per unit time; capability of estimation or confirmation of the half-life of the short-lived nuclide; and determination of the degree of homogeneity of a sample. However, the application of CAA is limited by the number of elements determined, because only some of the elements determined by conventional activation analysis can be determined by this method. In addition, dead time and pileup are serious problems in CAA and must be corrected. The principle, selection of optimal experiment conditions, detection limit, analytical precision of CAA, as well as the dead time and pileup corrections are discussed in this article. Some applications of this method are highlighted.

1 INTRODUCTION

Activation analysis has been well documented as a powerful technique for elemental analysis (*see* **Instrumental Neutron Activation Analysis**, **Charged Particle Activation Analysis**). Its high sensitivity and accuracy, simultaneous multielement analytical capability, and matrix-free and nondestructive analysis make it wide applicable in a variety of research fields such as biological, environmental, geological, industrial, archeological, and forensic studies. Of all kinds of activation analyses, reactor neutron activation has been the mostly used, and as many as 70 elements can be easily determined. However, for many elements, an analytical period from days to weeks is needed because the nuclides detected, which are produced by reaction of their stable isotope with neutron, have longer half-life, such as ^{75}Se (120 days), ^{46}Sc (84 days), ^{110m}Ag (250 days), and ^{181}Hf (45 days). To attain the required sensitivity, a long-time irradiation, delay, and counting are necessary. This reduces its competitive capability to other analytical techniques, such as inductively coupled plasma mass spectrometry (ICP–MS). The use of the short-lived nuclides of element in preference to their longer-lived activation products can significantly reduce the total experimental time, which in turn can reduce the analytical period and increase the number of samples measured per day, and make activation analysis more cost-effective and competitive. On the other hand, several elements can be determined

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

by neutron activation analysis (NAA) only through the measurement of short-lived nuclide, such as ^{20}F (11 s), ^{19}O (27 s), and ^{207m}Pb (0.8 s). Therefore, NAA using short-lived nuclides has been given more attention.

The saturation activation of short-lived radionuclides is reached quickly on irradiation, and there is no further increase in activity with time. Similarly, the counting is also limited by half-life since the activity of a short-lived radionuclide decays very quickly, and further counting cannot accumulate any more counts of radionuclides of interest; on the contrary, it increases the background derived from long-lived interfering nuclides. In addition, owing to the short half-life, the number of radioactive nuclei (N^*) formed in saturation is small ($N^* = N\sigma\Phi/\lambda = N\sigma\Phi t_{1/2}/\ln 2$, where N is the number of target nuclei, σ is the activation cross section, Φ is the neutron flux, λ is the decay constant, and $t_{1/2}$ the half-life of the formed radioactive nuclide). Consequently, the counting has a large statistical error, and both the analytical accuracy and detection limit are unsatisfied. It is possible, however, to overcome this by reirradiating sample and counting it again, once the initial activity has been decayed away. The number of counts of the nuclide of interest can be significantly increased by repeating this process of irradiation-decay-counting-decay and irradiation again for a number of cycles, i.e. cyclic activation, and summing the spectra from each counting to give a final spectrum. By this CAA, the sensitivity, precision, and accuracy will be considerably improved.

The history of CAA can be traced back to the beginning of 1960s. In 1960 and 1961,[1,2] Anders first reported a technique where a sample is cycled between the irradiating source and the radiation detector, in order to improve both the sensitivity and statistical certainty of the elemental analysis. In his experiment, he repeatedly irradiated the sample with neutrons from a Be target of a 2-MeV Van de Graff accelerator and counted on an NaI(Tl) detector for two consecutive periods of equal length. By subtracting the spectra acquired from each other, the resulting γ-ray difference spectrum represented the contribution from only very short-lived nuclides and suppressed the contribution of longer-lived components in the γ-spectra of the activated sample. The method was documented by determining fluorine using faster neutron reaction of ^{19}F(n, α)^{16}N, and Se, Ag, Hf, F, and O by thermal neutron (n, γ) reactions. Preliminary investigation indicated the usefulness of the technique for the determination of 18 elements, such as O, F, Na, Sc, Ge, Se, Br, Y, Rb, Rh, Ag, In, Er, Hf, W, Yb, Ir, and Au. However, the term *cyclic activation analysis* is not mentioned in this article. In 1966, Caldwell et al.[3] suggested the use of a similar technique as part of a combination neutron experiment for remote elemental analysis. The term *cyclic activation analysis* was first used by Givens et al. in 1968,[4] when they measured ^{16}N and ^{24m}Na using a technique similar to that of Anders. However, the irradiating source, rather than the sample, was cycled electronically.[5] Using this facility, the nuclides with half-life down to a few milliseconds, such as ^{24m}Na(20 ms), can be utilized for analytical purposes. In 1969, Tani et al.[6] used a similar experimental facility to observe the γ-spectra of ^{205}Pb ($\sim$4 ms) and ^{207m}Pb (800 ms).

In order to obtain the high sensitivity required while using CNAA as a useful trace element analysis technique, Ozek[7] and Spyrou et al.,[8] in 1973, first used nuclear reactor as a neutron source to supply a neutron flux more than three orders of magnitude higher than other sources used by the earlier workers, such as an accelerator and isotope sources for CNAA. They published their results on the measurement of lead through ^{207m}Pb in environmental samples.[8] Since 1970s, many facilities for CAA have been set up, and methodology and application studies were carried out.[9–16] This article is an updated version of the previous one published in the first edition of this book in 2000.[17]

2 THEORY

2.1 Equation of Cyclic Activation Analysis

The principle of CAA has been described by Given et al.[4] and Spyrou.[18] In this method, the sample is irradiated for a short period of time, and after a delay period from the end of irradiation, the radiation emitted is counted for a short period of time, then the sample is irradiated again and the entire process is repeated for a number of cycles (Figure 1). The detected radiations at each counting period are summed and finally a total cumulative detector response is obtained. The cycle period T is given by Equation (1):

$$T = t_i + t_d + t_c + t_w \tag{1}$$

where t_i is the time of irradiation, t_d is the delay time, i.e. the time between the end of irradiation and the start of counting, which is usually the time required to transfer the sample from the irradiation position to the counting station, or to transfer the irradiation source to the sample station, t_c is the counting time, and t_w is the waiting time, i.e. the time between the end of counting and the start of the next irradiation.

The detector response (or the number of counts) for the first cycle is given by Equation (2):

$$D_1 = \frac{N\sigma\Phi I\varepsilon}{\lambda}(1 - e^{-\lambda t_i})e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \tag{2}$$

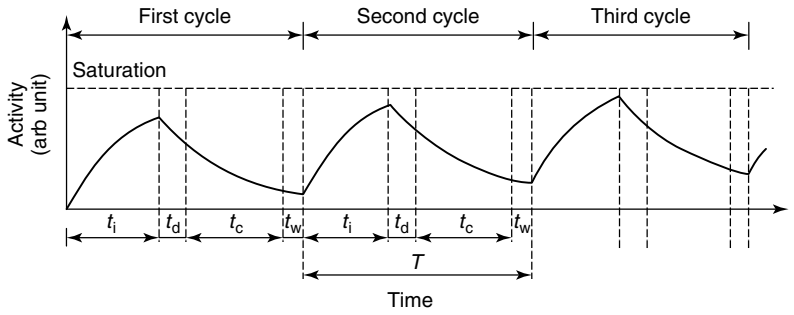

Figure 1 Time parameters of CAA and the variation of the activity of nuclide with time and cyclic number.

where the definitions of σ, Φ, and λ are the same as above, I is the intensity of the radiation of interest, and ε is the counting efficiency of the detector.

In the second counting period, the detector response is the same number of counts due to the second irradiation in addition to what was left from the first irradiation, expressed by Equation (3):

$$D_2 = D_1 + D_1 e^{-\lambda T} = D_1(1 + e^{-\lambda T}) \tag{3}$$

Similarly, for the nth cycle, which can be expressed by Equation (4)

$$D_n = D_1(1 + e^{-\lambda T} + e^{-2\lambda T} + e^{-3\lambda T} + \cdots + e^{-(n-1)\lambda T}) \tag{4}$$

Then the total cumulative detector response in all n cycles is given by Equation (5):

$$\begin{aligned} D_c &= \sum_{i=1}^{n} D_i \\ &= D_1 \left[\frac{n}{1 - e^{-\lambda T}} - \frac{e^{-\lambda T}(1 - e^{-n\lambda T})}{(1 - e^{-\lambda T})^2} \right] \\ &= \frac{N \Phi \sigma I \varepsilon}{\lambda} \left[\frac{n}{1 - e^{-\lambda T}} - \frac{e^{-\lambda T}(1 - e^{-n\lambda T})}{(1 - e^{-\lambda T})^2} \right] \\ &\quad \times (1 - e^{-\lambda t_i}) e^{-\lambda t_d} (1 - e^{-\lambda t_c}) \end{aligned} \tag{5}$$

Equation (5) gives the basic relationship for CAA.

2.2 Selection of Time Parameters

The detector response may be maximized by a proper selection of the parameters in Equation (5). For a given total experiment time, T_t ($T_t = nT = mt_{1/2}$), the maximal value of D_c occurs when $t_d = t_w = 0$ and $t_i = t_c$. However, the fact is that t_d and t_w are not zero and their values ultimately depend on the transfer system used. When the transfer is done manually, the transfer time can be relatively long in the value of 10 s to 1 min; in this case, the method sometimes is referred to as *pseudocyclic activation analysis* (PCAA).[12,19,20] In some fast transfer system, where the transfer is completed electronically and through computer control, the transfer time is of the order of 0.1 s and less.[21,22] The transfer time can be made very small when there is no physical transfer of the sample but a pulsed irradiation is used, which was suggested as the "real" CAA.[3,4] For the following discussion, unless otherwise stated, it is assumed that $t_d = t_w = 0$ and $t_i = t_c$.

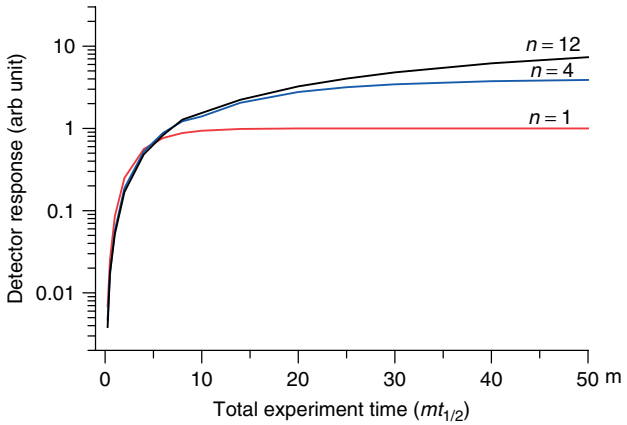

Figure 2 Variation of the conventional and cumulative signal with total experiment time, $t_d = t_w = 0$, $t_i = t_c$ and n is the number of cycles.

Figure 2 shows the variation of detector response for the nuclides of interest with the total experimental time ($T_t = mt_{1/2} = nT$) in conventional and cyclic case with different cyclic numbers. It is clear that below a certain experimental time, conventional activation is to be preferred to cyclic activation, and the cross over point of curves between conventional and cyclic activation shifts to longer experimental time with an increasing number of cycles. After the cross over point, cyclic activation has a higher detector response than that of conventional one-shot irradiation. For a certain total experiment time, such as $T_t = 10t_{1/2}$, the cumulative detector response increases with increasing number of cycles in the beginning, and there is an optimum cyclic number, after which the detector response does not increase and slowly decreases (Figure 3). In addition, for conventional single-shot activation analysis, the detector response increases quickly

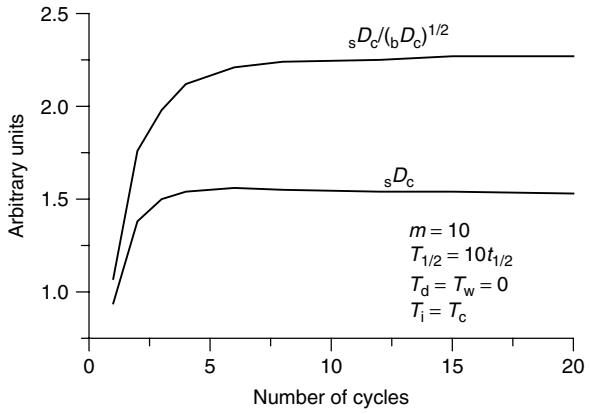

Figure 3 Variation of signal and signal-to-noise ratio $({}_sD_c/\sqrt{{}_bD_c})$ for cyclic and conventional activation with the number of cycles. Here, the underlying background to signal of interest is provided by the activity of nuclide with half-life equal to 10 times half-life of nuclide of interest ($T_{1/2} = 10_{1/2}$).

with increase in the total experiment time in the beginning and 95% of saturation activity is obtained before $5t_{1/2}$, following which a further increase in the total experiment time becomes insignificant (Figure 2). However, for CAA, this time point becomes longer. For $n = 4$, 95% of saturation activity is obtained after $15t_{1/2}$, and for $n = 12$, after a total experiment time of $50t_{1/2}$, the activity still increases with the increase in total experiment time.

However in CAA, the radionuclides produced from elements in the matrix, which has a longer half-life than the nuclide of interest, will decay slowly between irradiations and therefore the underlying background will increase through successive irradiation. Consequently, the detection of an element does not depend solely on the signal from the nuclide of interest, but also depends on the background contribution from the matrix elements.

The purpose of CAA is to improve the analytical precision and detection limit for the elements of interest. The analytical precision is usually expressed by relative standard deviation of a signal S measured above a background B (Equation 6):

$$\delta = \frac{\sqrt{S + 2B}}{S} \tag{6}$$

For the cumulative spectrum of n cycles, it is expressed as Equation (7):

$$\delta n = \frac{\sqrt{({}_sD_c + 2{}_bD_c)}}{{}_sD_c} \tag{7}$$

where ${}_sD_c$ is the cumulative detector response for the nuclide of interest, ${}_bD_c$ is the cumulative detector response for the nuclide making the major background contribution to the signal from the nuclide of interest.

The detection limit L_d is usually given by Equation (8):

$$L_d = 2\sqrt{B} \qquad \text{(counts)}$$

$$\text{or} \quad L_d = \frac{2\sqrt{B}}{S/m} = \frac{2m\sqrt{B}}{S} \qquad (\mu g) \tag{8}$$

where S/m is the counts per unit mass element of interest.

For CAA with n cycles, the detection limit is given by Equation (9):

$$L_d = \frac{2\sqrt{{}_bD_c}}{{}_sD_c/m} = \frac{2m\sqrt{{}_bD_c}}{{}_sD_c} \tag{9}$$

Therefore, $(\sqrt{{}_bD_c}/{}_sD_c)$ or $({}_sD_c/\sqrt{{}_bD_c})$ and $(\sqrt{{}_sD_c + 2{}_bD_c}/{}_sD_c)$ or $({}_sD_c/\sqrt{{}_sD_c + 2{}_bD_c})$, not the detector response for the nuclide of interest, must be optimized, when the most suitable cyclic timing parameters are chosen for a nuclide.

In Figures 4 and 5, the signal-to-noise ratio $({}_sD_c/\sqrt{{}_bD_c})$, and the ratio of signal to the square root of sum of ${}_sD_c$ and $2{}_bD_c$, $({}_sD_c/\sqrt{{}_sD_c + 2{}_bD_c})$, are plotted versus total experiment time $(mt_{1/2})$ for comparison of cyclic with conventional activation analysis. In this case, for illustrative purpose, assuming that only one long-lived nuclide with a half-life $(T_{1/2})$ of 10 times half-life of nuclide of interest $(T_{1/2} = 10t_{1/2})$ is considered to contribute to the underlying background of the signal. It can be seen that the background plays a significant role in emphasizing the difference between conventional and cyclic cases. With increase in the total experiment time, $({}_sD_c/\sqrt{{}_bD_c})$ and $({}_sD_c/\sqrt{{}_sD_c + 2{}_bD_c})$ have a maximum and this maximum shift to a longer experimental time with increasing number of cycles (Figures 4 and 5). Comparing with Figure 2, the cross over point between conventional and cyclic activation in Figure 4 is shifted to a lower value of m,

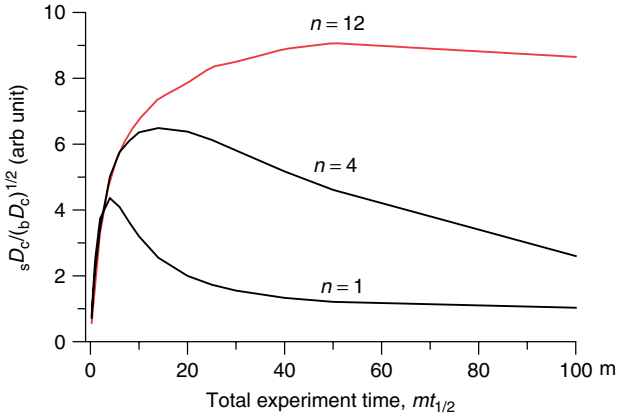

Figure 4 Variation of signal-to-noise ratio $({}_sD_c/\sqrt{{}_bD_c})$ for cyclic and conventional activation with the total experiment time $(T_{1/2} = 10t_{1/2})$.

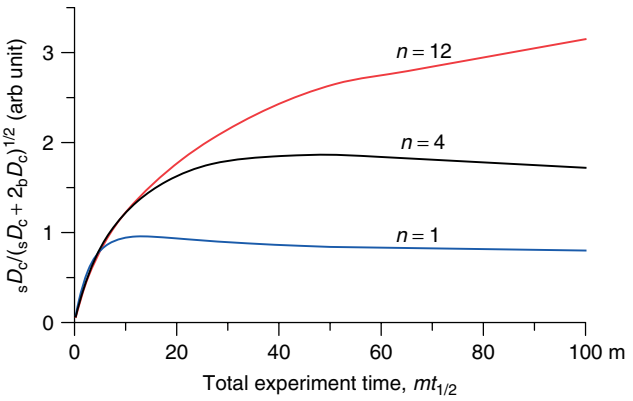

Figure 5 Variation of $({}_{s}D_{c}/\sqrt{{}_{s}D_{c}+2{}_{b}D_{c}})$ for cyclic and conventional activation with the total experiment time $(T_{1/2}=10t_{1/2})$.

i.e. cyclic activation becomes advantageous for shorter total experiment time than the case just considering the detector response.

Figure 6 shows the variation of signal-to-noise with the total experimental time (number of cycle) under a certain cycle period $(T=2t_{1/2})$. It is clear that in this case, signal-to-noise ratio increases always with increasing total experiment time.

To obtain a better cyclic advantage factor over conventional activation, it is obvious that one must increase the total experiment time $(mt_{1/2})$. Choice of the best total experiment time for cyclic activation is therefore a compromise between the time available for sample analysis and the detection limit to be achieved. Whenever the total experiment time is chosen, the number of cycles (n) or cycle period, $T=T_{t}/n$, has to be decided. In Figure 3, the variation of the signal $({}_{s}D_{c})$ and the signal-to-noise ratio $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ were plotted as a function of the number of cycles $(n=T_{t}/T)$. It can be seen that after a certain cycle, $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ and ${}_{s}D_{c}$ increase very slowly, and with further increase in the number of cycles these become

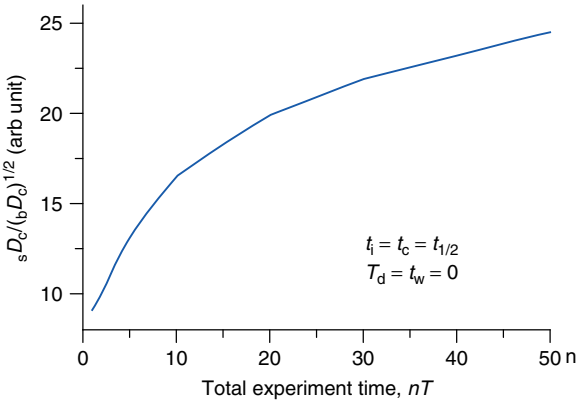

Figure 6 Variation of signal to noise $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ with the total experiment time (nT) for a certain cycle period $(T=2t_{1/2})$.

insignificant. After a maximum occurs, ${}_{s}D_{c}$ decrease slowly. For example, if $T=10t_{1/2}$ and $T_{1/2}=10t_{1/2}$, 95% of saturation activity of nuclides can be obtained after four cycles and 95% of a maximum of $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ after eight cycles. In addition, in order to maximize signal $({}_{s}D_{c})$, the optimal number of cycles is always smaller than that required for maximizing the signal-to-noise ratio $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$, and it also increases with the increase in the half-life of the background for the same total experiment time. However, it becomes less pronounced for the half-life of the background more than $100t_{1/2}$.

Spyrou et al.[18] and Spyrou and Kerr[23] investigated the effect of the transfer time (t_{d} and t_{w}) on $({}_{s}D_{c})$ and $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$, they found that the cross over point between cyclic and conventional activation was shifted to a longer experiment time with increasing transfer time. It means that CAA becomes more preferable to the conventional method with decreasing transfer time, and the transfer time is an important factor effecting the detection limit and precision.

In real situation, many long-lived nuclides from matrix elements, not just one, will contribute to the underlying background of the peak of the nuclide of interest after activation. For calculating these contributions and optimizing the parameters of cyclic activation, many computer programs have been developed.[24–26] In Tout and Chatt's program,[25] by entering total experiment time, transfer time (t_{d}, t_{w}), the mass of sample and relevant nuclear data of matrix elements contributing to the background activity, as well as the half-life of the nuclide of interest, $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ can be calculated for individual matrix elements and the total background. The optimal irradiation and counting time (or number of cycles) and its ranges were finally given for maximum and 95% of maximum of $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$. Their results indicated that a single value of t_{i} and t_{c} can be used to produce results close to the maximal $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ for nuclides with a large range of half-lives, i.e. for a total experiment time of 5 min, $t_{i}=t_{c}=6$–7 s will produce values of more than 90% of maximal $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ for nuclides with half-lives in the ranges 3–22 s. This means that a number of short-lived nuclides can be measured at close to their best detection limits by using only one set of CAA.

In Al-Mugrabi and Spyrou's program,[26] besides the contribution from the photopeaks of nuclides produced by matrix elements to the background, the Compton continuum, single and double escape peaks, and bremsstrahlung were also considered. By entering the information in Tout and Chatt's program, as well as by entering neutron flux, mass of sample and nuclear data of nuclide of interest and detector parameters, not only the optimized activation conditions were produced but also simulated γ-spectrum, detection limits, and precision for the elements of interest were presented.

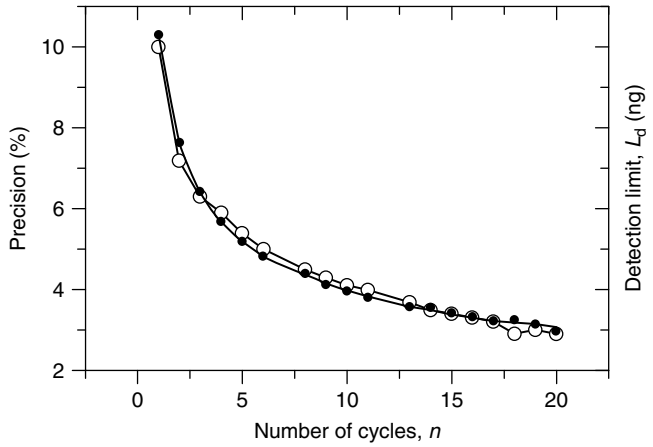

Figure 7 Variation of analytical precision and detection limit of CNAA for Se in human hair. $t_i = t_c = 30$ s, $t_d = t_w = 3$ s.

Figure 7 shows the experimentally measured variation of detection limit and analytical precision of CNAA for Se in human hair, with cyclic period of $T = t_i + t_d + t_c + t_w = 30 + 3 + 30 + 3 = 66$ s. It can be clearly seen that CNAA can significantly improve the detection limit and precision.

2.3 Estimation or Confirmation of the Half-life of Nuclide

Another function of CAA, estimation and confirmation of the half-life of nuclide by using the data obtained during the CAA, was suggested by Spyrou.[18,27,28]

For large n, $(1 - e^{-n\lambda T})$ tends to unity, Equation (5) can be reduced as

$$D_c = D_1 \left[\frac{n}{(1 - e^{-\lambda T})} - \frac{e^{-\lambda T}}{(1 - e^{-\lambda T})^2} \right] = \frac{nD_1}{(1 - e^{-\lambda T})} - \frac{D_1 e^{-\lambda T}}{(1 - e^{-\lambda T})^2} \quad (10)$$

where D_c is a linear function of n with all other parameters constant. Plotting D_c as a function of n, the slope of the line is given by Equation (11):

$$a = \frac{D_1}{(1 - e^{-\lambda T})} \quad (11)$$

and the intercept by Equation (12):

$$b = \frac{D_1 e^{-\lambda T}}{(1 - e^{-\lambda T})^2} \quad (12)$$

The half-life of the nuclide measured can be calculated by Equation (13):

$$t_{1/2} = \frac{T \ln 2}{\ln (1 - a/b)} \quad (13)$$

For this purpose, it is necessary to store the spectra of each individual cycle instead of using the accumulative spectrum.

2.4 Dead Time and Pileup Correction

CAA is usually employed for the determination of short-lived nuclides in a matrix of long-lived activation products. The activity of a sample not only changes considerably during a counting period owing to short-lived nuclide decay but also increases from cycle to cycle because of the matrix activity owing to the accumulation of longer-lived products. Consequently, a rapidly changing dead time is encountered in a counting period. The basic Equation (14) for dead time correction was given by Schonfeld[29]:

$$C = \int_0^{t_c} A_0 e^{-\lambda t} [1 - DT(t)] \, dt \quad (14)$$

where C is the actual acquired net counts in photopeak of interest, A_0 is the true initial photopeak count rate, and $DT(t)$ is the fractional analyzer dead time at time t. In order to implement this correction, the variation of fractional dead time during the counting period must be known. By the least square fitting of the experimental data, Egan et al.[30] found that $DT(t)$ is an exponential function of t, and it can be expressed by Equation (15):

$$DT(t) = B + Ce^{-kt} \quad (15)$$

where B, C, and k are constants. Hence, a correction factor (f) for dead time in a counting period can be obtained using Equation (16).

$$f = \frac{\int_0^{t_c} A_0 e^{-\lambda t} \, dt}{\int_0^{t_c} A_0 e^{-\lambda T} (1 - B - Ce^{-kt}) \, dt} \quad (16)$$

For cyclic activation with n cycles, the correction factor (F_n) can be expressed by Equation (17):

$$F_n = \frac{\sum_{n=1}^{n} \int_0^{t_c} A_0 e^{-\lambda t} dt}{\sum_{n=1}^{n} \int_0^{t_c} A_0 e^{-\lambda t} (1 - B - Ce^{-kt})_n dt} \quad (17)$$

$DT(t)$ not only varies with t in a counting period but also differs from one cycle to another. F_n is also a function of n.

It was observed that the correction factor for the dead time, f, is approximately equal to the ratio of clock time

to live time obtained from the multichannel analyzer clock data in each counting period.[30] Hence, it is usually used to correct for dead time after each cycle. In this case, it is important to store the γ-spectra of each cycle period and correct the counts of interest nuclide for dead time before summation of individual spectrum.

With increasing analyzer dead time, the pulse pileup loss (also called *summing effects* or *coincidence losses*) will become more serious. It can be overcome satisfactorily by introducing a pulser with constant frequency into the system.[30] The ratio of the areas under the pulser peaks, after dead time correction, was measured and the pileup correction factor was defined as F_p (Equation 18):

$$F_p = \frac{\text{Pulser peak area (at } \sim 0\% \text{ dead time)}}{\text{Pulser peak area (at sample dead time)}} \qquad (18)$$

However, pileup correction, unlike dead time, is implemented on the cumulative spectrum.

Wyttenbach[31] suggested another way to correct the pileup loss by using the real time (cock time), t_c, and the live time of the measurement, T_c (Equation 19):

$$F_p = \frac{P}{P_0} = 1 - \frac{2\tau}{\nu}\left(\frac{t_c}{T_c} - 1\right) \qquad (19)$$

where P_0 is the true photopeak count rate without coincidence losses, P is the actual photopeak count rate including pileup, ν is a constant of the detector system, and τ/ν can be measured by plotting P/P_0 versus t_c/T_c and subsequently be used to correct other spectra. However, it should be considered that actually τ depends on the energy of the γ ray measured.[32] In addition, the use of a loss-free counting system with pileup rejection will avoid this problem.[33]

2.5 Cumulative Activation Analysis

Guinn[34] suggested another method similar to CAA, which is based on the irradiation and counting of a number of replicates of the same sample and the individual spectra are summed to give one total spectrum. As the activity derived from activated matrix elements is the same for each cycle, the background count and the dead time in each cycle does not increase with the increasing number of replicates. Hence, this will lead to a better precision and detection limit than that of normal CAA for short-lived nuclides. The main advantage of this method is that a large representative sample can be measured even in the case of material, which activates to form high radioactive products.

By this method, Parry[35,36] measured rhodium, silver, and gold in some geological samples and reference materials and found that the detection limit was improved by a factor of 4.4–4.7 using 20 replicates in accordance with the theoretical factor of $n^{1/2}$. The main disadvantage of this method is the long time required for the preparation of n samples for each material to be analyzed, and the application is limited by the amount of samples available.

For overcoming the relatively high error resulted from inhomogeneity of sample, Farooqi et al.[37] suggested combining this technique (called *mass fraction*) with normal CAA to analyze fluorine in diet sample. In addition to good precision, a lower detection was also obtained.

3 CLASSIFICATION AND FACILITIES OF CYCLIC ACTIVATION ANALYSIS

According to source-target-detector system, Spyrou[18] classified CAA in the following five types:

1. An ''external'' beam of incident radiation, which can be mechanically chopped or electronically pulsed with stationary target, such as pulsed neutron source (or called *neutron generator*) as used by Givens et al.,[4,5] in which the waiting time can be as short as 0.5 ms. This type of CAA was termed *real* CAA. However, the radiation flux supplied in this system is usually low, i.e. the total 4π neutron flux is only $1.16 \times 10^{11}\,\text{n}\,\text{s}^{-1}$ in the facility used by Given et al.[4,5]
2. An ''external'' beam providing a continuous source of incident radiation, with the target moving into and out of the beam either linearly or by rotation. The typical source of this type of CAA is neutron source produced by an accelerator with which, in addition to fast neutron, thermal neutrons can be supplied. Further, the energy of neutron can also be chosen, which is very useful to avoid some interfering nuclear reactions. The neutron flux supplied usually ranges from 10^6 to $5 \times 10^{11}\,\text{n}\,\text{cm}^{-2}\,\text{s}^{-1}$. Most fast neutron CAA has been carried out in this kind of facility.[1,2,14,38,39]
3. An ''enclosed'' beam where the target is made to oscillate either cyclically or linearly into and out of the beam. The main radiation source in this type of CAA is the core of reactor. Not only can a high neutron flux (10^{11}–$10^{15}\,\text{n}\,\text{cm}^{-2}\,\text{s}^{-1}$) be used but also the thermal, epithermal, and fast neutron activation can be carried out. By detection of γ ray by coaxial HpGe (or Ge(Li)) and low-energy photon detector and delay neutron by neutron detector, most elements can be determined with a high accuracy and a good detection limit. Most CAA work has been implemented using this type of system.[7–13,18–20,23,25,27,30,37] Figure 8 shows a schematic diagram of this system. Similar facilities

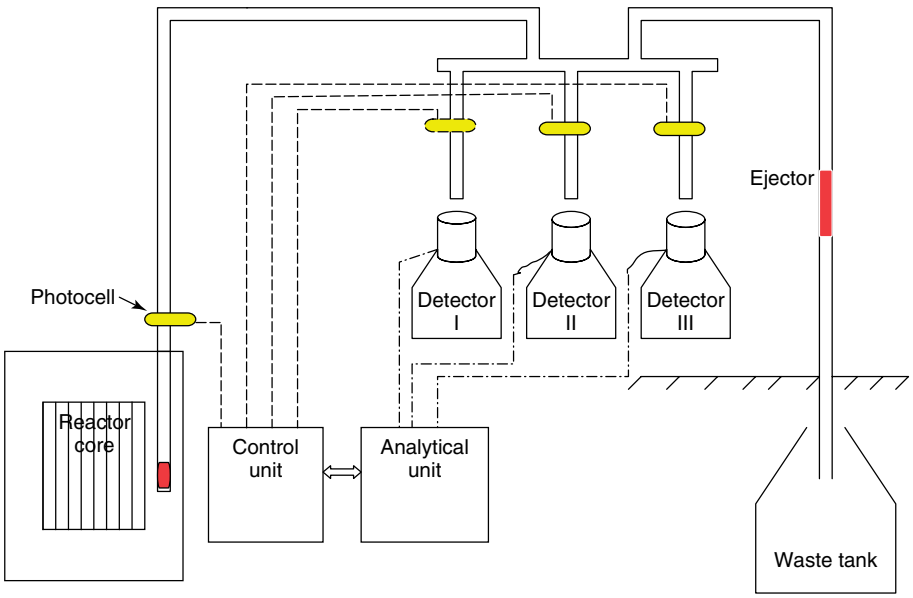

Figure 8 Schematic diagram of CNAA system.

have been set up for CNAA in such institution as University of London Reactor Center, UK,[10] Dalhousie University, Canada,[9] MNSR, China Institute of Atomic Energy,[13] MNSR, Syria,[40] MNSR, Shenzhen University, China (Personal communication), JRRS-SM, Japan Atomic Energy Research Institute, Japan,[41] Atominstitut, Vienna,[12] University of Missouri, USA.[39]

4. The case of an isotopic source, where the target remains stationary but the source is made to oscillate to and from the target position. The most commonly used isotope neutron source is Am–Be and ^{252}Cf neutron source.[42,43] A useful application of this system is in vivo NAA. The neutron flux supplied in this system is only 10^4–10^6 n cm^{-2} s^{-1}, and the detection limit is therefore high.
5. A primary source creating a secondary source, which in turn provides the irradiation beam for the target. An example of this case is the neutron interrogation system described by Mardani,[44] where a neutron beam is made to impinge on nuclear fuel, which on fission produces neutrons that act as the secondary source for the activation of a continuously circulating fluid (freon) containing fluorine. The reaction ^{19}F(n, α)^{16}N was used to determine the amount of fissile material in the fuel element by cyclic measurement of ^{16}N (7.2 s).

In all types of CAA, most work has been carried out using CNAA. Depending on the neutron source used, CNAA can be classified as reactor, neutrons produced by accelerator and neutron generator, and isotopic source CNAA. On the basis of the energy of neutron, it can be classified as thermal, epithermal, and fast neutron CAA; on the basis of the radiation emitted from the activated nuclide, it can be classified as γ ray and delay neutron CAA. Table 1 lists the elements that can be determined using different types of CNAA. The nuclear parameters of elements probably detected by cyclic thermal NAA and cyclic fast neutron activation analysis (FNAA) using their short-lived nuclides ($T_{1/2} < 100$ s) are listed in Tables 2 and 3, respectively. In addition to the elements listed in Tables 2 and 3, some elements with nuclides of medium half-lives ($t_{1/2} > 100$ s), such as Al, Mg, Ca, V, Cu, S, Mn, Na, Cl, K, Ti, Co, Br, I, In, Sr, U, and Th, may occur in the cumulative γ-spectrum or can be further counted a few minutes after the last cycle. These can, therefore, also be determined simultaneously.

4 APPLICATION OF CYCLIC ACTIVATION ANALYSIS

CAA was first suggested for the analysis of geological materials,[3–5] but with its development and maturity, this technique has been used in every field of elemental analysis, and there are numerous articles dealing with the application of CAA. Some main applications are discussed below.

4.1 Determination of Se, F, and O in Biological Materials

Selenium is known as a *biological essential trace element*, its concentration in normal biological tissues is lower than 1.0 μg g^{-1}, and a sensitive analytical technique is

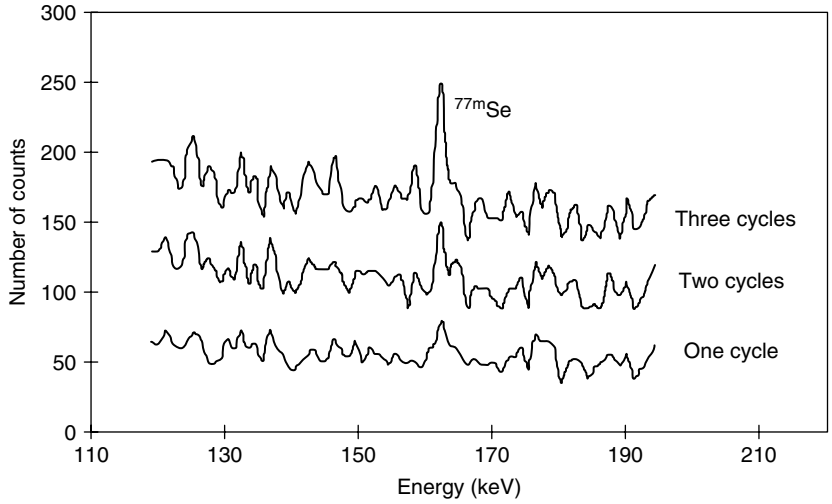

Figure 9 γ spectra of biological sample for ^{77m}Se by single-shot NAA and CNAA.

Table 1 Classification of CNAA and elements detected

Neutron	Source	Radiation measured	Elements detected
Thermal	Reactor, accelerator	γ ray	O, F, Sc, Ge, Se, Rb, In, Rh, Pd, Ag, Sb, Dy, Hf, Ce, Er, Yb, Ir, W, Pb, U
Thermal	Reactor	Delay neutron	U, Th
Epithermal	Reactor	γ ray	O, F, Ge, Rh, Ag, Ce, Er, Hf, Pt, Ir
Fast	Reactor, accelerator, isotopic source, neutron generator	γ ray	B, O, F, S, Cl, Si, Na, Mg, Ti, Ge, As, Se, Br, Rb, Y, In, Ba, Ce, Hf, Nd, Sm, Er, Tb, W, Au, Pb

required for its quantitative determination. NAA is a sensitive method for the determination of Se. In most cases, analysis is performed by measuring ^{75}Se, which has a longer half-life (120 days), and therefore needs a relatively long analytical period (weeks to months). The uses of short-lived isotope ^{77m}Se can considerably reduce the analytical time to a few minutes and thus the analytical expense. However, the detection limit and analytical precision are usually not good enough by conventional single-shot irradiation due to quick decay of ^{77m}Se. CNAA satisfactorily solved this problem, and therefore has been widely used for this purpose.[13,19,20,30] Figure 9 shows γ-spectra of ^{77m}Se in biological sample acquired by conventional single-shot NAA and CNAA. Although the background is increased, the signal of ^{77m}Se and the signal-to-background ratio are significantly improved by CNAA. Shi et al.[48] compared the conventional NAA with CNAA for the determination of Se in different food samples. In general, the detection limit of Se by measuring ^{77m}Se ($t_i : t_d : t_c = 30\,s : 10\,s : 30\,s$) is better than that by measuring ^{75}Se ($t_i : t_d : t_c = 17\,h : 16\,days : 12\,h$) using the same neutron flux by a factor of 1.5, and the detection limit using CNAA with three cycles is two times better than the single-shot NAA.

Figure 7 shows the detection limit and analytical precision for Se in human hair by conventional and cyclic NAA, respectively. It can be clearly seen that both detection limit and analytical precision were significantly improved by CNAA. Considering the interference from ^{38m}Cl ($t_{1/2} = 0.715\,s$), the optimal time parameters of CNAA of biological samples for Se given by Egan et al.[30] are $t_i = 19.5\,s$, $t_d = 3\,s$, $t_c = 18\,s$, and $t_w = 1\,s$. Under these conditions, the samples with Se concentrations as low as $10\,ng\,g^{-1}$ have been analyzed,[49] and usually 10 samples can be treated in 1 h. CNAA has also been used for the determination of Se in urine[50] and nail,[51] and as low as $1.8\,ng\,g^{-1}$ Se in urine samples has been measured.[50] In MNSR, China Institute of Atomic Energy, more than 2000 varieties of biological samples, such as human and animal tissues and body fluids, diet, and vegetation, have been analyzed for Se using this method in the last few years.

Fluorine is another human essential trace element and a bone seeker. Dental caries and osteoporosis were thought to be related to the intake of F and

Table 2 Nuclear data of elements (nuclides) detected by thermal and epithermal neutron CAA using short-lived nuclides ($t_{1/2} < 100\,s$)[45,46]

Element	Nuclei produced	Abundance of target nuclei	Half-life of produced nuclei (s)	σ_0 (b)	I_0 (b)	Main γ-ray energy (keV)
O	^{19}O	0.002	26.9	0.000161	0.00081	197.14
F	^{20}F	1.00	11.03	0.0095	0.039	1633.60
Cl	^{38m}Cl	0.242	0.715	0.05	0.38	671.3
Sc	^{46m}Sc	1.00	18.75	9.6	4.9	142.53
Ge	^{75m}Ge	0.365	47.7	0.143	0.35	139.6
Ge	^{77m}Ge	0.078	52.9	0.092	1.0	215.48
Se	^{77m}Se	0.090	17.45	21	16	161.93
Rb	^{86m}Rb	0.722	61.2	0.05	1.16	556.17
Tc	^{100}Tc	–	15.8	20	30	539.5
Rh	^{104}Rh	1.00	42.3	134	1275	555.8
Pd	^{107m}Pd	0.273	20.9	0.013	0.2	214.9
Ag	^{110}Ag	0.482	24.6	89	112	657.76
In	^{116m}In	0.957	14.1	87	–	1293.6
Sb	^{124m}Sb	0.427	93.0	0.035	0.93	645.86
Ce	^{139m}Ce	0.0025	56.4	0.15	2	757.0
Dy	^{165m}Dy	0.282	75.6	1698	425	515.5
Er	^{167m}Er	0.336	2.27	15	10	207.8
Yb	^{177m}Yb	0.127	6.41	3.8	–	104
Pt	^{199m}Pt	0.072	13.6	0.3	–	319
Hf	^{179m}Hf	0.273	18.68	53	1039	216
Ir	^{192m}Ir	0.373	87	300	1060	58
W	^{183}W	0.265	5.65	20	600	107.9
Pb	^{207m}Pb	0.241	0.80	0.03	0.1	570

the concentration in the tissues. Only one isotope, ^{20}F, can be used for its determination by NAA. Owing to very short half-life of ^{20}F (11.2 s), CNAA was therefore widely used for its determination.[9,11,37,52] Under the conditions of $t_i = t_c = 10\,s$ and $t_d = t_w = 2\,s$, a detection limit of $1–10\,\mu g\,g^{-1}$ for diet sample can be obtained by 14 cycles.[37] For the same condition, the detection limit of F in bone was reported to be $52\,\mu g\,g^{-1}$ [52]; this is mainly attributed to the high background under ^{20}F γ-ray peak (1633 keV), which is due to high ^{28}Al signal (1779 keV) in the spectra for bone samples. Since the reactions $^{23}Na(n, \alpha)^{20}F$ and $^{20}Ne(n, p)^{20}F$ can also produce ^{20}F, and the concentration of Na is much high in biological samples, the sodium interference must be corrected for. In addition, this interference can be significantly reduced using well-thermalized neutrons, such as irradiating in a thermal neutron irradiation facility in research reactors. Fluorine was actually the first element to be determined in CAA.[1,2] In addition to Se and F, De Silva and Chatt[20] also determined other elements, such as Rb, Ag, Sc, Al, Ca, Cl, Cu, K, Mg, Na, V, and I, in biological samples simultaneously by using one analytical condition; Cheng et al.[45] has also determined Ag and Sc in nail samples using CNAA.

Oxygen is a major component element of human and animal tissues and it is closely related with the metabolism of energy and mass. On the basis of low natural abundance of ^{18}O, Hou and Jiang[53] studied the metabolism of oxygen in mouse body by tracer with enriched isotope of ^{18}O and CNAA by measuring ^{19}O (26.9 s) produced by $^{18}O(n, \gamma)^{19}O$ reaction. Under time parameters of $t_i = t_c = 20\,s$, $t_d = t_w = 3\,s$ and six cycles, 0.01% of variation of the concentration of oxygen in animal blood was observed. Fast neutron reaction $^{19}F(n, p)^{19}O$ may interfere with the determination of ^{18}O; however, the low concentration of F in tissues (except bone and tooth) and low flux of fast neutron in the reactor make this interference not a serious problem.

4.2 Determination of Pb, F, Sc, and Ag in Environmental Samples

Lead is well known as a *toxic element* and its distribution as a pollutant is widespread. It is, therefore, of interest to determine its concentration in a variety of environmental materials. Lead cannot be determined by thermal NAA unless a very short-lived radionuclide ^{207m}Pb was used.[54] However, the half-life of ^{207m}Pb of only 0.8 s makes it difficult to be measured using conventional NAA method. By using a fast transfer pneumatic system, Egan and Spyrou[22] determined lead in some environmental samples. Under the conditions, $t_i = t_c = 2\,s$, $t_d = t_w = 1\,s$, and 50 cycles, the sensitivity in an interference-free matrix

Table 3 Nuclear data for elements (nuclides) detected by fast neutron CAA using short-lived nuclides ($t_{1/2} < 100\,s$)[46,47]

Element	Reaction	Abundance of target nuclei	Half-life of produced nuclei (s)	σ (mb)	Main γ-ray energy (keV)
B	$^{10}B(n, p)^{11}Be$	0.200	13.8	3.3	2125
O	$^{16}O(n, p)^{16}N$	0.99756	7.1	40	6129
F	$^{19}F(n, p)^{19}O$	1.00	27.1	20	197.14
Na	$^{23}Na(n, p)^{23}Ne$	1.00	38	44	440
	$^{23}Na(n, \alpha)^{20}F$	1.00	11.0	150	1633.60
Mg	$^{25}Mg(n, p)^{25}Na$	0.10	59.6	45	583
	$^{26}Mg(n, \alpha)^{23}Ne$	0.11	37.6	75	440
S	$^{34}S(n, p)^{34}P$	0.042	12.4	73	2127
Cl	$^{37}Cl(n, \alpha)^{24}P$	0.2423	12.4	101	2127
Ti	$^{46}Ti(n, p)^{46m}Sc$	0.080	18.7	48	142.53
Cr	$^{53}Cr(n, p)^{53}V$	0.0950	96.0	48	1006.2
Fe	$^{58}Fe(n, p)^{58}Mn$	0.0031	65	7	810.8
Ge	$^{76}Ge(n, p)^{76}Ga$	0.077	29	9.5	562.9
	$^{76}Ge(n, 2n)^{75m}Ge$	0.078	48	1000	139.6
As	$^{75}As(n, p)^{75m}Ge$	1.00	48	16	139.6
Se	$^{80}Se(n, p)^{80}As$	0.50	16	16	665.9
	$^{78}Se(n, 2n)^{77m}Se$	0.235	17.5	840	161.93
Br	$^{79}Br(n, n')^{79m}Br$	0.5069	4.9	653	207
Rb	$^{87}Rb(n, 2n)^{86m}Rb$	0.2783	61.2	750	556.17
Ru	$^{100}Ru(n, p)^{100}Tc$	0.126	15.8	15	539.5
Rh	$^{103}Rh(n, \alpha)^{100}Tc$	1.00	15.8	11	539.5
Pd	$^{104}Pd(n, p)^{104}Rh$	0.11	42.3	2.7	555.8
	$^{108}Pd(n, 2n)^{107m}Pd$	0.267	21.3	517	215
Nd	$^{141}Nd(n, 2n)^{141m}Nd$	0.2713	62.4	0.5	756.5
Sm	$^{144}Sm(n, 2n)^{143m}Sm$	0.031	66	540	754
Tb	$^{159}Tb(n, 2n)^{158m}Tb$	1.00	10.5	451	110
Er	$^{168}Er(n, 2n)^{167m}Er$	0.271	2.28	700	207.8
Au	$^{197}Au(n, n')^{197m}Au$	1.00	7.7	291	278
Hf	$^{180}Hf(n, 2n)^{179m}Hf$	0.352	18.7	600	216
Pb	$^{208}Pb(n, 2n)^{207m}Pb$	0.523	0.8	1650	570

was found to be 5 μg, and a detection limit of 30 μg g^{-1} was reported for biological matrix.

Owing to a high-resonance integral cross section of reaction $^{19}F(n, \gamma)^{20}F$, Parry et al.[55] determined fluorine in moss and soil sample with an improved detection limit by using epithermal neutron and cyclic activation.

Silver and scandium are usually determined by using their long-lived nuclides, ^{110m}Ag and ^{46}Sc. However, the use of CNAA not only reduces the analytical time but also improves the detection limit of these two elements by using their short-lived nuclides, ^{110}Ag and ^{46m}Sc. Chatt et al.[9,20] and Spyrou et al.[23] have used CNAA to analyze some environmental samples for Sc and Ag as well as other elements, such as Dy, Hf, Se, Al, V, etc.

In addition, fast neutron CAA was also used for the determination of Pb and F in environmental samples by using reactions $^{208}Pb(n, 2n)^{207m}Pb$ and $^{19}F(n, p)^{19}O$; a comparable detection was reported for the determination of F by epithermal neutron CAA and fast neutron CAA.[55] In addition to Pb and F, fast neutron CAA has also been used for many other elements such as As, Ce, Nd, and Hf.[39]

4.3 Determination of O, Sc, Si, Al, Ag, Au, Rh, Hf, Pb, and U in Geological and Industrial Materials

CAA was proposed early by Caldwell et al.[3] for the elemental analysis of lunar and planetary samples. Using this technique, [4,5] they successfully detected O, Si, Mg, and Al by fast neutron reactions, $^{16}O(n, p)^{16}N$, $^{28}Si(n, p)^{28}Al$, $^{24}Mg(n, p)^{24m}Na$, and $^{27}Al(n, \alpha)^{24m}Na$. Joshi and Agrawal[16] determined Pb and Na in soil samples by using fast neutron reactions $^{208}Pb(n, 2n)^{207m}Pb$ and $^{23}Na(n, \alpha)^{20}F$ under the condition of $t_i = t_c = 2.5\,s$, $t_d = t_w = 0.4\,s$ and 14 cycles. Their results indicated that besides Pb and Na, other eight elements, such as O, Si, Mg, As, Ce, Hf, Nd, and In, could also be determined by this method. Chatt et al.[9] determined Hf in Zirconium wire and Pb in brass by thermal neutron CAA, the detection limits reported are 0.93 μg g^{-1} and 3.6%, respectively. Parry[35,36] analyzed some geological materials for Rh,

Ag by thermal neutron CAA and Au by fast neutron CAA; she also analyzed F in industrial samples by used fast neutron CAA.[55] Owrang et al.[56] reported a fast neutron CAA method for the in situ determination of oxygen in oil by using a neutron generator, a few milligrams of oxygen in oil sample has been determined, using this technique, the concentration of oxygen in metal oxides could also be measured.

Uranium is usually determined by measuring ^{239}Np produced by ^{238}U(n, γ)^{239}U(β^-)^{239}Np reaction. However, by measuring delay neutron after thermal neutron irradiation and cyclic activation, the detection limit can be significantly improved. In addition to geological sample, it can also analyze super-pure material, such as single-crystal silicon for uranium in nanogram per gram level. Song[57] used this technique to analyze a variety of geological and industrial samples for uranium; a detection limit of 0.1 ng g^{-1} was reported.

5 CONCLUSION AND PERSPECTIVE

Since it was first proposed in the beginning of 1960s and rapidly developed in 1970s and 1980s, CAA has become a useful and mature technique for elemental analysis. Although some new CAA facilities were installed, in recent years, only some limited improvement on methodology was made. However, with the increase in CAA facilities, this technique is being applied in more wide fields, and with the increasing interest in using short-lived nuclides for activation analysis, CAA is attracting more attention. However, CAA, as one of the activation analytical techniques, can only be a supplement and not instead of the conventional activation analysis.

Comparing with conventional activation analytical techniques, CAA offers several advantages: (i) significant improvement of detection limit, analytical precision and accuracy for elements determined by using short-lived nuclides; (ii) short experimental time, hence reduced analytical expense and increased analytical capacity; (iii) capability of determining the degree of homogeneity of a sample by analyzing several portions of it for single cycles (cumulative activation); (iv) simultaneous determination of short- and medium-lived nuclides; and (v) estimating or confirmation of half-life of short-lived nuclides. However, CAA also has some disadvantages, e.g. (i) limited number of elements analyzed, i.e. only <50% of elements analyzed by conventional NAA can be determined by CNAA; (ii) some special equipment or changes of normal equipment used in conventional activation analysis are needed; (iii) dead time and pileup correction must be made; (iv) interference from impurity elements in shuttle rabbit material cannot be avoided and a special compressed gas (such as N_2, not air) should be used to transfer sample for reduced interference from ^{41}Ar and other nuclides in activated air.

In comparing with nonnuclear analytical techniques, such as ICP–MS, owing to rapidity, nondestructive analysis, and in situ analysis, CAA as one of the activation analysis techniques is more competitive. However, with the improvement of ICP–MS technique (especially nondestructive analysis by laser ablation inductively coupled plasma mass spectrometry (LA–ICP–MS)) and its popularity, the application of CAA (including all activation analysis techniques) is becoming less.

ABBREVIATIONS AND ACRONYMS

σ	Activation cross section
C	Actual acquired net counts in photopeak of interest
f	Correction factor for dead time in one cycle
F_n	Correction factor for dead time of n cycles
F_p	Correction factor for pileup
t_c	Counting time
S/m	Counts per unit mass element of interest
T	Cycle period ($T = t_i + t_d + t_c + t_w$)
CAA	Cyclic activation analysis
CNAA	Cyclic neutron activation analysis
λ	Decay constant ($\lambda = \ln 2/t_{1/2}$)
t_d	Delay time (the time interval between the end of irradiation and the start of counting)
D_1	Detector response (or the number of counts)
D_n	Detector response in nth cycle
ε	Efficiency of the detector
FNAA	Fast neutron activation analysis
DT(t)	Fractional analyzer dead time at time t
$t_{1/2}$	Half-life of the nuclide of interest
ICP–MS	Inductively coupled plasma mass spectrometry
I	Intensity of the radiation of interest
LA–ICP–MS	Laser ablation inductively coupled plasma mass spectrometry
NAA	Neutron activation analysis
n	Number of cycles
N	Number of target nuclei
N^*	Number of radioactive nuclei
PCAA	Pseudocyclic activation analysis
L_d	Qualitative detection limit

Φ	Radiation (neutron) flux intensity
δ	Relative standard deviation
I_0	Resonance integral cross section
σ_0	Thermal neutron activation cross section
P	The actual photopeak count rates including pileup
P_0	The true photopeak count rates without coincidence losses
T_c	The live time of the measurement
${}_bD_c$	The cumulative detector response for the nuclide making the major background contribution to the signal from the nuclide of interest
t_i	Time of irradiation
D_c ($_sD_c$)	Total cumulative detector response in all n cycles
T_t	Total experiment time
A_0	True initial photopeak count rate
t_w	Waiting time (the time interval between the end of counting and the start of irradiation)

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides

Mass Spectrometry **(Volume 13)**
Inorganic Substances, Mass Spectrometric in the Analysis of

REFERENCES

1. O.U. Anders, 'Determination of Fluorine by Neutron Activation', *Anal. Chem.*, **32**(10), 1368–1369 (1960).
2. O.U. Anders, 'Use of Very Short-lived Isotopes in Activation Analysis', *Anal. Chem.*, **33**(12), 1706–1709 (1961).
3. R.L. Caldwell, W.R. Mills Jr, L.S. Allen, P.R. Bell, R.L. Heath, 'Combination Neutron Experiment for Remote Analysis', *Science*, **152**, 457–465 (1966).
4. W.W. Givens, W.R. Mills, R.L. Caldwell, 'Cyclic Activation Analysis', *Proceedings of International Conference on Modern Trends in Activation Analysis, 1968*, National Bureau of Standard, Gaithersburg, Maryland, 139–147, 1969.
5. W.W. Givens, W.R. Mills, R.L. Caldwell, 'Cyclic Activation Analysis', *Nucl. Instrum. Methods*, **80**, 95–103 (1970).
6. A. Tani, Y. Matsuda, Y. Yuasa, N. Kawai, 'Repeat Pulse Activation Analysis for Measurement of Short Half Life Radionuclides', *Radiochem. Radioanal. Lett.*, **1**, 155–161 (1969).
7. F. Ozek, *'Cyclic Activation and its Applications'*, University of Surrey, M.Sc. Thesis, 1973.
8. N.M. Spyrou, K. Ingle, F. Ozek, 'Cyclic Activation in Trace Element Analysis of Environmental Samples', *Proceedings of 2nd International Conference on Nuclear Methods in Environmental Research*, Columbia, Missouri, 106–120, 1974.
9. A. Chatt, K.N. DeSilva, J. Holzbecher, D.C. Stuart, R.E. Tout, D.E. Ryan, 'Cyclic Neutron Activation Analysis of Biological and Metallurgical Samples', *Can. J. Chem.*, **59**, 1660–1664 (1981).
10. G.D. Burholt, E.A.Y. Caesar, T.C. Jones, 'The Fast Cyclic Activation System for Neutron Activation Analysis in the University of London Reactor', *Nucl. Inst. Method*, **204**, 231–134 (1982).
11. V.L. Spate, J.S. Morris, C.K. Baskett, M.M. Mason, C.L. Reams, D. Hunter, W.C. Willett, 'Determination of Fluorine in Human Nails Via Cyclic Instrumental Neutron Activation Analysis', *J. Radioanal. Nucl. Chem., Articles*, **179**(1), 27–33 (1994).
12. S.S. Ismail, F. Grass, G.P. Westphal, 'A New Efficient Set-up for Cyclic and Pseudocyclic Short Time Activation Analysis With High Count Rates', *J. Radioanal. Nucl. Chem.*, **233**(1–2), 149–153 (1998).
13. X.L. Hou, H.A. Das, 'Accuracy, Precision and Sensitivity in Cyclic INAA with Short-lived Radionuclides', *J. Radioanal. Nucl. Chem.*, **223**(1–2), 67–72 (1997).
14. P. Ila, 'Determination of Total Oxygen in Standard Rocks by Cyclic Activation Analysis Using 14 MeV Neutron Generator', *J. Radioanal. Nucl. Chem., Articles*, **148**(2), 415–422 (1991).
15. S.E.B. Nonie, K. Randle, A.G. Blackband, 'Development of a computer-based system for Fast Neutron Cyclic Activation Analysis', *J. Radioanal. Nucl. Chem., Articles*, **167**(1), 89–95 (1993).
16. G.C. Joshi, H.M. Agrawal, 'Multielemental Analysis of Soil Samples by Fast Neutron Activation', *J. Radioanal. Nucl. Chem., Articles*, **198**(2), 475–485 (1995).
17. X.L. Hou, 'Cyclic Activation Analysis', in *Encyclopedia of Analytical Chemistry*, ed. R.A. Meyers, John Wiley & Sons Ltd, Chichester, 12447–12459, 2002.
18. N.M. Spyrou, 'Cyclic Activation Analysis, A Review', *J. Radioanal. Chem.*, **61**(1–2), 211–242 (1981).

19. A. Chattopadhyay, K.N. DeSilva, 'Pseudo-Cyclic Neutron Activation analysis of Ag, F, Rb, Sc and Se in Biological Samples', *Trans. Am. Nucl. Soc.*, **32**, 185–186 (1979).

20. K.N. DeSilva, A. Chatt, 'A Method to Improve Precision and Detection Limits for Measuring Trace Elements Through Short-lived Nuclides', *J. Trace Microprobe Technol.*, **1**(3), 307–337 (1982–1983).

21. H.U. Fanger, R. Pepelnik, W. Michaelis, 'Fast Neutron Activation Analysis with Short-lived Nuclides', *J. Radioanal. Chem.*, **61**, 147–163 (1981).

22. A. Egan, N.M. Spyrou, 'Determination of Lead Via Lead-207m Using Cyclic Activation and A Modified Sum-coincidence System', *Anal. Chem.*, **48**(13), 1959–1962 (1976).

23. N.M. Spyrou, S.A. Kerr, 'Cyclic Activation: the Measurement of Short-lived Isotope in the Analysis of Biological and Environmental Samples', *J. Radioanal. Chem.*, **48**, 169–183 (1979).

24. M. Colin, C. Friedli, P. Lerch, 'Heavy Ions Activation Analysis: Theoretical Approach for Optimizing Cyclic Analysis', *Nucl. Instrum. Methods, Phys. Res.*, **B10**, 1062–1064 (1985).

25. R.E. Tout, A. Chatt, 'The Effect of Sample Matrix on Selection of Optimum Timing Parameters in Cyclic Neutron Activation Analysis', *Anal. Chim. Acta*, **133**, 409–419 (1981).

26. M.A. Al-Mugrabi, N.M. Spyrou, 'The use of Simulation for the Optimization of the Signal-to-noise Ratio in Cyclic Activation Analysis', *J. Radioanal. Nucl. Chem.*, **110**(1), 67–77 (1987).

27. N.M. Spyrou, C. Adesanmi, M. Kidd, L.G. Stephens-Newsham, A.Z. Ortaoval, F. Ozek, 'Usefulness of Thermal and Epithermal Cyclic Activation Analysis with Reactor System', *J. Radioanal. Chem.*, **72**(1–2), 155–182 (1982).

28. A.Z. Ortaovali, F. Ozek, I. Celenk, N.M. Spyrou, 'Study on Some Problems and Application of Cyclic Activation Analysis', *J. Radioanal. Chem.*, **61**(1–2), 175–182 (1981).

29. E. Schonfeld, 'Alpha? A Computer Program for the Determination of Radioisotopes by Least Squares Resolution of the Gamma-ray Spectra', *Nucl. Instrum. Methods*, **42**, 213–218 (1966).

30. A. Egan, S.A. Kerr, M.J. Minski, 'Determination of Selenium in Biological Materials Using ^{77m}Se and Cyclic Activation Analysis', *Radiochem. Radioanal. Lett.*, **28**(5–6), 369–378 (1977).

31. A. Wyttenbach, 'Coincidence Losses in Activation Analysis', *J. Radioanal. Chem.*, **8**, 335–343 (1971).

32. Z.B. Alfassi, A. Tsechansky, A.P. Kushelevsky, 'On the Correction of the Self Coincidence in Gamma Ray Spectra Dependence of Resolution Time on Energy', *J. Radioanal. Chem.*, **55**, 135–139 (1980).

33. K. Heydron, E. Damsgaard, 'Validation of a Loss-free Counting System for Neutron Activation Analysis with Short-lived Indicators', *J. Radioanal. Nucl. Chem.*, **215**(2), 157–160 (1997).

34. V.P. Guinn, 'Cyclic Nuclear Activation Analysis', *Radiochem. Radioanal., Lett.*, **44**, 133–140 (1980).

35. S.J. Parry, 'Cumulative Neutron Activation Analysis for the Improved Detection of Short-lived Nuclides', *J. Radioanal. Chem.*, **75**(1–2), 253–258 (1982).

36. S.J. Parry, 'Cumulative Neutron Activation Analysis of Short-lived Radionuclides for the Analysis of Large Samples in Mineral Exploration', *J. Radioanal. Nucl. Chem.*, **112**(2), 383–386 (1987).

37. A.S. Farooqi, W. Arshed, O.A. Akanle, C. Jeynes, N.M. Spyrou, 'Fluorine Determination in Diet Samples Using Cyclic INAA and PIGE Analysis', *J. Radioanal. Nucl. Chem.*, **161**(1), 71–78 (1992).

38. S.E. Nonie, K. Randle, 'The Multielement Potential of Fast Neutron Cyclic Activation Analysis', *J. Radioanal. Nucl. Chem.*, **185**(1), 35–44 (1994).

39. G.C. Joshi, H.M. Agrawal, 'Multielemental Analysis of Soil Samples by Fast Neutron Activation', *J. Radioanal. Nucl. Chem.*, **198**(2), 475–484 (1995).

40. I. Khamis, I. Othman, M. Nasri, M. Bakkour, 'Transfer Time Optimization of a Rapid Cyclic Instrumental Neutron Activation Analysis for Trace Element Detection', *Review of Scientific Instruments*, **77**(2), 1492–1494 (2001).

41. C. Yonezawa, T. Kurosawa, H. Matsue, 'Rapid Determination of Trace Amounts of Fluorine in Silicon Nitride by Instrumental Neutron Activation Analysis', *Bunseki Kagaku*, **47**(1), 829–834 (1998).

42. I.P. Matthews, N.M. Spyrou, 'Multielement Analysis of Bulk Matrices by Measurement of Prompt and Delay Gamma-ray as Well as Cyclic Activation Using An Isotopic Neutron Source', *Int. J. Appl. Radiat. Isot.*, **33**(1), 61–68 (1982).

43. S.J. Ryde, W.D. Morgan, A. Sivyer, C.J. Evans, J. Dutton, 'A Clinical Instrument for Multielement in vivo Analysis by Prompt, Delayed and Cyclic Neutron Activation Using ^{252}Cf'', *Phys. Med. Biol.*, **32**(10), 1251–1271 (1987).

44. K. Mardani, 'Neutron Interrogation of Fissile Material Using A Remote Detection System', University of Surrey, M.Sc. Thesis, 1978.

45. IEAE, 'Practical Aspects of Operation, A Neutron Activation Analysis Laboratory', IAEA-TECDO-564, Vennia, 1991.

46. J.R. Parrington, H.D. Knox, S.L. Breneman, E.M. Baum, F. Feiner, *'Chart of the Nuclides'*, 15th edition, General Electric Co. and KAPL Inc., 1996.

47. Z.B. Alfassi, *Activation Analysis*, Vol. II, CRC Press, Boca Raton, FL, 1990.

48. Y. Shi, E.E. Sullivan, J. Holzbecher, A. Chatt, 'Determination of Selenium in Canadian Food Items by Cyclic

Instrumental Neutron Activation Analysis', *Biol. Trace Elem. Res.*, **71–72**, 377–386 (1999).

49. A. Chatt, H.S. Dang, B.B. Fong, C.K. Jayawickreme, L.S. MaDowell, D.L. Pegg, 'Determination of Trace Elements in Food by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **124**(1), 65–77 (1988).
50. M. Navarrete, L. Cabrera, T. Martinez, G. Revel, J.P. Meyer, J. Guarner, 'Relationship Between the Cervical Uterine Cancer Evolution and Selenium Concentration in Urine Determined by NAA', *J. Radioanal. Nucl. Chem.*, **217**(2), 167–169 (1997).
51. T.P. Cheng, J.S. Morris, S.R. Koirtyohann, V.L. Spate, C.K. Baskett, 'The Analysis of Human Nail for 24 Elements via k0 and Cyclic Neutron Activation Analysis', *Nucl. Instrum. Methods*, **A353**, 457–460 (1994).
52. S.A. Kerr, N.M. Spyrou, 'Fluorine Analysis of Bone and Other Biological Materials: A Cyclic Activation Method', *J. Radioanal. Chem.*, **44**, 159–173 (1978).
53. X.L. Hou, W.J. Jiang, *Study of Metabolism of Oxygen in Mice by ^{18}O Tracer and Cyclic Neutron Activation Analysis*, Annual Report of China Institute of Atomic Energy, Chinese Atomic Energy Press, Beijing, 1993.
54. M. Wiemick, S. Amiel, 'Use of Very Short-lived Nuclide in Non-destructive Activation Analysis with A Fast Shuttle Rabbit, II Determination of Lead by Means of ^{207m}Pb', *J. Radioanal. Chem.*, **3**, 393–403 (1969).
55. S.J. Parry, R. Benzing, K.L. Bolstad, E. Steinnes, 'Epithermal/ Fast Neutron Cyclic Activation Analysis for the Determination of Fluorine in Environmental and Industrial Samples', *J. Radioanal. Nucl. Chem.*, **244**(14), 67–72 (2000).
56. F. Owrang, H. Mattsson, A. Nordlund, 'A Method for In-situ Quantification of Oxygen in Oil Using fast Neutron Activation Analysis', *Kerntechnik*, **69**(1–2), 51–57 (2004).
57. Q.X. Song, 'Improvement in the Determination of Trace Uranium by Repeating Activation and Delayed Neutron Counting, Determination of The Uranium Content of Multi-crystalline and Single-Crystal Silicon Samples', *Chinese J. Nucl. Techn.*, **10**(6), 35–46 (1987).

Direct Reading Instruments for the Determination of Aerosols and Particulates

David Y.H. Pui and Da-Ren Chen
University of Minnesota, Minneapolis, USA

1 Introduction **1**
1.1 Aerosols and Particles 1
1.2 Direct-reading Instruments 2
2 Particle Size Distributions and Instrument Measuring Ranges **2**
3 Size Distribution Measurement **3**
3.1 Differential Mobility Analyzer 4
3.2 Diffusion Batteries 5
3.3 Optical Particle Counter 6
3.4 Aerodynamic Particle Sizer and Aerosizer 9
3.5 Electrical Low Pressure Impactor 10
3.6 Phase Doppler Analyzer 10
4 Total (Integral) Concentration Detectors **11**
4.1 Condensation Particle Counter 11
4.2 Light-scattering Photometers and Nephelometers 12
4.3 Quartz-crystal Microbalance and Surface Acoustic Wave Microbalance 13
4.4 Tapered-element Oscillating Microbalance 13
5 Monodisperse Aerosol Standards **14**
5.1 Electrospray Aerosol Generator 14
5.2 Vibrating Orifice Aerosol Generator 15
5.3 Electrostatic Classifier Aerosol Generator 15
6 Summary **16**
Abbreviations and Acronyms **18**
References **18**

Solid and liquid particles suspended in a gas are referred to as aerosols. The suspending medium can be air, or some other gases, such as argon, nitrogen, helium, etc. Aerosols have found numerous scientific and technical applications. Beneficial applications include pharmaceutical aerosols for medical inhalation therapy or diagnosis, industrial aerosols for coating wave guides, refractories, gas sensors and semiconductors, as well as particles of modern materials for forming advanced ceramics, optical and optoelectronic devices, cosmetic products, paint pigments, and catalysts. Harmful effects of aerosols include atmospheric aerosols that contribute to urban air pollution and global radiation balance, dust particles ingested by engines causing premature failure of the equipment, and particles deposited on semiconductor wafers causing device failure and product yield loss.

Aerosols and particles can be analyzed by collecting them on a filter or a sampling surface for subsequent microscopic, gravimetric, or chemical analyses. The collection by the samplers gives an integrated sample and a time averaged concentration over a period of hours or days. Direct reading instruments carry out the sampling and analysis within the instrument and the property of interest can be obtained in near real-time. Direct reading instruments are available to cover particles in the size range of 0.002 to 100 μm. These instruments have fast time response and can follow rapid changes in both particle size and concentration. Good counting statistics can also be obtained because repeated measurements can be performed in a short time. However, these instruments usually rely on indirect sensing techniques and more calibration efforts are usually required.

1 INTRODUCTION

1.1 Aerosols and Particles

Solid and liquid particles suspended in a gas are referred to as aerosols. The suspending medium can be air, or some other gas, such as argon, nitrogen, helium, etc. The branch of science dealing with such gas-borne particles is known as aerosol science.

Aerosol is one of three types of dispersed systems found in nature and in various technical applications: aerosol (solid or liquid particles dispersed in a gas), hydrosol (solid or liquid particles suspended in a liquid), and foam (small gas bubbles dispersed in a solid or liquid).

Aerosols have found numerous scientific and technical applications. Atmospheric aerosols due to air pollution, e.g. smog, are the principal cause of visibility degradation. They contribute to the cloud formation process and affect global radiation balance. Airborne lead, asbestos, and other harmful substances can lead to serious health consequences when inhaled. Atmospheric dust particles ingested by an engine, a gas turbine, or other power generating equipment can cause wear and the premature failure of the equipment. Particles depositing on semiconductor wafers during manufacturing of large-scale integrated circuit chips can cause defects, leading to device failure and loss of product yield.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Aerosols are used in many beneficial applications such as the production of advanced materials and the use of aerosols for medical inhalation therapy or diagnosis. Aerosols are used in material synthesis to produce powders and bulk materials,[1] and for coating optical and microwave waveguides.[2] Owing to their simplicity and freedom from contamination, aerosol processes are used in the production of a variety of modern materials, such as advanced ceramics, carbon black, paint pigments, and high-temperature superconductors, and for coating waveguides, refractories, gas sensors, and semiconductors. An exciting new area for the materials industry is the production of nanostructure materials consisting of nanometer-size particles. Nanoparticles possess many desirable properties, including improved hardness, reduced internal friction (i.e. improved ductility), reduced melting point, and special optical and magnetic properties.[3] These properties are the basis of many so-called ''high-tech'' applications, including quantum dots, drill bit coatings, fuel cells, and tunable lasers.

Biotechnology is an emerging area for aerosol applications. Determination of the molecular weights of biopolymers, including oligonucleotides and proteins, is of fundamental importance to biotechnology development. These biomolecules with molecular weights up to 130 000 are, in essence, nanometer particles. Aerosol technique is used as a supplement to mass spectrometry for detecting and measuring these macromolecule particles.

1.2 Direct-reading Instruments

Particle concentration, size distribution and chemical composition are used to characterize an aerosol system. The concentration can be weighted by number (particles cm^{-3}), surface ($\mu m^2\,cm^{-3}$), volume ($\mu m^3\,cm^{-3}$), or mass ($\mu g\,cm^{-3}$) in a unit volume of gas medium. Other physical parameters, such as the total light scattering coefficient of the aerosol plume, may also be important and need to be measured.

Aerosols can be characterized by using sampling devices or by direct reading instruments. In the sampling devices, aerosols are collected on a substrate for subsequent gravimetric and/or chemical analysis, e.g. wet-chemical technique or proton induced X-ray fluorescence technique. The collection is usually done with a filter sampler to give an integrated sample and a time averaged concentration measurement. Attention must be given to the possibility of chemical interactions among the collected particles and of artifact formation due to reaction between the substrate and the more reactive chemical species in the gas phase, e.g. the reaction of glass fiber filters with SO_2 in air.[4]

Direct-reading instruments involve sampling and analyzing aerosols within the instruments, and properties of interest can often be obtained in real-time. Direct-reading instruments are available to cover particles in the size range of 0.002 to 100 μm. These instruments have fast time response and can follow rapid changes in both particle size and concentration. Recent advances enable near real-time measurements of particle chemical compositions. Good counting statistics can also be obtained because repeated measurements can be performed in a short time. However, these instruments usually rely on indirect sensing techniques and more calibration efforts are usually required.

2 PARTICLE SIZE DISTRIBUTIONS AND INSTRUMENT MEASURING RANGES

Aerosol particles may consist of microscopic bits of materials in the form of molecular clusters of only a few nanometers in diameter, to particles of a near macroscopic size of a fraction of a millimeter in diameter. The nominal size range of interest in aerosol science is 0.002 to 100 μm. The lower size limit, 2 nm, corresponds to the smallest particles that can be detected by a direct reading instrument, e.g. the ultrafine condensation particle counter (UCPC). The largest particles are those that would begin to fall out from a gas medium too quickly, e.g. rain drops, to be considered an aerosol. Particle size determination is important because many aerosol properties are a function of particle size. Particle deposition on wafers in semiconductor manufacturing and deposition in different regions of the human respiratory tract in health studies are determined by particle size.

Particle size distribution is usually presented as particle concentration in number, surface or volume (mass) weightings as a function of particle size. Vast databases have been collected on atmospheric size distributions from many air pollution studies. As a result, there is now a general understanding of the aerosol formation and removal mechanisms in the atmosphere. Figure 1 shows a volume-concentration size distribution of atmospheric aerosols measured near a combustion source, i.e. automobile traffic. Atmospheric aerosol is seen to consist of three modes and cover a size range of 0.002–100 μm. The three principal modes are referred to as the nuclei mode (0.002–0.05 μm), the accumulation mode (0.05–2 μm) and the coarse particle mode (2–50 μm). The accumulation mode and the coarse particle mode can be observed in nearly every atmospheric measurement, whereas the nuclei mode can only be measured near an aerosol source involving high temperature and other gas-to-particle conversion processes, e.g. combustion and photochemical

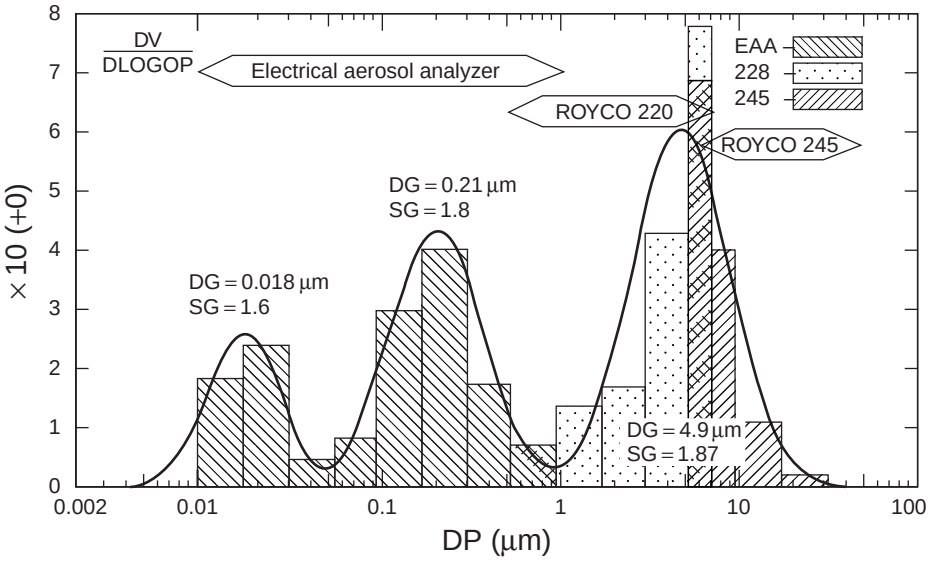

Figure 1 Particle size distribution weighted by volume showing the trimodal nature of atmospheric aerosols.[5]

aerosols. The relative amounts of material in the different modes depend on the locality and weather and on the pollution by local industry and motor vehicle traffic.

The chemical compositions of aerosol particles are different for the three modes. Particles in the coarse particle mode generally contain various crustal elements such as silicon, iron, aluminum, calcium, sea salt and plant particles. Particles in the nuclei and accumulation modes contain sulfates, nitrates, elemental and organic carbon, ammonium, lead and other trace constituents. This evidence suggests that the formation mechanisms are different for the fine and coarse particle modes. It is expected that the removal mechanisms will also be different owing to the wide particle size range covered by the three modes. A comprehensive review of the formation and removal mechanisms of atmospheric aerosols can be found in the paper by Whitby.[5]

A variety of instruments are available to measure particle size distribution and concentration in the entire size range from 0.002 to 100 μm. Figure 2 shows the available particle size analyzers and the size ranges they cover. Since most measuring instruments have a dynamic range of approximately two decades in particle size, several instruments based on different operating principles are needed to cover the five decades of particle size in the atmosphere.

Different direct-reading instruments measure different aerosol properties. Although many instruments provide data in particle ''size'', this ''size'' is actually derived from one of many possible particle properties, such as its gravimetric, optical, aerodynamic, mechanical, or force field mobility behaviors. It is expected that the measured particle size will be different for different instruments used. The diameter of an equivalent sphere is usually defined to specify the particle size measured by instruments based on different principles. For example, the aerodynamic equivalent sphere or Stokes equivalent sphere may be used to designate the same particle, and refer to the diameter of a unit density sphere or the same density sphere, respectively, having the same settling speed as the particle in question. Other equivalent sphere diameters can be based on other particle properties, e.g. diffusion, electrical mobility, and light scattering. Thus, the particle sizes measured by different principles may not be directly compared without some correction of the data to account for the differences.

3 SIZE DISTRIBUTION MEASUREMENT

Figure 2 shows the available commercial direct reading instruments for size distribution measurements. They cover the entire size range of interest from 2 nm to 100 μm. The measured size distributions can be integrated to obtain the total number, surface, and volume (mass) concentrations. Alternatively, these total concentrations can be measured by compact, relatively inexpensive integral concentration detectors (Figure 2). These detectors also cover most of the five decades in particle size range. The upper limits are often set by the efficiency of the sampling inlets. Since most direct reading instruments rely on particle detection by electronic transducers, monodisperse aerosol standards are needed to calibrate these instruments. Some of the principal methods for generating these aerosol standards in the five decades of particle size range are also shown in Figure 2.

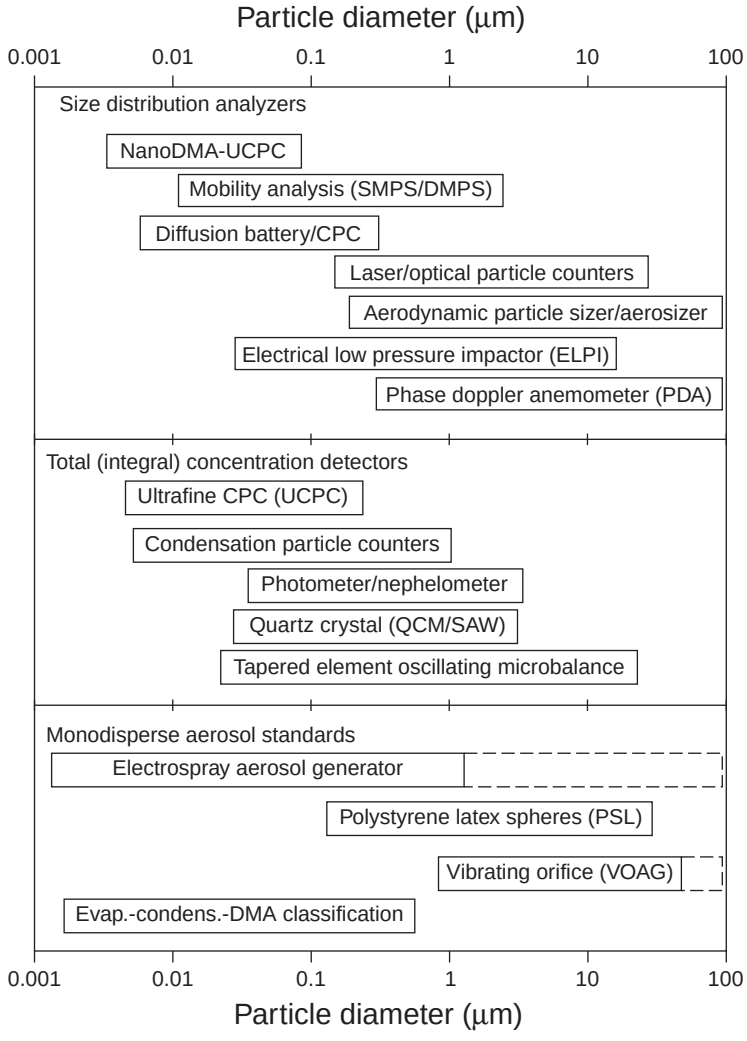

Figure 2 Particle size range of size distribution analyzers, total (integral) concentration detectors, and monodisperse aerosol standards.

3.1 Differential Mobility Analyzer

The high electric mobility of submicrometer particles in an electric field makes it possible to separate and classify electrically charged aerosol particles. If the electric mobility is a monotonic function of the particle size, size classification or size distribution, measurement can be made on the basis of the particle electrical mobility.

Figure 3 shows the principle of the differential mobility analysis technique. Charged aerosol and clean sheath air are introduced into the annular space between two concentric cylinders. Clean air flows along the inner cylinder that consists of a narrow slit on the downstream side of the cylinder. A high voltage is applied to the inner cylinder relative to the outer cylinder that is electrically grounded. The electric field between the two cylindrical electrodes causes charged particles to be attracted to the inner cylinder. Small particles with higher electrical mobility will be collected above the slit, while the large particles with lower electrical mobility will be collected below the slit or be carried away by the main flow. Only those particles with the correct electric mobility and of narrow mobility range (monodisperse) will be extracted by the slit flow and detected by a particle counter downstream. By varying the voltage on the inner cylinder and measuring the corresponding particle concentration, a mobility distribution can be obtained which can be inverted to obtain the particle size distribution.

Knutson and Whitby[6] obtained an analytical equation to describe particle motion inside the differential mobility analyzer (DMA). They showed that the resolution of the instrument is a function of aerosol/sheath flow ratio in the DMA. High resolution can be obtained by reducing the flow ratio. Recent numerical studies by Chen and Pui[7] and Chen et al.[8] have demonstrated that the DMA

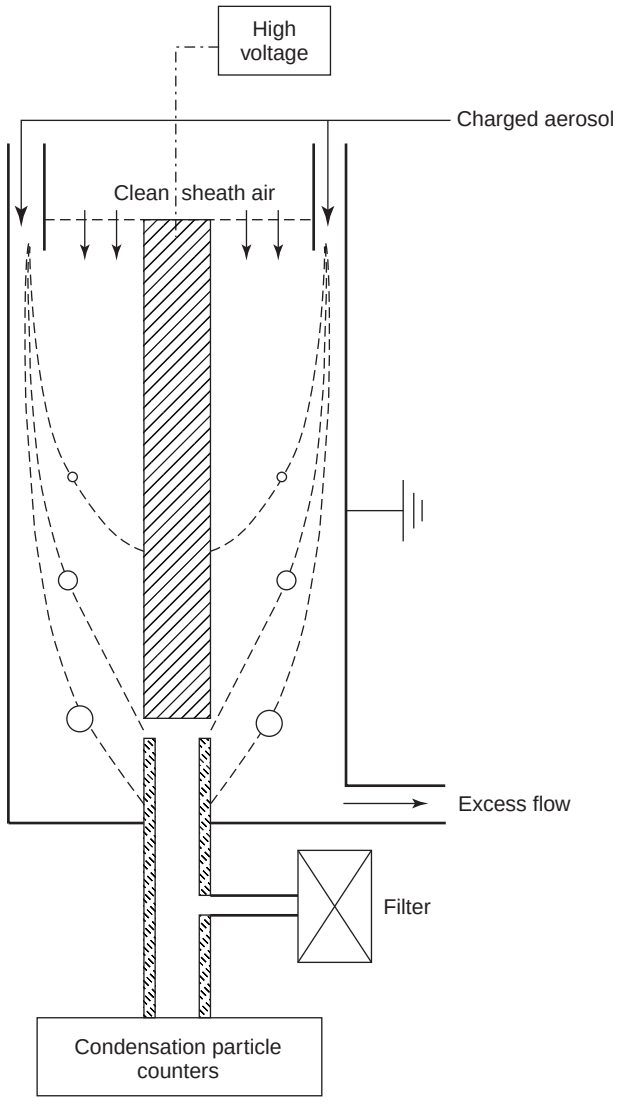

Figure 3 Principle of the differential mobility analysis technique.

is capable of resolving particle size distribution with a geometrical standard deviation of 1.01. The numerical codes also facilitated the design of a new generation of DMAs such as the nanometer aerosol differential mobility analyzer (Nano-DMA) used for measuring particle size distribution down to 3 nm.[9]

Figure 4 shows a schematic diagram of the available commercial DMAs in three different configurations, namely the TSI-DMA,[10] the Hauke 3/150[11] and the Spectromete de Mobilité Electrique Circulaire (SMEC).[12] Another radial-design DMA similar to that of SMEC was developed by Zhang et al.[13] Fissan et al.[14] evaluated the three different types of DMAs and found good performance for particle size larger than 10 nm. Below 10 nm, the instrument resolution and sensitivity (particle loss within the DMA) were degraded due to the high diffusivity of the nanoparticles. Subsequent development by Chen et al.[9] has resulted in a new Nano-DMA that is optimized for the nanoparticle size range. The Nano-DMA, shown schematically in Figure 5, provides a three-fold increase in the resolution and a three-fold decrease in the diffusional loss than the best available DMA.

One significant advance on using the DMA for particle size distribution measurement was developed by Wang and Flagan.[15] They made use of a scanning electric field, in place of changing the electric field in discrete steps in the traditional differential mobility particle sizer (DMPS) technique, to speed up the cycle time of the mobility analyzer considerably (from a cycle time of 3 min to less than 10 s). Using an exponential ramp in the field strength, the particles are classified in the time-varying electric field while maintaining a one-to-one correspondence between the time a particle enters the classifier and the time it leaves. The scanning mobility particle sizer (SMPS) is a commercial instrument that makes use of this scanning feature.

The DMA technique requires that the classified particles be electrically charged to a known level. Particles can acquire charges by exposing them to high concentration of either unipolar ions or bipolar ions. Devices are available based on a unipolar charging technique[10] or a bipolar charging technique.[16] Recent advances include the development of a high efficiency, high throughput unipolar aerosol charger for nanoparticles by Chen and Pui.[17]

3.2 Diffusion Batteries

Since the rate of diffusion of aerosol particles to a solid surface is a function of particle size, particle loss in a diffusion collector can be used for size distribution measurement. The technique has been used for many years for size distribution measurements of small particles below 0.1 μm.[18] A simple diffusion battery design consists of a single capillary tube, or a capillary tube bundle through which the aerosol is passed. A condensation particle counter (CPC) is then used to measure the upstream and downstream aerosol concentrations. The aerosol penetration through the diffusion collector can then be taken as a measure of particle size. By arranging a number of these diffusion collectors in series, the size distribution of the aerosol can be measured.

Several novel diffusion batteries have been developed by Sinclair,[19] who used metal disks containing uniform parallel holes of a finite length – the so-called collimated hole structure – and layers of fine stainless steel screens as diffusion collectors. Detailed studies of wire screens

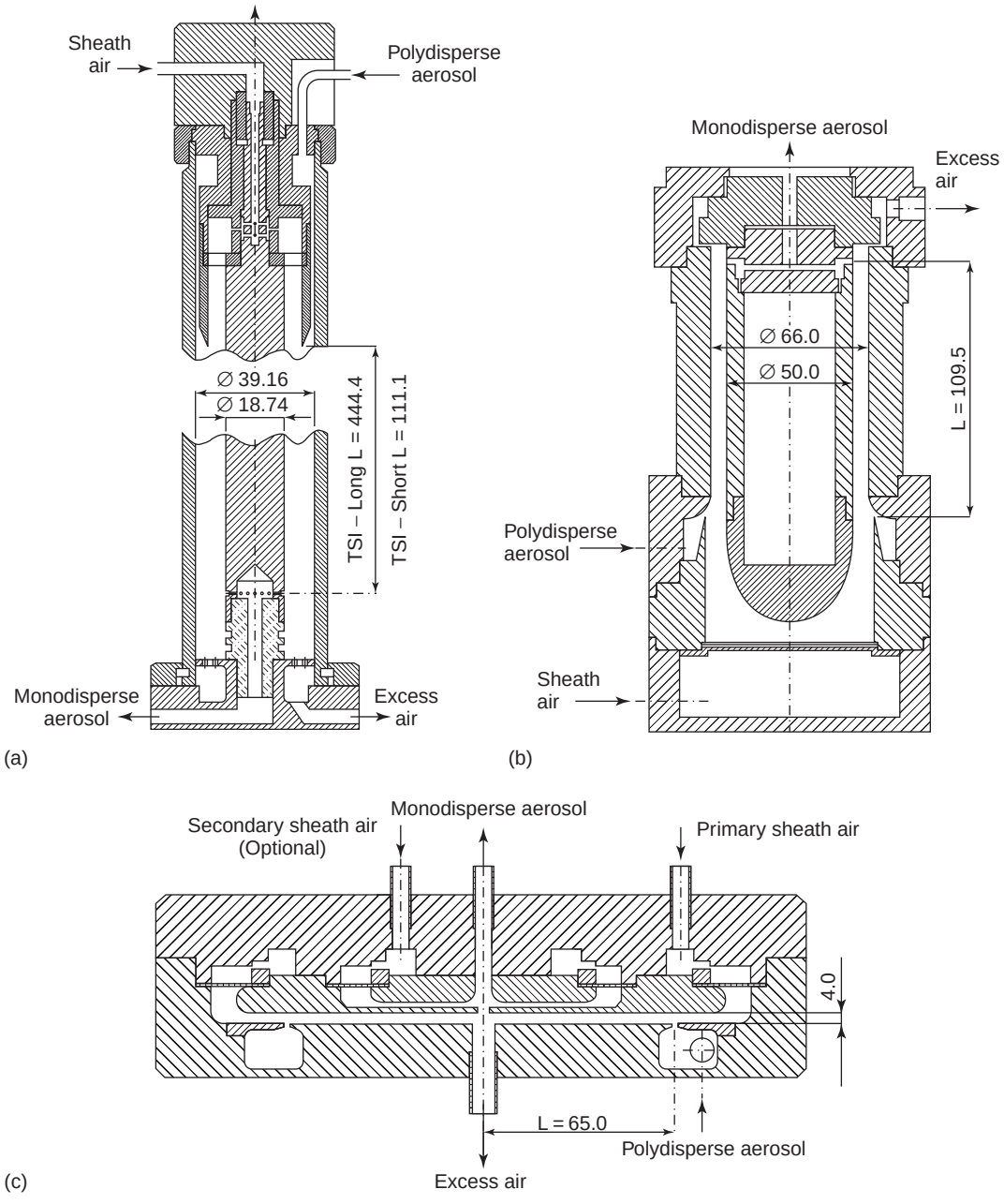

Figure 4 Schematic diagram of the available commercial DMAs in three different configurations: (a) the TSI-DMA;[10] (b) the Hauke 3/150;[11] and (c) the SMEC.[12]

as diffusion collectors have been reported.[20–22] Various sophisticated data reduction techniques based on Simplex and other minimization techniques for the diffusion battery have also been reported. The diffusion battery/condensation particle counter (DB/CPC) technique of aerosol size distribution measurement is now widely accepted.

The diffusion battery has also been used as a particle separator for size selective particle sampling. Lundgren and Rangaraj[23] use it for in-stack particle sampling. George[24] and Sinclair et al.[25] made use of the screen diffusion batteries to measure the submicron radioactive aerosols. One of the requirements for accurate diffusion battery measurement is that the counting efficiency of the CPC should be well characterized.

3.3 Optical Particle Counter

The optical particle counter (OPC) is widely used for size distribution measurements in both indoor and outdoor

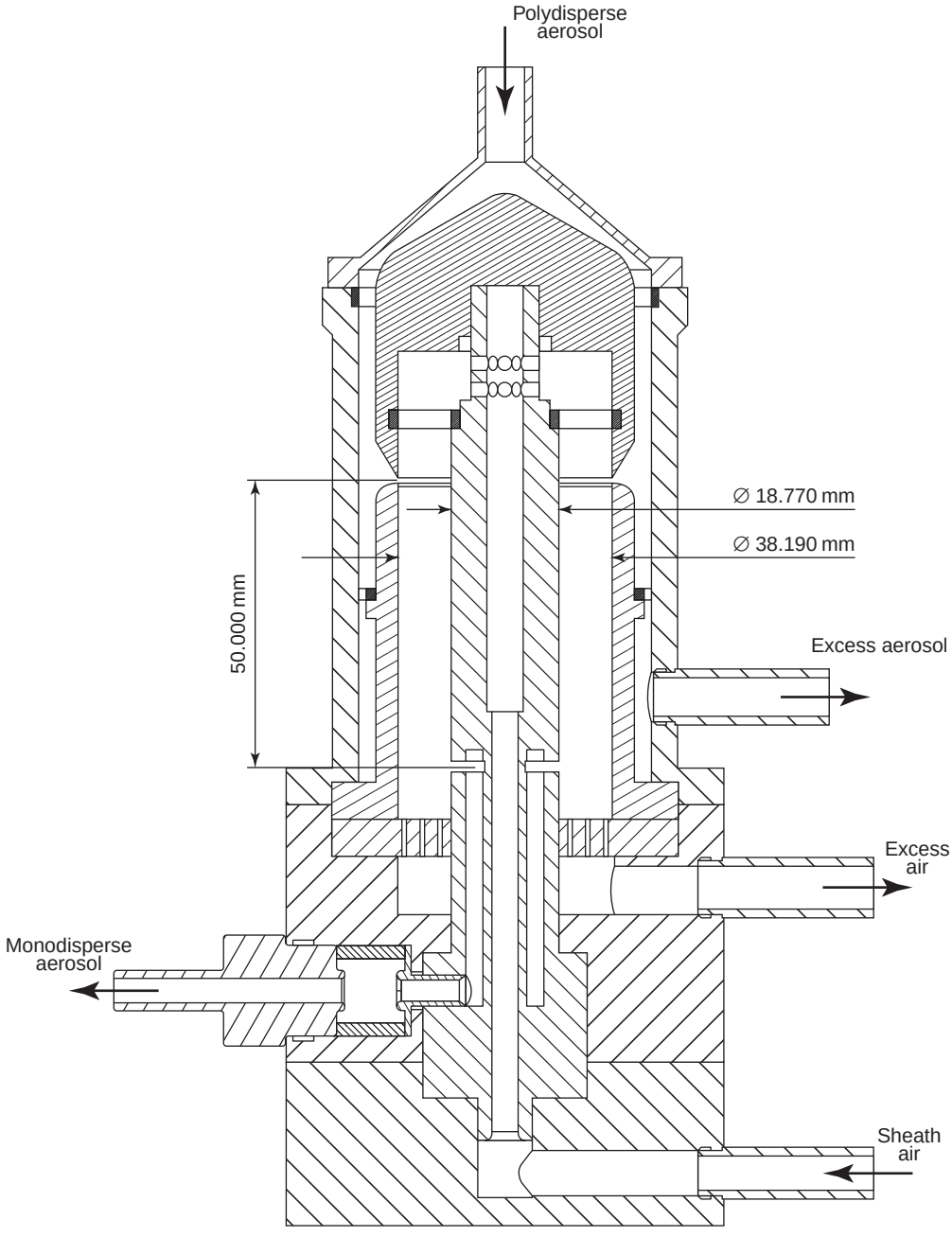

Figure 5 Schematic diagram of the Nano-DMA.[9]

environments. Figure 6 shows the operating principle of the OPC. Single, individual particles are carried by an air stream through an illuminated viewing volume in the instrument and cause light to be scattered to a photodetector. The photodetector generates a voltage pulse in response to each particle passing through the viewing volume. The pulse amplitude is then taken as a measure of particle size. The pulse is then counted and processed electronically to yield a pulse-height histogram, which is then converted to a histogram for particle size distribution using an appropriate calibration curve. Many commercial counters using an incandescent light source have been developed for particle size distribution measurement in the range 0.3 μm to approximately 10 μm. Recent advances include the use of laser illumination to achieve lower detection limits down to 0.05 μm.

OPCs differ widely in their design and performance characteristics. Table 1 gives a selected list of OPCs available commercially. The PMS counter uses the ''active scattering'' principle, in which the particles are passed

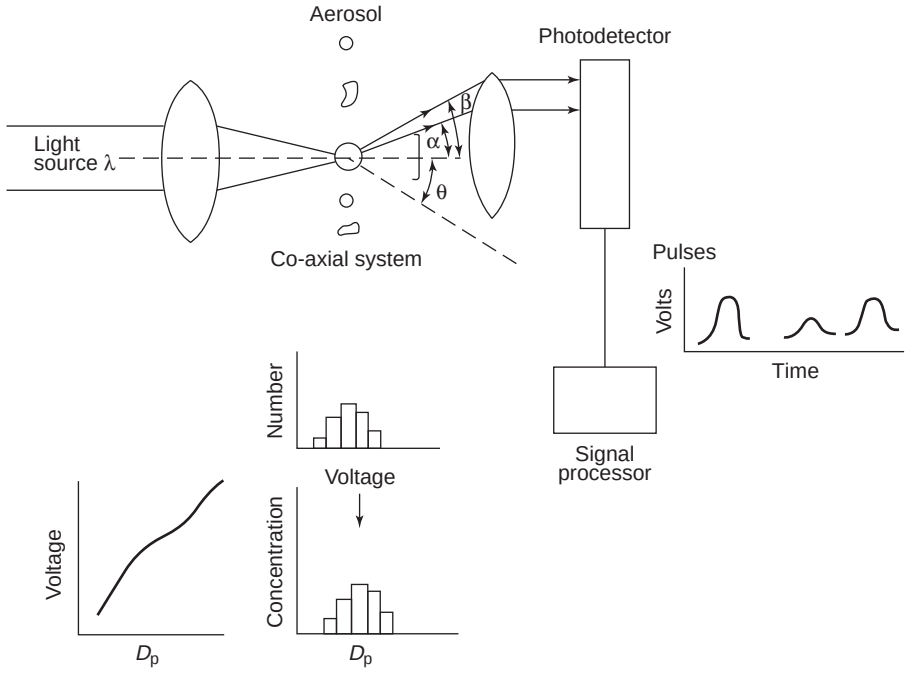

Figure 6 Operating principle of the optical (laser) particle counter.

Table 1 List of selected commercial optical and laser particle counters

Company	Model	Flow rate (cfm)	Size range (μm), (no. of channels)	Optical design	Light source[a]	Weight (kg)
CLIMET	CI-7600	0.1	0.1–5.0 (4)	90° (wide)	mm	14
	CI-500	1.0	0.3–25.0 (6)	15°–150°	laser	10
	CI-4224	1.0	0.3–10.0 (4)	15°–150°	LD	11
	CI-4200	0.1	0.3–5.0 (2)	forw. ang.	LD	3
HIAC/ROYCO	2250	1.0	0.5–5.0 (2)	forw. ang.	LD	2
	5109	1.0	0.09–1.0 (6)	wide ang.	active L	15
MET ONE	237B	0.1	0.3–5.0 (6)	forw. ang.	LD	3
	A2320	1.0	0.3–10.0 (6)	forw. ang.	LD	14
	A2200	1.0	0.14–3.0 (6)	forw. ang.	active L	17
	227B	0.1	>0.3/var.	forw. ang.	LD	1
PALAS	PC2000	–	0.25–25 (256)	2 × 90°	Xe bulb	20
PMS	LAS-X	0.1	0.1–7.5 (15)	wide ang.	active L	19
	LAS-HS	0.1	0.06–1.0 (8)	wide ang.	active L	23
	LAS151	1.0	0.15–5.0 (8)	wide ang.	passive	19
	LAS201	0.1	0.2/0.5 (2)	forw. ang.	LD	2
TOPAS	LAPS	0.1	0.3–20 (128)	forw. ang.	laser	10

[a] mm, multimode; LD, laser diode; L, laser; Xe, xenon.

through the resonant cavity of a helium–neon laser. The Hiac/Royco instrument uses a helium–neon laser and an external scattering volume. In contrast, the TSI instrument uses a solid-state laser diode to obtain a small, light-weight portable sensor. They can be differentiated by the light source (laser, laser diode, or white light) they use, the sampling flow rate of the instrument, the number of channels of data the instrument provides, and other distinguishing characteristics, such as portability and ability to be interfaced with computers.

In general, an instrument using a laser source, particularly of the ''active scattering'' type, can detect smaller particles than a corresponding instrument using an incandescent light source because of the higher

illuminating intensity of the laser. The lower detection limit of the white-light and laser diode counters is usually around 0.3 μm, whereas ''active scattering'' laser counters can detect particles as small as 0.1 μm and below. A higher flow rate instrument can count more particles in a given time period than an instrument of a lower sampling flow rate. However, the particle coincidence level of high flow rate instruments is usually much lower. OPCs of this type are particularly important for particle counting in a low concentration environment, such as in cleanrooms. A lower flow rate instrument has higher resolution and can detect smaller particles than the high flow counter. A sampling flow rate of 1 cubic foot per minute (cfm) is usually considered high and a flow rate of 0.01 cfm is usually considered low.

The OPC response, which gives a functional relationship between the pulse height and the particle size, depends on both the instrument properties and the particle properties. The former includes optical design, illumination source, and electronics gain, and the latter includes particle size and shape, refractive index, and orientation of nonspherical particles with the incident beam. The relative response of the OPC as a function of particle size can be calculated by means of the theory of electromagnetic scattering developed by Mie. The calculation for some white light counters has been reported by Cook and Kerker.[26] More recent studies have concentrated on laser particle counters.[27]

To determine the absolute voltage-size response of the OPCs, as well as other instrument characteristics, such as resolution, count coincidence, response to irregular particles, and inlet efficiency, experimental studies are generally required. Liu et al.[27] reported on the evaluation of several commercially available white-light counters using monodisperse spherical particles. Wen and Kasper[28] and Liu and Szymanski[29] evaluated the counting efficiencies of several commercial OPCs. A novel technique to determine the OPC response to irregular coal dust particles has been developed by Liu et al.[30] Marple and Rubow[31] made use of inertial impactors to obtain aerodynamic particle size calibration of the OPCs. Recent work mostly involves the evaluation of laser OPCs.[32–38] Comprehensive discussions on the principle and application of the OPC may be found in the papers by Willeke and Liu,[39] Knollenberg and Luehr,[40] and by Gebhart et al.[41]

Recent developments in OPCs are to increase the coincidence level while keeping high sampling flow rate, to measure the refractive index, and to determine the particle shape using the scattering intensity from multi-angle light scattering.[42–45] A review of recent developments was presented at the 5th International Congress on Optical Particle Sizing.[46]

3.4 Aerodynamic Particle Sizer and Aerosizer

In addition to direct light scattering measurement described above, light scattering can also be used in combination with other measurement principles to extend the measurement capabilities of the light scattering technique. An example of this is the use of an accelerating nozzle in combination with light scattering measurement.

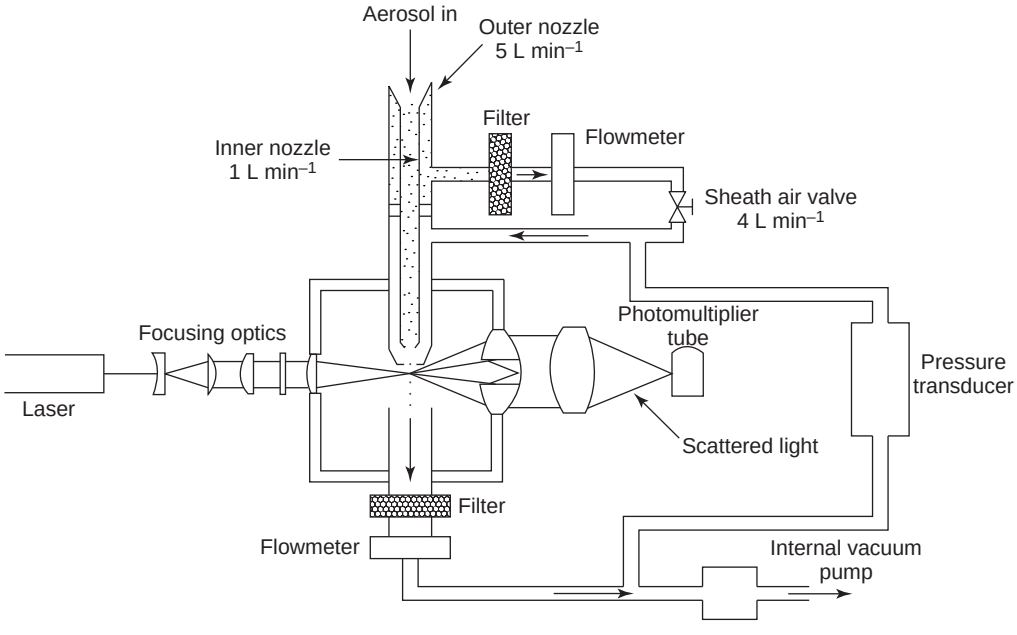

Figure 7 Schematic diagram of the APS described by Agarwal et al.[47]

Figure 7 shows a schematic diagram of a commercially available aerodynamic particle sizer (APS) described by Agarwal et al.[47] In this instrument, the particles are accelerated through a small nozzle to different speeds. The larger the particle size, the lower the speed of the particle due to particle inertia. The particle velocity at the nozzle exit is then measured by detecting the time required to pass through two laser beams with a fixed separation to provide a measure of particle size. This principle enables the ''aerodynamic size'' of the particles, in the size range 0.5 to 30 μm, to be measured, which is related to the settling speed of the aerosol and to particle deposition in the lung. Calibration studies on the APS have been reported by Chen et al.[48] and Baron.[49]

Current developments in APS include an ultraviolet (UV) pulse laser to detect the viability of bio-aerosol, and a high energy laser,[50] such as an Nd : YAG (neodymium : yttrium aluminum garnet) laser, and mass spectrometry for in-situ particle composition measurement.[51]

A second commercial aerodynamic sizing instrument, the Aerosizer, operates under the same time-of-flight principle as the APS. One significant difference is that particles are accelerated at sonic flow through a critical nozzle in the Aerosizer whereas subsonic flow acceleration is used in the APS. The Aerosizer is capable of measuring particles in a wider size range (0.5 to 2000 μm) and higher concentration (up to 1100 particles cm^{-3}) than the APS. However, the calibration curve for particle size is strongly dependent on particle density. Cheng et al.[52] calibrated two Aerosizers using uniform-sized spherical polystyrene latex particles (PSL) and glass beads and with nonspherical natrojarosite particles.

3.5 Electrical Low Pressure Impactor

By combining a conventional particle size classification method, e.g. the cascade impactor, with a direct reading particle detector, a near real-time particle size distribution analyzer can be constructed. Traditionally, the operation of cascade impactors involves depositing particles on a substrate and weighing the substrate before and after each sampling period. Consequently, traditional cascade impactors, such as the micro-orifice uniform deposition impactor (MOUDI),[53] are not considered as direct-reading instruments. By installing a quartz crystal sensor on the collecting stages of the cascade impactor, it is possible to perform nearly real-time size distribution measurement. A commercial instrument, the quartz-crystal microbalance (QCM), has been developed making use of this principle.[54] Two drawbacks of this technique are that the crystals need frequent cleaning and the concentration range is limited in order to avoid overloading the crystal stages.

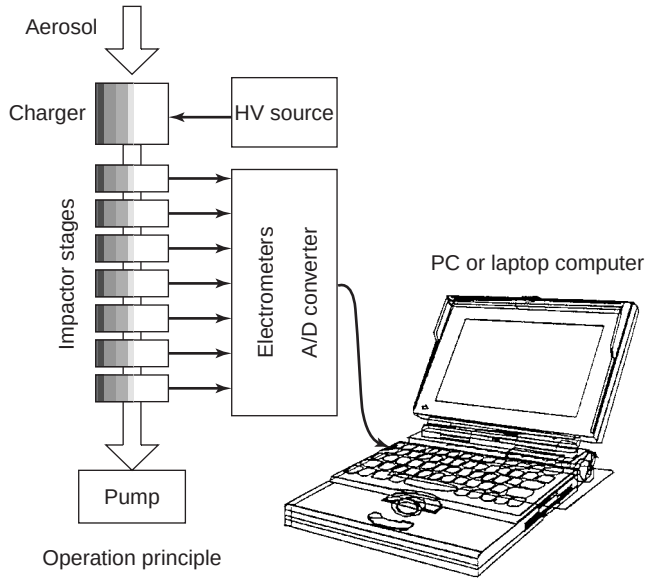

Figure 8 Schematic diagram of the ELPI.[55]

This is partially remedied by the recent development of an electrical low pressure impactor (ELPI), shown schematically in Figure 8.[55] The ELPI make use of a diffusion charger to charge the aerosol particles electrically. The charged particles are then impacted on the collection stages equipped with a current-detecting electrometer. The impactor stages provide the cut-size information and the currents provide the particle concentration information. The parallel current measurements allow a fast response compared with the DMPS and reduced maintenance compared with the QCM cascade impactors.

3.6 Phase Doppler Analyzer

The optical techniques described above involve extracting aerosol from the environment and transporting it to a sensing zone for measurement. In-situ techniques, often optically based, are noninvasive and measure aerosol in its natural state without extractive sampling. The noninvasive measurement is accomplished by locating the sensing volume external to the instrument, thereby eliminating the need for extractive sampling. The techniques are most suitable for measuring aerosols in hostile environments of extreme pressure and temperature ranges, and in reactive or corrosive environments.

Rader and O'Hern[56] summarized the techniques in a review paper. They classified the techniques into the following major categories: single-particle counters that

are intensity-based, phase-based, or imaging; and ensemble techniques of particle field imaging, Fraunhofer diffraction, or dynamic light scattering. To determine the particle size, single-particle counters measure the scattering behavior of an individual particle as it passes through a well-defined sensing zone formed by two crossing laser beams, while the ensemble techniques analyze the collective scattering of a large number of particles. They can be used to measure individual particle sizes from about 0.25 to above 1000 μm, concentrations as high as 10^6 particles cm^{-3}, and speeds in the kilometer per second range. With the ensemble techniques, particle mean diameters as low as 0.01 μm can be measured.

An example of a single-particle counter based on the phase Doppler principle is shown in Figure 9.[56] The principles of a phase Doppler particle analyzer (PDPA) were described in the paper by Bachalo and Houser.[57] The system consists of a laser and transmitting optics, and a receiver optics package with multiple photo-detectors to measure the spatial and temporal frequency of the Doppler-shifted light scattered by individual particles passing through the measuring volume. The spatial frequency gives a measure of the particle diameter, refractive index and receiver geometry, while the temporal frequency gives a measure of the particle velocity. A commercial instrument gives the specifications of either one or two velocity component measurements in addition to particle size, a particle diameter range of 1–8000 μm and a dynamic range of 35 : 1, and a velocity range from 1 to 200 $m\,s^{-1}$. The maximum measurable number concentration is 10^6 cm^{-3} which is based on the number of particles passing through a calculated size-dependent measurement volume. More information on the recent development of a phase Doppler particle-sizing velocimeter can be found in a special issue of *Particle and Particle Systems Characterization*.[58]

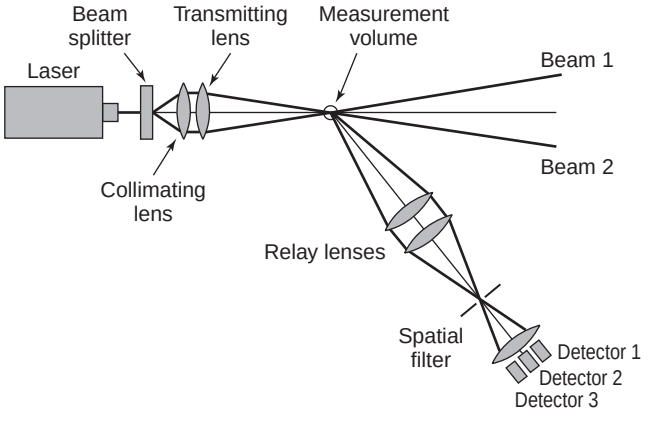

Figure 9 Schematic diagram of the phase Doppler analyzer.[56]

4 TOTAL (INTEGRAL) CONCENTRATION DETECTORS

Integral concentration detectors are those that can be used to measure some integral parameters of an aerosol over its entire size distribution, such as the total number or mass concentration or total light scattering or extinction coefficients. Several of the more widely used instruments are described below.

4.1 Condensation Particle Counter

The CPC, or condensation nucleus counter (CNC), is widely used to measure particles in the diameter range from approximately 0.005 to 1.0 μm. The instrument operates by passing the aerosol stream through a vapor-supersaturated region produced either by the adiabatic expansion or direct contact cooling to cause the vapor condensation on the particles. The particles are then grown to a size where they can be detected optically by light scattering. Recent advances include the development of a continuous flow, direct contact type CPC[59–61] and the mixing of a hot vapor stream and a cool aerosol stream to achieve a supersaturation condition.[62]

A schematic diagram of a commercially available, continuous flow CPC is shown in Figure 10.[63] In this instrument, butyl alcohol is used as the working fluid. An air stream is saturated with this vapor in a saturator kept at 35 °C. The subsequent cooling of this alcohol-vapor laden air stream in a thermoelectrically cooled condenser tube kept at 10 °C produces the required supersaturation for the vapor condensation on the particles. Particles emerging from the condenser tube at a size of approximately 12 μm are then detected optically by light scattering. For low particle concentrations, individual particles are

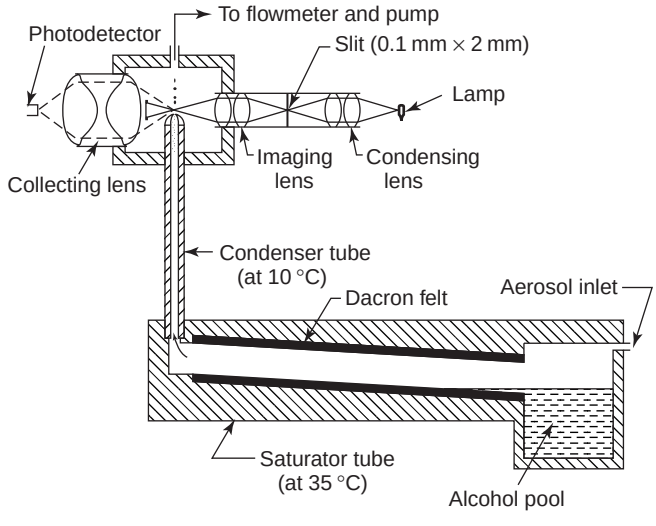

Figure 10 Schematic diagram of the CPC.[63]

counted. Above a particle concentration of 1000 particles cm^{-3}, the total light scattering from the droplet cloud is detected in a ''photometric mode'' to measure the total particle concentration. The concentration range of the instrument is from less than 0.01 particles cm^{-3} to more than 10^6 particles cm^{-3}.

Detailed calibration studies of the CPC have shown that below a particle size of 0.005 μm the response of the instrument begins to drop off as a function of particle size.[63–71] The counting efficiency decrease can be attributed to particle loss in the flow passages in the instrument due to diffusion and the lack of 100% activation due to inhomogeneous vapor concentration distribution in the condenser. By introducing clean sheath air around the aerosol stream in the CPC, Stolzenburg and McMurry[67] were able to increase the counting efficiency of the instrument to over 70% at a particle size of 0.003 μm. Figure 11 shows the details of such a sheath air CPC UCPC. Wilson et al.[72] using a similar design have developed a low pressure CPC for stratospheric aerosol measurements. Niessner et al.[73] have shown that by changing the supersaturation ratios in steps, the dependence of particle size on critical supersaturation for vapor condensation can be used for size distribution measurement.

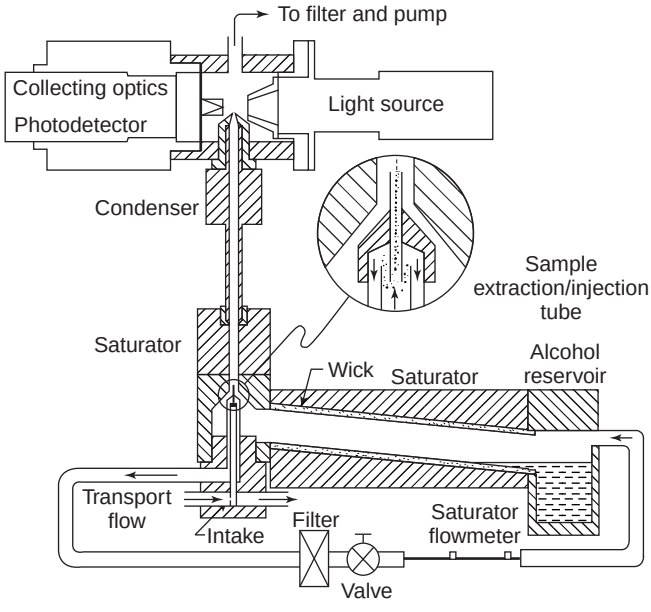

Figure 11 Schematic diagram of the UCPC.[67]

4.2 Light-scattering Photometers and Nephelometers

For atmospheric studies, the total light scattering coefficient of the airborne particles is important as it is related to atmospheric visibility or visual range. Measurement of the total light scattering coefficient is usually made with a photometer or integrating nephelometer. For aerosols that differ only in concentration and with the same size distribution, the integral light scattering measurement can be converted to mass concentration. Examples of such a correlation between total scattering and atmospheric mass concentration are given by Waggoner and Charlson.[74]

In the integrating nephelometer, shown schematically in Figure 12, the particles are illuminated in a sensing volume of approximately 1.0 L and scattered light from the particles reaches the photoreceptor at angles from 8° to 170° off axis. This simplifies the complex angular scattering relationship by summing the scattering over nearly the entire range of angles. Although the instrument was originally used to measure visual range, it has found its application in the study of the urban and rural atmospheric aerosol. In some cases, the scattering has been shown to be well correlated with the atmospheric mass concentration.[74,75] The instrument is simple in construction and has been used in automobiles and aircraft for mapping the concentration of particles in the 0.1 to 1.0 μm range. These particles are chiefly responsible for degraded urban visibility. Some caution must be exercised when using the nephelometer in an environment with sooty particles since the scattering will

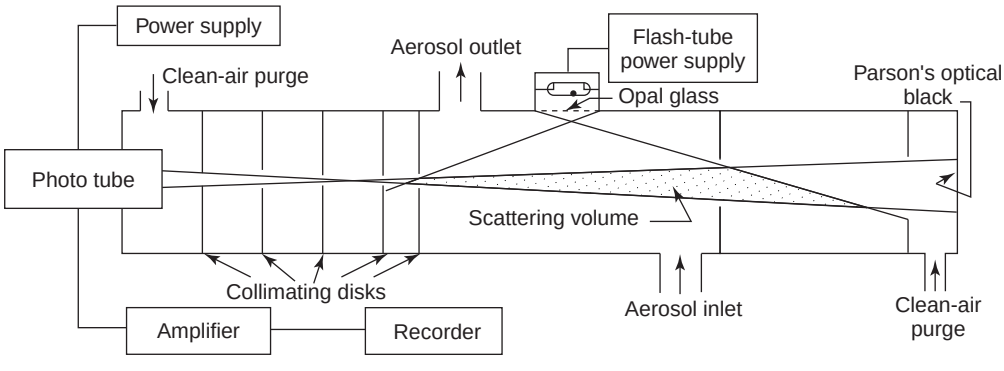

Figure 12 Schematic diagram of an integrating nephelometer.

be attenuated because of light absorption. In this case the apparent concentration will be lower than expected.

Forward-scattering photometers, which employ a laser or incandescent light source and optics similar to dark field microscopy, have been commercially produced. A narrow cone of light converges on the aerosol cloud, but it is prevented from falling directly on the photoreceptor by a dark stop; only light scattered in the near forward direction falls on the receptor. The readout of these instruments is in mass or number concentration, but the calibration may change with composition and size distribution of the particles. Based on the solutions to Maxwell's equations, forward scattering photometers are, however, less sensitive to the change in the refractive index of particles than are photometers at other commonly used sensing angles such as 30°, 45°, or 90°.

A forward-scattering photometer (45°–95°) was developed as a passive personal monitor for airborne particles.[76] This instrument displays the real-time particle mass concentration for time intervals as small as 10 s and calculates time-weighted averages (TWAs) for up to a full shift in a workday for display or readout.

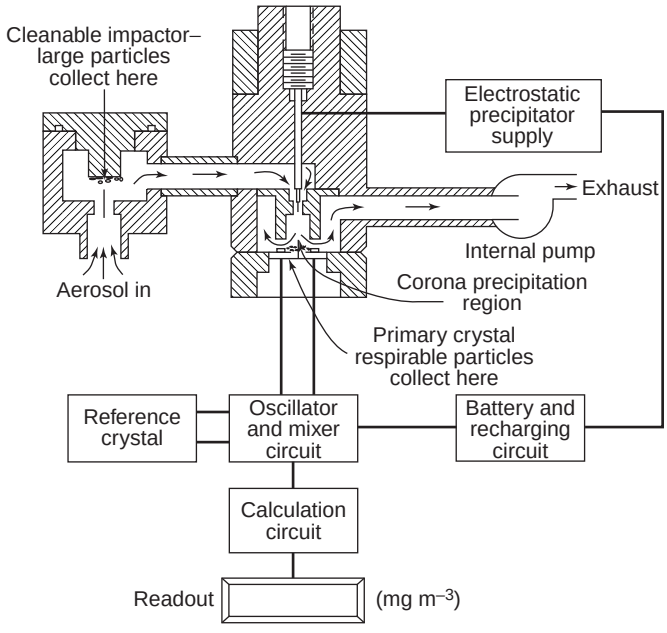

Figure 13 Schematic diagram of a portable quartz-crystal piezoelectric microbalance.[78]

4.3 Quartz-crystal Microbalance and Surface Acoustic Wave Microbalance

Several sensors for near real time mass concentration measurements have been developed. By depositing the particles on a quartz crystal the natural vibrating frequency of the crystal can be affected and used as a measure of the deposited particle mass. The deposited particle mass is proportional to the frequency shift. The particle deposition can be achieved either by electrostatic precipitation[77,78] or by inertial impaction.[54] The sensitivity of the QCM is approximately 10^9 Hz g^{-1}, which corresponds to a frequency shift of 1 Hz for a 10 MHz AT-cut quartz crystal. Figure 13 shows a schematic diagram of a battery operable piezoelectric microbalance for respirable aerosol detection.[79] The instrument incorporates a respirable impactor at the inlet to remove the nonrespirable particles, allowing the respirable particles to be deposited by electrostatic precipitation on the quartz crystal for measurement. The instrument can measure particle concentrations in the range of approximately 0.05–5.5 mg mL^{-1}.

A new development in QCMs enables the mass sensitivity to be significantly increased relative to the ones using AT-cut crystals. By modifying the quartz crystal and applying an electric field between two electrodes on the same surface (instead of through the crystal thickness) several micrometers apart, the mode of excitation results in a natural frequency of up to 300 MHz.[80] The vibration on the same surface is called the surface acoustic wave (SAW) mode. The mass sensitivity of the SAW devices can be up to 10^{10}–10^{11} Hz g^{-1}.

4.4 Tapered-element Oscillating Microbalance

Recent advances in the vibrating mass sensing technique include the use of a low frequency vibrating mass in the form of a hollow tapered element[81] coupled to a filter collector or impactor. The operation principle of the tapered-element oscillating microbalance (TEOM) is the same as for the QCM. Instead of relying on the natural frequency of quartz crystal, the vibration of the hollow tapered element is initiated and maintained by an electronic feedback system. The oscillation of the tapered element is then monitored by a light-emitting diode and phototransistor aligned perpendicularly to the oscillation plane of the tapered element. Figure 14 shows a typical arrangement for a TEOM. The aerosol stream is drawn from the ambient air and passed through a filter installed on the top of the tapered element. The collected particle mass is then inferred from the frequency difference before and after each sampling. Unlike QCMs, the collected particle mass is not directly proportional to the frequency shift. It is proportional to the difference of the inverse square of frequencies before and after sampling. This design extends the measurement range of the technique to mass concentration levels in the gram per cubic meter range. The filter can usually hold particle masses up to 2–6 mg. The application of the technique to particle measurement in high-temperature and high-pressure gas

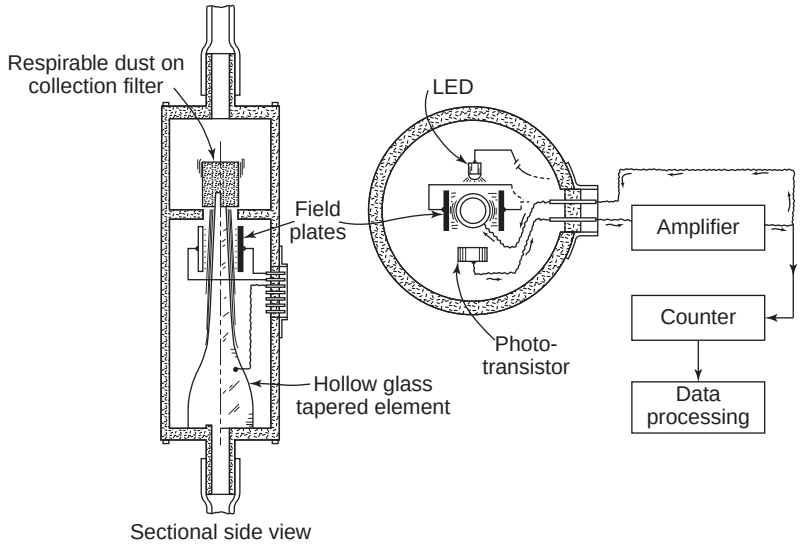

Figure 14 Schematic diagram of the TEOM.

streams of the pressurized fluidized bed combustor has been reported by Wang.[82]

5 MONODISPERSE AEROSOL STANDARDS

In order to correctly interpret the experimental data and to obtain the highest possible accuracy from a given instrument, a detailed calibration of the instrument with a monodisperse aerosol is required. In some cases, the availability of these aerosol standards has made it possible to develop new aerosol sensors for which the underlying theory is not well understood. Figure 2 shows some of the available monodisperse aerosol standards for calibrating the direct reading instruments.

5.1 Electrospray Aerosol Generator

Figure 15 shows a schematic diagram of the electrospray aerosol generator.[83] The spraying chamber is in the point-to-plate configuration with the capillary tube facing the plate. An orifice is located on the center of the plate, allowing the produced particles to enter the neutralization chamber. A coaxial tube allows CO_2 to flow as a sheath surrounding the capillary tube for suppressing possible corona discharge. The compressed air is supplied from the top of the chamber in order to transport the particles through the orifice. The liquid is fed from a syringe driven by a syringe pump. The flow rate is controlled by a programmable syringe pump. Once the droplets are produced, they are transported into the neutralization chamber quickly for neutralizing the highly charged droplets produced by electrospraying.

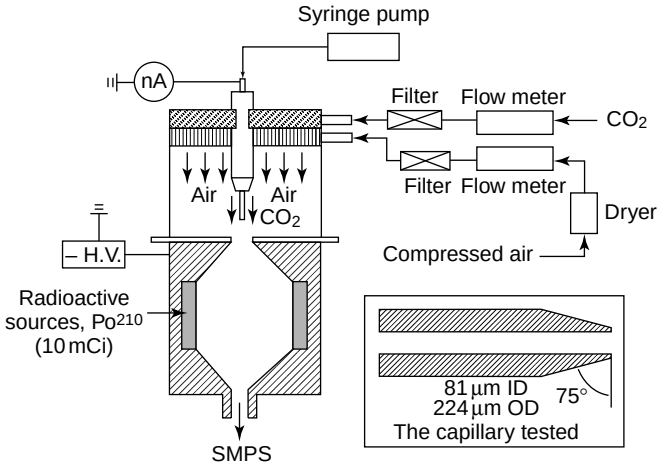

Figure 15 Schematic diagram of an electrospray aerosol generator.[83]

A negative high voltage is applied to the plate and the neutralization chamber. The capillary tube is connected to an electrometer which is used to measure the spraying current. The size of the produced liquid droplets is further reduced by the evaporation process. Using this technique, monodisperse aerosol particles in the range 3 nm up to 1 μm have been successfully produced.

The electrospray can also be used to aerosolize colloidal particle suspensions. For example, PSL and colloidal gold particles have been successfully produced using the electrospray. The droplet sizes produced by the electrospray are small compared to those produced by mechanical atomization techniques. It therefore produces much smaller residue particles in spraying the colloidal suspensions.

5.2 Vibrating Orifice Aerosol Generator

The vibrating orifice aerosol generator (VOAG), shown schematically in Figure 16, was developed by Berglund and Liu.[84] The generator is based on the instability and uniform break-up of a liquid jet under a uniform, periodic mechanical disturbance. The generator is composed of a droplet generation and dispersion system and an aerosol dilution, neutralization, and transport system.

To generate an aerosol with this system, a solution of a known solute concentration is supplied to the droplet generator by a syringe pump. The liquid flows through a small (approximately 10 μm) orifice in the droplet generator and is broken up into uniform droplets by the mechanical vibration produced by a piezoelectric ceramic. The uniform droplet stream is then quickly dispersed by a turbulent air jet to prevent droplet collision. The dispersed droplets are then mixed with a filtered dilution air to evaporate the solvent from the solution droplets. The aerosol is subsequently passed through a radioactive ^{85}Kr neutralizer to discharge the electrostatic charges on the particles prior to using it as a calibration aerosol.

The diameter of the aerosol particles produced by the vibrating orifice generator can be calculated from the equation for the conservation of mass. The overall uncertainty of the calculated particle size is about 1%.

5.3 Electrostatic Classifier Aerosol Generator

For producing submicron monodisperse aerosol standards in the size range from below 0.005 μm to 1.0 μm, an electrostatic classifier aerosol generator has been developed by Liu and Pui.[85] The method is based on the monotonic relationship between the electrical mobility of a singly charged particle and particle size. By passing aerosol with predominantly singly charged particles through the mobility classifier, a monodisperse fraction can be extracted according to electrical mobility. The generator, shown schematically in Figure 17, consists of a polydisperse aerosol generator, a ^{85}Kr charge neutralizer, a DMA, and an electrometer current sensor.

To generate a monodisperse aerosol with this system, a polydisperse feed aerosol is first produced either by atomization or by a vaporization-condensation technique. The aerosol is then passed through the ^{85}Kr charge neutralizer to obtain a Boltzmann equilibrium charge on the particles.[85] For submicron particles, an aerosol in Boltzmann charge equilibrium will contain both positively and negatively charged particles of a low level charge. This aerosol is then classified electrostatically by the DMA to obtain a monodisperse, singly charged aerosol.

The electrical mobility of the extracted monodisperse aerosol can be calculated from the operating condition of the DMA.[85] Knowing the aerosol mobility, the particle diameter can then be calculated. An accuracy of 2% in the calculated particle diameter can be achieved from carefully measuring the aerosol and clean air flow rates in

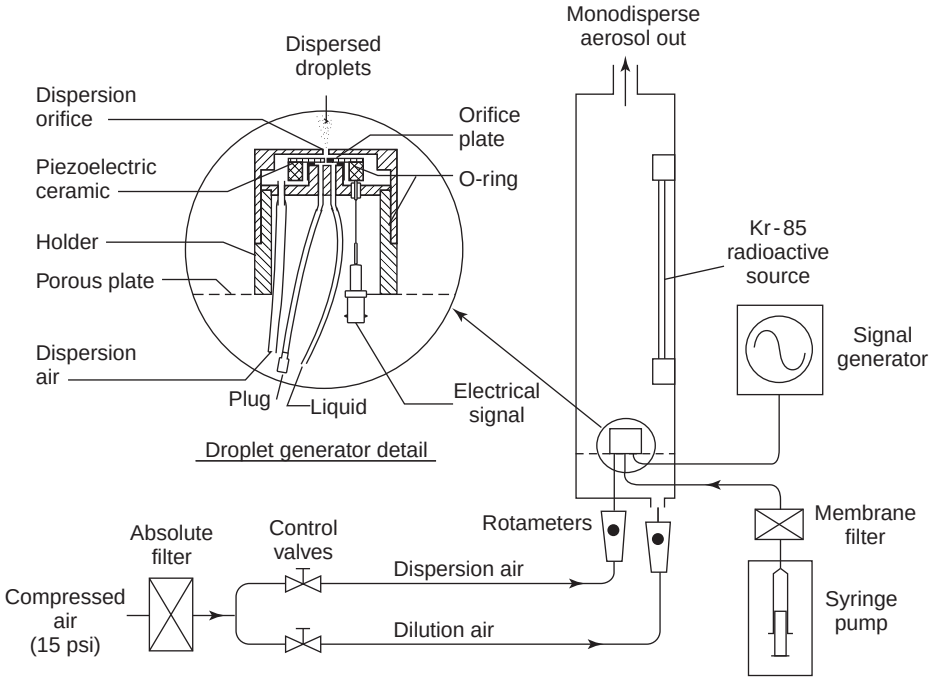

Figure 16 Schematic diagram of the VOAG.[84]

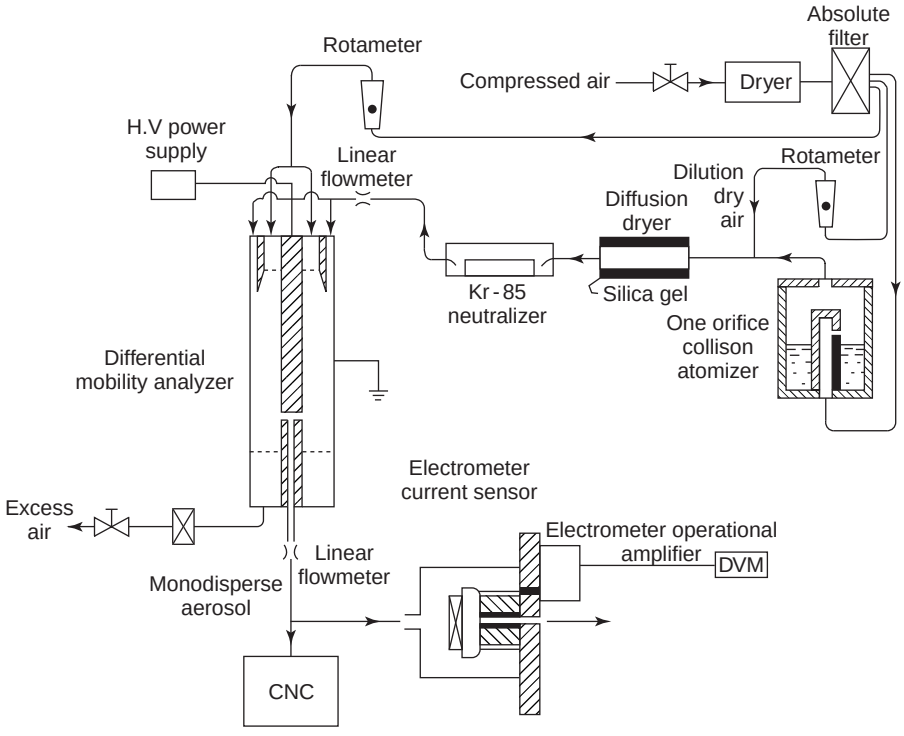

Figure 17 Schematic diagram of the electrostatic classifier aerosol generator.[85]

the DMA and the applied voltage on the central electrode in the DMA.

By sampling these singly charged particles into a Faraday cup (the electrometer current sensor) and measuring the corresponding current flow, the particle concentration can be calculated.

The electrostatic classification technique was used as a secondary transfer standard for the certification of the 0.1 μm standard reference material (SRM) by the National Institute of Standards and Technology (NIST). The method involved ''calibrating'' the DMA by two established primary calibration standards 0.269 μm PSL (NIST SRM 1691) and 0.895 μm PSL (NIST SRM 1690). Once the flow rate and the voltage of the DMA were adjusted according to the two primary standards, it was then used to measure the 0.1 μm PSL for establishing the 0.1 μm SRM.

6 SUMMARY

Considerable advances have been made in recent years in the development of direct-reading instruments for analyzing airborne particles. Instruments are now available to measure aerosol number concentration up to 10^6 particles cm^{-3}, mass concentration up to 1000 mg m^{-3}, and size distribution over a particle size range of 0.003 μm to over 100 μm. Advances in instrumentation are such that many of the measurement problems that were considered too difficult only 10 years ago can now be performed routinely with good experimental accuracy.

With a wide array of available commercial instruments, it is necessary for the practitioners to understand well the operating principles of the instruments and the aerosol system under study. Some of the criteria for selecting an appropriate instrument include the particle size range of interest; the system parameters to be studied, e.g. mass or number concentration versus particle size distribution, aerodynamic property versus light scattering property, etc.; and the cost and compactness of the instrument. Several useful references can be used to help select the appropriate instruments for the specific applications. Pui and Swift[86] provided a comprehensive review of the available instruments, and the specifications and commercial sources for these instruments. A detailed state-of-the-art review of instruments of different principles is contained in the book edited by Willeke and Baron.[87]

Figure 18 shows a flow chart for selecting a direct-reading instrument for analyzing airborne particles. The first step is to decide whether to perform an integral concentration measurement or a complete size distribution measurement. In principle, the integral concentration, i.e. mass concentration or number concentration, can be obtained by integrating

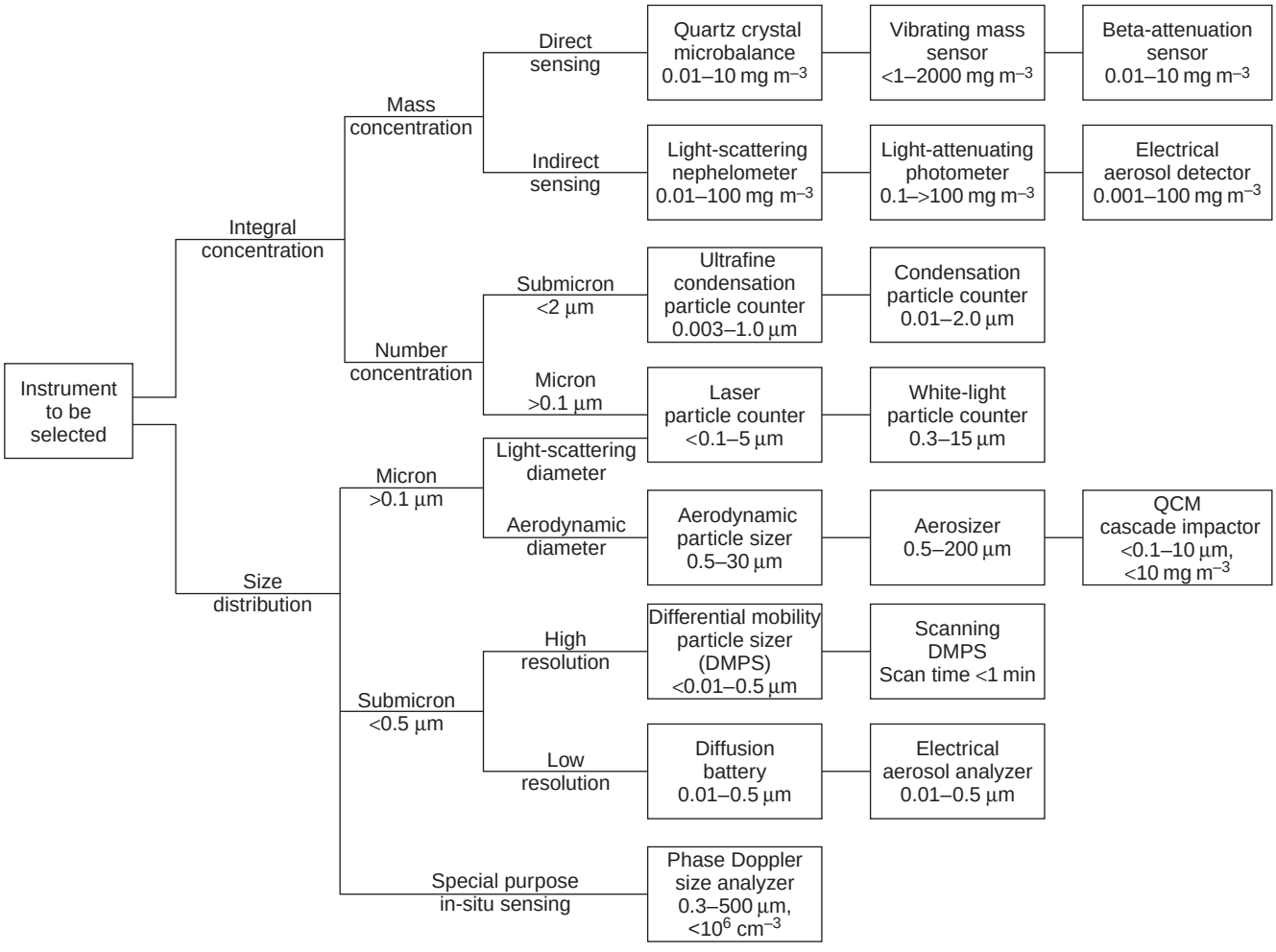

Figure 18 Flow chart for selecting a direct-reading instrument for analyzing airborne particles.[86]

the size distribution over the appropriate size range. However, there are many compact, inexpensive instruments available for integral concentration measurement compared to the more sophisticated and expensive instruments for size distribution measurement. Considerable cost saving can be realized if an integral concentration detector is deemed appropriate for the application.

For mass concentration measurement, as required in many industrial hygiene applications, the particles may be directly captured on a surface for measurement by the quartz-crystal sensing technique or by the beta-attenuation sensing technique. These techniques generally require longer sampling time and frequent cleaning of the surface. A quicker method would be to measure indirectly the light scattering intensity of the aerosol and to infer its mass concentration through calibration. The accuracy of the technique depends on the measured aerosols having nearly the same size distribution and differing only in concentration. For number concentration measurement, as required in many cleanroom applications, two classes of instruments may be used depending on the size range of interest. For particle diameter less than 2 μm, several CPCs may be used. For particle diameter larger than 0.1 μm, a large array of white-light or laser OPCs may be used.

For size distribution measurement, the instruments are divided into two major classes depending on their measuring size ranges. For particles larger than 0.1 μm, the OPCs or the particle relaxation size analyzers may be used. The latter are capable of measuring the aerodynamic particle size, which is important for deposition studies in the respiratory system. For submicron aerosol, the high-resolution DMPS or the low-resolution diffusion battery and the electrical aerosol analyzer may be used. A number of special purpose instruments are also available for difficult applications. For example, the PDPA measures aerosol in its natural state without extractive sampling. The noninvasive measurement is

accomplished by locating the sensing volume external to the instrument, thereby eliminating the need for extractive sampling. The technique is most suitable for measuring aerosols in hostile environments of extreme pressure and temperature ranges, and in reactive or corrosive environments. The cost of such a system, however, will be many times that of an integral concentration detector.

Instruments are also developed to meet specific applications. Examples include the development of a portable CPC for respirator fit testing, the adaptation of APS for bioaerosol/particle component measurement and the development of high flow rate laser particle counters for cleanroom applications. Computer direct-reading instruments have been developed with modern electronics and transducer technologies. Significant efforts have also been devoted to improve the software for on-line data reduction. These hardware and software developments will make the next generation of aerosol instruments more sensitive, accurate, compact and user friendly.

ABBREVIATIONS AND ACRONYMS

APS	Aerodynamic Particle Sizer
CNC	Condensation Nucleus Counter
CPC	Condensation Particle Counter
DB/CPC	Diffusion Battery/Condensation Particle Counter
DMA	Differential Mobility Analyzer
DMPS	Differential Mobility Particle Sizer
ELPI	Electrical Low Pressure Impactor
MOUDI	Micro-orifice Uniform Deposition Impactor
Nano-DMA	Nanometer Aerosol Differential Mobility Analyzer
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
NIST	National Institute of Standards and Technology
OPC	Optical Particle Counter
PDPA	Phase Doppler Particle Analyzer
PSL	Polystyrene Latex Particles
QCM	Quartz-crystal Microbalance
SAW	Surface Acoustic Wave
SMEC	Spectromete de Mobilité Electrique Circulaire
SMPS	Scanning Mobility Particle Sizer
SRM	Standard Reference Material
TEOM	Tapered-element Oscillating Microbalance
TWAs	Time-weighted Averages
UCPC	Ultrafine Condensation Particle Counter
UV	Ultraviolet
VOAG	Vibrating Orifice Aerosol Generator

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Airborne Instrumentation for Aerosol Measurements

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air

REFERENCES

1. H.K. Bowen, *Mater. Sci. Engr.*, **44**, 1–56 (1980).
2. M. Schober, *Ceram. Bull.*, **66**, 1698–1699 (1987).
3. R.W. Siegel, S. Ramasamy, H. Hahn, L. Zongquan, L. Ting, *J. Mater. Res.*, **3**, 1367–1373 (1988).
4. B.R. Appel, S.M. Wall, Y. Takiwa, M. Haik, *Atmos. Environ.*, **13**, 543–558 (1979).
5. K.T. Whitby, in *Sulfur in the Atmosphere*, eds. R.B. Huscr et al., Pergamon Press, New York, 1978.
6. E.O. Knutson, K.T. Whitby, 'Aerosol Classification by Electrical Mobility: Apparatus, Theory and Applications', *J. Aerosol Sci.*, **6**, 443 (1975).
7. D.-R. Chen, D.Y.H. Pui, 'Numerical Modeling of the Performance of Differential Mobility Analyzers for Nanometer Aerosol Measurements', *J. Aerosol Sci.*, **28**(6), 985–1004 (1997).
8. D.-R. Chen, D.Y.H. Pui, G.W. Mulholland, M. Fernandez, 'A High Efficiency, High Throughput Unipolar Aerosol Charger for Nanoparticles', *J. Nanoparticle Res.*, **1**, 115–126 (1999).
9. D.-R. Chen, D.Y.H. Pui, D. Hummes, H. Fissan, F.R. Quant, G.J. Sem, 'Design and Evaluation of a Nanometer Aerosol Differential Mobility Analyzer (Nano-DMA)', *J. Aerosol Sci.*, **29**(5/6), 497–510 (1998).
10. B.Y.H. Liu, D.Y.H. Pui, 'On the Performance of the Electrical Aerosol Analyzer', *J. Aerosol Sci.*, **6**, 249–264 (1975).
11. W. Winklmayr, G.P. Reischl, A.O. Lindner, A. Berner, 'A New Electromobility Analyzer for the Measurement of Aerosol Size Distributions in the Size Range from 1 to 1000 nm', *J. Aerosol Sci.*, **22**, 289 (1991).
12. M. Pourprix, Sélecteur de particules chargées, a haute sensibilité, Brevet francais No. 94 06273, 24 Mai, 1994.
13. S.-H. Zhang, Y. Akutsu, L.M. Russell, R.C. Flagan, J.H. Seinfeld, 'Radial Differential Mobility Analyzer', *Aerosol Sci. Technol.*, **23**, 357–372 (1995).
14. H. Fissan, D. Hummes, F. Stratman, P. Buscher, S. Neumann, D.Y.H. Pui, D. Chen, 'Experimental Comparison of Four Differential Mobility Analyzers for Nanometer

Aerosol Measurements', *Aerosol Sci. Technol.*, **24**, 1–13 (1996).
15. S.C. Wang, R.C. Flagan, 'Scanning Electrical Mobility Spectrometer', *Aerosol Sci. Technol.*, **13**, 230 (1990).
16. B.Y.H. Liu, D.Y.H. Pui, 'A Submicron Aerosol Standard and the Primary, Absolute Calibration of the Condensation Nucleus Counter', *J. Colloid Interface Sci.*, **47**, 155 (1974).
17. D.-R. Chen, D.Y.H. Pui, 'Design and Testing of an Aerosol/Sheath Inlet for High Resolution Measurements with a DMA', *J. Aerosol. Sci.*, **30**, 983–999 (1999).
18. L.W. Pollak, A.L. Metnieks, *Geofis. Pura Applicat.*, **37**, 183 (1957).
19. D. Sinclair, 'Measurement of Nanometer Aerosols', *Aerosol Sci. Technol.*, **5**, 187 (1986).
20. H.C. Yeh, Y.S. Cheng, M.M. Orman, 'Evaluation of Various Types of Wire Screens as Diffusion Battery Cells', *J. Colloid Interface Sci.*, **86**, 12 (1982).
21. H.G. Scheibel, J. Porstendorfer, 'Penetration Measurements for Tube and Screen-type Diffusion Batteries in the Ultrafine Particle Size Range', *J. Aerosol Sci.*, **15**, 673 (1984).
22. Y.S. Cheng, J.A. Keating, G.M. Kanapilly, 'Theory and Calibration of a Screen-type Diffusion Battery', *J. Aerosol Sci.*, **11**, 549 (1980).
23. D.A. Lundgren, C.N. Rangaraj, ESL-TR-81-04, Defence Technical Information Center, Virginia, 1981.
24. A.C. George, *Health Phys.*, **23**, 390 (1972).
25. D. Sinclair, A.C. George, E.O. Knutson, in *Airborne Radioactivity*, American Nuclear Society, La Grange Park, IL, 103, 1978.
26. D.D. Cooke, M. Kerker, 'Response Calculations for Light Scattering Aerosol Particle Counters', *Appl. Opt.*, **14**, 734 (1975).
27. B.Y.H. Liu, W.W. Szymanski, D.Y.H. Pui, 'Response of Laser Optical Particle Counter to Transparent and Light Absorbing Particles', *ASHRAE Trans.*, **92** (1986).
28. H.Y. Wen, G. Kasper, 'Counting Efficiencies of Six Commercial Particle Counters', *J. Aerosol Sci.*, **17**, 947–961 (1986).
29. B.Y.H. Liu, W.W. Szymanski, 'Counting Efficiency, Lower Detection Limit and Noise Levels of Optical Particle Counters', in *Proceedings of the 33rd Annual Technical Meeting*, Institute of Environmental Sciences, CA, 417–421, 1987.
30. B.Y.H. Liu, V.A. Marple, K.T. Whitby, N.J. Barsic, 'Size Distribution Measurement of Airborne Coal Dust by Optical Particle Counters', *Am. Ind. Hyg. Assoc. J.*, **8**, 443 (1974).
31. V.A. Marple, K.L. Rubow, 'Aerodynamic Particle Size Calibration of Optical Particle Counters', *J. Aerosol Sci.*, **7**, 425 (1976).
32. W.C. Hinds, G. Kraske, 'Performance of PMS Model LAS-X Optical Particle Counter', *J. Aerosol Sci.*, **17**, 67 (1986).
33. Y.J. Kim, J.F. Boatman, 'Size Calibration Corrections for the Active Scattering Aerosol Spectrometer Probe (ASASP-100X)', *Aerosol Sci. Technol.*, **12**, 665 (1990).
34. A. van der Meulen, B.G. van Elzakker, 'Size Resolution of Laser Optical Particle Counters', *Aerosol Sci. Technol.*, **5**, 313 (1986).
35. Y. Yamada, K. Miyamoto, A. Koizumi, 'Size Measurements of Latex Particles by Laser Aerosol Spectrometer', *Aerosol Sci. Technol.*, **5**, 377 (1986).
36. D.M. Garvey, R.G. Pinnick, 'Response Characteristics of the Particle Measuring Systems Active Scattering Aerosol Spectrometer Probe (ASASP-X)', *Aerosol Sci. Technol.*, **2**, 477–488 (1983).
37. B.T. Chen, Y.S. Cheng, H.C. Yeh, 'Experimental Responses of Two Optical Particle Counters', *J. Aerosol Sci.*, **15**, 457 (1984).
38. W.W. Szymanski, B.Y.H. Liu, 'On the Sizing Accuracy of Laser Optical Particle Counters', *Part. Charact.*, **3**, 1–7 (1986).
39. K. Willeke, B.Y.H. Liu, 'Single Particle Optical Counter: Principle and Application', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 698–730, 1976.
40. R.C. Knollenberg, R. Luehr, 'Open Cavity Laser 'Active' Scattering Particle Spectrometry from 0.05 to 5 microns', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 793–816, 1976.
41. J. Gebhart, J. Heyder, C. Rolh, W. Stahlhofen, 'Optical Aerosol Size Spectrometry Below and Above the Wavelength of Light – A Comparison', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 793–816, 1976.
42. B. Sachweh, H. Umhauer, F. Ebert, H. Buttner, R. Friehmelt, 'In Situ Optical Particle Counter with Improved Coincidence Error Correction for Number Concentrations up to 10^7 Particle cm^{-3}', *J. Aerosol Sci.*, **29**(9), 1075–1086 (1998).
43. W.D. Dick, P.H. McMurry, J.R. Bottiger, 'Size- and Composition-dependent Response of DAWN-A Multiangle Single-particle Optical Detector', *Aerosol Sci. Technol.*, **20**, 345–362 (1994).
44. B.A. Sachweh, W.D. Dick, P.H. McMurry, 'Distinguishing between Spherical and Nonspherical Particles by Measuring the Variability in Azimuthal Light Scattering', *Aerosol Science Technol.*, **23**, 373–391 (1995).
45. W.W. Szymanski, C. Schindler, 'Response Characteristics of A New Dual Optics Particle Spectrometer', in *Proceedings of 5th International Congress on Optical Particle Sizing*, Minneapolis, MN, 217–220, 1998.
46. P.H. McMurry, A.A. Naqwi, in *Proceedings of 5th International Congress on Optical Particle Sizing*, Minneapolis, MN, 1998.
47. J.K. Agarwal, R.J. Remiaz, F.J. Quant, G.J. Sem, 'Real-time Aerodynamic Particle Size Analyzer', *J. Aerosol Sci.*, **13**, 222 (1982).
48. B.T. Chen, Y.S. Cheng, H.C. Yeh, 'Performance of a TSI Aerodynamic Particle Sizer', *Aerosol Sci. Technol.*, **4**, 89 (1985).

49. P.A. Baron, 'Calibration and Use of the Aerodynamic Particle Sizer (APS-3300)', *Aerosol Sci. Technol.*, **5**, 55 (1986).
50. P.P. Hairstone, Jim Ho, F.R. Quant, 'Design of an Instrument for Real-time Detection of Bioaerosols Using Simultaneous Measurement of Particle Aerodynamic Size and Intrinsic Fluorescence', *J. Aerosol Sci.*, **28**(3), 471–482 (1997).
51. A.C. Noble, K. Prather, 'A Real-time Measurement of Correlated Size and Composition Profiles of Individual Atmospheric Aerosol Particles', *Environ. Sci. Technol.*, **30**(9), 2667–2680 (1996).
52. Y.S. Cheng, E.B. Barr, I.A. Marshall, J.P. Mitchell, 'Calibration and Performance of an API Aerosizer', *J. Aerosol Sci.*, **24**, 501 (1993).
53. V.A. Marple, K.L. Rubow, G. Ananth, H. Fissan, 'Microorifice Uniform Deposit Impactor', *J. Aerosol Sci.*, **17**, 489 (1986).
54. R.L. Chuan, 'Rapid Measurement of Particulate Size Distribution in the Atmosphere', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 763–776, 1976.
55. M. Marjamaki, J. Keskinen, O.-R. Chen, D.Y.H. Pui, 'Performance Evaluation of Electrostatic Low Pressure Impactor', *J. Aerosol. Sci.*, **31**, 249–261 (2000).
56. D.J. Rader, T.J. O'Hern, 'Optical Direct-reading Techniques: In Situ Sensing', in *Aerosol Measurement – Principles, Techniques and Applications*, eds. K. Willeke, P.A. Baron, Van Nostrand Reinhold, New York, 345–376, 1993.
57. W.D. Bachalo, M.J. Houser, 'Phase Doppler Spray Analyzer for Simultaneous Measurements of Drop Size and Velocity Distribution', *Opt. Eng.*, **23**, 583–590 (1984).
58. K. Leschonski, B. Scarlett, F. Durst, G. Jimbo (eds.) *Particle and Particle Systems Characterization*, **13**(2), 57–176 (1996).
59. J. Bricard, P. Delattre, G. Madelaine, M. Pourprix, 'Detection of Ultra-fine Particles by Means of a Continuous Flux Condensation Nuclei Counter', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 560–580, 1976.
60. D. Sinclair, G.S. Hoopes, 'A Continuous Flow Condensation Nucleus Counter', *J. Aerosol Sci.*, **6**, 1–13 (1975).
61. D. Sinclair, P.C. Yue, 'Continuous Flow Condensation Nucleus Counter II', *Aerosol Sci. Technol.*, **1**, 217 (1982).
62. Y. Kousaka, T. Niida, K. Okyuama, H. Tanaka, 'Development of a Mixing Type Condensation Nucleus Counter', *J. Aerosol Sci.*, **13**, 231 (1982).
63. J.K. Agarwal, G.J. Sem, 'Continuous Flow, Single-particle-counting Condensation Nucleus Counter', *J. Aerosol Sci.*, **11**, 343 (1980).
64. B.Y.H. Liu, C.S. Kim, *Atmos. Environ.*, **11**, 1097 (1977).
65. Z.Q. Zhang, B.Y.H. Liu, 'Dependence of the Performance of TSI 3020 Condensation Nucleus Counter on Pressure, Flow Rate and Temperature', *Aerosol Sci. Technol.*, **13**, 493 (1990).
66. Y.F. Su, Y.S. Cheng, G.J. Newton, H.C. Yeh, 'Counting Efficiency of the TSI Model 3020 Condensation Nucleus Counter', *Aerosol Sci. Technol.*, **12**, 1050 (1991).
67. M.R. Stolzenburg, P.H. McMurry, 'An Ultrafine Aerosol Condensation Nucleus Counter', *Aerosol Sci. Technol.*, **14**, 48 (1991).
68. W.T. McDermott, R.C. Ockovic, M.R. Stolzenburg, 'Counting Efficiency of an Improved 30A Condensation Nucleus Counter', *Aerosol Sci. Technol.*, **14**, 278 (1991).
69. J. Keston, A. Reineking, I. Porstendorfer, 'Calibration of a TSI Model 3025 Ultrafine Condensation Nucleus Counter', *Aerosol Sci. Technol.*, **15**, 107 (1991).
70. D. Sinclair, *Atmos. Environ.*, **16**, 955 (1982).
71. K.J. Noone, H.C. Hansson, 'Calibration of the TSI 3760 Condensation Nucleus Counter for Nonstandard Operating Conditions', *Aerosol Sci. Technol.*, **13**, 478 (1990).
72. J.C. Wilson, J.H. Hyun, E.D. Blackshear, 'The Function and Response of an Improved Stratospheric Condensation Nucleus Counter', *J. Geophys. Res.*, **88**, 6781 (1983).
73. R. Niessner, C. Helsper, G. Roenicke, 'Application of a Multistep Condensation Nuclei Counter as a Detector for Particle Surface Composition', in *Aerosols*, eds. B.Y.H. Liu, D.Y.H. Pui, H. Fissan, Elsevier, New York, 63–66, 1984.
74. A.P. Waggoner, R.J. Charlson, 'Measurement of Aerosol Optical Parameters', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 511–534, 1976.
75. S.S. Butcher, R.J. Charlson, *An Introduction to Air Chemistry*, Academic Press, New York, 1972.
76. P. Lilienfeld, 'Current Mine Dust Monitoring Instrument Development', in *Aerosols in the Mining and Industrial Work Environment*, eds. V.A. Marple, B.Y.H. Liu, Ann Arbor Science, Ann Arbor, ML, 733–758, 1982.
77. D.A. Lundgren, L.D. Carter, P.S. Daley, 'Aerosol Mass Measurement Using Piezoelectric Crystal Sensors', in *Fine Particles*, ed. B.Y.H. Liu, Academic Press, New York, 485–510, 1976.
78. G.J. Sem, K. Tsurubayashi, 'A New Mass Sensor for Respirable Dust Measurement', *J. Am. Ind. Hyg. Assoc.*, **36**, 791 (1975).
79. G.J. Sem, K. Tsurubayashi, K. Homma, 'Performance of the Piezoelectric Microbalance Respirable Aerosol Sensor', *J. Am. Ind. Hyg. Assoc.*, **38**, 580 (1977).
80. Bowers, Chuan, 'Surface Acoustic Wave Piezoelectric Crystal Aerosol Mass Monitor', *Rev. Sci. Instrum.*, **60**(7), 1297–1302 (1989).
81. H. Patashnick, G. Rupprechl, 'Personal Dust Exposure Monitor Based on Tapered Element Oscillating Microscope', *BuMines OFR*, NTIS PB 84-173749, 58–84, 1983.
82. J.C.F. Wang, 'A Real-time Particle Monitor for Mass and Size Fraction Measurements in a Pressurized Fluidized-bed Combustor Exhaust', *Aerosol Sci. Technol.*, **4**, 301 (1985).
83. D.-R. Chen, D.Y.H. Pui, S.L. Kaufman, 'Electrospraying of Conducting Liquids for Monodisperse Aerosol

Generation in the 4 nm to 1.8 μm Diameter Range', *J. Aerosol Sci.*, **26**, 963 (1995).
84. R.N. Berglund, B.Y.H. Liu, 'Generation of Monodisperse Aerosol Standard', *Environ. Sci. Technol.*, **7**, 147 (1973).
85. B.Y.H. Liu, D.Y.H. Pui, 'A Submicron Aerosol Standard and the Primary, Absolute Calibration of the Condensation Nuclei Counter', *J. Colloid Interface Sci.*, **47**, 155 (1974).
86. D.Y.H. Pui, D.L. Swift, 'Direct-reading Instruments for Airborne Particles', in *Air Sampling Instruments for Evaluation of Atmospheric Contamination*, 8th edition, ACGIH, Cincinnati, OH, 1995.
87. K. Willeke, P.A. Baron, *Aerosol Measurement – Principles, Techniques and Applications*, Van Nostrand Reinhold, New York, 1993.

Dust, Measurement of Trace Elements in

Rod Barratt
The Open University, Milton Keynes, MK7 6AA, UK

1 Introduction **1**
1.1 What is Dust? 2
1.2 Define the Objectives 3
2 Sampling **4**
2.1 Collecting the Sample 7
3 Selection of Analytical Methods **8**
3.1 Particle Size 8
3.2 Sample Decomposition and Dissolution 9
3.3 Metal Determination 14
4 Data **17**
4.1 Relating Pollutants in Dusts to Sources 20
4.2 Is the Environmental Quality of Dust Acceptable? 21
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Dust is a source of contamination of land, water and the atmosphere, and is of special concern owing to its trace metal content. There is a significant body of literature describing the use of deposited dusts as indicators of environmental pollution, particularly by trace metals. While several authors have identified the need for standardized guidelines on sampling techniques and on analytical procedures, few have addressed the problems in sufficient depth to produce such guidelines. Since some of the early papers on this topic, more is known about the chemical nature of dusts and the association of certain metals with components of the matrix. This information provides useful guidance on analytical techniques and methods for dissolution of samples.

In the analysis of trace elements in dusts, a prerequisite is to prepare a solution for analysis. A variety of methods have been used to digest dust samples for determination of trace metals, and this can make it difficult to compare the results from different studies. Whether total or available metal content is determined is just one uncertainty needing resolution. Time requirements range from 15 min to 24 h, and this is another variability that may make results incomparable, while the great amounts of acids and long digestion times may result in contamination. For over 20 years, microwave ovens have found increasing use and have been used to digest various kinds of samples, such as rocks, sediments, ores, soils, coal, fly ash, water, sludge and biological and botanical samples. This article describes the efficiency of various acids for the digestion of dust, both by traditional methods and using a microwave oven. The latter technique offers an advantageous route for the digestion of dust samples in preparation for trace metal analysis. Speed and good recovery coupled with less risk of contamination through the limited amounts of reagents required combine to make this an excellent choice in the suite of techniques for environmental analysis.

General reviews of analytical methods demonstrate the dominance of atomic absorption methods for the analysis of trace metals such as lead and cadmium in a variety of samples. This pattern is generally found for work on metals in dusts, with flame atomic absorption predominating for Pb, Cd, Zn, Ni and Cu. Other techniques have included neutron activation (NA) and inductively coupled plasma atomic emission spectroscopy (ICPAES). Polarography in the form of differential-pulse anodic stripping voltammetry (DPASV) is suitable for the simultaneous determination of trace quantities of Pb and Cd in dust. The method gives good reproducibility and has the added advantages of being relatively cheap, quick and simple. Interferences and means for overcoming their effects will be outlined. The aim of analysis of dusts is to represent environmental quality. This necessitates appropriate presentation of the data from the analysis, and some issues relating to this will be outlined.

Dusts will always present a potential route of contamination by trace metals and organic substances and the need for future work in this area will not diminish. Only by the application of good quality control, the use of reference materials and reliable techniques and an understanding of the significance of the results can confidence be placed in a useful approach to environmental quality assessment.

1 INTRODUCTION

A good understanding of the quality of the environment is a precursor to rational decision making for environmental management (Figure 1). A variety of indicators serve in monitoring environmental quality, and in recent years dusts have been popular environmental samples for study. As a first impression, the concept of analyzing dusts may appear naive, but for many years investigators have used this medium as a convenient indicator of pollution levels. Dust analysis continues to be attractive to environmental scientists, influenced no doubt by its ubiquity

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

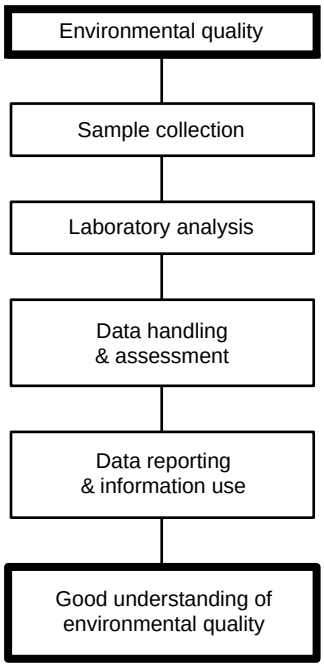

Figure 1 Understanding the quality of the environment is a precursor to rational decision making for environmental management.

and hence apparent ease of sampling. Dust is, indeed, a common material throughout the environment, although the degree of accumulation and its composition depend on many variables, including weather conditions, building construction materials, traffic density and industrial activity.

The role of dust as a potential vehicle for the intake of trace metals by children, especially in their home environment, has prompted many studies on dust composition. Its relevance is identified, inter alia, in a US Statute[1] following recognition by the US Congress of the following issues:

1. Low-level lead poisoning is widespread among American children, afflicting as many as 3 000 000 children under age 6, with minority and low-income communities disproportionately affected.
2. At low levels, lead poisoning in children causes intelligence quotient deficiencies, reading and learning disabilities, impaired hearing, reduced attention span, hyperactivity and behavior problems.
3. Pre-1980 American housing stock contains more than 3 000 000 tons of lead in the form of lead-based paint, with most homes built before 1950 containing substantial amounts of lead-based paint.
4. The ingestion of household dust containing lead from deteriorating or abraded lead-based paint is the most common cause of lead poisoning in children.
5. The health and development of children living in as many as 3 800 000 American homes is endangered by chipping or peeling lead paint, or excessive amounts of lead-contaminated dust in their homes.
6. The danger posed by lead-based paint hazards can be reduced by abating lead-based paint or by taking interim measures to prevent paint deterioration and limit children's exposure to lead dust and chips.

Similar observations on the role of paint have been made elsewhere.[2]

Most investigations on dust have been concerned with the lead content, with emphasis on sources from industry, traffic or paint. However, studies have included a broad spectrum of metals.

Careful sampling and analysis of dust are fundamental to any of these investigations, but the inhomogeneous nature of dust presents many problems to the investigator. Several of these aspects are covered in this article, with particular emphasis on obtaining data that are comparable with those from other studies and guidelines, such as exist. Clearly, the potential for harm depends largely on the bioavailability of the metals. Some studies have aimed at differentiating the available from the total metal content, and this emphasizes the care that is needed in comparing data from different studies.

1.1 What is Dust?

According to the International Union of Pure and Applied Chemistry (IUPAC), dust is[3]

> Small, dry, solid particles projected into the air by natural forces, such as wind, volcanic eruption, and by mechanical or manmade processes such as crushing, grinding, milling, drilling, demolition, shovelling, conveying, screening, bagging, and sweeping. Dust particles are usually in the size range from about 1 to 100 μm in diameter, and they settle slowly under the influence of gravity.

In the context of air pollution there is no definition of dust in the legal provisions of the UK Clean Air Acts. ''Grit'' is defined in regulations[4] as ''particles exceeding 76 μm in diameter'', whereas ''fume'' is defined in the Clean Air Act 1993 as ''solid particulate matter smaller than dust''. Hence a legal interpretation may be that dust is particulate matter smaller than 76 μm but with a lower limit of 1 μm.

These categories are supported by other guidelines:

1. *grit*, which is defined as ''particles of solid matter which are retained in a sieve of 75 μm nominal aperture in conformity with BS 410: 1986 of the British Standards Institution'';

2. *dust* is particles of solid matter which will pass through a sieve of 75 μm nominal aperture in conformity with BS 410: 1986;
3. *fume* is usually regarded as solid particulate matter smaller than dust, i.e. <1 μm in particle diameter.

Such a size limitation may be applied by those involved in air pollution control, but more general environmental investigations tend to follow a definition such as that in the Shorter Oxford English Dictionary:

> Earth or other solid matter so contaminated as to be easily raised and carried in a cloud by the wind; any substance pulverized; powder. Often extended to include ashes etc. from a house.

Some authors use the terms ''dirt'' or ''sediment'' as synonyms for dust, but all are referring to the general heterogeneous mixture of particulate matter deposited on surfaces. Clearly, the sources contributing to deposited dust are legion, but Fergusson and Schroeder[5] suggest that approximately 60% of house dust is non-organic in nature, and this probably contains most of the lead. Of the inorganic material in house dust, 30–40% comes from soil, 25–30% from street dust and 1–2% from aerosol that has settled in the house. Figures 2 and 3 illustrate X-ray spectra of two particles in dust. The first particle in a sample from a home (Figure 2) may have come from paintwork, as suggested by the high titanium level. The second particle in a sample from a domestic garage (Figure 3) clearly has a different origin. Absolute specification of their sources cannot be given without an unambiguous indicator species being identifiable. This approach will be referred to again later in this article.

Fergusson and Ryan[6] suggested that soil contributes more to larger than to smaller dust particles, and hence nonsoil particles probably contribute more to smaller dust particles. As around 50% of dust occurs as large particles (963–213 μm), soil is a major component of street dust. This was so for all cities studied, although the proportion was greater for some (75–90% for Halifax, Christchurch and Kingston) than others (60% in London and 57% in New York). Subsequently, Thornton[7] claimed that there is no reliable experimental method by which to quantify the soil component of house dust.

Clearly, soil is one of the major components of dust, whether from within or outside buildings, but the distinction between whether an external sample is dust or soil is problematic. It may be argued that size is not the critical factor in the definition of both types of particulate matter. To some extent, the term dust is dependent upon the source of the sample, so those samples from the roadside or inside buildings are called dust. Further, soil

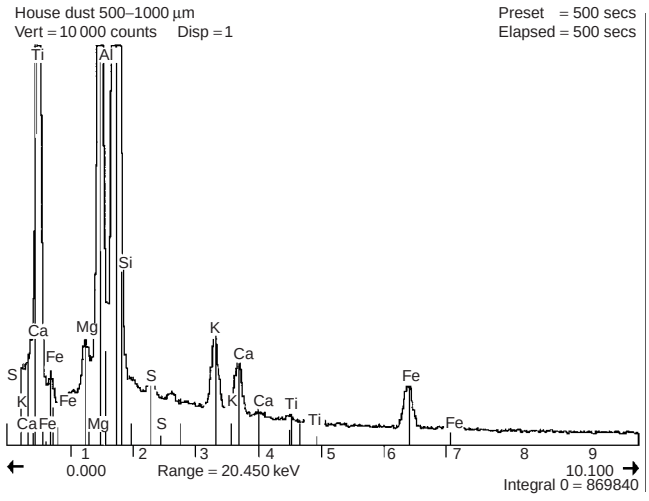

Figure 2 X-ray spectrum of particle in dust from a house showing a high Ti content.

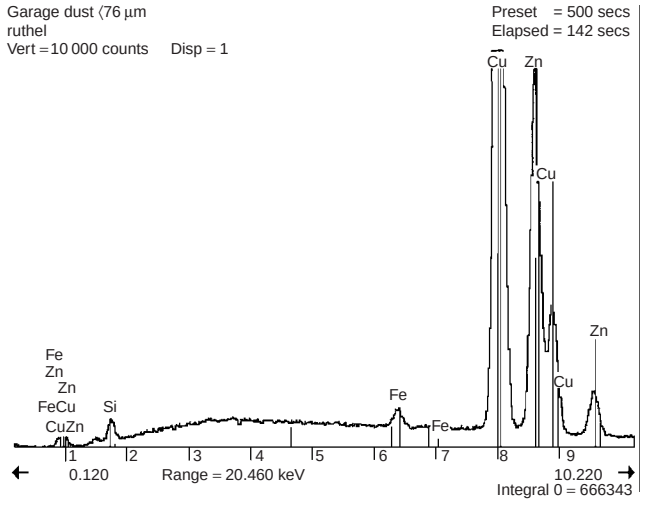

Figure 3 X-ray spectrum of particle in dust from a domestic garage.

is particulate matter which contains nutrients for plant growth; dust does not have this quality. Studies on the composition of soil in dust do not use this criterion.

This article is concerned with deposited dust collected from hard external surfaces and from inside buildings. Materials of an organic nature such as bacteria are not relevant to this article.

1.2 Define the Objectives

A variety of objectives have driven investigations involving dusts. These have included health risk assessment in domestic, workplace and general environmental settings. Other studies have focused on city-wide surveys

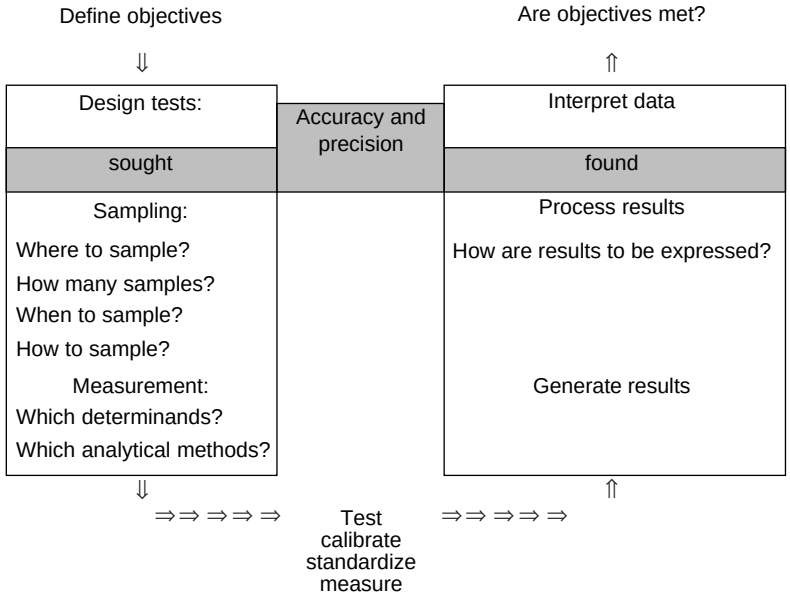

Figure 4 A sequential consideration of the factors involved in the design of a dust measurement program. (Reproduced with permission from John Wiley & Sons.)

and comparisons between cities. It is by no means certain that the design of the research programs has been properly based on achieving these objectives. A clear, precise definition of objectives makes the subsequent stages in an investigation capable of being designed efficiently, and ensures no wasted effort or wasted resources. It also makes it more likely that the results of the analysis will satisfy the requirements of the study on environmental quality.

Three broad objectives summarize the general reasons for carrying out dust analyses:

1. Investigations of the chemical and physical characteristics of dust, to determine factors such as its composition, content, solubility to certain solvents and size of particles.
2. General surveys of the distribution of trace elements in dust in certain areas, to elucidate the relationships between the trace element levels and factors such as traffic, road surface, meteorology. In this category the concerns are usually to evaluate the quality of the environment and to identify sources of pollution.
3. Evaluation of the quality of dust in relation to possible effects on health. Such studies often focus on the content of certain toxic metals in dust in areas where exposure of humans, and especially children, may occur.

The profile of an environmental issue which has often been claimed to be relatively neglected was raised following the publication of the House of Commons Environment Committee Sixth Report (1990). Indoor exposure to many air pollutants can often exceed outdoor exposures. The report notes that the relative neglect is because indoor pollution is often chronic and not directly or immediately life-threatening. Certainly this is true for pollution by metals in dusts, and there have been many studies looking at dusts collected within buildings, where there is greater potential for exposure of more sensitive subjects, namely children.

Organic material may be present in surface dust from a variety of sources, many of which have been listed in detail in the context of potential contamination of water through surface runoff.[8] However, relatively few studies have examined the organic composition of surface dusts. Nevertheless, fine dust re-suspended from the road has been linked to ambient levels of polycyclic aromatic hydrocarbons (PAHs) in air,[9] although there are suggestions that there are significant particle size implications.

A sequential consideration of the factors involved in the design of a dust measurement programme may follow the scheme in Figure 4. Many of these factors will be addressed in this article.

2 SAMPLING

Sampling is an essential precursor to analysis. The aim of sampling is to collect samples whose quality represents the quality of the whole. To achieve this, it is first necessary

for the sample to represent truly the temporal and spatial variation of the quality of the whole, and this requires consideration of the sampling locations and the time and frequency of sampling. A second prerequisite is that the analyte(s) of interest have the same values as in the original. Hence it is essential to consider methods of sample collection, transportation and storage, with implications for representativeness again, and also for avoiding contamination or analyte loss. Ultimately the results of analysis must be expressed and interpreted to relate to the objectives, as well as being comparable with other work. Elwood[10] commented that relevance of environmental measures to the actual exposure of subjects is probably one of the most difficult aspects of environmental research. While some compromise is usually necessary, some of the sampling procedures are utterly irrelevant. In particular, he cited that dust samples taken from a roadside gutter rather than from a pavement or from under a carpet are of unlikely relevance to human exposure. He concluded that the need for rigor increases as a field is progressively explored. This rigor starts by clarifying the objectives and following a sequence such as that outlined in Figure 4.

One of the earliest dust surveys[11] compared samples from urban and rural areas. Another study of similar date[12] had the objective of producing a city-wide survey of external and internal samples. External locations were classified according to road type, distance from industrial source or other characteristics according to a specified scheme. Some locations were repeatedly sampled on a seasonal basis while others were selected to give wide coverage. A later city-wide survey[13] included sampling at 2×2 km grid square intersections and subsequently on a 1×1 km grid.

As this program was concerned with background levels, no samples were collected within 20 m of busy roads, 10 m of building facades or 100 m of factories. However, Archer and Barratt[12] had demonstrated that the impact of an industrial source could be identified within 400 m of the source (Figure 5). Clearly, the height of discharge of dust will have an influence on the distance of contamination, and hence on the area for study or exclusion, depending on the objective. An elegant demonstration of dust sampling around a source was given by Hutton et al.[14] in relation to Pb and Cd from a municipal incinerator. Samples were obtained at varying distances along radii at 50° intervals from 0 to 90° in the downward area around the stack. Roadside sites were excluded from this sampling scheme. Unfortunately the report was marred by there being no description of the sampling procedure.

Where roadside contamination is of interest, the fundamentals of dispersion from low-level sources suggest that sampling within shorter distances may be more appropriate than when elevated sources are involved.

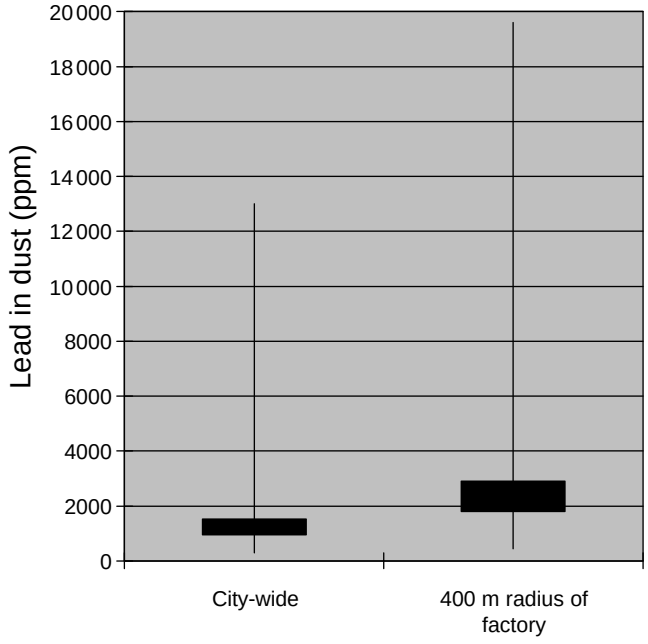

Figure 5 The impact within 400 m of an industrial source (vertical lines show high value of range, mean, median, low value of range.)

For example, Archer and Barratt[12] noted the decline in Pb concentration on moving from the gutter to the pavement (Figure 6). Grottker[15] reported that 95.85% of street dust solids were found within 0.5 m of the kerb (gutter section) and that larger particles were more predominant in this gutter section than in the asphalt section. In contrast, Hamilton et al.[16] found that for heavily used roads, lead concentrations were uniformly distributed over all surfaces sampled. This, they explained, was due to the rapid speed of the vehicles resulting in the movement of many of the Pb-rich fine particles towards the side of the road, where ''dilution'' with soil-derived material gave an even distribution of Pb. At one site, however, these authors did find elevated levels of Zn and Cd in the gutter.

The uncertainty of spatial variations was a question raised by Archer and Barratt[12] and discussed by Duggan,[17] who recognized that it was a problem difficult to address. No apparent reason could be found for differences in pavement samples spaced about 40 m apart, although samples collected on a 10 m grid pattern within playgrounds did not show significant differences. Fergusson and Ryan[6] reported that over a limited area, individual dust samples were reasonably representative of the area. For example, 10 samples collected 0.5 m apart and within a total distance of 5 m had a mean lead level of $3430\,\mu g\,g^{-1}$ and a standard deviation of $270\,\mu g\,g^{-1}$. Data were also presented for Cr, Cu, Zn

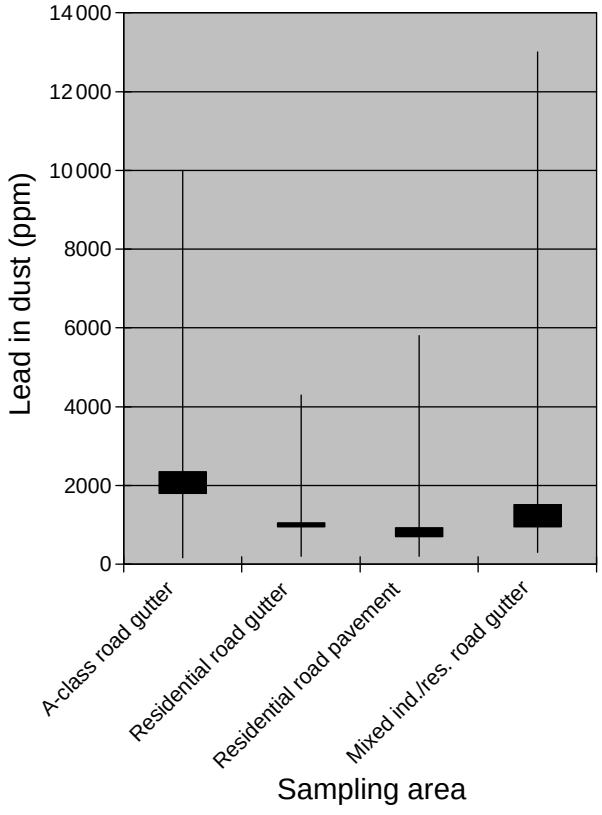

Figure 6 The changes in Pb concentration according to roadside position (vertical lines show high value of range, mean, median, low value of range.)

and Cd to demonstrate reasonable uniformity and to justify characterizing various cities by only two or three samples of dust. As these authors analyzed a wide range of elements (26) and found that the concentration of the major elements (Al, Cu, K, Mg, Ti) and several trace elements were similar, they suggested that this justified their approach. More work would seem to be needed in this area, for as stated by Hamilton et al.[16] it is an implicit assumption in all studies involving metals in dusts that representative samples have been taken.

In the context of soil sampling, Allen[18] noted that an investigator must allow for the inherent variability of soil samples. A high degree of spatial variability may exist over small areas and has been demonstrated by Ball and Williams.[19] Often the extent of this variation is unknown beforehand and hence sufficient samples must be taken to obtain a reliable measure of it. The absence of any indications of site variability can make it difficult to assess the significance of change such as seasonal variations that may be sampled by the data. Allen suggests that in practice between 5 and 10 replicates should be regarded as a minimum. However, if it is desired to estimate site variability with the same precision as might be obtained under more uniform laboratory conditions, at least 20 samples might be required.

By comparison, the scientific rigor usually applied to soil studies has often not been applied to dusts. Views on the numbers of samples required to produce meaningful results for dusts are as varied as the views on other aspects of dust sampling and analysis. Duggan[20] recommended that a considerable number of samples was necessary owing to the large spread of values expected. This broad spread was demonstrated by data in one of the largest of the early studies,[12] with typical ranges from a single city being reported as in Table 1.

In contrast, however, Fergusson and Ryan[6] compared pollution in six cities by taking two dust samples from each and three from a seventh. Each sample was obtained from an area not greater than 1 m^2. They noted the need for representative samples and commented that while few samples were taken, the ''similar'' concentrations for the major elements (e.g. Al, Cu, Na, K and Ti) and a number of trace elements suggested that the samples were reasonably representative. While this was true for some of the metals examined, close inspection reveals that some variations were wide by the standards of other studies. Thus, for example, lead levels ranged from 2008 to 4053 ppm in one city and from 887 ppm to 1.07% in another. Few other studies have relied on such small sample populations. Duggan[17] explored in more detail the random variations over short distances. Repeated systematic sampling within school playgrounds (total area 5000 m^2) demonstrated wide ranges and standard deviations, but no significant difference between the three means on the three occasions. He concluded that despite sharp variations in concentrations over fairly short distances, valid results could be obtained by taking a comparatively small number of large-area samples, 5 m^2 being indicated. While this modified view on sample numbers lends some support to the approach taken by Fergusson and Ryan, their sampling area of less than 1 m^2 is small.

While there appears to be some doubt about the validity of using limited numbers of samples for wide-area and even city-wide studies, investigations of specific

Table 1 Lead concentrations in urban dust samples

Sample type	No. of samples	Lead concentration ($mg\,g^{-1}$)	
		Median	Range
All roadside samples	1061	1200	160–50 000
Mixed residential–industrial area roadside gutter	161	950	300–1300
Residential road gutter	85	950	220–4300
Residential road pavement	180	700	200–5800
External surfaces other than roadside	455	1200	100–6700

sources clearly require more samples over a limited area. Thus Hutton et al.[14] sampled at 15 points between 0.15 and 5 km along 15° radii between 0 and 90° in the downwind direction from an incinerator. Schwar and Alexander[21] collected 5–12 samples in a "corridor" close to the external face of school buildings, while Rundle and Duggan[22] collected 7–17 samples in each school playground. Over 600 samples (internal and external) collected over a long period of time within a 400 m radius of a major lead-using works were considered by Archer and Barratt.[12]

The availability and cost of analytical facilities may determine the upper limit of the number of samples. One solution to the difficulty of analyzing large numbers of samples involves bulking, in which individual samples are combined to give a composite mixture. This method may give a satisfactory mean value, but conceals any site variability. The number of individuals making up the composite sample determines the validity of the mean. Allen[18] cites that in variable soils even 30 individual samples may not be sufficient. Bulking procedures must be standardized by taking samples of equal volume from similar situations. Davies et al.[23] used composite samples comprising 25 subsamples from pavements and roads. The subsamples were collected over an unspecified area but outside individual houses under investigation. One form of bulking may be regarded as sampling dust by sweeping a relatively large area. This approach was advocated by Duggan,[17] who indicated 5 m^2 as appropriate, whereas Schwar et al.[24] swept areas of 1–5 m^2 to help overcome problems of localized spatial variation.

2.1 Collecting the Sample

Implicit in this last reference is sweeping as a sampling technique. It is one of several options available, although none is without its problems.

There are recommendations of standard methods for the collection of dustfall[25] and suggested guidelines for assessing nuisance from deposited ambient dust.[26] Similarly, standards for collection of dust from flue gases[27,28] have been adopted and followed for many years. No standard guidelines exist for deposited dust. Thus, as with many other aspects of dust analyses, the variety of methods used make it difficult to compare results and this problem is compounded by the failure of many to give details of procedures used in their work.

In general, sample collection methods may be grouped as brushing or vacuum methods. Further dividing these options into areas of applicability[29] suggests

- single samples from vacuum cleaners for household dusts
- sweeping settled dusts from internal surfaces using a soft brush and thin card and storage in a self-sealing plastic bag
- bulked external samples collected with dustpan and brush.

Brushing involves using a soft brush to sweep dust from hard surfaces directly into plastic containers or on to paper and thence into containers or plastic bags. The need to avoid cross-contamination from the brush cannot be overemphasized. It is perhaps for this reason that some investigators prefer to use a spatula, which is also advantageous for damp samples. Suction techniques are another option, often using adapted domestic vacuum cleaners with collection of material on cellulose filters in an appropriate holder. Yet another option involves the use of moistened wipes to collect dust from surfaces, including hands.[30,31] The defined sampling area is wiped with a single, opened, premoistened tissue, which is then placed in a clean, labeled, self-sealing plastic bag. The person sampling wears polyethene gloves and a standard procedure is followed throughout.

Fergusson and Schroeder[32] gave one of the more detailed descriptions of sampling and suggested collection by suction (using a small diaphragm pump with vacuum of 27 kPa) from an area of 0.25 m^2 carpet on to glass-fiber filters held at 6 mm above the carpet. Another detailed description included an assessment of collection efficiency.[33] A different approach involves taking samples from the dust bag of domestic cleaners as included in the list above. This option was used, for example, in the Omaha childhood blood and environmental lead study[34] and in a national survey in the UK.[35] Implicit in all of these studies is the assumption that representative samples have been taken and that the approach has been standardized. Some claim to have done this.[36] There is also the issue of comparability, addressed by Archer and Barratt[12] in a matched set of brush and vacuum samples collected at each of 50 sites within a single office building. Their results were as set out in Table 2.

The results suggested to these authors that suction collected small particles such as fibers that effectively diluted the lead concentration, but more significant was the conclusion that this was an important area for further

Table 2 Comparison of brush and vacuum dust sampling techniques

Technique	Lead concentration (mg g^{-1})	
	Mean	Range
Vacuum	2700	1200–6300
Brush	3700	1300–11 000

work on sampling methodology. It is regrettable that part of a national survey of house dusts used techniques different from many others, and notably involved samples collected by both vacuum and brush techniques without confirming comparability.[23]

3 SELECTION OF ANALYTICAL METHODS

Measurements of toxic metals such as lead and cadmium in dust are necessary for their hazard evaluation and also for the identification of their natural and anthropogenic sources. Reliable, convenient and economic, sensitive analytical methods play an extremely important role in determining the content of toxic metals in dust. The options for sample analysis are varied and published descriptions of methods are sometimes full of detail and sometimes too brief to repeat. Some of these procedures are presented in Figure 7.

There are many possibilities in each step of the sequence, and several steps may be interchanged or associated. However, the first stage invariably involves preparing the sample in a form suitable for the subsequent analysis. This preparation stage will depend somewhat on the analytical method to be used later.

3.1 Particle Size

The inhomogeneous nature of dust was highlighted as a problem in the Introduction, and this is of no greater importance than in the context of particle size. The variation in metal concentration according to particle size may be seen by a cursory glance at the data for dusts collected from around an incinerator as summarized in Table 3. Notice that these data refer to "levels" expressed as micrograms per gram of trace metal measured in the sample. There are other options to consider, as will be outlined later.

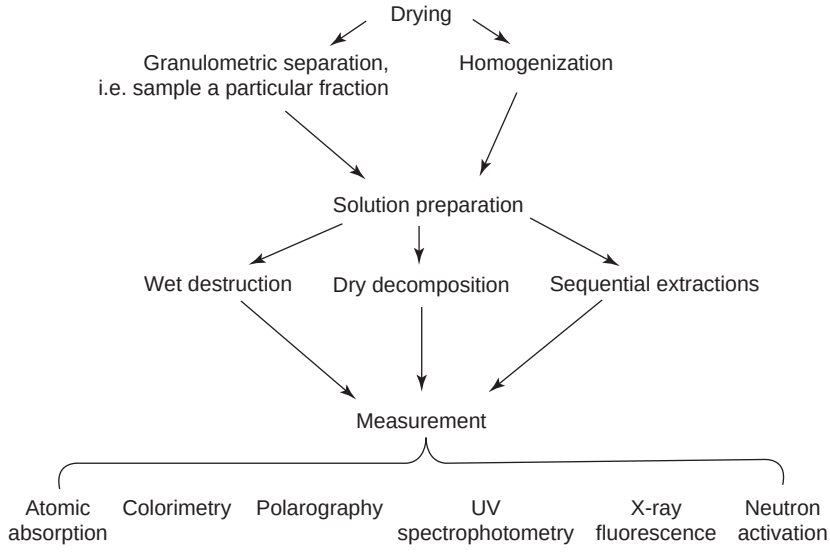

Figure 7 Analytical options for dusts. (Reproduced by permission from John Wiley & Sons.)

Table 3 Lead and cadmium levels ($\mu g\,g^{-1}$) in deposited dust around an incinerator

	Size					
	<75 μm		75–500 μm		500–1000 μm	
	Pb	Cd	Pb	Cd	Pb	Cd
Median	230.2	2.17	96.19	0.86	71.54	0.99
Mean	277.5	2.60	153.5	1.0	856.4	1.29
Range	24.3–565.6	0.46–8.81	10.28–613.5	0.04–3.21	8.4–8340	0.28–6.03
Downwind mean	232.4	2.64	124.4	0.88	734.5	1.32
	$\sigma = 113.5$	$\sigma = 2.13$	$\sigma = 115.6$	$\sigma = 0.76$	$\sigma = 219.0$	$\sigma = 1.45$
Upwind mean	346.6	2.52	197.7	1.18	1043	1.25
	$\sigma = 203.7$	$\sigma = 1.39$	$\sigma = 181.7$	$\sigma = 0.97$	$\sigma = 1875$	$\sigma = 1.18$

These observations are reflected in the wider literature. Fergusson and Ryan[6] examined nine particle size fractions in the range 33 ± 8 to 963 ± 38 μm. They found that the amount of each element in each size fraction decreases with the particle size because of the decline in the amount of dust in the smaller size fractions. Typically, 38–51% of the amounts of all elements for all samples occurred in the coarse fraction, defined as 213–963 μm. However, for most of the elements, their concentration increases with decrease in particle size, although the extent to which this occurred varied considerably from site to site. This distinction between amount and concentration will be developed later in this article in the context of levels and loadings of metals.

The significance of the particle size effect in terms of environmental effects includes the greater contamination of storm-water sediments through runoff flushing the finer particles into the drainage systems, and the greater risk of contamination of children's fingers leading to a greater risk of elevated body burdens of lead. From the analytical point of view, however, the problem is one of selecting a representative sample of the inhomogeneous mass.

The first step may be to reduce the sample bulk. Traditionally, mixing, heaping and quartering the air-dried sample has been the approach. The two opposite quarters are mixed and the process repeated to give the required sample size. It is usual to pick out large stones before proceeding with this process. For dusts, it seems unlikely that bulk reduction will be necessary unless a large cumulative sample is obtained. However, even if not reduced in bulk, the sample must be brought to a relatively homogeneous state by thorough mixing before being treated further.

The next problem concerns taking a specified size fraction for analysis or grinding the whole. Clearly if the latter route is chosen, care to avoid contamination of the sample, such as from cadmium in pigment of plastic grinding equipment, or to avoid differential loss of fine dust is essential. Hamilton et al.[16] stated that there was some evidence that grinding increases the measured metal level, presumably because there is a greater surface area exposed which makes metal extractions easier. This statement contradicts evidence presented by Jones and McDonald[37] that showed that ground samples gave lower values than sieved samples. However, both groups agree that grinding is inappropriate when dealing with environmental samples, as it is not the constituents of the core that are of interest. This important issue reflects the bioavailability of the metals, as mentioned earlier.

Despite this view, several groups have homogenized samples by grinding while others have selected only the dust fractions below a certain particle size. Again, the nonuniform approach makes it difficult to compare results, especially in view of the relationship between particle size and metal concentration. Thus, some groups have examined particle sizes below 250 μm, some below 500 μm, some below 1 mm, some below 2 mm and others have no specified size. It appears that the size is often selected merely because it relates to previous work in a city or within a group, but this does not facilitate wider comparison of data. This difficulty must be borne in mind when comparing data.

3.2 Sample Decomposition and Dissolution

This stage of sample preparation is an important area of trace analysis. Its aim is to present the sample to the analytical instrument in a form suitable for the analysis, and usually this is in the form of a particle-free aqueous solution. Generally, this will involve destruction of organic compounds and dissolution of inorganic solids. However, some argue that in terms of potential for harm, the bioavailability is the appropriate measure rather than total content. The bioavailable fraction can be considered as that fraction of the metal content that can be absorbed into the bloodstream after ingestion of the dust. The physical and chemical properties of the dust and also the particle size will all influence the bioavailability.

3.2.1 Extraction

It is possible to extract metals selectively from the sample according to the degree of bonding to the dust matrix. Five fractions are often considered as reflecting the strength of bonding to the matrix, and hence the environmental availability.

A particularly detailed study of sequential extractions for the speciation of particulate trace metals in sediments was carried out by Tessier et al.,[38] who adapted methods of soil analysis. The concept of partitioning the solid material into specific fractions proposes that these can be selectively extracted with appropriate reagents. The five fractions selected are summarized in Table 4.

In sequential extraction, the exchangeable fraction represents the surface-bound material, which is readily subject to sorption and desorption. Next, the carbonate phase represents moderate availability, which term also describes the subsequent fraction known as the Fe–Mn oxide phase. Some have further subdivided this latter phase according to the strength of adsorption to the Fe–Mn oxides. One extraction sequence for measurement of metals in these fractions of dusts is set out in Figure 8.

Thus, there may be an "easily reducible" fraction (residue shaken with 0.1 M $NH_2OH \cdot HCl$–0.01 M HNO_3 (20 mL), 30 min at 20 °C) followed by a moderately reducible fraction (residue shaken with 1 M $NH_2OH \cdot HCl$–25% CH_3COOH (20 mL), 4 h at 20 °C). The

Table 4 Fractions for partitioning of trace metals

1	Exchangeable fraction	Surface adsorption of trace metals. Changes in ionic composition (e.g. pH, Cl^-) are likely to influence sorption/desorption processes
2	Carbonate fraction	Trace metals associated with sediment carbonates, and hence susceptible to pH changes
3	Fraction bound to iron and manganese oxides	Iron and manganese oxides are present as nodules, concretions and cement between particles or as a coating on particles. They are excellent scavengers for trace metals and are thermodynamically unstable under anoxic conditions
4	Fraction bound to organic matter	Trace metals may be bound to organic matter present as organisms, detritus and coatings on mineral particles, etc., through complexations, peptization or bioaccumulation, etc. Under oxidizing conditions the organic matter can be degraded, which releases the metals
5	Residual fraction	Following removal of the previous four fractions, the residual solid may hold trace metals within the crystal structure. Normal environmental conditions will not release these metals

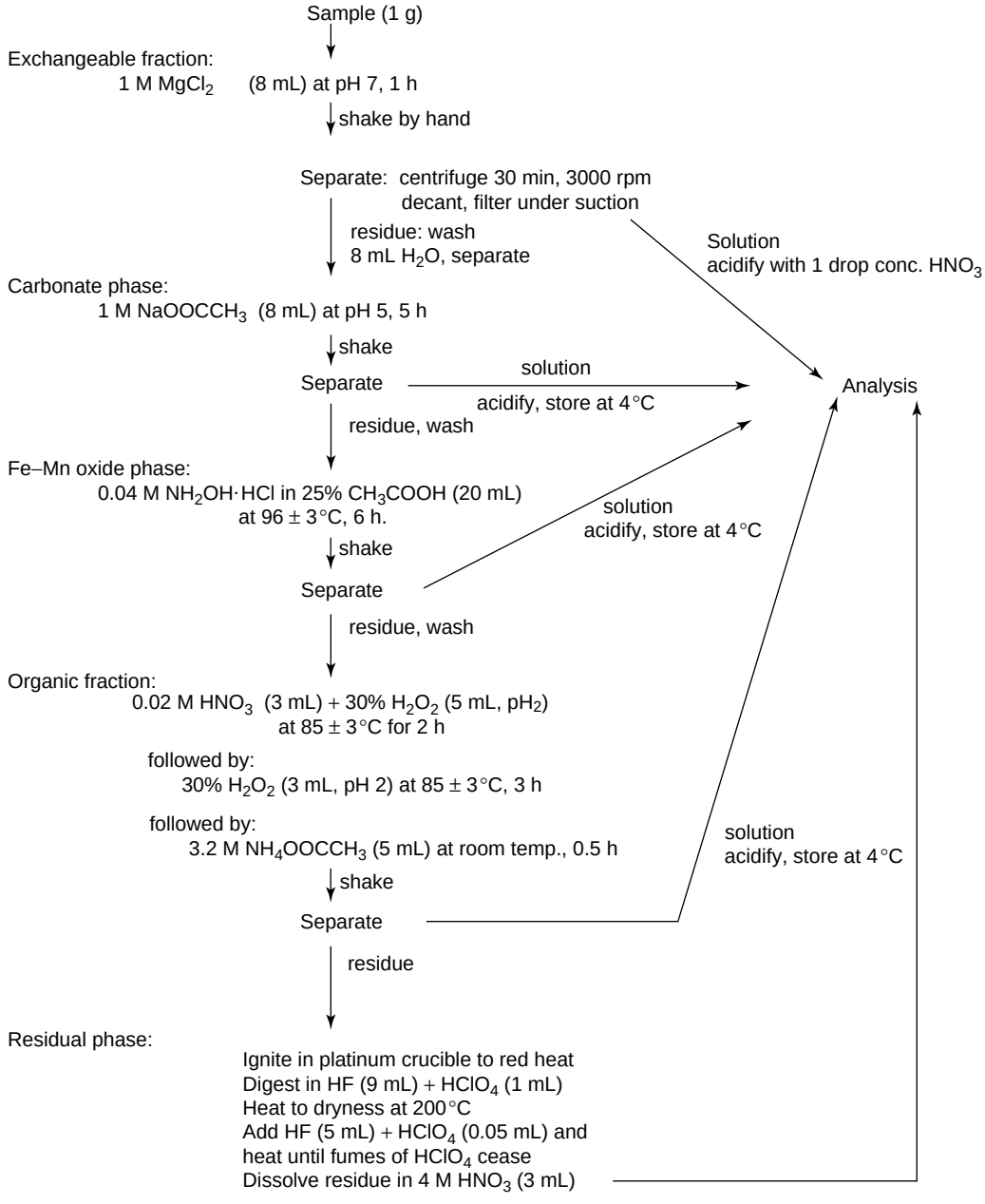

Figure 8 Extraction scheme for dusts. (Reproduced by permission from John Wiley & Sons.)

general scheme continues with the organic fraction and concludes with the residual phase, which represents the environmentally immobile fraction.

For dusts from selected cities across the world, Fergusson and Ryan[6] found that Pb, Zn and Cd showed similar extraction patterns (except for exchangeable Cd), the patterns for Mn and Fe were similar, while the extraction patterns for Cu were unique. Typically between 8.9 and 30.4% Cd was exchangeable whereas for the other metals the exchangeable component was relatively low (0–12.1%). Cu was present mainly in the organic fraction (32.6–74.3%) whereas the other metals were all present at less than 16.9% in this fraction. Pb and Cd were about equally distributed between carbonate and Mn–Fe oxide bound forms, while Zn was bound mainly to the Mn–Fe oxides. The major difference between the cities was for organic bound Cu, which related to the greater organic content of the dust in the bigger cities.

Gibson and Farmer[39] reported broadly similar partitions for Pb, Zn, Cu and Cd across the phases with the exception of Zn, which they found in the greatest proportion (26%) in the residual phase. However, they reported that a common feature of the four elements was a decline in the relative significance of the residual fraction with increasing concentration. This was not unexpected in contaminated environments with high input fluxes of metals from polluting sources. Leaching of samples with hydrochloric acid (0.07 M) to represent human stomach acidity indicated that an average 41% Pb, 44% Zn, 28% Cu and 47% Cd was released. The results suggested that Pb, Zn and Cd associated with the exchangeable, carbonate and easily reducible fractions were quantitatively released by the dilute hydrochloric acid, with some additional zinc being removed from the moderately reducible fraction.

Tessier et al.'s method is that used most widely in studies on deposited dust, but many have debated its merits and sought improved approaches. One serious problem with Tessier et al.'s approach is that each fraction cannot be separated completely. So, for example, the sum of Pb and Cd in each fraction always exceeds the total amounts of Pb and Cd determined by digestion in mineral acids. Several of the alternative sequential extraction schemes have been discussed elsewhere.[40]

3.2.2 Total Dissolution

Turning to the options for total destruction of the matrix, there is an extensive volume of literature on methods available for sample decomposition, which broadly divides into wet and dry techniques. In general terms, wet decomposition requires less equipment and there is less risk of volatilization and retention losses than for dry methods. Unfortunately, wet methods often require more reagents, more manipulations and hence present greater risks of contamination than dry decomposition. Dry methods require few reagents, little manipulation and bulky samples present few problems, although there is the volatilization risk which is particularly relevant to metals such as cadmium and lead.

There are, of course, inevitable exceptions. Alkaline fusion as a dry method requires more reagent, whereas wet digestion in the pressurized or microwave mode require little reagent, and such reagent is less contaminated. Such techniques do not appear to have found wide use in dust analyses. Reflecting on the relative merits of the options for sample destruction, it is perhaps no surprise that most investigators select wet methods. While it is hoped that reagent blanks and appropriate steps to control contamination and to purify reagents are included, it is unfortunate that the scant information on methods reported in many papers fail to mention these and other factors.

Table 5 illustrates a variety of techniques for sample destruction that have been used for dusts in the context of

Table 5 Decomposition of dusts for trace metal determination

Reagent	Time	Conditions	Metals
HNO_3–$HClO_4$ (9:1)	Overnight	–	Cd, Cu, Pb, Zn
HNO_3–$HClO_4$ (4:1)	24 h	50–190 °C	Pb
HNO_3–HCl	–	boiling	Pb, Cd, Cr, Co, Cu
HNO_3 (2 M)	30 min	–	Ni, Zn
HNO_3 (8 M)	4 h	85 °C	Pb
HNO_3–H_2O (1:1)	–	–	Pb, Cd
HNO_3–H_2O (2:1)	–	–	Cd, Cu, Pb, Zn, Ni
HNO_3 (4 M)	30 min	–	Pb
HNO_3 (25%)	15 min	Reflux after dry decomposition	Pb
Conc. HNO_3	–	–	Cd, Pb
Hot aqua regia	–	–	Pb
NH_2OH (0.25 M) in 25% CH_3COOH	Overnight	60 °C	Pb
HF–$HClO_4$ (10:1)	–	200–230 °C	Pb, Zn, Ni, Mn, Cu, Fe, Cr

a total or single representative concentration of the metal content. The gaps in the table indicate typical omissions of authors to give full details of procedures carried out.

It is clear from Table 5 that a variety of methods for sample dissolution have been used to produce a total or representative concentration of metal in dust. At a first impression, this range of methods appears to make comparisons suspect. Only one report describes dry decomposition, with dissolution of the residue in nitric acid.[12] Losses of metal (Pb) in the dry decomposition stage were avoided by maintaining the temperature below 550 °C. The dry method was attractive in view of the large numbers of samples being processed. Of the wet decomposition methods available, Fergusson and Ryan[6] found that acid extraction of samples (0.1–1 g) for 30 min in HNO_3 (4 M, 10 mL) recovered more than 90% of lead and cadmium, and over 80% copper. These estimates were made by comparison with tests on samples completely dissolved in $HF–HClO_4$, and agreed with previous work. Farmer and Lyon[41] analyzed soils and dusts using a single extraction procedure involving acetic acid and hydroxylamine hydrochloride. They found that for soils it was over 90% effective in recovering lead; the results were usually 5–10% less than those obtained when 8 M HNO_3 was used. Comparable studies on dusts were not quoted. The authors attributed the greater recoveries using HNO_3 to its ability to attack organic matter, which in other studies they carried out with H_2O_2. This reagent recovered less than 5% of dust and soil lead, and so they concluded that only minor amounts of lead were associated with the organic phase. It must be remembered that a ''good'' recovery for soils is not necessarily indicative of the behavior of dusts; the trace pollutant bound to the organic fraction of soils can be stripped off relatively easily. Many dusts may be silicaceous while others contain aluminates, the exact chemical species depending on the origin of the components. Thus wet digestion without HF is unlikely to be complete, and should not be expected to give the total metal content. While some studies have demonstrated comparability of wet digestion with and without HF, it cannot be assumed that all dusts will behave similarly. Clearly, the varying nature of dust samples from different sites may cause recoveries to differ from dust to dust, unless a reliable ''total'' decomposition method is used. This implies that all except the last method in Table 5 are subject to uncertainty.

In the absence of a standard method for extracting lead from dust, Jones and McDonald[37] compared the extraction efficiencies of four methods used in city-wide surveys in the UK. The authors did note that the method reported for samples from London was insufficiently detailed and precise to enable replication. From the methods tested, however, it was concluded that the differences in the analytical methods may account for the different lead values found in the different cities. Considerable caution was therefore recommended when comparing data from different surveys, and it was concluded that the problem will persist until there is agreement on a standard method of extraction of lead from street dust. This view echoed that of Archer and Barratt[12] 7 years previously. Only a total digestion procedure will allow comparison of the results of analyses of different dusts with confidence.

The sample preparation methods considered so far are relatively simple, although up to 24 h may be needed for some procedures. The main problem is selecting a reagent that is effective in releasing the metal of interest in a form that represents the potential release in relation to the exposure of a subject in the environment. It may be argued that attempts to seek the ''total'' metal in the sample give the worst case of release into the environment. The methods considered so far, with the exception of HF digestion, should be called leaching methods. Virtually by definition, leaching methods give information on the way that metals are bound in the matrix, and by definition they do not give the total content.

3.2.3 The Microwave Oven Option

Microwave oven techniques have been used to digest various kinds of samples for many years.[42–45] Compared with classical heating techniques, microwave methods are many times more efficient. The heating process is speeded up considerably and the sample preparation time can be reduced up to 10-fold. Two types of microwave digester are readily available, closed systems that use multiple pressurized containers, and open systems that focus the microwaves on the sample at atmospheric pressure. The conventional technique involves microwave heating of the sample with acid contained in a closed vessel comprising a body, cap and pressure release valve as illustrated in Figure 9. The whole of the vessel is constructed of Teflon®, which is relatively transparent to microwaves. A carousel within the oven holds several sample vessels. Heating causes a buildup of pressure within the closed vessel, and this facilitates breakdown of the sample owing to the higher boiling point of the reagents. This method has been used for digesting dust samples with different acids. Digestion in a microwave oven with different acids for various kinds of samples has been shown effective for Pb and Cd contents. The high pressure achieved in sealed vessels undoubtedly accounts for the effective digestion. The program used in the microwave oven method can bring the samples to 175–180 °C at a peak pressure of around 6 atm for most samples.

When HF has been used in the digestion procedure, it is essential to remove HF from the residue, otherwise

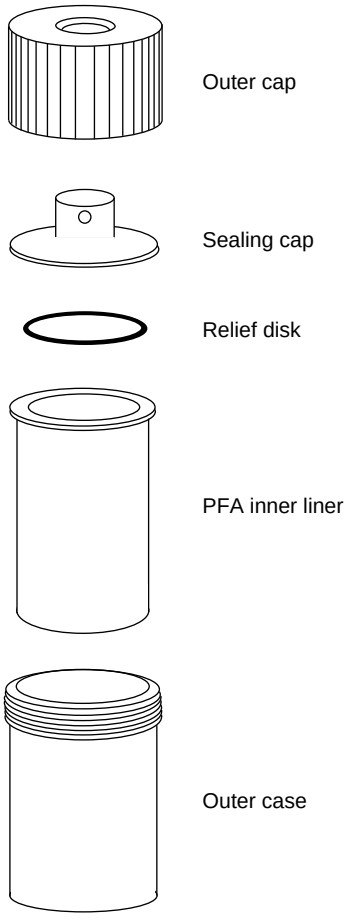

Figure 9 Microwave oven digestion cells. (Reproduced by permission from John Wiley & Sons.)

it will not only etch glassware and electrodes used in DPASV but also interfere with the determination of Pb by atomic absorption spectroscopy (AAS). A suitable removal procedure involves reaction of HF with boric acid according to a procedure we have reported elsewhere.[46] To remove acids from open vessels, it is preferable to use low power, otherwise loss of elements may occur. The risk of explosion in a sealed vessel precludes the use of perchloric acid in the microwave oven. Similarly, the use of sulfuric acid is not recommended owing to its high boiling point with the consequent risk of damage to the Teflon® vessels.

Three standard reference materials were used to validate this method of digestion (Table 6). Table 7 shows results by traditional digestion methods using different acids. Table 8 shows the recoveries using the microwave oven technique, and Table 9 compares the results by traditional digestion and the microwave oven method. It is clear that lower losses of metals occur in the microwave method, and when combined with the additional advantages of speed, lower acid requirements and lower blank values, the technique can be seen to be very attractive. However, there are limitations of the method. High-pressure digestion carries an inherent risk of explosion and there may be temperature limitations on the vessel used. In addition, multiple stage treatment is difficult when reagents have to be added during the dissolution procedure. Open microwave digestion systems overcome some of these problems, and can also handle larger sample sizes and avoid the cooling down period before pressurized vessels may be opened safely.

Table 6 Certified reference samples used for dust analyses

Metal	B1-76 (urban waste incineration ash)	N16-48 (urban particulate matter)	C83-01 (river sediment)
Pb ($mg\,g^{-1}$)	10.87	6.55	0.079
Cd ($\mu g\,g^{-1}$)	470	75	2.45

Table 7 Results of traditional digestion with different acids

Metal	Recovery (%)		
	HNO_3	$HCl + HNO_3$	$HNO_3 + HClO_4$
Pb	74.50 ± 3.91	97.95 ± 11.81	49.17 ± 2.73
Cd	89.38 ± 2.36	101.96 ± 9.91	88.64 ± 2.91

Table 8 Digestion efficiency (%) in microwave oven ($n = 3–6$)

Sample	Metal	HNO_3[a]	$HCl + HNO_3$[b]	$HNO_3 + HF$[b]
B1-76 (urban waste incineration ash)	Pb	93.23 ± 2.29	98.71 ± 11.78	98.82 ± 8.02
	Cd	94.58 ± 6.13	95.72 ± 5.84	93.69 ± 6.11
N16-48 (urban particulate matter)	Pb	86.46 ± 12.86	88.14 ± 7.39	98.98 ± 20.77
	Cd	87.38 ± 6.07	77.66 ± 3.58	92.81 ± 10.91
C83-01 (river sediment)	Pb	89.88 ± 12.97	95.92 ± 7.81	81.51 ± 6.36
	Cd	88.76 ± 5.79	91.85 ± 12.57	82.27 ± 8.69

[a] Data from AAS.
[b] Data from DPASV.

Table 9 Comparison of traditional wet digestion with microwave oven methods

Acid	Traditional wet digestion		Microwave oven digestion	
	Pb ($mg\,g^{-1}$)	Cd ($\mu g\,g^{-1}$)	Pb ($mg\,g^{-1}$)	Cd ($\mu g\,g^{-1}$)
HNO_3[a]	9.34 ± 0.54	414.73 ± 8.63	10.11 ± 0.25	444.6 ± 28.8
$HCl + HNO_3$[b]	10.69 ± 1.02	449.0 ± 8.39	10.73 ± 1.28	449.9 ± 27.5

[a] Data from AAS.
[b] Data from DPASV.

Table 10 Common analytical methods for trace metal determinations

Pollutant	AAS	Colorimetry	Polarography	UV spectrophotometry	X-ray fluorescence	NA analysis
Pb (%)	93	0.5	4	2.5	<0.1	2.5
Cd (%)	88	2.5	6	0.7		

3.2.4 Residue Analysis

As commented upon previously, rarely is complete dissolution achieved with different acids. Even HF leaves a residue, particularly for river sediment, which has high silica content. X-ray spectrometry was used to check the residue to provide further evidence of digestion efficiency with different acids. This further demonstrated the utility of the microwave oven approach.

In summary, a microwave oven offers a route with many advantages for the digestion of dust samples. Speed and good recovery coupled with less risk of contamination through the limited amounts of reagents required combine to make this an excellent choice in the suite of techniques for environmental analysis. It is important to bear in mind, however, that although microwave digestion has these merits from the chemical analysis perspective, this may not be the conclusion in terms of assessing environmental health effects, for which biological digestion is more appropriate. This caveat has to be remembered in the context of assessing the significance of the analytical data at a later stage.

3.3 Metal Determination

A general review of analytical methods used throughout Europe[47] for principal pollutants demonstrated the dominance of atomic absorption methods for trace metals such as lead and cadmium, as Table 10 indicates. This distribution of methods is generally found in reports on subsequent work on metals in dusts, with flame atomic absorption predominating for Pb, Cd, Zn, Ni and Cu. Polarography was used for 10% of samples in studies on lead by Day et al.,[11] who found agreement with atomic absorption analysis better than ±3%. Hamilton et al. used the same two techniques, with polarography in the form of DPASV.[48] Other techniques have included NA analysis and ICPAES.

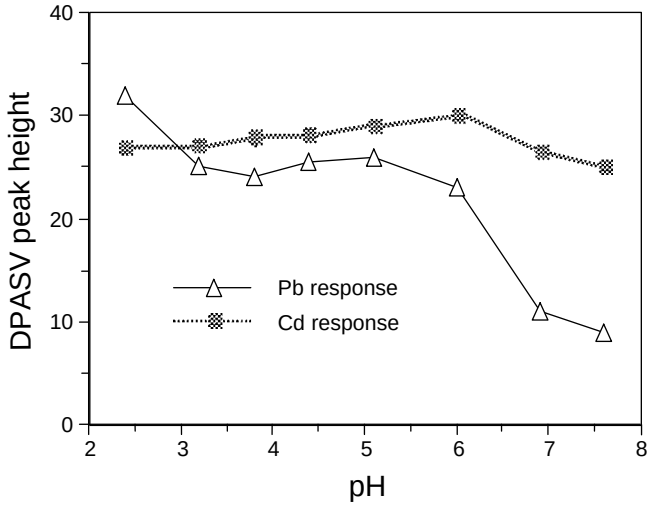

Figure 10 DPASV peak height as a function of pH of the solution.

The literature on the use of DPASV for the determination of Pb and Cd in dust is limited; the method has been used primarily for the determination of metals in airborne particles and in soil. By comparison with AAS, DPASV offers rapid and simultaneous determination of Pb and Cd, and also Zn and Cu, if desired. A simpler treatment of dust samples and less expensive equipment are further advantages. The results of studies to optimize the conditions for dust analyses are summarized below.

3.3.1 The Effect of pH and Buffer System

It is clear from Figure 10 that for Pb and Cd the oxidation current peaks ($[Pb^{2+}] = [Cd^{2+}] = 4$ ppb) are independent of the pH of the solution in the acid range up to pH 6.

Above pH 6, the oxidation current decreases sharply and this is attributed to the formation of hydroxides.

In order to keep the ionic strength of solution constant, different supporting electrolytes are used. No significant differences have been found in the use of $KClO_4$, KCl, KNO_3 and sodium ethanoate. In practice, after digestion, HCl and NaOH were used to adjust the pH of the solution to neutrality, then 5 mL of 0.5 M ethanoic acid–sodium ethanoate buffer (pH 4–5) were added to keep the pH and ionic strength of the solution constant.

3.3.2 The Effect of Deposition Time on Peak Current

Prolonging the deposition time can increase the sensitivity of the stripping determination. However, this must not be too long, i.e. the concentration of the substances in the solution must not perceptibly decrease, or repeated results cannot be obtained. For Pb and Cd in the range 0.1–0.01 ppm, a deposition time of 2 min is enough to obtain a strong signal. For Pb and Cd concentrations of 0.01–0.001 ppm a deposition time of 5 min is suitable. Concentrations below 0.001 ppm demand a deposition time of 10–15 min in order to obtain a strong enough signal that is distinguishable from that of a blank solution.

3.3.3 Interference Studies

As noted previously, hydrofluoric acid offers the only sure method of digesting the whole of the dust matrix to release all trace metals. However, if dust samples are digested with this reagent, all residual acid must be removed subsequently, otherwise it will not only etch the glassware and electrodes but will also interfere with the determination of Pb and Cd. When the concentration of HF in the solution to be analyzed is below 0.04%, no obvious effects were observed. At a concentration of 0.16% the oxidation current for Pb increased more and more as the determination time increased, while the oxidation current for Cd decreased to a constant value. Boric acid can be added to react with HF to form tetrafluoboric acid.[49] This method has been used in inductively coupled plasma emission spectroscopy (ICPES).[50–52]

Up to 3.2% of H_3BO_3 has no effect on the determination of Pb and Cd by DPASV. The efficiency of HF elimination by boric acid in the solution of 0.40 ppm Pb and 4 ppb Cd is shown in Table 11.

Other potential interferences were studied[40] by determining 0.04 ppm Pb and 0.02 ppm Cd in the presence of the following metal ions: Ag^+, Bi^{3+}, Ce^{3+}, Co^{2+}, Cr^{3+}, Cu^{2+}, Fe^{3+}, In^{3+}, Mn^{2+}, Ni^{2+}, Sb^{3+}, Sn^{2+}, Tl^+ and Zn^{2+}. All were introduced at 4 ppm, which in most cases will be much higher than their concentration in the solution after digestion. Of these ions, only In^{3+}, Sn^{2+}, Cr^{3+} and Tl^+ showed significant interferences, as indicated in Table 12.

Table 11 Efficiency of hydrofluoric acid elimination by boric acid

H_3BO_3 (%)	HF (%)	Pb (ppm) ± SD	Cd (ppb) ± SD
0.8	0.576	0.40 ± 0.003	3.99 ± 0.05
1.6	0.96	0.39 ± 0.01	3.93 ± 0.26
3.2	1.92	0.41 ± 0.01	4.09 ± 0.10
3.2	2.88	0.42 ± 0.02	4.06 ± 0.53
3.2	3.84	0.42 ± 0.03	3.95 ± 0.27

Table 12 Principal interferences in the determination of Pb and Cd by DPASV

Ion	Interference	Limiting concentration (ppm) to avoid interference with Cd or Pb
Cr^{3+}	Ill-formed peak at −0.65 to −0.75 V which interferes with Cd peak at −0.58 V	4
In^{3+}	Peak occurred at −0.53 V and interferes with Cd peak	0.02
Sn^{2+}	Peak at −0.4 V interferes with Pb peak	0.05
Tl^+	Peak at −0.48 V interferes with Pb	

The available information for levels in dust suggests that In^{3+}, Sn^{2+} and Tl^+ are present at very low levels, while the Cr^{3+} level of 20–200 $\mu g\, g^{-1}$ is beyond the interference range. Organic substances can also interfere in these determinations, but a technique for overcoming such problems uses the addition of potassium iodide.[53] Adsorption of organic species such as surfactants may shift the peak potential in the anodic direction and may decrease the peak height. Removal or destruction of organic matter before the stripping analysis is desirable for samples rich in organic matter. Ultraviolet (UV) irradiation or ozone oxidation often accomplishes this, but both are time consuming. Other options have included ion-exchange columns in a flow-injection system, silica for adsorptive stripping voltammetry and poly(ester–sulfonic acid)-coated mercury film electrodes for DPASV. The use of KI not only reduces the effect of adsorption caused by some organics, but also gives the added benefit of enhancing the limit of detection (LOD) of Pb and Cd by DPASV. The effect is illustrated in Figure 11.

3.3.4 Limits of Detection, Accuracy and Precision

The LOD of the technique is dependent on the blank value, and is improved by purifying the chemicals used in the determinations. Using the voltammetric conditions

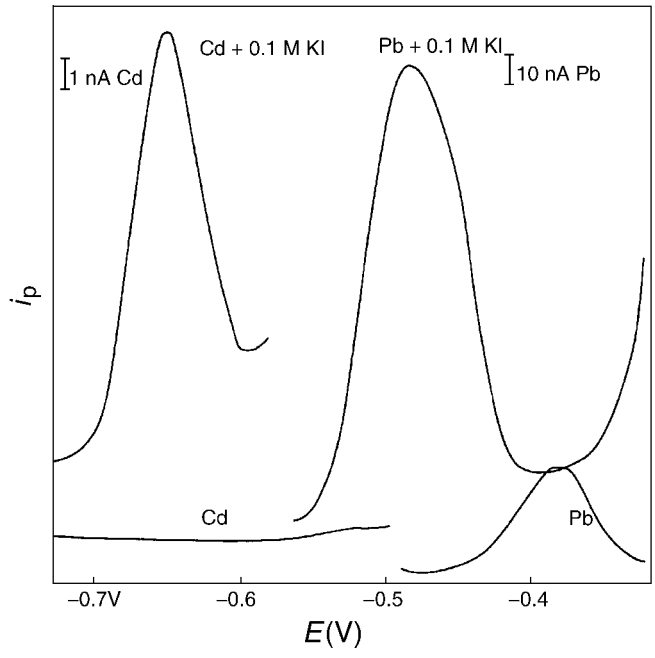

Figure 11 Benefits of KI addition in DPASV analysis.

Table 13 LODs

Blank	Pb (ppb) ± SD	Cd (ppb) ± SD
HNO_3–HF–H_3BO_3	0.20 ± 0.17	0.10 ± 0.08
HNO_3 or aqua regia	0.10 ± 0.08	0.11 ± 0.09
Ethanoic acid–sodium ethanoate	0.06 ± 0.05	0.07 ± 0.06

outlined in the procedure, the LODs were established as set out in Table 13. The accuracy and precision of the method were evaluated by replicate analyses of certified reference materials (CRMs) from the Laboratory of the Government Chemist (LGC). Table 14 indicates that the standards of performance achieved are satisfactory with the coefficient of variation ranging from 4 to 10%. The recovery for both Pb and Cd from all three samples is more than 90%.

A comparison between DPASV and AAS (Table 15) showed no significant difference (*t*-test) between the two methods.

3.3.5 Convenience

Determining Pb and Cd by AAS usually involves a dilute HNO_3 medium, because other Pb and Cd compounds, such as $PbCl_2$, $CdCl_2$, PbF_2 and CdF_2, are not as stable as $Pb(NO_3)_2$ and $Cd(NO_3)_2$. After digestion, concentrated acid must be removed by evaporation before analysis by AAS, whereas for DPASV the determination can be performed without evaporation. That makes the determination simple and rapid.

Table 16 Typical measuring parameters for Pb and Cd determination by DPASV

Hg drop size	Medium
Degassing	5–10 min
Deposition potential	−0.8 V
Deposition time	2 min for 0.1–0.01 ppm 5 min for 0.01–0.001 ppm 10 min for <0.001 ppm
Delay	10–30 s
Sweep rate	10 mV s^{-1}
Sweep range	−0.8 to −0.3 V
Pulse amplitude	50 mV

Table 14 Recovery tests ($n = 3$–6)

Sample	Certificate value		Determined value		Recovery (%)	
	Pb (mg g^{-1})	Cd (μg g^{-1})	Pb (mg g^{-1})	Cd (μg g^{-1})	Pb	Cd
B1-76 (urban waste incineration ash)	10.87 ± 0.17	470 ± 9	10.74 ± 0.86	440.34 ± 26.90	98.8	93.7
N16-48 (urban particulate matter)	6.55 ± 0.08	75 ± 7	6.48 ± 1.35	69.61 ± 7.59	98.9	92.8
C83-01 (river sediment)	0.079 ± 0.012	2.45 ± 0.3	0.073 ± 0.01	2.25 ± 0.20	92.4	91.8

Table 15 Comparison of analytical results obtained by DPASV and AAS ($n = 3$–6)

Sample and analysis	DPASV		AAS	
	Pb (mg g^{-1}) ± SD	Cd (μg g^{-1}) ± SD	Pb (mg g^{-1}) ± SD	Cd (μg g^{-1}) ± SD
B1-76 digested with HNO_3 in microwave oven	9.39 ± 1.64 9.52 ± 0.72	393.9 ± 59.2 392.2 ± 28.5	10.32 ± 0.58 9.96 ± 0.48	473.4 ± 43.1 431.9 ± 27.7
B1-76 digested with HCl + HNO_3 by traditional method	9.74 ± 1.95 11.55 ± 1.57	446.3 ± 7.3 512.1 ± 88.6	10.26 ± 0.31 11.77 ± 0.62	432.4 ± 34.7 460.1 ± 10.3
Garage dust digested with HCl + HNO_3 in microwave oven	0.83 ± 0.10 0.16 ± 0.05	3.15 ± 0.13 0.44 ± 0.04	0.66 ± 0.10 0.15 ± 0.02	3.11 ± 0.10 0.31 ± 0.19

The effective conditions for the determination of trace metals in dusts using DPASV are summarized in Table 16.

4 DATA

On inspection of attempts to represent environmental quality by the composition of dusts, it is apparent that different authors have used different ways of representing analytical data.

Analyses of environmental samples may be carried out on the material as received, air dried or oven dried, but the final results are best reported on an oven-dry basis. This is usually equivalent to drying to constant weight at 105 °C. Many investigators have dried samples at 105 °C, while others dried at temperatures ranging from 30 °C (air drying) to 80 °C. Some equilibrated samples to constant temperature and humidity before weighing, while too many authors in the literature have failed to mention whether samples were dried and on what basis their results were presented; their data are therefore of limited value to others working in the field.

Most workers have expressed total or extractable concentrations in units of $\mu g\,g^{-1}$ or similar derivatives, although some use ppm. Some specify units of $\mu g\,m^{-2}$, as discussed later. As most studies involve many samples, the representative values may be quoted as arithmetic means (although in one case of two samples), as ranges, as geometric means or medians. Archer and Barratt demonstrated the positive skewness of the frequency distributions of lead concentrations in dusts,[12] with the log-normal pattern characteristic of many pollution data being evident in the figures presented. Similar frequency distributions have been found for cadmium, copper, lead, zinc and iron. It is clear from many observations that the use of median or geometric mean concentrations is a preferred representation of the ''average'' value of the metal concentration. It has been suggested that the median concentration for dust samples from several sites varies with time by less than 30% for all metals. Hence, although individual results may vary widely, as ranges indicate, the median of a number of samples can provide a reliable picture of metal distributions. There appears to be little significance in consistent ratios of geometric and arithmetic means claimed by several investigators, however. Doubt on this was cast by Archer and Barratt,[12] whose data indicated ratios varying from 1.36 to 4.95. The ratio is consistently much greater close to buildings, and even more inside buildings.

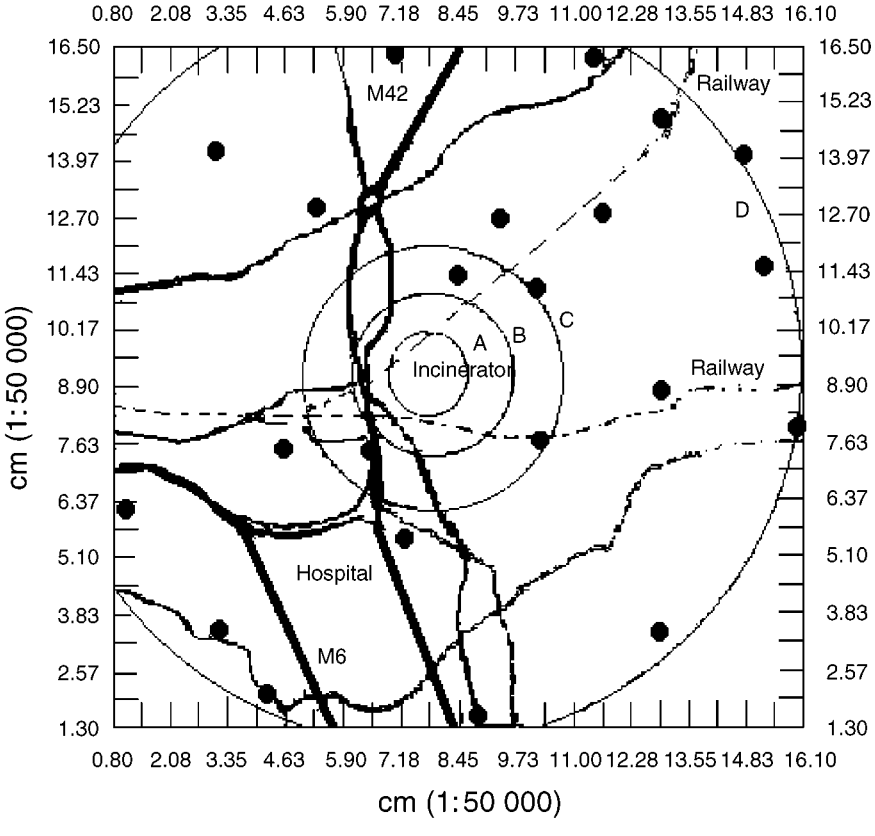

Figure 12 The incinerator and its surroundings.

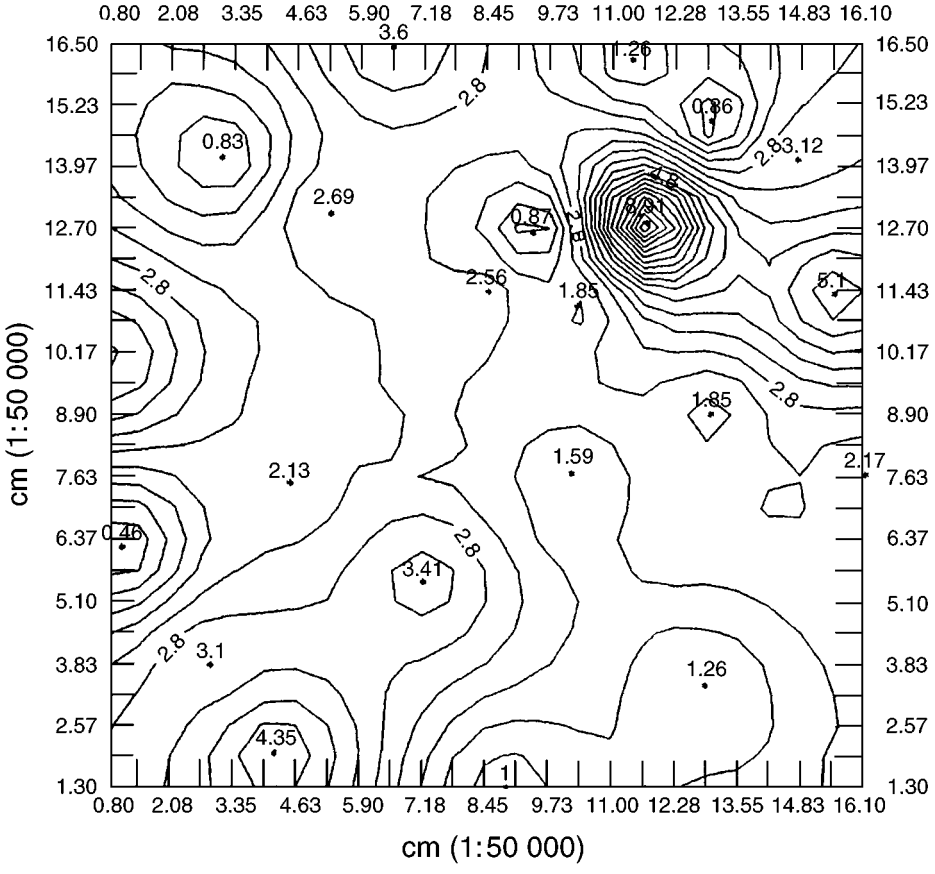

Figure 13 Isopleths of cadmium levels in dusts collected around an incinerator. (Reproduced with permission from John Wiley & Sons.)

Another issue concerns the concepts of *level* (in $\mu g\,g^{-1}$) and *loading* (in $\mu g\,m^{-2}$) that are used. Such different representations of data make comparisons difficult and may lead to different conclusions. This problem is best illustrated by an example involving two hypothetical samples. If sample 1 is collected from an area of $1\,m^2$, while sample 2 is collected from $10\,m^2$, the dust loading for both may be $1000\,mg\,m^{-2}$ and both levels of lead $1000\,\mu g\,g^{-1}$. However, the loading of lead for sample 1 is $1000\,\mu g\,m^{-2}$ whereas that for sample 2 is $100\,\mu g\,m^{-2}$. Depending on the method of presentation of the results, it may be concluded that the quality of dust expressed by the level of lead is the same, but it may be argued that in terms of the quality of environment area 2 is better than area 1.

Unfortunately, relatively few studies have acknowledged problems in data presentation such as that illustrated above, with few authors representing data as loading of trace elements. Those who use loading often fail to draw the reader's attention to the significance of the terminology.

The difficulties may be illustrated using data from the investigation around the incinerator referred to previously. A representation of the source and its surroundings is given in Figure 12. The levels (i.e. concentrations) of Pb and Cd in different sized dusts were determined and gave the results shown earlier in Table 3. The results indicate that Cd levels in fine dust exceed those in coarse dust. Pb levels change more widely as particle size increases. Figures 13 and 14 represent cadmium and lead levels as isopleth maps, and while statistically insignificant on the basis of mean upwind and downwind results, it may still be seen from Figure 13 that the highest Cd level is in the downwind direction from the incinerator, about 2.25–2.5 km distant. This is the region where the greatest contribution by airborne dispersion from the incinerator to ground level concentrations was predicted under the most common stability category in the UK, namely class D. It is noteworthy that the same conclusion would be drawn from an analysis of the levels in the largest sized particles, although data from the mid-range size would suggest the greatest concentration elsewhere.

The levels of Pb shown in Figure 14 suggest no obvious relationship with the presence of the incinerator. Areas with the highest levels are in opposite directions, both upwind and downwind of the incinerator. Of special note

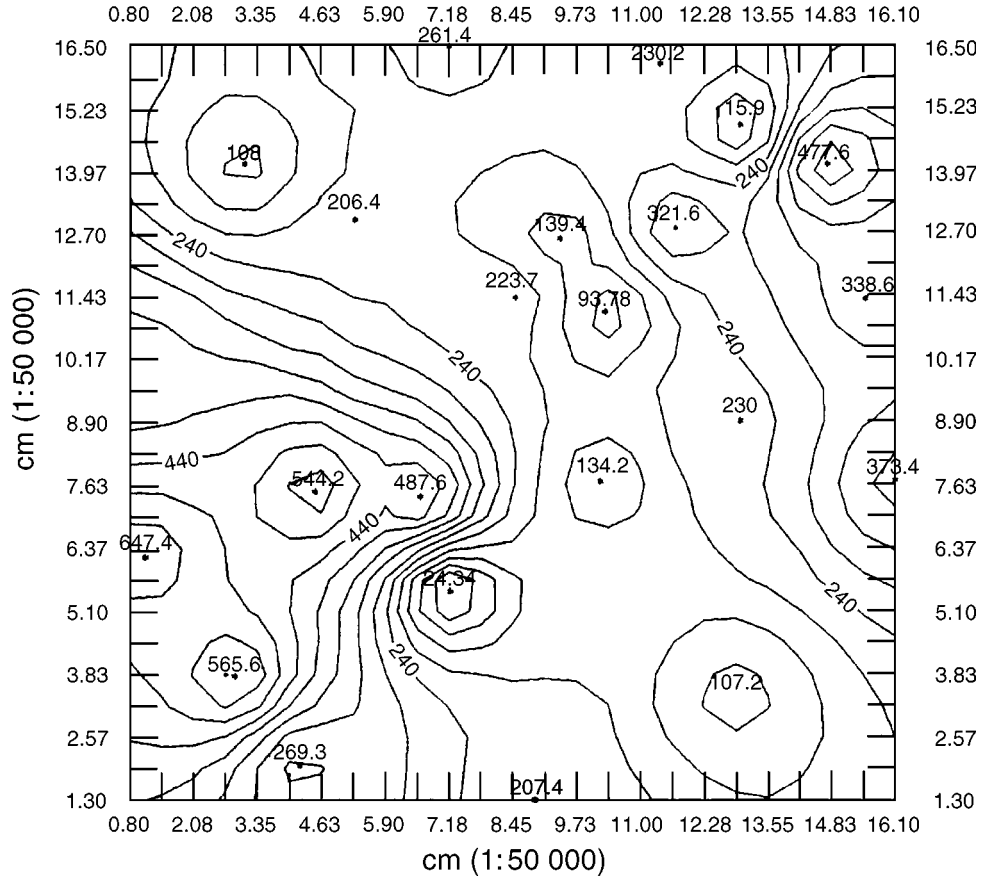

Figure 14 Isopleths of lead levels in dusts collected around an incinerator. (Reproduced with permission from John Wiley & Sons.)

is the observation that the highest levels occur in coarse dust fractions, suggesting that the incinerator is not the source of the metal in these samples, which were collected at greater distances from the stack. These results suggest that on the basis of pollutant levels, particulate matter released from the stack of the incinerator contributes to cadmium contamination of dust on the ground, although the contribution appears relatively small, with other pollution sources predominating. No impact on lead levels is indicated.

However, following the earlier logic, to evaluate the quality of environment it is necessary to consider the quantities of both metals and dust. A parallel for this is in relation to emissions to atmosphere, for which an assessment of mass emission may be more meaningful than the use of concentrations alone. Dust is only one environmental medium in which metals may be found, but in many respects it is unique. It is neither like air, which disperses widely in space, nor like water, in which dispersion within rivers, lakes and seas is possible. Since dust does not disperse uniformly, metals contained within the dust matrix cannot be expected to spread uniformly, even in the absence of a local source of pollution. A representation of the dust loading is given in Figure 15. The loading of a metal pollutant is a parameter to describe the quality of environment concerned with dust spread.

By comparing the spatial distributions, it was evident that the loadings of cadmium closely matched the locations for the highest loading of dust, whereas the highest lead loading was found at only one of these locations. The highest Cd levels are to be found to the northeast of the incinerator in the line of the prevailing wind direction, as discussed above. However the sites showing the highest Cd loading are located to the southwest of the incinerator, where great amounts of dust are deposited on the ground in the area of a discussed hospital. This is evident on comparing Figures 12 and 15. The highest Pb loading is northwest of the incinerator, at a car park and near a road junction. It was present in a great amount of fine deposited dust indicated in Figure 15. While some high Pb level sites are found elsewhere, the Pb loading is low because the amount of deposited dust is relatively low. The significance of these observations is that they demonstrate the care that is necessary in representing and interpreting data on environmental quality.

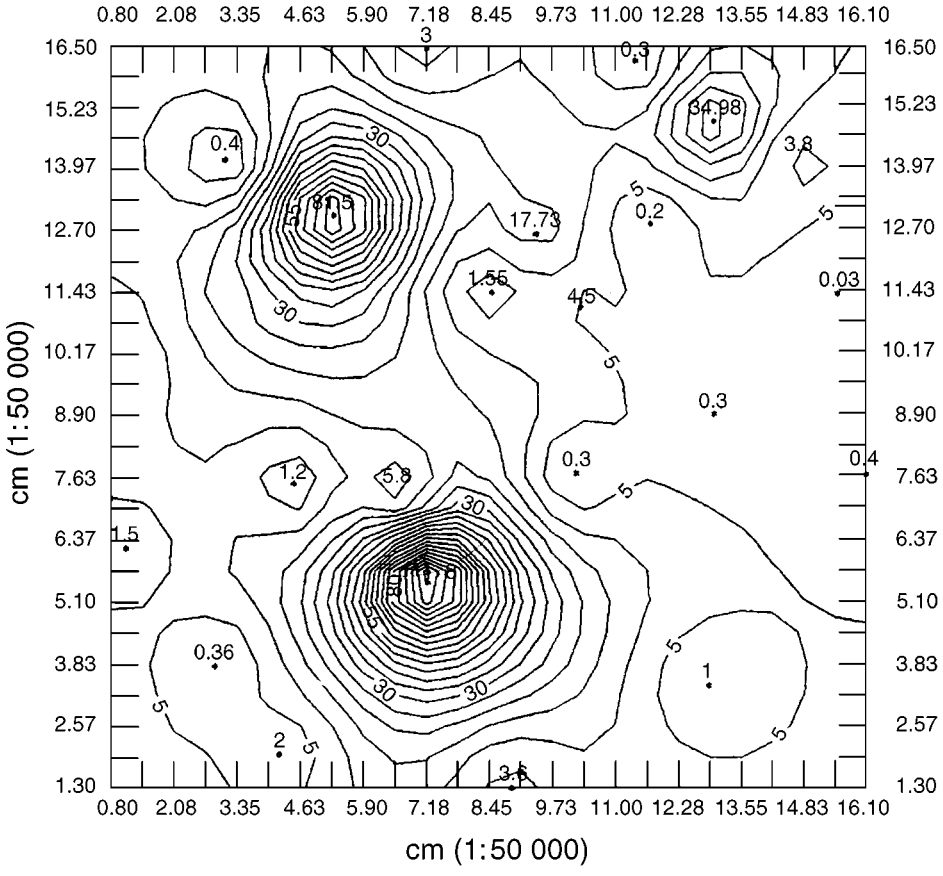

Figure 15 Dust loading around an incinerator. (Reproduced by permission from John Wiley & Sons.)

4.1 Relating Pollutants in Dusts to Sources

When applying dust data to environmental management, there is clearly a need to identify the source of the particulate matter. As the example of the incinerator demonstrated, simply identifying the presence of a contaminant does not associate it with the source.

Typical procedures involve:

- determining the topographic distribution to establish concentration profiles that relate sources and exposure points; this is similar to the approach around the incinerator;
- determining the relationships between elements, one of which may be a "tracer" characteristic of a specific source;
- determination of isotopic ratios to identify contributions from sources with different isotopic ratios.

One scheme for identifying further distinguishing features involves a sequence of steps (Figure 16).[54] Scanning electron microscopic examination and energy-dispersive X-ray analysis of individual particles in each subsample follow the fractionation of each dust sample according to particle size, density and ferromagnetic characteristics. Figures 2 and 3 illustrated the results from

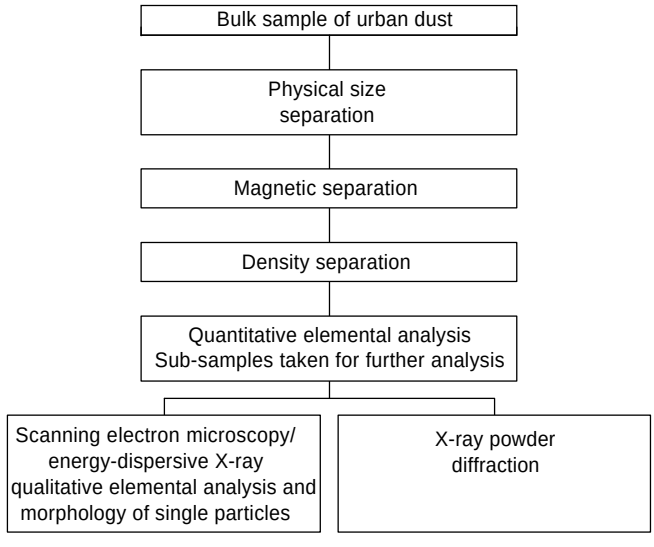

Figure 16 Sequence for fractionation of dust samples to identify sources. (Reproduced by permission from John Wiley & Sons.)

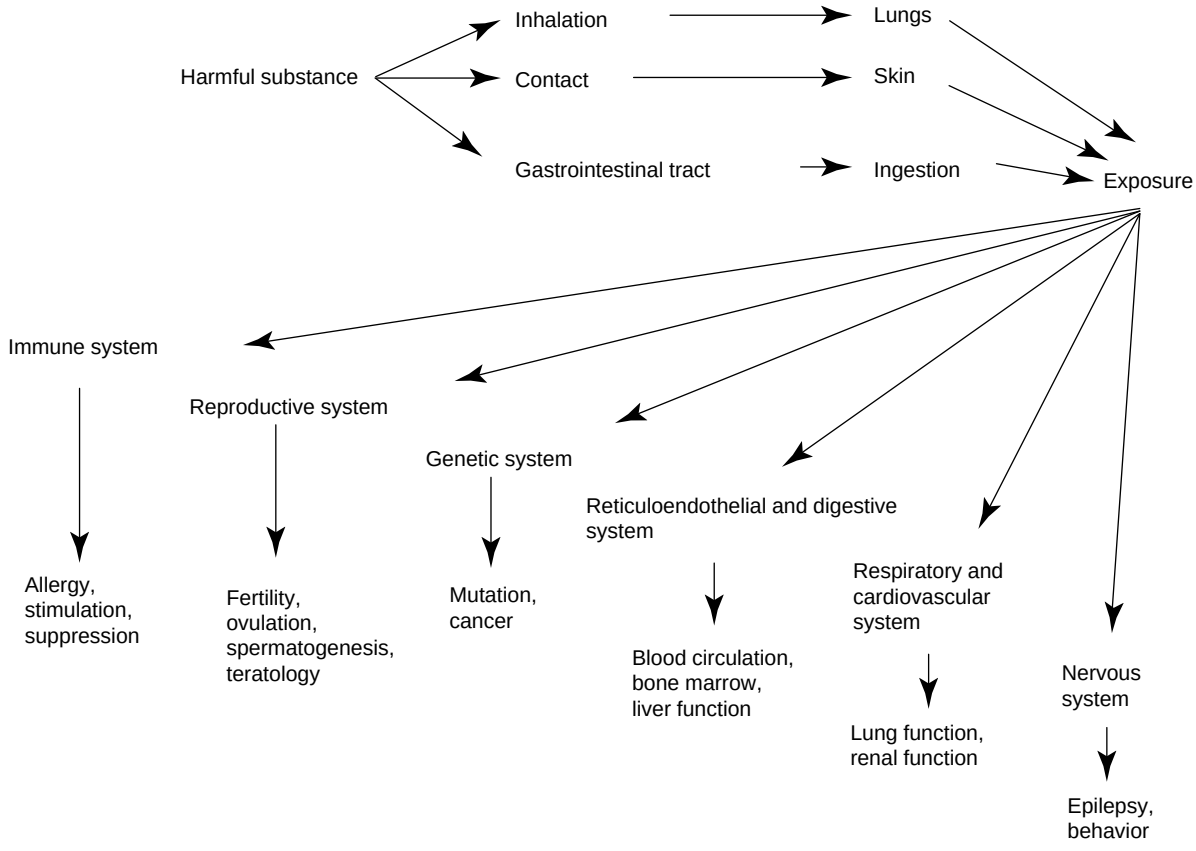

Figure 17 The potential of chemical substances in dust to cause harm. (Reproduced by permission from John Wiley & Sons.)

such analyses. Hence the morphological and chemical characteristics of each sample to be determined can be related to those observed in the particulate matter from suspect sources. Associations of elements in specific particles may suggest the origin of the particle. For example, particles containing lead and bromine can be linked to car exhaust emissions,[55] and lead to silicon ratios linked to soil, while lead to potassium can be linked to refuse burning.[56] Other techniques for source characterization include the use of stable isotope ratio.[57]

4.2 Is the Environmental Quality of Dust Acceptable?

In terms of cause and effect, an incident involving the release of a potentially harmful substance into the environment can be represented as

$$\text{event} \longrightarrow \text{impact on receptor} \longrightarrow \text{consequences or effects}$$

Land contamination, harm and loss of human life represent potential consequences at a site where the release has an impact. The cross-boundary dimension is represented by dispersion of particulate matter into the atmosphere, and subsequent deposition. The effects of an incident will depend on many factors including primarily the nature of the substance(s) concerned, its toxicity or other harmful effects, its pathways through the environment and its fate in the human body or other receptor. Clearly, biological responses depend on the duration of exposure and the ability of the substance(s) to reach target organs of the human body, for example, and all of these considerations are outside the scope of this article, which is concerned primarily with analytical methods for metals in dusts. However, some routes and their effects are shown in Figure 17. Clearly, in dealing with deposited dusts, this review is concerned with only a limited number of these routes of entry, and further is confined to human health effects. However, it would be remiss if awareness of the wider implications of dust contamination were not pointed out. Inhalation is an obvious route of entry as dust is transported through the atmosphere on its way to deposition. For airborne dusts there are well-established standards for health protection.

The lack of nationally agreed reference levels for judging the acceptability of trace metal concentrations in dust has been noted.[58] Standards on pollution by trace metals in dust do not exist, although in their absence, tentative guidelines have been used by organizations

such as the former Greater London Council. A guideline concentration for lead was adopted as $500\,\mu g\,g^{-1}$ to indicate the need for further investigation, while the level of $5000\,\mu g\,g^{-1}$ indicated a need for control action. The derivation and use of these guidelines are discussed further elsewhere.[59]

A second proposal for guidelines for both soils and dusts starts from the assumption that a single value is inappropriate.[60] Thus, a guideline in an area deemed acceptable in an industrial area would be unacceptable in a school playground. Blood lead is taken as the measure of tolerability and the guideline is derived from the model (Equation 1):

$$S = \frac{|T/G^n - B|}{\delta} \times 1000 \qquad (1)$$

where S = the guideline, a geometric mean concentration (micrograms of lead per gram of dust or parts per million), T = blood lead guideline (micrograms of lead per deciliter of whole blood), G = geometric standard deviation of blood lead distribution, typically in the range 1.3–1.5, B = the background or baseline blood lead concentration in the population from sources other than soil or dust, n = the number of standard deviations corresponding to the protection required for the population and δ = the slope of the response of the blood lead to dust lead relationship, with units of micrograms of lead per deciliter of blood increase per 1000 ppm increment of dust lead.

Examples of guidelines are shown in Table 17, although Wixson and Davies[60] should be consulted for more detailed examples.

Deposited dusts will always present a potential route of contamination by trace metals and organic substances and the need for future work in this area will not diminish. Only by the application of good quality control, reference materials and appropriate techniques for decomposition of the matrix to match the objectives of the study can data be viewed with confidence and compared with such guidelines as currently exist.

Table 17 Guidelines for dust lead according to target blood lead concentrations and the degree of protection required

Target blood lead ($\mu g\,dL^{-1}$)	Dust lead standard[a] for % of the population < target lead in blood			
	95%	98%	99%	99.9%
10	880	500	300	–
20	3750	3000	2600	1600
25	5200	4250	3700	2500

[a] Assumptions: $\delta = 2$; background blood lead = $4\,\mu g\,dL^{-1}$; geometric standard deviation = 1.4.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
CRM	Certified Reference Material
DPASV	Differential-pulse Anodic Stripping Voltammetry
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPES	Inductively Coupled Plasma Emission Spectroscopy
IUPAC	International Union of Pure and Applied Chemistry
LGC	Laboratory of the Government Chemist
LOD	Limit of Detection
NA	Neutron Activation
PAH	Polycyclic Aromatic Hydrocarbon
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Soil Sampling for the Characterization of Hazardous Waste Sites

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air

Electroanalytical Methods **(Volume 11)**
Pulse Voltammetry

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis

REFERENCES

1. *The Residential Lead-Based Paint Hazard Reduction Act of 1992*, US Congress, Washington, DC, 1992.
2. L. Johnson, P.J. Barlow, R.S. Barratt, 'Paint lead levels in a restricted environment', *J. R. Soc. Health*, **104**, 96–98 (1984).
3. *Glossary of Atmospheric Chemistry Terms*, IUPAC, 1989, Oxford, 1989.
4. *Clean Air (Emission of Grit and Dust from Furnaces) Regulations*, 1971, SI 1971, HMSO, London, 1971.
5. J.E. Fergusson, R.J. Schroeder, 'Lead in House Dust of Christchurch, New Zealand: Sampling, Levels and Sources', *Sci. Total Environ.*, **46**, 61–72 (1985).
6. J.E. Fergusson, D.E. Ryan, 'The Elemental Composition of Street Dust From Large and Small Urban Areas

Related to City Type, Source and Particle Size', *Sci. Total Environ.*, **34**, 101–116 (1984).
7. I. Thornton, 'Metal Content of Soils and Dusts', *Sci. Total Environ.*, **75**, 21–29 (1988).
8. R.E. Pitt, G. Amy, 'Toxic Materials Analysis of Street Surface Contaminants', Report EPA-R2-73-283, USEPA, Washington, DC, 1973.
9. R.M. Harrison, D.J.T. Smith, L. Luhana, 'Source Apportionment of Atmospheric Polycyclic Aromatic Hydrocarbons Collected from an Urban Location in Birmingham UK', *Environ. Sci. Technol.*, **30**, 825–832 (1996).
10. P.C. Elwood, in *Toxic Hazard Assessment of Chemicals*, ed. M.L. Richardson, Royal Society of Chemistry, London, 1986.
11. J.P. Day, M. Hart, M.S. Robinson, 'Lead in Urban Street Dust', *Nature*, **35**, 343–345 (1975).
12. A. Archer, R.S. Barratt, 'Lead Levels in Birmingham Dust', *Sci. Total Environ.*, **6**, 275–285 (1976).
13. M.J.R. Schwar, J.S. Moorcroft, D.P.H. Laxen, M. Thompson, C. Armorgie, 'Baseline Metal in Dust Concentrations in Greater London', *Sci. Total Environ.*, **68**, 25–43 (1988).
14. M. Hutton, A. Wadge, P.J. Milligan, 'Environmental Levels of Cadmium and Lead in the Vicinity of a Major Refuse Incinerator', *Atmos. Environ.*, **22**, 411–416 (1988).
15. M. Grottker, 'Runoff Quality From a Street With Medium Traffic Loading', *Sci. Total Environ.*, **59**, 457–466 (1987).
16. R.S. Hamilton, D.M. Revitt, R.S. Warren, 'Level and Physico-chemical Associations of Cd, Cu, Pb and Zn in Road Sediments', *Sci. Total Environ.*, **33**, 59–74 (1984).
17. M.J. Duggan, 'Temporal and Spatial Variations of Lead in Air and in Surface Dust–Implications for Monitoring', *Sci. Total Environ.*, **33**, 37–48 (1984).
18. S.E. Allen (ed.), *Chemical Analysis of Ecological Materials*, Blackwell, Oxford, 1974.
19. D.F. Ball, W.M. Williams, 'Variability of Soil Chemical Properties in Two Uncultivated Brown Earths', *J. Soil Sci.*, **19**, 379–391 (1968).
20. M.J. Duggan, 'Lead in Urban Dust: an Assessment', *Water, Air Soil Poll.*, **14**, 309–321 (1980).
21. M.T.R. Schwar, D.J. Alexander, 'Redecoration of External Leaded Paintwork and Lead in Dust Concentrations in School Playgrounds', *Sci. Total Environ.*, **68**, 45–59 (1988).
22. S.A. Rundle, M.J. Duggan, 'Lead Pollution From the External Redecoration of Old Buildings', *Sci. Total Environ.*, **57**, 181–190 (1986).
23. D.J.A. Davies, J.M. Watt, I. Thornton, 'Lead Levels in Birmingham Dusts and Soils', *Sci. Total Environ.*, **67**, 177–185 (1987).
24. M.J.R. Schwar, J.S. Moorcroft, D.P.H. Laxen, M. Thompson, C. Armorgie, 'Baseline Metal in Dust Concentrations in Greater London', *Sci. Total Environ.*, **68**, 25 (1988).
25. 'BS 1747: Method for the Measurement of Air Pollution. Part I. Deposit Gauges', British Standards Institution, London, 1969.
26. H.W. Vallack, D.E. Shillito, 'Suggested Guidelines for Deposited Ambient Dust', *Atmos. Environ.*, **32**, 2737–2744 (1998).
27. 'BS 893: Method for the Measurement of the Concentration of Particulate Material in Ducts carrying Gases', British Standards Institution, London, 1978.
28. 'BS 3405: Method for the Measurement of Particulate Emissions Containing Grit and Dust (Simplified Method)', British Standards Institution, London, 1983.
29. J. Watt, S. Moorcroft, K. Brooks, E. Culbard, I. Thornton, 'Metal Contamination of Dusts and Soils in Urban and Rural Households in the United Kingdom 1: Sampling and Analytical Techniques for Household and External Dusts', *Trace Subs. Environ. Health*, **22**, 229–235 (1983).
30. A.S. Al-Radady, B.E. Davies, M.J. French, 'Distribution of Lead Inside the Home: Case Studies in the North of England', *Sci. Total Environ.*, **145**, 143–156 (1994).
31. R. Reeves, T. Kjellström, M. Dallow, P. Mullins, 'Analysis of Lead in Blood, Paint, Soil and House Dust for the Assessment of Human Lead Exposure in Aukland', *NZ J of Science*, **25**, 221–227 (1982).
32. J.E. Fergusson, R.J. Schroeder, 'Lead in House Dust of Christchurch, New Zealand: Sampling, Levels and Sources', *Sci. Total Environ.*, **46**, 61 (1985).
33. S.S.Q. Hee, B. Peace, C.S. Clark, J.R. Boyle, R.L. Bornschein, P.B. Hammond, 'Evolution of Efficient Methods to Sample Lead Sources Such as House Dust and Hand Dust in the Homes of Children', *Environ. Res.*, **38**, 77–95 (1985).
34. C.R. Angle, A. Marcus, I-H. Cheng, M.S. McIntire, 'Omaha Childhood Blood Lead and Environmental Lead', *Environ. Res.*, **35**, 160–170 (1984).
35. I. Thornton, E. Culbard, S. Moorcroft, J. Watt, M. Wheatley, M. Thompson, J.F.A. Thomas, 'Metal Content of Soil and Dusts', *Environ. Technol. Lett.*, **6**, 137–144 (1985).
36. G.M. Raab, D.P.H. Laxen, M. Fulton, 'Lead from Dust and Water as Exposure Sources for Children', *Env. Geochem. Health*, **9**, 80–85 (1987).
37. K. Jones, A. McDonald, 'The Efficiency of Different Methods of Extracting Lead From Street Dust', *Environ. Poll. (Series B)*, **6**, 133–143 (1983).
38. A. Tessier, P.G.C. Campbell, M. Bisson, 'Sequential Extraction Procedure for the Speciation of Particulate Trace Metals', *Anal. Chem.*, **51**, 844–851 (1979).
39. M.J. Gibson, J.G. Farmer, 'Chemical Partitioning of Trace Metal Contaminates in Urban Street Dirt', *Sci. Total Environ.*, **33**, 49–57 (1984).
40. Y. Feng, *Methodological Aspects of Analysing Pb and Cd in Ground Dust and the Application in the Investigation of Dust Contamination*, PhD Thesis, the Open University, Milton Keynes, UK, 1994.
41. J.G. Farmer, T.D.B. Lyon, 'Lead in Glasgow Street Dirt and Soil', *Sci. Total Environ.*, **8**, 89–93 (1977).
42. Y. Feng, R.S. Barratt, 'The Determination of Pb and Cd in Dust by DPASV', International Conference on

Electrochemistry, Energy and Environment, Karaikudi, India, 10–12 February, 1993.
43. F. Smith, B. Cousins, J. Bozic, W. Flora, 'The Acid Dissolution of Sulfide Mineral Samples Under Pressure in a Microwave Oven', *Anal. Chim. Acta*, **177**, 243–245 (1985).
44. P. Barrett, L.J. Davidowski, Jr, 'Microwave Oven-based Digestion Technique', *Anal. Chem.*, **50**(7), 1021–1023 (1978).
45. G.M. Kimber, S. Kokot, 'Practical Options for Microwave Digestions', *Trends Anal. Chem.*, **9**(6), 203–207 (1990).
46. Y. Feng, R.S. Barratt, 'Digestion of Dust Samples in a Microwave Oven', *Sci. Total Environ.*, **143**, 157–161 (1994).
47. Doc. Eur 5730c, Levels of Pollution of the Environment by the Principal Pollutants, CEC, 1977.
48. R.S. Hamilton, D.M. Revitt, R.S. Warren, M.J. Duggan, 'Metal Mass Balance Studies Within a Small Highway Dominated Catchment', *Sci. Total Environ.*, **59**, 365–368 (1987).
49. Zdenek Sulcek, Pavel Povondra, *Methods of Decomposition in Inorganic Analysis*, CRC Press, Boca Raton, FL, 1989.
50. R.A. Nadkarni, 'Applications of Microwave Oven Sample Dissolution in Analysis', *Anal.Chem.*, **56**, 2233–2237 (1984).
51. M. Bettinelli, U. Baroni, N. Pastorelli, 'Analysis of Coal Ash and Environmental Materials by Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Atom. Spect.*, **2**, 485–489 (1987).
52. P.J. Lamothe, T.L. Fries, J.J. Consul, 'Evaluation of a Microwave Oven System for the Dissolution of Geologic Samples', *Anal. Chem.*, **58**, 1881–1886 (1986).
53. Y. Feng, R.S. Barratt, 'Effect of Potassium Iodide on Reducing the Adsorptive Interference of Surfactants and Organics in the Determination of Lead and Cadmium in Environmental Samples by Differential-pulse Anodic Stripping Voltammetry', *Analyst*, **119**, 2805–2808 (1994).
54. R.W. Linton, D.F.S. Natusch, R.L. Solomon, C.A. Evans, Jr, 'Physicochemical Characterization of Lead in Urban Dusts', *Environ. Sci. Technol.*, **14**, 159–164 (1980).
55. J.M. Watt, D.L. Johnson, 'Characterisation of Dusts by Scanning Electron Microscopy with Energy Dispersive X-ray Spectroscopy', in *Lead in the Home Environment*, eds. I. Thornton, E. Culbard, Science Reviews, Northwood UK, 85–96, 1987.
56. W. Van Borm, L. Wouters, R. Van Grieken, F. Adams, 'Lead Particles in an Urban Atmosphere: An Individual Particle Approach', *Sci. Total Environ.*, **90**, 55–66 (1990).
57. Y. Yaffe, C.P. Flessel, J.J. Wesolowski, A. Del Rosario, G.N. Guirguis, V. Matias, T.E. Degarmo, G.C. Coleman, J.W. Gramlich, W.R. Kelly, 'Identification of Lead Sources in California Children Using the Stable Isotope Ratio Technique', *Arch. Environ. Health*, **38**, 237–245 (1983).
58. R.S. Barratt, 'An Assessment of Dust Analyses', *Int. J. Environ. Anal. Chem.*, **40**, 77–97 (1990).
59. R.S. Barratt, Y. Feng, 'Dust, Measurement of Trace Elements in', in *Encyclopedia of Environmental Analysis and Remediation*, ed. Robert A. Meyers, John Wiley & Sons, Chichester, UK, 1421–1442, 1998.
60. B.G. Wixson, B.E. Davies, 'Guidelines for Lead in Soil', *Environ. Sci. Technol.*, **28**, 26A–31A (1994).

Elastic Recoil Detection Analysis

Patrick Trocellier
Commissariat à l'Énergie Atomique – Service de Recherches de Métallurgie Physique, Centre d'Études Nucléaires de Saclay, 91191 Gif sur Yvette, France

Timo Sajavaara
Department of Physics, University of Jyväskylä, Finland

1 Introduction **2**
2 Scattering Process and Conversion of Energy Spectrum to Depth Profile in Elastic Recoil Detection Analysis **2**
2.1 Kinematics 2
2.2 Cross Section 3
2.3 Energy–Depth Relationships 4
2.4 Recoil Yield 5
2.5 Depth Resolution 5
3 Conventional Elastic Recoil Spectrometry **5**
3.1 Glancing (Reflection) Geometry 5
3.2 Transmission Geometry 5
3.3 Data Processing 6
3.4 Standard Matrices and Ion-beam-induced Damage 6
3.5 Microbeam-induced Elastic Recoil Detection Analysis 7
4 Variant Recoil Methods **7**
4.1 $\Delta E-E$ Telescope 7
4.2 Time-of-flight Elastic Recoil Detection Analysis 8
4.3 Coincidence Spectrometry 9
4.4 Electromagnetic Cross Filters ($E \times B$) 10
5 High-energy and Heavy-ion Elastic Recoil Spectrometry **11**
5.1 High-energy ^{4}He-induced Elastic Recoil 11
5.2 High-energy Heavy-ion-induced Elastic Recoil 11
6 Application Examples of Elastic Recoil Detection Analysis for Profiling Hydrogen and other Elements **15**
6.1 Polymer Sciences 15
6.2 Semiconductor Materials 16
6.3 Thin Films 17
6.4 Interface Reactivity 17
6.5 Other Application Examples 18
7 Conclusion **18**
Acknowledgments **19**
Abbreviations and Acronyms **19**
Related Articles **19**
References **19**
Further Reading **24**

In 1976, a Canadian group described in detail for the first time a new ion beam analytical method based on the elastic recoil of target nuclei collided with high-energy heavy incident ions. In this case, 25–40-MeV ^{35}Cl impinged on a multilayer C or Cu (backing)/LiF or LiOH/Cu (30–150 nm)/LiF or LiOH and H, Li, O, and F recoiled atoms were detected. These exemplified the main characteristics of elastic recoil detection analysis (ERDA): its sensitivity to depth distribution and its ability to detect light elements in heavy substrates. In 1979, the use of megaelectronvolt energy ^{4}He beams permitted the use of ERDA to be extended to depth profiling of hydrogen isotopes in the near-surface region of solids.

ERDA has rapidly been revealed to be an excellent alternative to resonant nuclear reaction spectrometry (see **Nuclear Reaction Analysis***) for hydrogen determination in solids. Despite its less advantageous performance with respect to its lower depth resolution, lower analyzable depth, comparable sensitivity, and more restricting irradiation and detection geometry, some ERDA features have made its development in ion beam analysis (IBA) laboratories worldwide easier; these are simultaneous access to ^{1}H and ^{2}H depth distributions, access to single-ended Van de Graaff accelerators compared with tandem accelerators or cyclotrons, and the ability to be combined with Rutherford backscattering spectrometry (RBS) (see* **Rutherford Backscattering Spectroscopy***).*

The development of detection devices and the analytical capabilities offered by high-energy heavy-ion-induced ERDA in material sciences for profiling light, medium, and high mass number elements give this method a wide area in which to progress. The main advantage of heavy-ion ERDA and quite unique feature among analysis techniques is the fact that all sample elements can be depth profiled in one measurement by single detector telescope. By means of Monte Carlo (MC) simulations, the interpretation and reliability of the results have increased greatly over the last few years.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

1 INTRODUCTION

In 1976, L'Ecuyer et al.,[1] from the INRS-Energie Varennes (Québec), presented a paper that described in detail for the first time a new ion beam analytical method based on the elastic recoil of target nuclei collided with high-energy heavy incident ions, in this case, 25–40-MeV ^{35}Cl impinging on a multilayer C or Cu (backing)/LiF or LiOH/Cu (30–150 nm)/LiF or LiOH.[1] The work exhibited two of the main characteristics that established the technique of ERDA: sensitivity to depth distribution and the ability to detect light elements in heavy substrates.

The need to use high-energy heavy ions was a limitation to the development of this new analytical tool. In 1979, Doyle and Peercy published the first real example of the use of ERDA to determine hydrogen isotope depth profiles in the near-surface region of solids.[2]

At the same time, resonant nuclear reactions induced by energetic heavy ions were strongly promoted for the quantitative measurement of ^{1}H. ^{7}Li, ^{11}B, ^{15}N, and ^{19}F ion beams between 3 and 20 MeV were successfully shown to be very efficient for hydrogen depth profiling.[3–7]

ERDA has been revealed to be an excellent alternative to resonant nuclear reaction spectrometry for hydrogen determination in solids. Despite its less advantageous performance with respect to lower analyzable depth, comparable sensitivity, and more restricting irradiation and detection geometry, some ERDA features have made its use and development in IBA laboratories worldwide fast; these are simultaneous detection of not only ^{1}H and ^{2}H but also other elements, rapid development of detectors, and the great availability of small single-ended and tandem accelerators. As a result of detector development, especially in the case of TOF- and gas-ionization detectors, energies even below 10 MeV can now be used for high-resolution depth profiling.

In October 1996, Tirira, Serruys, and Trocellier, a group of French authors from Saclay published the first ion beam monograph entirely dedicated to ERDA.[8] This book has been used as the basis for the technical part of the present article and also as a bibliographical source.

In the following, we establish the theoretical basis of elastic recoil phenomenon, describe the experimental configurations generally adopted for elastic recoil spectrometry, and illustrate the analytical capabilities of ERDA for quantitative measurements in solids. Section 2 is devoted to general considerations of elastic collisions induced by megaelectronvolt ^{4}He and heavier incident ions in a hydrogenated material. Section 3 presents the practical characteristics of elastic recoil spectrometry in terms of irradiation and detection geometry, analytical performance, and data processing. Section 4 gives details of the main methodological variations from conventional ERDA using a single surface barrier detector and more recent experimental improvements. Section 5 is devoted to the use of high-energy ^{4}He ions and heavy ions for carrying out light-element profiling in solids. Section 6 contains some application examples of ERDA in a wide variety of fields from polymer science to semiconductor or thin-film technology.[9–11]

The concluding section of this article attempts to focus on the relevance of ERDA in comparison with resonant nuclear reaction spectrometry and secondary ion mass spectrometry (SIMS) (*see* **Secondary Ion Mass Spectrometry as Related to Surface Analysis**) for hydrogen determination in the near-surface region of solids. It also indicates the major trends in ERDA development since 1997.

2 SCATTERING PROCESS AND CONVERSION OF ENERGY SPECTRUM TO DEPTH PROFILE IN ELASTIC RECOIL DETECTION ANALYSIS

2.1 Kinematics

When an incident ion (mass m) with kinetic energy E_0 (a momentum P_0) interacts with a target atom (mass M) at rest through a purely elastic collision, the application of the conservation principles for total energy and momentum allows us to write Equations (1) and (2) (Figure 1):

$$E_0 = E + E' \tag{1}$$

$$P_0 = p + p' \tag{2}$$

with E, p, E', and p' the respective kinetic energies and momenta of the scattered incident ion and the recoiled target atom. Introducing θ and ϕ, the respective laboratory angles of scattering for m and M (Figure 1), we obtain Equations (3) and (4) given by

$$E' = K' E_0 \tag{3}$$

where

$$K' = \frac{4 m M \cos^2 \phi}{(m + M)^2} \tag{4}$$

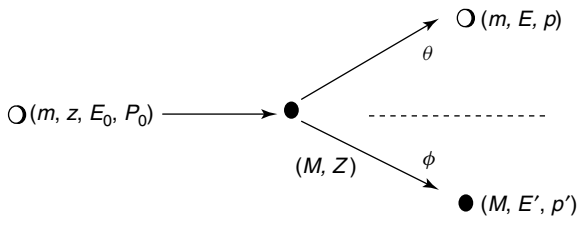

Figure 1 Schematic view of the elastic collision m/M.

and the angular relationship, Equation (5):

$$\tan\theta = \frac{\sin(2\phi)}{(m/M) - \cos(2\theta)} \tag{5}$$

If $m > M$, as for example when ^{1}H is the target atom and $^4He^+$ is the incident ion, the incident ion cannot be scattered from its incident direction through an angle greater than θ_{max}. This value is directly derived from the above equations, Equation (6):

$$\sin\theta_{max} = \frac{M}{m} \tag{6}$$

For a ^{4}He/^{1}H elastic collision, $\sin\theta_{max} \approx 0.26$ and $\theta_{max} = 14.47^\circ$. For a usual recoil angle $\phi = 30^\circ$, $K' = 0.48$, thus 2-MeV $^4He^+$ gives protons recoiling from the target surface an energy of 960 keV.

The kinematics of the elastic collision is described in detail in two IBA handbooks by Chu[12] and Tirira.[8]

2.2 Cross Section

2.2.1 Rutherford Scattering

The classical Rutherford formalism developed for a purely coulombian scattering leads to the well-known basic formula, Equation (7):

$$\left(\frac{d\sigma}{d\Omega}\right)^{cm} = \left[\frac{zZe^2}{4E_0^{cm}\sin^2(\xi/2)}\right]^2 \tag{7}$$

where $(d\sigma/d\Omega)^{cm}$ is the differential scattering cross section expressed in the center of mass system, z and Z are the respective atomic numbers of the incident ion and the target atom, e the elementary charge, E_0^{cm} the kinetic energy of the incident ion in the center of mass, and ξ the scattering angle in the center of mass.

In the laboratory system, Equation (7) becomes Equation (8):

$$\delta = \left(\frac{d\sigma}{d\Omega}\right)^{L} = \left(\frac{zZe^2}{4E_0}\right)^2 \left[\frac{m}{M} + 1\right]^2 \frac{1}{\cos^3\phi} \tag{8}$$

It is obvious that

- δ exhibits a minimum value for $\phi \approx 0$;
- δ tends to infinity when ϕ tends to $\pi/2$; in this case, the projectile/target interaction cannot be described by the Rutherford model;
- δ rapidly decreases as E_0 increases; and
- when $M \gg m$, δ is independent of the mass ratio.

In ERDA, Rutherford behavior is mainly observed for heavy-ion beams colliding with light target atoms (see Section 5) and for a ^{4}He ion beam below 1 MeV.

2.2.2 Non-Rutherford Cross Section

The Rutherford formula cannot be applied for elastic scattering/recoil spectrometry when the interaction potential between the incident ion and the target atom deviates from a pure coulomb potential. This can happen in two different cases. First, if the energy is high enough and secondly, in the low-energy case, if the two nuclei are screened by inner-shell electrons. The high energy at which elastic scattering/recoil cross sections begin to deviate from classic Rutherford behavior is called the *threshold energy*. It corresponds to the limit above which nuclear forces affect the interaction potential. Several authors have reviewed theoretical models combining a pure coulomb potential with a weak-perturbing nuclear contribution, as for example a Yukawa-like potential.[13] Experimental measurements of the threshold energy have been made for several projectile–target nucleus couples[14] and analytical formulations have been extracted.[15]

Using R_0, the distance of closest approach as a free parameter, the threshold energy is expressed by Equations (9) and (10):

$$E_{th} = \left(\frac{zZe^2}{2R_0}\right)\left(1 + \frac{m}{M}\right)\left(1 + \frac{1}{\cos\phi}\right) \tag{9}$$

where, for example,

$$R_0 = C_1 + C_2(m^{1/3} + M^{1/3}) \tag{10}$$

$C_1 = 6.003 \times 10^{-1}$m, $C_2 = 0.864 \times 10^{-15}$m, and $5.2 < (m^{1/3} + M^{1/3}) < 8.6$ (after Räisänen).[15] C_1 and C_2 are constants and m and M are the mass numbers. Applying this formalism, the threshold energy is 57.37 MeV for the couple ^{35}Cl/^{12}C and 469.60 MeV for the couple ^{127}I/^{16}O.[15]

In the case of heavy incident ions, the smaller cross sections due to the screening effect in low-energy scattering needs to be taken into account in the analysis. According to the analytical formula derived by Andersen et al.[16] the deviation from the pure Rutherford cross section is 21.7, 10.4, and 1.5% for Cu recoils detected at 40° and ejected by 10 MeV ^{197}Au, ^{127}I, and ^{35}Cl incident ions, respectively.

2.2.3 ^{4}He/^{1}H Collision

Historically, the scattering of protons by ^{4}He target nuclei was first investigated in the range 1–20 MeV for E_p. A phase shift analysis derived from quantum mechanics was thus applied to evaluate the cross section.[17,18]

The reverse collision can be treated by the same approach in the range 1–4 MeV for ^{4}He. The decomposition of wavefunctions in partial waves can be restricted

Table 1 Parameters obtained for A_l and Γ_l in the case of a $^4He/^1H$ elastic collision up to $E_{He} = 4\,MeV$

Partial wave	Scattering length A_l (fm)	Effective range Γ_l (fm)
$S_{1/2}$	3.35 ± 0.01	6.45 ± 0.01
$P_{1/2}$	-17.8 ± 0.8	-0.43 ± 0.05
$P_{3/2}$	-37.6 ± 1.5	-0.62 ± 0.03

to $S_{1/2}$, $P_{1/2}$, and $P_{3/2}$ contributions to describe the scattering of ^4He by ^1H. This method consists in determining nuclear phase shifts by using physical models such as the optical model or the *R*-matrix theory.[19–23] Tirira and Bodart have used experimental data below 4 MeV to extend previous phase shift studies.[24]

The best set of theoretical parameters (scattering length and effective range), derived by Tirira and Bodart[25] is presented in Table 1.

For ^4He energy values below 3 MeV and ϕ values below 30°, i.e. the most frequently used experimental conditions, several points with respect to the functional dependence of the non-Rutherford cross section $\delta_{nR} = (d\sigma/d\Omega)^L$ should be noted:

- δ_{nR} exhibits a minimum value for an energy E_{min} in the range 2.4–3 MeV for a recoil angle less than 30° ($E_{min} = 2.41\,MeV$ at 0°). As the angle increases from 0°, the cross-section minimum moves toward larger energy values.
- δ_{nR} has its lowest value at $\phi = 0^\circ$ for $E_{He} < E_{min}$.
- δ_{nR} has its highest value at 0° for $E_{He} > E_{min}$.
- δ_{nR} appears to reach the same constant value around $E_{He} = 3\,MeV$ for $0 < \phi < 40^\circ$.
- δ_{nR} varies very slightly near 0° with respect to E_{He}.
- δ_{nR} varies slightly in the $0–30^\circ$ range between 2 and 3 MeV.
- δ_{nR} shows a quasi-Rutherford behavior for $E_{He} <$ 1 MeV.

Comparing previously published data sets and their own experimental results, Tirira and Bodart derived a simple analytical expression allowing the elastic recoil differential cross section in the laboratory to be calculated for any $E_{He} - \phi$ couple, Equation (11)[25]:

$$\ln\left(\frac{d\sigma}{d\Omega}\right)^L = A_1 E_{He} + A_2 + A_3 E_{He}^{-1} + A_4 E_{He}^{-2} \qquad (11)$$

Table 2 contains different data sets for A_i values, for $0 \le \phi \le 40^\circ$ and $1 \le E_{He} \le 4\,MeV$.

Other binary collisions such as $^4He/^2H$, $^4He/^3H$, and $^3He/^1H$ have been treated in the same way.[26–30] A detailed discussion on this subject and the corresponding analytical expressions are given by Tirira et al.[8]

Table 2 Practical data sets for a rapid elastic recoil differential cross-section calculation[25]

ϕ ($^\circ$)	A_1	A_2	A_3	A_4
0	0.7651	1.7201	5.6116	−1.7011
5	0.7581	1.7321	5.6302	−1.7148
10	0.7366	1.7716	5.6797	−1.7527
15	0.6994	1.8492	5.7417	−1.8049
20	0.6449	1.9807	5.7890	−1.8568
25	0.5732	2.1840	5.7888	−1.8906
30	0.4779	2.4758	5.7117	−1.8897
35	0.3651	2.8682	5.5349	−1.8408
40	0.2349	3.3687	5.2445	−1.7350

From ref 25. © Elsevier 2006.

2.3 Energy–Depth Relationships

An incident ion with energy E_0 impinges on a target surface with an angle of incidence α with respect to the normal to the surface. Let us consider an elastic collision occurring at depth x beneath the target surface, the scattering direction makes an angle θ with the incident direction and the recoil angle is ϕ. The path of the incoming ion before collision is simply expressed by Equation (12):

$$L_{in} = \frac{x}{\cos\alpha} \qquad (12)$$

Its outgoing path is given by Equation (13):

$$L_{out} = \frac{x}{\cos\alpha} \qquad (13)$$

where the outgoing angle has been chosen to be equal to the incoming one. The path of the recoiled target atom is then given by Equation (14):

$$L_r = \frac{x}{\cos\beta} \qquad (14)$$

where $\beta = \pi - \phi - \alpha$.

The energy of the incident ion just before the collision is given by Equation (15):

$$E_0(x) = E_0 - \int_0^{x/\cos\alpha} S(E)\,dl \qquad (15)$$

where dl is the differential path. The energy of the scattered incident ion leaving the target is given by Equation (16):

$$E_1(x) = K E_0(x) - \int_0^{x/\cos\alpha} S(E)\,dl \qquad (16)$$

and the energy of the recoiled target atom is given by Equation (17):

$$E_2(x) = K' E_0(x) - \int_0^{x/\cos\beta} S_r(E)\,dl \qquad (17)$$

where $S(E)$ and $S_r(E)$ are the stopping powers for the incident ion and the recoiled target atom, respectively, in the target material considered.

2.4 Recoil Yield

For a thin target, the number of recoiled target atoms emitted in the depth range $x - x + \mathrm{d}x$ through an angle ϕ in the solid angle $\Delta\Omega$ is given by Equation (18):

$$\frac{\mathrm{d}N(x)}{\mathrm{d}x} = \frac{N_0 Q\, \mathrm{d}\sigma(E,\phi)}{\mathrm{d}\Omega\Delta\Omega} \tag{18}$$

where N_0 is the target atomic density and Q the total number of incident ions. The total recoil yield is then obtained by integrating Equation (18) on the thickness t of the target sample, considering the differential cross section as constant along t.

For a thick target, the analytical expression giving the recoil yield is slightly more complex because it has to take into account the variation of the differential cross sections with energy and all the energy spreading effects induced by particles traveling in matter.

2.5 Depth Resolution

Depth resolution δR_x strongly depends on total energy resolution of the system δE_T. This factor includes

- the detector energy resolution δE_d;
- the fluctuations in projectile energy δE_p;
- the energy spreading due to angular deviations δE_θ;
- the geometrical energy spreading δE_g;
- the energy straggling during incoming and outgoing paths δE_s;
- the multiple scattering contribution δE_{ms};
- the contribution from Doppler effect δE_D;
- the energy spreading due to the absorber δE_a; and
- the energy spreading due to surface roughness δE_{rou}.

A detailed description of the methods used for evaluating each contribution has been given by Tirira et al.[8]

3 CONVENTIONAL ELASTIC RECOIL SPECTROMETRY

3.1 Glancing (Reflection) Geometry

Figure 2 summarizes the experimental configuration called *glancing* or *reflection* geometry.

Several points should be noted:

- The beam spot size is enhanced by a factor $(1/\cos\alpha)$ with respect to normal incidence owing to the oblique incidence α.

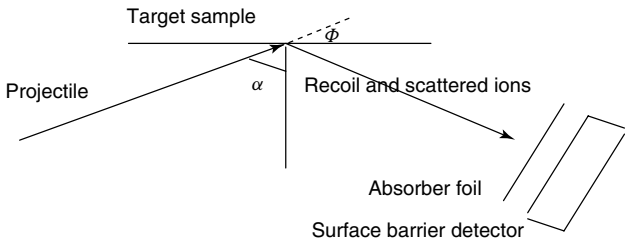

Figure 2 Schematic view of glancing geometry.

- The surface barrier detector is preceded by an absorber foil dedicated to stopping the ions scattered in the direction ϕ.
- The analyzable depth is rather limited owing to the glancing geometry. The range from the surface is obtained by multiplying the true incident ion range in the target by a factor $(\cos\alpha)$.
- This experimental configuration is absolutely prohibited with rough surface targets.

3.2 Transmission Geometry

Figure 3 summarizes the experimental configuration called *transmission* geometry.

Transmission geometry was the first experimental configuration used for ERDA measurements.[1,31–33]

Several points should be noted:

- The thickness of the target and the incident energy of the ^{4}He beam are strongly dependent, typically 20–25 μm for 3–3.5 MeV.[34]
- The presence of an absorber foil in front of the surface barrier detector is not necessary because the target itself slows down the scattered ions while they travel along their path until they are stopped.
- Practically, the surface barrier detector has to be placed as close as possible to the 0° direction, but not directly in view of the incident beam direction to avoid severe irreversible damage in the case of target breaking.
- The analyzable depth is only limited by the range of recoiling protons; it can reach up to 6–7 μm for polymer or silica-based materials.[34,35]

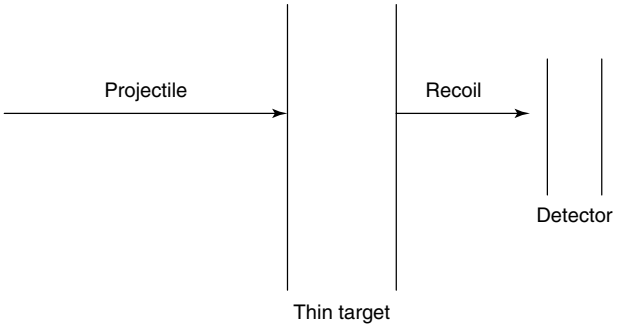

Figure 3 Schematic view of transmission geometry.

Table 3 Performances of ERDA hydrogen measurement

ERDA geometry	Analyzable depth (μm)	Depth resolution (nm)	
		Surface	Deep
Glancing	0.5–1	10	40
Transmission	6–8	30	25

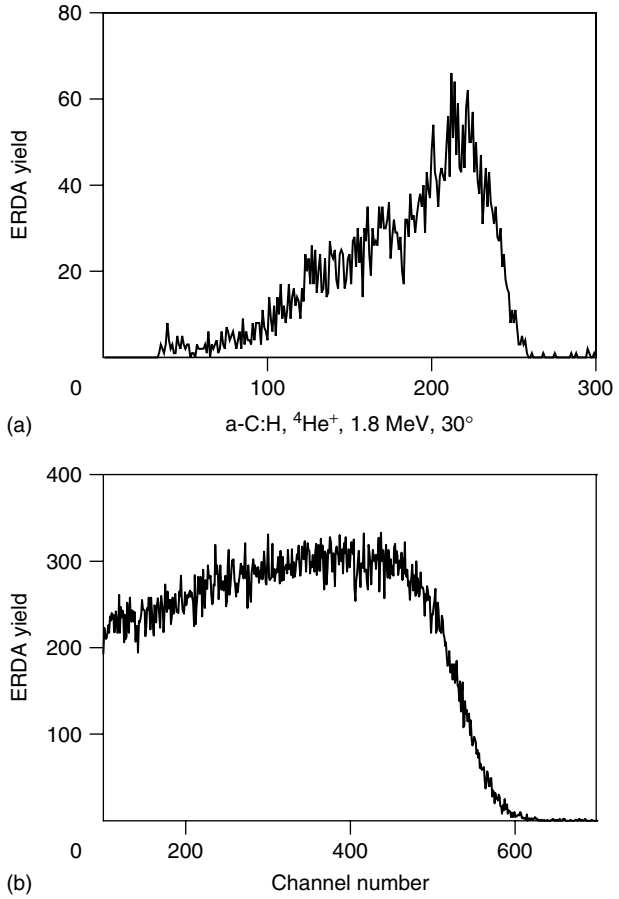

Figure 4 Examples of ERDA spectra obtained with a 1.8 or a 3.5-MeV ^{4}He ion millibeam. (a) Thin hydrogenated NiC/Ti multilayer coated on a silicon substrate in reflection geometry and (b) thick hydrogenated amorphous carbon layer in reflection geometry.

Table 3 permits comparison of the analytical performances offered by the two available geometries for hydrogen measurement.

Figure 4 gives two different examples of ERDA spectra obtained in glancing geometry for a hydrogenated NiC/Ti multilayer and a thick amorphous hydrogenated carbon layer.

3.3 Data Processing

Conventional ERDA data are treated in the same way as RBS, using a reconstitution-comparison-iteration computer program. The first step in the process is to define the experimental configuration (incident energy, geometry, detector characteristics, and target description). Then a simulated spectrum is built using the collision kinematic equation, the energy-loss formalism, the elastic recoil cross-section table, and straggling and multiple scattering subroutines.[36] Finally, the simulated spectrum is compared to the experimental data and adjusted in terms of elemental distributions, using a classical least-squares method until it matches correctly.

Several authors have proposed specific computer codes dedicated to ERDA interpretation.[36–38] Our purpose, in this section, is not to describe in detail any specific approach. The references listed above constitute excellent starting points to get to know classical IBA tools. An IAEA intercomparison[39] of some of the most used computer codes was published recently.

In the case of heavy-ion elastic recoil detection analysis (HI-ERDA), another approach known as *slab analysis* is often used instead of simulation and comparison. In the energy spectrum, a channel corresponds to a certain depth slice. The depth is calculated using scattering kinematics and either experimental or, more often, semiempirical stopping powers. The yield at different depths is then normalized using stopping powers and scattering cross sections for a given incident ion and sample atoms. If all the sample elements can be analyzed, the final depth profiles can be normalized to unity to obtain the atomic ratios. All the effects degrading depth resolution that are mentioned above are still visible in the depth profiles after slab analysis.

3.4 Standard Matrices and Ion-beam-induced Damage

IBA techniques are considered to be quantitative even without the use of reference samples provided that the differential cross section of the interaction observed is accurately known in the energy range investigated and also provided the experimental geometry (target and detector) is fully described.

However, the quantitative determination of any elemental depth distribution in a complex target material generally requires the use of a reference sample. This standard (or model sample) is investigated either to simplify the calculation procedure or to assess the profile extraction method. Moreover, finding a good hydrogen reference sample remains a particular problem owing to its high mobility in solids even under standard temperature conditions. The use of reference samples is highly recommended for accurate results in the case of HI-ERDA also; even in the case of thin films when all the elements present can be detected and concentrations normalized to unity.

Polymers would represent the most practical type of hydrogen reference samples because of their large H

content, their variety of formulation, their thickness range (from 3 μm up to several hundred micrometers) and their well-known core composition homogeneity. Nevertheless, there are two severe limitations:

- the thickness heterogeneity of thin films having thickness less than 20 μm and
- the sensitivity of polymer structure to ion beam bombardment.

Elemental losses are impossible to prevent and in practise polymers have to be replaced by hydrogen-implanted monocrystals (Si or Ge, for example). The versatility of implantation conditions (energy and fluence) is of major interest but conservation of the hydrogen profile for long periods of time is not guaranteed. The hydrogen mobility in silicon at 25 °C is far from negligible. Metal hydrides such as TiH_2 or Ti_xD_y (with $x = 1$ and $y \approx 1.8$) could be an alternative to implanted crystals,[40] but control of the stoichiometry is rather difficult. Hydrogenated amorphous carbon or silicon-carbide layers appear now to be a better compromise than hydride layers.[11]

The choice of a multielemental reference sample requires a complete description of the investigated sample. Routine methods such as electron microprobe analysis (*see* **Scanning Electron Microscopy in Analysis of Surfaces**) must be very efficient to ensure the lateral composition homogeneity of the target material. The most important criterion is to choose a reference material with a similar composition in order to be able to neglect corrections for stopping power. Ceramics, glasses, or thin layers are the most useful materials for IBA standards provided that their sensitivity to ion bombardment has been checked before use.

Several effects can be inferred by an incident ion beam impinging on a radiation-sensitive material: electron excitation, ionization, atomic displacement, bond breaking, formation of charge defect centers, temperature increase, growth of an electrical gradient, and surface sputtering.[38,41] The main consequences of these phenomena is the modification of the target composition and particularly the loss of certain elements of interest such as hydrogen, alkali metals, or halogen elements.[42] An extensive review of beam damage is presented by Tirira et al.[36] A comprehensive bibliography is given and concrete examples discussed.

3.5 Microbeam-induced Elastic Recoil Detection Analysis

Microbeam-induced elastic recoil detection analysis (μERDA) is the only routinely available technique for hydrogen determination using a nuclear microprobe. Experiments in transmission geometry are, in general, easier to carry out than experiments in reflection geometry.

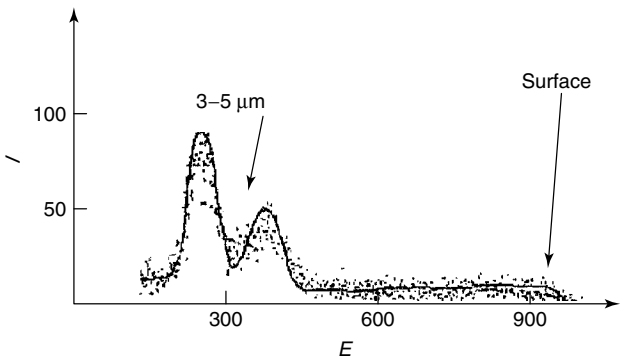

Figure 5 Thin polyimide foil (25 μm) investigated by μERDA in transmission geometry after high-energy heavy-ion irradiation.

Transmission geometry is used in polymer science applications or in geological applications[34,43–45] and reflection geometry is chosen for thin-layer characterization.[11]

Figure 5 gives an example of an ERDA spectrum obtained in transmission geometry on a thin polyimide film (25 μm) preirradiated with 50-MeV ^{32}S ions and measured with a 3.05-MeV $^4He^+$ microbeam (5×5 μm^2, 200 pA). The heavy-ion irradiation induces the formation of hydrogen bubbles in the core of the modified polymer structure. Two of these hydrogen-rich zones are found in the spectrum between 3 and 5 μm from the surface.

4 VARIANT RECOIL METHODS

The energy discrimination between a scattered incident ion and a recoiled target atom is one of the major problems to be solved in ERDA. It can be based on the difference of energy loss in a given thickness of a well-known material. When $m > M$ ($^4He/^1H$ collision), the scattered ^4He loses a higher fraction of its energy than the recoil ^1H reaching the detector. The use of an absorber foil (polymer or Al) in front of the detector allows discrimination by stopping the scattered events. This simple technique is currently the most adopted one for ERDA measurement (see Section 3) in transmission geometry, in which the target itself plays the role of the absorber foil. When $m < M$, this system is completely inefficient.

4.1 ΔE–E Telescope

The second way to solve the scattered/recoil event ambiguity is to replace the absorber foil by a thin detector (10–20 μm). The device constituted by a couple of detectors is called a solid-state $\Delta E - E$ telescope.[46,47]

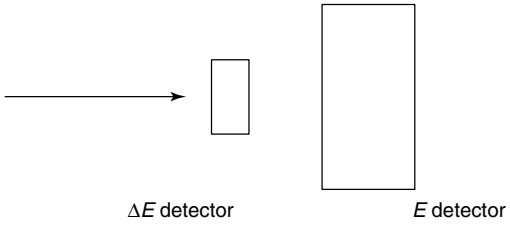

Figure 6 Schematic design of a $\Delta E-E$ detector.

Scattered ions are stopped in the ΔE detector and recoil events are detected in the E detector ($\geq 100\,\mu m$). Figure 6 is a scheme of such a detection device.

The solid ΔE detector can be replaced by a gas-ionization chamber.[47–49] Telescopes have been used for ERDA measurements with ^{4}He beams[46] as well as with heavy-ion beams.[50–53]

The mass and charge separation ability of a telescope are simply described by Equations (19) and (20):

$$\frac{\Delta M}{M} = \left[\left(\frac{\Delta E_R}{E_R}\right)^2 + \left(\frac{\Delta dE}{dE}\right)^2 + \left(\frac{\Delta dx}{dx}\right)^2\right]^{1/2} \quad (19)$$

$$\frac{\Delta Z}{Z} = \frac{1}{2}\left[\left(\frac{\Delta E_R}{E_R}\right)^2 + \left(\frac{\Delta dE}{dE}\right)^2 + \left(\frac{\Delta dx}{dx}\right)^2\right]^{1/2} \quad (20)$$

where dE is the energy loss in the thin detector and dx is the thickness of the thin detector.

The presence of an absorber foil in front of the detector in conventional ERDA induces a high straggling contribution in the total depth resolution. The use of a $\Delta E-E$ telescope eliminates this straggling contribution because the recoil energy can be simply deduced by summing the energy loss in the thin detector ΔE and the residual energy E_R. Only the straggling contribution from the detector dead layers cannot be suppressed. For example, a depth resolution better than 25 nm has been obtained for oxygen determination in native oxide layers using a 43-MeV ^{63}Cu beam.[51] Siegele et al. have obtained a depth resolution better than 20 nm on stainless steel with 130-MeV ^{127}I ions.[54] A value of 10 nm has been obtained by Assman.[55]

Figure 7 illustrates the increased selectivity offered by the use of a solid-state telescope for simultaneous detection of ^{1}H, ^{2}H, and ^{3}H with 4-MeV ^{4}He$^+$ in a titanium hydride layer.[8]

4.2 Time-of-flight Elastic Recoil Detection Analysis

The determination of mass separation between scattered and recoil ions by measuring the difference in their time of flight (TOF) for a given path length was first introduced in the mid-1980s.[56–58] The principle of TOF spectrometry is based on the measurement of particle

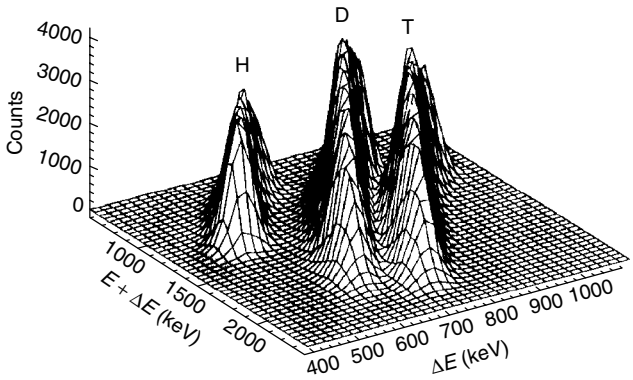

Figure 7 Three-dimensional plot of hydrogen, deuterium, and tritium distributions in a titanium hydride sample bombarded with 4-MeV helium ions. (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

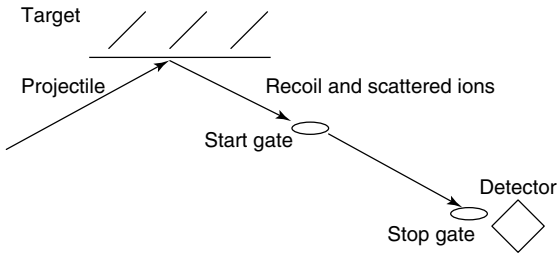

Figure 8 Schematic view of a TOF spectrometer.

flight time between a start gate and a stop gate giving two distinct signals. Simultaneously, the total energy of the particles is measured by means of a solid-state or gas-ionization detector. Figure 8 schematically describes a TOF spectrometer.

The flight time $t_f = t_2 - t_1$ of a particle with mass M and energy E_R traveling along a path of length L is given by Equation (21):

$$t_f = t_2 - t_1 = L\left(\frac{M}{2E_R}\right)^{1/2} \quad (21)$$

If the TOF spectrometer is set up with an entrance window of thickness x, Equation (20) becomes Equation (22):

$$t_f = L\left\{\frac{M}{2[E_R - S(E)x]}\right\}^{1/2} \quad (22)$$

where $S(E)$ is the stopping power of the window material for recoil ion M at energy E_R.

The time detector is generally an electrostatic mirror protected by a thin carbon foil ($0.5–20\,\mu g\,cm^{-2}$) so that the energy recorded by the surface barrier detector, which is also often used as the stop-signal detector, is given by Equation (23):

$$E = E_R - S(E)x \quad (23)$$

where x is the thickness of one or, in the case of two timing gates, two carbon foils. The carbon foil of the first electrostatic mirror serves as the entrance window for the TOF spectrometer. Typical flight time values are in the range 30–100 ns for a flight path of about 50 cm. Normally, the energy spectra used in the analysis is converted from timing signal. The two main reasons for this are (i) the normally better resolution for other elements than for H and He and (ii) the linear calibration for all the elements, which is not the case for a solid-state detector. A well-known limitation of the TOF-E technique is the poor detection efficiency (typically 10–60%) of timing gates for H, which is strongly stopping force and therefore, energy dependent. An elegant way to get around this limitation is to use the fact that for elements heavier than B, the detection efficiency is very close to 100%, and if the timing window is long enough, almost all noncoincident events seen only by the E detector are hydrogen. By summing up coincident events from the H curve and noncoincident events, 100% detection efficiency is obtained for H. This technique is described in detail in two round-robin characterizations where H was also studied.[59,60]

Figure 9 gives an example of the application of a TOF spectrometer for the ERDA characterization of a glass ceramic with an 84-MeV ^{127}I beam.[8] It can be seen that the mass resolution is less than 1 for low masses, and that ^{10}B and ^{11}B isotopes are clearly distinguishable. Nevertheless, mass resolution decreases with increasing mass so that ^{27}Al and ^{28}Si cannot be separated. This is mostly due to poor energy resolution of the solid-state detector. The mass resolution can be drastically improved by introducing a gas-ionization detector equipped with a thin SiN entrance window.[61] However, Al and Si can be readily separated with a $\Delta E-E$ gas detector as shown by Siegele et al.[54]

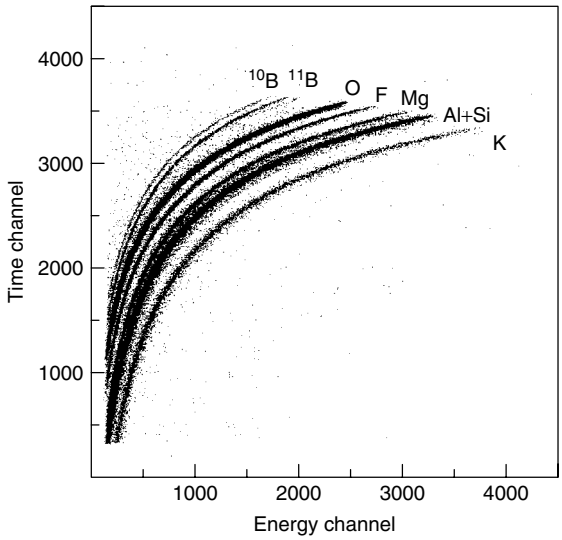

Figure 9 TOF spectrum obtained for a glass ceramic bombarded by a 84-MeV ^{127}I beam. (Reproduced with the kind permission of Plenum Press.)

A characteristic feature for TOF-ERDA is the fact that depth resolution at the surface improves when the energy of incident ion is lowered. This is due to the fact that the elemental energy spectra used in the analysis are converted from TOF spectra. Even with a simple TOF-E telescope, all elements below Ne can be separated from each other with projectile (Cl, Br, and I) energies below 20 MeV.

4.3 Coincidence Spectrometry

In the first ERDA experiments, from Cohen in 1972[31] to L'Ecuyer in 1976,[1] transmission geometry together with coincident detection of scattered and recoil particles were applied to achieve mass selectivity, to reduce the background, and to improve the sensitivity.

By detecting both particles, four parameters can easily be obtained: E_S, E_R, θ, and ϕ. Thus the mass of the recoil nucleus M and the interaction depth x can be deduced according to Figure 10. One variation of this technique, proton-proton scattering utilized by Reichart et al., allows sensitive 3D profiling of hydrogen down to ppm-levels.[62]

Two families of ERDA coincidence methods can be distinguished: those carefully adjusting θ and ϕ to achieve mass selectivity and to reduce the background, and those carefully measuring E_S and E_R to eliminate θ and ϕ.

Depth information is the ultimate goal of the first family of methods; this can be deduced from energy measurement. Large solid detectors are generally used in the second family of methods, and as a result, depth resolution is sacrificed. Figures 11 and 12 give one example of the application of each method.

Partial separation of ^{63}Cu and ^{65}Cu has been obtained by Klein et al.[63] using a pair of position-sensitive detectors at the GSI heavy-ion microprobe (Darmstadt,

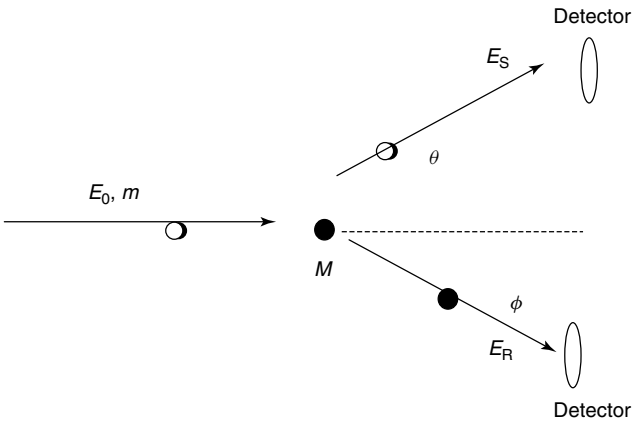

Figure 10 Typical coincidence configuration.

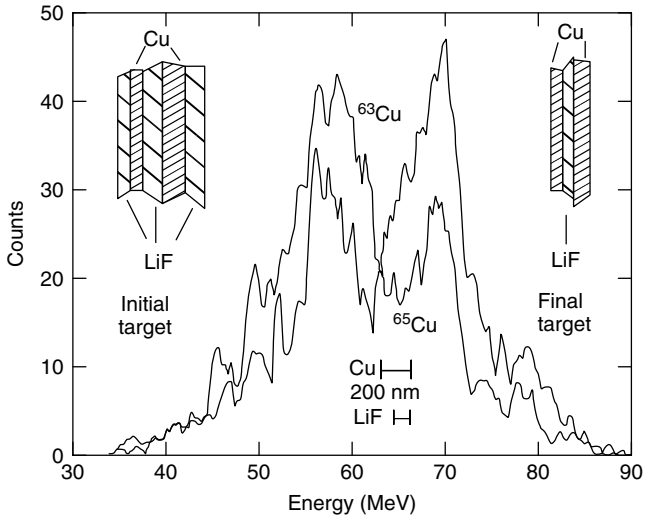

Figure 11 Coincident elastic recoil detection spectra of two thin Cu films separated by LiF measured with ^{58}Ni ions.[63] The final target is different from the initial one because the outer LiF layers have been partially decomposed during the irradiation. (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

Germany) on a sandwich target (LiF/Cu/LiF/Cu/LiF) with 1.4-MeV amu^{-1} ^{58}Ni ions. The dominant peak in Figure 12(a) corresponds to recoiled carbon atoms; recoiled oxygen is seen in the spectrum at higher He energies. Figure 12(b) is similar to Figure 12(a), but rotated by 180°, the surface line corresponds to the scattering process at the rear surface of the target.

4.4 Electromagnetic Cross Filters ($E \times B$)

The Canadian group from Varennes was the first to propose the use of an electromagnetic cross filter as an achromatic mass and charge selector and not only as a Wien filter.[65,66] A typical $E \times B$ (electric and magnetic crossed fields) device for ERDA measurement is shown in Figure 13.

When a particle passes through the $E \times B$ filter, Equations (24) and (25) describe its exact movement:

$$m\frac{dv_y}{dt} = q(E - v_x B) \tag{24}$$

$$m\frac{dv_x}{dt} = q_{v_y} B \tag{25}$$

Assuming $v_x = v_0$ and $v_y = 0$, they reduce to Equation (26):

$$m\frac{dv_y}{dt} = q(E - v_0 B) \tag{26}$$

At the exit of the filter, the particle is deviated by an angle α given by Equation (27):

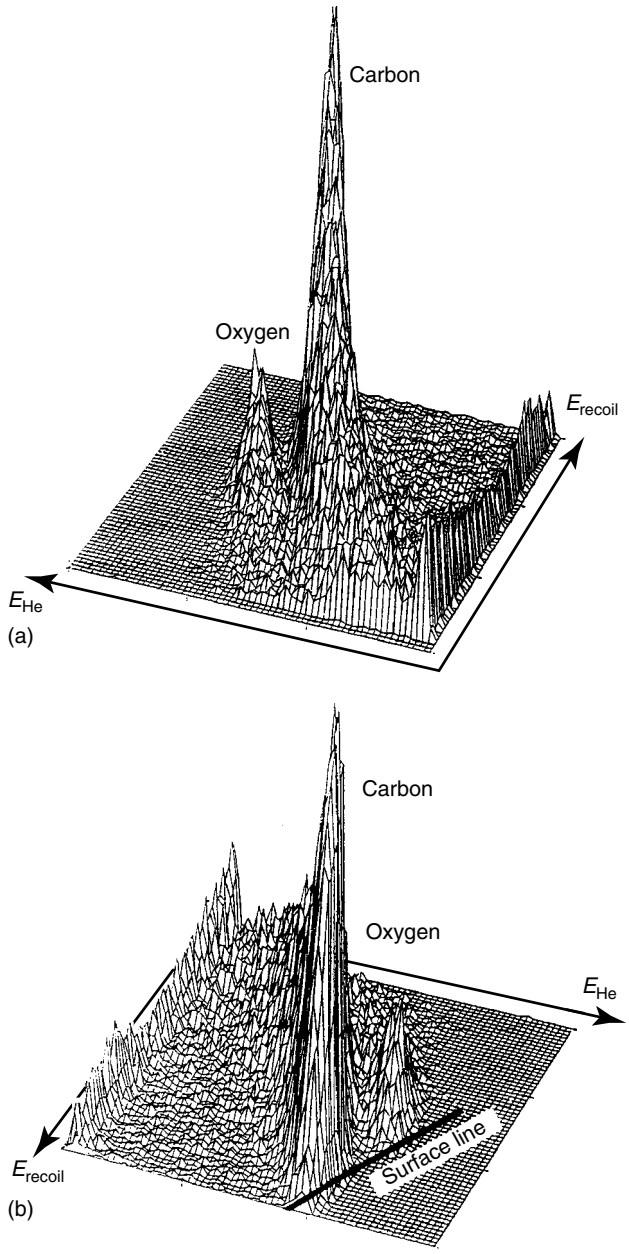

Figure 12 Coincidence recoil spectrum of a free-standing polycarbonate film obtained using a ^{4}He beam.[64] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

$$\sin\alpha \approx \alpha = \frac{qL(E - v_0 B)}{mv_0^2} \tag{27}$$

where L is the total length of the $E \times B$ filter.

The filter should deflect particles of all energies by the same angle; this condition is satisfied when Equation (28) applies:

$$\frac{d\alpha}{dE} = 0 \quad \text{or} \quad E = \frac{Bv_0}{2} \tag{28}$$

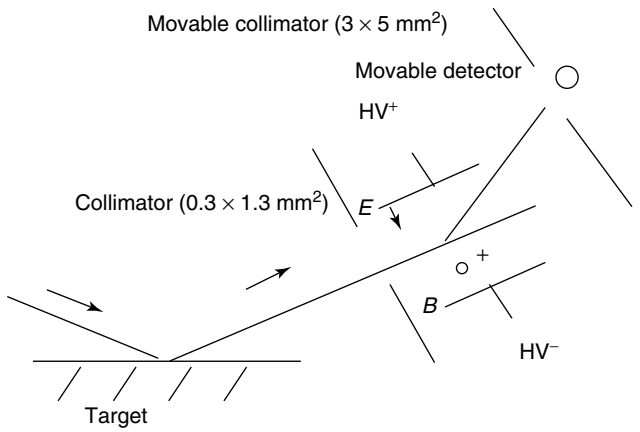

Figure 13 Schematic view of a $E \times B$ device used in ERDA.

Typical values for L, B, and E are, respectively, 70–80 mm, 1 kG, and 1.5–2 kV cm^{-1}.

Electromagnetic cross filters have been evolving since 1995 and more sophisticated devices have appeared such as the $B - E \times B - B$ filter proposed by Serruys, which has been described previously.[8]

The Canadian group from INRS Varennes and A. Chevarier's group in France (Lyon University) have collaborated together on the use of the $E \times B$ filter to characterize amorphous hydrogenated layers. They have shown that hydrogen desorption induced by 2.5-MeV $^4He^+$ with or without previous deuterium implantation was smaller for a-SiC:H than for a-C:H layers.[11] Figure 14 compares the hydrogen distributions in a-SiC:H and a-C:H layers after a deuterium implantation at a level between 2.2 and 3.5×10^{16} D cm^{-2}.

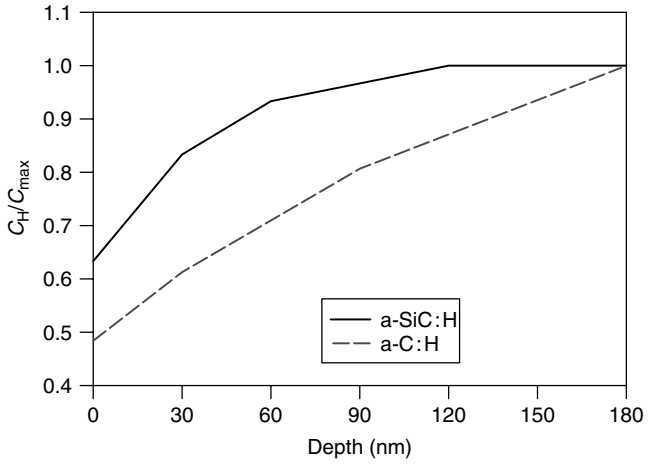

Figure 14 Comparison of the hydrogen distributions in a-SiC:H and a-C:H layers after a deuterium implantation between 2.2 and 3.5×10^{16} D cm^{-2} measured with a 2.5-MeV ^{4}He beam.

5 HIGH-ENERGY AND HEAVY-ION ELASTIC RECOIL SPECTROMETRY

As discussed in Section 2, low-energy ^{4}He (2–4 MeV)-induced ERDA is adapted to profile target nuclei lighter than the projectile: generally ^{1}H and ^{2}H. Using a specific coincidence detection device allows the researcher to extend the profiled mass range to target nuclei heavier than the projectile (see Section 3). There are two other means to improve the range of analyzable elements and the total explored depth. The first consists in increasing the energy of the incident ^{4}He beam to improve the energy separation between the contributions of the different target components and also to take advantage of the increase in the non-Rutherford cross section for light elements.[67,68] The second consists in increasing both the mass and the energy of the incident ion beam; this technique is called *high-energy HI-ERDA*.[48,69–71] The selectivity is thus largely improved; this was the configuration adopted by L'Ecuyer et al. in 1976 to investigate multilayered structures.[1]

5.1 High-energy ^{4}He-induced Elastic Recoil

When ^{4}He incident energy increases above 1 MeV, Rutherford scattering is no longer an applicable model. High-energy ^{4}He scattering in light elements has a non-Rutherford behavior and cross-section resonances exist for several couples ^{4}He/target nucleus, as shown in Figure 15 for carbon and oxygen.

Figure 16 shows the ERDA spectrum for ^{12}C recoil of an Al/C sandwich target, obtained using a 30 MeV α-particle beam in transmission geometry.

5.2 High-energy Heavy-ion-induced Elastic Recoil

Equation (4), describing the kinematics of the collision m/M, with $m \gg M$ becomes Equation (29):

$$E' = 4E_0 \cos^2 \phi \frac{M}{m} \tag{29}$$

The interaction cross section follows Rutherford behavior as long as the incident energy is below the threshold energy defined by Equation (9).

The analytical expression of the differential recoil cross section given by Equation (8) becomes Equation (30):

$$\frac{d\sigma}{d\Omega} = \left(\frac{zZe^2}{2E_0}\right)^2 \left[\frac{(m/M)^2}{\cos^3 \phi}\right] \tag{30}$$

Several points should be noted:

- The differential cross section is nearly independent of the recoil mass because the (Z/M) ratio ranges

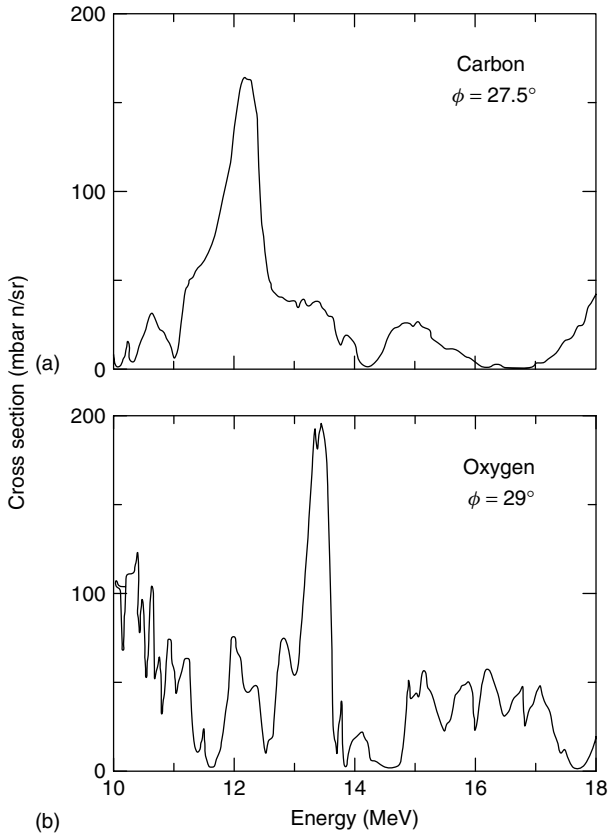

Figure 15 Scattering cross section for (a) $^4He/^{12}C$ and (b) $^4He/^{16}O$ collisions between 10 and 18 MeV.[72–74] (Reproduced by courtesy of H. Rijken and with the kind permission of Plenum Press.)

between 0.4 and 0.5 for all the elements except hydrogen.

- For hydrogen recoils $(Z/M) = 1$ and the differential cross section is increased by a factor of four relative to the other recoiled target atoms.
- The sensitivity of high-energy HI-ERDA increases strongly with increasing z. Anyhow, if a sophisticated detection system like $\Delta E - E$ or TOF-E is used in the analysis, the high-z projectile does not necessarily provide the lowest detection limit, but that can be obtained with lighter projectiles giving more energy and therefore better separation in the 2-D histogram for the recoils.
- The selectivity of high-energy HI-ERDA decreases with increasing Z. For example, C, N, and O can be separated well in stainless steel using a 136-MeV ^{127}I beam,[75] but transition elements such as Cr, Fe, and Ni cannot be properly separated using 136-MeV ^{127}I and the Bragg counter as shown in Figure 17. Nevertheless, Siegele et al. have shown that this separation is possible at 136 MeV, using a telescope with a gas chamber as the ΔE detector.[76]

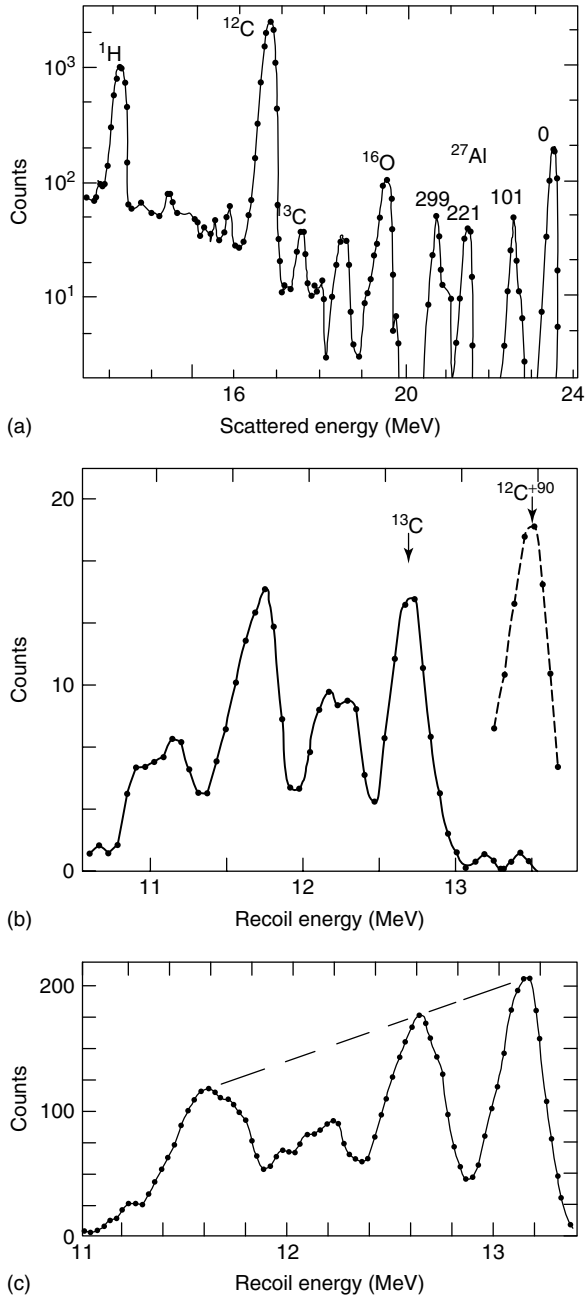

Figure 16 (a) Scattered ^4He spectrum, (b) ERDA spectrum for ^{13}C recoils, and (c) ERDA spectrum for ^{12}C recoils of an Al/C sandwich target, obtained using a 30-MeV α-particle beam in transmission geometry.[67] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

- The differential cross section increases with increasing recoil angle ϕ (factor 5 between 30 and 60°), but simultaneously the recoil energy is reduced. Generally, the detector system is located around 40°.

High-energy HI-ERDA requires specific detection devices such as those described in Section 4 (TOF

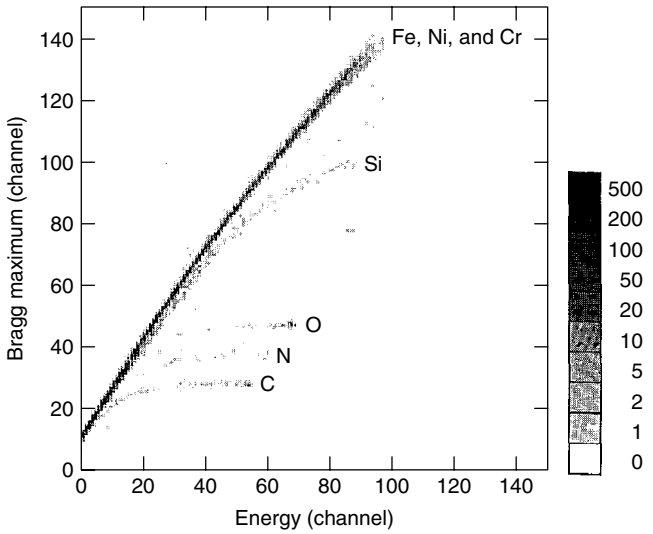

Figure 17 Two-dimensional plot (maximum energy loss/recoil energy) obtained on stainless steel with 136-MeV ^{127}I.[75] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

spectrometer or telescope). Electrostatic or magnetic spectrometers and ionization chambers are also in use.[68] Data processing is generally based on two or more parameter data acquisition and the elemental information is expressed in matrix form.

The analytical performance offered by high-energy HI-ERDA, particularly depth resolution, depends strongly on the total energy resolution of the detection system, on the nature and energy of incident ions and recoil particles, and on the depth where the collision has occurred. Dollinger et al. have reported a surface depth resolution of 8 nm for hydrogen determination in a thin carbon film bombarded by 120-MeV ^{197}Au ions in transmission geometry using a large magnetic spectrograph.[69] For carbon/boron multilayers deposited on a thick silicon substrate, a surface depth resolution of less than 1 nm is achieved in reflection geometry using the same beam parameters.[69] With low-energy ions and standard TOF-E telescope Brijs et al. demonstrated a depth resolution of 2 nm at the surface for H, N, and Si when characterizing very thin $SiO_2/Si_3N_4/SiO_2$ film stacks using glancing angles.[77] Measurement using 6-MeV ^{35}Cl incident ions and very glancing $3 + 35 = 38$ angle in reflection geometry is presented in Figure 18.

Normal ERDA geometry with the detector around 40° also offers good possibilities for the use of forward scattered incident ions, in addition to recoils, in the analysis. One possibility is to use light incident ion and do forward scattering analysis. The benefit of this is the high cross section and good depth resolution because of the small energy spread due to multiple scattering and kinematics. An example of this is the

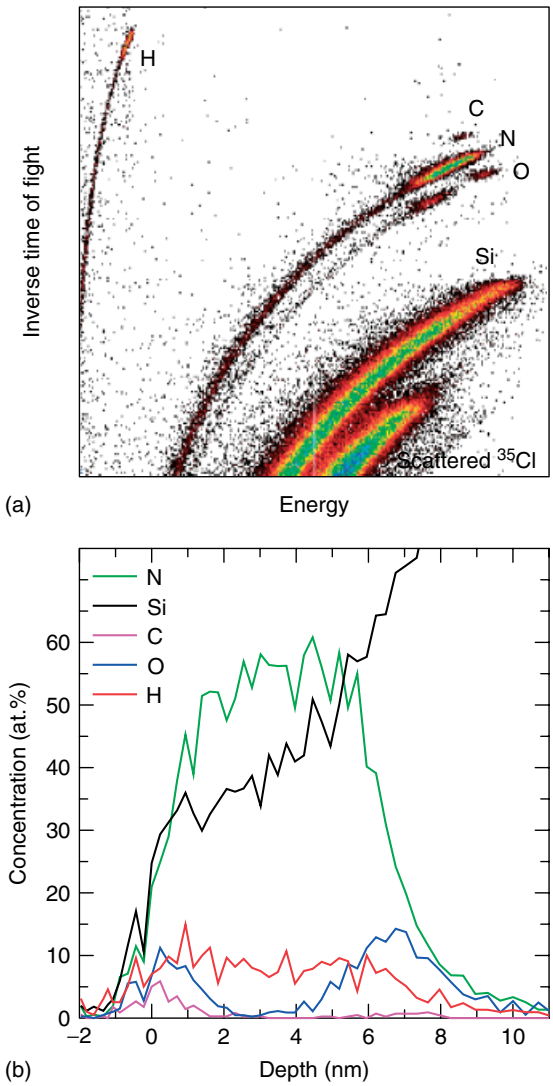

Figure 18 A metalorganic chemical vapor deposited Si_3N_4 film with some H and O impurities measured with 6-MeV ^{35}Cl beam using 3.2° incident and 35° recoiling angles. (a) A TOF-E histogram of the measurement showing all the sample elements and (b) the depth profiles converted from elemental TOF spectra using slab analysis approach. Density of 3.3 g cm^{-3} was used in conversion of the depth scale to nanometers.

study of 6-nm HfSiO and 25-nm NiSi films on silicon using 3–8-MeV ^{16}O ions as projectiles.[78] The use of scattered incident ions with recoiled target atoms in the analysis is demonstrated in Figure 19 where an atomic layer deposited Li–La–O film was measured using a 16-MeV ^{63}Cu beam. The recoiled La atoms suffered greatly from multiple scattering and could not be used in the analysis without proper MC simulations, but as La is the only heavy element present, scattered ^{63}Cu could be used instead. As two different scattering cross sections were used, the scattering angle had to be known very accurately.

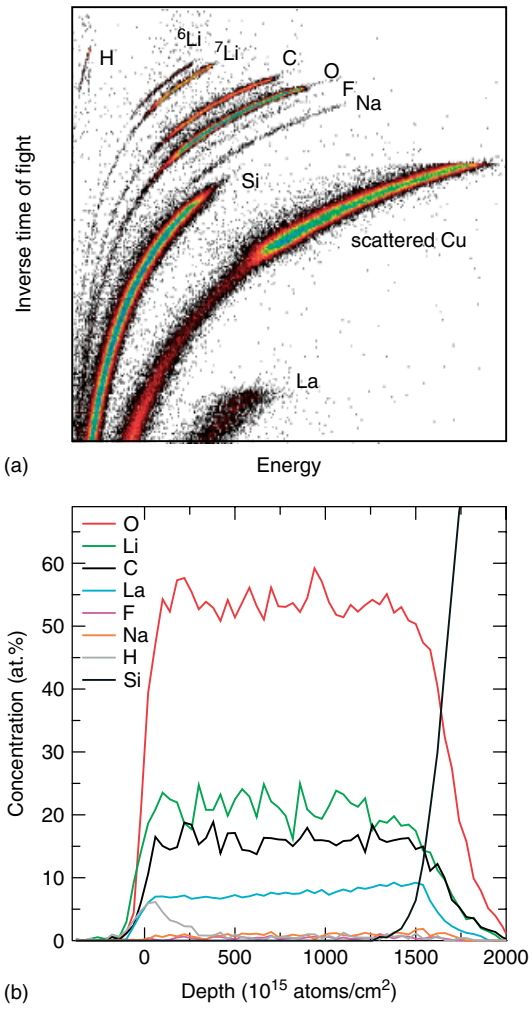

Figure 19 An atomic layer deposited Li–La–O film with nominal thickness of 165 nm measured with a 16-MeV ^{63}Cu beam and 19° incident and recoiling angles. (a) A TOF-E histogram of the measurement showing all the sample elements and (b) corresponding depth profiles. The scattered incident ions were used in the analysis of La profile. Note the low-energy tails due to multiple scattering.

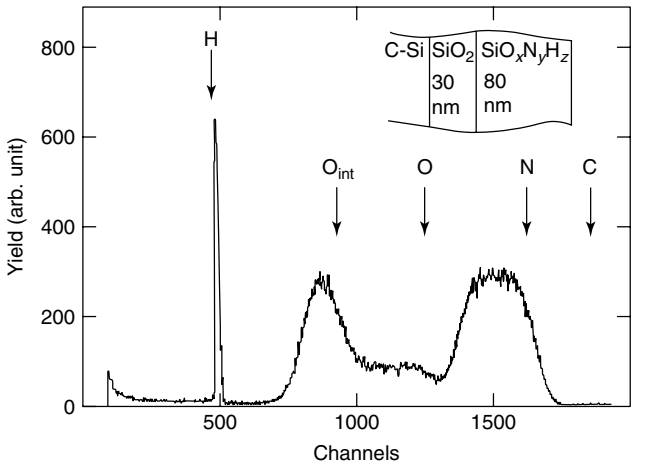

Figure 20 ERDA spectrum of an 80-nm silicon oxynitride film coated on 30-nm SiO_2 deposited on Si obtained with 30-MeV ^{28}Si ions.[80] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

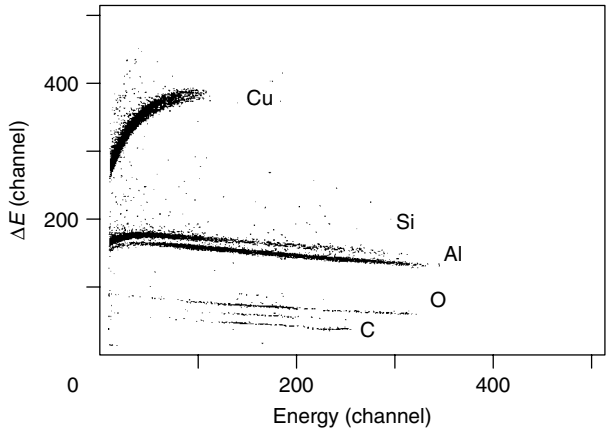

Figure 21 $\Delta E-E$ matrix and recoil energy spectra for a 15-nm Al/Cu multilayer deposited on silicon obtained by using a 200-MeV ^{197}Au ion beam.[81] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

The main limitation on analytical performance in high-energy HI-ERDA (depth resolution, sensitivity, or analyzed depth) is attributed to the behavior of the target material under ion bombardment. In the case of films with a thickness of the order of nanometers, the sputtering phenomenon, which is directly proportional to the nuclear energy loss, has to be taken into account when analyzing the measurements. In the case of thicker films, especially of polymers, a much more visible effect is the loss of light elements throughout the film thickness. In the case of porous low-*k* materials, the losses were found to be directly proportional to the energy loss of incident ions in the films and therefore to minimize the losses, the heaviest possible ion had to be used in the analysis.[79]

Figures 20 and 21 give two different examples of the use of high-energy HI-ERDA. The first example deals with the analysis of an 80-nm silicon oxynitride film coated on 30-nm SiO_2 deposited on Si with 30-MeV ^{28}Si ions.[80] The second example is dedicated to the characterization of a 15-nm Al/Cu multilayer deposited on silicon by using a 200-MeV ^{197}Au ion beam.[81]

Another general feature of HI-ERDA that has a great effect on the analysis reliability is multiple scattering, which, for more detailed analysis, can be divided to multiple and plural scattering. This multiple scattering causes angular spread, which can be described with analytical functions.[82–84,85] Equally important in the analysis are events where the scattering occurs closer to the sample atom nucleus resulting in a large angle

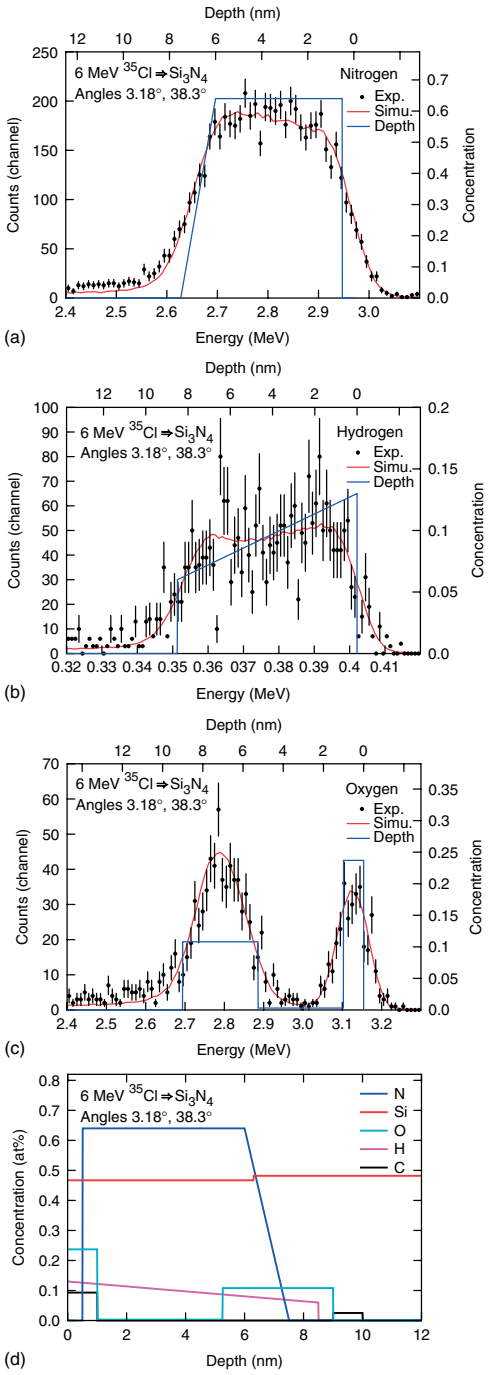

Figure 22 Monte Carlo simulations of the energy spectra for (a) nitrogen, (b) hydrogen, and (c) oxygen for the same measurement and data as in Figure 18. Black dots are experimental yields, red lines MC simulations, and blue lines depth profiles used for MC simulations. In (d) all depth profiles used for reproducing experimental spectra are collected in one depth profile.

scattering. Although these events are scarce, they can play an important role in the analysis. These scatterings are referred as *plural scattering*. Due to the nature of these events, no analytical function has been used to simulate them and they require full MC simulations. Multiple and plural scattering effects are strongly enhanced if the studied sample contains heavy elements; to carry out a very detailed analysis, the MC-simulation program should be used. The most sophisticated one at the moment is MCERD by Kai Arstila,[86] which can, due to several computational enhancements, do full simulations in roughly the same time as it takes to measure the sample. In Figure 22 an example of the use of MCERD in the analysis is shown for the same measurement as in Figure 18. The experimental energy spectra are well reproduced; the low-energy tails, and reliable depth profiles are also obtained. The conclusion of rising Si content deeper in the film drawn on the basis of slab analysis (Figure 18) can be proven to be wrong by MC-based analysis. In addition to MC simulations, also Molecular Dynamics simulations have a great potential in interpreting IBA results.[87]

6 APPLICATION EXAMPLES OF ELASTIC RECOIL DETECTION ANALYSIS FOR PROFILING HYDROGEN AND OTHER ELEMENTS

Four main application fields can be essentially distinguished for ERDA measurement, using either a low-energy ^{4}He beam, higher energy ^{4}He, or heavier ions: polymer science, microelectronics, thin films, and interface reactions. Each of these fields is illustrated in this section.

6.1 Polymer Sciences

Polymers are the most hydrogenated media encountered among solids. Three complementary topics of interest have given rise to the development of ERDA-based studies: the investigation of polymer surface properties, the hydrogen distribution in polymer blends and the irradiation behavior of polymers.

Many application examples have been presented and discussed by Tirira et al.[8] Only some of them are presented in this article.

For example, Chou et al. have studied the modifications induced by radio frequency or microwave treatment under a D_2O atmosphere of polyimide films. They showed that the polyimide surface exhibited a *D* uptake proportional to the time of treatment.[88] Moreover, a correlation has appeared between the surface concentration of hydroxy groups and the development of surface morphology during plasma treatment.

Green et al. studied the segregation of deuterated copolymer (molecular weight *M* in the range 87 000–520 000) at the interface of a homopolymer blend

(polystyrene/poly(vinyl methyl ether) with a molecular weight between 100 000 and 200 000) using 2.8-MeV ^{4}He ions.[89] They have shown that the diffusion coefficient of deuterated polymer varies as M^{-2}.

Composto et al. measured the diffusion coefficient of polymer molecules (deuterated polystyrene and deuterated poly(xylenyl ether) chains) in a complex polymeric structure using D as a tracer.[90] Diffusion coefficients have shown to be independent of the matrix molecular weight and to decrease as M^{-2}.

Several authors have applied ERDA to evaluate the degradation of polymer composition under charged-particle bombardment.[91] Wang et al. have investigated the modification of the electrical conductivity of high-temperature and high-performance polyetheretherketone (PEEK) and polyethersulfone polymers by ion implantation (50 and 180 keV As and Xe ions). A two-component regime has been found in which dominance depends on the implantation fluence ($<10^{15}$ ions cm^{-2} or $>10^{17}$ ions cm^{-2}). Another example of ERDA study of polymer degradation was given in Section 3. Figure 5 corresponds to the application of μERDA for polymer characterization after high-energy heavy-ion bombardment. In this case, 50-MeV ^{32}S ions were previously used to irradiate thin polyimide foils (25 μm). Bombardment-induced bond breaking leads to the formation of hydrogen-rich zones (micrometric bubbles or association of nanometric bubbles) trapped in the degraded polymer structure, as seen in Figure 5.

6.2 Semiconductor Materials

In this area, ERDA measurements are generally dedicated to profiling successive thin layers coated on a crystalline semiconductor substrate (Si, Ge, or GaAs) having different compositions and thicknesses.

The hydrogen profile is measured using a conventional approach with ^{4}He ions as shown by Wang et al.[92] or Barbour et al.[93] Other light elements such as C, N, and O have been profiled using high-energy HI-ERDA.[80]

Wang et al. have profiled hydrogen in semi-insulating polysilicon films. Figures 23 and 24 show the hydrogen recoil spectra obtained on films immersed in a gas flow reactant mixture (N_2O/SiH_4) in various ratios or annealed between 700 and 1000 °C for 30 min[92] for a given gas reactant ratio. The film deposition rate decreases when the N_2O/SiH_4 ratio increases owing to the increase in oxygen concentration. The H content of the film decreases when the annealing temperature increases owing to the densification and restructuring of the film.

Barbour et al. have shown that for silicon nitride film deposition by electron cyclotron resonance, the hydrogen content of the film was minimized when the stoichiometry was closest to Si_3N_4.[93]

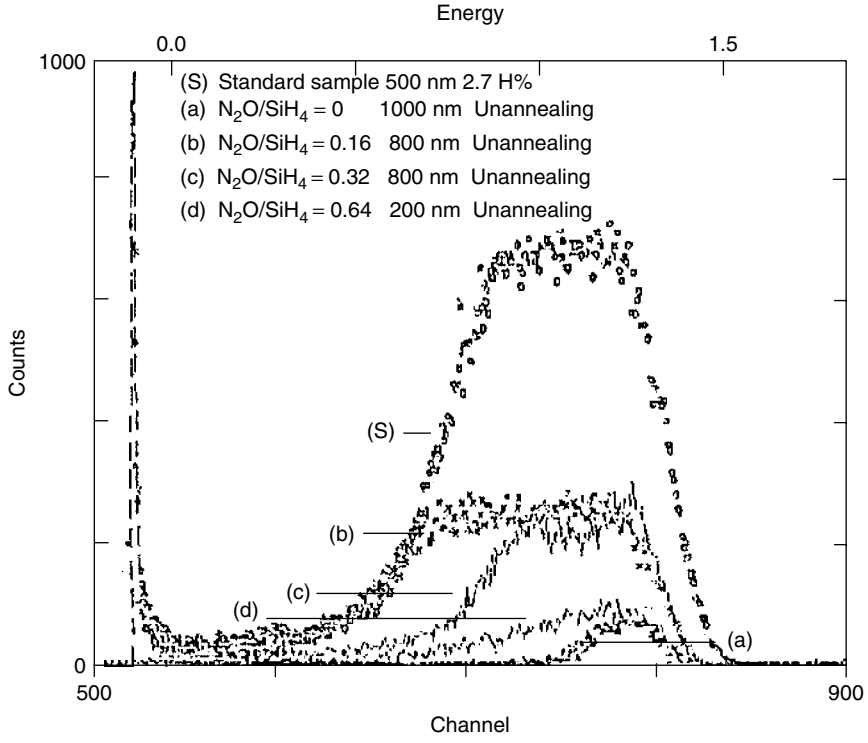

Figure 23 Hydrogen recoil spectra for polysilicon samples immersed in a N_2O/SiH_4 flow with various N_2O/SiH_4 ratios. (Data reproduced from Wang et al.,[92] with the kind permission of Elsevier Science and Plenum Press.)

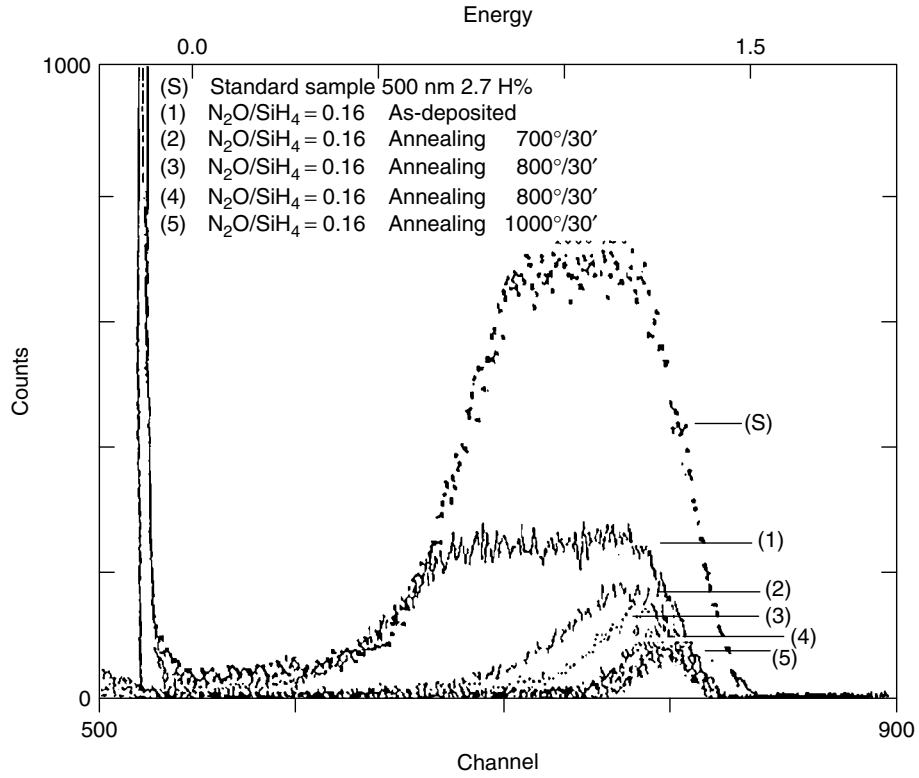

Figure 24 Hydrogen recoil spectra obtained for various annealing temperatures. (Data reproduced from Wang et al.,[92] with the kind permission of Elsevier Science and Plenum Press.)

6.3 Thin Films

In the case of thin-film monolayered or multilayered structures, ion beam methods have been revealed to be powerful characterization tools, particularly considering the HI-ERDA[94] and coupling of RBS with ERDA.[95–97]

The development of materials based on hydrogenated amorphous carbon ''diamond-like'' or silicon-carbide layers, because of their promising tribological properties, strongly influenced the interest in ERDA in the early 1990s.[98–101]

The relationship between the optical and electrical properties of hydrogenated amorphous silicon carbon films and their stoichiometry have been determined combining RBS and ^{4}He-induced ERDA by Compagnini et al.[95] A hydrogen concentration of 30 atom% was shown to increase the optical energy gap by a factor of two and the electrical resistivity by several orders of magnitude.

Figure 25 gives three recoil spectra obtained on a:C–H or a:C–D layers. Hydrogen implantation in the a:C–D layer was shown to induce a large surface D release,[99] extending to at least 80% of the analyzed depth (0–0.3 μm). Microelectronics is an important driving force in the development of this method. IBA has played an important role in the development of films replacing SiO_2 as a gate oxide[102] and the role of IBA in the film development has been very important. Another important thin-film application related to microelectronics has been Cu diffusion barriers like TiN[103] and TaN.[104,105] During the development process, deposited films normally contain substantial amounts of H, C, and O due to the precursors used – HI-ERDA can easily provide this information.

6.4 Interface Reactivity

This field of application corresponds to the experimental study of atomic transport mechanisms near a solid/solid or a solid/liquid interface, for example, oxide corrosion behavior,[106,107] wear properties of treated metal, amorphous layers, or glass surfaces.[108–113] Intensive use of ERDA has been reported by the Canadian group from Varennes[114] and by the French group from the University of Lyon.[11] These two groups have collaborated in studying some candidate materials for fusion reactor device coatings that act as facings for the plasma. Essentially, they studied hydrogen loss under irradiation and annealing conditions.

Figure 26 shows the erosion of the hydrogen depth profile in an a:C–H layer after $^4He^+$ bombardment at

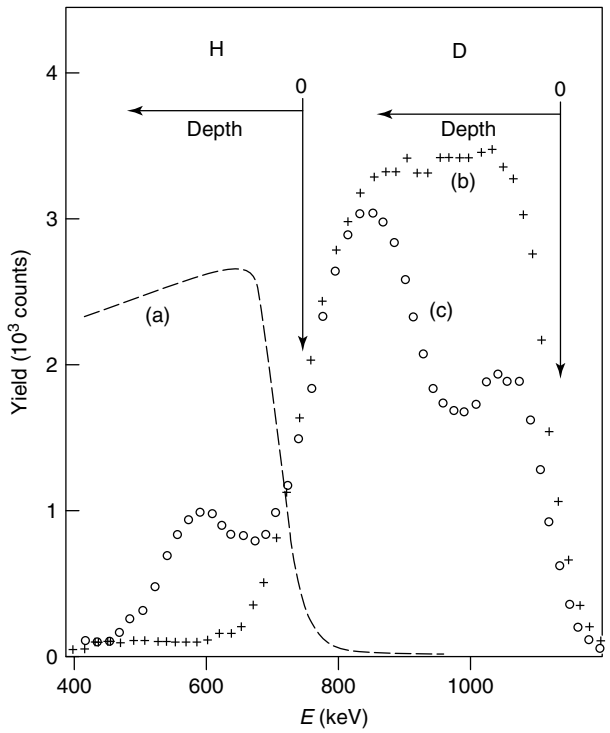

Figure 25 Energy spectrum of recoil protons and deuterons obtained with a 2.6-MeV $^4He^+$ beam (incidence angle = 75°, recoil angle = 30°). (a) Proton reference from unimplanted a:C–H layer, (b) deuteron reference from unimplanted a:C–D layer, (c) implanted area of the same layer in (b) after implantation with 11-keV H^+ at a fluence of 1.8×10^{17} ions cm^{-2}. The depth scale is approximately 0–0.3 μm. (Data from Boutard et al.,[99] with the kind permission of Plenum Press.)

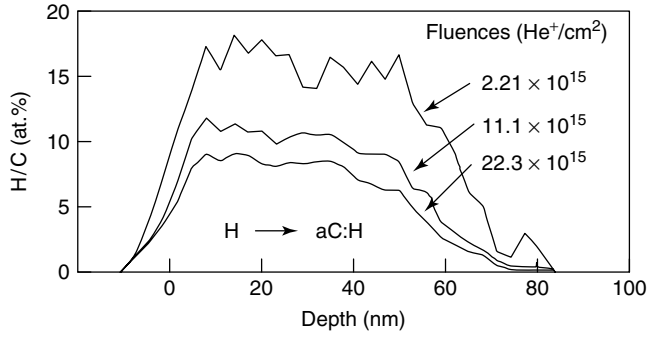

Figure 26 Evolution of the hydrogen depth profile measured by the $E \times B$ technique in an a:C–H layer bombarded with $^4He^+$ between 2.21×10^{15} and 2.23×10^{16} ions cm^{-2}. (Data reproduced from Ross and Richard,[115] with the kind permission of Plenum Press.)

different fluences. The $E \times B$ technique is used with a 350-keV incident helium beam.[115]

Ross and Richard have shown that hydrogen depletion is negligible in Si, SiC, and Sc, higher in Be, Si_3N_4, and TiC, and very high in C.

6.5 Other Application Examples

Other conventional ERDA or high-energy HI-ERDA applications have been reported since the late 1980s. For example, Mosbah et al. successfully proposed the use of ERDA in transmission geometry to measure the water content of melt inclusions trapped in minerals.[44] Their first published data concerned inclusions from Pantelleria, San Pietro, and Guadeloupe: water contents ranged between 3.4 and 11.8 wt% and agreed well with theoretical calculations. Naitoh et al. have studied the reordering of an aluminum-adsorbed silicon surface like $\sqrt{3} \times \sqrt{3}$ Al/Si (111) induced by hydrogen sorption.[116] They found that saturation occurs for a hydrogen coverage of about 1.3×10^{15} H cm^{-2}. One totally new application for the TOF-ERDA telescope was introduced by Zhang et al. in 2001.[117] For doing stopping force measurements, they installed a movable thin film between the second timing gate and energy detector and could then obtain continuous stopping curves for this film. This idea has now been used in many places for stopping force measurements, and has given better stopping values for the IBA community.

7 CONCLUSION

Three of the most striking characteristics of the ERDA method are its wide usable energy range from 350 keV to 1 GeV, its strong versatility in terms of detection configurations and variants, and with HI-ERDA, the possibility of depth profiling all the sample elements from H to Bi in a single measurement.

Besides the fact that it remains a delicate method, ERDA has featured in IBA-related conferences as the most promising and evolving technique for quantitative determination of elemental depth profiles in solids, from hydrogen to medium and even high mass numbers. In this approach, no prior knowledge of the sample is required and good results can be obtained without standard samples. The basic physics of the elastic collision process has made substantial progress since the late 1980s, especially when considering differential cross-section measurements and models or data-processing methods.

Moreover, the spectacular development of high-energy HI-ERDA since the beginning of the 1990s must be emphasized. It is based on the strong theoretical and experimental capabilities derived from nuclear physics research (experimental devices, technical procedures, and theoretical models) that have brought a constant source of improvement to ERDA users for materials science applications. More and more IBA laboratories worldwide have developed specific facilities for applying ERDA to various problems in this field. In addition to

high-energy HI-ERDA, special TOF-E telescopes have been constructed for laboratories with smaller (1.7–3-MV terminal voltage) tandem accelerators, making this unique analysis technique available to more users.

The different application examples presented and discussed in the above sections constitute a good sample of the use of ERDA.

ERDA is the most practical IBA technique available for profiling hydrogen isotopes in the near-surface region of solids from 100 nm to 10 μm with a depth resolution of 20 nm for conventional ERDA and even 1 nm for HI-ERDA at the surface. The use of resonant nuclear reactions induced by high-energy heavy-ion beams to profile hydrogen seems to be restricted to specific applications and particularly to samples able to support high fluences of 0.4–0.5-MeV amu^{-1} ^{15}N ions.[118] A round-robin characterization of ultrathin to thin (1–100 nm) aluminum oxynitride (AINO) films was organized and the participants used different IBA techniques.[60] This paper gives a good overview of the possibilities and limitations of present IBA techniques in the case of thin films. Considering the analytical capabilities of SIMS, it must be noted that besides its excellent depth resolution and its very strong sensitivity, SIMS is a destructive technique and the calibration of the erosion rate of the target under investigation remains difficult. Nevertheless, the technical improvements encountered with the SIMS machines of the last generation tended to enhance the competition between these two methods.

The future of ERDA looks bright. In three decades, it has matured from something new and exciting to a widely used and trusted technique. In the beginning, the technique was utilized in big laboratories mostly concentrating on nuclear physics, but having plenty of useful instruments around. In the next phase, smaller laboratories, which had mostly concentrated on RBS adopted the conventional ERDA technique, but the instrumentation development was still concentrated on bigger institutes that had mostly high-energy heavy beams available. This millennium, the development has again been toward smaller accelerators, but now using HI-ERDA usually with TOF-E detectors. In these laboratories, the use of 5–20-MeV heavy incident beams have replaced, in many applications, the previously intensively used RBS and nuclear reaction analysis (NRA). This development toward lower energies is likely to continue as large tandem accelerators are not constructed anymore and old ones are being dismantled.

It is not difficult to predict that, in the future, accurate ERD analysis will be performed by means of MC-based simulation software capable of handling properly both multiple and plural scattering, and also all the instrumental effects that broaden the energy spectra. With the aid of this type of analysis, the surface depth resolution, and especially the reliability of analysis deeper in the sample, will be pushed to a new level.

ACKNOWLEDGMENTS

Dr. Kai Arstila is acknowledged for the MC simulations.

ABBREVIATIONS AND ACRONYMS

ERDA	Elastic Recoil Detection Analysis
HI-ERDA	Heavy-ion Elastic Recoil Detection Analysis
IBA	Ion Beam Analysis
MC	Monte Carlo
μERDA	Microbeam-induced Elastic Recoil Detection Analysis
NRA	Nuclear Reaction Analysis
PEEK	Polyetheretherketone
RBS	Rutherford Backscattering Spectrometry
SIMS	Secondary Ion Mass Spectrometry
TOF	Time-of-flight

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission) • Rutherford Backscattering Spectroscopy

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

REFERENCES

1. J. L'Ecuyer, C. Brassard, C. Cardinal, J. Chabbal, L. Deschênes, J.P. Labrie, B. Terreault, J.G. Martel, R. Saint Jacques, 'An Accurate and Sensitive Method for the Determination of the Depth Distribution of Light Elements in Heavy Materials', *J. Appl. Phys.*, **47**, 381–382 (1976).
2. B.L. Doyle, P.S. Peercy, 'Technique for Profiling ^{1}H with 2.5 MeV Van de Graaff Accelerators', *Appl. Phys. Lett.*, **34**, 811–813 (1979).
3. G.M. Padawer, E.J. Schneid, 'A Lithium Microprobe Technique for Measuring the Concentration Profile of Hydrogen at Surfaces', *Trans. Am. Nucl. Soc.*, **12**, 493–494 (1969).

4. E. Ligeon, A. Guivarch, 'A New Utilization of ^{11}B Ion Beams: Hydrogen Analysis by ^{1}H(^{11}B, α)$\alpha\alpha$ reaction', *Radiat. Eff.*, **22**, 101–105 (1974).
5. W.A. Lanford, H.P. Trautvetter, J.F. Ziegler, J. Keller, 'New Precision Technique for Measuring the Concentration Versus Depth of Hydrogen in Solids', *Appl. Phys. Lett.*, **28**, 566–568 (1976).
6. D.A. Leich, T.A. Tombrello, 'A Technique for Measuring Hydrogen Concentration Versus Depth in Solid Samples', *Nucl. Instrum. Methods*, **108**, 67–71 (1973).
7. J.P. Thomas, C. Pijolat, M. Fallavier, 'Emploi de la Réaction Résonnante ^{1}H(^{15}N, $\alpha\gamma$)^{12}C pour L'obtention de Profils de Concentration de l'hydrogène dans les Matériaux', *Rev. Phys. Appliquée*, **13**, 433–439 (1978).
8. J. Tirira, Y. Serruys, P. Trocellier, *Forward Recoil Spectrometry: Applications to Hydrogen Determination in Solids*, Plenum, New York, 1996.
9. L. Calcagno, G. Foti, 'Interdiffusion in Polystyrene Crosslinked by keV ion Irradiation', *J. Appl. Phys.*, **71**, 3216–3217 (1992).
10. T. Sajavaara, J. Jokinen, K. Arstila, J. Keinonen, 'TOF-ERDA Spectrometry Applied for the Analysis of Be Migration in (100) GaAs', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **139**, 225–229 (1998).
11. G. Barbier, '*Etude du Comportement de l'hydrogène dans des Matériaux Amorphes Hydrogénés de Type a-C: H et a-SiC: H Devant Faire Face au Plasma des Réacteurs à Fusion*', Thèse, Université Lyon I, 10 April 1997.
12. W.K. Chu, J.W. Mayer, M.A. Nicolet, *Backscattering Spectrometry*, Academic Press, New York, 1978.
13. M. Bozoian, 'Threshold of Non-Rutherford Nuclear Cross-section for Ion Beam Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **56/57**, 740–743 (1991).
14. P. Goppelt, B. Gebauer, D. Fink, M. Wilpert, Th. Wilpert, W. Bohne, 'High Energy ERDA with Very Heavy Ions Using Mass and Energy Dispersive Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **68**, 235–240 (1992).
15. J. Räisänen, E. Rauhala, J. Knox, J.F. Harmon, 'Non-Rutherford Cross-section in Heavy Ion Elastic Recoil Spectrometry: 40–70 MeV ^{32}S Ions on Carbon, Nitrogen and Oxygen', *J. Appl. Phys.*, **75**, 3273–3277 (1994).
16. H.H. Andersen, F. Besenbacher, P. Loftager, W. Möller, 'Large-Angle Scattering of Light Ions in the Weakly Screened Rutherford Region', *Phys. Rev. A*, **21**, 1891–1901 (1980).
17. G. Freier, E. Lampi, W. Sleator, J.H. Williams, 'Angular Distribution of 1 to 3.5 MeV Protons Scattered by ^{4}He', *Phys. Rev.*, **75**, 1345–1347 (1949).
18. G.R. Miller, G.C. Phillips, 'Scattering of Protons from Helium and Level Parameters in ^{5}Li', *Phys. Rev.*, **112**, 2043–2047 (1958).
19. R.A. Arndt, D.D. Long, L.D. Roper, 'Nucleon–alpha Elastic-scattering Analyses', *Nucl. Phys. A*, **209**, 429–446 (1973).
20. G.R. Satchler, W. Owen, A.J. Elwyn, G.L. Morgan, R. Walte, 'An Optical Model for the Scattering of Nucleons from ^{4}He at Energies Below 20 MeV', *Nucl. Phys. A*, **112**, 1–30 (1968).
21. C.L. Critchfield, D.C. Dodder, 'Phase Shifts in Protonalpha Scattering', *Phys. Rev.*, **76**, 602–605 (1949).
22. D.C. Dodder, G.M. Halle, N. Jarmie, K. Witte, 'Tables of Phase Shifts and Experimental Observables for ^{4}He Elastic Scattering', Report LA-6389-MS, Los Alamos Research Laboratories, 1968.
23. T. Stammbach, R. Walter, '*R*-matrix Formulation and Phase Shifts for N-^{4}He Scattering for Energies up to 20 MeV', *Nucl. Phys. A*, **180**, 225–240 (1972).
24. J. Tirira, P. Trocellier, J.P. Frontier, P. Trouslard, 'Theoretical and Experimental Study of Low-energy ^{4}He-induced ^{1}H Elastic Recoil with Application to Hydrogen Behavior in Solids', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 203–207 (1990).
25. J. Tirira, F. Bodart, 'Alpha-proton Elastic Scattering Analyses up to 4 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **74**, 496–502 (1996).
26. G. Terwagne, G.G. Ross, L. Leblanc, 'Cross-section Measurements of the ^{1}H(^{3}He, ^{1}H)^{3}He Reaction at 20° and 30° between 1.9 and 3 MeV', *J. Appl. Phys.*, **79**, 8886–8891 (1996).
27. V. Quillet, F. Abel, M. Schott, 'Absolute Cross-section Measurements for H and D Elastic Recoil Using 1 to 2.5 MeV ^{4}He Ions and for the ^{12}C(d, p)^{13}C and ^{16}O(d, p_1)^{17}O Nuclear Reactions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **83**, 47–61 (1993).
28. A. Galonsky, R.A. Douglas, W. Haeberli, M.T. McEllistrem, H.T. Richards, 'Deuteron-Helium Differential Scattering Cross-section', *Phys. Rev.*, **98**, 586–599 (1955).
29. F. Besenbacher, I. Stensgaard, P. Vase, 'Absolute Cross-section for Recoil Detection of Deuterium', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 459–463 (1986).
30. J.A. Sawicki, 'Measurements of the Differential Cross-sections for Recoil Tritons in ^{4}He-^{3}T Scattering at Energies between 0.5 and 2.5 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 123–127 (1988).
31. B.L. Cohen, C.L. Fink, J.H. Degnan, 'Nondestructive Analysis for Trace Amounts of Hydrogen', *J. Appl. Phys.*, **43**, 19–25 (1972).
32. F.A. Smidt, A.G. Pieper Jr, 'Studies of the Mobility of Helium in Vanadium', *J. Nucl. Mater.*, **51**, 361–365 (1974).
33. J.A. Moore, I.V. Mitchell, M.J. Hollis, J.A. Davies, L.M. Howe, 'Detection of Low-mass Impurities in Thin Films

Using MeV Heavy-ion Elastic Scattering and Coincidence Detection Techniques', *J. Appl. Phys.*, **46**, 52–61 (1975).

34. J. Tirira, P. Trocellier, J.P. Frontier, 'Analytical Capabilities of ERDA in Transmission Geometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 147–150 (1990).
35. J. Tirira, P. Trocellier, J.P. Frontier, P. Massiot, J.M. Costantini, V. Mori, '3D Hydrogen Profiling by Elastic Recoil Detection Analysis in Transmission Geometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **50**, 135–139 (1990).
36. J. Tirira, J.P. Frontier, P. Trocellier, Ph. Trouslard, 'Development of a Simulation Algorithm for Energy Spectra of Elastic Recoil Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **54**, 328–333 (1991).
37. E. Kótai, 'Computer Methods for Analysis and Simulation of RBS and ERDA Spectra', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **85**, 588–596 (1994).
38. M. Mayer, 'SIMNRA Users Guide', Report IPP 9/113, Max-Planck-Institut für Plasmaphysik, Garching, 1997.
39. N.P. Barradas, K. Arstila, G. Battistig, M. Bianconi, N. Dytlewski, C. Jeynes, E. Kótai, G. Lulli, M. Mayer, E. Rauhala, E. Szilágyi, M. Thompson, 'Summary of ''IAEA intercomparison of IBA software''', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **266**, 1338–1342 (2008).
40. P. Trocellier, M. Le Pioufle, 'La Microsonde Nucléaire du Centre d'Etudes de Bruyéres-le-Châtel en Sciences des Matériaux', *Rev. Sci. Technique de la Défense*, **1**, 137–146 (1994).
41. O. Auciello, R. Kelly, *Ion Bombardment Modification of Surfaces*, Elsevier, New York, 1984.
42. A. Dunlop, F. Rullier-Albenque, C. Jaouen, C. Templier, J. Davenas, *Materials under Irradiation*, Transactions Technical Publications, Brookfield, 1993.
43. B.P. Doyle, R.D. Maclear, S.H. Connell, P. Formenti, I.Z. Machi, J.E. Butler, P. Schaaff, J.P.F. Sellschop, E. Sideras-Haddad, K. Baruth-Ram, '3D Micro-ERDA Microscopy of Trace Hydrogen Distributions in Diamond Using a 2-D-PSD with Event Reconstruction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **130**, 204–210 (1997).
44. M. Mosbah, R. Clocchiatti, J. Tirira, J. Gosset, P. Massiot, P. Trocellier, 'Study of Hydrogen in Melt Inclusions Trapped in Quartz with a Nuclear Microprobe', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **54**, 298–303 (1991).
45. R.J. Sweeney, V.M. Prozesky, K.A. Springhorn, 'The Quantitative Determination of H in Geological Materials (Silicates) by Elastic Recoil Detection Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **130**, 623–627 (1997).
46. V.M. Prozesky, C.L. Churms, J.V. Pilcher, K.A. Springhorn, R. Behrisch, 'ERDA Measurement of Hydrogen Isotopes with a $\Delta E - E$ Telescope', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **84**, 373–379 (1994).
47. D.K. Avasthi, D. Kabiraj, A. Bhagwat, G.K. Mehta, V.D. Vankar, S.B. Ogale, 'Simultaneous Detection of Light Elements by ERDA with Gas-ionization/Si $\Delta E - E$ Telescope', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **93**, 480 (1994).
48. J.P. Stoquert, G. Guillaume, M. Hage-Ali, J.J. Grob, C. Ganter, P. Siffert, 'Determination of Concentration Profiles by Elastic Recoil Detection with a $\Delta E - E$ gas Telescope and High Energy Incident Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **44**, 184–194 (1989).
49. H. Timmers, R.G. Elliman, G.R. Palmer, T.R. Ophel, D.J. O'Connor, 'The Development of a Facility for Heavy-ion Elastic Recoil Detection Analysis at the Australian National University', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 611–615 (1998).
50. W.M. Arnoldbik, C.T.A.M. de Laat, F.H.M.P. Habraken, 'On the Use of a $\Delta E - E$ Telescope in Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 832–835 (1992).
51. W.M. Arnoldbik, F.H.P.M. Habraken, 'Elastic Recoil Detection', *Rep. Prog. Phys.*, **56**, 859–902 (1993).
52. J.A. Davies, J.S. Forster, S.R. Walker, 'Elastic Recoil Detection Analysis with Heavy Ion Beams', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 594–602 (1998).
53. D. Plachke, H.D. Carstanjen, R.M. Emrick, H. Nolte, H. Huber, W. Assmann, R. Grötzschel, 'Channeling and Blocking Measurements on Quasicrystals', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 816–821 (1998).
54. R. Siegele, J.A. Davies, J.S. Forster, H.R. Andrews, 'Forward Elastic Recoil Measurements Using Heavy Ions', *J. Appl. Phys.*, **76**, 4524–4532 (1994).
55. W. Assmann, H. Huber, C. Steinhausen, M. Dobler, H. Glucker, A. Weidiger, 'Elastic Recoil Detection Analysis with Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **89**, 131–139 (1994).
56. R. Groleau, S.C. Gujrathi, J.P. Martin, 'Time-of-flight System for Profiling Recoiled Light Elements', *Nucl. Instrum. Methods Phys. Res.*, **218**, 11–15 (1983).
57. J.P. Thomas, M. Fallavier, D. Ramdane, N. Chevarier, A. Chevarier, 'High-resolution Depth Profiling of Light Elements in High Atomic Mass Materials', *Nucl. Instrum. Methods Phys. Res.*, **218**, 125–128 (1983).
58. H.J. Whitlow, G. Possnert, C.S. Petersson, 'Quantitative Mass and Energy-dispersive Elastic Recoil Spectrometry: Resolution and Efficiency Considerations', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **27**, 448–457 (1987).
59. G. Boudreault, R.G. Elliman, R. Grötzschel, S.C. Gujrathi, C. Jeynes, W.N. Lennard, E. Rauhala, T. Sajavaara, H. Timmers, Y.Q. Wang, T.D.M. Weijers, 'Round Robin: Measurement of H Implantation Distributions in Si by Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **222**, 547–566 (2004).

60. N.P. Barradas, N. Added, W.M. Arnoldbik, I. Bogdanović-Radović, W. Bohne, S. Cardoso, C. Danner, N. Dytlewski, P.P. Freitas, M. Jakšić, C. Jeynes, C. Krug, W.N. Lennard, S. Lindner, Ch. Linsmeier, Z. Medunić, P. Pelicon, R.P. Pezzi, C. Radtke, J. Röhrich, T. Sajavaara, T.D.M. Salgado, F.C. Stedile, M.H. Tabacniks, I. Vickridge, 'A Round Robin Characterisation of the Thickness and Composition of Thin to Ultra-thin AINO Films', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **227**, 397–419 (2005).

61. C. Kottler, M. Döbeli, F. Glaus, M. Suter, 'A Spectrometer for Low Energy Heavy Ion ERDA', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **248**, 155–162 (2006).

62. P. Reichart, G. Dollinger, A. Bergmaier, G. Datzmann, A. Hauptner, H.J. Körner, 'Sensitive 3D hydrogen microscopy by proton proton scattering', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **219–220**, 550–554 (2004).

63. S.S. Klein, P.H.A. Mutsaers, B.E. Fischer, 'Mass Selection and Depth Profiling by Coincident Recoil Detection for Nuclei in the Middle-mass Region', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **50**, 150–153 (1990).

64. H.C. Hofsäss, N.R. Parikh, M.L. Swanson, W.K. Chu, 'Depth Profiling of Light Elements Using Elastic Recoil Coincidence Spectrometry (ERCS)', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 151–156 (1990).

65. G.G. Ross, B. Terreault, G. Gobeil, G. Abel, C. Boucher, G. Veilleux, 'Inexpensive, Quantitative Hydrogen Depth Profiling for Surface Probes', *J. Nucl. Mater.*, **128/129**, 730–733 (1984).

66. G.G. Ross, L. Leblanc, 'Depth Profiling of Hydrogen and Helium Isotopes by Means of the ERD E × B Technique', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **62**, 484–492 (1992).

67. S.S. Klein, 'Separate Determination of Concentration Profiles for Atoms with Different Masses by Simultaneous Measurement of Scattered Projectile and Recoil Energies', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 464–468 (1986).

68. H.A. Rijken, S.S. Klein, L.J. Van Ijzendorn, M.J.A. de Voigt, 'Elastic Recoil Detection Analysis with High Energy Alpha Beams', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **79**, 532–535 (1993).

69. G. Dollinger, T. Faestermann, P. Maier-Komor, 'High Resolution Depth Profiling of Light Elements', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 422–427 (1992).

70. B. Terreault, J.G. Martel, R.G. Saint Jacques, J. L'Ecuyer, 'Depth Profiling of Light Elements in Materials with High Energy Ion Beams', *J. Vac. Sci. Technol.*, **14**, 492–500 (1977).

71. B.L. Doyle, N.D. Wing, 'The Sandia Nuclear Microprobe', SANDIA Report SAND82-2393, 1982.

72. E.B. Carter, G.E. Mitchell, R.H. Davis, 'Elastic Scattering of Alpha Particles by ^{12}C in the Bombarding Energy Range 10–19 MeV', *Phys. Rev.*, **133**, B1421–B1433 (1964).

73. M.K. Mehta, W.E. Hunt, R.H. Davis, 'Scattering of Alpha Particles by Oxygen II: Bombarding Energy Range 10–19 MeV', *Phys. Rev.*, **160**, 791–802 (1967).

74. H.A. Rijken, *'Detection Methods for Depth Profiling of Light Elements Using High Energy Alpha Particles'*, Thesis, University of Eindhoven, 1993.

75. R. Siegele, J.A. Davies, J.S. Forster, H.R. Andrews, 'Forward Elastic Recoil Measurements Using Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **90**, 606–610 (1994).

76. R. Siegele, H.K. Haugen, J.A. Davies, J.S. Forster, H.R. Andrews, 'Forward Elastic Recoil Measurements Using Heavy Ions', *J. Appl. Phys.*, **76**(8), 4524–4532 (1994).

77. B. Brijs, T. Sajavaara, S. Giangrandi, T. Janssens, T. Conard, K. Arstila, K. Nakajima, K. Kimura, A. Bergmaier, G. Dollinger, A. Vantomme, W. Vandervorst, 'The Analysis of a thin $SiO_2/Si_3N_4/SiO_2$ Stack: A Comparative Study of Low-energy Heavy-ion Elastic Recoil Detection, High-resolution Rutherford Backscattering and Secondary Ion Mass Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **249**, 847–850 (2006).

78. T. Sajavaara, B. Brijs, S. Giangrandi, K. Arstila, A. Vantomme, W. Vandervorst, 'Analysis of Thin High-k and Silicide Films by Means of Heavy Ion Time-of-flight Forward-scattering Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **249**, 292–296 (2006).

79. S. Giangrandi, B. Brijs, T. Sajavaara, H. Bender, F. Iacopi, A. Vantomme, W. Vandervorst, 'Irradiation-induced Damage in Porous Low-k Materials During Low-energy Heavy-ion Elastic Recoil Detection Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **249**, 189–192 (2006).

80. F.H.P.M. Habraken, 'Light Element Depth Profiling Using Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **68**, 181–189 (1992).

81. W. Assmann, H. Huber, Ch. Steinhausen, M. Dobler, H. Glückler, A. Weidinger, 'Elastic Recoil Detection Analysis with Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **89**, 131–139 (1994).

82. P. Sigmund, K. Winterborn, 'Small-Angle Multiple Scattering of Ions in the Screened Coulomb Region. I. Angular Distributions', *Nucl. Instrum. Methods*, **119**, 541–557 (1974).

83. E. Szilágyi, F. Paszti, 'Theoretical calculation of the depth resolution of IBA methods', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **85**, 616–620 (1994).

84. E. Szilágyi, L. Wielunski, F. Pászti, 'Theoretical Approximation of Energy Distribution of Elastically Recoiled Hydrogen Atoms', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 701–706 (1998).

85. G. Amsel, G. Battistig, A. L'Hoir, 'Small angle multiple scattering of fast ions, physics, stochastic theory and numerical calculations', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **201**, 325–388 (2003).

86. K. Arstila, T. Sajavaara, J. Keinonen, 'Monte Carlo Simulation of Multiple and Plural Scattering in Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **174**, 163–172 (2001).

87. K. Nordlund, 'Molecular dynamics for ion beam analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **266**, 1886–1891 (2008).

88. N.J. Chou, A.D. Marwick, R.D. Goldblatt, L. Li, G. Coleman, J.E. Heidenreich, J.R. Paraszczak, 'An Isotope Approach to Characterization of Microwave-Water-Plasma-Modified Polyimide Surfaces', *J. Vac. Sci. Technol., A*, **10**, 248–253 (1992).

89. P.F. Green, D.B. Adolf, L.R. Gilliom, 'Dynamics of Polystyrene/Poly(Vinyl Methyl Ether) Blends', *Macromolecules*, **24**, 3377–3382 (1991).

90. R.J. Composto, E.J. Kramer, D.M. White, 'Reptation in Polymer Blends', *Polymer*, **31**, 2320–2328 (1990).

91. Y. Wang, S.S. Mohite, L.B. Bridwell, R.E. Gield, C.J. Sofield, 'Modification of High-temperature and High-performance Polymers by Ion Implantation', *J. Mater. Res.*, **8**, 388–403 (1993).

92. Y. Wang, B. Huang, D. Cao, J. Cao, D. Zhu, K. Shen, 'Analysis of Hydrogen in Oxygen-doped Polysilicon by $^4He^+$ – H Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **84**, 111–113 (1994).

93. J.C. Barbour, H.J. Stein, O.A. Popov, M. Yoder, C.A. Outten, 'Silicon Nitride Formation from a Silane-Nitrogen Electron Cyclotron Resonance Plasma', *J. Vac. Sci. Technol., A*, **9**, 480–484 (1991).

94. M. Putkonen, T. Sajavaara, L. Niinisto, J. Keinonen, 'Analysis of ALD-processed Thin Films by Ion-beam Techniques', *Anal. Bioanal. Chem.*, **382**, 1791–1799 (2005).

95. G. Compagnini, L. Calcagno, G. Foti, 'Properties of Fully Implanted Si_xC_{1-x}: H Alloys', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **80/81**, 978–981 (1993).

96. R.G. Gordon, U. Riaz, D.M. Hoffman, 'Chemical Vapor Deposition of Aluminum Nitride Thin Films', *J. Mater. Res.*, **7**, 1679–1684 (1992).

97. H.Z. Wu, T.C. Chou, A. Mishra, D.R. Anderson, J.K. Lampert, 'Characterization of Titanium Nitride Thin Films', *Thin Solid Films*, **191**, 55–67 (1990).

98. M. Wang, K. Schmidt, K. Reichelt, H. Dimigen, H. Hübsch, 'The Properties of Titanium-containing Amorphous Hydrogenated Carbon Films', *Surf. Coat. Technol.*, **47**, 691–695 (1991).

99. D. Boutard, W. Möller, B.M.U. Scherzer, 'Hydrogen Isotope Exchange in Amorphous Hydrocarbon Layers: Plasma-deposited a-C:H and Ion Saturated Carbon', *Radiat. Eff. Defects Solids*, **114**, 281–289 (1990).

100. R.W. Paynter, G.G. Ross, C. Boucher, J.F. Pageau, B. Stansfield, 'Analysis of a-C/B:H Layers Deposited in TdeV', *J. Nucl. Mater.*, **196–198**, 553–557 (1992).

101. A. Raveh, L. Martinu, S.C. Gujrathi, J.E. Klemberg-Sapieha, M.R. Wertheimer, 'Structure–Property Relationships in Dual-frequency Plasma Deposited on Hard a-C:H Films', *Surf. Coat. Technol.*, **53**, 275–282 (1992).

102. G. Dollinger, A. Bergmaier, L. Goergens, P. Neumaier, W. Vandervorst, S. Jakschik, 'High Resolution Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **219**, 333–343 (2004).

103. M. Juppo, P. Alen, M. Ritala, T. Sajavaara, J. Keinonen, M. Leskelä, 'Atomic Layer Deposition of Titanium Nitride Thin Films Using Tert-butylamine and Allylamine as Reductive Nitrogen Sources', *Electrochem. Solid State Lett.*, **5**, C4–C6 (2002).

104. P. Alen, M. Juppo, M. Ritala, T. Sajavaara, J. Keinonen, M. Leskelä, 'Tert-butylamine and Allylamine as Reductive Nitrogen Sources in Atomic Layer Deposition of TaN Thin Films', *J. Mater. Chem.*, **17**, 107–114 (2002).

105. P. Alen, M. Juppo, M. Ritala, T. Sajavaara, J. Keinonen, M. Leskelä, 'Atomic Layer Deposition of Ta(Al)N(C) Thin Films Using Trimethylaluminum as a Reducing Agent', *J. Electrochem. Soc.*, **148**, G566–G571 (2001).

106. G.W. Arnold, H.R. Westrich, W.H. Casey, 'Application of Ion Beam Analysis (RBS and ERD) to the Surface Chemistry Study of Leached Minerals', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 542–546 (1992).

107. H.J. Matzke, A. Turos, 'Mechanisms and Kinetics of Leaching of UO_2 in Water', *Solid State Ionics*, **49**, 189–194 (1991).

108. C. Neelmeijer, R. Grötzschel, E. Hentschel, R. Klabes, A. Kolitsch, E. Richter, 'Ion Beam Analysis of Steel Surfaces Modified by Nitrogen Ion Implantation', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 242–249 (1992).

109. H.J. Whitlow, E. Johansson, P. Anders Ingemarsson, S. Hogmark, 'Recoil Spectrometry of Oil-additive-associated Compositional Changes in Sliding Metal Surfaces', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **63**, 445–450 (1992).

110. J. Rickards, R. Trejo-Luna, M.E. Ortiz, E. Andrade, E. Chavez, E.P. Zironi, H. del Castillo, M. Sanchez, 'Hydrogen Pick-up in Ion Nitrided Steels', *Mater. Sci. Technol.*, **9**, 536–537 (1993).

111. R. Lappalainen, E.L. Wang, J.P. Hirvonen, 'Application of Nuclear Methods to the Study of Transfer Wear of PTFE on Stainless Steel and Silicon', *Surf. Interface Anal.*, **16**, 385–389 (1990).

112. H. Ronkainen, J. Likonen, J. Koskinen, 'Tribological Properties of Hard-carbon Films Produced by the Pulsed

Vacuum Arc Discharge Method', *Surf. Coat. Technol.*, **54/55**, 570–575 (1992).

113. F.L. Freire, C.M. Lepienski, C.A. Achete, Jr 'Characterization by Rutherford Backscattering, Elastic Recoil and Nuclear Reaction Analysis of Near-surface Modifications of Glasses Submitted to a DC Potential', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **68**, 227–230 (1992).

114. F. Schiettekatte, G.G. Ross, A. Chevarier, N. Chevarier, A. Plantier, 'Ion Beam Induced Depth Profile Modification of H, D and He Implanted in Be, C and Si', *Nucl. Instrum. Methods Phys. Res., Sect. B.*, **132**, 607–619 (1997).

115. G.G. Ross, I. Richard, 'Influence of the Ion-Beam-induced Desorption on the Quantitative Depth Profiling of Hydrogen in a Variety of Materials', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 603–607 (1992).

116. M. Naitoh, H. Ohnishi, Y. Ozaki, F. Shoji, K. Oura, 'Hydrogen-induced Reordering of the Si-(111)-$\sqrt{3} \times \sqrt{3}$-Al Surface Studied by ERDA/LEED', *Appl. Surf. Sci.*, **60/61**, 190–194 (1992).

117. Y. Zhang, G. Possnert, H.J. Whitlow, 'Measurements of the Mean Energy-Loss of Swift Heavy Ions in Carbon with High Precision', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **183**, 34–47 (2001).

118. M.W. Ruckman, M. Strongin, W.A. Lanford, W.C. Turner, '^{15}N Hydrogen Depth Profiling Measurements of Candidate Superconducting Supercollider Beam Pipe Materials', *J. Vac. Sci. Technol., A*, **13**(4), 1994 (1995).

FURTHER READING

D.K. Avasthi, W. Assmann, H. Huber, H.D. Mieskes, H. Nolte, 'Online Monitoring of Ion Induced Modifications by ERDA Using a Large Area Position-sensitive Detector Telescope', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **142**, 117–121 (1998).

W. Bohne, J. Röhrich, G. Röschert, 'The New Time-of-flight ERDA Set-up at the HMI-Berlin', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **139**, 219–224 (1998).

J.A. Davies, J. Forster, S.R. Walker, 'High-energy Heavy Ion Beams in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **139**, 120–127 (1998).

M.B. Lewis, E.H. Lee, 'Residual Gas and Ion Beam Analysis of Ion-irradiated Polymers', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **61**, 457–465 (1991).

A.M. Stoneham, 'Radiation Effects in Insulators', *Nucl. Instrum. Methods Phys. Res., Sect. A*, **91**, 1–11 (1994).

Elemental Analysis by Isotope Dilution

T.M. Bahrainwala and Z.R. Turel
The Institute of Science, Mumbai, India

1 Introduction 1
2 History 2
3 Theory 2
3.1 Principle of Isotope Dilution Analysis 2
3.2 Instruments Used 4
4 Statistical Evaluation of the Method with Respect to Accuracy, Precision and Sensitivity of Isotope Dilution Analysis 5
4.1 Factors Affecting Precision 5
4.2 Factors Affecting Accuracy 5
4.3 Factors Affecting Sensitivity 5
5 Limitations of the Method 5
6 Application of Isotope Dilution Analysis to the Determination of Some Representative Elements 5
6.1 Boron 5
6.2 Carbon 6
6.3 Nitrogen 6
6.4 Oxygen 6
6.5 Phosphorus 6
6.6 Sulfur 6
6.7 Argon 6
6.8 Potassium 7
6.9 Calcium 7
6.10 Chromium 7
6.11 Iron 7
6.12 Cobalt 7
6.13 Copper 8
6.14 Zinc 8
6.15 Arsenic 8
6.16 Selenium 8
6.17 Silver 8
6.18 Cadmium 9
6.19 Indium 9
6.20 Antimony 9
6.21 Tellurium 9
6.22 Iodine 9
6.23 Lanthanum 9
6.24 Hafnium 10
6.25 Tungsten 10
6.26 Rhenium 10
6.27 Gold 10
6.28 Mercury 10
6.29 Thallium 11
6.30 Lead 11
6.31 Uranium 11
6.32 Plutonium 12
7 Application of Isotope Dilution Analysis to Some Multielemental Determinations 12
Acknowledgments 14
Abbreviations and Acronyms 14
References 14

The measurement of extremely small traces of materials, which cannot be determined by normal analytical techniques, becomes possible and even relatively rapid and simple if the element can be made radioactive by the action of neutrons or charged particles. Nuclear techniques possess exceptional advantages in having extraordinary sensitivity, selectivity in detection and determination, speed and simplicity.

One of the nuclear techniques applied in trace and ultratrace analysis is isotope dilution analysis (IDA) which was first introduced by Hevesy and Hobbie in 1932. The great advantage of the method lies in the fact that it can be used even when the procedure involves loss of material or does not allow a complete separation of the element or compound. The isotope dilution method has found several applications in the determination of geological ages, fission yields and other problems in nuclear physics and chemistry.

Various techniques involved in IDA such as substoichiometry, liquid chromatography (LC) and radiometric titration are discussed along with the application of this technique to the estimation of elements in complex matrices as reported in the literature.

1 INTRODUCTION

The measurement of extremely small traces of materials which cannot be determined by normal analytical techniques becomes possible and even relatively simple and rapid if the element can be made radioactive by the action of neutrons or charged particles.

Nuclear techniques possess exceptional advantages in having extraordinary sensitivity in detection and determination, speed and simplicity. The certainty in identification is excellent as it involves half-life measurement, type of disintegration and energy measurement of the emitted radiation, which are unique properties of an isotope. The sensitivity in many cases is as high as 10^{-10} g and in special cases it attains 10^{-12} g.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Isotopes are elements (nuclides) having the same atomic number and different mass number. They are classified as stable isotopes (SIs) and radioisotopes (RIs). In IDA, a RI is used as the 'marked' analyte. It has the same behavior as the analyte and is called a 'tracer', 'spike' or 'label'.

One of the nuclear techniques applied in trace and ultratrace analysis is IDA. In the isotope dilution method the isotopic composition of an element remains constant throughout a series of chemical reactions. A SI may be used instead of a radioactive isotope, in which case the concentration must be determined by mass spectrometry (MS) or by some other appropriate procedure. The great advantage of the method lies in the fact that it can be used even when the procedure involves loss of material or does not allow a complete separation of the element or compound. It is therefore particularly useful in the measurement of rare earths, fission products and complex mixtures of organic or biomedical compounds.

IDA and reverse isotope dilution analysis (RIDA) have found useful applications with SIs and radioactive isotopes. With the development of MS, and the availability of separated SIs and radioactive isotopes, the isotope dilution method has found several applications in the determination of geological ages, fission yields and in other problems in nuclear physics. Accuracies of a few per cent may be obtained for some elements that are present in concentrations as low as parts per billion (ppb) or parts per trillion (ppt).

2 HISTORY

The term IDA was first introduced in 1940 by Rittenburg and Foster. Hevesy and Hobbie[1,2] first introduced the method for determining traces of lead present in a mineral in about one part in 10^6; lead was measured by adding Ra D and then depositing PbO_2 anodically. Haissinsky[3] used a similar technique for measuring potassium in the presence of other alkali metals. Similarly, ^{35}S was used to analyze a mixture of benzyl derivatives of sulfides, sulfones and sulfoxides.[4] Other measurements reported in the literature during the initial stages of the development of the method include naphthalene in coal tar,[5] γ-hexachlorobenzene,[6] penicillin,[7] vitamin B_{12} and other organic compounds. In the case of penicillin, the molecule was labelled with ^{13}C.

A great advantage of IDA is that it does not require complete separation of the analyte from the sample. However, the quantity of the analyte separated from the matrix must be determined using an appropriate method, e.g. by gravimetry or spectrometry. This is a disadvantageous feature of the classical form of IDA. To overcome this problem, substoichiometric isotope dilution analysis (Subst-IDA) was proposed in 1958 by Ruzicka and Stary[8] where only the measurement of radioactivities is required. The procedure is rapid and easy, and effects of interfering substances are lowered.[8]

In 1974, the 'sub- and superequivalence (SSE) method' was introduced into IDA.[9] It does not need such a strict reliance on the stability constant as the Subst-IDA mentioned above. Various analytical techniques, such as solvent extraction, ion exchange and precipitation, have been employed for the separation.

3 THEORY

3.1 Principle of Isotope Dilution Analysis

The various types of IDA such as classical IDA, Subst-IDA and sub- and superequivalence isotope dilution analysis (SSEIDA) are explained below using mathematical equations. Their characteristics are also explained.

3.1.1 Classical Isotope Dilution Analysis

In this method, an unknown amount (x) of analyte in the sample is spiked with a RI (A = radioactivity; weight is negligible compared with the analyte in the sample). The specific activity of this analyte is $S_x = A/x$. To the sample, a known amount (y) of the analyte is added, so the specific activity of the mixture becomes $S_{x+y} = A/(x + y)$. An appropriate amount of reagent which reacts with the analyte is added next, and the pure product is separated. To remove any impurity from the matrix, purification must be carried out repeatedly. After that, the weight (m_{x+y}) and the radioactivity (a_{x+y}) of the product are measured. As the specific activity before and after the separation does not change, we have Equation (1):

$$\frac{A}{x+y} = \frac{a_{x+y}}{m_{x+y}} \tag{1}$$

The unknown quantity x to be determined is calculated from Equation (2):

$$x = \frac{Am_{x+y}}{a_{x+y} - y} \tag{2}$$

When the sample contains other components having properties similar to the analyte, it is difficult to obtain complete (100%) recovery of the pure analyte. In such a case, a yield correction would be made in the conventional analysis. However, this correction is not needed in IDA because only the specific activity needs to be known, as shown in the above equations. Therefore, complete separation and recovery are not required and, as a result, repeated purification steps may be made

to remove impurities. This is the major advantage of IDA. The weight m_{x+y} of the pure separated analyte must be determined by an appropriate method; however, this operation is troublesome and time-consuming. This situation is improved by Subst-IDA.

3.1.2 Substoichiometric Isotope Dilution Analysis

In this method, two samples containing an unknown amount (x) of the analyte are spiked ($S_x = A/x$) as in classical IDA, and standard solutions containing known amounts (y, $3y$) of the analyte are added to them. The specific activities of the diluted solutions become $S_{x+y} = A/(x+y)$ and $S_{x+3y} = A/(x+3y)$. To each, substoichiometric amounts (M) of the reagent are added, then the products are separated from unreacted species. Assuming that the quantities (m_{x+y}, m_{x+3y}) separated from each sample are the same, namely $m_{x+y} = m_{x+3y} = m$, as the reagent used reacted quantitatively with the reactants, and the radioactivities obtained as a_{x+y}, a_{x+3y}, Equations (3) and (4) hold good as the specific activity does not change before and after the separation:

$$\frac{A}{x+y} = \frac{a_{x+y}}{m} \tag{3}$$

$$\frac{A}{x+3y} = \frac{a_{x+3y}}{m} \tag{4}$$

From these equations, we obtain Equation (5):

$$x = \frac{y(3a_{x+3y} - a_{x+y})}{a_{x+y} - a_{x+3y}} \tag{5}$$

Equation (5) shows that x is obtained from the radioactivity without measurement of the weight m. This is a major advantage of Subst-IDA. Another point is that the amount of reagent used is substoichiometric compared with the quantity of the analyte and this leads to a decrease in consumption of the reagent by the interfering substances which react with reagent in the same way as the analyte, resulting in improvements in accuracy, precision and sensitivity.

The reagent must react with the analyte almost completely (nearly 100%) and be separated completely if the assumption $m_{x+y} = m_{x+2y}$ is to hold good. For this purpose, the stability constant of the product analyte and the reagent must be sufficiently large. Not many reagents satisfy this condition. This is a disadvantage of Subst-IDA, and to overcome this problem, SSEIDA was proposed.

3.1.3 Sub- and Superequivalence Isotope Dilution Analysis

The sample containing an unknown amount (x) of the spiked analyte (A = radioactivity, weight is negligible compared with x) is placed in each test tube in the first series. They are isotopically diluted by the addition of regularly increasing known amounts iy ($i = 0, 1, 2, \ldots$) of the analyte so that their specific activities become $A/(x+iy)$. In the second series, k times the unknown amount of the first series is placed in each test tube (kA = radioactivity, kx = amount of analyte); they are not isotopically diluted. All solutions in both series are brought to the same volume and same acidity by adding the appropriate amount of solvent.

A definite amount of the reagents is added to each of them. The quantity of reagent is not always necessarily substoichiometric and is allowed to be larger (superequivalent) than that of the analyte. After the separation of the products (first series, m_{x+iy}; and second series, m_{kx}), their radioactivities are measured (first series, a_{x+iy}; second series, a_{kx}). As the specific activities do not change before and after the separation, we have Equations (6) and (7)

$$\frac{A}{x+iy} = \frac{a_{x+iy}}{m_{x+iy}} \tag{6}$$

$$\frac{kA}{kx} = \frac{a_{kx}}{m_{kx}} \tag{7}$$

for the first and second series, respectively. From these equations, we obtain Equation (8):

$$\frac{a_{kx}}{a_{x+iy}} = \left(\frac{m_{kx}}{m_{x+iy}} \quad \left(\frac{1}{x} \quad iy + \frac{m_{kx}}{m_{x+iy}}\right.\right. \tag{8}$$

The quantities m_{x+iy} and m_{kx} generally depend on the concentration of the analyte in the solution, namely $(x+iy)/V$ and kx/V. However, if one sample (jth increment) in the first series has the same concentration (isoconcentration point) of the analyte as that in the second series, the degree of reaction with the reagent and the degree of separation must be the same. This means that Equation (9) holds:

$$m_{x+jy} = m_{kx} \tag{9}$$

In this special case (at the isoconcentration point), Equation (10) is derived from Equations (6) and (7):

$$\frac{a_{kx}}{a_{x+jy}} = k \tag{10}$$

Substituting Equations (9) and (10) into Equation (8), Equation (11) is obtained:

$$x = jy(k-1) \tag{11}$$

Thus, to determine x, it is necessary to find j. The value j is virtual and does not necessarily have to be an integer (the value of x is obtained by a graphical method). The number of samples of the second series is not necessarily

large, but several samples are generally used to obtain a reliable value by averaging.

As SSEIDA does not impose strict conditions on the reactivity and the quantity of the reagent, it has a wide range of choice of reagents compared with Subst-IDA. On the other hand, determination by classic IDA and Subst-IDA is possible using one or two samples (except for the multiple method), but more than four samples are necessary in SSEIDA to find the isoconcentration point. Therefore, this method demands more time and effort, and is not suitable for the analysis of small quantities of samples.

As is known from Equations (6) and (7), the effect of interfering substances is not negligible because their quantities in the matrix are different in the first and second series. Apart from the this disadvantage, SSEIDA seems to be an excellent method when sensitive, precise and accurate determination by Subst-IDA is difficult.

3.1.4 Stable Isotope Dilution Analysis

The principle of stable IDA is almost the same as that of radioisotope dilution analysis (radio-IDA). The difference is that radio-IDA makes use of the change of specific activity but stable IDA must depend on the change in the isotope ratio. A mass spectrometer is the only instrument sensitive enough to measure the isotope ratio, so this method is also called isotope dilution mass spectrometry (IDMS).

Two isotopes (1 and 2) are selected from the various isotopes comprising the analyte to monitor the isotope ratio. Their isotopic abundance in a sample containing an unknown amount (x) of the analyte are taken as $\theta_{x(1)}$, $\theta_{x(2)}$. To the sample, a known amount (y) of the analyte having an enriched SI of composition different from that of the sample is added, where the isotopic abundances of the SI are $\theta_{y(1)}$, $\theta_{y(2)}$, respectively. The isotope ratio, R_{x+y}, of 1 and 2 in the mixture after mixing with the diluent is calculated using Equation (12):

$$R_{x+y} = \frac{x\theta_{x(1)} + y\theta_{y(1)}}{x\theta_{x(2)} + y\theta_{y(2)}} \tag{12}$$

From this equation, we obtain Equation (13):

$$x = \frac{[\theta_{y(1)} - R_{x+y}\theta_{y(2)}]y}{R_{x+y}\theta_{x(2)} - \theta_{x(1)}} \tag{13}$$

As $\theta_{x(1)}$, $\theta_{x(2)}$, $\theta_{y(1)}$, $\theta_{y(2)}$ and R_{x+y} $\{= [m_{x+y}, \theta_{x+y(1)}] / [m_{x+y}, \theta_{x+y(2)}]\}$ of product m_{x+y}, which is separated from matrix elements, are determined by MS and y is known, x is calculated from Equation (13).

3.2 Instruments Used

3.2.1 Radioisotope Dilution Analysis

Some radio-IDA needs to use RIs, so instruments for measuring radioactivity are necessary. An appropriate instrument is selected depending on the kind and energy of radiation.

For example, a NaI(Tl) scintillation counter is commonly used to measure γ-rays from RIs such as ^{125}I and a liquid scintillation counter is used for low-energy β-rays from ^{3}H, ^{14}C, etc. A Geiger–Müller counter is used for high-energy β-rays from ^{204}Tl etc. A high-resolution HPGe semiconductor detector is able to measure γ-rays of various energies at the same time.

3.2.2 Stable Isotope Dilution Analysis

Mass spectrometers are used in stable IDA to measure isotopic ratios. The instrument mainly consist of three parts, for ionization, separation of ions and detection. For ionization, glow discharge, inductively coupled plasma (ICP) and corona discharge are suitable for solid, liquid and gaseous samples. MS combined with gas chromatography (GC) and LC are useful for preliminary separation and concentration of the analyte.

An example of Subst-IDA for the determination of iodine in common salts using ^{131}I as tracer is given below.

About 1 g, accurately weighed, of powdered salt sample was dissolved in 25 mL of doubly distilled water. Carrier solution (10–60 μg of iodide as NaI) and a suitable aliquot of ^{131}I were added and the whole system was warmed to about 50 °C. After cooling, 5 mL of tartaric acid and 2 mL of potassium iodate solutions were added. This mixture acts as a mild oxidizing agent and oxidizes only I^- to I_2 and leaves chloride as such. Initially an attempt to use H_2SO_4 instead of tartaric acid was made, but this resulted in large-scale oxidation of chloride, causing interference in the substoichiometric stage. The iodine liberated was immediately extracted into CCl_4. To the organic phase 0.1 mL of $AgNO_3$ solution (in substoichiometric amount) and 10 mL water were added. Typically about 15 μg or its integral multiple amounts of Ag were added for substoichiometric extraction. This gave approximately 80% extraction. After thorough mixing, a colloidal solution of AgI was formed in the aqueous phase, to which, after separation, 1 mL of sodium thiosulfate solution was added. This resulted in the formation of NaI in the aqueous phase, which was then counted on a 1.75×2.00 in a well-type NaI(Tl) detector.

When analyzing actual samples of common salts, however, a standard solution of NaCl with a known concentration of iodine (as NaI) was simultaneously analyzed according to the standard procedure. Finally, the specific activities of the standard and the sample

aliquots were compared and the concentration of iodine was calculated.

In order to ascertain the accuracy and precision of the method, several suitable mixtures of NaCl and NaI were analyzed. To a solution of nearly 1 g of NaCl, an aqueous solution containing 3–60 μg of NaI was added and its iodine content was determined.

Iodine is important as an essential trace element for humans, and its deficiency leads to goiter, a prominent disorder of the thyroid gland.[10,11] The important sources of iodine are foodstuffs and beverages that contain seafoods and common salt. In order to circumvent the deficiency of iodine in many parts of the world, several countries have launched a program of dietary supplementation. Even such programs have not always been successful in countering goiter in some countries.[12] In India, the consumption of seafood is very low and common salt is the only source of iodine. The Government of India has now launched a program for the consumption of iodized salts by the general public. Iodine has been determined by other methods also.[13–15]

4 STATISTICAL EVALUATION OF THE METHOD WITH RESPECT TO ACCURACY, PRECISION AND SENSITIVITY OF ISOTOPE DILUTION ANALYSIS

4.1 Factors Affecting Precision

The precision of the weighing of the separated analyte is important. However, with Subst-IDA weighing is not required.

Errors in counting radioactivity also affect precision. The measurement should last long enough to lower the error E to below 1% according to equation (14):

$$T = \frac{10\,000}{E^2 r} \tag{14}$$

where T = period of measurement and r = counting rate.

4.2 Factors Affecting Accuracy

The tracer and carrier (diluent) must behave in the same manner as the analyte in the sample in order to obtain a constant specific activity throughout the procedure. For this purpose, properties such as valency state of the element and the kind of reagent must be made uniform at an early stage of analysis. Isotopic equilibrium between the tracer and the analyte must be complete.

Loss of analyte (in unknown amount) may be caused during dissolution by adsorption on the container wall, vaporization, etc. When an unknown amount of an analyte is lost before dilution with the tracer, it strongly affects the accuracy of the method. Therefore, it is desirable to mix the tracer and diluent with the sample in the earliest stages of the operation. Loss after exchange does not affect the accuracy.

In each analysis, a simple clean room is desirable to prevent contamination from the reagents, apparatus, atmosphere, etc. which affects the accuracy. It must be ensured that the radioactivity comes from only one type of nuclide, i.e. the sample is radiochemically pure. When there is another element which emits radiation similar to that of the tracer in the sample, it affects the accuracy. In these cases, the tracer should be purified.

4.3 Factors Affecting Sensitivity

The volume of the reaction mixture should be minimized to increase the sensitivity. The presence of interfering ions affects the sensitivity, but the effect is less in Subst-IDA and SSEIDA, so the sensitivity of these methods is higher than that of classical IDA.

5 LIMITATIONS OF THE METHOD

The use of the proposed method for the determination of amounts of metals larger than 1 mg is not advantageous, because physicochemical analytical methods may be used whose precision is, under these conditions, better than the precision of radiometric methods, which is controlled by the statistical character of the radioactive decay. The sensitivity of the method is limited by two factors: (a) the organic reagent employed cannot be diluted infinitely and (b) at present, RIs of sufficiently high specific activity are not at hand for all cases. The latter limitation has been solved to a certain extent by the development of procedures for the preparation of carrier-free RIs.

6 APPLICATION OF ISOTOPE DILUTION ANALYSIS TO THE DETERMINATION OF SOME REPRESENTATIVE ELEMENTS

6.1 Boron

A method has been described for determining traces of boron in silicon.[16] The sample was dissolved in aqueous sodium hydroxide in the presence of ^{10}B tracer, and boron was separated by modification of Morrison and Rupp's method of electrolysis through a cation-exchange membrane. Boron was finally determined by making isotope dilution measurements with a mass spectrometer with a thermal ionization source. The method has the advantage that quantitative separation of boron was not

necessary. Amounts of boron down to 0.001 μg can be determined to within ±30%.

6.2 Carbon

The sensitivity and precision of field-desorption MS are greatly improved by using an isotopically labelled standard compound and averaging the m/z values for a large number (e.g. 100 or 210) of spectral scans.[17] Cationization involves the production of molecular ions with enhanced m/z values, e.g. when a mixture of glucose, [^{13}C] glucose and NaI was dissolved in water, and the solution was applied to a high-temperature field-desorption emitter, the spectrum contained peaks for the species $(M + Na)^+$. The use of such peaks, with correction for [^{13}C] glucose in the nominally unlabelled glucose, permits the determination of glucose with a mean error of ±7%. The method is considered to show promise for biomedical analysis.

Small amounts of carbon in metals have been determined by the isotope dilution method by using an omegatron.[18] From 0.4 ppm to 6% of carbon in steel, iron, copper, tin, nickel, tungsten or molybdenum or in tungsten carbide was determined by heating the sample (in the form of a thin ribbon) at 1400 K in an atmosphere containing a known concentration of ^{13}CO. After 3 h the concentration of ^{13}CO and ^{12}CO in this atmosphere were determined by MS using a dynamic mass spectrometer (omegatron). The coefficients of variation were <10%.

A method utilizing IDMS has been developed for the determination of carbon in sodium.[19] The details of the method are described for the determination of elemental carbon employing Van Slyke oxidation, although, by choice of spike material, other forms of carbon may be determined. Accuracy and precision in the ranges of 50 ± 10 to 150 ± 25 μg of carbon have been demonstrated. The determination was not affected by the loss of generated carbon dioxide or by contamination from any source except carbon.

6.3 Nitrogen

The method described has been applied to the determination of nitrogen in Nb, W, Mo, La, Ce, Re, Fe, Cr and Ni.[20] It is based on MS measurement of changes (effected by dilution with the nitrogen from the sample) in the isotopic composition of ^{15}N used as an indicator. Nb was used as carrier for the indicator (forming nitrides) for improving the isotopic exchange between indicator and sample. The sensitivity was 1 ppm and the precision was 3% for 5 g samples containing 10 ppm of nitrogen.

6.4 Oxygen

Oxygen has been determined in iron and steel by a modified isotope dilution technique.[21] Previous discrepancies in the determination of oxygen in steel by an isotope dilution and vacuum fusion technique were traced to adsorption of C^{18}O on deposits in the furnace. By treating the specimen at dull-red heat with a measured volume of oxygen enriched in ^{18}O and allowing it to cool before lowering it into the graphite crucible, good agreement was obtained between the two methods.

6.5 Phosphorus

In the method described, ^{32}P-labeled sodium phosphate was used to determine 10–50 ppm of P in 3–5 g samples of triuranium octaoxide.[22] The error was <3% and the reproducibility was satisfactory. The method depends on the precipitation of the P as molybdophosphate, which is converted into $MgNH_4PO_4$ for counting.

The use of the isotope dilution technique to eliminate problems associated with demands for a high load capacity in column-coupling capillary isotachophoresis was studied.[23] Determinations of phosphate present in a model mixture and in white wine served for the evaluation of this approach to quantitative analysis in isotachophoresis. A high selectivity of the analysis with acceptable accuracy and precision of the analysis of complex ionic mixtures in a short time are obvious advantages of this approach.

6.6 Sulfur

Adsorbed sulfate in soil was determined by adding a small quantity of carrier-free ^{35}S as sulfate to a solution in equilibrium with the soil.[24] The adsorbed sulfate was calculated using the isotopic dilution law from total added activity and the activity in the equilibrium solution. Total added activity and activity in the equilibrium fractions were determined by suspending soil in water overnight, adding a solution containing ^{35}S, diluted to an activity close to 6.7 kBq, shaking the suspension for 2 h and centrifugation at 9000 rpm for 10 min. One part of the supernatant was used for liquid scintillation analysis and mixed with Quick Safe A. The remaining supernatant was filtered through a 0.2 μm membrane filter and sulfate was determined using a Dionex 2000i ion chromatograph equipped with a AS4 anion-exchange separation column. Phosphate-extractable sulfate was extracted with 10 mM calcium dihydrogenphosphate. Part of the supernatant was filtered through a 0.2 μm filter and analyzed by ion chromatography as before. The two methods gave corresponding results for fresh soil but for air-dried soil the extraction method yielded approximately 25% more adsorbed sulfate than the ^{35}S method.

6.7 Argon

Radiogenic Ar was determined in rock samples by the method of isotope dilution with an $AlCl_3$ standard

containing ^{38}Ar.[25] After exposure to a neutron flux, 1 g of $AlCl_3$ contained $(4.48 \pm 0.14) \times 10^{-5}$ mL of ^{38}Ar. Rock samples mixed with 0.5 g of irradiated $AlCl_3$ were placed in a silica tube attached to a vacuum system, and the isotope content of the released Ar was determined in a mass spectrometer. The ^{40}Ar content of the rock sample was calculated with a correction for atmospheric Ar.

6.8 Potassium

Small amounts of K have been determined by IDA, involving an exhaustive chemical procedure.[26] A method for the determination of K in minerals difficult to decompose by chemical means is described. Procedures for beryl and tourmaline are given. A method has also been described for the mass spectrometric determination of K in minerals containing <1% of K. The results agree reasonably well with those obtained by chemical methods.

6.9 Calcium

The IDMS determination of low concentrations of Ca in minerals has been discussed.[27] The sample was decomposed with $HF-HClO_4$ after adding ^{42}Ca, then total Ca was separated from K and various other metals on a column of strongly acidic ion-exchange resin, complete removal of K being essential. The Ca content was calculated from the ratio of the peak intensity of ^{40}Ca to that of ^{42}Ca, as determined by MS. The dependence of the analytical error on the MS measurements is discussed. Calcium (0.15–0.004%) was determined in two feldspars and one lepidolite. The results (together with their precision) are reported.

An isotope dilution technique utilizing thermal ionization MS has been developed for the accurate determination of calcium in synthetic and serum samples at the 100 mg g^{-1} level.[28] Calcium was separated from a serum matrix by destruction of the organic matter with $HClO_4$ and HNO_3 followed by ion-exchange separation from interferences using AG50W-X8 100–200-mesh resin. An MS isotopic analysis procedure was developed using a $Ca(NO_3)_2$ solution deposited on Re sample filaments in a triple-filament ion source. The relative error between calculated and experimentally determined concentrations in synthetic calcium solutions was $\leq$0.1%. The 95% limit of error for a single analysis was approximately 0.2% for synthetic and serum samples. A comparison of the isotope dilution data with concurrently determined atomic absorption data from several clinical and independent laboratories is presented.

6.10 Chromium

A method has been described for the determination of chromium and its enriched SI in human urine by IDMS.[29] A volatile chelate was formed using trifluoroacetylacetone (TFA) and the fragment ions corresponding to $Cr(TFA)^{2+}$ in the m/z 356–360 region were monitored. The chelate was thermally stable and exhibited no memory effects when the isotope ratios changed. The detection limit for the method was 0.03 ng g^{-1} of Cr and the accuracy was verified using certified reference materials (CRMs) and by an independent method. The method was highly specific for chromium, owing to the combined properties of the chelating agent, chromatographic column and mass-specific detector. In addition to total chromium determinations, the method can also be used to quantitate enriched SIs of chromium used as metabolic tags in tracer experiments in human nutrition studies.

6.11 Iron

IDA by ion exchange has been used for the determination of traces of iron.[30] A theoretical evaluation of this method is given, from which the conditions for the determination of a number of metals can be predicted. The selectivity of the method was further increased by the use of masking and precipitating agents. The proposed method has been verified by the determination of iron using EDTA and Dowex 50 cation exchanger. The average precision of mean values for amounts of 10^{-7} g of Fe per 5 mL was 1.2%, for amounts of 10^{-8} g of Fe per 5 mL it was 4.0% and for amounts of 10^{-9} g of Fe per 5 mL it was 4.1%. Even large excesses of many metals did not interfere in the analysis. The method has been applied to the analysis of NaI(Tl) crystals. The procedure which has been developed is very simple, and is far more sensitive than colorimetry, activation analysis or other methods used up to now for determining traces of iron.

6.12 Cobalt

IDA has been combined with solvent extraction and spectrophotometry to produce an effective general method for the determination of constituents that will not extract quantitatively.[31] Combining isotope dilution with extraction of the cobalt–2-nitroso-1-naphthol complex into chloroform resulted in an accurate and reliable method for the determination of small amounts of cobalt in alloys of high copper content and in ingot irons. ^{60}Co was used as a radioactive tracer.

An isotope dilution method for the determination of cobalt employing ^{60}Co as the tracer and a spectrophotometric procedure based on the extraction of cobalt thiocyanate with methyl isobutyl ketone (MIBK) has

been developed.[32] The extraction was carried out under neutral or slightly basic pH conditions and this made the procedure virtually specific for cobalt. The method was suitable for the determination of cobalt at concentrations from a few parts per million upwards in biological materials, vitamin B_{12}, salts, metals, etc. When a well-type scintillation detector was used for measuring the radiation of ^{60}Co in liquid samples, the procedure was relatively simple and rapid. The results obtained in testing the behavior of a number of metals and the various stages of the extraction of cobalt thiocyanate with MIBK in the presence of ^{60}Co as tracer are also presented.

6.13 Copper

Copper has been determined in serum samples by radio-IDA and incomplete extraction.[33] On decomposition of serum from a venous blood sample, Cu was labeled with ^{64}Cu. Dithizone in CCl_4 was used for extracting Cu at substoichiometric levels. The original Cu concentration was calculated from Ruzicka's equation. The method is compared with a colorimetric method.

6.14 Zinc

A new separation method using ion exchange based on complexation in a resin matrix has been used for the substoichiometric determination of Zn by the isotope dilution method using a chelating agent-loaded resin.[34] A resin loaded with 8-quinolinol-5-sulfonic acid was also shown to be applicable to the selective preconcentration of chalcophile elements in natural water samples. An additional advantage of this resin was that the ion-exchange capacity can be varied according to the desired purpose and therefore substoichiometric separation may be achieved. The possibility of the determination of zinc with substoichiometric amounts of 8-quinolinol-5-sulfonic acid was investigated.

The comparative extractability of Zn with potassium salts of ethyl, propyl, butyl, pentyl and benzyl xanthates in the pH range 3.5–9.0 into chloroform has been studied, employing a sensitive and rapid substoichiometric radiochemical method.[35] The extent of reproducibility was tested in each case. The effect of associated ions on the extraction was studied. The amount of Zn present in the standard solutions was determined employing each xanthate separately. The Zn content present in geological water samples in and around Tirupati was determined by the method developed and compared with the values obtained by atomic absorption spectroscopy (AAS).

6.15 Arsenic

Arsenic has been determined in biological and environmental standard reference materials (SRMs) by substoichiometric solvent extraction using toluene-3,4-dithiol in benzene.[36] A radiochemical solvent extraction procedure has been developed for the determination of As(III) using ^{76}As tracer. It is based on the complexation of As(III) with toluene-3,4-dithiol at pH 2 and subsequent extraction in benzene. The effects of various parameters such as pH, time of equilibration, nature of solvent, quantitative character and interferences were studied. The method has been further developed into Subst-IDA for the determination of As at <1 mg levels and employed for the analysis of several environmental and biological SRMs from the National Institute of Standards and Technology (NIST) (USA), the International Atomic Energy Agency (IAEA) (Vienna) and NIES (Japan).

6.16 Selenium

Selenium has been determined in plant materials by a fluorimetric method employing ^{73}Se.[37] An exhaustive chemical procedure is described. A 0.05% solution of purified naphthalene-2,3-diamine in 0.1 M HCl (0.5 mL) was employed for the fluorimetric method. The fluorescence was compared against a reagent blank. The isotope dilution results were used to correct for losses during the combustion and extraction. Of 20 metal ions tested, serious interference was observed from Cr^{3+}, Sb^{3+} and Sn^{4+}. The standard deviation (SD) for the determination of 0.2 μg of Se was 0.04 μg.

The determination of Se by the SSE method of IDA involves the extraction of a test solution of Se labeled with ^{75}Se.[38] Sodiumdiethyldithiocarbamate was used as reagent and CCl_4 as a solvent. The activity was counted on an NaI(Tl) detector. The Se concentration was calculated from a calibration graph of concentration against the ratio A to A' where A = activity in the organic phase and A' = repeated measurement in a fresh aliquot. For 1.92 ppm of Se, the SD was 0.1371 ppm.

A method is described for the determination of 10^{-2} ppm levels of Se in organic samples.[39] The method consists of a micromodification of the spectrophotometric method employing 3,3′-diaminobenzidine as reagent for extracting selenium(IV), and incorporates an isotope dilution technique to compensate for the unavoidable losses that occur during the course of the determination of Se.

6.17 Silver

Isotope dilution electroanalysis has been used for the determination of silver at microgram levels.[40] The method consists in electrolysis at a controlled potential with a silver-wire cathode and a platinum wire anode and the measurement of the specific activity of the Ag deposited on the cathode. Two microelectrolysis cells were connected in series, one containing an amount

of inactive Ag (x) plus a known amount of ^{110}Ag (y) and the other containing the same amount (y) of ^{110}Ag. In the presence of Cu^{2+}, Ni^{2+} and Cd^{2+}, Ag may be determined in the range 10^{-3}–10^{-6} g mL^{-1} of electrolyte with a coefficient of variation of 2.3% at the 10^{-5} g mL^{-1} level.

6.18 Cadmium

Cadmium has been determined in SRMs.[41] The method involves four polyethylene samples with nominal Cd contents from 40 to 400 ppm, which were supplied as small wire cuts each of about 10 mg. Each sample was mixed for 1 h and bottled in 30 g units. Homogeneity was tested by taking 15 pieces of 60–250 μg from each of the four bottles and analyzing them by solid-sampling Zeeman-corrected AAS, with measurement at 326.1 nm. These results were tested statistically. Samples VDA-001, -002 and -003 gave normally distributed results and homogeneity factors and minimum representative sample weights (13–27 mg) were calculated. The fourth sample, VDA-004, was not normally distributed, with some microheterogeneity, but could still be used if sample weights ≥1 g were used. For certification, the Cd contents were determined by IDMS. Samples were oxidized with H_2SO_4–HNO_3–H_2O_2, the Cd was plated on to Pt and the deposit uncertainties were in the range of 1.2–12 ppm. Certified values were given and were confirmed.

The feasibility of isotope dilution inductively coupled plasma mass spectrometry (IDICPMS) has been established by determination of Cd in soil extracts from the ^{106}Cd to ^{111}Cd isotope ratio measured before and after addition of ^{106}Cd,[42] reasonable agreement with results by graphite furnace atomic absorption spectroscopy (GFAAS) being obtained. A program in Basic for individual and off-line isotope dilution calculations is appended.

6.19 Indium

Subst-IDA has been described employing an ion-exchange method for the determination of traces of indium.[43] A highly selective, substoichiometric method for the determination of indium has been developed. The procedure consists in the addition of a substoichiometric amount of EDTA to the test solution labeled with a known amount of radioindium, followed by separation of the complex formed from the excess of unreacted indium on a Dowex-50 column. Using a preliminary cupferron extraction, the procedure became highly selective. The average precision of the determination of indium for an amount of 1.1 μg in 10 mL was 1.6% in the presence of an excess of many metals. The maximum sensitivity of the method was about 5×10^{-5} μg per 10 mL.

6.20 Antimony

A rapid and selective substoichiometric radiochemical method has been developed for the determination of microgram amounts of antimony employing potassium ethyl xanthate as a reagent and chloroform as an extractant from sulfuric acid medium.[44] The effect of associated ions on the extraction was studied. The method developed was successfully applied to determine the antimony content in standard solutions and synthetic mixture with an average error of ±2.07%.

6.21 Tellurium

A radiometric method based on redox substoichiometry has been developed for the determination of tellurium.[45] The oxidation of tellurium(IV) to tellurium(VI) with potassium dichromate was employed as the substoichiometric reaction, followed by tributyl phosphate extraction of the unreacted tellurium(IV) for substoichiometric isolation. The oxidation of tellurium(VI) with dichromate under the substoichiometric conditions was incomplete. The chemical yield of tellurium(VI) produced by the oxidation of tellurium(IV) with dichromate was 83%. It was found that a tellurium content of 50–160 μg could be determined with an accuracy of ±1.3% by means of the comparison method.

6.22 Iodine

A Subst-IDA method was developed for the determination of iodine in different brands of common salt.[46] An aqueous salt solution containing ^{131}I tracer and NaI as carrier was oxidized by tartaric acid and KIO_3 and the liberated iodine was extracted with CCl_4. To the extract an aqueous solution of $AgNO_3$ was added in substoichiometric amount to obtain a colloidal solution of AgI. On adding sodium thiosulfate solution, the NaI so formed passed into aqueous solution, which was then counted. Several different brands of salt were analyzed. The method was especially suitable for the determination of microgram amounts of iodide in the presence of excess of chloride.

6.23 Lanthanum

The Subst-IDA method has been used for the determination of La in environmental samples using cupferron in chloroform.[47] Dried powder was fused with sodium peroxide and 99.9% scandium oxide as holdback carrier. The fusion mixture was allowed to melt with occasional stirring, the resulting cake was cooled in H_2O_2 and 6 M HCl before dilution with water. The ^{140}La tracer was treated with La carrier solution, water and 0.02 M aqueous cupferron. Ammonium buffer of pH 9.0 was added

and the resulting solution was shaken with chloroform for 2 min. After equilibration, a brown-red complex was formed with clear phase separation. A portion of the organic and aqueous phases were counted at the 487 and 1596 MeV photopeaks of ^{140}La using an NaI(Tl) well-type detector in conjunction with a multichannel analyzer (MCA). The calibration graphs were linear for 22–550 ng and 1.75–44 μg of La. The method was applied substoichiometrically to four cement dust particulate samples and SRMs of environmental importance; the relative SD was <8%. Interference masking and elimination methods are discussed.

6.24 Hafnium

Hafnium has been determined in Ze-4 and Ze-6 zirconium metal standards by stable IDA.[48] Hf values obtained for Ze-4 and Ze-6 standards were 77.4 ± 3.9 (tentative) and 166 ± 17 ppm, respectively.

Hafnium has been separated from zirconium by a chromatographic method and determined by IDA.[49] The method is based on the different stabilities of the sulfate complexes. The best separation was obtained by using 0.65 N H_2SO_4 for washing the cationic column. With 0.7 N H_2SO_4 the separation was significantly quicker, although not quantitative, but this did not interfere with the method of isotope dilution. The relative error was <1% of Hf in mixture containing approximately 10% of Hf, but higher if the Hf concentration was <3%. The method was used for the determination of Hf in eudialyte and Zr and also in mixed oxides of Zr and Hf of various concentrations.

6.25 Tungsten

An isotope dilution procedure using ^{185}W for the determination of W in high-alloy steels has been described.[50] Tungsten was determined spectrophotometrically with quinol in H_2SO_4 after homogeneous precipitation from HNO_3–H_2O_2. No further chemical separation was required for Mo–V–W steels, but for steels containing Nb, Ta and/or Ti, chemical separation of the elements that would interfere with subsequent spectrophotometric determination was necessary. The method was successfully tested on a number of National Bureau of Standards (NBS) steels to which various amounts of Ta and Ti were added.

6.26 Rhenium

Rhenium has been determined by a surface ionization isotope dilution method.[51] Down to 1 μg of Re can be determined by MS by heating (at 2200 °C) a tungsten filament on which the metal has been deposited. Re in 10 mg samples of Re–W alloys (3–25% of Re) has been determined by an isotope dilution method (with use of ^{187}Re) with a SD of ±2%.

A radiotracer method has been used for the determination of Re.[52] Radiometric correction and interpolation methods have been suggested for rhenium determination using a ^{188}Re (16.9 h) radiotracer. The radioactive isotope was obtained in an ^{188}Re generator consisting of a glass column filled with aluminum oxide with a parent isotope ^{188}W (69.4 days) firmly adsorbed. The isotope ^{188}Re formed was eluted from the generator with a 0.9% solution of sodium chloride. Isotopic equilibrium among stable and radioactive rhenium (in the form of perrhenate ions) was obtained just after mixing the two solutions. A scintillation detector with an NaI crystal was used for γ-counting.

Rhenium has been extracted with a tetraphenylphosphonium complex into dichloroethane. Liquid scintillation counting was used. By this method, 1–100 μg of rhenium(VII) in the analyzed sample was determined. The relative SD was 0.05–0.09% over the whole concentration range.

6.27 Gold

Gold has been determined by radioactivation analysis accompanied by the isotope dilution method.[53] The sample (containing M_x of the element to be determined) and the standard (containing M_s) were irradiated under identical conditions and then each was mixed with a large amount of carrier, an aliquot was separated and their activities A_x and A_s were measured. The value of M_x is expressed by $M_x = M_s A_x / A_s$. This relationship has been exemplified by the determination of Au (0.03–3.4 ppm) in metallic Cu with rhodamine B (CI Basic Violet 10) solution as extracting agent.

6.28 Mercury

IDA by solvent extraction has been employed for the highly selective determination of trace amounts of Hg.[54] Mercury has been determined in amounts of 10^{-6}–10^{-7} g mL^{-1} with an average precision of ±0.5% and in amounts of 10^{-8}–10^{-9} g mL^{-1} with an average precision of ±2.5%. The determination of smaller amounts of mercury was subject to large errors because of the lower specific activity of the radiomercury employed. In the analysis, even significant excesses of metals which also form extractable complexes with dithizone do not interfere. The method is very rapid and simple, because it consists of a single extraction of the analyzed solution and measurement of the activity of the extract. The procedure developed is far more precise and sensitive than the colorimetric method.

A new principle for the substoichiometric determination of traces of Hg by IDA has been reported.[55] The

method was based on the dithizone–carbon tetrachloride extraction of Hg from the test solution (e.g. a solution of an ore or a mineral water). The necessity for determining the chemical yield was avoided. The main advantage of the isotope dilution method is the use of low activities.

Malonic anilide has been employed for the substoichiometric extraction of Hg into a mixture of ether and MIBK and for the subsequent determinations in industrial effluents from the Durgapur area, W.B., India.[56]

A radiometric method based on Subst-IDA has been developed for the determination of trace amounts of Hg using α-thiopicolin-*o*-anisylamide as a substoichiometric reagent.[57] A preconcentration procedure has also been developed based on the adsorption of Hg(II) as a dithizone complex on microcrystalline naphthalene which, when used in conjunction with the proposed radiometric method, enhances its sensitivity and selectivity. As little as 0.2 μg of Hg could be determined in an aqueous phase of 80 mL. The method has been successfully applied to the determination of trace amounts of Hg in city waste incineration ash, zinc metal and cadmium oxide samples.

6.29 Thallium

Thallium has been determined by SSEIDA.[58] The general method has been applied to the determination of Tl. A constant substoichiometric amount of EDTA was used, and the unchelated Tl^{3+} was removed by precipitation with aqueous ammonia before an equal volume of the solution was removed for counting. Common anions did not interfere, and interfering cations were eliminated by extracting Tl(III) from 6 M HCl into diethyl ether.

Thallium has been determined using brilliant green by SSEIDA.[59] The 1:1 complex between $TlCl_4$ and the brilliant green (**1**) cation was extracted into benzene; this avoided the need for large excess of the reagent and yielded a stable colour with a high *E* value. In the subequivalence test, the aqueous medium, containing radioactive plus inactive Tl (carrier), was made 2.5 M in HCl, **1** was added in substoichiometric amount and the extraction was carried out for 9 min. The benzene extract was evaporated and the residue was counted with a Geiger–Müller tube. In the superequivalence variant, twice as much radioactive (but no inactive) Tl and more **1** was used. The amount of active Tl was found from the point of intersection of isotope dilution curves. For 0.5–105 μg of active Tl(III), the errors ranged from ±1.2 to 2.3%. Preliminary extraction of Tl into diethyl ether from 6 M HCl medium and washing the extract with 3 M HCl prevented interference from Fe(III), Ga(III) and In(III). Interference from Au(III) and Sb(V) was avoided by reducing these with metallic Cu before extraction. Sulfate, nitrate and acetate interfered in the formation of the Tl complex.

6.30 Lead

Ultramicro amounts of lead have been determined by IDMS.[60] After isotopic equilibration with approximately 0.5 μg of ^{208}Pb, the sample, e.g. pure water or snow (20–100 g), was mixed with silica gel (13 μg) and H_3PO_4 (127 μg) as stabilizing agents, and part of the mixture was loaded on to a rhenium filament for MS. The Pb concentration of the sample was calculated from the ratio of $^{206}Pb^+$ to $^{208}Pb^+$. The coefficient of variation was approximately 3% for several measurements of added ^{208}Pb (0.003–1.8 μg), 3–7% for river water (approximately 0.07 ppb) and 1.3% for mountain snow (0.86 ppb). The instrument produced 1 fA of ion current from each nanogram of added $^{208}Pb^+$, the detection limit being 0.1 aA. The method was also applicable to the analysis of reagents, e.g. ammonia solution, hydrochloric acid and perchloric acid.

The application of displacement and exchange reactions in IDA of Pb has been studied.[61] The efficiency of separation of metal ions combining a displacement or exchange reaction and a liquid–liquid extraction procedure was calculated as a function of parameters such as pH, the ratio of the concentration of reagent and analyte and the ratio of the concentration of reagent and interfering ions. The resulting equations can be used to calculate favorable conditions for multielement IDA.

6.31 Uranium

A procedure and instrumentation for the automatic measurement of isotopic ratios with a double collector magnetic mass spectrometer have been used for U-isotope analysis.[62] Essentially, measurement was achieved by discriminating in favor of the desired peak plateau for ion current integration during the high-voltage variation. The peak resolution was virtually unaffected by off-set drift in the electronic system. Before each ratio measurement, the off-set values were determined and digitally stored for final compensation when isotopic ratios were calculated with an arithmetic unit.

Traces of uranium have been determined in metallic sodium by IDA and a spectrophotometric method.[63] Sodium (5 g) was evaporated (within 5 h) at a pressure of 10^{-4} Torr at 350 °C in a stainless-steel crucible. There was no loss of U (approximately 1 μg) under these conditions. An isotope dilution method was used to determine 0.001–1 ppm of U with the use of an NBS standard isotope. For >1 ppm of U, the U was reduced with Zn amalgam to U(IV), which was determined spectrophotometrically with arsenazo III after extracting

Fe(III) with diethyl ether. No interference was caused by the small amount of Fe dissolved from the crucible.

6.32 Plutonium

Plutonium has been determined in a NIST SRM by IDA.[64] SRM 4351 human lung was acknowledged to be inhomogeneous. The value of a single analysis for $^{239}Pu + ^{240}Pu$ could lie within a wide range, but the accuracy of the analysis can be corroborated by comparing the corresponding $^{238}Pu/^{239}Pu + ^{240}Pu$ or $^{240}Pu/^{239}Pu$ values with the relationships between these ratios and the total plutonium concentration. The $^{238}Pu/^{239}Pu + ^{240}Pu$ ratio was reported by activity and the $^{240}Pu/^{239}Pu$ ratio by atoms in aliquots of the SRM 4351. The data give an experimental insight into the inhomogeneity of the material, augmenting the purely statistical indicators that have existed so far. An intercomparison exercise on $^{239}Pu + ^{240}Pu$ in SRM 4351 was carried out. The objectives were to evaluate the capabilities of the participating laboratories for Pu measurements in human tissues and to improve the characterization of the $^{239}Pu + ^{240}Pu$ concentration value for the SRM. The results were published, but the values of ^{239}Pu and of ^{240}Pu as distinct from $^{239}Pu + ^{240}Pu$ figures were not reported.

7 APPLICATION OF ISOTOPE DILUTION ANALYSIS TO SOME MULTIELEMENTAL DETERMINATIONS

IDMS has been used for the certification of SRMs.[65] Isotope dilution has been applied at the NBS, employing thermal ionization or spark source MS, to determine accurately the concentration of trace elements in natural materials. These techniques have been applied, in conjunction with separation by ion exchange, chelation by resins, electrodeposition and chemical extraction. The certification of a range of biological and environmental SRMs, including the contents of Pb and trace elements in water, 15 elements in coal and fly ash, U in bovine liver and Hg in water, is presented.

IDMS has been used for the trace analysis of microelectronically relevant heavy metals in high-purity titanium.[66] Because titanium is increasingly used in microelectronic devices, an IDMS method has been developed for the reliable determination of traces of U, Th, Cu, Pb, Cd, Cr, Ni and Fe in high-purity titanium primary materials. The measurement of isotope ratios was carried out with a thermal ionization quadrupole mass spectrometer using positive thermal ions formed by a single- or double-filament ion source, except for thorium, where an ICP mass spectrometer was applied. Different separation techniques (ion exchange, chromatography, extraction, electrolytic deposition, coprecipitation) were used for the trace element–matrix separation and for the specific isolation of the trace elements to be determined. The detection limits obtained were U, Th = 0.07, Cu = 1, Cd = 1.7, Ni = 4, Pb = 6 and Fe = 35 ng g^{-1}. Three titanium samples of different purity were analyzed with concentrations in the following ranges: U, Th = $<0.07 \times 10^{-3}$–0.09, Cd = <0.002–0.7, Cu, Ni, Pb, Cr = 0.01–30 and Fe = $7–6 \times 10^{3}$ ng g^{-1}. The IDMS results for one titanium sputter target were compared with those for two different laboratories which showed the urgent necessity for the application of independent and reliable analytical methods.

Multielemental IDA by means of radiometric titration for the determination of Cu and Pb has been discussed.[67] A theoretical concept was derived for multielement IDA in a liquid–liquid extraction system. The practical performance was based on radiometric titration. For this purpose, a prototype of a titration/extraction vessel with flow injection was constructed. For calculations of the initial concentrations of the elements of interest from the experimental data, a computer program was developed.

Niobium and tantalum have been quantitatively determined by IDA.[68] Preliminary separations of Nb and Ta from each other and from Ti and Zr can be made by precipitation with benzene–selenous acid and extraction with 8-hydroxyquinoline before determination by means of the radioactive isotopes ^{95}Nb and ^{182}Ta. Nb was separated from Zr in 1 M HCl containing 0.5% of citric acid and 1.5% of ammonium selenate and Ta from Zr by increasing the citric acid concentration to 1.5%. Nb was separated from Ta, Zr and Ti by extraction at pH 4.0 to 4.5 with 2% 8-hydroxyquinoline solution in chloroform.

Uranium–plutonium metallic spikes for IDMS accountancy measurements, preparation and characterization have been discussed.[69] The advantages were discussed of the use of a solid $^{235}U–^{239}Pu$ alloy rather than the conventional $^{233}U–^{242}Pu$ solution to provide tracers for the IDMS determination of U and Pu in spent nuclear fuel solutions for reprocessing. The compositions [all percentages (w/w)] U–0.5% Pu–50% Ti, U–0.5% Pu–25% each Nb and Zr and U–1.0% Pu–2.3% Nb were chosen on account of their good cold formability, and were prepared by inductive levitation melting, subsequently being cold-rolled to a thickness of approximately 0.1 mm. Metallography of these alloys showed no significant heterogeneity, and γ-ray spectrometry showed that the relative SD of the Pu distribution in portions composed of 0.1 g of U and 1 mg of Pu (the usual added masses) was 0.4% for the U–Pu–Nb–Zr alloy and 0.1–0.2% for the other two.

Sample preparation approaches have been discussed for the IDICPMS certification of reference materials.[70] The application of IDA to inductively coupled plasma

mass spectrometry (ICPMS) is discussed. The requirements for accurate sample preparation and instrumental analyte–matrix–plasma interaction for the attainment of precise isotope ratio measurements are presented. The method is illustrated by the determination of Mg, Cd, Mo and Pb in synthetic samples, Cu and Cd in Zn ore and Cu and Mo in sewage sludge. The integration of procedures into the certification program at NIST is described.

Simultaneous radiochemical determination of chemically similar elements by the concentration-dependent distribution method has been described.[71] The method with two substoichiometric systems can be characterized by the following features. The selectivity of the two systems must be different, and desirably each of them should react preferentially with a different element. It is sufficient to use only one tracer for the two systems, but the use of two tracers could be advantageous in some cases. The analysis is based on the fact that if phase distributions in the two systems occur under strictly standard conditions, a set of two distribution ratios corresponds to the two values of the unknown concentrations. The calculation of the unknown concentrations from the two distribution ratios can be carried out using (i) theoretically derived equations based on the equilibria involved, (ii) empirical or semiempirical relationships and (iii) calibration plots. The calibration plots can be used in any case; methods (i) and (ii) can be utilized under conditions required for sufficiently strict validity of the corresponding equations. The method was applied for the determination of Rb and Cs.

The partial isotope separation of U and Fe by means of solvent extraction has been discussed.[72] The separation of isotopes is of great practical importance and, hence, it is worthwhile to explore the possibility of isotope separation by means of the simple method of solvent extraction.

Subst-IDA has been applied for the determination of trace elements in liquid samples, of the carrier content in RI solutions and of concentrations of organic reagents.[73] Cu in mineral acids and in a ZnSe single crystal was determined by substoichiometric extraction with dithizone. Values of 1.8 and 0.018 ppm in nitric acid and of 1.4 and 0.44 ppm in ZnSe were obtained. Cu and P carrier contents in ^{64}Cu and ^{32}P solutions were determined by substoichiometric extractions with dithizone in carbon tetrachloride and with molybdic acid into MIBK in a series of solutions, adding various amounts of Cu and P carrier. An analogous method was applied for the analysis of dithizone and diethyldithiocarbamate solutions. The method was also applied to the determination of ^{60}Co radioactivity in environmental samples. The analytical results for water samples are described.

Molybdenum and nickel have been determined in SRMs using the internal normalization technique for high-accuracy IDA.[74] General exact equations and iteration techniques have been developed for internal normalization to eliminate the effect of thermal fractionation on isotope ratio measurements, and therefore IDA, by thermal ionization MS. The techniques were applicable to more than 20 elements, and have been extensively applied to the determination of Mo in ore concentrates (55% Mo) and silicate trace standards (50 and 500 ppm Mo). The SD of all internally corrected Mo isotope ratio measurements was 0.1%. The Mo sample size was 40 μg, but the normalization technique should apply to microgram and smaller samples with a more sensitive ion detection system. Procedures are described for the chemical separation of Mo from matrix interferences and for the MS determination of Mo. Application of the techniques to Ni in three pollution SRMs is described.

A new variation of IDA based on double labeling carried out on filter-paper has been described.[75] The determination of microgram and submicrogram amounts of silver (labeled with ^{110m}Ag) by precipitation with iodide (labeled with ^{131}I) and of Ca (labeled with ^{45}Ca) by precipitation with phosphate (labeled with ^{32}P) are given as examples. In the first example, the two radioelements (^{110m}Ag and ^{131}I) were measured by γ-spectrometry and in the second method (^{45}Ca and ^{32}P) by a β-absorption method. The results show the usefulness of the method.

Coprecipitation has been used as a means of separation in Subst-IDA.[76] It is shown with examples that by applying coprecipitation for separation and by using complexometric and oxidimetric or reductometric reactions as the substoichiometric reaction, a wide range of complexometric determinations of metal ions (e.g. with EDTA) and many types of redox determinations are opened up for further investigation. The method is simple, not time-consuming and offers many possibilities owing to the freedom to select a suitable precipitate.

A semiquantitative method of isotope dilution in the microgram and submicrogram ranges using the ring oven method has been described.[77] Radioactive tracer elements were added to the sample solution and separated by the ring oven technique. The separated ring of precipitate was counted and the original concentration determined from a calibration graph. Examples are cited for the determination of calcium, iron and phosphate.

Ca and Mg have been determined by isotope dilution at the microgram level.[78] Natural calcium contains 96.97% of ^{40}Ca and 2.06% of ^{44}Ca and the tracer isotope used contained 2.9% of ^{40}Ca and 97% of ^{44}Ca. Precautions were necessary to avoid contamination, e.g. from atmospheric dust and from the vessels used. A solution of known concentration (1 mL) mixed with a known amount of tracer isotope was evaporated under nitrogen to one drop, which was applied to the double Ta filament of the spectrometer and evaporated to dryness.

The residue was covered with ammonium oxalate (as ionization activator) by similar evaporation of three drops of a saturated solution, and the ratio of the two isotopes was determined. Amounts of approximately 1 μg of Ca could be determined readily with a SD of 2.5%. The sensitivity under the conditions used was approximately 0.01 μg. Determination of 10^{-4}–10^{-5} μg is considered possible. Interference due to ^{40}K can be eliminated by heating the filament to a temperature below that necessary for the determination of Ca. Natural Mg (78.6% of ^{24}Mg and 10.11% of ^{25}Mg) was determined by mixing $Mg(NO_3)_2$ solution (1 mL) with the tracer isotope (98.5% of ^{25}Mg) and BeO (1 mg) as activator, and applying the concentrated solution to a double rhenium filament. From solutions containing 30.34 and 3.034 μg mL^{-1} of natural Mg, recoveries were 30.44 ± 0.39 and 3.016 ± 0.086 μg mL^{-1}, respectively. Under suitable conditions, the measurement of approximately 10^{-4} μg was considered possible. Use of a double instead of a single filament increased the sensitivity approximately 100-fold.

Application to the determination of elements and chemical compounds by SI dilution has been reported.[79] A review is presented on the use of this technique for the determination of H, rare gases, C, N, O, S, alkali, alkaline earth and heavy metals, the actinium series and compounds containing these elements.

ACKNOWLEDGMENTS

Thanks are due to Dr J.P. Mittal, Director, Chemical Group, BARC and Dr D.D. Sood, Director, Radiochemistry and Isotope Group, BARC for some of the papers on IDA required for this work. Sincere thanks are due to Shri Taher for typing the manuscript in spite of his busy schedule.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
CRM	Certified Reference Material
GC	Gas Chromatography
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
IAEA	International Atomic Energy Agency
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IDA	Isotope Dilution Analysis
IDICPMS	Isotope Dilution Inductively Coupled Plasma Mass Spectrometry
IDMS	Isotope Dilution Mass Spectrometry
LC	Liquid Chromatography
MCA	Multichannel Analyzer
MIBK	Methyl Isobutyl Ketone
MS	Mass Spectrometry
NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
radio-IDA	Radioisotope Dilution Analysis
RI	Radioisotope
RIDA	Reverse Isotope Dilution Analysis
SD	Standard Deviation
SI	Stable Isotope
SRM	Standard Reference Material
SSE	Sub- and Superequivalence
SSEIDA	Sub- and Superequivalence Isotope Dilution Analysis
Subst-IDA	Substoichiometric Isotope Dilution Analysis
TFA	Trifluoroacetylacetone

REFERENCES

1. G. Hevesy, R. Hobbie, 'Lead Content of Rocks', *Nature*, **128**, 1083 (1931).
2. G. Hevesy, R. Hobbie, 'Determination of the Lead Content of Rocks. Application of the Methods of Radioactive Indicators to Analytical Chemistry', *Z. Anal. Chem.*, **88**, 1 (1932).
3–7. M. Haissinsky, *Nuclear Chemistry and Its Applications*, Addison-Wesley, Reading, MA, 1964.
8. J. Ruzicka, J. Stary, *Substoichiometry in Radiochemical Analysis*, Pergamon Press, Oxford, 1958.
9. H. Yoshioka, K. Hasegawa, 'Radioanalytical Application of Sub- and Super-equivalence Method of Isotope Dilution Analysis', *J. Radioanal. Nucl. Chem.*, **160**, 417–423 (1992).
10. E.J. Underwood, *Trace Elements in Human and Animal Nutrition*, Academic Press, New York, 281–322, 1977.
11. A.S. Prasad, *Trace Elements and Iron in Human Metabolism*, John Wiley & Sons, Chichester, 66–75, 1978.
12. S.V. Neveda, U.P. Kulkarni, S.M. Rao, *Trace Analysis and Technological Development*, ed. M. Sankar Das, Wiley Eastern, New Delhi, 257–274, 1983.
13. I.M. Kolthoff, P.J. Elving, E.B. Sandell, *Treatise on Analytical Chemistry*, Part II, Section A, Interscience, New York, 371–372, 1961.
14. P. Allain, Y. Mauras, C. Douge, L. Jaunault, T. Delaforte, C. Beaugrand, 'Determination of Iodine and Bromine in Plasma and Urine by ICP-MS', *Analyst*, **813**, 115 (1990).
15. K. Han, W.F. Koch, K.W. Pratt, 'Improved Procedure for the Determination of Iodide by Ion-chromatography

with Electro-chemical Detection', *Anal. Chem.*, **731**, 59 (1987).
16. D.C. Newton, J. Sanders, A.C. Tyrrell, 'The Determination of Boron in Silicon by Isotope Dilution', *Analyst*, **85**, 870–877 (1960).
17. W.D. Lehmann, H.R. Schulten, 'Quantitative Field Desorption Mass Spectrometry. III. Stable Isotope Dilution and Cationisation', *Angew. Chem.*, **16**(3), 184–185 (1977).
18. O.Yu. Begak, V.V. Fedorov, 'Determination of Small Contents of Carbon in Metals by Isotope Dilution Method with Use of an Omegatron', *Zh. Anal. Khim.*, **32**, 624–626 (1977).
19. K.Y. Eng, R.A. Meyer, C.D. Bingham, 'Determination of Carbon in Sodium by Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **36**, 1832–1834 (1964).
20. K.B. Orlova, E.N. Vitol, 'Determination of Nitrogen in Metals by Isotope Dilution Method', *Zh. Anal. Khim.*, **20**, 694–699 (1965).
21. J.P. Burden, C.R. Masson, 'Modified Isotope Dilution Technique for the Determination of Oxygen in Iron and Steel', *J. Iron Steel Inst.*, **202**(1), 28–31 (1964).
22. F.W. Lima, L.T. Atalla, 'Isotope Dilution Method for the Determination of Phosphorus in Uranium Compounds at the ppm Level', *An. Acad. Bras. Cienc.*, **33**(2), 149–160 (1961).
23. D. Kaniansky, P. Rajec, P. Havasi, A. Svec, F. Macasek, 'Isotope Dilution Analysis in Column Coupling Isotachophoresis', *J. Radioanal. Nucl. Chem. Lett.*, **104**(3), 157–182 (1986).
24. E. Karlton, 'Determination of Adsorbed Sulfate with Isotope Dilution of Sulfur (^{35}S) with Calcium Dihydrogen Phosphate Extraction', *Commun. Soil Sci. Plant Anal.*, **25**(3–4), 207–214 (1994).
25. Yu.M. Artemov, K.G. Knorre, V.P. Strizhov, 'Isotope Dilution with Use of a Standard in the Solid Phase (Applied to Argon)', *Geokhimya*, **12**, 1155–1156 (1963).
26. Yu.A. Shukolyukov, I.I. Matreeva, 'Determination of Small Amounts of Potassium by Isotope Dilution Method', *Zh. Anal. Khim.*, **16**, 544–548 (1961).
27. K.G. Heumann, E. Kubassek, W. Schwabenbauer, 'Mass Spectrometric Isotope Dilution Determination of Low Concentration of Calcium in Minerals', *Fresenius Z. Anal. Chem.*, **287**(2–3), 121–127 (1977).
28. L.J. Moore, L.A. Machlan, 'High Accuracy Determination of Ca in Blood Serum by Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **44**, 2291–2296 (1972).
29. C. Veillon, K.Y. Patterson, 'Determination of Natural and Isotopically Enriched Chromium in Urine by Isotope Dilution Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **66**, 856–860 (1994).
30. J. Stary, J. Ruzicka, 'Isotope Dilution Analysis by Ion Exchange. I. Determination of Traces of Iron', *Talanta*, **8**, 775–781 (1961).
31. W.D. Ralph, T.R. Sweet, I. Mencis, 'Determination of Small Amounts of Cobalt by Isotope Dilution Analysis', *Anal. Chem.*, **34**, 1992–1994 (1962).
32. K.F. Sporek, 'Determination of Small Amounts of Cobalt Using Isotope Dilution with Cobalt-60', *Anal. Chem.*, **33**, 754–758 (1961).
33. C.A. Bizollon, 'Determination of Copper in Serum by Radioisotope Dilution and Incomplete Extraction', *Ann. Pharm. Fr.*, **23**, 461–472 (1965).
34. H. Akaiwa, H. Kawamoto, K. Ogura, 'Substoichiometric Determination of Zinc by the Isotope Dilution Method Using a Chelating Agent-loaded Resin', *Talanta*, **24**, 394–395 (1977).
35. P.C. Reddy, B. Rangamannar, 'Comparative Substoichiometric Extraction of Zinc with Potassium Salts of Ethyl, Propyl, Butyl, Pentyl and Benzyl Xanthates', *J. Radioanal. Nucl. Chem. Lett.*, **200**(3), 181–190 (1995).
36. N.L. Chutke, M.N. Ambulkar, R.G. Weginwar, A.N. Garg, 'Substoichiometric Isotope Dilution Analysis of Arsenic in Biological and Environmental Standard Reference Materials by Solvent Extraction Using Toluene-3,4-dithiol in Benzene', *J. Radioanal. Nucl. Chem. Art.*, **185**(1), 145–156 (1994).
37. P. Cukor, J. Walzcyk, P.K. Lott, 'Application of Isotope Dilution Analysis to the Fluorimetric Determination of Selenium in Plant Material', *Anal. Chim. Acta*, **30**, 473–482 (1964).
38. J. Lesny, J. Tolgyessy, E.H. Klehr, 'Sub and Super Equivalence Method of Isotope Dilution Analysis. IV. Determination of Selenium', *Radiochem. Radioanal. Lett.*, **28**(1), 77–82 (1977).
39. W.J. Kelleher, M.J. Johnson, 'Determination of Traces of Selenium in Organic Matter. Combined Spectrophotometric and Isotope Dilution Method', *Anal. Chem.*, **33**, 1429–1432 (1961).
40. J. Ruzicka, P. Benes, 'Isotope Dilution Electroanalysis. II. Microgram Scale Determination of Silver', *Collect Czech. Chem. Commun.*, **26**, 1784–1791 (1961).
41. J. Pauwel, A. Lamberty, P. De Bievre, K.H. Grobecker, C. Bauspiess, 'Certified Reference Materials for the Determination of Cadmium in Polyethylene', *Fresenius' J. Anal. Chem.*, **349**, 409–411 (1994).
42. S.E. Long, 'Isotope Dilution Inductively Coupled Mass Spectrometry', *UK At. Energy Res. Establ. Memo.*, AERE-M 3540, **16** (1986).
43. J. Ruzicka, J. Stary, 'Isotope Dilution Analysis by Ion Exchange. II. Substoichiometric Determination of Traces of Indium', *Talanta*, **11**, 691–696 (1964).
44. P.C. Reddy, B. Rangamannar, 'Substoichiometric Radiochemical Determination of Antimony with Potassium Ethyl Xanthate', *J. Radioanal. Nucl. Chem. Lett.*, **187**, 107–112 (1994).
45. J. Suzuki, H. Yoshioka, A. Ohya, T. Kambara, 'Radiometric Analysis Based on Redox Substoichiometry. I.

Principles and Application for the Determination of Tellurium', *J. Radioanal. Nucl. Chem., Art.*, **109**(1), 101–107 (1987).
46. V. Singh, A.N. Garg, 'Substoichiometric Isotope Dilution Analysis Method for the Determination of Iodine in Common Salts Using ^{131}I Tracer', *Analyst*, **119**, 1417–1419 (1994).
47. R.G. Weginwar, N.L. Chutke, A.N. Garg, 'Substoichiometric Isotope Dilution Analysis of La in Environmental Samples Using Cupferron in Chloroform', *Sci. Total Environ.*, **130**, 219–230 (1993).
48. F.W. Walker, 'Determination of Hf in Ze-4 and Ze-6 Zirconium Metal Standards by the Stable Isotope Dilution and Neutron Activation Method', *US At. Energy Comm. Rep.*, KAPL-M-FW-4, **26** (1961).
49. G.M. Kolosova, Y.P. Cheng, M.M. Senyarin, 'Chromatographic Separation of Hf from Zr and the Determination of Hf by Isotope Dilution', *Zh. Anal. Khim.*, **15**, 346–366 (1960).
50. G. Leliaert, J. Hoste, Z. Eeckhaut, 'Determination of Tungsten in High Alloy Steels by Isotope Dilution', *Recl. Trav. Chim. Pays-Bas*, **79**, 557–566 (1960).
51. R.L. Walker, 'Rhenium Analysis by Surface Ionisation Isotope Dilution Method', *Int. J. Mass Spectrom. Ion Phys.*, **10**(2), 207–209 (1972).
52. M.N. Belkova, A.M. Chekmarev, V.I. Shamaev, I.D. Troshkina, 'Determination of Rhenium by Radiotracer Methods of Analysis', *J. Radioanal. Nucl. Chem., Art.*, **173**(1), 3–8 (1993).
53. K. Kudo, N. Suzuki, 'Radioactivation Analysis Accompanied by the Isotope Dilution Method. I. Determination of Traces of Gold', *J. Chem. Soc. Jpn. Pure Chem. Sect.*, **86**, 1050–1053 (1965).
54. J. Ruzicka, J. Stary, 'Isotopic Dilution Analysis by Solvent Extraction. III. Highly Selective Determination of Trace Amounts of Hg', *Talanta*, **8**, 535–538 (1961).
55. J. Ruzicka, J. Stary, A. Zeeman, 'New Principle of Activation Analysis Separations. VI. Substoichiometric Determination of Traces of Hg by Activation Analysis and by Isotope Dilution', *Zh. Anal. Khim.*, **19**, 932–936 (1964).
56. J. Dutta, S. Basu, 'Substoichiometric Extraction of Mercury with Malonic Anilide', *J. Radioanal. Nucl. Chem. Lett.*, **200**, 491–495 (1995).
57. D. Sandhya, M.S. Subramanian, 'Determination of Mercury by Substoichiometric Isotope Dilution Analysis', *Radiochim. Acta*, **65**, 105–109 (1994).
58. V.R.S. Rao, P.C. Rao, G. Tataiah, 'Determination of Tl by the Sub and Super Equivalence Method of Isotope Dilution Analysis', *Radiochem. Radioanal. Lett.*, **29**(5), 261–266 (1977).
59. V.S.R. Rao, P.C. Rao, G. Tataiah, 'Determination of Tl Using Brilliant Green [C.I. Basic Green 1] by the Sub and Super Equivalence Method of Isotope Dilution Analysis', *Radiochem. Radioanal. Lett.*, **29**(5), 261–266 (1977).
60. M. Masayo, N. Seiji, F. Kiyoshi, H. Makoto, 'ID-MS Determination of Ultra-micro Amounts of Lead', *Bunseki Kagaku*, **26**(2), 81–85 (1977).
61. P.C.A. Ooms, U.A.Th. Brinkman, H.A. Das, 'Applications of Displacement and Exchange Reactions in Neutron Activation and Isotope Dilution Analysis', *J. Radioanal. Chem.*, **46**, 255–264 (1978).
62. W. Ehrfeld, D. Froehlich, O. Fromheim, U. Knapp, 'Automatic Isotope Ratio Measurements with a Double-collector Magnetic Mass Spectrometer', *Kernforschungszentrum Karlsruhe Ber.*, KFK-2344, **14** (1977).
63. T. Masao, M. Yuji, T. Yukio, O. Misao, K. Takuji, 'Determination of Traces of Uranium in Metallic Sodium', *Bunseki Kagaku*, **20**, 1085–1091 (1971).
64. D.S. Popplewell, W. McCarthy, G.J. Ham, 'The Isotopic Composition of Plutonium in the National Institute of Standards and Technology Standard Reference Material 4351, Human Lung', *Radioanal. Nucl. Chem.*, **172**(2), 299–304 (1993).
65. L.J. Moore, H.M. Kingston, T.J. Murphy, P.J. Paulsen, 'Use of Isotope Dilution Mass Spectrometry for Certification of Standard Reference Materials', *Environ. Int.*, **10**(2), 169–173 (1984).
66. B. Beer, K.G. Heumann, 'Trace Analysis of Microelectronically Relevant Heavy Metals in High Purity Titanium with Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **65**, 3199–3203 (1993).
67. P.C.A. Ooms, G.P. Leendertse, H.A. Das, U.A.Th. Brinkman, 'Multielement Isotope Dilution Analysis by Means of Radiometric Titration', *J. Radioanal. Chem.*, **67**(1), 5–14 (1981).
68. L.P. Alimarin, G.N. Bilimovich, 'Quantitative Determination of Niobium and Tantalum by Isotope Dilution Method', *Collect. Czech. Chem. Commun.*, **26**(1), 255–261 (1961).
69. J.M. Orea Rocha, C.D. Ingelbrecht, A.J. Criado Portal, 'Uranium–Plutonium Metallic Spikes for Isotope Dilution Mass Spectrometry Accountancy Characterization', *Nucl. Instrum. Methods Phys. Res., Sect. A*, **334**, 96–100 (1993).
70. E.S. Beary, P.J. Paulsen, J.D. Fassett, 'Sample Preparation Approaches for Isotope Dilution Inductively Coupled Plasma Mass Spectrometric Certification of Reference Materials', *J. Anal. At. Spectrom.*, **9**, 1363–1369 (1994).
71. M. Kyrs, J. Halova, 'Simultaneous Radiochemical Determination of Chemically Similar Elements by the Method of Concentration Dependent Distribution. I. Principles and Theory of the Method of Two Substoichiometric Systems', *J. Radioanal. Chem.*, **52**(1), 53–67 (1979).
72. R.H. Betts, J. Bron, 'A Discussion of Partial Isotope Separation by Means of Solvent Extraction', *Sep. Sci.*, **12**, 635–639 (1977).

73. K. Kudo, K. Kobayashi, 'Determination of Trace Elements in Radioactive Materials by Substoichiometric Isotope Dilution Analysis', *J. Radioanal. Chem.*, **53**(1–2), 163–172 (1979).
74. L.J. Moore, L.A. Machlan, W.R. Shields, E.L. Garner, 'International Normalization Techniques for High Accuracy Isotope Dilution Analysis – Application to Molybdenum and Nickel in Standard Reference Materials', *Anal. Chem.*, **46**, 1082–1089 (1974).
75. H. Weisz, V. Krivan, 'Isotope Dilution Analysis in the Micro and Submicro Range by Double Labelling and Precipitation on Paper', *Talanta*, **16**, 823–826 (1969).
76. H.L. Polak, 'Coprecipitation as a Means of Separation in Substoichiometric Isotope Dilution Analysis. I. General Considerations and Introduction', *J. Radioanal. Chem.*, **9**, 241–249 (1971).
77. H. Weisz, D. Klockow, 'Semi-quantitative Isotope Dilution Analysis in the Microgram and Sub-microgram Ranges by the Ring Oven Method', *Mikrochim. Ichnoanal. Acta*, **5**, 1082–1093 (1963).
78. R. Delwiche, M. Vosters, 'Determination of Calcium and Magnesium on the Microgram Scale by Isotope Dilution', *Bull. Soc. Chim. Belg.*, **73**(3–4), 255–267 (1964).
79. Z. Dobkowski, 'Stable Isotope Dilution Analysis. II. Application to the Determination of Elements and Chemical Compounds', *Chem. Anal. (Warsaw)*, **10**, 519–530 (1965).

Instrumental Neutron Activation Analysis: Gamma Lines Table

S. Mark and Z.B. Alfassi
Ben Gurion University, Beer Sheva, Israel

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation[a]	Gamma line intensity	Other lines (keV) (intensities, %)
100	Nd-150	Pm-151	28.40 h		2.53	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
100.1	Ta-181	Ta-182	115.0 d	2099.74	14.1	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
100.9	Eu-153	Eu-154m	46.0 m		25.0	31.8 (5.7), 35.8 (14), 40.9 (8.1), 41.5 (35), 47 (4.5), 68.2 (37)
101.8	Pd-110	Pd-111m	5.5 h	0.43	0.27	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
102.1	Te-130	Te-131m	30 h	0.676	7.91	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
102.3	Gd-160	Gd-161	3.70 m	16.832	14.0	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 165.2 (2.60), 283.6 (6), 314.9 (22.9), 360.9 (60.6), 480.1 (2.70)
102.5	W-182	W-183m	5.15 s		2.35	46.5 (5.9), 52.6 (6.7), 58.0 (36.5), 59.3 (63.5), 67.2 (21.5), 69.1 (5.46), 99.1 (8.6), 107.9 (18.2), 160.5 (4.9)
103.1	Se-80	Se-81m	57.3 m	3.968	9.79	275.9 (1.3), 290.1 (1.0), 552.4 (0.15), 566 (0.38), 828.3 (0.47)
103.2	Sm-152	Sm-153	46.7 h	5500.2	28.3	40.9 (17.4), 41.5 (31.3), 47.0 (9.57), 48.3 (2.74), 69.7 (5.25)
103.2	Gd-152	Gd-153	242 d	220	21.8	40.9 (34), 41.5 (61.5), 47.0 (18.7), 48.3 (5.40), 69.7 (2.34), 83.4 (0.21), 97.4 (30.1)
104.3	Sm-154	Sm-155	22.1 m	124.85	75.1	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 141.4 (2.03), 245.7 (3.76)

[a] Probability of formation = cross-section of formation × natural abundance of appropriate isotope.

(*continued overleaf*)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
103.2	Gd-152	Gd-153	242 d	220	21.8	40.9 (34), 41.5 (61.5), 47.0 (18.7), 48.3 (5.40), 69.7 (2.34), 83.4 (0.21), 97.4 (30.1)
104.3	Sm-154	Sm-155	22.1 m	124.85	75.1	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 141.4 (2.03), 245.7 (3.76)
104.5	Yb-176	Yb-177m	6.41 m		76.6	51.4 (21.2), 52.4 (37.5), 59.3 (12.2), 61.0 (3.26), 227.0 (12.2)
104.7	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
104.8	Nd-150	Pm-151	28.40 h		3.52	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
105.3	Sm-154	Eu-155	4.68 yr		21.8	6.3 (8.5), 42.3 (6.97), 43.0 (12.5), 48.7 (3.87), 86.5 (32.7)
105.4	Hf-176	Hf-177m$_1$	1.08 s		16.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
105.4	Lu-176	Hf-177m$_1$	1.08 s		16.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
105.4	Lu-176	Lu-177m	160.9 d	18.2	12.2	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
105.4	Hf-176	Hf-177m$_2$	51.4 m		16.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
105.9	Mo-100	Mo-101	14.6 m	1.916	0.25	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
105.9	Re-187	Re-188m	18.6 m	100.16	10.8	59.7 (18), 61.1 (31), 63.6 (21.3), 69.2 (11), 71.2 (2.7), 92.5 (5.15), 156.0 (0.6), 169.5 (0.1)
107.9	W-182	W-183m	5.15 s		18.2	46.5 (5.9), 52.6 (6.7), 58.0 (36.5), 59.3 (63.5), 67.2 (21.5), 69.1 (5.46), 99.1 (8.6), 102.5 (2.35), 160.5 (4.9)
108.1	Th-232	Th-233	22.3 m	740	0.36	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
108.2	Dy-164	Dy-165m	1.26 m	47 940	3.01	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 153.8 (0.242), 361.7 (0.534), 515.5 (1.53)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
108.5	Ba-130	Ba-131m	14.6 m	0.265	55.2	31.8 (13.7), 32.2 (25.4), 36.4 (7.5), 37.3 (1.79), 79 (1.19)
108.8	Ru-96	Ru-97	2.9 d	1.6	0.11	18.3 (58.5), 20.7 (11.7), 215.7 (85.8), 324.5 (10.2), 460.6 (0.12), 569.3 (0.88)
109.3	Te-124	Te-125m	58.0 d	0.192	0.330	27.2 (33.8), 27.5 (63), 31 (18.2), 31.7 (3.68), 35.5 (6.66)
109.8	Yb-168	Yb-169	32.02 d		18.0	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 118.2 (1.93), 130.5 (11.5), 177.2 (22), 198 (36), 307.7 (11.1)
109.8	Er-168	Er-169	9.30 d	52.26	0.0013	8.4 (0.017), 49.8 (0.0036), 50.7 (0.0064), 57.5 (0.0021), 59.1 (0.0006), 118.2 (0.001)
109.9	O-18	O-19	27.1 s	0.000032	2.54	197.1 (95.9), 1356.8 (50.4), 1444.1 (2.64), 1554 (1.39)
111	Xe-124	Xe-125m	57 s	2.8	61.8	29.5 (17.9), 29.8 (33.2), 33.6 (9.75), 34.4 (2.14), 141 (19.6)
111.5	Th-232	Th-233	22.3 m	740	0.11	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
111.6	Er-170	Er-171	7.52 h	84.93	20.5	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 116.7 (2.30), 124.0 (9.10), 295.9 (28.9), 308.3 (64.4)
111.7	Ir-193	Ir-194m	171 d	376.2	8.85	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
112.5	Ga-71	Ga-72	14.1 h	187.929	0.12	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
113	Hf-176	Hf-177m_2	51.4 m		27.3	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
113	Lu-176	Lu-177m	160.9 d	18.2	21.7	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
113	Hf-176	Hf-177m_1	1.08 s		27.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
113	Lu-176	Lu-177	6.71 d	5460	6.60	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 208.4 (11), 249.7 (0.21), 321.3 (0.22)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
113	Lu-176	Hf-177m$_1$	1.08 s		27.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
113	Yb-176	Lu-177	6.71 d		6.60	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 208.4 (11), 249.7 (0.21), 321.3 (0.22)
113.8	Yb-174	Yb-175	4.19 d	2067	1.93	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 137.7 (0.12), 144.9 (0.33), 282.5 (3.08), 396.3 (6.55)
113.8	Hf-174	Hf-175	70 d	62.4	0.31	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 229.6 (0.77), 318.9 (0.17), 343.4 (86.6), 353.6 (0.23), 432.8 (1.56)
114	Mo-92	Mo-93m	6.85 h	0.089	0.68	17.4 (18.8), 19.7 (3.7), 263.1 (56.7), 684.7 (99.7), 949.8 (0.12), 1363.0 (0.79), 1477.1 (99.1)
114.3	Nd-148	Nd-149	1.73 h	14.4	19.0	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
116.7	Er-170	Er-171	7.52 h	84.93	2.30	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 124.0 (9.10), 295.9 (28.9), 308.3 (64.4)
116.8	Nd-150	Nd-151	12.4 m	6.768	45.9	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
118.2	Yb-168	Yb-169	32.02 d		1.93	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 130.5 (11.5), 177.2 (22), 198 (36), 307.7 (11.1)
118.2	Er-168	Er-169	9.30 d	52.26	0.001	8.4 (0.017), 49.8 (0.0036), 50.7 (0.0064), 57.5 (0.0021), 59.1 (0.0006), 109.8 (0.0013)
121.1	Se-74	Se-75	119.8 d	46.62	17.3	66.1 (1.14), 96.7 (3.48), 136 (59), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
121.3	Nd-146	Pm-147	2.62 yr		0.0028	39.5 (0.0006), 40.1 (0.0011), 45.4 (0.0003), 46.6 (0.0001), 76.2 (0.0001), 197.4 (0.0001)
121.5	Zn-70	Zn-71g	2.4 m	0.0514	3.0	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
121.5	Zn-70	Zn-71m	3.94 h	0.005	2.9	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
121.6	Lu-176	Lu-177m	160.9 d	18.2	5.87	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
121.6	Yb-176	Yb-177	1.9 h		3.4	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 150.4 (20), 1080.1 (5.5), 1241.4 (3.4)
121.8	Eu-151	Eu-152	13.33 yr	2 82 020	28.4	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
121.8	Eu-151	Eu-152m$_1$	9.32 h	1 57 740	7.16	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 344.3 (2.44), 841.6 (14.5), 963.4 (11.9)
122.7	Hf-178	Hf-179m$_2$	25.1 d	867.2	26.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
122.7	Re-185	Re-186	3.78 d	4188.8	0.72	58.0 (1.55), 59.3 (2.69), 61.5 (1.11), 63.0 (1.92), 67.2 (0.91), 69.1 (0.23), 71.3 (0.66), 73.4 (0.17), 122.7 (0.72), 137.2 (9.20), 767.5 (0.029)
123.1	Eu-153	Eu-154	8.8 yr		40.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
123.8	Ba-130	Ba-131	11.8 d	0.9328	29.2	30.6 (27.9), 31 (51.6), 35 (15.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
124	Er-170	Er-171	7.52 h	84.93	9.10	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 116.7 (2.30), 295.9 (28.9), 308.3 (64.4)
124.6	Xe-126	Xe-127m	69.2 s	0.0405	69.1	29.5 (15.4), 29.8 (28.6), 33.6 (8.43), 34.4 (1.85), 172.5 (37.9)
125.4	W-184	W-185	75.1 d	55.206	0.019	59.7 (0.015), 61.1 (0.025), 69.2 (0.008), 71.2 (0.002)
126.1	Cr-54	Cr-55	3.5 m	0.8496	0.0017	1528 (0.037), 2252.5 (0.0031)
127.2	Mo-100	Tc-101	14.2 m		2.86	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627 (0.39), 715.6 (0.69), 720 (0.23), 842.8 (0.23), 928.6 (0.13)
127.5	Cs-133	Cs-134m	2.91 h	250	12.6	11.2 (0.94), 30.6 (8.95), 31.0 (16.5), 35.0 (4.88), 35.8 (1.10), 138.7 (0.0039)
128.5	Lu-176	Hf-177m$_1$	1.08 s		20.2	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
128.5	Hf-176	Hf-177m_2	51.4 m		20.2	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
128.5	Hf-176	Hf-177m_1	1.08 s		20.2	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
128.5	Lu-176	Lu-177m	160.9 d	18.2	15.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
129.4	Os-190	Os-191	15.4 d	102.96	25.7	63.3 (16), 64.9 (27.6), 73.5 (9.51), 75.6 (2.45)
129.4	Pt-190	Ir-191m	4.94 s		25.7	41.8 (0.0056), 63.3 (16), 64.9 (27.6), 73.5 (9.51), 75.6 (2.45), 82.4 (0.02)
129.4	Pt-190	Pt-191	2.90 d	1.95	3.2	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
129.4	Os-190	Ir-191m	4.94 s		25.7	41.8 (0.0056), 63.3 (16), 64.9 (27.6), 73.5 (9.51), 75.6 (2.45), 82.4 (0.02)
129.6	Ru-104	Rh-105m	45 s		20.0	20.2 (34.5), 22.8 (7.14)
129.6	Ru-104	Ru-105	4.44 h	5.984	5.6	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
129.7	Kr-78	Kr-79m	50 s	0.0595	27.5	12.6 (31.8), 14.1 (5.48)
129.7	Pt-194	Pt-195m	4.02 d	2.961	2.83	30.9 (2.28), 65.1 (22.6), 66.8 (38.7), 75.7 (13.3), 77.9 (3.51), 98.9 (11.4)
129.8	Sr-84	Sr-85m	67.66 m	0.336	0.15	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 151.2 (12.8), 231.9 (84.4), 238.8 (0.275)
129.8	Kr-84	Kr-85m	4.48 h	5.13	0.30	12.6 (3.39), 13.4 (1.80), 14.1 (0.58), 15.0 (0.32), 151.2 (75), 304.9 (14)
130.2	Hg-196	Au-197m	7.80 s		3.12	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 201.8 (1.1), 279.0 (70.9), 409.1 (0.11)
130.2	Pt-196	Au-197m	7.80 s		3.12	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 201.8 (1.1), 279.0 (70.9), 409.1 (0.11)
130.2	Pt-196	Pt-197m	1.57 h	1.265	0.10	53.0 (1.08), 65.1 (13.6), 66.8 (23.8), 68.8 (0.40), 75.7 (8), 77.9 (2.26), 279.0 (2.3), 346.5 (11.2)
130.4	Hg-196	Hg-197m	23.8 h	18	0.22	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 133.9 (34.1), 165.0 (0.27), 279.0 (5.0)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
130.5	Yb-168	Yb-169	32.02 d		11.5	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 177.2 (22), 198 (36), 307.7 (11.1)
131.5	W-184	W-185m	1.67 m	0.061	4.34	9 (12), 58.0 (2.2), 59.3 (3.9), 65.9 (5.82), 173.7 (3.30)
133	Hf-180	Hf-181	42.4 d	443.52	43.0	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 136.2 (6.1), 345.9 (14.0), 482.0 (85.5)
133.6	Ba-130	Ba-131	11.8 d	0.9328	2.17	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
133.9	Hg-196	Hg-197m	23.8 h	18	34.1	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 130.4 (0.22), 165.0 (0.27), 279.0 (5.0)
134.2	W-186	W-187	23.9 h	1081.08	10.3	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 479.5 (25.3), 551.6 (5.89), 618.4 (7.27), 685.8 (31.6), 772.9 (4.77)
135.5	Pt-192	Pt-193m	4.33 d	1.738	0.11	12.6 (0.66), 65.1 (4.30), 66.8 (7.38), 75.7 (2.53), 77.9 (0.67)
136	Kr-78	Kr-79g	35.04 h	2.1	1.0	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
136	Se-74	Se-75	119.8 d	46.62	59.0	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
136	Ge-74	Ge-75m	47.7 s	6.205	0.020	9.9 (23.6), 11.0 (3.47), 61.9 (0.12), 139.7 (38.8)
136.2	Hf-180	Hf-181	42.4 d	443.52	6.1	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 133.0 (43.0), 345.9 (14.0), 482.0 (85.5)
137.2	Re-185	Re-186m	2.0×10^5 yr	12.716	9.20	8.6 (16.0), 10.0 (25.0), 11.7 (3.8), 59.0 (18.6), 59.3 (2.96), 63.0 (1.92)
137.2	Re-185	Re-186	3.78 d	4188.8	9.20	58.0 (1.55), 59.3 (2.69), 61.5 (1.11), 63.0 (1.92), 67.2 (0.91), 69.1 (0.23), 71.3 (0.66), 73.4 (0.17), 122.7 (0.72), 137.2 (9.20), 767.5 (0.029)
137.4	Br-81	Br-82g	35.30 h	12.8206	0.14	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
137.7	Yb-174	Yb-175	4.19 d	2067	0.12	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 144.9 (0.33), 282.5 (3.08), 396.3 (6.55)
138.3	In-115	In-116m_1	54.15 m	7753.32	3.29	416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
138.7	Cs-133	Cs-134m	2.91 h	250	0.0039	11.2 (0.94), 30.6 (8.95), 31.0 (16.5), 35.0 (4.88), 35.8 (1.10), 127.5 (12.6)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
138.9	Nd-150	Nd-151	12.4 m	6.768	8.72	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
138.9	Os-192	Os-193	30.5 h	82	4.27	63.3 (3.74), 64.9 (6.46), 73.0 (3.2), 73.5 (2.22), 460.5 (3.95)
139.7	Ge-74	Ge-75m	47.7 s	6.205	38.8	9.9 (23.6), 11.0 (3.47), 61.9 (0.12), 136.0 (0.020)
140.5	Mo-98	Tc-99m	6.01 h		87.7	18.3 (6.72), 20.7 (1.35), 142.6 (0.041)
140.5	Mo-98	Mo-99	2.75 d	3.1369	90.7	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
141	Xe-124	Xe-125m	57 s	2.8	19.6	29.5 (17.9), 29.8 (33.2), 33.6 (9.75), 34.4 (2.14), 111 (61.8)
141.4	Sm-154	Sm-155	22.1 m	124.85	2.03	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 104.3 (75.1), 245.7 (3.76)
142.5	Fe-58	Fe-59	44.503 d	0.3968	1.03	192.2 (3.11), 334.8 (0.26), 1099.3 (56.5), 1291.6 (43.2)
142.5	Sc-45	Sc-46m	18.7 s	980	56.0	4.1 (6.6), 4.5 (0.88), 889.2 (100), 1120.5 (100)
142.6	Zn-70	Zn-71m	3.94 h	0.005	5.6	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
142.6	Mo-98	Tc-99m	6.01 h		0.041	18.3 (6.72), 20.7 (1.35), 140.5 (87.7)
144.9	Yb-174	Yb-175	4.19 d	2067	0.33	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 137.7 (0.12), 282.5 (3.08), 396.3 (6.55)
145.3	Xe-126	Xe-127	36.41	0.27	4.29	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 172.1 (25.5), 202.9 (68.3), 375 (17.2)
145.4	Ce-140	Ce-141	32.50 d	50.433	48.4	35.5 (4.88), 36 (8.91), 40.7 (2.66), 41.8 (0.7)
146.1	Hf-178	Hf-179m$_2$	25.1 d	867.2	26.3	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
146.8	Ta-181	Ta-182m	15.8 m	1.029	35.6	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 171.6 (46.8), 185.0 (23.4), 318.4 (6.55)
147.2	Lu-176	Lu-177m	160.9 d	18.2	3.45	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
149.1	Ru-104	Ru-105	4.44 h	5.984	1.74	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
149.7	Te-130	Te-131m	30 h	0.676	20.5	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
149.7	Te-130	Te-131	25.0 m	9.126	68.9	28.3 (3.7), 28.6 (6.9), 32.3 (2), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)
150.4	Yb-176	Yb-177	1.9 h		20.0	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 121.6 (3.4), 1080.1 (5.5), 1241.4 (3.4)
150.8	Cd-110	Cd-111m	48.6 m	1.75	30.3	23.1 (34.1), 26.2 (7.25), 245.4 (94)
151.2	Kr-84	Kr-85 m	4.48 h	5.13	75.0	12.6 (3.39), 13.4 (1.80), 14.1 (0.58), 15.0 (0.32), 129.8 (0.30), 304.9 (14)
151.2	Sr-84	Sr-85g	64.84 d	0.196	0.001	13.4 (50.2), 15.0 (8.80), 514 (99.3), 868.1 (0.12)
151.2	Kr-84	Kr-85g	10.72 yr	2.394	0.0001	362.8 (0.0001), 514.0 (0.434)
151.2	Sr-84	Sr-85m	67.66 m	0.336	12.8	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 129.8 (0.15), 231.9 (84.4), 238.8 (0.275)
152.4	Ta-181	Ta-182	115.0 d	2099.74	7.17	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
153.3	Hf-176	Hf-177m_1	1.08 s		21.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
153.3	Lu-176	Hf-177m_1	1.08 s		21.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
153.3	Hf-176	Hf-177m_2	51.4 m		21.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
153.3	Lu-176	Lu-177m	160.9 d	18.2	16.6	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
153.8	Dy-164	Dy-165m	1.26 m	47 940	0.242	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 108.2 (3.01), 361.7 (0.534), 515.5 (1.53)
155	Re-187	Re-188	16.98 h	4569.8	14.9	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
155.9	Nd-148	Nd-149	1.73 h	14.4	5.93	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
156	Re-187	Re-188m	18.6 m	100.16	0.6	59.7 (18), 61.1 (31), 63.6 (21.3), 69.2 (11), 71.2 (2.7), 92.5 (5.15), 105.9 (10.8), 169.5 (0.1)
156	Sn-116	Sn-117m	13.61 d	0.0882	2.11	25 (18.7), 25.3 (35.1), 28.5 (9.92), 29.1 (1.91), 158.6 (86.4)
156.4	Ge-76	Ge-77	11.3 h	0.468	0.76	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
156.4	Ta-181	Ta-182	115.0 d	2099.74	2.73	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
158.2	Xe-134	Xe-135	9.08 h	2.704	0.29	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 249.8 (90.4), 358.4 (0.22), 408 (0.36), 608.2 (2.91)
158.4	Pt-198	Au-199	3.14 d		36.9	49.8 (0.33), 68.9 (4.47), 70.8 (7.59), 80.2 (2.62), 82.5 (0.72), 208.2 (8.38)
158.4	Hg-198	Hg-199m	42.6 m	0.1818	52.5	68.9 (19.0), 70.8 (32.3), 80.2 (11.1), 82.5 (3.08), 374.1 (13.8), 413.4 (0.027)
158.6	Cd-116	Cd-117m	3.36 h	0.1867	109.0	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
158.6	Sn-116	Sn-117m	13.61 d	0.0882	86.4	25 (18.7), 25.3 (35.1), 28.5 (9.92), 29.1 (1.91), 156 (2.11)
159	Te-122	Te-123m	119.7 d	2.86	84.0	27.2 (14.1), 27.5 (26.2), 31 (7.55), 31.7 (1.53), 88.5 (0.087), 247.5 (0.0003)
159.1	Ge-76	Ge-77	11.3 h	0.468	0.22	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
159.4	Ca-46	Sc-47	3.34 d	0.00296	68.0	–
159.7	Ge-76	Ge-77m	52.9 s	0.78	11.3	9.9 (3.75), 10.5 (0.13), 11.0 (0.55), 194.8 (0.48), 215.5 (20.9)
160.3	Sn-122	Sn-123	129.2 d	0.0046	0.002	1021.0 (0.002), 1030.2 (0.031), 1088.6 (0.6)
160.3	Sn-122	Sn-123m	40.08 m	0.828	85.6	26.1 (3.06), 26.4 (5.72), 29.7 (1.63), 30.4 (0.32), 381.4 (0.042), 541.8 (0.02)
160.5	W-182	W-183m	5.15 s		4.9	46.5 (5.9), 52.6 (6.7), 58.0 (36.5), 59.3 (63.5), 67.2 (21.5), 69.1 (5.46), 99.1 (8.6), 102.5 (2.35), 107.9 (18.2)
160.6	Xe-132	Xe-133	5.29 d	10.76	0.064	30.6 (13.3), 31 (24.5), 35 (7.25), 35.8 (1.64), 79.6 (0.6), 81 (35.9)
160.7	Hf-178	Hf-179m$_1$	18.7 s	1436.3	2.79	54.6 (16), 55.8 (28.1), 63.2 (9.39), 65.0 (2.4), 214.1 (95.2), 374.8 (0.005)
162	Se-76	Se-77m	17.5 s	198	52.4	11.2 (20.0), 12.5 (3.14)
162	Ge-76	As-77	38.83 h		0.16	88 (0.27), 239.1 (1.59), 250 (0.41), 520.8 (0.42)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
162.4	In-115	In-116m_2	2.18 s	7753.32	36.6	24.1 (28.3), 27.4 (6.13)
162.5	Th-232	Th-233	22.3 m	740	0.32	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
163.5	Ru-104	Ru-105	4.44 h	5.984	0.15	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
163.9	Xe-130	Xe-131m	11.9 d	1.845	1.96	29.5 (15.1), 29.8 (28.1), 33.6 (8.26), 34.4 (1.81)
163.9	Te-130	Xe-131m	11.9 d		1.96	29.5 (15.1), 29.8 (28.1), 33.6 (8.26), 34.4 (1.81)
165	Hg-196	Hg-197m	23.8 h	18	0.27	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 130.4 (0.22), 133.9 (34.1), 279.0 (5.0)
165.2	Gd-160	Gd-161	3.70 m	16.832	2.60	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 283.6 (6), 314.9 (22.9), 360.9 (60.6), 480.1 (2.70)
165.8	Ce-138	Ce-139	137.7 d	0.275	79.9	33 (22.6), 33.4 (41.5), 37.8 (12.3), 38.7 (3.02)
165.8	Ba-138	Ba-139	1.38 h	25.812	23.8	33 (1.38), 33.4 (2.54), 37.8 (0.76), 38.7 (0.19), 1254.7 (0.033), 1420.5 (0.261)
167.7	Nd-150	Pm-151	28.40 h		7.84	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
169.1	Th-232	Th-233	22.3 m	740	0.34	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
169.1	Pd-110	Pd-111m	5.5 h	0.43	0.42	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
169.5	Re-187	Re-188m	18.6 m	100.16	0.1	59.7 (18), 61.1 (31), 63.6 (21.3), 69.2 (11), 71.2 (2.7), 92.5 (5.15), 105.9 (10.8), 156.0 (0.6)
169.8	Hf-178	Hf-179m_2	25.1 d	867.2	18.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
170.7	Th-232	Th-233	22.3 m	740	0.13	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
170.7	Mg-26	Mg-27	9.46 m	0.42058	0.84	843.7 (71.8), 1014.4 (28.2)
170.7	Nd-150	Nd-151	12.4 m	6.768	3.81	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
171.3	Pd-110	Ag-111m	64.8 s		0.11	22.1 (15.6), 25.0 (3.30), 59.8 (0.531), 245.4 (0.46), 620.1 (0.11)
171.6	Ta-181	Ta-182m	15.8 m	1.029	46.8	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 146.8 (35.6), 185.0 (23.4), 318.4 (6.55)
171.9	Lu-176	Lu-177m	160.9 d	18.2	4.75	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
172.1	Xe-126	Xe-127	36.41	0.27	25.5	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 145.3 (4.29), 202.9 (68.3), 375 (17.2)
172.2	Pt-190	Pt-191	2.90 d	1.95	3.52	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
172.2	Pd-110	Pd-111m	5.5 h	0.43	33.5	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
172.5	Xe-126	Xe-127m	69.2 s	0.0405	37.9	29.5 (15.4), 29.8 (28.6), 33.6 (8.43), 34.4 (1.85), 124.6 (69.1)
173.3	Pd-110	Pd-111	23.4 m	2.22	0.11	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
173.7	W-184	W-185m	1.67 m	0.061	3.30	9 (12), 58.0 (2.2), 59.3 (3.9), 65.9 (5.82), 131.5 (4.34)
174.4	Lu-176	Hf-177m$_1$	1.08 s		16.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
174.4	Hf-176	Hf-177m$_2$	51.4 m		16.4	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
174.4	Hf-176	Hf-177m$_1$	1.08 s		16.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
175.1	Nd-150	Nd-151	12.4 m	6.768	7.19	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
176.2	Ga-69	Ga-70	21.15 m	100.968	0.297	1039.2 (0.673)
176.3	Sn-124	Sb-125	2.73 yr		6.79	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
177	Hf-176	Hf-177m$_1$	1.08 s		4.47	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
177	Lu-176	Lu-177m	160.9 d	18.2	3.40	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
177	Hf-176	Hf-177m$_2$	51.4 m		4.47	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
177	Lu-176	Hf-177m$_1$	1.08 s		4.47	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
177.2	Nd-150	Pm-151	28.40 h		3.58	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
177.2	Yb-168	Yb-169	32.02 d		22.0	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 130.5 (11.5), 198 (36), 307.7 (11.1)
177.3	Ge-76	Ge-77	11.3 h	0.468	0.17	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
179.4	Ta-181	Ta-182	115.0 d	2099.74	3.15	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
179.7	Mo-100	Tc-101	14.2 m		0.64	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
180.2	Kr-78	Kr-79g	35.04 h	2.1	0.10	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
180.3	Au-197	Au-198m	2.30 d		50.8	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 204.1 (41.5), 214.9 (76.9), 333.8 (15.0)
181.1	Mo-98	Mo-99	2.75 d	3.1369	6.08	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
182.2	Dy-156	Dy-157	8.10 h	1.98	2.2	43.7 (24.1), 44.5 (43.2), 50.3 (13.8), 51.7 (3.8), 60.8 (0.4), 83.0 (0.6), 265.3 (0.3), 326.2 (93.2)
183.6	Ru-104	Ru-105	4.44 h	5.984	0.10	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
184.1	Mo-100	Tc-101	14.2 m		1.69	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
184.4	Ho-165	Ho-166m	1.2×10^3 yr	350	73.9	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
185	Ta-181	Ta-182m	15.8 m	1.029	23.4	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 146.8 (35.6), 171.6 (46.8), 318.4 (6.55)
185.8	Pt-198	Pt-199	30.8 m	26.64	3.26	68.8 (2.0), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
186.7	Os-189	Os-190m	9.90 m	0.0041	69.9	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 361.1 (95.2), 502.6 (97.8), 616.1 (98.5)
187.4	Mo-100	Mo-101	14.6 m	1.916	0.47	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
188.4	Xe-124	Xe-125	16.9 h	13.7	54.9	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 243.4 (28.8), 453.8 (4.23), 846.5 (1.03)
188.9	Pd-108	Pd-109m	4.69 m	4.76	55.7	21.1 (21.7), 23.9 (4.56)
188.9	Se-82	Se-83m	70.4 s	0.3666	0.17	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
190.3	Kr-80	Kr-81m	13.0 s	10.2375	67.0	12.6 (14.9), 14.1 (2.56)
190.5	Th-232	Th-233	22.3 m	740	0.13	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
191.4	Pt-196	Pt-197	18.3 h	17.71	3.68	67.0 (0.99), 68.8 (1.69), 77.3 (17.2), 77.9 (0.58), 80.2 (0.16), 268.8 (0.23)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
191.4	Hg-196	Hg-197	2.67 d	462	0.49	67.0 (21.2), 68.8 (36.2), 77.3 (18.1), 77.9 (12.5), 80.2 (3.37), 268.7 (0.038)
191.6	In-113	In-114m	49.51 d	34.668	16.0	23.1 (3.91), 24.1 (28.1), 27.4 (6.08), 558 (3.4), 725.2 (3.4)
191.7	Pt-198	Pt-199	30.8 m	26.64	2.38	68.8 (2.0), 185.8 (3.26), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
191.9	Mo-100	Mo-101	14.6 m	1.916	18.8	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
192.2	Fe-58	Fe-59	44.503 d	0.3968	3.11	142.5 (1.03), 334.8 (0.26), 1099.3 (56.5), 1291.6 (43.2)
192.8	Hf-178	Hf-179m$_2$	25.1 d	867.2	20.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
194.8	Ge-76	Ge-77m	52.9 s	0.78	0.48	9.9 (3.75), 10.5 (0.13), 11.0 (0.55), 159.7 (11.3), 215.5 (20.9)
194.8	Ge-76	Ge-77	11.3 h	0.468	1.68	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
194.9	Th-232	Th-233	22.3 m	740	0.16	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
195.9	Mo-100	Mo-101	14.6 m	1.916	2.86	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
196.6	Xe-128	Xe-129m	8.87 d	0.91	4.59	29.5 (36.1), 29.8 (66.9), 33.6 (19.7), 34.4 (4.32), 39.6 (7.52)
197	Tb-159	Tb-160	72.3 d	2550	5.24	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
197.1	O-18	O-19	27.1 s	0.000032	95.9	109.9 (2.54), 1356.8 (50.4), 1444.1 (2.64), 1554 (1.39)
197.4	Nd-146	Pm-147	2.62 yr		0.0001	39.5 (0.0006), 40.1 (0.0011), 45.4 (0.0003), 46.6 (0.0001), 76.2 (0.0001), 121.3 (0.0028)
198	Yb-168	Yb-169	32.02 d		36.0	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 130.5 (11.5), 177.2 (22), 307.7 (11.1)
198.6	Ge-74	Ge-75	83 m	12.41	1.17	66.0 (0.11), 264.6 (11.3), 419.1 (0.18), 468.8 (0.22), 617.7 (0.11)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
198.6	Se-74	Se-75	119.8 d	46.62	1.47	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
200.6	Te-130	Te-131m	30 h	0.676	7.53	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
201.8	Lu-175	Lu-176	3.7×10^{10} yr	681.8	84.7	54.6 (9.45), 55.8 (16.5), 63.2 (5.53), 65.0 (1.42), 88.3 (13.1), 306.9 (93.3), 401.1 (0.84)
201.8	Pt-196	Au-197m	7.80 s		1.1	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 279.0 (70.9), 409.1 (0.11)
201.8	Hg-196	Au-197m	7.80 s		1.1	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 279.0 (70.9), 409.1 (0.11)
202.5	Y-89	Y-90m	3.19 h	0.1	96.5	479.5 (90.6), 682.0 (0.32)
202.9	Xe-126	Xe-127	36.41	0.27	68.3	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 145.3 (4.29), 172.1 (25.5), 375 (17.2)
202.9	Te-126	Te-127m	109 d	2.55	0.057	27.2 (10.4), 27.5 (19.3), 28.3 (0.4), 28.6 (0.75), 31 (5.56), 31.7 (1.13), 32.3 (0.22), 57.6 (0.5), 88.3 (0.084), 360.3 (0.132), 417.9 (0.969)
202.9	Te-126	Te-127	9.35 h	17.055	0.058	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 215.1 (0.039), 360.3 (0.135), 417.9 (0.993)
203.7	Hg-204	Hg-205	5.20 m	2.924	2.2	415.6 (0.13), 1218.7 (0.006)
204.1	Zr-94	Nb-95m	3.61 d		2.36	16.6 (36), 18.6 (6.97), 235.7 (25.1), 582.1 (0.055)
204.1	Zr-94	Nb-95	34.98 d		0.015	561.9 (0.015), 765.8 (99.9)
204.1	Au-197	Au-198m	2.30 d		41.5	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 180.3 (50.8), 214.9 (76.9), 333.8 (15.0)
204.2	Lu-176	Hf-177m$_1$	1.08 s		17.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
204.2	Hf-176	Hf-177m$_1$	1.08 s		17.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
204.2	Lu-176	Lu-177m	160.9 d	18.2	13.6	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
204.2	Hf-176	Hf-177m$_2$	51.4 m		17.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
205.8	Ir-191	Ir-192	73.83 d	23 275.2	3.20	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
205.8	Ir-191	$Ir-192m_2$	241 yr	3.73	3.20	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
207.8	Er-166	Er-167m	2.28 s	504	41.7	48.2 (5.40), 49.1 (9.60), 55.6 (3.07), 57.2 (0.85)
208.1	Nd-148	Nd-149	1.73 h	14.4	2.54	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
208.2	Pt-198	Au-199	3.14 d		8.38	49.8 (0.33), 68.9 (4.47), 70.8 (7.59), 80.2 (2.62), 82.5 (0.72), 158.4 (36.9)
208.3	Se-82	Se-83g	22.5 m	0.04888	1.85	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
208.4	Lu-176	Lu-177m	160.9 d	18.2	59.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
208.4	Lu-176	$Hf-177m_1$	1.08 s		73.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
208.4	Hf-176	$Hf-177m_2$	51.4 m		73.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
208.4	Lu-176	Lu-177	6.71 d	5460	11.0	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 249.7 (0.21), 321.3 (0.22)
208.4	Hf-176	$Hf-177m_1$	1.08 s		73.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
208.4	Yb-176	Lu-177	6.71 d		11.0	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 249.7 (0.21), 321.3 (0.22)
208.5	Kr-78	Kr-79g	35.04 h	2.1	0.78	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
209	Ge-76	Ge-77	11.3 h	0.468	0.89	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
209	Te-128	Te-129	69.6 m	6.34	0.17	27.8 (16.64), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
211	Ge-76	Ge-77	11.3 h	0.468	29.2	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
211.3	Nd-148	Nd-149	1.73 h	14.4	25.9	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
212	Mo-100	Mo-101	14.6 m	1.916	0.51	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
212.2	Te-120	Te-121m	154.0 d	0.032	81.4	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
213.4	Hf-177	Hf-178m_2	31 yr	0.00000186	80.9	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
213.4	Hf-177	Hf-178m_1	4.0 s	20.46	80.9	54.6 (20.1), 55.8 (35.2), 63.2 (11.8), 65.0 (3.01), 88.9 (62), 93.2 (17.3), 325.6 (93.9), 426.4 (96.9)
214	Hf-176	Hf-177m_2	51.4 m		40.9	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
214.1	Hf-178	Hf-179m_1	18.7 s	1436.3	95.2	54.6 (16), 55.8 (28.1), 63.2 (9.39), 65.0 (2.4), 160.7 (2.79), 374.8 (0.005)
214.5	Hf-176	Hf-177m_2	51.4 m		8.56	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
214.5	Hf-176	Hf-177m_1	1.08 s		8.56	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
214.5	Lu-176	Lu-177m	160.9 d	18.2	6.52	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
214.5	Lu-176	Hf-177m$_1$	1.08 s		8.56	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
214.9	Pd-106	Pd-107m	21.3 s	0.355	68.7	21.1 (16.0), 23.9 (3.35)
214.9	Au-197	Au-198m	2.30 d		76.9	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 180.3 (50.8), 204.1 (41.5), 333.8 (15.0)
215.1	Te-126	Te-127	9.35 h	17.055	0.039	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 202.9 (0.058), 360.3 (0.135), 417.9 (0.993)
215.2	Hf-179	Hf-180m	5.5 h	4.67	81.4	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 332.3 (94.4), 443.2 (82.8), 500.7 (14.5)
215.5	Ge-76	Ge-77	11.3 h	0.468	27.1	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
215.5	Ge-76	Ge-77m	52.9 s	0.78	20.9	9.9 (3.75), 10.5 (0.13), 11.0 (0.55), 159.7 (11.3), 194.8 (0.48)
215.6	Tb-159	Tb-160	72.3 d	2550	4.02	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
215.7	Ru-96	Ru-97	2.9 d	1.6	85.8	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 324.5 (10.2), 460.6 (0.12), 569.3 (0.88)
215.9	Ho-165	Ho-166m	1.2×10^3 yr	350	2.63	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
216.1	Ba-130	Ba-131	11.8 d	0.9328	19.9	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
216.4	Tl-205	Tl-206m	3.75 m		83.6	247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
216.7	Hf-177	Hf-178m$_2$	31 yr	0.00000186	63.7	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
217	Hf-178	Hf-179m$_2$	25.1 d	867.2	8.78	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
217	Kr-78	Kr-79g	35.04 h	2.1	2.40	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
218.1	Lu-176	Lu-177m	160.9 d	18.2	3.28	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
218.9	Zr-96	Zr-97	17.0 h	0.0632	0.17	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
219.1	Ge-76	Ge-77	11.3 h	0.468	0.1	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
221.5	Br-81	Br-82g	35.30 h	12.8206	2.25	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
221.8	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
222.1	Ta-181	Ta-182	115.0 d	2099.74	7.56	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
225.1	Ru-104	Ru-105	4.44 h	5.984	0.12	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
225.2	Se-82	Se-83g	22.5 m	0.04888	31.9	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
226	Gd-158	Gd-159	18.56 h	62.1	0.21	43.7 (5.54), 44.5 (9.93), 50.3 (3.06), 51.7 (0.89), 58.0 (2.27), 348.2 (0.22), 363.6 (10.8)
227	Yb-176	Yb-177m	6.41 m		12.2	51.4 (21.2), 52.4 (37.5), 59.3 (12.2), 61.0 (3.26), 104.5 (76.6)
228.5	Hf-176	Hf-177m$_2$	51.4 m		48.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
228.5	Lu-176	Hf-177m$_1$	1.08 s		48.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
228.5	Hf-176	Hf-177m$_1$	1.08 s		48.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
228.5	Lu-176	Lu-177m	160.9 d	18.2	36.6	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
229.3	Ta-181	Ta-182	115.0 d	2099.74	3.64	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
229.6	Hf-174	Hf-175	70 d	62.4	0.77	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 318.9 (0.17), 343.4 (86.6), 353.6 (0.23), 432.8 (1.56)
231.5	Se-82	Se-83m	70.4 s	0.3666	0.33	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
231.6	Ce-142	Ce-143	33.0 h	10.526	2.04	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 293.3 (42.8), 350.6 (3.27), 490.4 (2.13), 664.5 (5.6), 722 (5.32)
231.9	Sr-84	Sr-85m	67.66 m	0.336	84.4	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 129.8 (0.15), 151.2 (12.8), 238.8 (0.275)
233.2	Xe-132	Xe-133m	2.19 d	1.345	9.95	29.5 (16), 29.8 (29.7), 33.6 (8.73), 34.4 (1.92)
233.7	Mo-100	Tc-101	14.2 m		0.27	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
233.9	Hf-176	Hf-177m$_2$	51.4 m		7.24	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
233.9	Lu-176	Hf-177m$_1$	1.08 s		7.24	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
233.9	Lu-176	Lu-177m	160.9 d	18.2	5.51	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
233.9	Hf-176	Hf-177m$_1$	1.08 s		7.24	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
235.7	Zr-94	Zr-95	64.03 d	0.86227	0.24	16.6 (0.42), 18.6 (0.66), 724.2 (43.7), 756.7 (55.4)
235.7	Zr-94	Nb-95m	3.61 d		25.1	16.6 (36), 18.6 (6.97), 204.1 (2.36), 582.1 (0.055)
236.6	Hf-178	Hf-179m$_2$	25.1 d	867.2	18.3	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
237.4	Hf-177	Hf-178m$_2$	31 yr	0.00000186	8.79	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
238.3	Mo-100	Tc-101	14.2 m		0.31	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
238.8	Sr-84	Sr-85m	67.66 m	0.336	0.275	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 129.8 (0.15), 151.2 (12.8), 231.9 (84.4)
239.1	Ge-76	As-77	38.83 h		1.59	88 (0.27), 162 (0.16), 250 (0.41), 520.8 (0.42)
239.6	Ba-130	Ba-131	11.8 d	0.9328	2.42	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
240.1	Nd-150	Pm-151	28.40 h		3.61	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
240.2	Nd-148	Nd-149	1.73 h	14.4	3.94	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
240.9	Te-130	Te-131m	30 h	0.676	7.57	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
243.4	Xe-124	Xe-125	16.9 h	13.7	28.8	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 188.4 (54.9), 453.8 (4.23), 846.5 (1.03)
244.7	Eu-151	Eu-152	13.33 yr	2 82 020	7.51	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
245.4	Pd-110	Pd-111	23.4 m	2.22	0.48	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
245.4	Pd-110	Ag-111m	64.8 s		0.46	22.1 (15.6), 25.0 (3.30), 59.8 (0.531), 171.3 (0.11), 620.1 (0.11)
245.4	Pd-110	Pd-111m	5.5 h	0.43	0.45	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
245.4	Pd-110	Ag-111	7.45 d		1.24	23.1 (0.18), 96.7 (0.2), 342.1 (6.68)
245.4	Cd-110	Cd-111m	48.6 m	1.75	94.0	23.1 (34.1), 26.2 (7.25), 150.8 (30.3)
245.7	Sm-154	Sm-155	22.1 m	124.85	3.76	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 104.3 (75.1), 141.4 (2.03)
246.5	Pt-198	Pt-199	30.8 m	26.64	2.16	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
247.2	Tl-205	Tl-206m	3.75 m		12.0	216.4 (83.6), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
247.5	Te-122	Te-123m	119.7 d	2.86	0.0003	27.2 (14.1), 27.5 (26.2), 31 (7.55), 31.7 (1.53), 88.5 (0.087), 159 (84)
247.9	Eu-153	Eu-154	8.8 yr		6.60	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
249.4	Ba-130	Ba-131	11.8 d	0.9328	2.84	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
249.7	Lu-176	Lu-177m	160.9 d	18.2	6.08	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
249.7	Lu-176	Hf-177m$_1$	1.08 s		7.97	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
249.7	Hf-176	Hf-177m$_1$	1.08 s		7.97	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
249.7	Hf-176	Hf-177m$_2$	51.4 m		7.97	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
249.7	Yb-176	Lu-177	6.71 d		0.21	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 321.3 (0.22)
249.7	Lu-176	Lu-177	6.71 d	5460	0.21	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 321.3 (0.22)
249.8	Xe-134	Xe-135	9.08 h	2.704	90.4	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 358.4 (0.22), 408 (0.36), 608.2 (2.91)
250	Ge-76	As-77	38.83 h		0.41	88 (0.27), 162 (0.16), 239.1 (1.59), 520.8 (0.42)
250.6	Te-128	Te-129	69.6 m	6.34	0.35	27.8 (16.64), 209 (0.17), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
254.2	Zr-96	Zr-97	17.0 h	0.0632	1.10	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
254.3	Ce-136	Ce-137m	34.4 h	0.1805	10.9	33 (29.5), 33.4 (54.2), 34.3 (15.6), 34.7 (28.7), 37.8 (16.1), 38.7 (3.94), 39.2 (8.53), 40.2 (2.2), 447.2 (3.03)
254.7	Ge-76	Ge-77	11.3 h	0.468	0.20	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
255	Sn -112	Sn-113	115.1 d	0.71	1.85	24.1 (79.8), 27.4 (17.3), 391.7 (64.2), 638 (0.001)
255.7	Nd-150	Nd-151	12.4 m	6.768	15.3	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
257.5	Hf-178	Hf-179m_2	25.1 d	867.2	3.2	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
257.6	Hf-177	Hf-178m_2	31 yr	0.00000186	16.6	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
260.9	Cd-114	Cd-115	2.23 d	86.16	1.94	24.1 (33.1), 27.4 (7.15), 336.2 (49.7), 492.4 (8.03), 527.9 (27.5)
261.3	Kr-78	Kr-79g	35.04 h	2.1	12.7	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
262.8	Ru-104	Ru-105	4.44 h	5.984	6.49	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
263.1	Mo-92	Mo-93m	6.85 h	0.089	56.7	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 684.7 (99.7), 949.8 (0.12), 1363.0 (0.79), 1477.1 (99.1)
263.7	Cd-112	Cd-113m	14.1 yr	0.96	0.023	23.1 (0.035), 26.2 (0.0074)
264.1	Ta-181	Ta-182	115.0 d	2099.74	3.64	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
264.4	Ge-76	Ge-77	11.3 h	0.468	51.0	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
264.6	Se-74	Se-75	119.8 d	46.62	59.1	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
264.6	Ge-74	Ge-75	83 m	12.41	11.3	66.0 (0.11), 198.6 (1.17), 419.1 (0.18), 468.8 (0.22), 617.7 (0.11)
265.3	Dy-156	Dy-157	8.10 h	1.98	0.3	43.7 (24.1), 44.5 (43.2), 50.3 (13.8), 51.7 (3.8), 60.8 (0.4), 83.0 (0.6), 182.2 (2.2), 326.2 (93.2)
265.7	Bi-209	Bi-210m	3.0×10^6 yr	1.4	51.0	70.8 (3.9), 72.9 (6.6), 82.5 (2.3), 304.8 (28.0), 649.8 (2.9)
265.7	Bi-209	Bi-210	5.01 d	1.9	0.0001	11.1 (0.023), 13.4 (0.023), 15.8 (0.0043), 76.9 (0.003), 79.3 (0.006), 89.6 (0.002), 304.8 (0.0001)
265.7	Tl-205	Tl-206m	3.75 m		86.2	216.4 (83.6), 247.2 (12), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
267.7	Nd-148	Nd-149	1.73 h	14.4	6.03	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
268.1	Ge-76	Ge-77	11.3 h	0.468	0.3	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
268.2	Ba-134	Ba-135m	28.7 h	0.3872	15.6	31.8 (15.4), 32.2 (28.4), 36.4 (8.4), 37.3 (2.01)
268.7	Hg-196	Hg-197	2.67 d	462	0.038	67.0 (21.2), 68.8 (36.2), 77.3 (18.1), 77.9 (12.5), 80.2 (3.37), 191.4 (0.49)
268.8	Pt-196	Pt-197	18.3 h	17.71	0.23	67.0 (0.99), 68.8 (1.69), 77.3 (17.2), 77.9 (0.58), 80.2 (0.16), 191.4 (3.68)
268.8	Lu-176	Lu-177m	160.9 d	18.2	3.43	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
268.9	Hf-178	Hf-179m_2	25.1 d	867.2	11.0	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
270.2	Nd-148	Nd-149	1.73 h	14.4	10.7	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
272	Pd-110	Pd-111m	5.5 h	0.43	0.16	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
272.3	Zr-96	Zr-97	17.0 h	0.0632	0.25	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
273.3	Cd-116	Cd-117	2.49 h	0.3735	27.9	24.1 (4.58), 89.7 (3.26), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
273.5	Br-81	Br-82g	35.30 h	12.8206	0.79	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
275.2	Nd-150	Pm-151	28.40 h		6.61	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
275.9	Se-80	Se-81	18.5 m	26.288	0.87	290.1 (0.68), 552.4 (0.10), 566.0 (0.26), 828.3 (0.32)
275.9	Se-80	Se-81m	57.3 m	3.968	1.3	103.1 (9.79), 290.1 (1.0), 552.4 (0.15), 566 (0.38), 828.3 (0.47)
276	Kr-80	Kr-81g	2.1×10^5 yr	15.75	3.6	11.9 (46.2), 13.3 (7.3)
276.1	Ba-132	Ba-133m	38.9 h	0.0505	17.5	12.3 (1.5), 31.8 (15.3), 32.2 (28.2), 36.4 (8.34), 37.3 (1.99), 632.5 (0.01)
276.4	Ba-132	Ba-133	10.54 yr	0.6565	7.29	30.6 (34.4), 31 (63.5), 35 (18.8), 35.8 (4.24), 53.2 (2.2), 79.6 (2.43), 81 (32.8), 302.9 (18.6), 356 (62.3)
277.3	Hf-176	Hf-177m_2	51.4 m		75.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
278.4	Te-128	Te-129	69.6 m	6.34	0.53	27.8 (16.64), 209 (0.17), 250.6 (0.35), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
279	Hg-196	Hg-197m	23.8 h	18	5.0	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 130.4 (0.22), 133.9 (34.1), 165.0 (0.27)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
279	Hg-196	Au-197m	7.80 s		70.9	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 409.1 (0.11)
279	Pt-196	Au-197m	7.80 s		70.9	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 409.1 (0.11)
279	Pt-196	Pt-197m	1.57 h	1.265	2.3	53.0 (1.08), 65.1 (13.6), 66.8 (23.8), 68.8 (0.40), 75.7 (8), 77.9 (2.26), 130.2 (0.1), 346.5 (11.2)
279.2	Hg-202	Hg-203	46.58 d	145.53	81.5	70.8 (3.75), 72.9 (6.35), 82.5 (2.2), 84.9 (0.63)
279.5	Se-74	Se-75	119.8 d	46.62	25.2	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 264.6 (59.1), 303.9 (1.34), 400.6 (11.6)
279.8	Dy-164	Dy-165	2.33 h	28 200	0.50	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
280.1	Ru-104	Rh-105	35.36 h		0.17	21.1 (0.31), 306.1 (5.13), 318.9 (19.2)
280.5	Ho-165	Ho-166m	1.2×10^3 yr	350	30.1	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
281.3	Te-128	Te-129	69.6 m	6.34	0.15	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
281.8	Lu-176	Lu-177m	160.9 d	18.2	14.0	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
281.8	Lu-176	$Hf\text{-}177m_1$	1.08 s		18.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
281.8	Hf-176	$Hf\text{-}177m_2$	51.4 m		18.4	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
281.8	Hf-176	$Hf\text{-}177m_1$	1.08 s		18.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
282.5	Yb-174	Yb-175	4.19 d	2067	3.08	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 137.7 (0.12), 144.9 (0.33), 396.3 (6.55)
283.6	Gd-160	Gd-161	3.70 m	16.832	6.00	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						165.2 (2.60), 314.9 (22.9), 360.9 (60.6), 480.1 (2.70)
284.3	Te-130	I-131	8.02 d		6.06	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 364.5 (81.2), 637 (7.27), 722.9 (1.8)
285.9	Nd-148	Pm-149	2.21 d		3.10	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 590.9 (0.68), 830.5 (0.03), 833.2 (0.03)
289.3	Ga-71	Ga-72	14.1 h	187.929	0.19	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
289.8	Pd-110	Pd-111	23.4 m	2.22	0.10	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
289.8	Pd-110	Pd-111m	5.5 h	0.43	0.99	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
290.1	Se-80	Se-81	18.5 m	26.288	0.68	275.9 (0.87), 552.4 (0.10), 566.0 (0.26), 828.3 (0.32)
290.1	Se-80	Se-81m	57.3 m	3.968	1.0	103.1 (9.79), 275.9 (1.3), 552.4 (0.15), 566 (0.38), 828.3 (0.47)
293.3	Ce-142	Ce-143	33.0 h	10.526	42.8	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 350.6 (3.27), 490.4 (2.13), 664.5 (5.6), 722 (5.32)
293.5	Ir-193	Ir-194	19.15 h	6897	2.54	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
295	Ru-102	Ru-103	39.25 d	38.24	0.25	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 443.8 (0.32), 497.1 (89.5), 557.0 (0.832), 610.3 (5.64)
295.1	Hf-176	Hf-177m_2	51.4 m		69.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
295.9	Er-170	Er-171	7.52 h	84.93	28.9	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 116.7 (2.30), 124.0 (9.10), 308.3 (64.4)
295.9	Ir-191	Ir-192m_1	1.44 m	11 190	0.0018	8 (0.35), 9.2 (8.0), 10.8 (11.6), 12.5 (2.1), 58 (0.039), 316.5 (0.0097)
295.9	Ir-191	Ir-192m_2	241 yr	3.73	28.7	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
295.9	Ir-191	Ir-192	73.83 d	23 275.2	28.7	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
296.1	Se-82	Se-83g	22.5 m	0.04888	0.27	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
296.2	Hf-176	Hf-177m_2	51.4 m		6.51	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
296.2	Lu-176	Lu-177m	160.9 d	18.2	4.96	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
296.5	Lu-176	Hf-177m_1	1.08 s		6.51	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
296.5	Hf-176	Hf-177m_1	1.08 s		6.51	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
296.8	Hf-177	Hf-178m_2	31 yr	0.00000186	9.83	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
297.9	Er-162	Ho-163m	1.09 s		77.5	46.7 (3.19), 47.5 (5.69), 53.8 (1.79), 55.3 (0.50)
298	Xe-136	Xe-137	3.83 m	2.314	0.12	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
298.6	Tb-159	Tb-160	72.3 d	2550	27.4	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
299	Lu-176	Hf-177m_1	1.08 s		2.36	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
299	Hf-176	Hf-177m_2	51.4 m		23.6	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
299	Hf-176	Hf-177m_1	1.08 s		2.36	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
299.5	Kr-78	Kr-79g	35.04 h	2.1	1.57	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
300.6	Nd-150	Nd-151	12.4 m	6.768	1.93	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
300.7	Ho-165	Ho-166m	1.2×10^3 yr	350	3.79	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
300.7	Ir-193	Ir-194	19.15 h	6897	0.35	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
302.9	Ba-132	Ba-133	10.54 yr	0.6565	18.6	30.6 (34.4), 31 (63.5), 35 (18.8), 35.8 (4.24), 53.2 (2.2), 79.6 (2.43), 81 (32.8), 276.4 (7.29), 356 (62.3)
303.9	Se-74	Se-75	119.8 d	46.62	1.34	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 400.6 (11.6)
304.8	Bi-209	Bi-210	5.01 d	1.9	0.0001	11.1 (0.023), 13.4 (0.023), 15.8 (0.0043), 76.9 (0.003), 79.3 (0.006), 89.6 (0.002), 265.7 (0.0001)
304.8	Bi-209	Bi-210m	3.0×10^6 yr	1.4	28.0	70.8 (3.9), 72.9 (6.6), 82.5 (2.3), 265.7 (51.0), 649.8 (2.9)
304.9	Kr-84	Kr-85 m	4.48 h	5.13	14.0	12.6 (3.39), 13.4 (1.80), 14.1 (0.58), 15.0 (0.32), 129.8 (0.30), 151.2 (75)
305.5	Hf-176	Hf-177m_1	1.08 s		2.38	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
305.5	Lu-176	Hf-177m_1	1.08 s		2.38	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
305.5	Hf-176	Hf-177m_2	51.4 m		22.38	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
306.1	Ru-104	Rh-105	35.36 h		5.13	21.1 (0.31), 280.1 (0.17), 318.9 (19.2)
306.3	Kr-78	Kr-79g	35.04 h	2.1	2.60	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
306.8	Mo-100	Tc-101	14.2 m		88.0	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
306.9	Lu-175	Lu-176	3.7×10^{10} yr	681.8	93.3	54.6 (9.45), 55.8 (16.5), 63.2 (5.53), 65.0 (1.42), 88.3 (13.1), 201.8 (84.7), 401.1 (0.84)
307.7	Yb-168	Yb-169	32.02 d		11.1	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 130.5 (11.5), 177.2 (22), 198 (36)
308.1	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
308.3	Er-170	Er-171	7.52 h	84.93	64.4	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 116.7 (2.30), 124.0 (9.10), 295.9 (28.9)
308.4	Ir-191	Ir-192m_2	241 yr	3.73	29.7	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
308.4	Ir-191	Ir-192	73.83 d	23 275.2	29.7	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
311.3	Mo-100	Tc-101	14.2 m		0.24	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
311.4	Pd-108	Pd-109	13.46 h	219.6	0.032	22.1 (28.5), 25.0 (6.02), 88.0 (3.61), 647.3 (0.024)
311.5	Hf-176	Hf-177m_2	51.4 m		58.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
312.7	K-41	K-42	12.36 h	9.8258	0.35	899 (0.053), 1524.6 (18.6)
314.9	Gd-160	Gd-161	3.70 m	16.832	22.9	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 165.2 (2.60), 283.6 (6), 360.9 (60.6), 480.1 (2.70)
316	Hf-178	Hf-179m_2	25.1 d	867.2	19.7	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
316.4	Ru-104	Ru-105	4.44 h	5.984	10.9	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
316.5	Ir-191	Ir-192m_1	1.44 m	11 190	0.0097	8 (0.35), 9.2 (8.0), 10.8 (11.6), 12.5 (2.1), 58 (0.039), 295.9 (0.0018)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
316.5	Ir-191	Ir-192	73.83 d	23 275.2	82.9	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
316.5	Ir-191	Ir-192m$_2$	241 yr	3.73	82.9	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
316.8	Tl-205	Tl-206m	3.75 m		9.5	216.4 (83.6), 247.2 (12), 265.7 (86.2), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
317	Pt-198	Pt-199	30.8 m	26.64	4.87	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
318	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
318.4	Ta-181	Ta-182m	15.8 m	1.029	6.55	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 146.8 (35.6), 171.6 (46.8), 185.0 (23.4)
318.6	Zn-68	Zn-69	56 m	18.8	0.0012	872 (0.0002)
318.6	Zn-68	Zn-69m	13.76 h	1.353	0.0013	8.6 (1.90), 9.6 (0.26), 438.6 (94.8), 574.1 (0.033)
318.9	Ru-104	Rh-105	35.36 h		19.2	21.1 (0.31), 280.1 (0.17), 306.1 (5.13)
318.9	Hf-174	Hf-175	70 d	62.4	0.17	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 343.4 (86.6), 353.6 (0.23), 432.8 (1.56)
319	Lu-176	Lu-177m	160.9 d	18.2	10.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
319.4	Nd-146	Nd-147	10.98 d	24.066	1.95	38.2 (13), 38.7 (23.5), 43.8 (7.05), 44.9 (1.96), 91.1 (27.9), 531 (13)
320.1	Cr-50	Cr-51	27.7 d	69.165	9.83	–
320.1	Ti-50	Ti-51	5.8 m	0.9308	93.0	608.5 (1.18), 928.6 (6.88)
321.3	Yb-176	Lu-177	6.71 d		0.22	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 249.7 (0.21)
321.3	Lu-176	Lu-177	6.71 d	5460	0.22	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 249.7 (0.21)
324	Ir-193	Ir-194m	171 d	376.2	2.0	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
324.5	Ru-96	Ru-97	2.9 d	1.6	10.2	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 215.7 (85.8), 460.6 (0.12), 569.3 (0.88)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
325.6	Hf-177	Hf-178m$_1$	4.0 s	20.46	93.9	54.6 (20.1), 55.8 (35.2), 63.2 (11.8), 65.0 (3.01), 88.9 (62), 93.2 (17.3), 213.4 (80.9), 426.4 (96.9)
325.6	Hf-177	Hf-178m$_2$	31 yr	0.00000186	93.9	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
326.1	Ru-104	Ru-105	4.44 h	5.984	1.05	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
326.2	Dy-156	Dy-157	8.10 h	1.98	93.2	43.7 (24.1), 44.5 (43.2), 50.3 (13.8), 51.7 (3.8), 60.8 (0.4), 83.0 (0.6), 182.2 (2.2), 265.3 (0.3)
326.5	Nd-148	Nd-149	1.73 h	14.4	4.56	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
326.7	Hf-176	Hf-177m$_2$	51.4 m		65.4	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
327.7	Hf-176	Hf-177m$_1$	1.08 s		23.7	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
327.7	Lu-176	Hf-177m$_1$	1.08 s		23.7	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
327.7	Lu-176	Lu-177m	160.9 d	18.2	18.1	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
327.7	Hf-176	Hf-177m$_2$	51.4 m		23.7	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
327.7	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
328.4	Ir-193	Ir-194	19.15 h	6897	13.0	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
328.4	Ir-193	Ir-194m	171 d	376.2	92.8	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
328.8	La-139	La-140	40.28 h	892.12	20.7	432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
330.4	Zr-96	Zr-97	17.0 h	0.0632	0.11	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
330.9	Ru-104	Ru-105	4.44 h	5.984	0.66	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
332.1	Sn-124	Sn-125m	9.52 m	0.728	97.5	26.1 (0.53), 26.4 (1), 29.7 (0.28), 589.6 (0.2), 643 (0.16), 1404 (0.7), 1483.9 (0.18), 1615.3 (0.12)
332.1	Sn-124	Sn-125	9.64 d	0.0224	1.31	469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
332.3	Hf-179	Hf-180m	5.5 h	4.67	94.4	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 215.2 (81.4), 443.2 (82.8), 500.7 (14.5)
333.7	Mo-100	Mo-101	14.6 m	1.916	0.78	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
333.8	Au-197	Au-198m	2.30 d		15.0	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 180.3 (50.8), 204.1 (41.5), 214.9 (76.9)
334.3	Te-130	Te-131m	30 h	0.676	9.53	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
334.8	Fe-58	Fe-59	44.503 d	0.3968	0.26	142.5 (1.03), 192.2 (3.11), 1099.3 (56.5), 1291.6 (43.2)
336.2	Cd-114	Cd-115	2.23 d	86.16	49.7	24.1 (33.1), 27.4 (7.15), 260.9 (1.94), 492.4 (8.03), 527.9 (27.5)
336.2	Cd-114	In-115m	4.49 h		45.8	24.1 (27.7), 27.4 (6), 497.4 (0.047)
336.6	Ga-71	Ga-72	14.1 h	187.929	0.11	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
337.6	Ge-76	Ge-77	11.3 h	0.468	0.22	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
338.7	Ge-76	Ge-77	11.3 h	0.468	0.63	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
338.8	Ir-193	Ir-194m	171 d	376.2	55.1	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
340.1	Nd-150	Pm-151	28.40 h		22.4	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
340.2	Se-82	Se-83g	22.5 m	0.04888	0.41	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
341.6	Hf-176	Hf-177m_1	1.08 s		2.19	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
341.6	Lu-176	Hf-177m_1	1.08 s		2.19	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
341.6	Hf-176	Hf-177m_2	51.4 m		2.19	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
342.1	Pd-110	Ag-111	7.45 d		6.68	23.1 (0.18), 96.7 (0.2), 245.4 (1.24)
343.4	Hf-174	Hf-175	70 d	62.4	86.6	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 318.9 (0.17), 353.6 (0.23), 432.8 (1.56)
344.3	Eu-151	Eu-152	13.33 yr	2 82 020	26.6	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
344.3	Eu-151	Eu-152m_1	9.32 h	1 57 740	2.44	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 121.8 (7.16), 841.6 (14.5), 963.4 (11.9)
344.5	Cd-116	Cd-117	2.49 h	0.3735	17.9	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
344.7	Kr-78	Kr-79g	35.04 h	2.1	0.24	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
344.9	Nd-150	Pm-151	28.40 h		2.10	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
344.9	Zn-64	Zn-65	244.3 d	36.936	0.003	770.6 (0.003), 1115.5 (50.7)
345.9	Hf-180	Hf-181	42.4 d	443.52	14.0	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 133.0 (43.0), 136.2 (6.1), 482.0 (85.5)
346.5	Pt-196	Pt-197m	1.57 h	1.265	11.2	53.0 (1.08), 65.1 (13.6), 66.8 (23.8), 68.8 (0.40), 75.7 (8), 77.9 (2.26), 130.2 (0.1), 279.0 (2.3)
346.9	Co-59	Co-60	5.272 yr	1680	0.0076	1173.2 (99.9), 1332.5 (100)
347.6	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
348.2	Dy-158	Dy-159	144.4 d	4.3	0.001	43.7 (26.9), 44.5 (48.2), 50.3 (14.9), 51.7 (4.25), 58.0 (2.22), 79.5 (0.005)
348.2	Gd-158	Gd-159	18.56 h	62.1	0.22	43.7 (5.54), 44.5 (9.93), 50.3 (3.06), 51.7 (0.89), 58.0 (2.27), 226.0 (0.21), 363.6 (10.8)
350	Ru-104	Ru-105	4.44 h	5.984	0.28	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
350.2	Ru-104	Ru-105	4.44 h	5.984	1.0	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
350.6	Ce-142	Ce-143	33.0 h	10.526	3.27	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 490.4 (2.13), 664.5 (5.6), 722 (5.32)
351.2	Pt-190	Pt-191	2.90 d	1.95	3.36	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
353	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
353.6	Hf-174	Hf-175	70 d	62.4	0.23	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 318.9 (0.17), 343.4 (86.6), 432.8 (1.56)
355.4	Zr-96	Zr-97	17.0 h	0.0632	2.27	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
356	Ba-132	Ba-133	10.54 yr	0.6565	62.3	30.6 (34.4), 31 (63.5), 35 (18.8), 35.8 (4.24), 53.2 (2.2), 79.6 (2.43), 81 (32.8), 276.4 (7.29), 302.9 (18.6)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
356.7	Se-82	Se-83m	70.4 s	0.3666	17.3	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
356.7	Se-82	Se-83g	22.5 m	0.04888	68.6	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
357.5	Pd-102	Pd-103	16.96 d	3.468	0.0221	20.1 (63.8), 22.8 (13.2), 39.7 (0.0683), 357.5 (0.0221), 497.1 (0.004)
357.9	Pd-110	Pd-111m	5.5 h	0.43	0.40	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
358.1	Rh-103	Rh-104	42.3 s	13 500	0.016	19.2 (0.24), 21.7 (0.050), 555.8 (1.99), 1237 (0.066)
358.4	Xe-134	Xe-135	9.08 h	2.704	0.22	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 249.8 (90.4), 408 (0.36), 608.2 (2.91)
359.9	Pt-190	Pt-191	2.90 d	1.95	6.00	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
359.9	Th-232	Th-233	22.3 m	740	0.12	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
360.3	Te-126	Te-127m	109 d	2.55	0.132	27.2 (10.4), 27.5 (19.3), 28.3 (0.4), 28.6 (0.75), 31 (5.56), 31.7 (1.13), 32.3 (0.22), 57.6 (0.5), 88.3 (0.084), 202.9 (0.057), 417.9 (0.969)
360.3	Te-126	Te-127	9.35 h	17.055	0.135	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 202.9 (0.058), 215.1 (0.039), 417.9 (0.993)
360.9	Gd-160	Gd-161	3.70 m	16.832	60.6	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 165.2 (2.60), 283.6 (6), 314.9 (22.9), 480.1 (2.70)
361.1	Os-189	Os-190m	9.90 m	0.0041	95.2	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 186.7 (69.9), 502.6 (97.8), 616.1 (98.5)
361.7	Dy-164	Dy-165	2.33 h	28 200	0.84	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
361.7	Dy-164	Dy-165m	1.26 m	47 940	0.534	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 108.2 (3.01), 153.8 (0.242), 515.5 (1.53)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
362.6	Hf-178	Hf-179m_2	25.1 d	867.2	38.5	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
362.8	Kr-84	Kr-85g	10.72 yr	2.394	0.0001	151.2 (0.0001), 514.0 (0.434)
363.6	Gd-158	Gd-159	18.56 h	62.1	10.8	43.7 (5.54), 44.5 (9.93), 50.3 (3.06), 51.7 (0.89), 58.0 (2.27), 226.0 (0.21), 348.2 (0.22)
364.5	Te-130	I-131	8.02 d		81.2	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 284.3 (6.06), 637 (7.27), 722.9 (1.8)
365.4	Ag-109	Ag-110m	249.9 d	226.4	0.11	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
365.7	Ho-165	Ho-166m	1.2×10^3 yr	350	2.55	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
366.3	Ni-64	Ni-65	2.52 h	1.3832	4.61	507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
366.5	Mo-98	Mo-99	2.75 d	3.1369	1.16	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
366.9	Cd-116	Cd-117m	3.36 h	0.1867	3.33	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
367.4	Lu-176	Lu-177m	160.9 d	18.2	3.17	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
367.4	Ge-76	Ge-77	11.3 h	0.468	13.3	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
367.9	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
370.1	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
371.6	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
371.6	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
373.2	Ba-130	Ba-131	11.8 d	0.9328	14.1	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 486.5 (2.09), 496.3 (47.1)
374.1	Hg-198	Hg-199m	42.6 m	0.1818	13.8	68.9 (19.0), 70.8 (32.3), 80.2 (11.1), 82.5 (3.08), 158.4 (52.5), 413.4 (0.027)
374.8	Hf-178	Hf-179m$_1$	18.7 s	1436.3	0.005	54.6 (16), 55.8 (28.1), 63.2 (9.39), 65.0 (2.4), 160.7 (2.79), 214.1 (95.2)
375	Xe-126	Xe-127	36.41	0.27	17.2	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 145.3 (4.29), 172.1 (25.5), 202.9 (68.3)
376.7	Pd-110	Pd-111m	5.5 h	0.43	0.97	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
376.7	Pd-110	Pd-111	23.4 m	2.22	0.44	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
377.9	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
378.5	Lu-176	Lu-177m	160.9 d	18.2	29.9	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
378.5	Lu-176	Hf-177m$_1$	1.08 s		39.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
378.5	Hf-176	Hf-177m$_1$	1.08 s		39.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
378.5	Hf-176	Hf-177m$_2$	51.4 m		39.3	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
379.3	Mo-100	Mo-101	14.6 m	1.916	0.32	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
381.2	Mo-100	Mo-101	14.6 m	1.916	0.30	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
381.2	Ga-71	Ga-72	14.1 h	187.929	0.27	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
381.4	Sn-122	Sn-123m	40.08 m	0.828	0.042	26.1 (3.06), 26.4 (5.72), 29.7 (1.63), 30.4 (0.32), 160.3 (85.6), 541.8 (0.02)
385	Lu-176	Lu-177m	160.9 d	18.2	3.16	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
385	Hf-176	Hf-177m_2	51.4 m		4.15	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
385	Hf-176	Hf-177m_1	1.08 s		4.15	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
385	Lu-176	Hf-177m_1	1.08 s		4.15	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
386.3	Zn-70	Zn-71m	3.94 h	0.005	93.0	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
388.4	Sr-86	Sr-87m	2.80 h	8.2824	81.8	13.4 (0.15), 14.1 (8.75), 15.9 (1.58)
389	Kr-78	Kr-79g	35.04 h	2.1	1.52	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
389.2	Se-82	Se-83g	22.5 m	0.04888	0.62	225.2 (31.9), 356.7 (68.6), 485.7 (2.26), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
389.9	Zn-70	Zn-71m	3.94 h	0.005	2.6	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
390	Zn-70	Zn-71g	2.4 m	0.0514	3.8	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
390.8	Ir-193	Ir-194m	171 d	376.2	35.1	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
391.3	Pd-110	Pd-111m	5.5 h	0.43	5.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
391.4	Se-82	Se-83m	70.4 s	0.3666	0.10	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
391.7	Sn -112	Sn-113	115.1 d	0.71	64.2	24.1 (79.8), 27.4 (17.3), 255 (1.85), 638 (0.001)
391.7	Sn -112	In-113m	1.66 h		64.2	24.1 (20.1), 27.4 (4.35)
391.9	Pt-198	Pt-199m	13.6 s	0.194	84.7	32 (2.78), 65.1 (2.01), 66.8 (3.45), 75.7 (1.18), 77.9 (0.31)
393.3	Mo-100	Tc-101	14.2 m		0.10	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
393.4	Ru-104	Ru-105	4.44 h	5.984	3.73	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
393.4	Xe-136	Xe-137	3.83 m	2.314	0.14	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
396.3	Yb-174	Yb-175	4.19 d	2067	6.55	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 137.7 (0.12), 144.9 (0.33), 282.5 (3.08)
397.6	Kr-78	Kr-79g	35.04 h	2.1	9.50	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
398.6	Zn-70	Zn-71g	2.4 m	0.0514	0.61	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
398.9	Mo-100	Mo-101	14.6 m	1.916	0.90	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
400.4	Zr-96	Zr-97	17.0 h	0.0632	0.32	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
400.6	Se-74	Se-75	119.8 d	46.62	11.6	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
401.1	Lu-175	Lu-176	3.7×10^{10} yr	681.8	0.84	54.6 (9.45), 55.8 (16.5), 63.2 (5.53), 65.0 (1.42), 88.3 (13.1), 201.8 (84.7), 306.9 (93.3)
402.3	Nd-150	Nd-151	12.4 m	6.768	1.97	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
402.6	Kr-86	Kr-87	76.3 m	0.0519	49.6	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
408	Xe-134	Xe-135	9.08 h	2.704	0.36	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 249.8 (90.4), 358.4 (0.22), 608.2 (2.91)
408.7	Mo-100	Mo-101	14.6 m	1.916	1.60	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
409.1	Hg-196	Au-197m	7.80 s		0.11	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 279.0 (70.9)
409.1	Pt-196	Au-197m	7.80 s		0.11	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 279.0 (70.9)
409.4	Pt-190	Pt-191	2.90 d	1.95	8.00	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 456.5 (3.36), 538.9 (13.7)
409.8	Hf-178	Hf-179m_2	25.1 d	867.2	20.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
410.9	Ho-165	Ho-166m	1.2×10^3 yr	350	11.7	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
411.1	Eu-151	Eu-152	13.33 yr	2 82 020	2.23	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
411.8	Au-197	Au-198	2.70 d		95.5	68.9 (0.81), 70.8 (1.38), 80.2 (0.47), 82.5 (0.13), 675.9 (1.06), 1087.7 (0.23)
413.4	Hg-198	Hg-199m	42.6 m	0.1818	0.027	68.9 (19.0), 70.8 (32.3), 80.2 (11.1), 82.5 (3.08), 158.4 (52.5), 374.1 (13.8)
413.5	Pd-110	Pd-111m	5.5 h	0.43	1.7	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
413.5	Ru-104	Ru-105	4.44 h	5.984	2.22	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
413.7	Lu-176	Lu-177m	160.9 d	18.2	17.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
415.5	Pd-110	Pd-111m	5.5 h	0.43	1.5	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
415.6	Hg-204	Hg-205	5.20 m	2.924	0.13	203.7 (2.2), 1218.7 (0.006)
416.3	Ge-76	Ge-77	11.3 h	0.468	20.6	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
416.9	In-115	In-116m$_1$	54.15 m	7753.32	29.2	138.3 (3.29), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
417.9	Te-126	Te-127m	109 d	2.55	0.969	27.2 (10.4), 27.5 (19.3), 28.3 (0.4), 28.6 (0.75), 31 (5.56), 31.7 (1.13), 32.3 (0.22), 57.6 (0.5), 88.3 (0.084), 202.9 (0.057), 360.3 (0.132)
417.9	Te-126	Te-127	9.35 h	17.055	0.993	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 202.9 (0.058), 215.1 (0.039), 360.3 (0.135)
418.5	Hf-176	Hf-177m$_1$	1.08 s		28.1	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
418.5	Hf-176	Hf-177m$_2$	51.4 m		28.1	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
418.5	Lu-176	Lu-177m	160.9 d	18.2	21.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
418.5	Lu-176	Hf-177m$_1$	1.08 s		28.1	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
419.1	Ge-74	Ge-75	83 m	12.41	0.18	66.0 (0.11), 198.6 (1.17), 264.6 (11.3), 468.8 (0.22), 617.7 (0.11)
419.8	Ge-76	Ge-77	11.3 h	0.468	1.16	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
421.7	Mo-100	Mo-101	14.6 m	1.916	0.56	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
423.5	Nd-148	Nd-149	1.73 h	14.4	7.48	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
423.5	Nd-150	Nd-151	12.4 m	6.768	6.43	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
426.4	Hf-177	Hf-178m$_2$	31 yr	0.00000186	96.9	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
426.4	Hf-177	Hf-178m$_1$	4.0 s	20.46	96.9	54.6 (20.1), 55.8 (35.2), 63.2 (11.8), 65.0 (3.01), 88.9 (62), 93.2 (17.3), 213.4 (80.9), 325.6 (93.9)
427.9	Sn-124	Sb-125	2.73 yr		29.4	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
428.4	Ga-71	Ga-72	14.1 h	187.929	0.20	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
432.5	La-139	La-140	40.28 h	892.12	2.99	328.8 (20.7), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
432.6	Mo-100	Mo-101	14.6 m	1.916	0.12	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
432.8	Hf-174	Hf-175	70 d	62.4	1.56	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 318.9 (0.17), 343.4 (86.6), 353.6 (0.23)
433.2	Ce-136	Ce-137	9.0 h	1.197	0.065	10.6 (0.6), 33.0 (21.8), 33.4 (40.1), 37.8 (11.9), 38.7 (2.92), 436.6 (0.334), 447.2 (2.24)
433.9	Ag-107	Ag-108	2.41 m	1932.14	0.48	21.1 (1.29), 23.9 (0.27), 618.9 (0.25), 633 (1.75)
433.9	Ag-107	Ag-108m	127 yr	17.1	90.7	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 614.4 (91.2), 633.0 (0.15), 722.9 (91.3)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
434.2	Cd-116	Cd-117	2.49 h	0.3735	9.79	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
436.1	Er-162	Er-163	75.0 m	2.66	0.028	46.7 (22.4), 47.5 (39.8), 53.8 (12.5), 55.3 (3.52), 439.9 (0.027), 1113.5 (0.049)
436.6	Ce-136	Ce-137	9.0 h	1.197	0.334	10.6 (0.6), 33.0 (21.8), 33.4 (40.1), 37.8 (11.9), 38.7 (2.92), 433.2 (0.065), 447.2 (2.24)
438.6	Zn-68	Zn-69m	13.76 h	1.353	94.8	8.6 (1.90), 9.6 (0.26), 318.6 (0.0013), 574.1 (0.033)
439.3	Pd-110	Pd-111m	5.5 h	0.43	0.23	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
439.4	Ge-76	Ge-77	11.3 h	0.468	0.19	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
439.9	Er-162	Er-163	75.0 m	2.66	0.027	46.7 (22.4), 47.5 (39.8), 53.8 (12.5), 55.3 (3.52), 436.1 (0.028), 1113.5 (0.049)
439.9	Ne-22	Ne-23	37.2 s	0.42412	33.0	1636.5 (1), 2076.4 (0.1)
441	Th-232	Th-233	22.3 m	740	0.23	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
442.4	Se-82	Se-83g	22.5 m	0.04888	1.09	225.2 (31.9), 356.7 (68.6), 485.7 (2.26), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
443.2	Hf-179	Hf-180m	5.5 h	4.67	82.8	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 215.2 (81.4), 332.3 (94.4), 500.7 (14.5)
443.8	Ru-102	Ru-103	39.25 d	38.24	0.32	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 497.1 (89.5), 557.0 (0.832), 610.3 (5.64)
444	Eu-151	Eu-152	13.33 yr	2 82 020	3.12	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
445.7	Nd-150	Pm-151	28.40 h		4.01	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
446.8	Ag-109	Ag-110m	249.9 d	226.4	3.66	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
447.2	Ce-136	Ce-137m	34.4 h	0.1805	3.03	33 (29.5), 33.4 (54.2), 34.3 (15.6), 34.7 (28.7), 37.8 (16.1), 38.7 (3.94), 39.2 (8.53), 40.2 (2.2), 254.3 (10.9)
447.2	Ce-136	Ce-137	9.0 h	1.197	2.24	10.6 (0.6), 33.0 (21.8), 33.4 (40.1), 37.8 (11.9), 38.7 (2.92), 433.2 (0.065), 436.6 (0.334)
447.7	Th-232	Th-233	22.3 m	740	0.15	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
448.7	Mo-100	Mo-101	14.6 m	1.916	0.69	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
451.5	Ho-165	Ho-166m	1.2×10^3 yr	350	3.09	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
451.6	Se-82	Se-83g	22.5 m	0.04888	0.82	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
452.3	Te-130	Te-131m	30 h	0.676	5.6	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
452.3	Te-130	Te-131	25.0 m	9.126	18.2	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)
453.1	Zn-70	Zn-71g	2.4 m	0.0514	0.18	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
453.1	Zn-70	Zn-71m	3.94 h	0.005	1.1	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
453.3	Tl-205	Tl-206m	3.75 m		81.0	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
453.7	Hf-178	Hf-179m_2	25.1 d	867.2	66.0	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
453.8	Xe-124	Xe-125	16.9 h	13.7	4.23	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 188.4 (54.9), 243.4 (28.8), 846.5 (1.03)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
454	Hf-177	Hf-178m$_2$	31 yr	0.00000186	16.3	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
454.4	Pd-110	Pd-111m	5.5 h	0.43	1.3	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
455.5	Xe-136	Xe-137	3.83 m	2.314	31.2	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
456.5	Pt-190	Pt-191	2.90 d	1.95	3.36	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 538.9 (13.7)
457.2	Tl-205	Tl-206m	3.75 m		21.5	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
457.4	Se-82	Se-83g	22.5 m	0.04888	3.50	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
459.2	Th-232	Th-233	22.3 m	740	0.15	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
459.6	Te-128	Te-129m	33.6 d	0.4755	4.54	27.2 (8.18), 27.5 (15.2), 27.8 (10.4), 31 (4.39), 695.9 (2.9)
459.6	Te-128	Te-129	69.6 m	6.34	7.14	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
460.5	Os-192	Os-193	30.5 h	82	3.95	63.3 (3.74), 64.9 (6.46), 73.0 (3.2), 73.5 (2.22), 138.9 (4.27)
460.6	Ru-96	Ru-97	2.9 d	1.6	0.12	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 215.7 (85.8), 324.5 (10.2), 569.3 (0.88)
461.4	Ge-76	Ge-77	11.3 h	0.468	1.20	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
463.1	In-115	In-116	14.10 s	3828.8	0.25	1252.6 (0.031), 1293.5 (1.30)
463.4	Sn-124	Sb-125	2.73 yr		10.5	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
465.8	Lu-176	Hf-177m$_1$	1.08 s		3.11	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
465.8	Hf-176	Hf-177m$_1$	1.08 s		3.11	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
465.8	Hf-176	Hf-177m$_2$	51.4 m		3.11	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
468.1	Ir-191	Ir-192m$_2$	241 yr	3.73	48.1	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
468.1	Ir-191	Ir-192	73.83 d	23 275.2	48.1	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
468.8	Ge-74	Ge-75	83 m	12.41	0.22	66.0 (0.11), 198.6 (1.17), 264.6 (11.3), 419.1 (0.18), 617.7 (0.11)
469	Mo-100	Mo-101	14.6 m	1.916	0.12	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
469.4	Ru-104	Ru-105	4.44 h	5.984	17.3	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
469.8	Sn-124	Sn-125	9.64 d	0.0224	1.38	332.1 (1.31), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
470.1	Ru-104	Ru-105	4.44 h	5.984	0.18	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
470.5	Te-120	Te-121m	154.0 d	0.032	1.4	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
470.5	Te-120	Te-121	16.8 d	0.192	1.4	26.1 (21.5), 26.4 (48.6), 29.7 (11.5), 30.4 (2.26), 37.1 (0.12), 65.5 (0.26), 507.6 (17.7), 573.1 (80.3)
472.7	Se-82	Se-83g	22.5 m	0.04888	0.14	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
475.4	Ge-76	Ge-77	11.3 h	0.468	0.94	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
475.4	Cs-133	Cs-134	2.06 yr	2650	1.46	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
478	Re-187	Re-188	16.98 h	4569.8	1.04	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
479.5	W-186	W-187	23.9 h	1081.08	25.3	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 551.6 (5.89), 618.4 (7.27), 685.8 (31.6), 772.9 (4.77)
479.5	Y-89	Y-90m	3.19 h	0.1	90.6	202.5 (96.5), 682 (0.32)
480.1	Gd-160	Gd-161	3.70 m	16.832	2.70	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 165.2 (2.60), 283.6 (6), 314.9 (22.9), 360.9 (60.6)
482	Hf-180	Hf-181	42.4 d	443.52	85.5	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 133.0 (43.0), 136.2 (6.1), 345.9 (14.0)
482.9	Ir-193	Ir-194m	171 d	376.2	96.9	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
484.5	Cd-114	Cd-115m	44.6 d	1.03	0.290	933.8 (2), 1290.6 (0.89)
484.6	Ir-191	Ir-192	73.83 d	23 275.2	2.17	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
484.6	Ir-191	$Ir\text{-}192m_2$	241 yr	3.73	3.16	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
485.7	Se-82	Se-83g	22.5 m	0.04888	2.26	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
485.9	Pd-110	Pd-111m	5.5 h	0.43	0.49	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
486.5	Ba-130	Ba-131	11.8 d	0.9328	2.09	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 496.3 (47.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
487	La-139	La-140	40.28 h	892.12	45.9	328.8 (20.7), 432.5 (2.99), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
487.3	Zn-70	Zn-71g	2.4 m	0.0514	0.12	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
487.3	Zn-70	Zn-71m	3.94 h	0.005	62.3	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
487.4	Te-128	Te-129	69.6 m	6.34	1.32	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
489.2	Ca-46	Ca-47	4.54 d	0.00296	6.74	530.4 (0.1), 767 (0.19), 807.9 (6.89), 1297.1 (74.9)
489.5	Ru-104	Ru-105	4.44 h	5.984	0.54	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
490.4	Ce-142	Ce-143	33.0 h	10.526	2.13	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 350.6 (3.27), 664.5 (5.6), 722 (5.32)
490.8	Th-232	Th-233	22.3 m	740	0.17	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
492.4	Sm-144	Sm-145	340 d	2.17	0.0031	38.2 (39.7), 38.7 (72), 43.8 (21.6), 44.9 (6), 61.3 (12.4)
492.4	Cd-114	Cd-115	2.23 d	86.16	8.03	24.1 (33.1), 27.4 (7.15), 260.9 (1.94), 336.2 (49.7), 527.9 (27.5)
492.7	Te-130	Te-131	25.0 m	9.126	4.84	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 602 (4.2), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)
493.8	Pt-198	Pt-199	30.8 m	26.64	5.73	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 543.0 (14.8), 714.5 (1.86)
495	Hf-177	Hf-178m_2	31 yr	0.00000186	68.7	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
496.3	Ba-130	Ba-131	11.8 d	0.9328	47.1	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09)
497	Mo-100	Mo-101	14.6 m	1.916	0.56	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
497.1	Ru-102	Ru-103	39.25 d	38.24	89.5	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 443.8 (0.32), 557.0 (0.832), 610.3 (5.64)
497.1	Pd-102	Pd-103	16.96 d	3.468	0.0040	20.1 (63.8), 22.8 (13.2), 39.7 (0.0683), 357.5 (0.0221), 497.1 (0.004)
497.4	Cd-114	In-115m	4.49 h		0.047	24.1 (27.7), 27.4 (6), 336.2 (45.8)
498.4	Sb-123	Sb-124m_2	20.2 m	0.811	22.0	602.7 (22), 645.8 (22), 1101.0 (0.3)
498.4	Sb-123	Sb-124m_1	60.2 d	1.58	20.0	602.7 (20), 645.8 (20), 1101.0 (0.3)
499	Th-232	Th-233	22.3 m	740	0.21	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
499.3	Ru-104	Ru-105	4.44 h	5.984	2.03	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
499.7	Mo-100	Mo-101	14.6 m	1.916	1.47	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
500.1	Ru-104	Ru-105	4.44 h	5.984	0.55	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
500.7	Hf-179	Hf-180m	5.5 h	4.67	14.5	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 215.2 (81.4), 332.3 (94.4), 443.2 (82.8)
502.6	Os-189	Os-190m	9.90 m	0.0041	97.8	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 186.7 (69.9), 361.1 (95.2), 616.1 (98.5)
505.1	Mo-100	Mo-101	14.6 m	1.916	1.2	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
505.9	Mo-100	Mo-101	14.6 m	1.916	11.8	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
507.6	Te-120	Te-121m	154.0 d	0.032	17.7	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
507.6	Zr-96	Zr-97	17.0 h	0.0632	5.05	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
507.6	Te-120	Te-121	16.8 d	0.192	17.7	26.1 (21.5), 26.4 (48.6), 29.7 (11.5), 30.4 (2.26), 37.1 (0.12), 65.5 (0.26), 470.5 (1.4), 573.1 (80.3)
507.8	Ni-64	Ni-65	2.52 h	1.3832	0.29	366.6 (4.61), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
508.8	Pr-141	Pr-142	19.13 h	760	0.023	641.2 (0.002), 1575.7 (3.7)
508.9	Pd-110	Pd-111m	5.5 h	0.43	0.16	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
508.9	Pd-110	Pd-111	23.4 m	2.22	0.21	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
510	Se-82	Se-83g	22.5 m	0.04888	44.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
510.1	Mo-100	Mo-101	14.6 m	1.916	0.98	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
511	Cu-63	Cu-64	12.7 h	311.265	35.8	1345.8 (0.48)
511.6	Zn-70	Zn-71m	3.94 h	0.005	28.4	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
511.6	Zn-70	Zn-71g	2.4 m	0.0514	32.0	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
512.8	Mo-100	Mo-101	14.6 m	1.916	1.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
513.4	Zr-96	Zr-97	17.0 h	0.0632	0.55	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
513.7	Ru-104	Ru-105	4.44 h	5.984	0.20	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
514	Kr-84	Kr-85g	10.72 yr	2.394	0.434	151.2 (0.0001), 362.8 (0.0001)
514	Sr-84	Sr-85g	64.84 d	0.196	99.3	13.4 (50.2), 15.0 (8.80), 151.2 (0.001), 868.1 (0.12)
514.1	Mo-100	Mo-101	14.6 m	1.916	0.81	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
515.4	Mo-100	Mo-101	14.6 m	1.916	0.51	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
515.5	Dy-164	Dy-165m	1.26 m	47 940	1.53	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 108.2 (3.01), 153.8 (0.242), 361.7 (0.534)
516.1	Mo-100	Tc-101	14.2 m		0.11	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
519.3	Pd-110	Pd-111m	5.5 h	0.43	0.14	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
520.3	Ge-76	Ge-77	11.3 h	0.468	0.28	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
520.4	Se-82	Br-83	2.39 h		0.067	529.6 (1.4), 552.6 (0.017)
520.8	Ge-76	As-77	38.83 h		0.42	88 (0.27), 162 (0.16), 239.1 (1.59), 250 (0.41)
523	Kr-78	Kr-79g	35.04 h	2.1	0.25	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
524.2	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
525.3	Kr-78	Kr-79g	35.04 h	2.1	0.43	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
525.6	Pd-110	Pd-111m	5.5 h	0.43	1.2	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
526.6	Xe-134	Xe-135m	15.6 m	0.0312	80.5	29.5 (3.92), 29.8 (7.28), 33.6 (2.14), 34.4 (0.47), 786.9 (0.004)
527.4	Hf-176	Hf-177m_2	51.4 m		7.14	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
527.9	Cd-114	Cd-115	2.23 d	86.16	27.5	24.1 (33.1), 27.4 (7.15), 260.9 (1.94), 336.2 (49.7), 492.4 (8.03)
529.6	Se-82	Br-83	2.39 h		1.40	520.4 (0.067), 552.6 (0.017)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
529.8	Ho-165	Ho-166m	1.2×10^3 yr	350	10.3	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
530.4	Ca-46	Ca-47	4.54 d	0.00296	0.1	489.2 (6.74), 767 (0.19), 807.9 (6.89), 1297.1 (74.9)
531	Nd-146	Nd-147	10.98 d	24.066	13.0	38.2 (13), 38.7 (23.5), 43.8 (7.05), 44.9 (1.96), 91.1 (27.9), 319.4 (1.95)
531.4	Mo-100	Tc-101	14.2 m		1.02	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
533.1	Se-82	Se-83g	22.5 m	0.04888	4.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
533.5	Mo-100	Mo-101	14.6 m	1.916	0.40	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
535	Hf-177	Hf-178m_2	31 yr	0.00000186	8.86	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
538.9	Pt-190	Pt-191	2.90 d	1.95	13.7	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36)
539.3	Ru-104	Ru-105	4.44 h	5.984	0.11	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
540.5	Nd-148	Nd-149	1.73 h	14.4	6.58	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
541.8	Sn-122	Sn-123m	40.08 m	0.828	0.020	26.1 (3.06), 26.4 (5.72), 29.7 (1.63), 30.4 (0.32), 160.3 (85.6), 381.4 (0.042)
543	Pt-198	Pt-199	30.8 m	26.64	14.8	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 714.5 (1.86)
545.1	Mo-100	Tc-101	14.2 m		5.98	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
545.8	Dy-164	Dy-165	2.33 h	28 200	0.16	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						361.7 (0.84), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
547	Pd-110	Pd-111m	5.5 h	0.43	0.27	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
547	Pd-110	Pd-111	23.4 m	2.22	0.37	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
551.6	W-186	W-187	23.9 h	1081.08	5.89	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 618.4 (7.27), 685.8 (31.6), 772.9 (4.77)
552.2	Pd-110	Pd-111m	5.5 h	0.43	0.26	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
552.4	Se-80	Se-81m	57.3 m	3.968	0.15	103.1 (9.79), 275.9 (1.3), 290.1 (1.0), 566 (0.38), 828.3 (0.47)
552.4	Se-80	Se-81	18.5 m	26.288	0.10	275.9 (0.87), 290.1 (0.68), 566.0 (0.26), 828.3 (0.32)
552.6	Se-82	Br-83	2.39 h		0.017	520.4 (0.067), 529.6 (1.4)
552.9	Cd-116	Cd-117m	3.36 h	0.1867	125.0	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
554.3	Br-81	Br-82g	35.30 h	12.8206	70.6	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
555.8	Rh-103	Rh-104m	4.34 m	1000	2.38	19.2 (0.28), 20.2 (55.1), 22.8 (11.4), 51.4 (48.2), 77.5 (2.07), 97.1 (3.0), 767.8 (0.10)
555.8	Rh-103	Rh-104	42.3 s	13 500	1.99	19.2 (0.24), 21.7 (0.050), 358.1 (0.016), 1237 (0.066)
556.1	Rb-85	Rb-86m	1.02 m	3.82501	98.2	13.4 (0.88), 15.0 (0.16)
556.4	Pd-110	Pd-111m	5.5 h	0.43	0.19	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
557	Ru-102	Ru-103	39.25 d	38.24	0.832	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 443.8 (0.32), 497.1 (89.5), 610.3 (5.64)
558	In-113	In-114m	49.51 d	34.668	3.40	23.1 (3.91), 24.1 (28.1), 27.4 (6.08), 191.6 (16), 725.2 (3.4)
558	In-113	In-114	71.9 s	16.69	0.07	23.1 (2), 26.2 (0.42), 575.7 (0.004), 1300 (0.14)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
558	Ge-76	Ge-77	11.3 h	0.468	15.2	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
559.1	As-75	As-76	26.32 h	450	44.7	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
559.2	Ru-104	Ru-105	4.44 h	5.984	0.11	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
561.9	Zr-94	Nb-95	34.98 d		0.015	204.1 (0.015), 765.8 (99.9)
562.4	Ir-193	Ir-194m	171 d	376.2	69.9	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
563.2	Cs-133	Cs-134	2.06 yr	2650	8.38	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
563.2	As-75	As-76	26.32 h	450	1.17	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
564	Sb-121	Sb-122	2.70 d	335.2	70.8	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 692.6 (3.68), 1141.1 (0.57), 1256.8 (0.77)
564.2	Tl-205	Tl-206m	3.75 m		10.0	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
564.4	Cd-116	Cd-117m	3.36 h	0.1867	14.7	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
565.7	Dy-164	Dy-165	2.33 h	28 200	0.13	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
566	Se-80	Se-81	18.5 m	26.288	0.26	275.9 (0.87), 290.1 (0.68), 552.4 (0.10), 828.3 (0.32)
566	Se-80	Se-81m	57.3 m	3.968	0.38	103.1 (9.79), 275.9 (1.3), 290.1 (1.0), 552.4 (0.15), 828.3 (0.47)
566.2	Zn-70	Zn-71m	3.94 h	0.005	0.19	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
566.6	Mo-100	Mo-101	14.6 m	1.916	0.73	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
569.3	Ru-96	Ru-97	2.9 d	1.6	0.88	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 215.7 (85.8), 324.5 (10.2), 460.6 (0.12)
569.3	Cs-133	Cs-134	2.06 yr	2650	15.4	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
571	Ho-165	Ho-166m	1.2×10^3 yr	350	5.81	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
571.3	As-75	As-76	26.32 h	450	0.14	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
571.7	Mo-100	Mo-101	14.6 m	1.916	0.19	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
571.9	Se-82	Se-83g	22.5 m	0.04888	4.46	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
573.1	Te-120	Te-121m	154.0 d	0.032	80.3	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
573.1	Te-120	Te-121	16.8 d	0.192	80.3	26.1 (21.5), 26.4 (48.6), 29.7 (11.5), 30.4 (2.26), 37.1 (0.12), 65.5 (0.26), 470.5 (1.4), 507.6 (17.7)
574.1	Zn-68	Zn-69m	13.76 h	1.353	0.033	8.6 (1.90), 9.6 (0.26), 318.6 (0.0013), 438.6 (94.8)
574.2	Hf-177	Hf-178m_2	31 yr	0.00000186	83.6	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
574.9	Zn-70	Zn-71m	3.94 h	0.005	0.11	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
575	Pd-110	Pd-111m	5.5 h	0.43	3.0	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
575.1	Ru-104	Ru-105	4.44 h	5.984	0.84	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
575.7	In-113	In-114	71.9 s	16.69	0.004	23.1 (2), 26.2 (0.42), 558 (0.07), 1300 (0.14)
580	Pd-110	Pd-111m	5.5 h	0.43	0.63	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
580	Pd-110	Pd-111	23.4 m	2.22	0.837	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
581.6	Se-82	Se-83g	22.5 m	0.04888	0.34	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
582.1	Zr-94	Nb-95m	3.61 d		0.055	16.6 (36), 18.6 (6.97), 204.1 (2.36), 235.7 (25.1)
582.5	Ge-76	Ge-77	11.3 h	0.468	0.74	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
583.6	Pd-110	Pd-111m	5.5 h	0.43	0.25	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
586.3	Te-130	Te-131m	30 h	0.676	1.97	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
587.4	Ga-71	Ga-72	14.1 h	187.929	0.12	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
588.6	Ir-191	Ir-192m_2	241 yr	3.73	4.58	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 604.4 (8.33), 612.4 (5.43)
588.6	Ir-191	Ir-192	73.83 d	23 275.2	4.58	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 604.4 (8.33), 612.4 (5.43)
589.2	Ir-193	Ir-194	19.15 h	6897	0.14	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
589.6	Sn-124	Sn-125m	9.52 m	0.728	0.2	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 643 (0.16), 1404 (0.7), 1483.9 (0.18), 1615.3 (0.12)
590.1	Mo-100	Mo-101	14.6 m	1.916	5.6	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
590.9	Mo-100	Mo-101	14.6 m	1.916	16.4	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
590.9	Nd-148	Pm-149	2.21 d		0.68	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 285.9 (3.10), 830.5 (0.03), 833.2 (0.03)
591.8	Eu-153	Eu-154	8.8 yr		4.83	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
593.4	Se-82	Se-83g	22.5 m	0.04888	0.75	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
595.2	Th-232	Th-233	22.3 m	740	0.16	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
595.4	Pd-110	Pd-111m	5.5 h	0.43	0.12	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
596.1	Zn-70	Zn-71m	3.94 h	0.005	27.9	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
600.5	Ir-193	Ir-194m	171 d	376.2	62.3	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
600.6	Sn-124	Sb-125	2.73 yr		17.8	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
600.9	Ga-71	Ga-72	14.1 h	187.929	5.54	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
602	Te-130	Te-131	25.0 m	9.126	4.20	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
602.4	Zr-96	Zr-97	17.0 h	0.0632	1.4	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
602.7	Sb-123	Sb-124	60.2 d	175.07	98.4	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
602.7	Sb-123	$Sb\text{-}124m_1$	60.2 d	1.58	20.0	498.4 (20), 645.8 (20), 1101.0 (0.3)
602.7	Sb-123	$Sb\text{-}124m_2$	20.2 m	0.811	22.0	498.4 (22), 645.8 (22), 1101.0 (0.3)
603	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
604.4	Ir-191	$Ir\text{-}192m_2$	241 yr	3.73	8.33	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 612.4 (5.43)
604.4	Ir-191	Ir-192	73.83 d	23 275.2	8.33	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 604.4 (8.33), 612.4 (5.43)
604.7	Cs-133	Cs-134	2.06 yr	2650	97.6	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
606.1	Kr-78	Kr-79g	35.04 h	2.1	8.10	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
606.3	Br-81	Br-82g	35.30 h	12.8206	1.17	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
606.5	Hf-176	$Hf\text{-}177m_2$	51.4 m		11.6	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
606.6	Sn-124	Sb-125	2.73 yr		5.02	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
606.8	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
608.2	Xe-134	Xe-135	9.08 h	2.704	2.91	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 249.8 (90.4), 358.4 (0.22), 408 (0.36)
608.3	Mo-100	Mo-101	14.6 m	1.916	1.07	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
608.5	Ti-50	Ti-51	5.8 m	0.9308	1.18	320.1 (93), 928.6 (6.88)
609.2	Se-82	Se-83g	22.5 m	0.04888	3.09	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
609.3	Ni-64	Ni-65	2.52 h	1.3832	0.14	366.6 (4.61), 507.8 (0.29), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
610.3	Ru-102	Ru-103	39.25 d	38.24	5.64	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 443.8 (0.32), 497.1 (89.5), 557.0 (0.832)
611.6	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
612.4	Ir-191	Ir-192m_2	241 yr	3.73	5.43	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33)
612.4	Ir-191	Ir-192	73.83 d	23 275.2	5.43	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33)
614.4	Ag-107	Ag-108m	127 yr	17.1	91.2	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 433.9 (90.7), 633.0 (0.15), 722.9 (91.3)
614.4	Ge-76	Ge-77	11.3 h	0.468	0.56	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
616.1	Os-189	Os-190m	9.90 m	0.0041	98.5	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 186.7 (69.9), 361.1 (95.2), 502.6 (97.8)
616.2	Br-79	Br-80g	17.4 m	435.934	6.70	639.2 (0.24), 665.6 (1.15), 704.0 (0.19)
616.2	Br-79	Br-80m	4.42 h	121.656	7.17	37.1 (39.1), 48.9 (0.33), 639.2 (0.26), 665.6 (1.23), 704 (0.2)
617.7	Ge-74	Ge-75	83 m	12.41	0.11	66.0 (0.11), 198.6 (1.17), 264.6 (11.3), 419.1 (0.18), 468.8 (0.22)
618.4	W-186	W-187	23.9 h	1081.08	7.27	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 551.6 (5.89), 685.8 (31.6), 772.9 (4.77)
618.9	Ag-107	Ag-108	2.41 m	1932.14	0.25	21.1 (1.29), 23.9 (0.27), 433.9 (0.48), 633 (1.75)
619.1	Br-81	Br-82g	35.30 h	12.8206	43.1	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
620.1	Pd-110	Pd-111	23.4 m	2.22	0.11	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
620.1	Pd-110	Ag-111m	64.8 s		0.11	22.1 (15.6), 25.0 (3.30), 59.8 (0.531), 171.3 (0.11), 245.4 (0.46)
620.2	Zn-70	Zn-71m	3.94 h	0.005	56.7	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
620.3	Ag-109	Ag-110m	249.9 d	226.4	2.78	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
620.6	Dy-164	Dy-165	2.33 h	28 200	0.10	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 633.4 (0.57), 715.3 (0.53)
621.6	Se-82	Se-83g	22.5 m	0.04888	3.22	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
622	Ir-193	Ir-194	19.15 h	6897	0.33	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
623.2	Pd-110	Pd-111m	5.5 h	0.43	0.59	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
623.2	Pd-110	Pd-111	23.4 m	2.22	0.28	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
624.8	Ge-76	Ge-77	11.3 h	0.468	0.17	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
625.6	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
626.2	Ag-109	Ag-110m	249.9 d	226.4	0.23	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
627	Mo-100	Tc-101	14.2 m		0.39	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
630	Ga-71	Ga-72	14.1 h	187.929	24.8	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
631.2	Se-82	Se-83m	70.4 s	0.3666	0.47	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
631.8	Cd-116	Cd-117m	3.36 h	0.1867	2.80	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
631.8	Ge-76	Ge-77	11.3 h	0.468	6.59	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
632.3	Ru-104	Ru-105	4.44 h	5.984	0.15	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
632.5	Ba-132	Ba-133m	38.9 h	0.0505	0.01	12.3 (1.5), 31.8 (15.3), 32.2 (28.2), 36.4 (8.34), 37.3 (1.99), 276.1 (17.5)
632.8	Pd-110	Pd-111m	5.5 h	0.43	3.4	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
633	Ag-107	Ag-108	2.41 m	1932.14	1.75	21.1 (1.29), 23.9 (0.27), 433.9 (0.48), 618.9 (0.25)
633	Ag-107	Ag-108m	127 yr	17.1	0.15	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 433.9 (90.7), 614.4 (91.2), 722.9 (91.3)
633	Re-187	Re-188	16.98 h	4569.8	1.25	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
633.4	Dy-164	Dy-165	2.33 h	28 200	0.57	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 715.3 (0.53)
634.4	Ge-76	Ge-77	11.3 h	0.468	1.97	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
635.1	Re-187	Re-188	16.98 h	4569.8	0.15	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
635.9	Sn-124	Sb-125	2.73 yr		11.3	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
637	Te-130	I-131	8.02 d		7.27	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 284.3 (6.06), 364.5 (81.2), 722.9 (1.8)
638	Sn -112	Sn-113	115.1 d	0.71	0.0010	24.1 (79.8), 27.4 (17.3), 255 (1.85), 391.7 (64.2)
638.2	Hf-176	Hf-177m_2	51.4 m		20.1	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
638.7	Ru-104	Ru-105	4.44 h	5.984	0.22	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
639.2	Br-79	Br-80g	17.4 m	435.934	0.24	616.2 (6.70), 665.6 (1.15), 704.0 (0.19)
639.2	Br-79	Br-80m	4.42 h	121.656	0.26	37.1 (39.1), 48.9 (0.33), 616.2 (7.17), 665.6 (1.23), 704 (0.2)
641.2	Pr-141	Pr-142	19.13 h	760	0.002	508.8 (0.023), 1575.7 (3.7)
642.8	Mo-100	Mo-101	14.6 m	1.916	1.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
643	Sn-124	Sn-125m	9.52 m	0.728	0.16	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 1404 (0.7), 1483.9 (0.18), 1615.3 (0.12)
645.1	Ir-193	Ir-194	19.15 h	6897	1.16	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
645.6	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
645.8	Sb-123	Sb-124	60.2 d	175.07	7.24	602.7 (98.4), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
645.8	Sb-123	Sb-124m_1	60.2 d	1.58	20.0	498.4 (20), 602.7 (20), 1101.0 (0.3)
645.8	Sb-123	Sb-124m_2	20.2 m	0.811	22.0	498.4 (22), 602.7 (22), 1101.0 (0.3)
646.1	Os-184	Os-185	93.6 d	60	80.8	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 717.4 (4.11), 874.8 (6.59), 880.5 (4.98)
647.3	Pd-108	Pd-109	13.46 h	219.6	0.024	22.1 (28.5), 25.0 (6.02), 88.0 (3.61), 311.4 (0.032)
649.8	Bi-209	Bi-210m	3.0×10^6 yr	1.4	2.9	70.8 (3.9), 72.9 (6.6), 82.5 (2.3), 265.7 (51.0), 304.8 (28.0)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
650.4	Pd-110	Pd-111m	5.5 h	0.43	0.41	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
650.4	Pd-110	Pd-111	23.4 m	2.22	0.55	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
652.7	Ru-104	Ru-105	4.44 h	5.984	0.30	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
654.7	Pd-110	Pd-111m	5.5 h	0.43	0.14	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
654.8	Nd-148	Nd-149	1.73 h	14.4	7.95	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
656.2	Ru-104	Ru-105	4.44 h	5.984	2.03	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
657	As-75	As-76	26.32 h	450	0.14	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
657.7	Ag-109	Ag-110m	249.9 d	226.4	94.7	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
657.7	Ag-109	Ag-110	24.6 s	4157.07	4.49	21.1 (0.17), 23.9 (0.037), 815.3 (0.038), 1125.7 (0.015)
657.9	Zr-96	Zr-97	17.0 h	0.0632	106.0	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
657.9	Zr-96	Nb-97	72.1 m		98.5	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
660.7	Mo-100	Mo-101	14.6 m	1.916	0.22	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
661.6	Xe-136	Cs-137	30.14 yr		85.1	31.8 (2.02), 32.2 (3.72), 36.4 (1.1), 37.3 (0.26)
661.6	Xe-136	Ba-137m	2.55 m		89.9	31.8 (2.13), 32.2 (3.93), 36.4 (1.16), 37.3 (0.28)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
661.6	Ba-136	Ba-137m	2.55 m	0.0785	89.9	31.8 (2.13), 32.2 (3.93), 36.4 (1.16), 37.3 (0.28)
664.5	Ce-142	Ce-143	33.0 h	10.526	5.60	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 350.6 (3.27), 490.4 (2.13), 722 (5.32)
664.8	Se-82	Se-83g	22.5 m	0.04888	3.22	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
665	Te-130	Te-131m	30 h	0.676	4.32	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
665.3	As-75	As-76	26.32 h	450	0.39	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
665.6	Br-79	Br-80m	4.42 h	121.656	1.23	37.1 (39.1), 48.9 (0.33), 616.2 (7.17), 639.2 (0.26), 704 (0.2)
665.6	Br-79	Br-80g	17.4 m	435.934	1.15	616.2 (6.70), 639.2 (0.24), 704.0 (0.19)
666.8	Zn-70	Zn-71g	2.4 m	0.0514	0.90	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
668.5	Pd-110	Pd-111m	5.5 h	0.43	0.93	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
669.8	Th-232	Th-233	22.3 m	740	0.68	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
670.5	Ho-165	Ho-166m	1.2×10^3 yr	350	5.83	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
671.4	Sn-124	Sb-125	2.73 yr		1.80	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
672.5	Re-187	Re-188	16.98 h	4569.8	0.11	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
673.1	Ge-76	Ge-77	11.3 h	0.468	0.63	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
673.8	Kr-86	Kr-87	76.3 m	0.0519	1.89	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
673.9	Se-82	Se-83m	70.4 s	0.3666	15.1	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
675.9	Au-197	Au-198	2.70 d		1.06	68.9 (0.81), 70.8 (1.38), 80.2 (0.47), 82.5 (0.13), 411.8 (95.5), 1087.7 (0.23)
676.4	Ru-104	Ru-105	4.44 h	5.984	15.5	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
676.6	Ag-109	Ag-110m	249.9 d	226.4	0.14	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
677.6	Ag-109	Ag-110m	249.9 d	226.4	10.7	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
677.8	Nd-150	Nd-151	12.4 m	6.768	2.72	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
679.4	Se-82	Se-83g	22.5 m	0.04888	1.10	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
680.5	Tl-205	Tl-206m	3.75 m		79.3	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
682	Y-89	Y-90m	3.19 h	0.1	0.32	202.5 (96.5), 479.5 (90.6)
684.7	Mo-92	Mo-93m	6.85 h	0.089	99.7	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 949.8 (0.12), 1363.0 (0.79), 1477.1 (99.1)
685.8	W-186	W-187	23.9 h	1081.08	31.6	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 551.6 (5.89), 618.4 (7.27), 772.9 (4.77)
687	Ag-109	Ag-110m	249.9 d	226.4	6.49	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
687.8	Ir-193	Ir-194m	171 d	376.2	59.1	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
690.6	Zr-96	Zr-97	17.0 h	0.0632	0.25	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
692.6	Sb-121	Sb-122	2.70 d	335.2	3.68	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 564 (70.8), 1141.1 (0.57), 1256.8 (0.77)
694.2	Pd-110	Pd-111m	5.5 h	0.43	1.9	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
695.6	Mo-100	Mo-101	14.6 m	1.916	7.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
695.9	Te-128	Te-129m	33.6 d	0.4755	2.90	27.2 (8.18), 27.5 (15.2), 27.8 (10.4), 31 (4.39), 459.6 (4.54)
698.3	Br-81	Br-82m	6.13 m	119.8233	0.024	46.0 (0.24), 776.5 (0.2), 1474.8 (0.016)
698.3	Br-81	Br-82g	35.30 h	12.8206	27.9	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
698.5	Ge-76	Ge-77	11.3 h	0.468	0.22	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
699.2	Zr-96	Zr-97	17.0 h	0.0632	0.12	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
701.8	Mo-100	Mo-101	14.6 m	1.916	0.38	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
702.6	Nb-93	Nb-94	2.03×10^4 yr	100	100.0	871.1 (100)
702.6	Nb-93	Nb-94m	6.26 m	15	0.0030	16.6 (37.1), 18.6 (7.18), 41.0 (0.08), 871.1 (0.48)
703.8	Pd-110	Pd-111m	5.5 h	0.43	0.63	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
703.8	Zr-96	Zr-97	17.0 h	0.0632	0.93	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
704	Br-79	Br-80m	4.42 h	121.656	0.20	37.1 (39.1), 48.9 (0.33), 616.2 (7.17), 639.2 (0.26), 665.6 (1.23)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
704	Br-79	Br-80g	17.4 m	435.934	0.19	616.2 (6.70), 639.2 (0.24), 665.6 (1.15)
705.2	Ge-76	Ge-77	11.3 h	0.468	0.10	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
706.1	Se-82	Se-83g	22.5 m	0.04888	0.2	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
706.7	Ag-109	Ag-110m	249.9 d	226.4	16.7	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
708.1	Ag-109	Ag-110m	249.9 d	226.4	0.28	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
709.8	Pd-110	Pd-111	23.4 m	2.22	0.13	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
711.7	Ho-165	Ho-166m	1.2×10^3 yr	350	59.3	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
712.1	Se-82	Se-83g	22.5 m	0.04888	2.67	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
712.3	Ge-76	Ge-77	11.3 h	0.468	0.78	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
713	Mo-100	Mo-101	14.6 m	1.916	3.38	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
713.8	Sb-123	Sb-124	60.2 d	175.07	2.40	602.7 (98.4), 645.8 (7.24), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
714.3	Ge-76	Ge-77	11.3 h	0.468	6.77	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
714.5	Pt-198	Pt-199	30.8 m	26.64	1.86	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
715.3	Dy-164	Dy-165	2.33 h	28 200	0.53	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57)
715.6	Mo-100	Tc-101	14.2 m		0.69	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
717.4	Os-184	Os-185	93.6 d	60	4.11	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 646.1 (80.8), 874.8 (6.59), 880.5 (4.98)
717.6	Nd-150	Pm-151	28.40 h		3.99	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
718	Se-82	Se-83g	22.5 m	0.04888	16.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
718.9	Pd-110	Pd-111m	5.5 h	0.43	0.17	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
720	Mo-100	Tc-101	14.2 m		0.23	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
721.4	Zn-70	Zn-71g	2.4 m	0.0514	0.54	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
722	Ce-142	Ce-143	33.0 h	10.526	5.32	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 350.6 (3.27), 490.4 (2.13), 664.5 (5.6)
722.8	Sb-123	Sb-124	60.2 d	175.07	11.3	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
722.9	Te-130	I-131	8.02 d		1.80	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 284.3 (6.06), 364.5 (81.2), 637 (7.27)
722.9	Ag-107	Ag-108m	127 yr	17.1	91.3	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 433.9 (90.7), 614.4 (91.2), 633 (0.15)
723.3	Eu-153	Eu-154	8.8 yr		19.7	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
724.2	Ru-104	Ru-105	4.44 h	5.984	46.7	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
724.2	Zr-94	Zr-95	64.03 d	0.86227	43.7	16.6 (0.42), 18.6 (0.66), 235.7 (0.24), 756.7 (55.4)
724.8	Pd-110	Pd-111m	5.5 h	0.43	0.27	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
725.2	In-113	In-114m	49.51 d	34.668	3.4	23.1 (3.91), 24.1 (28.1), 27.4 (6.08), 191.6 (16), 558 (3.4)
733	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
735.1	Se-82	Se-83g	22.5 m	0.04888	0.82	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
735.7	Ga-71	Ga-72	14.1 h	187.929	0.37	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
736.4	Nd-150	Nd-151	12.4 m	6.768	7.19	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
739.4	Mo-98	Mo-99	2.75 d	3.1369	12.1	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
739.5	Mo-100	Mo-101	14.6 m	1.916	0.30	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
740.1	As-75	As-76	26.32 h	450	0.12	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
742	Ce-142	Pr-143	13.58 d		0.0001	–
743.4	Zr-96	Nb-97m	53 s		98.0	16.6 (1.10), 18.7 (0.21)
743.4	Zr-96	Zr-97	17.0 h	0.0632	92.6	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
743.6	Ge-76	Ge-77	11.3 h	0.468	0.17	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
744.3	Ag-109	Ag-110m	249.9 d	226.4	4.66	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
745.6	Pd-110	Pd-111m	5.5 h	0.43	0.12	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
745.7	Ge-76	Ge-77	11.3 h	0.468	0.91	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
748.1	Cd-116	Cd-117m	3.36 h	0.1867	4.5	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
749.9	Ge-76	Ge-77	11.3 h	0.468	0.84	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
751.8	La-139	La-140	40.28 h	892.12	4.41	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
752.3	Ho-165	Ho-166m	1.2×10^3 yr	350	13.2	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
753	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
753.4	Zn-70	Zn-71m	3.94 h	0.005	3.3	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
754.2	Ce-138	Ce-139m	56.4 s	0.00375	92.4	34.3 (1.58), 34.7 (2.9), 39.2 (0.86), 40.2 (0.22)
756.7	Zr-94	Zr-95	64.03 d	0.86227	55.4	16.6 (0.42), 18.6 (0.66), 235.7 (0.24), 724.2 (43.7)
756.9	Eu-153	Eu-154	8.8 yr		40.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
762.2	Pd-110	Pd-111m	5.5 h	0.43	1.2	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
763.9	Ag-109	Ag-110m	249.9 d	226.4	22.4	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
764.4	Th-232	Th-233	22.3 m	740	0.12	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
765.3	Tb-159	Tb-160	72.3 d	2550	2.04	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
765.8	Zr-94	Nb-95	34.98 d		99.9	204.1 (0.015), 561.9 (0.015)
766.7	Ge-76	Ge-77	11.3 h	0.468	0.74	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
767	Ca-46	Ca-47	4.54 d	0.00296	0.19	489.2 (6.74), 530.4 (0.1), 807.9 (6.89), 1297.1 (74.9)
767.5	Re-185	Re-186	3.78 d	4188.8	0.029	58.0 (1.55), 59.3 (2.69), 61.5 (1.11), 63.0 (1.92), 67.2 (0.91), 69.1 (0.23), 71.3 (0.66), 73.4 (0.17), 122.7 (0.72), 137.2 (9.20)
767.8	Rh-103	Rh-104m	4.34 m	1000	0.10	19.2 (0.28), 20.2 (55.1), 22.8 (11.4), 51.4 (48.2), 77.5 (2.07), 97.1 (3.0), 555.8 (2.38)
770.6	Zn-64	Zn-65	244.3 d	36.936	0.003	344.9 (0.003), 1115.5 (50.7)
771.3	Zn-70	Zn-71m	3.94 h	0.005	2.0	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
771.8	As-75	As-76	26.32 h	450	0.12	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
772.9	W-186	W-187	23.9 h	1081.08	4.77	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 551.6 (5.89), 618.4 (7.27), 685.8 (31.6)
773.7	Te-130	Te-131m	30 h	0.676	38.1	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
774.2	Mo-100	Mo-101	14.6 m	1.916	0.33	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
775.8	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
776.5	Br-81	Br-82m	6.13 m	119.8233	0.20	46.0 (0.24), 698.3 (0.024), 1474.8 (0.016)
776.5	Br-81	Br-82g	35.30 h	12.8206	83.4	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
777.8	Mo-98	Mo-99	2.75 d	3.1369	4.36	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 822.8 (0.13), 961.0 (0.10)
778.3	Mo-100	Mo-101	14.6 m	1.916	0.96	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
778.8	Ho-165	Ho-166m	1.2×10^3 yr	350	3.33	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
778.9	Eu-151	Eu-152	13.33 yr	2 82 020	13.0	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
781.3	Ge-76	Ge-77	11.3 h	0.468	0.96	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
782.5	Te-130	Te-131m	30 h	0.676	7.76	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
784.8	Ge-76	Ge-77	11.3 h	0.468	1.24	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
786.4	Ga-71	Ga-72	14.1 h	187.929	3.20	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
786.9	Xe-134	Xe-135m	15.6 m	0.0312	0.004	29.5 (3.92), 29.8 (7.28), 33.6 (2.14), 34.4 (0.47), 526.6 (80.5)
790.1	Mo-100	Mo-101	14.6 m	1.916	0.13	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
793.8	Te-130	Te-131m	30 h	0.676	13.8	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
794.3	Ge-76	Ge-77	11.3 h	0.468	0.26	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
795.8	Cs-133	Cs-134	2.06 yr	2650	85.4	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
796.5	Cd-106	Cd-107	6.50 h	0.25	0.065	22.1 (89), 25 (18.8), 93.1 (4.68), 796.5 (0.065), 828.9 (0.163)
797.5	Nd-150	Nd-151	12.4 m	6.768	5.51	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
797.8	Pd-110	Pd-111m	5.5 h	0.43	0.97	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
799	Se-82	Se-83g	22.5 m	0.04888	16.0	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
799	Se-82	Se-83m	70.4 s	0.3666	1.23	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
800.3	Sn-124	Sn-125	9.64 d	0.0224	1.0	332.1 (1.31), 469.8 (1.38), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
801.9	Cs-133	Cs-134	2.06 yr	2650	8.73	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
802.1	Te-128	Te-129	69.6 m	6.34	0.18	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 1083.9 (0.46), 1111.6 (0.18)
803.1	Tl-205	Tl-206	4.20 m		0.0055	–
803.1	Bi-209	Po-210	138.38 d		0.0012	10.5 (0.013), 12.6 (0.11), 14.9 (0.0026)
803.1	Bi-209	Tl-206	4.20 m		0.0055	–
804.3	Mo-100	Mo-101	14.6 m	1.916	1.00	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
804.5	Zr-96	Zr-97	17.0 h	0.0632	0.65	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
807.9	Ca-46	Ca-47	4.54 d	0.00296	6.89	489.2 (6.74), 530.4 (0.1), 767 (0.19), 1297.1 (74.9)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
810.2	Ga-71	Ga-72	14.1 h	187.929	2.01	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
810.3	Ho-165	Ho-166m	1.2×10^3 yr	350	63.3	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
810.4	Ge-76	Ge-77	11.3 h	0.468	2.15	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
813.4	Ge-76	Ge-77	11.3 h	0.468	0.12	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
814.3	Kr-86	Kr-87	76.3 m	0.0519	0.16	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
815.3	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
815.3	Ag-109	Ag-110	24.6 s	4157.07	0.038	21.1 (0.17), 23.9 (0.037), 657.7 (4.49), 1125.7 (0.015)
815.8	La-139	La-140	40.28 h	892.12	23.6	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
818	Ag-109	Ag-110m	249.9 d	226.4	7.32	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
818.7	In-115	In-116m$_1$	54.15 m	7753.32	11.5	138.3 (3.29), 416.9 (29.2), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
822	Ru-104	Ru-105	4.44 h	5.984	845.9	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
822.5	Sn-124	Sn-125	9.64 d	0.0224	3.99	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
822.8	Mo-98	Mo-99	2.75 d	3.1369	0.13	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 961.0 (0.10)
822.8	Te-130	Te-131m	30 h	0.676	6.10	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
823.1	Ge-76	Ge-77	11.3 h	0.468	0.57	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
826.2	Co-59	Co-60m	10.47 m	2040	0.0083	6.9 (27.2), 7.7 (3.67), 58.6 (2.02), 1332.5 (0.25)
827.8	Br-81	Br-82g	35.30 h	12.8206	24.2	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
828.3	Se-80	Se-81m	57.3 m	3.968	0.47	103.1 (9.79), 275.9 (1.3), 290.1 (1.0), 552.4 (0.15), 566.0 (0.38)
828.3	Se-80	Se-81	18.5 m	26.288	0.32	275.9 (0.87), 290.1 (0.68), 552.4 (0.10), 566 (0.26)
828.3	Pd-110	Pd-111m	5.5 h	0.43	0.16	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
828.9	Cd-106	Cd-107	6.50 h	0.25	0.163	22.1 (89), 25 (18.8), 93.1 (4.68), 796.5 (0.065)
829.5	Re-187	Re-188	16.98 h	4569.8	0.41	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 931.3 (0.56), 1610.4 (0.10)
829.8	Zr-96	Zr-97	17.0 h	0.0632	0.22	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
830.5	Nd-148	Pm-149	2.21 d		0.030	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 285.9 (3.10), 590.9 (0.68), 833.2 (0.03)
830.6	Ho-165	Ho-166m	1.2×10^3 yr	350	10.7	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
831.8	Cd-116	Cd-117	2.49 h	0.3735	2.26	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
832	Kr-78	Kr-79g	35.04 h	2.1	1.26	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
833.2	Nd-148	Pm-149	2.21 d		0.030	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 285.9 (3.10), 590.9 (0.68), 830.5 (0.03)
833.4	Cu-65	Cu-66	5.1 m	66.9011	0.18	1039.2 (8), 1333.2 (0.003)
834	Ga-71	Ga-72	14.1 h	187.929	95.6	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
835.7	Pd-110	Pd-111m	5.5 h	0.43	0.20	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
835.7	Pd-110	Pd-111	23.4 m	2.22	0.27	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
836.4	Kr-86	Kr-87	76.3 m	0.0519	0.77	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
836.5	Se-82	Se-83g	22.5 m	0.04888	15.9	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
841.6	Eu-151	Eu-152m$_1$	9.32 h	1 57 740	14.5	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 121.8 (7.16), 344.3 (2.44), 963.4 (11.9)
842.8	Mo-100	Tc-101	14.2 m		0.23	127.2 (2.86), 184.1 (1.69), 306.8 (88), 531.4 (1.02), 545.1 (5.98), 627 (0.39), 715.6 (0.69), 720 (0.23), 842.8 (0.23), 928.6 (0.13)
843.2	Ge-76	Ge-77	11.3 h	0.468	0.20	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
843.7	Mg-26	Mg-27	9.46 m	0.42058	71.8	170.7 (0.84), 1014.4 (28.2)
845.4	Kr-86	Kr-87	76.3 m	0.0519	7.34	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
845.9	Ru-104	Ru-105	4.44 h	5.984	0.62	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
846.5	Xe-124	Xe-125	16.9 h	13.7	1.03	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 188.4 (54.9), 243.4 (28.8), 453.8 (4.23)
846.8	Mn-55	Mn-56	2.58 h	1330	98.9	1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
848.9	Xe-136	Xe-137	3.83 m	2.314	0.62	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
852	Ru-104	Ru-105	4.44 h	5.984	0.15	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
852.2	Te-130	Te-131m	30 h	0.676	21.0	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
853.2	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
854.9	Zr-96	Zr-97	17.0 h	0.0632	0.33	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
856.1	Ca-48	Ca-49	8.72 m	0.2057	0.13	1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
859.1	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
860.4	Cd-116	Cd-117m	3.36 h	0.1867	7.89	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
861.1	Ga-71	Ga-72	14.1 h	187.929	0.91	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
862.8	Pd-110	Pd-111m	5.5 h	0.43	0.14	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
866.6	Se-82	Se-83g	22.5 m	0.04888	8.85	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
867.4	Eu-151	Eu-152	13.33 yr	2 82 020	4.21	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
867.6	As-75	As-76	26.32 h	450	0.13	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
867.8	La-139	La-140	40.28 h	892.12	5.59	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
868.1	Sr-84	Sr-85g	64.84 d	0.196	0.12	13.4 (50.2), 15.0 (8.80), 151.2 (0.001), 514 (99.3)
869.7	Mo-100	Mo-101	14.6 m	1.916	0.34	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
871.1	Nb-93	Nb-94m	6.26 m	15	0.48	16.6 (37.1), 18.6 (7.18), 41 (0.08), 702.6 (0.003)
871.1	Mo-100	Mo-101	14.6 m	1.916	1.80	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
871.1	Nb-93	Nb-94	2.03×10^4 yr	100	100.0	702.6 (100)
872	Zn-68	Zn-69	56 m	18.8	0.0002	318.6 (0.0012)
873.2	Eu-153	Eu-154	8.8 yr		11.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
874.8	Os-184	Os-185	93.6 d	60	6.59	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 646.1 (80.8), 717.4 (4.11), 880.5 (4.98)
875.2	Ge-76	Ge-77	11.3 h	0.468	0.74	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
875.9	Ru-104	Ru-105	4.44 h	5.984	2.47	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822 (845.9), 875.9 (2.47), 969.4 (2.08)
877.4	Mo-100	Mo-101	14.6 m	1.916	3.40	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
878.2	Ru-104	Ru-105	4.44 h	5.984	0.47	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
879.4	Tb-159	Tb-160	72.3 d	2550	30.0	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
880.5	Os-184	Os-185	93.6 d	60	4.98	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 646.1 (80.8), 717.4 (4.11), 874.8 (6.59)
880.7	Cd-116	Cd-117	2.49 h	0.3735	3.96	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
882.1	Pd-110	Pd-111m	5.5 h	0.43	0.20	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
883.5	Mo-100	Mo-101	14.6 m	1.916	0.70	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
883.6	Se-82	Se-83g	22.5 m	0.04888	7.75	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
884.7	Ag-109	Ag-110m	249.9 d	226.4	72.9	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
887	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
887.8	Se-82	Se-83g	22.5 m	0.04888	4.73	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
888.7	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
889.2	Sc-45	Sc-46	83.82 d	1740	100	4.1 (6.6), 4.5 (0.88), 142.5 (56), 1120.5 (100)
890.1	Th-232	Th-233	22.3 m	740	0.14	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
894.2	Ga-71	Ga-72	14.1 h	187.929	9.88	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
896.3	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
896.5	Ge-76	Ge-77	11.3 h	0.468	0.12	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
898	Rb-87	Rb-88	17.8 m	3.339	14.5	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
899	K-41	K-42	12.36 h	9.8258	0.053	312.7 (0.35), 1524.6 (18.6)
901	Ge-76	Ge-77	11.3 h	0.468	0.11	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
903.6	Mo-100	Mo-101	14.6 m	1.916	0.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
907	Ge-76	Ge-77	11.3 h	0.468	0.90	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
907.3	S-36	S-37	5.0 m	0.003	0.058	3103.3 (94.1), 3740.4 (0.23)
907.6	Ru-104	Ru-105	4.44 h	5.984	0.52	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
909.2	Sr-88	Sr-89	50.55 d	0.47908	0.0095	–
909.2	Sr-88	Y-89m	16.1 s		99.1	14.9 (0.45), 16.8 (0.083)
910	Te-130	Te-131m	30 h	0.676	3.28	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
910.1	Zn-70	Zn-71m	3.94 h	0.005	0.31	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
910.3	Zn-70	Zn-71g	2.4 m	0.0514	7.8	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
913.8	Ge-76	Ge-77	11.3 h	0.468	0.35	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
915.5	Sn-124	Sn-125	9.64 d	0.0224	3.85	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
919.6	La-139	La-140	40.28 h	892.12	2.68	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
923.1	Ge-76	Ge-77	11.3 h	0.468	0.65	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
924.2	Ga-71	Ga-72	14.1 h	187.929	0.14	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
925.2	La-139	La-140	40.28 h	892.12	2.68	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 1596.5 (95.4), 2521.7 (3.43)
925.5	Ge-76	Ge-77	11.3 h	0.468	0.74	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
928.6	Ti-50	Ti-51	5.8 m	0.9308	6.88	320.1 (93), 608.5 (1.18)
928.6	Mo-100	Tc-101	14.2 m		0.13	127.2 (2.86), 184.1 (1.69), 306.8 (88), 531.4 (1.02), 545.1 (5.98), 627 (0.39), 715.6 (0.69), 720 (0.23), 842.8 (0.23), 928.6 (0.13)
928.9	Ge-76	Ge-77	11.3 h	0.468	0.99	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
931.3	Re-187	Re-188	16.98 h	4569.8	0.56	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 1610.4 (0.10)
931.4	Cd-116	Cd-117m	3.36 h	0.1867	3.64	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
933.3	Mo-100	Mo-101	14.6 m	1.916	0.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
933.7	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
933.8	Cd-114	Cd-115m	44.6 d	1.03	2.00	484.5 (0.29), 1290.6 (0.89)
934.2	Mo-100	Mo-101	14.6 m	1.916	3.40	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
934.8	Kr-78	Kr-79g	35.04 h	2.1	0.13	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
937.5	Ag-109	Ag-110m	249.9 d	226.4	34.3	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
938.7	Ir-193	Ir-194	19.15 h	6897	0.59	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
939.4	Ga-71	Ga-72	14.1 h	187.929	0.26	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
939.4	Ge-76	Ge-77	11.3 h	0.468	0.27	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
943.3	Se-82	Se-83g	22.5 m	0.04888	0.89	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
943.5	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
944.7	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
946.7	Kr-86	Kr-87	76.3 m	0.0519	0.13	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
948.5	Te-130	Te-131	25.0 m	9.126	2.26	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 997.2 (3.34), 1147 (4.96)
949.8	Mo-92	Mo-93m	6.85 h	0.089	0.12	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 684.7 (99.7), 1363.0 (0.79), 1477.1 (99.1)
950.9	Ho-165	Ho-166m	1.2×10^3 yr	350	3.07	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
952.1	Br-81	Br-82g	35.30 h	12.8206	0.37	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
956.7	Zn-70	Zn-71m	3.94 h	0.005	0.19	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
961	Mo-98	Mo-99	2.75 d	3.1369	0.10	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
962.3	Tb-159	Tb-160	72.3 d	2550	10.0	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
963.4	Eu-151	Eu-152m$_1$	9.32 h	1 57 740	11.9	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 121.8 (7.16), 344.3 (2.44), 841.6 (14.5)
964.1	Eu-151	Eu-152	13.33 yr	2 82 020	14.6	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
964.7	Zn-70	Zn-71m	3.94 h	0.005	4.7	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
964.8	Zn-70	Zn-71g	2.4 m	0.0514	0.77	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
966.2	Tb-159	Tb-160	72.3 d	2550	25.5	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
968.3	Sb-123	Sb-124	60.2 d	175.07	1.84	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
969.4	Ru-104	Ru-105	4.44 h	5.984	2.08	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822 (845.9), 875.9 (2.47), 969.4 (2.08)
970.5	Ga-71	Ga-72	14.1 h	187.929	1.10	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
971.4	Zr-96	Zr-97	17.0 h	0.0632	1.10	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
974.7	Zn-70	Zn-71m	3.94 h	0.005	0.35	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
975.2	Pd-110	Pd-111m	5.5 h	0.43	0.15	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
980.6	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
982.2	Xe-136	Xe-137	3.83 m	2.314	0.21	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
987.9	Se-82	Se-83g	22.5 m	0.04888	0.62	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
987.9	Se-82	Se-83m	70.4 s	0.3666	15.3	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
988.2	Mo-100	Mo-101	14.6 m	1.916	0.19	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
988.6	Zn-70	Zn-71m	3.94 h	0.005	1.2	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
995.9	Se-82	Se-83g	22.5 m	0.04888	1.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
996.3	Eu-153	Eu-154	8.8 yr		10.3	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
996.3	Pd-110	Pd-111m	5.5 h	0.43	0.25	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
996.5	Ge-76	Ge-77	11.3 h	0.468	0.10	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
997.2	Te-130	Te-131	25.0 m	9.126	3.34	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 1147 (4.96)
997.2	Ag-109	Ag-110m	249.9 d	226.4	0.13	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
997.6	Se-82	Se-83m	70.4 s	0.3666	1.28	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
999.9	Ga-71	Ga-72	14.1 h	187.929	0.80	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1001.7	Ta-181	Ta-182	115.0 d	2099.74	2.09	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1004.8	Eu-153	Eu-154	8.8 yr		17.9	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
1006.5	Zn-70	Zn-71m	3.94 h	0.005	0.74	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1007.4	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1007.6	Br-81	Br-82g	35.30 h	12.8206	1.31	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1011	Mo-100	Mo-101	14.6 m	1.916	2.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1011.4	Zn-70	Zn-71m	3.94 h	0.005	0.68	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1011.8	Ir-193	Ir-194m	171 d	376.2	3.6	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
1012.5	Mo-100	Mo-101	14.6 m	1.916	12.8	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1014.4	Mg-26	Mg-27	9.46 m	0.42058	28.2	170.7 (0.84), 843.7 (71.8)
1016.4	Nd-150	Nd-151	12.4 m	6.768	2.92	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
1017.5	Ru-104	Ru-105	4.44 h	5.984	0.32	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
1018.6	Mo-100	Mo-101	14.6 m	1.916	0.64	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1020	Mo-100	Mo-101	14.6 m	1.916	0.47	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1020.6	Se-82	Se-83m	70.4 s	0.3666	1.97	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660 (1.78), 2051.4 (11)
1021	Sn-122	Sn-123	129.2 d	0.0046	0.002	160.3 (0.002), 1030.2 (0.031), 1088.6 (0.6)
1021.3	Zr-96	Zr-97	17.0 h	0.0632	1.34	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1021.5	Tl-205	Tl-206m	3.75 m		62.9	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1139.9 (5.2)
1022.5	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1024.5	Zr-96	Zr-97	17.0 h	0.0632	1.16	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1024.5	Zr-96	Nb-97	72.1 m		1.08	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
1025.7	Kr-78	Kr-79g	35.04 h	2.1	0.16	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
1029.1	Cd-116	Cd-117m	3.36 h	0.1867	11.7	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1030.2	Sn-122	Sn-123	129.2 d	0.0046	0.031	160.3 (0.002), 1021.0 (0.002), 1088.6 (0.6)
1030.5	Se-82	Se-83m	70.4 s	0.3666	20.9	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1036.5	Se-82	Se-83g	22.5 m	0.04888	0.2	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1038.6	Cs-133	Cs-134	2.06 yr	2650	1.00	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1167.9 (1.8), 1365.1 (3.04)
1039.2	Ga-69	Ga-70	21.15 m	100.968	0.673	176.2 (0.297)
1039.2	Cu-65	Cu-66	5.1 m	66.9011	8	833.4 (0.18), 1333.2 (0.003)
1042.1	Se-82	Se-83g	22.5 m	0.04888	1.03	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1044	Br-81	Br-82g	35.30 h	12.8206	27.4	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1045.2	Sb-123	Sb-124	60.2 d	175.07	1.85	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
1049.8	Mo-100	Mo-101	14.6 m	1.916	0.35	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1050.7	Ga-71	Ga-72	14.1 h	187.929	6.91	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1051.7	Cd-116	Cd-117	2.49 h	0.3735	3.79	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
1053.6	Se-82	Se-83m	70.4 s	0.3666	1.49	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1061.4	Lu-175	Lu-176m	3.65 h	1597.36	0.0007	54.6 (2.87), 55.8 (5.02), 63.2 (1.68), 65.0 (0.43), 82.1 (0.007), 88.3 (8.86), 1159.3 (0.0014)
1061.7	Ge-76	Ge-77	11.3 h	0.468	0.14	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1063.4	Se-82	Se-83m	70.4 s	0.3666	3.39	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1063.4	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1064.1	Se-82	Se-83g	22.5 m	0.04888	5.90	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1064.2	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1065.9	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1066	Cd-116	Cd-117m	3.36 h	0.1867	23.1	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1067.1	Sn-124	Sn-125	9.64 d	0.0224	9.05	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
1076.6	Rb-85	Rb-86	18.66 d	30.81659	8.78	–
1080.1	Yb-176	Yb-177	1.9 h		5.5	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 121.6 (3.4), 150.4 (20), 1241.4 (3.4)
1080.8	Ge-76	Ge-77	11.3 h	0.468	0.23	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1081.4	Br-81	Br-82g	35.30 h	12.8206	0.63	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1082	Se-82	Se-83g	22.5 m	0.04888	2.68	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1083.9	Te-128	Te-129	69.6 m	6.34	0.46	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1111.6 (0.18)
1085.2	Ge-76	Ge-77	11.3 h	0.468	5.72	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1085.9	Eu-151	Eu-152	13.33 yr	2 82 020	9.92	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
1087.7	Sn-124	Sn-125	9.64 d	0.0224	1.11	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1089.2 (4.28), 2001.8 (1.79)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1087.7	Au-197	Au-198	2.70 d		0.23	68.9 (0.81), 70.8 (1.38), 80.2 (0.47), 82.5 (0.13), 411.8 (95.5), 675.9 (1.06)
1088	Pd-110	Pd-111m	5.5 h	0.43	0.19	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1088.6	Sn-122	Sn-123	129.2 d	0.0046	0.6	160.3 (0.002), 1021.0 (0.002), 1030.2 (0.031)
1089.2	Sn-124	Sn-125	9.64 d	0.0224	4.28	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 2001.8 (1.79)
1097.3	In-115	In-116m$_1$	54.15 m	7753.32	56.2	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
1098.5	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1099.3	Fe-58	Fe-59	44.503 d	0.3968	56.5	142.5 (1.03), 192.2 (3.11), 334.8 (0.26), 1291.6 (43.2)
1101	Sb-123	Sb-124m$_1$	60.2 d	1.58	0.3	498.4 (20), 602.7 (20), 645.8 (20)
1101	Sb-123	Sb-124m$_2$	20.2 m	0.811	0.3	498.4 (22), 602.7 (22), 645.8 (22)
1102.1	Te-120	Te-121m	154.0 d	0.032	2.54	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
1107.4	Zn-70	Zn-71m	3.94 h	0.005	2.8	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1109.3	Zn-70	Zn-71g	2.4 m	0.0514	0.16	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1110.3	Se-82	Se-83g	22.5 m	0.04888	0.41	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1110.5	Zr-96	Zr-97	17.0 h	0.0632	0.11	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1111.6	Te-128	Te-129	69.6 m	6.34	0.18	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46)
1112.1	Eu-151	Eu-152	13.33 yr	2 82 020	13.6	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
1113.5	Er-162	Er-163	75.0 m	2.66	0.049	46.7 (22.4), 47.5 (39.8), 53.8 (12.5), 55.3 (3.52), 436.1 (0.028), 439.9 (0.027)
1115.1	Kr-78	Kr-79g	35.04 h	2.1	0.37	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
1115.5	Ni-64	Ni-65	2.52 h	1.3832	14.8	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
1115.5	Zn-64	Zn-65	244.3 d	36.936	50.7	344.9 (0.003), 770.6 (0.003)
1115.9	Pd-110	Pd-111m	5.5 h	0.43	1.0	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1116	Se-82	Se-83m	70.4 s	0.3666	0.54	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1119.3	Xe-136	Xe-137	3.83 m	2.314	0.11	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1120	Zn-70	Zn-71g	2.4 m	0.0514	2.2	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1120.5	Sc-45	Sc-46	83.82 d	1740	100	4.1 (6.6), 4.5 (0.88), 142.5 (56), 889.2 (100)
1121.3	Ta-181	Ta-182	115.0 d	2099.74	35.0	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1122.1	Nd-150	Nd-151	12.4 m	6.768	4.59	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
1125	Ge-76	Ge-77	11.3 h	0.468	0.11	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1125.5	Te-130	Te-131m	30 h	0.676	11.4	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
1125.7	Ag-109	Ag-110	24.6 s	4157.07	0.015	21.1 (0.17), 23.9 (0.037), 657.7 (4.49), 815.3 (0.038)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1129.9	As-75	As-76	26.32 h	450	0.14	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1139.8	Zn-70	Zn-71m	3.94 h	0.005	0.2	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1139.9	Tl-205	Tl-206m	3.75 m		5.2	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9)
1141.1	Sb-121	Sb-122	2.70 d	335.2	0.57	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 564 (70.8), 692.6 (3.68), 1256.8 (0.77)
1144.5	Ca-48	Ca-49	8.72 m	0.2057	0.11	856.1 (0.13), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
1147	Te-130	Te-131	25.0 m	9.126	4.96	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 997.2 (3.34)
1147.9	Zr-96	Zr-97	17.0 h	0.0632	2.64	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1148.9	Te-130	Te-131m	30 h	0.676	1.9	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
1150.8	Ir-193	Ir-194	19.15 h	6897	0.59	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
1151.8	Ge-76	Ge-77	11.3 h	0.468	0.18	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1159.3	Lu-175	Lu-176m	3.65 h	1597.36	0.0014	54.6 (2.87), 55.8 (5.02), 63.2 (1.68), 65.0 (0.43), 82.1 (0.007), 88.3 (8.86), 1061.4 (0.0007)
1161	Mo-100	Mo-101	14.6 m	1.916	3.97	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1163.3	Pd-110	Pd-111m	5.5 h	0.43	0.32	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1167.9	Cs-133	Cs-134	2.06 yr	2650	1.80	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1365.1 (3.04)
1169	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1173.2	Co-59	Co-60	5.272 yr	1680	99.9	346.9 (0.0076), 1332.5 (100)
1175.4	Kr-86	Kr-87	76.3 m	0.0519	1.11	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1177.9	Tb-159	Tb-160	72.3 d	2550	15.5	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
1180.6	Nd-150	Nd-151	12.4 m	6.768	15.3	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
1183.5	Ir-193	Ir-194	19.15 h	6897	0.30	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
1184.2	Mo-100	Mo-101	14.6 m	1.916	0.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1186.6	Mo-100	Mo-101	14.6 m	1.916	1.03	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1189	Ta-181	Ta-182	115.0 d	2099.74	16.4	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1191.7	Se-82	Se-83g	22.5 m	0.04888	4.18	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1193.3	Ge-76	Ge-77	11.3 h	0.468	2.43	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1199.9	Tb-159	Tb-160	72.3 d	2550	2.36	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1200	Mo-100	Mo-101	14.6 m	1.916	1.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1200.1	Pd-110	Pd-111m	5.5 h	0.43	0.30	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1206.6	Te-130	Te-131m	30 h	0.676	9.73	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
1206.9	Se-82	Se-83g	22.5 m	0.04888	0.89	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1209.9	Mo-100	Mo-101	14.6 m	1.916	0.13	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1212.7	As-75	As-76	26.32 h	450	1.63	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1215.2	Ga-71	Ga-72	14.1 h	187.929	0.79	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1215.4	Ge-76	Ge-77	11.3 h	0.468	0.12	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1216	As-75	As-76	26.32 h	450	3.84	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1218.7	Hg-204	Hg-205	5.20 m	2.924	0.006	203.7 (2.2), 415.6 (0.13)
1221.4	Ta-181	Ta-182	115.0 d	2099.74	27.4	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1225.9	Se-82	Se-83g	22.5 m	0.04888	1.2	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1228.5	As-75	As-76	26.32 h	450	1.39	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1230.9	Ga-71	Ga-72	14.1 h	187.929	1.45	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1231	Ta-181	Ta-182	115.0 d	2099.74	11.6	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1234.6	Cd-116	Cd-117m	3.36 h	0.1867	11.0	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1237	Rh-103	Rh-104	42.3 s	13 500	0.066	19.2 (0.24), 21.7 (0.05), 358.1 (0.016), 555.8 (1.99)
1241.4	Yb-176	Yb-177	1.9 h		3.4	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 121.6 (3.4), 150.4 (20), 1080.1 (5.5)
1242.2	Ge-76	Ge-77	11.3 h	0.468	0.38	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1245.2	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1249.4	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1251.1	Mo-100	Mo-101	14.6 m	1.916	4.61	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1252.6	In-115	In-116	14.10 s	3828.8	0.031	463.1 (0.25), 1293.5 (1.30)
1254.7	Ba-138	Ba-139	1.38 h	25.812	0.033	33 (1.38), 33.4 (2.54), 37.8 (0.76), 38.7 (0.19), 165.8 (23.8), 1420.5 (0.261)
1256.8	Sb-121	Sb-122	2.70 d	335.2	0.77	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 564 (70.8), 692.6 (3.68), 1141.1 (0.57)
1259.3	Se-82	Se-83g	22.5 m	0.04888	1.85	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1260.1	Ga-71	Ga-72	14.1 h	187.929	1.13	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1260.2	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1263.9	Ge-76	Ge-77	11.3 h	0.468	0.80	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1266.1	Si-30	Si-31	2.62 h	0.3317	0.07	–
1268.6	Zr-96	Zr-97	17.0 h	0.0632	0.17	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1268.6	Zr-96	Nb-97	72.1 m		0.16	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
1271.9	Tb-159	Tb-160	72.3 d	2550	7.60	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
1273.2	Xe-136	Xe-137	3.83 m	2.314	0.23	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1274.5	Eu-153	Eu-154	8.8 yr		35.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
1276.1	Zr-96	Zr-97	17.0 h	0.0632	0.97	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1276.8	Ga-71	Ga-72	14.1 h	187.929	1.56	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1280	Ge-76	Ge-77	11.3 h	0.468	0.16	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1282.5	Pd-110	Pd-111m	5.5 h	0.43	1.0	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1282.7	Zn-70	Zn-71m	3.94 h	0.005	0.27	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1286.3	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1290.6	Cd-114	Cd-115m	44.6 d	1.03	0.890	484.5 (0.29), 933.8 (2)
1290.7	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1291.6	Fe-58	Fe-59	44.503 d	0.3968	43.2	142.5 (1.03), 192.2 (3.11), 334.8 (0.26), 1099.3 (56.5)
1293.3	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1293.5	In-115	In-116	14.10 s	3828.8	1.30	463.1 (0.25), 1252.6 (0.031)
1293.5	In-115	In-116m$_1$	54.15 m	7753.32	84.4	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1507.4 (9.96), 1753.8 (2.46)
1293.7	Ar-40	Ar-41	1.83 h	65.736	99.1	1677.2 (0.052)
1293.8	Se-82	Se-83g	22.5 m	0.04888	1.6	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1297.1	Ca-46	Ca-47	4.54 d	0.00296	74.9	489.2 (6.74), 530.4 (0.1), 767 (0.19), 807.9 (6.89)
1299.1	Se-82	Se-83g	22.5 m	0.04888	5.83	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1300	In-113	In-114	71.9 s	16.69	0.14	23.1 (2), 26.2 (0.42), 558 (0.07), 575.7 (0.004)
1303	Se-82	Se-83m	70.4 s	0.3666	0.93	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11)
1303.3	Cd-116	Cd-117	2.49 h	0.3735	18.4	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1576.6 (11.2), 1723.1 (2.01)
1304	Mo-100	Mo-101	14.6 m	1.916	2.78	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1305.9	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1306.7	Zn-70	Zn-71m	3.94 h	0.005	0.11	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1309.3	Ge-76	Ge-77	11.3 h	0.468	0.46	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1311.4	Zn-70	Zn-71m	3.94 h	0.005	0.10	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1312.1	Tb-159	Tb-160	72.3 d	2550	2.97	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
1312.8	Ge-76	Ge-77	11.3 h	0.468	0.34	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1314.3	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1316.9	Se-82	Se-83g	22.5 m	0.04888	4.12	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1317.5	Br-81	Br-82g	35.30 h	12.8206	26.9	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1319.7	Ge-76	Ge-77	11.3 h	0.468	0.28	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1321.3	Ru-104	Ru-105	4.44 h	5.984	0.20	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
1322.2	Zn-70	Zn-71m	3.94 h	0.005	0.23	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1325.7	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1332.1	Kr-78	Kr-79g	35.04 h	2.1	0.44	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
1332.5	Co-59	Co-60m	10.47 m	2040	0.25	6.9 (27.2), 7.7 (3.67), 58.6 (2.02), 826.2 (0.0083)
1332.5	Co-59	Co-60	5.272 yr	1680	100	346.9 (0.0076), 1173.2 (99.9)
1333.2	Cu-65	Cu-66	5.1 m	66.9011	0.003	833.4 (0.18), 1039.2 (8)
1333.6	V-51	V-52	3.75 m	488.775	0.59	1434.1 (100), 1530.7 (0.12)
1334.3	Ag-109	Ag-110m	249.9 d	226.4	0.13	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1336.3	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1338	Kr-86	Kr-87	76.3 m	0.0519	0.63	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1339.3	Cd-116	Cd-117m	3.36 h	0.1867	2.1	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1339.4	Mo-100	Mo-101	14.6 m	1.916	0.12	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1341.2	Se-82	Se-83g	22.5 m	0.04888	5.70	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1345.8	Cu-63	Cu-64	12.7 h	311.265	0.48	511 (35.8)
1346.1	Mo-100	Mo-101	14.6 m	1.916	1.03	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1352.5	Se-82	Se-83g	22.5 m	0.04888	4.80	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1355.9	Mo-100	Mo-101	14.6 m	1.916	1.67	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1356.8	O-18	O-19	27.1 s	0.000032	50.4	109.9 (2.54), 197.1 (95.9), 1444.1 (2.64), 1554 (1.39)
1362.7	Zr-96	Zr-97	17.0 h	0.0632	1.35	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1363	Mo-92	Mo-93m	6.85 h	0.089	0.79	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 684.7 (99.7), 949.8 (0.12), 1477.1 (99.1)
1365.1	Cs-133	Cs-134	2.06 yr	2650	3.04	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8)
1366.3	Rb-87	Rb-88	17.8 m	3.339	0.11	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1368.2	Sb-123	Sb-124	60.2 d	175.07	2.36	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1691.0 (49.0), 2091.0 (5.61)
1368.3	Ge-76	Ge-77	11.3 h	0.468	3.18	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1368.5	Na-23	Na-24	14.96 h	53	100	2753.9 (99.9), 3867.3 (0.052)
1377.7	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1379.4	Ho-165	Ho-166	26.80 h	6300	0.93	48.2 (2.86), 49.1 (5.07), 55.6 (1.62), 57.2 (0.45), 80.6 (6.20), 1581.9 (0.18), 1662.4 (0.12)
1380.4	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1380.8	Zn-70	Zn-71m	3.94 h	0.005	0.36	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1382.4	Rb-87	Rb-88	17.8 m	3.339	0.77	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1382.6	Kr-86	Kr-87	76.3 m	0.0519	0.29	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1382.7	Mo-100	Mo-101	14.6 m	1.916	1.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1384.3	Ag-109	Ag-110m	249.9 d	226.4	24.3	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1388.5	Pd-110	Pd-111m	5.5 h	0.43	0.41	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1388.5	Pd-110	Pd-111	23.4 m	2.22	0.54	22.1 (17.0), 25 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
1389.9	Kr-86	Kr-87	76.3 m	0.0519	0.19	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1394.8	Mo-100	Mo-101	14.6 m	1.916	0.61	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1404	Sn-124	Sn-125m	9.52 m	0.728	0.7	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 643 (0.16), 1483.9 (0.18), 1615.3 (0.12)
1408	Eu-151	Eu-152	13.33 yr	2 82 020	20.8	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
1408.9	Ca-48	Ca-49	8.72 m	0.2057	0.63	856.1 (0.13), 1144.5 (0.11), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
1414.2	Mo-100	Mo-101	14.6 m	1.916	0.50	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1418.6	Mo-100	Mo-101	14.6 m	1.916	0.88	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1420.5	Ba-138	Ba-139	1.38 h	25.812	0.261	33 (1.38), 33.4 (2.54), 37.8 (0.76), 38.7 (0.19), 165.8 (23.8), 1254.7 (0.033)
1420.6	Se-82	Se-83g	22.5 m	0.04888	4.10	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1430	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1431.6	Mo-100	Mo-101	14.6 m	1.916	0.36	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1432.9	Cd-116	Cd-117m	3.36 h	0.1867	13.4	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1434.1	V-51	V-52	3.75 m	488.775	100	1333.6 (0.59), 1530.7 (0.12)
1435.8	Se-82	Se-83g	22.5 m	0.04888	0.82	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1439.1	As-75	As-76	26.32 h	450	0.33	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1440.8	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1444.1	O-18	O-19	27.1 s	0.000032	2.64	109.9 (2.54), 197.1 (95.9), 1356.8 (50.4), 1554 (1.39)
1447.4	Se-82	Se-83g	22.5 m	0.04888	0.48	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1452.6	Ge-76	Ge-77	11.3 h	0.468	0.11	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1453.6	As-75	As-76	26.32 h	450	0.13	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1459	Pd-110	Pd-111	23.4 m	2.22	0.56	22.1 (17.0), 25 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
1459	Pd-110	Pd-111m	5.5 h	0.43	0.42	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1460.8	K-39	K-40	1.28×10^9 yr	195.846	10.7	3 (0.95)
1464	Ga-71	Ga-72	14.1 h	187.929	3.55	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1468.9	Ir-193	Ir-194	19.15 h	6897	0.19	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
1474.8	Br-81	Br-82m	6.13 m	119.8233	0.016	46.0 (0.24), 698.3 (0.024), 776.5 (0.2)
1474.8	Br-81	Br-82g	35.30 h	12.8206	16.6	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1475.2	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1475.8	Ag-109	Ag-110m	249.9 d	226.4	3.99	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1476	Zn-70	Zn-71m	3.94 h	0.005	0.6	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1476.5	Ge-76	Ge-77	11.3 h	0.468	0.23	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1477.1	Mo-92	Mo-93m	6.85 h	0.089	99.1	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 684.7 (99.7), 949.8 (0.12), 1363.0 (0.79)
1479	Ge-76	Ge-77	11.3 h	0.468	0.20	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1481.8	Ni-64	Ni-65	2.52 h	1.3832	23.5	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1623.4 (0.47), 1724.9 (0.39)
1483.9	Sn-124	Sn-125m	9.52 m	0.728	0.18	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 643 (0.16), 1404 (0.7), 1615.3 (0.12)
1485.9	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1495.6	Ge-76	Ge-77	11.3 h	0.468	0.47	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1505	Ag-109	Ag-110m	249.9 d	226.4	13.1	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1507.4	In-115	In-116m$_1$	54.15 m	7753.32	9.96	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1753.8 (2.46)
1514.1	Mo-100	Mo-101	14.6 m	1.916	0.19	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1515.6	Zr-96	Zr-97	17.0 h	0.0632	0.13	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1515.6	Zr-96	Nb-97	72.1 m		0.12	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
1517.8	Mo-100	Mo-101	14.6 m	1.916	0.22	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1520.4	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1523	Mo-100	Mo-101	14.6 m	1.916	0.29	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1524.6	K-41	K-42	12.36 h	9.8258	18.3	312.7 (0.35), 899 (0.053)
1526.6	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1528	Cr-54	Cr-55	3.5 m	0.8496	0.037	126.1 (0.0017), 2252.5 (0.0031)
1530.3	Mo-100	Mo-101	14.6 m	1.916	0.27	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1530.7	V-51	V-52	3.75 m	488.775	0.12	1333.6 (0.59), 1434.1 (100)
1531.2	Kr-86	Kr-87	76.3 m	0.0519	0.36	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1532.5	Mo-100	Mo-101	14.6 m	1.916	5.96	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1538.8	Ge-76	Ge-77	11.3 h	0.468	0.13	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1548.7	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1554	O-18	O-19	27.1 s	0.000032	1.39	109.9 (2.54), 197.1 (95.9), 1356.8 (50.4), 1444.1 (2.64)
1554.8	Se-82	Se-83g	22.5 m	0.04888	2.54	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1558.5	Se-82	Se-83m	70.4 s	0.3666	1.20	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1562.3	Ag-109	Ag-110m	249.9 d	226.4	1.18	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1568.1	Ga-71	Ga-72	14.1 h	187.929	0.20	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1571.6	Ga-71	Ga-72	14.1 h	187.929	0.82	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1573.7	Ge-76	Ge-77	11.3 h	0.468	0.62	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1575.7	Pr-141	Pr-142	19.13 h	760	3.7	508.8 (0.023), 641.2 (0.002)
1576.6	Cd-116	Cd-117	2.49 h	0.3735	11.2	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1723.1 (2.01)
1576.8	Xe-136	Xe-137	3.83 m	2.314	0.10	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1578	Kr-86	Kr-87	76.3 m	0.0519	0.13	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1581.9	Ho-165	Ho-166	26.80 h	6300	0.18	48.2 (2.86), 49.1 (5.07), 55.6 (1.62), 57.2 (0.45), 80.6 (6.20), 1379.4 (0.93), 1662.4 (0.12)
1589.7	Mo-100	Mo-101	14.6 m	1.916	0.28	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1596.5	La-139	La-140	40.28 h	892.12	95.4	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 2521.7 (3.43)
1596.5	Eu-153	Eu-154	8.8 yr		1.83	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
1596.7	Ga-71	Ga-72	14.1 h	187.929	4.24	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1599.3	Mo-100	Mo-101	14.6 m	1.916	1.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1609.2	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1610.4	Re-187	Re-188	16.98 h	4569.8	0.10	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56)
1611.2	Kr-86	Kr-87	76.3 m	0.0519	0.10	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1612.5	Xe-136	Xe-137	3.83 m	2.314	0.12	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1615.3	Sn-124	Sn-125m	9.52 m	0.728	0.12	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 643 (0.16), 1404 (0.7), 1483.9 (0.18)
1622.6	Ca-48	Sc-49	57.4 m		0.01	1761.9 (0.05)
1623.4	Ni-64	Ni-65	2.52 h	1.3832	0.47	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1724.9 (0.39)
1631.6	Zn-70	Zn-71g	2.4 m	0.0514	0.38	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1633.7	F-19	F-20	11.0 s	0.96	100.0	3332.5 (0.009)
1636.5	Ne-22	Ne-23	37.2 s	0.42412	1.0	439.9 (33), 2076.4 (0.1)
1642.4	Cl-37	Cl-38	37.18 m	10.49159	31.6	2167.6 (42.4)
1650.3	Br-81	Br-82g	35.30 h	12.8206	0.79	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1651.3	Pd-110	Pd-111m	5.5 h	0.43	0.68	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1660	Se-82	Se-83m	70.4 s	0.3666	1.78	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660 (1.78), 2051.4 (11.0)
1662.4	Ho-165	Ho-166	26.80 h	6300	0.12	48.2 (2.86), 49.1 (5.07), 55.6 (1.62), 57.2 (0.45), 80.6 (6.20), 1379.4 (0.93), 1581.9 (0.18)
1662.4	Mo-100	Mo-101	14.6 m	1.916	0.68	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1664.6	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1673.9	Mo-100	Mo-101	14.6 m	1.916	1.69	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1677.2	Ar-40	Ar-41	1.83 h	65.736	0.052	1293.7 (99.1)
1680.8	Ga-71	Ga-72	14.1 h	187.929	0.90	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1684.2	Se-82	Se-83g	22.5 m	0.04888	0.2	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1691	Sb-123	Sb-124	60.2 d	175.07	49.0	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 2091.0 (5.61)
1691.1	Pd-110	Pd-111m	5.5 h	0.43	1.2	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1694.5	Se-82	Se-83m	70.4 s	0.3666	0.74	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1709.8	Ge-76	Ge-77	11.3 h	0.468	0.29	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1710.9	Ga-71	Ga-72	14.1 h	187.929	0.39	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1712.8	Mo-100	Mo-101	14.6 m	1.916	0.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1715.9	Se-82	Se-83g	22.5 m	0.04888	0.6	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1719.7	Ge-76	Ge-77	11.3 h	0.468	0.38	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1721.9	Pd-110	Pd-111m	5.5 h	0.43	0.32	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1723.1	Cd-116	Cd-117	2.49 h	0.3735	2.01	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2)
1724.9	Ni-64	Ni-65	2.52 h	1.3832	0.39	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47)
1727.2	Ge-76	Ge-77	11.3 h	0.468	0.14	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1740.5	Kr-86	Kr-87	76.3 m	0.0519	2.04	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1750.5	Zr-96	Zr-97	17.0 h	0.0632	1.34	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1753.8	In-115	In-116m$_1$	54.15 m	7753.32	2.46	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96)
1754.9	Mo-100	Mo-101	14.6 m	1.916	0.46	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1755	N-15	N-16	7.13 s	0.00000888	0.13	2741 (0.76), 6129 (68.8), 7115.2 (5)
1759.6	Zn-70	Zn-71m	3.94 h	0.005	0.93	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1759.8	Mo-100	Mo-101	14.6 m	1.916	0.98	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1761.9	Ca-48	Sc-49	57.4 m		0.05	1622.6 (0.01)
1768.2	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1775.2	Pd-110	Pd-111m	5.5 h	0.43	0.44	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1779	Se-82	Se-83m	70.4 s	0.3666	0.71	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1779	Al-27	Al-28	2.246 m	23.1	100.0	–
1779.6	Br-81	Br-82g	35.30 h	12.8206	0.12	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1779.8	Rb-87	Rb-88	17.8 m	3.339	0.22	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1779.9	Se-82	Se-83g	22.5 m	0.04888	1.9	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1783.4	Xe-136	Xe-137	3.83 m	2.314	0.41	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1787.7	As-75	As-76	26.32 h	450	0.33	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1810.7	Mn-55	Mn-56	2.58 h	1330	27.2	846.8 (98.9), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
1827.1	Se-82	Se-83g	22.5 m	0.04888	1.4	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1836	Rb-87	Rb-88	17.8 m	3.339	22.1	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1837.6	Ga-71	Ga-72	14.1 h	187.929	0.21	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1840.3	Mo-100	Mo-101	14.6 m	1.916	1.37	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1842.6	Kr-86	Kr-87	76.3 m	0.0519	0.14	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1846.4	Ge-76	Ge-77	11.3 h	0.468	0.16	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1847.6	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1851.6	Zr-96	Zr-97	17.0 h	0.0632	0.35	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1854.4	Se-82	Se-83g	22.5 m	0.04888	1.5	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1861.1	Ga-71	Ga-72	14.1 h	187.929	5.25	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1871.2	Se-82	Se-83g	22.5 m	0.04888	1.4	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1877.9	Ga-71	Ga-72	14.1 h	187.929	0.23	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1894.8	Se-82	Se-83g	22.5 m	0.04888	7.75	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1904.4	Zn-70	Zn-71g	2.4 m	0.0514	0.17	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1916.3	Xe-136	Xe-137	3.83 m	2.314	0.10	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1920.2	Ga-71	Ga-72	14.1 h	187.929	0.16	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1970.8	Pd-110	Pd-111m	5.5 h	0.43	0.59	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1973.2	Se-82	Se-83g	22.5 m	0.04888	0.62	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1991.3	Ga-71	Ga-72	14.1 h	187.929	0.11	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1997.3	Cd-116	Cd-117m	3.36 h	0.1867	26.2	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
2000.1	Ge-76	Ge-77	11.3 h	0.468	0.53	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2000.9	Te-130	Te-131m	30 h	0.676	2.01	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
2001.8	Sn-124	Sn-125	9.64 d	0.0224	1.79	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28)
2011.9	Kr-86	Kr-87	76.3 m	0.0519	2.88	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2028.1	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2029.4	Ga-71	Ga-72	14.1 h	187.929	0.12	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2032.1	Mo-100	Mo-101	14.6 m	1.916	6.94	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2038.4	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2041.2	Mo-100	Mo-101	14.6 m	1.916	2.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2045.2	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2051.4	Se-82	Se-83m	70.4 s	0.3666	11.0	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
2072.4	Se-82	Se-83g	22.5 m	0.04888	0.27	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2076.4	Ne-22	Ne-23	37.2 s	0.42412	0.1	439.9 (33), 1636.5 (1)
2077.2	Ge-76	Ge-77	11.3 h	0.468	0.22	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2085.4	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2088.8	Mo-100	Mo-101	14.6 m	1.916	0.79	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2089.6	Ge-76	Ge-77	11.3 h	0.468	0.33	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2091	Sb-123	Sb-124	60.2 d	175.07	5.61	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0)
2096.3	As-75	As-76	26.32 h	450	0.66	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
2096.4	Cd-116	Cd-117m	3.36 h	0.1867	7.44	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
2109.5	Ga-71	Ga-72	14.1 h	187.929	1.04	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2110.8	As-75	As-76	26.32 h	450	0.39	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
2111.2	Rb-87	Rb-88	17.8 m	3.339	0.12	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2112.8	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2113	Mn-55	Mn-56	2.58 h	1330	14.3	846.8 (98.9), 1810.7 (27.2), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
2114.3	Mo-100	Mo-101	14.6 m	1.916	0.45	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2118.9	Rb-87	Rb-88	17.8 m	3.339	0.44	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2126.1	Ge-76	Ge-77	11.3 h	0.468	0.19	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2167.3	Se-82	Se-83g	22.5 m	0.04888	0.34	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2167.6	Cl-37	Cl-38	37.18 m	10.49159	42.4	1642.4 (31.6)
2201.7	Ga-71	Ga-72	14.1 h	187.929	25.9	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2214.3	Ga-71	Ga-72	14.1 h	187.929	0.18	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
2223.2	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2228.9	Ca-48	Ca-49	8.72 m	0.2057	0.19	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
2252.5	Cr-54	Cr-55	3.5 m	0.8496	0.0031	126.1 (0.0017), 1528 (0.037)
2290.2	Se-82	Se-83g	22.5 m	0.04888	9.26	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2317.7	Zn-70	Zn-71m	3.94 h	0.005	0.65	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
2322.8	Cd-116	Cd-117m	3.36 h	0.1867	7.86	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
2337.4	Se-82	Se-83g	22.5 m	0.04888	3.43	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2341.6	Ge-76	Ge-77	11.3 h	0.468	0.45	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2371.7	Ca-48	Ca-49	8.72 m	0.2057	0.49	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
2408.5	Kr-86	Kr-87	76.3 m	0.0519	0.23	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2419.9	Se-82	Se-83g	22.5 m	0.04888	0.41	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2491	Ga-71	Ga-72	14.1 h	187.929	7.68	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2507.8	Ga-71	Ga-72	14.1 h	187.929	12.8	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
2515	Ga-71	Ga-72	14.1 h	187.929	0.25	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2521.7	La-139	La-140	40.28 h	892.12	3.43	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4)
2522.9	Mn-55	Mn-56	2.58 h	1330	0.99	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
2554.8	Kr-86	Kr-87	76.3 m	0.0519	9.23	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2558.1	Kr-86	Kr-87	76.3 m	0.0519	3.92	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2577.7	Rb-87	Rb-88	17.8 m	3.339	0.19	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2621.1	Ga-71	Ga-72	14.1 h	187.929	0.13	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2657.5	Mn-55	Mn-56	2.58 h	1330	0.65	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2959.8 (0.31), 3369.6 (0.17)
2677.9	Rb-87	Rb-88	17.8 m	3.339	2.02	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2734	Rb-87	Rb-88	17.8 m	3.339	0.11	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2741	N-15	N-16	7.13 s	0.00000888	0.76	1755 (0.13), 6129 (68.8), 7115.2 (5)
2753.9	Na-23	Na-24	14.96 h	53	99.9	1368.5 (100), 3867.3 (0.052)
2811.4	Kr-86	Kr-87	76.3 m	0.0519	0.32	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2844	Ga-71	Ga-72	14.1 h	187.929	0.43	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2849.8	Xe-136	Xe-137	3.83 m	2.314	0.18	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
2959.8	Mn-55	Mn-56	2.58 h	1330	0.31	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 3369.6 (0.17)
3009.4	Rb-87	Rb-88	17.8 m	3.339	0.25	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
3084.4	Ca-48	Ca-49	8.72 m	0.2057	92.1	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 4071.9 (7), 4738.2 (0.21)
3103.3	S-36	S-37	5.0 m	0.003	94.1	907.3 (0.058), 3740.4 (0.23)
3214	B-11	B-12	2.02×10^{-2} s	0.44	0.0006	4437.1 (1.28)
3218.5	Rb-87	Rb-88	17.8 m	3.339	0.22	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
3308.5	Kr-86	Kr-87	76.3 m	0.0519	0.45	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
3332.5	F-19	F-20	11.0 s	0.96	0.009	1633.7 (100)
3369.6	Mn-55	Mn-56	2.58 h	1330	0.17	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31)
3486.5	Rb-87	Rb-88	17.8 m	3.339	0.13	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
3740.4	S-36	S-37	5.0 m	0.003	0.23	907.3 (0.058), 3103.3 (94.1)
3867.3	Na-23	Na-24	14.96 h	53	0.052	1368.5 (100), 2753.9 (99.9)
4071.9	Ca-48	Ca-49	8.72 m	0.2057	7	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4738.2 (0.21)
4437.1	B-11	B-12	2.02×10^{-2} s	0.44	1.28	3214 (0.0006)
4738.2	Ca-48	Ca-49	8.72 m	0.2057	0.21	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7)
4742.7	Rb-87	Rb-88	17.8 m	3.339	0.15	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
6129	N-15	N-16	7.13 s	0.00000888	68.8	1755 (0.13), 2741 (0.76), 7115.2 (5)
7115.2	N-15	N-16	7.13 s	0.00000888	5.0	1755 (0.13), 2741 (0.76), 6129 (68.8)

Instrumental Neutron Activation Analysis

Zeev B. Alfassi
Ben Gurion University, Beer Sheva, Israel

1 Introduction 1

2 Basic Nuclear Physics 1
2.1 Radioactive Decay 1
2.2 Kinetics of Decay of Radioactive Nuclides 3
2.3 Kinetics of Chain Decays 4
2.4 Kinetics of Formation of Radioactive Nuclides by Irradiation 4

3 The Shape of a γ Spectrum 7

4 Neutron Sources 9
4.1 Sample Introduction 10

5 Instrumental Neutron Activation Analysis 10
5.1 Techniques 10
5.2 Calculation 11
5.3 Nuclear Interferences 13
5.4 A Test Case: Instrumental Neutron Activation Analysis of Trace Elements in Silicon 14
5.5 An Example of Interference Removal in Instrumental Neutron Activation Analysis: Determination of Platinum in Aerosols 15
5.6 Epithermal Instrumental Neutron Activation Analysis 16
5.7 Fast Neutron Instrumental Neutron Activation Analysis 20
5.8 Instrumental Neutron Activation Analysis with Emissions Other than γ Emission 22
5.9 Depth Profiling by Instrumental Neutron Activation Analysis 23
5.10 Practical Cases of Instrumental Neutron Activation Analysis 24

Abbreviations and Acronyms 27

Related Articles 28

References 28

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Analysis of the elemental content of a sample by irradiation with neutrons and measurement of the induced radioactivity are described.

1 INTRODUCTION

Chemical analysis by nuclear activation is an elemental analysis, i.e. it determines the contents of the various elements in the analyte sample, but cannot tell in what chemical form (compounds, valence states, etc.) they are present. The analysis is based on a reaction of the analyte element with nuclear projectiles (neutrons, accelerated small charged particles, e.g. protons, or γ-photons) [Equation 1]:

$$\text{target} + \text{projectile} \longrightarrow \text{light product} + \text{heavy product} \quad (1)$$

In instrumental neutron activation analysis (INAA) the projectile is a thermal neutron and the measurement of the concentration of the element is done via the heavy product, if it is a radionuclide which emits γ-rays. The amount produced of the radionuclide is proportional to the number of the target atoms.

2 BASIC NUCLEAR PHYSICS

This section gives only a brief description of basic knowledge. Readers who are interested in further detail should consult common textbooks.[1–6]

2.1 Radioactive Decay

Most radioactive decays occur via the transformations neutron → proton and proton → neutron and are called β-decays. They are characterized by the mass number, ΔA, which remains unchanged, whereas the atomic number is changed by ±1. There are three transformations of this type. A neutron is transformed into a proton by a β^- process [Equation 2]:

$$n \longrightarrow p^+ + \beta^- + \bar{\nu} \quad (2)$$

The proton remains in the nucleus, whereas the electron (β^-) and the antineutrino ($\bar{\nu}$) are emitted from the nucleus, with high kinetic energies. The electron is written as β^- and not e^-, to show that it is not an atomic electron. A proton can be transformed into a neutron by two processes. One process is β^+ decay, in which a proton is transformed into a neutron and a positron ($\beta^+ = e^+$), an entity similar in mass to an electron but which has an opposite charge [Equation 3]:

$$p^+ \longrightarrow n + \beta^+ + \nu \quad (3)$$

In this process there is also an emission of a neutrino (ν). Another process is the reaction of a proton with one of the surrounding electrons in the atom to yield a neutron. This process is called electron capture (EC). The neutrino and antineutrino are very small particles with no charge, and consequently they are very difficult to detect. They are not important in activation analysis, except for their effect on the β energies, which will be explained later.

Most radionuclides are not pure β emitters, but they emit simultaneously also a γ-photon (more accurately, the γ emission occurs in a very short time of $<10^{-10}$ s after the β^- emission). γ-Rays are electromagnetic waves, like light and radiowaves, but they have much higher energies (much shorter wavelengths). They have usually higher energies than X-rays, although the main difference between X-rays and γ-rays is their sources. X-rays are due to atomic transitions (transitions between different energy levels of electrons), whereas γ-rays are due to nuclear transitions (transitions between different energy levels of the nucleons). In most cases β^- decay does not yield the ground state of the product nuclide, but forms it in an excited state. The excited-state nuclide decays very rapidly (in most cases) to the ground state by the emission of either γ-rays or atomic electrons called conversion electrons. The emitted conversion electrons have no importance in activation analysis and we can neglect their emission and refer only to the emission of γ-rays.

A very important difference between β processes and γ decay (besides the difference in entities of the emitted particles, electrons vs photons, and the difference in ΔA, ± 1 in the β process and 0 in the γ decay) is the fact that in β processes there is emission of two particles, β^- and $\bar{\nu}$ or β^+ and ν, whereas in γ decay there is emission of only one photon. There are cases of the emission of a few photons from the same nuclide, but these are successive emissions and not a simultaneous emission as in the β processes. The importance of this difference lies in the fact that each nuclear decay is a transformation between two discrete energy states, resulting in a definite energy released in the process. If only one particle is emitted from the nucleus, this particle has a defined energy. When two particles are emitted, the released energy is distributed between them, and each particle can have a spectrum of energies ranging from zero up to a maximum energy, equal to the released energy. Since the emitted γ-photons have definite energies, they can be used in order to identify their emitters, by the measurement of the energy of the photons. As β-particles do not have definite energies, they usually cannot be used to identify their emitters.

In INAA, we need to identify the different radionuclides (in order to know the nuclide from which they were formed), in addition to the measurement of their activities. This is the reason why INAA employs almost exclusively the measurement of the spectrum of emitted γ-photons. In a few cases, where the indicator radionuclide (IRN) is a pure β^- emitter, it can be measured only in cases where there are only few (2–3) nonspecific γ-emitter IRNs in the sample with different half-lives. Their separate activities can be measured by measuring the activities at different times (measuring the decay curve) and extracting the various activities from the time dependence of the measured activity.

We said that the γ-photons emitted are due to the de-excitation of the nuclide produced in the β process. However, this de-excitation does not always have to be by only one photon. Moreover, not all β decays lead to the same level of excitation. These two facts together with the de-excitation by emission of conversion electrons (mainly from low-lying levels) lead to the possibilities that one radionuclide can have more than one kind (energy) of photon and that the number of photons should not be equal to the number of nuclides which have decayed (disintegrated). The number of photons of specific energy emitted per 100 disintegrated nuclides is called the intensity of the γ line, expressed as a percentage. Let us consider different cases of γ line intensity in four examples of radionuclides.

1. ^{28}Al decays completely to the 1.778 MeV excited state of ^{28}Si. This level decays to the ground state by only one photon. This means that ^{28}Al has only one γ line of 1.778 MeV with an intensity of 100%. This can be expressed by a diagram (called a decay scheme) of decreasing energies as shown in Scheme 1.

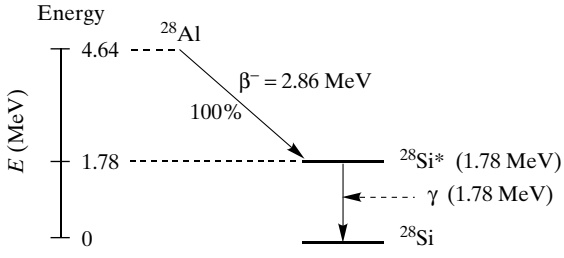

Scheme 1

2. ^{24}Na decays almost completely (>99.8%) to one level of ^{24}Mg (Scheme 2). However, this level decays to the ground state only by two successive photons. The first photon of 2.75 MeV is followed by a second photon of 1.39 MeV. This decay scheme explains why ^{24}Na has two γ lines of 2.75 and 1.39 MeV, each with almost 100% intensity.

3. ^{27}Mg decays to two different excited levels, resulting in two β^- with different energies, 1.59 MeV (31%) and 1.75 MeV (69%) (Scheme 3). The higher excited level (1.01 MeV) decays only 98% directly to the

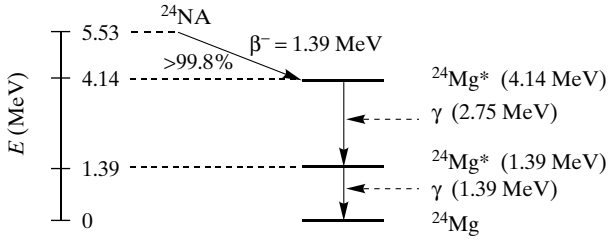

Scheme 2

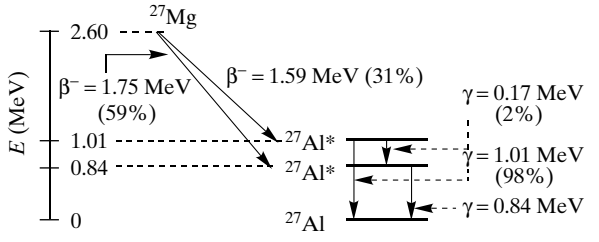

Scheme 3

ground state while 2% ($0.31 \times 0.02 = 0.006 = 0.6\%$ from the total nuclides disintegrated) decays to the lower excited state. The lower excited level decays completely to the ground state by one photon. This decay scheme explains why ^{27}Mg has three γ lines of 1.01 MeV (31%), 0.84 MeV (69%) and 0.18 MeV (0.6%).

4. ^{38}Cl decays directly to three different states of ^{38}Ar, both to two excited states of ^{38}Ar and to the ground state of ^{38}Ar (Scheme 4). This is the reason for ^{38}Cl having three different energy β^- particles (remember, it is only maximum energy, since β^- leads to a spectrum of β^- energies, up to E_{maximum}), 4.81 MeV (53%), 2.77 MeV (9%) and 1.77 MeV (38%). The higher excited state (3.77 MeV) decays almost completely to the lower excited state and only very little (0.06% of the 38%) decays directly to the ground state.

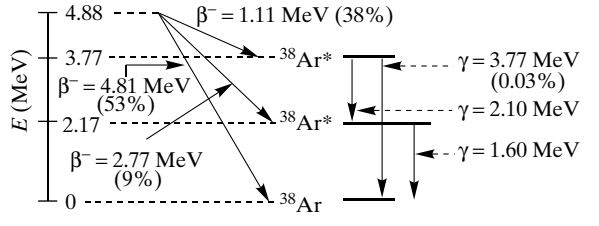

Scheme 4

Consequently ^{38}Cl has three γ lines, 3.77 MeV (0.03%), 2.10 MeV (38%) and 1.60 MeV (38% + 53% = 91%). These data together with more physical data on the nuclides can be found in books collecting all the decay schemes of the nuclides,[7] but for application of radionuclides it is sufficient to use the collection of all the γ lines of the radionuclides without the detailed decay scheme.[8–10] These collections are for all known radionuclides, but for most of cases in activation analysis it is better to use a smaller collection of radionuclides, only of those potentially formed by the specific kind of activation analysis.[11–13]

2.2 Kinetics of Decay of Radioactive Nuclides

The decay of radioactive nuclides is a statistical process. Thus, the number of atoms decaying per unit time (rate of decay) is proportional to the number of atoms of that specific radionuclide present in the sample. If the number of nuclei (atoms) of a specific radionuclide at time t is N^*, their rate of disappearance (decay or disintegration) is given by Equation (4):

$$-\frac{dN^*}{dt} = \lambda N^* \tag{4}$$

The constant λ, which is different for each radionuclide, is called the decay constant of this radionuclide. Integration of Equation (4) leads to the decay equation of radionuclides, i.e. the equation describing the time dependence of the number of atoms of the specific radionuclide [Equation 5]:

$$N^*(t) = N_0^* e^{-\lambda t} \tag{5}$$

where N_0^* is the number of atoms at time chosen as $t = 0$. Equation (5), which is the same as the integrated equation of any first-order chemical process, indicates the existence of a constant lifetime for the half of the atoms, which is called the half-life. It means that independently of the value of N_0^*, it takes the same time (for a specific radionuclide) for disintegration of half of the atoms, leaving only $N_0^*/2$ atoms of that radionuclide. Substituting $N^*(t)$ by $N_0^*/2$ in Equation (5) leads to the correlation between the half-life ($t_{1/2}$) and the decay constant (λ) [Equation 6]:

$$t_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda} \tag{6}$$

Equation (5), the decay equation, can be written also with $t_{1/2}$ instead of λ [Equation 7]:

$$N^*(t) = N_0^*(0.5)^{t/t_{1/2}} \tag{7}$$

The advantage of Equation (7) over Equation (5) is in that it gives a better 'feeling' for the extent of the decay. Thus, for example, if the half-life is $t_{1/2} = 2.5$ days, we know that after 2.5 days only half of the original atoms remain, and after 5 days the number of original atoms is reduced to one quarter. Equation (5) was easier to use when people were using tables of exponents. Nowadays

with computers there is no difference which equation is used. In tables of data, only $t_{1/2}$, the half-life, is given. In order to use Equation (5), λ should be calculated from Equation (6).

2.3 Kinetics of Chain Decays

In some cases a radionuclide decays not to a stable nuclide but to another radionuclide setting a chain of decays. Each chain is terminated in a stable nuclide. For a chain of decays as shown in Equation (8):

$$\mathrm{R}_1 \xrightarrow{\lambda_1} \mathrm{R}_2 \xrightarrow{\lambda_2} \mathrm{R}_3 \xrightarrow{\lambda_3} \cdots \longrightarrow \mathrm{R}_n \xrightarrow{\lambda_n} \mathrm{S} \qquad (8)$$

where R denotes a radionuclide and S stands for a stable isotope, the number of atoms R_k at time t assuming that at time zero only $R_1 \neq 0$ and is equal to R_1^0 is given by the Bateman equation [Equations 9 and 10]:

$$R_n = C_1 \mathrm{e}^{-\lambda_1 t} + C_2 \mathrm{e}^{-\lambda_2 t} + \cdots + C_n \mathrm{e}^{-\lambda_n t} \qquad (9)$$

$$R_n = \sum_{i=1}^{n} C_i \mathrm{e}^{-\lambda_i t} \qquad (10)$$

where [Equations 11 and 12]:

$$C_1 = \frac{\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_{n-1} N_1^0}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1) \cdots (\lambda_n - \lambda_1)} \qquad (11)$$

$$C_2 = \frac{\lambda_1 \lambda_2 \cdots \lambda_{n-1} N_1^0}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2) \cdots (\lambda_n - \lambda_2)} \qquad (12)$$

and generally [Equation 13]:

$$C_k = \frac{\prod_{i=1}^{n-1} \lambda_i}{\prod_{i=1,\neq \mathrm{k}}^{n} (\lambda_i - \lambda_k)} N_1^0 \qquad (13)$$

This is the solution of n simultaneous differential equations [Equations 14 and 15]:

$$\frac{\mathrm{d}R_1}{\mathrm{d}t} = -\lambda_1 R_1 \qquad (14)$$

$$\frac{\mathrm{d}R_k}{\mathrm{d}t} = \lambda_{k-1} R_{k-1} - \lambda_k R_k \quad (k = 2, 3, \ldots, n) \qquad (15)$$

2.4 Kinetics of Formation of Radioactive Nuclides by Irradiation

When a thin target is bombarded with a beam of projectiles, the rate of nuclear transformation (the number of nuclides produced per unit time) is proportional to the beam intensity (I = number of incident particles per unit time), to the target nucleus density (n = number of target nuclei per unit volume) and to the thickness of the target ($\mathrm{d}x$). The proportionality to the thickness of the target is limited to sufficiently small $\mathrm{d}x$, such that both the intensity of the beam and the energy of the projectiles remain practically unchanged. The proportionality constant is denoted σ and is called the reaction cross-section. This term comes from a simple model, assuming that each geometrical collision leads to a nuclear reaction. In this case, σ is the geometrical cross-section of the collision pair, $p(r_1 + r_2)^2$ [Equation 16]:

$$\frac{\mathrm{d}N^*}{\mathrm{d}t} = \sigma I n \, \mathrm{d}x \qquad (16)$$

Since the nucleus radius is a few femtometers (10^{-13} cm), the cross-sections are of the order of 10^{-24} cm^2. This is the reason why it has been customary to express cross-sections in units of barns, where 1 barn (b) = 10^{-24} cm^2.

Equation (16) assumes that the beam cross-section is smaller than the target size facing the beam, and consequently each projectile particle is transversing through the target. In the case that the target size is smaller than the beam cross-section, the beam intensity, I, should be replaced by the product ϕA, where ϕ is the beam flux (number of projectile particles per unit area and per unit time) and A is the area of the target facing the beam [Equation 17]:

$$\frac{\mathrm{d}N^*}{\mathrm{d}t} = \sigma \phi A n \, \mathrm{d}x \qquad (17)$$

$A \, \mathrm{d}x$ is the volume of the target and consequently $nA \, \mathrm{d}x$ is the total number of target atoms, N [Equation 18]:

$$\frac{\mathrm{d}N^*}{\mathrm{d}t} = \sigma \phi N \qquad (18)$$

Since we are speaking of a reaction of a specific nuclide, N is the total number of the atoms of that nuclide. Equation (18) is used mainly for irradiation in a nuclear reactor, where the target is located inside a uniform flux of neutrons. Equation (10) is used in cases of a narrow beam from charged particle accelerators.

In the case of prompt activation analysis, the only important number is the number of nuclei which were transformed by the nuclear reaction. Hence the number of nuclei transformed is given by the integration of Equation (16) or (18) over the irradiation time. However, for delayed activation analysis, the important factor is the number of nuclei decaying during the measurement period, which is delayed and done some time after the end of irradiation. In calculating the number of nuclei of the newly formed radionuclide at the end of the irradiation, it would be wrong to integrate those equations. The radionuclides are not only formed during the irradiation, but also are decaying. Combining Equations (18) and (4) leads to the complete equation for the rate of the change

of the number of radioactive nuclei during the irradiation in a constant flux of projectiles (ϕ) [Equation 19]:

$$\frac{dN^*}{dt} = \sigma\phi N - \lambda N^* \tag{19}$$

Integration of Equation (19) with the initial condition $N_0^* = 0$ leads to the equation for the number of radioactive nuclei at the end of irradiation for time t_i [Equation 20]:

$$N_{EOI}^* = \frac{\sigma\phi N}{\lambda}(1 - e^{-\lambda t_i}) \tag{20}$$

where t_i is the irradiation time and EOI stands for end of irradiation. In the product λt_i, λ should have the same units as t_i, whereas in the term before the parentheses the units should be the same as the time unit of ϕ.

For cases where $\lambda t_i \ll 1 (t_i \ll t_{1/2})$, the term in parentheses is approximately equal to λt_i ($e^{-\lambda t_i} \approx 1 - \lambda t_i$ for $\lambda t_i \ll 1$) and N_{EOI}^* is equal to $\sigma\phi N t_i$, i.e. the rate of formation multiplied by the irradiation time, since the decay is negligible.

For an irradiation time that is long relative to the half-life of decay ($t_i \gg t_{1/2}$, $\lambda t_i \gg 1$), the term in parentheses is equal to 1, which is its maximum value. This is the maximum number of nuclei which can be formed. Longer irradiation does not increase the number of radioactive nuclei, since the rate of formation by the nuclear reaction is equal to the rate of disappearance by radioactive decay. This statement assumes that N does not change substantially during the irradiation, which is common for the usual values of ϕ (10^{15} n s^{-1} cm^{-2}), σ (1 b) and t_i (<10 days). For the extreme cases $\Delta N/N$ is still usually $\leq 10^{-3}$. This maximum number of radionuclei formed is termed the saturation value or the number at saturation, N_{sat}^* [Equation 21]:

$$N_{sat}^* = \frac{\sigma\phi N}{\lambda} \tag{21}$$

The saturation value can be increased by increasing the flux of the bombarding particles or the amount of the irradiated nuclei (the mass of the target), but not by longer irradiation. Usually we speak of the mass of a specific element in the target and not of the number of atoms of this element, so it is desirable to transform N, the number of atoms, into m, the mass of specific element in the activated samples. The number of atoms N in m mg of element with atomic weight M is $(m/M)N_A$, where N_A is Avogadro's number. In contrast to the usual chemical measurements, where all natural isotopes of the same element react in almost the same way, this is not true for nuclear activation. For example, let us consider chlorine. Chlorine has two stable natural isotopes, ^{35}Cl with 75.77% abundance and ^{37}Cl with 24.23% abundance. Irradiation with thermal neutrons leads to an (n,γ) reaction forming the two radionuclides, ^{36}Cl and ^{38}Cl. ^{36}Cl cannot be used as IRN for activation analysis owing to absence of γ radiation (^{36}Cl is a pure β emitter) and its long half-life (the IRN is a radionuclide used for the identification and quantification of an element in activation analysis). Hence, it is not the total number of chlorine atoms that should be used in Equations (18–21) but rather the number of ^{37}Cl atoms alone. This is done by multiplying N by the abundance of the parent of the IRN (0.2423 for ^{37}Cl). Using f for the abundance, Equation (21) takes the form of Equation (22):

$$N_{sat}^* = \frac{m}{M} N_A f \frac{\sigma\phi}{\lambda} \tag{22}$$

The number of the nuclei of the IRN at the end of irradiation is obtained from Equation (16) as Equation (23):

$$N_{EOI}^* = N_{sat}^*(1 - e^{-\lambda t_i}) \tag{23}$$

The measurement of the γ activity (the rate that the γ-rays are emitted by the activated sample) is done not immediately after the end of irradiation, but rather after some time, which is called the decay time, t_d. The decay time (called also 'cooling' time) can be fairly short, being only the time required to transfer the sample from the irradiation port to the measurement station, or can be long. Long decay times are used when we are interested in IRNs with medium to long half-lives (they are also called medium- to long-lived IRNs). The high activity of the short-lived radionuclides at short times after irradiation obscures the activity of the longer lived radionuclides. Consequently, in order to measure medium- and long-lived IRNs, their measurement is delayed. This is the reason why the decay time is called also 'cooling time'. For determining the concentrations of many elements in the sample, the γ activity in the irradiated sample is measured several times (at least three) after different decay times. Since the number of atoms of the IRN at the beginning of the decay time is N_{EOI}, the number at the end of the decay period, t_d, is given by Equation (24) [according to Equation 5]:

$$N_{EOD}^* = N_{SOC}^* = N_{EOI}^* e^{-\lambda t_d} = N_{sat}^*(1 - e^{-\lambda t_i})e^{-\lambda t_d} \tag{24}$$

The subscripts EOD and SOC stand for end of decay and start of counting, respectively. The measurement of the γ activity is called counting, since it is done by the counting of the pulses formed in the γ-ray detection systems, due to the interaction of the γ-photons with the detector.

The counting is done during a time t_c called the counting time. The number of atoms of the radionuclides still undecayed at the end of the counting time, N_{SOC}^*, is given by Equation (25) [according to Equation 5]:

$$N_{EOC}^* = N_{SOC}^* e^{-\lambda t_c} = N_{sat}^*(1 - e^{-\lambda t_i})e^{-\lambda t_d}e^{-\lambda t_c} \tag{25}$$

The number of atoms decaying in the counting period (ΔN_c) is given by the difference in the number of atoms at the beginning of the counting period, N^*_{SOC}, and the number of atoms at the end of this period, N^*_{EOC} [Equation 26]:

$$\Delta N_c = N^*_{SOC} - N^*_{EOC} = N^*_{sat}(1 - e^{-\lambda t_i})e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \quad (26)$$

The actual number of events of decay recorded by the detection system, is smaller than ΔN_c owing to three different factors: (i) the intensity of the specific γ line (number of photons emitted per 100 atoms decayed) is in many cases $< 100\%$; (ii) not every γ-photon emitted reaches the detector, as the photons are emitted isotropically and only those in the direction of the detector will hit the detector; and (iii) some of the photons reaching the detector will not interact with it, and others will lose only part of their energy in the detector, and consequently will not be recognized as originating from the specific IRN. The last two factors are combined in a factor called the geometric efficiency of the detector, ε. This factor depends on the detector, the energy of the γ line and the distance of the sample from the detector. For large samples, ε depends also on the sample size. The factor ε is determined experimentally by counting the activity of calibrated standards (which can be bought commercially), for which the rate of disintegration is known accurately. The rate of decay of the standards is given on its certificate at specific date and time. Its activity at the time of calibrating our system is calculated using Equation (5). Note that Equation (5) applies both to the number of atoms in the sample and to the activity of the sample since the rate is $-dN^*/dt$, which is given by λN^*. Hence the number of counts recorded by the detection system, C, is given by Equation (27):

$$C = \varepsilon I_\gamma \Delta N_c = \varepsilon I_\gamma N^*_{sat}(1 - e^{-\lambda t_i})e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \quad (27)$$

where I_γ is the intensity of the specific γ line. Thus, for activation in a uniform flux, the number of measured counts is given by Equation (28):

$$C = mN_A f\sigma\varepsilon I_\gamma \phi \frac{1 - e^{-\lambda t_i}}{\lambda} e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \quad (28)$$

For a beam smaller than the target, ϕ is replaced by I/A, where A is the area of the target, usually expressed by $m_{sample}/(\rho\, dx)$. It should be noted that here the mass is the total mass of the sample and not the mass of the determined element; $\rho\, dx$ is the common way in nuclear physics to express thickness and its units are grams per square centimeter.

Equation (28) shows that the number of counts recorded is proportional to the mass of the element responsible for the specific γ line, and this is the basis of the activation method since the measurement of the specific γ line yields the mass in the sample of the specific original nuclide.

For some elements, mainly gold and tantalum, the radionuclides produced by the (n,γ) reaction with the reactor neutrons have exceptionally high cross-sections (e.g. for ^{198}Au the cross-section is 2.5×10^4 b). Owing to the high cross-section and although their concentration is very limited, a radionuclide is formed by absorption of two neutrons (in two consecutive steps and not in one interaction) in the stable nuclide. For example, for gold, which has only one stable isotope, ^{197}Au, the process is illustrated by Equation (29):

$$\begin{array}{ccccc} ^{197}\text{Au} & \longrightarrow & ^{198}\text{Au} & \longrightarrow & ^{199}\text{Au} \\ & & ^{198}\text{Hg(stable)} & & ^{199}\text{Hg(stable)} \end{array} \quad (29)$$

For ^{198}Au, the rate of formation is given by Equation (30):

$$\frac{d[^{198}\text{Au}]}{dt} = [^{197}\text{Au}]\sigma_1\phi - (\lambda_1 + \sigma_2\phi)[^{198}\text{Au}] \quad (30)$$

Equation (30) is the same as Equation (19) substituting λ for $\alpha_1 = \lambda_1 + \sigma_2\phi$. Hence the number of ^{198}Au atoms at the end of irradiation is given by Equation (31):

$$\begin{aligned} N(^{198}\text{Au})_{EOI} &= \frac{\sigma_1\phi[^{197}\text{Au}]}{\lambda_1 + \sigma_2\phi}[1 - e^{-(\lambda_1+\sigma_2\phi)t}] \\ &= \frac{\sigma_1\phi[^{197}\text{Au}]}{\alpha_1}(1 - e^{-\alpha_1 t}) \end{aligned} \quad (31)$$

For ^{199}Au, the rate of formation is given by Equation (32):

$$\frac{d[^{199}\text{Au}]}{dt} = \sigma_2\phi[^{198}\text{Au}] - \lambda_2[^{199}\text{Au}] \quad (32)$$

This is similar to the set of the differential Equations (15), differing in that the constants in the positive term in one equation differ from those in the negative term in the next one, as not all of the disappearance of R_i leads to R_{i+1}; however the same method of Bateman can be used also here. The solution of Equation (32), substituting in it Equation (31), yields Equation (33):

$$[^{199}\text{Au}] = \frac{\sigma_1\sigma_2\phi_2}{\alpha_1\lambda_2}[^{197}\text{Au}^0]\left(1 - \frac{\lambda_2 e^{-\alpha_1 t} - \alpha_1 e^{-\lambda_2 t}}{\lambda_2 - \alpha_1}\right. \quad (33)$$

This equation neglects the disappearance of ^{199}Au due to the absorption of neutrons. An equation which includes it was given by Heydorn,[14] but its contribution is negligible for all fluxes used in activation analysis. The general treatment can be found in Friedlander et al.[3] (p. 201).

3 THE SHAPE OF A γ SPECTRUM

In optical (absorption and emission) spectrometry we know that if a species has only one excited state its spectrum will appear as a single line without any background. This is due to the yes/no quantum characteristic of the energy transfer. The energy difference between the two states is either transferred or not transferred, but it cannot be transferred in parts. A photon of light in the infrared, visible or ultraviolet range can be absorbed totally or not at all, but it cannot lose only part of its energy. The situation is different for high-energy γ-rays. Low-energy X-rays interact with matter by the photoelectric process. In this process the γ-photon interacts with one electron of the material, losing all its energy to this electron. The transferred energy is higher than the binding energy of the electron, and thus the electron is ejected from the atom with kinetic energy equal to this difference ($h\nu_{\text{photon}}$ − binding energy of the electron). However, since the range of high kinetic energy electrons is considerably shorter than that of γ-rays, the electrons will lose their kinetic energy inside the detector material and from the point of view of the detector the complete energy of the photon have been absorbed in it, resulting in the same energy absorbed for all photons (assuming that all photons are of the same energy). The situation is different for higher energy γ-rays. The cross-section (probablity of reaction = rate constant) for the photoelectric process decreases with increasing energy of the photon more than that of the Compton scattering. For higher energy photons the main interaction with matter is via Compton scattering. In germanium the photoelectric effect is the dominant process for γ interaction up to about 200 keV, whereas from 200 keV upwards the Compton scattering becomes more important. It should be stressed that also below 200 keV the Compton scattering occurs but its contribution to the interaction of the photon with the germanium is less than that of the photoelectric process. Its contribution increases with the energy although its cross-section decreases with the energy, since it decreases less than that of the photoelectric process. In the Compton scattering the photon interacts with what might be called a free electron (mainly electrons from outer shells whereas the photoelectric process occurs with inner shell electrons, mainly from the innermost one, the K shell). The photon transmits only part of its energy. The energy that the electron receives is more than sufficient to eject it from the atom and it moves with the excess energy as kinetic energy, losing it by collisions in a very short distance. The photon retains part of its energy, by changing its frequency since $E = h\nu$, and scattered in a different direction. In the Compton process the energy of the original photon is shared between two particles (the ejected electron and the scattered photon), and consequently there is a continuous distribution of energies of the scattered photon. In order to conserve both energy and momentum, the γ-photon cannot lose all its energy (in the photoelectric effect the momentum conservation is compensated by the recoiling of the atom from which the electron was ejected, whereas in Compton scattering it is an interaction with a 'free' electron). The maximum energy that the photon can lose is given by Equation (34):

$$\frac{E_\gamma}{1 + m_0c^2/2E_\gamma} \tag{34}$$

where E_γ is the energy of the initial photon and m_0 is the electron rest mass; m_0c^2 is the rest mass of an electron in energy units and it is equal to 0.511 MeV. The main difference between the photoelectric and Compton processes with respect to their responses in the detector material is that the electrons (in both processes) lose all their energy in the detector, owing to their short range, while the scattered photon might escape from the crystal without further interaction. In γ-ray spectrometry we are measuring the full-energy peak, called the photopeak, as this energy will be the only one to appear if the only interaction process is photoelectric absorption. The photopeak results either from the γ-photons losing all their energy by photoelectric absorption or from a Compton scattering process followed by photoelectric absorption of the scattered lower energy γ-photon in the detector (the photoelectric absorption can be after one scattering or several scatterings, all of them within the detector material). However, the scattered γ-photon from the Compton process (actually from the several consecutive Compton processes) might escape from the detector crystal, leaving in the detector less energy than the full energy peak. This escape of scattered photons not only reduces the photopeak, but its main disadvantage is the formation of a background for lower energy peaks, since part of the energy (those given to the electron) is absorbed in the detector and the detector sees them as counts of lower γ energy than the actual energy of the γ-photon emitted by the IRN. The scattered photons do not have discrete energy and can be between a minimum given by Equation (34) and a maximum given by the original energy of the photon (minus the small value of the binding energy of the photon, which is not completely a free electron). For one γ energy source the Compton scattering produces a continuum background ranging from zero up to a maximum, called the Compton edge, given by Equation (35). This results from the fact that the minimum energy that the scattered photon can have is not zero but rather is given by Equation (35):

$$E_{\text{minimum of scattered}\,\gamma} = E_\gamma - \frac{E_\gamma}{1 + m_0c^2/2E_\gamma} = \frac{E_\gamma m_0c^2}{2E_\gamma + m_0c^2} \tag{35}$$

For $E_\gamma \gg m_0c^2$, E_{minimum} approaches a value of $m_0c^2/2$, i.e. 256 keV. Thus, for high-energy γ the Compton edge will be separated from the photopeak by about 256 keV. For a one-energy source the range between the photopeak and the Compton edge is almost free of counts. However, most real samples have many IRNs having different γ energies and the Compton continuum stretches from zero up to the Compton edge of the highest γ energy.

If the energy of the measured photon is >1.022 MeV, a third process can also be operative in the interaction of the photon with the detector – a process called pair production. In this process, the photon energy is transformed, under the influence of the field of a nucleus, into matter in the form of an electron–positron pair (transformation of energy into a matter–antimatter pair). Since the rest mass of each electron or positron is 0.511 MeV, the threshold of this reaction is 1.022 MeV to conserve energy. Although the threshold energy is 1.022 MeV, the cross-section for pair production is very low, for the energy range of <1.5–1.6 MeV, and can be ignored. In this process of pair production, the excess energy is shared between the kinetic energies of the electron and the positron. Since the ranges of positrons and electrons are very short, this excess energy ($E_\gamma - 1.022$ MeV) will be deposited in the detector. When the positron has lost all its kinetic energy, it reacts with an electron in an annihilation reaction to form two γ-photons of 0.511 MeV each (transformation of matter–antimatter into electromagnetic energy). Each of these 511-keV photons can either escape from the detector without any interaction, or lose part of its energy by Compton scattering, or lose all its energy in the detector (either by photoelectric absorption or by succesive Compton scatterings and photoelectric absorption). If one of the 511-keV photons deposits all its energy in the detector, while the other one totally escapes from the detector, the energy absorbed in the detector will be 511 keV less than the full energy peak. With γ-photons of above 1.6 MeV we usually see also this peak of $E_\gamma - 511$ keV, where E_γ is the photopeak energy. This peak is called single-escape (SE) peak, due to one 511-keV photon escaping from the detector. Another peak in the spectrum is the double-escape (DE) peak, rising from the two 511-keV photons escaping from the detector without any interaction with the detector material. The energy of this peak is $E_\gamma - 1.022$ MeV. When analyzing a γ-ray spectrum to find the radionuclides in the analyzed spectrum, every time we find a peak with energy $E_\gamma > 1.6$ MeV, which we know belongs to some radionuclide, we have to remember that in the list of the found peaks there are some which do not belong to other radionuclides but are either SE or DE peaks. We saw that ^{28}Al has a γ energy of 1.778 MeV. Hence we expect to see in the ^{28}Al γ-ray spectrum peaks of 1.778 MeV, 1.269 MeV (SE = 1.778 − 0.511) and 0.758 MeV (DE peak). ^{24}Na has two γ lines of 2.75 and 1.39 MeV. We expect to have four lines in the γ-ray spectrum, 2.75 MeV, 2.24 MeV (SE), 1.73 MeV (DE) and 1.39 MeV. Although 1.39 MeV is above the threshold of 1.022 MeV for pair production, its cross-section for this process is low and pair production contributes very little to the interaction of the 1.39-MeV photon with the detector. Hence in most cases we will not see the SE and DE peaks of this photon.

Figure 1 shows the measured γ-ray spectrum of ^{24}Na. The ratio of escape peaks to the photopeak depends on the detector size (influencing the probability of the γ escape) and on the energy of the γ-photon (affecting the chance that the photon will interact by pair production). In order to know whether a peak is a photopeak or SE or DE, it is preferable to start the analysis of the spectrum from the highest energy peak. For each peak we can remove from the list of peaks its SE and DE peaks.

An important artefact in many γ-ray spectra is the sum peak, a peak which is due to the summation of two γ-photons. If two photons are interacting with the detector, and the time between them reaching the crystal is smaller than a time characteristic of the detector, the detector treats both of them as a single photon and the energy deposited is the sum of the energies of the two photons. The two photons can be due to different radionuclides,

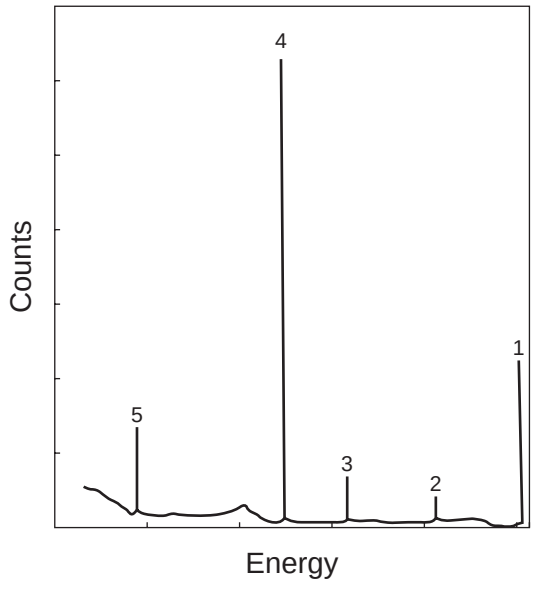

Figure 1 γ-Ray spectrum of ^{24}Na radioisotope taken with an HPGe detector. (1) Photopeak of 2754 keV; (2) SE peak of 2243 keV (2754–511 keV); (3) DE peak at 1732 keV (2754–1022 keV); (4) photopeak at 1369 keV; (5) 511-keV positron annihilation.

or to the same radionuclide which has more than one γ line. In the case of different radionuclides the sum peak is refer to 'pile-up' and can be reduced (relative to the photopeaks) by removing the measured sample further from the detector or by analyzing a smaller sample. A lower rate of counting increases the average time between two photons reaching the detector, thus decreasing the chances of pile-up. In the spectrum of only one γ line there is also a peak with energy given by Equation (35). This peak is due to 180° backscattering from external sources, mainly the shielding of the detector, of the original photon. The backscattered Compton photon reacts with the detector, giving a peak of its full energy. For a source with various γ-ray energies, this backscatter peak can range from 70 to 256 keV, thus adding to the Compton continuum rather than forming a distinct peak.

The Compton background is an intrinsic background due to the physical characteristics of the detector and the interaction of radiation with matter. Another source of background is external, due either to radioactive sources other than our sample or to cosmic radiation. This external background can be minimized by the use of shielding, done mainly with lead owing to its high Z and mainly its high density. Usually the detector is located inside a lead 'castle'. Lead shielding configurations suitable for a germanium detector are available commercially, usually in the form of a hollow cylinder with a sliding lid or revolving door for access.

In many cases where the Compton continuum interferes with the measurements of mainly small peaks, it can be reduced by the use of the Compton suppression method. In this method, the germanium detector is located inside an annular scintillation detector or inside a ring of scintillation detectors. Many of the scattered Compton γ-rays which escape from the germanium detector (in the case of the Compton continuum) deposit some energy in the surrounding detectors, whereas those photons which lose all their energy in the germanium detector do not cause any signal in the scintillation detector(s). The system is operated in the anticoincidence mode, i.e. the multichannel analyzer (MCA) is recording only the events which occur only in the germanium detectors. If within a fixed time the pulse in the germanium detector is followed with a pulse in the scintillation detector (indicating that the germanium pulse is in the Compton continuum and not part of a photopeak), this event is discarded. The scintillation detector(s) are usually NaI(Tl), although in some systems it is a crystal of bismuth germanate (BGO) ($Bi_4Ge_3O_{12}$) owing to its higher efficiency for absorption of γ-rays (since it has a high-Z element). However, BGO crystals are more expensive and are not used in most systems. A schematic diagram of a Compton suppression system is given in Figure 2.

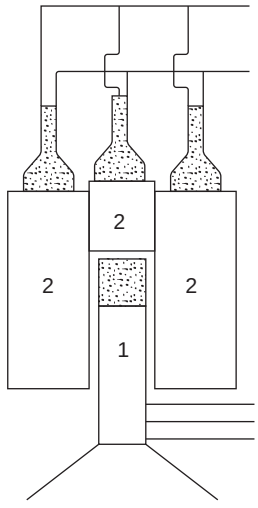

Figure 2 Schematic diagram of anticoincidence Compton suppression γ-spectrometer. (1) HPGe detector with liquid nitrogen cooling; (2) NaI(Tl) detectors with photomultipliers.

4 NEUTRON SOURCES

Three main sources of neutrons are available: (1) research nuclear reactors (nuclear reactors used as neutron sources), (2) ion and electron accelerators including neutron generators and (3) radioactive sources.

Research nuclear reactors have the highest neutron fluxes, but are limited concerning their price and availability. Consequently, nuclear reactors will be used for INAA of very minute amounts, which is the main subject of this article. Research nuclear reactors are usually large devices in which fissionable material, almost exclusively ^{235}U, is fissioned into two nuclides with simultaneous emission of neutrons which induce further fissions in a chain reaction. The fission-produced neutrons are very energetic. The cross-section for neutron-induced fission of fissionable nuclides increases with decreasing energy of the neutrons, and in order to increase the neutron activity, moderators which slow the neutrons are added to the reactor. To reflect back some of the neutrons which leak from the reactor core, reflectors are used. The fission process releases large amounts of energy, mainly due to the stopping of the two recoiling fissioned particles, and the system is cooled by a coolant (either liquid or gas). Nuclear reactors are categorized according to their fuel, moderator, coolant, reflector and configuration.

Almost all research nuclear reactors (neutron sources) are heterogeneous reactors in which the fuel is in the form of rods. The fuel is enriched ^{235}U (natural uranium has only 0.7% of ^{235}U, the fissile material). Most research reactors have 93–99% ^{235}U. Many of the reactors have rods which are U–Al alloys, but some

of the newer designs (mainly those converted to 20% ^{235}U) are of the uranium silicide type. TRIGA reactors operate with uranium–zirconium hydride fuel, which, owing to their large negative temperature coefficient of reactivity, allow the operation of the reactor in pulses. In the light water reactor (LWR), ordinary water (H_2O) is used both as a moderator and as a coolant. The reflector is mainly graphite but there are also Be and H_2O reflected reactors. The construction is either pool type or tank-in-pool type. Owing to the relative high cross-section for capturing thermal neutrons by H atoms and therefore the relatively small amount of moderator, the flux of neutrons in LWRs always contains a large fraction of fast and epithermal neutrons. The available powers are in the range 10–5000 kW with neutron fluxes of 5×10^{10}–1.5×10^{14} n cm^{-2} s^{-1}. Many reactors are unique in their design, but there are some commercial types which are more common, e.g. the American TRIGA and the Canadian Slowpoke. The TRIGA reactor is a popular multipurpose research reactor. About 50 of them are operating with power levels of 18 kW–3 MW fluxes of 7×10^{11}–3×10^{13} n cm^{-2} s^{-1}. The most common types are those of 250 kW and 1 MW. They are of the pool type, graphite reflected with uranium–zirconium hydride fuel with ^{235}U enrichment of 10–70%. The Slowpoke reactor is a low-power (20 kW) reactor designed specifically as a teaching aid with additional purposes of activation analysis and the production of small amounts of radioisotopes. The system is designed to operate remotely. It can be provided with up to five irradiation sites in the core with a flux of 10^{12} n cm^{-2} s^{-1} and five further tubes outside the reflector with half of that flux.

Heavy water research reactors are tank types. They contain usually enriched uranium fuel. Heavy water is used as both moderator and cooler, and neutrons are reflected from the heavy water and graphite. Owing to the low cross-section for thermal neutron absorption by D and O, and owing to the higher mass of moderator, they are characterized by a well-thermalized neutron flux (very little epithermal and fast neutron fluxes except inside the core). Owing to the lower moderation power of D compared with H, the physical size of heavy water reactors is larger and hence they have a large available irradiation volume. Their power is usually between 10 and 26 MW (fluxes of up to 2×10^{14} n cm^{-2} s^{-1}).

4.1 Sample Introduction

The way of introducing a sample into the neutron flux depends on the physical structure of the reactor. It is essential that the introduction of the sample will not affect the operation of the reactor. The irradiation site may be within the reactor core or outside in the moderator/reflector region. If the reactor is an open-pool type with access from above, vertical tubes or ropes can be installed to lower the samples down inside the core or close to the side. The samples are closed in sealed ampules in order not to be in contact with the water surrounding the core. The ampules are usually made of aluminum owing to its corrosion resistance and short-lived activation products. For short irradiations a polyethylene capsule can also be used. The manual loading of samples is neither quick nor reproducible in time when short irradiations are performed. In order to have quick and reproducible sample introduction, mechanical systems are used. Two types of mechanical systems are used: chain-driven racks and pneumatic devices. The latter is the most common, owing to the shorter loading and unloading time and less maintenance and failures of operation. In the pneumatic device, the sample is pushed along a tube with pressurized gas (air or nitrogen). The transfer time depends on the pressure and the distance transferred. In many systems the transit time is 1 s or less. The pneumatic device can be automated very easily. Figure 3 depicts schematically the pneumatic system.

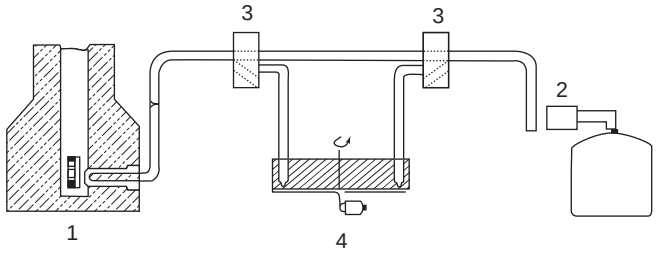

Figure 3 Schematic layout of a fast transfer system for INAA. (1) Reactor; (2) γ-spectrometer; (3) valves; (4) sample changer; the valves can allow also direct transfer of rabbit from reactor to detector or indirect transfer through the sample changer for decay.

With closed-tank reactors, the irradiation is done either by a pneumatic device or with neutron beams.

5 INSTRUMENTAL NEUTRON ACTIVATION ANALYSIS

5.1 Techniques

INAA is one of the simplest techniques for trace element analysis. Samples are inserted into polyethylene or polypropylene cans for short irradiation or into a quartz vial which is inserted in an aluminum can for long irradiation. PVC cans or glass vials cannot be used, since they will be become very radioactive owing to the formation of ^{38}Cl or ^{24}Na, respectively. For a short irradiation time each sample is 'sent' alone

to the reactor by means of a pneumatically operated tube – the rabbit system. For long irradiation times, many samples can be lowered into the reactor and left there for the required time. After the samples have been removed from the reactors, they are counted (this means taking a γ-ray spectrum) immediately or the γ-ray spectrum is taken after some 'cooling' (decay) time, to reduce the high radioactivity level due to short-lived radionuclides. Before counting, the vials are washed to remove surface contaminants. In some cases the samples are counted inside the vial, whereas in others the samples are taken out of the vial before the measurement, owing to interferences from trace elements in the vial.

5.2 Calculation

In general, the first calculation step consists of peak identification, i.e. identification of which peaks are photopeaks (discarding SE, DE and sum peaks) and identification of the radioactive species which are producing them. There are tables of γ-ray energies listed in order of increasing energies for ease of identification.[8–13] Some tables give the energies of all known radionuclides whereas others give them only for radionuclides which can be formed in one form of nuclear activation, e.g. in a nuclear reactor. For better identification the tables give the half-life of this radionuclide, its probability of formation and other γ lines of the same radionuclide. Thus, if we find one γ peak of some radionuclide, but not its other line/lines, it means a wrong identification. The different lines also have to be checked to be in the reported ratios, taking into account the relative intensities of the γ lines and the efficiency of the detector at those energies. Nowadays most MCAs are computer (PC)-based and have in their data a nuclide library. However, most libraries check only for the energy of the photopeak with a preset tolerance of 1–2 keV. Some commercial computer programs check also for the presence (in the right ratios) of the other lines of the same radionuclide. Most programs do not treat the identification through half-lives. It should be remembered that SE and DE energies are not listed in most libraries and this can lead to misidentifications. Figure 4 gives an example of γ-ray spectra measured from an activated sample.

The second step in the calculation is calculation of the net counts under the peak (plotted as number of counts in each channel per channel number). Various methods exist for the evaluation of the net peak 'area' (really it is the net peak counts). The simpler methods define the channels which belong to this peak and sum all the counts in that peak [Equation 36]:

$$\text{total} = \sum_{i=L}^{u} C_i \tag{36}$$

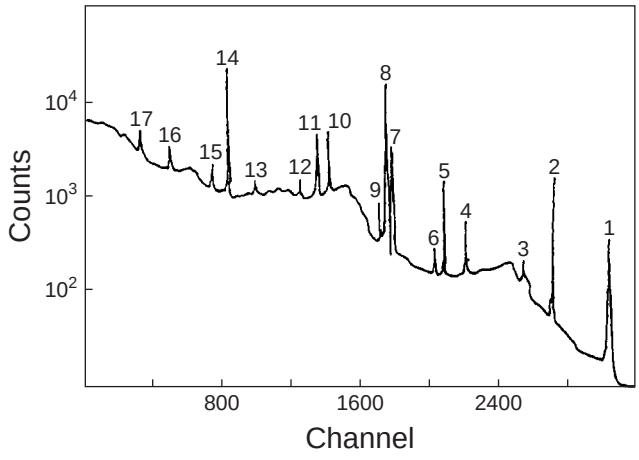

Figure 4 γ-Ray spectrum of a neutron-activated geological sample measured with an HPGe system 15 min after the end of irradiation. (1) 3084 keV ^{49}Ca; (2) 2754 keV ^{24}Na; (3) 2573 keV SE of ^{49}Ca; (4) 2243 keV SE of ^{24}Na; (5) 2113 keV ^{56}Mn; (6) 2062 keV DE of ^{49}Ca; (7) 1811 keV ^{56}Mn; (8) 1779 keV ^{28}Al; (9) 1732 keV DE of ^{24}Na; (10) 1434 keV ^{52}V; (11) 1369 keV ^{24}Na; (12) 1268 keV SE of ^{28}Al; (13) 1014 keV ^{27}Mg; (14) 846 keV ^{56}Mn + ^{27}Mg; (15) 757 keV DE of ^{28}Al; (16) 511 keV β^+ annihilation; (17) 320 keV ^{51}Ti.

where C_i is the counts in channel number i, and L and u are the lower and upper channels belonging to the peak. However, this total count value contains also the background (Compton and external), which should be subtracted. The background is estimated by taking the average number of counts in 3–5 channels on both sides of the peak, multiplying by the number of channels in the peak [Equation 37]:

$$\text{background} = \frac{u - L + 1}{2 \times 3} \left(\sum_{i=L-3}^{L-1} C_i + \sum_{i=u+1}^{u+3} C_i \right) \tag{37}$$

The value of 3 in this equation can be increased or decreased, but it is about the optimum value; lower numbers will increase the statistical error in the evaluation of the background, and larger numbers will increase the possibility that the following channels belong already to other peaks. The radioactive decay is not an absolute process but rather of a statistical nature, and consequently the larger the number of counts the more accurate is the result. The standard deviation of each decay count (as long as the counting time is small compared with the half-life of the measured radionuclide) is equal to the square root of the number of counts. The number of net counts is given by Equation (38):

$$N = \text{net counts} = \text{total} - \text{background} = T - B \tag{38}$$

The standard deviation of the net counts, σ_N, and the relative standard deviation, $\%\sigma_N$, are given by

Equations (39) and (40):

$$\sigma_N = \sqrt{T+B} = \sqrt{N+2B} \tag{39}$$

$$\%\sigma_N = \frac{100\sqrt{T+B}}{N} = \frac{100\sqrt{N+2B}}{N} \tag{40}$$

It should be kept in mind that the rate of counting is limited both by the rate of the process in the detector (which causes 'pile-up' or a sum peak in the case of high rates of counts) and by the rate of the sorting-out of pulses by the MCA and the electronic system. During the 'busy' period of the electronic system, they cannot acquire another signal, so that each signal reaching the system during its 'busy' period will be discarded. This 'busy' period is called the dead time and the actual measurement time (live time) is smaller than the real time (measured by the clock). Most MCA systems contain a device for correction of the dead time. However, these devices are accurate only up to about 10% dead time. It is usually recommended not to measure with dead times above 10%. In spite of this recommendation, there are cases where measurements with high dead times are inevitable, e.g. the measurement of a small amount of a short-lived IRN in the presence of a considerably higher activity of longer lived radionuclide.[15] In this case the correct live time can be measured[16] by the use of a pulse generator with a known rate at which its pulses are introduced to the same electronic system (the height of the pulse is chosen in a range were there are no γ signals), calculating the live time from the ratio of the measured to expected counts from the pulse generator. Instead of using a pulse generator, the longer lived radionuclide can be measured after partial decay or preferably at larger distance.[17]

5.2.1 Calibration and Calculation

The amount of a nuclide in a sample can be calculated from the counts of the radionuclide produced from its neutron activation by the use of Equation (28) if the physical constants $\sigma, f, I_\gamma, \phi$ and λ are known and the geometry factor ε is measured by using calibrated radioactive standards, which can be bought commercially. These standards are calibrated in a 4π detector, i.e. a detector which surrounds the source from all directions such that any β-particle or photon interacts with the detector. The geometry factor ε depends on the distance of the source from the detector and the energy of the γ line. For a fixed distance, ε is a maximum type function of the energy, the maximum being at 100–150 keV. For higher energies than the maximum, $\log\varepsilon$ is approximately a linear function of log(energy), although for more accurate interpolation a quadratic dependence of $\log\varepsilon$ on log(energy) is used. Standard sources often used for calibration are ^{241}Am (60 keV), ^{131}Ba (81, 302, 356 and 383 keV), ^{137}Cs (661 keV), ^{60}Co (1773 and 1332 keV), ^{22}Na (1275 keV) and ^{88}Y (898 and 1836 keV). These standard radioactive sources are used also for the calibration of the number of channels vs. the appropriate γ energy. The standard radioactive sources are supplied with a certificate, certifying their activity at a fixed date. The activity is given either in microcuries [$1\,\mu\text{Ci} = 3.7 \times 10^4$ dps (disintegrations per second)] or in kilobecquerels ($1\,\text{kBq} = 10^3$ Bq, 1 Bq = 1 dps, $1\,\mu\text{Ci} =$ 37 kBq). The activity of the source at the measurement is calculated by the use of Equation (41) which is derived from Equation (2):

$$A(t) = A_0 \exp(-\lambda t) \tag{41}$$

where $A(t)$ is the activity at the measurement time, A_0 is the certified activity and t is the time elapsed from the date of certification to the measurement.

The parameters f, I_γ and λ are tabulated and known very accurately. However, σ is a function of the energy of the neutrons and the real term in Equation (28) should not be $\sigma\phi$ but rather the integral $\int_0^\infty \sigma(E)n(E)\,\mathrm{d}E$, where $n(E)$ is the energy – flux density or the flux distribution function (number of neutrons per second per square centimeter for which their energy is in the range $E + \mathrm{d}E$ to E), $\phi = \int_0^\infty n(E)\,\mathrm{d}E$. These functions $n(E)$ and $\sigma(E)$ are not always known and sometimes they change with time and with location in the reactor. In order to overcome this problem, the use of comparators has been applied to calculate the results of neutron activation analysis (NAA). A comparator is a standard which contains a known amount of the element to be determined in the sample. If the sample (denoted by subscript s) and the comparator (denoted by subscript c) are irradiated simultaneously and counted under the same conditions, then Equation (26) gives Equation (42):

$$m_\mathrm{s} = m_\mathrm{c}\frac{C_\mathrm{s}}{C_\mathrm{c}}\frac{D_\mathrm{c}}{D_\mathrm{s}}\frac{M_\mathrm{c}}{M_\mathrm{s}} \tag{42}$$

where D is the decay factor [$D = \exp(-\lambda t_\mathrm{d})$] and M is the measurement time factor [$M = 1 - \exp(-\lambda t_\mathrm{c})$], where t_c is the counting time, not related to the subscript c representing the comparator. This method of comparators is the most accurate one, and it is the best one to use if the concentration of a single element (or even a few elements) in the sample is required. It should be noted that the sample and the comparators are of similar size and composition to compensate for unequal self-absorption of the measured γ-photons in the sample or comparator. This problem of self-absorption is almost negligible if both sample and comparators are small. The main problem in the use of a comparator for each element is when INAA is used for multielement analysis, which is the common case. In some samples more than 25 elements

are determined simultaneously (more accurately, by three irradiations for different periods of times followed by 5–6 countings at different times after the irradiations for various periods of times).[18–23] For these multielement determinations it is difficult to prepare comparators for all the elements, and further it become impractical owing to the large amount of space required for all the comparators. Thus for multielement analysis either multielemental comparators have to be used, or one or two element comparators are used for calculation of other elements. Some multielement standards (comparators) of different types (geological, biological, etc.) have been prepared either by groups of scientists or by governmental or international organizations. However, such a standard [standard reference material (SRM)] is not recommended as a primary standard owing to some large uncertainties in some elements and problems with homogeneity of small samples (of the SRMs) used in INAA.

The use of one or two elements as comparators for all elements has become more widely used in recent years. The advantage of using two elements or one element with two isotopes which are activated is that the use of two activated nuclides, if chosen well, can give more information on the energy distribution of the neutron flux in the reactor.[24–26] The most popular method is the so-called k_0 method, in which k_0 values were determined for most of the elements according to Equation (43):

$$k_0 = \frac{B_c}{B_s} \tag{43}$$

where [Equation 44]

$$B = \frac{C}{fI_\gamma \sigma_{th}} \tag{44}$$

and the subscripts c and s stand for comparator and sample/element, respectively, and σ_{th} is the cross-section for reaction with thermal neutrons. The most popular comparator is Zr, which has two activatable isotopes, used to monitor the thermal and epithermal flux simultaneously. The calculation of the amount of the different elements involve both the thermal and epithermal neutrons.

5.3 Nuclear Interferences

There are two main kinds of interferences in the calculation of trace element concentration by INAA. The first is the formation of the same radionuclide from two different elements. This would be almost impossible if all neutrons were thermal neutrons since then the only possible reaction is (n,γ). However, all reactors have some higher energy neutrons, which can induce other reactions. Examples of this case are the determination of Mg in the presence of Al, Cr in the presence of Fe or Al in the presence of Si or P. Mg can be determined only through its less abundant isotope $^{26}Mg(n,\gamma)^{27}Mg$. The cross-section for this reaction is rather low, making this measurement not very sensitive. It suffers interference since ^{27}Mg can be formed from nonthermal neutrons reacting with Al in an (n,p) reaction, $^{27}Al(n,p)^{27}Mg$. Similarly, the determination of chromium by the $^{50}Cr(n,\gamma)^{51}Cr$ reaction suffers interference by the nonthermal neutron reaction $^{54}Fe(n,\alpha)^{51}Cr$. Al is determined by the ^{28}Al,$^{27}Al(n,\gamma)^{28}Al$ reaction, which can be produced also by the reactions $^{31}P(n,\alpha)^{28}Al$ and $^{28}Si(n,p)^{28}Al$. In a flesh sample it was found that the contribution of P to ^{28}Al is more than 10 times that of the ^{28}Al from Al, in most reactors.[27–31]

Even for only thermal neutrons in few cases a specific IRN can be formed from two elements, in cases where the IRN can be formed also by consecutive reaction with two neutrons. Pt has several IRNs but all of them except one are short-lived and suffer from interferences from the major element background. Hence, for minute amounts of Pt it is determined by ^{199}Au formed in the process shown by Equation (45):

$$^{198}Pt \longrightarrow {}^{199}Pt \longrightarrow {}^{199}Au(3.139\,d,\ 158.3\,keV\gamma line) \tag{45}$$

However ^{199}Au is formed also by a two-neutron reaction with ^{197}Au, since ^{198}Au has a very high cross-section for neutron capture (2.65×10^4 b)[32,33] [Equation 46]:

$$^{197}Au \longrightarrow {}^{198}Au \longrightarrow {}^{199}Au \tag{46}$$

Another kind of interference is from two radionuclides having very close γ lines. An example of this kind of interference is the 846.8-keV line of ^{56}Mn and the 843.8-keV line of ^{27}Mg. Nowadays very good spectrometers are able to separate these two peaks if they have comparable activities but this is impossible for low-resolution germanium detectors or when one activity is much higher. If they cannot be separated by the spectrometer there are two ways to overcome this interference:

1. Since the half-life of ^{56}Mn (2.56 h) is longer than that of ^{27}Mg (9.45 min), decay of 2 h will leave practically only ^{56}Mn (the activity of ^{27}Mg will decrease by a factor of 6647 whereas the ^{56}Mn activity will be decreased by a factor of less than two). Thus measurements of the sample after the irradiation and after decay of 2–3 h will give the activities of both ^{56}Mn (from the delayed measurement) and ^{27}Mg (by subtracting the corrected activity of ^{56}Mn from the first measurement).
2. The activities of both radionuclides can be calculated using the fact that ^{27}Mg also has another γ line at 1014 keV. The ratio of the activities of the two lines

can be measured in a pure sample of Mg, and thus the 843.8-keV activity of ^{27}Mg can be calculated from the 1014-keV activity. Subtracting this activity from the combined (843.8 + 846.8)-keV peak yields the activity of ^{56}Mn 846.8-keV peak.

5.4 A Test Case: Instrumental Neutron Activation Analysis of Trace Elements in Silicon

Silicon has three naturally stable isotopes, ^{28}Si (92.21%), ^{29}Si (4.70%) and ^{30}Si (3.09%). Thus only ^{30}Si produces a radionuclide by the usual (n,γ) reaction of thermal neutrons. ^{31}Si produced by the ^{30}Si(n,γ)^{31}Si reaction has a 2.62-h half-life and very few γ-rays, with an intensity of only 0.07% (1.26 MeV). Considering the low abundance, the relatively low cross-section (0.28 b) and mainly the low intensity of the γ-rays, ^{31}Si does not present a large problem in instrumental analysis by γ-ray spectrometry. However, owing to the presence of fast and epithermal neutrons in the reactor, the main activity induced in silicon is due to ^{28}Al, ^{29}Al and ^{27}Mg formed by the reactions ^{28}Si(n,p)^{28}Al, ^{29}Si(n,p)^{29}Al and ^{30}Si(n,α)^{27}Mg. As the half-lives of these radionuclides are short (2.3, 6.6 and 9.5 min, respectively), INAA is possible for radionuclides with half-lives longer than 1–2 h. In any case, the activity induced in nuclides with shorter half-lives is usually not sufficient to measure the required purity for electronic-grade semiconductors, owing to the small amount of radioactive atoms which can be formed until reaching saturation. As there is no background from the Si matrix after about a day or two and since the concentrations of the trace elements are low, the ways to increase the sensitivity (decrease the lower limit of detection) are to irradiate with a larger total number of neutrons (higher flux, longer irradiation times or both). In the high flux reactor (HFR) at Petten, The

Table 1 Limits of detection of trace elements in Si by INAA (ppb by weight)[a]

Element	Ref. 35	Ref. 36	Refs. 37 and 38	Ref. 39	Element	Ref. 35	Ref. 36	Refs. 37 and 38	Ref. 39
Na	0.5	0.1	0.3	0.1	Sb	<0.18	0.01	0.02	0.006
Mg		3000			Te	0.004		0.2	
Al		500			I	8			
Cl		5			Cs	0.0005	0.05		
K	0.15	100	0.6		Ba	0.06	10	1.5	
Ca	10	2000			La	0.00015	0.01	0.0007	
Sc	0.00003	0.003	<0.002		Ce	0.0009		0.1	
Ti	3	1000	30		Pr	0.006			
V			10		Nd	0.01			
Cr	0.002	1	0.02	0.1	Sm	0.00003	0.01	0.0004	
Mn	15	0.1			Eu	0.00007	0.05	0.002	
Fe	0.3	50	4	1.6	Gd	0.006			
Co	0.0005	0.1	0.1	0.001	Tb	0.0001		0.006	
Ni	0.15	2000	3		Dy	300			
Cu	<0.56	0.5	0.1	0.1	Ho	0.003			
Zn	0.015	1.0	0.2	0.3	Er	0.04			
Ga	0.0015		0.05		Tm	0.003			
Ge	1				Yb	0.0002		0.001	
As	<0.033	0.02	0.001	0.008	Lu	0.0004			
Se	0.002		0.1		Hf	0.003	20	0.005	
Br	<0.0024	0.02	0.01		Ta	0.005	0.05	0.01	
Rb	0.01				W	<0.025	0.02	0.0001	
Sr	0.15		4		Re	0.0003			
Y	200				Os	0.0007			
Zr	0.15	500	2		Ir	0.000004		0.0002	
Nb	1.5				Pt	0.004		0.005	
Mo	0.006		0.01		Au	<0.0016	0.0005	0.0001	0.0008
Ru	0.0015		0.1		Hg	0.0006		0.7	
Pd	0.06				Tl	3			
Ag	0.003	0.01	0.08		Pb	300			
Cd	0.015		0.3		Th	0.0002	0.1	0.0015	
In	0.004		0.3		U	0.001	0.3	0.02	
Sn	0.2								

[a] Verheijke et al.[35] irradiated wafers of 15 mm diameter and 0.5 mm thickness (20.6 g) with a flux of 4×10^{13} n cm^{-2} s^{-1} for 72–96 h and counted for 6–8 h; Lindstrom[36] irradiated samples of 100 mg for up to 6 h at 5×10^{13} n cm^{-2} s^{-1}; Revel et al.[37,38] used 1 g of Si and irradiated for 72 h at 2.3×10^{14} n cm^{-2} s^{-1}; Fujingawa and Kudo[39] irradiated 5–10-g samples for 11 d at 2×10^{13} n cm^{-2} s^{-1}.

Netherlands, a special irradiation facility was constructed for irradiation of silicon samples from Philips, in which silicon wafers of up to 15 cm in diameter can be irradiated with 4×10^{13} thermal $n\,cm^{-2}\,s^{-1}$ and the irradiation is done for 72–96 h.[34,35] They also used large Ge(Li) detectors (100–150 cm^3) and counted for 8–16 h, and obtained the lowest detection limits. Table 1 summarizes some of the limits of detection given in the literature for an Si matrix.[35–39] For some elements, e.g. Cl and Mn, shorter irradiations were used, between 0.5 and 2.0 h. A special correction has to be done in the case of the determination of Na owing to the reaction $^{28}Si(n,p\alpha)^{24}Na$, as was shown by Niese[40] by irradiating very pure Si in a reactor core with and without a Cd cover to absorb the thermal neutrons. The contribution of the $^{28}Si(n,p\alpha)^{24}Na$ reaction to ^{24}Na in a silicon matrix depends on the spectrum of the neutrons in the reactor core, as very high energy is required for this reaction and the Q value exceeds that of a 14-MeV neutron generator.[40] Niese found[40] that the contribution of the $^{28}Si(n,p\alpha)^{24}Na$ reaction is equal to 0.7 ppb. For the HFR at Petten, The Netherlands, the contribution is 0.55[35] or 0.6 ppb.[34] Revel et al.[37] found this contribution of the $^{28}Si(n,p\alpha)^{24}Na$ reaction to be <0.4 ppb. Haas et al.[41] warned against the common use of wrapping the sample for irradiation in Al foil, as they showed that there is recoil of ^{24}Na from the Al wrapping to the sample, due to the high recoil energy from the $^{27}Al(n,\alpha)^{24}Na$ reaction. In many other uses it is not important, but owing to the very low concentration of Na in pure Si, this contribution might be considerable.

Several important elements cannot be determined in this method, including mainly the light elements H, Li, Be, B, C, N, O and F as they do not form radioisotopes or they are too short lived. Phosphorus cannot be determined in this way owing to a lack of γ emission from the only produced ^{32}P; however, ^{32}P β-rays can be determined destructively by the extraction of the phosphomolybdic complex in the presence of hold-back carriers of Ta and Au.[42] The chemical yield of the extraction was found to be $83.8 \pm 4\%$ and the decontamination factors were 79–6500 for the different elements studied. The limit of detection was found to be 3×10^{-11} g. They carried out their study[42] for phosphorus-doped silicon and Alfassi and Yang[43] showed that for this case, there is no need for chemical separation and 10–20 d of cooling are sufficient to ensure that the only β emitter is ^{32}P (with an accuracy of 1–2%, which is much better than the reproducibility of the extraction yield). The limit of detection should be about the same, as both methods use liquid scintillation counting for the measurement of the β activity, although Alfassi and Yang used a discriminator which reduces the detection limit by a factor of two.

5.5 An Example of Interference Removal in Instrumental Neutron Activation Analysis: Determination of Platinum in Aerosols[33]

Pt is emitted into the air by corrosion of the catalytic converters used in cars to reduce the emission of CO, NO_x and unburned hydrocarbons. In order to determine the concentration of Pt in the air, a high-volume air-pump pumps air through a filter and the amount deposited on the filter (together with the aerosols) is determined. Pt can be determined via several radionuclides. ^{191}Pt suffers from very low abundance (0.01%) of the parent stable nuclide, ^{190}Pt. ^{197}Pt and ^{199}Pt suffer from relatively short half-lives (18.3 h and 30.8 min). If a measurement is made shortly after the irradiation, these IRNs could not be observed owing to the large amounts of Compton-scattered photons from the major elements Na, Cl and Br. Thus Pt can be determined in minute amounts only via ^{199}Au [3.14 d, 158.4 keV (36.9%)], which is formed by the sequence of reactions in Equation (47):

$$^{198}Pt(7.2\%) \longrightarrow {}^{199}Pt \longrightarrow {}^{199}Au \qquad (47)$$

^{199}Au suffers from the two kinds of interferences mentioned earlier: (1) it can be formed also from other elements besides Pt and (2) another radionuclide formed from other elements have a very close γ line. In the following paragraphs we will show how these interferences can be overcome.

1. ^{199}Au can be formed also from the successive absorption of two neutrons by ^{197}Au owing to the very high cross-section for the (n,γ) reaction of ^{198}Au (26 500 b) and the high cross-section for the (n,γ) reaction of ^{197}Au (98.7 b) [Equation 48]:

$$^{197}Au \longrightarrow {}^{198}Au \longrightarrow {}^{199}Au \qquad (48)$$

   Since ^{198}Au is a radionuclide with a close half-life to ^{199}Au (2.69 d), it can be measured simultaneously with ^{199}Au, and can be used to calculate the concentration of Au in the sample and hence to calculate the contribution of Au to the measured peak of ^{199}Au. Subtracting this contribution, we obtain the counts of ^{199}Au formed from ^{198}Pt and hence the concentration of Pt.
2. Another problem is the very close γ lines of ^{199}Au (158.4 keV) and ^{47}Ca (159.3 keV). If similar counts are obtained for the two radionuclides, the γ spectrum shows clearly two distinct peaks, although the lower one (158.4 keV) does not go to zero, and hence there is lower accuracy in the calculation of the area under each peak (i.e. the counts due to each radionuclide). The situation is worse when the number of counts

of one of the radionuclides is considerably higher (ratio > 4), in which case the spectrum does not show two distinct peaks but one peak with a shoulder, more or less visible. In order to separate this composite peak into its two contributions, two methods can be used. An important remark concerning this measurement is not to use a computer convolution program to calculate the area of this peak, as it will be programmed with a fitting function which is characteristic for a pure photopeak. In this case it is better to calculate the area by a simple integration of the counts under the peak, subtracting a trapezoid for the background correction.[15] The two methods for separation of the contributions are based on (i) the different temporal behaviors of the two contributions and (ii) another γ line which allows the calculation of the contribution of one radionuclide to the mixed 158.4–159.3 keV peak, facilitating the calculation of the contribution of the other one by subtraction.

(a) Temporal separation: although ^{199}Au and ^{47}Sc have close half-lives which usually will not allow the temporal separation of their contributions, their parents [as none of them is produced directly from the (n,γ) reaction and are products of β decay of the (n,γ) products] have considerably different half-lives and hence the two contributions can be separated by measuring twice and solving the two equations (obtained by writing equations for the expected counts, writing x and y for the masses of Pt and Ca) with the two unknowns m_{Pt} and m_{ca} (or x and y, $x = m_{Pt}$ and $y = m_{ca}$) [Equations 49 and 50]:

$$^{198}\text{Pt} \longrightarrow {}^{199}\text{Pt} \longrightarrow {}^{199}\text{Au} \qquad (49)$$

$$^{46}\text{Ca} \longrightarrow {}^{47}\text{Ca} \longrightarrow {}^{47}\text{Sc} \qquad (50)$$

The disadvantage of this method is the need to make two measurements, hence the counting time required is more than doubled, since due to the decay between the two measurements the counting time of the second one must be longer than the first one in order to obtain the same statistical accuracy.

(b) ^{47}Sc does not have a γ peak without interference to calculate its contribution to the composite peak; however, its parent ^{47}Ca has a good γ line at 1297 keV without interference and this γ line can be used to measure the concentration of Ca and consequently its contribution via ^{47}Sc to the 158–159-keV peak.

The above contributory interferences are the only ones in a well-moderated nuclear reactor (heavy water moderated), while in most research reactors (light water moderated) the high-energy neutron flux will lead to additional interferences by the ^{47}Ti(n,p)^{47}Sc and ^{199}Hg(n,p)^{199}Au reactions. These interferences can be removed by additional irradiation inside a box made of B or Cd, which are thermal neutron absorbers, as explained in sections 5.6 and 5.7. However, this correction is not mandatory in most cases. The ^{199}Hg(n,p)^{199}Au reaction can be neglected as the cross-section is lower than for (n,γ) and the fast flux is smaller than the thermal flux and the Hg concentration is low. The ^{47}Ti(n,p)^{47}Sc reaction can be a greater problem owing to the higher concentration of the more abundant Ti. However, since Ti has several isotopes, the measurement of ^{48}Sc, which can be formed only by the ^{48}Ti(n,p)^{48}Sc reaction, can be used to calculate the contribution of the ^{47}Ti(n,p)^{47}Sc reaction.

5.6 Epithermal Instrumental Neutron Activation Analysis

In the usual INAA, the whole reactor neutron energy spectrum is used. However, in some cases, the use of part of the neutron spectrum is preferable; these systems are characterized by large differences in the activation cross-sections for the desired and the interfering nuclides in the various parts of the energy spectrum. The required trace elements are activated with part of the neutron spectrum while the interfering major elements are activated more strongly with the other parts of the spectrum, and thus we prefer to avoid this second part.

The neutron energy spectrum in a nuclear reactor is usually divided, for convenience, into three portions, and their relative abundances are dependent on the reactor structure. The most abundant fraction is the one of thermal neutrons, i.e. those neutrons which are in thermal equilibrium with the moderator atoms. Their most expected energy is equal to kT (where k is Boltzmann's constant and T is the neutron temperature), which at room temperature is equal to about 0.025 eV. The neutrons with energy above that of the thermal neutrons are divided into fast neutrons, those which are directly from fission and have not been moderated at all with energy mainly above 1 MeV, and epithermal neutrons, i.e. partly moderated and having energy between tenths of an electronvolt and 1 MeV. When the whole reactor neutron energy spectrum is used for activation, the main contribution is from the thermal neutrons, also owing to their usually higher cross-section [(n,γ) reactions].

A special case where advantage can be taken of the epithermal and fast neutrons is the case where the required trace elements are activated more strongly

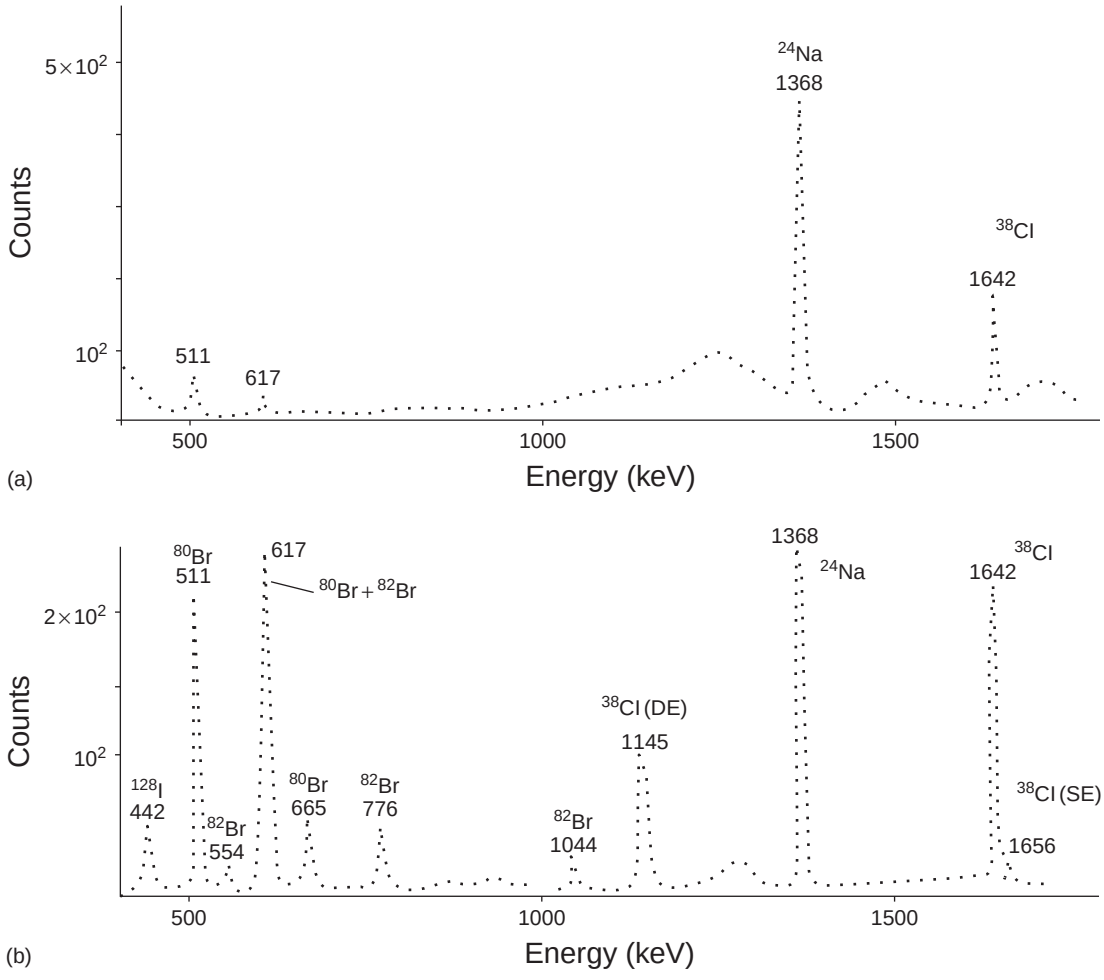

Figure 5 γ-Ray spectrum of a neutron-activated blood sample taken with a Ge(Li) detector. (a) Irradiated bare with reactor neutrons; (b) irradiated in a Cd shield.

relative to the major elements by the epithermal or fast neutrons. Most major elements in geology and biology follow the $1/v$ cross-section rule (their activation cross-section is inversely proportional to the square root of the neutron energy) throughout the whole energy spectrum. On the other hand, many of the less abundant elements have, in addition to their thermal activation, large activation cross-section resonances in the epithermal energy region and consequently can be activated preferentially in this region. Similarly, several of the less common elements can be activated by other neutron reactions besides the common (n,γ) reaction. These (n,p) and (n,α) reactions require higher energy than thermal reactions (and in most cases they do not occur also in the epithermal region and are induced only by fast neutrons).

A simple example of the advantage of using neutron filters (thermal neutron absorbers) in activation analysis is seen in Figure 5(a) and (b), which shows the γ-ray spectra of a sample of blood serum, activated with reactor neutrons once within a cadmium wrapping and once without any absorber (bare irradiation).[44] In the case of activation without a Cd absorber, the Compton-scattered photons of the major elements Na and Cl cover the peaks of bromine and iodine (bromine can be determined only after a delay of several days while the shorter-lived iodine cannot be determined instrumentally and can be determined only after chemical separation). The activation with epithermal neutrons (Cd cover) shows clearly the peaks of Br and I. A more complicated (geological) sample irradiated with and without a Cd cover is shown in Figure 6.

Owing to the presence of both thermal and epithermal neutrons, the product $\sigma\phi$ in Equation (28) should be replaced by the term $\phi_{th}\sigma_{th} + \phi_{epi}I_0$, where ϕ_{th} and σ_{th} are the flux and cross-section of the reaction with thermal neutrons and ϕ_{epi} and I_0 [the resonance integral of the

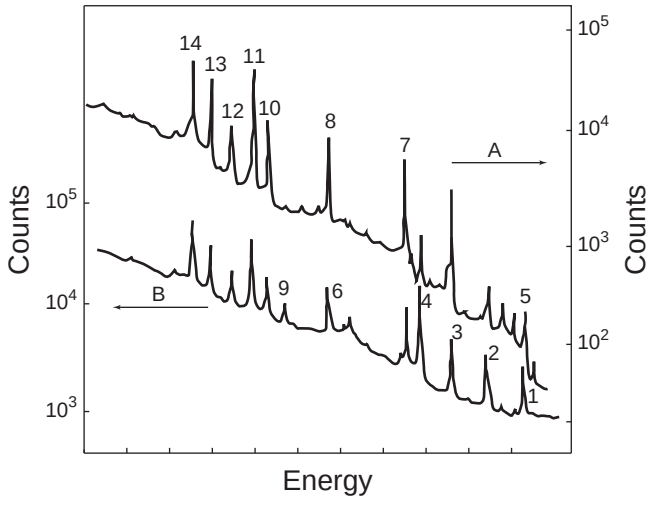

Figure 6 γ-Ray spectrum of neutron-activated air-dust after 10 d of cooling. (A) Inside a Cd sheet; (B) irradiated bare. (1) 1732 keV DE of ^{24}Na; (2) 1595 keV ^{140}La; (3) 1474 keV ^{82}Br; (4) 1369 keV ^{24}Na; (5) 1691 keV ^{124}Sb; (6) 1099 keV ^{59}Fe; (7) 1317 keV ^{82}Br; (8) 1044 keV ^{82}Br; (9) 889 keV ^{46}Sc; (10) 828 keV ^{82}Br; (11) 777 keV ^{82}Br; (12) 690 keV ^{82}Br; (13) 619 keV ^{82}Br; (14) 554 keV ^{82}Br.

(n,γ) reaction] are defined by Equations (51) and (52):

$$\phi_{epi} = \frac{\text{total epithermal flux}}{\int_{0.55\,eV}^{1\,MeV} dE/E} \tag{51}$$

$$I_0 = \int_{0.55\,eV}^{1\,MeV} \frac{\sigma(E)}{E}\,dE \tag{52}$$

The boundaries of 0.55 eV and 1 MeV were chosen since neutrons with energies above 1 MeV are categorized as fast neutrons rather than epithermal neutrons, and 0.55 eV as it is the cut-off in neutron absorption by Cd, the main filter used to capture thermal neutrons.

The epithermal activation properties of a nuclide can be expressed by the ratio I_0/σ_{th} or by means of the absorber ratio. The absorber ratio is the ratio of the activities of this nuclide irradiated once bare (without a cover) and once covered by an absorber of thermal neutrons. Thus the cadmium ratio is given by Equation (53):

$$R_{Cd} = \frac{\phi_{th}\sigma_{th} + \phi_{epi}I_0}{\phi_{th}I_0} = 1 + \frac{\phi_{th}\sigma_{th}}{\phi_{epi}I_0} \tag{53}$$

The most important fact is not only how the insertion of an absorber of the thermal neutrons influences the activity of the specific measured nuclide, but also how it influences the interfering nuclides. Hence it would be advantageous to analyze an element by epithermal neutron activation analysis (ENAA) rather than by using the whole spectrum of reactor neutrons for activation [reactor neutron activation analysis (RNAA)] only if its ratio of resonance integral to thermal neutron cross-section, I_0/σ_{th}, is larger than this ratio for the interfering elements. Several criteria have been suggested to measure the advantage of ENAA over RNAA, as follows.

5.6.1 Brune and Jirlow's Advantage Factor

Brune and Jirlow[45] suggested the use as an advantage factor for ENAA activation of the ratio between the cadmium ratio of the measured nuclide and the interfering nuclide [Equation 54]:

$$F_{BJ} = \frac{R_{Cd}^{i}}{R_{Cd}^{0}} \tag{54}$$

where R_{Cd} is the cadmium ratio as defined previously and the superscripts 0 and i stand for the measured and the interfering nuclides, respectively. This is the most commonly used advantage factor and several tables of this factor for many nuclides have appeared in the literature for cadmium absorbers and for boron absorbers.[11,45–48]

5.6.2 Parry's 'Improvement Factor'

Parry[49] pointed out that while the advantage factor describes well the increase in the signal-to-noise ratio, it does not consider the decrease in the activity of the analyzed element due to the elimination of the activation by the thermal neutrons and hence does not treat the larger error resulting from the lower counting statistics. Parry suggested that the true criterion should be the improvement in the detection sensitivity. The lower detection limit, L_D, for a radioactivity measurement, i.e. the minimum signal which can be detected above the background at the 95% confidence level is given by Equation (55):[50]

$$L_D = 2.33\sqrt{B} \tag{55}$$

where B is the background activity. The minimum detected mass in activation analysis (sensitivity) is given by Equation (56):

$$M_D = \frac{A}{L_D} = \frac{A}{2.33\sqrt{B}} \tag{56}$$

where A is the specific activity of the analyzed element under the experimental conditions for the activation and detection. Since A is proportional to the activity of the analyzed element and since B is due mainly to the interfering nuclide, Parry suggested that the improvement factor, f_p, is given by Equation (57):

$$f_p = \overline{\frac{R_{Cd}^{i}}{R_{Cd}^{0}}} \tag{57}$$

5.6.3 Bem and Ryan's Advantage Factor

Bem and Ryan[51] in 1981 followed the same trend of thinking as Parry; however, they suggested that the advantage factor should describe the improvement of the relative standard deviation of the counts. If the net counts (baseline corrected counts) is N and the background count is B, the advantage factor is given by Equation (58):

$$f_{BR} = \frac{\left(\frac{1}{N} + \frac{2B}{N^2}\right)^{1/2}_{0}}{\left(\frac{1}{N} + \frac{2B}{N^2}\right)^{1/2}_{i}} \tag{58}$$

where the subscripts i and 0 have the same meaning as the superscripts earlier.

When the background is mainly due to the interfering radionuclide, Bem and Ryan's advantage factor is equal to that of Parry. The correlation of Parry's advantage factor with that of Brune and Jirlow can be seen from Equation (59):

$$f_p = \frac{f_{BJ}}{\sqrt{R^i_{Cd}}} \tag{59}$$

Therefore, Parry's advantage factor is always smaller than that of Brune and Jirlow.

5.6.4 Tian and Ehmann's Generalized Advantage Factor

Tian and Ehmann[52] criticized Bem and Ryan's criterion (and consequently Parry's) on the grounds that in practice, in INAA, the limit on the number of counts is not due to the activity of the analyte, but rather to problems associated with the high counting rate. A high count rate leads to inferior resolution and also to problems of pile-up and dead time of the electronic system. In order to overcome these problems, the samples are measured far from the detector or are irradiated for shorter times. When the sample is activated with only epithermal neutrons, the total activity of the sample is reduced considerably, and hence the counting efficiency can be increased by using a smaller sample–detector distance, or the total counts can be increased by using larger samples or longer irradiation times. If the increase in counts (due to either count efficiency or size of the sample or length of irradiation) is given by G^2, then the generalized Tian and Ehmann's advantage factor is given by Equation (60):

$$f_{TE} = G\ \frac{\overline{R^i_{Cd}}}{R^0_{Cd}} \tag{60}$$

If $G = 1$, $f_{TE} = f_p$, and if $G = \sqrt{R^i_{Cd}}$, then $f_{TE} = f_{BJ}$.

The last case is the practical one, since both in the thermal activation and in the epithermal activation, the counting efficiency is usually chosen to obtain the maximum total counting rate allowable by the dead-time correction device. This generalized approach of Tian and Ehmann[52] gives a firmer basis for the widely used definition of Brune and Jirlow.

5.6.5 Thermal Neutron Absorbers

The main absorbers for the thermal neutrons are cadmium and boron owing to their high cross-section for reaction with thermal neutrons. The variation of the cross-section with the energy is different for Cd and for B, hence an intelligent choice of an absorber (also sometimes called a filter) will lead to optimized detection of some nuclides.[53] In some cases, one absorber is used and in others a combination of two absorbers, e.g. Cd + B.[49,53,54] In some experiments, filters which absorb the epithermal neutrons in some regions are used, allowing more selectivity for some elements.[55]

Since the main absorbers are boron and cadmium, it is very important to compare them from the technical point of view. The absorber can be used as a covering sheet, wrapping the sample in it, mainly in the case of cadmium for which metallic sheets are commercially available, or as a capsule built from these materials, or as a mixture with the sample, used with B_2O_3,[56] or as a permanent lined installation inside one of the irradiation ports of the reactor. The use of absorber-lined ports has the disadvantages that (1) scattered thermal neutrons can come from angles which are not covered by the lining of the absorber, leading to lower absorber ratios, whereas when the absorber is used as a capsule or wrapping, it is covered from all angles, and (2) the capsule in which the sample is held during the pneumatic transfer and which usually is made of polyethylene leads to partial thermalization of the epithermal neutrons. On the other hand, the use of an absorber-built capsule or wrapping suffers from the disadvantages of the absorption reactions. Cadmium is activated to short- and long-lived nuclides and the unloading and unpacking of the sample for medium- and long-lived radionuclide measurement faces radiation safety problems owing to the high radiation dose. Short-lived ($t_{1/2} < 20$–30 s) radionuclides cannot be measured at all in a cadmium capsule since the short half-life prohibits the safe unpacking of the vessel and the activity of the absorber is too high to allow measurement together with the filter.

While the absorption of neutrons by ^{10}B does not lead to radioactive products, the reaction $^{10}B(n,\alpha)^7Li$ is very exoergic ($Q = 2.792$ MeV) and the samples are heated considerably. Stroube et al.[57] found that in the 20-MW reactor at the former US National Bureau of

Standards (NBS), thermal heating of the boron nitride vessel limited the length of irradiation for freeze-dried foods to 4 s and prevented completely safe irradiation of wet food. When biological samples are irradiated, this heating accelerates the thermal decomposition of organic compounds, producing high pressure in the sample container, when airtight sealed, which may explode and contaminate or even ruin the irradiation port. In other cases, elements may be volatilized and lost. In order to avoid these effects, the time of irradiation should be limited. Williamson et al.[54] measured the temperature inside a polyethylene rabbit inserted into a Cd-lined irradiation port and found that the temperature reached an equilibrium value of 90 °C in about 7 min. When a BN capsule was irradiated in the same position, the measured temperature was 120 °C in about 3 min and continued to rise. Ehmann et al.[58] irradiated rock samples in a boron carbide filter for 20 h, keeping the sample in heat-sealed quartz ampules. Quartz, being a poor thermal conductor, keeps the sample from being highly heated; this solution is good for geological samples but will probably not suffice for biological samples. Chisela et al.[59] studied the temperature in a sintered BC capsule in an air-cooled irradiation facility and found the capsule to reach steady-state temperatures of 163 and 194 °C for 4.0- and 5.0-MW reactors, respectively. When a permanent installation from powdered B_4C was made with water coolant, the temperature reached not more than 50 °C.

The use of a permanent lining of absorber also has the disadvantage of reducing the total flux of the neutrons in the reactor and of excessive use of the nuclear fuel. A possible advantage of boron over cadmium is the reuse of the same filter in subsequent irradiation. Cadmium filters usually cannot be reused, at least immediately, owing to the long-lived radioactivity produced in cadmium during irradiation. The activity produced in boron filters is small and is due only to contamination in the boron. However, the use of boron filters is limited in many cases to not high total doses of neutrons owing to structural failure of the capsule, probably owing to excessive formation of helium gas from the $^{10}B(n,\alpha)^7Li$ reaction. Cadmium filters are easily made from metallic cadmium sheets of about 1 mm thickness. Boron is a difficult material to machine and Stuart and Ryan[60] prepared boron shields by forming a mixture of boron carbide powder and paraffin wax. The mixture was heated to 70 °C (paraffin melting point = 56 °C) and cast into cylindrical forms. A central hole was made in the form as it solidified by using a heated rod of appropriate diameter. The hole was not made through the whole length of the cylinder, in order to obtain a cylindrical capsule with a central cavity closed at one end. The other end was closed with a top made from the same material. Parry[47] used the same method, but instead of using boron carbide, she used B powder. It should be mentioned that the paraffin causes a small thermalization of the epithermal neutrons. The best machinable refractory boron compound is BN,[61] and consequently many of the studies with boron filters were made using BN capsules. Ehmann et al.[58] suggested that BN should not be used owing to the relatively high cross-section of 1.81 b for the $^{14}N(n,p)^{14}C$ reaction which leads to the formation of an appreciable amount of the long-lived radioactive ^{14}C, but rather boron carbide, another refractory compound of boron, should be used. Unfortunately, boron carbide is extremely hard and is not machinable, and a capsule can be made only by the hot pressing process.

One of the disadvantages of boron filters is the impurities found in boron powder, as discussed in detail by Bem and Ryan.[51] However, if a boron capsule is used together with a permanently installed Cd lining, the interferences due to the activities of ^{28}Al, ^{56}Mn and ^{38}Cl from the boron contaminants are significantly reduced.[47] Both cadmium and boron have high absorption cross-sections for low-energy neutrons; however, the energy dependences of the cross-sections differ considerably. Cadmium approaches a perfect sharp filter for the thermal region and has some resonance in the epithermal range, whereas boron behaves as almost a perfect $1/v$ absorber with no sharp energy cut-off. Although the cross-section for neutron capture by boron is lower than the cross-section for absorption by cadmium in the lower energy range of 0.01–1 eV, it can be compensated for by using thicker boron absorbers. A 0.25 cm thick boron shield is sufficient to stop practically all the thermal neutrons. The effective cut-off energy is almost independent of the thickness of the Cd absorber whereas it increases considerably with increasing thickness of the boron absorber.

5.7 Fast Neutron Instrumental Neutron Activation Analysis

The reactions that occur with fast neutrons with energy usually in the megaelectronvolt range should be looked upon in two ways: (1) the use of these reactions for the determination of some elements and (2) the possible interference of these reactions in the determination of some elements by (n,γ) reactions, due to the formation of the same nuclide as was shown earlier dealing with nuclear interferences. These interferences can be solved only by the use of double irradiation, one with a bare core and one inside a Cd or B filter, and calculating the contribution of each element. The same treatment is usually applied with the use of (n,p) and (n,α) reactions in the determination of some elements. The main advantage of these reactions is that they produce nuclides different from those produced by (n,γ) reactions. Consequently,

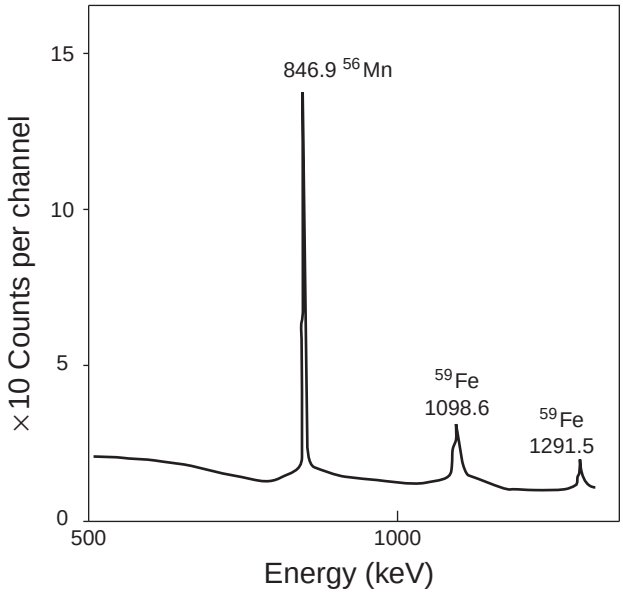

Figure 7 γ-Ray spectrum of iron irradiated with reactor neutrons (irradiation for 10 min, decay for 5 min and count for 6 min).

they may lead to a faster determination in the case of producing a short-lived nuclide rather than the long-lived nuclide produced in (n,γ) reactions. In other cases, they may allow the determination of elements which cannot be measured via (n,γ) reactions since the radionuclide produced is only a β emitter.

5.7.1 Rapid Determination of Iron

Figure 7 shows the γ-ray spectrum of an iron sample irradiated for a short time.[62] As can be seen, the peak of 847-keV γ-rays of ^{56}Mn due to the ^{56}Fe(n,p)^{56}Mn reaction is considerably higher than the ^{58}Fe(n,γ)^{59}Fe 1099-keV peak. Hence the use of (n,p) reactions for the determination of iron has a higher sensitivity in the case of short irradiation and counting. However, ^{56}Mn is formed also from manganese by the ^{55}Mn(n,γ)^{56}Mn reaction. The concentrations of both iron and manganese can be found by double irradiation, one sample with reactor neutrons (without any filter) and one sample with epithermal neutrons (with a cadmium absorber). If the specific activity (measured counts under the experimental set-up per gram of the element) for irradiation with reactor neutrons will be F_R and M_R for iron and manganese, respectively, and similarly for epithermal neutrons F_E and M_E, then the activity of 1 g of sample containing P_F% of iron and P_M% of manganese will be given by Equations (61) and (62):

$$C_R = \frac{F_R P_F + M_R P_M}{100} \tag{61}$$

$$C_E = \frac{F_E P_F + M_E P_M}{100} \tag{62}$$

where C_R and C_E are the activities induced by reactor neutrons and epithermal neutrons, respectively. The solution of these two equation gives Equations (63) and (64):

$$P_F = \frac{M_E C_R - M_R C_E}{\Delta} \tag{63}$$

$$P_M = \frac{F_R C_E - F_E C_R}{\Delta} \tag{64}$$

where $\Delta = 100(F_R M_E - F_E M_R)$.

5.7.2 Determination of Phosphorus and Silicon

Thermal neutron activation cannot be used for the determination of phosphorus and silicon. A radiative capture (n,γ) reaction with the only stable isotope of P leads to the formation of ^{32}P, which is a pure β emitter. In the case of silicon [stable isotopes ^{28}Si (92.2%), ^{29}Si (4.7%) and ^{30}Si (3.1%)], the only radionuclide produced by the (n,γ) reaction is ^{31}Si, which is almost only a β emitter. Its very low intensity of γ-rays (1266 keV, 0.07%) together with the low abundance of ^{30}Si and the low cross-section for radiative capture (0.11 b) permit only the determination of relatively large amounts of silicon. However, activation with epithermal neutrons leads also to the formation of ^{28}Al via both ^{31}P(n,α)^{28}Al and ^{28}Si(n,p)^{28}Al reactions and of ^{29}Al by the ^{29}Si(n,p)^{29}Al reaction. ^{28}Al is also produced by the ^{27}Al(n,γ)^{28}Al reaction. This leads to a procedure for the determination of Si from the activity of ^{29}Al, which is produced only from silicon, using the activities of ^{28}Al from activation with a Cd filter and without a filter to determine the concentration of both aluminum and phosphorus.

However, the activity of ^{29}Al produced for Si is almost two orders of magnitude less than the activity of ^{28}Al produced from it. Hence the use of ^{29}Al will both limit the minimum amount of silicon that can be determined and will reduce the accuracy of the measurement. Another problem associated with the measurement of ^{29}Al is that its main γ line is the 1273-keV line, which suffers from the interference of the SE peak of the more abundant ^{28}Al at 1268 keV.

However, when the concentration of Si is high, silicon can be determined by the ^{29}Si(n,p)^{29}Al reaction, as was done by Hancock[63] for the measurement of silicon in pottery using a Cd shield to decrease the formation of ^{28}Al. The samples were allowed to decay for 17–20 min before counting, to decrease the ^{28}Al activity further. Ördögh et al.[64] measured in that way the concentration of silicon in very small inhomogeneous lymph-node samples. The concentration

of P was determined spectrophotometrically by the molybdenum blue method and hence the concentration of Si and Al can be found from the ^{28}Al activity induced by both epithermal activation (Cd cover) and reactor neutron irradiation. Another way was suggested by Alfassi and Lavi,[65] who used the simultaneous determination of ^{27}Mg and ^{28}Al, each of them for both reactor activation and irradiation with only epithermal neutrons, to measure simultaneously Mg, Al, Si and P. Each of these radionuclides can be formed by three reactions (Scheme 5).

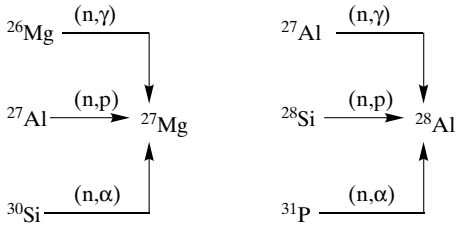

Scheme 5

The concentration of the four elements Mg, Al, Si and P can be obtained from a solution of the four equations for the four measured activities (844 keV due to ^{27}Mg and 1778 keV due to ^{28}Al, each without a filter and with a cadmium absorber). The specific activity (measured counts per gram of the element under the experimental set-up) will be assigned by three letters, the first giving the target element, the second the element formed and the third (R or E) indicating whether the activation has been done by reactor neutrons (without an absorber) or by epithermal neutrons (with a cadmium absorber). Thus, for example, SAR means the specific activity of ^{28}Al produced from silicon by reactor neutrons. The activities measured per gram of sample are also represented by three letters, the first one is always C, and the second and the third have the same meanings as for specific activities. If the concentrations of Mg, Al, Si and P are given by f_{Mg}, f_{Al}, f_{Si} and f_P in weight fractions, then the four appropriate equations are represented by Equations (65–68):

$$CMR = MMRf_{Mg} + AMRf_{Al} + SMRf_{Si} \quad (65)$$

$$CME = MMEf_{Mg} + AMEf_{Al} + SMEf_{Si} \quad (66)$$

$$CAR = AARf_{Al} + SARf_{Si} + PARf_P \quad (67)$$

$$CAE = AAEf_{Al} + SAEf_{Si} + PAEf_P \quad (68)$$

5.8 Instrumental Neutron Activation Analysis with Emissions Other than γ Emission

As written earlier, in most delayed activation analysis only the γ emission is measured, since it is difficult to distinguish between β^- emitters, owing to their extended spectra rather than monoenergetic emission, X-ray emission is less penetrating and neutron emission is very rare. However, all these emissions can be used in special cases. For β^-, it was shown that both for biological samples[66] and most geological samples,[67] owing to the higher concentration of P the main relatively long-lived β^- emitter is ^{32}P, and after 10 days more than 95% of the β^- emission is due to ^{32}P, and hence the total β^- measurement can be used for the determination of P.

5.8.1 X-Rays[68]

X-rays suffer from two disadvantages, which can be turned to advantage in special cases. First, the range of X-rays is fairly short and only those emitted from the first few micrometers of the sample can be detected. This is disadvantageous when the whole bulk concentration is needed. However, it is advantageous when only the first few micrometers are important and when in them the concentrations of the measured elements are higher. Second, fewer nuclides emit X-rays (mainly after EC or internal conversion) than those emitting γ-rays. This is disadvantageous since fewer elements can be determined in this method. However, if the major elements of the sample, which interfere with the measurement of the trace elements, do not emit X-rays, the few elements which emit X-rays can be determined even at low concentration owing to the lower background.

Owing to the low penetration of X-rays, the samples used for X-ray measurements should be very thin. The X-rays can be due to internal conversion (X-rays of the Z of the IRN), EC (X-rays of $Z-1$ element) or internal conversion after a β decay ($Z+1$ element). In γ-ray spectrometry the electrons emitted by the active sample are stopped by a thin Perspex or aluminum foil as the electrons emitted by the active sample are less penetrating than γ-rays. In the case of low-energy X-rays, they are less penetrating than some of the electrons and the electrons should be removed by a magnetic field.[69–71] X-ray measurements have been used mainly for the determination of rare earths in geological samples,[72–74] copper[75] and bromine[76] in biological and geological samples, but also other elements, such as Nb in steel[77] and Co and Hg in Dead Sea water,[78] have been determined.

5.8.2 Neutrons[79]

Delayed neutrons are emitted from only very few radionuclides. In many cases the radionuclide decays by both β^- and neutron emission. For example, ^{87}Br decays 2.8% by neutron emission while ^{88}Br decays 6.6% by neutron emission and ^{88}Se only 0.8%. ^{90}Br decays 25% by neutron emission. At least 57 radionuclides which emit neutrons are known, but none of them is

formed by an (n,γ) reaction since the nuclides should be more than one neutron further from the line of stability in order to emit neutrons. However, those radionuclides can be formed in a neutron-induced fission of fissile nuclides such as ^{235}U, since the heavy elements have a higher percentage of neutrons in the nucleus. This is the reason why activation analysis followed by measurement of neutrons is used for the determination of only U and Th. Since other elements do not produce neutron emitters, the background from other elements is very small. The neutrons are measured (after their moderation in a paraffin layer) by BF_3 counters. These are gaseous ionization chambers which measure the α-particles formed by the ^{10}B(n,α)^{7}Li reaction. Outside α-particles are not able to penetrate the counter walls. At least 57 delayed neutron emitters are formed in the neutron-induced fission with half-lives ranging from 0.08 to 55.6 s. In order to get rid of the ^{17}N (4.16 s) and to reduce the error due to wrong time determination, which is more prominent for short-lived radionuclides, the determination of the emitted neutrons is done only for those which have half-lives longer than about 10 s, by delaying the counting for 30–60 s.[80,81]

5.9 Depth Profiling by Instrumental Neutron Activation Analysis

Depth profiling involves the measurement of the concentration of a trace element as a function of its distance from the surface of the sample. This can be done by a destructive method in which the whole sample is activated and then divided into a number of sections of varying distances from the surface, either by gradual slow dissolution or by gradual milling and grinding (for each layer, its thickness and γ or β activity of the various elements are measured). Many samples are too important to be destroyed in the measurement. Using (n,α) or (n,p) reactions, the depth profiling can be done nondestructively. The emitted charged particle (α or p) has a definite kinetic energy. This particle, while transversing in the material, from its place of formation inside the sample to the detector positioned in front of the sample, loses part of its energy, depending on the distance it traveled in the sample. From the measured energy, the distance from the surface can be calculated. Thus, the energy spectrum measured for this particle is related to the concentration depth profile of the element, which is responsible for this nuclear reaction.

In order to allow the measurement of low concentrations, the nuclear reaction should have a high cross-section of the order of at least 10 b. These high cross-sections rule out most elements and only very few elements can be determined.

The method was developed in 1972 by Ziegler et al.[82] for the determination of boron due to the high cross-section of the ^{10}B(n,α)^{7}Li ($\sigma = 3837$ b). This cross-section is five orders of magnitude larger than that of almost any other stable nuclide for (n,α) reactions (except ^{6}Li). The measured sample is placed in a vacuum chamber (pressure $< 5 \times 10^{-6}$ Torr). They needed only about 10^{11} neutrons to detect a concentration of 100 ppm of boron in a 2-cm Si wafer. The ^{7}Li is produced in both the ground state (6%) and in the 479-keV excited state (94%). The energy is distributed between the α-particle and the ^{7}Li according to their masses (1471 and 839 keV for ^{4}He and ^{7}Li in the reaction leading to the excited states). The bombarding thermal neutron passes through the Si wafer (170 µm thick) with negligible attenuation, so that the reaction efficiency is independent of the depth. The neutron beam was collimated to produce a spot 1 cm radius at the target chamber (neutron flux 2.3×10^8 n cm^{-2} s^{-1}). The targets were mounted at a 45° angle to the neutron beam. The α-particles were detected by a surface barrier detector subtending an angle of 19 msr and was specially fabricated to have a low γ and electron background and to be able to measure the spectrum of the α-particles in the presence of ^{7}Li, although this led to poorer energy resolution for the α-particles [full width at half-maximum (fwhm) $\approx$ 20 keV instead of 12–14 keV]. The depth of the boron is determined by the energy loss of the 1471-keV α-particles using the known specific energy loss, dE/dx. An energy, E_f, parallel to depth is given by Equation (69):

$$\text{depth} = \int_{1471}^{E_f} \left(\frac{dE}{dx}\right)^{-1} dE \qquad (69)$$

Ziegler et al.[82] measured the energy spectra of 10^{15}–10^{16} ^{10}B atoms per square centimeter and calculated a concentration resolution of 3 ppm and a depth resolution of 20 nm. Müller et al.[83] used the high-flux reactor of ILL, Grenoble, France, with a collimated flux of 1.04×10^9 n cm^{-2} s^{-1} and 1.5 cm diameter, a detector with resolution of better than 16 keV for the 1471-keV peak. For a counting time of 1000 min, 10^{11} ^{10}B atoms per square centimeter will lead to about 110 integral counts. This is enough for measuring total boron concentration but not for profile determination and they estimated that for profile determination the minimum required concentration is 10^{12} ^{10}B atoms per square centimeter. They measured the concentration depth profiles of boron in silicon wafers implanted with 10^{12}, 10^{13} and 10^{14} ^{10}B atoms per square centimeter and the accuracy of the depth determination was about 25 nm. It should be remembered that this is for ^{10}B atoms whereas they constitute only 20% of the natural boron atoms. Kvitek et al.[84] studied the boron concentration profile in Si using the spectra of both

α-particles and ^{7}Li particles. As can be expected from the higher dE/dx, the spatial resolution obtained with the 839-keV ^{7}Li particle is better than that obtained by using the 1431-keV ^{4}He line (by about 30%). However, in a later work, they explained[85] that in spite of the better resolution of the ^{7}Li profiling, the fact that the ^{7}Li line is located in the high-background region of the measured spectra often makes the profile analysis by this method difficult. In order to obtain better resolution they suggested measuring the α-spectra with different tilting angles.

Biersack et al.[86] discussed the achievable depth resolution and the main sources of uncertainties. They pointed out four main inherent uncertainties: (1) the uncertainty in the angle θ between the point of reaction and point of detection as the particle travel distance L, $L = x/\cos\theta$, where x is the depth of the target atom; (2) the uncertainty in the energy resolution of the detector; (3) the uncertainty in energy loss due to energy loss straggling, assuming a fixed pathlength, L; and (4) uncertainty due to multiple angular scattering. They concluded that for (n,α) reactions, a depth resolution of 5 nm is feasible, and 10-nm depth resolution is feasible for (n,p) reactions.

Apart from B, the only element which can be determined in this way using the natural abundance is lithium with 7.4% ^{7}Li, having a cross-section of 940 b for the reaction ^{6}Li(n,α) T, with an α energy of 2055 keV and a T energy of 2728 keV. Helium can be determined not for natural He but for ^{3}He-doped substrates (stable isotope, natural abundance 0.0001%). Similarly, Na and Be can be determined with the radioactive isotopes ^{22}Na and ^{7}Be, respectively, owing to their high cross-section.

Downing et al.[87] gave a detailed description of the neutron depth profiling (NDP) system and summarized the available data in various systems throughout the world. Special attention was paid to minimizing the contamination of the thermal neutron beams with fast neutrons and γ-rays. The amount of the γ-rays and fast neutrons was reduced by filtering the beam through 200 mm of sapphire single crystal and an 80-mm long silicon single crystal, which led to a ratio of cadmium to gold of $>10^{4}$ and a γ intensity of 200 mrad h^{-1}.

The NDP technique is quantitative and has few interferences in silicon-based materials. As a result, it is often used for calibration for other techniques. Other instrumental methods, such as secondary ion mass spectrometry (SIMS), Fourier transform infrared (FTIR) spectrometry, and Auger electron spectroscopy (AES) have greater sensitivity than NDP in most applications and better resolution in some matrices, but each of the other techniques is subject to highly interactive chemically and electronically induced artifacts. NDP can be used to calibrate the other methods and, after calibration, the better resolution of the other method can be used. Rutherford backscattering spectroscopy (RBS) and NDP complement each other as they both nondestructively analyze different compositions: RBS measures the concentration profile of heavy atoms in matrices of light atoms whereas NDP is limited to some light elements.

5.10 Practical Cases of Instrumental Neutron Activation Analysis

In section 5.4 a test case of INAA of Si was given. Si is one of the best matrices for INAA, owing to the low radioactivity induced in it by neutron irradiation, and consequently very low detection limits were obtained. This is not the case for all matrices. The following section gives details of several examples of analyses of various matrices by INAA.

Greenberg et al.[88] described the procedures adopted by the NAA group of the National Institute of Standards and Technology (NIST) in the USA. The samples are first packaged in a cleaned container. The container, depending on the type of the sample and the length of irradiation, can be polyethylene bags or vials, sealed by heating, or flame-sealed high-quality quartz (to eliminate as much as possible of the activatable sodium) or aluminum cans. The samples are irradiated in a well-moderated (by D_2O) neutron flux of 10^{14} n cm^{-2} s^{-1} using a pneumatic tube for transferring the sample to the reactor. The length of the irradiation, decay and counting depends on the analyzed elements. For one or a few elements the timings are optimized for these elements. Thus, for the determination of mercury in urine the samples (in quartz ampules) are irradiated for 1 h and counted after decay both of 5 and 20 d in order to measure the Hg content via the formation of both ^{197}Hg ($t_{1/2} = 64.1$ h, $E_\gamma = 77$ and 191 keV) and ^{203}Hg ($t_{1/2} = 46.59$ d, $E_\gamma = 279$ keV). The long decay is used to reduce the activities of Na, Cl and Br.

To measure Cl and Br in various oils, they used short irradiation of a few minutes and measured the activity of ^{38}Cl ($t_{1/2} = 37.2$ min, $E_\gamma = 1642.2$ and 2167.6 keV) after a decay of several minutes. The concentration of Br was determined via ^{82}Br ($t_{1/2} = 35.3$ h, $E_\gamma = 554.3$ and 776.5 keV) after a decay of 24 h. For lower detection limits of bromine, irradiation from several hours up to 24 h was applied.

For a complete multielement analysis (20–40 elements which can be determined by NAA), the elements are divided into three groups according to the half-lives of their IRNs: 2 min –15 h, 0.5–5 d and >5 d.

Some elements appear in more than one group. The usual scheme of irradiations, decays and counting is followed. The same sample can be used for the two irradiations.

Table 2 Elemental concentrations (ppm, unless % is indicated) determined by INAA vs NBS (NIST) certified values

Element	SRM 1648 certified		SRM 1572 certified		Method
	INAA	value	INAA	value	
Ag	6.5 ± 0.5				C
Al(%)	3.5 ± 0.1	3.42 ± 0.11			A
As	119 ± 5	115 ± 10	31 ± 0.3	3.1 ± 0.3	B
Ba	750 ± 60		21.4 ± 1.2		A
Br	515 ± 30		8.2 ± 0.3		A, B
Ca(%)	5.9 ± 0.5		3.19 ± 0.10	3.15 ± 0.10	A
Cd	72 ± 6	75 ± 7			B
Ce	54 ± 3		0.28 ± 0.08		C
Cl(%)	0.46 ± 0.02				A
Co	17.9 ± 0.5		0.02 ± 0.006		C
Cr	410 ± 10	403 ± 12	0.74 ± 0.18	0.8 ± 0.2	C
Cs	3.5 ± 0.2		0.098 ± 0.010		C
Eu	0.81 ± 0.08		0.010 ± 0.003		C
Fe	$39\,200 \pm 800$	$39\,100 \pm 1000$	92 ± 13	90 ± 10	C
Hf	4.3 ± 0.3				C
Hg			0.079 ± 0.016	0.08 ± 0.02	A
I	20 ± 5				A
In	1.00 ± 0.07				A
K(%)	1.00 ± 0.06	1.05 ± 0.01	1.76 ± 0.03	1.82 ± 0.06	A, B
La	43 ± 2		0.19 ± 0.03		B
Mg(%)	0.85 ± 0.08				A
Mn	800 ± 20				A
Na(%)	0.41 ± 0.02	0.425 ± 0.002			B
Rb	53 ± 9		5.0 ± 0.3	4.84 ± 0.06	C
Sb	46 ± 2		0.041 ± 0.010		B, C
Sc	6.3 ± 0.2		0.011 ± 0.001		C
Se	27 ± 2	27 ± 1			C
Sm	4.2 ± 0.4		0.052 ± 0.006		B
Sr			94 ± 3	100 ± 2	C
Th	7.5 ± 0.3		0.015 ± 0.004		C
Ti(%)	0.41 ± 0.02				A
V	133 ± 7	140 ± 3			A
W	4.9 ± 0.6				B
Zn	4800 ± 200	4760 ± 140	28.3 ± 0.8	29 ± 2	C

The accuracy and precision of their determination can be found in Table 2, which gives their results for two NBS SRMs together with the certified values. The uncertainties listed represent the estimated overall analytical uncertainties at the 95% confidence level. SRM 1648 is Urban Dust and SRM 1572 is Citrus Leaves. In the column for the method, A, B and C refer to the previous scheme.

Cunningham and Stroube[89] applied INAA to 240 food composites. They irradiated with a neutron flux of 4.9×10^{13} n cm^{-2} s^{-1}. Two irradiations were done on each sample, one for 15 s and one for 4–5 h. For the short irradiation, a 2-min decay time and a 10-min counting time were used. For long irradiation, the decay times varied from 5 d to 4 weeks and counting times were 6–8 h. In the short irradiation they determined Ca, Cl, Cu, Ga, I, In, K, Mg, Mn, Na, S, Sr, Tl and V and in the long irradiation Ag, As, Au, Ba, Br, Cd, Ce, Co, Cr, Cs, Eu, Fe, Hf, La, Lu, Mo, Rb, Sb, Sc, Se, Sm, Ta, Th, W, Yb and Zn.

Sun and Jervis[90] used a Slowpoke reactor, which is a nuclear reactor specially designed mainly for teaching, with activation analysis and the production of small amounts of radioisotopes in mind. The reactor operates at a low power of 20 kW and has a maximum flux of 10^{12} n cm^{-2} s^{-1} (Sun and Jervis used a flux of 2.5×10^{11} n cm^{-2} s^{-1}). An advantage of this type of reactor is its high stability (1–2% over long periods), which allows the irradiation of standards only every few days and not necessarily with each sample. Sun and Jervis[91] determined 35 elements in rocks and soils using two irradiations and three countings. Al, Ba, Ca, Cl, Dy, I, Mg, Mn, Sr, Ti, U and V were determined by 1-min irradiation followed by 5-min counting after a decay of 10–15 min. Long irradiations of 16 h were followed by 30-min counting after 4–9-d decay (As, Br, K, La, Na,

Sb, Sm, W and Yb) and 100-min counting after a decay of 15–20 d (Ce, Co, Cr, Cs, En, Fe, Hf, Ln, Nd, Rb, Sb, Sc, Se, Sn, Ta, Tb, Th, Yb and Zn). Large relative standard deviations between several samples were found for Cl (32%), Sb (24%), Ti (18%), Ba (16%) and Br (14%), but for most elements the average relative standard deviation was <10%. A similar scheme was used by the same authors for samples of coal and ashes, except that the short irradiation time for coal samples was increased to 5 min.

Iskander[92] measured the concentrations of 28 elements in tobacco and cigarette paper. Samples were irradiated for two irradiation periods, the first for 3 min and the second for 8 h with a neutron flux of 2×10^{12} n cm^{-2} s^{-1} from a TRIGA Mark I reactor. The sample was irradiated for 3 min and was counted for 3 min after a 1-min decay (to measure Al, V, Ti, Ca and Mg) and after an additional decay of 26 min (total decay time 30 min) was counted again for 1000 s (to determine Cl and Mn). Also, the long-irradiated samples were counted twice: 4000 s after a decay of 12 h (to determine K, Na, As, Br and La) and 40 000 s after a decay of 21 d (Ba, Rb, Th, Cr, Ce, Hf, Fe, Sb, Sr, Ni, Sc, Se, Zn, Cs, Co and Eu).

McOrist et al.[93,94] studied trace element concentrations in various Australian opals. They irradiated their samples twice, for 1 min at a thermal flux of 2.5×10^{13} n cm^{-2} s^{-1} and for 18 h at a thermal flux of 5×10^{12} n cm^{-2} s^{-1}. Each irradiated sample was counted twice. The short-irradiated samples were counted for 10 min after a 20-min decay and for 45 min after a decay of 24 h. After the long irradiation, the samples were counted for 2 and 4 h after 7 and 28 d of decay, respectively.

Wallner and Katzlberger[95] studied the concentrations of 14 elements in spruce tree rings and needles from an area where an Al refinery used to be. Al, Mn, Ca and Cl were determined by 3–7 min of irradiation at a thermal neutron flux of 5×10^{10} n cm^{-2} s^{-1}. They calculated that for Al, assuming that the main interference for its 1779-keV line is from the Compton-scattered photons of the

Table 3 INAA timings for various elements in biological samples

Element	IRN	γ-ray of IRN	Half-life (keV) of IRN	Timings system	Interfering reactions
Na	24Na	15.02 h	1369	2, 3	24Mg(n,p), 27Al(n,a)
Mg	27Mg	9.46 m	1014	1	27Al(n,p)
Al	28Al	2.24 m	1779	1	28Si(n,p), 31P(n,a)
Si	31Si	2.62 h	1266	2	31P(n,p), 34S(n,a)
S	37S	5.0 min	3103	1	
Cl	38Cl	37.3 min	1642	1, 2	
K	42K	12.36 h	1525	2	42Ca(n,p)
Ca	49Ca	8.72 m	3084	1	
Sc	46Sc	83.8 d	889	4	46Ti(n,p)
V	52V	3.76 m	1434	1	
Cr	51Cr	27.70 d	320	4	54Fe(n,a)
Mn	56Mn	2.58 h	847	2	56Fe(n,p)
Fe	59Fe	44.6 d	1099	4	
Co	60Co	5.27 y	1173	4	63Cu(n,a)
Ni	58Co	70.8 d	811	4	
Cu	66Cu	5.10 m	1039	1	66Zn(n,p)
Zn	65Zn	244.1 d	1116	4	
As	76As	26.3 h	559	3	79Br(n,a)
Se	75Se	118.5 d	265	4	
Br	82Br	35.34 h	776	3	
Rb	86Rb	18.8 d	1077	4	
Sr	87mSr	2.80 h	388	2	
Mo	99Mo-99mTc	66.02 h	141	3	
Cd	115Cd-115In	53.4 h	336	3	
Sb	122Sb	2.68 d	564	3	
Sb	124Sb	60.20 d	1691	4	
I	128I	24.99 m	443	2	
Cs	134Cs	2.06 y	796	4	
Ba	139Ba	82.9 min	166	2	
Hg	203Hg	46.8 d	279	4	

The timing systems are:	system:	1	2	3	4
	Time of irradiation:	1–10 min	2–20 min	10 h	10 h
	Time of counting:	1–2 min	1 h	5–8 d	30 d
	Time of decay:	2–5 min	10 min	1–2 h	3–5 h

Table 4 Detection limits found for INAA of various biological samples

Biological sample	Irradiation time[a]	Decay time	Counting time	Detection limits (ppb)
1. Human blood[b]	5 d	7 mon	2 h	Co(4), Cs(4), Fe(10,000), Se(20)
	24 h, E	21 d	2 h	Br(1000), Fe(1500), Rb(300), Se(100), Zn(500)
2. Hair[c]	30 s	10 s	20 s	Ag(140), Cl(13,000), F(23,000), Se(160)
	10 m	5 m	5 m	Al(1600), Ba(7600), Ca(52,000), Cu(3,800), I(260) K(16,000), Mn(750), Na(4,100), S(0.28%), Zn(3,500)
	16 h	2 d	50 m	As(46), Au(2), Sb(45)
	16 h	21 d	50 m	Co(97), Hg(200), Ag(140), Ba(7,600), Se(160), Zn(3,500)
3. Meat, fish and poultry[d]	15 s	2 m	10 m	Ca(82,000), Cl(2,400), Cu(8,300), K(0.1%), Mg(0.013%), Mn(1,900), Na(260), V(81)
	4 h	5–28d	7 h	Ag(120), As(650), Br(130), Cd(170), Co(23) Cr(350), Cs(39), Eu(3.7), Fe(6,800), Rb(430), Sb(32), Sc(1.4), Se(300), Zn(460)

[a] d means days, h means hours, s = seconds, m = minutes, mon = months, E means irradiation with epithermal neutrons.
[b] F. Chisela and P. Bratter, *Anal. Chim. Acta.*, **188**, 85 (1986).
[c] A. Chatt, M. Sayjad, K.N. De Silva and C.A. Secord in *IAEA Health-related Monitoring of Trace Element Polluants using Nuclear Techniques*, IAEA TECDOC-330, IAEA, Vienna 1985, p. 33.
[d] W.C. Cunningham and W.B. Stroube, Jr, *Sci. Tot. Environ.* **63**, 29 (1987).

2113-keV line of ^{56}Mn, the optimum irradiation time is 4 min. However, for samples with a low Ca content longer irradiations up to 7 min were used. Na, K, Rb, Ba, Zn, Fe, Sc, Cr, Co and Br were determined by 10 h of irradiation with a thermal neutron flux of 8×10^{13} n cm^{-2} s^{-1}. The samples were counted three times, after a decay of 4–5 d (Na and K), a decay of 7 d (Br and Ca) and a decay of 14 d (Ba, Co, Cr, Fe, Rb, Sc and Zn).

Oddone et al.[96] studied the elemental content of Anatolian obsidians, volcanic glasses, by INAA using both thermal and epithermal neutrons. Mg, K, Ca, Ti, Mn, As and Dy were determined by 2 min of irradiation at a thermal neutron flux of 1×10^{12} n cm^{-2} s^{-1} in a TRIGA Mark II reactor, 250 kW. The sample was counted six times after decays of 0.5, 10, 30, 120, 480 and 960 min. The longer the decay time, the longer was the counting. Al and Si were determined by repeating the short-time irradiation for a sample wrapped in a 10-mm thick Cd foil. Both Al and Si were determined from the 1779-keV line of ^{28}Al, using the epithermal neutron activation reaction ^{28}Si(n,p)^{28}Al. Na, Ca, Sc, Cr, Fe, Co, Zn, Se, Rb, Zr, Nb, Sb, Ba, La, Ce, Nd, Sm, Eu, Gd, Tb, Ho, Tm, Yb, Lu, Hf, Ta and Th were determined by irradiation for 20 h. The samples were counted five times after decays of 3, 6, 24, 48 and 96 d. Epithermal neutron INAA was applied also with long irradiation (using samples inside a 1.5-mm thick Cd box) in order to reduce the interferences from elements with low resonance integrals and high thermal cross-section (small I_0/σ ratio). The samples were irradiated for 40 h and counted five times after decays of 3, 6, 12, 24 and 48 d. The epithermal neutron INAA improved the sensitivity and accuracy of the measurement of Zn, Se, Rb, Zr, Nb, Sb, Cs, Ba, Gd, Tb, Ho, Tm, Yb, Hf, Ta and Th. Ni, Sr and U, which could not be measured by thermal INAA, could be determined by the epithermal INAA.

Sato[97] suggested four different schemes of timings for the determination of various elements in biological samples (three times of irradiation with the last one being counted twice, after different decay times). His suggestion for the various elements is given in Table 3. Table 4 gives three examples from the literature, for the irradiation of biological samples together with their timings and the limits of detection of some elements via INAA.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
BGO	Bismuth Germanate
DE	Double-escape
EC	Electron Capture
ENAA	Epithermal Neutron Activation Analysis
FTIR	Fourier Transform Infrared
fwhm	full width at half-maximum
HFR	High Flux Reactor
INAA	Instrumental Neutron Activation Analysis
IRN	Indicator Radionuclide
LWR	Light Water Reactor
MCA	Multichannel Analyzer
NAA	Neutron Activation Analysis
NBS	National Bureau of Standards
NDP	Neutron Depth Profiling
NIST	National Institute of Standards and Technology
RBS	Rutherford Backscattering Spectroscopy
RNAA	Reactor Neutron Activation Analysis

SE	Single-escape
SIMS	Secondary Ion Mass Spectrometry
SRM	Standard Reference Material

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis

Nuclear Methods **(Volume 14)**
Instrumental Neutron Activation Analysis: Gamma Lines Table • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides

REFERENCES

1. K.S. Krane, *Introductory Nuclear Physics*, John Wiley, New York, 1988.
2. B.G. Harvey, *Introduction to Nuclear Physics and Chemistry*, 2nd edition, Prentice Hall, Englewood Cliffs, NJ, 1969.
3. G. Friedlander, J.W. Kennedy, E.S. Macias, J.M. Miller, *Nuclear and Radiochemistry*, 3rd edition, John Wiley, New York, 1981.
4. G.R. Choppin, J. Rydberg, *Nuclear Chemistry*, Pergamon Press, Oxford, 1980.
5. W.D. Ehmann, D.E. Vance, *Radiochemistry and Nuclear Methods of Analysis*, John Wiley, New York, 1991.
6. A. Vertes, I. Kiss, *Nuclear Chemistry*, Elsevier, Amsterdam, 1987.
7. R.B. Firestone, V.S. Shirley, C.M. Baglin, Y.S. Chu, J. Zipkin, *Table of Isotopes*, 8th edition, John Wiley, New York, 1996.
8. G. Erdtmann, W. Soyka, *The Gamma Rays of the Radionuclides*, Verlag Chemie, Weinheim, 1979.
9. U. Reus, W. Westmeier, *At. Data Nucl. Data Tables*, **29**, 1–192, 193–406 (1983).
10. E. Browne, R.B. Firestone, V.S. Shirley, *Table of Radioactive Isotopes*, John Wiley, New York, 1986.
11. Z.B. Alfassi, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 35, Vol. II, 1990.
12. H.P. Yule, in *Nondestructive Activation Analysis*, ed. S. Amiel, Elsevier, Amsterdam, 113, 1981.
13. M.D. Glascock, *Tables for Neutron Activation Analysis*, 4th edition, University of Missouri, Columbia, MO, 1996.
14. K. Heydorn, *Radiochim. Acta*, **3**, 161 (1964).
15. Z.B. Alfassi, B. Rietz, *Analyst*, **119**, 2407 (1994).
16. O.U. Anders, *Anal. Chem.*, **32**, 1368 (1960); **33**, 1706 (1961).
17. Z.B. Alfassi, N. Lavi, *Nucl. Instrum. Meth. Phys. Res. A*, **365**, 190 (1995).
18. L.I. Zhuk, A.A. Kist, I.N. Mikholskaya, N.S. Osinskaya, T. Tilayev, S.I. Tursunbayev, S.V. Agzamova, *J. Radioanal. Nucl. Chem.*, **120**, 369 (1988).
19. R.G.V. Hancock, M.D. Grynpas, B. Alpert, *J. Radioanal. Nucl. Chem.*, **110**, 283 (1987).
20. W.C. Cunningham, W.B. Stroube, Jr, *Sci. Total Environ.*, **63**, 29 (1987).
21. R. Dams, J.A. Robbins, K.A. Rahn, J.W. Winchester, *Anal. Chem.*, **42**, 861 (1970).
22. A. Wyttenbach, S. Bajo, L. Tobler, *J. Radioanal. Nucl. Chem.*, **114**, 137 (1987).
23. A. Simons, S. Landsberger, *J. Radioanal. Nucl. Chem.*, **110**, 555 (1985).
24. F. De Corte, A. Speecke, J. Hoste, *J. Radioanal. Chem.*, **3**, 205 (1969).
25. A. Simonits, F. De Corte, J. Hoste, *J. Radioanal. Chem.*, **31**, 467 (1976).
26. F. De Corte, S. Jovanovic, A. Simonits, L. Moens, J. Hoste, *Atomkernenergie, Suppl.*, **44**, 641 (1984).
27. Z.B. Alfassi, N. Lavi, *J. Radioanal. Nucl. Chem.*, **150**, 183 (1991).
28. M.D. Glascock, W.Z. Tian, W.D. Ehmann, *J. Radioanal. Nucl. Chem.*, **92**, 379 (1985).
29. Z.B. Alfassi, N. Lavi, *Analyst*, **109**, 959 (1984).
30. S. Landsberger, A. Arendt, B. Keck, M. Glascock, *Trans. Am. Nucl. Soc.*, **56**, 230 (1988).
31. B. Kratochvil, N. Motkosky, M.S.M. Duke, D. Ng, *Can. J. Chem.*, **65**, 1047 (1987).
32. X. Lin, K. Heydorn, B. Rietz, *J. Radioanal. Nucl. Chem.*, **160**, 85 (1992).
33. Z.B. Alfassi, T.U. Probst, B. Rietz, *Anal. Chim. Acta*, **360**, 243 (1998).
34. M.J.J. Theunissen, H.J.J. Jaspers, J.M.G. Hansen, M.L. Verheijke, *J. Radioanal. Nucl. Chem.*, **113**, 391 (1987).
35. M.L. Verheijke, H.J.J. Jaspers, J.M.G. Hansen, M.J.J. Theunissen, *J. Radioanal. Nucl. Chem.*, **113**, 397 (1987).
36. R.M. Lindstrom, in *Microelectronic Processing: Inorganic Materials Characterization*, ed. L.A. Casper, AES Symposium Series, No. 295, American Chemical Society, Washington, DC, 294, 1986.
37. G. Revel, N. Deschamps, C. Dardenne, J.L. Pastol, B. Hania, H. Dinh Ngugen, *J. Radioanal. Nucl. Chem.*, **85**, 137 (1984).
38. G. Revel, in *Euroanalysis VII Conference Proceedings 1987*, ed. E. Roth, Editions Physiques, Paris, 167, 1988.
39. K. Fujingawa, K. Kudo, *J. Radioanal. Chem.*, **52**, 411 (1979).
40. S. Niese, *J. Radioanal. Chem.*, **38**, 37 (1977).
41. E.W. Haas, R. Hofmann, F. Richter, *J. Radioanal. Chem.*, **69**, 219 (1982).
42. H. Jaskolska, L. Rowlinska, *J. Radioanal. Chem.*, **26**, 31 (1975).

43. Z.B. Alfassi, M.H. Yang, *J. Radioanal. Nucl. Chem.*, **132**, 99 (1989).
44. N. Lavi, Z.B. Alfassi, *Analyst*, **109**, 361 (1984); Z. Etzion, Z. Alfassi, N. Lavi, R. Yagil, *Biol. Trace Elem. Res.*, **12**, 411 (1987).
45. D. Brune, K. Jirlow, *Nukleonik*, **6**, 242 (1964).
46. E.S. Gladney, D.R. Perrin, J.P. Balagna, C.L. Warner, *Anal. Chem.*, **52**, 2128 (1980).
47. S.J. Parry, *J. Radioanal. Chem.*, **72**, 195 (1982); *J. Radioanal. Nucl. Chem.*, **81**, 143 (1984).
48. Z.B. Alfassi, *J. Radioanal. Nucl. Chem.*, **90**, 151 (1985).
49. S.J. Parry, *J. Radioanal. Chem.*, **59**, 423 (1980).
50. L.A. Currie, *Anal. Chem.*, **40**, 587 (1986).
51. H. Bem, D.E. Ryan, *Anal. Chim. Acta*, **124**, 373 (1981).
52. W.Z. Tian, W.D. Ehmann, *J. Radioanal. Nucl. Chem.*, **84**, 89 (1984).
53. F. Rossitto, M. Terrani, S. Terrani, *Nucl. Instrum. Methods*, **103**, 77 (1972).
54. T.G. Williamson, P.E. Penneche, B. Hostica, J.S. Brenizer, T.L. Ngugen, *J. Radioanal. Nucl. Chem.*, **114**, 387 (1987).
55. R.K. Tokay, M. Skalberg, G. Skarnemak, *J. Radioanal. Nucl. Chem.*, **96**, 265 (1985).
56. R.J. Rosenberg, *J. Radioanal. Chem.*, **62**, 145 (1981).
57. W.B. Stroube, Jr, W.C. Cunningham, G.J. Lutz, *J. Radioanal. Nucl. Chem.*, **112**, 341 (1987).
58. W.D. Ehmann, J. Brückner, D.M. McKown, *J. Radioanal. Chem.*, **57**, 491 (1980).
59. F. Chisela, D. Gawlik, P. Bratter, *J. Radioanal. Nucl. Chem.*, **98**, 133 (1987).
60. D.C. Stuart, D.E. Ryan, *Can. J. Chem.*, **59**, 1470 (1981).
61. H. Soreq, H.C. Griffin, *J. Radioanal. Chem.*, **79**, 135 (1983).
62. Z.B. Alfassi, N. Lavi, *J. Radioanal. Nucl. Chem.*, **84**, 363 (1984).
63. R.G.V. Hancock, *J. Radioanal. Chem.*, **69**, 313 (1982).
64. M. Ördögh, E. Orban, G. Miskovitz, J. Appel, E. Szabo, *Int. J. Appl. Radiat. Isot.*, **25**, 61 (1974).
65. Z.B. Alfassi, N. Lavi, *Analyst*, **109**, 959 (1984).
66. E. Steinnes, *Talanta*, **14**, 753 (1967).
67. E. Steinnes, *Anal. Chim. Acta*, **57**, 457 (1971).
68. M. Mantel, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 111, Vol. 1, 1990.
69. M. Mantel, S. Amiel, Z.B. Alfassi, *Anal. Chem.*, **50**, 441 (1978).
70. Z.B. Alfassi, T. Biran-Izak, M. Mantel, *Nucl. Instrum. Methods*, **151**, 227 (1978).
71. L.B. Graman, *J. Radioanal. Nucl. Chem.*, **99**, 75 (1986).
72. J. Hertogen, R. Gijbels, *Anal. Chim. Acta*, **56**, 61 (1971).
73. M. Mantel, S. Amiel, *J. Radioanal. Chem.*, **16**, 127 (1973).
74. P.A. Baedecker, J.J. Rowe, E. Steinnes, *J. Radioanal. Chem.*, **40**, 115 (1977).
75. D.R. Allen, W. Brookhart, *Anal. Chem.*, **46**, 1297 (1974).
76. M.S. Rapaport, M. Mantel, R. Nothman, *Anal. Chem.*, **51**, 1356 (1979).
77. M. Mantel, C. Shenberg, M.S. Rapaport, *J. Radioanal. Chem.*, **75**, 145 (1982).
78. M. Stiller, M. Mantel, M.S. Rapaport, *J. Radioanal. Nucl. Chem.*, **83**, 345 (1984).
79. Z.B. Alfassi, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 97, Vol. 1, 1990.
80. S. Amiel, *Anal. Chem.*, **34**, 1683 (1963).
81. C. Shenberg, Y. Nirel, Z. Alfassi, Y. Shiloni, *J. Radioanal. Nucl. Chem.*, **114**, 207 (1988).
82. J.F. Ziegler, G.W. Cole, J.E.E. Baglin, *J. Appl. Phys.*, **43**, 3809 (1972).
83. K. Müller, R. Henkelmann, H. Borofka, *Nucl. Instrum. Methods*, **129**, 557 (1979).
84. J. Kvitek, V. Hnatowicz, P. Kotas, *Radiochem. Radioanal. Lett.*, **24**, 205 (1976).
85. J. Cerevna, V. Hnatowicz, J. Hoffmann, Z. Kosina, J. Kvitek, P. Onheiser, *Nucl. Instrum. Methods*, **188**, 185 (1981).
86. J.P. Biersack, D. Fink, R. Henkelmann, K. Müller, *Nucl. Instrum. Methods*, **149**, 93 (1978).
87. R.G. Downing, R.F. Fleming, J.K. Langland, D.H. Vincent, *Nucl. Instrum. Methods*, **218**, 47 (1983).
88. R.R. Greenberg, R.F. Fleming, R. Zeisler, *Environ. Int.*, **10**, 129 (1984).
89. C.C. Cunningham, W.B. Stroube, Jr, *Sci. Total Environ.*, **63**, 29 (1987).
90. J.X. Sun, R.E. Jervis, *J. Radioanal. Nucl. Chem.*, **89**, 553 (1985).
91. J.X. Sun, R.E. Jervis, *J. Radioanal. Nucl. Chem.*, **114**, 89 (1987).
92. F.Y. Iskander, *J. Radioanal. Nucl. Chem.*, **89**, 511 (1985).
93. G.D. McOrist, A. Smallwood, J.J. Fardy, *J. Radioanal. Nucl. Chem.*, **185**, 293 (1994).
94. G.D. McOrist, A.S. Smallwood, *J. Radioanal. Nucl. Chem.*, **223**, 9 (1997).
95. G. Wallner, Ch. Katzlberger, *J. Radioanal. Nucl. Chem.*, **223**, 139 (1997).
96. M. Oddone, Z. Yegingil, G. Bigazzi, T. Ercan, M. Ozdogan, *J. Radioanal. Nucl. Chem.*, **224**, 27 (1997).
97. T. Sato, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 323–358, Vol. II, 1990.

Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Paul A. Ullucci
ESA Laboratories, Inc., Chelmsford, USA

1 Introduction 2
2 Background 2
3 Air Monitoring Versus Biological Monitoring 2
4 American Conference of Government Industrial Hygienists Biological Exposure Indices 3
5 Deutsch Forschungsgemeinschaft Biologischer Arbeitsstoff-Toleranz-Wert Values 4
6 Lead 4
6.1 Biological Monitoring Analytical Methods for Lead 4
6.2 Graphite Furnace Atomic Absorption Spectroscopy 4
6.3 Anodic Stripping Voltammetry 5
6.4 Inductively Coupled Plasma Mass Spectrometry 5
6.5 Other Methods 5
7 Cadmium 5
7.1 Biological Monitoring Analytical Methods for Cadmium 5
7.2 Graphite Furnace Atomic Absorption 5
7.3 Other Methods 6
8 Mercury 6
8.1 Biological Monitoring Analytical Methods for Mercury 6
8.2 Cold Vapor Atomic Absorption 6
9 Chromium 7
9.1 Biological Monitoring Analytical Methods for Chromium 8
10 Cobalt 8
10.1 Biological Monitoring Methods for Cobalt 8
11 Arsenic 8
11.1 Biological Monitoring for Arsenic 8
11.2 Direct Hydride Reduction Methods 9
11.3 Hydride Generation/Low Temperature Separation 10
11.4 Graphite Furnace Methods 10
11.5 High-performance Liquid Chromatography/Hydride Generation/Atomic Absorption 10
11.6 High-performance Liquid Chromatography/Hydride Generation/Atomic Fluorescence 10
11.7 High-performance Liquid Chromatography/Inductively Coupled Plasma Atomic Emission Spectroscopy 10
11.8 High-performance Liquid Chromatography/Inductively Coupled Plasma Mass Spectrometry 10
11.9 Dietary Arsenic: Arsenosugars 10
12 Multiple Metals Methods 11
13 Other Metals 11
14 Sample Collection 11
14.1 Introduction 11
14.2 Urine Collection 12
14.3 Blood Samples 12
15 Quality Control/Quality Assurance 13
16 The Future 15
Abbreviations and Acronyms 15
Related Articles 16
Further Reading 16
Basic Biological Monitoring Principles 16
Applications and Analytical Methods 16
Reviews 16
References 16

Biological monitoring uses sophisticated analytical techniques for the accurate and precise determination of metals in blood and urine. This provides occupational health professionals with the means to monitor worker exposure to toxic metals. For metals such as lead, cadmium, and mercury, good correlations have been found between the level of the metal in blood and/or urine and health impairment.

Analytical techniques such as graphite furnace atomic absorption spectroscopy (GFAAS) and inductively coupled plasma mass spectrometry (ICP/MS) have been widely used to determine metals in biological samples. Since metals may exist in different forms in blood and urine, chromatographic separation combined with spectroscopic detection has been used to determine metal species. Biological monitoring is an example of how modern analytical chemistry has provided the tools needed to protect the health of workers.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Exposure to metals occurs in a number of industries from the mining, roasting, and smelting of ores, to steel and metal alloy production, to the manufacture of batteries and silicon chips. The industries may be very large, such as factories producing thousands of lead batteries each day, or a small one-room electroplating shop, or a dentist's surgery filling a tooth with a mercury amalgam. The literature is replete with examples of toxic exposures in the above-mentioned cases (and many others). Unfortunately, the toxic exposure is not limited to workers using metals in industry but can occur in homeowners growing vegetables in a backyard garden or in a child playing in his room. The easily identifiable villain in these two cases is lead, a metal now ubiquitous in our environment. In fact, many of the metals of concern in industrial exposure, arsenic, cadmium, chromium, mercury, selenium, and vanadium, are also of concern in the environment owing to the activities of humankind throughout the course of human history.

Biological monitoring provides health care professionals with an important tool that can be used to assess metal exposure.[1] Advances in analytical chemistry have played a major role in the identification and quantification of metals in environmental and biological samples. Detection limits/quantitation limits continue to improve along with the accuracy and the reliability of the measurements. These improvements allow us not only to better assess the exposure of workers in a manufacturing plant, but also to assess the exposure of those individuals that live in the vicinity of that plant.

2 BACKGROUND

Biological monitoring can be defined as the measurement of chemical markers in biological media (blood, urine, expired air) that are indicative of external exposure to chemical agents.[2] The chemical marker may be the originally absorbed chemical, (e.g. Pb in the case of a battery worker), a metabolite of the absorbed chemical (most common in the case of organic solvent exposure), or a biochemical response induced by the chemical, (e.g. zinc protoporphyrin (ZPP)).[3] The first two cases are referred to as biological markers of exposure, and the latter is a biological marker of effect. Ideally, for every biomarker of exposure there should be a biomarker of effect. Two excellent examples of this are the two metals for which the United States Occupational Safety and Health Administration (OSHA) has mandated biological monitoring. The OSHA lead Standard (became law in 1979) requires blood-lead determinations as well as ZPP determinations.[4] ZPP is a biomarker of effect and is an indicator of the interference by lead in the biosynthesis of heme. It is used to monitor the biochemical response in workers where the exposure is usually chronic in nature.

The other metal for which OSHA has established biological monitoring guidelines is cadmium.[5] Blood and urine cadmium are the markers of exposure and β-2-microglobulin (β-2M) is the marker of effect. In the USA, lead and cadmium are the only cases where biological monitoring is mandated by law. Biological monitoring for lead is required by many countries throughout the world including Germany, Great Britain, Sweden, Finland, Denmark, Norway, France, and Australia. Germany has standards set for aluminum, mercury (inorganic and organic), and vanadium. The World Health Organization (WHO) has set standards for mercury in blood and urine and cadmium in urine.

3 AIR MONITORING VERSUS BIOLOGICAL MONITORING

Air sampling has been and continues to be the most common method of monitoring for chemical exposure in the workplace. In fact, air sampling has become so prominent that it is synonymous with the field of industrial hygiene. There are a number of advantages to air monitoring:

1. Air monitoring is a long-standing tradition (the first air exposure limits were set in the 1940s).
2. A large number of air-monitoring standards which have the force of law have been established.
3. Generally, there is worker acceptance of the method.
4. There are rugged analytical methods available from a number of qualified laboratories.

There are a number of significant weaknesses in using air monitoring as the only mechanism for monitoring worker exposure.[6] These weaknesses include the inability to account for dermal and oral exposure, the inability to account for individual variation in absorption of airborne contaminants, the inability to monitor the effectiveness of personal protective equipment (PPE), and the inability to assess exposure outside the workplace. Biological monitoring has the ability to monitor worker exposure via all routes of exposure including air, dermal, and oral. For many chemicals dermal absorption can be a significant route of exposure which is completely overlooked if air monitoring alone is used. Chemicals like aromatic amines (aniline, methylenedianiline), chlorophenol, dimethylformamide, most pesticides,

and even some metals can pose a serious dermal absorption problem. OSHA is now trying to address this problem by having inspectors monitor surface contamination by taking wipe samples.

Another route often ignored is the oral route of exposure. This route can be significant for metals when workers are in a dusty environment. Contamination of the hands during work followed by eating and smoking can be a source of oral exposure. A study of cadmium workers whose hands were checked for Cd just before entering the lunch room showed significant Cd contamination. Biological monitoring is a measure of total exposure; all forms of absorption will be accounted for. Because biological monitoring uses the individual as the dosimeter of exposure, it reflects individual differences in exposure, absorption, distribution, and/or metabolism.[7] There are significant differences in the absorption of chemicals by different people as well as by the same person under different conditions. The rate of breathing has a dramatic effect on the absorption. The harder a person works the more air he or she takes in and the more contaminant is absorbed. From light work, to moderate work, to heavy work a significant increase in absorption occurs. PPE is often the only barrier between the worker and serious health damage. However, does the worker make use of PPE all the time? Does the PPE fit properly? Is the PPE working properly, the right glove? Biological monitoring can answer these questions. Most important, there is a closer relationship between biological monitoring and health outcome than there is with external dose markers such as air monitoring.

There are some disadvantages to biological monitoring. Biological monitoring is often an invasive method, especially when the collection of blood is required. Even when biological monitoring involves the collection of urine only, the workers view this as an invasion of privacy and are concerned that the sample will be used for drug testing. The relationship between exposure to toxins in the air and the presence of toxins in the body is not fully appreciated. When all of these issues are addressed beforehand by the health professional, many of these problems disappear, and informed workers may demand this type of monitoring as protective of their health.

The last impediment to the use of biological monitoring involves the lack of biological monitoring standards. As mentioned, in the USA OSHA has mandated only two standards. The apparent reluctance of OSHA to create biological monitoring Standards was addressed in a recent paper by Rappaport.[8] In this paper, the author points out two problem areas where biological monitoring could provide solutions. One controversial area is associated with the dose rate which is involved in setting OSHA STELs (short-term exposure limits). Using biological monitoring to study the exposure–biomarker relationship could help solve this issue. The other problem area has to do with exposure to chemicals where accumulation occurs over months of exposure. This is the case for lead. Blood lead is known to be a better indicator of exposure and disease than air lead. This argument could be extended to almost every other metal for which accumulation occurs, as well as a number of organic compounds which are long-term exposure hazards. Adding to this our ever-growing knowledge of the effects of even low-level metal exposures, the need and value of biological monitoring will become more fully appreciated.

4 AMERICAN CONFERENCE OF GOVERNMENT INDUSTRIAL HYGIENISTS BIOLOGICAL EXPOSURE INDICES™

The American Conference of Government Industrial Hygienists (ACGIH) has established a number of biological exposure indices (BEI™). The BEIs™ are reference values of biological determinants. They are:

> Reference values intended as guidelines for the evaluation of potential health hazards in the practice of industrial hygiene. BEIs™ represent the levels of determinants which are most likely to be observed in specimens collected from a healthy worker who has been exposed to chemicals to the same extent as a worker with inhalation exposure to the TLV™ (threshold limit value™).[9]

The TLVs™ are air monitoring standards established by ACGIH. The values represent the maximum airborne exposure to which a worker can be exposed for 8 h day^{-1}, 50 weeks per year, without experiencing any loss of health and/or well-being.

BEIs™ provide a means of assessing exposure from all sources. They give an individual dimension to air sampling and they can be used to determine the effectiveness of PPE. BEIs™ are based on human data from both experimental and field studies. They take into account the relationship between external and internal doses at the TLV™, as well as the relationship between internal dose and reversible health effects.[9]

To date the ACGIH has proposed BEIs™ for the metals arsenic and soluble compounds including arsine, cadmium and inorganic compounds, chromium water-soluble fumes, cobalt, lead, mercury, and vanadium pentoxide. The following metals are under study for the establishment of BEIs™: aluminum, organic lead, manganese, nickel, and selenium.

5 DEUTSCH FORSCHUNGSGEMEINSCHAFT BIOLOGISCHER ARBEITSSTOFF-TOLERANZ-WERT VALUES

The German government established the first biological monitoring standards in 1981 with the creation of BAT values. The BAT value ''Biologischer Arbeitsstoff-Toleranz-Wert'' (biological tolerance value for occupational exposures) is defined as the maximum permissible quantity of a chemical substance or its metabolites or the maximum permissible deviation from the norm of biological parameters induced by these substances in exposed humans. The BAT value is established on the basis of currently available scientific data and reflects concentrations that generally do not adversely affect the health of the worker. They are for an exposure of $8\,h\,day^{-1}$, 40 h per week.[10,11] BAT values have been established for the following metals: aluminum, cadmium, lead, mercury (inorganic and organic), and vanadium pentoxide. BAT values are not established for substances which are known to be carcinogenic because it is not possible to specify safe levels for such compounds. For these compounds correlations between the concentration of a substance in the air of the work area and concentrations of the substance or its metabolites in biological media have been established. The following metals are included in this list: alkali chromates, arsenic trioxide, cobalt, and nickel.

6 LEAD

Monitoring exposure to inorganic lead in the workplace involves the determination of lead in whole blood as well as ZPP. ZPP is a biomarker of effect of lead on heme synthesis in the body. The OSHA limits of lead and ZPP are: Pb $< 50\,\mu g\,dL^{-1}$, ZPP $< 60\,\mu g\,dL^{-1}$. The ACGIH BEI™ for lead is $30\,\mu g\,dL^{-1}$. The BAT for lead is $70\,\mu g\,dL^{-1}$ ($30\,\mu g\,dL^{-1}$, women < 45 years). Determination of lead in urine is not a good indicator of inorganic lead exposure and is seldom used except to monitor chelation therapy. Urinary lead is useful if the exposure is due to organolead compounds. There is a BAT for organolead compounds in urine.

Table 1 OSHA-approved blood lead proficiency testing programs

CDC and Prevention
CAP
New York State Department of Health
Pennsylvania Department of Health
Wisconsin State Laboratory of Hygiene
Puerto Rico Department of Health
Centre de Toxicologie du Quebec

6.1 Biological Monitoring Analytical Methods for Lead

The determination of lead in blood is the most frequent biological monitoring analysis performed. A good way to determine the analytical methods that are used for blood lead is to refer to the various blood lead proficiency testing programs (see Table 1) that are available. Laboratories are required to participate in one or more of these programs in order to be certified able to perform this analysis. Participant reports from the various programs include information about the blood-lead method used by each laboratory. These programs include: Wisconsin/Centers for Disease Control (CDC), the College of American Pathologists (CAP), the State of New York, and le Centre de Toxicologie du Quebec. By far the most common method used is GFAAS followed by anodic stripping voltammetry (ASV). Other methods include ICP/MS, chelation extraction/flame atomic absorption spectroscopy (FAAS), and Delves Cup.

6.2 Graphite Furnace Atomic Absorption Spectroscopy

GFAAS methods involve the dilution of blood with a matrix modifier solution containing a surfactant (typically Triton X-100), ammonium phosphate, and nitric acid.[12,14–18] Instrumental requirements include automated sample injection, atomization using either ''wall''[13] or stabilized temperature platform (also called L'vov platform)[12,14–18], and most important, a background correction method. For blood-lead deuterium background correction,[12,14,16,17] the Smith–Hieftje background correction[18] or the Zeeman background correction[12,13,15–17,19] is used. Newer instrumentation using transverse Zeeman and longitudinal heating are also being used.[15] Graphite furnace conditions will vary from instrument to instrument but all involve a drying step to remove water, followed by an ashing step to destroy as much organic matter as possible (the phosphate modifier allows the destruction of the blood matrix without loss of the analyte, while nitric acid would facilitate the destruction by oxidation), followed by atomization. A final high-temperature ''cleaning burn'' completes the cycle. An important aspect of the blood-lead determination involves calibration: aqueous standards versus matrix matched standards. Direct aqueous standardization has been used by a number of authors for calibration.[15,16,18] Others have found aqueous standards unacceptable and have relied on matrix-matched standards for calibration.[12–14,17,19] The ability to use aqueous standards seems to be dependent on the design of the furnace system and, especially, the graphite tube design. It is this author's experience that matrix-matched standards provide the best accuracy for blood-lead determinations. The National Institute of Standards and Technology (NIST) has a standard reference material

for lead in blood (SRM 955b). The SRM contains blood standards at four different levels which can be used as primary standards or controls.

A number of variations in the GFAAS determination of Pb in blood have occurred which are worthy of note. Jacobson et al. did a protein precipitation step prior to furnace analysis which seemed to improve accuracy.[13] D'Haese et al. used a molybdenum-treated L'vov platform to reduce matrix and spectral interferences when using deuterium background correction.[14]

Modern graphite furnace systems completely automate the analysis process using a computer to control the automatic calibration step, the quality control (QC) checks, and the calculation of the final result. All of the foregoing are capable of being imported directly into a Laboratory Information Management System (LIMS). QC checks can initiate automatic recalibration and sample reanalysis should a control fall out of the acceptable range.

6.3 Anodic Stripping Voltammetry

ASV has been used for blood determinations since the early 1970s and was the first micromethod for blood lead. The technique involves the dilution of blood with a special reagent to free the lead from the red cell, the plating (reduction reaction) of the lead onto a mercury-coated graphite electrode, and the stripping (oxidation reaction) of the lead from the electrode. The potential at which plating and stripping occurs is specific for lead, whereas the current generated is proportional to the amount of lead.[20,21]

Recently, a hand-held battery-operated instrument was approved by the US Food and Drug Administration (FDA) for the determination of lead in blood from finger stick samples. The method involves dilution of the blood with a special reagent and then placing an aliquot of the sample onto the electrode. The cycle of plating and stripping occurs and a final answer is provided. The instrument was designed for the doctors' surgery and for the screening of subjects in the field. Correlation with graphite furnace and classical ASV methods was excellent.[22]

6.4 Inductively Coupled Plasma Mass Spectrometry

ICP/MS has been applied to the determination of lead in blood. The high solids content of blood can cause significant matrix problems, but the greatly improved detection limits of ICP/MS make this an ideal method of measuring lead in blood by simple dilution. Various diluents have been used including nitric acid, ammonia, and tetramethylammonium hydroxide followed by direct analysis.[23–26] Isotope dilution ICP/MS, which involves adding a known amount of a rare isotope of lead to the blood sample before sample treatment, has become the reference method for lead in blood. NIST SRM 955b (lead in blood) uses isotope dilution ICP/MS to establish the true value. The method of Paschal et al. involves microwave digestion of the blood with added isotope using nitric acid followed by direct ICP/MS analysis.[27] Stroh used simple dilution with an ammonia/Triton X-100 solution for the isotope dilution ICP/MS determination of Pb.[28] Another interesting use of the isotope dilution method and measuring several isotopes of lead simultaneously is the fingerprint capability. Lead from different sources exhibits different isotope ratios. These ratios can be used to pinpoint the source of the lead. This procedure has been used in cases of childhood lead poisoning where the source of the contamination was in doubt.[29]

6.5 Other Methods

Older methods involving chelation extraction FAAS[30,31] and Delves Cup[32] are rarely used being replaced by GFAAS and ASV. Other interesting blood-lead methods include capacitively coupled microwave plasma atomic emission[33] and electron-capture negative chemical ionization gas chromatography/mass spectrometry (GC/MS).[34]

7 CADMIUM

The OSHA Cadmium Standard requires monitoring for Cd in blood and urine. Cd in blood reflects recent exposure while Cd in urine reflects long-term exposure. In addition to the measurement of Cd, β-2M, a marker of renal tubular damage, is also required. The OSHA Cd Standard is blood Cd, $<5\,\mu g\,L^{-1}$, urine Cd, $<3\,\mu g\,g^{-1}$ creatinine, and β-2M, $<300\,\mu g\,g^{-1}$ creatinine. The ACGIH BEI™ for Cd is blood Cd, $5\,\mu g\,L^{-1}$; urine, $5\,\mu g\,g^{-1}$ creatinine. The BAT for Cd is $15\,\mu g\,L^{-1}$ for blood and urine.

7.1 Biological Monitoring Analytical Methods for Cadmium

For the OSHA Cadmium Standard, two proficiency testing programs (see Table 2) have been established: the CAP and le Centre de Toxicologie du Quebec. Methods used for Cd are GFAAS, ICP/MS, and extraction/FAAS (two of 50).

7.2 Graphite Furnace Atomic Absorption

By far the most common methods of determining cadmium in blood and urine involve GFAAS. The

Table 2 Proficiency testing programs for cadmium

CAP	
	Blood cadmium Urine cadmium Urine β-2M Urine creatinine
	Three samples, bimonthly OSHA and CLIA HCFA[a]
Centre de Toxicologie du Quebec	
	Blood cadmium Urine cadmium Urine β-2M
	Three samples, bimonthly OSHA

[a] CLIA is the Clinical Laboratory Improvement Act. HCFA is the Health Care Financing Administration.

determination of Cd in urine involves simple dilution of the sample with dilute nitric acid followed by GFAAS without a matrix modifier[35,36] or with a matrix modifier.[14,37,38] Without the use of a matrix modifier, the ashing/charring temperature must be carefully controlled (maximum of 400 °C) or loss of Cd will result.[35] Use of an ammonium phosphate modifier minimizes this problem while at the same time reducing the background signal. Most methods for Cd in urine use an ammonium phosphate modifier.[37,38] This includes the OSHA Cadmium Standard recommended method.[37] Calibration for the nonmodifier procedures involve the method of standard addition or, more commonly, preparation of urine-based standards.[35,36] Employing a matrix modifier allows the use of aqueous standards for calibration.[37,38] Chelation/extraction methods have also been used for the determination of Cd in urine. In this method, urinary Cd is chelated with hexamethylene ammonium/hexamethylene dithiocarbamidate, extracted into a mixture of diisopropyl ketone and xylene, and determined by GFAAS.[39] Aqueous standards prepared in glutathione and chelated/extracted like the samples are used for calibration.

For cadmium in blood, the procedures fall into two major categories: protein precipitation with nitric acid followed by direct GFAAS without matrix modifier[36,40,41] or direct analysis with matrix modifier.[14,42–44] The OSHA-recommended method uses protein precipitation (blood) with nitric acid and direct analysis without a matrix modifier.[36] The major disadvantages of the precipitation method are the additional sample handling required and the increased possibility of contamination. It is for this reason that many laboratories prefer to use the simple dilution and matrix modification method.[14,42–44] Standardization of Cd in blood is similar to that for urine. Nonmodifier methods use standard additions or standards prepared in blood, while modifier methods use direct aqueous calibrators in most cases, although one author recommended matrix-matched standards for blood Cd using deuterium background correction.[14]

7.3 Other Methods

Employing a column preconcentration technique and a slotted quartz tube mounted on the burner, Karakaya et al. successfully determined Cd in blood and urine by FAAS. The method had a detection limit of $1\,\mu g\,L^{-1}$.[45]

Many of the ICP/MS methods mentioned earlier for the determination of lead in blood and urine are also applicable to Cd in blood and urine. The methods involve simple dilution with various diluents and ICP/MS or isotope dilution ICP/MS determination.[19,31,32,46]

Cd and Pb in urine were determined by differential pulse ASV following high-pressure digestion of urine with a detection limit of $1\,\mu g\,L^{-1}$.[47]

8 MERCURY

The ACGIH BEI™ for total inorganic mercury in urine is $35\,\mu g\,g^{-1}$ creatinine for preshift samples. The BEI™ for total inorganic mercury in blood is $15\,\mu g\,L^{-1}$ for samples collected at the end of shift at the end of the work week. The BAT for total inorganic mercury in urine is $200\,\mu g\,L^{-1}$; blood is $50\,\mu g\,L^{-1}$. There is no BEI™ for organic mercury exposure, for example alkyl mercury compounds; however, there is a BAT for organic mercury, which is $100\,\mu g\,L^{-1}$. The organomercury compounds are associated with the red cells and, therefore, must be monitored in whole blood.

8.1 Biological Monitoring Analytical Methods for Mercury

The proficiency testing program of le Centre de Toxicologie du Quebec includes both blood and urine mercury (see Table 3). Methods used by participating laboratories include cold vapor atomic absorption with acid digestion or selective reduction, cold vapor with gold film concentration atomic absorption spectroscopy (AAS), and ICP/MS.

8.2 Cold Vapor Atomic Absorption

While the volatility of elemental mercury is the major reason for the potentially easy absorption of this metal, this same characteristic provides a unique method for its measurement. Most methods for mercury involve

Table 3 Proficiency testing programs for trace metals

CAP
Plasma/serum control
Aluminum
Chromium
Copper
Selenium
Zinc
Three samples, three times per year
Centre de Toxicologie du Quebec
Aluminum, plasma
Arsenic, urine, total and inorganic species
Cadmium, blood and urine
Chromium, plasma
Copper, plasma and urine
Fluoride, urine
Lead, blood
Mercury, blood and urine
Selenium, plasma and urine
Zinc, plasma
Three samples, bimonthly

chemical generation and measurement of mercury vapor by atomic absorption without flame or heat of any kind, thus the name cold vapor atomic absorption spectroscopy (CVAAS). Sample preparation prior to the CVAAS determination falls into two major categories, digestion and direct reduction. One common digestion procedure involves digestion of the sample with sulfuric acid and potassium permanganate followed by reduction with stannous chloride.[48]

The US Environmental Protection Agency (EPA) Method 245.1 (manual method) and 245.2 (automated method) have been modified for the determination of Hg in urine. Digestion is achieved with a combination of nitric acid, sulfuric acid, potassium permanganate, and potassium persulfate at a temperature of 95 °C. Reduction with stannous chloride follows.[49] The method can be used to determine Hg in a variety of biological samples including blood, urine, hair, and tissue. A flow injection procedure with on-line oxidation of the organic mercury species has been described.[50] This method can also be used to distinguish between organic and inorganic mercury. For Hg in urine, destruction of urine with nitric acid alone just prior to stannous chloride reduction can also be used.[51,52] While these methods are commonly used, a major interference in methods which use acid digestion and/or acidic reduction methods has been noted. Iodine (I^-) interferes with the digestion/reduction method and can cause total suppression of the Hg signal.[53] This effect is not observed in the nondigestion alkaline reduction methods.

The selective reduction methods (blood and urine) use a stannous chloride–cadmium chloride reductant under alkaline conditions to produce Hg vapor from both inorganic and organic mercury yielding total mercury.[54–56] Inorganic Hg can be determined using a reductant of stannous chloride alone, also under alkaline conditions.[54–56] Sodium borohydride under alkaline conditions has been used as a reductant for total Hg in blood.[57]

For increased sensitivity a commercially available amalgamation device coupled to a hydride generation device using acidic borohydride reduction has been used for Hg in blood and urine. The Hg vapor from the reduction reaction is trapped onto gold/platinum gauze to form a Hg amalgam. The amalgam is heated to release the Hg vapor.[58] A highly sensitive adaptation of the amalgam approach using a gold-coated quartz sand trap has been reported.[59] In this method biological samples are digested in sealed Teflon vessels prior to reduction with stannous chloride and concentration on the gold-coated sand trap. The trap is then removed from the apparatus for later analysis. The trap is then placed in a special heating unit to release the trapped Hg and to determine it by atomic fluorescence. The detection limit is 50 pmol L^{-1}.

ICP/MS has also been successfully used for Hg in blood by simple dilution of the sample with tetramethylammonium hydroxide.[25]

Alkyl Hg compounds like methyl mercury and ethyl mercury are far more toxic than elemental and inorganic mercury compounds. Thus, it is important to be able to differentiate between the various chemical forms of mercury. A number of novel approaches have been used. Chromatographic separation of Hg^{2+}, CH_3Hg^+, and $C_2H_5Hg^+$ by reversed phase liquid chromatography followed by borohydride reduction/CVAAS has been developed.[60] An interesting procedure for the determination of inorganic mercury and monomethyl mercury in a variety of biological samples uses alkaline digestion followed by an aqueous-phase ethylation reaction with sodium tetraethylborate. The ethyl derivatives are precollected on a trapping column. The trapping column is thermally desorbed onto a packed GC (gas chromatography) column for separation. The compounds are decomposed to Hg^0 with detection by CVAAS.[61]

The applicability of capillary electrophoresis (CE) to mercury speciation has been described.[62] Hg^{2+}, CH_3Hg^+, and $C_2H_5Hg^+$, were separated by CE and detected amperometrically at the parts per billion level.

9 CHROMIUM

The ACGIH BEI™ for chromium(VI), water-soluble fume is 10 μg g^{-1} creatinine for the increase during shift,

and $30\,\mu g\,g^{-1}$ creatinine, for end of shift at the end of the work week. Because it is considered a carcinogenic substance, there is no BAT for Cr(VI). However, the relationship between air levels and the corresponding level of Cr in whole blood and in urine is given. For example, at an air exposure level of $0.05\,mg\,m^{-3}$ (the ACGIH TLV™) the corresponding level of Cr in whole blood is $17\,\mu g\,L^{-1}$ and $20\,\mu g\,L^{-1}$ in urine.

9.1 Biological Monitoring Analytical Methods for Chromium

The Quebec proficiency testing program includes urine chromium as one of the analytes. Of the 36 laboratories reporting values for chromium, all of them used GFAAS with either Zeeman or deuterium background correction. (Note that CAP includes serum Cr in their trace metal program).

Sample preparation for the determination of Cr in urine involves simple dilution of the sample with a dilute nitric acid/Triton X-100 solution.[63–67] For Cr in whole blood, simple dilution is again the method of choice.[63,65] Because of the very low volatility of Cr, an ashing temperature of 1200 °C is used and a matrix modifier is not necessary. Zeeman or deuterium background correction have been found to be effective. Granadillo et al. did an extensive study including sample preparation methods, deuterium and Zeeman correction, pyrolysis/atomization temperature, and wall versus platform atomization and found that no background correction is necessary for Cr in a variety of biological samples.[63] The method of calibration is a major source of variation for urinary and blood Cr determinations. There are three major methods, aqueous standards,[63,67] the method of standard addition,[64,65] and matrix-matched standards.[66] This author prefers the last method which involves preparation of standards in a urine matrix. The only difficulty in this method is to obtain a urine sample with low or nondetectable levels of Cr. Chromium is an ubiquitous metal and great care must be exercised to minimize contamination in every step of the analytical procedure from sample collection to final analysis.

10 COBALT

The ACGIH BEI™ for cobalt in urine is $15\,\mu g\,L^{-1}$, the BEI™ for blood is $1\,\mu g\,L^{-1}$ for samples collected at the end of shift, at the end of the work week. As with chromium, there is a relationship between air exposure and a biological level for Co and no BAT. At an air level of $0.05\,mg\,m^{-3}$, a whole blood Co of $2.5\,\mu g\,L^{-1}$ and a urine level of $30\,\mu g\,L^{-1}$ are noted.

10.1 Biological Monitoring Methods for Cobalt

Methods for the determination of Co in urine use GFAAS with either direct analysis of diluted urine[68,69] or GFAAS determination following Co chelation and extraction.[70,71] Although not as sensitive as the chelation/extraction method, direct analysis has the advantage of simplicity, and most important, a decrease in potential for contamination from reagents used in sample preparation.

The determination of Co in blood involves direct analysis of diluted whole blood.[72] Because the difficulties of this determination include an increased detection limit compared with urine, the WHO recommends urine Co as the method of choice for Co monitoring.[73]

11 ARSENIC

The BEI™ for inorganic arsenic and metabolites in urine is $50\,\mu g\,g^{-1}$ creatinine for samples collected at the end of the work week. The As BEI™ will change to $35\,\mu g\,L^{-1}$ in 1999. There is no BAT for As because it is considered to be a carcinogen like Cr and Co. Urinary levels of $50\,\mu g\,L^{-1}$ are observed after exposure to As at $0.01\,mg\,m^{-3}$.

11.1 Biological Monitoring for Arsenic

Of all of the metals, As is certainly the most interesting, most complicated, and least understood. The metabolism of As is unique among all metals in proceeding by a stepwise methylation reaction.[74,75] An unusual aspect of As metabolism is the fact that no animal model exists which can be used to study human metabolism. A wide variation among species exists ranging from certain animals where the methylation reaction is nonexistent to species where methylation is the major route and everything in between. In humans the metabolism of inorganic As has been studied extensively and proceeds by Scheme 1:

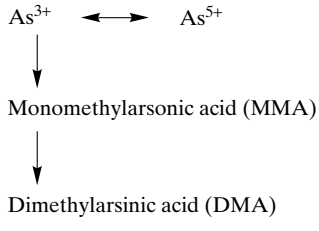

Scheme 1

This process apparently takes place in the liver and has always been considered to be a detoxification mechanism. Recent studies dispute this long-held belief.[76,77] Arsenic

Table 4 Proportions of arsenic metabolites being secreted

Metabolite	Controls ($n = 9$)(%)	Workers ($n = 34$)(%)
As^{3+}, As^{5+}	5.9 ± 5.2	23.3 ± 10.3
MMA	9.2 ± 8.4	16.9 ± 5.2
DMA	83.9 ± 10.8	59.9 ± 10.8

has been shown to cause both DNA hypomethylation in rat liver epithelial cells and DNA hypermethylation in the human p53 gene.

Proportions of each metabolite excreted for control subjects and exposed workers are shown in Table 4.[78]

Methylation efficiency decreases at very high dose levels. The time course of an As^{3+} ingestion begins with the excretion of mainly inorganic arsenic for the first 0–12 h. This rapidly changes to a preponderance of dimethylarsinic acid (DMA) with DMA comprising 70–80% of the As excreted 72–96 h postingestion. In an attempted suicide at very high As^{3+} levels, 85% of the excreted As was in the inorganic form (As^{3+} and As^{5+}) at 1 day postingestion, 93% at 2 days postingestion, 63% at 3 days postingestion, 49% at 4 days postingestion, and 14% at 8 days postingestion. The corresponding DMA was 0% at 1 day, 1% at 2 days, 35% at 3 and 4 days, and 69% at 8 days postingestion.[78]

Because of the good correlation between the level of exposure and urinary excretion (45–60% of the dose is excreted in 2 days), routine determination of As in urine is the best approach for monitoring occupational exposure to inorganic As. Measurement of As in urine is complicated, however, by the presence of high concentrations of As in a number of seafood products. These organoarsenicals (mainly arsenobetaine and arsenocholine) are relatively nontoxic and are rapidly excreted unchanged. The most common methods for measuring As in urine involve a vigorous acid digestion procedure prior to hydride generation atomic absorption or direct graphite furnace analysis. These methods would measure dietary arsenic as well as inorganic As from occupational exposure. Following a single seafood meal, total As levels as high as 1000 $\mu g\,L^{-1}$ have been observed. Thus measurement of total As in urine as means of assessing As exposure in the presence of dietary As is meaningless. Fortunately, the problem has been overcome by the development of analytical methods which can differentiate between As from the diet and As from occupational exposure. Many different approaches have been used, from direct hydride generation techniques to the extreme of hyphenated technology, high-performance liquid chromatography (HPLC)/hydride generation/photochemical reactor/ICP/AES (inductively coupled plasma atomic emission spectroscopy).

11.2 Direct Hydride Reduction Methods

Hydride generation atomic absorption is an ideal method for determining As and inorganic As metabolites. This is due to the fact that As^{3+}, As^{5+}, monomethylarsonic acid (MMA), and DMA all readily produce gaseous hydrides. However, the organic forms of As present in seafood (arsenobetaine and arsenocholine) do not form hydrides directly. Hydride generation AAS provides a method of distinguishing between the inorganic and organic forms. The method involves generation of hydrogen from acid and sodium borohydride which reacts with As^{3+} to form arsine (AsH_3), with MMA to produce methylarsine, and with DMA to produce dimethylarsine. The arsine gas formed is directed to a flame-heated quartz tube placed in the light beam of the atomic absorption spectrophotometer. The heat converts AsH_3 to As^0 for subsequent AAS detection.

The formation of the gaseous hydride for each of the As species depends on a number of factors: the acid used, the concentration of the acid, the concentration of the borohydride, the concentration of sodium hydroxide used to prepare and stabilize the borohydride solution, and the design of the hydride generator. The last point is most interesting. There is wide variation in the ability of commercial hydride generation systems to form, for example, arsine from As^{5+}. Continuous flow systems that mix the sample, the acid, and the borohydride in a small chamber just prior to the gas liquid separator (which delivers the arsine gas to the detector and the liquid to waste) discriminate significantly against As^{5+}. Without prereduction of As^{5+} to As^{3+}, no signal will be observed for As^{5+}. Nonflow systems that allow the reactants to mix together for a period of time prior to release of the gas will give a significant but not equal signal for As^{5+} compared with As^{3+}. This is due to reduction of the +5 form to the +3 form prior to formation of the hydride. Flow-through systems which include mixing coils of various lengths placed before the gas/liquid separator can improve the response for +5. One group of researchers has studied this reaction in detail with the result that selective reduction can be used to measure many of the forms independently of each other.[79,80]

In an occupational monitoring setting, the total of the inorganic species, As^{3+}, As^{5+}, MMA, DMA is most important. By adjusting acid concentrations and borohydride concentrations it is possible to obtain an equal response from each of the four As species. This method has been used for the determination of inorganic As and metabolites in urine.[81–84] Ion-exchange chromatography using disposable solid-phase extraction columns has been used to separate the inorganic species from the organic prior to hydride generation.[85] In another method using ion-exchange chromatography, $As^{3+,5+}$, MMA,

and DMA were separated and analyzed individually by hydride generation atomic absorption.[86]

11.3 Hydride Generation/Low Temperature Separation

Hydride generation/liquid nitrogen trapping/thermal separation is one of the earliest methods known to determine the various As species. In this method, the various forms of As are reduced to their hydrides and trapped in a liquid nitrogen cold trap. The trap is allowed slowly to return to room temperature during which time arsine and the methylarsines distill off one at a time. Direct current discharge atomic emission and AAS have been used for detection.[87–89]

11.4 Graphite Furnace Methods

GFAAS has been used for the determination of As in urine. Direct analysis of urine would give total As and be subject to the difficulties mentioned earlier. However, there are cases when total As in urine may be of interest.[90] Combining GFAAS with a column chromatographic separation of inorganic and organic forms prior to furnace determination has been used.[91,92] The cation-exchange chromatography step separates As^{3+}, As^{5+}, MMA, and DMA from arsenobetaine and other dietary organoarsenicals. Another approach is the combination of solvent extraction and GFAAS.[93,94] The method involves extraction of As species from acidified urine following reduction with potassium iodide. Only the As^{3+}, As^{5+}, MMA, and DMA are extracted under these conditions allowing elimination of the other organo arsenicals.

11.5 High-performance Liquid Chromatography/ Hydride Generation/Atomic Absorption

In this procedure the As species are separated by HPLC, and the effluent from the column is directed into a continuous hydride generation system with on-line atomic absorption detection. Separation of the As species typically involves anion-exchange chromatography.[95–97] Reversed phase ion-pair chromatographic separation on a C_{18} column has also been used.[98] Novel modifications of this method have appeared. Using low-power ultraviolet irradiation after column separation and before hydride generation allows the determination of a number of As species including arsenobetaine and other organoarsenicals.[99] The organic As compounds were also separated and determined using thermochemical hydride generation.[100]

11.6 High-performance Liquid Chromatography/ Hydride Generation/Atomic Fluorescence

HPLC/hydride generation with atomic fluorescence detection has also been applied to As speciation.[101] This method has been improved using short-column liquid chromatography to allow separation of the four As species in 3 min.[102]

11.7 High-performance Liquid Chromatography/ Inductively Coupled Plasma Atomic Emission Spectroscopy

Direct coupling of HPLC to ICP/AES has been used for As speciation.[103] HPLC separation by anion- and cation-exchange chromatography was compared. Ion-pairing reversed phase separation has also been used.[104] A column-switching valve system was used combining both anion-exchange and reversed phase separation prior to ICP/AES detection.[105] An interesting system adds hydride generation and a photoreactor to ICP/AES to improve detection limits.[106] In this system the effluent from the column is mixed with photooxidation chemicals, followed by a specially designed photoreactor, followed by the hydride generation system, followed by ICP/AES detection. The system allows detection of the four inorganic As species as well as arsenobetaine and arsenocholine.

11.8 High-performance Liquid Chromatography/ Inductively Coupled Plasma Mass Spectrometry

The direct coupling of HPLC and ICP/MS has been used to determine various As compounds in urine.[107] By combining anion-exchange and cation-exchange chromatography, the four anionic As species as well as the four cationic species were detected. This arrangement also eliminated the argon chloride interference common in ICP/MS.[108] Microbore HPLC and microscale flow injection coupled to ICP/MS allowed sub-parts per billion determination of As species.[109]

11.9 Dietary Arsenic: Arsenosugars

Unfortunately, the contribution of dietary As cannot be completely eliminated even when As speciation methods are employed. Certain marine animals especially crustacea have the ability to convert organoarsenicals to DMA. The contribution is small (10%) and not significant in occupational exposure but would be of importance when assessing low-level environmental exposure. More significant is the contribution due to ingestion of certain forms of seaweed which have been found to contain high levels of arsenosugars as well as much smaller amounts of DMA and As^{5+}. Arsenic species in various marine organisms have been determined using the techniques mentioned above: arsenosugars and inorganic As compounds by HPLC/ICP/MS.[110] HPLC/ICP/MS has been used to study urinary As excretion following ingestion of seaweed, crab, and shrimp.[111] In the case

of crab and shrimp virtually all of the As excreted was arsenobetaine. However, in the case of seaweed significant levels of DMA, As^{5+}, and three unidentified metabolites were observed. In a recent paper using ion-pair liquid chromatography with hydride generation and atomic fluorescence detection, a substantial increase in urinary DMA was observed following ingestion of seaweed.[112] The authors point out that the commonly used biomarkers of exposure to inorganic As may not be reliable when arsenosugars are ingested.

12 MULTIPLE METALS METHODS

A number of methods for the determination of multiple metals in biological samples have been published. Most of these methods involve ICP/AES or ICP/MS as the detection sources. Two National Institute for Occupational Safety and Health (NIOSH) methods have been published using ICP/AES. One method (NIOSH 8310) involves the determination of 16 metals in urine using a polydithiocarbamate resin extraction technique.[113] The detection limit is about $1\,\mu g\,L^{-1}$. For the determination of 20 different metals in whole blood or tissue, NIOSH Method 8005 uses a digestion of blood prior to ICP/AES determination.[114] The detection limit is approximately $10\,\mu g\,L^{-1}$ for each metal.

As mentioned earlier, the great sensitivity of ICP/MS coupled with multielement capacity makes this an ideal method for the determination of metals in blood or urine. Review articles by Houk on ICP/MS[115] and, especially, Durrant[116] on ICP/MS in biological analyses provide a detailed look at this most promising technique. Methods for the determination of Cd, Pb, and Hg in whole blood,[25] as well as Al, As, Cd, Pb, and Zn[26] by simple dilution were previously mentioned.

Many other methods have been published using ICP/MS. Only a few of them will be mentioned here: Fe, Co, Cu, Zn, Rb, Mo, and Ce were determined in human serum by a method that involved nitric acid dilution.[117] Interferences from polyatomic ions were corrected by a matrix blank solution. Direct determination after simple dilution of urine with nitric acid was studied by Mulligan et al.[118] A number of metals were studied for detection limit, linearity, and interferences. Detection limits were around $1\,\mu g\,L^{-1}$ for most metals. The method was compared to established methods for Sb, Cd, and Hg with good results. A number of interferences were noted in the 10 to 80 mass range. Good stability was observed during continuous working day operation. Vaughan et al. looked at a ten-fold dilution of biological samples and were able to detect 13 metals reliably in serum, 15 metals in whole blood, and 25 elements in saliva.[119]

Mass overlap was a problem for some metals (Cr, Mn, and V) as well as matrix-induced effects. Some of the matrix effects were minimized by using matrix-matched standards and the method of standard addition. Hsiung et al. studied in detail interferences encountered when determining metals in biological samples.[120] Sample preparation involved simple dilution in the case of urine and nitric acid protein precipitation for serum. The importance of selecting the right isotope to minimize polyatomic interferences and the use of external and internal calibration to minimize matrix effects were emphasized. A completely automated method that included all aspects of the analysis from sample preparation to final analysis was published by Huang et al.[121] Sample preparation involved flow-through microwave digestion, followed by on-line resin column extraction to remove matrix interferences, followed by column elution and ICP/MS determination. Elements studied included Fe, Ni, Cu, Zn, and Pb.

Simultaneous multielement GFAAS has been used to determine up to four metals in plasma. Al, Cr, Mn, and Ni were determined in serum and plasma using this technique.[122]

ASV has been applied to the simultaneous determination of Zn, Cd, Pb, and Cu in urine[123] and Cu, Pb, and Cd in whole blood.[21]

13 OTHER METALS

Aluminum: there is no BEI™ for aluminum; however, there is a BAT for Al in urine of $200\,\mu g\,L^{-1}$. GFAAS has been used to determine Al urine in both plasma and urine.[124,140]

Manganese: the validity of using either blood or urine monitoring to assess Mn exposure is debatable. However, numerous GFAAS methods have been used to measure Mn in biological samples both from a nutritional and an occupational exposure perspective.[125–128]

Nickel: there is no BEI™ or BAT. At an air exposure level of $0.10\,mg\,m^{-3}$, the observed value for Ni in urine is $15\,\mu g\,L^{-1}$. The most common method for Ni in urine is GFAAS[129,130]

Vanadium: there is a BEI™ for vanadium pentoxide which is total V in urine of $50\,\mu g\,g^{-1}$ creatinine. V in urine can be determined by GFAAS[131,132]

14 SAMPLE COLLECTION

14.1 Introduction

Sample collection is one of the most important issues in any analytical determination. It is critical to any determination involving trace metal analysis because sample

contamination is always a major concern. The subject has been covered in detail by Aitio and Jarvisalo.[133] Sources of potential contamination include the sample collection device (needles for blood collection), sample collection containers, the environment where the sample was collected, and the hands of the person collecting the sample. These considerations, as well as those associated with sample preservation and shipping, must be addressed to insure an unadulterated sample.

14.2 Urine Collection

An accurate 24 h urine sample is the ideal specimen for any determination involving urine. However, this type of sample is not feasible for occupational exposure monitoring. For this reason the ACGIH developed the BEIs™, based upon timed spot-urine collections. Samples are typically collected at preshift, postshift at the end of the day, postshift at the end of the week, and so on. The disadvantage of a spot-urine sample is the variation in dilution that occurs due to the variability of hydration of the worker. The most common way to address this variability is to normalize the urine values by using a creatinine correction for each specimen. Creatinine is a normal endogenous constituent in urine. Specific gravity can also be used but there are limits to the effectiveness of this correction factor. The creatinine level should be within the range of $0.5–3\,g\,L^{-1}$. Outside this range, the concentration of the appropriate analytes will be overcorrected. For instance, in the case of a dilute urine of $0.2\,g\,L^{-1}$, the uncorrected concentration will be multiplied by a factor of five often resulting in a falsely elevated result. To avoid erroneous results, one approach is to check the specific gravity at the site of collection with a urine test strip. If the urine specific gravity is greater than 1.015, the creatinine is usually greater than $0.5\,g\,L^{-1}$. This will avoid inconsistent results and retesting. The calculated value is obtained by dividing the analyte value in milligrams per liter (or micrograms per liter) by the creatinine value which is in grams per liter to obtain a value of milligrams per gram (or micrograms per gram) of creatinine.

Urine samples are collected in plastic cups or bottles that have been acid washed or have been analyzed to insure that they are free of trace metals. Preservation of urine samples for trace metals usually involves acidification with dilute acid. To avoid the possibility of contamination, acidification should never be done in the field. One way to avoid this situation is to send the samples to the laboratory, refrigerated by overnight courier, and to acidify the samples at the laboratory. Another approach is for the laboratory to provide two containers, a wide-mouth acid-washed container for sample collection and a leak-proof transport bottle containing an acid preservative for shipping. Samples should be surrounded by cold packs in a cooler that meets the regulations covering the shipping of biological samples. Samples should be sent to the laboratory by next day or second day delivery.

14.3 Blood Samples

The most convenient way of collecting a blood sample for the determination of trace metals in blood is the use of an evacuated tube, Vacutainer®, Venoject®, and so on. While convenient, these tubes are known to be potential sources of contamination for a number of metals, particularly aluminum and zinc. For metals such as Pb, Cd, Hg, and As, contamination is not an issue. Evacuated tubes containing ethylenediaminetetraacetic acid (EDTA) as an anticoagulant are routinely used, for example Purple top Vacutainer® EDTA binds with the calcium in the blood sample and interferes with the clotting process. This provides a longer anticoagulant action than heparin and is recommended when there is a delay between the sample collection and the analysis. Stability is not an issue for these metals, although samples should always be shipped in insulated containers to avoid temperature extremes that could result in an aesthetically displeasing sample. Note that temperature extremes can have a significant effect on ZPP determinations often collected along with blood lead.

For metals such as Al, Co, Cr, Ni, and Mn extreme caution must be taken to avoid contamination in all aspects of the sample collection process. Glass Vacutainer® tubes should not be used for these metals. Plastic Vacutainer® tubes have been found to be virtually free of metals contamination and are recommended for exposure monitoring for these metals. However, for nutritional monitoring purpose these tubes may not be acceptable for all metals. The reader is referred to the National Committee for Clinical Laboratory Standards (NCCLS) document on this subject.[134] Contamination of the blood sample with metals from stainless steel needles is also an issue of concern.[135] For occupational exposure purposes flushing of the needle by taking a preliminary blood sample before the actual trace metal sample is taken can be effective in limiting contamination from the needle.

For determinations requiring plasma or serum, an additional source of contamination must be considered. After collection in an appropriate tube, the sample is centrifuged and the plasma/serum layer transferred to another tube. This step must be performed using a plastic transfer pipette and a plastic tube known to be metal free in a laboratory environment. It should not be attempted in the field. Glass transfer pipettes and glass tubes should never be used.

All containers, evacuated tubes, transfer pipettes, transport tubes, and so on should be checked on a regular

basis for metal contamination. This is accomplished by analyzing several samples from each manufacturers' lot of materials.

15 QUALITY CONTROL/QUALITY ASSURANCE

QC and quality assurance (QA) are an essential part of any analytical determination no matter what matrix is being analyzed or which chemical is being measured. Every laboratory must have a formal QA/QC program in place to insure the validity of their data. This is true for all laboratories performing biological monitoring analyses. There are a number of preanalytical and analytical parameters involved in every determination. While a presentation on QA/QC is not the purpose here, two preanalytical QA parameters and two analytical QC parameters that affect biological monitoring should be mentioned. The subject of QA/QC in biological monitoring has been dealt with in previous publications.[136,137] Preanalytical parameters include laboratory certification and proficiency testing. All laboratory testing falls under the jurisdiction of some certifying or accrediting organization. The same is true of laboratories performing biological monitoring.

For example, in the USA biological monitoring laboratories fall under the CLIA'88. Any laboratory analyzing biological samples must obtain a CLIA certificate. Laboratories performing biological monitoring analyses will require a high-complexity testing certificate. Laboratories can be accredited by the CAP, by individual states, or by any other accrediting agency approved by CLIA. In addition, laboratories must be

Table 5 Proficiency testing programs: non-US programs

German Society of Occupational Medicine	
Chlorinated solvents in blood	
Six PCB isomers	
Organic solvent metabolites in blood	
Occupational and environmental exposure levels	
Two annual rounds	
Japan Federation of Occupational Health Organizations	
Blood lead	Free erythrocyte protoporphyrin
Urine aminolevulinic acid	Urine hippuric acid
Urine methylhippuric acid	Urine total trichlorocompounds
Urine trichloroacetic acid	Urine mandelic acid
Urine monomethylformamide	Urine 2,5-hexanedione
Annual rounds	
United Kingdom External Quality Assessment Scheme	
Blood lead	Blood cadmium
Biweekly samples	
Robens Institute of Health and Safety[a]	
Serum: copper, zinc, gold, aluminum, selenium	
Blood: lead, cadmium	
Urine: cadmium, mercury	
Monthly sample rounds	
Austria	
Blood: lead, cadmium, chromium, mercury	
Urine: lead, cadmium, chromium, mercury, aminolevulinic acid, fluoride, phenol, trichloroacetic acid, hippuric acid, mandelic acid, *trans*, *trans*-muconic acid, 1-hydroxypyrene	
Finnish Institute of Occupational Health	
Mandelic acid	Phenol Trichloroacetic acid
Methylhippuric acids	2,5-Hexanedione
Four sample rounds per year	

[a] Robens Institute of Industrial and Environmental Health and Safety, University of Surrey, Guildford, Surrey GU2 5XH, UK.

licensed in their individual state or any state where they obtain samples, for example laboratories analyzing samples from New York State must be licensed by New York State. Similar programs exist in other countries.

The second preanalytical parameter is proficiency testing. In many countries laboratories are required to participate in proficiency testing. For example, the OSHA Lead Standard requires laboratories to participate successfully in an approved blood-lead proficiency program. See Table 1 for a list of OSHA approved programs and an OSHA document for approval requirements.[138] Tables 2, 3, and 5 list the various proficiency testing programs available throughout the world for biological monitoring for metals.

The two analytical parameters of great interest to biological monitoring laboratories are calibration standards and QC materials. Calibration Standards are important since quantitative determinations are only as accurate as the Standard on which they are based. In the area of trace metal determinations, the NIST provides 67 different SRMs consisting of individual certified metal standard solutions which are used as primary metal standards. In addition, standard solutions of most metals are available from a number of manufacturers at various concentrations which are certified against these NIST SRMs. NIST also provides a blood-lead standard SRM (955b) consisting of four whole-blood calibrators which can be used as calibration standards or QC samples. Calibration verification standards are also available from a number of sources, all NIST traceable.

Matrix-matched QC materials provide the best means of assessing the overall accuracy of a biological monitoring determination and should be included in every analytical run. For the determination of trace metals in blood/urine/serum/plasma, a number of QC materials are available. Table 6 lists those materials available from NIST. In addition to the NIST blood-lead standards/controls mentioned earlier, urine-based standards are available for a number of metals and fluoride. NIST materials are the gold standard and can be used with confidence. The bilevel controls include a normal control and an elevated control. For any other level, the elevated control can be diluted with the normal control to obtain the desired concentration. Tables 7–10 list many of the various commercial QCs that are available for blood lead, urine metals, whole-blood metals, and serum metals, respectively. These controls are also bilevel including a normal and elevated level.

Matrix spike analysis is a valuable method for checking the accuracy of any analytical method. This is especially true in cases were matrix-matched QC materials are not available. A matrix spike is a sample to which a known amount of the analyte of interest is added. The original sample and this spiked sample are then analyzed.

Table 6 NIST SRMs for biological monitoring

SRM 955b	Lead in blood: four levels
SRM 2670	Toxic and constituent elements in freeze-dried urine: toxic and normal levels As, Cd, Cr, Cu, Pb, Hg, Se (certified) Al, Be, Au, Mn, Ni, Pt, V (noncertified)
SRM 2671a	Fluoride in urine: toxic and normal
SRM 2672a	Mercury in urine: toxic and normal
SRM 1589	PCBs in human serum

PCB = polychlorinated biphenyl

Table 7 Blood-lead QC materials

CDC and Prevention
Bio-Rad, Inc.
Kaulson Laboratories, Inc.
New York State Department of Health
Commission of European Communities
Ciba-Corning Diagnostics
Seronorm™ Controls, Nycomed Pharma AS, Norway
Utak Laboratories, Inc.

Table 8 Urine metals QC materials

Bio-Rad Lyphochek® bilevel control									
Al	Sb	As	Cd	Cr	Co	Cu	Pb	Mg	Mn
Hg	Ni	Se	Tl	Zn	F	Aminolevulinic acid			
Instrumentation Laboratory Urichem bilevel control									
As	Ca	Cu	Pb	Mg	Hg	K	Na	Zn	
Creatinine		Aminolevulinic acid							
Kaulson Laboratories Contox® bilevel control									
Heavy metals control:				As	Cd	Hg	Pb		
Trace metals control:				As	Cd	Hg	Pb	Cu	Fe
				Zn	Al	Mg	Mn	Be	Co
Utak Urine Metals Control bilevel control									
Al	As	Cd	Cr	Co	Cu	F	Pb	Mn	
Hg	Mo	Ni	Se	V	Zn	Creatinine			

Table 9 Whole-blood metals QC materials

Kaulson Laboratories Contox® bilevel control								
As	Cd	Hg	Pb					
Seronorm™ trace elements trilevel control								
Cd	Pb	Hg	Al	Sb	As	Be	Bi	Co
Cr	F	Mn	Mo	Ni	Tl	V		
Commission of European Communities Community Bureau of Reference								
Cd		Pb						

Table 10 Serum trace metals QC materials

Kaulson Laboratories Contox® bilevel control								
Cu	Fe	Zn	Al	Mg	Mn	Be	Co	Cr
Ni	Se	Sb	Bi	Zn				
Utak bilevel serum trace elements control bilevel								
Al	Cd	Cr	Cu	Mg	Mn	Se	Zn	

The difference should correspond to the amount spiked. Acceptable recovery ranges from 75 to 125% in most cases. Spiking should be done at two or three different concentration levels. While 10% spikes is normal, for highly critical analyses it is not unusual to perform 20% spikes (every fifth sample).

16 THE FUTURE

Biological monitoring will continue to provide new information about metal exposure and enable us to protect the health of workers better. Analytical chemistry will have an important role in this effort. Analytical methods are improving daily with lower levels of detection coupled with good accuracy and precision. Analytical methods like ICP/MS which combines graphite furnace detection limits with multielement capability will find wider application. In addition, advances in automation and computer control will make ICP/MS instrumentation easier to use, more reliable, and more cost effective. GFAAS instrumentation continues to improve. Systems are now faster, more reliable, and at lower cost.

One particularly exciting aspect of trace metal analyses has already been mentioned. This is the area of metal speciation. Speciation methods were discussed for arsenic and mercury but are applicable to many other metals. Considerable work has been done in the area of trace metals of nutritional importance, for example selenium and chromium. These methods all combine a chromatographic separation technique, such as HPLC, with an elemental detection technique, such as ICP/MS. These studies will be expanded to the toxicologically important metals as well. A publication dealing with the various aspects of trace metal speciation has appeared.[139]

The biochemistry of trace metals has always been and will always be an exciting and active field.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
ACGIH	American Conference of Government Industrial Hygienists
ASV	Anodic Stripping Voltammetry
BAT	Biologischer Arbeitsstoff-Toleranz-Wert
BEI	Biological Exposure Indices
CAP	College of American Pathologists
CDC	Centers for Disease Control
CE	Capillary Electrophoresis
CLIA	Clinical Laboratory Improvement Act
CVAAS	Cold Vapor Atomic Absorption Spectroscopy
DMA	Dimethylarsinic Acid
EDTA	Ethylenediaminetetraacetic acid
EPA	Environmental Protection Agency
FAAS	Flame Atomic Absorption Spectroscopy
FDA	Food and Drug Administration
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
HCFA	Health Care Financing Administration
HPLC	High-performance Liquid Chromatography
ICP/AES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
LIMS	Laboratory Information Management System
MMA	Monomethylarsonic Acid
NCCLS	National Committee for Clinical Laboratory Standards
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
OSHA	Occupational Safety and Health Administration
PCB	Polychlorinated Biphenyl
PPE	Personal Protective Equipment
QA	Quality Assurance
QC	Quality Control
SRM	Standard Reference Material
STEL	Short-term Exposure Limit
TLV	Threshold Limit Value

WHO	World Health Organization
ZPP	Zinc Protoporphyrin
β-2M	β-2-Microglobulin

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Sample Preparation for Elemental Analysis of Biological Samples in the Environment

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Industrial Hygiene **(Volume 6)**
Parent and Progeny Compounds in Exhaled Breath, Determination of

Atomic Spectroscopy **(Volume 11)**
Background Correction Methods in Atomic Absorption Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

FURTHER READING

Basic Biological Monitoring Principles

American Conference of Governmental Industrial Hygienists, *Documentation of the Biological Exposure Indices*, 6th edition, ACGIH, Cincinnati, OH, 1991.

Shane Qee Hee, *Biological Monitoring, An Introduction*, Van Nostrand Reinhold, New York, 1993.

R.R. Lauwerys, P. Hoet, *Industrial Chemical Exposure, Guidelines for Biological Monitoring*, 2nd edition, Lewis Publishers, Ann Arbor, 1993.

Applications and Analytical Methods

Biological Monitoring of Chemical Exposure in the Workplace, World Health Organization, Geneva, Vols. I and II, 1996.

J. Angerer, K. Schaller (eds.), *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, Germany, Vol. 1, 1985.

J. Angerer, K. Schaller (eds.), *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, Vol. 2, 1988.

V. Fiserova-Bergerova, M. Ogata (eds.), *Biological Monitoring of Exposure to Industrial Chemicals*, American Conference of Governmental Industrial Hygienists, Cincinnati, 1990.

T. Kneip, J. Crable (eds.), *Methods for Biological Monitoring*, American Public Health Association, Washington, DC, 1988.

G. V. Iyengar, K.S. Subramanian, J.R.W. Woittiez, *Element Analysis of Biological Samples, Principles and Practice*, CRC Press, New York, 1998.

Reviews

S. Taylor et al., *Journal of Analytical Spectrometry* publishes a bi-annual 'Atomic Spectrometry Update – Clinical and Biological Materials, Food and Beverages', *J. Anal. Spect.*, **12**(4), 119R–221R (1997).

M. Harper, C.R. Glowacki, P.R. Michael, *Analytical Chemistry* publishes a bi-annual review of industrial hygiene which includes a section on biological monitoring. *Anal. Chem.*, **69**(12), 307R–327R (1997).

REFERENCES

1. American Conference of Governmental Industrial Hygienists, 'Introduction to the Biological Exposure Indices', in *Documentation of the Biological Exposure Indices*, 6th edition, ACGIH, Cincinnati, OH, BEI-1–BEI-3, 1993.
2. S. Qee Hee, 'Concepts in Chemistry', in *Biological Monitoring, An Introduction*, Van Nostrand Reinhold, New York, 3, 1993.
3. R.R. Lauwerys, P. Hoet, *Industrial Chemical Exposure, Guidelines for Biological Monitoring*, 2nd edition, Lewis Publishers, Ann Arbor, 2–4, 1993.
4. OSHA Pb Standard: Code of Federal Regulations, Title 29, Part 1910.1025 Lead, March 1, 1979.
5. OSHA Cd Standard: Code of Federal Regulations, Title 29, Part 1910.1027 Cadmium, September 14, 1992.
6. R.R. Lauwerys, P. Hoet, 'Introduction', in *Industrial Chemical Exposure, Guidelines for Biological Monitoring*, 2nd edition, Lewis Publishers, Ann Arbor, 5–7, 1993.
7. S. Qee Hee, 'Route of Exposure, Distribution, Metabolism, and Biological Monitoring Media', in *Biological Monitoring, An Introduction*, Van Nostrand Reinhold, New York, 71, 1993.
8. S.M. Rappaport, 'Biological Monitoring and Standard Setting in the USA: A Critical Appraisal', *Toxicol. Lett.*, **77**, 171–182 (1995).

9. American Conference of Governmental Industrial Hygienists, 'Biological Exposure Indices', in *TLVs®/BEIs® for 1998*, ACGIH, Cincinnati, OH, 91–107, 1998.
10. Deutsche Forschungsgemeinschaft, *List of MAK and BAT Values, 1995*, VCH, Weinhem, 138–149, 1995.
11. D. Henschler, 'Biological Monitoring and Exposure Limits as Perceived in Germany', in *Biological Monitoring of Exposure to Industrial Chemicals*, eds. V. Fiserova-Bergerova, M. Ogata, American Conference of Governmental Industrial Hygienists, Cincinnati, 25–30, 1990.
12. D. Miller, D.C. Paschal, E.W. Gunter, P.E. Stroud, J. D'Angelo, 'Determination of Lead in Blood Using Electrothermal Atomisation Atomic Absorption and Spectrometry with a L'vov Platform Matrix Modifier', *Analyst*, **112**, 1701–1704 (1987).
13. B.E. Jacobson, G. Lockitch, G. Quigley, 'Improved Sample Preparation for Accurate Determination of Low Concentrations of Lead in Whole Blood by Graphite Furnace Analysis', *Clin. Chem.*, **37**(4), 515–519 (1991).
14. P.C. D'Haese, L.V. Lamberts, L. Liang, F.L. Van de Vyver, M.E. De Broe, 'Elimination of Matrix and Spectral Interferences in the Measurement of Lead and Cadmium in Urine and Blood by Electrothermal Atomic Absorption Spectrometry with Deuterium Background Correction', *Clin. Chem.*, **37**(9), 1583–1588 (1991).
15. P.J. Parsons, W. Slavin, 'A Rapid Zeeman Graphite Furnace Atomic Absorption Spectrometric Method for the Determination of Lead in Blood', *Spectrochim. Acta. Part B*, **48B**, 925–939 (1993).
16. H. Qiao, P.J. Parsons, W. Slavin, 'Transferability of Blood Lead Determinations by Furnace Atomic Absorption Spectrophotometry and Continuum Background Correction', *Clin. Chem.*, **41**(10), 1451–1454 (1995).
17. C.E. Flaynik, D. Shrader, 'Determining Lead in Blood: Evaluating Deuterium and Zeeman Background Correction', *Am. Clin. Lab.*, **13**, 45–47 (1994).
18. J.P. Helfrich, 'A Workstation-based Quick Furnace Method for the Determination of Lead in Whole Blood', *Am. Clin. Lab.*, **13**, 14–15 (1994).
19. D.I. Bannon, C. Murashchik, C.R. Zapf, M.R. Farfel, J.J. Chisolm, Jr., 'Graphite Furnace Atomic Absorption Spectroscopic Measurement of Blood Lead in Matrix-Matched Standards', *Clin. Chem.*, **40**(9), 1730–1734 (1994).
20. S.M. Roda, R.D. Greenland, R.L. Bornschein, P.B. Hammond, 'Anodic Stripping Voltammetry Procedure Modified for Improved Accuracy of Blood Lead Analysis', *Clin. Chem.*, **34**(3), 563–567 (1988).
21. Kh. Braininam, H. Schafer, A. Ivanova, R. Khanina, 'Determination of Copper, Lead and Cadmium in Whole Blood by Stripping Voltammetry with the use of Graphite Electrodes', *Anal. Chim. Acta.*, **330**, 175–181 (1996).
22. M. Shannon, N. Rifai, 'The Accuracy of a Portable Instrument for Analysis of Blood Lead in Children', *Ambulatory Child Health*, **3**, 249–254 (1997).
23. D.E. Nixon, T.P. Moyer, 'Routine Clinical Determination of Lead, Arsenic, Cadmium, and Thallium in Urine and Whole Blood by Inductively Coupled Plasma Mass Spectrometry', *Spectrochim. Acta. Part B*, **51**, 13–25 (1996).
24. M. Viczian, A. Lasztity, R.M. Barnes, 'Identification of Potential Environmental Sources of Childhood Lead Poisoning by Inductively Coupled Plasma Mass Spectrometry. Verification and Case Studies', *J. Anal. Atom. Spect.*, **5**, 293–300 (1990).
25. E. Bakowska, D. Potter, 'The Determination of Cadmium, Lead and Mercury in Whole Blood by Inductively Coupled Plasma Mass Spectrometry', HP Application Note 228–301, Hewlett Packard Publication # (43) 5963–5489E, 1995.
26. E. Bakowska, 'Superior Stability of HP 4500 ICP/MS in Clinical Applications – Practical Approach', HP Application Note, Hewlett Packard Publication # 5965-90-307E, 1998.
27. D.C. Paschal, K.L. Caldwell, B.G. Ting, 'Determination of Lead in Whole Blood Using Inductively Coupled Argon Plasma Mass Spectrometry with Isotope Dilution', *J. Anal. At. Spectrom.*, **10**(5), 367 (1995).
28. A. Stroh, 'Determination of Pb and Cd in Whole Blood using Isotope Dilution ICP/MS', *Atom. Spectrosc.*, **14**(5), 141–143 (1993).
29. M.J. Campbell, H.T. Delves, 'Accurate and Precise Determination of Lead Isotope Ratios in Clinical and Environmental Samples Using Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **4**(3), 235 (1989).
30. National Institute for Occupational Safety and Health, 'Lead in Blood and Urine (Method 8003, issue 2 8/15/94)', *NIOSH Manual of Analytical Methods*: 4th edition, ed. P. Ehler, DHHS (NIOSH), NIOSH Cincinnati, OH, 1994.
31. J. Angerer, K. Schaller, 'Lead', in *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 155–164, Vol. 1, 1985.
32. D.G. Mitchell, K.M. Aldous, F.J. Ryan, 'Mass Screening for Lead Poisoning: Capillary Blood Sampling and Automated Delves-Cup Atomic Absorption Analysis', *New York J. Med.*, **74**, 1599–1603 (1974).
33. M.W. Wensing, B.W. Smith, J.D. Winefordner, 'Capacitively Coupled Microwave Plasma Atomic Emission Spectrometer for the Determination of Lead in Whole Blood', *Anal. Chem.*, **66**(4), 531–535 (1994).
34. G.S. Baird, R.L. Fitzgerald, S.K. Aggarwal, D.A. Heron, 'Determination of Blood Lead by Electron-capture Negative Chemical Ionization Gas Chromatography–Mass Spectrometry', *Clin. Chem.*, **42**(2), 286–291 (1996).
35. P. Dube, C. Krause, L. Windmuller, 'Direct Determination of Cadmium in Urine using Graphite Furnace Atomic Absorption Spectrometry with Zeeman-effect Background Correction', *Analyst*, **114**(10), 1240–1253 (1989).

36. K. Stoeppler, M. Brandt, 'Contributions to Automated Trace Metal Analysis. Part V. Determination of Cadmium in Whole Blood and Urine by Electrothermal Atomic Absorption Spectrophotometry', *Fresenius' Z. Anal. Chem.*, **300**, 372–380 (1980).
37. E. Pruszkowsha, G. Carnick, W. Slavin, 'Direct Determination of Cadmium in Urine with the Use of a Stabilized Temperature Platform Furnace and Zeeman Background Correction', *Clin. Chem.*, **29**, 477–480 (1983).
38. K.S. Subramanian, J.-C. Meranger, J.E. MacKeen, 'Graphite Furnace Atomic Absorption Spectrometry with Matrix Modification for Determination of Cadmium and Lead in Human Urine', *Anal. Chem.*, **55**(7), 1064–1067 (1983).
39. J. Angerer, K. Schaller, *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 85–96, Vol. 2, 1988.
40. J. Angerer, K. Schaller, 'Cadmium', in *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 79–92, Vol. 1, 1985.
41. T. Kneip, J. Crable (eds.), 'Method 107. Cadmium in Blood', *Method for Biological Monitoring*, American Public Health Association, Washington, DC, 161–164, 1988.
42. K.S. Subramanian, J.C. Meranger, 'A Rapid Electrothermal Atomic Absorption Spectrophotometric Method for Cadmium and Lead in Human Whole Blood', *Clin. Chem.*, **27**(11), 1866–1871 (1981).
43. Z. DeBenzo, R. Fraile, N. Carrion, 'Electrochemical Atomization Atomic Absorption Spectrometry with Stabilized Aqueous Standards for the Determination of Cadmium in Whole Blood', *Anal. Chim. Acta.*, **231**, 283–288 (1990).
44. C. Robert, J. Clark, 'Improved Determination of Cadmium in Blood and Plasma by Flameless Atomic Absorption Spectroscopy', *Bull. Envir. Contam. Toxicol.*, **36**, 496–499 (1986).
45. A. Karakaya, S. Suezen, B. Yuecesoy, 'A Biological Monitoring Method for Cadmium', *J. Anal. Toxicol.*, **16**(6), 403 (1992).
46. S.F. Durrant, 'Inductively Coupled Plasma–Mass Spectrometry for Biological Analysis', *Trends Anal. Chem.*, **11**, 68–73 (1992).
47. A. Nur Ona, A. Temizer, 'Determination of Lead and Cadmium in Urine by Differential-pulse Anodic Stripping Voltammetry', *Analyst*, **112**(3), 227–229 (1987).
48. R. Linstedt, 'Determination of Mercury in Urine by Cold Vapor Atomic Absorption', *Analyst*, **95**, 264 (1970).
49. EPA Methods 245.1 and 245.2, *Methods for Chemical Analysis of Water and Wastes: US Environmental Protection Agency*, Environmental Monitoring Systems Laboratory-Cincinnati (EMSL-Cl) EPA-600/4-79-020, Revised March 1983.
50. C.P. Hanna, J.F. Tyson, S. McIntosh, 'Determination of Total Mercury in Waters and Urine by Flow Injection Atomic Absorption Spectrometry. Procedures Involving On- and Off-line Oxidation of Organomercury Species', *Anal. Chem.*, **65**(5), 653–656 (1993).
51. D.A. Robbins, 'Method 116. Mercury in Urine', in *Methods for Biological Monitoring*, eds. T. Kneip, J. Crable, American Public Health Association, Washington, DC, 207–211, 1988.
52. R. Henderson, H.P. Shotwell, L.A. Krause, 'Analysis for Total, Ionic, and Elemental Mercury in Urine as a Basis for a Biologic Standard', *Am. Ind. Hyg. Assoc. J.*, **35**, 576–580 (1974).
53. J.-P. Weber, 'Interference Encountered for Urine Mercury', Le Centre de Toxicologie du Quebec Interlaboratory Comparison Program Memorandum to Participants, June 25, 1997.
54. T. Kneip, J. Crable (eds.), 'Method 153. Total and Methyl Mercury in Urine and Blood', in *Methods for Biological Monitoring*, American Public Health Association, Washington, DC, 283–289, 1988.
55. L. Magos, 'Selective Atomic Absorption Determination of Inorganic Mercury and Methylmercury in Undigested Biological Samples, *Analyst*, **96**, 847–853 (1971).
56. D. Littlejohn, G.S. Fell, J.M. Ottaway, 'Modified Determination of Total and Inorganic Mercury in Urine by Cold Vapor Atomic Absorption Spectrometry', *Clin. Chem.*, **22**(10), 1719–1723 (1976).
57. D.C. Sharma, P.S. Davis, 'Direct Determination of Mercury in Blood by Use of Sodium Borohydride Reduction and Atomic Absorption Spectrophotometry', *Clin. Chem.*, **25**(5), 769–772 (1979).
58. J. Angerer, K. Schaller, 'Mercury' *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 195–211, Vol. 2, 1988.
59. S.A. Winfield, N.D. Boyd, M.J. Vimy, F.L. Lorscheider, 'Measurement of Total Mercury in Biological Specimens by Cold Vapor Atomic Fluorescence Spectrometry', *Clin. Chem.*, **40**(2), 206–210 (1994).
60. M. Fujita, E. Takabatake, 'Continous Flow Reducing Vessel in Determination of Mercuric Compounds by Liquid Chromatography/Cold Vapor Atomic Absorption Spectrometry', *Anal. Chem.*, **55**(3), 454–457 (1983).
61. L. Liang, N.S. Bloom, M. Horvat, 'Simultaneous Determination of Mercury Speciation in Biological Materials by GC/CVAFS After Ethylation and Room-temperature Precollection', *Clin. Chem.*, **40**(4), 602–607 (1994).
62. E.P.C. Lai, E. Dabek-Zlotorzynska, 'Capillary Electrophoresis with Amperometric Detection for Mercury Speciation', *American Environmental Laboratory*, Feature Article, June (1996).
63. V.A. Granadillo, L.P. de Machado, R.A. Romero, 'Determination of Total Chromium in Whole Blood, Blood Components, Bone, and Urine by Fast Furnace Program Electrothermal Atomization AAS and Using Neither Analyte Isoformation nor Background Correction', *Anal. Chem.*, **66**, 3624–3631 (1994).

64. J. Angerer, K.H. Schaller, 'Chromium', *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 97–115, Vol. 2, 1988.
65. K.H. Schaller, 'Chromium' in *Biological Monitoring of Chemical Exposure in the Workplace*, World Health Organization, Geneva, 91–111, Vol. I, 1996.
66. J.N Marks, M.A. White, A.R. Boran, 'Direct Determination of Chromium in Urine by Graphite Furnace Atomic Absorption Spectrophotometry', *Atomic Spectrosc.*, **9**, 73–75, (1988).
67. D.C. Paschal, G.G. Bailey, 'Determination of Chromium in Urine with Graphite Furnace Atomic Absorption Spectroscopy with Zeeman Correction', *Atomic Spectrosc.*, **12**, 151–154 (1991).
68. A.A. Bouman, A.J. Platenkamp, F.D. Posma, 'Determination of Cobalt in Blood and Urine by Flameless Atomic Absorption Spectroscopy. Comparison of Direct Analysis using Zeeman Background Correction and Indirect Analysis using Extraction in Organic Solutions', *Ann. Clin. Biochem.*, **23**, 346–350 (1986).
69. D. Templeton, 'Cobalt' in *Biological Monitoring of Chemical Exposure in the Workplace*, World Health Organization, Geneva, 35–50, Vol. II, 1996.
70. J. Angerer, K. Schaller, 'Cobalt', in *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 141–153, Vol. 1, 1985.
71. T. Kneip, J. Crable (eds.), 'Method 108. Cobalt in Urine', *Methods for Biological Monitoring*, American Public Health Association, Washington, DC, 165–170, 1988.
72. J. Angerer, K.H. Schaller, 'Cobalt', *Analyses of Hazardous Substances in Biological Materials*, VCH, Weinheim, 120–131, Vol. 2, 1988.
73. D. Templeton, 'Cobalt', in *Biological Monitoring of Chemical Exposure in the Workplace*, World Health Organization, Geneva, 46, Vol. II, 1996.
74. F.E. Brinckman, G.E. Parris, W.R. Blair, K.L. Jewett, W.P. Iverson, J.M. Bellama, 'Questions Concerning Environmental Mobility of Arsenic: Needs for a Chemical Data Base and Means for Speciation of Trace Organoarsenicals', *Environ. Health Perspectives*, **19**, 11–24 (1977).
75. D.J. Thomas, 'Arsenic Toxicity in Humans: Research Problems and Prospects', *Environ. Geochem. and Health*, **16**(3/4), 107–111 (1994).
76. M.P. Waalkes, Z. Zhao, 'The Association of Arsenic-induced Malignant Transformation with DNA Hypomethylation and Aberrant Gene Expression', Paper No. 1737, 1998 Society of Toxicology Conference, Seattle, WA, March, 1998.
77. M.J. Mass, 'Mechanisms of Arsenic Induced Cancer: A Role for Hypermethylation', Paper No. 1738, Society of Toxicology, Seattle, WA, March, 1998.
78. J.P. Buchet, R. Lauwerys, H. Roels, 'Comparison of Several Methods for the Determination of Arsenic Compounds in Water and Urine, Their Application for the Study of Arsenic Metabolism and for the Monitoring of Workers Exposed to Arsenic', *Int. Arch. Occup. Environ. Health*, **46**, 11–20 (1980).
79. R.K. Anderson, M. Thompson, E. Culbard, 'Selective Reduction of Arsenic Species by Continuous Hydride Generation, Part I. Reaction Media', *Analyst*, **111**, 1143–1151 (1986).
80. R.K. Anderson, M. Thompson, E. Culbard, 'Selective Reduction of Arsenic Species by Continuous Hydride Generation. Part II. Validation of Methods for Application to Natural Waters', *Analyst*, **111**, 1153–1160 (1986).
81. H. Norin, M. Vahter, 'A Rapid Method for the Selective Analysis of Total Urinary Metabolites of Inorganic Arsenic', *Scand. J. Work Environ. Health*, **7**(1), 38–44 (1981).
82. M. Vahter, L. Friberg, B. Rahnster, A. Nygren, P. Nolinder, 'Airborne Arsenic and Urinary Excretion of Metabolites of Inorganic Arsenic among Smelter Workers', *Int. Arch. Occup. Environ. Health*, **57**(1), 79–91 (1986).
83. R. Torralba, M. Bonilla, L. V. Perez-Arribas, A. Palacios, C. Camara, 'Speciation and Simultaneous Determination of Arsenic(III), Arsenic(V), Monomethylarsonate and Dimethylarsinate by Atomic Absorption using Inverse Least Squares Multivariate Calibration', *Spectrochim. Acta Part B*, **49**(9), 893–899 (1994).
84. X-C Le, W.R. Cullen, K.J. Reimer, 'Determination of Urinary Arsenic and Impact of Dietary Arsenic Intake', *Talanta*, **40**(2), 185–193 (1993).
85. C.P. Hanna, J.F. Tyson, S. McInosh, 'Determination of Inorganic Arsenic and Its Organic Metabolites in Urine by Flow-injection Hydride Generation Atomic Absorption Spectrometry', *Clin. Chem.*, **39**(8), 1662–1667 (1993).
86. O.J. DeBlais, S.V. Gonzalez, R.S. Rodriguez, J.H. Mendez, 'Determination and Speciation of Arsenic in Human Urine by Ion-exchange Chromatography/Flow Injection Analysis with Hydride Generation/Atomic Absorption Spectroscopy', *J. AOAC Int.*, **77**(2), 441–445 (1994).
87. R.S. Braman, D.L. Johnson, C.C. Foreback, J.M. Ammons, J.L. Bricker, 'Separation and Determination of Nanogram Amounts of Inorganic Arsenic and Methylarsenic Compounds', *Anal. Chem.*, **49**(4), 621–625 (1977).
88. E.A. Crecelius, 'Modification of the Arsenic Speciation Technique Using Hydride Generation', *Anal. Chem.*, **50**(6), 826–827 (1978).
89. J.S. Edmonds, K.A. Francesconi, 'Estimation of Methylated Arsenicals by Vapor Generation Atomic Absorption Spectrometry', *Anal. Chem.*, **48**(13), 2019–2020 (1976).
90. D.E. Nixon, G.V. Mussmann, S.J. Eckdahl, T.P. Moyer, 'Total Arsenic in Urine: Palladium Persulfate vs Nickel as a Matrix Modifier for Graphite Furnace Atomic Absorption Spectrophotometry', *Clin. Chem.*, **37**(9), 1575–1579 (1991).

91. A.A. Grabinski, 'Determination of Arsenic(III), Arsenic(V), Monomethylarsonate, and Dimethylarsinate by Ion-exchange Chromatography with Flameless Atomic Absorption Spectrometric Detection', *Anal. Chem.*, **53**(7), 966–968 (1981).
92. D.E. Nixon, T.P. Moyer, 'Arsenic Analysis II. Rapid Separation and Quantification of Inorganic Arsenic Plus Metabolites and Arsenobetaine from Urine', *Clin. Chem.*, **38**(12), 2479–2483 (1992).
93. P. Bavazzano, A. Perico, K. Rosendahl, P. Apostoli, 'Determination of Urinary Arsenic by Solvent Extraction and Electrochemical Atomic Absorption Spectrometry. A Comparison with Directly Coupled High-performance Liquid Chromatography–Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atomic Spectrosc.*, **11**(7), 521–524 (1996).
94. A. LeBlanc, 'A Method for the Determination of Arsenic in Urine' Le Centre de Toxicologie du Quebec, personal communication.
95. G.R. Ricci, L.S. Shepard, G. Colovos, N.E. Hester, 'Ion Chromatography with Atomic Absorption Spectrometric Detection for Determination of Organic and Inorganic Arsenic Species', *Anal. Chem.*, **53**(4), 610–613 (1981).
96. T. Maitani, S. Uchiyama, Y. Saito, 'Hydride Generation–Flame Atomic-Absorption Spectrometry as an Arsenic Detector for High-performance Liquid Chromatography', *J. Chromatogr.*, **391**, 161–168 (1987).
97. B.S. Chana, N.J. Smith, 'Urinary Arsenic Speciation by High-Performance Liquid Chromatography/Atomic Absorption Spectrometry for Monitoring Occupational Exposure to Inorganic Arsenic', *Anal. Chim. Acta.*, **197**, 177–186 (1987).
98. E. Hakala, L. Pyy, 'Selective Determination of Toxicologically Important Arsenic Species in Urine by High-performance Liquid Chromatography–Hydride Generation Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **7**(3), 191–196 (1992).
99. A.G. Howard, L.E. Hunt, 'Coupled Photooxidation–Hydride AAS Detector for the HPLC of Arsenic Compounds', *Anal. Chem.*, **65**(21), 2995–2998 (1993).
100. J-S. Blais, G-M. Momplaisir, W.D. Marshall, 'Determination of Arsenobetaine, Arsenocholine, and Tetramethylarsonium Cations by Liquid Chromatography–Thermochemical Hydride Generation–Atomic Absorption Spectrometry', *Anal. Chem.*, **62**(11), 1161–1166 (1990).
101. X.C. Le, M. Ma, N.A. Wong, 'Speciation of Arsenic Compounds Using High-performance Liquid Chromatography at Elevated Temperature and Selective Hydride Generation Atomic Fluorescence Detection', *Anal. Chem.*, **68**(24), 4501–4506 (1996).
102. X.C. Le, M. Ma, 'Short-column Liquid Chromatography with Hydride Generation Atomic Fluorescence Detection for the Speciation of Arsenic', *Anal. Chem.*, **70**(9), 1926–1933 (1998).
103. M. Morita, T. Uehiro, K. Fuwa, 'Determination of Arsenic Compounds in Biological Samples by Liquid Chromatography with Inductively Coupled Argon Plasma–Atomic Emission Spectrometric Detection', *Anal. Chem.*, **53**(12), 1806–1808 (1981).
104. K.E. LaFreniere, V.A. Fassel, D.E. Eckels, 'Elemental Speciation via High-performance Liquid Chromatography Combined with Inductively Coupled Plasma Atomic Emission Spectroscopic Detection: Application of a Direct Injection Nebulizer', *Anal. Chem.*, **59**(6), 879–887 (1987).
105. G.K.C. Low, G.E. Batley, S.J. Buchanan, 'Application of Column Switching in High-performance Liquid Chromatography to Arsenic Speciation Analysis with Inductively Coupled Argon Plasma Spectrometric Detection', *J. Chromatogr.*, **386**, 423–426 (1986).
106. R. Rubio, A. Padro, J. Alberti, G. Rauret, 'Determination of Arsenic Speciation by Liquid Chromatography–Hydride Generation Inductively Coupled Plasma Atomic Emission Spectrometry with On-line UV Photooxidation', *Anal. Chim. Acta*, **283**, 160–166 (1993).
107. E. Larsen, G. Pritzl, S.H. Hansen, 'Speciation of Eight Arsenic Compounds in Human Urine by High-performance Liquid Chromatography with Inductively Coupled Plasma Mass Spectrometric Detection Using Antimonate for Internal Chromatographic Standardization', *J. Anal. At. Spectrom.*, **8**(6), 557–563 (1993).
108. M.M. Kershisnik, R. Kalamegham, K.O. Ash, D.E. Nixon, E.R. Ashwood, 'Using $^{16}O^{35}Cl$ to Correct for Chloride Interference Improves Accuracy of Urine Arsenic Determinations by Inductively Coupled Plasma Mass Spectrometry', *Clin. Chem.*, **38**(11), 2197–2202 (1992).
109. S.A. Pergantis, E.M. Heithmar, T.A. Hinners, 'Microscale Flow Injection and Microbore High-performance Liquid Chromatography Coupled with Inductively Coupled Plasma Mass Spectrometry via a High-efficiency Nebulizer', *Anal. Chem.*, **67**(24), 4530–4535 (1995).
110. S.X.C. Le, W.R. Cullen, K.J. Relmer, 'Speciation of Arsenic Compounds in Some Marine Organisms', *Environ. Sci. Technol.*, **28**(9), 1598–1604 (1994).
111. X.C. Le, W.R. Cullen, K.J. Reimer, 'Human Urinary Arsenic Excretion After One-time Ingestion of Seaweed, Crab and Shrimp', *Clin. Chem.*, **40**(4), 617–625 (1994).
112. M. Ma, X.C. Le, 'Effect of Arsenosugar Ingestion on Urinary Arsenic Speciation', *Clin. Chem.*, **44**(3), 539–550 (1998).
113. National Institute for Occupational Safety and Health, 'Metals in Urine, Method 8310', *NIOSH Manual of Analytical Methods*, Issue 2, (8/15/94), 4th edition, P. Ehler (ed.) DHHS (NIOSH), NIOSH Cincinnati, OH, 1994.
114. National Institute for Occupational Safety and Health, 'Elements in Blood and Tissue, Method 8005', *NIOSH Manual of Analytical Methods*, Issue 2, (8/15/94),

4th edition, P. Ehler (ed.) DHHS (NIOSH), NIOSH Cincinnati, OH, 1994.
115. R.S. Houk, 'Mass Spectrometry of Inductively Coupled Plasma', *Anal. Chem.*, **58**(1), 97A–104A (1986).
116. S.F. Durrant, 'Inductively Coupled Plasma–Mass Spectrometry for Biological Analysis', *Trends Anal. Chem.*, **11**(2), 68–73, (1992).
117. H. Vanhoe, C. Vandecasteele, J. Versieck, R. Dams, 'Determination of Iron, Cobalt, Copper, Zinc, Rubidium, Molybdenum, and Cesium in Human Serum by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **61**(17), 1851–1857 (1989).
118. K.J. Mulligan, T.M. Davidson, J.A. Caruso, 'Feasibility of the Direct Analysis of Urine by Inductively Coupled Argon Plasma Mass Spectrometry for Biological Monitoring of Exposure to Metals', *J. Anal. At. Spectrom.*, **5**(6), 301–306 (1990).
119. M.A. Vaughan, A.D. Baines, D.M. Templeton, 'Multielement Analysis of Biological Samples by Inductively Coupled Plasma–Mass Spectrometry. II. Rapid Survey Method for Profiling Trace Elements in Body Fluids', *Clin. Chem.*, **37**(2), 210–215 (1991).
120. C.-S. Hsiung, J.D. Andrade, R. Costa, K.O. Ash, 'Minimizing Interferences in the Quantitative Multielement Analysis of Trace Elements in Biological Fluids by Inductively Coupled Plasma Mass Spectrometry', *Clin. Chem.*, **43**(12), 2303–2311 (1997).
121. C.-C. Huang, M.-H. Yang, T.-S. Shih, 'Automated Online Sample Pretreatment System for the Determination of Trace Metals in Biological Samples by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **69**(19), 3930–3939 (1997).
122. D. Agness, F. Portala, I.L. Shuttler, M. Leyrer, 'Simultaneous ETAAS Determination of Aluminum, Chromium, Manganese, and Nickel in Serum and Plasma', Perkin-Elmer SIMAA 6000 Application Note, Perkin-Elmer UK, Beaconsfield, 1995.
123. T.C. Pan, C.J. Horng, S.R. Lin, T.H. Lin, C.W. Huang, 'Simultaneous Determination of Zn, Cd, Pb, and Cu in Urine of Patients with Blackfoot Disease Using Anodic Stripping Voltammetry', *Biol. Trace Elem. Res.*, **38**, 233–241 (1993).
124. C. Bradley, F.Y. Leung, 'Aluminum Determined in Plasma and Urine by Atomic Absorption Spectroscopy with a Transversely Heated Graphite Atomized Furnace', *Clin. Chem.*, **40**(3), 431–434 (1994).
125. F. Baruthio, O. Guillard, J. Arnaud, F. Pierre, R. Zawislak, 'Determination of Manganese in Biological Materials by Electrothermal Atomic Absorption: a Review', *Clin. Chem.*, **34**(2), 227–234 (1988).
126. D.C. Paschal, G.G. Bailey, 'Determination of Manganese in Serum with Platform Furnace using Zeeman Correction', *Atom. Spectosc.*, **8**(5), 150–152 (1987).
127. G.A. Hams, J.K. Fabri, 'An Analysis for Blood Manganese Used to Assess Environmental Exposure', *Clin. Chem.*, **34**(6), 1121–1123 (1988).
128. J. Neve, N. Leclercq, 'Factors Affecting Determinations of Manganese in Serum by Atomic Absorption Spectrometry', *Clin. Chem.*, **37**(5), 723–728 (1991).
129. F.W. Sunderman, Jr., S.M. Hopfer, M.C. Crisostomo, 'Nickel by Electrothermal Atomic Absorption Analysis', *Method Enzymol*, **158**, 382–91 (1988).
130. D.E. Nixon, T.P. Moyer, D.P. Squillace, J.T. McCarthy, 'Determination of Serum Nickel by Graphite Furnace Atomic Absorption with Zeeman-effect Background Correction. Values in a Normal Population and a Population Undergoing Dialysis', *Analyst*, **114**, 1671–1674 (1989).
131. G. Heinemann, W. Vogt, 'Quantification of Vanadium in Serum by Electrothermal Atomic Absorption Spectrometry', *Clin. Chem.*, **42**(8), 1275–1282 (1996).
132. M.W. Airbouine, N.J. Smith, 'Determination of Vanadium in Urine and Its Application to Biological Monitoring of Occupationally Exposed Workers', *Atom. Spectrosc.*, **12**(2), 454–457 (1991).
133. A. Aitio, J. Jarvisalo, 'Biological Monitoring of Occupational Exposure to Toxic Chemicals, Collection, Processing and Storage of Specimens', *Annal. Clin. Lab. Sci.*, **15**(2), 121–139 (1985).
134. NCCLS Approved Standard, 'Control of Preanalytical Variation in Trace Element Determinations; Approved Guideline', Document No. C38-A, NCCLS, Wayne, PA, 1997.
135. K.S. Subramanian, 'Storage and Preservation of Biomedical Specimens', in *Element Analysis of Biological Samples, Principles and Practice*, eds. G.V. Iyengar, K.S. Subramanian, J.R.W. Woittiez, CRC Press, New York, 80–83, 1998.
136. C. Bradley, F.Y. Leung, 'Quality Control Procedures for Monitoring Whole Blood and Plasma Trace Metals by Atomic Absorption Spectrophotometry', *Sci. Total Environ.*, **89**, 353–359 (1989).
137. American Industrial Hygiene Association Sampling and Laboratory Analysis Committee, *Laboratory Quality Assurance Manual*, 2nd edition, AIHA Press, Fairfax, VA, 1997.
138. US Department of Labor, *OSHA List of Laboratories Approved for Blood Lead Analysis*, Occupational Safety and Health Administration, Salt Lake City, UT, January 2, 1998.
139. 'Analysis of Trace Element Species', in *Element Analysis of Biological Samples, Principles and Practice*, eds. G.V. Iyengar, K.S. Subramanian, J.R.W. Woittiez, CRC Press, New York, 175–202, 1998.
140. S. Valkonen, A. Aitio, 'Analysis of Aluminum in Serum and Urine for the Biomonitoring of Occupational Exposure', *Sci. Total Environ.*, **199**(1–2), 103–110 (1997).

Neutron Radiography for the Analysis of Plant–Soil Interactions

Brett H. Robinson, Ahmad Moradi and Rainer Schulin
Institute of Terrestrial Ecosystems, ETH, Zürich, Switzerland

Eberhard Lehmann
Paul Scherrer Institut, Villigen, Switzerland

Anders Kaestner
Paul Scherrer Institut, Villigen, Switzerland

1 Introduction 1
2 Neutron Radiography Configuration 1
3 Neutron Interactions with the Plant–Soil System 2
4 Configuration of Soil–Plant System For Optimal Imaging with Neutron Radiography 3
5 Practical Considerations 5
6 Application 6
7 Conclusions and Outlook 6
Related Articles 7
References 7

Neutron radiography (NR) can be used to quantify the spatial distribution of water in the soil–plant system with high precision and good spatial resolution. This property of neutron imaging results from the high interaction probability of hydrogen nuclei with slow neutrons. If there is a sufficient difference between the water content of the soil and roots, neutron radiographs can reveal plant roots and show root development. NR is noninvasive, and the radiation dose needed to image plant roots in soil does not affect plant development. Quantification of the soil's water content often requires correction for neutron-scattering artifacts. Root visibility is proportional to root thickness, and is inversely related to the width of the sample container and the water and organic matter contents of the ambient soil. Ideally, the soil should have low organic matter content and low water content but still permit the normal development of plant roots. Currently, the availability of neutron-imaging facilities limits the widespread application of NR to soil and root studies. However, technological development and increased investment will result in NR becoming a standard method for some soil–plant analyses.

1 INTRODUCTION

Terrestrial plants provide humanity with food and oxygen. Above- and belowground plant processes are equally important: while assimilation through photosynthesis occurs in the aboveground portions, the uptake of water and mineral nutrients occurs belowground. We have a limited understanding of some belowground processes because soil hinders the observation of plant roots. It is difficult to measure root development and water flux in the root zone without disturbing root growth or using artificial systems. Existing techniques include minirhizotrons, which are transparent plastic tubes inserted into the ground to view the roots, e.g. by using a video camera.[1] Minirhizotrons interfere with the root environment and only provide an incomplete picture. X-ray radiography has insufficient contrast to reveal root–water interactions.[2]

Willatt et al.[3] showed that NR could reveal roots and root zone processes, without greatly perturbing the system, thus allowing sequential measurements. This is a critical advantage of NR over other techniques. However, the available technology in the 1970s gave images of insufficient quality and exposed the plants to radiation doses that were potentially harmful. Recent technological developments, especially improved beam collimation, detection systems, and image-processing techniques, have allowed the production of images with much higher contrast and spatial resolution while reducing the plants' radiation exposure. These advances open up the possibility of using this technique to study root system development in soils and simultaneously monitor soil moisture distribution in near real time. Here, we describe the state of the art of NR as it relates to the analysis of plant–soil interactions.

2 NEUTRON RADIOGRAPHY CONFIGURATION

Figure 1 shows the configuration of a typical NR facility. NR requires a neutron source, which may be a reactor, the target of an elemental particle accelerator, or a neutron-emitting isotope. Thermal (12–100 meV) and cold (0.12–12 meV) neutrons are preferable for NR investigations. Therefore, epithermal, intermediate, and fast neutrons from the neutron source must be slowed

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

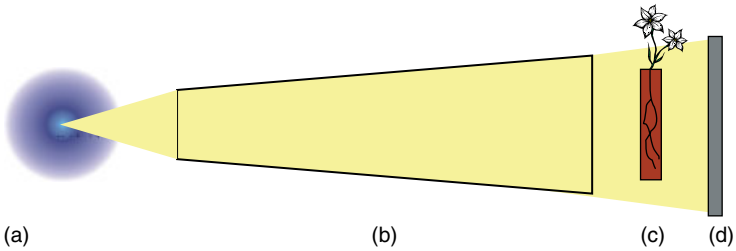

Figure 1 Configuration required for neutron imaging: (a) neutron source, (b) collimator, (c) sample, and (d) detector.

down using a moderator such as heavy water. Neutrons enter a collimator that forms a neutron beam with specific geometric properties. The collimator may also contain filters that modify the energy spectrum of the beam or reduce the beam's content of γ rays. The neutron beam is transmitted through the sample onto a plane position-sensitive detector, which is usually a scintillation screen. A charge-coupled device (CCD) camera coupled to the detector records a two-dimensional image that is a projection of the object on the detector plane. The advent of high-resolution digital cameras has allowed fast imaging and increased resolution. Digital imaging techniques that use a CCD camera combined with image-processing tools nowadays yield quantifiable images with a resolution of about 100 μm. Lehmann et al.[4] obtained a much higher spatial resolution with a special setup for microtomography. The required exposure time for such images is in the order of seconds. Images acquired using imaging plates or films (used until 1995) required several minutes of exposure to the potentially damaging neutron beam. The radiation dose received per image using modern techniques is about 0.003 mSv,[5] which is some two orders of magnitude less than the minimum value of 0.2 mSv h^{-1} found to affect plant growth.[6]

3 NEUTRON INTERACTIONS WITH THE PLANT–SOIL SYSTEM

NR is based on the Beer–Lambert exponential law of attenuation of radiation passing through matter[7]:

$$I = I_0 \exp(-\Sigma_{\text{sample}} d) \quad (1)$$

where I is the attenuated radiation (neutron) flux (cm^{-2} s^{-1}), after an incident neutron flux I_0 passes through a material of thickness d (cm) with an attenuating coefficient Σ (cm^{-1}), which is a characteristic of the material. The attenuation coefficient, also called the *macroscopic cross section*, is related to the tabulated microscopic cross section σ (cm^2) as

$$\Sigma = N\sigma \quad (2)$$

with a nuclear density N (M). When a sample is placed in a neutron beam, heterogeneities in the composition and thickness of the sample result in variations in the intensity of the transmitted beam. Unlike X rays, neutron radiation interacts with atomic nuclei. There is no systematic change in the neutron attenuation coefficient with atomic number or mass. Each isotope has a specific neutron cross section, σ, which is also energy dependent. Hydrogen has a neutron cross section some 10 times greater than deuterium and also greater than many other elements in the soil–plant system.

NR reveals structures in plant–soil systems owing to differences in the Σ values of the system's components. Table 1 shows a list of the chemical elements in the plant–soil system, along with their abundances and relative neutron cross sections. In both plant and soil, hydrogen is responsible for more than 90% of the neutron attenuation.

Some hydrogen is associated with organic molecules in the system; however, most hydrogen is water borne. Plant roots may thus be distinguished from soil due to their higher water content (θ). The gravimetric water content, θ ($g\,g^{-1}$) of plant roots generally ranges between 0.7 and 0.95 $g\,g^{-1}$, while that of soils at field capacity usually ranges between 0.12 and 0.3 $g\,g^{-1}$. The structures that NR reveals in the soil–plant system are sensitive to θ.

The high neutron attenuation coefficient of hydrogen is an important advantage of NR over X rays when applied to soil–plant system because the difference in water content allows the visualization of roots. The X-ray attenuation coefficients of root and soil components are less distinct (data not shown); therefore, the resulting radiograph has less contrast. Figure 2 shows that NR provides a better contrast between roots and soil than X rays.

The attenuation coefficient (Σ) results from two types of neutron interactions with matter: absorption (Σ_a) and scattering (Σ_s) (Figure 3a). Hydrogen attenuates neutrons primarily by noncoherent elastic scattering.[11] Neutron scattering causes deviations from the exponential law of attenuation for thicker samples (more than a few millimeters) because some neutrons are multiple scattered into the detector plane, thus producing

Table 1 Chemical elements, listed in order of abundance, in the plant[8]–soil[9] system, along with their relative neutron attenuation coefficients.[10] The plant and soil are assumed to have water contents of 0.8 and 0.2 g g^{-1}, respectively

	Plant (mol kg^{-1})	Soil (mol kg^{-1})	Element Σ (cm^{-1})	Plant neutron attenuation (cm^{-1})	Soil neutron attenuation (cm^{-1})
Hydrogen	100	24	3.4	344	82
Oxygen	50	35	0.17	8.5	6
Silicon	71E–4	9.3	0.11	<0.1	1.02
Carbon	7.5	1.1	0.56	4.2	0.6
Aluminum	7.4E–4	2.1	0.1	<0.1	0.21
Potassium	5.1E–2	0.93	0.06	<0.1	<0.1
Sodium	8.7E–4	0.82	0.09	<0.1	<0.1
Nitrogen	0.21	3.6E–2	0.43	<0.1	<0.1
Calcium	2.5E–2	0.19	0.08	<0.1	<0.1
Iron	3.6E–4	0.19	1.2	<0.1	0.22
Magnesium	1.6E–2	7.8E–2	0.15	<0.1	<0.1
Phosphorus	1.3E–2	1.2E–2	0.12	<0.1	<0.1
Titanium	4.1E–6	2.4E–2	0.6	<0.1	<0.1
Sulfur	6.2E–3	2.3E–3	0.06	<0.1	<0.1
Manganese	1.8E–4	2.8E–3	1.2	<0.1	<0.1
Chlorine	5.6E–4	1.6E–3	1.3	<0.1	<0.1
Zinc	6.1E–5	5.3E–4	0.35	<0.1	<0.1
Boron	3.7E–4	1.8E–4	102	<0.1	<0.1
Copper	1.9E–5	1.6E–4	1.1	<0.1	<0.1
Molybdenum	2.1E–7	5.2E–5	0.52	<0.1	<0.1
Total	158	74		356	91

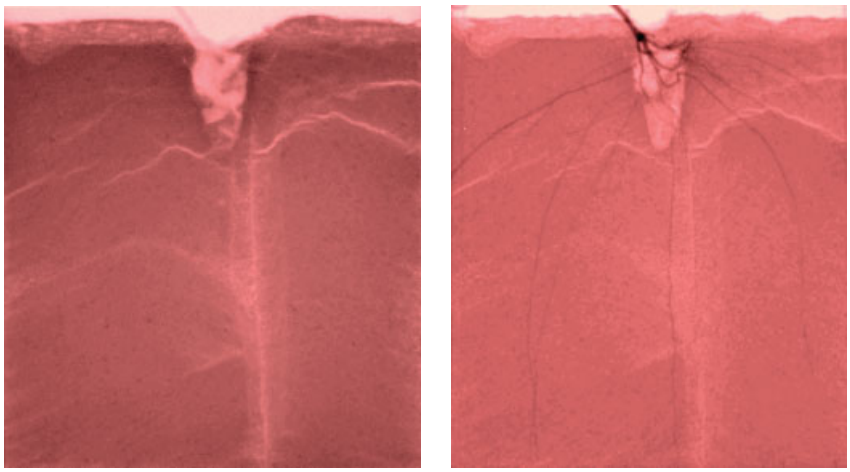

(a) (b)

Figure 2 Comparison of contrast between roots and soil in X-ray (120 keV) radiograph (a) and neutron radiograph (b) of the same sample. While soil heterogeneity and soil cracks are more visible in the X-ray radiograph, neutron radiography provides better contrast between soil and roots.

an artificially high signal. While the significance of neutron scattering is of little importance in the detection of roots in soil, it can be problematic when quantifying θ of the system. Algorithms such as the quantitative neutron imaging (QNI)[12] correct for the nonlinearity arising from neutron scattering. The QNI iteratively reconstructs the image by overlapping point-scattered functions calculated using a Monte Carlo simulation.

4 CONFIGURATION OF SOIL–PLANT SYSTEM FOR OPTIMAL IMAGING WITH NEUTRON RADIOGRAPHY

A neutron radiograph is the result of all the neutron attenuation processes that occur when the neutron beam passes through a sample. The Σ value of the soil depends on θ and the neutron attenuation properties of its solid components. Soils that are high in iron or organic matter

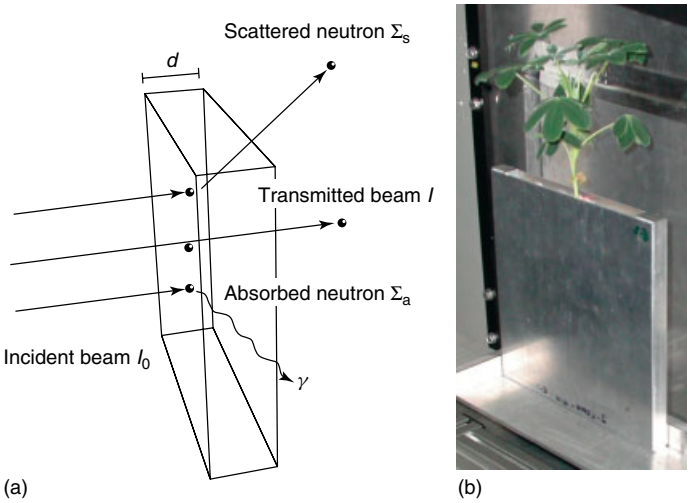

Figure 3 Neutron interactions with the sample (a) and an example of a plant sample (*Lupinus albis*) mounted in front of the detector (b). The dimensions of the sample container are 150 mm × 150 mm × 12 mm.

Table 2 Properties of various plant growth media with regard to NR

	Bulk density ($g\,cm^{-3}$)	Water content at 1 bar ($g\,g^{-1}$)	Σ (dry material) (cm^{-1})	Σ (water content at −1 bar) (cm^{-1})	Notes
Perlite	0.125	1.08	0.1	4.5	–
Porous glass beads	0.49	0.16	1.6	2.5	High Σ due to boron in glass
Ferrous mine tailings	1.4	0.08	1	1.6	–
Loam	1.2	0.17	0.5	1.5	–
Peat	0.58	0.09	0.75	1.4	–
Loamy sand	1.3	0.09	0.3	0.75	Normal root development
Fine quartz sand	1.5	0.01	0.25	0.35	Root development perturbed
Coarse quartz sand	1.45	0.01	0.25	0.3	

are unsuitable for NR because their high Σ values obscure the visualization of plant roots. The value of θ of the soil at the time of measurement should be as low as possible without inducing water stress in the plants. In practice, this represents the soil's θ at a water potential of −1 bar. Silica sands have a low inherent Σ and a low θ at −1 bar. However, Menon et al.[5] showed that the high density and sharp edges of this material perturb normal root growth. Moradi et al. (unpublished data) tested a variety of plant growth media for their suitability in NR (Table 2).

Table 2 shows that loamy sand, which permits normal root development, has a relatively low Σ at a θ at a water potential of −1 bar. Loamy sand has a higher θ at field capacity (ca. $0.35\,g\,g^{-1}$) than quartz sand (ca. $0.15\,g\,g^{-1}$). Plants can thus be left longer in loamy sand without irrigation.

Root visibility is proportional to root thickness, and inversely related to the width of the sample container.[13–17] The minimum detectable root thickness increases exponentially as the thickness of the soil profile increases. Moradi et al. (unpublished data) showed that using the loamy sand in Table 2, with a soil θ of $0.16\,g\,g^{-1}$ (ca. −1 bar) and an average root θ of $0.85\,g\,g^{-1}$, the minimum detectable root thickness R (mm) is related empirically to the slab thickness T (mm) according to the function:

$$R = 0.0034T^{1.68} \quad (3)$$

Roots are most easily visualized in thin containers. However, this condition is not conducive to plant growth since the walls of the container restrict root development. Therefore, one needs to find a balance between the ease of root visualization and the restriction of normal plant development.

For many herbaceous species, a soil profile with a thickness of 12 mm provides enough space for relatively normal root development.[5] This gives a minimal detectable root thickness of 0.22 mm (Equation 2). Such a setup allows the visualization of the skeleton of the root system; however, fine roots (<2 mm in diameter) are undetectable.

5 PRACTICAL CONSIDERATIONS

The sample container should be made of a material with a low Σ, such as aluminum ($\Sigma = 0.1\,\mathrm{cm}^{-1}$). Importantly, the system should not contain high concentrations of cobalt, which can form the persistent radioactive isotope Cobalt-60, with a half-life of about five years, upon exposure to neutron radiation.

Filling the sample container with soil as homogeneously as possible provides better contrast for root visualization. Filling from the side of the container rather than from the top results in less structural heterogeneity associated with variations in the particle size distribution and thus pore size distribution, which are visible in the resulting neutron radiograph because of variations in soil water content. Figure 4 shows the effect of filling patterns caused by pouring soil from the top of the container.

The suitability of NR to investigate root systems differs among plant species. As described earlier, thick roots with a high θ are more easily resolved than finer, drier roots. In a 10-mm-thick soil profile, the minimum detectable root size is ca. 0.16 mm (Equation 2). However, many plant species produce finer roots, which would not be visible in this system. Decreasing the thickness of the soil profile may permit the visualization of these roots, but may cause unnatural root growth patterns due to confinement. Moradi et al. (unpublished data) reported good root visibility in some members of the Fabaceae and Asteraceae families, whereas resolution was insufficient in Brassicacae, Solonaceae, and Poaceae. In principle, larger species such as small trees could be investigated using NR. However, this would require a larger container for nonperturbed growth. Consequently, the spatial resolution would decrease and one could only resolve large structural roots.

The water content of the soil and the plant should be monitored so that the soil water content is low enough at the time of imaging to provide adequate contrast yet not so low that the plant becomes water stressed. If a series of measurements are to be taken, then one can calculate the soil θ in the zone of interest from the results of each radiograph. This requires scattering correction and a water quantification algorithm calibrated for the particular plant–soil system. Water quantification using NR has the advantage over a gravimetric measurement because it can be used to determine the water content of the soil in the zone of interest, rather than providing an average value of a soil profile that may have a heterogeneous moisture distribution.

The neutron beam formed in the collimator is not perfectly homogeneous. This also holds for the detector system. Therefore, there are spatial variations in intensity of the resulting radiography that are caused by the beam, rather than the sample. This requires that each radiograph be corrected by normalizing the image by a "flat field" or open-beam image with no sample. Similarly, noise generated by the camera assembly should be removed. A

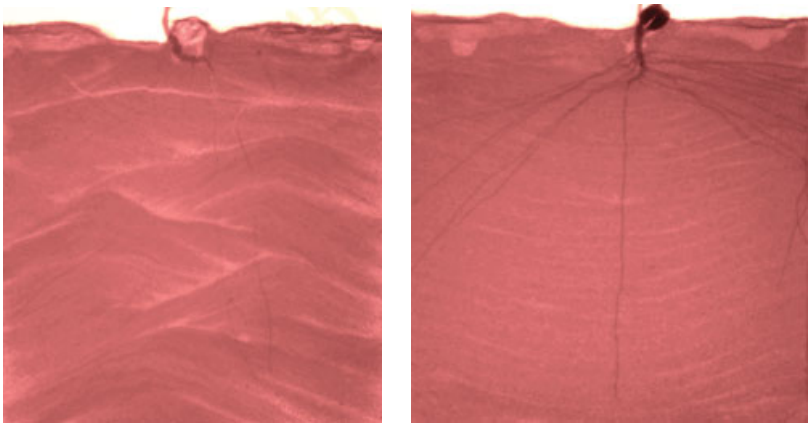

(a) (b)

Figure 4 Soil heterogeneity resulting from filling patterns can affect the image quality. The radiographs show the roots of *Cicer arietinum* in identical soil with high heterogeneity (a) and low heterogeneity (b).

standard first step in image analysis is normalization, i.e. to transform recorded raw information relative values:

$$I' = NORM\frac{I_{raw} - I_{dark}}{I_{openbeam} - I_{dark}} \quad (4)$$

where I_{raw} is the image as registered by the camera, I_{dark} is the dark noise image without the beam, $I_{openbeam}$ contains the spatial field variation of the beam without object, and NORM is a factor to bring the resulting value into the valid range of the image-processing tools.

The background gray-level intensity within a neutron radiograph, caused, for example, by variations in the soil water content, may vary so much that a single global threshold cannot differentiate the roots satisfactorily. Menon et al.[5] overcame this problem by modifying an algorithm originally developed for resolving blood vessels in retinal images.[18]

6 APPLICATION

The application of NR can enhance the study of root development and root–soil interactions by revealing the location of roots over time without disturbance. Plant root development is a function of the plant species and the nature of the soil into which they penetrate. Soil components, such as organic matter, nutrients, and contaminants, occur heterogeneously. NR is an ideal tool to study how roots interact with patches of low or high concentration. When roots encounter a patch or discontinuity in soil, they may proliferate, wither, or continue growing unaffected. Figure 5(a) and (b) shows how NR can reveal plant responses to a patch of nickel, a toxic heavy metal, in soil. Such experiments, for example, may reveal species that avoid contaminant hotspots and reduce the risk of plant contaminant uptake. Similarly, understanding the mechanisms by which crop plants interact with patches of nutrients may aid the development of treatments to improve crop, and subsequently human, nutrition.

Perhaps more importantly, NR can quantify the spatial distribution and flux of water in the plant–soil system in near real time, in combination with root imaging. This permits the study of root water uptake and the effect of roots on the passage of water through soil. Figure 6 shows a series of images detailing water infiltration into a profile and water uptake by plant roots. Such studies have a wide variety of potential applications, such as selecting vegetation to minimize leaching from contaminated sites and the optimization of irrigation and fertilization systems.

7 CONCLUSIONS AND OUTLOOK

The quantification of root mass and soil water content using NR of plant–soil interactions has considerable scope for improvement, particularly, the processing of raw images obtained at the neutron facility. Refinement of root segmentation algorithms described by Menon et al.[5] would greatly enhance the accuracy and precision of the technique. One obvious drawback of NR is that, at present, it requires access to a specialized facility, of which there are only a few available worldwide with the desirable performance. Competition for beam time is fierce, since NR finds applications in many fields of science. However, the usefulness of NR indicates that it may follow the same pattern of development as synchrotron radiation facilities. Initially, financial

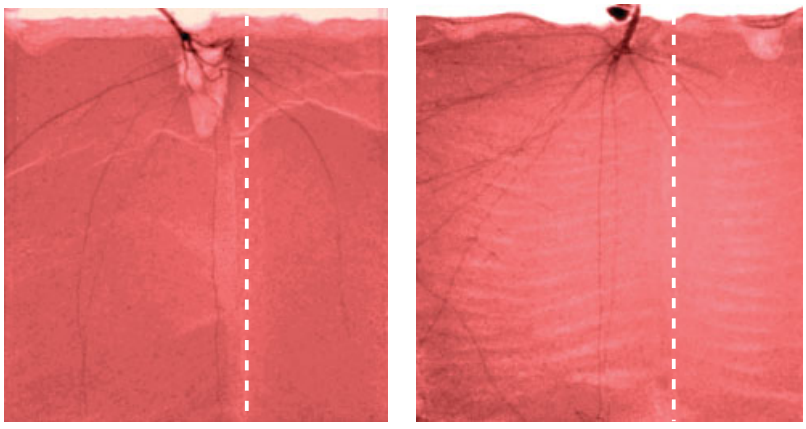

Figure 5 Neutron radiographs of two 150 mm × 150 mm × 120 mm slabs filled with sandy loam. The area to the right of the dotted line was spiked with 125 mg kg^{-1} Ni. The roots of *Cicer arietinum* (b) avoid the high-Ni zone, while the roots of *Berkheya coddii* (a) are unaffected. Neither plant showed any differences in the aboveground portions relative to their respective controls.

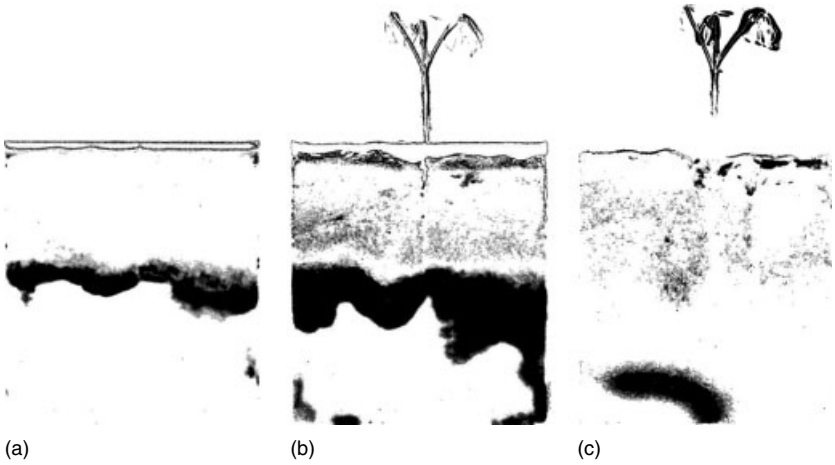

Figure 6 The change in water content (both positive and negative) of a 150 mm × 150 mm × 120 mm container, in which *Lupinus albus* was growing, after the infiltration of 10 mL of water ((a) 1–2 min, (b) 2–5 min, (c) 28–40 min). The wet front shows up as a dark band, while root water uptake is visible in b and c as a discontinuous gray area.

and technical constraints limited their application, but their usefulness ensured subsequent technological development and capital injection, and thus they became commonplace and the standard equipment for some analyses.

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Nuclear Reaction Analysis • Prompt γ-Neutron Activation Analysis • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis • Soil Instrumental Methods

REFERENCES

1. R.O. Kuchenbuch, K.T. Ingram, 'Image Analysis for Non-destructive and Non-invasive Quantification of Root Growth and Soil Water Content in Rhizotrons', *J. Plant Nutr. Soil Sc.*, **165**(5), 573–581 (2002).
2. P.J. Gregory, D.J. Hutchison, D.B. Read, P.M. Jenneson, W.B. Gilboy, E.J. Morton, 'Non-invasive Imaging of Roots with High Resolution X-ray Micro-tomography', *Plant Soil*, **255**(1), 351–359 (2003).
3. S.T. Willatt, R.G. Struss, H.M. Taylor, 'In situ Root Studies Using Neutron Radiography', *Agron. J.*, **70**(4), 581–586 (1978).
4. E.H. Lehmann, G. Frei, G. Kuhne, P. Boillat, 'The Micro-setup for Neutron Imaging: A Major Step Forward to Improve the Spatial Resolution', *Nucl. Instrum. Methods A*, **576**(2–3), 389–396 (2007).
5. M. Menon, B. Robinson, S.E. Oswald, A. Kaestner, K.C. Abbaspour, E. Lehmann, R. Schulin, 'Visualization of Root Growth in Heterogeneously Contaminated Soil Using Neutron Radiography', *Eur. J. Soil Sci.*, **58**(3), 802–810 (2007).
6. A. Real, S. Sundell-Bergman, J.F. Knowles, D.S. Woodhead, I. Zinger, 'Effects of Ionizing Radiation Exposure on Plants, Fish and Mammals: Relevant Data for Environmental Radiation Protection', *J. Radiol. Prot.*, **24**(4A), A123–A137 (2004).
7. S. Kasperl, P. Vontobel, 'Application of an Iterative Artefact Reduction Method to Neutron Tomography', *Nucl. Instrum. Meth. A*, **542**(1–3), 392–398 (2005).
8. F.B. Salisbury, C.W. Ross, *Plant Physiology*, Wadsworth, Belmont, CA, 1992.
9. B. Mason, C. Moore, *Principles of Geochemistry*, John Wiley & Sons, Hong Kong, 1982.
10. PSI (2008) '*Comparison Between Neutron and X-ray Attenuation*', from http://neutra.web.psi.ch/images/diag_x_n_e.gif.
11. D.G. Allis, B.S. Hudson, 'Inelastic Neutron Scattering Spectrum of Cs-2[B12H12]: Reproduction of its Solid-state Vibrational Spectrum by Periodic DFT', *J. Phys. Chem. A*, **110**(10), 3744–3749 (2006).
12. R.K. Hassanein, *Correction Methods for the Quantitative Evaluation of Thermal Neutron Tomography*, ETH Zurich, Switzerland, 2006.
13. T.M. Nakanishi, M. Matsubayashi, 'Nondestructive Water Imaging by Neutron Beam Analysis in Living Plants', *J. Plant Physiol.*, **151**(4), 442–445 (1997).

14. J. Furukawa, T.M. Nakanishi, H. Matsubayashi, 'Neutron Radiography of a Root Growing in Soil with Vanadium', *Nucl. Instrum. Meth. A*, **424**(1), 116–121 (1999).

15. T.M. Nakanishi, Y. Okuni, J. Furukawa, K. Tanoi, H. Yokota, N. Ikeue, M. Matsubayashi, H. Uchida, A. Tsiji, 'Water Movement in a Plant Sample by Neutron Beam Analysis as well as Positron Emission Tracer Imaging System', *J. Radioanal. Nucl. Chem.*, **255**(1), 149–153 (2003).

16. U. Matsushima, Y. Kawabata, M. Hino, P. Geltenbort, B.M. Nicolai, 'Measurement of Changes in Water Thickness in Plant Materials Using Very Low-energy Neutron Radiography', *Nucl. Instrum. Meth. A*, **542**(1–3), 76–80 (2005).

17. T.M. Nakanishi, Y. Okuni, Y. Hayashi, H. Nishiyama, 'Water Gradient Profiles at Bean Plant Roots Determined by Neutron Beam Analysis', *J. Radioanal. Nucl. Chem.*, **264**(2), 313–317 (2005).

18. A. Hoover, V. Kouznetsova, M. Goldbaum, 'Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response', *IEEE Trans. Med. Imag.*, **19**(3), 203–210 (2000).

Nuclear Reaction Analysis

Guy Demortier
Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium

1 Introduction: Role of Nuclear Reaction Analysis in Comparison with Other Ion Beam Analysis Techniques 1
2 Fundamentals of Nuclear Reactions 4
3 Kinematics of Nuclear Reactions 5
4 Identification of Signals for Analytical Purposes 8
5 Examples of Application 11
5.1 Analysis of Hydrogen in Various Materials 11
5.2 Depth Profiles of F, Na in SnO_2 Coated Glass 14
5.3 Stoichiometry of Y–Ba–Cu–O Superconductors 15
5.4 Three-dimensional Microanalysis of a Gold–Silicon Eutectic Alloy 16
5.5 Determination of Traces of Sulfur in Ancient Gold Artefacts 19
5.6 Particle-induced X-ray Emission, Particle-induced γ-ray Emission and Nuclear Reaction Analysis for Elemental Analysis of a Biological Reference Material 23
Abbreviations and Acronyms 23
Related Articles 23
References 23

Nuclear reaction analysis (NRA) is one among the numerous analytical techniques involving the irradiation of a material with energetic particles. In NRA methods we exclude any remnant radioactivity together with interactions leading to the scattering of the incident particle only: this last procedure leads to RBS (Rutherford backscattering spectroscopy) and ERD (elastic recoil detection).

In NRA the nature of the emitted particle used as an analytical signal is obviously different from that of the incident one. In order to avoid permanent radioactivity, the energy of the projectile is maintained at a sufficiently low value (below 6 MeV, easily available using a small accelerator). Consequently, due to the large coulomb barrier around heavy nuclei, only light nuclei may be easily identified (atomic mass less than or equal to 30).

After discussing the kinematics and fundamentals of the interactions, we present examples of applications in material sciences, archaeological and biological materials.

1 INTRODUCTION: ROLE OF NUCLEAR REACTION ANALYSIS IN COMPARISON WITH OTHER ION BEAM ANALYSIS TECHNIQUES

When a material is irradiated with charged particles like protons, deuterons, α-particles or other heavier ions, at energies ranging from a few hundred kiloelectron volts to a few megaelectron volts, the emission of photons, charged particles and neutrons gives rise to various spectroscopic methods of elemental analysis. The great majority of interactions of the incident charged particles with atoms of the irradiated material take place with atomic electrons: the incident particle progressively loses its energy ($-\mathrm{d}E$) in the material. The energy loss $\mathrm{d}E$ is expressed by Bethe's law of stopping power $S(E)$ Equation (1):

$$S(E) = -\frac{\mathrm{d}E}{\rho\,\mathrm{d}x} = K\frac{z^2 m}{E_\mathrm{i}}\frac{Z}{A}\ln\frac{aE}{I} \qquad (1)$$

where z, m and E_i are respectively the charge, mass and the energy of the incident particle, where ρ, I, Z and A, are respectively the density, mean ionization potential, atomic number and the atomic mass of the material, and a is an adjustable parameter. K comprises only physical constants. When the incident particle has sufficient energy to ionize atoms from any shell, $S(E)$ decreases with the increase in the incident energy (Table 1). Exceptions to this general rule occur at low incident particle energy, as seen for He^4 and He^3 ions in Table 1, when the projectile cannot eject inner shell electrons, the main consequence being that $Z_\mathrm{eff} < Z$. Therefore $S(E)$, in those cases, is smaller than the expected value.

The incident particle comes to rest after crossing a distance R, known as the range. This incident particle does not experience an appreciable deviation when interacting with atomic electrons. Interaction with one atomic nucleus only rarely causes deviation of the incident particle from a straight line. The range is almost the total distance that the particle travels following a straight line in the material. The range (R) increases with the incident projectile energy ($R \sim E^{1.8}$), and decreases with an increase in its charge and its mass. Table 2 gives some data on ranges of protons, deuterons, ^{3}He and α-particles (R) for a selection of light, medium and high Z elements.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Stopping power of light ions in materials in keV mg^{-1} cm^{-2}

Energy (MeV)	C	Al	Cu	Ag	Au
			Protons		
0.2	5.9 (E + 2)	3.6 (E + 2)	2.0 (E + 2)	1.5 (E + 2)	1.1 (E + 2)
0.5	3.7 (E + 2)	2.5 (E + 2)	1.6 (E + 2)	1.1 (E + 2)	7.6 (E + 1)
1	2.3 (E + 2)	1.7 (E + 2)	1.2 (E + 2)	8.7 (E + 1)	6.0 (E + 1)
2	1.4 (E + 2)	1.1 (E + 2)	8.1 (E + 1)	6.3 (E + 1)	4.5 (E + 1)
3	1.0 (E + 2)	8.4 (E + 1)	6.3 (E + 1)	5.0 (E + 1)	3.7 (E + 1)
5	7.0 (E + 1)	5.8 (E + 1)	4.4 (E + 1)	3.7 (E + 1)	2.8 (E + 1)
7	5.4 (E + 1)	4.5 (E + 1)	3.5 (E + 1)	2.9 (E + 1)	2.3 (E + 1)
			Deuterons		
0.2	7.5 (E + 2)	4.2 (E + 2)	2.4 (E + 2)	1.7 (E + 2)	1.2 (E + 2)
0.5	5.4 (E + 2)	3.4 (E + 2)	1.9 (E + 2)	1.4 (E + 2)	9.5 (E + 1)
1	3.7 (E + 2)	2.5 (E + 2)	1.6 (E + 2)	1.1 (E + 2)	7.6 (E + 1)
2	2.3 (E + 2)	1.7 (E + 2)	1.2 (E + 2)	8.7 (E + 1)	6.0 (E + 1)
3	1.8 (E + 2)	1.4 (E + 2)	9.4 (E + 1)	7.2 (E + 1)	5.1 (E + 1)
5	1.2 (E + 2)	9.5 (E + 1)	7.0 (E + 1)	5.6 (E + 1)	4.0 (E + 1)
7	9.3 (E + 1)	7.5 (E + 1)	5.7 (E + 1)	4.6 (E + 1)	3.4 (E + 1)
			^{3}He		
0.2	2.1 (E + 3)	1.2 (E + 3)	7.0 (E + 2)	5.2 (E + 2)	3.7 (E + 2)
0.5	2.2 (E + 3)	1.3 (E + 3)	7.8 (E + 2)	5.6 (E + 2)	4.0 (E + 2)
1	1.8 (E + 3)	1.2 (E + 3)	7.0 (E + 2)	5.1 (E + 2)	3.5 (E + 2)
2	1.2 (E + 3)	8.7 (E + 2)	5.6 (E + 2)	4.1 (E + 2)	2.9 (E + 2)
3	9.3 (E + 2)	7.0 (E + 2)	4.7 (E + 2)	3.5 (E + 2)	2.5 (E + 2)
5	6.5 (E + 2)	5.0 (E + 2)	3.6 (E + 2)	2.8 (E + 2)	2.0 (E + 2)
7	5.0 (E + 2)	4.0 (E + 2)	2.9 (E + 2)	2.3 (E + 2)	1.7 (E + 2)
			^{4}He		
0.2	2.0 (E + 3)	1.1 (E + 3)	6.4 (E + 2)	4.7 (E + 2)	3.5 (E + 2)
0.5	2.3 (E + 3)	1.4 (E + 3)	7.7 (E + 2)	5.7 (E + 2)	4.0 (E + 2)
1	2.0 (E + 3)	1.3 (E + 3)	7.5 (E + 2)	5.4 (E + 2)	3.8 (E + 2)
2	1.4 (E + 3)	1.0 (E + 3)	6.2 (E + 2)	4.5 (E + 2)	3.1 (E + 2)
3	1.1 (E + 3)	8.2 (E + 2)	5.3 (E + 2)	3.9 (E + 2)	2.7 (E + 2)
5	8.0 (E + 2)	6.1 (E + 2)	4.2 (E + 2)	3.2 (E + 2)	2.2 (E + 2)
7	6.2 (E + 2)	4.9 (E + 2)	3.5 (E + 2)	2.7 (E + 2)	1.9 (E + 2)

Note that the particle range increases with the atomic weight of the material owing to a dependence on Z/A.

During this slowing down of the projectile, a statistical transfer of energy to the atomic electrons of the material takes place. This statistical behavior of the energy loss of the projectile gives rise to energy straggling: an unequal energy loss from identical particles crossing the same target material under identical conditions. The relation between the energy loss and the achieved depth is therefore subject to some inaccuracy. For a very small energy loss, straggling can be calculated using the Landau and Vavilov theory.[1] Energy straggling is not symmetrical around the average value. For monoenergetic protons of 1 MeV crossing a thickness of 2.4×10^{-5} g cm^{-2} ($\approx$90 nm) of pure Al, the mean energy loss is 4.2 keV and straggling on this energy loss has a width at half maximum of 2.7 keV (Figure 1a).[2] For 10^{-3} g cm^{-2}, the energy loss of 1 MeV protons is 400 keV and straggling gives rise to a Gaussian distribution with a width at half maximum of 27 keV (Figure 1b). The relation between energy and depth achieved in a material is then defined in this case with an accuracy of 7%, which is relatively much lower than the inaccuracy at low incident energy: the greater the energy loss, the smaller the statistical uncertainty in the relation between depth with respect to loss. In most practical cases achievable by NRA, the depth profiling analysis is limited to regions of a few hundred nanometers, where the Landau and Vavilov calculation theory is valid, to a few micrometers where the Gaussian profile is effective. In this last case, the straggling, δ, is simply given by Equation (2):

$$\delta = 0.93z\sqrt{\frac{Z}{A}}\sqrt{x} \tag{2}$$

where x is the depth achieved in μm, z is the atomic number of the incident particle, Z and A are the atomic number and the atomic weight of the target, respectively.[3,4] Straggling is then the same for all particles of equal z (deuteron or proton, α or ^{3}He). The dependence of straggling on the energy of the particle is

Table 2 Range of light ions in materials in mg cm^{-2}

Energy (MeV)	C	Al	Cu	Ag	Au
			Protons		
0.2	3.1 (E − 1)	5.4 (E − 1)	9.4 (E − 1)	1.3 (E 0)	1.8 (E 0)
0.5	9.6 (E − 1)	1.5 (E 0)	2.6 (E 0)	3.7 (E 0)	5.2 (E 0)
1	2.7 (E 0)	3.9 (E 0)	6.4 (E 0)	8.8 (E 0)	1.3 (E + 1)
2	8.4 (E 0)	1.1 (E + 1)	1.7 (E + 1)	2.3 (E + 1)	3.2 (E + 1)
3	1.7 (E + 1)	2.2 (E + 1)	3.1 (E + 1)	4.0 (E + 1)	5.7 (E + 1)
5	4.1 (E + 1)	5.1 (E + 1)	7.0 (E + 1)	8.8 (E + 1)	1.2 (E + 2)
7	7.3 (E + 1)	9.1 (E + 1)	1.2 (E + 2)	1.5 (E + 2)	2.0 (E + 2)
			Deuterons		
0.2	3.2 (E − 1)	5.8 (E − 1)	9.9 (E − 1)	1.3 (E 0)	1.8 (E 0)
0.5	7.9 (E − 1)	1.4 (E 0)	2.4 (E 0)	3.3 (E 0)	4.6 (E 0)
1	1.9 (E 0)	3.1 (E 0)	5.3 (E 0)	7.3 (E 0)	1.0 (E + 1)
2	5.4 (E 0)	7.9 (E 0)	1.3 (E + 1)	1.8 (E + 1)	2.5 (E + 1)
3	1.0 (E + 1)	1.4 (E + 1)	2.2 (E + 1)	3.0 (E + 1)	4.4 (E + 1)
5	2.4 (E + 1)	3.2 (E + 1)	4.7 (E + 1)	6.2 (E + 1)	8.8 (E + 1)
7	4.4 (E + 1)	5.6 (E + 1)	7.9 (E + 1)	1.1 (E + 2)	1.4 (E + 2)
			^{3}He		
0.2	2.2 (E − 1)	3.8 (E − 1)	6.1 (E − 1)	7.9 (E − 1)	1.0 (E 0)
0.5	3.3 (E − 1)	5.7 (E − 1)	9.6 (E − 1)	1.3 (E 0)	1.7 (E 0)
1	5.7 (E − 1)	9.4 (E − 1)	1.6 (E 0)	2.2 (E 0)	3.0 (E 0)
2	1.2 (E 0)	1.9 (E 0)	3.2 (E 0)	4.3 (E 0)	6.1 (E 0)
3	2.2 (E 0)	3.2 (E 0)	5.1 (E 0)	6.9 (E 0)	9.9 (E 0)
5	4.8 (E 0)	6.6 (E 0)	1.1 (E + 1)	1.3 (E + 1)	1.9 (E + 1)
7	8.3 (E 0)	1.1 (E + 1)	1.6 (E + 1)	2.1 (E + 1)	3.0 (E + 1)
			^{4}He		
0.2	2.6 (E − 1)	4.5 (E − 1)	7.3 (E − 1)	9.3 (E − 1)	1.2 (E 0)
0.5	3.7 (E − 1)	6.5 (E − 1)	1.1 (E 0)	1.4 (E 0)	1.9 (E 0)
1	5.9 (E − 1)	9.9 (E − 1)	1.7 (E 0)	2.3 (E 0)	3.1 (E 0)
2	1.2 (E 0)	1.8 (E 0)	3.1 (E 0)	4.2 (E 0)	6.0 (E 0)
3	1.9 (E 0)	2.9 (E 0)	4.8 (E 0)	6.6 (E 0)	9.3 (E 0)
5	4.0 (E 0)	5.8 (E 0)	9.1 (E 0)	1.2 (E + 1)	1.7 (E + 1)
7	6.9 (E 0)	9.5 (E 0)	1.4 (E + 1)	1.9 (E + 1)	2.7 (E + 1)

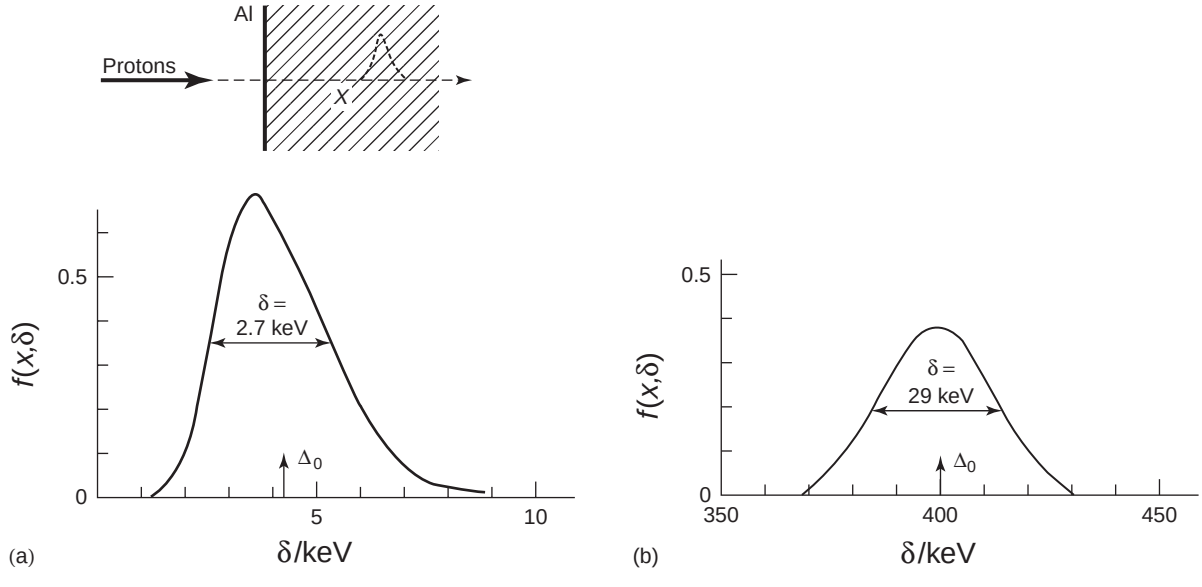

Figure 1 Energy straggling of 1 MeV protons for an energy loss of 4.2 keV (a) and 400 keV (b) after crossing a thickness of 2.4×10^{-5} g cm^{-2} (a) or 2×10^{-3} g cm^{-2} (b) in Al. (Adapted from G. Deconninck,[2] page 72.)

totally included in the value of x. For compound materials, Z and A are simply replaced by the mean values of Z and A.

The interaction of the incident particle with the atomic electrons of the target induces ionization. When ionization takes place in an inner electronic shell, subsequent rearrangement in order to return to atomic equilibrium leads to the emission of characteristic X-rays: the analytical technique based on this atomic process is called PIXE (particle-induced X-ray emission) (see **PIXE (Particle-induced X-ray Emission)**). PIXE is a very sensitive method for bulk and trace analysis of elements with an atomic number Z greater than 20.[5] PIXE may be also used to study elements from Na to K, but the absorption of low energy characteristic X-rays (1–3 keV) from these light elements into thick samples generally restricts the potential of the method. Analysis of low Z elements would then only be satisfactory for thin samples or to probe these elements in a very narrow layer at the surface of a bulk material.

When the incident particle passes close to the atomic nucleus, it may be elastically scattered. At very low impact parameter, a particle of mass m lower than the mass, M, of the collided nucleus, is scattered at an angle close to 180°. The spectroscopy of this backscattered particle (RBS) may easily be used for depth profiling of elements up to thicknesses close to 30% of their own range. The energy of the scattered particle at an angle close to 180° is related to the mass of the collided nucleus, Equation (3)

$$E_{\text{scatt}} = \left(\frac{M-m}{M+m}\right)^2 E_{\text{i}} \qquad (3)$$

and to the depth in the material at which the scattering took place. This depth can be determined using the stopping power of the target material. Furthermore, as the cross-section for such elastic scattering is proportional to the Z^2 of the target atoms, RBS is particularly powerful for the study of high Z elements (see **Rutherford Backscattering Spectroscopy**).

The probability of coulomb excitation of a target nucleus leading to the emission of a characteristic γ-rays[6] is much lower than for the backscattering and inner shell ionization phenomena. Consequently, PIGE (particle-induced γ-ray emission) is of particular interest for the study of light elements (see **Particle-induced γ-Ray Emission**) where PIXE and RBS suffer from a lack of sensitivity and accuracy. For these light target elements, the characteristic X-rays are of such low energy that they are absorbed into the material itself and RBS offers a much lower cross-section (because it is proportional to Z^2). When using heavy projectiles, the forward elastic scattering of light target nuclei give rise to another method of analysis called ERD (see **Elastic Recoil Detection Analysis**). If the energy of the incident particle is maintained below 5–10 MeV, no delayed radioactivity may be induced as would be the case in CPAA (charged particle activation analysis) typically performed at incident energies greater than 10 MeV (see **Charged Particle Activation Analysis**). By using all these analytical techniques simultaneously (or sequentially), powerful characterization of many materials may be achieved. In sections 5.3 and 5.6 of this article, we will discuss several complementary applications of nuclear reactions with other ion beam analysis (IBA) techniques.

2 FUNDAMENTALS OF NUCLEAR REACTIONS

In discussing NRA we will concentrate on prompt reactions leading to a transmutation of the target atom and then excluding RBS (see **Rutherford Backscattering Spectroscopy**), ERD (see **Elastic Recoil Detection Analysis**) and part of PIGE (see **Particle-induced γ-Ray Emission**) where the target nucleus does not change during the interaction. We will also exclude delayed γ-ray emission from a transmuted nucleus (see **Charged Particle Activation Analysis**).

The concise form of a nuclear reaction as expressed by nuclear physicists is

$$\text{A (a, bc) B}$$

where A is the target nucleus, a the incident projectile, b (and possibly c) the emitted particles or photons, B the residual nucleus which is generally a heavy particle carrying a very low energy which remains in the target material and cannot be detected.

In this respect, a reaction named

$$^{27}\text{Al(p, }\gamma)^{28}\text{Si}$$

would represent the capture of a proton in an ^{27}Al nucleus leading to the formation of an excited ^{28}Si which immediately decays to a stable ^{28}Si by the emission of a cascade of γ-rays. The spectroscopy of these γ-rays may be used as analytical signals.

The reaction

$$^{15}\text{N(p, }\alpha\gamma)^{12}\text{C}$$

would be induced by protons on ^{15}N leading to the production of a ^{12}C nucleus in an excited state after the emission of an α-particle. The spectroscopy of these γ-rays emitted by the excited ^{12}C or/and the spectroscopy of the emitted α-particle could be used for analytical purposes.

The interactions of a loosely bound compound projectile (such as a deuteron) give rise to the emission of energetic protons or neutrons after stripping of the incident deuteron. The breakdown of the incident deuteron (stripping reaction) may be written as

$$^{12}C(d, p_0)^{13}C \quad \text{or} \quad ^{12}C(d, p_i\gamma)^{13}C^*$$

The former represents the stripping of an incident deuteron by a ^{12}C nucleus, leading to the capture of a neutron in ^{12}C to produce a ^{13}C nucleus in its fundamental state. The emitted proton has a maximum energy $E(p_0)$. Various other ways could leave the residual nucleus of ^{13}C in one of its ($i = 1, 2, 3, \ldots$) excited states. The corresponding energies $E(p_i)$ of the p_i protons are then lower than $E(p_0)$. The stored energy in the excited $^{13}C^*$ nuclei will shortly afterwards appear as a cascade of characteristic γ-rays of this residual nucleus of ^{13}C. The energy of the emitted protons $p_0, p_1, p_2, p_3, \ldots$ is quantized and the differences in their energies corresponds to the various quantum states of the residual nucleus. Several groups of protons of different energies, as well as γ-rays, may be used for analytical purposes. Stripping of a deuteron with the capture of the proton is also possible. A neutron is then emitted, but in practice the spectroscopy of these emitted neutrons is much less easy than the spectroscopy of charged particles or photons, so its use in NRA is very limited.

3 KINEMATICS OF NUCLEAR REACTIONS

A fundamental parameter of a nuclear reaction is the Q value. It represents the balance between the total rest mass of interacting particles and those of emitted ones, Equation (4):

$$Q = (m_A + m_a - m_B - m_b)c^2 \tag{4}$$

where c is the velocity of light in vaccum. A positive Q value is a characteristic of a nuclear reaction in which some mass is converted into kinetic energy of emitted particles. This Q value added to the incident particle energy E_a is shared by both outgoing particles B and b, Equation (5):

$$Q = E_b + E_B - E_a \tag{5}$$

The partition between E_B and E_b may be calculated by using conservation laws: conservation of total momentum because the interaction takes place without any influence of external forces; conservation of total mass–energy because the isolated system formed by the incident particle and the target nucleus is being considered. In terms of E_a and E_b, one finds Equation (6):

$$Q = E_b\left(1 + \frac{m_b}{m_B}\right) - E_a\left(1 - \frac{m_a}{m_B}\right) - \frac{2}{m_B}\left[E_a E_b m_a m_b\right]^{1/2}\cos\theta \tag{6}$$

When b is the lightest emitted particle, its energy E_b in the direction θ relative to the incident beam of particles a is given by Equation (7):

$$E_b^{1/2} = \frac{\cos\theta(m_a m_b E_a)^{1/2} \pm \{m_a m_b E_a \cos^2\theta + (m_B + m_b)[m_B Q + (m_B - m_a)E_a]\}^{1/2}}{m_B + m_b} \tag{7}$$

As $E_b^{1/2}$ must be a positive quantity, the plus sign must only be kept if the following condition is met:

$$(m_B + m_b)|m_B Q + (m_B - m_a)E_a| > 0$$

This is always the case for positive Q values. For negative Q values, the minimum E_a energy giving rise to the emission of b at $\theta = 0°$ is given by Equation (8):

$$E_{a\,\min} = \|Q\|\frac{m_B + m_b}{m_B + m_b - m_a} \tag{8}$$

For higher E_a energies, two different energies may be attributed to the same particle b emitted at the same angle θ but in a limited energy range of E_a and only for the ejection of b in the forward direction. We will ignore this special behavior which could be inconvenient for elemental analysis.

A list of selected reactions for the study of light isotopes ($A \leq 20$) is given in Table 3. Resonant reactions which may be used for depth profiling (see sections 5.1 and 5.2 below) are particularly interesting and a suggestion for the best E_a incident energies is given.

For hydrogen, helium and lithium isotopes in a material, the detection of coincidental b and B particles may sometimes be used to solve chemical interference problems (superposition of signals from particles emitted with the same energy but arising from different B nuclei). As E_b energies are also dependent on the emission angle θ, a specific geometrical assembly may be required for solving interference problems.

The geometry of a nuclear reaction used for analysis in thick target materials is given in Figure 2. The incident particle "a" hits the sample at point I, crosses a thickness $x/\cos\alpha_i$, losing some of its energy (mainly by interaction with atomic shell electrons) and induces a nuclear reaction at P. The emitted particle with an energy $E_b(\theta)$ to be calculated for $E_a = E_a(0) - \Delta E(x/\cos\alpha_i)$ is produced at P and has to cross the distance $x/\cos\alpha_o$ before emerging from the sample surface at O. The outgoing energy at O is equal to $E_b(\theta) - \Delta E_b\,(x/\cos\alpha_o)$.

Table 3 Useful nuclear reactions for the analysis of light elements (hydrogen to fluorine)[a]

Element	Reaction	Q (MeV)	Recommended energy for depth profiling (MeV)	Refs.
Hydrogen	$D(d,p)T$	4.033		8
	$D(^3He,p)^4He$	18.352	0.64	9
	$D(^3He,\alpha)^1H$	18.352	0.64	
Helium	$^3He(d,p)^4He$	18.352	0.40	9
	$^3He(d,\alpha)^1H$	18.352	0.40	
Lithium	$^6Li(p,\alpha)^3He$	4.02		10
	$^6Li(p,^3He)^4He$	4.02		
	$^6Li(d,\alpha)^4He$	22.374		11
	$^6Li(^3He,p)^8Be$	18.786		12
	$^7Li(p,\alpha)^4He$	17.347		13
Beryllium	$^9Be(p,\alpha)^6Li$	2.125	0.33	14
	$^9Be(d,t)^8Be$	4.496		15
	$^9Be(d,\alpha)^7Li$	7.153		16
	$^9Be(^3He,p)^{11}B$	10.322		17
	$^9Be(^3He,\alpha)^8Be$	18.912		18
	$^{10}Be(^3He,p)^{12}C$	19.693		12
Boron	$^{10}B(d,\alpha)^8Be$	17.818		19
	$^{10}B(\alpha,p)^{13}C$	4.063	1.567	20, 21
	$^{11}B(p,\alpha)^8Be$	8.582	0.163 and 0.650	22, 23
	$^{11}B(^3He,p)^{13}C$	13.185		24
	$^{11}B(^3He,d)^{12}C$	10.464		24
	$^{11}B(\alpha,p)^{14}C$	0.784	Higher than 2.5 MeV owing to low Q value	21, 25
Carbon	$^{12}C(d,p)^{13}C$	2.722		26
	$^{12}C(^3He,p)^{14}N$	4.779		27
	$^{12}C(^3He,\alpha)^{11}C$	1.856		28
	$^{13}C(d,p)^{14}C$	5.951		29
Nitrogen	$^{14}N(d,p)^{15}N$	8.610		30
	$^{14}N(d,\alpha)^{12}C$	13.574		31, 32
	$^{15}N(p,\alpha)^{12}C$	4.966	0.429	33
	$^{15}N(d,\alpha)^{13}C$	7.687		34
Oxygen	$^{16}O(d,p)^{17}O$	1.917		35
	$^{16}O(d,\alpha)^{14}N$	3.11		36
	$^{16}O(^3He,p)^{18}F$	2.032		27, 37
	$^{16}O(^3He,\alpha)^{15}O$	4.914		37
	$^{18}O(p,\alpha)^{15}N$	3.980	0.629	11, 38
	$^{18}O(d,\alpha)^{16}N$	4.247		38
Fluorine	$^{19}F(p,\alpha)^{16}O$	8.114	0.340; 1.347	39
	$^{19}F(d,p)^{20}F$	4.374		40
	$^{19}F(d,\alpha)^{17}O$	10.031		40
	$^{19}F(\alpha,p)^{22}Ne$	1.675	2.45	41

[a] For applications of these nuclear reactions to elemental analysis see Demortier.[60]

The angle θ is simply $\alpha_i + \alpha_o$. As the most intense fraction of the emitted particles are elastically scattered a particles (see **Rutherford Backscattering Spectroscopy**), some calibrated absorber has to be inserted between the surface sample and the detector in order to stop them all and then to accept only b particles induced by a true nuclear reaction, when the target nucleus A is transmuted to another one, B. Reactions with large Q values and detection of b particles of mass lower than the mass of a are particularly useful characteristics of nuclear reactions suitable for analysis. A thin absorber will stop all the scattered a particles but will not greatly affect the energy of the lighter emitted b particles (see Equation 1).

In special cases corresponding to a large Q value and a low E_a energy, one may expect to detect b particles selectively with a higher mass than the incident particle a even after crossing the dedicated absorber. Reactions $^6Li(d,\alpha)^4He$, $^7Li(p,\alpha)^4He$, $^9Be(^3He,\alpha)^8Be$, $^{10}B(d,\alpha)^8Be$,

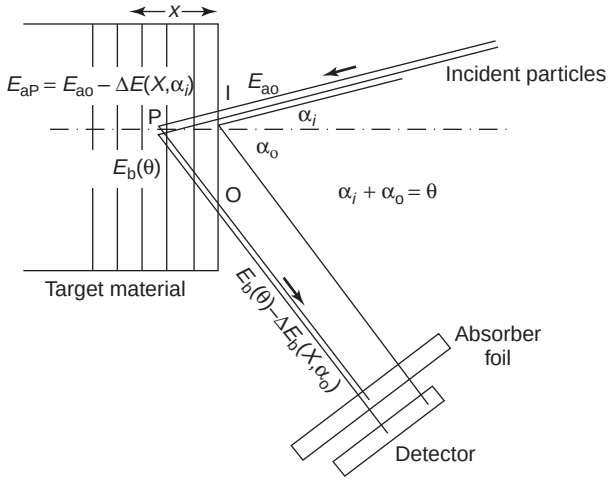

Figure 2 Parameters involved in the study of thick materials by the nuclear reaction A(a, b)B.

$^{14}N(d,\alpha)^{12}C$, and $^{19}F(p,\alpha)^{16}O$ shown in Table 3 would offer this possibility for an incident E_a energy in the region of 1 MeV, as would $^{11}B(p,\alpha)^{8}Be$, $^{18}O(p,\alpha)^{15}N$, $^{19}F(p,\alpha)^{16}O$, when using incident protons in the range of a few hundred kiloelectron volts.

The energetic scale of the compound nucleus C* made up with all the nucleons of A and a (and then also B and b) may be represented schematically (Figure 3).

Quantum states have their own energy width $\Gamma = \Delta E$ (and their associated lifetime ΔT related by $\Gamma \Delta T \geq \hbar$). Low-lying states (bound states) can only decay through γ-ray emissions: their width is of the order of 10^{-3} eV. At much higher excitation energies, known as positive energy states, the width is typically of the order of 0.1–100 keV. These levels may be formed by several typical configurations of nucleons, in particular (A + a), (B + b) or (C_i^*). Each of these configurations is associated with a typical width Γ_a, Γ_b, Γ_γ. The sum of all possible configurations Γ_i is related to the total width Γ: the higher Γ, the shorter the lifetime of the level.

When the incident particle a is sent on the nucleus A with a precise energy E_{ar}, the probability of forming an excited nucleus is greater than for any other E_a energy slightly greater or lower than E_{ar}. For particles of energy E_{ar}, the probability of the emission of a particle b (and/or a cascade of characteristic γ-rays) is also enhanced.

Owing to competing phase shifts in the quantum mechanical behavior of nuclear reactions, the cross-section may sometimes be very low in the vicinity of a resonant energy. The discontinuities in the cross-section values of nuclear reactions induced by a on A give rise to some potential for depth profiling of A nuclei: signals induced in the detector by b or c particles will be particularly intense when arising from those regions in the target material where $E_a = E_{ar}$. If a material is irradiated with particles a of energy greater than E_{ar}, the region where E_a will reach this particular energy E_{ar} will be achieved at a defined depth below the surface. This depth may be calculated by using the $S(E_a)$ values of

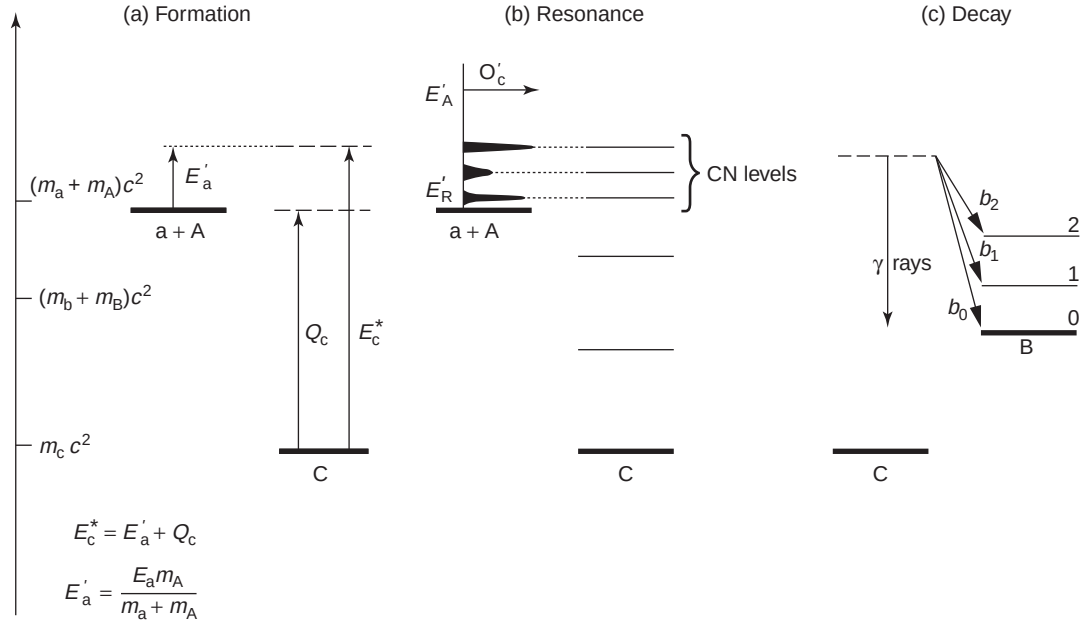

Figure 3 Mass–energy scheme of a resonant nuclear reaction. The formation in various levels of an excited compound nucleus (CN) (b) by the interaction of an incident particle a on a target nucleus A takes place only for discrete incident E'_a energies (E'_a calculated in the center of mass system). Particles b of various energies (b_0, b_1, ...) are consequently emitted to give rise to a final nucleus B which could appear in various states (c). (From G. Deconninck, LARN.[2])

Table 1 and the corresponding detected E_b values may be calculated by using Equation (7) and the necessary correction for the energy loss in the outgoing direction (see Figure 2) which is to be calculated by using the corresponding $S(E_b)$ values of Table 1.

The depth resolution is directly related to the total width of the resonance: the narrower the resonance, the more precise will be the depth resolution. Useful resonant energies E_{ar} for depth profiling are limited for several reasons:

1. The cross-section at the resonant energy must be intense in comparison with its value outside the resonance.
2. The resonant energy must be well isolated in order to avoid superposition of signals originating from various regions below the sample surface where several resonant energies may be reached.
3. The resonance width must be of the order of (or less than) the energy resolution of the whole experimental set-up (taking into account the accelerator energy resolution, detector resolution, detector solid angle accepting particles emitted with various θ and then various energies ...).
4. The energy straggling of the incident particles into a buried layer under the surface must be sufficiently small, as should be the energy straggling of the emitted particles along its outgoing path and into the necessary absorber (inserted between the target and the detector to stop elastically scattered a particles).

The detection of emitted γ-rays from the residual nucleus does not suffer from the last uncertainty (straggling in energy of the emitted particle) but γ-rays may be affected by Doppler shift when the emitting residual nucleus decays before being stopped in the material. Doppler broadening of emitted γ-rays only arises if the lifetime of the recoiling C^* or B^* nuclei is shorter than the time required to stop this recoil nucleus in the sample target.

4 IDENTIFICATION OF SIGNALS FOR ANALYTICAL PURPOSES

Modern particle detectors of protons, deuterons, ^{3}He and α-particles with energies in the megaelectron volt range are now exclusively solid-state detectors: p–n junctions made with crystals of silicon.[7]

The full energy of the detected particle is converted into pairs of electrons and holes: the number of pairs is strictly proportional to the deposited energy. A reverse polarization of the p–n junction is maintained by an adjustable bias across the desired detector depth, which is at least the range of the particles of maximum energy to be detected. This range may vary from 5 μm for 1-MeV α-particles to 0.6 mm for 7-MeV protons. The energy resolution (width at half maximum of a peak induced by monoenergetic particles) is generally in the order of 8–15 keV for the whole range of these useful energies.

For the highest energies, a succession of detectors mounted as a telescope may be necessary. The associated electronics of these telescopes is tailored in order to collect all the signals induced in the successive detectors coincidentally. For energetic particles, the time delay between the signals produced in two distant thin detectors can be taken into account by measuring the time between the two short signals; in this last case, the time-of-flight of the particle is used. For distant detectors, this time interval is used to identify the mass of the particle only. A third detector, or a set of successive detectors which stop the particle completely, is necessary to determine the energy of the charged particle. The resolution of time-of-flight systems is indeed insufficient to measure the energy of the energetic b particles accurately.

The kinetic energy (converted into a proportional signal in the solid-state detector and possibly the time-of-flight) is a characteristic quantity which allows experimentalists to identify the nucleus which was participating in the nuclear reaction. If this nuclear reaction took place at the surface of the sample and if the emitted particle were directly collected in the backward direction, the energy of this detected particle would be simply given by Equation (7), where E_a is the incident energy delivered by the particle accelerator. In contrast if the reaction took place below the surface, the incident particle reaches this region with a lower energy $E'_a < E_a$. E'_a can be calculated by using the target stopping power along the entrance pathlength $x/\cos\alpha_1$ and E'_b can be recalculated using the same Equation (7). The emitted b particle would also lose some of its energy along the outgoing trajectory at a distance $x/\cos\alpha_o$ which can also be calculated using the appropriate stopping power of the material for this particle b. Dividing the whole distribution of signals from $E_{b\,max}$ to $E_{b\,min}$ into energy intervals with a width of the order of the energy resolution of the experimental set-up (including energy straggling for both incident and emitted particles), it is possible to determine the depth profile of thicknesses ranging from 10 nm to a few microns. The experimentalist must have the ability to certify the origin of signals and to exclude eventual interfering elements (see Figure 4).

Figure 4 shows that particles with identical energies may be emitted with the same incident energy E_a and at the same angle θ by two different A nuclei. Changing

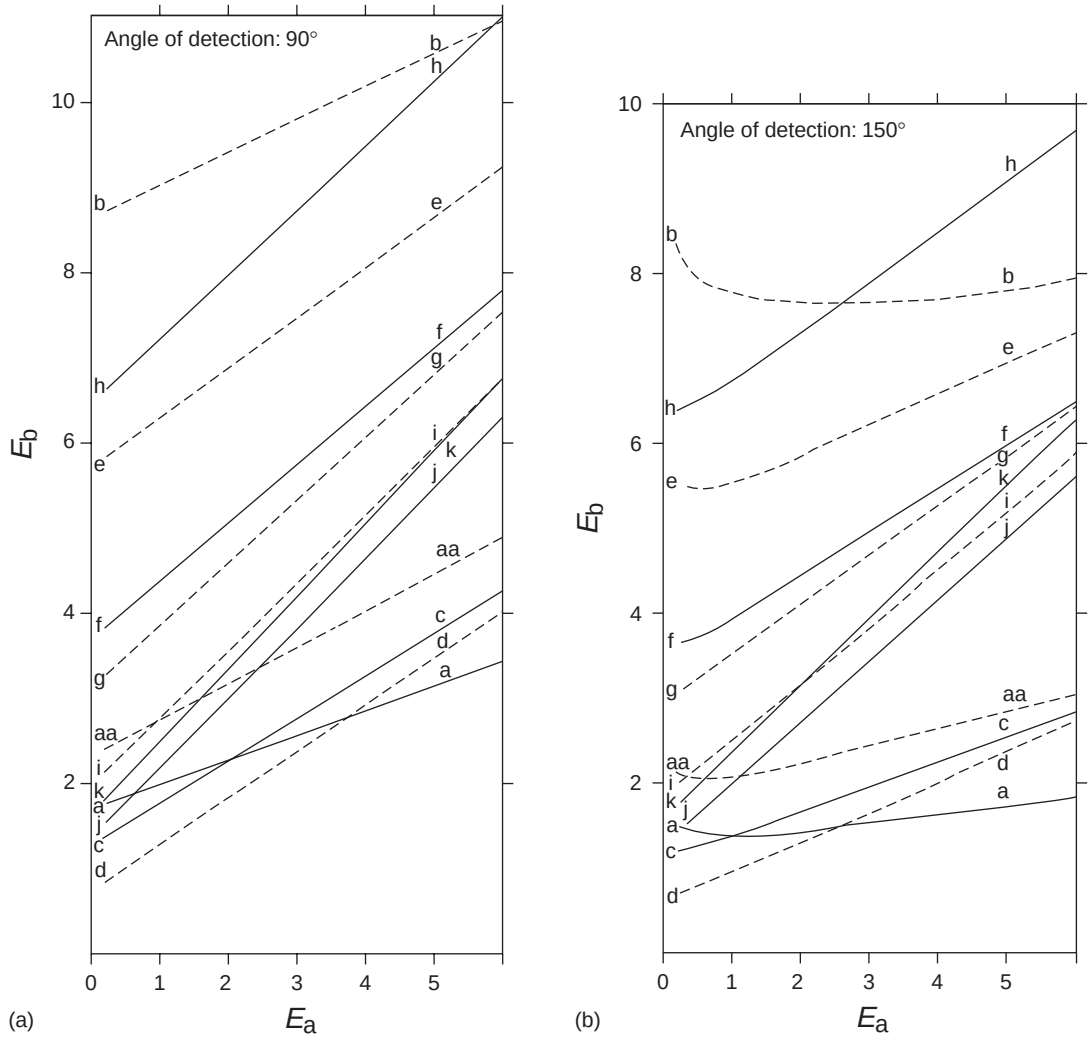

Figure 4 (a) Relation between the energy E_b of emitted α-particles at $\theta = 90°$ and the incident energy E_a for the most useful nuclear reactions induced by protons on ^{6}Li(a), ^{7}Li(b), ^{9}Be(c), ^{10}B(d), ^{11}B(e), ^{15}N(f), ^{18}O(g), ^{19}F(h), ^{23}Na(i), ^{27}Al(j) and ^{31}P(k). Line aa refers to reaction ^{6}Li(p,^{3}He)α which appears simultaneously with reaction ^{6}Li(p,α)^{3}He. (b) Relation between the energy E_b of emitted α-particles at $\theta = 150°$ and the incident energy E_a for the most useful nuclear reactions induced by protons on ^{6}Li(a), ^{7}Li(b), ^{9}Be(c), ^{10}B(d), ^{11}B(e), ^{15}N(f), ^{18}O(g), ^{19}F(h), ^{23}Na(i), ^{27}Al(j) and ^{31}P(k). Line aa refers to reaction ^{6}Li(p,^{3}He)α which appears simultaneously with reaction ^{6}Li(p,α)^{3}He.

the energy E_a or the angle of detection are alternatives that may be used to deal with the interferences. It may be helpful to use two detectors at two different angles simultaneously for complicated analyses: interferences between reactions at two different A nuclei do not take place at both angles (see Figure 4a and b).

In addition to the measurement of the energy of b particles which identify the target nuclei, the measurement of the intensity of signals with identical energies gives the concentration of these nuclei of interest. The intensity of such signals is governed by the cross-section of the particular nuclear reaction: the greater the cross-section, the better the capability of the nuclear reaction to perform accurate (due to statistics) and sensitive elemental determinations. Knowledge of all characteristics of the nuclear reaction (presence of resonances, angular distributions of emitted particles and elimination of interfering reactions by selecting the best angle of detection) is of prime importance to achieving confident analyses.

All these data may be found in various nuclear data tables[42–45] and handbooks.[46,47]

For γ-ray detection, large volume (40–100 cm^3) Ge(Li) photon detectors of high energy resolution (1.2 keV at 1.5 MeV) are used. They are cooled at liquid nitrogen temperature. Various processes of interaction (photoelectric effect, Compton effect, pair creations) give rise to an extended spectrum from which only the full

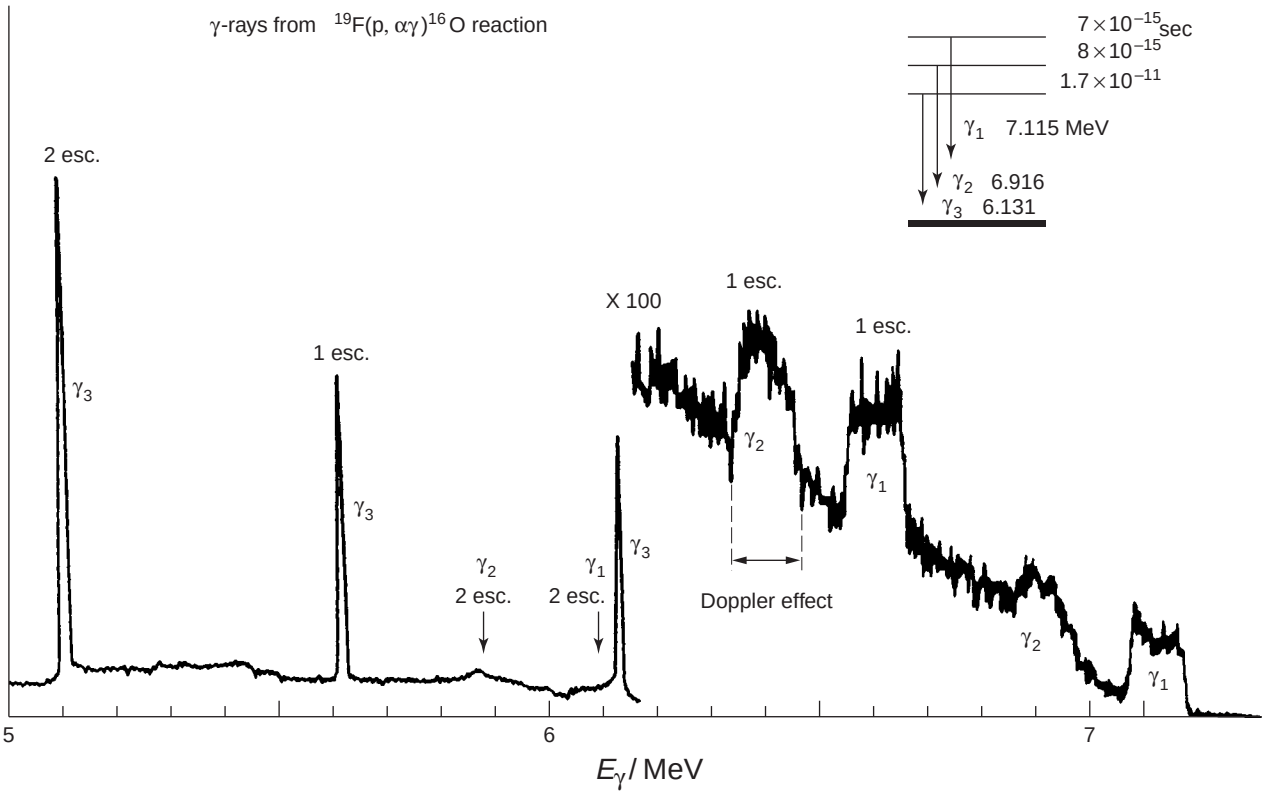

Figure 5 High-energy part of a γ-ray spectrum obtained with a solid-state Ge(Li) detector during the irradiation of a material containing fluorine. The γ_1, γ_2 and γ_3 peaks are produced by (p,αγ) reaction leading to excited ^{16}O. Both γ_1 and γ_2 peaks are broadened by the Doppler effect because of the short lifetime of the corresponding levels of the oxygen nucleus which decays before being at rest in the material. The sharp γ_3 peaks do not suffer this broadening. Each characteristic γ-ray gives rise to signals in three regions: the full energy peak and two escape peaks corresponding to the loss of one or two of the annihilation photons (511 keV) produced after the initial pair creation. (From G. Deconninck, LARN.[2])

energy peak is selected for γ-ray identification. The efficiency of such a detector is quite high for γ-rays of 2 MeV or less but may also be used at higher energies (Figure 5).

An alternative would be a large NaI or BGO (bismuth germanate oxide) detector (about 1000 cm^3) with a very high detection efficiency but with a much lower energy resolution. This kind of detector is particularly useful for depth profiling using (p,γ) or (p,αγ) resonant reactions involving γ-ray energies of more than 2 MeV. In contrast to particle spectroscopy whose energy is a complicated function depending on $E_a, \theta, x/\cos\alpha_1$ and $x/\cos\alpha_2$, γ-ray spectroscopy only concerns measurement of the energy of the excited state of the promptly decaying excited B^* nucleus. To profile one element which could give rise to a resonant nuclear reaction (see Table 3) at a particular E_{ar} energy, the E_a energy is progressively increased to allow this resonant reaction to take place at various depths below the surface. For increasing values of $(E_a - E_{ar})$ the resonant reaction would produce γ-rays at depths calculated using the stopping power (Table 1) of the incident particle a only.

Examples of depth profiling using γ-ray or particle spectroscopy will be given in sections 5.1 and 5.6 of this article.

The analytical investigation of large specimens or fragile materials which cannot be introduced in the vacuum sometimes requires the use of external beam geometry. This technique is now widely familiar to PIXE (see **PIXE (Particle-induced X-ray Emission)**) and PIGE users (see **Particle-induced γ-Ray Emission**) but may be also tailored for NRA when the outgoing particle has a sufficiently high energy (i.e. for (d,p) reactions) by comparison with the energy of the incident particle. The incident beam crosses a thin foil of metal or of mylar before reaching the sample situated outside the vacuum vessel of the accelerator. The distance in the air between the target and the detector may be finely tuned in order to stop the elastically backscattered a particles completely and allow only b particles to reach this detector. The energy straggling of this incident particle in the exit window and in the air, and the energy straggling of the emitted particle in the sample and also in the air does not allow, in this particular case, depth profiling to be performed.

5 EXAMPLES OF APPLICATION

Many laboratories in the world use NRA in the field of materials science. We have selected a few typical results. The first concerns the study of hydrogen distribution in the depth of materials by using a methodology involving the detection of γ-rays produced by irradiation of energetic ^{15}N ions to induce a resonant nuclear reaction at various depths in the sample. General problems related to energy straggling, cross-sections and performances of γ-ray detectors will be discussed.

The second results concern the depth profiling of F and Na in SnO_2 coated glass (an insulating material) using simultaneous RBS and NRA techniques at two particular resonant proton energies for (p,$\alpha\gamma$) reactions. The third application benefits from simultaneous PIXE and NRA for a complete characterization of all the components of a YBaCuO superconductor using a microbeam of low energy deuterons.

We will then present the use of a (d,p) reaction for the study of microregions in the grain boundaries of Au-Si alloys and for the determination of traces of sulfur in ancient gold jewellery artefacts. Nuclear reactions are indeed not sensitive to high Z nuclei owing to the high coulomb barrier, but they offer a unique technique (free of interference) for studying light elements in heavy matrices. The last application concerns the determination of all the major and minor elements and a majority of trace elements (in the region of a few micrograms per gram) in a reference material of biological interest.

5.1 Analysis of Hydrogen in Various Materials

The use of megaelectron volt ion beams to analyze materials for hydrogen has been more interesting and more important than even the early researchers in this field realised. The reasons for this are simple:

1. Hydrogen is probably the most common contaminant element, especially in thin film materials.
2. Hydrogen has important effects on the chemical, physical and electrical properties of many materials.
3. The analysis for hydrogen is difficult or impossible by most traditional analytical methods.

Megaelectron volt ion beams to probe for hydrogen may be used in two different approaches. The first is NRA, in which megaelectron volt ions bombard a sample, inducing nuclear reactions between the incident ion and hydrogen in the target. The second is ERD in which megaelectron volt ions bombard a sample and through elastic collisions cause hydrogen ions to recoil out of the sample with the number of these recoils used to determine the amount of hydrogen in the target. This last problem (of ERD) is treated in **Elastic Recoil Detection Analysis**. The basic principle for hydrogen analysis using NRA is to use nuclear reactions between protons and light nuclei but in the reverse direction: light nuclei are incident particles and hydrogen is the collided nucleus.

While many nuclear reactions are possible for hydrogen analysis, the reactions most commonly used to probe for ^{1}H in material are those induced by ^{15}N and ^{19}F.[43] These reactions are similar in many ways: they both have low-energy isolated resonances above a few megaelectron volts. Another reaction which is capable of profiling much deeper into samples uses a ^{7}Li beam. The important parameters for most useful reactions for H analysis are given in Table 4.

The use of the ^{15}N reaction as a probe for hydrogen is given here as an illustration of this approach. The nuclear reaction is:

$$H(^{15}N, \alpha\gamma)^{12}C$$

The very narrow and intense resonance is shown in Figure 6. The γ-ray intensity at the resonant energy is four orders of magnitude higher than the one off-resonance.

A beam of ^{15}N ions is incident on the sample being analyzed and the yield of characteristic γ-rays of ^{12}C is measured. Because this is a resonant reaction, the cross-section is large at the resonance energy (E_{res}) and small off-resonance. Therefore, if a sample is bombarded with ^{15}N ions at the resonance energy, the yield of γ-rays is proportional to the hydrogen only present at the surface

Table 4 Nuclear reaction cross-section characteristics

Reaction	$^{7}Li + H$ (ref. 49)	$^{15}N + H$ (ref. 50)	$^{15}N + H$ (ref. 50)	$^{19}F + H$ (ref. 50)	$^{19}F + H$ (ref. 50)
Resonance energy (MeV)	3.07	6.385	13.35	6.418	16.44
Cross-section (σ_0) at resonance (mb)	4.8	1650	1050	88	440
Resonance width (Γ) (keV)	81	1.8	25.4	44	86
$\sigma_0\Gamma$(mb keV)	389	2970	26 700	3870	37 800
Relative yield	0.13	1.000	9.0	1.3	12.7
Energy of next resonance (MeV)	7.11	13.35	18.0	9.1	17.6
γ-Ray energy (MeV)	17.7, 14.7	4.43	4.43	6.13, 6.98, 7.12	6.13, 6.98, 7.12

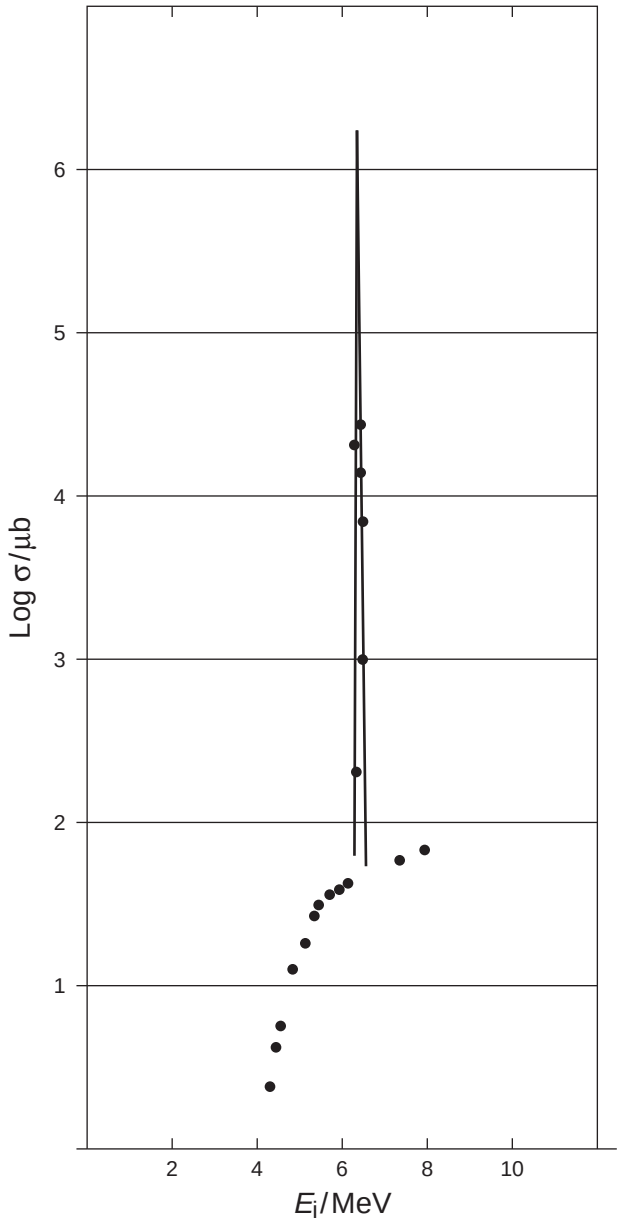

Figure 6 Variation of the cross-section of the reaction $H(^{15}N,\alpha\gamma)^{12}C$ for increasing ^{15}N incident ions. The cross-section increases by at least four orders of magnitude at the resonance energy ($E_i = 6.385$ MeV). (Data are taken from Horn and Lanford.[51])

of the sample. If the beam energy is raised above the resonant energy, there is no longer a reaction with surface hydrogen (because the ^{15}N ions are above the resonant energy) but as the ions lose energy when penetrating the sample, they reach the resonant energy at a particular depth: the yield of γ-rays is then proportional to the hydrogen content at this depth.

Characteristic γ-rays from the reaction are detected by a large (and therefore efficient) scintillation detector (NaI), or a BGO detector, located close behind the irradiated samples.

The γ-ray yield is proportional to the energy integrated area under the Breit–Wigner formulae for a resonant cross-section, i.e. proportional to $\sigma_0\Gamma$, where σ_0 is the cross-section at the resonance energy and Γ is the width of the resonance. The yield of γ-rays is also inversely proportional to the energy loss of the incident ions. This is because the yield of γ-rays is proportional to the amount of H per cubic centimeter within the resonance detection window (i.e. when the energy is on-resonance) and this is proportional to the thickness of the resonance detection window. This thickness is inversely proportional to dE/dx.

As in all resonant nuclear reaction profiling, a measurement of the number of γ-ray signals induced by the nuclear reactions versus incident energy is made. A known standard material containing a known quantity of hydrogen at a well-defined depth is used for quantitative determination of depth profiles in the material to be studied.

The matrix of this reference material may be different from the matrix of the material under investigation. All the matrix effects are contained in dE/dx which is matrix-dependent.

The depth resolution (Δx) of the method depends on the measured full width at half maximum (fwhm) of the hydrogen profile, for an infinitely thin layer of hydrogen at a certain depth in the sample. This is then determined by the width of the reaction resonance (Γ) and the fluctuations in the energy (ΔE) of the bombarding ions at this depth in the sample. ΔE has contributions from the energy spread in the incident proton beam (ΔE_b), which depends on the performance of the accelerator, energy straggle (ΔE_s) (see Equation 2) and, for detection of γ-rays, the Doppler energy broadening (ΔE_D). Since in most cases these are all Gaussian distributions, the total energy width is obtained by adding them in quadrature, Equation (9):

$$\Delta E^2 = \Delta E_b^2 + \Delta E_s^2 + \Delta E_D^2 \qquad (9)$$

For the depth profile at the extreme surface, the energy straggling is not Gaussian and another more sophisticated procedure is involved.[52] Some additional explanation is given for Doppler effect. Hydrogen atoms in materials are not stationary but vibrate. For a narrow nuclear resonance, this motion can shift the reaction on- and off-resonance. For example, the width of the ^{15}N resonance is 1.8 keV, whereas the Doppler broadening is in the range 5–15 keV, depending on the strength of the hydrogen bond. This contribution is nevertheless usually small compared to straggle, except when analyzing very thin hydrogen surface layers where the Doppler effect is the dominant source of broadening.

The ideal nuclear reaction for profiling would be one with an infinitely narrow resonance reaction with a zero cross-section off-resonance. The ^{15}N nuclear reaction is nearly this ideal. The off-resonance cross-section is more than four orders of magnitude smaller than the peak resonance cross-section. For other resonance reactions, the off-resonance cross-sections are generally much larger.

However, even in this favorable case, there are situations where the nonzero off-resonance cross-section needs to be taken into account. Thus, a layer of TiH_2 on clean Si was profiled.[51] The results of this profile are shown in Figure 7. The counts above 7.2 MeV essentially all come from the off-resonance cross-section.

The correction for such off-resonance effects is simple. Since the off-resonance yield is proportional to the on-resonance energy γ-ray yield (integrated between 6.4 and 7.2 MeV in Figure 6) times the off-resonance cross-section, once the proportionality constant is known, these off-resonance counts are subtracted from the total counts. In the case of ^{15}N NRA, this correction is Equation (10):

$$Y_{\text{off res}} = (1.28 \times 10^{-5})I \tag{10}$$

where I is the energy integrated on-resonance yield (in counts keV).[51]

This correction is simply due to the fact that no other resonance at higher energy (the next one being at $E_{\text{inc}} = 18$ MeV) in the $^{15}N + p$ reaction gives a contribution.

The problem of sensitivity or limit of detection is discussed now. When making NRA measurements where the reaction product is a high-energy γ-ray, the detectors are also sensitive to background radiations, such as cosmic rays and, in some cases, accelerator-generated radiation. This background is the main factor in the limitation of the sensitivity of NRA. There are several ways in which sensitivity can be improved, the first being by high beam current. Since the external background is independent of the beam current whereas the real count rate increases linearly with this beam current, sensitivity improves linearly with beam current. However, this approach can be used only up to the limit of the beam current available in a particular accelerator or the limit of beam current

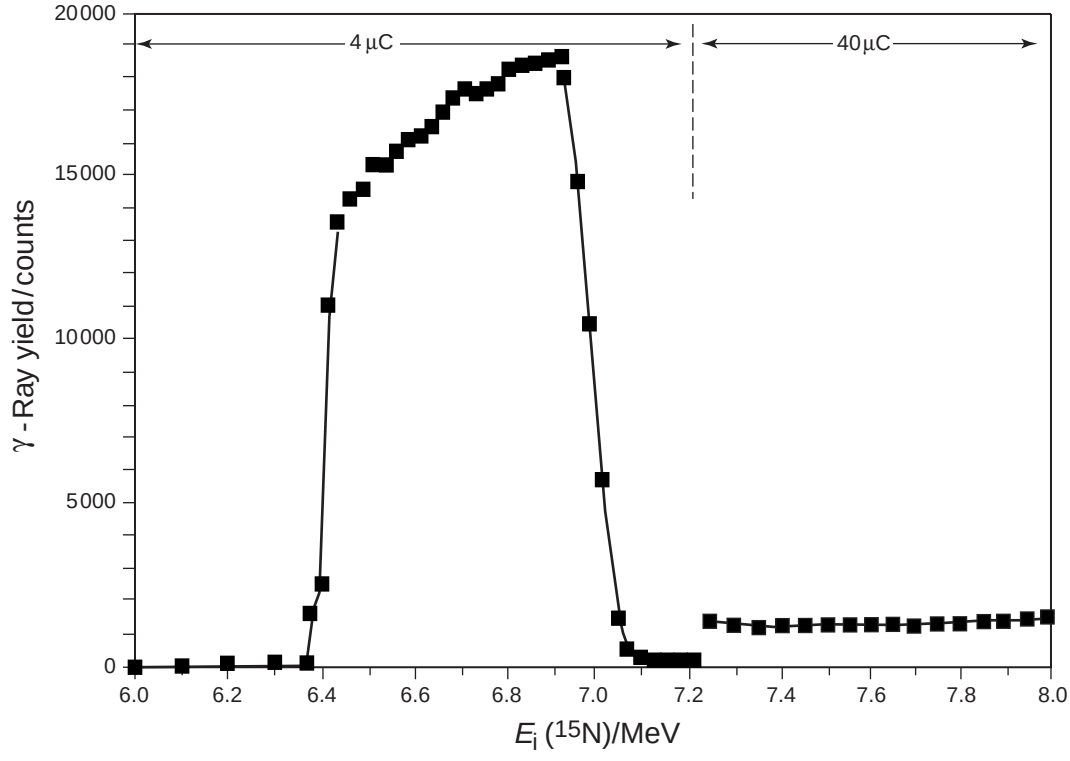

Figure 7 Use of the resonant energy of ^{15}N ions for hydrogen depth profiling. The sample is a coating of TiH_2 on a Si substrate. The sharp increase in the γ-ray yield appears when the incident ^{15}N energy reaches the resonant energy: the production of γ-rays takes place at the sample surface. When the ^{15}N ions incident energy exceeds 7 MeV, the resonant energy is only reached in the Si substrate. At energies higher than 7.2 MeV, the number of incident ^{15}N ions was increased by a factor of ten to appreciate the continuum (out-resonance) counting rate. (Reprinted from W. Lanford, 'Analysis for Hydrogen by Nuclear Reaction and Energy Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 65–82. Copyright (1992), with permission from Elsevier Science.[48])

that a particular sample will withstand. The last limitation may be avoided by rastering the beam along a large area on the surface for samples where the surface homogeneity is guaranteed.

Another approach is to build material and electronic shields to reduce the background in the γ-ray detectors. This approach is now much easier with the advent of bismuth germinate BGO scintillation detectors which are progressively replacing the usual NaI scintillator. BGO detectors are much more efficient (for a given size) than NaI and are much less sensitive to neutrons generated by cosmic rays.

In good cases, sensitivities of the order of micrograms per gram and depth resolutions of a few nanometers are possible. However, the practical limit achievable is dependent on the nature of the sample being analyzed and on the quality of the accelerator (energy resolution, beam current intensity, etc.) and its associated equipment.

Historically, NRA has been used in many studies of hydrogen in materials. However, because of the wide availability of new facilities for He ERD and because of the potential of high-energy heavy ion ERD, this alternative method is becoming more popular (see **Elastic Recoil Detection Analysis**).

Nuclear reactions may also be used for the study of the other isotopes of hydrogen. Some of the reactions that have been used include: ^{2}H(d,p)T, ^{2}H(d,n)^{3}He, ^{3}He(d,p)^{4}He, ^{3}H(d,n)^{4}He and ^{3}H(p,n)^{3}He. For all reactions giving rise to neutrons, the analytical technique involves the detection of the associated charged particle.

The most widely used reaction for deuterium analysis is the D(^{3}He,p)^{4}He reaction[53] which has a positive Q-value (18.35 MeV) making identification of the reaction products easy and the elimination of interfering signals comfortable using a coincidence technique with the detection of both particles simultaneously emitted: p and ^{4}He. Depth information is obtained by the energy loss of the outgoing ^{4}He particles. The cross-section for this reaction is not large (70 mb at 0.75 MeV) compared to the elastic cross-section (see **Elastic Recoil Detection Analysis**), so count rate problems are of concern.

5.2 Depth Profiles of F, Na in SnO_2 Coated Glass

A typical application of simultaneous RBS and resonant NRA is the study of fluorine and sodium migration in a SnO_2 layer deposited on a glass substrate.[54] The depth profiles are measured by proton-induced resonant reactions. The depth calibration which is essential to localize the interface is done by proton backscattering on Sn. As the SnO_2 layer is thin (200 nm), the same stopping power may be used for incoming protons and outgoing elastically scattered protons on Sn, the heaviest element in the sample. The fwhm of the Sn signals is therefore twice the actual energy loss (ΔE_p) of the incoming protons in the SnO_2 layer. ΔE_p then gives the exact position of the interface of SnO_2 with the bulk glass in terms of energy, without using the stopping power parameters. RBS is therefore used to calibrate the resonant reaction excitation curves.

Profiling of sodium is carried out by using the ^{23}Na(p,$\alpha\gamma$)^{20}Ne at the resonance energy $E_R = 1010.5$ keV (resonance width $\Gamma_R < 0.5$ keV). γ-Rays of 1634 keV are detected in a 4 in $\times$ 4 in well NaI detector offering maximum efficiency. The target sample is irradiated into the well of the NaI detector. As it is essential to reduce the irradiation duration to prevent sodium migration induced by the beam, the beam is rastered on the sample in order to avoid local overheating of the glass. The experimental excitation curve is displayed in Figure 8(b). A small surface concentration is observed followed by a steep edge at $E_p = E_R + 14.6$ keV, and then a constant concentration of 16 atom% (the bulk concentration of Na in glass). The RBS spectrum was collected simultaneously. Figure 8(a) shows this RBS spectrum for the SnO_2 layer on glass under the same experimental conditions as used for the profiling, i.e. $E_p = 1.015$ MeV. The width $2\Delta E_p$ of the Sn signal as measured on the RBS spectrum is 29.2 keV (twice 14.6 keV) and can be related directly to the step in the Na depth profile. The exact localization of the concentration step with respect to the interface is therefore known directly, without any calculation using physical parameters such as the stopping power of the Sn and O mixture. Therefore, any other element that could be present in the layer and whose measurement is not accessible by either RBS or NRA methods will not disturb the measurement.

Profiling of fluorine is done using the ^{19}F(p,$\alpha\gamma$)^{16}O at resonance energy $E_r = 340$ keV ($\Gamma_R = 2.4$ keV). The γ-rays in the range of 6–8 MeV are also detected using the large NaI scintillator. The excitation curve is displayed in Figure 8(d) showing a shallow concentration followed by a huge peak localized in the SnO_2 layer, close to the interface with the glass substrate, and just in front of the sodium step. The energy loss $\Delta E'_p$ corresponding to the interface was, once again, measured by the backscattering spectrum which was collected simultaneously in the same way as in Figure 8(c). Once more, the width of the Sn peak is twice the energy loss of the protons measured by the new RBS spectrum. Using these two resonant nuclear reactions, it was then possible to localize the relative position of (i) the fluoride layer (completely in the SnO_2 surface layer), (ii) the interface between SnO_2 and the glass, and (iii) the sodium step which is located entirely in the glass substrate.

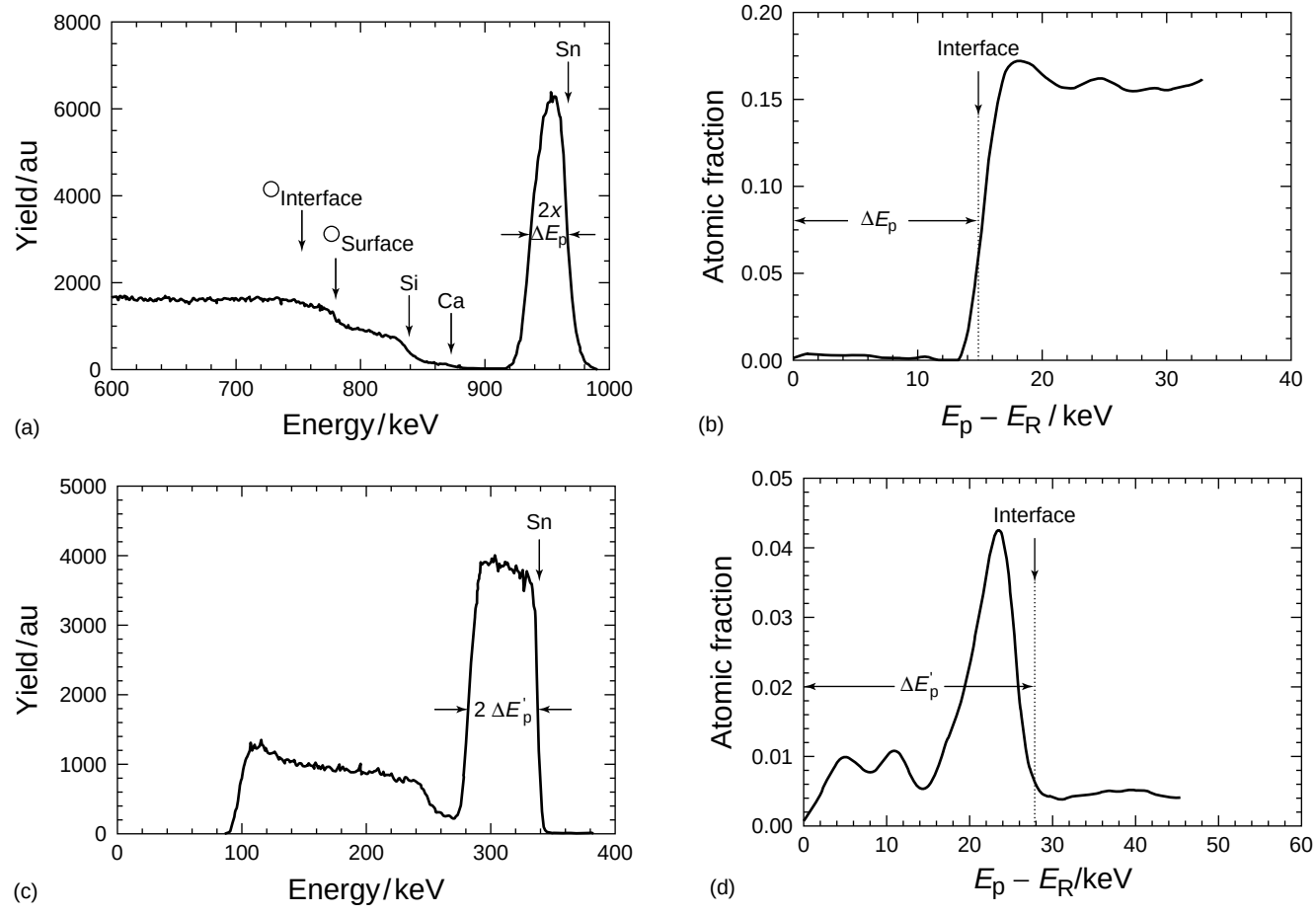

Figure 8 Determination of Na profile by the $^{23}Na(p,\alpha\gamma)^{20}Ne$ reaction around the resonant energy (1010.5 keV). The Sn thickness was determined at 1 MeV by RBS (a) and the step in the γ-ray production indicated that Na is only present in the glass substrate behind the SnO_2 coating (b). Similarly, the F profile is determined by the $^{19}F(p,\alpha\gamma)^{16}O$ reaction around the resonant energy at 340 keV. The corresponding RBS spectrum at 340 keV (c) gives the same SnO_2 thickness as in (a) and the F profile (d) indicates that this element is mainly present in the SnO_2 coating. (Reprinted from G. Terwagne, G. Deconninck, 'Concentration Profile of Light Elements Near an Interface', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 153–155, Copyright (1992), with permission from Elsevier Science.[54])

5.3 Stoichiometry of Y–Ba–Cu–O Superconductors

In addition to crystallographic characterization and measurement of thermal and electrical properties of superconducting materials, rapid and nondestructive analytical techniques involving milliprobe facilities were developed to search for the best conditions for their synthesis and to test the stoichiometric composition of $YBa_2Cu_3O_{6.5+x}$.

Irradiation of pellets made from a powder superconductor with 2 MeV deuterons may give, at the same time, the relative bulk concentration of all the elements.[55] A beam of deuterons is focused on a narrow region (5 μm) of the pellet. A solid-state particle detector, situated at 135° relative to the incident beam, collects protons from (d,p) reactions on O and C. A low carbon content is always possible from the precursors used in the synthesis ($BaCO_3$) or from a deposit on the sample surface by the incident deuteron beam of some residual atoms from oil used in the pumping set-up. The elastically scattered deuterons are stopped in a mylar foil of appropriate thickness inserted between the irradiated surface and the detector. Simultaneously, a Si(Li) detector collects X-rays, L X-rays of Ba and K X-rays of Cu and Y which are used as analytical signals for the rest of the elements. A second solid-state detector is installed at 170° relative to the incident beam and is used to collect the elastically backscattered deuterons in order to check the homogeneity of the Y, Ba and Cu distribution at depth. A flat plateau of Ba, Y and Cu in the RBS spectrum would certify this homogeneity. The deuteron beam is rastered on the surface. The integrated time of the whole measurement is of the order of 1 h to create a map of the four elements in an area of 40 × 40 spots of 5 μm wide.

A typical proton spectrum is shown in Figure 9(a). Carbon signals from surface contamination appear in region D. Carbon signals from the bulk material (region C) are to be subtracted from regions of oxygen signals (regions A and B). The subtraction of the carbon contribution in regions A and B of oxygen in the bulk (A) and at the surface (B) is performed by calculating the relative contribution of carbon in (C) and in the corresponding regions A and B of the proton spectrum collected on a pure carbon homogeneous standard sample (Figure 9b).

Using the pure RBS cross-section on Cu, Y and Ba, the stoichiometric determination of heavy compounds (expected to be $Y_1Ba_2Cu_3$) was $Y_{1.01}Ba_2Cu_{3.01}$ (Ba_2 as reference).

The accuracy of the measurement of O concentration (7 ± 1) by RBS was not sufficient because of superposition of the O signals on to the huge contribution of the heavier elements. The accuracy of the O content using results of the (d,p) reaction, is of the order of 3% (statistical accuracy).

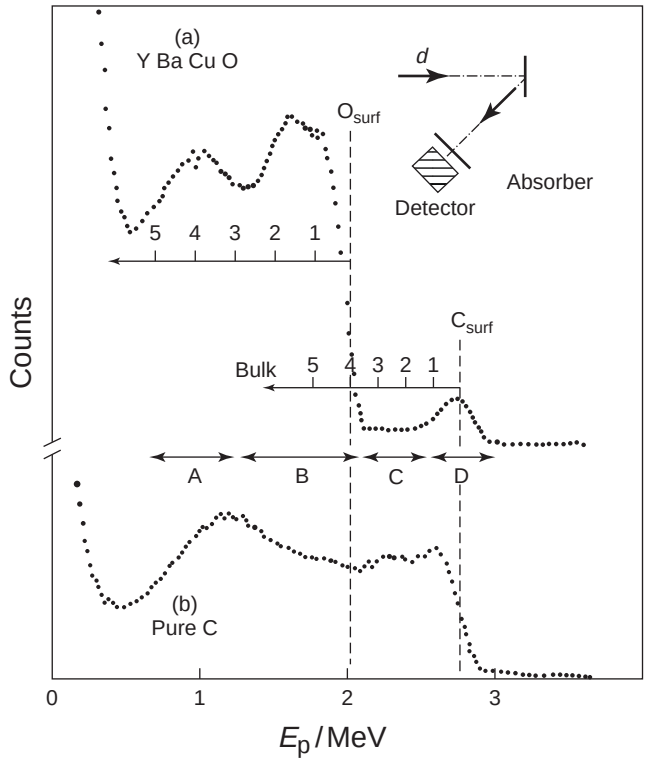

Figure 9 Proton spectra from (d,p) reactions on C and O. The incident deuteron energy is 2 MeV and the mylar absorber thickness is 10 mg cm^{-2}. Oxygen may be profiled down to 5–6 mg cm^{-2} below the surface. (Reprinted from G. Demortier et al., 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors With Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496, Copyright (1988), with permission from Elsevier Science.[56])

Materials showing the strongest Meissner effect (levitation of a small magnet) at liquid nitrogen temperature are also those showing the best homogeneity in Cu, Ba and Y concentrations in $5 \times 5 \times 5$ μm regions analyzed (Figure 10a and b). Possible local nonhomogeneities in several samples show that Ba and Y are related.[56]

Ternary diagrams of three of the four components (Y, Ba, Cu, O) indicate that, as far as the most homogeneous material is concerned, local discrepancies from the ideal $Y_1Ba_2Cu_3$ composition are mainly observed in the Y content, the reason lies only in the poor statistical accuracy on K X-rays of Y induced by 2 MeV deuterons.[56] An improvement in the accuracy of Y determination is achieved when the same sample is studied by PIXE induced by 3 MeV protons. The distribution of the points in a ternary diagram involving O, Cu and Ba as extracted from numerical results on the homogeneous material (Figure 11a) is not of statistical origin, but indicates some nonhomogeneity in the microscopic distribution of oxygen. For the second (nonhomogeneous) sample, the nonhomogeneity is so great that the points are scattered far from the ideal position: the center of the triangle. Nonhomogeneities in the O concentration of Figure 11(b) cannot be correlated with the presence of any specific oxide (CuO, Y_2O_3, BaO, etc.). For the homogeneous sample, the local concentration of oxygen exhibits variations of about 10% around the expected value of 6.7. The reason could be the presence, in the synthesized sample, of copper in two different valence states (2^+ and 3^+).

The experimental accuracy on the relative concentrations of Cu, Ba and O, using simultaneously PIXE and (d,p) reactions, was less than 3%.

5.4 Three-dimensional Microanalysis of a Gold–Silicon Eutectic Alloy

A film of silicon was deposited on pure polycrystalline gold foils, previously rolled down to thicknesses ranging from 10 to 20 μm. The deposition was performed with an electron gun at low pressure. The amount of evaporated silicon was adjusted in order to obtain films up to 1 μm thick on the gold substrate. This substrate was maintained at 400 °C during the silicon deposition, a temperature which is slightly higher than the eutectic temperature (363 °C). At this temperature a liquid Au–Si mixture diffuses, during the deposition procedure, in the gold foils (Figure 12) but only along well-defined paths, the gold grain boundaries. K X-rays of Si induced by PIXE **PIXE (Particle-induced X-ray Emission)** clearly indicate that a gold–silicon alloy has diffused on the surface of the foil by forming characteristic leaf-shaped decorations starting from the boundaries of the gold grains which were flattened by the rolling procedure. Several of these

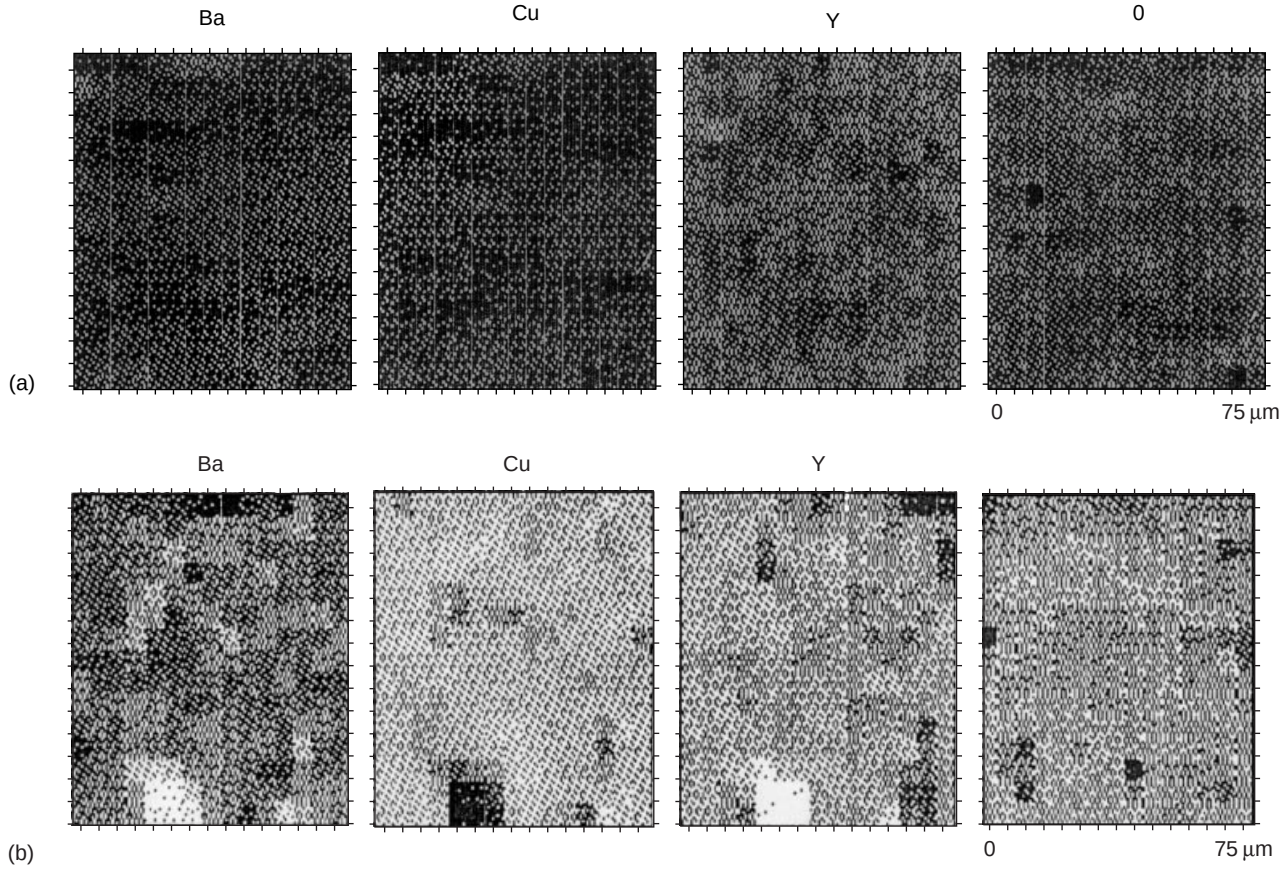

Figure 10 Concentration maps of Ba, Cu, Y and O in bulk superconductors. Concentrations are presented by 64 gradations of gray. The maximum values are, respectively, in each pixel: 8192 for Ba(L), 4096 for Cu(K_α), 256 for Y(K_α) and 1024 for O (protons). The beam diameter is less than 5 μm and the distance between steps is 10 μm. (a) Homogeneous YBaCuO sample; (b) Nonhomogeneous YBaCuO sample. (Reprinted from G. Demortier et al., 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors With Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496. Copyright (1988), with permission from Elsevier Science.[56])

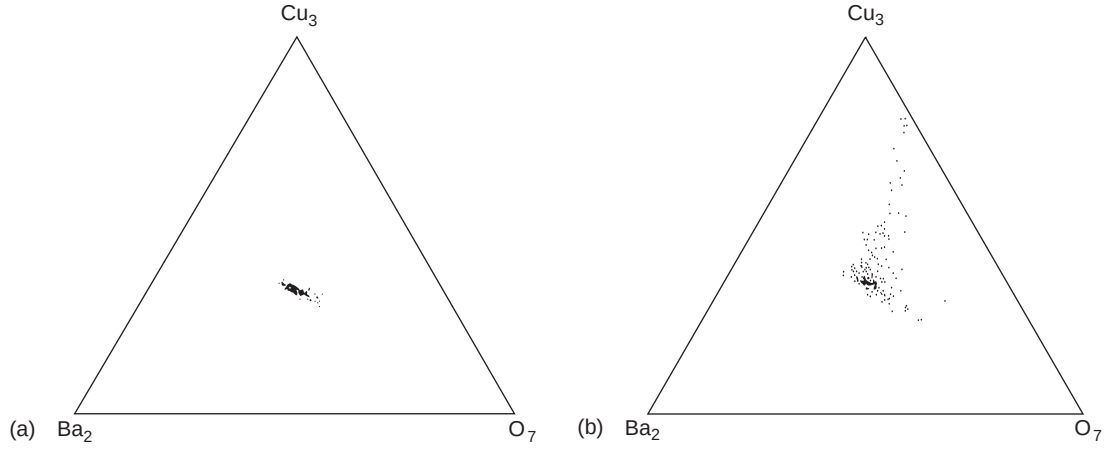

Figure 11 Ternary diagram of Cu, Ba and O for a homogeneous sample (a) and a nonhomogeneous one (b). (Reprinted from G. Demortier et al., 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors With Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496. Copyright (1988), with permission from Elsevier Science.[56])

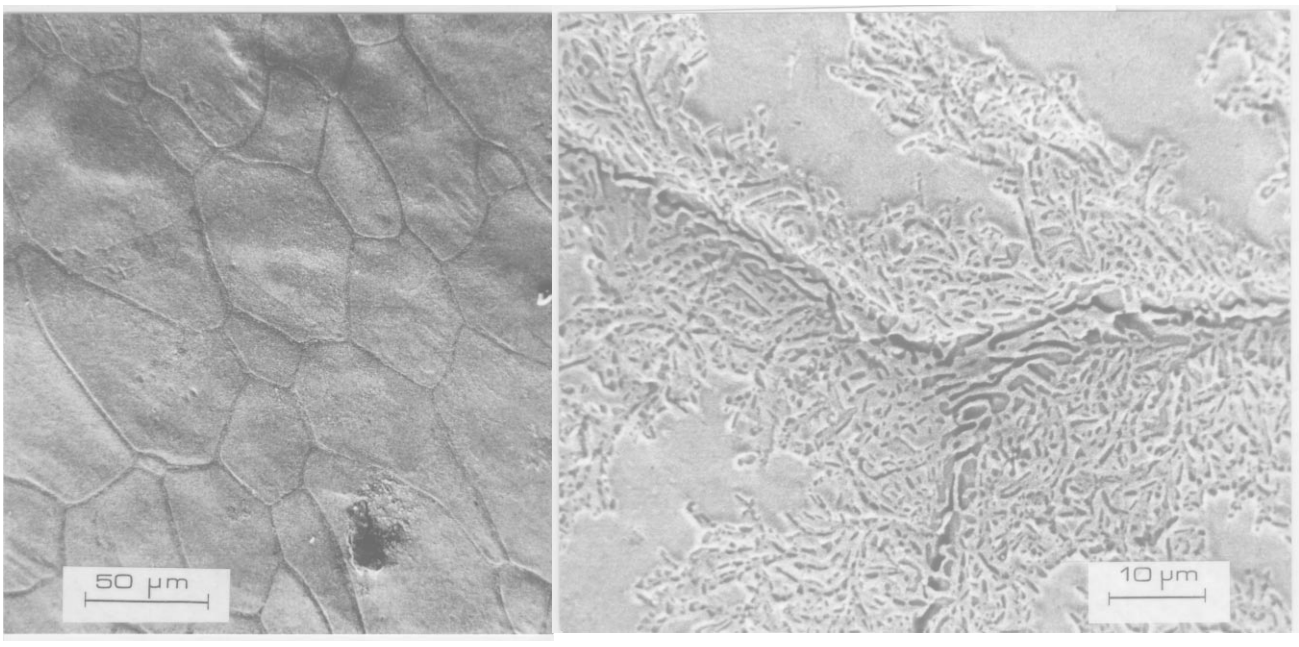

Figure 12 Electron micrographs of a Au–Si eutectic alloy formed by Si deposition in the grain boundaries of a 12 μm thick sheet of gold. (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

flattened grains may occupy the whole thickness of the foil. Therefore, the diffusion of the liquid phase took place through the whole gold foil thickness. The diffusion is so fast (a few minutes at such a low temperature cannot be explained by a process of solid state diffusion) that nothing other than a liquid phase of gold–silicon eutectic alloy could diffuse. The atomic concentration of Si in this eutectic alloy is 19% (about 3.2% Si by weight).

The depth profile of Si in gold is taken using a (d,p) reaction.[57] The nuclear reaction $^{28}Si(d,p_0)^{29}Si$ ($Q = 6.25\,keV$) (leaving the residual ^{29}Si nuclei in their fundamental energy level) is mainly governed by neutron capture involving no exchange of angular momentum ($\ell = 0$). As a consequence, the maximum intensity of the proton emission is achieved in the forward direction. The cross-section at $\theta = 0^\circ$ of this nuclear reaction at $E_d = 2.8\,MeV$ is quite large at $6\,mb\,sr^{-1}$. We then chose to detect the emitted protons in the forward direction. In this special arrangement, the proton detector is situated into the incident deuteron beam (Figure 13). As incident deuterons of 2.8 MeV cannot cross the coulomb barrier of a Au nucleus, the (d,p) nuclear reaction can only take place with light nuclei. Therefore, the detected proton spectra contain only information on Si. A 20 μm thick absorber of pure gold is inserted between the sample to be studied and the detector in order to stop the incident deuteron beam completely even when the beam would cross a hole in the gold–silicon sample. The choice of a material containing only a heavy element (Au) as the absorber prevents any production of additional protons in this gold absorber during the deuteron irradiation.

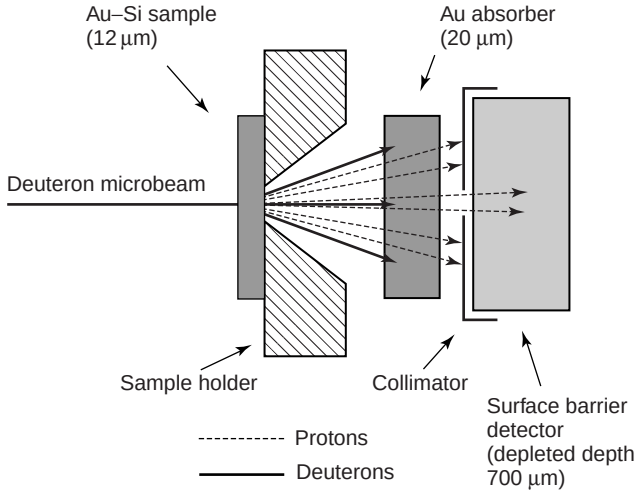

Figure 13 Special experimental set-up used to perform depth profiling of Si in a thin gold foil treated for Si deposition. (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

The quantitative determination of silicon at different depths below the surface is based on the comparison of proton spectra obtained when ultrapure silicon and gold–silicon samples in the same geometrical

arrangement are bombarded. The silicon reference material is a thin foil (75 μm thick) of Si used in semiconductor technology. The corresponding depths in pure silicon and in gold–silicon alloys are calculated using the relative stopping power of pure silicon and that of the eutectic gold–silicon alloy (Figure 14). To scan the sample, a proton microbeam (5 μm × 5 μm) is rastered on the surface along 50 μm × 50 μm regions, a surface involving more than the area of one single flattened grain.

If the deuteron beam hits a gold grain boundary in which the diffusion of the eutectic alloy took place, a proton spectrum like that of Figure 14(a) may be observed

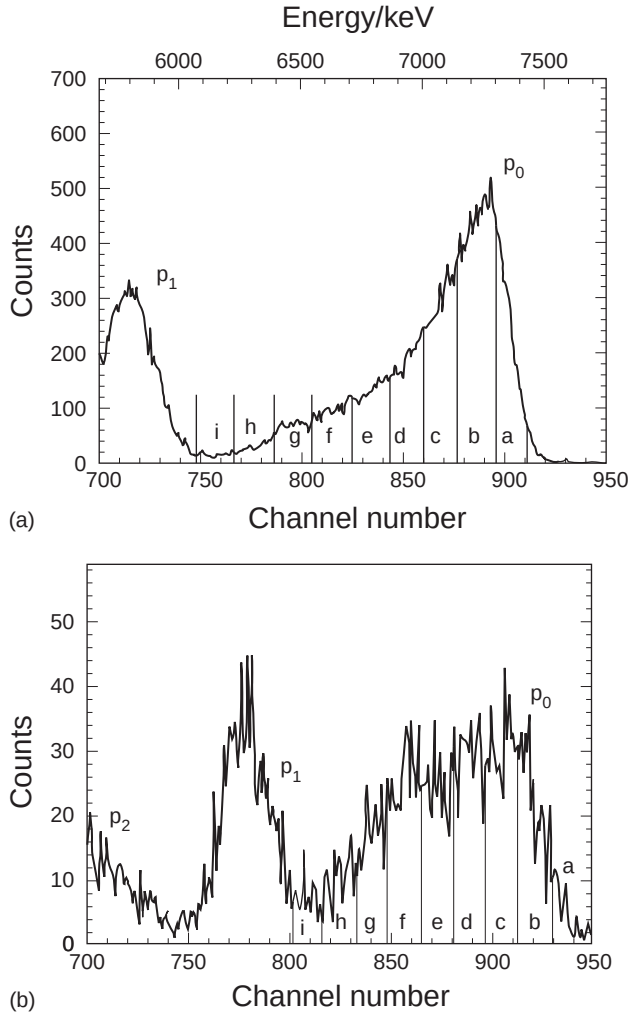

Figure 14 Experimental procedure to obtain depth profiling of silicon from the decomposition in nine slices from the p_0 contribution. (a) Pure 75 μm Si sample; (b) 12 μm thick Au foil processed for silicon diffusion (irradiation of the rear face with 2.8 MeV deuterons). (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

showing a more important contribution in the medium and the low energy parts of the p_0 broad ''peak'' than when this deuteron beam hits a uniform silicon sample (Figure 14b). The depth distribution of Si in this particular case indicates that the relative concentration of Si is increasing from the surface to a region buried at several microns below the surface.

The map of Figure 15 corresponds to the silicon distribution at various depths below the surface irradiated at the rear part of the prepared gold foil. Figure 15(a) reflects the Si concentration at the bottom of the prepared foil which is reached by the most energetic protons. The regions of the proton spectrum shown in Figure 14(a) show schematically the corresponding regions in the depth of the 12-μm thick irradiated specimens. The incident beam was directed onto the face of the gold foil opposite to the face on which the Si deposition was made. In any region the silicon concentration does not exceed a few percent, a concentration well below that of the eutectic composition. This eutectic concentration is observed in Figure 15(i) which corresponds to grain boundary regions at the point where deuterons exit and at the surface where the silicon was deposited.

Nevertheless, in regions where silicon is detected only a liquid phase with a composition close to that of the eutectic could be admitted. The diffusion takes place so rapidly (a few minutes) that no appreciable solid state diffusion could be expected. Each concentration of silicon lower than 19% (as observed in nearly all parts of each map) can be understood by considering that the size of a grain boundary is narrower than the area of the incident deuteron beam. Scans like those of Figure 15 may then be converted into measurements of the width of the grain boundary filled with the eutectic alloy, the ratio of the measured silicon concentration on the eutectic composition being the ratio of the width of the grain boundary to the beam diameter.

5.5 Determination of Traces of Sulfur in Ancient Gold Artefacts

Investigations of the composition of archaeological artefacts made from precious metals require the use of nondestructive analytical techniques. PIXE and PIGE have been successfully applied in our laboratory for topographical elemental analysis of Fe, Cu, Zn, Ag, Cd, Sn and Au in uncorroded surfaces of jewellery items. The composition of the material in narrow regions of tiny solders on antique gold artefacts may give valuable information about the workmanship of ancient goldsmiths. In various papers, the hypothesis that in ancient times cadmium sulfide (a yellow mineral with a color close to that of gold) may have been ''alloyed'' with gold to obtain a material suitable for brazing has

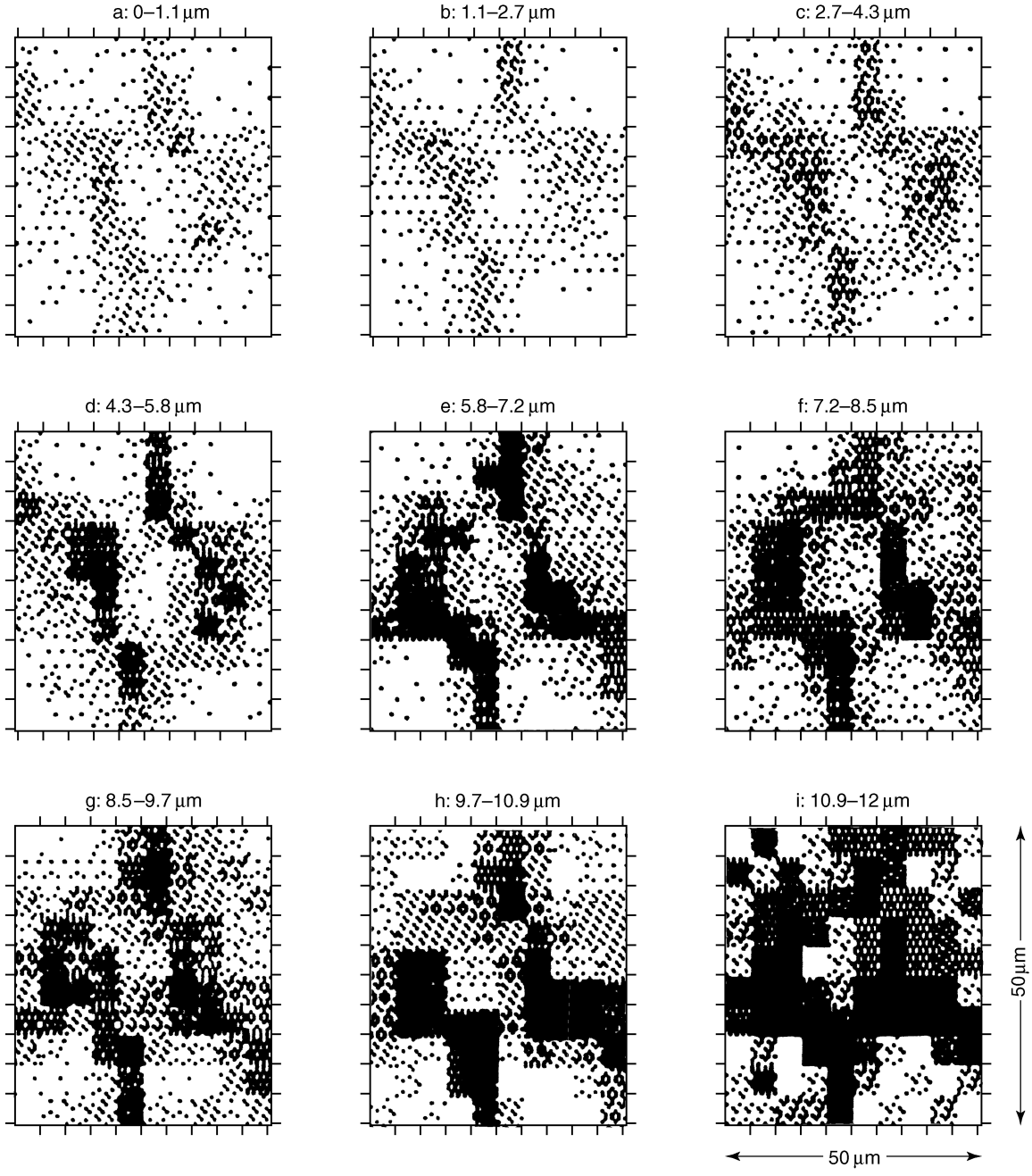

Figure 15 Set of nine maps corresponding to the nine slices of Figure 14. Symbols (a) to (i) refer to the nine layers in the 12 μm thick sample. (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

been discussed.[57] To test this hypothesis an experimental procedure for analyzing traces of sulfur in gold artefacts was developed. It is indeed expected that during the antique alloying procedure, traces of sulfur from cadmium sulfide were introduced in the brazing alloy.

Detection of energetic protons induced by deuteron bombardment of the specimen under investigation was achieved. The detector was situated at an angle of 135° with respect to the incident deuteron beam in order to facilitate the study of thick specimens often with an irregular shape. Sheets of mylar of uniform thickness (2.9 mg cm^{-2} each) were inserted between the target and the detector in order to stop the scattered deuterons completely. Owing to the fact that most of the (d,p) reactions have a large positive Q value, the proton energies are not significantly reduced by these absorbers.

The best experimental condition for sulfur determination is obtained at $E_d = 1.9$ MeV, in which case the cross-section is sufficient and the p_0 and p_1 peaks are well resolved.[57] Since the cross-section for 1.7 MeV deuterons is greater than at 1.9 MeV, working conditions at 1.9 MeV (on a thick sample) allow us to check for the presence of sulfur in the bulk and not only at the surface. Possible interfering elements (with their appropriate Q values) are within the scope of (d,p) reactions on S: Si, Mg, Cl, Ca (for p_0) and Al, P, Cl, K, Mg (for p_1).

The choice of detection angle and incident deuteron energy was then a determining factor in eliminating the

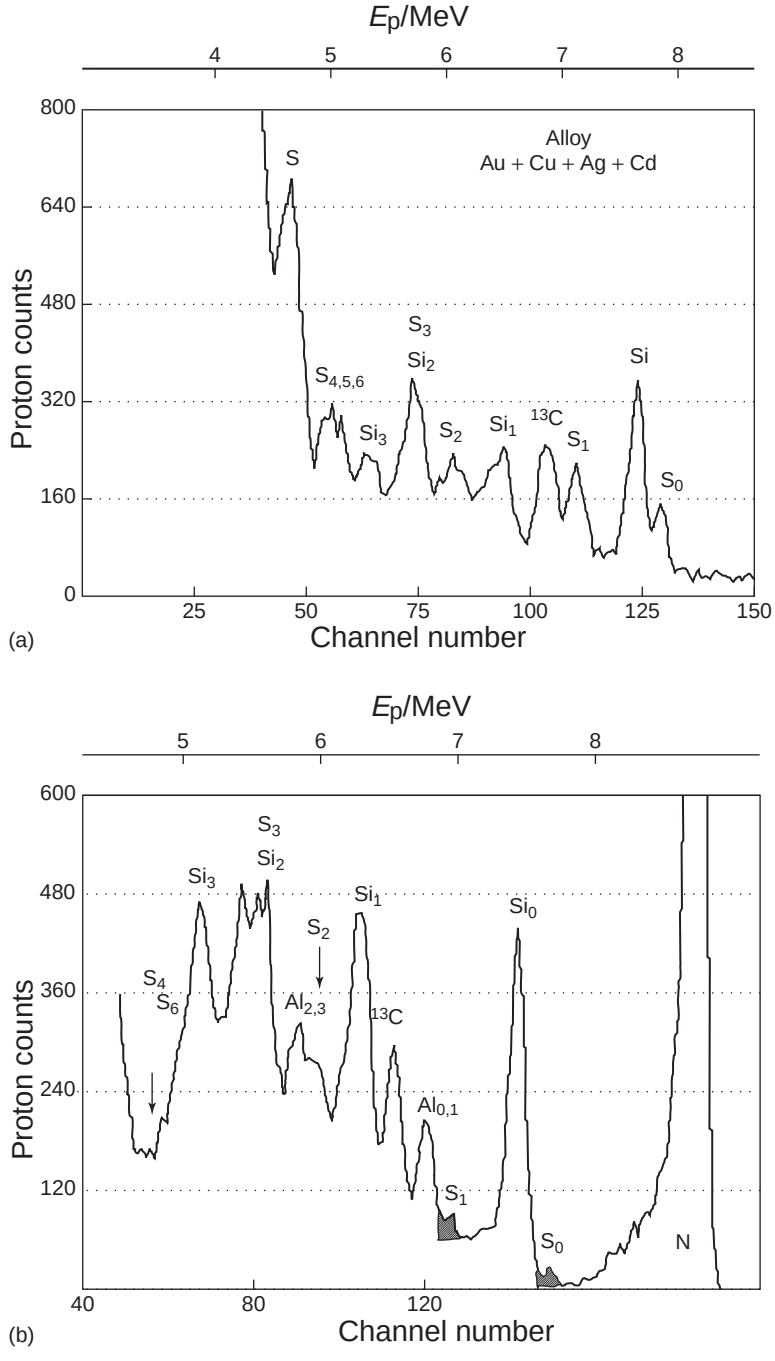

Figure 16 Experimental results from a (d,p) reaction induced on gold alloys prepared by an antique procedure (a) and in a soldering region of an antique jewellery artefact of the 9th century A.D. (b) Several peaks indicate the presence of Si and S. (Reprinted from G. Demortier, A. Gilson, 'Determination of Traces of Light Elements in Gold Artifacts Using Nuclear Reactions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **18**, 286–290. Copyright (1987) with permission from Elsevier Science.[57])

possible interferences. Detection of p_0 and p_1 protons provided a check for an interference-free investigation. This is because of the presence of both p_0 and p_1 signals for S outside the regions for Al($p_0 + p_1$) and (Si p_0) and the relative intensities of these peaks in comparison with a reference sample.

Sulfur and silicon (which were always introduced in metals when heating ores in antique metallurgy) are of crucial importance in understanding ancient technologies. Sulfur (in blende and greenockite) and silicon (in many minerals) may be expected to be present in narrow parts of the solder formed during the antique brazing of gold. Figure 16(a) shows the analysis of a soldering alloy that we prepared using this antique procedure: gold–cadmium alloys (with low melting point) are obtained by direct dissolution of CdS in a previously melted drop of gold. The concentrations of S and Si detected are, respectively, 120 ± 35 ppm and 50 ± 15 ppm. Note also the presence of C (from ^{13}C) as an extra signal arising from the experimental set-up, i.e. a carbon (from oil in the pumping system) deposit on the target during the deuteron irradiation. The mylar absorber must have a thickness of 25 mg cm^{-2} in order to eliminate also protons from

Figure 17 Small spherical pearl (diameter 11 mm) (9th century A.D.) decorated with granules, and which could be a part of a necklace, found in Syria with other Byzantine jewellery items.

Table 5 Analysis of biological reference materials by NRA, PIGE and PIXE

Element	IBA techniques				Bowen's values[58]	
	E_i in keV	Reaction	Particle or photon detected	Concentration	Mean	Ranges
B	450	(p,α)		46.2 ± 5 ppm	48.3	39–56
	700	(p,α)		45.1 ± 4 ppm		
C	1020	(d,p)	p_0	$43.7 \pm 2\%$	41.8%	
N	1500	(d,α)	α_0	$4.76 \pm 0.3\%$	4.31%	
O	900	(d,p)	p_0	$43 \pm 3\%$	5.15%[a]	
F	2000	(p,p′γ)	γ110 keV	7.14 ± 2 ppm	4.92	4.2–6.2
	2000	(p,p′γ)	γ197 keV	6.23 ± 2 ppm		
	2200	(p,αγ)	γ5100 keV	6.8 ± 3 ppm		
	2200	(p,αγ)	γ6100 keV	6.1 ± 3 ppm		
Na	1500	(p,p′γ)	γ439 keV	2390 ± 120 ppm	2506	1220–3250
	1900	(p,p′γ)	γ439 keV	2630 ± 100 ppm		
	2000	(p,p′γ)	γ439 keV	2510 ± 80 ppm		
	2100	(p,p′γ)	γ439 keV	2480 ± 70 ppm		
	2200	(p,p′γ)	γ439 keV	2540 ± 70 ppm		
	2000	(p,p′γ)	γ585 keV	1920 ± 250 ppm	1572	1370–1700
	2200	(p,p′γ)	γ585 keV	1580 ± 150 ppm		
P	700	PIXE	K_α	$0.47 \pm 0.04\%$	0.4489	0.402–0.481
Cl	700	PIXE	K_α	$0.37 \pm 0.02\%$	0.3415	0.218–0.445
K	700	PIXE	K_α	$2.70 \pm 0.2\%$	2.46	2.06–2.93
Ca	700	PIXE	K_α	$4.05 \pm 0.1\%$	4.085	
Mn	1700	PIXE	K_α	13.0 ± 2 ppm	14.7	12.6–18.0
Fe	1700	PIXE	K_α	113 ± 7 ppm	118.3	88–157
Cu	1700	PIXE	K_α	4.9 ± 0.5 ppm	4.99	3.6–6.5
Zn	1700	PIXE	K_α	32.5 ± 2 ppm	33.2	30–38
Br	1700	PIXE	K_α	26.0 ± 2 ppm	26.1	
Rb	1700	PIXE	K_α	49.0 ± 4 ppm	52.2	49–57
Sr	1700	PIXE	K_α	100.2 ± 5 ppm	98.9	65–150

[a] This low oxygen concentration, determined by neutron activation analysis, is not reliable with the expected value for a biological sample.

(d,p) reactions on ^{12}C. The concentration of Cd could be determined simultaneously by detection of characteristic X-rays induced by the same deuteron beam. The presence of residual sulfur (only at trace level) indicates that at least 95% of S was lost during the alloying procedure.

Figure 16(b) was obtained during the irradiation of a region of a soldered gold granule on a hollow gold pearl found in Syria (Hauran) with other more prestigious objects of Byzantine jewellery (9th century A.D.) (Figure 17).

The information from S (at concentration 30 ± 7 ppm) cannot be easily isolated from the large contribution of Si at the surface. The width of the peaks of Si present in this ancient artefact is much narrower than those observed in pure Si and in the soldering alloys formed by the methods available in antiquity (Figure 16a). The presence of traces of Si and often also traces of Al, can be attributed to dust inclusions in narrow regions of ''porous'' solders.

5.6 Particle-induced X-ray Emission, Particle-induced γ-ray Emission and Nuclear Reaction Analysis for Elemental Analysis of a Biological Reference Material

Bowen's kale,[58] a powder obtained by dry freezing this vegetable, was widely distributed in various laboratories to be certified for a reference material. Using various IBA techniques, it was possible to determine most of the main elements and traces quantitatively. Table 5 summarizes these results.[59]

NRA induced by protons and deuterons was used for the determination of B, C, N, F and Na. The choice of a specific incident energy and the selected identification signal were determined for interference-free determination.

Compositions of various elements (from P or Ca) were obtained at low energy PIXE in order to minimize the X-ray absorption into the target itself. Traces of higher Z elements were found using PIXE at 1.7 MeV in order to improve the sensitivity. Determination of hydrogen was not possible due to the rapid destruction of the sample powder during heavy ion irradiation (for ERD or NRA using ^{15}N). The concentration of oxygen was established at 43%, a more reliable value than the previous determination using neutron activation analysis. The total content of this reference material using PIXE, PIGE and NRA exclusively (hydrogen not included) is $98.9 \pm 3.5\%$.

ABBREVIATIONS AND ACRONYMS

BGO	Bismuth Germanate Oxide
CPAA	Charged Particle Activation Analysis
ERD	Elastic Recoil Detection
fwhm	Full Width at Half Maximum
IBA	Ion Beam Analysis
NRA	Nuclear Reaction Analysis
PIGE	Particle-induced γ-ray Emission
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in • Noble Metals, Analytical Chemistry of

Surfaces **(Volume 10)**
Ion Scattering Spectroscopy in Analysis of Surfaces

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission) • Rutherford Backscattering Spectroscopy

REFERENCES

1. P. Vavilov, 'Ionization Losses of High Energy Heavy Particles', *Sov. Phys. JETP*, **5**, 749–751 (1997).
2. G. Deconninck, *Introduction to Radioanalytical Physics*, Elsevier Scientific, Amsterdam, 1978.
3. G. Deconninck, Y. Fouilhe, 'Energy Spreading Calculations and Consequences', in *Ion Beam Surface Layer Analysis*, eds. O. Meyer, K. Käppeler, G. Linker, Plenum Press, 97, Vol. 1, 1975.
4. J.M. Harris, M.A. Nicolet, 'Energy Straggling of ^4He Ions Below 2 MeV in Al, Ni and Au', *Phys. Rev. B*, **11**, 1013 (1975).
5. S.A.E. Johannsson, J.L. Campbell, K.G. Malmqvist (eds.), *Particle-induced X-ray Emission Spectrometry (PIXE)*, John Wiley & Sons, New York, 1995.
6. G. Deconninck, G. Demortier, F. Bodart, 'Prompt Gamma-ray Spectroscopy and its use in Elemental Analysis', *Atomic Energy Rev.*, Suppl. 2, 1981.
7. W.S.C. Williams, *Nuclear and Particle Physics*, Clarendon Press, Oxford, 1991.
8. N. Jarmie, J. Seogrove, 'Nuclear Data Tables', Los Alamos Scientific Laboratory Report, LA-2014, 1957.
9. J.L. Yarnell, R.H. Lovberg, W.R. Stratton, 'Angular Distribution of the Reaction $^3He(d,p)^4He$ Between 240 keV and 3.65 MeV', *Phys. Rev.*, **90**, 292 (1953).

10. A.J. Elwyn, R.E. Holland, C.N. Davids, L. Meyer-Shützmeister, F.P. Moorning, W. Ray, Jr, 'Cross Sections for ^{6}Li(p,^{4}He)^{3}He Reaction at Energies Between 0.1 and 3.0 MeV', *Phys. Rev. C*, **20**, 1984 (1979).
11. B. Maurel, 'Théorie stochastique de la perte d'énergie dans la matière des particules chargées. Application aux courbes d'excitation autour de résonances nucléaires étroites et à la mesure des profils de concentration près de la surface des solides', PhD Thesis, University of Paris VII, 1981.
12. J.P. Schiffer, T.W. Bonner, R.H. Davis, F.W. Prosser, Jr, 'Study of the Reaction Mechanism for (^{3}He,p) Reactions with ^{6}Li; ^{11}B and ^{13}C', *Phys. Rev.*, **104**, 1064 (1980).
13. Y. Cassagnou, J.M.F. Jeronymo, A. Sadeghi, P.D. Forsyth, 'The ^{7}Li(p,α)α Reaction', *Nucl. Phys.*, **33**, 449 (1962).
14. A.J. Sierk, T.A. Tombrello, 'The ^{9}Be(p,α) and (p,d) Cross Sections at Low Energies', *Nucl. Phys. A*, **210**, 341 (1973).
15. M. Shaanan, R. Kalish, V. Richter, 'Measurement of Profiles by ^{9}Be(p,α)^{6}Li Reaction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **16**, 56 (1986).
16. J.A. Biggerstaff, R.F. Hood, H. Scott, M.T. McEllistrem, 'Differential Cross Sections for (d,α) and (d,t) Reactions in ^{9}Be', *Nucl. Phys.*, **36**, 631 (1962).
17. E.A. Wolicki, H.D. Holmgren, R.L. Johnston, E. Geer-Illsley, 'Differential Cross Sections for the ^{9}Be(^{3}He,p)^{11}B Reaction', *Phys. Rev.*, **116**, 1585 (1959).
18. B. Bilwes, R. Bilwes, J.L. Ferrero, A. Garcia, 'Mécanisme de la réaction ^{9}Be(^{3}He,α)^{8}Be(0) et résonances géantes isoscalaires du ^{12}C', *J. Physique*, **39**, 805 (1978).
19. K.H. Purser, B.H. Wildenthal, 'The B^{10}(d,α)Be^{8} Reaction and the 2.94 MeV State in ^{8}Be', *Nucl. Phys.*, **44**, 22 (1963).
20. L.C. McIntyre, Jr, J.A. Leavitt, M.D. Ashbaugh, Z. Lin, J.O. Stoner, Jr, 'Determination of Boron Using the B(α,p)C Nuclear Reaction at Incident Energies near 3 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 221 (1992).
21. A. Armigliato, G.G. Bentini, G. Ruffini, G. Battaglin, G. Della Mea, A.V. Drigo, 'Analysis of Boron Predeposited Silicon Wafers by Combined Ion Beam Techniques and X-ray Microanalysis', *Nucl. Instrum. Methods*, **149**, 653 (1978).
22. E. Ligeon, A. Bontemps, 'Nuclear Reaction Analysis of Boron and Oxygen in Silicon', *J. Radioanal. Chem.*, **12**, 335 (1972).
23. J.P. Scanlow, G. Farell, M.C. Ridgway, R. Valizadeh, 'Nuclear Reaction Analysis of Shallow Boron Implanted in Silicon', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **16**, 479 (1986).
24. H.D. Holmgren, E.A. Wolicki, R.L. Johnston, '^{11}Be(^{3}He,p)^{13}C and ^{11}B(^{3}He,d)^{12}C Reactions', *Phys. Rev.*, **114**, 1281 (1959).
25. C. Olivier, J.W. McMillan, T.B. Pierce, 'An Investigation of the Use of the Nuclear Microprobe for Examining Distributions of Boron Isotopes', *J. Radioanal. Chem.*, **31**, 515 (1976).
26. E. Kashy, R.R. Perry, J.R. Risser, 'Excited States in ^{14}N from ^{12}C(d,d)^{12}C, ^{12}C(d,p$_0$)^{13}C and ^{12}C(d,p$_1$)^{13}C*', *Phys. Rev.*, **117**, 1289 (1960).
27. G. Terwagne, 'Cross Section Measurements of the Reactions Induced by ^{3}He Particles on Carbon', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **122**, 1 (1997).
28. H.M. Kuan, T.W. Bonner, J.R. Risser, 'An Investigation of the ^{12}C + ^{3}He Reactions at Bombarding Energies Between 1.8 and 5.4 MeV', *Nucl. Phys.*, **51**, 481 (1964).
29. J.B. Marion, G. Weber, 'Compound Nucleus Effects in Deuteron Reactions ^{12}C(d,p)^{13}C', *Phys. Rev.*, **103**, 167 (1956).
30. G. Amsel, D. David, 'La microanalyse de l'azote par l'observation directe de réactions nucléaires: applications', *Rev. Phys. Appl.*, **4**, 383 (1969).
31. W. Koenig, A. Riccato, R. Stock, P. Cuzzocrea, E. Perillo, M. Sandolini, G. Spadaccini, 'Cross Section Measurements and Angular Distribution Study of ^{14}N(d,α)^{12}C Nuclear Reaction', *II Nuovo Cimento A*, **39**, 9 (1977).
32. T. Ishimatsu, T. Nakashima, N. Kato, S. Morita, 'The ^{14}N(d,α_0) and ^{14}N(d,α_1) Reactions from 1 to 3.5 MeV', *J. Phys. Soc. Jpn.*, **17**, 1189 (1962).
33. F.B. Hagendorn, J.B. Marion, 'Further Investigation of ^{15}N + p Reactions', *Phys. Rev.*, **108**, 1015 (1957).
34. J.A. Sawicki, J.A. Davies, T.E. Jackman, 'Absolute Calibration of the ^{14}N(d,α_0)^{13}C and ^{15}N(p,α_0)^{12}C Reaction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 530 (1986).
35. F. Seiler, C.H. Jones, W.J. Anzick, D.F. Herring, K.W. Jones, 'The Elastic Scattering of Deuterons by ^{16}O from 0.65 to 2.0 MeV', *Nucl. Phys.*, **45**, 647 (1963).
36. A. Turos, L. Wielunski, A. Barcz, 'Use of the Nuclear Reaction ^{16}O(d,α)^{14}N in the Microanalysis of Oxide Surface Layers', *Nucl. Instrum. Methods*, **111**, 605 (1973).
37. S.Y. Tong, W.N. Lennard, P.F.A. Alkemada, I.V. Mitchell, 'Absolute Surface Carbon Coverage Determination via the ^{12}C(^{3}He,p)N^{14} Reaction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 91 (1990).
38. G. Amsel, 'Spectroscopie par détecteurs à semidconducteur des réactions ^{16}O + d, ^{18}O + p', *Ann.Phys.*, **9**, 297 (1964).
39. D. Dieumegard, B. Maurel, G. Amsel, 'Microanalysis of Fluorine by Nuclear Reactions. I. ^{19}F(p,α_0)^{16}O and ^{19}F(p,$\alpha\gamma$)^{16}O Reactions', *Nucl. Instrum. Methods*, **168**, 93 (1980).
40. B. Maurel, G. Amsel, D. Dieumegard, 'Microanalysis of Fluorine by Nuclear Reactions. II. ^{19}F + d Reactions', *Nucl. Instrum. Methods*, **191**, 349 (1981).
41. M. Vos, D.O. Boerma, P.J.M. Smulders, S. Oosterhoff, 'Defect and Dopent Depth Profiles in Boron-implanted Silicon Studied with Channeling and Nuclear Reaction Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **17**, 234 (1986).
42. S. Fiarman, S.S. Hanna, 'Energy Levels of Light Nuclei, A = 3', *Nucl. Phys. A*, **251**, 1–64 (1975).
43. S. Fiarman, W.E. Meyerhof, 'Energy Levels of Light Nuclei, A = 4', *Nucl. Phys. A*, **206**, 1–64 (1973).

44. P.M. Endt, C. Van der Leun, 'Energy Levels of Light Nuclei, A = 21–44', *Nucl. Phys. A*, **310**, 1–752 (1978).
45. F. Ajzenberg-Selove, 'Energy Levels of Light Nuclei, A = 5–10', *Nucl. Phys. A*, **227**, 1–244 (1974); 'Energy Levels of Light Nuclei, A = 16–17', **281**, 1–148 (1977); 'Energy Levels of Light Nuclei, A = 18–20', **475**, 1–198 (1987); 'Energy Levels of Light Nuclei, A = 11–12', **506**, 1–158 (1990); and 'Energy Levels of Light Nuclei, A = 13–15', **523**, 1–196 (1991).
46. J.W. Mayer, E. Rimini (eds.), *Ion Beam Handbook for Material Analysis*, Academic Press, New York, 1997.
47. J. Tesmer, M. Nastasi (eds.), *Handbook of Modern Ion Beam Materials Analysis*, MRS, Pittsburgh, 1995.
48. W. Lanford, 'Analysis for Hydrogen by Nuclear Reaction and Energy Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 65–82 (1992).
49. P. Trocellier, Ch. Engelmann, 'Hydrogen Depth Profile Measurements Using Resonant Nuclear Reactions: a Review', *J. Radioanal. Nucl. Chem.*, **100**, 117 (1986).
50. F. Xiong, F. Rauch, C. Shi, Z. Zhou, R.P. Livi, T.A. Tombrello, 'Hydrogen Depth Profiling in Solids: a Comparison of Several Resonant Nuclear Reaction Techniques', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **27**, 432 (1987).
51. K.M. Horn, W. Lanford, 'Observation of the Bond-dependent Doppler Broadening of the $p(^{15}N,\alpha\gamma)^{12}C$ Nuclear Reaction', *Nucl. Instrum. Methods*, **29**, 609 (1983).
52. G. Amsel, B. Maurel, 'High Resolution Techniques for Nuclear Reaction Narrow Resonance Width Measurements and for Shallow Depth Profiling', *Nucl. Instrum. Methods*, **218**, 183 (1988).
53. P.A. Pronko, J.G. Pronko, 'Depth Profiling of ^{3}He and ^{2}H in Solids Using the $^3He(d,p)^4He$ Resonance', *Phys. Rev. B*, **9**, 2870 (1974).
54. G. Terwagne, G. Deconninck, 'Concentration Profile of Light Elements Near an Interface', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 153–155 (1992).
55. S. Mathot, G. Demortier, 'Silicon Diffusion in Gold Grain Boundaries', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **50**, 52–56 (1990); and 'Three-dimensional Nuclear Microanalysis in Materials Science', **77**, 312–319 (1993).
56. G. Demortier, F. Bodart, G. Deconninck, G. Terwagne, Z. Gabelica, E.G. Derouane, 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors with Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496 (1988).
57. G. Demortier, A. Gilson, 'Determination of Traces of Light Elements in Gold Artifacts Using Nuclear Reactions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **18**, 286–290 (1987).
58. H.J.M. Bowen, 'Problems in the Elementary Analysis of Standard Biological Materials', *J. Radioanal. Chem.*, **19**, 215 (1972).
59. J. Baijot-Stroobants, G. Debras, G. Demortier, 'Analyse quantitative globale et distribution superficielle du carbone, de l'azote et de l'oxygène dans des matrices biologiques par réactions nucléaires de basse énergie', *Radioanal. Radiochem. Lett.*, **27**, 271–280 (1976).
60. G.Demortier (ed.), 'Microanalysis of Light Elements (Hydrogen to Neon) Using Charged Particle Accelerators', Proceedings of the Third International Conference on Chemical Analysis, Namur, July 1991, *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 1–309, 1992.

Parent and Progeny Compounds in Exhaled Breath, Determination of

Shane S. Que Hee
UCLA School of Public Health, Los Angeles, USA

1 History 1
2 The Respiratory System 2
2.1 The Nose 2
2.2 The Mouth and the Throat 2
2.3 The Windpipe 2
2.4 The Bronchi 2
2.5 The Lungs 3
2.6 Breathing 3
2.7 Gas/Vapor Exchange 3
2.8 Reference Breathing Parameters 4
3 Breath 5
3.1 Gross Chemical Composition of Exhaled Breath 5
3.2 Markers of Xenobiotic Exposure in Exhaled Breath 6
3.3 Medical Breath Testing 7
3.4 Dental Breath Tests 8
4 Breath Sampling 9
4.1 Gas Bags 9
4.2 Summa Cannisters 10
4.3 Solid Sorbent Sampling 10
4.4 Glass Tubes 10
4.5 Other Methods 10
5 Chemical Analysis of Breath 10
5.1 Introduction 10
5.2 Gas Chromatography and Mass Spectrometry Methods 11
5.3 Infrared Methods 11
5.4 Other Methods 12
Acknowledgments 12
Abbreviations and Acronyms 12
Related Articles 12
References 13

Breath analysis is the sampling and analysis of an analyte in the exhaled breath of living animals, and the subsequent interpretation of the results.[1] Breath sampling is noninvasive of the body as is urine sampling, unlike blood sampling. All three modes of biological monitoring can be very useful and often are necessary. Breath sampling is especially useful for chemically inert volatile organic compounds (VOCs).

The related branches of scientific endeavor include:

- *Pulmonary function testing is the measurement of pulmonary parameters to diagnose pulmonary disorders.[2]*
- *Pulmonary toxicology is the study of adverse effects of inspired toxicants on the respiratory system.*
- *Inhalation toxicology is the study of adverse effects of inspired toxicants on body target organs and tissues. Pulmonary toxicology is thus a subset of inhalation toxicology. The health consequences of ''sniffing'' chemicals and ''poppers'' are included. Many scientists use inhalation toxicology as a term to describe animal experimentation.*
- *Anesthesiology is the intentional exposure of anesthetics in nonadverse doses to humans to aid medical operations where pain would be excessive.*
- *Taste and odor science is the investigation of taste and odor in inspired breath and oral ingestion from natural sources (decaying vegetation, decomposing bodies, and environmental odors) and synthetic sources (food, drink, drugs, cosmetics, perfumes, toiletries, wine tasting and bouquet, beer taste and odor, sewer headspaces and emissions, pesticide applications, garbage and waste sites, natural and potable waters), and environmental and indoor air quality situations. Often taste and odor panels are a part of the evaluation of such phenomena. When used for risk assessment purposes, taste, odor, and other aesthetic considerations are called organoleptic criteria.*

This article will focus on human exhaled breath analysis, except where specified otherwise.

1 HISTORY

Breath analysis has been used for many years:[1–9]

- As a diagnostic technique of good health, as part of medical monitoring. The Hippocratic physicians (Hippocrates of Cos 469–399 BC) and Aristotle (384–322 BC) recommended that physicans sniff the breath of their patients as part of health assessment. Halitosis is defined as bad breath. Modern pulmonary function testing dates from 1846 from the invention of the water spirometer to predict people who would later develop tuberculosis.[2] The maximum voluntary ventilation test was described in 1933. Reference ventilation values for Caucasians were

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

first designated in 1948, but it was not until the late 1950s that the technique was applied to detect occupational lung diseases after the invention of waterless spirometers.

- To assess blood alcohol content, for example for drink-driving testing and in forensics. The alcohol breath test was developed in the 1950s.
- To determine if exposure has occurred to xenobiotics as part of biological monitoring in the workplace. The biological exposure indices (BEIs)[6] of the American Conference of Governmental Industrial Hygienists (ACGIH) were first formulated in 1983.[1]
- To determine if exposure has occurred to xenobiotics as part of biological monitoring in indoor air. The Total Exposure Assessment Methodology (TEAM) studies of EPA have involved breath testing in at least 800 residents since 1979.[1]

2 THE RESPIRATORY SYSTEM

The purpose of the respiratory system (respiration means to breathe again) is to provide inhaled oxygen O_2 in air to oxygenate the blood, and to allow exhalation of waste gases such as the ultimate gaseous products of metabolic oxidation (carbon dioxide CO_2, carbon monoxide CO, sulfur dioxide SO_2, and nitrogen oxides NO_x), and the ultimate gaseous products of anaerobic metabolism (hydrogen H_2, methane CH_4, hydrogen sulfide H_2S, and ammonia NH_3). Xenobiotics are breathed in and out adventitiously.[1]

The respiratory system consists[3] of two lungs and the tubes through which air reaches them and is also expelled. An inspired breath through the nose flows through the nasal cavities to the pharynx, to the larynx or voice box, and then into the trachea or windpipe. Air then continues into the bronchi, one going to each lung. The bronchi subdivide into bronchioles and keep subdividing until the smallest microscopic air sacs are reached, the alveoli. The alveoli are the sites of O_2 and CO_2 exchange with the blood across a thin double membrane, the intermembrane fluid being piped into the lymph system. Mouth breathing directs outside air into the pharynx.

The easiest access to the blood by xenobiotics is when they are in the form of gases (1 mole of the compound at ambient conditions is completely in the form of a gas) or vapors (the gaseous phase of a compound that is a liquid or solid at ambient conditions). Aerosols are deposited in the respiratory system according to particle size. All aerosol sizes that can be conducted through the nasal passages are termed ''inhalable'' and are $<100\,\mu m$ in aerodynamic diameter (the diameter of the equivalent sphere of density $1\,g\,cm^{-3}$ of the same terminal velocity as the test particle). Those that penetrate into the alveoli ($<10\,\mu m$ in aerodynamic diameter) are termed ''respirable''. The fraction that is inhalable but not respirable is termed ''thoracic''. Particles $\leq 0.1\,\mu m$ and between 0.5 and $5.0\,\mu m$ in aerodynamic diameter deposit efficiently in the alveoli. Those between 0.1 and $0.5\,\mu m$ in aerodynamic diameter are not efficiently deposited and may also be breathed out.[1]

The compartments of the nose, pharynx, larnyx, trachea, and bronchi contain mucus cells. They contain cells that are ciliated to sweep deposited foreign particles towards the esophagus to be swallowed with the saliva. Thus large particles are ingested rather than being absorbed through the lungs or exhaled.

2.1 The Nose

The function of the nose is to humidify and filter the air, and bring it to body temperature.The nose contains the receptors for odor. The two nostrils contain hairs which prevent large particles from entering beyond, to the two nasal cavities that are separated by the nasal septum. Large aerosols $>10\,\mu m$ in aerodynamic diameter deposit in the nose. Three bony projections, the conchae, project from the back walls of the nose. All the surface cells in the nasal cavities are ciliated, and beat towards the throat. The sinuses in the bones of the skull near the nose are connected to the nasal cavities.

2.2 The Mouth and the Throat

The nasal cavities connect directly to the pharynx (throat) via the nasopharynx, then to the oropharynx directly behind the mouth, and then to the laryngopharynx that leads to the larynx. Just down from the opening to the larnyx (the glottis) is the smaller opening into the esophagus, part of the digestive tract. Large particles deposit on the tongue and on the walls of the mouth and pharynx in mouth breathing.

The larnyx or voice box is supported by rings of cartilage in its walls. It contains the vocal cords which vibrate during exhalation. During swallowing, the larynx is sealed off by the epiglottis. When particles enter the larynx, coughing is the reflex.

2.3 The Windpipe

The trachea or windpipe extends from the larynx to the middle of the chest and is also supported by C-shaped rings of cartilage. Irritation causes sneezing.

2.4 The Bronchi

The trachea divides into the right and left bronchus, each leading to a lung. The bronchus enters each lung

at a depression called the hilus, where blood vessels and nerves also enter each lung. The progressive subdivision of primary bronchi, to secondary bronchi, to bronchioles, to the alveoli is referred to as the bronchial tree. The bronchi and bronchioles are supported by C-shaped rings of cartilage like the trachea.

2.5 The Lungs

Each lung has more than a million bronchioles. Each bronchiole leads to a cluster of alveoli. The wall of each alveolus consists of a single layer of epithelial cells and elastic fibers that allow stretching and contraction. Each alveolus is surrounded by a network of capillaries, the deoxygenated part being from the pulmonary artery, and the oxygenated part emptying into the pulmonary vein. The distance across the alveolar membrane and the capillary membrane is about 500 nm. Alveoli are coated with lung surfactant (mostly C_{16} fatty acid lecithins) to prevent them from collapsing. The major mechanism for removal of microorganisms and particles in the alveoli is engulfment by pulmonary alveolar macrophages. The effective surface area of the alveoli is about 70 m^2.

The lungs occupy the thoracic cavity, the right lung having three lobes and the left, two. The lungs are separated physically by the mediastinum that contains the heart, esophagus, thymus gland, and parts of other organs. Each lung is covered with a pleural membrane, the visceral pleura, and the thoracic cavity is lined with the parietal pleura. The space between the two pleura is the pleural cavity and is filled with a fluid that reduces friction between the two pleura during breathing. The thoracic cavity is bounded on the front and sides by the chest wall that contains the ribs. Its floor is a strong, dome-shaped muscle, the diaphragm.

2.6 Breathing

Pulmonary ventilation is the movement of air in and out of the lungs.[3] Breathing in is inspiration or inhalation. Breathing out is expiration or exhalation. During inhalation the diaphragm contracts and flattens and the intercostal muscles contract. Since the fluid on the pleural membranes causes the lung surfaces to remain in contact with the thoracic cavity wall, the increased size of the chest cavity causes the space within each lung to increase. This causes a partial vacuum causing air to rush in to equalize inside and outside pressures.

Exhalation occurs when the diaphragm and intercostal muscles relax, allowing the elastic tissues to recoil and the thoracic cavity to decrease in size. Air is pushed out of the lung until the inside and outside pressures are again equal. Not all the air is pushed out of the lungs since the pleural cavity fluid is maintained at a pressure lower than atmospheric.

2.7 Gas/Vapor Exchange

The circulatory system delivers oxygenated blood to target organs and tissues, and delivers back blood saturated with CO_2 for exhalation. Each gas diffuses through the thin lining of the capillary and the thin lining of the alveolus. The usual concentration of O_2 in inhaled air is[3] about 20.9% (v/v), and in exhaled alveolar air is about 14%.[2] Similarly, the concentration of CO_2 is 0.040% (or 400 ppmv (parts per million by volume)), and 5.0–5.6%, respectively.[3] The concentration of N_2 in both remains approximately constant at about 79%, as does argon concentration at 0.9% (9000 ppmv). The relative humidity (RH) of exhaled alveolar air is close to 100% or about 6.5% of alveolar air volume.

Thus mixed exhaled air and end-exhaled air can be distinguished on the basis of CO_2 and water vapor concentrations. It has been suggested that alveolar air concentrations of analytes be normalized to 5% CO_2 as denominator for comparison purposes,[5] just as creatinine or specific gravity is utilized for urine analyte concentrations to correct for dilution effects. A water vapor criterion is impractical because of great variation of relative humidities in ambient air.

The above implies that vapors and gases of xenobiotics will also be absorbed adventitiously into the blood by diffusion, and will be similarly exhaled. The blood air partition coefficient K_{ba} at equilibrium at 37 °C is defined for an analyte of equilibrium concentrations c_{blood} and c_{air} in Equation (1) as:[1]

$$K_{ba} = \frac{c_{blood}}{c_{air}} \tag{1}$$

If analyte $K_{ba} < 5$, alveolar air does not reflect mixed venous blood analyte concentration.

The inverse of K_{ba} determines analyte alveolar breath concentration after analyte absorption, but is not the same as the dimensionless Henry's law constant $K_H = K_{aw}$ for an analyte in air/water at 25 °C (Equation 2), though the two are related:

$$K_H = K_{aw} = \frac{c_{air}}{c_{water}} \tag{2}$$

Thus compounds that can be exhaled in the gaseous state must be gases, or the vapors of volatile liquids or solids with high vapor pressures.

Other characteristics propitious for breath monitoring for exposing analytes include:

- little potential for metabolism, for example, most anesthetics, aliphatic amines, and chlorofluorohydrocarbons;
- low blood and water solubilities, Equations (1) and (2), respectively;

- a low fat/blood partition coefficient K_{fb} defined in Equation (3):

$$K_{fb} = \frac{c_{fat}}{c_{blood}} \quad (3)$$

K_{fb} is related[1] to the octanol/water partition coefficient K_{ow} for chlorinated/brominated hydrocarbons and unsubstituted hydrocarbons by Equation (4):

$$K_{fb} = 70.9 \log K_{ow} - 127 \quad (4)$$

for compounds of $\log K_{ow} > 2.0$. The relationship of analyte K_{ow} to its water solubility c_{water} in micromolar units at 20–25 °C is provided[4] in Equation (5):

$$\log K_{ow} = 5.00 - 0.670 \log c_{water} \quad (5)$$

for compounds with $\log K_{ow}$ over the range 1.0–6.7.

For example, the relatively inert chemical perchloroethylene at 25 °C has the following measured characteristics:[1] $\log K_{ow} = 2.60$, $c_{water} = 2.413$ mM; and vapor pressure = 2.53 kPa corresponding to a c_{air} of $19/(0.08205 \times 298 \times 760) = 1$ mmol L^{-1}, assuming an ideal gas. At 37 °C, the K_{fa} is $(1.0–1.5) \times 10^3$, and K_{ba} is 9–15. Since perchloroethylene is not much metabolized (1–3%), breath analysis might be potentially useful, sensitive, and selective.

The calculated K_H at 25 °C from Equation (2) is $1/2.413 = 0.414$ compared with the $1/K_{ba}$ range of 0.067–0.11 at 37 °C. This shows that the solubility of the analyte in blood is at least four times that in water, though the temperatures are different. The calculated $\log K_{ow}$ using Equation (5) is 2.73. The measured value is 2.60, a positive bias of 5.0% using the log scale, but one of 35% using transformed values. The calculated K_{fb} from Equation (4) using the measured K_{ow} is 57, compared with the theoretical value of $K_{fa}/K_{ba} = K_{fb}$ of 67–167, not far from the lowest part of the range. Thus most of the estimates are within an order of magnitude of measured values.

2.8 Reference Breathing Parameters

The lung volumes breathed in and out are dependent on physical activity, height, age, gender, and ethnicity, and show large interindividual differences. Volumes are all corrected to body temperature, body pressure, and saturated water vapor. These volumes must be measured before interpretation of breath concentrations can be made.

The vital capacity (VC) of the lung is the difference in volume between the total lung volume and the dead space (residual volume) without forcing exhalation. The dead space is the volume occupied by the nasal and oral cavities, larynx, trachea, and bronchi. The greater the physical activity the greater is the air volume breathed in and out for a given person. VC within a specific ethnic group for the same age increases linearly with increasing height.[5] For a given height for adult men of the same age, the order of decreasing VC is: Swedes, other Europeans, African-Americans, and Indians/Chinese/Bantu/Other Africans.[5] For example for 40-year-old males of 170 cm height, the average VCs were respectively in liters: 5.0, 4.2–4.3, 4.1 and 3.35–3.80.[5] Thus people with smaller lung VCs who are of the same age and height but are of different ethnicity will show higher mixed exhaled breath concentrations than those with higher VCs after exposure to the same air concentration of analyte, or to the same absorbed dose. For males of the same height and ethnicity, VCs decrease with increasing age after adolescence. Weight does not affect VC much because of the competing effects of fat deposition and a large frame. Women have smaller VCs than men of the same height, age, and ethnicity. The threshold for a significant effect on VC is 10% of baseline.[7] The forced expired volume in 1 s, (FEV_1), is the most widely used pulmonary function test because it is short, easily performed, and highly reproducible. The threshold for a significant effect on FEV_1 is a 20% change in baseline.[7] The FEV_1/FVC ratio is also often used, where FVC is the forced vital capacity. Most Caucasian male adults doing light physical activity breathe about 12–20 times per minute, about 20 000 times per day, and inhale about 35 pounds (13.4 m^3) of dry air per day.[3] Each breath therefore has an average volume of about 672 mL.

The major predictive equations for males and females are known for Caucasians.[8]

For males ≥ 18 years of age, for height h in inches, age a in years and volumes in L Equations (6–8) give:

$$VC = 0.113h - 0.032a - 1.86 \quad (6)$$

$$FVC = 0.148h - 0.025a - 4.241 \quad (7)$$

$$FEV_1 = 0.092h - 0.032a - 1.26 \quad (8)$$

For Caucasian males < 18 years of age, for h in cm Equations (9) and (10) give:

$$VC = FVC = 4.4 \times 10^{-6} h^{2.6727} \quad (9)$$

$$FEV_1 = 0.86\ VC \quad (10)$$

For Caucasian females ≥ 18 years of age, for h in inches, Equations (11–13) give:

$$VC = 0.1h - 0.012a - 2.8 \quad (11)$$

$$FVC = 0.115h - 0.024a - 2.852 \quad (12)$$

$$FEV_1 = 0.089h - 0.025a - 1.932 \quad (13)$$

For Caucasian females < 18 years of age, for h in cm Equations (14) and (15) give:

$$VC = FVC = 3.3 \times 10^{-6}\, h^{2.7194} \quad (14)$$

$$FEV_1 = 0.86\, VC \quad (15)$$

Both VC and FEV_1 decrease during metal fume fever, and during acute exposures to toluene diisocyanate, red cedar dust, detergent enzymes, tobacco smoke, and formaldehyde.[9]

Exhaled breath samples are generally taken voluntarily from sitting adults at rest for a set time of 10 s for each breath sampled, except in exercise physiology or kinesiology. The average dead space volume in Caucasian adults is about 150 mL.[1,6] This volume is 30% (range 20–40%) of the total tidal volume of average volume 500 mL (400–600 mL range). Thus for the case of an analyte excreted in the breath from the blood, the highest concentration of analyte is in the alveolar air, assuming no analyte is breathed in. To ensure the absence of the analyte in room air where the exhaled breath is being taken, some investigators provide pure air for inspiration by valved respirators connected to compressed air before collection of each exhaled breath. This procedure is highly recommended, though it could also be done more conveniently by inspiring pure air contained in a gas bag.

When the dead volume and alveolar air are both sampled, this is termed ''mixed exhaled air''. When the dead space is expelled and only alveolar air is sampled, this is termed ''end-exhaled air''. Correspondingly, if the analyte is being breathed in and is chemically inert, the highest concentration is found in the dead space, and the most dilute in alveolar air assuming no excretion of analyte from the blood.

Thus, the interpretation of the concentration of analyte found in breath samples in addition to volume-dependent variables depends on:

- past history of exposure;
- whether high physical activity is involved or not if the major exposure was by inhalation;
- when the last inhalation exposure occurred;
- type of analyte; if the latter is reactive (for example 2,4-toluene diisocyanate and phosphine) or very water soluble (for example formaldehyde and methanol), breath concentrations are not interpreted simply;
- smoking history; the heavier the smoker of the same age, height, and ethnicity, the lower the VC. The average smoker loses VC at the rate of 50 mL $year^{-1}$;
- medical conditions; the following lower VC: acute or chronic obstructive lung disease, asbestosis (lung damage caused by exposure to asbestos), asthma, beryllosis (lung damage caused by beryllium exposure), bronchitis, byssinosis (lung damage caused by cotton dust and microbial exposure), emphysema, fibroses, interstitial lung disease, lung cancer, lung allergy, pleurisy, pneumoconioses, pneumonia, pneumothorax, pulmonary edema, rhinitis, sarcoidosis, silicosis (lung damage caused by silica exposure), and tuberculosis. Hypertension raises VC;
- medications; use of bronchodilators (like methylxanthines, β-agonists, and anticholinergics) and bronchoconstrictors (such as histamine or methacholine) affect VC and the dead space volume.

3 BREATH

3.1 Gross Chemical Composition of Exhaled Breath

Table 1 shows the major compounds in expired breath apart from the inhaled components of pure air, methane, and hydrogen. The compounds are[10] mostly gases and

Table 1 Ranges of endogenous vapors in exhaled breath[10]

Compound (CAS-RN)	Range ($\mu g\, h^{-1}$)	Compound (CAS-RN)	Range ($\mu g\, h^{-1}$)
Acetone (67-64-1)	1.3–2000	Ethanol (64-17-5)	11–440
n-Butane (106-97-8)	0.24–9.4	Furan (110-00-9)	0.25–98
1-Butene (106-98-9)	1.3–510	Indene (95-13-6)	0.14–83
Dimethyldifluorosilane	1.2–140	Isoprene (78-79-5)	15–390
Dimethyl sulfide (75-18-3)	0.54–53	Isopropyl alcohol (67-63-0)	2.0–160

These compounds occurred in at least seven out of 10 samples from eight volunteers not overly exposed to xenobiotic vapors and gases. Other compounds detected less frequently included: acetaldehyde (75-07-0), 2-butene(*cis* [590-18-1] and *trans* [624-64-6]), *n*-butyl alcohol (71-36-3), caprylic alcohol (111-87-5), carbon disulfide (75-15-0), cyclohexyl alcohol (108-93-0), dimethyl disulfide (624-92-0), ethyl acetate (141-78-6), ethyl acetylene (503-17-3), ethylene (74-85-1), 1-hexene (592-41-6), 2-hexene (592-43-8), isobutyl alcohol (78-83-1), methanol (67-56-1), methyl acetylene (74-99-7), methyl cyclohexane (108-87-2), methyl cyclopentane (96-37-7), 1-octyne (629-05-0), 1-pentene (109-67-1), propane (74-98-6), and *n*-propyl alcohol (71-23-8).

vapors from the metabolic processes of the body or those emitted by symbiotic microorganisms.

The products of mammalian metabolism include:

- the volatile ketone body acetone (fruity smell of odor threshold about 400 ppbv (parts per billion by volume[4]) when fat breakdown predominates, or in diabetes. Breath concentrations in excess of 2 ppmv have been measured;
- isoprene from the dephosphorylation of isopentenyl pyrophosphate and dimethylallyl pyrophosphate, and during the synthesis of geranyl (farnesyl) pyrophosphate during the synthesis of squalene in the cholesterol biosynthetic pathway;
- excess aldehydes like acetaldehyde (odor threshold about 4 ppbv[4]) released from transketolase activated aldehyde transfers. Biological oxidation processes via cytochrome- P450 produce formaldehyde, acetaldehyde, acetone, propanal, 2-butanone, *n*-butanal, *n*-pentanal, and *n*-hexanal in human breath,[12] as well as acetone, 2-butanone, 2-pentanone, 2-hexanone, and 2-heptanone in rat breath.[13] Cytochrome 2E1 is largely responsible for the formation of short chain methyl ketones in rats;[13]
- nonmethane alkanes like ethane (odor threshold about 12 ppmv[4]) from n-3 polyunsaturated fatty acids, *n*-butane from n-5 polyunsaturated fatty acids, *n*-pentane[11] (odor threshold about 2 ppmv) from n-6 polyunsaturated fatty acids, and *n*-hexane from n-7 polyunsaturated fatty acids by lipid peroxidation starting from the methyl end of the fatty acid. Unsaturated aliphatic hydrocarbons except for acetylene are also produced by lipid peroxidation.

Those compounds emitted by symbiotic microorganisms include:

- acetaldehyde from the decarboxylation of pyruvate from glycolysis to form acetaldehyde, and its further reduction product ethanol from fungi, this process being in competition with L-lactate formation;
- sulfides and disulfides from anaerobic bacteria. Dimethyl sulfide has an odor threshold of about 200 pptv (parts per trillion by volume);[4]
- acetylene from anaerobic bacteria, along with methane and hydrogen.

Ethanol and acetaldehyde could be of both metabolic, microbial, or xenobiotic origin. The origins of dimethyldifluorosilane, indene, and furan are unknown. For chromatographic analyses of xenobiotics, the background peaks must not interfere with analytes of interest. Other endogenous compounds have been found including hydrogen cyanide from the oxidation of thiocyanate by salivary peroxidase,[4] and nitric oxide formed mostly from the endothelial cells in the nasal cavity (0.750–1.60 ppm).[15] Breath also contains particles up to 4 particles cm^{-3}, usually below 300 nm aerodynamic diameter, with <2% of the particles being of diameter >1 μm.[16] The proportion of particles increases with increasing physical activity. Particulate matter and excess water are usually caught in an empty impinger just before the sampling device.

3.2 Markers of Xenobiotic Exposure in Exhaled Breath

The major set of recommendations for xenobiotics in breath in the United States are the BEIs of ACGIH provided in Table 2.[17] They are all used for exposing xenobiotics rather than their metabolic progeny. The breath BEIs are the biological equivalent values after absorption of their corresponding threshold limit values (TLV) for workplace physical activity. The TLV is the time-weighted average air concentration in the personal

Table 2 Compounds with ACGIH Breath Testing Recommendations for 1998[17]

Compound (CAS-RN)	MW	TLV (ppmv)	TLV basis	BEI (ppmv)	BEI sampling conditions
Carbon monoxide (630-08-0)	28.01	25	Anoxia	20	End of shift eea
Ethyl benzene (100-41-4)	106.16	100	Irritation	Sq	End of shift at end of work week eea
n-Hexane (110-54-3)	86.18	50	Neuropathy	Sq	End of shift eea
Methyl chloroform (71-55-6)	133.42	350	Anesthesia	40	Prior to last shift of work week
Perchloroethylene (127-18-4)	165.80	25	Irritation	5	Prior to last shift of work week eea
Trichloroethylene (79-01-6)	131.40	50	CNS	Sq	

CNS, Central nervous system; MW, molecular weight; ppmv = mg m^{-3} × 24.46/MW; Sq, semiquantitative only and to confirm exposure; eea, end-exhaled air.

breathing zone samples that are taken for 8 h per work day over 5 d per week over a working lifetime that is deemed nontoxic. The BEIs measure the excretion of markers of inhaled absorbed chemicals during a workshift. BEIs do not account for nonwork exposures, or workplace skin or oral absorption.

The distinction between excretion of absorbed marker and exhalation of inhaled chemical is better distinguished with end-exhaled air samples than with mixed exhaled air samples. Thus the BEI breath recommendations are for the use of the former. However when the marker of absorbed dose is the same as the inhaled compound, it is still possible that what is being measured in the breath is the last high exposure or the most recent exposure rather than excretion of the marker. Thus direct reading instrument air concentration measurement should be done to define the last inhalation exposure. This is not common practice, however. The problem does not arise when the marker is different from the inhaled compound, or if only oral or skin absorption of the exposing compound is definitely involved. Breathing in pure air only has been the solution adopted by some investigators.[18]

The breath studies of the TEAM studies of EPA have been reviewed.[1,18–21] Breath samples were taken after 12 h of exposure, the previous 12 h being monitored by breathing zone personal sampling (the sampler was located on the lapel with the pump for the solid sorbent located on the belt). The increased exposure to benzene and styrene for smokers was first proven in these studies. Chamber studies involving 10 h exposures at parts per billion by volume concentrations followed by 24 h breath excretion for humans were also done.[19] The excretion had four phases with measured median times of α, 3–12 min; β, 0.3–2 h; γ, 2–5 h and δ, 1–4 d. The data were analyzed by a four-compartment model. Representative compounds were chosen for study. The fractions of the absorbed dose excreted in the breath at equilibrium were nonpolar aromatics and *n*-decane, 0.06–0.16; chlorinated aliphatics (trichloroethylene and dichloromethane), 0.22–0.23; *n*-hexane, 0.35; and methyl chloroform, 0.88.

Lauwerys and Hoet[22] have provided breath threshold guidelines for 18 chemicals (Table 3). In principle, any volatile inert chemical of low K_{ba} or $\log K_{ow} > 1.0$ can be sensitively sampled in exhaled air after inhalation exposure, as shown in section 4. Table 3 also provides some reference retention, breath excretion, and breath half-time data for these inhaled compounds.

Progeny organic compounds that could constitute positive interferences must also be considered. The major progeny compound markers are listed.[22]

- Breath CO can arise from overexposure to dichloromethane (methylene chloride) or dibromomethane. An 8 h exposure to 150 ppmv of dichloromethane is equivalent to an 8 h exposure to 35 ppmv CO, both producing about 5% (w/w) carboxyhemoglobin HbCO relative to total hemoglobin.
- Breath acetone also arises from the metabolism of isopropyl alcohol,[22] and from the metabolism of C_1 and C_2 halogenated compounds.[1]
- Breath 2,2,2-trichloroethanol can arise from metabolism of methyl chloroform, trichloroethylene, and perchloroethylene as well as trichloroethanol exposure. Since methyl chloroform and perchloroethylene are not much metabolized, the major interference is from trichloroethylene. The pulmonary $t_{0.5}$ (the breath half-time) also differ.
- Smoking is a major contributor to breath concentrations of over 6000 chemicals. Important ones are benzene, CO, ethylene oxide, and styrene.

The following metal markers can be analyzed in breath:

- Mercury: mercury vapor is 74–80% absorbed after inhalation and has been detected in exhaled breath for a few hours after the end of exposure.[22] The excretion half-times in breath are 1 min, and 18 h.[6] Mercury vapor has been detected in the mouth and from exhaled air from mercury amalgam tooth fillings.[23] Mercury is also liberated by metabolism of aryl and alkoxyalkyl mercury derivatives.
- Hydrides: metal hydrides such as arsine, hydrogen mercuride, hydrogen selenide, stannane, stibine, plumbane, and hydrogen telluride should be amenable for detection in breath after exposure to the hydride. Hydrogen selenide has been detected in breath.[24]
- Alkylated metal derivatives: the alkyl derivatives of lead and mercury are also probably partially eliminated in the exhaled breath after inhalation, although this has not been specifically shown. Dimethylmercury may also be exhaled after endogenous formation. Dimethylselenide is excreted in exhaled breath at high selenium absorbed doses.[22] Dimethyl telluride (garlicky odor) is similarly excreted.[22]
- Other covalent metal species: some examples may include include beryllium chloride and probably stable volatile metal esters.

3.3 Medical Breath Testing

There is an extensive literature. Some reviews and experimental studies are cited in Table 4, which gives details of the current uses of medical breath testing.[1,25–42] These tests are particularly important for children, and where

Table 3 Characteristics of xenobiotics sampled in breath discussed by Lauwerys and Hoet[22]a

Compound (CAS-RN)	RT (%)	EX (%)	$t_{0.5}$ (h)	Guideline after 8 h TWA exposure (end-exhaled air)
Acetone (67-64-21)	45	20	4.0	42 ppmv(es) for 150 ppmv
Benzene (71-43-2)	50	10–50	0.42, 2.5, 30	22 ppbv(ds) for 1 ppmv 120 ppbv(ps) for 10 ppmv
Carbon disulfide (75-15-0)	80	5–30	0.018, 0.92	Unsuitable (high interindividual variation)
Carbon monoxide (630-08-0)	80–90	10–20	5(2–7)	12 ppmv(es) for 25 ppmv (nonsmokers)
Cyclohexane (110-82-7)	>90	<10	0.19, 0.96	220 ppmv(ds) for 300 ppmv
Dichloromethane (75-09-2)	55–70			15 ppmv(es) for 50 ppmv
Dimethylformamide (68-12-2)	65–75		3, 7	2.8 ppmv(es) for 10 ppmv
Ethylene oxide (75-21-8)	75–80			0.24 ppmv(ds) for 1 ppmv
Halothane (151-67-7)				0.5 ppmv(ds) for 5 ppmv
n-Hexane (110-54-3)	14–18	10	0.18, 1.7	50 ppmv(ds) for 50 ppmv
(2-Methylpentane) (107-83-5)				420 ppmv (ds) for 500 ppmv
(3-Methylpentane) (96-14-0)				420 ppm(ds) for 500 ppmv
Isopropanol (67-63-0)				210 ppmv(ds) for 400 ppmv
Styrene (100-42-5)	42–90	<5	0.22–0.87, 4–20	10 ppmv(ds) for 50 ppmv
Tetrachloroethylene (127-18-4)	47–90	38	0.25, 4, 96	8 ppmv(ps) for 50 ppmv 60 ppm(ds) for 50 ppmv
Toluene (108-88-3)	40–60	20	3.8(2.6–6)	20 ppmv(ds) for 100 ppmv
1,1,1-Trichloroethane (71-55-6)	24–40	97	1–9	40 ppmv(ps) end of work week 6–20, >26 for 350 ppmv
Trichloroethylene (79-01-6)	45–75	<10	0.33, 3, 30	0.5 ppmv(ps) for 50 ppm 12 ppmv(es) for 50 ppm
Vinyl chloride (75-01-4)	40			Only validated for >25 ppmv
Xylene (1330-20-7)	55–65	3–11	1, 20	Variable

%, Percent of TWA concentration × exposure time; (ds), during shift; (es), end of shift; (ps), 16 h postshift
[a] Some data are from ACGIH[6] for lung retention (RT), breath excretion after absorption (EX), and breath half-time $t_{0.5}$, after exposure to the time-weighted average concentration (TWA) for 8 h.

a patient has lost much blood. There are four types of medical breath tests:

- respiratory gas measurements, for O_2, N_2, argon, water vapor, and CO_2. Of these only carbon dioxide is considered here[28,29] since it is related to malabsorption and bacterial overgrowth breath tests;
- tests for unlabeled markers of symbiotic organism presence. These include the classic hydrogen and methane breath tests;[1,25,26]
- tests for lipid peroxidation.[1,30,31] These include the standard ethane and *n*-pentane breath tests, and overlap the tests for endogenous breath chemicals;
- tests for labeled CO_2 in breath from labeled substrates.[32–42] There are two subdivisions

  (a) Endogenous target organ function such as the aminopyrine and caffeine breath tests of liver function.[1,34–37] Some selective cytochrome-P450 liver function breath tests include: caffeine for 1A2, chlorzoxazone for 2E1, lidocaine for 3A, midazolam for 3A, and erythromycin for 3A inducing systems;[43]

  (b) malabsorption of carbon sources or symbiotic organism overgrowth using a huge number of substrates.[32,33,38–42,44]

The use of compounds with radioactive labels previously inhibited the common use of labeled CO_2 tests. Explosive growth is now occurring in this research area with the advent of ^{13}C-enriched commercial substrates, the common use of gas chromatography/mass spectrometry (GC/MS), and the lack of human subject and human fears about radiolabeled compounds.[44]

3.4 Dental Breath Tests

A major use of breath tests is to define the sources of halitosis (bad breath or malodor). The dental breath

Table 4 Some medical breath tests and their uses

Analyte (CAS-RN)	Use	Refs.
Hydrogen (1333-74-0)	Malabsorption of nutrients such as carbohydrates and essential fats. 79% excreted in flatus and 21% excreted in exhaled breath. Bacterial overgrowth detection. Positive if over 20 ppm or two-fold baseline	1, 25, 26
Methane (74-82-8)	Malabsorption of nutrients such as carbohydrates and essential fats. Some people excrete more methane than hydrogen. Bacterial overgrowth detection	1, 25, 27
Carbon dioxide (124-38-9)	Effectiveness of anesthesia by capnography	28, 29
Ethane (74-84-0) or Pentane (78-78-4)	Lipid peroxidation	11, 30, 31
$^{13}C-$ or $^{14}C-$Urea (58069-82-2 and 57-13-6)	Labeled CO_2 from bacterial infection, especially *Helicobacter pylori* and *Bifidobacteria* via product of urease activity	32, 33
$^{13}C-$ or $^{14}C-$Aminopyrine (1606-67-3)	Labeled CO_2 from liver demethylation	1, 34, 35
$^{13}C-$ or $^{14}C-$Caffeine (78072-66-9 and 58-08-2)	Labeled CO_2 for liver enzymes sensitive for 3-methyl cholanthrene (Cytochrome P450 1A2)	1, 34, 36, 37
$^{13}C-$ or $^{14}C-$Octanoic acid (59669-16-8 and 124-07-2)	Labeled CO_2 measured every 15 or 30 min for 4 or 6 h for gastric emptying rate of solids	38, 39
$^{13}C-$ or $^{14}C-$Triolein (82005-46-7)	Labeled CO_2 for effectiveness of restorative proctocolectomy via ileoanal pouches	40
$^{13}C-$ or $^{14}C-$Triglycerides	Labeled CO_2 for measurement of exocrine pancreatic function	41
$^{13}C-$ or $^{14}C-$Xylose (70849-21-7 and 58-86-6)	Labeled CO_2 for measurement of bacterial overgrowth in lower bowel	42

tests concentrate on odors from the mouth, the tongue, the oropharynx, the teeth, and the gums.[45–49] Bacterial activity in the oral cavity is the usual reason for mouth odor. The bacterial culprits are usually Gram-negative, but especially *Klebsiella* spp and *Enterobacter* spp. Volatile sulfides and disulfides, cadaverine, putresine, hydrogen sulfide, and volatile acids have been postulated as the odoriferous agents.[47–49] The most effective treatments are by use of zinc acetate impregnated chewing gums which inhibit thiol compound emissions, effective mouth hygiene that controls plaque, and mouth washing with chlorhexidine gluconate which is a bactericide.

4 BREATH SAMPLING

Breath analysis can be direct or a sample of breath enclosed in a container can be taken for laboratory analysis. The direct approach saves time but is more expensive. The container must be large enough to enable sensitive detection of the analyte, the instrumental detection limit being exceeded by increasing sample and subsampling sizes. The more sensitive the analyzer the more likely that it can be used for direct measurement without intermediate containers. Another alternative is to concentrate breath directly or from the breath sample container by using a solid sorbent or impinger technique before later laboratory analysis.

4.1 Gas Bags

Many breath samples are collected in clean nonleaking Tedlar™ or Teflon™ gas bags equipped with an open/shut valve of stainless steel or Teflon™, and a separate or an integrated injection valve suitable for using gas-tight syringe subsampling and also injection of liquids to generate defined concentration atmospheres. An evaluation of 1-L Tedlar™ bags in containing a typical 500 mL single-breath sample has been published.[50] The bags contained separately acetone, 2-butanone, methanol, methyl chloroform, perchloroethylene, and *m*-xylene in either an atmosphere of saturated water vapor, or dry nitrogen. No significant difference occurred at $p \leq 0.05$ for each compound except for methanol, where the wet-bag reading was 10% lower than the dry bag. When excess water was present beyond three times saturation, acetone and 2-butanone also gave lower wet bag readings. Concentrations of many organic compounds and mixtures in Tedlar bags have also been shown to provide the same responses to detectors not sensitive to water vapor at 90% RH and dry conditions.[51,52]

A typical breath sampling set-up consists of the bag connected to an impinger or Buchner flask that catches spittle and drool.[53] All connections must be of clean Teflon™ tubing of minimum length and diameter wide enough to minimize resistance. The sterilized mouthpiece that is connected to the impinger or Buchner flask must be able to fit into the pursed lips for air blowing purposes. The

total possible volume of the bag should be chosen so that the bag is at least half-filled with breath sample. The bag should never be completely full since this promotes leaks at seams. During injection or subsampling the analyst must be careful not to puncture the bag walls opposite the injection or subsampling port. The gas bag technique is favored by field industrial hygienists,[53] but has also been used by clinicians[54,55] to sample for hydrogen, nitrous oxide, and other gases. The gas-tight syringe mode of introduction to the analyzer minimizes any matrix effects of breath since the air constituents elute first.

4.2 Summa Cannisters

Summa cannisters are stainless steel cannisters which can be evacuated.[18,52] The stainless steel interior is relatively inert and some cannisters are also coated inside with Teflon™. Some disadvantages relative to Tedlar™ gas bags are their weight, greater expense for the same volume, being opaque, and leaking often at their on/off valves. The non-Teflon™ coated models are more prone to corrosion and analyte adsorption/desorption cycles.

The Summa cannister technique is favored by many samplers, including during EPA's TEAM studies.[18,56] Some more recent examples include:

- sampling of methyl tertiary butyl ether (MTBE) in gasoline stations;[57]
- sampling of vinyl chloride, *cis*-1,2-dichloroethene, chloroform, bromodichloromethane, and trichloroethylene.[58]

4.3 Solid Sorbent Sampling

This is probably the major mode of sampling of vapors because of its low weight and better storage and transport characteristics. The concentrated analyte has to be stored and transported, rather than the breath sample itself. There are two general modes: physical adsorption and chemisorption. Both forms are amenable to thermal desorption for maximum sensitivity, or solvent desorption at high concentrations and long sampling times. Subsequent thermal desorption also minimizes any matrix effects of breath. However, the solid sorbent must be thermally conditioned beforehand. Solvent desorption will cause masking of breath constituents where the solvent and its impurities elute on chromatographic columns. Since a fraction of the sample is injected on-column, its sensitivity is less than thermal desorption techniques.

Some recent examples of physical adsorption of breath compounds include:

- 1259 breath VOCs[59]
- Ethanol, acetone, and isoprene[60]
- VOCs in microgravity[61]
- Benzene.[62]

A recent example of chemisorption of breath compounds involved the reaction of aldehydes and ketones with 2,4-dinitrophenylhydrazine coated on silica gel.[63]

These techniques can be used directly by blowing through the solid sorbent tube connected to a dry gas meter to record the volume, or after collection of breath in a container where the breath can be subsampled at a constant flow rate for a given time.[53] The solid sorbents must be able to tolerate high relative humidities. Thus solid sorbents like Tenax™ TA, XAD resins, and Chromosorb porous polymers are favored for both physical adsorption and chemisorption.[53]

4.4 Glass Tubes

Though these are not favored because of their fragility, 40 mL glass tubes equipped with Teflon™ stopcocks and stoppers, and silanized to prevent wall adsorption have been used to contain breath samples containing methyl chloroform, carbon tetrachloride, and trichloroethylene vapors.[64] Glass tubes have the advantage of being evacuable and transparent.

4.5 Other Methods

There are many other breath collection methods. Analyte impinger collection is usually not done because the contents are spilled easily in the field, on storage, and during transport. Impingers are also fragile. A cheap homemade device has captured breath samples containing chloroform and allows transfer to an evacuated headspace vial for subsequent analysis.[65] Cannulas connected to direct reading instruments have been inserted into the nose for sampling purposes, but this is invasive.[15] Another device also allows thermal desorption.[66] Cryogenic concentration has been reported to concentrate inflammation mediators[67] and was used initially in the field.[10] Thorium and radon progeny have been collected on double filter cassettes.[68]

5 CHEMICAL ANALYSIS OF BREATH

5.1 Introduction

There are two types of analysis:

- those for single analytes involving chromatographic or very selective methods
- those for a class of analytes, for example, direct reading methods.

As for the analysis of single VOCs in other branches of chemistry, capillary GC/MS is the gold standard, followed by capillary gas chromatography/Fourier transform infrared spectrometry (GC/FT/IR). Capillary and packed column GC/detection are screening techniques that provide maximum breath concentrations. High performance liquid chromatography (HPLC) methods are rare except for the 2,4-dinitrophenylhydrazones of aldehydes and ketones and, although an ultraviolet detector is used, specificity is conferred by the selectivity of the derivatization.[63] Direct reading methods are useful screening techniques that may signal the need for a second look at breath sampling with single compound analysis. However, sensitivity and selectivity may be major problems with direct reading methods.

5.2 Gas Chromatography and Mass Spectrometry Methods

GC/MS has been used to identify, quantify, and confirm most of the VOCs in breath beginning in the early 1970s.[10,69] The early work was done on packed GC columns and used magnetic sector mass spectrometers.[10] Most of the recent work utilizes capillary GC and quadrupole or ion trap MS (mass spectrometry). Table 1 summarizes the major analytes quantified and confirmed by GC/MS.[10]

Among selected breath components quantified and confirmed by capillary GC/MS within the last five years are the following:

- organic solvents in the breaths of workers[10,66,70]
- organic solvents in the breaths of home dwellers or in the ambient environment[18,19,21,56–59,71]
- Endogenous VOCs;[11,31,59,60] the most sophisticated of these studies[59] involves adsorption of exhaled breath on a solid sorbent, thermal desorption into a cryotrap, and then a second thermal desorption into the GC/MS. Multivariate techniques allowed identification of patients having characteristic breath volatiles (large concentrations of *n*-pentane and carbon disulfide) for schizophrenia[72]
- $^{13}CO_2$ from ^{13}C-labeled metabolism studies (see references 35–39, 41–44)
- organic solvents in the breaths of volunteers exposed to known solvents in chamber studies or after internal injection.[71]

As expected, screening breath analysis has been done with detectors less expensive than MS. The breath concentrations so found are maximal ones and are not confirmed.[62,64,65,73] This is not so if a solvent of known purity is exposed to volunteers in chamber studies, or after internally provided doses of pure solvents.[65,71]

5.2.1 The Selected Ion Flow Tube Method

A major new variation of the MS technique that is direct reading is the selected ion flow tube (SIFT) technique.[74–78] The breath trace gases but not the major gases in air and breath are subjected to chemical ionization by ion–molecule reactions using "soft" proton transfer from preselected ions (H_3O^+, NO^+, and O_2^+) in a fast flow of inert carrier gas. The sensitivity for most endogenous compounds is about 10 ppbv. The time response is 20 ms, allowing concentration changes to be followed during a normal breathing cycle. Compounds such as acetone, nitric oxide, nitrogen dioxide, ammonia, isoprene, and methanol can be analyzed simultaneously directly or from gas bags without a GC column. The concentration of NH_3 was shown to be proportional to the infection of *Helicobacter pylori*,[76] usually measured through labeled CO_2 produced from the urease metabolism of labeled urea (section 3.3). The breath concentration of ammonia also increased ten times during kidney failure and the effectiveness of dialysis treatment can be determined by monitoring ammonia.[77] Isoprene concentration increased during stress.[77] The volatiles of tobacco smoke, crushed garlic, and masticated mint were also detected in breath samples.[76] Breath concentrations of *n*-butylamine (fishy odor) and other amines were also correlated to kidney problems.[77] The SIFT technique was able to quantify a mixture of eight chemicals of different MW in 20 ms without a column.[78] The response vs concentration curves were linear from 10 ppbv to 30 ppmv. The technique is being miniaturized to make it portable.[77,78] It cannot resolve isomers like those of xylene. It may not resolve low MW compounds when higher MW compounds are also present that can give rise to ions of the same m/z (mass-to-charge ratio).

Another GC/MS avenue for breath sampling is to use the only commercially available portable direct reading GC/MS, the Hapsite.[79] No such uses have been reported.

5.3 Infrared Methods

FT/IR methods have been used to analyze some breath organics and CO_2. An example is the analysis of methanol in the end-exhaled breath collected in a Mylar gas bag after inhalation of 0, 100, 200, 400, and 800 ppmv, and after dermal exposure by immersing one hand in the pure solvent.[80] The spectrometer was equipped with a mercury–cadmium-telluride nitrogen-cooled detector. The detection limits were 500 ppb for methanol and 2000 ppm for CO_2. The concentration of methanol

in end-exhaled air did not become proportional to blood methanol concentration until 2 h after each 8 h inhalation exposure ended. Methanol peaked in the breath 15 min after a 16 min dermal exposure of one hand. The headspace GC/flame ionization detection (FID) technique for measuring methanol in blood may have been suspect in this study.

Direct reading infrared (IR) methods are utilized directly on breath or breath in gas bags for quantitation of CO_2 by capnometers,[1,28,29,81] and ethyl alcohol by the standard Intoxilyzer 5000S instrument.[82–85] Any compounds that absorb at the same wavelength as the analyte are positive interferences. Thus, for breath alcohol using the Intoxilyzer 5000S, the endogenously produced acetone, isopropyl alcohol, and methyl ethyl ketone all interfere as shown by parallel GC headspace analysis.[82] Solvent vapors that may interfere at high concentrations are toluene, *m*-xylene, *o*-xylene, methanol, and isopropyl alcohol.[85]

5.4 Other Methods

Nitric oxide is detected by a portable chemiluminescence detector.[15,54,81] Portable direct reading electrochemical detectors have been used to detect breath ethyl alcohol,[84] and carbon monoxide.[53,86] The volatile mouth total thiols and disulfides thought to be responsible for halitosis can be measured directly by a halimeter.[48,49] Gamma counting is used to detect exhaled thorium collected on filters.[68] Colorimetric detector tubes can be used for specific gases.[87] The accuracy of the latter is ±25% and the interferences must be tested for separately.

Gases of high concentration such as hydrogen,[55] methane, oxygen, nitrogen, argon, water vapor, and carbon dioxide can be quantified in the same sample by the standard method of GC/thermal conductivity detection using molecular sieve or porous-layer open tubular columns.[88]

ACKNOWLEDGMENTS

Salary support is acknowledged from the UCLA Center for Occupational and Environmental Health.

ABBREVIATIONS AND ACRONYMS

ACGIH	American Conference of Governmental Industrial Hygienists
BEI	Biological Exposure Indices
CNS	Central nervous system
EX	Breath Excretion after Absorption
FEV_1	Forced Expired Volume in 1 s
FID	Flame Ionization Detection
FT/IR	Fourier Transform Infrared Spectrometry
FVC	Forced Vital Capacity
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High Performance Liquid Chromatography
IR	Infrared
MS	Mass Spectrometry
MTBE	Methyl Tertiary Butyl Ether
MW	Molecular Weight
RH	Relative Humidity
RT	Lung Retention
SIFT	Selected Ion Flow Tube
TEAM	Total Exposure Assessment Methodology
TLV	Threshold Limit Value
TWA	Time-weighted Average Concentration
VC	Vital Capacity
VOC	Volatile Organic Compound

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Laboratory Instruments in Clinical Chemistry, Principles of • Urinalysis and Other Bodily Fluids

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments • Photoacoustic Spectroscopy in Trace Gas Monitoring • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Formaldehyde, Environmental Analysis of

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Electrochemical Sensors for Field Measurements of Gases and Vapors • Field-based Analysis of Organic Vapors in Air

Field-portable Instrumentation cont'd **(Volume 5)**
Microelectromechanical Systems Technology Applied to the Miniaturization of Field Instrumentation • Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements • Radon, Indoor and Remote Measurement of • Solid-phase Microextraction in Analysis of Pollutants in the Field • Solid-state Sensors for Field Measurements of Gases and Vapors

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air • Sensors in the Measurement of Toxic Gases in the Air

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Theory of Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

Radiochemical Methods **(Volume 14)**
Actinides and other Alpha-emitters, Determination of • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. S.S. Que Hee, *Biological Monitoring: An Introduction*, Van Nostrand Reinhold, New York, 1993.
2. A.F. Wilson (ed.), *Pulmonary Function Testing Indications and Recommendations*, Grune and Stratton, New York, 1985.
3. E.P. Solomon, *Introduction to Human Anatomy and Physiology*, W.B. Saunders Company, Philadelphia, 1992.
4. K. Verschueren, *Handbook of Environmental Data on Chemicals*, 2nd edition, Van Nostrand Reinhold, New York, 1983.
5. J.F. Murray, *The Limitations of Pulmonary Function Testing*, Pulmonary Function Testing Indications and Recommendations, ed. A.F. Wilson, Grune Stratton, New York, 1–8, 1985.
6. American Conference of Governmental Industrial Hygienists (ACGIH), *Documentation of the Biological Exposure Indices*, ACGIH, Cincinnati, 1991.
7. W.M. Alberts, J.W. Ramsdell, 'Bronchial Challenge', in *Pulmonary Function Testing Indications and Recommendations*, ed. A.F. Wilson, Grune Stratton, New York, 57–68, 1985.
8. J.G. Mohler, G.A. Wolff, 'Computers in Pulmonary Medicine', in *Pulmonary Function Testing Indications and Recommendations*, ed. A.F. Wilson, Grune Stratton, New York, 255–274, 1985.
9. K.H. Kilburn, R. Warshaw, 'Pulmonary Function Testing for Occupational Epidemiology and Disability', in *Pulmonary Function Testing Indications and Recommendations*, ed. A.F. Wilson, Grune Stratton, New York, 329–337, 1985.
10. J.P. Conkle, B.J. Camp, B.E. Welch, 'Trace Composition of Human Respiratory Gas', *Arch. Environ. Health*, **30**(6), 290–295 (1975).
11. M. Phillips, J. Greenberg, M. Sabas, 'Alveolar Gradient of Pentane in Normal Human Breath', *Free Radical Res.*, **20**(5), 333–337 (1994).
12. Y. Lin, S.R. Dueker, A.D. Jones, S.E. Ebeler, A.J. Clifford, 'Protocol for Collection and HPLC Analysis of Volatile Carbonyl Compounds in Breath', *Clin. Chem.*, **41**(7), 1028–1032 (1995).
13. J.M. Mathews, J.H. Raymer, A.S. Etheridge, G.R. Velez, J.R. Bucher, 'Do Endogenous Volatile Organic Chemicals Measured in Breath Reflect and Maintain CYP2E1 Levels in Vivo?', *Toxicol. Appl. Pharmacol.*, **146**(2), 255–260 (1997).
14. P. Lundquist, H. Rosling, B. Sorbo, 'The Origin of Hydrogen Cyanide in Breath', *Arch. Toxicol.*, **61**(4), 270–274 (1988).
15. E.A. Ferguson, R. Eccles, 'Relationship between Nasal Nitric Oxide Concentration and Nasal Airway Resistance', *Rhinology*, **35**(3), 120–123 (1997).
16. C.I. Fairchild, J.F. Stampfer, 'Particle Concentration in Exhaled Breath', *Am. Ind. Hyg. Assoc. J.*, **48**(11), 948–949 (1987).
17. American Conference of Government Industrial Hygienists, *1998 TLVs and BEIs*, ACGIH, Cincinnati, 1998.
18. K.W. Thomas, E.D. Pellizari, S.D. Cooper, 'A Canister-based Method for Collection and GC/MS Analysis of Volatile Organic Compounds in Human Breath', *J. Anal. Toxicol.*, **15**(1), 54–59 (1991).
19. L. Wallace, T. Buckley, E. Pellizzari, S. Gordon, 'Breath Measurements as Volatile Organic Compound Biomarkers', *Environ. Health Perspect.*, **104** (Suppl. 5), 861–869 (1996).
20. L.A. Wallace, W.C. Nelson, E.D. Pellizzari, J.H. Raymer, 'Uptake and Decay of Volatile Organic Compounds at Environmental Concentrations: Application of a Four-compartment Model to a Chamber Study of Five Human Subjects', *J. Expos. Anal. Environ. Epidemiol.*, **7**(2), 141–163 (1997).
21. A.B. Lindstrom, J.D. Pleil, 'A Methodological Approach for Exposure Assessment Studies in Residences Using

Volatile Organic Compound-contaminated Water', *J. Air Waste Management Assoc.*, **46**(11), 1058–1066 (1996).
22. R.R. Lauwerys, P. Hoet, *Industrial Chemical Exposure: Guidelines for Biological Monitoring*, 2nd edition, Lewis Publishers, Boca Raton, 1993.
23. B.B. Eley, 'The Future of Dental Amalgam: a Review of the Literature. Part 3: Mercury Exposure from Amalgam Restorations in Dental Patients', *Br. Dental J.*, **182**(9), 333–338 (1997).
24. L.C. Alderman, J.J. Bergin, 'Hydrogen Selenide Poisoning: an Illustrative Case with Review of the Literature', *Arch. Environ. Health.*, **41**(6), 354–358 (1986).
25. J.J. Rumessen, 'Hydrogen and Methane Breath Tests for Evaluation of Resistant Carbohydrates', *Eur. J. Clin. Nutrit.*, **46** (Suppl. 2), S77–S90 (1992).
26. T.H. Florin, 'Alkyl Halides, Super Hydrogen Production and the Pathogenesis of *Pneumatosis cystoides* coli', *Gut*, **41**(6), 778–784 (1997).
27. T.D. Bolin, J.R. Genge, V.M. Duncombe, Soe-Aung, Myo-Khin, 'Patterns of Methane Production in a Burmese (Myanmar) Population', *J. Gastroenterol. Hepatol.*, **11**(1), 71–76 (1996).
28. K.R. Wren, 'Ventilation Monitoring During Monitored Anesthesia Care: a Review', *Aana J.*, **62**(6), 521–526 (1994).
29. S.G. Saisch, 'An Introduction to Capnography', *Biofeedback and Self Regulation*, **19**(2), 115–134 (1994).
30. A. Favier, 'Oxidative Stress: Value of its Demonstration in Medical Biology and Problems Raised by the Choice of Marker', *Ann. Biol. Clinique.*, **55**(1), 9–16 (1997).
31. C.M. Kneepkens, G. Lepage, C.C. Roy, 'The Potential of the Hydrocarbon Breath Test as a Measure of Lipid Peroxidation', *Free Radical Biol. Med.*, **17**(2), 127–160 (1994) and **17**(6), 6091 (1994).
32. A.F. Goddard, R.P. Logan, 'Review Article: Urea Breath Tests for Detecting *Helicobacter pylori*', *Aliment. Pharmacol. Ther.*, **11**(4), 641–649 (1997).
33. G.R. Gibson, E.R. Beatty, X. Wang, J.H. Cummings, 'Selective Stimulation of Bifidobacteria in the Human Colon by Oligofructose and Inulin', *Gastroenterology*, **108**(4), 975–982 (1995).
34. J.A. Young, D.I. Cook, E.W. van Lennap, M. Roberts, in *Physiology of the Digestive Tract*, ed. L.R. Johnston, Raven Press, New York, 773–815, 1987.
35. F. Mion, P.E. Queneau, M. Rousseau, J.L. Brazier, P. Paliard, Y. Minaire, 'Aminopyrine Breath Test: Development of a 13C-Breath Test for Quantitative Assessment of Liver Function in Humans', *Hepato-Gastroenterology*, **42**(6), 931–938 (1995).
36. U. Fuhr, K.L. Rost, 'Simple and Reliable CYP1A2 Phenotyping by the Paraxanthine/Caffeine Ratio in Plasma and in Saliva', *Pharmacogenetics*, **4**(3), 109–116 (1994).
37. A.C. Parker, P. Pritchard, T. Preston, I. Choonara, 'Induction of CYP1A2 Acitivity by Carbamazepine in Children Using the Caffeine Breath Test', *Br. J. Clin. Pharmacol.*, **45**(2), 176–178 (1998).
38. B.D. Maes, M.I. Hiele, B.J. Geypens, P.J. Rutgeerts, Y.F. Ghoos, G. Vantrappen, 'Pharmacological Modulation of Gastric Emptying Rate of Solids as Measured by the Carbon Labelled Octanoic Acid Breath Test Influence of Erythromycin and Propantheline', *Gut*, **35**(3), 333–337 (1994) and **35**(6), 8661 (1994).
39. M.G. Choi, M. Camilleri, D.D. Burton, A.R. Zinsmeister, L.A. Forstrom, K.S. Nair, 'Reproducibility and Simplification of the 13C-Octanoic Acid Breath Test for Gastric Emptying of Solids', *Am. J. Gastroenterology*, **93**(1), 92–98 (1998).
40. A.E. M'Koma, K. Lindquist, L. Liljeqvist, 'Biochemical Laboratory Data in Patients Before and After Restorative Proctocolectomy. A Study on 83 Patients with a Follow-up of 36 months', *Ann. Chirurgie*, **48**(6), 525–534 (1994).
41. G.R. Swart, E.A. Baartman, J.L. Wattimena, T. Rietveld, S.E. Overbeek, J.W. van den Berg, 'Evaluation Studies of the 13C-Mixed Triglyceride Breath Test in Healthy Controls and Adult Cystic Fibrosis Patients with Exocrine Pancreatic Insufficiency', *Digestion*, **58**(5), 415–420 (1997).
42. S.J. Lewis, G. Young, M. Mann, S. Franco, S.J. O'Keefe, 'Improvement in Specificity of [14C]d-Xylose Breath Test for Bacterial Overgrowth', *Digestive Dis. Sci.*, **42**(8), 1587–1592 (1997).
43. E. Tanaka, D.D. Breimer, 'In-Vivo Function Tests of Hepatic Drug-oxidizing Capacity in Patients with Liver Disease', *J. Clin. Pharm. Therapeut.*, **22**(4), 237–249 (1997).
44. D. Rating, C.D. Langhans, 'Breath Tests: Concepts, Applications and Limitations', *Eur. J. Pediatrics*, **156**(Suppl. 1), S18–S23 (1997).
45. G.T. Clark, S. Nachnani, D.V. Messadi, 'Detecting and Treating Oral and Nonoral Malodors', *J. Calif. Dental Assoc.*, **25**(2), 133–144 (1997).
46. D.V. Messadi, 'Oral and Nonoral Sources of Halitosis', *J. Calif. Dental Assoc.*, **25**(2), 127–131 (1997).
47. S. Goldberg, H. Cardash, H. Browning III, H. Sahly, M. Rosenberg, 'Isolation of *Enterobacteriaceae* from the Mouth and Potential Association with Malodor', *J. Dental Res.*, **76**(11), 1770–1775 (1997).
48. S.M. Waler, 'The Effect of Zinc-containing Chewing Gum on Volatile Sulfur-containing Compounds in the Oral Cavity', *Anata Odont. Scand.*, **55**(3), 198–200 (1997).
49. A. Bosy, 'Oral Malodor: Philosophical and Practical Aspects', *J. Can. Dental Assoc.*, **63**(3), 196–201 (1997).
50. W.A. Groves, E.T. Zellers, 'Investigation of Organic Vapor Losses to Condensed Water Vapor in Tedlar Bags used for Exhaled Breath Sampling', *Am. Indust. Hyg. Assoc. J.*, **57**(3), 257–263 (1996).
51. J.B. Barsky, S.S. Que Hee, C.S. Clark, 'An Evaluation of the Response of Some Portable Direct-reading 10.2 eV and 11.8 eV Photoionization Detectors and a Flame Ionization Gas Chromatograph for Organic Vapors in High Humidity', *Am. Indust. Hyg. Assoc. J.*, **46**(1), 9–14 (1985).

52. I.N. Lee, S.S. Que Hee, C.S. Clark, 'Additivity of Detector Responses of a Portable Direct-reading 10.2 eV Photoionization Detector and a Flame Ionization Gas Chromatograph for Atmospheres of Multicomponent Organics: Use of FID/PID Ratios', *Am. Indust. Hyg. Assoc. J.*, **48**(5), 437–441 (1987).
53. S.S. Que Hee, 'Biological Monitoring', in *The Occupational Environment-its Evaluation and Control*, ed. S.R. DiNardi, AIHA Press, Fairfax, VA, 263–282, 1997.
54. W.C. Dillon, V. Hampl, P.J. Shultz, J.B. Rubins, S.L. Archer, 'Origins of Breath Nitric Oxide in Humans', *Chest*, **110**(4), 930–938 (1996).
55. F.H. Wong, C.Y. Yeung, K.W. Fung, A.Y. Tam, 'Breath Hydrogen (H_2) Analysis in Southern Chinese Children and Infants by Gas Chromatography and a Novel Automatic Sampling System', *Sing. Med. J.*, **37**(1), 72–81 (1996).
56. J.D. Pleil, A.B. Lindstrom, 'Collection of a Single Alveolar Exhaled Breath for a Volatile Organic Compounds Analysis', *Am. J. Indust. Med.*, **28**(1), 109–121 (1995).
57. A.B. Lindstrom, J.D. Pleil, 'Alveolar Breath Sampling and Analysis to Assess Exposures to Methyl Tertiary Butyl Ether (MTBE) During Motor Vehicle Refueling', *J. Air Waste Manag. Assoc.*, **46**(7), 676–682 (1996).
58. J.D. Pleil, A.B. Lindstrom, 'Exhaled Human Breath Measurement Method for Assessing Exposure to Halogenated Volatile Organic Compounds', *Clin. Chem.*, **43**(5), 723–730 (1997).
59. M. Phillips, 'Method for the Collection and Assay of Volatile Organic Compounds in Breath', *Anal. Biochem.*, **247**(2), 272–278 (1997).
60. C. Grote, J. Pawliszyn, 'Solid-phase Microextraction for the Analysis of Human Breath', *Anal. Chem.*, **69**(4), 587–596 (1997).
61. J.H. Raymer, K.W. Thomas, E.D. Pellizzari, D.A. Whitaker, S.D. Cooper, T. Limero, J.T. James, 'A Breath Sampling Device for Measuring Human Exposure to Volatile Organic Compounds in Microgravity', *Av. Space Environ. Med.*, **65**(4), 353–360 (1994).
62. G.N. Ljungkvist, R.G. Nordlinder, 'A Field Method for Sampling Benzene in End-exhaled Air', *Am. Ind. Hyg. Asso. J.*, **56**(7), 693–697 (1995).
63. Y. Lin, S.R. Dueker, A.D. Jones, S.E. Ebeler, A.J. Clifford, 'Protocol for Collection and HPLC Analysis of Volatile Carbonyl Compounds in Breath', *Clin. Chem.*, **41**(7), 1028–1032 (1995).
64. V.B. Stein, R.S. Narang, L. Wilson, K.M. Aldous, 'A Simple, Reliable Method for the Determination of Chlorinated Volatile Organics in Human Breath and Air Using Glass Sampling Tubes', *J. Anal. Toxicol.*, **20**(3), 145–150 (1996).
65. A. LeBlanc, B. Levesque, S. Allaire, 'Rapid, Sensitive, and Noninvasive Sampling Technique for Determination of Chloroform in Breath', *J. Anal. Toxicol.*, **19**(1), 56–57 (1995).
66. D. Dyne, J. Cocker, H.K. Wilson, 'A Novel Device for Capturing Breath Samples for Solvent Analysis', *Sci. Total Environ.*, **199**(1–2), 83–89 (1997).
67. G. Becher, K. Winsel, E. Beck, G. Neubauer, E. Stresemann, 'Breath Condensate as a Method of Noninvasive Assessment of Inflammation Mediators from the Lower Airways', *Pneumologie*, **51**(Suppl. 2), 456–459 (1997).
68. K.W. Terry, G.S. Hewson, 'Thorium Lung Burdens of Mineral Sands Workers', *Health Phys.*, **69**(2), 233–242 (1995).
69. R.D. Stewart, 'The Use of Breath Analysis in Clinical Toxicology', in *Essays in Toxicology*, Academic Press, New York, Chapter 5, 1974.
70. K. Jones, D. Dyne, J. Cocker, H.K. Wilson, 'A Biological Monitoring Study of 1-Methoxy-2-Propanol: Analytical Method Development and a Human Volunteer Study', *Sci. Total Environ.*, **199**(1–2), 23–30 (1997).
71. P. Kostrzewski, A. Wiaderna-Brycht, B. Czerski, 'Biological Monitoring of Experimental Human Exposure to Trimethylbenzene', *Sci. Total Environ.*, **199**(1–2), 73–81 (1997).
72. M. Phillips, M. Sabas, J. Greenberg, 'Increased Pentane and Carbon Disulfide in the Breath of Patients with Schizophrenia', *J. Clin. Pathol.*, **46**(9), 861–864 (1993) and **47**(9), 870 (1994).
73. N. Ong, C.Y. Shi, S.E. Chia, S.C. Chua, H.Y. Ong, B.L. Lee, T.P. Ng, K. Teramoto, 'Biological Monitoring of Exposure to Low Concentrations of Styrene', *Am. J. Indust. Med.*, **25**(5), 719–730 (1994).
74. P. Spanel, P. Rolfe, B. Rajan, D. Smith, 'The Selected Ion Flow Tube (SIFT) – A Novel Technique for Biological Monitoring', *Ann. Occup. Hyg.*, **40**(6), 615–626 (1996).
75. P. Spanel, D. Smith, 'Selected Ion Flow Tube: a Technique for Quantitative Trace Gas Analysis of Air and Breath', *Med. Biol. Eng. Comp.*, **34**(6), 409–419 (1996).
76. D. Smith, P. Spanel, 'The Novel Selected-ion Flow Tube Approach to Trace Gas Analysis of Air and Breath', *Rapid Commun. Mass Spectrom.*, **10**(10), 1183–1198 (1996).
77. R.F. Service, 'Breathalyzer Device Sniffs for Disease', *Science*, **281**(Sept 4), 1431 (1998).
78. D. Smith, P. Spanel, J.M. Thompson, B. Rajan, J. Cocker, P. Rolfe, 'The Selected Ion Flow Tube Method for Workplace Analyses of Trace Gases in Air and Breath: its Scope, Validation, and Applications', *Appl. Occup. Environ. Hyg.*, **13**(12), 817–823 (1998).
79. Inficon, *Hapsite Portable GC/MS*, Publication BR43A75K, East Syracuse, NY, 1997.
80. A. Franzblau, S. Batterman, J.B. D'Arcy, N.E. Sargent, K.B. Gross, R.M. Schreck, 'Breath Monitoring of Inhalation and Dermal Methanol Exposure', *Appl. Occup. Environ. Hyg.*, **10**(10), 833–839 (1995).
81. C.A. Byrnes, S. Dinarevic, C.A. Busst, E.A. Shinebourne, A. Bush, 'Effect of Measurement Conditions on Measured Levels of Peak Exhaled Nitric Oxide', *Thorax*, **52**(8), 697–701 (1997).

82. A.W. Jones, 'Interfering Substances Identified in the Breath of Drinking Drivers with Intoxilyzer 5000S', *J. Anal. Toxicol.*, **20**(7), 522–527 (1996).
83. A.W. Jones, L. Andersson, 'Variability of the Blood/Breath Alcohol Ratio in Drinking Drivers', *J. Forensic Sci.*, **41**(6), 916–921 (1996).
84. H. Gehring, W. Nahm, J. Baerwald, E. Konecny, P. Schmucker, 'Breath Alcohol Analyzers with Electrochemical Sensors – Accuracy of Measurements in Lung Model Ventilation', *Biomed. Technik*, **41**(3), 54–59 (1996).
85. J.P. Caldwell, N.D. Kim, 'The Response of the Intoxilyzer 5000 to Five Potential Interfering Substances', *J. Forensic Sci.*, **42**(6), 1080–1087 (1997).
86. H.J. Vreman, L.M. Baxter, R.T. Stone, D.K. Stevenson, 'Evaluation of a Fully Automated End-tidal Carbon Monoxide Instrument for Breath Analysis', *Clin. Chem.*, **42**(1), 50–56 (1996).
87. Drager, *Drager–Tube Handbook*, 11th edition, Drager Sicherheitstechnik GmbH, Lubeck, Germany, 1998.
88. K. Schlaht, 'Analysis of Aviator's Breathing Oxygen by Gas Chromatography', *LC/GC*, **16**(9), 854–859 (1998).

Particle-induced γ-Ray Emission

P.H.A. Mutsaers and M.J.A. de Voigt
Cyclotron Laboratory, Department of Applied Physics, Eindhoven University of Technology, and Center of Plasma and Radiation Physics, Den Dolech 2, P.O.Box 513, 5600 MB, The Netherlands

1 Introduction 1
2 Basic Nuclear Physics 2
2.1 Principles of Nuclear Reactions 2
2.2 γ-Ray Emission 5
2.3 Principles of Depth Profiling 7
3 Particle-induced γ-Ray Emission 7
3.1 Basic Principles of Particle-induced γ-Ray Emission 7
3.2 Possible Reactions and Their γ-Ray Energies 8
3.3 Sensitivity 8
3.4 Background 12
3.5 Experimental Set-up and Procedures 12
3.6 Depth Profiling 14
4 Application of the Particle-induced γ-Ray Emission Technique 15
5 Conclusions 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Particle-induced γ-ray emission (PIGE) is commonly applied for quantitative analysis of materials. γ-Rays, characteristic of the elements, are emitted after nuclear reactions. These reactions are induced by particle beams (e.g. protons) from accelerators. The reaction conditions and probabilities are explained, as well as the γ-ray properties and probabilities of emission. Knowledge of those parameters enables concentrations of elements to be deduced from the γ-ray spectra, which under certain conditions are also depth dependent. Therefore the samples must be small and flat. The experimental set-up and procedures are given together with a table with the most intense proton reactions and the energies of the emitted γ-rays. Various examples are given to illustrate the method in different applications.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

PIGE or prompt radiation analysis (PRA) is an analytical technique making use of γ-rays that are emitted during or immediately after bombardment of a sample with an ion beam. Here, discussion is restricted to prompt γ-ray emission. These are γ-rays that are detected during irradiation, in contrast to activation techniques where radiation is detected after irradiation.[1] The technique has been used since the 1960s for analytical purposes. Sippel and Glover[2] showed that PIGE with protons could be used for the determination of Mg and F in geological samples. Pierce et al.[3] showed that deuterons could be used to determine C and Pierce et al.[4] used 4 MeV protons to determine the amount of Si in several kinds of steel. At that time, the detection of γ-rays was mostly performed with NaI(Tl) detectors which have a relatively poor energy resolution, thereby limiting the analytical technique to the determination of one or two elements simultaneously. The development of Ge(Li) detectors with a much better energy resolution around 1970 made it possible to determine a number of elements in the sample simultaneously. Since then, studies have been carried out[5–9] with light as well as heavy projectiles that showed that PIGE with a Ge(Li) detector is a powerful and sensitive method of multielemental analysis. In general, protons and α-particles are used as projectiles, but measurements have also been done with d, t, ^{3}He, Li, B, N, F and Cl projectiles.

Besides the determination of concentrations in (thick) samples, PIGE can also be used to determine depth profiles by using narrow resonances in a cross-section of nuclear reactions. In general, PIGE is a technique that is less often used than other ion-beam analysis techniques like particle-induced X-ray emission (PIXE) or Rutherford backscattering spectroscopy (RBS). In PIXE or RBS spectra, elements appear in a relatively simple and systematic way, making interpretation and analysis of the spectra straightforward. In PIGE, the situation is more complex since even different isotopes of the same element result in completely different γ-spectra. Knowledge about the energy of resonances, the energy of the γ-rays, and interfering reactions in the sample must be available for each nuclide of interest. Interpretation and analysis of PIGE spectra can thus be much more complex than for PIXE or RBS spectra. This means that PIXE and RBS can be seen as general techniques to determine the composition of an unknown sample, while PIGE is used when the composition of the sample is more or less known and a γ-spectrum can be used to quantify the concentration or determine a concentration profile. PIXE (see **PIXE (Particle-induced X-ray Emission)**) is a technique that is generally used to determine elements

with $Z > 12$. PIGE is often used as a complementary technique to PIXE to determine light elements.

Although PIGE is regarded as a less general and versatile technique compared to PIXE, PIGE can have strong advantages. This is particularly the case for light elements ($Z < 11$) that cannot be analyzed or are hard to analyze with PIXE. Moreover, PIGE produces totally different spectra for elements of neighboring Z, removing peak-overlap problems between neighboring elements in the spectra. PIGE is only sensitive to specific isotopes, limiting the general applicability. It is, however, a powerful tool when isotopic analysis is required. Also, compared to similar PIXE spectra, minimal pile-up problems are present in PIGE spectra when there is no significant background radiation. Although PIGE is typically less sensitive than PIXE, the peaks in a PIGE spectrum are well separated and there is no need for absorption corrections for the γ-rays leaving the sample. This simplifies the experimental arrangement and the analysis of the spectra and can make the technique more attractive.

2 BASIC NUCLEAR PHYSICS

2.1 Principles of Nuclear Reactions

PIGE is based on nuclear reactions induced by accelerated beams that hit the sample of interest. After the reaction, nuclei in the sample remain in the excited states and decay by emitting γ-rays. The energies of these γ-rays are characteristic of the elements (isotopes). The two measured quantities thus yield the elemental identification and concentration, respectively. The technique is thus quantitative, or semi-quantitative when standards are used.

For a nuclear reaction to occur between a beam particle and a sample nucleus it is necessary that the beam energy exceeds the Coulomb barrier. The barrier energy (in MeV) can be estimated from Equations (1) and (2):

$$E_c = 1.44 \frac{Z_1 Z_2}{d} \tag{1}$$

$$\text{with} \qquad d = 1.25 \;\; A_1^{1/3} + A_2^{1/3} \tag{2}$$

where Z_1 and Z_2 are the proton numbers of the beam particles and the sample nuclei, respectively. The quantity d in fm (1 fm $= 10^{-15}$ m) is approximately the sum of the two respective radii, with A_1 and A_2 the respective mass numbers. The approximation is based on the notion that nuclear forces have very short ranges ($\sim$1 fm). Thus the two colliding nuclei must almost touch each other. If the beam energy is smaller than E_c then the beam particle will be deflected by the nucleus. γ-Rays can still be produced in this case by Coulomb excitation of the nucleus. This, however, is not very likely for light ion beams and will not be discussed further because light-ion induced reactions will be treated as applying most frequently. These reactions encompass inelastic scattering, capture (or compound nuclear) reactions and particle transfer. For protons they are denoted as: (p,p$'\gamma$), (p,γ) and, for instance, (p,$\alpha\gamma$), respectively.

At low energy, $< E_c$, only atomic excitations occur, giving rise to the emission of X-rays. This process is exploited in the PIXE method, as discussed in **PIXE (Particle-induced X-ray Emission)**.

Equation (1) results in $E_c = 1.2$ MeV for protons on ^{7}Li, 4 MeV on ^{28}Si, 5.2 MeV on ^{40}Ca and 13.6 MeV on ^{208}Pb. For α-particles on ^{28}Si, a beam energy of at least 7 MeV is required. It is concluded that a proton beam energy of at least 1 MeV is required and single-ended Van de Graaff accelerators, tandem Van de Graaff accelerators and cyclotrons are used for PIGE experiments.

The most commonly applied nuclear process in PIGE is the compound nuclear reaction, schematically written as Equation (3)

$$\mathrm{a} + \mathrm{A} \longrightarrow \mathrm{C}^* \longrightarrow \mathrm{B} + \mathrm{b}, \tag{3}$$

with a representing the beam particle, A the target nucleus, C* the compound nucleus (CN) in excited state that may decay by γ-rays or form a final product nucleus B and particle b. The product nucleus B, when formed in an excited state, B*, decays by γ-rays. The latter γ-rays are called reaction γ-rays, while those of C* are γ-rays of the CN. Those γ-rays are characteristic of the sample nucleus A (see Figure 1), in particular from the well-known discrete low-lying states.

The CN model is based on the assumption that the probability of decay into any specific set of final products is independent of the means of formation of the CN. The decay of the CN level only depends on its energy. If the excitation energy is above the particle binding energy then it surely will decay via the particle channel, because particle decay, owing to its strong interaction, is much more likely than γ-ray decay with its electromagnetic interaction.

The excitation energy, E_x, of the CN C* can be calculated from Equations (4) and (5):

$$E_x = Q + \frac{E_p m_A}{m_A + m_a} \tag{4}$$

$$\text{with} \qquad Q = [m_C - (m_A + m_a)]c^2 \tag{5}$$

with Q the reaction Q value in MeV, when the masses are expressed in atomic mass units, u, and the speed of light squared as $c^2 = 931.5\ \mathrm{MeV\,u^{-1}}$. In exothermic reactions mass is converted into energy and Q is positive, while

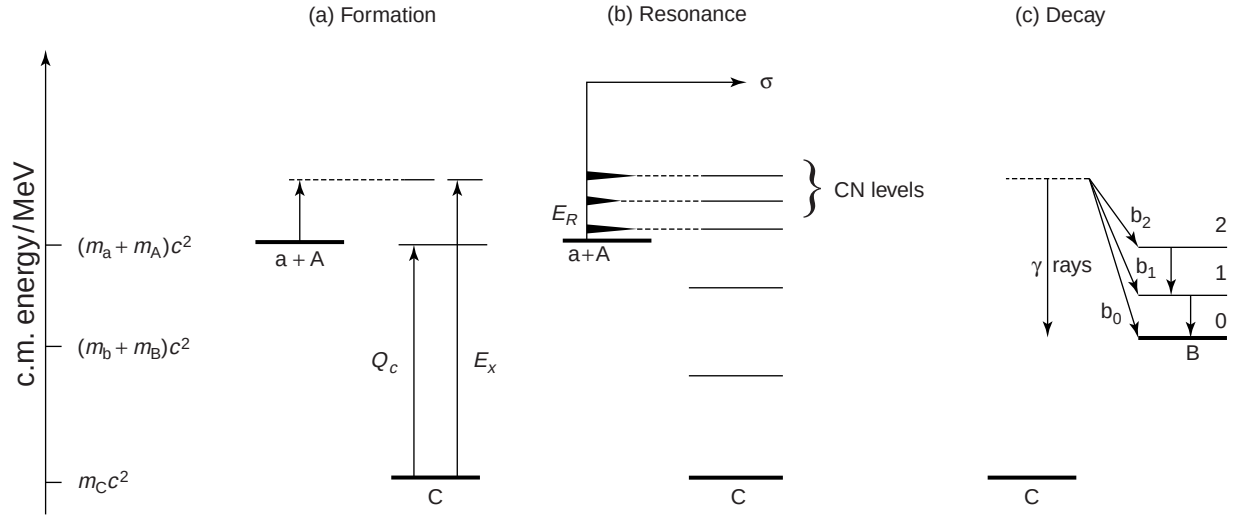

Figure 1 Formation (a) of CN C* in an excited state (CN level) (b) and its decay either by γ-rays or by particle emission b (c) to form the reaction product B* in excited states or in the ground state B. In (b) resonance levels are indicated with their schematic cross-sections σ. On the vertical axis the center-of-mass (c.m.) energy is indicated. (The figure is adapted from Deconninck.[1])

in endothermic reactions Q is negative. In Equation (4) the laboratory beam energy E_p is corrected by the mass factor of the two initial particles to obtain the energy in the center-of-mass system as an internal energy. The external energy or kinetic energy, E_{kin}, of the center-of-mass itself is written as Equation (6)

$$E_{kin} = \frac{E_p m_a}{m_A + m_a} \tag{6}$$

This energy is completely transferred to the recoiling CN as kinetic energy. The recoil energy may be of importance if the nuclear states involved in the decay of the CN have lifetimes shorter than the slowing down time of the recoil in the sample. In that case the γ-rays are emitted in flight and may show a shift and broadening in energy (see section 2.2).

Another important factor in PIGE analysis is the yield of γ-rays, which is directly connected to the reaction and emission probabilities. Here the reaction probability will be treated, while γ-ray emission is discussed in section 2.2.

The reaction probability is commonly expressed by the reaction cross-section σ, which represents a fictive surface at the nucleus through which the initial particle has to penetrate to induce a particular reaction. In many cases the cross-section exhibits a resonance-like behavior as a function of beam energy (Breit–Wigner shape), such as in proton capture or (p,γ) reactions at relatively low energy. The width Γ (in eV) of the resonance is a measure of the probability of forming that particular CN level and of decaying via γ-ray emission. The width Γ is related to the lifetime, τ, of the CN level via Equation (7), the uncertainty relation of Heisenberg:

$$\Gamma\tau = \frac{h}{2\pi} \tag{7}$$

To distinguish the formation of the CN level from the decay probability, Γ_p is introduced as the formation width (or probability) and Γ_γ as the decay width. The reaction cross-section can then be expressed as a function of beam energy in Equation (8)

$$\sigma_{p,\gamma}(E) = \pi\lambda\!\!\!^{-2} g \frac{\Gamma_p \Gamma_\gamma}{(E - E_R) + \Gamma^2/4} \tag{8}$$

Here E_R is the resonance energy at the maximum of the cross-section, $\lambda\!\!\!^{-}$ is the reduced wavelength of the initial particle and g is a statistical factor, determined by the spins, I, j and J of the nuclear particles A, a, and C*, respectively, Equation (9):

$$g = \frac{2J + 1}{(2j + 1)(2I + 1)} \tag{9}$$

For details see Deconninck.[1] If the resolution of the beam is larger than the total width Γ, the total area I_m under the resonance is obtained by integrating the cross-section over a broad energy range around the resonance, Equation (10):

$$I_m = 2\pi\lambda\!\!\!^{-2} g \frac{\Gamma_p \Gamma_\gamma}{\Gamma} \tag{10}$$

A Breit–Wigner resonance and its integral are shown in Figure 2.

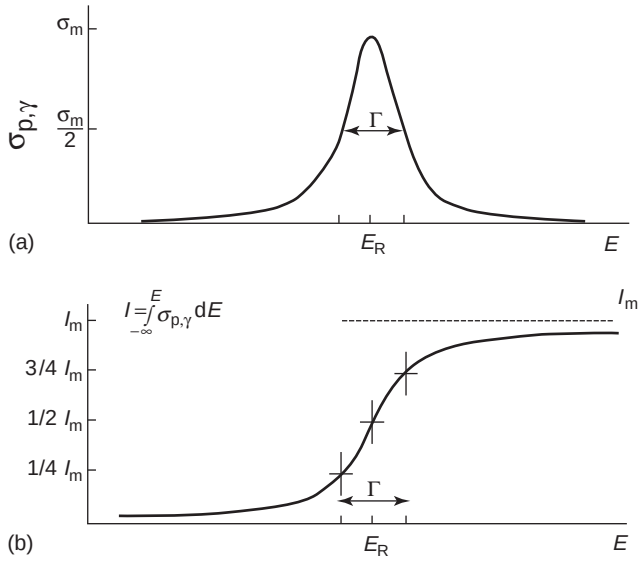

Figure 2 Breit–Wigner resonance shape curve (a) and its integral curve (b) are illustrated along with some of the relevant parameters discussed in the text. The maximum cross-section at the resonance energy E_R is indicated by σ_m. (The figure is taken from Deconninck.[1])

In the case where Γ is small (keV) the factor $(2J+1)\Gamma_P\Gamma_\gamma/\Gamma$ is proportional to the integrated cross-section. Thus the factor is proportional to the sensitivity of the resonance reaction. The total width Γ is contained in several nuclear data tables for many resonance levels in nuclei.[9–11] From this it can be concluded that the value of the cross-section is not only important for sensitivity, but also for knowledge of resonance structures (Equation 8) and, for practical applications, for the beam energy spread.

There are many other possible nuclear reactions yielding γ-rays that are resonant or nonresonant. Examples often used for elemental analysis are the inelastic scattering (p,p′γ), and further (p,$\alpha\gamma$) and (p,nγ) with protons. With other beam particles, commonly used reactions are (d,pγ), (d,nγ) and (α,nγ). At relatively low beam energies, in the few megaelectron volts region, the (p,p′γ) and the (p,$\alpha\gamma$) reactions also show resonant structures. Therefore Equation (8) can be used to find the cross-section by replacing Γ_γ for the exit channel by Γ_p and Γ_α, respectively. In the cases where the cross-section does not show resonant behavior, but varies only slowly with beam energy, the cross-section can be taken to be constant or averaged over the relevant energy region.

For nonresonant reactions, Equation (11) can be used to calculate the γ-ray yield $Y(E_0)$ at an incident energy E_0 of particle x on a target containing element A with atomic mass A_A (in g mol^{-1}), concentration c_A and with a cross-section $\sigma_{x,\gamma}(E)$, according to Equation (8), to produce particular γ rays:

$$Y(E_0) = N_p c_A A_A^{-1} N_{av} e(E_\gamma) \frac{\Omega}{4\pi} \int_{E_0}^{E_f} \frac{\sigma_{x,\gamma}(E)}{\varepsilon(E)}\, dE \qquad (11)$$

where N_p is the number of particles x hitting the target, which can be deduced from the measured beam current and time. N_{av} is the Avogadro number, $e(E_\gamma)$ is the efficiency of the detector for the particular γ-ray and Ω is the solid angle of the detector (in steradians). The cross-section $\sigma_{x,\gamma}(E)$ (cm^2, note 1 barn $= 10^{-24}$ cm^2) and the stopping cross-section $\varepsilon(E)$ (MeV g^{-1} cm^2) are integrated over the energy from E_0 to E_f, the energy of the particle x when leaving the target at the back side. Equation (11) also represents the thick-target yield, by putting $E_f = 0$, corresponding to the incident particle coming to a complete stop in the sample. The thick-target yield in literature is often expressed in units of $(\mu C\,sr)^{-1}$, that is in counts per quantity of electric charge on the target and per unit of solid angle. This can be derived easily from Equation (11) by dividing out N_p (expressed as charge) and Ω. The cross-sections and the stopping cross-sections of particle x slowed down in the sample have been tabulated in various publications for many particles, energies and stopping materials (see e.g. Nuclear Data Tables[12] for reaction cross-sections and Ziegler et al.[13] for stopping powers). If they are not tabulated then the known values must be interpolated or, in the worse case, measured. In Equation (11) the attenuation of γ-rays in the target, in possible absorbers and in the detector window is neglected, but it must be taken into account when γ-ray energies are low.

For detailed formulas in cases of sharp and broad resonances the reader is referred to Deconninck.[1] To circumvent the problems involved in the use of absolute cross-sections and stopping powers, internal or external standards are often used to compare with the measurement on the real sample. In the case where internal standard elements are added, their concentrations and cross-sections must be known precisely. In the case of an external standard, a sample is prepared with exactly the same elements as the sample of interest, at known concentrations. In the first case only one measurement is needed, but integrations for the different elements (Equation 11) are still necessary. However, beam current and detector solid angle are canceled. In the second case the integrations are canceled, but beam current (and solid angle, if different) need to be measured.

For thin samples the cross-section can be considered constant. It is convenient to add a standard element with mass number A_s and with known concentration c_s to the sample. When comparing the γ-ray yields of element A and standard S, not only are the solid angles and number of beam particles canceled, but so too are the stopping powers because in both cases they are the same. The

unknown concentration c_A can easily be calculated from the measured γ-ray yields from Equation (12):

$$c_A = c_s \frac{Y_A \sigma_s e_s A_A}{Y_s \sigma_A e_A A_s} \tag{12}$$

2.2 γ-Ray Emission

γ-Ray emission is governed by the electromagnetic interaction and transitions are described by connecting the initial and final nuclear quantum states via the electromagnetic multipole operator. The electromagnetic field can be expanded in terms of multipoles, i.e. monopole, dipole, quadrupole, octupole, and so on. They can be of electric or magnetic character, with certain constraints. The multipole moments have the form of Equation (13):

$$M_l \propto r^l Y_l^m(\theta, \phi) \tag{13}$$

here l represents the order of the multipole moment and $Y_l^m(\theta, \phi)$ is the spherical harmonic function. The electromagnetic operator, which acts on the initial and final states involved and gives rise to the γ-ray transition, also has the basic form of Equation (13). For a detailed treatment of the electromagnetic features important for γ-ray emission the reader is referred to text books, for example by Krane[14] and Ejiri and de Voigt.[15] Here we will restrict ourselves to the more practical rules that determine the nature and intensity of γ-rays.

The electromagnetic transition probability $T(E(M), l)$, according to Equation (7) is given as Equation (14):

$$T(E, (M), l, if) = \frac{\Gamma(E, (M), l, if)}{\hbar} = \frac{\Gamma_\gamma}{\hbar} = \frac{1}{\tau} \tag{14}$$

The radiation width Γ_γ of Equation (7) is thus equivalent to the transition probability T, which depends on the electric (E) or magnetic (M) character, on the multipolarity l of the radiation and on the structure of the initial and final nuclear states (if). The multipolarity l of the transition corresponds to the order l of the multipole moment and of the multipole operator (Equation 13). The transition probability depends on external factors, i.e. the energy E_γ of the transition, the multipolarity l and the nuclear radius R, as well as on the internal factors (i.e. the nuclear structure of the initial and final states). The latter information is not always available and varies strongly from nucleus to nucleus and even within a nucleus from state to state. Therefore we will discuss the general consequences of the external factors and refer the reader to the various nuclear data compilations for the internal factors; i.e. for nuclei with mass 21–44 to the compilation of Endt[10] (and references quoted therein) and for nuclei below mass 21 to the compilation of Ajzenberg–Selove[11] (and references quoted therein).

The importance of the external and internal factors is that they also determine the value of Γ_γ and thereby the cross-section $\sigma_{x,\gamma}$, according to Equation (8) and thus the yield $Y(E_0)$, according to Equation (11). A simplified indication for the external factor T_{ex} can be obtained from Equations (15) and (16):

$$T_{ex}(E) \propto \left(\frac{E_\gamma}{\hbar c}\right)^{2l+1} R^{2l} \tag{15}$$

and

$$T_{ex}(M) \propto \left(\frac{E_\gamma}{\hbar c}\right)^{2l+1} R^{2l-2} \tag{16}$$

Noting that $\hbar c = 197\,\text{MeV fm}$, that E_γ is of the order of 1 MeV and that the nuclear radius is about 3 fm for light nuclei, one concludes that the transition probability decreases rather strongly with increasing multipolarity of the radiation. For each unit of l the difference is about a factor of 10^4. The electric transitions are more probable than the magnetic ones of the same multipolarity by a factor of R^2, which is about a factor of 10 for light nuclei. Note that these are only rough estimates and that the internal factor may change the picture quite drastically. For instance in collectively rotating nuclei, in the strongly deformed regions, the electric quadrupole (E_2) transition is strongly enhanced. Nevertheless, it is probably safe to state that in general the sequence of probability for transitions to occur is E_1, M_1, E_2, M_2, E_3, M_3, E_4, M_4, and so on. The transition probability increases strongly with the transition energy, particularly for the higher multipolarities. This is illustrated in Figure 3, where absolute transition probabilities are given for one-particle transitions, i.e. transitions caused by one (valence) proton in the nucleus.

The energy of the γ-ray is simply the energy difference between the initial and final states, Equation (17):

$$E_\gamma = E_i - E_f, \tag{17}$$

here E_i and E_f represent the excitation energies of the initial and the final states, respectively. In this formula the recoil energy of the de-exciting nucleus is neglected. It can be calculated from the γ-ray energy and mass, M, of the nucleus as Equation (18):

$$E_{rec} = \frac{E_\gamma^2}{2Mc^2} \tag{18}$$

Noting that the numerator is of the order of megaelectron volts and the denominator of gigaelectron volts, E_{rec} is only about $10^{-3} E_\gamma$, thus mostly of the order of kiloelectron volts, which can be neglected.

The multipolarity of the γ-radiation is restricted by the spins (and parities) of the initial and final states I_i (π_i)

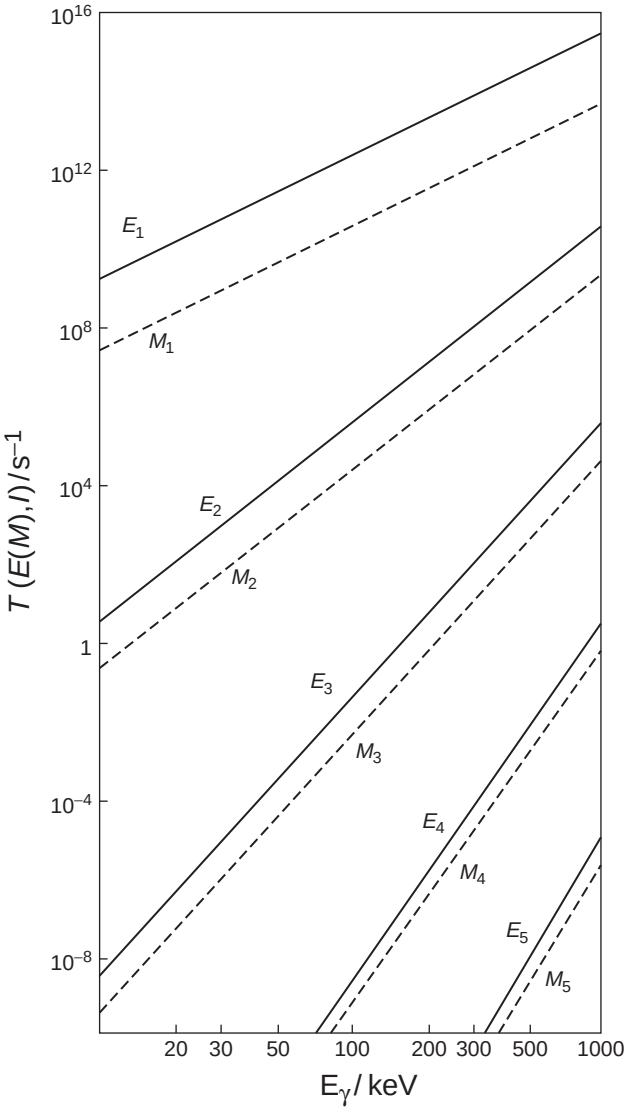

Figure 3 One-particle transition probabilities as functions of the γ-ray energies for various multipole radiations.

and I_f (π_f), respectively. The multipole radiation of order l (Equation 13) carries l units ($\hbar$) of angular momentum per photon. Conservation of angular momentum and the notion that its couplings are vectorwise leads to the rule, Equation (19):

$$|I_i - I_f| \leq l \leq I_i + I_f \tag{19}$$

It thus appears that for certain spin combinations more than one multipolarity is allowed. In that case the lowest allowed multipolarity is the most probable one with mixing of the next higher one, for which l is one unit larger. The parities of the initial and final states, π_i and π_f, are also important for the electric or magnetic character. Conservation of parity, with π_l the parity of the emitted γ-ray with multipolarity (angular momentum) l, leads to the rule, Equation (20):

$$\pi_l = \pi_i \pi_f \tag{20}$$

The parity π_l of electric transitions is determined by Equation (21):

$$\pi_l(E) = (-1)^l \tag{21}$$

The parity π_l of magnetic transitions is determined by Equation (22):

$$\pi_l(M) = (-1)^{l+1} \tag{22}$$

The consequences of the latter two rules are that electric transitions with even multipolarity l and magnetic transitions with odd l have even parity; thus M_1, E_2, M_3, E_4, M_5, E_6, and so on have even parity. Consequently E_1, M_2, E_3, M_4, E_5, M_6, and so on have odd parity. With the rules shown in Equations (19)–(22) and from the known spins and parities of the initial and final states the character of the radiation can be determined and thereby the probability of the radiation and its intensity as far as these factors are concerned. The following examples will illustrate the simplicity of the rules:

Transition
$1^+ \rightarrow 0^+$ M_1 radiation only possible
$1^- \rightarrow 0^+$ E_1
$2^+ \rightarrow 0^+$ E_2
$3^+ \rightarrow 2^+$ $M_1, E_2, \ldots, M_5$. In practice only M_1, E_2 mixed.
$1^- \rightarrow 2^+$ E_1, M_2, E_3. In practice only E_1, M_2 mixed.

Note that the transition $0^+ \rightarrow 0^+$ cannot proceed via γ-radiation because the intrinsic spin of a photon is 1^-. In this particular case the transition is fully converted to conversion electrons. All other transitions are also for a minor fraction converted, but the conversion coefficients are rather small, particularly for low multipolarity radiation and high transition energies. For instance, at 300 keV for E_1 radiation the conversion is already below 1%, but for M_4 radiation, electrons and γ-rays have about the same probability. For further detail see, for example, Krane[14] and Ejiri and de Voigt.[15]

It should be realized that γ-rays generally are not emitted isotropically in space. Asymmetry can be as much as 50%. The shape of the angular distribution depends on the angular momentum transfer in the reaction and on the population of the initial state and the final nucleus, as well as on the character (multipolarity) of the radiation. Quadrupole radiation, for instance, has its maximum along the direction of the beam, while dipole radiation peaks in directions perpendicular to the beam direction. For absolute values, measured angular distributions must be used or can be calculated on the basis of known

or assumed angular momentum alignments (see e.g. Krane[14]).

Finally, the energy shift and broadening of the γ-ray lines in the spectrum, which are due to the Doppler effect, should be mentioned. When a γ-ray is emitted during the flight of the excited nucleus through the sample or in vacuum behind the sample, the energy measured will be Doppler shifted. This is the case when the lifetime of the nuclear excited state is smaller than the average slowing down time of the recoiling nucleus in the sample. This time depends on the initial velocity of the recoil and on the slowing down material, but generally has values in the range of 10^{-8}–10^{-14} s. The measured γ-ray energy is then given by Equation (23):

$$E_\gamma = E_{\gamma 0} \quad 1 + \frac{v}{c}\cos\theta \tag{23}$$

where $E_{\gamma 0}$ is the unshifted energy, θ the angle of observation with respect of the beam direction and v is the velocity of the recoiling nucleus. In the case of compound nuclear reactions the initial velocity can be calculated easily from Equation (6) as Equation (24):

$$\frac{v}{c} = 0.046\,35\frac{\sqrt{m_a E_p}}{m_A + m_a} \tag{24}$$

The masses are given in mass numbers and the energy in MeV, resulting in the ratio v/c. Almost all nuclear lifetimes of compound states, mostly highly excited states, are very short, of the order of 10^{-14}–10^{-15} s. The corresponding de-exciting γ-rays are thus fully Doppler shifted and broadened, depending on the angle of emission and on the detector solid angle. Low-lying excited states may have much longer lifetimes and the corresponding γ-rays will not be affected by the Doppler effect. The resulting sharp lines are generally more useful for elemental analysis.

2.3 Principles of Depth Profiling

Charged-particle induced resonance reactions offer the possibility of obtaining the depth in the sample from which the γ-radiation is emitted and thus determining the depth profile of the elements. The leading principle is based on the fact that the beam particles are slowed down in the sample and thus the beam energy has to be increased to match the value of the resonance. The amount of increase ΔE (keV) depends on the depth Δx (cm) and the two quantities are related via the stopping power S (keV cm^{-1}) by Equation (25):

$$\Delta x = \frac{\Delta E}{S} \tag{25}$$

In compilations (see e.g. Ziegler et al.[13]) the stopping cross-section ε is often given in units of keV atom^{-1} cm^2. Here ε is defined by Equation (26) as:

$$\varepsilon = \frac{\Delta E}{N\Delta x} \tag{26}$$

where N is the atomic density in the sample (at cm^{-3}). It should be noted that the depth resolution depends on the straggling in the sample, on the resonance width and on the spread in beam energy. It is therefore important to select resonances that are as narrow as the beam spread (keV).

Thus, the depth Δx can be calculated from the difference between the resonance and the beam energy, the stopping cross-section and the atomic density. The atomic density can be expressed in terms of the mass density ρ (g cm^{-3}) as Equation (27):

$$N = \frac{\rho N_{av}}{M} \tag{27}$$

where N_{av} is the Avogadro number (6.022045×10^{23} mol^{-1}) and M is the molecular or atomic weight.

The stopping cross-sections are tabulated (for example in Ziegler et al.[13]) for elements. In the case of a compound a simple rule called Braggs rule may be applied, Equation (28):

$$\varepsilon = \sum_i c_i \varepsilon_i \tag{28}$$

where c_i are the mass concentrations or the atomic fractions in a molecule.

3 PARTICLE-INDUCED γ-RAY EMISSION

3.1 Basic Principles of Particle-induced γ-Ray Emission

In PIGE, samples are bombarded with a beam consisting of light particles (protons, deuterons, tritons, ^{3}He or ^{4}He (α)) or heavy ions. The energy of the incident particles is usually in the range of 1–4 MeV amu^{-1}. The beam enters a vacuum chamber, in which the sample is placed, through a set of diaphragms and the emitted γ-rays are detected by, for example, a hyperpure Ge or Ge(Li) detector. The pulses from the detector are analyzed by a multichannel analyzer (MCA). A more detailed description of the experimental set-up can be found in section 3.5. A typical PIGE energy spectrum is shown in Figure 4.

This type of energy spectrum consists of a number of peaks corresponding to the γ-rays related to the elements present in the sample. These characteristic peaks are superimposed on a background that originates from Compton scattering of γ-rays in the detector. The number of counts in a peak of a certain element corresponds

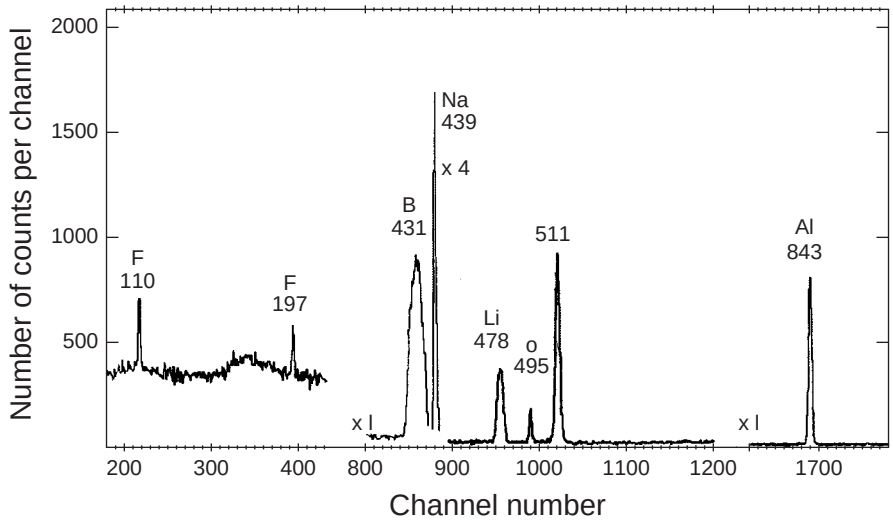

Figure 4 Example of a PIGE energy spectrum of a potassium feldspar (ST-2B) sample. The sample was measured with 1.8 MeV protons. Energies marked on peaks are in keV.[8]

to the amount of that element in the sample under investigation. All the parameters that determine the number of counts in a certain peak are known or can be measured. Using computer programs and databases, an absolute determination of the amounts of the elements in the sample is possible.

A complication in the spectra can be that the same γ-rays could result from two different reactions. For example, consider the reactions $^{26}Mg(p,\gamma)^{27}Al$ and the $^{27}Al(p,p'\gamma)^{27}Al$ which both give γ-rays with an energy of 843 keV and 1023 keV. Magnesium should then be identified by the $^{25}Mg(p,p'\gamma)^{25}Mg$ reaction which produces γ-rays of 586 keV. Aluminum can be identified by the $^{27}Al(p,\gamma)^{28}Si$ reaction with γ-rays of 1778 and 2836 keV.

Prompt γ-rays can originate from different types of nuclear reactions such as $(p,p'\gamma)$, $(p,\alpha\gamma)$, or (p,γ) for protons, and $(d,n\gamma)$, $(\alpha,\alpha'\gamma)$ and so on for other projectiles. When an element consists of more than one isotope, γ-rays can be produced in any of the isotopes and this will show up in the accompanying spectra as more lines belonging to one element but different isotopes. This leads to isotopic determination.

3.2 Possible Reactions and Their γ-Ray Energies

In a PIGE spectrum, the peaks are characteristic of different elements or even different isotopes of elements present in the sample under investigation. In Table 1, the most intense proton reactions together with the energy of the emitted γ-rays are presented for elements with $4 < Z < 20$. In Table 2, the most intense proton reactions for a number of heavy element isotopes are presented. The elemental contents are obtained from isotopic contents, taking the natural abundances into account.

3.3 Sensitivity

A general discussion about the sensitivity of PIGE for different kinds of sample is difficult to give since it depends on the content of the major elements (the matrix), the minor elements and the trace elements in the sample. There are several ways to define the sensitivity of an analytical method.

The first and commonly used way to define sensitivity is by the minimal detectable concentration or limit of detection (LOD). It is defined by Equation (29)

$$\mathrm{LOD} = k\sigma_b, \tag{29}$$

where σ_b is the standard deviation of the background under a (certain interval of a) peak of interest in a spectrum, and k is a factor reflecting the desired confidence level.[19] It can be interpreted more easily with the help of Figure 5.

In this figure, the energy interval ΔE is taken as the detector resolution and is equal to the full width at half-maximum (fwhm) of the peak. I_Z is the number of counts in the peak and I_b is the number of counts in the background under the peak. Both are the number of counts in the energy interval ΔE. The standard deviation σ_b is taken to be the standard deviation of the background counts I_b. The confidence level $k = 3$, gives a confidence level of 99.98% for normally distributed measurements. This means that a peak is considered to be detectable if, Equation (30):

$$I_Z \geq 3\sqrt{I_b} \tag{30}$$

Table 1 Most intense possible proton reactions for light element isotopes with $4 < Z < 20$, together with the energy of the emitted γ-rays[16,17]

Isotope	Reaction	E_γ	Isotope	Reaction	E_γ	Isotope	Reaction	E_γ
^{7}Li	^{7}Li(p,nγ)^{7}Be	429	^{25}Mg	^{25}Mg(p,p′γ)^{25}Mg	390	^{33}S	^{33}S(p,p′γ)^{33}S	841
	^{7}Li(p,p′γ)^{7}Li	478		,,	585		,,	1472
^{9}Be	^{9}Be(p,γ)^{10}B	415		,,	975	^{34}S	^{34}S(p,γ)^{35}Cl	1219
	,,	718		,,	990		^{34}S(p,p′γ)^{34}S	2127
	,,	1022		,,	1380	^{35}Cl	^{35}Cl(p,p′γ)^{35}Cl	1219
	,,	7477		,,	1612		,,	1763
	^{9}Be(p,αγ)^{6}Li	3562		,,	1965		^{35}Cl(p,γ)^{36}Ar	1970
^{10}B	^{10}B(p,αγ)^{7}Be	429		^{25}Mg(p,γ)^{26}Al	417		,,	2208
	^{10}B(p,p′γ)^{10}B	718	^{26}Mg	^{26}Mg(p,γ)^{27}Al	844	^{37}Cl	^{37}Cl(p,nγ)^{37}Ar	1410
^{11}B	^{11}B(p,p′γ)^{11}B	2125		,,	1014		,,	1611
^{12}C	^{12}C(p,γ)^{13}N	2366		^{26}Mg(p,p′γ)^{26}Mg	1809		^{37}Cl(p,p′γ)^{37}Cl	1727
	,,	3511	^{27}Al	^{27}Al(p,p′γ)^{27}Al	171		^{37}Cl(p,αγ)^{38}Ar	2127
^{13}C	^{13}C(p,γ)^{14}N	1635		,,	844		^{37}Cl(p,γ)^{38}Ar	670
	,,	2313		,,	1014		,,	1642
	,,	8062		,,	1720		,,	2168
	^{13}C(p,p′γ)^{13}C	3089		,,	2211	^{39}K	^{39}K(p,γ)^{40}Ca	755
^{14}N	^{14}N(p,p′γ)^{14}N	2313		,,	2734		^{39}K(p,nγ)^{41}Ca	2522
	^{14}N(p,γ)^{15}O	6793		,,	2981	^{41}K	^{41}K(p,p′γ)^{41}K	980
^{15}N	^{15}N(p,αγ)^{12}C	4439		,,	3004		,,	1294
^{16}O	^{16}O(p,γ)^{17}F	495		^{27}Al(p,γ)^{28}Si	1779		^{41}K(p,γ)^{42}Ca	313
^{17}O	^{17}O(p,p′γ)^{17}O	871		,,	2839		,,	899
^{18}O	^{18}O(p,p′γ)^{18}O	1982		^{27}Al(p,αγ)^{24}Mg	1369		,,	1525
^{19}F	^{19}F(p,p′γ)^{19}F	110	^{28}Si	^{28}Si(p,γ)^{29}P	1384		^{41}K(p,αγ)^{38}Ar	2168
	,,	197		,,	4343		^{41}K(p,nγ)^{41}Ca	1943
	,,	1236		^{28}Si(p,pγ)^{28}Si	1779		,,	2010
	,,	1349	^{29}Si	^{29}Si(p,γ)^{30}P	677		,,	2576
	,,	1357		,,	709		,,	2604
	,,	1459		^{29}Si(p,p′γ)^{29}Si	755	^{42}Ca	^{42}Ca(p,p′γ)^{42}Ca	1525
	^{19}F(p,αγ)^{16}O	6129		,,	1273	^{43}Ca	^{43}Ca(p,p′γ)^{43}Ca	373
	,,	6917		,,	2028	^{44}Ca	^{44}Ca(p,p′γ)^{44}Ca	1157
	,,	7114	^{30}Si	^{30}Si(p,γ)^{31}P	1266		^{44}Ca(p,γ)^{45}Sc	364
^{23}Na	^{23}Na(p,p′γ)^{23}Na	440		,,	2233		,,	531
	,,	1636		^{30}Si(p,p′γ)^{30}Si	2235		,,	543
	,,	1951	^{31}P	^{31}P(p,p′γ)^{31}P	1266		,,	720
	,,	2391		,,	2233	^{48}Ca	^{48}Ca(p,nγ)^{48}Sc	371
	^{23}Na(p,αγ)^{20}Ne	1634		^{31}P(p,γ)^{32}S	2230		,,	520
	^{23}Na(p,γ)^{24}Mg	1369		^{31}P(p,αγ)^{28}Si	1779		,,	779
	,,	4238	^{32}S	^{32}S(p,γ)^{33}Cl	811			
^{24}Mg	^{24}Mg(p,γ)^{25}Al	452		,,	2035			
	^{24}Mg(p,p′γ)^{24}Mg	1369		^{32}S(p,p′γ)^{32}S	2230			

Systematic studies concerning sensitivities have been performed by Deconninck and Demortier[5,6] with protons, Clark et al.[20] with protons and α-particles, Borderie et al.[9,21] with tritons and α-particles, Giles and Peisach[7] with α-particles and Borderie et al.[22] with heavy ions. Of about 70 elements studied, 40 can be determined with high sensitivity (<100 ppm). Trace element sensitivity (ppm range) can be obtained for the light elements Li, Be, B, N, O, F, Na and P.[23]

In Figure 6, it can be seen that protons and α-particles are the best choice for the determination of light elements. Heavy ions can be used for the determination of medium and heavy elements at a level around 50 ppm. Accurate determination of concentrations in samples depends on the LOD and on the presence of major and minor elements in the sample. As a rule of thumb, accurate determinations can be performed at concentration levels around 10 times the LOD.

The second way to define sensitivity is to express it as the minimum absolute amount that can be detected. The absolute amount can be calculated from the concentration of an element if the areal mass density (g/cm^2) of the sample is known. The minimal amount of light elements is in the range 10^{-10}–10^{-11} g. The lower limit in the absolute amount is obtained with the ion microprobe technique (see section 3.5). Here, a beam diameter below 1 μm can

Table 2 Most intense possible proton reactions for elements (isotopes) with $Z > 30$ together with the energy of the emitted γ-rays[18]

Isotope	Reaction	E_γ	Isotope	Reaction	E_γ	Isotope	Reaction	E_γ
^{69}Ga	^{69}Ga(p,γ)^{70}Ge	175	^{86}Sr	^{86}Sr(p,γ)^{87}Y	793	^{111}Cd	^{111}Cd(p,p$'\gamma$)^{111}Cd	342
	,,	668	^{87}Sr	^{87}Sr(p,γ)^{88}Y	232	^{112}Cd	^{112}Cd(p,p$'\gamma$)^{112}Cd	617
	,,	1040	^{88}Sr	^{88}Sr(p,γ)^{89}Y	909	^{113}Cd	^{113}Cd(p,p$'\gamma$)^{113}Cd	298
	,,	1708		,,	1313	^{114}Cd	^{114}Cd(p,p$'\gamma$)^{114}Cd	558
	^{69}Ga(p,p$'\gamma$)^{69}Ga	318		,,	1507	^{115}In	^{115}In(p,nγ)^{115}In	115
^{71}Ga	^{71}Ga(p,nγ)^{71}Ge	175		,,	1745		,,	497
	,,	327	^{89}Y	^{89}Sr(p,γ)^{90}Zr	420	^{121}Sb	^{121}Sb(p,p$'\gamma$)^{121}Sb	508
	,,	391		,,	562		,,	573
	,,	500		,,	891	^{123}Sb	^{123}Sb(p,p$'\gamma$)^{121}Sb	160
	^{71}Ga(p,p$'\gamma$)^{71}Ga	391		,,	2186		,,	382
^{70}Ge	^{70}Ge(p,γ)^{71}As	147		,,	2319		,,	542
^{72}Ge	^{72}Ge(p,γ)^{73}As	254	^{90}Zr	^{90}Zr(p,γ)^{91}Nb	1082	^{126}Te	^{126}Te(p,p$'\gamma$)^{126}Te	666
	,,	361			1208	^{127}I	^{127}I(p,p$'\gamma$)^{127}I	145
^{73}Ge	^{73}Ge(p,nγ)^{73}As	254	^{92}Zr	^{92}Zr(p,γ)^{93}Nb	656		,,	172
	,,	361		,,	744		,,	203
^{74}Ge	^{74}Ge(p,γ)^{75}As	199	^{94}Zr	^{94}Zr(p,p$'\gamma$)^{94}Zr	919		,,	375
	,,	265	^{93}Nb	^{93}Nb(p,γ)^{94}Mo	449		,,	418
	,,	280		,,	703		,,	629
	,,	596		,,	742	^{181}Ta	^{181}Ta(p,p$'\gamma$)^{181}Ta	136
	^{74}Ge(p,p$'\gamma$)^{74}Ge	596		,,	850		,,	165
^{76}Ge	^{76}Ge(p,p$'\gamma$)^{76}Ge	563		,,	871		,,	302
^{75}As	^{75}As(p,p$'\gamma$)^{75}As	199		^{93}Nb(p,p$'\gamma$)^{93}Nb	742	^{182}W	^{182}W(p,p$'\gamma$)^{182}W	100
	,,	280	^{95}Mo	^{95}Mo(p,p$'\gamma$)^{95}Mo	204	^{183}W	^{183}W(p,p$'\gamma$)^{183}W	292
	,,	573	^{96}Mo	^{96}Mo(p,p$'\gamma$)^{96}Mo	778	^{184}W	^{184}W(p,p$'\gamma$)^{184}W	111
	^{75}As(p,γ)^{76}Se	559	^{97}Mo	^{97}Mo(p,p$'\gamma$)^{97}Mo	481	^{186}W	^{186}W(p,p$'\gamma$)^{186}W	123
	^{75}As(p,nγ)^{75}Se	112	^{98}Mo	^{98}Mo(p,p$'\gamma$)^{98}Mo	787	^{194}Pt	^{194}Pt(p,p$'\gamma$)^{194}Pt	329
	,,	287	^{100}Mo	^{100}Mo(p,p$'\gamma$)^{100}Mo	535	^{195}Pt	^{195}Pt(p,p$'\gamma$)^{195}Pt	211
^{76}Se	^{76}Se(p,p$'\gamma$)^{76}Se	559	^{99}Ru	^{99}Ru(p,p$'\gamma$)^{99}Ru	90		,,	239
^{77}Se	^{77}Se(p,p$'\gamma$)^{77}Se	239	^{100}Ru	^{100}Ru(p,p$'\gamma$)^{100}Ru	540	^{196}Pt	^{196}Pt(p,p$'\gamma$)^{196}Pt	356
^{78}Se	^{78}Se(p,p$'\gamma$)^{78}Se	614	^{101}Ru	^{101}Ru(p,p$'\gamma$)^{101}Ru	127	^{198}Pt	^{198}Pt(p,p$'\gamma$)^{198}Pt	407
^{80}Se	^{80}Se(p,p$'\gamma$)^{80}Se	666	^{102}Ru	^{102}Ru(p,p$'\gamma$)^{102}Ru	475	^{197}Au	^{197}Au(p,p$'\gamma$)^{197}Au	191
	,,	276	^{104}Ru	^{104}Ru(p,p$'\gamma$)^{104}Ru	358		,,	269
^{79}Br	^{79}Br(p,p$'\gamma$)^{79}Br	217	^{103}Rh	^{103}Rh(p,p$'\gamma$)^{103}Rh	295		,,	279
	,,	306		,,	357		,,	548
	,,	523	^{102}Pd	^{102}Pd(p,p$'\gamma$)^{102}Pd	556	^{198}Hg	^{198}Hg(p,p$'\gamma$)^{198}Hg	412
	^{79}Br(p,γ)^{80}Kr	617	^{104}Pd	^{104}Pd(p,p$'\gamma$)^{104}Pd	556	^{199}Hg	^{199}Hg(p,p$'\gamma$)^{199}Hg	158
^{81}Br	^{81}Br(p,p$'\gamma$)^{81}Br	276	^{105}Pd	^{105}Pd(p,p$'\gamma$)^{105}Pd	280		,,	208
	^{81}Br(p,nγ)^{81}Kr	190		,,	442	^{200}Hg	^{200}Hg(p,p$'\gamma$)^{200}Hg	368
^{85}Rb	^{85}Rb(p,p$'\gamma$)^{85}Rb	130	^{108}Pd	^{108}Pd(p,p$'\gamma$)^{108}Pd	434	^{203}Tl	^{203}Tl(p,p$'\gamma$)^{203}Tl	279
	,,	151	^{110}Pd	^{110}Pd(p,p$'\gamma$)^{110}Pd	374	^{205}Tl	^{205}Tl(p,p$'\gamma$)^{205}Tl	204
	^{85}Rb(p,γ)^{86}Sr	1077	^{107}Ag	^{107}Ag(p,p$'\gamma$)^{107}Ag	325		,,	416
	,,	1854		,,	423			
	^{85}Rb(p,nγ)^{85}Sr	232	^{109}Ag	^{109}Ag(p,p$'\gamma$)^{109}Ag	311			
^{87}Rb	^{87}Rb(p,nγ)^{87}Sr	485		,,	415			
	,,	1229	^{110}Cd	^{110}Cd(p,p$'\gamma$)^{110}Cd	658			

be used,[24] making detection of an absolute amount of 10^{-15}–10^{-16} g possible, although measurement times can be very high if concentration distributions have to be determined.

The third way is to use the thick-target yield (Equation 11) as a way to compare different elements, projectiles, energies and so on. The advantage of this method is that PIGE is most often used for the analysis of thick samples. A thick sample means a sample in which the incident particle is completely stopped. Generally, in thin samples the sensitivity is poor compared to thick samples. There is a strong energy dependence of the excitation curve, and the sample weight and composition should be known accurately.[25]

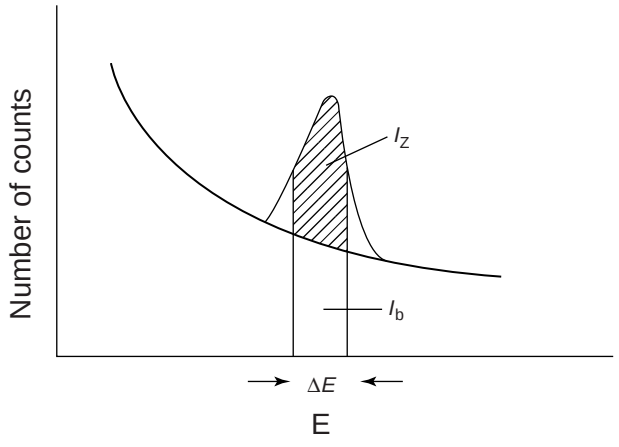

Figure 5 Limit of detection.

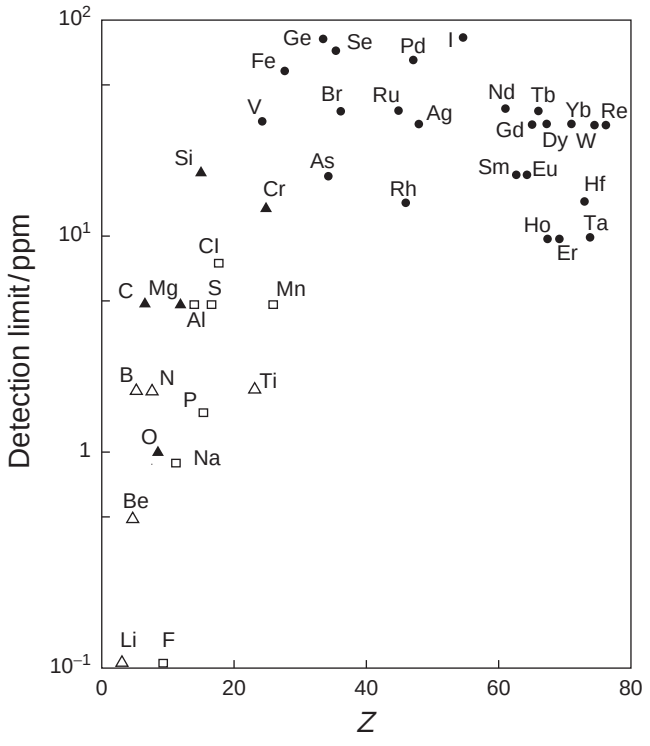

Figure 6 LOD of a number of elements using PIGE for different projectiles (adapted from Borderie[23] and references therein). ●, ^{35}Cl at 55 MeV; , p; , t; △, α; light projectiles at $E \leq 5$ MeV.

Several authors have published tables with thick-target yields.[7,16–18,25,26]

Kenny et al.[25] measured the absolute thick-target yield for several elements at incident proton energies of 2.0 and 2.5 MeV. They measured the yields for several γ-ray lines for elements ranging from F to Au. In Figure 7 the thick-target yield is plotted as a function of the atomic number Z for the strongest γ-ray line of an element. It can be seen that the yield, and thus the sensitivity, is highest for the light elements.

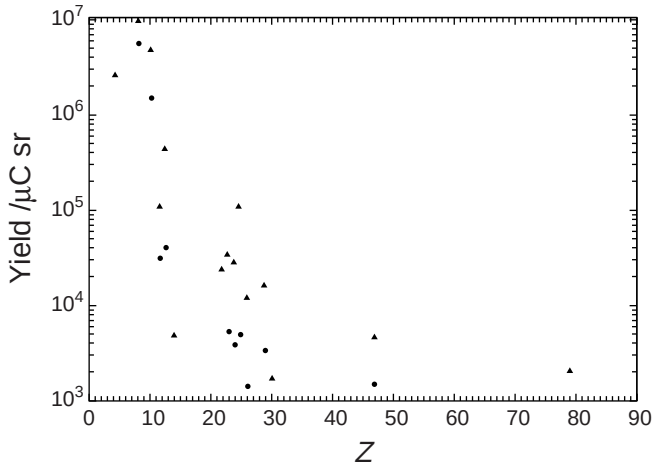

Figure 7 Thick-target yields of the strongest γ-rays for a proton energy E_p of 2 MeV(●) and 2.5 MeV( ). (The figure is drawn with data taken from Kenny et al.[25])

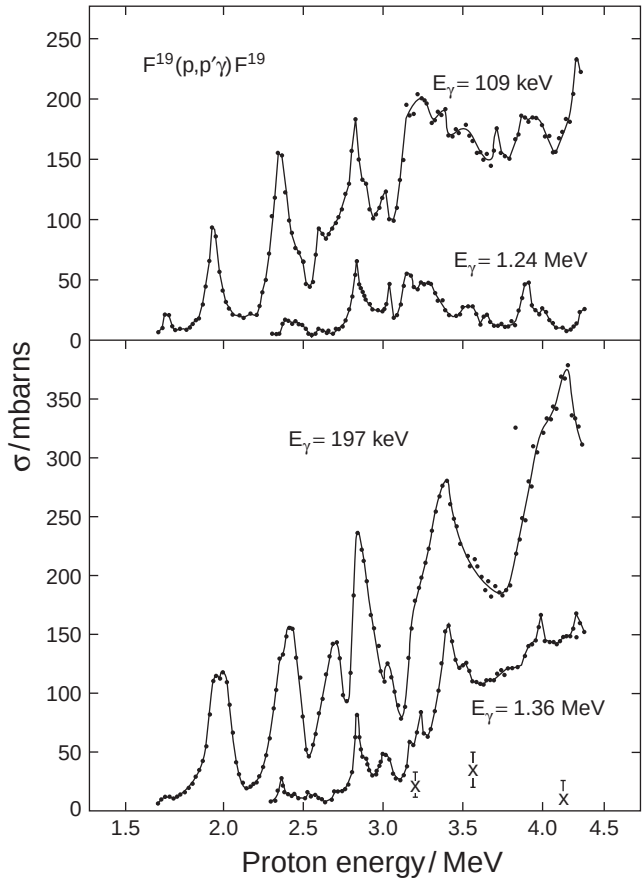

Figure 8 Cross-section for the production of 0.109, 0.197, 1.24 and 1.36 MeV γ-rays from the inelastic proton scattering reaction on fluorine (from Ranken et al.[27]).

Other datasets are presented by Anttila et al.[16] and Kiss et al.[17] They present thick-target yields for several proton energies between 1.0 and 4.2 MeV for isotopes with $4 < Z < 21$. Thick-target γ-ray yields for heavy elements ($Z > 30$) were determined by Räisänen and Hänninen[18] for 1.7 and 2.4 MeV protons. Thick-target γ-ray yields for light elements with 2.4 MeV He^+ are given by Lappalainen[26] and prompt γ-rays generated by 5 MeV α-particles were investigated by Giles and Peisach[7] who determined the sensitivity of elements ranging from lithium to hafnium.

In the following, a few examples of excitation curves (cross-section versus particle energy) for light elements will be given. In Figure 8 the cross-section for the $^{19}F(p,p'\gamma)^{19}F$ inelastic scattering reaction is given[27] for four different γ-rays (109 keV, 197 keV, 1.24 MeV and 1.36 MeV).

For magnesium, excitation functions of ^{24}Mg, ^{25}Mg and ^{26}Mg were determined by Boni et al.[28] Taking into account the isotopic abundances, the 1369 keV γ-ray from $^{24}Mg(p,p'\gamma)^{24}Mg$ has the highest cross-section. As an example, this excitation function is given in Figure 9 for beam energies from 2.8 to 3.8 MeV.

A proper excitation curve is measured such that angular distribution effects are averaged out. For this purpose a large solid-angle detector is often positioned at an angle of 55° with respect to the beam direction.

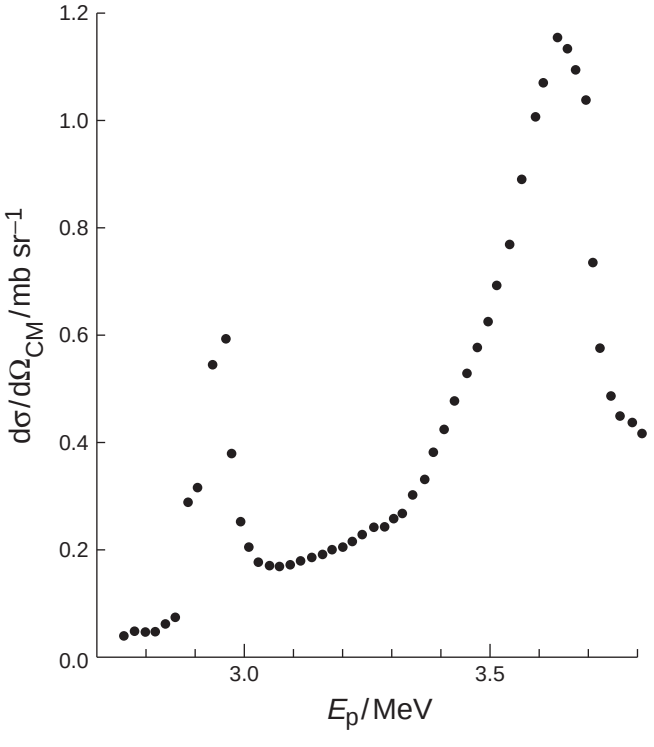

Figure 9 Excitation function for the 1369 keV γ-ray from $^{24}Mg(p,p'\gamma)^{24}Mg$, $\theta_{lab} = 90°$ (from Boni et al.[28]).

3.4 Background

The LOD of an element can be calculated with the help of Equation (30) and is determined by the background radiation under the peak. The background can be caused by minor or major elements in the sample under investigation but also by the experimental set-up and/or the detector.

The background caused by the minor and major elements is dependent on the type and energy of the projectiles. The background is mainly caused by Compton scattering in the detector, depending on the response function and on the γ-rays reaching the detector. These γ-rays can be generated by competing nuclear reactions in the sample and in the surrounding materials.

γ-Radiation can also be produced by reactions taking place in the diaphragms in front of the vacuum chamber, the (aluminum) beam pipe, the vacuum chamber itself, and the Faraday cup used to measure the beam current. In the design of the vacuum chamber care has to be taken to use materials that do not produce large numbers of γ-rays. Light ions, particularly those induced by deuterons, cause more severe background effects than heavy ions. This is mainly due to the lower Coulomb barrier and to the many reaction channels that are open. Precautions have to be taken so that materials do not contain contaminations that can produce γ-rays. Another problem could be natural radioactivity in the surroundings of the experimental set-up. Although the intensity generally is too low to be of any influence, in time-consuming experiments or in cases where a very low detection limit is required, this background can be of importance. To prevent any influence of the surroundings, lead shielding of the detector should be used. Aged lead exhibits less activity than new material.

3.5 Experimental Set-up and Procedures

In this section a typical experimental arrangement to perform PIGE experiments will be presented together with some experimental procedures.

3.5.1 Accelerator and Beam Guidance System

Most PIGE experiments on light elements are performed with energies between 1 and 4 MeV amu^{-1}. This energy is also ideally suited to performing simultaneous measurements with PIXE and/or RBS. In this case a single-ended Van de Graaff accelerator or pelletron accelerator is well suited to perform the PIGE experiments. If reactions are to be used where higher energies and/or heavy ions such as ^{15}N are necessary, tandem accelerators or cyclotrons should be used. In order for the accelerator to perform depth profiling, an easily changeable beam energy and

good energy resolution are required. This excludes a cyclotron for depth profiling, unless improvement of the beam quality is obtained, for example by means of a dispersive system.[24]

Usually, the accelerator is connected to a so-called analyzing magnet which selects a beam with a desired energy for a specific ion. The calibration of the analyzing magnet is usually done by means of a suitable resonant nuclear reaction. Nonlinearity and hysteresis effects of the magnet can make the calibration procedure and the setting of the magnetic field to the correct value critical if great accuracy in the beam energy is required.

Sometimes a proton beam is used[29] with an increased energy spread to make the excitation function more or less flat without any resonance structure. In this case it is not necessary to know the exact value of the beam energy from the accelerator.

The beam is transported to the irradiation chamber through evacuated pipes by means of bending magnets which bend the beam and electrostatic quadrupole lenses which focus the beam. The size of the beam on the target is usually in the range 0.5–5 mm in diameter but can be focused down to below 1 μm in a microprobe set-up.

3.5.2 *Irradiation Chamber*

Many possible layouts of the irradiation chamber are used to perform PIGE analyses. In Figure 10, a possible layout is presented where most of the commonly used features are included. The targets can be mounted on a one- or two-dimensional translation stage. This allows for multiple samples to be mounted and measured without breaking the vacuum by the application of micrometers or stepper motors. Moreover, the surface of the sample can be scanned if inhomogeneities in the sample are expected. One target position is mostly occupied with fluorescent quartz. It can be used to observe and align the beam spot on the target. For the majority of applications it is useful to be able to detect several types of radiation and particles from the sample. Besides a Ge(Li) detector to detect high energy γ-rays (PIGE), a Si(Li) detector to detect X-rays (PIXE) and a surface barrier detector to detect scattered particles (RBS) are usually mounted in the irradiation chamber.

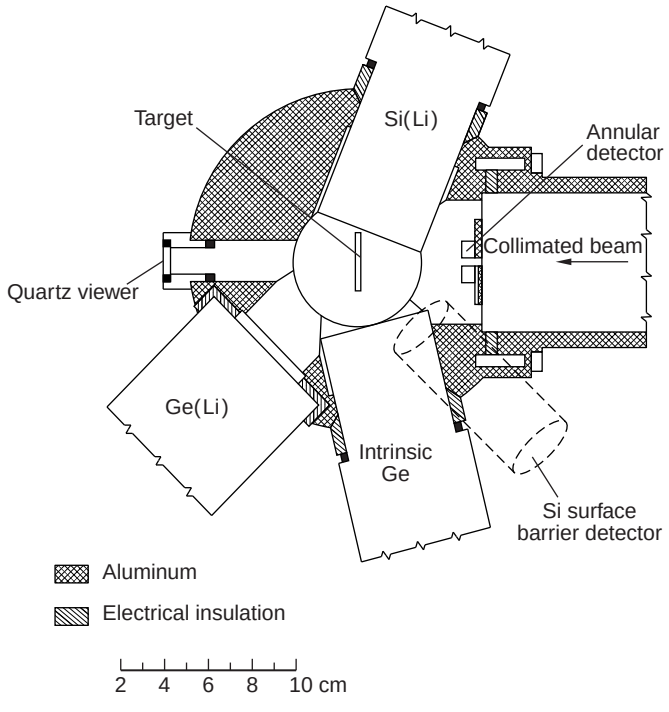

Figure 10 Example layout of a vacuum chamber to perform PIGE experiments (from Borderie[23] and references therein).

For quantitative measurements, the beam charge can be determined in several ways. With thin samples in which the incident particles are not stopped, a Faraday cup behind the sample can be used to determine the beam charge accurately. With thick samples a rotating vane in front of the sample can be used or the beam charge can be determined from a beam current measurement from the sample itself. In all cases, care has to be taken to prevent secondary electrons from the target reaching the Faraday cup by applying secondary electron suppressor rings near the target and/or in front of the Faraday cup.

3.5.3 *Detectors*

The γ-rays in a PIGE experiment can be measured either by a scintillator (NaI(Tl)) or by bismuth–germanium–oxide (BGO), a Ge or a Ge(Li) detector. A detailed description of these detectors can be found in a textbook by Knoll.[30] The NaI(Tl) detector has a relatively poor energy resolution (about 6% at 662 keV) combined with a high efficiency. It can be used in experiments where the energy resolution is not critical but efficiency is important, for example in depth profiling experiments where high efficiency is needed. BGO detectors have the advantage of a higher efficiency than NaI(Tl) for a given size crystal and a better peak-to-Compton ratio. The compact set-up makes BGO well suited to exploiting weak resonances. The superior energy resolution of a Ge detector (<2 keV at 1.33 MeV), a relatively high efficiency and a large peak-to-Compton ratio means that the Ge detector is mainly used in experiments to determine the bulk concentration of complex samples. The energy resolution is high enough to distinguish γ-rays from competing reactions and to analyze interfering peaks in the spectrum. For low-energy γ-rays, use of a planar hyperpure germanium detector can be very useful.[23] It has an energy resolution of about 500 eV at 122 keV.

3.5.4 Electronics

The detectors are connected to charge-sensitive preamplifiers and the resulting pulses are amplified in a spectroscopy amplifier. Shaping and filtering in these amplifiers are used to improve the signal-to-noise ratio. Next, the pulses are fed into a MCA. The analog-to-digital converter (ADC) part of the MCA usually has a resolution of 12 bit (4096 channels), although sometimes a higher resolution is used. The conversion time of the ADC has to be fast enough to prevent dead time problems, especially during high-count-rate experiments. Usually, a conversion time below 10 μs is applied.

3.5.5 Calibration and Spectrum Analysis

First, the γ-ray detection system has to be calibrated. This can be carried out quite easily by using a set radioactive source such as ^{60}Co, ^{208}Tl or ^{207}Bi, depending on the energy region of interest. As an example, in Figure 11, a calibration spectrum is presented. Several peak-fitting routines exist to find the significant peaks and to determine accurately their (channel) positions in the spectrum. To analyze a spectrum, first the peaks have to be identified and possible overlap problems solved. The high energy resolution of a Ge(Li) detector and the limited number of γ-rays means that the peaks in a PIGE spectrum are usually well separated without overlap problems. The actual calibration is usually performed by fitting a linear (or higher) polynomial function between the energy of the peaks in the spectrum and their corresponding positions. The second step is to determine the intensities of the peaks. This is generally performed by a computer code. This computer code is usually also equipped with an automatic peak search algorithm making analysis of PIGE spectra easy to perform. Finally, the intensities of the peaks combined with a sensitivity calibration give the concentration of the isotopes present in the sample.

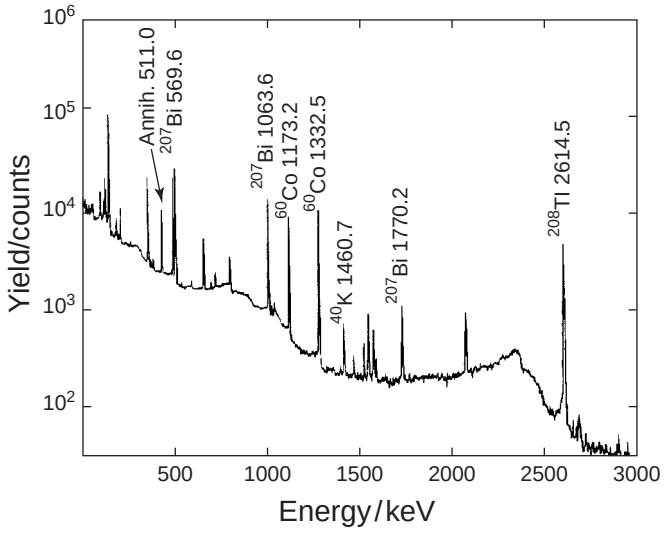

Figure 11 γ-Ray energy calibration spectrum of a Ge(Li) detector using several radioactive sources as indicated at the relevant peaks (from Hirvonen[31]).

The precise beam energy has to be determined, particularly for depth profiling. This can be done by a number of reactions with narrow and strong resonances, depending on the energy of interest; e.g. the reaction ^{27}Al(p,γ)^{28}Si with a resonance energy of 991.90 ± 0.04 keV and ^{15}N(p,$\alpha\gamma$)^{12}C with 429.57 ± 0.09 keV.

3.5.6 Target Preparation

Generally, the target preparation of PIGE samples is easy: just use them as they are. Some geological samples can be polished into sections, 20 μm thick, which allow investigation with an optical microscope.[23] Powdered samples can be pressed into pellets and analyzed directly. Biological materials can be freeze-dried and measured. When liquids have to be analyzed, more complicated procedures have to be followed or an external beam has to be used (see below). In microbeam analysis, care has to be taken not to lose the integrity of the sample during preparation. In the analysis of light elements in biological samples this can be especially difficult to accomplish.

3.5.7 External Beam

Samples such as liquids, wet biological samples or objects with large dimensions cannot be placed inside an evacuated irradiation chamber. These samples can be analyzed with an external beam set-up where the beam exits the beam pipe through a thin foil which is often cooled. The sample is often placed in a helium atmosphere to decrease target heating and to increase the particle range of the protons (and thus diminish straggling) since stopping for hydrogen in helium is low.[23]

3.5.8 Microbeam

A focused microbeam can be used to determine lateral distributions of elemental concentrations on a scale of about 1 μm. A microbeam set-up requires an object and aperture diaphragm combined with a number (2–4) of magnetic quadrupoles to focus the beam down to micrometer sizes.[24] The focusing principles, problems, and applications can be found in a textbook by Watt and Grime.[32]

3.6 Depth Profiling

Several isotopes have the potential for depth profiling, for instance ^{1}H, ^{13}C, ^{15}N, ^{18}O, ^{19}F, ^{22}Ne, ^{23}Na, ^{24}Mg,

^{26}Mg, ^{27}Al, ^{29}Si, ^{30}Si. Most of these light nuclei have sharp and strong resonances in the excitation curve. The instrumentation for performing depth profiling is discussed in section 3.5. Additional requirements for the accelerator are a good energy resolution and the capability to change the energy easily. High-energy resolution of the beam is required in applications where narrow resonances are used. A detailed description of depth profiling can be found in Hirvonen.[31]

4 APPLICATION OF THE PARTICLE-INDUCED γ-RAY EMISSION TECHNIQUE

Possible applications range from geological and archeological samples, steel samples, dust and aerosol samples to biomedical samples. In this section only a few examples are presented of PIGE analyses to serve as illustrations.

Van IJzendoorn et al.[33] used the PIGE technique to quantify thin layers of SiF_x that were a result of reactive ion etching of Si wafers with a CF_4 plasma. The quantification is important to understand the etching process. The ^{19}F(p,p′γ)^{19}F reaction was used to determine F on the Si wafer. A proton energy of 2.78 MeV was chosen to suppress a Si reaction and thus limit the Compton background. An example of a γ-ray spectrum is given in Figure 12.

In Figure 13 an example of the analysis of a hafnium plate is presented.[18] The hafnium plate is bombarded with 10 mC of 2.4 MeV protons. In the spectrum the following elements can be seen: O (150 ppm), Na (0.3 ppm), Al (30 ppm) and P (5 ppm). Also heavier elements including Zr (2.8%), Fe (100 ppm) and Cu (<50 ppm) can be observed.

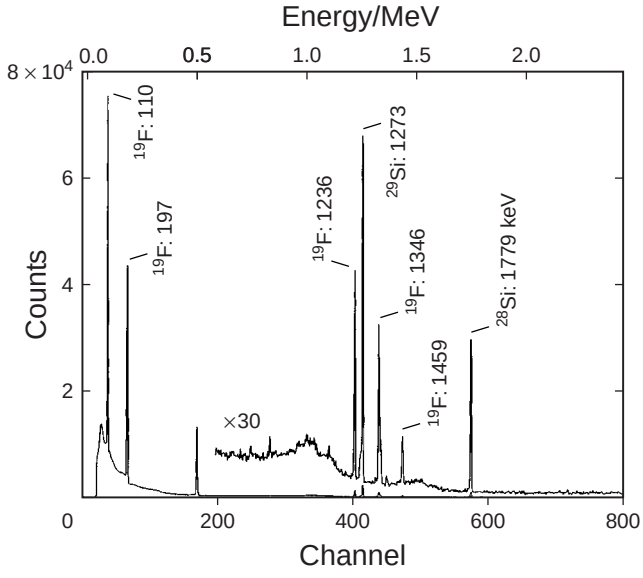

Figure 12 γ-Ray spectrum of a 120-nm thick CF_2 layer on Si measured with 2.8-MeV protons (from van IJzendoorn et al.[33]).

An example of depth profiling with the ^{15}N(p,αγ)^{12}C reaction is given in Figure 14. Here, the diffusion of 40 keV ^{15}N in evaporation-deposited Ni was measured.[34] The nitrogen segregates to the surface and to the Ni–Ta interface at increased temperatures. The nitrogen distribution is very narrow and the broadening is due to experimental resolution at this depth. This is clearly seen at the measured depth profiles.

5 CONCLUSIONS

We have shown that PIGE offers a simple, sensitive and quantitative multielemental analysis technique. With

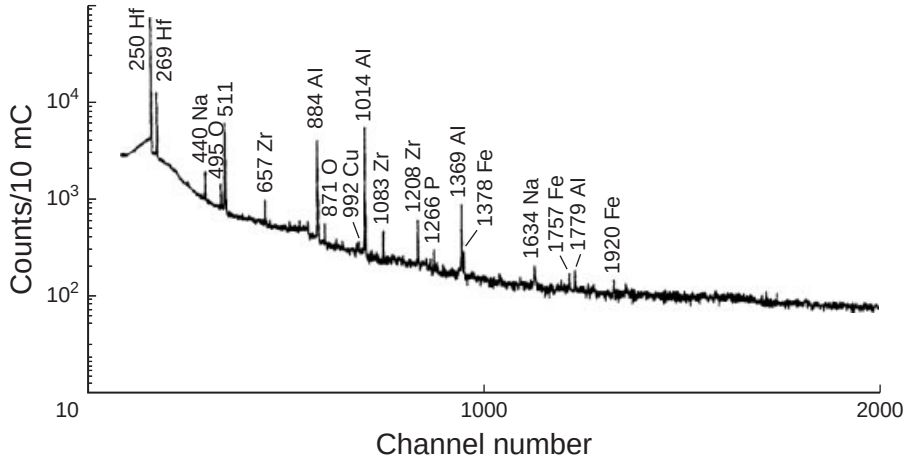

Figure 13 γ-Ray spectrum of a hafnium plate after bombardment with 2.4-MeV protons. A lead absorber in front of the detector was used to decrease the low-energy γ-rays (from Räisänen and Hänninen[18]). The annihilation peak is 511 keV.

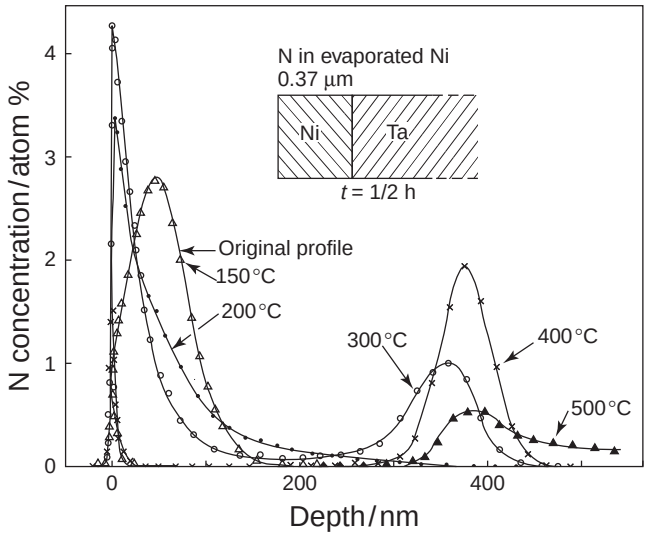

Figure 14 ^{15}N profiles in evaporated Ni on Ta after annealing. The original profile is that of the 40 keV ^{15}N implants (from Hirvonen[31] and Lappalainen and Anttila[34]). The top layer of Ni is 370-nm thick. The measured ^{15}N profiles show increased yields at the surface and at the Ni–Ta interface at increased temperatures (from Lappalainen and Anttila[34]).

tunable accelerators and a good beam energy resolution, depth profiles can be obtained as deep as several micrometers, depending on the material and beam energy. The technique can be combined easily with PIXE and RBS for simultaneous additional information. The strength of PIGE is found particularly for very light elements where other techniques fail or experience difficulties. Sensitivities can reach the parts per million range, depending on the element of interest and on the matrix composition.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
BGO	Bismuth–Germanium–Oxide
CN	Compound Nucleus
fwhm	Full Width at Half-maximum
LOD	Limit of Detection
MCA	Multichannel Analyzer
PIGE	Particle-induced γ-Ray Emission
PIXE	Particle-induced X-Ray Emission
PRA	Prompt Radiation Analysis
RBS	Rutherford Backscattering Spectroscopy

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • PIXE (Particle-induced X-ray Emission)

REFERENCES

1. G. Deconninck, *Introduction to Radioanalytical Physics*, Akademia Kiado, Budapest, 1978.
2. R.F. Sippel, E.D. Glover, 'Sedimentary Rock Analysis by Charged Particle Bombardment', *Nucl. Instrum. Methods*, **9**, 37–48 (1960).
3. T.B. Pierce, P.F. Peck, W.M. Henry, 'The Rapid Determination of Carbon in Steels by Measurement of the Prompt Radiation Emitted During Deuteron Bombardment', *Analyst*, **90**, 339–345 (1965).
4. T.B. Pierce, P.F. Peck, D.R.A. Cuff, 'Application of Inelastic Proton Scattering to the Rapid Determination of Silicon in Steels', *Anal. Chim. Acta*, **39**, 433–436 (1967).
5. G. Deconninck, 'Analysis by Prompt X- and γ-ray Detection', *J. Radioanal. Chem.*, **12**, 157–169 (1972).
6. G. Deconninck, G. Demortier, Laboratoire Interfacultaire d'Analyses par Reactions Nucleaires Reports 725 and 752, LARN, Namur, Belgium, 1972 and 1975.
7. I.S. Giles, M. Peisach, 'A Survey of the Analytical Significance of Prompt γ-rays Induced by 5 MeV Alpha-particles', *J. Radioanal. Chem.*, **50**, 307–360 (1979).
8. R.B. Boulton, G.T. Ewan, 'Simultaneous Analysis of Light Elements Using Prompt Nuclear Reaction γ-rays', *Anal. Chem.*, **49**, 1297–1304 (1977).
9. B. Borderie, J.N. Barrandon, 'New Analytical Developments in Prompt γ-ray Spectrometry with Low-energy Tritons and Alpha Particles', *Nucl. Instrum. Methods*, **156**, 483–492 (1978).
10. P.M. Endt, 'Energy Levels of $A = 21–44$ Nuclei (VII)', *Nucl. Phys. A*, **521**, 1–830 (1990).
11. F. Ajzenberg-Selove, 'Energy Levels of Light Nuclei $A = 18–20$', *Nucl. Phys. A*, **392**, 1–216 (1983).
12. 'Nuclear Data and Nuclear Data Sheets', Academic Press, New York and London.
13. J.F. Ziegler, B.F. Biersack, U. Littmark, *The Stopping and Ranges of Ions in Solids*, Pergamon Press, New York, 1985.
14. K.S. Krane, *Introductory Nuclear Physics*, J. Wiley & Sons, New York, 1998.
15. H. Ejiri, M.J.A. de Voigt, *γ-Ray and Electron Spectroscopy in Nuclear Physics*, Clarendon Press, Oxford, 1989.
16. A. Anttila, R. Hänninen, J. Räisänen, 'Proton-induced Thick-target γ-Ray Yields for the Elemental Analysis of the $Z = 3–9$, 11–21 Elements', *J. Radioanal. Chem.*, **62**, 293–306 (1981).
17. A.Z. Kiss, E. Koltay, B. Nyako, E. Somorjai, A. Anttila, J. Räisänen, 'Measurements of Relative Thick Target

Yields for PIGE Analysis on Light Elements in the Proton Energy Interval 2.4–4.2 MeV', *J. Radioanal. Chem.*, **89**, 123–141 (1985).
18. J. Räisänen, R. Hänninen, 'Heavy Element ($Z > 30$) Thick-target γ-Ray Yields Induced by 1.7 and 2.4 MeV Protons', *Nucl. Instrum. Methods*, **205**, 259–268 (1983).
19. International Union of Pure and Applied Chemistry, 'Nomenclature, Symbols, Units and their Usage in Spectrochemical Analysis', *Spectrochim. Acta*, **33**, 241–245 (1978).
20. P.J. Clark, G.F. Neal, R.O. Allen, 'Quantitative Multielement Analysis Using High Energy Particle Bombardment', *Anal. Chem.*, **47**, 650–658 (1975).
21. B. Borderie, J.L. Pinault, J.N. Barrandon, *Analusis*, **5**, 282 (1977).
22. B. Borderie, J.N. Barrandon, B. Delaunay, M. Basutçu, 'Use of Coulomb Excitation by Heavy Ions (^{35}Cl, 55 MeV) for Analytical Purposes: Possibilities and Quantitative Analysis', *Nucl. Instrum. Methods*, **163**, 441–451 (1979).
23. B. Borderie, 'Present Possibilities for Bulk Analysis in Prompt γ-Ray Spectrometry with Charged Particles', *Nucl. Instrum. Methods*, **175**, 465–482 (1980).
24. P.H.A. Mutsaers, 'Nuclear Microprobe Design', *Nucl. Instrum. Methods, B*, **113**, 323–329 (1996).
25. M.J. Kenny, J.R. Bird, E. Clayton, 'Proton Induced γ-Ray Yields', *Nucl. Instrum. Methods*, **168**, 115–120 (1980).
26. R. Lappalainen, A. Anttila, J. Räisänen, 'Absolute α-induced Thick-target γ-Ray Yields for the Elemental Analysis of Light Elements', *Nucl. Instrum. Methods*, **212**, 441–444 (1983).
27. W.A. Ranken, T.W. Bonner, J.H. McCrary, 'Energy Dependence of $F^{19} + p$ Reactions', *Phys. Rev.*, **109**, 1646–1651 (1958).
28. C. Boni, E. Cereda, G.M. Braga Marcazzan, V. de Tomasi, 'Prompt γ Emission Excitation Functions for PIGE Analysis of Li, B, F, Mg, Al, Si and P in Thin Samples', *Nucl. Instrum. Methods B*, **35**, 80–86 (1988).
29. C. Boni, A. Caridi, E. Cereda, G.M. Braga Marcazzan, 'PIXE–PIGE Analysis of Thin Fly-ash Samples', *Nucl. Instrum. Methods, B*, **45**, 352–355 (1990).
30. G.F. Knoll, *Radiation Detection and Measurement*, J. Wiley & Sons, New York, 1979.
31. J.-P. Hirvonen, 'Nuclear Reaction Analysis: Particle-γ Reaction', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, PA, 1995.
32. F. Watt, G.W. Grime, *Principles and Applications of High-energy Ion Microbeams*, Hilger, Bristol, 1987.
33. L.J. van IJzendoorn, M. Haverlag, A. Verbeek, J. Politiek, M.J.A. de Voigt, 'Quantification of Fluorine in SiF_x Etch Residues on Silicon with Proton Induced γ-Emission', *Nucl. Instrum. Methods, B*, **113**, 411–414 (1996).
34. R. Lappalainen, A. Anttila, 'Annealing Behaviour of Implanted Nitrogen in Bulk and Evaporated Ni', *Appl. Phys. A*, **42**, 263–267 (1987).

Photon Activation Analysis

Kazuyoshi Masumoto
High Energy Accelerator Research Organization, Tokyo, Japan

Christian Segebade
Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany

1 Introduction **2**
1.1 Analytical Use of Electron Accelerators 2
1.2 Principle of Photon Activation Analysis 2
1.3 Short History of Photon Activation Analysis 3
1.4 Textbook and Data Collections for Photon Activation Analysis 4

2 Photonuclear Reaction **4**
2.1 Characteristics of Photonuclear Reactions 4
2.2 Bremsstrahlung for Photon Activation 4
2.3 Photonuclear Reaction Yield 5

3 Analytical Application of Photon Activation **7**
3.1 Conditions of Photon Activation Analysis 7
3.2 Characteristics of Photon Activation Analysis 8

4 Analytical Procedures **8**
4.1 Procedure for Light-element Analysis 8
4.2 Procedure for Multielement Analysis 9
4.3 Sample Preparation, Standard and Flux Monitoring 9
4.4 Irradiation 10
4.5 Measurement and Evaluation 13

5 Interference Management **14**
5.1 Interference Reactions 14
5.2 Peak Overlapping 14
5.3 Other Errors 15

6 Quantitative Methods for Quality Assurance **16**
6.1 Internal Standard Method 16
6.2 Internal Standard Method Coupled with the Standard Addition Method 17
6.3 The Stable Isotope Dilution Method 18

7 Application **19**
7.1 Analysis of Light Elements 19
7.2 Multielement Analysis of Biological Materials 19
7.3 Analysis of Geological and Environmental Materials 19
7.4 Multielement Analysis of Raw Materials and Industrial Products 20

8 Comparison with Other Activation Analysis Methods **20**
8.1 Irradiation 20
8.2 Comparison of Nuclear Reactions Induced 21
8.3 Sensitivity and Selectivity of Elements to be Determined and Applicability to Various Kinds of Materials 21

9 Conclusion **22**

Abbreviations and Acronyms **22**

Related Articles **22**

References **22**

From the large number of analytical methods, activation analysis techniques are the only ones which are based upon nuclear reaction. The material sample studied is exposed to high-energy radiation which can be partly absorbed by a nucleus in the sample. Thus the nucleus is excited to a high energy level which can decay through quasi-prompt emission of a nuclear particle or photon. The product nuclide produced is mostly radioactive, and so emits delayed radiation. Both this and the aforementioned prompt radiation can be measured using appropriate radiation detectors. By evaluating the energy and the count rate of the particles detected, qualitative and quantitative analyses of the target material under study can be performed. Thus it is clear that elements only, not chemical species, can be determined directly. A large variety of particles can be used for activation, namely uncharged ones (neutrons, photons) or charged particles like protons, deuterons, tritons and even heavier ones. Mostly thermal neutrons from nuclear research reactors are used since this technique offers the highest average analytical sensitivity. During photon activation, the target nucleus is activated by photonuclear reaction. This is induced to ''normal'' material at high energies, usually not below about 10 MeV. The photonuclear reaction data of the elements suggest an activation energy around 30 MeV with respect to analytical sensitivity and interfering reactions, respectively. This energy is best achievable with bremsstrahlung sources like high-power linear accelerators or microtrons.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Favorable irradiation parameters are: 30 MeV electron energy at 100–150 μA mean electron beam current. With the help of suitable radiation spectrometers, e.g. high-resolution germanium detectors connected to appropriate pulse processing electronics, photon (γ or characteristic X-ray) spectra can be taken by which simultaneous multicomponent analyses can be carried out without chemical separations, sometimes even nondestructively. Moreover, partly extreme sensitivities can be achieved, and some elements can be analyzed whose determinations are difficult or impossible using other techniques, e.g. light elements like carbon, nitrogen, oxygen and fluorine. A further advantage is the relative freedom from blanks in many cases; after bremsstrahlung exposure, undesirable surface contaminants can be removed from the sample, and the recontamination that eventually occurs is inactive, and thus can be disregarded. Since the activation and measuring process is independent of the chemical status of the component studied, a large variety of matrices can be analysed. Photon activation has been applied in several areas including:

- *geo- and cosmochemistry;*
- *environmental, biological and medical science;*
- *industrial product and high-purity material analysis;*
- *archaeological and forensic science;*
- *certification of reference materials.*

The disadvantages of the method are common to all activation analysis techniques, e.g. the instrumental equipment costs. The cost of a high-performance germanium spectrometer is about US $30 000, and this does not include the permanent costs of maintenance and liquid nitrogen supply. Also, additional personnel qualifications are required for radioactive laboratory work. Finally, the handling of radioactive waste unavoidably produced during activation analysis might be problematic in some cases.

1 INTRODUCTION

1.1 Analytical Use of Electron Accelerators

Electron accelerators have been used not only for nuclear and particle physics but also in many other areas, such as in material, biological, geological and archeological sciences. In the analytical use of electron accelerators, photons such as synchrotron radiation and bremsstrahlung induced by high-energy electrons have been used more frequently than an electron beam itself.

Synchrotron radiation is generated by high-energy electrons circulating in the synchrotron. Synchrotron radiation energy is distributed continuously covering the infrared (IR) to X-ray region and interacts with molecules or atoms. In particular, the X-ray region of synchrotron radiation has been applied extensively to the field of analytical science using X-ray photoelectron spectroscopy (XPS), extended X-ray absorption fine structure (EXAFS), and X-ray fluorescence (XRF); these methods are useful to determine elements and their chemical states. Recently, many facilities using electron storage rings have been constructed for the exclusive use of synchrotron radiation.

However, bremsstrahlung is radiated in a forward direction of the electron beam when the electron is decelerated by Coulomb interactions in a material. Bremsstrahlung has a continuous energy distribution with a maximum energy equal to the electron acceleration energy. As the photon energy is significantly higher than that of synchrotron radiation the bremsstrahlung photon interacts with the nucleus. The cross-sections of photonuclear reactions are about a hundred times larger than that of nuclear reaction induced by electrons of the same energy. Therefore, bremsstrahlung from high-energy electrons can well be used for radioisotope production and particularly for activation analysis. Usually, photon activation analysis (PAA) is carried out using an electron linear accelerator or a microtron which can supply mean electron beam currents of some tens to hundreds of microamperes at energies between 6 to say 60 MeV.

1.2 Principle of Photon Activation Analysis

In PAA, unknown samples are irradiated with high-energy photons. Then, radionuclides produced by photonuclear reactions in the sample (the (γ,n)-type reaction occurs most probably; see below, section 2) are identified by their radiation emission energy, and the intensity of radioactivity is measured. Each radionuclide represents an element originally present in the sample studied, and the intensity of radioactivity is proportional to the abundance of this element.

In general, unknown samples and comparative reference standards are irradiated simultaneously or sequentially, equipped with flux monitors. After correction of the activating photon flux received by the sample and the standard, respectively, the contents of the components of the sample can be calculated by the specific activity ratios of the sample and the comparative standard. Usually, a γ-ray spectrum of sample is measured with a high-resolution spectrometer as described below (see section 1.2.2). In the spectrum, the elements present in the sample are represented by respective peaks whose areas are proportional to the radioactivity of the element to be determined. The peak area ratio of the sample and the standard is commonly used for the calculation of the elemental content. In the following, the different radiation measurement devices commonly used in PAA are described briefly.

1.2.1 Scintillation Spectrometer

A scintillation detector consists of a large, cylindrical thallium (Tl)-doped NaI crystal which is shielded by aluminum housing (3 in. both in height and diameter being the standard size). Incident photon energy emitted from a radioactive sample is converted into visible light by the crystal. A photocathode plus photomultiplier (PM) tube is connected to the crystal. These convert the light flash into an electric pulse signal whose height is proportional to the absorbed photon-ray energy. These signals are processed further (amplified, reshaped and more) by a sequence of amplifiers (preamplifier, spectroscopy amplifier). The signals thus produced have values of several volts, depending upon the incident energy, are Gauss-shaped and short enough (4 μs in the normal approach) to enable undisturbed processing of high pulse frequencies, generated by higher activities of the measured material. The signals are processed further in a multichannel analyzer which produces a spectrum whose lines represent the respective energies emitted by the sample measured.

1.2.2 Coaxial Germanium Spectrometer

This device is used for high-resolution measurement of γ-rays. The basic set-up is similar to that of the NaI spectrometer described above. A high-purity germanium single crystal is applied instead the scintillation crystal plus PM tube. The other pulse-processing electronics are the same as described above. The resolution power exceeds that of an NaI spectrometer by about two orders of magnitude. Therefore, it is indispensable for the measurement of complex multicomponent spectra as mostly occur during instrumental multielement activation analysis.

1.2.3 Low-energy Photon Spectrometer

A low-energy photon detector (LEPD) is commonly used for detecting characteristic X-rays and γ-rays below 100 kV. The LEPD is made of a thin crystal of germanium (typically 10 mm thickness, 30 mm diameter). All the other devices in the spectrometer are the same as used in the coaxial spectrometer described above. The detection efficiency for X-rays is high, because X-rays are fully absorbed even by a thin crystal. Most of the photoactivation product isotopes of medium to heavy elements decay by electron capture (EC), because the (γ,n) reaction products are mostly neutron-deficient nuclides. EC decay is always accompanied by characteristic X-ray emissions whose energies are those of the respective decay products. The γ-ray spectra of medium to heavy elements are often more complex than the X-ray spectra, as shown by Segebade and Weise.[1] Therefore, low-energy photon spectrometry offers a complementary method to γ-ray spectrometry. The LEPD detection sensitivities for 36 elements were compared with that of a usual Ge detector.[2] For quantitative analysis, self-attenuation of X-rays in a sample should be always estimated by the factor

$$F(E) = \frac{1 - \exp[-\mu(E)d]}{\mu(E)d}$$

where d is the sample thickness and $\mu(E)$ is the total attenuation coefficient of the sample material at a given energy. Sato et al.[3] carefully checked the self-absorption effect for the analysis of biological materials and the validity of the method was verified by analyzing several biological materials.

1.3 Short History of Photon Activation Analysis

The pioneering work on photonuclear reactions was done by Chadwick and Goldhaber[4] in 1934. In this work, the photodisintegration of deuterons was induced by 2.62 MeV γ-rays from ^{208}Tl as the excitation source. The radioisotope excitation method can only be applied to limited examples such as several isomers produced by (γ,γ′)-reactions and the photodisintegration of deuterium and beryllium.

Since 1950, electron accelerators have been constructed in many research institutions. These primarily served nuclear physics purposes, but an intense systematic study of PAA also began. In 1954, Basile et al.[5] used a betatron for the determination of oxygen in organic acids. In the beginning, PAA was recognized as an analytical tool for light elements such as C, N and O, which cannot be determined by neutron activation analysis (NAA) using nuclear reactors.

First, Geiger-Müller counters were used for product nuclide radiation counting. In the 1940s, the energy resolution power of scintillation crystal detectors was recognized. Mostly Tl-doped NaI crystals were used for γ-counting, and they are still used for special applications (e.g. light element analysis).

Since the beginning of the 1960s, high-resolution γ-ray spectrometry developed dramatically with the introduction of the Ge-based semiconductor detector, the multichannel analyzer and, somewhat later, the computer. In activation analysis, multielement analysis frequently can be performed easily without chemical separation. PAA has been extensively applied to a large variety of materials as a complementary tool for NAA, and it has been recognized that PAA is suitable for multielement analysis of geological, biological and environmental materials. The analysis of light elements in highly purified materials has been also very important in the field of material science and technology and will be of urgent interest also in the future.

Electron accelerators of comparatively low energy (a few tens of megaelectron-volts) and high power (a few kilowatts) have been used for PAA. However, a limited number of electron accelerators suitable for PAA are available for analysts since modern accelerators are designed for the production of much higher energies to primarily meet the demands of nuclear and particle physicists.

1.4 Textbook and Data Collections for Photon Activation Analysis

In 1972, Engelmann[6] introduced PAA in his book *Advances in Activation Analysis*, in which the principle of PAA and the application to light element analysis were explained in detail. *Photon Activation Analysis* by Segebade et al.[7] is a comprehensive textbook of PAA, containing 1206 references and many useful tables, such as a list of usable and competing photonuclear reactions and a list of γ-rays and X-rays. Kushelevsky[8] presented a short guide of PAA in *Activation Analysis*, where the principles and applications of the method are concisely summarized.

Toms[9] compiled the photonuclear reaction products and their γ-ray energies whilst Williams et al.[10,11] collected γ-ray spectra of 40–44 MeV photon activation products. Lutz and Segebade[12] published a two-dimensional (half-life versus γ-ray energy) map of photonuclear reaction products and Kato[13] obtained sensitivities of 71 elements in PAA using 30 MeV bremsstrahlung. These basic data have been very helpful for users of PAA.

2 PHOTONUCLEAR REACTION

2.1 Characteristics of Photonuclear Reactions

A photonuclear reaction is the interaction between a nucleus and photon. The cross-section of a photonuclear reaction is mainly dependent on the probability of photon absorption. A schematic representation of the cross-section of the photon absorption is shown in Figure 1.[7] The inelastic reaction which is expressed by (γ,γ') can be induced by the irradiation of photon energies even below a few mega electron volts and target nuclides are excited from the ground state to some isomer levels. The cross-section integrals of (γ,γ') reactions are very small, and only a few radioisotopes which are useful for PAA are produced. A photonuclear reaction is usually characterized by the ''giant dipole resonance'' or ''giant resonance'', because this type of reaction has the largest cross-section among several types of photonuclear reactions. When the wavelength of the photon becomes similar to the diameter of nucleus, the photon can be absorbed by the target nucleus through the electric dipole resonance mechanism with a high degree of probability. This phenomenon is interpreted by the fact that a collective oscillation of all protons against all neutrons is induced by the electromagnetic wave. The maximum cross-section region is located at about 14 MeV for heavy elements to about 25 MeV for light elements. After photon absorption, one or two neutrons and/or protons or higher-order particles can be released because the binding energy of one nucleon is about 10 MeV. At about 60 MeV where the wavelength approximately equals the diameter of the deuteron, a neutron–proton pair in a nucleus is likely to absorb the photon. As the result of this reaction, called a ''quasi-deuteron reaction'', a proton and neutron pair is emitted. In this energy region, two or more nucleons can also be released from the target nucleus by a direct process and many kinds of reaction products are formed. The cross-sections of the direct process are much smaller than that of the (γ,n) reaction of the giant resonance region. Above 150 MeV, the total photoabsorption cross-section increases again because of the photo-meson production. As the spallation and fragmentation reactions are also induced, various reaction products are produced. However, the latter two types of photoreaction are not suitable for analytical purposes, but rather have to be considered as sources of interference.

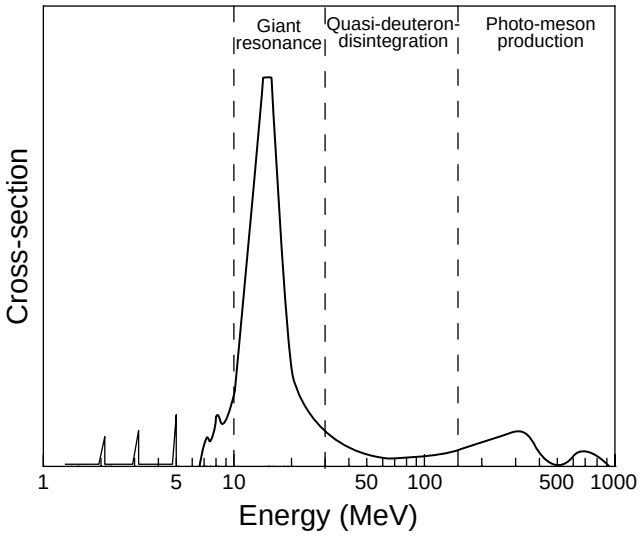

Figure 1 Schematic representation of the total cross-section for the absorption of photons by nuclei. (Reproduced by permission from Walter de Gruyter GmbH & Co.)

2.2 Bremsstrahlung for Photon Activation

Several radioisotopes (e.g. ^{124}Sb) were used as γ-ray excitation sources and applied to the analysis of deuterium, beryllium and several isomeric states. Photon fluxes from

such radioisotopes are usually not sufficient for practical elemental analysis purposes.

Bremsstrahlung is used preferably for activation in PAA. Bethe and Heitler[14] theoretically discussed the radiation loss of electrons. Interaction between electrons and matter can be explained as follows. For low atomic number (Z) target material and low-energy electrons, the major partition of energy loss is due to ionization. For high-Z target and high-energy electron, radiation energy loss is increased. As the radiation loss is proportional to Z^2, elements with higher atomic number more efficiently produce bremsstrahlung. Radiation length is defined as the thickness in which the energy of the electron is reduced to $1/e$ of its incident energy. The radiation lengths of air, aluminum and lead are 304 m (36.7 g cm^{-2}), 8.9 cm (24.0 g cm^{-2}) and 5.6 mm (6.37 g cm^{-2}) respectively. The cross-sections of ionization and radiation loss become equal at a critical energy which is about $800/Z$ MeV, where Z is the atomic number of the target. Therefore radiation loss becomes predominant for electrons above 10 MeV, when a heavy element, e.g. platinum, is used as target material.

In the case of a thin target, bremsstrahlung is radiated strongly to the forward direction, i.e. the direction of the incident electron beam. A thick target is used in practice for activation analysis. In this case, the angular distribution of bremsstrahlung is broadened, and the percentage of low-energy photons increases because of the multiple scattering of the electron in a thick target. Figure 2 shows the calculated photon flux for electron energies of 25, 30 and 35 MeV and a mean electron beam current of 100 μA.[15] The spectrum of bremsstrahlung is calculated by Schiff's equation.[16] The tungsten target is 6 mm (two radiation lengths) thick, which is the range of 35 MeV electrons (thus, all incident electrons are absorbed). The spectrum is integrated over 5° from the direction of the electron beam.

In PAA, the most suitable irradiation energy is 20–30 MeV because the radiated bremsstrahlung spectrum covers the giant resonance region of most of the nuclides. Under such irradiation conditions, reaction products induced by the (γ,n) and (γ,p) reactions can be conveniently used for quantitative element determination.

2.3 Photonuclear Reaction Yield

In PAA, the reaction yield is used for the estimation of sensitivity of the pertinent element and the magnitude of nuclear interference. When a photonuclear reaction is induced by the irradiation of bremsstrahlung of a certain maximum energy the photonuclear reaction yield Y can be derived by the integral in Equation (1) by using the bremsstrahlung spectrum and the excitation function for the relevant reaction,

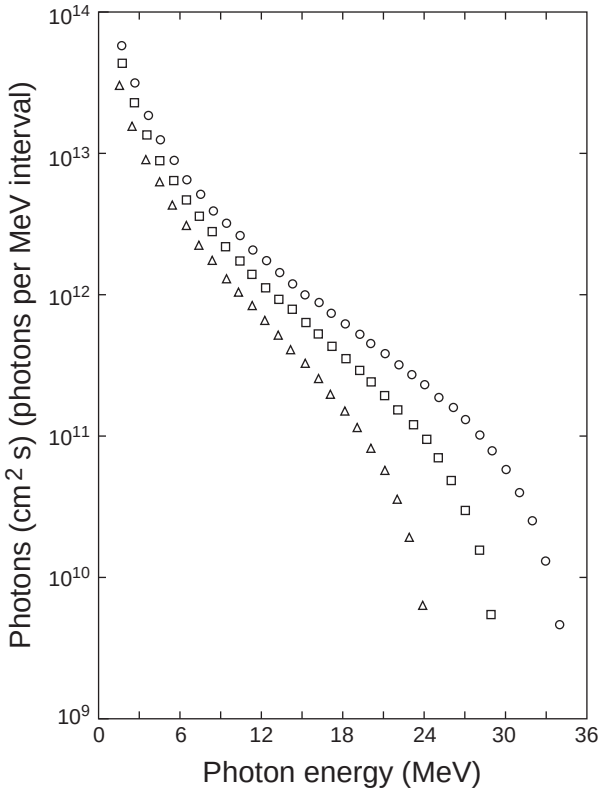

Figure 2 Calculated flux for (○) 35-, ( ) 30-, and (△) 25-MeV electrons for 100 μA average beam current and a subtended angle of 5°. (Reprinted with permission from G.J. Lutz, 'Calculation Sensitivities in Photon Activation Analysis', *Anal. Chem.*, **41**, 426. Copyright (1969) American Chemical Society.)

$$Y = N_0 \int_{E_{th}}^{E_{max}} \sigma(k)\Phi(k, E_{max})k^{-1}\,dk \qquad (1)$$

where N_0 is Avogadro's number, E_{th} is the threshold energy for the reaction, E_{max} is the bremsstrahlung maximum energy, $\sigma(k)$ is the cross-section in square centimeters per nucleus, and $\Phi(k, E_{max})/k$ is the number of photons at a given energy k. If the bremsstrahlung spectrum is normalized to 1 roentgen and a target amount of 1 mole is used, a yield can be obtained per mole per roentgen.

When equivalent quanta, Q, are introduced, the yield per equivalent quanta for the given reaction, σ_Q, is given by Equation (2),

$$\sigma_Q = Q^{-1} \int_{E_{th}}^{E_{max}} \sigma(k)\Phi(k, E_{max})k^{-1}\,dk \qquad (2)$$

where Q is given by

$$Q = E_{max}^{-1} \int_{0}^{E_{max}} \Phi(k, E_{max})\,dk \qquad (3)$$

According to the definition, Q is a hypothetical number of photons per second per square centimeter assuming maximum energy of all photons.

The general feature of a photonuclear reaction yield at the bremsstrahlung maximum energy of 30 MeV is shown in Figure 3 where the yields of various types of photonuclear reactions are plotted against the atomic number.[13] The yields (mol^{-1} R^{-1}) of (γ,n), (γ,2n) and (γ,3n) reactions increase with mass number. The (γ,2n) reaction yield is about one order of magnitude smaller than that of (γ,n) reaction for heavy elements. The (γ,p) reaction shows maximum yield at atomic number 20 and decreases with atomic number. In the case of higher-order charged particle emission reaction, yield curves show a similar shape as that of (γ,p) reaction.

The reaction yields of the (γ,n),[17] (γ,p)[18] and (γ,α)[19] reactions were measured with 20 MeV bremsstrahlung irradiation. The difference of the yields between 20 and 30 MeV are small. The reaction yields were also measured systematically at 60 MeV irradiation.[20] The yields of the single particle emission reactions, such as (γ,n), (γ,p) and (γ,α), are almost equal to the results of 30 MeV irradiation. The yields of (γ,2n), (γ,3n), (γ,4n), (γ,pn) and (γ,αn) reactions are remarkably higher. The (γ,2p) reactions are observed for several light elements. The emission of more particles, up to nine, was observed at 60 MeV.

To observe the dependency of the photonuclear reaction yield upon energy and mass number, the functions of the yield (σ_Q; mb) for the (γ,n), (γ,p) and (γ,pn) reactions are shown in the form of two-dimensional graphs in Figures 4, 5 and 6. In these figures, x- and y-axes show the target mass number and bremsstrahlung

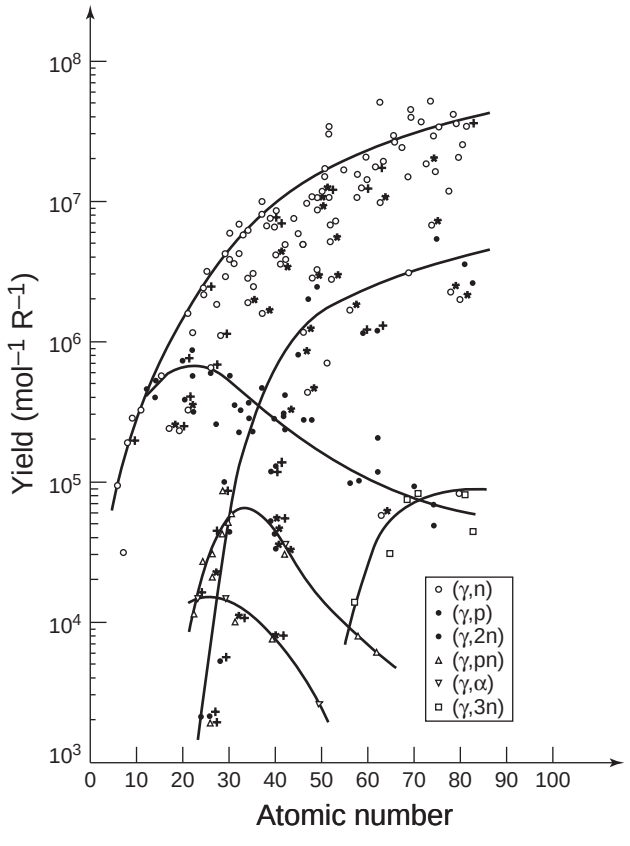

Figure 3 Yields of photonuclear reactions as a function of atomic number with 30 MeV bremsstrahlung. (Reproduced by permission of Akademica Kiado, Budapest.)

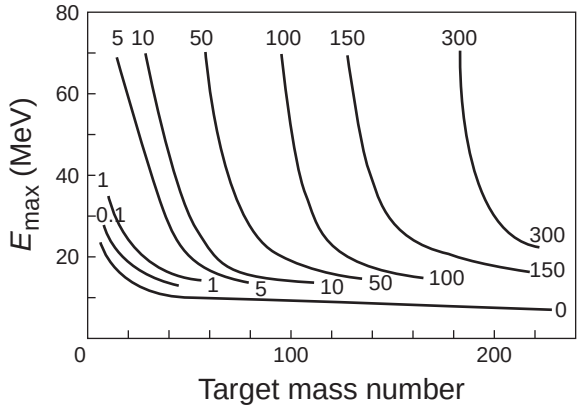

Figure 4 Yields of the (γ,n) reactions as a function of bremsstrahlung maximum energy and target mass number. The numerical values in the figure are yields per equivalent quanta (in mb). (Reproduced from K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 573 (1978) by permission of North-Holland Publishing Co.)

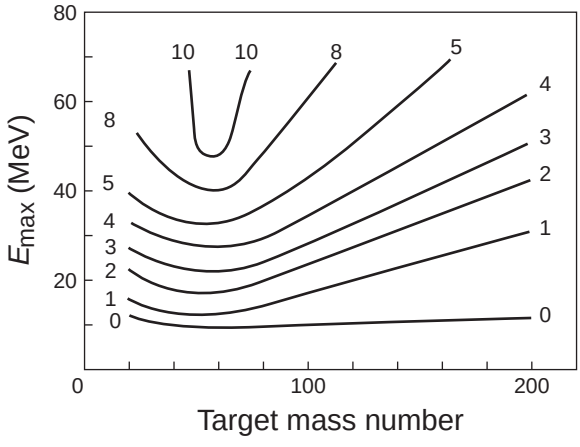

Figure 5 Yields of the (γ,p) reactions as a function of bremsstrahlung maximum energy and target mass number. The numerical values in the figure are yields per equivalent quanta (in mb). (Reproduced from K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 573 (1978) by permission of North-Holland Publishing Co.)

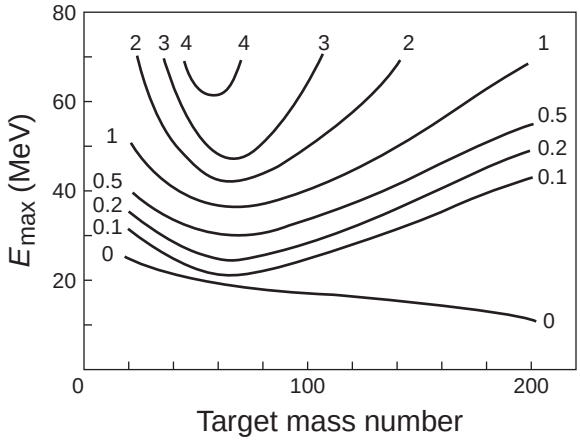

Figure 6 Yields of the (γ,pn) reactions as a function of bremsstrahlung maximum energy and target mass number. The numerical values in the figure are yields per equivalent quanta (in mb). (Reproduced from K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 573 (1978) by permission of North-Holland Publishing Co.)

maximum energy, respectively.[21] In each figure, function graphs of the respective reaction yields are drawn. The yield "zero" marks the calculated mass threshold.

The (γ,n) reactions start to take place in the 10 MeV region, and the yield increases rapidly with energy up to 30 MeV. The curves are relatively parallel to the energy axis above 30 MeV. It is found that the yield is dependent on the mass number rather than on the excitation energy in this energy region.

The charged particle emission reactions such as (γ,p) and (γ,pn) show quite different behavior. The functions run through a maximum for mass numbers between 50 and 60. The reaction yields decrease with increasing mass number at the given bremsstrahlung maximum energy. This is due to the Coulomb potential barrier restriction of high-Z nuclei. Below 50 MeV, the yield of certain mass numbers is dependent on the bremsstrahlung maximum energy. The yield of the (γ,pn) reactions shows similar patterns to that of the (γ,p) reactions though the former have higher threshold and lower yields than the latter.

Similar charts of photonuclear reaction cross-section and the trends of integrated cross-section were also given by Barbier.[22] He also applied the simple calculation method of induced activity. The excitation curves for the photofission reaction, the photoproduction of π-mesons and other various reactions up to the gigaelectron-volt region are shown in this paper.

The (γ,n) reaction is useful for the determination of light to heavy elements and the (γ,p) reaction is favorable for the analysis of medium-Z (from Na up to Sn) elements.

3 ANALYTICAL APPLICATION OF PHOTON ACTIVATION

3.1 Conditions of Photon Activation Analysis

At constant photon flux during irradiation, radioactivity induced in an element to be determined can be calculated as shown in Equation (4):

$$A = \frac{wN_0\theta}{AW}\sigma_Q Q[1 - \exp(-\lambda t)] \qquad (4)$$

where w is the amount of element, θ is the isotopic abundance of the target isotope, AW is the relative atomic mass of element, σ_Q is the yield (cm^2 per equivalent quanta), Q is the flux density (equivalent quanta $cm^{-2}\,s^{-1}$), λ is the decay constant of the induced radioisotope, and t is the irradiation time.

In order to produce bremsstrahlung efficiently, metals of high Z and high melting point, such as platinum and tungsten, are recommended as a converter material. Electron energy which is higher than about 15 MeV is indispensable to induce giant resonance reactions. Beam currents of several tens to hundreds of microamperes are necessary to induce radioactivity giving sufficient detection sensitivity. Betatrons and synchrotrons are not suitable for PAA due to their low beam currents. In Van de Graaf accelerators, beam current is in the microampere region, but the acceleration energy is low. However, many types of electron linear accelerators and microtrons are suitable for PAA.

As an example, consider the induced radioactivity estimated using Equation (1). The equivalent quanta obtained under the irradiation conditions mentioned above are about $10^{14}\,cm^{-2}\,s^{-1}$. The yield of (γ,n) reaction is around 100 mb ($10^{-25}\,cm^2$) for elements in the relative atomic mass region of around 100 (say molybdenum through silver). When $1\,\mu g\,g^{-1}$ of an element in a sample of 100 mg is analyzed, the number of target nuclides is about 10^{14}. After irradiation for a duration equal to 1/10 of the half-life of the radioisotope, induced activity is about 10^3 disintegrations s^{-1}. When the detection efficiency is 1%, 10 counts s^{-1} can be obtained. To obtain γ-ray peak counts within a counting error of 1%, a counting time of 1000 s is necessary at least. Therefore, it seems to be efficient to simply increase the irradiation time, flux, sample mass, detector efficiency and counting time. However, in the instrumental approach (analysis without radiochemical separation), the detection limit cannot always be improved so simply since the background activity might also be increased. The choice of irradiation energy and the cooling time is more effective to attain better detection limits.

Sensitivity data have been published which were obtained for several conditions of irradiation and

measurement, e.g. irradiation energies of 20 MeV,[17] 30 MeV,[13] 40–44 MeV[23] and 110 MeV.[24] Lutz calculated sensitivities for electron energies of 25, 30 and 35 MeV.[15]

3.2 Characteristics of Photon Activation Analysis

Since PAA is a complementary method to the more familiar NAA, the characteristics of both are compared in the following.

3.2.1 Transmittance of High-energy Photons

As high-energy photons can penetrate and activate mostly without notable matrix effects, PAA is suitable for bulk analysis of various kinds of materials and for simultaneous irradiation of several samples. The photon flux decreases with distance from the conversion target. Moreover, the bremsstrahlung beam exhibits a sharp flux density gradient perpendicular to the beam axis. Thus the activating radiation doses of samples and comparison standards should be monitored.

3.2.2 Systematic Trend of Photonuclear Reaction Yields

The (γ,n) reactions are mostly utilized in PAA. As shown in Figures 3 and 4, the threshold energies of the (γ,n) reactions are about 10 MeV for heavy elements and about 20 MeV for light elements. The (γ,n) reaction yields of heavy elements are up to two orders of magnitude higher than those of light elements at the bremsstrahlung maximum energy of 30 MeV. However, light elements are usually major components and heavy elements exist at trace levels in natural products. As the radioactivity induced in an element is roughly proportional to the product of the yield and the abundance of the pertinent element, the differences in radioactivities induced in the light and heavy element fractions are frequently of the same order of magnitude. Such a systematic trend of photonuclear reaction yields gives the well-balanced sensitivities for the instrumental analysis of natural samples.

3.2.3 Interference by Self-shielding Effect, Matrix Activity and Fission Products

In NAA, the self-shielding effect and intense radioactivity sometimes interfere with trace element analysis, as there are several nuclides which have anomalously large neutron capture cross-sections. Typical examples of self-shielding are Cd and B. In NAA, after neutron exposure, ^{24}Na, ^{32}P and ^{82}Br are major activities in biological materials and ^{56}Mn and ^{24}Na are major activities in rock, dust, soil and related samples because of their large neutron activation cross-sections. Rare-earth metals such as Eu and transition metals such as Co also lead to intense radioactivities and self-shielding. In PAA, such a self-shielding effect is mostly negligible, and the problem of matrix activity is not so frequent because of the penetrability of bremsstrahlung and, respectively, the systematic character of photonuclear reaction yields as mentioned above.

When U and/or Th are present in a sample in large amounts, neutron-induced fission products often interfere with the analysis of medium to heavy elements because some of the nuclides formed by fission are the same as those used for NAA of another element. A respective interference might also occur in PAA, but not to the extent as in NAA.

In NAA, the determination of Mg, Ti, Ni, Y, Zr, Nb, I, Tl, Pb and the light elements is very difficult. However, these elements are easily determined by PAA because photonuclear reaction products of these elements have suitable nuclear characteristics for photon spectrometry. More than 30 elements in geological and environmental samples can be determined instrumentally. To reduce radioactivities from light elements and to improve the signal-to-background ratio for heavy elements, 20 MeV irradiation is sometimes preferable to 30 MeV irradiation. Most of the medium to heavy elements can be determined at concentration levels of micrograms per gram and below.

3.2.4 Characteristics of the (γ,n) Reaction Products

Radioisotopes induced by the (γ,n) reaction decay with positron emission and/or EC in the most cases. Several light elements, such as carbon, nitrogen, oxygen and fluorine, lead to purely positron-emitting nuclides. These can be determined by the coincidence counting method (see below, section 4.1). Many of the (γ,n) products from medium to heavy elements decay by EC. In EC decay, the detection of characteristic X-rays of the decay product is convenient for analytical purposes. Low-energy photon spectrometers (see above) are useful to detect characteristic X-rays and low-energy γ-rays.

4 ANALYTICAL PROCEDURES

4.1 Procedure for Light-element Analysis

As mentioned above, the photonuclear activation products of some light elements like C, N, O and F are pure positron emitters without any characteristic photon emission. The annihilation photon quanta ($E = 511$ keV) produced through reaction of positrons with electrons in close vicinity can be measured most favorably with a pair of γ-ray detectors facing one another and electronically switched in coincidence.[6,7] Since the annihilation

quanta, each representing the rest mass of an electron, are emitted from the source at an angle of 180°, the pulse processing electronic system is set up to exclusively count quanta coincidentally absorbed in both detectors. Thus, radiation background is suppressed to a great extent. Because the photon energy is unspecific the activation product (^{11}C, ^{13}N, ^{15}O, ^{18}F respectively) usually has to be separated from the sample matrix after bremsstrahlung exposure. The different appropriate radiochemical separation procedures are described in detail in the literature.[3,4] Heat extraction of radiocarbon, radionitrogen and radiooxygen followed by collection of the respective radiofractions with appropriate agents, as well as vapor distillation of radiofluorine, are utilized mostly. The radiochemical purity of the separated fraction is checked by recording a decay curve during photon counting and comparing the half-life thus obtained with the literature value. Analyses without chemical separations are possible in exceptionally favorable cases only.

The practical analytical procedure typically might be as follows: The analytical sample is prepurified and wrapped in aluminum foil. The comparison standard (usually the pure element, e.g. graphite in the case of carbon analysis, or a stoichiometrically well-determined compound) is also wrapped, and both samples are stacked in an irradiation container, e.g. an aluminum rabbit. This is then transported pneumatically to the irradiation position. The exposure period is adjusted according to the respective half-life and to the expected concentration to be determined. After irradiation, the rabbit is transported back into the laboratory where the samples are unpacked, taken out of the aluminum foil and eventually etched with a cleaning agent to remove surface contaminants. Thereafter, the sample is subject to radiochemical separation. Then the collecting substance containing the radionuclide to be measured is brought into the counting position between the detector pair, followed by multiple countings at a fixed cycle. Counting cycle and data acquisition are performed by a personal computer. The comparison standard is then treated likewise, eventually omitting the cleaning step. The element content is then calculated by comparison of the integral count numbers of the sample and the standard, respectively.

4.2 Procedure for Multielement Analysis

Multielement analysis of natural samples by the comparative method is described as a typical case of PAA. The comparative method (see section 4.1) is mostly used in PAA. When a sample is irradiated together with a comparative standard that contains w_{std} of the element to be determined the unknown amount (w_x) of element in the sample is obtained by Equation (5):

$$w_x = w_{std} \left(\frac{A_x}{A_{std}} \right) \left(\frac{\phi_{std}}{\phi_x} \right) \quad (5)$$

where A_x and A_{std} are the activities induced from the element to be determined and ϕ_x and ϕ_{std} are fluxes of bremsstrahlung irradiated on the sample and the comparative standard respectively.

Usually, the γ-ray peak area ratio of the sample and the standard is used instead of the activity ratio (A_x/A_{std}). In this case, the counting geometry (such as the shapes of sample and standard, the distance from sample and detector) must be equal. As the radioactivity decays according to its half-life after irradiation and also during measurement, the radioactivity of sample and standard should be normalized to the same time, usually the end of irradiation. When the obtained peak count number (C) is normalized to the end of irradiation, the initial activity (A) is calculated by Equation (6):

$$A = \frac{C\lambda}{\exp(-\lambda t_c)[1 - \exp(-\lambda t_m)]} \quad (6)$$

where λ is the decay constant $[\ln(2)/t_{1/2}]$ for the radioisotope, t_c is the cooling time after irradiation, and t_m is the measuring time.

4.3 Sample Preparation, Standard and Flux Monitoring

For the comparative method, a standard containing a well-known amount of each component to be analyzed must be available. This might be a multielement or a single element comparator. It is desirable that such a standard and samples have similar elemental compositions. These materials should be prepared under strictly controlled conditions following preparation procedures which enable maximum accuracy of the respective element concentrations. Certified reference materials (CRMs) should not be used for highly precise concentration calculations, but rather as materials analysed for quality control of the obtained results, the accuracy in particular.

As various kinds of CRMs are issued by many organizations in the world, a standard suitable for nearly each analytical purpose can be chosen. These CRMs are analyzed by several analytical methods at several laboratories. However, most of these certified values are averaged values with standard deviations. Usually only limited numbers of elements are certified. There are several examples where the concentration of the element to be determined is not certified or its concentration is largely different from that in the sample. In order to avoid this, two or more CRMs can be used for cross-checking. When natural products are used as CRMs, homogeneity and stability should be checked carefully. Also limited

amounts of CRMs have been supplied. In most cases the use of CRMs as comparators is not recommended, since high-precision results as required in certification analyses cannot be expected.

Laboratory standards have been useful to solve the above problems. A filter paper doped with several standard solutions of elements to be determined is often used in NAA. This method cannot be recommended for PAA because a small amount of residual acid might decompose the cellulose paper during irradiation. Moreover, the homogeneity of filter paper is questionable. Phenol–formaldehyde resin[25] and gelatine[26] were used as the base materials for activation analysis of biological materials. An artificial standard of polyacrylamide gel matrix was also developed by utilizing the copolymerization reaction of acrylamide in a mixed solution containing the appropriate amounts of elements.[27] Polyacrylamide has high tolerance for radiation and a similar matrix composition to biological materials. For the analysis of rock samples, multielement solution was gelatinized by hydrolysis of tetraethylsilicate.[28,29] As the gelatinization takes place in acidic conditions, most of the standard solution of metals can be used directly. Silica gel has high stability for irradiation. The elemental abundance in such synthetic standards is easily and accurately obtained as the weight ratio of the added amount of element to the final weight of gel, and elemental composition can be adjusted freely according to samples to be analyzed and elements to be determined. The homogeneity of these standards is usually satisfactory because the mixed solution containing several elements is solidified as is.

As photon flux and energy spectrum show pronounced gradients, bremsstrahlung irradiated on each sample should be monitored individually. To estimate flux, sample and standard are stacked alternately or irradiated with flux monitors. The accuracy of the former method is good because the specific activity of each element in the standard is plotted against the irradiation position and the specific activity induced in the sample can be estimated. However, the measurement of samples and standards irradiated is very time-consuming. In the latter method, the monitor reaction data are largely different from the reaction for analysis of each element. Therefore, it is recommended to choose the monitor reaction that has a similar excitation function to the reaction for analysis, and a product radioisotope that has a half-life similar to those of the radioisotopes to be determined.

4.4 Irradiation

An example of an irradiation set-up[30] is shown in Figure 7. The beam profile can be observed with a thin beryllium oxide disk of 0.2 mm in thickness and the beam position is precisely adjusted with a platinum position monitor made up of a thin wire and a small flat tip connected to an electronic microammeter. The platinum tip is positively charged by secondary electron emission when hit by the electron beam. Therefore, the beam position can be pointed exactly to the center of sample by scanning for the maximum current of the microammeter. The beam intensity is indirectly monitored with a ferrite-core-loaded toroidal coil coupled with a current integrator after rectification, because electric current is induced on the toroidal coil when a pulsed beam is passing through the core. This core monitor has been calibrated with a pulse generator. The induced pulses are also monitored

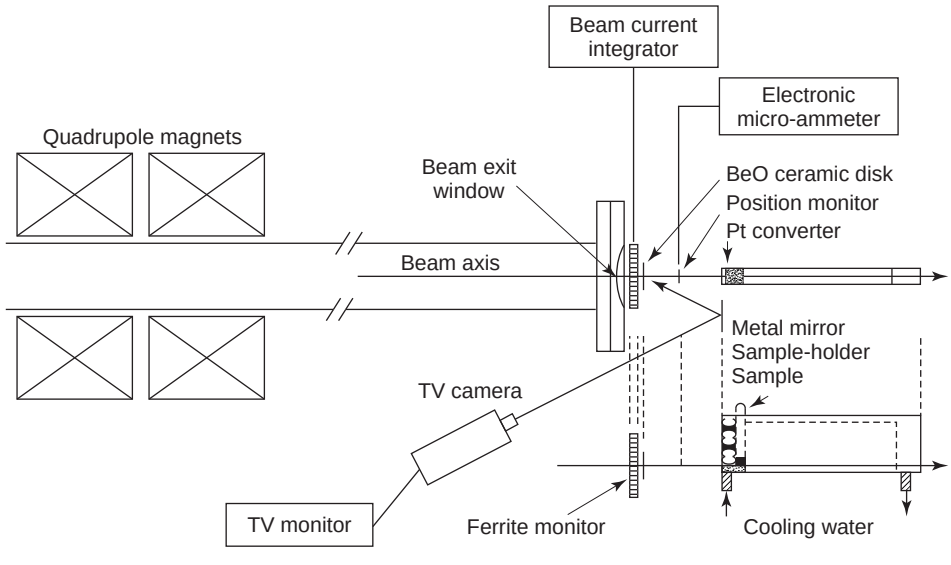

Figure 7 Schematic drawing of the irradiation system. (Reproduced by permission of Akademica Kiado, Budapest.)

with an oscilloscope. Four pieces of platinum plate of 0.5 mm thickness are placed inside the sample holder to serve as the conversion target. The converter and sample are cooled with water to avoid melting of materials caused by high incident beam power. The beam power, which is the product of electron energy and average beam current, turns to heat finally. In this facility, irradiation, for activation analysis has usually been performed at 30 MeV electron energy and 150 μA average current. Hence, 4.5 kW of beam power is focused within a diameter of a few millimeters.

Bremsstrahlung does not cause such significant radiation damage on the sample. Electrons penetrating the converter cause radiation damage and local heating which creates severe problems especially for biological samples. To reduce effects caused by penetrating electrons, a graphite block has been inserted between the converter and sample. Light elements absorb electrons efficiently without reducing the flux of bremsstrahlung from the converter considerably. Also, a bending magnet placed behind the converter to deflect electrons passing through the converter has been applied. In this case, samples should be kept at a good distance from the converter to avoid unwanted electron bombardment. Another possibility to avoid electron-induced damage is to use a thick target which absorbs all electrons (see above).

Secondary neutrons are mainly emitted from the converter and also from surrounding materials. These neutrons are predominantly induced by photonuclear reactions in the giant resonance region. It is known that energy spectra of photoneutrons are predicted by the statistical model of the nucleus, and the angular distribution is mainly isotropic.[31] Photoneutrons are retarded by the collision with surrounding materials and thermalized by water and other light material in the vicinity. Such neutrons cause the (n,γ), (n,p) and (n,α) reactions in a sample.

The spatial distributions of bremsstrahlung and photoneutrons are shown in Figure 8.[32] Figure 8 shows the results of Pt (2 mm) plus graphite (10 mm) and Pt (2 mm), W (3 mm) and graphite (10 mm). The solid lines show the distribution of activity of ^{196}Au induced by the ^{197}Au(γ,n)^{196}Au reaction which corresponds to the flux of bremsstrahlung. In Figure 8, the spatial distribution of neutrons is drawn by dotted lines, according to the ^{198}Au activity induced by the ^{197}Au(n,γ)^{198}Au reaction. According to the result of Engelmann,[6] the best thickness of

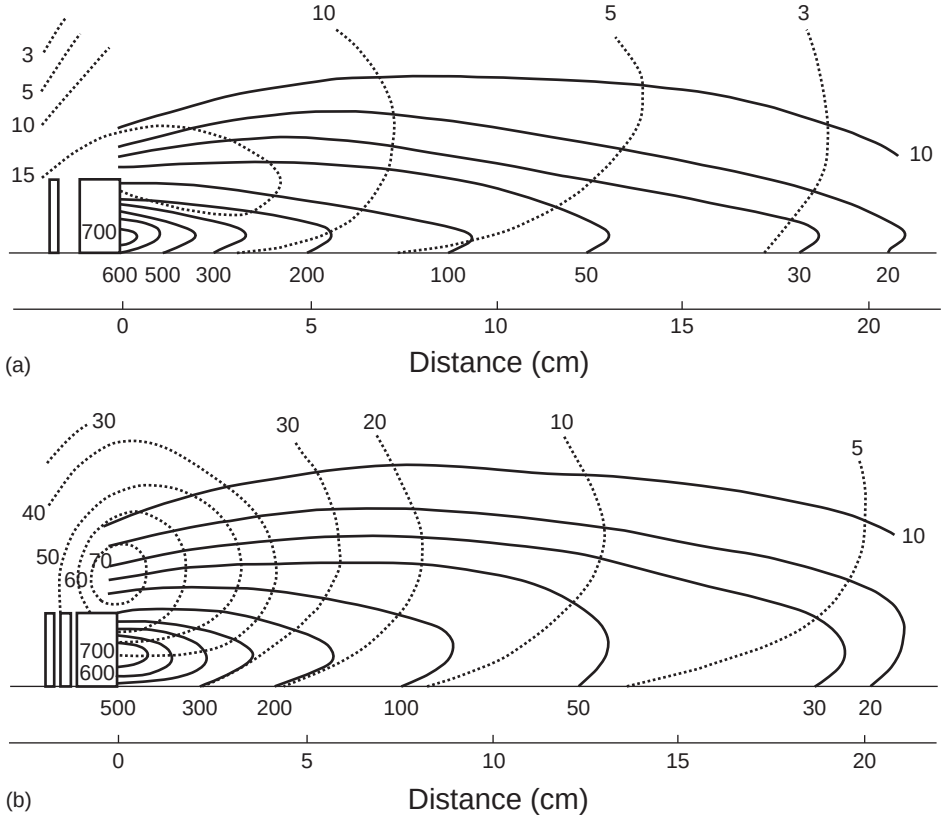

Figure 8 Distribution of the intensity of bremsstrahlung and photoneutrons in sample holder by the irradiation of 30 MeV electrons: (a) 2 mm Pt converter + 10 mm graphite; (b) 2 mm Pt + 3 mm W converter + 10 mm graphite. Straight line: ^{196}Au activity (354 keV γ-ray peak counts $s^{-1}mg^{-1}$). Dotted line: ^{198}Au activity (412 keV γ-ray peak counts $s^{-1}mg^{-1}$).

platinum converter to induce (γ,n) reactions is 3 g cm^{-2} at 30 MeV and 4.5 g cm^{-2} at 45 MeV. As the thickness of 2 mm is 4.3 g cm^{-2}, the converter thickness presented in Figure 8(a) is suitable for activation. In this case, electron energy is reduced by half of the initial energy. The target combination of Figure 8(b) is used for the irradiation of biological materials. The additional converter of 3 mm thick W plate can reduce the initial electron energy by a factor of six, and the graphite block of 10 mm in thickness can fully absorb electrons passing through the converter. As this graphite block absorbs only 3 to 4% of photons at the energy of 10 to 30 MeV, photon flux is not suppressed significantly. As shown in Figure 8(b), the bremsstrahlung beam produced by the converter of 5 mm in thickness is broadened, and the photoneutron flux is increased. The full widths at half-maximum of bremsstrahlung induced by 2- and 5-mm-thick converters are 12 and 16 mm, respectively, directly behind

Table 1 Pertinent nuclear data for PAA of natural products

Element	Target nuclide	Isotopic abundance (%)	Reaction	Product nuclide	Half-life[a]	γ-Ray energy (keV)
C	^{12}C	98.89	(γ,n)	^{11}C	20.39 m	511
N	^{14}N	99.63	(γ,n)	^{13}N	9.97 m	511
O	^{16}O	99.76	(γ,n)	^{15}O	1.18 m	511
F	^{19}F	100	(γ,n)	^{18}F	1.830 h	511
Na	^{23}Na	100	(γ,n)	^{22}Na	2.602 y	1274.5
Mg	^{25}Mg	10.0	(γ,p)	^{24}Na	14.66 h	1368.6
Si	^{30}Si	3.1	(γ,p)	^{29}Al	6.6 m	1273.4
P	^{31}P	100	(γ,n)	^{30}P	2.50 m	511
Cl	^{35}Cl	75.77	(γ,n)	^{34m}Cl	32.2 m	146.4, 2127.7
K	^{39}K	93.26	(γ,n)	^{38}K	7.64 m	2167.7
Ca	^{44}Ca	2.09	(γ,p)	^{43}K	22.3 h	372.8, 617.5
Ca	^{48}Ca	0.187	(γ,n)	^{47}Ca	4.536 d	1297.1
Sc	^{45}Sc	100	(γ,n)	^{44}Sc	3.93 h	1157
Ti	^{47}Ti	7.4	(γ,p)	^{46}Sc	83.83 d	889.3, 1120.5
Ti	^{48}Ti	73.7	(γ,p)	^{47}Sc	3.341 d	159.4
Ti	^{49}Ti	5.4	(γ,p)	^{48}Sc	1.821 d	983.5, 1037.5, 1312.0
Cr	^{52}Cr	73.79	(γ,n)	^{51}Cr	27.704 d	320.1
Mn	^{55}Mn	100	(γ,n)	^{54}Mn	312.2 d	834.8
Fe	^{57}Fe	2.15	(γ,p)	^{56}Mn	2.578 h	846.8
Co	^{59}Co	100	(γ,n)	^{58}Co	70.92 d	810.8
Ni	^{58}Ni	68.3	(γ,n)	^{57}Ni	1.503 d	1377.6
Cu	^{65}Cu	30.8	(γ,n)	^{64}Cu	12.701 h	511.0, 1345.8
Zn	^{66}Zn	27.9	(γ,n)	^{65}Zn	244.1 d	1115.5
Zn	^{68}Zn	18.8	(γ,p)	^{67}Cu	2.580 d	184.6
Ga	^{69}Ga	60.1	(γ,2n)	^{67}Ga	3.261 d	184.6, 300.2
As	^{75}As	100	(γ,n)	^{74}As	17.78 d	595.9
Se	^{76}Se	9.0	(γ,n)	^{75}Se	119.77 d	136.0, 264.7
Br	^{79}Br	50.69	(γ,2n)	^{77}Br	2.3765 d	239.0, 520.6
Rb	^{85}Rb	72.17	(γ,n)	^{84}Rb	32.9 d	881.7
Sr	^{88}Sr	82.6	(γ,n)	^{87m}sr	2.80 h	388.4
Y	^{89}Y	100	(γ,n)	^{88}Y	106.61 d	898.1, 1836.1
Zr	^{90}Zr	51.5	(γ,n)	^{89}Zr	3.268 d	909.2
Nb	^{93}Nb	100	(γ,n)	^{92m}Nb	10.15 d	934.5
Mo	^{100}Mo	9.6	(γ,n)	^{99}Mo	2.7477 d	140.5
Sb	^{123}Sb	42.7	(γ,n)	^{122}Sb	2.70 d	564.4
I	^{127}I	100	(γ,n)	^{126}I	13.0 d	388.6, 666.4
Cs	^{133}Cs	100	(γ,n)	^{132}Cs	6.48 d	667.7
Ba	^{136}Ba	7.85	(γ,n)	^{135m}Ba	1.20 d	268.3
Ce	^{140}Ce	88.5	(γ,n)	^{139}Ce	137.66 d	165.9
Tm	^{169}Tm	100	(γ,n)	^{168}Tm	93.1 d	198.2, 815.9
Tl	^{203}Tl	29.5	(γ,n)	^{202}Tl	12.23 d	439.6
Pb	^{204}Pb	1.42	(γ,n)	^{203}Pb	2.169 d	279.2
U	^{238}U	99.275	(γ,n)	^{237}U	6.75 d	208

[a] y, years; m, months; d, days, h, hours.

the graphite block. Through quasi-elastical scattering the photoneutrons are also distributed behind the graphite block. The spatial distribution and neutron thermalization are influenced by the irradiation conditions. To estimate the interference caused by photoneutrons it is recommended that neutron monitors be applied adjacent to the samples during each radiation cycle.

High irradiation energy is preferable to obtain better sensitivity. However, as several types of photonuclear reactions are increasingly induced at the incident electron energy above 30 MeV, nuclear interference problems arise at higher irradiation energy.

4.5 Measurement and Evaluation

The elements and nuclear data for PAA of typical natural samples are listed in Table 1. Since each radioisotope has a specific half-life, detection sensitivity and signal-to-background ratio, which depend upon both matrix activity and radioactivity to be analyzed, are strongly decay-time dependent. Figure 9 shows the time dependency of detection limits of nine elements in coal fly ash (NIST Standard Reference Material (SRM) 1633a) irradiated at 25 MeV for 2 h. The detection limit is defined as the concentration of element which yields the count number equal to 3σ of background counts under pertinent γ-ray photopeak. As shown in Figure 9,[32] the detection limit strongly varies with cooling time. For multielement analysis of natural samples, γ-ray spectra of samples should be measured several times at appropriate intervals after irradiation. In Figure 10,[32] the detection limits of 26 nuclides are plotted against a suitable cooling time in the case of coal fly ash quoted above. The degree of interference from γ-ray overlapping also changes according to the half-lives of pertinent nuclides. Therefore, the cooling time should be carefully adjusted in order to analyze each element under the best possible conditions.

As a result, typical γ-ray measurements for the analysis of the type of natural products studies are performed preferably in the following periods: (1) a few minutes after irradiation for K, Cl; (2) a few hours after irradiation for Ti(^{47}Sc), Sc, Fe, Sr; (3) a few days after irradiation for Mg, Ca(^{43}K), Ti(^{48}Sc), Ni, Ba, Pb; (4) 1 week after irradiation for Ca(^{47}Ca), Zn(^{67}Cu), Cs, Zr, Mo, Sb, U; and (5) several weeks after irradiation for Na, Ti(^{46}Sc), Cr, Mn, Co, Zn(^{65}Zn), As, Rb, Nb, Y, I, Ce, Tl. Two or more radioisotopes are available for the determination of several elements, e.g. Ca, Ti and Zn.

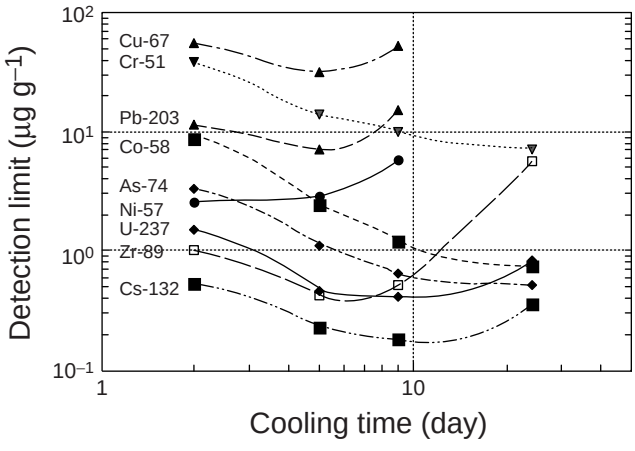

Figure 9 Cooling time dependence of the detection limit of nine elements in coal fly ash.

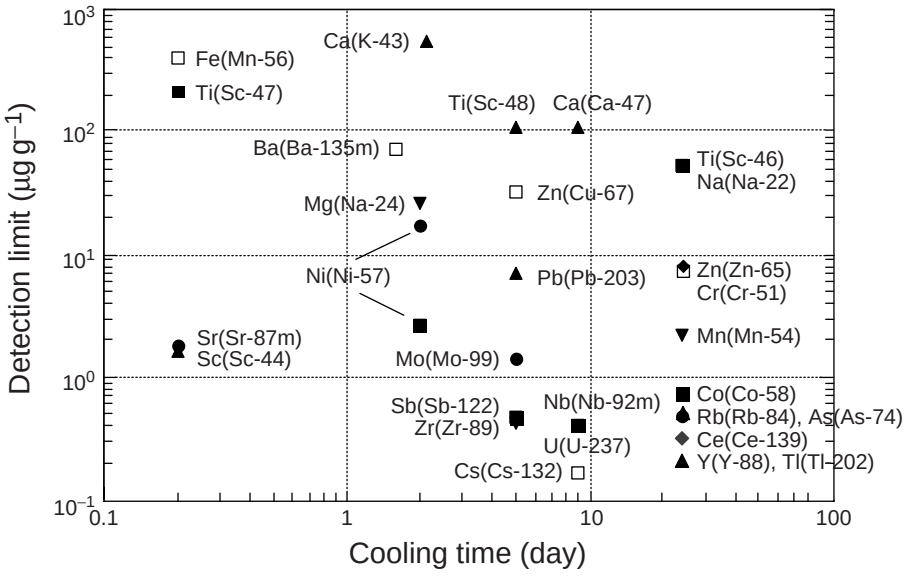

Figure 10 Detection limits of 26 nuclides obtained at a suitable cooling time in coal fly ash.

5 INTERFERENCE MANAGEMENT

5.1 Interference Reactions

In PAA, a radioisotope used for determination can be produced by another element. The interference reactions have been discussed by Kato et al.,[20,21] Williams and Hislop,[23] Segebade et al.[33] and Miyamoto et al.[34]

Typical examples of nuclear interference are listed in Table 2. For example, $^{55}Mn(\gamma,n)^{54}Mn$ and $^{52}Cr(\gamma,n)^{51}Cr$ reactions are used for the determination of Mn and Cr. However, ^{54}Mn and ^{51}Cr can be produced by $^{56}Fe(\gamma,pn)^{54}Mn$ and $^{56}Fe(\gamma,\alpha n)^{51}Cr$ reactions from Fe. The degree of interference is strongly dependent upon the irradiation conditions and, in particular, the Mn/Fe and Cr/Fe abundance ratios. Hence, when natural materials are analyzed the interference of Cr and Mn might be crucial since usually Fe is in significantly higher concentrations than the former elements in these materials. For the determination of Na(^{22}Na) through $^{23}Na(\gamma,n)^{22}Na$ the analytical reaction suffers interference from $^{24}Mg(\gamma,pn)^{22}Na$ and $^{27}Al(\gamma,\alpha n)^{22}Na$. The interferences from Mg and Al should be checked as well. The interference reactions are always accompanied by charged particle emission. As these reactions are suppressed by the nuclear Coulomb barrier, interference reactions are not so severe for heavier elements.

Photoneutrons also cause interference problems for the analysis of several elements. For example, ^{24}Na from $^{25}Mg(\gamma,p)^{24}Na$ for the determination of Mg and ^{56}Mn from $^{57}Fe(\gamma,p)^{56}Mn$ for the determination of Fe suffer interference from $^{23}Na(n,\gamma)^{24}Na$ and $^{55}Mn(n,\gamma)^{56}Mn$, respectively. The distribution of photoneutrons is critically dependent on the irradiation condition and set-up as shown in Figure 8. Table 3 shows the weight ratio of the amount to yield the same activity from interference and analytical reaction. The weight ratios (yield of analytical reaction/yield of interfering reaction; calculated for equal element masses) were obtained at several irradiation positions using a 5 mm thick converter. The weight ratios in interference cases caused by photonuclear reactions decrease with the distance from the graphite block. Hence, the interference yields increase with the distance from the converter. In this interference case, the change of the weight ratio reflects the change of bremsstrahlung spectrum. This is due to the larger scattering cross-section of lower photon energies. In the case of photoneutron interference, the fast neutron flux rapidly decreases with the distance from the converter; consequently, the interference yield by slow neutrons becomes more significant than that by fast neutrons.

The nuclear reaction yields of the analytical reaction and the interference reaction were studied systematically.[21] The results are listed in Table 4 for the SRM orchard leaves (NIST SRM 1571) and in Table 5 for the geochemical reference sample JB-1 (basalt) issued by the Geological Survey of Japan. The magnitude of interference, expressed as the percentage of the activity induced by the interference reaction to the total activity, increases with the irradiation energy.

Table 2 Interference reactions in PAA

Element	Analytical reaction	*Q*-value (MeV)	Interfering reaction	*Q*-value (MeV)
C	$^{12}C(\gamma,n)^{11}C$	18.7	$^{16}O(\gamma,\alpha n)^{11}C$	25.9
			$^{14}N(\gamma,p2n)^{11}C$	31.2
F	$^{19}F(\gamma,n)^{18}F$	10.4	$^{23}Na(\gamma,\alpha n)18F$	20.9
Na	$^{23}Na(\gamma,n)^{22}Na$	12.4	$^{24}Mg(\gamma,pn)^{22}Na$	19.2
			$^{27}Al(\gamma,\alpha n)^{22}Na$	22.5
Mg	$^{25}Mg(\gamma,p)^{24}Na$	12.1	$^{27}Al(\gamma,2pn)^{24}Na$	31.4
			$^{27}Al(n,\alpha)^{24}Na$	–[a]
			$^{23}Na(n,\gamma)^{24}Na$	–[a]
Cl	$^{35}Cl(\gamma,n)^{34m}Cl$	12.6	$^{39}K(\gamma,\alpha n)^{34m}Cl$	19.9
K	$^{39}K(\gamma,n)^{38}K$	13.1	$^{40}Ca(\gamma,pn)^{38}K$	21.4
Sc	$^{45}Sc(\gamma,n)^{44}Sc$	11.3	$^{46}Ti(\gamma,pn)^{44}Sc$	21.7
Ti	$^{48}Ti(\gamma,p)^{47}Sc$	11.4	$^{51}V(\gamma,\alpha)^{47}Sc$	10.3
			$^{48}Ca(\gamma,n)^{47}Ca\rightarrow{}^{47}Sc$	10.0
Cr	$^{52}Cr(\gamma,n)^{51}Cr$	12.0	$^{56}Fe(\gamma,\alpha n)^{51}Cr$	19.7
			$^{54}Fe(n,\alpha)^{51}Cr$	–[a]
Mn	$^{55}Mn(\gamma,n)^{54}Mn$	10.2	$^{56}Fe(\gamma,pn)^{54}Mn$	20.9
			$^{54}Fe(n,p)^{54}Mn$	–[a]
Fe	$^{57}Fe(\gamma,p)^{56}Mn$	10.6	$^{55}Mn(n,\gamma)^{56}Mn$	–[a]
			$^{59}Co(\gamma,2pn)^{56}Mn$	28.0
Co	$^{59}Co(\gamma,n)^{58}Co$	10.5	$^{60}Ni(\gamma,pn)^{58}Co$	20.0
			$^{63}Cu(\gamma,\alpha n)^{58}Co$	16.2
Zn	$^{68}Zn(\gamma,p)^{67}Cu$	10.0	$^{71}Ga(\gamma,\alpha)^{67}Cu$	5.3
As	$^{75}As(\gamma,n)^{74}As$	10.2	$^{76}Se(\gamma,pn)^{74}As$	19.8
Rb	$^{85}Rb(\gamma,n)^{84}Rb$	10.5	$^{86}Sr(\gamma,pn)^{84}Rb$	20.1
Sr	$^{88}Sr(\gamma,n)^{87m}Sr$	11.1	$^{89}Y(\gamma,pn)^{87m}Sr$	18.2
Y	$^{89}Y(\gamma,n)^{88}Y$	11.5	$^{90}Zr(\gamma,pn)^{88}Y$	19.8

[a] Photoneutron reaction.

5.2 Peak Overlapping

Another type of interference is the spectral overlap of peaks whose energies are so close that the spectrometer cannot resolve them. These might be due to signals of different nuclides or one product nuclide whereby the former case is more critical than the latter. However,

Table 3 Weight ratio of several interference reactions (converter thickness = 5 mm)

Interfering reaction	Analytical reaction	Distance from graphite block (cm)		
		0	6	14
$^{56}Fe(\gamma,pn)^{54}Mn$	$^{55}Mn(\gamma,n)^{54}Mn$	92	71	60
$^{55}Mn(n,\gamma)^{56}Mn$	$^{57}Fe(\gamma,p)^{56}Mn$	0.29	0.20	0.18
$^{23}Na(n,\gamma)^{24}Na$	$^{25}Mg(\gamma,p)^{24}Na$	29	19	17
$^{27}Al(n,\alpha)^{24}Na$	$^{25}Mg(\gamma,p)^{24}Na$	204	185	165

Table 4 Interference in PAA of NIST SRM 1571 orchard leaves

Element	Abundance ratio	Magnitude of interference (%)[a] E_{max}		
		30 MeV	40 MeV	60 MeV
Na(^{22}Na)	Mg/Na = 75.6	51.9	88.7	92.5
	Al/Na = 4.99	0.38	6.2	13.9
Mg(^{24}Na)	Al/Mg = 0.066	0.031	0.09	0.56
Mn(^{54}Mn)	Fe/Mn = 3.3	2.0	13.0	20.2
Cr(^{51}Cr)	Fe/Cr = 130	0.76	37.3	54.2
Zr(^{89}Zr)	Mo/Zr = 15.4	0.073	0.95	7.1
Fe(^{56}Mn)	Mn/Fe = 0.303	9.77[b]		
Mg(^{24}Na)	Na/Mg = 0.0132	0.000059[b]		

[a] (Activity from interfering element)/(Total activity) × 100.
[b] Interference caused by photoneutron.

Table 5 Interference in PAA of JB-1 (basalt) issued from the Geological Survey of Japan

Element	Abundance ratio	Magnitude of interference (%)[a] E_{max}		
		30 MeV	40 MeV	60 MeV
Na(^{22}Na)	Mg/Na = 2.25	2.83	17.7	26.7
	Al/Na = 3.70	0.28	4.69	10.7
Mg(^{24}Na)	Al/Mg = 1.65	0.78	2.2	12.1
Ti(^{47}Sc)	Ca/Ti = 8.32	12.2	10.8	8.1
	V/Ti = 0.0266	0.12	0.13	0.13
Mn(^{54}Mn)	Fe/Mn = 54.1	31.1	71.1	80.6
Cr(^{51}Cr)	Fe/Cr = 155	0.96	41.3	58.5
Y(^{88}Y)	Zr/Y = 6.00	0.83	–	4.1
	Nb/Y = 0.59	–	–	0.067
Zr(^{89}Zr)	Mo/Zr = 0.15	0.00071	0.0094	0.075
Nb(^{92m}Nb)	Mo/Nb = 1.5	0.096	0.53	0.95
Fe(^{56}Mn)	Mn/Fe = 0.303	0.66[b]		
Mg(^{24}Na)	Na/Mg = 0.0132	0.20[b]		

[a] (Activity from interfering element)/(Total activity) × 100.
[b] Interference caused by photoneutron.

the use of doublet or multiplet peaks from one product nuclide should also be avoided. Typical examples for peak overlap interference by different nuclides are listed in Table 6. This type of interference mainly occurs during analysis of medium to heavy elements. To avoid the interference the difference of half-lives of product nuclides can be effectively used in the following cases: Co/Ca, As/Ca, Tl/Zn and Pb/Hg. However, this frequently entails excessive time consumption in the analytical task. On the other hand, if one nuclide emits two or more γ-ray energies, the peak ratio method can be applied to subtract the interfering peak area. Typical examples are Zn/Ga, Nb/Fe, Cs/I, I/Y and U/Ga. The latter method reduces the time consumption, but a greater analytical error is induced.

Nowadays the γ-ray overlapping interference is not so problematic due to the dramatic improvement of the energy resolution power of modern photon spectrometers; a modern Ge detector offers peak half-widths of 1.8 keV or less at 1333 keV (^{60}Co). Moreover, several highly efficient peak deconvolution programs have been developed.

5.3 Other Errors

As mentioned in section 3.1, the induced radioactivity is proportional not only to sample mass but also to the cross-section and isotopic abundance of the sample components, flux of bremsstrahlung and saturation factor. In the case of γ-ray measurement, the peak count number obtained is dependent on the counting time, γ-ray branching ratio of radioisotope, detection efficiency of γ-ray and others. In the above factors, sample mass, irradiation time and counting period can be adjusted to reduce statistical errors and to improve precision and detection limit. However, the improvement of these factors is often restricted by the experimental conditions, such as number of samples and limited access to irradiation and measurement facilities. It is important to store the peak counts until the counting error becomes small enough.

Activation analysis includes the following steps: (1) sample preparation; (2) irradiation; (3) measurement; and (4) analysis. To improve accuracy, systematic errors

Table 6 Typical interferences caused by γ-ray overlapping

Element	Analytical reaction	Half-life	γ-ray (keV)	Interfering reaction	Half-life	γ-ray (keV)	Other γ-ray (keV)
Co	^{59}Co(γ,n)^{58}Co	70.92 d	810.8	^{48}Ca(γ,n)^{47}Ca	4.536 d	807.9	1297.1
Zn	^{68}Zn(γ,p)^{67}Cu	2.580 d	184.6	^{69}Ga(γ,2n)^{67}Ga	3.261 d	184.6	300.2
As	^{75}As(γ,n)^{74}As	17.78 d	595.9	^{44}Ca(γ,p)^{43}K	22.3 h	593.4	372.8, 617.5
Nb	^{93}Nb(γ,n)^{92m}Nb	10.15 d	934.5	^{54}Fe(γ,pn)^{52}Mn	5.591 d	935.5	744.2, 1434.1
I	^{127}I(γ,n)^{126}I	13.0 d	388.6	^{89}Y(γ,2n)^{87}Y	3.346 d	388.4	484.9
Cs	^{133}Cs(γ,n)^{132}Cs	6.48 d	667.7	^{127}I(γ,n)^{126}I	13.0 d	666.4	388.6
Tl	^{203}Tl(γ,n)^{202}Tl	12.23 d	439.6	^{70}Zn(γ,n)^{69m}Zn	13.76 h	438.6	
Pb	^{204}Pb(γ,n)^{203}Pb	2.169 d	279.2	^{204}Hg(γ,n)^{203}Hg	46.60 d	279.2	
U	^{238}U(γ,n)^{237}U	6.75 d	208.0	^{69}Ga(γ,2n)^{67}Ga	3.261 d	209.0	300.2

in each step should be checked carefully. Nuclear interference and γ-ray overlapping are avoidable, but some systematic errors, such as flux monitoring and counting geometry, cannot be cancelled out by quantitative methods. In the following section, the internal standard method, the standard addition method and the stable isotope dilution activation analysis used for highly accurate analyses are described.

6 QUANTITATIVE METHODS FOR QUALITY ASSURANCE

6.1 Internal Standard Method

The most reliable method for monitoring the activating flux received by a sample is the method using an internal standard because the flux monitoring component is inherent in both the sample and the comparative standard. This method was called the internal reference method by Tsai et al.[35] and Yagi and Masumoto[36] or the internal monitor method by Oka et al.,[37] in order to differentiate the slight change of the definition of each technique applied.

First, the comparative standard in which the abundance ratio (K) of the element (a) to be determined and the internal standard element (b) are accurately known is prepared. It is not necessary to know the absolute abundance of each element. The accurately weighed amount (w_b) of the internal standard is added to the sample and mixed homogeneously. Then, appropriate amounts of the sample and the comparative standard are irradiated. The activity ratio of a pertinent element to the internal standard is measured for the sample (R) and the comparative standard (R^*), respectively, and the relevant equations are shown as Equations (7) and (8).

$$R = \frac{A_a}{A_b} = \frac{(w_a N_0 \theta_a / AW_a)\sigma_a Q[1 - \exp(-\lambda_a t_a)]}{(w_b N_0 \theta_b / AW_b)\sigma_b Q[1 - \exp(-\lambda_b t_b)]} \quad (7)$$

$$R^* = \frac{A_a^*}{A_b^*} = \frac{(w_a^* N_0 \theta_a / AW_a)\sigma_a Q^*[1 - \exp(-\lambda_a t_a)]}{(w_b^* N_0 \theta_b / AW_b)\sigma_b Q^*[1 - \exp(-\lambda_b t_b)]} \quad (8)$$

The sample weight and flux (Q) are eliminated by introducing the activity ratio. Again, by including the ratio of the two activity ratios (R^*/R) mentioned above, nuclear reaction yields, saturation factor, isotope abundance and relative atomic mass can be cancelled out. Using Equation (9), abundance ratios of the sample (w_b/w_a) and the comparative standard (K) are left.

$$\frac{R^*}{R} = K\frac{w_b}{w_a} \quad (9)$$

Therefore, the abundance of element a (w_a) can be calculated by Equation (10):

$$w_a = K w_b \left(\frac{R^*}{R} \right. \quad (10)$$

When the ratio is obtained as the γ-ray peak area ratio, Equation (11) can be used.

$$\frac{R^*}{R} = \frac{(C_a^*/\phi(\gamma_a)\varepsilon(\gamma_a)\Omega^*)/(C_b^*/\phi(\gamma_b)\varepsilon(\gamma_b)\Omega^*)}{(C_a/\phi(\gamma_a)\varepsilon(\gamma_a)\Omega)/(C_b/\phi(\gamma_b)\varepsilon(\gamma_b)\Omega)} \quad (11)$$

where $\phi(\gamma)$ is the branching ratio of the γ-ray, $\varepsilon(\gamma)$ is the detector efficiency at the γ-ray energy, Ω is the solid angle of the measurement, and C is the peak area. The asterisk designates the standard. Finally, all these factors can be cancelled out by applying the double ratio. The double ratio can be expressed as two γ-ray peak area ratios.

In this method, it is only necessary to know the abundance ratio (K) of the comparative standard and amount of internal standard element (w_b) added to the sample, respectively. It is not necessary to weigh the sample and the comparative standard irradiated, to monitor the photon flux on sample and standard by the flux monitors, or to adjust strictly the geometrical condition of measurement, because this formula only needs the activity ratio. However, the latter statement is valid in limits only since the functions of counting efficiencies of the analytical signal and the internal monitor signal over the distance of the sample from the detector might differ. This difference can be significant particularly in the close vicinity of the detector.

It is essential that the irradiation energy for both the sample and the standard is the same. The difference of matrix effect between the sample and the comparative standard cannot be compensated by this method.

As to the comparison of the internal standard method with the single comparator method[38] as frequently utilized in NAA: in the single comparator method, the same amount of an element (a) to be determined and the comparator element (b) are irradiated and the activity ratio of element a to b is measured in advance. The activity ratio (C), which is regarded as constant, can be expressed as shown in Equation (12).

$$C = \frac{(\theta_a / AW_a)\sigma_a[1 - \exp(-\lambda_a t_a)]}{(\theta_b / AW_b)\sigma_b[1 - \exp(-\lambda_b t_b)]} \quad (12)$$

To determine an element in the sample, a sample and single comparator pair is irradiated simultaneously and the abundance of the element (w_a) can be calculated by the activity ratio (A_a/A_b), the amount of comparator (w_b) and constant as in Equation (13):

$$w_a = C w_b \left(\frac{A_a}{A_b} \right. \quad (13)$$

Although the quantitative equation of the internal standard method is analogous to that of the single comparator method, the principles of both methods are quite different from each other. The accuracy of the single comparator method is mostly unsatisfactory in PAA, because it is very difficult to keep the activity ratio constant. The energy spectrum for every irradiation and the beam current during one irradiation cannot be considered constant for every accelerator.

For highly pure materials, one of the main matrix components can be used conveniently as internal standard without adding an exotic element. This method was applied for the analysis of impurity or minor constituent in metals or alloys by Kato et al.[35,37] If a matrix element (atomic number Z) leads to isotopes of $(Z-1)$ or $(Z-2)$ by the $(\gamma,\mathrm{p}x\mathrm{n})$ or $(\gamma,\alpha x\mathrm{n})$ reactions, these radioisotopes can be used for internal standardization. Then, a $(\gamma,x\mathrm{n})$ product from neighboring elements (atomic number $Z-1$ or $Z-2$) can be chemically separated together with isotopes of $(Z-1)$ or $(Z-2)$ induced from the matrix. As these internal standards chemically behave like the $(\gamma,x\mathrm{n})$ products from impurity, their activity ratio is always constant. Therefore, the unknown amount can be determined without correction of chemical yield.

When the internal standard is added to the sample, the following conditions are required: none or a negligibly small amount of internal standard exists in the sample; only one or two γ-rays emit from the internal standard; and γ-rays from the internal standard do not interfere with γ-rays to be analyzed. Sc, Y and Sm have been used routinely as internal standards for a variety of samples by Segebade et al.,[39] and Ce was used for the trace element analysis of aluminum alloys by Yagi and Masumoto.[36]

To maintain maximum accuracy it is desirable that the half-life and γ-ray energy of the internal standard are similar to those of radioisotopes to be analyzed, and that the cross-section of the photonuclear reaction of the internal standard is similar to that of an element to be determined because the full energy spectrum of bremsstrahlung cannot be accurately monitored by using only one photonuclear reaction and also the photon flux tends to fluctuate in each irradiation.

6.2 Internal Standard Method Coupled with the Standard Addition Method

To achieve higher analytical accuracy without using any comparative standard and to compensate matrix effects arising from irradiation and measurement, the internal standard method coupled with the standard addition method has been proposed. This method was initiated by Lariaert et al.[40] in NAA as the internal standard method. After Lariaert et al. applied it to NAA of minor elements in high alloy steels,[41] Yagi and Masumoto[42] revived it by the full use of high-resolution γ-ray spectrometry and a new sample preparation technique. In order to distinguish it from the method dealt with in the previous section, this method is called the internal standard method coupled with the standard addition method. In this section, the definition and characteristics of this method are described in some detail.

In this method, duplicate samples are taken. The accurate amount (w_0) of element to be determined is added to one portion of sample. The sample and the standard addition sample are irradiated simultaneously and the activity ratios of the pertinent element and internal standard are measured for both sample (R) and standard addition sample (R^*).

The activity ratio of the standard addition sample can be expressed by Equation (14):

$$R^* = \frac{A_a^*}{A_b^*} = \frac{[(w_a + w_0)N_0\theta_a/AW_a]\sigma_a Q^*[1-\exp(-\lambda_a t_a)]}{(w_b N_0\theta_b/AW_b)\sigma_b Q^*[1-\exp(-\lambda_b t_b)]} \tag{14}$$

Then, by calculating the double ratio (R^*/R), the amount of internal standard element (w_b) is also cancelled out. As shown in Equation (15), a very simple relation is obtained.

$$\frac{R^*}{R} = \frac{w_a + w_0}{w_a} \tag{15}$$

The unknown amount of w_a is calculated using Equation (16):

$$w_a = \frac{w_0}{[(R^*/R) - 1]} \tag{16}$$

In this method, it is not necessary to know the amount of internal standard present in the sample (w_b). Only the amount of element added to the sample (w_0) is used for calculation. If an appropriate internal standard does not exist in the sample, the same amount of external element is added to the sample and the standard addition sample.

This method was applied to multielement analyses and one or more elements were used simultaneously as internal standards. For PAA of biological, geological and environmental samples, Na, Mg, Ca, Ti and Fe, usually inherent in natural samples, are useful internal standards because the concentrations of these elements are relatively high and half-lives and γ-rays of reaction products are convenient for the measurement. This method has been applied to PAA and charged particle activation analysis (CPAA) of biological materials,[42–44] sediment,[45,46] soil,[47,48] fly ash[49] and alloys.[50–52]

The use of two or more internal standards is effective to cross-check the analytical results.[53] The homogeneity of elements and internal standards in the sample and the standard addition sample is always checked by comparing with the standard deviations of activity ratios (R^* and R).

If $\sigma(R^*)$ is always larger than $\sigma(R)$ of a certain element, only this element was not mixed homogeneously at the standard addition step. If $\sigma(R)$ is always larger than $\sigma(R^*)$ of a certain element, this element is not homogeneous in the sample itself.

This method combines the characteristics of the internal standard method and that of the standard addition method. In the standard addition technique, comparative standards are not necessary. This method can eliminate several systematic errors, which are involved in the comparative method, such as the matrix composition of sample, the geometrical conditions of irradiation and measurement, and the analysis of the γ-ray spectrum.

In order to check whether loss or contamination of an element is caused during sample preparation, another portion of sample is prepared in the same procedure as the standard addition sample without adding an element to be determined and the double ratio (R^*/R) is also measured. When this ratio is <1, the pertinent element might be lost during the standard addition procedure. When this ratio is >1, contamination has to be considered. The analysis of CRMs by this method is useful to check the degree of interference reaction. If the content of one element is systematically higher than the certified value, the positive deviation indicates interference by other component(s) in the sample. As this method is capable of self-checking it is very useful for the analysis of reference materials.

6.3 The Stable Isotope Dilution Method

Stable isotope dilution activation analysis utilizes isotope analysis. For example, Ce has two stable isotopes, ^{140}Ce (88.48%) and ^{142}Ce (11.08%), which can produce ^{139}Ce and ^{141}Ce, respectively, by (γ,n) reactions. When an enriched isotope ^{142}Ce is added to the sample, the activity ratio of ^{139}Ce and ^{141}Ce changes. The degree of the activity ratio change is relative to the abundance of Ce in the sample and the added amount of ^{142}Ce. The conditions of this method are that the elements to be determined consist of at least two stable isotopes, and these two or more isotopes produce radioisotopes that can be detected separately by γ-ray spectrometry.

Stable isotope dilution activation analysis is performed using the following steps: the enriched isotope j is used as a spike, in which the isotopic abundances (θ_s) of i and j are known; the enriched isotope (w_s) is added to the sample; and the sample and the spiked sample are irradiated simultaneously. The activity ratio induced by two stable isotopes (i and j) is measured for the sample (R_x) and the spiked sample (R_m) using Equations (17) and (18).

$$R_x = \frac{A_{ix}}{A_{jx}} = \frac{(w_x\theta_{ix}/AW_x)\sigma_i[1-\exp(-\lambda_i t_i)]}{(w_x\theta_{jx}/AW_x)\sigma_j[1-\exp(-\lambda_j t_j)]} \quad (17)$$

$$R_m = \frac{A_{im}}{A_{jm}} = \frac{(w_x\theta_{ix}/AW_x + w_s\theta_{is}/AW_s)\sigma_i[1-\exp(-\lambda_i t_i)]}{(w_x\theta_{jx}/AW_x + w_s\theta_{js}/AW_s)\sigma_j[1-\exp(-\lambda_j t_j)]} \quad (18)$$

where the subscripts x, s and m indicate the sample, an enriched isotope and the spiked sample, respectively, w_x is the weight of element in the sample, w_s is the spiked amount of an enriched isotope, AW is the relative atomic mass, R is the activity ratio (isotope i/isotope j), and θ_j and θ_i are the isotopic abundances of isotopes i and j respectively.

Again the ratio of these activity ratios is used as shown in Equation (19):

$$\frac{R_m}{R_x} = \frac{(w_x\theta_{ix}/AW_x + w_s\theta_{is}/AW_s)(w_x\theta_{ix}/AW_x)}{(w_x\theta_{jx}/AW_x + w_s\theta_{js}/AW_s)(w_x\theta_{jx}/AW_x)} \quad (19)$$

Then, the equation is slightly simplified to Equation (20):

$$\frac{R_m}{R_x} = \frac{1 + (w_s/w_x)(\theta_{is}/\theta_{ix})(AW_x/AW_s)}{1 + (w_s/w_x)(\theta_{js}/\theta_{jx})(AW_x/AW_s)} \quad (20)$$

From Equation (20), the unknown amount (w_x) of element can be obtained using Equation (21):

$$w_x = w_s\left(\frac{AW_x}{AW_s}\right)\left(\frac{\theta_{js}}{\theta_{jx}}\right)\frac{(R_m/R_x) - (\theta_i/\theta_j)_s/(\theta_i/\theta_j)_x}{1-(R_m/R_x)} \quad (21)$$

This equation was introduced by Masumoto and Yagi[54] in 1983.

When highly enriched isotopes can be used as spikes, the abundance ratio $(\theta_i/\theta_j)_s/(\theta_i/\theta_j)_x$ becomes much smaller than the activity ratio (R_m/R_x). Then, a simple equation can be obtained, as shown by Equation (22):

$$w_x = w_s\left(\frac{AW_x}{AW_s}\right)\left(\frac{\theta_{js}}{\theta_{jx}}\right)\frac{1}{(R_x/R_m)-1} \quad (22)$$

As the unknown amount of an element can be determined by the activity ratio of two radioisotopes, the stable isotope dilution method is regarded as another type of internal standard method. In order to reduce error, the ratio (R_m/R_x) should preferably range from 0.3 to 0.7.[55]

In PAA, this technique was proposed for the determination of Ca, Zn and Ce.[54] Furthermore, it was applied to PAA[56] and CPAA[57] of Sr in biological materials. In the case of isotope dilution, it is not necessary to determine the chemical yield when the element to be determined is separated before or after irradiation. Hence, Sr was separated chemically before irradiation. Simultaneous determination of Ca, Rb, Sr and Ce was tried in PAA[58] in order to show the applicability of this technique in multielement analysis.

7 APPLICATION

7.1 Analysis of Light Elements

One of the important features of PAA is the analysis of light elements, such as C, N, O and F. In the book by Engelmann et al.,[59] PAA techniques for light element analysis have been described in detail and compared with chemical methods and CPAA. The radioisotopes produced by light elements are positron emitters. In order to analyze such elements, chemical separation after irradiation is mostly necessary. Fedoroff et al.[60] systematically studied the separation of radiocarbon in molten salts and expanded their studies to the simultaneous determination of C and N.[61]

In this section, some analytical examples of carbon and fluorine are described.

7.1.1 Determination of Carbon

Yoshioka et al.[62,63] determined the carbon impurity in copper and gallium arsenide. In PAA, trace amounts of carbon can be determined by detecting ^{11}C produced by $^{12}C(\gamma,n)^{11}C$. As ^{11}C is a positron emitter with a half-life of 20.39 min, a rapid chemical separation of radioactive carbon is necessary after irradiation (yet faster separation schemes have to be applied for oxygen analyses via ^{15}O; half-life = 2 min). Copper samples were oxidized with Sn as a flux material, and GaAs with Pb_3O_4. $^{11}CO_2$ produced in the combustion process was collected as $Ba^{11}CO_3$. The separation procedure was completed within one half-life of ^{11}C after irradiation. In the case of Cu and GaAs, the sensitivity is 0.02 $\mu g\,g^{-1}$ or 10^{15} atoms cm^{-3}, respectively.

Carbon impurity in a highly purified semiconductor has been usually analyzed with IR spectrophotometry. This method uses the relationship between the IR absorbance and the carbon concentration. The absorption of certain wavelengths of IR light is effected by the resonance with a vibration mode of a chemical bond between carbon and the substrate. A nuclear reaction, however, is not influenced by the chemical state of carbon. This fact suggests that the calibration factor between carbon concentration versus absorbance of IR, which had been used formerly, should be reconsidered.

7.1.2 Determination of Fluorine

The ion-selective electrode (ISE) is commonly used for determining fluorine in an aqueous solution. However, trace amounts of fluorine are difficult to detect by an ISE. In PAA, fluorine can be determined by the $^{19}F(\gamma,n)^{18}F$ reaction. In the case of high-level sodium contents in the sample, e.g. geological material, the irradiation energy should be reduced to 20 MeV because of the interfering nuclear reaction $^{23}Na(\gamma,\alpha n)^{18}F$. For the sensitive analysis of fluorine, pyrolysis and lanthanide precipitation of fluorine-18 after 20 MeV irradiation have been proposed.[64] The detectable concentration level was 0.06 $\mu g\,g^{-1}$ and rocks containing less than 10 $\mu g\,g^{-1}$ of fluorine were analyzed.

7.2 Multielement Analysis of Biological Materials

PAA of biological materials was reviewed by Kuttemperoor[65] and Sato.[66] By the development of high-resolution Ge detectors, PAA has been widely recognized as an efficient multielement determination technique of several biological materials, as reported by Kato et al.[67,68]

Biological samples, especially several organs of animals and blood, usually contain Na and P at high concentration levels. As the neutron activation cross-section of $^{23}Na(n,\gamma)^{24}Na$ is very large and the Compton scattering by γ-rays from ^{24}Na causes very high background in the γ-ray spectrum, ^{24}Na often masks other γ-rays emitted by minor and trace elements in NAA. Furthermore, since ^{32}P due to phosphorus produces high-energy β-ray emission followed by bremsstrahlung, detection of trace elements is frequently perturbed, especially in blood and bone samples.

In the case of PAA, high intensity radiation produced in biological samples is due to ^{11}C, ^{13}N and ^{15}O induced by the (γ,n) reaction from carbon, nitrogen and oxygen, respectively. Fortunately, γ-ray measurement can be carried out after a comparatively short cooling period, say about 3 h after irradiation, as the longest half-life of the aforementioned nuclides is 20 min (^{11}C). Even after 3 h, several short- to medium-lived radioisotopes such as ^{29}Al from Si, ^{38}K from K, ^{34m}Cl from Cl, ^{56}Mn from Fe and ^{87m}Sr from Sr can often be measured. Moreover, the carbon reaction can be suppressed by activating at an energy below its (γ,n) threshold (about 20 MeV). Many elements which can hardly be determined by NAA, such as Mg, Ca, Ni, Zn, Sr, Zr, Mo, I and Pb, are easily analyzed by PAA. In 1984, PAA of Zr in biological samples was performed without chemical separation.[69] The radioisotopes of medium half-life, such as ^{24}Na from Mg and ^{43}K and ^{47}Ca from Ca, sometimes show up as a source of undesired Compton background. To remove background activities, Yamashita and Suzuki[70] proposed a group separation scheme using complex formation of metals with tropolone-5-sulfonic acid and collecting on anion-exchange resin.

7.3 Analysis of Geological and Environmental Materials

PAA has been applied for the analysis of geological and environmental materials. About 20 to 30 elements can be determined instrumentally and simultaneously. The list of determinable elements covers major to trace

elements, essential to toxic elements, transition metals, heavy metals, alkali and alkaline-earth metals, halogens and others. These elements are Na, Mg, Si, Cl, K, Ca, Sc, Ti, Cr, Mn, Fe, Co, Ni, Zn, As, Rb, Sr, Y, Zr, Nb, Mo, Sb, I, Cs, Ba, Ce, Tl, Pb and U. The pioneering work of multielement analysis of rock and soil was done by Kato et al.[71,72] and Chattopadhyay and Jervis[73] respectively. PAA of atmospheric matter was reported by Aras et al.[74] and Kato et al.[75]

Geological and environmental samples contain various elements with high neutron capture cross-section, such as Na, Sc, Mn, Co, Br and Eu. In NAA, these elements induce high intense radioactivity even at trace levels. In PAA, radioactivities induced by major elements such as Na, Mg, Ca, Ti, Mn and Fe are relatively low in most cases. As the (γ,p) reactions are the most efficient reactions of Mg, Ca, Ti and Fe, the radioactivities induced by these elements can be reduced by lowering the irradiation energy well below 30 MeV.

Several selected elements advantageously detectable by PAA were also determined. Determination of Ni by a coincidence technique,[76] determination of Nb and Y for the characterization of standard rocks,[77] and determination of I, Tl and U in environmental materials[78] are typical examples.

Deep-sea sediment[79,80] and volcanic rocks[81] were analysed by PAA to discuss the sedimentation process and the distribution of minor elements during the formation of rock, respectively.

7.4 Multielement Analysis of Raw Materials and Industrial Products

Gijbels and Hertogen[82] reviwed PAA of ore and minerals. Albert[83] discussed trace analysis using PAA in a review article of nuclear methods. As mentioned before, NAA of certain matrices is sometimes very difficult, because of the self-shielding problem and high radioactivity caused. Typical examples are rare-earth elements, cadmium and boron, which also have large neutron capture cross-sections, and uranium ore and phosphate, which lead to fission products. To verify the capability of PAA in these fields, lanthanides in boron carbide were determined instrumentally by PAA.[84] Detection limits of these elements were found in the submicrogram range. Twenty-eight impurities in boron were also determined by PAA.[85]

8 COMPARISON WITH OTHER ACTIVATION ANALYSIS METHODS

In this section, the different features, advantages and disadvantages of the three activation analysis techniques PAA, CPAA and NAA are compared regarding the irradiation technique, nuclear reactions induced as well as sensitivity and selectivity achievable. Only the basic methods are considered; special techniques like nuclide-induced photodisintegration or 14 MeV NAA are excluded from discussion.

8.1 Irradiation

8.1.1 Photon Activation Analysis

The energy distribution of bremsstrahlung is continuous from zero to the incident electron energy. As high-energy photons have a high penetration power for most materials, a stack of several samples can be irradiated simultaneously. The matrix effect normally is very small in PAA. The energy distribution is not changed drastically during penetration. As the photon flux is decreased exponentially by scattering and absorption in the matter, the photon flux density on each sample should be monitored. Although the incident electron beam is focused, bremsstrahlung cannot be focused. As already touched on, bremsstrahlung has a nonisotropic spatial distribution which depends on the converter thickness and the geometry of the electron beam.

8.1.2 Charged Particle Activation Analysis

As the charged particles are strongly retarded in a sample by scattering, particles are stopped at a certain range. Both the induced radioactivity and the predominant reaction type induced change along with the depth since both flux and energy of the respective particle change. Therefore, range or stopping power of charged particles in each material should be taken into account for the quantitative analysis. It is difficult to irradiate a stack of samples. However, CPAA can efficiently be applied to analyze surface layers of materials. Hence, the surface contamination problem is more imminent than in NAA or PAA. Local heating of the materials by charged particle is a severe problem for the analysis of organic and other sensitive materials. Charged particles can be focused even on extremely small areas. Thus CPAA can be used favorably for the analysis of small spots, e.g. in the study of areal elemental distributions.

8.1.3 Neutron Activation Analysis

Since the neutron has no electrical charge it can penetrate the bulk of samples more easily. The absorption cross-sections of neutrons are strongly dependent on the nuclear properties of each element. The energy distribution of reactor neutrons shows Maxwell–Boltzmann distribution according to the temperature in the reactor. In most cases, many samples can be irradiated simultaneously

and homogeneously in a capsule. The self-shielding effect of certain elements, such as Cd, should be considered because the neutron flux is locally depressed in the presence of these elements. Neutron beams cannot be focused except, to a certain limited extent, with a neutron beam tube installed at the reactor.

8.2 Comparison of Nuclear Reactions Induced

8.2.1 Photon Activation Analysis

As mentioned above, photonuclear reactions are mainly controlled by the photoabsorption process (giant resonance, see above) giving rise to the (γ,n)-type reaction. The resonance energy ranges from about 10 MeV for heavy elements to about 20 MeV for light elements. The yield of the photoneutron reaction increases with atomic number. For analysis, (γ,n) and (γ,p) reactions are utilized normally. Higher-order reactions such as (γ,αn), (γ,np) or (γ,T) mostly appear as sources of interference. In practice, 30 MeV bremsstrahlung has proven to be optimal in most applications.

8.2.2 Charged Particle Activation Analysis

In CPAA, protons, deuterons, α and ^{3}He particles are commonly used as probes. In the case of charged particle activation, the Coulomb barrier effect should be considered. For heavier elements, higher excitation energies are needed to penetrate the coulomb barrier and to induce nuclear reactions. As several reaction types can be induced at this excitation energy many different radioisotopes are observed and the probability of nuclear interference is considerable. Since the higher atomic number elements are more likely to be subject to nuclear interference than light ones, CPAA is not normally suitable for the analysis of heavy elements.

In the case of deuteron and ^{3}He activation, many nuclear reactions are exothermic and can be induced at low excitation energy. Such nuclear reactions have been applied to the analysis of light elements. The irradiation particle and irradiation energy can be selected quasi-freely for the selective analysis of certain elements.

8.2.3 Neutron Activation Analysis

In NAA, thermal and epithermal neutrons are used. Neutrons are absorbed by elements according to the $1/v$ law and by resonance reactions. The thermal neutron capture cross-section is inversely proportional to the velocity of the neutron, and this is known as "the $1/v$ law". The resonance reactions induced in the epithermal region are often used for the analysis of specific elements using cadmium shielding of the sample because cadmium has very high absorption cross-section for thermal neutrons. The analysis of light element is very difficult. Unlike in PAA, there is no regularity in the neutron absorption reactions. Nonetheless, NAA can be used for multielement analysis of a large variety of materials.

8.3 Sensitivity and Selectivity of Elements to be Determined and Applicability to Various Kinds of Materials

8.3.1 Photon Activation Analysis

PAA is used for the analysis of light elements such as C, N, O and F, medium elements such as Ti, Ni, Sr, Zr, Nb, and Y, and heavy elements such as Pb, which are very difficult to analyze by NAA. Furthermore, PAA is suitable for the analysis of matrices which contain some elements of high neutron capture cross-section, e.g. boron and noble metals. As photonuclear reaction cross-sections systematically increase with atomic number, PAA can be applied to the analysis of heavier elements in light matrices, such as natural materials (rocks, soils, dusts, organics, etc.). The sensitivity for trace analyses is mostly sufficient (microgram per gram region and below), although not as good as in NAA for the average case. Using appropriate equipment and with the exception of a few interference cases (see above), PAA is highly selective due to discrete energy signals of the respective activation products.

8.3.2 Charged Particle Activation Analysis

CPAA is advantageous particularly for light element analysis. CPAA has often been applied to the impurity analysis of industrial materials. The shape of the sample is critical, so the application is somewhat restricted. The average sensitivity is comparable to that of PAA whilst the selectivity is frequently limited due to numerous nuclear reactions eventually induced in the individual element.

8.3.3 Neutron Activation Analysis

NAA can detect many elements instrumentally, except for light elements. Extremely low detection limits are achievable for a good number of elements, e.g. several rare earth elements, noble metals and others; the respective detection limits are lower partly by orders of magnitude, as compared with PAA and CPAA. Thus, NAA is suitable for sensitive analysis of trace elements in various types of materials. However, interference through eventual excessive matrix activities have to be taken into account. Moreover, in the normal case, only small sample masses should be irradiated because of the intense radioactivity induced. Thus, problems of representativeness might arise. As to the selectivity, see section 8.3.1.

9 CONCLUSION

PAA offers sufficient sensitivity for the analysis of a great number of elements covering the entire periodic table. A large variety of sample types can be analyzed. The quality of PAA results can be improved by application of quantitative methods, such as the internal standard method or the stable isotope dilution method. In the case of the classical comparative method, the use of synthetic multielement standards of acrylamide or silica matrix yields significant improvement of accuracy.

In PAA (also in NAA), samples can be measured repeatedly. A laboratory automation system using a small robot for sample changing[86,87] has released analysts from the tedious work of sample handling and data acquisition during γ-ray measurement. Moreover, many ''human'' sources of error can be avoided. Such systems also assist researchers of many other fields using activation analysis and other radioactivity measurements.

Although other, nonactive analytical methods underwent dramatic improvement in recent years concerning sensitivity and other features, PAA and other activation techniques can offer specific advantages, e.g. the analysis of elements difficult to analyse by other methods. Moreover, purely instrumental and even nondestructive analyses are state-of-the-art in PAA.

Thus, as trace characterization is still important in material science, environmental science and other fields, PAA can contribute to the analysis of light to heavy elements.

ABBREVIATIONS AND ACRONYMS

CPAA	Charged Particle Activation Analysis
CRM	Certified Reference Material
EC	Electron Capture
EXAFS	Extended X-ray Absorption Fine Structure
IR	Infrared
ISE	Ion-selective Electrode
LEPD	Low-energy Photon Detector
NAA	Neutron Activation Analysis
PAA	Photon Activation Analysis
PM	Photomultiplier
SRM	Standard Reference Material
XPS	X-ray Photoelectron Spectroscopy
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis

Nuclear Methods **(Volume 14)**
Charged Particle Activation Analysis • Elemental Analysis by Isotope Dilution • Instrumental Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
γ-Spectrometry, High-resolution, for Radionuclide Determination • Nuclear Detection Methods and Instrumentation

REFERENCES

1. Ch. Segebade, H.P. Weise, 'Comparison of Sensitivity Estimates for Low Energy Photon and Classical Gamma-ray Spectroscopy Applied to Photon Activation Analysis', *J. Radioanal. Chem.*, **45**, 209–220 (1978).
2. T. Kato, T. Sato, K. Mori, 'Analytical Sensitivities for Thirty-six Elements in Cellulose Matrix by Means of High-energy Photon Activation and Low-energy Photon Detection', *Anal. Sci.*, **6**, 103–106 (1990).
3. T. Sato, K. Mori, T. Kato, 'Self-absorption Correction for Low-energy Photons in Instrumental Photon Activation Analysis of Biological Materials', *Anal. Sci.*, **82**, 217–221 (1991).
4. J. Chadwick, M. Goldhaber, 'A Nuclear Photo-effect: Disintegration of the Diplon by γ-rays', *Nature*, **134**, 237 (1934).
5. R. Basile, J. Hure, P. Leveque, C. Schuhl, *C. R. Acad. Sci.*, **239**, 422 (1954).
6. C. Engelmann, 'Photon Activation Analysis', in *Advances in Activation Analysis*, eds. J.M.A. Lenihan, V.P. Guinn, Academic Press, New York, 1–55, 1972.
7. C. Segebade, H.-P. Weise, G.J. Lutz, *Photon Activation Analysis*, Walter de Gruyter, Berlin, 1988.
8. A.P. Kushelevsky, 'Photon Activation Analysis', in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, 219–237, 1990.
9. M.E. Toms, 'A Compilation of Photonuclear Reaction Products and Associated Gamma-energies', *J. Radioanal. Chem.*, **20**, 177–381 (1974).
10. P. Anderson, J.S. Hislop, D.R. Williams, 'High Resolution γ-spectra of 40–44 MeV γ Photon Activation Products. Part 1. The Elements Sodium to Molybdenum', AERE – R 7823, HMSO, London, 1974.
11. D.R. Williams, 'High Resolution γ-spectra of 40–44 MeV γ Photon Activation Products. Part 2. The Elements Ruthenium to Uranium', AERE – R 9021, HMSO, London, 1978.
12. G.J. Lutz, Ch. Segebade, 'Two-dimensional (Half-life vs. Energy) Map of Gamma-rays Emitted by Radionuclides Produced Through Photonuclear Reactions', *J. Radioanal. Chem.*, **33**, 303–337 (1976).
13. T. Kato, 'Quantitative Evaluation of the Photon-activation Method in Analysis', *J. Radioanal. Chem.*, **16**, 307–327 (1973).

14. H. Bethe, W. Heitler, 'On the Stopping of Fast Particle and on the Creation of Positive Electrons', *Proc. R. Soc. London*, **A134**, 83–112 (1934).
15. G.J. Lutz, 'Calculation Sensitivities in Photon Activation Analysis', *Anal. Chem.*, **41**, 424–427 (1969).
16. L.I. Schiff, 'Energy-angle Distribution of Thin Target Bremsstrahlung', *Phys. Rev.*, **83**, 252–253 (1951).
17. Y. Oka, T. Kato, K. Nomura, T. Saito, 'The Yields of Radioactivities Induced by (γ,n) Reactions with 20 MeV Bremsstrahlung', *Bull. Chem. Soc. Jpn.*, **40**, 575–579 (1967).
18. Y. Oka, T. Kato, K. Nomura, T. Saito, H.-T. Tsai, 'A Study on the Yields of (γ,p) Reactions with 20 MeV Bremsstrahlung', *Bull. Chem. Soc. Jpn.*, **41**, 380–384 (1968).
19. Y. Oka, T. Kato, K. Nomura, T. Saito, H.-T. Tsai, 'A Study on the Yields of (γ,α) Reactions with 20 MeV Bremsstrahlung', *Bull. Chem. Soc. Jpn.*, **41**, 2660–2662 (1968).
20. T. Kato, K. Masumoto, N. Sato, N. Suzuki, 'The Yields of Photonuclear Reactions for Multielement Photon Activation Analysis', *J. Radioanal. Chem.*, **32**, 51–70 (1976).
21. K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 567–577 (1978).
22. M. Barbier, 'Radioactivity Induced by High Energy Electrons and Photons', in *Induced Radioactivity*, North-Holland, Amsterdam, 214–249, 1969.
23. D.R. Williams, J.S. Hislop, 'High Resolution γ-spectra of 40–44 MeV γ Photon Activation Products. Part 3. Data Section', AERE – R 9022, HMSO, London, 1978.
24. E. Ricci, 'Sensitivities for Photon Activation Analysis with Thick-target, 110-MeV Electron Bremsstrahlung', *Anal. Chem.*, **46**, 615–618 (1974).
25. M.A. Kolmi'tsev, T.S. Ambardanashvili, V.Yu. Dundua, 'Comparison Standards for Neutron Activation Analysis on a Base of Phenol–Formaldehyde Resin', *J. Radioanal. Chem.*, **20**, 549–557 (1974).
26. D.H. Anderson, J.J. Murphy, W.W. White, 'Gelatin Multicomponent Trace Element Reference Material', *Anal. Chem.*, **48**, 116–117 (1976).
27. K. Masumoto, N. Suzuki, 'A New Trace Element Standard Reference Material with Polyacrylamide Gel Matrix for Radioactivation Analysis', *Radiochem. Radioanal. Lett.*, **42**, 99–106 (1980).
28. J.W. Mitchell, L.D. Blitzer, T.Y. Kometani, T. Gills, L. Clark, Jr, 'Homogeneously Doped Silica Matrices for Trace Element Standards in Neutron Activation Analysis', *J. Radioanal. Chem.*, **39**, 335–342 (1977).
29. A.R. Date, 'Preparation of Trace Element Reference Materials by a Co-precipitated Gel Technique', *Analyst*, **103**, 84–92 (1978).
30. M. Yagi, K. Kondo, 'A New Irradiation System for Photon Activation Analysis by Electron Accelerator', *J. Radioanal. Chem.*, **36**, 247–253 (1977).
31. G.A. Price, 'Energy Spectra and Angular Distributions of Photoneutrons from Heavy Nuclei', *Phys. Rev.*, **93**, 1279–1285 (1954).
32. K. Masumoto, K. Sakamoto, Y. Miyamoto, J.H. Zaid, A. Kajikawa, H. Haba, 'The Application of KURRI-Linac to Photon Activation Analysis of Environmental Materials', Abstract from the 40th Symposiumon Radiochemistry, Riken, Japan, 192–193, 1996.
33. Ch. Segebade, G.J. Lutz, H.P. Weise, 'Quantitative Evaluation of Interference Reactions in Photon Activation Analysis. The Production of Identical Radionuclides', *J. Radioanal. Chem.*, **39**, 179–184 (1977).
34. Y. Miyamoto, H. Haba, A. Kajikawa, K. Masumoto, T. Nakanishi, K. Sakamoto, 'Interferences in Neutron and Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **239**, 165–175 (1999).
35. H.-T. Tsai, K. Kato, Y. Oka, 'Photon Activation Analysis for Iodine by Using Cesium as the Internal-reference Element', *Bull. Chem. Soc. Jpn.*, **43**, 2482–2486 (1970).
36. M. Yagi, K. Masumoto, 'A New Internal Reference Method for Activation Analysis and its Application. Determination of Ti, Cr, Fe, Cu, Ga and Zr in Aluminum Alloys by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **84**, 369–380 (1984).
37. Y. Oka, T. Kato, H.-T. Tsai, K. Nomura, 'An Internal-monitor Method for Activation Analysis of Silver in Palladium with Thermal Neutrons', *Bull. Chem. Soc. Jpn.*, **41**, 329–333 (1968).
38. F. Girrardi, G. Guzzi, J. Pauly, *Anal. Chem.*, **37**, 1085–1092 (1965).
39. Ch. Segebade, M. Kuhl, B.F. Schmitt, R. Neider, 'Some Remarks on the State of Photon Activation Analysis and the Use of Internal Standards', *J. Radioanal. Chem.*, **72**, 665–696 (1982).
40. G. Laliaert, J. Hoste, Z. Eeckhaut, 'Neutron Activation Analysis Using an Internal Standard', *Nature*, **182**, 600 (1958).
41. G. Laliaert, J. Hoste, Z. Eeckhaut, 'Activation Analysis of Vanadium in High Alloy Steels Using Manganese as Internal Standard', *Anal. Chim. Acta*, **19**, 100–107 (1958).
42. M. Yagi, K. Masumoto, 'A New Internal Standard Method for Activation Analysis and its Application. Determination of Co, Ni, Rb and Sr in Pepperbush by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **84**, 316–331 (1984).
43. K. Masumoto, M. Yagi, 'Simultaneous Determination of P, Cl, K and Ca in Several Control Serums by Alpha-particle Activation Analysis Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **109**, 449–458 (1987).
44. M. Yagi, K. Masumoto, 'Instrumental Charged-particle Activation Analysis of Several Selected Elements in Biological Materials Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **111**, 359–369 (1987).

45. K. Masumoto, M. Yagi, 'Highly Accurate and Precise Multielement Determination of Environmental Samples by Means of Photon Activation Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **100**, 287–301 (1986).
46. M. Yagi, K. Masumoto, 'Instrumental Photon Activation Analysis of Environmental Materials Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **109**, 337–346 (1987).
47. K. Masumoto, M. Yagi, 'Instrumental Photon Activation Analysis of Soil Samples Using the Internal Standard Method Coupled with the Standard Addition Method', *J. Radioanal. Nucl. Chem.*, **116**, 375–388 (1987).
48. K. Masumoto, M. Yagi, 'Revaluation of the Internal Standard Method Coupled with the Standard Addition Method Applied to Soil Samples by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **121**, 131–139 (1988).
49. K. Masumoto, M. Yagi, 'Instrumental Photon Activation Analysis of Coal Fly Ashes Using the Internal Standard Method Coupled with the Standard Addition Method', *J. Radioanal. Nucl. Chem.*, **152**, 55–66 (1991).
50. M. Yagi, K. Masumoto, 'Simultaneous Determination of Ti, Cr, Fe, Cu, Ga and Zr in Aluminum Alloys by Charged Particle Activation Analysis Using Internal Standard', *J. Radioanal. Nucl. Chem.*, **91**, 379–387 (1985).
51. K. Masumoto, M. Yagi, 'Determination of Phosphorus in Low-alloy Steels by Alpha-activation Analysis', *J. Radioanal. Nucl. Chem.*, **130**, 243–250 (1989).
52. K. Masumoto, M. Hara, D. Hasegawa, E. Iino, M. Yagi, 'Photon and Proton Activation Analysis of Iron and Steel Standards Using the Internal Standard Method Coupled with the Standard Addition Method', *J. Radioanal. Nucl. Chem.*, **217**, 247–253 (1997).
53. K. Masumoto, 'Application of the Internal Standard Method Coupled with the Standard Addition Method in Photon Activation Analysis to the Trace Characterization of Environmental Materials', *J. Radioanal. Nucl. Chem.*, **203**, 37–46 (1996).
54. K. Masumoto, M. Yagi, 'Stable-isotope Dilution Activation Analysis and Determination of Ca, Zn and Ce by Means of Photon Activation', *J. Radioanal. Chem.*, **79**, 57–68 (1983).
55. M. Yagi, 'Stable Isotope Dilution Activation Analysis', in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, 131–143, Vol. 1, 1990.
56. M. Yagi, K. Masumoto, 'Isotope Dilution Activation Analysis for Special Samples in which the Self-shielding Effect is Negligible. Determination of Sr by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **90**, 91–103 (1985).
57. K. Masumoto, M. Yagi, 'Determination of Sr in Biological Materials by Charged Particle Activation Analysis Using Stable-isotope Dilution Method', *J. Radioanal. Nucl. Chem.*, **91**, 369–378 (1985).
58. K. Masumoto, M. Yagi, 'Simultaneous Determination of Ca, Rb, Sr and Ce in Environmental Materials by Photon Activation Analysis Using the Stable Isotope Dilution Method', *J. Radioanal. Nucl. Chem.*, **99**, 287–297 (1986).
59. C. Engelmann, G. Kraft, J. Pauwels, C. Vandecasteele, *Modern Methods for the Determination of Non-metals in Non-ferrous Metals*, Walter de Gruyter, Berlin, 1985.
60. M. Fedoroff, C. Loos-Neskovic, G. Revel, 'Radiochemical Separations in Molten Salts. Determination of Carbon in Metals by Photon Activation', *J. Radioanal. Chem.*, **55**, 219–232 (1980).
61. M. Fedoroff, C. Loos-Neskovic, J.C. Rouchaud, V.N. Samosyuk, B.A. Chapyzhnikov, 'Simultaneous Determination of Nitrogen and Carbon in Silicon by Photon Activation', *J. Radioanal. Chem.*, **88**, 45–49 (1985).
62. A. Yoshioka, K. Nomura, M. Takeya, K. Shimura, K. Masumoto, M. Yagi, 'Determination of Carbon in Copper by Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **122**, 175–182 (1988).
63. A. Yoshioka, K. Nomura, O. Kawakami, K. Shimura, K. Masumoto, M. Yagi, 'Determination of Carbon Impurity in Gallium Arsenide Crystals by Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **148**, 201–209 (1991).
64. H. Sakurai, M. Fukaya, Y. Hayashibe, Y. Sayama, K. Masumoto, T. Ohtsuki, 'Determination of Fluorine in Standard Rocks by Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **217**, 267–271 (1997).
65. V.Z. Kuttemperoor, 'Photon Activation Analysis of Biological Materials', in *CRC Handbook Series in Clinical Laboratory Science*, ed. D. Seligson, CRC Press, Boca Raton, 209–284, 1977.
66. T. Sato, 'Activation Analysis of Biological Materials', in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, 344–345, 1990.
67. T. Kato, N. Sato, N. Suzuki, 'Multielement Photon Activation Analysis of Biological Materials', *Anal. Chim. Acta*, **81**, 337–347 (1976).
68. N. Sato, T. Kato, N. Suzuki, 'Multielement Determination in Tobacco Leaves by Photon Activation Analysis', *J. Radioanal. Chem.*, **36**, 221–238 (1977).
69. T. Sato, 'The Determination of Zirconium in Biological Materials by Photon Activation Analysis', *J. Radioanal. Nucl. Chem. Lett.*, **86**, 141–150 (1984).
70. M. Yamashita, N. Suzuki, 'Photon Activation Analysis of Trace Metals in Biological Materials Via Collection of Metal Tropolone-5-sulfonate Complexes onto Anion-exchange Resin', *J. Radioanal. Chem.*, **60**, 73–85 (1980).
71. T. Kato, I. Morita, N. Sato, 'Non-destructive Photon-activation Analysis of Standard Rocks with 30 MeV Bremsstrahlung', *J. Radioanal. Chem.*, **18**, 97–110 (1973).

72. N. Sato, T. Kato, N. Suzuki, 'Multielement Photon-activation Analysis of Rocks Materials with 30 MeV Bremsstrahlung', *Radiochim. Acta*, **21**, 63–68 (1974).
73. A. Chattpadhyay, R.E. Jervis, 'Multielement Determination in Market-garden Soils by Instrumental Photon Activation Analysis', *Anal. Chem.*, **46**, 1630–1639 (1974).
74. N.K. Aras, W.H. Zoller, G.E. Gordon, G.J. Lutz, 'Instrumental Photon Activation Analysis of Atmospheric Particulate Material', *Anal. Chem.*, **45**, 1481 (1974).
75. T. Kato, N. Sato, N. Suzuki, 'Non-destructive Multielement Photon-activation Analysis of Environmental Materials', *Talanta*, **23**, 517–524 (1976).
76. K. Masumoto, N. Suzuki, 'Selective Coincidence Spectrometry in the Nondestructive Determination of Nickel in Geological Materials with High-energy Photon Activation', *J. Radioanal. Chem.*, **46**, 121–135 (1978).
77. T. Kato, K. Masumoto, 'Determination of Niobium and Yttrium in GSJ Basalt by Photon Activation Analysis', *Geostand Newslett.*, **5**, 167–170 (1981).
78. K. Masumoto, T. Ohtsuki, Y. Miyamoto, J.H. Zaidi, A. Kajikawa, H. Haba, K. Sakamoto, 'Photon Activation Analysis of Iodine, Thallium and Uranium in Environmental Materials', *J. Radioanal. Nucl. Chem.*, **239**, 495–500 (1999).
79. T. Kato, N. Sato, N. Suzuki, 'Multielement Analysis of Deep-sea Sediments by Photon Activation', *Bull. Chem. Soc. Jpn.*, **50**, 1930–1933 (1977).
80. T. Kato, M. Kato, 'Multielement Determination in Floor Sediments of the Japan Sea by Nondestructive Photon Activation Analysis', *J. Radioanal. Chem.*, **57**, 105–112 (1980).
81. T. Yoshida, K. Masumoto, K.-I. Aoki, 'Photon-activation Analysis of Standard Rocks Using an Automatic γ-ray Counting System with a Micro-robot', *J. Jpn. Assoc. Min. Pet. Econ. Geol.*, **81**, 406–412 (1986).
82. R.H. Gijbels, J. Hertogen, 'Activation Analysis of Ores and Minerals', *Pure Appl. Chem.*, **49**, 1555–1573 (1977).
83. Ph. Albert, 'Nuclear Methods in Trace and Ultra-trace Analysis: Past Experiences and Future Possibilities', *Pure Appl. Chem.*, **54**, 689–713 (1982).
84. Ch. Berthelot, G. Carraro, V. Verdingh, 'The Use of Photon Activation Analysis for the Determination of Sm, Eu, Gd and Dy in Boron Carbide', *J. Radioanal. Chem.*, **50**, 185–194 (1979).
85. T. Kato, K. Nagase, H. Yokobayashi, 'Determination of Trace Elements in Boron by Instrumental Photon Activation Analysis', *Anal. Sci.*, **1**, 307–311 (1985).
86. M. Yagi, K. Masumoto, M. Muto, 'An Automatic Gamma-ray Spectrometer Equipped with a Micro-robot for Sample Changing', *J. Radioanal. Nucl. Chem.*, **98**, 31–38 (1986).
87. K. Masumoto, S.-I. Ushino, M. Mutoh, M. Yagi, T. Yoshida, 'An Automatic Measurement System for Radiochemical Study Using Small Robots and PC-LAN', *Lab. Robotics Automat.*, **6**, 21–27 (1994).

PIXE (Particle-induced X-ray Emission)

P.A. Mandò
Dipartimento di Fisica dell'Università and INFN, Firenze, Italy

1 Introduction **1**
1.1 General Features 1
1.2 Brief Historical Overview and Review Literature 3
1.3 Comparison with other X-ray-based Analytical Techniques 3

2 In-depth Review of the Technique **5**
2.1 Basic Aspects of X-ray Emission from Atomic Shells 5
2.2 Quantitative PIXE Analysis 6
2.3 Minimum Detection Limits in PIXE Analysis 11
2.4 Ion Accelerators and Sample Irradiation 13
2.5 External Beam Set-ups 15
2.6 Detection of X-rays, Data Acquisition and Reduction 17
2.7 Micro-PIXE 21

3 Survey of the Main Analytical Applications **22**
3.1 Environmental Pollution Studies 22
3.2 Biomedical Applications 24
3.3 Applications to Earth Science 26
3.4 Applications to Art and Archaeology 27
3.5 Applications to Historical Problems 29

Abbreviations and Acronyms **30**

Related Articles **30**

References **30**

Particle-induced X-ray emission (PIXE) is the most popular among the ion beam analysis (IBA) techniques, which are based on the use of the specimen to be analyzed as a target for a beam of accelerated particles. The detection of the radiation induced by the beam bombardment is then used to discriminate and quantify the presence of the different elements in the specimen. In PIXE, what is exploited is in particular the X-rays emitted from the target, whose energies are characteristic of the emitting atomic species. After a general, simple description of the main features of PIXE, with a short historical overview and a brief comparison to other X-ray fluorescence (XRF) techniques, the article covers in greater depth all the specific aspects of this ''nuclear'' technique. The basic aspects of X-ray emission from the atoms are first recalled; then, the extraction of quantitative compositional data from thin and thick specimens is explained, and a discussion is given of the excellent performance of PIXE in terms of minimum detection limits (MDLs), and of the factors affecting them. A technical description then follows of how proper beams for PIXE are produced and of the experimental set-ups commonly used, with particular emphasis on the external-beam arrangements. The X-ray detector characteristics, the electronics for constructing the energy spectra, and the software processes for their deconvolution, leading to the extraction of quantitative data, are then briefly described. The last section surveys the main analytical applications of PIXE in various fields (environmental monitoring, biomedicine, earth sciences, cultural heritage), with no intent of exhaustiveness but rather with the purpose of focusing on when and why PIXE may be particularly suitable in each of them.

1 INTRODUCTION

1.1 General Features

PIXE is a technique for the elemental analysis of a sample. This is used as a target for the bombardment with a beam of accelerated particles; the interactions of the beam particles with the target atoms lead to the emission of X-rays of characteristic energies, through the detection of which the target composition can be deduced. The PIXE technique is often also referred to as *proton*-induced X-ray emission, owing to the fact that in this technique protons are almost universally chosen to induce X-ray emission. For this reason, the following discussion always considers the case of protons; no difference of either theoretical or practical relevance occurs anyway with beams of other particles. The PIXE technique belongs to the category of IBA techniques, and is perhaps the most widely used among them.

The method makes it possible to detect simultaneously and with great sensitivity almost the full range of elements present in the analyzed sample. In practice, some limitations partially restrict the number of detectable elements, but the multi-element capability is actually retained to a very large extent. As a rule of thumb (to be better specified below), one might state that in a single measurement – typically lasting few minutes – all elements with $Z > 10$ are detected with MDLs down to trace levels.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The working principle of PIXE is quite simple. Using a small (usually electrostatic) accelerator, a beam of particles is produced with an energy of typically 2–3 MeV. The beam particles, when impinging on a target, interact with atoms and nuclei of the target itself: in particular, the inner shells of the target atoms may be ionized. The inner shell vacancy is promptly filled (within a time interval of the order of 10^{-15} s) by an outer electron, and the difference $\Delta E = E_o - E_i$ of the electron binding energies E_o and E_i in the outer and inner shell is released. One of the processes through which the energy ΔE can be released is just the emission of an X-ray of that energy. A competing mechanism (the Auger effect) is the energy transfer to another electron of an outer shell, which is consequently ejected with a kinetic energy $E_k = \Delta E - E_j$, where E_j is the binding energy of the electron to which ΔE is transferred. Figure 1 shows schematically the processes of X-ray and Auger electron emission.

The energy of the emitted X-ray (or Auger electron) is clearly characteristic of the emitting atomic species, being the difference of electron binding energies that are indeed characteristic of the atomic species. Thus, the energy spectroscopy of the X-rays from the sample during particle beam bombardment provides an analysis of its composition. Such an analysis is *multi-elemental*, because all the atomic species present in the target undergo interactions with the beam particles, and therefore emit characteristic X-rays. It is also *quantitative*, because during an irradiation the number of interactions of the beam particles with the atoms of a given species is obviously proportional to the abundance of atoms of that species in the target. Under most circumstances (better specified in the following) the analysis is also totally *nondestructive*, because the beam particle–target atom interactions do not produce significant alterations in the target and involve only a tiny fraction of the target atoms. In any case, the possibility of further analysis (by the same or different techniques) is guaranteed for the same specimen.

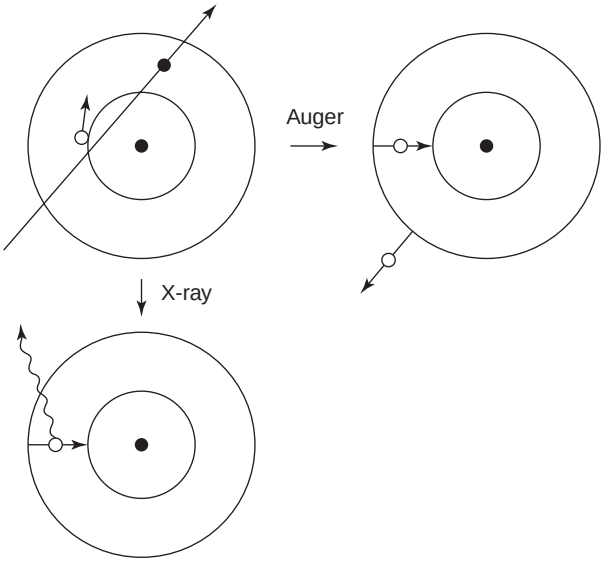

Figure 1 Diagram of an inner shell ionization induced by proton bombardment, followed by Auger electron or X-ray emission (alternative to each other). (Reproduced by permission from Johansson et al.[1])

The principle of operation of PIXE is the same as that exploited to perform elemental analysis through electron probe microanalysis (EPMA) or XRF, but with a different ionizing agent (protons rather than electrons or primary electromagnetic radiation, respectively). This difference has consequences in terms of important analytical parameters such as MDLs, range of detected elements, or sample depth probed (as briefly discussed later).

The possibility of performing PIXE while keeping the sample in the atmosphere (*nonvacuum* or *external* PIXE), and that of miniaturizing the beam size down to 1 μm or even less (micro-PIXE), make it a very versatile technique. All kinds of samples may be quantitatively analyzed: solid, gaseous, or liquid (the latter two with an external beam). Elemental mapping may also be performed by scanning the beam on the sample during analysis (or moving the sample under the beam).

In PIXE, the detection and discrimination of X-rays emitted under beam bombardment is usually performed through the use of solid state Si(Li) detectors (or, much less frequently, hyperpure Ge detectors), which discriminate the different X-rays by their different energies. The detector response signals are proportional to the energy releases within the active volume of the detector; this way of discriminating different X-rays is therefore called energy-dispersive spectroscopy (EDS). The energy resolution is generally adequate for unambiguous discrimination. Conventionally, resolution is given as the full width at half-maximum (fwhm) of the peak corresponding to the 5.9-keV X-ray from Mn. A typical value is around 160 eV, but even better resolution may be reached, especially for small-surface detectors and when a long time constant is used for the linear shaping of the detector signals. To give an idea of the kind of information collected, Figure 2 shows two typical X-ray spectra obtained with Si(Li) detectors in a PIXE measurement of a sample of aerosol. We return later to a discussion of the characteristics of the spectra (in particular the background shape) and to the way of treating partially overlapping lines. Here the intention is to point out that the multi-element analytical capability is a real feature of PIXE with EDS, and not just a theoretical claim.

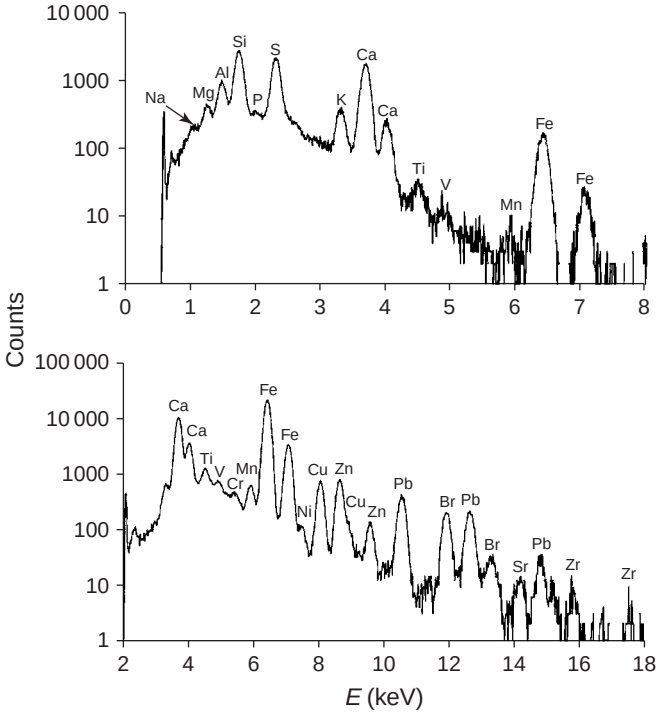

Figure 2 PIXE spectra from a sample of aerosol collected on a Millipore® filter. The spectra refer to the same sample, but have been recorded simultaneously by two Si(Li) detectors in different geometries (see Figure 11, and the discussion of two-detector set-ups in section 2.6.3). The spectrum acquisition time was about 7 min, at a proton beam current of 5 nA.

By using Bragg spectrometers that exploit X-ray diffraction from a crystal, wavelength-dispersive spectroscopy (WDS) may also be performed, which allows a much better energy resolution (by a factor of 50 or so). The WDS technique is clearly at a premium to solve close multiplets of X-rays of similar energies from different elements, but a serious drawback is a drastic drop of detection efficiency. In typical PIXE set-ups for analytical purposes, WDS is therefore never accomplished, because one of the great analytical pros of the technique, i.e. its high sensitivity, would be lost. After all, even with EDS most ambiguities may be solved by using other X-rays from the same element as markers of its presence. Thus, the use of WDS is left in general to studies of basic atomic physics, such as changes of electron binding energies as a function of chemical state, rather than to analytical applications. For this reason, we will not discuss WDS any further in this context.

1.2 Brief Historical Overview and Review Literature

The PIXE technique was established as a universal analytical technique by Johansson et al. in the early 1970s. Its popularity and the range of applications have constantly increased over the years. Today (1999), tens of laboratories throughout the world currently use PIXE for elemental analysis in fields as diverse as environmental pollution, biology and medicine, earth sciences, material sciences, art and archaeology, historical document examination, and forensic science. Comprehensive descriptions of the theoretical basis of the technique and its practical use in the main fields of application are available.[1,2] Dedicated conferences on ''Particle-induced X-ray Emission and its Analytical Applications'' have been held every third year since 1977; PIXE sections have been and are present in many other, more general, series conferences (e.g. International Conference on Ion Beam Analysis, European Conference on Accelerators Applied to Research and Technology, and others). Specific conferences on nuclear microprobes (where micro-PIXE applications are widely represented), and more recently on the applications in the field of biomedicine (bio-PIXE Symposia), are also periodically held.

The literature on PIXE work is distributed over several journals dealing with the fields of the specific applications. However, the above-mentioned Proceedings of the International PIXE Conferences – published by North Holland as separate issues of the review *Nuclear Instruments & Methods in Physics Research, Section B* – are a good source of concentrated information concerning both applications and basic physical and technological advancements. The journal *International Journal of PIXE* (published by World Scientific Publishing Co Pte Ltd, Singapore) contains topical and review articles on the various fields of application.

1.3 Comparison with other X-ray-based Analytical Techniques

1.3.1 Electron Probe Microanalysis

1.3.1.1 Minimum Detection Limits The absence of a primary bremsstrahlung (BS) continuum in the PIXE spectra (section 2.3.1), with the consequent low background in the energy spectra, is the main reason for much better (i.e. lower) detection limits in PIXE with respect to EPMA. In the latter, primary BS gives rise to a considerable, continuous background that extends to 20–30 keV (20–30 kV being the typical accelerating voltage used in the electron gun). This background is therefore present over all the energy region of interest for elemental analysis by X-ray line spectra. As X-ray production cross-sections for EPMA and PIXE are comparable under optimum conditions, the much lower backgrounds in PIXE spectra translate into better MDLs, typically lower by more than two orders of magnitude in PIXE with respect to EPMA.

1.3.1.2 Probed Depth The range of electrons typically used in EPMA ($E_e = 20–30$ keV) is of the order

of a few micrometers at most, in light-Z matrices. This is consequently also the investigated depth from surface with EPMA, for all elements. As discussed below, in typical PIXE ($E_p = 2 - 3$ MeV), the range traversed by protons is about 10 times larger; for higher-Z elements (detected through higher-energy X-rays) the probed depth is consequently much larger in PIXE than for EPMA. For the detection of lighter elements the effects of X-ray self-absorption (section 2.2) become dominant after only a few micrometers and the difference in the probed depth between the two techniques becomes smaller.

1.3.1.3 Space Lateral Resolution With an electron beam it is relatively standard to obtain beam spots on the target of less than 1 μm. Elemental mapping by scanning electron microscopy (SEM) is therefore easily performed at an EPMA facility. Standard PIXE is performed with larger beam cross-sections, but the possibility of micro-PIXE down to beam sizes below one micrometer exists, as briefly described in section 2.7. This makes it possible to perform trace-element two-dimensional mapping of a sample. However, it should be pointed out that EPMA yields excellent space resolution as a standard feature, whereas micro-PIXE requires sophisticated tuning and cannot therefore be considered a routine analytical technique.

1.3.1.4 Possibility of Nonvacuum Analysis With EPMA, it is mandatory to analyze the sample in vacuum. When large objects are investigated, this implies taking a sample of the material, which may be problematic when analyzing precious items. With PIXE, external beams are easily used (most practical aspects of the experiment get easier) which makes the analysis of large objects feasible with no need to take samples and no damage at all.

1.3.2 X-ray Fluorescence

1.3.2.1 Minimum Detection Limits Both PIXE and XRF can be used for trace-element analysis down to parts per million (ppm) levels. The difference between the two is the behavior of the sensitivity curve with Z. Generally speaking, due to the different ionization mechanisms (and the consequent trends of X-ray production cross-sections) XRF has lower MDLs for higher-Z elements than PIXE, whereas the analytical capability of PIXE extends to lower-Z elements. However, for both techniques the sensitivity curve can be varied by changing the energy of the ionizing agent (primary electromagnetic radiation and protons, respectively). The XRF method may become totally blind to higher-Z elements when choosing an exciting radiation suitable for obtaining lower MDLs for lower-Z elements, and vice-versa, whereas PIXE preserves the total range of detected elements even when optimized (by varying the proton energy) with a better sensitivity to lighter or heavier elements.

1.3.2.2 Probed Depth XRF probes considerably greater depths than standard energy PIXE as far as the heavier elements are concerned. For the lighter-Z elements (under conditions when XRF can detect them) it is again the self-absorption of low-energy X-rays – rather than the excitation mechanism – that determines the effective probed depth. Thus the difference is smaller.

1.3.2.3 Lateral Resolution The possibility of obtaining real microbeams for PIXE has been mentioned in section 1.3.1. As far as XRF is concerned, until recently the typical spot over which XRF analysis was carried was of the order of square centimeters. The analysis of much smaller spots required strong collimation, and this drastically decreased the exciting radiation intensity and therefore made measurement times unduly long. However, the recent development of focusing techniques for X-radiation (based on capillary light-guides suitable in the range of X-ray wavelengths) has opened up new possibilities of miniaturizing the exciting X-ray beams for XRF, while still preserving acceptable intensities. The useful spot size obtained with these techniques is now of the order of some tenths of a square millimeter.

1.3.2.4 Nonvacuum Analysis As with PIXE, XRF can be easily accomplished in nonvacuum conditions (although commercial XRF equipment is usually designed for in-vacuum analysis). In this respect there is no technical difference between the two techniques.

1.3.2.5 Nonlaboratory Analysis Portable XRF equipment has been developed, mostly based on the use of an X-ray tube as the source of electromagnetic exciting radiation. Although some limitations exist when performing fully quantitative analyses – especially for light elements – the opportunity to perform in situ analyses on non-transportable materials is considerable. The need for a particle accelerator to produce the beam that excites X-ray emission makes conventional PIXE an intrinsically nonportable technique. However, in the late 1990s portable PIXE equipment, based on the use of a purely α-emitting source (^{210}Po), has been developed and tested with success.[3,4] The present limitation of this apparatus is the low intensity of the α-beam due to isotropic emission of the particles from the source. This makes the measurements much more time-consuming than in the laboratory. Further developments in the relative source–sample–detector geometry are underway to improve the performance of this interesting variant to standard PIXE.

2 IN-DEPTH REVIEW OF THE TECHNIQUE

2.1 Basic Aspects of X-ray Emission from Atomic Shells

Let us first recall the spectroscopic nomenclature for atomic shells and for X-rays corresponding to electron transitions between them.

The potential energy is set to zero for free electrons, whereas electrons in the atomic shells have increasingly negative energies the more bound they are. The innermost electronic shell (principal quantum number $n = 1$) is the K shell, and corresponds to the most tightly bound electrons; therefore, for any given element, K electrons have the minimum energy value. Examples of K-level energies are −1072 eV for Na, −4038 eV for Ca, −7113 eV for Fe, −11 103 eV for Ge, −25 514 eV for Ag, −37 441 eV for Ba, and −88 005 eV for Pb.

The second shell (principal quantum number $n = 2$) is the L shell, with subshells L_1, L_2 and L_3 corresponding to $(l = 0,\ j = 1/2)$, $(l = 1,\ j = 1/2)$, $(l = 1,\ j = 3/2)$, respectively. Examples of level L_1, L_2, and L_3 energies are: −63.3, −31.1, and −31.0 eV for Na; −438, −350, and −346 eV for Ca; −846, −721, and −708 eV for Fe; −1414, −1248, and −1217 for Ge; −3806, −3524, and −3351 eV for Ag; −5988, −5624, and −5247 eV for Ba; and −15 860, −15 200, and −13 035 eV for Pb.

The sequence of other shells is then M, N, and O, with principal quantum numbers $n = 3$, 4, and 5, respectively, with less tightly bound electrons (this applies to higher-Z elements having a large number of electrons).

Once a vacancy is created in an inner shell (by any ionizing agent), not all electronic transitions from outer shells or subshells are possible. Transition selection rules arise from quantum mechanical laws, so that only certain transitions are permitted, each having a well defined probability.

In the case of ionization from the L or superior shells, the Coster–Kronig effect must be taken into account. For the L shell, this consists of a nonradiative transfer of a vacancy, primarily created in the L_1 or L_2 subshell, to a higher L subshell (L_2 or L_3 for primary L_1 ionization, L_3 for primary L_2 ionization). Coster–Kronig transitions occur before the filling from an outer electron takes place; the electron transition energy is then varied. The probabilities of L_1 to L_2 (f_{12}), L_1 to L_3 (f_{13}) and L_2 to L_3 (f_{23}) Coster–Kronig transitions are high (up to over 50% for the L_1 to L_3 transition), and vary with atomic number. Similar and even more complicated effects take place when the primary ionization occurs in a shell higher than L. Although the Coster–Kronig effect is important from the point of view of basic atomic physics, it has no practical consequence for quantitative analyses based on induced X-ray classification, such as PIXE, EPMA, and XRF. In fact, it can be treated as an effective change of ionization probability of the various subshells and automatically taken into account when performing quantitative PIXE measurements, for which the knowledge of ionization cross-sections is essential (section 2.2).

Once an ionization has been produced in an inner shell, deexcitation of the atom can be accomplished by X-ray or Auger electron emission. The ratio of the number of X-ray emissions to the total number of ionizations produced in the given shell is called the fluorescence yield relative to that shell and is usually indicated with w (w_K, w_L, etc.). For the L and higher shells, subshell ionizations and consequently subshell fluorescence yields can be considered separately.

Fluorescence yields may be calculated theoretically and measured experimentally.[5–8] They are a function of the atomic number Z and Figure 3 shows the behavior of w_K and w_L versus Z.

The nomenclature for X-rays emitted as a consequence of an outer-electron transition following inner-shell ionizations recalls both the primary vacancy location and the filling electron provenance. These are referred to as K-series, L-series, M-series, etc. X-rays, depending on the initially ionized shell. Within each series, the further specification is then related to the provenance shell of the electron. K_α and K_β X-rays correspond to the filling electron coming from the L and superior shells, respectively, with still further distinctions ($K_{\alpha1}$, $K_{\alpha2}$, $K_{\beta1}$, $K_{\beta2}$, $K_{\beta3}$) depending on the provenance subshells. Figure 4 gives a summary of the X-ray nomenclature for the main lines of the K and L series.

Because electron binding energies increase with Z, X-ray energies also increase with the atomic number. Figure 5 shows the trend for X-ray energies of the K and L main lines as a function of Z.

Within each X-ray series, the intensity ratio of the different transitions varies with Z.

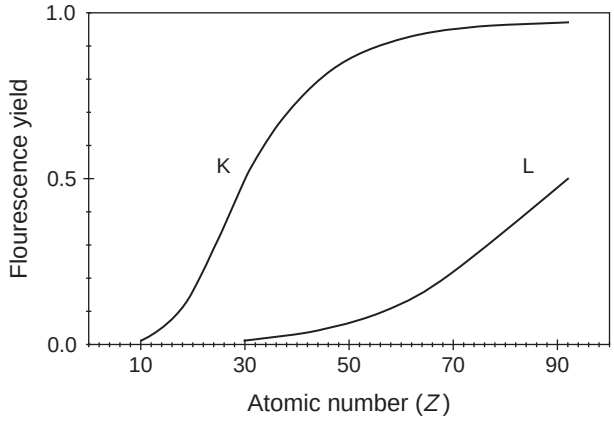

Figure 3 Trend of the K- and L-fluorescence yields versus Z. (Reproduced by permission from Johansson and Campbell.[2])

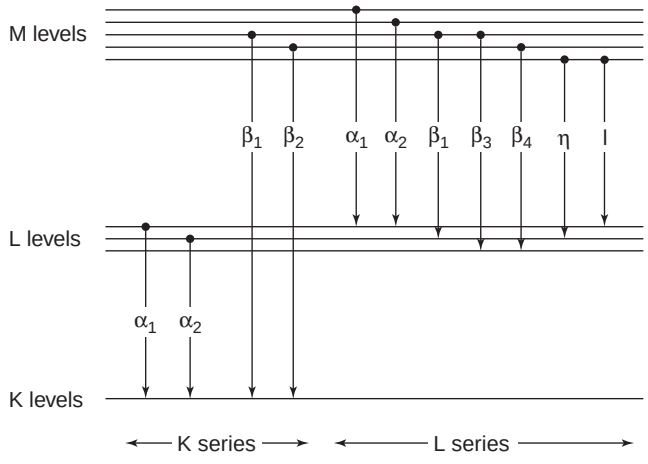

Figure 4 Summary of the main X-ray transitions to the K and L levels and of the corresponding nomenclature. (Reproduced by permission from Johansson et al.[1])

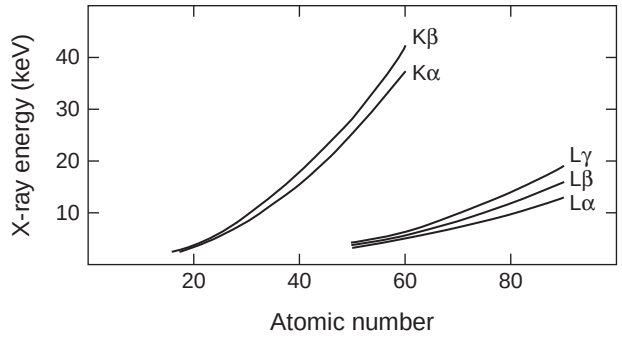

Figure 5 Energies of the main K- and L-series X-rays as a function of the element Z. (Reproduced by permission from Johansson et al.[1])

The ratio $I(K_\beta)/I(K_\alpha)$ of the two main groups (β and α) for K-series X-rays increases rapidly from 1.3% for Al (the lowest-Z element where K_β rays are possible) to 12% for K, then much more slowly (14% for Zn, 19% for Zr, 22% for Sn, etc.). The ratio $I(L_\beta)/I(L_\alpha)$ of the two main groups of the L-series X-rays decreases from about 70% in Sn to about 55% in the highest-Z elements; contrary to the K-series case, the $I(L_\beta)/I(L_\alpha)$ ratio also displays a dependence on the ionizing particle energy, but this effect is not very pronounced and is always known with sufficient accuracy for the purpose of spectra deconvolution.

2.2 Quantitative PIXE Analysis

Although a unified mathematical treatment of the procedure for a quantitative PIXE analysis might be given for all kinds of samples, it is easier to understand the principle by first examining the case of thin samples. An operational definition of "thin" and "thick" for a PIXE measurement is given below; for the moment, when referring to a thin specimen let us think of a few atomic layers.

2.2.1 Thin Specimens

The proton interactions with the inner shell electrons of the target atoms produce ionizations. The number of ionizations $N_{\mathrm{ion},j}(Z)$ produced in a given shell j ($j = \mathrm{K,L,M,\ldots}$) for a given atomic species Z in the target is obviously proportional to the target concentration of that atomic species, to the number of particles N_p having passed through the target and to the thickness t of the sample (Equation 1):

$$N_{\mathrm{ion},j}(Z) = \sigma_{\mathrm{ion},j} N_Z t N_\mathrm{p} \tag{1}$$

where N_Z is the number of atoms of atomic number Z per unit volume. The term $\sigma_{\mathrm{ion},j}$ is the *ionization cross-section* for the given shell j. As seen from Equation (1), it has the physical dimensions of an area, and is clearly a measure of the probability of the interaction giving rise to the specific ionization. Except for the case of the K shell, ionization cross-sections are made up of more terms, corresponding to ionizations of the various subshells, for instance σ_{L_1}, σ_{L_2}, and σ_{L_3}.

Measured values of K- and L- ionization cross-sections have been reported in a large number of papers; a much smaller amount of data is available for the M-shell case.

For the K shell, a collection of measured values is available[9] and a set of reference values based on statistical analysis of a compendium of published data has been compiled.[10] With the exception of very low atomic numbers, the results of the so-called ECPSSR (see below for a justification of the acronym) theory of the particle–electron interaction[11,12] agree within about 3% [13] with the best experimental data. This theory (developed from the pioneering work of Brandt and Lapicki[14,15]), treats the deflection and velocity change of the particles due to the Coulomb (C) field, the perturbation of the atomic stationary state (PSS) by the projectile, relativistic effects (R), and energy loss (E) during the collisions between beam particles and atomic electrons. For each element, the cross-section data (reference or theoretical) may be fitted as a function of proton energy E_p with a fifth-order logarithmic polynomial.[16,17] A universal treatment over a wide range of elements and proton energies with just six fitted parameters altogether, yielded – in the early stages of development of PIXE – a reasonable although not completely satisfactory reproduction of the tabulated data. However, using a distinct set of polynomial coefficients for each atomic number, tabulations can be reproduced within 1%.

The situation is more complex for the L shell, in which the ionization cross-sections differ and have a different dependence on the proton energy for the three subshells. As a consequence, the relative intensities between the various L lines also have a dependence on proton energy, which is not the case for K lines. Coster–Kronig transitions further complicate the overall picture; thus, comparison between experimental values (extracted for instance from the compilation of L X-ray production cross-sections in Cohen[18]) and theoretical predictions (such as the ECPSSR approach[19]) is more difficult and the agreement less satisfactory.

The trend of $\sigma_{\mathrm{ion,K}}$ and $\sigma_{\mathrm{ion,L}}$ with proton bombarding energy, taken from ECPSSR calculations for some elements, is shown in Figure 6. A strong decrease with increasing Z is apparent, especially for the K-shell ionization cross-sections. A strongly increasing trend with increasing E_{p} occurs within the range of proton bombarding energies usually adopted. The cross-section reaches a maximum when the projectile velocity matches that of the ejected electron. Further increase of projectile energy results in a slow decrease in cross-section. It is apparent from Figure 6 that the increase with E_{p} is larger for higher values of Z so that, in general, working at higher energies makes the analysis of heavier elements comparatively more sensitive.

For the M shell, much less systematic work has been done. However, although little use is normally made in PIXE of the detected M-series X-rays for the quantification of the high-Z elements, knowledge of M-subshell ionization cross-sections, and consequently of the intensity ratios of M lines, would be important.

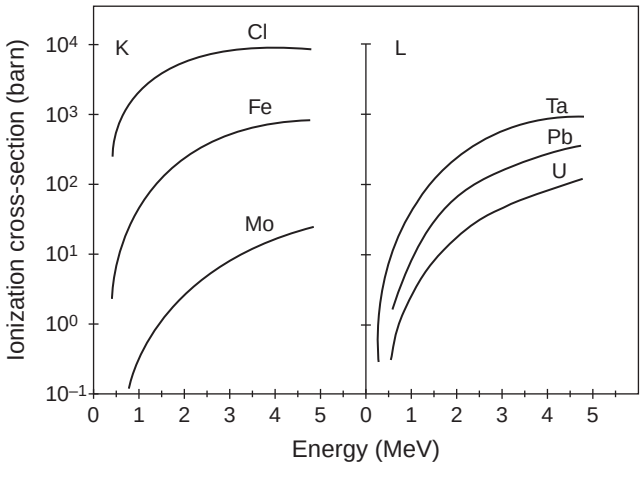

Figure 6 K- and total L-ionization cross-sections (in barn; 1 barn = 10^{-24} cm^2) for some elements versus proton bombarding energy. The displayed values are an ECPSSR calculation, which however is quite representative of the reference values from a compilation of the available experimental data. (Reproduced by permission from Johansson and Campbell.[2])

Indeed, in X-ray spectra from complex samples, the M lines from high-Z elements overlap the K or L lines of lighter elements so that a safe deconvolution of the multiplets must rely also on (experimental or theoretical) knowledge of the intensity ratios of the various M lines.

The product of ionization cross-section and fluorescence yield for a given shell gives the cross-section σ_{X_j} for the production of an X-ray of the corresponding series; from what has been said it follows that σ_{X_j} is also a function of Z and E_{p}. As stated above, within each series the branching ratio among the different lines is a slowly varying function of Z and, for the L shell, this also depends a little on the particle bombarding energy.

In terms of X-rays produced, the equivalent of Equation (1) is given by Equation (2):

$$N_{\mathrm{X}_j}(Z) = \sigma_{\mathrm{X}_j} N_Z t N_{\mathrm{p}} \tag{2}$$

The angular distribution in the emission of the produced X-rays is isotropic with respect to the beam incidence.

Equation (2) can be rewritten in terms of areal density $\rho_Z t$ of the element Z (mass per unit area normal to the beam, with ρ_Z standard volumic mass density) as:

$$N_{\mathrm{X}_j}(Z) = \sigma_{\mathrm{X}_j} \frac{N_{\mathrm{Av}} \rho_Z t}{A} \frac{Q}{e} \tag{3}$$

where N_{Av} is Avogadro's number and A the atomic mass of the species. In Equation (3), the number of protons N_{p} passing through the target is more conveniently expressed by integrated beam charge Q divided by unit charge e of the particles. Indeed, Q is usually a measurable quantity (see section 2.4.3).

The capability of performing a fully quantitative analysis with PIXE in essence derives from Equation (3).

In a thin target, for any given element the X-ray production cross-sections are constant throughout the target (because the beam energy loss is negligible, i.e. all interactions happen at the same proton energy, say E_0), and the number of detected X-rays is a well-defined fraction of those produced (isotropic emission of all X-rays). From Equation (3), one obtains the number Y_0 of X-rays from the element Z, detected during a measurement, simply by taking the overall detection efficiency into account. The latter is the product of a geometrical factor (related to the solid angle Ω covered by the detector) and of an intrinsic efficiency $\varepsilon_{\mathrm{det}}$, a function of the X-ray energy (and therefore of Z) depending on the detector characteristics (section 2.6.1):

$$Y_{0\mathrm{X}_j}(Z) = \sigma_{\mathrm{X}_j}(Z, E_0) \frac{\Omega}{4\pi} \varepsilon_{\mathrm{det}} \frac{N_{\mathrm{Av}} \rho_Z t}{A} \frac{Q}{e} \tag{4}$$

Equation (4) can be reorganized as follows (Equation 5):

$$Y_{0X_j}(Z) = \left(\sigma_{X_j}(Z, E_0)\frac{\Omega}{4\pi}\varepsilon_{\text{det}}\frac{N_{\text{Av}}}{Ae}\right) Q\rho_Z t \qquad (5)$$

or (Equation 6):

$$Y_{0X_j}(Z) = \eta Q \rho_Z t \qquad (6)$$

by defining an efficiency factor η for the given X-ray, beam energy, detector and geometry (Equation 7):

$$\eta = \eta(Z) = \sigma_{X_j}(Z, E_0)\frac{\Omega}{4\pi}\varepsilon_{\text{det}}\frac{N_{\text{Av}}}{Ae} \qquad (7)$$

From Equation (6) it is easily seen that – *provided* η *is known* – the areal density $\rho_Z t$ for a given element can be obtained by measuring the corresponding X-ray yield on the detector and the integrated beam charge on the target. Knowledge of η is in principle possible even in absolute terms, because the physical factors in Equation (7) are either known (σ, see above) or measurable (Ω, ε_{det}, see section 2.6). In practice, rather than relying on the absolute knowledge of the terms in Equation (7) for determining the efficiency factors $\eta(Z)$, the latter are obtained experimentally through Equation (5) itself. This is performed by bombarding a series of thin standards containing known quantities $\rho_Z t$ of the various elements, and measuring the corresponding X-ray yields and the integrated beam charge. The measurements on the standards must be performed using exactly the same geometry, with the same absorption layers (if any) and at the same beam energy as on the samples. The analytical procedure for the determination of elemental concentrations in unknown thin samples is therefore a relative procedure.

Figures 7 and 8 show the curves for the efficiency factor as a function of X-ray energy (i.e. Z), for the K and L series, respectively, in a typical detector geometry. As a result of the very large variation of X-ray production cross-section over Z, the efficiency factor in a given geometry also spans a wide range of values, drastically decreasing for the detection of X-rays of higher-Z elements within the same series. In addition, the intrinsic efficiency of Si(Li) detectors progressively drops at X-ray energies above 20–25 keV, which corresponds to K-series X-rays of elements around $Z = 50$. This would create a loss of sensitivity to the detection of heavier elements; the problem is to a large extent overcome by the fact that L-series X-rays (produced with much larger cross-sections, see Figure 6) can be used as markers of their presence. At these values of Z the L-series X-rays are well within the useful range for an Si(Li) detector

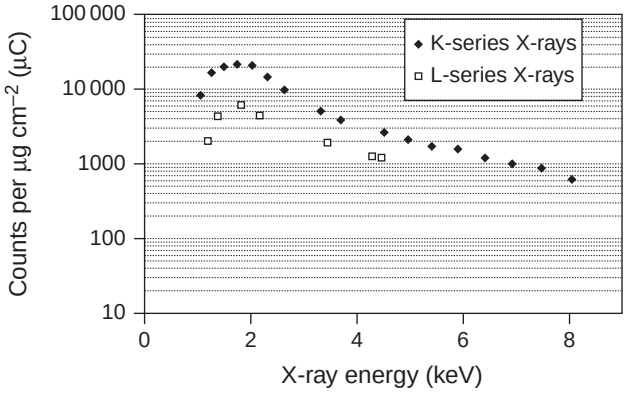

Figure 7 Trend of efficiency factors versus X-ray energies, for K-series and L-series X-rays, with an Si(Li) detector covering a solid angle of 2.5 msr (~12.5 mm^2 active area, ~7 cm distance) and no explicit absorbers. Experimental data obtained with thin standards certified to ±5% accuracy.

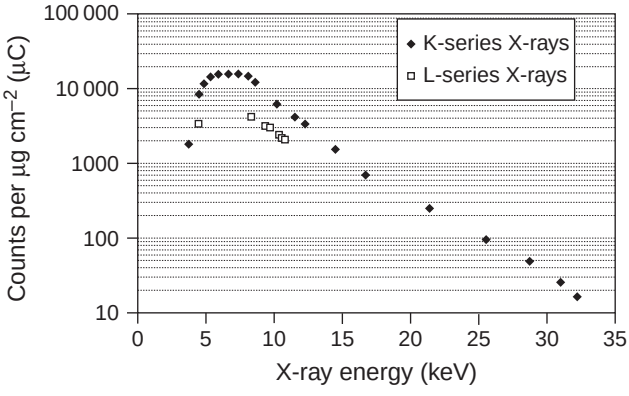

Figure 8 Same as Figure 7, but for a larger solid angle Si(Li) detector (~70 msr) having a 400-µm Mylar® absorber in front.

(see section 2.6.1), and this holds true for L-series X-rays up to the highest Z values. It will also be seen that in the higher-energy region of the spectrum the background is much lower, thus improving detection of weak lines. Further expedients to make detection limits more homogeneous through the widest possible range of Z are discussed in section 2.6.3.

2.2.2 Nonthin Specimens

Before seeing how quantitative analysis may be performed in the general case, it is useful to consider more carefully what "thin" and "nonthin" mean for PIXE analysis. The two physical effects that concur to determine whether a specimen is nonthin as far as PIXE analysis is concerned are:

- the beam energy loss while penetrating the sample;
- the self-absorption of X-rays produced in the interior of the sample.

A sample is thin if one may neglect these effects (the so-called matrix effects) and assume therefore:

- that all interactions happen at the same beam energy;
- that X-rays produced by the beam inside the target, emitted within the acceptance solid angle of the detector, have negligible probability of being lost to detection due to interactions inside the sample material itself.

By the expression ''thin sample'', we mean that the quantitative results obtained from its PIXE analysis by neglecting matrix effects are correct. Of course, this is never absolutely true and the validity of the thin-target approximation depends on the degree of accuracy one is looking for. Another important fact to bear in mind is that the thickness below which a sample may be considered thin depends both on its composition and on the X-ray energy that is observed (i.e. on the element, the abundance of which one is quantifying). Indeed, both beam stopping power and X-ray attenuation due to self-absorption depend on the average Z of the material; X-ray attenuation also strongly depends on the X-ray energy, increasing dramatically with decreasing X-ray energy. As a result of the latter point, a given physical thickness may be still considered ''thin'' as far as detection of a higher-energy X-ray is concerned, whereas it is already ''nonthin'' for the detection of a lower-energy X-ray from a lighter element. Approximately below 10–20 $\mu g\,cm^{-2}$ a sample of a medium-Z matrix can be assumed to be thin (to about 1% or better) for the quantification of all elements usually detected in PIXE (X-ray energies above 1 keV, corresponding to Na K_α). For the detection of higher-Z elements, detected through X-rays in the energy range above 6–7 keV, even a thickness of 2–300 $\mu g\,cm^{-2}$ might be considered thin.

When the sample thickness is such that the consequences of beam energy loss and X-ray absorption cannot be neglected, the extraction of quantitative information from the X-ray intensities in PIXE spectra is more complicated, but can still be dealt with. Indeed, both the physics of beam particle–target atom collisions, producing the energy loss, and that of X-ray interactions, producing the self-absorption effect, are well known and can be properly taken into account. The X-ray yield is obtained by integration of the contributions from the different depths inside the target. For each of them, one has to consider the proper X-ray production cross-section (which is a function of E_p and therefore of the depth from surface) and the reduction in the detected X-ray intensity, which also depends on the depth at which they are produced, because those X-rays produced deeper in the sample are more affected by self-absorption.

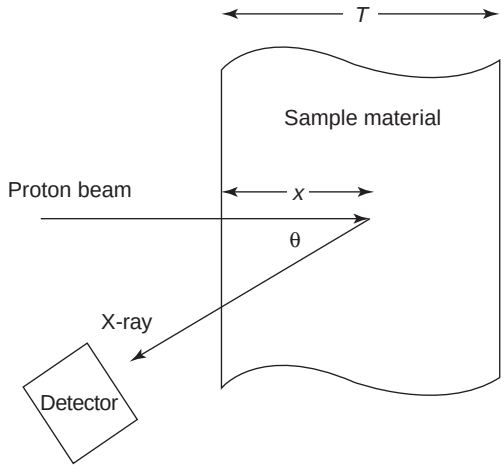

Figure 9 Diagram of a thick target analysis (see Equation 8).

The detected X-ray yield for a given element is obtained by the following generalization of Equation (4):

$$Y_{X_j}(Z) = \frac{\Omega}{4\pi}\varepsilon_{det}\frac{Q}{e}\frac{N_{Av}}{A}\rho_Z \int_0^T \sigma_{X_j}(Z, E_p)e^{-\frac{\mu x(E_p)}{\cos\vartheta}}\,dx \tag{8}$$

In Equation (8) T is the finite thickness of the sample and the exponential term accounts for the self-absorption of X-rays produced at a general depth x from the surface. The X-ray intensity attenuation coefficient μ is a function of the matrix composition and of the X-ray energy. The usual detection geometry (Figure 9) has been assumed in Equation (8), where ϑ is the angle at which the detector is placed with respect to the target normal.

In Equation (8), one should consider that x and E_p can each be expressed as a function of the other through the beam mass stopping power $S = dE_p/\rho dx$ (a function of the matrix composition and of E_p). Equation (8) can therefore be expressed as Equation (9),

$$Y_{X_j}(Z) = \frac{\Omega}{4\pi}\varepsilon_{det}\frac{Q}{e}\frac{N_{Av}}{A}\frac{\rho_Z}{\rho}\int_{E_0}^{E_0-\Delta E}\sigma_{X_j}(Z, E_p) \times e^{-\frac{\mu x(E_p)}{\cos\vartheta}}\frac{dE_p}{S(E_p)} \tag{9}$$

where E_0 is the initial beam energy (at target entrance) and ΔE is the total beam energy loss through the target thickness T (Equation 10):

$$\Delta E = \int_0^T S\,dx \tag{10}$$

In general, for any X-ray, a matrix correction factor F_{X_j} is introduced, defined as the ratio between the ideal yield Y_0 that would be obtained from the same target, in the absence of matrix effects, and the actual yield Y

(Equation 11):

$$F_{X_j}(Z) = \frac{Y_{0X_j}(Z)}{Y_{X_j}(Z)} \tag{11}$$

Therefore, from Equations (4) and (8), one obtains

$$F_{X_j}(Z) = \frac{T\sigma_{X_j}(Z, E_0)}{\int_0^T \sigma_{X_j}(Z, E_p(x))e^{-\frac{\mu x}{\cos\vartheta}}dx} \tag{12}$$

By recalling Equation (6) for Y_0, Equation (11) implies (Equation 13):

$$(\rho t)_Z = \frac{Y_{X_j}(Z)}{Q\eta_{Z,E_0}}F_{X_j}(Z) \tag{13}$$

where η is the efficiency factor, resulting from the measurements on thin standards at the surface beam energy E_0. In principle, therefore, the quantitative analysis of nonthin samples can be obtained from the detected yields $Y_{X_j}(Z)$, by considering the samples as thin (use of thin target efficiency factors) and then applying the proper correction factors F to account for matrix effects.

The problem remains as to how to obtain proper correction factors. This is a priori a problem, but can be solved quite well in many cases, such as when analyzing a homogeneous target of known thickness. The equivalent to the exact knowledge of target thickness T is a much more frequent case than might be thought at first sight. In fact, whenever the thickness, although unknown, is larger than the proton range R_p (of the order of 50–100 $\mu g\,cm^{-2}$), for the calculation of F it can be replaced by R_p itself, or by any larger value. In this case one talks about "infinitely thick" targets, where of course this term only implies that $T > R_p$. Even before this limit is reached, at least for the lower-Z elements, the exponential term in Equations (8) and (12) may be so small after a certain depth that the expression under the integral goes to zero and contributions from deeper layers can be neglected. In simple terms, for the lower-Z elements the dominant matrix effect is self-absorption rather than the decrease in production cross-section due to the beam stopping down; their X-rays continue to be produced at depths from which they will not reach the detector due to the self-absorption effect. Actually, as far as the detection of lower-Z elements is concerned, the target becomes infinite much before the condition $T > R_p$ is fulfilled.

It is also necessary to make an assumption about the matrix composition to evaluate the integral in Equation (12), because both μ and S are matrix dependent. If this composition is not known, as is usually the case in analytical work, the adopted technique then becomes an iterative procedure. The zeroth-order approximation of the matrix composition is obtained by neglecting all the matrix effects, i.e. by using Equation (13), with $F(Z) = 1$ for all values of Z. The matrix composition obtained by the areal densities thus deduced for the detected elements is then used to compute the stopping power and absorption coefficients to be used in Equation (12). From that, a first-order approximation of the matrix composition is obtained by reworking Equation (13), and this composition is then used again to compute the stopping power and absorption coefficients to be used in Equation (12), and so on iteratively. The procedure rapidly converges and a good estimate of the actual composition can be given. Two aspects must be pointed out in this respect.

First, the assumption is made of a homogeneous composition as a function of depth through the explored depth range (recall that we are talking about tens of micrometers at most). However, this assumption may be completely wrong, and the only way to check its validity is to perform further independent analyses with other techniques (or even simultaneously with other IBA techniques, or performing subsequent PIXE runs at different proton bombarding energies).

Second, the various approximations of the matrix composition, if based on just the PIXE-detected elements, can be largely defective because very light elements are hidden from the analysis (typically, this happens for $Z < 11$, but the actual limit depends on the specific set-up). The undetected fraction can be important and affects the actual matrix effects, which depend on all the elements present in the material. In many practically important circumstances the undetected fraction may be known independently, on the basis of separate measurements such as simultaneous measurements with other IBA techniques. In other cases, it can be indirectly deduced by the concentrations of the detected fraction itself; this is the case for instance in the analysis of rocks of volcanic origin. Here all elements can be considered to be present in the form of oxides, so that each element bears an amount of oxygen in a well-defined proportion to the detected element's abundance. Under these circumstances, the undetected fraction is well estimated and a good approximation of the whole matrix composition can be used to compute the factors F from Equation (12). Finally, it should be remembered that both the stopping power and the absorption coefficients depend on the matrix composition, but not so dramatically as to alter significantly the final values obtained by the iterative procedure of Equation (12) if a minor change in the matrix composition assumption is made. Obviously, this affects the different elements in a different manner: as a rule of thumb, lighter elements are more affected by incorrect assumptions about matrix composition. As for any other quantitative analytical technique, it must be

kept in mind here that the more unverified assumptions are made, the less accurate may be the final quantitative result.

If the sample thickness is larger than about 50 μm, the uncertainty about its value is irrelevant, because from the point of view of PIXE this is actually an infinitely thick target for the quantification of all elements. However, for a target of an intermediate unknown thickness, the quantitative analysis becomes problematic because, in principle, there may be different combinations of composition and thickness leading to the same detected X-ray yields. Some independent guess often may be made on either of the two, thus leading to reasonable results also in this case; however, it must be stressed that this remains the hardest situation to deal with for an accurate quantitative analysis.

2.3 Minimum Detection Limits in PIXE Analysis

A high value of the efficiency factor η for a given element, leading to large statistics for the corresponding X-ray peak in the PIXE spectrum (Equation 6), is obviously important for a sensitive determination of that element. However, this is not the only requisite, as the MDL depends on the signal-to-noise ratio.

In general, the performance in terms of the MDLs improves for any analytical method when the signal is detected with higher efficiency and when the noise is low. Because the cross-sections for the production of X-rays by accelerated particles are high, the number of X-rays produced by PIXE per unit quantity of a given atomic species is large. This means that the signal is good. The noise from which the signal must emerge is, in the case of PIXE (as well as in any other spectroscopic technique), the continuous background in the spectra. As can be seen in Figure 2, a continuous background is actually present in the X-ray energy spectra from a PIXE measurement, and characteristic X-ray peaks must be discriminated from the statistical fluctuations of this background. Let us therefore examine in some detail the reasons for the presence of this continuous background and the possibilities for the operator to reduce it.

2.3.1 Background in PIXE Spectra

There are several physical reasons for the presence of the continuous background in the PIXE spectra.

In the lower-energy part the continuous spectrum is higher and arises mainly from atomic electrons ejected by the interactions of the beam particles with the target. During their stopping down within the target, these electrons emit BS radiation. This is a consequence of a basic law of electromagnetism, that when a particle bearing electric charge undergoes an acceleration, it emits energy in the form of electromagnetic radiation. The intensity of the emitted BS is proportional to the square of the acceleration; its energy has a continuous spectrum extending up to the initial energy of the emitting particle. The electron BS in the lower part of the PIXE spectrum is called secondary BS, as opposed to the primary BS produced directly by the protons in their stopping down. It is easy to understand that:

- In PIXE, secondary BS has a much higher intensity than primary BS, which is essentially negligible. The forces responsible for both proton and secondary electron acceleration have equal intensity (being the Coulomb forces of interaction with atomic electrons) but the mass of protons is much larger than the mass of electrons. The proton acceleration and the corresponding BS intensity are consequently much lower.
- The maximum energy of the secondary BS is in the ratio $4m_e : M$ to that of the incident particles (with m_e and M being the electron and proton mass, respectively), this being the maximum energy transfer from incident particles to free electrons. Indeed, the maximum velocity acquired by the electrons, occurring for central collisions, is equal to twice that of the incident particles, so that Equation (14) holds:

$$E_{\max,e} = \frac{1}{2} m_e 4 v_p^2 = 4 \frac{m_e}{M} \frac{1}{2} M v_p^2 = \quad 4 \frac{m_e}{M} \quad E_p \qquad (14)$$

In the case of protons, the maximum energy value of the secondary BS is therefore around 4.3 and 6.5 keV for $E_p = 2$ MeV and 3 MeV, respectively. Consequently, under typical measuring conditions the background due to BS is practically zero in the higher part of the energy spectrum, i.e. the region of the X-ray lines used for the quantification of medium- and high-Z elements. This is just the region where X-rays are produced with lower cross-sections by PIXE. Therefore, where the signal is lower, the noise is fortunately lower also, and good MDLs are preserved. Many studies (both theoretical and experimental) have been carried out on the physics leading to the BS emission in PIXE.[20,21] An important practical feature is that BS emission is not isotropic, but has a maximum at around 90° to the beam direction. Therefore, to improve the performance in terms of MDLs, one should avoid placing the X-ray detector at around this angle. The most frequently used geometry is with the detector as backwards as possible to the beam direction, typically at 135° or more.

Some background continuum exists even in the higher energy part of the PIXE spectrum. It originates mainly from the Compton interactions of the higher-energy gamma rays produced by the incident particles either

in the target or in other materials before hitting the target (collimators, exit window in the external beam set-up, etc.). Compton interactions produce a partial energy release of the electromagnetic radiation in the material where they occur; the fraction of the full gamma energy that is released depends on the scattering angle in the Compton interaction. The resulting spectrum is therefore continuous and practically flat throughout the energy region of interest for PIXE, i.e. 1–30 keV.

The subject of gamma ray production in a PIXE experiment has been covered in detail.[22] The intensity of Compton background is generally low, and depends on:

- The detector material and active volume. It increases with increasing Z and this is one reason why silicon detectors are in general preferred to germanium ones. As far as the detector's volume is concerned, in the range of detector sizes that are used, larger ones produce a higher Compton background because of an increased probability of gamma ray interactions within their volume.
- The composition of the target itself and of the other materials with which the beam particles can interact. For low-Z materials, nuclear interactions with the beam protons are much more probable due to the closer approaches of the protons to the target nuclei, which in turn derives from the lower Coulomb repulsion. At the proton bombarding energies used in PIXE, most nuclei of low-Z elements can be excited inelastically and emit as a consequence gamma rays, which produce the Compton background in the PIXE spectra. The intensity of such emissions depends on the element, but also on the thickness of the emitting material and beam energy, due to the existence of resonances in the nuclear interaction cross-sections. Under typical conditions, fluorine and sodium are the elements giving rise to the higher gamma ray yield.

The Compton background arising from the production of gamma rays of elements within the target is unavoidable and the only expedient to lower it is to avoid the presence of unnecessary material around the detector, to decrease the probability of Compton scattering of the gamma rays in the surrounding materials. More can be done instead to decrease the contribution due to gamma rays originating from primary beam interactions on materials other than the target itself. First, low-Z materials such as aluminum should be avoided as constituents of chambers, collimators, scatterers, exit windows (in external beam set-ups), etc. Second, the points along the beam path that may constitute possible gamma ray sources arising from beam interactions should be effectively shielded towards the X-ray detector.

2.3.2 X-ray Line Interference

In addition to the presence of the continuous background, there is another factor that may worsen – in specific circumstances only – the MDL of some elements. This problem is common to all analytical techniques based on X-ray spectrometry (i.e. also XRF and EPMA) and it is recalled here just for the sake of completeness. This limiting factor consists of energy overlaps between X-rays originating from different elements. When the energy difference between two X-rays is much smaller than the detector resolution, they can hardly be discriminated by the spectra deconvolution codes. If one of the two elements giving rise to the overlapping X-rays is present in much larger abundance than the other, it can mask the presence of the weaker one even in the case of somewhat larger energy differences, within the tails of the strong peak. Thus, the possibility of detection of the weaker line is considerably worsened. Examples of this problem are:

- K_β lines of element Z with K_α of element $(Z+1)$, for $20 < Z < 30$;
- K line of Na (1.041 keV) with L lines of Zn (1.020 keV);
- K line of Mg (1.254 keV) with L lines of As (1.282 keV);
- K_α line of S (2.308 keV) with M lines of Pb (centroid at 2.346 keV) and to a lesser extent with those of Hg (centroid at 2.224 keV);
- K_α line of Ti (4.509 keV) with L_α lines of Ba (centroid at 4.465 keV);
- K_α line of As (10.532 keV) with L_α lines of Pb (centroid at 10.551 keV).

When the interference between two lines cannot be safely deconvoluted, the quantification of the two elements must be done on the basis of other lines of their spectra. This often implies that minor-intensity X-ray lines come into play and the MDL for the corresponding element increases.

2.3.3 Minimum Detection Limits

In the light of all the above considerations concerning high X-ray production cross-sections and detection efficiencies, and low continuous background in the energy spectra, it turns out that PIXE is basically a very sensitive technique. Figure 10 gives an idea of the MDLs of PIXE under typical analytical conditions. It should be emphasized that this is just a general indication, and that in any given analytical situation the actual MDL for a given element may be different. The MDL can depend on specific features of the sample under analysis, such as whether it is thinner or thicker, or the specific matrix material. Experimental conditions also have an

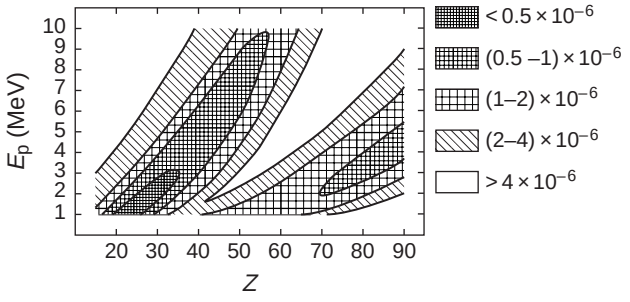

Figure 10 Attainable MDL as a function of proton bombarding energy and atomic number, in typical conditions. (Reproduced by permission from Johansson and Johansson.[17])

effect: beam energy, beam current intensity, and total time of target exposure to the beam all affect the analytical performance in terms of the MDL. Under certain circumstances, the attainable MDL can be much better than suggested by Figure 10. An example is when searching for ultratrace elements of medium Z in a very low-Z matrix. The opposite occurs in the presence of a high abundance of elements that give rise to X-rays interfering with those of the elements searched for.

When dealing with MDLs, both relative (in terms of minimum detectable mass over the total mass of the sample) and absolute values should be considered. Indeed another great quality of PIXE is that absolute quantities as small as picograms of the elements in the detectable range can be quantified in measurements lasting only a few minutes. This makes PIXE a very useful technique when nondestructivity is a mandatory requirement (as in the case of precious items such as works of art), or when the quantity of material to be analyzed is very small (as in some applications related to environmental pollutants).

2.4 Ion Accelerators and Sample Irradiation

2.4.1 Beam Production and Transportation

Most of the particle beams for PIXE analysis are produced by small Van de Graaff accelerators, with maximum terminal voltages of 2–4 MV. Accelerators of this kind often exist in nuclear physics divisions of universities or other research institutions, being the equipment used for nuclear spectroscopy studies in the 1960s and 1970s. In these situations the initial cost of setting up centers for IBA and specifically PIXE has been limited to the dedicated irradiation chambers (or the specimen holders for external set-ups, see below) and to the detection systems, which are not especially expensive. This certainly favored, in the 1980s, the start-up of PIXE activities in many laboratories. However, these accelerators are becoming less supported by their manufacturers. A new generation of small electrostatic accelerators (based on voltage generators of both Van de Graaff and Cockroft-Walton type) have been installed in some laboratories either to replace older ones or to start up new PIXE (and other IBA) activities. Most of these new accelerators are of the tandem type. In tandem accelerators, two acceleration tubes are placed on opposite sides of the positive high-voltage terminal, located in the centre of the machine, inside a tank which is gas filled at high pressure (this is needed to avoid voltage discharges to ground). The ions are produced by an external negative-ion source and then injected into the first acceleration tube from the low-voltage side. After acceleration to the positive voltage terminal, the ions undergo a charge inversion through the stripping of the outer electrons (produced by passing the beam through proper thin carbon foils or gas flow), so that they are further accelerated in the second tube towards ground. In this way, a lower terminal voltage is sufficient to produce a beam of sufficient energy. For instance, with a 2-MV tandem, proton beams up to 4 MeV, or α-beams up to 6 MeV can be produced. Further advantages of tandems derive from the presence of the external ion source (contrary to the case of single-ended electrostatic accelerators, where the ion source must be on the high-voltage terminal inside the high-pressure tank). With different sources, it is possible to accelerate any kind of ions, including the heavier ions that may be used for other IBA techniques. In addition, one can perform accelerator mass spectroscopy (AMS) for the search of ultratrace elements or rare isotopes in a sample. For this purpose, the sample is used as the material to be ionized in the source; the best-known case is the quantification of ^{14}C for archaeological dating. Together, these factors have initiated start-up of several new laboratories (or the modernization of previously established ones) based on tandem machines, where both PIXE and other ion-beam-based applications are performed.

In standard PIXE, the beam from the accelerator is led to hit the sample by conventional beam transportation systems, mostly based on magnetic deflection and focusing. The typical energy stability of beams produced by this kind of accelerator is quite good, with a residual voltage ripple normally less than 1 kV rms at maximum voltage. For PIXE applications this is not a real need, however, because the trend of X-ray production cross-sections with beam energy is rather smooth. Also, the beam currents that the accelerators might produce (up to tens of microamps) are not exploited, due to the large value of the production cross-sections. Typically, PIXE measurements are performed at beam current intensities of 20–30 nA at most (but much less than 1 nA for certain applications in which precious samples might otherwise be damaged by a more intense beam; see section 3).

Thus, in both respects (energy stability and current intensity) the accelerator performance is never a limitation for standard PIXE. This is less true for micro-PIXE set-ups, in which a much higher beam brightness is needed and high energy stability is at a premium. However, the standard performance of typical accelerators is already adequate for micro-PIXE; thus, a new accelerator of special design is not required just to implement a micro-PIXE facility.

The beam transportation line is obviously under vacuum. Typical values are around 10^{-6} mbar, although worse vacuum levels may be tolerated for PIXE applications (but not for micro-PIXE). Visual inspection of the beam cross-section is often provided along the beam line by retractable quartz viewers, whose fluorescence is observed through viewports by means of TV cameras. This makes the tuning of focusing and deflection systems much easier. For the same purpose, retractable Faraday cups are often available at critical points along the beam line.

The beam cross-section at the target may need to be either as small as possible (for a microbeam, section 2.5), or larger so as to average over a wider area of interest, which may be even some tens of square millimeters. In the latter case, it is necessary to have a uniform distribution of the beam current intensity over the probed area, in order to correctly average out inhomogeneities in the target composition.

A uniform beam current density can be achieved in several ways. One possibility is to pass the beam through a thin metal foil (some hundreds of micrograms per square centimeter) placed one or two meters upstream. Then the multiple small-angle scattering within the foil widens the angular aperture of the beam, and an adjustable collimator picks up its central, homogeneous part over the desired size. A side effect with this technique is an energy loss and straggling of the beam, but this is not a problem in PIXE and can be perfectly taken into account in the quantitative analysis.

Alternatively, the beam may be defocused by untuning the magnetic quadrupoles present on the beam line for focusing, and then collimating close to the target. Although this technique is simpler, it is less effective because it is difficult to monitor the actual homogeneity of the beam intensity over the selected area. A third method uses a well focused beam, and rasters with an x-y scan over the selected sample area. The disadvantage of this technique is that, in the presence of target inhomogeneities, the X-ray count rate from the different parts may vary considerably during the raster. Consequently, the pile-up corrections in fitting the PIXE spectra (section 2.6.2) may become problematic, as the corrections applied by the data reduction software are based on the assumption of a time-uniform count rate.

All collimators along the beam transportation lines are generally made in materials producing no detectable X- or gamma rays. Tantalum is the usual choice far from the target point because of its good dissipation of the heat generated by beam stopping. The production of K-series X-rays (around 60 keV) and of gamma rays due to Coulomb excitation of the nucleus is not intense, whereas copious L X-rays produced are easily absorbed by placing lead shielding in the line of sight to the detectors. Close to the target, the preferred material is graphite, which does not emit X-rays if sufficiently pure.

2.4.2 Irradiation Chambers

In conventional PIXE, irradiation is performed with the sample in vacuum. This is a limitation in several important kinds of application, such as when large samples are to be analyzed and taking samples of material is not allowed, or when keeping a specimen in vacuum could alter its composition. To overcome these problems, many laboratories have implemented external beam irradiation facilities (section 2.5).

The irradiation chamber is often made in aluminum, for economical reasons, but this material is not best suited because of its relatively high cross-section for the production of gamma rays. The regions of the chamber (as well as of the interior of the beam line) that are exposed to the beam (due to scattering) should be lined with nonemitting materials, such as ultrapure graphite.

Most chambers have a multisample capability in order to make the analysis faster. More or less sophisticated means of automatic sample changing are available. Usually a glass window allows the operator to view the samples inside the chamber with a video camera, both during sample changing and during irradiation, to check sample integrity under the beam.

The usual detection geometry is with the X-ray detector as backwards as possible, because the secondary BS is lower (section 2.3.1). The X-ray detector may be placed outside the vacuum chamber, viewing the sample through a window made as thin as possible so as to limit absorption of low-energy X-rays. Alternatively the detector cryostat may even be incorporated within the chamber, sometimes with the possibility of a windowless geometry based on a sliding vacuum seal (section 2.6.1). A means of inserting a variety of X-ray absorbers of different thickness and material must be provided, to enable changes to the detection efficiency curve versus X-ray energy according to the specific analytical demands of the experiment (section 2.6.2).

Additional detectors are often present in the chamber to complement PIXE analysis with the simultaneous use of other types of ion-beam-based techniques. These usually include one or more surface barrier silicon detectors

with which to implement backscattering spectroscopy. When using protons as projectiles, it is not always safe to assume ''Rutherford'' elastic scattering as valid, i.e. cross-sections may not always be computed from purely Coulomb projectile–nucleus interactions. In the proton elastic scattering on light-Z elements, the effects of the nuclear force on cross-sections may be nonnegligible or even dominant, due to the close approaches that can take place between protons and target nuclei. In spite of this, however, useful quantitative information can be extracted (corroborating and sometimes adding to that obtained by PIXE), provided that experimental data (including the nuclear interaction effects) are available for the cross-sections. Inelastic scattering can also provide useful information. A germanium detector may be faced to the irradiation chamber for the simultaneous detection of gamma rays: PIGE. Because gamma ray energies are much higher than X-ray energies, and consequently the absorption of electromagnetic radiation is not a problem, the germanium detector need not be placed inside the chamber.

2.4.3 Measurement of Beam Current and Charge Integration

During an irradiation, it is important to monitor the beam current because the sample may be damaged if safe levels of intensity are not maintained. The safety level varies with sample types, being much lower in general for electrically and thermally nonconductive materials. Equally important is accurate integration of the beam current during each run, as this is often a fundamental parameter for correct quantification of the obtained data (see Equations (3), (6), and (13)).

With thin specimens that allow the beam to pass through them with negligible energy loss, the traditional technique may be used of both measuring and integrating the beam current in a Faraday cup. The cup must be obviously insulated from the beam line. The cup material or lining must be such as to avoid emission of detectable X- and gamma rays, for the same reasons given above for the materials of beam line, collimator, and irradiation chamber. Typically graphite is used. Care should be taken that the cup receives all the particles that have traversed the beam; with samples of intermediate thickness, a relevant overall scattering may occur and the Faraday cup geometry must ensure an adequate acceptance angle. Also, the phenomenon of secondary electron escape from the cup may be relevant, which could give rise to an overestimate of the actual beam current. To prevent this effect, a negatively charged ring may be used as a guard electrode at the entrance to the cup, to reject outgoing electrons. Alternatively they can be deflected and thus recollected through the use of small magnet assemblies (electrical bias and small magnets will not affect the much heavier and energetic incoming particles). An accuracy as good as 1% can be achieved through the use of good-quality current meters and charge integrators.

With thick samples that stop the beam, the current must be measured and integrated on the sample itself, which requires the samples to be made electrically conductive if not already so (in order to avoid sample charging and consequent abrupt discharges). This may be achieved by a thin surface carbon coating, as in the case of EPMA. Secondary electrons escaping from the sample may be a much more relevant effect than with Faraday cups, due to the open geometry; they must be recollected by using a negatively biased suppressor electrode at a few hundred volts potential. Alternatively, the target itself may be positively biased with respect to the chamber.

With thick samples – or even with samples of intermediate thickness, causing a large scattering of the beam – alternatives to the Faraday cup may be used to measure the beam current. A thin self-supporting foil can be inserted along the beam path before the sample, and the count rate of backscattered particles, measured by a proper particle detector, can be taken as an indirect measurement of beam current intensity. Alternatively, the yield of some other interaction product of the beam particles with the foil material – such as X-rays themselves – may be used as an indicator of the beam current. In some external beam set-ups, the beam line exit window itself is used. When using such procedures, the problems of correct quantification of the current connected to secondary electron emission are completely overcome. After a calibration they may even provide a good absolute measurement of the current, provided that the metal foil does not change its thickness during irradiation due to beam-induced damage. A variant of this method is to measure the yield of backscattered particles (or other beam-induced radiation) on a rotating vane intercepting the beam at a sufficiently high frequency that, at the end of the measurement, the current sampling procedure is significant. Typically, 1 Hz is sufficient since run times are of the order of minutes.

2.5 External Beam Set-ups

An increasingly popular variant of conventional PIXE is external or ''in-air'' PIXE. This, as well as other IBA techniques, can be performed simply by passing the particle beam through a thin window, thus hitting the sample in its natural environment. The beam size can be determined by an aperture in the last section of the beam path within the vacuum, whereas the X-ray detectors are external. The sample is placed a few millimeters from the window and, by using a positioning stage (sometimes with micrometric control), can be moved in the plane

perpendicular to the beam for selective analysis. The external environment is often saturated with helium, to reduce beam energy loss and straggling with respect to air. The helium may be simply flooded in front of the exit window, with no containment, or be held in a simple closed environment. The use of helium has the further significant advantage that the transmission of low-energy X-rays is much higher than in air, which increases the possibility of detecting lighter elements. As an example, more than 96% of the 1.04-keV K X-rays of sodium are still transmitted when traversing 3 cm of helium, whereas for a similar distance in air the transmission would be only 2.4×10^{-6}; for the X-rays of sulphur (2.31 keV), the transmissions in 3 cm of helium and air are 99.7% and 28%, respectively. A disadvantage is that helium can penetrate the thin beryllium window of the Si(Li) detectors (section 2.6.1). This produces a progressive worsening of the vacuum levels inside the cryostat, which may necessitate a pump and bake procedure, typically after some months of use. To overcome this problem, the use of hydrogen rather than helium has also been reported (detector windows are much less permeated by the hydrogen molecules), but has not gained popularity for the fear of accidents due to its inflammability.

Figure 11 is a sketch of a simple external beam set-up, with helium flooding in an unenclosed environment and two Si(Li) detectors.

Topics relevant to external beam analysis have been reviewed in several articles.[23–25]

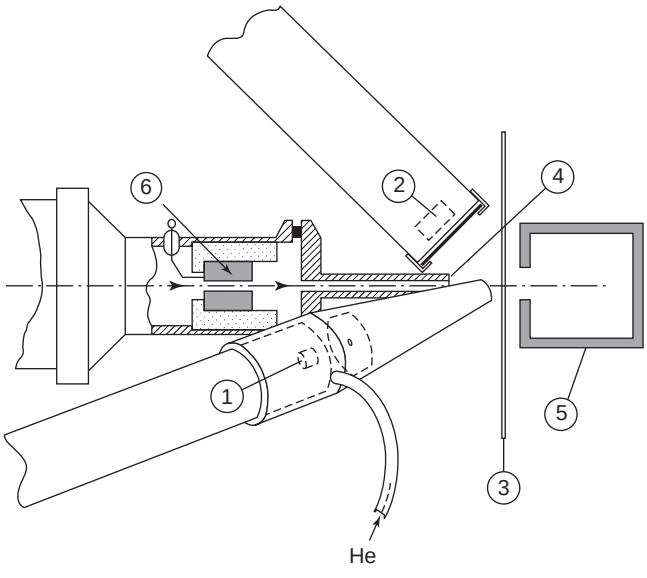

Figure 11 Diagram of a two-detector, external beam set-up (viewed from above): (1, 2) X-ray detectors, (3) sample for analysis, (4) beam exit window, (5) Faraday cup, (6) graphite collimator.

The main advantages of an external beam set-up are:

- ease of handling and changing of the sample – even large objects may be analyzed, and the specimen positioning is not problematic;
- drastically reduced risk of beam-induced damage to the sample – the problems of local heating are much smaller, especially when using a helium flow, which is a very efficient heat remover;
- electrical charging of insulated samples does not occur – this avoids the need for surface coating by graphite, as is usually done with samples in vacuum;
- ease of installation and change to the set-up (type, number and position of detectors, tilting angle of the sample, provisions for monitoring camera, etc.).

The first three of these points have encouraged the use of external beam set-ups for the analysis of fragile and/or precious objects, such as items of artistic, archaeological, and historical interest (sections 3.4 and 3.5). However, the external beam set-up is advantageous for most PIXE applications. Only for applications requiring a microbeam below about 10–20 μm is the external beam set-up not appropriate. Indeed, the scattering in the window and in the external path produces a beam halo. Thus, even with the best focusing, the actual beam spot size on the target cannot be smaller than a minimum value (depending on several factors, mainly the distance travelled by the beam in atmosphere). Some inconvenience with an external beam may also arise for other IBA techniques, such as Rutherford backscattering spectroscopy (RBS) or nuclear reaction analysis (NRA), when operated simultaneously with PIXE. As in those techniques the accuracy of the analytical information is directly related to the beam energy, the energy straggling due to the exit window and to the residual path in the atmosphere worsens the quality of the data obtained. However, in the late 1990s thinner beam exit windows have become available (see below). This, and the use of helium outside the window, has enabled the advantages of the external beam set-up to be retained without excessive drawbacks.

In external beam set-ups, the exit windows for the beam must be mechanically strong enough to sustain the pressure drop from the vacuum environment of the accelerator beam lines to atmosphere, and also exhibit a high resistance to radiation damage. Certain materials, such as Mylar®, which are mechanically strong enough, cannot tolerate even weak beams without early rupture. A typical material for beam exit windows is Kapton®, a polymer of composition $(C_{22}H_{10}N_2O_4)_n$. At a thickness of 8 μm, Kapton® typically withstands a beam of 10 nA with a cross-section of 1 mm^2 for 2–3 days, after which a change of the window is a wise precaution. The energy losses in

8 μm of Kapton® are 90 keV and 125 keV, respectively, for 3 MeV and 2 MeV protons; the fwhm energy straggling is about 20 keV (uncorrected Bohr estimate). Further energy loss and straggling before hitting the target are typically of the same order in air, or much smaller in helium. The overall values for both the energy loss and straggling (the former sum up linearly, the latter combine quadratically) are therefore of little consequence to PIXE, due to the smooth trend of X-ray production cross-sections with beam energy. Therefore, the possibility of a quantitative analysis is fully retained with external PIXE.

Another material sometimes used for exit windows is aluminum (about 10 μm thick), with a durability of months before changing the window. However, a higher continuous background is produced in the spectra due to beam-induced gamma rays from the window. Its use is therefore not recommended when bombarding very thin samples, where the X-ray yield is relatively weaker. With thick samples, an aluminum window may be a good choice.

The successful use of ultrathin windows of Si_3N_4, available down to a thickness of 0.1 μm, has been reported.[26] This material exhibits a very high resistance to the beam and the thin foils still withstand the pressure drop for a surface area of a few square millimeters. It has been employed therefore to extract focused proton microbeams down to 20 μm, with negligible halo due to the extreme thinness and to a very compact geometry with the specimen quite close to the window, in a helium atmosphere.

2.6 Detection of X-rays, Data Acquisition and Reduction

2.6.1 Working Principle and Characteristics of Si(Li) Detectors

Typical detectors used in PIXE are lithium drifted silicon crystals acting as diodes. In these solid-state nuclear detectors a free charge-depleted volume is created on one side of the junction by applying a reverse bias with a convenient high voltage power supply (some hundreds of volts, up to 1 kV). The ionizing radiation that interacts in this volume produces free hole–electron pairs that are promptly swept by the electric field – due to the voltage bias – towards the collecting electrodes. As the number of electron–hole pairs created inside the active volume is proportional to the energy release of the ionizing radiation (an average energy value of 3.6 eV is needed to create a hole–electron pair), the total charge collected on the electrodes is also proportional to the released energy. The current pulses through the detector produce, at the preamplifier output, voltage steps of amplitudes proportional to the energy releases. These voltage signals, when appropriately amplified and shaped, are then amplitude-analyzed to construct the detected X-ray energy spectra (section 2.6.2).

The Si(Li) detectors must work at a low temperature (typically −170 °C). Otherwise, when voltage biased, the inverse current of the diode (increasing exponentially with increasing temperature) would damage the detector itself and/or the first stage of signal preamplification. In particular, the compensation due to the presence of the drifted Li ions might be lost in a short time due to their mobilization (induced by the electric field). Si(Li) detectors can instead be safely stored at room temperature when they are not under bias. Protection circuits should always be used (and are indeed available in modern detectors) to prevent the high voltage supply from actually applying the bias until a sufficiently low diode temperature is reached.

Because of the need for a low operation temperature, the crystals are kept inside a cryostat under vacuum and cooling is typically achieved by conduction through a cold finger kept in contact with liquid nitrogen in a Dewar. The option of a Peltier cooling system is also available, which makes the assembly much more compact at the expense of less-effective cooling (which implies a slightly worse energy resolution). A vacuum is first created in the cryostat at time of manufacture. A proper amount of molecular sieves, in thermal contact with the cold finger within the cryostat, passively maintain a sufficient level of vacuum when the detector is cooled. In general, this situation may be maintained for several years, after which servicing (pump and bake procedure) is needed to restore the vacuum.

The cryostat wall in front of the detector (the entrance window) must be very thin if low-energy X-rays are to be transmitted with negligible probability of being absorbed by the window itself before reaching the detector's active volume. The window should also be light-tight because silicon detectors are photosensitive, producing a large response to ambient light. Traditional cryostat windows are made of beryllium, of a thickness down to 8 μm. They can stand the pressure drop over a surface area of about 1 cm^2, but they must be handled with great care because they are very fragile. Any puncture of the window may be catastrophic, with sudden ventilation of the detector's cryostat. This in turn may have very serious consequences in that the detector system can be destroyed if the diode was under bias – the sudden vapor condensation on its surface gives rise to current paths, which may spoil the passivated surface that maintains the bias voltage. In addition, the large currents may burn the entrance FET (field effect transistor) of the preamplifier.

Typical active areas of the Si(Li) detectors are from 10 to 100 mm^2 and the thickness of the depleted volume is from 3 to 6 mm. As far as the intrinsic efficiency

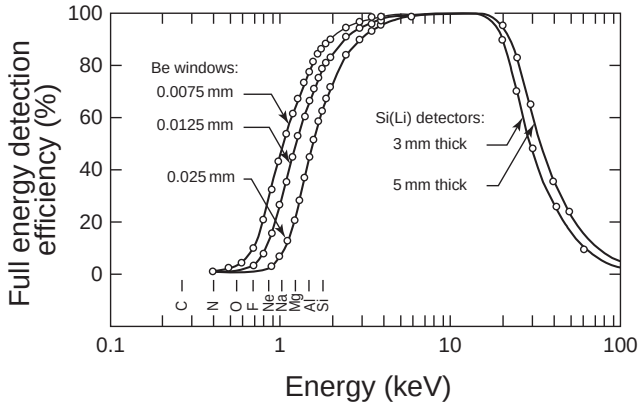

Figure 12 Intrinsic efficiency of Si(Li) detectors versus X-ray energy. The three curves on the left side refer to detector entrance windows of different thickness; the two curves on the right refer to different overall thickness of the detector's active volume. (Reproduced by permission of EG&G Ortec.)

ε_{det} is concerned (see Equation (4) and the following discussion), Si(Li) detectors have a 100% intrinsic efficiency over an X-ray range from 2–3 keV to about 20 keV. This means that in this energy range an X-ray reaching the entrance window will lose all its energy in interactions inside the active volume, thus producing a full-energy peak in the spectrum. Figure 12 shows the intrinsic efficiency of typical Si(Li) detectors as a function of X-ray energy, for various entrance windows and detector active thicknesses. The detection efficiency decreases at both lower and higher energies. On the high-energy side, this is in connection with the limited active thickness of the detector – with smaller detector thicknesses and higher X-ray energies, the probability of losing all the energy within the active volume decreases. On the low-energy side, the intrinsic efficiency decrease is due to the increasing probability of X-ray absorption in the detector entrance window.

Materials other than beryllium have also been used to build extra-low absorption windows that make the detection of very light elements possible, down to carbon. An alternative when the operator wishes to extend analysis to these elements is the possibility of window removal. This option can only be used when working with an in-vacuum set-up, and through some mechanical design to allow for vacuum-tight insertion of the cryostat inside the irradiation chamber. Removal of the window can only be done once a sufficiently good level of vacuum is reached in the chamber, in order to avoid vapor condensation on the cold detector surface.

It should be remembered that the severe problems of self-absorption within the sample itself (section 2.2.2) for the very-low-energy X-rays involved in the detection of these light elements (276 eV for carbon) makes their analysis only qualitative anyway. This consideration is more critical the lower the atomic number. As a general rule, relatively good quantification can be made only when starting from Na ($E_X = 1041$ eV); however, this also depends on the kind of sample (whether it is extremely thin or not, the nature of matrix, etc.) and on the degree of accuracy required.

Besides the effect of the cryostat window, at lower X-ray energies the detection efficiency is very sensitive to any absorption layers introduced intentionally or due to dead surface layers on the detector. The latter may occur when the vacuum level inside the cryostat worsens because of small leaks – a layer of condensed vapor is formed on the front surface of the cold detector inside the cryostat. Low-energy efficiency can also be drastically reduced by purposely interposing absorption filters along the path from target to detector. For example, with a Mylar® absorber of only 10-μm thickness the transmission of the 1.04 keV X-rays of Na is 3.3%, that of the 1.74 keV X-rays of Si is 47%, that of the 2.3 keV X-rays of S is 72%, and 92% of the 3.69 keV X-rays of Ca are transmitted. With 100 μm of Mylar®, transmissions are $<10^{-14}$, 0.05%, 3.5%, and 43%, respectively, for the X-rays of the four elements. It is obvious that by using different absorber material and thickness one can drastically modify the curve of detection efficiency versus X-ray energy. As explained in section 2.6.3, this property is often exploited to make the MDLs more homogeneous over a wider range of Z.

2.6.2 Spectra Acquisition and Data Reduction

The voltage signals produced by the charge preamplifiers are first shaped and amplified to make them suitable for amplitude classification, which is the basis for the acquisition of the energy spectra. A quasi-Gaussian pulse shape is normally used, which can be synthesized by different linear networks including successive stages of differentiation and integration of the original quasistep signal from the preamplifier.

The time constant used must be long enough to allow for complete collection of the charge created by the ionizing radiation inside the active volume of the detector. Incomplete charge collection would result in the so-called ballistic deficit effect, which produces low-energy tails to the line shapes in the energy spectra, increasing linearly with peak energy. Also, because of the prevalence of series noise, long time constants favor a better signal-to-noise ratio, i.e. better energy resolution.

At the same time, the time constant should be short enough to avoid a long time occupancy of the signal baseline by the pulse, in order to make handling of higher count rates possible. For the usual quasi-Gaussian pulse shapes a baseline recovery to better than 1% is achieved in a time equal to 3–5 times the time constant.

When the average time interval between successive pulses gets smaller than about 10–20 times the baseline occupation of a single pulse, there begins to be a non-negligible, increasing probability that a pulse arises before the previous one has completed its recovery to the baseline. This generates a wrong classification of the pulse amplitude (the pile-up effect) with consequent distortion of the accumulated energy spectra. Minor distortions may also be generated by imperfect tunings of the electronic chain, such as improper adjustment of the pole-zero cancellation or of the baseline active recovery circuits.

The choice of time constant is dictated by the main requirement in terms of performance (as outlined above, high count rate capability and good energy resolution are somehow conflicting). In general, a compromise between the two needs is given by a time constant value of 6–10 μs, but longer values also may be used when extremely good resolution is the main goal and the low count rate makes this possible. Of course, everything depends on the specific experimental situation – target composition and thickness, beam current, detector size and geometry, etc.

The linearly shaped pulses are amplitude analyzed by an analog-to-digital converter (ADC), which is then coupled to a computer where a distribution histogram of the digitized amplitudes is constructed. Because all the signal handling processes are linear, the histogram represents the spectrum of the radiation energy losses inside the detector. Thus, the full-energy X-ray interactions give rise to peaks in the spectrum. The energy scale of the spectrum can be calibrated by using a source of X- or gamma rays of known energies for a spectrum acquisition.

To extract quantitative information about elemental abundance in the sample, the peak areas in the X-ray energy spectra must be evaluated as accurately as possible. The various aspects of the spectrum deconvolution procedures are essential for quantitative analysis. In particular, modeling of the continuous background shape and deconvolution of peak multiplets have been widely studied. Several software packages are available to the analyst, including provisions for taking thick target matrix effects into account. Extended reviews of these aspects are available.[2,27]

The first step of the process is in general the modeling of the background. As the background is higher in the lower-energy X-ray region (section 2.3.1), where in addition the X-ray density (and hence the probability of overlaps) is higher, this is the most crucial energy region to be considered. Basically, two different strategies may be adopted: the assumption of an analytical (in general polynomial) function for the shape of the continuum; and an empirical subtraction of its contribution by a smoothing–filtering procedure that eliminates the variations of the counts per channel that are slower than a preset value. To be considered as not arising from true lines, these variations must meet certain criteria, which take into account the amplitude of the statistical fluctuations and the assumed peak widths. Once the contribution of the continuous background has been eliminated by this filtering procedure, the peak areas are determined from the filtered spectra (rather than from the direct spectra).

For the purpose of the detection of peaks, even in the presence of multiplets, and of the evaluation of peak areas, two aspects are essential: (a) knowledge of the detector's response to a single line, i.e. the peak shape; and (b) access to a complete and accurate database of X-ray energies and relative intensities for all the elements in the periodic table, so that after energy calibration the peaks can be assigned to the corresponding elements.

Considering the first point above, the line shapes of the full-energy X-ray peaks in the spectrum are basically Gaussian. The effect of electronic noise that adds randomly to the signal pulse height, and the stochastic nature of the process of electron–hole pair creation in the detector's active volume both contribute to this shape. X-rays losing identical energies inside the active volume do not always produce exactly the same number of pairs, and this fluctuation in turn results in voltage pulses of variable amplitudes and therefore in peak widening. Due to the statistical nature of pair production, this contribution to the fluctuations is proportional to the square root of the number of electron–hole pairs created, i.e. of the energy loss. The total width of a peak in the X-ray spectrum is then given by an expression such as Equation (15),

$$\Delta E = \sqrt{A + BE_X} \tag{15}$$

where A is the square of the electronic noise contribution, B is the square of the so-called Fano-factor relating fluctuations to the square root of energy loss, and E_X is the X-ray energy corresponding to the peak. This behavior of ΔE versus X-ray energy is taken into account when deconvoluting complex spectra.

However, the actual spectral line shape may be more complicated than Gaussian if ballistic deficit and/or pulse pile-up effects are present. Even in the absence of these effects, for high precision deconvolution an accurate representation of the detector's response is needed. Even minor effects give rise to lineshape distortions, which have to be taken into account, in particular to properly detect and quantify small peaks on the tails of larger ones. Figure 13 illustrates the main contributions to the lineshape. Several deviations from the ideal Gaussian shape can occur to a larger or smaller extent depending on the specific detector.[28]

A major effect that must be mentioned is the presence of an additional satellite peak – an escape peak – at an energy lower by 1.74 keV than the full energy peak.

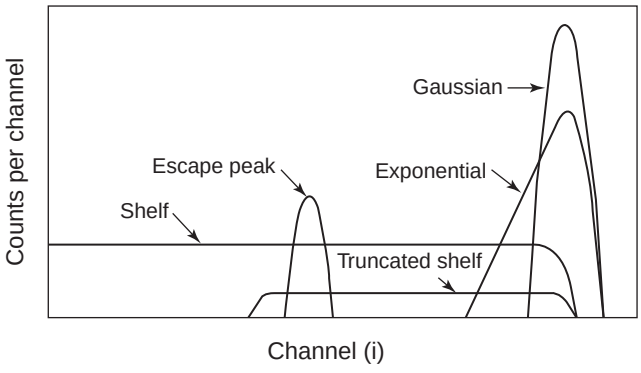

Figure 13 Diagram of the different lineshape components for a Si(Li) detector in response to an X-ray. (Reproduced by permission from Campbell.[28])

Escape peaks result from primary X-ray interactions that ionized – by fluorescence – the K shell of the silicon atoms in the detector crystal. The resulting K-series X-rays of silicon (1.74 keV) may then escape the detector before releasing their energy to the crystal. In these cases, the total energy released to the crystal is 1.74 keV below the primary X-ray energy.

The relative intensity of the escape peak with respect to the corresponding full-energy peak increases with decreasing X-ray energy. This is because the average depth of interaction of the primary X-ray from the detector surface is lower at lower X-ray energies, and it is then less probable that the silicon X-rays produced by fluorescence are recaptured within the crystal. The relative intensity of the Si escape peak to the full-energy peak falls from about 1% at 5 keV to about 0.1% at 10 keV. It is important to keep in mind the presence and intensity of the escape peaks because they may overlap the full-energy peaks of other elements, giving rise to some confusion in the spectra deconvolution if not properly taken into account. Examples of possible mistakes between an escape peak and a full-energy peak which may be important to an evaluation are:

- Ca K_α escape ($3.69 - 1.74 = 1.95$ keV) with the P K_α line (2.00 keV)
- Cr K_α escape ($5.41 - 1.74 = 3.67$ keV) with the Ca K_α line (3.69 keV)
- Fe K_β escape ($7.05 - 1.74 = 5.31$ keV) with the Cr K_α line (5.41 keV).

2.6.3 Optimization of Detection Limits over a Wide Range of Z

As discussed in section 2.2.1 (Figures 7 and 8), the physics of X-ray production in PIXE unbalances the detection efficiency in favor of the low- and medium-Z elements. This holds true even after using the favorable circumstance that L-series rather than K-series X-rays can be used for the quantification of higher-Z elements. This unbalance is a problem when detection of low or trace quantities of high-Z elements is required in the presence of a lower-Z matrix. Figure 7 shows that, if the count rate due to the low-energy X-rays emitted by low-Z elements is kept within tolerable limits, the total number of counts of X-rays from medium- and high-Z elements is too small within reasonable measurement times. In this situation two strategies may be adopted.

The first uses two X-ray detectors simultaneously, which complement each other in terms of the efficiency curve and the solid angle covered. One detector is chosen to cover the widest possible solid angle (which increases efficiency regardless of the X-ray energy), but is prevented from obtaining a large count rate from low-energy X-rays by shielding its surface with an absorber of appropriate thickness (this reduces the efficiency for just the low-energy X-rays). This detector is thus optimized for the medium- and high-Z elements but, due to the absorber, it is practically blind to the lightest elements (see the efficiency curve in Figure 8). The second detector, with no absorber but covering a much smaller solid angle, is then used to simultaneously provide analytical information concerning the low-Z elements. This case is represented by the efficiency curve of Figure 7. The combined effect of the two-detector system, when both L- and K-series lines are used, is that a much broader range of Z is covered with more homogeneity in the detection efficiency.

By changing the absorber on the higher-Z detector and the solid angle of the lower-Z detector (by using appropriate collimators), it is possible to optimize the global detection efficiency over the two detectors to the widest range of Z, tailored to the particular analytical problem (which trace elements are of interest, the matrix composition, etc.).

A cheaper alternative to a two-detector system is the use of a so-called funny filter on just one detector. A funny filter is a nonhomogeneous absorber, which can be constructed from a stack of foils having central apertures of different sizes. In this way different absorptions can be achieved for different solid angles. Thus the detection efficiency for higher-energy X-rays (which are little or not at all affected by the presence of the absorption layers) can be actually made larger than for the lower energies. In other terms, referring to Equation (7), the parameter Ω becomes an increasing function of X-ray energy, getting higher for higher-Z elements. A judicious choice of the funny filter geometry and composition can compensate effectively for the drop in X-ray production cross-section at higher Z. The drawback is that the actual shape of the function η versus E_X becomes more irregular than with a homogeneous absorber (or no

explicit absorber). This results in a more difficult modeling of the efficiency curve $\eta(E_X)$, which in turn is essential to obtain accurate quantitative results with PIXE. In practice, when using a funny filter a large number of standard reference samples (one for each element over a much larger Z range) is needed to experimentally determine the detailed behavior of $\eta(E_X)$. Even so, the uncertainty remains larger because the impact of possible inaccuracies in the knowledge of the areal densities of the individual standards is no more attenuated by the effect of ''averaging'' with neighboring-Z standards that is instead allowed when in the presence of a smooth behavior with Z.

2.7 Micro-PIXE

Depending on the specific analytical problem, one may need a relatively large size of probing proton beam (up to tens of square millimeters, when it is necessary to get an average composition on a wide surface), or a very small one (when the surface distribution of the elemental concentrations has to be mapped with the highest possible detail). When the beam diameter is reduced to about 100 μm or less, one generally talks about microbeams. In the simplest microbeam set-ups, the cross-section of the beam on the target is defined by collimation, by placing slits or apertures upstream, not far from the target. By this procedure, beam sizes as small as 100–200 μm can be obtained; this represents a sufficiently small size for many applications requiring detailed information on spatial distribution of elemental concentrations.

Sometimes an even smaller beam cross-section is necessary, and any further reduction of the beam size by brute collimation is no longer effective. At very small apertures, scattering of the particles by the slit or collimator edges produces significant beam halos over an area larger than that geometrically defined by the aperture. To obtain beam sizes down to 1 μm or below, strong focusing is necessary. The production of well-defined and sufficiently intense beams (1 nA of proton current is not exceptional) of the order of 1 μm is now a standard feature of set-ups based on strong focusing and the technical aspects of these set-ups have been widely reviewed.[29–31] Essentially, a microbeam is obtained through a beam-optics system which includes an object (a small, micrometrically adjustable slit aperture illuminated by the beam from the accelerator) and a set (doublet, triplet or quadruplet) of magnetic quadrupoles acting as a whole on the particle beam as a positive lens. The beam is focused on the target (image of the system), the distance of which from the quadrupoles is much smaller than the object–lens distance. In this way, high demagnification is obtained. In order to reduce the geometrical aberrations, the beam divergence at the entrance of the quadrupole system is strongly limited by collimator slits (Figure 14).

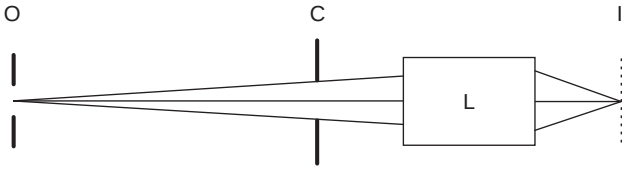

Figure 14 Diagram of a microbeam-forming system, with object slits O, collimator slits C to limit beam divergence, positive lens L (obtained with a doublet, triplet or quadruplet of quadrupoles) and image plane I where the beam is focused and the target must be placed. (Reproduced by permission from Grime and Watt.[30])

Almost invariably, the microbeam facility includes a system to fast scan the beam two-dimensionally over the sample, with a raster area having sides of the order of up to several hundred micrometers. This is accomplished by magnetic deflection with coils prior to the quadrupole focusing. Figure 15 is a photograph of the final section of a proton microprobe, including collimator slits, scanning coils, triplet of magnetic quadrupoles and irradiation camera, with micrometric adjustment of the target position and viewing microscope from outside.

In the presence of scanning coils, data acquisition is performed in the so-called list-mode. This means that when an X-ray is detected, a triplet of data is stored: the X-ray energy and the two coordinates of the beam spot position at the detection instant (deduced by the current values of the scanning coils at the instant of X-ray detection). The analysis of the triplets, which can be performed on line, includes the possibility of obtaining elemental maps (plotted as the number of detected X-rays of any given energy as a function of position over the whole scanned area). Alternatively, it is possible to

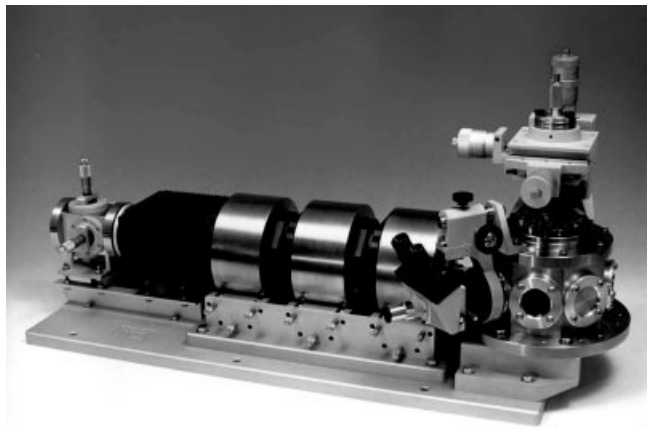

Figure 15 Photograph of a proton microprobe system, showing the final section. From left to right: the collimator slits (micrometrically adjustable), the scanning coils, the lens (made in this case of a triplet of quadrupoles), the target chamber. (Reproduced by permission of Oxford Microbeams Ltd.)

reconstruct the X-ray spectrum detected over any portion of the scanned area (down to pixels that correspond to the limits of lateral resolution, i.e. to 1 μm or even less). This is similar to the result of EPMA; however, the better MDLs offered by PIXE (section 1.3.1) make it possible to extend elemental mapping even to trace elements.

When pushing the limits of lateral resolution to their extremes, a number of factors come into play which limit the possibilities of obtaining ever smaller beam spot sizes.

First, the complex aberration effects of the probe-forming lens system have to be diagnosed and minimized. These include:

- astigmatism, due to the lack of cylindrical symmetry of the quadrupoles;
- higher-order multipole contamination of the quadrupole field, deriving from inaccuracies of construction of the quadrupoles (deviations from fourfold symmetry);
- spherical aberrations, coming into play only if higher currents are needed, which implies the acceptance of a larger beam divergence;
- rotational, translational and tilt misalignments of the quadrupole multiplet assembly.

Second, external sources can also produce broadening of the beam spot. Among these, an important role is played by the accelerator characteristics:

- beam brightness at the object slits, which in turn depends on many factors connected with the ion source, accelerator, and up-stream beam optics design;
- accelerator energy instabilities, which produce the equivalent of chromatic aberrations.

Other relevant factors affecting the quality of the microbeam are:

- scattering phenomena, both at the slits and due to residual gas in the beam line due to poor pumping;
- stray electromagnetic fields that may be caused by several sources in the vicinity of the beam line (transformers, motors, power supplies, etc.);
- mechanical effects such as vibrations (e.g. induced by pumps) in the target area, or even slow drifts of the target during irradiations, such as those caused by thermal effects.

One final point concerning PIXE with proton microbeam set-ups is that although a lateral resolution as small as 1 μm or less can be obtained, one should not forget that the depth investigated by PIXE at standard bombarding energies (2–3 MeV) is up to some tens of micrometers (sections 1.3 and 2.2). Consequently, in thick targets the compositional information is an averaged value over a larger size. Thus full advantage of ultrahigh lateral resolution is only obtained with very thin targets (such as with certain biological specimens), or in samples where the elemental map of the surface layer is also representative of the structure in depth (as in the analysis of electronic devices).

3 SURVEY OF THE MAIN ANALYTICAL APPLICATIONS

The analytical applications of PIXE are innumerable. Regular international conferences are held to give scientists of the IBA community the opportunity to exchange ideas on the technical developments of PIXE (improved databases of fundamental parameters, detection systems, data reduction codes, etc.), and on new analytical applications. New ideas continue to be proposed, especially when PIXE is combined with other IBA and nonIBA techniques, which is performed more and more. The following discussions outline some general considerations concerning groups of applications, with no claim to being exhaustive. The intention is to point out the peculiar analytical aspects of PIXE that are specifically important in each of the main applications.

3.1 Environmental Pollution Studies

These applications were among the first to be successfully performed during the early years of PIXE development. Both water pollution and air pollution are currently investigated, with more emphasis on the latter. The importance of aerosol analysis is obvious, both in view of air quality assessment in urban and industrial areas (impact on the health of humans, animals and plants), and as a tool in the investigation of general atmospheric phenomena such as short- and long-distance air transportation, visibility, and climate.

Since the early stages of aerosol investigations by PIXE, several reasons for a widespread use of this technique became apparent.

- Variations in the atmospheric aerosol composition can take place over very short timescales (of the order of an hour) or on longer, seasonal, timescales. The same applies to spatial variations: air transportation and deposition phenomena can produce extremely different results depending on the site of measurement. This large variability calls for vast numbers of analyses, and a method such as PIXE, capable of yielding fast quantitative results, lends itself to these

investigations much better than does conventional chemical analysis.

- All elements are potentially important in aerosol analysis, because many can be toxic, and because any compound, bound in particles of the appropriate size, can have an effect on visibility (by the scattering of light). Also, the simultaneous detection of groups of elements, occurring in particular ratios, can be a signature of the source of pollution. Their detection at different sites may constitute an important indicator of long-range air transportation phenomena. In performing the necessary multi-element quantitative analyses, the PIXE method offers advantages in terms of time and cost, with respect to chemical methods or optical spectroscopies. Particularly important is the possibility of complementing PIXE with the simultaneous use of other IBA techniques to extend the range of detectable elements, especially the lighter ones.
- The total quantity of aerosol mass available for compositional analysis is quite small. Typical aerosol mass concentrations in air range from a few to over one hundred micrograms per cubic meter. Traditional ways of collecting aerosol are by filtration over long periods of time, using large pumps and large-size filters in order to collect at least one or two milligrams of aerosol for the analysis of only some elements or compounds. The sensitivity of PIXE, for which overall quantities of about a microgram are adequate for a multi-element analysis, considerably reduces collection times and makes the use of much smaller pumping set-ups possible.
- It is important to obtain separate information on the composition of aerosols of different sizes. The particle size distribution is not homogeneous, and the relevance of particles of different sizes is quite different. As far as the impact on health is concerned, this is clearly connected to the penetration of particles into the breathing apparatus, with smaller ones reaching deeper levels and therefore being potentially more dangerous. As to visibility problems created by aerosols, the size-dependence of their effect, being connected to the scattering of light, is even more obvious. Finally, the size distribution also provides information on the sources of pollution. For all these reasons, sampling methodologies providing size-fractionation of the aerosol are highly desirable. These have been developed using inertial impactors that exploit the differing inertias of different sized particles when following air flow paths in special collectors. These samplers may also provide time-sequence deposition of the size-fractionated aerosol on streaks (typically, about 1 cm high) whose total length may correspond up to 1 week of aerosol collection or even more; each section of about 1 mm width corresponds to a short sampling period of typically 1 h. PIXE can be used to analyze these sections selectively, thereby reconstructing the time behavior of the aerosol composition. The overall quantity of aerosol deposited on each section is minute, typically of the order of 10 μg. The high absolute sensitivity of PIXE is therefore exploited to the highest degree. No other technique enables the analyst to obtain such a high time-resolution of the variation in aerosol composition, with simultaneous sensitive detection of a very large number of elements in runs lasting a few minutes. In addition, after PIXE measurements the samples remain available for alternative analyses, to corroborate or complement the data obtained by PIXE.
- The analysis of aerosol collected on ''traditional'' filters is also important. Daily sampling with these filters is routinely performed by public authorities in many western cities, but in most cases the only parameter measured is the total mass of aerosol deposited over the day. The collected aerosol is analyzed compositionally only for specific campaigns aimed at obtaining information about a few specific components; in these cases, with traditional analysis, the samples are destroyed. The nondestructive PIXE analysis of time sequences of these filters allows instead the trend of air pollution to be followed over long timescales without losing them for further tests. The reconstruction of aerosol composition has been reported in urban environments for periods of months or years.[32,33] It is obvious that over these timescales the use of 1-h resolution streakers would result in unmanageable quantities of data. Also with traditional filtering procedures, therefore, PIXE provides a precious tool in both widening the range of elements detected from each filter and preserving the samples for further investigation.

Let us now summarize the typical procedure for a PIXE analysis of aerosols.

After sampling on streak (for high time-resolution analysis), or on one-day filters (for studies of trends over long timescales), proton irradiation is performed with runs typically lasting a few minutes per sample.

For the analysis of streaks, the ''sample'' is each section of the streak and use is made of relatively small-size beams (typically 1 mm; microbeams are not needed). After completion of a run, the next section is immediately exposed to the beam by simply shifting the streaker by the proper distance, often in an automated process. For the daily filters, larger beams (made homogeneous over their cross-section by one of the methods outlined at the

end of section 2.4.1) are used and/or scans are performed over the whole area.

The currents used are typically 5–40 nA. Both in-vacuum and external beams have been used for this kind of measurement. External beams are preferable both to facilitate positioning, changing and/or scanning of the samples, and to minimize the risk of selective loss of some aerosol components when under the beam.

The spectra are normalized to the integrated beam charge on a Faraday cup behind the sample (the aerosol deposit plus substrate are thin enough to let the beam pass through), or to alternative normalization parameters (section 2.4.3).

The deduction of areal densities of the different elements on the deposit is then performed by simple comparison with thin standards. In the thin-target approximation, only the very lowest-Z detected elements may be slightly underestimated, due to the onset of self-absorption effects (section 2.2.2) within the particle grains. The latter effects can be of some relevance (up to 50%) only for the larger-size fraction of the particulate and for elements from Na to Si. However corrections can be applied, and/or these elements may be simultaneously detected by alternative IBA techniques, such as particle-induced gamma ray emission (PIGE), which also may produce data on lower-Z elements undetected by PIXE, such as fluorine.[34,35]

The elemental concentration in the air is finally obtained through knowledge of the sampling parameters (area of pumping aperture, flow rate, duration of sampling).

In a typical collection campaign, the amount of data accumulated is huge (30–40 elemental concentrations from each sample, separation according to particle size, hundreds of streaker sections (or daily filters) analyzed). Therefore, a statistical analysis must be performed. Principal component analysis (PCA) or factor analysis (FA)[36,37] are routinely performed on the data sets. These processes group the detected elements into factors according to the similarity of the behavior over time of their concentrations, thereby indicating pollution sources, reconstructing their relative weights in the total particulate mass, and also disentangling possible contributions from different sources to the same element.

Owing to the large amount of data that can be collected in a relatively easy way, large-scale monitoring of environmental quality makes wide use of PIXE data. For instance, the US monitoring network ''IMPROVE'' is largely based on PIXE-determined elemental concentrations in air.

Hundreds of papers have been published on PIXE measurements of aerosols. Specific results and review articles are published on the occasions of the triennial International PIXE Conferences.[38] Several reviews are available.[2,39,40]

3.2 Biomedical Applications

Biomedical applications of PIXE have been reviewed in detail together with exhaustive lists of references.[2,41,42] Also for biomedical work, a large number of specific papers are published in the Proceedings of the International PIXE Conferences,[38] Nuclear Microprobe Conferences, and the more recent topical series of bio-PIXE Conferences.[43,44]

The main reason for the great success of PIXE in biomedical applications is linked to the very nature of biological materials, where light elements – undetected by PIXE – are the dominant component. This favors attaining the best results in terms of MDLs for higher-Z elements. Provided that the samples can tolerate the relatively intense beam currents, MDLs well below the ppm level are easily reached. This trace element capability makes it possible to obtain precious information on physiological and pathological processes in biomedicine, where the oligo-elements often play a key role. The problem is in the correct preparation of the samples, because the biological matrix is easily damaged by irradiation and elements such as hydrogen and oxygen may be lost to a large extent due to heating. However, when these problems are adequately tackled, biomedical applications of PIXE can yield important results.

The second feature of PIXE that is important in biomedical applications is the high spatial resolution, i.e. the possibility of mapping trace element distributions in the tissues with a high definition. If a microbeam is used, the intracellular distribution of trace elements may be revealed.

To obtain good results in biomedical PIXE applications, close interdisciplinary collaboration between physicists and biologists is important at all stages of the experiment. This includes experiment design, sample preparation, data collection and processing, and interpretation of the results.

Traditionally, applications of PIXE in biomedicine have focused on the detection of trace quantities of transition elements and heavier elements in tissues. More recently, interest has shifted to the detection of lighter elements, such as aluminum, potassium and calcium. Often, the low concentration levels of these elements, which are well within the analytical capabilities of PIXE, are not easy to detect with alternative techniques, especially when dealing with small tissue samples or at the cellular level.

The sampling strategy is crucial to obtaining significant biodata from any analytical technique and must be adapted to the specific features of the chosen methodology. An essential point to bear in mind is that a large number of samples is necessary because of the large variability of trace element concentrations in biological

tissues. In this respect, PIXE is a good choice because it can analyze large numbers of samples in a relatively short time, in addition to preserving them for further complementary analysis. The latter point is an advantage in biomedical investigations because the bio-availability of trace elements may be quite different, depending on the chemical form in which the elements are bound. Thus further analysis of the same sample can be performed to yield this essential chemical bonding information.

Bio-PIXE experiments are usually performed on samples collected from living matter (biopsies). To avoid or minimize ethical problems connected with the acquiring of such samples, many investigations are performed on dermatological samples, such as nails or hair, as biological indicators of a physiological or pathological state of the individuals from which they were taken. The use of body fluids, such as sweat, tears, urine or blood, also implies minimum invasion in the sampling procedure. In any case, even when dealing with biopsies, the high absolute sensitivity of PIXE, requiring the collection of only minute amounts of tissue, is a good reason for selecting this technique. However, due to the very small amounts of sampled material involved, special care must be taken to avoid contamination from the surgical instruments used for the biopsies.

The need for small amounts of sampled material is beneficial to studies aimed at clarifying the processes of forest decline. Drill cores are collected from trees, small enough as not to harm them, and the radial trend of contaminant elements from environmental agents studied along the tree rings. One exploits here the fact that, in plants, only the external layer is living matter, interacting with the environment, and that each annual tree ring records the elemental content resulting from equilibrium with the environment for the corresponding year. Past environmental conditions can be thus reconstructed by simple PIXE scans on the small samples collected.

Sample preparation for bio-PIXE analysis depends on whether macro-analysis or micro-analysis is intended.

In macro-analysis, the desired spatial resolution in the detected elemental concentration is no smaller than a fraction of a millimeter; sometimes an average over a larger dimension is requested. When this is the case, many biological materials can be analyzed with a minimum of preparation, especially when an external beam set-up is used. Examples of materials analyzed in this way are botanical samples such as leaves, wood, and tree rings, or animal (or human) specimens such as bones, nails, hair, and teeth. Direct beam bombardment is possible in these cases, but no internal standard can be included and the absolute quantitative analysis may be difficult with PIXE alone because the matrix material is to a large extent undetected. However, in these studies it is often the relative behavior of minor or trace element concentrations that is searched for, so the matrix composition can be unimportant. Another point to be kept in mind when extracting analytical information from these untreated thick samples is that the total quantity of interrogated material volume is small, in the milligram range, due to the essentially surface character of PIXE analysis (section 2.2.2). Caution must be used in evaluating the representativeness of the results, which may be questioned if the material is not homogeneous to the macroscopic scale. Finally, surface roughness can also cause inaccuracies in the quantitative data, especially as far as lower-Z elements are concerned.

Soft tissues are often dried before analysis. The technique of lyophilization by freeze-drying is most common and performed by commercially available equipment. It has been shown that, for macro-analysis, the original composition is not affected by this procedure (whereas the spatial distribution of elements in micro-analysis could be affected if crystals are formed or thawing occurs). After drying, the tissues may be pulverized, homogenized and compressed to pellets for analysis. In this process, the material can be mixed with ultra-pure carbon powder, to ensure that the sample is electrically conductive. This avoids problems connected with insulated samples during bombardment. Throughout all these steps, great care must be taken to avoid contamination, such as from the pressing surfaces of the pelletizer.

To further increase sensitivity to the trace elements, the organic matrix mass can be reduced by well-established techniques such as wet ashing, dry ashing, low-temperature ashing, and acid digestion. The possible selective loss of some elements (e.g. chlorine and bromine in dry ashing) must be considered with care.

The preparation of biological samples for micro-PIXE analysis is different. It is clear that, for the result to be significant in terms of biological information about trace element distribution at the cellular or sub-cellular level, the spatial structure must be preserved during sample preparation and irradiation. Experience gained in the preparation of samples for electron microscopy is only of partial help here, because in those preparations chemicals are often added to increase visual contrast, and these may contaminate or selectively remove certain elements. Cryo-preparation techniques have been developed instead that lead to almost instantaneous freezing of the cells with preservation of a life-like state in terms of the elemental distribution. The frozen material must then be maintained at low temperature (below 130 °C) and the analysis performed on thin sections obtained with an ultracryomicrotome. As these sections show very little image contrast in using optical microscopy, their positioning under the microbeam with traditional optical methods is difficult. Contrast enhancement by

histochemical treatments is ruled out for fear of contamination, so alternative techniques have been developed. For instance, because the samples are thin, scanning transmission ion microscopy (STIM) can be used to properly direct the beam scan over the sample for microanalysis.

3.3 Applications to Earth Science

Review treatments of PIXE applications in the earth sciences have been given in several articles. (See references 2 [Chapter 3], 45, 46)

The impact of PIXE for the earth sciences has increased during the 1990s, with developments in microbeam technology. Micro-PIXE has early been acknowledged by geologists as an ideal complement to traditional compositional investigations using EPMA. The very low MDLs attainable by PIXE, even on the micrometric scale, make it very attractive to extend to trace elements the studies of elemental partitioning between different phases in a rock, or the zoning effects within a single crystal.

Whereas the MDLs of EPMA can be pushed at most to hundreds of ppm, micro-PIXE with beams of the order of 10 μm can reach, with relative ease, a factor of more than 100-fold better, i.e. ppm sensitivity. The attraction for Earth scientists is that this extra sensitivity is achieved with the same sample preparation (thin sections glued on a glass support) and using the same analytical principle used for EPMA (i.e. X-ray spectroscopy), with which they are so familiar. In the preparation of thin sections for PIXE, it must be remembered that the penetration of 2–3 MeV protons is larger than for 20–30 keV electrons (section 1.3.1). Therefore, thicker sections must be cut (about 70–80 μm; the standard thickness for EPMA is 30 μm), to prevent the protons from reaching the backing. With this precaution, both techniques can be used on the very same sample. In addition to providing additional information on trace elements, PIXE also provides quantification of the major and minor elements so that the results can be compared to EPMA data for the latter elements. What is crucially important is that the additional trace-element analysis provided by PIXE is achieved in a nondestructive manner, unlike alternative highly sensitive techniques with high spatial resolution, such as mass spectroscopies SIMS (secondary ion mass spectroscopy) or LAMS (laser ablation mass spectroscopy).

In comparing EMPA and PIXE results for major and minor elements, the differences in the probed depth and in spatial resolution (section 1.3.1) must be considered. This comparison is useful in that it adds reliability to the overall analytical data. The quantification of elemental concentrations from the PIXE spectra can be as accurate as 1–2%, for two reasons. First, the sample thickness is infinite and the problems connected with knowledge of the integration interval, discussed in section 2.2.2, are avoided. Second, the undetected fraction (very light elements) is in most cases (such as for silicate rocks) easy to estimate just on the basis of the detected fraction; or even all elements are actually detected, as in sulfide rocks. In silicate rocks the elements can be considered to be present in the form of oxides, thus the only undetected element is oxygen and its abundance can be estimated from the stoichiometric ratios to the detected elements. In practice, an initial estimate is done of the total oxygen abundance (which is often around 40%), in order to make a first approximation of the matrix effects in the computations described in section 2.2.2. Iterative procedures then lead to a self-consistent dataset of all elemental abundances, oxygen included.

Typical problems that have been addressed by micro-PIXE in the field of earth sciences are trace element partitioning between phases, which provides important clues to the physical and chemical parameters of the environment at the time of rock formation, and mineral zoning effects as regards trace elements. The latter frequently do not follow the pattern of major elements, and the different behavior is of relevance to geologists. These investigations require a beam of small dimensions; however, this does not necessarily imply the use of a microbeam. Beam sizes of a few tens of microns may often be adequate. One should keep in mind that the depth within the mineral, interrogated by the proton beam, is of the order of 10 μm, or more for the medium-to-heavy elements typically searched for in trace quantities. A higher lateral resolution would therefore bring no benefit (section 2.7). The relatively broad size of microbeam required means that no specially refined tuning is needed on the set-up.

More traditional, broader beam sizes have been used to investigate rocks to obtain bulk composition data. Here, the obvious competitor to PIXE is XRF which, when operating on a larger spatial scale, is capable of yielding trace element abundance. Indeed, the same kind of problems are addressed by XRF and PIXE, and even the same samples can be used. As the problem in bulk composition analysis is the inhomogeneous structure of the rocks, the latter have to be thoroughly ground to a very fine powder and homogenized. The final sample preparation step is usually pressing the material into pellets, which are then exposed to the X-ray or proton beam. Due to the trace element capability of both techniques, great care must be taken to avoid contamination by the instruments used for grinding and pressing. Although in principle both XRF and PIXE are capable of yielding global information on composition, it must be said that the former technique has better MDLs, if properly applied, for the higher-Z elements, whereas PIXE is better for the lighter ones. Another difference lies in the different depth investigated, but

this is not relevant for homogenized samples because the pellets are effectively infinitely thick for both techniques. Finally, a common feature of XRF and PIXE analyses, which is disappointing to geologists, is that the important problem of rare earth patterns cannot be investigated. On the one hand, the detection efficiency for the K-lines of rare earth elements (which are sufficiently energy separated from each other) is too low, so that they cannot be detected at their low levels of concentrations. On the other hand, the L-series X-rays, which might instead be detected with more efficiency, cannot be safely deconvoluted because the overlaps of their multiplets are too close. Some work has been done to overcome these problems. Higher-energy PIXE has been attempted, but the advantage of the increase in K-series X-ray production cross-sections is accompanied by a much higher continuous background, induced in the spectra by the onset of many nuclear reactions. The latter lead to emission of gamma rays, producing Compton interactions in the detector (section 3.1). Also the attempt for a proper deconvolution of the close-lying multiplets of L-series X-rays was basically unsuccessful, as even small differences in the assumed X-ray lineshape in the spectrum fitting procedure alter the deduced relative intensities of the various elements.

When considering broad-beam PIXE applications in the earth sciences, not only XRF but all the other traditional techniques can usually be applied as well. This is because there is usually plenty of material to analyze and the nondestructive character of XRF and PIXE becomes less important. The advantage of XRF and PIXE is in their multi-element capability, which makes the compositional characterization of samples from bulk materials much faster. However, as a summary consideration one should admit that the real benefit of PIXE to the earth sciences is only when using microbeams for trace-element distribution studies.

3.4 Applications to Art and Archaeology

Recent surveys of PIXE applications in the field of art and archaeology can be found.[47,48] These applications probably started later than those mentioned above, but have now been developed to a large extent as a result of more widespread use of external beam set-ups. The importance of PIXE measurements (and IBA in general) in this domain is demonstrated by the installation of a dedicated accelerator in the Laboratoire de Recherche des Musées de France, in the basements of the buildings of the Louvre Museum in Paris.[49–51] With the prominent use of PIXE analyses among others, this accelerator has produced significant contributions to a better understanding of the materials and techniques used in the past for the production of works of art. Intense and significant activity in this field has also been performed in many other laboratories throughout the world.

The problems addressed by PIXE analysis in this field fall into two categories: first, an advancement of the knowledge of materials and production technologies used in works of art in ancient times; second, the investigation of materials prior to restorations, so as to appropriately choose the restoring procedure, and avoid incompatibilities and irreversible effects between materials.

The motivating forces within the first category can be many. Reconstructing the material composition of artifacts may lead to:

- deeper insight into the technical developments of the past, allowing the historian to better understand the advance in the technological skills of artisans at various sites and times;
- discovery of the sources of supply of raw materials, thus contributing to the solution of historical problems, such as the existence of certain trade routes at the time of manufacture of the ancient objects under analysis;
- indirect dating through compositional analogies with dated materials and, under favorable circumstances, lending support to attributions to specific artists, providing a criterion for authentication (or refutation).

To the second category of problems, namely the restoration of works, should be added the significant contribution to a better understanding of the deterioration of the works of art.

Whatever the reason for analyzing the material of an object of artistic or archaeological relevance, in almost all cases a mandatory requirement is that no damage must occur, and quite often the physical removal of some material for analysis is forbidden. This raises a difficulty in that the inherent inhomogeneity of the materials used to produce works of art (whether one is dealing with paints, or metal alloys, or terracotta, or any other) calls for an analysis extended to many samples. The representativeness of the results might be otherwise questioned. Hence there is a need to perform such analyses with techniques capable of yielding a full compositional characterization of artifacts in a fast and nondestructive way. PIXE meets these requirements probably better than any other analytical technique. With the exception of metal objects, external set-ups are needed so that the nondamaging character of the technique is thus fully preserved provided that sufficiently low beam currents are used. Low beam intensities, on the other hand, are feasible in most investigations because of the nonthin targets (i.e. the quantity of material is relatively large) and because the problem is often not the

search for trace elements. Typical beam currents used in this kind of application are well below 1 nA, sometimes down to a few picoamps.

In terms of risks of damage, the most sensitive precious materials are glass, porcelains, and any vitrified surface. With these materials, even with external beams, currents of a few tens of picoamps can sometimes produce a visible stain after a measurement. This sort of damage can often be tolerated, but the curator of the object must be warned about the risk. This damage is attributed to atomic dislocations in the structure of the material, giving rise to color centres. Stains induced in this manner have been often reported to fade away after some days or weeks, due to a self-annealing process, even at room temperature. To minimize the effect, whenever the sample size makes it possible, a diffused beam over a larger area may be employed, or a smaller-size beam may be scanned across the surface (or the surface moved under the beam).

Although all kinds of materials may be analyzed by PIXE, most published work has concentrated on gold artifacts (coins, jewels), temperas used in miniatures in illuminated manuscripts or incunabula, and terracotta. Relatively little PIXE activity has concerned the materials of traditional paint on wood or canvas. This is because PIXE is a surface analysis technique, and little or nothing can be discovered about the deeper layers. Thus, if PIXE is performed on the paintings as they are, the protective varnish, which is invariably laid over the color in paintings on wood or canvas, prevents access to the paint materials. Material sampling is seldom allowed; alternatively, the varnish could be removed, but this is only done on the occasion of restorations, and is not normally permitted for general investigations, such as when trying to characterize the color palette of a given painter or school.

For paints, the field where PIXE is best exploited is that of illuminated ancient books. Here, no varnish was used because the paints were believed to be naturally protected by remaining unexposed to direct light and atmospheric agents for most of the time. Much PIXE work has concentrated on this subject,[25,52–54] and unexpected results of art-historical importance have been found; for instance, the widespread use of ultramarine blue has been detected in Italian scriptoria dating from early medieval times. Until this finding it was believed that this pigment (obtained by a sophisticated refining technique from lapislazuli, a semiprecious stone only found in Afghanistan) was reserved for very important works, at least until the early Renaissance, because of the high cost of import from a distant site.

All PIXE work on miniature paints has been performed with external beam set-ups. This is essential with manuscripts, because of handling considerations and because only by the use of an external beam (in a helium flow) can the measurements be really nondamaging, in all senses.

The surface character of PIXE analysis also explains why most of the published work on metal products concerns gold. The surface patina layer covering ancient objects made of other metals is a problem to PIXE, unless the material under study is the patina itself (which can in fact be the case sometimes). As far as gold items are important work includes that by Demortier,[55–57] which led among other things to the discovery of skilful soldering techniques as early as 3000 years ago. This work demonstrated the use of various gold alloys with decreasing melting temperatures, which avoided the unsoldering of nearby previously soldered details. These measurements have been performed both with external beams and with in-vacuum microbeams. For this kind of object, indeed, no disadvantage (damage, or difficult handling due to large size) comes from the use of a vacuum set-up, whereas the use of an in-vacuum microbeam provides sometimes a better tool of analysis.

Terracotta has been often investigated by PIXE analysis. The information searched for in this case is usually a provenance clue, often based on the trace-element content of the material. Thus, the PIXE quality most exploited is its very low MDLs. When dealing with terracotta, the compositional analysis is usually performed on fragments, whose esthetical value is irrelevant; therefore, higher beam currents may be used with no fear of possible visible damage, and the sensitivity to trace elements achieved in runs lasting a few minutes. However, due to the surface character of PIXE analysis, caution must be exercised when deriving conclusions from the detection of traces – surface contamination might be the reason for finding trace concentrations of unexpected elements. These risks must be carefully evaluated and procedures for reducing them adopted (such as analysis of cleaved surfaces of the fragments, or preliminary cleaning with appropriate noncontaminant processes). The statistical analysis of detected trace-element abundances over large series of samples may also help in discriminating significant results from accidental findings. An alternative to direct irradiation is to sample some material, which in small amounts is often allowed when dealing with fragments. In these cases, the material can be powdered and pressed into pellets for analysis. The same considerations as for the analysis of rocks with macro-PIXE (section 3.3) apply in this respect, in particular about the alternatives to PIXE for these analyses.

An example of extensive PIXE investigations of this kind is the analysis of pottery and obsidian artifacts performed in Australia.[58] A large number of pottery samples have been analyzed with a PIXE/PIGE set-up; their composition was clearly correlated to local

sources of clay, which made it possible to establish relationships between prehistoric settlements and their trade connections. The obsidian artifacts, which were largely used by prehistoric populations as tools for everyday life or for hunting, were prepared using material collected from local deposits of volcanic origin. The compositional characterization of hundreds of these tools by PIXE and PIGE, followed by statistical analysis for pattern recognition, made clear separations into different groups possible. Because the populations took their tools with them when migrating from one land to another (even over distances of thousands of kilometers), the archaeologists were able to reconstruct the processes of migration by matching the composition of obsidian tools to that of volcanic flows in quite different places.

3.5 Applications to Historical Problems

Although related to the previous section, the studies described below have been separated out because, for them, the use of PIXE has been particularly useful and in certain respects unique. The study of ancient written documents, namely of the composition of paper, parchment, and especially ink, is a tool to characterize and differentiate them. Such characterizations address problems of chronology, authenticity, or technical procedures for the production of books in ancient times. For the study of inks in particular, PIXE is an exceptionally suited tool owing to its high absolute sensitivity, its nondestructivity, the possibility of using beams well below 1 mm in diameter in external set-ups and to perform simultaneous quantitative analysis of many elements. All these features make it really useful for these problems.

In a study of one exemplar of the Gutenberg Bible,[59] the paper and ink were analyzed by an external 1-mm^2 beam. The ink was notable for its high Cu and Pb content, and their weight ratio was used to characterize the ink. It was thus possible to reconstruct the sequence of printing of the various pages of the book.

Another study of the same group[60] concerned the so-called Vinland map, which is believed to be a pre-Columbus document giving the first cartographic evidence of North America. However, its authenticity had been debated and after a chemical analysis (which was performed on a forcedly limited number of tiny samples of ink, as they had to be picked up from the map), it was concluded that it must be a forgery because of a high content of TiO_2, which has become available only in recent times. An extensive PIXE analysis of ink from this document, performed instead on a very large number of spots, with no need for picking up material and no damage, showed that the amount of titanium was in fact quite small, thus re-opening the possibility that the map is authentic. In this case, the nondestructive character of a PIXE analysis in situ was the winning card.

Ink analysis by PIXE has also been used to indirectly reconstruct the chronology of documents, as in the following example. During his life Galileo produced, besides the published works, hundreds of notes, hand-written on loose sheets of paper, where he wrote propositions and sketched diagrams about his ideas on the problems of motion and on astronomy. These sheets are obviously undated and conflicting opinions have been raised within the community of science historians about their chronological order. Establishing their dates on the basis of an objective criterion is important for the reconstruction of the development of Galilean thought. Analysis with PIXE of the ink composition in some of Galileo's dated hand-written documents – such as letters – first showed[61] that a correlation can be established between different ink compositions and periods of Galileo's life. The inks are of the so-called metallo-gallic type and the quantitative ratios between the various metallic components (Fe, Ni, Cu, Zn, Pb) vary from one manuscript to another, so that they can be used to characterize different periods. The idea is then to characterize the ink of the undated propositions in the loose sheets, and attribute them to the different periods on the basis of composition matches to the dated documents. Even when an absolute date cannot be established, an important objective is a relative dating, i.e. differentiation between the inks of the various undated documents. Indeed, most of the controversies concerned whether certain notes had been written at the same or at different times. Using these criteria, both absolute and relative dating have been possible for a number of these loose sheets and controversies between historians of science have been solved.[62,63] It is important to keep in mind that this contribution is to be considered in the frame of a global approach where the historical and textual evidence is of course of primary importance. This calls for the need for integrated teams of researchers from the various fields to work together throughout all the phases of the investigation.

From a technical point of view, PIXE analysis of the inks in documents is easily performed with external beams. The typical beam size used is between 200 μm and 1 mm, with no need for microbeam set-ups. Because the elements of interest are not trace elements in the inks, but are major or minor components, the beam currents may be kept very low (typically 100–200 pA) and no damage occurs. In extracting quantitative data, two circumstances must be considered. First, the beam will also excite X-rays from the paper (or parchment) substrate, so that the result refers to an effective target that is a combination of ink and substrate. This is taken into account by measurements on spots of uncovered substrate, and appropriate subtraction of the backing contribution, after beam charge normalization of the

ink plus substrate and substrate-only spectra. Second, the ink actually penetrates the substrate, to an extent depending on many factors such as substrate porosity, ink fluidity at the very instant of writing, and even ageing phenomena. These effects can lead to ambiguities in that the target thickness and the matrix composition of the effective target are to some extent uncertain. However, as far as the ratio of the elements of interest (medium or heavy metals) is concerned, it can be shown that the value deduced with no matrix correction is only weakly dependent on these parameters. Therefore, the accuracy of the thin target approximation (sections 2.2.1 and 2.2.2) remains sufficient for a meaningful ink discrimination (about 10% for the measured ratios). The differences between different inks are much larger than that, even with changes of orders of magnitude having been observed.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
AMS	Accelerator Mass Spectroscopy
BS	Bremsstrahlung
EDS	Energy-dispersive Spectroscopy
EPMA	Electron Probe Microanalysis
FA	Factor Analysis
FET	Field Effect Transistor
fwhm	Full Width at Half-maximum
IBA	Ion Beam Analysis
LAMS	Laser Ablation Mass Spectroscopy
MDL	Minimum Detection Limit
NRA	Nuclear Reaction Analysis
PCA	Principal Component Analysis
PIGE	Particle-induced Gamma Ray Emission
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectroscopy
STIM	Scanning Transmission Ion Microscopy
WDS	Wavelength-dispersive Spectroscopy
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Proton-induced X-ray Emission in Environmental Analysis

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • Ion Scattering Spectroscopy in Analysis of Surfaces

Mass Spectrometry **(Volume 13)**
Tandem Mass Spectrometry: Fundamentals and Instrumentation

Nuclear Methods **(Volume 14)**
Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Rutherford Backscattering Spectroscopy

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist (eds.), *Particle-induced X-Ray Emission Spectrometry (PIXE)*, John Wiley & Sons Inc., Chichester, 1995.
2. S.A.E. Johansson, J.L. Campbell, *PIXE: a Novel Technique for Elemental Analysis*, John Wiley & Sons Inc., Chichester, 1988.
3. G. Pappalardo, J. De Sanoit, A. Musumarra, C. Calvi, C. Marchetta, 'Feasibility Study of a Portable PIXE System Using a ^{210}Po Alpha Source', *Nucl. Instrum. Methods.*, **B109/110**, 214 (1996).
4. L. Pappalardo, 'A Portable PIXE System for the In-situ Characterisation of Black and Red Pigments in Neolithic, Copper Age and Bronze Age Pottery', *Nucl. Instrum. Methods.*, **B150**, 576 (1999).
5. W. Bambynek, B. Crausemann, R.W. Fink, H.U. Freund, H. Mark, C.D. Swift, R.E. Price, P. Venugopalo Rao, 'X-ray Fluorescence Yields, Auger, and Coster-Kronig Transition Probabilities', *Rev. Mod. Phys.*, **44**, 716 (1972).
6. M.H. Chen, B. Crasemann, H. Mark, 'Relativistic K-shell Auger Rates, Level Widths, and Fluorescence Yields', *Phys. Rev.*, **A21**, 436 (1980).
7. M.O. Krause, 'Atomic Radiative and Radiationless Yields for K and L Shells', *J. Phys. Chem. Ref. Data*, **8**, 307 (1979).

8. J.H. Hubbel, National Institute of Standards and Technology Report NISTIR 89-4144, U.S. Dept. of Commerce, Gaithersburg, MD, 1989.
9. G. Lapicki, 'Cross-sections for K-shell X-ray Production by Hydrogen and Helium Ions in Elements from Beryllium to Uranium', *J. Phys. Chem. Ref. Data*, **18**, 111 (1989).
10. H. Paul, J. Sacher, 'Fitted Empirical Reference Cross Sections for K-shell Ionization by Protons', *At. Data Nucl. Data Tables*, **42**, 105 (1989).
11. M.H. Chen, B. Crasemann, 'Relativistic Cross Sections for Atomic K- and L-shell Ionisation by Protons, Calculated from a Dirac–Hartree–Slater Model', *At. Data Nucl. Data Tables*, **32**, 217 (1985).
12. M.H. Chen, B. Crasemann, 'Atomic K-, L- and M-shell Cross Sections for Ionization by Protons: a Relativistic Hartree–Slater Calculation', *At. Data Nucl. Data Tables*, **41**, 257 (1989).
13. J.L. Campbell, W.J. Teesdale, J.-X. Wang, 'Accuracy of Thick-target Micro-PIXE Analysis', *Nucl. Instrum. Methods.*, **B50**, 189 (1990).
14. W. Brandt, G. Lapicki, 'L-shell Coulomb Ionization by Heavy Charged Particles', *Phys.Rev.*, **A20**, 465 (1979).
15. W. Brandt, G. Lapicki, 'Energy-loss Effect in Inner-shell Coulomb Ionization by Heavy Charged Particles', *Phys.Rev.*, **A23**, 1717 (1981).
16. K.R. Akselsson, S.A.E. Johansson, 'X-ray Production by 1.5–11 MeV Protons', *Z. Phys.*, **266**, 245 (1974).
17. S.A.E. Johansson, T.B. Johansson, 'Analytical Applications of Particle-induced X-ray Emission', *Nucl. Instrum. Methods.*, **137**, 473 (1976).
18. D.D. Cohen, 'K and L Shell X-ray Cross Sections for Use in PIXE Analysis Systems', *Nucl. Instrum. Methods.*, **B49**, 1 (1990).
19. D.D. Cohen, M. Harrigan, 'K- and L-shell Ionisation Cross Sections for Protons and Helium Ions Calculated in the ECPSSR Theory', *At. Data Nucl. Data Tables*, **33**, 255 (1985).
20. F. Folkmann, C. Gaarde, T. Huus, K. Kemp, *Nucl. Instrum. Methods.*, **116**, 487 (1974).
21. K. Ishii, S. Morita, 'Continuous Backgrounds in PIXE', *Int. J. PIXE*, **1**(1), 1 (1990).
22. S.A.E. Johansson, 'Optimization of the Sensitivity in PIXE Analysis', *Int. J. PIXE*, **2**(1), 33 (1992).
23. E.T. Williams, 'PIXE Analysis with External Beams: Systems and Applications', *Nucl. Instrum. Methods.*, **B3**, 211 (1984).
24. B.L. Doyle, D.S. Walsh, S.R. Lee, 'External Micro-ion-beam Analysis (X-MIBA)', *Nucl. Instrum. Methods.*, **B54**, 244 (1991).
25. P.A.Mandò, 'Advantages and Limitations of External Beams in Applications to Arts and Archaeology, Geology and Environmental Problems', *Nucl. Instrum. Methods.*, **B85**, 815 (1994).
26. T. Calligaro, J.-C. Dran, B. Moignard, J. Salomon, 'Development of an External Beam Nuclear Microprobe on the AGLAE Facility of the Louvre Museum', *Nucl. Instr. Meth.*, **B161–163**, 328 (2000).
27. J.L. Campbell, 'Instrumentation, Fundamentals, and Quantification – Extraction of X-ray Intensities from PIXE Spectra', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 2.6, 1995.
28. J.L. Campbell, 'Instrumentation, Fundamentals, and Quantification – Detector Lineshape and Escape Peaks', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 2.4.3, 1995.
29. F. Watt, G.W. Grime (eds.), *Principles and Applications of High Energy Ion Microbeams*, Hilger, Bristol, UK, 1987.
30. G.W. Grime, F. Watt, *Beam Optics of Quadrupole Probe Forming Systems*, Hilger, Bristol, UK, 1984.
31. F. Watt, G.W. Grime, 'The High-energy Ion Microprobe', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 3, 1995.
32. J. Miranda, 'Studies of Atmospheric Aerosols in Large Urban Areas Using PIXE: an Overview', *Nucl. Instrum. Methods.*, **B109/110**, 439 (1996).
33. I. Orlic, X. Wen, T.H. Ng, S.M. Tang, 'Two Years of Aerosol Pollution Monitoring in Singapore: a Review', *Nucl. Instrum. Methods.*, **B150**, 457 (1999).
34. C. Boni, A. Caridi, E. Cereda, G.M.B. Marcazzan, 'A PIXE/PIGE Setup for the Analysis of Thin Samples', *Nucl. Instrum. Methods.*, **B47**, 133 (1990).
35. F. Calastrini, P. Del Carmine, F. Lucarelli, P.A. Mandò, P. Prati, A. Zucchiatti, 'External Beam PIGE for Fluorine Determination in Atmospheric Aerosol', *Nucl. Instrum. Methods.*, **B136/138**, 975 (1998).
36. G.D. Thurston, J.D. Spengler, 'A Quantitative Assessment of Source Contributions to Inhalable Particulate Matter Pollution in Metropolitan Boston', *Atmos. Environ.*, **19**(1), 9 (1985).
37. E. Swietlicki, S. Puri, H.C. Hansson, 'Urban Air Pollution Source Apportionment Using a Combination of Aerosol and Gas Monitoring Techniques', *Atmos. Environ.*, **30**(15), 2795 (1996).
38. K.G. Malmqvist (ed.), *Proceedings of the Eighth International Conference on PIXE and its Analytical Applications*, Lund (Sweden), June 1998, *Nucl. Instrum. Methods.*, **B150**, 1999.
39. E. Koltay, 'Elemental Analysis of Atmospheric Aerosols: Results and Perspectives of the PIXE Technique', *Int. J. PIXE*, **1**(2), 93 (1990).
40. T.A. Cahill, 'Compositional Analysis of Atmospheric Aerosols', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 5, 1995.

41. K.G. Malmqvist, 'Quantitative PIXE Analysis of Biomedical Material – Sample Preparation, Irradiation and Quality Control', *Nucl. Instrum. Methods.*, **B49**, 184 (1990).
42. K.G. Malmqvist, 'Biological and Medical Applications', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 4, 1995.
43. K. Ishii, H. Orihara, S. Morita (eds.), *Proceedings of the International Symposium on Bio-PIXE*, Sendai (Japan), July 1992, in *Int. J. PIXE*, **2**(3,4), 1992.
44. Z. Zengh, Z. Hu, L. Wei (eds.), *Proceedings of the International Symposium on Bio-PIXE*, Beijing (China), August 1996, in *Int. J. PIXE*, **6**(1,2), 1996.
45. J.D. MacArthur, Xin-Pei Ma, 'A Review of Particle-induced X-ray Emission in Geology', *Int. J. PIXE*, **1**(4), 311 (1991).
46. J.L. Campbell, 'Applications in Earth Sciences', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 6, 1995.
47. K.G. Malmqvist, 'Applications in Art and Archaeology', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 7, 1995.
48. M.A. Respaldiza, J. Gómez Camacho (eds.), *Applications of Ion Beam Analysis to Arts and Archaeometry*, Secretariado de Publicaciones de la Universidad de Sevilla, 1997.
49. Ch. Lahanier, G. Amsel, Ch. Heitz, M. Menu, H.H. Andresen (eds.), *Proceedings of the International Workshop on Ion Beam Analysis in the Arts and Archaeology*, Pont-à-Mousson (France), February 1985, *Nucl. Instrum. Methods.*, **B14**, 1 (1986).
50. G. Amsel, M. Menu, J. Moulin, J. Salomon, 'The 2 MV Tandem Pelletron Accelerator of The Louvre Museum', *Nucl. Instrum. Methods.*, **B45**, 296 (1990).
51. M. Menu, T. Calligaro, J. Salomon, G. Amsel, J. Moulin, 'The Dedicated Accelerator-based IBA Facility AGLAE at the Louvre', *Nucl. Instrum. Methods.*, **B45**, 610 (1990).
52. P. Del Carmine, M. Grange, F. Lucarelli, P.A. Mandò, 'Particle-induced X-ray Emission with an External Beam: a Non-destructive Technique for Material Analysis in the Study of Ancient Manuscripts', in *Studi e Testi*, eds. M. Maniaci, P.F. Munafò, Biblioteca Apostolica Vaticana, Città del Vaticano, 7–27, Vol. 358, 1993.
53. M. Bernasconi, R. Cambria, L. Dal Poz, P. Del Carmine, M. Grange, F. Lucarelli, J.D. MacArthur, P.A. Mandò, 'Analyse des couleurs dans un groupe de manuscrits enluminés du XII au XV siècle avec l'emploi de la technique PIXE', in *Studi e Testi*, eds. M. Maniaci, P.F. Munafò, Biblioteca Apostolica Vaticana, Città del Vaticano, 57–101, Vol. 358, 1993.
54. C. Neelmeijer, W. Wagner, H.P. Schramm, U. Thiel, 'De re metallica (G.Agricola) – IBA on air', *Nucl. Instrum. Methods.*, **B99**, 390 (1995).
55. G. Demortier, 'Microbeam Applications in Archaeology', in *Principles and Applications of High Energy Ion Microbeams*, eds. F. Watt, G.W. Grime, Hilger, Bristol, UK, 1987.
56. G. Demortier, 'Ion Beam Analysis of Gold Jewelry', *Nucl. Instrum. Methods.*, **B64**, 481 (1992).
57. G. Demortier, 'Ion Beam Studies of Archaeological Gold Jewellery Items', *Nucl. Instrum. Methods.*, **B113**, 347 (1996).
58. P. Duerden, E. Clayton, J.R. Bird, D.D. Cohen, 'Recent Ion Beam Studies on Archaeology and Art', *Nucl. Instrum. Methods.*, **B14**, 50 (1986).
59. B.H. Kusko, T.A. Cahill, R.A. Eldred, R.N. Schwab, 'Proton Milliprobe Analyses of the Gutenberg Bible', *Nucl. Instrum. Methods.*, **B3**, 689 (1984).
60. T.A. Cahill, R.N. Schwab, B.H. Kusko, R.A. Eldred, G. Möller, D. Dutschke, D.L. Wick, A.S. Pooley, 'The Vinland map, Revisited: New Compositional Evidence on Its Inks and Parchment', *Anal. Chem.*, **59**, 829 (1987).
61. L. Giuntini, F. Lucarelli, P.A. Mandò, W. Hooper, P.H. Barker, 'Galileo's Writings: Chronology by PIXE', *Nucl. Instrum. Methods.*, **B95**, 389 (1995).
62. F. Lucarelli, P.A. Mandò, 'Recent Applications to the Study of Ancient Inks with the Florence External-PIXE Facility', *Nucl. Instrum. Methods.*, **B109–110**, 644 (1996).
63. P. Del Carmine, L. Giuntini, W. Hooper, F. Lucarelli, P.A. Mandò, 'Further Results from PIXE Analysis of Inks in Galileo's Notes on Motion', *Nucl. Instrum. Methods.*, **B113**, 354 (1996).

Prompt γ-Neutron Activation Analysis

Carlos Oliveira, José Salgado, and Frederico G. Carvalho
ITN-Nuclear and Technological Institute, Sacavém, Portugal

1 Introduction **1**
2 Theory and Applicability of the Method **2**
2.1 Physical Principles 2
2.2 Detectability and Sensitivity 5
2.3 Prompt Versus Delayed Neutron Activation Analysis 6
3 Equipment **7**
3.1 General 7
3.2 Neutron Sources 7
3.3 Shielding 9
3.4 Detectors 11
3.5 Signal Processing and Data Handling 11
4 Instrument Design and Calibration **13**
4.1 Experimental Calibration 13
4.2 k_0 Method 13
4.3 Monte Carlo Library Least Squares Technique 14
4.4 Monte Carlo Simulation of Calibration Curves 14
4.5 Design Optimization 15
5 Applications of the Method **15**
5.1 Introduction 15
5.2 Borehole Logging 15
5.3 Industrial On-line Applications 16
5.4 Medical In Vivo Applications 17
5.5 Determination of Environmental Contaminants 19
5.6 Explosives and Drugs Detection Systems 19
5.7 Other Applications 19
6 Present Trends of Development **20**
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **21**

Prompt γ-neutron activation analysis (PGNAA) is a multielemental analytical technique based on the detection of γ-rays emitted by excited nuclei following neutron radioactive capture or neutron inelastic scattering. The accumulated count rate relative to the γ-ray energy characteristic of a certain isotope depends on its concentration. Measurements take place during sample irradiation. The technique can provide accurate information on the elemental composition of materials in a broad range of cases: raw materials, mineral exploration, contaminants, detection of forbidden products, such as drugs and explosives, inspection of containers and applications in health care. Among the main advantages of the PGNAA technique are: faster results; nondestructiveness; flexible sample requirements; bulk sample analysis; in situ application by direct on-line measurement; detectable unique signatures for most elements; improved accuracy for short-lived elements; low residual activity. On the other hand, the most significant limitations are: higher detection limits (DLs) as compared with delayed γ-neutron activation analysis (DGNAA); longer irradiation times; single (one-cycle) sample irradiation; impracticability of enhancing a specific isotope contribution by radiochemical or physical (half-life) separation techniques; requirement for relatively larger samples.

1 INTRODUCTION

PGNAA is a powerful analytical technique that can provide accurate information on the elemental composition of materials in a broad range of cases: raw materials, mineral exploration, contaminants, detection of forbidden products, such as drugs and explosives, inspection of containers (detection of plutonium) and applications in health care.

When a material is bombarded with neutrons, nuclear interactions in the sample may lead to the formation of excited compound nuclei that lose their excitation energy by the emission of one or more γ-rays (prompt γ-rays) and eventually one or more particles. PGNAA is based on the detection of the emitted γ-rays. The γ-ray energy identifies the element while the intensity of the line is a measure of the element concentration.

The analysis is the result of a combination of a number of different processes or steps:

- neutron emission by a neutron source (radioisotope, neutron generator, nuclear reactor, spallation source);
- neutron interaction with the sample's constituent nuclei;
- emission of γ-rays following neutron capture or neutron inelastic scattering;
- γ-ray attenuation in the material;

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

- detection of γ-rays (solid state or scintillation detector);
- conversion of detector counts in percent weight for the elements being analyzed based on a previous calibration of the system.

These steps and processes will be addressed in the following sections.

The PGNAA technique is best suited for measuring elements: (1) with very high neutron capture cross-section; (2) elements that decay too rapidly to be measured by conventional DGNAA; and (3) elements that through neutron reactions generate only stable isotopes or radioactive isotopes with a low probability of γ-decay.

The main advantages of PGNAA are: faster results; nondestructiveness (sample preparation is not needed); flexible sample requirements; bulk sample analysis (representative average for nonhomogeneous materials); in situ application by direct on-line measurement; detectable unique signatures for most elements; improved accuracy for short-lived elements; low residual sample activity.

On the other hand, the most significant handicaps are: higher DLs as compared with DGNAA, which restricts the number of elements that can be measured at trace concentrations; longer irradiation times; single (one-cycle) sample irradiation; impracticability of radiochemical or physical (half-life) separation to enhance specific isotope contributions; requirement for relatively larger samples.

The gauge response of a PGNAA analyzer is determined in a complex manner by the system parameters – geometry, source activity and neutron energy spectrum, and detector efficiency – and by certain properties of the analyzed sample such as bulk density and water content. It follows that the count rate caused by a given γ-ray line is a nonlinear function of the concentration of the element to be measured. There are, in principle, several methods for calibrating a PGNAA gauge. Experimental calibration methods use standards of known composition. An alternative method is the use of Monte Carlo codes to simulate the sample spectra and/or to establish calibration curves.

Trends in the evolution of PGNAA relate to both the instruments of the technique and the range of its applications.

The use of advanced neutron sources, improved spectrometers (Compton suppression and pair spectrometer), nuclear imaging techniques, sophisticated calibration methods and the extension of PGNAA to new materials and sample types are some of the trends which indicate that this technique has a promising future.

2 THEORY AND APPLICABILITY OF THE METHOD

Neutrons are electrically neutral; they are not affected by atomic electrons or by the positive charge of the nucleus. Neutrons can pass through the atomic cloud and interact directly with the nucleus; they are highly penetrating in most materials. On the other hand, the attenuation coefficient of γ-rays for most elements decreases monotonically with energy in the range from 0.1 to 10 MeV so that γ-rays with energy > 2 MeV are also relatively penetrating. Thus neutron–gamma based techniques do not suffer from the limitations of penetration of X-rays and β-rays and are favored for investigating bulky materials. Sample preparation is not required. All the major elements can be assayed provided a high-resolution detector is used.

Neutron activation analysis (NAA) is an important elemental analytical technique with many applications in different fields: industry, medicine, resource exploration, forensic work, and so on.

2.1 Physical Principles

When a material is bombarded with neutrons, a neutron can be captured by a nucleus, forming a compound nucleus. The compound nucleus remains for a certain time in a highly excited state due to the addition of the binding energy of the neutron to the nucleus. It will lose this energy by the emission of one or more γ-rays (prompt γ-rays) and eventually one or more particles,

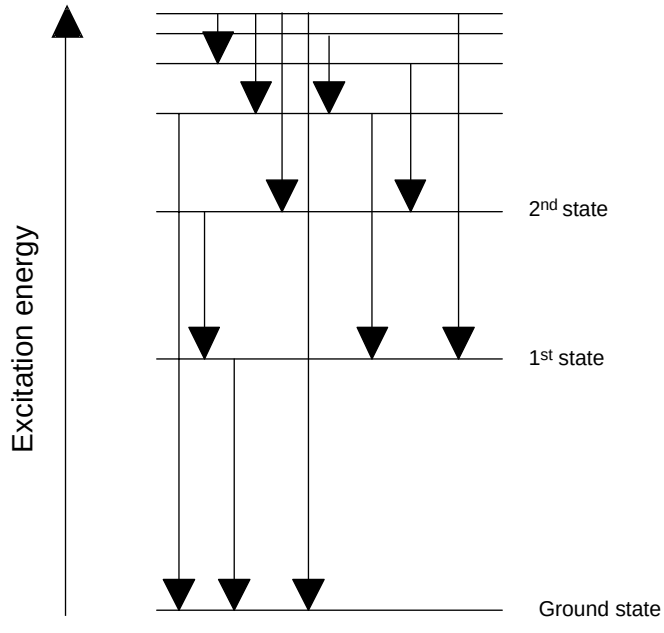

Figure 1 Decay scheme of an excited nucleus.

namely one neutron (inelastic scattering), which can be accompanied by several γ-rays. After the emission of this primary radiation, the daughter nucleus can remain unstable (radioactive), decaying to the ground state by a β-decay process, followed by γ-emission (delayed γ-rays).

As shown in Figure 1, the de-excitation of the compound nucleus may take the form of a stepwise process, which can be very complex, originating a spectrum with a large number of lines (89 for Si, 449 for Cl, 113 for Ca and 187 for Fe). Most of these lines have a very small relative intensity. The large number of emitted photons is due to (1) the high binding energy of the compound nucleus and (2) the narrowing of the energy gap between levels with increasing excitation energy. Two other causes of the increase in the number of lines in a spectrum must also be taken into account: (1) every isotope has its own decay scheme and (2) the high-energy γ-photons may interact through pair production in the detector. The subsequent positron annihilation originates the emission of two 0.511 MeV γ-rays. One or both of these γ-rays may escape the detector. This means that every γ-line may be split in three lines (the full f, with energy E_γ; the single escape s, with energy $E_\gamma - 0.511$ MeV; and the double escape d, with energy $E_\gamma - 1.022$ MeV. The γ-ray spectrum is a signature of the sample elemental composition. The intensity of the lines is proportional to the weight content of the respective element.

Prompt γ-rays are emitted within 10^{-15}–10^{-14} s of the interaction; delayed γ-rays originate in the decay of induced radioactivity. Two types of activation analysis can, then, be used: PGNAA when information is derived from prompt γ-rays and conventional instrumental neutron activation analysis (INAA) or DGNAA, where information is derived from delayed γ-rays (Figure 2). In the second case, the irradiated samples are normally removed from the irradiation position to detect the delayed γ-rays.

PGNAA is based on the analysis of a few γ-lines for each element. The selection of these lines takes into account their intensity and interference from other lines.

Two physical mechanisms predominate in PGNAA: thermal neutron capture reactions (n,γ) and fast neutron inelastic scattering (n,n′γ).

Thermal neutron capture	${}^{A}_{Z}\mathrm{X}(\mathrm{n},\gamma){}^{A+1}_{Z}\mathrm{X}^{*}$	${}^{27}_{13}\mathrm{Al}(\mathrm{n},\gamma){}^{28}_{13}\mathrm{Al}$ $E_\gamma = 7.724$ MeV

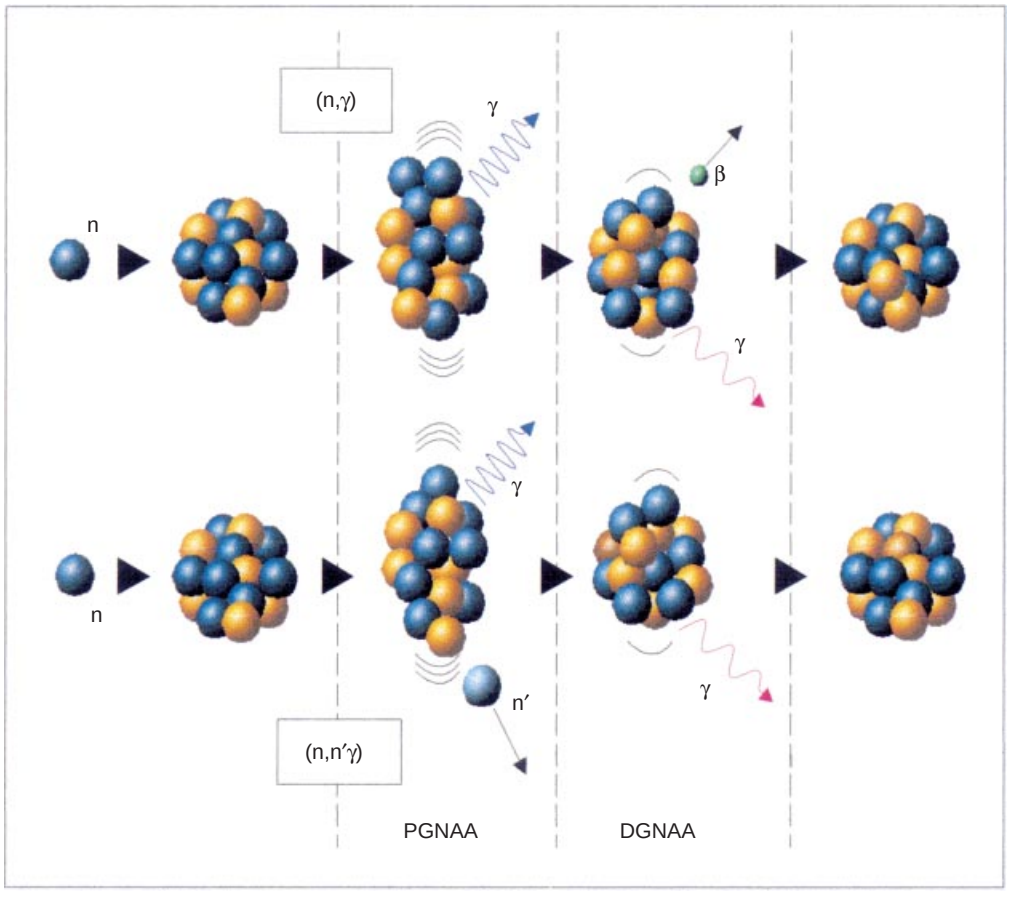

Figure 2 Basic processes in PGNAA and DGNAA.

Table 1 Microscopic cross-section of selected nuclei for thermal neutron capture and 14 MeV inelastic neutron scattering reactions and prompt γ-ray energies[a]

Reaction	$\sigma_{c,th}$ (barn)	$\sigma_{i,14\,MeV}$ (barn)	Dominant line energies E_γ (MeV)
$^{1}_{1}H(n,\gamma)^{2}_{1}H$	0.32	–	2.223
$^{12}_{6}C(n,\gamma)^{13}_{6}C$	0.0034	–	4.945, 3.684, 1.261
$^{12}_{6}C(n,n'\gamma)^{12}_{6}C$	–	0.425	4.439
$^{16}_{8}O(n,n'\gamma)^{16}_{8}O$	–	0.474	6.128, 6.049, 7.117
$^{14}_{7}N(n,\gamma)^{15}_{7}N$	0.075	–	10.828, 6.321, 5.532
$^{32}_{16}S(n,\gamma)^{33}_{16}S$	0.491		0.841, 5.420, 2.380
$^{32}_{16}S(n,n'\gamma)^{32}_{16}S$		0.500	2.230, 4.282
$^{27}_{13}Al(n,\gamma)^{28}_{13}Al$	0.232	–	7.724
$^{56}_{26}Fe(n,\gamma)^{57}_{26}Fe$	2.26	–	7.646, 7.632
$^{56}_{26}Fe(n,n'\gamma)^{56}_{26}Fe$		0.749	0.847, 1.238
$^{40}_{20}Ca(n,\gamma)^{41}_{20}Ca$	0.43		1.943, 6.420
$^{40}_{20}Ca(n,n'\gamma)^{40}_{20}Ca$	–	0.243	3.736, 3.904

$1\text{barn} = 10^{-28}\,\text{m}^2$.
[a] Reproduced from Mikesell et al.[1] with permission from IAEA.

Inelastic scattering $\quad {}^{A}_{Z}X(n, n'\gamma)^{A}_{Z}X^{*} \quad {}^{208}_{82}Pb(n,n'\gamma)^{208}_{82}Pb$

$$E_\gamma = 2.615\,\text{MeV}$$

For every nucleus there is a definite probability (cross-section) of thermal neutron capture ($\sigma_{c,th}$) and/or inelastic scattering ($\sigma_{i,14\,MeV}$), as is shown in Table 1.

The larger the cross-section the larger the number of nuclear reactions for a given isotopic abundance in the sample and a given neutron flux and, generally, the higher the sensitivity of the method.

For a homogeneous sample the count rate, R_{ij}, in γ-peak i from element j is given by the expression in Equation (1):

$$R_{ij} = \frac{N_A w_j}{4\pi M_j}\rho I_{ij}\varepsilon_i \int_{E_n}\int_{V}\Phi(E_n,\vec{r})\sigma_j(E_n) \times \int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dE_n\,dV \quad (1)$$

where N_A is the Avogadro number; w_j, the weight content of element j; M_j, the atomic mass; ρ, the bulk density ($m_j = w_j \times \rho$ is the mass of element j); I_{ij}, the number of photons pertaining to element j, with energy E_i, emitted per neutron captured; ε_i the detector efficiency for photons of energy E_i; $\Phi(E_n, \vec{r})$, the energy-dependent neutron flux, at position $\vec{r}$; $\sigma_j(E_n)$, the average element microscopic cross-section for the nuclear reaction that originates the γ-line of interest; $\vec{r}_d$, the path in the medium along the line drawn through volume element dV and the detection position; $\mu(E_i)$, the mass attenuation coefficient; and S, the detector surface.

When the energy dependence of the neutron flux inside the medium does not vary with position, the flux dependence on E_n and r can be separated. This happens, for instance, throughout the volume of a moderating material such as coal[2,3] when using ^{241}Am-Be, ^{252}Cf or 14 MeV neutron sources. Also the thermal and fast neutron fluxes can be formally separated and the neutron flux can be written as Equation (2):

$$\Phi(E_n,\vec{r}) = \Phi^0_{th}\Phi_{th}(E_n)\Phi_{th}(\vec{r}) + \Phi^0_F\Phi_F(E_n)\Phi_F(\vec{r}) \quad (2)$$

where th and F refer, respectively, to the thermal and the fast neutron fluxes. The energy dependence of the thermal component can be approximated by a Maxwell distribution and this will be indicated in the following formula by substituting M for th in $\Phi_{th}(E_n)$.

Equation (1) can thus be rewritten as Equation (3):

$$R^{I,C}_{ij} = \frac{N_A m_j}{4\pi M_j} I_{ij}\varepsilon_i\Phi^0_{th,F}\int_{E_n}\Phi_{M,F}(E_n)\sigma_j(E_n)\,dE_n \times \int_{V}\Phi_{th,F}(\vec{r})\int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dV \quad (3)$$

where the superscripts I and C refer to (n,n′γ) and (n,γ) reactions, respectively. Equation (3) gives the count rate, R^I_{ij}, arising from the fast component of the neutron flux or gives the count rate, R^C_{ij}, coming from the thermal component of the neutron flux. For most elements the thermal neutron capture predominates. In fact, in most applications the information derived from the detector count rate originates in thermal neutron reactions (thermal PGNAA); in certain special cases, however, fast neutrons are used to obtain information on the sample composition (fast PGNAA). Taking into account that the Maxwellian distribution has the expression in Equation (4):

$$\Phi_M(E_n) = \frac{1}{(kT)^2}E\exp\left\{-\frac{E}{kT}\right\} \quad (4)$$

where k is the Boltzmann constant and T the neutron temperature, Equation (3) takes the form, Equation (5):

$$R^C_{ij} = \frac{N_A m_j}{4\pi M_j} I_{ij}\varepsilon_i\Phi^0_{th}\int_{E_n}\frac{1}{(kT)^2}E_n\exp\left\{-\frac{E_n}{kT}\right\}\sigma_j(E_n)\,dE_n \times \int_{V}\Phi_1(\vec{r})\int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dV = \frac{N_A m_j}{4\pi}\frac{\sigma_{j,0}I_{ij}}{M_j}\varepsilon_i\Phi^0_{th}\int_{V}\Phi_1(\vec{r})\int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dV \quad (5)$$

where $\sigma_{j,0}$ is the element averaged $2200\,\text{m s}^{-1}$ neutron capture cross-section.

This equation shows that the count rate from a given line is proportional to the sample content of element j. However when a perturbation of the fast neutron flux is

caused by the sample, as happens currently in applications using isotopic neutron sources and large samples, the volume integral depends on the sample bulk density and composition, particularly on the water content. Fast neutrons emitted by the neutron source are generally moderated inside the sample. An increase of the bulk density and/or of the water content (or hydrogen content), will increase the thermal neutron flux and, consequently, the reaction rate. On the other hand, increasing the bulk density will increase the attenuation of photons inside the sample. The two effects have, thus, opposite signs and can cancel each other. An optimized design of the measuring arrangement (geometry, proper choice of sample size, amount of external thermalization) can make the volume integral essentially independent of sample density and composition.

In certain cases it is important to take into account the contribution of the epithermal neutrons to the detector count rate. To this end the thermal neutron cross-section can be substituted by the quantity shown in Equation (6):

$$g\sigma_{j,0} + I'_0 \frac{\Phi_{\text{epi}}}{\Phi_{\text{th}}} \tag{6}$$

where Φ_{epi} is the epithermal flux, g is the Wescott constant and I'_0 the resonance integral, Equation (7):

$$I'_0 = \int_{E_{\text{th}}}^{\infty} \sigma(E) \frac{\mathrm{d}E}{E} \tag{7}$$

2.2 Detectability and Sensitivity

In Equation (5) the quantity $S_{\text{f}} = I\sigma/M$ is constant for a given γ-ray pertaining to a given element. S_{f} is called *sensitivity factor* or *index* and is a useful measure of the relative sensitivity of the method for different elements under the same experimental conditions. Table 2 shows selected sensitivity factors of different elements when using gamma energies above 2.223 MeV.

Examination of the data shows that certain elements can be very easily detected by PGNAA. These include H, B, Cl, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, Gd, Dy, Au and Hg.

For thermal PGNAA to be appropriate for the detection of an element it is not sufficient that the element has a high sensitivity factor. It is important that the useful gamma lines of the element show a high spectral contrast,

Table 2 Sensitivity factor of various elements for thermal PGNAA

Element	Cross-section σ (barn)	Atomic mass, M	Energy (MeV)	Intensity, I (γ/n)	Sensitivity
H	0.332	1.008	2.223	1.00	0.329
B[a]	0.103	10.811	4.443	0.7596	0.00724
C	0.0034	12.011	4.945	0.6764	0.00019
N	0.0747	14.007	10.829	0.1412	0.00075
Na	0.400	22.990	6.395	0.2218	0.00386
Al	0.230	26.981	7.724	0.2743	0.00234
Si	0.160	28.086	4.934	0.6269	0.00357
Cl	33.2	35.453	6.111	0.2000	0.1873
K	2.10	39.102	7.770	0.0670	0.00360
Ca	0.430	40.08	6.420	0.3889	0.00417
Sc	27.2	44.96	8.175	0.1183	0.0716
Ti	6.10	47.90	6.760	0.2417	0.0308
V	5.04	50.94	7.163	0.1314	0.0130
Cr	3.10	52.00	8.884	0.2697	0.0161
Mn	13.3	54.94	7.244	0.1213	0.0294
Fe	2.55	55.85	7.645	0.2412	0.0110
			7.631	0.2851	0.0130
Co	37.2	58.93	6.876	0.0821	0.0518
Ni	4.43	58.71	8.999	0.3774	0.0284
Cu	3.79	63.54	7.915	0.3082	0.0184
Zn	1.10	65.37	7.863	0.1058	0.00178
Ag	63.6	107.87	5.698	0.0116	0.00683
Cd	2450	112.40	5.824	0.0213	0.464
Gd	4.9×10^4	157.25	6.750	0.0225	70.1
Dy	930	162.50	5.607	0.0293	0.168
Au	98.8	196.97	6.251	0.0552	0.0278
Hg	376	200.59	5.966	0.1386	0.2598
Pb	0.17	207.19	7.368	0.9405	0.00077

[a] In the case of B, the sensitivity can be enhanced (approx. 73) if the 0.48 MeV γ-ray from the reaction ^{10}B(n,α) is detected. This is usually only done when the low energy gamma background does not disturb the measurement.

i.e. a high ratio of the line intensity to that of neighboring lines pertaining to other elements.

The sensitivity factor gives qualitative information only: elements with high sensitivity factors will be more easily detected in a complex matrix. However, it is necessary to quantify the feasibility of measuring an element. Two quantities are normally considered: the absolute sensitivity (in counts $s^{-1}\,mg^{-1}$) and the DL (in $\mu g\,g^{-1}$).

The DL is related to the absolute sensitivity, S, through Equation (8):[4]

$$DL = 3.29\frac{\sqrt{R_b/t}}{S} \tag{8}$$

where R_b is the background count rate and t is the counting time.

Actual values of DL (or S) for a given arrangement and element are determined experimentally.

The two quantities, S and DL, depend on intrinsic sample characteristics related to nuclear properties of the elements (cross-section; number of γ-rays emitted and atomic mass; matrix interferences) and on parameters of the measuring arrangement (neutron spectrum and flux, sample dimensions, detector efficiency and resolution, background). For example, a reactor-based facility using a cold neutron source, guide tubes and a high-resolution germanium detector with Compton suppression, will normally show higher absolute sensitivity (lower DLs) than an arrangement using an isotopic neutron source for on-line measurement of large samples conveyed on a belt.

The actual DLs vary from one arrangement to another, depending on the neutron flux in the sample, the distance between sample and detector and the background due to shielding. The DL of an element can be affected by interferences from other elements present in the sample or in shielding or structural materials.

The elemental DLs for reactor-based PGNAA facilities are usually much lower than those of facilities with isotopic sources, used in on-line measurements.

As an example, Table 3 shows the DLs of the PGNAA facility installed at the Cold Neutron Research Facility of the National Institute of Standards and Technology (NIST).[5]

Table 3 Estimated spectral interference-free DLs for a 1 g sample counted for 24 h

Range ($\mu g\,g^{-1}$)	Element
0.01–0.1	B, Cd, Sm, Gd
0.1–1	Hg
1–10	H, Cl, In, Nd
10–100	Na, S, K, Ti, V, Cr, Mn, Co, Ni, Cu, As, Se, Br, Mo, Au
100–1000	Mg, Al, Si, P, Ca, Fe, Zn
1000–10 000	C, N, F, Pb

2.3 Prompt Versus Delayed Neutron Activation Analysis

As a laboratory analytical technique for routine sample analysis, PGNAA is generally complementary to the conventional delayed NAA systems installed at most research reactors. In contrast to DGNAA, the method is in principle applicable to all elements. As irradiation and measurement are carried out simultaneously, true multielement analysis is possible in a single measurement, at the expense of extremely complex γ-ray spectra.

DGNAA however is used intensively and is preferred for the determination of a large number of elements where no restrictions apply to handling and preparation of the sample. This is so because of the possibility of using higher neutron fluxes and of postirradiation counting that allow for background reduction and lower DLs.

In PGNAA arrangements installed at reactor facilities a wide variety of samples, e.g. liquids, solids, gases, museum objects, hazardous materials, can be irradiated using extracted beams that have fewer limitations on sample size and composition. In all cases neutron fluxes used in PGNAA are currently one or more orders of magnitude lower than those used in DGNAA.

The PGNAA technique is best suited for measuring elements: (1) with extremely high neutron capture cross-section; (2) elements which decay too rapidly to be measured by DGNAA; and (3) elements which produce only stable isotopes or radioactive isotopes with low gamma decay probability.

Sample preparation is not needed for PGNAA, and the results are obtained immediately. The irradiated samples may be large, which is advantageous when a representative average for nonhomogeneous material is needed. (The accuracy of the results may be affected by the presence of inhomogeneities that influence significantly the neutron flux distribution in the sample bulk and/or the position-dependent gamma attenuation.) However, large-sized samples may lead to problems due to neutron self-shielding, and specially due to neutron scattering, particularly in hydrogen.

The main advantages of the PGNAA technique are summarized in Table 4. Foremost among the advantages are nondestructiveness and the fact that nearly all elements yield a detectable signal. On the other hand, the most significant handicap is higher DLs, which restricts the number of elements that can be measured at trace concentrations and the long irradiation times often required.

Table 4 Advantages and disadvantages of PGNAA compared with DGNAA

Advantages	Faster results
	Nondestructiveness
	Low residual activity
	Flexible sample requirements
	Bulk sample analysis
	In situ application (on-line measurements)
	Detectable unique signatures for most elements
	Improved accuracy for short-lived elements
Disadvantages	Single (one-cycle) sample irradiation
	Longer irradiation time required
	Cannot use half-lives to enhance some particular isotope contribution
	Most elements not detectable at trace levels
	Relatively large sample required

While comparing different NAA techniques mention should be made of pulsed NAA, a technique that borrows some features from both PGNAA and for DGNAA. Pulsed neutron sources, e.g. neutron generators, are used so that between pulses the delayed gamma emission from the sample can be detected separately from the prompt gammas. This approach can lead to enhanced signal-to-background ratios in specific situations.

3 EQUIPMENT

3.1 General

According to the physical principle of PGNAA an operational arrangement must include a neutron source, a detection system with appropriate electronics and adequate shielding for radiological safety purposes. Generally, a dedicated computer forms the interface between the equipment and the user.

The specific objective for which the PGNAA device is designed (nature of the samples, elements to be measured, type of nuclear reactions, expected γ-ray spectrum) determines the type of source, the detector and the electronics as well as the manner in which they are assembled. An arrangement for on-line control of an industrial process in a factory or an analyzer installed at the end of a neutron guide tube have quite different designs.

3.2 Neutron Sources

The choice of the neutron source is determined not only by requirements concerning nuclear parameters such as neutron energy and flux, and half-life (when an isotopic source is considered), but also by the specific conditions regarding its utilization. Different kinds of sources are available: isotopic sources, neutron generators, nuclear reactors and spallation sources.

3.2.1 Isotopic Neutron Sources

There are two types of isotopic neutron source used in PGNAA: (α,n) sources and fission sources.

Sources of the first type are based in (α,n) reactions of isotopes that have large cross-sections for neutron production. Beryllium is an outstanding example. The sources consist of a mixture of beryllium and an α-emitter such as polonium or americium, the whole contained in a sealed capsule:

$$^{4}_{2}\mathrm{He} + ^{9}_{4}\mathrm{Be} \longrightarrow ^{1}_{0}\mathrm{n} + ^{12}_{6}\mathrm{C}$$

However, only a small number of α-particles (1 in 10^4) is effective in the production of neutrons so that the yield of the sources is relatively small. Also a background of γ-rays is present that is inconvenient for the application of the technique. Table 5 shows the main characteristics of several beryllium (α,n) sources.

Taking into account the neutron yield, the half-life of the α-emitter and the γ-background, the ^{241}Am isotope is currently preferred. There are ^{241}Am-Be sources commercially available up to 925 GBq.[6]

Table 5 Be (α,n) source characteristics

Nuclide	$T_{1/2}$	$E\alpha$ (MeV)	Yield ($n\,s^{-1}/10^6\,\alpha$)	Yield ($n\,s^{-1}\,Ci^{-1}$)	Yield ($n\,s^{-1}\,g^{-1}$)	Av. energy (MeV)	γ-dose at 1 m for $10^6\,n\,s^{-1}$ ($mGy\,h^{-1}$)
^{238}Pu	89 y	5.50	114	2.8×10^6	4.5×10^7	4.0	<0.01
^{239}Pu	24.4 y	5.14	65	1.6×10^6	1.2×10^5	4.59	<0.01
^{210}Po	138 d	5.30	73	2.5×10^6	1.1×10^{10}	4.54	<0.001
^{241}Am	458 y	5.48	82	2.2×10^6	6.5×10^8	4.46	0.01
^{244}Cm	18.1 y	5.79	100	3×10^6	2.5×10^8	4.31	<0.01
^{242}Cm	163 d	6.10	118	4×10^6	$\approx 10^{10}$	4.16	<0.01
^{226}Ra	1620 y	7.7 to 4.8	502	1.1×10^7	1.5×10^7	3.94	0.5
^{227}Ac	22 y	7.4 to 5.7	702	1.5×10^7	1.7×10^9	3.87	0.07

Spontaneous fission sources can be found in the region of the transuranium elements where the α-decay instability competes with spontaneous fission. However the half-lives of the transuranium elements are generally very short. The only nuclide with an acceptable half-life for use in applications is ^{252}Cf ($T_{1/2} = 2.645$ y). This nuclide decays by α emission in 97% of the cases and by spontaneous fission in the remaining 3%. The average number of neutrons emitted per fission is 3.76. A source containing 3.7×10^{10} Bq of ^{252}Cf has a neutron yield of 4.3×10^{9} n s^{-1}, which is three orders of magnitude higher than that of the more interesting (α,n) sources. The γ-dose at 1 m from a source emitting 10^{6} n s^{-1} is 0.0007 mGy h^{-1} (a factor of 10 smaller than for a ^{241}Am-Be source). The average neutron energy of a ^{252}Cf source is 2.3 MeV.

In Figure 3 neutron spectra and γ-spectra for Cf and Am-Be sources are plotted for comparison.[7]

The high specific activity of the Cf source is an advantage over the (α,n) sources. Other advantages are small size and low heat and gas generation, which decreases the risk of source damage for long-term operation.[8] These radioisotope sources are used in PGNAA devices designed for field work.

3.2.2 Neutron Generators

Sealed tube neutron generators[9,10] are based on the following reactions:

$$^{2}_{1}\mathrm{H} + {}^{3}_{1}\mathrm{H} \longrightarrow {}^{4}_{2}\mathrm{He}(3.6\,\mathrm{MeV}) + {}^{1}_{0}\mathrm{n}(14.1\,\mathrm{MeV})$$

$$^{2}_{1}\mathrm{H} + {}^{2}_{1}\mathrm{H} \longrightarrow {}^{3}_{2}\mathrm{He}(0.8\,\mathrm{MeV}) + {}^{1}_{0}\mathrm{n}(2.45\,\mathrm{MeV})$$

In generators emitting 14.1-MeV neutrons, a deuterium ion beam (or a mixture of deuterium and tritium ions) is accelerated up to between 70 and 300 kV and made to impinge on a tritiated target (or a deuterated and tritiated target). In the latter case 2.45-MeV neutrons are also produced. However, the emission of 14.1-MeV neutrons dominates. It is possible to obtain an emission rate of 10^{11} n s^{-1}.

Tubes emitting 2.45-MeV neutrons are based on deuterium–deuterium reactions, i.e. a beam of deuterium ions is accelerated up to a few hundred kilovolts and impinges on a deuterium target. Neutron emission of such tubes may reach 10^{10} n s^{-1}.

A great advantage of neutron generator tubes is the possibility of on/off switching, i.e. the production of neutrons is stopped whenever required. Another interesting characteristic is the possibility of pulsing the neutron beam.

The lifetime of a sealed tube is limited by target heating and target sputtering from ion beam impurities.

A new neutron generator source has been presented: the inertial electrostatic confinement (IEC) device. The

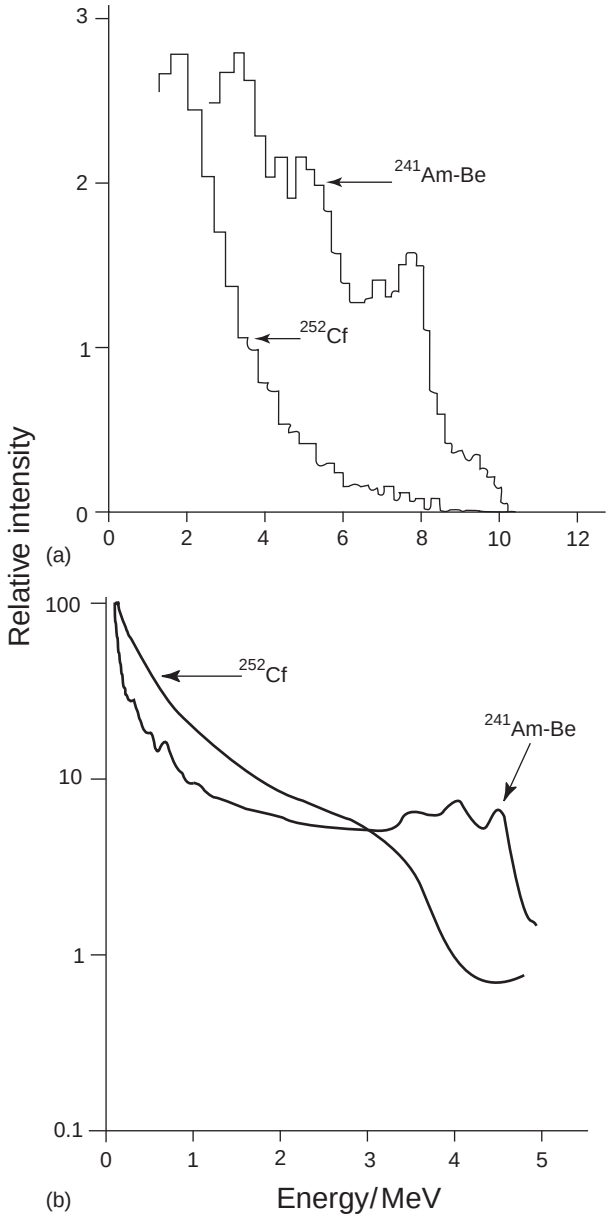

Figure 3 Neutron spectrum (a) and gamma spectrum (b) for ^{241}Am-Be and ^{252}Cf neutron sources. (Reproduced by permission from Amersham Buchler.)

IEC device is a neutron source where deuteron ions are accelerated producing fusion reactions as they react with a deuterium plasma target. Commercially available units deliver 10^{7} of 2.45-MeV n s^{-1} during steady-state operation. The neutron emission is isotropic. It is a practical low-to-medium-yield source for laboratory and industrial applications.[11]

3.2.3 Nuclear Reactors

Nuclear reactors are still the most important neutron sources. The thermal neutron flux of a reactor source

is of the order of 10^{13} n cm^{-2} s^{-1} up to 10^{15} n cm^{-2} s^{-1}. The neutron energy spectrum can be divided into three components: fast, epithermal and thermal. Their relative importance is dependent on the reactor structure. Usually, the thermal neutrons constitute the most intense fraction. Thermal neutrons are in thermal equilibrium with the atoms of the reactor moderator; their most probable energy equals kT (where k is Boltzmann's constant and T is the neutron temperature). At room temperature the most probable energy is 0.025 eV. Epithermal neutrons have energies from $5kT$ to 0.8 MeV. Fast neutrons have energies >0.8 MeV.

PGNAA facilities are usually installed at the extraction end of a neutron beam tube (sometimes using a neutron guide).[5,12–26]

The performance can be enhanced by using subthermal or "cold" neutrons.[12–14,16,19] These neutrons have energies lower than 10 meV. Cold neutrons are extracted from a cooled moderator placed in the neighborhood of the reactor core. The moderator may be liquid hydrogen or deuterium at about 30 K. Occasionally, liquid methane is also used at temperatures of about 100 K. The cold neutron absorption cross-section for most elements is about two to three times greater than that for thermal neutrons ($1/v$ behavior).

Owing to their long wavelength, about 0.3 nm, cold neutrons can be bent out of the flight path followed by higher energy radiations by using curved neutron guides. As a consequence, such cold neutron beams display a much lower γ and fast neutron background. Figure 4 shows the Cold Neutron Research Facility at NIST.

3.2.4 Neutron Spallation Sources

The operation of a neutron spallation source is, in principle, similar to that of an X-ray tube.[27,28]

Protons are accelerated to a high energy, in the range of hundreds of megaelectronvolts (typically 800 MeV up to 1 GeV) and directed on to a heavy metal target, such as uranium, tungsten, lead or a liquid metal (mercury). The result is excitation of the target nuclei. Their deexcitation is accomplished by the emission of several neutrons per incident proton (the number shows an energy dependence and may reach 30). Most of the neutrons produced have energies similar to those of a fission spectrum.

Neutron spallation sources can be continuous or pulsed. A typical value of the neutron flux from a continuous spallation source is 10^{13}–10^{14} n s^{-1} cm^{-2}.

Figure 5 shows the Paul Scherrer Institute spallation neutron source SINQ.[29]

3.3 Shielding

The shielding of a PGNAA facility must be designed according to the neutron source used. Shielding problems for on-line systems are completely different from those of an analyzer installed at the end of a neutron guide tube in a reactor hall. The shielding has two major functions: reducing the detector background so as to enhance the signal to background ratio and guaranteeing radiological safety.

Radiation coming directly from the neutron source as well as that originating in interactions within the sample and components of the system (structure, shielding) can be classified in four groups: (1) fast and epithermal neutrons, (2) thermal and cold neutrons, (3) primary γ-radiation and (4) secondary γ-radiation.

3.3.1 Fast and Epithermal Neutrons

The absorption cross-section of fast and epithermal neutrons is small for all materials. For this reason, it is necessary to slow down the neutrons before capture. The most effective shield for these neutrons is a mixture of a hydrogenous material and a thermal neutron absorber.

3.3.2 Thermal and Cold Neutrons

A few elements, such as boron, lithium, cadmium and gadolinium, have a high cross-section for thermal and cold neutrons. Compounds of these elements can be used to obtain an effective shielding against these neutrons. Neutron absorption in boron and cadmium is followed by the emission of γ-rays. This emission must be taken into account in the design of the shielding of the facility.

3.3.3 Primary γ-rays

Primary γ-rays are emitted directly by the neutron source. The problem is particularly acute in on-line analyzers using isotopic neutron sources. The source is generally surrounded by a layer of a strong γ-ray absorber such as lead or bismuth.

3.3.4 Secondary γ-rays

Secondary γ-rays originate in (n,γ) or (n,n′γ) reactions that take place within the sample or in shielding and structural materials. Only γ-rays from the sample directed towards the detector are wanted for the analysis. Adequate shielding must be provided to absorb all other γ-rays.

3.3.5 Composite Shielding

A good shielding material for neutrons is a mixture of a hydrogenous material and a good thermal neutron

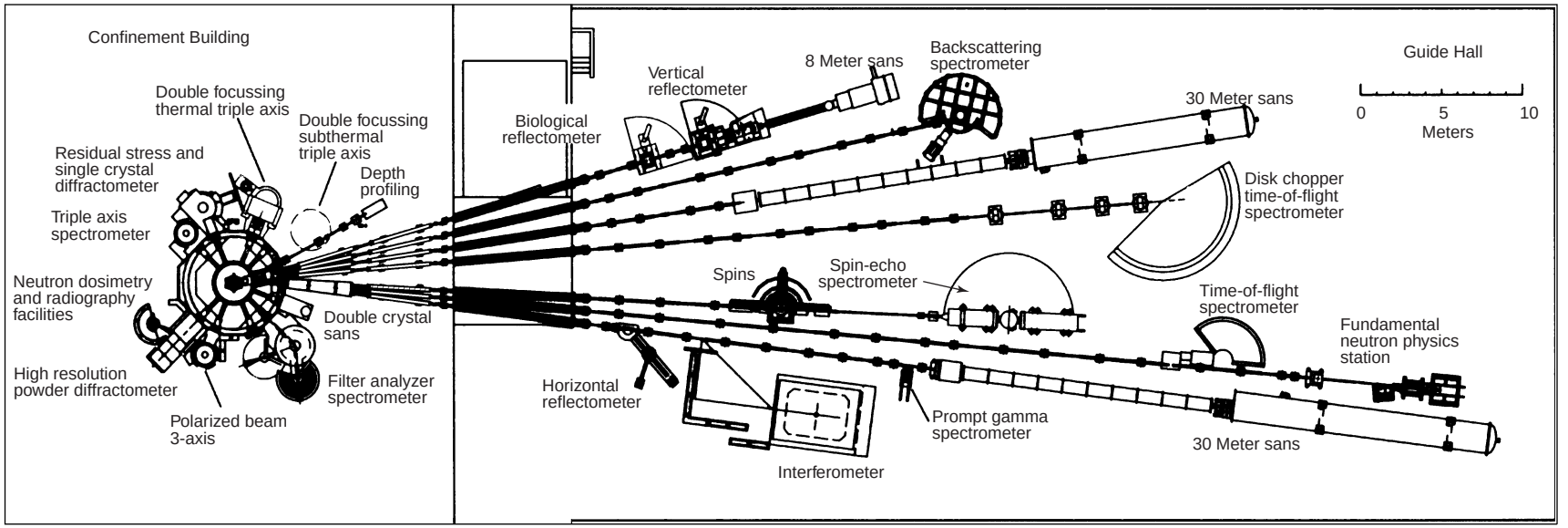

Figure 4 Experimental facilities around the NIST Reactor. A PGNAA facility is represented. (Reproduced from Paul et al.[12] with permission.)

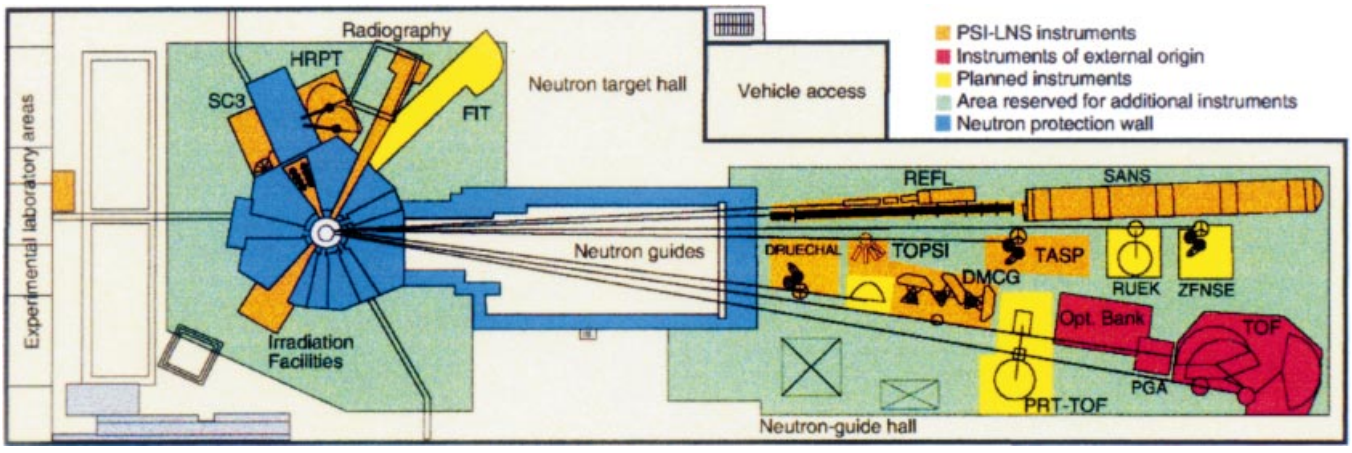

Figure 5 Experimental facilities around the Paul Sherrer Institute spallation neutronen quelle (SINQ). A PGNAA facility is represented on the right-hand side. (Reproduced from Paul Scherrer Institute Annual Report[29] with permission.)

absorber. Polyethylene with addition of boron or lithium has proven to make a good shielding because of the physical and chemical properties of the composite: high density of hydrogen atoms, chemical inertness and ease of machining. Lithium–polyethylene materials are very useful as detector shielding because lithium generates no capture γ-rays. Other hydrogenous materials, such as water or paraffin, are not advisable for industrial applications because of fire hazards and the possibility of leakage that can jeopardize safety.

A combination of lead and boron–polyethylene makes a material with good shielding properties for mixed neutron–γ fields.

3.4 Detectors

The criteria for the choice of a γ-ray detector must take into consideration efficiency, resolution, sensitivity to neutron damage and maintenance.

Essentially two types of detector are used in PGNAA. Scintillation detectors based on a thallium-doped sodium iodide detector, (NaI(Tl)), or a bismuth germanate detector ($Bi_4Ge_3O_{12}$), (BGO), and solid-state germanium detectors, normally the n-type hyper-pure germanium detector (HPGe). Resolution, efficiency and maintenance care are different for the two detector types.

The HPGe detector shows the best resolution, approximately 0.1% fwhm (full width at half-maximum). The NaI(Tl) scintillator has a resolution of about 5% at 2 MeV, or 3% at 10 MeV; the BGO scintillator has the poorest resolution, about a factor of two lower than that of NaI(Tl). The resolution of the HPGe detector enables the separation of closely spaced γ-lines in complex spectra.

Regarding efficiency, HPGe detector units currently used normally have lower efficiencies than BGO and NaI(Tl) detector units. However, HPGe detectors can have relative efficiencies (compared to a standard NaI(Tl) detector) of the order of 1, if a large enough germanium crystal is used, with the corresponding high cost. The HPGe detector is the most sensitive to neutron interactions within the crystal that promote the formation of traps causing resolution degradation. Owing to the mechanism of transport of charged particles inside the crystal, the n-type detector is less sensitive to neutron damage than the p-type.

Finally, HPGe detectors must be operated at low temperature (100 K) requiring expensive maintenance services. This is a drawback for on-line applications in industry.

Frequently an association of different detectors is used to optimize the arrangement (HPGe/NaI(Tl) or HPGe/BGO).

Detectors used for on-line industrial PGNAA are in general scintillation detectors. Reactor-based analyzers predominantly use HPGe detectors.

3.5 Signal Processing and Data Handling

The main objectives of signal processing and data handling at a PGNAA facility are to maximize the pulse throughput rates and minimize the γ-ray background without detriment to the energy resolution.

The characteristics of the signal processing system must be adapted to the type of detector and to the expected γ-ray count rates. Signal processing and data handling are carried out by a γ-spectroscopy unit.

The basic components of the γ-spectroscopy unit are the preamplifier, amplifier, analog to digital converter (ADC) and multichannel analyzer (MCA).

The function of a preamplifier is to extract the signal from the detector without significantly degrading

the intrinsic signal-to-noise ratio. The preamplifier is located as close as possible to the detector to minimize noise.

A germanium detector represents an extremely high capacitive impedance source (10 GΩ, 10–50 pF), generating a small charge ($Q = 2.7 \times 10^{-13}$ C) in the detection of a 5-MeV particle. The small output signal requires that the preamplifier does not add appreciable noise.

The absorption of a γ-ray by the detector produces a current pulse at the preamplifier input. The preamplifiers are typically of charge sensitive configuration and are coupled to the detector through the gate of a field effect transistor (FET). Two types of these preamplifiers can be used with Ge detectors: one employs the dynamic charge restoration method (resistor capacitor feedback) and the other the pulsed charge restoration method (pulsed optical or transistor reset) to discharge the integrator.

In many high count rate and high-resolution applications the transistor reset preamplifiers (TRP) offer the best performance.

Nowadays advanced hybrid-electronics manufacturing capability allows the incorporation of the detector element, preamplifier and high-voltage filter into a compact streamlined system.

Where the resolution is not a vital system parameter it is not necessary to use Ge detectors. In their place scintillation detectors can be used. A scintillation detector coupled to a photomultiplier tube generates moderately large signals at the output thus relaxing restrictions on the noise contribution from the preamplifier. They can be also associated with charge sensitive preamplifiers.

A main amplifier with sufficient linearity and large enough bandwidth extracts the energy information from the preamplifier output and converts it to a shape that can be easily analyzed by an ADC. The first stage of the amplifier (differentiator) reduces the decay time and blocks the low frequency or direct current components in the input signal. The following stage (integrator) recovers the energy information from the signal. This yields Gaussian-shaped pulses. The amplitude of each pulse is directly proportional to the energy of the γ-radiation.

To process an incoming pulse completely the amplifier must be adjusted for a time constant (TC) about 10 times the input pulse rise time. A shorter TC stops the processing before it has reached its maximum amplitude, which means loss of resolution in the data. But a longer TC means longer amplifier processing time, which reduces the throughput capabilities of the amplifier: if a new pulse reaches the amplifier before processing the previous one, a pile-up occurs and the piled up pulses are invalid events. Although there are ways to prevent the processing of these pulses, the net effect is decreased system output. As a shorter shaping TC minimizes pile-up, a TC of 2 μs is recommended for good resolution with Ge detectors. Its maximum throughput capability is just under 20 kcps (kilocounts per second), which is achieved at an input count rate of about 50 kcps. If the input rate exceeds that value, throughput actually decreases owing to pulse pile-up.

For high count rates the use of a gated integrator (GI) will allow a gain of the order of 10 kcps in throughput.

In order to ensure that the piled-up events are not processed by the ADC, additional circuits must usually be considered. These circuits are called pile-up rejector/live time corrector circuit (PUR/LTC). When a new pulse arrives before the previous one is processed, the new pulse is rejected.

The ADC unit converts the amplitude of each event into a digital value to be processed by the MCA or computer. Two types of ADC are available: the Wilkinson ADC and the fixed conversion time (FCT) ADC. The FCT ADCs can offer conversion times significantly shorter than those from the Wilkinson type at least for higher energy events. The choice depends on the specific application taking into account channel resolution and signal amplitude.

The MCA or a personal computer-based MCA acquisition board collects and stores the data, provides its visualization on a monitor, performs mathematical operations to generate the final results and transfers the raw and/or the treated data to an external storage device.

Current cheap memory circuitry provides adequate numbers of data channels and counting capacities as well as short enough data storage times (no more than a couple of microseconds) to deal with the highest count rates encountered in practice.

Hanna et al.[30] and Wormald[31] proposed the use of compound detectors for Compton suppression and pair spectrometry. The use of coincidence/anticoincidence techniques with compound detectors can provide simpler spectra by selecting only the useful events. Pair production and Compton scattering are the most important interactions for high-energy γ-rays. If the interaction occurs within the sensitive volume of the detector and the resulting secondary γ-rays escape the detector, a smaller amplitude pulse is recorded.

Schematically, pair spectrometers use three detectors: a central one in which the primary radiation is detected by pair production, and two others for detecting the 0.511-MeV annihilation photons. An event is recorded by using coincidence techniques if a photon of $E_\gamma - 1.022$ MeV is registered in the central detector and, simultaneously, two 0.511 MeV photons are detected in the others. This method improves the high-energy tail of the spectrum. Wormald proposed a close packing array of cylinders or

hexagonal prisms to form a large collecting area adapted to the shape of the sample.

The Compton suppression spectrometer uses anticoincidence techniques to eliminate events resulting from Compton scattering: only signals which do not have a coincidence signal in another detector are recorded. This technique greatly simplifies the low-energy part of the spectra. A γ-ray spectrometer operating simultaneously in Compton suppression and pair production modes acquires information in three modes: single mode, Compton suppression mode and pair mode. Analysis of the complex γ-ray spectrum can be performed by comparing the three spectra. The spectrometer performance is enhanced.

A wide variety of application programs is available for data handling of the spectra (nuclide identification, background and interference corrections, peak intensity and percent weight content calculation).

4 INSTRUMENT DESIGN AND CALIBRATION

The gauge response of a PGNAA analyzer is determined in a complex manner by the system parameters – geometry, source activity and neutron energy spectrum, and detector efficiency – and by certain properties of the analyzed sample such as bulk density and water content. It follows that the count rate caused by a given γ-ray line is a nonlinear function of the concentration of the element to be measured.

The density and the dimensions of the sample have a direct effect on reaction rates in the sample volume and on the attenuation of γ-rays and thus on the γ-count rate at the detector. For a fixed geometry and composition of the sample the thermal neutron flux depends on the bulk density. When the density increases, the thermal neutron distribution is ''compressed'' spatially with a higher maximum shifted towards the region of the sample adjacent to the neutron source. On the other hand, the relevant γ-rays on average travel a longer way to the detector inside the sample and are subject to a stronger attenuation. These two effects act in opposite directions upon the γ-count rate and can cancel each other for particular geometries.

Furthermore, the hydrogen content of the sample may also have a significant effect on the neutron flux distribution.

In principle, the composition of an unknown sample can be derived by solving Equations (3) or (5). In this direct or absolute method the concentration of each element is calculated using fundamental parameters, such as cross-sections, number of photons emitted by the element per neutron captured, detector efficiency, γ-ray attenuation coefficients in the sample and solid angle viewed by the detector.[32,33] However, this method is only applicable in a small number of cases owing to uncertainties in the available microscopic nuclear parameters, the unavailability of others and the influence of composition, density and sample volume on the count rate from a given γ-ray line. In activation analysis of large samples a number of correction factors must be applied (neutron absorption and moderation, γ-ray attenuation within the sample and detector efficiencies).

The simplest case is that of thin samples irradiated in a thermal or cold neutron beam. Neutron scattering and γ-absorption in the sample are, then, negligible thus reducing the influence of sample composition and homogeneity on the count rate.

The problem becomes more complicated for on-line gauges using isotopic neutron sources and large samples. Generally, there is neutron thermalization inside the sample so that the thermal neutron flux distribution is dependent on the sample composition, particularly on hydrogen content (moisture) and bulk density.

There are, in principle, several methods for calibrating a PGNAA gauge.

4.1 Experimental Calibration

Experimental calibration establishes the correlation between the γ-ray peak areas and the known chemical composition, density and water content of the samples. The experimental determination of the calibration curves demands long laboratory work, using a large number of standards with well-known compositions, matrix and geometry that closely match those of the sample.[34] The variation in a given parameter can affect the γ-count rate due to a given element, although the concentration of the element remains unchanged. It is a very time-consuming approach, because new standards must be prepared for each sample type.

4.2 k_0 Method

The k_0 method is used extensively in delayed NAA.[35] The k_0-factor is the ratio of the absolute sensitivity divided by the detection efficiency of the γ-rays of the element to be measured compared with that of a comparator element, s, Equation (9):

$$k_{0j} = \frac{R_j/m_j\varepsilon_j}{R_s/m_s\varepsilon_s} = \frac{\theta_j I_j \sigma_j/M_j}{\theta_s I_s \sigma_s/M_s} \tag{9}$$

where θ is the isotopic abundance of the nuclide.

In the absence of significant γ-ray self-absorption and for elements with neutron cross-section varying as $1/v$,

the k_0-factor is independent of neutron energy, sample matrix and geometry.

Measurement of k_0-factors allows in principle accurate determinations of relative concentrations of every element contained in the sample without the use of the comparator. In effect, the relative concentration of element j to element i in the sample is given by the formula, Equation (10):

$$\rho_{j,i} = \frac{m_j}{m_i} = \frac{R_{\gamma j}/\varepsilon_j k_{0j}}{R_{\gamma i}/\varepsilon_i k_{0i}} \tag{10}$$

When the concentration of one of the elements is known, the absolute concentration of any other element can be assessed. Otherwise it is necessary to measure only the sample and a single standard of similar properties (geometry, neutron scattering and γ-absorption), containing the comparator element to determine absolute concentrations.

There are some problems that can affect the accuracy of the k_0 measurements. Among them are deviations from the $1/v$ behavior of the neutron cross-sections, gamma attenuation in thick samples or with large concentrations of high-Z elements and heterogeneities in the sample.

The k_0 method has been applied by several authors to PGNAA measurements.[36–39]

4.3 Monte Carlo Library Least Squares Technique

Gardner et al. have developed the Monte Carlo library least squares method (MCLLS) for PGNAA.[40–42]

According to this model the PGNAA gauge is a ''black box''. The gauge response, R is represented by the function, Equation (11):

$$R = f(x, u, v, w, \ldots) \tag{11}$$

where x is the parameter to be measured and u, v and w represent variables that characterize the measurement interferences. Then, Equation (12):

$$x = F(R, u, v, w, \ldots) \tag{12}$$

In the linear least squares (LLS) model the spectrum of the sample is a linear combination of individual components. The content of each channel can be expressed as, Equation (13):

$$b_i = \sum_{j=1}^{m} x_j a_{ij} + E_i \quad i = 1, n \tag{13}$$

where b_i is the unknown sample spectrum or counting rate in channel i of the sample; x_j is the amount of component j in the sample; a_{ij} are the library spectra or counting rate per unit amount of component j in channel i; and E_i is the random error in channel i due to statistical counting rate fluctuations.

If the system is linear, i.e. if the gauge response to an individual component does not depend on other parameters, the concentration, x_i can be obtained by minimizing the reduced chi-square with respect to each amount.

However, as the PGNAA analyzers are usually nonlinear, the MCLLS approach has been developed. This is an iterative method that requires that a very accurate Monte Carlo code enables the calculation of the spectra of a sample given its composition, density and geometry. To this end the entire arrangement and the sample are simulated.

Simulations are performed in order to generate the elemental library spectra of all elements present in the sample. The model consists of the following steps:

1. The Monte Carlo code generates the pulse height spectrum of a sample whose composition and density are similar to that of the unknown sample.
2. The prompt γ-ray spectrum for each element is recorded separately to serve as a library spectrum.
3. The linear library least squares analysis is then performed using the spectra of the unknown sample and the calculated library spectra. The elemental composition obtained is compared with the initially assumed values.
4. If the calculated elemental composition of the unknown sample is close enough to the values assumed for the Monte Carlo simulation, these are taken as the final measured values. If not, the LLS values are taken for the next iteration and steps 1 through 3 are repeated until the linearity assumption is valid.

4.4 Monte Carlo Simulation of Calibration Curves

Another method relies on the Monte Carlo simulation of the measuring arrangement with samples having different compositions, densities, water contents and thicknesses, in order to establish the correlation between the count rate for unit concentration of an arbitrary element and the sample parameters.[43] A comparison is then carried out between the simulated calibration curves and the experimental results obtained with a few standards. A regression analysis is carried out to adjust the simulated calibration curves to the experimental points. Although Monte Carlo simulation studies do not entirely suppress the need to carry out experimental calibration work, they considerably reduce the number of calibration standards required. This method has been used by

Oliveira et al.[44,45] using the MCNP (Monte Carlo N-particle) code[46] for PGNAA analysis of coals and cement raw materials.

4.5 Design Optimization

The crucial part of the design of a PGNAA system is that of ensuring a well-defined correspondence between the output of the system and the elemental composition of the sample. Design optimization is particularly important in the case of equipment for field use with isotopic neutron sources. With fast neutron sources, the thermal flux within the sample volume strongly depends on the moderating properties of the sample, which in turn depend on its density and the number of hydrogen atoms per unit volume, i.e. the water content. On the other hand, the average path length to the detector of γ-rays generated in the sample strongly depends on the sample dimensions. Thus the γ-ray count rate at the detector depends on the sample bulk density and dimensions and on the spatial distribution of the thermal flux.

The purpose of design optimization is to reduce the complexity of the functional dependence of the system response relative to the chosen calibration parameters. This can be achieved in a certain range of sample parameters by acting on the sample geometry, namely the sample thickness, and on the effective neutron source spectrum modified by the introduction of a moderating material around the source or between source and detector, which will substantially influence the thermal neutron flux and its spatial distribution inside the sample.

Monte Carlo simulation is a useful method for speeding up design work and system optimization. The influence of the different parameters on the specific count rate from an element can be established by means of the simulation. Optimum design parameters mean that the gauge response is less sensitive to measurement interferences, such as the ones referred to above (fluctuations of density and water content). To reach this objective, the variance of the measured variables must be minimized with respect to both the gauge response variance and the variances of each measurement interference (see Equation 12).

5 APPLICATIONS OF THE METHOD

5.1 Introduction

PGNAA is a powerful analytical technique that can provide accurate information about the elemental composition of materials in a broad range of cases: raw materials, mineral exploration, contaminants, detection of forbidden products, such as drugs and explosives, inspection of containers (detection of plutonium) and applications in health care. In the following sections the main fields of application of the method are briefly reviewed.

5.2 Borehole Logging

Since the early 1980s, the most interesting advances in nuclear logging techniques concerned the development of spectroscopy techniques, especially for the analysis of γ-ray spectra.[47]

Measurements of the energy spectrum of γ-rays were first applied to the detection of natural radioactivity from potassium and of unstable isotopes of the thorium and uranium decay series. Later on, neutron sources coupled to energy-sensitive detectors were used to observe γ-rays from neutron-induced reactions.

Nowadays, spectroscopy of γ-rays from thermal neutron capture and the inelastic scattering of fast neutrons is commonly used in the study of wellbore geophysics. In this type of well logging, a gauge consisting of a neutron generator or a radioisotopic fast neutron source, and γ-ray detectors is introduced through the formations in boreholes. There are applications for exploration and exploitation of mineral resources (oil, gas and coal) as well as for the study of geochemical processes.

The main interest of the oil and gas industry in this technique derives from its ability to locate porous zones that may contain hydrocarbons and are free of clay minerals whose presence would inhibit fluid flow. At the same time it must be ensured that oil and gas production do not contaminate potable water supplies. The coal industry is interested in in situ control of coal quality.

The set of detectable elements provides valuable information on the nature of rock matrix or geological formation. Previously, this information could only be obtained through discrete core analyses.

However, the proper determination of elemental concentrations requires reliable information on geophysical parameters as well as specific spectrometric information related to the individual elements. Several approaches have been developed.[48–50]

Elemental concentrations obtained from logging measurements have been routinely used to provide sedimentary mineralogy, cation-exchange capacity, permeability and other rock and fluid properties. A comparison between mineralogy derived from logging measurements and that derived from laboratory analyses of core samples is shown in Figure 6.

Information from borehole logging during coal exploration can give an early indication of coal quality by allowing the direct determination of the concentration of sensitive elements and the derivation of important

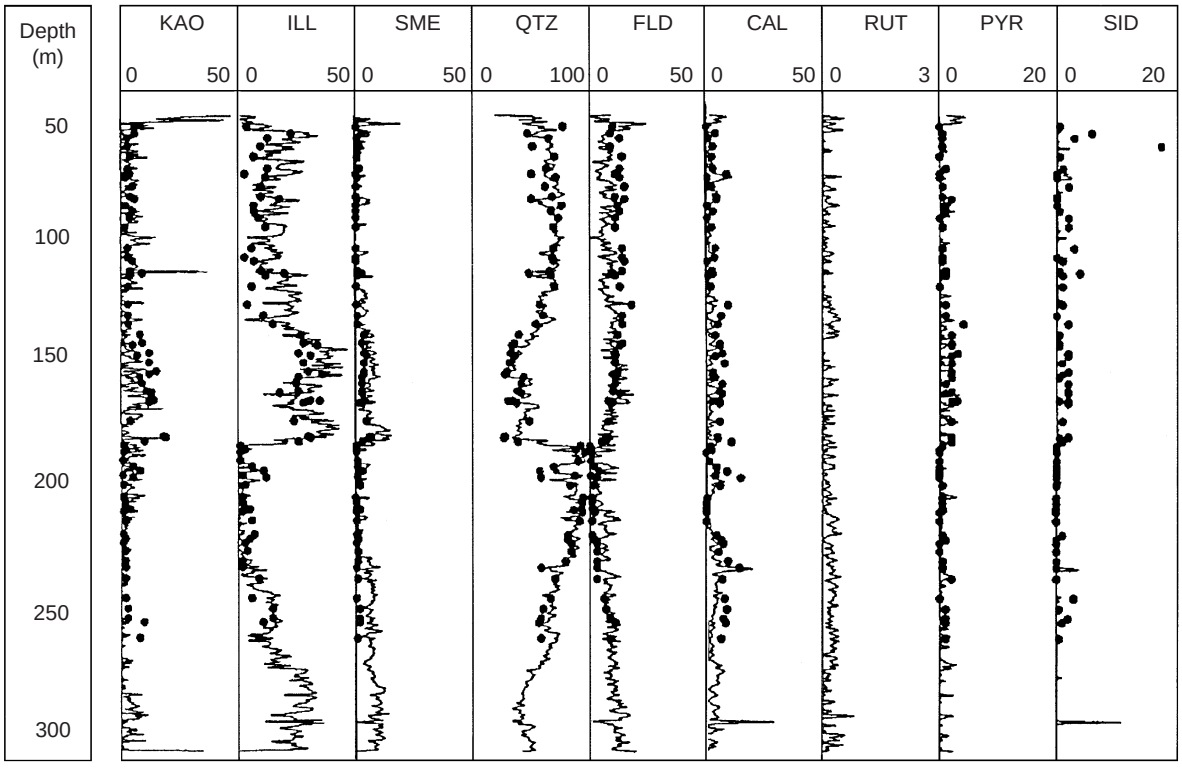

Figure 6 Example of the comparison between mineralogy derived from logging measurements and that derived from laboratory analyses of core samples. The continuous curve is the result from elemental logging measurements. The circles are the results from laboratory analyses of core material. KAO, kaolinite; ILL, illite; SME, smectite; FLD, feldspar; QTZ, quartz; CAL, calcite; RUT, rutile; PYR, pyrite; SID, siderite. (Reproduced from Clayton and Schweiter, *Nucl. Geophys.*, **7**, 158, Copyright (1993)[47] with permission from Elsevier Science.)

economic parameters such as ash content and calorific value.

5.3 Industrial On-line Applications

On-line industrial applications of PGNAA are particularly important in the quality control of coal and in the fabrication of cement. Equipment for bulk analysis of coal and cement is presently commercially available. Other on-line applications in raw material control have been implemented, e.g. the determination of Al in bauxite ores[51] or conveyor belt sorting of aluminum alloys in scrap.[52]

The PGNAA technique surpasses the X-ray fluorescence (XRF) technique owing to the higher penetrating power of the neutron–gamma radiation, which as a consequence has the potential to analyze the whole bulk of the material stream. In these applications radioisotopic sources (generally ^{252}Cf) or neutron generators are used.

The raw material can flow through a vertical chute where the analysis process occurs; generally the chute is lined with an extremely low-friction material to ensure plug free operation. Alternatively, the raw material can be conveyed on a belt. In this case physical integration of the equipment in a previously existing process line is simpler and less costly. Problems related to ''dead zones'' which always appear in the vertical chute are also avoided in this configuration.

The use of coal as an energy source has severe environmental consequences (ash and SO_2 emissions). Good quality control in all phases of the coal fuel cycle will limit the adverse effects. PGNAA can be used in all phases of the coal cycle: mining, coal preparation and power station operation, for the determination of ash, moisture content and specific energy. Among the major coal elements, hydrogen and iron give the best response; sulfur and many of the ash elements (Si, Al, Ti, Ca, Mg, Na, K) are also satisfactorily detected. Carbon responds poorly; however, as its concentration is high it can be detected. Oxygen can hardly be detected in the γ-spectrum due to the element's very small thermal neutron capture cross-section. However, oxygen and carbon can be easily detected using (n,n$'\gamma$) reactions, whose threshold energies are, respectively, 6.13 MeV and 4.44 MeV. This implies the use of a 14-MeV neutron generator or of ^{241}Am-Be as the neutron source.[53]

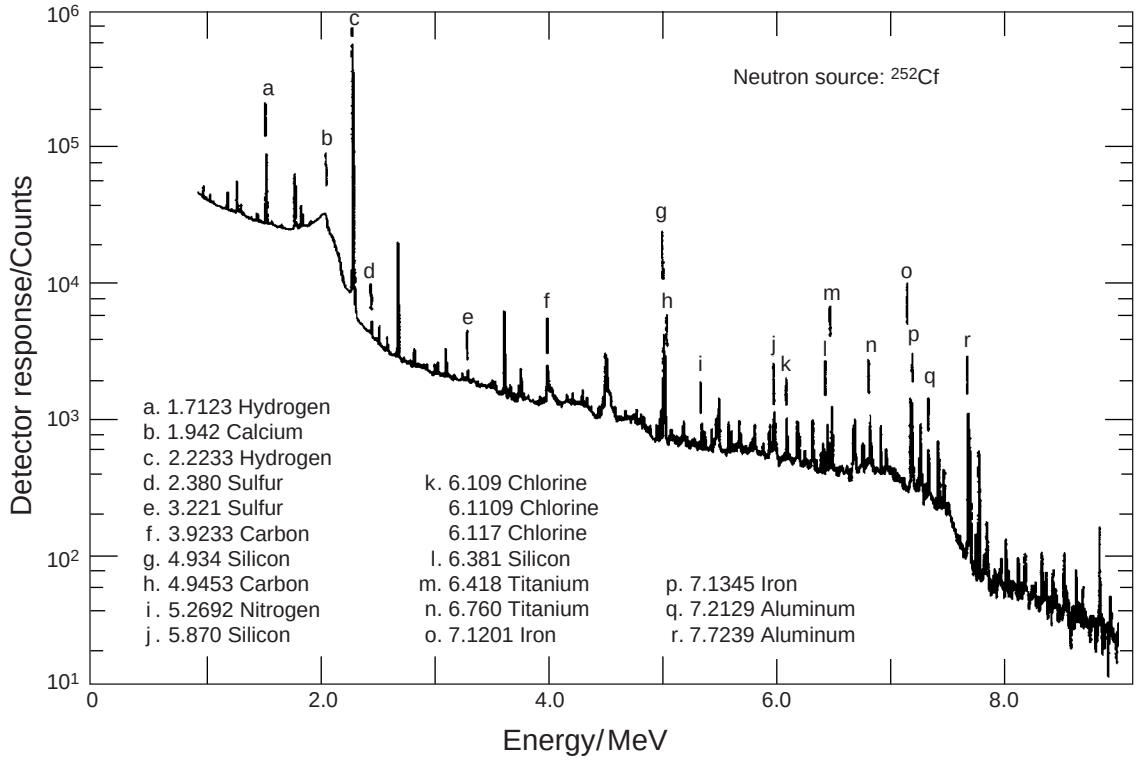

Figure 7 An experimental prompt γ-ray spectrum of a sub-bituminous coal. (Reproduced with permission from Y.L. Yuan, R.P. Gardner, K. Verghese, *Nucl. Technol.*, **77**(1), 105 (1987), Copyright 1987 by the American Nuclear Society, La Grange Park, Illinois.)

Figure 7 shows an experimental prompt γ-ray spectrum of a coal.

The cement industry is fundamentally interested in the measurement of silicon, aluminum, iron and calcium oxides. All the elements involved (except oxygen) can be determined by PGNAA.[54]

The most common application in the fabrication of cement is raw mix proportioning, with the analyzer positioned just ahead of the raw mill. Real-time results can be used to make frequent adjustments of the component feeders to keep the raw mix close to reference composition; another application is the analysis of raw material as it enters the preblending pile. As an alternative to mechanical samplers, analyzers can reduce maintenance and labor costs. A third application is the sorting of raw material coming from the quarry. When synchronized with a downstream diverter, the analyzer can be used to sort one-minute increments of material into save or reject piles.

5.4 Medical In Vivo Applications

In vivo neutron activation analysis (IVNAA) is a unique direct in vivo method for the multielemental analysis of the living human body. Various other radiation-based techniques (radiography, isotopic imaging, radiotracer dilution) reflect only relative changes in tissue density or volume. Table 6 shows the elements measured by activation analysis.[55]

The prompt γ-variety of IVNAA can provide data on the total or partial body content of the following elements: oxygen, carbon, hydrogen, nitrogen, calcium, sodium, chlorine, silicon, iron, cadmium and mercury.

Although PGNAA instruments were initially developed in nuclear research laboratories, methodologies have advanced to a stage where facilities were designed and built exclusively for clinical application in hospitals and health centres.

The choice of the neutron source is based on elements to be measured, the required degree of uniformity of the activation, the acceptable level of accuracy and the admissible radiation dose. Cf and Pu-Be are the most used sources; some facilities are equipped with neutron generators or cyclotrons. Large NaI(Tl) scintillators are generally used as detectors, although in certain facilities germanium or BGO detectors are preferred.

Considerable effort has been expended in reducing the absorbed dose without decreasing the precision of results. The substitution of Pu-Be sources by Cf contributes to this end. The range of exposure doses in existing facilities depends in part on the element to

Table 6 Elements measured by in vivo activation analysis[a]

Body element	Reference man		Technique	E_γ (MeV)
	g	%		
Total body				
Oxygen	43 000	61	pulsed	6.10
Carbon	16 000	23	pulsed	4.44
Hydrogen	7000	10	prompt	2.23
Nitrogen	1800	2.6	prompt	10.83
Calcium	1000	1.4	delayed	3.10
			prompt	6.42
Phosphorus	780	1.1	delayed	1.78
Sodium	100	0.14	delayed	2.75
			prompt	6.40
Chlorine	95	0.12	delayed	6.11
			prompt	2.23
Partial body				
Silicon	18	trace	prompt	1.78
Iron	4.2	trace	delayed	0.85
			prompt	0.847
Cadmium	<0.15	trace	prompt	0.56
Aluminum	<0.08	trace	delayed	1.78
Mercury	<0.002	trace	prompt	0.37
Lithium	<0.01	trace	prompt	($^3_1$H counting)

[a] Reproduced from Ellis[55] with permission.

be analyzed, the precision required and the detection sensibility of the counting system. The Swansea system is shown in Figure 8.[56] The dose is often less than, or comparable to, many routine diagnostic radiographic procedures.

The initial application that has influenced the subsequent development of IVNAA is the determination of total body calcium content in patients suffering from various metabolic diseases. Since calcium is predominantly located in the skeleton, a measure of this element would provide a direct examination of the total skeletal mass. Clinical research studies have investigated a number of metabolic disorders in which there is a primary or secondary alteration of mineral metabolism: osteoporosis, osteomalacia, renal osteodystrophy, Paget's disease, Cushing's syndrome, acromegaly, thyroid and parathyroid disorders, myotonic dystrophy, thalassemia and alcoholic cirrhosis.

The body nitrogen measurement provides a direct determination of body protein, which is the essential component of the body's lean tissue mass. Changes in protein mass can reflect changes in nutritional status. Hence clinical studies involving nitrogen measurements in patients with progressive diseases may be important in the following cases: postoperative surgery, protein malnutrition, cancer renal failure, cardiovascular disease, total parental nutrition, obesity and growth deficiencies.

Body carbon measurements can provide an index of the body fat.

Cadmium measurements have been carried out in industrial workers. Kidney concentrates cadmium and has been identified as the initial target organ in toxicological studies. The association between cadmium burdens and various diseases related to kidney malfunctions have also

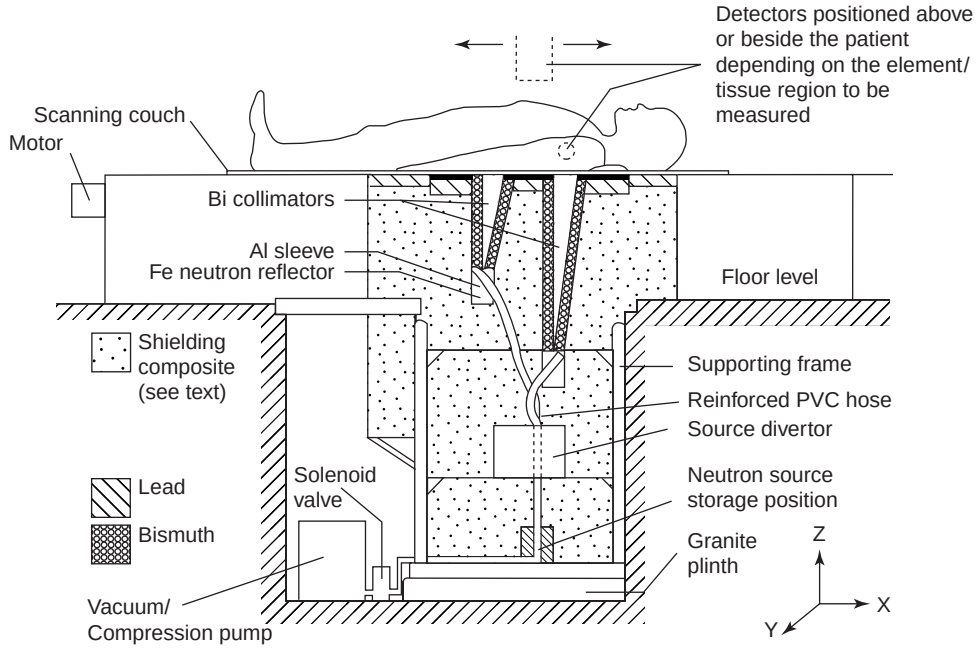

Figure 8 The Swansea IVNAA system. (Reproduced from Natto et al., *Int. J. Appl. Radiat. Isot.*, **49**, 546, Copyright (1998)[56] with permission from Elsevier Science.)

been investigated. Several studies involving nondialyzed patients in renal failure have shown that cadmium was a cofactor in the cause of the renal disease.

5.5 Determination of Environmental Contaminants

Elemental analysis of toxic elements in contaminated materials is essential for solving environmental problems. PGNAA can also be applied to this purpose in the determination of major and minor elemental components and even of trace elements.

Dulloo et al.[57] used a 14-MeV pulsed generator for the nondestructive assay of mercury, cadmium and lead. The DLs for a 600 s run are 15 ppm, 170 ppm and 8600 ppm for Cd, Hg and Pb respectively.

Yonezawa et al.[23] carried out multielement analysis of environmental samples by PGNAA with cold and thermal neutrons in a reactor facility. Cd was detected in mussels, river sediments, pond sediments, vehicle exhaust particulates and unpolished rice flour; Hg was detected in human hair.

Hiep and Tan[22] used the technique for analyzing the concentration of Al, Fe, Ti, Si, Ca and K in airborne dust samples.

5.6 Explosives and Drugs Detection Systems

Considerable progress has been made in research and development of nuclear-based techniques for the detection of explosives and drugs. The technique must be noninvasive, sensitive to the smallest effective amounts of explosives and drugs and give rise to a minimum of false alarms. An effective contraband inspection system must be capable of determining the relevant chemical elements and their spatial distribution in the inspected object.[58]

Chemical explosives have high nitrogen and oxygen contents. Pure narcotics are characterized by high content of H and C, moderate content of O and small content of N. However, narcotics hydrochlorides also contain small amounts of Cl. While high nitrogen density is a good indicator of explosives, the detection of a low oxygen to carbon ratio combined with the presence of N and Cl is a strong indication of the presence of drugs.

The PGNAA technique enables the three-dimensional mapping of explosives and narcotics in luggage. This is accomplished by irradiating parts of the interrogated object with short pulses of fast monoenergetic neutrons (Figure 9). The three-dimensional distribution can be obtained by scanning the sample.[59]

5.7 Other Applications

Many other applications have been carried out in different fields using PGNAA.

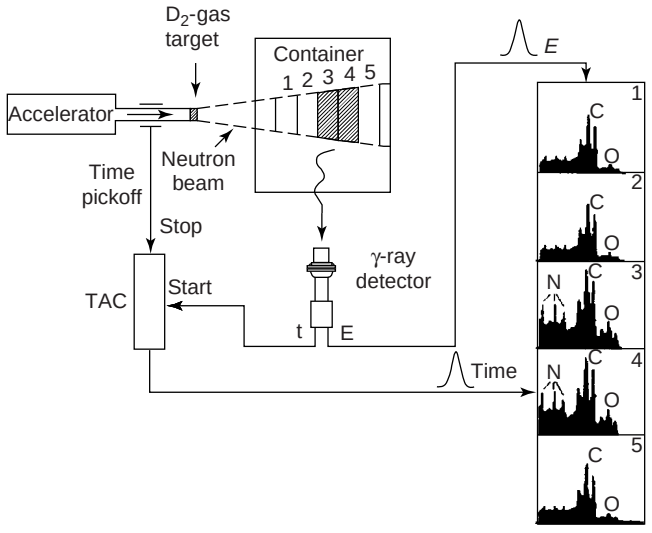

Figure 9 Principle of three-dimensional multielemental mapping. The γ-ray spectra shown on the right are measured in the five indicated layers. (Reproduced from Sawa[58] with permission.)

Anderson et al.[60] described the determination of H, B, Cl, K, Na, S, Ca and Cd in foods using a thermal neutron beam. The advantage of this type of analysis lies in its sensitivity to elements of nutritional or toxicological interest while no sample preparation is needed, thus eliminating the risk of volatilization losses.

Kuno et al.[61] applied PGNAA to the determination of the concentration of B, H, S and Si in on-shore sediments in order to investigate the distribution and migration of chemical elements. The measurements were complemented by DGNAA results.

Caffrey et al.[62] applied a portable PGNAA device with an isotopic source to the identification of the contents of munitions, chemical storage containers, chemical warfare materials in stockpiles, etc.

Prettyman et al.[63] investigated the use of the technique as a diagnostic tool related to matrix characterization for nondestructive assay and plutonium surveillance. The technique can help to confirm item descriptions (for example, plutonium chloride versus plutonium oxide).

Spyrou et al.[64] proposed a method to determine the concentration of a large number of elements by measuring the prompt or delayed γ-rays and applying the principles of reconstructive tomography to obtain the spatial distribution of elemental concentration in a selected plane or slice through the sample. This method has been employed in the examination of radioactive waste as well as in elemental analysis of biomedical and bioenvironmental samples.

The contents of major and some trace elements in large archaeological materials have been determinated by Sueki et al.[24]

6 PRESENT TRENDS OF DEVELOPMENT

According to Gardner[52] ''PGNAA is a rapidly emerging important new technology and measurement approach''. However, further development of its applications in different branches of human activity require improvements in neutron sources, signal processing and data handling as well as in nuclear data.

The introduction of a new point source generator based on the IEC principle may have a promising future in on-line field applications of PGNAA. Apart from sharing the on-off switching capability with conventional neutron generator devices, it offers longer life and lower costs.

Industrial on-line process control requires short response times. For a PGNAA analyzer this means, normally, high count rates. High-rate data acquisition systems must be improved.

Enhancement of signal to background noise ratio leads to lower statistical uncertainty of the results thus contributing to reducing response times. This objective can be achieved by proper system design optimization. Monte Carlo methods can play an important role in this respect.

More accurate measurements and thus better data regarding microscopic nuclear parameters, such as cross-sections, number of photons emitted per neutron captured by the relevant isotope and γ-ray attenuation coefficients in different materials are important for simulation studies carried out with a view to optimizing design and calibration of PGNAA analyzers.

Although certain physical constraints (neutron flux perturbations and γ-ray attenuation) are difficult to overcome, the development and application of the k_0 method to large samples would be welcome.

The MCLLS technique, which has been applied to coal, cement raw materials, oil well logging, vitrified nuclear waste and conveyor belt sorting of aluminum alloys in scrap, can also be successfully applied to other fields.

The nuclear imaging technique using a pulsed fast neutron analyzer for the detection of explosives concealed in airline luggage has potential application in the future to a variety of other situations.

ACKNOWLEDGMENTS

The authors wish to thank our colleagues Carlos Cruz for helpful discussions concerning signal processing and data handling, António Falcão for his contribution in graphics design of Figures 1 and 2 and Teresa Pires for her improvements in graphics in Figures 3, 6, 7, 8 and 9.

ABBREVIATIONS AND ACRONYMS

ADC	Analog to Digital Converter
BGO	Bismuth Germanate Detector
DGNAA	Delayed γ-Neutron Activation Analysis
DL	Detection Limit
FCT	Fixed Conversion Time
FET	Field Effect Transistor
fwhm	Full Width at Half-maximum
GI	Gated Integrator
HPGe	Hyper-pure Germanium Detector
IEC	Inertial Electrostatic Confinement
INAA	Instrumental Neutron Activation Analysis
IVNAA	In Vivo Neutron Activation Analysis
LLS	Linear Least Squares
MCA	Multichannel Analyzer
MCLLS	Monte Carlo Library Least Squares Method
MCNP	Monte Carlo N-particle
NAA	Neutron Activation Analysis
NaI(Tl)	Thallium-doped Sodium Iodide Detector
NIST	National Institute of Standards and Technology
PGNAA	Prompt γ-Neutron Activation Analysis
PUR/LTC	Pile-up Rejector/Live Time Corrector Circuit
SINQ	Spallation Neutronen Quelle
TC	Time Constant
TRP	Transistor Reset Preamplifiers
XRF	X-ray Fluorescence

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Radiochemical Methods **(Volume 14)**
Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation

REFERENCES

1. J.L. Mikesell, F.E. Senftle, A.B. Tanner, 'Borehole Elemental Analysis of Coal Using Fast Neutrons, Gamma, X-ray and Neutron Techniques for the Coal Industry', Proceedings of an Advisory Group Meeting, Vienna, 1984, IAEA, Vienna, 25–42, 1986.
2. M.R. Wormald, C.G. Clayton, 'In-situ Analysis of Coal by Measurement of Neutron-induced Prompt γ-rays', *Int. J. Appl. Radiat. Isot.*, **34**(1), 71–82 (1983).
3. C. Oliveira, J. Salgado, 'A Monte Carlo Simulation Study of the Space Dependence of Prompt Gamma-ray Volume Emission Rate in Bulk Coal Irradiated with an Am-Be Source', Proceedings of the Conference on Radioisotope Application and Radiation Processing in Industry, Leipzig, ed. D. Luther, Institute of Isotope and Radiation Research (2), 737–743, 1989.
4. D.L. Anderson, M.P. Failey, W.H. Zoller, W.B. Walters, G.E. Gordon, R.M. Lindstrom, 'Facility for Non-destructive Analysis for Major and Trace Elements Using Neutron-capture Gamma-ray Spectrometry', *J. Radioanal. Chem.*, **63**(1), 97–119 (1981).
5. R.L. Paul, R.M. Lindstrom, A.E. Heald, 'Cold Neutron Prompt Gamma-ray Activation Analysis at NIST – Recent Developments', *J. Radioanal. Nucl. Chem.*, **215**(1), 63–68 (1997).
6. 'Industrial Gauging and Analytical Instrumentation Sources', Amersham Catalogue S96/89, Amersham, UK.
7. E.A. Lorch, 'Neutron Spectra of ^{241}Am/B, ^{241}Am/Be, ^{241}Am/F, ^{242}Cm/Be, ^{238}Pu/^{13}C and ^{252}Cf Isotopic Neutron Sources', *Int. J. Appl. Radiat. Isot.*, **24**, 585–591 (1973).
8. Y.A. Karelin, Y.N. Gordeev, V.I. Karasev, V.M. Radchenko, Y.V. Schimbarev, R.A. Kuznetsov, 'Californium-252 Neutron Sources', *Appl. Radiat. Isot.*, **48**(10–12), 1563–1566 (1997).
9. P. Bach, 'Sealed Tubes Neutron Generators', *International Conference Proceedings on Nuclear Techniques for Analytical and Industrial Applications*, Mykonos, June 1992, eds. G. Vourvopoulos, T. Paradellis, Western Kentucky University, Bowling Green, KY, 397–412, 1992.
10. J.J. Kehayias, H. Zhuang, 'Use of a Pulsed Neutron Generator for the In Vivo Measurement of Carbon, Oxygen and Hydrogen Content in Humans', *International Conference Proceedings on Nuclear Techniques for Analytical and Industrial Applications*, Mykonos, June 1992, eds. G. Vourvopoulos, T. Paradellis, Western Kentucky University, Bowling Green, KY, 265–284, 1992.
11. G.H. Miley, J. Sved, 'The IEC Plasma-target-based Neutron Source', *Appl. Radiat. Isot.*, **48**(10–12), 1557–1561 (1997).
12. R.L. Paul, R.M. Lindstrom, D.H. Vincent, 'Cold Neutron Prompt Gamma Activation Analysis at NIST: A Progress Report', *J. Radioanal. Nucl. Chem.*, **180**(2), 263–269 (1994).
13. C. Yonezawa, M. Margara, H. Sawahata, M. Hoshi, Y. Iyo, E. Tachikawa, 'Prompt Gamma-ray Analysis using Cold and Thermal Guided Neutron Beams', *J. Radioanal. Nucl. Chem.*, **193**(1), 171–178 (1995).
14. K. Unlu, C.R. Martinez, B.W. Wehring, 'Prompt Gamma Activation Analysis with the Texas Cold Neutron Source', *J. Radioanal. Nucl. Chem.*, **193**(1), 145–154 (1995).
15. G. Molnar, T. Belgya, L. Dabolczi, B. Fazekas, Z.S. Revay, A. Veres, I. Bikit, Z. Kiss, J. Östör, 'The New Prompt Gamma-activation Analysis Facility at Budapest', *J. Radioanal. Nucl. Chem.*, **215**(1), 111–115 (1997).
16. B.W. Wehring, K. Unlu, C.R. Martinez, 'Application of Cold Neutron Prompt Gamma Activation Analysis at the University of Texas Reactor', *Appl. Radiat. Isot.*, **48**(10–12), 1343–1348 (1997).
17. S. Alam, M.A. Zaman, S.M. Islam, M.H. Ahsan, 'Extraction of Pure Thermal Neutron Beam for the Proposed PGNAA Facility at the Triga Research Reactor of AERE, Savar, Bangladesh', *Nucl. Method Instrum.*, **B83**, 235–244 (1993).
18. O.K. Harling, J.M. Chabeuf, F. Lambert, G. Yasuda, 'A Prompt Gamma Neutron Activation Analysis Facility Using a Diffracted Beam', *Nucl. Method Instrum.*, **B83**, 557–562 (1993).
19. C. Yonezawa, A.K. Wood, 'Prompt γ-ray Analysis of Boron with Cold and Thermal Neutron Guided Beams', *Anal. Chem.*, **67**(24), 4466–4470 (1995).
20. K. Sueki, K. Kobayashi, W. Sato, H. Nakahara, T. Tomizawa, 'Nondestructive Determination of Major Elements in a Large Sample by Prompt γ-ray Neutron Activation Analysis', *Anal. Chem.*, **68**(13), 2203–2209 (1996).
21. V.H. Tan, 'Utilisation of Horizontal Experimental Channels of the Dalat Reactor for Research and Applications', Proceedings 10th Pacific Basin Nuclear Conference, Kobe, Japan, Atomic Energy Society of Japan and Japan Atomic Industrial Forum, 10–16, 1996.
22. N.T. Hiep, V.H. Tan, 'Thermal Neutron Capture Prompt Gamma Analysis for the Element Determination in Environmental Objects', Proceedings 10th Pacific Basin Nuclear Conference, Kobe, Japan, Atomic Energy Society of Japan and Japan Atomic Industrial Forum, 1353–1357, 1996.
23. C. Yonezawa, H. Matsue, M. Hoshi, 'Multi-element Analysis of Environmental Samples by Cold and Thermal Guided Neutron Induced Prompt Gamma-ray Measurement', *J. Radioanal. Nucl. Chem.*, **215**(1), 81–85 (1997).
24. K. Sueki, Y. Oura, W. Sato, H. Nakahara, T. Tomizawa, 'Analysis of Archaeological Samples by the Internal Monostandard Method of PGAA', *J. Radioanal. Nucl. Chem.*, **234**(1–2), 27–31 (1998).
25. K. Riley, O. Harling, 'An Improved Prompt Gamma Neutron Activation Analysis Facility using Focused Diffracted Neutron Beam', *Nucl. Method Instrum.*, **B143**, 414–421 (1998).
26. Y. Oura, H. Nakahara, K. Sueki, W. Sato, T. Tomizawa, 'Completely Non-destructive Elemental Analysis of Bulky Samples by PGA (Prompt Gamma Analysis)', 13th

Radiochemical Conference, Marianske Lazne-Jachymov, Czech Republic, Czech Chemical Society, I.M. Marci Spectroscopic Society and Czech Radiological Society, Abstract, 136, 1998.
27. M.S. Zucker, N. Tsoupas, P.E. Vanier, U. von Wimmersperg, S.F. Mughabghab, E. Schmidt, 'Spallation Neutron Production Measurements', *Nucl. Sci. Eng.*, **129**, 180–186 (1998).
28. J.R. Haines, J. Alonso, B.R. Appleton, T.A. Gabriel, 'Overview of the National Spallation Neutron Source', *Am. Nucl. Soc.*, **76**, 377–378 (1997).
29. 'Theme 1996: Neutrons for research; SINQ' Paul Scherrer Institute Annual Report 1996, 19–33 (1996).
30. A.G. Hanna, R.M. Brugger, M.D. Glascock, 'The Prompt Gamma Neutron Activation Analysis Facility at MURR', *Nucl. Method Instrum.*, **188**, 619–627 (1981).
31. M.R. Wormald, 'Pair Spectrometer NaI(Tl) Array for Neutron-induced Prompt Gamma-ray Analysis', *Nucl. Geophys.*, **3**(4), 373–380 (1989).
32. T. Gozani, G. Reynolds, H. Bozorgmanesh, E. Elias, T. Maung, V. Orphan, 'Nuclear Assay of Coal', Electric Power Research Institute, Palo Alto, CA, Vol. 1, 1980.
33. T. Gozani, 'Physics of Recent Applications of PGNAA for On-line Analysis of Bulk Minerals', Proceedings of the 5th International Symposium on Capture Gamma-ray Spectroscopy Related Topics, Knoxville, TN, American Institute of Physics, Woodbury, NY, 1984.
34. W. Herzog, 'Prompt Gamma Neutron Activation Analysis of Hard Coal, Raw Lignite and a Raw Glass Mixture', *Nucl. Geophys.*, **3**(4), 467–473 (1989).
35. F. de Corte, 'The k_0-standardization Method. A Move to the Optimization of Neutron Activation Analysis', Aggrégé Thesis, University of Gent, 1987.
36. R.L. Paul, 'The Use of Element Ratios to Eliminate Analytical Bias in Cold Neutron Prompt Gamma-ray Activation Analysis', *J. Radioanal. Nucl. Chem.*, **191**(2), 245–256 (1995).
37. R.L. Paul, R.M. Lindstrom, 'Measurement of k_0-Factors of Prompt Gamma-ray Activation Analysis', 2nd International k_0 Users Workshop, Ljubliana, Jozef Stefan Institute, Ljubliana, Slovenia, 55, 1996.
38. R.M. Lindstrom, R. Fleming, R. Paul, E. Mackey, 'The k_0 Approach in Cold Neutron Activation Analysis', Proceedings of the International k_0 Users Workshop, Gent, ed. F. de Corte, Institute of Nuclear Science, University of Gent, 121, 1992.
39. M. Rossbach, F. Corte, 'Introducing the k_0-Factor into Prompt Gamma Cold Neutron Activation Analysis', Proceedings of the International k_0 Users Workshop, Gent, ed. F. de Corte, Institute of Nuclear Science, University of Gent, 125, 1992.
40. C.M. Shyu, R.P. Gardner, K. Verghese, 'Development of the Monte Carlo-library Least Squares Method of Analysis for Neutron Capture Prompt Gamma-ray Analyzers', *Nucl. Geophys.*, **7**(2), 241–267 (1993).
41. R.P. Gardner, P. Guo, Q. Ao, C.L. Dobbs, 'Black Box Radiation Gauges and Analyzers: Dream Or Reality?', *Appl. Radiat. Isot.*, **48**(10–12), 1273–1288 (1997).
42. R.P. Gardner, A. Sood, Y.Y. Wang, L. Liu, P. Guo, R.J. Gehrke, 'Single Peak Versus Library Least-squares Analysis Methods for the PGNAA Analysis of Vitrified Waste', *Appl. Radiat. Isot.*, **48**(10–12), 1331–1335 (1997).
43. C.G. Clayton, C.F. Coleman, 'Analysis of Neutron Flux Distributions Generated by Sources of Fast Neutrons in a Range of Bituminous Coals', *Int. J. Appl. Radiat. Isot.*, **36**(10), 757–788 (1985).
44. C. Oliveira, J. Salgado, F.G. Carvalho, 'Optimization of PGNAA Instrument Design for Bulk Coal Analysis', *Nucl. Geophys.*, **7**(2), 285–303 (1993).
45. C. Oliveira, J. Salgado, 'Calibration Curves of a PGNAA System for Cement Raw Material Analysis using the MCNP Code', *Appl. Radiat. Isot.*, **40**(12), 1685–1689 (1998).
46. J.E. Briesmeister (ed.), 'MCNP – A General Monte Carlo Code for N-Particle Transport Code, LA-12625-M, Version 4B Manual. RSICC, Oak Ridge, TN, 1997.
47. C.G. Clayton, J.S. Schweitzer, 'A Review of Aspects of Nuclear Geophysics'. *Nucl. Geophys.*, **7**(2), 143–171 (1993).
48. R.C. Odom, S.M. Bailey, R.D. Wilson, 'Benchmarking Computer Simulations of Neutron-induced Gamma-ray Spectroscopy for Well Logging', *Appl. Radiat. Isot.*, **48**(10–12), 1321–1328 (1997).
49. R. Hertzog, L. Colson, B. Seeman, M. O'Brien, D. McKeon, P. Wraight, J. Grau, E. Ellis, J. Schweitzer, M. Herron, 'Geochemical Logging with Spectrometry Tools', Technical Symposium, Society of Petroleum Engineers, Dallas, Paper SPE16792, Richardson, TX, 1987.
50. J.A. Grau, J.S. Schweitzer, D.V. Ellis, R.C. Hertzog, 'A Geological Model for Gamma-ray Spectroscopy Logging Measurements', *Nucl. Geophys.*, **3**(4), 351–359 (1989).
51. Y. Idris, I.I. Funtua, I.M. Umar, S.B. Elegba, 'Neutron Activation Analysis of Aluminium with an Am-Be Neutron Source-application for Bauxite Investigation in Nigeria', *Appl. Radiat. Isot.*, **49**(1–2), 41–42 (1998).
52. R.P. Gardner, P. Guo, A. Sood, C.W. Mayo, R.J. Gehrke, C.L. Dobbs, 'Monte Carlo Aided Treatments of the Non-linear Inverse PGNAA Measurement Problem for Various Continuous On-line Applications', *J. Radioanal. Nucl. Chem.*, **233**(1–2), 105–107 (1998).
53. 'Nuclear Techniques in the Coal Industry', Proceedings of a Final Research Coordination Meeting Held in Krakow, IAEA-TECDOC-845, 1994.
54. R.C. Woodward, 'On-line Elemental Analysis', *World Cement*, **20**(8), 272–274 (1989).
55. K.J. Ellis, 'In Vivo Activation Analysis: Present and Future Prospects', *J. Radioanal. Nucl. Chem.*, **169**(2), 291–300 (1993).
56. S.A. Natto, D.G. Lewis, S.J.S. Ryde, 'Benchmarking the MCNP Code for Monte Carlo Modelling of an In Vivo

Neutron Activation Analysis System', *Appl. Radiat. Isot.*, **49**(5–6), 545–547 (1998).

57. A.R. Dulloo, F.H. Ruddy, T.V. Congedo, J.G. Seidel, R.J. Gehrke, 'Detection Limits of a Laboratory Pulsed Gamma Neutron Activation Analysis System for the Non-destructive Assay of Mercury, Cadmium, and Lead', *Nucl. Technol.*, **123**, 103–112 (1998).
58. Z.P. Sawa, 'PFN GASGA Techniques for Detection of Explosives and Narcotics', International Conference Proceedings on Nuclear Techniques for Analytical and Industrial Applications, eds. G. Vourvopoulos, T. Paradellis, Mykonos, June, 1992, 335–358, 1992.
59. T. Gozani, 'Neutron Based Non-intrusive Inspection Techniques', International Conference on Neutron in Research and Industry, ed. G. Vourvopoulos, Mykonos, June 1996, 174–181, 1997.
60. D.L. Anderson, W.C. Cunningham, G.H. Alvarez, 'Multielement Analysis of Food by Neutron Capture Prompt Gamma-ray Activation Analysis', *J. Radioanal. Nucl. Chem.*, **167**(1), 139–144 (1993).
61. A. Kuno, M. Matsuo, B. Takano, C. Yonezawa, H. Matsue, H. Sawahata, 'Neutron Induced Prompt Gamma-ray and Instrumental Neutron Activation Analyzes of Urban Estuarine Sediments', *J. Radioanal. Nucl. Chem.*, **218**(2), 169–176 (1997).
62. A.J. Caffrey, R.J. Gehrke, R.C. Greenwood, J.K. Hartwell, K.M. Krebs, G.D. McLaughlin, 'US Army's Experience with the PINS Chemical Assay System', *Trans. Am. Nucl. Soc.*, **72**, 117 (1995).
63. T.H. Prettyman, L.A. Foster, S. Parish, 'Applications of PGNAA to Plutonium Surveillance', *Trans. Am. Nucl. Soc.*, **77**, 146–147 (1997).
64. N.M. Spyrou, J.M. Sharaf, S. Rajeswaran, E. Mesbahi, 'Determination of the Elemental Distribution in a Sample Using Neutron Induced Gamma-ray Emission Tomography', *J. Radioanal. Nucl. Chem.*, **217**(2), 273–241 (1997).

Radiochemical Neutron Activation Analysis

Kaj Heydorn*
Risø National Laboratory, Roskilde, Denmark

1 Introduction **1**
1.1 Principle and Definition of Radiochemical Neutron Activation Analysis 1
1.2 Relationship to Instrumental Neutron Activation Analysis 2
1.3 Advantages and Limitations 2
2 Theory and Operating Principles **2**
2.1 Sample and Comparator 3
2.2 The Activation Process 4
2.3 Radiochemical Separation 4
2.4 Methodology of Counting 5
2.5 Data Processing 6
2.6 Recovery Correction 7
2.7 Final Calculation of Analytical Result 8
3 Instrumentation **10**
3.1 Irradiation Systems 10
3.2 Counting Equipment 11
4 Quality Assurance **11**
4.1 Statistical Control 12
4.2 Verification Methods 13
4.3 Traceability Chain 14
4.4 Uncertainty Budget 15
4.5 Statement of Uncertainty 16
5 Perspective and Future Developments **16**
5.1 Analytical Metrology 17
5.2 Certification Analysis 17
5.3 Sampling Constants 17
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Radiochemical neutron activation analysis (RNAA) is a method for trace-element determination based on the measurement of an indicator radionuclide, chemically separated from a neutron-activated sample. Unlike instrumental neutron activation analysis (INAA) the RNAA is therefore not a nondestructive method, and at the same time the amount of work put into the analysis of each sample makes RNAA much more expensive than INAA. This is the cost that has to be paid for determining selected elements at much lower concentrations and superior accuracy, while maintaining matrix independence and most of the other excellent characteristics of INAA. One of these characteristics is the small self-shielding during irradiation with neutrons, which opens the possibility of analyzing also large samples without any pretreatment.

Particularly for the lowest levels of trace elements RNAA provides several important advantages over alternative analytical methods:

a) the absence of a reagent blank, and
b) the corresponding insensitivity to contamination after activation
c) the controlled addition of carrier to avoid losses by adsorption etc., and
d) the corresponding possibility of determining chemical yield or recovery.

When carefully and thoughtfully carried out, RNAA qualifies as a so-called definitive method, one of the reasons being its direct traceability to a comparator of the pure element to be determined.

1 INTRODUCTION

The use of neutrons for analytical purposes goes back to the classical paper by Hevesy and Levi,[1] who discovered in 1936 in Copenhagen that activation of a sample of presumably pure Y_2O_3 by means of a Ra–Be neutron source gave rise to a radionuclide that could only be attributed to Dy; by comparing with samples of pure Dy they were even able to determine the actual quantity of Dy in the particular sample.

1.1 Principle and Definition of Radiochemical Neutron Activation Analysis

Although himself a chemist Hevesy never applied radiochemistry to any activated samples for analytical purposes; this was first done by Seaborg and Livingood,[2] who determined Ga after the addition of carrier and radiochemical separation of iron activated at the Berkeley cyclotron.

RNAA is a method for the determination of elements, and its full power was not realized until after the second World War, when the availability of nuclear reactors was brought to the attention of the analytical community. A clear presentation of the principles of RNAA, including carrier addition and yield determination, was

* Current affiliation: Technical University of Denmark, Lyngby, Denmark

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

first made by Boyd,[3] who still provides the fundamental description of what was later referred to as *comprehensive* RNAA:[4]

1. activation with neutrons
2. carrier addition
3. radiochemical separation
4. measurement of activity
5. determination of yield.

After half a century RNAA has reached maturity, the principles are well understood and no major technological developments are expected to take place in the foreseeable future. At the same time, it is also improbable that any other method will be found that could provide the same special characteristics so valuable for ascertaining the quality of trace element analysis.

1.2 Relationship to Instrumental Neutron Activation Analysis

The outstanding characteristic of neutron activation analysis in comparison with other methods of analysis is the timescale associated with the lifespan of the excited species or characteristic radionuclides, often referred to as *indicators*, which permit the introduction of a chemical separation between the end of activation and the beginning of measurement.

In purely instrumental neutron activation analysis this decay period is used only for simple physical operations, such as separating the sample from its container, taking an aliquot by volume or weight, cleaning the surface of the sample or adjusting its shape or volume before measurement is begun. In INAA the preservation of the integrity of the sample permits the determination of many elements in the same sample.

The radiochemical separation carried out during the decay period is characteristic of RNAA and serves to reduce or eliminate unwanted radionuclides interfering with the measurement of the indicator, so as to minimize the uncertainty associated with the determination of a particular element. RNAA is therefore basically a method for the determination of only a small number of elements in the same sample.

Over the years successful attempts have been made to reduce the need for radiochemical separations by advances in instrumentation for counting, supplemented with the use of γ-radiation or charged particles rather than neutrons for activation.[5] For highest accuracy at the lowest levels RNAA is still needed as a reference method.

1.3 Advantages and Limitations

RNAA is clearly not a nondestructive method, and at the same time the amount of work put into the analysis of each sample makes RNAA much more expensive than INAA. This is the cost that has to be paid for determining selected elements at much lower concentrations and superior accuracy in comparison with INAA, while maintaining most of the other excellent characteristics of this method, see **Instrumental Neutron Activation Analysis**, including the small self-shielding during irradiation with neutrons, which opens up the possibility of analyzing large samples as well without any pretreatment.[6]

Particularly for the lowest levels of trace elements RNAA provides several important advantages over alternative analytical methods:

1. absence of a reagent blank
2. corresponding insensitivity to contamination after activation
3. controlled addition of carrier to avoid losses by adsorption etc.
4. corresponding possibility of determining chemical yield or recovery.

The presence of a controlled amount of carrier during the radiochemical separation process, regardless of the original content in the sample, makes the a priori precision in the uncertainty budget independent of the concentration of the element to be determined, the *determinand*, see section 4.4. Any method that has been found to be useful for chemical separation of the indicator element from the material to be analyzed by any other analytical method, could also be used in RNAA. In this article we are therefore concerned only with the evaluation of the *performance characteristics* of the potential separation methods (see section 2.3) not with the methods themselves.

When carefully and thoughtfully carried out, RNAA qualifies as a so-called definitive method,[7] one of the reasons being its direct traceability to a comparator of the pure element to be determined.

2 THEORY AND OPERATING PRINCIPLES

The special features of RNAA can be expressed as a combination of separate, almost independent unit operations, and Figure 1[8] illustrates the entire process from the sample to be analyzed to the result of the analysis.

The RNAA sensitivity S given in Equation (1) for a particular determinand element X is expressed as the number of counts per unit weight of X, and it depends upon the half-life of its indicator $T_{1/2}$, the duration of the activation t_i, the decay time t_c, and the counting time t_m

$$S = \frac{C_x(1 - e^{-\lambda t_i})\, e^{-\lambda t_c}\, (1 - e^{-\lambda t_m})}{\lambda} \qquad (1)$$

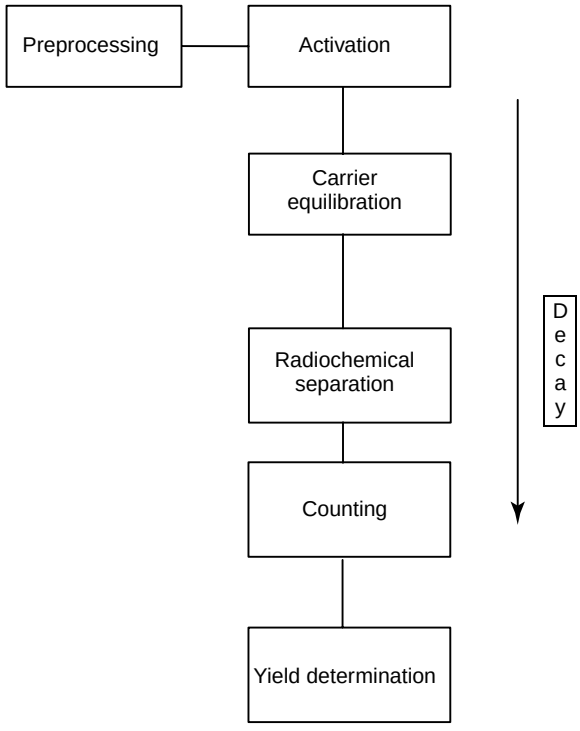

Figure 1 Schematic representation of RNAA methodology.

where $\lambda T_{1/2} = \ln 2$, and the rate constant (Equation 2)

$$C_x = P_i R_c P_m \qquad (2)$$

has to be determined experimentally from a known quantity of *comparator*.

The rate of formation P_i is proportional to the neutron flux density during activation, see section 2.2, but its magnitude varies by orders of magnitude from element to element. The recovery R_c and the detection efficiency P_m will be discussed later.

Table 1 lists those elements with corresponding indicators which have rate constants larger than that of Zn.[9a]

Table 1 Elements with high sensitivity in neutron activation analysis

Element	Indicator	Half-life	γ-Energies (keV)
Aluminum	^{28}Al	2.2 min	1779
Antimony	^{122}Sb	2.8 d	564
	^{124}Sb	60.2 d	603, 1691
Arsenic	^{76}As	26.5 h	559, 657
Bromine	^{82}Br	35.5 h	554, 777, 619
Cesium	^{134}Cs	2.06 y	569, 605, 796
Cobalt	^{60}Co	5.27 y	1173, 1333
Copper	^{64}Cu	12.7 h	511
Dysprosium	^{165}Dy	2.33 h	95, 280, 362
Europium	^{154}Eu	8.59 y	723, 873, 1275
Gallium	^{72}Ga	14.1 h	834, 2202
Gold	^{198}Au	2.69 d	412
Hafnium	^{175}Hf	70 d	344
	^{181}Hf	42.4 d	134, 482
Holmium	^{166}Ho	26.9 h	81, 1379
Indium	^{116}In^{m}	54 min	1097, 1294
Iodine	^{128}I	25.1 min	443
Lanthanum	^{140}La	40.2 h	329, 487, 1596
Manganese	^{56}Mn	2.58 h	847, 1811, 2113
Mercury	^{203}Hg	46.6 d	279
Praseodymium	^{142}Pr	19.1 h	1576
Rhenium	^{186}Re	90.6 h	137
Ruthenium	^{103}Ru	39.3 d	497
Samarium	^{153}Sm	46.8 h	103
Scandium	^{46}Sc	83.8 d	889, 1121
Selenium	^{75}Se	120 d	136, 265
Silver	^{110}Ag^{m}	250 d	658, 885, 1384
Sodium	^{24}Na	15.0 h	1368, 2754
Tantalum	^{182}Ta	115 d	1121, 1221
Terbium	^{160}Tb	73 d	299, 879, 965
Thorium	^{233}Pa	27 d	312
Tungsten	^{187}W	24 h	480, 686
Vanadium	^{52}V	3.8 min	1434
Ytterbium	^{175}Yb	4.2 d	283, 396
Zinc	^{65}Zn	244 d	1115

2.1 Sample and Comparator

In RNAA the analytical sample is just as prone to contamination before activation, as it is in all other analytical methods. The characteristic insensitivity to contamination applies only after completion of the activation process, and it is therefore important that any conditioning or preprocessing of a sample be kept at an absolute minimum.

Some reactors require biological samples to be dried or lyophilized before activation in order to reduce formation of gases by radiolysis of the water content; this may, however, require special tracer studies to ascertain that no loss of determinand is incurred.[10] At low levels of determinand any such preprocessing necessitates the introduction of a blank to be analyzed together with a sample in order to monitor possible contamination occurring before activation.

Although it is possible to carry out neutron activation analysis without activating an isotopic comparator together with the sample,[9a] this is rarely done for RNAA. The limited number of elements that can be determined in the same sample makes the use of individual comparators much more attractive, because of the greater freedom it gives with respect to both activation and counting conditions.

A comparator is prepared from a weighed quantity of the element to be determined or a compound with an accurately known stoichiometry,[11] chosen so that no other significant activity will be present at the time of measurement.

The validity of the comparator requires that it be activated under exactly the same conditions as the sample with respect to both time and space. It must be physically as similar as possible to the sample at the time of counting, i.e. after completion of the radiochemical separation: same counting vial, same volume, similar matrix, etc. However, the comparator should *not* be processed, in order to eliminate any possible risk of loss, which means that $R_c \equiv 1$, and the rate constant for the comparator C_x^* is reduced to Equation (3)

$$C_x^* = P_i P_m \tag{3}$$

The comparator need not be identical or even similar to the sample, as long as it is exposed to the same fluence of neutrons. The sample, therefore, can be of any size or shape from up to perhaps several kilograms[12] to practically weightless quantities, still using the same comparator. In some cases it may be appropriate to weigh the analytical sample, as well as to take an aliquot of the comparator after completion of the activation.

2.2 The Activation Process

Activation of a sample by neutrons could be carried out with a number of devices, such as accelerators, cyclotrons, radioactive sources,[13] or even spallation sources; nuclear reactors remain, however, the most suitable source of neutrons for RNAA.

Research reactors are found in many countries,[14] and most of them offer irradiation services for neutron activation analysis with neutron flux densities up to 10^{14} neutrons $cm^{-2}\,s^{-1}$, which is more than sufficient for the determination of the elements listed in Table 1. More important is the spatial flux homogeneity that is needed to ascertain that samples and comparator irradiated together are exposed to exactly the same number of neutrons.

Irradiation of an element with a relative atomic mass A_r at a constant neutron flux density φ for a period of time t_i, produces a specific activity of:

$$\frac{N_A \sigma_a \varphi (1 - e^{-\lambda t_i})}{A_r} \text{Bq g}^{-1}$$

where σ_a is the appropriate cross-section of the element for the formation of a specific radionuclide, and N_A is the Avogadro number.

The *saturation activity* achieved after an infinite irradiation time is numerically equal to the rate of formation of the indicator nuclide (Equation 4)

$$P_i = \frac{N_A \sigma_a \varphi}{A_r} \tag{4}$$

Some small reactors with modest neutron flux densities like the Canadian Slowpoke type[15] may be so stable in time that sample and comparator need not even be irradiated simultaneously. In most nuclear reactors, however, changes in neutron flux exposure during irradiation may take place as a result of the introduction or removal of other samples, movement of control rods, change of power level, moderator temperature, or virtually intermittent operation, as well as many other things beyond the control of the analyst. These changes will affect indicator isotopes differently because of their different half-lives; however, the simultaneous irradiation of a comparator of the element to be determined, the determinand, will automatically account for such changes.

Subject to the needed sensitivity the duration of irradiation should be kept as short as possible[9b] in order to:

1. keep the total activity of the sample low enough for proper handling without excessive shielding and time-consuming use of remote handling equipment;
2. minimize radiation decomposition of the sample, which may result in the loss of halogens, mercury,[16] or other elements, when the irradiation capsule is opened;
3. reduce the influence of long-lived activated matrix components on the choice of counting conditions;
4. limit the nuclear interference from double neutron capture in some elements.[17]

2.3 Radiochemical Separation

During the period between the end of activation and the beginning of measurement the analytical sample is subjected to a radiochemical separation, which should be completed preferably within one mean life of the indicator.[9b] The performance characteristics of the radiochemical separation therefore must include the time needed to complete the task, which in practice limits the use of RNAA to elements with indicator half-lives in excess of a few minutes.

Sample decomposition in RNAA presents the same problems as in all other methods of chemical analysis;[18] to ascertain carrier equilibration the procedure should include a complete oxidation–reduction cycle. The process can hardly be completed in less than several minutes, which is just feasible for V in human serum with an indicator half-life of 3.8 min,[19] but barely enough for Al with 2.2 min half-life.[20]

The absorption of a neutron gives rise to the emission of one or more capture γ-rays with energies of several MeV, which causes a nuclear recoil energy of $536 \times E_\gamma^2 / A_r$ eV which is enough to break a chemical bond. Because of this Szilard–Chalmers effect the activated atom no longer retains its original position in a chemical entity, and RNAA is therefore incapable of determining any speciation of an element. By

the same token, however, postirradiation carrier equilibrium does not necessarily require complete sample decomposition, but only complete dissolution. This is an important difference from other reference methods, such as isotope dilution mass spectrometry (IDMS), that rely on complete isotopic exchange between an added isotopic tracer and the total content of an element in the sample.

The addition of carrier is required not only for all elements that need to be determined, or rather their indicators, but also for some interfering elements, from which a good separation is important. Such a *hold-back carrier* could prevent adsorption or coprecipitation that might result in poor and irreproducible separation from interfering radionuclides.

The aim of a radiochemical separation is to achieve a high and reproducible recovery of the indicator radionuclide at the same time as a very low yield of all other radionuclides. The efficiency of the separation is expressed by the *separation factor*,[21] K, which, as shown in Equation (5), is the ratio between the yield of the interfering species, r_i, and the recovery of the determinand

$$K = \frac{r_i}{R_c} \tag{5}$$

K is usually determined by adding a suitable radioactive tracer together with carriers to an unirradiated sample which is then processed and measured exactly as a normal sample.[22] Experimentally determined separation factors are presented in Table 2[9c] for a simple radiochemical separation of Cu in biological materials. A good separation will have separation factors better than 10^{-4}, but in extreme cases a separation factor better than 10^{-7} has been demonstrated.[38] It is worth pointing out that the use of highly effective radiochemical separations with good separation factors makes possible the determination of several other elements with less sensitivity than those listed in Table 1, such as Cr, Fe, Mo, and Cd.[22,23]

Complete separation of the indicator from all other radionuclides in the activated sample is required when the indicator has to be measured by its emission of β-particles in order to achieve sufficient sensitivity.[24] In most cases of γ-emitting indicators a high separation factor is needed only for the dominant interfering elements. Such *interference removal* is often needed for activated matrix components such as ^{24}Na, ^{42}K, and ^{82}Br that dominate the γ-spectra of activated biological samples for days or weeks after the end of irradiation. Successful removal of ^{24}Na is possible by means of hydrated antimony pentoxide,[25] and other matrix indicators may be removed by precipitation with thioacetamide, extraction of Fe with ether, or other conventional methods of chemical separation.[26] It should be recalled that nuclear interference from matrix elements cannot be eliminated by a radiochemical separation, but requires a preirradiation separation.

Each combination of determinand element and matrix composition may have its own optimum radiochemical separation method that minimizes the effort and maximizes the quality of the final result. A general discussion of the factors involved in such an optimization must be based on the separation factors needed for the particular analytical task. This paves the way for the use of less specific *group separations*, the use of *scavengers* to collect the elements to be determined, adsorption on solid oxides, activated carbon etc. Some of these methods seem to work best without carrier addition, but without determination of the recovery of the indicator these methods would suffer from an intrinsic lack of credibility.

2.4 Methodology of Counting

A radiochemically separated sample that contains only the indicator radionuclide could be measured by any suitable instrument with a sufficiently high detection efficiency, and in this respect it would be hard to beat a simple Geiger-Müller (GM) counter or liquid scintillation counter which detects β-particles with close

Table 2 Experimentally determined interferences in a method for the determination of Cu in biological materials by RNAA[9c]

Interfering element	Radionuclide	Separation factor K	Effective value f	mg of element equivalent to 1 µg of Cu
Sodium	^{24}Na	1.4×10^{-5}	0.033	2000
Zinc	^{65}Zn	1.4×10^{-3}	−0.016	40
Gallium	^{72}Ga	3.7×10^{-1}	−0.13	0.02
Arsenic	^{76}As	1.3×10^{-3}	0.93	0.8
Bromine	^{80}Br^{m}	5.0×10^{-5}	0.32	60
Antimony	^{122}Sb	2.6×10^{-2}	0.27	0.15
Tungsten	^{187}W	1.5×10^{-3}	0.96	0.7

to 100% efficiency. It is, however, possible only in very special cases[24,27] to prove that the sample does not contain any other radionuclides that might influence the measurement. In other cases the required high radionuclidic purity may be achieved only at the cost of a lower recovery of the indicator.

Most radiochemically separated samples therefore contain other radionuclides than the indicator and therefore must be measured with instruments that can distinguish between the radionuclides present, usually by measurement of a complete γ-spectrum, where individual radionuclides can be identified by their different γ-energies.

Among such detectors the Na(Tl) scintillation detector is available in the largest sizes, which may even be used to surround the sample; it therefore offers the highest detection efficiency, P_m. Better γ-energy resolution is obtained with the germanium semiconductor detector, which will reduce interference from accompanying radionuclides, but in return carries a higher cost and a lower efficiency.

From the point of view of minimum interference the *optimum decay time*[9b] from end of activation to measurement is given by Equation (6)

$$t_c = \frac{1}{\lambda} - t_i \tag{6}$$

when the indicator is closer to saturation than any of the other radionuclides.

For the sensitivity of the analysis a lower efficiency P_m can be compensated by an increase in *counting time* t_m. However, the increasing influence of radionuclides with longer half-lives on the measurement increases the uncertainty of the measurement of the indicator, so that for a small signal on a high base line nothing is gained[9d] by counting for more than approximately 1.81 half-lives.

Good detection efficiency requires both samples and comparators to be counted close to the detector; in some cases it may even be expedient to count in a well counter. This results in considerable γ-ray and X-ray coincidence losses and corresponding, but identical, changes in the spectra of sample and comparator. Unless such coincidences give rise to spectral interference, the close counting geometry creates no bias in the determination of the indicator. However, P_m becomes highly dependent upon small differences in counting geometry, as well as on the uniformity of the materials. The reproducibility of the counting geometry should therefore be carefully checked,[28] and the identity of the sample and comparator geometry must be ascertained. For optimum results it is recommended that all samples and comparators be counted in liquid form, so as to ensure homogeneity; by measuring the same volumes in identical containers the counting geometry can be kept constant within <1%.

2.5 Data Processing

The processing of γ-spectra in RNAA is very different from the methods employed in INAA (see **Instrumental Neutron Activation Analysis**).

First, we are only interested in the indicator nuclide and therefore need not identify all other radionuclides in the spectrum, unless they contribute to the spectrum in the vicinity of the indicator γ-energies, such as those listed in Table 1.

We are, however, interested in the best possible quantitation of the indicator with the minimum uncertainty and without interference, even close to the limit of detection. In the evaluation of small-to-invisible photo peaks on a high Compton continuum, the determination of the number of counts that can with a high degree of certainty be attributed to the indicator is best carried out by simple peak integration according to Covell, and not by fitting the small number of counts to an assumed Gaussian or other peak shape.

The optimum choice of peak boundaries for peak area evaluation depends upon the full width at half-maximum (FWHM) and the relative height of the photo peak to the base continuum. The optimum half-width has a limiting value of 0.90758 FWHM for photo peak heights approaching zero, and Figure 2 shows optimum half-widths for peak area integration as a function of relative peak heights and FWHM.[29]

The FWHM is a function of γ-energy E

$$(\mathrm{FWHM})^2 = wE + w_0 \tag{7}$$

where the coefficients w and w_0 in Equation (7) are determined from the actual spectrum by weighted linear regression of FWHM^2 for the largest photo peaks on their corresponding γ-energies or channel numbers. These peaks would be the same as used for energy calibration by simple linear regression of γ-energies on channel numbers and do not necessarily include any photo peak of the indicator nuclide.

The FWHM corresponding to the chosen photo peak of the indicator is now used in combination with Figure 2 to select the peak boundaries for optimum precision of the partial peak area evaluation according to Covell. The partial peak area is then converted to the full peak area by assuming a Gaussian shape with the assumed FWHM, which is a purely mathematical operation that does not change the relative precision of the measurement. The final result of the data processing is the full peak area in counts together with its associated standard deviation, calculated on the basis of a Poisson distribution for each individual channel content. An implementation of this approach has been verified by comparison with IAEA (International Atomic Energy Agency) test spectra with

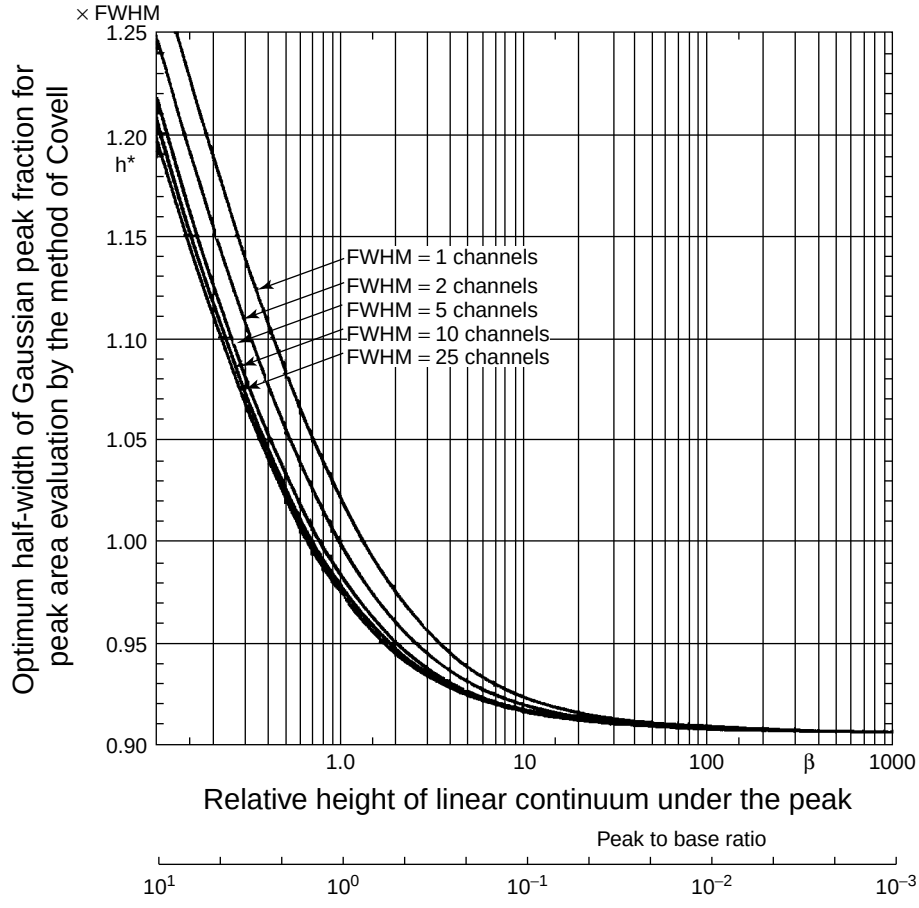

Figure 2 Optimum half-width of Gaussian peak fraction for peak area evaluation by the method of Covell as a function of height of linear continuum under the peak for selected values of FWHM. (Reprinted with permission from K. Heydorn, W. Lada, *Anal. Chem.*, **44**, 2314.[29] Copyright 1972, American Chemical Society.)

known peak areas, as well as with actual spectra with known ratios.[28]

The possible interference from other radionuclides is not estimated by attempting to resolve doublets or multiplets, but by the measurement of the effective values f of potentially interfering elements.[30] By irradiating a known quantity of an interfering element exactly like a sample together with a comparator of the determinand, and using the peak boundaries for the determinand, the *effective value* is determined as the ratio between counts per unit mass of interfering element relative to determinand. For elements with effective values in excess of 0.01 separation factors should be determined in order to specify the assumptions of the RNAA.

Table 2 lists effective values for seven elements suspected of interfering with the determination of Cu by RNAA.[9c] This allows the user to judge the performance characteristics of the method for the analysis of actual samples.

Interference from the formation of the indicator nuclide during activation from another element by (n,p), (n,α) reactions, or by double neutron capture, is usually referred to as *nuclear interference* and can be corrected for only by the simultaneous activation of an additional comparator of the interfering element.[17] A particularly treacherous interference occurs when the indicator nuclide is also produced by fission of uranium, which may necessitate a separate determination of this element[31] in order to correct for this positive bias.

Sometimes the absence of interference may be demonstrated by the determination of the same element using two different indicator nuclides.[32]

2.6 Recovery Correction

The determination of recovery or chemical yield of the analytical separation is an outstanding property of RNAA and enables the method to estimate the true content of the determinand in the sample.

In routine applications of highly standardized radiochemical separations the chemical yields are often determined by adding radioactive tracers of the determinands

to an inactive sample before performing the separation, exactly as described for the determination of interferences. The use of such a priori chemical yields instead of individual recoveries is justified only when the yields are close to 100% and the correction therefore quite small.[33]

In research applications, where every analysis might differ from the rest, recovery is usually not quantitative and is susceptible to considerable variation from sample to sample, which means that recoveries must be determined for each individual analysis. In that case much larger corrections for chemical yield can be accepted, while still maintaining statistical control.

A comparison of the amount of carrier in the separated sample with the originally added quantity by almost any analytical method could be used for the determination of recovery, and in many applications gravimetry has been used. However, such chemical methods might jeopardize the principle of *not* processing the added carrier reference comparator sample, and therefore a direct instrumental method without any additional treatment and a minimum of handling of the separated sample is clearly preferable, and simple radioanalytical methods based on either a short re-irradiation or the addition of a radioactive tracer, are often the methods of choice:

1. With carrier added at the milligram level the irradiation time can be very short, and the measurement of the induced radioactivity can be very simple.[34] By simultaneous irradiation and direct comparison between added and recovered carrier there is no need to know the amount of carrier accurately, provided the content in the sample is negligible in comparison to the added quantity of carrier.
2. Addition of a radioactive tracer together with the carrier opens the possibility of determining recovery at the same time as measurement of the indicator. Clearly, the tracer should not be produced by neutron activation and the amount added should be chosen carefully in order not to disturb the measurement.[9e] A list of radionuclides used as tracers for the determination of individual recoveries of added carrier is presented in Table 3. If they are truly carrier-free they may even be added to the sample before activation, so that any losses that might occur as a result of preirradiation processing would be included in the recovery.[10]
3. Instead of adding the radioactive tracer to the sample before separation it may be added after completion of the separation and the measurement of the indicator, in order not to cause spectral interference. It is now possible to correct for recovery during separation by comparing the specific activities of the carriers added to the separated sample and the carrier reference comparator sample by using *substoichiometric* methods.[35]

It has been proposed that an internal standard of another element be used to correct for recovery;[36] however, this is limited to the cases where this element yields a radionuclide isotopic with the indicator nuclide.

If carefully thought out, the individual determination of recovery for each sample requires very little additional effort in comparison to the radiochemical separation.[37]

2.7 Final Calculation of Analytical Result

The calculation of results obtained by RNAA is basically similar to that of INAA; however, the proper utilization of the various measurements for correcting initial results and estimating their uncertainty may become quite complicated. It is, therefore, strongly recommended that

Table 3 Indicators and tracers for individual yield corrections in RNAA

Determinand	Indicator	Half-life	γ-Energy (keV)	Tracer	Half-life	γ-Energy (keV)
Antimony	^{122}Sb	2.8 d	564	^{125}Sb	2.8 y	408
Arsenic	^{76}As	26 h	559	^{74}As	17.8 d	596
Cadmium	^{115}Cd	53.5 h	528	^{109}Cd	463 d	88
Cobalt	^{60}Co	5.3 y	1332	^{57}Co	271 d	122
Copper	^{64}Cu	12.7 h	511	^{67}Cu	62 h	184
Gallium	^{72}Ga	14 h	834	^{67}Ga	78 h	93
Gold	^{198}Au	2.7 d	412	^{195}Au	183 d	99
Iodine	^{128}I	25 min	443	^{126}I	13.0 d	388
Manganese	^{56}Mn	2.6 h	847	^{54}Mn	312 d	835
Mercury	^{197}Hg	65 h	78	^{203}Hg	47 d	279
Selenium	^{75}Se	120 d	265	^{81}Se^{m}	57 min	103
Tin	^{123}Sn^{m}	40 min	160	^{113}Sn	115 d	392
Tungsten	^{187}W	24 h	480	^{181}W	121 d	65
Uranium	^{239}Np	56 h	β	^{237}Np	2×10^{6} y	α
Vanadium	^{52}V	3.8 min	1434	^{48}V	16.0 d	984

calculation be done in the form of a spread-sheet, not only in order to avoid making mistakes, but also because it facilitates the implementation of quality control, as well as the smooth propagation of errors from one stage to the next.

As an example let us pretend that we have carried out RNAA for two determinands in the same sample, and for simplicity let us assume that both can be determined from the same γ-spectrum, but with some interference from one indicator to the other. We also assume that comparators of both determinands were irradiated together with the sample. The data available for calculating the concentration of determinands in a sample may be organized as shown in Table 4.

Calculations are now carried out in the following sequence:

1. Photopeak counts are corrected to count rates at the time of pile-out by multiplication with a decay correction for

Table 4 Data used to calculate concentration of determinands

Cell content	Column
Sample identification	A
Irradiation time t_i	B
Date and time of pile-out	C
Weight of sample	D
Recovery 1	E
estimated standard deviation	F
Recovery 2	G
estimated standard deviation	H
Counting time t_m	K
Date and time of counting	L
Photopeak counts 1. determinand	M
estimated standard deviation	N
Photopeak counts 2. determinand	Q
estimated standard deviation	R
Photopeak counts 3. determinand	U
estimated standard deviation	V
Comparator 1. determinand	AA
Content of determinand	AD
Counting time t_m^*	AK
Date and time of counting	AL
Photopeak-1 counts	AM
estimated standard deviation	AN
Photopeak-2 counts	AQ
estimated standard deviation	AR
Comparator 2. determinand	BA
Content of determinand	BD
Counting time $t_m^\#$	BK
Date and time of counting	BL
Photopeak-1 counts	BM
estimated standard deviation	BN
Photopeak-2 counts	BQ
estimated standard deviation	BR

t_c = date and time of counting − date and time of pile-out

$$\text{decay correction} = \frac{\lambda e^{\lambda t_c}}{1 - e^{-\lambda t_m}} \quad (8)$$

In Equation (8) λ is the decay constant for the radionuclide associated with the photopeak in question.

Decay corrections are conveniently placed in the columns immediately after the counting data:

	Columns		
• Sample	O	S	X
• Comparator	AO	AS	
• Comparator	BO	BS	etc.

2. The uncorrected content of 1 determinand in the separated sample and its corresponding standard deviation are calculated as:

$$\text{content} = \text{AD} \times \frac{\text{M}}{\text{AM}} \times \frac{\text{O}}{\text{AO}}$$

and placed in column CA

$$\text{CV}^2 = \left(\frac{\text{N}}{\text{M}}\right)^2 + \left(\frac{\text{AN}}{\text{AM}}\right)^2$$

$$\text{standard deviation} = \text{SQRT}(\text{CV}^2) \times \text{content}$$

and placed in column CB

3. The uncorrected content of 2 determinand in the separated sample and its corresponding standard deviation are calculated as:

$$\text{content} = \text{BD} \times \frac{\text{Q}}{\text{BQ}} \times \frac{\text{S}}{\text{BS}}$$

and placed in column CC

$$\text{CV}^2 = \left(\frac{\text{R}}{\text{Q}}\right)^2 + \left(\frac{\text{BR}}{\text{BQ}}\right)^2$$

$$\text{standard deviation} = \text{SQRT}(\text{CV}^2) \times \text{content}$$

and placed in column CD

4. Correction of the influence of 2 determinand on the result for the 1 determinand requires the calculation of the effective value:

$$\text{effective value} = \frac{\text{AD}}{\text{BD}} \times \frac{\text{BM}}{\text{AM}} \times \frac{\text{BO}}{\text{AO}}$$

and placed in column CE

$$\text{CV}^2 = \left(\frac{\text{AN}}{\text{AM}}\right)^2 + \left(\frac{\text{BN}}{\text{BM}}\right)^2$$

$$\text{standard deviation} = \text{SQRT}(\text{CV}^2) \times \text{effective value}$$

and placed in column CF

5. The corrected content of the 1 determinand in the separated sample is now calculated as:

   content = CA − CC × CE

   and placed in column CG

   $$CV^2 = \left(\frac{R}{Q}\right)^2 + \left(\frac{BR}{BQ}\right)^2 + \left(\frac{BN}{BM}\right)^2$$

   $$\text{standard deviation} = \text{SQRT}\left(CA \times \frac{N}{M}\right)^2 + CV^2(CC \times CE)^2 + \left(CG \times \frac{AN}{AM}\right)^2$$

   and placed in column CH

   At this stage it should be checked whether CG is negligible in comparison with the quantity of added carrier; if not the recovery E should be reduced by subtracting their ratio.

6. The real concentration of determinand in the sample material and its uncertainty are finally calculated from the original sample weight and the recovery of the appropriate carrier in the separated sample:

   $$\text{concentration} = \frac{CG}{D \times E}$$ is placed in column CK

   $$CV^2 = \left(\frac{F}{E}\right)^2 + \left(\frac{CH}{CG}\right)^2$$

   standard deviation = SQRT(CV^2) × ABS(CK)

   and placed in column CL

An example of actual results obtained from the determination of Pt in a reference material from the Bureau Communautaire de Réference (BCR) certified reference material (CRM) 186 pig kidney[38] is shown in Table 5, which also indicates the spreadsheet columns, from which these results are taken. The indicator for Pt is ^{199}Au, formed by the decay of ^{199}Pt, but also from neutron capture in ^{198}Au, the indicator for gold. The correction for nuclear interference is in this case comparable to the total amount of ^{199}Au and therefore leads to very large relative uncertainties in the final results for Pt.

Corrections for nuclear interference from uranium in the determination of La, Ce, Sm, Nd, Zr, Mo, Ru, Te, etc. may be carried out in a similar way,[9f] except that the concentration of U in the sample has to be determined separately.

This paradigm can easily be modified to accommodate other situations in RNAA with additional determinands and γ-spectra, using the same general approach to calculation of results and their uncertainties.

Corrections for possible blank values, such as those associated with preirradiation treatment of the sample,[33] must be carried out by calculating the blank value exactly as if it were a sample and then subtracting from the final result for the sample itself.

3 INSTRUMENTATION

Radiochemical separation requires a radioanalytical laboratory of Type C,[39] which is essentially just a good quality chemical laboratory with a radiochemical hood and a little shielding to reduce radiation exposure.

Requirements for irradiation facilities and counting equipment do not go beyond those needed for INAA, but the emphasis on some of the features and specifications is shifted because of the nature of the separated sample and the low levels of activity of the indicator nuclides.

3.1 Irradiation Systems

In order to benefit most from the properties of RNAA it should be possible to activate rather large samples without any preceding treatment that might lead to losses or contamination. As with INAA it is important that the entire sample and associated comparators be exposed to identical neutron flux densities for exactly the same time.

This is best achieved in a system that permits rapid transport in and out of the irradiation zone, while providing rotation of the irradiation container during activation.[40] In order to minimize sample decomposition

Table 5 Determination of the platinum content in BCR CRM 186 pig kidney by RNAA[38]

Sample size(g)	Recovery (%)	Apparent platinum (ng) ± standard deviation	Observed gold (ng) ± standard deviation	Effective value ratio ± standard deviation	Actual platinum (ng) ± standard deviation
2.23	76.4	13.6 ± 0.8	3.206 ± 0.004	4.50 ± 0.04	−0.9 ± 0.8
2.38	92.5	13.6 ± 0.5	3.713 ± 0.004	3.50 ± 0.03	0.6 ± 0.5
2.19	72.2	10.7 ± 0.9	3.161 ± 0.003	3.84 ± 0.03	−1.4 ± 0.9
2.25	95.7	18.5 ± 0.9	4.436 ± 0.005	4.36 ± 0.05	−0.8 ± 0.9
D	E	CA ± CB	CC ± CD	CE ± CF	CG ± CH

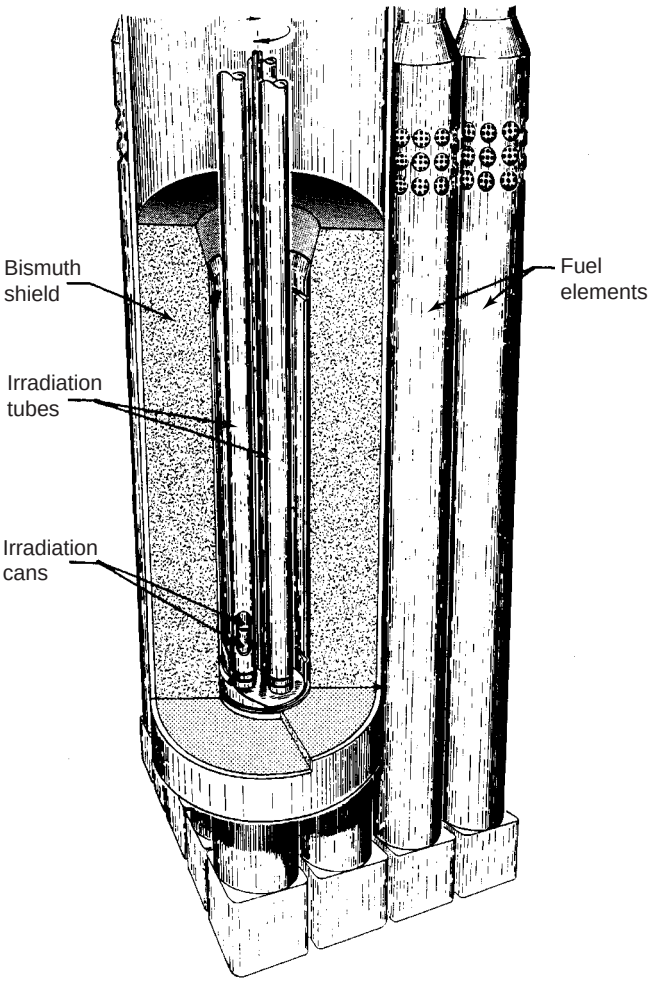

Figure 3 Cut-away of the rotating rig facility in the Danish DR 3 reactor. The three irradiation tubes in the rotating assembly are surrounded by 50 mm of bismuth in order to reduce γ-ray exposure. Carbon dioxide gas is used for transportation of the irradiation cans in and out of the reactor during operation.[9] (Reprinted with permission from K. Heydorn, *Neutron Activation Analysis*, Chapter 3, Uniscience 1984. Copyright CRC Press, Boca Raton, FL.)

the thermal neutrons should be accompanied by a minimum of epithermal and fast neutrons, as well as γ-radiation.

In Figure 3 is shown a drawing of such a facility in the Danish Reactor DR 3, which permits simultaneous irradiation of three containers of about 30 cm^3, while they are rotated in a thermal neutron flux density of 3×10^{13} neutrons cm^{-2} s^{-1} shielded by 50 mm of bismuth to reduce the γ-ray intensity.

3.2 Counting Equipment

A complete radiochemical separation of the indicator nuclide would allow use of the very sensitive but *nonspecific* detection of β-particles by very simple counting equipment. However, no separation is perfect, and in practice the less sensitive, but high-resolution detection of γ-rays is preferable, wherever possible.

With the availability of germanium semiconductor detectors having efficiencies comparable to, or even surpassing, that of the $3'' \times 3''$ sodium-iodide scintillation detector, the instrumentation for RNAA becomes similar to that needed for INAA with the same electronic equipment and multichannel analyzer with at least 16 K (K = 1024) channels.

The *detection efficiency* P_m determines the sensitivity of determination S, while an *improved resolution* of the detection system improves both the precision and accuracy of the determination by:

1. facilitating the detection of unexpected interferences
2. minimizing the correction for known spectral interferences
3. reducing the standard deviation of the photopeak area
4. improving the detection limit.

Similar improvements may sometimes be achieved by the use of two or more detectors in combination with coincidence/anticoincidence circuitry to obtain spectra providing improved detection limits with only a slight loss of sensitivity.[41]

As previously discussed the close counting geometry of the separated sample must be exactly equal to that of the comparator, calling for an accurately reproducible counting position to ascertain exactly equal counting efficiency and coincidence losses for both. A counting assembly with such reproducible counting positions for liquid samples is shown in Figure 4.

The usually low activity of the separated samples makes corrections for dead time and pile-up less critical, and a simple 50–60 Hz pulser with a narrow peak in the highest channels of the spectrum is a reliable way of making such corrections. The use of digital signal processing (DSP) improves resolution at high count rates, but seems to give no particular advantages at low count rates.

4 QUALITY ASSURANCE

While RNAA is potentially capable of providing analytical results without bias and with known uncertainty, the method clearly requires no less skill and experience than other methods. Quality assurance is therefore a sine qua non, and fortunately the method is particularly well suited for modern methods of quality control/quality assurance (QC/QA) based on the so-called BIPM (Bureau International des Poids et Mesures) philosophy.[42]

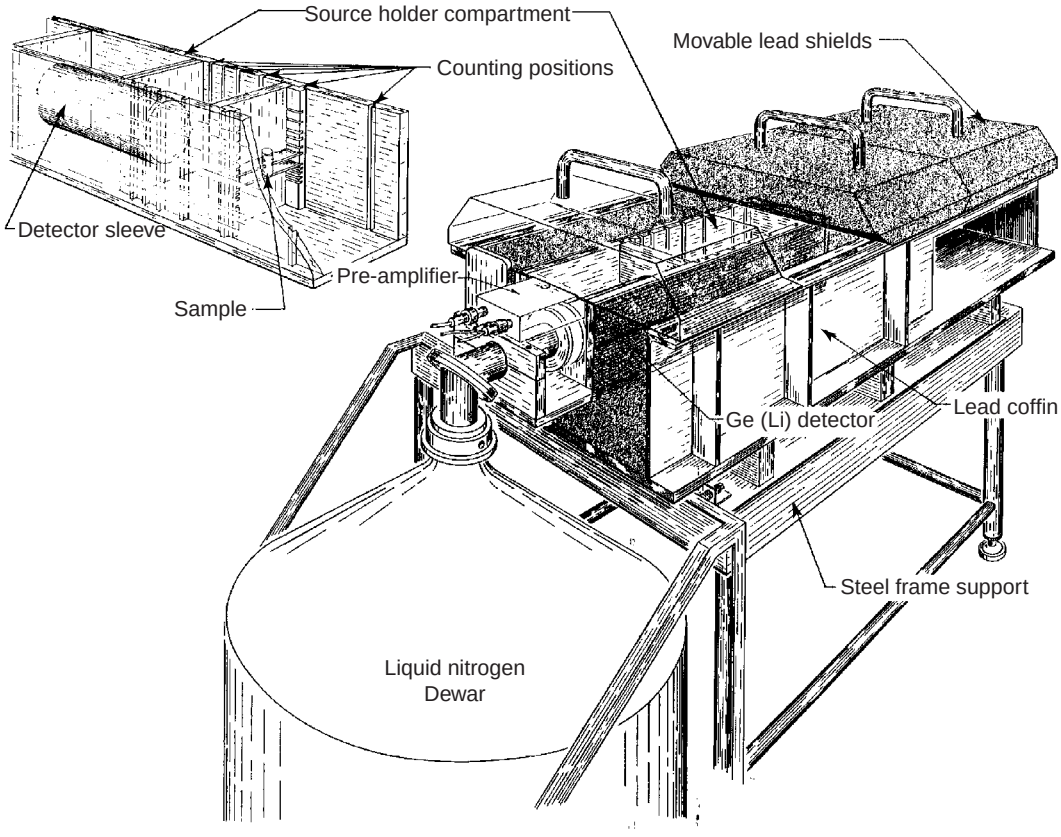

Figure 4 Horizontal lead coffin for counting with a germanium semiconductor detector. The perspex box shown as an insert illustrates that the modifications required by using different detectors do not affect the source compartment. The coffin offers 100 mm of lead shielding, but is open at the detector end that usually faces a wall. (Reprinted with permission from K. Heydorn, *Neutron Activation Analysis*, Chapter 3, Uniscience 1984.[9] Copyright CRC Press, Boca Raton, FL.)

The implementation of this line of thought is described in a basic ISO (International Organization for Standardization) document,[43] and for use in analytical chemistry in a Eurachem Guide;[44] none of these documents, however, provide any guidance with respect to RNAA or to activation analysis in general.

4.1 Statistical Control

The analytical method is in a state of statistical control when its known sources of variation fully account for the observed variability among replicate determinations.

The *a posteriori variability* of n replicate analyses is expressed as the variance s^2 of results y_i in Equation (9)

$$s^2 = \frac{\sum^{n}(y_i - y_0)^2}{n-1} \tag{9}$$

where y_0 is their mean value.

When the analytical method is completely defined, it is possible to identify and evaluate all sources of variation by statistical or other methods, referred to as uncertainty components of type A or B.[43] Their influence on the analytical results is calculated by the law of error propagation on the basis of variances $\widehat{\sigma}^2$ of p individual components x_i (Equation 10)

$$\widehat{\sigma}_0^2 = \sum_{1}^{p} \left(\frac{\partial y}{\partial x_i} \right)^2 \widehat{\sigma}_i^2 \tag{10}$$

This resulting variability may be calculated without actually carrying out any analyses and is therefore referred to as the *a priori precision*.[45]

In RNAA and other methods where the process of counting stochastic events is part of the analytical method the statistics of counting has to be compounded with the a priori precision, even though its contribution is not known until an actual measurement has been carried out. The calculation of *counting statistics* is based on the Poisson distribution for the number of counts detected from a radioactive source, and this assumption has been amply demonstrated to be fulfilled in neutron activation analysis. Each photopeak area is associated with a standard

deviation, σ_C, calculated from the same spectrum and by the same computer program. The total expected a priori variability of radioanalytical results then becomes Equation (11)

$$\widehat{\sigma}^2 = \widehat{\sigma}_0^2 + \sigma_C^2 \tag{11}$$

where σ_C^2 expresses the contribution from counting statistics.

Agreement between a priori and a posteriori variability means that the analytical method is in a state of statistical control; this is tested by calculating the statistic T by Equation (12)

$$T = \sum_1^n \frac{(y_i - \widehat{\mu})^2}{\widehat{\sigma}_i^2} \tag{12}$$

where $\widehat{\mu}$ is calculated by Equation (13)

$$\widehat{\mu} = \frac{\sum_1^n y_i \widehat{\sigma}_i^{-2}}{\sum_1^n \widehat{\sigma}_i^{-2}} \tag{13}$$

as the weighted mean of n observations. If there is no significant difference between the expected–a priori–and the actual–a posteriori–standard deviations the statistic T is closely approximated by a χ^2-distribution with $n-1$ degrees of freedom.[46] Significant deviations from this distribution show that there is a disagreement between the sources of variability expected to influence the result according to Equation (11) and the observed variability. The process of identifying the sources and magnitudes of variability that influence the actual analytical results is called the Analysis of Precision.[47]

4.2 Verification Methods

Verification means to verify the absence of errors, and the analysis of precision is an important tool for ascertaining the absence of unknown sources of variability. It is also possible to use this statistic to demonstrate the absence of systematic errors.[48]

The most important condition for relying on the analysis of precision is that the standard deviations σ_i of the analytical results are correctly estimated for each individual measurement. Hence, it must be ascertained that the program used to calculate the results and its associated uncertainty correctly takes into account the contributions from counting statistics and other known uncertainty components. This was tested by a series of experiments involving replicate activation and counting of identical samples under a variety of conditions, covering the entire field of potential application.[28]

Table 6 Verification of calculated counting statistics for photopeak areas covering almost two orders of magnitude and having FWHM ranging from 2.8 to 15 channels

γ-Energy (keV)	Total peak area counts ± SEM[a]	Replicates n	Statistic T
245	43 440 ± 72	29	23.23
344	103 578 ± 95	29	45.95
368	3 166 ± 38	29	26.29
411	7 281 ± 34	29	22.02
444	9 244 ± 39	29	27.40
689	1 673 ± 30	29	23.07
779	21 074 ± 52	27	30.55
867	6 150 ± 44	23	25.25
964	19 102 ± 49	19	29.64
1112	15 362 ± 48	17	25.66
1213	1 505 ± 25	14	10.58
1299	1 629 ± 19	13	8.53
1408	19 319 ± 45	13	16.01
All energies		300	314.18

[a] SEM = Standard error of the mean = Standard deviation/$\sqrt{n}$.

As an example it was considered necessary to ascertain that counting statistics were correctly calculated for spectra recorded with different gains, which means that identical photopeaks had different widths expressed as FWHM. By counting the same ^{152}Eu source up to 30 times at different gain settings and using the most intense γ-line at 122 keV as a comparator the results shown in Table 6 were obtained. For each set of photopeak areas and associated counting statistics, σ_C, the statistic T was calculated in accordance with Equations (12) and (13). All photopeak areas covering an intensity range of two orders of magnitude and values of FWHM from 2.8 to 15 channels were found to be in good statistical control with the possible exception of the 344 keV photopeak. This photopeak offers the highest precision next to the comparator peak and is therefore providing the highest sensitivity for the detection of unexpected sources of variability; however, no correlation with FWHM could be observed and it was therefore concluded that, although there might be a small additional random source of variation, there was no systematic effect associated with the changes in photopeak width.

Once it has been verified that all known sources of variation are correctly accounted for in the calculated standard uncertainty of the analytical result, it becomes possible to detect and identify unknown sources of variability by judicious use of the analysis of precision.[47] Verification of the absence of unknown sources of variability in the analytical results, however, does not exclude the presence of a constant bias or proportional error. Only comparison with results obtained by independent methods or the analysis of materials with known composition can be used for such verification.

Table 7 Results for selenium in mg/kg by neutron activation analysis of a series of BCR certified biological reference materials

Certified reference material			Analytical result	
BCR code	Type of material	Confidence interval	Mean ± SEM	Ratio ± Standard deviation
CRM 184	Bovine muscle	183 ± 12	179 ± 3	0.98 ± 0.03
CRM 185	Bovine liver	446 ± 13	453 ± 27	1.02 ± 0.06
CRM 186	Pig kidney	10 300 ± 500	10 860 ± 170	1.05 ± 0.03
CRM 189	Wholemeal flour	132 ± 10	127 ± 3	0.96 ± 0.04
CRM 278	Mussel tissue	1660 ± 40	1647 ± 25	0.99 ± 0.02
CRM 279	Sea lettuce	593 ± 32	600 ± 22	1.01 ± 0.05
CRM 281	Rye grass	28 ± 4	22.2 ± 2.5	0.79 ± 0.11
			Weighted mean	1.00 ± 0.01
			T	8.17
			Degrees of freedom	6
			$P(\chi^2 \geq T)$	0.23

The preferred way of verifying the absence of such errors is to analyze a series of matrix-matched reference materials with certified concentrations of the determinand for the entire range of concentrations of interest. Results of such a verification analysis for selenium by INAA and RNAA are shown in Table 7 giving results for seven biological reference materials certified by BCR and covering a concentration range of more than two decades.[28] On the assumption that the confidence intervals for the certified values represent the standard uncertainty with a coverage factor $k = 2$, we can combine the uncertainties with the standard uncertainties for the actual results. The ratios between the observed and the expected results are calculated together with their associated standard deviations and presented in the last column of Table 7.

The weighted mean of these ratios is calculated according to Equation (13) and the distribution of the results is tested by the T-test from Equation (12). In this case the test shows no disagreement with the expected ratio of one, and therefore no detectable systematic error. Within the range of concentrations and matrices investigated we can therefore disregard blank or contamination problems, as well as interference or calibration errors.

4.3 Traceability Chain

Traceability is characterized by an *unbroken chain* of comparisons that relates the analytical result to a national or international realization of the appropriate SI units.[49] Each step of the traceability chain must be associated with an expression of uncertainty in accordance with the ISO document,[43] so that the combined uncertainty of the analytical result can take this into account.

Traceability is not so difficult to establish in the physical measurement of a property of a particular sample, including elemental analysis by INAA. Traceability in chemical analysis is much more complicated for two reasons:

- The traceability chain is broken as soon as a chemical separation takes its beginning, i.e. already at the sample dissolution stage.
- There are a huge number of chemical species that may require determination, some being unavailable in pure condition.

In this respect RNAA is superior to other analytical methods in keeping the traceability chain unbroken because:

- carrier addition and equilibration take place right from the dissolution stage, thereby restoring the traceability chain by the determination of the recovery of added carrier;
- the number of elements that can be determined by RNAA is not more than approximately 70; the majority of these can be used directly as comparators in the form of pure elements or compounds with known stoichiometry.[11]

It must, however, be remembered that RNAA determines a particular isotope of the element to be determined and therefore relates directly only to the same isotope in the comparator. Thus, the indicator radionuclide represents the element as such only if the isotopic composition of the comparator exactly equals that in the sample. For the most sensitive elements listed in Table 1 this condition is essentially always fulfilled; but in special cases the isotopic traceability becomes of major concern.

Neutron activation analysis was for example used for the determination of the specific activity of ^{125}I[50] and ^{129}I[51] using ^{126}I, ^{128}I, and ^{130}I as indicators. In this case total iodine content had a completely different isotopic composition from naturally occurring iodine and could

only be determined by using individual comparators for all nuclides, each having its own traceability chain.

The uncertainty with which a result can be reported is thus limited by the uncertainty of the traceability chain, which should therefore be kept at a level where its contribution is small in comparison with other uncertainty components. This is usually achieved by the use of certified reference materials from internationally recognized sources like BCR or NIST (National Institute of Standards and Technology).

In a method like RNAA with inherently small uncertainties and a short traceability chain the contribution to the uncertainty from the traceability chain can be kept at a level of less than 1% when several different pure isotopic comparators are used.[52] The use of the so-called k_0 method[53] based on the use of comparators of elements different from the determinand also represents a break in the traceability chain and is therefore discouraged in the context of using RNAA for certification purposes.

4.4 Uncertainty Budget

When standard deviations, σ_i, have been estimated for all uncertainty components, p, discussed in section 2, the uncertainty budget is composed simply by listing the individual items of the summation in Equation (10). Common to all RNAA budgets are sample weight, quantity of comparator, half-life of indicator, neutron flux density variation, measurements of decay and counting times, amount of carrier, and recovery. Another important common feature is that the radiochemical separation as such does not contribute to the uncertainty in any other way than expressed by the recovery, which means that the entire a priori standard uncertainty is independent of the concentration of the determinand.

All these major uncertainty components are evaluated by statistical methods and therefore belong to Type A.[43] Other uncertainty components may have to be evaluated by inference from other methods or by previous experience; such contributions are referred to as Type B. In RNAA, uncertainty components of Type B might include several potential losses or gains of determinand or comparator that are not otherwise accounted for, such as

- sample contamination before or during irradiation
- loss of determinand before complete carrier equilibration
- leaching from or adsorption to the irradiation container walls
- influence of moisture etc. on sample weight
- effects of radiolysis
- incomplete sample dissolution.

Many of these uncertainty components apply to both sample and comparator, and estimated corrections should in principle be applied to the results of the analysis. The comparator should have a much higher concentration of determinand than the sample and therefore be much less susceptible to contamination than the sample, but losses might occur from

- incomplete transfer from the irradiation container
- diffusion, adsorption, or precipitation during irradiation.

The experienced analyst should be able to reduce the magnitude of all these effects to the extent that even with an estimated Type B combined standard uncertainty of 100% of the correction to be applied, there would be no significant influence on the uncertainty budget.

In a budget for the determination of chlorine by RNAA[48] all Type A contributions are listed with their individual standard uncertainties, as shown in Table 8. Based on the analysis of precision and the application of the T-statistic from Equation (12) it had been concluded that no detectable contribution from Type B uncertainty components could be detected.

Since RNAA is characterized by the absence of reagent blanks the uncertainty budget contains only relative sources of variation, which facilitates the comparison of the items: it is here immediately recognized that the variation in fluence between sample and comparator is the main contributor to the combined uncertainty, whereas weighings have no detectable effect. The influence of half-life depends entirely on the time lapse between the measurement of the separated sample and the comparator. In this particular case the recovery is determined by re-irradiation, which means that the contributors from its measurement are exactly the same as for the sample itself.

Even in a complicated method like RNAA the combined uncertainty is almost always dominated by one or two uncertainty components, one of which is often the contribution from counting statistics. As pointed out above, this contribution is not known until an actual analysis has been carried out; therefore it does not strictly

Table 8 Sources of variability and their estimated standard uncertainties[48]

Parameter	Mean	σ_a	σ_r (%)
Sample weight	2 g	0.2 mg	
Comparator	2 g	0.2 mg	
Concentration	0.2 mg g^{-1}	–	<0.5
Half-life	$T_{1/2} = 37.24$ min	0.05 min	
Neutron flux density	2.5×10^{13} n cm^{-2} s^{-1}	–	2.3
Decay correction	$t_1 - t_2 = 40$ min	0.1 min	
Carrier	0.2 cm^3	0.6 mm^3	
Photopeak area	>22 000 counts	≅150 counts	<0.7

belong to the a priori precision, but must of course be included in the combined uncertainty.

At very low levels of the determinand the a priori precision becomes insignificant, and the uncertainty is completely attributed to the counting statistics; this contribution can be influenced by the choice of detector, counting geometry and counting time, thereby making an a priori estimation less meaningful.

4.5 Statement of Uncertainty

Results presented without uncertainty are useless because no conclusions can be based on them; results with incorrect uncertainty are dangerous because false and misleading conclusions may be drawn.

The tradition of quoting only uncertainty from counting statistics when reporting results obtained by neutron activation analysis is therefore dangerous and should be supplemented with the a priori precision obtained from the uncertainty budget.

Both the analytical result and its uncertainty must be given in SI units with the appropriate subdivisions, and it is preferable to state the *standard uncertainty*, so that statistical tests and inferences can be made directly from the numbers presented. This applies to all results, including those that are not significantly different from zero, or even negative numbers.

As an example we shall take the numbers for platinum in radiochemically separated samples of BCR 186 that were presented in Table 4. These numbers are typical for analytical results, obtained as a difference between two almost equally large numbers: without any real difference between these numbers the probability of getting positive or negative results is the same, and getting three negative and one positive result from four experiments as in this case, has a probability of 25%. The quoted standard uncertainties confirm that the individual numbers are not significantly different from zero.

Thus, there is no reason to regard these numbers as in any way suspicious and therefore they should also be treated in exactly the same way as all other results. Concentrations are calculated exactly as usual, and results are presented in Table 9 together with their uncertainties. Application of the T-test from Equation (12) shows no significant difference between the results, which can therefore be pooled in accordance with Equation (13). The final result is closer to zero than any of the individual results and its standard uncertainty is smaller.

The standard uncertainty provides a confidence interval corresponding to approximately 67% for a normal distribution and this changes very little with deviations from this distribution. However, in many practical situations this confidence interval is not generally appreciated, and it has been decided to multiply the standard uncertainty with a coverage factor, k,[43] which is frequently taken to be a factor of 2. This gives a coverage of 95% for a normal distribution, but this number may change considerably with deviations from this assumption. Under all circumstances the coverage factor must be reported for all analytical results, so that standard uncertainties can be recalculated before use of the data.

Table 9 Concentrations of Pt in BCR 186 pig kidney found by RNAA[38]

Sample size (g)	Recovery (%)	Actual platinum (ng) ± standard deviation	Calculated concentration ($ng\,g^{-1}$) ± standard deviation
2.23	76.4	-0.9 ± 0.8	-0.5 ± 0.5
2.38	92.5	0.6 ± 0.5	0.3 ± 0.2
2.19	72.2	-1.4 ± 0.9	-0.9 ± 0.6
2.25	95.7	-0.8 ± 0.9	-0.4 ± 0.4
D	E	CG ± CH	CK ± CL
		Weighted mean	-0.06 ± 0.18
		$T = 5.70$ at 3 degrees of freedom	

Instead of reporting negative results or results with large uncertainties some laboratories prefer to present their results simply as *n.d. (not detected)*. In other cases, compliance with legal requirements or limits necessitates reporting safe *upper limits*, which are conventionally set as three times the standard uncertainty. Such deflation of the information content is, however, strongly discouraged and should always be accompanied by the reporting of the original data, positive or negative, with their associated standard uncertainty.

Proper reporting of analytical results makes possible the direct application of the analysis of precision and other statistical tools to help in deciding whether a difference between sets of measurement has been observed or not and thereby helps to avoid drawing false conclusions.

5 PERSPECTIVE AND FUTURE DEVELOPMENTS

A top priority in the analytical perspective is the introduction of the BIPM philosophy[42] in analytical chemistry courses at the university level and the implementation of the corresponding ISO Norm[43] at the practical level, so that analytical results can assume their rightful place in the hierarchy of measurement science. This process has been hindered by the traditionally inexact and ambiguous chemical terminology,[54] and the pontification of incommensurate terms and concepts in some ISO standards.[55] Publication of the EURACHEM guide[44] has been a

very significant step forward, but its recommendations are not yet mandatory even for laboratories seeking accreditation. The methodology of RNAA has contributed significantly to revealing such inadequacies[56] and continues to pioneer the introduction of generally recognized measurement concepts and definitions in analytical chemistry.

5.1 Analytical Metrology

The purpose of analytical metrology is the introduction and implementation of metrology in analysis, based on the successful use of the BIPM philosophy in physical measurements, which at any rate are assuming a dominant position in contemporary analysis. In metrology we require correction for all known errors before reporting a measurement, while the uncertainty of the correction must be included in the uncertainty budget. This is often neglected in analytical measurement, where even corrections for blank value or recovery are not always made.[57]

This has been brought out very clearly by comparisons between traditional chemical analyses and purely instrumental methods, where the latter frequently yields higher results for the content of trace elements or other determinands, because of incomplete dissolution of the sample or loss of elements. At the lowest levels this has sometimes been compensated by the inadvertent contamination of the sample during chemical processing or inadequate blank correction.

Correction for incomplete recovery that can always be made in RNAA by measurement of the chemical yield of the carrier is much less straightforward in alternative methods, but nevertheless has to be done during the development stage of the analytical method by using radioactive or stable tracers.[58]

Correction for incomplete sample dissolution can be made by separate analysis of the solid residue by an instrumental method such as INAA.[59]

If such corrections are small their contribution to the uncertainty budget is also small and may be based on sound judgement as a Type B component to the combined uncertainty.

5.2 Certification Analysis

The certification of reference materials for elemental content is required for purposes of international traceability and the mutual recognition of laboratory accreditation needed in environmental and commercial activities. No effort is spared by the leading suppliers of CRMs to make sure that the certified concentrations are reliable, and each of them have developed procedures for performing certification analyses that minimize the risk of erroneous or misleading certification.[60]

The aim is to certify the *total amount* of an element present in the material, regardless of its chemical or physical form, and the certified value should be as close as possible to the true value, which means that corrections must be made for all known or assumed biases. In order to minimize the risk of overlooking unknown or unexpected biases, the use of several independent analytical methods for determining the same element is the basis for all serious certifications.

The absence of a reagent blank, the possibility of determining the recovery of added carrier, and the maintenance of statistical control make neutron activation analysis very attractive for the certification of low concentrations of the elements listed in Table 1. Together with its direct traceability to the pure element this makes neutron activation analysis a potentially definitive method.[7]

The combination of RNAA with INAA using different indicators and different comparators even opens up the possibility of self-validation[32] by fulfilling most of the requirements for completely independent methods of analysis. This introduces the possibility of certifying reference materials for elements for which no alternative reliable methods are available.

These properties of RNAA further emphasize the applicability of this method for certification purposes, and Table 10 presents examples of elements determined in a variety of reference materials, certified by BCR and covering the entire range of trace element concentrations from $0.01\,mg\,kg^{-1}$ to 0.01%. Particularly since the introduction of natural materials for certification, NIST and all other major suppliers have counted on neutron activation analysis not only for elements with

Table 10 Use of RNAA by BCR for the certification of reference materials in the European Union

Element	Materials	Examples of certification	
		BCR code	$mg\,kg^{-1}$
Antimony	Animal feed	CRM 281	0.047
	Metallic copper	CRM 074	0.58
Arsenic	Animal feed	CRM 274	0.132
	Animal tissue	CRM 278	5.9
	Metallic copper	CRM 074	0.78
	Coal	CRM 181	27.7
	Plants	CRM 279	3.09
	Sediment	CRM 320	76.7
Cadmium	Animal tissue	CRM 422	0.017
	Coal	CRM 040	0.11
	Milk powder	CRM 151	0.101
	Hair	CRM 397	0.521
	Sewage sludge	CRM 146	77.7
	Incineration ash	CRM 176	470

(*continued overleaf*)

Table 10 (*continued*)

Element	Materials	Examples of certification	
		BCR code	mg kg^{-1}
Calcium	Milk powder	CRM 063	12 600
Chlorine	Fresh water	CRM 399	50.5
Chromium	Animal feed	CRM 281	2.14
	Coal	CRM 040	31.3
	Leaves	CRM 100	8.0
	Sewage sludge	CRM 597	203
	Fly ash	CRM 038	192
Cobalt	Animal feed	CRM 402	0.178
	Coal	CRM 040	7.8
Copper	Animal feed	CRM 414	29.5
	Animal tissue	CRM 186	31.9
	Cereals	CRM 191	2.6
	Milk powder	CRM 150	2.23
	Plants	CRM 279	13.14
	Sewage sludge	CRM 144	713
Iodine	Animal feed	CRM 129	0.167
	Animal tissue	CRM 422	4.95
	Milk powder	CRM R63	0.81
Iron	Fresh water	CRM 398	0.029
	Milk powder	CRM 150	11.8
Manganese	Animal tissue	CRM 422	0.543
Mercury	Animal feed	CRM 482	0.48
	Animal tissue	CRM 185	0.044
	Coal	CRM 180	0.123
	Milk powder	CRM 151	0.101
	Sediment	CRM 277	1.77
	Soil	CRM 141R	0.25
	Sewage sludge	CRM 146R	8.62
	Fly ash	CRM 038	2.10
Molybdenum	Animal feed	CRM 281	0.84
Phosphorus	Metallic copper	CRM 017	6.9
	Milk powder	CRM R63	11 100
	Leaves	CRM 101	1690
Selenium	Animal feed	CRM 274	1.03
	Animal tissue	CRM 184	0.183
	Metallic copper	CRM 074	0.37
	Coal	CRM 182	0.68
	Cereals	CRM 189	0.132
	Plants	CRM 279	0.593
	Sediment	CRM 320	0.214
	Incineration ash	CRM 176	41.2
Thallium	Incineration ash	CRM 176	2.85
Zinc	Animal tissue	CRM 278	76
	Coal	CRM 040	30.2
	Plants	CRM 279	51.3
	Soil	CRM 143R	1063
	Sewage sludge	CRM 145R	2137

high sensitivity, but also for elements where alternative methods are not entirely satisfactory;[61] this situation is not likely to change in the foreseeable future.

5.3 Sampling Constants

The characterization of a natural CRM must include information about possible heterogeneity, either in the form of a minimum sample size or – preferably – as a statement of the maximum sampling constant.[62] All reference materials are carefully homogenized during their preparation, which means that they may be considered uniform and therefore can be characterized by the Ingamells' sampling constant[63] (Equation 14)

$$K_S = R^2 W \tag{14}$$

where R is the coefficient of variation observed by the analysis of replicate samples with weight W. With a known sampling constant the contribution of heterogeneity to the variability of results is accounted for by adding a term to the a priori precision from Equation (11) (Equation 15)

$$\hat{\sigma}_w^2 = \hat{\sigma}^2 + \hat{\mu}^2 \frac{K_S}{10^4 W} \tag{15}$$

which also shows that the sampling constant can be determined by analyzing a number of samples by an analytical method in statistical control.[47]

Present reference materials have sampling constants of the order of 1 g,[40] but modern methods of analysis use much smaller samples and therefore need materials with smaller sampling constants. Sampling constants need to be determined separately for each determinand,[64] and the use of methods that are in statistical control regardless of sample size is therefore in great demand. For many elements both RNAA and INAA can be used on samples in the milligram to microgram range without problems and therefore can be used to advantage in future verification of very small sampling constants just by activating the samples in very high neutron flux densities.

ABBREVIATIONS AND ACRONYMS

BCR	Bureau Communautaire de Réference
BIPM	Bureau International des Poids et Mesures
CRM	Certified Reference Material
DSP	Digital Signal Processing
FWHM	Full Width at Half-Maximum
GM	Geiger-Müller
IDMS	Isotope Dilution Mass Spectrometry
INAA	Instrumental Neutron Activation Analysis
ISO	International Organization for Standardization
NIST	National Institute of Standards and Technology
QC/QA	Quality Control/Quality Assurance
RNAA	Radiochemical Neutron Activation Analysis
SEM	Standard Error of the Mean

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Instrumental Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Radiotracer Methods

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Nuclear Detection Methods and Instrumentation

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. G. Hevesy, H. Levi, 'The Action of Neutrons on the Rare Earth Elements', *K. Dan. Vidensk. Selsk., Mat.-Fys. Medd.*, **XIV**(5), 3–34 (1936).
2. G.T. Seaborg, J.J. Livingood, 'Artificial Radioactivity as a Test for Minute Traces of Elements', *J. Am. Chem. Soc.*, **60**, 1784–1786 (1938).
3. G.E. Boyd, 'Method of Activation Analysis', *Anal. Chem.*, **21**, 335–347 (1949).
4. 'Definitions and Terms', *J. Radioanal. Nucl. Chem.*, **113**, 539–541 (1987).
5. W.D. Ehmann, D.E. Vance, *Radiochemistry and Nuclear Methods of Analysis*, J. Wiley & Sons, New York, 253–308, 1991.
6. K. Heydorn, 'INAA – Application and Limitation', *Fresenius' J. Anal. Chem.*, **337**, 498–502 (1990).
7. R. Dybczynski, B. Danko, 'A Definitive Method for the Determination of Trace Amounts of Cobalt in Biological Materials by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **182**, 43–59 (1994).
8. K. Heydorn, 'Neutron Activation Analysis' in *Encyclopedia of Medical Devices and Instrumentation*, ed. J.G. Webster, J. Wiley & Sons, New York, Vol. 3, 2041–2046, 1988.
9. K. Heydorn, *Neutron Activation Analysis for Clinical Trace Element Research*, CRC-Press, Boca Raton, FL, (a) 75–83, Vol. 1, (b) 101–103, Vol. 1, (c) 120–122, Vol. 1, (d) 129–131, Vol. 1, (e) 145–149, Vol. 1, (f) 42–45, Vol. 2, 1984.
10. K. Heydorn, E. Damsgaard, B. Rietz, 'Systematic Differences in the Determination of Vanadium in Standard Reference Material 1571 Orchard Leaves', *Anal. Chem.*, **52**, 1045–1049 (1980).
11. J.R. Moody, R.R. Greenberg, K.W. Pratt, T.C. Rains, 'Recommended Inorganic Chemicals for Calibration', *Anal. Chem.*, **60**, 1203A–1218A (1988).
12. P. Bode, R.M.W. Overwater, J.J.M. De Goeij, 'Large-sample Neutron Activation Analysis: Present Status and Prospects', *J. Radioanal. Nucl. Chem.*, **216**, 5–11 (1997).
13. K. Heydorn, *Handbook on Metals in Clinical and Analytical Chemistry*, Marcel Dekker, Basle, 181–194, 1994.
14. *Nuclear Research Reactors in the World*, International Atomic Energy Agency, Vienna, 1998.
15. D.E. Ryan, A. Chatt, J. Holzbecher, 'Analysis for Trace Elements with a SLOWPOKE Reactor', *Anal. Chim. Acta*, **200**, 89–100 (1987).
16. K. Heydorn, E. Damsgaard, 'Gains or Losses of Ultratrace Elements in Polyethylene Containers', *Talanta*, **29**, 1019–1024 (1982).
17. Lin Xilei, K. Heydorn, B. Rietz, 'Limit of Detection for the Determination of Pt in Biological Material by RNAA using Electrolytic Separation of Gold', *J. Radioanal. Nucl. Chem.*, **160**, 85–99 (1992).
18. R.R. Greenberg, H.M. Kingston, R.L. Watters, K.W. Pratt, 'Dissolution Problems with Botanical Reference Materials', *Fresenius' J. Anal. Chem.*, **338**, 394–398 (1990).
19. A.R. Byrne, J. Versieck, 'Vanadium Determination at the Ultratrace Level in Biological Reference Materials and Serum by Radiochemical Neutron Activation Analysis', *Biol. Trace Elem. Res.*, **27**, 529–540 (1990).
20. A.J. Blotcky, J.P. Claasen, F.R. Roman, E.P. Rack, S. Badakhsh, 'Determination of Aluminum by Chemical and Instrumental NAA in Biological Standard Reference Materials and Human Brain Tissue', *Anal. Chem.*, **64**, 2910–2913 (1992).
21. A. Mizuike, 'Separations and Preconcentrations', in *Trace Analysis, Physical Methods*, ed. G.H. Morrison, Interscience, New York, 103–159, 1965.
22. D. van Renterghem, R. Cornelis, R. Vanholder, 'Radiochemical Determination of Twelve Trace Elements in Human Blood Serum', *Anal. Chim. Acta*, **257**, 1–5 (1992).
23. R.R. Greenberg, R. Zeisler, 'A Radiochemical Procedure for Ultratrace Determination of Chromium in Biological Materials', *J. Radioanal. Nucl. Chem.*, **124**, 5–20 (1988).
24. J. Kucera, M. Vobecky, L. Soukal, D. Zákoucky, D. Vénos, 'Low Level Determination of Thallium in Biological and Environmental Reference Materials by RNAA Using Several Counting Methods', *J. Radioanal. Nucl. Chem.*, **217**, 131–137 (1997).
25. F. Girardi, E. Sabbioni, 'Selective Removal of Radiosodium from Neutron-activated Materials by Retention on Hydrated Antimony Pentoxide', *J. Radioanal. Chem.*, **1**, 169–178 (1968).
26. Z.B. Alfassi, C.M. Wai (eds.), *Preconcentration Techniques for Trace Elements*, CRC Press, Boca Raton, FL, 1992.
27. E. Damsgaard, K. Heydorn, 'Synthesis of Precision for the Certification of Phosphorus in Biological Materials by

Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **113**, 267–275 (1987).
28. K. Heydorn, 'Validation of Neutron Activation Analysis Techniques', in *Quality Assurance for Environmental Analysis*, ed. Ph. Quevauviller et al., Elsevier, Amsterdam, 89–110, 1995.
29. K. Heydorn, W. Lada, 'Peak Boundary Selection in Photopeak Integration by the Method of Covell', *Anal. Chem.*, **44**, 2313–2317 (1972).
30. K. Heydorn, E. Damsgaard, 'Simultaneous Determination of Arsenic, Manganese, and Selenium in Biological Materials by Neutron Activation Analysis', *Talanta*, **20**, 1–11 (1973).
31. B. Danko, R. Dybczynski, 'Determination of Molybdenum and Uranium in Biological Materials by Radiochemical Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **216**, 51–57 (1997).
32. A.R. Byrne, 'Some Improvements in the Quality of NAA Procedures and the Reliability of the Results', *Biol. Trace Elem. Res.*, **44**, 529–537 (1994).
33. R.R. Greenberg, R. Zeisler, H.M. Kingston, T.M. Sullivan, 'Neutron Activation Analysis of the NIST Bovine Serum Standard Reference Material using Chemical Separations', *Fresenius' J. Anal. Chem.*, **332**, 652–656 (1988).
34. K. Heydorn, 'Use of the National Physical Laboratory Ionization Chamber Type 1383A in Neutron Activation Analysis', *J. Radioanal. Chem.*, **10**, 245–256 (1972).
35. K. Masumoto, 'Isotope Dilution Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, John Wiley & Sons, New York, 511–523, 1994.
36. A.R. Byrne, M. Dermelj, 'An Endogeneous Standard, Radioisotopic Ratio Method in NAA', *J. Radioanal. Nucl. Chem.*, **223**, 55–60 (1997).
37. A. Lagerwaard, J.R.W. Woittiez, J.J.M. de Goeij, 'An Independent Accurate Reference Method for the Determination of Chromium in Biological Materials', *Fresenius' J. Anal. Chem.*, **351**, 786–789 (1995).
38. B. Rietz, K. Heydorn, 'Determination of Gold and Platinum in Biological Materials by Radiochemical Neutron Activation Analysis using Electrolytic Separation of Gold', *J. Radioanal. Nucl. Chem.*, **174**, 49–56 (1993).
39. *Safe Handling of Radionuclides*, Safety Series No. 1, International Atomic Energy Agency, Vienna, 1973.
40. K. Heydorn, E. Damsgaard, 'The Determination of Sampling Constants by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **110**, 539–553 (1987).
41. S. Landsberger, S. Peshev, 'Compton Suppression Neutron Activation Analysis: Past, Present and Future', *J. Radioanal. Nucl. Chem.*, **202**, 201–224 (1996).
42. P. Giacomo, 'News from Bureau International de Poids et Mesures', *Metrologia*, **17**, 73 (1981).
43. *Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1993.
44. Eurachem Working Group, *Quantifying Uncertainty in Analytical Measurement*, BSI, London, 1995.
45. K. Heydorn, 'Detection of Systematic Errors by the Analysis of Precision', *Accuracy in Trace Analysis*, National Bureau of Standards Publication 422, Washington, DC, 127–139, 1976.
46. K. Heydorn, K. Nørgaard, 'Analysis of Precision of Activation Analysis Methods', *Talanta*, **20**, 835–842 (1973).
47. K. Heydorn, 'Quality Assurance and Statistical Control', *Microchim. Acta*, **III**, 1–10, (1991).
48. K. Heydorn, 'Detecting Errors in Micro and Trace Analysis by Using Statistics', *Anal. Chim. Acta*, **283**, 494–499 (1993).
49. *Certification of Reference Materials – General and Statistical Principles*, ISO Guide 35, ISO, Geneva, 1989.
50. K. Heydorn, 'Determination of the Specific Activity of Carrier-free ^{125}I Preparations by Neutron Activation Analysis', *J. Radioanal. Chem.*, **3**, 225–232 (1969).
51. Susan Parry et al., 'The Determination of ^{129}I in Milk and Vegetation Using Neutron Activation Analysis', *Sci. Total Environ.*, **174**, 351–360 (1995).
52. R.R. Greenberg, 'Accuracy in Standards Preparation for Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **179**, 131–139 (1994).
53. F. de Corte, 'The Traceability of the Neutron Activation Analysis k_0-Standardization Method', *J. Trace Microprobe Tech.*, **5**, 115–134 (1987).
54. K. Heydorn, 'Letter to the Editor', *Trends Anal. Chem.*, **11**, II (1992).
55. 'Accuracy (trueness and precision) of Measurement Methods and Result – Part 1: General Principles and Definitions', ISO Standard 5725-1, International Organization for Standardization, Geneva, 1994.
56. K. Heydorn, 'Comments on Definitions of Analytical Terms', *Accred. Qual. Assur.*, **2**, 204–205 (1997).
57. P. Willetts, R. Wood, 'Measurement Uncertainty – A Reliable Concept in Food Analysis and for the Use of Recovery Data?', *Accred. Qual. Assur.*, **3**, 231–236 (1998).
58. IUPAC, 'Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement', 1995 IUPAC Technical Report, 1998.
59. K. Heydorn, E. Damsgaard, 'Indirect Determination of Soluble Trace Elements in Certified Reference Materials by Neutron Activation Analysis', *Mikrochim. Acta*, **119**, 297–304 (1995).
60. K. Heydorn, 'Procedures for Certification of BCR Environmental Reference Materials', in *Harmonization of Health Related Environmental Measurements using Nuclear and Isotopic Techniques*, IAEA, Vienna, 157–173, 1997.
61. R.R. Greenberg, 'The Role of Neutron Activation Analysis in the Development and Certification of NIST

Environmental Standard Reference Materials', in *Harmonization of Health Related Environmental Measurements using Nuclear and Isotopic Techniques*, IAEA, Vienna, 137–144, 1997.

62. 'Contents of Certificates of Reference Materials', *ISO Guide 31*, International Organization for Standardization, Geneva, 1981.
63. C.O. Ingamells, P. Switzer, 'A Proposed Sampling Constant for Use in Geochemical Analysis', *Talanta*, **20**, 547–568 (1973).
64. A. Chatt, R.R. Rao, C.K. Jayawickreme, L.S. McDowell, 'Determination of Sampling Constants for Selenium in Biological Reference Materials by Neutron Activation', *Fresenius' J. Anal. Chem.*, **338**, 399–407 (1990).

Radiochemical Separation Schemes for Multielement Determination

Zeev B. Alfassi
Ben Gurion University of the Negev, Beer Sheva, Israel

1 Introduction 1
2 Separation in Material Sciences 1
3 Geological and Environmental Samples 5
4 Biological Samples 12
Abbreviations and Acronyms 14
Related Articles 14
References 14

Radiochemical activation analysis is a more laborious analytical method than instrumental analysis, but can be used to determine elements at lower concentrations. In most cases, the sample to be analyzed is separated into groups of elements rather than individual elements, so that the elements do not interfere with one another during the measurement of radionuclides content. Schemes for separation into multielement groups are given for metals, biological, geological and environmental samples.

1 INTRODUCTION

While instrumental neutron activation analysis (INAA) is more popular than RNAA (radiochemical neutron activation analysis), owing to its simplicity, RNAA can determine much lower concentrations than INAA, because the interferences from other radionuclides are reduced considerably. Whereas INAA involves only activation and activity measurement, RNAA also includes dissolution of the sample and chemical separation of the sample solution. In some cases, the interferences can be removed by decay (''cooling'') and INAA can be used. However, in many cases the IRNs (indicator radionuclides) are shorter lived than those of the interferences, and in other cases the required decay time is too long to be practical. Owing to the chemical dissolution, separation and recovery yield determination, RNAA involves more person hours and should be avoided when it is possible. Heydorn, in his article (**Radiochemical Neutron Activation Analysis**), shows why RNAA is so important in verification of the concentration of trace elements in reference materials. However, RNAA is not just used for validation of reference materials, although this is clearly one of its main uses. RNAA involves irradiation and counting, dealt with in the article **Instrumental Neutron Activation Analysis**, as well as dissolution of the samples and separation of the dissolved samples into various groups. The dissolution step is much simpler than for other methods of analysis, because the final measurement is the determination of the radionuclides formed by the irradiation. As the dissolution step is done after the irradiation is carried out, no contamination during the dissolution step can influence the final results. In addition, as the dissolution of the sample is done together with that of carriers, the danger of losing trace elements by adsorption is considerably lower. On the other hand, owing to the radioactivity of the irradiated sample, the dissolution, if performed many times, should be done behind lead shields. In principle every chemical separation that is used in any other analytical method can be used in RNAA. But since the separation very rarely involves a single element, an appropriate group of elements which do not interfere with each other is different for various determination methods. This is the reason that it was decided that it was worth collecting together in this article the main methods of chemical separation into groups that are used in RNAA. In some cases, even the separation into groups is not needed, and it is sufficient to remove the major interferences. Egger and Krivan[1] in their determination of trace elements in aluminum, removed only the interference of ^{24}Na (by sorption on a hydrated antimony pentoxide, HAP, column) and were then able to measure 38 trace elements, all in concentrations of less than 1 ppb. For most other analytical methods the bulk of Al might need to be removed.

An additional aspect which is different for RNAA with respect to all other determination methods, is that in RNAA the final determination can be performed while the trace elements are adsorbed onto a separation column, whereas all other methods require the elution of the trace elements from the column.

2 SEPARATION IN MATERIAL SCIENCES

Park et al.[2] studied the concentration of 13 trace elements in high-purity molybdenum by separating them into three groups, separated from the major interferences which are Mo and W. A cation exchange column (Dowex™ 50) in dilute HCl medium is used to remove Mo and W, which are only weakly adsorbed on this column, while many of the impurity elements are strongly absorbed. Na and K are eluted by using additional dilute

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

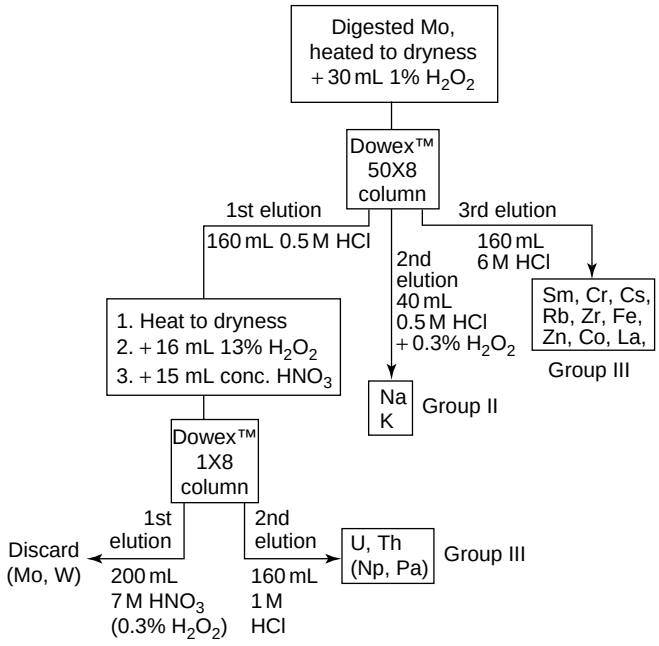

Figure 1 Park et al.[2] method for separation of 13 elements into three groups for determination of trace elements in molybdenum. (Reproduced from Park et al.[2] with permission.)

HCl. Nine trace elements were eluted with 6-M HCl. Together with Mo and W, ^{239}Np and ^{233}Pa, which are the IRNs for U and Th, due to β^- decay following the (n,γ) reaction, were also eluted, respectively, Equation (1)

$$^{238}\mathrm{U(n,\gamma)}^{239}\mathrm{U} \xrightarrow{\beta^-} {}^{239}\mathrm{Np};$$
$$^{232}\mathrm{Th(n,\gamma)}^{233}\mathrm{Th} \xrightarrow{\beta^-} {}^{233}\mathrm{Pa} \qquad (1)$$

Np and Pa were separated from Mo and W using anion exchange (Dowex™-1) in strong nitric acid media. Pa and Np are strongly absorbed on the anion exchange resin, while molybdenum and tungsten are not. Figure 1 gives a schematic representation of the process of separation of Park et al.[2] Before counting, each eluate was concentrated by partial evaporation.

Theimer and Krivan[3] developed a group separation for the determination of 20 elements in high-purity molybdenum (Figure 2). They developed two modifications. The first one involves removal of Mo and its daughter Tc and the main contamination of W, by absorption on anion exchange resin (Dowex™ 1X8). Five other elements are also absorbed on the column, but 20 elements are eluted and can be determined simultaneously in the eluate. The second modification allows determination of only 11 elements, although separation is effected into two groups. The advantage of this special modification is the separation of ^{233}Pa (the IRN of Th) with only Sc, enabling determination of small amounts of Th. Both modifications

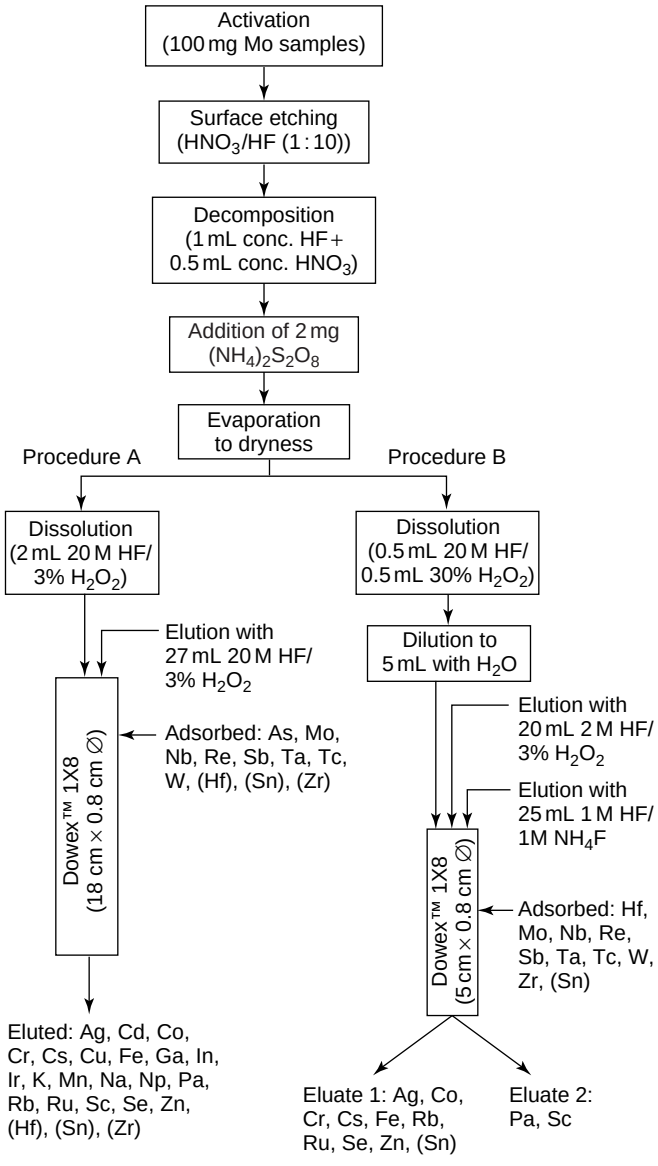

Figure 2 Flow chart of the two procedures for radiochemical separation of trace elements in molybdenum. (Reproduced from Theimer and Krivan,[3] with permission.)

are based on elution with various concentrations of HF, similar to all the procedures developed by Krivan et al. for determination of trace elements in metals. The flow chart for the two radiochemical separation procedures is given in Figure 2. The two methods are based on Mo, Tc, and W being absorbed on an anion exchange column in HF media, while the trace elements are eluted with HF. For selective separation of Pa (together with Sc only), lower concentrations of HF are used. The elements given in parentheses are eluted but not completely, so that they are found both on the column and in the eluate. The elements determined via medium-lived IRNs and which had

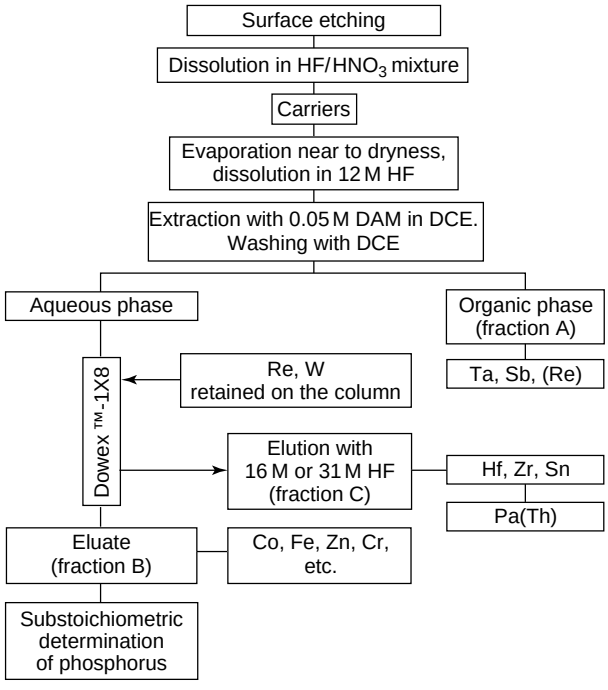

Figure 3 Flow chart for the radiochemical separation of trace elements in tungsten. (Reproduced from Caletka et al.[4] with permission.)

to be in eluate 1 according to their distribution ratios, as for example Cu, Ga, K, Mn and Na, do not appear in procedure B, because about 1% of the technetium is eluted with this fraction, masking the activity of these IRNs. However, if $(NH_4)_2S_2O_8$ is added to the sample solution prior to its evaporation, in order to oxidize all technetium species to the heptavalent state, less than 0.01% of the technetium is found in fraction 1, and the other elements can also be determined in this fraction.

A similar technique was developed by Krivan's group to determine the content of 19 elements in high-purity tungsten.[4] However, a modification was done in order to measure the concentrations of Ta and Sb, which in the previous processes were adsorbed on the anion exchange resins together with Mo, Tc and W, and could not be determined. Ta and Sb are extracted with organic solvent (dichloroethane, DCE) after forming complexes with diantipyrilmethane (DAM), prior to the introduction of the solution of the dissolved irradiated tungsten onto the anion exchange column. In order to measure the P content via ^{32}P, a β^--only emitter, ^{32}P is substoichiometrically extracted as molybdophosphate with tetraphenylarsonium chloride in dichloromethane.[5] The flow chart for this separation is given in Figure 3.

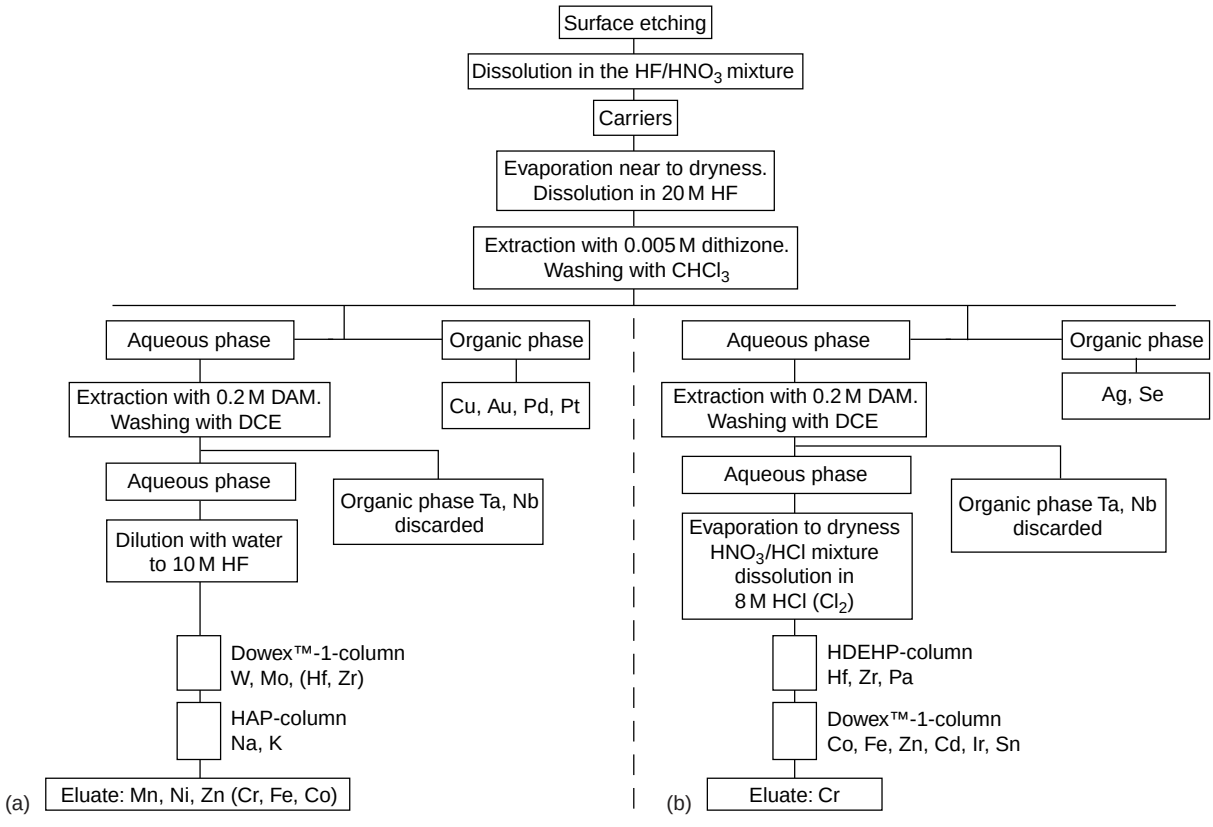

Figure 4 Two procedures for RNAA of trace elements in niobium (a) for medium-lived IRNs; (b) for long-lived IRNs. HDEHP, di(2-ethylhexyl)-orthophosphoric acid. (Based on figures from Caletka et al.[6] with permission.)

Caletka et al.[6] developed a RNAA procedure for determination of 26 elements in niobium by irradiation and processing of two samples, one for measurement of medium-lived IRNs (the radionuclide used for the determination of an element in NAA), by irradiation for 12 h. The second sample was irradiated for 5 days in order to determine the long-lived IRNs. In the determination of medium-lived IRNs, they wanted to measure ^{56}Mn ($t_{1/2} = 2.58$ h) and ^{65}Ni ($t_{1/2} = 2.52$ h) as well. This is possible because the main activity produced from the matrix (^{94m}Nb) is short lived ($t_{1/2} = 6.24$ min) and one hour of cooling leaves a sample that is not too radioactive to handle. For matrixes where the main activity is longer lived, IRNs with half-lives of a couple of hours can be determined only by pre-irradiation separation of the matrix, or treatment in special hot laboratories, where very high radioactivity can be handled. Different schemes of separation were devised for the medium- and long-lived IRNs, however, they have many common features. Both use solvent extractions followed by separation on columns. The same two solvent extractions were used in the two

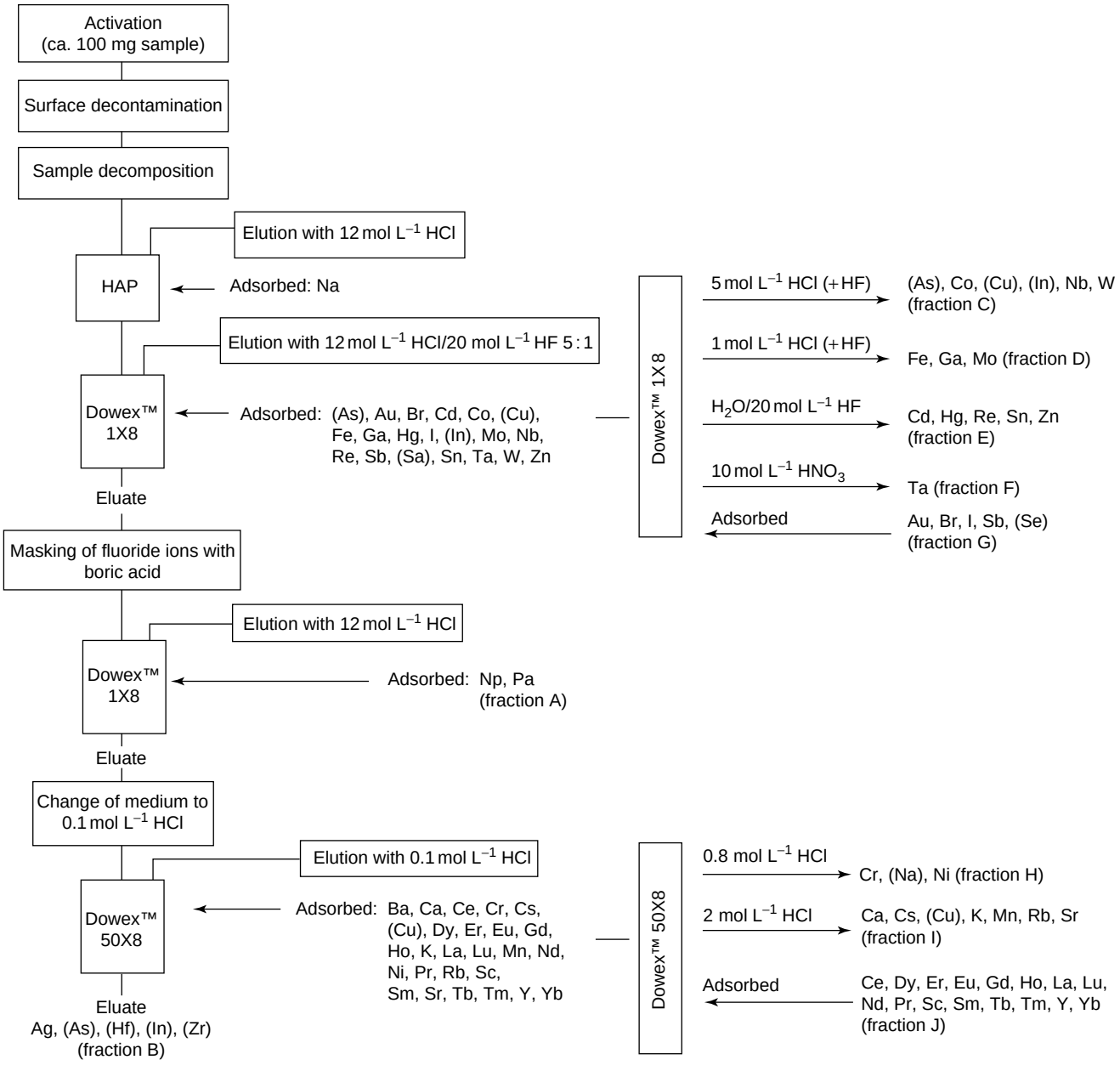

Figure 5 Procedure for separation of 52 elements into 10 fractions in irradiated ultrapure aluminum. (Reproduced from Egger and Krivan[7] with permission.)

procedures, and the only difference is in the columns and acids used for separation of the last aqueous phase into three groups. The first extraction step is for cations which form chelates with dithizone and are extracted with chloroform. This fraction includes Cu, Au, Pd and Pt in the medium-lived IRNs, and Ag, Se in the long-lived IRN samples. The second extraction of metals is for those which form complexes with DAM. The organic solvent is DCE. This extraction is used to remove tantalum and niobium, and the organic phase containing them is discarded. The elements remaining in the aqueous phase are separated into three groups using two columns in series. One group is adsorbed on each column, and the third group contains the elements which were eluted from the two columns. For the long-lived samples, the first column is the reversed phase di-(2-ethylhexyl) orthophosphoric acid on a solid support and then a column of anion exchange resin-Dowex™-1. For the medium-lived IRNs, the anion exchange-Dowex™-1 is the first column, followed by a HAP column. In high-concentration HCl solution, HAP retains only the Na ions; however, under the conditions used here, K also remained on the HAP column. The flow chart of the separation is given in Figure 4.

A thorough determination of the trace elements in ultrapure metal was performed for aluminum.[7] Previously, we mentioned that 38 elements were determined in levels of parts per billion, by selective removal of ^{24}Na with a HAP column. Later, the same group extended this study, and in order to improve the limits of detection, an additional separation into 10 groups was done.[7] Forty-three elements have limits of detection below 10 ppb, and for U and Th the limits of detection are 50 ppt (5×10^{-11}). The separation scheme is given in Figure 5. First, Na^+ is retained on HAP. With 12-M HCl elution, this is the only cation to be retained. The eluate from the HAP column is loaded onto an anion exchange and washed with 12-M HCl–20-M HF, 5 : 1. Twenty elements from the eluate are adsorbed on the anion exchanger resin (Dowex™ 1X8), and later divided into five groups, by elution of four different groups with different eluants. The fifth group is the one that is retained on the column after the fourth elution. The original eluate from the Dowex™ 1X8 column is masked by addition of boric acid and the solution is introduced into another Dowex™ 1X8 anion-exchanger column. The masking of the F^- anions by boric acid leads to retention of Pa and Np (IRNs of Th and U) on the column. Further elements are separated on a cation exchanger. The solution is evaporated to dryness, dissolved in 0.1-M HCl and introduced into a cation exchanger Dowex™ 50X8 and washed with 0.1-M HCl. The eluate is one group of five elements, while the 27 elements remaining on the column are separated into three groups by two elutions. The flow chart for the separation (Figure 5) shows the separation of 52 elements as studied by radiotracers. Not all of them were found in the ultrapure Al. Only those not in brackets were actually found in the irradiated Al.

3 GEOLOGICAL AND ENVIRONMENTAL SAMPLES

Many studies have been done on the determination of precious metals in geological samples. Nadkarni and Morrison[8,9] digested the irradiated sample

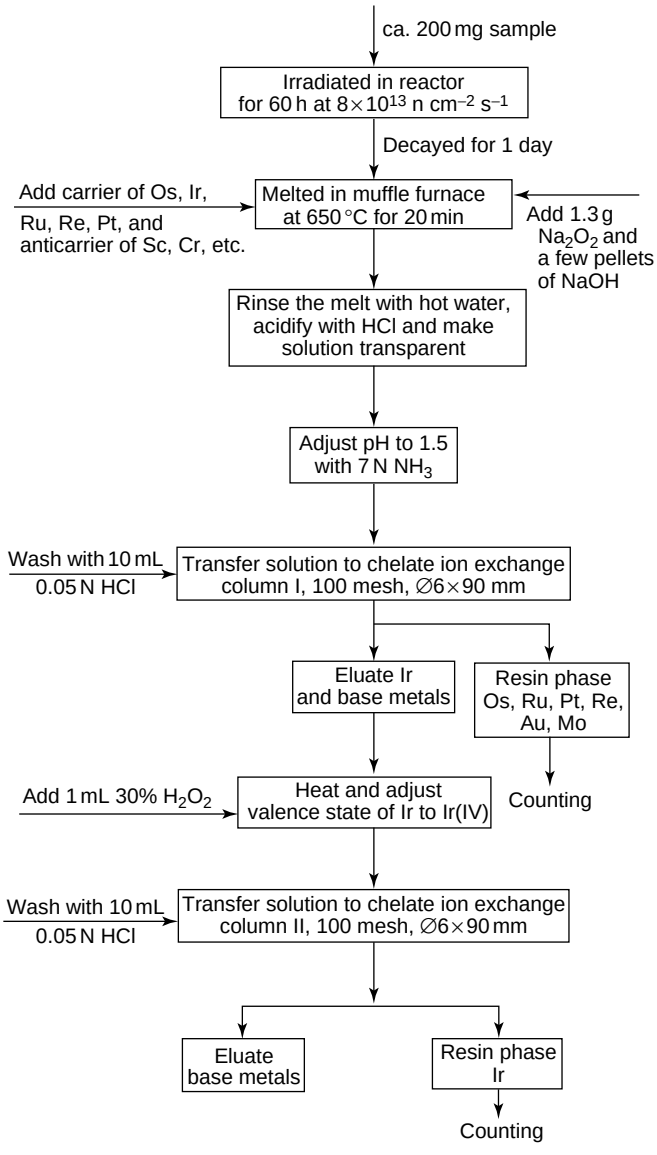

Figure 6 Radiochemical separation procedure of noble metals based on chelate ion-exchange resin. (Reproduced from Chai et al.[14] with permission.)

(200–500 mg) by peroxide fusion, and separated the noble metals on a Srafion NMRR selective chelating ion-exchange resin. However, yields are not well reproducible and reirradiation of each sample is necessary to determine them.[10] Only Pd, Pt, Ir and Au can be determined by this process. The use of the ion exchanger Srafion NMRR was questioned by Stockman,[11] who could not repeat the results of Nadkarni and Morrison and attributed this to changes in the manufacture of the resin. There are also some previous data which give contrasting results for the specificity of this resin for d^8 ions.[12,13]

Chai et al.[14] developed two different methods for radiochemical separation of the noble metals. They found that Nadkarni and Morrison were right and the Srafion NMRR resin can be used to retain Os and Ir. Ten milliliters of 0.05-M HCl eluted 100% of Sc, Cs, and Fe, whereas only 10% of Os was eluted. Nadkarni and Morrison reported that thiourea can be used to elute the noble metals, whereas Stockman argued that thiourea elutes only gold, while Ir remains on the resin. Chai et al.[14] found that Ir and Re can be eluted with NH_3, whereas only one-quarter of Os was eluted and no Au or Pt. Chai et al.[14] found that all platinum group elements (PGE), including Au, are retained by various productions of this resin. Chai et al.[14] suggested a scheme for radiochemical separation of noble metals, which is given in Figure 6. They found that their digestion process forms mainly Ir(III), which is not absorbed onto the chelate resin. However, oxidation of the chelated Ir with hot H_2O_2 transforms all Ir to the +4 oxidation state, which is absorbed on another column with the same resin.

An alternative process for the separation of the noble metals, suggested by Chai et al.,[14] is based on extraction of noble metals with long-chain primary amines (19–23 carbon atoms) in DCE. All the noble metals are extracted in high yields into the organic phase. The base metals were only negligibly extracted.

Another group besides the noble metals, which is determined many times by RNAA, is the rare earth elements (REE) group. Zilliacus et al.[15] described two procedures for the separation of the REE group from irradiated geological samples. A simple and faster method was developed for samples with concentrations above 0.5 ppm, while a method with more separation steps was developed for samples with lower concentrations. A more time-consuming process is required to obtain a clean REE fraction. The two processes are described in Figure 7. Both methods are based on fusion with

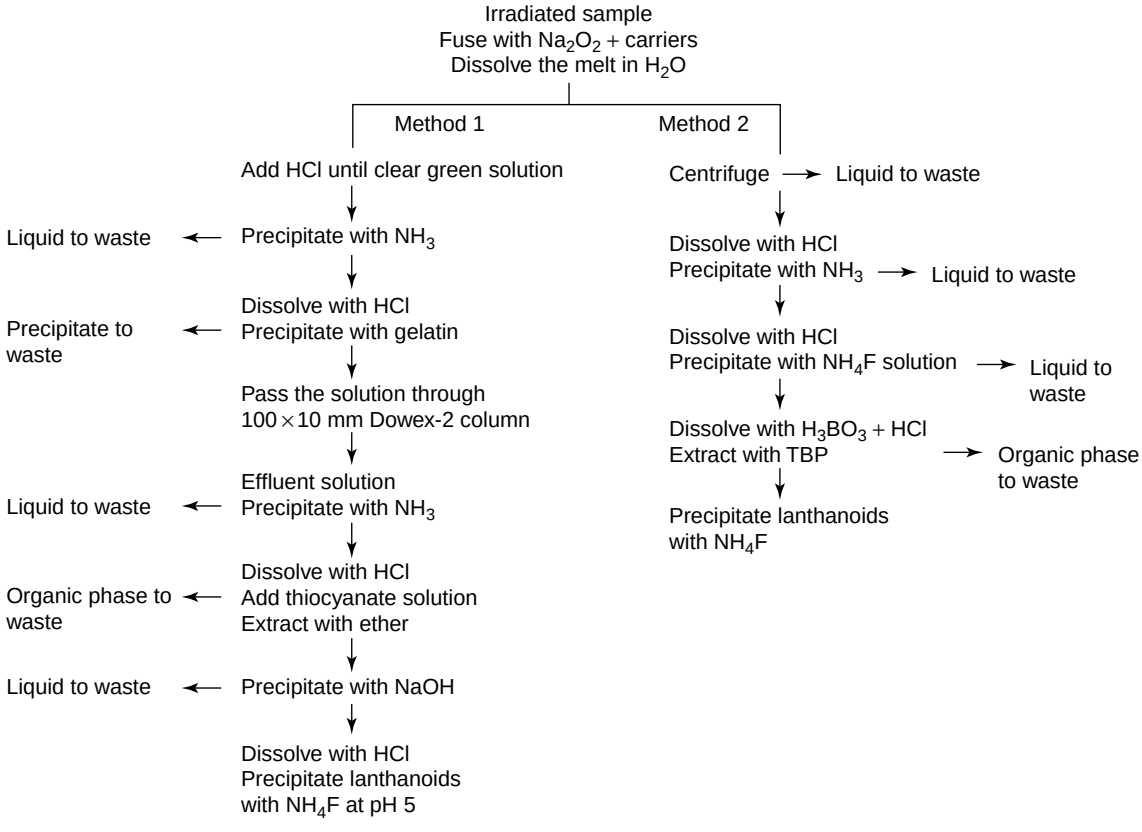

Figure 7 Separation scheme for the analysis of REE in geological samples (Reproduced from Zilliacus et al.[15] with permission.)

Na_2O_2 and cycles of precipitation as hydroxides and as fluorides. In the simple method there is one more step for removal of Sc and other metals by extraction with tributylphosphate (TBP). For the lower concentration samples, SiO_2 is removed by precipitation with gelatin in acidic solution, the main interferences are adsorbed on an anion exchange column in a HCl medium, while the REE are eluted. Sc is removed by extraction with ether from SCN^- solution. Purification of the REE fraction is done by a cycle of precipitation of hydroxides and fluorides. A similar process was also developed by Laul et al.[16] They did not use the Sc removal step by extraction; however, in the final step, they performed three cycles of hydroxides–fluoride precipitation.

A similar method was developed by Wandless and Morgan.[17] The method is based on the precipitation of the REE hydroxides both at high pH and at pH 9 in ammoniacal solution. SiO_2 is removed by precipitation with gelatin in acidic solution. Fe is removed by an anion exchanger in the Cl^- form from 8-M HCl solution. Zr, Sc and Hf are removed by an anion exchanger in the SCN^- form from 0.8-M SCN^- + 0.5-M Cl^- solution. The only remaining interference is Cr^{3+}, which is removed by precipitation of the REE with excess 8-M NaOH. The complete scheme is given in Figure 8.

Parthasarathy et al.[18] separated the REE into two groups rather than into one group, as in all other processes. They separated a group of light REE and a group of heavy REE. The process is different from all those previously described, and the flow chart is given in Figure 9.

Morrison et al.[19] described a procedure for chemical group separations, which, with the use of a Compton suppression Ge(Li) detector, can determine 45 elements in geological samples. The flow chart for this process is given in Figure 10. The first group is the one including volatile elements and volatile fluorides. Elements adsorbed to HAP in 8-M HCl formed the second group. Those adsorbed to anion exchange resin in 8-M HCl formed the third and fourth groups. The separation between these two groups is done by elution with 0.5-M HCl. The elements not adsorbed on the anion exchange in the 8-M HCl are separated into two groups by extraction with TBP. Laul et al.[20] developed a RNAA method for trace elements in terrestrial rocks and stony meteorites. They stated that their main emphasis was to minimize chemical procedure and maximize the γ-ray spectrometry aspect. However, their separation is longer than that of Morrison et al.,[19] and they measure only 16 elements. The flow chart of their separation is given in Figure 11. The

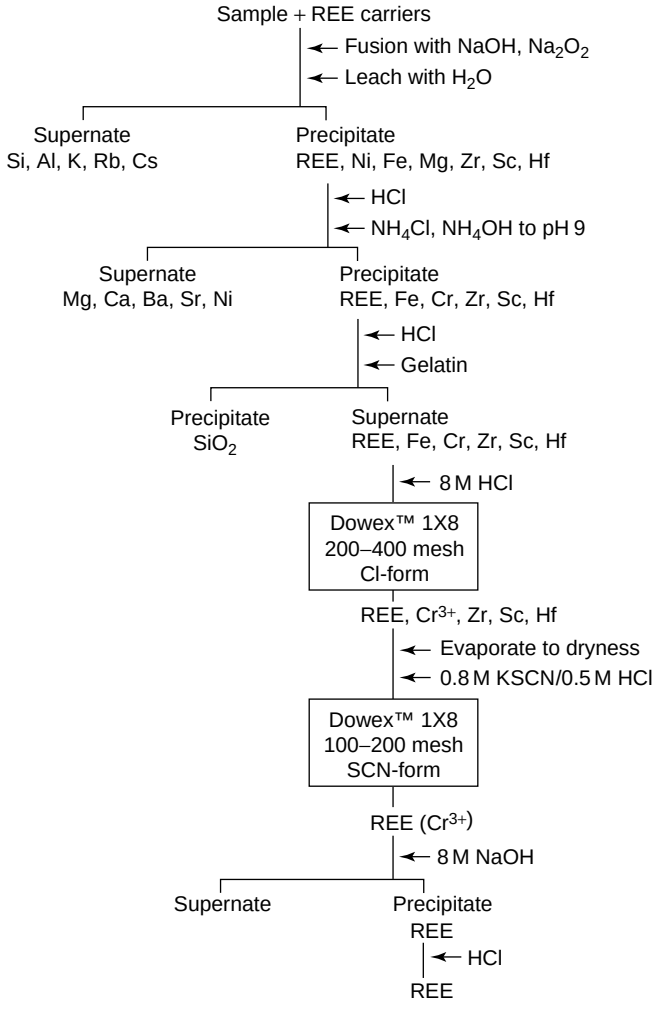

Figure 8 Separation procedure flow chart for RNAA of REE in geological samples. (Reproduced from Wandless and Morgan[17] with permission.)

scheme involves four different separation techniques: distillation, solvent extraction, ion-exchange columns, and precipitation.

Smets et al.[21] developed a group separation for RNAA determination of 24 elements in a wide variety of silicate rocks and minerals. The samples were decomposed in a $HF-HNO_3$ mixture in a PTFE (polytetrafluoroethylene) lined bomb, and were separated into soluble and insoluble fluorides. The soluble fluorides were separated into three groups by sequential elution (0.1-M HF, 3-M HCl + acetone, 12-M HCl) from a cation exchange column. The insoluble fluorides (Ca, Sr, Ba, REE and part of Rb and Cs) were dissolved and purified from iron and scandium activities by extraction with TBP. The separation scheme is given in Figure 12. The elements written in large type were determined, whereas those elements in small

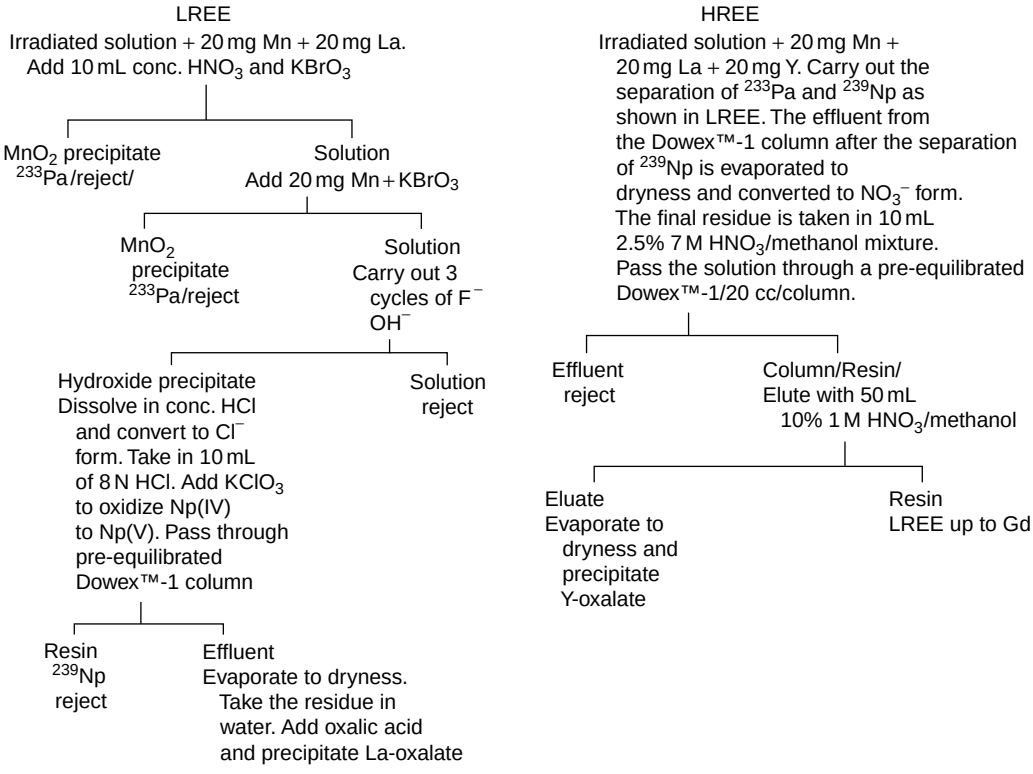

Figure 9 Flow chart for RNAA of REE by separation of two fractions, light and heavy REE (LREE and HREE). (Reproduced from Parthasarathy et al.[18] with permission.)

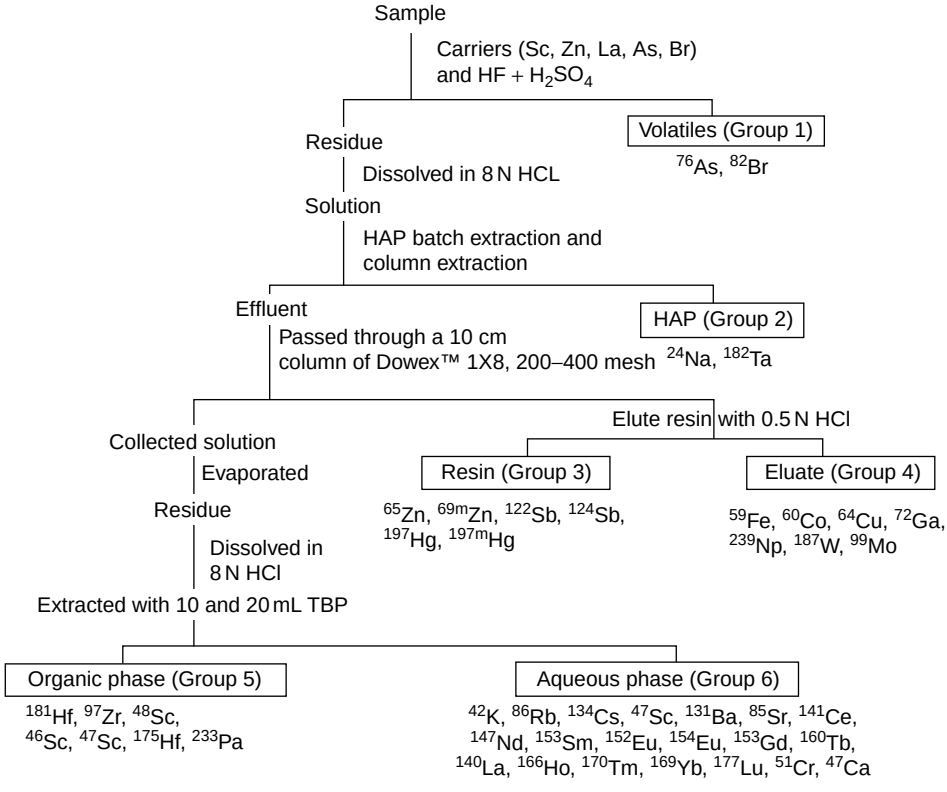

Figure 10 Flow chart for chemical separation of 45 trace elements in geological samples into six groups appropriate for simultaneous counting. (Reproduced from Morrison et al.[19] with permission.)

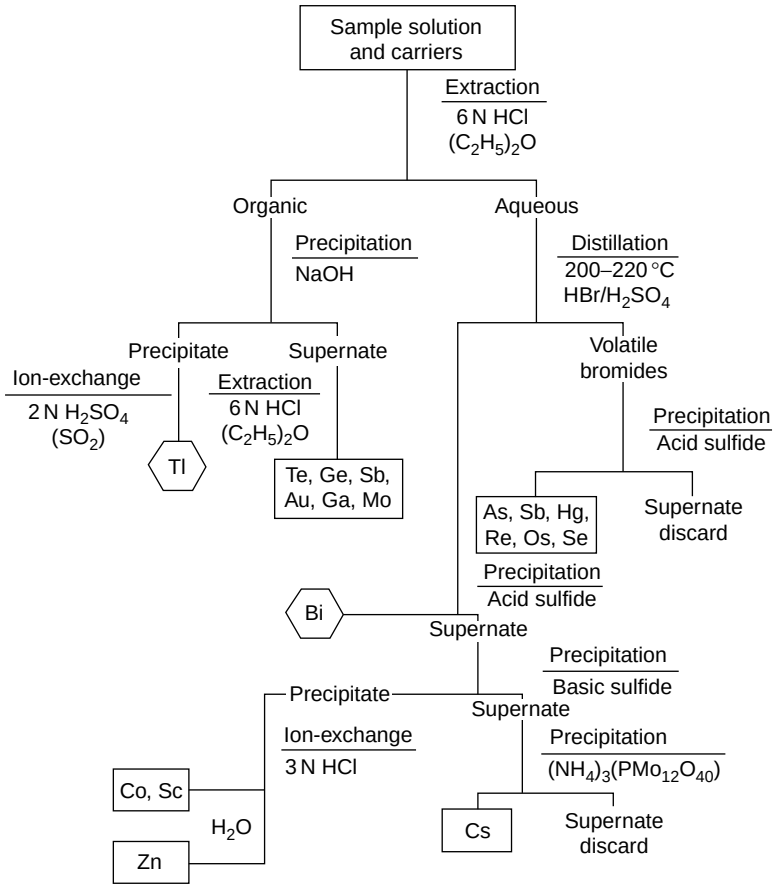

Figure 11 Chemical separation of trace elements in geological samples, leading to determination of 16 elements. (Reproduced from Laul et al.[20] with permission.)

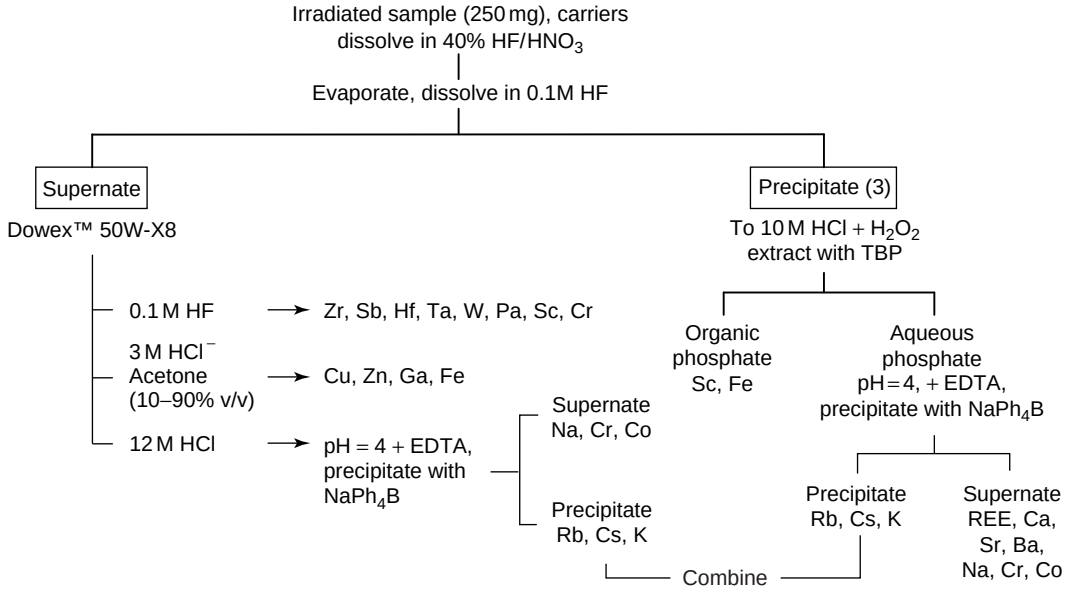

Figure 12 Chemical separation procedure for groups of trace elements in geological samples. (Reproduced from Pietra et al.[22] with permission.)

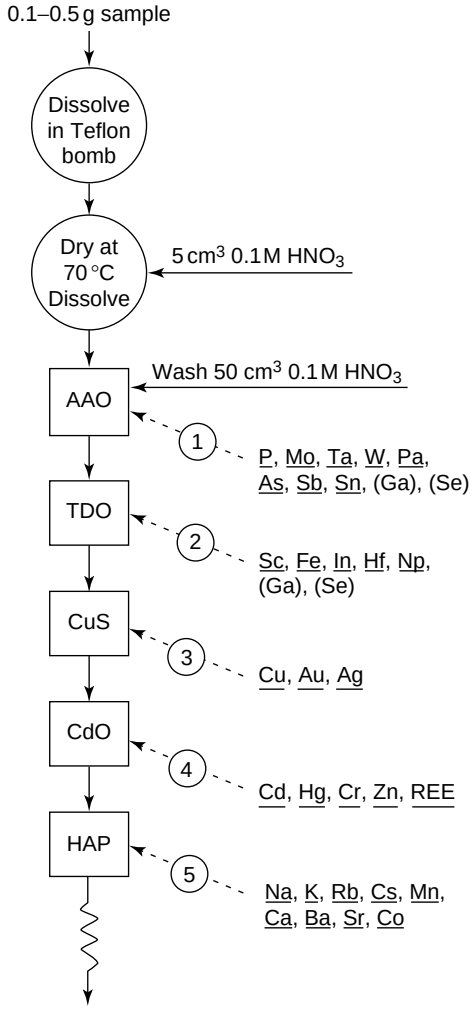

Figure 13 Radiochemical separation scheme for the determination of 39 elements from geological samples. (Reproduced from Pietra et al.[22] with permission.)

type are found in more than one fraction and cannot be determined. In some samples their concentration is too high and they might interfere with the determination of the other elements in the group. For these cases, the authors developed a more thorough scheme, in which for each fraction there is a method of removing the interference. However, these further steps should be done only if γ-spectrometry will not be sufficient.

Pietra et al.[22] developed 22 different RNAA procedures for environmental and biological samples. Many of these procedures are for several elements or for single elements; however, some are for group separations. The most extensive procedure (50 elements) was used only for biological samples. The next most extensive (39 elements) procedure was used for geological samples, and the separation scheme is depicted in Figure 13. The samples were dissolved in acid mixtures in a Teflon bomb, and, after drying, dissolved in 0.1-M HNO_3 which was also used as the only eluant. The trace elements were separated into five groups by the use of five different columns: acidic aluminum oxide (AAO), tin dioxide (TDO), copper sulfide (CuS), cadmium oxide (CdO) and HAP. The separation time for this procedure is about 2 h. All the IRNs were counted in the columns without elution. In another scheme where the geological material was dissolved in 6-M HF, the HAP column was used first, followed by the cation exchange column, and later an anion exchange column can be used to separate the mixture into four groups in which 35 elements can be determined. As in the previous scheme, in order to simplify the procedure and to make it easier for automation, only one eluant (6-M HF) is used.

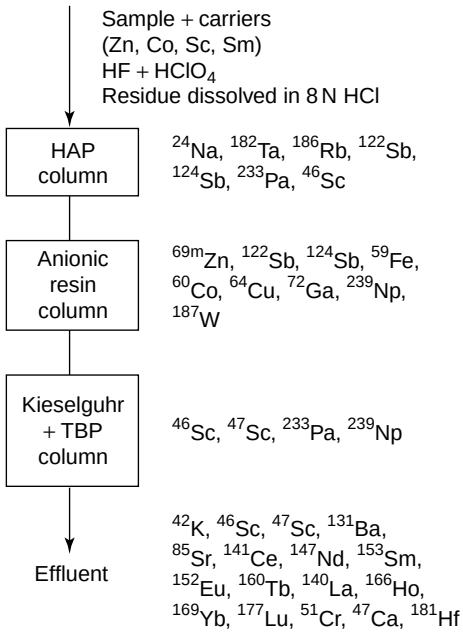

Figure 14 Simple group separation scheme for geological samples. (Reproduced from Vasconcellos and Lima[23] with permission.)

Vasconcellos and Lima[23] developed a procedure that is between several elements and a group separation. It is based on the use of three columns: HAP, anionic resin column and a reverse phase chromatography column with TBP on a Kieselguhr support. Several elements appeared in more than one fraction, but all the REE appear in one group and can be measured. The same is true for Na, Ta, Zn, Fe, Co, Cu, Ga, K, Ba, and Cr. The flow chart for the separation is given in Figure 14.

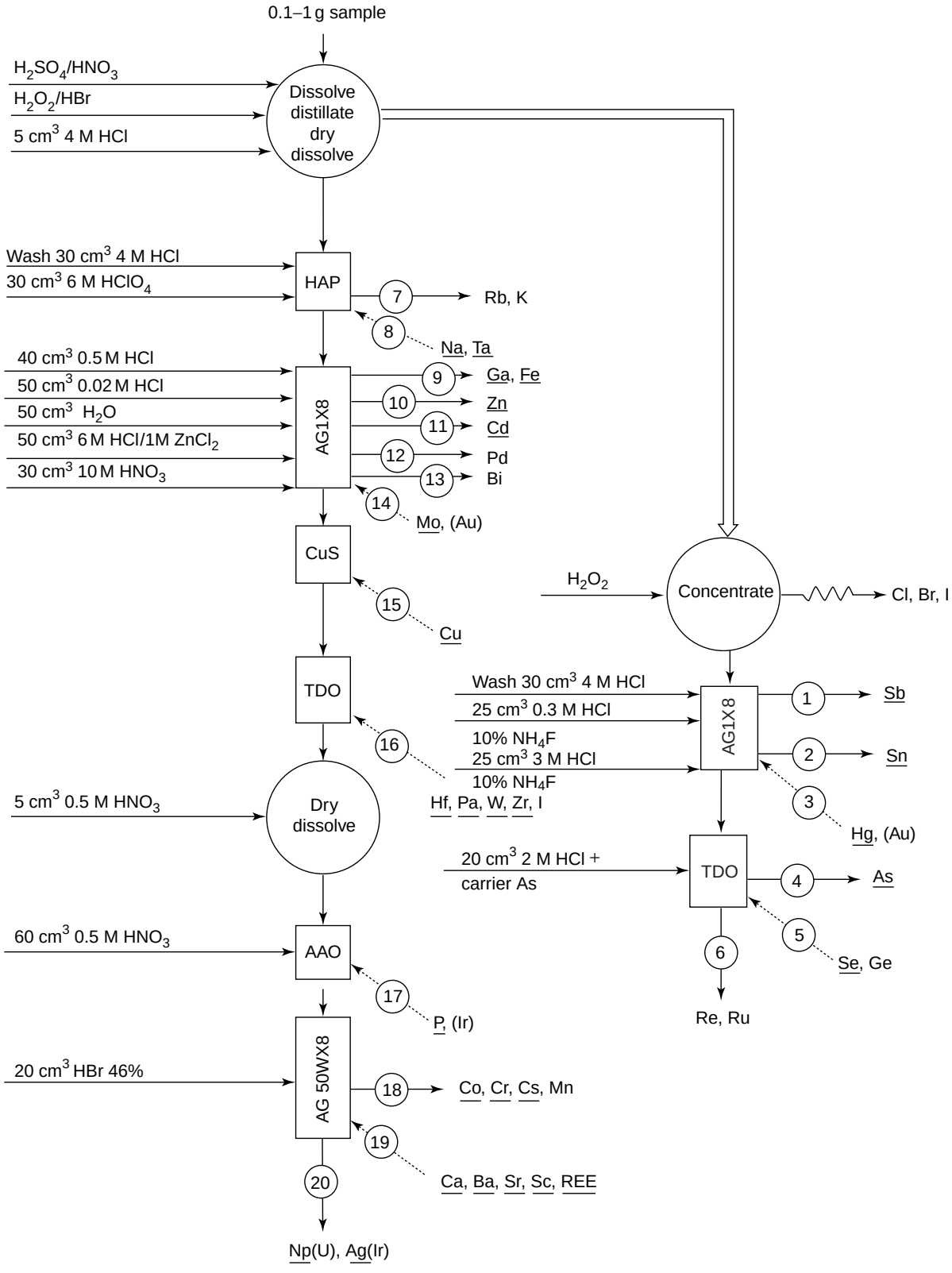

Figure 15 Pietra et al.[22] scheme for radiochemical separation for the determination of 50 elements in biological samples. (Reproduced from Pietra et al.[22] with permission.)

4 BIOLOGICAL SAMPLES

Many group separations were developed for biological materials. Some of them divide the sample into a few groups measured on a Ge(Li) or HPGe detector, while others perform very elaborate separations into many groups to enable activity measurements on very small amounts using a well-type NaI(Tl) detector. Figure 15 shows the flow chart of the Pietra et al.[22] procedure for the determination of 50 elements. The procedure starts with distillation of the elements (Cl, Br, I, Os) and the chlorides–bromides (Sb, Sn, Hg, Au, As, Se, Ge, Re, Ru). The chlorides–bromides are separated into six groups (not more than two elements in the same group) by the use of two columns (first an anion exchanger and then TDO) with successive elutions. An arrow from the column means elution. Elements with arrows to the column are retained on the column. Elements in brackets are present in more than one group. The nonvolatile elements/chlorides/bromides are separated using successive elutions with six columns: anion exchanger, CuS, TDO, AAO, and cation exchanger. The procedure of Pietra et al.[22] for the determination of 39 elements using five

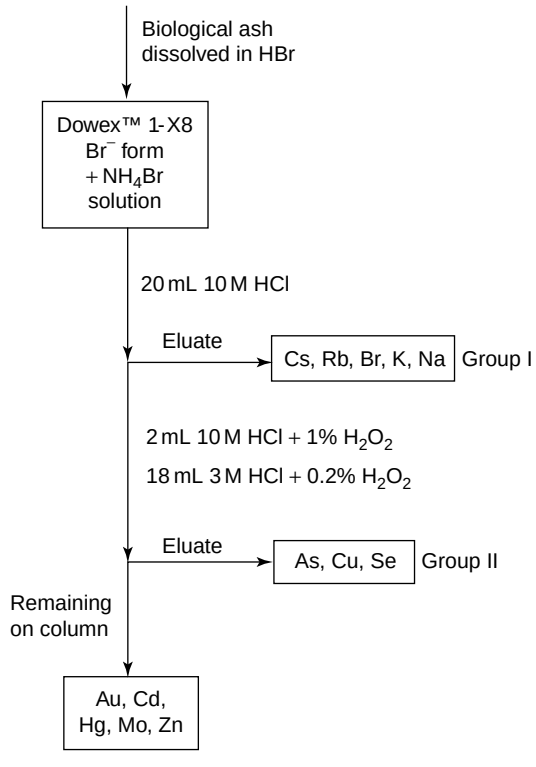

Figure 16 Determination of 10 trace elements in biological specimens by separation into three groups[24] (Na, K and Br were not determined).

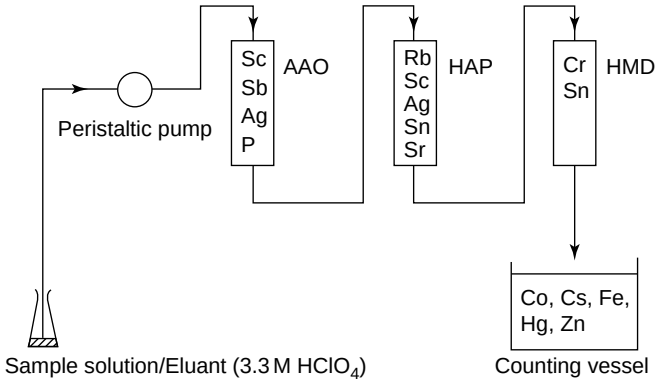

Figure 17 Determination of 16 trace elements in biological samples by separation into four groups using three columns. AAO, acidic aluminum oxide; HAP, hydrated antimony pentoxide; HMD, hydrated manganese dioxide. (Reproduced from Xilei et al.[25] with permission.)

columns with a single eluant, which is described for geological material, was also used for biological material (Figure 13).

Although 50 elements can be separated and determined, most studies do not need so much information, and most studies measured 10–20 elements by more simple methods. Pietra et al.[22] have several schemes for this purpose. van Renterghem and Cornelis[24] determined 10 elements in human serum. Their procedure divides 13 elements (Na, K, and Br were not determined in the serum) into three groups by using an anion exchange column in the Br^- form with two elutions. The scheme is shown in Figure 16.

Sixteen trace elements were determined by the same group using three columns and one eluant,[25] as can be seen in Figure 17.

Schuhmacher et al.[26] developed an almost complete scheme, in order to enable the activities to be measured with a well-type NaI(Tl) detector, to measure the contents of 25 trace elements in small biological samples (1–15 mg). The scheme is based on 14 consecutive columns divided into three Groups (3,7,4). For all columns in one group, the same eluant was used. Between the second and third group the eluate is automatically titrated to pH 2. The scheme is shown in Figure 18.

Figure 19 gives the scheme of Yeh et al.[27] for group separation, measuring 21 elements, not including Na and Br, which can also be measured. This scheme includes distillation of Br as an element, a HAP column, As and Se distillation as bromides, and an AAO column. A more extensive review on RNAA can be found in Alfassi.[28]

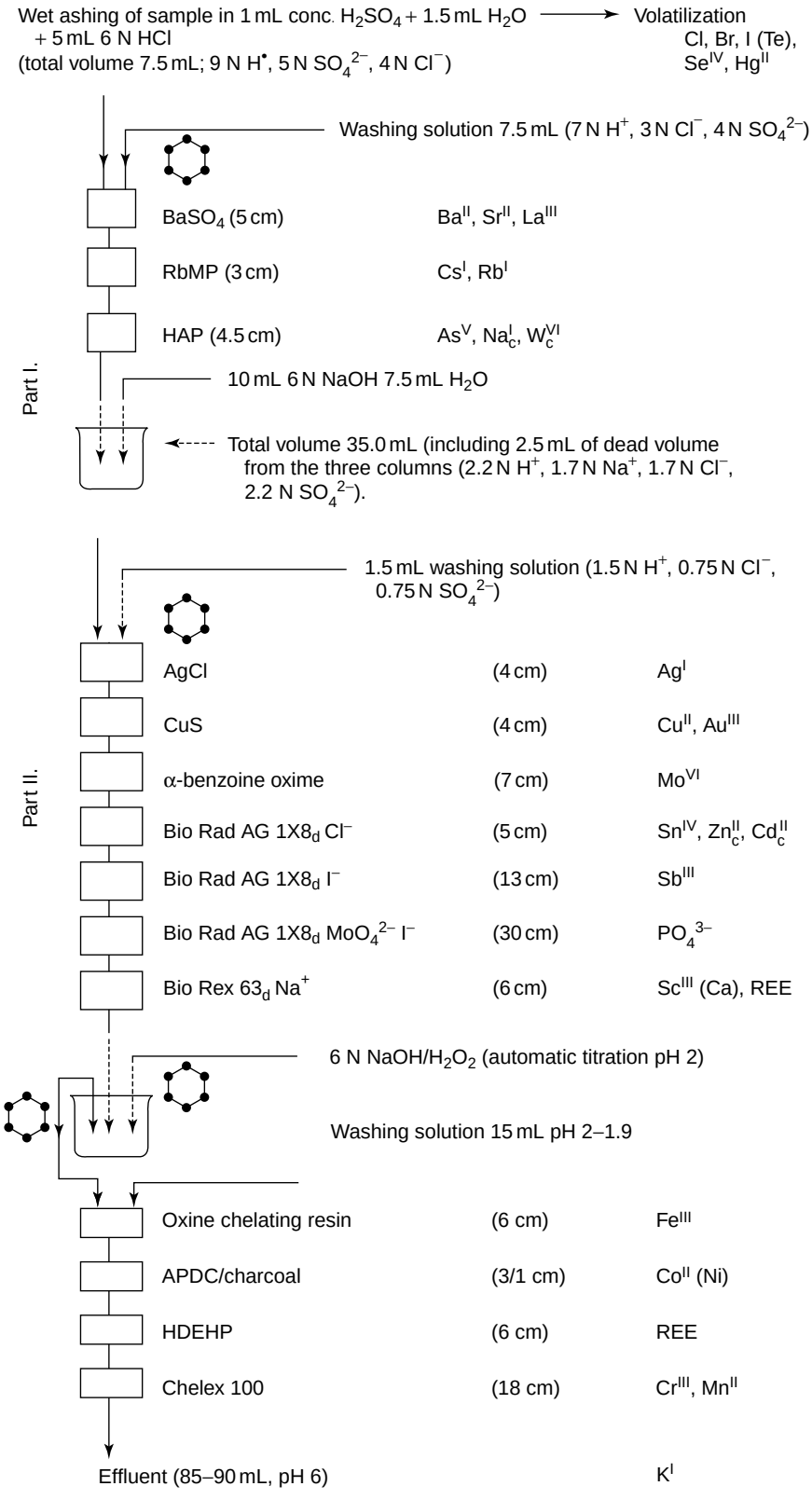

Figure 18 Half automated procedure for almost complete separation of 25 elements in biological samples. RbMP, Rubidium meta phosphate; APDC, ammonium pyrrolidine dithiocarbamate. (Reproduced from Schuhmacher et al.[26] with permission.)

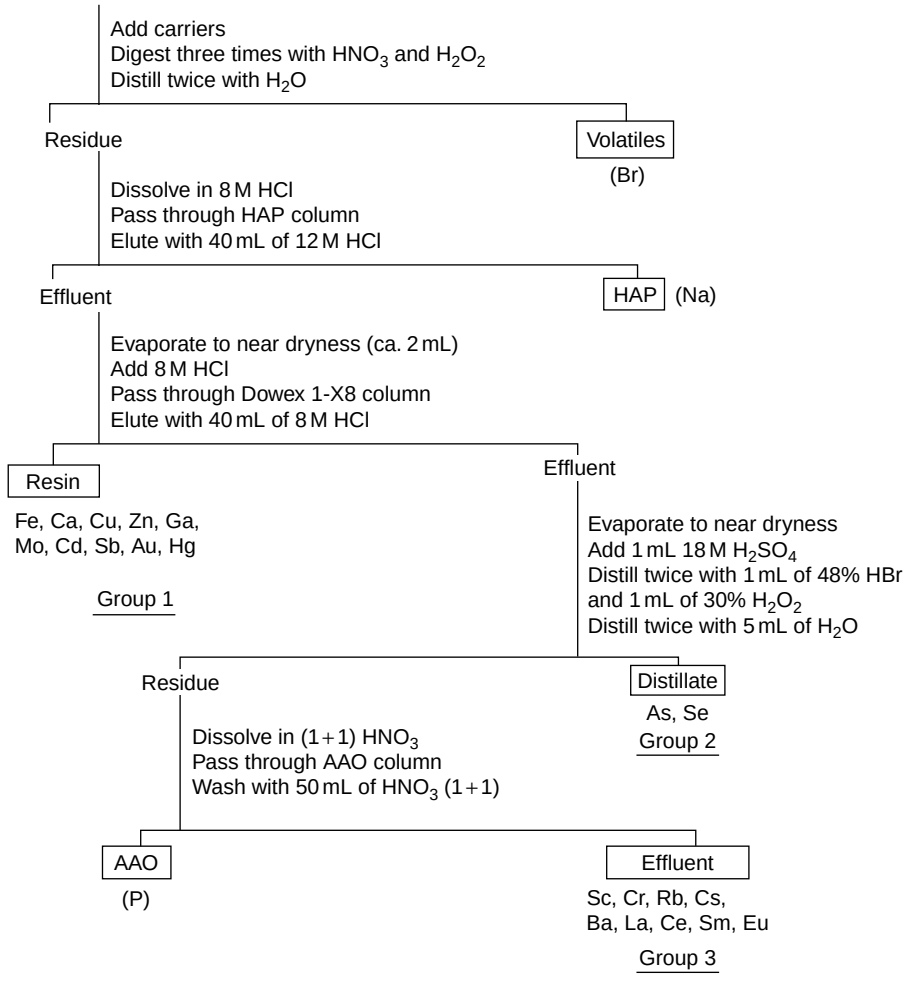

Figure 19 Yeh et al.[27] scheme for RNAA determination of 21 elements. (Reproduced from Yeh et al.[27] with permission.)

ABBREVIATIONS AND ACRONYMS

AAO	Acidic Aluminum Oxide
APDC	Ammonium Pyrrolidine Dithiocarbamate
DAM	Diantipyrilmethane
DCE	Dichloroethane
HAP	Hydrated Antimony Pentoxide
HDEHP	Di(2-ethylhexyl)-orthophosphoric Acid
HMD	Hydrated Manganese Dioxide
INAA	Instrumental Neutron Activation Analysis
IRN	Indicator Radionuclide
PGE	Platinum Group Elements
RbMP	Rubidium Meta Phosphate
REE	Rare Earth Elements
RNAA	Radiochemical Neutron Activation Analysis
TBP	Tributylphosphate
TDO	Tin Dioxide

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

REFERENCES

1. K.P. Egger, V. Krivan, *Fresenius' Z. Anal. Chem.*, **323**, 827 (1986).

2. K.S. Park, N.B. Kim, Y.S. Kim, K.Y. Lee, H.W. Choi, Y.Y. Yoon, *J. Radioanal. Nucl. Chem.*, **123**, 585 (1988).
3. K.H. Theimer, V. Krivan, *Anal. Chem.*, **62**, 2722 (1990).
4. R. Caletka, R. Hausbeck, V. Krivan, *J. Radioanal. Nucl. Chem.*, **120**, 305 (1988).
5. R. Caletka, C. Vorwalter, V. Krivan, *Anal. Chim. Acta*, **141**, 393 (1982).
6. R. Caletka, W.G. Faix, V. Krivan, *J. Radioanal. Chem.*, **72**, 109 (1982).
7. K.P. Egger, V. Krivan, *Fresenius' Z. Anal. Chem.*, **331**, 394 (1988).
8. R.A. Nadkarni, G.H. Morrison, *Anal. Chem.*, **46**, 232 (1974).
9. R.A. Nadkarni, G.H. Morrison, *J. Radioanal. Chem.*, **38**, 435 (1977).
10. S.J. Parry, *Analyst*, **105**, 1157 (1980).
11. H.W. Stockman, *J. Radioanal. Chem.*, **78**, 307 (1983).
12. S.L. Law, *Science*, **174**, 285 (1971).
13. R.A.A. Muzzareli, R. Rocchetti, *Anal. Chim. Acta*, **70**, 465 (1974).
14. C.F. Chai, S.L. Ma, X.Y. Mao, K.N. Liao, W.C. Liu, *J. Radioanal. Nucl. Chem.*, **114**, 281 (1987).
15. R. Zilliacus, M. Kaistilla, R.J. Rosenberg, *J. Radioanal. Chem.*, **71**, 323 (1982).
16. J.C. Laul, E.A. Lepel, W.C. Weimer, N.A. Wognan, *J. Radioanal. Chem.*, **69**, 181 (1982).
17. G.A. Wandless, J.W. Morgan, *J. Radioanal. Nucl. Chem.*, **92**, 273 (1985).
18. P. Parthasarathy, H.B. Desai, S.R. Kaysasth, *J. Radioanal. Nucl. Chem.*, **105**, 277 (1986).
19. G.H. Morrison, J.T. Gerard, A. Travesi, R.L. Currie, S.F. Peterson, N.M. Potter, *Anal. Chem.*, **41**, 1633 (1969).
20. J.C. Laul, D.R. Case, M. Wechter, F. Schmidt-Bleek, M.E. Lipschutz, *J. Radioanal. Chem.*, **4**, 241 (1970).
21. T. Smets, J. Hertogen, R. Gijbels, J. Hoste, *Anal. Chim. Acta*, **101**, 45 (1978).
22. R. Pietra, E. Sabbioni, M. Gallorini, E. Orvini, *J. Radioanal. Nucl. Chem.*, **102**, 69 (1986).
23. M.B. Vasconcellos, W.F. Lima, *J. Radioanal. Chem.*, **44**, 55 (1978).
24. D. van Renterghem, R. Cornelis, in *Trace Element Analytical Chemistry in Biology and Medicine*, eds. P. Bräter, P. Schramel, Walter de Gruyter, Berlin, 55, 1988.
25. L. Xilei, D. van Renterghem, R. Cornellis, L. Mees, *Anal. Chim. Acta*, **211**, 231 (1988).
26. J. Schuhmacher, W. Maier-Borst, H. Hauser, *J. Radioanal. Chem.*, **37**, 503 (1977).
27. S.J. Yeh, P.Y. Chen, C.N. Ke, S.T. Hsu, S. Tanaka, *Anal. Chim. Acta*, **87**, 119 (1976).
28. Z.B. Alfassi, in *Determination of Trace Elements*, ed. Z.B. Alfassi, VCH, Weinheim, 309, 1994.

Radiotracer Methods

Hendrik A. Das
Netherlands Energy Research Foundation (ECN), and University of Amsterdam, Amsterdam, The Netherlands

1 Introduction 1
2 Principle and Formulation 1
2.1 General 1
2.2 Radiotracer Experiments in Closed Systems 2
2.3 Radiotracer Experiments in Open Systems 3
2.4 Limitations of Radiotracer Methods 3
2.5 Practical Applications of Radiotracers 4
3 Measurement of Spatial Distributions 5
3.1 Survey 5
3.2 Profiling in Solids 5
3.3 Radiography and Tomography 6
3.4 Diffusion in Wetted Granular Solids 6
3.5 Diffusion in Liquids 8
4 Observations on Phase Equilibria and Related Kinetics 8
4.1 Survey 8
4.2 Speciation-controlled Diffusion 8
4.3 Kinetics of Colloid Association 9
4.4 Evaporation Kinetics of Traces of Organic Pollutants 9
Related Articles 10
References 10

Small amounts of radionuclides of suitable half-life can be used to observe the kinetics of mass transport and isotopic exchange in the equilibrium distribution of a (micro-)component between coexistent phases.

Quantitative assay of radioactivity is sensitive, precise and accurate and usually performed instrumentally. Though handling of radionuclides is restricted to authorized laboratories only, the combination of its potential and simplicity makes the radiotracer method an indispensable part of modern analytical chemistry.

1 INTRODUCTION

Small changes in isotopic composition by addition of minute quantities of either stable isotopes or radionuclides offers the possibility of labeling ions and molecules without significant shifts in their thermodynamic properties. This ''tracer'' principle enables the observation of kinetics, including isotopic exchange and equilibrium distributions.

The measurement of stable tracers is performed by mass spectrometry, usually on small aliquots of solution. As a rule, assay of radiotracers may be achieved by direct β or γ counting of relatively large subsamples. In case of α-active radiotracers, sample pretreatment is mandatory to obtain a thin-layered counting aliquot.

Radiotracers are applied in authorized laboratories only. Owing to the sensitivity of detection, the amount of activity, expressed in disintegrations per second per Becquerel (Bq) can be at quite a modest rate, usually below 10^5 Bq. Such quantities can be with handled with simple precautions in the many low-level tracer laboratories in hospitals, industry and university laboratories.

The most frequent application of the method is in medical diagnosis, followed by metabolic studies, environmental transport and availability experiments, including pilot plant simulations. The optimization of benchtop analytical procedures constitutes another established range of uses. This text concentrates on the principles of the radiotracer method and its nonmedical applications. A few illustrative cases have been taken from environmental studies, material science and analytical procedure development.

2 PRINCIPLE AND FORMULATION

2.1 General

The basic idea of the radiotracer method is the supposed identical behavior of stable and radioactive nuclides. Thus the presence of a tiny amount of an isotopic radionuclide, negligible in weight but well able to be detected by its radiation, enables observation of the ''labeled'' element. The chemical species of the radiotracer should be equal to that of the bulk mass in question. Thus both ions and (labeled organic) molecules are used. The small influence of the mass difference can be neglected in the large majority of applications. It becomes significant for light elements only, from about oxygen to elements of lower atomic mass, and in some artificial arrangements for isotope separation.

The signal obtained from a radiotracer depends on the concentration and the specific activity of the compound or phase which is being measured. In the case of a closed system, a radiotracer experiment gives information on the rate of the net mass-transport which affects the concentration, and on the rate of isotopic exchange which

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

influences the specific activity of the compound of interest. In open systems this distinction is not generally made but the variation of the signal is considered to be proportional to the changes in concentration.

2.2 Radiotracer Experiments in Closed Systems

The aim of radiotracer experiments in closed systems is to observe the interaction between two or more phases and compounds and, eventually, their equilibrium situation. Usually an amount of radioactivity is added as a component of one phase or compound and its quantity measured as a function of time. If the reaction between two compounds in one phase is examined, separation has to be performed prior to the measurement. Often an aliquot is taken for counting.

The result is a number of counts, A, related to the specific activity, a (disintegrations $s^{-1} g^{-1}$ or disintegrations $s^{-1} mol^{-1}$) and the concentration, c ($g\,L^{-1}$ or M) by Equation (1)

$$A = \gamma\, a\, c \tag{1}$$

The constant of proportionality, γ (in counts disintegration^{-1} s L or counts disintegrations^{-1} s mol), accounts for the volume of the aliquot and the counting conditions. At the beginning of the experiment Equation (2) follows

$$A_0 = \gamma\, a_0\, c_0 \tag{2}$$

and thus, Equation (3)

$$\frac{A}{A_0} = \frac{a}{a_0}\frac{c}{c_0} \tag{3}$$

If we now write Equation (4)

$$f_A = \frac{A}{A_0}, \quad f_a = \frac{a}{a_0} \quad \text{and} \quad f_c = \frac{c}{c_0} \tag{4}$$

it follows that Equation (5)

$$f_A = f_a f_c \tag{5}$$

and thus, with $df/dt = f'$, Equation (6)

$$f'_A = f_a f'_c + f_c f'_a \tag{6}$$

Equation (6), the activity balance, together with the mass balance, forms the basis of all radiotracer experiments in a closed system. In kinetic studies the third obvious equation is the rate balance: at equilibrium the net rate of two opposing reactions must be zero.

Without net mass transport $f_c = 1$ and $f'_c = 0$ and thus $f'_A = f'_a$ and $f_A = f_a$. In this case, a gradual isotopic exchange can be observed. If on the other hand there is no isotopic exchange, the reverse holds and $f_A = f_c$, enabling the measurement of net mass transport.

When both a and c change during the experiment, extra information is needed to solve Equation (6). This may take the form of a separate determination of f_c or a logical prediction of f_a. The various possibilities are summarized in Figure 1.

A clear distinction must be made between the amount of radiotracer which is present in the system and that which is available: in leaching experiments on radioactivated solids the available amount increases nonlinearly with time.

In general, the moment of introduction of the radioactivity does not coincide with the beginning of the

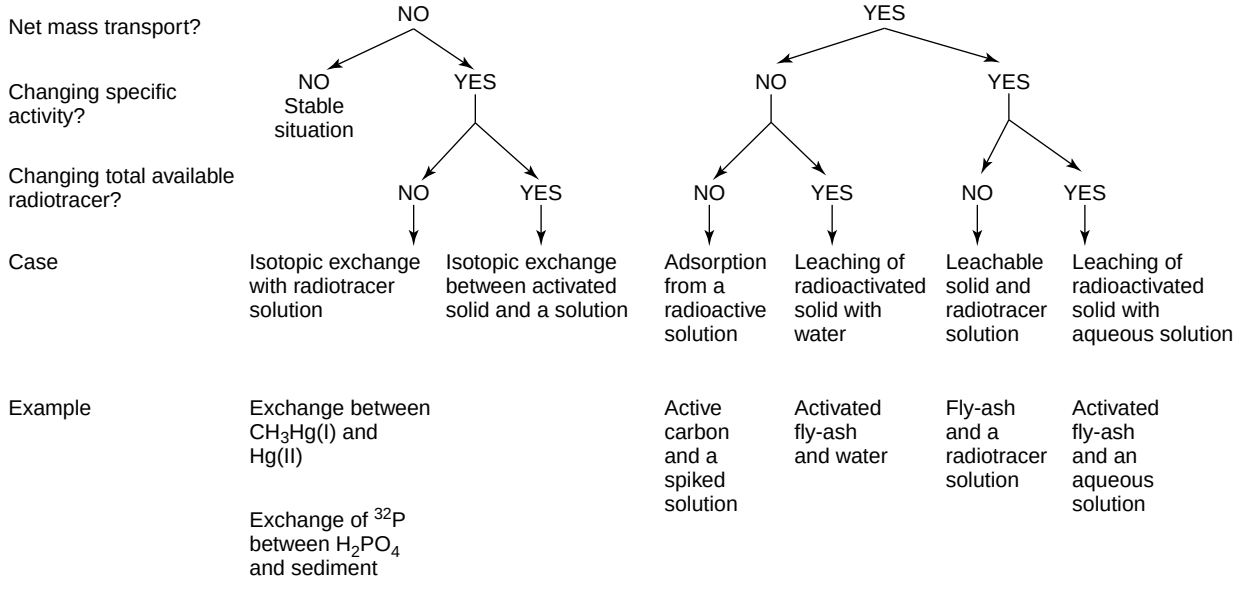

Figure 1 Survey of radiotracer experiments in a closed system.

interaction. By varying that moment, a (decreasing) ratio between net mass transfer and isotopic exchange may be observed.

Applications of radiotracer experiments to closed systems refer to interactions between compounds in one phase or different, adjacent, phases. As isotopic exchange or net mass transfer may be requested, this makes four possible combinations.

2.2.1 Isotopic Exchange

Generally, the reaction is of the type $AB + A^*C \longleftrightarrow A^*B + AC$. The asterisk indicates the radiotracer. By definition, the time of its introduction is taken as zero. By plotting the logarithm of the exchanged fraction against time, the rate constant can be obtained.[1–3] A similar formalism is applicable to isotopic exchange between two adjacent phases.

2.2.2 Net Mass Transport and Final Spatial Distribution

The case of net mass transport at a constant specific activity is by far the most frequent among radiotracer applications. A complication arises when the total available amount of radioactivity changes with time, as in the leaching of radioactivated or labeled solids. The activity of the leachant is measured as a function of time. It is obvious that the specific activity will remain constant only when the fresh eluent does not contain this element of interest. The mass balance may be written as Equation (7)

$$V\frac{\mathrm{d}c}{\mathrm{d}t} = GF(t) \tag{7}$$

where V is the volume of eluent (mL), c is the concentration in the eluent ($\mathrm{g\,mL^{-1}}$), t is the time (h), G is the amount of solid (g) and $F(t)$ is the specific mass transfer function ($\mathrm{g\,g^{-1}\,h^{-1}}$). Measurements of c as a function of t yields $F(t)$.

Table 1 gives the empirical expression for $F(t)$ and those for f_A, the number of counts as a function of time for a leachant without or with some initial concentration of the analyte.

2.3 Radiotracer Experiments in Open Systems

Radiotracers are used in open environmental systems to measure flow rate and/or time of residence as well as their distributions. This is invariably done by a one-time (δ-function) injection. Either the radioactivity is added to the system or induced by activation in small aliquots taken from the system. Unlike experiments in a closed system, f_c only is considered. Thus $f_a = 1$ and $f_A = f_c$, implying that the radiotracer remains in its original phase.

In principle, there are no special rules involved in the use of a radiotracer as the tagging agent. Thus the formalism developed by Danckwerts[4] and expanded in many textbooks, like that of Levenspiel,[5] may be used. The distribution of the radiotracer over the effluent, $f(t)$, is determined. From this the first and the second moment, average residence time and its variance are calculated. These measurements are mandatory in column leaching studies as the flow pattern influences the activity elution curve.

2.4 Limitations of Radiotracer Methods

Applications of radiotracers are restricted by fundamental as well as practical reasons. The limited choice of radiotracers and their half-lives and the isotope effects in reaction rates and equilibrium concentrations can be regarded as fundamental restrictions. The requirements for the operation of a radiochemical laboratory and the mandatory safety precautions in field experiments are of a practical nature.

The finite number of useful radionuclides and the limitations in their availability are reflected by any commercial catalogue. The most important consequence is the predominance of β-counting, usually by liquid scintillation, in physiological tracer experiments.

The influence of the mass difference between the natural isotopic mixture and the radiotracer can be expressed in terms of empirically determined correction factors for equilibria and reaction rates.[6,7] In elements with atomic masses greater than oxygen, the effect is usually lower than 5%.

Table 1 Functions f_c and f_A for the interaction between an activated leachable solid and an aqueous solution

c_0	$= 0$	$\neq 0$
$F(t)$	$k_1 e^{-\alpha t} - k_2(c - c_{eq})$	
$f_c = c/c_{eq}$	$\frac{k_1}{c_{eq}} \frac{e^{-\alpha t} - e^{\frac{-k_2 G}{V} t}}{\left(\frac{k_2 G}{V} - \alpha\right)} + \left(1 - e^{\frac{-k_2 G}{V} t}\right.$	$\frac{k_1}{c_{eq}} \frac{G}{V} \frac{e^{-\alpha t} - e^{\frac{-k_2 G}{V} t}}{\left(\frac{k_2 G}{V} - \alpha\right)} + \left(1 - e^{\frac{-k_2 G}{V} t}\right. + \frac{c_0}{c_{eq}} e^{\frac{-k_2 G}{V} t}$
$f_A = A/A_{eq}$	identical to f_c	$\left\{\frac{[c_0 V/G + k_1/\alpha](1 - e^{-\alpha t})}{c_0 V/G + k_1/\alpha(1 - e^{-\alpha t})}\right\} f_c$

2.5 Practical Applications of Radiotracers

The numerous applications of radiotracers can be systematized by division into three broad categories:

- optimization and performance of laboratory determinations by chemical separation
- measurement of spatial distributions
- observation of phase equilibria and related kinetics.

Some major trends are given below.

2.5.1 Application of Radiotracers in Laboratory Determinations by Chemical Separation

Optimization and, eventually, actual performance of trace determinations in the analytical laboratory may be done efficiently by means of radiotracers. Inorganic applications usually rely on γ-emitting radionuclides, while organic (i.e. biological and medical) experiments are mostly based on β-emitters. A second distinction is that between off- and on-line measurements. Obviously, the latter are preferred whenever they are feasible.

An example of an inorganic application is met in the determination of Hg^{II} in water by reduction with $SnCl_2$ and aeration; the escaping mercury vapor is trapped on active carbon.[8] Figure 2(a–c) illustrate to this procedure.

Radionuclides used in the separation and determination of organic trace compounds by high-performance liquid chromatography and off- or on-line counting are listed in Table 2; Figure 3 shows the flow-sheet of the apparatus.[9,10]

The choice between off- and on-line counting depends on the scope of the experiment and the specific activity of the radiotracer. The advantages of off-line counting are the variable counting time, the stable and low count rate of the background and the absence of any influence of chemiluminescence. The accompanying disadvantages comprise the inevitable ''breaking-up'' of the originally continuous chromatogram, the risk of irreproducible fractions and the time-consuming laborious procedure. Off-line scintillation counting is convenient when a small number of well-separated peaks are expected and if the amount of radioactivity, governed by the specific activity, is low.

Flow-through detection with a water-miscible liquid scintillator implies its continuous addition, usually at flow rates between 0.1 and 1 mL min^{-1} and subsequent mixing. The detector cell is ~50 µL in volume, thus appreciably larger than the usual ~10 µL cells of other detection methods. The advantages of flow-through counting are the automatic procedure and the high counting efficiency. The obvious disadvantage is band broadening in the cell.

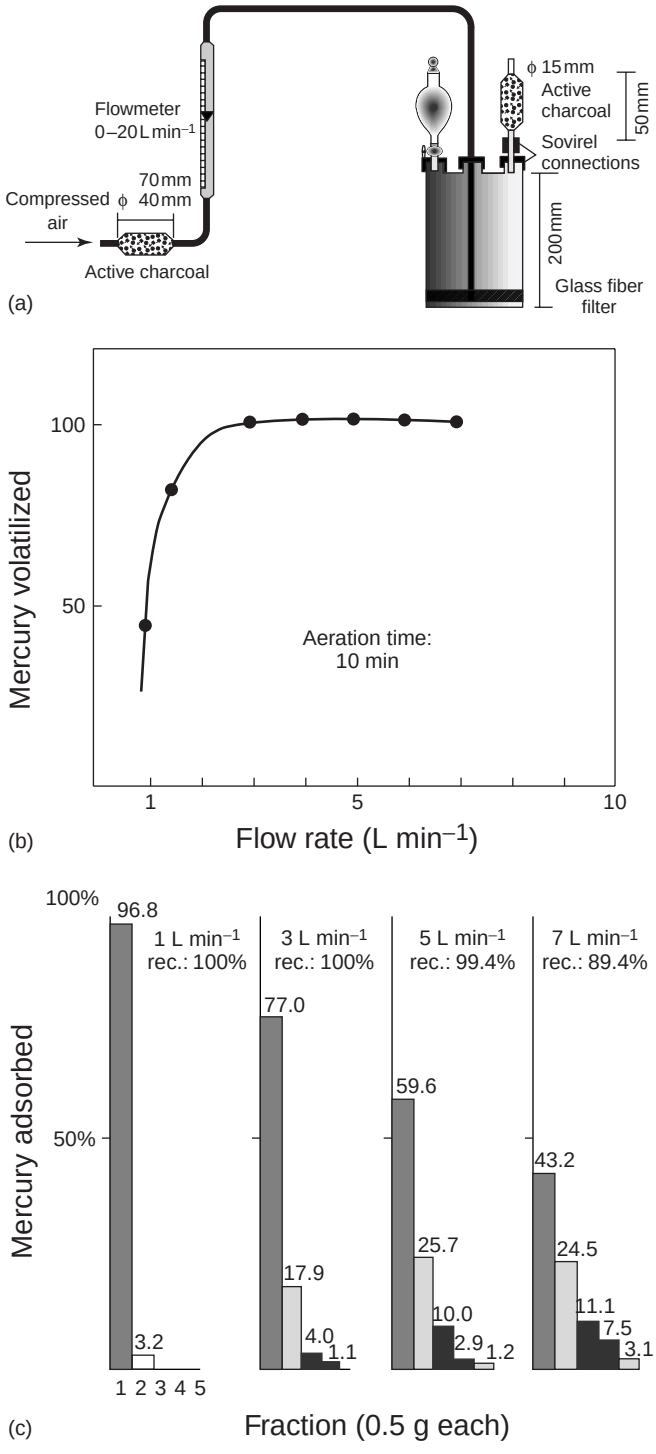

Figure 2 Determination of mercury in water samples. (a) The aeration vessel. (b) Influence of flow rate on volatilization of mercury. (c) Influence of flow rate on adsorption of mercury on activated charcoal. rec. = recovery.

Solvent segmentation based on postcolumn extraction/segmentation of the column eluate with a suitable liquid scintillation is the best option.[11]

Table 2 Radionuclides used in liquid scintillation counting

	E_γ(MeV)	E_β(keV)	$T_{1/2}$	Specific activity[a] (TBq mmol^{-1})	Range[b] (μm)
^{3}H		19	12.4 years	1	4
^{11}C	0.511		20.4 min		
^{14}C		155	5760 years	2 GBq mmol^{-1}	264
^{13}N	0.511		9.96 min		
^{15}O	0.511		2.03 min		
^{18}F	0.511		109.7 min		
^{32}P		1710	14.3 days	200	7870
^{35}S		167	87.4 days	30	302
^{36}Cl		700	3×10^5 years	3 KBq mmol^{-1}	2707
^{45}Ca		258	164 days	45 GBq mmol^{-1}	636
^{59}Fe	1.1; 1.2	500; 1600	44.6 days	40 GBq mmol^{-1}	
^{63}Ni		67	100 years	20 GBq mmol^{-1}	54
^{67}Ga	0.18; 0.30		78 h	25	
^{99m}Tc	0.14		6 h		
^{125}I	0.035	(X-ray)	60 days	75	
^{131}I	0.36	610	8 days	550	
^{140}La	0.49; 1.6	1380	40 h	5 GBq mmol^{-1}	

[a] Specific activities given are about the maximum values specified for commercially available radiolabeled compounds or elements. TBq mmol^{-1} unless stated otherwise.
[b] The maximum range given is calculated for pure β-emitters in water from the formula: Range in g cm^{-2} $= 0.11[(1 + 22.4\,E_\beta^2)^{1/2} - 1]$ with E_β in MeV.

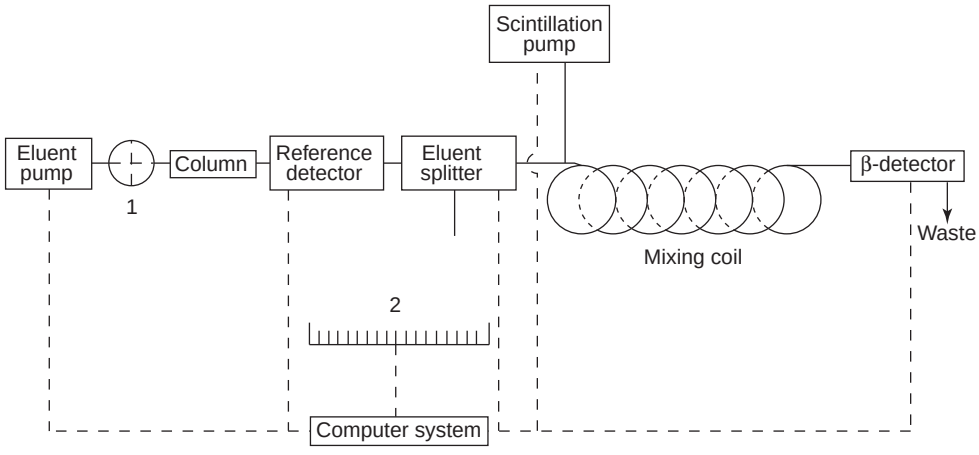

Figure 3 Apparatus for flow-through liquid scintillation counting. 1 = injector, 2 = fraction collector.

3 MEASUREMENT OF SPATIAL DISTRIBUTIONS

3.1 Survey

In assessing spatial distribution, radiotracers are most often used to measure net mass transfer of a previously spiked component within a known matrix. Thus radioanalysis is just one of the options. Conversely, the measurement of isotopic exchange is open to isotopic methods only. Here radioanalysis has to compete with mass spectrometry in measuring coefficients of self-diffusion.

The oldest application in spatial distributions is the study of self-diffusion in metals.[1–3] Relevant uses are met in the profiling of doping agents in semiconductors, tomography and the determination of diffusion coefficients in wetted granular solids and in liquids. In industrial pilot plants and factories, radiotracers are used to measure linear velocity and dispersion as mentioned in section 2.3.

3.2 Profiling in Solids

Flat solid surfaces of spiked samples are probed for their trace concentration profiles by abrasion or controlled

dissolution, followed by measurements on the material removed or the remaining aliquot. In general the first approach gives better precision but the actual advantage over measurement on the remaining aliquot depends strongly on the concentration profile.

The thickness of the layer removed in one step usually varies between 0.5 and 10 μm with a lower limit of ~0.1 μm. The radiotracer method thus fills the gap between macroscopic removal of surface layers and the physical methods for the submicron range.

A severe limitation is imposed by the required properties of the radiotracer and the mandatory combination of a radiochemical and solid-state laboratory which is found only in some university institutes and a few industrial laboratories. The use of special polishing apparatus with copper discs is also implied to obtain a precision of less than 1 μm.[11] The amount removed is determined by microbalance weighing at a precision of ≤10 μg.

Counting of the remaining aliquot is restricted to β-emitting radionuclides and is performed with a solid anthracene detector. Collection of the abrased material on the copper discs and the surface of the remaining aliquot is feasible within 1%. In the case of silicon HF dissolution is necessary which precludes the use of a liquid scintillator for β-emitters and restricts detection to Cerenkov counting and individual quenching correction.

3.3 Radiography and Tomography

There is no possibility or need to discuss the whole field of present day radiotracer-based radiography. Its use in biology and medicine as thin layer radiochromatography is obvious. As a relatively new feature, the use of storage phosphor screens can be mentioned, as this has increased the sensitivity and widened the applicability of double labeling procedures.[12]

From the beginning of the 1980s, computer-aided tomography of three-dimensional objects became a new branch of radionuclide application.[13] As with radiography, this text does not endeavor to treat this new specialization in detail; only principles are signaled.

Tomography is either based on the variation of the object's own γ-count rate with the position towards the detector or on the change in attenuation of transmitted photons or neutrons. The first approach may be based on previous neutron activation[14] or an added tracer; the last option belongs to the medical imaging techniques which will not be considered here. Industrial and nuclear waste tomography use the same principle.[15] The full width at half maximum of the scanning of a ^{198}Au-source was ~7 mm in the early experiments. Much better spatial resolution is obtainable in examining used fuel elements through a <1 mm slit of ~1 m length, as performed at ECN, Petten.

Using a ^{241}Am (59.5 keV) and a ^{137}Cs (661 keV) source in conjunction with collimators of 1–5 mm and distances of 50–150 mm resulted in full width at half maximum of 5–40 mm.[16] Comparable values (9–18 mm) were obtained in transmission and ^{131}I-based emission experiments with phantoms.[17] Off-line measurements on a neutron-activated bone sample by way of ^{24}Na, $T^{1/2} = 15$ h, gave a spatial resolution of 1–2 mm.[18]

On-line measurements in the 1.8×10^{12} cm^{-2} s^{-1} neutron beam at the Grenoble HFR gave ~7 mm full width at half maximum for $CdCl_2$ test pellets of that size.[19]

Apart from the routine measurements on used fuel elements, as performed at ECN, no systematic application of radionuclide-based tomography is reported: simulated fuel[19] and a test loop experiments with ^{24}Na[20] are scarce examples. In sharp contrast, the use of radioactive sources in industrial trouble-shooting as "go devils" is an established practice.

3.4 Diffusion in Wetted Granular Solids

Probably the most practical consequence of differences in element speciation is the variation in environmental mobility. Obviously the water content and the apparent diffusion coefficient of an inert tracer, i.e. one which does not interact with the solid, have to be determined. These data will then serve as reference values.

The use of a miniaturized system for diffusion measurements was first demonstrated in the case of traces of ^{239}Pu.[21,22] Then it was applied to self-diffusion under equilibrium conditions.[23]

Two identical cylinders are pressed together in a syringe. One part contains the radionuclide; the diffusion into the other section is measured. If a "weightless" spike is used, operation is under equilibrium conditions. This procedure yields the coefficient of self-diffusion, which for trace constituents is equal to the effective diffusion coefficient. When a component is involved which is present in solution only the apparent diffusion coefficient(s) obtained reflect(s) the tortuosity and constructivity of the void fraction(s). Experimentally, these two effects combine into one "restriction factor".

If a component is present in both the solid and the liquid phase and if the added radiotracer exchanges rapidly over the phase border, its diffusion will reflect the distribution ratio. When the element involved is present in more than one chemical form the result of the diffusion experiment will depend on the rate of isotopic exchange between the added spike and the (other) species. Only in the limiting cases of either very rapid exchange or no exchange at all, radiotracer diffusion will yield valuable information. Rapid exchange leads to the determination of the mass-averaged apparent diffusion coefficient. No exchange at

all implies the determination of the apparent diffusion coefficient of the spiked chemical form. The simple diffusion Equation (8) applies to a radiotracer which is exclusively present in solution and in one chemical form only and for the case of semi-infinite media.

$$\left(\frac{\partial(ac)}{\partial t}\right)_x = D\left(\frac{\partial^2(ac)}{\partial x^2}\right)_t \tag{8}$$

where a is the specific activity ($\mathrm{Bq\,g^{-1}}$), c is the concentration ($\mathrm{g\,L^{-1}}$), D is the coefficient of self-diffusion ($\mathrm{m^2\,s^{-1}}$), x is the position (m) and t is time (s).

At chemical equilibrium, for a ''weightless'' spike, this expression reduces to Equation (9)

$$\left(\frac{\partial a}{\partial t}\right)_x = D\left(\frac{\partial^2 a}{\partial x^2}\right)_t \tag{9}$$

From a plot of $\ln a$ against x^2 and curve fitting to the solution of Equation (9) for a finite system[23] the apparent diffusion coefficient together with information about the eventual coexistence of more than one flow-channel and their relative importance are obtained.

Under equilibrium conditions and with rapid isotopic exchange Equation (10) gives the experimentally apparent D-value

$$D_e = \frac{D/r}{1 + (m_S/m_L)} \tag{10}$$

where D refers to the ''true'' or unretarded diffusion coefficient in an aqueous solution, r to the mentioned restriction factor and m_S/m_L to the ratio of the exchangeable masses of the element involved in the solid and liquid phases. Whether the assumption of fast isotopic exchange is valid or not can be verified by varying the time lag between the application of the radiotracer to one compartment and the actual joining of the two sections. If this has no influence on the result, the condition is fulfilled.

As an inert radiotracer, tritiated water is used. Environmental investigations refer most often to anions as these feature the highest mobilities. Suitable radiotracers are $^{74}AsO_3^{III}$, $^{75}SeO_3^{II}$, and $^{75}SeO_4^{II}$, $^{99}MoO_4^{II}$ and $^{124}SbO_3^{III}$. It can be concluded that a specific activity of $\leq$10 mCi g^{-1} is sufficient for adding a ''weightless'' spike of $\leq$20 µg g^{-1}.

Figure 4 summarizes the procedure. About 3 g of the sample material is mixed with 0.9 mL of tracer solution and carefully homogenized. The diffusion tube consists of polythene with internal diameter 8 mm and length 70 mm. The interface of the two segments is gently marked with a tiny drop of ^{170}Tm solution.

After a storage time of 5–15 days, the combined segments are taken out, frozen in liquid nitrogen and sliced into 0.5–1 mm coupes. These are dried, weighted and counted. Data are processed to find the most likely D-value. Figure 5 gives the result for the diffusion of tritiated water through a fly-ash. The estimated value of the apparent diffusion coefficient is expressed as pDe. The usual error is 0.1 unit.

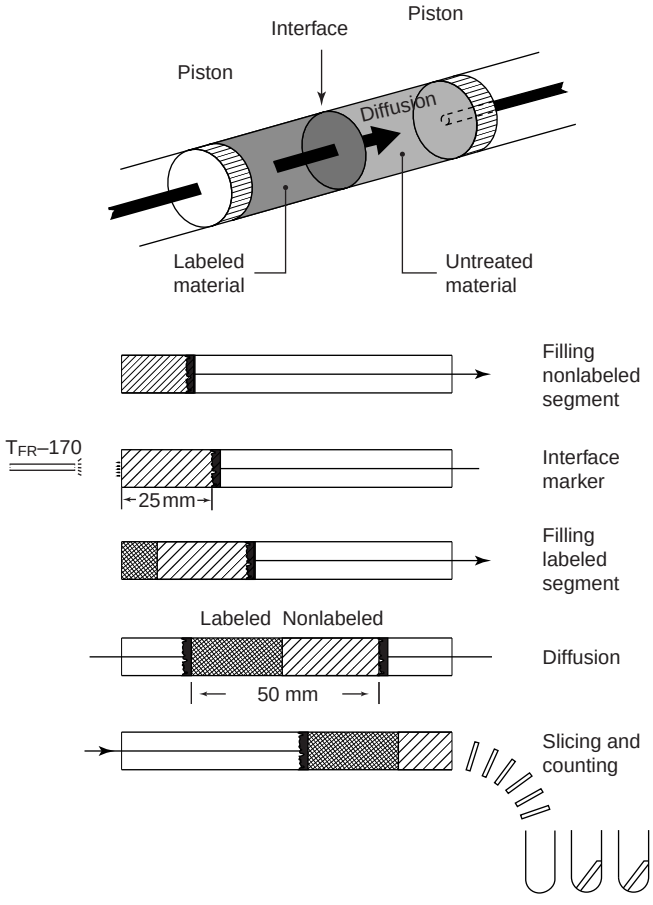

Figure 4 Measurement of the apparent diffusion coefficient After Das.[23]

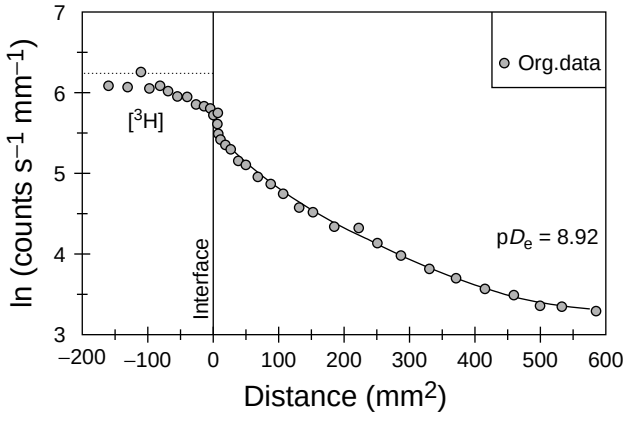

Figure 5 The diffusion of tritiated water in fly-ash. After Das.[23]

3.5 Diffusion in Liquids

The precise and accurate determination of diffusion coefficients of trace constituents (either of ionic or of molecular nature) in dilute, mostly aqueous, solutions is one of the obvious uses of radiotracers in the field of spatial speciation.

The main advantage of radiotracer experiments here, as in wet granular solids, is the possibility of measuring at zero diffusion gradient.[24] Two techniques have been put forward, based respectively on a glass frit diaphragm[25] and an open capillary within a large tank.[24]

Initially, a diaphragm cell of $\sim$100 mL was used.[28] Later, the capillary procedure came into use, offering the advantage of a small spiked volume of $\geq$10 μL per capillary. It implies stirring the outside solution and thus requires rigorous precautions against turbulent flow and mechanical dragging of the spiked liquid from the capillaries. In the final version just one capillary is used, mounted within the well of a bored-through plastic scintillator which monitors continuously the remaining radioactivity. The whole apparatus is mounted in a thermostated room. With a (small) correction for the variation of the counting geometry with proceeding diffusion, a precision of $\sim$0.2% can be reached.[26,27]

Reports on the use of the capillary procedure are found in Kepak et al.[29] and Podhajecky et al.,[30] while a reexamination of the diaphragm procedure is given by Podhajecky et al.[31]

4 OBSERVATIONS ON PHASE EQUILIBRIA AND RELATED KINETICS

4.1 Survey

Within a system in (apparent) equilibrium, elements often occur in different physical phases. The potential use of radiotracers in measuring these distributions, and the kinetics with which they are reached, is obvious. The early, often ''classical'', applications are mentioned in most textbooks on radiochemistry. The number of literature reports on the subsequent use in all branches of science, technology and medicine is enormous. Virtually all applications refer to the measurement of net mass transport (cf. section 2.2).

Current literature abounds with (more or less) defined procedures for physical speciation,[32–41] particulate matter and colloids in natural waters[42–54] and availability measurements on trace metals in granular solid wastes.[55–57] Eventually these investigations pertain to the occurrence and distribution of radionuclides from nuclear bomb tests and waste reprocessing.[58–67] Quoted references give some examples only.

A methodical application is comparison of analytical techniques on water containing particulate matter.[68] Equally, preconcentration and storage have been examined with the help of radiotracers. The contact time between water sample and container, whether polythene or glass, should not exceed more than a few minutes. In situ immobilization, eventually combined with preconcentration, is imperative. Scavenging with active carbon after adsorption of a suitable chelating agent is a possibility.[42] Collection on Chelex-100 or on a combination of an anion and cation exchanger is well known; a survey is given by Florence.[69]

Freezing, in combination with freeze-drying, is the best way of preserving the sample in its original state.[70] Acidification to pH $\sim$ 1–2 is justified only after thorough elimination of particulate matter, to avoid desorption. Using ^{203}Hg, it was found that Hg^{II} desorbs completely at pH = 0.1 with a half-life of $\sim$30 min.[42]

The quest for well-defined and generally applicable speciation schemes has been successful in a few cases only. The separation of colloids from fresh water samples by the hollow fiber technique[47,52] is an example; the standard leaching test on coal fly-ash is another,[55] although a recent comparison revealed some systematic discrepancies between the various modes.[71] Sequential leaching of soils by a series of increasingly aggressive solutions has been applied in a multitrade of schemes, all of limited significance. Several articles[72–82] giving examples have been published.

Applications of radiotracers to adsorption and desorption measurements and in isotopic exchange experiments were mentioned in section 2. Three practical applications are considered here:

- measurement of speciation controlled diffusion through soils[82,84]
- kinetic association experiments with colloids in water[85]
- evaporation kinetics of traces or organic pollutants from surface water.[86]

4.2 Speciation-controlled Diffusion

Figure 6 outlines the apparatus used. The ''steady state'' experiment imposes a linear gradient over the test sample but it takes a considerable time to reach it.[83] Typical experiments with sand samples take $\sim 10^3$ h. The radionuclide concentration in the unspiked reservoir is always much lower than that in the spiked compartment. This eliminates back diffusion.

The concentration in the unspiked cell is monitored until its rate of increase becomes constant. Relevant mathematics are given by Crank.[83] The crucial feature is that $\partial^2 c/\partial x^2$ is zero and thus $\partial c/\partial x$ is constant.

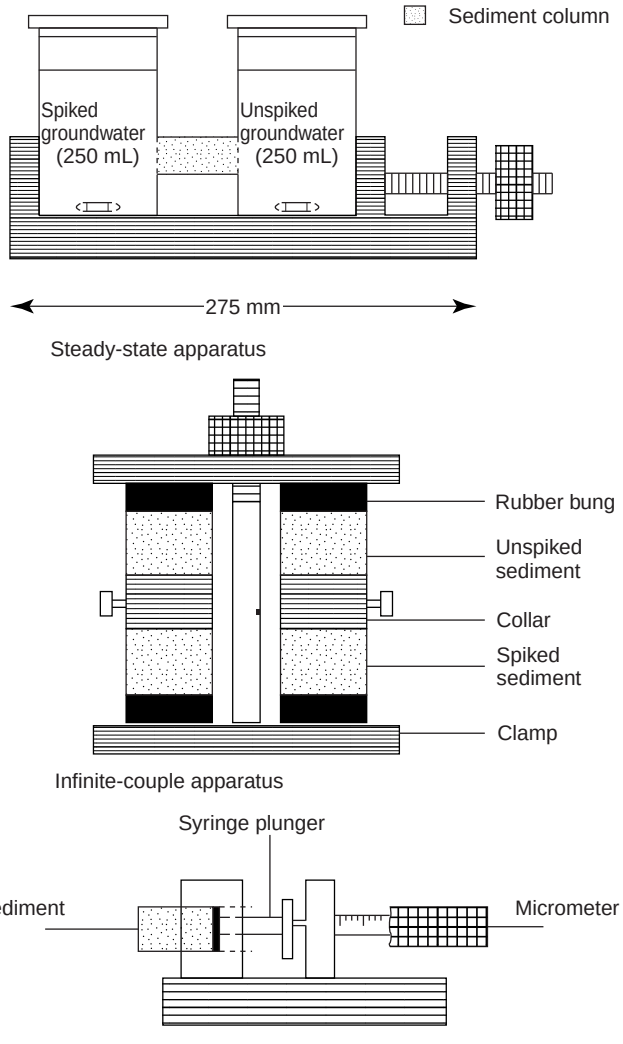

Figure 6 Equipment for steady state diffusion through wet sand.[82] (J.J.W. Higgo, Iodine Speciation and Diffusion in a Sand–Groundwater System, DG Science, Research. Development, 1991. Reproduced by permission of the British Geological Survey.)

"Infinite couple" or "back-to-back" geometry has two cylinders of material, one of them spiked, pressed together. It thus resembles the array discussed in section 4.1, except for its much larger dimensions. These enable the use of large aliquots of groundwater, taken from a separate equilibration tank.[86]

The effective diffusion coefficient and distribution constant was determined for both I^- and IO_3^-, spiked with $^{125}I(T^{1/2} = 60\,\text{day})$, as a function of the solid/solute ratio. Data were found to follow Freundlich isotherms.

4.3 Kinetics of Colloid Association

The use of spikes in radiotracers to determine recoveries in physical and chemical separations is obvious. Their most apparent application is met in the experiments on

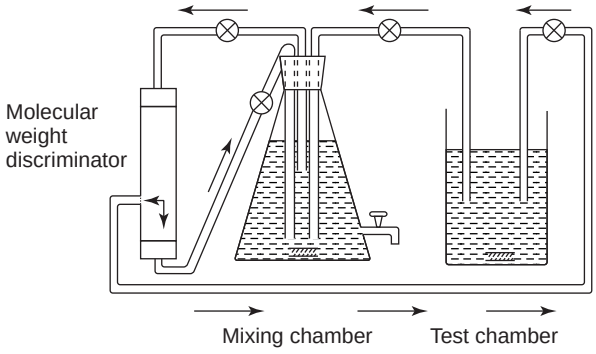

Figure 7 Equipment for the measurement of trace element association with colloids. After Salbu.[47]

hollow fiber preconcentration, using an "Amicon" apparatus, of naturally occurring radionuclides.[47,48,66,85–88] Cartridges with nominal molecular weight cut-off of 10^5, 10^4 and 10^3 are applied. Isolated fractions are analyzed by γ-spectrometry.

Kinetic association experiments are performed with the apparatus shown schematically in Figure 7. The sample is introduced into the test chamber and transported by a peristaltic pump to a mixing chamber and separator. The ultrafiltrate returns to the test chamber while the retained colloids go to the mixing chamber. The decrease in tracer radioactivity in the test solution reflects dilution, association with naturally occurring colloids and, eventually, sorption to the equipment.

It is found that equilibrium is usually reached in 1–2 h. The majority of naturally occurring ^{60}Co and added ^{65}Zn is in the $>10^4$ D fraction.

An alternative to dialysis may be found in electrophoresis.[33,64] Free liquid electrophoresis was applied to speciation studies of radium in natural waters using ^{224}Ra as the radiotracer.[64] Results indicate that in 10^{-2} M chloride solutions the Ra^{II} cation is the main chemical species between pH = 2 and pH = 7.

4.4 Evaporation Kinetics of Traces of Organic Pollutants

Radiotracers are applied to environmentally important kinetics. In evaporation measurements, the overall mass transfer coefficient, K, in $\text{mol}\,\text{min}^{-1}\,\text{cm}^{-2}$ has to be obtained.[89] To this end, evaporation experiments are conducted with labeled trace constituents from open vessels with a known gas–liquid surface area. The concentration of the (volatile) trace compound decreases exponentially with a characteristic constant KS/V_L for a given surface area S (cm^2) and liquid volume V_L (mL).

In a closed flask, the equilibrium concentration in the solution reflects the thermodynamic potential of the

trace constituent in the liquid phase. By measuring this equilibrium as a function of the initial concentration, eventual deviation from ideality can be detected.

''Carrier-free'' ^{3}H- or ^{14}C-labeled compounds are used to spike stock solutions of the chemicals of interest. With commercially available specific activities of $\sim$1 TBq ^{3}H (m mol^{-1}), a liquid volume of $\sim$50 mL, an aliquot size of 2 mL and a counting time of 100 min, the limit of determination at $\sigma_{rel} \sim 1\%$ is 10^{-2} ppb. For ^{14}C-labeled compounds of $\sim$2 GBq ^{4}C (m mol^{-1}) the lower limit is $\sim$5 ppb.

RELATED ARTICLES

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Nuclear Methods **(Volume 14)**
Elemental Analysis by Isotope Dilution • Instrumental Neutron Activation Analysis • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Rutherford Backscattering Spectroscopy • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Actinides and other Alpha-emitters, Determination of • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Mass Spectrometry of Long-lived Radionuclides

REFERENCES

1. M. Haïssinsky, *Nuclear Chemistry and its Applications* (English translation), Addison-Wesley, Reading MA, 536–565, 1964.
2. H.A.C. McKay, *Principles of Radiochemistry*, Butterworths, London, 393–445, 1971.
3. G. Friedlander, J.W. Kennedy, J.M. Miller, *Nuclear and Radiochemistry*, 2nd edition, J. Wiley & Sons, New York, 191–220, 1964.
4. J.M. Smith, *Chemical Engineering Kinetics*, 3rd edition, McGraw-Hill Int. Book Company, London, 268–297, 1981.
5. O. Levenspiel, *Chemical Reaction Engineering*, J. Wiley & Sons, New York, 242–308, 1962.
6. L. Melander, *Isotope Effects on Reaction Rates*, Ronald Press, New York, 46–64, 1960.
7. M. Haïssinsky, *Nuclear Chemistry and its Applications* (English translation), Addison-Wesley, Reading, MA, 640–669, 1964.
8. H.A. van der Sloot, H.A. Das, 'The Determination of Mercury in Water Samples from the Environment by Neutron Activation Analysis', *Anal. Chim. Acta*, **73**, 235–244 (1974).
9. A.C. Veltkamp, H.A. Das, R.W.Frei, U.A.Th. Brinkman, E.C.N. Petten, 'On-line Low-level Radiometric Detection of [14C]Remoxipride in Liquid Chromatographic Effluents. Application to Urine Samples', *J. Chromatogr.*, **384**, 357–369 (1987).
10. A.C. Veltkamp, H.A. Das, U.A.T. Brinkman, 'Direct and Reverse Flow-through Heterogeneous Scintillation Counting of Radiolabeled Amino Acids Using Post-column Solvent Segmentation', *J. Chromatogr.*, **508**(2), 301–308 (1990).
11. L.M.L.J. Leblans, M.L. Vetcheyke, *Philips Tech. J.*, **25**, 103–106 (1963).
12. R.R. Johnston, S.C. Pickett, D.L. Barker, 'Quantitative Double-label Autoradiography Using Storage Phosphor Imaging', *Electrophoresis*, **11**, 355–360 (1990).
13. G. Davies, N.M. Spyrou, I.G. Hutchinson, J. Huddleston, 'Applications of Emission Tomography in the Nuclear Industry', *Nucl. Instrum. Methods Phys. Res., Sect. A*, **A242**(3), 615–619 (1986).
14. T.B. Pierce, J. Huddleston, I.G. Hutchinson, 'A Study of the Possible use of Tomographic Techniques for Examining the Distribution of Activity in Irradiated Samples', *Radiochem. Radioanal. Lett.*, **48**, 81–92 (1981).
15. D.J. Gentle, N.M. Spyrou, 'Region of Interest of Tomography in Industrial Applications', *Nucl. Instr. Methods*, **A299**, 534–537 (1992).
16. T.B. Pierce, J. Huddleston, I.G. Hutchinson, 'The Preliminary Investigation into the Potential Use of Tomographic Techniques for the Examination of Mechanical Components', *J. Radioanal. Chem.*, **79**, 287–301 (1983).
17. J.M. Sanders, N.M. Spyrou, 'Improvement of Image Quality in Emission and Transmission Tomography by Radiation of Scattered Photons', *Nucl. Instr. Methods*, **A221**, 93–97 (1984).
18. N.M. Spyrou, 'Prompt and Delayed Radiation Measurements in the Elemental Analysis of Biological Materials: The Case for Neutron Induced Gamma-ray Emission Tomography', *J. Radioanal. Nucl. Chem.*, **110**(2), 641–653, (1987).
19. N.M. Spyrou, Kusminazto, E.G. Nicolau, '2D Reconstruction of Elemental Distribution Within a Sample Using Neutron Capture Prompt Gamma-rays', *J. Radioanal. Nucl. Chem.*, **112**, 57–64 (1987).
20. D.J. Finnigan, F. McFall, J. Huddleston, I.G. Hutchinson, 'Application of Emission Tomography to the Study of Solute Concentration Processes in PWR Steam Generators', *Water Chem. Nucl. React. Syst.*, **5**(2), 182–183 (1990).

21. F. Schreiner, S. Fried, A.M. Friedman, 'Measurement of Radionuclide Diffusion in Ocean Floor Sediment and Clay', *Nucl. Technol.*, **59**, 429–438 (1982).
22. B. Torstenfeld, 'Migration of the Fission Products Strontium, Technetium, Iodine and Cesium in Clay', *Radiochim. Acta*, **39**(2), 97–104 (1986).
23. H.A. Das, H.A. van der Sloot, G.J. de Groot, R.C.H. Steneker, 'Determination of the Apparent Diffusion Coefficients in Water Containing Granular Solids', *Int. J. Appl. Radiot. Isot.*, **40**, 615–619 (1989).
24. H.J.M. Bowen, *Chemical Applications of Radioisotopes*, London, Methuen, 1969.
25. R. Mills, 'A Remeasurement of the Self-diffusion Coefficients of Sodium Ion in Aqueous Sodium Chloride Solutions', *J. Am. Chem. Soc.*, **77**, 6116–6119 (1955).
26. R. Mills, E.W. Godbole, 'The Precision Measurement of Single Ion Diffusion Coefficients', *Austr. J. Chem.*, **11**, 1–8 (1958).
27. R. Mills, E.W. Godbole, 'The Precision Measurement of Single Ion Diffusion coefficients', *Austr. J. Chem.*, **12**, 102–103 (1959).
28. R. Mills, A.W. Adamson, 'The Measurement of Self-diffusion in Electrolyte Solutions', *J. Am. Chem. Soc.*, **77**, 3454–3458 (1955).
29. F. Kepak, J. Teply, J. Kriva, 'The Study of the Effect of Gramma Radiation on the Ionic and Colloidal Solutions of ^{144}Ce(III) at Trace Concentration by the Method of Self-diffusion and Centrifugation', *Radiochem. Radioanal. Lett.*, **21**, 365–370 (1975).
30. P. Podhajecky, P. Benes, A. Gosman, 'Radiotracer Analysis of the Physicochemical State of Trace Elements in Aqueous Solutions. I. Capillary Diffusion Method', *J. Radioanal. Chem.*, **51**(2), 315–324 (1979).
31. P. Podhajecky, A. Gosman, P. Benes, 'Radiotracer Analysis of the Physicochemical State of Trace Elements in Aqueous Solutions. IV. Diaphragm Diffusion Method', *J. Radioanal. Chem.*, **57**(2), 253–262 (1980).
32. M. Kersten, U. Foerstner, 'Effect of Sample Pretreatment on the Reliability of Solid Speciation Data of Heavy Metals – Implications for the Study of Early Diagenetic Processes', *Mar. Chem.*, **22**(2–4), 299–312 (1987).
33. P. Benes, M. Kuncova, J. Slovak, P. Ramos, 'Analysis of the Interaction of Radionuclides with Solid Phase in Surface Waters Using Laboratory Model Experiments: Methodical Problems', *J. Radioanal. Nucl. Chem.*, **125**(2), 295–315 (1988).
34. V. Spevackova, J. Kucera, 'Trace Element Speciation in Contaminated Soils Studied by Atomic Absorption Spectrometry and Neutron Activation Analysis', *Int. J. Environ. Anal. Chem.*, **35**(4), 241–251 (1989).
35. R. Chester, F.J. Lin, K.J.T. Murphy, 'A Three Stage Sequential Leaching Scheme for the Characterization of the Sources and Environmental Mobility of Trace Metals in the Marine Aerosol', *Environ. Technol. Lett.*, **10**(10), 887–900 (1989).
36. A. Bermond, G. Sommer, 'Simulation of Heavy Metals Extraction in Soil Samples Compared with Experimental Results', *Environ. Technol. Lett.*, **10**(11), 989–994 (1989).
37. P.M.V. Nirel, F.M.M. Morel, 'Pitfalls of Sequential Extractions', *Water Res.*, **24**(8), 1055–1056 (1990).
38. M. Sager, R. Belocky, R. Pucsko, 'On the Determination of the Speciation of Major and Trace Elements in Sediments by Sequential Solution Techniques', *Acta Hydrochim. Hydrobiol.*, **18**(2), 157–173 (1990).
39. S. Dudka, A. Chlopecka, 'Effect of Solid-phase Speciation on Metal Mobility and Phytoavailability in Sludge-amended Soil', *Water, Air, Soil Pollut.*, **51**(1–2), 153–160 (1990).
40. A.M. Ure, 'Trace Elements in Soil: Their Determination and Speciation', *Fresenius' J. Anal. Chem.*, **337**(5), 577–581 (1990).
41. A.M. Ure, 'Trace Element Speciation in Soils, Soil Extracts, and Solutions', *Mikrochim. Acta*, **2**(1–6), 49–57 (1991).
42. H.A. van der Sloot, 'Neutron Activation Analysis of Trace Elements in Water Samples After Preconcentration on Activated Carbon', Report ECN-1, 1976, Petten, The Netherlands.
43. J.C. Duinker, R.F. Nolting, H.A. van der Sloot, 'The Determination of Suspended Metals in Coastal Waters by Different Sampling and Processing Techniques (Filtration, Centrifugation)', *Neth. J. Sea Res.*, **13**(2), 282–297 (1979).
44. B. Salbu, H.E. Bjoernstad, N.S. Lindstroem, E.M. Brevik, J.P. Rambaek, J.O. Englund, K.F. Meyer, H. Hovind, P.E. Paus, 'Particle Discrimination Effects in the Determination of Trace Metals in Natural Fresh Waters', *Anal. Chim. Acta*, **167**, 161–170 (1985).
45. A. Mizuike, 'Recent Developments in Trace Metal Speciation in Fresh Water', *Pure Appl. Chem.*, **59**(4), 555–564 (1987).
46. J.H.D. Vangenechten, N.A. Chughtai, J. Bierkens, O.L.J. Vanderborght, 'Similarity of Americium-241 and Iron-59 Speciation in Selected Freshwaters and of their Adsorption on Crayfish Exoskeleton', *J. Environ. Radioact.*, **5**(4), 275–286 (1987).
47. B. Salbu, H.E. Bjoernstad, E. Lydersen, A.C. Pappas, 'Determination of Radionuclides Associated with Colloids in Natural Waters', *J. Radioanal. Nucl. Chem.*, **115**(1), 113–123 (1987).
48. B. Salbu, 'Radioactive Tracer Techniques in Speciation Studies', *Environ. Technol. Lett.*, **8**(8), 381–392 (1987).
49. N. Breward, D. Peachy, 'The Development of Portable Equipment to Study Physical and Chemical Phases in Natural Waters', DOE Report RW/888, 102.
50. J. John, B. Salbu, E.T. Gjessing, H.E. Bjoernstad, 'Effect of pH, Humus Concentration, and Molecular Weight on Conditional Stability Constants of Cadmium', *Water Res.*, **22**(11), 1381–1388 (1988).

51. M. Obdrzalek, P. Benes, 'Analysis of the Speciation of Lead in River Waters Using a Radiotracer Method', *Radioisotopy*, **29**(1–2), 202–209 (1988).
52. S.C. Apte, M.J. Gardner, D.T.E. Hunt, 'An Evaluation of Dialysis as a Size-based Separation Method for the Study of Trace Metal Speciation in Natural Waters', *Environ. Technol. Lett.*, **10**(2), 201–212 (1989).
53. A. Mizuike, 'Preconcentration of Trace Elements in Water for Speciation Studies', *J. Chin. Chem. Soc.*, (Taipei), **37**(2), 117–124 (1990).
54. K. Hirose, 'Chemical Speciation of Trace Metals in Seawater: Implication of Particulate Trace Metals', *Mar. Chem.*, **28**(4), 267–274 (1990).
55. ANS Standards Committee, Working Group ANS 16.1. Report ANSI-ANS 16-1-1986.
56. H.A. van der Sloot, 'A Standard Leaching Test for Combustion Residues', Report BEOP-31, Arnhem, The Netherlands, 1984. Available from ECN.
57. IAEA-Tecdoc 626, 'Guidelines for Leaching Studies on Coal Fly Ash and Other Solid Wastes', Vienna, July 1991.
58. H.A. Vos, 'The Speciation of Radionuclides in Sediment and Soils', Report ARL TRO 58, 1981, Yallambie, Victoria 3085.
59. I. Neretnieks, *Diffusivities of Some Dissolved Constituents in Compacted Wet Bentonite Clay*, Royal Institute of Technology, Stockholm, Sweden, 1982.
60. A. van Dalen, J. Wijkstra, 'Distribution and Diffusion Measurements in Deep Ocean Sediment Samples', *Radiochim. Acta*, **38**(3), 141–144 (1985).
61. A. van Dalen, J. Wijkstra, 'Neptunium and Americium Distribution and Diffusion Measurements in Deep-ocean Sediment Samples', *J. Less-Common Met.*, **122**, 563–567 (1986).
62. B. Törstenfelt, 'Migration of the Fission Products Strontium, Technetium, Iodine and Cesium in Clay', *Radiochim. Acta*, **39**(2), 97–104 (1986).
63. P. Vilks, D.J. Drew, 'The Effect of Colloids on Actinide Migration', in *Proceedings of 2nd International Conference of Radioactive Waste Management*, Canadian Nuclear Society, Toronto, 667–673, 1986.
64. P. Benes, 'Speciation Procedures', in *The Environmental Behavior of Radium*, IAEA, Vienna, 279–283, 1990.
65. G. Riise, H.E. Bjoernstad, H.N. Lien, D.H. Oughton, B. Salbu, 'A Study on Radionuclide Association with Soil Components Using a Sequential Extraction Procedure', *J. Radioanal. Nucl. Chem.*, **142**(2), 531–538 (1990).
66. B. Salbu, 'Preconcentration and Fractionation Techniques in the Determination of Trace Elements in Natural Waters, their Concentration and Physico-chemical Form', University of Oslo, Norway, 1984.
67. 'Materials Research Society Symposium Proceedings: Scientific Basis for Nuclear Waste Management', MRS, Pittsburgh, Pennsylvania, 1982.
68. J. Smith, J. Nelissen, R. van Griekeren, 'Comparison of Preconcentration Procedures for Trace Metals in Natural Waters', *Anal. Chim. Acta*, **111**, 215–226 (1979).
69. T.M. Florence, 'The Speciation of Trace Elements in Waters', *Talanta*, **29**, 345–364 (1982).
70. G. Tölg, 'Extreme Trace Analysis of the Elements', *Talanta*, **19**, 1409–1521 (1972).
71. H.A. van der Sloot et al., 'Comparison of Different Regulatory Leaching Test Procedures for Waste Materials', Report ECN-C-91-082. Petten (NH), The Netherlands, 1991.
72. A. Tessier, P.G.C. Campbell, M. Bisson, 'Sequential Extraction Procedure for the Speciation of Particulate Trace Metals', *Anal. Chem.*, **51**(7), 844–851 (1979).
73. V. Spevackova, J. Kucera, 'Trace Element Speciation in Contaminated Soils Studied by Atomic Absorption Spectrometry and Neutron Activation Analysis', *Int. J. Environ. Anal. Chem.*, **35**(4), 241–251 (1989).
74. R. Chester, F.J. Lin, K.J.T. Murphy, 'A Three Stage Sequential Leaching Scheme for the Characterization of the Sources and Environmental Mobility of Trace Metals in the Marine Aerosol', *Environ. Technol. Lett.*, **10**(10), 887–900 (1989).
75. A. Bermond, G. Sommer, 'Simulation of Heavy Metals Extraction in Soil Samples Compared with Experimental Results', *Environ. Technol. Lett.*, **10**(11), 989–994 (1989).
76. P.M.V. Nirel, F.M.M. Morel, 'Pitfalls of Sequential Extractions', *Water. Res.*, **24**(8), 1055–1056 (1990).
77. M. Sager, R. Belocky, R. Pucsko, 'On the Determination of the Speciation of Major and Trace Elements in Sediments by Sequential Solution Techniques', *Acta Hydrochim. Hydrobiol.*, **18**(2), 157–173 (1990).
78. J.M. Martin, P. Nirel, A.J. Thomas, 'Sequential Extraction Techniques, Promises and Problems', *Mar. Chem.*, **22**, 313–341 (1987).
79. M. Kersten, U. Foerstner, 'Effect of Sample Pretreatment on the Reliability of Solid Speciation Data of Heavy Waters', *Mar. Chem.*, **22**, 299–312 (1987).
80. H. Umeki, A. Suzuki, R. Kiyose, 'A Leach Model For Safety Assessment', *Adv. Ceramics*, **20**, 523–529 (1986).
81. D.C. Adriano (ed.), *Biogeochemistry of Trace Metals*, Lewis Publishers, Boca Raton, 1992.
82. J.J.W. Higgo, 'Iodine Speciation and Diffusion in a Sand–Groundwater System', Report EUR 13277 EN, 1991.
83. J. Crank, *The Mathematics of Diffusion*, Clarendon Press, Oxford, 1975.
84. D. Haigh, 'The Groundwater Chemistry of the BGS In Situ Migration Experiment', British Geological Survey, Report WE/08/16.
85. B. Salbu, 'Size Fractionation Techniques Combined with INAA for Speciation Purposes', *J. Radioanal. Nucl. Chem.*, **112**(1), 169–174 (1987).

86. H.A. Das, W.J.M. Hegeman, C.H. van der Weyden, 'Radiotracer Measurement of the Volatalization of Organic Trace Constituents from Water', *Isotopes Environ. Health Stud.*, **32**, 341–346 (1996).
87. C.J.M. Kramey, 'Approaches in Chemical Speciation Studies in Marine Waters', *Elsevier Oceangr. Series*, **43**, 3–36 (1986).
88. J.A. Cox, K. Slonawska, D.K. Gatchell, A.G. Hiebert, 'Metal Speciation by Donnan Dialysis', *Anal. Chem.*, **56**, 650–657 (1984).
89. R.B. Bird, W.E. Stewart, E.N. Lightfoot, *Transport Phenomena*, J. Wiley & Sons, New York, 1960.

Rutherford Backscattering Spectroscopy

Kenji Kimura
Kyoto University, Kyoto, Japan

1 Introduction **1**
2 Fundamentals **2**
2.1 Kinematics of Ion–Atom Collisions 2
2.2 Rutherford Cross-section 3
2.3 Non-Rutherford Cross-sections 4
2.4 Stopping Power 5
2.5 Energy-loss Straggling 6
3 Standard Rutherford Backscattering Spectroscopy **6**
3.1 Apparatus 6
3.2 Thin Film Analysis 7
3.3 Depth Profiling 7
3.4 Computer Simulation 8
3.5 Channeling Effect 8
3.6 Comparison to Related Techniques 9
4 Advanced Techniques **10**
4.1 Heavy Ion Rutherford Backscattering Spectroscopy 10
4.2 High-resolution Rutherford Backscattering Spectroscopy 11
4.3 Microprobe Rutherford Backscattering Spectroscopy 12
4.4 Light Element Analysis by Non-Rutherford Scattering 12
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

Rutherford backscattering spectroscopy (RBS) is one of the most powerful techniques for measuring elemental depth profiles. It allows quantitative and nondestructive analysis with a reasonable depth resolution. In conventional RBS (combination of 1–4 MeV He ions with a silicon surface-barrier detector), the typical depth resolution is about 10 nm and the typical mass resolution is about 1 u for light elements (mass number $M < 20$) and about 20 u for heavy elements (M ca. 200). Sensitivity of 100 ppm can be easily obtained for heavy elements in light hosts although analysis of light elements in heavy hosts is difficult. In the present article, physical concepts which RBS relies on are briefly introduced. Some examples of analyses are demonstrated and several advanced techniques which have been developed to improve capabilities of the conventional RBS are discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The analysis of the surface and near-surface region of solids has become an important field of science and technology. RBS is one of the most powerful methods among a number of analyzing techniques. In RBS, high-energy ion beams, usually H or He ions with energies in the range 1–4 MeV, are used as probes. A sample is irradiated by the ion beam. Almost all ions penetrate deep inside the sample up to ca. 10 μm until they entirely lose their kinetic energy. During the penetration some ions collide with the target atoms and are subject to elastic Coulomb scattering (Rutherford scattering) between the projectile and the target nuclei and may be backscattered from the sample. The energy of the ion backscattered from the target atom depends on the target atom mass. This allows compositional analysis of a surface region of several micrometers by measuring the energy spectrum of the backscattered ions. Figure 1 illustrates a schematic set-up of RBS and an example of an energy spectrum for a two-element thin film on a low-mass substrate. The ions scattered from each element form a separated peak. The number of target atoms in the film can be derived from the peak yields and the peak width gives the film thickness. Table 1 summarizes the features of standard RBS and can be used as a rough guide.

The method was first employed by nuclear physicists to analyze their targets in 1951.[1] Extensive application to material science began in the late 1960s. The technique has been refined and now constitutes one of the most common methods for compositional analysis of the surface region. RBS has many outstanding features:

1. RBS is a nondestructive method of analysis.
2. It allows quantitative analysis of all elements simultaneously except for hydrogen.
3. RBS has a depth resolution typically of about 10 nm, which can be improved up to monolayer resolution in the surface region with special equipment.
4. The measurement time is typically as short as several minutes.
5. Sensitivity of RBS is about 100 ppm for heavy elements although it is not so good for light elements.

RBS relies on the following physical concepts:

1. The kinematic factor of the elastic scattering which describes the reduction of incident energy in a collision between the probe ion and the target atom.

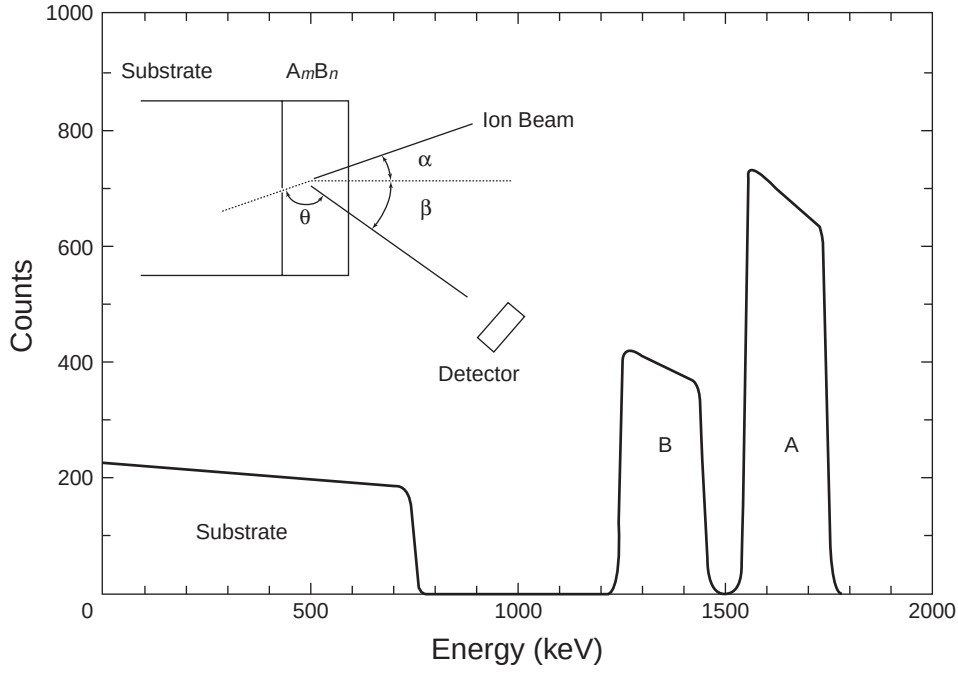

Figure 1 RBS spectrum for a two-element compound film (A_mB_n) on a low-mass substrate. The inset shows the schematic set-up of RBS.

Table 1 Important parameters and capabilities of standard RBS

Probe ion	H, He
Ion energy	1–4 MeV
Beam current	1–100 nA
Detector	SSB detector
Depth resolution	ca. 10 nm (easily improved to 2 nm using a grazing-angle technique)
Range	ca. 10 μm
Element range	>1 for H ion >4 for He ion
Mass resolution	1 u for light elements ($M_2 < 40$ u) with 2 MeV He ion 20 u for heavy elements (M_2 ca. 200 u) with 2 MeV He ion
Sensitivity	ca. 100 ppm for heavy elements in light hosts

SSB, silicon surface barrier.

The resulting energy of the scattered ion increases with target atom mass. This allows identification of the target atom by measuring the scattered ion energy.

2. The differential scattering cross-section which gives the probability of scattering. This allows basic quantitative analysis without a standard sample.
3. The stopping power which is defined by the energy loss of the ion per unit path length inside the target. The energy of the backscattered ion depends on the depth from which the ion was scattered because the path length is proportional to the depth. This allows the depth profiling of elements in the target.
4. The energy-loss straggling which is the fluctuation of the energy loss arising from the statistical feature of the energy-loss process. This determines the intrinsic depth resolution.

There are numerous reviews on RBS[2–4] and for the state of the art reader can refer to relevant conference proceedings such as ion beam analysis (IBA), Conference on the Application of Accelerators in Research and Industry (CAARI), European Conference on Accelerator Applications in Research and Technology (ECAART) and Nuclear Microprobe Technology and Applications (NMTA). Most have been published in *Nuclear Instruments and Methods* in the Physics Research section B.[5]

2 FUNDAMENTALS

2.1 Kinematics of Ion–Atom Collisions

When an ion is scattered from a target atom at an angle θ, the ratio of the scattered-ion energy E to the incident energy E_0 can be calculated using the laws of conservation of energy and momentum, Equation (1)

$$K = \frac{E}{E_0} = \frac{(M_2^2 - M_1^2 \sin^2\theta)^{1/2} + M_1 \cos\theta}{M_1 + M_2}^2 \qquad (1)$$

where M_1 and M_2 are masses of the incident ion and the target atom, respectively. The binding energy of the target atom and its thermal vibration are neglected in the derivation of Equation (1). These effects are negligibly small in collisions between ions with megaelectron volt energy and target atoms. It is clear that the energy ratio K, called the kinematic factor, is determined by the mass ratio M_2/M_1 and the scattering angle θ. The basic idea of RBS relies on this equation. Figure 2 shows the kinematic factor as a function of the mass ratio M_2/M_1. The kinematic factor, and thus the energy of the scattered ion, increases monotonically with the mass ratio indicating that the target atom mass can be determined from the observed energy of the scattered ion.

The mass resolution is defined by Equation (2)

$$\delta M_2 = \frac{\delta E}{E_0} \left(\frac{\mathrm{d}K}{\mathrm{d}M_2} \right)^{-1} \tag{2}$$

where δE is the overall energy resolution of the RBS system. The dominant contribution to δE comes from the energy resolution of a SSB detector which is commonly used in RBS. Figure 3 displays the calculated mass resolution for 2-MeV H, He, and C ions, where the typical energy resolutions of the SSB detector (i.e. $\delta E = 10$, 15, and 50 keV for H, He, and C ions, respectively) are used for δE. The results at different scattering angles are also shown for 2-MeV He ions. A mass difference of 1 u can be resolved for light elements ($M_2 < 40$ u) with 2-MeV He ions. The resolution, however, becomes worse very rapidly with increasing M_2 and reaches δM_2 ca. 20 u at M_2 ca. 200 u.

For better mass resolution, heavy ions are usually employed. Although the energy resolution of the SSB detector is worse for heavy ions, the resulting mass resolution for C ions is better than for He ions and $\delta M_2 =$

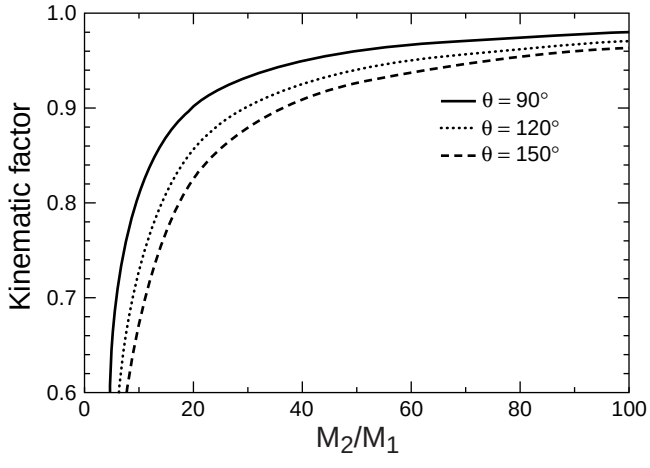

Figure 2 Kinematic factor K as a function of the ratio of the target mass to the ion mass M_2/M_1.

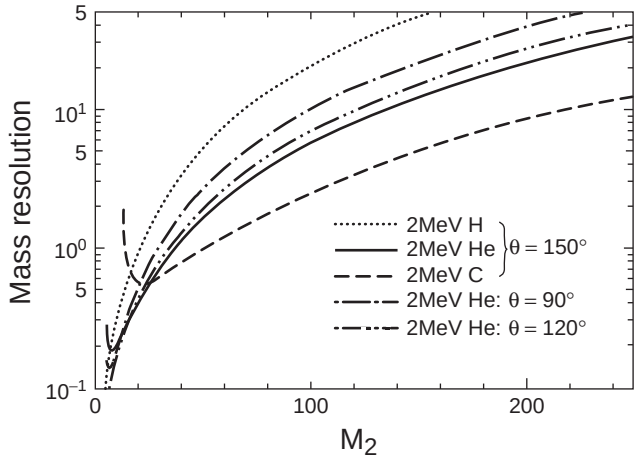

Figure 3 Mass resolution for 2-MeV H, He and C ion beams as a function of target mass M_2. Typical energy resolution of the SSB detector is used in the calculation.

1 u is achieved up to M_2 ca. 50 u with 2-MeV C ions. The resolution can be improved considerably using the time-of-flight (TOF) technique for energy analysis instead of the SSB detector in heavy ion RBS (HIRBS). The details of the heavy ion RBS will be discussed in section 4.1.

2.2 Rutherford Cross-section

The scattering cross-section σ is an effective area associated with the colliding particles. When projectile ions impinge on a thin target film of thickness t, the scattering yield is given by Equation (3)

$$Y_{\mathrm{tot}} = \frac{Nt\sigma Q}{\cos\alpha} \tag{3}$$

where N is the volume density of the target atom, Q the number of incident ions, and α is the incident angle (the definition of α is shown in Figure 1). In RBS, the ions scattered at a particular scattering angle θ are measured by a detector with a finite acceptance solid angle $\Delta\Omega$. The number of these ions can be calculated (Equation 4) with an angular differential cross-section $\mathrm{d}\sigma/\mathrm{d}\Omega$

$$Y(\theta) = Nt \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \frac{\Delta\Omega Q}{\cos\alpha} \tag{4}$$

The differential cross-section for Coulomb scattering (Rutherford cross-section) is given by an analytical formula (Rutherford formula), Equation (5)

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \right)_{\mathrm{R}} = \left(\frac{Z_1 Z_2 e^2}{4E} \right)^2 \times \frac{4 \left\{ \sqrt{(M_2^2 - M_1^2 \sin^2\theta)} + M_2 \cos\theta \right\}^2}{M_2 \sin^4\theta \sqrt{M_2^2 - M_1^2 \sin^2\theta}} \tag{5}$$

where Z_1, Z_2 are atomic numbers of incident ion and target atom, respectively and E is the incident energy. This simple formula guarantees to extract quantitative data about abundances of elements from RBS spectra. The cross-section is proportional to Z_2^2, indicating high sensitivity for heavy elements in RBS. A useful number in evaluating the Rutherford cross-section is $e^2 = 14.4\,\text{eV Å}$ ($1\,\text{Å} = 10^{-10}\,\text{m}$). For example, 1 MeV He^+ ions of total dose 1 μC are incident on a Si film of $t = 1000\,\text{Å}$ at $\alpha = 0°$. The differential cross-section is calculated to be $(d\sigma/d\Omega)_R = 1.126 \times 10^{-8}\,\text{Å}^2/sr$ at $\theta = 150°$. Using a typical acceptance angle of SSB detectors, $\Delta\Omega = 5\,\text{msr}$ the yield is given by Equation (6)

$$Y = 4.994 \times 10^{-2}\,\text{Å}^{-3} \times 1000\,\text{Å} \times 1.126 \times 10^{-8}\,\text{Å}^2/sr^{-1} \times 5 \times 10^{-3} sr \times \frac{1 \times 10^{-6}}{1.602 \times 10^{-19}} \approx 17\,500 \quad (6)$$

The associated statistical error is $1/\sqrt{17\,500} = 0.8\%$. This means the thickness of the film can be determined with an accuracy less than 1% in this case.

2.3 Non-Rutherford Cross-sections

The Rutherford cross-section can be used in situations where the incident ion energy is in the order of $1\,\text{MeV u}^{-1}$. The cross-section deviates from the Rutherford formula in both high-energy and low-energy regions. On the low-energy side, the screening effect due to the bound electrons in the target atom reduces the cross-section from the Rutherford formula. The cross-section cannot be given by a simple analytical formula like Equation (5). The screened Coulomb potential is given by Equation (7)

$$V(r) = \frac{Z_1 Z_2 e^2}{r}\phi \quad \frac{r}{a} \quad (7)$$

where ϕ is the screening function and a is the screening length. There are several analytical formulas for the screening function given by a sum of exponential terms, Equation (8)

$$\Phi(z) = \sum_i^n \alpha_i \, e^{-\beta_i z} \quad (8)$$

where $\{\alpha_i\} = \{0.1, 0.55, 0.35\}$, $\{\beta_i\} = \{6.0, 1.2, 0.3\}$ for Molière's screening function[6] ($a = 0.8853 a_B (Z_1^{1/2} + Z_2^{1/2})^{-2/3}$ is commonly used) and $\{\alpha_i\} = \{0.181\,75, 0.509\,86, 0.280\,22, 0.028\,171\}$, $\{\beta_i\} = \{3.1998, 0.942\,29, 0.402\,90, 0.201\,62\}$ for Ziegler–Biersack–Littmark's screening function ($a = 0.8853 a_B (Z_1^{0.23} + Z_2^{0.23})^{-1}$ is used).[7]

Using the screening function, the differential cross-section can be calculated numerically. Algorithms for rapid computations are proposed by Mendenhall and Weller.[8] A useful analytical formula for the screening correction valid for large scattering angles is given by L'Ecuyer et al.,[9] Equation (9)

$$\frac{d\sigma/d\Omega}{(d\sigma/d\Omega)_R} = 1 - \frac{0.049 Z_1 Z_2^{4/3}}{E_{CM}} \quad (9)$$

where E_{CM} is the center-of-mass kinetic energy in kilo-electron volts. Examples of the results calculated with Equation (9) together with the result of exact numerical calculation are shown in Figure 4. The deviation from the Rutherford cross-section is large for heavy elements and the correction formula for Equation (9) is reasonably good for both heavy and light elements. Although the deviation for 1-MeV He–Au scattering is about 4% and the Rutherford formula may not be used in a precise measurement, Equation (9) can be safely used even in this situation. It must, however, be noted that Equation (9) does not contain the scattering angle θ and is valid only for large θ. At smaller scattering angles, the screening effect becomes large.

On the high-energy side, the incident ion may approach to the target nuclei within the range of the nuclear force. This causes deviation from the Rutherford cross-section. The deviation is large for light elements because the Coulomb barrier of the nucleus is low for light elements. An empirical formula to estimate the threshold energy, above which the deviation from the Rutherford cross-section for $160° < \theta < 180°$ is larger than 4%, is given by Leavitt and McIntyre,[10] Equations (10–12)

$$E_{th} = (0.12 \pm 0.01)Z_2 - (0.5 \pm 0.1)\,\text{MeV for H} \quad (10)$$

$$E_{th} = (0.25 \pm 0.01)Z_2 + (0.4 \pm 0.2)\,\text{MeV for He} \quad (11)$$

$$E_{th} = (0.330 \pm 0.005)Z_2 + (1.4 \pm 0.1)\,\text{MeV for Li} \quad (12)$$

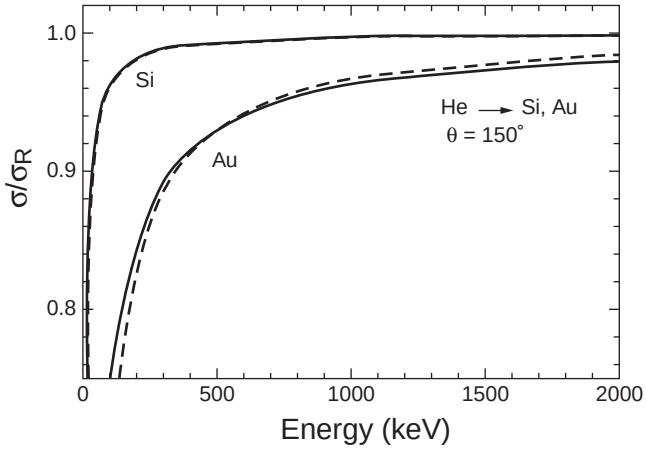

Figure 4 Screening correction for the scattering cross-section as a function of the incident ion energy. The results of the numerical calculation (solid lines) as well as those of the analytical formula (dashed lines) are shown for He–Si and He–Au scattering.

In contrast to the correction for screening effect, no practical methods exist for rapid accurate estimation of the nuclear force correction. The cross-section depends on both the ion energy and the scattering angle in a complicated manner (e.g. Figure 14). Nevertheless, high-energy ions are useful in some cases, especially in analysis of light elements. The standard RBS is less sensitive for light elements due to the small Rutherford cross-sections of light elements (the cross-section is proportional to the square of the target atomic number). The non-Rutherford cross-sections for light elements are frequently enhanced by one or two orders of magnitude while those for heavy elements may remain as a Rutherford cross-section. This improves the sensitivity for light elements. The details will be discussed in section 4.4.

2.4 Stopping Power

An energetic ion passing through a solid loses its energy primarily through excitation of electrons. The rate of the energy loss per unit path length, $\mathrm{d}E/\mathrm{d}x$, is called stopping power and understanding the stopping power has been a subject of extensive studies. Using the first-order perturbation theory, the stopping power is given by[11] Equation (13)

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{4\pi Z_1^2 Z_2 N e^4}{mV^2} \ln\left(\frac{mV^2}{I}\right. \tag{13}$$

where m is the electron mass, V the velocity of the ion, and I ($\approx 10 \times Z_2$ eV) is the mean excitation energy of the target electrons. This equation, called the Bethe–Bloch formula, is used as a guide in semi-empirical fitting of stopping power data to provide semi-empirical formulas. A number of tables are available to calculate stopping powers reasonably accurately.[7,12–14] Figure 5 shows examples of the stopping powers of Si and Au for H and He ions calculated with the tabulated values given by Andersen and Ziegler.[12,13] The stopping power has a broad maximum around 1 MeV for He ions and around 100 keV for H ions. The maximum energy and the maximum stopping power increases with both projectile and target atomic number. The stopping power for megaelectron-volt He ions is several times larger than that for megaelectron-volt H ions. This indicates that better depth resolution is available for He ions.

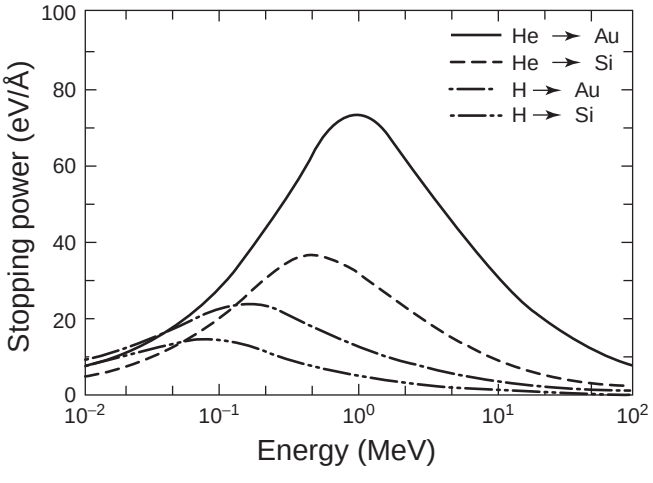

Figure 5 Stopping powers of Si and Au for H and He ions as a function of the ion energy.

In order to estimate the stopping power of a compound, it is convenient to introduce stopping cross-section defined by Equation (14)

$$\varepsilon = \frac{1}{N}\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right. \tag{14}$$

The stopping power of a compound $\mathrm{A}_m\mathrm{B}_n$ can be calculated using Bragg's rule, which is based on the simple assumption that the constituent elements act independently in the energy loss process, Equation (15)

$$\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right._{\mathrm{A}_m\mathrm{B}_n} = N_{\mathrm{mol}}[m\varepsilon(\mathrm{A}) + n\varepsilon(\mathrm{B})] \tag{15}$$

where $\varepsilon(\mathrm{A})$ and $\varepsilon(\mathrm{B})$ are the stopping cross-sections of the constituents and N_{mol} is the molecular density. This neglects the effect of chemical bonding but gives reasonably accurate values for megaelectron-volt light ions.

The relationship between the energy of the scattered ion and the depth at which the ion is scattered can be calculated using the stopping power, the kinematic factor, and the geometry of the scattering. Neglecting the variation of the stopping power along the ion path, the energy of the ion scattered from a target atom at a depth t is given by Equation (16)

$$E(t) = K \quad E_0 - \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right._{\mathrm{in}} \frac{t}{\cos\alpha} \quad - \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right._{\mathrm{out}} \frac{t}{\cos\beta} \tag{16}$$

where the subscripts in and out refer to the inward path and outward path, respectively, and α and β are the incident and exit angles, respectively. For a thin surface layer, the variation of the stopping power is small and the use of an average value is a good approximation. The variation in energy δE corresponds to a variation in depth δt through Equation (17)

$$\delta E = -[S]\delta t \tag{17}$$

where $[S]$ is commonly called the energy-loss factor and is given by Equation (18)

$$[S] = \quad K\frac{(\mathrm{d}E/\mathrm{d}x)_{\mathrm{in}}}{\cos\alpha} + \frac{(\mathrm{d}E/\mathrm{d}x)_{\mathrm{out}}}{\cos\beta} \tag{18}$$

The depth resolution depends on the overall energy resolution, the stopping power, and the geometry. The energy resolution is normally composed of three contributions: the energy spread of the incident beam δE_{in}, the detector resolution δE_d, and the energy loss straggling δE_{s} (the last of these, δE_{s} will be discussed in detail in the next section). Assuming these contributions are independent, the overall resolution is given by Equation (19)

$$(\delta E)^2 = (\delta E_{\text{in}})^2 + (\delta E_{\text{d}})^2 + (\delta E_{\text{s}})^2 \tag{19}$$

For example, in the case of 2-MeV He on Au at $\alpha = 0°$ and $\beta = 30°$ ($\theta = 150°$), the energy-loss factor is calculated to be 135 eVÅ^{-1}. Using a typical energy resolution in the standard RBS, $\delta E = 15$ keV and a depth resolution of 11 nm is obtained. The depth resolution can be improved up to ca. 2 nm simply by using a grazing-angle technique, usually with a grazing exit angle, for example β ca. 85°. In the grazing-angle geometry, the acceptance angle of the SSB detector should be much smaller than the grazing angle. Otherwise the distribution of β and with it the distribution of $[S]$ makes the depth resolution worse.

2.5 Energy-loss Straggling

The energy-loss process described in the previous section has a stochastic feature. The slowing down of the ion is accompanied by energy spreading which is called energy-loss straggling. For long path length, the energy distribution approaches a Gaussian distribution. The standard deviation Ω of the distribution for the high-energy limit was given by Bohr,[15] Equation (20)

$$\Omega_{\text{B}}^2 = 4\pi Z_1^2 Z_2 e^4 N t \tag{20}$$

and has been extended to lower energies by many authors. A useful analytical formula was given by Lindhard and Scharff,[16] Equation (21)

$$\frac{\Omega^2}{\Omega_{\text{B}}^2} = L(\chi) = \begin{cases} 0.68\chi^{1/2} - 0.08\chi^{3/2} & \chi \leq 3 \\ 1 & \chi \geq 3 \end{cases} \tag{21}$$

where $\chi = (V/v_{\text{B}})^2/Z_2$, V is the ion velocity, and v_{B} ($= 2.2 \times 10^8$ cm s^{-1}) the Bohr velocity. More refined models were developed by Bonderup and Hvelplund,[17] Chu,[18] and Besenbacher et al.[19]

For compounds, an additive rule similar to Bragg's rule was proposed,[18] Equation (22)

$$\frac{\Omega^2(\text{A}_m\text{B}_n)}{N_{\text{mol}} t} = \frac{m\Omega^2(\text{A})}{N_{\text{A}} t} + \frac{n\Omega^2(\text{B})}{N_{\text{B}} t} \tag{22}$$

The contribution of the energy-loss straggling to the energy resolution is given by Equation (23)

$$(\delta E_{\text{s}})^2 = 8 \ln 2 \left(K^2 \Omega_{\text{in}}^2 + \Omega_{\text{out}}^2\right) \tag{23}$$

where $8 \ln 2$ is a conversion factor from the standard deviation to the full width at half maximum, and Ω_{in} and Ω_{out} are the straggling in inward and outward paths. For megaelectron-volt H and He ions incident on layers $\leq$100 nm, the straggling is small (e.g. Ω_{B} ca. 2.7 keV for He–Si at $t = 100$ nm) compared to the resolution of the SSB detector (δE_d ca. 15 keV for megaelectron-volt He) and can be neglected in the obtainable depth resolution in the standard RBS. The contribution of the energy-loss straggling, however, increases with $t^{1/2}$ and the straggling sets a fundamental limit to the depth resolution at deeper layers.

3 STANDARD RUTHERFORD BACKSCATTERING SPECTROSCOPY

3.1 Apparatus

In the standard RBS, 1–4-MeV He ion beams and a SSB detector are commonly used. Although both the mass and depth resolutions are poorer for H ions than for He ions, H ions are sometimes used in RBS in order to analyze deeper regions. Ions generated by an ion source are accelerated by an electric field. The most widely used devices for the production of a megaelectron-volt ion beam are the Van de Graaff-type and Cockcroft- and Walton-type accelerators. The accelerated ions are analyzed by a magnet to eliminate contaminant ion species and to select the ion energy. The calibration of the analyzing magnet is usually done using resonant nuclear reactions. After passing the analyzing magnet, the ion beam is collimated by apertures to a size of ca. 1 mm and introduced to a scattering chamber where targets are mounted on a manipulator. The beam current is usually monitored by target current or a beam monitor system installed between the final aperture and the target. The typical beam current is of the order of 10 nA. For a channeling measurement, the beam divergence angle and the precision of the manipulator should be less than 1 mrad.

A SSB detector is installed inside the chamber to measure the energy spectrum of scattered ions. The SSB detector is basically a Schottky barrier diode. A reverse bias is applied to make a depletion region of the order of 100 μm which is greater than the range of the mega-electron-volt light ions. The electron-hole pairs created by the ion in the depletion region are swept by the high electric field and collected. The quantity of charge collected bears a linear relationship to the total energy spent to create electron–hole pairs (about 3.6–3.7 eV per electron–hole pair). The linearity between the collected charge and the ion energy is quite good for megaelectron-volt light ions. The signal from the detector is amplified

by a preamplifier and a linear amplifier and sent to a pulse-height analyzer (PHA). The PHA stores pulses of a given height in a given channel. The relation between the channel number and the ion energy can be calibrated by measuring ions of known energy, for example α particles from ^{241}Am. It should be noted that the energy lost in the entrance window as well as the energy lost to nuclear scattering do not contribute to the pulse height. The equivalent thickness of the window is about 100 nm, which corresponds to about 30 keV loss for 1-MeV He. These effects are serious for heavy ions and must be taken into account in the conversion procedure from pulse height to ion energy, especially in the low-energy region.

3.2 Thin Film Analysis

An example of RBS spectrum for a high-T_c superconductor thin film deposited on a Si substrate is depicted in Figure 6.[20] The film was prepared by ion beam sputtering of $Bi_2Sr_2CaCu_2O_7$. The peaks for Bi, Sr, Cu, and Ca are well separated and the oxygen peak can be clearly seen on top of the Si signal. The integrated counts Y_i for each element can be accurately determined. The areal density of element i is obtained from Equation (4) as

$$n_i = N_i t = \frac{Y_i \cos\alpha}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_i \Delta\Omega Q} \tag{24}$$

Although precise measurements of the ion energy E, the scattering angle θ, the detector acceptance angle $\Delta\Omega$, and the incident ion number Q are necessary to estimate the absolute areal density n_i, an average stoichiometric ratio can be obtained without these values, Equation (25)

$$\frac{N_i}{N_j} = \frac{Y_i}{Y_j}\frac{(\mathrm{d}\sigma/\mathrm{d}\Omega)_j}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_i} \approx \frac{Y_i}{Y_j}\left(\frac{Z_j}{Z_i}\right)^2 \tag{25}$$

The composition of the film shown in Figure 6 can be determined to be $Bi_1Sr_{1.6}Ca_{0.9}Cu_{1.4}O_7$ using this equation.

In order to estimate the absolute areal density n_i using Equation (24), a standard sample implanted with a heavy element of a known areal density N_{ref} is sometimes used, Equation (26)

$$n_i = N_i t = \frac{Y_i(\mathrm{d}\sigma/\mathrm{d}\Omega)_{\mathrm{ref}}}{Y_{\mathrm{ref}}(\mathrm{d}\sigma/\mathrm{d}\Omega)_i} N_{\mathrm{ref}} \approx \frac{Y_i}{Y_{\mathrm{ref}}}\left(\frac{Z_{\mathrm{ref}}}{Z_i}\right)^2 N_{\mathrm{ref}} \tag{26}$$

where Y_{ref} is the yield of the standard sample measured under the same conditions. An alternative conventional method to estimate the absolute areal density is to make use of the signal of substrate material, Equation (27)

$$n_i = N_i t = \frac{Y_i}{y_s}\frac{\delta E_{\mathrm{ch}}}{[S]_S} N_s \frac{(\mathrm{d}\sigma/\mathrm{d}\Omega)_s}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_i} \approx \frac{Y_i}{y_s}\frac{\delta E_{\mathrm{ch}}}{[S]_S} N_s \left(\frac{Z_s}{Z_i}\right)^2 \tag{27}$$

where y_s is the height of the energy spectrum at the leading edge of the substrate, δE_{ch} the corresponding energy for one channel, N_s the atomic density of the substrate, $(\mathrm{d}\sigma/\mathrm{d}\Omega)_s$ the differential cross-section of the substrate atom and $[S]_S$ the energy-loss factor for the substrate.

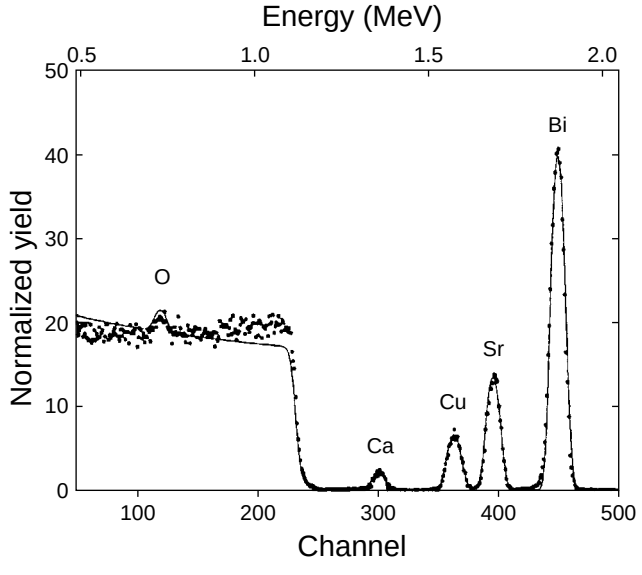

Figure 6 RBS spectrum of a high-T_c superconductor thin film $Bi_1Sr_{1.6}Ca_{0.9}Cu_{1.4}O_7$ deposited on a Si substrate. Signals of Bi, Sr, Cu, and Ca are seen as separated peaks. (Reproduced by permission of Elsevier Science B.V. from Climent-Font et al.[20])

3.3 Depth Profiling

When the stoichiometry changes with depth, the variation of the composition can be determined from the RBS spectrum. The stoichiometry at a depth t can be derived from the corresponding yields of all compositional elements at t, Equation (28). In Equation (28) F_i is the atomic fraction of element i at a depth t.

$$F_i = \frac{y_i[S]_i \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_i}{\sum_j y_j[S]_j \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_j} \approx \frac{y_i[S]_i / Z_i^2}{\sum_j y_j[S]_j / Z_j^2} \tag{28}$$

This procedure is not simple because the energy-loss factor $[S]_i$ is a function of the composition. For dilute concentrations of an impurity, the stopping power is simply determined by the matrix and the depth profiling

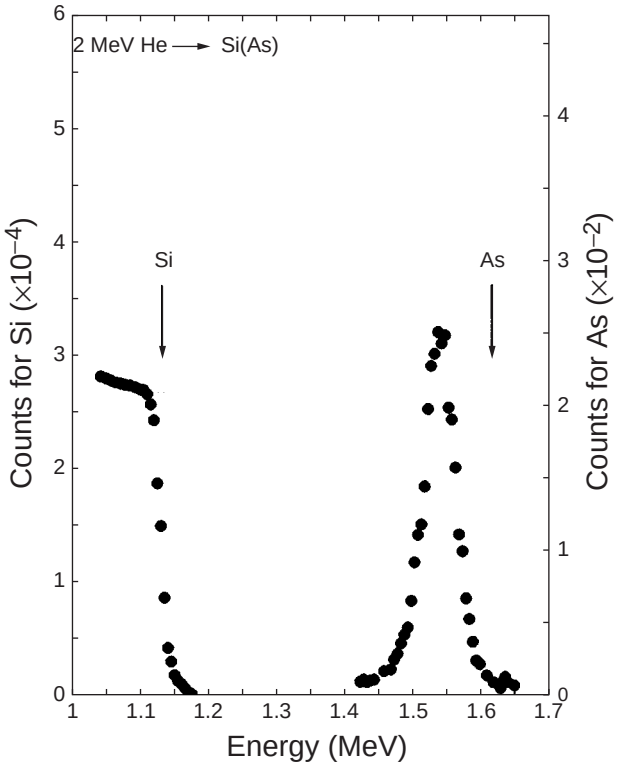

Figure 7 RBS spectrum of 2-MeV He ions backscattered from Si implanted with 1.2×10^{15} As ions cm^{-2} at 200 keV. The arrows indicate the energy of ions scattered from surface Si and As atoms. (Reproduced by permission of Plenum Press from Mayer et al.[2])

is straightforward. Figure 7 shows an example of such an RBS spectrum for Si implanted with 1.2×10^{15} As ions cm^{-2} at 200 keV measured with 2-MeV He ions.[2] There is a separated peak of As at ca. 1.55 MeV which corresponds to t ca. 150 nm. The width of the As peak at ca. 60 keV is much larger than the energy resolution of the SSB detector showing that the observed As peak profile reflects the actual As distribution in Si. The energy-loss factors for Si and As are calculated to be 44.3 and 46.0 eV Å^{-1} respectively. Using the peak yield of As, $y_{As} = 250$ and the Si yield at the corresponding depth, $y_{Si} = 28\,000$, the peak concentration can be calculated from Equation (28) as Equation (29)

$$F_{As} \approx \frac{250 \times 46/33^2}{(250 \times 46/33^2 + 28\,000 \times 44.3/14^2)} = 0.0017 \tag{29}$$

Note that the sensitivity of RBS can easily reach 100 ppm level for heavy elements.

The RBS spectrum shown in Figure 7 also provides a good example of how to estimate areal density of the implanted atoms using Equation (27). Using the total yield of the As peak $Y_{As} = 3350$ and $\delta E_{ch} = 5$ keV, the areal density of As is given by Equation (30)

$$n_{As} \approx \frac{3350}{28\,000} \frac{5\,\text{keV}}{44.3\,\text{eV Å}^{-1}} 4.994 \times 10^{-2}\,\text{Å}^{-3} \left(\frac{14}{33}\right)^2 = 0.12\,\text{Å}^{-2} \tag{30}$$

The value obtained agrees well with the number of implanted As ions measured in the implantation procedure. This good agreement clearly shows the capability for quantitative analysis in RBS.

3.4 Computer Simulation

When the signals from different elements overlap in a spectrum, derivation of the target composition as a function of depth is not straightforward although a basic derivation can be obtained using Equation (28). A powerful approach to analyzing RBS spectra is by computer simulation. There are a number of computer simulation programs to analyze the RBS spectrum such as RUMP,[21] GISA,[22] and SCATT/HYPRA.[23] In these programs, the target composition is changed until the calculated spectrum matches the measured spectrum. As with other inverse problems, this procedure may obtain a local minimum. Recently, a simulated annealing algorithm was successfully applied to the analysis of RBS spectra.[24] It tends asymptotically to the absolute minimum rather than to a local minimum with any initial compositions. This allows fully automatic analysis without time-consuming human intervention. Figure 8 shows examples of the different steps of the fit to a RBS spectrum of a multilayered iron–cobalt silicide generated by the simulated annealing algorithm. The final result agrees reasonably well with the original spectrum with a simple initial composition (Figure 8a) generated automatically by the program.

3.5 Channeling Effect

The energy spectrum from a single crystal aligned with a low-index axis or plane parallel to the incident beam direction is very different from those discussed before. The ion experiences a series of correlated small angle collisions with crystal atoms and cannot approach to the atomic row or plane over a certain distance. This phenomenon, called channeling, drastically reduces the yield of Rutherford scattering. The analyzing procedures described in the previous sections cannot be applied to the RBS spectra measured under channeling conditions. In a simple RBS measurement, a single crystal target is usually rotated during measurement to avoid channeling effects.

Although the channeling is undesirable for RBS, it provides a unique method for studying lattice defects.

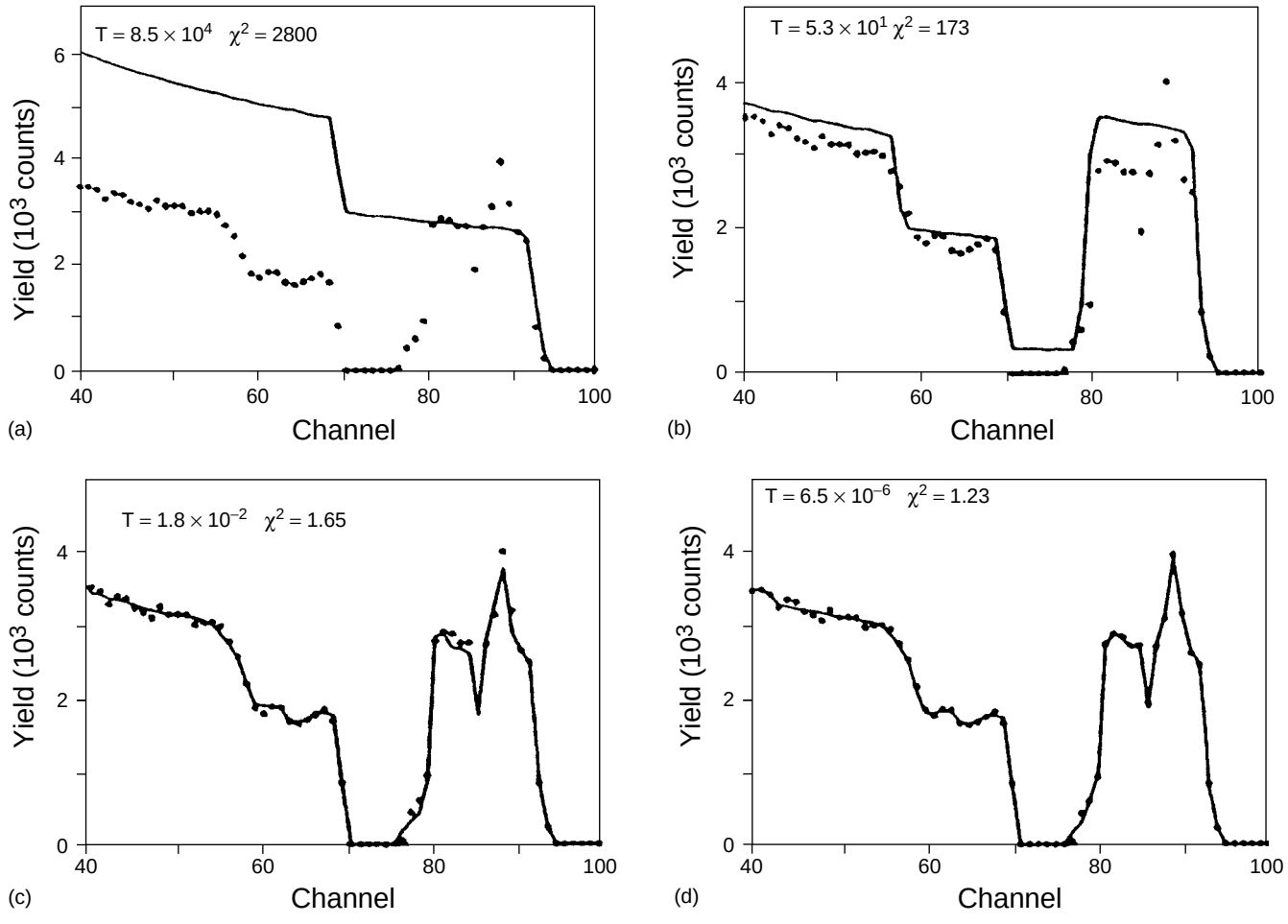

Figure 8 Differential steps of the fit from (a) the initial guess to (d) the final results obtained in simulated annealing analysis of a RBS spectrum of a multilayered iron–cobalt silicide. (Reproduced by permission of American Institute of Physics from Barradas et al.[24])

Channeling ions might stop channeling upon scattering by lattice defects (dechanneling) or even be directly backscattered from the defects (direct scattering). These effects depend on the structure of the defect. Impurity atoms, for example, may cause direct scattering depending on both the location of the impurity atom and the channeling axis as shown in Figure 9.[25] The substituted atoms cannot be seen by channeling ions but the interstitial atoms may be seen by the ions channeling through particular channels. Thus the site of the impurity atom can be roughly determined by channeling measurements at several major axes. For more detailed information, the lattice site of the interstitial atom can be determined within an accuracy of 0.01 nm by measuring the incident-angle dependence of the backscattering yield from the impurity atoms around the channeling direction.

Other defects also cause dechanneling and the energy dependence of the dechanneling probability is different in each case, that is $\propto \sqrt{E}$ for dislocation, $\propto E^0$ for stacking fault, and $\propto E^{-1}$ for point defect.[26] Thus the type of major defect can be determined by measuring the energy dependence of the dechanneling probability. Other useful applications of channeling are described in the literature.[27–29]

3.6 Comparison to Related Techniques

The low-energy version of RBS is called ion scattering spectroscopy (ISS).[3] Rare gas ions with energy of several kiloelectron volts and electrostatic analyzers are commonly used. ISS is sensitive to surface atoms because of the large neutralization probability of ions scattered from subsurface atoms and that is used extensively for compositional and structure analysis of surfaces. In order to study subsurface regions, the TOF technique is employed to measure energy spectra of both ions and neutral atoms. An alternative method is the use of alkali

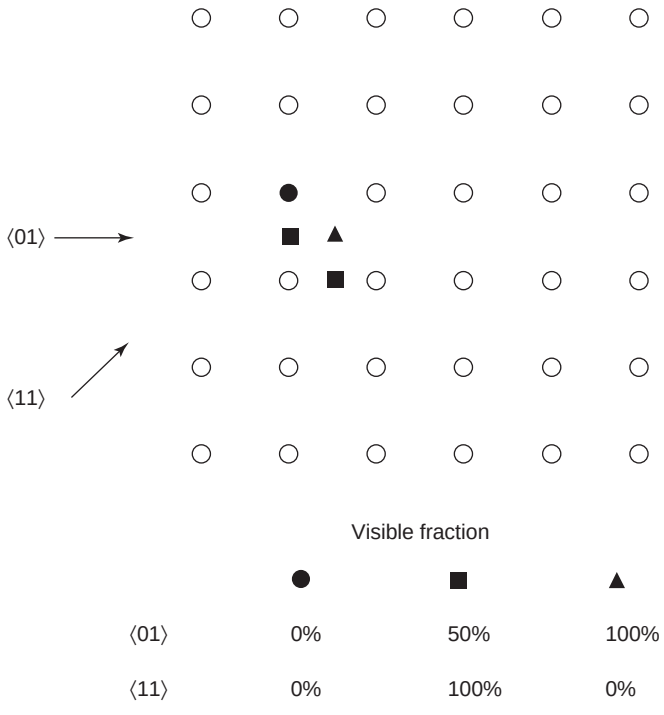

Figure 9 Two-dimensional model illustrating how the channeling effect can be used to locate foreign atoms in a crystal.

metal ions for which the neutralization probability is small even for the ions scattered from the subsurface region.[30]

The sensitivity of RBS is not sufficient for determination of trace elements. Particle-induced X-ray emission (PIXE) provides quantitative analysis of trace elements present at parts per million levels.[31] Also see the article **PIXE (Particle-induced X-ray Emission)** in this publication. H ions of 1–3 MeV are used to produce inner-shell vacancies in target atoms. Characteristic X-rays emitted via refilling the inner-shell vacancies by outer-shell electrons are detected by an X-ray detector. Compared with the similar technique of electron probe microanalysis (EPMA), the background due to the brehmsstrahlung is many orders of magnitude smaller and a sensitivity of better than 0.1 ppm is easily obtained for many elements in low-mass materials. Although the sensitivity of PIXE is excellent, PIXE does not have depth resolution.

Elastic Recoil Detection (ERD) is a similar technique to RBS.[32] Also see the article **Elastic Recoil Detection Analysis** in this publication. In ERD, recoiled ions are detected instead of the scattered ions. While RBS is suitable for analysis of heavy elements, ERD provides quantitative depth profiles of light elements, especially hydrogen. Before the first introduction of ERD,[33] nuclear reaction analysis (NRA) was predominantly used to measure depth profiles of light elements.[34] Also see the article **Nuclear Reaction Analysis** in this publication. Nowadays, ERD is commonly used because of its simplicity and ease of use although the standard ERD has poorer sensitivity and depth resolution than NRA.

4 ADVANCED TECHNIQUES

4.1 Heavy Ion Rutherford Backscattering Spectroscopy

As was discussed in section 2.1, use of heavy ions improves mass resolution. Heavy ions also have an advantage of better depth resolution than light ions because the stopping power is proportional to the square of the ion charge. Figure 10 shows an example of heavy ion RBS (HIRBS) spectrum of a thin $Y_1Ba_2Cu_3O_{7-x}$ film using 22-MeV ^{12}C ions and a SSB detector.[35] Although the energy resolution of the SSB detector is degraded for C ions, the separation of each element is perfect and, moreover, two copper isotopes, ^{63}Cu and ^{65}Cu, are clearly resolved, which is usually impossible in the standard RBS with megaelectron-volt He ions.

If the TOF technique is employed to measure the energy spectrum instead of the SSB detector, the energy resolution and thus both the mass and depth resolutions are improved considerably (typical energy resolution of TOF detectors is better than 1%). Figure 11 depicts an example of a HIRBS spectrum (25 MeV ^{35}Cl beam) of a 48 nm layer of GaAs capped by a 26 nm Pd layer on Si substrate measured with a TOF detector.[36] The two Ga isotopes as well as several Pd isotopes are well-resolved.

Although the excellent mass resolution is clear in HIRBS, there are several disadvantages of HIRBS: (1) The cross-section deviates from the Rutherford formula particularly for heavy target elements. (2) Light

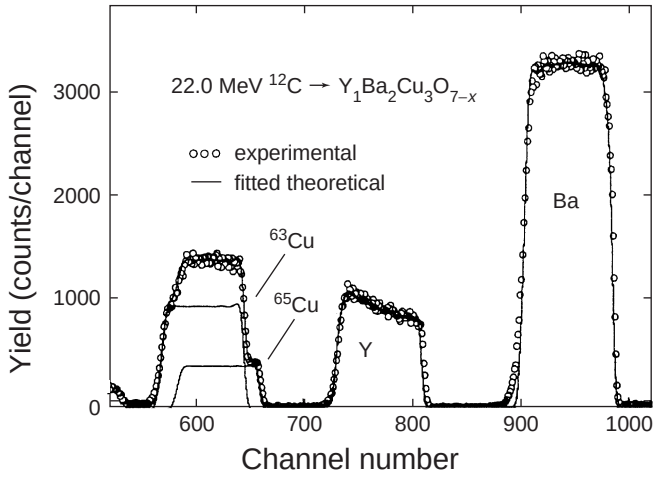

Figure 10 Heavy ion RBS spectrum of the $Y_1Ba_2Cu_3O_{7-x}$ high-T_c superconductor film measured with 22 MeV C ions. (Reproduced by permission of Elsevier Science B.V. from Rauhara et al.[35])

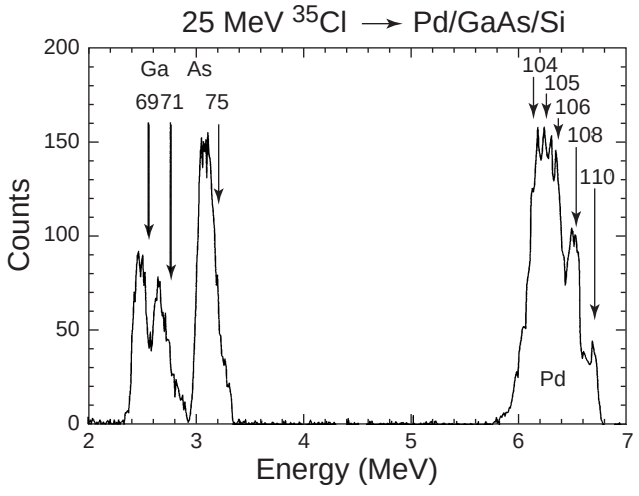

Figure 11 Heavy ion RBS spectrum (25-MeV ^{35}Cl beam) of a 48-nm layer of GaAs covered by 26 nm of Pd measured with a TOF detector. (Reproduced by permission of Elsevier Science B.V. from Döbeli et al.[36])

elements are less sensitive than standard RBS; in particular, elements lighter than the probe ion cannot be analyzed. (3) Irradiation of heavy ions causes significant radiation damage not only in targets but also in the SSB detectors.

4.2 High-resolution Rutherford Backscattering Spectroscopy

Improvement in the depth resolution up to atomic level has been tried by several groups.[37–46] An overview of existing high-resolution spectrometers for IBA is given in the literature.[45] It was shown that both a high-resolution spectrometer (energy resolution of ca. 0.1% is easily achieved for electrostatic or magnetic spectrometers) and a grazing-angle technique are essential for monolayer resolution in RBS.[44] Figure 12 shows examples of HRBS spectra of PbTe (100) and PbTe (111) measured with a magnetic spectrometer. The arrows labeled Pb and Te indicate the energies of the ions elastically scattered from Pb and Te atoms. There are several peaks at ca. 285.5, ca. 283.3 and ca. 281.1 keV in Figure 12(a). These peaks corresponds to the Pb atoms in the first, second, and third atomic layers, respectively. Note that even the ions scattered from the topmost atomic layer show inelastic energy loss because the electronic surface is outside the atomic surface. There are also small peaks at ca. 278 and ca. 275.8 keV corresponding to the Te atoms in the first and second layers. The yield ratio of the Pb peak to the Te peak is close to the cross-section ratio indicating that the PbTe (100) is stoichiometric. The spectrum of PbTe (111) shown in Figure 12(b) is basically the same as the PbTe (100) except for the yield of the first Pb

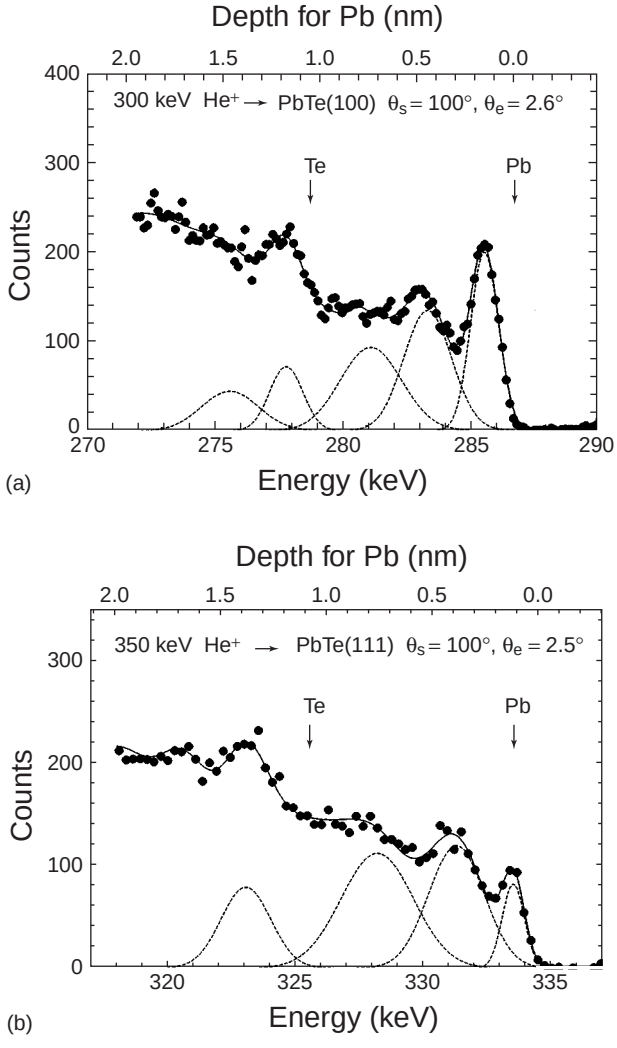

Figure 12 High-resolution rutherford backscattering spectroscopy (HRBS) spectra of (a) PbTe (001) and (b) PbTe (111) surfaces. Surface atomic layers can be resolved as separated peaks.

peak. The yield of the first Pb peak is about 25% of the second and third Pb peaks and the inelastic energy loss for the first Pb peak (ca. 1 keV) is smaller than that for the first Te peak (ca. 2.1 keV). From these results, it can be concluded that the PbTe (111) surface is terminated by the Pb layer and the atomic density of the topmost Pb layer is reduced to about 25% of the bulk value.

Combination of heavy ions with a high-resolution spectrometer may provide better depth resolution. In addition to the large radiation damage, however, the charge-state distribution of the scattered heavy ions complicates the analysis because the scattered ions of all charge states cannot simultaneously be measured by the electrostatic or magnetic spectrometers. An optimal probe ion for HRBS has been discussed and

the sub-megaelectron-volt He ion was concluded to be the best one.[46]

Disadvantages of HRBS are the relatively small acceptance angle $\Delta\Omega$ (typically less than 1 msr) and a narrow energy window of the spectrometer (typically several percent of the ion energy). These lead to a long acquisition time and possible radiation damage. Spectrometers with a wide acceptance angle and a wide energy window have been developed by several groups[47,48] and not only a high-resolution spectrometer but also a high-resolution RBS system with a compact sub-megaelectron-volt ion accelerator are now commercially available.

4.3 Microprobe Rutherford Backscattering Spectroscopy

Finely focused high-energy ion beams turn ion beam spectroscopy into ion beam microscopy.[49] Development of a high-energy ion microprobe was started as early as the late 1960s. There are more than 40 facilities which have high-energy ion microprobes. Magnetic quadrupole lenses are commonly used to form microprobes of the order of ca. 1 μm at a beam current of ca. 100 pA. In scanning the microprobe across a specimen and detecting a variety of ion-scattering processes, such as RBS, PIXE elastic recoil, nuclear reaction, and so on, microscope images are obtained. These images display features of the specimen that cannot readily be imaged by other techniques. The combination of the microprobe with RBS provides three-dimensional mapping of elements in the specimen.

Figure 13 shows an example of the RBS mapping and tomographic image of 1 μm size WSi_x wiring lines.[50] A beam of 300-keV Be^{2+} ions (beam size ca. 0.1 μm) was rastered over a $30 \times 30\,\mu m^2$ area and RBS spectra were measured at 128×128 points. The mapping image is made with the W signal of the observed RBS spectra.

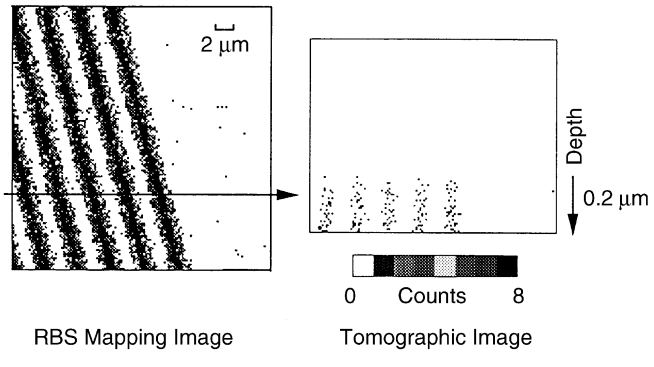

Figure 13 RBS mapping and tomographic image of 1-μm-sized WSi_x wiring lines measured with a microprobe. (Reproduced by permission of Elsevier Science B.V. from Park et al.[50])

The tomographic image shows the cross-sectional image along the line shown in the mapping image, which is obtained by converting the each RBS spectrum to a depth distribution of W. The WSi_x lines are clearly resolved in both the mapping and the tomographic images.

4.4 Light Element Analysis by Non-Rutherford Scattering

When the ion energy approaches the Coulomb barrier height of the target nucleus, the ion penetrates the Coulomb barrier and a compound nucleus in a highly excited state may be formed. If the incident energy of the ion matches one of the excited levels, the reaction probability is considerably changed and often enhanced. The excited compound nucleus thus produced may decay via re-emission of the incident particle. This phenomenon, called elastic resonance scattering, has a large cross-section sometimes several orders of magnitude larger than the Rutherford cross-section. Figure 14 displays an example of the non-Rutherford cross-section for $^4He-^{12}O$ scattering as a function of the incident He energy.[51] There are several resonances with enhancement factors of 100–200. Because the cross-sections for heavy elements remain Rutherford values in this energy region, the relative sensitivity for oxygen is improved by a factor of ca. 100. However, the cross-section varies in a very complicated manner with energy and also with the scattering angle. The analysis of energy spectra has to be performed with care. There is an energy region, 8.3–8.9 MeV, where the cross-section varies slowly. Using this energy region the analysis of oxygen is as simple as standard RBS with an enhancement factor of about 35.

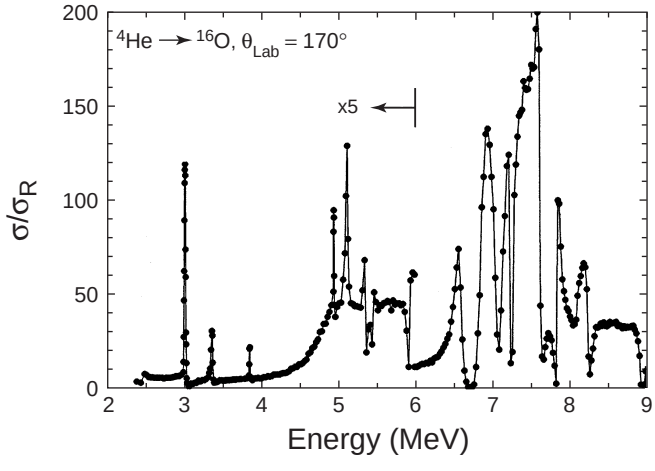

Figure 14 Measured 170° laboratory cross-sections for $^4He-^{16}O$ expressed in terms of the Rutherford cross-sections. (Reproduced by permission of Elsevier Science B.V. from Cheng et al.[51])

A sharp and isolated resonance is seen at $E_R = 3.045$ MeV. When the energy of the incident He ions is larger than E_R, the resonance occurs at a certain depth where the ion energy matches E_R after energy loss in the incoming path, that is, only the oxygen at a certain depth can be seen by the resonance. Thus a sharp resonance can be used for depth profiling of oxygen by sweeping the incident ion energy. Other examples of sharp resonances used in non-RBS are the 4.265 MeV resonance in the $^4He-^{12}C$ cross-section and 2.525 MeV resonance in the $^1H-^9Be$ cross-section. Useful data on non-Rutherford cross-sections for H and He incident ions are compiled in the literature.[52]

ABBREVIATIONS AND ACRONYMS

CAARI	Conference on the Application of Accelerators in Research and Industry
ECAART	European Conference on Accelerator Applications in Research and Technology
EPMA	Electron Probe Microanalysis
ERD	Elastic Recoil Detection
HIRBS	Heavy Ion Rutherford Backscattering Spectroscopy
HRBS	High-resolution Rutherford Backscattering Spectroscopy
IBA	Ion Beam Analysis
ISS	Ion Scattering Spectroscopy
NMTA	Nuclear Microprobe Technology and Applications
NRA	Nuclear Reaction Analysis
PHA	Pulse-height Analyzer
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy
SSB	Silicon Surface Barrier
TOF	Time-of-Flight

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Elastic Recoil Detection Analysis • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission)

REFERENCES

1. A.B. Brown, C.W. Snyder, W.A. Fowler, C.C. Lauritsen, 'Excited States of the Mirror Nuclei, Li^7 and Be^7', *Phys. Rev.*, **82**, 159–181 (1951).
2. J.W. Mayer, M.-A. Nicolet, W.-K. Chu, in *Material Characterization Using Ion Beams*, eds. J.P. Thomas, A. Cuchard, Plenum Press, New York, 333–366, 1978.
3. L.C. Feldman, J.W. Mayer, *Fundamentals of Surface and Thin Film Analysis*, North-Holland, Amsterdam, 1986.
4. J.R. Tesmer, M. Nastasi (eds.), *Handbook of Modern Ion Beam Materials Analysis*, Material Research Society, Pittsburgh, 1995.
5. M.F. De Silva, J.C. Soares (eds.), 'Proceedings of the 13th International Conference on Ion Beam Analysis', Lisboa/Portugal 1997, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 136–138.
6. G. Molière, 'Theorie der Streuung schneller geladener Teilchen I Einzelstreuung am abgeshirmten Coulomb-Feld', *Z. Naturforsch, A: Phys. Sci.*, **2**, 133–145 (1947).
7. J.F. Ziegler, J.P. Biersack, U. Littmark, 'The Stopping and Ranges of Ions in Solids', in *The Stopping and Ranges of Ions in Matter*, ed. J.F. Ziegler, Pergamon Press, New York, Vol. 1, 1985.
8. M.H. Mendenhall, R.A. Weller, 'Algorithms for the Rapid Computation of Classical Cross-Sections for Screened Coulomb Collisions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **58**, 11–17 (1991).
9. J. L'Ecuyer, J.A. Davies, N. Matsunami, 'How Accurate Are Absolute Rutherford Backscattering Yields', *Nucl. Instrum. Methods*, **160**, 337–346 (1979).
10. J.A. Leavitt, L.C. McIntyre, Jr., 'Backscattering Spectrometry', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 37–81, 1995.
11. F. Bloch, 'Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie', *Ann. Phys. (Leipzig)*, **16**, 285–320 (1933).
12. H.H. Andersen, J.F. Ziegler, 'Hydrogen Stopping Powers and Ranges in All Elements', in *The Stopping and Ranges of Ions in Matter*, ed. J.F. Ziegler, Pergamon Press, New York, Vol. 3, 1977.
13. J.F. Ziegler, 'Helium Stopping Powers and Ranges in All Elemental Matter', in *The Stopping and Ranges of Ions in Matter*, ed. J.F. Ziegler, Pergamon Press, New York, Vol. 4, 1977.
14. ICRU Report 49, 'Stopping Powers and Ranges for Protons and Alpha Particles', International Commision on Radiation Units and Measurements, Bethesda, 1993.
15. N. Bohr, 'The Penetration of Atomic Particles through Matter', *Kgl. Danske Videnskab. Selskab. Matt-Fys. Medd.*, **18**(8), (1948).
16. J. Lindhard, M. Scharff, 'Energy Loss in Matter by Fast Particles of Low Charge', *Kgl. Danske Videnskab. Selskab. Matt-Fys. Medd.* **27**(15), (1953).
17. E. Bonderup, P. Hvelpund, 'Stopping Power and Energy Straggling for Swift Protons', *Phys. Rev. A*, **4**, 562–569 (1971).

18. W.K. Chu, 'Calculation of Energy Straggling for Protons and Helium Ions', *Pys. Rev. A*, **13**, 2057–2060 (1976).
19. F. Besenbacher, J.U. Andersen, E. Bonderup, 'Straggling in Energy Loss of Energetic Hydrogen and Helium Ions', *Nucl. Instrum. Methods*, **168**, 1–16 (1980).
20. A. Climent-Font, C. Palacio, M.T. Fernández-Jiménez, D. Díaz, 'Characterization of BiSrCaCuO Thin Films by the Complementary Use of IBA and AES Analytical Techniques', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **118**, 108–112 (1996).
21. L.R. Doolittle, 'A semi-automatic Algorithm for Rutherford Backscattering Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 227–231 (1986).
22. J. Saarilahti, E. Rauhala, 'Interactive Personal Computer Data Analysis of Ion Backscattering Spectra', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 734–738 (1992).
23. M.D. Strathman, S. Bauman, 'Angle-Resolved Imaging of Single Crystal Materials with MeV Helium Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 840–845 (1992).
24. N.P. Barradas, C. Jeynes, R.P. Webb, 'Simulated Annealing Analysis of Rutherford Backscattering Data', *Appl. Phys. Lett.*, **71**, 291–293 (1997).
25. J.W. Mayer, L. Eriksson, J.A. Davies, *Ion Implantation in Semiconductors*, Academic Press, New York, 1970.
26. Y. Quéré, 'Dechanneling of Fast Particles by Lattice Defects', *J. Nucl. Mater.*, **53**, 262–267 (1974).
27. L.C. Feldman, J.W. Mayer, S.T. Picrau, *Materials Analysis by Ion Channeling*, Academic Press, New York, 1982.
28. D.S. Gemmel, 'Channeling and Related Effects in the Motion of Charged Particles through Crystals', *Rev. Mod. Phys.*, **46**, 129–227 (1974).
29. M.L. Swanson, 'Channeling', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 231–300, 1995.
30. E. Taglauer, W. Englert, W. Heiland, D.P. Jackson, 'Scattering of Low-Energy Ions from Clean Surfaces: Comparison of Alkali- and Rare-Gas-Ion Scattering', *Phys. Rev. Lett.*, **45**, 740–743 (1980).
31. S.A.E. Johansson, J.L. Campbell, *PIXE: A Novel Technique for Elemental Analysis*, Wiley, New York, 1988.
32. W.A. Arnold Bik, F.H.P.M. Habraken, 'Elastic Recoil Detection', *Rep. Prog. Phys.*, **56**, 859–902 (1993).
33. J. L'Ecuyer, C. Brassard, C. Cardinal, J. Chabbal, L. Deschenes, J.P. Labrie, B. Terrault, J.G. Martel, R. St. Jacques, 'An Accurate and Sensitive Method for the Determination of the Depth Distribution of Light Elements in Heavy Materials', *J. Appl. Phys.*, **47**, 381–382 (1976).
34. W.A. Lanford, 'Nuclear Reactions for Hydrogen Analysis' in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 193–204, 1995.
35. E. Rauhara, J. Saarilahti, N. Nath, 'Protons, Helium and Heavy ions in Backscattering Analysis of High-T_c Superconductors', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **61**, 83–90 (1991).
36. M. Döbeli, P.C. Haubert, R.P. Livi, S.J. Spicklemire, D.L. Weathers, T.A. Tombrello, 'A Time-of-Flight Detector for Heavy Ion RBS', *Nucl. Instrum. Methods Phys. Res. Sect. B*, **56/57**, 764–767 (1991).
37. E. Bøgh, 'Rutherford Scattering of Protons in the Surface Layers of a Tungsten Single Crystal', *Phys. Rev. Lett.*, **19** 61–63 (1967).
38 J.K. Hirvonen, G.K. Hulbler, 'Application of a High-Resolution Magnetic Spectrometer to Near-Surface Material Analysis', in *Ion Beam Surface Layer Analysis*, eds. O. Meyer, G. Linker, F. Käppeler, Plenum, New York, 457–469, 1976.
39. A. Feuerstein, H. Grahmann, S. Kalbitzer, H. Oetzmann, 'Rutherford Backscattering Analysis with Very High Depth Resolution using an Electrostatic Analyzing System', in *Ion Beam Surface Layer Analysis*, eds. O. Meyer, G. Linker, F. Käppeler, Plenum, New York, 471–481, 1976.
40. T. Oku, J. Kanasaki, N. Matsunami, N. Itoh, K. Matsuda, M. Aoki, 'Proton Energy Loss Spectroscopy for Surface Layer Analysis in the Monolayer Regime', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 142–145 (1986).
41. J. Vrijmoeth, P.M. Zagwijn, J.F.M. Frenken, J.F. van der Veen, 'Monolayer Resolution in Medium-Energy Ion-Scattering Experiments on the NiSi2 (111) Surface', *Phys. Rev. Lett.*, **67**, 1134–1137 (1991).
42. Th. Enders, M. Rilli, H.D. Carstanjen, 'A High-Resolution Electrostatic Spectrometer for the Investigation of Near-Surface Layers in Solids by High-Resolution Rutherford Backscattering with MeV Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 817–824 (1992).
43. W.M. Arnoldbik, W. Wolfswinkel, D.K. Inia, V.G.G. Verleun, S. Lobner, J.A. Reinders, F. Labohm, D.O. Boerma, 'A High Resolution Magnetic Spectrograph for Ion Beam Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **118**, 567–572 (1996).
44. K. Kimura, K. Ohshima, M. Mannami, 'Monolayer Analysis in Rutherford Backscattering Spectroscopy', *Appl. Phys. Lett.*, **64**, 2232–2234 (1994).
45. D.O. Boerma, 'Report on the Round Table Session on: Near Surface Analysis and Charge Exchange Studies with High Resolution Spectrographs', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 42–46 (1998).
46. K. Kimura, K. Nakajima, M. Mannami, 'Some Applications of High-resolution RBS and ERD using a Magnetic Spectrometer', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 1196–1202 (1998).
47. W.A. Lanford, B. Anderberg, H. Enge, B. Hjorvarsson, 'Compact Broad Range Magnetic Spectrometer for Use in Ion Beam Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 1177–1182 (1998).

48. K. Kimura, M. Kimura, Y. Mori, M. Maehara, H. Fukuyama, 'Development of a Compact High-resolution RBS System for Monolayer Analysis', in *Application of Accelerators in Research and Industry*, eds. J.L. Duggan, I.L. Morgan, American Institute of Physics, Woodbury, 500–503, 1999.
49. M.B.H. Breese, D.N. Jamieson, P.J.C. King, *Material Analysis Using a Nuclear Microprobe*, Wiley, New York, 1996.
50. Y.K. Park, T. Kishimoto, M. Takai, 'Microanalysis of a submicron patterned WSi_x Structure using a Nuclear Microprobe', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **130**, 534–538 (1997).
51. H.-S. Cheng, H. Shen, J. Tang, F. Yang, 'Cross Sections for 170° Backscattering of ^{4}He from Oxygen in the Energy Range of 2.0–9.0 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **83**, 449–453 (1993).
52. R.P. Cox, J.A. Leavitt, L.C. McIntyre, Jr, 'Non-Rutherford Elastic Backscattering Cross Sections', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 481–508, 1995.

Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Richard H. Brown
Health and Safety Laboratory, Sheffield, UK

Martin Harper
University of Alabama at Birmingham, Birmingham, USA

1 Introduction 2
2 Selection of Sampling Devices and Measurement Methods 2
3 Whole-air Samplers 3
3.1 Methods of Sampling and Analysis 3
3.2 Canisters 3
3.3 Flexible Plastic Containers 4
4 Active Sorbent Samplers 4
4.1 Sampling Pumps 4
4.2 Liquid Absorbers 5
4.3 Solid Sorbents 5
4.4 Coated Sorbents and Filters 13
5 Diffusive Samplers 14
5.1 Operating Principles 14
5.2 Environmental Factors Affecting Sampler Performance 15
6 Sampling Devices with Direct Readout 17
6.1 Operation of Detector Tubes 17
6.2 Chemical Reactions used in Detector Tubes 18
6.3 Theory of Operation 18
6.4 Certification of Detector Tubes 18
7 Calculations 18
7.1 Absorber 19
7.2 Sorbent Tube 19
7.3 Volume Fraction 19
7.4 Diffusive Sampler 19
8 Standards 19
8.1 European Standards 19
8.2 US Standards 20
8.3 Standards Issued by International Organization for Standardization 20
9 Quality Systems and Quality Control 21
Acknowledgments 21
List of Symbols 22
Abbreviations and Acronyms 22
Related Articles 22
References 23

It is possible to determine many gases and vapors in workplace air through the use of direct-reading instruments. However, these instruments are often expensive, and not suited to personal monitoring as required by standards and regulations. They also may be insufficiently selective to monitor atmospheres contaminated by a mixture of components, or insufficiently sensitive to determine the low concentrations of, for example, regulated carcinogens. A more common technique is to remove a sample of air for subsequent analysis, typically in a laboratory, but sometimes in the field. This article covers sampling techniques for the determination of gases and vapors in air. It also covers recovery techniques that are applicable to the laboratory. It is not possible to completely divorce these subjects as the selection of sampling method is often dependent on the choice of recovery technique, and vice versa. Taking a sample of air improves the sensitivity of analysis, as well as integrating the result over the time-period of sample collection. Exposure limit (EL) values are normally given as time-weighted average concentrations over work-shifts or shorter periods. It is possible in most cases to choose a sampling method to cover a full work-shift with a single sample. By far the commonest technique is to concentrate the sample by collecting the contaminants of interest using a solid or liquid sorbent, where the mechanism of collection is adsorption, chemical reaction, or solution. The collection procedure involves either pulling air through the sorbing medium (active sampling) or allowing the molecules to diffuse to the medium under certain conditions (passive, or diffusive, sampling).

The commonest technique for recovery is solvent desorption. An aliquot of the solution is then presented for analysis. For volatile organic compounds (VOCs), the commonest technique is to use a charcoal sorbent medium contained in a tube, through which air is drawn, followed by desorption using carbon disulfide, either alone or mixed with other solvents. Probably more than 75% of workplace air samples for VOCs are taken this way. Thermal desorption, using heat to release the sampled VOCs, rather than a solvent, is of rapidly growing importance because of its higher sensitivity. The selection of sorbent is more critical, and most recent protocols are using multiple beds of different sorbents. Whole-air samples in polymer bags have been taken for industrial monitoring of permanent gases, whereas whole-air samples in passivated metal canisters have been used commonly in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © Crown Copyright

ambient air monitoring of VOCs. Similar, but smaller, metal canisters may also be used in future for industrial hygiene work. Industrial hygiene air sampling will exist for as long as chemicals are used in the workplace, to determine the existence of hazards and the extent of risks, as well as to ensure the effectiveness of engineering controls. In many countries, the use of air sampling is a requirement of health and safety programs in order to compare workers' exposures to regulatory guidelines or legal limits.

1 INTRODUCTION

This article discusses the sampling and recovery of gases and vapors commonly found in the workplace environment. Many of these techniques are equally applicable to indoor or ambient-air monitoring outside of the workplace. It is limited to descriptions of sampling methods for on-site or subsequent laboratory analysis. It does not, therefore, include any discussions of direct-reading instruments or laboratory analytical procedures (other than sample recovery), which can be found in the appropriate related articles.

2 SELECTION OF SAMPLING DEVICES AND MEASUREMENT METHODS

There are two basic methods for collecting gas and vapor samples. In one, called whole-air sampling, an actual sample of contaminated air is taken in a flask, bottle, bag, or other suitable container. In the other, called sorbent sampling, gases or vapors are removed from the air and concentrated by contact with a sorbing medium.

The first method is often used for the collection of instantaneous or short-term (grab) samples, usually within a few seconds or a minute, but the method is easily adapted to sampling over longer periods. Many direct-reading on-the-spot devices also involve grab sampling. Grab sampling is of questionable value when the contaminant or contaminant concentration varies with time and a time-weighted average exposure is desired. In such circumstances, integrated sampling is used instead, and the most popular method used in occupational hygiene today is sorbent sampling. The gas or vapor is extracted from air and concentrated by (a) dissolution in a sorbing liquid, (b) reaction with a sorbing solution (or reagent therein), or (c) collection onto a solid sorbent. Any of these methods can be combined with analytical reagents to provide an on-site determination. For each method the collection efficiency should be accurately known and should be close to 100%. For samples analyzed by laboratory techniques sample recovery efficiency and storage stability should also be known.

There are thus frequently many alternative approaches to the determination of gases and vapors, and the best one will depend on the circumstances.

Factors that will need to be taken into account include:

1. the measurement task;
2. the concentration to be determined;
3. the time resolution required;
4. selectivity to the target gas or vapor and sensitivity to interfering gases and vapors;
5. bias, precision and overall uncertainty required;
6. susceptibility of the sampler to environmental factors;
7. fitness for purpose, such as weight, size, durability, etc. (and especially the suitability of the equipment for personal monitoring);
8. training requirements for the reliable operation, maintenance and calibration;
9. the total cost of purchase and operation, including calibration and maintenance;
10. compliance with the performance requirements of appropriate national regulations;
11. conformity to the user's quality system.

Having established the requirements, the next step in the selection of a sampling device and analytical procedure is to search the available literature. Primary sources are the compendia of methods recommended by the regulatory authorities or governmental agencies in Europe:

- the UK *Methods for the Determination of Hazardous Substances*;[1]
- the German *Analytische Methoden*[2] and *BIA Arbeitsmappe Messung von Gefahrstoffen*;[3]
- the Swedish *Principer och Methoden*;[4]
- the US *NIOSH Manual of Analytical Methods*,[5] *OSHA Analytical Methods Manual*,[6] or *USEPA (United States Environmental Protection Agency) Compendium of Methods*.[7]

Secondary sources are published literature references in, for example, the *Analyst*, the *Journal of Environmental Monitoring*, the *Annals of Occupational Hygiene, Staub, Reinhaltung der Luft*, the *American Industrial Hygiene Association Journal*, or *Applied Occupational and Environmental Hygiene*. Further valuable resources include the Methods Exchange Network of the American Industrial Hygiene Association (AIHA) and standards published by the American Society for Testing and Materials (ASTM; see section 8).

If a published procedure is not available, then one can be devised from theoretical considerations. However, its suitability must be established experimentally before application. Protocols for these experimental evaluations are available from the government agencies and standards organizations mentioned above. The final stage is to review the performance characteristics of the available methods against the selection criteria already established. Important information on the performance characteristics (items 5 and 10 above) of devices or procedures can be obtained from various sources. These include:

- the manufacturer's instructions for use;
- published commercial technical information;
- technical and research publications;
- national and international standards (see section 8);
- user groups.

3 WHOLE-AIR SAMPLERS

3.1 Methods of Sampling and Analysis

Whole-air samplers are containers of varying capacity and configurations. They can have a rigid construction (typically metal), in which case, the internal pressure of the container is reduced, either to near zero (<0.1 mbar) or to a known absolute pressure, and sampling takes place by filling through a critical orifice. If the container size is carefully matched with the size of the critical orifice, samples of various volumes can be obtained over different time-periods. It is possible to obtain sufficient sample for replicate analyses from a container whose volume is low enough (200–600 mL) for the container to be worn as a personal sampler. Such a container can be filled over long time periods with the appropriate choice of critical orifice. For example, a sampling rate of 0.4 mL min^{-1} will two-thirds fill a 300 mL container over an 8-hour work-shift. As the vacuum is depleted the orifice will become subcritical and the subsequent rate of fill nonlinear. Two-thirds fill is a typical limitation of the linear range of critical-orifice controllers, although some are now available with modifications so that the fill-rate remains linear with only small pressure differences. Fine filters are required to ensure the flow controllers are not blocked by particles. This method has the advantages that a personal sampling pump is not required and that the containers normally can be re-used.

Alternatively, the containers can be of flexible plastic, in which case the internal volume is close to zero when sampling begins, and sample is either fed into the container, or drawn in by a vacuum applied to the exterior of the container. The former technique can lead to bias either through contamination of the sample or loss from the sample during passage through the sampling pump. One manufacturer supplies a sampling pump specifically designed to minimize these biases. The latter technique involves a pump and a rigid box that can be evacuated and the assembly is too cumbersome for personal sampling.

Both types of containers are generally removed to a laboratory for analysis, although it is possible to achieve field readability if the proper equipment and direct-reading instrument are available. The NIOSH ((US) National Institute for Occupational Safety and Health) has published several methods for the field determination of vapors after collection with sampling bags.[5] In the laboratory, the containers can be analyzed by direct injection into the analytical device. For sampling bags this is usually by withdrawing an aliquot through a septum fitting on the bag with a gas-tight syringe. For rigid containers the procedure is more complex. Typically, the pressure within the container is measured and the container is then overpressured with clean gas (air, nitrogen, or helium) and the pressure measured again. From the pressure difference the dilution of the sample can be determined. An aliquot of the sample is then bled into a gas-sampling valve and passed to the analytical device. This procedure has the advantage of allowing several analyses of the same sample, but the disadvantage of diluting it. It is also possible to overpressure the sample in the field, using a pump. However, if the water vapor partial pressure within the canister exceeds its saturation vapor pressure then water vapor condenses on the canister wall. This can have a marked effect on the recovery of polar compounds and on the analysis because the amount of water vapor entering the gas chromatography (GC) column will vary as sample is withdrawn from the container.[8]

An alternative procedure for the analysis of canisters is to pump out the entire contents of the sample through a trap (cold trap, sorbent trap, or a combination) and to analyze the contents of the trap. This only allows a single shot at the analysis, but eliminates the dilution factor, and the problem of moisture condensation within the can, as the water will be dealt with at the focusing stage.

3.2 Canisters

Glass containers originally used to collect samples have been largely superseded by stainless steel containers that have been specially treated to reduce sorption effects. Such treatments include electropolishing (removal of free surface iron), the SUMMA® process (deposition of mixed nickel/chromium oxide), and SilicoSteel® lining (deposition of fused silica layer). Electropolished canisters are seldom seen today. SUMMA® canisters are the most widespread, but they show poor stability for some reactive compounds, and give poor recoveries, especially for

polar organic compounds, at low humidities. Water can be injected into the can to alleviate this problem. Fused silica-lined cans show greater stability for some reactive compounds and better recovery for polar compounds and are becoming very popular. All types of canister have been used extensively for collecting trace organic gases in ambient air sampling,[9,10] but their potential advantages have only recently been recognized in occupational sampling. Advantages over the traditional pump/sorbent tube method include:

- applicability to low-molecular-weight compounds (e.g. C_1 and C_2 hydrocarbons, permanent gases) and some reactive compounds (e.g. reduced sulfur gases);
- minimal interference from ozone (inside the container ozone has a short life-time because of autoreaction);
- independence from batteries (e.g. can be used in cold climates).

The size and weight of containers is greater than the micropumps used for low flow-rate sampling, but not much greater than the larger personal pumps used for sampling aerosols. The cost of container plus orifice controller is equivalent to that of a high-end sampling pump. Larger containers are held in a harness suspended from the workers' shoulder, smaller containers may be equipped with a belt-clip.

A recent study[11] has shown some significant under- and overestimations of certain compounds by the canister method compared to a continuous cycling gas chromatograph used for ambient air analysis.

3.3 Flexible Plastic Containers

Plastic bags are used to collect air samples and prepare known concentrations that can range from parts per billion to more than 10% by volume in air. The bags are commercially available in sizes up to 250 L. However, 1–15 L bags are the most useful.

These bags are constructed from a number of materials, including polyester, poly(vinylidene chloride), Teflon®, aluminized Mylar®, or other fluorocarbons (e.g. Tedlar®). Tedlar® bags are the most popular. Bags have the advantages of being light, nonbreakable, inexpensive to ship, and simple to use. However, they should be used with caution as many factors affect precision and accuracy. Storage stability is probably the most critical factor affecting the use of bags. Instability may be due to permeation through the material of the bag, adsorption by the bag or fittings and/or reaction with other sampled components. Permeation through the bag is a function of the ratio of bag surface area to filled volume, the nature and thickness of the material, and the temperature. Although permeation is relatively rapid for permanent gases it is not the main source of long-term losses for less-VOCs. Losses of these compounds can be relatively rapid (half-life of hours) or slow (half-life of days) and depend more on the nature of the bag material and fittings (for example, many compounds are more stable in the presence of polypropylene fittings than stainless steel fittings).[12] Photochemical reactions have been observed inside clear bags, and opaque bags are often recommended for reactive compounds. An example of a popular sampling method using Tedlar® bags is the method for reduced sulfur gases published by the National Council for Air and Stream Improvement (NCASI) in the USA.[13]

Plastic bags should be tested before they are used. Such testing should be done under ambient conditions that approximate those of the sampling environment. Some general recommendations are available in the published literature[14–17] for the use of such bags for air sampling. A good review of specific applications up to 1967 is Schuette;[18] other specific applications have been discussed.[19–21] Posner and Woodfin[21] made a useful systematic study of five bag types and six organic vapors; they conclude that Tedlar® bags are best for short-term sampling, whereas aluminized Mylar® bags are better for long-term storage prior to analysis. Storage properties, decay curves, and other factors, however, will vary considerably from those reported for a given gas or vapor because sampling conditions are rarely identical. Each bag, therefore, should be evaluated for the specific gas or gas mixture for which it will be used. In theory, these bags are reusable, but caution must be exercised as memory effects are common, and small holes or other leaks can appear after even a single use. Bags should not be filled to bursting, and filled bags should not be transported in depressurised aircraft holds. In addition, before transporting hazardous atmospheres in any kind of containers it is necessary to check whether the containers comply with transportation safety regulations.

4 ACTIVE SORBENT SAMPLERS

4.1 Sampling Pumps

Except for evacuated canisters (described in section 3) and diffusive samplers (described in section 5), sampling devices are used in conjunction with a sampling pump and air-metering device. To avoid contaminating the metering device and pump, these are usually placed downstream of the sampler during the sampling period. However, because many samplers introduce back-pressure, the sampling train should be precalibrated by the use of an external flow-meter upstream of the sampling head.[22]

There are a large number of manufacturers of sampling pumps, although the mechanisms are rather similar and

not far removed from the fish-tank pumps they were developed from. Typically, personal sampling pumps operate through movement of a piston or diaphragm, and often contain some form of surge or pulsation dampener to ensure even flow (within 5% of specification).[23] A feed-back circuit may be included to cause pumps to speed up or slow down in response to back-pressure changes. Although the latest generation of pumps has sophisticated means to measure the flow, government regulations normally require the flow to be checked before and after sampling (and the final result to be within 5% of the initial reading) using a standard method such as a bubble flow-meter. Periodic calibration during the sampling period is recommended for most accurate results. Some form of visual flow-fault indicator is often included to alert the wearer to failure of the sampling train. Most pumps, whether piston or diaphragm, can be assumed to deliver a fixed volume that does not change with temperature and pressure. However, that is not always true for the calibration device. Whereas bubble flow-meters also have a fixed volume, the calibration of a rotameter depends on air density and, therefore, on temperature and pressure. A pump should be calibrated using a bubble flow-meter for most accurate results without need of correction.

Since the early 1970s the size, weight and noise emission of pumps has been reduced to the point of being an equivalent burden to the worker as one of today's cellular telephones. Sampling pumps are sold at a moderate cost, practically insignificant when amortized over a large number of samples, and when compared to the cost of analysis of those samples. Added to this cost, however, is the requirement for maintenance and calibration. Using an occupational hygiene professional only for pump calibration is a possibly unnecessary extra expense to a sampling program.

4.2 Liquid Absorbers

The sorption theory of gases and vapors from air by solution, as developed by Elkins et al.,[24] assumes that gases and vapors behave like perfect gases and dissolve to give a perfect solution. The concentration of the vapor in solution is increased during air sampling until an equilibrium is established with the concentration of vapor in the air. Sorption is never complete, however, because the vapor pressure of the material is not reduced to zero but is only lowered by the solvent effect of the sorbing liquid. Some vapor will escape with continued sampling, but it is replaced. Continued sampling will not increase the concentration of vapor in solution once equilibrium is established.

According to formulas developed by Elkins et al.,[24] and verified by Gage[25] in his experiments with ethylene oxide, the efficiency of vapor collection depends on (a) the volume of air sampled, (b) the volume of the sorbing liquid, and (c) the volatility of the contaminant being collected. Efficiency of collection, therefore, can be increased by cooling the sampling solution (reducing the volatility of the contaminant), increasing the solution volume by adding two or more bubblers in series, or altering the design of the sampling device. Sampling rate and concentration of the vapor in air are not primary factors that determine collection efficiency.

Sorption of gases and vapors by chemical reaction depends on the size of the air bubbles produced in the bubbler, the interaction of contaminant with reagent molecules, the rapidity of the reaction, and a sufficient excess of reagent solution. If the reaction is rapid and a sufficient excess of reagent is maintained in the liquid, complete retention of the contaminant is achieved regardless of the volume of air sampled. If the reaction is slow and the sampling rate is not low enough, collection efficiency will decrease.

Four basic sorbers used for the collection of gases and vapors are simple gas-washing bottles, spiral and helical sorbers, fritted bubblers, and glass-bead columns. Their function is to provide sufficient contact between the contaminant in the air and the sorbing liquid.

Petri, Dreschsel, and midget impingers are examples of simple gas-washing bottles. They function by applying a suction to an outlet tube, which causes sample air to be drawn through an inlet tube into the lower portion of the liquids contained in these absorbers. They are suitable for collecting nonreactive gases and vapors that are highly soluble in the sorbing liquid. The sorption of methanol and butanol in water, esters in alcohol, and organic chlorides in butyl alcohol are examples. They are also used for collecting gases and vapors that react rapidly with a reagent in the sampling media. High collection efficiency is achieved, for example, when toluene diisocyanate is hydrolyzed to toluene diamine in Marcali[26] solution. Hydrogen sulfide reaction with cadmium sulfate and ammonia neutralized by dilute sulfuric acid are other examples.

4.3 Solid Sorbents

Solid materials will remove gases and vapors from air with which they are in contact. This is a result of the physical forces of adsorption (particularly the London dispersion force of the Van der Waals interactions),[27] or it may be a result of the stronger attraction of hydrogen bonding, or even of complete chemical reaction. In the case of physical adsorption, the strength of the adsorption and the capacity of the adsorbent are related to its surface area and porosity, as well as to the chemical nature and polarizability of both the surface and the adsorbed

molecules. However, even such relatively nonporous materials as glass beads and stainless steel surfaces are able to adsorb significant numbers of molecules, and this is the cause of sample losses in some systems. In order to obtain sufficient molecules for analysis it is necessary to effectively trap all of the contaminant molecules in samples of up to several liters of air, and preferably in a small quantity of sorbent that can be deployed conveniently for personal sampling. This requires a sorbent with a relatively large surface area (from 10 to above 1000 $m^2 g^{-1}$).

The sorption capacity of a sampler, i.e. the volume of air that can be collected without loss of contaminant, depends on the sampling rate, the quantity of sorbent, the sorbent surface area, the density of active sites and bulk density, the volatility of the contaminant, and the concentration of contaminant in the air. Sampling pumps are limited to an accurate lower flow-rate of around 20 mL min^{-1}, resulting in a 10 L sample over 8 h. For many organic vapors, this sample volume can be collected even at flow-rates up to 1 L min^{-1}, without significant loss using glass tubes containing 100–400 mg sorbent in a primary adsorptive layer. A secondary back-up layer, typically containing half the sorbent of the primary layer is included to detect breakthrough of sampled material. A breakthrough of more than 20% of the total sample collected found in the back-up section indicates that some of the sample probably was lost. Optimum sample volumes are found in validated procedures.[5,6]

It is normally best to refer to an established procedure for proper sampling rates and air sample volumes. In the absence of such information, breakthrough experiments must be performed before field sampling is attempted. Normally, these experiments are conducted using dynamic standard atmospheres prepared at twice the EL or threshold limit value (TLV®) to give a suitable margin of safety to the measured breakthrough volume.[28] In addition, such tests should be performed at high humidities, or else a substantial correction factor should be used to account for the effects of atmospheric water vapor. Some sorbents such as carbon beads for methanol, 2,4-dinitrophenylhydrazine-coated silica gel for formaldehyde, and sulfuric acid-coated carbon beads for phosphine are less effective at low humidity, but this is less commonly encountered. Several methods for testing the collection efficiency of a sorbing device are available:

- by sampling from a dynamic standard atmosphere or from a gas-tight chamber or tank containing a known gas or vapor concentration (as mentioned);
- by series testing where enough samplers are arranged in series so that the last sampler does not recover any of the test gas or vapor;
- by comparing results obtained with a device known to be accurate;
- by introducing a known amount of gas or vapor into a sampling train containing the absorber being tested.

In addition to the collection efficiency it is also necessary to define the recovery efficiency and, as this frequently varies with chemical loading on the sorbent, the procedure normally involves dosing the sampler with varying quantities of the analyte. This experiment is repeated after allowing the chemicals to equilibrate with the sorbent for the longest period of practical storage time prior to analysis (2–3 weeks).

After the procedure has been validated, field sampling may be performed. Immediately before sampling, the ends of the sorbent tube are broken, rubber or Tygon® tubing is connected to the back-up end of the sorbent tube, and air is drawn through the sampling train with a calibrated pump. A personal breathing zone or area sample may be collected. The duration of the sampling may be several minutes or up to 8 h or longer, depending on the information desired. When sampling is completed, plastic, rather than rubber, caps are placed on the ends of the tube.

4.3.1 Choice of Desorption Procedure

An early drawback to using sorbent tubes for air sampling of organic compounds was the difficulty in recovering samples for analysis. Although the weight-difference before and after sampling could be used as an estimate of the pollution, this is a very crude and nonselective measure. A solvent can be used to displace the collected compounds, and an aliquot of the solvent can then be injected into a chromatograph to separate the components followed by a detector to quantify them. The ideal solvent should consist of small molecules to enable penetration of pores in the nanometre size range, a high heat of adsorption to ensure release of the collected compounds, and good solvating power for a wide range of organic compounds. In addition, it should be easily separated from the chemicals of interest, and not interfere with their analysis.

Alternatively, collected compounds can be released by the application of heat by a procedure known as thermal desorption. Unfortunately, some sorbents exhibit catalytic properties at high temperatures, and some tend to adsorb water from the air, which render them less suitable for thermal desorption, but other sorbents are available that are much more useful. Almost any single vapor could be collected on a sorbent and removed using either desorption procedure. The final choice is quite complex and depends on many factors, not least of which is the availability of the equipment.

4.3.1.1 Solvent Desorption The most frequently used solvent for liquid desorption, especially from charcoal (see section 4.3.2.1), is carbon disulfide. Carbon disulfide has many advantages. It penetrates even the smallest pores of charcoal and has a very high heat of adsorption. It is a good solvent for nonpolar compounds, although not so good for polar compounds. It does not react with most displaced compounds (except amines). Best of all, it is easily separated from most other compounds and has a reduced response to flame ionization detection (FID), although remnant ions will quench the response of closely eluting chlorinated hydrocarbons. Carbon disulfide does not always completely remove all compounds from charcoal. Recovery varies for each chemical and for each batch of charcoal used. The extent of individual recoveries must be determined experimentally and a correction for desorption efficiency applied to the analytical result.[29] Over a narrow range of analyte concentrations, as used in the NIOSH validations,[30] this desorption efficiency is essentially constant, but it may vary widely over larger concentration ranges, particularly for polar compounds.[31,32] Desorption efficiency can also be affected by the presence of water vapor and other contaminants.[33–35] The NIOSH[5] recommends that methods be used only where the desorption efficiency is greater than 75%; ideally, it should be greater than 90%. The recovery of polar compounds may be improved by the addition of a polar modifier (typically an alcohol) or by the use of an alternative solvent such as a mixture of methylene chloride and methanol, or dimethylformamide (either alone or in combination with carbon disulfide). Dimethylformamide has the advantage of being miscible with both carbon disulfide and any collected water. Although it has a significant FID response, because of its high boiling point (172 °C) and high polarity, it elutes in a region of the chromatogram with little else of interest, and thus does not interfere with most analyses.

The practical desorption step in charcoal analysis is also critical because, upon the addition of carbon disulfide to charcoal, the initial heat of reaction may drive off the more volatile components of the sample. This can be minimized by adding charcoal slowly to precooled carbon disulfide. Another technique is to transfer the charcoal sample to vials lined with Teflon® septum caps and to introduce the carbon disulfide with an injection needle. The sealed vial will prevent the loss of any volatilized sample. Headspace analysis is then possible.

It should be emphasized that carbon disulfide is a highly toxic solvent that produces serious effects on the cardiovascular and nervous systems. Care should be exercised in handling the solvent, and the analytical procedure should be performed in a well-ventilated area. Dimethylformamide has a very low vapor pressure and therefore presents less of a hazard from breathing, but there is also potential toxic hazard from dermal exposure.

Solvents for other sorbents are discussed under the sorbent type. Solvent desorption has the principal disadvantage of diluting the sample before analysis, but the principal advantage of allowing more than one analysis per sample. Solvent desorption requires only normal laboratory glassware and chemicals. It can be automated to handle high sample throughput.

4.3.1.2 Thermal Desorption Because of the high toxicity and flammability of carbon disulfide and the labour-intensive nature of the solvent desorption procedure, a useful alternative is to desorb the collected analyte thermally.[36–38] Except in a few cases, this is not practical with charcoal as sorbent because the temperature needed for desorption (e.g. 300 °C) would result in some catalytic decomposition of the analytes. Weaker sorbents are used instead, so larger tubes are required as more sorbent is required for efficient collection; usually 200–500 mg of sorbent, depending on type. Desorption can be made fully automatic, and analysis is usually carried out by GC. Some desorbers also allow automatic selection of sample tubes from a multiple-sample carousel. The whole sample can be transferred to the gas chromatograph, resulting in greatly increased sensitivity compared with the solvent desorption method. Alternatively, some desorbers allow the desorbed sample to be held in a reservoir from which aliquots are withdrawn for analysis, but then the concentrating advantage is reduced.

Because the least volatile components of the atmosphere are collected at the front of the tube, and the most volatile components penetrate further, sorbents tubes are normally desorbed in the reverse direction from sampling, or in back-flush mode. This becomes very important when combinations of sorbents are used in the tube. Although the sample could, in theory, be entered directly into the chromatograph upon desorption, this is rarely done in practice as broad peaks would result. Secondary trapping involves either adsorption on a smaller, secondary sorbent bed, or cryofocusing, or a combination of both. The equipment is expensive and represents a significant add-on to the standard cost of a gas chromatograph.

The main disadvantage of thermal desorption directly with an analyzer is that it is essentially a one-shot technique; normally, the whole sample is analyzed. This is why many such methods are linked to mass spectrometry. However, with capillary chromatography, it is usually possible to split the desorbed sample before analysis and, if desired, the vented split can be collected and reanalyzed.[39] Alternatively, the desorbed sample can be split between two capillary columns of differing polarity.[40]

Desorption efficiency is usually 100% for the majority of common solvents and similar compounds in a boiling range of approximately 50–250 °C.[41] However, if a wide boiling range is to be covered, more than one sorbent may be required. Thus, gasoline may be monitored by a Chromosorb 106® tube and carbon tube in series.[42] Extensive lists of recommended sampling volumes and minimum desorption temperatures for Tenax® and other sorbents are given by Brown and Purnell[38] and the UK Health and Safety Executive (HSE) Method MDHS (methods for the determination of hazardous substances) 72.[43]

Thermal desorption has been adopted as a (non-exclusive) recommended method for the determination of VOCs in the UK,[1] Germany,[2] and the Netherlands,[44] but it is less widely accepted elsewhere. The NIOSH[5] has relatively few methods based on thermal desorption (compared to those which use solvent desorption). The USEPA[7] has a number of methods based on thermal desorption and mass spectrometry for the lower concentrations found in ambient and indoor air. As regulated limits for occupational exposure reach 1 ppm or less (e.g. benzene, acrylonitrile, vinyl chloride, 1,3-butadiene, ethylene oxide, etc.) the potential of thermal desorption for analysis of a concentrated sample without dilution becomes more appealing.

4.3.1.3 Other Techniques for Desorption Microwave desorption is somewhat similar to thermal desorption except that it has some potential advantages based on the fact that microwave energy is selectively adsorbed by the chemicals of interest rather than by the sampling tube and sorbent. This efficiency of heating should lead to a faster plug transfer to the GC. In addition, there is a potential for using charcoal as the sorbent. Supercritical fluid extraction (SFE) is somewhat similar to solvent desorption except that the fluid is in its critical state and the desorption is carried out under pressure, ultimately leading to a better extraction efficiency. A further advantage is if a gas is used as the solvent, as it can be evaporated off to leave a concentrated sample. Unfortunately, supercritical carbon dioxide, the commonest solvent for SFE, does not give particularly good desorption efficiency of nonpolar chemicals from charcoal, unless doped with a nonpolar modifier (e.g. carbon disulfide). Other solvent systems have not been tried. Both techniques remain little more than curiosities at this time, although both show promise.

4.3.2 Solid Sorbents Used with Solvent Desorption

4.3.2.1 Activated Charcoal Charcoal is an amorphous form of carbon formed by partially burning wood, nutshells, animal bones, and other carbonaceous materials. A wide variety of charcoals are available; some are more suitable for liquid purification, some for decolorization, and others for air purification and air sampling. Ordinary charcoal becomes activated charcoal by heating it with steam to 800–900 °C. During this treatment, a porous, submicroscopic internal structure is formed that gives it an extensive internal surface area, as large as $1000\,m^2\,g^{-1}$ charcoal. This greatly enhances its sorption capacity. Activated charcoal is an excellent sorbent for most organic vapors. During the 1930s and 1940s, it was used in the then well-known activated charcoal apparatus[24] for the collection and analysis of solvent vapor. The quantity of vapor in the air sample was determined by a gain in weight of the charcoal tube. Lack of specificity, accuracy, and sensitivity of the analysis and the difficult task of equilibrating the charcoal tube, however, discouraged use. Renewed interest in activated charcoal as a sorbent for sampling organic vapors appeared in the 1960s.[29,45,46] The ease with which carbon disulfide extracts organic vapors from activated charcoal and the capability of microanalysis by GC are the reasons for its current popularity. Air sampling procedures using activated charcoal are widely used by industrial hygienists[47–49] and form the basis of the majority of the official analytical methods for VOCs recommended by the HSE,[1] NIOSH[5] and OSHA (Occupational Safety and Health Administration).[6]

Analytical information on selected NIOSH procedures is given in Table 1. In general, the NIOSH procedures use a 100 mg charcoal tube (with 50 mg back-up), but very volatile analytes may require a larger tube. The NIOSH studies showed that the charcoal tube method is generally adequate for hydrocarbons, halogenated hydrocarbons, esters, ethers, alcohols, ketones, and glycol ethers that are commonly used as industrial solvents. Compounds with low vapor pressure and reactive compounds (e.g. amines, phenols, nitro compounds, aldehydes, and anhydrides) generally have low desorption efficiencies from charcoal and require alternative sorbents such as silica gel or porous polymers for collection, or alternative reagent systems for recovery. Compounds containing a carbon–carbon double bond (e.g. vinyl compounds and acrylates) often show recoveries dependent on loading, which is partially a function of the presence of active sites of strong adsorption on the carbon surface. These sites may be stabilized by the presence of a free-radical inhibitor such as hydroquinone or *tert*-butyl catechol.

Inorganic compounds, such as ozone, nitrogen dioxide, chlorine, hydrogen sulfide, and sulfur dioxide, react chemically with activated charcoal and cannot be collected for analysis by this method.

Even for substances recommended for sampling on charcoal, this sorbent may not always be ideal. Reference

to Table 1 will indicate that carbon disulfide is the recommended desorption solvent for nonpolar compounds, whereas a variety of desorption cocktails are required for the more polar compounds. Difficulties arise, therefore, when sampling mixtures of polar and nonpolar compounds because each will give poor recoveries with the other's desorption solvent. One option is to use a two-phase carbon disulfide–water mixture for desorption with analysis of both phases.[50] Several alternative and more universal solvents have been investigated,[51–52] but

Table 1 NIOSH methods that give further information on sampling and analysis details for specific VOCs

Method name	Test compounds	Sorbent[a]	Desorption solvent	Method no.
Acetone cyanohydrin		QS	Ethyl acetate	2506[b]
Acetic acid		C	Formic acid	1603
Acetonitrile		C[c]	85 : 15 Methylene chloride–methanol	1606
Acrylonitrile		C	98 : 2 CS_2–acetone	1604
Alcohols I	*tert*-Butyl alcohol, isopropyl alcohol, ethanol	C	99 : 1 CS_2–2-butanol	1400
Alcohols II	*n*-Butyl alcohol, isobutyl alcohol, *sec*-butyl alcohol, *n*-propyl alcohol	C	99 : 1 CS_2–2-propanol	1401
Alcohols III	Allyl alcohol, isoamyl alcohol, methyl isobutyl carbinol, cyclohexanol, diacetone alcohol	C	95 : 5 CS_2–2-propanol	1402
Alcohols IV	2-Butoxyethanol, 2-ethoxyethanol, 2-methoxyethanol	C	95 : 5 CH_2Cl_2–methanol	1403
Allyl chloride		C	Benzene	1000
Amines Aliphatic	Diethylamine, diethylamine	S	dil. H_2SO_4 in 10% aq. methanol	2010
Amines Aromatic	Aniline, *o*-toluidine, 2,4-xylidine, *N*,*N*,-dimethyl-*p*-toluidine, *N*,*N*,-dimethylaniline	S	95% Ethanol	2002
Aminoethanol Compounds I	2-Aminoethanol, 2-dibutylaminoethanol, 2-diethylaminoethanol	S[d]	80% Methanol	2007
Anisidine		XAD-2[e]	Methanol	2514[f]
Butadiene		C[c]	Methylene chloride	1024
Butyl glycidyl ether		C	CS_2	1616
Carbon disulfide		C[g]	Toluene	1600[h]
β-Chloroprene		C	CS_2	1002
Cresols and phenol	*o*-Cresol, *m*-cresol, *p*-cresol, phenol	XAD-7	Methanol	2546
Dichlorodifluoromethane, etc.	Dichlorodifluoromethane, 1,2-dichlorotetrafluoroethane, chlorodifluoromethane	C[i]	Methylene chloride	1018
Dichlorofluoromethane		C[j]	CS_2	2516
1,1-Dichloro-1-nitroethane		C	CS_2	1601
Difluorodibromomethane		C[k]	2-Propanol	1012
Dimethylacetamide		S	Methanol	2004
Dimethyl sulfate		P	Diethyl ether	2524[l]
Dioxane		C	CS_2	1602
Diphenyl		T[m]	Carbon tetrachloride	2530
Epichlorohydrin		C	CS_2	1010
Esters I	*n*-Amyl acetate, *n*-butyl acetate, 2-ethoxyethyl acetate, ethyl acrylate, methyl isoamyl acetate, *n*-propyl acetate, isobutyl acetate, *sec*-amyl acetate, *sec*-butyl acetate, isoamyl acetate	C	CS_2	1450
Ethyl acetate		C	CS_2	1457

(*continued overleaf*)

Table 1 (*continued*)

Method name	Test compounds	Sorbent[a]	Desorption solvent	Method no.
Ethyl bromide		C	2-Propanol	1011
Ethyl chloride		C[c]	CS_2	2519
Ethylene chlorohydrin		C[n]	95 : 5 CS_2 – 2-propanol	2513
Ethylene dibromide		C	99 : 1 Benzene – methanol I	1008
Ethyl ether		C	Ethyl acetate	1610
Ethyl formate		C	CS_2	1452
Fluorotrichloromethane		C[c]	CS_2	1006
Furfuryl alcohol		Q	Acetone	2505
Glycidol		C	Tetrahydrofuran	1608
Hexachlorobutadiene		XAD-2	Hexane	2543[o]
Hexachloro-1,3-cyclo-pentadiene		T[p]	Hexane	2518[o]
Hydrocarbons BP 36–126°	Benzene, toluene, pentane, hexane, heptane, octane, cyclohexane, methylcyclo-hexane	C	CS_2	1500
Hydrocarbons Aromatic	Benzene, cumene, naphthalene, α-methylstyrene, styrene, vinyltoluene, *p-tert*-butyl-toluene, ethylbenzene, toluene, xylene	C	CS_2	1501
Hydrocarbons Halogenated	Chloroform, tetrachloroethylene, *o*- & *p*-dichlorobenzene, bromoform, methyl chloroform, chlorobromomethane, 1,2-dichloroethane, 1,1,1-trichloroethane, 1,1,2-trichloroethane, 1,2-dichloroethylene, tetrachloroethylene, carbon tetrachloride, ethylene dichloride, chlorobenzene, hexachloroethane, 1,2,3-trichloropropane	C	CS_2	1003
Isophorone		C[n]	CS_2	2508
Isopropyl acetate		C	CS_2	1454
Isopropyl ether		C	CS_2	1618
Isopropyl glycidyl ether		C	CS_2	1620
Ketones I	Acetone, cyclohexanone, diisobutyl ketone, 2-hexanone, methyl isobutyl ketone, 2-pentanone	C	CS_2	1300
Ketones II	Camphor, ethyl butyl ketone, mesityl oxide, 5-methyl-3-heptanone, methyl *n*-amyl ketone	C	99 : 1 CS_2 – methanol	1301
Methanol		S[q]	5% 2-Propanol in water	2000
Methyl acetate		C	CS_2	1458
Methyl acrylate		C	CS_2	1459
Methylal		C	Hexane	1611
Methyl cellosolve acetate	2-Methoxyethyl acetate	C	CS_2	1451
Methyl chloride		C[i]	Methylene chloride	1001
Methyl cyclohexanol		C	Methylene chloride	1401
Methyl cyclohexanone		Q	Acetone	2521
Methylene chloride	Dichloromethane	C[r]	CS_2	1005
Methyl methacrylate		XAD-2[c]	CS_2	2537
Methyl *tert*-butyl ether		C[c]	CS_2	1615
Methyl iodide		C	Toluene	1014
Naphthas	Kerosine, petroleum ether, rubber solvent, Stoddard solvent, etc.	C	CS_2	1550

Table 1 (*continued*)

Method name	Test compounds	Sorbent[a]	Desorption solvent	Method no.
Nicotine		XAD-2	Ethyl acetate	2544[b]
Nicotine		XAD-4	Ethyl acetate + 0.01% triethylamine	2551[b]
Nitroaromatic compounds	Nitrobenzene, nitrotoluene(s), 4-chloronitrotoluene	S	Methanol	2005
Nitroethane		XAD-2[s]	Ethyl acetate	2526
2-Nitropropane		106	Ethyl acetate	2528
Nitroglycerin and ethylene glycol dinitrate		T	Ethanol	2507
N-Methyl-2-pyrrolidone		C	95 : 5 Methylene chloride–methanol	1302
1-Octane thiol		T	Acetone	2510[b]
Pentachloroethane		R	Hexane	2517[o]
Phenyl ether		C	CS_2	1617
Phenyl ether/diphenyl mixture		S	Benzene	2013
Phenyl glycidyl ether		C	CS_2	1619
Propylene oxide		C	CS_2	1612
Propylene dichloride	1,2-Dichloropropane	C[n]	85 : 15 Cyclohexane–acetone	1013[l]
Pyridine		C	Methylene chloride	1613
Terpenes	Limonene, α-pinene, β-pinene, 3-carene	C	CS_2	1552
1,1,2,2-Tetrabromoethane		S	Tetrahydrofuran	2003
1,1,1,2-Tetrachloro-2,2-difluoroethane, etc.	1,1,1,2-Tetrachloro-2,2-difluoroethane, 1,1,2,2-tetrachloro-1,2-difluoroethane	C	CS_2	1016
1,1,2,2-Tetrachloroethane		C[n]	CS_2	1019
Tetrahydrofuran		C	CS_2	1609
Trichloroethylene		C	CS_2	1022
1,1,2-Trichloro-1,2,2-trifluoroethane		C	CS_2	1020
Trifluorobromomethane		C[i]	Methylene chloride	1017
Turpentine		C	CS_2	1551
Vinyl acetate		CMS	95 : 5 Methylene chloride–methanol	1453
Vinyl bromide		C[c]	Ethanol	1009
Vinyl chloride		C[k]	CS_2	1007
Vinylidine chloride	1,1-Dichloroethene	C	CS_2	1015

[a] Abbreviations are as follows: C, charcoal (100 + 50 mg tube); CMS, carbon molecular sieve (160 + 80 mg tube); P, Porapak P® (100 + 50 mg tube); Q, Porapak Q® (150 + 75 mg tube); QS, Porapak QS® (100 + 50 mg tube); R, Porapak R® (70 + 35 mg tube); S, Silica gel (150 + 75 mg tube); T, Tenax® (100 + 50 mg tube); 106, Chromosorb 106® (100 + 50 mg tube); XAD-2 (100 + 50 mg tube); XAD-4 (80 + 40 mg tube); XAD-7 (100 + 50 mg tube).
[b] Analysis by GC/Nitrogen/phosphorus detector.
[c] 400 + 200 mg tube.
[d] 300 + 150 mg tube.
[e] 150 + 75 mg tube.
[f] Analysis by high-performance liquid chromatography/ultraviolet.
[g] Used a sodium sulfate drying tube in front.
[h] Analysis by GC/flame photometric detection.
[i] Two tubes 400/200 + 100/50 mg.
[j] Two tubes 400/200 + 400/200 mg.
[k] Two tubes, 150 + 150 mg.
[l] Analysis by GC/conductivity (Hall).
[m] 20 + 10 mg tube.
[n] Petroleum-based carbon.
[o] Analysis by GC/electron capture detection.
[p] Two tubes 75 + 25 mg.
[q] 100 + 50 mg tube.
[r] Two tubes 100/50 + 100/50 mg.
[s] 600 + 400 mg tube.

none of these has achieved wide recognition. In such circumstances, it may be necessary to take two samples and desorb each one with a different solvent.

4.3.2.2 Silica Gel Silica gel is an amorphous form of silica derived from the interaction of sodium silicate and sulfuric acid. It has some advantages over activated charcoal for sampling certain gases and vapors. Polar contaminants, including alcohols, amines, amides, and phenols are more easily removed from the sorbent by polar solvents such as water, methanol or acetonitrile, and some inorganic substances, including acid gases and vapors (e.g. HF, HCl, formic and acetic acids) for which charcoal is unsuitable can be collected.

One disadvantage of silica gel is that it will adsorb water. Silica gel is electrically polar, and polar substances are preferentially attracted to active sites on its surface. Water is highly polar and is tenaciously held. If enough moisture is present in the air or if sampling is continued long enough, water will displace organic solvents (which are relatively nonpolar in comparison) from the silica gel surface. With water vapor at the head of the list, compounds in descending order of polarizability are alcohols, aldehydes, ketones, esters, aromatic hydrocarbons, alkenes, and alkanes. It is obvious, therefore, that the volume of moisturized air that can be effectively passed over silica gel is limited.

In spite of this limitation, silica gel has proved to be an effective sorbent for collecting many gases and vapors. Even under conditions of 90% humidity, relatively high concentrations of benzene, toluene, and trichloroethylene are quantitatively sorbed on 10 g of silica gel from air samples collected at the rate of 2.5 L min^{-1} for periods of at least 20 min or longer.[53,54] Under normal conditions, hydrocarbon mixtures of 2–5-carbon alkanes, low-molecular-weight sulfur compounds (H_2S, SO_2, mercaptans), and alkenes concentrate on silica gel at dry ice–acetone temperature if the sample volume does not exceed 10 L.[55] Significant losses of ethylene, methane, ethane, and other light hydrocarbons occur if the sampling volume is extended to 30 L.

4.3.2.3 Polymeric Sorbents A variety of organic polymers have been used for sample collection followed by solvent desorption. The commonest are polymers derived from styrene, often with admixture from similar aromatic compounds such as ethyl vinyl benzene or divinyl benzene. These can be obtained from several different manufacturers, and are available under the manufacturers' trade names (e.g. Amberlite®, Chromosorb®, Porapak®, Hayesep®). These are generally either microporous (surface areas of 600–800 $m^2 g^{-1}$), such as Amberlite® XAD-4, Chromosorb 106®, Porapak Q® or Hayesep D®, or mesoporous (surface areas of 200–400 $m^2 g^{-1}$), such as Amberlite® XAD-2 or Chromosorb 102®. The latter group are commonly used for the collection of large-molecular-weight, semivolatile organic compounds, including pesticides[56] and polyaromatic hydrocarbons,[57] whereas the former are used for VOCs.[58] These polymers are compatible with most solvent systems. Benzene, cyclohexane, hexane, acetonitrile, methylene chloride, methanol, diethyl ether, and carbon disulfide have all been used, either alone or in combination, depending on the analytes being sampled. These sorbents are hydrophobic so that sampling, storage, or recovery of organic compounds are not affected by ambient water vapor.

The polymer Tenax® has also been used for sampling certain compounds such as dinitrotoluene and trinitrotoluene (TNT) with subsequent solvent desorption,[59] although it is more normally used with thermal desorption.

4.3.2.4 Molecular Sieves Molecular sieves are sorbents tailored to have uniform pore dimensions. The name is usually referred to the silicate minerals known as zeolites, and carbons made from the carbonization of polymer precursors. In general use, very small pores in the low-nanometer range are implied. The surface area of these ultramicroporous materials is very high, and retention is enhanced by three-dimensional interactions with the walls of the pores and with nearby molecules. Thus small, highly volatile molecules can be retained, but molecules larger than the pore dimensions will not enter the sorbent (hence the sieve effect). The main disadvantage of these sorbents is that they adsorb water molecules, which also fit into the pores. Although charcoal is technically hydrophobic, the adsorption of even a few water molecules causes further molecules to bind to them by hydrogen bonding. Carbon molecular sieves are found in several OSHA methods[6] for methylene chloride, acetone, 2-butanone and vinyl chloride.

4.3.3 Sorbents Used with Thermal Desorption

Graphitized carbons and carbon molecular sieves may be used with thermal desorption, either alone or in combinations. Porous polymer sorbents (e.g. Tenax®, Porapak Q® and Chromosorb 106®) are also used[60–62] alone, or, in the case of Tenax®, which has a high thermal stability, in combination with carbon sorbents.

4.3.3.1 Polymeric Sorbents The same polymers mentioned under solvent desorption are used in thermal desorption procedures. As thermal desorption has been focused traditionally on collecting samples from relatively clean environments (indoor and ambient air quality investigations), there has been much concern over blank levels

in the sorbent and how low blank levels can be maintained during storage and transportation. Tenax® has been shown to have the lowest blank levels if conditioned properly, and stored appropriately. Other issues of sampling low concentrations include potential interference from artifact formation. Tenax® has been shown to produce oxygenated aromatics such as acetophenone, benzophenone, benzaldehyde, and phenol through reaction with ambient ozone. Nevertheless, Tenax® is one of the most popular sorbents for sampling atmospheres of relatively low concentration (low ppm to ppb total volatile organic compounds (tVOC)).

Workplace atmospheres typically contain higher concentrations (1–1000 ppm). Background becomes less of a problem when sampling these concentrations but capacity more so. Tenax® has a very low surface area and small capacity, and may be overwhelmed by high concentrations. Unfortunately, as thermal desorption tubes are often used without any form of back-up, such breakthrough may go unnoticed. Chromosorb 106® has much higher capacity and is generally regarded as the best all-round sorbent for occupational hygiene investigations.[63]

An alternative solution to the problem of capacity is to use tubes containing multiple sorbent beds. The sorbent with the least capacity for very volatile compounds is the first sorbent through which the air passes. Although the volatile compounds will pass through this bed, less-volatile compounds will be adsorbed, and on a sorbent from which they can be released easily. Sorbents are arranged in order of increasing strength of adsorption so that the final bed adsorbs the most volatile compounds. Desorption must take place in the opposite direction for best results. Typically two or three beds are used. The sorbents used must be thermally compatible with each other. In practice, Tenax® or Tenax® GR (a composite of Tenax® and graphitized carbon) are the only polymers with sufficient thermal stability to be placed alongside carbonaceous sorbents.

The use of multibed tubes is relatively new, and issues of collection efficiency and sample storage stability still require careful evaluation.

4.3.3.2 Graphitized Carbons Graphitized carbons are granular forms of amorphous graphitized carbon black. Several are available with different surface areas, such as Carbotrap (surface area around $100\,m^2\,g^{-1}$) or Carbotrap C (surface area around $10\,m^2\,g^{-1}$). They can be conditioned to have low blank levels, and they are compatible with the highest temperatures normally employed in desorption. Their main drawback is that they are friable, and dust caused by poor handling can increase the pressure-drop across the tube, or escape to block capillary lines. In addition, catalytic reactions have been observed with some reactive compounds such as terpenes. Because of their thermal compatibility with each other and with carbon molecular sieve adsorbents, they are frequently used in multibed tubes such as that designed for the semiquantitative screening method 2549 published by the NIOSH[5] or those developed for use by the USEPA.[64]

4.3.3.3 Molecular Sieves Carbon molecular sieves are often used as the final stage of two- or three-stage sorbent tubes to collect the most volatile components. The water that they also collect requires management. One method is to dry the sorbent tube by the passage at ambient temperature of around 300 mL of dry carrier gas prior to desorption. This has the benefit of removing water without disturbing the collected organic vapors.[65]

4.4 Coated Sorbents and Filters

Many highly reactive compounds are unsuitable for sampling directly onto sorbents, either because they are unstable or cannot be recovered efficiently. In addition, some compounds may be analyzed more easily, or with greater sensitivity, by derivatizing them first, which can sometimes be achieved during the sampling stage. Typical common examples are the collection of ethylene oxide as bromoethanol with hydrobromic acid-coated charcoal tubes, and the collection of aldehydes as their derivatives with hydroxmethyl piperidine or 2,4-dinitrophenylhydrazine. Reaction rates may be slow so that lower sampling rates may be required to provide greater contact time between the air being sampled and the sorbent. The reaction may also require specific pH or humidity levels for completion. Coated sorbents and filters have become very popular in recent years as a replacement for sampling with liquid absorbers which have problems with breakage, spillage, or liquid evaporation.

4.4.1 Coated Sorbents

The first sorbent available with a coating was silica gel coated with an inorganic acid for more efficient collection of organic amines and ammonia. Others are given in Table 2. The most commonly encountered are those used for formaldehyde. Charcoal may be coated to restrict reaction at active sites (see section 4.3.2.1).

4.4.2 Coated Filters

When reaction rates are sufficiently fast, the depth of a glass-fiber filter whose fibers are coated with a thin layer of reagent is adequate for efficient sample collection. In a standard closed-face three-piece cassette a second filter may be incorporated to detect breakthrough. If it is desired to use the cassette open face, either a second

Table 2 Collection and analysis of gases and vapors (coated sorbents and filters)

Test compounds	Coating reagent	Matrix[a]	Method No.
Acetic anhydrides	Veratrylamine	F	OSHA 102
Aldehydes	2-(Hydroxymethyl)piperidine	T	NIOSH 2501, 2539, 2541, OSHA 52, 68
Aldehydes	2,4-Dinitrophenylhydrazine	T, F	NIOSH 2016, OSHA 64
Amines	NBD[b] chloride	T	OSHA 34, 41
Amines	NITC	T	NIOSH 2540, OSHA 60
Ammonia	Sulfuric acid	T	NIOSH 6016, OSHA ID 188
Aromatic amines	Sulfuric acid	F	NIOSH 2017, OSHA 65, 71, 73, 78, 87, 93, 105
Arsenic trioxide, fluorides, sulfur dioxide	Sodium carbonate	F	NIOSH 7901, 7902, 7906, 6004, OSHA ID 110
Cyanoacrylates	Phosphoric acid	T	OSHA 55
Diisocyanates	1-(2-Pyridyl)piperazine	F	OSHA 42
Ethylene oxide	Hydrobromic acid	T	NIOSH 1614, OSHA 50
Hydrazine	Sulfuric acid	F	OSHA 108
Iodine, phosphine, sulfur dioxide	Potassium hydroxide	T	OSHA ID 177, 180, 200, 212

[a] T, sorbent tube; F, filter.
[b] NBD, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole; NITC, naphthyl isothiocyanate.

ring-piece may be added for the extra filter, or two cassettes can be used in series. Although the range of available coated filters is quite large, the most popular methods are those for hydrogen fluoride, glutaraldehyde, aromatic amines, and isocyanates. Coated filters are able to collect particulates less than 10 μm in diameter, in addition to vapors, so are useful when the contaminant may be present in a mixed aerosol/vapor phase (e.g. isocyanates in spray-painting). Commonly used coated filters are listed in Table 2.

5 DIFFUSIVE SAMPLERS

A diffusive sampler is a device that is capable of taking samples of gas or vapor pollutants from the atmosphere at a rate controlled by a physical process, such as diffusion through a static air layer or permeation through a membrane, but does not involve the active movement of the air through the sampler. A number of publications are available which review the use of diffusive samplers for workplace air sampling.[66–72] It should be noted that in the USA, the adjective ''passive'' is sometimes used in describing these samplers and should be regarded as synonymous with ''diffusive''.

This type of diffusive sampler should not be confused with the annular or aerosol denuders, which not only rely on diffusion to collect the gas or vapors but also upon the air in question being simultaneously drawn through the annular inlet into the sampler. Aerosol particles have diffusion coefficients too low to be collected on the annular inlet and are trapped on a back-up filter.

5.1 Operating Principles

The mass of the analyte which can diffuse to a suitable sorbent within a certain time is determined by Equation (1) which is derived from Fick's first law of diffusion:

$$m_s = \frac{AD(\rho_1 - \rho_2)t}{l} \quad (1)$$

where A is the cross-sectional area of diffusion path (cm^2), D is the coefficient of diffusion ($cm^2\ s^{-1}$), l is the length of diffusion path (cm), m_s is the mass of analyte sorbed by diffusion (ng), t is the sampling time (s), ρ_1 is the actual mass concentration at the beginning of the diffusion layer ($l = 0$) ($mg\ m^{-3}$), and ρ_2 is the actual mass concentration at the end of the diffusion layer ($mg\ m^{-3}$).

Ideally ρ_1 is equal to the concentration of the given analyte in the air outside the diffusive sampler (ρ), and ρ_2 equals zero (zero-sink condition). In that case the magnitude of the diffusive uptake rate, $A \times D/l$, is dependent only on the diffusion coefficient of the given analyte and on the geometry of the diffusive sampler used.

For some combinations of sampler and contaminant there are factors that can give rise to nonideal behaviour, so that (Equation 2)

$$m_s = \frac{AD\rho tk}{l} \quad (2)$$

where k is the correction factor for nonideal behavior and ρ is the actual mass concentration of analyte in air ($mg\ m^{-3}$).

5.1.1 Dimensions of Diffusive Uptake Rate

For a given concentration ρ in milligrams per cubic metre of gas or vapor, the diffusive uptake rate is given by Equation (3),

$$U = \frac{m_s}{\rho t'} \tag{3}$$

where U is the sampling rate ($cm^3\,min^{-1}$) and t' is the sampling time (min).

Although the uptake rate U has dimensions of volume per unit time, this does not indicate a real volumetric flow of air containing the contaminant.

Diffusive uptake rates are sometimes quoted in units of $ng\,ppm^{-1}\,min^{-1}$. These are practical units, as most environmental analysts use ppm for concentrations of gases and vapors. The dependency of uptake rates on temperature and pressure is explained later. Thus for a given concentration (ppm) of gas or vapor, the sampling rate is given by Equation (4),

$$U' = \frac{m_s}{\phi t'} \tag{4}$$

where U' is the sampling rate ($ng\,ppm^{-1}\,min^{-1}$) and ϕ is the volume concentration of analyte in air (ppm, v/v).

Ideal and practical diffusive uptake rates are related by Equation (5),

$$U' = \frac{UMT_0P_1}{MVT_1P_0} \tag{5}$$

where M is the molecular weight of the analyte, MV is the molar volume at normal temperature and pressure (NTP) conditions, T_0 is the normal temperature (typically 298 K in the USA, 293 K in the UK), T_1 is the temperature at the sampling site, P_0 is the normal pressure (typically 101 kPa), P_1 is the pressure at the sampling site.

Note that this equation does not fully take into account the temperature dependence of the diffusion coefficient (see section 5.2.1).

5.1.2 Bias due to the Selection of a Nonideal Sorbent

The performance of a diffusive sampler depends critically on the selection and use of a sorbent or collection medium which has high sorption efficiency. The residual vapor pressure of the sampled compound at the sorbent surface (ρ_2) will then be very small in comparison to the ambient concentration, and the observed uptake rate will be close to its ideal steady-state value, which can usually be calculated from the geometry of the sampler and the diffusion coefficient of the analyte in air.

In the case where a weak sorbent is used, then ρ_2 in Equation (1) is nonzero and m_s/t will decrease with the time of sampling. Hence U in Equation (2) will also decrease with the time of sampling. The magnitude of this effect is dependent on the sorption isotherm of the analyte and sorbent concerned, and may be calculated with the aid of computer models.[73,74]

Another manifestation of the same effect is back-diffusion, sometimes called reverse diffusion. This can happen where, some time after sampling has started, the vapor pressure of the analyte at the sorbent surface, ρ_2, is greater than the external concentration, ρ_1, for example if a sampler is first exposed to a high concentration and then to a much lower or even zero concentration. This type of exposure profile can occur in certain applications, and the magnitude of any error introduced will depend on whether the period of high concentration occurs at the beginning, middle, or end of the sampling period. The phenomenon has been discussed in detail by Bartley and colleagues[75–77] and a simple test proposed[78] to give an estimate of the maximum bias to be expected between a pulsed exposure and an exposure to a constant concentration, which latter normally provides the basis for the sampler calibration. The extent of back-diffusion can also be modelled theoretically.[73,79]

It is therefore desirable to choose a sorbent with high sorption capacity and low vapor pressure of the sorbed material or of the reaction product formed by a reactive sorbent.

5.2 Environmental Factors Affecting Sampler Performance

5.2.1 Temperature and Pressure

For an ideal diffusive sampler, the dependence of U on absolute temperature and pressure is governed by that of the diffusion coefficient of the analyte. The latter dependence is given by Equation (6),

$$D = f(T^{n+1}, P^{-1}) \tag{6}$$

with $0.5 < n < 1.0$.

Hence, the dependence of U, expressed in units of $cm^3\,min^{-1}$ or equivalent is (Equation 7),

$$U = f(T^{n+1}, P^{-1}) \tag{7}$$

when U' is expressed in units of $ng\,ppm^{-1}\,min^{-1}$ or equivalent by application of Equation (6), then the dependence is given by Equation (8):

$$U' = f(T^n) \tag{8}$$

In the latter case, the dependence will be of the order of 0.2–0.4% K^{-1}. In the case of a nonideal sampler, the temperature dependence of U' may be compensated by the temperature dependence of the sorption coefficient

of the analyte.[80] In any case, accurate knowledge of the average temperature and pressure during the sampling period is important for a correct application of Equations (3) and (4).

5.2.2 Humidity

High humidity can affect the sorption capacity of hydrophilic sorbents, such as charcoal and molecular sieve. This will normally reduce the sampling time (at a given concentration) before saturation of the sorbent occurs, when sampling becomes nonlinear because of a significant ρ_2 term in Equation (1). High humidity can also alter the sorption behavior of the exposed inner wall of tube-type samplers or draught screen, particularly if condensation occurs.

5.2.3 Transients

Simple derivations of Fick's law assume steady-state conditions, but in the practical use of diffusive samplers, the ambient level of pollutants is likely to vary widely. The question then arises as to whether a sampler will give a truly integrated response (ignoring sorbent effects) or will miss short-lived transients before they have had a chance to be trapped by the sorbent. The passage of molecules into the diffusion path, followed by their diffusion back out of the sampler when the external concentration has dropped to zero and before the molecules have reached the sorbent bed has also been termed 'reverse diffusion'. Although the effects are similar the cause is different from that referred to in the section on nonideal behavior listed above. In that case the cause is related to the strength of the sorbent, in this case to the geometry of the sampler. The problem has been discussed theoretically[75,81–83] and practically[81,84,85] and shown not to be a problem, provided the total sampling time is well in excess of (say 10 times) the time constant of the diffusive sampler, i.e. the time a molecule takes to diffuse into the sampler under steady-state conditions. The time constant τ for most commercial samplers is between about 1 and 10 s and is given by (Equation 9):

$$\tau = \frac{l^2}{D} \tag{9}$$

5.2.4 Air Velocity

5.2.4.1 Effect of Low and High Wind Speeds Ambient air face velocity and orientation can affect the performance of a diffusive sampler because they may influence the effective diffusion pathlength.[86–89] The diffusive mass uptake of a sampler (Equation 2) is a function of the length l and the cross-sectional area A of the diffusion gap within the sampler. The nominal diffusion pathlength is defined by the geometry of the sampler and is the distance between the sorbent surface and the external face of the sampler. The cross-sectional area is also defined by the geometry of the sampler and if the cross-section of the diffusion gap is not constant along its length, is defined by the narrowest portion. The effective length, l, is not necessarily the same as the nominal length, and may be greater or less, depending on circumstances.

Under conditions of low external wind speeds, the effective diffusion pathlength may be increased.[88,89] This is because a boundary layer[86,87] exists between the stagnant air within the sampler and the turbulent air outside and contributes to the effective diffusion pathlength l. In reality, there is an area outside the sampler where there is a transition between static air and turbulent air, but this is equivalent to an extra length (δl) of static air which must be included in the value of l. The value of δl depends on the external geometry of the sampler, being roughly proportional to the linear cross-section of the sampler collection surface, where this surface is flat. It also decreases with increasing air velocity. Its significance depends on the value of the nominal pathlength of the diffusive sampler. Thus a sampler with a small cross-section and long internal air gap will be relatively unaffected by air velocity, whereas a short, fat sampler will be significantly affected. This is borne out in practice, as has been demonstrated with samplers of varying length.[88,89] Low sampling rates are observed at low air velocities, but increase to a plateau value as the boundary layer effect becomes insignificant.

Under conditions of high external wind speeds, the effective diffusion pathlength may be decreased.[90–96] This is because turbulent air disturbs the static air layer within the sampler, which reduces the effective air gap by a factor δl. The value of δl is small, provided the length to diameter ratio of the sampler air gap is greater than 2.5–3,[90] or it can be avoided, or greatly reduced, by incorporating a draught shield, such as a stainless steel screen or plastic membrane.

The overall effect is therefore sinusoidal.

5.2.4.2 Consequence for Different Sampler Geometries Tube-type samplers are typically unaffected by low air velocities[81,97,98] but those without a draught shield may be affected by high speeds.

Badge-type samplers generally have a large surface area and small air gap, so that they may be more affected by air velocity than tube designs and typically require a minimum face velocity of between 0.2 and 0.5 m s^{-1}.[67,99–101] Some badges with an inadequate draught shield are also affected at high air velocities.[98,99,102]

Radial diffusive samplers[103] require a minimum face velocity of about 0.25 m s^{-1}.

5.2.5 Sources of Methods and Uptake Rates

In general, the regulatory authorities have been reluctant to accept diffusive monitoring methods, except in some countries, for example the UK and the Netherlands.[1,44] Extensive lists of recommended sampling rates for solvent desorption methods are given in *The Diffusive Monitor*[104] and the UK HSE Method MDHS 88.[105] Extensive lists of recommended sampling rates for thermal desorption methods are given in *The Diffusive Monitor*[106] and the UK HSE Method MDHS 80.[107] The situation with regard to regulatory acceptance is changing slowly. In the USA, the OSHA Technical Center has begun a program of sampler validation and has published the first of a series of methods (Method 111, toluene) featuring diffusive samplers.

6 SAMPLING DEVICES WITH DIRECT READOUT

There are two types of sampling devices which provide immediate or on-site read-out covered in this section, and a third that combines elements of the other two. In the first type, colorimetric diffusive samplers, uptake is by diffusion (see section 5) and a progressive reaction of an air contaminant with a chemical coating on the substrate develops a color that may be read by eye or by a portable instrument. In the second type, detector tubes, air is pulled through a chemically coated sorbent in a graduated tube. The concentration of a contaminant is related to the length of stain caused by the reaction between the contaminant and the sorbent coating. These tubes are typically for short-term almost instantaneous measurements, but diffusive types are also available that can be used for long-term measurements. This type has elements of both of the other types. All of these devices are especially suitable for simple, reactive contaminants that have few interferences. Examples of such chemicals include phosgene, ammonia, chlorine, and formaldehyde. The detector tube devices can be useful for more complex chemicals or mixtures by incorporation of multiple layers, including sections to humidify or dry the sample, activate the chemical, or to remove interferences (precleanse layer). Detector tubes are very useful for identifying hazardous atmospheres but are less often used for exposure monitoring because of their short sampling periods and the high cost of multiple samples to cover a longer period. Diffusive devices are popular with large industries which have performed in-house validations and are satisfied with the sampler performance, but only very few have been tested by government agencies. A very useful reference work on these devices has been published.[108] A manual of proper practice for detector tubes also is available.[109] The theoretical aspects of sampling with diffusive samplers (section 5) extends also to those samplers that develop a color. Measurement of the developed color can be done qualitatively, by comparison with colored standards, or quantitatively using a spectrophotometer, normally in reflectance mode. The propagation of light within an opaque medium is a very complex process that can be solved mathematically if certain assumptions are made, and the resulting calibration is inherently nonlinear. However, there is usually a narrow range of concentration over which the calibration can be considered linear, and this can be extended by appropriate mathematical transformation of the raw data. Relatively inexpensive devices are available for this purpose. Where the color develops as a length of stain, as is the case with detector tubes, the situation is somewhat different, and is described below.

6.1 Operation of Detector Tubes

The two ends of the tube are broken by a device provided by the manufacturer which ensures an opening of sufficient width not to limit flow into the tube. The tube is then placed in the manufacturer's holder which is fitted with a squeeze bellows or piston pump for short-term grab samples, or connected to a pump for long-term samples. Diffusive-type detector tubes are placed in their own special holder. The correct amount of air is drawn through the short-term samplers by applying the appropriate number of strokes of the bellows or piston pump, each stroke being typically 100 mL. Care must be taken to allow adequate time for the vacuum to reach equilibrium with the ambient pressure, and also to make sure all connections are leak-tight. Manufacturer's instructions should be followed exactly; most manufacturers and applicable standards advise against substituting pumps from different manufacturers, as the pump and tube are certified as a unit.[110,111] The number of strokes varies from one to as many as 100, and the number usually can be increased to increase sensitivity. Some pumps are equipped with stroke counters to ensure the correct number is taken. One hundred strokes usually require a considerable time to accomplish and such samples are time-weighted averages. Short-term detector tubes can be used to determine time-weighted average concentrations for up to 8 h or more, provided a sufficient number of samples is taken, but this will usually not prove to be cost-effective. Long-term tubes are used in the same way as sorbent tubes; diffusive versions have a defined diffusion path. In all cases, the length of color change is read against a calibration marked on the outside of the tube. The type of color change depends on the reaction chemistry, and may be more or less easy to determine by eye. Detector tubes are carefully made to provide

sharp reaction fronts where possible, but sometimes the boundary is diffuse and difficult to identify. Reading the concentration value is the single largest source of error associated with these devices, and the accuracy may not reach the requirements normally required for monitoring for compliance purposes. Recently, electronic devices have been made available to reduce the variation associated with reading tubes by eye.

6.2 Chemical Reactions used in Detector Tubes

Several useful chemical reactions are widely employed in detector tubes. These reactions include reductions (e.g. of chromate or iodine pentoxide), color change of pH indicators, and reactions to form dyes (e.g. with *o*-tolidine or tetraphenylbenzidine). However, these reactions are often not specific to a single contaminant, and cross-sensitivity must be taken into account when using these devices. Sometimes the same reaction may be involved in two different tubes, only the calibration being different. The reacting chemicals are impregnated on inert supporting material, such as silica gel, alumina, ground glass, pumice, or resin. Precleanse layers may be used to remove water vapor or other interferences (e.g. hydrogen chloride when measuring hydrogen cyanide). In addition, there may be reactive layers that produce a product that is measured by the color reaction (e.g. liberating halogen from a halogenated hydrocarbon). Because many of these chemicals are very reactive, long-term storage may not be possible. A shelf-life of at least 2 years is desirable; in general, shelf-life may be improved by refrigerated storage.

6.3 Theory of Operation

In the general case where the sampling period is relatively short and the flow-rate is relatively high, complete equilibrium is not reached between the contaminant gas or vapor and the chemical reactants, and the length of stain is determined by the kinetic rate of reaction. In theory,[112] the stain length is proportional to the logarithm of the product of gas concentration and sample volume (Equation 10):

$$\frac{L}{H} = \ln(C'V) + \ln\frac{K}{H} \tag{10}$$

where L is the stain length (cm), K is a constant that varies with type of tube and analyte gas or vapor (cm^{-2}), and H is the mass transfer proportionality factor (cm). Thus a plot of stain length L against the logarithm of the product $C'V$ yields a straight line of slope H.

If the reaction rate is sufficiently fast that equilibrium can be attained, the general Equation (10) reduces to

$$L = KC'V \tag{11}$$

Equation (11) generally holds true for long-term detector tubes, where the concentration is calculated from the volume sampled. For stain-length passive (diffusive) dosimeters, Fick's law (Equation 1) is combined with Equation (11) to give Equation (12):

$$rL + L^2 = KDAC't \tag{12}$$

where r is the diffusive pathlength (cm).

6.4 Certification of Detector Tubes

A performance evaluation program of detector tubes was initiated by the NIOSH in the late 1960s and early 1970s. Many tubes were found at that time to have accuracies in the range ±25% to ±35%. A formal certification program, with testing by the NIOSH in Morgantown, WV, was put in place,[113,114] but was dropped 10 years later for lack of funding. In 1986, the Safety Equipment Institute (SEI) announced a voluntary program for third-party certification,[115] which continues today. However, in a recent air sampling instruments workshop,[116] the program was criticized mainly on the grounds of having too narrow a scope of testing. Detector tubes are the subject of many national and international standards; some examples are given from the USA, Japan, and the UK.[117–119]

7 CALCULATIONS

Prior to calculating concentration from a sampling result, it is important to keep in mind the use to which the result will be put. The results of most sampling exercises are to be compared to a regulated or recommended limit value. The ultimate source of most of the limit values in use around the world are the TLV®s of the American Conference of Governmental Industrial Hygienists (ACGIH). These values are given in units of parts per million by volume (ppm) at NTP (or 25 °C, one atmosphere, in the USA). The results obtained from pumped sampling exercises are in units of mass per unit volume (e.g. $mg\,m^{-3}$), which can be compared to the limit value by using the molar volume of the chemical at NTP in the calculation. The handling of results from diffusive sampling exercises is more complex, as the diffusion coefficient varies with pressure and temperature, whereas, as explained above, the volume of a pump does not. In order to compare a result to the ACGIH TLV® it is necessary either to correct the uptake rate (which is normally given for NTP) for pressure and temperature and use the molar volume at NTP, or to use the molar volume at the sampling site conditions of pressure and temperature. This procedure is to be used, for example, for comparing results to OSHA permissible exposure limits (PELs). Other

government regulations may vary. These corrections are not insignificant. At an altitude of 6000 ft (approximately 2000 m) the correction is around 17%. Outdoors during freezing weather the correction is around 12%.

The collected sample is analyzed, either directly if it is a gas-phase or impinger sample, or after desorption if it is collected on a solid sorbent, using appropriate gas or liquid standard solutions to calibrate the analytical instrument. The mass concentration of the analyte in the air sample is then calculated using the following equations.

7.1 Absorber

This is given by Equation (13):

$$C = \frac{m - m_{\text{blank}}}{E_s V} \tag{13}$$

where C is the mass concentration of analyte in the air (mg m^{-3}), m is the mass of analyte in the sample (μg), E_s is the sampling efficiency, m_{blank} is the mass of analyte in blank (μg), and V is the volume of air sample (litres).

7.2 Sorbent Tube

This is given by Equation (14):

$$C = \frac{m_1 + m_2 - m_{\text{blank}}}{E_d V} \tag{14}$$

where m_1 is the mass of analyte in first tube section (μg), m_2 is the mass of analyte in back-up tube section (if used) (μg), and E_d is the desorption efficiency corresponding to m_1.

7.3 Volume Fraction

If it is desired to obtain a result in ppm at NTP that can be compared directly to an ACGIH TLV® or OSHA PEL, Equation (15) is used:

$$C' = C\frac{MV}{M} \tag{15}$$

7.4 Diffusive Sampler

The method of calculation of atmospheric concentrations is essentially the same as for pumped samplers, i.e. the collected sample is analyzed and the total weight of analyte on the sampler is determined (Equation 16):

$$C = \frac{m_1 + m_2 - m_{\text{blank}}}{E_d V} \tag{16}$$

Note that m_2 is relevant only to samplers with a back-up section, and an additional multiplication factor may be needed to account for differing diffusion pathlengths to primary and back-up sections.

The total sample volume V is calculated from the product of the effective sampling rate (L min^{-1}) and the time of exposure (min). The sampling rate should be corrected for the ambient temperature and pressure as stated above. This calculation gives C' in ppm at NTP:

$$C' = \frac{C \times \text{MV}}{M(T_1/T_0)^{1.5}(P_0/P_1)} \tag{17}$$

Equation (17) provides results that can be directly compared to an ACGIH TLV® or OSHA PEL.

8 STANDARDS

8.1 European Standards

The task of developing appropriate standards for workplace air quality measurements within the European Community has been carried forward by working groups (WGs) of CEN (European Committee for Standardization) Technical Committee TC 137. This committee took the view that air quality assessment standards should take the form of performance requirements rather than prescribed methods. This approach has the advantage of allowing any method to be used which meets these requirements without stifling innovation and development. The actual writing of standard measurement procedures was taken to be the role of ISO (International Organization for Standardization) or member state regulatory bodies, such as HSE (for example the MDHS series).

The approach taken by WG2 has been to develop a hierarchy of standards with a general performance requirements document at the top and a series of specialized standards under this umbrella.

8.1.1 Umbrella Standard

The umbrella standard[120] provides, among other things, definitions and minimum requirements for unambiguity (the uniqueness of the result), selectivity (which depends on whether detailed knowledge of the air composition is known in advance) and overall uncertainty (a combination of precision and bias).

8.1.2 Second-tier Standards

These standards contain specific (minimum) performance requirements for measuring devices, together with the appropriate test methods. So far, the standards for diffusive samplers,[121] pumped sorbent tubes,[122] detector tubes,[123] and low-volume sampling pumps[124] have been published. Standards are being prepared for high-volume pumps, sampling and analysis of metal species (or, more generally, of chemical agents in airborne particles), samplers for mixed aerosols and vapors and, jointly with a

CENELEC (European Electrotechnical Committee for Standardization) WG, for direct-reading electrical apparatus. Guidance is also being prepared for chemically impregnated systems and for the selection of procedures and devices.

8.2 US Standards

8.2.1 Government Standards

Agency standards include those issued by the NIOSH, OSHA, and the USEPA. The NIOSH has published a manual on sampling strategy[125] and also guidance for method development and evaluation.[126] In addition, the introduction to the *Manual of Analytical Methods*[5] contains much valuable information. NIOSH methods are considered to be guidance, rather than regulations, and may be modified as necessary, provided suitable evidence of equivalent performance has been documented. OSHA inspectors and analytical laboratories are required to use OSHA methods where possible, and NIOSH or other methods where these do not exist. There is no obligation on private parties to use OSHA or NIOSH methods, and the existence of data gathered using these methods will not necessarily prevent an OSHA assessment. The USEPA has regulations concerning the assessment of exposures to pesticides and has published guidelines for exposure monitoring. The USEPA does not regulate indoor air quality and the *Compendium* methods are for guidance only.[127]

8.2.2 Voluntary Consensus Standards

Public Law 104-113 includes a requirement for federal government agencies to make use of voluntary consensus standards where possible, rather than duplicating the standard-setting process themselves. The two major organizations involved in standardization are the ASTM (see section 2) and the American National Standards Institute (ANSI). The ASTM has published many standards on workplace air measurement through its committee D22, which are published in the annual book of standards.[128] A selection of different standard guides, practices and test methods is given in Table 3. The ANSI has many standards on worker protection, such as those concerning respirators, but has relatively few[111,129] that deal with air sampling or analysis. This situation may change as the AIHA has become a secretariat to the ANSI. The AIHA is in the process of collating and publishing a compendium of standards and good practices in occupational hygiene, which will include a chapter on sampling and analysis.

Table 3 ASTM standards relevant to workplace sampling and analysis (1998)

Terminology
- D 1356-98: Sampling and Analysis of Atmospheres

Guides
- E 1370-96: Air Sampling Strategies for Worker and Workplace Protection
- E 3614-97: Laboratories Engaged in Sampling and Analysis of Atmospheres and Emissions

Practices
- D 3687-95: Analysis of Organic Compound Vapors Collected by the Activated Charcoal Tube Adsorption Method
- D 6246-98: Evaluating the Performance of Diffusive Samplers
- D 5337-97: Flow Rate of Personal Sampling Pumps
- D 3686-95: Sampling Atmospheres to Collect Organic Compound Vapors (Activated Charcoal Adsorption Method)
- D 4597-97: Sampling Workplace Atmospheres to Collect Organic Gases or Vapors with Solid Sorbent Diffusive Samplers
- D 6196-97: Selection of Sorbents and Pumped Sampling/Thermal Desorption Analysis Procedures for VOCs in Air
- D 4490-96: Toxic Gases or Vapors Using Detector Tubes, Measuring the Concentration of
- D 4599-97: Toxic Gases or Vapors Using Length of Stain Dosimeter, Measuring the Concentration of

Test methods
- D 4413-88: Ethylene Oxide in Workplace Atmospheres (Charcoal Tube Methodology), Determination of (editorial change 1995)
- D 5578-94: Ethylene Oxide in Workplace Atmospheres (HBr Derivatization Method), Determination of
- D 4765-98: Fluorides in Workplace Atmospheres
- D 5197-97: Formaldehyde and Other Carbonyl Compounds in Air (Active Sampler Methodology), Determination of
- D 4913-89: Hydrogen Sulfide by Direct Reading, Length of Stain, Visual Chemical Detectors, Determining Concentration of
- D 5932-96: 2,4-Toluene Diisocyanate (2,4-TDI) and 2,6-Toluene Diisocyanate (2,6-TDI) in Workplace Atmospheres (with 9-(*N*-Methylaminomethyl) Anthracene Method) (MAMA) in the Workplace, Determination of
- D 5836-95: 2,4-Toluene Diisocyanate (2,4-TDI) and 2,6-Toluene Diisocyanate (2,6-TDI) in Workplace Atmospheres (1-2 PP Method), Determination of
- D 4766-98: Vinyl Chloride in Workplace Atmospheres (Charcoal Tube Method)
- D 5466-95: Volatile Organic Chemicals in Atmospheres (Canister Sampling Method), Determination of

8.3 Standards Issued by International Organization for Standardization

Several ISO standards have been developed for monitoring VOCs. These include ISO 8762: 1988 (vinyl chloride by charcoal tube),[130] ISO 9486: 1991 (chlorinated hydrocarbons by charcoal tube),[131] and ISO 9487:

1991 (aromatic hydrocarbons by charcoal tube).[132] More recently, a generic method for VOCs using thermal desorption has been developed by TC146/SC6/WG5 as draft ISO/DIS 16017. This is in two parts, pumped sampling and diffusive sampling and is applicable to workplace, indoor and ambient air. An equivalent solvent desorption method is being developed by SC2/WG4 as ISO/DIS 16200, but this is applicable only to workplace air.

Standards are being developed in TC146/SC6/WG3 for the measurement of VOCs in emissions from building materials and for formaldehyde in indoor air (both pumped and diffusive).

Standards for the measurement of inorganic gases in ambient air are also available. These include ISO 7996: 1985 (nitrogen dioxide)[133] and ISO 6767: 1990 (sulfur dioxide by the *p*-rosaniline method).[134] There are also two detector tube methods for the measurement of inorganic gases in workplace atmospheres: ISO 8760: 1982 (carbon monoxide)[135] and ISO 8761: 1982 (nitrogen dioxide).[136]

9 QUALITY SYSTEMS AND QUALITY CONTROL

The primary control of method quality is validation. Methods are validated by government agencies, academic researchers, corporate occupational hygiene departments, independent research institutes, or equipment manufacturers. Protocols for performing method validation are published, and fall into two types – laboratory and field. Laboratory-based protocols allow reference to atmospheres of traceable standard concentration, whereas field protocols involve comparison to a previously validated reference method. Validation studies are limited to a defined set of conditions and use of the sampling method beyond the tested range is appropriate only if the results are treated with caution.

Sorbents that are treated with reactive chemicals may show considerable batch-to-batch variation. These variations may be magnified where the measurement is based on some property of the chemical, such as a color reaction. In length-of-stain color detector tubes it is very important to check each batch for potential biases caused by variations in sorbent mesh-size, packing, and coating, and also in the measurement marks applied to the outside of the tube. As mentioned above, the US SEI oversees a testing program for detector tube certification[115] based on a program originally designed by the NIOSH.[114]

Quality systems are available for canister sampling. It is possible to fill canisters with inert gas and check the background for memory effects from prior samples. It is possible to add certified quantities of the analyte(s) of interest to the canister after sampling (matrix addition) or to add an internal standard that will not interfere with the analysis, and it is possible to obtain canisters containing certified quantities of analyte(s) as external standards. It is normally possible to analyze more than one aliquot of the same sample. If a secondary trap is used in the analysis, it may be possible for standards to be added to the trap to determine the recovery efficiency. Round-robin programs are not available for the occupational use of canisters but may be available for environmental levels.

Several quality assurance schemes have been developed that apply to the commonest methods of workplace air sampling: sorbent tubes and diffusive samplers for VOCs, and filters for metal particulates, silica and fibers. The Sampling and Laboratory Analysis Committee of the AIHA has published a *Laboratory Quality Assurance Manual* in both English[137] and Spanish. It recommends the use of reagent and matrix blanks, calibration standards, internal standards, external reference materials, matrix spike additions, and replicate analyses where appropriate. It also discusses how the results from these measurements can be used in an on-going control program. Round-robin and proficiency schemes are also important. One such is the Proficiency Analytical Testing (PAT) Program[138] of the Analytical Accreditation Committees of the AIHA.[139] Another is the UK HSE Workplace Analysis Scheme for Proficiency (WASP). Details of these programs may be obtained from The Laboratory Accreditation Co-ordinator, AIHA, 2700 Prosperity Ave., Suite 250, Fairfax, Virginia 22031, and the WASP Co-ordinator, Health and Safety Laboratory, Broad Lane, Sheffield, S3 7HQ, United Kingdom.

The WASP scheme also includes test samples appropriate to the thermal desorption technique, at both occupational and ambient concentration levels. Certified reference materials are available from the EC (European Commission) BCR (Community Bureau of Reference) for aromatic hydrocarbons (CRM 112)[140] and chlorinated hydrocarbons (CRM 555).[141]

An occupational sampling program is of little value unless proper care has been taken in the selection of sampling and analytical method, the timing and placement of sampling, and the interpretation of results. Many countries have organizations for the training and certification of hygiene professionals.

ACKNOWLEDGMENTS

Parts of this work have appeared previously in the ACGIH *Air Sampling Instruments*, 8th edition (eds. B.S. Cohen and S.V. Hering), 1995 and in *Occupational Hygiene*, 2nd edition (eds. J.M. Harrington and K. Gardiner), 1995 and in R.H. Brown: "New European legislation and its relation to methods performance

criteria and standardization'', ACGIH 1998 Applied Workshop, Chapel Hill, NC, February 1998, but in all cases, Crown Copyright is reserved.

Financial support for Richard H. Brown from the HSE, Field Operations Division, is acknowledged (project JE1100012).

LIST OF SYMBOLS

ρ	actual mass concentration of analyte in air ($mg\,m^{-3}$)
ρ_1	actual mass concentration at the beginning of the diffusion layer ($l = 0$) ($mg\,m^{-3}$)
ρ_2	actual mass concentration at the end of the diffusion layer ($mg\,m^{-3}$)
ϕ	actual volume concentration of analyte in air (ppm, v/v)
τ	time constant of diffusive sampler (s)
A	cross-sectional area of diffusion path (cm^2)
C	measured mass concentration of analyte in air ($mg\,m^{-3}$)
C'	measured volume fraction of the analyte in air in (ppm, v/v)
D	coefficient of diffusion ($cm^2\,s^{-1}$)
E_d	desorption efficiency corresponding to m_1
E_s	sampling efficiency
H	mass transfer proportionality factor (cm)
K	constant for detector tubes (cm^{-2})
L	length of stain for a detector tube (cm)
M	molecular mass of the analyte of interest ($g\,mol^{-1}$)
MV	molar volume at NTP (cm^3)
P	pressure of air sampled (kPa)
P_0	normal pressure of air (kPa)
P_1	actual pressure of air at the sampling site (kPa)
T	absolute temperature of air sampled (K)
T_0	normal temperature (K)
T_1	temperature at the sampling site (K)
U	sampling rate ($cm^3\,min^{-1}$)
U'	sampling rate ($ng\,ppm^{-1}\,min^{-1}$)
V	volume of air sample (mL)
k	correction factor for nonideal behavior
l	length of diffusion path (cm)
m	mass of analyte in sample (ng)
m_1	mass of analyte on first tube section (ng)
m_2	mass of analyte on back-up tube section (if used) (ng)
m_{blank}	mass of analyte in blank (ng)
m_s	mass of analyte sorbed by diffusion (ng)
t	sampling time (s)
t'	sampling time (min)

ABBREVIATIONS AND ACRONYMS

ACGIH	American Conference of Governmental Industrial Hygienists
AIHA	American Industrial Hygiene Association
ANSI	American National Standards Institute
ASTM	American Society for Testing and Materials
BCR	Community Bureau of Reference
CEN	European Committee for Standardization
CENELEC	European Electrotechnical Committee for Standardization
EC	European Commission
EL	Exposure Limit
FID	Flame Ionization Detection
GC	Gas Chromatography
HSE	Health and Safety Executive
ISO	International Organization for Standardization
MDHS	Methods for the Determination of Hazardous Substances
NCASI	National Council for Air and Stream Improvement
NIOSH	(US) National Institute for Occupational Safety and Health
NTP	Normal Temperature and Pressure
OSHA	Occupational Safety and Health Administration
PAT	Proficiency Analytical Testing
PEL	Permissible Exposure Limit
SEI	Safety Equipment Institute
SFE	Supercritical Fluid Extraction
TLV®	Threshold Limit Value
TNT	Trinitrotoluene
tVOC	Total Volatile Organic Compounds
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
WASP	Workplace Analysis Scheme for Proficiency
WG	Working Group

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Industrial Hygiene **(Volume 5)**
Carcinogens, Monitoring of Indoor Air • Chromatographic Techniques in Industrial Hygiene

Industrial Hygiene cont'd **(Volume 6)**
Parent and Progeny Compounds in Exhaled Breath, Determination of

REFERENCES

1. Health and Safety Executive, *Methods for the Determination of Hazardous Substances*, HSE Books, Sudbury, UK, in series, 1981–98.
2. Deutsche Forschungsgemeinschaft, *Analytische Methoden zur Prufung Gesundheitsschadlicher Arbeitsstoffe*, DFG, Verlag Chemie, Weinheim, Germany, 1985.
3. Berufsgenossenschaftliches Institut fur Arbeitssicherheit, *BIA Arbeitsmappe Messung von Gefahrstoffen*, Erich Schmidt Verlag, Bielefeld, Germany, 1989.
4. Arbetslivsinstitutet, *Principer och Metoder for Provtagning och Analys av Ämnen Upptagna på Listan över Hygieniska Gränsvården*, Arbete och Hålsa, Vetenskaplig Skriftserie 1997: 6, Solna, Sweden, 1997.
5. (US) National Institute for Occupational Safety and Health, *NIOSH Manual of Analytical Methods*, 2nd edition, DHEW (NIOSH) Pub. No. 75–121, 1975; 3rd edition, DHEW (NIOSH) Pub. No. 84–100, 1984, revised 1990; 4th edition, DHHS (NIOSH) Pub. No. 94–113, 1994 (updated regularly).
6. Occupational Safety and Health Administration, *OSHA Analytical Methods Manual*, OSHA Analytical Laboratories, Salt Lake City. Available from ACGIH, Cincinnati, OH, USA (various methods, 1979–1995, updated regularly).
7. US Environmental Protection Agency, *EPA Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, USEPA, Washington, DC, 1984 (updated regularly).
8. W.A. McClenny, S.M. Schmidt, K.G. Kronmiller, 'Variation of the Relative Humidity of Air Released from Canisters After Ambient Sampling', *J. Air Waste Manage. Assoc.*, **49**, 64–69 (1999).
9. T.J. Kelly, M.W. Holdren, 'Applicability of Canisters for Sample Storage in the Determination of Hazardous Air Pollutants', *Atmos. Environ.*, **29**, 2595 (1995).
10. W.A. McClenny, J.D. Pleil, G.F. Evans, K.D. Oliver, M.W. Holdren, W.T. Winberry, 'Canister-based Methods for Monitoring Toxic VOCs in Ambient Air', *J. Air Waste Manage. Assoc.*, **41**, 1038–1318 (1991).
11. J.L. Varns, J.D. Mulik, D. Williams, 'Passive Sampling Devices and Canisters: Their Comparison in Measuring Air Toxics During a Field Study', in *Proceedings of the 1990 EPA/A&WMA International Symposium, Measurement of Toxic and Related Air Pollutants*, 219–223, 1990.
12. M. Harper, unpublished results.
13. National Council of the Paper Industry for Air and Stream Improvement, 'A Study of the Use of Tedlar® Bag Sampling for the Determination of Reduced Sulfur Gas Concentrations in Workplace Atmospheres', Technical Bulletin 656, NCASI, Research Triangle Park, NC, 1993.
14. Intersociety Committee, *Methods of Air Sampling and Analysis*, 3rd edition, Lewis Publishers, Inc., Chelsea, MI, 1988.
15. G.O. Nelson, *Controlled Test Atmospheres, Principles and Techniques*, Ann Arbor Science Publishers, Ann Arbor, MI, 1971.
16. E.D. Pellizzari, W.F. Gutknecht, S. Cooper, D. Hardison, *Evaluation of Sampling Methods for Gaseous Atmospheric Samples*, EPA 600/3-84-062, USEPA, Research Triangle Park, NC, 1984.
17. A.G. Apol, W.A. Cook, E.F. Lawrence, 'Plastic Bags for Calibration of Air Sampling Devices – Determination of Precision of the Method', *Am. Ind. Hyg. Assoc. J.*, **27**, 149 (1966).
18. F.J. Schuette, 'Plastic Bags for Collection of Gas Samples', *Atmos. Environ.*, **1**, 515–519 (1967).
19. R.L. Seila, W.A. Lonneman, S.A. Meeks, 'Evaluation of Polyvinyl Fluoride as a Container Material for Air Pollution Studies', *J. Environ. Sci. Health A*, **11**, 121 (1976).
20. R. Rothwell, A.D. Mitchell, 'Plastic Bags for Sampling of C2–C6 Hydrocarbons', *Clean Air*, **7**, 35–36 (1977).
21. J.C. Posner, W.J. Woodfin, 'Sampling with Gas Bags 1: Losses of Analyte with Time', *Appl. Ind. Hyg.*, **1**, 163–168 (1986).
22. American Society for Testing and Materials, Standard D 5337-97, *Standard Test Method for Flow Rate of Personal Sampling Pumps*, ASTM, West Conshohocken, PA.
23. K.L. Rubow, 'Air Movers and Samplers', in *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, eds. B.S. Cohen, S.V. Hering, 8th edition, ACGIH, Cincinnati, OH, 203–245, 1995.
24. H.B. Elkins, A.K. Hobby, J.E. Fuller, 'The Determination of Atmospheric Contaminants I: Organic Halogen Compounds', *J. Ind. Hyg.*, **19**, 474–485 (1937).
25. J.C. Gage, 'The Efficiency of Absorbers in Industrial Hygiene Air Analysis', *Analyst*, **85**, 196–203 (1960).
26. K. Marcali, 'Microdetermination of Toluene Diisocyanates in the Atmosphere', *Anal. Chem.*, **29**, 552 (1957).
27. S.J. Gregg, K.S.W. Sing, *Adsorption, Surface Area and Porosity*, 2nd edition, Academic Press, London, 1982.
28. M. Harper, 'Methods of Characterizing Sorbents for Air Sampling Purposes', in *Proc. IVth Int. Conf. on Fundamentals of Adsorption*, Kyoto, 17–22 May 1992, International Adsorption Society, 267, 1992.
29. E.J. Otterson, C.U. Guy, 'A Method of Atmospheric Solvent Vapor Sampling on Activated Charcoal in Connection with Gas Chromatography', in *Transactions of the 26th Annual Meeting*, Philadelphia, PA, American Conference of Governmental Industrial Hygienists, Cincinnati, OH, 37, 1964.

30. National Institute for Occupational Safety and Health, *Documentation of the NIOSH Validation Tests*, DHEW (NIOSH) Publ. No. 77–185, 1977.
31. J.C. Posner, J.R. Okenfuss, 'Desorption of Organic Analytes from Activated Carbon I: Factors Affecting the Process', *Am. Ind. Hyg. Assoc. J.*, **42**, 643–646 (1981).
32. J.C. Posner, J.R. Okenfuss, 'Desorption of Organic Analytes from Activated Carbon II: Dealing with the Problem', *Am. Ind. Hyg. Assoc. J.*, **42**, 647–652 (1981).
33. J. Rudling, 'Organic Solvent Vapor Analysis of Workplace Air', Arbete och Halsa. Vetenskaplig Skriftserie 1987: 11. Arbetarskyddsverket, Solna, Sweden, 1987.
34. B. Callan, K. Walsh, P. Dowding, 'Industrial Hygiene VOC Measurement Interference', *Chem. Ind.*, **X**, 250–2 (1991).
35. M. Harper, L.V. Guild, 'Desorption Efficiency – A Critical Factor Affecting the Accuracy of Air Sampling Methods', in *Clean Air at Work*, eds. R.H. Brown, M. Curtis, K.J. Saunders, S. Vandendriessche, EC Publ no EUR 14214, Brussels-Luxembourg, 186–189, 1992.
36. A. Zlatkis, H.A. Lichtenstein, A. Tishbee, 'Concentration and Analysis of Volatile Organics in Gases and Biological Fluids with a New Solid Absorbent', *Chromatographia*, **6**, 67–70 (1973).
37. E.D. Pellizzari, J.E. Bunch, B.H. Carpenter, E. Sawicki, 'Collection and Analysis of Trace Organic Vapor Pollutants in Ambient Atmospheres', *Environ. Sci. Technol.*, **9**, 552–560 (1975).
38. R.H. Brown, C.J. Purnell, 'Collection and Analysis of Trace Organic Vapor Pollutants in Ambient Atmospheres. The Performance of a Tenax®–GC Adsorbent Tube', *J. Chromatogr.*, **178**, 79–90 (1979).
39. J. Kristensson, *Diffusive Sampling and Gas Chromatographic Analysis of Volatile Compounds*, PhD Thesis, University of Stockholm, Sweden, 1987.
40. M.D. Wright, 'A Dual-capillary Column System for Automated Analysis of Workplace Contaminants by Thermal Desorption – Gas Chromatography', *Anal. Proc.*, **24**, 309–311 (1987).
41. R.H. Brown (ed.), *Study of Sorbing Agents for the Sampling of Volatile Compounds from Air*, EC Contract MAT1-CT92-0038, Health and Safety Laboratory, 1995.
42. CONCAWE, *Method for Monitoring Gasoline Vapor in Air*, CONCAWE Report 8/86, Den Haag, The Netherlands, 1986.
43. Health and Safety Executive, *Methods for the Determination of Hazardous Substances. Volatile Organic Compounds in Air*, MDHS 72, HSE Books, Sudbury, UK, 1993.
44. Nederlands Normalissatie-Instituut, *Methods in Nederlandse Voornorm Series (Luchtkwaliteit Werkplekatmosfeer)*, NNI, Delft, The Netherlands (in series, 1986–92).
45. C.L. Fraust, E.R. Hermann, 'The Adsorption of Aliphatic Acetate Vapors Onto Activated Carbon', *Am. Ind. Hyg. Assoc. J.*, **30**, 494–499 (1969).
46. F.H. Reid, W.R. Halpin, 'Determination of Halogenated and Aromatic Hydrocarbons in Air by Charcoal Tube and Gas Chromatography', *Am. Ind. Hyg. Assoc. J.*, **29**, 390–396 (1968).
47. L.D. White, D.G. Taylor, P.A. Mauer, R.E. Kupel, 'A Convenient Optimized Method for the Analysis of Selected Solvent Vapors in the Industrial Atmosphere', *Am. Ind. Hyg. Assoc. J.*, **31**, 225–232 (1970).
48. C.L. Fraust, 'The Use of Activated Carbon for Sampling Industrial Environs', *Am. Ind. Hyg. Assoc. J.*, **36**, 278 (1975).
49. L.R. Reckner, J. Sacher, *Charcoal Sampling Tubes for Several Organic Solvents*, Pub. No. 75–184, DHEW (NIOSH), Morgantown, MV, June 1975.
50. P.W. Langvardt, R.G. Melcher, 'Simultaneous Determination of Polar and Non-polar Solvents in Air Using a Two-phase Desorption from Charcoal', *Am. Ind. Hyg. Assoc. J.*, **40**, 1006–1012 (1979).
51. I. Johansen, F. Wendelboe, 'Dimethylformamide and Carbon Disulfide Desorption Efficiencies for Organic Vapors on Gas-sampling Charcoal Tube: Analyses with a Gas Chromatographic Backflush Technique', *J. Chromatogr.*, **217**, 317–326 (1981).
52. J.C. Posner, 'Comments on 'Phase Equilibrium Method for Determination of Desorption Efficiencies' and Some Extensions for Use in Methods Development', *Am. Ind. Hyg. Assoc. J.*, **41**, 63–66 (1980).
53. H.B. Elkins, L.D. Pagnotto, E.M. Comproni, 'The Ultraviolet Spectrophotometric Determination of Benzene in Air Samples Adsorbed on Silica Gel', *Anal. Chem.*, **34**, 1797 (1962).
54. J.H.C. Van Mourik, 'Experiences with Silica Gel as Absorbent', *Am. Ind. Hyg. Assoc. J.*, **26**, 498 (1965).
55. A.P. Altshuller, T.A. Bellar, C.A. Clemons, 'Concentration of Hydrocarbons on Silica Gel Prior to Gas Chromatographic Analysis', *Am. Ind. Hyg. Assoc. J.*, **23**, 164 (1962).
56. National Institute for Occupational Safety and Health, *NIOSH Manual of Analytical Methods*, 4th edition, Pub. No. 94–113, Method 5600, DHHS (NIOSH), Morgantown, WV, August, 1994.
57. J.C. Chuang, S.W. Hannan, N.K. Wilson, 'Field Comparison of Polyurethane Foam and XAD-2 Resin for Air Sampling for Polynuclear Aromatic Hydrocarbons', *Environ. Sci. Technol.*, **21**, 798–804 (1987).
58. M. Harper, 'Evaluation of Solid Sorbent Sampling Methods by Breakthrough Volume Studies', *Ann. Occup. Hyg.*, **37**, 65–88 (1993).
59. Occupational Safety and Health Administration, *OSHA Analytical Methods Manual*, OSHA Analytical Laboratories, Salt Lake City, UT, Method 44, October, 1983.
60. G. Bertoni, F. Bruner, A. Liberti, C. Perrino, 'Some Critical Parameters in Collection Recovery and Gas Chromatographic Analysis of Organic Pollutants in Ambient Air Using Light Adsorbents', *J. Chromatogr.*, **203**, 263–270 (1981).

61. P. Ciccioli, A. Cecinato, E. Bancaleoni, M. Frattoni, 'Use of Carbon Adsorption Traps Combined with High Resolution Gas Chromatography – Mass Spectrometry for the Analysis of Polar and Non-polar C4–C14 Hydrocarbons Involved in Photochemical Smog Formation', *J. High Res. Chromatogr.*, **15**, 75 (1992).
62. H. Rothweiler, 'Active Sampling of VOC in Non-industrial Buildings', in *Clean Air at Work*, eds. R.H. Brown, M. Curtis, K.J. Saunders, S. Vandendriesche, CEC Pub. No. EUR 14214, Brussels-Luxembourg, 1992.
63. R.H. Brown, 'What is the Best Sorbent for Pumped Sampling-thermal Desorption of Volatile Organic Compounds? Experience with the EC Sorbents Project', *Analyst*, **121**, 1171–1175 (1996).
64. J.F. Pankow, W. Luo, L.M. Isabelle, D.A. Bender, R.J. Baker, 'Determination of a Wide Range of Volatile Organic Compounds in Ambient Air Using a Multisorbent Adsorption/Thermal Desorption and Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **70**, 5213–5221 (1998).
65. W.A. McClenny, K.D. Oliver, E.H. Daughtrey, Jr, 'Dry Purging of Solid Adsorbent Traps to Remove Water Vapor Before Thermal Desorption of Trace Organic Gases', *J. Air Waste Manage. Assoc.*, **45**, 792–800 (1995).
66. V.E. Rose, J.L. Perkins, 'Passive Dosimetry – State of the Art Review', *Am. Ind. Hyg. Assoc. J.*, **43**, 605–621 (1982).
67. M. Harper, C.J. Purnell, 'Diffusive Sampling – A Review', *Am. Ind. Hyg. Assoc. J.*, **48**, 214–218 (1987).
68. A. Berlin, R.H. Brown, K.J. Saunders (eds.), *Diffusive Sampling: An Alternative Approach to Workplace Air Monitoring*, CEC Pub. No. 10555EN, Commission of the European Communities, Brussels-Luxembourg, 1987.
69. R.H. Brown, 'Diffusive Sampling', in *Clean Air at Work*, eds. R.H. Brown, M. Curtis, K.J. Saunders, S. Vandendriessche, EC Publ no EUR 14214, Brussels-Luxembourg, 141–148, 1992.
70. R.H. Brown, 'The Use of Diffusive Samplers for Monitoring of Ambient Air', *Pure Appl. Chem.*, **65**, 1869–1874 (1993).
71. M. Harper, 'Personal Monitoring b) Passive', in *The Encyclopedia of Analytical Science*, Academic Press, London, 3776–3782, 1995.
72. M. Harper, 'Diffusive Sampling', *Appl. Occup. Environ. Hyg.*, **13**, 759–763 (1998).
73. N. Van den Hoed, O.L.J. van Asselen, 'A Computer Model for Calculating Effective Uptake Rates of Tube-type Diffusive Air Samplers', *Ann. Occup. Hyg.*, **35**, 273–285 (1991).
74. E. Nordstrand, J. Kristensson, 'A Computer Model for Simulating the Performance of Thick-bed Diffusive Samplers', *Am. Ind. Hyg. Assoc. J.*, **55**, 935–941 (1994).
75. D.L. Bartley, L.J. Doemeny, D.G. Taylor, 'Diffusive Monitoring of Fluctuating Concentrations', *Am. Ind. Hyg. Assoc. J.*, **44**, 241–247 (1983).
76. D.L. Bartley, 'Diffusive Monitoring of Fluctuating Concentrations Using Weak Sorbents', *Am. Ind. Hyg. Assoc. J.*, **44**, 879–885 (1983).
77. D.L. Bartley, M.L. Woebkenberg, J.C. Posner, 'Performance of Thick-sorbent Diffusive Samplers', *Ann. Occup. Hyg.*, **32**, 333–343 (1988).
78. D.L. Bartley, G.J. Deye, M.L. Woebkenberg, 'Diffusive Monitor Test: Performance Under Transient Conditions', *Appl. Ind. Hyg.*, **2**(3), 119–122 (1987).
79. J.C. Posner, G. Moore, 'A Thermodynamic Treatment of Passive Monitors', *Am. Ind. Hyg. Assoc. J.*, **46**, 277–285 (1985).
80. H.-U. Pfeffer, L. Breuer, K. Ellermann, *Validierung von Passivsammlern für Immissionsmessungen von Kohlenwasserstoffen*, Materialien Nr. 46, Landesumweltamt NRW, Essen, 1998.
81. R.H. Brown, J. Charlton, K.J. Saunders, 'The Development of an Improved Diffusive Sampler', *Am. Ind. Hyg. Assoc. J.*, **42**, 865–869 (1981).
82. F.J. Hearl, M.P. Manning, 'Transient Response of Diffusion Dosimeters', *Am. Ind. Hyg. Assoc. J.*, **41**, 778–783 (1980).
83. D.W. Underhill, 'Unbiased Passive Sampling', *Am Ind. Hyg. Assoc. J.*, **44**, 237–239 (1983).
84. H. Hori, I. Tanaka, 'Response Characteristics of the Diffusive Sampler at Fluctuating Vapor Concentrations', *Am. Ind. Hyg. Assoc. J.*, **54**, 95–101 (1993).
85. J.R. Compton, G.A. Dwiggins, C.E. Feigley, D.A. Ludwig, 'The Effect of Square-wave Exposure Profiles Upon the Performance of Passive Organic Vapor Monitoring Badges', *Am. Ind. Hyg. Assoc. J.*, **45**, 446–450 (1984).
86. F.C. Tompkins, R.L. Goldsmith, 'A New Personal Dosimeter for the Monitoring of Industrial Pollutants', *Am. Ind. Hyg. Assoc. J.*, **38**, 371–377 (1977).
87. D.W. Underhill, C.E. Feigley, 'Boundary Layer Effect in Diffusive Monitoring', *Anal. Chem.*, **63**, 1011–1013 (1991).
88. L. Pozzoli, D. Cottica, 'An Overview of the Effects of Temperature, Pressure, Humidity, Storage and Face Velocity', in *Diffusive Sampling: An Alternative Approach to Workplace Air Monitoring*, eds. A. Berlin, R.H. Brown, K.J. Saunders, CEC Pub. No. 10555 EN, Commission of European Communities, Brussels-Luxembourg, 1987.
89. N. Zurlo, F. Andreoletti, 'Effect of Air Turbulence on Diffusive Sampling', in *Diffusive Sampling: An Alternative Approach to Workplace Air Monitoring*, eds. A. Berlin, R.H. Brown, K.J. Saunders, EC Publ. no. 10555 EN, Brussels-Luxembourg, 174–176, 1987.
90. S.R. Coleman, 'A Tube-type Diffusive Monitor for Sulphur Dioxide', *Am. Ind. Hyg. Assoc. J.*, **44**, 929–936 (1983).
91. E.D. Palmes, A.F. Gunnison, J. DiMattio, C. Tomczyk, 'Personal Sampler for Nitrogen Dioxide', *Am. Ind. Hyg. Assoc. J.*, **37**, 570–577 (1976).

92. C.H.F. Atkins, J. Sandalls, D.V. Law, A.M. Hough, K. Stevenson, *The Measurement of Nitrogen Dioxide in the Outdoor Environment Using Passive Diffusion Tube Samplers*, AERE-R12133, Harwell Laboratory, UK, 1986.
93. A.J. Gair, S.A. Penkett, 'The Effects of Wind Speed and Turbulence on the Performance of Diffusion Tube Samplers', *Atmos. Environ.*, **29**, 2529–2533 (1995).
94. C.E.C. Downing, G.W. Campbell, J.C. Bailey, *A Survey of Sulphur Dioxide, Ammonia and Hydrocarbon Concentrations in the United Kingdom Using Diffusion Tubes: July to December 1992*, Report LR 964, Warren Spring Laboratory, Stevenage, UK, 1994.
95. M. Ferm, *A Sensitive Diffusional Sampler*, Report B1020, Institutet for Vatten och Luftvardsforskning, Sweden, 1991.
96. W. Frenzel, E. Grimm, G. Gruetzmacher, 'Evaluation of a Diffusive Sampling Method for the Determination of Atmospheric Ammonia', *Fresenius' J. Anal. Chem.*, **351**, 19–26 (1995).
97. K.H. Pannwitz, 'Influence of Air Currents on the Sampling of Organic Solvent Vapors with Diffusive Samplers', in *Diffusive Sampling, an Alternative Approach to Workplace Air Monitoring*, eds. A. Berlin, R.H. Brown, K.J. Saunders, EC Publ. no. 10555 EN, Brussels-Luxembourg, 157–160, 1987.
98. D. Mark, A. Robertson, H. Gibson, G. Borzucki, B. Cheme, W.M. Maclaren, *Evaluation of Diffusive Samplers for Monitoring Toxic Gases and Vapors in Coalmines*, Report TM/90/11, Institute of Occupational Medicine, Edinburgh, 1990.
99. H. Hori, I. Tanaka, 'Effect of Face Velocity on Performance of Diffusive Samplers', *Ann. Occup. Hyg.*, **40**, 467–475 (1996).
100. R.G. Lewis, J.D. Mulik, R.W. Coutant, G.W. Wooten, C.R. McMillin, 'Thermally Desorbable Passive Sampling Device for Volatile Organic Chemicals in Ambient Air', *Anal. Chem.*, **57**, 214–219 (1985).
101. A. Kasper, H. Puxbaum, 'A Badge-type Passive Sampler for Monitoring Ambient Ammonia Concentrations', *Fresenius' J. Anal. Chem.*, **350**, 448–453 (1994).
102. B.S. Samini, 'The Effect of Face Air Velocity on the Rate of Sampling of Air Contaminants by a Diffusive Sampler', in *Diffusive Sampling, an Alternative Approach to Workplace Air Monitoring*, eds. A. Berlin, R.H. Brown, K.J. Saunders, EC Publ. no. 10555 EN, Brussels-Luxembourg, 166–169, 1987.
103. V. Cocheo, C. Boaretto, P. Sacco, 'High Uptake Rate Radial Diffusive Sampler Suitable for Both Solvent and Thermal Desorption', *Am. Ind. Hyg. Assoc. J.*, **57**, 897–904 (1996).
104. Health and Safety Executive, 'Diffusive Uptake Rates on Draeger ORSA-5, 3M 3500/20, SKC 575 and Radiello Diffusive Samplers', *The Diffusive Monitor*, **9**, 14–19 (1997).
105. Health and Safety Executive, *Methods for the Determination of Hazardous Substances. Volatile Organic Compounds in Air*, MDHS 88, HSE Books, Sudbury, UK, 1997.
106. Health and Safety Executive, 'Diffusive Uptake Rates on Perkin-Elmer Sorbent Tubes', *The Diffusive Monitor*, **8**, 14–16 (1996).
107. Health and Safety Executive, *Methods for the Determination of Hazardous Substances. Volatile Organic Compounds in Air*, MDHS 80, HSE Books, Sudbury, UK, 1995.
108. B.E. Saltzman, P.E. Caplan, 'Detector Tubes, Direct-reading Passive Badges, and Dosimeter Tubes', in *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, 8th edition, eds. B.S. Cohen, S.V. Hering, American Conference of Governmental Industrial Hygienists, Cincinnati, OH, 401–437, 1995.
109. American Industrial Hygiene Association, *Direct Reading Colorimetric Tubes – A Manual of Recommended Practices*, AIHA, 2700 Prosperity Avenue, Fairfax, VA, 1977.
110. F.H. Colen, 'A Study of the Interchangeability of Gas Detector Tubes and Pumps', *Am. Ind. Hyg. Assoc. J.*, **38**, 686–694 (1977).
111. American National Standards Institute, *American National Standard 102-1990 (R1998): Gas Detector Tube Units – Short-term Type for Toxic Gases and Vapors in Working Environments*, ANSI, 1998.
112. B.E. Saltzman, 'Basic Theory of Gas Indicator Tube Calibrations', *Am. Ind. Hyg. Assoc. J.*, **23**, 112–126 (1962).
113. National Institute for Occupational Safety and Health, *Certification of Gas Detector Tube Units*, Federal register 38: 11 458 (May 8, 1973); also 43 CFR 84.
114. C.P. Roper, 'The NIOSH Detector Tube Certification Program', *Am. Ind. Hyg. Assoc. J.*, **35**, 438–442 (1974).
115. F.E. Wilcher, Jr, 'SEI Gas Detector Tube Certification', *Appl. Ind. Hyg.*, **3**, R-7 (August, 1988).
116. K. Acer, 'Report of Workshop #7 on Detector Tubes for Gases and Vapors', *Appl. Occup. Environ. Hyg.*, **8**, 409 (1993).
117. US Department of Labor, *Directive 73–4: Use of Detector Tubes*, Washington DC, March, 1973.
118. Japanese Standards Association, *Japanese Industrial Standard: Detector Tube Type Gas Measuring Instruments*, JIS K 0804-1985, Japanese Standards Association, 1–24, Akasaka 4, Minato-ku, Tokyo 107, Japan, 1986.
119. British Standards Association, *British Standard: Gas Detector Tubes BS5343. Part 1. Specification for Short Term Gas Detector Tubes. British Standards Institution*, BSI, 2 Park St., London W1A 2BS, UK, 1986.
120. *EN 482: 1994, Workplace Atmospheres – General Requirements for the Performance of Procedures for the Measurement of Chemical Agents*, European Committee for Standardization, Brussels, Belgium.

121. *EN 838: 1995, Workplace Atmospheres – Requirements and Test Methods for Diffusive Samplers for the Determination of Gases and Vapors*, European Committee for Standardization, Brussels, Belgium.
122. *EN 1076: 1997, Workplace Atmospheres – Requirements and Test Methods for Pumped Sorbent Tubes for the Determination of Gases and Vapors*, European Committee for Standardization, Brussels, Belgium.
123. *EN 1231: 1996, Workplace Atmospheres – Requirements and Test Methods for Short-term Detector Tube Systems*, European Committee for Standardization, Brussels, Belgium.
124. *EN 1232: 1997, Workplace Atmospheres – Requirements and Test Methods for Pumps for Personal Sampling of Chemical Agents*, European Committee for Standardization, Brussels, Belgium.
125. N.A. Leidel, K.A. Busch, J.R. Lynch, *Occupational Exposure Sampling Strategy Manual*, Publication No. 77–173, DHHS (NIOSH), Morgantown, WV, 1977.
126. E.R. Kennedy, T.J. Fischbach, R. Song, P.M. Eller, S.A. Shulman, *Guidelines for Air Sampling and Analytical Method Development and Evaluation*, Publication No. 95–117, DHHS (NIOSH), Morgantown, WV, 1995.
127. US Environmental Protection Agency, Office of Pesticide Programs, *Pesticide Assessment Guidelines, Subdivision U, Applicator Exposure Monitoring*, Report 540/9-87-127, USEPA, 1987.
128. American Society for Testing and Materials, *Annual Book of ASTM Standards*, Vol. 11.03, Atmospheric Analysis; Occupational Health and Safety; Protective Clothing, ASTM, West Conshohocken, PA (updated annually).
129. American National Standards Institute, *American National Standard 104-1998: Air Sampling Devices – Diffusive Types for Gases and Vapors in Working Environments*, ANSI, 1998.
130. *ISO 8762: 1988, Workplace Air – Determination of Vinyl Chloride – Charcoal Tube/Gas Chromatographic Method*, ISO, 1988.
131. *ISO 9486: 1991, Workplace Air – Determination of Vaporous Aromatic Hydrocarbons – Charcoal Tube/Solvent Desorption/Gas Chromatographic Method*, ISO, 1991.
132. *ISO 9487: 1991, Workplace Air – Determination of Vaporous Chlorinated Hydrocarbons – Charcoal Tube/Solvent Desorption/Gas Chromatographic Method*, ISO, 1991.
133. *ISO 7996: 1985, Air Quality – Determination of Mass Concentration of Nitrogen Dioxide*, ISO, 1985.
134. *ISO 6767: 1990, Air Quality – Determination of Mass Concentration of Sulfur Dioxide by the p-Rosaniline Method*, ISO, 1990.
135. *ISO 8760: 1982, Air Quality – Carbon Monoxide – Direct Indication Using Detector Tubes*.
136. *ISO 8761: 1982, Air Quality – Nitrogen Dioxide – Direct Indication Using Detector Tubes*.
137. American Industrial Hygiene Association, Sampling & Laboratory Analysis Committee, *Laboratory Quality Assurance Manual*, 2nd edition, AIHA, Fairfax, VA, 1997.
138. National Institute for Occupational Safety and Health, *NIOSH Proficiency Analytical Testing (PAT) Program*, DHEW (NIOSH) Pub. No. 77–173, 1977.
139. American Industrial Hygiene Association, *Laboratory Accreditation*, AIHA, Fairfax, VA, 1985.
140. S. Vandendriessche, B. Griepink, J.C.Th. Hollander, J.W.J. Gielen, F.G.G.M. Langelaan, K.J. Saunders, R.H. Brown, 'Certification of a Reference Material for Aromatic Hydrocarbons in Tenax® Samplers', *Analyst*, **116**, 437–441 (1991).
141. Th.L. Hafkenscheid, 'CRN 555: A Reference Material for the Determination of Chlorinated Hydrocarbons in Air Using Thermal Desorption–Gas Chromatography', *The Diffusive Monitor*, **9**, 4–5 (1997).

Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

M. Borsaru
CSIRO-Exploration and Mining, Kenmore, Queensland, Australia

1 Introduction 1
2 Scattering Cross-section 2
3 Scattering and Absorption of γ-Radiation 2
3.1 The Three Principal Processes 2
3.2 Other Interactions of γ-Rays With Matter 3
3.3 Density and Composition Measurements With γ-Rays 4
4 Natural γ-Radiation 5
5 Scattering and Absorption of Neutrons 6
5.1 Neutron Sources 7
5.2 Interaction of Neutrons with Matter 7
5.3 Macroscopic Treatment of Slowing Down 8
6 Radiation Detectors 8
7 The Determination of Water Content in Soil and Boreholes Using Neutrons 9
7.1 Definition of Porosity 9
8 The Monte Carlo Method 9
9 Nuclear Techniques for In Situ Analysis 10
9.1 Passive Logging – Natural γ-Ray 10
9.2 Active Logging 11
9.3 Oil Well Logging 11
9.4 Uranium 13
9.5 Coal 14
9.6 Iron Ore 14
Abbreviations and Acronyms 14
Related Articles 15
Appendix 1 15
The Compton Effect 15
Elastic Scattering of Fast Neutrons 15
References 15

Nuclear techniques are widely used in geological and mineral applications. Owing to the high penetration of γ-radiation and neutrons, nuclear techniques have found many applications, including well logging, bulk analysis and on-stream analysis. The first part of this article presents the physics of the interaction of γ-rays and neutrons with matter. Applications based on these interactions to well logging for oil, uranium, coal and iron ore are presented in the second part of the article. Applications based on the natural radioactivity of rocks are also covered.

1 INTRODUCTION

The interactions of γ-radiation and neutrons with matter form the basis of a large number of nuclear techniques developed for chemical analysis or used to infer properties of interest in logging. Nuclear techniques have the ability to measure properties of materials without contact and, because of the high penetration of both γ-radiation and neutrons, enable measurements to be made through walls. In borehole-logging applications, these two types of penetrating radiation can traverse the pressure housing of the logging tools and the formation and return a signal carrying information about the surrounding formation and the rock matrix.

This article presents the principles of the interaction of γ-rays and neutrons with matter and outlines geological/mineral applications. The emphasis is on nuclear borehole logging. Because of the deep penetration of neutrons and γ-rays, nuclear techniques are well suited to borehole-logging applications. The volume of matter investigated by nuclear radiation is much larger than the core analyzed from a cored borehole, so a more representative sample is analyzed.

Well logging grew from the need of the petroleum industry to evaluate hydrocarbon formations. Today, borehole logging is a mature technology, and nuclear logging is well established in the oil, gas, uranium and coal industries. It is also making inroads in the mineral industry. It is now very difficult to imagine these industries without logging.

Nuclear logging, employing in situ γ-ray spectroscopy, can measure the chemical composition of the earth formation, thereby providing an indication of the dominant mineralogy. This is vital information in oil well logging. In situ γ-ray spectroscopy replaces the time-consuming chemical analysis of the core in a laboratory. Bulk density, which is another important property of an earth formation, is measured by γ-ray scattering. Some minerals contain the natural radioactive elements potassium, thorium and uranium. The detection of their γ-radiation indicates the presence of these elements, and quantitative measurements can be made of their concentration.

Nuclear applications are an integral part of today's measurement technologies, contributing to increased

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

efficiency and large financial savings in industry. Although a mature technology, new applications are still being found.

2 SCATTERING CROSS-SECTION

For the study of neutron and γ-ray scattering it is convenient to introduce the scattering cross-section. For a very thin target bombarded with neutrons or γ-rays, the number of scattered particles is very small compared with the number of particles in the beam, as shown in Equation (1):

$$-\frac{\mathrm{d}N}{N} = A\,\mathrm{d}x = n\sigma\,\mathrm{d}x \tag{1}$$

where $-\mathrm{d}N$ is the number of scattered particles, N is the number of incoming particles, $\mathrm{d}x$ is the thickness of the target and A is a proportionality constant dependent on the number of nuclei per unit volume n and the nature of the nuclei, represented by σ, the scattering cross-section. For a thicker target, the number of particles passing through undeflected, N, is obtained by integrating Equation (1), illustrated in Equation (2):

$$N = N_0\,\mathrm{e}^{-n\sigma x} \tag{2}$$

where N_0 is the total number of particles in the beam which hit the target. The total number of particles scattered from the beam is shown in Equation (3):

$$N_{\mathrm{sc}} = N_0(1 - \mathrm{e}^{-n\sigma x}) \tag{3}$$

When the target contains several isotopes of different cross-sections, $n\sigma$ is replaced by $n_1\sigma_1 + n_2\sigma_2 + n_3\sigma_3 + \cdots$

3 SCATTERING AND ABSORPTION OF γ-RADIATION

γ-Rays are electromagnetic radiation, as are X-rays, light, radio waves, etc. For γ-rays or neutrons, the processes by which they interact with matter are such that a single event will remove the γ-ray or neutron from the beam. According to Equation (2), the intensity of the beam hitting a target will decrease exponentially, as illustrated in Equation (4):

$$N = N_0\,\mathrm{e}^{-n\sigma x} = N_0\,\mathrm{e}^{-\mu x} \tag{4}$$

where the constant μ is called the absorption coefficient and has dimension $(\text{length})^{-1}$. It is also common to measure absorber thickness in $\mathrm{g\,cm^{-2}}$ and write Equation (4) as Equation (5):

$$N = N_0\,\mathrm{e}^{-\mu_{\mathrm{m}}\xi} \tag{5}$$

where $\xi = \rho x$ is the mass per unit area and where the mass absorption coefficient μ_{m} is, as illustrated in Equation (6):

$$\mu_{\mathrm{m}} = \frac{\mu}{\rho} = \frac{n\sigma}{\rho} = \frac{N_{\mathrm{A}}\sigma}{M_{\mathrm{A}}} \tag{6}$$

N_{A} is Avogadro's number, M_{A} is the atomic weight of the element in the absorber and ρ is the density. If more elements are present in the absorber, then Equation (7) applies:

$$\mu_{\mathrm{m}} = N_{\mathrm{A}} \sum_i \frac{\alpha_i\sigma_i}{M_{\mathrm{A}i}} \tag{7}$$

where α_i is the abundance ratio of the ith element.

The interaction of γ-radiation with matter in the energy range from 50 keV to 50 MeV is dominated by three processes: (1) photoelectric effect, (2) Compton effect and (3) pair production.

3.1 The Three Principal Processes

3.1.1 The Photoelectric Effect

In the photoelectric effect, predominant at low energies, the γ-ray (photon) gives all its energy to a bound electron. The electron, to overcome its binding energy in the atom, uses part of the energy and the rest is taken as kinetic energy. The energy of the emitted electron is therefore $E_{\mathrm{e}} = E_\gamma - E_{\mathrm{b}}$, where E_γ is the γ-ray energy and E_{b} is the binding energy of the electron in the atom, usually the k-shell energy. The probability of photon interacting with an electron in a given orbit is higher when the energy of the photon is equal to the binding energy and decreases with increasing energy. The electron gives up its energy to other atoms and eventually falls back into an orbit. The X-rays emitted following the capture of the electrons into orbits interact with the matter again and again until all the energy is absorbed either at very low-energy orbital levels or other mechanisms which are outside the scope of this article.

The cross-section of the photoelectric absorption per atom is a function of photon energy, E, illustrated in Equation (8), and is approximated[1] as being proportional to:

$$\sigma(E) \approx \frac{Z^{4.5}}{E^n},\ 2.5 \le n \le 3.5 \tag{8}$$

The mass absorption coefficient for the photoelectric effect is shown in Equation (9):

$$\mu_{\mathrm{mph}} = \rho N_{\mathrm{A}} \sum_i \frac{\alpha_i\sigma_i(E)}{M_{\mathrm{A}i}} \tag{9}$$

The contribution of the photoelectric effect at low energies is significant for heavy elements because of

their high atomic number. For lighter elements, the contribution of the photoelectric effect is reduced as a result of the linear relationship between its cross-section and $Z^{4.5}$.

3.1.2 The Compton Effect

The Compton effect predominates around 1 MeV. It implies scattering of photons by the atomic electrons. The photon is deflected from its trajectory with or without loss of energy, as shown schematically in Figure 1 (ν is the frequency of the photon of wavelength λ, where $\nu = 1/\lambda$). This happens at photon energies much larger than the electron binding energies so that, theoretically, the electrons can be considered free. Expressions for the energy $h\nu'$ of the scattered photon, the energy T of the scattered electron and the relationship between the angles of scattering shown in Figure 1 are given in Appendix 1. The cross-section for the Compton scattering was calculated by Klein and Nishina.[2] It is well described by Davisson[3] and many other books of nuclear physics or quantum mechanics.

3.1.3 The Pair Production Effect

In the pair production effect, a photon disappears with the creation of an electron/positron pair. The total kinetic energy of the resultant particles is equal to the photon energy minus the mass energy of the electron/positron pair created by the interaction (1.02 MeV) and therefore the pair production effect cannot take place below 1.02 MeV. Figure 2 shows schematically the pair production effect. The pair production takes place in the field of a nucleus or electron with no change of the state of the nucleus or its electrons. The nucleus must be present to absorb some of the momentum of the photon. Pair production cannot take place in the absence of the nucleus or electron because the conservation of energy and conservation of momentum cannot be simultaneously observed.

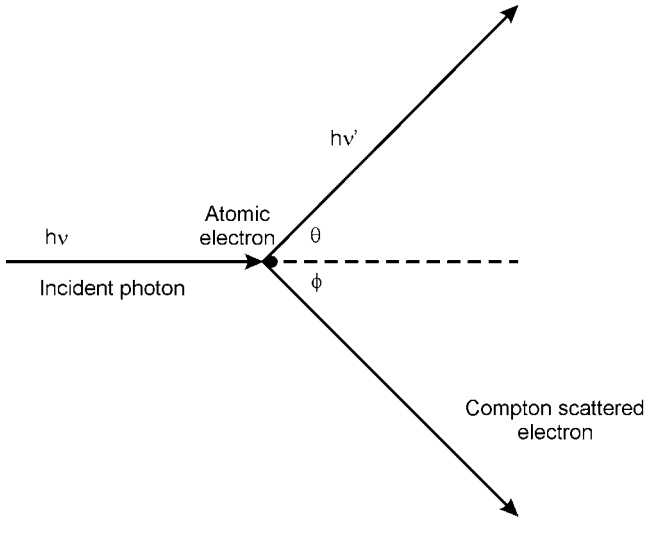

Figure 1 Compton effect.

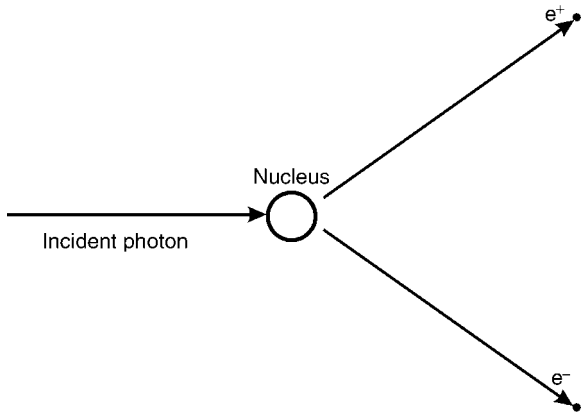

Figure 2 Pair production effect.

3.2 Other Interactions of γ-Rays With Matter

The photoelectric, Compton and pair production effects are the major interactions of γ-rays with matter and contribute most of the total γ-ray attenuation coefficient in the energy region being considered. However, they are not the only interactions. Other interactions of interest are resonance scattering of γ-rays and photonuclear reactions, also called γ-activation. There are other interactions (Thomson, Delbrück, Rayleigh) which occur at low energies. They are treated in more specialized books and are outside the scope of this general article.

3.2.1 Resonance Scattering

γ-Ray resonance scattering is an elastic process that takes place via an excited state of a stable nucleus. If the energy of the γ-ray corresponds exactly to the energy level of the stable nucleus, the nucleus can absorb this γ-ray and become an unstable excited nuclear state. The nucleus will regain its stable state by emitting a γ-ray that is theoretically of the same energy as the one absorbed. In practice, the γ-ray emitted by the excited nucleus at rest has less energy than the photon that excited the nucleus, E_r, because of the recoil energy taken by the nucleus. If M is the mass of the emitting atom, conservation of energy and momentum leads to the expression for the energy of the emitted photon $h\nu$: $h\nu = E_r - (E^2/2Mc^2)$. The same amount of kinetic energy is transferred to the nucleus which is excited by the γ-ray, so that the total energy displacement is $\Delta E = E^2/Mc^2$. In general, ΔE is large by comparison with the width of the energy level Γ and the system is out of resonance, in other words

the cross-section for resonance absorption is small. From Heisenberg's principle, $\Gamma(\text{eV}) = 6.58 \times 10^{-16}/\tau$, where τ is the lifetime of the excited state in seconds. This is why it took a long time to prove the existence of γ-ray resonance scattering.

Mössbauer discovered that an atom embedded in a crystal lattice can emit or absorb γ-radiation without loss from the recoil because the momentum is taken up by the crystal as a whole with negligible energy loss.

There are a few methods of producing resonance scattering. The change of quantum energies can be achieved by moving the absorber and the radiation source with respect to each other (based on the Doppler effect). Another practical technique to obtain γ-rays of the precisely defined energy for resonance scattering is to choose a radioisotope source that decays via the excited state of the chosen element. For example, if the element chosen for γ-ray scattering is ^{60}Ni, a ^{60}Co source may be used. ^{60}Co decays by β- and γ-ray emission via excited states of ^{60}Ni. Matching the radioactive source to the chosen element is essential for γ-ray resonance. Resonance scattering does not take place when the γ-ray source is in a solid state because the recoil energy losses during the emission of γ-rays make it deficient in energy by several tens of keV. This deficiency may be overcome by using a gaseous source. γ-Rays from gaseous sources are Doppler broadened so that about 1% of the γ-rays are in resonance.

γ-Ray resonance scattering has found practical applications. Mössbauer spectroscopy is a powerful technique for elemental analysis based on this effect. Copper and nickel are two favorable elements for analysis by γ-ray scattering using vapor sources of $^{65}ZnI_2$ and $^{60}CoBr_2$ respectively.

3.2.2 Photonuclear Reactions

γ-Rays of high energies can produce nuclear reactions with atomic nuclei. If the incident γ-ray energy exceeds a threshold energy, it is possible to remove particles from stable nuclei. Each element is characterized by a particular threshold energy. The most frequent nuclear reactions induced by high-energy γ-rays in the energy range of interest are (γ,n) processes. The maximum cross-section for light elements is at 20–25 MeV and $\cong$15 MeV with other elements. The photoneutron reaction has found practical applications mostly for the determination of low-atomic-number elements. The technique is selective if the (γ,n) threshold energies are selected appropriately.

The only reactions to have threshold energies less than 2.5 MeV are the (γ,n) reactions for ^{9}Be (threshold = 1.67 MeV) and ^{2}H (2.22 MeV). The (γ,n) reaction with beryllium has been used in different techniques of analysis for this element. ^{124}Sb is a suitable γ-ray source for beryllium analysis. It has a half-life of 60 days and emits γ-rays of 1.69 and 2.09 MeV. The average energy of neutrons from the ^{9}Be(γ,n) reaction using a ^{124}Sb source is approximately 24 keV. The neutrons are moderated in a hydrogenous moderator and detected with thermal neutron detectors such as $^{10}BF_3$ or ^{3}He counters.

3.3 Density and Composition Measurements With γ-Rays

The three major interactions of γ-radiation with matter have found wide industrial applications. Here we review the basis of density and composition determination.

The Compton mass attenuation coefficient μ_c for photon energies much larger than the electron binding energies can be expressed as (Equation 10):

$$\mu_c = \left(\frac{\rho N_A Z}{M_A}\right) e_\mu \tag{10}$$

where $\rho N_A/M_A$ represents the number of atoms per cubic centimeter, Z is the number of electrons in the atom and e_μ is the Compton scattering cross-section per electron. For a γ-ray of given energy (Equation 11):

$$\mu_c = \frac{Z}{M_A}\rho \tag{11}$$

Except for hydrogen, the Z/M_A ratio for low-Z elements (which form the Earth's crust) is very nearly a constant equal to $\frac{1}{2}$. For hydrogen, $Z/M_A = 1$, and this introduces an error when hydrogenous material is present in matter investigated. If the energy of γ-rays is greater than 300 keV, one can neglect the photoelectric effect and assume that the interaction is by the Compton effect. Equation (4) can be used to derive Equation (12):

$$\log N = \log N_0 - \mu_c x = \log N_0 - \frac{Z}{M_A}\rho x \tag{12}$$

For a radioactive source of long half-life, N_0 is constant. By plotting a graph of $\log N$ vs density (ρ), one obtains a straight line if the thickness χ of the materials of different densities is constant. The density is extracted from this linear relationship.

Concept of Equivalent Atomic Number, Z_{eq} In order to facilitate measurement of the chemical composition of a medium by γ-scattering, Czubek[1] introduced a quantity called the equivalent atomic number (Z_{eq}) of the scattering medium in terms of photoelectric and Compton coefficients. We have seen that the Compton and photoelectric absorption coefficients (μ_c, μ_{ph}) for given γ-ray energies are proportional to $(Z/M_A)\rho$ and $(Z^{4.5}/M_A)\rho$ respectively. Therefore, $\mu_{ph}/\mu_c = \text{const.}Z^{3.5}$. One can

show that, for a mixture of elements (Equation 13):

$$Z_{\mathrm{eq}} = Z = \sqrt[3.5]{\frac{\sum_i W_i \frac{Z_i^{4.5}}{M_{\mathrm{A}i}}}{\sum_i W_i \frac{Z_i}{M_{\mathrm{A}i}}}} \tag{13}$$

where Z_{eq} is a substitute for Z for a multielement medium. With the assumption that $Z_i/M_{\mathrm{A}i}$ is a constant for low-Z elements,

$$Z_{\mathrm{eq}} = \sqrt[3.5]{\sum_i W_i Z_i^{3.5}} \tag{14}$$

Equation (14) shows the strong dependence of the Z_{eq} of a medium on the constituent element with the highest atomic number. This dependence can be used to determine the concentration of a high-Z element in a matrix of low-Z elements, e.g. the percentage of Fe in iron ore, ash in coal, etc. (%Fe or %ash are proportional to Z_{eq}).

Practical Methods of Density and Z_{eq} Measurements When there is access to both sides of the medium under investigation, the density and Z_{eq} are determined by measuring the direct attenuation of γ-radiation passing through the medium. A collimated γ-ray beam of energy greater than 300 keV is employed for density measurements. The γ-ray sources used are usually ^{60}Co and ^{137}Cs. By using γ-rays of both low energy and high energy in a medium with a high photoelectric absorption coefficient, one can determine both the photoelectric and Compton absorption coefficients of the medium and thus Z_{eq} from Equation (13).

When only one side of the medium is available, e.g. boreholes, roads or thick walls, the detector and the source are placed on the same side of the medium. Shielding is used to stop the γ-rays from the source reaching the detector. Figure 3 shows schematically the geometry suitable for this application together with the γ-ray energy spectrum recorded by the detector. The γ-rays from the source undergo successive Compton scattering, losing energy. The γ-rays can reach the detector after one or multiple scatterings. Below 300 keV the photoelectric absorption increases with decreasing γ-ray energy. This spectrum is called the backscattered spectrum. It ranges from almost the source energy down to an energy at which photoelectric absorption reduces the spectral intensity to almost zero. The high-energy region of the spectrum, above 300 keV, is a function of the electronic density of the medium, and below 300 keV both photoelectric and Compton effects are present. Theory shows that the ratios of the count rates recorded in two energy windows, at high and low energy, is a function only of the Z_{eq} of the medium. Borehole logging techniques for the determination of Fe grade in iron ore deposits or ash in coal have been developed on this basis.[4,5]

4 NATURAL γ-RADIATION

All rocks and soils contain a great number of radioactive elements that emit γ-radiation. The three main sources of natural γ-rays are potassium, decay products in the uranium series and decay products in the thorium series.

Of the three naturally occurring isotopes of potassium, only ^{40}K is radioactive; it has a half-life of 1.28×10^9 years and isotopic abundance of 0.0118%, which gives a specific activity of 31.4 Bq per g of natural potassium. ^{40}K decays to ^{40}Ca and emits a γ-ray of energy 1.46 MeV. Because many rock-forming minerals contain potassium, it is the commonest natural γ-ray encountered in nature.

^{238}U and ^{235}U are more abundant among the isotopes of uranium, with ^{238}U making up 99.3% of the total. Each isotope of uranium decays by α-emission to an isotope of thorium. Thorium and its daughters are also radioactive and a decay chain of the parent isotope is produced. ^{238}U has a half-life much longer than its radioactive decay products, and in an undisturbed medium a secular equilibrium is established. A radioactive decay series is said to be in a state of secular equilibrium when the number of atoms of each daughter being produced in the series is equal to the number of atoms of that daughter being lost by radioactive decay. ^{214}Bi is one of the isotopes produced in the ^{238}U chain and decays emitting a 1.765 MeV γ-ray. This γ-ray can be used for the determination of uranium concentration in a uranium deposit if the deposit is in equilibrium. Uranium is widespread in the environment. Most crustal rocks contain some uranium, averaging about 2.7 ppm or 33 Bq kg^{-1}. Phosphate-bearing rocks used as fertilizer contain high concentrations of uranium. It is also found in high concentration in oceans, ranging from about 2 to 3.7 mg m^{-3}.

^{232}Th has a decay chain similar to the decay chains of uranium. One prominent γ-ray of energy 2.6 MeV, emitted by the daughter product ^{208}Tl, is usually used for the determination of thorium concentration. Like uranium, thorium is widely found in rocks. The average thorium content of the continental upper crust is 9–10 ppm, about four times greater than that of uranium. However, having lower activity than uranium, the radioactivity concentration of thorium and uranium is about the same. High concentrations of thorium ($\cong$10%) are found in monazite, a rare-earth mineral. Owing to its high level of radioactivity, people living close to monazite sands can be exposed to high doses of radiation.

Measurement of natural radiation has found many applications. Because of the high level of natural

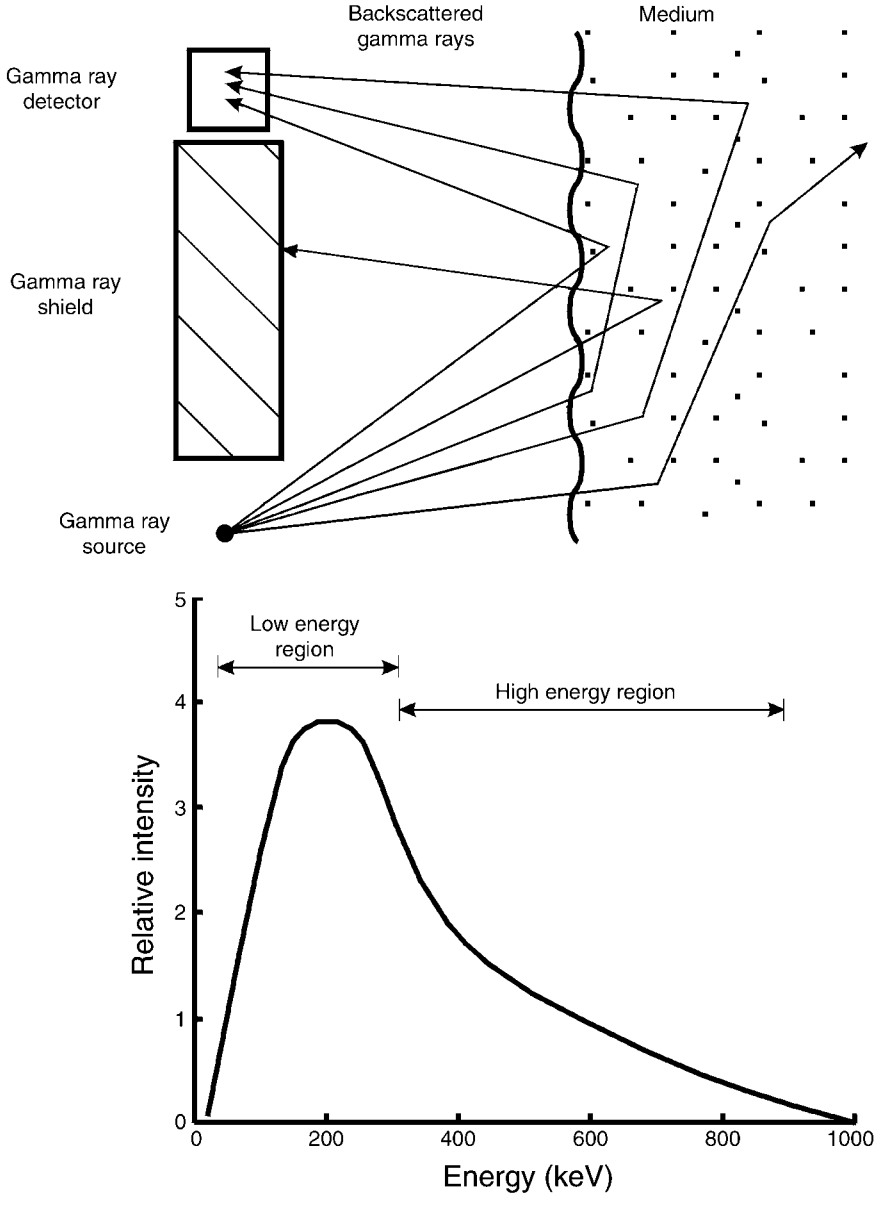

Figure 3 Schematic representation of the backscattered γ-rays and a backscattered γ-ray spectrum.

radioactivity in shales it is used to distinguish between shales and other sediments. It is also used in the exploration and mining of uranium. A technique to determine the ash content of coal on conveyor belts has also been developed. It is based on the fact that uranium, thorium and potassium are associated with the mineral mater in coal and the γ-ray activity of coal is correlated with the ash content.[6] An interface gauge for the simultaneous monitoring of the green liquor and red mud interface and the profile of the solids concentration has been developed for the alumina industry. It has been shown that, subsequent to the digestion of bauxite in caustic soda, traces of radioactivity originally present in bauxite remain with the red mud and the green liquor is free of radioactivity.[7] The gauge consists of a number of γ-ray counters arranged inside a Teflon-coated steel tube, which is suspended vertically inside settling tanks.

5 SCATTERING AND ABSORPTION OF NEUTRONS

The neutron is a neutral particle having a mass approximately equal to the mass of a proton (about 1800 times the rest mass of an electron). The neutron has a magnetic moment and therefore there is an electromagnetic

interaction between the neutron and the atomic electrons. However, this interaction is small and can be neglected when studying the scattering and stopping of neutrons in matter. In the present discussion we will assume that the neutron interacts exclusively with the atomic nuclei. Consequently, neutron interactions in matter are rarer than γ-ray interactions and neutron ranges are longer.

The neutron energies are arbitrarily classified as:

- fast neutrons – above 500 keV
- intermediate – 1 keV to 500 keV
- slow – below 1 keV.

Slow neutrons are further subdivided as

- epithermal neutrons – from 0.1 eV to 1 keV
- thermal neutrons – below 0.1 eV.

5.1 Neutron Sources

The most common neutron sources used in applications are: (1) isotopic sources, (2) neutron generators and (3) nuclear reactors.

5.1.1 Isotopic Sources

Isotopic sources use (i) (α,n) reactions, (ii) photo-neutron reactions or (iii) the fission process.

Neutron Sources Based on the (α,n) Reaction The principal reaction leading to neutron emission is the reaction: $^{9}\mathrm{Be} + \alpha \longrightarrow {}^{12}\mathrm{C} + \mathrm{n}$. This source consists of an α-emitting nuclide, e.g. Am, Pu, Po or Ra, mixed with beryllium. The α-emitter and the beryllium are usually in powder form and are doubly sealed in stainless-steel capsules.

Neutrons Emitted by Photonuclear Reactions The binding energy of the last neutron in ^{9}Be is only 1.67 MeV, and neutrons can therefore be released from ^{9}Be by photodisintegration, i.e. by bombardment with γ-rays of energy 1.67 MeV or more.

Fission Sources The spontaneous fission of ^{252}Cf, which is an artificial isotope produced in nuclear reactors, is a source of neutrons. The Cf nucleus splits spontaneously into two fission products and emits one or more neutrons.

5.1.2 Neutron generators

Intense neutron radiation can be obtained by bombarding various targets with charged particles from a Van de Graaff accelerator or from a cyclotron. Commercially available neutron generators use one of the following reactions: $^{2}\mathrm{H} + {}^{2}\mathrm{H} \longrightarrow {}^{3}\mathrm{H} + \mathrm{n}$ (D–D reaction) or $^{2}\mathrm{H} + {}^{3}\mathrm{H} \longrightarrow {}^{4}\mathrm{He} + \mathrm{n}$ (D–T reaction). Neutron generators based on these reactions produce neutrons of energies 2.6 MeV and 14 MeV respectively.

5.1.3 The Nuclear Fission Reactor

The nuclear fission reactor provides neutron fluxes of high intensity close to the reactor core.

5.2 Interaction of Neutrons with Matter

A fast neutron suffers many interactions in the process of losing its energy to reach thermal energy. This process is called slowing down. The neutron may disappear during this process as a result of nuclear reactions when the neutron impinges on nuclei forming the scattering medium. The probability of scattering from or reaction with a single nucleus is called cross-section (σ). The unit for σ is the barn and is equal to 10^{-24} cm^2.

There are four types of interactions that can occur between the neutron and the surrounding nuclei: (1) particle reactions (n,*x*), e.g. (n,α), (n,p), (n,*x*n) (for neutrons of high energy only), (2) inelastic scattering (n,n′γ) (the neutron energy must be above a threshold characteristic for each element, normally above 1 MeV), (3) elastic scattering and (4) radiative capture.

5.2.1 (n,x) Reactions

A nuclear reaction involves the release of a charged particle or more than one neutron. The final reaction product cannot be the same as the target nucleus. Characteristic γ-rays are emitted and the neutron reactions can be used to identify the presence of particular elements in materials or for the determination of their concentration.

5.2.2 Inelastic Scattering

A neutron can interact with a nucleus of atomic number Z and mass number A to form a compound nucleus in an excited state of mass $A + 1$. The compound nucleus decays very rapidly to the ground state by emitting a γ-ray, and it then emits a neutron. The γ-rays that are produced are characteristic of the energy levels of the compound nucleus and can be used to identify the target nucleus.

5.2.3 Elastic Scattering

Elastic scattering is the most important interaction for neutrons produced by isotopic sources (average energy below 4.5 MeV). This interaction is also known as ''billiard ball'' scattering. Although the neutron reactions and inelastic scattering reduce the neutron population, their importance is modest by comparison with neutron scattering. Inelastically scattered neutrons can suffer large

energy loss; however, they constitute a small fraction of the neutrons slowing down to low energies. Most of the neutrons reach low energies through repeated elastic collisions. In elastic scattering, a percentage of the incident neutron's kinetic energy is transferred to the recoiling nucleus and the neutron will have less kinetic energy after collision.

A useful quantity in the theory of the neutron slowing-down process is the percentage decrease in the neutron energy as the result of a collision. This is usually expressed as the *average logarithmic energy decrement* ξ, also known as neutron *lethargy*. For nuclei with mass $A > 10$, which are common in the Earth's crust, $\xi = 2/(A + 2/3)$. For hydrogen $\xi = 1$. Typical values of ξ for a few common nuclei are: H(1), C(0.158), O(0.12), Al(0.072), Si(0.070), Ca(0.049), Fe(0.035). The lethargy is used to calculate the mean number of scatterings required to slow down the neutron. For example, the average number of scatterings calculated for a few elements to slow down a neutron of 14 MeV, produced by a neutron generator, to thermal energy (0.01 eV) are: H(19), C(112), O(154), Al(290), Si(297), Fe(539). These examples show the unique position of hydrogen in slowing down neutrons. Its position is further enhanced when taking into account scattering cross-sections. Most of the nuclei abundant in the Earth's crust have scattering cross-sections of several barns. The proton's scattering cross-section between $\approx 10^5$ eV and 0.5 eV is about 20 barn.

The cross-section for elastic scattering, the relationship between the neutron energies before and after scattering and the theoretical expressions for lethargy are given in Appendix 1.

5.2.4 Radiative Capture

Once slowed down to thermal energies, neutrons diffuse through the medium without further loss of energy until their life is terminated by other processes such as radiative capture. In this process, the thermal neutron enters the nucleus, producing a compound nucleus in an excited state, which then decays to the ground state by the emission of one or more γ-rays that are characteristic of the capture nucleus. The radiation capture process only takes 10^{-11} s, which is virtually instantaneous compared with the initial slowing down and diffusion process, which may take several hundred microseconds. The product nucleus may be stable or it may decay (with a half-life of between a few microseconds and a number of years) to another product nucleus, often with β-particle emission. This is called thermal neutron activation.

5.3 Macroscopic Treatment of Slowing Down

For a moderating medium consisting of a homogeneous collection of different elements, the elastic macroscopic scattering cross-section Σ_s can be written as $\Sigma_s = \sum_i n_i \sigma_{si}$, where n_i is the number of nuclei of the ith element per cubic centimeter and i represents the existing elements. The same applies to the absorption cross-section $\Sigma_a \cdot \Sigma$ is analogous to μ in γ-ray attenuation. For a beam of neutrons passing through a material of thickness x, $N = N_0 e^{-\Sigma x}$ where N represents the number of neutrons that have had no interaction. The mean free path λ is defined as $\lambda = 1/\Sigma$. For a neutron density of N neutrons per cubic centimeter, the number of interactions per cubic centimeter is $N\nu\Sigma$, where ν is the velocity of the neutron. The quantity $N\nu$ is called the flux ϕ and has units of neutrons per square centimeter per second. $\langle\xi\rangle$ for the moderating medium is the weighted-average cross-section taken over all i nuclei: $\langle\xi\rangle = \sum_i (n_i \sigma_{si} \xi_i)/\Sigma_s$.

The theory of neutron slowing down and diffusion is presented in books dealing with nuclear reactor theory. It was developed by Glasstone and Edlund[8] for nuclear reactors. The theory was later developed for nuclear logging by other authors. Neutron slowing down ends when the neutrons reach thermal equilibrium with the atoms and molecules of the medium. Once neutrons have reached thermal energies, they diffuse through a medium in the same way that one gas diffuses through another, except that their lifetime is limited by absorption. If there was no absorption, the distribution of neutron energies would follow the Maxwell–Boltzmann gas distribution law; the most probable energy, defined as the energy corresponding to the peak of the distribution curve, would be kT, where T is the temperature in Kelvin (absolute temperature) and k is the Boltzmann constant (8.61×10^{-5} eV K^{-1}). For a temperature of 25 °C, kT is about 0.025 eV. The neutron population is given by the diffusion equation which describes conservation of neutrons: Production + scattering in − absorption − leakage = 0.

The Boltzmann transport equation derived for gases also applies to neutrons. Solving the Boltzmann equation is not an easy task. For more accurate results, the neutrons are treated as belonging to several energy groups (multigroup diffusion theory).

6 RADIATION DETECTORS

The γ-ray detectors most commonly used in nucleonic gauges can be divided into three categories: gas-filled counters, scintillation detectors and solid-state detectors. The gas-filled counter is one of the oldest detectors for nuclear radiation. The Geiger–Mueller counter is still used in nuclear applications for its reliability and low cost. The scintillation detector is one of the most widely used detectors. Popular scintillators are NaI(Tl), CsI(Tl)

or CsI(Na) and $Bi_4Ge_3O_{12}$ (commonly referred to as BGO). The solid state detector is used in applications when high-energy resolution is required.

Common neutron detectors are the boron counter, based on the (n,α) reaction $^{10}B + n \longrightarrow {}^{7}Li + \alpha + \gamma$, the ^{3}He detector, based on the (n,p) reaction $n + {}^{3}He \longrightarrow {}^{3}H + p$ and special scintillators developed for neutron detection.

7 THE DETERMINATION OF WATER CONTENT IN SOIL AND BOREHOLES USING NEUTRONS

The principle of the determination of water content is the slowing down of fast neutrons. Fast neutrons emitted from a neutron source successively undergo the processes of slowing down, thermalization and diffusion. Hydrogen primarily determines the slowing down power of the medium, but it does not dominate its neutron transport property because of the presence of elements with very large neutron absorption cross-sections (e.g. Cl, Cd, In, Sm, Gd, B). These elements serve to keep the scattered neutrons in the vicinity of the source.

7.1 Definition of Porosity

It is common to express the water content in terms of porosity. The bulk density of an ideal formation consisting of uniformly distributed, fluid-filled pores in a rock matrix is $\rho_B = \Phi\rho_f + (1 - \Phi)\rho_{ma}$. Here, ρ_f and ρ_{ma} are the fluid and rock matrix densities, respectively, and Φ is the porosity. The porosity is given by Equation (15):

$$\Phi = \frac{\rho_{ma} - \rho_B}{\rho_{ma} - \rho_f} \tag{15}$$

When neutrons travel through matter from the point source, their fluxes decrease with distance. The rate of decrease is governed by a ''scaling unit'', which is a property of the medium. The scaling units are the slowing-down length L_s, which is a measure of the ability of the bulk medium to reduce the neutron energy to a final energy in a given distance, the diffusion length of thermal neutrons L_d (the distance that thermal neutrons travel until they are absorbed), and the migration length L_m of the neutrons from their original state as primary neutrons to the point of their absorption in the medium. These distance variables are related by the equation: $L_m^2 = L_s^2 + L_d^2$. Experimental data show that the values of L_s and L_d depend mainly on the porosity of the rocks. This is because there is a certain degree of saturation of the pore space by water (or hydrocarbons when applied to well logging).

A neutron moisture gauge consists of a neutron source and one or two neutron detectors. The instruments are normally calibrated in terms of porosity or water content. The shape of the calibration curve relating detector count rate to porosity or water content is influenced by source–detector distance/geometry. There are two basic approaches to measuring the porosity: the short-spaced (source–detector) sonde used for soil moisture determinations and the long-spaced sonde used for borehole logging.

8 THE MONTE CARLO METHOD

The Monte Carlo method is a numerical procedure based on statistical theory. It is used to calculate the distribution of γ-rays or neutrons produced by one or more γ-ray or neutron sources in a medium of known chemical composition. Particle fluxes and response functions (reaction rates) in selected volume elements can be calculated with fair accuracies. The technique is useful in the design of nucleonic gauges to predict the optimum source–detector–shielding configuration. The gauge response for a particular configuration can be tested in a short time at low cost on PCs using Monte Carlo calculations. Thus, tedious, time-consuming and expensive laboratory measurements can be bypassed.

The concept of the Monte Carlo technique is tracking the transport through matter of neutrons or γ-rays by simulating the statistical nature of the interaction processes. This is based on the fact that the macroscopic cross-sections may be interpreted as a probability of interaction per unit distance traveled by the neutron or γ-ray. In Monte Carlo calculations, a set of neutrons/γ-rays histories is generated by following individual neutrons/γ-rays through successive collisions. The locations and results of collisions are determined from the range of possibilities by sets of random numbers. A large library of experimental nuclear data is used in the program to select the random numbers that influence the particle history. These include the direction of the first flight on leaving the source, the choice of interaction from those possible at each collision, the mean free path between collisions and the angle of emergence after collision. Tracking the particle finishes if the particle is either absorbed, passes a lower energy limit set by the user or crosses an outer boundary. Then the history of a new particle starts. In order to decrease the statistical errors of predictions obtained from Monte Carlo calculations, the tracking of many particle histories is needed. This is becoming more and more affordable, in both time and cost, with the advent of the new generation of powerful PCs. Codes for Monte Carlo calculations on PCs are readily available and

the technique has become a powerful tool in nucleonic gauge design and nuclear applications.

9 NUCLEAR TECHNIQUES FOR IN SITU ANALYSIS

Nuclear logging is well established and used routinely in the oil, gas, uranium and coal industries. It is essential for the oil and gas industries, in which very deep holes are drilled (thousands of meters). Owing to the deep penetration of γ-radiation and neutrons, nuclear logging can locate the presence of oil or gas behind the well casing.

The laboratory analysis of core samples retrieved from boreholes and nuclear logging are complementary. Although the core can provide all the information that can be extracted from a borehole, nuclear logging is able to provide information almost instantaneously. The volume of rock sampled by nuclear borehole logging is also much larger than the core samples and thus provides better sampling statistics, especially in heterogeneous deposits. There are a number of books published in this field: Hallenburg;[9] Hearst and Nelson;[10] Tittman;[11] Ellis.[12] There are other books dealing with more specific aspects of nuclear logging. More recently, review papers have been published by Ellis et al.,[13] Mills et al.,[14] and Borsaru.[15]

Nuclear borehole logging techniques are either passive (natural γ) or active. In passive logging, the natural radiation in the borehole is measured by an appropriate detector, whereas in active logging an artificial radioactive source provides the radiation measured by the detector. Figure 4 shows a schematic diagram of a basic borehole logging system for nuclear logging. The borehole logging tool shown in the figure is fitted with one radiation detector. The source can be either a γ-ray source or a neutron source. In the absence of the source the tool is used for passive logging. Nuclear logging can be classified according to the radioactive source employed in the logging tool.

9.1 Passive Logging – Natural γ-Ray

The *γ-ray spectral log* technique provides a spectrometric measurement of the γ-radiation of rocks whereas the γ-log measures the intensity of the total natural γ-radiation

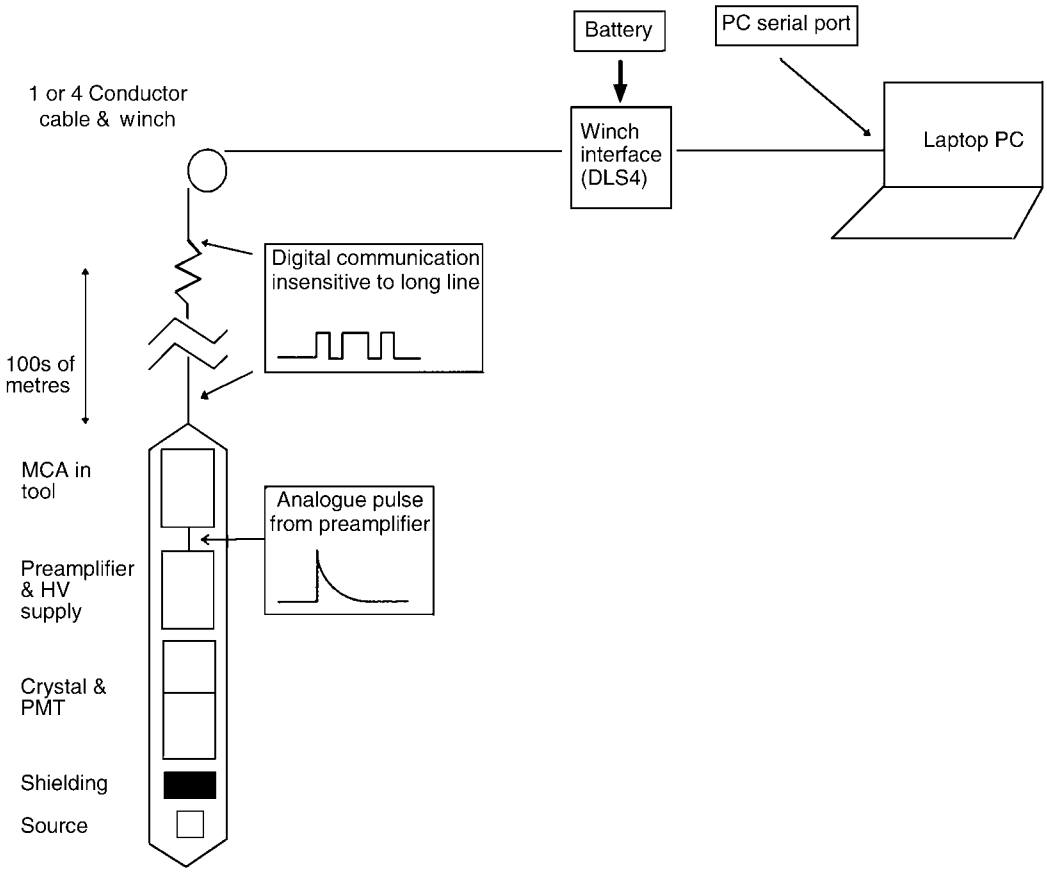

Figure 4 Schematic representation of a nuclear logging system.

of the rock. These techniques have widespread applications in petroleum and uranium exploration, potash, phosphate, coal, heavy minerals, iron ore and other mineralizations. In the γ-ray spectral log, the potassium, uranium and thorium components of the natural radiation are determined and the K, U, Th, U/K, Th/K and U/Th measurements are used for borehole lithology determinations and correlations. A review of γ-ray spectral logging has been written by Fertl.[16]

9.2 Active Logging

The backscattered γ-ray technique was developed for measuring the rock density, providing information on the type of rock encountered by the borehole, and in special cases for grade control. Figure 5 shows schematically a backscattered γ-ray spectrum obtained with a ^{60}Co γ-ray source. The backscattered spectrum is dependent on the intensity and energy of the primary γ-ray source employed in the logging tool and the type of rock intersected by the borehole. Czubek[1] reviewed the theory of γ–γ logging. The theory shows that the spectral intensity at high energy (>180 keV) is determined by the electron density, and at low energies by both the electron density and Z_{eq} (see Equations 13 and 14). The count rate in the density region is related to the bulk density of the rock interrogated by the γ-radiation. The physical basis for the density measurement is that the Compton cross-section per electron is essentially independent of the atom in which the electron is bound, and in this energy region the only interaction of consequence is Compton scattering. The electron density is directly proportional to bulk density. The ratio of counts recorded in energy windows at high and low energies is proportional to Z_{eq}. Since Z_{eq} is strongly dependent on the element of highest atomic number, e.g. Fe in iron ore, even a small variation of that element in a low-atomic-number matrix will cause significant variation in Z_{eq}. This is the principle of determining Fe grade in iron ore, or ash in coal, by the γ–γ technique. Mathew and Anderson[17] have shown that the ratio of count rates in two energy windows at high energies (>500 keV) is proportional to the borehole diameter. In order to extract the most information from the backscattered γ-ray spectrum, the γ–γ logging must be spectrometric and automatic gain stabilization must be incorporated in the system.

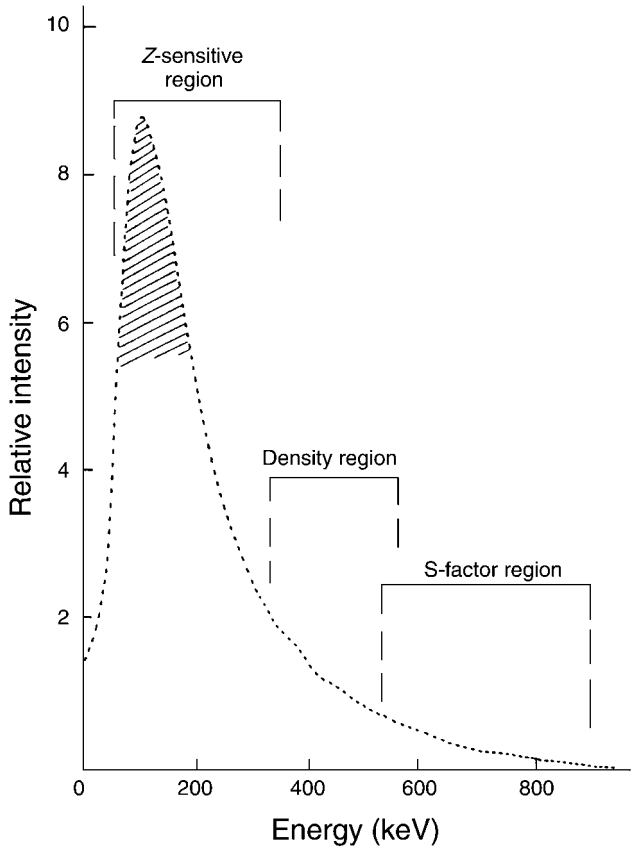

Figure 5 Typical γ–γ spectrum obtained with a ^{60}Co γ-ray source. (Reproduced from 'Use of Backscattered Gamma Radiation for Determining Grade of Iron Ores in Blast Holes and Development Drill Holes', *Proceedings of the Australasian Institute of Mining and Metallurgy*, (1977). Reprinted with permission of the Australasian Institute of Mining and Metallurgy.)

Prompt γ neutron activation, neutron inelastic scattering and *neutron activation* are useful techniques in borehole logging. They require the use of a neutron source. The type of technique chosen depends on the particular application, the nuclei involved and the energy of the neutrons. When the neutron source is a neutron generator, both neutron inelastic scattering and prompt γ neutron activation can be performed simultaneously with the neutron generator operating in a pulsed mode. This technique is largely used in oil well logging.

9.3 Oil Well Logging

9.3.1 γ-Ray Source

The bulk density of rock formations is an important parameter in oil well logging. The new generation density-logging tool, the litho-density tool, measures both the formation's bulk density and its photoelectric absorption, using a ^{137}Cs γ-ray source. This is carried out by recording the count rates in the Z-sensitive and density regions of the backscattered γ-ray spectrum shown in Figure 5. The photoelectric absorption P_e gives a measure of the average atomic number of the formation, which provides an indication of the lithology (sandstone, limestone, dolomite, etc.). The tool has two detectors at different spacings from the γ-ray source and has gain stabilization. Porosity can also be determined from the bulk density and P_e.

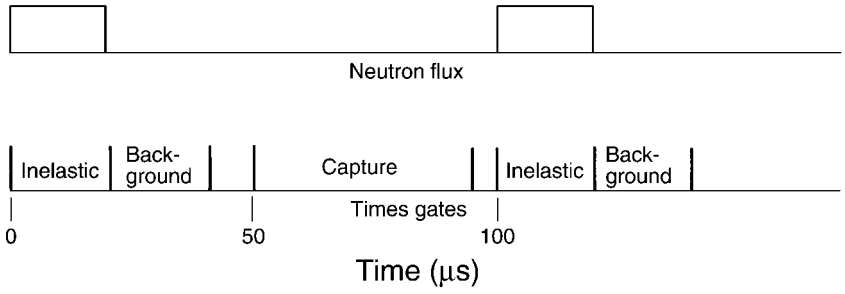

Figure 6 Timing of gates for a logging tool using a neutron generator.

9.3.2 Neutron Generator

The neutron generator is well established in oil well logging. Owing to the high energy of the neutrons produced by the generator (14 MeV), the depth of investigation is greater than that for radioactive neutron sources. It can also excite many nuclear reactions that cannot be produced by steady-state sources. However, probably the most important feature of the neutron generator is that it can be pulsed, and this allows the development of special techniques that cannot be achieved with radioactive sources.

Neutron Die-away or Neutron Lifetime This is an old technique developed for the determination of the chlorine content of formation fluid. A pulsed neutron source is needed and the neutron generator is an ideal one. The basis of this technique is to measure the change in neutron flux with time after a neutron burst. The fast neutrons emitted during the short neutron burst are slowed down to thermal energy and diffuse in the formation. They die when absorbed in a capture process by nuclei from the formation, and capture γ-rays are released. The number of capture γ-rays is proportional to the number of thermal neutrons present in the formation. The determination of neutron die-away is based on measuring the ratio of γ-ray count rates at two different times after the fast neutrons produced by the burst have decayed to thermal neutrons.[18] The die-away log provides information to distinguish oil (low-capture cross-section with resulting slow γ-ray die-away rate) from saline water (high-capture cross-section and fast die-away rate) behind the steel casing in wells.

γ-ray Spectroscopy Using a Neutron Generator Using a neutron generator in a pulsed mode makes it possible to separate the γ-rays released by neutron inelastic scattering from the γ-rays generated by the neutron capture process. Inelastic scattering γ-rays are emitted within a few nanoseconds of the fast neutron interaction and occurs during the neutron pulse. Capture γ-rays are produced after tens of microseconds by thermal neutrons traveling through the formation. Figure 6 shows schematically the gate timing of this tool. The tool is used to measure the carbon/oxygen and calcium/silicon ratios. By measuring both ratios, one can distinguish carbon in calcium carbonates from that in hydrocarbons and thus the presence of oil can be established. The energy of γ-rays released from neutron inelastic scattering with C, O, Ca and Si are 4.44, 6.1, 3.7 and 1.78 MeV respectively. A background spectrum is recorded in a time window of the same length, but with the generator turned off (see Figure 6). This is a capture spectrum and is subtracted from the inelastic spectrum.[19]

Schlumberger developed the Geochemical Logging Tool (GLT™). This tool makes the following spectral measurements: natural γ-rays produced by K, Th and U; Al by neutron activation with a ^{252}Cf neutron source; and elements in the rock formation measured by capture γ-ray spectrometry using a pulsed neutron generator.[20] The tool incorporates three scintillation detectors, a ^{252}Cf neutron source and a neutron generator. Grau and Schweitzer[21] developed a technique capable of measuring the absolute concentration of a number of elements present in the formation. The technique assumes that the composite spectrum measured in boreholes can be represented by a linear combination of known standard spectra. The weighted least-squares method of fitting the entire spectrum with the set of standard spectra extracts the chemical composition. The measured spectrum is decomposed into elemental standards. Carbon and oxygen are not measured by the GLT™ but are accounted for by the chemical combination of each element measured in the capture spectrum as a single oxide or carbonate. The tool is used to derive mineralogy and petrophysical parameters from the elemental concentrations.

9.3.3 Steady-state Neutron Sources

Am–Be is used for porosity measurements. Single-detector tools employing mostly epithermal neutron detection or two-detector neutron sondes employing thermal neutron detection are commonly used in oil well logging for porosity measurements.

9.3.4 Natural γ-Ray Spectral Logging

Natural γ-ray spectral logging provides information about the presence of clays in the formation matrix. The tendency is to replace the natural γ-ray tools recording the count rates in the three K/U/Th windows with fully spectrometric tools, which record the whole spectrum in 256 channels.

9.3.5 Measurements While Drilling

Measurements while drilling are used to gather information on the formation being drilled in real time. The detectors are mounted in the drilling string near the bit. The detectors are either scintillators or Geiger–Mueller tubes that have been ruggedized to withstand the drilling process. γ-Ray, neutron porosity and density measurement-while-drilling (MWD) tools have been developed and are widely used.

9.4 Uranium

9.4.1 Natural γ-Ray Logging

γ-Rays emitted from the uranium minerals are routinely used for both quantitative and qualitative determination of uranium in exploration or ore-body evaluation. The γ-rays are ideal to use as intrinsic indicators for uranium as they occur naturally, sample relatively large volumes of material and are simple to detect and identify. The decay series of ^{238}U (excluding some short-lived isotopes) is:

$$^{238}U \xrightarrow{4.5\times10^{9}\,a} {}^{234}U \xrightarrow{2.5\times10^{5}\,a} {}^{230}Th \xrightarrow{8\times10^{4}\,a} {}^{226}Ra \xrightarrow{1620\,a} {}^{222}Rn \xrightarrow{3.8\,d} \ldots {}^{214}Pb \xrightarrow{27\,min} {}^{210}Pb \xrightarrow{20\,a} {}^{206}Pb(\text{stable})$$

When secular equilibrium is reached after some 10^6 years, the activity of each isotope, except ^{206}Pb, in the decay series is equal. Disequilibrium can occur throughout the chain if one or more of the daughter products is lost by any process other than radioactive decay. Since each daughter product is an element with its own characteristic physical and chemical properties, it may behave differently within a given environment. For example, ^{222}Rn is a gas. Also, the solubilities of radium, uranium and thorium differ, and preferential leaching of elements may occur.

Total γ-ray Logging This is probably the easiest logging technique for quantitative analysis. The total number of γ-rays can be recorded with either a Geiger counter or a scintillator. The logging probes are calibrated in special uranium test pits. Since this method records γ-rays from many daughters of uranium, it can only be applied for quantitative analysis if both the test pit and uranium ore body are in secular equilibrium.

Accurate Spectral γ-ray Logging The principle of this technique for the in situ quantitative determination of uranium[22] is measurement of the γ-radiation (1001 keV) produced during the decay of protactinium-234 (^{234}Pa) to uranium-234. In the uranium decay series, ^{238}U first decays to ^{234}Th, which decays to ^{234}Pa, which in turn decays to ^{234}U. Because the half-lives of all three reactions are short, it is unlikely that in nature the ^{238}U and ^{234}Pa will become separated. Thus, the determination of the ^{234}Pa grade from the count rate of the 1001 keV γ-ray gives a measure of the ^{238}U grade in the ore body. This method is independent of the amount of disequilibrium in the ore body. A high resolution (intrinsic Ge detector) is necessary to measure the count rate of the 1001 keV, which is weak in comparison with other γ-ray lines in the natural γ-spectrum.

9.4.2 Delayed Fission Neutrons

Fission occurs when uranium nuclei interact with neutrons and the nuclei split into fission products, emitting neutrons. Two to three percent of the fission products are unstable and β-decay, forming other nuclei. By doing this, each fission product emits one neutron. These are called delayed neutrons and are emitted with different half-lives, between 0.2 and 50 s. This is the basis of the delayed fission neutron technique for quantitative analysis of uranium and is independent of the state of equilibrium of the uranium ore body. The technique consists of irradiating the uranium ore with neutrons, removing the neutron source and counting the delayed neutrons after the source neutrons disappear.

The neutron generator is well suited for this application.[23] The generator is pulsed and then turned off. When the neutrons produced by the generator die away, the delayed neutrons are counted. The effect of diffusion of the neutrons in the medium must be accounted for. Background subtraction is achieved using the reaction $^{17}O(n,p)^{17}N \longrightarrow {}^{16}O + n$.

^{252}Cf is also used as a source of neutrons[24] by shuttling the source inside the moving probe between two detectors; while one detector is counting the delayed neutrons, the source is irradiating the ore body. Owing to the much lower energy of the neutrons produced by ^{252}Cf, no background subtraction is needed (the energy of the neutrons is below the threshold energy of the $^{17}O(n,p)$ reaction). Correction for diffusion of the neutrons in the uranium ore is needed.

9.4.3 Prompt Fission Neutrons

The neutrons emitted immediately after fission can also be used to measure the uranium concentration.[25,26] The

technique uses a neutron generator as the primary source of neutrons. An epithermal neutron detector is gated off during a neutron pulse and starts counting epithermal neutrons immediately after the pulse, while thermal neutrons are present and causing fission. The uranium concentration is determined from the ratio of epithermal neutrons to (i) thermal neutrons or (ii) thermal neutron die-away time.

9.5 Coal

Nuclear borehole logging is used routinely in the exploration for coal and its acceptance in the mining stage for coal quality control is increasing. The parameters determining the economic value of coal are the thickness and depth of the coal seams and the quality of coal. Coal quality is determined by the ''raw ash'' content, calorific value, rank, moisture content, volatiles and the elemental composition of the ash. These parameters are determined from the coal core in the laboratory. However, this complete information is not always needed. In many cases it is sufficient to measure only thickness, depth and the raw ash content of the coal seams, and this can be achieved by nuclear logging in ''easier to drill open holes''. Natural-γ, $\gamma-\gamma$ and prompt neutron$-\gamma$ techniques are all used for coal logging. Natural-γ is used to delineate the coal seams based on the fact that coal is low in natural radiation compared with the interseam sediments.

The $\gamma-\gamma$ Technique This technique was developed for the coal mining industry in the 1970s.[27,28] It is used mostly for the delineation of coal seams based on the difference in density between coal and sediments. The ash content of the coal seams can be derived indirectly from the correlation between ash and density. Good density–ash correlations are found for many coal deposits. However, this correlation is not universal. A direct measurement of ash in coal is achieved using the spectrometric $\gamma-\gamma$ technique.[5] This technique relies on the fact that variations in the ash content of coal correspond to changes in Z_{eq}. When the ash content and Z_{eq} are uniquely correlated, ash content can be measured by measuring changes in Z_{eq}. This is carried out by calculating the P_z ratio between count rates recorded in two broad spectral windows at high and low energies in the backscattered γ-ray spectrum. This measurement requires gain stabilization.

The Prompt Neutron$-\gamma$ Method for Determination of Ash in Coal Owing to its high hydrogen content, coal is an excellent medium for the neutron capture technique. The neutrons emitted by the neutron source are thermalized by colliding with the hydrogen nuclei present in coal and subsequently interact with the nuclei in the coal matrix. The basis for ash determination is the correlation that exists between ash and the main constituents of ash, i.e. Al, Si and Fe or a combination of two of these elements. Charbucinski et al.[29] developed a method based on the neutron$-\gamma$ technique for the determination of ash content, depth and thickness of coal strata. The technique was later extended for the measurement of Fe in coal.[30]

Coal Face Ash Analyzer Coal ash determination at the coal face falls into the category of in situ determination. A portable instrument capable of monitoring the quality of exposed coal seams and differentiating between coal seams and ''look-alike'' coal sediments is useful for the coalmining industry. Wesolinski and de Jesus[31] developed a technique for the determination of ash on the coal face based on the simultaneous measurement of the backscattered γ-radiation at two different energies. They developed a portable instrument that can locate the coal–seam interface and measure the ash content on the coal face. The activity of the γ-ray sources employed in the instrument was 185 MBq. Borsaru et al.[32] developed a coal face analyzer based on the measurement of natural γ-rays. The technique relies on the correlation between the natural γ-radiation of the coal and its ash content. Another coal face analyzer based on the backscattered $\gamma-\gamma$ technique was developed more recently by Borsaru et al.[33] It uses two microsources (^{133}Ba and ^{137}Cs) of total activity 2.2 MBq. By employing such low-activity γ-ray sources, the analyzer does not need special shielding and the user is not exposed to unacceptable levels of radiation.

9.6 Iron Ore

The backscattered $\gamma-\gamma$ technique has been employed for the simultaneous measurement of iron grade, density and borehole diameter in wide (310 and 380 mm) dry blast holes in iron ore deposits.[4] The primary γ-ray source was 165 MBq ^{60}Co. The backscattered γ-ray spectrum was stabilized using a ^{60}Co microsource. The borehole diameter was measured using the ratio (S-factor) between count rates in energy windows at intermediate energies (400–700 keV) and at high energies (700–950 keV).[17] The source-to-detector separation was 70 cm. The iron grade determination was based on the P_z technique described earlier. The accuracy for ore grade obtained in 1.5 m splits was 1.9% Fe in ore in the range from 35% to 69% Fe and 3% Fe for ores in the range of iron content 9–69% Fe.

ABBREVIATIONS AND ACRONYMS

MWD Measurement-while-drilling

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
Iron Ore, Sample Preparation and Analysis of

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Photon Activation Analysis • Prompt γ-Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

APPENDIX 1

The Compton Effect

The following expressions (Equations 16–18) can be derived for the energy $h\nu'$ of the scattered photon, the energy T of the scattered electron and the relationship between the angles of scattering shown in Figure 1:

$$h\nu' = \frac{h\nu}{1+\alpha(1-\cos\theta)} \tag{16}$$

$$T = h\nu\frac{\alpha(1-\cos\theta)}{1+\alpha(1-\cos\theta)} \tag{17}$$

$$ctg\phi = (1+\alpha)tg\tfrac{1}{2}\theta \tag{18}$$

where $\alpha = h\nu/m_0c^2$ and m_0 is the rest energy of the electron. According to Equation (16), the energy after scattering is not independent of the energy before scattering. The mean relative energy loss decreases with decreasing γ-ray energy.[1] Equation (16) shows that the energy of the scattered γ-ray is a function of the angle of scattering, and the maximum energy that an electron can acquire is less than the total energy of the γ-ray: it is, at most, 225 keV.

Elastic Scattering of Fast Neutrons

The cross-section for elastic scattering for fast neutrons[34], as shown in Equation (19), is approximately:

$$\sigma_{el} \approx \pi\left(R + \frac{\lambda}{2\pi}\right)^2 \tag{19}$$

where $R_0 = 1.4 \times 10^{-13}$ cm and $R = R_0A^{1/3}$ is approximately equal to the radius of the nucleus with mass number A. $\lambda/2\pi$ is the wavelength equal to the radius of the neutron and is equal to $4.44 \times 10^{-13}/E_n$, where E_n is the neutron energy in MeV. When a neutron interacts elastically with a nucleus, the relation between the neutron energy before and after scattering, E_1 and E_2 respectively, is illustrated in Equation (20) as:

$$\frac{E_2}{E_1} = \frac{A^2 + 2A\cos\theta + 1}{(A+1)^2} \tag{20}$$

where A is the mass of the target nucleus and θ is the scattering angle in the center-of-mass system. Equation (20) reveals four features of elastic scattering: the fractional energy loss depends only on the mass ratio and scattering angle and is independent of energy. The minimum possible value $E_{2\,min}$ of E_2 (corresponding to $\theta = 180^\circ$) is $E_{2\,min} = [(A-1)/(A+1)]^2$ and is the result of a head-on collision. For $\theta = 0$, a glancing collision, $E_2 = E_1$, and no energy loss occurs. When $A = 1$, corresponding to collision with a hydrogen nucleus, for $\theta = 180^\circ$ (head-on collision), $E_2 = 0$ and all the neutron's kinetic energy is transferred to the recoiling proton.

The percentage decrease in the neutron energy as the result of a collision, expressed as the *average logarithmic energy decrement* ξ, is known as neutron *lethargy* and is defined in Equation (21) as:

$$\xi = \ln\left(\frac{\overline{E_1}}{\overline{E_2}}\right) = \frac{\int_{E_1}^{E_{2\,min}} [\ln(E_1/E_2)]\,P(E_2)\,dE_2}{\int_{E_1}^{E_{2\,min}} P(E_2)\,dE_2} \tag{21}$$

where $P(E_2)$ is the probability that the scattered neutron will have energy E_2. When the integrations are performed, it is found that the mean lethargy increase per collision, ξ, illustrated in Equation (22) is:

$$\xi = 1 + \frac{(A-1)^2}{2A}\ln\frac{A-1}{A+1} \tag{22}$$

REFERENCES

1. J.A. Czubek, 'Advances in Gamma-gamma Logging', in *Nuclear Geophysics, Selected Papers on Applications of Nuclear Techniques in Mineral Exploration, Mining and Process Control*, ed. C.G. Clayton, Pergamon Press, Exeter, 153–172, 1983.
2. O. Klein, Y.Z. Nishina, 'Über die streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac', *Z. Physik*, **52**, 853 (1929).
3. C.M. Davisson, 'Interaction of γ-radiation with Matter', in *Alpha-, Beta- and Gamma-ray Spectroscopy*, ed. K. Siegbahn, North-Holland, Amsterdam, 33–78, Vol. 1, 1965.
4. P.L. Eisler, J. Charbucinski, M. Borsaru, A.P. Indrans, 'Spectrometric Gamma-gamma Borehole Methods for

Iron Ore Mining', Proceedings of the Second International Symposium on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications, 6–8 October, Golden, Colorado, USA, 189–201, 1987.
5. M. Borsaru, J. Charbucinski, P.L. Eisler, S.F. Youl, 'Determination of Ash Content in Coal by Borehole Logging in Dry Boreholes using Gamma-gamma Methods', *Geoexploration*, **23**, 503–518 (1985).
6. P.J. Mathew, 'On-stream Coal Ash Analysis Based on Natural Gamma Ray Activity', Nuclear Techniques in the Exploration and Exploitation of Energy and Mineral Resources, IAEA-SM-308/89, 93–103, 1991.
7. P.J. Mathew, 'A Natural γ-Ray Gauge for Green Liquor and Red Mud', Proceedings of the International Conference on Applications of Radioisotopes and Radiation in Industrial Development, 1–3 March, Bombay, 1984, eds. S.M. Rao, A.B. Majali, R.G. Deshpande, T.S. Murthy, Wiley Eastern, Bombay, 154–163, 1984.
8. S. Glasstone, M.C. Edlund, *The Elements of Nuclear Reactor Theory*, Van Nostrand, Princeton, NJ, 1952.
9. J.K. Hallenburg, *Geophysical Logging for Mineral and Engineering Applications*, Pennwell Books, Penwell Publishing, Tulsa, OK, 1984.
10. J.R. Hearst, P.H. Nelson, *Well Logging for Physical Properties*, McGraw-Hill, New York, 1985.
11. J. Tittman, *Geophysical Well Logging*, Academic Press, Orlando, FL, 1986.
12. D. Ellis, *Well Logging for Earth Scientists*, Elsevier, New York, 1987.
13. D.V. Ellis, J.S. Schweitzer, J.J. Ullo, 'Nuclear Techniques for Subsurface Geology', *Ann. Rev. Nucl. Part. Sci.*, **37**, 213–241 (1987).
14. W.R. Mills, D.C. Stromswold, L.S. Allen, 'Advances in Nuclear Oil Logging', *Nucl. Geophys.*, **5**, 209–227 (1991).
15. M. Borsaru, 'Nuclear Techniques for In Situ Evaluation of Coal and Mineral Deposits', *Nucl. Geophys.*, **7**, 555–574 (1993).
16. W.H. Fertl, 'Gamma-ray Spectral Logging: A New Evaluation Frontier', *World Oil*, **196**(4), 79–91, **196**(5), 87–98, **196**(6), 147–155, **196**(7), 189–204, **197**(2), 57–63, **197**(4), 99–112 (1983).
17. P.J. Mathew, M.R. Anderson, 'A Gamma-gamma Method for Measuring the Diameters of Air-filled Boreholes', Borehole Geophysics for Mining and Geotechnical Applications, ed. P.G. Killeen, paper 85-27, 269–275, Geological Survey of Canada, Canada, 1986.
18. A.A. Youmans, E.C. Hopkinson, R.A. Bergen, H.L. Oshry, 'Neutron Lifetime: A New Nuclear Log', *J. Petr. Tech.*, **16**, 319 (1964).
19. W.E. Schultz, H.D. Smith, 'Laboratory and Field Evaluation of a Carbon/Oxygen (C/O) Well Logging System', *J. Petr. Tech.*, **26**, 1103–1110 (1974).
20. R. Hertzog, L. Colson, B. Seeman, M. O'Brien, H. Scott, D. McKeon, P. Wraight, J. Grau, D. Ellis, J. Schweitzer, M. Herron, 'Geochemical Logging with Spectrometry Tools', *SPE Form. Eval.*, **4**, 153–162 (1989).
21. J.A. Grau, J.S. Schweitzer, 'Elemental Concentrations from Thermal Neutro Capture Gamma-ray Spectra in Geological Formations', *Nucl. Geophys.*, **3**, 1–9 (1989).
22. B. Tanner, R.M. Moxham, F.E. Senftle, 'Assay for Uranium and Determination of Disequilibrium by Means of In Situ High-resolution γ-Ray Spectrometry', US Geol. Survey Open-File Rept., USA, 77–571, 1977.
23. W.W. Givens, W.R. Mills, C.L. Dennis, R.L. Caldwell, 'Uranium Assay Logging using a Pulsed 14-MeV Neutron Source and Detection of Delayed Fission Neutrons', *Geophysics*, **41**, 468 (1976).
24. D.K. Steinman, J. Stokes, J.A. Adams, R. Voigt, C.S. Pepper, K.V. Laubscher, W.J. Smith, J. John, '252Cf-based Delayed-neutron Logging System for In Situ Assay of Uranium Ore', Trans. Soc. Prof. Well Log Analysts 21st Logging Symp., paper I, 1980.
25. R.W. Barnard, 'Interpretation of Prompt Fission Neutron Uranium Borehole Logs', *Trans. Am. Nuc. Soc.*, **28**, 107 (1978).
26. D.R. Humphreys, R.W. Barnard, H.M. Bivens, D.H. Jensen, W.A. Stephenson, J.H. Weinlein, 'Uranium Logging with Prompt Fission Neutrons', *Int. J. Appl. Rad. Isotopes*, **34**, 231 (1983).
27. J.R. Samworth, 'The Radiation Density Log Applied to the Resolution of Thin Beds in Coal Measures', SPWLA 3rd European Symposium, London, paper R, 1974.
28. D.R. Reeves, 'Application of Wireline Logging Techniques to Coal Exploration', Proceedings of the First International Coal Exploration Symposium, London, 18–21 May, ed. W.L.G. Muir, 112–128, 1976.
29. J. Charbucinski, S.F. Youl, P.L. Eisler, M. Borsaru, 'Prompt Neutron-gamma Logging for Coal Ash in Water-filled Boreholes', *Geophysics*, **51**, 1110–1118 (1986).
30. M. Borsaru, M.S. Biggs, W.J.F. Nichols, 'Neutron-gamma Logging for Iron in Coal and Implications for Estimating the Ash Fusion Characteristics at Callide Mine', *Nucl. Geophys.*, **7**, 539–545 (1993).
31. E.S. Wesolinski, A.S.M. de Jesus, *A Portable Coal Face Ash Monitor Based on Dual Energy Gamma Radiation*, Nuclear Techniques in the Exploration and Exploitation of Energy and Mineral Resources, IAEA, Vienna, 33–46, 1991.
32. M. Borsaru, C. Ceravolo, P. Waddington, Gu Wenhao, 'A Coal Face Ash Analyzer Based on Natural Gamma-ray Activity', *Nucl. Geophys.*, **6**, 383–390 (1992).
33. M. Borsaru, C. Ceravolo, G. Carson, T. Tchen, 'Low Radioactivity Portable Coal Face Ash Analyzer', *Appl. Radiat. Isot.*, **48**, 715–720 (1997).
34. R.D. Evans, *The Atomic Nucleus*, McGraw-Hill, New York, 1955.

Sensors in the Measurement of Toxic Gases in the Air

William A. Groves
University of Iowa, Iowa City, USA

1 Introduction **1**
1.1 Defining Chemical Sensors 2
1.2 Performance Characteristics 3
1.3 Defining Toxic Gases and Vapors 5
2 Sensor Types **6**
2.1 Electrochemical Sensors 6
2.2 Mass Sensors 9
2.3 Optical Sensors 12
2.4 Biosensors 13
3 Applications **14**
3.1 Industrial Hygiene Monitoring 15
3.2 Confined Space Monitoring 15
3.3 Process Emission Monitoring 16
3.4 Leak Detection 17
4 Quality Control and Maintenance **18**
4.1 Data Quality Objectives 18
4.2 Calibration 19
4.3 Maintenance 20
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

The use of sensors to measure toxic gases and vapors in air has increased rapidly due to several factors, including a growing public awareness of the hazards associated with many airborne chemicals, increasing regulatory requirements for industries using hazardous chemicals, and continuous advancements in the technology and instrumentation available to measure these compounds. The principal advantage of sensor systems for toxic gases and vapors relative to traditional sampling and analytical methods, is the ability to measure chemicals over short time periods, that is, in a matter of seconds or minutes, thus providing the information necessary to protect workers from acute health effects. Modern sensor-based instrumentation provides many options for processing and storing measurements, thereby allowing extreme flexibility in evaluating exposures to toxic gases and vapors relative to ceiling limits, short-term exposure limits (STELs), or longer term average concentrations based on user selectable time periods. This extreme range of options for processing data makes sensors an attractive, and in some cases, the only alternative for measuring toxic gases and vapors in air. Most of these devices can be broadly categorized as: (1) electrochemical sensors *which constitute the vast majority of devices currently available for detection of toxic gases and vapors; (2)* mass sensors *that are based on piezoelectric materials including surface acoustic wave (SAW) devices; or (3)* optical sensors *that rely on the interaction of electromagnetic radiation with an analyte, resulting in some alteration of the properties of the radiation. The most common applications of sensors for measuring toxic gases and vapors in air include industrial hygiene monitoring, confined space monitoring, process emission monitoring, and leak detection.*

1 INTRODUCTION

The use of sensors to measure toxic gases and vapors in air has increased considerably since the late 1960s due to a number of factors, including a growing public awareness of the hazards of many airborne chemicals and the resulting legislation designed to protect people and the environment. However, without rapid advances made over the same time period in the area of electronics technology, the current widespread use of sensors and modern instrumentation for measuring toxic gases and vapors would not be possible. Continuous development in the area of microlithography has led to significant decreases in the size and power requirements of the integrated circuitry used to process, display, and store the data generated by sensor-based instrumentation. Improvements in microfabrication techniques have also been applied to the sensors themselves, resulting in the ability to design and fabricate sensor elements with physical features on a submicron scale.[1] Interdigital transducers (IDTs), microelectrodes, chemically sensitive field-effect transistors (CHEMFETs), and SAW devices are just a few examples of sensing components that owe their existence to microfabrication techniques. Research and development in related fields also has led to new approaches and applications for sensing technologies, including the adaptation of optical fibers for use in sensor systems and the relatively recent introduction of biologically active materials that may be used to enhance the selectivity of many different sensing techniques. These ongoing advances in the field of sensor development have led to an increased capability to detect selectively and quantify concentrations of chemical species found in a variety of media. Given the industrial nature of our society and the large potential for exposure to a broad range of chemicals that exert toxic effects in an indiscriminate fashion, the importance of chemical sensors becomes

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

quite clear and justifies the considerable effort devoted to developing these devices.

1.1 Defining Chemical Sensors

As a result of the diverse backgrounds of researchers engaged in sensor research, there has been some ambiguity in the nomenclature applied to the field. The terms ''sensor'' and ''transducer'' have often been used interchangeably to refer to a device which provides an electrical output signal in response to some physical quantity, property, or condition which can be measured.[2] However, some researchers prefer to make a distinction between the transducer, which is usually defined as the element that converts one form of energy to another, and the sensor as a whole, which consists of the transducer and any associated processing circuitry needed to produce an electrical output related to the physical input.[3] In order to address this inconsistency, in the USA the National Research Council Committee on New Sensor Technologies[2] has adopted definitions for the following three terms: the *sensor element* performs the fundamental transduction mechanism that converts one form of energy into another, the *sensor* consists of the sensor element and its physical packaging and external connections, and a *sensor system* comprises the sensor and associated signal processing hardware (analog or digital) located either in the same package or discrete from the sensor itself. These definitions are consistent with those developed by others[4] and will be used throughout this discussion.

Chemical sensors are devices designed to detect the presence and concentration or quantity of a given chemical species. These sensors can be based on any characteristic of the analyte that can be used to distinguish it from the surrounding sample media, including chemical reactivity, optical or electrical properties, and mass.[1] The ideal chemical sensor would be an inexpensive, portable, robust device that responds selectively, accurately, and instantaneously to any concentration of a particular analyte present in any medium. In practice, however, the large number of possible chemical species ($>10^6$), the susceptibility of most transduction mechanisms to interfering chemical compounds and changing environmental variables, and the physical limitations of the systems needed for sample introduction and data processing all serve to make attainment of this idealized performance impossible. For these reasons, development of chemical sensor systems most often focuses on specialized areas of application, such as the detection of toxic gases and vapors in air, for which the analytical challenges and acceptable levels of performance can be more clearly defined.

Chemical sensors can be classified on the basis of different criteria, including the field of application, the type of analyte, and the physicochemical operating principle of the device.[4] The latter criterion has been used most often and will be employed for this discussion. Using this approach in the context of detecting toxic gases and vapors, most devices can be broadly categorized as *electrochemical*, *mass*, or *optical* sensors. A general description of each of these chemical sensor categories will be presented as an introduction, with a more detailed examination following later in the chapter.

Electrochemical sensor systems are based upon the measurement of electrical signals that can be related to the concentration of a chemical species. These devices can be further classified as amperometric, conductometric, or potentiometric, depending on whether current, resistance, or voltage, respectively, is used as the measure of response. The most common type of electrochemical sensor is the electrochemical cell, which consists of two or more electrodes, an electrolyte, and an electrical system used to measure or control voltage or current.[5] The category of electrochemical sensors is very broad and also includes systems based on semiconducting materials, such as CHEMFETs and ion selective field-effect transistors (ISFETs). These devices are particularly well suited for miniaturization using microfabrication techniques and the resulting small package and low power requirements yield an attractive alternative for many sensing applications.

The category of mass sensors comprises a group of devices based on the acoustic properties of piezoelectric materials.[6–8] These sensors employ a set of metal transducers, which launch acoustic waves into the piezoelectric substrate at ultrasonic frequencies ranging from one to several hundred megahertz. The device will operate at some resonant frequency that is determined by the geometry of the transducers, the thickness of the material, and the orientation of the crystal structure. Any mass deposited on the surface of the sensor causes a change in the resonant frequency of the device, which can be measured electronically. The first of these devices developed became known as the quartz crystal microbalance (QCM); however, several other types of acoustic mass sensors have been described including SAW, flexural plate wave (FPW), and acoustic plate mode (APM) devices. These sensors are similar to the semiconductor-based electrochemical devices in that they can very easily be microfabricated, resulting in small sensors with low power consumption.

Optical chemical sensors are based on the interaction of electromagnetic radiation with an analyte, resulting in some alteration of the properties of the radiation. This class of devices includes optical fibers and waveguides, which can be used to transport a spectrometric source of light to the sample interface. Changes in optical properties related to index of refraction, absorbance, fluorescence,

or chemiluminescence can then be evaluated using an appropriate type of photodetector.[1] Optical sensors offer many advantages, including elimination of the signal "noise" associated with the transduction mechanisms of sensor systems based on electrical phenomena.

1.2 Performance Characteristics

Given the diverse nature of the transduction mechanisms upon which different chemical sensors systems are based, it may seem at first to be very difficult to compare performance. However, since the fundamental application of these devices is similar, that is, the identification and quantification of toxic gases and vapors, a set of performance criteria can be developed that allows comparisons between different sensor systems to be made. Table 1 lists a number of possible quantitative performance criteria, or figures of merit, that can be used to evaluate instrumentation for measuring toxic gases and vapors.[9] Each of these terms will be defined below.

Accuracy or bias provides a measure of the systematic error of a sensor and can be defined in absolute or relative terms. The absolute systematic error is simply equal to the difference between the average sensor result and the true value. For example, if the average reading, $\bar{x}$, for a sensor used to measure carbon monoxide is 12 parts per million (ppm) when the true concentration is 10 ppm the absolute systematic bias would simply be 12–10 ppm, or 2 ppm. However, since it is generally more useful to know what the bias is relative to the magnitude of the measurement, relative systematic bias is frequently reported and is defined as follows, Equation (1):

$$\text{relative systematic bias (\%)} = \left(\frac{\bar{x} - x_{\text{true}}}{x_{\text{true}}} \right) \times 100 \quad (1)$$

For the previous example, relative systematic bias = $[(12 - 10)/10]100 = 20\%$. In practice, every effort is made to identify and eliminate sources of systematic bias through the use of blanks and instrument calibration.

Table 1 Sensor performance characteristics

Characteristic	Measurement
Accuracy or bias	Absolute systematic error, relative systematic error
Precision	Standard deviation, RSD
Sensitivity	Calibration sensitivity, analytical sensitivity
Detection limit	Average blank plus three times the standard deviation
Working concentration range	LOQ to LOL
Selectivity	Coefficient of selectivity, correct classification rate

RSD, relative standard deviation; LOQ, limit of quantification; LOL, limit of linearity.

Precision describes the reproducibility of a measurement and is most often represented by the standard deviation or coefficient of variation (CV) calculated for a series of measurements recorded under the same conditions. The sample standard deviation, s, is a statistical measure defined as, Equation (2):

$$s = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N - 1}} \quad (2)$$

The RSD is defined as $s/\bar{x}$, and the CV is simply the RSD expressed as a percentage, or $100\% \times \text{RSD}$. These measures of precision provide an indication of the random error of a measurement and can be used to define confidence intervals or to estimate levels of confidence associated with a given measurement.

Sensitivity is a measure of an instrument's ability to detect small differences in analyte concentration. Sensitivity is the manifestation of two important sensor characteristics: the slope of the calibration curve and the precision or reproducibility of the measurement.[9] For two sensors having equal precision, the one with the steepest calibration curve would be the most sensitive. The simplest definition of sensitivity is *calibration sensitivity*, which for linear response is independent of concentration and equal to the slope of the calibration curve (m). However, the calibration sensitivity fails to account for the reproducibility of the measurement and so an alternative measure of sensitivity has been proposed. The *analytical sensitivity* is defined as the slope of the calibration curve (m) divided by the standard deviation (s) of the signal measurement. This approach offers the advantage of being relatively insensitive to amplification factors since ordinarily m and s would be affected proportionally, and also yields an estimate of sensitivity that is independent of the measurement units employed. One disadvantage of analytical sensitivity as a figure of merit results from the fact that s often varies with concentration, making the resulting estimate concentration-dependent.

The limit of detection (LOD) is defined as the minimum concentration of an analyte that can be detected with some known level of confidence.[9,10] This limit therefore depends on the ratio of the magnitude of the signal to the size of the statistical fluctuations, or "noise" in the signal for sample media that do not contain any analyte (blank). Once the analytical signal exceeds the blank response by more than some multiple, k, times the noise, an analyte is "detected". A value of three is most often employed for k, meaning that the minimum detectable response (S_m) is defined as the mean blank signal ($\bar{S}_0$) plus three times the standard deviation of the blank signal (s_0) or,

Equation (3):

$$S_m = \bar{S}_0 + 3s_0 \tag{3}$$

The LOD can then be calculated as, Equation (4):

$$\text{LOD} = \frac{S_m - \bar{S}_0}{m}, \tag{4}$$

or simply Equation (5):

$$\text{LOD} = \frac{3s_0}{m} \tag{5}$$

where m is the slope of the calibration curve. The LOQ is defined similarly, except that a value of 10 is typically used for k, Equation (6):

$$\text{LOQ} = \frac{10s_0}{m} \tag{6}$$

This has the effect of reducing the influence of random error on the signal measurement, and therefore yields a better estimate of the true value. Thus, the LOQ is a lower concentration limit, above which it is expected that the relative influence of random error or noise on a measurement will be below some acceptable level determined by the value of k employed.

The working concentration range or operating range for a sensor system is defined as the concentration range bounded at the low end by the LOQ, and at the upper end by the LOL, above which, sensor response becomes nonlinear (Figure 1).

Operation in the range below the LOQ results in larger errors due to the proportionately larger influence of noise on the signal, while operation at concentrations above the LOL complicates sensor calibration and interpretation of results. It is therefore desirable that a sensor system have as large a working concentration range as possible in order to avoid the need either to preconcentrate samples that would be below the LOQ or to dilute samples that fall above the LOL. In general, a working range of at least two orders of magnitude is needed for a sensor system to be useful.[9]

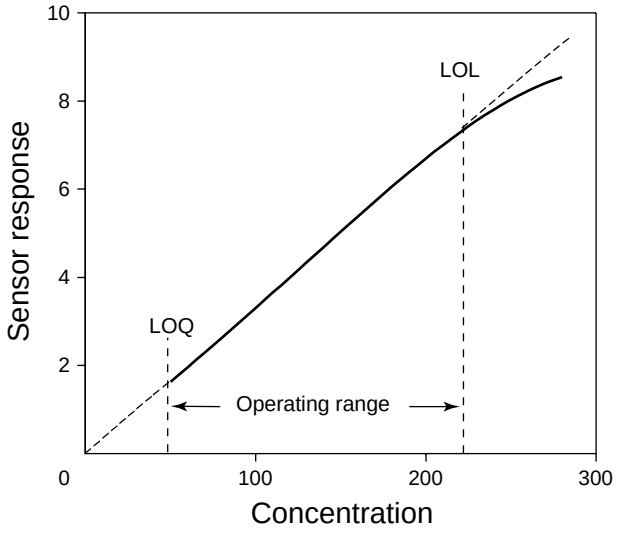

Figure 1 Sensor operating range defined by LOQ and LOL. —— actual, - - - - - ideal.

The selectivity of a sensor refers to the extent to which it responds without interference from other chemical species also present in the sample. No sensor is completely unaffected by the presence of other species, particularly those that are chemically similar to the analyte of interest. For this reason it is important that potential interferents be characterized and considered when selecting a sensor system for a given application. One way to evaluate selectivity for a sensor is by evaluating the relative response of the sensor for the analyte of interest and any interferents likely to be present. The relative response (S_r) is defined as the sensitivity of the sensor for the analyte of interest (m_a) divided by the sensitivity for the interferent (m_i), Equation (7):

$$S_r = \frac{m_a}{m_i} \tag{7}$$

A large value of S_r is indicative of better selectivity. For example, if a sensor system is to be used to measure hydrogen sulfide (analyte) in the presence of carbon monoxide (interferent) and S_r is reported to be 10^4, this means that the response of the sensor to hydrogen sulfide is 10 000 times that of carbon monoxide. Thus it would require 10 000 ppm of carbon monoxide to produce a response equivalent to 1 ppm hydrogen sulfide. In this case, it is unlikely that the presence of carbon monoxide would significantly interfere with the measurement of hydrogen sulfide and the sensor could be considered selective. However, for many analyte-interferent pairs, S_r may be much smaller or even less than one, in which case the sensor is not selective and an alternative method of analysis may be required.

In addition to these quantitative criteria, there are a number of other important characteristics that are used to evaluate sensor system capabilities for a given application (Table 2). One of the most important of these is response time, t_r, which is frequently defined as the time required for the sensor response to reach some fraction (typically 90%) of the final value. Response time is an important consideration for compounds that are extremely toxic, since it would obviously be important to detect these as quickly as possible. In this case a response time on the order of seconds would be a requirement, while for less critical applications such as process monitoring, a slower response time on the order of minutes may be satisfactory. Another important consideration for selecting sensor

Table 2 Qualitative sensor system performance characteristics

Response time	Time required to reach 90% of final value
Ease of use	Packaging and data handling features, calibration requirements
Operator skill	Level of training required to operate and obtain useful results
Costs	Cost of equipment, cost of consumables, cost of calibration supplies

systems is ease of use. Instruments vary widely in how they are packaged and in what types of feature are offered, including how data is displayed and stored, and how the instrument is calibrated for use. A related characteristic is the skill required to operate the instrument – the utility of a sensor system is decreased significantly if highly skilled operators are needed in order to obtain useful results. Finally, cost is always an important factor in selecting a sensor system, and again instruments can vary widely in the costs associated with operation. These costs can include periodic replacement or factory recalibration of the sensing element itself, consumables such as pre-filters or other sample handling accessories, and the gases required for calibration and quality control.

1.3 Defining Toxic Gases and Vapors

Toxic gases and vapors are by definition compounds that adversely affect living organisms.[11] However, Paracelsus noted in the 1500s that *all* substances are poisons and that only the dose differentiates a poison from a remedy, making the task of defining toxic substances somewhat more complex. Despite Paracelsus' enlightened view, a smaller but growing subset of readily recognized toxic compounds has developed with time, often as the result of some unfortunate encounter, or more recently, based upon epidemiological study or through the use of improved laboratory techniques for assessing the toxicity of various substances. This combined pool of human experience has lead to the development of a number of lists of known toxic chemicals, and in some cases, recommended exposure levels below which it is expected that the incidence of adverse health effects will be very low. For example, the American Conference of Governmental Industrial Hygienists (ACGIH) publishes an annual listing of threshold limit values (TLVs) which refer to airborne concentrations of chemical substances to be used as guidelines in the control of health hazards in the workplace.[12]

In addition to voluntary consensus standards such as the ACGIH TLVs, governmental agencies are frequently charged with developing legal standards applicable to toxic gases and vapors, which typically requires first developing a list of regulated substances. As part of legislation known as the Clean Air Act (CAA) of 1970 and CAA Amendments of 1990, the United States Environmental Protection Agency (USEPA), was charged with developing a list of hazardous air pollutants (HAPs) that are to be regulated. This list, which now consists of 188 toxic air pollutants, includes numerous gases and vapors such as benzene, hydrogen sulfide, formaldehyde, and vinyl chloride. The Occupational Safety and Health Administration (OSHA) is another US governmental agency that promulgates standards for a number of toxic gases and vapors in the form of both substance-specific standards which outline the legal concentration limits (permissible exposure limits or PELs) and practices and procedures for working with various chemicals, and a set of tables (known as the Z-Tables) which outline the OSHA PELs for numerous chemicals.

In reviewing the lists of chemicals considered to be toxic by different government agencies, the shortcomings of this approach to regulating human exposure becomes apparent. Organizations may focus on different aspects or characteristics of a chemical's toxicity according to the mission of that particular agency. For example, OSHA regulations are specifically designed to protect workers from exposures to chemicals arising out of their employment, and therefore focus on chemicals that historically have been of some importance in industry. The USEPA on the other hand, considers toxicity from a comprehensive environmental perspective and attempts to evaluate all mechanisms of exposure. These different approaches can yield lists of regulated toxic chemicals that are not consistent from agency to agency. This result reflects the compromises that must be made in developing finite lists of toxic compounds for the purpose of regulating exposures according to the goals and philosophy of a particular organization or government agency.

An important ramification of this regulatory approach is that efforts to control exposures are likely to be the greatest for chemicals that have been defined as toxic materials by regulatory agencies, despite the fact that there may be other equally toxic chemicals that have not yet been added to a list. Since regulatory processes typically take place in a political environment, decisions to include or exclude chemicals are not always made in a timely fashion, and neither are they always based entirely on sound toxicological data. Additional factors such as the nature of the expected adverse health effects, the number of people potentially exposed, and the feasibility and economics of controlling exposures may also be considered and can dictate whether a particular substance becomes regulated. Inclusion on a regulated list of substances can also have a significant impact on efforts to develop instrumentation for measuring a

particular toxic substance. If a large number of industries suddenly become legally and economically responsible for complying with a new regulation for a particular chemical, the demand for instrumentation to measure that chemical can increase sharply in a very short time. This logical chain of events demonstrates the interrelationship between regulatory processes and the development of sensor systems and re-emphasizes the implications of the necessarily subjective process of defining toxic chemicals.

Chemicals that have through some means been identified as toxic can be further classified in a number of ways, depending on the application of interest. For example, toxic chemicals can be grouped by their target organ (liver, kidney, reproductive), effects (CNS (central nervous system) depression, liver injury, cancer), or physical state (gas, liquid, solid).[11] Another characteristic of toxic chemicals that has relevance both in evaluating exposures and in developing regulatory limits, is the temporal relationship between exposure and toxic response. Acute effects are those that occur almost immediately and can be reversible or irreversible. Chronic effects on the other hand are those that have a very long latency period and may require 20 to 30 years before being observed. Exposure limits for acute toxins are usually based on a relatively short sampling period, for example, the ACGIH STELs are defined as a 15-min time weighted average (TWA) air concentration that should not be exceeded at any time during the workday. Ceiling limits, which are defined as concentrations that should never be exceeded during any part of a workday, are another type of limit developed specifically for controlling exposures to chemicals that exert acute health effects. While STELs and ceiling limits are generally employed to protect from *reversible* acute health effects, concentrations defined as immediately dangerous to life or health (IDLH) have also been developed for many toxic gases and vapors and are designed to protect from *irreversible* acute health effects. IDLH concentrations refer to conditions that: (1) pose an immediate or delayed threat to life or (2) would cause irreversible adverse health effects, or (3) interfere with an individual's ability to escape unaided from a hazardous situation.

Sensor systems are particularly well suited to evaluating exposures to toxic chemicals that exert acute health effects since they generally can provide near-real time results. Traditional analytical techniques, which involve collecting an integrated sample over some period of time, followed by laboratory analysis, do not provide the immediate feedback needed to make decisions regarding exposures to acute toxins. In addition, modern sensor-based instrumentation typically provides many options for processing and storing measurements, thereby allowing extreme flexibility in evaluating exposures relative to ceiling limits, STELs, or longer term average concentrations based on user selectable time periods. It is this extreme range of options for processing data that makes sensors such an attractive alternative for measuring toxic gases and vapors.

2 SENSOR TYPES

In this section a more detailed description of the most common types of sensor will be presented. Devices will be generally categorized as electrochemical, mass, or optical sensors, and the operating principles and important characteristics of each type will be described.

2.1 Electrochemical Sensors

The vast majority of chemical sensors currently available for detecting toxic gases and vapors are based upon electrochemical principles. Instruments employing these devices are available in a range of sizes, including those designed for both personal and fixed-point sampling, and there are hundreds of makes and models available. Electrochemical sensors include those that employ electrochemical cells, as well as solid-state devices based on semi-conducting properties of the sensing elements. These sensors can be subcategorized according to the type of electrical quantity measured, that is, if voltage is the measure of response, then the device would be referred to as a potentiometric process. Likewise, sensor systems that use current or resistance as the measure of response would be categorized as amperometric, or conductometric processes, respectively.

The most common type of chemical sensor is the electrochemical cell, which consists of two or more electrodes, an electrolyte, and an electrical system used to measure or control voltage or current. One of the electrodes is an indicating electrode which is designed to interact selectively with the chemical of interest, while the second electrode serves as a reference by providing a relatively constant potential during the measurement. In a three-electrode design (Figure 2), the working or sensing electrode usually consists of a catalytic metal such as platinum. Analytes diffuse through a porous barrier into the electrolyte and, depending on the voltage at the sensing electrode, may be oxidized or reduced. The electrons involved in the oxidation/reduction process will flow through the cell and constitute the output of the sensor. A reference electrode that is usually not exposed to the analyte provides a stable electrochemical potential in the electrolyte for comparison with the sensing electrode. The counter electrode is provided to complete the circuit, and is adjusted so that a current that is exactly equal but opposite to the working electrode will result. The concentration of the analyte is related to the

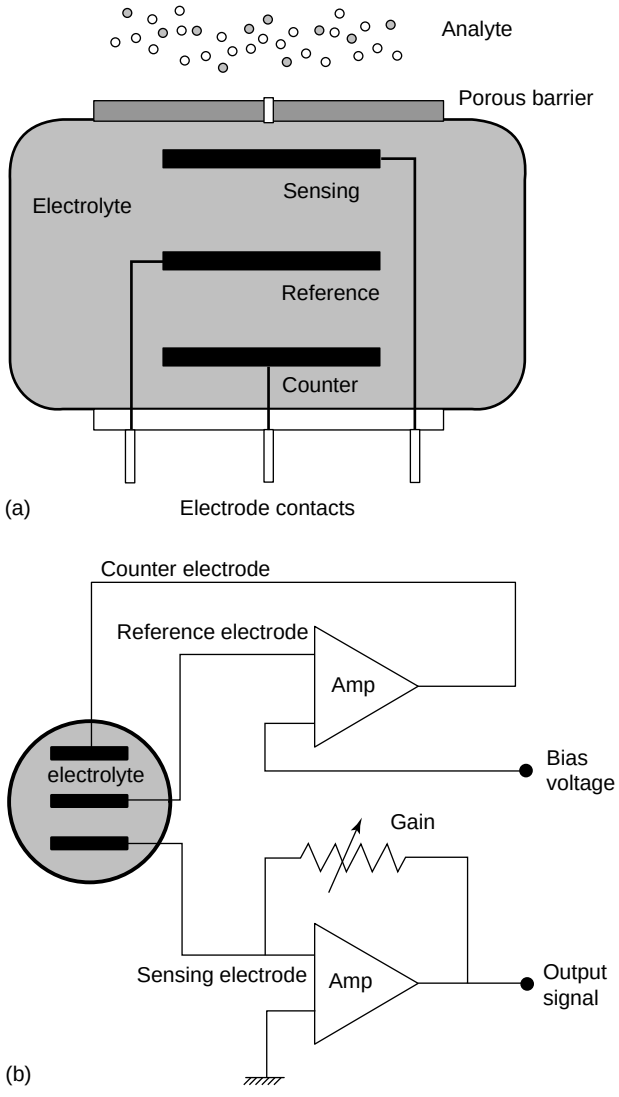

Figure 2 Simplified schematic for (a) an electrochemical-cell based sensor and (b) typical electronics circuitry.

voltage or current depending on how the electronics of the sensing system are designed.

Sensors based upon electrochemical cells are generally limited to use for detecting reactive inorganic chemicals; however, this class encompasses a broad range of contaminants that are frequently encountered in industrial workplace settings. Table 3 lists a number of toxic contaminants for which electrochemical cell-based sensor systems are available.

The specificity of electrochemical cell-type sensors can be quite good for a wide range of toxic gases and vapors. Selectivity is generally best for amperometric cells as a result of the ability to vary the voltage at the sensor electrodes, thereby controlling the extent to which analytes are polarized. This approach would more correctly be referred to as voltammetry, which is defined as the measurement of current in a cell as a function of the applied voltage. Through careful selection of the electrode system, catalysts, electrolyte, the voltage applied to the electrodes, and the polymer membrane through which the analyte diffuses, a considerable degree of selectivity can be obtained.[10] These types of sensor are generally stable, can have long service lives, require very little power, and for certain toxic gases may have LODs on the order of 0.1 ppm. The greatest limitation of electrochemical cell-type sensors is the effect of interfering gases on measurements. Despite the previously cited techniques for maximizing specificity, it is not possible to obtain complete selectivity and interfering compounds may cause positive or negative cross-sensitivities depending on the electrochemical nature of the analyte of interest.

Table 3 Toxic gases for which electrochemical sensors are available[10,13,14]

Ammonia (NH_3)	Ethylene oxide (C_2H_4O)	Nitric oxide (NO)
Arsenic pentafluoride (AsF_5)	Fluorine (F_2)	Nitrogen dioxide (NO_2)
Arsine (AsH_3)	Formaldehyde (CH_2O)	Ozone (O_3)
Boron trichloride (BCl_3)	Hydrazine (N_2H_4)	Phosgene ($COCl_2$)
Carbon dioxide (CO_2)	Hydrogen chloride (HCl)	Phosphine (PH_3)
Carbon monoxide (CO)	Hydrogen cyanide (HCN)	Silane (SiH_4)
Chlorine (Cl_2)	Hydrogen fluoride (HF)	Silicon tetrachloride ($SiCl_4$)
Diborane (B_2H_6)	Hydrogen sulfide (H_2S)	Sulfur dioxide (SO_2)

Instrumentation employing electrochemical cell-type sensors is widely available and has been continually improved over the years as a result of developments in microprocessor technology and increasing capabilities for miniaturization. Most manufacturers offer small lightweight battery-operated versions of their instruments that can be equipped with multiple sensors, thus allowing a single instrument to monitor several toxic gases simultaneously (Figure 3). This approach allows users to switch out sensor modules and affords a great deal of flexibility for addressing different monitoring scenarios. One of the most common applications of this type of instrument is confined space monitoring for which a typical instrument set-up might include hydrogen sulfide, carbon monoxide, combustible gas, and oxygen sensors.

Several types of electrochemical sensor are based on conductometric processes, which involve changes in resistance of a sensing element in relation to the concentration of an analyte. One of the most common devices of this type is known as a catalytic sensor, or pellistor, and is frequently employed in instrumentation

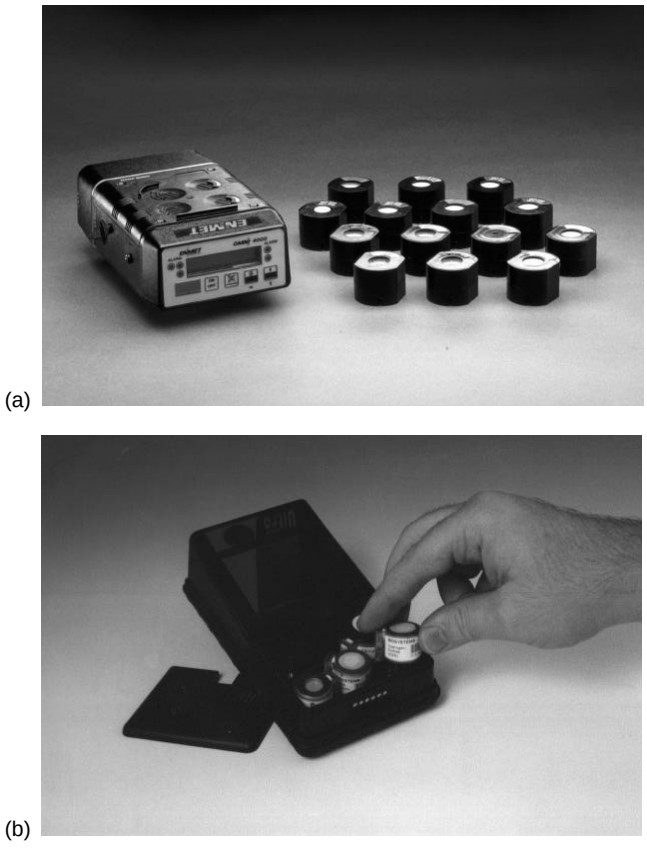

Figure 3 Electrochemical-cell based sensor systems with interchangeable sensors (a) Enmet; (b) Biosystems. (Reproduced by kind permission of Enmet Corp. and Biosystems Inc. respectively.)

designed to measure concentrations of combustible gases and vapors.[8,15] The principle of operation for the catalytic sensor is relatively simple. A wire coated with a catalyst is heated to an elevated temperature, which causes combustible gases to be oxidized. The resulting generation of heat changes the resistance of the wire, which is configured as part of a Wheatstone bridge circuit in order to measure the response (Figure 4). In these simple devices, the wire coil is used both to heat the sensor and as the resistive sensing element. The wire typically consists of a platinum coil embedded in a pellet of a catalytic material such as ThO_2/Al_2O_2, which is also coated with a porous catalytic metal, for example, palladium or platinum. The catalyst reduces the temperature at which combustion takes place, which serves both to minimize energy requirements and to reduce the rate at which the platinum coil deteriorates as a result of evaporation.[15] Although playing an important role in the process of detecting an analyte, the catalyst is susceptible to ''poisoning'' as a result of the deposition of solid decomposition products on the surface of the sensor or from irreversible adsorption of

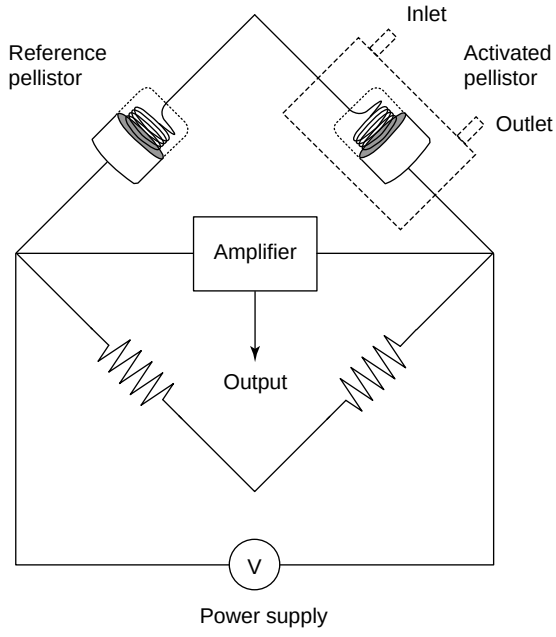

Figure 4 Typical catalytic sensor (pellistor) configuration in Wheatstone bridge circuit.

certain materials. Silicones, tetraethyl lead, and organic phosphorus compounds are frequently cited as materials that can poison catalytic sensors.

Although catalytic sensors have found wide use in combustible gas indicator (CGI) instruments, their use for measuring toxic gases is limited as a result of two factors. First, since the sensor will respond to any chemical that is oxidized, selectivity is poor. This is actually a necessary characteristic for the CGI since it would be desirable to detect the presence of any combustible gases. The second factor limiting catalytic sensor use for toxic gas detection is poor sensitivity. Although these devices work well for detecting combustible gases in the range of the lower explosion limits (LELs) and upper explosion limits (UELs) for many chemicals, such concentrations are generally several orders of magnitude greater than applicable occupational exposure limits (OELs) which are based on toxicological effects rather than the danger of explosion.

Metal oxide semiconductor (MOS) sensors, or solid-state sensors, are another type of conductometric sensor used for measuring toxic gases and vapors, or as a general survey instrument. In its simplest form, a MOS sensor consists of a semiconducting material such as tin oxide (SnO_2) on a sintered alumina bead (Figure 5). A heating coil is used to control the temperature of the sensor and a set of leads or electrodes are connected across the bead. Upon exposure to an analyte, the conductivity of the bead will change as the chemical is adsorbed due to the redistribution of electrons in the metal oxide material.[15]

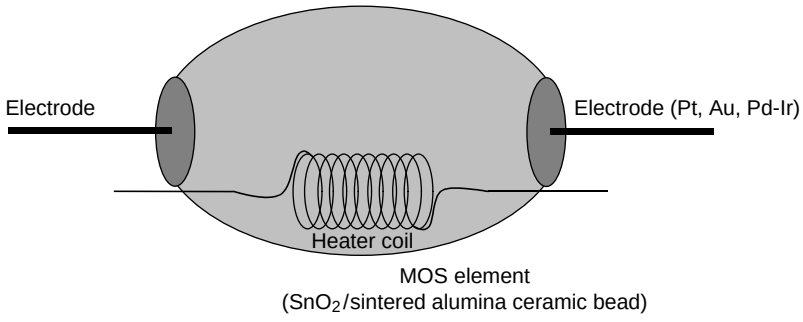

Figure 5 MOS sensor.

The relationship between MOS conductivity changes and temperature is complex and can be characteristic for a given chemical. This phenomenon has led to the development of arrays of MOS sensors operated at different temperatures and/or using different types of semiconducting materials in an effort to achieve greater selectivity than a single sensor could provide.[16–19] The characteristic response patterns can be determined for a number of chemicals of interest and then used to identify and quantify unknown samples using statistical pattern recognition techniques or neural networks. MOS sensors are particularly well suited for miniaturization, given their inherent use of semiconducting materials, and are easily incorporated into integrated circuits.[20] This means that fairly large arrays of sensors can be fabricated on a single chip, resulting in a very small size and minimal power requirements. Despite these promising developments, commercially available versions of MOS sensors are generally limited to the simplest form first described, and are most often used as a broad-range screening tool for detecting the widest possible range of toxic and combustible gases.

Sensors based on conducting polymers represent a relatively new approach involving a conductometric process for measuring toxic gases and vapors.[21,22] These sensors are based on the use of conducting polymers or polymer composites that are placed between two electrodes. The electrical resistance of the polymer is then monitored and used as the measure of response. When the sensor is exposed to a chemical, the resistance across the polymer changes due to the absorption of the analyte in the polymer. Responses are generally rapid and reversible and power requirements are small, making this sensor an attractive approach to measuring toxic gases and vapors. The polymers used can consist of electrochemically prepared or ''grown'' conducting polymers such as polypyrrole, polyaniline, and polythiophene. Alternatively, conducting polymer composites have been fabricated by mixing carbon black with various insulating polymers such as poly(vinyl acetate), poly(*N*-vinylpyrrolidone), or poly(styrene).[23] The resulting sensors are only partially selective – responses are based on the solubility properties of each analyte–polymer pair – so arrays of conductive polymer sensors are typically employed and have been referred to as an ''electronic nose''. Sensitivity for polymer-based sensors can vary widely as a result of differences in the extent to which an analyte will partition into the polymer layer. In cases where sensitivity is a problem, preconcentration techniques such as the use of granular adsorbents and thermal desorption can be used.

2.2 Mass Sensors

Detection and measurement of extremely small quantities of mass cannot be performed with conventional analytical balances, which typically have maximum sensitivities on the order of a tenth of a microgram, and employ mechanisms that do not translate well to use in portable instrumentation. However, microsensors based on acoustic wave properties have been shown to be capable of detecting mass changes in the range of nanograms.[6,8] This extreme sensitivity has led to extensive research examining the capabilities of many different types of acoustic wave sensors for a variety of applications including the measurement of toxic gases and vapors.[7,24–27]

The use of acoustic waves for sensing requires a device with a means for generating the waves, circuitry for operating the device, and methods for measuring the wave characteristics of interest, which may include changes in velocity, frequency, or amplitude. A typical device consists of a piezoelectric material such as quartz, with a set of metal transducers mounted on the surface(s). The transducers launch and receive acoustic waves into the quartz substrate at frequencies ranging from one to several hundred megahertz. Changes in surface mass caused by the presence of an analyte cause the properties of the acoustic wave to be altered, resulting in measurable changes in velocity, frequency, or amplitude, that can be used as the measure of response. Different types of acoustic sensors are categorized based on the mode of operation, which is determined by the orientation of the

acoustic wave in relation to the piezoelectric substrate and the metal transducers. The most common acoustic sensor categories are thickness shear mode (TSM), which includes devices often referred to as QCMs, SAW, and FPW.[7] Of these three types of acoustic sensor, the SAW device has received the most attention as a sensor for toxic gases and vapors owing to its excellent mass sensitivity, small size, and low-power requirements.[7,24,26,28,29]

SAW device technology resulted from an outgrowth of the development of IDTs in the 1960s.[24] The ability to fabricate IDTs on the surface of a piezoelectric material made it possible to generate a number of different types of acoustic wave, including the surface Rayleigh waves upon which SAW devices are based. The first application of SAW devices was primarily as radiofrequency (rf) filtering and processing circuit elements. However, it was soon discovered that these devices were extremely susceptible to environmental effects, which led to an examination of their use as chemical sensors.

SAW devices are produced using integrated circuit microfabrication techniques that result in a small package and low power requirements, making it possible to fabricate multiple sensors on a single sensor package (Figure 7). The mechanism of operation involves the generation of rf mechanical waves on the surface of a piezoelectric substrate such as single-crystal quartz.[7,28,29] The waves are launched and received by parallel sets of thin metal electrodes (IDTs) deposited on the surface of the substrate. Application of a sinusoidal voltage to alternately connected IDT electrodes causes a periodic electric field to be imposed on the crystal substrate. This electric field gives rise to a periodic strain field due to the piezoelectric effect resulting in the production of a standing SAW.[7,26,28,29] The wavelength and the frequency of the SAW are determined by the spacing of the IDT electrodes and the velocity of the wave in the substrate. Typical SAW device operating frequencies range from 30 to 400 MHz.[7]

SAW devices can be used as chemical sensors provided that two conditions are met. First, the target analyte must be trapped on or near the surface of the substrate. This is usually accomplished through a chemical reaction or through sorption with a partially selective coating.[29] And second, the presence of the analyte must produce a change in some property of the coating (e.g. modulus, mass, density) that results in a change in the amplitude or velocity of the SAW. SAW chemical sensors are usually employed in a dual-device configuration (Figure 6) in which one sensor acts as a reference device and a second coated or ''working'' sensor is exposed to the analyte of interest. This approach involves using the difference frequency of the two sensors as the measure of response, which has two benefits. First, it provides some compensation for environmental effects. For example, if a change in temperature occurs, both the reference and working sensors would be affected in a similar manner and the resulting difference frequency should be less affected than the working sensor alone. In addition, the difference frequencies, which are generally on the order of hertz to kilohertz, are much lower than the frequency of operation for the individual SAW devices (MHz) and can therefore be measured with less sophisticated electronic circuitry than would otherwise be possible. The dual-device approach can be applied to arrays of SAW chemical sensors by using a multiplex process in which a single reference sensor is used to produce a difference frequency for multiple working sensors.

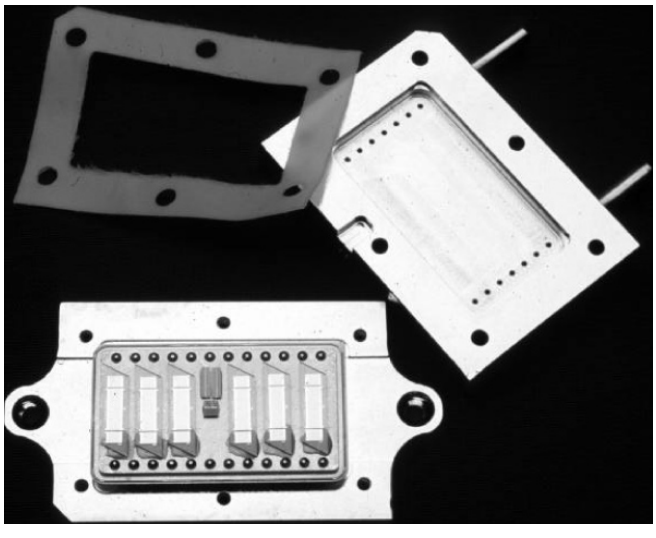

Figure 6 Fabrication of multiple SAW sensors on a single header/flow-cell. (Reproduced by kind permission of Sawtek Inc.)

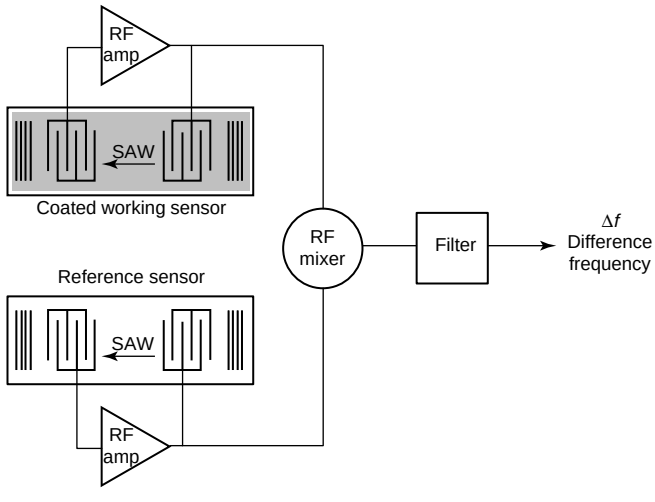

Figure 7 Typical dual-device configuration for polymer-coated SAW chemical sensor.

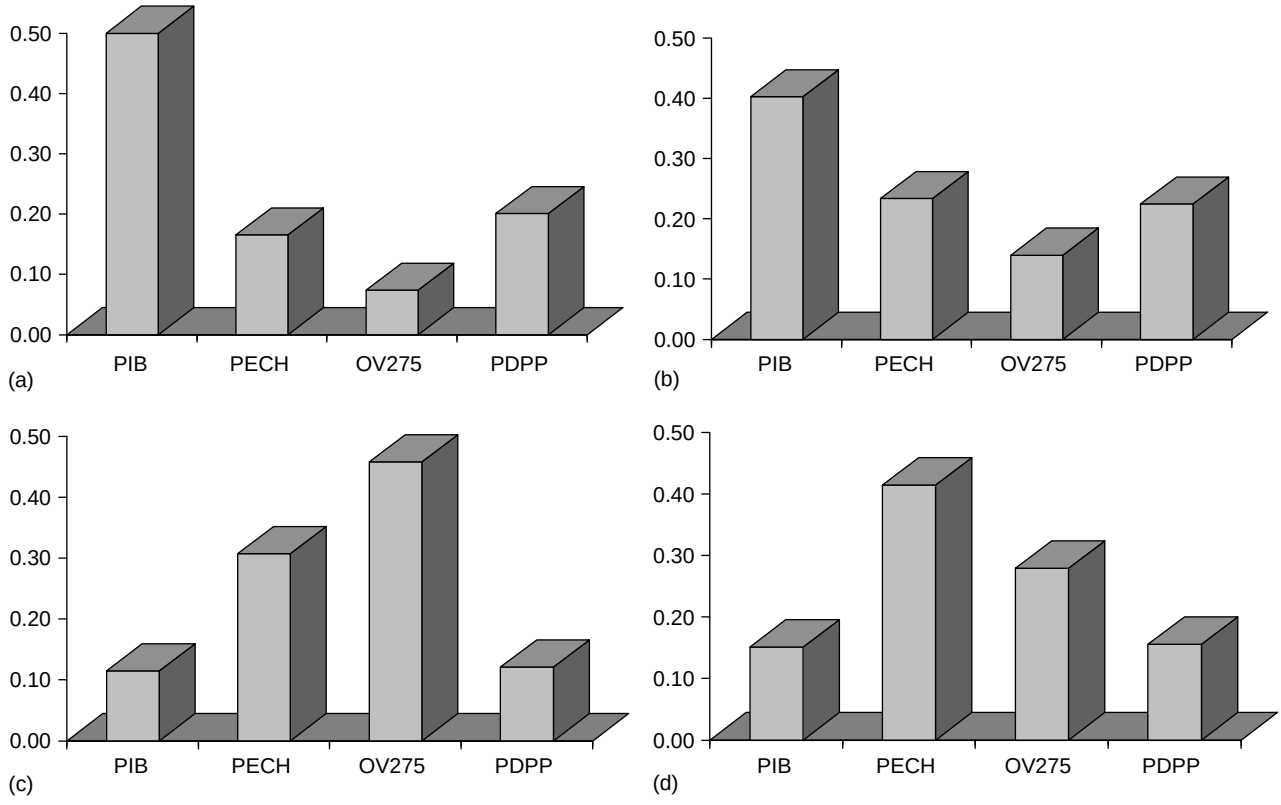

Figure 8 Chemical sensor array response pattern or ''fingerprint'' for several different vapors. (a) Perchloroethylene, (b) trichloroethylene, (c) 2-propanol, (d) 1,4-dioxane. Polymer coatings: PIB, polyisobutylene; PECH, polyepichlorhydrin; OV275, bis-cyanoallyl polysiloxane; PDPP, poly(diphenoxyphosphazine).

The use of a multisensor array comprising several SAW sensors coated with different sorptive polymers provides a means of achieving a high degree of selectivity while retaining the reversibility of a sorption process.[28,30,31] With this approach, each sensor has only a partial degree of selectivity, but a specific vapor will produce a characteristic response pattern or ''fingerprint'' (Figure 8) that can be interpreted using pattern recognition techniques. Classification criteria can be established using methods such as the k-nearest neighbor (KNN), soft independent modeling of class analogy (SIMCA), or extended disjoint principal component regression (EDPCR).[32–35] The selectivity of the polymer-coated sensor array can be optimized by selecting coatings that engage in different solubility interactions (e.g. dispersion, dipole–dipole, hydrogen bonding) thereby maximizing the amount of information provided by a given number of sensors.[24,34,36]

The sensitivity of polymer-coated SAW sensors is determined by the mass sensitivity of the bare device as well as the interaction between the polymer and analyte of interest. In cases where polymer–vapor interactions are strong, sensor responses will be relatively large and it may be possible to measure the analyte directly and compare results to relevant standards such as OSHA PELs. However, for many applications, lower LODs are needed, making it necessary to enhance sensor system performance through the use of a preconcentration stage such as a thermally desorbable granular adsorbent tube.[29,30,37] Using this approach, excellent sensitivity has been demonstrated for a variety of instruments and applications employing SAW sensor systems for measuring toxic gases and vapors.

Most research involving SAW sensors has focused on application of these devices to the detection and measurement of organic vapors, since many existing technologies, such as electrochemical-cell based sensors, are best suited to the measurement of reactive inorganic gases and vapors, rather than organic vapors. However, SAW sensor systems can generally be used to measure any gas or vapor, provided that an appropriate coating can be identified. The diversity of SAW sensor systems is demonstrated by the wide range of applications reported in the literature, including the measurement of carbon tetrachloride in process streams,[38] carbon dioxide and relative humidity,[39] nitrogen dioxide,[40] 1,3-butadiene and styrene,[27] ammonia,[41] numerous organic vapors,[29,31,37,42,43] and chemical agents such as

Figure 9 SAW sensor system for measuring organophosphorus and organosulfur nerve and blister agents. (Reproduced by kind permission of Microsensor Systems, Inc.)

organophosphorus and organosulfur nerve and blister agents (Figure 9).[30]

2.3 Optical Sensors

Optical sensors represent a subclass of chemical sensors in which electromagnetic radiation is directed to and from a sensing region that is exposed to the analyte of interest.[1,8,44] Interaction with the sensing media causes some measurable change in the properties of the radiation that can be related to the concentration of the analyte. A light source typically consisting of light emitting diodes (LEDs), tungsten–halogen lamps, or lasers, provides the incident radiation that is directed to the sensing region. The sensing media can consist of a variety of different types of material designed to enhance selectivity, including various reagents, polymers, and biologically active materials. Technological developments in the telecommunications field have led to the availability of many different types of optical device that can be used to direct and interface electromagnetic radiation in optical sensor systems, including optical fibers, hollow cylinders, and planar structures.[1,44,45] These devices are fabricated such that there is a difference in refractive index between the central core and outer cladding, resulting in almost complete internal reflection of optical energy. This allows a very efficient and low loss transmission of optical signals. Optical sensors offer a number of advantages relative to electrical sensors, including immunity to electrical interference, intrinsic safety, and the capability of transmitting signals over relatively large distances.[46]

Although a wide variety of configurations have been reported, there are generally two approaches to

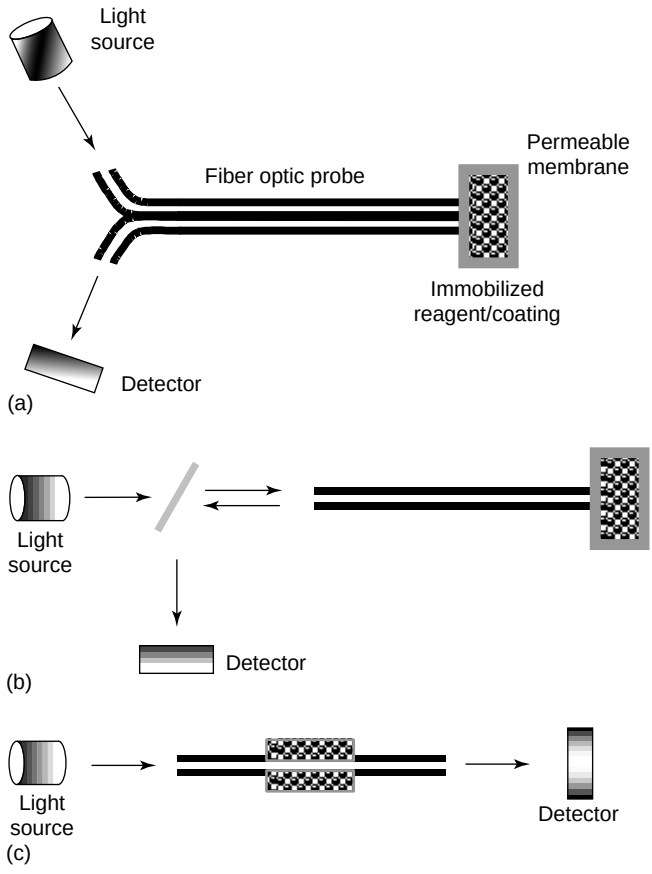

Figure 10 Common configurations for optical sensors.

employing optical sensors. In the first, the electromagnetic radiation is directed to the sensing region, which is typically located at the distal end of an optical fiber. The radiation interacts with the analyte and is returned to an appropriate detector through another section of optical fiber (Figure 10a). Alternatively, the same fiber can be used to transmit and receive the radiation (Figure 10b). Neither of these approaches requires direct contact between the sensing/reagent phase and the optical fiber, making it relatively easy to replace or change the sensing media in a sensor system.[46] The second common configuration employed in optical sensor systems involves removal of the cladding from a portion of the optical structure (fiber, hollow tube, planar substrate) and replacing it by a reagent or polymer coating (Figure 10c). This approach is based on the fact that when light is reflected in an optical structure, a portion of the radiation penetrates a small distance into the less dense reflecting medium. This penetrating radiation is known as the evanescent wave, which, if absorbed by the less dense medium, results in an attenuation of the beam at specific absorption band wavelengths.[9,44,46,47] Changes in the internally reflected light beam can be evaluated using a variety of optical

measurement techniques, including chemiluminescence, fluorescence, colorimetry, and absorption.

The light source and detector employed in an optical sensor system are related to the application and the optical parameter of interest. If signals need to be transmitted over large distances, then a high energy source such as an argon ion laser is usually required. However, laser-based sensor systems suffer from a number of limitations, including large power requirements, higher costs, possible photodecomposition of reagent/coatings, and the limited number of excitation wavelengths available.[46] LEDs are the simplest and least expensive light source, but generally provide low power and are also only available for a limited range of wavelengths (550–1800 nm).[44] Tungsten–halogen light sources can be used to provide higher intensities than LEDs over a broader range of wavelengths (340–2500 nm) thus providing an alternative to LED and laser sources. The detectors used in optical sensor systems are usually either a solid state diode or a photomultiplier tube. In many cases the incoming radiation may be modified by selecting specific wavelengths for detection. Solid state diodes are attractive for sensor systems, owing to their low costs and simplicity, but may not be sensitive enough for some applications. In these instances, photomultiplier tubes and photon counting can be used to enhance and quantify the signal.[44]

The reagent or coating used in the sensing region of an optical sensor is selected to interact with the analyte such that some measurable optical property is changed. These materials can generally be categorized by whether the interaction is reversible or irreversible. Reversible processes are preferred for most sensor applications for a number of reasons, including, at least in theory, an unlimited lifetime. The most common types of reversible analyte–reagent phase interactions employed in optical sensors are (1) concentration or preferential partitioning of the analyte into an immobilized phase on the sensor, (2) reversible reactions between analyte and the immobilized reagent phase, and (3) displacement of a ligand in the immobilized reagent phase by the analyte.[46] The first approach involves the use of partially selective coatings such as polymers into which the analyte partitions, based on the solubility properties of the polymer–analyte pair. Arrays of sensors can easily be prepared by applying different polymers, containing an appropriate dye, to the distal end of an optical fiber. The resulting bundle of fibers can then be employed as a sensor system by monitoring changes in an appropriate optical property, such as fluorescence, resulting from the partitioning of the analyte into the polymer–dye matrix.[45,48,49] Classification techniques such as neural networks are then used to interpret the resulting response patterns and identify and quantify analytes.[50,51] The development of reagent coatings that involve reversible reactions or displacement of a reagent-bound ligand offers the potential for highly selective optical sensors and has been the focus of much research.[52] However, these approaches require the availability of appropriate and thoroughly characterized reaction mechanisms/pathways for every analyte, and in many cases suffer from slow response times and reagent instabilities.[44,46]

The amount of information that can be conveyed by a single optical sensor is much greater than that of most other types of chemical sensors, owing to the ability to examine multiple wavelengths simultaneously. Although these capabilities have yet to be fully realized, numerous applications of optical sensors to the detection of toxic gases and vapors have been described, including the measurement of organic vapors,[48,50,51] ammonia[47] (Figure 11), nitric acid,[52] carbon dioxide, and hydrogen cyanide.[44] The inherent advantages of optical sensing systems are likely to make their use more prevalent in commercially available instrumentation in the future.

2.4 Biosensors

Biosensors are analytical devices that employ a biologically sensitive material to detect chemical species directly without the need for complex sample processing.[53] Although not a separate class of sensor from the standpoint of the transduction mechanisms, the unique nature of the approach taken to sensing analytes with these devices justifies the use of a separate category. Biosensors are typically made by attaching a biologically sensitive material such as an enzyme to an appropriate transducer which converts the biochemical response into a quantifiable electronic signal. The goal of this approach

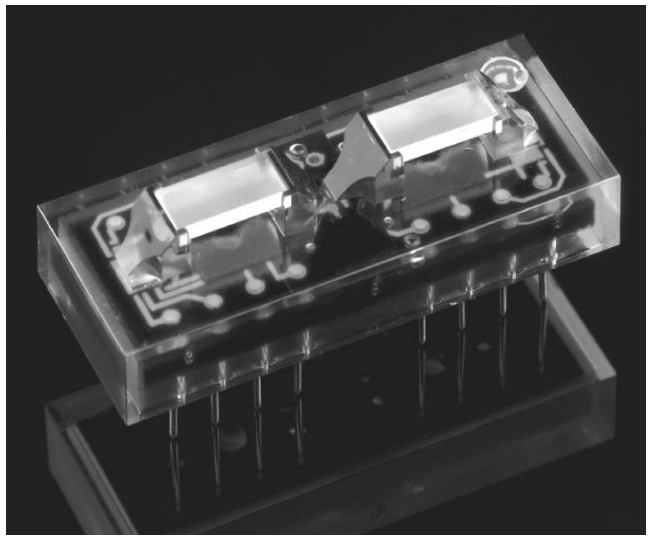

Figure 11 Planar waveguide optical ammonia sensor package. (Reproduced by kind permission of Industrial Scientific Corp.)

Table 4 Biosensor components[53–56]

Biologically sensitive materials	Transducer/mechanism
Enzymes	Electrochemical
Glucose	Potentiometric
oxidase	Amperometric
Antibodies	Conductometric
Immunoglobulins	Mass
Receptors	Acoustic wave –
Insulin	QCM, SAW, FPW
Adrenaline	Optical
Taste, smell	Absorbance
Whole cells	Luminescence
Microbial	Fluorescence

is to take advantage of the specific nature of many normal biological processes, thereby providing highly selective detection of the desired chemical species.[53–55] Table 4 lists some of the more common biologically sensitive materials and transducer mechanisms incorporated in biosensor systems.

The majority of biosensors described have been based on the use of enzymes, antibodies, receptors, or whole cells as the selective media.[53,54,56] Enzymes are proteins that catalyze specific chemical reactions in the body. They are found in all types of cell and although sharing a common polypeptide building block, specific enzymes vary considerably in their actual structure and catalytic role. Enzymes generally exhibit a greater degree of specificity compared to chemical catalysts. Antibodies are serum proteins produced by B lymphocytes and plasma cells in response to a foreign substance. In most cases the antibody will respond to only one substance – the antigen. In biosensors based on antibodies, the analyte can be either the antibody itself, or the corresponding antigen or a piece of the antigen (hapten). Even minor changes in the structure of an antigen can make it invisible to the antibody produced in response to the original antigen structure, thus the reactions are highly specific.[54] Molecular receptors are cellular proteins that bind specific chemicals, resulting in a conformational change in the protein structure that elicits a cellular response.[54] This ''triggering'' effect results in a biological amplification of a signal, and, when combined with the high degree of specificity, make receptors an attractive approach to biosensing. Biological cells are membrane-bound structures containing a variety of chemicals including enzymes, nucleic acids, and proteins. Higher-order organisms contain billions of cells, whereas the simplest living structures such as algae and bacteria consist of a single living cell. Even these simple cells can demonstrate physicochemical or biological changes in response to specific chemicals and are therefore potentially useful as biosensors.

The reliability of a biosensor depends on a number of factors, including how well the biological components can be incorporated into the sensor/transducer. Methods for fixing or immobilizing the biosensitive material include adsorption, gel entrapment, covalent coupling, and cross-linking.[53,56] Adsorption typically involves bringing an aqueous solution of the biosensitive material into contact with a suitable solid material such as silica gel, clay, or ion exchange resin, and then washing to remove residual material. Entrapment in polymeric gels allows small analyte molecules to penetrate the sensing region without loss of the sensing material. Covalent coupling involves reactions between nonessential parts of the biological material and a chemically activated surface such as glass, synthetic polymer, or cellulose. Cross-linking agents induce intermolecular binding or cross-linking to attach biological materials to solid supports or other structures. For each of these approaches the challenges are similar – to prepare a biologically active sensor system in a way that does not alter or inhibit the original nature of the biological process.

Biosensors based on each of the main sensor/transducer platforms, electrochemical, mass, and optical, have been developed and continue to be explored at a rapid pace.[53,56] Much of this work focuses on identifying appropriate biologically sensitive reagents/coatings for specific applications, and exploring immobilization techniques specific to different types of transducers. Current biosensor applications for measurement of toxic gases and vapors include both reactive inorganic and organic compounds such as pesticides, alcohols, ammonia, formaldehyde, BTEX (benzene, toluene, ethylbenzene, and xylene), and cyanide compounds.[57,58] Although biosensors potentially offer greater specificity than other types of chemical sensor, there are considerable challenges remaining to be addressed before their use in measuring toxic gases and vapors can become more widespread. The majority of these challenges are related to the biological materials employed, including issues such as identifying and synthesizing selective materials for each application/analyte, overcoming stability issues and the limited shelf lives that are typical for most biological media, addressing slow response times, and developing commercially viable sensor packages and instrumentation that can be used in the field. Given the many advantages of biosensors, it is likely that significant resources will continue to be focused on the remaining challenges so that the potential of these devices can eventually be realized.

3 APPLICATIONS

Given the wide variety of sensor systems available for measuring toxic gases and vapors, selection of an

appropriate type is typically based on the requirements of a specific application. Applications vary considerably with regard to the objectives for measurement and the specific characteristics of each scenario, including the type of analyte(s) to be measured, the nature of the processes involved, and the physical characteristics of the sampling environment. Despite these differences, it is possible and potentially useful to categorize toxic gas sensing applications generally by the purpose of the measurement. This approach can simplify the process of selecting an appropriate sensing system for applications that require similar sensor characteristics or levels of performance. In the following section, application categories consisting of industrial hygiene monitoring, confined space monitoring, process emission monitoring, and leak detection will be considered and a discussion of the specific needs of each will be presented. Although certainly not the only categories that could be created, these are likely to comprise the majority of sensing applications relevant to the detection and measurement of toxic gases and vapors.

3.1 Industrial Hygiene Monitoring

Industrial hygiene is generally defined as the discipline devoted to the anticipation, recognition, evaluation, and control of worker exposure to chemical, physical, and biological hazards in the workplace. Using this definition, the measurement of toxic gases and vapors would fall under the area of evaluation of chemical hazards, and in fact, many industrial hygienists spend a significant portion of their time carrying out this activity. The focus on controlling hazards that may affect the *employees in the workplace* serves to differentiate this activity from others involving the measurement of toxic gases and vapors for reasons related to safety or environmental protection.

In evaluating worker exposure to specific chemical hazards, industrial hygienists refer to OELs. These OELs may consist of legally defined concentration limits such as the United States OSHA PELs, the United Kingdom maximum exposure limits (MELs) and occupational exposure standards (OESs), or the German Maximale Arbeitsplatz Konzentrationen (MAK) values. Alternatively, consensus or recommended standards such as the ACGIH TLVs may be referred to in cases where a legal standard either does not exist, or is less protective compared to other available standards. Most OELs are similar in that they involve measurement of the concentration of contaminants in the breathing zone of the worker, which has obvious implications for the size and weight of the equipment used to perform the measurement. OELs are generally developed to address either acute or chronic health effects of toxic gases and vapors and so are based on different sampling/averaging times. Evaluation of exposure to chemicals for the purpose of controlling acute health effects is generally based on comparison to STELs, which typically involves averaging the airborne concentrations over time periods of 15–30 min. Ceiling limits are also often employed for acute hazards and represent concentrations that are never to be exceeded during the workday. For chemicals with chronic health effects, concentrations are generally averaged over the entire work shift and reported as an eight-hour TWA.

The nature of industrial hygiene applications presents many challenges for measuring toxic gases and vapors. This results from the need to measure a wide variety of analytes selectively in the breathing zone of workers at relatively low concentrations for average times ranging from a few minutes to several hours. The implications of these requirements are significant: (1) instrumentation must be lightweight and battery powered so that it can be worn by the worker, thus allowing measurements to be made in the breathing zone, (2) the sensor(s) employed must be capable of selectively analyzing the chemical of interest in the presence of other interferences, (3) response times must be fast enough to allow useful comparisons between measured concentrations and ceiling limits, and (4) data processing and display capabilities should be such that both short-term and long-term averaging can be performed, as well as continuous comparison of measured concentrations to applicable ceiling limits. In addition to these fundamental requirements, a number of practical considerations are also important, including costs, ease of use, and calibration and maintenance requirements.

Despite these challenges, instrument manufacturers have made significant progress towards addressing the needs of industrial hygiene applications. Dozens of manufacturers offer sensor-based instruments designed to perform personal monitoring for a number of toxic gases and vapors, including carbon monoxide, hydrogen sulfide, hydrogen cyanide, ammonia, chlorine, and nitrous oxide compounds. The vast majority of these instruments are based on electrochemical sensors, and in particular, electrochemical cell sensors, and so are generally limited to use for reactive inorganic compounds. However, research focusing on development of acoustic/mass sensors, optical sensors, and electrochemical sensors for industrial hygiene applications should eventually lead to the availability of instrumentation applicable to a wider variety of analytes, including both organic and inorganic compounds.

3.2 Confined Space Monitoring

Confined space monitoring is one of the most important applications of sensors for measuring toxic gases

and vapors. This process generally involves the use of instrumentation to evaluate the atmosphere in an enclosed space prior to the entry of workers and during the course of any activities, to ensure that contaminants are below acceptable levels. Confined spaces are defined as any space that (1) is large enough or configured such that an employee could enter and perform work, (2) has limited or restricted means for entry and exit, and (3) is not designed for continuous occupancy.[59] Examples of confined spaces include process vessels, tanks, and sewers, as well as large open pits, and excavations. In recognition of the hazards and potential for serious injury and loss of life, there are usually legally enforceable standards applicable to work that is conducted in confined spaces.

Procedures for entering confined spaces consist of a number of steps for ensuring worker safety, including identification of hazards in the following categories: (1) oxygen deficient atmospheres, (2) flammable atmospheres, (3) toxic atmospheres, and (4) mechanical and physical hazards. Sensor-based instrumentation is invaluable in evaluating hazards in the first three of these four categories. Verification of adequate oxygen is obviously important to sustain life but is also the first measurement made when evaluating a confined space for entry, since sensors used to evaluate flammable or toxic atmospheres may not function properly in oxygen-deficient environments. Atmospheres containing less than 19.5% oxygen are generally considered to be oxygen deficient. The level of oxygen in a confined space may be decreased by a number of mechanisms, including displacement by other gases, consumption by chemical or biological reactions, combustion processes, or as a result of normal respiration by the workers. Flammable atmospheres can result from the work being conducted in a confined space, including operations such as spray painting, coating, or the use of flammable solvents. Welding and cutting operations that involve the use of oxyacetylene equipment may also produce explosive atmospheres as a result of leaks. Toxic atmospheres may be present in a confined space as a result of the material stored in the space, decomposition of organic materials, the work being conducted, or the movement of toxic gases from adjacent areas into the confined space.

Instruments designed specifically for confined space monitoring are offered by many manufacturers and come in numerous configurations, with various features. However, these instruments generally have several common characteristics due to the requirements of the application. Most are relatively small, battery-operated devices, equipped with multiple chemical specific sensors. Sensor configurations vary but generally include an oxygen sensor, and some type of flammable/combustible gas indicating sensor. The capability to add additional sensors for specific toxic gases is offered by most manufacturers, and allows users to customize the instrument for their specific application. Carbon monoxide and hydrogen sulfide are two acutely toxic contaminants that occur widely in confined spaces. For this reason, the ''stock'' sensor configuration for a confined space monitoring application very often consists of sensors for detection of oxygen, combustible gases, carbon monoxide, and hydrogen sulfide. Electrochemical-cell type sensors are most often employed for the measurement of oxygen, carbon monoxide, and hydrogen sulfide, while a simple pellistor or MOS-type sensor may be used to evaluate combustible gases. Confined space monitoring equipment is usually equipped with alarms in order to warn workers immediately when the atmosphere becomes hazardous, based on user-definable criteria or set points. Equipment may rely on passive diffusion or active sampling with pumps in order to bring the contaminants into contact with sensors. The latter configuration allows for the use of longer probes or sampling lines, which can be lowered into a confined space to evaluate the atmosphere prior to workers entering. Once a space has been cleared for entry, smaller personal type sensor systems are often worn by workers to monitor continuously for hazardous atmospheres.

3.3 Process Emission Monitoring

The increased attention paid to the release of toxic gases and vapors into the environment has led to the development of numerous regulatory requirements for monitoring process emissions.[60] In many cases these regulations can be satisfied through the use of continuous emission monitoring systems (CEMSs), which generally involve measuring the concentration of specific contaminants in an exhaust stack.[61,62] These systems can be either in situ, which involves the use of a fixed sensor inside the exhaust stream, or extractive, in which a probe is used to extract a sample which is then analyzed using a remote sensor system. In situ systems are best suited for applications requiring measurement of a small number of toxic gases or vapors simultaneously, since standard sensor packages are available and equipment and installation costs are relatively low. Extractive systems are most often employed when a larger number of analytes are to be measured, or if the conditions inside the exhaust stream are so harsh that the use of in situ equipment is not feasible. In these cases, a remote sensor system allows for the use of a larger more sophisticated instrument, without exposure to the conditions of the process stream. Extractive systems can be more expensive than in situ ones as a result of the additional sampling lines required and the higher equipment and installation costs, but may be the only alternative for applications requiring

continuous monitoring of multiple analytes under harsh conditions.

The types of sensor used for process emission monitoring are similar to those used in other applications involving measurement of toxic gases and vapors, the main difference being the packaging and configuration of the complete systems. Systems can vary greatly in complexity from single point, single analyte systems, to multipoint systems for the measurement of multiple contaminants. Electrochemical sensors are probably the most common type of device employed for process emission monitoring, however the use of optical and piezoelectric sensors is increasing. Although generally not considered to be sensor systems, traditional benchtop analytical methods such as gas chromatography, and infrared spectrometry are also used with extractive systems for process emission monitoring. The packaging and configuration of sensor systems used for process emission monitoring depends on the requirements of the specific application. For in situ monitoring of a single analyte, a single sensor mounted in the process stream would be adequate. The output from the sensor could be displayed directly at the location or could be routed to a central control location for processing and display. Extractive systems are more likely to involve the use of multiple sensors and locations, and therefore would require more sophisticated systems for transmitting and monitoring sensor outputs.

3.4 Leak Detection

Detection of leaks or ''fugitive emissions'' is another important application of toxic gas and vapor sensor systems. Environmental regulations may require that leak detection be performed on a regular basis for processes involving toxic gases and vapors.[63] Leaks may also represent a significant cost to operations in the form of lost product or raw materials, which is why many organizations perform leak testing on a routine basis. In addition, sudden leaks of acutely toxic materials can pose an immediate threat to life so operations involving these types of material may have fixed leak detection systems in place to protect from this hazard. Potential sources of leaks include valves, pumps, flange joints, fittings, pressure relief devices, vents, access doors, seals, pumps, and compressors.

There are generally two approaches to monitoring for leaks of toxic gases and vapors. The first involves mounting fixed sensors in areas where leaks are likely to occur or where the consequence of a leak is severe (see Figure 12). An example of this scenario would be the use of a fixed sensor mounted in a semiconductor processing area to detect any release of toxic gases such as arsine (AsH_3). One problem associated with this approach is how to

(a)

(b)

Figure 12 Fixed-point (a) area monitor with visual alarm (reproduced by kind permission of Biosystems Inc.) and (b) sensor transmitter in industrial housing. (Reproduced by kind permission of Industrial Scientific Corp.)

determine the number of sensors required for a given area and where to locate the sensors, since these decisions have implications for the time it takes for a sensor to respond to a potentially life threatening leak. Alternatively, sensors could be placed at every probable leak site, for example fittings, flanges, and seals, although this method may not be feasible for operations where the number of potential leak sites is very large. The second approach to leak detection involves the use of portable instrumentation for periodic or routine examination of all potential leak sites. This method is appropriate when the consequence of a leak is less severe and when it is acceptable for a leak to go undetected for some period of time. The requirements for sensor systems employed in either of these approaches to leak testing are similar. In both cases it is desirable that response times be relatively fast (30 s or less) so that (1) in the case of life threatening releases of toxic gases, adequate warning is provided, and (2) when using portable instrumentation to screen for leaks, the operator can complete multiple measurements more quickly. Selectivity may not be important for many leak testing applications, since there often is only a single gas or vapor to be measured. However, in cases where multiple contaminants are potentially present and one or more are extremely toxic, sensor selectivity would be a paramount consideration. For applications where selectivity is not crucial, MOS and pellistor-type sensors may be employed, provided that sensitivity is adequate for the toxic gas of interest. When selectivity is important, more specific sensor systems, such as those employing electrochemical cells, coated piezoelectric, or optical sensors, could be used.

4 QUALITY CONTROL AND MAINTENANCE

Quality control is a fundamental component of most activities in business and industry, but is particularly important in the measurement of toxic gases and vapors, given the serious nature of the consequences that could result from unreliable data. For this reason, sound procedures for calibrating and maintaining sensor systems are crucial to ensure that the data obtained using these devices is of acceptable quality and that the lives and health of those affected are ultimately protected. In this section, the requirements of different toxic gas and vapor sensing applications will be considered from the standpoint of data quality, and typical procedures for the calibration and maintenance of sensor systems will be described.

4.1 Data Quality Objectives

The quality of the data obtained by different sensor systems for measuring toxic gases and vapors is likely to be variable, owing to a number of factors. Various sensing platforms have different strengths and weaknesses and are therefore likely to be better suited to certain applications. Alternatively, when employed in different manufacturer's instruments, the same type of sensor may produce output of varying quality as a result of design differences. Another factor affecting the quality of data obtained using sensor systems is the fact that these devices are often configured in personal or portable monitoring systems, which are then employed in harsh environments. Under these conditions, it is inevitable that some sacrifices are made in terms of data quality, particularly when compared to analytical laboratory-based instrumentation. Further, even the most rigorously designed sampling and analytical methods for toxic gases and vapors have measurable levels of error and uncertainty in the resulting data, and so it is reasonable to expect these in sensor systems as well. Given these constraints, the challenge becomes one of matching the data quality objectives of a specific application to the capabilities of available toxic gas and vapor sensor systems. In the following paragraphs, the requirements of several toxic gas and vapor measurement applications will be considered and used to illustrate differences in the resulting data quality objectives for sensor systems.

4.1.1 Industrial Hygiene Monitoring

As previously described, industrial hygiene monitoring frequently involves the measurement of toxic gas and vapor concentrations in order to evaluate worker exposures by comparing results to legal limits such as the OSHA PELs. Validated analytical methods that have met specific requirements for accuracy and precision are required for this compliance monitoring. OSHA stipulated accuracy requirements for the measurement of toxic gases and vapors, are generally ±35% for concentrations at or below 0.5 times the PEL and ±25% for concentrations at or above the PEL. The National Institute for Occupational Safety and Health (NIOSH) also develops and validates methods specifically for workplace exposure assessment. NIOSH has established the criterion that on average, over a concentration range of 0.2–2 times the OEL, a method should provide a result that is within ±25% of the true value 95% of the time.[64] Validation requires a determination of the bias and precision of a method, which are then used together to calculate a confidence interval for the estimate of overall accuracy. Although the use of sensor-based direct reading instrumentation for OSHA compliance monitoring is not prohibited, the vast majority of validated methods involve collection of an integrated air sample onto sample media over some period of time, with subsequent laboratory analysis. This is probably because OELs for toxic gases

and vapors are typically in the low parts per million concentration range and thus challenge the sensitivity of sensor systems. However, there is increasing recognition of the capabilities of sensor-based direct reading instruments, particularly for evaluating OELs based on short averaging times such as ceiling limits and STELs. OSHA has validated a sensor-based instrument for the measurement of carbon monoxide and several OSHA standards, including those for ethylene oxide (29 CFR 1910.1047) and formaldehyde (29 CFR 1910.1048) specifically cite direct-reading instruments as an alternative method of analysis, thus paving the way for increasing use of sensor systems for industrial hygiene monitoring.

4.1.2 Confined Space Monitoring

Like industrial hygiene monitoring, confined space monitoring is usually conducted in order to comply with governmental regulations and so the data quality objectives are generally defined by applicable regulations. However, referring to the OSHA Permit-required Confined Spaces Standard (29 CFR 1910.146),[59] there is little specific information given with regard to the performance of instrumentation used for detecting hazardous atmospheres. The standard simply states that before an employee enters a permit-required confined space, the internal atmosphere shall be tested with a calibrated direct-reading instrument for (1) oxygen content: 19.5–23.5%, (2) flammable gases and vapors: concentrations must be less than 0.1 times the LEL, and (3) potential toxic air contaminants: concentrations must be below the applicable PEL and/or IDLH. Without OSHA, specified levels of accuracy for the measurement of toxic gases and vapors, it seems reasonable that the general requirements for industrial hygiene compliance monitoring would be appropriate (i.e. ±35%) for concentrations at or below 0.5 times the PEL and ±25% for concentrations at or above the PEL.

In the case of measuring flammable gases and vapors, performance standards have been developed for sensor systems. The primary standard for these devices is American National Standards Institute/Instrument Society of America (ANSI/ISA)-S12.13.1-1986, which specifies that the instrument must indicate the true test concentration within ±3% of the full scale reading or ±10% of the applied concentration, whichever is greater. The standard also has performance specifications related to temperature and humidity effects, response times, and effects of air velocity, vibration, electromagnetic interference, and long-term storage. These relatively rigorous requirements are necessary due to the potentially severe consequence of failing to detect a hazardous environment, but are also more easily achieved given the relatively high concentrations to be measured – typical LEL concentrations are of the order of 1–3%, meaning that 0.1 times the LEL would be 1000–3000 ppm.

4.1.3 Process Emission Monitoring

In most instances where process emission monitoring is required by environmental regulations, data quality objectives are clearly defined and strict procedures for ensuring the attainment of these objectives are enforced. In the United States of America, the CAA Amendments, Resource Conservation and Recovery Act (RCRA), and state environmental programs are intended to reduce air pollution and provide for an accurate characterization of emissions of specific toxic substances, including many gases and vapors. The CAA Amendments apply to a variety of industries, and air emission measurements are a key part of most facility's compliance strategies. Depending on the nature of an industry and the quantity of specific pollutants emitted, USEPA regulations may require either a CEMS or the use of manual source testing methods which involve collection of source effluent samples using filters, or other sampling media and subsequent laboratory analysis. An examination of the data quality objectives for the numerous USEPA-required measurements of toxic gases and vapors is beyond the scope of this section. However, these specifications are available in the Federal Register, and in many cases are online, so that those affected are able to gather reliable data according to USEPA regulations. The individual subparts of the USEPA rules specify the reference methods used to evaluate the accuracy and precision of CEMS and include detailed performance specifications outlining all aspects of sampling and analysis systems, including procedures and specifications for calibration and quality assurance/quality control (QA/QC).

4.2 Calibration

All sensor systems require periodic calibration in order to ensure that the resulting measurements of toxic gases and vapors meet the required data quality objectives for a specific application. This process generally involves the use of calibration or ''span'' gases for which the true concentration is known to some degree of certainty. The span gas is sampled and the instrument is adjusted using a span control so that the output is equal to the known span gas concentration. Span gases are readily available for many commonly encountered toxic gases at a variety of concentrations; however, the gases are consumed in the calibration process and may represent a considerable expense if the instrument is used on a regular basis. In recognition of the expense and potential complexity of calibration, instrument manufacturers have gone to great

lengths to simplify the process. Unfortunately, calibration requirements have become a part of the marketing process for instrumentation, which may result in some manufacturers making exaggerated claims regarding the acceptable amount of time between calibrations for their particular system. Although data quality objectives vary by application, for the purpose of measuring toxic gases and vapors where the consequences of unreliable data are significant, sensor system accuracy should be verified on a daily basis.

Most sensor systems are designed to be calibrated using a single span gas concentration. Calibration then involves zeroing the instrument either by sampling in an environment that is known to be contaminant-free, or using a cylinder containing ''zero'' air. The output of the instrument is adjusted to read zero and then the calibration gas is sampled and the span control is used to adjust the output to read the known concentration. This process which is usually referred to as a one-point calibration, actually defines the calibration curve for the instrument using two points, zero and the span gas concentration. A potential problem with this approach is that unless the instrument response is absolutely linear, some error is introduced when concentrations away from the calibration concentration are measured (Figure 13). Manufacturers try to minimize deviations from linear response through sensor design and electronic processing of sensor outputs; however, the problem still exists to varying degrees. For this reason, the selection of the span gas concentration is important and it is usually desirable to use a concentration close to the exposure limit of interest in order to maximize the accuracy of the instrument.

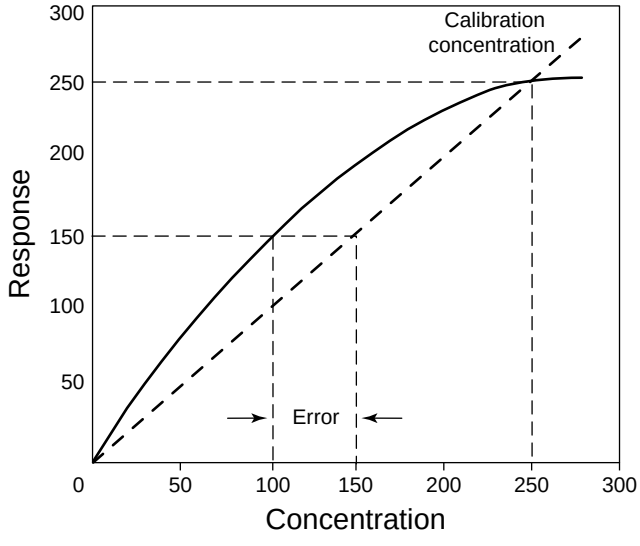

Figure 13 Potential error introduced by using one point calibration and assuming linear response. ———, actual response; - - - -, linear response.

Figure 14 Instrument with calibration kit (reproduced by kind permission of Biosystems).

Calibration kits for combustible and toxic gases are commercially available and are usually provided when an instrument is initially purchased.[65] These kits often consist of one or more cylinders of known concentrations of specific air–gas mixtures, a pressure regulator and valve, and a hose adapter for connecting the cylinders to the instrument (Figure 14). Instruments can also be calibrated using static test-atmospheres generated in either rigid containers (e.g. glass, stainless steel) or in flexible bags (Tedlar®, Mylar®) by adding a known quantity of the chemical of interest to the known volume of air in the container. Alternatively, dynamic test-atmosphere generation systems can be used to prepare known concentrations of toxic gases and vapors; however, these approaches become increasing complex, and may not be cost effective or efficient for many users.

4.3 Maintenance

Maintenance is an important part of quality control for toxic gas and vapor sensor systems. Although the requirements are generally not complicated, the consequences of equipment failure can be serious or even fatal, so a proactive approach to maintenance is a crucial part of any measurement process. Maintenance tasks can be categorized by the individual components of the instrument, including items such as batteries or power supplies, sample tubing and pneumatics, filters, flashback arresters, sensors, and electronics.[66]

4.3.1 Batteries

Portable instruments usually operate on battery power and are designed to provide at least 8 h of service. However, many rechargeable batteries have a memory

effect, and also eventually fail, so it is important that the batteries have been cycled and fully charged to ensure optimum performance.

4.3.2 Pneumatics

Many instruments contain tubing and various valves or fittings to direct air flow. Tubing can become worn or cracked as a result of exposure to harsh environments, and may eventually leak, so it is important to examine these systems both visually and functionally to determine if they are in working order. Any pumps and valves should also be examined for blockage or signs of corrosion or contamination by particulates, to ensure proper operation. Filters are often used to prevent dust and gross particulates from entering a sensor system, but may become clogged. This can cause reduced airflow through the instrument, which may result in decreased accuracy or detrimental effects on instrument response time. Therefore, filter replacement should be a routine part of instrument maintenance.

4.3.3 Intrinsic Safety Components

Sensor systems for measuring toxic gases and vapors are often used in hazardous environments that require an intrinsically safe design. For this reason, it is vitally important that maintenance activities are consistent with those recommended by the manufacturer, so that the instrument remains intrinsically safe. Although different manufacturers' sensor systems may appear to be similar, users should never interchange parts, since this can eliminate the intrinsically safe rating of an instrument both from a legal and practical standpoint. Flashback arresters, if required, should be examined to ensure that they are present and in good condition.

4.3.4 Sensors

The design and function of sensors varies significantly, so the manufacturer's recommendations for maintenance should be referred to for specific guidance. In general, instruments are often used in atmospheres containing corrosive environments, under conditions of temperature and humidity extremes that could damage sensors. Although the most obvious signs of these effects is likely to be erratic readings or complete failure, sensors should be examined periodically for signs of corrosion, condensation, leakage, and to evaluate the general physical condition of the sensor assembly.

4.3.5 Electronics

Erratic readings can indicate dirty or faulty potentiometers or corroded electrical connections. Oxidation of contacts can also occur if an instrument is not used for extended periods. Contacts can be cleaned using physical means such as a piece of fine abrasive; however, in general, electrical maintenance should be handled by the instrument manufacturer.

ABBREVIATIONS AND ACRONYMS

ACGIH	American Conference of Governmental Industrial Hygienists
ANSI/ISA	American National Standards Institute/Instrument Society of America
APM	Acoustic Plate Mode
BTEX	Benzene, Toluene, Ethylbenzene, and Xylene
CAA	Clean Air Act
CEMS	Continuous Emission Monitoring System
CGI	Combustible Gas Indicator
CHEMFET	Chemically Sensitive Field-effect Transistor
CNS	Central Nervous System
CV	Coefficient of Variation
EDPCR	Extended Disjoint Principal Component Regression
FPW	Flexural Plate Wave
HAP	Hazardous Air Pollutant
IDLH	Immediately Dangerous to Life or Health
IDT	Interdigital Transducer
ISFET	Ion Selective Field-effect Transistor
KNN	k-nearest Neighbor
LED	Light Emitting Diode
LEL	Lower Explosion Limit
LOD	Limit of Detection
LOL	Limit of Linearity
LOQ	Limit of Quantification
MAK	Maximale Arbeitsplatz Konzentrationen
MEL	Maximum Exposure Limit
MOS	Metal Oxide Semiconductor
NIOSH	National Institute for Occupational Safety and Health
OEL	Occupational Exposure Limit
OES	Occupational Exposure Standard
OSHA	Occupational Safety and Health Administration
OV275	bis-Cyanoallyl Polysiloxane
PDPP	Poly(diphenoxyphosphazine)
PECH	Polyepichlorhydrin
PEL	Permissible Exposure Limit
PIB	Polyisobutylene
QA/QC	Quality Assurance/Quality Control
QCM	Quartz Crystal Microbalance
RCRA	Resource Conservation and Recovery Act

rf	Radiofrequency
RSD	Relative Standard Deviation
SAW	Surface Acoustic Wave
SIMCA	Soft Independent Modeling of Class Analogy
STEL	Short-term Exposure Limit
TLV	Threshold Limit Value
TSM	Thickness Shear Mode
TWA	Time Weighted Average
UEL	Upper Explosion Limit
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

REFERENCES

1. R.W. Murray, 'Chemical Sensors and Microinstrumentation: An Overview', in *Chemical Sensors and Microinstrumentation: ACS Symposium Series #403*, eds. R.W. Murray, R.E. Dessey, W.R. Heineman, J. Janata, W.R. Seitz, American Chemical Society, Washington, DC, 1–19, 1989.
2. Committee on New Sensor Technologies, 'Materials and Applications; Expanding the Vision of Sensor Materials', National Academy Press, Washington, DC, 9–18, 1995.
3. J. Fraden, *Handbook of Modern Sensors*, American Institute of Physics, Woodbury, NY, 505–510, 1997.
4. P. Hauptmann, *Sensors: Principles and Applications*, Prentice Hall, Englewood Cliffs, NJ, 4–7, 1993.
5. W.R. Heineman, 'Introduction: Electrochemical Sensors', in *Chemical Sensors and Microinstrumentation: ACS Symposium Series #403*, eds. R.W. Murray, R.E. Dessey, W.R. Heineman, J. Janata, W.R. Seitz, American Chemical Society, Washington, DC, 22–25, 1989.
6. D.S. Ballantine, R.M. White, S.J. Martin, A.J. Rico, E.T. Zellers, G.C. Frye, H. Wohltjen, *Acoustic Wave Sensors: Theory, Design, and Physico-Chemical Applications*, Academic Press, San Diego, CA, 3–4, 1997.
7. J.W. Grate, S.J. Martin, R.M. White, 'Acoustic Wave Microsensors: Part I', *Anal. Chem.*, **65**, 940–948A (1993).
8. J. Fraden, *Handbook of Modern Sensors*, American Institute of Physics, Woodbury, NY, 1–9, 1997.
9. D.A. Skoog, J.L. Leary, 'Principles of Instrumental Analysis, 4th edition, Saunders College Publishing, Fort Worth, TX, 5–9, 1992.
10. J.L. Perkins, 'Modern Industrial Hygiene: Recognition and Evaluation of Chemical Agents', Van Nostrand Reinhold, New York, 604, 1997.
11. C.D. Klaasen, D.L. Eaton, 'Principles of Toxicology', in *Casarett and Doull's Toxicology – The Basic Science of Poisons*, eds. M.O. Amdur, J. Doull, C.D. Klaasen, 4th edition, Pergamon Press, Elmsford, NY, 12–49, 1991.
12. ACGIH Threshold Limit Values Committee, *1998 TLVs and BEIs – Threshold Limit Values for Chemical Substances and Physical Agents; Biological Exposure Indices*, The American Conference of Governmental Industrial Hygienists, Cincinnati, OH, 1998.
13. L.A. Todd, 'Direct-reading Instrumental Methods for Gases and Vapors', in *The Occupational Environment – Its Evaluation and Control*, ed. S.R. DiNardi, AIHA Press, Fairfax, VA, 177–201, 1997.
14. N.O. Korolkoff, 'Survey of Toxic Gas Sensors and Monitoring Systems', *Solid State Technol.*, **32**, 49–64 (1989).
15. C.F. Cullis, J.G. Firth, 'Detection and Measurement of Hazardous Gases', Heinemann, London, 33–52, 1981.
16. B. Bott, T.A. Jones, 'The Use of Multisensor Systems in Monitoring Hazardous Atmospheres', *Sens. Actuators*, **9**, 19–25 (1986).
17. J.W. Gardner, 'Detection of Vapours and Odours from a Multisensor Array Using Pattern Recognition Part 1. Principal Component Cluster Analysis', *Sens. Actuators B*, **4**, 109–115 (1991).
18. U. Weimar, D. Schierbaum, W. Gopel, 'Pattern Recognition Methods for Gas Mixture Analysis: Applications to Sensor Arrays Based on SnO_2', *Sens. Actuators B1*, **1**, 93–96 (1990).
19. R. Muller, E. Lange, 'Multidimensional Sensor for Gas Analysis', *Sens. Actuators*, **1**, 39–48 (1986).
20. I. Lundstrom, C. Svensson, 'Gas-sensitive Metal Gate Semiconductor Devices, in *Solid State Chemical Sensors*, eds. J. Janata, R.J. Huber, Academic Press, New York, 2–61, 1985.
21. D. Hodgins, 'The ''Electronic Nose'' Using Conducting Polymer Sensors', *Sensor Rev.*, **14**, 28–31 (1994).

22. P.N. Bartlett, S.K. Ling-Chung, 'Conducting Polymer Gas Sensors Part III: Results for Four Different Polymers and Five Different Vapors', *Sens. Actuators*, **20**, 287–292 (1989).
23. B.J. Doleman, M.C. Lonergan, E.J. Severin, T.P. Vaid, N.S. Lewis, 'Quantitative Study of the Resolving Power of Arrays of Carbon Black–Polymer Composites in Various Vapor Sensing Tasks', *Anal. Chem.*, **70**, 4177–4190 (1998).
24. D.S. Ballantine, Jr, H. Wohltjen, 'Surface Acoustic Wave Devices for Chemical Analysis', *Anal. Chem.*, **61**, 704–712A (1989).
25. J.W. Grate, S.N. Kaganove, S.J. Patrash, B.M. Wise, 'Polymer-coated Surface Acoustic Wave Sensor Arrays for Distinguishing Chemically Similar Organic Vapors', presented at 193rd Meeting of the Electrochemical Society – Sensors for Environmental Monitoring and Occupational Safety, San Diego, CA, May 6, 1998.
26. J.W. Grate, M. Klusty, 'Surface Acoustic Wave Vapor Sensors Based on Resonator Devices', *Anal. Chem.*, **63**, 1719–1727 (1991).
27. G.-Z. Zhang, E.T. Zellers, 'Portable Instrument Employing a Surface Acoustic Wave Sensor with a Regenerable Reagent Coating for Direct Measurement of 1,3-Butadiene and Styrene', *Rev. Sci. Instrum.*, **66**, 239–246 (1995).
28. H. Wohltjen, 'Mechanism of Operation and Design Considerations for Surface Acoustic Wave Device Vapour Sensors', *Sens. Actuators*, **5**, 307–325 (1984).
29. H. Wohltjen, D.S. Ballantine, Jr, N.L. Jarvis, 'Vapor Detection with Surface Acoustic Wave Microsensors', in *Chemical Sensors and Microinstrumentation: ACS Symposium Series #403*, eds. R.W. Murray, R.E. Dessey, W.R. Heineman, J. Janata, W.R. Seitz, American Chemical Society, Washington, DC, 157–175, 1989.
30. J.W. Grate, S.L. Rose-Pehrsson, D.L. Venezky, M. Klusty, H. Wohltjen, 'Smart Sensor System for Trace Organophosphorous and Organosulfur Vapor Detection Employing a Temperature-controlled Array of Surface Acoustic Wave Sensors, Automated Sample Preconcentration, and Pattern Recognition', *Anal. Chem.*, **65**, 1868–1881 (1993).
31. S.L. Rose-Pehrsson, J.W. Grate, D.S. Ballantine, Jr, P.C. Jurs, 'Detection of Hazardous Vapors Including Mixtures Using Pattern Recognition Analysis of Responses from Surface Acoustic Wave Devices', *Anal. Chem.*, **60**, 2801–2811 (1988).
32. D.S. Ballantine, R.M. White, S.J. Martin, A.J. Rico, E.T. Zellers, G.C. Frye, H. Wohltjen, *Acoustic Wave Sensors: Theory, Design, and Physico-Chemical Applications*, Academic Press, San Diego, CA, 70–98, 1997.
33. D.L. Massart, B.G. VanDeginste, S.M. Demming, Y. Michotte, L. Kaufman, 'Chemometrics – A Textbook, Elsevier, Amsterdam, 403–407, 1988.
34. E.T. Zellers, S.A. Batterman, M. Han, S.J. Patrash, *Anal. Chem.*, **67**, 1092–1106 (1995).
35. E.T. Zellers, T. Pan, S.J. Patrash, M. Han, S.A. Batterman, 'Extended Disjoint Principal Components Regression Analysis of SAW Vapor Sensor-array Responses', *Sens. Actuators B*, **12**, 123–133 (1993).
36. A.R. McGill, M.H. Abraham, J.W. Grate, 'Choosing Polymer Coatings for Chemical Sensors', *Chemtech.*, **24**, 27–37 (1994).
37. W.A. Groves, E.T. Zellers, 'Prototype Instrument Employing a Microsensor Array for the Analysis of Organic Vapors in Exhaled Breath', *Am. Ind. Hyg. Assoc. J.*, **57**, 1103–1108 (1996).
38. K.B. Pfeifer, J.L. Sprung, T.R. Galloway, 'Polymer-coated Surface Acoustic Wave Monitoring of CCl_4 in a Steam Reforming Reactor', *Sens. Actuators B*, **22**, 37–45 (1994).
39. A.E. Hoyt, A.J. Ricco, J.W. Bartholomew, G.C. Osbourn, 'SAW Sensors for the Room-temperature Measurement of CO_2 and Relative Humidity', *Anal. Chem.*, **70**, 2137–2145 (1998).
40. M.S. Nieuwenhuizen, A.J. Nederlof, 'Surface Acoustic Wave Gas Sensor for Nitrogen Dioxide Using Phthalocyanines as Chemical Interfaces. Effects of Nitric Oxide, Halogen Gases, and Prolonged Heat Treatment', *Anal. Chem.*, **60**, 236–240 (1988).
41. M. Penza, E. Milella, V.I. Anisimkin, 'Monitoring of NH_3 Gas by LB Polypyrrole-based SAW Sensor', *Sens. Actuators B*, **47**, 218–224 (1988).
42. J. Park, G.Z. Zhang, E.T. Zellers, 'Personal Monitoring Instrument for the Selective Measurement of Multiple Organic Vapors', *Am. Ind. Hyg. Assoc. J.*, submitted.
43. R. Kottenstette, G. Frye, E. Heller, J. Roark, M. Wilhelm, D. Newby, 'A Hand Held SAW Array Sensor System for Volatile Organic Compounds', presented at 193rd Meeting of the Electrochemical Society – Sensors for Environmental Monitoring and Occupational Safety, San Diego, CA, May 6, 1998.
44. M.A. Arnold, 'Fiber-Optic Chemical Sensors', *Anal. Chem.*, **64**, 1015–1025 (1992).
45. D.R. Walt, C. Munkholm, P. Yuan, S. Luo, S. Barnard, 'Design, Preparation, and Applications of Fiber-Optic Chemical Sensors for Continuous Monitoring', in *Chemical Sensors and Microinstrumentation: ACS Symposium Series #403*, eds. R.W. Murray, R.E. Dessey, W.R. Heineman, J. Janata, W.R. Seitz, American Chemical Society, Washington, DC, 252–272, 1989.
46. R.W. Seitz, 'Chemical Sensors Based on Fiber Optics', *Sensors*, August, 6–9 (1985).
47. A.Q. Wang, 'Gas and Vapour Detection Using Fibre Optic Sensing Technology', *Int. Environ. Technol.*, **8**, 26–27 (1998).
48. T.A. Dickinson, J. White, J.S. Kauer, D.R. Walt, 'A Chemical-detecting System Based on a Cross-reactive Optical Sensor Array', *Nature*, **382**, 697–700 (1996).
49. B.G. Healey, D.R. Walt, 'Fast Temporal Response Fiber-optic Chemical Sensors Based on the Photodeposition of Micrometer-scale Polymer Arrays', *Anal. Chem.*, **69**, 2213–2216 (1997).

50. J.M. Sutter, P.C. Jurs, 'Neural Network Classification and Quantification of Organic Vapors Based on Fluorescence Data from a Fiber-optic Sensor Array', *Anal. Chem.*, **69**, 856–862 (1997).
51. S.R. Johnson, J.M. Sutter, H.L. Engelhardt, P.C. Jurs, J. White, J.S. Kauer, T.A. Dickenson, D.R. Walt, 'Identification of Multiple Analytes Using an Optical Sensor Array and Pattern Recognition Neural Networks', *Anal. Chem.*, **69**, 4641–4648 (1997).
52. J. Janata, M. Josowicz, P. Vanysek, M.D. DeVaney, 'Chemical Sensors', *Anal. Chem.*, **70**, 179R–208R (1998).
53. R.S. Sethi, 'Transducer Aspects of Biosensors', *Biosens. Bioelectron.*, **9**, 243–264 (1994).
54. M.P. Byfield, R.A. Abuknesha, 'Biochemical Aspects of Biosensors', *Biosens. Bioelectron.*, **9**, 373–400 (1994).
55. B. Hock, 'Antibodies for Immunosensors – A Review', *Anal. Chim. Acta*, **347**, 177–186 (1997).
56. F. Scheller, F. Schubert, *Biosensors: Techniques and Instrumentation in Analytical Chemistry*, Elsevier, New York, 11, 1992.
57. K.R. Rogers, C.L. Gerlach, 'Environmental Biosensors – A Status Report', *Environ. Sci. Technol.*, **30**, 486A–491A (1996).
58. K. Rogers, 'Biosensors for Environmental Applications', *Biosens. Bioelectron.*, **10**, 533–541 (1995).
59. Code of Federal Regulations, Title 29, Volume 5, Parts 1900 to 1910. 999 Revised as of July 1, 1999, US Government Printing Office, Washington, DC 20401.
60. B. Maisel, 'Making the Most of Air Emission Requirements', *Pollution Eng.*, **4**, 74–75 (1996).
61. J.R. White, 'Technologies for Enhanced Monitoring', *Pollution Eng.*, **27**(6), 46–50 (1995).
62. J. Cermenaro, 'How to Buy a CEM System', *Pollution Eng.*, **27**(8), 46–47 (1995).
63. E.M. Zdankiewicz, 'Choosing Detection Devices for VOC Leak Testing', *Pollution Eng.*, **30**(7), 42–44 (1998).
64. National Institute for Occupational Safety and Health, *NIOSH Manual of Analytical Methods* – 4th edition, NIOSH, Cincinnati, OH, 1994.
65. R.M.A. Hahne, 'Direct-reading Instruments for Gases, Vapors, and Particulates', in *Fundamentals of Industrial Hygiene*, eds. B.A. Plog, J. Niland, P.J. Quinlan, 4th edition, National Safety Council, Itasca, IL, 509–527, 1996.
66. AIHA Gas and Vapor Detection Systems Committee, Manual of Recommended Practice for Combustible Gas Indicators and Portable Direct-reading Hydrocarbon Detectors, ed. S.F. Chelton, 2nd edition, American Industrial Hygiene Association, Fairfax, VA, 1993.

Spectroscopic Techniques in Industrial Hygiene

Jin Wang, Paul D. Siegel, Daniel M. Lewis, Evanly Vo, and William E. Wallace
US Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Morgantown, USA

Kevin Ashley and Lloyd E. Stettler
US Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, USA

1 Introduction **2**
2 Mass Spectrometry **4**
2.1 Introduction 4
2.2 Gas Chromatography/Mass Spectrometry 4
2.3 Liquid Chromatography/Mass Spectrometry, Liquid Chromatography Tandem Mass Spectrometry and Liquid Chromatography Ion Trap Mass Spectrometry 7
2.4 Inductively Coupled Plasma Mass Spectrometry 11
2.5 Time-of-flight Mass Spectrometry, Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Future Trends 12
3 Scanning Electron Microscopy and X-ray Microanalysis **13**
3.1 Introduction 13
3.2 Sample Preparation 15
3.3 Lung Particle Analysis 16
4 Atomic Spectrometry **18**
4.1 Introduction 18
4.2 Flame Atomic Absorption Spectrometry 19
4.3 Electrothermal-, Hydride Generation- and Cold Vapor Atomic Absorption Spectrometry 20
4.4 Inductively Coupled Plasma Atomic Emission Spectrometry 22
5 Chromatographic Spectrophotometric Detectors **23**
5.1 Introduction 23
5.2 Spectrophotometric Detectors for High-performance Liquid Chromatography 23
5.3 Spectrophotometric Detectors for Gas Chromatography 25
6 Infrared and Raman Spectroscopies **25**
6.1 Introduction 25
6.2 Infrared Spectroscopy 25
6.3 Raman Spectroscopy 27
7 Field-portable Spectroscopy **28**
7.1 Introduction 28
7.2 Portable Gas and Vapor Analyzers 28
7.3 Portable Aerosol Monitors 30
Abbreviations and Acronyms **31**
Related Articles **32**
References **32**

Over the past few decades the pace of change in spectroscopic techniques has been remarkable. Spectroscopic techniques are emerging as important, powerful, and versatile tools in determining exposure levels of hazards generated in working environments. Occupational safety and health studies employ spectroscopic techniques to analyze hazardous chemicals, biomarkers, and particulate matters of exposure. In comparison with many traditional detection techniques such as gravimetric methods, spectrometric techniques are much more sensitive, selective and accurate. The major spectroscopic techniques used in industrial hygiene include mass spectrometry (MS), scanning electron microscopy (SEM), X-ray microanalysis (XM), atomic spectrometry (AS), ultraviolet/visible (UV/VIS) photometry, fluorescent spectrometry (FS), Fourier transform infrared (FTIR) spectroscopy, and Raman spectroscopy (RS). Interest in using MS in industrial hygiene is driven by its value in understanding basic physical, chemical, and biological processes related to workers' exposure to occupational hazards, and in devising new methodologies to monitor exposures. SEM has become particularly useful in the study of pneumoconioses and workplace environmental particles since being complemented with energy dispersive X-ray (EDX) analysis and automated image analysis capabilities. SEM and EDX have been used extensively to characterize particles found in lung tissues. Atomic spectrometric methods are used widely for occupational health evaluation of inorganic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

metals. The development of inductively coupled plasma atomic emission spectrometry (ICPAES) techniques has become increasingly attractive, and has been applicable to analysis of nearly all the elements. FTIR and Raman spectroscopies are employed to detect highly toxic gas and vapor mixtures. Additionally, field-portable methods for monitoring airborne workplace contaminants and toxins have received increasing attention. To date, highly specific, selective, and sensitive spectroscopic technologies have allowed for the development of novel methodologies and new indicators for exposure characterization. Assessment of actual body burden of chemicals, which are more directly related to potential adverse occupational health effects, can be accomplished. The major spectroscopic techniques and their applications to industrial hygiene are described in this article.

1 INTRODUCTION

The pace of change in spectroscopic techniques has been remarkable. They are emerging as important, powerful, and versatile tools, and have increasing applications in the workplace. For occupational safety and health studies, spectroscopic techniques are primarily used to detect and analyze hazardous chemicals, biomarkers (e.g. metabolites, DNA adducts, protein conjugates, and allergens), and to investigate pneumoconioses and workplace environmental particles. They play an important role in determining exposure levels of hazards generated in the working environment, and in finding the adverse effects of exposures and their mechanisms of action. In comparison with many traditional detection techniques such as gravimetric methods, spectrometric or spectrophotometric techniques are much more sensitive, selective and powerful. The major spectroscopic techniques used in industrial hygiene studies include MS, SEM, XM, AS, UV/VIS photometry, FS, FTIR, and RS. In this article, the basic principles of these major spectroscopic techniques and their applications to industrial hygiene are described.

The interest in using MS by scientists in studies pertinent to industrial hygiene is driven by the need to understand the basic physical, chemical, and biological processes related to workers' exposure to occupational hazards and to devise new methodologies to monitor exposures in the work environment. MS is capable of looking at the details of exposures, and as the most sensitive tool, to analyze small molecules and macromolecules in biological systems. With tandem mass spectrometry (MS/MS) techniques, MS brought special capabilities to providing specific characterization of molecular structures and to detecting target analytes at trace levels. The development of new ionization methods such as electrospray ionization, atmospheric pressure chemical ionization (APCI), and matrix-assisted laser desorption ionization (MALDI) have contributed to great strides forward in the study of biological macromolecules. The ability to ionize polar, labile and involatile species has been the fundamental basis upon which MS extends its applications. It is important for us to recognize the breakthroughs in ionization techniques of the 1980s and 1990s. These advances have allowed application of MS to progress and develop rapidly. Additionally, combining gas chromatography (GC) or liquid chromatography (LC) with MS offers the possibility of taking advantage of both chromatography as a powerful separation technique and MS as a powerful and sensitive detection and identification technique. Furthermore, Fourier transform ion cyclotron resonance mass spectrometry (FTICRMS), and a new generation of high performance time-of-flight mass spectrometry (TOFMS), which includes nanoflow electrospray hybrid quadrupole time-of-flight mass spectrometry (QTOFMS) and matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI/TOFMS) technologies, are just beginning to add much needed analytical power to our arsenal for macromolecular identification and characterization. These developments will bring important methodology to scientists and thus accelerate the integration of MS into work-related research strategies. The range of applications in environmental health and industrial hygiene studies includes development of qualitative and quantitative analytical procedures, structural determination of aberrant proteins, investigation of biomarkers for exposure to hazards, characterization of allergens related to occupational asthma, and identification of microbial products in complex samples. MS and its related techniques and applications are described in section 2.

SEM is introduced in section 3. It is ideally suited to study pneumoconioses and workplace environmental particles when it is equipped with an EDX analysis system and an image analyzer (IA). Particle matter has been recognized as a cause of various lung diseases for many years. Specific pneumoconioses such as coalworker's pneumoconiosis, asbestosis, and silicosis may result from inhalation of particles in the workplaces. SEM and EDX have been used extensively to characterize particles found in lung tissues. While it is not possible to compare a measured lung dust burden to the actual exposure due to the lack of exposure data and to particle clearance, the measured lung particulate burden does represent retained particle dose. The SEM and EDX analysis may be used for a number of applications in occupational safety and health including the analysis of particles collected on air filter samples taken from the working environment, analysis of individual particles found in bulk dusts, and in the study of pneumoconioses.

Bulk compositional analyses of respirable particulate material, and even particle-by-particle compositional analyses, may not always be sufficient to predict the level of biological activity of respirable particles or the health risks of exposure to them. Toxicants may be located on the particle surface or may be more heavily concentrated there, resulting in heightened biological availability and expression of toxic activity. Or, in some cases, surface coating materials may have a prophylactic effect on expression of the toxicity of the underlying bulk of the particle. An example of the latter effect is the modified activity of the quartz component of some mixed composition mineral dust exposures for causing mixed dust pneumoconioses. SEM with EDX or wavelength dispersive X-ray spectroscopy can provide some information about the elemental composition with depth into a particle, by acquiring X-ray spectra at two or more electron beam accelerating voltages. This method has been used to detect thin submicrometer aluminosilicate coatings or clay ''surface occlusion'' on respirable quartz particles. Applications of SEM to the study of pneumoconioses, including determination of lung particulate burden, particle chemistry, and respirable particle surface characterization, are described in section 3 of this article.

Atomic spectrometric methods continue to be used widely for occupational health evaluation of inorganic metals and are documented briefly in section 4. The development of ICPAES techniques has been increasing, and has been applied to elemental analysis of nearly all the elements. ICPAES offers a simultaneous or rapid sequential multielement determination capability at the major, minor, and certain trace concentration levels. It has become established as a widely accepted methods for the analysis of metallic aerosols, powders of metals, dusts, and fly ashes. A vast amount of published atomic spectrometric methodology is available, covering a wide range of application areas. This is because trace elements released into the atmosphere in industrial processes have aroused great interest. Major advantages include specificity, speed, and ease of use. However, for ultratrace or trace level analysis atomic spectrometric methods have been increasingly replaced by inductively coupled plasma mass spectrometry (ICPMS). AS and its related techniques and applications are described in section 4.

Spectrophotometric detectors coupled to chromatographic separation techniques are often used to characterize the workplace environment and are discussed briefly in section 5. These techniques are employed to confirm the accuracy of direct reading instruments or when direct reading instruments are not available for the environmental contaminant, when the workplace environment is complex and contains multiple chemicals that need to be measured, for regulatory documentation of exposure levels, and for biological monitoring. Spectrophotometric detectors that are coupled to high-performance liquid chromatography (HPLC) or GC include UV/VIS, fluorescent, light scattering, refractive index, diode array, MS, infrared (IR), radioactivity, and luminescence. The choice of detector is dependent on the analyte's spectral properties, and the required sensitivity and selectivity. Both the sensitivity and selectivity of spectroscopic detectors are dependent, in part, on wavelength(s) employed. In addition, it is desirable to choose a detector with a wide linear working range to be able to assess both major and minor chemical components in the workplace environment. In monitoring the workplace environment, the most commonly employed HPLC detectors are the UV/VIS and fluorescent photometers, whereas, the most commonly used GC detectors are the flame photometric detector (FPD) and the mass spectral detector (MSD).

With the increased use of highly toxic gas and vapor mixtures in science and industry, FTIR and Raman spectroscopies have played an important role in industrial hygiene monitoring. Although FTIR and Raman spectroscopies are similar in that both techniques provide information on vibrational frequencies, there are many differences between the two techniques. For example, some vibrations are only Raman active while others are only IR active; the vibration of a heteropolar diatomic molecule is IR active, whereas that of a homopolar diatomic molecule is not IR active. The method for determining quartz content in respirable coal dust is often based on dispersive IR or FTIR spectroscopy. Like IR, RS is a powerful technique and has a variety of applications. They are complementary, and both are utilized whenever possible. IR and Raman spectroscopes and their related techniques and applications are discussed briefly in section 6.

Finally, field-portable methods, which have received increasing attention, are summarized in section 7. Field-portable methods for monitoring airborne workplace contaminants and toxins have received increasing attention. A number of portable monitors for airborne contaminants have been commercially available for many years, but new developments may provide for on-site compliance monitoring, which has heretofore been more the exception than the rule. The ability to conduct measurements on-site in the occupational setting offers significant advantages. Field-portable methods are often desired so that decisions regarding worker protection, engineering controls, and so on can be made quickly. The capability for rapid decision-making offered by on-site monitoring can help to save costs, and also offers a means to assess, and thereby prevent, worker overexposures to toxic substances in a timely manner. Field-based monitoring is especially useful for applications in the construction industry, in agriculture, and in other situations where

jobs may be short-term and the workforce is transient. On-site techniques can also be beneficial in instances where short-term monitoring is desired. In section 7, field-portable spectrometric techniques are covered, and some applications are presented.

To date, highly specific, selective, and sensitive spectroscopic technologies have allowed for the development of novel methodologies and new indicators of exposures. Therefore, the assessment of the actual body burden of chemicals, which is directly related to potential adverse occupational health effects, can be accomplished.

2 MASS SPECTROMETRY

2.1 Introduction

The mass spectrometer is an instrument capable of producing a beam of ions by converting neutral molecules into gaseous ions, and then separating these ions according to their mass-to-charge ratio and recording the relative abundances of the separated ion species as a mass spectrum.[1] Today MS has brought special capabilities to a wide variety of scientific research by providing specific analyses of substances, their metabolites, and biological macromolecules, often with structural information. The range of applications includes qualitative and quantitative analytical procedures employed in environmental health and industrial hygiene.

In instrumentation, the mass detector has been developed in a variety of types, shapes, and sizes.[2–4] The selection of the detector is based on the needs of the user and its functions:[1] the quadrupole analyzer is employed to provide an electron ionization (EI) or chemical ionization (CI);[2] the ion trap spectrometer, which is a highly geometrically modified quadrupole analyzer, is adequate for low-energy collision;[3] the ion cyclotron resonance spectrometer, which is the basis for Fourier transform MS, has high resolution;[4] the time-of-flight (TOF) spectrometer, which separates ions in time rather than space and has an almost unlimited mass range, is used for detection of macromolecules; and[5] the magnetic sector instrument, in which ion separation is achieved spatially by the application of a magnetic field (sometimes coupled with an electrostatic field), is suited to conduct collisionally activated dissociation at high energy. Furthermore, by coupling of two or more types of the above-mentioned mass analyzers, an MS/MS or an ion trap mass spectrometry (MS^n) technique is achieved. MS/MS is a key instrumental development in analytical and bioanalytical chemistry. It is widely applied in the characterization of molecular structures and in the trace analysis of targeted analytes. The most widely used activation method is called collisionally induced dissociation, in which precursor ions are selected in the first MS of multistage MS for repulsive collisions with inert gases such as helium or argon. In addition, the coupling of MS with separation techniques and the development of new ionization methods such as electrospray ionization, APCI, and MALDI have contributed to great strides forward in this field. The ability to ionize polar, labile and involatile species has been the fundamental basis upon which MS extends its applications. It is important for us to recognize the breakthroughs in ionization techniques of the 1980s and 1990s. These advances have allowed application of MS to progress and develop rapidly.

The interest in using MS by scientists in studies pertinent to industrial hygiene is driven by the need to understand the basic physical, chemical, and biological processes related to workers' exposure to hazards and to devise new methodologies to control exposures in the work environment. MS is capable of looking at the details of exposure, and is a highly sensitive tool for structural determination of aberrant proteins, development of biomarkers for exposures, identification of microbial products in complex samples, and characterization of hazards in occupational and environmental health evaluations.

2.2 Gas Chromatography/Mass Spectrometry

2.2.1 Principles and Instrumentation

The coupling of GC with MS was first achieved in 1957.[5] It is a combination of two microanalytical techniques: a separation technique, GC, and an identification technique, MS. The gas chromatography/mass spectrometry (GC/MS) combination overcomes certain deficiencies or limitations caused by using each technique individually, and gives a two-dimensional identification consisting of both a GC retention time and a mass spectrum for each component of the mixture. This combination has several advantages. First, it can separate components of a complex mixture so that mass spectra of individual compounds can be obtained for qualitative purposes; second, it can provide quantitative information on these same compounds. GC/MS can provide a complete mass spectrum from as little as 1 pmol of an analyte, which gives direct evidence for the molecular weight and a characteristic fragmentation pattern or chemical fingerprint that can be used as the basis for identification. Although the direct GC/MS method is limited to the analysis of those compounds that can be made volatile without thermal decomposition, many compounds that are nonvolatile can be handled successfully after chemical derivatization. The instrumentation of GC/MS consists essentially of three components: the gas chromatograph, the mass spectrometer and a data system. GC/MS has developed

into one of the most sensitive and selective analytical techniques for the separation, identification and quantification of components of complex mixtures.

The GC/MS technique has been utilized in a diverse range of applications such as toxicology, environmental monitoring, molecular biology, clinical health, and industrial hygiene, as well as many others. In these applications, the analyte is often present in a complex matrix consisting of a great number of compounds which may mask its presence or otherwise inhibit its detection. One of the most common ways in which GC/MS is used in industrial hygiene is as a diagnostic tool, which is particularly useful in the analysis of complex mixtures where the analytes are present in low quantities. GC/MS is also widely used in determining hazardous materials, derived metabolites, and protein/DNA adducts in biological fluids (e.g. blood or urine) from workers exposed to specific hazards.

2.2.2 Methodology

2.2.2.1 Sample Preparation Sample preparation is important for successful analyses by GC/MS. Industrial airborne hazards include gases, vapors, liquids, and particulates. Air sampling is of course a crucial step within the total scheme of air analysis. There are a variety of methods available for the collection of airborne particles. Generally, a method selected will often depend on the purpose for which the sample is being taken, and also depend on the type of compounds to be analyzed. A few preprocessing steps may be needed to manipulate the sample into a form ready for analysis. In addition, the physical state of sample material will affect the method to be used for the introduction of sample into the spectrometer. A wide variety of techniques are available for processing gaseous, liquid and solid samples. In industrial hygiene studies, one prevalent strategy in dealing with air samples consists of filter collection of particles, followed by an appropriate second stage adsorption of organic chemical vapor or fumes onto XAD-2 sorbent [treated with 2-(hydroxymethyl)piperidine] or charcoal tubes, and then solvent desorption and GC/MS analysis. There have been considerable improvements in air sampling and GC/MS detection.[6,7]

Another commonly used strategy of GC/MS analysis is the derivative method. Derivatization is an approach for increasing volatility of target analytes which is particularly useful for biological samples. In those cases where compounds are too polar or thermally unstable to be amenable to GC analysis, the situation can be improved by the formation of a suitable derivative with characteristics that render it more amenable to GC. In addition, derivatization can enhance sensitivity and selectivity by altering the fragmentation mechanism of the molecule. A further beneficial effect of derivatization is that it normally results in a compound possessing characteristic ions at mass-to-charge ratios. In tissue analysis, the sample usually requires dialysis after homogenization to remove free monomeric sugars and amino acids. Sample pretreatment of microorganisms for GC/MS analysis generally involves extraction of the class of compounds of interest (e.g. lipids, proteins, or carbohydrates) followed by derivatization.

2.2.2.2 Electron Ionization and Chemical Ionization EI is one of the main ionization methods employed in GC/MS system,[8] while CI has also been employed occasionally. In EI, energy sufficient for ionization and fragmentation of the analyte molecules is acquired by interaction with electrons (ca. 70 eV) from a hot filament. Some structural features of the analyte molecule can be deduced from the fragmentation pattern of the molecular ion. During the ionization process, in addition to the production of positive ions, a small number of molecules undergo addition of one or more electrons to form negative ions. At the operating ionizing energy of the analytical mass spectrometer (60–100 eV) the sensitivity of negative ion formation is several orders of magnitude less than that for positive ion production. Compared with EI, CI is a soft ionization technique. It achieves ionization of the analyte by collision with reagent ions (usually proton-rich ions), but without transferring excessive energy to the nascent analyte ions. The result is the formation of abundant adduct ions, often protonated molecules, that contain the intact molecular species of the analyte. In this way, CI and EI are complementary. Because of the soft ionization process, the even-electron molecular adduct ions undergo little fragmentation compared with that of the odd-electron molecular ion produced during EI. Often, CI spectra of individual analytes are sufficiently simple to allow the direct analysis of mixtures, and therefore can be used to provide structural information that is not available from an EI spectrum.

2.2.2.3 Quantitative Analysis and Selected Ion Monitoring The quantitative applications of MS are based on comparison between the ion current obtained from the analyte in the sample matrix and the ion current from another compound chosen as an internal standard, or the ion current obtained from analyses of standard aliquots of the pure analyte. The method of recording the ion current is usually by selected ion monitoring (SIM), although the technique of repetitive scanning over a narrow mass range is sometimes employed. The technique of SIM is one of the most versatile and commonly used spectrometric methods. It allows a mass spectrometer to record simultaneously the intensities of a limited number of chosen ions only. This allows the instrument to dwell for a greater proportion of the analysis time on those mass-to-charge ratios of

greatest significance in the mass spectrum of the analyte of interest, resulting in an increase in sensitivity. A major application of this technique is to provide quantitative analysis of compounds at low concentrations.[9]

2.2.3 *Applications*

In the 1990s, the application of GC/MS and related techniques for the characterization and quantitation of organic compounds and biomarkers has grown spectacularly in environmental health and occupational exposure studies. One such example is the use of GC/MS in the analysis of the health effects of the environmental aromatic hydrocarbons. Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous, and some of them are potentially carcinogenic substances, to which humans are exposed in the environment and in certain workplaces. Estimation of the resulting health risk is therefore of great occupational health importance. The determination of PAHs and their metabolites or protein/DNA adducts as biomarkers is the most suitable way to assess the exposure and to estimate this risk.

Sturaro et al.[10] have reported a GC/MS investigation of polycyclic aromatic compounds (PACs) in the manufacture of rubber tubes, using a two-stage air sampler and a GC/MS method to monitor eight PACs. With this method, detection limits of 8–15 ng mL^{-1}, corresponding to a level of 12–23 ng m^{-3} in the workplace, were achieved. Workers exposed to airborne aromatic amines in another rubber manufacturing process were evaluated by Menichini et al.[11] Samples were collected on a glass fiber filter, processed through a silica gel tube, and analyzed by GC/MS/SIM. Their work outlines a procedure for the determination of occupational exposure to airborne aromatic amines in the rubber industry and the application of this method in a tire manufacturing plant. In related work, Menichini et al.[12] determined PAHs in mineral oils and oil aerosols in glass manufacturing. PAHs were analyzed by GC/MS in graphited mineral oils used for mold lubrication and in aerosols emitted during their application in two plants. High boiling PAHs were detected in oils but generally not in air. Volatile PAHs were found in oil samples and to a lesser extent in air near the emission source. Additionally, Bundt et al.[13] have investigated structure-type separation of diesel fuels by solid-phase extraction (SPE) and identification of the two- and three-ring aromatics. Commercially available standards were used for identification. Sulfur-containing PAHs in diesel fuel are mainly represented by methyl-substituted dibenzothiophenes. Cooper[14] has developed a GC/MS method to confirm the presence of *N*-nitrosamines in workplace air samples. Detection limits of the three procedures and retention time precision of both SIM techniques are good. Typical examples of the use of these techniques for confirmation of *N*-nitrosamines are described by Cooper.[14] More applications are summarized in Table 1.

Urinary naphthols (1- and 2-naphthol) have been suggested as route-specific biomarkers for exposure to airborne PAHs.[27] The application of urinary naphthol levels as biomarkers in 119 Japanese male workers was reported. The urinary 1- and 2-naphthol levels were observed three and sevenfold higher, respectively,

Table 1 Typical applications based on GC/MS methods

No.	Chemical hazards	Sample type	Ref.
1	PAHs (glass manufacturing)	Mineral oils and oil aerosols	12
2	PAHs in diesel fuel	Petroleum	13
3	PAHs (cooking plant and foundries)	Workplace air, and dust	21
4	PACs (energy laboratory)	Coal combustion product	16
5	PACs (manufacture of rubber tubes)	Workplace air	10
6	Perchloroethylene (five dry cleaning firms)	Workplace air, and blood of exposed workers	19
7	Airborne aromatic amines (rubber manufacturing)	Workplace air	11
8	*N*-Nitrosamines	Workplace air	14
9	VOCs	Indoor air	18
10	Polychlorinated biphenyls (electrical workers)	Workroom surfaces, tools, palms, and blood	22
11	Polybrominated dioxins, dibenzofurans, octachlorostyrene	Combustion, and flame retardants	20
12	Glycol ethers (newspaper printing plant)	Workplace air	17
13	Organic vapor pollutants	Ambient atmospheres	15
14	Resin acid compounds	Emitted from rosin in soldering flux	23
15	Semivolatile organic compounds	Workplace atmospheres	24
16	Airborne chemical agents	Air	25
17	Coal tar pitch volatiles	Cold tar pitch fume	26

VOCs, volatile organic compounds.

among smokers than among nonsmokers. Also the ratios of urinary 2-naphthol to 1-naphthol were significantly higher among smokers than nonsmokers. In another study, a method for the simultaneous determination of urinary phenanthrene, fluoranthene, pyrene, chrysene and benzo[*a*]pyrene metabolites has been developed for individual risk assessment at a PAH-burdened workplace.[28] The method allows the determination of 25 different components. The PAH exposure of coke plant workers during several consecutive days resulted in fairly constant individual urinary metabolite profiles.[29] It was also demonstrated that in the case of coke plant workers there is a correlation between inhaled PAHs and metabolites excreted. Mass relationships between inhaled PAHs and metabolites excreted were found to differ from one individual to another.

The exposures of agricultural workers to organochlorine pesticides were studied by Guardino et al.[30] The chlorinated pesticides and their metabolites in whole blood samples from 30 farmers and 24 nonoccupationally exposed workers were determined by GC/MS. The potential that sawmill workers might be exposed to chlorophenols was investigated by Kontsas et al.[31] A GC/MS procedure for the determination of chlorophenols in urine was developed. The concentrations of urinary chlorophenols in previously exposed workers were of the same magnitude as those found in nonexposed controls and in the general population. The feasibility of using plasma, blood and hemoglobin (Hb) adducts for monitoring occupational exposure to the suspected human carcinogen 4,4′-methylenebis(2-chloroaniline) (MOCA) was investigated by Vaughan et al.[32] The levels of MOCA in the blood and urine of five individuals who were exposed to MOCA during the manufacture were determined by the GC/MS method. It was found that the use of blood samples for monitoring exposure to MOCA offers advantages over the currently used urinary MOCA measurements.

Occupational exposure to toluene diisocyanate (TDI) among workers in a polyurethane foam factory[33] was studied during 48-h periods and biological samples from nine subjects. Five workers were found to show high average urinary elimination rates of TDI. The elimination rate curves for all of the subjects studied had a linear relationship with exposure to TDI. The study indicates that it is possible to monitor exposure to TDI by monitoring urinary concentrations of TDI by GC/MS. A study of exposure to benzidrine was reported by Hsu et al.[34] Exposure to benzidine, which is subsequently acetylated to *N*-acetylbenzidine and *N*,*N*′-diacetylbenzidine, has been implicated in the development of bladder cancer in humans. In this study, an isotope dilution GC/negative ion CI MS method was developed to quantitate urine concentrations of benzidine and its acetylated metabolites. The method is applicable to the measurement of other aromatic amines and their acetylated metabolites. Worker exposure to sawing fumes from pine was investigated by Eriksson et al.[35] Three metabolites from α-pinene have been identified in human urine after occupational exposure to it. Urine was enzymatically hydrolyzed, and metabolites were identified by GC/MS using EI and CI with isobutane as the reagent gas. The use of Hb and serum–protein adducts of hazard reagents as biomarkers for occupational and environmental exposure assessment has received increasing interest. The environmental pollutant 2,4,6-trinitrotoluene (TNT) is an important occupational health hazard, and is taken up through the skin and by inhalation. It is therefore essential to have fast and reliable methods to monitor human exposure. In a related work, a GC/MS method, which quantifies Hb adducts of TNT for 50 workers and controls from a Chinese munition factory, was reported by Sabbioni et al.[36] The Hb adduct levels ranged from 3.7 to 522 ng. However, in control samples no adducts could be found. In another study,[37] alachlor–protein adducts were examined as potential biomarkers of alachlor exposure, a genotoxic and carcinogenic herbicide. The method developed was based on the observation that cleavage of *S*-cysteinyl alachlor–protein adducts by methanesulfonic acid gave the rearrangement product.

Hb samples from ethylene oxide-exposed workers and nonexposed referents were analyzed by Farmer et al.[38] GC/MS was used to determine an Hb adduct as its methyl ester heptafluorobutyryl derivative, after hydrolysis of the protein and isolation of the alkylated amino acid. Ranasinghe et al.[39] have reported an application of GC/electron capture negative CI high-resolution MS for characterization and quantitation of DNA and protein adducts. The method has adequate sensitivity and specificity to measure accurately DNA and protein adducts as low as endogenous concentrations in rodent and human tissues. Additional applications are listed in Table 2.

2.3 Liquid Chromatography/Mass Spectrometry, Liquid Chromatography Tandem Mass Spectrometry and Liquid Chromatography Ion Trap Mass Spectrometry

2.3.1 Principles and Instrumentation

The history of liquid chromatography/mass spectrometry (LC/MS) starts in the early 1970s,[62] and since then the technique has been developed rapidly. There are several general reviews[63–66] and books published on this subject.[67–70] The combination of LC and MS offers the possibility of taking advantage of both LC as a powerful separation technique and MS as a selective and sensitive detector. A considerable number of LC/MS

Table 2 Typical biomonitoring based on GC/MS methods

No.	Biomarkers	Sample type	Ref.
1	Alachlor–protein adducts (alachlor exposure)	In vitro and in vivo	37
2	Hb adducts (ethylene oxide exposure)	Blood	38
3	Hb adducts (methyl bromide exposure)	In vitro erythrocytes	40
4	Trinitrotoluene and metabolites	Urine	42
5	Hb adducts (3,3-dichlorobenzidine exposure)	Rat erythrocytes	43
6	PAHs and metabolites (PAH-exposed workers)	Urine	28
7	DNA adducts (2,3-epoxy-4-hydroxynonanal exposure)	Calf thymus DNA	44
8	Metabolites (monoterpenes α-pinene and β-pinene exposure)	Urine	35
9	Hb adducts (TNT)	Blood	36
10	Benzidine and metabolites (benzidine exposure)	Urine	34
11	Hb [(4,4-methylenebis(2-chloroaniline)]	Blood, plasma, and urine	32
12	Hb adducts and metabolites (MDI exposure)	Urine	45
13	PAHs (at various workplaces)	Urine	29
14	DDT and related compounds (agricultural workers)	Whole blood	30
15	Chlorophenols (sawmill workers)	Urine	31
16	VOCs	Whole blood	41
17	MDA and metabolites (MDA exposure)	Urine and blood plasma	46
18	Hb adduct	Rat erythrocytes	47
19	Pesticide metabolites (pesticides exposure)	Urine	48
20	*N*-Phenylvaline (benzene exposure)	Blood	49
21	*S*-Benzyl-*N*-acetyl-L-cysteine (toluene exposure)	Urine	50
22	DNA damage (benzo[*a*]pyrene exposure)	Human lymphocytes	51
23	Toluenediamine (TDI)	Urine and plasma	33
24	Polyaromatic carcinogen–DNA adducts	Review (humans)	52
25	Hb adducts (PAH exposure)	Blood	53
26	Phenol and metabolites (phenol exposure)	Urine and plasma	54
27	Hb adducts (ethylene oxide exposure)	Blood	55
28	*S*-(2-Carboxyethyl)cysteine (acrylamide exposure)	Blood	56
29	DNA–protein adducts (PAH exposure)	Review	57
30	DNA adducts (trace organic exposure)	Review	58
31	PAH–DNA adducts (PAH exposure)	Review	59
32	DNA adducts (exposure assessment)	Review	60
33	DNA adducts (chemical carcinogenesis)	Review	61

DDT, 1,1,1-trichloro-2,2-bis[*p*-chlorophenyl]ethane; MDI, methylenediphenyl diisocyanate; MDA, methylenebisaniline.

interfaces have been developed. Widely used interfaces included particle beam, thermospray, continuous-flow fast atom bombardment (FAB), electrospray and APCI. Electrospray ionization was one of the most important ionization methods in the 1990s, and is extensively used in LC/MS systems. Electrospray is the result of charging a liquid at a needle tip by applying a high potential. With the increase in the potential, the droplet size is reduced and the droplets begin to have a horizontal component in their movement as well as a higher speed. Nowadays most types of atmospheric pressure ion sources can be used in combination of an electrospray and APCI interface, which is the most widely adapted interface to LC/MS systems.

As a hyphenated technique, LC coupling to MS results in a powerful and versatile analytical tool. Furthermore, LC coupling to a tandem or an ion trap mass spectrometer (LC/MS/MS or LC/MS^n) plays another important role in the development of new analytical strategies. MS/MS, generally, includes a collision cell in which deliberate fragmentation of parent ions can be achieved. Collision-induced fragmentation provides daughter ions by collisional energy transfer between parent ions and a collision gas, normally helium, nitrogen or argon, at an elevated pressure. The most widely used MS/MS configuration is the triple quadrupole instrument, where mass analysis is performed in the first and third quadrupole while the second quadrupole is used as a collision cell. It offers a major advantage for structural characterization of components in mixtures and trace analysis. Successful interfacing of this technique has created an effective methodology in the analysis of nonvolatile, labile, and macromolecular compounds. A wide variety of important applications include determination of large portions of protein sequences with either on-line or off-line enzymatic hydrolysis, and development of molecular biomarkers, which represent a more accurate determinant of potential risk of exposure to hazards or carcinogens than those which only assess external exposure.

2.3.2 Methodology

2.3.2.1 Sample Preparation Both on-line and off-line sample pretreatment by SPE are commonly employed in LC/MS, LC/MS/MS, and LC/MS^n methods. In quantitative studies of metabolites, and protein–DNA adducts, off-line sample pretreatment appears to be preferred. This is partly due to the composition of biological samples, where the presence of proteins may cause clogging of the SPE columns or cartridges used. For example, a typical Hb and DNA adducts assay involves extraction of DNA or Hb from its biological matrix (e.g. blood or tissue), cleavage of the macromolecule to smaller components, and the use of enrichment techniques, such as derivatization and/or SPE preconcentration. To speed up a complete procedure, the sample pretreatment can be performed on an automated batch-scale SPE. Sometimes, improving the speed and/or performance of the sample pretreatment prior to LC/MS analysis is a matter of considerable research.

After determination of the molecular mass and the elemental composition, the next step in the qualitative analysis is the interpretation of fragment ion peaks in the mass spectrum to achieve structure elucidation. A powerful tool in structure elucidation is the use of MS/MS. The information from a measured spectrum is reduced to a small number of the most significant peaks and then compared with the library spectra. Computerized library searching is very useful as it provides ideas on which direction to search when a completely unknown analyte must be identified, or provides adequate confirmation when the presence of a compound is to be confirmed.

2.3.2.2 Interface Technology In the 1970s, 1980s and 1990s, LC/MS development has resulted in a considerable number of different interfaces. A major effort for this research was to improve the instrument capability to ionize analytes including highly polar, labile and biomacromolecules directly from the liquid phase. The most widely used interfaces are electrospray, APCI, thermospray, particle beam, and continuous-flow FAB. Detailed interfacing strategies are available from a book by Niessen.[62] The electrospray interface provides one of the most promising interfaces for LC/MS in that it disposes of the mobile phase during spray generation and offers detection limits into the femtomole range. It has the following advantages: (1) direct ionization from solution; (2) production of multiply charged ions which extend the effective mass range of the mass analyzer; (3) introduction of methods to aid in desolvation of the analyte; and (4) low background from the ionization process. It is remarkable that we can obtain the mass spectrum of a protein with a large molecular mass, and consume an amount of sample of the order of picomoles to femtomoles. APCI is also used for continuous monitoring of an HPLC column. The vaporized eluate from the HPLC is forced through the APCI source with a slow heated stream of nitrogen gas.

2.3.3 Applications

Application of LC/MS in the analysis of workplace and environmental samples has been increasing, especially in the development of exposure indicators. For example, direct analysis of DNA adducts by using LC combined with electrospray ionization MS can obviate those problems arising from the employment of chemical derivatization needed for GC/MS. That is because DNA adducts are usually very polar, and the derivatization is often difficult and less successful. A number of studies have demonstrated that protein–DNA adducts in biological samples can be quantified precisely and accurately by using LC coupling to MS or MS/MS.[71–75] Workers are occupationally exposed to a wide array of chemical compounds. The compounds or their metabolites can interact with biological macromolecules such as proteins, RNA, and DNA. These interactions can result in covalent bonding between the chemicals and macromolecules, leading to DNA damage and the formation of DNA adducts. If these damages are not enzymatically repaired, they can be the cause of mutations and might lead to chemically induced carcinogenesis. The structural elucidation of these DNA adducts is an important research topic in cancer prevention. The use of DNA adducts as biological markers for risk assessment and occupational and environmental monitoring has generated great interest because they represent direct indications of primary damage to genetic material by chemicals. Therefore, protein–DNA adducts may prove to be more accurate and reliable than measurement of external exposure. To accomplish these goals, a sensitive and specific LC/MS method is capable of detecting low picomole quantities of adduct in relatively small complex biological samples, and is specific enough to confirm the structure of the adducts.

The coupling of LC to MS and MS/MS pertinent to industrial hygiene studies has been reported by a number of researchers. Some of them are summarized in Table 3. In one study,[76] sensitive and specific isotope dilution LC/MS and LC/MS/MS methods were developed for the detection and quantitation of DNA adducts formed upon exposure of animals to carcinogenic 1,2-dihaloethanes, 1,2-dichloroethane and 1,2-dibromoethane. These are important industrial chemicals used as additives in gasoline, as intermediates in the production of vinyl chloride, vinyl bromide, and other halogenated organics, as components of grain or soil fumigants, and as solvents for cleaners and other industrial products. In another study, an HPLC/electrospray MS method was developed for the analysis of 7-(2-hydroxyethyl)guanine, the major

Table 3 Typical applications based on LC/MS, LC/MS/MS, and LC/MS^n methods

No.	Chemical hazards and/or biomarkers	Methods	Ref.
1	DNA adduct (1,2-dichloroethane and 1,2-dibromoethane exposure)	LC/MS/MS	76
2	Benzene metabolites (benzene exposure)	LC/MS/MS	82
3	Aromatic sulfonates (textile industry or construction)	LC/MS	71
4	CP and IF (health care personnel occupationally exposure)	LC/MS/MS	81
5	DNA adducts (1,3-butadiene exposure)	LC/MS/MS	80
6	Hb adducts (methyl bromide exposure)	LC/MS/MS	83
7	DNA adducts (treated with *N*-nitrosodiethylamine)	LC/MS	79
8	PAH–DNA adducts (in vitro reaction with PAHs)	LC/MS	72
9	DNA adducts (in vitro reaction with bisphenol A diglycidyl ether)	LC/MS	78
10	Ethylene oxide–DNA adduct (ethylene oxide exposure)	LC/MS	73
11	DNA adducts (vinyl chloride exposure)	LC/MS	77
12	PAH metabolites (PAH exposure)	Review	74
13	DNA adducts (in vitro reaction with bisphenol A diglycidyl ether)	LC/MS/MS	75
14	Hazardous industrial chemicals	LC/MS	85
15	Aliphatic isocyanates	LC/MS	86
16	Polymeric MDI and other isocyanates	LC/MS	87
17	TDI	LC/MS	88

CP, cyclophosphamide; IF, ifosfamide; MDI, methylenediphenyl diisocyanate.

DNA adduct formed after exposure to ethylene oxide. The method is based on DNA neutral thermal hydrolysis, adduct microconcentration, and final characterization and quantification by HPLC coupled to single-ion monitoring electrospray MS. The method was found to be selective, sensitive, and easy to handle with no need for enzymatic digestion or previous sample derivatization. Yen et al.[77] have developed a method to quantify *N*,3-ethenoguanine, a promutagenic DNA adduct of vinyl chloride exposure. The applicability of the method was established by determining DNA adduct in rats treated with chloroethylene oxide and an unexposed human liver. It was observed that the concentration of DNA adduct in the rat livers increased with increasing dose, but was inversely related to the time after exposure. This trend suggests rapid DNA repair and that adducts reduce in rat livers.

Vanhoutte et al.[78] developed a nanoscale LC/electrospray MS methodology for the detection and identification of DNA adducts by in vitro reaction mixture resulting from the interaction of calf thymus DNA with bisphenol A diglycidyl ether. In other work, Singh et al.[79] developed an HPLC/electrospray MS method to detect and characterize two major ethylated DNA adducts for monitoring exposure to genotoxic ethylating agents. This approach was shown to be capable of detecting the DNA adduct in liver tissue from mice treated intraperitoneally with *N*-nitrosodiethylamine. A major chemical in rubber and plastics manufacture, 1,3-butadiene, inducing DNA adducts in vivo and in vitro, was investigated by Tretyakova et al.[80] The LC/electrospray MS/MS methods developed in this work provide the means to study accumulation, repair and dose–response relationships of 1,3-butadiene–DNA adducts in vivo.

Occupational exposure to cyclophosphamide (CP) and ifosfamide (IF) was investigated by Minoia et al.[81] An LC/MS/MS system was employed to monitor CP and IF exposure of 24 workers. The extent of exposure was assessed by the analysis of air samples, wipe samples, dermal pads, and urinary excretion at the beginning and at the end of the work shift. The results of this investigation demonstrate that higher risk may be caused by incorrectly using airflow hoods.

Benzene is an important industrial chemical and ubiquitous environmental pollutant. It is used in the manufacturing of a wide variety of consumer products. Melikian et al.[82] developed a sensitive and specific LC/MS/MS assay for determination of urinary benzene metabolites. The objective of this study was to investigate how various levels of exposure affect the metabolic activation pathways of benzene in humans and to examine the relationship between urinary metabolites and other biological markers. Ferranti et al.[83] used LC/MS/MS for the structural study of adducts formed in human Hb by in vitro exposure of erythrocytes to the alkylating agent methyl bromide (MeBr). MeBr is a highly toxic gas widely used as a fumigant of field soil for control of a wide spectrum of pests and diseases. Peptide mapping by this method allowed location of methylated amino acids within the protein sequence. The results demonstrated the usefulness of the analytical approach for the characterization of Hb adducts with methyl bromide or similar compounds, which can constitute the basis for biomonitoring of human exposure.

The LC/MS, LC/MS/MS, or LC/MS^n techniques have been developed for characterization and quantitation of pesticides, herbicides, and insecticides.[84] For example,

carbamate pesticides are used in large quantities. Since their thermal liability prohibits GC analysis, the analysis of these compounds and their metabolites is usually performed by LC/MS. They have been analyzed using most commercially available interfaces. Positive ion detection with a soft ionization technique is the method of choice. Phenylurea herbicide is one of a group of herbicides which are frequently analyzed by LC/MS or LC/MS/MS.[84] Chlorinated phenoxy acid herbicides have been extensively analyzed by means of particle-beam, thermospray, and electrospray LC/MS.[84]

2.4 Inductively Coupled Plasma Mass Spectrometry

2.4.1 Principles and Instrumentation

The inductively coupled plasma (ICP) was first utilized as an ion source for analytical MS by Montaser.[89] The degree of ionization of most elements is 90% or greater. Since its introduction, ICPMS has exhibited a large number of special attributes. The most important and unique characteristics of the argon ICP are as a viable source for simultaneous multielement analysis with high sensitivity. In ICPMS, the test sample is typically converted to an aerosol and transported into the plasma where the desolvation–vaporization–atomization–excitation–ionization processes occur. In comparison with classical combustion flames, the argon ICP exhibits a high gas temperature (4500–8000 K) and a high electron temperature (9800–10 000 K). Such conditions coupled with the relatively long plasma–sample interaction times lead to nearly complete vaporization–atomization of sample aerosol, and also reduce the chemical and physical interferences in the plasma.

In general the ICP source coupled with a quadrupole MS allows a multiple-elemental, multiple-isotope analysis to be performed. Increasingly, the argon ICPMS is replacing established atomic emission and atomic absorption spectrometries for trace or ultratrace element research. ICPMS offers simultaneous or rapid sequential multielement determination capability at major, minor, trace, and ultratrace concentration levels. Indeed, it is the most powerful means for determination of trace inorganic metals.

2.4.2 Methodology

2.4.2.1 Sample Preparation The sample preparation for ICPMS measurement is not substantially different from that encountered in other atomic spectrometric methods. Dissolution of a heterogeneous sample which provides homogeneity at the molecular level is usually required. Elemental analysis can be performed in nearly all kinds of matrices. The development of sample decomposition for ICPMS is an important step in sample preparation. It relies on sample type, specific analyte species of interest, and analyte molecular interactions. Most biological samples are prepared by decomposing the sample using thermal or chemical means, followed by dissolution of the ash residue and dilution to a specific volume prior to analysis. Microwave digestion is often used. Another common technique is a hot-plate dissolution procedure, but there are several primary limitations, such as long dissolution times and the potential loss of volatile elements. In a microwave digestion technique, closed vessels are utilized to decompose samples and minimize loss of volatile elements during the digestion process. During sample preparation, the elements being determined, their analysis requirements, and specific interferences that might be encountered for their determination dictate whether separation and preconcentration steps might be required. Sometimes, sample preparation may be a lengthy and complex process, depending on the form of the sample and the specific elements being determined.

2.4.2.2 Quantitative Analysis Quantitative analysis by ICPMS can be achieved by the use of a precise peak-hopping/signal integration procedure. Two modes of operation typically are used: rapid spectral scanning and peak hopping. The most commonly used mode, the rapid scanning method, covers the entire mass range. By closely matching the bulk chemical composition of the calibration standards to the known matrix of the samples, improved accuracy can be obtained. The use of internal standards is highly recommended to achieve maximum precision and accuracy. The calibration curves for selected elements, when plotted on logarithmic axes, demonstrate linearity over a wide dynamic range. Most modern ICPMS instruments offer a wide dynamic range. The ICPMS technique also provides sufficient isotope ratio precision and sensitivity to enable isotope dilution quantitation at trace concentration levels. Additionally, flow injection analysis (FIA) has been successfully used to improve the quantitative determination of trace elements in samples with high dissolved solids. Chemical modification of the sample can be performed by the addition of reagents to the transport line. Another significant advantage of the flow injection technique is the ability to make measurements on microsized samples, especially when high efficiency nebulizers are used.

2.4.2.3 Laser Ablation Laser ablation was first reported for sample introduction into ICPMS by Gray.[90] With this method, samples can be analyzed with minimal sample preparation. Usually specimens of metal alloys or similar materials are prepared by grinding or polishing a flat surface. Similar to other solid sampling techniques, it provides viable analytical results when suitable solid standards are available. Laser ablation ICPMS also gives important spatial resolution, which is particularly useful for the determination of the chemical composition

of grain boundaries or mineral inclusions. It is a very powerful method for the analysis of particulates or solid samples.

2.4.3 Applications

ICPMS has been utilized effectively in industrial work environments. It has been effectively used for trace analysis with high sensitivity, especially for the traditionally ''difficult-to-excite'' refractory elements such as molybdenum (Mo), vanadium (V) and zirconium (Zr), and with multielement detection capability. The applications include the determination of trace elements in air, exhaust, liquid, and dust in various working environments.

Monitoring occupational exposure to heavy metals with the ICPMS technique had been reported by several researchers, although atomic absorption spectrometry has until now been used most extensively in occupational and environmental health. In one study, Schramel et al.[91] established an ICPMS analytical method to determine the concentration of antimony (Sb), bismuth (Bi), lead (Pb), cadmium (Cd), mercury (Hg), palladium (Pd), platinum (Pt), tellurium (Te), tin (Sn), thallium (Ti) and tungsten (W) in urine. The aim of this work was to develop a method which is equally suitable for the determination of occupationally as well as environmentally caused metal excretion. In another study, Apostoli et al.[92] evaluated multiple exposure to metals in eight types of metal welding, such as manual metal arc for mild and stainless steel, continuous wire, submerged arc, and brazing. Environmental monitoring was carried out in eight different occupational situations and the ICPMS technique was adopted in order to characterize exposure to several elements simultaneously and with high accuracy. The results showed that up to 23 elements could be measured. The highest concentration was found for aluminum (Al), manganese (Mn), iron (Fe), nickel (Ni), chromium (Cr), copper (Cu) and zinc (Zn). Karpas et al.[93] presented a simple method, based on ICPMS, for determination of uranium in urine at levels that indicate occupational exposure. Sample preparation involves a 50-fold dilution of the urine by nitric acid (2% HNO_3) and no other chemical treatment or separation. The analytical procedure is fully automated so that over 100 analyses may be performed per day. Measurement by ICPMS of lead in plasma and whole blood of lead workers and controls was reported by Schutz et al.[94] The levels of lead in blood plasma and whole blood were measured by ICPMS in 43 male lead smelter workers and seven controls. The samples were handled under routine laboratory conditions. By a simple dilution procedure, lead in plasma may be determined accurately and with good precision down to the concentrations present in controls. It suggested that lead in blood plasma should be considered as a complement to current indicators of lead exposure and risk. Application of ICPMS to monitor radionuclide was reported by Vita and Mayfield.[95] An ICPMS method was developed to detect U-235 and U-238 in urine. The strong nitric and hydrochloric acid digestion of the urine and the application of the anion exchange resin for the uranium separation provided dependable recovery. In addition, multiple exposure to arsenic (As), antimony (Sb), and other elements in art glass manufacture was studied by Apostoli et al.[96] The results confirmed that arsenic, which is the main hazard in glass production, reaches a high air concentration.

2.5 Time-of-flight Mass Spectrometry, Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Future Trends

2.5.1 Introduction

TOFMS was introduced commercially in the late 1950s.[97] However, it was only in the 1990s that the high mass range and the high sensitivity multichannel recording capabilities were realized, which make this type of spectrometry an attractive instrument for contemporary research. A typical TOFMS system contains three main components: ion source, ion drift region, and detection system. A new generation of high-performance TOFMS instrumentation includes nanoflow electrospray hybrid quadrupole TOFMS and MALDI/TOFMS. Modern TOF instruments have many intrinsic advantages. They are ideally suited for pulsed ion sources because they have a sufficiently narrow pulse width, and a fast response. They can provide macromolecular measurements of proteins with considerably higher accuracy than gel electrophoresis, or can be used to map enzymatic digests, reveal post-translational modifications, determine the positions of disulfide bonds, assess carbohydrate heterogeneity in glycopeptides, or provide amino acid sequences.

FTICRMS is now an established viable analytical technique.[4] FTICRMS allows in principle an unlimited mass range to be reached with an extremely high resolution. As a technique based on Fourier transform, the resolution depends on the observation time, which is linked to the disappearance of the detected signal. In order to achieve high resolution, a very high cell vacuum is necessary, which is a major limitation of this technique.

The advancing technologies of hybrid QTOFMS, MALDI/TOFMS and FTICRMS are just beginning to add much needed analytical power to our arsenal for macromolecular identification and characterization. An application of more efficient collision-induced dissociation and product-ion detection is using a hybrid of quadrupole and TOF analyses in QTOFMS. These developments will bring important methodology to scientists and thus accelerate the integration of MS into

work-related research strategies. Research will include structural determination of aberrant proteins, development of biomarkers for specific hazard to exposed workers, characterization of protein allergens related to occupational asthma, and identification of microbial products in complex environmental samples.

2.5.2 *Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry*

A rapidly advancing area in which TOFMS plays a major role is in MALDI/TOFMS. It has developed into a powerful tool for characterizing biological macromolecules. In a MALDI/TOFMS experiment, the analyte of interest is mixed with an appropriate matrix material which facilitates the desorption and ionization of intact analyte molecules with masses up to several hundred thousand m/z. This absorbs strongly at the wavelength of the incident laser light. Many very large biomolecules have been mass analyzed by this technique. MALDI/TOFMS is able accurately to sequence single strands from DNA without the need for labels or primers,[98] and to identify single base mutations within a polymerase chain reaction (PCR) product. Although the effective range of sequencing may not be as wide as with conventional methods, MALDI/TOFMS can sequence short DNA strands in less time with the ability to identify all of the bases in the strand. Single strands and PCR products of up to 500 bases have also been detected, but it is not clear whether the mass accuracy and resolution will be sufficient for sequencing long strands. Peptide mapping by MALDI/TOFMS is gradually reaching a confidence level owing to improved mass accuracy and sample preparation methods and the availability of complete genomic information for a number of organisms.

MALDI/TOFMS has been applied and established as a valuable analytical technique for the detection of PAHs with high spatial resolution in the work environment. Bezabeh et al.[99] investigated nitrated PAH pollutants generated by incomplete combustion using negative ion laser desorption ionization TOFMS. In related work, spatially resolved laser desorption/laser ionization TOFMS has been used for the detection of PAH–picrate complexes,[100] and PAHs in individual micrometer-sized diesel particulate.[101]

2.5.3 *Hybrid Quadrupole Time-of-flight Mass Spectrometry*

The development of a hybrid QTOFMS instrument overcomes the mass range and precursor ion resolution limitations of a triple-quadrupole mass spectrometer.[102] A particular advantage that this hybrid has over the triple-quadrupole instrument is that it can provide high precursor ion resolution by virtue of its double focusing properties and it may be used to obtain MS/MS information from precursor ions of the same nominal mass. The sequencing of larger proteins by enzymatic cleavage and subsequent MS/MS analysis of the resulting peptides by QTOFMS is well documented. With the advantage of electrospray, also enabling an on-line nanoflow LC coupling to QTOFMS, this approach has found a few applications.[103] For example, Deforce et al.[104] characterize DNA oligonucleotides by coupling of capillary zone electrophoresis to electrospray ionization QTOFMS. A procedure for fast and precise molecular weight, purity, and base composition determination of oligonucleotides was described. This method has been useful not only for determination of the purity and the length of bases in oligonucleotides, but also for confirmation of the expected base composition, making this technique an extremely useful tool for quality control in the field of oligonucleotide research. Hybrid QTOFMS has just started to be applied to occupational health studies.

3 SCANNING ELECTRON MICROSCOPY AND X-RAY MICROANALYSIS

3.1 Introduction

The scanning electron microscope, when equipped with an EDX analysis system and an IA, is ideally suited to study pneumoconioses and workplace environmental particles. The salient features of this instrumentation which make it valuable in occupational safety and health applications are discussed very briefly below. For a comprehensive description of the instrumentation and theory, the reader is referred to the texts by Goldstein et al.[105] and Lee.[106] An excellent short description of SEM and X-ray analysis principles and instrumentation is given by Ingram et al.[107]

3.1.1 *Instrumentation*

In SEM, a focused beam of electrons is scanned in a raster pattern across a specimen of interest. Numerous complex events occur at each point where the electron beam impinges on the atoms of the specimen. Some, but by no means all, of the signals produced by the electron beam–specimen interaction are illustrated in Figure 1.

By utilizing an appropriate detector for a specific signal, an image of the specimen can be constructed on a point by point basis. For instance, the detection of secondary electrons, which are very low energy electrons arising from very near the surface of the specimen, is used to produce the secondary electron image (SEI). The SEI is the three-dimensional-like image

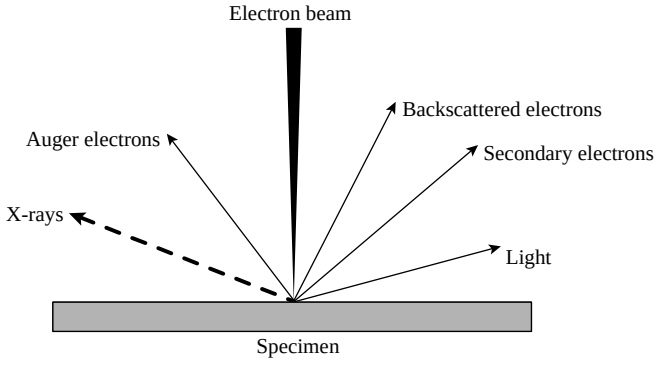

Figure 1 Some signals resulting from the interaction of the electron beam with a specimen.

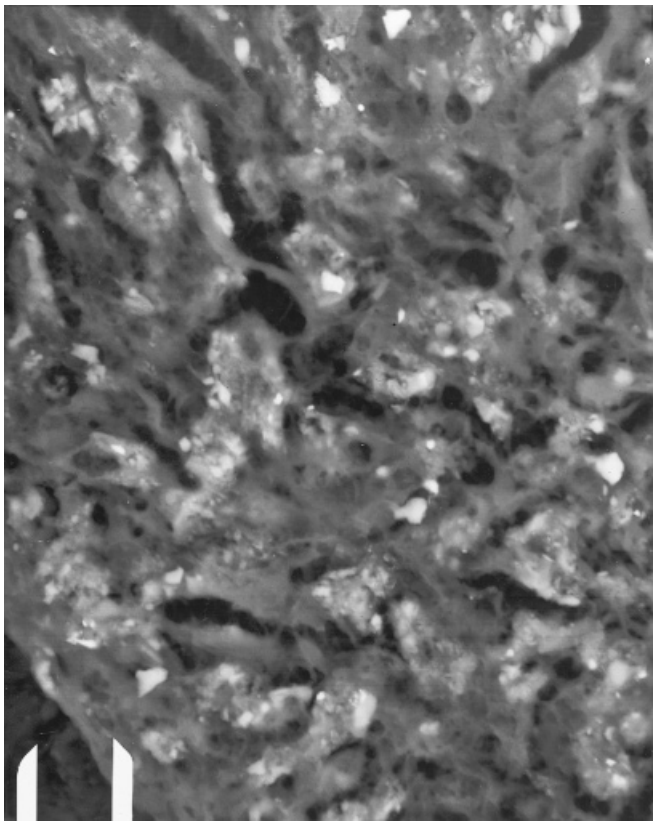

Figure 2 BEI of a 5-μm thick section of lung tissue from a subject with silicosis showing numerous particles (brighter areas). Marker = 10 μm.

typically associated with SEM. A second image which is particularly valuable in the study of pneumoconioses is the backscattered electron image (BEI). Backscattered electrons result when the original beam electrons are elastically scattered from the nuclei of atoms within the substrate. Hence, the backscattered electron signal is affected by atomic number. Furthermore, the BEI may provide some information about composition beneath the specimen surface. In our laboratory, we generally use the BEI for our pneumoconiosis studies since the exogenous particles generally are of higher average atomic number than the surrounding lung tissue and are easily identified. Figure 2 shows a BEI image from a 5-μm thick section of lung tissue from a subject with silicosis.

X-rays are generated by the inelastic interaction of the electron beam and atoms of the specimen. Two types of X-ray are produced: (1) white radiation or bremsstrahlung and (2) characteristic X-rays. The bremsstrahlung results from the interaction of the beam electrons with the nucleus of the specimen atoms, and forms a continuous background over which characteristic X-ray signals are superimposed. Characteristic X-rays result from the ejection of inner shell electrons of the specimen atoms by the electron beam. The generation of characteristic X-rays forms the basis for chemical analysis of the specimen since both the wavelength and the energy of the characteristic X-rays can provide definitive information about the atoms from which they emanate. In SEM, X-rays may be measured through the use of either a wavelength dispersive spectrometer (WDS) or an energy dispersive spectrometer (EDS). With the WDS, the wavelengths of X-rays are determined using crystal diffraction and Bragg's law. The EDS detects the energies of X-rays using a lithium-drifted Si crystal.The WDS has better spectral resolution compared with the EDS, and is capable of light element detection. However, only one element at a time can be determined with a WDS. The characteristic peaks of all elements with an atomic number of 9 (fluorine) or greater can be detected simultaneously with the standard EDS. Light element detection is also possible with EDS systems that have windowless detectors or detectors with ultrathin windows. The texts by Goldstein et al.,[105] Lee,[106] Heinrich,[108] and Friel[109] contain extensive discussions of X-ray analysis and instrumentation.

Another extremely valuable attachment for SEM is an IA. A computer-based IA permits rapid, real-time automated characterization of SEM image features, including size measurements and chemical analyses when used in conjunction with an EDS. The system currently used in our laboratory is PC-based (personal computers) and controls the scanning of the electron beam across the specimen. A regular grid point spacing pattern is used with a preset point density to locate features. At each point, the image signal (we use the backscattered electron signal) is compared to an adjustable threshold to determine whether the beam is on a feature of interest. Once a feature is found, a narrower grid point spacing pattern is used to determine physical parameters, and subsequently an X-ray analysis of the feature is performed. For more information on IAs, the reader is referred to the review paper by Lee and Kelley,[110] and the text by Friel.[109]

3.1.2 Listing of Occupational Safety and Health Applications

SEM and EDX analysis may be used for a number of applications in occupational safety and health including the analysis of particles collected on air filter samples taken from the working environment, analysis of individual particles found in bulk dusts, and in the study of pneumoconioses. Applications of SEM to the study of pneumoconioses, including determination of lung particulate burden, particle chemistry, and respirable particle surface characterization, will be described in the subsequent sections.

3.2 Sample Preparation

There are two ways to study particles found in lung tissue. One examines the particles in situ, i.e. particles may be located by SEM (usually using the BEI) and an EDX analysis performed while the particles are still in the tissue. An alternative procedure is to extract the particles from the tissue matrix followed by EDX analysis. Sample preparation procedures and the advantages and disadvantages of these two techniques are described briefly below. A complete discussion of SEM preparation techniques is given by Shelburne et al.[111] The paper by DeNee[112] also describes various preparation techniques.

3.2.1 Tissue Preparation for In Situ Particle Analysis

The key advantage of SEM in situ analysis in the study of pneumoconioses is the opportunity to correlate the location, types, and concentration of particles found with any tissue pathology seen by light microscopy. On the other hand, in situ analysis is very time-consuming because particles must be identified and analyzed manually. The preparation of lung (or other types of) tissue for light microscopic and SEM examination generally involves formalin fixation, followed by paraffin embedding. Serial sections (5-μm thick) are cut from the paraffin block both for light microscopy and SEM analysis. The sections for SEM are mounted on a carbon substrate (we use 1-inch, or 2.5-cm, diameter carbon planchets), de-paraffinized using xylene, and air dried. The final specimens generally do not need coating for conductivity. However, if specimen charging is a problem, the specimen may be coated with carbon using a vacuum evaporator. Metal coatings should be avoided because they may interfere with subsequent EDX analyses. As can be seen in Figure 2, particles are easily located in the 5-μm section; however, they may occur in aggregates, making individual particle analyses a challenge.

3.2.2 Tissue Particle Isolation Techniques

The key advantage of isolating particles from the tissue matrix is the speed with which subsequent analyses may be performed. Once the tissue matrix is removed, a filter containing the residual particles may be prepared, and the particles analyzed using an automated SEM/EDX/IA procedure. The analysis procedure used in our laboratory is described in section 3.3.1. The primary disadvantage of the particle isolation technique is that the opportunity to relate tissue pathology to particle type, location, and concentration is lost. Two procedures most often used to extract particles from the tissue matrix, low-temperature ashing and chemical digestion, are described below.

3.2.2.1 Low-temperature Ashing In the low-temperature ashing process, an oxygen plasma is used to oxidize the organic components of the tissue matrix. A review of the application of low-temperature ashing in preparing samples for SEM analysis is given by Thomas and Hollahan.[113] The procedure we use for lung tissue consists of the following steps: 1–2 g of fresh tissue is selected and diced into 1–2-mm cubes, and freeze dried to constant weight. Approximately 0.1–0.2 g of the dry tissue is accurately weighed, placed in a clean glass vial, and ashed in a low-temperature asher at 90 W for 7 h at an oxygen pressure of 2 Torr. The ash is suspended in 50 mL of a solution of 0.05% Aerosol OT® in filtered, deionized water, and then placed in an ultrasonic bath for 15 min. One milliliter of glacial acetic acid is added to the suspension which is then made up to a final volume of 100 mL with filtered, deionized water and allowed to stand overnight. Since the particulate burdens of lungs vary considerably, preparation of samples with the proper particle loading for automated analysis is accomplished by filtering varying aliquots of the final suspension onto 25-mm diameter, 0.1-μm pore size polycarbonate (Nuclepore®) filters. The filters are attached to carbon planchettes with colloidal graphite, and may be examined uncoated under SEM. Filters with proper loading have minimal particle overlap and generally contain approximately 50–100 particles per field of view at a magnification of 1000X. It should be noted that some particles present originally as aggregates in the tissue may not be broken apart during the low-temperature ashing procedure.[114]

3.2.2.2 Chemical Digestion Biological tissues may be chemically digested using a number of different agents including oxidizing agents (sodium hypochlorite), strong alkaline solutions (sodium or potassium hydroxide), and proteolytic enzymes (ficin). Digestion with each of these agents is reviewed briefly in the paper by Mastin et al.[115] In our laboratory, we have used an adaptation of the sodium hypochlorite digestion procedure described by

Coin et al.[116] to prepare lung tissue for particle analysis. Approximately 0.1–0.5 g of wet, formalin-fixed tissue is accurately weighed and placed in a clean glass vial. A second piece of wet tissue is also weighed and then dried to constant weight to determine the wet-to-dry weight ratio. Ten milliliters of triple-filtered sodium hypochlorite are added to the glass vial containing the wet tissue, and the vial is then shaken gently for 30–45 min. The digestate is filtered onto a 0.2-μm pore size polycarbonate filter and washed with 10 mL of deionized water. Lipids and other undigested debris are then extracted from the filtrate by two treatments with the following sequence of triple-filtered reagents: 10 mL of isopropanol, 10 mL of 7% oxalic acid, 10 mL of Clorox®, and 10 mL of deionized water. The filter may now be dried and processed as before for SEM analysis. We have found that this digestion procedure takes less time to prepare samples than low-temperature ashing. There are a couple of negatives, however. The washing steps may lead to an uneven redistribution of the particles on the filter. In addition, achieving the proper particle loading on the filter may require some experimentation with the initial sample wet weight or separation of the initial digestate into aliquots of varying volumes.

3.3 Lung Particle Analysis

Particulate matter has been recognized as a cause of various lung diseases, the pneumoconioses, for many years. Specific pneumoconioses such as coalworker's pneumoconiosis, asbestosis, and silicosis may result from inhalation of particles at the worksite. Numerous methods, as reviewed by Mastin et al.,[115] have been used to analyze particulate matter deposited in human lungs. SEM and EDX have been used extensively to characterize particles found in lung tissue (see reviews by Shelburne et al.[111] and Baker et al.[117]). While it is not possible to compare a measured lung dust burden to the actual exposure because of the lack of exposure data and to particle clearance, the measured lung particulate burden does represent retained particle dose.

Everyone has a background lung particulate burden resulting from inhalation of respirable particles present in the ambient environment. Hence, a good database for the particle contents of the lungs of subjects with no pneumoconioses and/or history of occupational exposures, i.e. ''normal'' lungs, is needed so that the particle analysis data from diseased lungs can be properly interpreted. The application of SEM and EDX to the determination of inorganic particulate burdens for normal and diseased lungs, as well as respirable particle surface characterization, will be discussed in the next sections.

3.3.1 Normal Lung Inorganic Particulate Burdens

We have used an automated SEM/EDX/IA method to determine the lung nonfibrous, inorganic particulate burdens for a large set of subjects with no overt pneumoconioses. This procedure is not suitable for fiber analyses because mineral fibers usually have very small diameters and do not produce a BEI with enough contrast for automated analysis. In addition, organic particles, such as coal, are destroyed in the low-temperature ashing process. Complete details of the automated method for nonfibrous particles are given in the paper by Stettler et al.[118] Briefly, a small portion of freeze-dried lung homogenate from each subject was ashed in a low-temperature asher and filters containing the residue prepared as described earlier in this report. The filter samples were analyzed in the electron miscroscope which was equipped with an EDS and an IA using the BEI image. After a field of view is selected, particles in the field are automatically detected by the IA, sized, and then analyzed for 31 elements using an X-ray spectrum acquire time of 5 s. In our procedure, a minimum of 1000 exogenous particles in a minimum of 20 randomly selected fields of view at a magnification of 1000X are analyzed. The IA grid point spacing density used in the procedure is set to find particles 0.2 μm in diameter and larger. After X-ray analysis, the particles are classified by the IA using a chemistry definition file which defines particle types by their major elemental components and net fractional X-ray intensities. The list of elements analyzed and a description of the chemistry definition file are given in the book chapter by Stettler et al.[119] More than 145 000 individual particles were analyzed in this study. The average (±standard deviation) exogenous particle concentration found in 87 lungs was $476 \pm 380 \times 10^6$ particles per gram (ppg) of dry lung with a range of $71–1860 \times 10^6$ ppg. On average, various aluminum silicates accounted for 38.1% and particles classified as silica accounted for 21.0% of the particles in the 87 lungs. Large numbers of various metal-containing particles, primarily titanium and iron occurring singly or in combination with Si, were also found in the lungs. The particle levels seen in the study by Stettler et al.[119] were similar to those seen in other studies of normal lungs by Churg and Wiggs[120,121] and by Paoletti et al.[122] Analytical transmission electron microscopy methods were used in the other studies with average particle concentrations in the range $180–465 \times 10^6$ ppg.

3.3.2 Inorganic Particulate Burdens of Lungs with Pneumoconioses

Although the automated SEM/EDX/IA method described above has been used to determine the particulate burdens in lungs with suspected occupational exposures

by Stettler et al.,[114] many studies have involved manual in situ analyses using 5-μm thick sections of lung tissue. Of particular note is the work of Abraham et al.[123–125] who have developed and used an in situ procedure to determine the nonfibrous, inorganic particulate burdens of over 400 lungs. Their procedure uses 5-μm thick sections mounted on carbon. Complete descriptions of their SEM analysis procedure are given elsewhere.[124,125] Briefly, a morphometric point counting approach using both the SEI and BEI at a magnification of 6000X is used to locate particles in randomly selected fields. The number of particles in 100 consecutive fields are counted. If fewer than 100 particles are counted, additional fields are analyzed. The particles are sized and analyzed using a 20-s X-ray analysis. Size and elemental data in the form of net X-ray counts per second for each element observed are recorded. EDX data are used to sort exogenous particles into three major classes: silica (showing only a silicon peak), silicates (showing silicon and other cations such as Al, Mg, K, Ca, Fe), and other (mostly metals, either singly or in combinations). Particle concentration data are reported in terms of particles per cubic centimeter. The average total exogenous particle concentration for 433 cases by this technique was $473 \pm 113 \times 10^6$ particles cm^{-3} with concentrations in the range $1–33\,450 \times 10^6$ particles cm^{-3}.[123] The major types of particles found included silica, aluminum silicates, metals, and talc.

3.3.3 Respirable Particle Surface Characterization

Bulk compositional analyses of respirable particulate material, and even particle-by-particle compositional analyses, may not always be sufficient to predict the level of biological activity of respirable particles or the health risks of exposure to them. Toxicants may be located on the particle surface or may be more heavily concentrated there, resulting in heightened biological availability and expression of toxic activity. Or, in some cases, surface coating materials may have a prophylactic effect on expression of the toxicity of the underlying bulk of the particle. An example of the latter effect is the modified activity of the quartz component of some mixed composition mineral dust exposures for causing mixed dust pneumoconioses by LeBouffant et al.,[126] Kriegseis and Scharmann,[127] and Harrison et al.[128]

3.3.3.1 Multiple-voltage Scanning Electron Microscopy with X-ray Spectroscopy: Experimental SEM with EDS or WDS X-ray spectroscopy can provide some information on the elemental composition with depth into a particle, by acquiring X-ray spectra at two or more electron beam accelerating voltages.[129] The depth of penetration of electrons and of the excitation of characteristic X-rays is dependent on the incident electron energy and on the density of the particle target. For densities in the range of most respirable particulate minerals, a 20–30 keV electron beam will excite spectra to a depth of the order of one to a few micrometers. This range is a function of electron voltage such that a 5-keV electron will excite spectra to a depth of the order of 0.1–0.01 μm in such materials. Thus, comparing spectra acquired at 20 keV and at 5 keV can provide some indication of the homogeneity or heterogeneity of elemental distribution with depth into a particle. This method has been used to detect thin submicrometer aluminosilicate coatings or clay "surface occlusion" on respirable quartz particles.[130] To control for correction factors, comparison is not made of the absolute values of elemental spectral line intensities; instead, the ratio of spectral line intensities of different elements in a particle is measured at one voltage and compared with the ratio measured at another excitation voltage. For example, the ratio of the silicon line intensity to the aluminum line intensity of a particle is measured at 20-keV excitation and then is measured at 5-keV excitation, or is measured at a series of voltages, e.g. 20, 11, 9, 7, 5 keV. The beam generally must be relocated on the particle of interest after a voltage change. This can be done in an automated mode with modern SEM/EDS computer-controlled systems. Using such systems, several fields of particles can be analyzed for several hundreds of particles at one voltage, and the same fields can be reanalyzed at a second voltage. Typically, some smaller particles identified at one voltage in computer-controlled data acquisition will be missed at the other voltage. But visual editing of the resultant files can easily assure proper matching of particles for comparison of the ratio of elemental line intensities between the two voltages. Obviously, the electron beam voltage must be greater than the energy (frequency) of the X-ray lines used in the analysis. This can be a problem for analyses at the low electron beam accelerating voltages: 5-keV electrons will not excite the K series lines for heavier elements. In some cases this can be circumvented by using the higher series lines (L, M, ...) which occur below 2 keV.

3.3.3.2 Multiple-voltage Scanning Electron Microscopy with X-ray Spectroscopy: Interpretation Change in measured spectral line intensity ratios versus electron beam accelerating voltage can be predicted as a function of particle density and size and structure with an analytical formula.[130] Diminution of the electron beam intensity with depth of penetration into a material can be modeled in one dimension as a Beer's law process, i.e. differential loss of intensity is proportional to intensity in a differential thickness of material.[131] Empirical relationships are available for this proportionality constant as a function

of electron voltage.[132] In the differential thickness of material the electron beam will stimulate characteristic X-rays. An X-ray line intensity is a function of the intensity of the electron beam with adequate energy to stimulate the given X-ray line, of the concentration of target atoms, of the cross-section for ionization of the target atom shell by those electrons, and of the fluorescence yield or probability that the electron-excited atom will decay by the specific X-ray emission. A simple empirically determined function provides a model for the effective scattering ionization cross-section versus incident electron energy and X-ray spectral line frequency.[133] Tabulated values of fluorescence yield are available.[134] These together can be used to describe the generation of X-rays with depth by the electron beam. Then a Beer's law model can describe extinction of the generated X-ray as it leaves the particle.[134] Together these models of electron stimulation of X-rays and of X-ray emission provide a differential equation for X-ray line intensity as a function of material thickness, material density, electron beam voltage, X-ray spectral line frequency, and fluorescence yields. This differential equation can be solved across the particle boundaries for alternative models of particle structure with depth, e.g. for a homogeneous mixed composition particle versus a heterogeneously structured particle. This provides a prediction of the change in measured elemental composition ratios versus electron beam accelerating voltage. These predictions then can be used to interpret measurements of X-ray elemental spectral line intensity ratios of a particle versus electron beam accelerating voltage in terms of particle compositional structure with depth. An example in the literature details application of the model to aluminosilicate-contaminated silica particles.[135] Predictive models for a silica particle homogeneously contaminated with aluminum and for a silica core particle with an aluminosilicate clay coating were compared with experimental data of silicon to aluminum line intensity ratios measured at 20–5-keV electron beam accelerating voltages. As an example of ''occluded'' particle behavior, for a 2-μm thick particle of 97% Si/(Si + Al) overall composition, the measured fraction of silicon drops to below 80% at 5 keV, in concert with the prediction for a clay-coated particle.[136]

3.3.3.3 Scanning Electron Microscopy with Scanning Auger Spectroscopy for Particle Surface Analyses

Scanning Auger spectroscopy uses electron imaging and excitation of particulate samples much as SEM does. However, electron-excited emission of Auger electrons rather than characteristic X-rays provides the mechanism for elemental analysis. The shallow escape depth of these Auger electrons results in information being obtained only from the depth of a few atomic layers into the sample.[137] This provides a very near surface analysis compared to that provided by SEM/EDS, essentially measuring elemental composition in a nanometer thickness compared to a micrometer thickness. Scanning Auger spectroscopy also provides a more shallow depth of analysis by one to two orders of magnitude than is reasonably achievable by multiple-voltage SEM/EDS, that is, 1 nm depth by Auger versus 10–100 nm minimum depth by multiple-voltage SEM/EDS. To some extent, sample preparation and purity constraints are alleviated in the multiple-voltage SEM method: incidental surface contamination on samples may need to be removed prior to Auger analysis by argon ion bombardment of the sample. For analysis of particles in lung tissue thin sections, scanning Auger spectroscopy is not appropriate, while multiple-voltage SEM/EDS can provide some identification of surface occluded mineral particles in 5-μm thick tissue sections. For dust samples, a combination of conventional (20–30-keV) SEM/EDS, of multiple-voltage SEM/EDS, and of scanning Auger spectroscopy provides the possibility of analyzing respirable particle samples for elemental composition from their bulk to their surface. Performing these joint analyses on explicit particles is challenging but possible.[138] Modern automated methods provide the capability for performance of that suite of analyses on a large population of respirable-sized particles with the same specific particles analyzed in all three regimens. This has been demonstrated for the combination of SEM/EDS and scanning Auger spectroscopy for cobalt, chromium, and tungsten analyses of respirable hard metal particles.[139]

4 ATOMIC SPECTROMETRY

4.1 Introduction

Atomic spectrometric techniques are the most widely used analytical techniques for quantitative analysis of inorganic metals. The techniques have been applicable to nearly all the elements, including most of the metals and semimetals.[140] The development of ICPAES techniques has been increasing. ICPAES offers a simultaneous or rapid sequential multielement determination capability at the major, minor, and certain trace concentration levels. It has become established as a widely accepted method for the analysis of metallic aerosols, powders of metals, dusts, and fly ashes. To date, a vast number of published atomic spectrometric methods are available, covering a wide range of applications. Atomic spectrometric methods are widely employed for occupational health evaluation, owing to the presence of trace elements in the atmosphere released from various industrial processes.

4.2 Flame Atomic Absorption Spectrometry

4.2.1 Principles and Instrumentation

Flame atomic absorption spectrometry (FAAS) has been used for the determination of about 65 elements with detection limits ranging from a few parts per billion to a few parts per million.[141] Instruments are reliable, robust and simple to use. It is a single channel instrument, but operating conditions can be changed rapidly so that several elements can be determined in one sample in an automated sequential run. Two types of flame mixtures are commonly used: air–acetylene and nitrous oxide–acetylene. The latter flame is significantly hotter and is needed to atomize some elements such as aluminum, which form refractory oxides. The sample solution is pneumatically nebulized into a spray chamber where a sample mist is formed and mixed with a fuel gas. Then, the sample aerosol enters a flame where dissociation and atomization occur. During the rather limited residence time in the flame, the droplets are dried and the resulting salt particles vaporized. The resulting molecular species may be atomized by thermolysis or by chemical reaction with reducing species such as carbon and carbon monoxide. Generally, FAAS is selective, rapid, and amenable to automation with adequate sensitivity.

4.2.2 Methodology

4.2.2.1 Sample Preparation In a typical FAAS assay, an analyte is dissolved into a solution. The sample preparation procedure depends strongly on the properties of both the target elements and the solvent matrices. The principal objective of sample preparation is to dissolve target analytes, and to remove interferences. The method of liquid–liquid extraction is a convenient way of preparing some samples in which the target analytes are easily dissolved in the solution. It is relatively simple, rapid, and favorable to FAAS. Organic solvent can also be used to enhance detection sensitivity. Methyl isobutyl ketone is the most popular solvent because of its extraction and nebulization efficiency as well as combustibility. Other solvents such as ketones or esters can be used as well. In the pulse-nebulization mode of FAAS, a number of elements can be successively determined in an extract. Another alternative sample preparation is to use microwave heating as a source of intense energy for rapid mineralization of liquid and solid samples. The reduction in the digestion time and the higher reaction speed may be due to the fact that the energy transfer is improved, and the microwave field has a specific chemical influence on organic molecules in acidic media. Generally, a few sources of systematic error should be considered. These include contamination, losses of trace elements, volatilization, and physical/chemical transformations of the samples.

4.2.2.2 Flow Injection Analysis FIA has nowadays become a powerful analytical tool for sample preparation and introduction before measurement. It offers a convenient and fast approach to enhance and automate preliminary steps for atomic spectrometric detectors. Moreover, flow manifolds can ease the well-known problem of sample introduction to atomizers or even expand the classical scope of atomic/elemental information. Flow injection strategies with atomic spectrometric detectors are used in research and analytical laboratories. For detailed information, the reader is referred to a book by Sanz-Medel.[142]

4.2.2.3 Interferences In FAAS assay, minor ionization interferences may be encountered with Rb and Li in an air–acetylene flame and additional elements such as Al, Ba, Be, Ga, and Si in a nitrous oxide–acetylene flame.[143] The interferences can be suppressed by adding ionization buffers. Background absorption is not a particular problem and is compensated for by using a deuterium background corrector. Sometimes, specific solute–volatization interferences are observed in the determination of B, Ba, Cr, Mo, Pt, and Sn. They can be overcome by adding suitable spectrochemical buffers. Transport interferences are encountered with viscous sample solutions. Various nonspecific matrix effects are observed with nebulizing solutions of high salt and/or acid content. In such cases, acid-matched calibration and standard addition checks are advisable.

4.2.3 Applications

FAAS is a rather selective instrument for many metals. It has been used in a number of studies associated with occupational and environmental health. In a field investigation carried out at a North American nickel alloy production facility, the levels of worker exposures to inhalable and total nickel-containing aerosol during nickel alloy production were studied by Tsai et al.[144] Worker exposures in a range of workplaces throughout the facility were assessed. The results showed that inhalable aerosol exposure levels for both overall aerosol and for total nickel were consistently and significantly higher than the corresponding total aerosol levels. In related work, Torjussen et al.[145] investigated the concentration and distribution of heavy metals in nickel-exposed workers and of controls. Biopsy specimens from 30 nickel-exposed individuals and six controls were analyzed by FAAS to determine the content of nickel, copper, cobalt, zinc and iron. Some differences in epithelial types between specimens from the nickel-exposed group and the control group were seen.

Pilger and Broder[146] reported a method that is suitable for assessing exposure to toxic metals in occupational indoor environments. The method has been evaluated for 21 metals including antimony, cadmium, chromium, cobalt, copper, lead, iron, zinc, and so on. The method is element specific. In another study, Bellido-Milla et al.[147] detected hygiene hazards involved in naval industry welding processes. The metal contents of welding fumes produced at the shipyard were investigated to assess the hygiene hazards. Personal and environmental samples were collected on cellulose or polyvinylchloride filters. Samples were analyzed to encompass every possible working condition. Quantitative metal determinations were carried out by FAAS. The results obtained for metals and particles were compared and conclusions were drawn according to the type of welding procedure, sampling place, and use of fume extractor.

Burguera et al.[148] reported the determination of lead in hair of exposed gas station workers and in unexposed adults by microwave-aided dissolution of samples and flow injection atomic absorption spectrometry (FIAAS). Lead content in head hair of 53 gas station workers together with an equal number of normal controls was determined. Samples of hair were washed with ethanol and water and were subject to microwave digestion prior to the determination of lead by FIAAS. The lead content in hair of the gas station workers ($48.7 \pm 17.5\,\mu g\,g^{-1}$) was significantly higher than that of the normal controls ($17.2 \pm 8.1\,\mu g\,g^{-1}$). The effects of washing and sample digestion procedures, head sampling site, hair color, age, smoking habits and duration of exposure to the metals were discussed. In related work, Othman[149] had reported a preliminary investigation of the lead level in whole blood of normal and occupationally exposed populations in Damascus City. Tsalev et al.[150] studied the manganese in whole blood of exposed workers and unexposed individuals in a manganese alloy plant. The purpose of this work was to elucidate the state of health of workers employed in a manganese industry.

4.3 Electrothermal-, Hydride Generation- and Cold Vapor Atomic Absorption Spectrometry

4.3.1 Principles and Instrumentation

An electrothermal atomic absorption spectrometry (ETAAS) system is equipped with a graphite tube, which is aligned in the spectrophotometer optical path, and is enclosed in an inert gas, usually argon atmosphere. The graphite tube furnace is electrically heated to preset ''dry'', ''ash'', ''atomize'', and ''clean-out'' temperatures, so that it sequentially removes the solvent, organic matter/volatile matrix constituents, volatizes/atomizes the analyte, and eventually expels matrix/analyte residues. A transient peak signal is thus produced. Although graphite tube furnaces are most popular, there are also some other atomization devices such as carbon cups, Ta ribbon, Mo microtube, and so on. Graphite furnace atomic absorption spectrometry systems provide a 50-fold to 500-fold improvement in sensitivity relative to FAAS. Generally, an important ETAAS technique is to apply it to trace element analysis of biological samples with low limit of detections (LODs), and small size of samples. A typical ETAAS analysis is at nanogram per milliliter levels.[140,141]

Hydride generation atomic absorption spectrometry (HGAAS) involves the generation of a volatile hydride of an analyte by means of a reducing agent added to a reaction vessel containing an acidic solution of the sample. The generated hydrides are transported to a heated atomizer cell, which can be a heated quartz tube or graphite tube. Sodium borohydride is most commonly used as a reducing agent. The thermally decomposed hydride produces atomic vapors that can be measured quantitatively. Since the hydride is separated from the matrix, advantages include high sensitivity and reduced interferences. However, this technique is only applicable to a limited number of elements. Those elements that form volatile hydrides include As, Se, Bi, Sb, and Te.[140,141]

Cold vapor atomic absorption spectrometry (CVAAS) analysis is used for specific determination of mercury, which can exist in an atomic state at ambient temperature owing to its high vapor pressure. In a manner similar to HGAAS, a reducing agent is added to a reaction vessel containing a sample with trace levels of ionic mercury. Stannous chloride and sodium borohydride are the most commonly used reducing agents. Other constructions as well as automated continuous flow devices have been applied to biological samples.[140,141]

4.3.2 Methodology

4.3.2.1 Sample Preparation The sample preparation for ETAAS and HGAAS measurements is not substantially different from that encountered in other atomic absorption spectrometry methods. Liquid–liquid extraction is highly suitable as a manual pretreatment procedure for HGAAS and speciation. The implementation of liquid–liquid extraction in a continuous fashion enables on-line coupling to AS instruments and contributes advantages inherent in automatic methods of analysis. A detailed technical description of continuous liquid–liquid extraction processes can be found in several books.[140–143] Most biological samples are prepared by decomposing the sample using thermal or chemical means. The microwave approach for sample preparation is becoming a powerful tool. Calibration with standard

addition and verification of modified procedures by means of certified reference materials are often used.

4.3.2.2 Interferences ETAAS and HGAAS are methods of choice for the majority of elements. However, they have some drawbacks and limitations. Matrix effects in ETAAS are common, pronounced and complex. Therefore, background correction must always be provided. An adequate calibration is then required to compensate for residual discrepancies. Matrix/analyte modification techniques are commonly used and are aimed at either decreasing the relative volatility of the analyte or increasing the volatility of the matrix, or both. HGAAS has been applied to the determination of several volatile elements such as As, Se, Sn, Cd, Cu, and so on, which are not easily determined by flame and ETAAS. This technique is relatively selective and amenable to automation. In operation, the organic matter should be completely oxidized and the analyte should be in an oxidation state. Some potential interferences may be expected owing to acids, oxidants, and ions of noble metals. Their nature and extent depend on many factors such as acidity or pH, and the presence of oxidation states of both the analyte and the interferent. Generally, the important interferences and adverse effects can be reduced or eliminated by properly optimizing the experimental procedures.

4.3.3 Applications

Numerous papers have appeared that described application of ETAAS, HGAAS, and CVAAS methods to workplace occupational exposure studies. Rollin and Nogueira[151] reported identification of aluminum fractions in serum by Zeeman atomic absorption spectrometry in order to ascertain the distribution of aluminum (Al) in normal and occupationally exposed sera. It was found that the relative distribution of Al between high molecular mass and low molecular mass fractions was statistically significantly different. This suggests that at high concentrations of total Al in serum, the percentage of the Al bound to the low molecular mass is lower, but the absolute quantity of Al circulating as the low molecular mass complex is increased. This low molecular mass Al complex is thought to play an important role in intracellular accumulation of Al. In related work, Gitelman et al.[152] measured serum Al and urinary Al/creatinine ratios in 235 Al workers and 44 controls in the Al industry to examine the association between occupational exposure to airborne Al and Al absorption. Serum and urine samples were taken before and after 3–5-day work shifts. Occupational exposure was estimated from Al measurements of respirable and total particulates in air. Median exposure values were 25 and 100 $\mu g\,m^{-3}$, respectively. These results are consistent with the systemic absorption of Al from occupational exposure, and suggest the presence of a sensitive uptake process for airway Al. In other related work, Rollin et al.[153] investigated the effect of exposure to Al on concentrations of essential metals in serum of foundry workers. The concentrations of Al in serum and urine of 33 volunteers exposed to inhalation of Al_2O_3 dust were measured. These were compared with results from 20 normal subjects not exposed. The Al concentration in serum was significantly raised in the subjects exposed to dust, but urine showed no significant difference from controls. This redistribution was selective, as the serum concentration of Cu was decreased whereas the serum concentration of Zn was increased. The serum concentration of Fe did not change significantly. Biological monitoring of occupational Al powder exposure was reported by Letzel et al.[154] Fifty-four workers from the exposed group were studied.

The measurement of salivary cadmium by ETAAS and its use as a biological indicator of occupational exposure has been reported by White et al.[155] The method has been developed and employed to measure cadmium levels in saliva samples collected by two different methods from a group of ex-workers previously exposed to cadmium, two groups of currently exposed workers, and an unexposed population as a control. Salivary cadmium levels were significantly raised in both of the groups of currently exposed individuals and in past workers with previous long-term exposure when compared with an unexposed population. In related work, Abernathy et al.[156] developed a method for measuring cellular Cd and DNA-bound Cd following micromolar exposures to cadmium dichloride. Following low-level exposure to cadmium dichloride, atomic absorption spectrometry with Zeeman background correction was used to measure total cell-associated Cd in wet-ashed cells. The lower LODs were determined to be 100 pg of Cd per 10^6 cells. This method is sensitive and reproducible, and is suitable for the detection of Cd in biological matrixes after low levels of Cd exposure. The determination of silver in whole blood and its application to biological monitoring of occupationally exposed groups were studied by Armitage et al.[157] Blood silver levels were determined in 98 occupationally exposed workers involved in bullion production, cutlery manufacture, chemical manufacture, jewellery production and silver reclamation. Other occupational applications of the ETAAS methods included study of urinary excretion of nickel in nickel–chromium electroplaters.[158]

HGAAS methods have been used in several studies of interest in occupational and environmental health. In a field investigation, Jensen et al.[159] reported the sum of concentrations of inorganic arsenic, methylarsonic acid and dimethylarsinic acid in urine from adults and children living in an unpolluted area. The results

from the unpolluted area were compared with the corresponding sum from adults and children living in an area polluted with arsenic, and the corresponding sum from persons occupationally exposed to arsenic. The median values for 22 adults and 10 children aged 3–10 years living in the unpolluted area were 9.3 and 19.8 nmol As $mmol^{-1}$ creatinine, respectively. The corresponding ranges were 3.2–27.9 and 7.7–57.8 nmol As $mmol^{-1}$ creatinine, respectively. The arsenic level in urine from adult workers handling arsenic-treated wood was approximately four-fold higher than controls. The arsenic levels in urine from two glass workers were nine- and two-fold higher, respectively. In another work, Blas et al.[160] developed a method for determination and speciation of arsenic in human urine by HGAAS. This method is applicable to urine samples in studies relating to arsenic exposure and its monitoring.

Another atomic spectrometric technique, the CVAAS method, has been used in mercury analyses. Martin et al.[161] reported a study of using spot urine samples for low-level occupational mercury exposure assessment, and demonstrated a relationship of Hg exposure with porphyrin and creatinine excretion rates. Hg and porphyrin levels in single void urine spot samples were compared with calculated 24-h urine levels in 35 practicing dentists who had been occupationally exposed to low levels of elemental Hg. The study aimed to determine the individual variability for Hg and porphyrin concentrations in spot samples over a 24-h period, and determine the time of day at which a spot sample would give an Hg concentration closest to the 24-h average concentration. The results confirmed previous reports of a first-order diurnal pattern with a mid-morning peak for Hg concentration. In other studies, the CVAAS method was also used in the determination and speciation of mercury, methylmercury, ethylmercury and phenylmercury concentrations in urine samples taken from students and staff of a dental workplace.[162]

4.4 Inductively Coupled Plasma Atomic Emission Spectrometry

4.4.1 Principles and Instrumentation

The ICPAES technique has been available commercially since the mid-1970s.[163] The technique is an emission spectroscopic method in which the sample is dissociated into its atomic form and excited to high energy levels including the ionic form by introducing the sample into the center of a gaseous plasma sustained inside an induction coil energized with a high frequency alternating current. The excited species then emit characteristic radiation as they relax back to the atomic and ionic ground states. Principally, in ICPMS analytes are atomized, excited, and ionized, and then identified by their optical spectrum.

Today ICPAES has become widely accepted, and has been applied to elemental analysis of nearly all the elements. It offers simultaneous or rapid sequential multielement determination capability at the major, minor, and trace concentration levels. Because of its simplicity of use, wide linear dynamic range, and accuracy of analysis, it has become established as the accepted method for the analysis of metallic aerosols, powders of metals, dusts, and fly ashes.

4.4.2 Methodology

4.4.2.1 Sample Preparation In multielement analysis by ICPAES, the dissolution of a sample can be quite complex. Like most classical chemical analysis and dissolution methods for atomic absorption spectrometry, the sample preparations are designed to bring the analytes into solution, without loss or gain. In general, there is no universal solvent or universal dissolution method. Knowledge of a wide range of sample types and dissolution procedures is needed to make a choice. When a large range of elements is under consideration, high temperatures can cause the loss of volatile elements, while low temperatures may result in the incomplete dissolution of refractory elements. Dry ashing methods can be used for large batches of samples, but there can be problems involving the loss of certain volatile elements. Acid digestion with either hot or cold is widely applicable to the analysis of a majority of elements. There are two groups of acids. One group includes oxidizing acids such as nitric, sulfuric and perchloric acids. Another group is nonoxidizing acids including hydrochloric, hydrobromic and hydrofluoric acids, and so on. Both groups can be used to dissolve metals, oxides, and carbonates. Microwave extraction or digestion is a safer, faster, cheaper procedure, and causes less contamination of sample prepared for trace analysis. It is increasingly replacing conventional techniques such as hotplate acid digestion. For a particular sample type, a specific method should be developed through experiments in which the acid mixture, the microwave power setting, and the heating period are varied to determine which combination gives the best results. Because ICPAES can tolerate high levels of organic material, the complete destruction of organic material is not necessary.

4.4.3 Applications

Many papers have been published that deal with the measurement of metals of interest in workplace air, or dust, and in biological monitoring by ICPAES. Lo and Arai[164] developed a rapid method for the simultaneous determination of 11 metals (As, Be, Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb and Zn) in urine by

ICPAES. Acidification of the urine was the only sample preparation required. This procedure has been applied for routine screening of workers for occupational exposure to toxic metals. Biological exposure evaluation and hair analysis in workers handling chromium compounds were reported by Kudo et al.[165] The hair of 40 workers exposed to chromium compounds was analyzed for 18 elements by ICPAES; 21 subjects worked in a factory that manufactured chromate pigments, 11 in a painting factory and 8 in a Cr plating factory. It was found that simultaneous measurements of Cr in the hair and urine were useful for determining the extent of exposure to Cr. Determination of multiple elements should be the best means of estimating the degree of exposure to Cr compounds in which they are found.

Olson et al.[166] described trace element analysis of As, Be, Cd, Cr, Cu, Fe, Pb, V, and Zn in airborne particulates and in human urine by ICPAES. The principle of the method, range and sensitivity, interferences, accuracy and precision, advantages and disadvantages are detailed in their paper. Franzlau[167] used ICPAES in screening for trace metal exposures in an industrial population. A heterogeneous group of asymptomatic industrial workers were examined and had hair and blood samples analyzed for 10 metals via the ICPAES technique. Hull[168] employed ICPAES for the multielement analysis of industrial hygiene samples. Experiments were conducted to define the lower limit of quantitative determination and the analytical range of each element for which an OSHA (occupational safety and health administration) personal exposure limit exists. The effects of varying solution matrices and interelement effects were investigated. Cadmium emissions from fumes were studied during processing of Cd-containing thermoplastics processing. Air sampling volumes of 120–388 L were collected in the door area, at the machine nozzle and mold vent, and 1 m away from the machine approximating the position of the operator, and analyzed for Cd by ICPAES.

5 CHROMATOGRAPHIC SPECTROPHOTOMETRIC DETECTORS

5.1 Introduction

Spectrophotometric detectors coupled to chromatographic separation techniques are often used to characterize the workplace environment. These techniques are employed[169] to confirm the accuracy of direct reading instruments or when direct reading instruments are not available or when the workplace environment is complex and contains multiple chemicals that need to be measured for regulatory documentation of exposure levels[170] and for biological monitoring.[171] Standardized methods are published by a variety of government agencies and societies, including the National Institute of Occupational Safety and Health (NIOSH), OSHA, Mine Safety and Health Administration (MSHA), Environmental Protection Administration (EPA), and American Society for Testing and Materials (ASTM). Many of these methods can now be obtained through the Internet. The NIOSH Manual of Analytical Methods can be found at www.cdc.gov/niosh/nmam/nmampub.html. This site also contains links to the other organizations mentioned. Harper et al.[169] have reviewed the literature published in the field of industrial hygiene. Many references to specific air monitoring and biological techniques are sited within this review. The following is a short overview of chromatographic spectrometric detectors, emphasizing UV/VIS and fluorescence detectors.

5.2 Spectrophotometric Detectors for High-performance Liquid Chromatography

Spectrophotometric detectors that can be coupled to HPLC include UV/VIS, fluorescent, light scattering, refractive index, diode array, MSD, IR detectors, radioactivity and luminescence. The choice of detector is dependent on the analyte's spectral properties, and required sensitivity and selectivity. Both the sensitivity and selectivity of spectroscopic detectors are dependent, in part, on wavelength(s) employed. The wavelength where maximum absorption occurs for the chemical of interest may not necessarily be used to quantify levels in a particular occupational environment if chromatographically coeluting contaminants absorb at that wavelength. The settings may be changed slightly or moved to a secondary absorbance peak of the analyte to eliminate the interference. Wavelength selection can also be limited by mobile-phase requirements for chromatographic separation of the environmental components of interest. Water, methanol and acetonitrile are mobile phases used in reversed-phase HPLC with ultraviolet cutoff points of $\leq$210 nm. Choice of detector can also be influenced by the linearity and dynamic range (concentration versus detector response) requirements. It is desirable to choose a detector with a wide linear working range to be able to assess both major and minor chemical components in the workplace environment. Comparison of detector sensitivities for a particular analyte can be made by using the LOD, which is reported as the amount of analyte that provides a signal that is two to three times that of background noise.

The most commonly employed HPLC detectors in monitoring the workplace environment are the UV/VIS photometric and fluorescent detectors. UV/VIS detectors are available as fixed, variable and scanning-wavelength detectors. Diode array detectors, which

Table 4 NIOSH Manual of Analytical Methods: HPLC–spectroscopic method[a,b]

Method number	Chemical/chemical class	Detector	Wavelength(s) (nm)	LOD/sample
3507	Acetaldehyde	UV	254	0.1 mg
2514	Anisidine	UV	254	0.35 μg
5031	Aspartame	UV	220	2 μg
5509	Benzidine, 3′,3′-dichlorobenicidine	UV	254	0.05 μg
5509	Benzoyl peroxide	UV	254	10 μg
5510	Bromoxynil, bromoxynil octanoate	UV	254	0.6, 0.3 μg
2014	*p*-Chlorophenol	UV	280	2.5 μg
5001	2,4-D and 2,4,5-T	UV	284	15 μg
5013	Dyes (*o*-dianisidine, *o*-tolidine, benzidine)	UV	280	3 μg
2540	Ethylenediamine	UV	254	0.9 μg
	Diethylenetriamine			0.16 μg
	Triethylenetetramine			0.3 μg
5700	Formaldehyde	UV	365	0.08 μg
2532	Glutaraldehyde	UV	365	0.3 μg
5004	Hydroquinone	UV	290	10 μg
5522	Isocyanate	Fluor	275/320	0.1–0.3 μg
3512	Maleic anhydride	UV	254	15 μg
5029	4,4′-Methylenedianiline	UV	254	0.12 μg
5033	*p*-Nitroaniline	UV	375	20 μg
5003	Paraquat	UV	254	10 μg
5512	Pentachlorophenol	UV	254	8 μg
5032	Pentamidine isethionate	Fluor	270/430	18 ng
5506	Polyaromatic hydrocarbons	UV/Fluor	254; 340/425	0.1–0.8 μg
5008	Pyrethium	UV	225	10 μg
5007	Rotenone	UV	290	4 μg
5016	Strychnine	UV	254	0.8 μg
5005	Thiram	UV	254	5 μg
5516	2,4- and 2,6-Toluene diamine	UV	229	0.1 μg
2535	Toluene-2,4-diisocyanate	UV	254	0.1 μg
5018	2,4,7-Trinitrofluoren-9-one	UV	280	0.04 μg
5002	Warfarin	UV	280	2.5 μg

[a] UV, ultraviolet; fluor, fluorescent; 2,4-D, 2,4-dichlorophenoxyacetic acid; 2,4,5-T, 2,4,5-trichlorophenoxyacetic acid.
[b] Fluorescent wavelengths are given for excitation/emission.

provide simultaneous spectral data on each peak eluted from the column, are increasingly being employed. Sensitivity of UV/VIS detectors is limited by the detector flow cell path with respect to volume restrictions. In general, the detector cell volume for standard column HPLC should be no greater than 8 μL cm^{-1} of the optical path length. Sensitivity for an analyte can be enhanced by conjugation of chemical functional groups to specific chromophores.

Fluorimetric detectors are considered, in general, to provide better selectivity and sensitivity than UV/VIS detectors, although both UV/VIS and fluorimetric detectors are available for HPLC detection. The emitted light is usually measured at right angles to the excitation source. The enhanced sensitivity in fluorescence techniques over absorption detectors is mainly due to the reduction in noise caused by background light. Flow cells for fluorimetric detectors must optimize excitation and emission collection efficiency for the small volume of the cell. This requires use of a high-intensity excitation light that is usually provided by mercury or xenon arc lamps. Lasers are becoming more commonly employed as the excitation source. They can focus a high intensity beam onto a small capillary cell and the intensity of the source does not decay with age as seen with traditional excitation lamps. Lasers that emit light in the lower ultraviolet ranges, at present, are very costly, which has limited the spectrum of analytes and conjugates that have been employed with this technique. Secondary excitation/emission wavelengths can be used if interferences are encountered and derivatizing agents can be conjugated to a nonfluorescent analyte to allow for detection. This can improve the LOD for an analyte from 1 to 3 orders of magnitude over UV/VIS detection.

Table 4 is a list of HPLC/ultraviolet and fluorescent methods from the NIOSH Manual of Analytical Methods (4th edition) for chemical analysis of air samples. The majority of the methods take advantage of UV and/or fluorescent spectral properties of the analyte. The 254-nm wavelength is standard on fixed wavelenth detectors, because of the broad range of chemicals that absorb at that wavelength. Many methods in the past

have been developed using 254 nm as the detection wavelength because of this. Derivatization to stabilize and/or enhance spectral properties is used in several of the methods. Naphthylisothiourea derivatives are employed for the measurement of ethylenediamine, diethylenetriamine, and triethylenetetramine (method number 2540), and 2-4-dinitrophenyldrazone formaldehyde conjugation in method number 5700. Isocyanates are derivatized to a stable form for analysis by either ultraviolet or fluorescent detectors. Newer reagents have been developed for trapping and detection of isocyanates, such as 1-(9-anthracenylmethyl)piperazine, 9-(methylaminomethyl)anthracene, 1-(2-methoxyphenyl)-piperazine and dibutylamine, because of the potential exposure to multiple polymeric forms. These reagents employ UV, fluorescent and/or MSDs to give a fingerprint of the multiple polymer isocyanate exposure.

5.3 Spectrophotometric Detectors for Gas Chromatography

Spectrophotometric detectors for GC include MSD, IR, FPDs, and chemiluminescence–redox detectors (CRDs). The majority of the GC applications for industrial hygiene are based on flame ionization detectors, however, the MSD is becoming increasingly common for both air and biological marker analysis. Flame photometric detectors (FPDs) are highly selective for both phosphorus and sulfur. Phosphorus-containing compounds are detected at 510 and 526 nm. Sulfur species emits light at around 394 nm. Applications for the FPD include analysis of pesticides and sulfur-containing gases and fuels.

FTIR is often used to complement the MSD and can provide functional group data and an IR fingerprint to help identify unknown components in the workplace environment or a chemical metabolite following exposure. The traditional light pipe FTIR instrument is relatively insensitive. This has limited its application to mainly qualitative analysis, e.g. identification of trichothecene mycotoxins in bulk samples. One commercially available FTIR instrument has employed direct deposition of the analyte onto a moving ZnSe window to improve detectability into the picogram (of dodecane) range. This increase in sensitivity should allow for the development of applications for qualitative and quantitative analysis of the workplace environment through both air and biological sampling.

CRDs are useful for compounds that are poorly detected by FPDs. The CRD detects the resultant chemiluminescence following specific redox reactions of compounds such as ammonia, sulfur dioxide, and thiols. A variety of other speciality spectroscopic detectors are available, including pyrochemiluminescent nitrogen and pyrofluorescent sulfur detectors that have been used for specific types of analyses. These detectors convert nitrogen-containing compounds to nitric oxide (NO) and sulfur-containing compounds to sulfur dioxide (SO_2). The NO is reacted with ozone to form the excited state of NO_2^*, which releases a photon of light.

6 INFRARED AND RAMAN SPECTROSCOPIES

6.1 Introduction

Emissions of hazardous chemical pollutants are concerns in both the environment and the workplace for safety reasons.[172,173] Identification of chemical pollutants in these places can be accomplished by various techniques.[174] IR spectroscopy, especially FTIR, and RS are the techniques of choice for the identification of the majority of chemical pollutants. Many pollutants have chemical groups of toxicological importance that can be identified and quantified by IR or RS, which measures the vibrational excitation of atoms around the bonds that connect them.

Although IR and Raman spectroscopies are similar in that both techniques provide information on vibrational frequencies, there are many differences between the two techniques. Some vibrations are only Raman active while others are only IR active; the vibration of a heteropolar diatomic molecule is IR active, whereas that of a homopolar diatomic molecule is not IR active. In molecules having a center of symmetry, a vibration is IR active, Raman active, or active in both; however, totally symmetric vibrations are always Raman active.

In IR and Raman, the spectra range from 4000 to 50 cm^{-1}. In both vibrational spectroscopies, the small quantity of samples needed, the speed with which a spectrum can be obtained, and the wide applicability of the methods combine to make IR and Raman spectroscopies two of the most useful tools available to the chemist, industrial hygiene chemist, and industrial hygienist. Both spectroscopies have a great potential because of their applicability to many different fields, such as in structural chemistry,[175,176] biology,[177] biochemistry,[178,179] medicine,[180] and industry.[181] In the following, some specific applications of FTIR and Raman spectroscopies in occupational and environmental hygiene chemistry are presented.

6.2 Infrared Spectroscopy

6.2.1 Fourier Transform Infrared Spectroscopy for Monitoring Airborne Gas and Vapor Contaminants

With the increased use of highly toxic gas and vapor mixtures in science and industry, the need has come for more

sensitive and versatile air monitoring technologies. Of the instruments that are appropriate for industrial hygiene monitoring of gases and vapors cited herein,[182–184] only FTIR spectroscopy is presented in the following discussion.

In the 1990s, a number of IR instruments have been developed, including filter, optical null and ratio recording grating, Fourier transform, and tune laser diode spectrophotometers.[185] FTIR has been shown to be particularly valuable in the monitoring of airborne gas and vapor contaminants.[186] The earliest applications of IR spectroscopy for the identification and quantification of pollutants in air were made using a long-path gas cell. Stephens et al. employed a prism spectrometer (with thermocouple detectors) to measure atmospheric chemistry related to smog pollution.[187] In later studies with an FTIR long-path system, Tuazon et al. studied trace pollutants in ambient air and synthetic atmosphere.[188] Herget and Levine[189] used an FTIR spectrometer with a 20-m gas cell as a near real-time monitoring method for semiconductor process gas emissions. Some other research has been carried out using FTIR for measurement of mobile source emissions,[190] polluted urban air,[191] and reactivity of hydrocarbons.[192] Many studies have utilized remote sensing FTIR to provide reliable air pollution information.[193,194] In remote systems, the gas cells are replaced by telescopes that are used to collimate, send, and receive IR light. With an interferometer system, which employed telescopic optics, the remote sensing FTIR has been used to measure the concentration of gaseous pollutants ranging from 10 ppm to 1 ppb over a 1-km path length.[195]

An open-path FTIR system has been developed to monitor gas and vapor contaminants quantitatively in the workplace.[196] This system was tested in the workplace to monitor numerous organic vapors, such as ammonia, methanol, methylene chloride, and sulfur hexafluoride.[197] An open-path FTIR system has also been applied to air pollution monitoring of trace gases in ambient air.[198] The ability of open-path FTIR instruments to monitor multiple compounds in real time makes it an ideal industrial air monitoring instrument.

6.2.2 On-filter Analysis of Quartz in Respirable Coal Dust Using Fourier Transform Infrared

The method for determining quartz content in respirable coal dust is based on dispersive IR spectroscopy.[199] This method has been adequate for many years with a working range capable of measuring 25–250 μg of quartz. A development of the FTIR method is that 10 μg of quartz can be detected in coal dust samples.[200] FTIR employs an interferometer to obtain information about the transmission of IR energy of all wavelengths emitted by the source and passing through the sample. Since there are no entrance or exit slits in FTIR, a greater amount of energy reaches the detector, resulting in increased sensitivity. The interferometer in FTIR contains a fixed mirror and a moving mirror, and the laser tracking of the moving mirror results in greater precision of the wavelength measurement. This permits multiple scans to be averaged, so the signal-to-noise ratio of the FTIR spectrum increases. The quantitative FTIR analysis of quartz in respirable coal dust samples collected on filters would provide the analyst with many benefits, such as speed, convenience, and productivity.[201]

Min-U-Sil 5 (US Silica Co.) quartz dust (pure quartz particles, 5 μm or smaller) is used as the standard for FTIR analysis.[199,202] Polyvinyl chloride or acrylonitrile copolymer filters used for this study have a pore size of 5 μm or less. The filter diameter must be appropriate to the sampler used. The dust samples are collected on the filters. After ashing the membrane filter, the dust samples are dispersed into isopropyl alcohol. Details of the sample treatments and filtration procedure are described by Tuchman.[200] Calibration samples, using Standard Reference Material,[203] are prepared over the range of 2–20 μg.[200] The filter samples are placed in the filter holder to be analyzed by FTIR. The routines for determining quartz using FTIR are the same as described by Tuchman.[200] The FTIR bands of the quartz spectrum have been reported at 1087, 799, 780, 695, 524, and 467 cm^{-1}.[200] These band intensities as shown in the quartz absorption spectrum[200] occur closer to the far-IR region and are not obscured by random noise. However, the absorption band at 1087 cm^{-1} is strong but very broad while the 695 cm^{-1} is weak; therefore, these two bands are problematic for quantifying quartz. The remaining four bands at 799, 780, 524, and 467 cm^{-1} are suitable for quantifying quartz.

6.2.3 Studies on Structural Changes of Collagen in Silicosis Using Fourier Transform Infrared

Silicosis is a well-known occupational disease that continues to spread even after exposure to dust stops, and there is no effective treatment. Therefore, prevention of exposure to dust pollutants and early detection in the work environment are required. Although studies on structural changes of collagen in silicosis can be accomplished by several techniques,[204,205] only studies on structural changes of silicotic collagen using FTIR are presented as follows. Test samples in the lung, lymph node, and other various tissues of normal and silicosis-affected humans or animals are used for FTIR studies.[206] The tissues are fixed in KBr as described by Yurui et al.[206] The wavenumber of FTIR spectroscopy

ranges from 4000 to 400 cm^{-1}. The bands of the normal spectrum have been reported at 3400–3200, 2900–1200, and 1200–1000 cm^{-1}.[206] These band intensities as shown in the normal spectrum[206] are broad, and their intensities are medium. The band positions and relative intensities of the differential spectrum between normal and silicosis spectra are compared to those of the normal spectrum. It has been found that there are higher band intensities in the differential spectrum at 3400–3200, 2900–1200, and 1200–1000 cm^{-1}.[206] The higher band intensities at 3400–3200 and 1200–1000 cm^{-1} correspond to increases in −OH and −Si−O−R groups,[207] respectively, while the higher band intensity at 2900–1200 cm^{-1} indicates the shortening of the −C−C− backbone.[206] The increase in −Si−O−R groups indicates that silica forms linking bridges between collagen which may be the cause of progressive enlargement of nodules.

6.3 Raman Spectroscopy

Like IR, RS is a powerful technique and has a variety of applications. IR and Raman spectroscopies are complementary, and both are utilized whenever possible. Since experimental techniques and applications of RS have been reviewed extensively, only microscopic inclusions caused by inhaled particles (including talc, rutile, α-quartz, and calcite) which have been identified in situ by Raman microspectroscopy are covered. Other experimental techniques and their applications should be referred to: nonlinear RS,[208] time-resolved RS,[209] matrix-isolation RS,[210] high-pressure RS,[211] Fourier transform RS,[212] surface-enhanced RS,[213] and Raman spectroelectrochemistry.[214]

6.3.1 Experimental

The sample being studied is placed on the stage of the microscope and illuminated by light from the transmission illuminator. The sample is focused and adjusted by viewing from the optical viewpoint. Then, the illuminator lamp is switched off and the laser beam is passed through filtering optics. The scattered light from the sample is collected by the objective and sent into the spectrometer. A cooled photomultiplier detector is used for detection.

6.3.2 Identification of Inhaled Particles, Including Talc, Rutile, α-Quartz, and Calcite

6.3.2.1 Identification of Talc A pure talc particle (talcum powder), used as a reference, and test particle samples in the foreign bodies in various tissues[215] that contain a talc particle (both are about the same size, 5–10 μm) are used for RS. The tissue samples are fixed in formaldehyde, dehydrated, and embedded in Paraplast™ as described by Mul et al.[216] Sections of paraffin-embedded material are placed on the stage of the microscope for Raman measurements and control purposes. The wavenumber for RS ranges from 4000 to 50 cm^{-1}. The Raman peaks of the reference spectrum have been reported at 115, 197, 366, 679, and 1049 cm^{-1}.[217] The 197 and 679 cm^{-1} bands as shown in the reference spectrum[217] are very strong, the 115 and 366 cm^{-1} bands are strong, and the 1049 cm^{-1} band is medium. A direct comparison between the test sample spectrum and the reference spectrum (the peak positions and their relative intensities) is sufficient to make an assignment for the talc particle.

6.3.2.2 Identification of Rutile (TiO_2) A pure rutile particle, used as a reference, and test particle samples in the foreign bodies in various tissues[215] that contain a rutile particle (both are about the same size, 5–10 μm) are used for RS. The tissue samples are fixed and placed on the stage of the microscope as described above. The wavenumber of RS ranges from 4000 to 50 cm^{-1}. The Raman peaks of the pure rutile spectrum have been reported at 240, 440, and 610 cm^{-1}.[218] The 240, 440, and 610 cm^{-1} bands as shown in the pure rutile spectrum[218] are broad, and their intensities are medium. A direct comparison between the test sample spectrum and the pure rutile spectrum (the frequencies, relative intensities, and linewidths) is sufficient to make an assignment for rutile in test particle samples.

6.3.2.3 Identification of α-Quartz (SiO_2) A reference particle and test sample particles in the foreign bodies in various tissues[215] that contain α-quartz particles (both are about the same size, 5–10 μm) are used for RS. The tissue samples are fixed and placed on the stage of the microscope as described above. The typical peaks of a reference spectrum have been reported at 128, 206, and 466 cm^{-1} band.[219] The 128 and 466 cm^{-1} bands as shown in the reference spectrum[219] are sharp, and the 206 cm^{-1} band is broad. A direct comparison between the test sample spectrum and the reference spectrum (the peak positions, shape, and their relative intensities) is sufficient to make an assignment for α-quartz samples.

6.3.2.4 Identification of Calcite Test particle samples and a reference particle in the foreign bodies in various tissues[215] that contain a calcite particle (both are about the same size, 5–10 μm) are used for RS. The tissue samples are fixed and placed on the stage of the microscope as described above. The Raman peaks of a reference spectrum have been reported at 156, 285, 715, and 1088 cm^{-1}.[220] A direct comparison between the test sample spectrum and the reference spectrum (the spectral positions and line shape) is sufficient to make an assignment for the calcite particle.

Table 5 Portable spectrometric techniques for monitoring gases and vapors

Instrumental technique	Applicable analyte(s)	Comments
IR photometers	CO, CO_2, NO_x, N_2O, SO_2, hydrocarbons, fluorocarbons, etc.	Detection limits sub-ppm to few percent range. Single-species or multigas devices
Colorimetric detection	Formaldehyde, TDI, HCN, Cl_2, H_2S, SO_2, NO_x, etc.	Sub-ppm detection limits for most species. Specific for certain target analytes
UV/VIS photometers	Hg vapor, O_3, SO_2, NO_x, NH_3, organic vapors, etc.	Sub-ppm to ppm detection limits. Analyte specific
Chemiluminescent detectors	Ozone, NO_x	Highly sensitive and selective. Detection limits $\sim$ 10 ppb
Photometric analyzers (includes devices based on flame photometry, fluorescence, other)	CO, SO_2, sulfur compounds., halogenated hydrocarbons, phosphorus compounds, etc.	Measurement in ppb to ppm range, depending on type. Single- and multispecies devices

7 FIELD-PORTABLE SPECTROSCOPY

7.1 Introduction

Field-portable methods for monitoring airborne workplace contaminants and toxins have received increasing attention. A number of portable monitors for airborne contaminants have been commercially available for many years, but new developments may provide for on-site compliance monitoring, which has heretofore been more the exception than the rule. The ability to conduct measurements on-site in the occupational setting offers significant advantages. Field-portable methods are often desired so that decisions regarding worker protection, engineering controls, and so on can be made quickly. The capability for rapid decision-making offered by on-site monitoring can help to save costs, and also offers a means to assess, and thereby provide timely prevention of, worker overexposures to toxic substances. Field-based monitoring is especially useful for applications in the construction industry, in agriculture, and in other situations where jobs may be short term and the workforce is transient. On-site techniques can also be beneficial in instances where short-term monitoring is desired. In this section, field-portable spectrometric techniques are covered, and some applications are presented. Because of limited space here, a general overview is presented; more specific information is available by consulting the literature referenced herein.

7.2 Portable Gas and Vapor Analyzers

While many commonly used portable gas and vapor analyzers are based on electrochemical or electrical measurement,[221–223] some rely on spectrometric means. Direct-reading instruments which rely on the use of IR, fluorescence, luminescence, or colorimetry have proven to be most popular.[222–225] Other spectrometries have also been used for on-site real-time gas and vapor monitoring. Some portable instruments, e.g. IR, allow for multigas detection, while others are designed for the measurement of single species of interest, such as carbon monoxide or ozone. Table 5 summarizes direct-reading spectrometric devices that are commonly used in the industrial hygiene field for monitoring gases and vapors. A wide variety of commercial instruments are available.

Most portable IR gas analyzers are nondispersive IR instruments that ordinarily require a plug-in power source.[222,223,226] Some dispersive instruments have also been introduced.[222] Battery-powered instruments have been produced, but these are generally species-specific; for instance, an IR photometer for monitoring carbon dioxide over a wide concentration range relies on Ni–Cd batteries as an optional power source. Portable multigas IR analyzers are becoming more popular as their applicability is enhanced through interfacing of the instrument with spectral libraries via computer. A wide variety of organic and inorganic gases can be monitored semiquantitatively or in some cases quantitatively. For example, portable IR monitors for CO are able to measure quantitatively this dangerous compound in the parts per million range, where such concentration levels are potentially hazardous to life and/or health.

Field-portable FTIR instruments are now available,[227,228] and a national voluntary consensus standard has been published which describes a portable FTIR method for determining gaseous compounds.[229] IR or FTIR monitoring allows for real-time or near real-time measurement of numerous toxic gases and vapors, and has applications in many occupational settings. An advantage of IR or FTIR monitoring is that sample preparation is minimal, and gases and vapors may be monitored following sampling by using a suitable sampler; or alternatively no samples may be needed at all (depending on the application). Open-path FTIR for real-time in situ monitoring of airborne gaseous pollutants has become popular for remote sensing, and also offers promise for applications in occupational settings.[230,231] A new IR spectral database standard, National Institute of Standards and Technology (NIST) SRM 79, has been made available recently.[232]

This NIST IR database contains absorption coefficient data for 21 hazardous air pollutants, and provides for quality assurance for quantitative FTIR open-path in situ measurements of these species.

Colorimetric and UV/VIS spectrophotometric monitors for gases and vapors are widely used in the industrial hygiene field (Table 5). Most of these monitoring devices, e.g. for ozone, mercury vapor, NO_x, ammonia, and sulfur dioxide, function by means of ultraviolet absorption.[222,233] Colorimetric monitors ordinarily employ a reaction between a selective reagent and the analyte of interest in order to form a colored complex which can be measured in the visible spectral range. Most commercially available colorimetric and UV/VIS photometers require an external power source for their operation. However, there are examples of such portable devices that can be operated using battery power, e.g. some Hg and SO_2 monitors. Portable, battery-powered devices are useful for personal monitoring, while monitors requiring an external power source are limited to use as area (static) monitors. UV/VIS and colorimetric instruments are able to detect most species of interest in the parts per million range, and many are equipped with alarms if readings are high. Many of the UV/VIS direct-reading instruments offer continuous monitoring capability, with response times of $\sim$1 s for measurements in the parts per million range.

Direct-reading chemiluminescent detectors offer a means for measuring ozone and oxides of nitrogen, with excellent specificity and high selectivity (Table 5). The operation of chemiluminescent detection involves excitation of O_3 molecules via chemical reaction and the subsequent detection of photon emission from the excited state species, which may either be an intermediate or a product of the reaction.[234] Ozone may function as a reagent for NO_x analysis or, of course, as the analyte of interest.[235] Luminol has also been used for field-portable monitoring of NO_2,[236] and ethylene has been employed as a reactant for field-portable monitoring of ozone.[237] Owing to high power needs, field chemiluminescence instruments generally require an external power source.[222,228]

Photometric analyzers are used widely in the industrial hygiene field for the on-site, real-time monitoring of numerous gas and vapor species (Table 5). The operation of the detector is via measurement of light emission from a high-temperature H_2 flame.[238] FPDs are useful for selective measurement of gaseous sulfur or phosphorus compounds, with detection limits in the parts per billion range for these species. The high power required for operation of these devices necessitates the use of an external power source.

Portable photoionization detectors (PIDs) for organic vapors are commercially available.[239] These require less power than the aforementioned photometric analyzers, and simple hand-held, battery-powered devices can be used to monitor volatile organic compounds (VOCs) in real time. However, for enhanced performance (such as minimizing interferences from more abundant hydrocarbons), PIDs can be used as detectors for portable GC instruments.[240]

Fluorescence analyzers are available for monitoring CO and SO_2 (Table 5). An Xe or Hg arc lamp is used to excite the analyte species, which give rise to sufficient fluorescence intensity so that detection limits in the few parts per billion range can be achieved.[222] Photomultiplier tubes are used for detection. Like photometric analyzers, there is a need for high power in order to facilitate source excitation and operation of photomultiplier detectors, and thus an external power source is required. The performance of the portable fluorimetric SO_2 analyzer has been shown to be equivalent to that of a colorimetric reference method promulgated by the EPA.[228]

Other types of photometric analyzer are commercially available, whereby alternative chemical strategies may be used to produce a spectral signal or color change.[222] Spectral intensity analyzers, for example, can be used for general, nonspecific monitoring of halogenated hydrocarbons. Other types of photometer allow for automated sampling by use of media which undergo a color change upon reaction with target analytes (e.g. like that already mentioned for monitoring SO_2). Reflectance may be used for the measurement of a variety of species such as ammonia, phosgene, HCN, and arsine, with detection limits in the parts per million range. In a few cases, field portability and on-site monitoring applications are enhanced by the use of battery power instead of an external power source.

New developments in spectrometric gas and vapor monitors have provided for better detection limits and other attributes. Fiber optic chemical sensors for continuous monitoring have been an area of wide interest,[241] and applications in the measurement of gases and vapors are widespread. Fiber optics have proven to be useful for the design of field-portable devices for optical,[242,243] fluorimetric,[244] and IR[245] monitoring of such species as VOCs, Hg, explosive agents, and so on. The use of optical sensor arrays for multispecies monitoring is also an area of significant promise.[246] Miniaturization of TOF mass spectrometric devices has allowed for the on-site monitoring of gaseous analytes, with excellent prospects for multispecies monitoring.[247] Hand-held ion mobility spectrometry (IMS) instruments have found applications in industrial hygiene and military applications.[248] A hand-held IMS device was used to monitor VOCs on-site in the workplace.[249] The

performance of IMS is enhanced when used as a detection scheme following GC separation.[250,251]

7.3 Portable Aerosol Monitors

Direct-reading portable aerosol monitors which are used for industrial hygiene purposes are often based on light-scattering or light-attenuation properties (Table 6). The most widely used are light-scattering devices known as aerosol photometers or nephelometers.[223] Light-attenuating photometers are also available, as are other real-time aerosol monitors that are not based on optical techniques.[252] Portable instruments have been developed which are applicable over different aerosol size ranges (Table 6) and each device has its own benefits and limitations.

Aerosol photometers operate by directing polychromatic light toward an aerosol as it is passed through an optical chamber, and by measuring the light which is scattered at a chosen scattering angle with respect to the incident light beam. Optical aerosol particle counters use a monochromatic light source such as a laser, photodiode or tungsten filament lamp to illuminate the aerosol sample. For both photometers and optical particle counters, photomultiplier tubes or photodiodes are generally used for detection of the scattered light. Smaller scattering angles are best for the detection of large particles, while a scattering angle of 90° offers maximum sensitivity for small particles.[223] Many factors contribute to the light scattering profile, e.g. wavelength of the incident light beam, size and shape of the particles, refractive index of the particle, density and concentration of the aerosol, and size distribution of the aerosol.[253] Thus it can be seen that these devices may suffer from numerous limitations which can restrict their applicability for quantitative monitoring, and therefore are generally not used for compliance monitoring purposes. Nevertheless, they are very useful for applications as on-site screening instruments and semiquantitative measurement of aerosol concentrations.

Table 6 Direct-reading optical aerosol monitors

Type of monitor	Applicable aerosol size range (μm)	Comments
Aerosol photometer or nephelometer	0.1–1.0	Integral light scattering; aerosols with same size distribution
Optical particle counter	~0.1–>10	Monochromatic or polychromatic source; light scattering
CNC	<0.01–1.0	Particles enlarged for photometric measurement

CNC, condensation nucleus counter.

Several popular instruments for estimation of aerosol concentrations in real-time are based on light-scattering methods, and a variety of techniques have been employed to improve their performance.[223,254] Techniques for sampling and analysis have been developed for both "extracted" samples and for in situ analysis.[252] Several devices utilize monochromatic IR or near-IR radiation to illuminate the sample, and some employ size-selective devices in order to isolate the aerosol range of interest. These devices are generally applicable for monitoring aerosols which are about 1.0 μm in diameter or greater, and are therefore not useful for the detection of small diameter aerosols. However, some have been claimed to give accurate aerosol concentration measurements over the range of 0.01–100 mg m^{-3} for aerosols of 0.1–20 μm in diameter.[223]

For measurement of very small aerosols, condensation nucleus counters (CNCs) are usually employed (Table 6). The CNC functions by actually enlarging the aerosol particles to a size which can be measured photometrically.[223,252] This is usually accomplished by subjecting the ultrafine aerosol to a vapor, and then cooling the mixture to cause supersaturation. In this manner the aerosol particles operate as condensation nuclei upon which the supersaturated vapor can nucleate and cause the aerosol particles to grow in diameter. In so doing, the intensity of scattered light can be used to measure the concentration of the enlarged particles. CNCs are widely used for testing high-efficiency particulate air filters in respirator fit-testing.[255]

With some exceptions, real-time aerosol monitors are survey instruments which can only be used to measure total concentrations of airborne particulates. Furthermore, the size ranges of aerosols which real-time aerosol monitors can measure tend to be limited.[256] Direct-reading, real-time aerosol monitoring instruments generally cannot give any information on the identities of airborne contaminants which may be present in the test aerosol. However, in recent instrument developments, efforts have been made to obtain more species-specific information using real-time optical monitoring. For example, a real-time monitor for respirable particles based on laser light scattering was said to be capable of detecting 0.1 fibers cm^{-3} with very short sampling times.[257] The instrument uses a diode array to sense scattered light and then assigns a particle to a particular class depending on its scattering characteristics. In most direct-reading instrumental applications, knowledge about the specific makeup of the aerosol being monitored is necessary before a survey instrument is employed.

Portable methods for measuring chemical species in captured aerosols ordinarily require that an aerosol sample be prepared and analyzed on-site in the field. In some cases sample preparation may be minimal, while

in other instances considerable sample treatment may be needed prior to on-site analysis. The option of using a field-portable analytical method depends on the needs of the user. If it is desirable to have an analytical result quickly, then it may be necessary to perform on-site analysis. Another possible reason for conducting analysis on-site has to do with the reactivity of target analytes which can exist in the sampled aerosol.[258] Reactive compounds may need to be analyzed quickly in the field using field-portable instrumentation, because the chemical form(s) of the analyte(s) of interest may change if samples are stored for too long a time period (e.g. for subsequent fixed-site laboratory analysis).

Various field-portable spectrometric techniques for on-site determination of heavy metals in collected aerosol samples have been evaluated. For instance, portable X-ray fluorescence (XRF) has been used to measure metal species in air filter samples.[259,260] In one study, portable XRF was used to determine metals in filters which were prepared from aerosolized metal oxides.[259] Excellent quantitative results for up to 18 metals were obtained, although detection limits for Pb and Cd were somewhat high compared to the action levels of interest. Very good correlations were obtained between portable XRF data and results obtained using a laboratory XRF instrument. Thus portable XRF offers the potential for on-site multielement monitoring of aerosol filter samples collected in the field. Since the method is nondestructive, the samples can be analyzed subsequently for metals content using a confirmatory technique such as ICPMS.

In related studies, modern portable XRF devices have been used to determine lead in filter samples, with a view toward obtaining lower detection limits for this metal. In one investigation, air filter samples were collected from construction sites where lead paint removal activities were undertaken.[261] The lead loading range of the data set was 0.1–1500 μg of lead per sample. Portable XRF measurements were conducted on the filter samples using a protocol which accounted for the variability in the density of the aerosol which was deposited on the filters. A NIOSH reference technique, graphite furnace atomic absorption spectrometry, was used for confirmatory analysis and method evaluation purposes.[262] For the portable XRF method, a lower detection limit of ~6 μg Pb/filter was determined, and the portable XRF method accuracy was ±16.4%. The performance of the portable XRF instrument indicated that the device can be used for the quantitative analysis of lead air filter samples over a wide concentration range. The practicing industrial hygienist can use portable XRF to produce a rapid on-site determination of lead exposure and immediately communicate to workers and help identify appropriate levels of personal protection.

Some other techniques that offer promise for on-site multimetals spectrochemical analysis include laser-induced breakdown spectroscopy (LIBS)[263] and spark-induced breakdown spectroscopy.[264] While neither instrument has been commercialized, prototypes of both have been evaluated for their ability to determine a number of heavy metals in air samples. For metals such as lead, LIBS may offer lower detection limits than portable XRF devices.[265] Efforts to make prototype LIBS devices more easily field portable have focused on the use of fiber optics.[265–267]

Spectrometric methods for the on-site analysis of species present in aerosols following a sample dissolution step have been published. For example, a field-portable method for the determination of airborne hexavalent chromium, Cr(VI), was developed and evaluated.[268] The procedure employed ultrasonic extraction[269,270] in order to solubilize Cr(VI) in test samples. Subsequently SPE using strong anion exchange was employed to separate Cr(VI), which is anionic, from Cr(III) and other metal cations. Following elution of the isolated Cr(VI), trace concentrations of Cr(VI) were measured using the diphenylcarbazide method[271,272] by means of a field-portable battery-powered spectrophotometer. It is expected that additional field-portable spectrometric measurement methods for more analytes will become more widely used in the future for industrial hygiene monitoring.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
AS	Atomic Spectrometry
ASTM	American Society for Testing and Materials
BEI	Backscattered Electron Image
CI	Chemical Ionization
CNC	Condensation Nucleus Counter
CP	Cyclophosphamide
CRD	Chemiluminescence–Redox Detector
CVAAS	Cold Vapor Atomic Absorption Spectrometry
EDS	Energy Dispersive Spectrometer
EDX	Energy Dispersive X-ray
EI	Electron Ionization
EPA	Environmental Protection Administration
ETAAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Flame Atomic Absorption Spectrometry

FAB	Fast Atom Bombardment
FIA	Flow Injection Analysis
FIAAS	Flow Injection Atomic Absorption Spectrometry
FPD	Flame Photometric Detector
FS	Fluorescent Spectrometry
FTICRMS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
Hb	Hemoglobin
HGAAS	Hydride Generation Atomic Absorption Spectrometry
HPLC	High-performance Liquid Chromatography
IA	Image Analyzer
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IF	Ifosfamide
IMS	Ion Mobility Spectrometry
IR	Infrared
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LIBS	Laser-induced Breakdown Spectroscopy
LOD	Limit of Detection
MALDI	Matrix-assisted Laser Desorption Ionization
MALDI/TOFMS	Matrix-assisted Laser Desorption Ionization/Time-of-flight Mass Spectrometry
MOCA	4,4′-Methylenebis-(2-chloroaniline)
MS	Mass Spectrometry
MS^n	Ion Trap Mass Spectrometry
MSD	Mass Spectral Detector
MSHA	Mine Safety and Health Administration
MS/MS	Tandem Mass Spectrometry
NIOSH	National Institute of Occupational Safety and Health
NIST	National Institute of Standards and Technology
OSHA	Occupational Safety and Health Administration
PAC	Polycyclic Aromatic Compound
PAH	Polycyclic Aromatic Hydrocarbon
PCR	Polymerase Chain Reaction
PID	Photoionization Detector
QTOFMS	Quadrupole Time-of-flight Mass Spectrometry
RS	Raman Spectroscopy
SEI	Secondary Electron Image
SEM	Scanning Electron Microscopy
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
TDI	Toluene Diisocyanate
TNT	2,4,6-Trinitrotoluene
TOF	Time-of-flight
TOFMS	Time-of-flight Mass Spectrometry
UV/VIS	Ultraviolet/Visible
VOC	Volatile Organic Compound
WDS	Wavelength Dispersive Spectrometer
XAD-2	2-(Hydroxymethyl)piperidine
XM	X-ray Microanalysis
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Infrared Spectroscopy in Environmental Analysis

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene • Dust, Measurement of Trace Elements in

Industrial Hygiene cont'd **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

REFERENCES

1. J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, New York, 1997.
2. M. Harper, C.R. Glowacki, P.R. Michael, 'Industrial Hygiene', *Anal. Chem.*, **69**, 307R–327R (1997).
3. J.E. Adkins, N.W. Henry, 'Industrial Hygiene', *Anal. Chem.*, **67**, 349R–376R (1995).
4. E.D. Hoffmann, J. Charette, V. Stroobant, *Mass Spectrometry – Principles and Applications*, John Wiley & Sons, New York, 1996.
5. M. McMaster, C. McMaster, *GC/MS – A Practical User's Guide*, Wiley-VCH, New York, 1998.

6. J.P. Lodge, Jr, *Methods of Air Sampling and Analysis*, 3rd edition, Lewis Publishers, Chelsea, MI, 1989.
7. D.R. Booker, W.D. Griffiths, D.M. Lyons, S.L. Upton, *Aerosol Sampling Guidelines*, AEA, UK, 1998.
8. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', *Anal. Chem.*, **70**, 747R–716R (1998).
9. E. Constantin, A. Schnell, *Mass Spectrometry*, Paramount Publishing, Amsterdam, 1991.
10. A. Sturaro, G. Parvoli, L. Doretti, G. Gori, G.B. Bartolucci, 'GC-MS Investigation of Polycyclic Aromatic Compounds in the Manufacture of Rubber Tubes', *Ann. Occup. Hyg.*, **37**(3), 327–334 (1993).
11. E. Menichini, L. Boniforti, S.D. Marzio, 'Occupational Exposure to Airborne Aromatic Amines in Rubber Manufacturing: Determination by GC-MS', *Toxicol. Environ. Chem.*, **22**, 9–16 (1989).
12. E. Menichini, L. Bonanni, F. Merli, 'Determination of Polycyclic Aromatic Hydrocarbons in Mineral Oils and Oil Aerosols in Glass Manufacturing', *Toxicol. Environ. Chem.*, **28**, 37–51 (1990).
13. J. Bundt, W. Herbel, H. Steinhart, 'Structure-type Separation of Diesel Fuels by Solid Phase Extraction and Identification of the Two- and Three-ring Aromatics by Capillary GC-MS', *J. High Resolut. Chromatogr.*, **14**(2), 91–98 (1991).
14. C.V. Cooper, 'GC-MS Analysis of Extracts of Workplace Air Samples for Nitrosamines', *Am. Ind. Hyg. Assoc. J.*, **48**(3), 265–270 (1987).
15. E.D. Pellizzari, J.E. Bunch, R.E. Berkley, J. McRae, 'Collection and Analysis of Trace Organic Vapor Pollutants in Ambient Atmospheres', *Anal. Lett.*, **9**(1), 45–63 (1976).
16. K.S. Chiu, K. Biemann, 'Structural Characterization of Polycyclic Aromatic Compounds by Combined GC-MS and GC-FTIR', *Anal. Chem.*, **56**, 1610–1615 (1984).
17. E.R. Dennedy, P.F. O'Connor, 'Application of Multidimensional GC-MS to the Determination of Glycol Ethers in Air', *J. Chromatogr.*, **522**, 303–313 (1990).
18. S. Dai, L. Zhang, Z. Bai, 'The Analysis of VOCs by GC/MS with Whole Column Cold Trapping on a Fused Silica Capillary Column in Indoor Environment', *Anal. Sci. Technol.*, **8**(4), 829–834 (1995).
19. R. Podeu, M. Einhaus, G. Rezniczek, 'Identification Von Inhalation Noxen am Arbeitsplatz mit Hilfe Von GC/MS', *Atemw-lungenkrkh*, **21**, S67–S69 (1995).
20. E.S. Lahaniatis, W. Bergheim, C. Rainer, 'Hazardous Halogenated Substances Formed During Combustion Processes', *Toxicol. Environ. Chem.*, **20–21**, 501–506 (1989).
21. E. Yrjanheikki, L. Pyy, E. Hakala, T. Lapinlampi, A. Lisko, K. Vahakangas, 'Exposure to Polycyclic Aromatic Hydrocarbons in New Cooking Plant', *Am. Ind. Hyg. Assoc. J.*, **56**, 782–787 (1995).
22. M. Maronj, A. Colombi, S. Cantoni, E. Ferioli, V. Foa, 'Occupational Exposure to Polychlorinated Biphenyls in Electrical Workers', *Br. J. Ind. Med.*, **38**, 49–54 (1981).
23. P.A. Smith, D.A. Gardner, D.B. Drown, G. Downs, W.W. Jederberg, W.W. Still, *Am. Ind. Hyg. Assoc. J.*, **58**, 868–875 (1997).
24. F. Mangani, L. Lattanzi, M. Maione, *Chromatographia*, **47**, 57–62 (1998).
25. G. Radolovich, D. Stewart, J. Pieri, M.J. Gooden, M. Chatfield, 'Development of a GCZ-MS/FPD System for the Simultaneous Analysis of VX, GB and HD', *Proc. ERDC Sci. Conf. Chem. Biol. Def. Res.*, ed. D.A. Berg, National Technical Information Service, Springfield, VA, 165–173, 1996.
26. E. Scobbie, D.W. Dabil, J.A. Groves, *Ann. Occup. Hyg.*, **42**, 45–59 (1998).
27. M. Yang, M. Koga, T. Katoh, T. Kawamoto, 'A Study for the Application of Urinary Naphthols, New Biomarkers for Airborne PAHs', *Arch. Environ. Toxicol.*, **36**, 99–108 (1999).
28. G. Grimmer, J. Jacob, J. Dettbarn, K.W. Naujack, 'Determination of Urinary Metabolites of Polycyclic Aromatic Hydrocarbons (PAH) for the Risk Assessment of PAH-exposed Workers', *Int. Arch. Occup. Environ. Health*, **69**, 231–239 (1997).
29. J.P. Buchet, J.P. Gennart, F. Mercado-Calderon, J.P. Delavignette, L. Cupers, R. Lauwerys, 'Evaluation of Exposure to Polycyclic Aromatic Hydrocarbons in a Coke Production and a Graphite Electrode Manufacturing Plant', *Br. J. Ind. Med.*, **49**, 761–768 (1992).
30. X. Guardino, C. Serra, J. Obiols, M.G. Rosell, M.J. Berenguer, F. Lopez, J. Brosa, 'Determination of DDT and Related Compounds in Blood Samples from Agricultural Workers', *J. Chromatogr. A*, **719**, 141–147 (1996).
31. H. Kontsas, C. Rosenberg, P. Pfaffli, P. Jappinen, 'GC-MS Determination of Chlorophenols in the Urine of Sawmill Workers with Past Use of Chlorophenol-containing Anti-stain Agents', *Analyst*, **120**(6), 1745–1749 (1995).
32. G.T. Vaughan, R.S. Kenyon, 'Monitoring for Occupational Exposure to 4,4′-Methylenebis(2-chloroaniline) by GC-MS Analysis of Hemoglobin Adducts, Blood, Plasma, and Urine', *J. Chromatogr. B*, **678**, 197–204 (1996).
33. P. Persson, M. Dalene, G. Skarping, M. Adamsson, L. Hagmar, 'Biological Monitoring of Occupational Exposure to Toluene Diisocyanate', *Br. J. Ind. Med.*, **50**, 1111–1118 (1993).
34. F.F. Hsu, V. Lakshmi, N. Rothman, V.D. Bhatnager, R.B. Hayes, R. Kashyap, D.J. Parikh, S.K. Kashyap, J. Turk, T. Zenser, B. Davis, 'Determination of Benzedrine, *N*-Acetylbenzidine, and *N*,*N*′-Diacetylbenzidine in Human Urine by Capillary GC/NICI-MS', *Anal. Biochem.*, **234**, 183–189 (1996).
35. K. Eriksson, J.-O. Levin, 'GC-MS Identification of Metabolites from α-Pinene in Human Urine after Occupational Exposure to Sawing Fumes', *J. Chromatogr. B*, **677**, 85–98 (1996).

36. G. Sabbioni, J. Wei, Y.-Y. Liu, 'Determination of Hemoglobin Adducts in Workers Exposed to 2,4,6-Trinitrotoluene', *J. Chromatogr. B*, **682**, 243–248 (1996).
37. G.R. Lambert, W.T. Padgett, M.H. George, K.T. Kitchin, S. Nesnow, 'Quantitative Analysis of Alachlor Protein Adducts by GC-MS', *Anal. Biochem.*, **268**, 289–296 (1999).
38. P.B. Farmer, E. Bailey, S.M. Gorf, M. Tornqvist, S. Osterman-Golkar, A. Kautiainen, D.P. Lewis-Enright, 'Monitoring Human Exposure to Ethylene Oxide by the Determination of Hemoglobin Adducts using GC-MS', *Carcinogenesis*, **7**(4), 637–640 (1986).
39. A. Ranasinghe, N. Scheller, K.Y. Wu, P.B. Upton, J.A. Swenberg, 'Application of GC/ECNCI-HRMS for Analysis of DNA and Protein Adducts', *Chem. Res. Toxicol.*, **11**, 520–526 (1998).
40. P. Ferranti, N. Sannolo, G. Mamone, I. Fiume, V. Carbone, M. Tornqvist, A. Bergman, A. Malorni, 'Structural Characterization by MS of Hemoglobin Adducts formed After In Vitro Exposure to Methyl Bromide', *Carcinogenesis*, **17**(12), 2661–2671 (1996).
41. D.L. Ashley, M.A. Bonin, F.L. Cardinali, J.M. McCraw, J.S. Holler, L.L. Needham, D.G. Patterson, 'Determining Volatile Organic Compounds in Human Blood from a Large Sample Population by using Purge and Trap GC/MS', *Anal. Chem.*, **64**, 1021–1029 (1992).
42. M. Coombs, V. Schillack, 'Determination of Trinitrotoluene and Metabolites in Urine by Means of GC-MS', *Int. Arch. Occup. Environ. Health*, **71**, S22–S25 (1998).
43. R. Joppich-Kuhn, R. Hanggi, P. Sagelsdorff, A.E. Smith, H.J. Weideli, M. Joppich, 'Determination of Dichlorobenzidine–Hemoglobin Adducts by GC/MS-NCI', *Int. Arch. Occup. Environ. Health*, **69**, 240–246 (1997).
44. H.-J.C. Chen, L. Zhang, J. Cox, J.A. Cunningham, F.-L. Chung, 'DNA Adducts of 2,3-Epoxy-4-hydroxynonanal: Detection of 7-(1′,2′-Dihydroxyheptyl)-3*H*-imidazo[2,1-*I*] Purine and 1,*N*-Ethenoadenine by GC/NICI-MS', *Chem. Res. Toxicol.*, **11**, 1474–1480 (1998).
45. D. Schutze, O. Sepai, J. Lewalter, L. Miksche, D. Henschler, G. Sabbioni, 'Biomonitoring of Workers Exposed to 4,4′-Methylenedianiline or 4,4′-Methylenediphenyl Diisocyanate', *Carcinogenesis*, **16**(3), 573–582 (1995).
46. P. Brunmark, M. Kalene, G. Skarping, 'GC/NICI-MS of Hydrolysed Human Urine and Blood Plasma for the Biomonitoring of Occupational Exposure to 4,4′-Methylenebisaniline', *Analyst*, **120**, 41–45 (1995).
47. M.W. Patrick, S. Shane, J.R. Froines, 'Determination of Hemoglobin Adduct Levels of the Carcinogen 2,4-Diaminotoluene using GC/EIPI-MS', *J. Chromatogr. B*, **667**, 166–172 (1995).
48. R.H. Hill, D.B. Shealy, S.L. Head, C.C. Williams, S.L. Bailey, M. Gregg, S.E. Baker, L.L. Needham, 'Determination of Pesticide Metabolites in Human Urine using an Isotope Dilution Technique and Tandem MS', *J. Anal. Toxicol.*, **19**, 323–329 (1995).
49. M. Bader, G. Lehnert, J. Angerer, 'GC/MS Determination of *N*-phenylvaline, a Possible Biomarker for Benzene Exposure in Human Hemoglobin by the *N*-alkyl Edman Method', *Int. Arch. Occup. Environ. Health*, **65**, 411–414 (1994).
50. S. Takahashi, M. Dagawa, K. Shiwaku, K. Matsubara, 'Determination of *S*-benzyl-*N*-acetyl-L-cysteine by GC/MS as a New Marker of Toluene Exposure', *J. Anal. Toxicol.*, **18**, 78–80 (1994).
51. K. Allam, S. Abdel-Baky, R.W. Giese, 'KO2 Chemical Transformation/MS Detection of Covalent Damage to the DNA of Cultured Human Lymphocytes Exposed to Benzo[*a*]pyrene', *Anal. Chem.*, **65**, 1723–1727 (1993).
52. G. Talaska, J.H. Roh, '32P-Postlabelling and MS Methods for Analysis of Bulky Polyaromatic Carcinogen-DNA Adducts in Humans', *J. Chromatogr.*, **580**, 293–323 (1992).
53. B.W. Day, S. Naylor, L.-S. Gan, Y. Sahali, T.T. Nguyen, P.L. Skipper, J.S. Wishnok, S.R. Tannenbaum, 'Molecular Dosimetry of Polycyclic Aromatic Hydrocarbon Epoxides and Diol Epoxides via Hemoglobin Adducts', *Cancer Res.*, **50**, 4611–4618 (1990).
54. W.M. Pierce, D.E. Nerland, 'Qualitative and Quantitative Analyses of Phenol Phenylglucuronide, and Phenylsulfate in Urine and Plasma by GC/MS', *J. Anal. Toxicol.*, **12**, 344–347 (1988).
55. P.B. Farmer, E. Bailey, S.M. Gorf, M. Tornqvist, S. Osterman-Golkar, A. Kautiainen, D.P. Lewis-Enright, 'Monitoring Human Exposure to Ethylene Oxide by the Determination of Hemoglobin Adducts using GC/MS', *Carcinogenesis*, **7**(4), 637–640 (1986).
56. E. Bailey, P.B. Farmer, I. Bird, J.H. Lamb, J.A. Peal, 'Monitoring Exposure to Acrylamide by the Determination of *S*-(2-Carboxyethyl) Cysteine in Hydrolyzed Hemoglobin by GC/MS', *Anal. Biochem.*, **157**, 241–248 (1986).
57. J. Angerer, C. Mannschreck, J. Gundel, 'Biological Monitoring and Biochemical Effect Monitoring of Exposure to Polycyclic Aromatic Hydrocarbons', *Int. Arch. Occup. Environ. Health*, **70**, 365–377 (1997).
58. R.W. Giese, M. Saha, S. Abdel-baky, K. Allam, 'Measuring DNA Adducts by GC/EC-MS Trace Organic Analysis', *Mass Spectrom.*, **271**, 505–523 (1997).
59. R.W. Giese, P. Vouros, 'Methods Development Toward the Measurement of Polyaromatic Hydrocarbon-DNA Adducts by MS', *Health Effects Institute Research Report*, Number 61, 1–36, 1993.
60. D.D. La, J.A. Swenberg, 'DNA Adducts: Biological Markers of Exposure and Potential Applications to Risk Assessment', *Mutat. Res.*, **365**, 129–146 (1996).
61. E. Kriek, M. Rojas, K. Alexandra, H. Bartsch, 'Polycyclic Aromatic Hydrocarbon–DNA Adducts in Humans: Relevance as Biomarkers for Exposure and Cancer Risk', *Mutat. Res.*, **400**, 215–231 (1998).

62. W.M.A. Niessen, *Liquid Chromatography–Mass Spectrometry*, 2nd edition (revised and expanded), Marcel Dekker, New York, 1999.
63. J.F. Garcia, D. Bacelo, 'LC-MS Interfacing Systems', *J. High Resolut. Chromatogr.*, **16**, 633–641 (1993).
64. P.J. Arpino, 'LC-MS, a Review', in *MS in the Biological Sciences*, ed. M.L. Gross, Kluwer Academic, The Netherlands, 253–267, 1992.
65. K. Tomer, C. Parker, 'Biochemical Application of LC-MS, *J. Chromatogr.*, **492**, 189–221 (1989).
66. D. Barcelo, 'Applications of LC-MS in Environmental Chemistry', *J. Chromatogr.*, **59**, 543 (1996).
67. A.L. Yergey, C.G. Edmonds, L.A.S. Lewis, M.L. Vestal, *LC/MS Techniques and Applications*, Plenum Press, London, 1990.
68. D.F. Parriott, *A Practical Guide to HPLC Detection*, Academic Press, San Diego, 1992.
69. G. Patonay, *HPLC Detection: Newer Methods*, VCH, Weinheim, 1992.
70. M.A. Brown, *LC-MS-Applications in Agricultural, Pharmaceutical, and Environmental Chemistry*, American Chemical Society, Washington, DC, 1990.
71. M.J. Suter, S. Riediker, W. Giger, 'Selective Determination of Aromatic Sulfonates in Landfill Leachates and Groundwater using Microbore LC Coupled with MS', *Anal. Chem.*, **71**, 897–904 (1999).
72. J. Ding, P. Vouros, 'Capillary Electrochromatography and Capillary Electrochromatography–MS for the Analysis of DNA Adduct Mixtures', *Anal. Chem.*, **69**, 379–384 (1997).
73. L. Leclercq, C. Laurent, E.D. Pauw, 'High-performance LC/Electrospray MS for the Analysis of Modified Bases in DNA', *Anal. Chem.*, **69**, 1952–1955 (1997).
74. P. Strickland, D. Kang, P. Sithisarankul, 'Polycyclic Aromatic Hydrocarbon Metabolites in Urine as Biomarkers of Exposure and Effect', *Environ. Health Perspect*, **104**, 927–932b (1996).
75. K. Vanhoutte, W.V. Dongen, E. Esmans, 'A Strategy for the Identification of 2′-Deoxynucleoside and 2′-Deoxynucleotide Adducts using Electrospray Tandem MS', *Eru. Mass Spectrom.*, **2**, 181–192 (1996).
76. H. Huan, J. Ahmedin, D. Connie, A. Steven, S.A. Barker, D.H. Swenson, J.C. Means, 'Analysis of DNA Adducts, *S*-[2-(*N*-guanyl)ethyl]glutathione by LC/MS and LC/Tandem MS', *Anal. Biochem.*, **265**, 139–150 (1998).
77. T.-Y. Yen, N.I. Christova-Gueoguieva, N. Scheller, S. Holt, J.A. Swenberg, M.J. Charles, 'Quantitative Analysis of the DNA Adduct *N*,-3-Ethenoguanine using LC/Electrospray Ionization MS', *J. Mass Spectrom.*, **31**, 1271–1276 (1996).
78. K. Vanhoutte, W.V. Dongen, H.I. Lemiere, E.L. Esmans, H.V. Onckelen, V. Eeckhout, R.E. Soest, A.J. Hudson, 'Development of a Nanoscale LC/Electrospray MS Methodology for the Detection and Identification of DNA Adducts', *Anal. Chem.*, **69**, 3161–3168 (1997).
79. R. Singh, G.M. Sweetman, P.B. Farmer, D.E. Shuker, K. Rich, 'Detection and Characterization of Two Major Ethylated Deoxyguanosine Adducts by HPLC/Electrospray MS, and ^{32}P-Postlabeling', *Chem. Res. Toxicol.*, **10**, 70–77 (1997).
80. N.Y. Tretyakova, S.-Y. Chiang, V. Walker, J.A. Swenber, 'Quantitative Analysis of 1,3-Butadiene-induced DNA Adducts In Vivo and In Vitro using LC/Electrospray Ionization Tandem MS', *J. Mass Spectrom.*, **33**, 363–376 (1998).
81. C. Minoia, R. Turci, C. Sottani, A. Schiavi, L. Perbellini, S. Angeleri, F. Draicchio, P. Apostoli, 'Application of HPLC/Tandem MS in Environmental and Biological Monitoring of Health Care Personnel Occupationally Exposed to Cyclophosphamide and Ifosfamide', *Rapid Commun. Mass Spectrom.*, **12**, 1485–1493 (1998).
82. A.A. Melikian, R. O'Connor, A.K. Prahalad, P. Hu, H. Li, M. Kagan, S. Thompson, 'Determination of the Urinary Benzene Metabolites *S*-phenylmercapturic Acid and *trans*,*trans*-Muconic Acid by LC/Tandem MS', *Carcinogenesis*, **20**(4), 719–726 (1999).
83. P. Ferranti, N. Sannolo, G. Mamone, F. Immacolata, V. Carbone, M. Tornqvist, A. Bergman, A. Malorni, 'Structural Characterization by MS of Hemoglobin Adducts Formed after In Vitro Exposure to Methyl Bromide', *Carcinogenesis*, **17**(12), 2661–2671 (1996).
84. J.D. Rosen, *Applications of New Mass Spectrometry Techniques in Pesticide Chemistry*, John Wiley & Sons, New York, 1987.
85. J. White, R.H. Brown, M.R. Clench, 'Particle Beam Liquid Chromatography/Mass Spectrometry Analysis of Hazardous Agricultural and Industrial Chemicals', *Rapid Commun. Mass Spectrom.*, **11**, 618–623 (1997).
86. D. Karlsson, 'Determination of Complex Mixtures of Airborne Isocyanates and Amines. Part 5: Determination of Low Molecular Weight Aliphatic Isocyanates as Dibutylamine Derivatives', *Analyst*, **123**, 1507–1512 (1998).
87. H. Tinnerber, M. Spanne, M. Dalene, G. Skarping, 'Determination of Complex Mixtures of Airborne Isocyanates and Amines. Part 3: Methylenediphenyl Diisocyanate, Methylenediphenylamino Isocyanate and Methylenediphenyl-diamine and Structural Analogs after Thermal Degradation of Polyurethane', *Analyst*, **122**, 275–278 (1997).
88. H. Tinnerberg, D. Karlsson, D. Dalene, M. Skarping, *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2207–2219 (1997).
89. A. Montaser, *Inductively Coupled Plasma MS*, Wiley-VCH, New York, 1998.
90. A.L. Gray, 'Solid Sample Introduction by Laser Ablation for ICP Source MS', *Analyst*, **110**, 551–556 (1985).
91. P. Schramel, I. Wendler, J. Angerer, 'The Determination of Metals (Antimony, Bismuth, Lead, Cadmium, Mercury, Palladium, Platinum, Tellurium, Thallium, Tin and Tungsten) in Urine Samples by ICP-MS', *Int. Arch. Occup. Environ. Health*, **69**, 219–223 (1997).

92. P. Apostoli, P.E. Brunelli, L. Alessio, 'Multiple Exposure to Metals in Eight Types of Welding', *G. Ital. Med. Lav. Frg.*, **19**(2), 8–14 (1997).
93. Z. Karpas, L. Halicz, J. Roiz, R. Marko, E. Katorza, A. Lorber, Z. Goldbart, 'ICP-MS as a Simple, Rapid and Inexpensive Method for Determination of Uranium in Urine and Fresh Water', *Health Phys.*, **71**(6), 879–885 (1996).
94. A. Schutz, I.A. Bergdahl, A. Ekholm, S. Skerfving, 'Measurement of ICP-MS of Lead in Plasma and Whole Blood of Lead Workers and Controls', *Occup. Environ. Med.*, **53**, 736–740 (1996).
95. O.A. Vita, K.C. Mayfield, 'Analysis of Urine for U-235 and U-238 by ICP-MS', in *Application of ICP-MS to Radionuclide Determinations*, ed. R.W. Morrow, J.S. Crain, ASTM Special Technical Publication (PCN, 04-012910-35), West Constohocken, PA, 1995.
96. P. Apostoli, S. Giusti, D. Bartoli, A. Perico, P. Bavazzano, L. Alessio, 'Multiple Exposure to Arsenic, Antimony, and Other Elements in Art Glass Manufacturing', *Am. J. Ind. Med.*, **34**, 65–72 (1998).
97. R.J. Cotter, *Time-of-Flight Mass Spectrometry Instrumentation and Applications in Biological Research*, Am. Chem. Soc., Washington, DC, 1997.
98. P.L. Ross, P.A. Davis, P. Belgrader, 'Analysis of DNA Fragments from Conventional and Microfabricated PCR Devices using Delayed Extraction MALDI/TOF-MS', *Anal. Chem.*, **70**, 2067–2078 (1998).
99. D.Z. Bezabeh, T.M. Allen, E.M. McCauley, P.B. Kelly, 'Negative Ion Laser Desorption Ionization TOF-MS of Nitrated PAH', *J. Am. Soc. Mass Spectrom.*, **8**, 630–636 (1997).
100. S.M. Hankin, P. John, G. Smith, 'Laser TOF-MS of PAH–Picrate Complexes, *Anal. Chem.*, **69**, 2927–2930 (1997).
101. S. Hankin, P. John, 'Laser TOF-MS Analysis of PAHs on Single Diesel Particulates', *Anal. Chem.*, **71**, 1100–1104 (1997).
102. M.J. Lee, S. Monte, J. Sanderson, N.J. Haskins, 'A Preliminary Study using Fast Gradient LC Coupled to a Q-TOF/MS', *Rapid Commun. Mass Spectrom.*, **13**, 216–221 (1999).
103. F.G. Hanisch, B.N. Green, R. Bateman, J. Peter-Katalinic, 'Localization of *o*-Glycosylation Sites of MUC1 Tendem Repeats by Q-TOF ESI Mass Spectrometry', *J. Mass Spectrom.*, **33**, 358–362 (1998).
104. D.L. Deforce, J. Raymackers, L. Meheus, V. Wijnendaele, A.D. Leenheer, V.D. Eeckhout, 'Characterization of DNA Oligonucleotides by Coupling of Capillary Zone Electrophoresis to Electrospray Ionization Q-TOF-MS', *Anal. Chem.*, **70**, 3060–3068 (1998).
105. J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, A.D. Romig, C.E. Lyman, C. Fiori, E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*, Plenum, New York, 1–820, 1992.
106. R.E. Lee, *Scanning Electron Microscopy and X-ray Microanalysis*, PTR Prentice Hall, Englewood Cliffs, NJ, 1–458, 1993.
107. P. Ingram, J.D. Shelburne, A. LeFurgey, 'Principles and Instrumentation', in *Microprobe Analysis in Medicine*, eds. P. Ingram, J.D. Shelburne, V.L. Roggli, Hemisphere Publishing, New York, 1–34, 1989.
108. K.F.J. Heinrich, *Electron Beam X-Ray Microanalysis*, Van Nostrand Reinhold, New York, 1–578, 1981.
109. J.J. Friel, *X-Ray and Image Analysis in Electron Microscopy*, Princeton-Gamma-Tech, Princeton, NJ, 1–97, 1998.
110. R.J. Lee, J.F. Kelley, 'Overview of SEM-based Automated Image Analysis', *Scanning Electron Microsc.*, **I**, 303–310 (1980).
111. J.D. Shelburne, P. Ingram, A. LeFurgey, R. Herlong, 'Preparatory Techniques', in *Microprobe Analysis in Medicine*, eds. P. Ingram, J.D. Shelburne, V.L. Roggli, Hemisphere Publishing, New York, 35–53, 1989.
112. P.B. DeNee, 'Identification and Analysis of Particles in Biological Tissue using SEM and Related Techniques', *Scanning Electron Microsc.*, **I**, 461–468 (1976).
113. R.S. Thomas, J.R. Hollahan, 'Use of Chemically-reactive Gas Plasmas in Preparing Specimens for Scanning Electron Microscopy and Electron Probe Microanalysis', *Scanning Electron Microsc.*, **I**, 83–92 (1974).
114. L.E. Stettler, D.H. Groth, S.F. Platek, 'Automated Characterization of Particles Extracted from Human Lungs: Three Cases of Tungsten Carbide Exposure', *Scanning Electron Miscrosc., I*, 439–448 (1983).
115. J.P. Mastin, L.E. Stettler, J.D. Shelburne, 'Quantitative Analysis of Particulate Burden in Lung Tissue', *Scanning Microsc.*, **2**(3), 1613–1629 (1988).
116. P.G. Coin, V.L. Roggli, A.R. Brody, 'Deposition, Clearance, and Translocation of Chrysotile Asbestos from Peripheral and Central Regions of the Rat Lung', *Environ. Res.*, **58**, 97–116 (1992).
117. D. Baker, K.G. Kupke, P. Ingram, V.L. Roggli, J.D. Shelburne, 'Microprobe Analysis in Human Pathology', *Scanning Electron Microsc.*, **II**, 659–680 (1985).
118. L.E. Stettler, S.F. Platek, R.D. Riley, J.P. Mastin, S.D. Simon, 'Lung Particulate Burdens of Subjects from the Cincinnati, Ohio Urban Area', *Scanning Microsc.*, **5**(1), 85–94 (1991).
119. L.E. Stettler, D.H. Groth, S.F. Platek, J.R. Burg, 'Particulate Concentrations in Urban Lungs', in *Microprobe Analysis in Medicine*, eds. P. Ingram, J.D. Shelburne, V.L. Roggli, Hemisphere Publishing, New York, 133–146, 1989.
120. A. Churg, B. Wiggs, 'Mineral Particles, Mineral Fibers, and Lung Cancer', *Environ Res.*, **37**, 364–372 (1985).
121. A. Churg, B. Wiggs, 'Types, Numbers, Sizes, and Distribution of Mineral Particles in the Lungs of Urban Male Cigarette Smokers', *Environ. Res.*, **42**, 121–129 (1987).
122. L. Paoletti, D. Batisti, S. Caiazza, M.G. Petrelli, F. Taggi, L. De Zorzi, M.A. Dina, G. Donelli, 'Mineral Particles

in the Lungs of Subjects Resident in the Rome Area and Not Occupationally Exposed to Mineral Dust', *Environ. Res.*, **44**, 18–28 (1987).

123. J.L. Abraham, B.R. Burnett, A.L. Hunt, 'Development and use of a Pneumoconiosis Database of Human Pulmonary Inorganic Particulate Burden in over 400 Lungs', *Scanning Microsc.*, **5**(1), 95–108 (1991).
124. J.L. Abraham, B.R. Burnett, 'Quantitative In Situ Analysis of Inorganic Particulate Burden in Tissue Sections–an Update', in *Microprobe Analysis in Medicine*, eds. P. Ingram, J.D. Shelburne, V.L. Roggli, Hemisphere Publishing, New York, 111–131, 1989.
125. J.L. Abraham, B.R. Burnett, 'Quantitative Analysis of Inorganic Particulate Burden In Situ in Tissue Sections', *Scanning Electron Microsc.*, **II**, 681–696 (1983).
126. L. LeBouffant, H. Daniel, J.C. Martin, S. Bruyere, 'Effects of Impurities and Associated Minerals on Quartz Toxicity', *Ann. Occup. Hyg.*, **26**, 625–634 (1982).
127. W. Kriegseis, A. Scharmann, 'Specific Harmfulness of Respirable Dusts from West German Coal Mines V: Influence of Mineral Surface Properties', *Ann. Occup. Hyg.*, **26**, 511–525 (1982).
128. J.C. Harrison, P.S. Brower, M.D. Attfield, C.B. Doak, M.J. Keane, R.L. Grayson, W.E. Wallace, 'Surface Composition of Respirable Silica Particles in a Set of US Anthracite and Bituminous Coal Mine Dusts', *J. Aerosol Sci.*, **28**, 689–696 (1997).
129. W.E. Wallace, M.J. Keane, 'Differential Surface Composition Analysis by Multiple-voltage Electron Beam X-ray Spectroscopy', US Patent 5 210 414, US Government Printing Office, Washington, DC, 1993.
130. W.E. Wallace, J. Harrison, M.J. Keane, P. Bolsaitis, D. Eppelsheimer, J. Poston, S. Page, 'Clay Occlusion of Respirable Quartz Particles Detected by Low Voltage Scanning Electron Microscopy–X-ray Analysis', *Ann. Occup. Hyg.*, **34**, 195–204 (1990).
131. V.E. Cosslett, R.N. Thomas, 'Multiple Scattering of 5–30 keV Electrons in Evaporated Metal Films', *Br. J. Appl. Phys.*, **15**, 883–907 (1944).
132. J.I. Goldstein, H. Yakowitz, *Practical Scanning Electron Microscopy, Electron and Ion Microprobe Analysis*, Plenum Press, New York, 83, 1975.
133. K.F.J. Heinrich, *Electron Beam X-Ray Microanalysis*, Van Nostrand Reinhold, Scarborough, CA, 240, 1981.
134. W.J. Veigele, 'Attenuation Cross-sections in cm^2/g at Selected Wavelengths for 94 Elements', in *Handbook of Spectroscopy*, ed. J.W. Robinson, CRC Press, Boca Raton, FL, Vol. 1, 1974.
135. W.E. Wallace, M.J. Keane, J.C. Harrison, J.W. Stephens, P.S. Brower, R.L. Grayson, M.D. Attfield, 'Surface Properties of Silica in Mixed Dusts', in *Silica and Silica-Induced Lung Diseases*, eds. V. Castranova, V. Vallyathan, W.E. Wallace, CRC Press, Boca Raton, FL, 1996.
136. W.E. Wallace, J.C. Harrison, R.L. Grayson, M.J. Keane, P. Bolsaitis, R.D. Kennedy, Q.A. Wearden, M.D. Attfield, 'Aluminosilicate Surface Contamination of Respirable Quartz Particles from Coal Mine Dusts and from Clay Works Dusts', *Ann. Occup. Hyg.*, **38**, 439–445 (1994).
137. K.D. Childs, D.L.A. LaVanier, P.M. Lindley, B.W. Scheuler, G. Mulholland, A.C. Diebold, 'Comparison of Submicrometer Particle Analysis by Auger Electron Microscopy, Time-of-Flight Secondary Ion Mass Spectrometry, and Secondary Electron Microscopy with Energy-dispersive X-ray Spectroscopy', *J. Vac. Sci. Technol., A*, **14**, 2392–2404 (1996).
138. A. Grekula, E. Ristolainen, V.P. Tanninen, H.K. Hyvarinen, P.L. Kalliomaki, 'Surface and Bulk Chemical Analysis on Metal Aerosols Generated by Manual Metal Arc Welding of Stainless Steel', *J. Aerosol Sci.*, **17**, 1–9 (1995).
139. J.W. Stephens, J.C. Harrison, W.E. Wallace, 'Correlating Auger Electron Spectroscopy with Scanning Electron Microscopy–Energy Dispersive Spectroscopy for the Analysis of Respirable Particles', *Scanning*, **20**, 302–310 (1998).
140. D.L. Tsalev, Z.K. Zaprianov, *Atomic Absorption Spectrometry in Occupational and Environmental Health Practice*, CRC Press, Florida, Vols I & II, 1984.
141. S.J. Haswell, *Atomic Absorption Spectrometry-Theory, Design and Applications*, Oxford, New York, 1991.
142. A. Sanz-Medel, *Flow Analysis with Atomic Spectrometric Detectors*, Lausanne, New York, 1999.
143. D.L. Tsalev, *Atomic Absorption Spectrometry in Occupational and Environmental Health Practice*, CRC Press, New York, Vol. III, 1995.
144. P.J. Tsai, J.H. Vincent, G.A. Wahl, G. Maldonado, 'Worker Exposures to Inhalable and Total Aerosol During Nickel Alloy Production', *Ann. Occup. Hyg.*, **40**(6), 651–659 (1996).
145. W. Torjussen, F.-M.S. Haug, I. Andersen, 'Concentration and Distribution of Heavy Metals in Nasal Mucosa of Nickel-exposed Workers and of Controls', *Acta Otolaryngol.*, **86**, 449–463 (1978).
146. C.W. Pilger, I. Broder, Determination of Metals on Airborne Particulates by AAS, Adapted from Method P & CAM 173, *NIOSH Manual of Analytical Methods*, Vol. I, 1977, and Method ID-121, OSHA, 1985.
147. D. Bellido-Milla, M.P. Hernandez-Artiga, J.L. Hidalgo-Hidalgo, J.A. Murioz-Leyva, 'Analytical Study of Hygiene Hazards Involved in Naval Industry Welding Processes', *Appl. Occup. Environ. Hyg.*, **10**(11), 921–925 (1995).
148. J.L. Burguera, M. Burguera, C.E. Rondon, C. Rivas, J.A. Burguera, O.M. Alarcon, 'Determination of Lead in Hair of Exposed Gas Station Workers and in Unexposed Adults by Microwave-aided Dissolution of Samples and FI/AAS', *J. Trace Elem. Electrolytes Health Dis.*, **1**, 21–26 (1987).
149. A. Othman, 'Preliminary Investigation of the Lead Level in Whole Blood of Normal and Occupationally Exposed

Populations in Damascus City', *Sci. Tot. Environ.*, **43**, 141–148 (1985).
150. D.L. Tsalev, F.J. Langmyhr, N. Gunderson, 'Direct Atomic Absorption Spectrometric Determination of Manganese in Whole Blood of Unexposed Individuals and Exposed Workers in a Norwegian Manganese Alloy Plant', *Bull. Environ. Contam. Toxicol.*, **17**(6), 660–666 (1977).
151. H.B. Rollin, C.M.C.A. Nogueira, 'Identification of Aluminium Fractions in Serum using the Techniques of HPLC, Ultrafiltration and Zeeman AAS', *Eur. J. Clin. Chem. Clin. Biochem.*, **35**, 215–222 (1997).
152. H.J. Gitelman, F.R. Alderman, M. Kurs-Lasky, H.E. Rockette, 'Serum and Urinary Aluminium Levels of Workers in the Aluminium Industry', *Am. Occup. Hyg.*, **39**(2), 181–191 (1995).
153. H.B. Rollin, P. Theodorou, T.A. Kilroe-Smith, 'The Effect of Exposure to Aluminium on Concentrations of Essential Metals in Serum of Foundry Workers', *Br. J. Ind. Med.*, **48**, 243–246 (1991).
154. S. Letzel, K.H. Schaller, J. Angerer, H. Drexler, A. Weber, K. Schmid, D. Weltle, 'Biological Monitoring of Occupational Aluminium Powder Exposure', *Occup. Hyg.*, **3**, 271–280 (1996).
155. M.A. White, S.A. O'Hagan, A.L. Wright, H.K. Wilson, 'The Measurement of Salivary Cadmium by Electrothermal AAS and its use as a Biological Indicator of Occupational Exposure', *J. Expos. Anal. Environ. Epid.*, **2**(2), 195–206 (1992).
156. T.V. Abernathy, K.B. Lee, R.J. Parker, E. Reed, 'The Measurement of Cadmium in Biological Materials using Graphite Furnace AAS with Zeeman Background Correction', *Oncol. Rep.*, **6**, 155–159 (1999).
157. S.A. Armitage, M.A. White, H.K. Wilson, 'The Determination of Silver in Whole Blood and its Application to Biological Monitoring of Occupationally Exposed Groups', *Am. Occup. Hyg.*, **40**(3), 331–338 (1996).
158. M.A. White, A.M. Boran, *Urinary Excretion of Nickel in Nickel-Chromium Electroplates, Nickel and Human Health: Current Perspectives*, eds. E. Nieboer, J.O. Nriagu, John Wiley & Sons, New York, 1992.
159. G.E. Jensen, J.M. Christensen, O.M. Poulsen, 'Occupational and Environmental Exposure to Arsenic-increased Urinary Arsenic Level in Children', *Sci. Tot. Environ.*, **107**, 169–177 (1991).
160. O.J. Blas, S.V. Gonzalez, R.S. Rodriguez, J.H. Mendez, *Determination and Speciation of Arsenic in Human Urine by Ion-exchange Chromatography/FIA with Hydride Generation/AAS*, ed. C.E. Van Hall, ASTM, Philadalphia, 302–323, 1994.
161. M.D. Martin, T. Mccann, C. Naleway, J.S. Woods, B.G. Leroux, A.M. Bollen, 'The Validity of Spot Urine Samples for Low-level Occupational Mercury Exposure Assessment and Relationship to Porphyrin and Creatinine Excretion Rates', *J. Pharmacol. Exp. Ther.*, **277**, 239–277 (1996).
162. M.A. Seckin, S. Aygun, Y. Ataman, 'Determination, Speciation of Mercury in a Dental Work Place by Cold Vapor AAS and Gas–Liquid Chromatography', *Int. J. Environ. Anal. Chem.*, **26**, 1–17 (1986).
163. G.L. Moore, *Introduction to ICP-AES-Analytical Spectroscopy Library*, New York, Vol. 3, 1989.
164. F.B. Lo, D.K. Arai, 'Biological Monitoring of Toxic Metals in Urine by Simultaneous ICP-AES', *Am. Ind. Hyg. Asoc. J.*, **50**(5), 245–251 (1989).
165. M. Kudo, Y. Aizawa, T. Takata, 'Biological Exposure Evaluation and Hair Analysis in Workers Handling Chromium Compounds', *Kitasato Lgaku*, **18**(1), 39–47 (1988).
166. K.W. Olson, W.J. Haas, V.A. Fassel, 'Analyses of Airborne Particulates and Human Urine by ICP-AES', *DHEW (NIOSH)*, Publication No. 79–110, NIOSH, Cincinnati, OH, 117, 1978.
167. A. Franzlau, 'Use of ICP-AES in Screening for Trace Metal Exposures in an Industrial Population', *Environ. Res.*, **46**(1), 15–24 (1988).
168. R.D. Hull, 'Multielement Analyses of Industrial Hygiene Samples', *American Industrial Hygiene Association Conference*, May 25–29, 1981.
169. M. Harper, C.R. Glowacki, P.R. Michael, 'Industrial Hygiene', *Anal. Chem.*, **69**, 307R–327R (1997).
170. P.M. Eller, M.E. Cassinelli (eds), *NIOSH Manual of Analytical Methods*, 4th edition, U.S. DHHS, PHS, CDC, NIOSH, DPSE, Cincinnati, OH, 1994.
171. J.A. Dean, *Analytical Chemistry Handbook*, McGraw-Hill, NY, 1995.
172. J.E. Davis, E.R. Stevens, D.C. Staff, *Bull. Environ. Contam. Toxicol.*, **31**, 631 (1983).
173. W.F. Durham, H.R. Wolfe, J.W. Elliott, 'Absorption and Excretion of Parathion by Spraymen', *Arch. Environ. Health*, **24**, 381 (1972).
174. J.E. Peterson, in *Industrial Hygiene Aspects of Plant Operations*, eds. L.J. Cralley, L.V. Cralley, MacMillan, New York, 222–247, Vol. 2, 1984.
175. N.S. Ham, A. Walsh, *Spectrochim. Acta*, **12**, 88 (1958).
176. G.W. King, in *Spectroscopy and Molecular Structure*, ed. R.J.H. Clark, Winston, New York, 1964.
177. M. Tasumi, M. Fujiwara, 'Vibrational Spectra of Chlorophylls', in *Advances in Spectroscopy*, eds. L.J. Cralley, R.E. Hester, John Wiley & Sons, New York, 1987.
178. M. Manfait, T. Theophanides, 'Drug–Nucleic Acid Interactions', in *Advances in Infrared and Raman Spectroscopy*, eds. L.J. Cralley, R.E. Hester, John Wiley, New York, Vol. 13, 1986.
179. W. Levin, 'Vibrational Spectroscopy in Membrane Assemblies', in *Advances in Infrared and Raman Spectroscopy*, eds. L.J. Cralley, R.E. Hester, John Wiley, New York, Vol. 11, 1984.
180. Y. Ozaki, K. Iriyama, 'Potential of Raman Spectroscopy in Medical Science', in *Laser Light Scattering Spectroscopy of Biological Objects*, eds. J. Stepanek, P. Anzenbacher, B. Sedlacek, Elsevier, Amsterdam, 1987.

181. V.A. Maroni, J.R. Ferraro, in *Practical Fourier Transform Infrared and Raman Spectroscopy-Industrial Laboratory Chemical Analysis*, eds. J.R. Ferraro, K. Krishman, Academic Press, San Diego, 1990.
182. National Institute for Occupational Safety and Health, *NIOSH Manual of Analytical Methods*, ed. P. Eller, Cincinnati, Ohio, 1986.
183. C.T. Helmers, 'Industrial Central Atmospheric Monitor', *Sensors*, **2**, 20–25 (1984).
184. B.S. Samimi, 'Calibration of MIRAN Gas Analyzer: Extent of Vapor Loss within a Closed Loop Calibration System', *Am. Ind. Assoc. J.*, **44**, 40–45 (1983).
185. H. Xiao, S.P. Levine, J.B. D'Arcy, G. Kinnes, D. Almaguer, 'Comparison of the Fourier Transform Infrared (FTIR) Spectrophotometer and the Miniature Infrared Analyzer (MIRAN) for the Determination of Trichloroethylene (TCE) in the Presence of Freon-133 in Workplace Air', *Am. Ind. Hyg. Assoc. J.*, **7**, 395–401 (1990).
186. C.R. Strang, S.P. Levine, W.F. Herget, 'Evaluation of the Fourier Transform Infrared (FTIR) Spectrometer as a Quantitative Air Monitor for Semiconductor Manufacturing Process Emissions', *Am. Ind. Hyg. Assoc. J.*, **50**, 70–77 (1989).
187. E.R. Stephens, W.E. Scott, P.L. Hanst, R.C. Doerr, 'Recent Developments in the Study of the Organic Chemistry of the Atmosphere', *J. Air Pollut. Control. Assoc.*, **6**, 159–165 (1956).
188. E.C. Tuazon, R.A. Graham, A.M. Winer, R.R. Easton, J.N. Pitts, P.L. Hanst, 'A Kilometer Pathlength Fourier-transform Infrared System for the Study of Trace Pollutants in Ambient and Synthetic Atmosphere', *Atmos. Environ.*, **12**, 865–875 (1978).
189. W.F. Herget, S.P. Levine, 'Fourier Transform Infrared (FTIR) Spectroscopy for Monitoring Semiconductor Process Gas Emissions', *Appl. Ind. Hyg.*, **1**, 110–112 (1986).
190. J.E. Sigsby, A. MaCarver, R. Snow, *Evaluation of a FTIR Mobile Source Measurement System*, EPA 600/S3/036, EPA, Washington, DC, 1989.
191. P.L. Hanst, W.E. Wilson, R.K. Patterson, B.W. Gay, L.W. Chaney, C.S. Burton, *A Spectroscopic Study of California Smog*, EPA 650/4-75-006, EPA, Washington, DC, 1975.
192. J.M. Heuss, W.A. Glasson, 'Hydrocarbon Reactivity and Eye Irritation', *Environ. Sci. Technol.*, **2**, 1109–1116 (1968).
193. G.M. Russwurm, R.H. Kagann, O.A. Simpson, W.A. MaClenny, W.F. Herget, 'Long-path FTIR Measurements of Volatile Organic Compounds in an Industrial Setting', *J. Air Waste Manag. Assoc.*, **41**, 1062–1066 (1991).
194. G.M. Russwurm, W.A. MaClenny, 'A Comparison of FTIR Open Path Ambient Data with Method TO-14 Canister Data', in *Proceeding of International Conference on Optical Remote Sensing*, Air and Waste Management Association, Pittsburgh, PA, 248–253, 1990.
195. W.F. Herget, J.D. Brasher, 'Remote Measurement of Gaseous Pollutant Concentrations using a Mobile Fourier Transform Interferometer System', *Appl. Opt.*, **18**, 3404–3420 (1979).
196. H. Xiao, S.P. Levine, J.B. D'Arcy, R. Spear, T. Pritchett, 'A Transportable Remote Sensing, Infrared Air-monitoring System', *Am. Ind. Hyg. Assoc. J.*, **52**, 449–457 (1991).
197. O.A. Simpson, R.H. Kagann, 'Measurement of Emissions at a Chemical Wastewater Treatment Site with an Open Path Remote Fourier Transform Interferometer', in *Proceeding of International Conference on Optical Remote Sensing*, Air and Waste Management Association, Pittsburgh, PA, 937–939, 1990.
198. L.T. Tso, W.C. Liao, S.I. Chang, 'Portable Long Open Path FTIR Applied in In Situ Measurement of Trace Gases of Ambient Air Pollution', in *Proceeding of International Conference on Optical Remote Sensing*, Air and Waste Management Association, Pittsburgh, PA, 95–106, 1992.
199. U.S. Mine Safety and Health Administration, 'Infrared Determination of Quartz in Respirable Coal Mine Dust', *Chemical Method No P-7*, 19, 1989.
200. D.P. Tuchman, *Research Toward Effects of Quartz Dust on Filters using FTIR Spectroscopy*, US Bureau of Mines, Washington, DC, 1–17, 1992.
201. C. Lorberau, 'Investigation of the Determination of Respirable Quartz on Filter Media using Fourier Transform Infrared Spectrophotometry', *Appl. Occup. Environ. Hyg.*, **5**, 348–350 (1990).
202. C.C. Anderson, 'Collaborative Tests of Two Methods for Determining Free Silica in Airborne Dust', in *NIOSH-210-79-0059*, Washington, DC, 157, 1983.
203. National Institute for Occupational Safety and Health, 'Development of Respirable Pure Quartz and Cristobalite Calibration Standard Reference Materials with known Particle Size Distribution', in *NIOSH-IA-80-36*, Washington, DC, 1–15, 1984.
204. H.K. Beard, R.R. Brown, H. Muir, 'Immunochemical Localization of Collagen Types and Proteoglycan in Pig Invertebral Discs', *Immunology*, **41**, 491 (1980).
205. R. Ganesh, L.J. Striker, L.D. Hudson, G.E. Striker, 'Extracellular Matrix in Normal and Fibrotic Human Lungs', *Am. Rev. Respir. Dis.*, **131**, 281–289 (1985).
206. L. Yurui, H. Xun, Y. Zhu, 'Studies on Structural Changes of Collagen in Silicosis', *Biomed. Environ. Sci.*, **7**, 302–306 (1994).
207. A. Leesmith, 'Infrared Spectroscopy', in *Practical Handbook of Spectroscopy*, ed. I.W. Robinson, CRC Press, Boca Raton, FL, 579, 1991.
208. J.J. Barrett, *Appl. Spectrosc. Rev.*, **21**, 419 (1985).
209. G.H. Atkinson, in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, John Wiley, New York, Vol. 9, 1982.

210. L.M. Proniewicz, I.R. Paeng, K. Nakamoto, *J. Am. Chem. Soc.*, **113**, 3294 (1991).
211. D.M. Dams, P.D. Hatton, A.C. Shaw, T.K. Tan, *J. Chem. Soc. Chem. Commun.*, 226 (1981).
212. B. Schrader, in *Practical Fourier Transform Infrared and Raman Spectroscopy-industrial Laboratory Chemical Analysis*, eds. J.R. Ferraro, K. Krishman, Academic Press, San Diego, 167–202, 1990.
213. R.L. Garrell, 'Surface-enhanced Raman Spectroscopy', *Anal. Chem.*, **61**, 401 (1989).
214. R.P. Cooney, M.R. Mahoney, A.J. McQuillan, in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Heyden, London, Vol. 9, 1982.
215. J.M. Lerner, F. Adar, *Laser Focus World*, Pennwell Publishing, Feb., 1989.
216. F.F.M.D. Mul, H. Buiteveld, J. Lankester, J. Mud, J. Greve, 'Raman Microspectroscopy in Human Pathology', *Hum. Pathol.*, **15**, 1062–1068 (1984).
217. G.L. Rosasco, J.J. Blaha, 'Raman Microprobe Spectra and Vibrational Assignments of Talo', *Appl. Spectrosc.*, **34**, 140 (1980).
218. R.J. Capwell, F. Spagnolo, M.A. Desesa, 'A Rapid Determination of Low Concentrations of Anatase in Rutile TiO_2 Pigments by Raman Spectroscopy', *Appl. Spectrosc.*, **26**, 537 (1972).
219. S.K. Sharma, J.F. Mammone, M.F. Nicol, 'Raman Investigation of Ring Confirmations in Vitreous Silica', *Nature*, **292**, 141 (1981).
220. H. Buiteveld, F.F.M.D. Mul, J. Mud, 'Identification of Inclusions in Lung Tissue with a Raman Microprobe', *Appl. Spectrosc.*, **38**, 304 (1984).
221. K. Ashley, 'Electroanalytical Applications in Occupational and Environmental Health', *Electroanalysis*, **6**, 805 (1994).
222. M.L. Woebkenberg, C.S. McCammon, 'Direct-reading Gas and Vapor Instruments', in *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, 8th edition, eds. B.S. Cohen, S.V. Hering, American Conference of Governmental Industrial Hygienists (ACGIH), Cincinnati, OH, 1995.
223. L.A. Todd, 'Direct-reading Instrumental Methods for Gases, Vapors, and Aerosols', in *The Occupational Environment – Its Evaluation and Control*, ed. S.R. DiNardi, American Industrial Hygiene Association, Fairfax, VA, 1997.
224. C.R. Strang, S.P. Levine, 'The Limits of Detection for the Monitoring of Semiconductor Manufacturing Gas and Vapor Emissions by Fourier Transform Infrared (FTIR) Spectroscopy', *Am. Ind. Hyg. Assoc. J.*, **50**, 78 (1989).
225. W.M. Draper, K. Ashley, C.R. Glowacki, P.R. Michael, 'Industrial Hygiene Chemistry: Keeping Pace with Rapid Change in the Workplace', *Anal. Chem.*, **71**, 33R (1999).
226. J.E. Yocom, S.M. McCarthy, *Measuring Indoor Air Quality*, Wiley, New York, 1991.
227. C.H. Lindh, B.A.G. Joensson, H.E. Wilander, 'Direct Measurement of Hexahydrophthalic Anhydride in Workplace Air with a Transportable Fourier Transform Infrared Spectrometer', *Am. Ind. Hyg. Assoc. J.*, **57**, 832 (1996).
228. V.D. Makepeace, C.W. Chase, C.T. Chaffin, T.L. Marshall, P.T. Jaakola, R.M. Hoffman, W.G. Fateley, *Mikrochim. Acta, Suppl.*, **14**, 563 (1997).
229. ASTM D6348, 'Standard Test Method for Determination of Gaseous Compounds by Extractive Direct Interface Fourier Transform Infrared (FTIR) Spectroscopy', in *Annual Book of ASTM Standards*, American Society for Testing and Materials (ASTM), West Conshohocken, PA, Vol. 11.03, 1998.
230. M. Simonde, H.K. Xiao, S.P. Levine, 'Optical Remote Sensing for Air Pollutants–Review', *Am. Ind. Hyg. Assoc. J.*, **55**, 953 (1994).
231. L.A. Todd, 'Evaluation of an Open-path Fourier Transform Infrared Spectrophotometer using an Exposure Chamber', *Appl. Occup. Environ. Hyg.*, **11**, 1327 (1996).
232. NIST, *SRM Quarterly: Standard Reference Materials Program*, National Institute of Standards and Technology, Gaithersburg, MD, 1999.
233. J. Janata, 'Optical Sensors', in *Principles of Chemical Sensors*, Plenum, New York, 241–284, 1989.
234. J.A. Hodgeson, 'Chemiluminescent Techniques for Air Pollution Monitoring', *Toxicol. Environ. Chem. Rev.*, **11**, 81 (1974).
235. USEPA, *Compendium of Methods for Determination of Airborne Pollutants in Indoor Air*, US Environmental Protection Agency, Research Triangle Park, NC, 1988.
236. G.J. Wendel, D.H. Stedman, C.A. Cantrell, L. Damrauer, 'Luminol-based Nitrogen Dioxide Detector', *Anal. Chem.*, **55**, 937 (1983).
237. N.L. Nagda, H.E. Rector, M.D. Koontz, *Guidelines for Monitoring Indoor Air Quality*, Hemisphere Publishing, New York, 1987.
238. P.N. Breysse, P.S.J. Lees, 'Analysis of Gases and Vapors', in *The Occupational Environment – Its Evaluation and Control*, ed. S.R. DiNardi, AIHA, Fairfax, VA, 1997.
239. J.M. Daisy, 'Real-time Portable Organic Vapor Sampling', in *Advances in Air Sampling*, Lewis Publishers, Chelsea, MI, 1988.
240. J.B. Barsky, S.S. Que Hee, C.S. Clark, 'An Evaluation of the Response of Some Portable, Direct-reading 10.2 eV and 11.8 eV Photoionization Detectors, and a Flame Ionization Gas Chromatograph for Organic Vapors in High Humidity Atmospheres', *Am. Ind. Hyg. Assoc. J.*, **46**, 9 (1985).
241. D.R. Walt, C. Munkholm, P. Yuan, S. Luo, S. Barnard, 'Design, Preparation, and Applications of Fiber-optic Chemical Sensors for Continuous Monitoring', in *Chemical Sensors and Microinstrumentation*, ACS

Symposium Series No. 403, eds. R.W. Murray, R.E. Dessy, W.R. Heineman, J. Janata, W.R. Seitz, American Chemical Society, Washington, DC, 1989.
242. F.L. Dickert, S. Thierer, 'Molecularly Imprinted Polymers for Optochemical Sensors', *Adv. Mater.*, **8**, 987 (1996).
243. J.D. Miller, 'Fiber Optic Systems in Potentially Explosive Atmospheres', *Adv. Instrum. Control*, **51**, 421 (1996).
244. J. Peck, M.L. Wilson, M. Rhodes, 'Fiber Optic Fugitive Emission Dosimeter', *Proc. SPIE – Int. Soc. Opt. Eng.*, **2836**, 140–146 (1996).
245. D.E. King, J.D. Webb, 'Infrared Waveguide Sensor with Functionalized Monolayer for Detection of Airborne Pollutants', *Proc. SPIE – Int. Soc. Opt. Eng.*, **2836**, 38–49 (1996).
246. T. Oyabu, S. Hirobayashi, H. Kimura, 'Detection of Gaseous Indoor-air Pollution using a Multisensor System', *Sens. Mater.*, **9**, 177–186 (1997).
247. A.J. White, M.G. Blamire, C.A. Corlett, B.W. Griffiths, D.M. Martin, S.B. Spencer, S.J. Mullock, 'Development of a Portable Time-of-Flight Membrane Inlet Mass Spectrometer for Environmental Analysis', *Rev. Sci. Instrum.*, **69**, 565–571 (1998).
248. G.A. Eiceman, Z. Karpas, *Ion Mobility Spectrometry*, CRC Press, Boca Raton, FL, 1994.
249. A. Graeme, *J. Autom. Chem.*, **20**, 1 (1998).
250. D. Young, C.L.P. Thomas, A.H. Brittain, G.A. Eiceman, 'Extending the Concentration and Linear Dynamic Range of Ion Mobility Spectrometry with a Sheath Flow Inlet', *Anal. Chim. Acta*, **381**, 69–83 (1999).
251. J.P. Dworzanski, M.-G. Kim, A.P. Snyder, N.S. Arnold, H.L.C. Meuzelaar, 'Performance Advances in Ion Mobility Spectrometry Through Combination with High Speed Vapor Sampling, Preconcentration and Separation Techniques', *Anal. Chim. Acta*, **293**, 219 (1994).
252. D.Y.H. Pui, D.L. Swift, 'Direct-reading Instruments for Airborne Particles', in *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, 8th edition, eds. B.S. Cohen, S.V. Hering, ACGIH, Cincinnati, OH, 1995.
253. J. Gebhart, 'Optical Direct-reading Techniques: Light Intensity Systems', in *Aerosol Measurement – Principles, Techniques, and Applications*, eds. K. Willeke, P.A. Baron, Van Nostrand Reinhold, New York, 1993.
254. P.A. Baron, M.K. Mazumder, Y.S. Cheng, 'Direct-reading Optical Techniques using Optical Particle Detection', in *Aerosol Measurement – Principles, Techniques, and Applications*, eds. K. Willeke, P.A. Baron, Van Nostrand Reinhold, New York, 1993.
255. P.A. Baron, 'Modern Real-time Aerosol Samplers', *Appl. Ind. Hyg.*, **3**, 97–103 (1988).
256. P.A. Baron, 'Modern Real-time Aerosol Samplers', in *Advances in Air Sampling*, Lewis Publishers, Chelsea, MI, 1988.
257. P. Kaye, E. Hirst, Z. Wang-Thomas, 'Neural-network-based Spatial Light-scattering Instrument for Hazardous Airborne Fiber Detection', *Appl. Opt.*, **36**, 6149–6156 (1997).
258. E.R. Kennedy, 'Sampling Reactive Aerosols', in *Advances in Air Sampling*, Lewis Publishers, Chelsea, MI, 1988.
259. A.A. Dost, 'Monitoring Surface and Airborne Inorganic Contamination in the Workplace by a Field Portable X-ray Fluorescence Spectrometer', *Ann. Occup. Hyg.*, **40**, 589–610 (1996).
260. M.B. Bernick, P.R. Campagna, 'Application of Field-portable X-ray Fluorescence Spectrometers for Field-screening Air Monitoring Filters for Metals', *J. Haz. Mater.*, **43**, 91–99 (1995).
261. J.C. Morley, C.S. Clark, J.A. Deddens, K. Ashley, S. Roda, 'Evaluation of a Portable X-ray Fluorescence Instrument for the Determination of Lead in Workplace Air Samples', *Appl. Occup. Environ. Hyg.*, **14**, 306–316 (1999).
262. P.M. Eller, M.E. Cassinelli (eds), *NIOSH Manual of Analytical Methods*, 4th edition, Method No. 7105, National Institute for Occupational Safety and Health, Cincinnati, OH, 1994.
263. Y.-I. Lee, Y.-J. Yoo, J. Sneddon, 'Recent Advances in Laser-induced Breakdown Spectrometry', *Spectroscopy*, **13**(7), 14–21 (1998).
264. M.E. Fraser, 'Spark-induced Breakdown Spectroscopy for Metals Analysis', presented at *Pittcon '98*, 1–5 Mar., New Oleans, LA, 1998.
265. K.Y. Yamamoto, D.A. Cremers, M.J. Ferris, L.E. Foster, 'Detection of Metals in the Environment using a Portable Laser-induced Breakdown Spectroscopy Instrument', *Appl. Spectrosc.*, **50**, 222–233 (1996).
266. D.A. Cremers, J.E. Barefield, A.C. Koskelo, 'Remote Elemental Analysis by Laser-induced Breakdown Spectroscopy using a Fiber-optic Cable', *Appl. Spectrosc.*, **49**, 857–860 (1995).
267. B.J. Marquardt, S.R. Goode, S.M. Angel, 'In Situ Determination of Lead in Paint by Laser-induced Breakdown Spectroscopy using a Fiber-optic Probe', *Anal. Chem.*, **68**, 977–981 (1997).
268. J. Wang, K. Ashley, D. Marlow, E.C. England, G. Carlton, 'Field Method for the Determination of Hexavalent Chromium by Ultrasonication and Strong Anion-exchange Solid-phase Extraction', *Anal. Chem.*, **71**, 1027 (1999).
269. J. Wang, K. Ashley, E.R. Kennedy, C. Neumeister, 'Determination of Hexavalent Chromium in Industrial Hygiene Samples using Ultrasonic Extraction and Flow Injection Analysis', *Analyst*, **122**, 1307 (1997).
270. K. Ashley, 'Ultrasonic Extraction of Heavy Metals from Environmental and Industrial Hygiene Samples for their Subsequent Determination', *Trends Anal. Chem.*, **17**, 366–372 (1998).

271. G.J. Williams, N.M. Blaton, O.M. Peeters, C.J. de Ranter, 'The Interaction of Chromium (VI), Chromium (III) and Chromium (II) with Diphenylcarbazide, Diphenylcarbazone and Diphenylcarbadiazone', *Anal. Chim. Acta*, **88**, 345–352 (1977).

272. B.R. James, J.C. Petura, R.J. Vitale, G.R. Mussoline, 'Hexavalent Chromium Extraction for Soils: a Comparison of Five Methods', *Environ. Sci. Technol.*, **29**, 2377–2381 (1995).

Surface and Dermal Monitoring

Shirley A. Ness
Chicago, USA

1 Introduction 1
2 Skin Factors that Influence Permeability 2
3 Sources of Dermal Exposure 3
4 Surface Contamination Evaluations 3
4.1 Standards for Wipe Samples 4
4.2 Surface Collection Methods 4
4.3 Sampling Media 4
4.4 Sampling Strategies 6
4.5 Wipe Sampling 7
4.6 Lioy–Weisel–Wainman Quantitative Surface Sampler 7
4.7 Adhesive Sampling 7
4.8 Vacuuming Methods 9
4.9 Personal Sampling Pump Method 9
4.10 Colorimetric Surface Swypes 9
4.11 Dislodgeable Residues 9
4.12 Volatile Surface Contamination 9
5 Dermal Monitoring 10
5.1 Interpretation of Dermal Sampling Results 11
5.2 Skin Sampling 11
5.3 Direct-reading Methods 12
5.4 Pad Dosimetry Methods 13
5.5 Clothing for Dosimetry 15
5.6 Whole-body Monitoring 16
6 Conclusions 17
Abbreviations and Acronyms 17
Related Articles 17
References 18

Skin contamination can produce the same types of toxic effects as inhalation if a chemical can significantly penetrate the skin. Surface contamination can be ingested orally, create airborne exposures when surface buildup is disturbed, and be transferred into other areas, including from the workplace into the home. When used to determine the potential for dermal exposure, methods are often categorized as indirect or direct. Indirect methods involve the use of filters, gauze pads, or swabs to wipe surfaces; the use of instruments to ''sniff'' surfaces for volatile compounds; and the use of vacuum methods and adhesive tape to lift dust off surfaces. Direct methods measure exposure to the body and include the use of gauze, cellulose, or charcoal pads for dosimetry; analysis of clothing, skin wipes; washing the skin with solvent; and monitoring for skin color change and fluorescence.

Current sampling strategies emphasize air exposures and overlook the contribution of dermal exposures and surface contamination. Integration of surface and dermal monitoring techniques into sampling strategies is important so that a better estimation of exposures and therefore controls to protect against them are developed.

1 INTRODUCTION

Historically, most toxic exposures have been categorized using air monitoring techniques because occupational and environmental standards for chemicals were based on air concentrations. However, air sampling will not measure the total exposure to chemicals capable of penetrating the skin. Skin contamination can occur as a result of fallout from aerosols, via direct immersion into the chemical, as a result of accidental spills onto the body, through vapor penetration of the skin, or from contact with contaminated surfaces. Such contamination can produce the same types of toxic effects as inhalation exposure if a chemical can significantly penetrate the skin. Surface contamination can be ingested orally, create airborne exposures when surface buildup is disturbed, and be transferred into other areas, including from the workplace to the home.

Monitoring for surface contamination has been a standard practice on work sites with radioactive materials since the 1950s, and a similar time span exists for dermal monitoring of pesticide applicators and harvesters. Only since the 1980s have these methods been applied to industrial and residential contamination with any regularity.

When used to determine the potential for dermal exposure, methods are often categorized as indirect or direct. Indirect methods are used to estimate the potential for dermal exposure but not whether an exposure has actually occurred. Such methods involve the use of filters, gauze pads, or swabs to wipe surfaces, the use of instruments to ''sniff'' surfaces for volatile compounds, and the use of vacuum methods and adhesive tape to lift dust from surfaces. Direct methods measure exposures to the body and include the use of gauze, cellulose, or charcoal pads for dosimetry, clothing, skin wipes, washing the skin with a solvent, and monitoring for skin color change and fluorescence.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Dermal studies can identify specific body areas that are being exposed, thus making sure the proper chemical protective clothing (CPC) is selected. They can also identify activities and work tasks within an operation that pose the highest risk of exposure, and assist in determining the proper types of controls needed. In some instances, dermal sampling will identify compounds to which workers are exposed that are not collected on air samples. For example, with polychlorinated biphenyls (PCBs), those with higher vapor pressures tend to volatilize whereas the heavier compounds remain on the skin. Therefore, the relative contributions of dermal and respiratory exposure routes can be assessed. Penetration of CPC and work clothing by contaminants can be determined, in part, by measuring contamination on the interior of garments and also on the skin that was to be protected. The effectiveness of decontamination procedures can also be assessed.

However, dermal absorption must take place in order for contact actually to produce an exposure. A central issue is determining when dermal exposure is likely to be the major route of entry compared with other routes of exposure. It is also important to determine what parts of the body are at greatest risk and to measure dose. The best program is often an integrated approach that uses surface and dermal sampling in conjunction with air and biological monitoring to assess not only the true dose an individual is receiving, but also the sources of exposure.

There is currently debate as to whether external measurement of dermal exposure or biological monitoring represents the best approach to determining exposure. Biological monitoring measurements of contaminants or their metabolites in body fluids (or breath) are considered better measurements of exposure in some cases because they represent an individual's actual *absorbed dose*. However, the individual must have been exposed to and absorbed the contaminant in order to use these methods to determine dose.

Although in some cases biological monitoring is the only effective means of determining whether skin absorption is a significant concern, including detection of contaminants that do not have reliable sampling methods, it is ineffective for other compounds that are rapidly metabolized in the body. Some chemicals have poor biological warning properties or biological indicators of exposure and therefore are not good choices for biological monitoring.[1] As an example, methylenediamine (MDA) easily enters the body through the skin and, once absorbed into the body, it is rapidly eliminated so that using a biological measurement such as the urine may not detect the apparent exposure. Therefore, the option of performing surface and dermal monitoring is important.

2 SKIN FACTORS THAT INFLUENCE PERMEABILITY

The focus of skin exposure determinations is often the permeability characteristics of the contaminant, but other factors relating to the anatomy and the environment are also important. The importance of dermal exposure often depends on its contribution relative to other routes of exposure. Some factors increase the rate or extent of dermal absorption of a compound without impacting on pulmonary or oral uptake, which increases the relative contribution of the dermal pathway to total dose. Therefore, performing dermal exposure evaluations is much more complicated than just identifying those chemicals with high permeation rates.

The structure and condition of the skin and the specific characteristics of a contaminant determine whether a harmful exposure takes place. The skin surface area, anatomical site exposed, duration and frequency of exposure, contaminant absorption rate, fraction absorbed through the skin, concentration of the contaminant, its physical state, and the mixture in which it is contained are all important factors that affect dermal exposure. Individual and job-specific factors include skin hydration from sweating, pre-existing dermatitis, abrasions, the effectiveness of CPC (if worn), and wearing contaminated work clothing for extended periods of time. In addition, the absence or presence of adequate chemical warning properties is important, such as visual stains on the skin or irritation that would make a person more likely to wash the skin promptly, thus limiting exposure.

Commonly called the acid mantle of the skin, a waxy type of mixture composed of sebum, breakdown products of keratin, and sweat coats the outer surface of the skin. It is believed that this coating protects against the penetration of water and water-soluble chemicals, but this protection is minimal and the layer is easily removed by soaps, solvents, and alkaline compounds.

The skin itself is composed of two layers: (1) the epidermis, a layer about 100 μm thick, and (2) the dermis, a layer about 500–3000 μm thick. The skin can absorb chemicals directly through the epidermal cells and through the hair and sweat follicles. The specific skin site that is exposed will also influence the degree of exposure. Certain areas such as the scrotum are more permeable than other areas of the body primarily owing to differences in stratum corneum thickness. Contaminants that permeate the skin can enter the systemic circulation via the capillaries in the dermis.

Important factors that influence the permeation of chemicals through the skin include the compound's solubility, molecular size, concentration gradient on the skin, polarity, pH, and other permeability characteristics that are measured experimentally. Except for compounds

with a molecular weight of 400 g mol^{-1} and more, the molecular size and weight of a compound appear to have less of an effect on the rate or extent of percutaneous absorption than lipophilicity. The best penetrants are those that are soluble in both lipids and water, traversing the skin via both the lipid and aqueous pathways, whereas compounds that are largely soluble in one medium or the other are not as good at penetrating the skin.[2]

3 SOURCES OF DERMAL EXPOSURE

Workers are exposed to a variety of physical states and concentrations of chemicals, including concentrated and dilute liquids, vapors in confined spaces, and contaminated soil and sludge on hazardous waste sites. Exposures do not have to be to large volumes of material to have a significant dermal effect. Many small exposures over a consistent period can build up over time, especially for chemicals that tend to have long half-lives in the body.

Dermal exposures to workers can be categorized as either primary or secondary.[3] Primary exposures occur as a result of direct involvement with the operation and include contact with process-generated contamination, equipment and materials used in the work area, or exposures generated as a result of work activities. The secondary exposures, which are more difficult to identify, occur when workers in turn contaminate surfaces remote from them.

The concern for dermal exposure does not end in the workplace. Environmental contaminants, such as lead and pesticides, can build up on surfaces such as carpets, walls, and counters inside homes and buildings regularly used by the community – schools, day care centers, and public office buildings – and may be associated with a number of different environmental media including water, soil, sediment, and consumer products. Contamination migrating from hazardous waste sites is also a concern. Workers wearing contaminated clothing may expose their families by close contact, contamination of furniture and rugs, contamination of other family clothing in the laundry, and contamination of family vehicles. Finally, hobbies and home industries can also contaminate residences.

The Adult Blood Lead Epidemiology and Surveillance (ABLES) program which is conducted in the USA draws on reports of high blood levels that medical laboratories transmit to health departments or other agencies in a number of states. It has been recognized that parents' exposure in the workplace can be a source of exposure to the children at home, either through lead brought home on personal items or through exposing the children to the workplace. When elevated blood levels are identified in the parents, further information is gathered on children and other at-risk persons in the home.

4 SURFACE CONTAMINATION EVALUATIONS

There are a variety of reasons why surface contamination, especially removable surface contamination, may have to be sampled. Selection of compounds for surface monitoring depends on their ability to cause systemic toxicity via penetration of the skin, dermal irritation, sensitization, cancer, or oral toxicity. Whenever persistent contaminants are present on surfaces or in the air, there is a possibility that workers' skin and clothing may become contaminated, and that these materials will be transported to surfaces in their personal vehicles and their homes, thus exposing their families. This is thought to be the primary reason why blood lead levels of household members residing in homes of workers with occupational lead exposures can be much higher than those of families of non-lead-exposed workers.

The variety of chemicals for which surface sampling can be conducted is large, although the most common groups are amines, metals and other toxic dusts, PCBs and tetrachlorodibenzodioxin (TCDD), and pesticides. Surface sampling for microorganisms is most commonly done to determine the potential for inhalation exposure either as a result of a heating, ventilating, and air conditioning system or by resuspension into the air after being disturbed on a surface.

Wipe sampling is useful for compounds that rapidly absorb through the skin as long as their physical characteristics are appropriate for this collection method. Surfaces that may be contacted by food or other materials that are ingested or placed in the mouth (chewing tobacco, gum cigarettes) may be wipe sampled (including hands and fingers) to identify contamination.[4] Wipe sampling methods have come under scrutiny recently with the result that the methodologies are producing more consistent results and, in at least one instance, the Lioy–Weisel–Wainman (LWW) Sampler may provide reliable quantitative results.

Chemicals for which wipe sampling is ineffective even though they can penetrate the skin include volatile solvents such a 1-butanol, because their evaporation from surfaces can be rapid and sampling will not provide adequate quantitation. Most gases such as bromine and boron trifluoride do not redeposit on surfaces, so sampling for them is ineffective; however, some gases such as arsine and stibine may revert to their metallic forms after contact with surfaces and remain. Owing to their volatile nature, most organic solvents are not suitable for wipes.

When accompanied by close observation of the operation in question, wipe sampling can help identify sources of contamination and poor work practices. Surface contamination may be readily apparent for colored compounds such as lead and cadmium pigments, but it may

not be evident for surface deposition from lead fumes in a radiator shop or bumper (chromium) plating facility, or on dark and colored surfaces. In addition, the effectiveness of CPC such as gloves, aprons, and respirators in these situations can be evaluated by wipe sampling the inner surfaces of the protective gear (and protected skin).

It can be difficult to determine whether equipment or instruments are effectively decontaminated after they have been used in a contaminated area, such as on a hazardous waste site or for sampling of concentrated materials. Wipe and/or ''sniff'' sampling can be useful in determining whether these procedures are effective.

Surface deposits of the volatile type generally cannot be determined through conventional surface sampling methods and other investigatory methods are therefore necessary. In many cases, a field evaluation will depend on whether there is an instrument capable of detecting the contaminant of interest.

4.1 Standards for Wipe Samples

The simplest method for setting up a wipe sample standard is to compare qualitatively areas for relative contamination or use a zero acceptable criterion for areas that are supposed to be clean, such as eating areas and personal vehicles. Guidelines or standards have been developed for some chemicals. In California the United States Occupational Safety and Health Administration (USOSHA) has set a surface sampling standard for 4,4′-methylenebis(2-chloroaniline) (MBOCA). In the USA, the Department of Housing and Urban Development (HUD) has developed a standard for lead abatement wipe samples and criteria have also been developed for asbestos.

Re-entry levels for pesticides on foliage may be appropriate to compare with other pesticide surface contamination such as on equipment and vehicles associated with manufacturing or work ancillary to the application of the material. Re-entry levels are often expressed as waiting intervals (days) to allow sufficient time for dislodgeable pesticide residues to dissipate to lower and thus safer concentrations. Some examples are included in Table 1. In some cases background levels have been used for surface cleanup standards.[5]

4.2 Surface Collection Methods

Surface sampling methods range from the use of hand pressure to move a filter against a surface, vacuuming, coupons, and mechanical devices designed to slide or roll for nonvolatile compounds to the use of direct-reading instruments to measure volatile surface contamination.

The most common surface sampling method used for chemicals is wipe sampling. For simplicity, the term wipe will be reserved for those methods using hand pressure to move media across a surface. Wipe sampling has also been termed swipe, smear, and smair sampling. Another method is the use of paper, cloth, tape, or microscope slides coated with an adhesive to press against the test surface to collect a sample of particles. Other collection media/methods include gauze, commercially available wet wipes, and vacuum cleaners.

Surface characteristics such as porosity or impermeability, smoothness or roughness, and the presence or absence of moisture have a marked effect on the potential for resuspension or release on contact of contaminants. Nonporous or impervious solid surfaces for United States Environmental Protection Agency (USEPA) purposes include metals, aluminum siding, and enameled or laminated surfaces. Porous surfaces include wood, concrete, asphalt, and plasterboard, and can allow contamination to permeate to a considerable depth. In these situations a core sample may be more appropriate than a wipe.

Adhesive tapes have been found to provide a better removal efficiency when taking samples from rough surfaces than wiping with filter papers. Waxes and polishes on surfaces can absorb contaminants and can be more difficult to decontaminate or to use for surface sample retrieval. In the case of a surface from which it is difficult to dislodge material, such as a carpet, vacuuming methods are better than wiping.

4.3 Sampling Media

Laboratory-grade filters are the most common media used. Generally there are two types of filters recommended by the USOSHA for collecting wipe samples.[4] Glass fiber (GF) filters (37 mm, 1.0 μm) are used for materials analyzed by high-performance liquid chromatography (HPLC) and often for substances analyzed by gas chromatography (GC). Methyl cellulose ester (MCE) air sampling filters or cellulose filtration-type filters are generally used for metals, and may be used for anything not analyzed by HPLC. Filters are fragile, and when using solvents extreme care must be used or they will be damaged. Filter papers vary in their ability to absorb materials. A wide variety of commercially available materials have also been used for wipe sampling.

Preweighing allows for the determination of concentrations (micrograms per gram) on media. However, tared wipe media can change characteristics in the field if not handled carefully. Moisture can be picked up or lost, and skin oils can contaminate media or add to the weight. Since MCE filters are very hygroscopic, they cannot be successfully tared for field use.

Media have been used with and without pre-extraction for many compounds, including lead and pesticides. A blank must be analyzed for background contamination prior to selecting the media, however, and along with

Table 1 Guidelines and standards for surface sample results

Compound	Limit	Criteria		Refs.
Asbestos	≥1%	Abatement		6
Asbestos	≥0.1%	Abatement	State of California	
Azinphos-methyl	3.1 μg cm^{-2}	Field re-entry		7
Beryllium	25 μg per 12 in^2	MAC		8
Beryllium	10^{-2} μg cm^{-2}	Inactive and low-activity areas		9
Beryllium	0.5 μg cm^{-2}	Control areas		9
Beryllium	$<5\times10^{-2}$ μg cm^{-2}	Contact surfaces		9
Chlordane	50 μg ft^{-2}, walls			10
Dialifor	0.8 μg cm^{-2}	Field re-entry		7, 11
3,3′-Dichlorobenzidine	10 μg per 100 cm^2			
Dimethoate	53 μg cm^{-2}	Field re-entry		7
Dioxin	55 μg m^{-2}			12
Lead	200 μg ft^{-2}, floor 500 μg ft^{-2}, window sills 800 μg ft^{-2}, window wells 200 μg ft^{-2}, exterior porch	Clearance for lead paint abatement; lead in settled dust in residences		13
MBOCA	100 μg cm^{-2}		USOSHA (CA)	
Mercury	1 μg m^{-2}	Detection limit of monitoring instrument		14
Methidathion	0.6 μg cm^{-2}	Field re-entry		7
Nor-nitrogen mustard hydrochloride	0.5 μg cm^{-2}	Internal company standard		15
Paraoxon	0.02 μg cm^{-2}	Field re-entry		7
Parathion	0.09 μg cm^{-2}	Field re-entry		7
PCBs	100 μg cm^{-2}			16
PCBs	0.5 μg per 100 cm^2	Office building cleanup standard		17
PCBs	1 mg m^{-2}	Decontamination standard for worker re-entry into transformer vault		18
PCBs	100 μg m^{-2}	Decontamination standard for offices in the same building		18
PCBs	10 μg per 100 cm^2	USEPA standard for low-contact indoor impervious surfaces		19
PCBs	10 μg per 100 cm^2	USEPA standard for high-contact solid surfaces indoors		19
PCBs	10 or 100 μg per 100 cm^2 plus encapsulation	USEPA standard for low contact indoor nonimpervious surfaces		19
PCBs	100 μg per 100 cm^2			19
PCBs	10 μg per 100 cm^2	Indoor solid surfaces and high-contact outdoor solid surfaces and indoor electical vault areas		19
PCBs	10 μg per 100 cm^2	Low-contact outdoor impervious solid surfaces		19
PCBs	100 μg m^{-2}	NIOSH standard for surfaces that routinely contact workers' unprotected skin		20
PCDFs	50 μg m^{-2}			21
Phosalone	7 μg cm^{-2}	Field re-entry		7
PCDFs	3 μg m^{-2}		State of California, in Schecter and Tiernan[22]	
PCDDs	3 μg m^{-2}		State of California, in Schecter and Tiernan[22]	
Sodium fluoroacetate	0.2 mg per 100 cm^2	Protection for workers		23
PCDD	≤0.75 μg m^{-2} , exterior building surfaces ≤0.01 μg m^{-2} , interior building surfaces	Decontamination of homes and schools after Seveso incident		24

(*continued overleaf*)

Table 1 (*continued*)

Compound	Limit	Criteria	Refs.
PCDD equivalents	$1\,\mu g\,m^{-2}$	Office building cleanup standard	17
PCDD/PCDF (total)	$24\,\mu g\,m^{-2}$	Decontamination standard for worker re-entry into transformer vault	18
PCDD/PCDF (total)	$3\,\mu g\,m^{-2}$	Decontamination standard for offices in the same building	18

PCDF, polychlorinated dibenzofuran; PCDD, polychlorinated dioxin; NIOSH, National Institute of Occupational Safety and Health; MAC, maximum allowable concentration.

the samples after collection. Media designed to be used as sterile bandages are often very clean anyway. While there have been some concerns about contamination in commercial media, these are considered to be minimal as long as the media are pretested for the contaminant to be sampled the first time they are used, and blanks of the media are included for analysis with the samples. Commercial premoistened hand wipes are frequently used for lead abatement sampling.

There is much debate over solvent selection. Solvents that increase the recovery of material from surfaces may not accurately mimic skin moisture. In this regard, a saline solution or water is considered the most appropriate solvent. Purity of solvents is important. Although they need not be of spectroscopic grade, they should be of sufficient purity not to contaminate the sample. Distilled or deionized water should be used rather than tap water. It is important that the same amount of solvent be used for each sample collection.

4.4 Sampling Strategies

The first step in developing a sampling strategy in which wipe sampling will be performed is to identify as closely as possible the type and expected concentration of contaminant(s). In any one survey it is best to be consistent and use media of the same size and shape. For comparing results from different surveys this is also best. The decision to use wipe sampling as the collection method in a survey should be made after evaluating its advantages and limitations. Wipe samples are inexpensive, fast, and simple to perform. They are useful as a screening tool, especially for trace elements. However, as a quantitative measurement they have several drawbacks, including operator bias and poor reproducibility. They are limited largely to trace elements and other inorganic analyses, and are sufficient for surface loading data (grams per square centimeter) but not for concentration (micrograms per gram) of the analyte in the dust. Wipes are not suited for sampling carpets or other absorbent materials, and the results may not reflect actual skin pickup.

In a single survey it is best to use the same size sampling area for all samples unless the surface or object to be sampled makes doing so impossible. If different surveys are to be compared, it is best once again to use the same sampling area for all chemicals. USOSHA methods use a 100-cm^2 sampling area, which is the most common, but any dimension can be used. HUD specifies a 1-ft^2 area for lead abatement sampling. Where different sized areas must be wiped, the size should be documented.

Templates are commonly used to assure consistency in the size of the area sampled. Both disposable and reusable templates can be used. If reusable, they must be decontaminated between samples, which is one reason why they are often constructed of galvanized or stainless steel rather than absorbent materials such as plastic.

Figure 1(a) shows the size of the template used by USOSHA and Figure 1(b) shows a template being used to take a sample inside of a tank. An alternative to using templates is to cut filter paper to the desired size of the sample, e.g. 10×10 cm, and place the paper over the exact surface to be sampled. The paper is then pressed evenly over the entire surface with a gloved hand.[25] If several or all fingers are used, then equal pressure should be applied by all.

Samples can become contaminated from the hands of the samplers or, worse, sampling professionals might expose themselves to the contaminant. Therefore, clean gloves appropriate for protection against the contaminant being sampled should always be worn. A new set of gloves should be used with each individual sample to avoid contamination of the filter from the hand. Most often surface samples are collected on a one time only basis in order to make a recommendation to improve a program; however, regular repeat surveys to determine the effectiveness of cleaning programs or engineering controls will provide better assurance that they are effective.

Areas that are supposed to be clean, such as changing rooms and lunchrooms, are important to sample regularly. In addition, a number of obvious contamination buildup points should be visited and checked regularly: frequently used doorways, finger plates around essential controls, changing-room surfaces on the plant side of the changing facilities, and operative monitoring equipment in the plants.

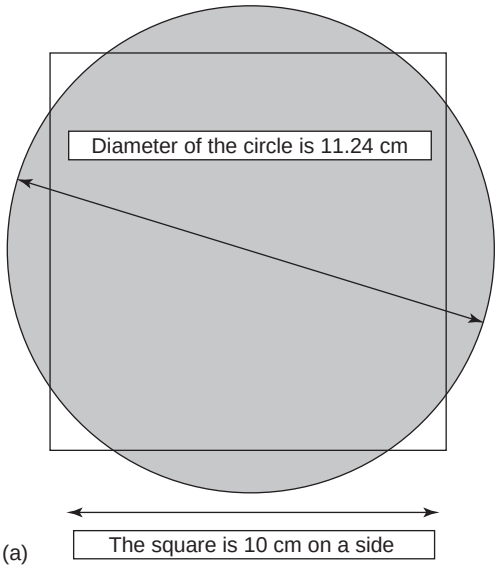

Figure 1 (a) A template for USOSHA sampling and (b) the template in use. (Source: (a) USOSHA[4] and (b) drawing by Renee Joslin.)

If multiple samples are to be taken at a work site, one should prepare a rough sketch of the area(s) or room(s) to be wipe sampled along with the locations of key surfaces. Often air sampling locations are also included on the same diagram. Another sampling approach is to set-up a grid system and collect a specific number of samples inside each grid box. This strategy is commonly used in buildings suspected of widespread contamination. If the intent is to characterize and estimate the exposure over a large surface such as the tables in a lunchroom, collecting many individual samples for separate analysis and calculating the average of all the samples may provide a more useful estimate of the potential exposure. Alternatively, after several samples have been collected they can be composited for a single analysis. Compositing can make it easier to compare the sampling results of several dwellings.

In situations where the detection level for a contaminant is very low, several wipes may be collected, composited, and treated as a single sample. Samples for dioxins and dibenzofurans are often collected in this manner. In some instances where a large area must be sampled, compositing can be done to represent an entire surface such as a ceiling or each wall in a room.

4.5 Wipe Sampling

Various strategies for wiping exist (Table 2): wiping across the surface in either an "S" or a "Z" motion; wiping across horizontally or up and down; using a series of progressively decreasing circles or squares to fill up a circle or a square. To some degree the shape of the template will dictate the motion: A circular template would be done with a circular motion, and a square with either a square, horizontal, or vertical motion. Care should be taken not to rewipe any sections that have been wiped previously.

4.6 Lioy–Weisel–Wainman Quantitative Surface Sampler

A flat surface wipe sampler has been developed to quantitatively measure the concentration (micrograms per gram) and surface loading (micrograms per square centimeter) of dust on flat surfaces.[30] The LWW wipe sampler consists of two sections, a template and a movable plate–filter retainer. It uses a template to map a specific area for the quantitative collection of dust and to control the movement of a collection plate. The moveable plate traces the rectangular opening within the template. The filter medium used to sample a surface is a ragged polyethylene drain disk material that is cut into rectangles of 3.8×6.35 cm. Self-closing forceps are used when handling the filters.

4.7 Adhesive Sampling

Cloth, tape, labels, and cellophane tape with adhesive backings have all been used for sample collection. The primary use of tape methods is to assess surface contamination of dusts, primarily in the asbestos industry, although the methodology was developed to collect dust for environmental and forensic purposes.[37]

Table 2 Wipe sampling methods[a]

Method	Surface type	Wipe media	Area	Sampling procedure[b]	Shipping container	Units	Ref.
Proposed ASTM Standard	Hard: floors, windows		Template (12 × 12 in) or measured area	Wipe with "S" motion L → R, F → B. Fold wipe. Repeat "S" motion L → R, F → B	Centrifuge tube (50 mL)	$\mu g\,ft^{-2}$	26
Georgia Tech	Hard	Little Ones baby wash cloths	Measured area	Wipe with "S" motion F → B, L → R. Fold wipe. Repeat "S" motion F → B, L → R	Centrifuge tube (50 mL)	$\mu g\,ft^{-2}$	27
Farfel	Hard	Nonalcohol wipes	Template (12 × 12 in) or measured area	Wipe with "S" motion F → B, L → R. Fold wipe inside. Repeat "S" motion	Ziploc™ bags	$\mu g\,ft^{-2}$	28
HUD	Hard	Commercial nonalcohol premoistened wipes	Template (12 × 12 in) or measured area	Wipe with "S" motion F → B, L → R. Fold wipe inside. Repeat "S" motion 90° to first wipe	Polypropylene tube (50 mL)	$\mu g\,ft^{-2}$	13
Rabinowitz		Preweighed filter paper, Whatman 541	930 cm^2		Resealable polyethylene bags	µg per wipe	29
LWW	Hard	Three preweighed polyethylene filters (3.8 × 6.35 cm)	Template (27.3 × 4 cm)	Wipe within template with filter No. 1 (moistened with Type 1 water). Wipe within template with filter No. 2 (moistened with Type 1 water). Wipe within template with filter No. 3 (dry)	Plastic bag	$\mu g\,ft^{-2}$ $\mu g\,g^{-1}$	30
NIOSH	Hard	Gauze (2 × 2 in) wetted with 1–2 mL distilled water. Wash'n Dri wipes (individually packaged)	Disposable template (10 × 10 cm)	Wipe with "S" motion three or four vertical strokes. Fold wipe inside. Wipe with "S" motion three or four horizontal strokes. Fold wipe inside. Wipe with "S" motion three or four vertical strokes	Ziploc™ bags	$\mu g\,ft^{-2}$	31
Stark		Preweighed cotton gauze				$\mu g\,g^{-1}$	32
State of Massachusetts	Hard	Individually wrapped alcohol medical wipes	Template (1 ft^2)	Wipe area once L → R	Plastic tube	$\mu g\,ft^{-2}$	33
State of North Carolina	Hard	Johnson & Johnson Baby Wipes	Template (1 ft^2) or measured area	Wipe with "S" motion. Rewipe with "S" motion 90° to first pass	Ziploc™ bags	$\mu g\,ft^{-2}$	34
Vostal	Hard (uncarpeted)	Disposable paper towel (14 × 20 cm) moistened with 20% denatured alcohol or 1 : 150 benzalkonium chloride	Template (1 ft^2)			$\mu g\,ft^{-2}$	35

[a] Source: USEPA.[36]
[b] L → R = left to right and F → B = front to back. ASTM, American Society for Testing and Materials.

4.8 Vacuuming Methods

Vacuum methodologies have become an important aspect of surface sampling especially for collecting residential samples, and, therefore, the methodologies have been refined at least as much as those for wipe sampling, if not more. Vacuum methods most often involve the use of an elaborate apparatus many of which were developed under USEPA contracts. Vacuum methods are more effective for sampling carpets than filter wiping methods.

The advantages of vacuum methods include the collection of adequately sized samples providing enough material for detailed analyses, larger sample volumes allowing for more repetitions during sampling, providing better sample area coverage, the direct comparison of indoor dust concentrations to those in outdoor soil, and the ability to obtain surface dust loading and concentration data. Disadvantages can include the collection of dust that is inaccessible (and, therefore, nonrepresentative) for exposure, variable sampling efficiency between devices and for some surfaces, the requirement for sampling apparatus that is more costly than typical wipe media, and the need for more sampling time (including time to clean and decontaminate the devices between samples).

4.9 Personal Sampling Pump Method

The personal air sampling pump and filter cassette method has been standardized and is often referred to as the University of Cincinnati (DVM) method. The sampling pump is attached with Tygon tubing (3 mm internal diameter) to a 37 mm diameter three-stage cassette with a 0.8 μm MCE or other filter inside. Flow rates range from 2 to 3 L min^{-1}. In general, this method is limited to the collection of stable pesticides and metals. The most common method has a second piece of Tygon tubing attached to the front of the cassette as a suction head.

4.10 Colorimetric Surface Swypes

Surface ''swypes'' are sold by CLI Colorimetric Laboratories. They are absorbent pads impregnated with colorimetric reagent indicators to provide a visual determination of contamination on work surfaces and skin. A color indication within 2–3 min of placing the wipe in solvent is a positive result. The solvent is drawn up through the pad via capillary action, concentrating the contaminant in the colorimetric detection zone. A clear, plastic backing on the pad allows for the attachment of a handle and prevents the solution from contacting the user's skin.

4.11 Dislodgeable Residues

Several methods are used to measure aerosol deposition: foam strips, tape, and surface vacuuming for dust or spray that has already settled. Tape samples are used primarily for surface asbestos contamination, but can also be used for other dusts. Vacuum methodologies have improved and, when used with prescribed methodologies, several devices can produce quantitative results. A standardized method for carpet samples has been developed by the American Society for Testing and Materials (ASTM). Occasionally a case can be made for collecting a bulk sample rather than wiping or vacuuming a surface, for example when a mechanical or manual device is used regularly to vacuum or sweep surfaces and the dust is derived from a floor or carpet. Dust has also been collected from carpet sweepers and used to represent surface contamination. Dislodgeable residues are collected using several methods, including the polyurethane foam (PUF) Roller, Dow Drag Sled, CFDA Roller, and, for foliage, coupons and leaf punches.

Acids and bases (or alkalis) can be detected by their reaction with pH paper. Dislodgeable residue methods measure that portion of surface contamination likely to be removed by human contact. They are frequently used to sample carpets in residential studies. An example of these apparatus is the PUF Roller. The PUF Roller consists of a weighted roller fitted with a PUF cover that simulates the pressure applied to a surface by a crawling child weighing 9 kg (20 lb). The matrix of PUF is more similar to skin than the media used with other devices for dislodgeable residues.

4.12 Volatile Surface Contamination

When does a measurement for a volatile compound fall into the surface category rather than air sampling? Although this is a gray area, when a measurement is 1–12 in (2.5–30 cm) from a surface, it can be assumed that the source is the surface – unless there is a process that is actively generating the contaminant nearby. If measurement is in the breathing zone or in the middle of a room (general area) away from a contaminated surface, it can be categorized as air sampling rather than surface sampling. As is the case with other contaminants, surface sampling of volatile materials is usually done to determine if skin or clothing exposures are a concern. For many volatile contaminants, a general survey monitor with a photoionization detector (PID) or a flame ionization detector (FID) can be used. In this case it is important to remember that unless the instrument is calibrated with the contaminant in question, the results are all relative to the response of the instrument. In some instances screening for the presence of volatile materials on a surface can be useful to determine if further testing or decontamination is necessary.

5 DERMAL MONITORING

There are two types of methods for sampling dermal exposures: those that assess exposure indirectly by monitoring for surface contamination and those that directly assess the degree of contamination on the individual's skin. Indirect methods, as discussed previously, include wipe sampling whereas direct methods are often experimental and have used cellulose, gauze, or charcoal pads, clothing for dosimetry, skin washes, and monitoring the skin for contaminants that fluoresce.

Pesticide manufacture, formulation, and field application were one of the first situations where dermal monitoring was performed, and the use of dermal monitoring has been expanded to virtually all areas involving pesticides, including field re-entry, harvesting, greenhouse work, office buildings, and residences. As a result, the methodologies that have been standardized for the use of dosimetry pads and whole-body dosimetry arose from pesticide monitoring.

Skin sampling can be done using methods that require laboratory analysis or using methods that produce either color changes upon exposure or fluorescence under ultraviolet (UV) light. Hands are the most common application for skin sampling methods when they are the primary source of exposure. Skin sampling is also useful to determine if cleaning methods are effectively removing skin contamination. Sampling under gloves and other CPC can identify previously unsuspected exposures and estimate the degree of protection achieved during actual field use.

Of the techniques that have been developed to monitor for dermal exposure, pad or patch dosimeters have been used the most. With pad monitors, contaminants are entrapped or absorbed by the cloth mesh of the multilayered materials of which the pads are composed. A common method involves placing absorbent pads at various points on a worker's clothing and skin and exposing them for a carefully measured interval while the worker performs routine functions (Figure 2). The pads are removed and transported to a laboratory, where the contaminant is extracted and analyzed. The amount of contaminant that comes into contact with the skin is calculated using an anatomical model that allows for extrapolation of the small area represented by the pad to one or more body parts.

Pads can measure the effectiveness of personal clothing for protection against chemicals, estimate skin exposure, and determine which parts of the body receive the highest exposure. Pads fall into several categories. The most common have been developed to collect a contaminant in specific physical form: spray or dust. Multiuse pads designed to collect either dusts or sprays have also been developed. More recently, chemical dopants and other

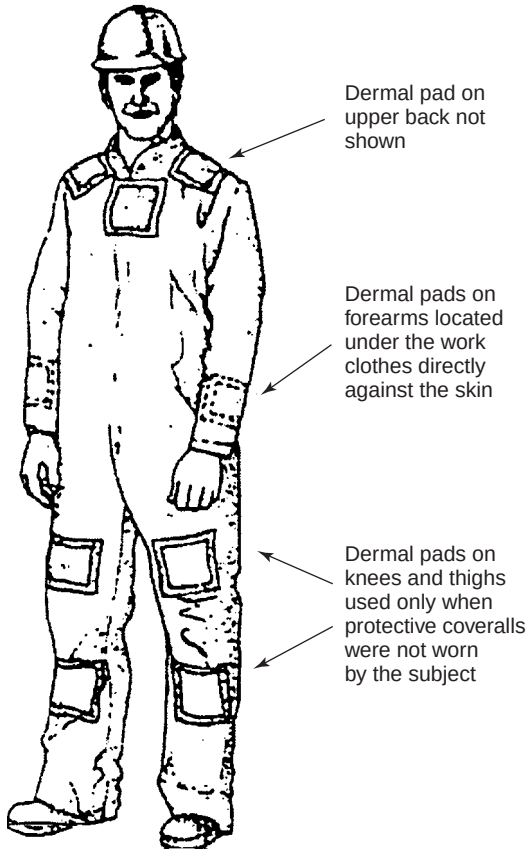

Figure 2 Badges attached to worker using USEPA configuration. (Source: USEPA.[38])

materials such as polyurethane have been used to collect specific chemicals or chemical families. α-Cellulose is commonly used for measuring sprays and surgical gauze is used for measuring dusts.

Pads have several advantages including their small size, portability, low cost, and passive sampling, which eliminates the need for pumps that require calibration and maintenance. The primary limitation of using absorbent pads is the assumption that exposures are uniform over various body parts. Since the pad technique collects a spot sample of the exposure to each body region, a potential source of error in using pads for sample collection is in the extrapolation from concentrations measured on a relatively small surface area compared to the entire body surface area.[39]

A variation of the absorbent pad method involves knitted cotton or other garments that cover the anatomical areas expected to have an exposure. These garments are then extracted in a laboratory for contaminants and the extract is analyzed. The most common method uses cotton gloves for collection and some studies have used long underwear.

Clothing samples fall into two categories: items deliberately prepared and selected to be used as monitoring media and items of normal work clothing that are identified during surveys as highly likely to contain contaminants. These are then collected and either sacrificed for sampling or subjected to a nondestructive method such as wiping, sniffing, or vacuuming which is used as a qualitative determination of the presence of contaminants.

The whole-body dosimetry method is best used for preventive sampling, that is, to determine if there is significant skin exposure and to what areas of the body it is occurring. There are two methods: (1) using undergarments of cotton, long underwear tops and bottoms, and socks to assess the amount of material penetrating outer clothing; and (2) using disposable overgarments such as Tyvek® suits, which estimate exposure as if the skin were bare.

An alternative is to use portions of clothing to cut down on the bulk of material requiring analysis. This usually involves cutting away the most likely areas to show contamination either before monitoring or afterward. Cotton sleeves (covering the arm from the top of the arm to the wrists) and cotton leggings (covering the leg from the top part of the thigh to the ankle) washed before use have been used for sampling workers who were picking strawberries.[40] In between these extreme lies the use of specific items of clothing, including gloves, T-shirts, and coveralls. If workers wear leather shoes it is possible for contamination to build up in them. In this case, foot exposure should also be assessed, which has been done by analyzing cotton socks.

A major obstacle is the lack of standard garments for sampling. In studies to date, garments have been selected for convenience rather than according to scientific criteria, and garment characteristics (fiber content, construction, finish, weight, thickness), of course, vary.[41]

5.1 Interpretation of Dermal Sampling Results

With few exceptions, there are no standards related to dermal exposures; each situation must be handled on a case-by-case basis. The type of health effects caused, the contribution of skin absorption or ingestion to the total dose, and air sampling results (where appropriate) must be considered when evaluating sample results. The methods for calculating dermal exposure are fairly simplistic and frequently do not account for the use of protective clothing and its effectiveness. However, when measurements of dermal exposure are used to estimate absorbed dose, the amount of the contaminant penetrating clothing and estimates of dermal absorption data obtained in animal experiments should be incorporated into the result.

When interpreting results for dermal samples, it is generally assumed that 100% skin absorption has occurred unless the experimenter has enough information to adjust penetration using experimentally derived factors such as partition constants (e.g. the oil/water partition coefficient, $K_{o/w}$). However, direct dermal absorption from contact with a treated surface or deposited soil, dust, or liquids containing contaminants is difficult to assess.

5.2 Skin Sampling

When planning to sample the skin it is extremely important that the skin is cleansed before the day's exposure begins and, following exposure, skin cleaning is even more important. Prior to surveys where wiping, swabbing, or washing will be used to sample the skin, subjects should be requested not to apply lotions or creams that day as they may interfere with the analysis or interact with the contaminant being sampled.

Sampling strategies most commonly attempt to identify whether a particular region of the body is being contaminated. Occasionally more than one area is sampled at a time, such as the forehead and hands. Most often skin samples are collected before lunch and at the end of a work shift or, in the case of community monitoring, on a one-time basis. However, another strategy that can be used in the workplace is to collect samples that reflect the exposure presented by specific tasks.

The neck is generally exposed and is not washed during the course of the day, hence it is a good sampling area. The forehead may have a higher exposure potential because chemical absorption may be increased due to the reduced skin thickness of this area. The USOSHA recommends that its compliance officers wipe only hands and fingers if doing skin wipes.[4] Before any skin wipe is taken, the compliance officer must explain why the sample is necessary and ask the employee about any possible skin allergies to the chemicals in the sampling filter or media.

A problem with any skin sampling procedure is to find a solvent that will provide adequate removal of residues without causing injury to the skin or increased absorption of a contaminant. Most often the solvent used to sample direct skin is water. Solvents should be selected that are not irritating, do not cause drying or cracking of the skin, and are not readily absorbed by the skin. In addition, they should not facilitate skin penetration of the material being sampled.

Commonly used solvents to rinse contaminated workers' hands are 95% ethanol, 10% 2-propanol, acetone, distilled water, or water containing detergent, or some combination of these three materials such as 50 : 50 ethanol–water.[42]

Wiping methods for skin use the same types of media as are used for surfaces, such as filters, gauze, and commercial premoistened wipes. Given the concerns regarding solvent selection in the previous section, the simplest option is to wipe the skin with a dry filter. In some instances the skin will be moist with perspiration and the use of a solvent will only dilute the sample. Depending on the compounds for which analysis will be performed, perspiration can present problems.

Similar to surface wipes, skin wipes should be done on equal-sized areas so that comparisons can be made not only of different areas on the same individual, but also of different individuals. One way to determine if the source of a contaminant on skin is related to the immediate environment is to wipe control subjects (who are in known contaminant-free environment) and compare these to skin samples of exposed persons. There may be a limitation to the number of times a hand can be wiped or washed per day because skin moisture and oils can be depleted.

A disadvantage of wipe methods to detect semivolatile mixtures such as polycyclic aromatic hydrocarbons (PAHs) is their inability to collect the most volatile fraction. Skin wipes can be collected using Whatman Smear Tabs (No. 50). The wipe samples are analyzed using the same procedure as that used on the filter wipes for surface contamination. A variety of premoistened commercial materials have been used to perform skin wipes. Hand and face wipes for lead exposure have been performed with a commercially available paper towel premoistened with benzalkonium and alcohol. The method that is frequently used to performed hand wipes, especially for lead, utilizes premoistened commercial baby wipes and is as follows.[31] The first wipe in the package is discarded. The subject is asked to pull the second wipe from the package. The subject is then instructed to wipe both hands thoroughly for 1 min, after which time the used baby wipe is placed into a clean labeled Zip-Loc™ bag. It is not necessary for the sampler to wear gloves during this procedure because the sample wipe only comes in contact with the subject.

The USEPA recommends skin washing for hand exposure estimations. Skin washing is considered most appropriate for chemicals that have low rates of dermal absorption. Skin washing has been used for the hand, wrist, arm, foot, and ankle, but most commonly the hands are washed. Some studies have rinsed the entire arm (wrist bone to shoulder), the torso, and portions of the legs, but swabs are usually used for larger parts of the body.

Rinsing techniques may be more representative of actual exposure than surface (pad, cotton glove) methods because the skin is subjected to more variables such as sweating and changes in body temperature. However, the effectiveness of this method is severely limited by the rapid rate at which the skin absorbs and retains certain compounds, such as organophosphates. Therefore, skin rinsing may underestimate the potential dose, because only the unabsorbed or slowly absorbed fractions of a contaminant may be collected.[43]

The bag rinse procedure is the most common rinsing method and involves the use of polyethylene bags containing a suitable solvent, such as distilled water, for washing the contaminant from the hands.[44] The most common intervals to select for rinse samples are morning break, lunch, afternoon break, and after work. Sometimes a sample is collected before the start of work to determine if there is residual contamination. Bags are selected that are sturdy enough to hold 200 mL of solvent. The bag rinse may be more effective than the swab method for removing hand contamination. Three sequential rinses have been shown to be optimum to assure that most of the contaminant has been removed from each hand. For analysis, rinses of the same hand are usually composited.

Hand washes can be performed as simply as releasing 250 mL of distilled water or other solvent from a separating funnel or other container held over the hands, and collecting it in a stainless-steel bowl.[45] The wash water is transferred to glass jars that were covered and sealed. Washing the hands in a stream of solvent or in a basin containing solvent is faster than shaking the hands in plastic bags containing solvent, but is not as standardized or as convenient as the bag rinse. Sunflower oil has been used to wash hands for PAH exposure. This may also be suitable to collect other oil-soluble, sticky materials which will not rinse off easily with a solvent. Hand rinse solutions can be quantitatively transferred to volumetrics and diluted to a 100 mL volume for analysis.

Preweighed adhesive labels have been used to sample a standard area on the palms of children for lead exposure from contaminated soil.[46] Tape has been used to collect fibers directly from skin when employees in a data-processing computer room were complaining of itching and localized rashes.[47]

5.3 Direct-reading Methods

Direct-reading methods are available for skin monitoring for PAHs, amines, and metals. At the simplest, monitoring can be examination of a worker's skin for obvious discoloration since certain compounds such as methylenedianiline will cause this effect. The best known direct-reading method is fluorescence, which has been used to identify dermal contamination in two situations: when compounds such as PAHs possess a natural fluorescence and when fluorescent compounds have been deliberately added to chemicals so that their contamination can be visibly traced. UV light is used to visualize the contamination in both cases. The technique consists

of scanning the body of a worker with a hand-held UV lamp and examining the fluorescence in a dark room. Photographs are often used to permit more careful study later and comparisons among subjects.

Video imaging has been integrated into studies using fluorescent tracers, and although it is not yet widely used, it may be a primary method in the future since air sampling studies are also incorporating this technique. An important advantage of the technique is its ability to provide an assessment of the affected body areas.

Although UV light has been used in several plants to detect skin contamination, there is still concern about the risk of skin sensitization and carcinogenic effects.

5.4 Pad Dosimetry Methods

Dermal sampling using small pads or patches attached to various areas of the skin or clothing is one of the most standardized methods for collecting chemicals. Pad samplers of cellulose materials such as filter paper and multilayered gauze are appropriate for sampling contaminants with low vapor pressure, when exposure is the result of airborne mists or dusts. For volatile materials, pads containing charcoal have been developed, and for certain semivolatile materials, polypropylene pads have been used. Although they are not as well tested as the cellulose pads, their use is promising because they may provide for greater collection efficiency and a broader number of compounds for sampling. Other variations include the use of multiple materials and adding chemical dopants to enhance collection.

Pads must be absorbent enough to retain all the liquid that contacts them or, if used for collecting dusts or dried residues, must be porous enough to collect these materials. The pads must be strong enough to hold up under the abuse that they will receive in the field. They must not contain additives such as sizing that will interfere with the chemical analysis of the extracted chemicals. Additives can usually be removed by pre-extraction of the pads; however, some interfering compounds cannot be adequately removed to allow analysis of low levels of contaminants. The following are variables in pad use: size, type of medium, combinations of media, backing(s), number of layers, pre-extracted versus not extracted, holder, means of attachment, number of pads used for monitoring, sampling locations, overlapping versus adjacent pads, sampling period, and whether combined or analyzed separately.

The USEPA recommends that pads be constructed from papermaking pulp or a similar material, approximately 1 mm thick, which is commonly called α-cellulose, especially if sprays are being collected. It resembles blotter paper. Gauze pads are recommended for dust collection. They can be assembled from individually cut layers of material or used as off-the-shelf medical supplies, such as surgical dressings or sponges. The compositon of a typical gauze pad for dermal sampling is layers (often 8, 9, 16, or 32) of surficial gauze with an impervious backing. Soft polypropylene has been used as an adsorbing material for PAHs and may also be useful for sampling other semivolatile materials.[48] Gauze pads have been doped with glycerin to enhance the collection of MBOCA.[49]

The size of pads can vary, although commonly used sizes are 3×3 in, 4×4 in, and 2×2 in (1 in = 2.54 cm). Different sized pads may be used for monitoring different sized body parts. The purpose of backings is to prohibit contamination of pads by skin oils and perspiration from beneath the dosimeter. Glassine (weighing) paper, polyethylene plastic (4–10 mL), filter paper, brown wrapping paper, and aluminum foil are commonly used as backings. A multiuse pad for situations where both dusts and sprays are being generated either at the same time or sequentially, such as when pesticides are mixed and then applied by the same person, consists of one or more layers of filter paper covered with multiple layers of gauze, e.g. eight-ply surgical gauze stapled to a double layer of filter paper with a backing of manila poster paper.[50] The filter paper readily absorbs liquid while the gauze collects impinging dry residues and the poster paper provides extra mechanical strength.

To hold the pad together during sampling there are a number of options, ranging from a simple staple in the center to surrounding it with tape or placing the pad inside a specially made holder. Holders have been made by cutting a hole in one side of a glassine paper envelope. The finished pad should be bound so that the maximum sampling area is exposed.

A protocol should be developed prior to the start of the study that sets the criteria for collecting samples taking into account the goals of the survey. In order to eliminate the contribution of previous contamination from workers' clothing to a new sampling day, a standardized set of laundered clothing can be issued to all subjects at the start of each day's operation.[51] At the end of the exposure period the clothing is collected along with the pad samplers. When pads are used on top of disposable clothing, this clothing is also replaced between applications. The number of workers selected to be sampled and, where multiple sites are involved, the number of sites selected to develop a representative sample must also be decided. In general, it is best to monitor as many different individuals as practicable at any given time.

There are four basic strategies for locating pad dosimeters: (1) measuring dermal exposure or deposition on the body due to the environment only, in which case dosimeters are placed outside the work clothing only; (2) measuring dermal dose to skin, in which

case dosimeters are placed on the skin under the clothing; (3) measuring both skin and work clothing exposure, which also allows one to measure the protection afforded certain clothing; and (4) measuring the degree of protection provided by CPC, in which case sometimes three layers of pads are used.

As for standardized methods, the USEPA has two different pad location scenarios: one for sampling workers *not* wearing impervious CPC and another for when CPC is worn. The World Health Organization (WHO) and the National Agricultural Chemicals Association (NACA) have also developed protocols (Table 3). When only work clothing and no impervious CPC is worn, the USEPA recommends that at least 10 dermal pads be attached at the following locations: in front of the legs just below the knees, in front of the thighs, at the back of the forearms, on top of the shoulders, at the back of the neck at the edge of the collar, and on the upper chest near the jugular notch.[54] If a CPC suit is being worn, only six pads are attached. The most common placement is to offset the interior pad so that it is right next to the exterior pad on top of the clothing so that both pads sample similar exposures. These extra pads should be attached to areas of the body that are expected to receive the highest exposure.

If pads are placed both inside and outside normal work clothing, the ability of the fabric to protect against the contaminant can be determined by comparing the outer and inner monitored values. It is important to note that the workers' shirts and trousers are often constructed of different materials and, therefore, comparisons of penetration between workers may be difficult. While the standardized methods for pad placement are useful, the specific needs of the survey may require other variations. The final selection of sites will depend on the potential for exposure during a specific activity. If exposure is expected to be limited to the arms, and it has been demonstrated in prior studies of the same task, then there is no benefit in sampling the head, chest, back, etc.

There are a variety of ways to attach pads to clothing. Some studies have utilized specially designed harnesses or lightweight vests fitted with open-front pockets to hold the shoulder, chest, and back pads in position. Chest pads have also been placed inside shirt pockets. When pads are placed under garments, various approaches are: to tape the pad to the skin under the garment, pin the pad under the same clothing to which the outer pad is attached, or pin or tape the pad to a T-shirt underneath the clothing. Surgical tape used to attach pads to the skin is generally nonirritating to perspiring skin and tolerable by the wearer.

Monitoring media should be changed during some natural break in the worker's schedule. Pad replacement is done when different tasks are to be monitored, when heavy buildup is suspected, or when gathering information on morning versus afternoon exposures.

Attaching pads to disposable jackets, coveralls, and hats minimizes the time required for changeovers for resampling. Another way to speed up pad exchange is to staple pads to the outside of fresh white cotton T-shirts and put the shirts over the subject's work clothing at the beginning of the exposure period.

A photograph of each worker in sampling garb is useful for estimating the amount of skin exposed, type of work clothing worn, and CPC worn. Extensive notes (Figure 3) should be taken during sampling including temperature, humidity, wind velocity and direction, duration of sampling, type, concentration, and amount of chemicals present, type of equipment used, and any other parameters relevant to the process. Typical observations often made of the individuals being sampled are physical characteristics (sex, age, weight, height, and body surface) and individual productivity. It has been suggested that age correlates positively with productivity, possibly because older workers are more experienced and motivated.[54]

At the end of the sampling period, the pads are carefully removed from the worker. Care should be taken not to touch the sampling portion of the absorbent pad. Samples can be handled by their outside perimeters, or clean latex gloves worn or forceps used to remove them. When feasible, fold the dosimeter onto itself before sliding the pad into its original prelabeled plastic bag, and seal or place two dosimeters into the bag face-to-face if they

Table 3 Comparison of USEPA, NACA, and WHO sampling locations[a]

Body region represented	USEPA[19]	NACA[53]	WHO[b][52]
Head	Shoulder, back, and chest	Front and back	Head
Face	Chest	–	–
Back of neck	Back	–	–
Front of neck	Chest	–	–
Chest/stomach	Chest	Chest	Sternum
Back	Back	Back	Back
Upper arm	Shoulder and forearm/ upper arm	Upper arm	–
Forearm	Forearm	Lower arm	Midway between elbow and wrist
Thigh	Thigh	Upper leg	Thigh
Lower leg	Shin	Lower leg	Lower leg
Feet	–	–	–

[a] Source: Curry and Iyengar.[53]
[b] In WHO guidelines, pads for lower leg, thigh, and lower arms are placed on only one side, depending on whether the subject is right- or left-handed.

Company Name				Contract Number	Date (Mm/Dd/Yr)
Site ID				Work Assignment No	Substance Monitored
EMPLOYEE AND WORK AREA DATA					
Employee Name				Job Title/Work Duties:	
PERSONAL PROTECTIVE EQUIPMENT					
Eye Protection				NOTE: Specify type, material, brand, etc.	
Boots/Shoes				Head Protection	
Respirator				Gloves	
Protective Suit/Coveralls					
DERMAL SAMPLING DATA					
Dermal Pad Media:					Lot No.
Body Region	Start Time	Stop Time	Duration	Sample Numbers	
Back					
Chest					
Forearms				Left	Right
Shoulders				Left	Right
Thighs				Left	Right
Shins				Left	Right
Hand Rinses Solvent: DISTILLED WATER				Left	Right
				Left	Right
				Left	Right
Signature				Date	Checked by:

Figure 3 Dermal assessment data sheet. (Source: USEPA.[53])

are to be analyzed as one sample. Place all the bags from a single test subject in a sealed container such as a wide-mouthed jar or another plastic bag.

There are two different ways of analyzing pads. One method is to combine pads from opposite sides of the body prior to extraction. For example, both forearms would be combined as well as both legs. However, if exposure values from particular pads are needed for calculations, these pads must be extracted individually. Calculations are generally performed in order to extrapolate the results of pads over the specific anatomical areas of the body their sampling locations represent.

5.5 Clothing for Dosimetry

Whole items of clothing used as samplers eliminate the need to extrapolate an exposure from a smaller collecting device such as a pad over the body surface. Clothing items are usually prepared and placed on the subject for the specific purpose of sampling, although workers' personal clothing has also been collected for analysis. Gloves have been used for the hands, short-sleeved undershirts for the upper part of the body, socks for the feet, and the tops of socks for both ankles and wrists. Work clothing has also been analyzed after laundering to determine the effectiveness of laundering. Whole items of clothing usually contain material that must be extracted prior to use and require large amounts of solvent for extraction of contamination.

It is very difficult to change clothing or underclothing after each exposure period without interfering with the work pace. Also, it is nearly impossible to remove these garments without contaminating them with residues from other areas of the body, especially those on the workers' hands and hair, unless special precautions are taken. Occasionally a garment may have to be cut off the subject in order to avoid its contamination by some part of the body. T-shirts fall into this category. Clean, disposable gloves should always be worn when removing and handling items of clothing used for sampling. Clothing used as samplers has the advantage of decreasing exposure to the skin; however, it can saturate and, once contaminated, become a source of continuing exposure. Lastly, covering contamination already present on the hands or body with dosimeter clothing may result in an increased exposure of the test subjects due to occlusion, so hygiene is important.

Absorbent gloves are a popular method for measuring contaminants contacting the hands. Two types of lightweight absorbent gloves are used most frequently: white 100% cotton gloves which are typically white cotton twill work gloves with knit wristlets and seamless palms, and white nylon gloves similar to those used for fruit picking. Nylon gloves are lint-free and some studies have found them easier to extract than cotton gloves.[56] Leather gloves have been used for sampling, especially if they were typically used in a specific job.

Prior to use, gloves are usually prepared in a manner similar to dosimetry pads to remove interferents. They may be laundered one or more times with detergent, bleached, and then extracted with a solvent such as acetone or a solvent blend (hexane–acetone) one or more times depending on the level of analytical detection required. The amount of time each subject wears the gloves should be recorded. Gloves should be replaced whenever obvious staining or moisture has soaked into them.

As with dosimetry pads, the gloves can be composited or extracted individually for analysis. Sometimes only sections are cut from the back and palm of gloves or they can also be cut into pieces.

Gloves are often used when hand rinses are not acceptable. Lightweight absorbent gloves also provide several advantages over rinsing or swabbing for the assessment of hand exposure. It is easy to exchange gloves in between monitoring periods. Absorbent gloves are the only method for monitoring the hands that will collect residues that would otherwise be absorbed into the skin during the exposure period, although if the exposure period is too long or concentrations are too great breakthrough may occur.

The primary advantages of using gloves are ease of use and the fact that the hands represent one of the most highly exposed dermal areas. The contamination over the entire surface of the hand is collected, so there is

no need to estimate using extrapolation methods. The use of gloves may have certain disadvantages, however. Cotton and nylon gloves have absorption characteristics that differ from those of the skin. They can get in the way and impede work where high dexterity is required. The wide variety of hand shapes found in the population (palm widths, finger lengths) makes it difficult to get a smooth fit in many instances because the sizes of gloves are limited. Gloves may absorb more material than would be retained by the skin, thus overestimating hand exposure.[57] Even smooth-surfaced nylon gloves may retain several times more residue than would have adhered to flesh. Gloves also contain foreign materials such as sizing that may be difficult to remove by pre-extraction and may interfere with chemical analysis of low levels of contaminants. This problem can be reduced by pre-extraction of gloves with an appropriate solvent.

Cotton or nylon socks are used to measure the amount of contaminant that enters the shoes in a method similar to using gloves for measuring hand exposure. Socks with the feet cut out have been used for measuring exposure around the forearms, lower legs, and ankles. Perspiration may be a greater concern when using socks to sample than when using gloves because boots will trap moisture. Therefore, if the boot material is suitable, rinses of the boot interiors may be more useful than monitoring using socks. Many parts of shoes have been wiped or ''sniffed'' with a direct-reading instrument, depending on the contaminant.

Paper caps have been used to measure head (scalp) exposure. Disposable paper respirators have been used to estimate facial exposures and will provide an estimate of contamination to the nose, chin, and cheeks. An advantage of this method is that the shape of these masks results in their touching the face only around the edges where the face seal is made. This eliminates the problem with sweating and picking up potentially eliminated contaminants that may have been stored in the skin, and also that of getting the sampling media damp. A disadvantage is that the mask surface area is much larger than the portion of the face that is covered, so estimates of the actual skin surface area being sampled may be difficult. Respirators can also be used to collect samples representing inhalation exposure.

Fabric tights (footless, white) made of 54% cotton, 36% polyester, and 10% spandex have been used for monitoring dermal exposure to carpet residue produced by indoor pesticide foggers. Blue jeans have been analyzed for exposure to pesticides (ethyl and methyl parathion) during field re-entry.[58] All jeans were purchased from the same manufacturer. Prior to use the jeans were laundered twice. Analysis was done by washing the jeans separately three times in hexane, soaking them for 10 min during each wash. The extracts were combined and concentrated or diluted, as required, for GC analysis.

Short-sleeved T-shirts have been used to measure the amount of a contaminant that penetrates through outer work clothing, both ordinary work garb and CPC.[59,60] All T-shirts were from the same manufacturer. Prior to use the T-shirts were laundered twice. Analysis was similar to that for blue jeans as described above. Underpants may be useful for studying exposures to compounds known to produce scrotal cancer, such as PAHs.

5.6 Whole-body Monitoring

Overalls and coveralls can be used along with gloves to approximate the total body exposure. As with other clothing, there is the option to either analyze a worker's existing clothing or to purchase and prepare clothing to use specifically for sampling. One application in which it is best to use the worker's personal clothing is when there is a question as to whether contamination is remaining in the clothing.

The current trend is to use long underwear, sometimes called a union suit, or disposable CPC body suits such as Tyvek®. When worn under regular clothing, a body suit covers the entire trunk and the extremities; alternatively, disposables can be worn over regular clothing. When used as underwear, the suits are often worn for 4 h and then collected. Removal must be done carefully so as not to disturb collected material. After removal the underwear is often placed in Zip-Loc™ bags and stored on dry ice until processing. Clothing can be cut up in order to identify exposures to specific parts of the body. In some instances after observing workers a decision is made to analyze only portions of the long underwear such as arms and legs.

A procedure for using disposable body suits has been developed by the WHO.[52] When using disposable coveralls to perform whole-body dosimetry the material should be checked first to determine its ability to collect the contaminant of interest. For example, some compounds such as Ethazol™ will readily penetrate Tyvek®.[61]

Clothing dosimetry is currently the only technique that can be used to measure contamination to the total body surface without having to use extrapolation from a smaller sampling area. A disadvantage of this method is the potential for heat stress to workers due to the additional layer of clothing. However, concerns still exist as to how accurately the residues picked up by the disposable suits or long underwear reflect the ability of the skin to collect and retain contaminants.

6 CONCLUSIONS

Current sampling strategies emphasize inhalation exposures and overlook the contribution of dermal exposures and surface contamination. Although in some cases biological monitoring is the only effective means of determining whether skin absorption is a significant concern, it is ineffective for other compounds that are rapidly metabolized in the body. Integration of surface and dermal monitoring techniques into sampling strategies will provide a better estimation of actual exposures and therefore ensure that appropriate controls are developed. The best program is often an integrated approach that uses surface and dermal sampling in conjunction with air and biological monitoring to assess not only the true dose an individual is receiving, but also the sources of exposure.

ABBREVIATIONS AND ACRONYMS

ABLES	Adult Blood Lead Epidemiology and Surveillance
ASTM	American Society for Testing and Materials
CPC	Chemical Protective Clothing
FID	Flame Ionization Detector
GC	Gas Chromatography
GF	Glass Fiber
HPLC	High-performance Liquid Chromatography
HUD	Department of Housing and Urban Development
LWW	Lioy–Weisel–Wainman
MAC	Maximum Allowable Concentration
MBOCA	4,4′-Methylenebis(2-chloroaniline)
MCE	Methyl Cellulose Ester
MDA	Methylenediamine
NACA	National Agricultural Chemicals Association
NIOSH	National Institute of Occupational Safety and Health
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PCDD	Polychlorinated Dioxin
PCDF	Polychlorinated Dibenzofuran
PID	Photoionization Detector
PUF	Polyurethane Foam
TCDD	Tetrachlorodibenzodioxin
USEPA	United States Environmental Protection Agency
USOSHA	United States Occupational Safety and Health Administration
UV	Ultraviolet
WHO	World Health Organization

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Airborne Instrumentation for Aerosol Measurements • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays

Environment: Water and Waste cont'd **(Volume 4)**
Luminescence in Environmental Analysis • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Sampling Considerations for Biomonitoring

Field-portable Instrumentation **(Volume 4)**
Aircraft-based Flux Sampling Strategies

Field-portable Instrumentation cont'd **(Volume 5)**
Radon, Indoor and Remote Measurement of

Industrial Hygiene **(Volume 5)**
Industrial Hygiene: Introduction

Industrial Hygiene cont'd **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure • Parent and Progeny Compounds in Exhaled Breath, Determination of • Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

General Articles **(Volume 15)**
Spot Test Analysis

REFERENCES

1. 'Occupational Exposure to 4,4′-Methylenedianiline (MDA). Final Rule. Preamble', Department of Labor, Washington, DC, 1992.
2. 'Dermal Exposure Assessment: Practices and Principles', Environmental Protection Agency, Washington, DC, 1992.
3. P.S.J. Lees, M. Corn, P.N. Breysse, 'Evidence for Dermal Absorption as the Major Route of Body Entry During Exposure of Transformer Maintenance and Repairmen to PCBs', *AIHA J.*, **48**(3), 257–264 (1987).
4. 'OSHA Technical Manual', 2nd edition, Department of Labor, Washington, DC, 1991.
5. L. Oatman, R. Roy, 'Surface and Indoor Air Levels of Polychlorinated Biphenyls in Public Buildings', *Bull. Environ. Contam. Toxicol.*, **37**, 461–466 (1986).
6. H. Beaulieu, 'Asbestos Contamination of Abatement Equipment Surfaces', *Appl. Ocup. Environ. Hyg.*, **6**(7), 583–587 (1991).
7. J.B. Knaak, 'Safe Levels of Foliar Residues', *Residue Rev.*, **75**, 81–96 (1980).
8. N.E. Bolton, *Industrial Hygiene, Quarterly Report, 1st Quarter, April, 1962*, Department of Energy, Oak Ridge National Laboratories, Oak Ridge, TN, 1985.
9. W.N. Saxby, J.A. Hole, 'Practical Aspects of Surface Contamination Control at A.W.R.E.', in *Surface Contamination*, eds. G.W. Royster, B.R. Fish, Pergamon Press, New York, 39–51, 1965.
10. NIOSH 'Health Hazard Evaluation Report', January, 1988. Planned Parenthood of South Central Indiana, Bloomington, IN', HETA 83-258-1859, National Institute of Occupational Safety and Health, Cincinnati, OH, 1988.
11. NIOSH 'Health Hazard Evaluation Report', May, 1986. Hilton-Davis Chemical Co., Cincinnati, OH', HETA 84-58-1700, National Institute of Occupational Safety and Health, Cincinnati, OH, 1986.
12. J. Bromley, D.C. Wilson, E.T. Smith, 'Remedial Measures Following Accidental Release of Dioxin', *Chemosphere*, **6**, 687–703 (1988).
13. 'Lead-based Paint: Interim Guidelines for Hazard Identification and Abatement in Public and Indian Housing', Office of Public and Indian Housing, Department of Housing and Urban Development, Washington, DC, 1990.
14. J. Zirschky, L. Witherell, 'Cleanup of Mercury Contamination of Thermometer Workers' Homes', *AIHA J.*, **48**(1), 81–84 (1987).
15. H. Thulin, 'Determination of Nor-nitrogen Mustard Hychloride in Air and on Surfaces by Using Filter or Wipe Sampling and a Derivatization–Gas Chromatographic Procedure', *AIHA J.*, **54**(2), 76–81 (1993).
16. J.M. Perdek, F.J. Freestone, D. Sandifer, A. Sherman, 'Closure of a Dioxin Incineration Facility', *Chemosphere*, **23**(8/10), 1295–1305 (1991).
17. NIOSH 'Health Hazard Evaluation Report', 1986. Page Belcher Federal Building, Tulsa, OK', HETA 85-289-1738, National Institute of Occupational Safety and Health, Cincinnati, OH, 1986.
18. 'Guide for Decontamination of Building Structures and Equipment at Superfund Sites', EPA 600/2-85-028, Environmental Protection Agency, Washington, DC, 1985.
19. 'Polychlorinated Biphenyls Spill Cleanup Policy: Final Rule', 40 CFR Pt. 761, FR 52:10, 688-10, 710, Environmental Protection Agency, Washington, DC, 1987.
20. NIOSH 'Health Hazard Evaluation Report', June 1991. LTV Steel, East Chicago, IN', HETA 91-27-2117, National Institute of Occupational Safety and Health, Cincinnati, OH, 1991.
21. S. Marklund, R. Anderson, M. Tysklind, C. Rappe, 'Emissions of PCDDs and PCDFs from a PVC-fire in Holmsund, Sweden', *Chemosphere*, **18**(1/6), 1031–1038 (1989).
22. A. Schecter, T. Tierman, 'Occupational Exposure to Polychlorinated Dioxins, Polychlorinated Furans, Polychlorinated Biphenyls, and Biphenylene After an Electrical Panel and Transformer Accident in an Office Building in Binghamton, NY', *Environ. Health Perspect.*, **60**, 305–315 (1985).
23. P.K. LaGoy, R.L. Bohrer, F.H. Halvorsen, 'The Development of Cleanup Criteria for an Acutely Toxic Pesticide at a Contaminated Industrial Facility', *AIHA J.*, **53**, 298–302 (1992).
24. A. Di Domenico, et al., 'Accidental Release of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) at Seveso, Italy: III. Monitoring of Residual TCDD Levels in Reclaimed Buildings', *Ecotoxicol. Environ. Saf.*, **4**, 321–326 (1980).
25. C. Chavalitnitikul, L. Levin, L. Chen, 'Study and Models of Total Lead Exposures of Battery Workers', *AIHA J.*, **45**(12), 802–808 (1984).
26. 'Proposed Standard Practice for the Field Collection of Dust Samples Using Wipe Sampling Methods for Lead Determination by Atomic Spectroscopy Techniques', American Society for Testing and Materials, Philadelphia, PA, 1993.
27. 'Wipe Sampling Procedures: Inspecting for Lead Hazards and Lead-based Paint Risk Assessment', Georgia Technical Research Institute, Athens, GA, 1993.
28. M.R. Farfel, J.J. Chisolm, 'Health and Environmental Outcomes of Traditional and Modified Practices for Abatement of Lead-based Paint', *Am. J. Publ. Health*, **80**, 1240 (1990).
29. M. Rabinowitz, A. Leviton, H. Needleman, D. Bellinger, C. Waternaux, 'Environmental Correlates of Infant Blood Lead Levels in Boston', *Environ. Res.*, **38**, 96 (1985).
30. P.J. Lioy, T. Wainman, C. Weisel, 'A Wipe Sampler for the Quantitative Measurement of Dust on Smooth

Surfaces: Laboratory Performance Studies', *J. Exp. Anal. Environ., Epidemiol.*, **3**, 315 (1993).
31. 'Lead in Surface Wipe Samples, Draft Method 0700', in *NIOSH Manual of Analytical Methods*, National Institute of Occupational Safety and Health, Cincinnati, OH, 1992.
32. A.D. Stark, R.E. Quah, J.W. Meigs, E.R. DeLouise, 'The Relationship of Environmental Lead to Blood-lead Levels in Children', *Environ. Res.*, **27**, 372 (1982).
33. R. Timperi, *Dust Collection Protocol*, State of Massachusetts, 1992.
34. J. Hayes, *Dust Sampling Protocol*, Government Printing Office, State of North Carolina, 1992.
35. J.J. Vostal, E. Taves, J.W. Sayre, E. Chamey, 'Lead Analysis of House Dust: A Method for the Detection of Another Source of Lead Exposure in Inner City Children', *Environ. Health Perspect.*, **7**, 91 (1974).
36. 'The National Human Exposure Assessment Survey (NHEXSUS) Workshop', Environmental Protection Agency, Research Triangle Park, NC, 1994.
37. G. Nichols, 'Scotch Magic Tape–An Aid to the Microscopist for Dust Examination', *Microscope*, **33**, 247–254 (1985).
38. 'Acrylamide Grouting Study', EPA 560/5-87-009, Environmental Protection Agency, Washington, DC, 1987.
39. R.A. Fenske, 'Nonuniform Dermal Deposition Patterns During Occupational Exposure to Pesticides', *Arch. Environ. Contam. Toxicol.*, **19**, 332–337 (1990).
40. G. Ritchey, R. Frank, F.L. McEwen, H.E. Braun, 'Captan Residues on Strawberries and Estimates of Exposure to Pickers', *Bull. Environ. Contam. Toxicol.*, **38**, 840–846 (1987).
41. R.A. Fenske, 'Dermal Exposure Assessment Techniques', *Ann. Occup. Hyg.*, **37**(6), 687–706 (1993).
42. K.T. Maddy, R.G. Wang, C.K. Winter, *Dermal Exposure Monitoring of Mixers, Loaders, and Applicators of Pesticides in California*, HS-1069, Worker Health and Safety Unit, State of California, 1–7, 1983.
43. R.C. Webster, H.I. Maibach, 'Regional Variation in Percutaneous Absorption', in *Percutaneous Absorption*, 2nd edition, eds. R.L. Bronaugh, H.I. Maibach, Marcel Dekker, New York, 1989.
44. W.F. Durham, H.R. Wolfe, 'Measurement of Exposure of Workers to Pesticides', *Bull. WHO*, **26**, 75 (1962).
45. J.B. Knaak, K.C. Jacobs, G.M. Wang, 'Estimating the Hazard to Humans Applying Nemacur 3EC with Rat Dermal Dose ChE Response Data', *Bull. Environ. Contam. Toxicol.*, **37**, 159–163 (1986).
46. M.L. Lepow, L. Bruckman, M. Gillette, et al., 'Investigations into Sources of Lead in the Environment of Urban Children', *Environ. Res.*, **10**, 415–426 (1975).
47. NIOSH 'Health Hazard Evaluation Report', April, 1984. King Kwit Minit Market, Inc., Cincinnati, OH', HETA 83-409-1450, National Institute of Occupational Safety and Health, Cincinnati, OH, 1984.
48. F.J. Jongeneelen, P.T.J. Scheepers, A. Groenendijk, S.J. Veenstra, 'Airborne Concentrations, Skin Contamination, and Urinary Metabolite Excretion of Polycyclic Aromatic Hydrocarbons Among Paving Workers Exposed to Coal Tar Derived Road Tars', *AIHA J.*, **49**(12), 600–607 (1988).
49. NIOSH 'Health Hazard Evaluation Report', September, 1985. Steinmetz and Sons, Moscow, PA', HETA 84-508-1626, National Institute of Occupational Safety and Health, Cincinnati, OH, 1985.
50. J.C. Maitlen, C.R. Sell, L.M. McDonough, 'Workers in the Agricultural Environment: Dermal Exposure to Carbaryl', in *Pesticide Residues and Exposure*, American Chemical Society, Washington, DC, 83–103, 1982.
51. R. Grover, H.J. Cesona, N.I. Muir, D. Riedel, C.A. Franklin, K. Yoshida, 'Factors Affecting the Exposure of Ground Rig Applicators to 2,4-Dimethylamine Salt', *Arch. Environ. Contam. Toxicol.*, **15**, 677–686 (1986).
52. 'Field Surveys of Exposure to Pesticides. Standard Protocol: VBC/82.1.', Division of Vector Biology and Control, World Health Organization, Geneva, 1982.
53. P. Curry, S. Iyengar, 'Comparison of Exposure Assessment Guidelines for Pesticides', *Rev. Environ. Contam. Toxicol.*, **129**, 79–83 (1992).
54. 'Pesticide Assessment Guidelines Subdivision U, Applicator Exposure Monitoring', EPA 540/9-87-127, Office of Pesticide Programs, Washington, DC, 1986.
55. 'Pesticide Hazard Assessment Project', *Harvester Exposure Monitoring Studies (1980–1986)*, Vol. 1, EPA 540/09-88-037, Environmental Protection Agency, Washington, DC, 1988.
56. 'Pesticide Hazard Assessment Project', *Harvester Exposure Monitoring Studies (1980–1986)*, Vol. 1, EPA 540/09-88-036, Environmental Protection Agency, Washington, DC, 1988.
57. J.E. Davis, E.R. Stevens, D.C. Staiff, 'Potential Exposure of Apple Thinners to Azinphos-methyl and Comparison of Two Methods for Assessment of Hand Exposure', *Bull. Environ. Contam. Toxicol.*, **31**, 631–638 (1983).
58. G.W. Ware, D.P. Morgan, B.J. Estessen, W.P. Cahill, 'Establishment of Re-entry Intervals for Organophosphate-treated Cotton Fields Based on Human Data. II. Azodrin, Ethyl and Methyl Parathion', *Arch. Environ. Contam. Toxicol.*, **2**(2), 117–129 (1974).
59. W.J. Hayes, *Toxicology of Pesticides*, Williams & Wilkins, Baltimore, MD, 1975.
60. G. Chester, R.J. Ward, 'Occupational Exposure and Drift Hazard During Aerial Application of Paraquat to Cotton', *Arch. Environ. Contam. Toxicol.*, **13**, 551–563 (1984).
61. J.H. Stamper, H.N. Nigg, W.D. Mahon, A.P. Nielson, M.D. Royer, 'Pesticide Exposure to Greenhouse Handgunners', *Arch. Environ. Contam. Toxicol.*, **18**, 515–529 (1989).

Nucleic Acids Structure and Mapping: Introduction

Cassandra L. Smith
Boston University, Boston, USA

The articles presented in the section, *Nucleic Acids Structure and Mapping*, illustrate the breadth of basic and applied research in this area today. The subjects range from basic chemical characteristics and three-dimensional structures of nucleic acids, to new high and low resolution analytical methods. Several articles illustrate the increasing interest in nucleic acids as biomaterials. The basics of many of these subjects have been discussed in my recently published textbook.[1] Information relative to nucleic acids may also be found in other sections (e.g. *Biomedical Spectroscopy*, *Biomolecules Analysis*, *Mass Spectrometry*, *Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy*, and **Scanning Probe Microscopy, Industrial Applications of**). Some articles in other sections relevant to nucleic acids are indicated below along with pointers to subjects covered in this section.

Nucleic acids were first isolated by Friedrich Miescher in the late 1800s from salmon sperm. Until the 1940s, DNA was believed to be a structural component within cells. In 1950, Chargoff noted that although the mole ratios of each of the bases in DNA from different organisms varied, while the G/C and T/A ratios were 1.

From 1944 to 1952, a series of experiments by Oswald Avery, Colin MacLeod and Maclyn McCarty and by Alfred Hershey and Martha Chase, using *Escherichia coli*, established DNA as the genetic material. In 1953, the X-ray diffraction experiments of Rosalind Franklin allowed James Watson and Francis Crick to described the canonical helical structure of double-stranded DNA. The ability to understand DNA function using the Watson–Crick model led to an explosion of functional studies on nucleic acids. These studies were aided by the ability to analyze and manipulate DNA first in model systems taking advantage of their genetic systems and then with DNA from virtually any organism using only a relatively few recombinant DNA methods.

Despite the long history and interest on nucleic acid research, and the existence of only five different monomers (A, T, U, C and G), today it is not possible to predict the precise three-dimensional structure of a non-canonical DNA structure (**DNA Molecules, Properties and Detection of Single; DNA Probes; DNA Structures of Biological Relevance, Studies of Unusual Sequences;** and **RNA Tertiary Structure**). This situation is in contrast to the progress that has been made in understanding and predicting the structure of proteins composed of 22 different amino acid monomers. Furthermore, there is an increasing renewed interest in understanding basic DNA chemistry promoted by research on peptide DNA (**PNA and Its Applications**) as well as other phosphodiester or base modified DNAs, and the proliferation of research on the selection and construction of nucleic acids with novel enzymatic activities or structural properties (**Structural Analysis of Ribozymes** and **Aptamers**). In addition, the search for cheaper and faster DNA sequencing and mapping methods (**Sequencing Strategies and Tactics in DNA and RNA Analysis**), prompted by the Human Genome Project, has also refocused attention on DNA fundamentals.

Hence, the first series of articles in this section are on basic chemical characteristics (**Nucleic Acid Structural Energetics**) and non-canonical nucleic acid structures and the methods that are used to analyze them. This includes X-ray diffraction methods (**X-ray Structures of Nucleic Acids**), nuclear magnetic resonance (**Nuclear Magnetic Resonance and Nucleic Acid Structures**), chemical **(Polycyclic Aromatic Compounds Mapping), Raman Spectroscopy in Analysis of Biomolecules** in the *Biomolecules Analysis* section, and enzymatic footprinting (**Structural Analysis of Ribozymes**) as well as some well established methods such as electrophoresis.

One obvious omission in this section is research using nucleic acids to build specific 3-dimensional homogenous and heterogenous molecular structures.[2,3] Nucleic acids (and peptides) are versatile tools for building molecular structures and have the advantage of providing fast prototyping since virtually any nucleic acid sequence can be rapidly and easily synthesized.

There are relatively few well-established methods that have been available to a large number of researchers for analyzing DNA. These include conventional DNA sequencing, hybridization (**Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes**), the polymerase chain reaction (**Polymerase Chain Reaction and Other Amplification Systems**), and electrophoresis. Many of the basic methods have made their way into clinical (in *Clinical Chemistry* section **Nucleic Acid Analysis in Clinical Chemistry** and forensic (in Forensic section **Polymerase Chain Reaction in the Forensic Analysis of DNA**) applications. Mass spectrometry (**Mass Spectrometry of Nucleic Acids** and **Mass Spectrometry in Structural Biology** in the *Biomolecules Analysis* section) promises to make a major impact in the areas traditionally served by these technologies.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Refinements in hybridization experiments have led to the production of DNA chips (in *Clinical Chemistry* section **DNA Arrays: Preparation and Application**) and sequencing by hybridization strategies (**Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes**) for DNA analysis. DNA hybridization can also form the basis of biosensors (**Fluorescence-based Biosensors** in *Biomolecules Analysis* section).

DNA chip or SBH experiments can be viewed either as DNA sequence fractionations (i.e. sorting by hybridization) or as DNA fingerprinting experiments. Some DNA fingerprinting methods including indexing sequencing methods that developed from the idea of accumulated sequence around restriction enzyme cleaved sites (sequence tagged restriction sites (STARs).[4] A DNA fingerprinting method focused on restriction enzyme cleavage sites (**Restriction Landmark Genomic and cDNA Scanning**) is presented here.

Recent developments in fast, high resolution DNA fractionation methods (e.g. **Capillary Electrophoresis of Nucleic Acids**) are also presented. A number of low resolution methods of analysis are presented, including the analysis of human chromosomes by fluorescent in situ hybridization (**Fluorescence In Situ Hybridization**) and electron tomography (**Electron Tomography of Chromosome Structure** and **Single Biomolecule Detection and Characterization**). One new low resolution method that directly observes and measures DNA length microscropically is being used to create genomic restriction maps (**Optical Mapping in Genomic Analysis**). An important omission in this section is direct DNA visualization using atomic force microscopy (ATM). Articles in other Encyclopedia sections include **Atomic Force Microscopy in Analysis of Polymers**, and the entire section entitled *Surfaces*. Some recent reviews on DNA have been published.[5,6]

Other articles in this section discuss radiation hybrid mapping (**Radiation Hybrid Mapping**) of and the construction and mapping of libraries composed of bacterial artificial chromosomes (**Genome Physical Mapping Using BACs**) or other cloned sequences. A cheap and easy alternative to these low resolution mapping approaches is one we have developed called cloneless libraries.[7] In the cloneless library approach, genomic restriction fragments purified from pulsed field gel electrophoresis fractionations are used in place of cloned sequence. This approach is cheap, and easy, and has the advantage that a cloneless library can easily be generated from any DNA source.

This section of the Encyclopedia closes with a review of comparative genomic methods focused on differential display and subtractive hybridization (**Comparative Genomics: Differential Display and Subtractive Hybridization**). These comparative methods are accessible to the largest number of people. This article serves to remind the reader that it is comparative genomics that will synthesize new biological information.

It should be apparent that the articles presented in this section do not exhaustively cover current research on nucleic acids. Rather, they are representative of a wide spectrum of possibilities. It is hoped that a reader will be able to initiate studies in specific areas with the articles presented here but also recognize that this field is rapidly changing and so should also be prepared to investigate other areas not reviewed in this section.

REFERENCES

1. C.R. Cantor, C.L. Smith, *Genomics*, *The Science and Technology Behind the Human Genome Project*, Wiley & Sons, New York, 1999.
2. N.C. Seeman, 'DNA Technology: Novel DNA Constructions', *Ann. Rev. Biophys. Biomol. Structure*, **27**, 225–248 (1998).
3. N.C. Seeman, 'Molecular Craftwork with DNA', *The Chemical Intelligencer*, 38–47 (1995).
4. R. Oliva, S.K. Lawrence, Y. Wue, C.L. Smith, 'Chromosomes: Molecular Studies', in *Encyclopedia of Human Biology*, ed. R. Dulbecco et al., Academic Press, San Diego, 475–488, 1991
5. H.G. Hansman, L. Pietrasants, 'Atomic Force Microscopy and other Scanning Probe Microscopies', *Curr. Opin. Chem. Biol.*, **2**(5), 579 (1998).
6. C. Bustamaente, G. Zuccheri, S.H. Leuba, G. Yang, B. Samori, 'Visualization and Analysis of Chromatin by Scanning Force Microscopy', *Methods*, **12**(1), 73–83 (1997) .
7. C.L. Smith, D. Mass, M. Lewis, Cloneless Libraries: Efficient Long Range Mapping of Chromosome 20, in preparation, 1999.

Aptamers

William James
University of Oxford, Oxford, UK

1 Introduction 1
2 In Vitro Evolution of Nucleic Acid Ligands 1
2.1 Outline of Process 1
2.2 History of Discovery and Development 3
2.3 Choice of Chemistry 3
2.4 Practical and Theoretical Considerations 5
3 Properties of Published Aptamers 8
3.1 Structures 8
3.2 Aptamer Size 9
3.3 Aptamer Targets 9
3.4 Affinity 10
3.5 Specificity 11
4 Applications and Downstream Technologies 11
4.1 Conjugation to Detectable Moieties 11
4.2 Aptamers in Biosensors and Other Detectors 12
4.3 Aptamer Combination 12
4.4 Use of Aptamers In Vivo 13
4.5 Studies on Nucleic Acid Structure 13
4.6 Isolation of New Catalysts 15
5 Perspectives and Future Developments 15
Acknowledgments 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Aptamers are artificial nucleic acid ligands that can be generated against amino acids, drugs, proteins and other molecules. They are isolated from complex libraries of synthetic nucleic acid by an iterative process of adsorption, recovery and reamplification. They have potential applications in analytical devices, including biosensors, and as therapeutic agents.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

A very neat job in a small space[1]

Aptamers range in size from approximately 6 to 40 kDa and sometimes have complex three-dimensional structures, produced by a combination of Watson–Crick and non-canonical intramolecular interactions. They bind to their targets with K_D typically in the low nanomolar range and can distinguish enantiomers of small molecules or minor sequence variants of macromolecules with frequently several orders of magnitude K_D ratio. They are typically composed of RNA, single-stranded DNA or a combination of these with non-natural nucleotides.

Aptamers are isolated from extremely complex libraries of nucleic acids, generated by combinatorial chemistry, by an iterative process of adsorption, recovery and reamplification. Additional sequence variation can be introduced at each cycle and the process becomes an in vitro paradigm of Darwinian evolution. After sufficient enrichment, aptamers can be cloned and studied as homogeneous sequence populations.

Aptamers can be used to analyze the natural processes of nucleic acid–protein recognition, to generate inhibitors of enzymes, hormones and toxins with potentially pharmacological uses, to detect the presence of target molecules in complex mixtures and to generate lead compounds for medicinal chemistry. Their advantages over alternative approaches include the relatively simple techniques and apparatus required for their isolation, the number of alternative molecules that can be screened (routinely of the order of 10^{15}) and their chemical simplicity. Disadvantages of aptamers include their pleiomorphism, their high molecular mass and the restricted range of target sites that appear to be suitable.

2 IN VITRO EVOLUTION OF NUCLEIC ACID LIGANDS

2.1 Outline of Process

Aptamers are ligands derived by a process of combinatorial chemistry, in which the desired property is identified by affinity chromatography and encoded genetically. The process is illustrated in Figure 1 and described below.

2.1.1 Library Synthesis and Complexity

Combinatorial chemistry is the production of a very large number of different molecules by the repeated

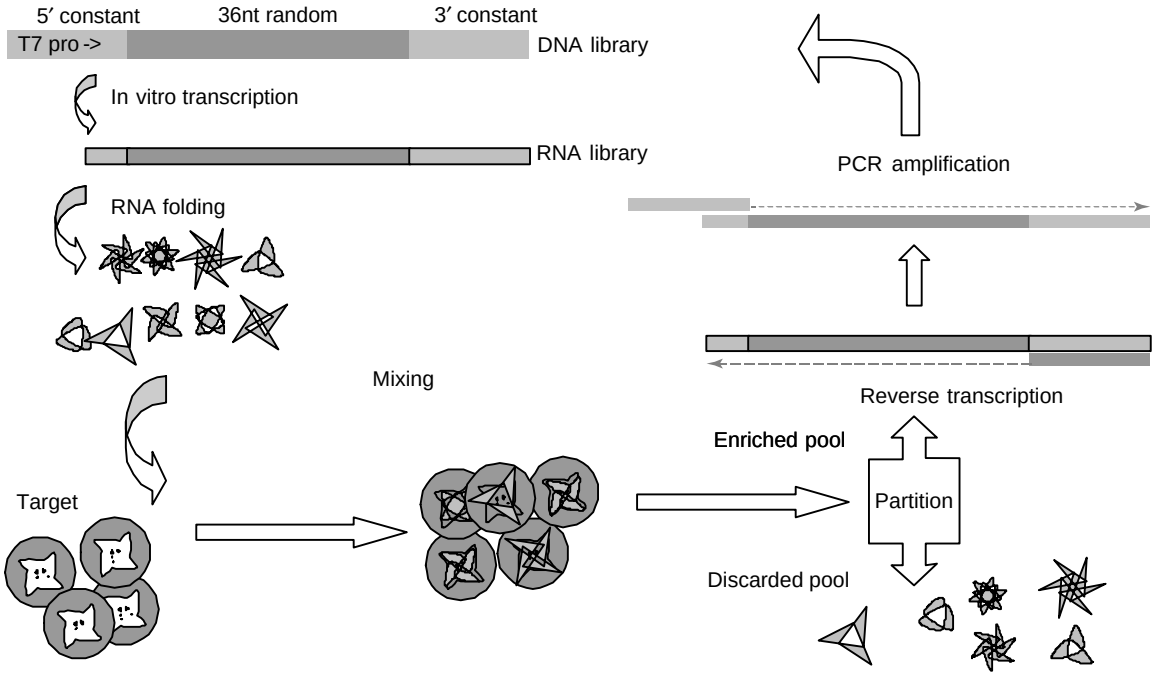

Figure 1 Schematic representation of the in vitro selection of nucleic acid ligands (aptamers).

use of a limited range of synthetic steps. It takes many forms, outside the scope of this article, but was inspired by the sequential, solid-phase synthesis of oligopeptides and oligonucleotides (oligos: **Combinatorial Chemistry Libraries, Analysis of**). An oligo of defined sequence is commonly synthesized on a solid support by a cycle of deprotection of the acceptor end of the growing chain and its derivitization by using an excess of the next activated monomer in the mobile phase. By modifying this process so that the activated monomer is a mixture, rather than a single species, oligos of randomized sequence may be readily obtained. For sequences of randomized length N and y alternative monomers at each position, a library of oligos of diversity y^N can, in principle, be synthesized. For example, a nucleic acid library of randomized length 40 has a maximum theoretical diversity of $4^{40} = 1.2 \times 10^{24}$. The theoretical diversity, otherwise known as *sequence space*, of such a library will often exceed by many orders of magnitude the number of molecules that can be handled in practice. Convenience and expense generally restrict the initial library to a sample of in the region of 10^{13}–10^{15} molecules taken, presumably randomly, from the larger sequence space.

2.1.2 Selection of the Desired Property

The desired property is, typically, the ability to bind to a molecule of interest. Depending on the anticipated application, the desired binding properties may be a fast association rate, slow dissociation rate, high affinity, low affinity to closely related molecules, or a combination of these. This property will be a function of the three-dimensional structure of the folded nucleic acid and will be a combination of its van der Waals surface contacts, hydrogen bonds, stacking interactions and other non-covalent bonds that can form between the aptamer and its target. It is a necessary assumption that, to a reasonable approximation, the three-dimensional structure of an aptamer is uniquely determined by the sequence of its bases. These assumptions will be explored below.

By mixing the solution-phase library with the target molecule and subsequently retrieving the target and removing unbound or loosely bound nucleic acids, the few nucleic acids that have the desired property can be recovered from the library. This partitioning step is a form of affinity chromatography and the methodology can be varied depending on the nature of the target and the exact property desired. This will be discussed in detail below. The selection step differs from screening methods used in other branches of combinatorial chemistry because of the numbers of candidate molecules that can be reviewed at one time. Conventional libraries of potential drugs may consist of the order of 10^6 molecules, which can be screened for activity by robotic means in a few months. Combinatorial libraries of greater than about 10^5 are impractical to screen by conventional, iterative methods, as vanishingly few molecules of each species would be present in each assay. Perhaps the most sensitive methods involve affinity partitioning followed by mass spectrometry (MS).[2]

2.1.3 Genetic Encoding

In contrast to other forms of combinatorial chemistry, where a variety of methods have been devised to encode the structure of each agent, or at least its synthetic recipe, in the solid-phase support, aptamers contain within themselves the genetic code for their own amplification and synthesis. Nucleic acid polymerase enzymes are used to convert the target-bound aptamer, if it was based on RNA, into DNA, to amplify the copy number exponentially (by a factor of at least 10^{10}) and, again if the aptamer is RNA, to transcribe the amplified DNA template back into RNA.

2.1.4 Enrichment and Evolution

Because the affinity-based partitioning methods are imperfect, the cycle of partitioning and amplification is normally repeated 6–12 times, sometimes more, before molecules of the desired property predominate in the population of nucleic acids. The process, as described above, represents a reduction in sequence diversity from the initial sample of sequence space (e.g. 10^{15} out of a space of 10^{24}) to a population of the order of typically 10 distinct sequences that show appreciable binding to the target. In addition, however, the error-prone nature of the enzymes used to amplify the selected sequences leads to the introduction of mutations that effectively allow the procedure to sample a greater proportion of the sequence space than was initially sampled. This leads to the appearance of clearly related but evolutionary divergent sequences within the final aptamer population, in a process akin to Darwinian evolution by natural selection.

2.2 History of Discovery and Development

While experiments involving the isolation of nucleic acids from artificial libraries on the basis of their biochemical properties were being widely discussed during 1988 and 1989, three groups independently published their results in 1990. First, the Joyce group reported the use of in vitro mutation, selection and amplification to isolate RNAs that were able to cleave DNA.[3] They began with a natural RNA-cleaving ribozyme, rather than a purely random library, and were looking for a novel enzymic activity rather than a selective ligand. However, their experiment had most of the essential features described in this article, including the repeated cycles of reaction performed in a single vessel. Second, the Gold group described experiments designed to identify the sequence requirements of T4 DNA polymerase, in which the library was based on the natural target hairpin structure but with the eight loop nucleotides randomized.[4] The process of in vitro selection, for which they coined the term 'SELEX' (selective expansion of ligands by exponential enrichment), was able to identify the natural target of the enzyme as the predominant, high-affinity ligand, with one major variant emerging with similar affinity. The authors patented their process. Less than a month later, the Szostak group reported the use of in vitro selection to isolate 'molecules with specific ligand-binding activities'.[5] More radically than the previous reports, they began with a library that was structurally unrelated to any known nucleic acid, having 100 nucleotides of randomized sequence, and chose targets that had no previously identified nucleic acid ligands. They used affinity chromatography to isolate RNAs with specific and selective binding characteristics for a number of organic dyes (chosen because of their potential as H-bond partners and their planar structures that might be expected to form stacking interactions with the nitrogenous bases of RNA). It was this group that coined the term 'aptamer' for such nucleic acid ligands.

2.3 Choice of Chemistry

Aptamers are nucleic acids that can be composed of naturally occurring monomers or chemically synthesized derivatives. For a review of this field, see Eaton and Pieken.[6] As outlined above, the nucleotides must be compatible with RNA-dependent DNA polymerases (reverse transcriptase) and with DNA-dependent RNA polymerases (such as T7 RNA polymerase) for the enzymatic steps required in the production of RNA-based aptamers, or with thermostable DNA-dependent DNA polymerases (such as Taq polymerase) for the production of DNA-based aptamers.[7,8] This restricts the available chemistry more strictly than is the case with purely synthetic oligonucleotides. Nevertheless, researchers have adopted many modifications of natural nucleotides in order to overcome two substantial problems posed by natural nucleic acids.

2.3.1 Phosphodiester Bond Hydrolysis

This is a particular problem in natural RNA, as the hydroxyl at the 2′-position is reactive, particularly at higher than neutral pH, and will attack the neighboring phosphodiester bond to produce a cyclic 2′,3′-phosphate, thereby breaking the nucleic acid backbone (see Figure 2). This reaction is catalyzed by many transition metals ions, particularly lead and iron, and by a range of ribonucleases found ubiquitously in biological samples. To overcome this problem, one has turned to modifications of either the 2′-position of the ribose moiety or modifications of the phosphodiester backbone. By far the most common approach is the use of nucleotides substituted at the 2′-position with either an amino group or fluorine atom. These modifications were pioneered

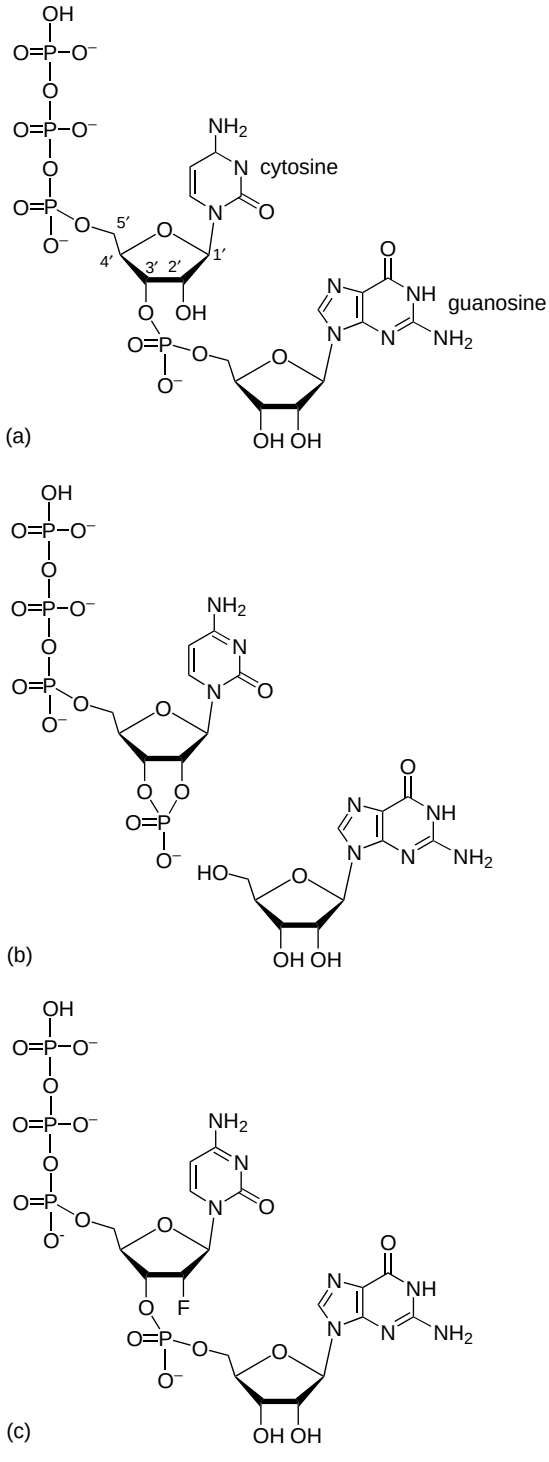

Figure 2 Representative RNA components. (a) Diagram of the structure of a CG dinucleotide, showing the numbering convention for carbons in the ribose moiety. (b) The 2′,3′-cyclic phosphate that forms as the result of nucleolysis. (c) A CG dinucleotide containing 2′-fluorine.

by the Eckstein group[9,10] and are compatible with T7 RNA polymerase for efficient in vitro transcription.[11] Where they have been compared, 2′-deoxy and 2′-amino chemistries have been found to produce aptamers with similar affinity, although different structure (for an example, against immunoglobulin E (IgE), see Wiegand et al.[12]). However, 2′-fluoro chemistry produces aptamers with greater thermal stability and probably higher affinity than 2′-deoxy or 2′-amino chemistry.[13] Although one cannot generally change the chemistry of the nucleotides after selection without changing the structure of the aptamer and frequently abolishing its properties as a ligand, this can often be done in a more selective manner (A. Tahiri-Alaoui, unpublished work). For example, further modifications at the 2′-position that give additional stability can be incorporated synthetically into aptamers with 2′-OH, -H, -NH_2 or -F that have been identified through in vitro evolution. For example, 2′-*O*-methyl has been introduced at certain purines in 2′-F-pyrimidine aptamers against the vascular endothelial growth factor (VEGF), with beneficial effects.[14] The alternative approach, to modify the phosphodiester backbone, is more challenging to the enzymology of in vitro evolution. Nevertheless, by using α-thio-substituted deoxynucleoside triphosphates (dNTPs), a phosphorothioate DNA library was successfully screened for aptamers against the transcription factor NF-IL6,[15] and some progress has been made toward the use of analogous, 'thio-RNA' aptamers.[16,17]

Another way to produce nuclease-resistant aptamers is to select an aptamer that binds the enantiomer of the eventual target, then synthesize the enantiomer of the aptamer as a nuclease-insensitive ligand of the normal target. Such 'spiegelmers' have been made in L-DNA against the peptide hormone, vasopressin[18] and in L-RNA against L-adenosine and L-arginine.[19,20]

2.3.2 Absence of Hydrophobic or Basic Residues

In contrast to proteins, nucleic acids are strikingly uniform in their hydrophilicity and low p*I*. In spite of these limitations, a surprising range of enzymatic activities are possible for nucleic acids and the hydrogen bond and stacking interactions of their component bases provide a diverse toolbox of structural motifs.[6]

Nevertheless, this uniformity almost certainly limits the ease with which aptamers can be made to certain targets and a greater range of side chains, including some with basic or hydrophobic character would be welcome additions to the armamentarium. The challenge, as with substitutions giving nuclease resistance, is to discover approaches that are consistent with the enzymes used during aptamer isolation. One approach, using 5-(1-pentynyl)-2′-deoxyuridine, succeeded in producing additional aptamers to thrombin,[21] though no reports

have emerged showing that this approach has produced aptamers against more refractory targets.

It has long been speculated that the precursor to our DNA-genome, protein-catalyst world of organisms was a world in which RNA functioned for genomes, ligands and enzymes[22] and this was consolidated by the discovery of ribozymes.[23,24] (In his paper on the 'Evolution of the Genetic Apparatus', Orgel[22] speculated on 'life based on nucleic acids without a genetic code' and said that 'it seems to me quite possible that polynucleotide chains could make primitive selection among organic molecules such as amino acids by forming stereospecific complexes stabilized by hydrogen-bonding and hydrophobic interactions'. He also speculated that polynucleotides might have the ability in early stages of life to catalyze chemical reactions but that this 'function would subsequently have been taken over by the much more versatile polypeptides'.) Indeed, it is proposed by some that the restricted range of amino acids extant in the current world of biology is dictated by those homologous with natural, prebiotic derivatives of 5-hydroxymethyluracil that had previously been a feature of the 'RNA world'.[25] Further, one should note that many biological RNAs of the present world have additional methyl groups added post-transcriptionally to their purines, which increase the local hydrophobicity of the nucleic acid (reviewed by Levy and Miller[26]). None of these potentially helpful modifications is reproduced during enzymatic replication of nucleic acids, though some of them, rather than acting as chain terminators, are ignored. For example, 5-methylation of cytosine is a common postreplication modification of eukaryotic DNA and 5-methyldeoxycytidine is replicated routinely to dG during the S phase of the eukaryotic cell cycle.

Instead of attempting to broaden the repertoire of nucleotides used during the replication and transcription phases of in vitro selection of aptamers, one might modify one or more nucleotides post-transcriptionally. It was found that derivitization of 2′-NH_2 groups on aminopyrimidine-RNA with succinimide did not prevent reverse transcription.[27] Although this might usefully open up a wide range of functional adducts, the efficiency was low, with only two out of 23 amino groups per molecule, on average, being successfully derivatized. In order to preserve the genetic encoding of aptamers, it is essential that such processes are either fully efficient or reproducibly inefficient (i.e. the same nucleotides being derivatized each cycle).

2.4 Practical and Theoretical Considerations

A comprehensive and practical introduction to the methodology of in vitro evolution of RNA ligands was provided by Fitzwater and Polisky[28] and so will not be repeated here. A mathematical description of the process, under assumptions of equilibration, was produced by Irvine et al.[29] and supplemented by a model for simultaneous selection against multiple targets by Vant-Hull et al.[30] We shall confine ourselves to a discussion of some of the most relevant details.

2.4.1 Library Complexity

When designing a library, the desire is to produce sequences that can be amplified and transcribed with high efficiency, of as great a diversity of structures as is practicable and at reasonable cost. One is bound to use substantial fixed, flanking regions for primer-based amplification and transcription, so one question is how long to make the randomized region (N) and how to ensure that the library produced contains as near to the theoretical maximum of sequences possible. First, although the sequence space available to the library is 4^N, the maximum number of molecules that can be manipulated in standard molecular biology laboratories (M) is of the order of 10^{15}, and errors of synthesis and workup have been estimated to reduce this diversity to 10^{13}–10^{14}. N need only be around 22–24 to reach this practical limit. However, in order to produce a structure capable of specific interaction with a target molecule, the RNA needs to be large enough to fold into a complex tertiary structure. It appears that many classes of RNA fold need to composed of significantly longer stretches of nucleic acid than this (see above), so most people use libraries where $N \geq 35$ and sometimes much more. As N increases, one is able to sample progressively a smaller fraction of the theoretical sequence space. For example, if $M = 10^{14}$ and $N = 24$, we could sample $(10^{14}/2.81) \times 10^{14} = 36\%$ of the sequence space, but if $N = 25$, this reduces to 8.9%. However, if we consider each 25-mer to be composed of two overlapping 24-mers which differ only at their termini, then the $N = 25$ library effectively contains 2×10^{14} 24-mers. This suggests that one should aim for larger rather than smaller values of N. Two considerations effectively limit the practical size of N. First, although the efficiency of oligonucleotide synthesis is constantly improving, errors accumulate with each cycle and a point is reached at which incorrectly synthesized molecules begin to have a significant impact on the quality of the library. The limit is imprecise but it seems unwise at present to make large-scale libraries using oligo lengths much greater than about 120 nt, suggesting a practical limit of around $N = 70$–80. Second, if RNA folds are typically between 25 and 50 nt long, libraries with N greater than approximately 50 nt may well contain sequences comprising two independent folds, complicating the process of in vitro selection.

The actual diversity of a library can be compromised by variations in the efficiency of coupling phosphoramidite

monomers during oligonucleotide synthesis, which in turn can depend on the particular practices of the manufacturer chosen. Although standard methods use equimolar mixtures at each N stage during synthesis, the Szostak group has reported the use of an A : C : G : T ratio of 3 : 3 : 2 : 2 and we have found it necessary to use ratios of 6 : 5 : 5 : 4 in order to achieve random incorporation. Gross bias during synthesis and enzymic manipulation can be estimated by checking the sequence of a dozen or so clones made from the library after a single round of amplification, transcription, reverse transcription and reamplification, without selective partitioning. The clones should be sequenced and the overall base composition within the random region tested for divergence from expectation. In addition, one should look for bias in the dinucleotide and trinucleotide composition and at base composition at each position.

What are the chances of finding a target-binding sequence during the in vitro exploration of sequence space? The answer seems to depend on one's target and the criteria (e.g. affinity threshold) used to identify binding. For example, within the $N = 100$ library described by Ellington and Szostak,[5] the authors estimated that 10% of the library contained aptamers for organic dye molecules. In contrast, Burke and Gold[31] estimated that perhaps one in 10^{11} sequences within their library contained adenosine-binding motifs. Our own experience with libraries of $N = 36$ is that protein-binding aptamers sequences exist at an initial frequency of typically 1 in 10^{11}–10^{12}. If these figures are representative, it shows that care needs to be taken during the early stages of library production and manipulation not to reduce the sequence complexity much below its theoretical maximum.

Tuerk[32] has made calculations based on computer modeling of the SELEX process and suggested that binding reactions that yield 6% of the total RNA bound to target would be optimal. Schneider et al.[33] showed that when reactions were arranged to select 10% of the RNA population bound to ρ protein (low stringency) or 1% (high stringency) during SELEX, ligands of similar sequences were produced by both selection strategies, although the speed of affinity enhancement of the pool was increased somewhat for the high stringency selection. This illustrates that the fine tuning of ligand–target ratios is not crucial for success in obtaining optimal binding ligands. Tuerk[32] suggests that the first binding reaction be conducted in large volumes (50 mL) with target molecule concentrations at 0.2 times the K_D of the original nucleic acid pool–target interaction and all of the starting nucleic acid. This would yield about 9% of the original nucleic acid population bound to a large number of target molecules, decreasing the chance that low-abundant, unique sequences of high affinity are lost in the first selection.

2.4.2 Mutagenesis and Affinity Maturation

Thus far, we have been considering merely the selection of ligands from an initially very small and approximately random sample of sequence space. However, the phrase 'in vitro evolution' implies that we are able to explore a much greater fraction of sequence space by allowing mutation to generate additional diversity and selection to retain a small fraction with favorable properties. By repeatedly selecting for the desired property, the path of exploration should tend towards optimum ligands; those that occupy segments of sequence space that correspond to affinity maxima. However, we can imagine many ways in which this favorable result would not have been reached. First, the initial sample of the library may not have contained any members that were within striking distance of a high-affinity region of sequence space. This will depend on the complexity of the library and the nature of the target, as described above. Second, the amount of mutation at each level may have been too little to allow significant progress towards the affinity maximum. Conversely, mutation might have been so high that the peaks would have been missed or overshot. Third, selection might not have been stringent enough, allowing the minority of high-affinity ligands to be swamped by their low-affinity brethren. Conversely, selection might have been too stringent, eliminating sequences of intermediate affinity before high-affinity sequences had been arrived at. This is a particularly acute problem during in vitro selection because one rarely has any way of knowing either the affinity topology of sequence space or the properties of early-cycle oligonucleotides.

There has been surprisingly little experimental analysis of this problem. Although we know that aptamers selected against a particular target usually fall into sequence groups which suggest phylogenetic relatedness, the polyphyletic origin of aptamers means that two similar sequences may have arisen by convergence from two independent sequences sampled in the initial library as easily as by divergence from a common ancestor. Nevertheless, in one study, explicit evolution was measured from an aptamer that had been previously selected against L-citrulline to others which could bind alternative amino acids.[34] Mutation was introduced at a single step by chemical resynthesis of the original aptamer to give an average of 30% mutation at each position. Binding to citrulline was recovered after three rounds of selection and to arginine after four rounds. However, no lysine- or glutamine-binding aptamers were obtained. The more usual method of introducing mutation is to encourage misincorporation of nucleotides during amplification by increasing the concentration of dNTPs and Mg^{2+} and using enzymes lacking proofreading activity.[28]

2.4.3 Partitioning Methodology

The separation of desired from undesired sequences is of critical importance in the selection and evolution of aptamers (see the discussion above and a mathematical analysis by Vant-Hull et al.[30]). Ideally, one would be able to allow the nucleic acid pool and the target to interact freely, be able to monitor the progress of the interaction and retain those nucleic acids whose desired property lay above a threshold determined in part by the properties of the pool. Most methods of partitioning fall short of these ideals. First, tethering the target to a solid support facilitates the separation of bound from unbound nucleic acids and facilitates procedures such as competitive elution which can be helpful in setting affinity and specificity parameters. This is the approach used in some of the earliest experiments, and has been particularly used in the identification of aptamers to small molecules.[5,34–36] Immobilization of target proteins to Sepharose has also been used successfully to facilitate partitioning[37,38] but mixing the nucleic acid and target protein in the soluble phase has significant advantages. First, many protein immobilization regimes involve derivitization of the ε-amino groups of lysine, thereby destroying a key feature of many aptatopes. Second, immobilization reduces the mobility of the protein, impeding its mixing with nucleic acid and may produce direct steric hindrance to aptamer binding. Third, a progressive reduction in the molar ratio of target protein to nucleic acid in successive cycles, in order to increase the stringency of selection, is very hard to achieve in an immobilized system. Consequently, one of the common approaches is to use an insoluble matrix that selectively adsorbs protein following nucleic acid–protein interaction in the fluid phase. Nitrocellulose filters are a cheap and convenient matrix that is widely used for this purpose.[4,39,40] However, we have found that nitrocellulose preparations are not as selective for protein as would be desired. We have compared the use of nitrocellulose filters, poly(vinylidene difluoride) (PVDF) membranes, activated Sepharose, octyl Sepharose and the deproteinizing matrix Strataclean™ resin for their ability to select for CD4-binding aptamers via interaction with the protein (F. Kesten, personal communication). We found that nitrocellulose and PVDF trapped substantial amounts of 2′-F-RNA nonspecifically and that octyl Sepharose was not an efficient protein binder under these circumstances (results not shown). However, CNBr-activated Sepharose and Strataclean™ resin were both efficient at pulling down 2′-F aptamer RNA in a target-dependent fashion (see Figure 3).

Models of the selection process also assume that equilibrium between RNA and protein has been reached before partitioning.[29] However, this is very unlikely to be the case during early rounds, when the concentration

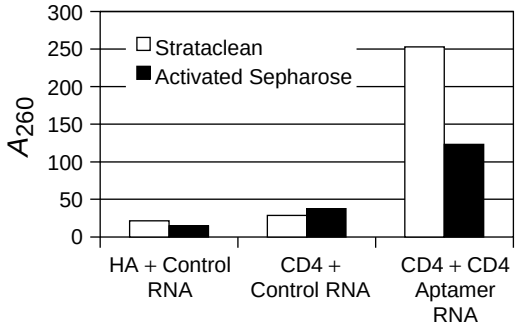

Figure 3 Selective partitioning of ligand RNA using either activated Sepharose or Strataclean™ resin.

of ligands is extremely small or at any stage when incubation times of the order of 10 min are used routinely and the most desirable aptamers have dissociation $t_{1/2}$ of the order of 1 h (J. Ibrahim and L. Frigotto, personal communication). In a recent study, we were able to show that increasing the incubation time from 30 min to 2.5 h substantially reduced the number of cycles required to isolate high-affinity aptamers.[41] Recent improvements in robotics technology have enabled the process of in vitro selection to be reduced to a matter of days.[42]

2.4.4 Structural Pleiomorphism

A central but unspoken assumption of the methods for in vitro selection and evolution of aptamers is that there is a direct and unique relationship between the sequence of an aptamer and its shape: the 'one sequence, one structure' assumption. In other words, if one selects a nucleic acid sequence s in round n on the basis that it binds to protein X, then one would expect that all copies of the amplified sequence s will bind to X in round $n+1$. However, if sequence s can fold into more than one conformation, only a proportion of the reamplified s will bind to X in round $n+1$. Worse, if each sequence is present at low copy number, for example during the early rounds of selection, the copy of sequence s in round n may not be in the right conformation to bind X and thus not be selected at all. How pleiomorphic should we expect most RNAs to be? If we were to answer this on the basis of the known folding of functional RNAs such as the *Tetrahymena* intron or the hammerhead ribozyme, which fold rapidly into their native states, we might be misled. Instead, if we look at most monoclonal aptamers, it is not uncommon to discover that they comprise distinct conformers, only a minority of which are competent ligands (A. Tahiri-Alaoui and E. Kraus, personal communication). During selection, one would expect the selective process to favor those RNAs that are minimally pleiomorphic, so the persistence of this property in aptamers suggests that

pleiomorphism is common in unselected oligonucleotides. RNA folding algorithms, admittedly still only able to predict approximately 75% of experimentally determined base pairings,[43,44] indicate that a given sequence is frequently capable of folding into several structures, each with similar ΔG. If the free energy barriers between any of these folded forms are great enough, multiple conformers would be expected. Natural evolution of, for example, ribozymes, would have imposed a selective pressure for sequences whose folding pathways are more uniform or for an association with chaperone-like proteins that would favor one conformation over another. Those working with aptamers have adopted a variety of strategies to minimize this problem – from snap-cooling nucleic acid pools from the denatured state to a slow refolding procedure – but none is entirely successful. Although technically challenging, it is becoming possible to analyse co-existing structures within a population of RNA molecules with identical sequence.[45]

Conformational flexibility has been observed in protein–aptamer interactions, for example, an induced fit in the target polypeptide following the binding of aptamers raised against the HIV-1 regulatory protein.[46] Conversely, the conformation of the nucleic acid was seen to change upon binding of an RNA aptamer to its amino acid target.[47]

3 PROPERTIES OF PUBLISHED APTAMERS

3.1 Structures

The structure of a large number of aptamers has been determined by enzymatic or chemical probing, nuclear magnetic resonance (NMR) and X-ray crystallography (see reviews by Feigon et al.[48] and Patel[49]). They are relatively amenable to NMR methodology because of their small size and their rigidity when complexed with target. Although the fundamental repertoire of nucleic acid secondary structures is radically different from that of proteins, the similarities between the interaction sites between protein ligands and their receptors can be strikingly similar to those of aptamers.[50] A few of the salient features are described below.

3.1.1 DNA Aptamers

One of the earliest aptamers studied structurally was the 15mer DNA aptamer against thrombin, d(GGTTGGTGTGGTTGG), isolated by Bock.[7] Using two-dimensional proton NMR, it was shown[51] that this short oligo folded tightly into a four-stranded structure, stabilized by two stacked G tetramers, with each of the two pairs of strands having a TT dinucleotide loop (see Figure 4). X-ray crystallography of the aptamer complexed with its target[52] showed that the aptamer was sandwiched between the two highly basic exosites of thrombin, making no contact with the active site at the base of the cleft between the two exosites. The X-ray and NMR models were very similar, but the polarity of the strands was different, with the two TT loops spanning the major groove in the X-ray model and the minor groove in the NMR model. These two structures are technically very hard to discriminate from the crystallographic data, and it is probable that the NMR model is, in fact, correct.[53] Another GT-rich oligo, d(G^{s}TGGTGGGTGGGTGGG^{s}T) (where the first and last phosphodiester bonds are replaced with phosphorothioate), was found to be an inhibitor of the integrase enzyme of HIV-1.[54] The structure of this aptamer is again four strands, stabilized by a pair of G quartets but in which the joining loops are all TG.[55] The G octet is not the only kind of structure adopted by DNA aptamers, however. For example, an aptamer against argininamide is a hairpin loop, which undergoes a rearrangement upon binding to its target, in which the loop is reflexed to trap the amino acid between the stem and the stabilized loop by a combination of H-bonds and stacking interactions.[56]

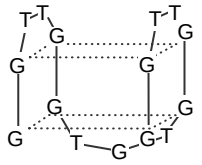

Figure 4 Representation of a thrombin-binding aptamer. The phosphodiester backbone is shown with solid gray lines. The noncanonical hydrogen bonds forming the G tetrads are shown as dotted lines.

3.1.2 RNA Aptamers

Aptamers based on RNA, or nucleotides with predominantly RNA-like properties, are capable of adopting a seemingly greater variety of structures (reviewed by Patel et al.[57] and Ferre-D'Amare and Doudna[58]). For example, a flavin mononucleotide (FMN)-binding aptamer consists of an asymmetric loop that binds the target, flanked by two A-form helices.[59] The asymmetric loop itself forms a widened, colinear helix comprising an *anti* purine pair and a base triple at one side, two *anti* purine pairs at the other of the intercalated isoalloxasine moeity of FMN (see Figure 5). Essentially, the target is acting as a highly abnormal intercalated base in a modified double helix. Another nucleotide-binding aptamer, which recognizes adenosine 5′-monophosphate (AMP),[60] incorporates its target even more intimately into its structure. Tetraloops of consensus sequence GNRA (where N is any nucleotide and R is a purine) are

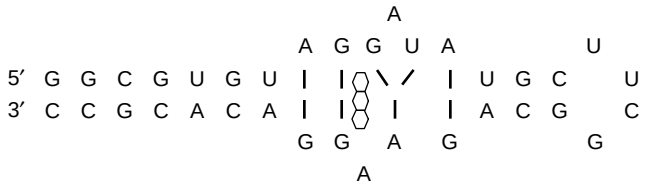

Figure 5 Secondary structure representation of an FMN-binding aptamer. Noncanonical base pairs are indicated by single solid lines and the base triplet by a set of three radiating lines. The FMN's isoalloxasine moiety is indicated by a triple hexagon.

common in natural RNAs and are particularly stable.[61,62] The AMP-binding aptamer was found to recruit the free AMP and thereby form a GNRA-like loop from a region that is unpaired in the absence of a target.[63–65]

The structures of protein-binding aptamers show further variations on these themes. For example, aptamers that bind the HIV-1 regulatory protein, Rev, accommodate an α-helix of their target within a modified major groove that is widened by the presence of a number of non-Watson–Crick base pairs.[66–68] In contrast, a class of aptamers that bind the reverse transcriptase of HIV-1 possess a compact form of pseudoknot structure[69,70] (see Figure 6) which is believed to occupy the groove within the enzyme normally occupied by its kinked template. A biotin-binding aptamer was also found to have a rigid pseudoknot structure.[71] The pseudoknot represents just one class of structural element or fold exhibited by functional RNAs. The common feature of these folds is that they allow the molecule to be more compact than the classical A-form double helix (reviewed by Ferre-D'Amare and Doudna[58]) and thereby take on a more globular form, which is perhaps more suited to ligand recognition.

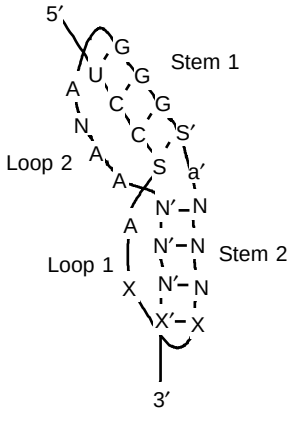

Figure 6 A reverse transcriptase-binding pseudoknot aptamer.[69] S–S′ represents a G : C or C : G base pair, X represents optional nucleotides and N–N′ represents any base pair.

3.2 Aptamer Size

It has already been noted that the sequence space available to most libraries used in aptamer isolation exceeds the practicable limit of molecules that can be handled. Moreover, for many applications it is desirable that the aptamer should be as small as possible, on costs grounds, reasons of target accessibility and so on. Nevertheless, because functional RNA folds have a finite size, the minimum length of aptamers is often larger than the size that generates a manageable sequence space. Further, the size of aptamer that is initially recovered from a process of in vitro selection is inflated by the flanking, fixed sequence regions, required for amplification and transcription, which may add up to 50 nt. A combination of deletion analysis, footprinting and in vitro synthesis can be used to determine the shortest stretch of nucleic acid that can bind to the target. The size range of minimal aptamers is fairly wide and the following give some indication. The minimum motif within VEGF aptamers was between 23 and 35 nt (depending on sequence family);[14] minimum xanthine- and guanine-binding aptamers were 32 nt long;[72] streptomycin-binding aptamers were 46 nt long;[73] and the minimum region of an aptamer that binds a serine protease of the blood (protein C) may be as much as 99 nt long.[74] This gives an M_r range of 7.5–32 kDa for aptamers, with 10 kDa being typical. The solvent-exposed surface area for a typical aptamer would be expected to be in the range 50–60 nm^2.

3.3 Aptamer Targets

Can aptamers be raised against any target molecule? Put another way, can the lack of chemical diversity among nucleic acids be compensated by the size of the sequence space that can be sampled and the resultant structural diversity of aptamers? The evidence is that aptamers can be generated against small ions, such as Zn^{2+},[36] to nucleotides such as adenosine triphosphate (ATP),[60,75] oligopeptides[76] and large glycoproteins such as CD4,[38] spanning the size range 65 Da–150 kDa, with no theoretical upper limit. The chemical classes of targets are reasonably diverse, including organic dyes,[5,77] neutral disaccharides,[78] aminoglycoside antibiotics (see Figure 7),[79] dopamine,[35] a porphyrin[80] and biotin.[71]

Nevertheless, there is evidence for a strong degree of bias, among protein-binding aptamers, for a limited number of sites. First, a surprisingly high proportion of proteins against which aptamers have been described are themselves ligands for polyanions such as nucleic acids or glycosaminoglycans. These include thrombin and other proteases of the clotting cascade,[7,74,81] a number of heparin-binding growth factors,[13,82–84] cellular transcription factors[85,86] and viral regulatory

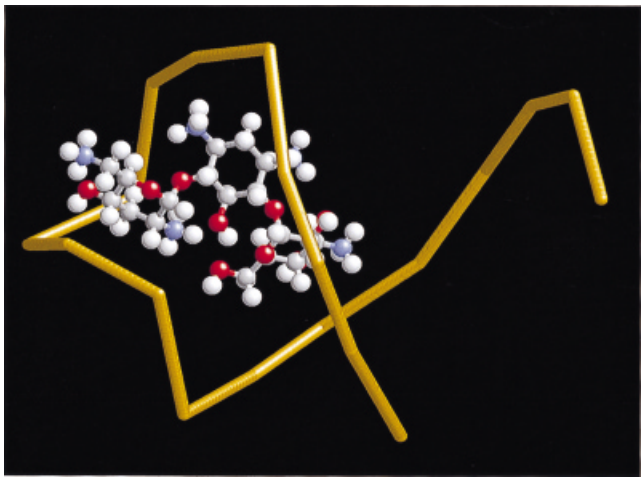

(a)

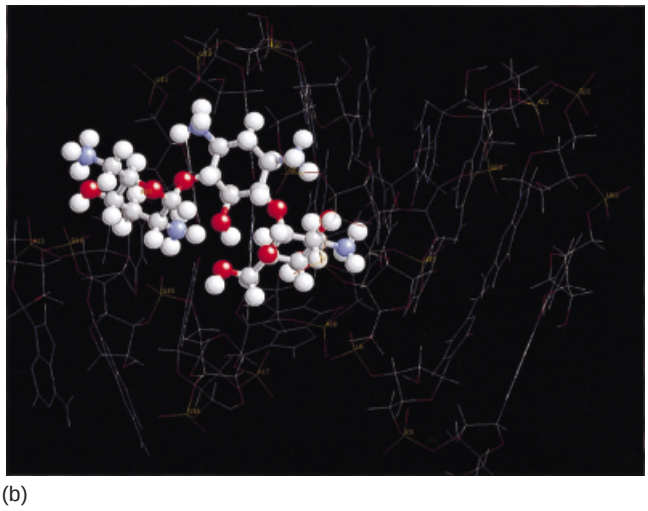

(b)

Figure 7 NMR-derived structure of a tobramycin-binding aptamer, bound to its target. (Data derived from conformer 1 of 13 published by L. Jiang and D.J. Patel, 'Solution Structure of the Tobramycin – RNA Aptamer Complex', *Nature Struct. Biol.*, **5**(9), 769–774 (1998).) (a) The phosphodiester backbone of the aptamer is depicted as an orange cylinder and the tobramycin is shown in ball-and-stick form. (b) The aptamer is shown as a wireframe model with the base identities and positions labeled.

proteins.[39,87–89] It has even been reported that some heparin-binding proteins, such as thrombin, may have natural plasma aptamers.[90]

Second, when the aptamer-binding sites (or aptatopes) on large target proteins are mapped, it is usually found that they are coincident, even if the aptamers fall into unrelated sequence families. For example, all six of the distinct classes of aptamer against VEGF were found to bind to the same region, competing with heparin and other natural ligands and being cross-linkable to the same cysteine,[14,84] and four structurally distinct aptamers against reverse transcriptase all bound the same region of the enzyme.[70] More strikingly, when five different sequence classes of aptamer against CD4 were analysed, all bound to a single region of one of the four immunoglobulin (Ig)-like domains,[38] in contrast to monoclonal antibodies, which define several epitopes in each domain.

This focusing of aptamer reactions on a small fraction of a large macromolecular target suggests that the process of in vitro selection does not proceed entirely according to simple models of multitarget partitioning.[30] Rather, it suggests that favorable and unfavorable aptatopes are distinguished by differences of affinity of many orders of magnitude. It seems likely that a major obstacle is that the Coulombic repulsion between the phosphate-containing backbone of nucleic acids and negatively charged amino acid and sugar residues at the surface of many proteins and glycoproteins produces extremely low association rates. This would mean that only a small proportion of the surface of the macromolecule would be 'visible' to aptamers, probably regions of high solvent exposure in which positively charged residues provided a degree of electrostatic steerage towards the aptatope. This has been most closely studied in the case of anti-thrombin DNA aptamers. By mutating a critical arginine to glutamate, the binding of aptamers was abolished.[91] The mutant form of thrombin was then used to raise further aptamers and these were found still to bind regions homologous to that recognized by the first-generation aptamers.[92]

3.4 Affinity

The affinity of published aptamers varies very widely. Generally, aptamers against small molecules have affinities in the micromolar range. For example, aptamers against amino acids such as citrulline and arginine range from 0.3 to 65 μM,[34,93] those against ATP and xanthine were 6 and 3.3 μM, respectively,[72,75] those against dopamine were 2.8 μM and those against vitamin B_{12} were 90 nM.[94] Aptamers to nucleic acid-binding molecules typically have affinities in the nanomolar range. For example, aptamers against retroviral integrase, reverse transcriptase and nucleocapsid proteins were 10–800, 0.3–20 and 2 nM, respectively,[70,95–97] and approximately 0.8 nM against the ribosomal RNA (rRNA)-binding aminoglycoside antibiotics.[98] Affinities in the nanomolar to subnanomolar range are found against heparin-binding proteins. Examples include platelet-derived growth factor (PDGF) (0.1 nM),[99] basic fibroblast growth factor (low nanomolar),[83] thrombin (25 nM),[7] VEGF (approximately 100 pM),[14] keratinocyte growth factor (approximately 1 pM)[13] and

the heparin-binding non-pancreatic, secretory phospholipase A2 (1.7 nM).[100] Aptamers against proteins that do not bind heparin or nucleic acids have typically lower affinities. For example, those against the oncoproteins K-Ras and Raf-1 were 0.14–1 μM and 300 nM, respectively,[101,102] one to substance P had a K_D of 190 nM[76] and one to the NS3 protein of hepatitis C virus had a K_D of 650 nM.[103] Immunoglobulin class G (IgG) superfamily domain-containing proteins seem to be able to elicit aptamers in the 2–40 nM range of affinity[12,38,104] and this might relate to their ability to interact with other cell-surface glycoproteins.

3.5 Specificity

When measured, it has been found that those targets that give rise to very high-affinity aptamers have measurable affinity for unselected RNA. For example, the *E. coli* ρ protein has an affinity for unselected RNA in the 1 μM range and can be used to isolate specific aptamers that bind with an affinity of approximately 1 nM.[33] This observation suggests that the assertion that high affinity necessarily produces high specificity[105] is unsafe, but the evidence suggests that aptamers can show a great deal of specificity. First, different enzymes with similar activity can be distinguished. For example, α-thrombin could be differentiated from γ-thrombin;[106] the reverse transcriptase of feline immunodeficiency virus could be distinguished from the homologous enzyme of three other retroviruses;[107] two isozymes of protein kinase C differing in just 23 residues could be readily distinguished;[108] and the CD4 glycoproteins of rat and mouse could be distinguished (73% identity).[38] Perhaps more impressive is the ability to discriminate between enantiomers, such as the threefold discrimination between L- and D-citrulline cit ($\Delta\Delta G = 2.5\,\mathrm{kJ\,mol^{-1}}$), sixfold difference in K_D between L- and D-arginine ($\Delta\Delta G = 4.6\,\mathrm{kJ\,mol^{-1}}$)[34] and, subsequently, the 12 000-fold ratio in K_D between L- and D-arginine reported by Geiger et al.[93] Aptamers against aminoglycoside antibiotics have been particularly specific, with anti-tobramycin aptamers having 3–4 orders of magnitude lower affinity to other aminoglycosides[98] and one against streptomycin having four orders of magnitude lower affinity to the closely related bluensomycin.[73] Strikingly, aptamers against caffeine had a 10 000-fold lower affinity for theophylline, which differs in just a methyl group.[109] However, there are limits to specificity, as one would expect. For example, aptamers against coenzyme A also recognize AMP,[90] those against xanthine recognize guanine, but not adenosine cytosine or uracil,[72] and those against the disaccharide cellobiose recognize cellulose, a polysaccharide containing the same repeating unit, but not the disaccharides lactose and maltose.[78]

4 APPLICATIONS AND DOWNSTREAM TECHNOLOGIES

In comparison with the comparable field of monoclonal antibodies, aptamers suffer from a relatively modest range of downstream technology which would enable them to be used more routinely in diagnostic and analytical assays or other practical applications. Accordingly, in the following sections, some of the advances that have been made in these areas are reviewed.

4.1 Conjugation to Detectable Moieties

Often, the first step in exploiting an antibody is its conjugation to a detectable moiety, such as a fluorochrome, an enzyme or a generic ligand such as biotin. Conjugation is most usually done by derivatizing ε-amino groups of lysines on the antibody molecule, but care has to be taken not to block lysine residues near or within the complementarity-determining regions. Conjugation can be done through carbohydrate side chains, but this is much less usual. Even more commonly, secondary antibodies, which recognize conserved epitopes on antibodies, irrespective of their antigen, are conjugated to detectable moieties, thereby simplifying the development of assays for new antigens. The challenge of aptamers is to develop a generic labeling system that does not tend to disrupt aptamer structure or hinder interaction with target.

Nucleic acids can be radioactively 'body labeled' during transcription or replication by the use of modified nucleotides. Where these are ^{3}H or ^{32}P body labeled, this does not affect the chemistry of the aptamer and thus does not affect its properties as a ligand. This is a very common way of quantitating the binding of an aptamer or a pool of nucleic acids to the target molecule.[28] In addition to body labeling, site-specific labeling with a radionuclide can be useful, for example where the aptamer is synthesized chemically. For example, in one case, a stannyl nucleotide was incorporated at the 5′ end of a DNA aptamer during chemical synthesis and tin was subsequently replaced by ^{123}I in an oxidation reaction.[110] The aim was to produce a ligand that can be used for in vivo imaging.

Aptamers can also be labeled using fluorescent or reactive groups. Here, the challenge is to introduce the detectable groups at positions that do not interfere with ligand properties. For example, a chemically synthesized aptamer against human neutrophil elastase was labeled at either the 3′ or 5′ end with fluorescein or biotin.[111] It was found that incorporation of the label reduced the affinity of the aptamer unless a substantial spacer was used. In vitro transcribed aptamers can have biotin or fluorochromes incorporates as body labels but this is generally at the expense of their properties as ligands. Labeling at the 3′ end can be achieved in a number of ways:

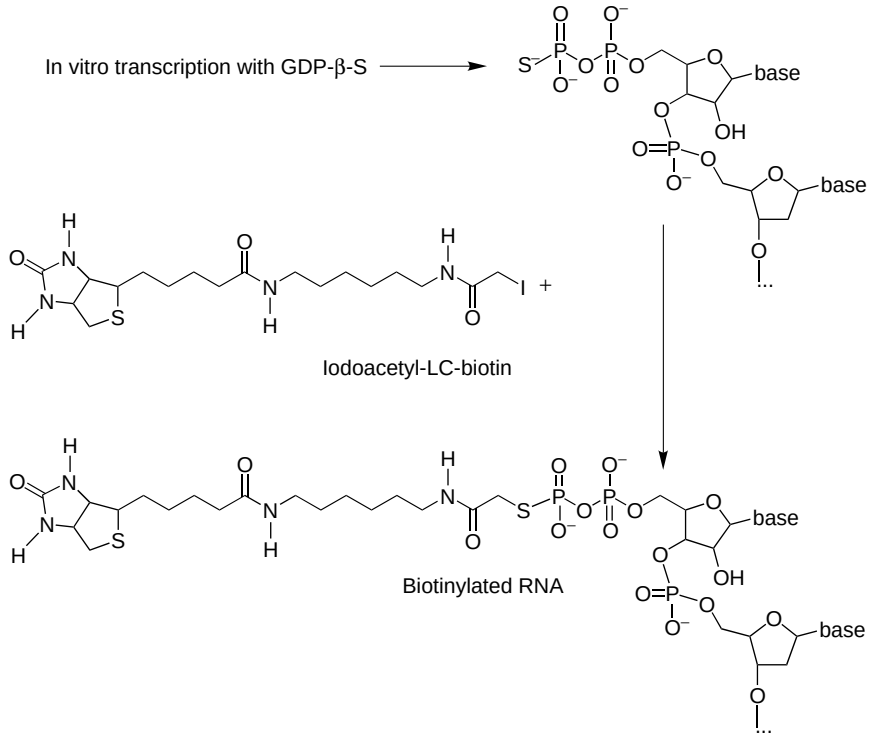

Figure 8 5′-Labeling of in vitro RNA transcripts using GTP-β-S (see text).

by templated extension using Klenow polymerase,[112] by T4 RNA ligase-mediated ligation[113] and by terminal deoxynucleotidyl transferase.[114] Labeling at the 5′ end can be achieved by the supplementation of the in vitro transcription mix with an excess of GTP-β-S, the thiol of which can then be used to attach biotin (see Figure 8).[115] This approach was used successfully to multimerize an anti-CD4 aptamer via streptavidin, which in turn allowed a fluorochrome to be attached to the aptamer for use in flow cyometry.[115] We have already discussed the problems associated with inefficient derivitization of internal 2′-amino groups.[27]

4.2 Aptamers in Biosensors and Other Detectors

As high-affinity, high-selectivity ligands, aptamers have potential in a range of detection systems, including biosensors (for reviews, see Osborne et al.[116] and Bier and Furste[117]). We have already discussed their application in flow cytometry[115,118] and they can also be used in sandwich assays akin to enzyme-linked immunosorbent assay (ELISA).[119] Fluorescently labeled aptamers have also been used to quantitate IgE and thrombin using a rapid method based on capillary electrophoresis/laser-induced fluorescence (CE/LIF), in which detection down to 50 pM was reported.[120]

Unlabeled aptamers can also be used in certain analytical methods. For example, surface plasmon resonance (SPR) methods, such as those exploited by the BIAcore system, have been used to detect activated 2′,5′-oligoadenylate synthetase and CD4.[38,121] In these cases, the aptamer was in the mobile phase and the target was immobilized. However, with appropriate derivitization and immobilization, aptamers could be used to detect specific molecules in complex mixtures using SPR.

Perhaps the most exciting method so far described also depends on an optical flow cell system and the evanascent wave from a total internal reflection event at the optical surface. However, in the application described by Potyrailo et al.,[122] the effect detected is not on plasmon resonance angle but on fluorescence anisotropy. They reported that fluorescent aptamers immobilized at the optical surface could be used to detect as little as 0.7 amol of thrombin in a 140-pL test volume in just a few minutes.

Other methods that might be applicable to aptamer-based detection in the future include thin-layer interference techniques which have been used to measure DNA – small molecule interactions in real time.[123]

4.3 Aptamer Combination

One approach to increasing the usefulness of aptamers is to combine them with another ligand. For example, a DNA aptamer to human neutrophil elastase was noninhibitory to the enzyme's protease activity but improved the K_i of a weak peptide inhibitor by five orders

of magnitude when the two were conjugated through an *N*-methoxysuccinyl link.[124] Alternatively, one can link aptamers together to form a dimeric or multimeric ligand. This can take the form of joining identical aptamers, perhaps to increase affinity by reducing the dissociation rate of the complex from a multimeric target, or of joining aptamers against different aptatopes, in order to change specificity. An example of the former approach was the homodimerization of aptamers against L-selectin, which resulted in ligands with K_D in the range of good monoclonal antibodies.[118] The procedure is not straightforward, however, as linking two aptamers in a single oligonucleotide can result in loss of function for one or both ligand moieties, either by steric hindrance or by disruption of the folding of each module. This problem was found with combinations of aptamers against coenzyme A, chloramphenicol and adenosine, but was overcome by the use of further rounds of in vitro selection on chimeric RNAs in which the junctions between aptamers were diversified.[125] A practical use for heterodimerization of aptamers has been the linking of aptamers that recognize thermostable polymerases but with different specificity.[126] The resultant chimeras had a broad specificity to enzymes from a range of thermophilic species and were potentially useful for improving the efficiency of PCR from low-copy-number templates. A further potential application of heterodimerization is to link an aptamer that recognizes a target molecule of analytical interest to a second aptamer that recognizes a detectable molecule, such as a fluorochrome.[127]

Aptamer modules have also been linked to ribozymes (catalytic RNAs) to provide a means for controlling enzyme activity through allosteric interactions. For example, the activity of a hammerhead autocatalytic ribozyme fused to an ATP-binding aptamer was shown to be regulated >100-fold by the presence of ATP.[128–130] This effect depends on the conformational change in the aptamer following ligand binding and steric hindrance between the folded aptamer and the ribozyme. These and other authors have extended this work to construct RNA-cleaving ribozymes whose activity is regulated by other small molecules, such as theophylleine and FMN,[129,131] and RNA-ligating ribozymes whose activity is regulated by ATP.[132] The general principle could form the basis of a new generation of in vitro assays for analytes.

4.4 Use of Aptamers In Vivo

Since the lead given to the field at the turn of the century by Paul Ehrlich (see, for example, Ehrlich[133]) (translation by Brock[134]), there has been a continuing desire to develop new classes of specific ligands, preferably by rational means, that could act as 'magic bullets', seeking out their molecular targets in vivo and destroying them. The possibility that aptamers might be candidate therapeutic agents has naturally received much attention (see review by Osborne et al.[116]). The main challenge to all these applications is the pharmacokinetics of oligonucleotides in vivo: the rate at which they are degraded by plasma nucleases; the rate at which they are sequestered by reticulo-endothelial cells and plasma proteins; the rate with which they are excreted by the kidney; and the rate at which they redistribute from the blood to the tissue or body fluid of interest. Natural, 2′-OH-RNA degrades very rapidly in human serum in vitro but this process can be slowed at least 1000-fold by 2′-amino modification.[135] In vivo studies on the $t_{1/2}$ of single-stranded DNA aptamers give values in the range from <2 min to approximately 8 min.[136–138] Conjugation of the aptamer to either lipids or polyethylene glycol has been reported to improve the stability and distribution kinetics of DNA aptamers sufficiently to produce therapeutic effects.[139,140] A wide range of potentially therapeutic targets for aptamers are summarized in Table 1, with a note of therapeutic effect or clinical benefit, if any.

More radically, it is conceivable that one could use aptamers in a form of gene therapy, to inhibit the function of intracellular proteins – either host-derived or viral – in vivo. This approach is fraught by all the problems that have dogged the gene therapy field for many years: inefficient and intrusive delivery systems, inefficient or transient expression, safety concerns, poor selectivity, and so on (for a review, see Palu et al.[141]). Worse, the rather low concentration of free divalent cations and the presence of high concentrations of RNA-binding proteins in the cell means that aptamers selected under the conventional in vitro conditions might well be inappropriately folded in the cell. Nevertheless, some positive reports have appeared. For example, in a study of the inducible expression of aptamers against RNA polymerase II in yeast, a significant effect on transcription was seen as a result of aptamer expression, but only in yeast strains with abnormally low levels of polymerase II expression.[142] More promisingly, U6 snRNA and $tRNA_{met}$ cassettes were used to express aptamers that bound the HIV-1 regulatory protein, Rev, in cells[143] and it was found that these, like Rev decoys derived from the natural Rev RNA ligand, were able to inhibit the expression of the virus. However, these experiments were done in vitro in a convenient but unphysiological cotransfection system, so one must be careful not to overextrapolate.

4.5 Studies on Nucleic Acid Structure

One of the first studies describing in vitro selection of nucleic acid ligands was designed to analyze the sequence requirements of a nucleic acid-binding

Table 1 Potentially therapeutic aptamers

Target	References	Comments
Toxins		
Ricin, pepocin, gypsphilin	144	Have different sequence from natural rRNA targets
Enzymes		
Thrombin	7	First identification
	138, 145	Use in vivo
	146	Better than heparin against clot-bound thrombin
	81	An RNA aptamer against thrombin
Activated plasma protein C	74	
HIV-1 reverse transcriptase	69, 96	
Other retroviral reverse transcriptases	107, 147	
HIV-1 integrase	95, 148	
Protein kinase C	108	
Human neutrophil elastase	124, 149, 150	Identification of inhibitory aptamer when conjugated to weak peptide inhibitor of enzyme. Effective in an animal model of lung injury
Hepatitis C virus NS3 protease/helicase	103, 151, 152	
Yersinia protein tyrosine phosphatase	153	
Phospholipase A2	100	Inhibited contractions of guinea pig pleural strip in vitro
2′,5′-Oligoadenylate synthetase	121	
Angiogenin (ribonuclease activity)	154	DNA aptamer prevents angiogenesis and cell proliferation
Adhesion and recognition molecules		
L-selectin	155	DNA aptamers inhibit cellular adhesion and rolling in vitro and trafficking in vivo
P-selectin	156	Affinity (20 pM) is 10^6-fold better than the natural ligand sialyl Lewis X. Unlike the latter, discriminate against E- and L-selectins by 10^4–10^5
CD4	38, 115	Disrupts immune recognition in vitro
Rhinovirus capsid protein	157	Intended to block infectivity of virus
Growth factors and hormones		
Basic fibroblast growth factor	83	
VEGF	14, 84, 140	Inhibit induction of permeability in vivo. 2′-F modification and lipid derivatization Improve stability and pharmacological properties
NGF[a]	82	
PDGF	99, 139	Effective in vivo as a PEG conjugate
Keratinocyte growth factor	13	
γ-Interferon	158, 159	
Substance P	76	
Vasopressin	18	
Regulatory proteins and oncogenes		
E2F transcription factor	86	Prevents binding to DNA and entry into S phase
K-ras	101	Raised against farnesylated peptide
Raf-1	102	
P210bcr-abl	160	DNA aptamer introduced into chronic myelogenous leukemia cells by electroporation reduced cell proliferation
MDM2 oncoprotein	161	
HIV-1 Rev	67, 88, 143, 162–164	These are Rev-responsive, even though unrelated to natural Rev-response element. Are inhibitory to HIV-1 growth in cell cultures
HTLV-I Tax	39	
Miscellaneous		
IgG against human insulin receptor	104	Possible approach to blocking a common form of antibody-mediated, autoimmune diabetes
IgE	12	Bind Fc portion of antibody, preventing interaction with cell surface receptor. Possible role in allergy therapy
Hamster PrP[a]	165	

[a] NGF, nerve growth factor; PrP, prion protein.

protein.[4] Aptamers have continued to be a powerful method for analyzing nucleic acid sequence binding requirement of a range of viral proteins. For example, aptamers have been used to dissect the nucleic acid binding sites of the coat proteins of bacteriophages R17,[166] MS2[167] and ϕ29,[168] the RNA-dependent RNA polymerase of bacteriophage Qβ,[169] the Epstein–Barr virus EBER1 RNA,[170] the regulatory, Rev-response element of HIV-1[136,171] and the HIV-1 nucleocapsid protein.[97]

The approach has been used productively to examine the interaction between components of the translational machinery and natural RNAs, e.g. the elongation factors EF-Tu,[172] elF-4B[173] and SelB,[174] the S8[175] and S1[176] ribosomal proteins from *E. coli*, the L32 ribosomal protein of yeast[177] and the decoding end of 16S rRNA.[178] Finally, transcription and post-transcriptional processing have been amenable to the aptamer approach, e.g. the requirements for the *E. coli* transcriptional terminator ρ,[33] the eukaryotic splicing factors SF2 and SC35[179] and GU-rich sequences for efficient polyadenylation.[180] It has been proposed that these methods could be used to identify all the protein–RNA interactions encoded in the genome,[181] although this is not yet an established approach.

4.6 Isolation of New Catalysts

The very first paper describing directed in vitro evolution of nucleic acids described the isolation of a totally new enzyme, a nuclease composed of DNA.[3,182] The possibility was thereby opened up that one could develop any form of useful new catalyst that could be conceived and for which a selection procedure could be devised. To an extent, this has been realized, with the isolation of enzymes that can ligate DNA in a Zn/Cu-dependent fashion[183] or ligate RNA with a similar mechanism to that of protein enzymes.[184–187] These new enzymes are not restricted to phosphodiester bond formation and breakage. For example, in vitro-selected nucleic acid enzymes have been described that cleave amide bonds[188] and alkylate halogenated peptides.[189] The combination of a fluorochrome aptamer and its target was found to have low-level oxidative activity, when coupled to its target,[77] and a hemin–aptamer complex was found to have levels of peroxidase activity comparable with protein peroxidases, and much higher than those of catalytic antibodies.[190] More significantly, RNA enzymes have been selected with amide bond-forming activity[191] in which a uridine was replaced by a 5′-imidazole derivative of uridine, capturing the chemical characteristics of histidine in nucleic acid. Also, a ribozyme was isolated that catalyzed a Diels–Alder cycloaddition reaction in which uridine is replaced by a 5-pyridyl derivative of uridine.[192] This RNA catalyst is therefore the first nucleic acid with carbon–carbon bond-forming activity.

5 PERSPECTIVES AND FUTURE DEVELOPMENTS

Clearly, the first decade of research using aptamers has opened up some very exciting avenues in basic and applied research. The power of the technique has surprised many and, even though the chemistry of nucleic acids is much less diverse than that of polypeptides and there are substantial restrictions to the technology as it stands, it is already a useful tool and promises to become more so. There are substantial advances in methods which use the principle of in vitro selection but apply it to the evolution of peptide ligands. For example, adaptations of the two-hybrid display methodology using randomized peptides displayed on thioredoxin in *E. coli* has been used to generate peptide aptamers against CDK-2 with nanomolar affinity[193] which blocks G1-S transition during the cell cycle of eukaryotic cells engineered to express it.[194,195] Similar approaches display peptides within the green fluorescent protein[196] or at the C-terminus of the lac repressor protein.[197] One problem with these methods, and the widely used phage display technology [see reviews by Burton[198] and Winter et al.,[199]] is that the critical step of passing through a living organism during the generation of the library reduces the number of different molecules that can be screened by many orders of magnitude. To overcome this, methods have been developed to link a nucleic acid based library with in vitro translation, thereby obviating the need for a living cell. The challenge is to preserve the 'genetic encoding' by maintaining the association between nucleic acid and the encoded polypeptide. One approach is to use polysomes, collections of ribosomes simultaneously translating a messenger RNA (mRNA).[200] More conveniently, a method in which translation is simultaneously arrested and the RNA linked covalently to the nascent polypeptide chain has been developed,[201] in which a puromycin group is incorporated at the 3′ end of the library RNA and is linked to the C-terminus of the polypeptide by the action of the ribosome's peptidyl transferase.

ACKNOWLEDGMENTS

I am grateful for stimulating discussions with Elmar Kraus, Jamal Ibrahim, Laura Frigotto and Abdessamad Tahiri-Alaoui and the opportunity to present their as yet

unpublished results. I thank Abdessamad for his critical reading of the manuscript.

ABBREVIATIONS AND ACRONYMS

AMP	Adenosine 5′-Monophosphate
ATP	Adenosine Triphosphate
CE/LIF	Capillary Electrophoresis/Laser-induced Fluorescence
dNTP	Deoxynucleoside Triphosphate
ELISA	Enzyme-linked Immunosorbent Assay
FMN	Flavin Mononucleotide
Ig	Immunoglobulin
IgE	Immunoglobulin E
IgG	Immunoglobulin class G
mRNA	Messenger RNA
MS	Mass Spectrometry
NGF	Nerve Growth Factor
NMR	Nuclear Magnetic Resonance
PDGF	Platelet-derived Growth Factor
PrP	Prion Protein
PVDF	Poly(vinylidene difluoride)
rRNA	Ribosomal RNA
SELEX	Selective Expansion of Ligands by Exponential Enrichment
SPR	Surface Plasmon Resonance
VEGF	Vascular Endothelial Growth Factor

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • DNA Arrays: Preparation and Application • Electroanalysis and Biosensors in Clinical Chemistry • Immunochemistry

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Molecules, Properties and Detection of Single • Nucleic Acid Structural Energetics • Polymerase Chain Reaction and Other Amplification Systems • RNA Tertiary Structure • X-ray Structures of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Protein–Oligonucleotide Interactions • Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Combinatorial Chemistry Libraries, Analysis of

REFERENCES

1. F.H. Crick, 'The Origin of the Genetic Code', *J. Mol. Biol.*, **38**(3), 367–379 (1968).
2. S. Kaur, L. McGuire, D. Tang, G. Dollinger, V. Huebner, 'Affinity Selection and Mass Spectrometry-based Strategies to Identify Lead Compounds in Combinatorial Libraries', *J. Protein Chem.*, **16**(5), 505–511 (1997).
3. D.L. Robertson, G.F. Joyce, 'Selection in Vitro of an RNA Enzyme that Specifically Cleaves Single-stranded DNA', *Nature (London)*, **344**(6265), 467–468 (1990).
4. C. Tuerk, L. Gold, 'Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase', *Science*, **249**(4968), 505–510 (1990).
5. A.D. Ellington, J.W. Szostak, 'In Vitro Selection of RNA Molecules that Bind Specific Ligands', *Nature (London)*, **346**(6287), 818–822 (1990).
6. B.E. Eaton, W.A. Pieken, 'Ribonucleosides and RNA', *Annu. Rev. Biochem.*, **64**, 837–863 (1995).
7. L.C. Bock, L.C. Griffin, J.A. Latham, E.H. Vermaas, J.J. Toole, 'Selection of Single-stranded DNA Molecules that Bind and Inhibit Human Thrombin', *Nature (London)*, **355**(6360), 564–566 (1992).
8. A.D. Ellington, J.W. Szostak, 'Selection in Vitro of Single-stranded DNA Molecules that Fold into Specific Ligand-binding Structures', *Nature (London)*, **355**(6363), 850–852 (1992).
9. W.A. Pieken, D.B. Olsen, F. Benseler, H. Aurup, F. Eckstein, 'Kinetic Characterization of Ribonuclease-resistant 2′-Modified Hammerhead Ribozymes', *Science*, **253**(5017), 314–317 (1991).
10. D.M. Williams, F. Benseler, F. Eckstein, 'Properties of 2′-Fluorothymidine-containing Oligonucleotides: Interaction with Restriction Endonuclease EcoRV', *Biochemistry*, **30**(16), 4001–4009 (1991).
11. R. Padilla, R. Sousa, 'Efficient Synthesis of Nucleic Acids Heavily Modified with Non-canonical Ribose 2′-Groups Using a Mutant T7 RNA Polymerase (RNAP)', *Nucleic Acids Res.*, **27**(6), 1561–1563 (1999).
12. T.W. Wiegand, P.B. Williams, S.C. Dreskin, M.H. Jouvin, J.P. Kinet, D. Tasset, 'High-affinity Oligonucleotide Ligands to Human IgE Inhibit Binding to Fc Epsilon Receptor I'. *J. Immunol.*, **157**(1), 221–230 (1996).
13. N.C. Pagratis, T. Fitzwater, D. Jellinek, C. Dang, 'Potent 2′-Amino- and 2′-Fluoro-2′-deoxyribonucleotide RNA Inhibitors of Keratinocyte Growth Factor', *Nature Biotechnol.*, **15**(1), 68–73 (1997).
14. J. Ruckman, L.S. Green, S. Waugh, W.L. Gillette, D.D. Henninger, L. Claesson-Welsh, N. Tanjic, '2′-Fluoropyrimidine RNA-based Aptamers to the 165-Amino Acid Form of Vascular Endothelial Growth Factor

(VEGF165). Inhibition of Receptor Binding and VEGF-induced Vascular Permeability Through Interactions Requiring the Exon 7-encoded Domain', *J. Biol. Chem.*, **273**(32), 20 556–20 567 (1998).
15. D.J. King, D.A. Ventura, A.R. Brasier, D.G. Gorenstein, 'Novel Combinatorial Selection of Phosphorothioate Oligonucleotide Aptamers', *Biochemistry*, **37**(47), 16 489–16 493 (1998).
16. S. Jhaveri, B. Olwin, A.D. Ellington, 'In Vitro Selection of Phosphorothiolated Aptamers', *Bioorg. Med. Chem. Lett.*, **8**(17), 2285–2290 (1998).
17. J. Kawakami, N. Sugimoto, 'Evolution of a Phosphorothioate RNA Library During in Vitro Selection', *Nucleic Acids Symp. Ser.*, (37), 201–202 (1997).
18. K.P. Williams, X.H. Liu, T.N. Schumacher, H.Y. Lin, D.A. Ausiello, P.S. Kim, D.P. Bartel, 'Bioactive and Nuclease-resistant L-DNA Ligand of Vasopressin', *Proc. Natl. Acad. Sci. USA*, **94**(21), 11 285–11 290 (1997).
19. S. Klussmann, A. Nolte, R. Bald, V.A. Erdmann, J.P. Furste, 'Mirror-image RNA that Binds D-Adenosine', *Nature Biotechnol.*, **14**(9), 1112–1115 (1996).
20. A. Nolte, S. Klussmann, R. Bald, V.A. Erdmann, J.P. Furste, 'Mirror-design of L-Oligonucleotide Ligands Binding to L-Arginine', *Nature Biotechnol.*, **14**(9), 1116–1119 (1996).
21. J.A. Latham, R. Johnson, J.J. Toole, 'The Application of a Modified Nucleotide in Aptamer Selection: Novel Thrombin Aptamers Containing 5-(1-Pentynyl)-2′-Deoxyuridine', *Nucleic Acids Res.*, **22**(14), 2817–2822 (1994).
22. L.E. Orgel, 'Evolution of the Genetic Apparatus', *J. Mol. Biol.*, **38**(3), 381–393 (1968).
23. T.R. Cech, 'Self-splicing RNA: Implications for Evolution', *Int. Rev. Cytol.*, **93**, 3–22 (1985).
24. K. Kruger, P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, T.R. Cech, 'Self-splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of *Tetrahymena*', *Cell*, **31**(1), 147–157 (1982).
25. M.P. Robertson, S.L. Miller, 'Prebiotic Synthesis of 5-Substituted Uracils: A Bridge Between the RNA World and the DNA–Protein World', *Science*, **268**(5211), 702–705 (1995).
26. M. Levy, S.L. Miller, 'The Prebiotic Synthesis of Modified Purines and Their Potential Role in the RNA World', *J. Mol. Evol.*, **48**(6), 631–637 (1999).
27. M.J. Kujau, S. Wölfl, 'Intramolecular Derivatization of 2′-Amino-pyrimidine Modified RNA with Functional Groups that is Compatible with Re-amplification', *Nucleic Acids Res.*, **26**(7), 1851–1853 (1998).
28. T. Fitzwater, B. Polisky, 'A SELEX Primer', *Methods Enzymol.*, **267**, 275–301 (1996).
29. D. Irvine, C. Tuerk, L. Gold, 'SELEXION. Systematic Evolution of Ligands by Exponential Enrichment with Integrated Optimization by Non-linear Analysis', *J. Mol. Biol.*, **222**(3), 739–761 (1991).
30. B. Vant-Hull, A. Payano-Baez, R.H. Davis, L. Gold, 'The Mathematics of SELEX Against Complex Targets', *J. Mol. Biol.*, **278**(3), 579–597 (1998).
31. D.H. Burke, L. Gold, 'RNA Aptamers to the Adenosine Moiety of *S*-Adenosyl Methionine: Structural Inferences from Variations on a Theme and the Reproducibility of SELEX', *Nucleic Acids Res.*, **25**(10), 2020–2024 (1997).
32. C. Tuerk, 'Using the SELEX Combinatorial Chemistry Process to Find High Affinity Nucleic Acid Ligands to Target Molecules', *Methods Mol. Biol.*, **67**, 219–230 (1997).
33. D. Schneider, L. Gold, T. Platt, 'Selective Enrichment of RNA Species for Tight Binding to *Escherichia coli* Rho Factor', *FASEB J.*, **7**(1), 201–207 (1993).
34. M. Famulok, 'Molecular Recognition of Amino Acids by RNA-aptamers – an L-Citrulline Binding RNA Motif and its Evolution into an L-Arginine Binder', *J. Am. Chem. Soc.*, **116**(5), 1698–1706 (1994).
35. C. Mannironi, A. Di Nardo, P. Fruscoloni, G.P. Tocchini-Valentini, 'In Vitro Selection of Dopamine RNA Ligands', *Biochemistry*, **36**(32), 9726–9734 (1997).
36. J. Ciesiolka, J. Gorski, M. Yarus, 'Selection of an RNA Domain that Binds Zn^{2+}', *RNA*, **1**(5), 538–550 (1995).
37. H. Takeno, T. Tanaka, Y. Kikuchi, 'RNA Aptamers to a Protease, Subtilisin', *Nucleic Acids Symp. Ser.*, (37), 249–250 (1997).
38. E. Kraus, W. James, A.N. Barclay, 'Novel RNA Ligands Able to Bind CD4 Antigen and Inhibit CD4+ T Lymphocyte Function', *J. Immunol.*, **160**(11), 5209–5212 (1998).
39. Y. Tian, N. Adya, S. Wagner, C.Z. Giam, M.R. Green, A.D. Ellington, 'Dissecting Protein : Protein Interactions Between Transcription Factors with an RNA Aptamer', *RNA*, **1**(3), 317–326 (1995).
40. R. Sägesser, E. Martinez, M. Tsagris, M. Tabler, 'Detection and Isolation of RNA-binding Proteins by RNA–Ligand Screening of a cDNA Expression Library', *Nucleic Acids Res.*, **25**(19), 3816–3822 (1997).
41. L. Frigotto, E. Kraus, A. Barclay, W. James, 'Affinity Maturation During in Vitro Selection of Aptamers: A Kinetic Study', *Nucleic Acids Res.*, submitted.
42. J.C. Cox, P. Rudolph, A.D. Ellington, 'Automated RNA Selection', *Biotechnol. Prog.*, **14**(6), 845–850 (1998).
43. M. Zuker, 'On Finding all Suboptimal Foldings of an RNA Molecule', *Science*, **244**(4900), 48–52 (1989).
44. D.H. Mathews, J. Sabina, M. Zuber, D.H. Turner, 'Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure', *J. Mol. Biol.*, **288**(5), 911–940 (1999).
45. A.R. Schroder, T. Baumstark, D. Riesner, 'Chemical Mapping of Co-existing RNA Structures', *Nucleic Acids Res.*, **26**(14), 3449–3450 (1998).
46. W. Xu, A.D. Ellington, 'Anti-peptide Aptamers Recognize Amino Acid Sequence and Bind a Protein Epitope', *Proc. Nat. Acad. Sci. USA*, **93**(15), 7475–7480 (1996).

47. Y. Yang, M. Kochoyan, P. Burgstaller, E. Westhof, M. Famulok, 'Structural Basis of Ligand Discrimination by Two Related RNA Aptamers Resolved by NMR Spectroscopy', *Science*, **272**(5266), 1343–1347 (1996).
48. J. Feigon, T. Dieckmann, F.W. Smith, 'Aptamer Structures from A to Zeta', *Chem. Biol.*, **3**(8), 611–617 (1996).
49. D.J. Patel, 'Structural Analysis of Nucleic Acid Aptamers', *Curr. Opin. Chem. Biol.*, **1**(1), 32–46 (1997).
50. K.A. Marshall, M.P. Robertson, A.D. Ellington, 'A Biopolymer by Any Other Name Would Bind as Well: A Comparison of the Ligand-binding Pockets of Nucleic Acids and Proteins', *Structure*, **5**(6), 729–734 (1997).
51. R.F. Macaya, P. Schultze, F.W. Smith, J.A. Roe, J. Feigon, 'Thrombin-binding DNA Aptamer Forms a Unimolecular Quadruplex Structure in Solution', *Proc. Natl. Acad. Sci. USA*, **90**(8), 3745–3749 (1993).
52. K. Padmanabhan, K.P. Padmanabhan, J.D. Ferrara, J.E. Sadler, A. Tulinsky, 'The Structure of Alpha-thrombin Inhibited by a 15-mer Single-stranded DNA Aptamer', *J. Biol. Chem.*, **268**(24), 17 651–17 654 (1993).
53. J.A. Kelly, J. Feigon, T.O. Yeates, 'Reconciliation of the X-ray and NMR Structures of the Thrombin-binding Aptamer d(GGTTGGTGTGGTTGG)', *J. Mol. Biol.*, **256**(3), 417–422 (1996).
54. R.F. Rando, J. Ojwang, A. Elbaggari, G.R. Reyes, R. Tinder, M.S. McGrath, M.E. Hogan, 'Suppression of Human Immunodeficiency Virus Type 1 Activity in Vitro by Oligonucleotides Which Form Intramolecular Tetrads', *J. Biol. Chem.*, **270**(4), 1754–1760 (1995).
55. N. Jing, M.E. Hogan, 'Structure–Activity of Tetrad-forming Oligonucleotides as a Potent anti-HIV Therapeutic Drug', *J. Biol. Chem.*, **273**(52), 34 992–34 999 (1998).
56. C.H. Lin, D.J. Patel, 'Encapsulating an Amino Acid in a DNA Fold', *Nature Struct. Biol.*, **3**(12), 1046–1050 (1996).
57. D.J. Patel, A.K.Suri, F. Jiang, L. Jiang, P. Fan, R.A. Kumar, S. Nonin, 'Structure, Recognition and Adaptive Binding in RNA Aptamer Complexes', *J. Mol. Biol.*, **272**(5), 645–664 (1997).
58. A.R. Ferre-D'Amare, J.A. Doudna, 'RNA Folds: Insights from Recent Crystal Structures', *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 57–73 (1999).
59. P. Fan, A.K. Suri, R. Fiala, D. Live, D.J. Patel, 'Molecular Recognition in the FMN–RNA Aptamer Complex', *J. Mol. Biol.*, **258**(3), 480–500 (1996).
60. M. Sassanfar, J.W. Szostak, 'An RNA Motif that Binds ATP', *Nature (London)*, **364**(6437), 550–553 (1993).
61. V.P. Antao, I. Tinoco, Jr, 'Thermodynamic Parameters for Loop Formation in RNA and DNA Hairpin Tetraloops', *Nucleic Acids Res.*, **20**(4), 819–824 (1992).
62. C.R. Woese, S. Winker, R.R. Gutell, 'Architecture of Ribosomal RNA: Constraints on the Sequence of ''Tetra-loops''', *Proc. Natl. Acad. Sci. USA*, **87**(21), 8467–8471 (1990).
63. T. Dieckmann, E. Suzuki, G.K. Nakamura, J. Feigon, 'Solution Structure of an ATP-binding RNA Aptamer Reveals a Novel Fold', *RNA*, **2**(7), 628–640 (1996).
64. F. Jiang, R. Fiala, D. Live, R.A. Kumar, D.J. Patel, 'RNA Folding Topology and Intermolecular Contacts in the AMP–RNA Aptamer Complex', *Biochemistry*, **35**(40), 13 250–13 266 (1996).
65. F. Jiang, R.A. Kumar, R.A. Jones, D.J. Patel, 'Structural Basis of RNA Folding and Recognition in an AMP–RNA Aptamer Complex', *Nature (London)*, **382**(6587), 183–186 (1996).
66. F. Leclerc, R. Cedergren, A.D. Ellington, 'A Three-dimensional Model of the Rev-binding Element of HIV-1 Derived from Analyses of Aptamers', *Nature Struct. Biol.*, **1**(5), 293–300 (1994).
67. A.D. Ellington, F. Leclerc, R. Cedergren, 'An RNA Groove', *Nature Struct. Biol.*, **3**(12), 981–984 (1996).
68. X. Ye, A. Gorin, A.D. Ellington, D.J. Patel, 'Deep Penetration of an Alpha-helix into a Widened RNA Major Groove in the HIV-1 Rev Peptide–RNA Aptamer Complex', *Nature Struct. Biol.*, **3**(12), 1026–1033 (1996).
69. C. Tuerk, S. MacDougal, L. Gold, 'RNA Pseudoknots that Inhibit Human Immunodeficiency Virus Type 1 Reverse Transcriptase', *Proc. Natl. Acad. Sci. USA*, **89**(15), 6988–6992 (1992).
70. D.H. Burke, L. Scates, K. Andrews, L. Gold, 'Bent Pseudoknots and Novel RNA Inhibitors of Type-1 Human-immunodeficiency-virus (HIV-1) Reverse-transcriptase', *J. Mol. Biol.*, **264**(4), 650–666 (1996).
71. C. Wilson, J. Nix, J. Szostak, 'Functional Requirements for Specific Ligand Recognition by a Biotin-binding RNA Pseudoknot', *Biochemistry*, **37**(41), 14 410–14 419 (1998).
72. D. Kiga, Y. Futamura, K. Sakamoto, S. Yokoyama, 'An RNA Aptamer to the Xanthine/Guanine Base with a Distinctive Mode of Purine Recognition', *Nucleic Acids Res.*, **26**(7), 1755–1760 (1998).
73. S.T. Wallace, R. Schroeder, 'In Vitro Selection and Characterization of Streptomycin-binding RNAs: Recognition Discrimination Between Antibiotics', *RNA*, **4**(1), 112–123 (1998).
74. S.W. Gal, S. Amontov, P.T. Urvil, D. Vishnuvardhan, F. Nishikawa, P.K. Kumar, S. Nishikawa, 'Selection of a RNA Aptamer that Binds to Human Activated Protein C and Inhibits its Protease Function', *Eur. J. Biochem.*, **252**(3), 553–562 (1998).
75. D.E. Huizenga, J.W. Szostak, 'A DNA Aptamer that Binds Adenosine and ATP', *Biochemistry*, **34**(2), 656–665 (1995).
76. D. Nieuwlandt, M. Wecker, L. Gold, 'In Vitro Selection of RNA Ligands to Substance P', *Biochemistry*, **34**(16), 5651–5659 (1995).
77. C. Wilson, J.W. Szostak, 'Isolation of a Fluorophore-specific DNA Aptamer with Weak Redox Activity', *Chem. Biol.*, **5**(11), 609–617 (1998).

78. Q. Yang, I.J. Goldstein, H.Y. Mei, D.R. Engelke, 'DNA Ligands that Bind Tightly and Selectively to Cellobiose', *Proc. Natl. Acad. Sci. USA*, **95**(10), 5462–5467 (1998).
79. M. Famulok, A. Huttenhofer, 'In Vitro Selection Analysis of Neomycin Binding RNAs with a Mutagenized Pool of Variants of the 16S rRNA Decoding Region', *Biochemistry*, **35**(14), 4265–4270 (1996).
80. Y. Li, C.R. Geyer, D. Sen, 'Recognition of Anionic Porphyrins by DNA Aptamers', *Biochemistry*, **35**(21), 6911–6922 (1996).
81. M.F. Kubik, A.W. Stephens, D. Schneider, R.A. Marlar, D. Tasset, 'High-affinity RNA Ligands to Human Alpha-thrombin', *Nucleic Acids Res.*, **22**(13), 2619–2626 (1994).
82. J. Binkley, P. Allen, D.M. Brown, L. Green, C. Tuerk, L. Gold, 'RNA Ligands to Human Nerve Growth Factor', *Nucleic Acids Res.*, **23**(16), 3198–3205 (1995).
83. D. Jellinek, C.K. Lynott, D.B. Rifkin, N. Janjic, 'High-affinity RNA Ligands to Basic Fibroblast Growth Factor Inhibit Receptor Binding', *Proc. Natl. Acad. Sci. USA*, **90**(23), 11 227–11 231 (1993).
84. D. Jellinek, L.S. Green, C. Bell, N. Janjic, 'Inhibition of Receptor Binding by High-affinity RNA Ligands to Vascular Endothelial Growth Factor', *Biochemistry*, **33**(34), 10 450–10 456 (1994).
85. L.L. Lebruska, L.J. Maher, 3rd, 'Selection and Characterization of an RNA Decoy for Transcription Factor NF-kappa B', *Biochemistry*, **38**(10), 3168–3174 (1999).
86. J. Ishizaki, J.R. Nevins, B.A. Sullenger, 'Inhibition of Cell Proliferation by an RNA Ligand that Selectively Blocks E2F Function', *Nature Med.*, **2**(12), 1386–1389 (1996).
87. P. Allen, B. Collins, L. Gold, 'Selex-Derived RNAs with High-affinity and Specificity to HIV-1 Integrase and Nucleocapsid Proteins', *FASEB J.*, **8**(7), A1268 (1994).
88. L. Giver, D.P. Bartel, M.L. Zapp, M.R. Green, A.D. Ellington, 'Selection and Design of High-affinity RNA Ligands for HIV-1 Rev', *Gene*, **137**(1), 19–24 (1993).
89. R. Yamamoto, S. Toyoda, P. Viljanen, K. Machida, S. Nishikawa, K. Murakami, K. Taira, P.K. Kumar, 'In Vitro Selection of RNA Aptamers that Can Bind Specifically to Tat Protein of HIV-1', *Nucleic Acids Symp. Ser.*, (34), 145–146 (1995).
90. D.H. Burke, D.C. Hoffman, 'A Novel Acidophilic RNA Motif that Recognizes Coenzyme A', *Biochemistry*, **37**(13), 4653–4663 (1998).
91. M. Tsiang, C.S. Gibbs, L.C. Griffin, K.E. Dunn, L.L. Leung, 'Selection of a Suppressor Mutation that Restores Affinity of an Oligonucleotide Inhibitor for Thrombin Using In Vitro Genetics', *J. Biol. Chem.*, **270**(33), 19 370–19 376 (1995).
92. M. Tsiang, A.K. Jain, K.E. Dunn, M.E. Rojas, L.L.K. Leung, C.S. Gibbs, 'Functional Mapping of the Surface Residues of Human Thrombin', *J. Biol. Chem.*, **270**(28), 16 854–16 863 (1995).
93. A. Geiger, P. Burgstaller, H. von der Eltz, A. Roeder, M. Famulok, 'RNA Aptamers that Bind L-Arginine with Sub-micromolar Dissociation Constants and High Enantioselectivity', *Nucleic Acids Res.*, **24**(6), 1029–1036 (1996).
94. J.R. Lorsch, J.W. Szostak, 'In Vitro Selection of RNA Aptamers Specific for Cyanocobalamin', *Biochemistry*, **33**(4), 973–982 (1994).
95. P. Allen, S. Worland, L. Gold, 'Isolation of High-affinity RNA Ligands to HIV-1 Integrase from a Random Pool', *Virology*, **209**(2), 327–336 (1995).
96. D.J. Schneider, J. Feigon, Z. Hostomsky, L. Gold, 'High-affinity ssDNA Inhibitors of the Reverse Transcriptase of Type 1 Human Immunodeficiency Virus', *Biochemistry*, **34**(29), 9599–9610 (1995).
97. P. Allen, B. Collins, D. Brown, Z. Hostomsky, L. Gold, 'A Specific RNA Structural Motif Mediates High Affinity Binding by the HIV-1 Nucleocapsid Protein (NCp7)', *Virology*, **225**(2), 306–315 (1996).
98. Y. Wang, J. Killian, K. Hamasaki, R.R. Rando, 'RNA Molecules that Specifically and Stoichiometrically Bind Aminoglycoside Antibiotics with High Affinities', *Biochemistry*, **35**(38), 12 338–12 346 (1996).
99. L.S. Green, D. Jellinek, R. Jenison, A. Ostman, C.H. Heldin, N. Janjic, 'Inhibitory DNA Ligands to Platelet-derived Growth Factor B-chain', *Biochemistry*, **35**(45), 14 413–14 424 (1996).
100. P. Bridonneau, Y.F. Chang, D. O'Connell, S.C. Gill, D.W. Snyder, L. Johnson, T. Goodson, Jr, D.K. Herron, D.H. Parma, 'High-affinity Aptamers Selectively Inhibit Human Nonpancreatic Secretory Phospholipase A2 (hnps-PLA2)', *J. Med. Chem.*, **41**(6), 778–786 (1998).
101. B.A. Gilbert, M. Sha, S.T. Wathen, R.R. Rando, 'RNA Aptamers that Specifically Bind to a K Ras-derived Farnesylated Peptide', *Bioorg. Med. Chem.*, **5**(6), 1115–1122 (1997).
102. M. Kimoto, K. Sakamoto, M. Shirouzu, I. Hirao, S. Yokoyama, 'RNA Aptamers that Specifically Bind to the Ras-binding Domain of Raf-1', *FEBS Lett.*, **441**(2), 322–326 (1998).
103. P.T. Urvil, N. Kakiuchi, D.M. Zhou, K. Shimotohno, P.K. Kumar, S. Nishikawa, 'Selection of RNA Aptamers that Bind Specifically to the NS3 Protease of Hepatitis C Virus', *Eur. J. Biochem.*, **248**(1), 130–138 (1997).
104. J.A. Doudna, T.R. Cech, B.A. Sullenger, 'Selection of an RNA Molecule that Mimics a Major Autoantigenic Epitope of Human Insulin-receptor', *Proc. Natl. Acad. Sci. USA*, **92**(6), 2355–2359 (1995).
105. B.E. Eaton, L. Gold, D.A. Zichi, 'Let's Get Specific: the Relationship Between Specificity and Affinity', *Chem. Biol.*, **2**(10), 633–638 (1995).
106. L.R. Paborsky, S.N. McCurdy, L.C. Griffin, J.J. Toole, L.L. Leung, 'The Single-stranded DNA Aptamer-binding Site of Human Thrombin', *J. Biol. Chem.*, **268**(28), 20 808–20 811 (1993).
107. H. Chen, D.G. McBroom, Y.Q. Zhu, L. Gold, T.W. North, 'Inhibitory RNA Ligand to Reverse Transcriptase

from Feline Immunodeficiency Virus', *Biochemistry*, **35**(21), 6923–6930 (1996).
108. R. Conrad, L.M. Keranen, A.D. Ellington, A.C. Newton, 'Isozyme-specific Inhibition of Protein Kinase C by RNA Aptamers', *J. Biol. Chem.*, **269**(51), 32 051–32 054 (1994).
109. R.D. Jenison, S.C. Gill, A. Pardi, B. Polisky, 'High-resolution Molecular Discrimination by RNA', *Science*, **263**(5152), 1425–1429 (1994).
110. H. Dougan, J.B. Hobbs, J.I. Weitz, D.M. Lyster, 'Synthesis and Radioiodination of a Stannyl Oligodeoxyribonucleotide', *Nucleic Acids Res.*, **25**(14), 2897–2901 (1997).
111. K.A. Davis, B. Abrams, Y. Lin, S.D. Jayasena, 'Use of a High Affinity DNA Ligand in Flow Cytometry', *Nucleic Acids Res.*, **24**(4), 702–706 (1996).
112. Z. Huang, J. Szostak, 'A Simple Method for 3′-Labeling of RNA', *Nucleic Acids Res.*, **24**(21), 4360–4361 (1996).
113. T. England, O. Uhlenbeck, '3′-Terminal Labelling of RNA with T4 RNA Ligase', *Nature (London)*, **275**, 560–561 (1978).
114. V. Rosemeyer, A. Laubrock, R. Seibl, 'Nonradioactive 3′-End Labeling of RNA Molecules of Different Lengths by Terminal Deoxynucleotidyl Transferase', *Anal. Biochem.*, **224**, 446–449 (1995).
115. K.A. Davis, Y. Lin, B. Abrams, S.D. Jayasena, 'Staining of Cell Surface Human CD4 with 2′-F-pyrimidine-containing RNA Aptamers for Flow Cytometry', *Nucleic Acids Res.*, **26**(17), 3915–3924 (1998).
116. S.E. Osborne, I. Matsumura, A.D. Ellington, 'Aptamers as Therapeutic and Diagnostic Reagents: Problems and Prospects', *Curr. Opin. Chem. Biol.*, **1**, 5–9 (1997).
117. F.F. Bier, J.P. Furste, 'Nucleic Acid Based Sensors', in *Frontiers in Biosensorics*, eds. F.W Scheller, F. Schubert, J. Fedrowitz, Birkhäuser, Basel, 97–120, 1997.
118. S. Ringquist, D. Parma, 'Anti-L-Selectin Oligonucleotide Ligands Recognize CD62L-positive Leukocytes: Binding Affinity and Specificity of Univalent and Bivalent Ligands', *Cytometry*, **33**(4), 394–405 (1998).
119. D.W. Drolet, L. Moon-Mcdermott, T.S. Romig, 'An Enzyme-linked Oligonucleotide Assay', *Nature Biotechnol.*, **14**(8), 1021–1027 (1996).
120. I. German, D.D. Buchanan, R.T. Kennedy, 'Aptamers as Ligands in Affinity Probe Capillary Electrophoresis', *Anal. Chem.*, **70**(21), 4540–4545 (1998).
121. R. Hartmann, P.L. Norby, P.M. Martensen, P. Jorgensen, M.C. James, C. Jacobsen, S.K. Moestrup, M.J. Clemens, J. Justesen, 'Activation of 2′–5′ Oligoadenylate Synthetase by Single-stranded and Double-stranded RNA Aptamers', *J. Biol. Chem.*, **273**(6), 3236–3246 (1998).
122. R.A. Potyrailo, R.C. Conrad, A.D. Ellington, G.M. Hieftje, 'Adapting Selected Nucleic Acid Ligands (Aptamers) to Biosensors', *Anal. Chem.*, **70**(16), 3419–3425 (1998).
123. J. Piehler, A. Brecht, G. Gauglitz, M. Zerlin, C. Maul, R. Thiericke, S. Grabley, 'Label-free Monitoring of DNA–Ligand Interactions', *Anal. Biochem.*, **249**(1), 94–102 (1997).
124. Y. Lin, A. Padmapriya, K.M. Morden, S.D. Jayasena, 'Peptide Conjugation to an in Vitro-selected DNA Ligand Improves Enzyme Inhibition', *Proc. Natl. Acad. Sci. USA*, **92**(24), 11 044–11 048 (1995).
125. D.H. Burke, J.H. Willis, 'Recombination, RNA Evolution, and Bifunctional RNA Molecules Isolated Through Chimeric SELEX', *RNA*, **4**(9), 1165–1175 (1998).
126. Y. Lin, S.D. Jayasena, 'Inhibition of Multiple Thermostable DNA Polymerases by a Heterodimeric Aptamer', *J. Mol. Biol.*, **271**(1), 100–111 (1997).
127. L.A. Holeman, S.L. Robinson, J.W. Szostak, C. Wilson, 'Isolation and Characterization of Fluorophore-binding RNA Aptamers', *Folding Des.*, **3**(6), 423–431 (1998).
128. J. Tang, R.R. Breaker, 'Mechanism for Allosteric Inhibition of an ATP-sensitive Ribozyme', *Nucleic Acids Res.*, **26**(18), 4214–4221 (1998).
129. J. Tang, R.R. Breaker, 'Examination of the Catalytic Fitness of the Hammerhead Ribozyme by in Vitro Selection', *RNA*, **3**(8), 914–925 (1997).
130. J. Tang, R.R. Breaker, 'Rational Design of Allosteric Ribozymes', *Chem. Biol.*, **4**(6), 453–459 (1997).
131. M. Araki, Y. Okuno, Y. Hara, Y. Sugiura, 'Allosteric Regulation of a Ribozyme Activity Through Ligand-induced Conformational Change', *Nucleic Acids Res.*, **26**(14), 3379–3384 (1998).
132. M.P. Robertson, A.D. Ellington, 'In Vitro Selection of an Allosteric Ribozyme that Transduces Analytes to Amplicons', *Nature Biotechnol.*, **17**(1), 62–66 (1999).
133. P. Ehrlich, 'Über moderne Chemotherapie', in *Beiträge zur Experimentellen Pathologie und Chemotherapie*, Akademische Verlagsgesellschaft, Leipzig, 167–202, 1909.
134. T. Brock, *Milestones in Microbiology*, American Society for Microbiology, Washington, DC, 176–184, 1961.
135. D. Jellinek, L.S. Green, C. Bell, C.K. Lynott, N. Gill, C. Vargeese, G. Kirschenheuter, D.P. McGee, P. Abesinghe, W.A. Pieken, R. Shapiro, D.B. Rifkin, D. Moscatelli, N. Janjic, 'Potent 2′-Amino-2′-deoxypyrimidine RNA Inhibitors of Basic Fibroblast Growth Factor', *Biochemistry*, **34**(36), 11 363–11 372 (1995).
136. L.C. Griffin, G.F. Tidmarsh, L.C. Bock, J.J. Toole, L.L. Leung, 'In Vivo Anticoagulant Properties of a Novel Nucleotide-based Thrombin Inhibitor and Demonstration of Regional Anticoagulation in Extracorporeal Circuits', *Blood*, **81**(12), 3271–3276 (1993).
137. L. Reyderman, S. Stavchansky, 'Pharmacokinetics and Biodistribution of a Nucleotide-based Thrombin Inhibitor in Rats', *Pharm. Res.*, **15**(6), 904–910 (1998).
138. A. DeAnda, Jr, S.E. Coutre, M.R. Moon, C.M. Vial, L.C. Griffin, V.S. Law, M. Komeda, L.L.K. Leung, D.C. Miller, 'Pilot Study of the Efficacy of a Thrombin Inhibitor for Use During Cardiopulmonary Bypass', *Ann. Thoracic Surg.*, **58**(2), 344–350 (1994).

139. J. Floege, T. Ostendorf, U. Janssen, M. Burg, H.H. Radeke, C. Vargeese, S.C. Gill, L.S. Green, N. Janjic, 'Novel Approach to Specific Growth Factor Inhibition in Vivo: Antagonism of Platelet-derived Growth Factor in Glomerulonephritis by Aptamers', *Am. J. Pathol.*, **154**(1), 169–179 (1999).
140. M.C. Willis, B.D. Collins, T. Zhang, L.S. Green, D.P. Sebesta, C. Bell, E. Kellogg, S.C. Gill, A. Magallanez, S. Knauer, R.A. Bendele, P.S. Gill, N. Janjic, B. Collins, 'Liposome-anchored Vascular Endothelial Growth Factor Aptamers', *Bioconjug. Chem.*, **9**(5), 573–582 (1998); Erratum: *Bioconj. Chem.*, **9**(5), 633 (1998).
141. G. Palu, R. Bonaguro, A. Marcello, 'In Pursuit of New Developments for Gene Therapy of Human Diseases', *J. Biotechnol.*, **68**(1), 1–13 (1999).
142. M. Thomas, S. Chédin, C. Carles, M. Riva, M. Famulok, A. Sentenac, 'Selective Targeting and Inhibition of Yeast RNA Polymerase II by RNA Aptamers', *J. Biol. Chem.*, **272**(44), 27 980–27 986 (1997).
143. P.D. Good, A.J. Krikop, S.X. Li, E. Bertrand, N.S. Lee, L. Giver, A. Ellington, J.A. Zaia, J.J. Rossi, D.R. Engelke, 'Expression of Small, Therapeutic RNAs in Human Cell Nuclei', *Gene Ther.*, **4**(1), 45–54 (1997).
144. I. Hirao, S. Yoshinari, S. Yokoyama, Y. Endo, A.D. Ellington, 'In Vitro Selection of Aptamers that Bind to Ribosome-inactivating Toxins', *Nucleic Acids Symp. Ser.*, (37), 283–284 (1997).
145. L.C. Griffin, J.J. Toole, L.L. Leung, 'The Discovery and Characterization of a Novel Nucleotide-based Thrombin Inhibitor', *Gene*, **137**(1), 25–31 (1993).
146. W.X. Li, A.V. Kaplan, G.W. Grant, J.J. Toole, L.L. Leung, 'A Novel Nucleotide-based Thrombin Inhibitor Inhibits Clot-bound Thrombin and Reduces Arterial Platelet Thrombus Formation', *Blood*, **83**(3), 677–682 (1994).
147. H. Chen, L. Gold, 'Selection of High-affinity RNA Ligands to Reverse Transcriptase: Inhibition of cDNA Synthesis and RNase H Activity', *Biochemistry*, **33**(29), 8746–8756 (1994).
148. N. Jing, R.F. Rando, Y. Pommier, M.E. Hogan, 'Ion Selective Folding of Loop Domains in a Potent Anti-HIV Oligonucleotide', *Biochemistry*, **36**(41), 12 498–12 505 (1997).
149. N.M. Bless, D. Smith, J. Charlton, B.J. Czermak, H. Schmal, H.P. Friedl, P.A. Ward, 'Protective Effects of an Aptamer Inhibitor of Neutrophil Elastase in Lung Inflammatory Injury', *Curr. Biol.*, **7**(11), 877–880 (1997).
150. J. Charlton, G.P. Kirschenheuter, D. Smith, 'Highly Potent Irreversible Inhibitors of Neutrophil Elastase Generated by Selection from a Randomized DNA–Valine Phosphonate Library', *Biochemistry*, **36**(10), 3018–3026 (1997).
151. K. Fukuda, D. Vishnuvardhan, S. Sekiya, N. Kakiuchi, K. Shimotohno, P.K. Kumar, S. Nishikawa, 'Specific RNA Aptamers to NS3 Protease Domain of Hepatitis C Virus', *Nucleic Acids Symp. Ser.*, (37), 237–238 (1997).
152. P.K. Kumar, K. Machida, P.T. Urvil, N. Kakiuchi, D. Vishnuvardhan, K. Shimotohno, K. Taira, S. Nishikawa, 'Isolation of RNA Aptamers Specific to the NS3 Protein of Hepatitis C Virus from a Pool of Completely Random RNA', *Virology*, **237**(2), 270–282 (1997).
153. S.D. Bell, J.M. Denu, J.E. Dixon, A.D. Ellington, 'RNA Molecules that Bind to and Inhibit the Active Site of a Tyrosine Phosphatase', *J. Biol. Chem.*, **273**(23), 14 309–14 314 (1998).
154. V. Nobile, N. Russo, G. Hu, J.F. Riordan, 'Inhibition of Human Angiogenin by DNA Aptamers: Nuclear Colocalization of an Angiogenin–Inhibitor Complex', *Biochemistry*, **37**(19), 6857–6863 (1998).
155. B.J. Hicke, S.R. Watson, A. Koenig, C.K. Lynott, R.F. Bargatze, Y.F. Chang, S. Ringquist, L. Moon-McDermott, S. Jennings, T. Fitzwater, H.L. Han, N. Varki, I. Albinana, M.C. Willis, A. Varki, D. Parma, 'DNA Aptamers Block L-Selectin Function in Vivo. Inhibition of Human Lymphocyte Trafficking in SCID Mice', *J. Clin. Invest.*, **98**(12), 2688–2692 (1996).
156. R.D. Jenison, S.D. Jennings, D.W. Walker, R.F. Bargatze, D. Parma, 'Oligonucleotide Inhibitors of P-Selectin-dependent Neutrophil–Platelet Adhesion', *Antisense Nucleic Acid Drug Dev.*, **8**(4), 265–279 (1998).
157. A. De Beuckelaer, J.P. Fürste, 'Selection of High Affinity RNA Ligands to a Synthetic Peptide of Human Rhinovirus 14 Coat Protein', *Nucleic Acids Symp. Ser.*, (33), 23 (1995).
158. P.P. Lee, M. Ramanathan, C.A. Hunt, M.R. Garovoy, 'An Oligonucleotide Blocks Interferon-gamma Signal Transduction', *Transplantation*, **62**(9), 1297–1301 (1996).
159. M.F. Kubik, C. Bell, T. Fitzwater, S.R. Watson, D.M. Tasset, 'Isolation and Characterization of 2′-Fluoro-, 2′-Amino-, and 2′-Fluoro/amino-modified RNA Ligands to Human IFN-gamma that Inhibit Receptor Binding', *J. Immunol.*, **159**(1), 259–267 (1997).
160. G.N. Schwartz, Y.Q. Liu, J. Tisdale, K. Walshe, D. Fowler, R. Gress, R.C. Bergan, 'Growth Inhibition of Chronic Myelogenous Leukemia Cells by ODN-1, an Aptameric Inhibitor of p210bcr-abl Tyrosine Kinase Activity', *Antisense Nucleic Acid Drug Dev.*, **8**(4), 329–339 (1998).
161. B. Elenbaas, M. Dobbelstein, J. Roth, T. Shenk, A.J. Levine, 'The MDM2 Oncoprotein Binds Specifically to RNA Through its Ring Finger Domain', *Mol. Med.*, **2**(4), 439–451 (1996).
162. K.B. Jensen, B.L. Atkinson, M.C. Willis, T.H. Koch, L. Gold, 'Using in Vitro Selection to Direct the Covalent Attachment of Human Immunodeficiency Virus Type 1 Rev Protein to High-affinity RNA Ligands', *Proc. Natl Acad. Sci. USA*, **92**(26), 12 220–12 224 (1995).
163. L. Giver, D. Bartel, M. Zapp, A. Pawul, M. Green, A.D. Ellington, 'Selective Optimization of the Rev-binding Element of HIV-1', *Nucleic Acids Res.*, **21**(23), 5509–5516 (1993).

164. T.L. Symensma, L. Giver, M. Zapp, G.B. Takle, A.D. Ellington, 'RNA Aptamers Selected to Bind Human Immunodeficiency Virus Type 1 Rev in Vitro are Rev Responsive in Vivo', *J. Virol.*, **70**(1), 179–187 (1996).
165. S. Weiss, D. Proske, M. Neumann, M.H. Groschup, H.A. Kretzschmar, M. Famulok, E.L. Winnacker, 'RNA Aptamers Specifically Interact with the Prion Protein PrP', *J. Virol.*, **71**(11), 8790–8797 (1997).
166. D. Schneider, C. Tuerk, L. Gold, 'Selection of High Affinity RNA Ligands to the Bacteriophage R17 Coat Protein', *J. Mol. Biol.*, **228**(3), 862–869 (1992).
167. M.A. Convery, S. Rowsell, N.J. Stonehouse, A.D. Ellington, I. Hirao, J.B. Murray, D.S. Peabody, S.E. Phillips, P.G. Stockley, 'Crystal Structure of an RNA Aptamer–Protein Complex at 2.8 Å Resolution', *Nature Struct. Biol.*, **5**(2), 133–139 (1998).
168. F. Zhang, D. Anderson, 'In Vitro Selection of Bacteriophage Phi29 Prohead RNA Aptamers for Prohead Binding', *J. Biol. Chem.*, **273**(5), 2947–2953 (1998).
169. D. Brown, L. Gold, 'Template Recognition by an RNA-dependent RNA Polymerase: Identification and Characterization of Two RNA Binding Sites on Qβ Replicase', *Biochemistry*, **34**(45), 14 765–14 774 (1995).
170. M. Dobbelstein, T. Shenk, 'In Vitro Selection of RNA Ligands for the Ribosomal L22 Protein Associated with Epstein–Barr Virus-expressed RNA by Using Randomized and cDNA-derived RNA Libraries', *J. Virol.*, **69**(12), 8027–8034 (1995).
171. K.B. Jensen, L. Green, S. MacDougal-Waugh, C. Tuerk, 'Characterization of an in Vitro-selected RNA Ligand to the HIV-1 Rev Protein', *J. Mol. Biol.*, **235**(1), 237–247 (1994).
172. V. Hornung, H.P. Hofmann, M. Sprinzl, 'In Vitro Selected RNA Molecules that Bind to Elongation Factor Tu', *Biochemistry*, **37**(20), 7260–7267 (1998).
173. N. Methot, G. Pickett, J.D. Keene, N. Sonenberg, 'In Vitro RNA Selection Identifies RNA Ligands that Specifically Bind to Eukaryotic Translation Initiation Factor 4B: the Role of the RNA Remotif', *RNA*, **2**(1), 38–50 (1996).
174. S.J. Klug, A. Hüttenhofer, M. Kromayer, M. Famulok, 'In Vitro and in Vivo Characterization of Novel mRNA Motifs that Bind Special Elongation Factor SelB', *Proc. Natl. Acad. Sci. USA*, **94**(13), 6676–6681 (1997).
175. H. Moine, C. Cachia, E. Westhof, B. Ehresmann, C. Ehresmann, 'The RNA Binding Site of S8 Ribosomal Protein of *Escherichia coli*: Selex and Hydroxyl Radical Probing Studies', *RNA*, **3**(3), 255–268 (1997).
176. S. Ringquist, T. Jones, E.E. Snyder, T. Gibson, I. Boni, L. Gold, 'High-affinity RNA Ligands to *Escherichia coli* Ribosomes and Ribosomal Protein S1: Comparison of Natural and Unnatural Binding Sites', *Biochemistry*, **34**(11), 3640–3648 (1995).
177. H. Li, S.A. White, 'RNA Aptamers for Yeast Ribosomal Protein L32 have a Conserved Purine-rich Internal Loop', *RNA*, **3**(3), 245–254 (1997).
178. P. Purohit, S. Stern, 'Interactions of a Small RNA with Antibiotic and RNA Ligands of the 30S Subunit', *Nature (London)*, **370**(6491), 659–662 (1994).
179. R. Tacke, J.L. Manley, 'The Human Splicing Factors ASF/SF2 and SC35 Possess Distinct, Functionally Significant RNA Binding Specificities', *EMBO J.*, **14**(14), 3540–3551 (1995).
180. K. Beyer, T. Dandekar, W. Keller, 'RNA Ligands Selected by Cleavage Stimulation Factor Contain Distinct Sequence Motifs that Function as Downstream Elements in 3′-end Processing of Pre-mRNA', *J. Biol. Chem.*, **272**(42), 26 769–26 779 (1997).
181. B.S. Singer, T. Shtatland, D. Brown, L. Gold, 'Libraries for Genomic SELEX', *Nucleic Acids Res.*, **25**(4), 781–786 (1997); Erratum: *Nucleic Acids Res.*, **25**(21), 4430 (1997).
182. A.A. Beaudry, G.F. Joyce, 'Directed Evolution of an RNA Enzyme', *Science*, **257**(5070), 635–641 (1992).
183. B. Cuenoud, J.W. Szostak, 'A DNA Metalloenzyme with DNA Ligase Activity', *Nature (London)*, **375**(6532), 611–614 (1995).
184. D.P. Bartel, J.W. Szostak, 'Isolation of New Ribozymes from a Large Pool of Random Sequences', *Science*, **261**(5127), 1411–1418 (1993).
185. E.H. Ekland, D.P. Bartel, 'The Secondary Structure and Sequence Optimization of an RNA Ligase Ribozyme', *Nucleic Acids Res.*, **23**(16), 3231–3238 (1995).
186. E.H. Ekland, J.W. Szostak, D.P. Bartel, 'Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences', *Science*, **269**(5222), 364–370 (1995).
187. A.J. Hager, J.W. Szostak, 'Isolation of Novel Ribozymes that Ligate AMP-activated RNA Substrates', *Chem. Biol.*, **4**(8), 607–617 (1997).
188. X. Dai, A. De Mesmaeker, G.F. Joyce, 'Cleavage of an Amide Bond by a Ribozyme', *Science*, **267**(5195), 237–240 (1995).
189. M. Wecker, D. Smith, L. Gold, 'In Vitro Selection of a Novel Catalytic RNA: Characterization of a Sulfur Alkylation Reaction and Interaction with a Small Peptide', *RNA*, **2**(10), 982–994 (1996).
190. P. Travascio, Y. Li, D. Sen, 'DNA-enhanced Peroxidase Activity of a DNA–Aptamer–Hemin Complex', *Chem. Biol.*, **5**(9), 505–517 (1998).
191. T.W. Wiegand, R.C. Janssen, B.E. Eaton, 'Selection of RNA Amide Synthases', *Chem. Biol.*, **4**(9), 675–683 (1997).
192. T.M. Tarasow, S.L. Tarasow, B.E. Eaton, 'RNA-catalysed Carbon–Carbon Bond Formation', *Nature (London)*, **389**(6646), 54–57 (1997).
193. P. Colas, B. Cohen, T. Jessen, I. Grishina, J. McCoy, R. Brent, 'Genetic Selection of Peptide Aptamers

that Recognize and Inhibit Cyclin-dependent Kinase 2', *Nature (London)*, **380**(6574), 548–550 (1996).
194. B.A. Cohen, P. Colas, R. Brent, 'An Artificial Cell-cycle Inhibitor Isolated from a Combinatorial Library', *Proc. Natl Acad. Sci. USA*, **95**(24), 14 272–14 277 (1998).
195. M.G. Kolonin, R.I. Finley, Jr, 'Targeting Cyclin-dependent Kinases in *Drosophila* with Peptide Aptamers', *Proc. Natl. Acad. Sci. USA*, **95**(24), 14 266–14 271 (1998).
196. M.R. Abedi, G. Caponigro, A. Kamb, 'Green Fluorescent Protein as a Scaffold for Intracellular Presentation of Peptides', *Nucleic Acids Res.*, **26**(2), 623–630 (1998).
197. C.M. Gates, W.P. Stemmer, R. Kaptein, P.J. Schatz, 'Affinity Selective Isolation of Ligands from Peptide Libraries Through Display on a Lac Repressor ''Headpiece Dimer''', *J. Mol. Biol.*, **255**(3), 373–386 (1996).
198. D.R. Burton, 'Phage Display', *Immunotechnology*, **1**(2), 87–94 (1995).
199. G. Winter, A.D. Griffiths, R.E. Hawkins, H.R. Hoogenboom, 'Making Antibodies by Phage Display Technology', *Annu. Rev. Immunol.*, **12**, 433–455 (1994).
200. G.M. Gersuk, M.J. Corey, E. Corey, J.E. Stray, G.H. Kawasaki, R.L. Vessella, 'High-affinity Peptide Ligands to Prostate-specific Antigen Identified by Polysome Selection', *Biochem. Biophys. Res. Commun.*, **232**(2), 578–582 (1997).
201. R.W. Roberts, J.W. Szostak, 'RNA–Peptide Fusions for the in Vitro Selection of Peptides and Proteins', *Proc. Natl. Acad. Sci. USA*, **94**(23), 12 297–12 302 (1997).

Capillary Electrophoresis of Nucleic Acids

Kunio Kawamura
Osaka Prefecture University, Osaka, Japan

1 Introduction **1**
2 History **2**
3 Instrumentation of Capillary Electrophoresis **3**
3.1 Hardware 3
3.2 Sample Preparation 3
4 Separation and Detection Methods **4**
4.1 Separation Mode and Theory of Nucleic Acids Migration 4
4.2 Detection Methods 8
5 Practical Analysis and Applications **10**
5.1 Chemical Structures 10
5.2 Size, Sequence and Conformation Analysis 12
5.3 Applications and Needs 15
6 Limitation and Comparison with Other Methods **15**
6.1 Comparison with High-performance Liquid Chromatography 15
6.2 Comparison with Slab Gel Electrophoresis 15
6.3 Limitations of Capillary Electrophoresis for Nucleic Acids Analysis 15
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Capillary electrophoresis (CE) is a powerful and relatively new technique for nucleic acid mapping, and can be used along with conventional slab electrophoresis and high-performance liquid chromatography (HPLC). There are several types of separation techniques of CE, for instance capillary zone electrophoresis (CZE), micellar electrokinetic chromatography (MEKC), capillary gel electrophoresis (CGE), and capillary affinity gel electrophoresis (CAGE), all of which are capable of being used for quantification and qualification of nucleic acids. CZE and MEKC are suitable for the separation of bases, nucleoside and nucleotide monomers, and short nucleic acids. Although the sensitivity of ultraviolet (UV) detection is not as high as HPLC, since the light path length of CE is more than 100 times smaller than that of HPLC, a laser-excited fluorescence detection system enables highly sensitive detection of nucleic acids. CGE and CAGE, in which electrophoresis is performed in molecular sieving materials of polymer gels, are used rather than HPLC and slab gel electrophoresis (SGE) for sequencing of large nucleic acids for several reasons: high-resolution power, high-speed sequencing, ease of automation, and small sample volume. Combining the advantages of CGE and laser fluorescence detection with a four- or two-color labeling method enables sequencing of more than 1000 bases within 1 h. Moreover, recent improvements in capillary array electrophoresis (CAE) with high-throughput and high-speed sequencing systems contribute to an expansion of the sequence analysis of DNA fragments in order to achieve the goal of the Human Genome Project (HGP).

1 INTRODUCTION

CE is a powerful and relatively new technique with strong resolution and selectivity for the quantification and qualification of nucleoside bases, nucleosides, nucleotides and nucleic acids, DNA and RNA. Several studies are being carried out to sequence more than 3 billion bases of human DNA by the year 2005, for which cost-effective, high-throughput, and high-fidelity automated sequencing systems are required. CE is an important technique for the analysis of nucleic acids along with conventional electrophoresis and HPLC. CE is becoming popular in nucleic acid analysis because of its extremely high resolution, rapid analysis, ease of automation, and capability of sequence analysis.

There are different CE techniques,[1] which are described elsewhere in this encyclopedia, and they are classified mainly into two groups, that is CE using free solution and electrochromatography. In this article, several types of CE using free solution for nucleic acids are highlighted. These CE methods are based on the electrophoretic separation mechanisms such as CZE, capillary isotachophoresis (CITP), MEKC, CGE, and CAGE. Since these different types of CE provide different information on nucleic acids and their components, the separation mode should be properly chosen to obtain the best performance. The separation mode should be selected on the basis of what information is required about the nucleic acids and what information can be obtained by the selected CE method.

Analysis of relatively small molecules, that is bases, nucleosides, and nucleotide monomers and sometimes even relatively short oligonucleotides, which have

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

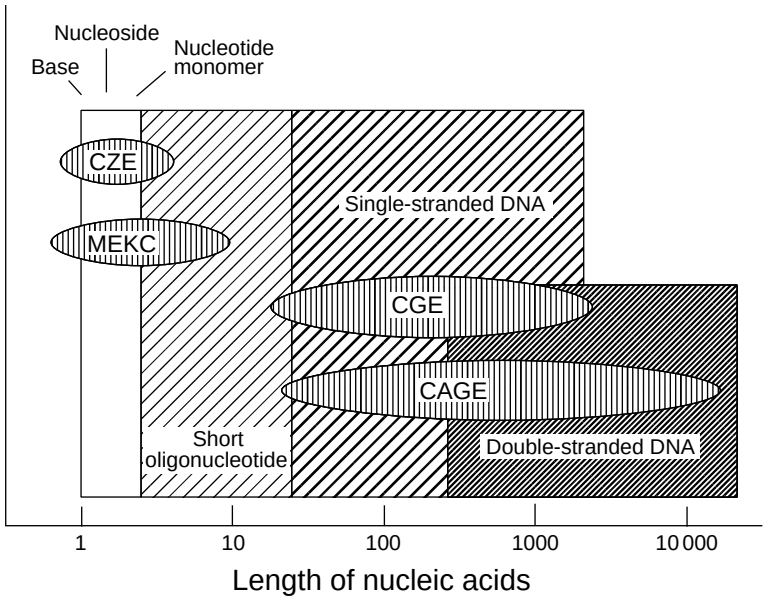

Figure 1 A guide to select the CE mode suitable for several sizes of nucleic acids. CZE, MEKC, CGE and CAGE cover characterization, size determination, and sequencing of nucleic acids of different sizes.

distinguishable mass/charge ratios (m/z), are best performed by CZE and MEKC. Figure 1 gives a rough guide of how to select the mode of CE for several sizes of nucleotides. The separation by CZE is based on electrophoretic mobility of analytes and electroosmotic flow (EOF), so that notable differences among the m/z of ionic solutes results in good resolution. Separation of bases and nucleosides is difficult or impossible using CZE since they have no charges on their moiety, which does not result in a difference of electrophoretic mobility. However, MEKC involves chromatographic phenomena of solutes which adsorb or partition to the pseudo-stationary micellar phase in the presence of surfactant. Thus, MEKC can be applied for the analysis of species with no charges and achieves better separation of bases, nucleosides, and also of mono- and short nucleotides.

CGE is powerful for the separation of large nucleic acid molecules that have a similar m/z and are not fundamentally suitable to be resolved by CZE and MEKC. Using CGE, electrophoresis is performed in molecular sieving materials of polymer gels such as cross-linked polyacrylamide gel (PAG). CGE allows the separation of large nucleic acids which have substantially the same m/z. Moreover, using CGE, sizing and sequencing of DNA are possible in a shorter time than SGE, since high electric fields can be applied.

CGE is a popular technique for genome mapping, which has been extensively investigated in the 1990s. Combining the advantages of CGE and laser fluorescence detection with the four- or two-color labeling methods allows the sequencing of more than 1000 bases within 1 h. In addition, CAE in which multiple CE separations are performed in parallel using a number of capillaries and scanned by a laser fluorescence detector is becoming a powerful automatic mapping technique of DNA.

2 HISTORY

The history of CE is briefly described from the viewpoint of the separation mode. A few early studies on isotachophoresis using a narrow-bore capillary were carried out to separate nucleotide monomers.[2,3] CZE using an open-tubular glass capillary was first carried out successfully in 1981 by Jorgenson et al.[4,5] This method is the physicochemical basis of applications of CE to nucleic acid analysis. The earliest successful application of CZE to nucleotides was in 1983.[6] In addition, a successful study for the analysis of short oligonucleotides using MEKC was accomplished in 1987.[7] There have been more than 50 papers published on the analysis of bases, nucleosides, and nucleotides using CZE and MEKC up to 1998, and the practical applications have been extensively investigated. For CZE and MEKC, several types of modification of electrophoretic eluents and additives have been investigated to achieve the best resolution of nucleic acids.

The first successful study of CGE was demonstrated by Cohen et al.[8] in 1988, in which 20 sequences were separated within 20 min using a cross-linked PAG-filled capillary column. In early studies, cross-linked PAG, which was popular for SGE, was used for CGE

in a fused-silica capillary of which the inner wall was coated. However, the gel-filling technique is often problematic in that bubbles form in the capillary tubing and cause unstable currents. To solve the problems, CGE using noncross-linked PAGs or entangled polymer solutions such as cellulose derivatives as molecular sieving materials have been developed instead of covalently bonded cross-linked PAGs.[9–11] CAGE is becoming a popular technique for genome mapping since the capillary is replaceable and yields sufficient resolution as high as conventional CGE.

In addition, detection methods have been extensively investigated. Since nucleotide bases have a UV absorption band at 250–270 nm, UV absorption detection is useful for the quantification of nucleic acids, but the sensitivity of UV detection is not high. Laser-excited fluorescence detection was developed and is as sensitive as radioisotopic detection.[12,13] Further, enhancement of fluorescence detection of DNA by intercalating dye was also developed for the analysis of double-stranded DNA (dsDNA), in which the fluorescence enhancer increases the sensitivity of DNA by a factor of over 1000.[14]

The sequence analysis of single-stranded DNA (ssDNA) using fluorescent dye labeling had been extensively developed on the basis of the sequencing technique for SGE.[13,15–18] Although CGE separations are rapid, the throughput is about the same as that of conventional sequencing by SGE if only one capillary can be run and detected at a time. The development of a high-speed and high-throughput DNA-sequencing technique was required for the success of the HGP,[19] and for this reason CAE was developed in 1992–1993.[20–23]

Most recently, there has been growing interest in the use of microfabricated devices filled with sieving media for electrophoretic separation of DNA.[24,25]

3 INSTRUMENTATION OF CAPILLARY ELECTROPHORESIS

3.1 Hardware

3.1.1 System

The hardware for CE is described elsewhere in this encyclopedia. A high-voltage power supply, sample injector, and detector are the essential components to run CE. In general, a conventional CE system is sufficient to run CZE, MEKC, and CGE. Normally, commercially available instruments cost approximately US$ 50 000–100 000. For DNA sequence analysis, a capillary array DNA sequencer is also commercially available.

3.1.2 Detector

Nucleic acids have UV absorption in the wavelength range 250–270 nm, which is due to absorption of nucleoside bases, so that quantification of bases, nucleosides, nucleotides, and size determination of nucleic acids can be performed using a UV detector. For sequence analysis by CGE, a laser-excited fluorescence labeling detection system is required.

3.1.3 Injection

Sample injection can be carried out by several methods. For CZE, MEKC, and sometimes for CAGE, samples are injected by siphoning, electroinjection, injection under vacuum or compression, etc. The electroinjection method involves a short period (a few seconds) of running the electrophoresis. This method is convenient and is sometimes the only possible method in cases where no other injection technique can be applied, such as for CGE. Mechanical injections under vacuum and/or compression injection methods are not suitable for CGE, since gel material is covalently bonded to fused-silica surfaces and these injection methods could destroy the gels in the capillary. Additionally, there is a problem with the electroinjection method in that it is not possible to know whether analytes are homogeneously introduced or not.[26]

3.1.4 Temperature Control

Temperature control is important to obtain the best reproducible data. The influence of temperature for nucleic acid analysis by CGE has been studied.[27]

3.1.5 Capillary

The bore size of the capillary used is frequently 25–100 μm inner diameter (ID) for CZE and MEKC and the coating of the inside of the capillary is not necessary for CZE, but is important for CGE. For CGE, a slightly wider bore capillary (50–100 μm ID) is advantageous to fill molecular sieving materials. The gel-filling techniques are described in a later section. Gel-filled capillaries are commercially available, although somewhat costly.

3.2 Sample Preparation

A general description of sample preparation for CE can be found elsewhere in this encyclopedia. The best performance of CE separation is achieved when analytes are resolved in the same buffer solution used as the running buffer. The concentration should be adjusted to 10^{-5}–10^{-3} M monomer units of nucleic acids (which corresponds to c. 10^{-3}–0.1 absorbance scale). For sequencing of DNA using CGA and CAGE with a laser-excited

fluorescence detector, the detection sensitivity is about 10^{-12} M.[20]

4 SEPARATION AND DETECTION METHODS

4.1 Separation Mode and Theory of Nucleic Acids Migration

4.1.1 Capillary Zone Electrophoresis

Nucleoside bases, nucleosides, and monomeric nucleotides, which are the components of nucleic acids, can be separated by CZE in free solution and MEKC.[28–30] In early studies of electrophoresis in narrow-bore capillaries, there had been a few investigations on analysis of nucleotide monomers by CITP,[2,3] and a small number of applications as a preconcentration method for CGE have been reported lately.[31,32] These methods are not as popular at present as CZE and MEKC.

In CZE, bases, nucleosides, nucleotide monomers and large nucleotides behave differently depending on the m/z. Basically, both electrophoretic flow (EPF) and EOF determine the mobility of analytes in CE. EOF towards the anode occurs if the inner surface of the capillary is negatively charged since an electrical double layer forms on the negatively charged fused-silica surface, which is a regular occurrence when fused-silica capillary tubing is used at neutral pH (Figure 2a). Here, the velocity (V_{os}) of EOF is expressed by Equation (1),[33]

$$V_{os} = -\frac{\varepsilon\zeta_1}{\eta} \quad E = \frac{l}{t_n} \tag{1}$$

where ε is the dielectric constant, ζ_1 is the zeta potential on the capillary wall, η is the viscosity of the liquid medium, E is the electric field strength, l is the effective length of capillary, and t_n is the migration time of a neutral species. Equation (1) indicates that V_{os} is positive since ζ_1 is negative for fused-silica capillary tubing at neutral pH. The EOF in fused-silica capillaries is fairly fast at neutral pH and dependent on the surface charge of the inner wall of the capillary (see general description on CE).[34]

Analytes migrate at their own electrophoretic speed, which is determined by the m/z. For example, the m/z

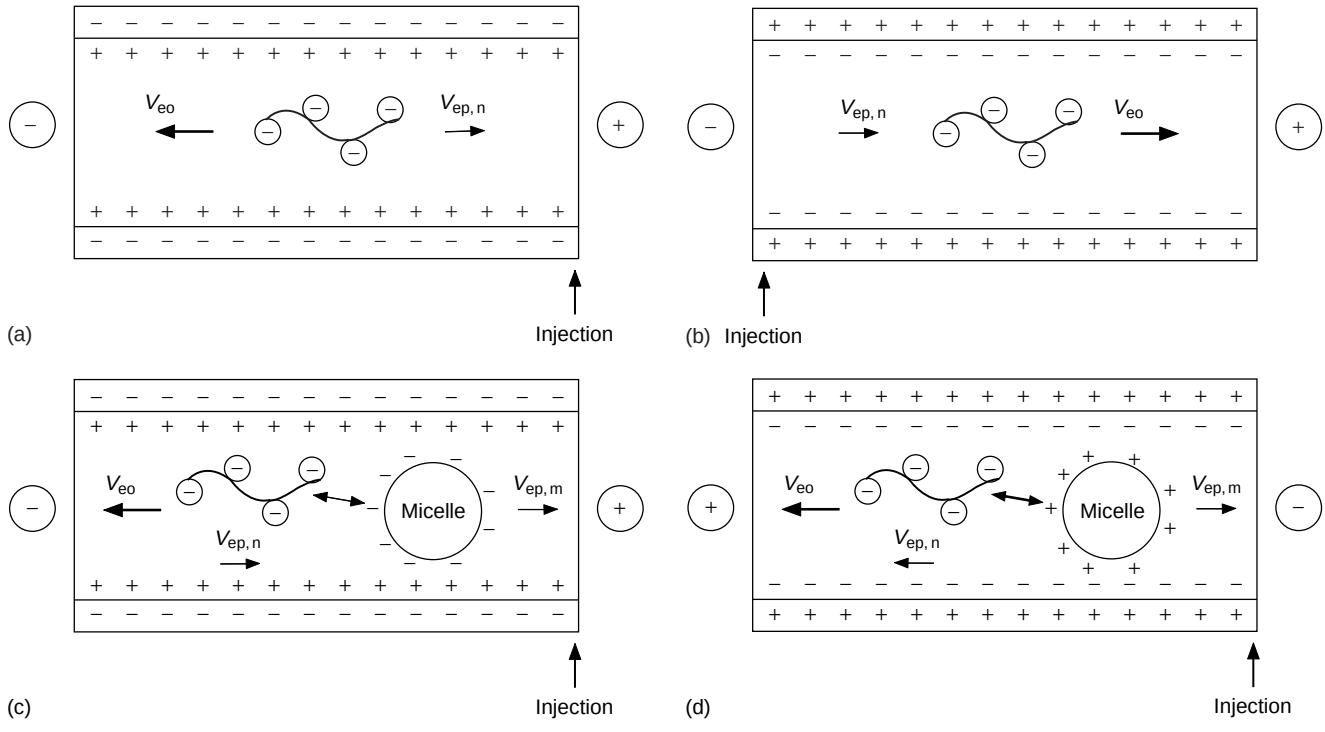

Figure 2 Model of migration of nucleic acids in CZE and MEKC. (a) CZE in negatively charged fused-silica surface in which EOF occurs towards the anode. The velocity of EOF is faster than the electrophoretic velocity of nucleic acids at neutral pH. Small nucleic acids migrate faster than large nucleic acids towards the anode. Samples are to be injected from the side of the cathode. (b) CZE in positively charged fused-silica surface in which EOF occurs towards the cathode. Large nucleic acids migrate faster than small nucleic acids. Samples are to be injected from the side of the anode. (c) MEKC in the presence of anionic surfactant in which EOF occurs towards the anode. Strong interaction between nucleic acids and micelles is not expected since both nucleic acids and micelles have negative charges. Samples are to be injected from the side of the cathode. (d) MEKC in the presence of cationic surfactant in which EOF occurs towards the cathode. Electrostatic interaction between nucleic acids and micelles is expected. Samples are to be injected from the side of the anode.

decreases in the order of 5′-ATP (adenosine triphosphate) >5′-ADP (adenosine diphosphate) >5′-AMP (adenosine monophosphate). The velocity of EPF is expressed in Equation (2),

$$V_{ep} = \frac{k_0 \varepsilon \zeta_2}{\eta} \, f(\kappa a) E \tag{2}$$

where k_0 is a constant, ζ_2 is the zeta potential of the solute, and $f(\kappa a)$ is the function depending on the particle shape. Thus, the mobility of the solutes is controlled by the sum of the electroosmotic and electrophoretic velocities. The apparent mobility of solutes (V_s) is expressed by Equation (3).[7]

$$V_s = V_{ep} + V_{os} \tag{3}$$

Thus, Equation (3) can be expressed by Equation (4) using Equations (1) and (2).

$$V_s = \left\{ -\zeta_1 + [k_0 \zeta_2 f(\kappa a)] \right\} \frac{\varepsilon E}{\eta} \tag{4}$$

Equation (4) indicates that V_s is proportional to E so that faster analysis is possible at higher electric fields. The analytical window between the solvent front and the last elution is narrow for CZE so that each peak appears very compact, and most small molecules are usually eluted in less than 30 min. If wider separation windows are used, short oligonucleotides can probably be separated.

The addition of cations, such as a cationic surfactant, slows the EOF since the electrical double layer on the fused-silica surface switches to the opposite polarity. Further, an excess of cations, such as metal ions and cationic surfactant, causes reverse flow of the EOF towards the cathode (reverse EOF) (Figure 2b). In CE using fused-silica capillary tubing, the EOF towards the anode (normal flow) increases with increasing pH.[34]

Factors affecting separation in CZE have been summarized in an early investigation.[4] Nucleic acid bases and nucleosides do not have charges at neutral pH, which presents problems for CZE since EOF is fairly fast, thus the separation of bases and nucleosides is difficult using CZE at neutral pH. However, for the same type of nucleotides having the same number of phosphate groups but different bases, CZE sometimes produces notable resolution, which results in sufficient separation.[35–38]

However, the separation of oligonucleotides by CZE is difficult for two reasons: first, the m/z becomes constant with increasing oligonucleotide length; and second, only a very narrow window size can be used since the EOF is fast when using CZE.

Since oligonucleotides have negative charges on the phosphate group, EPF occurs towards the cathode. A useful method is to slow the EOF, so that resolution of the oligonucleotides would be enhanced. Modifications of the electrophoretic running buffer, such as the addition of urea and a decrease in the pH, expand the CE window.[34,38,39]

4.1.2 Micellar Electrokinetic Chromatography

MEKC was established by Terabe et al.[40,41] in 1984. In MEKC, partitioning and adsorption to micelles, which are regarded as pseudo-stationary phases of liquid chromatography, act as an effective interaction, as well as EOF and EPF. Thus, even neutral species, that is bases and nucleosides, are potentially resolved by the chromatographic effect. There are two different MEKC methods for nucleotides that have different directions of EOF.

The separation model of nucleotides is shown in Figure 2(c), (d) for cases with EOF towards the anode and towards the cathode. In the presence of an anionic surfactant, such as sodium dodecyl sulfate (SDS), both the micelles and the nucleotides have negative charges, so that less interaction between micelles and nucleotides is expected (Figure 2c).[7,14,42–45] In the second case, a cationic surfactant such as dodecyltrimethylammonium bromide (DTAB) is used, and nucleotides interact with cationic micelles by electrostatic interaction (Figure 2d). In this case, EOF occurs in the reverse direction in the presence of an anionic surfactant. Thus, the samples have to be injected from the anode. In addition, nucleotides migrate, with partitioning of nucleotides onto the cationic micelles, which is notably effective since micelles and nucleotides have a different polarity, as well as regular EPF, towards the cathode. In these systems, the capacity factor k' defined by Equation (5) can be expressed by Equation (6),[7,41]

$$k' = \frac{n_{mc}}{n_{aq}} \tag{5}$$

$$= \frac{t_R - t_w}{[t_w(1 - t_R/t_{mc})]} \tag{6}$$

where n_{mc} is the total numbers of moles of solute in the micellar phase and n_{aq} is that in the aqueous phase, t_w is the retention time of the water molecule, t_{mc} is that of the micelle, and t_R is that of the analyte. Thus, k' can be written as Equation (7).

$$t_R = \left\{ \frac{1 + k'}{[1 + (t_w/t_{mc})k']} \right\} t_w \tag{7}$$

The resolution (R_s) between two solutes 1 and 2 is given by Equation (8),

$$R_s = \left(\frac{H^{0.5}}{4} \quad \left(\frac{\alpha - 1}{\alpha} \quad \frac{k'_2}{(1 + k'_2)} \right. \right. \times \left\{ \frac{(1 - t_w/t_{mc})}{[1 + (t_w/t_{mc})k'_1]} \right\} \tag{8}$$

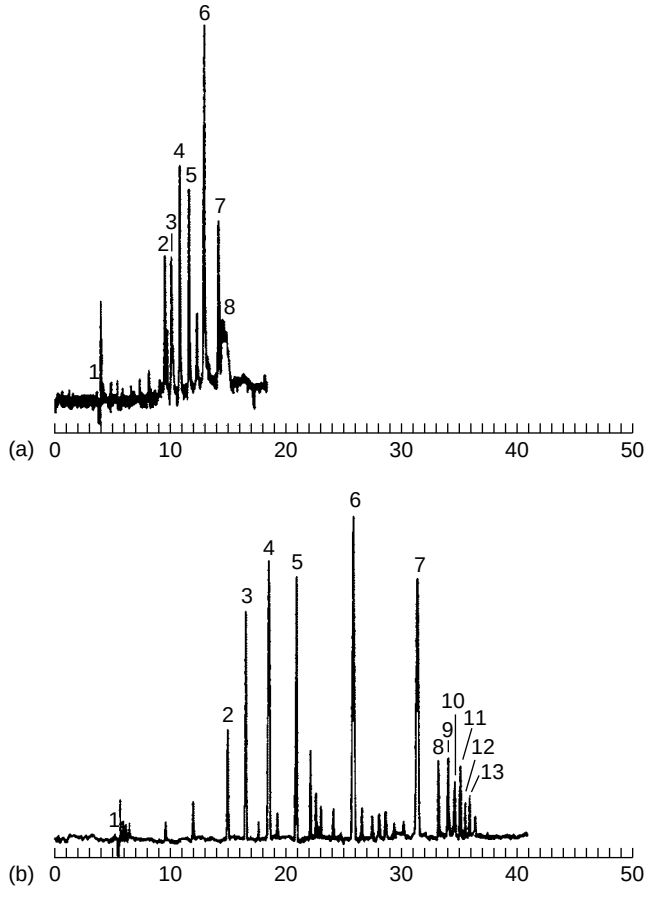

Figure 3 MEKC separation of oligothymidine mixture.[7] Electrophoretic conditions: (a) 5 mM Tris, 5 mM Na_2HPO_4, 7 M urea, pH 7, 50 mM SDS, 5 mM Mg(II); detection: 260 nm; capillary: 65 cm × 0.05 mm ID (effective length 45 cm), 20 kV, 25 °C. (b) Same as (a) except addition of 0.3 mM Cu(II) in the buffer. Oligothymidines: 1, solvent; 2, $d(pT)_2$; 3, $d(pT)_3$; 4, $d(pT)_4$; 5, $d(pT)_6$; 6, $d(pT)_{10}$; 7, $d(pT)_{12}$; 8, $d(pT)_{13}$; 9, $d(pT)_{14}$; 10, $d(pT)_{15}$; 11, $d(pT)_{16}$; 12, $d(pT)_{17}$; 13, $d(pT)_{18}$. [From Figure 5 on p. 1025 in the paper by A.S. Cohen, S. Terabe, J.A. Smith, B.L. Karger, *Anal. Chem.*, **59**(7), 1021–1027 (1987). Reproduced by permission from American Chemical Society.]

where H is the plate number and α is the separation factor (k_2'/k_1').

An example of resolution of oligonucleotides achieved by MEKC is shown in Figure 3. The addition of metal ions improves the resolution of oligothymidine, and the electrophoretic window is expanded with the addition of urea. Further, it has been demonstrated in a few cases that the addition of host–guest reagents improves the resolution of nucleotides only if there was originally a very small difference in CZE and MEKC. For example, nucleotide isomers 5′-AMP, 3′-AMP, and 2′-AMP, which have a small difference in their effective m/z, are separated by amplification of the difference in the presence of β-cyclodextrin.[37,46]

In summary, the benefit of using CZE and MEKC is that the analytical procedure is simple and preparation of the capillary is not required. Furthermore, CZE has extremely high resolution compared with HPLC.

4.1.3 Capillary Gel Electrophoresis

The separation by CGE is performed in a gel-filled capillary with a sieving matrix such as PAG. Although the m/z of DNA is considered to be independent of the size of oligonucleotide containing more than 10–15 bases, one chain length difference can be detected by the sieving effect at a speed of 1000 bases per hour.

The mechanism of CGE is schematically shown in Figure 4.[8,47,48] In capillaries filled with sieving polymer, small nucleic acids move faster through a sieving mesh than large nucleic acids do. The EOF is zero, or very low, in coated capillaries, so that the power supply polarity is the reverse of normal CZE. The gels are covalently bonded or not bonded on to the inner surface of fused-silica tubing.[49,50] The velocity of the nucleic acids through the sieving matrix is mainly dependent on the length of nucleic acid molecules as large nucleic acids migrate slower than small nucleic acids, but also it is influenced by their conformation in some cases.[51,52] The separation behavior is dependent on temperature,[27] buffer composition,[53] etc.[51,54,55]

On the basis of the Ogston model,[56,57] the relationship between the electrophoretic mobility of the solute (μ) and the gel concentration (T) is expressed by Equation (9),

$$\ln \mu = \ln \mu_0 - K_R T \tag{9}$$

where μ_0 is the migration time in free solution without gel and K_R is the retardation coefficient. The relationship can be applied to CAGE with entangled polymers and has also been experimentally established by means of a Ferguson plot.[58] When the size of the analyte is larger than the mesh size, the electrophoretic mobility

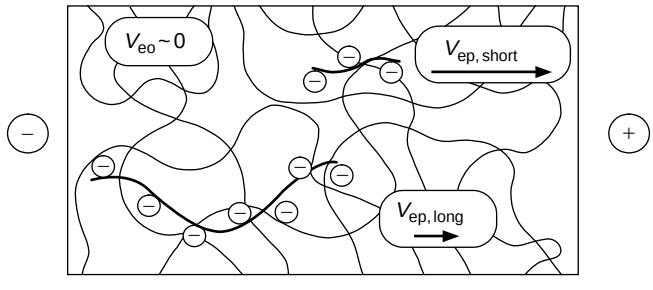

Figure 4 Model of migration behavior of nucleic acids in CGE such as PAGs. Large nucleic acids migrate slower than small nucleic acids. EOF is zero or less since the fused-silica surface is modified so that samples should be injected from the side of the anode.

is expressed by the reptation mechanism as shown in Equation(10),[59]

$$\mu \sim \frac{N}{N_2} = \frac{1}{N} \tag{10}$$

where N is the molecular size (base unit number of nucleic acids). A refinement on the model at large electric field was achieved by Lumpkin et al.[57] as expressed by Equation (11).

$$\mu \cong K\left(\frac{1}{N + bE^2}\right. \tag{11}$$

where K is a constant and b is a function of the mesh size of the polymer network. The relationship has been well established by several experiments.[10,11]

Although CGE using conventional PAG does not have the ability to resolve differences of sequence, specific base recognition of oligodeoxynucleotides is possible when using polyacrylamide poly(vinyl adenine)-conjugated gel[60–62] and other types of sequence-dependent polymer.[63]

4.1.3.1 Preparation of Gel-filled Capillary The preparation of gel-filled capillaries has been extensively investigated.[64–66] The following procedure is based on these studies.

First, a fused-silica surface may be treated since both a modified and a nonmodified fused-silica capillary can be used for preparation of gel-filled capillaries. Modification of a silica surface is carried out using uncross-linked PAG or by the formation of Si−C bonds, in which EOF is kept close to zero. Further, PAG modified on a silica surface is capable of being cross-linked with PAG that is to be used as the sieving matrix. In this case, the gel-filled capillary is, however, not replaceable.

In order to prepare an inner-surface-modified capillary, an appropriate length (50–100 μm ID) of capillary is rinsed with 1 M NaOH solution for about 15 min and with distilled water for 10 min. Acetic acid solution (6 mM) containing 3-methacryloxypropyltrimethoxysilane (0.4%) is put into the capillary for 10 min. The capillary is allowed to stand for 1 h and washed with distilled water for 10 min.

To polymerize acrylamide solutions, a stock solution, containing 40% acrylamide monomer and 0–5% *N*,*N*′-methylene-bis(acrylamide) as cross-linker, is diluted with buffer solution. The acrylamide solution is degassed thoroughly under vacuum, or by ultrasonic degassing, or degassing with inert gas bubbling. Ammonium peroxysulfate and *N*,*N*,*N*′,*N*′-tetramethylethylenediamine solutions are added into the acrylamide solution and flow into the capillary for 5 min. For gelation, the capillary is allowed to stand for a few hours overnight.

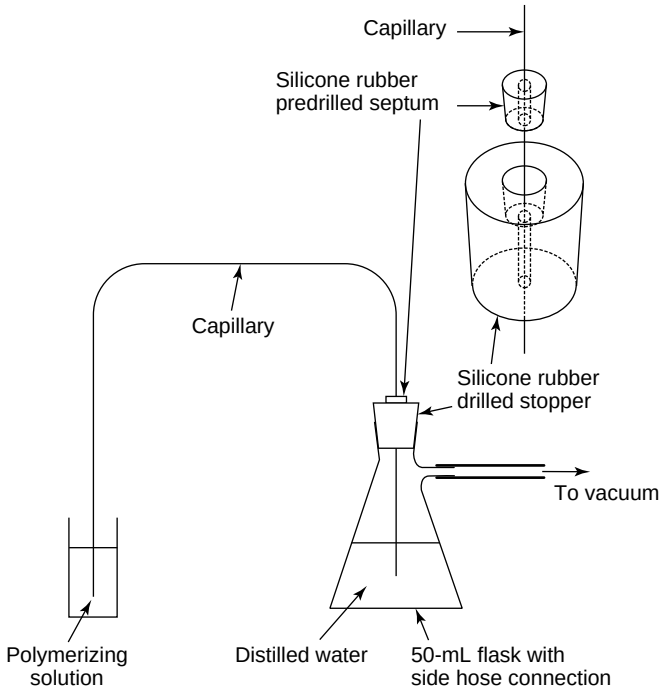

Figure 5 Equipment for preparation of PAG-filled capillaries in laboratories.[66] [From Figure 1 on p. 1222 in the paper by Y. Baba, T. Matsuura, K. Wakamoto, Y. Morita, Y. Nishitsu, M. Tsuhako, *Anal. Chem.*, **64**(11), 1221–1225 (1992). Reproduced by permission from American Chemical Society.]

The gel-filling technique requires careful operation and is shown in Figure 5.[66] The gel-filling procedure can be problematic because of bubble formation during gel polymerization. This can be avoided when the gel is polymerized at moderate concentrations and the polymerizing acrylamide solution is quickly introduced into capillaries.[65]

The sieving size of PAG is expressed by the gel concentration (%T) and the extent of cross-linking (%C). The definitions of %T and %C are given by Equations (12) and (13).[1,66]

$$\%T = \frac{(\text{mg acrylamide} + \text{mg cross-linker})}{(\text{mL buffer volume})} \times 0.01 \tag{12}$$

$$\%C = \frac{(\text{mg cross-linker})}{(\text{mg acrylamide} + \text{mg cross-linker})} \times 100 \tag{13}$$

%T defines the total amount of monomeric material in a given volume of gel, and %C the relative amount of cross-linker in the monomer content. Cross-linker concentrations of 2–7% are commonly used. The calibration of the capillaries can be carried out.[67] Effective ranges of separation of ssDNA and dsDNA at several %T are 50–1000 bases ssDNA and 100–10 000 dsDNA at 3%T,

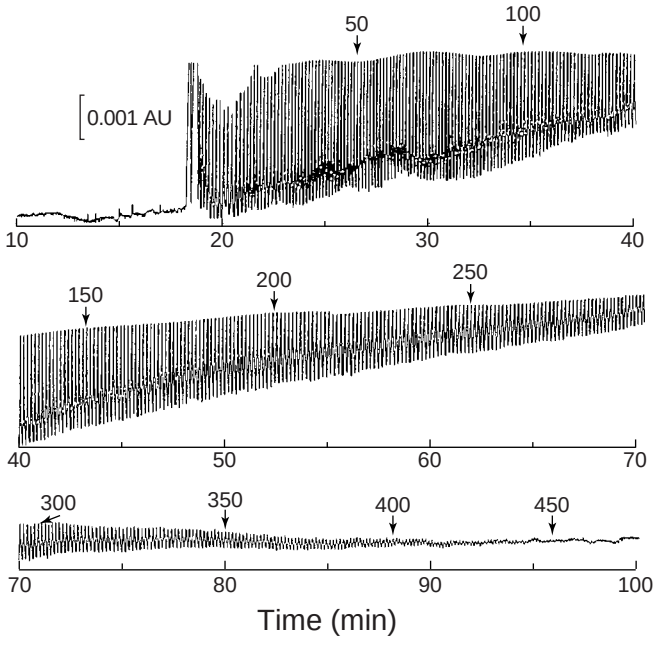

Figure 6 Size determination of polyadenylic acid digested by nuclease P1 by CGE using PAG.[66] Conditions: capillary 50 cm × 0.1 mm ID (effective length 30 cm); running buffer: 0.1 M Tris-borate, 7 M urea, pH 8.6; gel: 5%*T*, 1.5%*C*; 200 V cm^{-1}; injection: 5 kV × 1 s; detection: 260 nm. [From Figure 5 on p. 1224 in the paper by Y. Baba, T. Matsuura, K. Wakamoto, Y. Morita, Y. Nishitsu, M. Tsuhako, *Anal. Chem.*, **64**(11), 1221–1225 (1992). Reproduced by permission from American Chemical Society.]

20–500 ssDNA and 50–5000 dsDNA at 5%*T*, and 10–300 ssDNA and 10–1000 dsDNA at 8%*T*.[48]

The running buffer solution for ssDNA and dsDNA normally contains a high concentration of urea (6–8 M) or formamide (30% v/v), which is guaranteed to destroy hydrophobic interaction and then cause threadlike DNA. The electropherogram detected by UV absorbance is shown in Figure 6.[66]

4.1.4 Capillary Affinity Gel Electrophoresis

In spite of the high performance of CGE with cross-linked PAG, the routine preparation of a homogeneous gel-filled column is generally tedious and difficult owing to polymerization-induced shrinkage and the formation of bubbles inside the capillaries. This problem is solved by CAGE, in which entangled polymers are used as an alternative to cross-linked PAG.[59]

In 1989 Zhu et al.[9] first demonstrated that cross-linked PAG is not necessary for CGE. They showed that viscous solutions of hydrophobic polymers, such as cellulose derivatives, act as sieving matrices for DNA molecules. The separation mechanism of nucleotides in the sieving gel is basically the same as that of CGE, as shown in Figure 4. Theoretical studies on the electrophoretic behavior of nucleotides through sieving polymers with different field strength have been carried out.[59] There have been a number of studies on the conditions of CAGE[9,10,49,59] and the use of several types of sieving matrices, such as cellulose derivatives,[68–73] replaceable or noncross-linked linear polyacrylamide,[11,74–78] and N-substituted acrylamides,[79] poly(ethylene oxide) solutions,[80] poly(vinylpyrrolidone) matrix,[81] agarose gels,[82] glucomannan,[83] and poly(acryloylaminoethoxyethanol).[84] These polymer matrices are capable of being replaced by CGE for routine analysis. The most recent achievement is that sequence analysis can be run using a four-color laser fluorescence labeling method in which 1000 bases can be sequenced within 1 h.[76]

CAGE has additional advantages in the conformational analysis of nucleic acids, especially in the separation of double-stranded nucleic acids.[73,85–88]

4.2 Detection Methods

4.2.1 Ultraviolet Absorption Detection

While radioisotopic detection is popular in SGE, it is not easily applied for CE detection. For simple size determination of nucleic acids, UV detection is generally used due to its low cost and convenience. Nucleoside bases have absorption bands around 250–270 nm and their molar absorptivity (ε) is c. $1–1.5 \times 10^4$ mol^{-1} dm^3 cm^{-1}. Although quantification of nucleotides and size determinations are able to be performed by UV detection, the light path length (25–100 μm) of CE is more than 100 times smaller than that of HPLC (1 cm) so that the theoretical sensitivity of CE is 100–1000 times smaller than that of HPLC. For example, detection limits of concentration of nucleotide in CE, where ε is 1×10^4 mol^{-1} dm^{3}cm^{-1}, is 1×10^{-5} M when 0.001 absorbance unit is the detection limit in the UV detector, while the mole base detection limit[89] using the UV detector is generally reached at 10^{-18} mol of injected DNA.[86] Thus, diluted samples are not very efficiently detected, even though the absolute amount of detection limit is extremely low. This has been a significant problem in CE.

4.2.2 Fluorescence Detection

Laser-induced fluorescence detection is one of the solutions for improving low sensitivity in UV detection.[12,35,86,90,91] Single molecular detection was achieved for DNA fragments,[92] and detection of native fluorescence[93] was also effective for sensitive detection at neutral pH in CZE[12] and MEKC.[14] Indirect monitoring of nucleotide monomers[35] and ultrasensitive near-infrared (NIR) fluorescence detection in CGE[94] are also useful.

Although most CE analysis for dsDNA has relied on UV absorption for detection, detection by laser-excited fluorescence which is enhanced by addition of some intercalating dyes to dsDNA is also beneficial.[85,86,91,95] The fluorescence of thiazole orange homodimer (TOTO) **(1)** and oxazole yellow homodimer (YOYO) **(2)** is enhanced with intercalating to dsDNA by a factor of 1100 TOTO and 3200 YOYO to that of free TOTO and YOYO, respectively, in which the detection limits are c. 10^{-19} mol.

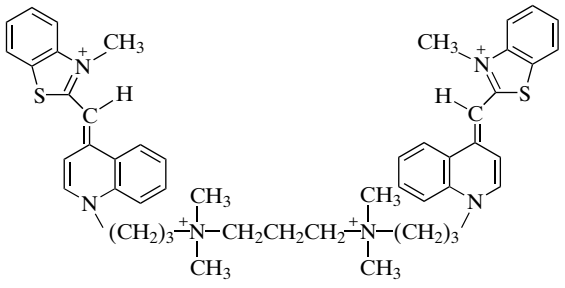

(1)

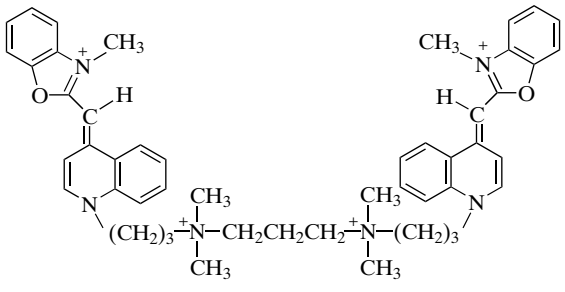

(2)

4.2.3 Sequence Analysis

A significant advance in sequencing technology occurred in 1986–87 with the development of Sanger's method of fluorescent labeling and laser detection. The basic techniques for sequencing were developed originally for SGE[90,96–104] and the application to CGE had succeeded by 1990–92.[13–16,105–107] Several types of laser detection systems, including a single spectral channel laser-induced system,[15] a single-lane with four-channel,[16] and a four- and two-color channel detection system[13] using CGE were developed.

The DNA sequencing procedure using four-color sequencing is described as follows. First, target DNA is incubated with primers in the presence of dideoxynucleotides, which are labeled by four types of fluorescence reagent corresponding to four bases, to polymerize the target DNA and terminate the polymerization with dideoxynucleotide.[108] A mixture of DNA terminated with fluorescently labeled dideoxynucelotides is obtained. Thus, the four-color fluorescence is detected to assign

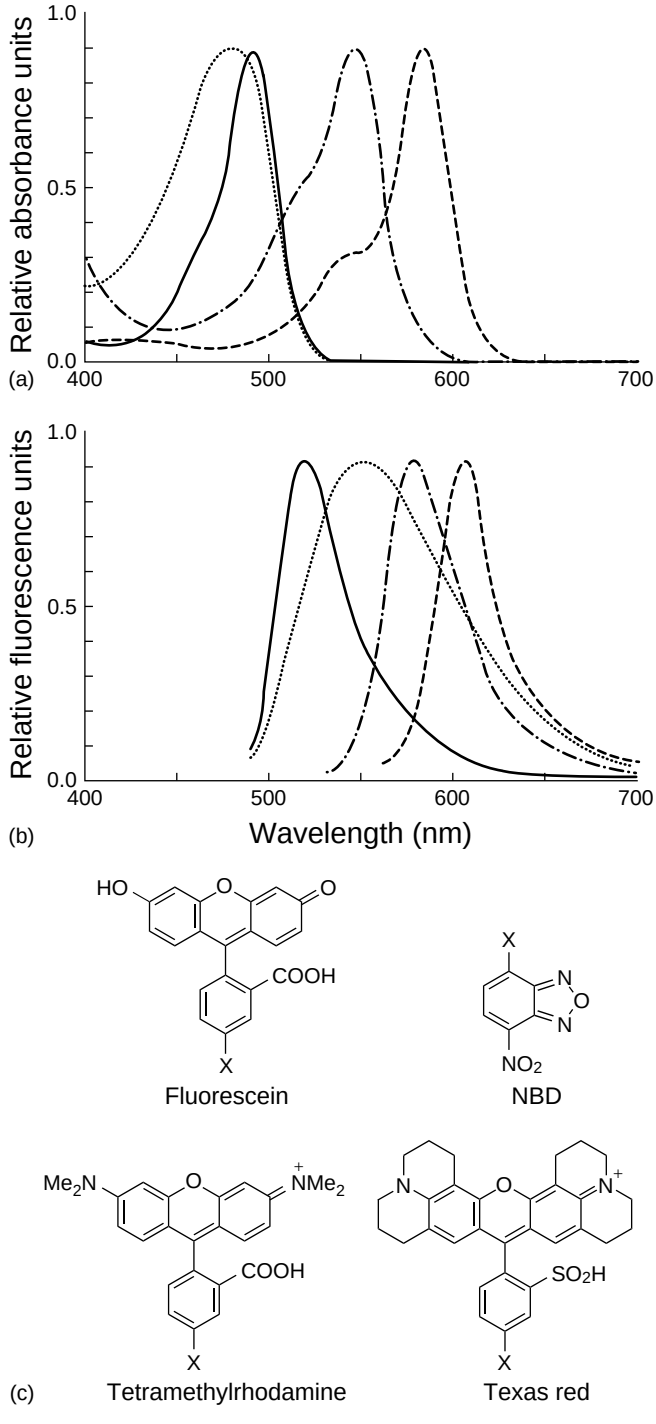

Figure 7 A set of four-color fluorescence labeling reagents for sequence analysis of DNA.[97] (a) Absorption spectra; (b) emission spectra. Spectra of fluorescent dyes of derivatives of fluorescein (——), NBD (· · ·), tetramethylrhodamine (–·–) and Texas red (- - -). X, Oligonucleotide primer to which the dye bound. (c) Structures. [From Figure 2 on p. 675 in the paper by L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B.H. Kent, L.E. Hood, *Nature*, **321**, 674–679 (1986). Reproduced by permission from Nature (Macmillan Magazines Ltd).]

the corresponding bases. Figure 7 shows the structure and fluorescence emission spectra of a set of four-color labeling dyes. Another fluorescence labeling system is also popular,[99] and more recently new types of fluorescence reagents with emissions close to infrared have been developed.[95]

An example of the four-spectral channel sequencing detection system is shown in Figure 8.[13] Two laser beams, 488 nm and 543.5 nm, are combined coaxially through a sector wheel which is rotated at 2 Hz. The fused-silica capillary tubing is placed at the beam focus and the fluorescence is collected and filtered by a filter wheel which is a synchronized sector wheel. The sector wheel is rotated at the same rate as the filter wheel to block the two laser beams alternately. Excitation of 488 nm causes emissions 540 and 560 nm, and excitation of 543.5 nm causes emissions 580 and 610 nm. The detection limits of this detector are 200 zmol for all four types of labeling reagents.

The sequencing technique can be applied to CAGE with noncross-linked polymers. Full automation of the process would be expected for sequencing of DNA since this methodology eliminates the gel-filling procedure. Based on the investigations of a number of sieving materials, a successful study has been demonstrated using replaceable linear polyacrylamide.[109] Sequencing of DNA using CAGE and the conventional labeling method[97] is shown in Figure 9 in which separation of 1000 bases is possible in less than 55 min at an accuracy greater than 99% with M13mp18 as template.[109]

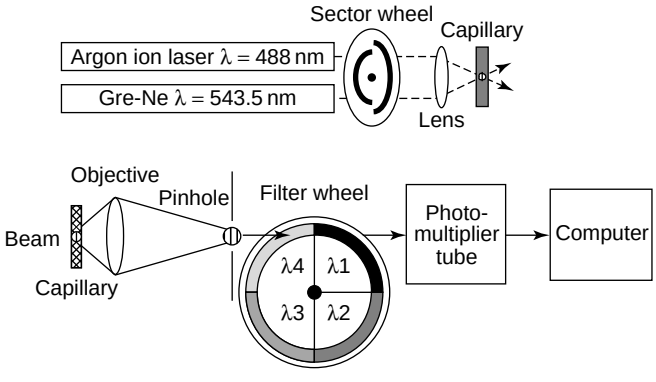

Figure 8 Equipment for laser-induced fluorescence detection for four-color sequencing.[13] Two laser beams are focused on a capillary and fluorescence is collected at right angles. A rotating filter wheel, synchronized with the sector wheel, is fitted to the emission bands of 540, 560, 580, and 610 nm of four dyes marketed by Applied Biosystems. The four dyes are derivatives of fluorescein (FAM), 2′,7′-dimethoxy-4′,5′-dichlorofluorescein (JOE), tetramethylrhodamine (TAMRA), and rhodamine X carboxylic or sulfonic acid (ROX). (Gre-Ne = green helium–neon laser.) [From Figure 1 on p. 2836 in the paper by H. Swerdlow, J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S. Wu, N.J. Dovichi, *Anal. Chem.*, **63**(24), 2835–2841 (1991). Reproduced by permission from American Chemical Society.]

4.2.4 Capillary Array Electrophoresis

Although CGE provides rapid, high-field, high-resolution separation with sequencing rate, overall throughput was about the same as that of automated slab-gel sequencing. Besides, SGE can easily run multiple separations. For the success of the HGP, both advantages of CGE and SGE are to be combined for high-speed and high-throughput DNA sequencing methods. To achieve this goal, CAE using a laser-excited fluorescence detection system was developed in 1992–93.[20–23] Figure 10 shows the first successful equipment of CAE in which a number of capillaries are scanned by laser detection. Instrumental improvement is still in progress[110–112] and CAE instruments are now commercially available.

4.2.5 Combination of Capillary Electrophoresis with other Techniques

A combination of CE with mass spectroscopy offers critical information of nucleic acids, such as conformation with adducts,[113,114] and has been applied for several types of CE such as CZE[115,116] and CAGE.[117]

Moreover, an enzymatic reactor is attachable to CE, for example, DNA can be cleaved to fragments by an on-line enzymatic reactor.[118,119]

5 PRACTICAL ANALYSIS AND APPLICATIONS

5.1 Chemical Structures

5.1.1 Application of Capillary Zone Electrophoresis and Micellar Electrokinetic Chromatography for Nucleic Acid Components

Applications of CZE and MEKC for bases and nucleosides, and charged species nucleotides and short oligonucleotides have been extensively investigated. As an example, separation of bases, nucleosides, and nucleotide monomers was achieved by MEKC with the addition of urea and metal additives, in which even oligonucleotides with 18 monomeric units were resolved.[7] Bases are also separable using several types of CZE methods[120,121] and applications of CZE to nucleosides in fish tissues[122] have been demonstrated. Determinations of monomeric nucleotides are feasible by CZE,[36,123] and by MEKC with anionic[14,45] and cationic surfactants.[43] Using CZE and MEKC, cyclodextrin as a host–guest reagent is useful for molecular recognition of monomeric and dimeric

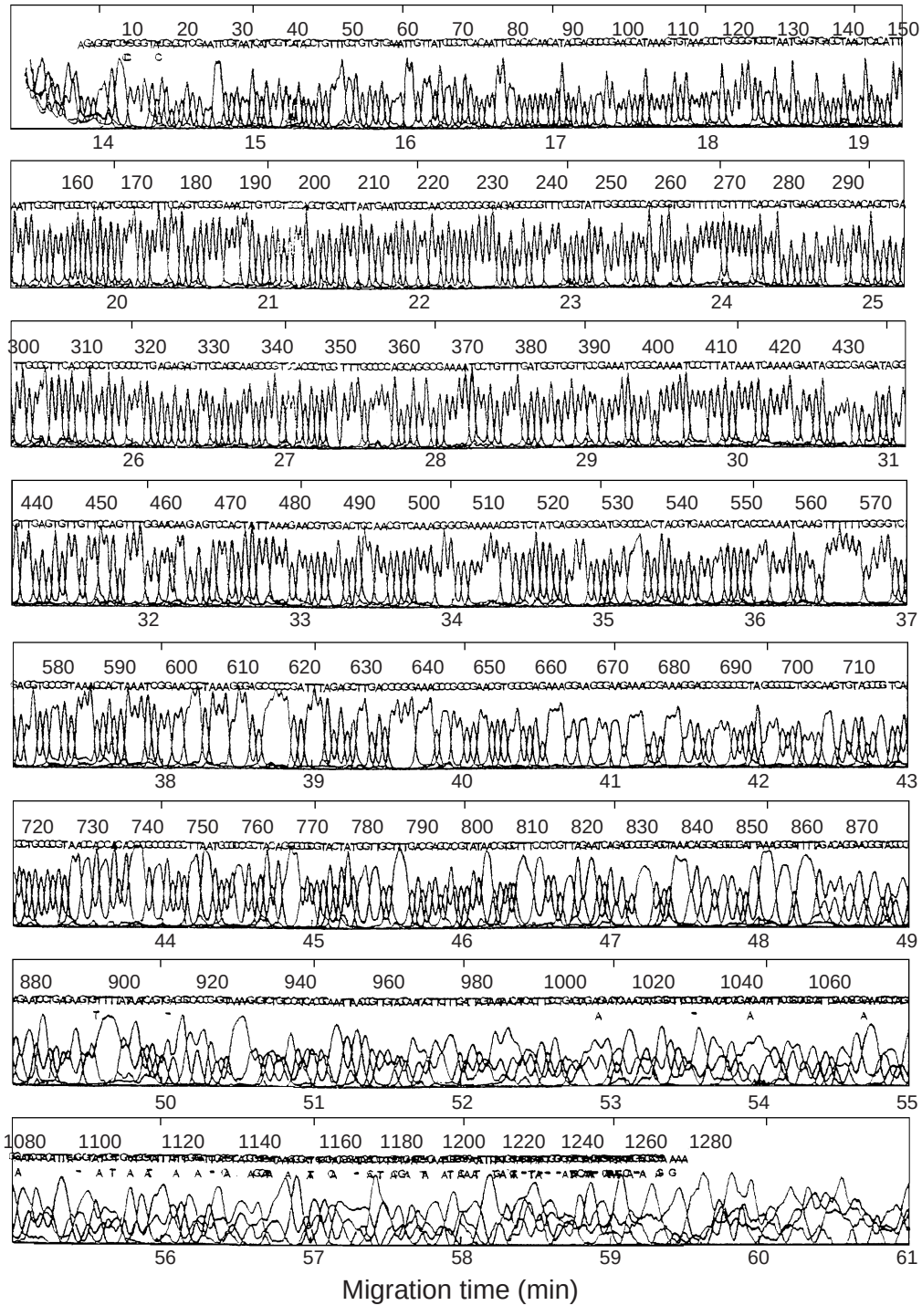

Figure 9 Electrophoretic sequencing by CAGE using replaceable linear PAG.[109] DNA sequencing fragments generated on M13mp18 at optimum conditions: 2% 9 MDa linear polyacrylamide mixture and 0.5% 50 kDa linear polyacrylamide mixture, 200 V cm^{-1}, 60 °C. Capillary, 45 cm × 0.75 mm ID, coated with poly(vinyl alcohol); running buffers 50 mM Tris, 50 mM TAPS, 2 mM EDTA, 7 M urea; 200 V cm^{-1} at 60 °C; injection: 25 V cm^{-1} × 10 s. [From Figure 3 on p. 4002 in the paper by O. Salas-Solano, E. Carrilho, L. Kotler, A.W. Miller, W. Goetzinger, Z. Sosic, B.L. Karger, *Anal. Chem.*, **70**(19), 3996–4003 (1998). Reproduced by permission from American Chemical Society.]

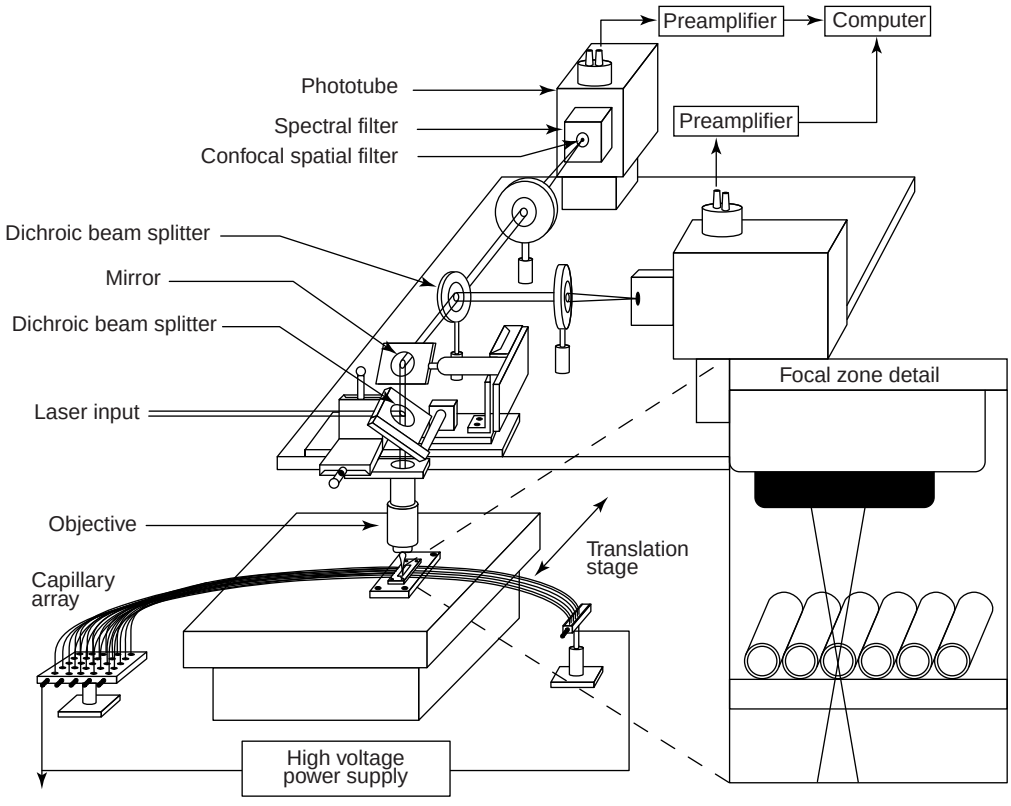

Figure 10 Schematic of the two-color, confocal-fluorescence capillary array scanner.[21] Excitation light from an argon laser phototube is brought to a 10-μm-diameter focus within the 100-μm-ID capillaries in the array. The fluorescence is passed back through the first beam splitter to a second beam splitter that separates the red and green channels. Sequencing is carried out using conventional two-color dye (FAM and JOE). [From Figure 1 on p. 2150 in the paper by X.C. Huang, M.A. Quesada, R.A. Mathies, *Anal. Chem.*, **64**(18), 2149–2154 (1992). Reproduced by permission from American Chemical Society.]

nucleotide isomers.[37,46] The separation of 3′,5′- and 2′,5′-linked adenylyladenosine monophosphate was achieved in the presence of β-cyclodextrin, which has a distinguishable interaction and results in a difference in electrophoretic mobility (Figure 11).[46] This method can be applied for the detection of both linked adenylyladenosine monophosphates in prebiotic products. Even though the separation of oligonucleotides is fundamentally difficult using CZE and MEKC, some trials have been carried out.[38,72]

The methods using CE and MEKC can be applied to analogs of nucleotides such as modified nucleotides.[42]

5.2 Size, Sequence and Conformation Analysis

5.2.1 Size Determination in Capillary Gel Electrophoresis

5.2.1.1 Single-stranded DNA Size determination by CE is best performed by CGE[49,63,105,124] and CAGE,[68] and the first size determination of nucleotides by CGE was demonstrated by Cohen et al.[8] in 1988. The aim was to analyze ssDNA and dsDNA so a comparison of resolution of ssDNA and dsDNA in CE was carried out.[70,87] ssDNA fragments can be separated and detected by UV[124] or laser fluorescence.[125] The separation in sequence analysis of DNA is directly based on the separation techniques of ssDNA. Thus, best resolution is obtained for large DNA with a single base unit.

5.2.1.2 Double-stranded DNA dsDNA has a wider range in molecular size than ssDNA so that the resolution of 10 base differences may be required. Highly sensitive detection methods for dsDNA were developed on the basis of the technique, which was improved for SGE.[95,126,127] A detection technique using intercalating dyes has been investigated using CAGE.[85,86] Figure 12 shows the dsDNA by CAGE using HPMC with detection using intercalating dyes.

5.2.1.3 RNA and Nucleotide Analogs The structure of RNA is almost the same as DNA except the hydroxyl group of the 2′-position of ribose in DNA is replaced by hydrogen. Thus, CE, MEKC, and CGE[66,128]

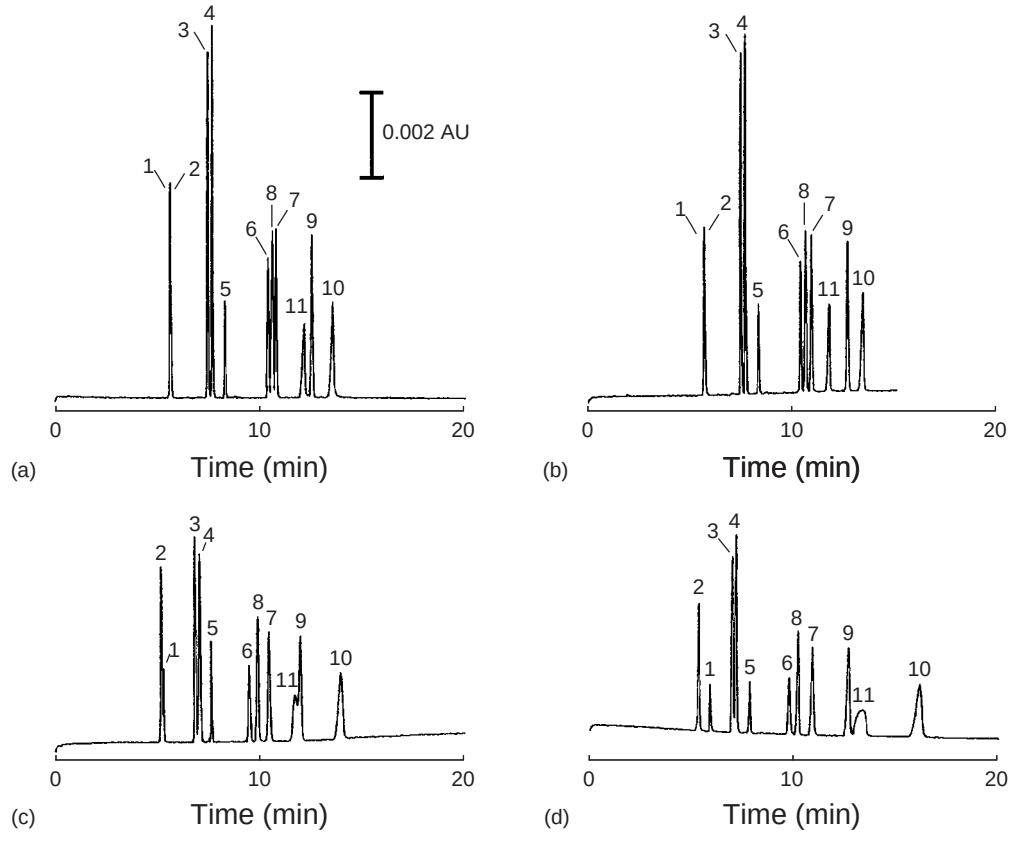

Figure 11 CZE separation of diadenosine monophosphate in the presence of β-cyclodextrin.[46] Operating conditions: 0.02 M Na_2HPO_4, 0.015 M β-cyclodextrin, 10^{-4} M $MgCl_2$, pH: (a) 7.0; (b) 7.5; (c) 8.0; (d) 9.0. Capillary: 61 cm × 0.05 mm (effective length 45 cm); 15 kV, 40 °C. Peaks: 1, adenine; 2, adenosine; 3, 3′,5′-adenylyladenosine monophosphate; 4, 2′,5′-adenylyladenosine monophosphate; 5, adenosine 3′,5′-cyclic monophosphate; 6, P_1,P_2-diadenosine 5′,5′-pyrophosphate; 7, adenosine 5′-monophosphate; 8, adenosine 3′-monophosphate; 9, adenosine 2′-monophosphate; 10, adenosine 5′-diphosphate, 11, adenosine 5′-triphosphate. [From Figure 6 on p. 175 in the paper by K. Kawamura, *J. Chromatogr., A*, **802**(1), 167–177 (1998). Reproduced by permission from Elsevier Science.]

methods are easily applied to RNA since the 2′-hydroxyl group is not important in determining the electrophoretic behavior. For example, application of CGE to transfer RNA and 5S ribosomal RNA has been demonstrated.[129]

DNA with linkage of phosphothioate is also able to be separated by similar methods to normal DNA. Successful applications of CGE and CAGE have been demonstrated,[130,131] and in particular sequencing of antisense DNA analogs is also possible by CGE.

5.2.2 Sequence Analysis by Capillary Gel Electrophoresis

Based on the fundamental studies and further improvement of the instruments, a number of application studies for sequencing nucleic acids have been carried out with electrophoretic conditions such as electric fields,[132,133] concentration of polyacrylamide,[75,133] etc. Further use of replaceable sieving matrices such as poly(ethylene oxide),[80] noncross-linked polyacrylamide,[76] poly(vinylpyrrolidone),[81] and others[69,134,135] is successful for sequencing.

Detection systems have also been investigated to achieve high sensitivity, reproducibility, high speed, and low cost.[136] NIR fluorescence,[94,137] labeling dyes,[138] and electrophore mass tag dideoxy DNA sequencing has also been developed.[139]

5.2.3 Capillary Array Electrophoresis

To detect a large number of parallel capillaries, several investigations, such as high-speed separation,[140] monitoring of 100 capillaries,[111] multiple sheath-flow method,[110] energy transferable primer,[112] multiple laser focusing,[141] and microfabrication,[142] have been carried out. Recently, size analysis of dsDNA using intercalating dyes was also carried out using CAE.[143]

Automation of DNA sequencing is also important for the HGP and has also been developed.[144]

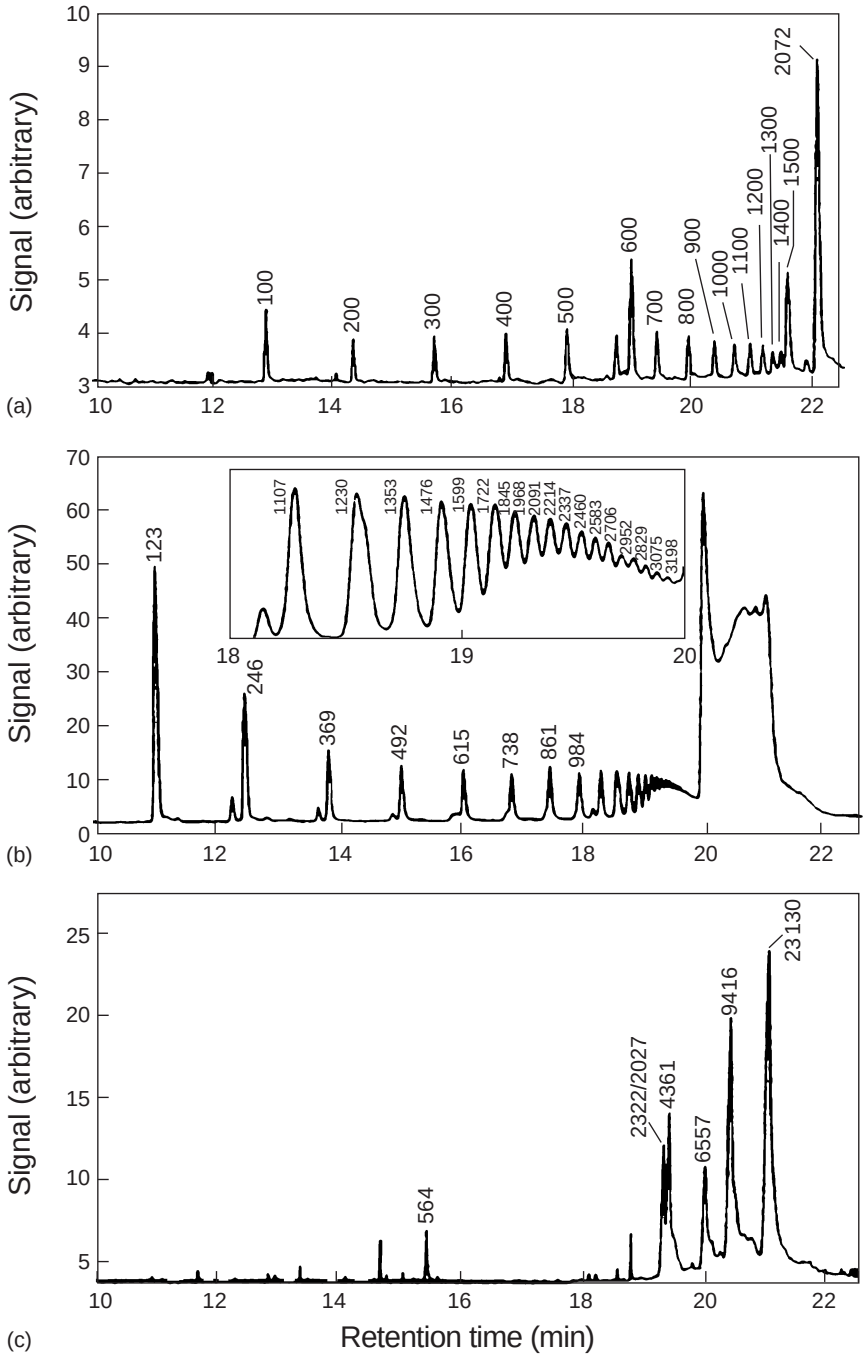

Figure 12 Separation of large dsDNA by CAGE using hydroxypropylmethyl cellulose (HPMC).[86] Operating conditions: HPMC 0.4%, 0.1 M NaCl, 0.54 g Tris, 0.275 boric acid, 0.1 mmol EDTA in 50 mL buffer; 10 or 32 μm ID capillary; 200 V cm^{-1}. (a) 100 base pair ladder, (b) 123 base pair ladder, (c) λ DNA Hind III intercalated with YOYO-3 (a benzoxazolium-4-quinolinium dimer with three carbon atoms bridging the aromatic rings of the unsymmetrical cyanines). [From Figure 8 on p. 214 in the paper by D. Figeys, E. Arriaga, A. Renborg, N.J. Dovichi, *J. Chromatogr., A*, **669**(1–2), 205–216 (1994). Reproduced by permission from Elsevier Science.]

It has been investigated for several types of rapid scanning methods of capillaries,[145–147] and a detection system using a photodiode array,[148] multipoint detection,[149] and an injection system of samples.[26] Moreover, miniaturization of the system is also important. An electrophoretic system on silicon devices[24,150] and a miniaturized injection method[151] are also interesting.

5.3 Applications and Needs

5.3.1 Medical

Since CE provides extremely high resolution and rapid analysis, its practical applications have been investigated widely for the HGP, for polymerase chain reaction (PCR) products and restriction fragment analysis. Several modes of CE, namely, CZE, MEKC, and CGE, have useful medical applications, such as AIDS virus (HIV-1) DNA in blood,[152] antisense DNA using microdevices,[153] nucleosides in urine,[154] clinical diagnostic DNA by CZE,[155] cancer diagnosis,[156] ssDNA in blood plasma,[157] DNA restriction fragments in cystic fibrosis and Gaucher's disease.[158]

5.3.2 Genetic

Many applications of CE in genetics are possible, such as diagnosis of the detection of point mutation in DNA,[159,160] gene diagnosis,[161] restriction mapping,[162] rat mRNA for gastric H^+,K^+-ATPase,[163] plasmid DNA,[164] and apolipoprotein E genotypes.[165] ssDNA analysis is beneficial for polymorphism analysis,[125] and DNA restriction fragments by CAGE,[166–170] and of the AIDS (HIV-1) virus in blood.[171] PCR products are able to be analyzed by conventional CGE and CAGE,[172] and by using a microfabricated device.[173]

6 LIMITATION AND COMPARISON WITH OTHER METHODS

6.1 Comparison with High-performance Liquid Chromatography

CE, HPLC, and conventional SGE are important techniques for nucleotide analysis and mapping. Comparative studies using several modes of CE and HPLC have been carried out for size determination of large fragments of DNA, and a systematic comparison has also been made using CE and HPLC.[48,128,174,175]

For the quantification of nucleotide monomers and short oligonucleotides, HPLC is superior to CE in terms of speed and selectivity.[176] MEKC is similar to conventional HPLC;[177] for simple analysis of nucleotides MEKC has proven to be superior to HPLC in terms of speed, small sample requirement, selectivity, and sensitivity, in which MEKC is a great time-saving method. Reproducibility of HPLC (1–2%) is somewhat better than CE (2–4%) for the routine analysis of oligonucleotides.[128]

The resolution power of CGE, which is much superior than that of HPLC, is significant for the analysis of large size DNA fragments. The theoretical plate number reaches over 10 million per meter, which is more than 10 times higher than that of HPLC.[16,47,66,109,128] CGE has a more rapid separation ability for DNA than for HPLC; for example, CGE was shown to be about four times faster than HPLC for the separation of the same number of DNAs.[174] Furthermore, the fragmentation of large DNA fragments by HPLC, CGE, and CAGE is equally sufficient for the separation of relatively short nucleotides up to 1000 bases. However, CGE and CAGE are superior to HPLC for the separation of larger DNA fragments.[131,178,179]

The absolute amount at detection limits of nucleic acids by CE using UV absorption is much smaller than that by HPLC, whereas concentration sensitivity by UV detection of HPLC is about 100 times higher than that of CGE. If one can use laser fluorescence detection, the detection limit by CE is much smaller than that by HPLC using UV.

6.2 Comparison with Slab Gel Electrophoresis

CE separations are rapid because of the high electric fields that can be applied, but the throughput is about the same as that of SGE. CAE solved this problem and is almost a prerequisite for the success of the HGP.[18] Further, CE is superior to SGE on the resolution power, on speed, being easy to fully automate, no preparation of gel, size determination and sequencing of nucleic acids. Comparison of conventional SGE and CGE has been carried out.[155,180]

6.3 Limitations of Capillary Electrophoresis for Nucleic Acids Analysis

Fragmentation by CE is rapid at a sizing speed of 1000 bases in less than 1 h and CAE is becoming a fully automated technique for DNA sequencing. These techniques are excellent and seem to have great capability for nucleotide mapping, but they have been primarily developed for the goal of the HGP. Therefore, CE approaches would potentially provide useful information in other fields which are not related to the HGP. For example, analysis of relatively small oligonucleotides having 10–20 sequences is not easy using either CGE or CZE. For CGE, highly concentrated and extraordinary cross-linked polymers are required to separate short oligonucleotides. Needless to say, CAE is available as a fully automated instrument, but it is too costly to run only for analyzing short nucleotides. Also, there are less practical methods of CZE and MEKC for the analysis of relatively short oligonucleotides, which are regarded as too large for CZE. As another example, analysis of nucleotides involving unusual linkages, such as pyrophosphate and 2′,5′-linked RNA, have been less studied by both CE and HPLC so that CE would be profitable for these analyses.[46]

Another limitation of CE is that more skill is required to use CE than HPLC and SGE. This is because CE involves techniques that require care to keep high reproducibility, for example pretreatement of a capillary is often required for each analysis and a small amount of impurity sometimes significantly affects electrophoretic behavior in CZE. These problems may be solved by full automation and by using commercially available laboratory kits for CE analysis.

ACKNOWLEDGMENTS

Permissions to reproduce figures were generously provided by the American Chemical Society, *Nature*, and Elsevier Science. The Library and Information Science Center in Osaka Prefecture University helped to prepare the manuscript.

ABBREVIATIONS AND ACRONYMS

ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
CAE	Capillary Array Electrophoresis
CAGE	Capillary Affinity Gel Electrophoresis
CE	Capillary Electrophoresis
CGE	Capillary Gel Electrophoresis
CITP	Capillary Isotachophoresis
CZE	Capillary Zone Electrophoresis
dsDNA	Double-stranded DNA
DTAB	Dodecyltrimethylammonium Bromide
EOF	Electroosmotic Flow
EPF	Electrophoretic Flow
HGP	Human Genome Project
HPLC	High-performance Liquid Chromatography
HPMC	Hydroxypropylmethyl Cellulose
ID	Inner Diameter
MEKC	Micellar Electrokinetic Chromatography
NIR	Near-infrared
PAG	Polyacrylamide Gel
PCR	Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
SGE	Slab Gel Electrophoresis
ssDNA	Single-stranded DNA
TOTO	Thiazole Orange Homodimer
UV	Ultraviolet
YOYO	Oxazole Yellow Homodimer

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction

Clinical Chemistry **(Volume 2)**
Nucleic Acid Analysis in Clinical Chemistry

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Mass Spectrometry of Nucleic Acids • Nuclear Magnetic Resonance and Nucleic Acid Structures • Nucleic Acid Structural Energetics • Polymerase Chain Reaction and Other Amplification Systems • Restriction Landmark Genomic and cDNA Scanning • Sequencing Strategies and Tactics in DNA and RNA Analysis

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Protein–Oligonucleotide Interactions

Liquid Chromatography **(Volume 13)**
Biopolymer Chromatography • Capillary Electrophoresis • Micellar Electrokinetic Chromatography

REFERENCES

1. G.J.M. Bruin, A. Paulus, 'Biopolymer Separations with Capillary Electrophoresis', *Anal. Meth. Instrum.*, **2**(1), 3–26 (1995).
2. I. Nukatsuka, H. Yoshida, 'Capillary Tube Isotachophoretic Separation of Nucleotides Using Complex-forming Equilibria', *J. Chromatogr.*, **237**(3), 506–510 (1982).
3. M. Goto, K. Irino, D. Ishii, 'Multichannel Spectrophotometric Detector for Fused-silica Capillary Tube Isotachophoresis', *J. Chromatogr.*, **346**, 167–176 (1985).
4. J.W. Jorgenson, K.D. Lukacs, 'Zone Electrophoresis in Open-tubular Glass Capillaries', *Anal. Chem.*, **53**(8), 1298–1302 (1981).
5. J.W. Jorgenson, K.D. Lukacs, 'Capillary Zone Electrophoresis', *Science*, **222**, 266–272 (1983).
6. T. Tsuda, G. Nakagawa, M. Sato, K. Yagi, 'Separation of Nucelotides by High-voltage Capillary Electrophoresis', *J. Appl. Biochem.*, **5**(4–5), 330–336 (1983).
7. A.S. Cohen, S. Terabe, J.A. Smith, B.L. Karger, 'High-performance Capillary Electrophoretic Separation of Bases, Nucleosides, and Oligonucleotides: Retention Manipulation via Micellar Solutions and Metal Additives', *Anal. Chem.*, **59**(7), 1021–1027 (1987).

8. A.S. Cohen, D.R. Najarian, A. Paulus, A. Guttman, J.A. Smith, B.L. Karger, 'Rapid Separation and Purification of Oligonucleotides by High-performance Capillary Gel Electrophoresis', *Proc. Natl. Acad. Sci. USA*, **85**(24), 9660–9663 (1988).
9. M. Zhu, D.L. Hansen, S. Burd, F. Gannon, 'Factors Affecting Free-zone Electrophoresis and Isoelectric Focusing in Capillary Electrophoresis', *J. Chromatogr.*, **480**, 311–319 (1989).
10. P.D. Grossman, D.S. Soane, 'Capillary Electrophoresis of DNA in Entangled Polymer Solutions', *J. Chromatogr.*, **559**(1–2), 257–266 (1991).
11. A.E. Barron, D.S. Soane, H.W. Blanch, 'Capillary Electrophoresis of DNA in Uncross-linked Polymer Solutions', *J. Chromatogr.*, **652**(1), 3–16 (1993).
12. Y.-F. Cheng, N. Dovichi, 'Subattomole Amino Acid Analysis by Capillary Zone Electrophoresis and Laser-induced Fluorescence', *Science*, **242**, 562–564 (1988).
13. H. Swerdlow, J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S. Wu, N.J. Dovichi, C. Fuller, 'Three DNA Sequencing Methods Using Capillary Gel Electrophoresis and Laser-induced Fluorescence', *Anal. Chem.*, **63**(24), 2835–2841 (1991).
14. T. Lee, E.S. Yeung, M. Sharma, 'Micellar Electrokinetic Capillary Chromatographic Separation and Laser-induced Fluorescence Detection of 2′-Deoxynucleoside 5′-Monophosphates of Normal and Modified Bases', *J. Chromatogr., Biomed. Appl.*, **565**(1–2), 197–206 (1991).
15. H. Swerdlow, R. Gesteland, 'Capillary Gel Electrophoresis for Rapid, High Resolution DNA Sequencing', *Nucleic Acids Res.*, **18**(6), 1415–1419 (1990).
16. J.A. Luckey, H. Drossman, A.J. Kostichka, D.A. Mead, J. D'Cunha, T.B. Norris, L.M. Smith, 'High Speed DNA Sequencing by Capillary Electrophoresis', *Nucleic Acids Res.*, **18**(15), 4417–4421 (1990).
17. H. Swerdlow, S. Wu, H. Harke, N.J. Dovichi, 'Capillary Gel Electrophoresis for DNA Sequencing. Laser-induced Fluorescence Detection with the Sheath Flow Cuvette', *J. Chromatogr.*, **516**, 61–67 (1990).
18. H. Drossman, J.A. Luckey, A.J. Kostichka, J. D'Cunha, L.M. Smith, 'High-speed Separation of DNA Sequencing Reactions by Capillary Electrophoresis', *Anal. Chem.*, **62**(9), 900–903 (1990).
19. T. Hunkapiller, R.J. Kaiser, B.F. Koop, L. Hood, 'Large-scale and Automated DNA Sequence Determination', *Science*, **254**, 59–67 (1991).
20. X.C. Huang, M.A. Quesada, R.A. Mathies, 'Capillary Array Electrophoresis Using Laser-excited Confocal Fluorescence Detection', *Anal. Chem.*, **64**(8), 967–972 (1992).
21. X.C. Huang, M.A. Quesada, R.A. Mathies, 'DNA Sequencing Using Capillary Array Electrophoresis', *Anal. Chem.*, **64**(18), 2149–2154 (1992).
22. R.A. Mathies, X.C. Huang, 'Capillary Array Electrophoresis: an Approach to High-speed, High-throughput DNA Sequencing', *Nature*, **359**, 167–169 (1992).
23. H. Kambara, S. Takahashi, 'Multiple-sheathflow Capillary Array DNA Analyser', *Nature*, **361**, 565–566 (1993).
24. C.S. Effenhauser, G.J.M. Bruin, A. Paulus, M. Ehrat, 'Integrated Capillary Electrophoresis on Flexible Silicone Microdevices: Analysis of DNA Restriction Fragments and Detection of Single DNA Molecules on Microchips', *Anal. Chem.*, **69**(17), 3451–3457 (1997).
25. D. Schmaizing, A. Adourian, L. Koutny, L. Ziaugra, P. Matsudaira, D. Ehrlich, 'DNA Sequencing on Microfabricated Electrophoretic Devices', *Anal. Chem.*, **70**(11), 2303–2310 (1998).
26. A. Guttman, H.E. Schwartz, 'Artifacts Related to Sample Introduction in Capillary Gel Electrophoresis Affecting Separation Performance and Quantitation', *Anal. Chem.*, **67**(13), 2279–2283 (1995).
27. A. Guttman, N. Cooke, 'Effect of Temperature on the Separation of DNA Restriction Fragments in Capillary Gel Electrophoresis', *J. Chromatogr.*, **559**(1–2), 285–294 (1991).
28. Z. Deyl, R. Struzinsky, 'Capillary Zone Electrophoresis: its Applicability and Potential in Biochemical Analysis', *J. Chromatogr., Biomed. Appl.*, **569**(1–2), 63–122 (1991).
29. P.G. Righetti, C. Gelfi, M. Perego, A.V. Stoyanov, A. Bossi, 'Capillary Zone Electrophoresis of Oligonucleotides and Peptides in Isoelectric Buffers. Theory and Methodology', *Electrophoresis*, **18**(12–13), 2145–2153 (1997).
30. P.G. Righetti, C. Gelfi, 'Recent Advances in Capillary Zone Electrophoresis of DNA', *Forensic Sci. Int.*, **92**(2–3), 239–250 (1998).
31. P.A. Walker, III, W.K. Kowalchyk, M.D. Morris, 'On-line Raman Spectroscopy of Ribonucleotides Preconcentrated by Capillary Isotachophoresis', *Anal. Chem.*, **67**(23), 4255–4260 (1995).
32. S. Auriola, I. Jaaskelainen, M. Regina, A. Urtti, 'Analysis of Oligonucleotides by On-column Transient Capillary Isotachophoresis and Capillary Electrophoresis in Poly(ethylene glycol)-filled Columns', *Anal. Chem.*, **68**(22), 3907–3911 (1996).
33. C.L. Rice, R. Whitehead, 'Electrokinetic Flow in a Narrow Cylindrical Capillary', *J. Phys. Chem.*, **69**(11), 4017–4024 (1965).
34. K.D. Lukacs, J.W. Joegenson, 'Capillary Zone Electrophoresis: Effect of Physical Parameters on Separation Efficiency and Quantitation', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **8**(8), 407–411 (1985).
35. L. Gross, E.S. Yeung, 'Indirect Fluorimetric Detection and Quantification in Capillary Zone Electrophoresis of Inorganic Anions and Nucleotides', *J. Chromatogr.*, **480**, 169–178 (1989).
36. R. Takigiku, R.E. Schneider, 'Reproducibility and Quantitation of Separation for Ribonucleoside Triphosphates and Deoxyribonucleoside Triphosphates by Capillary Zone Electrophoresis', *J. Chromatogr.*, **559**(1–2), 247–256 (1991).

37. T. Tadey, W.C. Purdy, 'Capillary Electrophoretic Separation of Nucleotide Isomer via Complexation with Cyclodextrin and Borate', *J. Chromatogr. B: Biomed. Appl.*, **657**(2), 365–372 (1994).
38. G. De Bellis, G. Salani, 'Oligonucleotide Analysis by Capillary Zone Electrophoresis at Low pH', *Anal. Chim. Acta*, **345**(1–3), 1–4 (1997).
39. J.T. Smith, Z. El Rassi, 'Capillary Zone Electrophoresis of Biological Substances with Surface-modified Fused-silica Capillaries with Switchable Electro-osmotic Flow', *J. High Resolut. Chromatogr.*, **15**(9), 573–578 (1992).
40. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Chromatography with Micellar Solutions and Open-tubular Capillary', *Anal. Chem.*, **56**(1), 111–113 (1984).
41. S. Terabe, K. Otsuka, T. Ando, 'Electrokinetic Chromatography with Micellar Solution and Open-tubular Capillary', *Anal. Chem.*, **57**(4), 834–841 (1985).
42. K.H. Row, W.H. Griest, M.P. Maskarinec, 'Separation of Modified Nucleic Acid Constituents by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr.*, **409**, 193–203 (1987).
43. J. Liu, J.F. Banks, Jr, M. Novotny, 'High-speed Micellar Electrokinetic Capillary Chromatography of the Common Phosphorylated Nucleosides', *J. Microcolumn Sep.*, **1**(3), 136–141 (1989).
44. T. Kaneta, S. Tanaka, M. Taga, H. Yoshida, 'Effect of Addition of Glucose on Micellar Electrokinetic Capillary Chromatography with Sodium Dodecyl Sulfate', *J. Chromatogr.*, **609**(1–2), 369–374 (1992).
45. A.-F. Lecoq, L. Montanarella, 'Separation of Nucleosides and Nucleotide-3′-monophosphates by Micellar Elecropkinetic Capillary Chromatography', *J. Microcolumn Sep.*, **5**(2), 105–118 (1993).
46. K. Kawamura, 'Capillary Electrophoretic Separation of Mono- and Dinucleotides of Adenosine Using Cyclodextrin Solutions with $MgCl_2$ as Additive', *J. Chromatogr. A*, **802**(1), 167–177 (1998).
47. A. Guttman, A.S. Cohen, D.N. Heiger, B.L. Karger, 'Analytical and Micropreparative Ultrahigh Resolution of Oligonucleotides by Polyacrylamide Gel High-performance Capillary Electrophoresis', *Anal. Chem.*, **62**(2), 137–141 (1990).
48. Y. Baba, 'Prediction of the Behaviour of Oligonucleotides in High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr., Biomed. Appl.*, **618**(1–2), 41–55 (1993).
49. A. Paulus, J.I. Ohms, 'Analysis of Oligonucleotides by Capillary Gel Electrophoresis', *J. Chromatogr.*, **507**, 113–123 (1990).
50. C. Heller, 'Capillary Electrophoresis of Proteins and Nucleic Acids in Gels and Entangled Polymer Solutions', *J. Chromatogr. A*, **698**(1–2), 19–31 (1995).
51. T. Satow, T. Akiyama, A. Machida, Y. Utagawa, H. Kobayashi, 'Simultaneous Determination of the Migration Coefficient of Each Base in Heterogeneous Oligo-DNA by Gel Filled Capillary Electrophoresis', *J. Chromatogr. A*, **652**(1), 23–30 (1993).
52. T. Satow, 'Analysis of DNA by Capillary Electrophoresis', *Bunseki Kagaku*, **43**(10), 731–743 (1994).
53. A. Guttman, A. Arai, K. Magyar, 'Influence of pH on the Migration Properties of Oligonucleotides in Capillary Gel Electrophoresis', *J. Chromatogr.*, **608**(1–2), 175–179 (1992).
54. D. Demorest, R. Dubrow, 'Factors Influencing the Resolution and Quantitation of Oligonucleotides Separated by Capillary Electrophoresis on a Gel-filled Capillary', *J. Chromatogr.*, **559**(1–2), 43–56 (1991).
55. Y. Cordier, O. Roch, P. Cordier, R. Bischoff, 'Capillary Gel Electrophoresis of Oligonucleotides: Prediction of Migration Times Using Base-specific Migration Coefficients', *J. Chromatogr. A*, **680**(2), 479–489 (1994).
56. A.G. Ogston, 'The Spaces in a Uniform Random Suspension of Fibers', *Trans. Faraday Soc.*, **54**(11), 1754–1757 (1958).
57. O.J. Lumpkin, P. Dejardin, B.H. Zimm, 'Theory of Gel Electrophoresis of DNA', *Biopolymers*, **24**(8), 1573–1593 (1985).
58. K.A. Ferguson, 'Starch-gel Electrophoresis; Application to the Classification of Pituitary Proteins and Polypeptides', *Metab., Clin. Exp.*, **13**(10), 985–1002 (1964).
59. P.D. Grossman, D.S. Soane, 'Experimental and Theoretical Studies of DNA Separations by Capillary Electrophoresis in Entangled Polymer Solutions', *Biopolymers*, **31**(10), 1221–1228 (1991).
60. Y. Baba, M. Tsuhako, T. Sawa, M. Akashi, E. Yashima, 'Specific Base Recognition of Oligodeoxynucleotides by Capillary Affinity Gel Electrophoresis Using Polyacrylamide Poly-(9-vinyladenine) Conjugated Gel', *Anal. Chem.*, **64**(17), 1920–1925 (1992).
61. E. Yashima, N. Suehiro, N. Miyauchi, M. Akashi, 'Affinity Gel Electrophoresis of Nucleic Acids. Nucleobase-selective Separation of DNA and RNA on Agarose-poly(9-vinyladenine) Conjugated Gel', *J. Chromatogr. A*, **654**, 151–158 (1993).
62. E. Yashima, N. Suehiro, N. Miyauchi, M. Akashi, 'Affinity Gel Electrophoresis of Nucleic Acids. Specific Base- and Shape-selective Separation of DNA and RNA on Polyacrylamide–Nucleobase Conjugated Gel', *J. Chromatogr. A*, **654**, 159–166 (1993).
63. A. Muscate, F. Natt, A. Paulus, M. Ehrat, 'Capillary Affinity Gel Electrophoresis for Combined Size- and Sequence-dependent Separation of Oligonucleotides', *Anal. Chem.*, **70**(7), 1419–1424 (1998).
64. T. Wang, G.J. Bruin, J.C. Kraak, H. Poppe, 'Preparation of Polyacrylamide Gel-filled Fused-silica Capillaries by Photopolymerization with Riboflavine as the Initiator', *Anal. Chem.*, **63**(19), 2207–2208 (1991).
65. Y. Baba, T. Matsuura, K. Wakamoto, M. Tsuhako, 'A Simple Method for the Preparation of Polyacrylamide Gel Filled Capillaries for High Performance Separation

of Polynucleotides by Using Capillary Gel Electrophoresis', *Chem. Lett.*, **3**, 371–374 (1991).
66. Y. Baba, T. Matsuura, K. Wakamoto, Y. Morita, Y. Nishitsu, M. Tsuhako, 'Preparation of Polyacrylamide Gel Filled Capillaries for Ultrahigh Resolution of Polyacrylamides by Capillary Gel Electrophoresis', *Anal. Chem.*, **64**(11), 1221–1225 (1992).
67. A. Paulus, E. Gassmann, M.J. Field, 'Calibration of Polyacrylamide Gel Columns for the Separation of Oligonucleotides by Capillary Electrophoresis', *Electrophoresis*, **11**(9), 702–708 (1990).
68. A. Guttman, N. Cooke, 'Capillary Gel Affinity Electrophoresis of DNA Fragments', *Anal. Chem.*, **63**(18), 2038–2042 (1991).
69. W.A. MacCrehan, H.T. Rasmussen, D.M. Northrop, 'Size-selective Capillary Electrophoresis (SSCE) Separation of DNA Fragments', *J. Liq. Chromatogr.*, **15**(6–7), 1063–1080 (1992).
70. Y. Baba, N. IshiMaru, K. Samata, M. Tsuhako, 'High-resolution Separation of DNA Restriction Fragments by Capillary Electrophoresis in Cellulose Derivative Solutions', *J. Chromatogr. A*, **653**(2), 329–335 (1993).
71. A.E. Barron, H.W. Blanch, D.S. Soane, 'A Transient Entanglement Coupling Mechanism for DNA Separation by Capillary Electrophoresis in Ultradilute Polymer Solutions', *Electrophoresis*, **15**(5), 597–615 (1994).
72. K. Khan, A. Van Schepdael, J. Hoogmartens, 'Capillary Electrophoresis of Oligonucleotides Using a Replaceable Sieving Buffer with Low Viscosity-grade Hydroxyethyl Cellulose', *J. Chromatogr. A*, **742**(1–2), 267–274 (1996).
73. L. Roed, I. Arsky, E. Lundances, T. Greibrokk, 'Rapid and Reproducible Capillary Electrophoretic Separation of Double-stranded DNA Fragments in Simple Methyl Cellulosic Sieving System', *Chromatographia*, **47**(3–4), 125–134 (1998).
74. M.C. Ruiz-Martinez, J. Berka, A. Belenkii, F. Foret, A.W. Miller, B.L. Karger, 'DNA Sequencing by Capillary Electrophoresis with Replaceable Linear Polyacrylamide and Laser-induced Fluorescence Detection', *Anal. Chem.*, **65**(20), 2851–2858 (1993).
75. N. Best, E. Arriaga, D.Y. Chen, N.J. Dovichi, 'Separation of Fragments up to 570 Bases in Length by Use of 6%*T* Non-cross-linked Polyacrylamide for DNA Sequencing in Capillary Electrophoresis', *Anal. Chem.*, **66**(22), 4063–4067 (1994).
76. J.Z. Zhang, Y. Fang, J.Y. Hou, H.J. Ren, R. Jiang, P. Roos, N.J. Dovichi, 'Use of Non-cross-linked Polyacrylamide for Four-color DNA Sequencing by Capillary Electrophoresis Separation of Fragments Up to 640 Bases in Length in Two Hours', *Anal. Chem.*, **67**(24), 4589–4593 (1995).
77. E. Carrilho, M.C. Ruiz-Martinez, J. Berka, I. Smirnov, W. Goetzinger, A.W. Miller, D. Bardy, B.L. Karger, 'Rapid DNA Sequencing of More Than 1000 Bases per Run by Capillary Electrophoresis Using Replaceable Linear Polyacrylamide Solutions', *Anal. Chem.*, **68**(19), 3305–3313 (1996).
78. S.A. Soper, D.C. Williams, Y. Xu, S.J. Lassiter, Y. Zhang, S.M. Ford, R.C. Bruch, 'Sanger DNA-sequencing Reactions Performed in a Solid-phase Nanoreactor Directly Coupled to Capillary Gel Electrophoresis', *Anal. Chem.*, **70**(19), 4036–4043 (1998).
79. E. Simo-Alfonso, C. Gelfi, M. Lucisano, P.G. Righetti, 'Performance of a Series of Novel N-substituted Acrylamides in Capillary Electrophoresis of DNA Fragments', *J. Chromatogr., A*, **756**(1–2), 255–261 (1996).
80. E.N. Fung, E.S. Yeung, 'High-speed DNA Sequencing by Using Mixed Poly(ethylene oxide) Solutions in Uncoated Capillary Columns', *Anal. Chem.*, **67**(13), 1913–1919 (1995).
81. Q. Gao, E.S. Yeung, 'A Matrix for DNA Separation: Genotyping and Sequencing Using Poly(vinylpyrrolidone) Solution in Uncoated Capillaries', *Anal. Chem.*, **70**(7), 1382–1388 (1998).
82. K. Kleparnik, S. Fanali, P. Bocek, 'Selectivity of the Separation of DNA Fragments by Capillary Zone Electrophoresis in Low-melting-point Agarose Sol', *J. Chromatogr.*, **638**(2), 283–292 (1993).
83. T. Izumi, M. Yamaguchi, K. Yoneda, T. Isobe, T. Okuyama, T. Shinoda, 'Use of Glucomannan for the Separation of DNA Fragments by Capillary Electrophoresis', *J. Chromatogr.*, **652**(1), 41–46 (1993).
84. K.W. Talmadge, A.K. Tan, M. Zhu, 'DNA Fragment Analysis by Capillary Polymer Sieving Electrophoresis Using Poly(acryloylaminoethoxyethanol)-coated Capillaries', *J. Chromatogr., A*, **781**(1–2), 335–345 (1997).
85. H. Zhu, S.M. Clark, S.C. Benson, H.S. Rye, A.N. Glazer, R.A. Mathies, 'High-sensitivity Capillary Electrophoresis of Double-stranded DNA Fragments Using Monomeric and Dimeric Fluorescent Intercalating Dyes', *Anal. Chem.*, **66**(13), 1941–1948 (1994).
86. D. Figeys, E. Arriaga, A. Renborg, N.J. Dovichi, 'Use of the Fluorescent Intercalating Dyes POPO-3, YOYO-3 and YOYO-1 for Ultrasensitive Detection of Double-stranded DNA Separated by Capillary Electrophoresis with Hydroxypropylmethyl Cellulose and Non-cross-linked Polyacrylamide', *J. Chromatogr., A*, **669**(1–2), 205–216 (1994).
87. M.J. van der Schans, A.W.H.M. Kuypers, A.D. Kloosterman, H.J.T. Janssen, F.M. Everaerts, 'Comparison of Resolution of Double-stranded and Single-stranded DNA in Capillary Electrophoresis', *J. Chromatogr., A*, **772**(1–2), 255–264 (1997).
88. H. Oana, R.W. Hammond, J.J. Schwinefus, S.-C. Wang, M. Doi, M.D. Morris, 'High-speed Separation of Linear and Supercoiled DNA by Capillary Electrophoresis. Buffer, Entangling Polymer, and Electric Field Effects', *Anal. Chem.*, **70**(3), 573–579 (1998).
89. J. Macek, U.R. Tjaden, J. Van der Greef, 'Resolution and Concentration Detection Limit in Capillary

Gel Electrophoresis', *J. Chromatogr.*, **545**(1), 177–182 (1991).
90. M.A. Quesada, H.S. Rye, J.C. Gingrich, A.N. Glazer, R.A. Mathies, 'High-sensitivity DNA Detection with a Laser-excited Confocal Fluorescence Gel Scanner', *BioTechniques*, **10**(5), 616–625 (1991).
91. A.N. Glazer, H.S. Rye, 'Stable Dye–DNA Intercalation Complexes as Reagents for High-sensitivity Fluorescence Detection', *Nature*, **359**, 859–861 (1992).
92. B.B. Haab, R.A. Mathies, 'Single Molecule Fluorescence Burst Detection of DNA Fragments Separated by Capillary Electrophoresis', *Anal. Chem.*, **67**(18), 3253–3260 (1995).
93. R.E. Milofsky, E.S. Yeung, 'Native Fluorescence Detection of Nucleic Acids and DNA Restriction Fragments in Capillary Electrophoresis', *Anal. Chem.*, **65**(2), 153–157 (1993).
94. D.C. Williams, S.A. Soper, 'Ultrasensitive Near-IR Fluorescence Detection for Capillary Gel Electrophoresis and DNA Sequencing Applications', *Anal. Chem.*, **67**(19), 3427–3432 (1995).
95. A.N. Glazer, K. Peck, R.A. Mathies, 'A Stable Double-stranded DNA–Ethidium Homodimer Complex: Application to Picogram Fluorescence Detection of DNA in Agarose Gels', *Proc. Natl. Acad. Sci. USA*, **87**(10), 3851–3855 (1990).
96. L.M. Smith, S. Fung, M.W. Hunkapiller, T.J. Hunkapiller, L.E. Hood, 'The Synthesis of Oligonucleotides Containing an Aliphatic Amino Group at the 5′ Terminus: Synthesis of Fluorescent DNA Primers for Use in DNA Sequence Analysis', *Nucleic Acids Res.*, **13**(7), 2399–2412 (1985).
97. L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B.H. Kent, L.E. Hood, 'Fluorescence Detection in Automated DNA Sequence Analysis', *Nature*, **321**, 674–679 (1986).
98. W. Ansorge, B. Sproat, J. Stergenmann, C. Schwager, M. Zenke, 'Automated DNA Sequencing: Ultrasensitive Detection of Fluorescent Bands during Electrophoresis', *Nucleic Acids Res.*, **15**(11), 4593–4602 (1987).
99. J.M. Prober, G.L. Trainor, R.J. Dam, F.W. Hobbs, C.W. Robertson, R.J. Zagursky, A.J. Cocuzza, M.A. Jensen, K. Baumeister, 'A System for Rapid DNA Sequencing with Fluorescent Chain-terminating Dideoxynucleotides', *Science*, **238**, 336–341 (1987).
100. G.L. Trainor, M.A. Jensen, 'A Procedure for the Preparation of Fluorescence-labeled DNA with Terminal Deoxynucleotidyl Transferase', *Nucleic Acids Res.*, **16**(24), 11846 (1988).
101. S. Tabor, C.C. Richardson, 'DNA Sequence Analysis with a Modified Bacteriophage T7 DNA Polymerase', *J. Biol. Chem.*, **265**(14), 8322–8328 (1990).
102. H. Voss, J. Zimmermann, C. Schwager, H. Erfle, J. Stegemann, K. Stucky, W. Ansorge, 'Automated Fluorescent Sequencing of Cosmid DNA', *Nucleic Acids Res.*, **18**(4), 1066 (1990).
103. J. Zimmermann, H. Voss, C. Schwager, J. Stegemann, H. Erfle, K. Stucky, K.T., W. Ansorge, 'A Simplified Protocol for Fast Plasmid DNA Sequencing', *Nucleic Acids Res.*, **18**(4), 1067 (1990).
104. W. Ansorge, J. Zimmermann, C. Schwager, J. Stergemann, H. Erfle, H. Voss, 'One Label, One Tube, Sanger DNA Sequencing in One and Two Lanes on a Gel', *Nucleic Acids Res.*, **18**(11), 3419–3420 (1992).
105. A.S. Cohen, D.R. Najarian, B. Karger, 'Separation and Analysis of DNA Sequence Reaction Products by Capillary Gel Electrophoresis', *J. Chromatogr.*, **516**, 49–60 (1990).
106. D.Y. Chen, H.R. Harke, N.J. Dovichi, 'Two-label Peak-height Encoded DNA sequencing by Capillary Gel Electrophoresis: Three Examples', *Nucleic Acids Res.*, **20**(18), 4873–4880 (1992).
107. D.Y. Chen, H.P. Swerdlow, H.R. Harke, J.Z. Zhang, N.J. Dovichi, 'Low-cost, High-sensitivity Laser-induced Fluorescence Detection for DNA Sequence by Capillary Gel Electrophoresis', *J. Chromatogr.*, **559**(1–2), 237–246 (1991).
108. F. Sanger, S. Nicklen, A.R. Coulson, 'DNA Sequencing with Chain-terminating Inhibitors', *Proc. Natl. Acad. Sci. USA*, **74**(12), 5463–5467 (1977).
109. O. Salas-Solano, E. Carrilho, L. Kotler, A.W. Miller, W. Goetzinger, Z. Sosic, B.L. Karger, 'Routine DNA Sequencing of 1000 Bases in Less Than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions', *Anal. Chem.*, **70**(19), 3996–4003 (1998).
110. S. Takahashi, K. Murakami, T. Anazawa, H. Kambara, 'Multiple Sheath-flow Gel Capillary-array Electrophoresis for Multicolor Fluorescent DNA Detection', *Anal. Chem.*, **66**(7), 1021–1026 (1994).
111. K. Ueno, E.S. Yeung, 'Simultaneous Monitoring of DNA Fragments Separated by Electrophoresis in a Multiplexed Array of 100 Capillaries', *Anal. Chem.*, **66**(9), 1424–1431 (1994).
112. Y. Wang, J. Ju, B.A. Carpenter, J.M. Atherton, G.F. Sensabaugh, R.A. Mathies, 'Rapid Sizing of Short Tandem Repeat Alleles Using Capillary Arrays Electrophoresis and Energy-transfer Fluorescent Primers', *Anal. Chem.*, **67**(7), 1197–1203 (1995).
113. S.M. Wolf, P. Vouros, 'Incorporation of Sample Stacking Techniques into the Capillary Electrophoresis CF-FAB Mass Spectrometric Analysis of DNA Adducts', *Anal. Chem.*, **67**(5), 891–900 (1995).
114. Z. Zhao, J.H. Wahl, H.R. Udseth, S.A. Hofstadler, A.F. Fuciarelli, R.D. Smith, 'Online Capillary Electrophoresis–Electrospray Ionization Mass Spectrometry of Nucleotides', *Electrophoresis*, **16**(3), 389–395 (1995).
115. D.L.D. Deforce, F.P.K. Ryniers, E.G. Van den Eeckhout, F. Lemiere, E.L. Esmans, 'Analysis of DNA

Adducts in DNA Hydrolysates by Capillary Zone Electrophoresis and Capillary Zone Electrophoresis-electrospray Mass Spectrometry', *Anal. Chem.*, **68**(20), 3575–3584 (1996).

116. D.L.D. Deforce, J. Raymackers, L. Meheus, F. Van Wijnendaele, A. De Leenheer, E.G. Van den Eeckhout, 'Characterization of DNA Oligonucleotides by Coupling of Capillary Zone Electrophoresis to Electrospray Ionization Q-TOF Mass Spectrometry', *Anal. Chem.*, **70**(14), 3060–3068 (1998).
117. A. Harsch, P. Vouros, 'Interfacing of CE in a PVP Matrix to Ion Trap Mass Spectrometry: Analysis of Isomeric and Structurally Related (N-Acetylamino)fluorene-modified Oligonucleotides', *Anal. Chem.*, **70**(14), 3021–3027 (1998).
118. W. Nashabeh, Z.E. Rassi, 'Enzymophoresis of Nucleic Acids by Tandem Capillary Enzyme Reactor–Capillary Zone Electrophoresis', *J. Chromatogr.*, **596**, 251–264 (1992).
119. J. Saevels, K. Huygens, A. Van Schepdael, J. Hoogmartens, 'In-line Coupling of the Enzymatic Degradation of Oligonucleotides with Capillary Polymer Sieving Electrophoresis', *Anal. Chem.*, **89**(16), 3299–3303 (1997).
120. D.-K. Xu, L. Hua, H.-Y. Chen, 'Determination of Purine Bases by Capillary Zone Electrophoresis with Wall-jet Amperometric Detection', *Anal. Chim. Acta*, **335**(1–2), 95–101 (1996).
121. W. Jin, H. Wei, X. Zhao, 'Determination of Adenine and Guanine by Capillary Zone Electrophoresis with End-column Amperometric Detection at a Carbon Fiber Microdisc Array Electrode', *Electrophoresis*, **9**(10), 770–774 (1997).
122. A.-L. Nguyen, J.H.T. Luong, C. Masson, 'Determination of Nucleosides in Fish Tissues Using Capillary Electrophoresis', *Anal. Chem.*, **62**(22), 2490–2493 (1990).
123. X. Huang, J.B. Shear, R.N. Zare, 'Quantitation of Ribonucleotides from Base-hydrolyzed RNA Using Capillary Zone Electrophoresis', *Anal. Chem.*, **62**(18), 2049–2051 (1990).
124. X.C. Huang, S.G. Stuart, P.F. Bente, III, T.M. Brennan, 'Capillary Gel Electrophoresis of Single-stranded DNA Fragments with UV Detection', *J. Chromatogr.*, **600**(2), 289–295 (1992).
125. H. Arakawa, A. Tsuji, M. Maeda, M. Kamahori, H. Kambara, 'Analysis of Single-strand Conformation Polymorphisms by Capillary Electrophoresis with Laser-induced Fluorescence Detection', *J. Pharm. Biomed. Appl.*, **15**(9–10), 1537–1544 (1997).
126. V. Murry, 'Improved Double-stranded DNA Sequencing Using the Linear Polymerase Chain Reaction', *Nucleic Acids Res.*, **17**(21), 8889 (1988).
127. H.S. Rye, M.A. Quesada, K. Peck, R. Mathies, A.N. Glazer, 'High-sensitivity Two-color Detection of Double-stranded DNA with a Confocal Fluorescence Gel Scanner Using Ethidium Homodimer and Thiazole Orange', *Nucleic Acids Res.*, **19**(2), 327–333 (1991).
128. Y. Baba, T. Matsuura, K. Wakamoto, M. Tsuhako, 'Comparison of High-performance Liquid Chromatography with Capillary Gel Electrophoresis in Single-base Resolution of Polynucleotides', *J. Chromatogr.*, **558**(1), 273–284 (1991).
129. E. Katsivela, M.G. Hoefle, 'Separation of Transfer RNA and 5S Ribosomal RNA Using Capillary Electrophoresis', *J. Chromatogr. A*, **700**(1–2), 125–136 (1995).
130. L. DeDionisio, 'Capillary Gel Electrophoresis and the Analysis of DNA Phosphorothioates', *J. Chromatogr.*, **652**(1), 101–108 (1993).
131. G.S. Srivatsa, P. Klopchin, M. Batt, M. Feldman, R.H. Carlson, D.L. Cole, 'Selectivity of Anion Exchange Chromatography and Capillary Gel Electrophoresis for the Analysis of Phosphorothioate Oligonucleotides', *J. Pharm. Biomed. Anal.*, **16**(4), 619–630 (1997).
132. J.A. Luckey, L.M. Smith, 'Optimization of Electric Field Strength for DNA Sequencing in Capillary Gel Electrophoresis', *Anal. Chem.*, **65**(20), 2841–2850 (1993).
133. T. Manabe, N. Chen, S. Terabe, M. Yohda, I. Endo, 'Effects of Linear Polyacrylamide Concentrations and Applied Voltages on the Separation of Oligonucleotides and DNA Sequencing Fragments by Capillary Electrophoresis', *Anal. Chem.*, **66**(23), 4243–4252 (1994).
134. M.C. Ruiz-Martinez, O. Salas-Solano, E. Carrilho, L. Kotler, B.L. Kargen, 'A Sample Purification Method for Rugged and High-performance DNA Sequencing by Capillary Electrophoresis Using Replaceable Polymer Solutions. A. Development of the Cleanup Protocol', *Anal. Chem.*, **70**(8), 1516–1527 (1998).
135. O. Salas-Solano, M.C. Ruiz-Martinez, E. Carrilho, L. Kotler, B.L. Kargen, 'A Sample Purification Method for Rugged and High-performance DNA Sequencing by Capillary Electrophoresis Using Replaceable Polymer Solution. B. Quantitative Determination of the Role of Sample Matrix Components on Sequencing Analysis', *Anal. Chem.*, **70**(8), 1528–1535 (1998).
136. H. Lu, E. Arriaga, D.Y. Chen, N.J. Dovichi, 'High-speed and High-accuracy DNA Sequencing by Capillary Gel Electrophoresis in a Simple, Low-cost Instrument. Two-colour Peak-height Encoded Sequencing at 40 °C', *J. Chromatogr. A*, **680**(2), 497–501 (1994).
137. J.H. Flanagan, Jr, C.V. Owens, S.E. Romero, E. Waddell, S.H. Kahn, R.P. Hammer, S.A. Soper, 'Near-infrared Heavy-atom-modified Fluorescence Dyes for Base-calling in DNA-sequencing Applications Using Temporal Discrimination', *Anal. Chem.*, **70**(13), 2676–2684 (1998).
138. B.K. Nunnally, H. He, L.C. Li, S.A. Tucker, L.B. McGown, 'Characterization of Visible Dyes for Four-decay Fluorescence Detection in DNA Sequencing', *Anal. Chem.*, **69**(13), 2392–2397 (1997).
139. L. Xu, N. Bian, Z. Wang, S. Abdel-Baky, S. Pillai, D. Magiera, V. Murugaiah, R.W. Giese, P. Wang,

T. O'Keefe, H. Abushamaa, L. Kutney, G. Church, S. Carson, D. Smith, M. Park, J. Wronka, F. Laukien, 'Electrophore Mass Tag Dideoxy DNA Sequencing', *Anal. Chem.*, **69**(17), 3595–3602 (1997).

140. S.M. Clark, R.A. Mathies, 'High-speed Parallel Separation of DNA Restriction Fragments Using Capillary Array Electrophoresis', *Anal. Biochem.*, **215**(2), 163–170 (1993).
141. T. Anazawa, S. Takahashi, H. Kambara, 'A Capillary Array Gel Electrophoresis System Using Multiple Laser Focusing for DNA Sequencing', *Anal. Chem.*, **68**(15), 2699–2704 (1996).
142. A.T. Woolley, G.T. Sensabaugh, R.A. Mathies, 'High-speed DNA Genotyping Using Microfabricated Capillary Array Electrophoresis Chips', *Anal. Chem.*, **69**(11), 2181–2186 (1997).
143. S.M. Clark, R.A. Marthies, 'Multiplex dsDNA Fragment Sizing Using Dimeric Intercalation Dyes and Capillary Array Electrophoresis: Ionic Effects on the Stability and Electrophoretic Mobility of DNA–Dye Complexes', *Anal. Chem.*, **69**(7), 1355–1363 (1997).
144. H. Swerdlow, B.J. Jones, C.T. Wittwer, 'Fully Automated DNA Reaction and Analysis in a Fluidic Capillary Instrument', Anal. Chem., **69**(5), 848–855 (1997).
145. S. Kim, H.J. Yoo, H.-G. Nam, J.H. Hahn, 'Fast Capillary-scanning System for Detecting Fluorescently Labeled DNA Sequencing Fragments Separated by Capillary Gel Electrophoresis', *Chromatographia*, **40**(5–6), 345–349 (1995).
146. S. Kim, H.J. Yoo, J.H. Hahn, 'Postelectrophoresis Capillary Scanning Method for DNA Sequencing', *Anal. Chem.*, **68**(5), 936–939 (1996).
147. H. Tan, E.S. Yeung, 'Integrated On-line System for DNA Sequencing by Capillary Electrophoresis: From Template to Called Bases', *Anal. Chem.*, **69**(4), 664–674 (1997).
148. Y. Baba, R. Tomisaki, M. Tsuhako, 'Three-dimensional Electropherogram for the Separation of Oligodeoxynucleotides and DNA Restriction Fragments Using Capillary Gel Electrophoresis with a Photodiode Array Detector', *J. Liq. Chromatogr.*, **18**(7), 1317–1324 (1995).
149. P. Sun, R.A. Hartwick, 'The Effect of Electric Fields on the Dispersion of Oligonucleotides Using a Multi-point Detection Method in Capillary Gel Electrophoresis', *J. Liq. Chromatogr.*, **17**(9), 1861–1875 (1994).
150. A.T. Woolley, R.A. Mathies, 'Ultra-high-speed DNA Sequencing Using Capillary Electrophoresis Chips', *Anal. Chem.*, **67**(20), 3676–3680 (1995).
151. T. Kaneta, H. Jeong, E.S. Yeung, 'Miniaturized Injection Method Using Silver-coated Capillaries in DNA Sequencing by Multiplexed Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **21**(5), 287–290 (1998).
152. R.G. Brownlee, F.J. Sunzeri, M.P. Busch, 'Application of Capillary DNA Chromatography to Detect AIDS Virus (HIV-1) DNA in Blood', *J. Chromatogr., Biomed. Appl.*, **533**, 87–96 (1990).
153. C.S. Effenhauser, A. Paulus, A. Manz, H.M. Widman, 'High-speed Separation of Antisense Oligonucleotides on a Micromachined Capillary Electrophoresis Device', *Anal. Chem.*, **66**(18), 2949–2953 (1994).
154. B. Krattiger, A.E. Bruno, H.M. Widmer, R. Daendliker, 'Hologram-based Thermooptical Absorbance Detection in Capillary Electrophoresis: Separation of Nucleosides and Nucleotides', *Anal. Chem.*, **67**(1), 124–130 (1995).
155. P.G. Righetti, C. Gelfi, 'Analysis of Clinically Relevant, Diagnostic DNA by Capillary Zone and Double-gradient Gel Slab Electrophoresis', *J. Chromatogr., A*, **806**(1), 97–112 (1998).
156. S.F. Liu, J.A. North, B.K. Murray, M.X. Huang, M.L. Lee, 'Analysis of Intracellular Ribonucleotide Pools in a Drug-treated Human Cancer Cell Line Using Capillary Zone Electrophoresis', *J. Microcolumn Sep.*, **6**(1), 49–54 (1994).
157. L. Reyderman, S. Stavchansky, 'Quantitative Determination of Short Single-stranded Oligonucleotides from Blood Plasma Using Capillary Electrophoresis with Laser-induced Fluorescence', *Anal. Chem.*, **69**(16), 3218–3222 (1997).
158. C. Bory, C. Chantin, D. Bozon, 'Capillary Electrophoretic Analysis of DNA Restriction Fragments and PCR Products for Polymorphism and Mutation Studies in Cystic Fibrosis and Gaucher's Disease', *J. Pharm Biomed. Anal.*, **13**(4–5), 511–514 (1995).
159. A.W.H.M. Kuypers, P.M.W. Willems, M.J. van der Shcans, P.C.M. Linssen, H.M.C. Wessels, C.H.M.M. de Bruijn, F.M. Everaerts, E.J.B.M. Mensink, 'Detection of Point Mutations in DNA Using Capillary Electrophoresis in a Polymer Network', *J. Chromatogr., Biomed. Appl.*, **132**(2), 149–156 (1993).
160. H. Arakawa, K. Uetanaka, M. Maeda, A. Tsuji, 'Analysis of Polymerase Chain Reaction Product by Capillary Electrophoresis and its Application to the Detection of Single Base Substitution in Genes', *J. Chromatogr. A*, **664**(1), 89–98 (1994).
161. Y. Baba, R. Tomisaki, C. Sumita, M. Tsuhako, T. Miki, T. Ogihara, 'High-resolution Separation of PCR Product and Gene Diagnosis by Capillary Gel Electrophoresis', *Biomed. Chromatogr.*, **8**(6), 291–293 (1994).
162. K. Peck, S.-L. Wung, G.-S. Chang, J.J.Y. Yen, Y.-Z. Hsieh, 'Restriction Mapping of Genes by Capillary Electrophoresis with Laser-induced Fluorescence Detection', *Anal. Chem.*, **69**(7), 1380–1384 (1997).
163. B.-M. Stalbom, A. Torven, L.G. Lundberg, 'Application of Capillary Electrophoresis to the Post-polymerase Chain Reaction Analysis of Rat mRNA for Gastric H^+, K^+-ATPase', *Anal. Biochem.*, **217**(1), 91–97 (1994).
164. K. Hebenbrock, K. Schuegerl, R. Freitag, 'Analysis of Plasmid-DNA and Cell Protein of Recombinant

Escherichia coli Using Capillary Gel Electrophoresis', *Electrophoresis*, **14**(8), 753–758 (1993).

165. E.S.C. Koay, M. Zhu, T. Wehr, M.L. Choong, M.C. Khaw, S.K. Sethi, T.C. Aw, 'Detection of Apolipoprotein E Genotypes by Capillary Electrophoresis', *Talanta*, **45**(4), 673–681 (1998).
166. M. Strege, A. Lagu, 'Separation of DNA Restriction Fragments by Capillary Electrophoresis Using Coated Fused Silica Capillaries', *Anal. Chem.*, **63**(13), 1233–1236 (1991).
167. A. Guttman, B. Wanders, N. Cooke, 'Enhanced Separation of DNA Restriction Fragments by Capillary Gel Electrophoresis Using Field Strength Gradients', *Anal. Chem.*, **64**(20), 2348–2351 (1992).
168. K.C. Chan, C.-W. Whang, E.S. Yeung, 'Separation of DNA Restriction Fragments Using Capillary Electrophoresis', *J. Liq. Chromatogr.*, **16**(9–10), 1941–1962 (1993).
169. J. Cheng, K.R. Mitchelson, 'Glycerol-enhanced Separation of DNA Fragments in Entangled Solution Capillary Electrophoresis', *Anal. Chem.*, **66**(23), 4210–4214 (1994).
170. C.V. Owens, Y.Y. Davidson, S. Kar, S.A. Soper, 'High-resolution Separation of DNA Restriction Fragments Using Capillary Electrophoresis with Near-IR, Diode-based, Laser-induced Fluorescence Detection', *Anal. Chem.*, **69**(6), 1256–1261 (1997).
171. H.E. Schwartz, K. Ulfelder, F.J. Sunzeri, M.P. Busch, R.G. Brownlee, 'Analysis of DNA Restriction Fragments and Polymerase Chain Reaction Products towards Detection of the AIDS (HIV-1) Virus in Blood', *J. Chromatogr.*, **559**(1–2), 267–283 (1991).
172. H.E. Schwartz, K.J. Ulfelder, 'Capillary Electrophoresis with Laser-induced Fluorescence Detection of PCR Fragments Using Thiazole Orange', *Anal. Chem.*, **64**(15), 1737–1740 (1992).
173. A.T. Woolley, D. Hadley, P. Landre, A.J. deMello, R.A. Mathies, M.A. Northrup, 'Functional Integration of PCR Amplification and Capillary Electrophoresis in a Microfabricated DNA Analysis Device', *Anal. Chem.*, **68**(23), 4081–4086 (1996).
174. Y. Baba, M. Tsuhako, S. Enomoto, A.M. Chin, R.S. Dubrow, 'High Performance Separation of Nucleic Acids Using Capillary Electrophoresis and High-performance Liquid Chromatography', *J. High Resolut. Chromatogr.*, **14**(3), 204–206 (1991).
175. A.M. Kamel, P.R. Brown, 'High-performance Liquid Chromatography versus Capillary Electrophoresis for the Separation and Analysis of Oligonucleotides', *Am. Lab.*, **28**(3), 42–46 (1996).
176. P.J. Oefner, G.K. Bonn, C.G. Huber, S. Nathakarnkitkool, 'Comparative Study of Capillary Zone Electrophoresis and High-performance Liquid Chromatography in the Analysis of Oligonucleotides and DNA', *J. Chromatogr.*, **625**(2), 331–340 (1992).
177. A. Loregian, C. Scremin, M. Schiavon, A. Marcello, G. Palu, 'Quantitative Analysis of Ribonucleotide Triphosphates in Cell Extracts by High-performance Liquid Chromatography and Micellar Electrokinetic Capillary: a Comparative Study', *Anal. Chem.*, **66**(18), 2981–2984 (1994).
178. Y. Baba, C. Sumita, K. Hide, N. Ishimaru, K. Samata, A. Tanaka, M. Tsuhako, 'Separation of DNA Fragments by High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Liq. Chromatogr.*, **16**(4), 955–965 (1993).
179. C. Sumita, Y. Baba, K. Hide, N. Ishimaru, K. Sumata, A. Tanaka, M. Tsuhako, 'Comparative Study of Nonporous Anion-exchange Chromatography, Capillary Gel Electrophoresis and Capillary Electrophoresis in Polymer Solutions in the Separation of DNA Restriction Fragments', *J. Chromatogr., A*, **661**(1–2), 297–303 (1994).
180. D.L.D. Deforce, R.E.M. Millecamps, D. Van Hoofstat, E.G. Van den Eeckhout, 'Comparison of Slab Gel Electrophoresis and Capillary Electrophoresis for the Detection of the Fluorescence Labeled Polymerase Chain Reaction Products of Short Tandem Repeat Fragments', *J. Chromatogr., A*, **806**(1), 149–155 (1998).

Comparative Genomics: Differential Display and Subtractive Hybridization

Cassandra L. Smith, Joseph Bouchard, Gregg Surdi, Giang Nguyen, K. Pimpis, and Linda G. Tolstoi
Boston University, Boston, USA

1 Introduction **1**
2 Options for Comparing Genomes **2**
2.1 DNA Fingerprinting Methods 2
3 Differential Display **3**
3.1 Principles 3
3.2 Options 4
3.3 Trouble Shooting 4
3.4 Applications 5
4 Subtractive Hybridization **5**
4.1 Principles of Subtractive Hybridization 5
4.2 Options 6
4.3 Trouble Shooting 7
4.4 Applications 7
5 Other Considerations **7**
6 Perspective **8**
Acknowledgments **8**
Abbreviations and Acronyms **8**
Related Articles **8**
References **8**

An understanding of genome function within the context of human cells requires new methodology. This article describes comparative approaches that are used for genomes and genes in situ. However, the major focus of the article is on differential display (DD) and subtractive hybridization (SH) because these two methods are most accessible to a large number of researchers. The basic principles of these methods are described along with some fundamental reminders about the important factors that must be kept in mind for successful experiments.

1 INTRODUCTION

There are an increasing number of genomic sequences becoming available. Within the next 5 years even the human genome sequence should be completed. Understanding the meaning of these DNA sequences will take much more work and involve many comparative experiments.

In the past, gene function was defined by comparing a cell containing a specific mutation with a cell that did not have the mutation. Mutations are also used quite effectively for dissecting biochemical mechanisms and pathways. The power of using mutations lies in the fact that a very well defined change is being studied. In the past, most research was confined to model organisms with well-developed genetic systems that allowed gene manipulations. This type of approach is quite useful for experimental organisms that have well-developed genetic systems which allow gene manipulation. It is quite clear that these types of experiments will continue in the future. However, it is also clear that we need to understand gene function in complex organisms where gene manipulation is not possible or easily done.

Comparative studies in the complex genomes are difficult. For instance, it is estimated that there may be over 100 000 genes in the human genome. Of course, there are many naturally occurring mutations that aid geneticists and other researchers to understand disease and other natural processes. However, even putting aside the ethics of manipulating human genomes for experimental purposes, humans are not good experimental organisms because their life span is too long. Hence, it is quite clear that comparative methods focused on complex organisms with long life spans must be independent of gene manipulations. Furthermore, even with ''good'' experimental organisms, the number of known genes and interactions are too great to be dissected by past methods using conventional recombinant DNA methods. In such experiments, genes are cloned, sequenced, modified and studied in detail out of their native genomic environments.

A large number of methods and variations have been developed for comparing whole genomes and entire repertoires of genes. Most methods can be applied to large and small genomes. The technical difficulty of a particular approach is usually a function of genome size. This article will review genomic comparative techniques, focusing on DD and SH because these methods are most accessible to a large number of researchers.

For simplicity, all calculations done will assume a random DNA sequence composition. However, for many organisms the amount of sequence information available in the databases allows for more accurate calculations of experimental results and for some completely sequenced genomes, a total modeling of the experimental results.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 OPTIONS FOR COMPARING GENOMES

There are several approaches for comparing, globally, genome structure and function. These studies can be divided into those that compare the primary structure of genomic DNA and those that compare gene expression profiles. It should be noted that gene expression profiling requires quantitative analysis of levels that vary by 10^5-fold. This means that quantitative studies on gene expression are more complicated than direct genome comparisons.

Gene expression profiling is currently most often done at the mRNA level. Such experiments do not provide information about post-transcriptional regulation of gene expression. Two-dimensional gel electrophoresis has been used for some time to establish and compare protein profiles from different samples. Some recent experiments compare mRNA and protein profiles from the same sample. In yeast, Gygi et al.[1] found little correlation between protein and mRNA expression levels. It will be interesting to see what post-transcriptional regulatory mechanisms are uncovered and how these initial observations are extended to other organisms and systems. These types of comparative experiments and protein profiling in general will not be discussed here.

2.1 DNA Fingerprinting Methods

Genomic sequence comparisons between different samples is the only reliable method to detect all differences. Gene expression profiling can be done using DNA sequencing methodology. In these approaches, random cDNAs are sequenced and the frequency of occurrence of specific sequences is recorded in different cells or under different conditions such as disease states. Some have termed this electronic profiling and several successful biotechnology companies specialize in this approach. These types of experiments form the new field of pharmacogenetics. (Pharmacogenetics promises to tailor treatment and medication to individuals.)

A variation on random cDNA sequencing is called serial analysis of gene expression (SAGE).[2] SAGE links short bits of cDNA sequence together in a single sequencing template to increase efficiency of the analysis. Even so, the financial cost of current technology prevents widespread use of large-scale sequencing experiments for global comparative studies.

DNA fingerprint methods have been used to compare both genomic DNA and cDNA sequences. The first method that was developed hybridized a multilocus simple, tandem repeat probe to restriction enzyme cleaved genomic DNA fractionated electrophoretically.[3] Since then a number of easier polymerase chain reaction (PCR) methods have been developed for DNA fingerprinting. In some cases, PCR is used to amplify unique sequences located between primer sequences composed of arbitrary primers (randomly amplified polymerase chain reaction, RAPD).[4,5] In other cases, the primer sequences consist of interspersed repeat sequences, for example Alu sequences in the human genome.[6] The pool of PCR amplified fragments are size fractionated electrophoretically to create a complex DNA fingerprint. The DNA fingerprints of two or more samples are compared to assess the similarity and differences between samples.

DD[7] compares the size distribution pattern of DNA fragments generated from random amplification of two mRNA samples. Here, we will distinguish cDNA differential display (cDD) from genomic differential display (GDD). A more detailed discussion of these methods is given below.

Methods such as sequencing by hybridization (SBH, also known as DNA chip arrays)[8,9] replace time-consuming electrophoretic size fractionations with a single-stranded DNA array composed of different sequences. Hybridization of a single-stranded sample to a DNA array produces a DNA fingerprint revealing which array sequences are complementary to a position of the sample DNA. In conventional SBH, the presence but not the location of the sequence is obtained. The sample sequence may be reconstructed from the pattern of short oligonucleotides that hybridize to the template. An alternative method, called positional sequencing by hybridization (PSBH),[10] revealed both the presence and location of the complementary sequence. In this protocol, the array targets are partially double stranded and have a single-stranded end. In this format, the single-stranded end of the target sequence hybridizes to the probe element with stacking interactions from the duplex region insuring a high level of matching bases. The specificity of this hybridization may be increased by adding enzymatic steps, such as a DNA ligation and/or polymerase extension reaction.

In SBH, oligonucleotides are typically immobilized in an array format on a silicon chip to which a labeled template is hybridized. The presence of the template on each array element is scored as positive or negative when conventional radioactive, fluorescent or chemiluminescent labels are used.

Conventional SBH experiments have been severely handicapped by the hybridization of mismatched sequences, especially end mismatches. PSBH reduces the level of mismatches at least 10-fold. Further enhancement may be obtained when SBH experiments are analyzed by mass spectrometry (MS).[11]

MS analysis does not require a label. MS not only detects the presence of a species as do the other detection methods but also provides the mass of the molecule. Since the masses of the bases are known, different

sequence compositions with distinguishable masses may be distinguished when they are hybridized to the same array element. Potentially, MS allows the experimental results to be precisely correlated to DNA sequence(s).

DNA arrays may also be composed of other molecules. For instance, cloned cDNAs,[12,13] or genomic sequences[14] or even uncloned single or pools of genomic restriction fragments[15] may be used as targets.

Comparative genome hybridization (CGH)[16] experiments hybridize pairs of differentially labeled genomic DNA or cDNAs to normal metaphase chromosomes. Usually, one sample is labeled with fluorescein and the second with 4′,6-diamidino-2-phenylindole (DAPI). The samples are mixed together at equimolar ratios. Hybridization to metaphase chromosome means that the DNA probes are sorted to chromosomal regions. The fluorescein/DAPI signal ratio will remain contant unless there is a region that is amplified or deleted in one of the genomic DNA samples. Likewise, when cDNA probes are analyzed, the fluorescein/DAPI signal ratio will remain constant except for those regions where there is a difference in gene expression. CGH provides a low resolution (~10 Mb) genomic localization of differences and does not provide directly the sequence causing the difference. The advantage is that a controlled pairwise comparison is done, providing a much more robust quantitative analysis. The disadvantage is that the complexities of this method do not allow it to be adapted by a large number of laboratories. A similar type of ''two-color'' approach has also been applied to analyzing DNA arrays.[17]

2.1.1 Isolation of DNA Differences Between Samples

All of the experiments discussed above may be classified as DNA fingerprinting methods. These methods present similarities and differences between samples. An alternative to these approaches are methods that provide information only about the differences between samples.

In SH, only ''candidate'' differences are provided to the experimenter. An advantage of the DNA fingerprinting methods are that they can be quantitative and provide ongoing information to the experimenter. An advantage of the SH experiments is that they provide directly the candidate differences between samples. The DNA fingerprinting experiments provide information about differences that then must be isolated for further testing. In both approaches, the first chore of the experimenter is to prove that the indicated differences are real. This will allow the level of false differences between samples (i.e. false positives) to be assessed. The level of false negatives (false similarities) between samples is more difficult to assess. Our own approach to this problem has been to verify our method using a model system where the experimental results can be compared with the theoretical results (see below) and where few, if any, differences should exist between samples.

DD and SH are technically demanding. The key to success lies in understanding the underlying biochemistry and biophysics. The goal of this article is to provide the user with a basic understanding of the principles of these methods and factors that affect reliability. Also discussed are some of the many options and applications.

3 DIFFERENTIAL DISPLAY

3.1 Principles

DD was originally developed by Liang and Pardee[18] as a PCR method for amplifying and labeling subsets of mRNA to identify differences in gene expression. A recent useful collation of a variety of DD articles can be found in Liang and Pardee.[7] The pool of amplified products <500 bp (base pairs) is size-fractionated on a high-resolution DNA sequencing gel. This produces a display consisting of the size-dependent banding pattern of the pool of PCR products. The display produced from two or more samples is compared to identify differences in gene expression patterns.

In the originally described method,[18] the starting material was purified mRNA which is converted to cDNA using conventional methods. Then PCR was used to amplify subsets of the cDNA library. One of the PCR primer pairs was complementary to the polyA tail of eukaryotic mRNA with additional 3′-anchor bases (e.g. $5'T_{11}CA3'$). The use of a homopolymer T primer with a unique dinucleotide 3′-end serves two purposes. First, the unique dinucleotides anchored the primer to the unique genomic sequences adjacent to the 5′-end of the polyA mRNA tail. This insured the polymerase extension reaction initiated at the sequence adjacent to the polyA tail and not randomly within the mRNA polyA tail. The anchor bases also control the complexity of the PCR amplification products. Twelve different dinucleotide (e.g. CA, CT, CG, CC, GA, GT, GG, GC, AA, AT, AG, AC) anchors are needed to amplify all mRNA specifically. In a random sequence, a dinucleotide anchored primer sequence would reduce the number of PCR amplification with a specific dinucleotide anchor products by 1/12.

The second PCR primer was a 10-mer composed of an arbitrary sequence (e.g. Ltk3, CTTGATTGCC). There are 1049 different possible 10-mer combinations and any one 10-mer sequence would be expected to occur once in 10^6 bp. Since an average mRNA is about 10^3 bp in length, a single 10-mer should hybridize to 1 in 10^3 cDNAs.

DD analysis is usually restricted to fragments less than 10^3, hence the number of amplified products should be even less. The arbitrary primers used in these experiments

were chosen after testing to give reproducible and specific amplification products with the polyT anchored primers. Subsequent studies have taken a more critical look at the theoretical distribution of primers so that the speed and efficiency of DD could be maximized.[19]

3.2 Options

3.2.1 Genomic Analysis

Genomic comparisons of simple genomes may be done directly. For instance, it is quite easy using pulsed field gradient (PFG) techniques to compare chromosomal fingerprints of small genomes like bacteria, yeast and protozoan.[20] The resolution of these methods can be improved by establishing fingerprints using restriction enzymes that cleave infrequently and in some cases restriction enzymes that cleave frequently. These types of approaches can be applied to organisms whose genomes range in size up to about 100 million bp. In contrast, the human genome is 1000-fold greater than this. With today's technology it is still impossible to compare directly total human genomes. Instead genome complexity must be reduced.

We[21–25] and others[26–28] have reported on the use of targeting in DD. In this approach, a class of fragments containing a specific sequence (a target sequence) is selected for analysis (see below). In contrast to others, our methods have focused on analyzing selected genomic fragments rather than cDNAs (targeted genomic differential display (TGDD) vs targeted cDNA differential display (TcDD) respectively). The focus on genomic DNA allows for an exact modeling of the expected results using DNA from genomes whose entire sequence is known and was easier to develop as a quantitative method because of the smaller dynamic range requirements (e.g. ~2-fold versus 10^5-fold for genomic vs cDNA studies, respectively).

Two methods for TGDD were established by us.[29] In both methods, oligonucleotides of known sequence are ligated to the ends of genomic restriction fragments. In Method I, there is an initial hybridization – capture step which is used to purify the restriction fragments that contain the target sequence away from the remaining restriction fragments. An immobilized oligonucleotide complementary to the target sequence is used for the hybridization – capture step. Subsequently, the captured fragments are amplified by PCR. A single PCR primer complementary to the adapter sequences ligated onto the ends of the genomic fragments may be used. Alternatively, the adapter primer may be used with a primer complementary to the target sequence. In the former case, sequences from each side of the target are amplified; in the latter case the sequence from only one side of the target is amplified.

Method I eliminates the capture step and uses long (40-mer) adapter primers to enhance self-annealing of the adapter primers, an effect called PCR suppression.[30] In this method, a primer to the target sequence and a short 20-mer primer to the adapter sequence are used. Only fragments that contain a target sequence can be amplified. This version of the method has the advantage that it is easier to perform. Although, the adaption of interspersed repetitive element-bubble PCR[31] with Method I should prevent self-annealing of the adapter sequences,

Targeting has been adapted to cDNA studies (Nguyen et al., unpublished results). In TcDD, the adapter primer was replaced by a primer complementary to the polyA tail.

3.2.2 Targeting

A detailed discussion of targeting issues can be found in Bouchard et al.[29] In brief, it is important to realize that targeting can be done with a simple repeating sequence (the polyA tail of eukaryotic cDNA is one example of this) or more complicated sequence coding for a protein motif, a *cis*-acting sequence such as transcriptional activation signal. The method can analyze the target sequence itself and/or the surrounding sequence depending on the PCR primer that is used. Usually, the target is interspersed throughout the genome. This means that the method samples sequences spread throughout the genome. In some cases the target may be more confined to a specific chromosomal region, for example telomeric sequences. In the case of simple repeating sequences, unique bases added to the 3′- or 5′-end of the primer serves to anchor the primer to the target end and to reduce complexity (see above).

The question of how to design more complex motif primers that target gene families is not yet answered. Most, if not all, gene families consensus sequences were developed comparing amino acid compositions and not DNA sequences. In fact, very little work has been done on characterizing gene families at the DNA sequence level because it is clear that there can be much more variability at the DNA sequence level because of the redundancy of the genetic code. Additionally, different preferential codon usage is observed in different organisms and an emerging complication is that codon usage in a particular species may reflect the expression level of the protein sequence and/or be protein class specific.

3.3 Trouble Shooting

The major problem with DD experiments is the high number of false positives. A suggested way of dealing with this issue is to reconfirm each alleged difference individually by hybridization. This requires that the

DNA fragment be isolated and/or sequenced so that a hybridization probe can be generated.

Our own experience with TGDD has suggested several ways of improving the experimental results. One of the most important factors is to analyze high-resolution Gaussian-like distributions of the fragment size fractionations. In most DD experiments, the sample is radioactively labeled and the fractionation results recorded by exposure of the polyacrylamide gel to film to produce a banding pattern. Although this type of analysis is useful, especially if the focus is kept on the differential presence of very dark bands, a much more realistic pattern of intensity changes is obtained when higher resolution data are collected in real time on an automated DNA sequencing instrument.

Our extensive characterization of the factors that influence the reliability of DD has led us to the conclusion that the most important aspect is to match the starting DNA concentrations. There are several reasons for doing this; in particular, the occurrence of unequal amplification of the PCR product during the late cycles. This problem is likely related to the specific sequence that needs to be amplified. For example, although the human genome is 60% A + T, usually equal amounts of the trinucleotide precursors are added to the PCR. This may mean that in each cycle of PCR there is a 20% difference in the use of trinucleotide precursors dATP and TTP vs dCTP and dGTP. Furthermore, the Michaels constant (K_m) of each trinucleotide is different. Hence, the cycle at which precursors become limiting will differ. Of course, this is only a theoretical consideration. The real inequities must be based on the base composition of each fragment in the amplified pools. Another cause of unequal amplification may be the rapid annealing of high concentration PCR products, which would then prevent the subsequent annealing or extension of primers. These considerations emphasize the importance of using samples with matched DNA concentrations and a low number of PCR cycles.

Another consideration in any experiment involving nucleic acids, and especially those involving DNA, is the accurate determination of the sample concentration. It is difficult to accurately determine DNA concentrations. In the cell, 90% of the nucleic acid is RNA. This means that the greatest detriment to accurately determining DNA concentration from a cell extract is contaminating RNA sequences. Although it is possible to purify DNA from RNA via density gradient centrifugations, this is not usually done because of the limiting amount of material that is available. RNA or DNA specific dyes can be used but they will bind to the alternative species with less affinity as pointed out by many of the manufacturers. The problem is that the amount of contaminating material is unknown. Hence, it is more useful to be consistent in the manner in which DNA concentrations are determined and correlating this with useful results. In fact, since the entire scientific community has this problem it might be best to try several initial (two-fold variations below and above the target) DNA concentrations. Additionally, one might consider varying the number of PCR cycles by two cycles around the target to minimize the number of experiments needed, especially when starting out.

An important issue that cannot be ignored is the ratio and concentration of the primers in the PCR. For small oligonucleotide primers, the annealing temperature will be concentration dependent in addition to being sequence dependent.[32] Other considerations in these types of multilocus are the frequency of occurrence of particular sequences in the template DNA as well as the sequence conservation. The best way to handle these variables is to determine the best ratio and concentration of primer pairs to use empirically. Typically we titrate multilocus primers against each other in the 0.3 to 2 mM range.

3.4 Applications

There are many applications for DD and an increasing number of publications are appearing using this technology. Most publications focus on cDNA applications. Our own studies using TGDD on closely matched genomes compare the genomes of twins, families and different tissues from the same individual. The goal of these experiments is to determine the effect of genome stability and heritability on phenotype.

4 SUBTRACTIVE HYBRIDIZATION

4.1 Principles of Subtractive Hybridization

The goal of SH experiments is to isolate a target sequence(s) (T) that is present in one sample, called the test DNA (S), and absent in a matched sample, called the driver DNA (D). There are a large number of published protocols. In all cases, a mixture of excess D mixed with S, is denatured and allowed to reanneal. In this reaction there are three single-stranded species which can be distinguished as shown in Figure 1.

The D and S DNAs can form parental homoduplexes or heteroduplexes with complementary strands. All

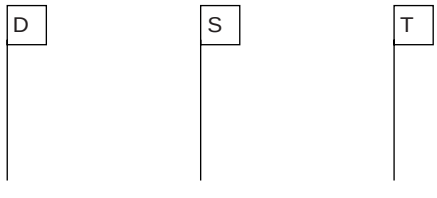

Figure 1 Single-stranded DNA.

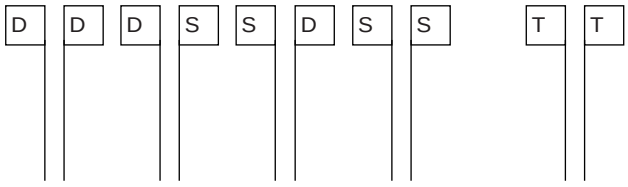

Figure 2 T DNA homoduplexes.

sequences are in common except for the T DNA. The T DNA can only form homoduplexes (Figure 2).

Equation (1) describes the kinetics of double-stranded DNA formation.[32]

$$f_{ds} = \frac{k_2 C_0 t}{1 + k_2 C_0 t} \tag{1}$$

where k_2 is the second-order rate constant, C_0 is the initial concentration of single-stranded DNA segments, and t is time. The fraction of DNA that has formed double strands, f_{ds}, can be calculated from the initial concentration as shown in Equation (2):

$$C_{ds} = f_{ds} C_0 = k_2 C_0^2 t \tag{2}$$

When this equation is applied to SH, two DNA samples must be considered. The first sample contains S- and T-type DNA; the second sample, D, contains only S-type DNA. Since T strands will only form double strands with other T strands the concentration of T : T, denoted C_{Tds}, can be determined from Equation (2). Equation (3) shows:

$$C_{Tds} = \frac{k_2 C_T^2 t}{1 + k_2 C_T t} \tag{3}$$

The formation of S : S, S : D, D : D can be determined by considering first the kinetics of D : D formation and then extracting the amount of S : S formed from the mole fraction of S : S, $X_{S:S}$ (see Equation 4). This can be done under the conditions where the concentration of S is negligible compared to the concentration of D.

$$X_{S:S} = \frac{C_{S1}^2}{C_{D1}^2} \tag{4}$$

where C_{S1} and C_{D1} denote the initial concentrations of S- and D-type strands respectively.

Equation (5) shows

$$(C_{Sds}) = (C_{D1}) \times X_{S:S} = \frac{k_2 C_{D1}^2 t \times C_{S1}^2}{(1 + k_2 C_{D1} t) C_{D1}^2}$$
$$= \frac{k_2 C_{S1}^2 t}{1 + k_2 C_{D1} t} \tag{5}$$

After the hybridization the next step is enrichment by PCR. Only the T : T and S : S strands have matching primers so only those strands will be exponentially amplified. This effectively removes S : D and D : D strands reducing the amount of S contaminants and enriching the amount of T strands. The ratio of enrichment resulting from the first subtraction round is given by Equation (6):

$$E = \frac{(C_{Tds})}{(C_{Sds})} = \frac{(k_2 C_T^2 t)}{(1 + k_2 C_T t)} \times \frac{(1 + k_2 C_{D1} t)}{k_2 C_{S1}^2 t}$$
$$= \frac{(C_T^2) \times (1 + k_2 C_{D1} t)}{(C_{S1}^2) \times (1 + k_2 C_T t)} \tag{6}$$

Since the T and the S strands come from the same genome they have the same concentration. Under this assumption the enrichment effectively simplifies to Equation (7):

$$E = \frac{(C_{Tds})}{(C_{Sds})} = \frac{C_{D1}}{C_T} \tag{7}$$

Therefore the enrichment ratio is increased or the amount of contaminating S strands is decreased simply by using a much higher concentration of D strands relative to T.

4.2 Options

Many variations and ''improvements'' on the SH concept have been published. Despite this, routine success has been elusive, although it is becoming more frequent. The major improvements in the technology included a minimization and simplifying the number of manipulations, use of PCR amplification, and developing more effective protocols through an improved understanding and modeling of the hybridization kinetics of a variety of approaches.

4.2.1 Polymerase Chain Reaction Subtraction

A commercial SH kit sold by Clontech is called ''PCR-Subtraction''. The principles but not the details of this procedure will not be presented here. (Details of this protocol can be found in material available from Clontech.)

This protocol was developed from modeling experiments of a variety of SBH protocols.[33–35] The authors considered several protocols: double-stranded DD and double-stranded SS (+TT); double-stranded DD and complementary single-stranded S (+T); single-stranded D and double-stranded SS (+TT); single-stranded D and complementary single-stranded SS (+TT). The initial modeling experiments focused on cDNAs and a single cycle protocol was developed. This method, called PCR subtraction, allows differential expressed genes to be isolated in a single round of subtraction for complex genomes and genomic differences to be isolated for microorganisms of genomes less than 10 Mb (Figure 3).

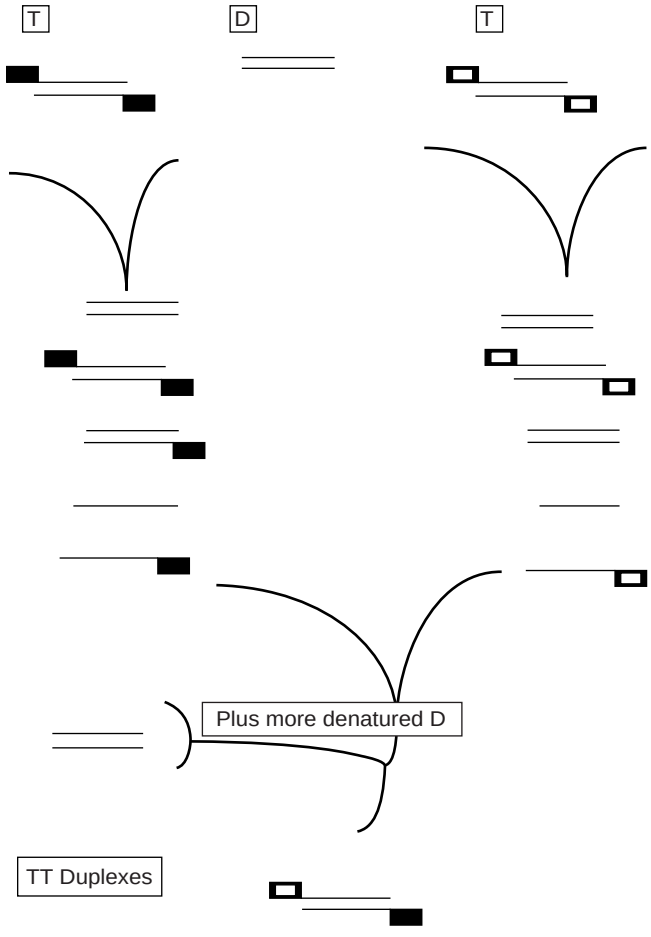

Figure 3 PCR subtraction.

In this protocol, first denatured 3 g D and 0.1 g S (+T) DNAs are annealed together in 1.5 L for 20 h at 68 °C. This allows an approximate equalization of the differentially expressed genes in the single-stranded fraction because of the second-order hybridization kinetics of the different cDNA abundance classes. In practice, two simultaneous initial SH reactions are carried out. The only difference in the reactions is that the S DNAs have different adapter sequences ligated onto their 5′-ends. The 5′-ends do not have a ligated adapter sequence because double-stranded unphosphorylated adapters were used. No adapters are ligated onto the D samples. The two reactions are mixed together and added, undenatured, to a new fraction of denatured D. During the second hybridization, the fragments shared between D and S are reduced further. The single-stranded T fragments in the two reactions will now form a TT with different 5′-end sequences. In a subsequent PCR the ends are filled in. Only duplexes with unmatched ends are amplified because the homoduplexes with complementary adapters self-anneal and preventing annealing of the shorter PCR primers, a phenomenon called suppression PCR.[30] This method takes advantage of rapid annealing of T DNA relative to the remaining S DNA in the single-stranded fraction.

4.3 Trouble Shooting

The major nonobvious difficulties in SH experiments, are similar to those encountered in DD experiments. Many of the problems can be alleviated by careful attention to DNA concentrations. However, a major difficulty is the assessment of DNA concentrations that are extremely critical and may best be handled empirically, especially when getting started.

4.4 Applications

There are many applications of SH experiments and progress is being made in identifying cDNAs differentially expressed. The major problems that need to be resolved are how to detect single base pair differences between samples and how to carry out total genomic subtraction experiments. In fact, more complex protocols have been developed using a complex reduction protocol with a gel electrophoresis step called in-gel competitive hybridization (IGCH)[36–38] and/or a preferential PCR amplification step called representational difference analysis (RDA).[39] In IGCH, excess D and T are mixed together, denatured and then fractionated using high-resolution electrophoresis. Individual size fractions are isolated from the gel and allowed to anneal.

5 OTHER CONSIDERATIONS

Most genome comparative experiments are not done on single cells because of the added difficulty of working with small amounts of DNA and in the case of genomic DNA, two molecules. Hence, the results present an average of several and usually, many cells. It is clear that it would be valuable to develop comparative methodology that allows the analysis of single cells. For instance, the analysis of single cells making up a tissue would provide information on the range of changes occurring in a tissue rather than the average of the changes.

There are several methods that can use the PCR to amplify DNA from low copy, including single molecules. One of the first methods was primer extension pre-amplification (PEP).[40] Reasonable success was reported with a primer pool of length $n = 15$, containing every possible 4^{15} primer. The concentration of any unique species is extremely low, but apparently high enough to insure amplification (e.g. 30 cycles) of most (estimated to be ~80%) of the genome after 50 cycles of PCR. A major

problem with this method is the potential for the primers to interact with each other.

An improvement to this method used a chimeric primer, with a variable portion and a conserved portion. This method was called tagged polymerase chain reaction (T-PCR)[41] because genomic fragments were tagged with a common adapter sequence during the PCR. In T-PCR, each primer had a 5′-end with a constant 17 base sequence and a 3′-end with a variable 9 base sequence. The first few rounds of PCR were done with the chimeric primer, then this primer was removed and the remaining PCR cycles were done with a primer solely composed of the constant region located at the ends of the tagged amplified fragments. A number of variations were developed, such as degenerative oligonucleotide-primed polymerase chain reaction (DOP-PCR).[42] Here the primer contains an internal variable portion (9 bases in length), a long 5′ constant region of 13 bases in length and a short 3′ constant region (3 bases in length). Cheung and Nelson[43] demonstrated that, with a slight modification, DOP-PCR could be used to amplify random genetic markers from small amounts of DNA.

6 PERSPECTIVE

Comparative genomics is a rapidly changing field. Perhaps the most important aspect in the application of a specific technology is adaptability. Not only is it important to stay advised of critical improvements in methods and conceptualization that create quantum improvement in the technology, but it is also important to be able to adapt new technologies as they become available.

The question of how one decides on a specific approach is more difficult. DD does not, in general, provide sequences of interest but rather shows you where they are so they may be isolated. DD provides quantitative information on the number and the extent of differences. The end results of an SH experiment are the sequences of interest. However, SH does not provide information about the number or the level of differences.

ACKNOWLEDGMENTS

This work was supported by grants NIH (1P50 HL55001) and DOA DAMD (17-94-J-414) to CLS.

ABBREVIATIONS AND ACRONYMS

bp	Base Pairs
cDD	cDNA Differential Display
CGH	Comparative Genome Hybridization
DAPI	4′,6-Diamidino-2-phenylindole
DD	Differential Display
DOP-PCR	Degenerative Oligonucleotide-primed Polymerase Chain Reaction
GDD	Genomic Differential Display
IGCH	In-gel Competitive Hybridization
MS	Mass Spectrometry
PCR	Polymerase Chain Reaction
PEP	Primer Extension Pre-amplification
PFG	Pulsed Field Gradient
PSBH	Positional Sequencing by Hybridization
RAPD	Randomly Amplified Polymerase Chain Reaction
RDA	Representational Difference Analysis
SAGE	Serial Analysis of Gene Expression
SBH	Sequencing by Hybridization
SH	Subtractive Hybridization
TcDD	Targeted cDNA Differential Display
TGDD	Targeted Genomic Differential Display
T-PCR	Tagged Polymerase Chain Reaction

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Nucleic Acid Analysis in Clinical Chemistry

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Capillary Electrophoresis of Nucleic Acids • DNA Molecules, Properties and Detection of Single • DNA Probes • Fluorescence In Situ Hybridization • Polymerase Chain Reaction and Other Amplification Systems • Sequencing Strategies and Tactics in DNA and RNA Analysis

REFERENCES

1. S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, 'Correlation Between Protein and mRNA Abundance in Yeast', *Mol.Cell. Biol.*, **19**(3), 1720–1730 (1999).
2. V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, 'Serial Analysis of Gene Expression', *Science*, **270**, 484–487 (1995).
3. A.J. Jeffreys, V. Wilson, S.L. Thein, 'Individual-specific 'Fingerprints' of Human DNA', *Nature*, **316**, 76–79 (1985).

4. J.G. Williams, A.R. Kubelik, K.J. Livak, J.A. Rafalski, S.V. Tingey, 'DNA Polymorphisms Amplified by Arbitrary Primers are Useful as Genetic Markers', *Nucleic Acids Res.*, **18**(22), 6531–6535 (1990).
5. J. Welsh, M. McClelland, 'Fingerprinting Genomes Using PCR with Arbitrary Primers', *Nucleic Acids Res.*, **18**(24), 7113–7218 (1990).
6. D.L. Nelson, S.A. Ledbetter, L. Cordo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbedder, C.T. Caskey, 'Alu Polymerase Chain Reaction: A Method for Rapid Isolation of Human-specific Sequences from Complex DNA Sources', *Proc. Natl. Acad. Sci. USA*, **86**(17), 6686–6690 (1989).
7. P. Liang, A. Pardee (eds.), *Methods in Molecular Biology, Differential Display Methods and Protocols*, Humana Press Inc., Totowa, Vol. 85, 1997.
8. S. Strezoska, T. Pauneska, D. Radosavljevic, I. Labat, R. Drmanac, R. Crkvenjakov, 'DNA Sequencing by Hybridization: 100 Bases Read by a Non-gel-based Method', *Proc. Natl. Acad. Sci. USA*, **88**(22), 10089–10093 (1991).
9. R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, L. Labat, M. Zeremski, J. Snoddy, W.K. Funhouser, B. Koop, L. Hood, R. Crkvenjakov, 'DNA Sequence Determine by Hybridization: A Strategy for Efficient Large-scale Sequencing', *Science*, **260**, 1649–1652 (1993).
10. N.E. Broude, T. Sano, C.L. Smith, C.R. Cantor, 'Enhanced DNA Sequencing by Hybridization', *Proc. Natl. Acad. Sci. USA*, **91**(8), 3072–3076 (1994).
11. H. Koster, K. Tang, D.J. Fu, A. Braun, D. Van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, 'A Strategy for Rapid and Efficient DNA Sequencing by Mass Spectrometry', *Natl. Biotechnol.*, **14**(9), 1123–1128 (1996).
12. L. Wodicka, H. Dong, M. Mittmann, M.H. Ho, D.J. Lockhart, 'Genome-wide Expression Monitoring in *Saccharomyces cerevisiae*', *Natl. Biotechnol.*, **15**(13), 1359–1367 (1997).
13. D.J. Duggan, M. Bittner, Y. Chen, P. Meltzer, J.M. Trent, 'Expression Profiling Using cDNA Microassays', *Natl. Genet.*, **21**(1Suppl.), 10–14 (1999).
14. D. Grotheus, C.R. Cantor, C.L. Smith, 'Top-down Construction of an Ordered Schizosaccharomyces *pombe* Cosmid Library', *Proc. Natl. Acad. Sci. USA*, **91**(10), 4461–4465 (1994).
15. C.L. Smith et al., 'Cloneless Libraries: Long Range Mapping of Chromosome 20 and cDNA Selection for the 20q13 Amplified in Cancer Cells', in preparation.
16. A. Kallioniemi, O.P. Kallioniemi, D. Sudar, D. Rutovita, J.W. Grey, F. Waldman, D. Pinkel, 'Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors', *Science*, **258**, 818–821 (1992).
17. D. Shalon, S.J. Smith, P.O. Brown, 'A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization', *Genome Res.*, **6**(7), 639–645 (1996).
18. P. Liang, A.B. Pardee, 'Differential Display of Eukaryotic Messenger RNA by Means of the Polymerase Chain Reaction', *Science*, **257**, 967–971 (1992).
19. D. Bauer, H. Muller, J. Reich, H. Riedl, V. Ahrenkiel, P. Warthoe, M. Strauss, 'Identification of Differentially Expressed mRNA Species by an Improved Display Technique (DDRT-PCR)', *Nucleic Acids Res.*, **21**(18), 4272–4280 (1993).
20. C.L. Smith, S. Klco, T.Y. Zhang, H. Fang, R. Oliva, J.B. Fan, M. Bremer, S. Lawrance, 'Analysis of Megabase DNA Using Pulsed Field Gel Electrophoresis', in *Methods in Molecular Genetics*, ed. K.W. Adolph, Academic Press, San Diego, 155–196, 1993.
21. N. Broude, A. Chandra, C.L. Smith, 'Differential Display of Genome Subsets Containing Specific Interspersed Repeats', *Proc. Natl. Acad. Sci. USA*, **94**(9), 4548–4553 (1997).
22. N. Broude, N. Storm, S. Malpel, J.H. Graber, S. Lykyanov, E. Sverdlov, C.L. Smith, 'PCR Based Targeted Genomic and cDNA Differential Display', *Genet. Anal.*, **15**(2), 51–63 (1999).
23. R.P. Oliveria, N.E. Broude, A.M. Macedo, C.R. Cantor, C.L. Smith, S.D.J. Pena, 'Probing the Genetic Population Structure of *Trypanosoma cruzi* with Polymorphic Microsatellites', *Proc. Natl. Acad. Sci. USA*, **95**(7), 3776–3780 (1998).
24. I. Lavrentieva, N. Broude, Y. Lebedev, I.I. Gottesman, S.A. Lukyanov, L. Smith, 'High Polymorphism Level of Genomic Sequences Flanking Insertion Sites of Human Endogenous Retroviral Long Terminal Repeats', *FEBS Lett.*, **443**(3), 341–347 (1999).
25. J. Bouchard, C. Foulon, G. Nguyen, N. Storm, C.L. Smith, 'Genomic Discordance in Mongozygotic Twins', in *Techniques in the Behavioral and Neural Sciences*, eds. W. Crusio, R. Gerlai, Elsevier, Amsterdam, 237–256, 1999.
26. R.P. Kandpal, G. Kandpal, S.M. Weissman, 'Construction of Libraries Enriched for Sequence Repeats and Jumping Clones, and Hybridization Selection for Region-specific Markers', *Proc. Natl. Acad. Sci. USA*, **91**(1), 88–92 (1994).
27. B. Stone, W. Wharton, 'Targeted RNA Fingerprinting: the Cloning of Differentially-expressed cDNA Fragments Enriched for Members of the Zinc Finger Gene Family', *Nucleic Acids Res.*, **22**(3), 2612–2618 (1994).
28. D. Graf, A.G. Fisher, M. Merkenschlager, 'Rational Primer Design Greatly Improves Differential Display-PCR (DD-PCR)', *Nucleic Acids Res.*, **25**(11), 2239–2240 (1997).
29. J. Bouchard, N. Storm, C.L. Smith, 'Enhancing Targeting Genomic and cDNA Differential Display Using a *Saccharomyces cerevisiae* Model System', in preparation.
30. P.D. Siebert, A. Chenchik, D.E. Kellog, K.A. Lukyanov, S.A. Lukyanov, 'An Improved PCR Method for Walking in Uncloned Genomic DNA', *Nucleic Acids Res.*, **23**(6), 1087–1088 (1995).

31. D.J. Munroe, M. Haas, E. Bric, T. Whitton, H. Aburatani, K. Hunter, D. Ward, D.E. Housman, 'IRE-Bubble PCR: A Rapid Method for Efficient and Representative Amplification of Human Genomic DNA Sequences from Complex Sources', *Genomics*, **19**(3), 506–514 (1994).
32. C.R. Cantor, C.L. Smith, *Genomics: The Science and Technology Behind the Human Genome Project*, Wiley & Sons, New York, 1999.
33. O.D. Ermolaeva, E.D. Sverdlov, 'Subtractive Hybridization, a Technique for Extraction of DNA Sequences Distinguishing Two Closely Related Genomes: Critical Analysis', *Genet. Anal.*, **13**(2), 49–58 (1996).
34. O.D. Ermolaeva, S.A. Lukyanov, E.D. Sverdlov, 'The Mathematical Model of Subtractive Hybridization and its Practical Application', *Ismb.*, **4**, 52–58 (1996).
35. N.G. Gurskaya, L. Diatechenko, A. Chencik, P.D. Siebert, G.L. Khaspekov, L.A. Lukyanov, L.L. Vagner, O.D. Ermolaeva, S.A. Lukyanov, E.D. Sverdlov, 'Equalizing cDNA Subtraction Based on Selective Suppression of Polymerase Chain Reaction: Cloning of Jurkat Cell Transcripts Induced by Phytoghemaglutinin and Phorbol 12-Myristate 13-Acetate', *Anal. Chem.*, **240**(1), 90–97 (1996).
36. H. Yokota, T. Iwasaki, M. Takahasi, M. Oishi, 'A Tissue-specific Change in Repetitive DNA in Rats', *Proc. Natl. Acad. Sci. USA*, **86**(23), 9233–9237 (1989).
37. H. Yokota, M. Oishi, 'Differential Cloning of Genomic DNA: Cloning of DNA with an Altered Primary Structure In-gel Competitive Reassociation', *Proc. Natl. Acad. Sci. USA*, **87**(16), 6398–6402 (1990).
38. H. Yokota, S. Amano, T. Yamane, K. Ataka, E. Kikyua, M. Oishi, 'A Differential Cloning Procedure of Complex Genomic DNA Fragments', *Anal. Biochem.*, **219**(1), 131–138 (1994).
39. N.A. Lisitsyn, N. Lisitsyn, M. Wigler, 'Cloning the Difference Between Two Complex Genomes', *Science*, **259**(5097), 946–951 (1993).
40. L. Zhang, X. Cui, K. Schmitt, R. Hubert, W. Navidi, N. Arnheim, 'Whole Genome Amplification from a Single Cell: Implications for Genetic Analysis', *Proc. Natl. Acad. Sci. USA*, **89**(13), 5847–5851 (1992).
41. D. Grothues, C.R. Cantor, C.L. Smith, 'PCR Amplification of Megabase DNA with Tagged Random Primers (T-PCR)', *Nucleic Acids Res.*, **21**(5), 1321–1322 (1993).
42. H. Telenius, N.P. Carter, C.E. Bebb, M. Nordenskjold, B.A. Ponder, A. Tunnacliffe, 'Degenerate Oligonucleotide Primed PCR: General Amplification of Target DNA by a Single Degenerate Primer', *Genomics*, **13**(3), 718–725 (1992).
43. V.G. Cheung, S.F. Nelson, 'Whole Genome Amplification Using a Degenerate Oligonucleotide Primer Allows Hundreds of Genotypes to be Performed on Less than One Nanogram of Genomic DNA', *Proc. Natl. Acad. Sci. USA*, **93**, 14 676–14 679 (1966).

DNA Molecules, Properties and Detection of Single

Zhifeng Long
Genetic Therapy, Inc., A Norvartis Company, Gaithersburg, USA

1 Introduction 1
2 Current Technologies of DNA Detection 1
2.1 Principles 1
2.2 Methods, Instrumentation and Applications 2
3 Parameters Affecting Detection Sensitivity 2
3.1 Definition of Sensitivity 2
3.2 Detection Method 2
3.3 Size of DNA 3
3.4 Assay Optimization 3
3.5 Levels of Noise or Impurities of the Test Matrix 5
4 Statistical Assessment of Limit of Detection: Properties of Single 5
4.1 Limit of Detection 5
4.2 Theoretical Limit 5
4.3 Properties of Single 6
4.4 Beyond the Limit 7
5 The Pros and Cons of DNA Amplification 7
Abbreviations and Acronyms 8
Related Articles 8
References 8

Technological advances have created more DNA detection systems with more power than ever before. By amplifying the target or its molecular counterpart, the probe, very minute amounts of DNA can be detected or even sequenced. Different technologies give rise to different DNA detection systems. Each system has its merits and limitation. In many applications, detection sensitivity is the parameter of interest. This article first gives an outline of different DNA detection systems and their sensitivities and limits. It then describes how to achieve and prove one-copy sensitivity. Finally, it discusses the testing strategy, supported by statistics to increase detection confidence. Although the limit of detection (LOD) can be improved with better technology and instrument, the highest sensitivity is set by probability laws. To detect one-copy of a DNA molecule, one needs to understand the properties of single.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The ability to detect deoxyribonucleic acid (DNA) molecules is crucial to the development of molecular biology and today's many analytical instruments for biological applications. From the beginning of the molecular biology era, the search for better ways to detect DNA has never stopped. Such endeavor has led to enormous advances in our understanding of DNA – the basic building block of life. It was the improvement in X-ray diffraction that generated the crystallographic data for DNA and formed the basis of the classical discovery of the double helix structure by Watson and Crick.[1] It was the DNA polymerase chain terminator sequencing method that allowed us to read the genetic make-up of life and revolutionized the field of automated sequencing.[2] It is the polymerase chain reaction (PCR) that makes DNA detection, cloning and sequencing so easy and so powerful.[3–6]

It used to take days to detect a target DNA by restriction enzyme digestion, gel electrophoresis and Southern blot hybridization.[7] Now this can be done in a matter of hours in a tiny, single tube. The thermally stable *Taq* DNA polymerase isolated from *Thermus aquaticus* can amplify DNA more easily by cycling through the temperatures of denaturing and annealing than billions of *Escherichia coli* bacteria used to do in conventional DNA cloning.

2 CURRENT TECHNOLOGIES OF DNA DETECTION

2.1 Principles

Nucleic acid probes interact with their complementary target sequences mainly through H-bonding. We call such an interaction molecular hybridization. Molecular hybridization between probe and target sequences is the basis for most current technologies of DNA detection. Mass spectrometry (MS), which detects DNA by separating ionized molecules according to their ratios of mass to electric charge, is an exception. There are direct and indirect detection methods. Direct detection usually involves an excessive amount of target DNA and specialized instruments or methods that detect DNA without modifying the original quantity. In many applications, however, the limited amount of target DNA prohibits direct detection. In this case, indirect detection involving either probe or target DNA amplification can be used to enhance the signal of the probe or target.

2.2 Methods, Instrumentation and Applications

DNA can be detected directly by a number of methods and instruments. Using a spectrophotometer with ultraviolet (UV) capability to measure optical absorbance (OD) at 260 nm, it is possible to detect DNA in aqueous solution at a concentration of 250 ng mL^{-1} and above. For accurate measurement, however, a concentration of 1 μg mL^{-1} or higher is preferred. Ethidium bromide (ETBr) forms a noncovalent complex with double-stranded DNA by intercalating between stacked base pairs. Upon UV illumination, the DNA–ETBr complex emits an orange fluorescence which allows the detection of DNA. This is usually used in conjunction with agarose or polyacrylamide gel electrophoresis (PAGE). DNA is first separated based on size by gel electrophoresis, then the DNA in the gel is visualized by staining with ETBr and illuminated with a UV transilluminator. Photographs of stained gels showing DNA bands are often taken with a Polaroid camera. As little as 10 ng of a 300 base pair (bp) DNA can be detected in this way. Using the molecular hybridization technique, one can transfer the DNA bands from agarose gels to a solid membrane and use a DNA as a probe to detect another DNA that has a homologous sequence.[7]

Both DNA and RNA molecules can be detected by molecular hybridization. Nucleic acid samples can also be spotted directly on a solid membrane and such a method is called dot blot or slot blot, depending on the way in which the sample is applied to the membrane. The probe can be labeled radioactively or nonradioactively, providing a means for detection following molecular hybridization. The molecular hybrid is visualized with the aid of an X-ray film when radioactive or chemiluminescence labeling is used, or by colorimetric detection of biotin/alkaline phosphatase–streptavidin conjugate.[8] As little as 10^6 copies of a DNA or RNA sequence can be detected by molecular hybridization methods. Subpicomolar levels of DNA can be detected directly with a mass spectrometer. However, it cannot distinguish a target-specific sequence in a sample containing mixed DNA molecules.

In addition to direct detection of DNA, there are indirect detection methods that involve amplification of either the target DNA or its probe. The most widely used target amplification technology is the PCR.[3–6] It is possible to detect one DNA molecule by PCR amplification.[6] Among a number of probe amplification techniques, the branched deoxyribonucleic acid (bDNA) system is the most successful.[9] It is possible to detect as few as 50 molecules with bDNA technology.[10]

Molecular hybridization has been widely used in screening DNA libraries in colony and phage plaque hybridization in recombinant DNA laboratories.[11] Traditional recombinant DNA applications include bacterial colony and phage plaque screening, plasmid DNA screening by agarose gel and/or Southern blot hybridization. Recent additions include in situ hybridization and fluorescence in situ hybridization (FISH), microarray or gene chip technology, DNA fingerprinting and its forensic applications. There are commercial kits based on DNA hybridization, including diagnostic kits for viral infection, cancers and inherited genetic mutations. More sophisticated diagnostic tools have entered the market in recent years, which involve amplification techniques such as PCR and bDNA. These usually require special equipment and/or expertise. DNA sequencing technology has now become a basic tool, not only for molecular biology, but also for many far-reaching projects such as the genome projects of human and other species. It is has also been used as a detection method in genotyping and pharmacogenetics.

3 PARAMETERS AFFECTING DETECTION SENSITIVITY

The methodology, the size and concentration of the target DNA and the levels of complexity and impurity of the test matrix can all affect the sensitivity of a detection system.

3.1 Definition of Sensitivity

A clear definition of sensitivity is essential for comparing one assay with another. A sensitivity unit should define both the lowest concentration of the target and the level of impurities in a test matrix in which the target is present. Often the target is a short stretch of DNA sequence and the test matrix is an aqueous solution of DNA of high complexity, such as genomic DNA. Thus, detecting one copy of the target DNA in 10^6 copies of genome of an organism is more sensitive than detecting one copy of the target DNA in 10^3 copies of the same genome because the former has a higher level of impurities than the latter.

3.2 Detection Method

The methodology often presets the upper limit of sensitivity of a detection system. The spectrophotometric method using OD at 260 nm detects only down to 250 ng mL^{-1} DNA. MS can detect subpicomolar levels of DNA, or a few hundred molecules in 0.1 μL, since this technique can use very little amount of sample. Although quick and direct if one has access to the equipment for these two techniques, neither of them can detect a sequence-specific target in a sample containing different DNA molecules. Agarose gel electrophoresis or PAGE can separate different DNA molecules by size,

and detect as little as 10 ng of DNA by ETBr staining. Combining agarose gel electrophoresis with Southern blot hybridization will increase the detection sensitivity to 10^6 molecules with very high specificity provided by the molecular hybridization with a sequence-specific probe.

To compare different methods more easily, we shall convert the mass of DNA into molar concentration. Assuming that the DNA target is a 300 bp fragment, it is interesting to find that 10 ng contains 3×10^{10} molecules. The size of DNA is in reciprocal relation to the number of molecules; for the same mass, the larger the DNA fragment, the smaller is the number of molecules.

The sensitivity and specificity of an indirect, amplification-based detection system is usually higher than that of a direct DNA detection method. The bDNA technique can detect as little as 50 molecules of a target-specific sequence in a sample containing different sequences. PCR can detect as little as one molecule of a target-specific sequence in 500 000 human cells or 10^6 copies of human haploid genome.[12] This sensitivity, if calculated for the mass ratio, is equivalent to detecting 1 in 7.2×10^{12} equivalent copies of DNA (the amplified PCR product is 417 bp. Since there are two genomes per diploid human cell and approximately 3×10^9 bp per genome, 500 000 human cells contain $2 \times 5 \times 10^5 \times 3 \times 10^9$ bp DNA. Therefore, 417 bp/$(2 \times 5 \times 10^5 \times 3 \times 10^9)$ bp $= 1/7.2 \times 10^{12}$, or 1 per 7.2×10^{12} equivalent copies of DNA). In this respect, PCR is the most sensitive and specific detection method in analytical chemistry.

3.3 Size of DNA

Depending on the method of the assay, the size of the target DNA may or may not affect the detection sensitivity. Spectrophotometry and agarose gel electrophoresis may detect 250 ng mL^{-1} and 10 ng per band, respectively, regardless of the size of DNA. Since the size of DNA is in a reciprocal relation to the number of molecules for the same amount of DNA, these methods have a higher sensitivity in detecting larger size DNA when the sensitivity is expressed as the smallest number of molecules detected. On the other hand, PCR is less sensitive in detecting large-size DNA because it is less efficient in amplifying larger DNA fragments. In detection systems based on the TaqMan® chemistry described below, it is desirable to choose a product size of 100–400 bp for optimum PCR amplification. However, it has been shown that PCR can detect one copy of a 1.4 kilobase (kb) DNA sequence in 10^6 copies of the human haploid genome, using the conventional gel electrophoresis and Southern blot hybridization method.[13]

3.4 Assay Optimization

It is relatively easy to optimize the detection conditions for direct detection methods. In spectrophotometry, an instrument that uses the least sample volume gives the highest sensitivity. Some current models of spectrophotometers can measure OD in a sample volume as small as 50 μL. Some models use a 96-well format quartz microtiter plate, offering both high throughput and sensitivity. A 50 μL volume of DNA at 250 ng mL^{-1} concentration contains only 12.5 ng of DNA. This sensitivity is equivalent to detection by agarose gel electrophoresis, although a more accurate OD measurement requires a DNA concentration of 1000 ng mL^{-1} or 50 ng of DNA in 50 μL.

PCR assay optimization involves both the amplification and postamplification steps. In assays that use agarose gel electrophoresis to detect PCR-amplified products, it is important to include the limitation of detection by gel electrophoresis. The smallest amount of DNA that can be detected in an ETBr-stained agarose gel is approximately 10 ng. As shown above, 10 ng of a 300-bp DNA fragment contains 3×10^{10} molecules. It requires a minimum of 37 cycles of PCR to amplify one molecule to 3×10^{10} molecules (assume 100% amplification efficiency. Let x be the number of PCR cycles required to increase the target sequence by 3×10^{10} fold, then we have $2^x = 3 \times 10^{10}$. Solve the equation to obtain $x \approx 35$. Since it takes two cycles to generate the first double-stranded PCR product bracketed by two PCR primers, it requires $n = 35 + 2 = 37$ cycles of PCR to amplify one molecule to 3×10^{10} molecules). This is very important information, because it tells us that unless we run 37 cycles of PCR we should not expect to see the target band in an agarose gel. Very often it is necessary to run more than 37 cycles to achieve one-copy sensitivity, because the amplification is rarely 100% efficient.

3.4.1 Real-time Polymerase Chain Reaction Detection of a Single DNA Molecule: One Example

The fluorescence detection of PCR product using the 7700 Sequence Detection System from PE Biosystems (Foster City, CA, USA) allows real-time detection and quantitation of DNA sequences without the need for labor-intensive processes such as gel electrophoresis and radioactive hybridization.[14,15] The 7700 instrument has a built-in 96-well format thermal cycler. The detection is based on the TaqMan® chemistry. The method uses a fluorescence oligonucleotide probe with a 5′ reporter dye and downstream 3′ quencher dye.[16] During PCR, the $5' \rightarrow 3'$ exonuclease activity of the *Taq* DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector

of the 7700 instrument. Using this system, a 417-bp sequence of the *Herpes simplex* virus thymidine kinase gene (HSV-Tk) was amplified using primers 5′-CGAGACAATCGCGAACATCTAC-3′ (forward) and 5′-GCCAGCATAGCCAGGTCAAG-3′ (reverse) and a TaqMan® probe, 5′-FAM-CCGGCACAAACATC-GTGT-TAMRA-3′, that contains a covalently linked reporter dye FAM at the 5′-end and quencher dye TAMRA at the 3′-end. It has been shown that as little as one copy of the HSV-Tk gene can be detected in a background of 3.3 µg of genomic DNA, which is equivalent to 500 000 human cells.[12]

As shown in Figure 1(a–c), detection of one copy of DNA requires more than 45 cycles of PCR amplification using the PE 7700 system. To evaluate the results obtained with this real-time PCR, the amplified products were analyzed on agarose gels, followed by Southern blot hybridization. The presence or absence of hybridization for all of the reactions matched completely the results from the 7700. Differences between the intensities in

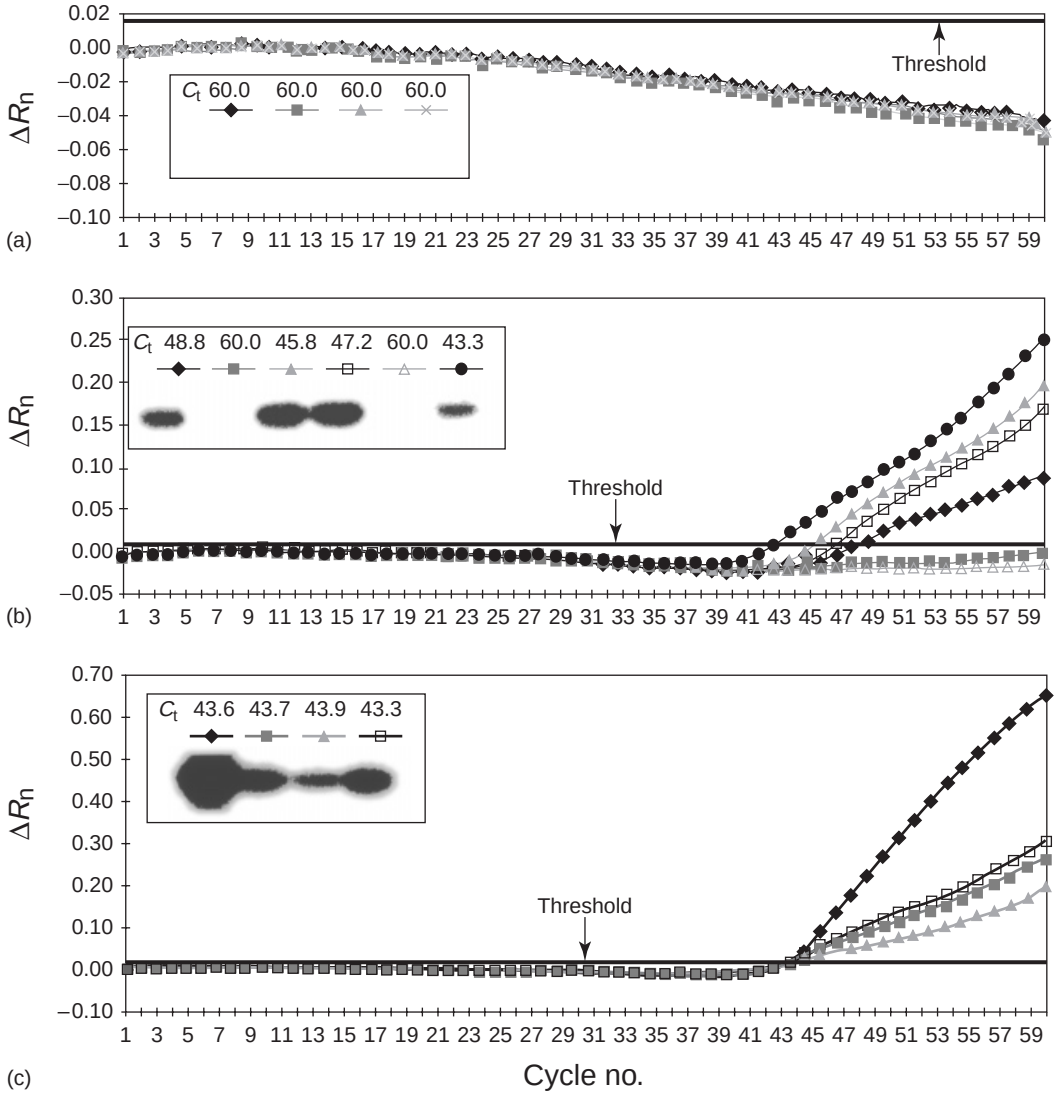

Figure 1 Amplification plots and hybridization results for (a) 0-, (b) 1- and (c) 10-copy of targets using the PE 7700 sequence detection system. Relative normalized fluorescence change (ΔR_n) is plotted vs time (PCR cycle number). (a) Four replicates of 3.3 µg of negative DNA (0 target); (b) six replicates of one target per 3.3 µg of DNA; (c) four replicates of 10 targets per 3.3 µg of DNA. The horizontal solid line spanning each plot depicts the threshold fluorescence intensity (0.013). Any sample that reached a fluorescence value that exceeded the threshold was considered to be positive, and the cycle at which this first occurred was defined as the threshold cycle (C_t). A sample for which no amplification occurred is given a C_t value of the total number of PCR cycles (60) performed. The isotopic Southern blot hybridization results of the amplified products at the end of the 60 cycles are shown in the insets, with the C_t values given. (Reproduced by permission of Eaton Publishing from Lockey et al.[12])

hybridization signals and the threshold cycle (C_t) values were probably caused by nonexponential amplification in the later cycles of PCR after the threshold. At a concentration of 10 copies per sample the C_t values of different samples were very close (43.3–43.9), but the intensities of hybridization signals varied significantly. This suggests that the use of C_t values for quantitation is more accurate than analyzing the product yields at the end of all PCR cycles. The former measures at the logarithmic phase of amplification. C_t values vary widely in the replicates of the one-copy samples. This suggests that quantitation is no longer accurate at such a low concentration. In other words, it is possible to detect one copy of DNA, but not possible to quantitate it by PCR. Unlike the detection of DNA in samples containing 10 copies of target molecules, detection of DNA in samples containing one copy of the target is not a 100% event. This will be discussed further later.

With proper assay development and validation, the PE 7700 automated amplification and detection system is able to achieve reproducibly one-copy sensitivity without the need for a time-consuming assay development process and highly experienced staff to analyze PCR products, making it an ideal system for clinicians and researchers who demand a high-throughput PCR assay. The method has been adapted to detect infection in patients treated with living pig tissues.[17] With the inclusion of a calibration curve based on known concentrations, it can be used for both high-sensitivity end-point detection and quantitation within one test.[17]

3.5 Levels of Noise or Impurities of the Test Matrix

The presence of impurities in a test sample may affect the specificity and sensitivity of the detection system. The OD method using spectrophotometry is greatly limited when impurities exist in similar amount to the target DNA, even if the impurities are not nucleic acids. The OD method is unable to detect the target sequence in the presence of other nucleic acids. For detection systems based on molecular hybridization, including PCR, it is possible to detect a target sequence in the presence of high levels of impurities and other nucleic acids.

During PCR, impurities of high concentrations of either nucleic acids or proteins may significantly reduce the detection sensitivity. It is possible to detect one copy of a target sequence in the presence of 6.6 μg of genomic DNA, which is equivalent to a DNA content of 10^6 mammalian cells in a reaction volume of 100 μL (Z. Long, unpublished data). When the reaction volume is reduced to 50 μL for the same PCR assay, the sensitivity is no longer one copy. The sensitivity can be restored to one copy when the nonspecific DNA level is reduced to 3.3 μg in a 50-μL reaction volume. Similar inhibitory effects on PCR have been observed in the presence of high concentrations of RNA, even when the target is DNA.

4 STATISTICAL ASSESSMENT OF LIMIT OF DETECTION: PROPERTIES OF SINGLE

4.1 Limit of Detection

The LOD is the lowest concentration of an analyte detectable in a sample by a well-defined assay with an acceptable detection confidence. Note that the level of analyte is not necessarily quantitated. Different assay systems may have very different LODs, and these have been described above. The LOD of a detection system is usually equal to the sensitivity of the system, except when the target copy numbers are low. For example, when the target copy number is one per sample and the detection sensitivity is one copy, the LOD may not equal one copy if only one replicate of the sample is tested because the detection confidence is only 63%. A confidence of less than 95% is usually unacceptable for statistical reasons. However, one can design a test strategy to analyze multiple replicates of a sample to achieve a one-copy LOD.

4.2 Theoretical Limit

Any detection method has a limit. It is important to understand whether the limit is caused by the instrumentation/methodology or by something more fundamental such as the probability laws. When an assay has demonstrated a sensitivity of one copy in an appropriate assay validation, the LOD has reached the theoretical limit. At the theoretical limit, detecting or not detecting the target is no longer a definite event, but an issue of probability or likelihood.

According to the Poisson probability distribution, if samples taken from a homogeneous solution have a concentration of one target molecule per sample volume, 36.79% of the samples will contain no target at all, and 63.21% will contain one or more targets (Table 1). Of the 63.21% positive samples, a good proportion of samples will contain two or more targets instead of one. As the target concentration increases, the chance of obtaining a positive sample also increases. To be almost certain that each sample will contain at least one target, the concentration has to be greater than nine targets per sample.

Theoretically, one can detect one target in a sample obtained from a solution that has a concentration of less than one target per sample because there is a chance that a sample may contain one target. However, one

Table 1 The Poisson probability distribution – average number of targets versus probability of a positive sample

Average number of targets	Probability of containing no target (%)	Probability of containing one or more targets (%)
1	36.79	63.21
2	13.53	86.47
3	4.98	95.02
4	1.83	98.17
5	0.67	99.33
6	0.25	99.75
7	0.09	99.91
8	0.03	99.97
9	0.01	99.99
10	<0.01	>99.99

cannot detect less than one target in a given sample. In a given sample, the number of targets follows a discrete distribution of whole number with a minimum of zero, but the concentrations of target in solutions are continuous. It is important to understand such a difference.

4.3 Properties of Single

The sensitivity of a detection system has reached its maximum when the assay has demonstrated its ability to detect one copy of DNA. To prove such a sensitivity, one needs to perform a series of tests using a number of replicate samples obtained from a homogeneous solution containing one copy of target DNA per sample volume. It may also be done with a two- or three-copy concentration. However, a higher copy concentration requires a larger number of replicate samples. The results from the tests can be analyzed by a Yates-corrected chi-square test, as described previously.[12] It is preferred that both the number of positive results and the number of negative results are ≥ 5 in performing the statistical test, in order to avoid misrepresentation of data. Theoretically, a higher target concentration corresponds to a smaller chance of obtaining a negative result. To reduce the sample size required to generate ≥ 5 negative and also ≥ 5 positive results, it is important to choose a concentration that has a similar probability of producing a balanced number of positive and negative results.

Although there is little that one can do to improve the sensitivity of a detection system once it has been optimized to detect one copy of DNA target in the appropriate test matrix, there is a lot one can do in designing the appropriate testing strategy to improve the probability of detection. The detection confidence of an assay is the probability of detecting a positive sample in one or more tests. At a concentration of one copy of target DNA per unit sample volume, the detection confidence can be increased to 95% by running three replicates of the test per sample. It can be further improved to a 98% detection confidence by using four replicates of the test (Figure 2a).

The relationships between the detection confidence and the number of replicates per test sample when the target concentrations are two and three copies per unit sample volume, are illustrated in Figure 2(b) and 2(c), respectively. For example, a 98% detection confidence can be obtained with only two replicates when the target concentration is two copies per unit sample volume. Up to 99.99% detection confidence is attainable with three replicates when the target concentration is three copies per unit sample volume. The probabilities of

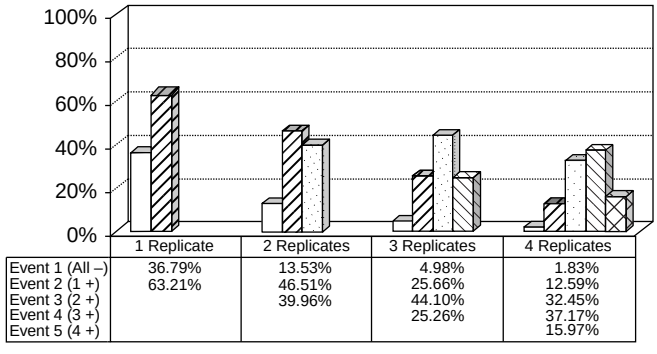

(a)

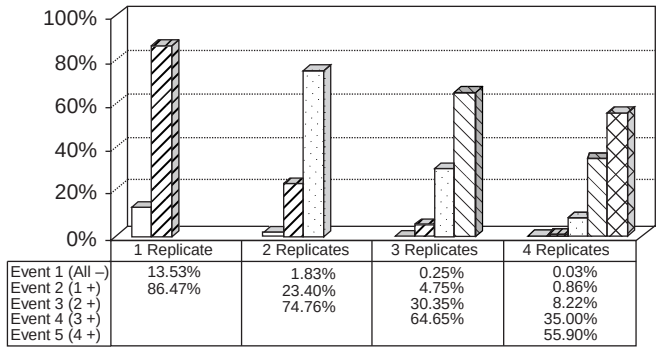

(b)

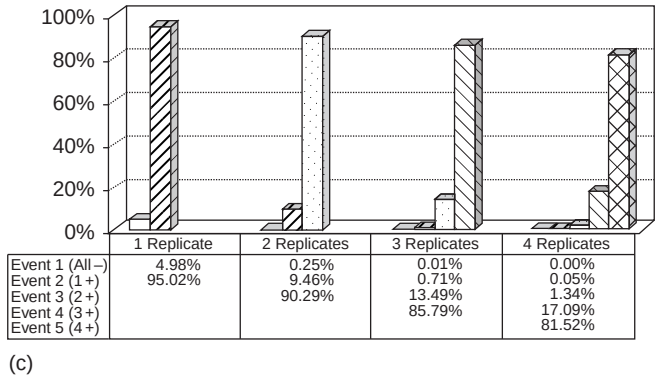

(c)

Figure 2 Probability of detection as a function of target numbers and test replicates per sample. (a) One, (b) two and (c) three targets per sample.

different events with different combinations of positive and negative results are indicated in Figure 2(a–c). One can use these probabilities to design an appropriate testing strategy to optimize data collection and interpretation and minimize waste of resources. A conclusion that can be drawn from the analysis is that it must test multiple replicates of sample in order to achieve a statistically acceptable detection confidence of detecting one copy of the target or less in a sample. Conversely, as shown in Table 1, if one can only afford to run one replicate per test sample and want to have a 99.99% confidence of detection, one can choose to lower the claimed LOD of the assay to 9–10 copies. It is important to realize that this can only be done when the detection system has already demonstrated a one-copy sensitivity. It is the claimed LOD that is reduced, not the sensitivity.

4.4 Beyond the Limit

Is it possible to go below one copy of detection? Is there a limit to detection at all, knowing that everything is probable?

For low levels of target concentration (less than one copy per sample), according to the Poisson probability distribution [Equation 1]:

$$P = 1 - e^{-x} \tag{1}$$

where P is the probability of detection and x is the number of targets per sample (concentration). It is not difficult to find that there is a good chance of obtaining a positive sample even if the concentration is below one copy. For example, when $x = 0.1$, the probability of obtaining a positive sample is 9.5%. This is not a bad chance. However, to ensure a 99.0% detection confidence when $x = 0.1$, one must test at least 46 replicates of each sample (let the sample size be N, then $(e^{-0.1})^N = 1 - 0.99$; solve the equation to obtain $N \approx 46$).

Hence the answer to the second question is that the resources are the limit to achieving less than one-copy sensitivity. Because of the cost and availability of materials, it is not practical to go beyond one-copy detection sensitivity. As discussed in section 3.4, it is possible to detect low levels (one copy or less) of DNA, but it is not possible to perform quantitation at similar levels. Hence detection can go beyond the limit but quantitation can not.

5 THE PROS AND CONS OF DNA AMPLIFICATION

The major advantage of DNA amplification is its unsurpassed sensitivity. Better than finding a needle in a haystack, PCR has proved to be extremely sensitive and reliable, once optimized and validated. However, because of its great sensitivity, there is a misconception among beginners that the sensitivity is the cause of false-positive results. PCR has been known to have a high propensity to generate false-positive results. There are many publications on how to control the problem of PCR contamination.[18–25] Opposition to using PCR in critical areas such as diagnostic/forensic studies and quality control testing argues that PCR is inherited with contamination, so that applications of PCR are bound with faults. Some users even set 30 cycles as the maximum number of PCR cycles in order to avoid contamination. However, one should realize that it is neither the number of PCR cycles nor the sensitivity itself that causes false positives, but the laboratory environment, sample cross-contamination, contaminated equipment, reagents or poor techniques of the operator. It requires a minimum number of cycles (>30) and optimization to achieve one-copy sensitivity.[12,13] Such sensitivity can be obtained without contamination if proper procedures are implemented and strictly followed by the operator.

To control false-positive results, it is necessary to implement physical barriers, proper air pressure and separate air handlers between the areas for clean reagents and source DNA materials and between pre-PCR and postamplification processes. Clean protective clothing including face-masks, and hair and shoe covers is highly recommended. Setting up proper sentinel and PCR reagent controls is essential to demonstrate contamination-free reagents and laboratory environment. Because both amplification and detection processes are contained within a closed system in the PE 7700 instrument, it minimizes the potential for product carryover contamination. The use of uracil N-glycosylase (UNG) prior to PCR further reduces false positives by digesting previously amplified DNA products that contain uracil instead of thymine and are therefore subjected to UNG digestion.

Accurate results are critical for diagnostics, forensics and quality control testing of biological samples. This applies to any detection system. A false-negative result could result in transmitting an infectious virus to patients who used a failed product. A false-positive result could mean a suspension of clinical trials or a halt in manufacture of a product. All these are very costly to the sponsor. The principle of the statistical considerations and the PCR assay optimization strategy described here can save these troubles. A properly optimized and executed PCR test can provide very high sensitivity, comparable to and sometimes even better than the time-consuming infectivity culture assays. The distinct characteristics of single target detection are not limited to PCR and may

be applied to other detection systems that require single-copy sensitivity.

ABBREVIATIONS AND ACRONYMS

bDNA	Branched Deoxyribonucleic Acid
ETBr	Ethidium Bromide
FISH	Fluorescence In Situ Hybridization
HSV-Tk	*Herpes Simplex* Virus Thymidine Kinase Gene
LOD	Limit of Detection
MS	Mass Spectrometry
OD	Optical Absorbance
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
UNG	Uracil *N*-Glycosylase
UV	Ultraviolet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry • Pharmacogenetic Testing • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids • DNA Probes • Fluorescence In Situ Hybridization • Mass Spectrometry of Nucleic Acids • Polymerase Chain Reaction and Other Amplification Systems • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Sequencing Strategies and Tactics in DNA and RNA Analysis • X-ray Structures of Nucleic Acids

Mass Spectrometry **(Volume 13)**
Time-of-flight Mass Spectrometry

REFERENCES

1. J.D. Watson, F.H.C. Crick, 'Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid', *Nature (London)*, Issue 4356 (1953); also *Nature*, **248**, 765 (1974).
2. F. Sanger, S. Nicklen, A.R. Coulson, 'DNA Sequencing with Chain Terminating Inhibitors', *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467 (1977).
3. K.B. Mullis, 'Process for Amplifying Nucleic Acid Sequences', US Patent, 4 683 202, 1987.
4. K.B. Mullis, H.A. Erlich, N. Arnheim, G.T. Horn, R.K. Saiki, S.J. Scharf, 'Process for Amplifying, Detecting and/or Cloning Nucleic Acid Sequences', US Patent, 4 683 195, 1987.
5. K.B. Mullis, F.A. Faloona, 'Specific Synthesis of DNA In Vitro via a Polymerase Catalyzed Chain Reaction', *Methods Enzymol.*, **155**, 335–350 (1987).
6. R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich, 'Primer-directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase', *Science*, **239**, 487–491 (1988).
7. E.M. Southern, 'Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis', *J. Mol. Biol.*, **98**, 503–517 (1975).
8. P.R. Langer, A.A. Waldrop, D.C. Ward, 'Enzymatic Synthesis of Biotin-labeled Polynucleotides: Novel Nucleic Acid Affinity Probes', *Proc. Natl. Acad. Sci. USA*, **78**, 6633–6637 (1981).
9. M.S. Urdea, J. Kolberg, J. Clyne, J.A. Running, D. Besemer, B. Warner, R. Sanchez-Pescador, 'Application of a Rapid Non-radioisotopic Nucleic Acid Analysis System to the Detection of Sexually Transmitted Disease-causing Organisms and Their Associated Antimicrobial Resistances', *Clin. Chem.*, **35**, 1571–1575 (1989).
10. T. Horn, C.A. Chang, M.S. Urdea, 'Chemical Synthesis and Characterization of Branched Oligodeoxyribonucleotides (bDNA) for Use as Signal Amplifers in Nucleic Acid Quantitation Assay', *Nucleic Acids Res.*, **25**, 4842–4849 (1997).
11. M. Grunstein, D.S. Hogness, 'Colony Hybridization: A Method for the Isolation of Cloned DNAs that Contain a Specific Gene', *Proc. Natl. Acad. Sci. USA*, **72**, 3961–3965 (1975).
12. C. Lockey, E. Otto, Z. Long, 'Real-time Fluorescence Detection of a Single DNA Molecule', *BioTechniques*, **24**, 744–746 (1998).
13. Z. Long, L. Li, T. Grooms, C. Lockey, K. Nader, I. Mychkovsky, S. Mueller, I. Burimski, P. Ryan, G. Kikuchi, D. Ennist, S. Marcus, E. Otto, G. McGarrity, 'Biosafety Monitoring of Patients Receiving Intracerebral Injection of Murine Retroviral Vector Producer Cells', *Hum. Gene Ther.*, **9**, 1165–1172 (1998).
14. K.J. Livak, J. Marmaro, J.A. Todd, 'Towards Fully Automated Genome-wide Polymorphism Screening', *Nature Genet.*, **9**, 341–342 (1995).
15. C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, 'Real Time Quantitative PCR', *Genomic Res.*, **6**, 986–994 (1996).
16. K.J. Livak, S.J. Flood, J. Marmaro, W. Giusti, K. Deetz, 'Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for

Detecting PCR Product and Nucleic Acid Hybridization', *PCR Methods Appl.*, **4**, 357–362 (1995).

17. K. Paradis, G. Langford, Z. Long, W. Heneine, P. Sandstrom, W.M. Switzer, L.E. Chapman, C. Lockey, D. Onions, E. Otto, 'Search for Cross-species Transmission of Porcine Endogenous Retrovirus in Patients Treated with Living Pig Tissue', *Science*, **285**, 1236–1241 (1999).
18. S. Kwok, R. Higuchi, 'Avoiding False Positives with PCR', *Nature (London)*, **339**, 237–238 (1989).
19. P.A. Kitchin, Z. Szotyori, C. Fromholc, N. Almond, 'Avoidance of False Positives', *Nature (London)*, **344**, 201 (1990).
20. G.D. Cimino, K.C. Metchette, S.T. Isaacs, Y.S. Zhu, 'More False-positive Problems', *Nature (London)*, **345**, 773–774 (1990).
21. M.C. Longo, M.S. Berninger, J.L. Hartley, 'Use of Uracil DNA Glycosylase to Control Carry-over Contamination in Polymerase Chain Reactions', *Gene*, **93**, 125–128 (1990).
22. G.D. Cimino, K.C. Metchette, J.W. Tessman, J.E. Hearst, S.T. Isaacs, 'Post-PCR Sterilization: A Method to Control Carryover Contamination for the Polymerase Chain Reaction', *Nucleic Acids Res.*, **19**, 99–107 (1991).
23. D. Malsch, Z. Long, C. Sutter, 'The Whys and Wherefores of DNA Amplification', *Nature Biotechnol.*, **11**, 940–942 (1993).
24. P.N. Rys, D.H. Persing, 'Preventing False Positives: Quantitative Evaluation of Three Protocols for Inactivation of Polymerase Chain Reaction Amplified Products', *J. Clin. Microbiol.*, **31**, 2356–2360 (1993).
25. M.J. Espy, T.F. Smith, D.H. Persing, 'Dependence of Polymerase Chain Reaction Product Inactivation Protocols on Amplicon Length and Sequence Composition', *J. Clin. Microbiol.*, **31**, 2361–2365 (1993).

DNA Probes

Sanjay Tyagi
Public Health Research Institute, New York, USA

1 Introduction 1
2 Hybridization Formats 2
3 Heterogeneous Hybridization Assays 2
3.1 Target Immobilization 3
3.2 Probe Immobilization 4
3.3 Hybrid Immobilization 4
3.4 Sandwich Hybridization 4
3.5 In Situ Hybridization 5
4 Labels 5
5 Homogeneous Hybridization Assays 6
6 Target Amplification 7
7 Probe Amplification 9
8 Mutation Detection 10
9 Taking DNA Probes into a Protein World 11
10 Future Prospects 12
Acknowledgments 12
Abbreviations and Acronyms 12
Related Articles 12
References 12

Hybridization of nucleic acid probes has been a key technique used in the exploration of the molecular biology of the cell. Use of this technique is now expanding into the field of clinical diagnostics for the detection of pathogens and genetic disorders. In a hybridization assay, a labeled nucleic acid probe of a particular sequence is used to detect the presence of a target nucleic acid sequence. The fundamental principals that guide the design of the nucleic acids hybridization probes and assays are reviewed in this article. New improvements that simplify the implementation of these assays and enhance their sensitivity are discussed. Coupling of hybridization with amplification of nucleic acids in vitro, which results in extremely sensitive detection of specific nucleic acids, are also described.

1 INTRODUCTION

Imagine that you have a magic reagent to which you add a droplet of a body fluid from a patient; you wait for a moment and a glow appears in the tube holding the mixture; the glow not only tells you which pathogen is responsible for the patient's illness, but also indicates which drugs to use to treat the disease. Also imagine that you can perform this diagnosis before any symptoms of the disease appear, improving the chances of success with the treatment, and you can perform this test on a large population with ease. The creation and development of such reagents are the promise of nucleic acid-based detection and are the aspiration of a diverse community of researchers. At the heart of nucleic acid-based detection is the process of hybridization, in which a strand of nucleic acid anneals to its complement. A nucleic acid probe can identify its complement in a sea of unrelated molecules. The bonds that form between complementary strands are among the strongest macromolecular associations known. This combination of high specificity and strong bonds gives the nucleic acid probes their advantage over conventional probes that are often based on antibody–antigen interactions.

When a macromolecular probe such as an antibody binds to its target, the forces that hold them together are electrostatic attractions, hydrogen bonds, salt bridges and hydrophobic interactions, and the bonds are located in a complex three-dimensional pattern. When one needs a protein with the ability to recognize a given target molecule, it is not possible to design it from its primary amino acid sequences, because the rules that predict the shape of a protein from its sequence are not well understood. Even after the rules of protein folding have become clear, this task will remain extremely complex. However, the diversity of the antibody repertoire within mammalian circulatory systems has allowed us to circumvent this hindrance by selecting antibodies that are specific for purified targets. On the other hand, the interaction between a nucleic acid probe and its complement is distinguished from the interaction between an antibody and an antigen by the number and organization of the contacts between the two molecules. A nucleic acid probe forms an ordered series of hydrogen bonds between the complementary nucleotides of the probe and the target strands, and the placement of the bonds is determined by the order of the nucleotides. The hydrogen bonds are placed in a fixed three-dimensional context by stacking interactions that occur between neighboring bases in each strand. These features permit the design of a nucleic acid probe for any given target nucleic acid by simply selecting a complementary sequence. Moreover, the strength and the specificity of the designed probes can be predicted with high precision.[1]

The technique of nucleic acids hybridization was first developed in the laboratory of Sol Speigelman.[2–4] A brief history of the key developments in hybridization technology has been described by Keller and Manak.[5]

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The most familiar format for performing hybridization reactions has been to label the probes with radioactive atoms, fluorophores, or reporter enzymes, incubate them with denatured and immobilized DNA, remove the excess probes by extensive washing, and then detect the signal that is associated with the surface on which the target is immobilized. The amount of target can be quantitatively inferred from the strength of the signal. One of the most significant developments occurred after the discovery of restriction endonucleases and molecular cloning, which made it possible also to obtain information on the size of the DNA that contains the target of the probe.[6,7] In these assays, a DNA sample is incubated with restriction enzymes, the variously sized fragments that are generated are fractionated within a gel by electrophoresis and then blotted on a nitrocellulose membrane, and labeled probes are hybridized to the fragments on the membrane.[8] The analytical power of Southern blotting has made it a mainstay of molecular biology research. The power of this approach has been enhanced by the capability of generating single-stranded probes of defined sequence and high density labeling by in vitro transcription.[9]

The technique of hybridization has been utilized to explore the flow of information within cells and has been instrumental in the discovery of a number of biological phenomena. For example, the very first hybridization experiment[2] demonstrated that messenger RNAs are complementary to their DNA templates. The discovery that a large fraction of mammalian DNA contains small sequences that are repeated hundreds of times was accomplished in a very simple experiment in which the kinetics of reassociation of denatured calf thymus DNA were measured.[10] Intragenic noncoding sequences (introns) were discovered in eukaryotic genes by hybridizing mRNAs to genomic DNA and then observing in electron micrographs that regions of the genomic DNA looped out from the heteroduplex because they did not have complementary sequences in the RNA.[11] Hybridization reactions also have often been used to isolate homologous genes from other species. For example, from the knowledge that the *engrailed* gene in fruit flies is responsible for segmentation, it is possible to construct probes that can be used to isolate the homologous gene from humans, and to understand the role that it plays in human embryogenesis.[12]

Recent developments in hybridization technology allow it to be used for the detection of pathogens, mutations, and molecular rearrangements in clinical diagnostic settings and for forensic analysis. It is now possible to determine the extent to which a human tissue is infected with a pathogen by probing for the nucleic acids of that pathogen. This mode of pathogen detection is different from conventional immunoassays where the immune response of the body to the pathogen is detected. For this reason, it is possible to detect pathogens in absence of, or before the onset of, the immune response. Any pathogen can be detected from the knowledge of its nucleic acid sequence. Similarly, mutations that are related to disease state can be detected by simple and automated procedures.

2 HYBRIDIZATION FORMATS

The physical properties of nucleic acids that exhibit measurable changes upon hybridization are viscosity, absorption of ultraviolet light, and circular dichroism. The magnitudes of these changes are so small that millimolar concentrations of nucleic acid are required to observe the changes. Therefore, in order to measure the degree of hybridization for targets that are present in smaller concentration range in a sample, it is necessary to label the probes with a label that can be detected with higher sensitivity. These labeled probes are incubated with the sample, the hybrids are immobilized on a solid surface, and the excess of the probes is removed by washing the surface. Finally, the quantity of the label that remains after the washing is determined. Since these assays utilize both liquid and solid phases, their format is referred to as heterogeneous. Recently, it has become possible to detect hybridization in a liquid phase without having to immobilize and isolate the hybrids. These assays are called homogeneous assays.

3 HETEROGENEOUS HYBRIDIZATION ASSAYS

In order to isolate probe–target hybrids, a number of different solid surfaces and immobilization formats have been explored. Among the solid surfaces that are utilized are filter papers, nitrocellulose and nylon membranes, plastic beads, the walls of test tubes, and paramagnetic particles. Each surface offers unique advantages. Nitrocellulose binds to denatured DNA by adsorption, and slows down its renaturation, but at the same time permits hybridization with probes that are in excess. Nylon membranes are mechanically stronger than nitrocellulose membranes and do not become brittle when subjected to heating and drying steps, and they can therefore be probed multiple times. Magnetic beads are microscopic ferromagnetic particles that are coated with silanes and coupled with specific oligonucleotides or with other affinity agents such as streptavidin. Hybrids attached to the paramagnetic particles via the affinity agents and then can be isolated with a permanent magnet which causes the particles to stick to the wall

of the test tube and the supernatant can be withdrawn by aspiration.[13] Another advantage of paramagnetic particles is their large surface area, which provides a high binding capacity. Sometimes the surface of plastic dipsticks, test tubes or microtiter wells is used to immobilize hybrids.[14] Although this markedly simplifies the assays, this mode of hybrid immobilization suffers from poor binding capacity. Glass surfaces, such as microscope slides, are commonly used when the probes have a fluorescent label.[15]

There are a number of different immobilization formats for isolating probe–target hybrids from the excess of labeled probes. The most common method is to immobilize the denatured target nucleic acid and to perform the hybridization by immersing the solid medium in a solution of labeled probes. This is followed by washing the surface under conditions that maintain the hybrids but remove the excess probes. A second method is to immobilize a series of different probes at predetermined locations on a surface and then add labeled targets. In these assays, the location of the signal on the surface identifies the targets that are present in the sample. The third method is to perform the hybridization in solution and then immobilize the hybrids on surfaces. A fourth method is to perform assays in which the target is sandwiched between two probes, one of which serves as a label probe and the other serves as a capture probe. In these assays, the presence of the signal indirectly indicates the presence of the target, since the target bridges the signal and the capture probes. Finally, fixed tissue sections can be used for in situ hybridization, where the target molecules are naturally immobilized within the cellular matrix and the probes are hybridized to the targets at their original location.

3.1 Target Immobilization

Southern blotting is the most traditional and commonly used hybridization format, named after the developer of the technique.[8] This format is used in molecular cloning experiments to identify bacterial colonies that contain a particular nucleic acid target sequence. It is also used in the determination of the size of the RNA or DNA that contains a particular target sequence. The applications of this format range from the 'fingerprinting' of human DNA for forensics to the study of changes in the level of expression for particular RNAs upon treatment with drugs. In performing Southern hybridizations, the targets are denatured in situ, i.e. within the bacterial colony or within the gel that is used to segregate the nucleic acids by size, and then transferred to a nitrocellulose or nylon membrane. The DNA attached to the membrane is then hybridized to a radioactively labeled probe, excess probes are washed away, and autoradiography is performed to locate the position of colonies or bands.[16]

The power of this method is illustrated by an example in which different *Mycobacterium tuberculosis* isolates are typed in order to track the transmission of the disease in human populations. In this method, *Mycobacterium tuberculosis* is cultured from infected patients and its genomic DNA is isolated. The DNA is incubated with the restriction enzyme *Pvu II*, which cleaves it at a number of specific locations, generating a number of fragments of different lengths. Digested DNA from *Mycobacterium tuberculosis* bacteria that are isolated from different individuals are electrophoresed in different lanes of an agarose slab gel, in which different fragments are separated according to their size. The fractionated DNAs are transferred to a nylon membrane and hybridized to a 325-nucleotide long radioactively labeled probe that is specific for the insertion sequence *IS6110*. This insertion element, which is present at multiple sites, has a tendency to hop randomly from one place to another in the genome of *Mycobacterium tuberculosis*. Owing to this tendency of the insertion sequence, different sets of fragments are labeled by the probe for different strains of the bacterium. Thus, each strain produces a characteristic pattern of bands in the autoradiogram (Figure 1). Using this 'fingerprint', it is possible to identify each isolate and to chart the spread of a strain as it is transmitted within a population.[17]

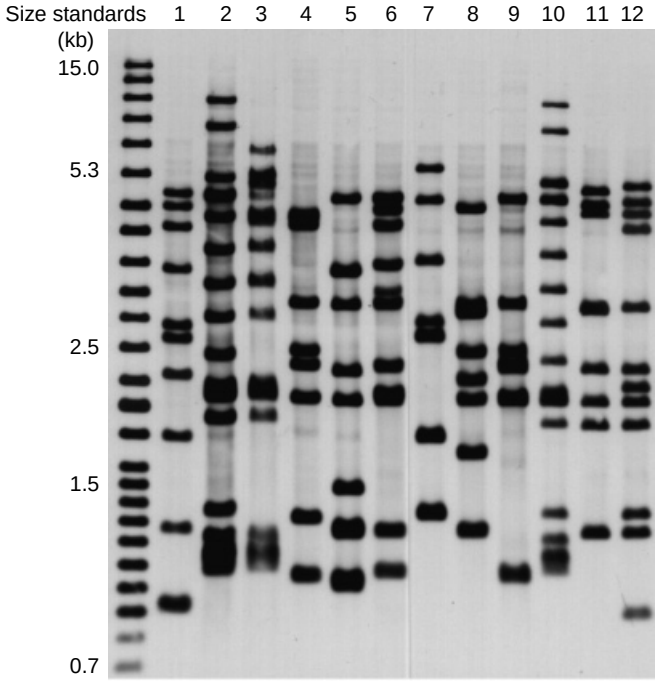

Figure 1 Fingerprinting isolates of *Mycobacterium tuberculosis* by Southern hybridization. (Courtesy of P.J. Bifani and B.N. Kreiswirth.)

3.2 Probe Immobilization

The second most common mode of isolation of probe–target hybrids is via immobilized probes. In this format the probes are linked to unique positions on a solid surface and the target is labeled. This format permits the interrogation of a sample for the presence of multiple sequences, as a number of different probes can be linked to different locations on the solid surface. The location of the signals indicate which sequences are present in the sample. With the advances in the chemistry of DNA synthesis and immobilization, it has become possible to attach thousands of different probes in an area as small as 1 cm^2 and inquire whether a target contains their complement.[18,19] These high-density arrays of immobilized probes are commonly known as 'DNA chips'.

3.3 Hybrid Immobilization

Specific immobilization of probe–target hybrids is uncommon because of the lack of physical methods that can distinguish probe–target duplexes from unrelated double-stranded DNA. However, in the case of RNA–DNA hybrids, it has been possible to raise an antibody that recognizes RNA–DNA heteroduplexes. This antibody is used to immobilize the hybrids on solid surfaces. In these assays, the antibody that recognizes the RNA–DNA hybrid is attached to the surface of microtiter wells, the hybridization mixture is added to the well, and excess probes are washed away (Figure 2). The RNA probes are constructed in such a way that each probe carries many biotin moieties. A second antibody (or, in some cases, streptavidin) that binds to the biotin moieties in the probes and is also attached to a reporter is added to develop the signals.[20]

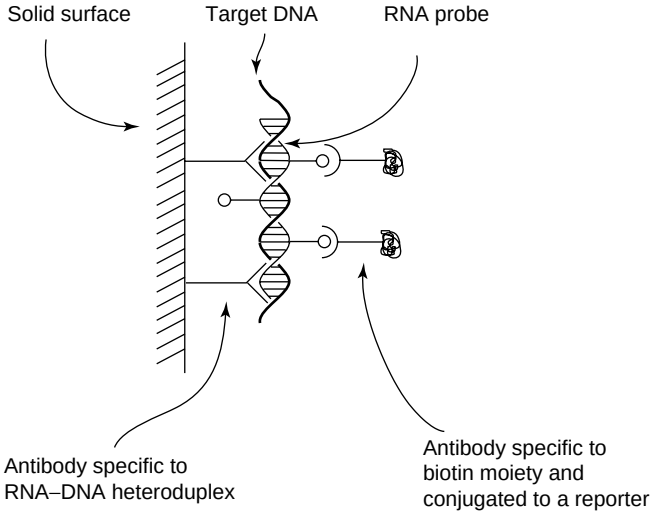

Figure 2 Detection of hybrids by immobilization of RNA–DNA heteroduplex.

3.4 Sandwich Hybridization

When hybridization is performed on solid surfaces with immobilized target or probes, the hybridization proceeds slowly and does not attain its full potential because the access to the targets or the probes is hindered.[21] In addition, the surfaces bind nonspecifically to labeled probes and thereby obscure the weak signals emanating from rare targets. One of the schemes that make the signals more dependent on the presence of the targets is sandwich hybridization, in which two probes are utilized simultaneously. The first probe serves as a capture probe, and can bind both to the solid surface and to the target. The second probe is a reporter probe, which can bind to the target at a second site and carries a reporter group. Thus, the association of the signal to the solid surface is strictly target dependent (Figure 3).

Usually, in sandwich hybridization assays the hybridizations are performed in solution and the hybrids are then transferred to a solid surface. Among the means for immobilization of the capture probes are biotin–streptavidin links or oligo-dT : oligo-dA links.[13,22] Sandwich hybridization reactions often decrease background signals to undetectable levels. However, when the detection is made highly sensitive, low-level background signals become significant. In order to reduce background signals further, a number of modifications in the basic sandwich hybridization format have been used. For example, in a reversible target capture scheme hybrids are eluted from paramagnetic particles after removal of the excess probes and recaptured on to the surface of a new set of paramagnetic particles. Nonspecifically bound probes which contribute to the background are discarded along with the first set of particles. The capture and release steps are repeated several times to attain extremely sensitive detection.[13]

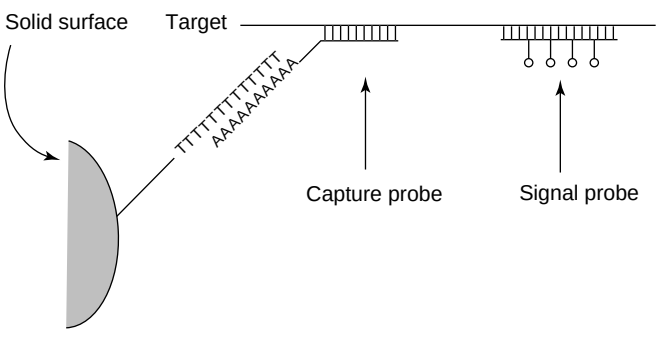

Figure 3 Sandwich hybridization.

3.5 In Situ Hybridization

In situ hybridization is utilized to determine the location of specific nucleic acid sequences within tissues. It can also be used to detect specific sequences in particular subcellular domains, such as the cytoplasm, the nucleolus or the chromosomes. In situ hybridization reactions in entire organisms, such as insects or rodent embryos, are also possible. During in situ hybridization, the nucleic acid targets remain trapped within the cellular matrix, while remaining accessible to probes. In order to make the nucleic acids targets accessible to probes, while maintaining the cellular and histological integrity of the specimens, they are 'fixed' with reagents that remove membranes and cross-link proteins. Fixed specimens can withstand the rigors of hybridization, washing, and signal development.

The probes for in situ hybridization are radioactively labeled or labeled with tags that can generate color or fluorescence at the site of hybridization. In order to confine the signal to the hybridization site in intracellular localization experiments, ^{35}S-labeled probes and emulsion autoradiography are often utilized. The radiation from the ^{35}S radionuclide can travel only short distances and the photosensitive emulsion is in intimate contact with the specimen ensuring high-resolution images.[23] High-resolution images of the target distribution can also be obtained by utilizing fluorescence labels in thin tissue sections.

In the following example, in situ hybridization is used to map the distribution of two particular mRNAs within embryos of the fruit fly, *Drosophila melanogaster*. The target nucleic acids were *bicoid* and *oskar* mRNAs that localize at the anterior and the posterior ends of early preblastoderm embryos. RNA probes that are complementary to each of these mRNAs are prepared by in vitro transcription of the cloned genes encoding these RNAs. Digoxygenin-bearing nucleotides are incorporated into these probes during in vitro transcription. After permeabilization and fixing, the embryos are incubated with the probes. Excess probes are removed by several rounds of washing. The signal is developed in a series of reactions in which the embryos are first incubated with an antibody that is conjugated to alkaline phosphatase that can bind to the digoxygenin moiety. Further washing removes the antibody–alkaline phosphatase conjugates that are not bound to digoxygenin moieties and then a chromogenic substrate of the alkaline phosphatase, nitroblue tetrazolium, is added. A bluish-purple precipitate is created by the alkaline phosphatase wherever the RNA is localized. The gradient of color stemming from the two poles of the embryo indicated the distribution of the two RNA molecules (Figure 4).

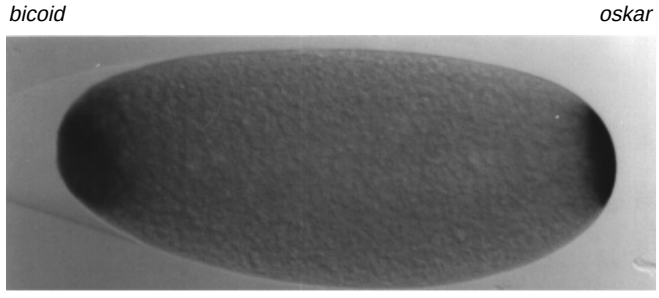

Figure 4 Detection of localized mRNAs in an early embryo of *Drosophila melanogaster* by in situ hybridization. The *bicoid* mRNA localizes at the anterior end and the *oskar* mRNA localizes at the posterior end of the embryo. (Courtesy of P. Tolias.)

4 LABELS

A number of different labeling schemes, including the use of radioactive atoms, fluorescent moieties, chemoluminescence generating systems, and enzymatic reporters, have been explored in the construction of DNA probes, some of them having been inherited from immunoassays. Radioactive ^{32}P is a natural label for nucleic acids since it can be incorporated into their backbone during synthesis with radioactive nucleoside triphosphates. In order to improve the resolution of the autoradiograms, ^{35}S is used since the radiation from ^{35}S can only travel short distances. ^{35}S can also be introduced into the probes via radioactive nucleotide triphosphates. For reasons of safety, stability, and waste disposal, nonradioactive labels have been utilized. Among the nonradioactive labels that can provide sensitivity equivalent to radioactive labels are enzymatic labels that turn a colorless substrate into a colored product or turn a nonfluorescent substrate into a fluorescent product. Enzymatic labels can produce several thousand detectable products due to catalytic turnover. Commonly used enzymatic labels are alkaline phosphatase, horseradish peroxidase, and β-galactosidase.[24,25] Most often, enzymatic labels are introduced indirectly in the assay schemes because hybridization conditions denature proteins, and the bulk of the proteins may interfere with hybridization. A common mechanism of indirect labeling is to incorporate a small hapten such as a biotin moiety during the synthesis of the hybridization probe, add an enzyme-linked antibody for the hapten, and then remove the excess antibodies. Signals are developed by adding the substrate for the enzyme. In addition to chromogenic substrates, chemicals that can participate in a light-generating reaction are also commonly employed. These chemiluminescent labels provide very high sensitivity and can be used just like radioactive labels because both autoradiography and scintillation counting can be performed to

detect the signal.[24] Recently, fluorophores have become common labeling moieties. Fluorescent labeling moieties do not offer a sensitivity as high as radioactivity and enzymatic labels, but they have become popular because rare targets can be made to be more abundant by nucleic acid amplification. Another reason for using fluorescent labels is that high sensitivity can be achieved when localized signals are viewed under the microscope.

5 HOMOGENEOUS HYBRIDIZATION ASSAYS

Although hybridization assays that rely on the isolation of probe–target hybrids offer high sensitivity, they are cumbersome to perform. In addition, the requirement for immobilization and washing precludes their use in the real-time monitoring of processes that produce nucleic acids. It is necessary to sample the reactions at different times and quantify the products at a later time. In addition, in order to detect nucleic acids within cells, the cells have to be fixed to permit removal of the probes. It is desirable to follow the progress of nucleic acid synthesis reactions without interruption and to detect nucleic acids in living cells and organisms without killing them. Several assays have recently been developed that can detect nucleic acids in homogeneous solutions without the need to isolate the probe–target hybrids. These assays utilize interactive fluorescence reporters that exchange energy in a distance-dependent manner.[26]

The earliest homogeneous hybridization assays based on fluorescence resonance energy transfer (FRET) utilized two oligonucleotides that bind to neighboring sites on a target.[27,28] One probe in this binary probe system carries a donor fluorescent moiety at its 3′ end, and the other probe carries an acceptor fluorescent moiety at its 5′ end. The donor and acceptor fluorophores are chosen so that efficient energy transfer can take place from the donor to the acceptor when the two probes are in close proximity, whereas no energy transfer should occur when the two probes are apart from each other and are free in solution. Energy transfer, which is measured by the decrease in donor fluorescence or the increase in acceptor fluorescence, indicates the extent of hybridization and can be used to follow hybridization in real time (Figure 5a). An additional benefit of this approach is higher specificity, as two probes, rather than one, must bind to the target in order to generate a signal.

A second approach that uses FRET in homogeneous hybridization assays also uses a pair of oligonucleotides, one of which is designed to be complementary both to the target strand and to the other oligonucleotide. The 5′ end of one oligonucleotide is labeled with a donor fluorophore and the 3′ end of the other oligonucleotide is labeled with an acceptor fluorophore, such that when the two oligonucleotides anneal to each other the two labels are next to each other and exhibit efficient energy transfer. Since the binding of two small oligonucleotides is a dynamic process, when the target is added some of the oligonucleotides bind to the target, resulting in a decrease in energy transfer (Figure 5b). Thus, the extent of increase in the fluorescence of the donor fluorophore and the extent of decrease in the fluorescence of the acceptor fluorophore indicate the extent of hybridization.[29,30]

In a third approach, the donor and acceptor fluorophores are attached at two ends of the same

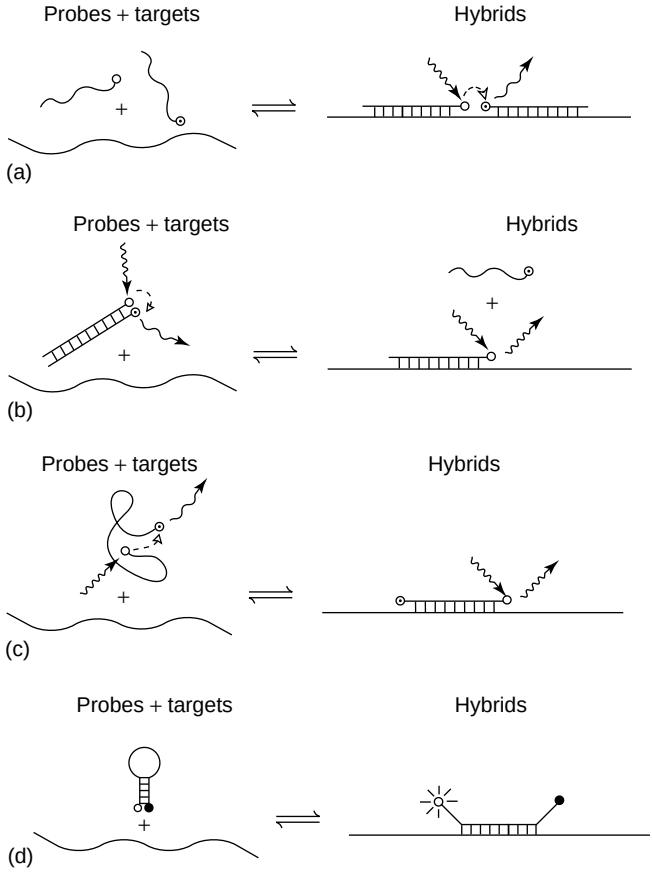

Figure 5 Different methods for detection of hybrids in a homogenous solution using interacting fluorophores. (a) Two probes bind to a target on adjacent sites and their fluorophores and interact via FRET. (b) Target binds to one of the oligonucleotides in a quenched pair of oligonucleotides and restores fluorescence. (c) The doubly labeled probe has a random coil conformation when it is free and a stretched-out conformation when it is bound to a target; FRET can take place in the free from, whereas it is precluded in the target-bound form. (d) On their own, molecular beacons are dark because their hairpin stem holds a quencher in close proximity to their fluorophore, whereas upon binding to a target they undergo a conformational reorganization that restores fluorescence.

oligonucleotide, which serves as the probe. Since an oligonucleotide in solution behaves like a random coil, the two ends sometimes come close enough to each other within the lifetime of the excited state of the donor fluorophore that some energy transfer occurs. However, when the probe binds to its target, the two ends are permanently kept away from each other, precluding energy transfer (Figure 5c). Although, the magnitude of the changes in energy transfer is relatively modest, it is possible to use it for monitoring hybridization.[31]

Another limitation of doubly labeled random-coil probes, in addition to the relatively small changes in fluorescence that occur upon hybridization, is that the magnitude of the changes decrease with the length of the probe. A fourth approach based on energy transfer solves these problems by introducing small complementary sequences on either end of the probe that brings the two labels in close proximity through the formation of a hairpin stem. The length of the stem sequences is chosen such that the formation of a probe–target hybrid readily disrupts the stem and removes the fluorophore from the vicinity of the quencher (Figure 5d). These probes, called molecular beacons, have an additional useful feature: the acceptor moiety is a nonfluorescent quencher dye.[32,33] As a consequence, nonhybridized molecular beacons are dark and hybridized ones are fluorescent (Figure 6). In addition, the hairpin stems of these probes bring the fluorophore and the quencher so close to each other that energy transfer takes place by more efficient mechanisms than FRET, resulting in a very high level of quenching. Furthermore, these close-distance energy transfer mechanisms do not require any particular relationship between the spectra of the fluorophore and the quencher, as FRET-based mechanisms do. As a consequence, well-quenched probes can be constructed from any fluorophore, allowing multiplex assays for many different targets in the same solution.

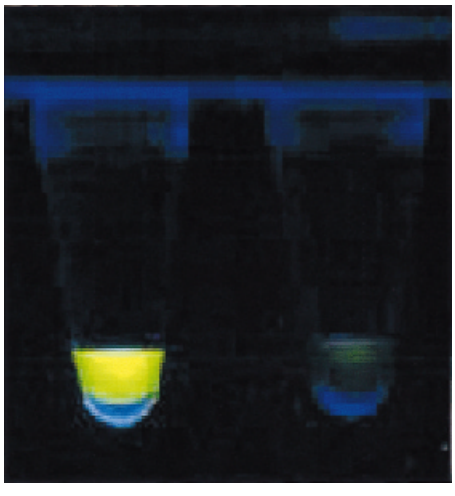

Figure 6 Detecting a specific nucleic acid in homogeneous solution with a molecular beacon. Two test tubes that contain a solution of the molecular beacon probe are illuminated with an ultraviolet lamp. The tube on the left also contains a target for the molecular beacon.

6 TARGET AMPLIFICATION

One of the reasons why hybridization-based assays did not find widespread utility in clinical laboratories for a long time is that targets that are clinically relevant are often rare constituents of biological samples. For example, pathogens, such as the human immunodeficiency virus, may be present in as few as one in 10^6 human cells, yet that individual may still be infectious. Another significant factor is the complexity of these assays. These limitations have been overcome by the development of techniques for the selective in vitro amplification of target nucleic acids. The most developed of these techniques is the polymerase chain reaction (PCR).[34] In this process, a specific nucleic acid sequence is amplified using two oligonucleotide primers, deoxyribonucleotides, a thermostable DNA polymerase, and a thermal cycler. The primers hybridize to opposite strands of the target sequence and initiate product strand synthesis by DNA polymerase and the thermal cycler denatures the double-stranded templates by raising the temperature. As the temperature of the reaction mixture is cycled between the stages of denaturation and annealing, the number of the DNA fragments defined by the two primers doubles every cycle. As a result of this exponential amplification, the target sequence is enriched more than 10^6-fold. After selective enrichment, the target sequences can readily be detected by hybridization.

In addition to the PCR, there are two other well-developed target amplification schemes: strand displacement amplification[35] and transcription-based amplification.[36] These techniques also utilize a pair of specific primers to amplify a region of the target. These methods accomplish the denaturation of double-stranded products by enzymatic means rather than via thermal denaturation, and the amplification can therefore be performed at one uniform temperature, eliminating the need for a thermal cycling device. Strand displacement amplification utilizes a DNA polymerase that displaces the DNA strand complementary to the template strand while extending the primer. The extendable 3′ ends are created by the nicking activity of a restriction endonuclease. The displaced strands serve as templates for further rounds of synthesis. As a result, the DNA is amplified exponentially as long as the primers and the enzymes remain in excess.[35]

Transcription-based amplification utilizes the RNA polymerase of bacteriophage T7, reverse transcriptase, and ribonuclease H in a concerted reaction. The initial RNA templates are copied by reverse transcription and one of the primers carries a tail corresponding to the bottom strand of the T7 RNA polymerase promoter. The original RNA in the resulting RNA–DNA hybrid is degraded by the action of ribonuclease H. The second primer then binds to the single-stranded DNA and is extended by the natural DNA polymerase activity of the reverse transcriptase, creating a functional promoter for the T7 RNA polymerase. The T7 RNA polymerase then generate more than 100 copies of the DNA template. Owing to the high turnover rate of T7 RNA polymerase, this reaction also produces large-scale amplification of the target.[36]

In order to result in amplification, both primers must bind to their targets in each cycle of PCR, and sometimes, especially with nucleic acids of high complexity, nonspecific products are also amplified. As a result, it is necessary to analyze the products of amplification by electrophoresis, followed by a hybridization step to identify the desired amplification product. In these confirmatory hybridization reactions, it is not necessary to use highly sensitive labels such as radioactive atoms or chromogenic enzymes. However, the assays remain labor intensive because of post-amplification analyses. In addition, the amplified DNA contaminates the laboratory and can become a source of false signals in future assays. Recently, the development of new homogeneous hybridization techniques has solved these problems by linking specific hybridization with nucleic acid amplification. It is now possible to perform extremely sensitive detection of specific sequences in a sealed assay tube.

The homogeneous hybridization techniques discussed above can be used to detect the product of amplification in the same tube in which the reaction is carried out. In addition, it is possible to monitor the progress of these reactions in real time, which enables quantitation of the initial template within an extremely wide dynamic range.[32,37,38] Although each of the target amplification schemes mentioned above can be monitored with a homogeneous hybridization method, we shall discuss the use of molecular beacons as an example of real-time monitoring of amplification reactions.

In order to monitor the progress of a PCR, a molecular beacon whose loop sequence is designed to be complementary to a region of the expected amplification product is included in the reaction mixture. The length of the arm sequences of molecular beacons is chosen so that the stem is stable at the annealing temperature of the PCR. The length of the loop sequence is chosen so that the probe–target hybrid would be stable at the annealing temperature also. During the denaturation step

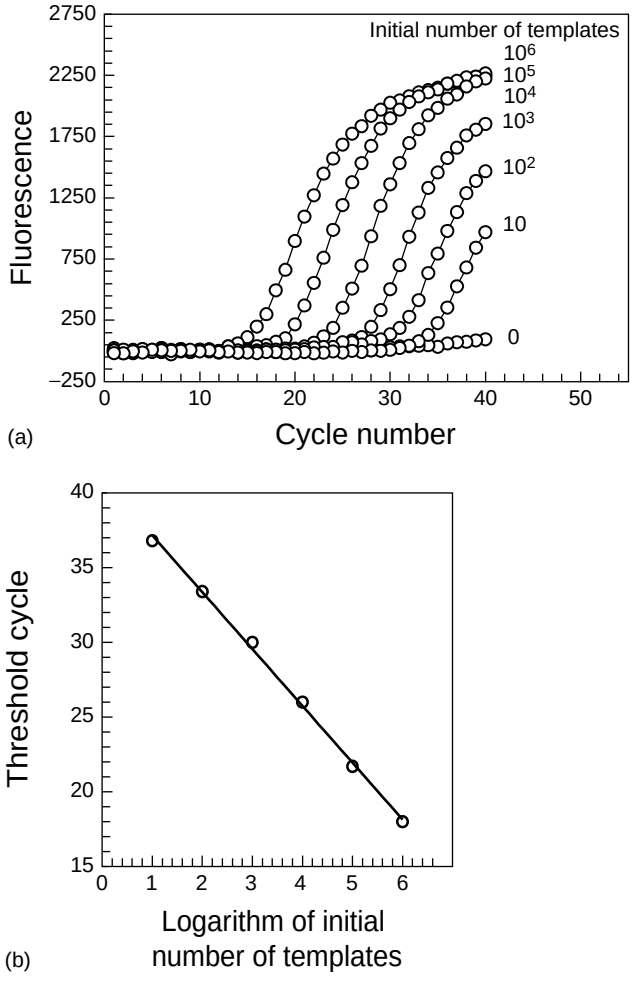

Figure 7 Real-time measurement of amplicon synthesis during PCR using molecular beacons. (a) Change in fluorescence in seven reactions, each initiated with a different number of template molecules (indicated), is plotted against the number of cycles. (b) The threshold cycle, the cycle at which the signal becomes detectable, is plotted against the logarithm of the initial number of templates.

of the PCR, the molecular beacons assume a random coil configuration and fluoresce. As the temperature is lowered to allow annealing of the primers, stem hybrids form rapidly, preventing fluorescence. However, at the annealing temperature, the molecular beacons bind to the complementary strand in the amplified DNA and generate fluorescence. When the temperature is raised to allow primer extension, the molecular beacons dissociate from their targets and do not interfere with polymerization. A new round of hybridization takes place in the annealing step of every cycle, and the intensity of the resulting fluorescence indicates the amount of accumulated amplified DNA (Figure 7a). The number of thermal cycles that it takes for the fluorescence intensity to reach a detectable value is inversely proportional to the

logarithm of the number of template molecules that are initially present in a reaction (Figure 7b). These extremely sensitive and quantitative assays can be performed in a single tube and require one operational step in which the reaction mixture is assembled.

In addition to molecular beacon probes, other homogeneous methods of hybridization can also be used to monitor the progress of amplification reactions. FRET-based binary probes can be used to perform real-time PCR.[27] The two probes that bind to adjacent sites on the DNA strand being amplified are added in the initial reaction mixture along with the primers. In order to monitor the progress of the reaction in real time, the extent of resonance energy transfer is measured continuously during the reaction.[39] Similarly, random-coil probes labeled with two interacting fluorophores can also be used to monitor the progress of PCRs in real time. These probes can draw additional benefit from the natural 5′–3′ endonucleolytic activity of DNA polymerase, since this activity results in cleavage of the probe that the polymerase finds sitting in its tract during polymerization.[40,41] Since the cleaved fragments diffuse away from each other are not able to participate in FRET, the fluorescence increases and provides a measure of the progress of the PCR.[37] With reference to the 5′–3′ endonucleolytic activity of the Taq DNA polymerase, these probes are called TaqMan probes. Interfacing these homogeneous hybridization methods with amplification reactions allows the simple and sensitive detection of rare nucleic acids that are present in complex nucleic acid mixtures.

7 PROBE AMPLIFICATION

In vitro amplification of target sequences is highly effective for the detection of rare sequences. An alternative strategy is to amplify the signals associated with the probes themselves. For example, by using reporter enzymes it is possible to generate 1000 detectable signal moieties for every probe that is bound to its target. However, higher levels of amplifications are required for the detection of rare pathogens in clinical samples. There are three strategies that provide the required degree of amplification: RNA hybridization probes that can be amplified by Qβ-replicase, branched-DNA probes that form dendrimeric complexes, and padlock probes that can initiate rolling circle amplification. Assays based on these signal amplification techniques are as sensitive as those based on target amplification techniques.

The Qβ replicase-mediated amplification system relies upon bifunctional recombinant RNA probes that can serve as hybridization probes and can be amplified exponentially upon incubation with the RNA-directed RNA polymerase isolated from *Escherichia coli* infected with bacteriophage Qβ.[42,43] This system exploits the ability of Qβ replicase to amplify naturally occurring RNA templates such as midivariant-1 (MDV-1) RNA. These RNAs have evolved to serve as templates for Qβ replicase. Their replication does not require primers and strand separation occurs naturally at 37 °C owing to the self-structure of the RNA and to the action of the replicase. RNAs that replicate exponentially possess an internal replicase binding site, a particular 3′-terminal sequence for the initiation of replication, and a particular 5′-terminal sequence that encodes the 3′-terminal initiation sequence needed for the initiation of synthesis from the product strand. Hybridization probe sequences are inserted within a region of the MDV-1 RNA that preserves the elements that are necessary for replication.[44] In order to perform assays that detect rare targets, these probes are hybridized to the target along with capture probes, the hybrids are collected on the surface of paramagnetic particles, excess of the probes is removed by washing, the hybrids are eluted into a solution, and finally Qβ replicase is added to the eluted hybrids, amplifying the probes 10^{11}-fold in a 30-min incubation.[45] The amplified RNA can be detected by fluorescence of an intercalating dye. In order to ensure that the nonspecifically bound probes that survived the stringent washing do not produce a signal, an additional target-dependent step has been interjected in these assays. This involves using recombinant RNA probes that are broken into two pieces in the middle of the probe region, neither of which can be amplified exponentially on its own. However, when the two segments of the probe are brought next to each other by binding to the target, they can be ligated together, generating fully functional replicating RNA (Figure 8). With binary hybridization probes that become replicatable by the mediation of the target, it is possible to detect as few as 100 copies of a target in a complex nucleic sample.[46]

A second approach which relies on the hybridization probes to give a signal that is substantially larger than the number of probes is based on the use of branched DNA probes. These probes bind specifically to their target and carry multiple branches that stem from the hydroxyl functionalities of the deoxyribose sugars. These branches themselves are oligonucleotide sequences to which other branched DNA oligonucleotides can be attached. Depending upon the valency of each branched oligonucleotide and upon the number of successive layers that are used, a very large dendrimeric structure can be assembled. Reporter enzymes are attached at multiple locations on the branches of these dendrimers. Thus, each target sequence results in a large number of detectable labels after the signals are developed.[47]

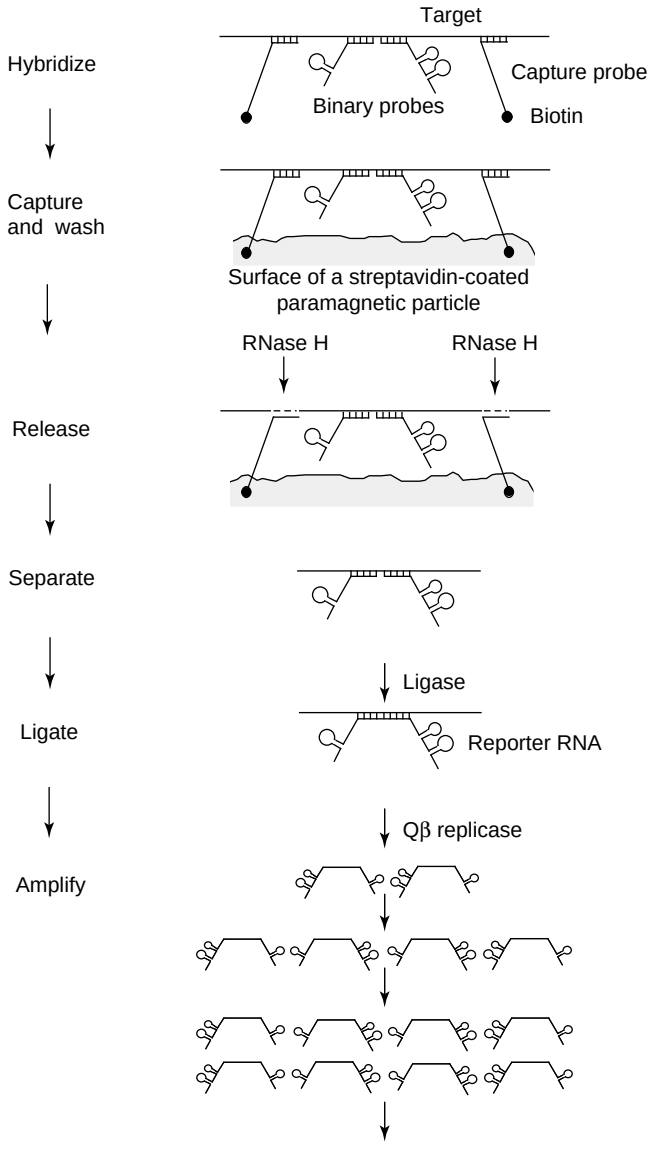

Figure 8 Detection of a rare target RNA by probe amplification with Qβ-replicase. The details of the physical and enzymatic steps are described in the text. (Reproduced by permission from Tyagi et al.[46] Copyright 1996 National Academy of Sciences, USA.)

In a third approach, circularizing oligonucleotide probes (padlock probes) are utilized. The 5′ and 3′ end regions of these linear oligonucleotides are designed to base pair next to each other on a target strand. If properly hybridized, the two ends can be joined by target-dependent enzymatic ligation, converting the linear probes to circular molecules that are catenated to the target sequence.[48] Owing to this topological linkage with the target, the hybrids can be washed with extremely high stringency, markedly enhancing the assay specificity. The probe remains associated with the target molecule even when the hybrid is disrupted. This permits localized detection of target sequences on a chromosome with fluorescence. In addition to permitting stringent washing, the circularized probes allow signal amplification by a rolling-circle replication mechanism. In this DNA synthesis process, a primer is annealed to the circle and extended with a DNA polymerase. DNA polymerase extends the primer using the circle as a template and when it reaches the tail of the primer the polymerase begins to displace the primer and the extended product. Product strand displacement continues indefinitely, resulting in the synthesis of a very long DNA containing a tendemly repeated copy of the circularized probe.[49] This multimeric linear copy of the probe can be exponentially copied by including a second primer that can bind to the product strand. When the second primer binds, a cascade of DNA synthesis and strand displacement ensues, resulting in the synthesis of a large amount of DNA. This serves as a signal indicating the presence of the target to which the circularizing probes were initially bound.[50,51]

8 MUTATION DETECTION

Accelerated by the human genome project, a large number of genetic variations are being discovered, some of which are linked to genetic disorders. Therefore, the task of detecting specific sequences often involves distinguishing closely related sequences that can differ from each other by only a single nucleotide substitution. Although oligonucleotide probes are exquisitely specific, their power to discriminate between closely related targets decreases as their length increases. As the length of the probe is increased, the free energy penalty resulting from a mismatched base pair in a probe–target duplex represents a smaller and smaller fraction of the total free energy change associated with binding. Consequently, allele-specific oligonucleotides for a target region have to be found by making them progressively smaller and by making the conditions of hybridization progressively more stringent. Several groups have described ways of improving the specificity of oligonucleotide probes by making them more structured.[52] This can be accomplished by embedding the oligonucleotides within hairpin stems or by circularizing them.[33,53] Alternatively, specific enzymes that recognize matched or mismatched base pairs are recruited to enhance the level of discrimination.[54,55]

In order to detect single-nucleotide variations in clinical diagnostic settings, or for the discovery of relationships between genotype and disease, it is important to be able to perform accurate assays on a large number of samples.

A number of techniques that have been conventionally used for the detection of mutations have been described by Cotton.[56] There are two recent techniques that can accurately detect mutations and can be used in formats that have high throughput: one is based on the use of homogeneous hybridization probes coupled to PCR, and the second is a cleavage assay that utilizes the specificity of 'flap' endonuclease from *Pyrococcus furiosus*. The first technique employs real-time PCRs to amplify a region of the DNA that contains the single-nucleotide variation of the interest in the presence of a pair of homogeneous hybridization probes. Either molecular beacons or Taq-Man probes can be employed in these assays. One of the probes is specific for one allele and is labeled with one color, whereas the second probe is specific to the other allele and is labeled with a different color. The color that develops indicates which allele is present.[57–59] Since 96 reactions can be performed in parallel and the analysis is performed as the PCR proceeds, the throughput of these assays is so high that the genotypes of thousands of samples can be determined every week.

The cleavage assays utilize a flap endonuclease from *Pyrococcus furiosus* to cleave Y-shaped structures that are created by hybridizing two oligonucleotides to the region of the target at the site of the mutation (Figure 9). These Y-shaped structures are akin to those that arise naturally during DNA synthesis when the 3′ end of a growing chain displaces the sequence that is bound to the template. The sequence in the displaced strand and the newly synthesized strand are the same and undergo back-and-forth branch migration. A number of enzymes

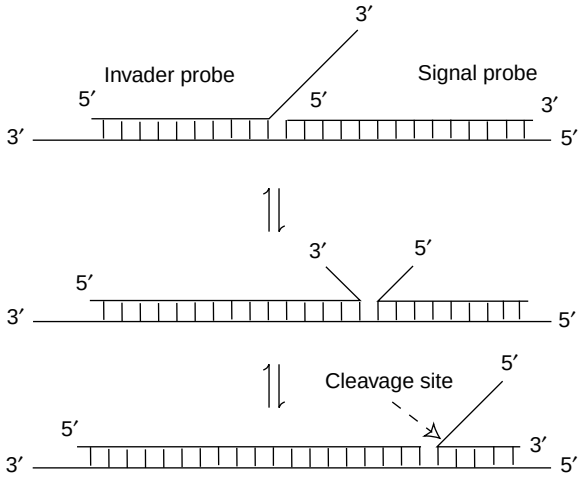

Figure 9 Schematic representation of the branch migration and cleavage in the mutation detection assay using flap endonucleases. The invader probe has the same sequence at its 3′ end as the signal probe has at its 5′ end. Flap endonuclease can cleave the signal probe only when the nucleotide at the 3′ end of the invader probe is perfectly paired.

associated with DNA repair, including DNA polymerase and flap endonuclease from *Pyrococcus furiosus*, have the capacity to excise these redundant portions of displaced DNA by nicking the displaced strand where it meets the template strand. The artificial Y structures are created by binding an 'invader oligonucleotide' and a 'signal oligonucleotide' to the target at the same time (Figure 9). The 3′ end of the invader oligonucleotide share the same sequence with the 5′ end of the signal oligonucleotide, permitting branch migration to create a substrate of the flap endonuclease. Cleavage results in a detectable signal, as FRET between the terminal labels of the signal probes is disrupted. Y-structures possessing a mismatched base pair at the junction of the two arms can not be cleaved. Hence, by using two variants of the signal probe where each is specific for a different allele and has a different reporter fluorophore, it is possible to discriminate between two alleles. The system is highly specific, because the specificity of the flap endonuclease is added to the specificity of hybridization probes.[55]

9 TAKING DNA PROBES INTO A PROTEIN WORLD

All of the hybridization assays described so far make use of the ability of nucleic acid strands to anneal specifically to their complements. A new kind of DNA probe is being developed that expands the utility of nucleic acids beyond their traditional role as hybridization probes. These oligonucleotide probes do not hybridize to a complementary nucleic acid, but rather they bind to a protein target.[60] These oligonucleotide probes assume a conformation that allows them to fit within a pocket of their protein target. These oligonucleotides are created and used in the image of antibodies. From the vast field of possible shapes and structures inherent in a partially or fully randomized collection of oligonucleotides, it is possible to isolate an oligonucleotide that can bind to any target of interest. A library of all possible sequences is created on a DNA synthesizer and an in vitro selection scheme is used for the isolation of a desired molecule. Methods of in vitro evolution and selection which allow for the isolation of a particular sequence from hundreds of billions of other possibilities were first introduced in 1967.[61] These methods have recently been invigorated by the use of PCRs. Using a scheme termed SELEX (systematic evolution of ligands by exponential enrichment),[62] it has been possible to find nucleic acid ligands for a wide variety of molecules including proteins and organic dyes.[63] In this scheme, a library of oligonucleotides is panned for the molecules that exhibit an affinity for a desired target. Those that are capable of binding to the target are

exponentially amplified. This process is repeated until a molecule of desired specificity and affinity is found. For example, SELEX has been used to find a DNA ligand for vascular endothelial growth factor, a protein involved in angiogenesis. Using this DNA ligand as a detector probe and an antibody as a capture probe, an assay has been developed for the detection of vascular endothelial growth factor.[64] The DNA ligand was tagged with a fluorescein moiety in order to produce a chemiluminescent signal via an antifluorescein antibody and an alkaline phosphatase detection system. This assay exhibits a specificity equivalent to common immunoassays.

The advantage of using nucleic acid ligands for proteins, as opposed to antibodies, is that animals are not required for the development of the diagnostic reagent. It may be possible to find ligands for poorly immunogenic targets. Knowledge of the sequence of a nucleic acid ligand allows its accurate reproduction anywhere in the world since its production is not subjected to animal-to-animal variations. For applications and detection methods in which the bulk of the antibodies is a limitation, the small size of nucleic acid ligands offers an advantage. In the conventional nucleic acid-based diagnostics, one can easily identify probes that characterize a pathogenic species or strain by using the knowledge of the nucleotide sequence of relatively invariant regions of genomes. Similarly, finding the antibodies for conventional immunoassays is also simple, because all one needs is an antigen, a syringe, and an animal. Based on the diversity of ligands isolated by SELEX, developers of this technique predict that nucleic acid ligands will also prove to be as general and easy to find as conventional nucleic acid probes and antibodies.

10 FUTURE PROSPECTS

The technique of hybridization has played a key role in the exploration of the molecular biology of the cell. Improvements that simplify its implementation and enhance its sensitivity are taking it to the realm of clinical diagnostics, where it is important that the technique is simple to perform and is amenable for automation. These developments will make it possible to map the distribution of mRNAs in living cells by microscopy and study gene expression in an ensemble of living cells with fluorescence-activated cell sorters.[32] We shall see innovative use of DNA–DNA hybridization in the development of new materials for the construction of nanostructures.[65] In addition, we shall witness the solution of computational problems by the utilization of the manifest specificity of oligonucleotides.[66]

ACKNOWLEDGMENTS

Many ideas expressed in this article were developed in a long-term collaboration with Fred Russell Kramer. This work was supported by National Institutes of Health grants HL-43521 and ES10536.

ABBREVIATIONS AND ACRONYMS

FRET	Fluorescence Resonance Energy Transfer
MDV-1	Midivariant-1
PCR	Polymerase Chain Reaction
SELEX	Systematic Evolution of Ligands by Exponential Enrichment

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Imaging

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Nucleic Acid Analysis in Clinical Chemistry

Forensic Science **(Volume 5)**
DNA Extraction Methods in Forensic Analysis • Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Aptamers • DNA Molecules, Properties and Detection of Single • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Fluorescence In Situ Hybridization • Nucleic Acid Structural Energetics • PNA and Its Applications • Polymerase Chain Reaction and Other Amplification Systems • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

REFERENCES

1. J. SantaLucia, Jr, 'A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics', *Proc. Natl. Acad. Sci. USA*, **95**, 1460–1465 (1998).
2. B.D. Hall, S. Spiegelman, 'Sequence Complementarity of T2-DNA and T2-specific RNA', *Proc. Natl. Acad. Sci. USA*, **47**, 137–146 (1961).

3. S. Spiegelman, 'Hybrid Nucleic Acids', *Sci. Am.*, May, 183–195 (1964).
4. D. Gillespie, S. Spiegelman, 'A Quantitative Assay for DNA–RNA Hybrids with DNA Immobilized on a Membrane', *J. Mol. Biol.*, **12**, 829–842 (1965).
5. G.H. Keller, M.M. Manak, *DNA Probes*, Stockton Press, New York, 1993.
6. D. Nathans, H.O. Smith, 'Restriction Endonucleases in the Analysis and Restructuring of DNA Molecules', *Annu. Rev. Biochem.*, **44**, 273–293 (1975).
7. S.N. Cohen, A.C. Chang, H.W. Boyer, R.B. Helling, 'Construction of Biologically Functional Bacterial Plasmids in Vitro', *Proc. Natl. Acad. Sci. USA*, **70**, 3240–3244 (1973).
8. E.M. Southern, 'Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis', *J. Mol. Biol.*, **98**, 503–517 (1975).
9. W.T. McAllister, C. Morris, A.H. Rosenberg, F.W. Studier, 'Utilization of Bacteriophage T7 Late Promoters in Recombinant Plasmids During Infection', *J. Mol. Biol.*, **153**, 527–544 (1981).
10. R.J. Britten, D.E. Kohne, 'Repeated Sequences in DNA', *Science*, **161**, 529–540 (1968).
11. S.M. Tilghman, D.C. Tiemeier, J.G. Seidman, B.M. Peterlin, M. Sullivan, J.V. Maizel, P. Leder, 'Intervening Sequence of DNA Identified in the Structural Portion of a Mouse Beta-globin Gene', *Proc. Natl. Acad. Sci. USA*, **75**, 725–729 (1978).
12. A.L. Joyner, T. Kornberg, K.G. Coleman, D.R. Cox, G.R. Martin, 'Expression During Embryogenesis of a Mouse Gene with Sequence Homology to the *Drosophila* Engrailed Gene', *Cell*, **43**, 29–37 (1985).
13. D.V. Morrissey, M. Lombardo, J.K. Eldredge, K.R. Kearney, E.P. Groody, M.L. Collins, 'Nucleic Acid Hybridization Assays Employing dA-tailed Capture Probes. I. Multiple Capture Methods', *Anal. Biochem.*, **181**, 345–359 (1989).
14. A. Lagerkvist, J. Stewart, M. Lagerstrom-Fermer, U. Landegren, 'Manifold Sequencing: Efficient Processing of Large Sets of Sequencing Reactions', *Proc. Natl. Acad. Sci. USA*, **91**, 2245–2249 (1994).
15. Z. Guo, R.A. Guilfoyle, A.J. Thiel, R. Wang, L.M. Smith, 'Direct Fluorescence Analysis of Genetic Polymorphisms by Hybridization with Oligonucleotide Arrays on Glass Supports', *Nucleic Acids Res.*, **22**, 5456–5465 (1994).
16. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
17. P.J. Bifani, B.B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M.L. Lutfey, S.L. Moghazeh, W. Eisner, T.M. Daniel, M.H. Kaplan, J.T. Crawford, J.M. Musser, B.N. Kreiswirth, 'Origin and Interstate Spread of a New York City Multidrug-resistant *Mycobacterium tuberculosis* Clone Family', *J. Am. Med. Assoc.*, **275**, 452–457 (1996).
18. A.C. Pease, D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, S.P. Fodor, 'Light-generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis', *Proc. Natl. Acad. Sci. USA*, **91**, 5022–5026 (1994).
19. S. Singh-Gasson, R.D. Green, Y. Yue, C. Nelson, F. Blattner, M.R. Sussman, F. Cerrina, 'Maskless Fabrication of Light-directed Oligonucleotide Microarrays Using a Digital Micromirror Array', *Nature Biotechnol.*, **17**, 974–978 (1999).
20. C.O. Yehle, W.L. Patterson, S.J. Boguslawski, J.P. Albarella, K.F. Yip, R.J. Carrico, 'A Solution Hybridization Assay for Ribosomal RNA from Bacteria Using Biotinylated DNA Probes and Enzyme-labeled Antibody to DNA : RNA', *Mol. Cell. Probes*, **1**, 177–193 (1987).
21. A.C. Syvanen, M. Laaksonen, H. Soderlund, 'Fast Quantification of Nucleic Acid Hybrids by Affinity-based Hybrid Collection', *Nucleic Acids Res.*, **25**, 5037–5048 (1986).
22. P.R. Langer, A.A. Waldrop, D.C. Ward, 'Enzymatic Synthesis of Biotin-labeled Polynucleotides: Novel Nucleic Acid Affinity Probes', *Proc. Natl. Acad. Sci. USA*, **78**, 6633–6637 (1981).
23. A.T. Haase, K. Henry, M. Zupancic, G. Sedgewick, R.A. Faust, H. Melroe, W. Cavert, K. Gebhard, K. Staskus, Z.Q. Zhang, P.J. Dailey, H.H. Balfour, Jr, A. Erice, A.S. Perelson, 'Quantitative Image Analysis of HIV-1 Infection in Lymphoid Tissue', *Science*, **274**, 985–989 (1996).
24. L.J. Kricka, 'Chemiluminescent and Bioluminescent Techniques', *Clin. Chem.*, **37**, 1472–1481 (1991).
25. J.H.M. Knoll, P. Lichter, 'In Situ Hybridization and Detection Using Nonisotopic Probes', *Curr. Protocols Mol. Biol.*, **3**, 14.7.1–14.7.14 (1998).
26. C.R. Cantor, 'Lighting Up Hybridization', *Nature Biotechnol.*, **14**, 264 (1996).
27. M.J. Heller, L.E. Morrison, 'Chemiluminescent and Fluorescent Probes for DNA Hybridization Systems', in *Rapid Detection and Identification of Infectious Agents*, eds. D.T. Kingsbury, S. Falkow, Academic Press, New York, 345–356, 1985.
28. R.A. Cardullo, S. Agrawal, C. Flores, P.C. Zamecnik, D.E. Wolf, 'Detection of Nucleic Acid Hybridization by Nonradiative Fluorescence Resonance Energy Transfer', *Proc. Natl. Acad. Sci. USA*, **85**, 8790–9794 (1988).
29. L.E. Morrison, T.C. Halder, L.M. Stols, 'Solution-phase Detection of Polynucleotides Using Interacting Fluorescent Labels and Competitive Hybridization', *Anal. Biochem.*, **183**, 231–244 (1989).
30. L.E. Morrison, L.M. Stols, 'Sensitive Fluorescence-based Thermodynamic and Kinetic Measurements of DNA Hybridization in Solution', *Biochemistry*, **32**, 3095–3104 (1993).
31. K.M. Parkhurst, L.J. Parkhurst, 'Kinetic Studies by Fluorescence Resonance Energy Transfer Employing a Double-labeled Oligonucleotide: Hybridization to the Oligonucleotide Complement and a Single-stranded DNA', *Biochemistry*, **34**, 285–292 (1995).

32. S. Tyagi, F.R. Kramer, 'Molecular Beacons: Probes that Fluoresce Upon Hybridization', *Nature Biotechnol.*, **14**, 303–308 (1996).
33. S. Tyagi, D.P. Bratu, F.R. Kramer, 'Multicolor Molecular Beacons for Allele Discrimination', *Nature Biotechnol.*, **16**, 49–53 (1998).
34. R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, 'Enzymatic Amplification of Beta-globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia', *Science*, **230**, 1350–1354 (1985).
35. G.T. Walker, M.C. Little, J.G. Nadeau, D.D. Shank, 'Isothermal In Vitro Amplification of DNA by a Restriction Enzyme/DNA Polymerase System', *Proc. Natl. Acad. Sci. USA*, **89**, 392–396 (1992).
36. D.Y. Kwoh, G.R. Davis, K.M. Whitfield, H.L. Chappelle, L.J. DiMichele, T.R. Gingeras, 'Transcription-based Amplification System and Detection of Amplified Human Immunodeficiency Virus Type 1 with a Bead-based Sandwich Hybridization Format', *Proc. Natl. Acad. Sci. USA*, **86**, 1173–1177 (1989).
37. C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, 'Real Time Quantitative PCR', *Genome Methods*, **6**, 986–994 (1996).
38. G. Leone, H. van Schijndel, B. van Gemen, F.R. Kramer, C.D. Schoen, 'Molecular Beacon Probes Combined with Amplification by NASBA Enable Homogeneous, Real Time Detection of RNA', *Nucleic Acids Res.*, **26**, 2150–2155 (1998).
39. M.J. Lay, C.T. Witter, 'Real Time Fluorescence Genotyping of Factor V Leiden Mutation During Rapid-cycle PCR', *Clin. Chem.*, **43**, 2262–2267 (1997).
40. P.M. Holland, R.D. Abramson, R. Watson, D.H. Gelfand, 'Detection of Specific Polymerase Chain Reaction Product by Utilizing the 5′–3′', Exonuclease Activity of *Thermus aquaticus* DNA Polymerase', *Proc. Natl. Acad. Sci. USA*, **88**, 7276–7280 (1991).
41. L.G. Lee, C.R. Connell, W. Bloch, 'Allelic Discrimination by Nick-translation PCR with Fluorogenic Probes', *Nucleic Acids Res.*, **21**, 3761–3766 (1993).
42. B.C. Chu, F.R. Kramer, L.E. Orgel, 'Synthesis of an Amplifiable Reporter RNA for Bioassays', *Nucleic Acids Res.*, **14**, 5591–5603 (1986).
43. F.R. Kramer, P.M. Lizardi, 'Replicatable RNA Reporters', *Nature (London)*, **339**, 401–402 (1989).
44. E.A. Miele, D.R. Mills, F.R. Kramer, 'Autocatalytic Replication of a Recombinant RNA', *J. Mol. Biol.*, **171**, 281–295 (1983).
45. H. Lomeli, S. Tyagi, C.G. Pritchard, P.M. Lizardi, F.R. Kramer, 'Quantitative Assays Based on the Use of Replicatable Hybridization Probes', *Clin. Chem.*, **35**, 1826–1831 (1989).
46. S. Tyagi, U. Landegren, M. Tazi, P.M. Lizardi, F.R. Kramer, 'Extremely Sensitive, Background-free Gene Detection Using Binary Probes and Beta Replicase', *Proc. Natl. Acad. Sci. USA*, **93**, 5395–5400 (1996).
47. M.S. Urdea, J. Kolberg, J. Clyne, J.A. Running, D. Besemer, B. Warner, R. Sanchez-Pescador, 'Application of a Rapid Non-radioisotopic Nucleic Acid Analysis System to the Detection of Sexually Transmitted Disease-causing Organisms and their Associated Antimicrobial Resistances', *Clin. Chem.*, **35**, 1571–1575 (1989).
48. M. Nilsson, H. Malmgren, M. Samiotaki, M. Kwiatkowski, B.P. Chowdhary, U. Landegren, 'Padlock Probes: Circularizing Oligonucleotides for Localized DNA Detection', *Science*, **265**, 2085–2088 (1994).
49. A. Fire, S.-Q. Xu, 'Rolling Replication of Short DNA Circles', *Proc. Natl. Acad. Sci. USA*, **92**, 4641–4645 (1995).
50. D. Zhang, M. Brandwein, T.C.H. Hsuih, H. Li, 'Amplification of Target-specific, Ligation-dependent Circular Probe', *Gene*, **211**, 277–285 (1998).
51. P.M. Lizardi, X. Huang, Z. Zhu, P. Bray-Ward, D.C. Thomas, D.C. Ward, 'Mutation Detection and Single-molecule Counting Using Isothermal Rolling-circle Amplification', *Nature Genet.*, **19**, 225–232 (1998).
52. G. Bonnet, S. Tyagi, A. Libchaber, F.R. Kramer, 'Thermodynamic Basis of the Enhanced Specificity of Structured DNA Probes', *Proc. Natl. Acad. Sci. USA*, **96**, 6171–6176 (1999).
53. S. Wang, A.E. Friedman, E.T. Kool, 'Origins of High Sequence Selectivity: a Stopped-flow Kinetics Study of DNA/RNA Hybridization by Duplex- and Triplex-forming Oligonucleotides', *Biochemistry*, **34**, 9774–9784 (1995).
54. R. Youil, B.W. Kemper, R.G. Cotton, 'Screening for Mutations by Enzyme Mismatch Cleavage with T4 Endonuclease VII', *Proc. Natl. Acad. Sci. USA*, **92**, 87–91 (1995).
55. V. Lyamichev, A.L. Mast, J.G. Hall, J.R. Prudent, M.W. Kaiser, T. Takova, R.W. Kwiatkowski, T.J. Sander, M. de Arruda, D.A. Arco, B.P. Neri, M.A. Brow, 'Polymorphism Identification and Quantitative Detection of Genomic DNA by Invasive Cleavage of Oligonucleotide Probes', *Nature Biotechnol.*, **17**, 292–296 (1999).
56. R. Cotton, 'Mutation Detection', Oxford University Press, Oxford, 1997.
57. L.G. Kostrikis, S. Tyagi, M.M. Mhlanga, D.D. Ho, F.R. Kramer, 'Spectral Genotyping of Human Alleles', *Science*, **279**, 1228–1229 (1998).
58. S.A. Marras, F.R. Kramer, S. Tyagi, 'Multiplex Detection of Single-nucleotide Variations Using Molecular Beacons', *Genet. Anal.*, **14**, 151–156 (1999).
59. P.A. Morin, R. Saiz, A. Monjazeb, 'High-throughput Single Nucleotide Polymorphism Genotyping by Fluorescent 5′ Exonuclease Assay', *Biotechniques*, **27**, 538–542 (1999).
60. S.D. Jayasena, 'Aptamers: an Emerging Class of Molecules that Rival Antibodies in Diagnostics', *Clin. Chem.*, **45**, 1628–1650 (1999).
61. D.R. Mills, R.L. Peterson, S. Spiegelman, 'An Extracellular Darwinian Experiment with a Self-duplicating Nucleic

Acid Molecule', *Proc. Natl. Acad. Sci. USA*, **58**, 217–224 (1967).
62. C. Tuerk, L. Gold, 'Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase', *Science*, **249**, 505–510 (1990).
63. L. Gold, B. Polisky, O. Uhlenbeck, M. Yarus, 'Diversity of Oligonucleotide Functions', *Annu. Rev. Biochem.*, **64**, 763–797 (1995).
64. D.W. Drolet, L. Moon-McDermott, T.S. Roming, 'An Enzyme-linked Oligonucleotide Assay', *Nature Biotechnol.*, **14**, 1021–1025 (1996).
65. J.J. Storhoff, C.A. Mirkin, 'Programmed Material Synthesis with DNA', *Chem. Rev.*, **99**, 1849–1862 (1999).
66. L.M. Adleman, 'Molecular Computation of Solutions to Combinatorial Problems', *Science*, **266**, 1021–1024 (1994).

DNA Structures of Biological Relevance, Studies of Unusual Sequences

Yuh-Hwa Wang
University of Medicine and Dentistry of New Jersey, Piscataway, USA

1 Introduction 1
2 Structure and Properties of DNA Containing Bulges and Mismatches 2
2.1 Gel Electrophoretic Analysis of DNAs Containing Base Bulges and Mismatches 2
2.2 Visualization of Three-base Bulged DNAs by Electron Microscopy and Estimation of the Kinking Angle Produced by the Bulges 3
2.3 Binding of RecA Protein to DNAs Containing Base Bulges and Mismatches 4
2.4 Summary 5
3 Trinucleotide Repeating DNA, Nucleosome Positioning, and Human Genetic Diseases 6
3.1 CTG Repeats in Myotonic Dystrophy 7
3.2 Expanded CCG Repeats in Fragile X Syndrome and Folate-sensitive Fragile Sites 12
3.3 Summary 19
Acknowledgments 19
Abbreviations and Acronyms 19
Related Articles 19
References 19

DNA molecules with unusual structures can have significant consequences in many DNA metabolic processes. This article summarizes structural studies of DNA molecules containing bulged and mismatched bases and trinucleotide repeating DNAs. Extra base bulges and mismatches in double-stranded DNA can arise from replication errors or imperfect recombination. If these lesions escape the repair system, they would subsequently cause mutagenesis. Using a combination of gel electrophoresis, electron microscopy (EM), chemical modification, and probing with RecA protein, the structure of bulged and mismatched DNAs was examined and factors such as base composition and flanking base sequence have been shown to influence the overall structure of these molecules.

Expansion of trinucleotide repeats is a major cause of several inherited neurodegenerative diseases. Described here are findings that show, by in vitro nucleosome assembly assay, that the expanded CTG triplet sequences derived from myotonic dystrophy (DM) patients form the most stable naturally occurring nucleosomes that are known. In contrast, long tracts of repeating CCG triplets (expanded in fragile X syndrome (FraX) patients) resist nucleosome assembly. This information could provide critical clues about how these unusual DNAs cause harmful biological effects in living cells.

1 INTRODUCTION

DNA molecules can possess structural polymorphism depending on base composition, sequence arrangement, and many other factors. Various structures of DNA molecules could determine their fate in DNA metabolic processes. Our research interest is to understand the structure and function of unusual DNA sequences which could cause mutagenesis and diseases in living cells.

Extra base bulges and mismatches in double-stranded DNA can arise from recombination between sequences that are not fully homologous or from errors of replication. Bulged bases resulting from replicative errors are considered to play an important role in frame-shift mutagenesis.[1] In *Escherichia coli*, mismatches can be repaired by the methyl-directed repair system and several short-patch mismatch repair pathways, and the repair efficiency depends on the type of mismatch and the sequence context.[2] Genetic studies have shown that single-base bulges can also be repaired by the methyl-directed repair system in *E. coli*[3] and, using cell extracts, this repair system has been shown to repair DNAs containing 1–3-base bulges as efficiently as a G·T mismatch.[4]

Human homologs of mismatch repair proteins have been identified: hMSH2-6 are bacterial MutS homologs and hMLH1, hPMS1, and hPMS2 are homologs of bacterial MutL.[5] Mutations of these genes have been shown to be responsible for almost all patients with hereditary nonpolyposis colorectal cancer, in which simple repetitive DNA sequences have a high incidence of mutation.[6] Biochemical and genetic studies have demonstrated that these human repair proteins possess different affinities for DNA containing bulges and mismatches depending upon the type and number of bases. Therefore, it is critical to understand the structure of DNA containing bulges and mismatches and examine the determining factors such as base composition, flanking

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

sequences and number of bases. This article describes our studies using a variety of biochemical assays to investigate the structure and properties of short duplex DNA molecules containing bulged and mismatched bases.

Since 1991, 12 human genetic diseases including DM and FraX have been shown to be caused by expansion of trinucleotide repeats.[7–10] In each disease, an expanded block of repeating trinucleotides is present in or near a gene whose loss or alteration leads to the disease. The repeating triplet blocks present in normal individuals at these loci are heterogeneous in size and have less than 35 copies. In the afflicted person there can be as many as 2000 repeats. The diseases have been divided into two groups based on the location of the expanded triplet within the disease gene. The first group includes DM, FraX and Friedreich's ataxia, in which the triplet blocks are not in the protein coding regions. In these diseases, the triplet blocks must expand beyond ~100 repeats for the full disease symptoms to appear and perfect triplet runs are much more deleterious than longer blocks that contain scattered interruptions.[10–15] In the second group, including Huntington's and Kennedy's diseases, the expanded triplet repeats are within the protein coding sequences. Thus, expansion by a smaller number of CAG/CTG repeats that encode polyglutamine tracts in the disease protein alter or abolish its activity. The triplet diseases frequently show unusual genetics (anticipation), in which the severity of the disease increases with succeeding generations. Disease symptoms also correlate with the size of the expanded triplet block.

For each triplet disease, two major questions need to be answered: what is the mechanism of expansion of triplet repeats?; and how does the expanded triplet repeat in each disease lead to clinical symptoms? In this article, we summarize our studies, showing that repeating sequences CTG (in DM) and CCG (in FraX) have extremely unusual properties in forming nucleosomes. Thus, the biological manifestations of DM and FraX diseases are likely to be due in part to the unusual properties of DNAs containing very long triplet repeats and the resulting unusual chromatin structures.

2 STRUCTURE AND PROPERTIES OF DNA CONTAINING BULGES AND MISMATCHES

Structural studies using nuclear magnetic resonance (NMR) spectroscopy suggest that bulged bases can exist in either a stacked-in or looped-out conformation depending on the base composition of the bulge, temperature, and other factors.[16–21] Gel electrophoretic studies in our laboratory (see below) and the laboratories of Lilley and Crothers have shown that bulges of 1–5-bases produce kinks in DNA, which result in decreased mobilities of bulged DNAs in polyacrylamide gels.[22–25] Whereas single-base bulges produce kinks in DNA, studies using NMR,[26] X-ray crystallography[27] and gel electrophoresis[22,23,28] show that single-base mismatches do not. Furthermore, DNA containing multiple base mismatches shows no retardation of gel mobility in 15% polyacrylamide gels.[22] These results suggest that mismatched bases remain stacked into the helix and produce at most only subtle perturbations in the DNA helix.[28]

We have applied a combination of gel electrophoresis, EM, chemical modification, and RecA probing techniques to understand further the structure and properties of short duplex DNA molecules containing either bulges or mismatches.

2.1 Gel Electrophoretic Analysis of DNAs Containing Base Bulges and Mismatches

Gel electrophoresis and EM have been used to examine the effects of 1–4 extra bases on the structure of DNA. DNA duplexes of 32 basepairs (bps) were synthesized containing unique single-stranded ends such that upon ligation the 32-mers can only join in a head-to-tail fashion. The 32-mers were synthesized to contain various arrangements of single or multiple extra bases. The effect of the extra bases on each 32-mer was amplified by the ligation process which places the extra bases in phase with the helix in the linear and circular multimers. It was observed that single extra bases caused a marked retardation of the electrophoretic mobility of the linear multimers, indicating that the bulged bases kinked the DNA. Analysis of the size distribution of the circular multimers confirmed this observation. DNAs with two to four extra bases at a single site showed even greater electrophoretic retardations and yielded smaller circular multimers. Placing the extra base(s) out of phase with the helix created kinks which canceled each other and produced no net retardation.[23]

The effects of bulge composition and flanking sequence on the kinking of DNA by bulged bases were further investigated in this way. Synthetic 30 bp duplex DNAs containing two single-base bulges spaced by 10 bp were ligated head-to-tail, and their electrophoretic behavior in highly cross-linked gels was examined. All bulge-containing DNAs showed marked electrophoretic retardations compared with non-bulge-containing DNA. Regardless of the sequence of the flanking bps, purine bulges produced greater retardations than pyrimidine bulges. Furthermore, bulged C and Ts produced the same retardations as did bulged G and As. Bulged DNA containing different flanking bps showed marked differences

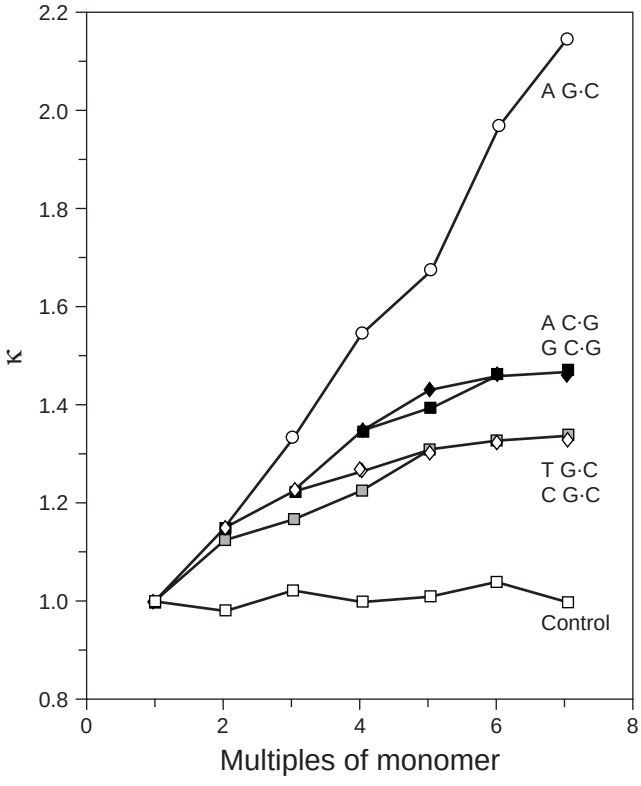

Figure 1 Effect of different bulged bases on the electrophoretic mobility of DNA. DNA duplexes were ligated head-to-tail, electrophoresed through a 15% polyacrylamide gel, and visualized by staining with ethidium bromide. The apparent size of the DNA in each band of a gel was determined relative to the mobility of *Hae*III-digested φX 174 DNA fragments. The κ value is defined as the ratio of the apparent size of a DNA fragment to its actual size. These κ values were plotted against the actual size of the DNA as multiples of the monomer length.

in electrophoretic mobility, for example, the G·C and C·G flanking bps were seen to have very different effects (Figure 1). Thus, these results imply an important role of base stacking in determining how neighboring bps influence the kinking of DNA by a single-base bulge.[28]

We have applied the same techniques to examine DNAs containing single-base mismatches. Twenty-four DNAs containing eight possible mismatches with three different flanking sequences (A·T, G·C, or a combination of the two) were analyzed on 15% polyacrylamide gels and showed no detectable retardation in electrophoretic mobility.[23]

2.2 Visualization of Three-base Bulged DNAs by Electron Microscopy and Estimation of the Kinking Angle Produced by the Bulges

The demonstration that bulged bases produce electrophoretic retardations has provided a means of screening for small chromosomal deletions or insertions. For example, nearly 70% of the carriers of cystic fibrosis (CF) in the western Caucasian population possess a single three-base deletion (ΔF_{508}) in the CF gene on chromosome 7.[29] Several groups have used the polymerase chain reaction (PCR) to amplify small regions of DNA (100 bp) around codon 508 of the CF gene;[30–32] if the DNA is from a carrier of the disease and thus a heterozygote, amplification produces four single-stranded DNA fragments, two that contain the deletion and two that do not. Upon annealing the four DNA strands, two homoduplexes and two heteroduplexes (one containing a CTT bulge and the other containing an AAG bulge) are produced, and the latter species can be detected by their retarded electrophoretic mobility. This approach has also been used to screen for a four-base insertion mutation in Tay–Sachs disease.[33,34]

For these genetic screens to be useful and to understand further the structure of bulge-containing DNA, it is important to measure directly the kinking angle produced in a DNA by a bulged base and to examine the distribution of kinking angles within the population of bulged DNAs. To approach this problem we obtained DNA from a CF carrier. The CF homoduplex and bulge-containing heteroduplex molecules were prepared as described above. These DNAs were examined by EM for the presence of visible kinks. Twice as many bulge-containing

Table 1 EM analysis of PCR-amplified CF DNA (2584 molecules scored)[a]

DNA	Molecules straight (%)	Molecules with a bend at the center (%)
PCR-amplified:		
Heteroduplexes	58	42 (14)
Homoduplexes	76	24 (3)
75 bp DNA	83	17 (2)
Synthetic:		
CTT-bulged DNA	62	38 (13)
AAG-bulged DNA	61	39 (12)
85 bp DNA	81	19 (3)
82 bp DNA	81	19 (3)

[a] DNAs amplified by PCR[35] from the DNA of a CF carrier were annealed to produce homoduplexes and heteroduplexes and fractionated on polyacrylamide gels. A 75 bp DNA with no bulged bases served as a control. In a separate experiment, four DNAs were prepared from synthetic oligonucleotides: a DNA with an AAG bulge in the one strand, a DNA with a CTT bulge in the other strand, an 85 bp DNA with no deletion in either strand, and an 82 bp DNA with the three-base deletion in both strands. The DNAs were prepared for EM as described elsewhere.[36] In fields of molecules, the number that were straight or bent by less than 30° were scored against those which appeared to have a central kink of 30° or greater, as judged by eye. Number in parentheses indicates the percentage of the total molecules that contain a kink 40–60% from one end of the DNA and in which the kink measures between 70 and 90°. A kink of 0° represents a perfectly straight molecule.

DNAs were seen to have a central kink as compared with the non-bulged controls (Table 1). Examination of the distribution of kinking angles showed that the bulged population contained 5–7-fold more molecules with a central kink of $80 \pm 10^{\circ}$ than did the control molecules (Figure 2a–c and Table 1).

Examination of the DNAs synthesized to contain just the CTT or AAG bulges and the respective homoduplexes revealed a similar distribution of kinked and straight molecules as seen in the PCR-amplified DNA (Table 1), and the angular distribution of kinks (data not shown) was nearly identical for both.

2.3 Binding of RecA Protein to DNAs Containing Base Bulges and Mismatches

Previous studies have shown that thymine dimers,[37] (6–4) photoproducts,[38] psoralen cross-links,[39] and intercalating drugs[40,41] can promote the nucleation of RecA protein binding on double-stranded DNAs in the presence of adenosine-5′-*O*-(3-thiotriphosphate) (ATPγS), a poorly hydrolyzed analog of ATP. Other structural variations such as B–Z junctions,[42,43] changes in DNA superhelicity,[44] and an A·T-rich sequence[45] also enhance the binding of RecA protein to double-stranded DNAs. These observations suggest that RecA protein recognizes perturbations in the DNA helix. To test the hypothesis that RecA protein is intimately involved in repair,[46,47] it would be of value to examine the interaction of RecA protein with two lesions that are frequent products of DNA metabolism: bulged bases and single-base mismatches.

Using a nitrocellulose filter binding assay, we examined DNA duplexes containing 1–4-base cytosine bulges and the control DNA. Upon incubation with RecA protein in the presence of ATPγS, DNA with a one-base cytosine bulge showed similar kinetics of RecA binding as compared with DNA lacking a bulge (Figure 3). RecA bound to DNA containing two- or three-base cytosine bulges with a similar kinetics and both showed higher levels of binding compared with the control DNA. DNA containing a four-base cytosine bulge showed even higher binding (~30% more than the control). These results suggest that RecA protein recognizes the perturbation produced by the bulged bases.

Gel electrophoretic studies have demonstrated that single purine bulges in DNA cause greater retardation than single pyrimidine bulges (see above). This suggests that purine bulges introduce kinks of a greater angle into the DNA helix as compared with pyrimidine bulges. To determine whether any correlation exists between the kinking angle and RecA binding, a set of DNAs containing 1–4-base adenine bulges was investigated. Interestly, we found that RecA bound to DNA containing

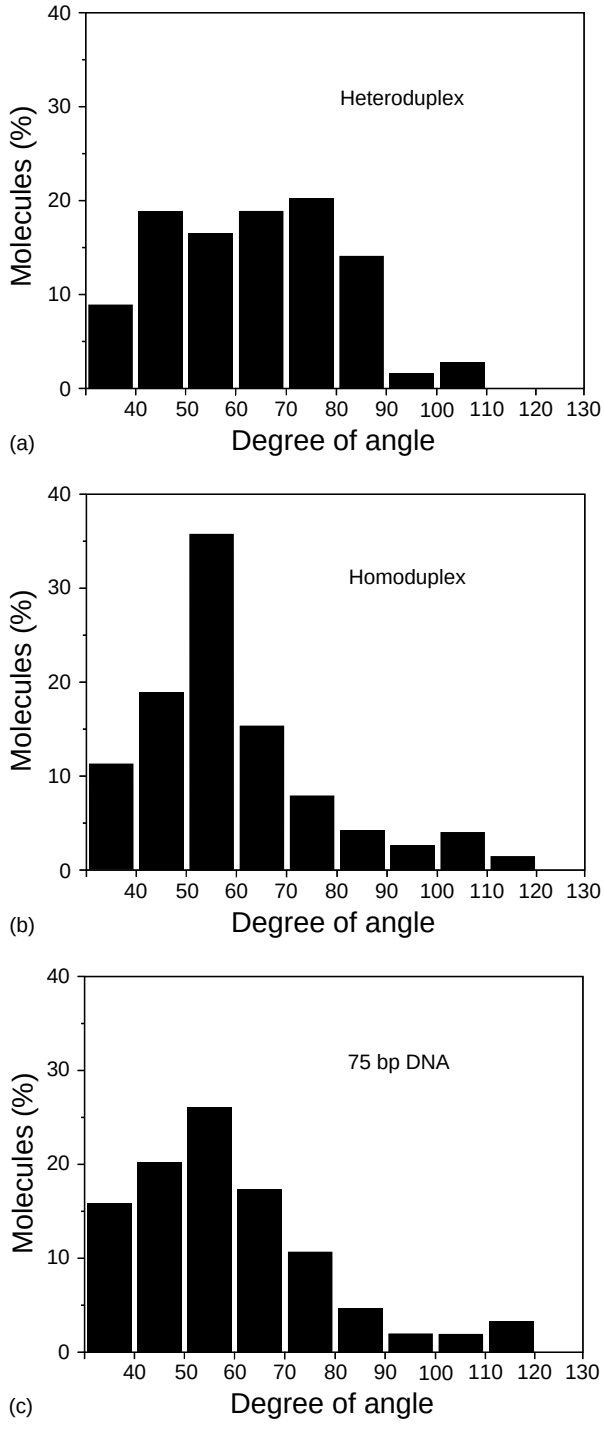

Figure 2 Angular distribution of kinking angles. The CF DNAs from PCR amplification, a pool of both heteroduplexes (a), both homoduplexes (b), and the 75 bp control DNA (c) were prepared for EM. From the micrographs, the kinking angle in each kinked DNA was measured, in addition to its position relative to the DNA ends. For those molecules in which the kink was within 10% of the center, the angle of the kink was plotted. Approximately 100 bent molecules for each sample were measured. A cut-off of angles less than 30° was imposed by eye.

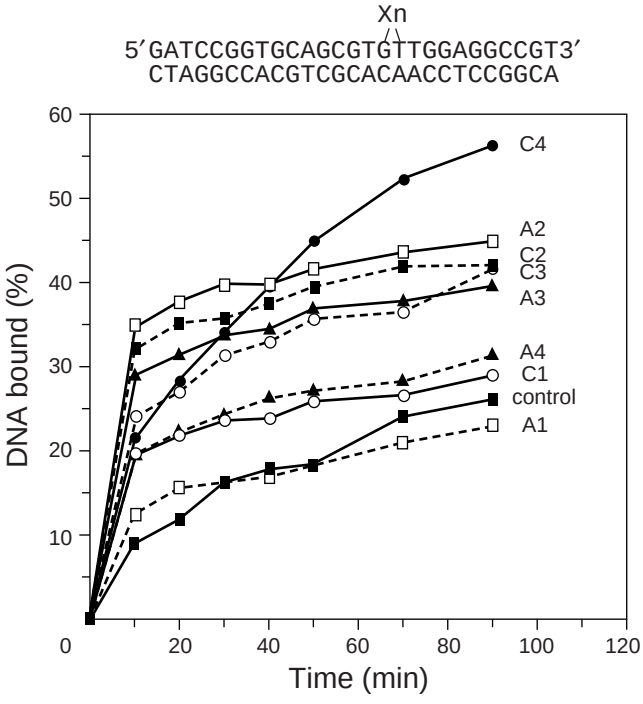

Figure 3 Binding of RecA to DNAs containing 1–4-base cytosine or adenine bulges. Duplex DNAs containing central 1–4-base cytosine bulges (C1, C2, C3, or C4), 1–4-base adenine bulges (A1, A2, A3, or A4), and the control DNA (control) were prepared by annealing the ^{32}P-labeled upper strand (the strand containing extra bases) with the complementary lower strand. The labeled DNA (4 μg mL^{-1}) was incubated with RecA protein at a concentration of 240 μg mL^{-1} in the presence of ATPγS at a final concentration of 200 μM. The efficiency of RecA binding was determined by the filter binding assay.

a four-base cytosine bulge much more efficiently than to DNA with a four-base adenine bulge, although four-base adenine bulges produce kinks of a greater angle than four-base cytosine bulges (Figure 3). Further, RecA bound to two- and three-base adenine bulged DNA with higher affinity than to the four-base adenine bulged DNA. The results suggest that the degree of kinking is not the only determinant of RecA binding. Comparison of the chemical sensitivity of two DNAs, one containing a four-base adenine bulge and the other a four-base cytosine bulge while bound by RecA protein revealed a similar chemical accessibility at the 5′ side of the bulges, but higher accessibility at the 3′ side of the bulge for the cytosine bulged DNA. We concluded that the degree of binding of RecA protein to bulged DNA was dependent on the conformation of the bulged bases and the kinking angles produced by the bulges as determined by the type and number of bases in the bulge (Figure 3).

Although a single-base mismatch does not kink DNA (see above), RecA protein showed preferential binding to DNAs containing certain single-base mismatches (Figure 4a and b). Moreover, DNAs containing single-base mismatches flanked by A·T bps showed equal or better efficiency of RecA binding than when the mismatches were flanked by G·C bps. Among the eight possible mismatches flanked by A·T bps, the A·C, A·A, G·A and C·C mismatches promoted RecA protein binding in presence of ATPγS. Interestly, an A·C mismatch flanked by A·T bps in a 28 bp DNA faciliated the binding of RecA protein to the same high level as when the 28 bp DNA contained a four-base cytosine bulge (Figure 4c). Chemical probing techniques were used to examine the structure of DNA within the RecA filament. It was found that upon binding of RecA protein, the DNA helix became accessible over at least 14 bp and the degree of chemical modification agreed with the binding efficiency of RecA protein (Figure 5).[48]

2.4 Summary

Gel electrophoretic studies have shown that DNAs containing 1–4-base bulges exhibit decreased mobilities in polyacrylamide gels, suggesting that bulged bases produce a kink in the DNA helix. The degree of kinking depends on the number of bases in the bulge, the base composition of the bulge, and the flanking bps. Visualization of a DNA containing a three-base bulge by EM has directly demonstrated kinks created by bulges and provided a measure of the distribution of the kinking angles. The binding of RecA protein to short DNAs containing lesions provides a sensitive means to detect perturbations in the helix. Binding of RecA protein to bulged DNA is influenced by the conformation of the bulged bases and by the kinking angles, which in turn are determined by the type and number of bases in the bulge. Although a single-base mismatch does not produce a kink in the DNA helix, we have found that certain single-base mismatches, for example an A·C mismatch flanked by A·T bps, facilitate the binding of RecA protein with the same high level as a four-base cytosine bulge. When these DNAs containing mismatches or base bulges were probed with chemicals, the degree of chemical accessibility agreed with the binding efficiency of RecA protein.

These studies provide critical structural information about DNAs containing bulges and mismatches. It will be important to correlate the structural information of these damaged DNAs with the relative rates of their correction by the methyl-directed mismatch repair system. Further, the binding of RecA to damaged DNA may facilitate recombinational repair by increasing the local concentration of RecA protein at the site of the damage. Therefore, the efficiency of RecA binding should also be taken into consideration in the overall repair/mutagenesis outcome of these damaged DNAs.

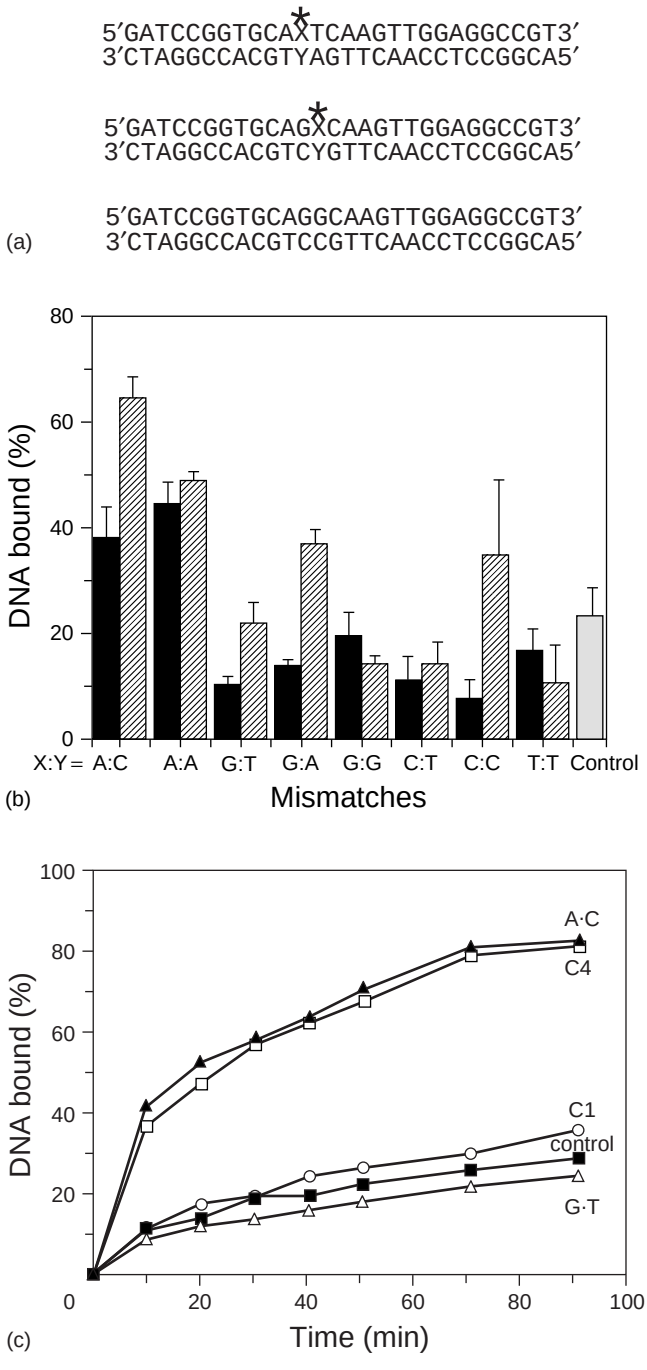

Figure 4 Comparison of RecA binding to bulged and mismatched DNAs. (a) Sequences of DNAs containing eight possible mismatches (positions indicated by X·Y) with two different flanking bps (A·T or G·C) and the control DNA (the lower sequence) are shown. (b) The 40 min time points of RecA binding to all 16 mismatched DNAs and the control DNA. The solid bars refer to mismatches flanked by G·C bps and the hatched bars to mismatches flanked by A·T bps. (c) DNAs containing: one-base cytosine bulges (C1), four-base cytosine bulges (C4), A·C mismatch flanked by A·T bps (A·C), G·T mismatch flanked by A·T bps (G·T), and the control DNA were compared using the filter binding assay as in Figure 3.

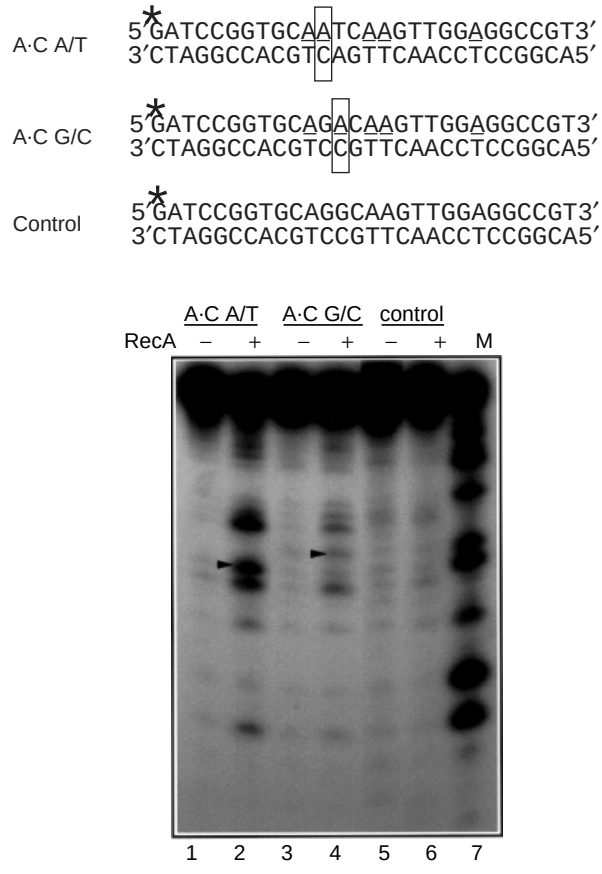

Figure 5 Diethyl pyrocarbonate (DEP) probing of DNAs containing A·C mismatches. DNAs containing an A·C mismatch flanked by A·T bps (lanes 1 and 2) or flanked by G·C bps (lanes 3 and 4) and the control DNA (lanes 5 and 6) were incubated without RecA protein (lanes 1, 3 and 5) or with RecA protein (lanes 2, 4, and 6) for 40 min at 37 °C. DEP was added at 37 °C for an additional 10 min and the modified DNAs were cleaved with piperidine and analyzed by electrophoresis in a 20% polyacrylamide–8.3 M urea gel. The positions of the ^{32}P label (∗), mismatched bp (boxed) and DEP-cleaved adenines (underlined) are indicated. In the autoradiogram, the mismatched adenines are indicated by arrows and a G-specific cleavage reaction of the control DNA serves as markers (M).

3 TRINUCLEOTIDE REPEATING DNA, NUCLEOSOME POSITIONING, AND HUMAN GENETIC DISEASES

DNA in the human cell is packaged progressively beginning with the most elementary particle termed the nucleosome. In a nucleosome 146 bps of DNA are wrapped 1.75 times around a protein core consisting of an octamer of the core histones (H2a, H2b, H3, H4). Each nucleosome is spaced in the human cell by ~50 bps of linker DNA to which histone H1 and other factors bind. This ''string of beads'', folded into larger structures by the action of H1, is eventually condensed into the metaphase

chromosomes observed by light microscopy. Different DNA sequences have been found to show differences in their affinity for wrapping around the core histone octamer and thus to be assembled into chromatin.

The ability of a variety of DNA sequences to form nucleosomes has been studied extensively. Repeated tracts of 4–6 adenines in-phase with the helix produce sequence-directed bends, and bent DNA preferentially assembles into nucleosomes,[49,50] presumably owing to the greater ease of wrapping of the DNA about the histone core. The sequence element $(A/T)_3NN(G/C)_3NN$ creates an anisotropically flexible wedge which also facilitates wrapping DNA about the histone octamer.[51] The 5S RNA gene from several species contains such flexible wedges and exhibits very strong sequence-directed nucleosome positioning.[52–55] Indeed, Shrader and Crothers[51] were able to create synthetic nucleosome positioning elements in which nucleosome formation was energetically favored over the 5S DNA by spacing up to 10 wedges in-phase with the helix.

Sequences also exist that inhibit nucleosome formation and in vivo such sequences, were they to occur in promoter or enhancer regions, could play an extremely important role in maintaining access to the DNA. DNA–RNA hybrids,[56] left-handed Z-DNA,[57] and DNA containing poly(dA : dT) runs have all been shown to exclude nucleosome formation in vitro.[58–60] Iyer and Struhl[61] found that poly(dA : dT) tracts in the yeast his 3 promoter stimulate Gcn4-activated transcription in vivo owing to nucleosome exclusion and increased accessibility of Gcn4-binding sites.

Described here are our findings that show that the expanded triplet sequences characteristic of DM and FraX, the two most common human triplet diseases, bracket the known range of affinities of DNAs for nucleosome assembly. While long tracts of repeating CTG triplets (expanded in DM) form the most stable known naturally occurring nucleosomes, long tracts of repeating CCG triplets (expanded in FraX) resist nucleosome assembly. In addition, methylation of the CCG repeats further represses the already weak ability of these (CCG) DNAs to assemble into chromatin. The physical basis for exclusion of nucleosomes by DNA containing long tracts of CCG repeats has been probed in experiments with model DNAs. Results from these studies revealed a new class of repeats in eukaryotic promoters, $[(C/G)_3NN]_n$, which may serve to loosen chromatin structure in the promoters of TATA-less genes.

3.1 CTG Repeats in Myotonic Dystrophy

DM is an autosomal dominant disease and the defect has been mapped to chromosome 19.[11,12] Two genes at the DM locus [the myotonic dystrophy protein kinase (DMPK) and the myotonic dystrophy locus-associated homeodomain protein (DMAHP) genes] have been shown to be affected by the expanded CTG repeat which is located in the 3′ untranslated region of the DMPK gene[62–65] and is about 1 kb from the major transcription start site of the DMAHP gene[66–68] (Figure 6). A block of 5–37 CTGs is frequently found in normal individuals, while DM patients with mild symptoms have triplet blocks of 50–80 repeats, and expansions from 100 to several thousand repeats can be found in individuals with full-blown symptoms. DM patients are heterozygous with only one of the two alleles containing an expanded triplet block, since homozygotes are rare and lethal. To date, the pathogenic mechanism of DM is not well understood.

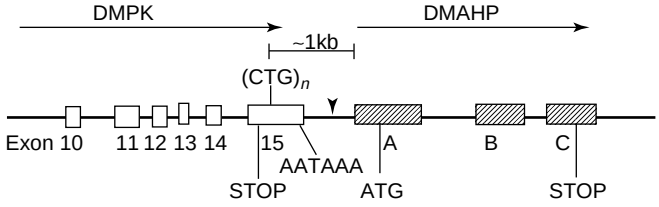

Figure 6 Gene organization of the DM locus. The location of the translation start site (ATG), the translation stop sites (STOP), the putative poly(A) site (AATAAA), and the CTG repeats $[(CTG)_n]$ are indicated. The region demonstrating hypersensitivity to DNase I[69] is indicated by an arrow.

3.1.1 Nucleosome Positioning by Long CTG Repeats

Knowledge that the threshold for the triplet diseases frequently approximates the amount of DNA in a nucleosome, and the general understanding that any short repeating sequence may amplify sequence signals for nucleosome positioning or exclusion, led us[70] to examine the ability of long repeating tracts of $(CTG)_n$ to assemble into chromatin. This work was carried out using a simple but well established method of chromatin assembly involving mixing DNA with purified histone octamers (H2a, H2b, H3, H4) in 2 M NaCl followed by a progressive lowering of the salt concentration to 0.6 M or less by stepwise dilution or dialysis.[50,51,70–73] In the approach used in Wang et al.,[70] a minimal amount of histone octamers was used such that each DNA would on the average, contain only a single assembled nucleosome. This simplified measurement of the nucleosome position since the foreshortening due to multiple nucleosomes would have greatly complicated analysis. In this study, direct EM was utilized to measure the position of several hundred individual nucleosomes along single DNAs and the data were summed to create a nucleosome positioning map. To distinguish one end of the linear DNA from the other, one DNA end was labeled with a biotin–streptavadin tag.

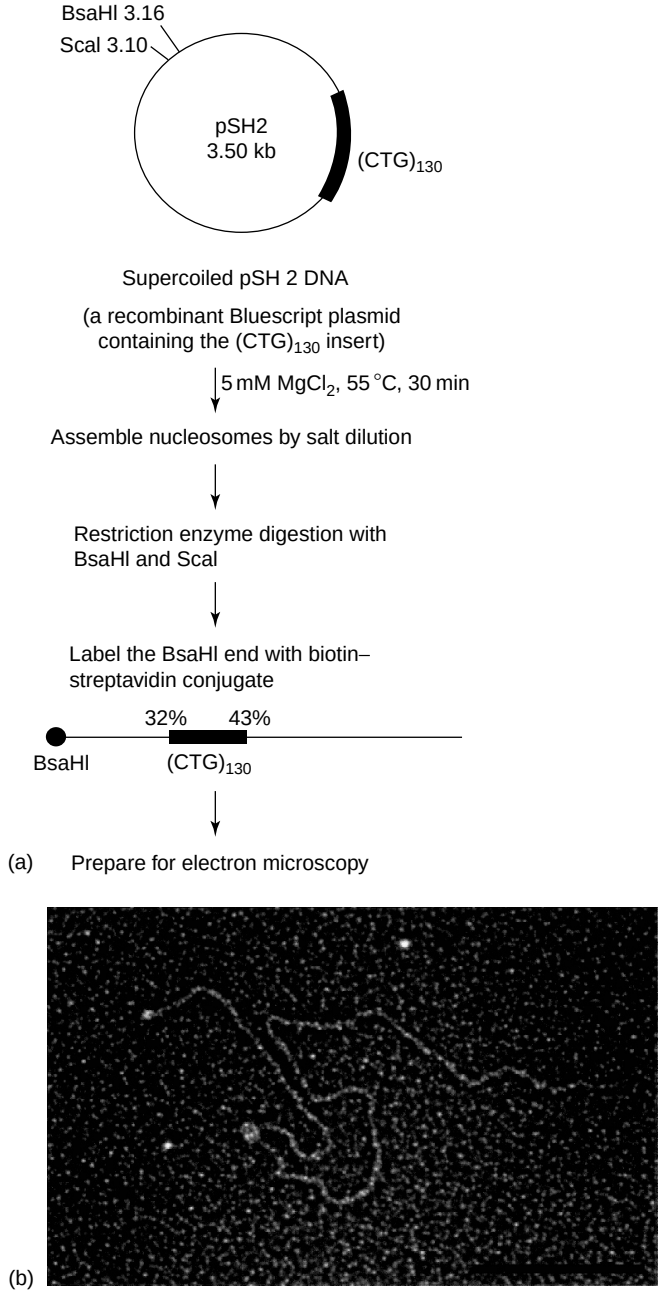

Figure 7 (a) Scheme for nucleosome assembly on supertwisted plasmid DNA and end labeling with biotin–streptavadin complexes. (b) Visualization of nucleosomes assembled on streptavadin-labeled pSH2 DNA containing a $(CTG)_{130}$ repeat derived from DM patients. The DNA (curved filament) contains streptavadin protein bound to one end and one nucleosome assembled near the center of the DNA. In the absence of added histone, no nucleosome-like objects are observed. Reconstitution of the DNA with histones is described in this article. Preparation for EM including rotary shadowcast with tungsten was described by Wang et al.[70] Bar, 130 nm.

For such analysis it was important to also consider the effects of magnesium and supercoiling of the template DNA. Numerous DNA structural transitions are facilitated by negative supercoiling or influenced by heating in the presence of magnesium. A supertwisted plasmid, pSH2 isolated from *E. coli*, containing 130 CTG repeats derived from a DM patient was treated with magnesium and heat and then reconstituted into chromatin. The DNA was then linearized to place the CTG triplets in pSH2 between 32 and 43 map units along the DNA (Figure 7a and b). The resulting nucleosome map (Figure 8a) revealed that 48% of all the nucleosomes were present in the region between 30 and 45 map units. In contrast, the map for the Bluescript vector lacking the CTG insert (Figure 8b) showed a uniform distribution of nucleosomes over the length of the DNA with only 10% of the nucleosomes present in the region between 30 and 45 map units. The use of linear or supercoiled DNA that had not been treated with heat or magnesium still resulted in 29% of all nucleosomes localized to the triplet repeat (Figure 8c and d). Further analysis using a set of six pUC19-based plasmids (pretreated with heat and magnesium) containing 26–250 contiguous CTG triplet repeats showed that as the size of the triplet block increased in size the efficiency of nucleosome formation at the repeats also increased up to a repeat size of 180. These data are summarized in a normalized form in Figure 9.

These data provided the first direct evidence that repeating CTG triplets can serve as a strong nucleosome-positioning element. Indeed, this increase in strength of nucleosome formation occurred over the range where the triplet blocks change from normal, to the intermediate, to disease-causing. The EM observations therefore suggest that expanded blocks may create unusually stable nucleosomes that could alter local chromatin structure. Microscopy, however, was unable to provide precise measurements of the strength of the CTG repeat blocks as nucleosome positioning elements, data that would be essential to models based on these observations.

A method for measuring the strength of nucleosome positioning elements has been developed by Shrader and Crothers,[51] termed competitive nucleosome reconstitution (Figure 10a). In this assay, a small amount of the specific DNA to be tested (0.1 μg) is labeled with ^{32}P and then mixed with a 200-fold excess of unlabeled calf thymus DNA (20 μg) and ~5 μg of calf thymus histone octamers[50] in a solution containing 2 M NaCl. The salt is slowly lowered to 0.1 M NaCl and the mixture is electrophoresed on a 5% polyacrylamide gel to separate free DNA from the nucleosome-assembled DNA. The amount of DNA in each band is quantified. Usually three or four separate but identical experiments are carried out for each set of DNAs. In such studies, a sequence found in the *Xenopus borealis* somatic 5S RNA gene is frequently used as a standard against which other sequences are measured. This 5S element was previously the strongest

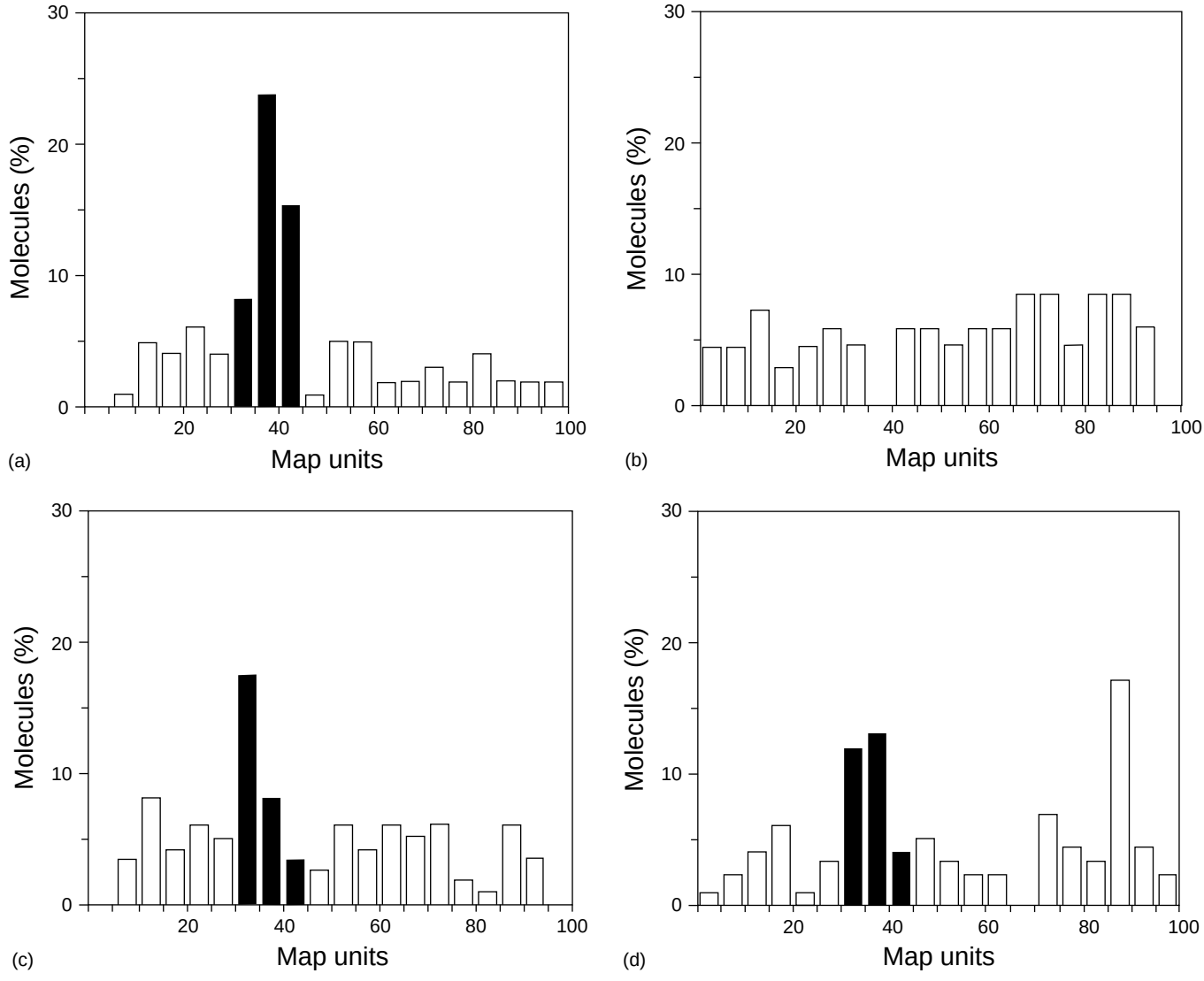

Figure 8 Distribution of nucleosomes reconstituted on to: (a) closed circular pSH2 DNA [which contains the $(CTG)_{130}$ repeat], (b) closed circular Bluescript vector, (c) linear pSH2 without magnesium and heat treatment, and (d) closed circular pSH2 without magnesium and heat treatment. Reconstitution and preparation for EM were performed as described.[70] At least 100 DNA molecules containing single nucleosomes were photographed and the position of each nucleosome from the streptavidin-labeled end was measured, and a histogram (length broken into 20 segments with each percentage of length being one map unit) showing the location of the nucleosomes along the DNA was generated. The filled bars indicate the position of the CTG repeat sequences along the DNA.

known natural nucleosome positioning element. Using this method, we[74] compared DNAs containing 75 and 130 CTG repeats to the *X. borealis* somatic 5S RNA gene. In such studies it is important that the DNAs being compared are of the same size. Thus two pairs of DNAs, ~285 and ~450 bp in size, were compared.

The results of such an analysis for the sets of DNAs of ~285 and ~450 bp are shown in Figure 10(b) and (c). The nucleosome-assembled DNA appears as a retarded band. One major retarded band was observed in both 5S RNA gene samples while two retarded bands were found in the ~285 bp $(CTG)_{75}$-containing DNA and possibly three closely spaced retarded bands were present in the ~450 bp $(CTG)_{130}$-containing DNA. The multiple retarded bands represent DNAs associated with two or three nucleosomes. The results revealed that the ratio of the fraction of the nucleosome-assembled DNA to free DNA is 5.6 ± 0.4-fold higher in the DNA containing the $(CTG)_{75}$ block as compared with the 5S RNA gene. The ~450 bp DNA containing the $(CTG)_{130}$ block is 8.7 ± 0.7 fold stronger than the 5S RNA gene (472 bp) in nucleosome assembly.

These assays allow one to calculate the free energy difference for nucleosome formation. The difference in

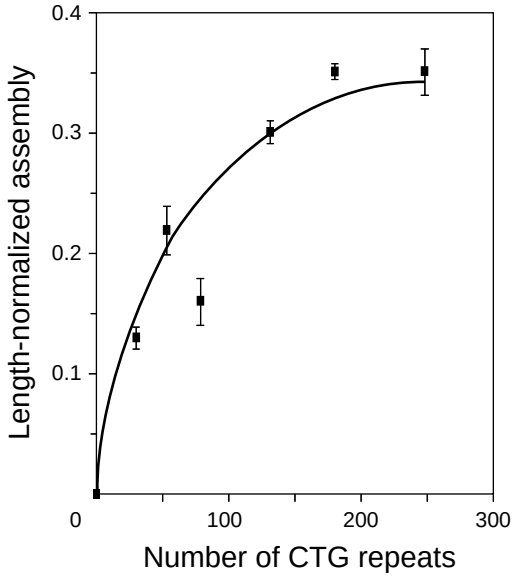

Figure 9 Strength of sequence-directed nucleosome positioning correlates with the size of the repeated CTG block. Data from Figure 3 of Wang et al.[70] were normalized as described.

free energy between the $(CTG)_{75}$ DNA and the 5S RNA gene was found to be 1015 ± 43 cal mol^{-1} and for the $(CTG)_{130}$ DNA the difference was 1276 ± 45 cal mol^{-1}. In Figure 11, these results are combined with data obtained by Shrader and Crothers[51] and further work described below. To match the different studies, the free energies for the vector DNA for each set of experiments was taken as zero.

These studies demonstrated that when CTG repeat blocks grow to sizes of $n \geq 75$ they become the strongest known natural nucleosome positioning elements, and (at $n = 130$) are nearly nine times stronger than the 5S element. The difference in binding free energy between the strongest and average DNA elements is seen to span a range of $\geq$100-fold. However, the total energy difference measured in high salt is relatively low, 3 kcal mol^{-1}.[70] Hence whether or not such differences measured in high salt would be physiologically relevant remained in question. In the cell, at physiological ionic strength and in the presence of histone H1 and nucleosome assembly factors, these differences could be great enough to dictate nucleosome positioning. The strongest evidence that these free energy differences do represent physiologically important differences was to come from model studies of the CCG triplet repeat from Fragile X expansions as described below.

Godde and Wolffe[78] also examined nucleosome assembly on $(CTG)_n$ repeats by competitive nucleosome reconstitution and mapped translational and rotational positions by nuclease and hydroxyl radical cleavage. They confirmed the observation of Wang and Griffith[74] that CTG repeats create sites for preferential nucleosome assembly. However, they did not observe the increase in efficiency of nucleosome assembly with increasing size of the repeat from 10 to 55 and 62 copies, which may be too short to display the effect. Probing of rotational position demonstrated a clear 10–11 bp repeating pattern over the DNA containing six copies of CTG repeat and the pattern remains similar in $(CTG)_{55}$ DNA. Moreover, the $(CTG)_6$-containing DNA favors a single translational position with the CTG sequence bordering the nucleosomal dyad whereas the $(CTG)_{55}$ DNA showed several nucleosome positions.

3.1.2 Implications of Arrays of Hyperstable Chromatin at the Myotonic Dystrophy Locus

The in vitro nucleosome reconstitution studies described above clearly show that repeating CTG triplets can generate very unusual chromatin structure. When the repeat tracts grow into the size range in which DM disease symptoms appear, these repeat tracts can generate chromatin segments with stabilities that are possibly nine times greater than the previously strongest known natural nucleosome positioning element, the 5S DNA. Although the precise physical reason why such repeating triplets generate such highly stable chromatin remains unclear, a number of possible biological implications of the appearance of such abnormally stable chromatin segments are worth mentioning.

Otten and Tapscott[69] have carried out in vivo mapping of nuclease hypersensitive sites with DNA samples of fibroblasts and myoblasts from three unrelated DM patients in which the DM alleles are heterozygote and the expansion is ~6 kb. Compared with the observations from an unaffected individual and the wild-type allele, the expanded allele showed loss of the DNAse I hypersensitive site located just 3′ of the CTG repeat (indicated by an arrowhead in Figure 6) and resistance to restriction endonuclease cleavage adjacent to the DNase I hypersensitive site. Recently, this hypersensitive site has been mapped to an enhancer region which regulates the expression of the DMAHP gene.[79] In cells of DM patients with a loss of the hypersensitive site, the amount of DMAHP transcript was reduced compared with the controls and the amount of transcript from the expanded allele was also greatly lowered compared with that from the wild-type allele. These results elegantly demonstrated that the expanded CTG repeat produces an altered chromatin structure, possibly resulting from the strong nucleosome positioning by the CTG repeat. The altered chromatin structure might suppress the transcription of the DMAHP gene which leads to the pathogenesis of DM.

Furthermore, the appearance of a segment of hyperstable chromatin could in itself promote further expansion

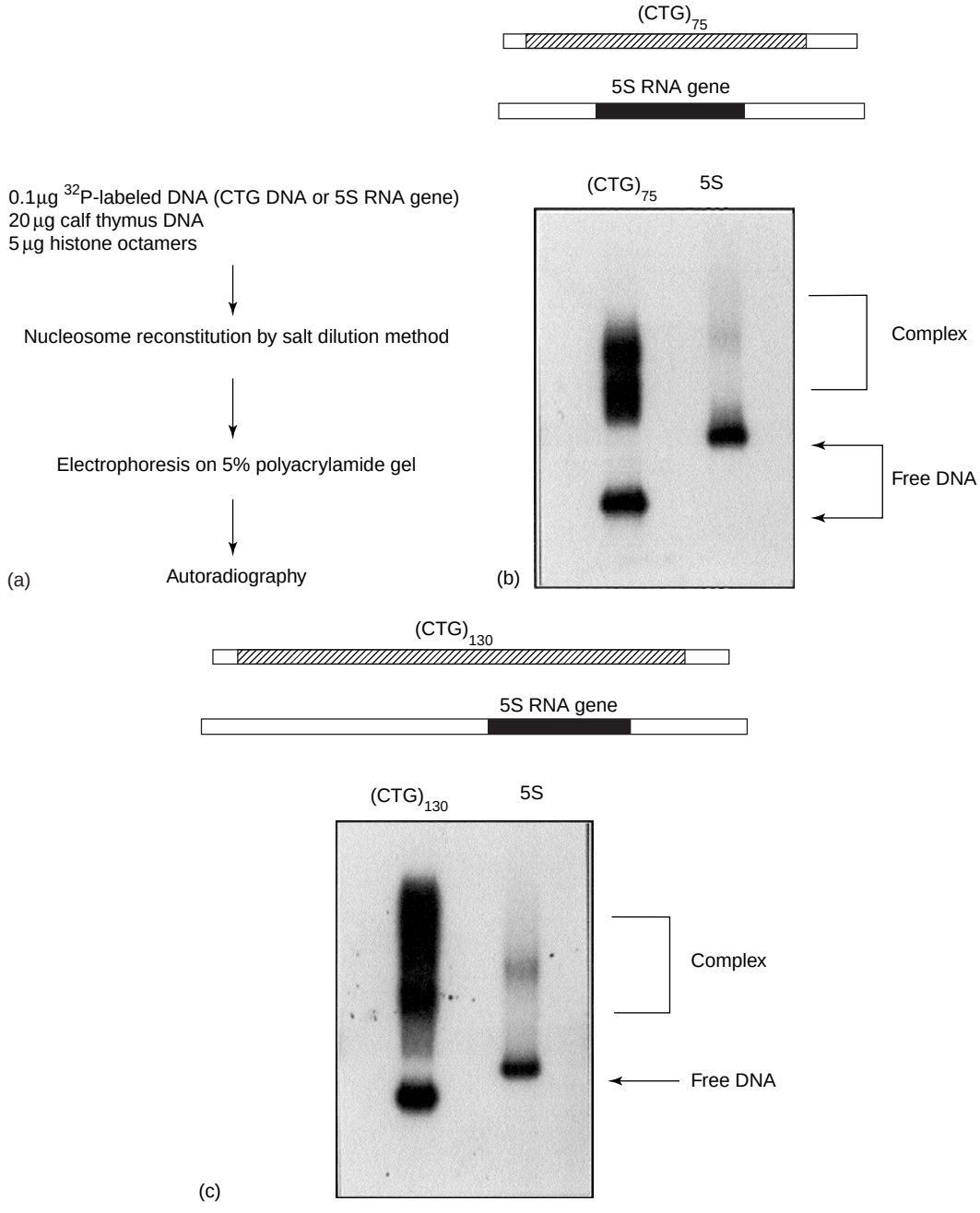

Figure 10 Competitive nucleosome reconstitutions with CTG triplet containing DNA. (a) Scheme for competitive nucleosome reconstitution. (b) The ~285 bp $(CTG)_{75}$-containing fragment or the 303 bp 5S RNA gene and (c) the ~450 bp $(CTG)_{130}$ DNA or the 472 bp 5S RNA gene labeled with ^{32}P (100 ng) were mixed with 20 μg of unlabeled calf thymus DNA and ~5 μg of calf thymus histone octamers in a solution containing 2 M NaCl, 100 μg mL^{-1} of bovine serum albumin and 0.1% Nonidet P-40. The total DNA added is about a 4-fold molar excess over the histone octamers. The salt was slowly lowered and the mixture then electrophoresed on a 5% polyacrylamide gel. Samples were visualized by autoradiography.

of the triplet tract. During the replication of eukaryotic DNA, the replication machinery must remove the DNA template from the histone octamer prior to, or possibly in concert with, the helicase action that separates the two DNA strands. Were the replication machinery to be severely hindered in this action by the presentation of an array of hyperstable nucleosomes, then the likelihood of polymerase slippage as the replication fork transited the triplet tract would be expected to be much greater.

As arrays of hyperstable nucleosomes might impede the replication machinery, they might also slow or stop

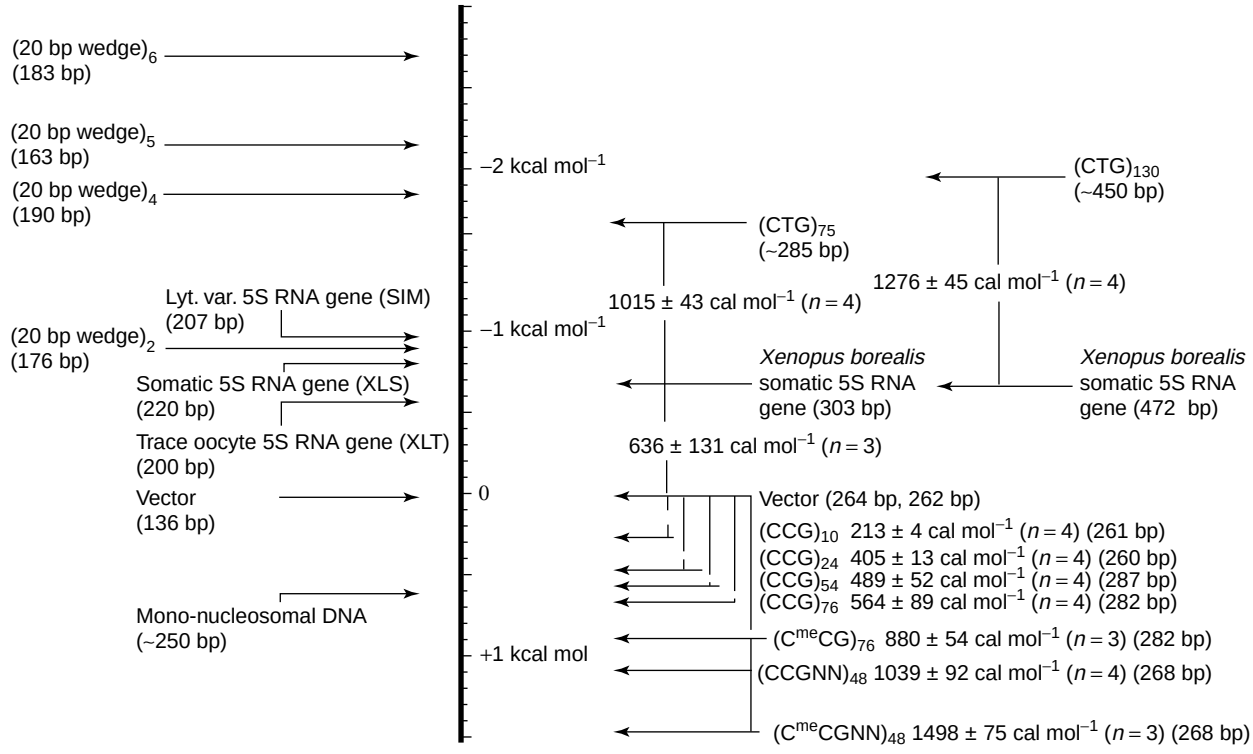

Figure 11 Comparison of free energies derived from several nucleosome positioning sequences. Data which are at the left of the free energy scale bar were adapted from Shrader and Crothers.[51] The sequence of the 20 bp anisotropically flexible wedge is TCGGGTTTAGAGCCTGTAAC. Results from studies of Wang and Griffith[74–77] are listed at the right side of the bar with the free energies of the vector DNA (pUC19 fragments) defined as zero cal mol^{-1}.

transcription complexes. When RNA polymerase stalls on DNA the nascent RNA chain is frequently cleaved, inactivating the RNA. Examination of transcription of the DMPK gene has provided conflicting results regarding the amount of DMPK transcript expressed by the expanded DM allele.[80–86] Whether or not expanded CTG tracts in DM protein kinase gene result in lowered mRNA levels is currently in question. It is possible that such an effect may only be evident when the tracts become longer than those investigated in some of the studies reported to date.

Finally, a global effect of arrays of hyperstable nucleosomes may be to generate domains of inactive chromatin. Such elements could inactivate genes either upstream or downstream of the domain and possibly over distances of many kilobases. Domains generated by an array of hyperstable nucleosomes would be similar to chromatin barriers that have been shown to silence genes moved into their proximity. Given the findings that the CTG tract in the DMPK gene is also in the upstream region of the DMAHP gene,[79,87] this presents one of the most attractive models of how this expanded CTG tract could exert its biological effect. It could be imagined that as the CTG tract grows, the size and thus strength of the array of hyperstable nucleosomes increase greatly and with each additional nucleosome added the effect of this very unusual hyperstable chromatin boundary could grow from just interference with transcription of the DMAHP gene to interference also with the DMPK gene and finally to interfering with all three genes including the upstream 59 protein gene. The clinical observations that the presentation of DM differs depending on the size of the CTG tract may reflect a progressive involvement of additional genes as the tract increases in size.

3.2 Expanded CCG Repeats in Fragile X Syndrome and Folate-sensitive Fragile Sites

Fragile sites are chromosomal loci that stain poorly, contain gaps and are frequent sites of DNA strand breakage. Thus they represent regions of unstable chromatin structure. More than 100 separate fragile sites have been identified in the human genome, classified as common or rare, and, further, they are divided according to the agents used to identify them.[88] Fragile sites are highly conserved and appear to have played key roles in the stepwise evolution of primate chromosomes. Recent work suggests that they are major sites of drug-induced chromosomal amplification.[89] A number of fragile sites have been cloned, the majority being members of the class termed rare, folate-sensitive sites which are seen

by light microscopy when cells are exposed to a folate starvation regime. Sequence analysis of five rare, folate-sensitive sites, FRAXA, FRAXE, FRAXF, FRA16A, and FRA11B, revealed long blocks of repeating CCG triplets together with the frequent methylation of nearby CpG islands.[90–96] Another fragile site, FRA16B, was sequenced and found to involve a highly A/T-rich element demonstrating that not all fragile sites involve repeating CCG triplets. Expansion of the triplet block in the 5′ untranslated region of the FMR-1 gene[97,98] at FRAXA from 20 to 50 repeats in most individuals to >200 CCG repeats has been linked to the FraX (see below). The expanded triplet block is the site of preferential breakage at FRAXA.[90,91] The FRAXE site on the X chromosome[93] is correlated with a rare form of mental retardation, and the FRA11B site is associated with Jacobsen's syndrome. Here, a portion of the long arm of chromosome 11 is lost, implying a link between this fragile site and chromosome breakage.[96] FRAXF and FRA16A are apparently without disease phenotype.[94,95]

FraX was named for the presence of a rare, folate-sensitive fragile site on the X chromosome (FRAXA) whose appearance correlates with the disease,[97,98] In contrast to DM, the FMR-1 gene is likely the only gene

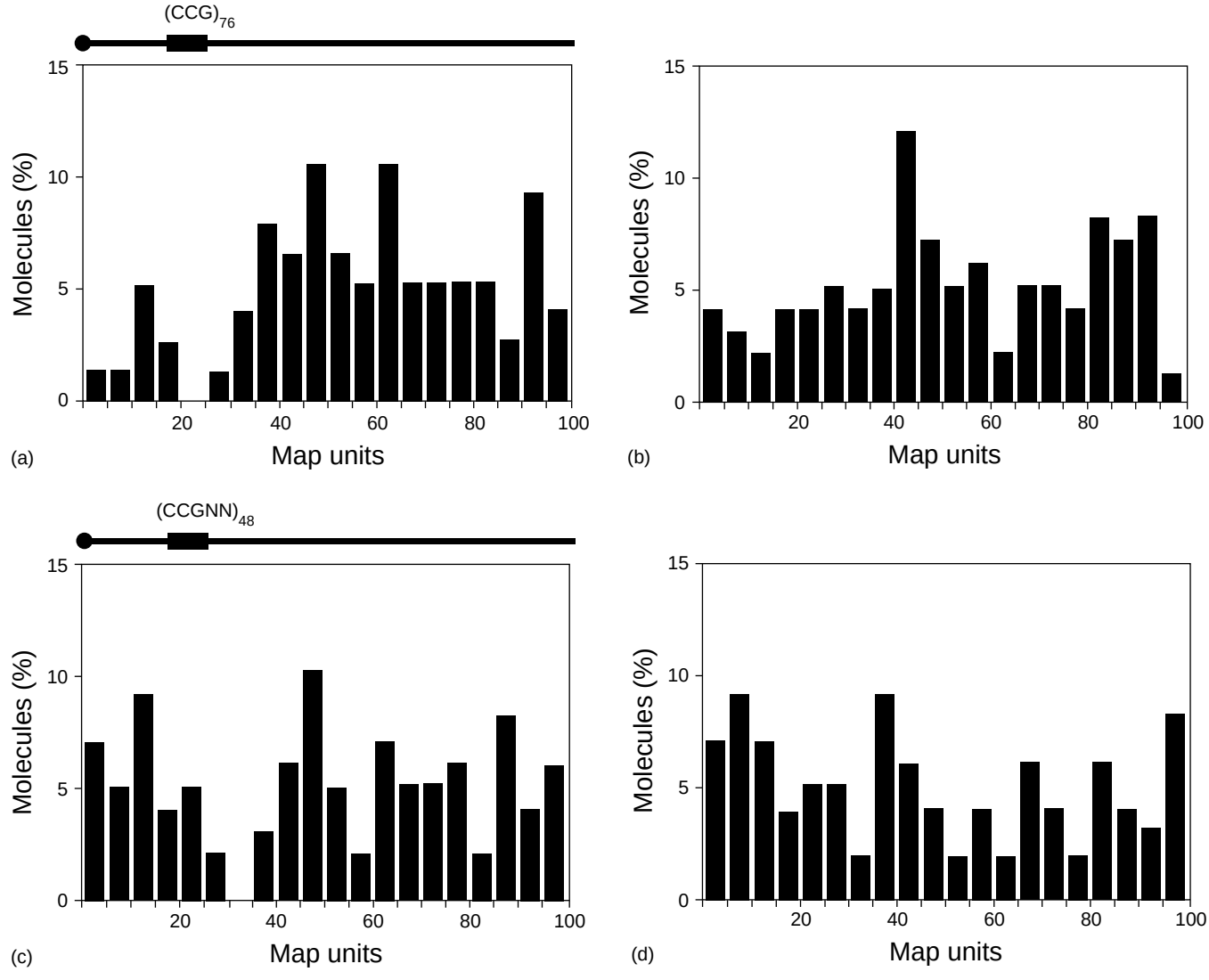

Figure 12 Distribution of nucleosomes assembled on closed, circular (a) pRW3376 containing the $(CCG)_{76}$ repeat block and (b) pT7 blue vector as a control; (c) $p(CCGNN)_{48}$ plasmid and (d) pGEM3zf(+) vector as a control. Nucleosome assembly and EM analysis were performed as described.[75,76] At least 100 DNA molecules containing single nucleosomes were photographed, the position of each nucleosome from the streptavidin-labeled end (filled circle) was measured, and a histogram was generated (with DNA length broken into 20 slices and each percentage of length equivalent to 1 map unit) showing the location of the nucleosomes along the DNA. The positions of the $(CCG)_{76}$ sequence in pRW3376 and the $(CCGNN)_{48}$ sequence in $p(CCGNN)_{48}$ are indicated.

involved in FraX. Studies of the FMR-1 gene have also shown reduced transcription from this locus in FraX patients[99] and rare FraX cases are seen in which there has been a deletion or point mutation in the FMR-1 protein, pointing to loss of protein function as a cause of the disease.[100] Moreover, the severity of the disease has been shown to be directly related to the size of the triplet repeat and also the purity of the repeat tract. Thus, moderate-sized pure CCG repeat tracts are more deleterious than longer repeat tracts containing interruptions. Finally, there is also a strong correlation between the severity of the disease symptoms and the level of methylation of the region. Hence any model for the molecular basis of FraX will have to account for these observations.

3.2.1 Nucleosome Exclusion by Repeating CCG Tracts

The discovery of hyperstable nucleosome formation by long tracts of CTG triplet repeats spurred a parallel analysis of DNAs containing long CCG triplet repeat tracts. This work was further impelled by the knowledge that the fragile sites containing CCG triplet blocks exhibit properties of unstable chromatin suggesting that significant effects on chromatin structure due to the presence of these repeats might be found. Indeed it had been speculated earlier[101] that fragile sites may result from the inability of DNA to fold compactly during metaphase.

In the work of Wang et al.,[75] a series of plasmids containing different sized tracts of repeating CCG triplets were employed. The plasmid pRW3376 contains 76 tandem CCG repeats[102] and this DNA was reconstituted with purified histone octamers, linearized, and one DNA end labeled with streptavadin to place the $(CCG)_{76}$ repeat block between 18% and 26% from the marked end in a manner similar to that illustrated in Figure 12(a). A nucleosome map was then prepared. Inspection of the data showed that only 2.6% of the DNAs had a nucleosome located between 15 and 20 map units from the marked end and no molecules had nucleosomes between 20 and 25 map units. This was in contrast to other segments of 5 map units length which showed from 5% to 10% of the DNA with nucleosomes assembled. The parent vector showed no regions of nucleosome exclusion (Figure 12b). These direct EM observations provided the first evidence that repeating CCG triplets could lead to nucleosome exclusion in contrast to nucleosome positioning observed with repeating CTG triplets. Control experiments showed that legitimate nucleosomes were being formed on normal sequence DNA in the same reactions.

Competitive nucleosome reconstitution was then used in this study to measure the energetics of nucleosome formation for DNA fragments containing CCG repeats of increasing size (Figure 13a and b). The ratio of nucleosome assembled DNA to free DNA for a pUC19 fragment and similar length fragments containing 10, 24, 54, and 76 CCG repeats was determined. It was found that as the length of the repeat block increased, the efficiency of nucleosome assembly decreased. Whereas a 261 bp DNA containing 10 tandem CCG repeats showed a 1.4-fold lower efficiency (30% decrease) in nucleosome assembly relative to the pUC19 fragment, the fragment containing 76 tandem CCG repeats was 2.6 ± 0.4-fold less effective (62% decrease), corresponding to a $564 \pm 89\,\text{cal mol}^{-1}$ difference in free energy (Figure 11). Compared with the free energy value for a similar sized DNA containing 75 CTG repeats (see above), this amounts to a 40-fold difference between the two repeating triplet DNAs. These observations clearly paralleled the in vivo observation of fragile sites being regions of unstable chromatin.

While the physical basis for repeating CTG triplets forming hyperstable nucleosomes remains unclear, experiments with model variants of the CCG motif provided clues as to why this expanded triplet may exclude nucleosomes and, further, revealed a possible function for these motifs in eukaryotic genomes.

3.2.2 Model Studies Involving the Repeat $[(C/G)_3NN]_n$

Considering of the properties of the sequence motif in the 5S RNA gene, $(G/C)_3NN(A/T)_3NN$, which forms very strong nucleosomes,[51,103] we postulated[76] that DNA containing long repeats in the form of $(G/C)_3NN(G/C)_3NN$ might exclude nucleosomes. This latter motif differs from the 5S RNA gene element in that each time minor groove compression is required, the DNA presents the histone octamer with a wedge that favors bending into the major groove. Shrader and Crothers[51] had noted that a triplet of As or Ts preferentially bends into the minor grove of DNA whereas a triplet of Gs or Cs bends preferentially into the major grove. Thus DNA containing a nucleosome-sized tract of repeating CCGNN pentanucleotides which is a member of the $(G/C)_3NN(G/C)_3NN$ motif family should resist nucleosome formation. Pure repeating CCG triplet tracts are also a member of this general family and if nucleosome exclusion were observed for DNA containing repeating CCGNN pentanucleotides this would provide an explanation for nucleosome exclusion by repeating CCG triplets.

To test the hypothesis that long tracts of $(G/C)_3NN$ repeats will exclude nucleosomes, three pGEM3zf(+)-based recombinant plasmids, $p(CCGNN)_{12}$, $p(CCGNN)_{24}$, and $p(CCGNN)_{48}$ containing, respectively, 12, 24, and 48 tandem copies of the CCGNN repeat with the Ns varied and rich in As and Ts, were generated. EM was then utilized together with in vitro nucleosome reconstitution onto supertwisted $p(CCGNN)_{48}$ plasmid

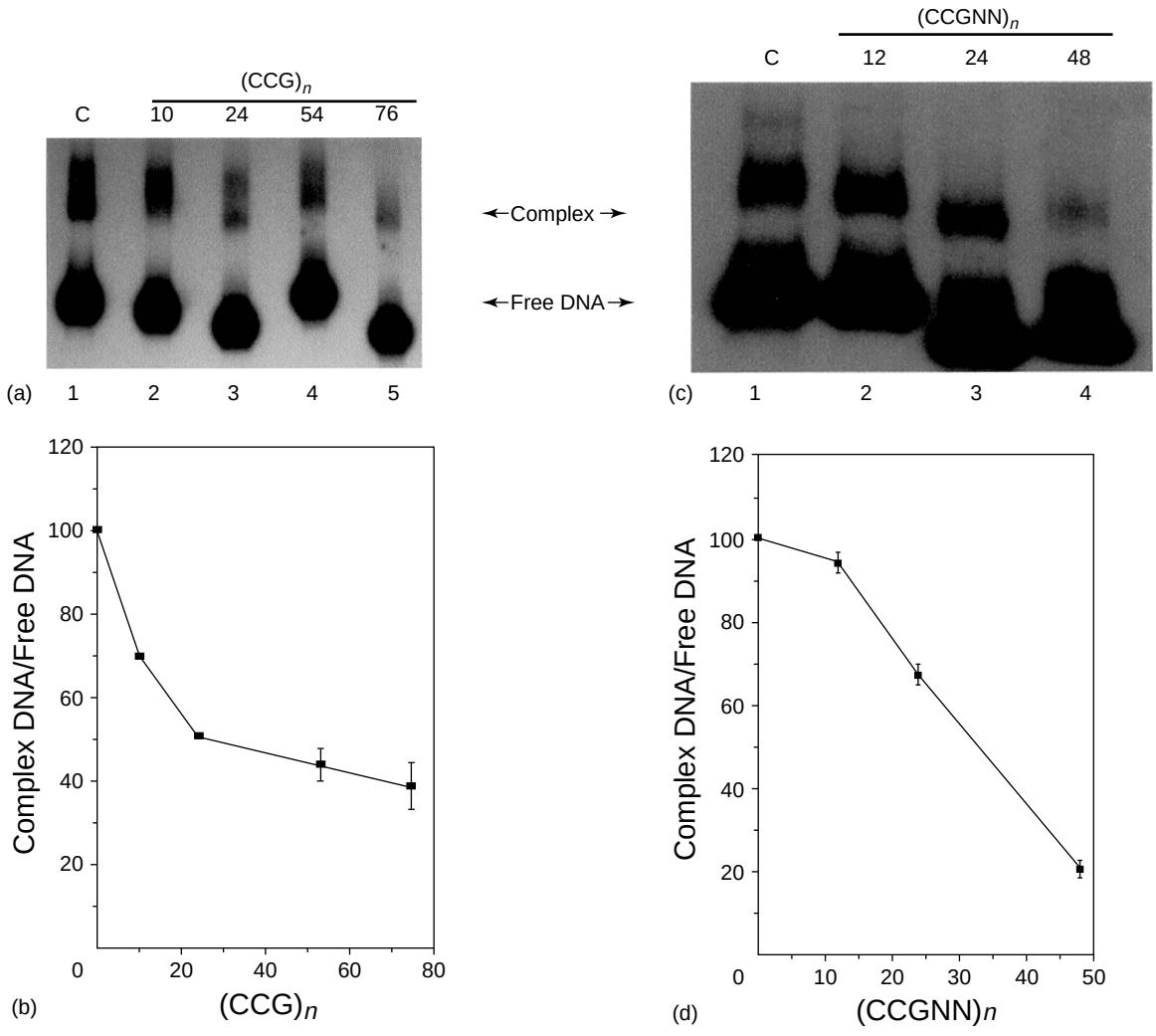

Figure 13 Competitive nucleosome reconstitution with (a, b) CCG and (c, d) CCGNN containing DNAs. (a) Autoradiogram of a competitive nucleosome reconstitution experiment. Lane 1, the 262 bp pUC19 fragment; lanes 2–5, DNA fragments containing 10, 24, 54, and 76 CCG repeats respectively. Note: the $(CCG)_{54}$ and $(CCG)_{76}$ repeat blocks contain AGG and CAG interruptions.[102] (b) Dependence of nucleosome assembly on the length of the repeat block. (c) Autoradiogram of a competitive nucleosome reconstitution experiment. Lane 1, the 262 bp pUC19 fragment; lanes 2–4, DNA fragments containing 12, 24, and 48 CCGNN repeats, respectively. (d) Dependence of nucleosome assembly on the length of the repeat block. DNA preparation, reconstitution of the fragments with histones, and gel electrophoresis were as described elsewhere.[75,76] Each DNA was reconstituted in three separate but identical experiments, and the fraction of DNA in the nucleosome-assembled and nucleosome-free DNA bands was measured by a PhosphorImager.

DNA to examine the propensity of the inserts to assemble into nucleosomes as described above. Subsequent to assembly the DNA was linearized placing the CCGNN block in $p(CCGNN)_{48}$ between 27% and 34% from the tagged end. Analysis revealed strong nucleosome exclusion over the $(CCGNN)_{48}$ insert with only 2% of all nucleosomes mapped located between 25 and 35 map units whereas in other segments of this length ~10% or more of the DNA contained a nucleosome (Figure 12c). In contrast, the parent pGEM3zf(+) vector showed no regions of strong nucleosome exclusion (Figure 12d).

Competitive nucleosome reconstitution assays were then employed as described[74] to quantify the degree of nucleosome exclusion over the insert in these plasmids. DNAs ~260 bp in length containing $(CCGNN)_n$ blocks of 12, 24, or 48 repeats were generated. As the length of the CCGNN repeat sequence increased from 12 to 48 repeats within a ~260 bp segment, the ability of the DNA to exclude nucleosomes increased in proportion to the length of the repeat block (Figure 13c and d). Analysis showed that the 268 bp DNA containing the $(CCGNN)_{48}$ repeat is 4.9 ± 0.6-fold less efficient in nucleosome assembly than the same sized pUC19 fragment. Indeed,

a 261 bp DNA containing 12 tandem CCGNN repeats assembled nucleosomes 94% as efficiently as a 262 bp pUC19 fragment. When these results were combined with results from earlier studies, it was estimated that the $(CCGNN)_{48}$ repeat block is ~78-fold less efficient than a similar length block of repeating CTG triplets. The difference in free energy between the $(CCGNN)_{48}$ fragment and the pUC19 DNA is 937 ± 79 and $\sim 2600\,cal\,mol^{-1}$ relative to the CTG_{75} repeating element (Figure 11).

Based on the findings described above, a computer search was carried out using the GenBank database to examine the prevalence of the general $[(G/C)_3NN]$ motif.[76] Searches against 240 bp of a $[(G/C)_3NN]_{48}$ continuously repeating sequence revealed many matches, and 75 examples showed ≥85% sequence match over 200 bp of this motif (Table 2). Many of these were present in or near the control regions for eukaryotic genes. Of the 75 with the greatest number of sequence matches, 31 genes were noted to contain the $(G/C)_3NN$ motif in the 5′ region upstream of the coding sequences. Such regions would be loci where nucleosome exclusion would be expected to provide favored access to sequence-specific proteins engaged in the regulation of gene expression. It was also observed that at least 20 of these 31 genes lack a TATA box in the promoter region and are "TATA-less" genes. In two of these genes, the 5′ regions have been mapped for nuclease hypersensitive sites. The −530 to −300 nucleotide (nt) upstream control region of the human dihydrofolate reductase gene contains the $[(G/C)_3NN]_{48}$ motif (with 87% sequence match) and this overlaps with a hypersensitive region mapped in vivo.[104] In the gene for the human ETS-2 nuclear phosphoprotein, there is a sequence upstream of the gene between nts

Table 2 Examples of genes containing regions showing >85% sequence matches to the $[(G/C)_3NN)]_{48}$ sequence taken from a GenBank search[a]

Location	Gene or encoded protein	Sequence matches[b]
5′ end (upstream of coding sequences)	Human glutamate dehydrogenase	222/235
	Human arginosuccinate lyase	213/230
	Human insulin receptor	214/237
	Wheat α-amylase	208/233
	Chicken hsp 90	205/231
	Human dihydrofolate reductase	205/228
	Pig nuclear factor 1	201/225
	Rat insulin-like growth factor binding protein	202/232
	Rabbit metallothionine	195/229
	Mouse neu proto-oncogene	222/236
	Human hematopoietic cell-specific protein	215/235
	Human fibronectin	205/233
	Rat phosphorylase kinase catalytic subunit	199/230
	Human vitronectin protein	196/229
	Mouse S16 ribosomal protein	202/232
	Rat neu oncogene	214/231
	Human gastrin releasing peptide	202/228
	Human ETS 2 oncogene	204/231
	Rat nucleolin	199/230
Coding region	Chicken c-fos proto-oncogene	208/231
	Chicken tropoelastin	206/224
	Human collagen-like protein	201/217
	Herpes simplex 1 UL 18	204/230
	EBV nuclear antigen 3C	204/234
	Human thymidine kinase	199/233
	Chicken ubiquitin	202/231
	Human translocation fusion protein (E2A-Prl)	202/232
	Human p120	199/233
3′ end (downstream of coding sequences)	Human cytochrome p450	207/232
	Mouse histone H2A·X	209/235
	Human lamin B2	206/232
	Chicken p53 oncoprotein	193/215
	Human α_1 collagen type 1	204/232

[a] At least 75 genes in the GenBank were found to have ≥85% sequence matches to $[(G/C)_3NN]_{48}$.
[b] The first number indicates the number of bases showing sequence matches with $[(G/C)_3NN]_{48}$ and the second number refers to the length of this sequence.

−195 and +45 containing the $[(G/C)_3NN]_{48}$ motif (with 86% sequence match). The promoter region of this gene maps between nts −159 and +141. Here studies of nucleosome positioning in the promoter region using S1 nuclease revealed a hypersensitive region from nts −150 to −50.[105] This 100 bp hypersensitive region exactly co-maps with the region containing the $[(G/C)_3NN]_{48}$ motif.

The discovery of long DNA sequences in the human genome with strong matches to the $[(G/C)_3NN]_n$ motif and the finding that in two cases these sites correspond to regions shown by others to be spared of nucleosomes have several important implications. First, it provides the most compelling evidence to date that the differences in free energy measured in vitro in high salt (which are in the range of a few kilocalories) are biologically significant. Indeed, the two most common human triplet diseases, DM and FraX, involve triplet expansions which bracket the range from strongest to weakest nucleosome assembly elements. The second and broader implication of this work is that the sites composed of long repeats of $(C/G)_3NN$ pentanucleotides present in the 5′ control regions of genes lacking TATA boxes may be a common occurrence in eukaryotic genomes. Here these elements would serve to loosen the chromatin structure, providing access to the DNA for the transcriptional machinery.

In FraX, AGG interruptions have been shown to stabilize the CCG repeats against expansion.[106–111] Comparison of the data presented in the papers of Wang and Griffith (Figure 13a–d),[76] reveal a greater resistance to nucleosome assembly for the $(CCGNN)_{48}$ DNA in contrast with the similar sized $(CCG)_{76}$ block. This suggests that the interruptions in the latter DNA may significantly impair the ability of this overall element to exclude nucleosomes. Thus the well characterized effects of these interruptions could be to moderate the ability of these long repeat blocks to alter the local chromatin structure.

3.2.3 *Methylation of CCG Enhances Nucleosome Exclusion*

In eukaryotic cells, methylation of CpG dinucleotides by DNA methyltransferase has been found to inhibit directly gene expression.[112] This enzyme places a methyl group at C-5 of the cytosine residues of 5′CpG3′ dinucleotides and has been shown to play an important role in embryonic development.[113] Repression of genes by DNA methylation has been studied extensively[112] and might serve as a regulatory step to turn off a gene in order to distinguish alleles of different parental origins (genomic imprinting), or when a gene is not needed, or is subject to dosage compensation (X chromosome inactivation). The inhibition of transcription by methylation may be mediated by the binding of a methyl-CpG binding protein to DNA sequences containing methylated CpG dinucleotides.[114,115] In the context of FraX, Sutcliffe et al. have demonstrated that long CCG repeats alone cannot repress FMR-1 transcription whereas DNA methylation is required for transcriptional repression of the FMR-1 gene.[116] Furthermore, DNA demethylation by 5-azadeoxycytidine has been shown to reactivate the FMR-1 gene in patient cell lines.[117] These observations prompted studies to determine how methylation of the CCG tracts would influence their already poor ability to assemble into chromatin.

SssI methylase was used in our studies[77] to place a methyl group on the C-5 position of cytosines within the 5′CpG3′ dinucleotides. Two sets of experiments were carried out in this study. In one, a 262 bp fragment from pUC19 was compared with a 282 bp DNA containing 76 tandem CCG repeats. In the second, the same pUC19 fragment was compared with a 268 bp fragment containing 48 contiguous CCGNN repeats as described above. The number of possible methylation sites for the pUC19 fragment, the CCG-containing DNA, and the CCGNN-containing DNA is 40, 164, and 102, respectively. Here the level of methylation was estimated from the cleavage pattern by the methylation-sensitive restriction endonuclease HhaI. Competitive nucleosome reconstitution was then used to measure the energetics of nucleosome formation over the CCG and CCGNN repeats as compared with a pUC 19 DNA of the same size. This analysis (Figure 14a–c) showed that DNA methylation had no measurable influence on nucleosome formation for the pUC 19 fragment, an observation noted in several previous studies.[118–120]

In contrast, the $(CCG)_{76}$ DNA methylated at 85% of the available sites was 2.0 ± 0.2-fold less effective in nucleosome assembly compared with the same $(CCG)_{76}$ DNA fragment that was unmethylated, and 4.4 ± 0.4-fold less effective relative to the pUC19 fragment. The difference in free energy between the methylated (at 85%) and the unmethylated $(CCG)_{76}$ DNA is 405 ± 44 cal mol^{-1}. With respect to the pUC19 DNA, the methylated (at 85%) $(CCG)_{76}$ DNA is 880 ± 54 cal mol^{-1} less favorable (Figure 11). Similar results were obtained when the pUC19 fragment was compared with the 268 bp fragment containing 48 tandem CCGNN repeats. The $(CCGNN)_{48}$ DNA fragment methylated at 62% of the available sites was 2.1 ± 0.3- and 12.6 ± 1.6-fold less effective in nucleosome formation compared with the unmethylated $(CCGNN)_{48}$ DNA and the pUC 19 fragment, respectively (Figures 11 and 14b). These efficiencies correspond to 425 ± 93 and $1498 + 75$ cal mol^{-1} differences in free energy, respectively. The greatest difference was observed with a fragment

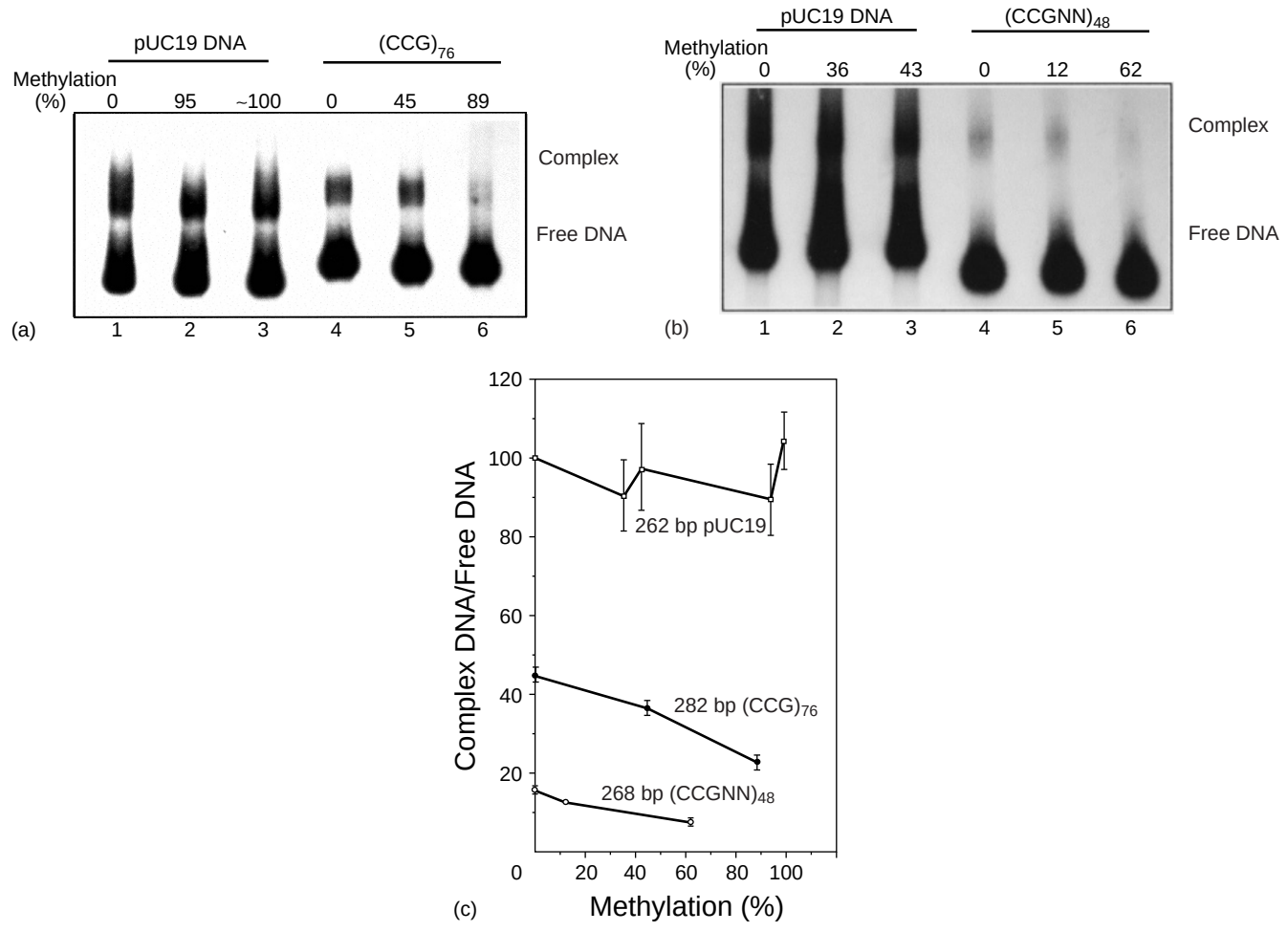

Figure 14 Competitive nucleosome reconstitution with methylated DNAs. Competitive nucleosome reconstitution was carried out comparing (a) a 262 bp pUC19 fragment with a 282 bp DNA containing 76 tandem CCG repeats at three different levels of methylation. Lanes 1–3, the pUC19 fragments methylated at 0, 95, and ~100% of the available sites respectively; lanes 4–6, the $(CCG)_{76}$ DNA fragment methylated at 0, 45, and 89% of the available sites respectively. (b) Comparison of the 262 bp pUC19 fragment to a 268 bp DNA containing 48 tandem CCGNN repeats at three different levels of methylation. Lanes 1–3, the pUC19 fragments methylated at 0, 36, and 43% of the available sites respectively; lanes 4–6, the $(CCGNN)_{48}$ DNA methylated at 0, 12, and 62% of the available sites respectively. (c) Dependence of nucleosome assembly on methylation of CpG dinucleotides. Data from the competitive nucleosome reconstitution experiments are shown. Each DNA was reconstituted in three separate but identical experiments, and the fraction of DNA in the nucleosome-assembled and nucleosome-free DNA bands was measured by PhosphorImager scanning.

containing the $(CCGNN)_{48}$ repeat block. When this DNA was highly methylated, it was nearly 13-fold less effective in nucleosome assembly compared with a mixed sequence DNA of the same size.

Wang and Griffith[77] presented a general model for the role of CCG repeats and methylation in the generation and expression of fragile sites containing these triplet repeats. It was proposed that long repeating CCG triplet blocks inhibit chromatinization of DNA which would increase the accessibility of these regions to DNA methyltransferase. Next, methylation of the repeats would further repress chromatinization and make the DNA more available for binding by methyl CpG binding proteins which have been shown to repress transcription. The net effect of these changes would be the expression of the fragile site and the genetic repression of nearby genes such as the FMR-1 gene.

Godde et al.[121] also examined the ability of CCG repeats to assemble into nucleosomes and found that DNA containing a short CCG repeat (13 copies), upon methylation, showed preferential nucleosome assembly with an affinity similar to that of the $(CTG)_{10}$-containing DNA. However, in a longer CCG DNA (74 copies), methylation represses the ability to assemble nucleosomes to a level similar to mixed sequence DNA. However, the mechanism for the different effects of

methylation and repeat number on nucleosome assembly is not clear.

3.3 Summary

The expanded CTG and CCG triplets were shown to generate nucleosomes with extremely unusual properties. We have demonstrated that DNAs containing 130 copies of CTG repeats assemble into nucleosomes with a stability nine times higher than any previously known natural DNA element. Moreover, expanded CCG repeats show nucleosome exclusion and methylation of the CCG repeats further enhances this exclusion, providing a possible molecular explanation for the basis of fragile sites in chromosomes. Finally, a DNA motif $(CCGNN)_n$, based on the CCG repeat, excludes nucleosomes in vitro and is a DNase I hypersensitive site in yeast (unpublished data).

These findings provide us with a hypothesis in which expanded CTG and CCG triplets result in chromatin with unusual properties which can cause harmful biological effects including the generation of fragile sites and alteration of gene expression. Further, studies of the effect of these simple repeats in forming chromatin and, in the future, the ability of such templates to be replicated or transcribed will contribute greatly to our knowledge of the role of chromatin structure in fundamental biological processes.

ACKNOWLEDGMENTS

I express my deepest appreciation to Dr Jack D. Griffith for his support. Most of the work described here was carried out in his laboratory.

ABBREVIATIONS AND ACRONYMS

ATPγS	Adenosine-5′-*O*-(3-thiotriphosphate)
bp	Basepair
CF	Cystic Fibrosis
DEP	Diethyl Pyrocarbonate
DM	Myotonic Dystrophy
DMAHP	Myotonic Dystrophy Locus-associated Homeodomain Protein
DMPK	Myotonic Dystrophy Protein Kinase
EM	Electron Microscopy
FraX	Fragile X Syndrome
NMR	Nuclear Magnetic Resonance
nt	Nucleotide
PCR	Polymerase Chain Reaction

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Molecules, Properties and Detection of Single • DNA Probes • Electron Tomography of Chromosome Structure • Nuclear Magnetic Resonance and Nucleic Acid Structures

REFERENCES

1. G. Streisinger, Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, M. Inouye, 'Frameshift Mutations and the Genetic Code', *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 77–84 (1966).
2. P. Modrich, 'Mechanisms and Biological Effects of Mismatch Repair', *Annu. Rev. Genet.*, **25**, 229–253 (1991).
3. C. Dohet, R. Wagner, M. Radman, 'Methyl-directed Repair of Frameshift Mutations in Heteroduplex DNA', *Proc. Natl. Acad. Sci. USA*, **83**, 3395–3397 (1986).
4. B.A. Learn, R.H. Grafstrom, 'Methyl-directed Repair of Frameshift Heteroduplexes in Cell Extracts from *Escherichia coli*', *J. Bacteriol.*, **171**, 6473–6481 (1989).
5. R. Fishel, 'Mismatch Repair, Molecular Switches, and Signal Transduction', *Genes Dev.*, **12**(14), 2096–2101 (1998).
6. P. Modrich, 'Mismatch Repair, Genetic Stability, and Cancer', *Science*, **266**, 1959–1960 (1994).
7. G.R. Sutherland, R.I. Richards, 'Simple Tandem DNA Repeats and Human Genetic Disease', *Proc. Natl. Acad. Sci. USA*, **92**, 3636–3641 (1995).
8. S.T. Warren, 'The Expanding World of Trinucleotide Repeats', *Science*, **271**, 1374–1375 (1996).
9. H.L. Paulson, K.H. Fischbeck, 'Trinucleotide Repeats in Neurogenetic Disorders', *Annu. Rev. Neurosci.*, **19**, 79–107 (1996).
10. V. Campuzano, L. Montermini, M.D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelle, F. Zara, J. Canizares, H. Koutnikova, S.I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P.I. Patel, S. Di Donato, J.L. Mandel, S. Cocozza, M. Koeig, M. Pandolfo, 'Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion', *Science*, **271**, 1423–1427 (1996).
11. M. Mahadevan, C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy, S. LeBlond, J. Earle-McDonald, P. DeJong, B. Wieringa, R. Korneluk, 'Myotonic Dystrophy

Mutation: an Unstable CTG Repeat in the 3′ Untranslated Region of the Gene', *Science*, **255**, 1253–1255 (1992).

12. Y. Fu, A. Pizzuti, R. Fenwich, J. King, S. Rajnarayan, P. Dunne, J. Duble, A. Nasser, T. Ashizawa, P. DeJong, B. Wieringa, R. Korneluk, M. Perryman, H. Epstein, T. Caskey, 'An Unstable Triplet Repeat in a Gene Related to Myotonic Dystrophy', *Science*, **255**, 1256–1258 (1992).
13. E. Kremer, M. Pritchard, M. Lynch, S. Yu, S. Holman, E. Baker, S. Warren, D. Schlessinger, G. Sutherland, R. Richards, 'Mapping of DNA Instability at the Fragile X to a Trinucleotide Repeat Sequence', *Science*, **252**, 1711–1718 (1991).
14. Y.H. Fu, D.P.A. Kuhl, A. Pizzuti, M. Pieretti, J.S. Sutcliffe, S. Richards, A.J.M.H. Verkerk, J.J.A. Holden, R.G. Fenwich, S.T. Warren, B.A. Oostra, D.L. Nelson, T. Caskey, 'Variation of the CGG Repeat at the Fragile X Site Results in Genetic Instability: Resolution of the Sherman Paradox', *Cell*, **67**, 1047–1058 (1991).
15. A.J.M.H. Verkerk, M. Pieretti, J.S. Sutcliffe, Y.H. Fu, D.P.A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F. Zhang, B.E. Eussen, G.J.B. van Ommen, L.A.J. Blonden, G.J. Riggins, J.L. Chastain, C.B. Kunst, H. Galaard, C.T. Caskey, D.L. Nelson, B.A. Oostra, S.T. Warren, 'Identification of a Gene (FMR-1) Containing a CGG Repeat Coincident With a Breakpoint Cluster Region Exhibiting Length Variation in Fragile X Syndrome', *Cell*, **65**, 905–914 (1991).
16. S.A. Woodson, D.M. Crothers, 'Structural Model for an Oligonucleotide Containing a Bulged Gaunosine by NMR and Energy Minimization', *Biochemistry*, **27**, 3130–3141 (1988).
17. Y.T. van den Hoogen, A.A. van Beuzekom, H. van den Elst, G.A. van der Marel, J.H. van Boom, C. Altona, 'Extra Thymidine Stacks into the d(CTGGTGCGG). d(CCGCCCAG) Duplex. An NMR and Model-building Study', *Nucleic Acids Res.*, **16**, 2971 (1988).
18. M.W. Kalnik, D.G. Norman, P.F. Swann, D.J. Patel, 'Conformation of Adenosine Bulge-containing Deoxytridecanucleotide Duplexes in Solution. Extra Adenosine Stacks into Duplex Independent of Flanking Sequence and Temperature', *J. Biol. Chem.*, **264**, 3702–3712 (1989).
19. M.W. Kalnik, D.G. Norman, M.G. Zagorski, P.F. Swann, D.J. Patel, 'Conformational Transitions in Cytidine Bulge-containing Deoxytridecanucleotide Duplexes: Extra Cytidine Equilibrates Between Looped Out (Low Temperature) and Stacked (Elevated Temperature) Conformations in Solution', *Biochemistry*, **28**, 294–303 (1989).
20. M.A. Rosen, D. Live, D.J. Patel, 'Comparative NMR Study of A(n)-bulge Loops in DNA Duplexes: Intrahelical Stacking of A, A–A, and A–A–A Bulge Loops', *Biochemistry*, **31**, 4004–4014 (1992).
21. M.A. Rosen, L. Shapiro, D.J. Patel, 'Solution Structure of a Trinucleotide A–T–A Bulge Loop Within a DNA Duplex', *Biochemistry*, **31**, 4015–4026 (1992).
22. A. Bhattacharyya, D.M.J. Lilley, 'The Contrasting Structures of Mismatched DNA Sequences Containing Looped-out Bases (Bulges) and Multiple Mismatches (Bubbles)', *Nucleic Acids Res.*, **17**, 6821–6840 (1989).
23. C.-H. Hsieh, J.D. Griffith, 'Deletions of Bases in One Strand of Duplex DNA, in Contrast to Single-base Mismatches, Produce Highly Kinked Molecules: Possible Relevance to the Folding of Single-stranded Nucleic Acids', *Proc. Natl. Acad. Sci. USA*, **86**, 4833–4837 (1989).
24. J.A. Rice, D.M. Crothers, 'DNA Bending by the Bulge Defect', *Biochemistry*, **28**, 4512–4516 (1989).
25. Y.H. Wang, J.D. Griffith, 'Effects of Bulge Composition and Flanking Sequence on the Kinking of DNA by Bulged Bases', *Biochemistry*, **30**, 1358–1363 (1991).
26. D. Patel, L. Shapiro, D. Hare, 'Conformation of DNA Base Pair Mismatches in Solution', in *Nucleic Acids and Molecular Biology*, eds. F. Eckstein, D.M.J. Lilley, Springer, Berlin, 70–84, 1987.
27. O. Kennard, 'The Molecular Structure of Base Pair Mismatches', in *Nucleic Acids and Molecular Biology*, eds. F. Eckstein, D.M.J. Lilley, Springer, Berlin, 25–52, 1987.
28. A. Bhattacharyya, D.M.J. Lilley, 'Single-base Mismatches in DNA Long and Short-range Structure Probed by Analysis of Axis Trajectory and Local Chemical Reactivity', *J. Mol. Biol.*, **209**, 583–597 (1989).
29. B. Kerem, J.M. Rommens, J.A. Buchanan, D. Markiewicz, T.K. Cox, A. Chakravarti, M. Buchwald, L.-C. Tsui, 'Identification of the Cystic Fibrosis Gene: Genetic Analysis', *Science*, **245**, 1073–1080 (1989).
30. F. Anglani, L. Picci, C. Camporese, F. Zacchello, 'Heteroduplex Formation in Polymerase Chain Reaction', *Am. J. Hum. Genet.*, **47**, 169–170 (1990).
31. P.E. Barker, 'Detection of Cystic Fibrosis Mutations on Nondenaturing Acrylamide Gels (NDGE)', in *Genome Mapping and Sequencing*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 23, 1990.
32. J. Rommens, B. Kerem, W. Greer, P. Chang, L.-C. Tsui, P. Ray, 'Rapid Nonradioactive Detection of the Major Cystic Fibrosis Mutation', *Am. J. Hum. Genet.*, **46**, 395–396 (1990).
33. B.L. Triggs-Raine, R.A. Gravel, 'Diagnostic Heteroduplexes: Simple Detection of Carriers of a 4-bp Insertion Mutation in Tay–Sachs Disease', *Am. J. Hum. Genet.*, **46**, 183–184 (1990).
34. S. Shore, R. Myerowitz, 'Polymerase Chain Reaction-generated Heteroduplexes from Ashkenazi Tay–Sachs Carriers with an Insertion Mutation Can be Detected on Agarose Gels', *Am. J. Hum. Genet.*, **47**, 169 (1990).
35. J.R. Riordan, J.M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J.-L. Chou, M.L. Drumm, M.C. Iannuzzi, F.S. Collins, L.-C. Tsui, 'Identification of the Cystic

Fibrosis Gene: Cloning and Characterization of Complementary DNA', *Science*, **245**(4922), 1066–1073 (1989).
36. J.D. Griffith, G. Christiansen, 'Electron Microscope Visualization of Chromatin and Other DNA–Protein Complexes', *Annu. Rev. Biophys. Bioeng.*, **7**, 19–35 (1978).
37. C. Lu, R.H. Scheuermann, H. Echols, 'Capacity of RecA Protein to Bind Preferentially to UV Lesions and Inhibit the Editing Subunit (Epsilon) of DNA Polymerase III: A Possible Mechanism for SOS-induced Targeted Mutagenesis', *Proc. Natl. Acad. Sci. USA*, **83**, 619–623 (1986).
38. M. Rosenberg, H. Echols, 'Differential Recognition of Ultraviolet Lesions by RecA Protein. Possible Mechanism for Preferential Targeting of SOS Mutagenesis to (6–4) Dipyrimidine Sites', *J. Biol. Chem.*, **265**, 20 641–20 645 (1990).
39. Y.B. Shi, J. Griffith, H. Gamper, J. Hearst, 'Evidence for Structural Deformation of the DNA Helix by a Psoralen Diadduct but not by a Monoadduct', *Nucleic Acids Res.*, **16**, 8945–8952 (1988).
40. M. Kojima, M. Suzuki, T. Morita, T. Ogawa, H. Ogawa, M. Tada, 'Interaction of RecA Protein with pBR322 DNA Modified by *N*-hydroxy-2-acetylaminofluorene and 4-hydroxyaminoquinoline 1-Oxide', *Nucleic Acids Res.*, **18**(9), 2707–2714 (1990).
41. R.J. Thresher, J.D. Griffith, 'Intercalators Promote the Binding of RecA Protein to Double-stranded DNA', *Proc. Natl. Acad. Sci. USA*, **87**(13), 5056–5060 (1990).
42. J.A. Blaho, R.D. Wells, 'Left-handed Z-DNA Binding by the RecA Protein of *Escherichia coli*', *J. Biol. Chem.*, **262**(13), 6082–6088 (1987).
43. J. Kim, J. Heuser, M.M. Cox, 'Enhanced RecA Protein Binding to Z DNA Represents a Kinetic Perturbation of a General Duplex DNA Binding Pathway', *J. Biol. Chem.*, **264**(36), 21 848–21 856 (1989).
44. B.F. Pugh, M.M. Cox, 'Stable Binding of RecA Protein to Duplex DNA. Unraveling a Paradox', *J. Biol. Chem.*, **262**(3), 1326–1336 (1987).
45. S.C. Kowalczykowski, J. Clow, R.A. Krupp, 'Properties of the Duplex DNA-dependent ATPase Activity of *Escherichia coli* RecA Protein and Its Role in Branch Migration', *Proc. Natl. Acad. Sci. USA*, **84**(10), 3127–3131 (1987).
46. A.I. Roca, M.M. Cox, 'The RecA Protein: Structure and Function', *Crit. Rev. Biochem. Mol. Biol.*, **25**(6), 415–456 (1990).
47. M.M. Cox, 'The RecA Protein as a Recombinational Repair System', *Mol. Microbiol.*, **5**(6), 1295–1299 (1991).
48. Y.-H. Wang, C.D. Bortner, J. Griffith, 'RecA Binding to Bulge- and Mismatch-containing DNAs. Certain Single Base Mismatches Provide Strong Signals for RecA Binding Equal to Multiple Base Bulges', *J. Biol. Chem.*, **268**(23), 17 571–17 577 (1993).
49. E.N. Trifonov, 'Curved DNA', *CRC Crit. Rev. Biochem.*, **l9**, 89–106 (l985).
50. C.-H. Hsieh, J.D. Griffith, 'The Terminus of SV40 Replication and Transcription Contains a Sharp Sequence-directed Curve', *Cell*, **52**, 535–544 (1988).
51. T.E. Shrader, D.M. Crothers, 'Artificial Nucleosome Positioning Sequences', *Proc. Natl. Acad. Sci. USA*, **86**, 7418–7422 (1989).
52. R.T. Simpson, D.W. Stafford, 'Structural Features of a Phased Nucleosome Core Particle', *Proc. Natl. Acad. Sci. USA*, **80**, 51–55 (1983).
53. D. Rhodes, 'Nucleosome Cores Reconstituted from Poly (dA-dT) and the Octamer of Histones', *Nucleic Acids Res.*, **6**, 1805–1816 (1979).
54. N. Ramsay, 'Deletion Analysis of a DNA Sequence that Positions Itself Precisely on the Nucleosome Core', *J. Mol. Biol.*, **189**, 179–188 (1986).
55. J.M. Gottesfeld, 'DNA Sequence-directed Nucleosome Reconstitution on 5S RNA Genes of *Xenopus lavevis*', *Mol. Cell. Biol.*, **7**, 1612–1622 (1987).
56. K. Dunn, J.D. Griffith, 'The Presence of RNA in a Double Helix Inhibits Its Interaction with Histone Protein', *Nucleic. Acids Res.*, **8**, 555–566 (1979).
57. J. Nickol, M. Behe, G. Felsenfeld, 'Effect of the B–Z Transition in Poly(dG-m^5dC).Poly(dG-m^5dC) on Nucleosome Formation', *Proc. Natl. Acad. Sci. USA*, **79**, 1771–1775 (1982).
58. R.T. Simpson, P. Kunzler, 'Chromatin and Core Particles Formed from the Inner Histones and Synthetic Polydeoxyribonucleotides of Defined Sequence', *Nucleic Acids Res.*, **6**, 1387–1415 (1979).
59. G.R. Kunkel, H.G. Martinson, 'Nucleosomes will not Form on Double-stranded RNA or Over Poly(dA) : Poly (dT) Tracts in Recombinant DNA', *Nucleic Acids Res.*, **9**, 6869–6888 (1981).
60. A. Prunell, 'Nucleosome Reconstitution on Plasmid-inserted Poly(dA) : Poly(dT)', *EMBO J.*, **1**, 173–179 (1982).
61. V. Iyer, K. Struhl, 'Poly(dA : dT), a Ubiquitous Promoter Element that Stimulates Transcription via Its Intrinsic DNA Structure', *EMBO J.*, **14**, 2570–2579 (1995).
62. H. Harley, J. Brook, S. Rundle, S. Crow, W. Reardon, A. Buckler, P. Harper, D. Housman, D. Shaw, 'Expansion of an Unstable DNA Region and Phenotypic Variation in Myotonic Dystrophy', *Nature*, **355**, 545–5446 (l992).
63. J. Brook, M. McCurrach, H. Harley, A. Buckler, D. Church, H. Aburatani, D. Hunter, V. Stanton, J. Thirion, T. Hudson, R. Sohn, B. Zemelman, R. Snell, S. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonen, K. Joyhnson, P. Harper, D. Shaw, D. Housman, 'Molecular Basis of Myotonic Dystrophy: Expansion of a Trinucleotide (CTG) Repeat at the 3′ End of a Transcript Encoding a Protein Kinase Family Member', *Cell*, **68**, 799–808 (l992).
64. C. Tsilfidis, A. McKenzie, G. Mettler, J. Barcelo, R. Korneluk, 'Correlation Between CTG Trinucleotide

Repeat Length and Frequency of Severe Congenital Myotonic Dystrophy', *Nat. Genet.*, **1**, 192–195 (l992).
65. J. Redman, G. Fenwick, Y.-H. Fu, A. Pizzuti, T. Casky, 'Relationship Between Parental Trinucleotide Repeat Length and Severely of Myotonic Dystrophy in Offspring', *J. Am. Med. Assoc.*, **269**, 1960–1965 (1993).
66. C.A. Boucher, S.K. King, N. Carey, R. Krahe, C.L. Winchester, S. Rahman, T. Creavin, P. Meghji, M.E. Bailey, F.L. Chartier, S.D. Brown, M.J. Siciliano, K. Johnson, 'A Novel Homeodomain-encoding Gene is Associated with a Large CpG Island Interrupted by the Myotonic Dystrophy Unstable $(CTG)_n$ Repeat', *Hum. Mol. Genet.*, **4**, 1919–1925 (1995).
67. T.R. Klesert, A.D. Otten, T.D. Bird, S.J. Tapscott, 'Trinucleotide Repeat Expansion at the Myotonic Dystrophy Locus Reduces Expression of DMAHP', *Nat. Genet.*, **16**, 402–406 (1997).
68. C.A. Thornton, J.P. Wymer, Z. Simmons, C. McClain, R.T. Moxley III, 'Expansion of the Myotonic Dystrophy CTG Repeat Reduces Expression of the Flanking DMAHP Gene', *Nat. Genet.*, **16**, 407–409 (1997).
69. A.D. Otten, S.J. Tapscott, 'Triplet Repeat Expansion in Myotonic Dystrophy Alters the Adjacent Chromatin Structure', *Proc. Natl. Acad. Sci. USA*, **92**, 5465–5469 (1995).
70. Y.-H. Wang, S. Amirhaeri, S. Kang, R.D. Wells, J.D. Griffith, 'Preferential Nucleosome Assembly at DNA Triplet Repeats from the Myotonic Dystrophy Gene', *Science*, **265**, 669–671 (l994).
71. J.J. Hayes, T.D. Tullius, A.P. Wolffe, 'The Structure of DNA in a Nucleosome', *Proc. Natl. Acad. Sci. USA*, **87**, 7405–7409 (l990).
72. J.J. Hayes, D.J. Clark, A.P. Wolfe, 'Histone Contributions to the Structure of DNA in the Nucleosome', *Proc. Natl. Acad. Sci. USA*, **88**, 6829–6833 (1991).
73. J.J. Hayes, J. Bashkin, T.D. Tullius, A.P. Wolfe, 'The Histone Core Exerts a Dominant Constraint on the Structure of DNA in a Nucleosome', *Biochemistry*, **30**, 8434–8440 (1991).
74. Y.-H. Wang, J.D. Griffith, 'Expanded CTG Triplet Blocks from the Myotonic Dystrophy Gene Create the Strongest Known Natural Nucleosome Positioning Elements', *Genomics*, **25**, 570–573 (1995).
75. Y.-H. Wang, R. Gellibolian, M. Shimizu, R.D. Wells, J.J.D. Griffith, 'Long CCG Triplet Repeat Blocks Exclude Nucleosomes: A Possible Mechanism for the Nature of Fragile Sites in Chromosomes', *J. Mol. Biol.*, **263**, 511–516 (1996).
76. Y.-H. Wang, J.D. Griffith, 'The $[(G/C)_3NN]_n$ Motif: A Common DNA Repeat that Excludes Nucleosomes', *Proc. Natl. Acad. Sci. USA*, **93**, 8863–8867 (1996).
77. Y.H. Wang, J.D. Griffith, 'Methylation of Expanded CCG Triplet Repeat DNA from Fragile X Syndrome Patients Enhances Nucleosome Exclusion', *J. Biol. Chem.*, **271**, 22 937–22 940 (1996).
78. J.S. Godde, A.P. Wolffe, 'Nucleosome Assembly on CTG Triplet Repeats', *J. Biol. Chem.*, **271**, 15 222–15 229 (1996).
79. T.R. Klesert, A.D. Otten, T.D. Bird, S.J. Tapscott, 'Trinucleotide Repeat Expansion at the Myotonic Dystrophy Locus Reduces Expression of DMAHP', *Nat. Genet.*, **16**, 402–406 (1997).
80. Y. Fu, D. Friedman, S. Richards, J. Pearlman, R. Gibbs, G. Scarlato, R. Fenwick, C. Caskey, 'Decreased Expression of Myotonin Protein Kinase mRNA and Protein in Adult Form of Myotonic Dystrophy', *Science*, **260**, 235–238 (l993).
81. P. Carango, E. Noble, H.G. Marks, V. Funanage, 'Absence of Myotonic Dystrophy Protein Kinase (DMPK) mRNA as a Result of a Triplet Repeat Expansion in Myotonic Dystrophy', *Genomics*, **18**, 340–348 (l993).
82. L. Sabouri, M. Mahadevan, M. Narang, D. Lee, L. Surh, R. Korneluk, 'Effect of the Myotonic Dystrophy Mutation on mRNA Levels of the DM Gene', *Nat. Genet.*, **4**, 233–237 (l993).
83. K.L. Taneja, M. McCurrach, M. Schalling, D. Housman, R.H. Singer, 'Foci of Trinucleotide Repeat Transcripts in Nuclei of Myotonic Dystrophy Cells and Tissues', *J. Cell Biol.*, **128**, 995–1002 (1995).
84. J. Wang, E. Pegoraro, E. Menegazzo, M. Gennarelli, R.C. Hoop, C. Angelini, E.P. Hoffman, 'Myotonic Dystrophy: Evidence for a Possible Dominant-negative RNA Mutation', *Hum. Mol. Genet.*, **4**, 599–606 (1995).
85. B.M. Davis, M.E. McCurrach, K.L. Taneja, R.H. Singer, D.E. Housman, 'Expansion of a CUG Trinucleotide Repeat in the 3′ Untranslated Region of Myotonic Dystrophy Protein Kinase Transcripts Results in Nuclear Retention of Transcripts', *Proc. Natl. Acad. Sci. USA*, **94**, 7388–7393 (1997).
86. M.G. Hamshere, E.E. Newman, M. Alwazzan, B.S. Athwal, J.D. Brook, 'Transcriptional Abnormality in Myotonic Dystrophy Affects DMPK but not Neighboring Genes', *Proc. Natl. Acad. Sci. USA*, **94**, 7394–7399 (1997).
87. S. Harris, C. Moncrieff, K. Johnson, 'Myotonic Dystrophy: Will the Real Gene Please Step Forward!', *Hum. Mol. Genet.*, **5**, 1417–1423 (1996).
88. G.R. Sutherland, 'Chromosomal Fragile Sites', *Genet. Anal. Tech. Appl.*, **8**, 161–166 (1991).
89. A. Coquelle, E. Pipiras, F. Toledo, G. Buttin, M. Debatisse, 'Expression of Fragile Sites Triggers Intrachromosomal Mammalian Gene Amplification and Sets Boundaries to Early Amplicons', *Cell*, **89**, 215–225 (1997).
90. E.J. Kremer, M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S.T. Warren, D. Schlessinger, G.R. Sutherland, R.I. Richards, 'Mapping of DNA Instability at the Fragile X to a Trinucleotide Repeat Sequence', *Science*, **252**, 1711–1718 (1991).

91. A.J.M.H. Verkerk, M. Pieretti, J.S. Sutcliffe, Y.-H. Fu, D.P.A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F. Zhang, B.E. Eussen, G.-J.B. van Ommen, L.A.J. Blonden, G.J. Riggins, J.L. Chastain, C.B. Kunst, H. Galaard, C.T. Caskey, D.L. Nelson, B.A. Oostra, S.T. Warren, 'Identification of a Gene (FMR-1) Containing a CGG Repeat Coincident with a Breakpoint Cluster Region Exhibiting Length Variation in Fragile X Syndrome', *Cell*, **65**, 905–914 (1991).
92. I.K. Hornstra, D.L. Nelson, S.T. Warren, T.P. Yang, 'High Resolution Methylation Analysis of the FMR-1 Gene Trinucleotide Repeat Region in Fragile X Syndrome', *Hum. Mol. Genet.*, **2**, 1659–1665 (1993).
93. S. Knight, A.V. Flannery, M.C. Hirst, L. Campbell, Z. Christodoulou, S.R. Phelps, J. Pointon, H.R. Middleton-Price, A. Barnicoat, M.E. Pembrey, J. Holland, B.A. Oostra, M. Bobrow, K.E. Davies, 'Trinucleotide Repeat Amplification and Hypermethylation of a CpG Island in FraXE Mental Retardation', *Cell*, **74**, 127–134 (1993).
94. J.E. Parrish, B.A. Oostra, A.J.M.H. Verkerk, C.S. Richards, J. Reynolds, A.S. Spikes, L.G. Shaffer, D.L. Nelson, 'Isolation of a GCC Repeat Showing Expansion in FRAXAF, a Fragile Site Distal to FRAXA and FRAXE', *Nat. Genet.*, **8**, 229–235 (1994).
95. J.K. Nancarrow, E. Kremer, K. Holman, H. Eyre, N.A. Doggett, D.L. Paslier, D.F. Callen, G.R. Sutherland, R.I. Richards, 'Implication of FRA16A Structure for the Mechanism of Chromosomal Fragile Site Genesis', *Science*, **264**, 1938–1941 (1994).
96. C. Jones, L. Penny, T. Mattina, S. Yu, E. Baker, L. Voullaire, W.Y. Langdon, G.R. Sutherland, R.I. Richards, A. Tunnacliffe, 'Association of a Chromosome Deletion Syndrome with a Fragile Site Within the Proto-oncogene CBL2', *Nature*, **376**, 145–149 (1995).
97. D.L. Nelson, 'The Fragile X Syndromes', *Semin. Cell Biol.*, **6**, 5–11 (1995).
98. G.R. Sutherland, R.I. Richards, 'The Molecular Basis of Fragile Site in Human Chromosomes', *Curr. Opin. Genet. Dev.*, **5**, 323–327 (1995).
99. H. Hinds, C. Ashley, J. Sutcliffe, D. Nelson, S. Warren, D. Housman, M. Schalling, 'Tissue Specific Expression of FMR-1 Provides Evidence for a Functional Role in Fragile X Syndrome', *Nat. Genet.*, **1**, 36–43 (l993).
100. A. Gedeon, E. Baker, H. Robinson, M. Partington, B. Gross, A. Manca, B. Korn, A. Poustka, S. Yu, R. Sutherland, 'Fragile X Syndrome Without CCG Amplification has an FMR1 Deletion', *Nat. Genet.*, **1**, 341–344 (l992).
101. J.P. Chaudhuri, 'On the Origin and Nature of Achromatic Lesions', *Chromosome Today*, **3**, 147–151 (1972).
102. M. Shimizu, R. Gellibolian, B.A. Oostra, R.D. Wells, 'Cloning, Characterization, and Properties of Plasmids Containing CGG Triplet Repeats from the FMR-1 Gene', *J. Mol. Biol.*, **258**, 614–626 (1996).
103. T.E. Shrader, D.M. Crothers, 'Effects of DNA Sequence and Histone–Histone Interactions on Nucleosome Placement', *J. Mol. Biol.*, **216**, 69–84 (1990).
104. T. Shimada, K. Inokuchi, A.W. Nienhuis, 'Chromatin Structure of the Human Dihydrofolate Reductase Gene Promoter', *J. Biol. Chem.*, **261**, 1445–1452.
105. G.J. Mavrothalassitis, D.K. Watson, T.S. Papas, 'The Human ETS-2 Gene Promoter: Molecular Dissection and Nuclease Hypersensitivity', *Oncogene*, **5**, 1337–1342 (1990).
106. E.E. Eichler, J.J. Holden, B.W. Popovich, A.L. Reiss, K. Snow, S.N. Thibodeau, C.S. Richards, P.A. Ward, D.L. Nelson, 'Length of Uninterrupted CCG Repeats Determines Instability in the FMR-1 Gene', *Nat. Genet.*, **8**, 8–94 (1994).
107. M.C. Hirst, P.K. Grewal, K.E. Davies, 'Precursor Arrays for Triplet Repeat Expansion at the Fragile X Locus', *Hum. Mol. Genet.*, **3**, 1553–1560 (1994).
108. C. Kunst, S.T. Warren, 'Cryptic and Polar Variation of the Fragile X Repeat Could Result in Predisposing Normal Alleles', *Cell*, **77**, 853–861 (l994).
109. A.L. Reiss, H.H. Kazazian, Jr, C.M. Krebs, A. McAughan, C.D. Boehm, M.T. Abrams, D.L. Nelson, 'Frequency and Stability of the Fragile X Syndrome', *Hum. Mol. Genet.*, **3**, 393–398 (1994).
110. K. Snow, D.J. Tester, K.E. Kruckeberg, D.J. Schaid, S.N. Thibodeau, 'Sequence Analysis of the Fragile X Trinucleotide Repeat: Implications for the Origin of the Fragile X Mutation', *Hum. Mol. Genet.*, **3**, 1543–1551 (1994).
111. E.E. Eichler, H.A. Hammond, J.N. MacPherson, P.A. Ward, D.L. Nelson, 'Population Survey of the Human FMR1 CGG Repeats Structure Suggests Biased Polarity for the Loss of AGG Interruption', *Hum. Mol. Genet.*, **4**, 2199–2208 (1995).
112. P.H. Tate, A.P. Bird, 'Effects of DNA Methylation on DNA-binding Proteins and Gene Expression', *Curr. Opin. Genet. Dev.*, **3**, 226–231 (1993).
113. E. Li, T.H. Bestor, R. Jaenisch, 'Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality', *Cell*, **69**, 915–926 (1992).
114. J. Boyes, A. Bird, 'DNA Methylation Inhibits Transcription Indirectly via a Methyl-CpG Binding Protein', *Cell*, **64**, 1123–1134 (1991).
115. J. Boyes, A. Bird, 'Repression of Genes by DNA Methylation Depends on CpG Density and Promoter Strength: Evidence for Involvement of a Methyl-CpG Binding Protein', *EMBO J.*, **11**, 327–333 (1992).
116. J.S. Sutcliffe, D.L. Nelson, F. Zhang, M. Pieretti, C.T. Caskey, D. Saxe, S.T. Warren, 'DNA Methylation Represses FMR-1 Transcription in Fragile X Syndrome', *Hum. Mol. Genet.*, **1**, 397–400 (1992).
117. P. Chiurazzi, M.G. Pomponi, R. Willemsen, B.A. Oostra, G. Neri, 'In vitro Reaction of the FMR1 Gene Involved in Fragile X Syndrome', *Hum. Mol. Genet.*, **7**, 109–113 (1998).

118. G. Felsenfeld, J. Nickol, M. Behe, J.D. McGhee, D. Jackson, 'Methylation and Chromatin Structure', *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 577–584 (1982).
119. H.R. Drew, M.J. McCall, 'Structural Analysis of a Reconstituted DNA Containing Three Histone Octamers and Histone H5', *J. Mol. Biol.*, **197**, 485–511 (1987).
120. E.W. Englander, A.P. Wolffe, B.H. Howard, 'Nucleosome Interactions with a Human Alu Element: Transcriptional Repression and Effects of Template Methylation', *J. Biol. Chem.*, **268**, 19 565–19 573 (1993).
121. J.S. Godde, S.U. Kass, M.C. Hirst, A.P. Wolffe, 'Nucleosome Assembly on Methylated CGG Triplet Repeats in the Fragile X Mental Retardation Gene 1 Promoter', *J. Biol. Chem.*, **271**, 24 325–24 328 (1996).

Electron Tomography of Chromosome Structure

Peter Engelhardt
Haartman Institute, University of Helsinki, Helsinki, Finland

1 Introduction **2**
1.1 Chromosome Structure 2
1.2 Electron Tomography Compared with Earlier Methods 9
2 Preparation Steps for Chromosomes **9**
2.1 General 9
2.2 Chromosome Isolation and Whole Mounts 10
2.3 Pretreatments: Salt Extraction and Enzyme Digestions 12
2.4 Fixation of Chromosomes 13
3 Preservation Methods for Three-dimensional Reconstructions **15**
3.1 Critical-point Drying 15
3.2 Drying from *tert*-Butanol 16
3.3 Freeze-drying 16
3.4 Cryo-electron Tomography 17
3.5 Section Preparations 17
4 Data Collection **18**
4.1 Recording of Tilt Series 18
4.2 Manual Recording 19
4.3 Digitalization of Photographic Negatives 19
4.4 Automatic Recording 20
5 Electron Tomography Methods **21**
5.1 Alignment of Tilt Series 21
5.2 Reconstruction Methods 22
5.3 Weighted Back-projection Method 22
5.4 Maximum Entropy Method 22
5.5 Dual-axis Tilting Method 23
6 Rendering of Reconstructed Chromosomes **24**
6.1 Visualization of Three-dimensional Reconstructions 24
6.2 Producing Stereo Animations from Three-dimensional Volumes 27
6.3 Computer-aided Virtual Environment 28
7 Immunoelectron Tomography **29**
8 In Situ Hybridization Electron Tomography **31**
9 Hard Copies, Production of Solid Three-dimensional Models **32**
10 Method Development **32**
Acknowledgments **32**
Abbreviations and Acronyms **33**
Related Articles **33**
References **34**

Three-dimensional (3-D) structures at the molecular level of chromosomes and other cell structures are far beyond the reach of ordinary or confocal (3-D) light microscopy (LM). The electron tomography method (ETM), i.e. transmission electron microscopy (TEM) tomography, is able to provide methods of 3-D reconstruction of macromolecular assemblies, providing insights into the qualitative and quantitative spatial comprehension of macromolecular structures. ETM is a powerful method for developing reliable immunoelectron tomography (IET) and in situ hybridization analysis at the molecular level.

With ETM the data are collected with TEM by taking tilt series, i.e. projection images from different viewing angles over the object. By combining such projections taken over a sufficient angular range (from 0 to $\pm 60^{\circ}$ manually, e.g. in 3° increments), a wholly transparent 3-D representation of the object is recovered with necessary details. This requires the use of suitable computational methods.

Automatic single-axis tilt-series data registration is possible today that will not only allow the collection of data with increments as small as 1°, for high-resolution reconstructions of bulky whole mounts, but will also protect the specimen from electron beam damage (dose reduced by a factor of 10–100). Only automation of ETM procedures makes it possible to collect tilt series of cryo-electron microscopy (EM) preparations of unfixed and unstained preparations in vitrified ice, i.e. fully hydrated samples close to the native state.

Novel methods and tools for electron tomography are available that have been developed in different laboratories. Our electron tomography programs are automated to a high degree and therefore very fast and easy to use, in comparison with other program kits that are available. Specifically, our arsenal includes advanced tomography procedures based on the maximum entropy method (MEM) (e.g. 16-bit gray-scale MEM).

A collection of preparative methods for preserving whole-mounted chromosomes and cells for ETM studies has been developed. Examples of 3-D reconstructions of human chromosomes are shown at high resolution (3–15 nm).

It is considered that ETM allows 3-D reconstructions, depending on the thickness of the preparation, with a

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

resolution in the range 2.5–8.5 nm. In this respect, some success was recently made as we were able to detect nanoprobes (1.4-nm immunogold markers) with the IET methods developed.

ETM of chromosomes involving localized genes with in situ hybridization and gold-labeling techniques is possible and would allow chromosome mapping in a 3-D configuration with molecular resolution much more accurately than any standard EM method which are affected by superimposing gold labels with their accompanying structures.

Concerning the structure and folding of the DNA in chromosomes, it is important to note that the linear sequence of bases in DNA is accompanied by higher level organizations, composed of loops, twists and folds in the long DNA chain, and, significantly, by DNA–protein interactions, e.g. nucleosomes, supranucleosomes, together with telomerase, topoisomerases and the new scaffold proteins [structural maintenance of proteins (SMC) family] recently described. These so-called 'higher order structures' are certainly very much involved in the regulation of gene expression in chromosomes. Further, it is easy to realize that genes distantly located in the linear sequence map could in fact be topologically associated when DNA is folded in chromosomes, and thus mutually regulated. Sequencing of the genomes does not include and cannot reveal higher order structures. It is only with ETM that genome maps can be brought in relation to 3-D molecular configurations, and it can be refined by combination with in situ hybridization to localize gene sequences and with IET to identify chromosomal proteins.

1 INTRODUCTION

1.1 Chromosome Structure

Higher order chromosome structure is an enigmatic and challenging problem. It is more than 100 years since the word chromosome first appeared in the literature,[1] and almost 50 years since the principle of the DNA structure was established.[2] The discovery of DNA structure was a turning point in biology and initiated the rapid development of molecular biology. We have been treated to detailed descriptions of various biological structures at the molecular level and various roles have been attributed to DNA. As for DNA, the overall principle of replication is clear, but many details even of replication have turned out to be uncompromisingly complex.[3] In addition, the organization of DNA and the regulation of its function in interphase nuclei and its condensation into metaphase chromosomes at mitosis remain a distressing tangle in cell biology.

Difficulties in solving chromosome structure largely stem from the huge amount of DNA wrapped into chromosomes. Yet the packing and organization of the DNA in many phages and viruses with basically less DNA are also unclear.[4] The length of DNA per diploid human somatic nuclei (diameter about 5–10 μm) is about 1.8 m. In mitosis twice as much is arranged into 46 chromosomes with a total length of about 200 μm. Consequently, the compaction proportion of the DNA in chromosomes is about $10^4 : 1$.[5]

In addition, the length and thickness of the chromosomes change dynamically when chromosomes condense in mitosis. Related changes are observed in meiosis, although the chromosomes are more attenuated and delicate, forming the so-called synaptonemal complexes when chromosomes pair up. In mitosis, early prophase chromosomes appear as long strands that reach full compression at metaphase. After mitosis, the chromosomes revert to their diffuse interphase form. In the interphase nucleus, chromosome positions tend to be conserved in their anaphase–telophase configurations. In this orientation, called the Rabl orientation,[6] centromeres and telomeres occupy opposite regions of the nucleus. The chromosomes somehow rearrange themselves before the next division. Regions of interphase chromosomes, in yeast and *Drosophila*, have recently been shown to be in Brownian motion limited to a subvolume no more than 1% of the total nuclear volume of living cells.[7]

What regulates, causes and lies behind the interphase chromosome movements and prophase contraction at the molecular level is not known, but such movements may be caused by proteins analogous to the contractile proteins of the cytoskeleton. Microtubules have not been observed, in higher eukaryotes, inside nuclei as long as the nuclear envelope (NE) remains intact. In lower eukaryotes, e.g. in yeast cells, bundles of microtubules are detected inside cell nucleus, at mitosis, as the NE remains intact in the budding. Bundles of striated microfilaments[8] have been described in meiosis along the lateral elements of the synaptonemal complex (SC) and shown by immunological methods[9] to react like actin, myosin and tubulin. The actual involvement in meiosis of these and many other immunologically indicated proteins is still controversial.[9] It should be noted that the above-mentioned restricted Brownian motion of chromosome regions has been shown to be inhibited only by the microtubule-depolymerizing agent nocodazole[7] and not by sodium azide, which is known to stop ATPase-driven processes.

In recent years, new groups of proteins have been isolated and shown to participate in chromosome contraction (the SMC family of proteins),[10] sister chromatid cohesion apparatus (cohesin complex) and sister chromatid separation (separins that destroy cohesion).[11]

In the folding of the chromosomes, different levels can be distinguished. Remarkably, there is a better understanding of how the DNA is organized in the chromosomes at the molecular level than at the gross chromosome level. This discrepancy may partly be due to the fact that chromosomes are too small for detailed studies at the LM level and too bulky for EM, that is, with conventional EM methods.

At the molecular level, an outstanding achievement was to reveal the association of DNA with histones to form particles called nucleosomes. The DNA molecule, the double helix itself, is 2.2 nm in diameter and makes a full helical turn of 3.4 nm (along the spiral axis) for each length of 10 nucleotide pairs of the double helix (10.6 pairs for free DNA and 10.2 pairs for nucleosomal DNA).[12] Almost two (1.75) full rounds of DNA totaling 2×83 nucleotide pairs are wound round the periphery of the nucleosome, which is roughly a cylinder 11 nm in diameter and 5.6 nm thick. It contains an octomer histone core, 6.6 nm in diameter and 5.6 nm thick, composed of two copies of each of the four histones H2A, H2B, H3 and H4. From the nucleosomes, now studied in crystallography down to a resolution of 2.8 Å,[12] polypeptide chains of the histones stick out. This is considered to play an important part in the regulation of the DNA–histone core assemblies (assembly vs disassembly), chiefly by acetylation and deacetylation of the histone tails. The DNA and histone cores are linked into a polynucleosome: an 11-nm thick 'beads-on-a-string' fiber. 'Linker DNA' of various length, from 0 to 80 nucleotide pairs, connects the nucleosomes. In this 'beads-on-a-string' association the DNA is packed lengthwise in a proportion of 6 : 1.

Chromatin, i.e. chromosome fibers, is biochemically a poorly defined entity varying with isolation sources and conditions and is composed of DNA with histones and chromosomal nonhistone proteins. The beaded-string organization is observed when chromatin is isolated and treated in buffers (pH 7.4) at low, i.e. hypotonic, salt concentration (10 mM), depleted of the divalent cations Mg^{2+} and Ca^{2+}.[13] The required deficiency of divalent cations in the isolation procedures indicates that the rather loose beaded-string configuration cannot persist under physiological conditions. However, it is not known whether living cells are capable of regulating the concentration of the divalent cations intranuclearly.

In the presence of divalent cations, 1–5 mM Mg^{2+} or/and Ca^{2+}, in neutral buffers, the isolated chromatin fibers observed have a diameter in the range 20–50 nm, depending on the isolation conditions.[13] The size is in better agreement with that of the chromatin and chromosome fibers observed in isolated whole mounts of, e.g. metaphase chromosomes.[13]

The prevalent concept of the folding of the DNA into the 30-nm chromatin fiber structure is likewise attributed only to histones. In this configuration the 11-nm polynucleosomal DNA fiber is anticipated to fold into a 30-nm solenoid with the incorporation of a molecule of histone H1 to each nucleosome. The binding site of H1 at the nucleosome has been defined but the arrangement of H1 in the 30-nm fiber organization is unclear. Recent data have positioned the globular domain of H1 close to the dyad axis, in a polar arrangement of H1 molecules along the nucleosome fiber that may influence gene regulation in addition to its structural role.[14]

The perplexity confronted in understanding the 30-nm chromatin or chromosome fiber structure is evident from various studies.[15] The observations have resulted in numerous models that can be roughly classified as (1) the original solenoid model, (2) some kind of a helical coil of the nucleosome fiber, (3) a beaded (superbead, supranucleosome, nucleomer) arrangement of the nucleosome subunits[16] and (4) a continuously variable zigzag nucleosomal ribbon.[4] The irregular zigzag nucleosomal ribbon has support from electron tomography studies,[4] cryo-electron tomography[17] and recently radiation experiments on chromatin in studies on living cells,[18] affirming linker lengths of 40 nucleotide pairs for an idealized zigzag model.

Estimating that 6–7 nucleosomes of the polynucleosomal string condense into one turn of a 30-nm thick chromatin fiber, the length of DNA is reduced in the proportion of 40 : 1. To obtain the 10^4 : 1 packing in metaphase chromosomes, further packing of 250 : 1 is required.

The prevalent opinion about packing the 30 nm chromosome fiber further into metaphase chromosomes holds that the fiber folds are packed into loops that are fixed to a chromosome scaffold structure composed of nonhistone proteins.[19] Topoisomerase II has been shown to be a main component of the scaffold proteins, making its contribution decisively important in sorting out tangling DNA strands after replication, in the condensing of chromatin loops and in the separation of the daughter chromosomes at mitosis. In addition, as mentioned above, certain new scaffold proteins have been shown to be involved in the condensation of chromosomes and in the separation of sister chromatids in mitosis.[20]

The chromosome scaffold was originally revealed in preparations of isolated metaphase chromosomes treated with high salt buffers (2 M NaCl solution) or dextran sulfate–heparin, to extract the histones and other soluble proteins. The extracted chromosomes were spread on a hypophase of a hypotonic solution. In these preparations clouds of looped dehistonized DNA could be observed surrounding a residual chromosome that principally persisted in the shape of the original metaphase chromosome. Only a residual chromosome scaffold remained after DNase digestion. With the same methods, an analogous nuclear matrix, nuclear skeleton

or cage of interphase nuclei was isolated, surrounded by clouds of looping dehistonized DNA, when digestion with nucleases was omitted.[19]

These observations suggested a looped folding of the chromatin fibers around nonhistone chromosomal scaffold proteins[19] as the principle of organizing the 30-nm chromatin fiber in chromosomes rather than a coiled-coil model. Consequently, the size, association and configuration of the regions [called scaffold-associated region (SAR) and matrix-associated region (MAR)] of looped DNA with the scaffolding proteins have been the object of various studies for the past decades, but contradictions and perplexities have hindered satisfactory concepts.[19] For example, only a slight change in the conditions for the isolation of an interphase nuclear matrix[19] resulted in a residual NE composed only of a nuclear pore complex (NPC) lamina fraction, depleted of an internal nuclear matrix. This may indicate the delicate and dynamic association of chromatin with the NE.[8,21]

It is important to recognize the apparently different organization units of the different levels of chromosome organization: (1) the polynucleosomal DNA fiber, (2) the 30-nm chromatin fiber and (3) looping of the chromatin fiber somehow around the chromosome scaffold, forming the metaphase chromosome. In interphase nuclei, chromatin is organized around analogous structures, variously termed the nuclear matrix, scaffold or skeleton, which seem to be dynamically integrated with the NPC lamina of the NE.[19,21,22]

Although the principal organization of the repeated building blocks at each organization level must be uniform, the variations and complexity in these repeats seem to increase from level to level. At the 'beads-on-a-string' level, the unit is the nucleosome with the linking DNA. For the 30-nm chromatin fiber the unit is 6–7 nucleosomes forming one turn, as assumed in the solenoid model, or their arrangement is more variable and flexible, namely an irregular nucleosomal ribbon according to the zigzag model or, as assumed in the supranucleosomal chromatin model, 7–42 nucleosomes are packed into various-sized 30–65-nm supranucleosomal units.[23] At the next higher organization level the units are the loops bound by the scaffolding structures. The loop with an associated scaffold element seems to be the elementary building block towards higher chromosome organization. Variations of the loop sizes and the clustering of the loops into, e.g. rosette-like configurations have been observed and suggested in different instances.[21,22]

How, then, do we understand the organization of the chromomeres,[13,21,22] as described in the classical studies with LM? Chromomeres are best observed in prophase chromosomes of meiosis and they must in some way be folded by the 30-nm chromatin-loop organization with scaffold proteins.

Also, the bands seen in the giant polytene chromosomes of salivary glands in, e.g. *Drosophila* must in some way fit in, i.e. be a natural consequence of the folding of the 30-nm chromatin fiber. As polytene chromosomes represent tight lateral alignments of bundles of interphase chromosomes (up to 1024 chromosomes in *Drosophila*), the bands distinguished are considered to be apposed chromomeres[13] of interphase chromosomes. Thus chromomeres seem to be organizational entities also in the interphase and not only consequences of early prophase chromosome condensation. This indicates that the folding of the 30-nm chromatin loops is actually preserved in interphase chromosomes, if we assume that polytene chromosomes are not distinctive exceptions. In addition, the size difference of the bands in polytene chromosomes must also be a natural consequence of an adequate chromosome model.[21,22]

Chromosome coiling, i.e. the coiling of the cromonemata[13] in classical descriptions from early LM studies,[24] is another example of large-scale structural chromosome features that must fit with any presumed looping of the 30-nm chromatin with scaffolding proteins, if the model concerned is to be acceptable. In animal cells, chromosome coiling is not usually distinguished after ordinary treatment, i.e. without specific procedures.[13] However, even without any specific treatments it is observed in plants with very large chromosomes (e.g. in *Tradescantia*), and also in living plant cells[24] and in unfixed animal cells processed for EM.[25]

Significantly, chromosome coiling also appears when ordinary chromosome preparations are squashed in 50%[13] or isolated with 60% acetic acid (AA),[22] indicating that chromosome preparations isolated with detergents under 'physiological' conditions or water spread[13] somehow lose or disassemble the coiling. We can also presume that the banding technique that is used today in karyotyping might have the structural background in the underlying chromosome coiling, as has been suggested.[26]

Immunological methods for staining topoisomerase II (a main scaffold protein) reveal a clearly spiralized internal core in chromosome preparations, similar to chromosome coiling.[27] Also, 3-D reconstructions by electron tomography of chromosomes, after histone extraction and DNA digestion, appear as a coiled residual chromosome structure, indicating that chromosome coiling is maintained by scaffold proteins.[28] Accordingly, in chromosome condensation, the coiling is somehow brought about in the folding of chromatin fibers by the scaffold proteins. It remains to be shown if scaffold proteins, when successfully isolated and dissolved, will self-assemble to coiled structures, especially when affiliated with DNA and histones.

Textbook diagrams and models for the higher order chromosome structure fail in giving substantial structural reasons for the several intermediate folding levels from the 30-nm chromatin fiber up to metaphase chromatids.[29] Recent discussions,[30,31] and earlier reviews of chromosome structure[22] have tried to fill this gap. Having their particular merits, most chromosome models do not meet the criteria for an adequate and unifying model – those of comprehensively accounting for different-sized chromomeres of interphase chromosomes (e.g. polytene) and meiotic prophase chromosomes, for chromosome coiling and for the origin of

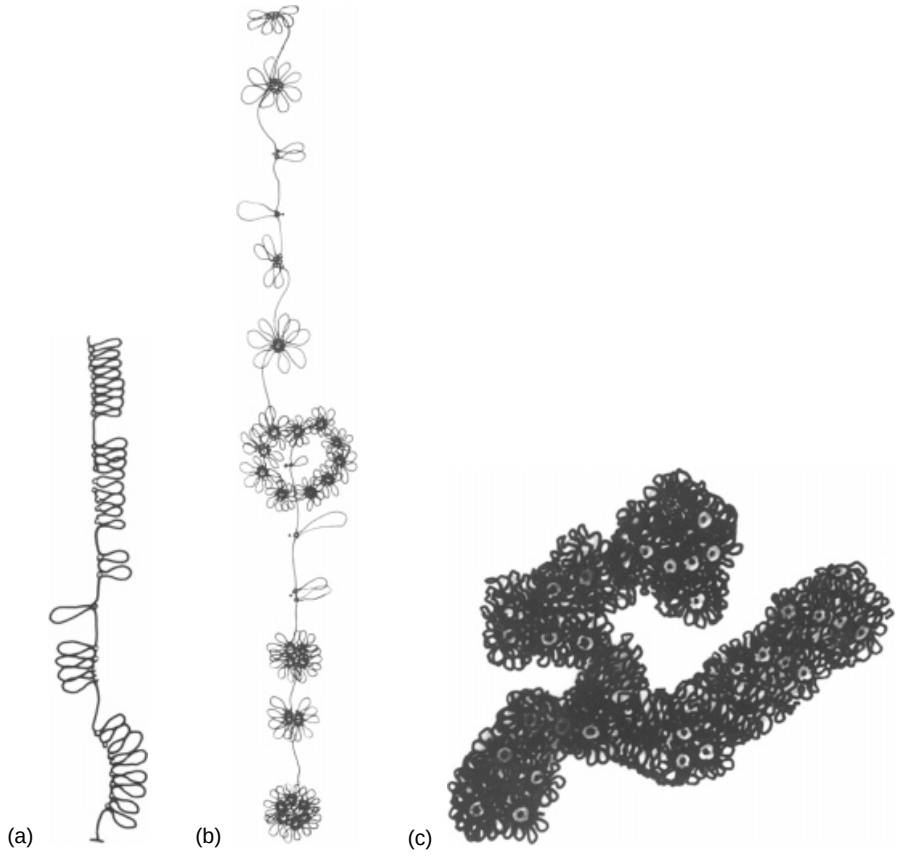

Figure 1 A unified model[8,21,22] of higher-order DNA (DNA loop or chromatin fiber) folding in chromosomes: the loop-and-rosette model. The different categories of data (e.g. polytene chromosomes, Figure 2) fit best with this model. From interphase to metaphase chromosomes. (a) Single (histone-depleted) DNA loops and accumulations (duplications) of these in linear form, i.e. total or strictly local unfolding of chromomeres in a chromatid. (b) Arrangement of various size classes and organization types of chromomeres in a partly stretched-out interphase chromatid. The folding of a continuous DNA by looping: single loops which would correspond to primary replicons; double loops, multiple loops and rosettes (around ring-like cores of scaffold elements, i.e. cyclomeres[8]) and their clustering; representing the origins for different-sized chromomeres during evolution. (c) The condensed form representing the metaphase chromosome. Ring-like structures are embedded in the chromatin. Note the similarity of appearance in a section of the premeiotic X chromosome of Acheta, (d), and in a stereo pair critical-point drying (CPD)-whole mount of a Chinese hamster ovary (CHO) metaphase chromosome (prepared as in section 2.4, Procedure: 1b) shown in (e). There is a complete analogy with a chromatid of the polytene chromosome. Only the condensation is different as no lateral apposition of chromatids occurs. Instead, a linear apposition causes chromatid coiling and linear condensation of the chromosome arms. This would be chiefly due, not to the condensation of interchromomeric DNA, but rather to the tendency of the chromomere core elements (scaffold elements, i.e. cyclomeres and their subunits) to interact and assemble. Proteins of these interactions have recently been identified as the SMC family among scaffold proteins.[10,20] This results in higher-order structures such as chromosome (chromatid) coils, in which chromomeres are arranged in analogy with nucleosomes in the models of 30 nm chromatin fiber, either as in the solenoid model[15] or as in the zigzag nucleosomal ribbon model.[4] (f) 3-D reconstruction of the chromosome shown in (e), from tilt series with 6° increments (section 4.2), with MEM (section 5.4), visualized with BOB (section 6.1), shown in stereo pairs (recorded with SNAPSHOT from Silicon Graphics Inc. (SGI) display). (g) Stereo views, with 10° increments, of the reconstructed chromosome in (f). Note that chromosome coiling is recognizable as chromosome banding and the impression of bands depends also on the viewing angle. (h) Stereo series as in (g). Details of the chromosome arms with coiling (arrowhead) and ring-like structures of cyclomere size (arrows), surrounded with looping fibers of hazy chromatin. Note that these fibers are not detected in the chromosome scaffold, Figure 8(a, e) (cf. the MEM/3-D reconstruction of chromosome scaffold Figure 8a, b, e). (Reproduced from Engelhardt.[22])

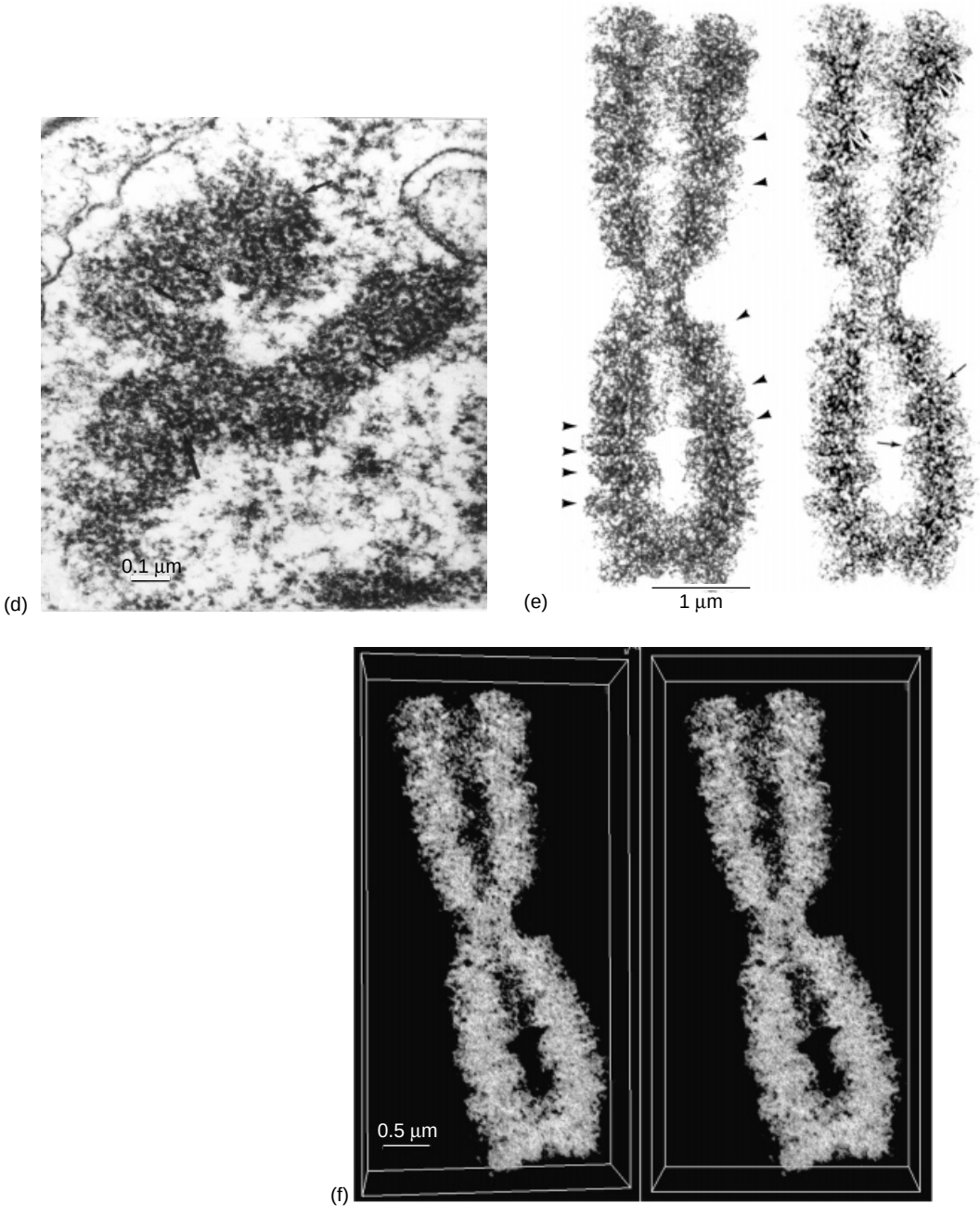

Figure 1 (*Continued*)

and interplay with the chromosome scaffold and nuclear matrix.

It is worth noting that, when the chromosome material (chromatin) coils and (therefore) compacts, the two levels of coiling, (1) the 11-nm fiber to the 30-nm fiber and (2) the extended chromonema to the compacted mitotic chromosome or chromatid, appear closely similar and suggest analogous configurations of condensation (solenoid and zigzag configuration, see the notes in the legend of chromosome model, Figure 1).

Our early investigation of chromosome structure,[8] with a unifying intent, started partly from theoretical postulates that motivated a search for adequate fixing and staining methods with ruthenium red (RR) (developed for intracellular staining, see section 3.5) for meiotic chromosomes in order to reveal details in the attachment of DNA to the NE and to the SC. The methods[8] revealed the distinct presence of ring-like cores (which we named cyclomeres) in the premeiotic X chromosome of *Acheta* and also along the lateral elements of the SC. The ring-like

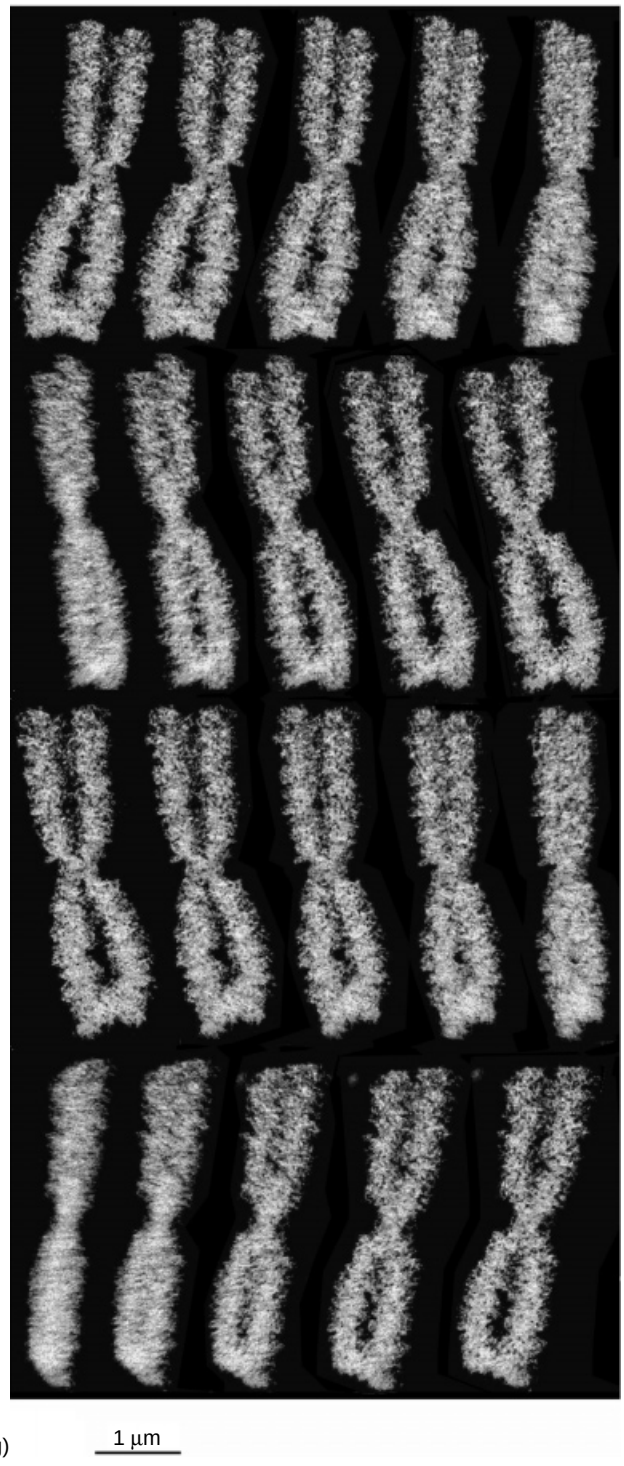

(g)

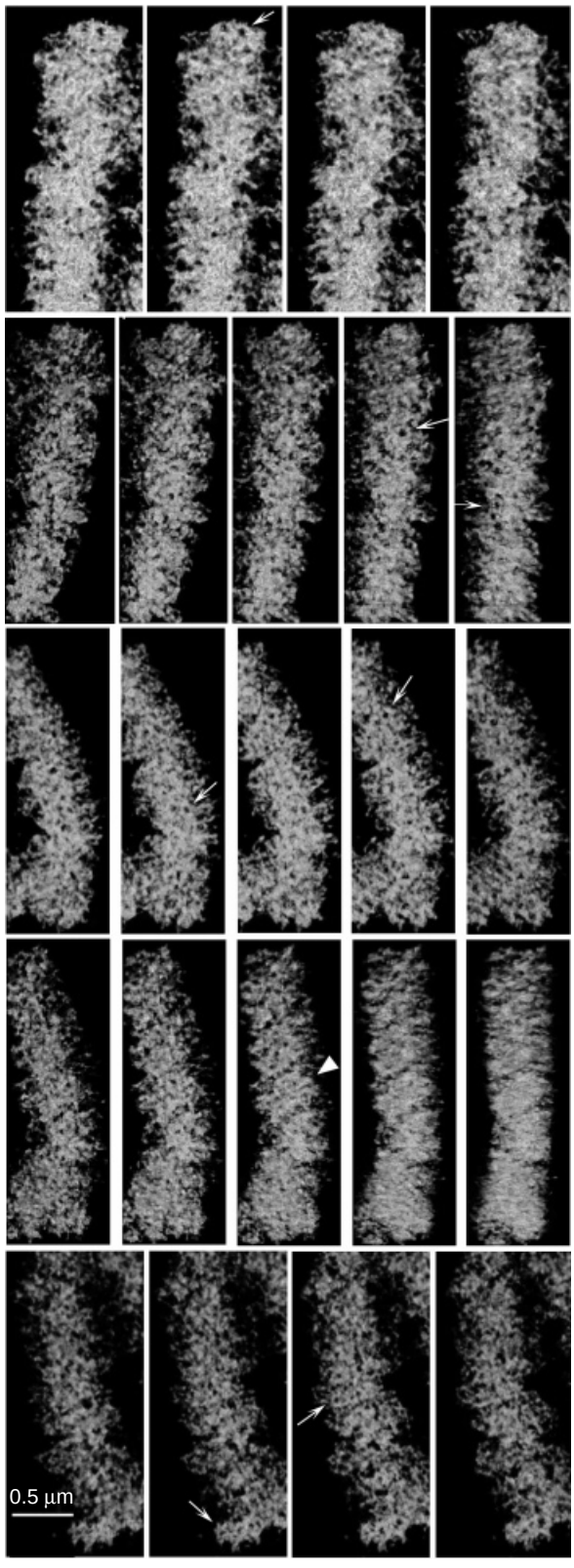

(h)

Figure 1 (*Continued*)

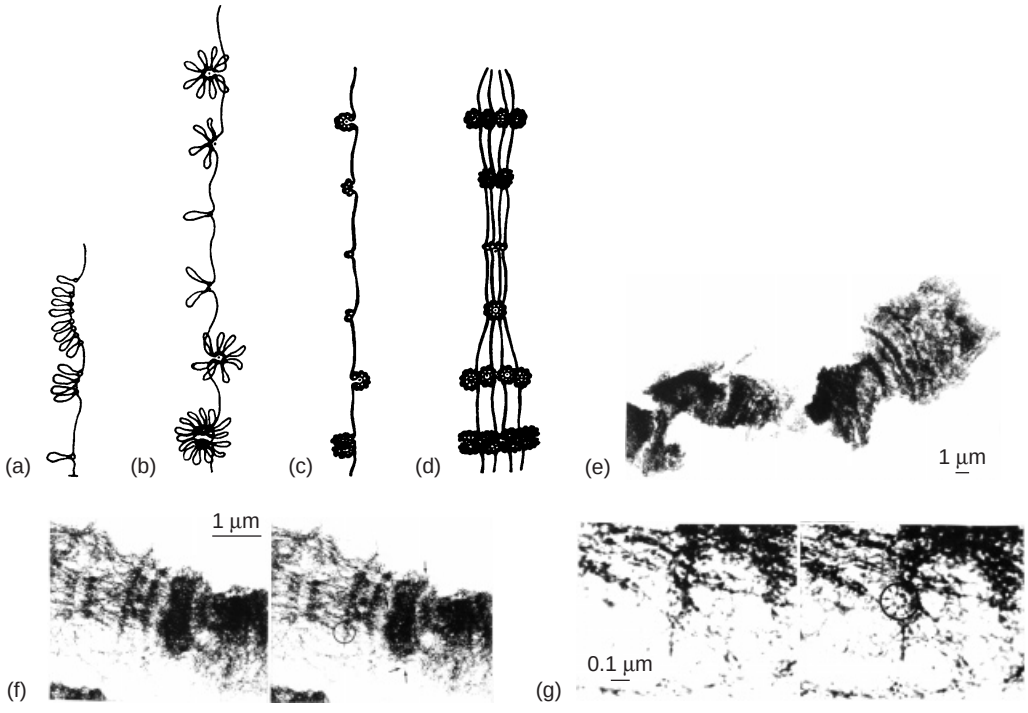

Figure 2 The loop-and-rosette model from interphase to polytene chromosomes. (a) As Figure 1. (b) As Figure 1. (c) As (b), showing that the small DNA loops (1.2–2.2 kb) as seen in polytene chromosomes,[21,22] are bound by 6–11 nucleosomes that form superbeads (30–60 nm in diameter, cf. metaphase chromosome, Figure 7), because of the short loop lengths associated with the scaffold elements. (d) Polytenisation: lateral apposition of chromatids (only four chromatids are drawn for clarity to show the principle), demonstrating the formation of the bands of different thickness. A constriction is interpreted as tendency for primary chromomeres (single loops) to fuse from separate chromatids, rather than as point of underreplication as has been formerly suggested (references in Engelhardt[22]). (e) General view of a whole mount of a CPD polytene chromosome (prepared as the metaphase chromosome in Figure 1e). (f) Stereo pairs of details at higher magnification. Ring-like structures (thin arrows) can be seen. In some rings (encircled) granular subunits are detected. (g) Ring-like alignments of scaffold-forming subunits (circles containing thin arrows) can be visualized in the bands, and they are similar to those found in metaphase chromosomes (cf. Figure 1e–h). This indicates that polytene chromosome bands are composed of the same scaffolding ring-like structures and subunits as metaphase chromosomes and this is best shown when the gross structure of the bands is also well preserved. (Reproduced from Engelhardt.[22])

structures could be traced to derivatives originating from the NE. More specifically, the structures appeared to be virtually identical with an apparently detachable nuclear part of the bipartite NPC. Striated microfilaments lining the lateral elements of the SC were revealed. Apparently, they connect neighboring cyclomeres and may represent the NE lamina. Certain lamina proteins have indeed recently been shown by immunoelectron microscopy (IEM) after DNase digestion to be present throughout the nuclear interior.[32]

It was suggested[8] that the DNA is folded around these ring-like cores, which function as the control elements for replication and transcription. The folding of DNA around these foci emerged as the most convenient way of packing DNA in interphase and metaphase chromosomes. The DNA folding (Figure 1) was presented as the loop-and-rosette model.[22] This implied a natural reason for the different sizes of chromomeres as arising naturally from replicon duplication and from the clustering of repeated replicons to rosettes. By the same principle, a model for polytene chromosomes was presented (Figure 2).[21,22] The details and consequences have been extensively discussed[22] in relation to the NE, chromosome scaffold and the nuclear matrix, and as resulting in the evolution of the chromosome structure.

This integrative model, in retrospect, involved details that strikingly resemble present-day concepts about the central control of replication machinery of eukaryotes. This machinery, with its rosette-like control elements, appears to be uniform from yeast to higher eukaryotes,[3,33] and the proposal of eukaryotic chromomeres as replication and transcription factories, also composed of rosette-form configurations, has been presented.[31] The protein composition of the replication foci involves different dynamically associated proteins, now identified.[33] A single (170-kDa) structural protein, called foci-forming activity (FFA-1), assembles as a ring-like core of the foci that generate the loop replicons. It is a novel protein according to limited sequence data.[33] It

remains to be seen if these proteins have any relationship with our early suggestions.[8]

1.2 Electron Tomography Compared with Earlier Methods

The theoretical resolution of TEM at 100 kV is 0.003 nm, which is sufficient for visualizing atoms (the van der Waals radius of the hydrogen atom being 0.12 nm and that of carbon 0.2 nm). In practice, owing to instrumental constraints and other restrictions that are difficult to overcome,[34] the resolution for modern TEM is at best 0.1 nm, which is still of the magnitude of atomic resolution accomplished, e.g. in TEM crystallography.[35] However, in biological samples the practical resolving power is often not better than about 1–2 nm. At least partly, this is in practice because details in images are in variable degrees superimposed owing to specimen thickness.

Inspection of stereo pairs greatly improves the actual resolution of structures under study, because it eliminates the superimposition of structures. This is accomplished with a mirror stereoscope (e.g. Old Delft, with auxiliary 4.5 × magnification) viewing photographic negatives in transmitted light. This technique has been found most useful and advantageous for interpreting bulky whole-mounted chromosome preparations after, e.g. digestion and treatment procedures.[22]

However, the modern 3-D approach is the ETM, i.e. TEM tomography,[36] to replace the static 3-D viewing of stereo pairs. TEM produces images that are orthogonal projections of the 3-D object under study. Tilt series of image projections of an object from 0 to $\pm 60^{\circ}$, with e.g. a 3° increment, are recovered with a eucentric goniometer specimen holder that is usually incorporated in a modern TEM instrument. The set of projections in tomography are named Radon transforms after Radon (1917), who first presented the mathematical (back-projection) method of rebuilding the original 3-D object from its projections.[36] It was only in 1968 that such an approach was put forward for 3-D reconstruction in TEM studies of biological samples,[37–39] authored in three independent papers.[40] At present ETM allows reproducible 3-D reconstructions with a resolution of 2.5–8.5 nm.[41]

Only in recent decades has ETM become a useful tool in structural biology[36,42] because of the great increase in computer power available. Chromosome studies have played a pivotal role in the development of ETM as described in a detailed review by Woodcock,[4] who also mentions that the MEM in ETM was first used in chromatin studies.[43] One of the first successful tomographic reconstructions of a whole-mounted chromosome (spread on water surface),[44] attempting to track[45] the folding of the chromatin fiber, was performed by a filtered back-projection method with an 'exact' filter algorithm.

The reasons why ETM has not yet become a widely used tool in structural biology were discussed at a recent meeting.[46] Earlier methods have been time-consuming in the extreme. It could take weeks or months to collect the amount of data that the present digital technology provides in days or hours. In addition, as the research community engaged in developing the methods has been rather small, the number of high-impact papers showing striking and convincing results is still relatively small. All this may soon change as programs, high-performance computing and instruments become available.

With the lack of efficient methods, biology has been working in a relatively flat world in trying to understand structures and functions at the molecular level. Yet these things occur in a dynamic 3-D environment. It is with ETM that we are beginning to have a closer look at this 3-D cell environment. In particular, there is at last a solid prospect of comprehending, with ETM combined with immunological,[47] in situ hybridization and with other methods, the higher order structure of chromosomes, the intricacies of which – as presented by the mingled 3-D hierarchy, the manifold of scaffold proteins,[10,11,19] and the DNA folding enigmatically connected with all this – have so far been next to impenetrable.

2 PREPARATION STEPS FOR CHROMOSOMES

2.1 General

As previously discussed,[4] there is in fact no consensus on how to prepare chromosomes for ETM investigation. In addition, ordinary EM fixation methods for, e.g. thin section studies are insufficient for revealing details in nuclear and chromosome configuration, as recognized earlier.[8] To show details of nuclei in sectioned material, staining and preservation with, e.g. various RR methods, as mentioned above, have proved to be more revealing.[8] Viruses with known surface decoration are useful as test objects for developing fixation and preservation methods for ETM.[28] Viruses can also be used as internal standards for evaluation of the 3-D preservation and resolution.

Most ordinary preparation methods used in TEM investigations are usually also convenient for ETM: (1) sections of epoxy resin embeddings, (2) whole mounts and (3) cryo-EM samples. All have their advantages and disadvantages. However, ETM has revealed shortcomings in some common preparative methods such as negative staining of, e.g. some viruses that appear very flattened (about 50%) in 3-D reconstructions.[28]

Reconstruction from sections of epoxy resin embeddings that are commonly used in EM may not be the best method. Sections, however, may often be the only choice to select an embedded specific object for reconstruction, e.g. chromosomes from certain tissue samples, cell aggregates or cultures. According to our experience, reconstructions obtained from sections are not usually as successful as those from whole mounts. The reason might be that sections suffer from various kinds of deformation that are not easily controlled, e.g. in cutting sections with the ultramicrotome and picking them on the grids, in addition to the observed thinning, up to 50%, of sections in EM, in the collection of tilt series.[4] Also, the signal-to-noise ratio is degraded in embeddings. Noise is a general problem also in cryo-TEM when unstained specimens are embedded in vitreous ice. However, the low signal-to-noise ratio may significantly be improved by removing the epoxy resin (or other embedding medium) from sections, thus making a kind of whole-mount preparation of these sections (section 3.5).

Whole mounts of, e.g. chromosomes are preferable for achieving a high signal-to-noise ratio and to secure that the whole content of the object under investigation is carried along. This is not always the situation for ultrathin sections. Sections of large organelles such as eukaryotic chromosomes and nuclei can, however, be of advantage when specimens such as whole mounts are too thick for ordinary 100–120-kV TEM; in earlier studies it was considered that 1 MV was the only choice for chromosomes,[13] for instance. Today, with modifications of preparative methods, 100–120-kV EM has proved sufficient for whole mounts, not only for whole-mounted chromosomes, but also various parts of whole-mounted cells.[47]

2.2 Chromosome Isolation and Whole Mounts

It is sufficient to present the main principles used in chromosome isolation methods because details of the various procedures developed over the years have been published and thoroughly reviewed earlier.[48,49] This equally applies to chromosome scaffold isolation,[19] where new chromosome isolation methods have been introduced.

If a high yield of chromosomes is required, chromosomes are isolated in bulk from synchronized cell cultures[49] of established cell lines that are used in cytological laboratories, e.g. HeLa (human) and CHO cells.

The procedure is as follows. Cell cultures in logarithmic growth phase, when the confluence of cells is 50–60%, are synchronized with the addition of (2 mM) thymidine in the cultivation medium and incubated for 16–22 h (depending on the cell cycle for cell lines used). The growth medium, with thymidine, is replaced with a fresh medium including a mitotic inhibitor, e.g. (0.06–0.1 $\mu g\,mL^{-1}$) Colcemid (nocodazole or vinblastine for certain cell lines), with more (10–20%) than ordinary (5%) calf serum, or fetal calf serum, and incubated for 8–12 h. Cells in mitosis are released, naturally or aided by gentle shaking, into the growth medium, because in the cell lines to be used (e.g. HeLa, CHO) mitotic cells are rounded up and detach from the bottle surface. Mitotic cells are collected from the medium by centrifugation.

It is generally essential for a successful isolation of metaphase chromosomes with most of the chromosome isolation methods used that mitotic cells are swollen by hypotony by dispersing the pellet of mitotic cells in a hypotonic solution, e.g. 75 mM KCl, for about 10 min, at room temperature (RT) or at 37 °C. Some cells need first to be washed with the hypotony solution to make the hypotony treatment effective. The swollen cells are then collected gently by centrifugation, e.g. at 800–1000 rpm for 5–10 min. The total treatment time, including the centrifugation, must not exceed 15 min. This is important when divalent cations (Mg^{2+} and Ca^{2+} or polyamines), that are essential for chromosome stabilization, are not present (but see Mendelsohn,[49] with divalent cations present).

After the hypotony treatment, any of the various chromosome isolation procedures,[19,48,49] 'physiological' or less physiological, are convenient.

The only general requirement for isolating chromosomes and nuclei and keeping them stable, with e.g. 5–10 mM (or lower) *N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid) (HEPES) buffer at pH 7.4, appears to be the presence of divalent cations (1–5 mM Ca^{2+} and/or Mg^{2+}) or polycations such as spermine (0.0125 mM) and spermidine (0.05 mM).[27] Monovalent salts only tend to cause aggregation.

Mitotic cells are also suited for the microspreading methods of mitotic[50] or meiotic chromosomes,[51] either without or with hypotony treatment.

The procedure is briefly as follows. In this method the cells are spread on water, or on hypotonic salt solution for more gentle spreading (0.05–0.1 M NaCl), to control better the spreading forces at the hypophase that break the cells and release the chromosomes. The chromosomes are picked from the hypophase surface on poly-L-lysine (PL) (below) or glow-discharged[51] carbon–Formvar-coated grids, after dehydration in methanol (MeOH), lightly stained with (0.001–0.002%) uranyl acetate (UA) in MeOH (30 s), washed in MeOH and dried by CPD (section 3.1).

With this elegant method, the chromosome preparations are ready for EM, without laborious purification and isolation procedures. There is the disadvantage that cytoplasmic obscuring components may also attach on the grid preparations. It is for the investigator to choose

the chromosomes on the grids, which does not seem to be too difficult a procedure as we can see in, e.g. DuPraw's splendid pictures of spread chromosomes.[5]

From the viewpoint of early chromosome morphology, these isolation methods yield the unexpected result that chromosome coiling is indistinct after any of the 'physiological' isolation or spreading methods, especially at neutral pH. The reason for this is not clearly understood.[13] Chromosomes squashed in 50% AA[13] after standard cytological fixation in MeOH–AA (3:1) do show distinct coiling.

Chromosomes isolated in bulk in 60% AA, after hypotony treatment, surprisingly show the same chromosome coiling[22] when the loose pellet of swollen cells is washed and dispersed in 60% AA. This method works also for cells that have been fixed, and stored at −20 °C, in MeOH–AA (3:1). The pellet of chromosomes, after centrifugation (1500–5000 rpm, 10–15 min), may appear indistinct depending on the initial amount of cells used. In any case, the pellet is washed in 60% AA, two or more times, and collected by centrifugation. After washing in 60% AA the chromosomes are diluted in a small amount of 60% AA and stored at +4 or −20 °C. After this procedure, a suspension of pure chromosomes, with 10–30% prophase and interphase nuclei, is found, with insignificant amount of cytoplasmic debris, when investigated by phase-contrast microscopy. Distinct chromosome coiling is visible with phase-contrast microscopy. The chromosomes do not aggregate when pelleted by centrifugation and the pellet is easily redispersed.

Polytene chromosomes can also be isolated by this method but to release them from nuclei they have to be spread on, e.g. 60% AA.[22] The chromosomes are then picked on polystyrene-coated grids (below), dehydrated, lightly stained with UA (above) and CPD (section 3.1) or *tert*-butanol dried (section 3.2).

In the AA-isolated metaphase chromosomes, the clearly preserved coiling matches classical chromosome descriptions. Thus, the AA-isolated chromosomes seem to be tempting objects to use to study the higher order chromosome structure related to chromosome coiling.

The metaphase chromosomes are mounted on EM grids by centrifugation (1–5 min, 500–1500 *g*) of a sample (5–10 µl of diluted chromosome solution, or 2–3 µl of the loose pellet) through a cushion of 60% AA, in microchambers (Figure 3) for grids (50–200 µl).

The centrifugation for mounting chromosomes on the grid is preferred as chromosomes are better preserved in their 3-D configuration than with other more direct application methods.

With AA, Formvar or carbon–Formvar-coated grids are not suitable, as they are not resistant to AA. However, grids coated with plastic (polystyrene) films (Petri dish pieces, e.g. Nunc, Nalge Nunc Inc., dissolved to make a 0.5% solution in 1,2-dichloroethane), without or with additional carbon coating, are resistant to AA.[22] These films are stronger than Formvar films but less adhesive. To prevent the film peeling off the grids, the latter may first be treated with an adhesive solution (0.01% polystyrene in benzene).[51]

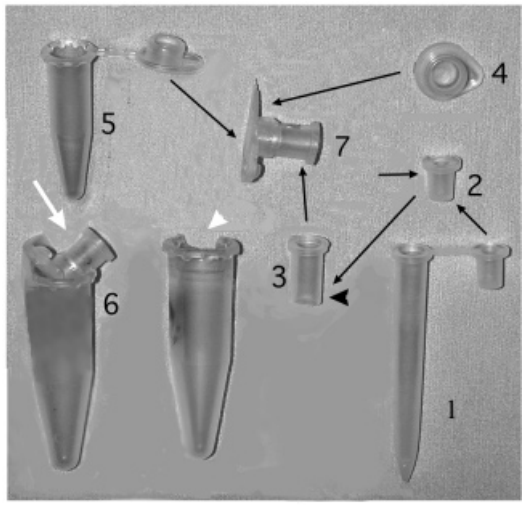

Figure 3 Grid microchambers are made of plastic hematocrit tubes (1). The plug (2) of the tube is cut off (arrow) so that the plug can be easily forced to the required depth (arrowhead), into the cut-off top part (3) of the hematocrit tube to function as the bottom for the grid. Chambers of suitable sizes are thus easily obtained, e.g. 50–200 µL. The caps of a 0.5 mL microcentrifuge (Eppendorf) tubes (4) are used as adapters for (3) fitted to 0.5 mL micro-centrifuge (Eppendorf) tubes (5) (which fit into to a larger tube for a swing-out desktop centrifuge, not shown). When a swing-out desktop centrifuge is not available, the 0.5 mL cap is positioned in a 45° angle (white arrow) in a 1.5 mL microcentrifuge (Eppendorf) tube (6) with a u-form cut (white arrowhead) for the grid microchambers (7), as a substitute in a swing-out position to have the sample evenly distributed over the grid. If high centrifugation speeds (over 1500 rpm) are used, the caps at 45° angle must be fixed (glued or melted) to be kept in place. These chambers are analogous to Miller's grid chambers,[53] but faster and easier to make.

Before mounting chromosomes on the grids, fiducial gold markers that are needed for the alignment of the images of the tilt series for ETM are applied on the grids. Ni or Au grids are preferable as in the presence of gold Cu will wither when stored. Colloidal gold markers are applied by dipping the grid, one or both sides, in a drop of, e.g. 5-, 10-, 15- or 40-nm colloidal gold solution diluted to, e.g. 1:10, with filtered double-distilled water ($FddH_2O$) (0.22 µm filter, Millipore) or with PL, e.g. 0.01–0.1 mg mL^{-1} in $FddH_2O$. Being uniformly more round and less liable to aggregate, commercial immunogold particles (e.g. expired) are more advisable than self-made particles. The grid, with adhering gold suspension, is immersed in 100% MeOH that precipitates the gold on the grid. This procedure is repeated, if

necessary, to obtain the amount of gold needed. The minimum number of gold markers is three [or four[52]], preferably 8–12 or more, for the image area and the magnification used. The same gold particles must be identified on each image in a tilt series (from 0 to ±60°, e.g. in 3° increments). The size of gold particles chosen depends on the magnification that will be used for the reconstruction. In addition, a mixture of different-sized gold particles is useful for different magnifications. With PL, the dilution of the gold suspension may be increased to a much higher degree. However, we observed the drawback that when embedded in PL the gold markers look less sharp and are difficult to distinguish, especially at high tilts.

2.3 Pretreatments: Salt Extraction and Enzyme Digestions

When the chromosomes are mounted on grids, they are ready for further procedures for EM. If salt treatments or enzyme digestions are required, they must be completed at this stage, before postfixation and staining are performed.

Note that the methods that follow are suited also for other than AA-employing chromosome isolation methods[48,49] and for microspreading methods.[50,51] The point is that the chromosomes are attached on grids for easy transfer to different solutions used.

The attachment of isolated chromosomes with centrifugation, in the isolation buffer used, is secured by treating the carbon- and film-coated grids with, e.g. PL (0.1–1 mg mL^{-1} in FddH_2O), for about 1–5 min, followed by washing in drops of FddH_2O. Also, glow discharging makes the carbon or plastic film coating hydrophilic and contributes to attaching chromosomes on the grid surface.[51] The same effect is obtained by UV radiation (10–30 min) or dipping in alcohol followed by air drying (AD).

These procedures are not necessary with AA-isolated chromosomes, as they do stick on hydrophobic polystyrene film-coated grids with carbon or not. Also, PL does not harm the attachment of AA isolated chromosomes.

To remove the grid easily from the microchambers after centrifugation, follow a simple procedure.[53] After centrifugation, add 60% AA to form a convex surface in the micro-chamber. Retrieve with curved anticapillary (and antimagnetic, for Ni grids) forceps by flipping the chamber upside down to let the grid float on the surface of the hanging drop. If the grid does not detach from the bottom by slightly rocking the microchamber, it can be released by touching with a thin needle or one arm of the thin forceps on the periphery of the grid.

Grid washings, etc. are accomplished in, e.g. a large Petri dish, with moistened filter-paper, overlaid with Parafilm. The Parafilm is preferably provided with drop-sized depressions (Figure 4).

The grids are washed in HM (10 mM HEPES, pH 7.4, 5 mM $MgCl_2$) or HC (10 mM HEPES, pH 7.4, 5 mM $CaCl_2$) buffer on the Parafilm, with several droplets (1–5 min in each) to remove the AA and to balance the pH to neutral. AA-isolated chromosomes retain their morphology when transferred to water solutions containing divalent cations at neutral pH. (For a routine, it is recommended to use double-distilled water, and filter (0.22 μm, e.g. Millipore) all solutions. Also use rubber gloves, wash if powdered.)

To isolate the chromosome scaffold, histones are extracted with, e.g. 2 M NaCl solution and digested with DNase and RNase.[21,22]

However, chromosomes isolated with 60% AA, according to our sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) analysis and immunoblotting, show no or negligible amounts of histones,[54] the bulk of which appears in the primary 60% AA extract.

Thus extraction with 2 M NaCl-solution apparently is not necessary to extract histones. In any case, digestion with DNase I (10 μg mL^{-1}) (Boehringer), with or without RNase A (10 μg) (Boehringer), completed in 2 M NaCl in HM buffer, shows better results when monitored with DNA-specific fluorochromes, e.g. DAPI (4′,6-diamidino-2-phenylindole) (Sigma), 0.25–0.5 μg mL^{-1}, in the digestion solution.[22,54]

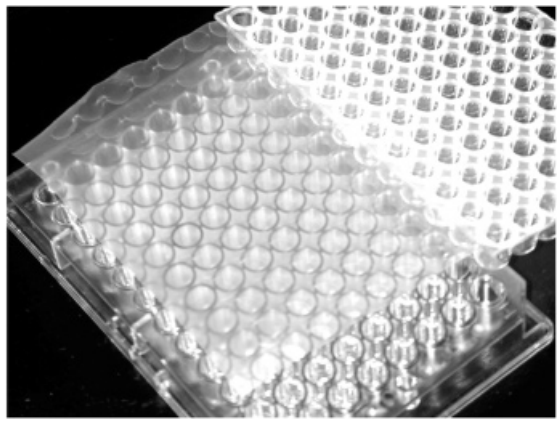

Figure 4 Depressions on Parafilm are obtained as follows. Take two incubation chambers with 96 wells, one with round-bottomed wells, frames removed. After inserting Parafilm (with its cover paper, not shown) between the two chambers, press together. The 'dimples' formed will ensure that the incubation drops (about 50 μL) do not float around during washing steps, when positioned in a large petri dish with moistened filter paper (to prevent the drops from drying during incubation).

After DNase I digestion in 2 M salt–HM buffer for 1–2 h at RT, no DNA-specific fluorescence can be detected after washings in 2 M salt–HM buffer followed by washings in HM buffer, and mounted in DAPI mounting medium (MM) (90% glycerine, 0.25–0.5 $\mu g\,mL^{-1}$ DAPI, in HM or HC, with antifading substances such as 10–20 mM dithiothreitol or 2-mercaptoethanol). Many commercial antifade/bleaching embedding products are available (note that the presence of divalent cations is necessary for the stability of isolated chromosomes).

Any other enzyme digestions desired are performed similarly [cf. hyaluronidase[21,22]], sequentially or if possible simultaneously, in the buffer solutions required.

Control preparations are made in the same salt buffers but without the enzyme, denatured enzymes or with enzyme inhibitors.

For enzyme digestion of resinless thick sections of whole-mounted cells, see section 3.5.

2.4 Fixation of Chromosomes

Chromosome preparations, with or without various pretreatments, preferably are chemically postfixed with glutaraldehyde (GA), a cross-linking substance and a reliable preserver of cellular ultrastructural details. Our purpose is to maintain the chromosome coiling for 3-D reconstruction. AA is not a cross-linking fixative but preserves the higher order chromosome structure, e.g. chromosome coiling, by acting as a precipitator of chromosomal nonhistone proteins. As mentioned, there is no consensus on how to preserve chromosomes for ETM,[4] so the field is open to the development of new fixation methods for testing and refining. Cautious fixation and staining are important especially after digestion–extraction steps when the residual delicate material left might easily be lost in washing steps or the configuration might change during dehydration steps before, e.g. CPD (section 3).

We have found it advantageous to use a set of various fixation procedures as a standard protocol to test different constituents and their combinations on the 3-D structural preservation and appearance. This concerns not only methods for structural chromosome studies but also any other material such as whole-mounted cells, e.g. in developing IET[47] methods (section 7).

Fixing and preserving whole mounts is a delicate procedure, as too much contrast, i.e. electron density, must be avoided to keep specimens transparent, when conventional 100–120-kV TEM is used for collecting the tilt series. The methods used in standard thin-sectioning studies, where high contrast is rarely a problem, will not work well if not modified. Little attention has been paid to developing preparative methods for whole mounts for TEM. The quality of ETM 3-D reconstructions depends as much on the development of programming methods and the performance of TEM as on the development of specimen preparative techniques suited for ETM, just as early EM depended on the instrument but equally on developing specimen preparative techniques.

CPD and other 3-D preserving methods (section 3) are used to avoid drying artifacts after the fixation procedures. It has been reported,[55] however, in scanning electron microscopy studies, that GA shrinks cells to some 45%, despite CPD. The shrinking by GA was prevented by using tannic acid (TA), after osmium tetraoxide (OT) fixation.

We tested the TA procedure using viruses as test objects. We had previously found that viruses fixed with GA appeared slightly smaller than recorded in the literature, after drying from solidified *tert*-butanol (analogous to CPD, section 3.2) for 3-D reconstructions by ETM.[28] With TA included in the fixation procedures, the size of the viruses appeared more in accordance with the reported sizes of the viruses (unpublished work).

Upon testing the TA method on chromosomes, the chromosome fibers appeared significantly thicker (40–60 nm) than chromatin fibers without TA (35–40 nm). In addition, the chromosome fibers appeared to be composed of supranucleosome-like particles of various sizes rather than a more regular fiber that the solenoid model predicts (cf. Figures 5 and 6 with Figure 7). In addition, the organization of the chromosome scaffold fibers changed appearance with TA.[54]

In our present standard fixation protocol, TA is included with GA, OT and UA with some modification of the original method.[55] Fixation controls, to test the effect of the different substances and their combinations, are used.

Dehydration and staining with UA (0.001–0.002%) in MeOH are performed in microcentrifuge tubes (Eppendorf) (1.5-mL test tubes). Curved anticapillary (and antimagnetic, for Ni grids) forceps are advantageous in the transferring steps.

The procedures are as follows. The fixation and preserving methods, for AA-isolated chromosomes and chromosome scaffolds, are accomplished with chromosomes mounted on grids:

1. (a) No postfixation;
   (b) boiling in 50% AA, for 30–60 s,[21]
   (c) wash, 3–5×, in HM or HC buffer, if the specimens are brought to salt and enzyme treatments, section 2.3;
   (d) dehydration (30, 50, 75, 100%, 10–30 s per step) in MeOH;

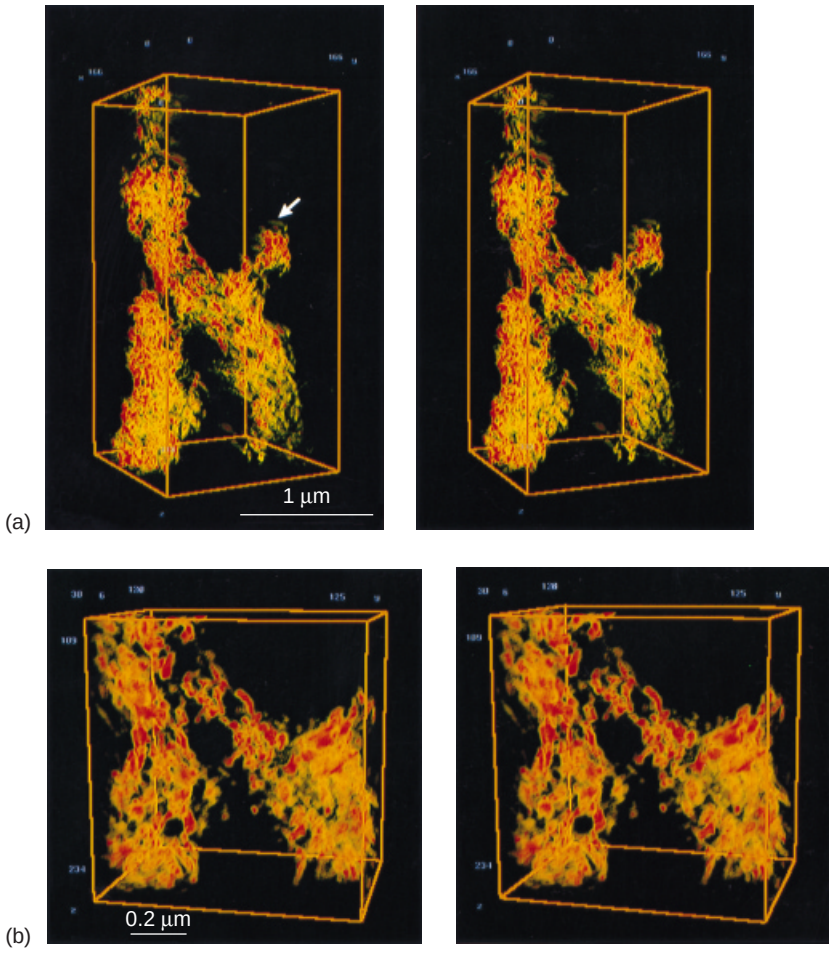

Figure 5 (a) Example of a low-resolution 3-D reconstruction with weighted back-projection method (WBM) of AA-isolated (HeLa) metaphase chromosome (prepared as in section 2.4, Procedure 1(b), and dried from *tert*-butanol, section 3.2), visualized with BOB (section 6.1; Figure 1f). The WBM has been low-pass filtered to 25 nm. Chromosome fibers of 25–60 nm are seen. Streaks seen in depth direction are artifacts that appear with WBM due to the missing data and other discrepancies (e.g. alignment inexactness, when tilting parameters are not accurately determined, especially with the angle of tilt axis being manually approximated, produces whirl-like streaks, arrow, in the reconstruction). [Reproduced by permission of the Center of Scientific Computing from Engelhardt et al.[28] and also from Engelhardt and Ruokolainen.[54]] This case of ETM, scanning, alignment and WBM was processed with the programs developed in Dr U. Skoglund's laboratory, Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden, in 1993. (b) Details, in stereo pair of (a), at higher magnification. Chromosome fibers (20–60 nm) appear smoothed (low-pass filtered) with no sub-details detectable.

(e) staining in (0.001–0.002%) UA in MeOH for 30 s, wash in MeOH (3×);

(f) drying, with 3-D preservation methods, e.g. CPD (section 3);

Comment: boiling in 50% AA[22] causes additional purification of chromosomes from histones and cytoplasmic fragments.

2. Rapid TA method for chromosomes, modified from:[55]

- washing in HM buffer (3–5×);
- fixing, for 2–5 min, in mixture (1 : 1) of:
  1% GA in HM buffer, pH 7.4;
  0.5% OT in 0.1 M sodium cacodylate (NC) buffer, pH 7.4;
- washing in HM buffer (3–5×, 0.5–5 min);
- 0.5% TA in 0.1 M NC/HM buffer (1 : 1), 5 min;
- washing in HM buffer (3–5×, 0.5–5 min);

(a) staining with (0.001–0.01%) UA in FddH_2O, 5–10 min;

(b) without UA in FddH_2O:

- dehydration, staining in UA in MeOH, and drying as in (a).

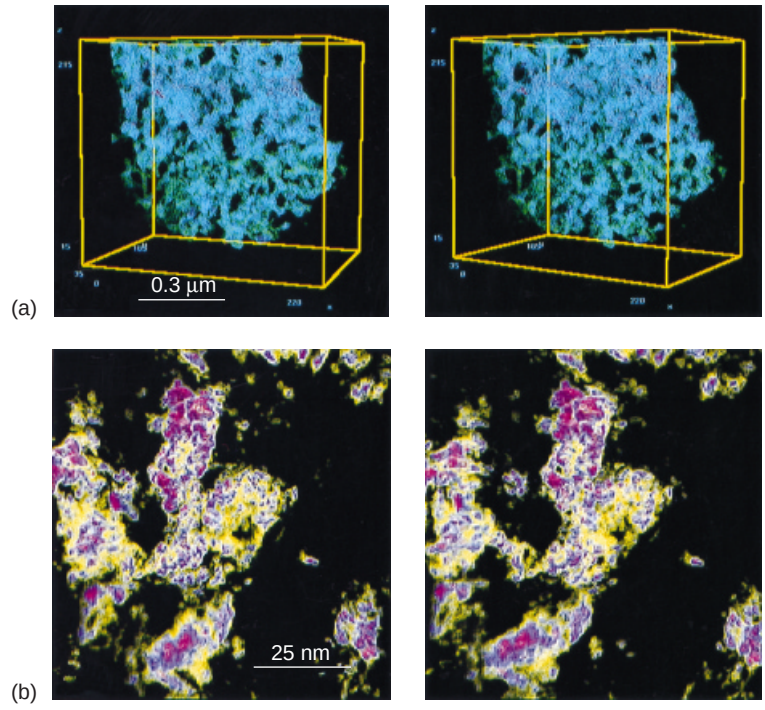

Figure 6 (a) Chromosome (HeLa), prepared as in Figure 5, showing 3-D reconstruction with MEM (section 5.4) of a chromosome telomere, in stereo (see above). (The original magnification of the tilt series was 48 000×, scanning pixel size 20 µm, giving the original resolution as 0.42 nm. With 6 × binning, the voxel size is 2.5 nm for the volume, i.e. the 'conservative estimate'[52] of resolution is not better than about (2–4) × 2.5 nm = 5–10 nm.) Some details (cf. Figure 5a) of chromosome fibers may be detected even at this low magnification. (b) At high resolution, the chromosome fibers appear decorated with details, but are difficult to interpret. Histones are extracted by AA but DNA is present (cf. immunoblotting analysis of histones and DNA fluorescence with DAPI in Figure 8d). [Reproduced from Engelhardt and Ruokolainen.[54]]

Controls and modifications:

3. as in 2, but without TA;
4. as in 2, but without OT;
5. as in 4, but without TA;
6. as in 3, but without UA.

3 PRESERVATION METHODS FOR THREE-DIMENSIONAL RECONSTRUCTIONS

3.1 Critical-point Drying

After fixation, the whole-mounted preparations have to be dried. AD causes artifacts, shrinkage, wrinkling and distortion that destroy the 3-D configuration. Drying artifacts are caused by disruptive effects of the surface tension of the fluid from which they are dried, e.g. water but also substances such as alcohol, used for dehydration, have a surface tension high enough to destroy delicate 3-D configurations.

The problem was recognized and solved in the early development (1950) of TEM techniques by T.F. Anderson, who introduced the CPD method.[50] CPD is conventionally used for whole-mount studies with scanning electron microscopy. CPD needs, however, an apparatus where the procedures are performed at high pressure.

CPD can be briefly described as follows.[50] After the grid samples have been gradually dehydrated in steps, e.g. in MeOH, the grids are put in a grid holder in the CPD apparatus, with some dehydration fluid, and the MeOH is completely replaced with liquid carbon dioxide (CO_2), at a high pressure (60–65 atm) in several (7–10) washing steps with liquid CO_2. CPD is completed, at an elevated pressure of about 100 atm ($P_c = 73.8$ bar), by increasing the temperature to ca. 40 °C. Keeping the temperature above the critical point temperature ($T_c = 31$ °C), the pressure is slowly decreased, regulated by a relief valve. After CPD, the specimens are ready for TEM.

However, CPD has some disadvantages such as causing thermal and pressure stresses in the specimens that may be difficult to avoid. An apparatus is needed for the procedure that may not always be available. The procedure needs liquid CO_2, preferably dried to remove

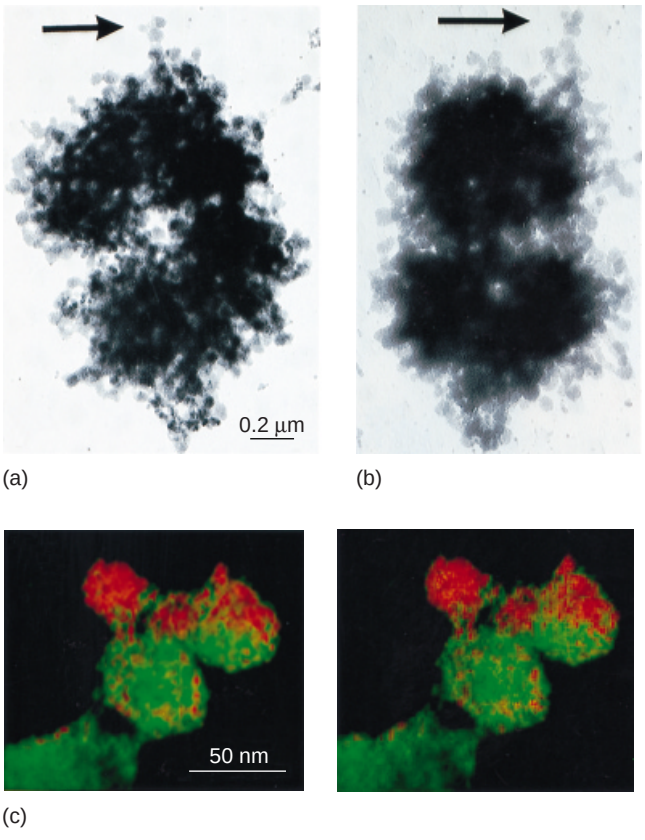

Figure 7 Chromosome (AA-isolated from HeLa cells), pretreated with TA before CPD to avoid shrinkage (section 2.4, procedure 2): (a) original tilt series at 0° tilt [reproduced from Engelhardt and Ruokolainen[54]] and (b) +60° tilt with mean diameter of the 30 nm chromatin fibers (arrow) increased to 30–60 nm with globular domains 30–60 nm in diameter. (c) Stereo pair of 3-D reconstruction with MEM (section 5.4) of the chromosome fiber [shown in (a) and (b) with an arrow]. The 30–60 globular domains of the chromosome fibers appear decorated and seem to be connected by two (2–4-nm) thin strands. [Reproduced from Engelhardt and Ruokolainen.[54]]

minute amounts of water that may be present.[56] The water present in liquid CO_2 was shown[56] to be the reason for drying artifacts such as thickening of cytoskeleton fibers in the microtrabecular lattice.[57] In addition, the CPD procedure takes time to finish.

3.2 Drying from *tert*-Butanol

When an apparatus for CPD is not available, a substitute for CPD is drying of the grid samples infused with *tert*-butanol. This method needs no expensive accessory, although such an apparatus has been constructed.[58] The procedure is much faster than CPD and is claimed to give analogous results. The method is easy to accomplish, with only a few restrictions, as ordinary grids coated with Formvar are not resistant to *tert*-butanol. However, plastic films (section 2.2) that are resistant to 60% AA used in chromosome isolation with an AA procedure are suitable for *tert*-butanol drying. In addition, grids with a pure carbon film coating are usable.

After the dehydration steps with, e.g. MeOH (and staining with UA in MeOH) of the grid samples, microcentrifuge tubes (Eppendorf) (1.5 mL) are used: one tube with a mixture of MeOH–*tert*-butanol (1 : 1), and then a series of 3–4 tubes with 100% *tert*-butanol (*tert*-butanol is liquid at temperatures above 25.5 °C). The grids are washed in each step for 10–60 s. After the last step each grid is dropped to the bottom in a separate microcentrifuge tube (Eppendorf) (1.5 mL) with about 0.5 mL of *tert*-butanol until all grids are ready. After infiltration with *tert*-butanol, the fluid is changed several times, if considered necessary, by gently pouring off the liquid, keeping the grid on the bottom. For the drying procedure a small amount of *tert*-butanol is left, letting the grid attach by capillary force to the wall of the tube, then sucking the fluid droplets left on the bottom with a pipet before closing the tube. The preparations are left for a short time in a cold place (below 25.5 °C), or briefly on ice, to let the *tert*-butanol solidify. The sample-containing tubes, with lid open, are brought to a vacuum drier (e.g. Savant Speed Vac Concentrator, with centrifugation off) to let the *tert*-butanol dry off. The grids are usually seen to drop to the bottom of the tubes when the drying is completed. The operation will not take longer than 10–30 min, depending on the amount of *tert*-butanol and the degree of vacuum used. The grids must be at RT when the preparations are lifted out from the vacuum dryer, to avoid water condensing on the specimens.

3.3 Freeze-drying

Freeze-drying (FD), like CPD, needs expensive equipment. The advantage of FD over CPD and drying from solidified *tert*-butanol is that the preparations are not dehydrated with organic substances that might be destructive, extracting membranes and denaturing proteins. Partial FD is used in the freeze-etching method, where cell preparations are coated with Pt–carbon to obtain replicas of, e.g. the cell cytoskeleton, showing sharp views of actin bundles and tubulin fibers in stereo.[59] However, replicas are not useful for ETM 3-D reconstructions as the entire preparation must be transparent and not only a surface view. When FD is adapted for ETM, the time must be extended (1–3 h) so that the drying of the specimen is completed. The specimens must be prewashed with $FddH_2O$ to remove salts that will otherwise crystallize. If samples cannot maintain their configuration without salt, a volatile salt, e.g. ammonium acetate, is used at a concentration that keeps the specimen stable. Negative staining methods, e.g. of viruses, combined with FD

look superior[60] for 3-D reconstructions, as the specimen configuration is maintained and not flattened as with AD.

In brief, the procedure to freeze the specimen for FD is analogous to cryo-TEM methods. The specimen is frozen in liquid nitrogen, with a container with cryogen (e.g. liquid ethane or propane), as an intermediate fluid that is functioning as an effective conductor, for rapid freezing of the water-embedded specimen into vitrified ice. The FD is performed by raising the temperature (to −80 °C). Drying out of the preparation at high vacuum will then be possible within a reasonable time,[60] after which the preparations are ready for ETM.

3.4 Cryo-electron Tomography

Automation programs for collecting tilt series in ETM have made it possible, for 3-D reconstructions, to use single-axis cryo-ETM to collect tilt series of frozen vitrified samples, i.e. fully hydrated specimens, which are chemically unfixed, i.e. close to native state.[40] However, with this almost ideal method for specimen preservation a computer-controlled TEM and goniometer stage with cryo-holders, cooled with liquid nitrogen and accessories for preparing vitrified cryo-samples, are needed. A disadvantage is that the signal-to-noise ratio of the images usually tends to be poor, especially of unfixed and unstained specimens embedded in the vitrified ice. The noise in images is also difficult to abolish for specimens such as chromosomes as images collected in tilt series with axial tomography cannot be improved by any averaging image processing, as is possible with single particle or icosahedral virus tomography.[36,42]

The procedure is as follows. Use isolated chromosomes mounted (section 2.2) on grids, coated with perforated carbon, or better with perforated plastic film,[61] on, e.g. 50–75-mesh Ni or Au grids with colloidal (e.g. 5, 10, or 15 nm) gold markers for collecting tilt series. The idea with perforated-film coatings is to collect images of chromosomes that are preferably spanning a hole, embedded only in vitrified ice without a support, thus reducing noise and improving image contrast. Using 50-, 75-mesh or even single-slot grids, the prospect of viewing the whole tilting range from $0 \pm 60^{\circ}$ is considerably improved. The grid–specimen is washed in HM or HC buffer (≤5 mM HEPES buffer), in several droplets, before the grid, after excess buffer has been blotted off with filter-paper, is quickly dropped in a container, inserted in liquid nitrogen containing liquid ethane or propane, or better, in a mixture of ethane and propane (3 : 1) that prevents solidification of the freeze-transferring cryogens. The grids can be stored in liquid nitrogen until examination. The grid–specimen is installed in a liquid nitrogen-cooled cryo-specimen holder that protects the specimen from atmospheric humidity before cryo-TEM.

With cryo samples, automatic collection of the tilt series is the only way to avoid melting and beam damage of the specimens, as automation will shorten the exposure 10–100-fold compared with manual methods.[40]

3.5 Section Preparations

Our 3-D reconstructions with ETM, accomplished on sections of epoxy-embedded material, are not so successful as with whole mounts. Many reasons may account for this difference. As mentioned, sections have been noticed to become up to 50% thinner,[4,36] in the collection of tilt series data. If thinning takes place gradually during the collection, the reconstruction operations will severely fail, owing to unrestrained distortions of structures. The noise is more marked especially in thick sections than for whole mounts. However, thin sections also become thicker at high tilts (twofold at 60°)[52] and images are blurred.

The advantage with the sectioning technique, however, is that most material can be embedded, sectioned and structures of interest selected that would otherwise be difficult to isolate or are too bulky to study as whole mounts.

Standard fixation methods used in EM are usually inadequate for chromosomes and nuclear structure studies.[8,22] The fixation and staining methods that we have found applicable for section studies of chromosomes and nuclei are based on using RR in different combinations with GA, OT and, in dehydration before embedding in plastic resin, with UA (stain) in MeOH.

The procedure is as follows.

1. Specimens are fixed in 1–2% GA in 0.1 M NC buffer, pH 7.0–7.4, or in Ringer solution (RS), with (a) RR (750–1500 ppm) or (b) without RR, at RT, for 1–2 h or longer, washed in the NC buffer or RS.
2. Further fixed in 1% OT in the 0.1 M NC buffer or RS (2–4 h), RT, (a) with RR (750–1500 ppm) or (b) without RR, washed in the NC buffer or RS.
3. Stained in 0.1 M NC buffer or RS, (a) with RR (750–1500 ppm) or (b) without RR, for 24 h, at +4–8 °C, washed with buffer solution, briefly in water, before gradual dehydration in MeOH, then stained in 1% UA in MeOH [overnight (o/n) to several days, at +4–8 °C], washed in MeOH, transferred to absolute ethanol, propylene oxide, and finally to propylene oxide–Durcupan ACM (1 : 1), o/n.
4. Embedded in Durcupan ACM, epoxy resin (Fluka).

Controls: without RR, and with RR in the different steps and combinations.

The penetration of RR into nuclei and chromosomes can be checked in semithick section for phase-contrast LM.

Sections are cut with an ultramicrotome and mounted on grids according to standard EM thin-sectioning techniques. The sections on the grids are stained with Reynold's lead citrate (10–30 min, in a Petri dish with solid KOH to avoid precipitation), washed with freshly boiled FddH_2O and AD before examination in TEM.

Comments: importantly, to enhance the intracellular contrast with RR, it is to be used after the OT fixation (step 3). This is because OT fixation will destroy the semipermeability of cell membranes and RR will then penetrate and stain nuclei and chromosomes. Before OT fixation the intracellular penetration of RR is almost nonexistent. However, if high contrast is required, after staining with RR (step 3), postfixation with OT is added.

To study, e.g. interphase chromosomes, whole mounts of nuclei are too thick for ordinary 100–120-kV TEM. Thin sections must be used to collect the tilt series. Section thickness is, however, limited to <0.1 μm for 100–120 kV.

If thick sections must be used to incorporate spatial details, the embedding medium can also be removed. The sections are therefore changed to a kind of whole mounts. This improves the signal-to-noise ratio and makes the thick embedded material more permeable to electrons.

The material is fixed only with GA, before embedding and sectioning, and after removing the embedding, brought stepwise to buffer solution and prepared as whole mounts (section 2.4) or then processed as in bulk embedding for sectioning, i.e. fixed, stained, embedded and sectioned using standard fixation with GA or OT, or using the RR methods for chromosomes and nuclei, as presented above.

Gold markers are introduced only on the section-free side of the grid, to ensure that they do not float off when the embedding medium is removed. It may also be advisable not to poststain the sections with, e.g. Reynold's lead citrate, that may form precipitates when the embedding medium is removed. The procedure for dissolving the embedding medium from sections is somewhat different, depending on the different embedding media used.[62,63]

Removal of resin from ordinary epoxy thin sections is carried out as follows:[62]

1. Preparation of the solution: prepare a solution of seven pellets of KOH in 100 mL of absolute ethanol (the color of the solution will be light orange–brown when ready, after 1–2 days).
2. Immerse the sections (0.25–0.5 μm or thicker), mounted on carbon–Formvar grids, in the solution for several minutes (to be time tested, depending on section thickness and epoxy resin used).
3. Wash in ethanol and MeOH, with several changes.
4. CPD (section 3).

If section material seems to be lost, PL-coated grids (section 2.2) are used to ensure better that the material will stick to the grids.

Note that the method seems to be mild enough to be used with immunological methods.[62]

In addition, for enzyme digestions a method has been introduced[32] for immunostaining, e.g. interphase chromosomes. Isolated whole cells are encapsulated in agarose microbeads for the enzyme digestion and immunostaining procedures. Thick, resinless EM sections are obtained that are excellent for ETM purposes (applications in sections 2.3 and 7).

4 DATA COLLECTION

4.1 Recording of Tilt Series

Images recorded in TEM are two-dimensional (2-D) projections of a 3-D object. Owing to the large depth of focusing in TEM, all the details are usually in focus and superimposed in thick preparations, whole mounts and sections. Chromosomes lack symmetry, the kind of which is characteristic of icosahedral viruses. Thus, for 3-D reconstructions with ETM, single-axis recording of tilt projections is accomplished that after the alignment procedure is processed with the WBM and MEM.

A conservative prediction[52] of the resolution (d) obtained with single-axis ETM is correlated with the number (theoretical) of projections (N) available (from 0 to ±90°) and the object thickness (D), and is given[4] by Equation (1):

$$d = \frac{\pi D}{N} \tag{1}$$

The actual resolution realized is, however, difficult to estimate, as the number of projections is limited by the tilting range from 0 to ±60°. This tilting range is obtainable with standard specimen holders and goniometers in TEM. Special specimen stages and holders have been available earlier[64] and recently introduced for cryo-EM, including new rectangular grids.[65] Such grids with a 'grill-like' pattern have long been available, but the new 'cryo-tilting grids' are in a way improved because of the square shape, with rectilinear rims perpendicular to the tilting axis that will free the tilting procedure from shading grid bars. The disadvantage with these new grid holders is that there is no way of rotating the grid specimen. The rotation of the specimen is most important for oblong whole-mounted structures such as

chromosomes to orient them parallel to the tilt axis to record tilt series with optimal resolution in ETM.

4.2 Manual Recording

Although manual recording of tilt series in axial ETM will in the future be replaced by automatic recording (section 4.4), there are probably few laboratories today that have access to or can afford digitally controlled TEM equipped with automation facilities. Manual collection of tilt series is, after some training, a fairly quick procedure when recorded on EM photographic film plates [actually a faster procedure than recorded manually or automatically with a slow-scan charge-coupled device (CCD) camera]. The EM plates also have the advantage that the recorded area is very much larger than that covered by a CCD camera, and owing to this advantage for film plates, more and larger structures can be collected at higher magnification. This may be important not only for the structures under investigation but also for collecting gold markers for the alignment procedure, as there may not always be enough markers available in the area that the CCD chip covers. Manually, increments of less than 3° are difficult to accomplish in practice. We have found in any case that tilt series, from 0 to ±60° with a 3° increment, are sufficient for ETM reconstruction of whole mounts, and have become a standard for us. Testing tilt series, recorded with a 2° increment, has not shown significant improvements in reconstructions. However, reconstructions with a 5° increment give significantly less satisfactory results (however cf. Figure 1f, g, h).

Manual recording of tilt series is carried out as follows:

1. Check that the TEM is under optimal conditions, especially for 100 or 120 kV, which might differ from alignment settings at 60 or 80 kV.
2. Preferably, use a rotating grid holder, to adjust that the tilt axis is oriented optimally, i.e. along the length of the chromosomes.
3. Test, at low magnification, that the tilting from 0 to ±60° is attainable and not hidden, e.g. by grid bars. It is not absolutely necessary to cover the whole tilting (lacking one or even two images will not be a disaster).
4. Importantly, test that the preparation is stable, i.e. will not drift significantly during the recording. The stabilization can be improved by slight evaporation of additional carbon on grid specimens or by keeping the specimen in the electron beam, at low magnification, for some time (e.g. 5–10 min). If nothing helps to stop the movement, choose another place on the same grid for a new object, or change for a new grid specimen. Unstable specimens will easily ruin a reconstruction by blurring the whole or a great part of a tilt series. Restoring blurred images for high resolution with image processing is difficult, laborious or impossible.
5. Choose preferably the largest objective aperture, especially for whole mounts to achieve a high transparency, and a suitable magnification. According to our experience, the highest magnification with which we can collect a tilt series without significant radiation damage is about 60 000×. The electron beam must be kept as weak as possible to avoid radiation damage to the preparation. If possible, use the 'minimum-dose focusing procedure' that is available in some microscopes. This may be very useful for sensitive samples especially at high magnification.
6. A eucentric goniometer is preferable for collecting tilt series. The eucentricity of the goniometer must be adjusted as accurately as possible to be of use in collecting the tilt series. If goniometer adjustment is completed the specimen will not move and focusing will not change significantly, at least at low magnifications, when the specimen is tilted within the medium tilting range. Adjustment of the goniometer is, however, not very reliable, especially at high magnifications and tilting, and the focusing and shifting must usually be readjusted. If focusing must be corrected and if it is important to keep the magnification for each image in a tilt series unchanged, the focusing is adjusted with the *z*-control screw of the goniometer and by shifting the specimen back to position. If focusing is adjusted in a normal way during tilting, the magnification of the images will change. However, this incident will usually not be of much significance as most ETM programs will correct the scaling, i.e. magnification of each image, in the alignment procedure (section 5.1).
7. Collect the tilt series starting from 0 to ±60°, with e.g. a 3° increment (our standard). Some instructions[52] recommend starting from one extreme tilt end and move through zero to the other tilt end, to avoid sudden jumps in data collection that may occur owing to insufficient precision of the goniometer.
8. It is advisable to adjust the exposure parameters for photographic EM plates so that the plates can be developed with a machine (e.g. Kodak, used for autoradiography films) available in most laboratories. With this procedure, a check of tilt series is rapidly accomplished, so that negatives with faults, not in focus or missing can be replaced while the specimen is still in position in the microscope.

4.3 Digitalization of Photographic Negatives

Digitalization, i.e. scanning of the tilt series, is necessary for computer processing. Scanning is a more laborious

procedure than manual collection of the tilt series data on photographic negatives. At present scanning is also largely manual. According to our knowledge, there is no scanning device available with automatic feeding of EM plates to the scanner. However, devices for 35-mm film strips and also framed slide pictures can be fed and scanned automatically but not film plates of the size used in EM (e.g. 6.5 × 9 cm).

One of the advantages with EM photographic plates, as mentioned, is the large area of the specimen that can be recorded. This may partly compensate for the tiresome scanning work, in comparison with the limited view covered with a CCD camera. In addition, there are more working prospects with film plates, as they can be scanned with the resolution needed and they allow the choice of other sectors of interest, on the large plates, for scanning and with, e.g. higher resolution.

Ordinary flat-bed scanners (e.g. Umax PowerLook 2000) that we use today have a maximum resolution of 1000 dpi (in both directions, optical and mechanical), i.e. a pixel size of 25.4 μm. Better but also much more expensive scanners are available, e.g. drum scanners, but the manual scanning is much too laborious (attaching many negatives on the drum) for the improved resolution. A high-precision flat-bed scanner (Zeiss, Scan1) is available that has a pixel size of 7.5 μm (precision 2 μm), but the price is much more than 20–30 times that of an ordinary flat-bed scanner.

At a magnification of a negative of, e.g. 50 000×, scanning with 1000 dpi, the pixel resolution in the negative will accordingly be about 0.5 nm. This pixel resolution will exceed the conservative estimate of the resolution achievable, of ca. 5–8 nm, for an object with a diameter of, e.g. 100 nm (with 3° increments), according to Equation (1). However, the actual resolution frequently experienced is better than the conservative theoretical estimate.[52] In theory, the resolution that is achieved, is sufficient for twice, or owing to the many procedures in the alignment, 3–4 times the original pixel size.[52] Thus, oversampling with a 2–4 times smaller pixel size is of advantage for achieving in practice a needed resolution. As the grain size for EM photographic plates is, after chemical development procedures, estimated to be about 5 μm, this will in any case be the limit to the resolution that can be extracted from photographic negatives with digital procedures.

Scanning of the photographic negatives is carried out as follows:

1. Before scanning, dust from the negatives is removed with a soft brush, along with gentle breathing on the negatives, which will help, when the humidity is not high enough, to remove static electricity that might be a problem in dry conditions.
2. Up to eight negatives (size 6.5 × 9 cm), with shiny side down, are positioned on a flat-bed scanner (e.g. Umax PowerLook 2000).
3. The scanning is accomplished with a semiautomatic procedure together with automatic correction of contrast, of histograms, gamma values, etc. for each negative. Manual correction will take longer and the results will not necessarily be better.
4. The area and resolution (1000 dpi, optimal) are selected for the negatives and the title number for the image (corresponding to the tilt angle) and format used (e.g. jpeg) are recorded. The program automatically collects the data of the regions, according to the mentioned settings for each negative, in sequence.

Today, even with the advantages of semiautomatic scanning procedures, the completion of the scanning of all (e.g. 41, 3° increments) plates will usually take longer than the manual collection of the tilt series with TEM.

4.4 Automatic Recording

The automatic collection of tilt series data[40] will in the future replace manual recording methods. Automation is the only manner for very sensitive specimen recordings such as cryo-EMT procedures (section 3.4), which involve embedding in vitreous ice, as the exposure to the electron beam is reduced 10–100-fold, as mentioned. Automated low-dose data collection of a tilt series uses adjacent areas for correction of changes in image shift and focus. The selected area chosen for the tilt series is recognized with the cross-correlation function process. The images are recorded using a high-performance digital camera system, i.e. a cooled slow-scan CCD camera (1024 × 1024 or 2048 × 2048 pixels).[40]

Because of the cost, most laboratories will not have access to such facilities in the near future. However, to improve the resolution in ETM reconstructions, tilt series with smaller increment values are needed, e.g. 1°. The collection of tilt series with 1° increments is for practical reasons only possible with automatic data collection in digital format. Regardless of the preparative methods used, specimens suffer less radiation damage with automatic recording. In addition, the recording of the dynamic gray-scale range and avoiding image distortions are claimed to be superior with a CCD camera, in comparison with photographic film recording. All these improvements involving automatic data collection procedures are dealt with in a recent review.[40]

5 ELECTRON TOMOGRAPHY METHODS

5.1 Alignment of Tilt Series

The tilt-series images collected must be most accurately aligned to be sufficient for ETM reconstructions, especially after manual recording and scanning procedures. Alignment of the tilt series can be accomplished without markers by using the cross-correlation function method as used in the automatic data collection process. However, before the cross-correlation or other automatic method has been better refined, the use of fiducial markers has proved to be a more accurate method for the aligning of tilt series data.[4] A semiautomatic method is available, part of the program IMOD,[66] for fiducial gold marking with the advantage that some gold tags may not need to be present in all the images. However, at first glance, the method appears more complicated than manual methods in use.

As mentioned earlier, different-sized colloidal gold particles on the grids are convenient for different magnifications. The minimum number to be marked is three (or four[52]), ideally 10–20. Corresponding gold particles must be identified on all images of the collected tilt series, and preferably evenly distributed around the object to be reconstructed. Commercial gold markers are preferred, e.g. immunogold, as they appear to be more uniformly round (U. Skoglund, Karolinska Institutet, Stockholm, Sweden, personal communication) than laboratory-made markers.[67] The spherical form is important, as the program searches the midpoint pixel of the gold as the target for coordinates registered. Nevertheless, the search parameters can be turned to a minute search area to respond to manual correction, if an automatic search does not give the appropriate exactness. If sufficient gold particles cannot be found, other naturally appearing granules, etc. are usable as alternative markers within the area, only to be tracked accurately enough. It may be useful to add that in the program we have developed, the gold particles can be situated on any side or level on the grid. They need not to be on the same plane.

Our program, JPEGANIM for SGI machines,[28] used for the procedure has been appropriately designed for convenience, using interactive graphic interfaces with menus. The images of the whole tilt series are run as a movie, so each image can be sequentially viewed, back and forth, for rapidly recognizing the same gold particles in each image. The images can be contrast-enhanced temporarily or permanently, enlarged and moved to different regions when the image size exceeds the display, and processed otherwise, e.g. with unsharp masking and other filters. All these features aid in making the search and gold marking an easy and rapid operation.

When sufficient markers have been picked up and the files saved, the next step will be to test the alignment accuracy, using SOLVE from the menu in JPEGANIM. The program uses a conjugate gradient method that gives a set of overdetermined nonlinear equations that can be solved.[28] After running, the program produces a plot with evaluation of alignment correlation, in pixels, with estimates for all images and gold markers used, as graphical scores. The refinement is accomplished, when necessary, by either removing the worst gold markers, in sequence, or trying to correct the position of poor gold markers in the images with the poorest scores and rerunning the program. A set of files is used in the next step for an evaluation, which includes corrected tilting angles, scaling, midpoints and the axial angles.

The alignment procedure will then be tested with ALIGN from the menu, e.g. on a small area with only one, or a few, gold particles. A midpoint is chosen and an area framed with the mouse, producing a sequence of pictures. The z coordinate for the midpoint is obtained by using the parallax method, i.e. marking the same point from two image tilts, the zero and an end tilting image, for the best accuracy. If the same point is difficult to mark in the endmost tilt, any other image can be used. An area is then framed for the object to be aligned, from the midpoint, and the program will produce a set of aligned images, twice the size of the frame chosen.

After running the program, the accuracy of the aligned images can be tested as a movie. The movie must run smoothly and keep the midpoint in place, to obtain an accurate reconstruction. Scaling, rotations and translation of the origin are all corrected by the program.

With the test procedure adequate, the object for reconstruction framed and aligned images saved, there is saved a file including the size of the object, in pixels, with corrected tilt angles and a filtering parameter used in the reconstruction procedure. Occasionally, if a set of aligned images show sudden jumps at certain places, the corrected tilt angles are replaced with the original angles chosen, i.e. ideal values with the increments for the tilt series. (Direct goniometer readings are more precise than computed values.) A realignment (2–3 times) procedure of aligned images may also be of help in correcting the alignment for difficult tilt series.

The aligned images can be viewed in stereo mode. Gold marking in stereo can be of great help for identification in difficult circumstances. The stereo mode is also most useful in the examination and interpretation of the original EM images as stereo movies, involving image processing, contrast and filtering procedures for comparison with the 3-D reconstructions produced with ETM.

5.2 Reconstruction Methods

The ETM reconstruction programs that are used need a set of accurately aligned images, i.e. Radon transforms. The aligned projections must be twice the area of the object region selected and, as mentioned, this is accomplished by the alignment program, to be implemented for an ETM 3-D reconstruction of the object.

A reconstruction is first produced, as a test, with the WBM, a quick procedure in SGI machines (Origin 2000, 128 R12k processors, 160 Gbyte RAM). If the reconstruction with WBM shows adequate results, the reconstruction is completed with MEM. Today we use COMPAQ AlphaServer 8400, eight EV6 processors, 12 Gbyte RAM. This is advantageous, as ETM with MEM is calculation intensive. (An earlier approximation, to compare the time taken by the programs to complete a reconstruction of a certain volume, was the following: WBM took the same time in minutes as MEM in hours. However, this match is not very valid for long, as the computers are steadily upgraded, and now the time needed for MEM processing is less than half of our earlier estimate.)

5.3 Weighted Back-projection Method

For some reasons, the programs available for ETM [e.g. IMOD (URL: http://www.soapdish.colorado.edu/imod/download.html) public domain program and SPIDER & WEB (URL: http://www.wadsworth.org/spider_doc/spider/docs/spider_license.html) by license] use WBM processing for ETM reconstructions. This may partly be due to earlier limits in computing capacity. Another reason might be that WBM[36] has a long traditional line going back to Radon (1917).

In practice, our program (Backproject) does not need very much attention by the ordinary user and the reconstruction, once the filtering parameter has been determined, is an automatic procedure. The filtering parameter may be varied for an influence on resolution, smoothness or noise level of the reconstruction. However, in optimal cases, the appearance and resolution might be improved, but real improvements in reconstruction are more affected by the original quality of the tilt series and accuracy of the aligned images.

After developing and acquainting ourselves with the MEM procedure, we are now using WBM only as a rapid scrutiny method having in fact rejected WBM as a reliable reconstruction method. It is therefore unnecessary to go into the theoretical contents of WBM, as they have been presented earlier in detail.[28,36]

5.4 Maximum Entropy Method

The MEM was first introduced in an astrophysical context for filtering telescope pictures from noise, and the method initially presented significant improvements in image quality. The introduction into 3-D reconstruction procedures in structural biology was accomplished for the first time for chromosome studies, as earlier mentioned.[43] After this initial work, relatively few investigators in ETM have shown an interest in using MEM. This must be partly due to a lack of computing power and long processing times, or to poor evaluation methods used in viewing 3-D reconstructions.

In cooperation with Juha Ruokolainen (CSC, Center for Scientific Computing, Espoo, Finland), we decided to develop a MEM version of our own that was ready in early 1995. Later we extended our MEM also to a 16-bit MEM version that uses aligned 16-bit gray-scale images.[68] We became intensely aware of the need for better 3-D quality when we recognized the relatively poor accuracy of the WBM reconstructions viewed with the high-quality stereo procedures for ETM that we had developed earlier.[28]

MEM yields uniformly better reconstructions than WBM (cf. e.g. Figure 1f, g, h with Figure 5a, b), especially for whole-mount samples of chromosomes: they appear distinct in stereo viewing of the object with the volume visualization programs BOB and FUNCS (section 6.1). In addition, the noise level in WBM-reconstructed objects, as measured by α-values, is often as high as 100 or more (of the 256 gray-scale levels) when using BOB or FUNCS. With MEM the α setting is much less, often close to zero.

Another way of comparing the quality of WBM and MEM reconstructions is to examine the shape of gold markers. With WBM the missing data wedges between 60 and 90° stretch structures along the electron beam (z-axis). According to the results, MEM corrects errors due to missing data more effectively than WBM, and the gold markers are clearly less oblong with MEM.

The disadvantage with MEM is the long calculation time needed for large volumes (e.g. $400 \times 400 \times 500$ pixels), only manageable for laboratories having access to supercomputer facilities. However, even desktop computers are today as fast as yesterday's supercomputers, so it should not be long before any laboratory will have the capacity to perform such computing. For smaller volumes (e.g. $100 \times 100 \times 100$ pixels) with an ordinary desktop computer this might already, or soon, be possible within a reasonable performance time.

A description of our MEM deconvolution is as follows.[68] Solution of the deconvolution with MEM aims to minimize the function that is composed by the square of the reconstruction error (Equation 2):

$$e^2 = \|\rho - Af\|^2 \tag{2}$$

where ρ is the images and A is the back-projection operator, and by the regularizing term. The regularizing term in MEM is the entropy of the object to be reconstructed (Equation 3):

$$E = -\sum_i \sum_j f_{ij} \ln f_{ij} \tag{3}$$

where $f_{ij} = f(x_i, y_j)$ is the object to be reconstructed. The quantity f must be positive. Combining Equations (2) and (3), the problem may now be presented as an optimizing problem: search for f_{ij} such that the expression $\mathbf{A}$ is at minimum (Equation 4):

$$\mathbf{A} = \sum_i \sum_j f_{ij} \ln f_{ij} + \lambda \|\rho - Af\|^2 \tag{4}$$

where A is the Radon transform operator representing the reconstruction function as back-projections of images. The Lagrange coefficient, λ, is a regularization parameter, which weights the reconstruction error with respect to the entropy.

As mentioned, the experiences with MEM are promising. For example, noise is notably eliminated, and so is the waving that the Fourier transformation also causes. In addition, according to the definition, MEM produces a positive reconstruction of the object, in contrast with the Fourier transformation method. The density quantity should be positive. Additional MEM deconvolution procedures can be found elsewhere[36,43] and in a recent application.[41]

Our MEM procedure (MaxEntropy) is almost as transparent for the user as WBM. In the file that is created in the saving of the aligned images, the regularization parameter is added after the filtering (default) value. If problems appear, often seen as regions of minor stripes or bands, then this regulating parameter is tested with other values until the reconstruction has been refined to remove the minor distortions.

There is an additional regularization parameter for our 16-bit MEM. This 16-bit MEM, as yet not much used, will be especially useful when 16-bit gray-scale images are accessible with high-performance 16-bit CCD cameras with automatic data collection of tilt series, especially in cryo-ETM procedures.

As already mentioned, MEM is executed on a COMPAQ AlphaServer (8400) SG 140 (eight processors, 12 Gbyte RAM), especially for large volumes. However, the data have to be converted to be usable for viewing the volumes in the SGI machines, as the bytes in the Alpha machines are in a different order. This inconsistency is corrected by using a conversion program in the Alpha machine to make the volumes applicable for SGI machines.

Actually, a set of different conversion and other subprograms are repeatedly used. There is no reason to go into all the details, as different programs and machine set-ups may require different modifications.

5.5 Dual-axis Tilting Method

In conventional single-axis tilt series, the tilts used are limited to the range from 0 to $\pm 60^\circ$. As already mentioned (section 4.1), grid holders are available for increasing the tilting range, e.g. from 0 to $\pm 70^\circ$ (or even up to 80°), but at the highest tilting ranges occultation easily occurs, and for sections the thickness will in general become too large, especially at very high tilts, e.g. two times thicker at 60° and three times thicker at 70°,[52] for collecting reliable data with conventional 100–120-kV TEM.

Methods have recently been introduced to reduce the missing data wedge, e.g. between 60 and 90°, by using dual-axis tomography tilt series,[69,70] The two programs transform the missing wedge to a missing pyramid, and combine somewhat differently the two orthogonal single-axis tilt series obtained by 90° rotation.

The dual-axis tilting approach has shown improvements in some situations, e.g. membrane vesicles appear restored from distortions in comparison with the single-axis method. Detailed instructions on the procedure, with or without a specimen-rotating grid holder, have recently been prescribed.[52]

It seems doubtful if any improvements can be obtained for oblong structures such as chromosomes and chromatin fibers, as microtubules oriented perpendicular to the tilt axis fade and are lost in reconstructions.[52] In any case, improvements have been claimed, at least in some cases, also for microtubules.[70] According to our experience, long structures, e.g. microvilli, when tilted perpendicular to the tilt axis, appear very unsharp and faded along the reconstruction. It is obvious that the dual-tilting axis method is useful only for fairly rounded objects that need not be specially oriented for collecting tilt series.

In our experience, to test cryo-preserved preparations (in cooperation with H. Cheng, Karolinska Institute, Stockholm, Sweden, and M. Marko, Wadsworth Center, Albany, NY, USA) using our MEM procedure and combining the dual-axis tilt reconstructions with IMOD,[70] no clearly distinct improvements emerged. However, all possible refinements have not yet been accomplished and the original tilt series from the sections used were, for some unknown reasons, difficult to align accurately. Such inaccuracies may have contributed to the unsatisfactory result, as the improvements usually detected with a dual-axis tilting method seem to be the most delicate[70] and advantageous when all basic parameters are refined to the optimum.

6 RENDERING OF RECONSTRUCTED CHROMOSOMES

6.1 Visualization of Three-dimensional Reconstructions

When a chromosome reconstruction has been completed with WBM or MEM, there are several ways to visualize 3-D volumes. In any case the reconstruction must be in the right format for the program. For such conversions another set of programs are used.

As indicated, we regularly use two programs for SGI machines, BOB (a public domain program, University of Minnesota, Graphic Visualization Laboratory, Army High Performance Computing Research Center) and FUNCS,[28] that efficiently display volumes of 3-D reconstructions of, e.g., chromosomes in stereo mode.

The SGI workstations (Octane, 1 Gbyte RAM, and Onyx2, 4 Gbyte RAM with 4 MIPS R10000 processors and InfiniteReality2E-graphics), with high-frequency displays, and provided with liquid-crystal stereo (Crystal Eyes) eyeglasses, synchronized by an infrared light device controlled by the computer, show images in color stereo mode. According to our experience, stereo viewing is the only way of even beginning to comprehend the geometry and 3-D structure of chromosomes (Figures 1f, g, h, 8a, c, 9a, b and 10b, c). Other methods might involve

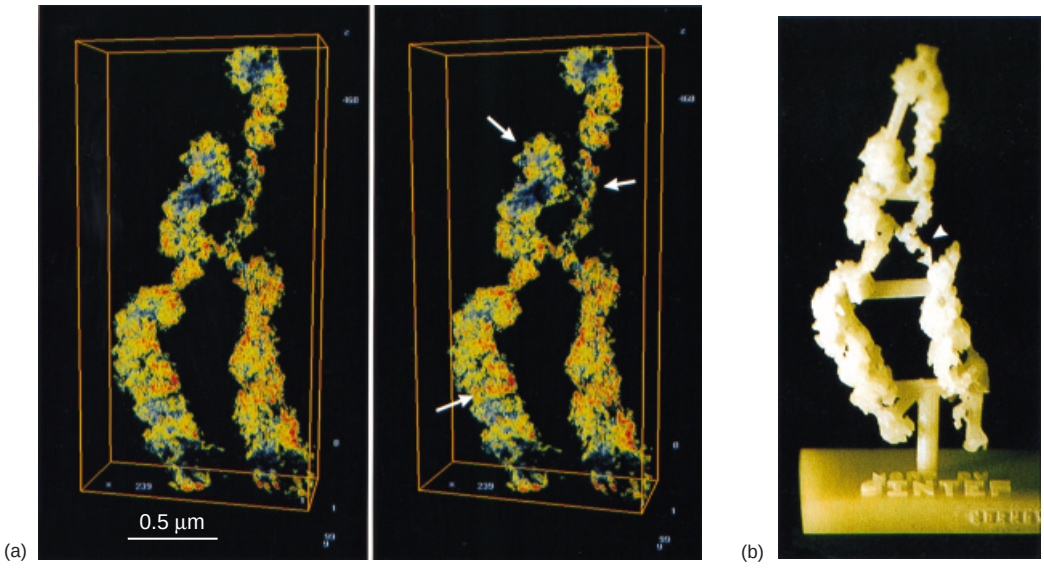

Figure 8 (a) Chromosome scaffold, stereo pair of 3-D reconstruction with MEM, section 5.4. [AA-isolated chromosomes from HeLa cells, prepared as in section 2.4, procedure: 1(b) and (c), then treated as in section 2.3, with DNase and RNase digestion; then as in section 2.4, Procedure 1(d) and dried from *tert*-butanol, section 3.2.] Typical coiling appears in chromosomes, although they are depleted of histones and DNA, showing that scaffold elements are maintaining the coiling. The pattern and structures that can be detected are wholly in agreement with the unified loop-and-rosette chromosome model (Figure 1). The unified model gives predictions about the structure of the chromosome core regions, i.e. scaffold, in all details. Ring-like structures that are in agreement with the size of cyclomeres are detected (arrows) in the chromosome scaffold. In addition, gyres of the scaffold macrocoils measure to fit two rows of cyclomeres, e.g. in a zigzag pattern [cf. zigzag ribbon model for arranging nucleosomes in the 30 nm chromosome fiber[4,17]], although minor coils,[25] e.g. solenoid arrangement of cyclomeres, can also be seen closer to the centromere region, indicating that the structure of the chromosome coiling is versatile. (b) Hard copy, i.e. solid 3-D plastic model (section 9) produced earlier[28] from a WBM 3-D reconstruction. Coiling and many of the ring-like structures are found in the hard copy and importantly arch-like bends (arrow) that the unified loop-and-rosette model also predicts (cf. Figures 1b and 2c). (c) Immunoblottings with antibody (Ab) to H1 and H2b (antibodies to histones H1 and H2b kindly provided by Dr M. Bustin, NIH, Bethesda, MD, and Dr B.D. Stollar, Department of Biochemistry, Tufts University School of Medicine, Boston, MA, USA, respectively): 60% AA-isolated chromosome pellet, 60% AA-isolated chromosome pellet that has been further boiled (0.5–2 min) in 50% AA, and as positive controls the supernatant of 60% AA extract and of whole cells. LMW (low molecular weight) protein markers (Bio-Rad). Text in each sample column. (d) AA-isolated chromosomes fluoresce brightly with the DNA-specific fluorochrome DAPI. Digestion with nucleases DNase I and RNase A (section 2.3) remove all detectable fluorescence, but chromosome scaffolds retain their original shape and can be seen in LM (text in each picture). [(a–d) partly published in Engelhardt et al.[28,54]] (e) Stereo series of the chromosome scaffold arm [see (a)] with 10° increments around the tilting axis for each row, to observe the chromosome coiling from various directions. Ring-like structures of cyclomere size (arrows) are present. Thicker gyres (arrowheads) of the scaffolding coils seem to be fitted with two rows of cyclomeres in a zigzag pattern as suggested in the unified loop-and-rosette chromosome model [Figure 1, and cf. (a)]. [There is resemblance to intact chromosomes, with histones and DNA, having 'large-scale domains, roughly 100–130 nm in width', as described in EM tomographic sections of metaphase and telophase chromosomes by Belmont;[30] cf. his Figure 1.]

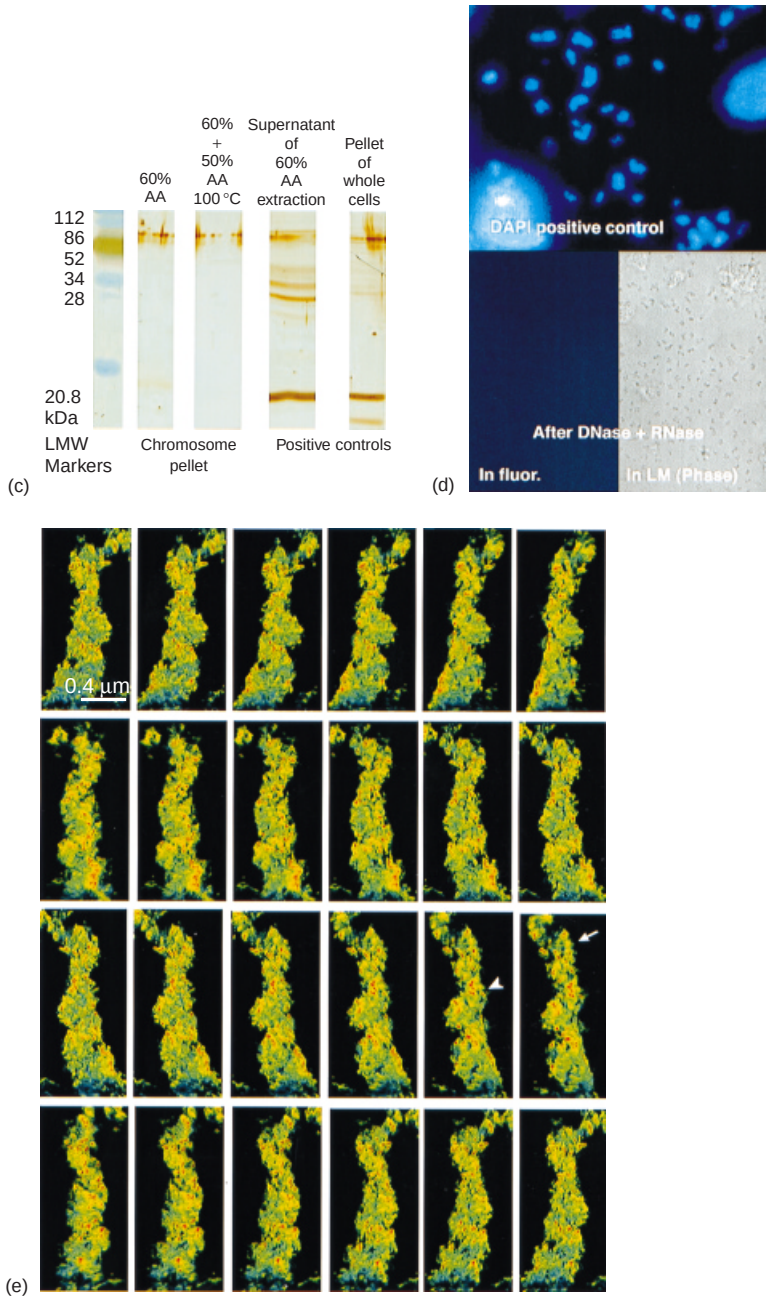

Figure 8 (*Continued*)

underestimation of the structural complexity and be rather pointless in really getting to grips with the 3-D reconstruction by ETM.

The 3-D reconstructions of chromosome are shown as translucent objects, regulated by alpha blending, and color may be introduced for different gray-scale ranges, composed of voxels, the size and density correlating with the pixel resolution of the original EM image projections. Viewed in mono and gray-scale, the reconstructed 3-D chromosomes look identical with the original 2-D EM images. Gray-scale viewing in stereo mode may actually be a more correct way to resolve complicated structures than replacing gray-scale values with various colors; although the colors look inspiring, they may actually confuse, not only beginners, rather than clear up complex structures. However, replacing gray-scale values with different colors can effectively intensify minute density differences for detection and analysis (e.g. Figure 9b).

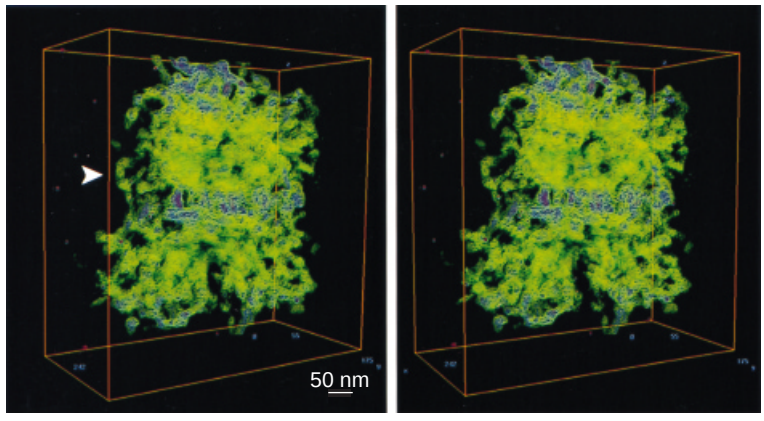

(a)

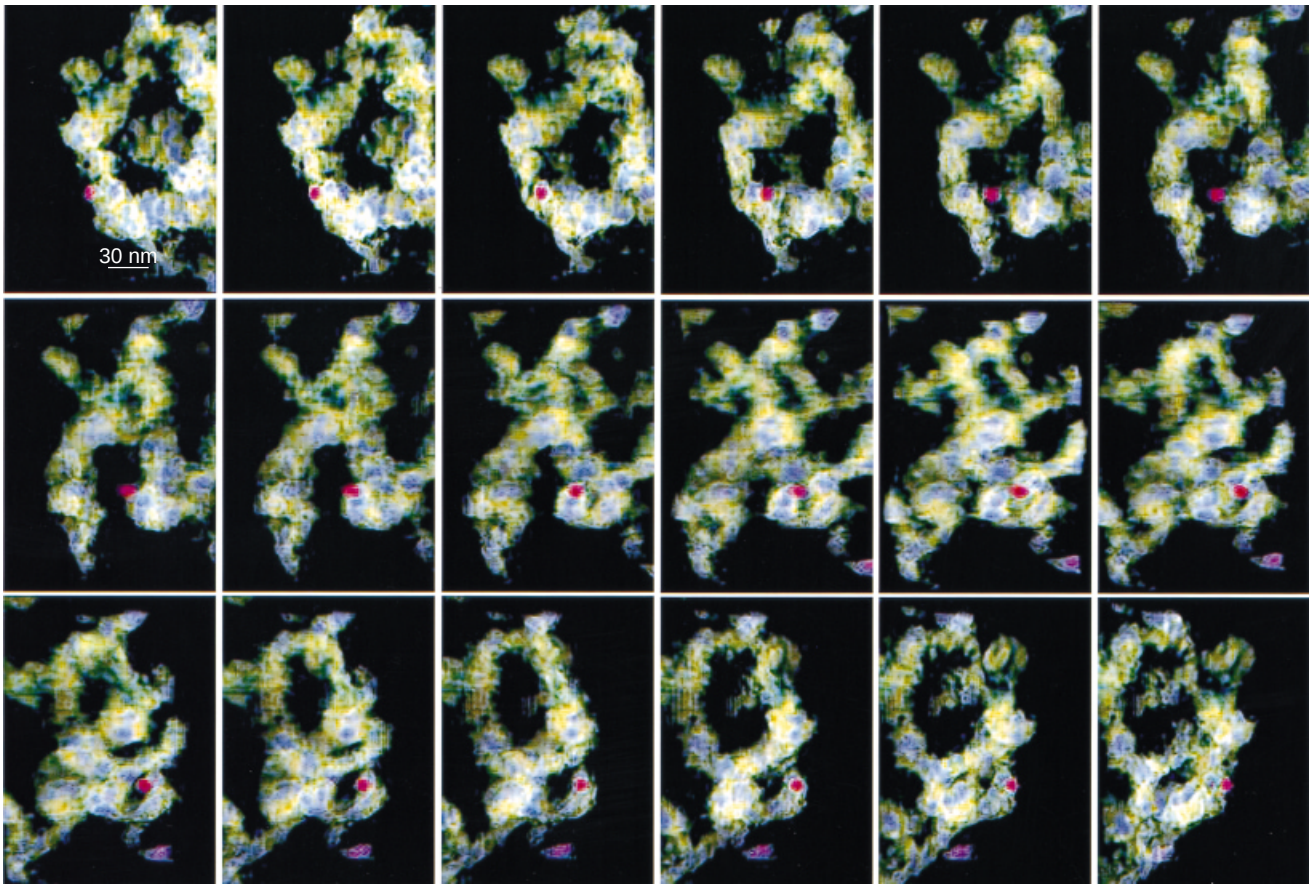

(b)

Figure 9 (a) Stereo pair of MEM 3-D reconstructed chromosome scaffold of a small (HeLa) chromosome (prepared as in Figure 8a) visualized with BOB (section 6.1; cf. Figure 1f). Coiling is obviously not prevalent in this size class of chromosomes, as it would require a certain length. However, scaffold fibers can be studied in more detail. Scaffold fibers appear as disordered fibers and arches or have an open ring-like form (arrow), as the loop-and-rosette model predicts. In the background fiducial gold markers (red color) are seen. [Reproduced from Engelhardt and Ruokolainen[54]]. (b) Detail from (a) (thick arrowhead), stereo series of scaffold fibers rotated along the tilting axis, with 10° increments. Along the fiber perforation-like structures are obvious (light blue). A 10-nm gold particle (red) is visible among the fibers. A description of the perforation pattern of the fiber structure has been presented earlier[54] as peripherally located toroid-like structures of about 8 nm in outer diameter with a central hole of about 4 nm, but also open toroids could be detected along the spirally wound scaffold fibers. It was suggested that the perforations (toroids) are manifestations of topoisomerase II (a main scaffold protein). [Stereo pair in (b) reproduced from Engelhardt and Ruokolainen.[54]]

Furthermore, a 3-D reconstruction can be rotated and moved in any direction, enlarged, sliced and details can be cut out for closer viewing, or joining of voxels can smoothly display large volumes.

We use the programs for different aims. BOB is an easy and fast way for regular viewing of large and small volumes in stereo mode. FUNCS is slightly more difficult to use, but includes more features, such as introducing different-colored lights from different directions, interpolation of volumes or showing the volume as surface view (isosurfaces). Most importantly, FUNCS allows high-quality animations (section 6.2).

Other methods, such as sectioning the volume from any direction into 2-D slices that can be viewed in series, as a slow or rapid movie, also in stereo mode, or from all directions simultaneously, are available for dissecting and analyzing reconstructions.[52] These and other features are found in such programs as IMOD and SPIDER & WEB (section 5.3).

6.2 Producing Stereo Animations from Three-dimensional Volumes

The problem, when viewing high-resolution reconstructions of, e.g. chromosomes, is that volumes (over 200 × 200 × 200 pixels) become slow, i.e. scenarios have to be recalculated by the computer, e.g. when the image is enlarged or the direction is slightly changed, e.g. in tracking structures in detail. The recalculation takes time, relative to the size of volume, which causes difficulties in following details when minor structural changes take place, i.e. causes loss of the track.

New effective high-end SGI workstations (as detailed in section 6.1) are available that may solve the calculation problems at least up to some volume sizes. However, the programs have to be updated (rewritten) for new workstations.

A much simpler and effective way to solve this kind of problem, which we have used earlier, is by creating high-resolution animations[28] around an axis, usually the tilting axis (Figures 1g, h, 8e, 9b, 10c). These animations can be produced from any size of reconstructions that are recovered with ETM, including interpolations and illumination from different directions, both of which improve significantly the appearance of details. Animations are produced with FUNCS, around the tilt axis with, e.g. 10° increments and three-time interpolations. However, animations with smaller increments, e.g. 1–5°, improve the tracking of details considerably. Depending on the size of the volume, the animations are produced in less than 1 h up to o/n.

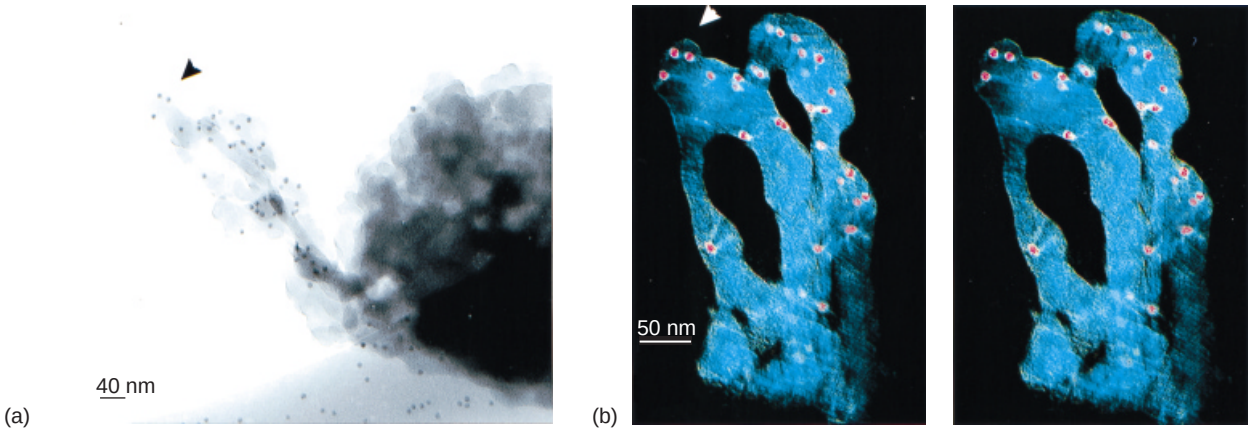

Figure 10 (a) Tilt picture (+45°) from the tilt series (0 to ±60°, in 3° increments) of chromosome scaffold (prepared as in Figure 8a). Overhanging scaffold fibers are seen originating from the chromosome scaffold torn from the carbon–Formvar film, visible at the picture base with 10-nm gold particles. The freely dangling scaffold fibers were used for the 3-D reconstruction to achieve the highest resolution. After high-resolution scanning the pixel size corresponds to 0.4 nm in the tilt series, collected at 58 000× magnification. [Another tilt picture from this series published in Engelhardt and Ruokolainen.[54]] (b) Stereo pair of MEM 3-D reconstruction of freely hanging scaffold fibers. Fiducial 10-nm gold particles are seen tagged to the fibers. (c) Stereo series, with 10° increment, of MEM 3-D reconstruction of top part [arrowhead in (a) and (b)] of the scaffold fibers, from high-resolution animation series (interpolating pixel vs voxel size to 0.1 nm). In stereo, an impression is made of a slender mostly peripheral lace-like structure, demarcating sharply the scaffold fiber periphery. [Outer mesh diameter 3.5 nm, inner mesh diameter and thickness about 1.5 nm. These values are rough as they are on or below the limit of the 'conservative estimate of resolution',[52] which is about 3 nm, Equation 1.] Stereo viewing reveals that the gold particles are inside the stretched fibers. This confirms that the fibers are very empty indeed. If the lace-like covering is real it could be part or derivative of the NE lamina that has recently been shown to be found inside the nuclei with IEM[32] after DNase digestion. [Stereo pair of this series published in Engelhardt and Ruokolainen.[54]]

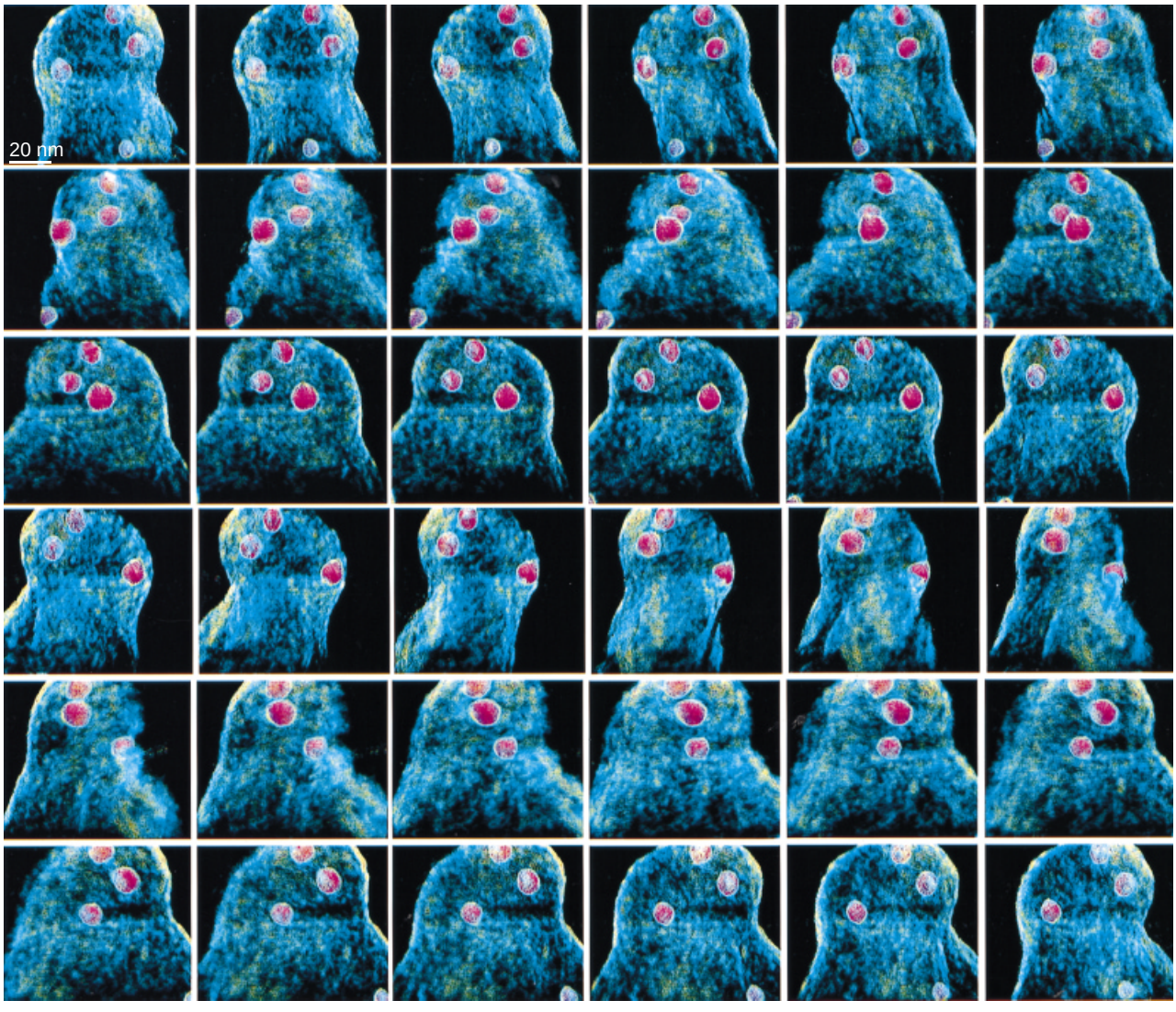

(c)

Figure 10 (*Continued*)

The images of an animation are saved in high-quality jpeg format and viewed with JPEGANIM, in stereo mode. Also, as already mentioned, from the menu of JPEGANIM, the movies can be magnified, moved to different regions, and structures can be measured with great precision, run with different speeds or very accurately by stepwise moving, smoothly, frame after frame. With these animations, the stereo viewing is not disturbed by blurring recalculations as the animations are sequences of ready-processed images that are all in the memory of the SGI workstation (as detailed in section 6.2).

6.3 Computer-aided Virtual Environment

Another means of 3-D visualization of volumes that we have recently used for viewing 3-D reconstructions of chromosomes, usable also for an audience, is the computer-aided virtual environment (CAVE) (available, e.g. at the Helsinki University of Technology).

The CAVE is a room, with wall-sized displays (three walls and the floor), controlled by a high-end SGI workstation (Onyx2, Infinitereality, 1 Gbyte RAM and four R10000 processors). High-end precision projectors, controlled by the SGI workstation, project the images of the volume on the walls. The viewers stand in the middle of the CAVE with liquid-crystal stereo (Crystal Eyes) eyeglasses. The rotating volume can be seen in 3-D form from all directions and from within at high resolution.

No doubt the 3-D view obtained is most impressive with very accurate details, as the whole volume is seen with all details and not limited by frames, as happens with

ordinary computer displays, when the object is magnified. The impression is that this is not only an excellent method for showing 3-D reconstructions for a group, to promote meaningful discussions among researchers, but also an important tool for perceiving structures that cannot be easily displayed in other ways.

7 IMMUNOELECTRON TOMOGRAPHY

The ultimate goal, in structural studies and particularly in 3-D reconstruction studies with ETM of biological specimens, would be to recognize different proteins, nucleic acids and other macromolecular assemblies straight from their structural 3-D configuration.

Before this ideal goal, achievable perhaps in the distant future, we have to use some other means of identification, such as immunological methods. Immunological methods work strikingly adequately in locating specific proteins at LM and EM levels.[67]

We have recently developed IET methods.[47] In principle, these methods do not differ much from ordinary IEM methods. We have used whole-mounted cells grown on EM grids for IET. Primarily, whole-mounted cells are not usually used in IEM in the way that thin sectioning or whole mounts of protein assemblies or objects of comparable size have been used. In whole-mounted cells or objects such as chromosomes, the immunolabeling would be difficult to distinguish because of the thickness and superimposition of immunolabeling in the 2-D images in TEM of such samples.

However, whole mounts are very suitable for IET. Instead of attaching gold markers on the grids for aligning the tilt series, immunogold labeling will be most adequate for the alignment. The different-sized immunogold markers will identify different proteins on the same preparations.

The problems in the development procedures were obtaining clean control preparations and preserving the cell ultrastructure, antigenic properties of the proteins and making cells permeable to immunogold labeling. In our IET reconstructions, we could easily see 5- and 10-nm gold conjugates. In addition, in distinct reconstructions, 1.4-nm immunogold (Nanoprobes Inc.) was detected especially clearly in overhanging microvilli, where the carbon–Formvar support had been torn away. This was to accomplish a set of high-resolution tilt series, so that the resolution was not obstructed by the support (Figure 11).

It remains to be seen how IET works for whole-mounted chromosomes, but we see no prominent difficulties. The steric hindrance by DNA present may cause penetration problems for immunogolds used. Thus, DNase-treated chromosomes, i.e. isolated chromosome scaffolds, would be the primary target in identifying scaffold proteins in 3-D reconstructions. The antigenicity of proteins may be hampered by 60% AA isolation of chromosomes, so that a more physiological method for chromosome scaffold isolation[19] might be immunologically more suitable. In any case, 60% AA-extracted histones had not lost their antigenic properties (Figure 8c) as shown in immunoblottings.[54]

IET procedures[47] modified for whole-mounted chromosomes are as follows. The starting material is chromosomes or chromosome scaffolds[19] mounted on PL-coated or glow-discharged carbon–Formvar–Ni grids[51] (section 2.2). The treatments are performed on Parafilms, with drop-sized depressions on wet filter-paper in Petri dishes (section 2.3). Washings and incubations are performed at RT, preferably on a tabletop shaker (gentle shaking). All solutions are filtered through a 0.22 μm (Millipore) filter. Small amounts that are difficult to filter are centrifuged, e.g. Ab mixtures, GA and OT solutions.

1. Prefixing in 0.125% GA in HM solution, 15 min.
2. Without prefixing:

   Wash in HM solution, several drops.

   Preincubation: wash and incubate o/n in TXHS (0.05% Triton X-100, 30 mM HEPES, pH 7.4, 0.1–0.5 M NaCl, 20 mM KCl, 5 mM $MgCl_2$, 20 mM glycine, 0.02% NaN_3), with (a) 0.5% fish gelatin (FG), or (b) no FG, 1–2 h at RT, and/or +4 °C o/n.

   (a) Incubation in primary Ab mixture (1 + 2):

      (i) Rabbit polyclonal Ab for protein 1 (LM-checked, e.g. $1:50–1:10^3$).

      (ii) Mouse monoclonal Ab for protein 2 (LM-checked, e.g. 1:25).

      Diluted in (a) TXHS + FG or (b) TXHS, 1–2 h at RT, and +4 °C o/n.

      Wash in TXHS, in several drops (1–10 min per drop).

   (b) Incubation in secondary gold-conjugated Ab mixtures; examples:

      (i) Mixture: goat anti-rabbit (GAR)–10 nm gold (1:25 dilution, Sigma) with GAR–1.4 nm gold (1:25 dilution, fluoronanogold, Nanoprobes Inc.) and goat anti-mouse (GAM)–5 nm gold (1:25 dilution, Sigma), diluted in (a) TXHS + FG or (b) TXHS, and/or

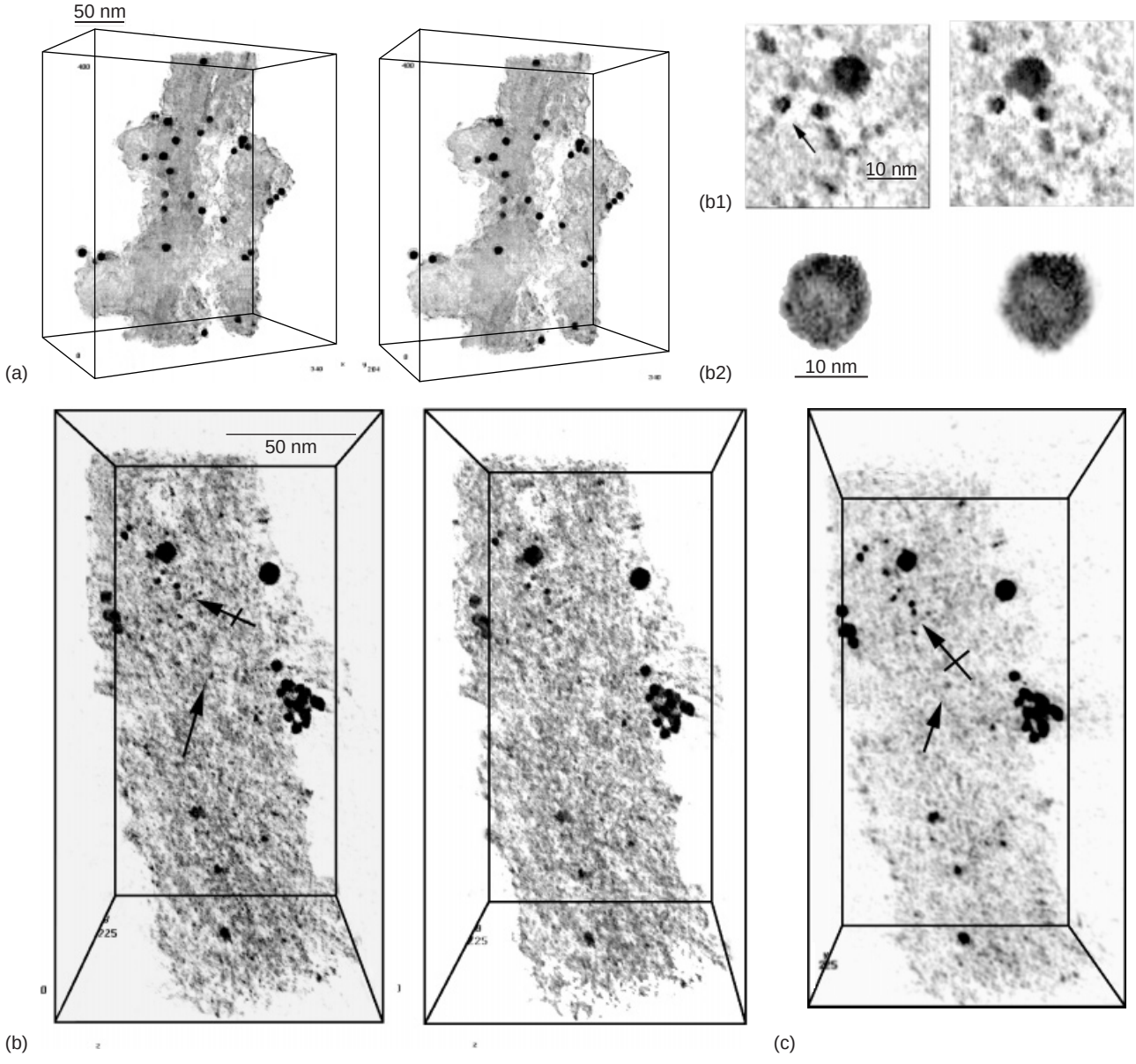

Figure 11 Examples of IET. Microvillar proteins in MRC-7 cells. Overhanging microvilli (without carbon–Formvar support) of MCF-7 cells,[47] will be representative samples of IET methods developed that may analogously be used for IET chromosome studies (section 7). (a) Stereo pair (BOB, cf. Figure 1f) of microvillar ezrin labeled with 10-nm gold particles. The specimen was postfixed with the GA, OT, TA method[47] (section 7). TA was used to minimize shrinkage of the samples after GA fixation (sections 2.4 and 7). The raw IET data (sections 4.2, 4.3 and 5.1) were processed with MEM (section 5.4). The 3-D preservation using TA shows more substructural details than without TA [cf. (d)]. (b) A stereo pair image [as in (a)] of microvillar ezrin and mucin. The specimen was postfixed with the TA method [as in (a)]. Ezrin was labeled with 10- and 1.4-nm gold conjugates and mucin with 5-nm gold conjugates. MEM 3-D reconstruction methods were used [cf. (a)]. Ezrin labeled with 1.4-nm gold particles is marked by arrows. Arrow with ribbon shows clusters of 1.4-nm gold particles. The intermediate-sized particles depict mucin labeled with 5-nm gold conjugates. (b1) Detail of (b) showing a stereo pair of 1.4-nm gold clusters (arrow) at higher magnification. (b2) Details in a single 10-nm gold particle. (c) IET reconstructed with WBM (section 5.3), for comparison of the same microvilli as in (b) with MEM procedure. Gold labels are seen but the 3-D reconstruction is less sharp and more noisy. (d) Stereo pair image [as in (a)] of MEM 3-D reconstructed microvillus as a control without TA (section 7). Ezrin was labeled with 10- and 1.4-nm gold conjugates and mucin with 5-nm gold conjugates. The 1.4-nm gold conjugates are present but more difficult to detect than in (b) with TA (presumably the detection of 1.4-nm particle size is more difficult owing to the shrinkage caused by GA fixation and CPD, which the treatment with TA seems to prevent.[55] [Reproduced from Engelhardt et al.[47]]

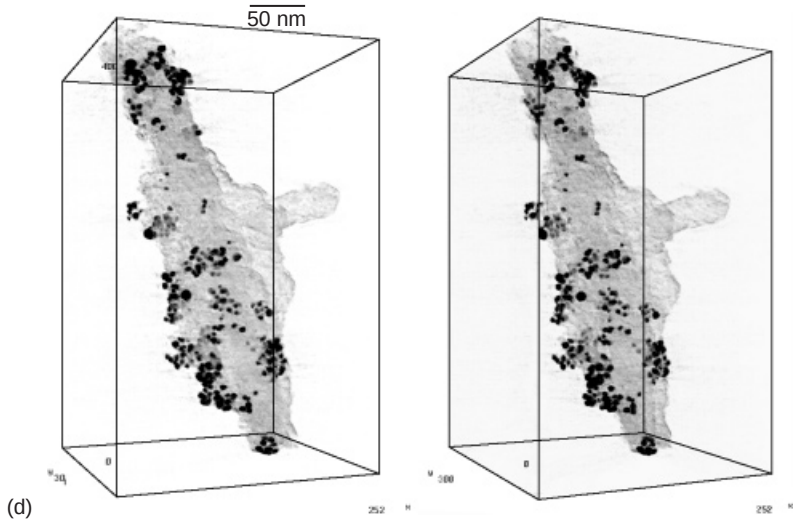

Figure 11 (*Continued*)

(ii) Mixture: GAR–10 nm gold (1 : 25 dilution) with GAM–5 nm gold (1 : 25 dilution) and GAM–1.4 nm gold (1 : 25 dilution), in (a) TXHS + FG or (b) TXHS.

Incubation: 1–2 h at RT and 4 °C o/n.

Wash in TXHS, several drops (1–10 min per drop).

Wash in HM, several drops.

3. Postfixing: 1.25% GA (1 h–o/n) in HM, (a) add OT (from e.g. 1% stock solution) in 0.1 M NC buffer, pH 7.4, to 0.3% OT as a mixture with GA in HM, 1–5 min, RT; (b) no OT.

   Wash in HM, several drops.

4. TA: (a) 1% TA in 0.1 M NC and HM (1 : 1), 5 min (for whole-mounted cells 1 h–o/n); (b) no TA.

   Wash in HM, several drops.

5. Dehydrated in 30, 50, 75, 100% MeOH.
6. Stain in 0.001–0.002% UA in MeOH (30 s), wash in MeOH.
7. CPD or other 3-D-preserving drying methods (section 3).

Controls: In IET controls, control (preimmune) serum was used or primary serum Abs were omitted and samples were postfixed only in GA (with or without TA), with no UA before step 7.

LM: immunofluorescence fixed in 3.5% paraformaldehyde in HM, using secondary GAR–fluorescein isothiocyanate (FITC) and/or GAM–Rhodamine, according to standard protocols,[51] mounted in DAPI/MM.

For immunolabeling of resinless thick sections, enzyme-digested or not, of whole-mounted cells, see section 3.5.

8 IN SITU HYBRIDIZATION ELECTRON TOMOGRAPHY

Analogous with IET methods, the same development concerns in situ hybridization electron tomography studies. In situ hybridization methods have been accomplished for EM earlier[67] and in microspreading of meiotic chromosomes, i.e. EMISH (electron microscopy in situ hybridization).[51] In principle the very same methods would be sufficient.

In addition, LM in situ hybridization techniques are standard procedures[51] only to replace, e.g. after the first biotin–avidin amplification procedures, the second one with gold-conjugated avidin, and replacing AD with, e.g. CPD (section 3).

FISH (fluorescence in situ hybridization) preparations are frequently contaminated with background signals that may be difficult to eliminate and spoil the accuracy required with in situ hybridization for EM preparations. In addition, cytoplasmic fragments regularly contaminate LM and microspreading preparations in EM studies.

It is therefore essential to strive for chromosome purification methods where cytoplasmic contaminations are eliminated, to be useful for in situ ETM. In addition, it is important to refine blocking methods to remove any background and unspecific labeling.

Superimposing contaminants and background labeling may disturb ordinary in situ hybridization EM preparations. ETM has the advantage that from the 3-D reconstructions, superimposed contaminating materials may be eliminated, as chromosomes with specific labeling can be cut out from surrounding material with ETM procedures and the pure 3-D chromosome reconstructions can be analyzed separately in detail.

9 HARD COPIES, PRODUCTION OF SOLID THREE-DIMENSIONAL MODELS

Hard copies, e.g. photographic prints, are produced of computer images, and hard copies of 3-D reconstructions of, e.g. chromosomes can also be produced, as we showed years ago.[28] This manufacturing procedure is called rapid prototyping technology (RPT). Today also the term laser stereolithography (laser photo-polymer technology) is used, including various other manufacturing procedures.

The solid 3-D model produced by the RPT method of our reconstructed chromosome scaffold is presented with all details. Our primary 3-D reproduction had to be supported by solid bars to maintain the spatial relationship of various parts. The material, from which the model is manufactured is nontransparent and somewhat fragile. The height is 13 cm and the width and thickness are relative to the computer reconstruction. Also, most of the details, although complicated, are reproduced with comparative accuracy that is limited only by the nontransparent material that is less useful for structural studies.

The technology to reproduce solid 3-D models of 3-D reconstructions is advancing rapidly. This will bring us solid 3-D models of more transparent material with different colors to accentuate, e.g. internal structures with the same density, i.e. isosurfaces. Such hard copies will be a great aid for complex structural studies that chromosome structures particularly require.

We must admit, however, that although the impression of the original chromosome scaffold appeared as a typical chromosomal coil, and the impression of a coiling structure prevailed in the viewing of the 3-D reconstruction in the computer, the solid 3-D model could neither verify nor disprove the coiling configuration.

Actually, we can realize that this apparent conflict may arise because the scaffold is not a uniform structure but is composed of thicker and thinner segments, i.e. coiling does not look very regular at a closer look. Owing to the irregular thickness, the coil appears also dually as sequences of bands and interbands, although the bands are actually just thicker clusters of the coiling scaffold structure and the interbands are thinner regions of the coiling scaffold, as the loop-and-rosette model predicts.

The higher level structure of chromosomes is indeed very complex. Neither the 3-D reconstruction (performed with WBM) nor the solid 3-D model allowed unambiguous structural interpretation of configurations. Further, the problematic situation of chromosome coiling (cf. Figures 1f, g, h and 8a, e) resembles the level of the 30-nm chromatin fiber.[4,17] Chromosome structures at all levels are much more complicated than we think and the methods are not yet sufficiently developed.

10 METHOD DEVELOPMENT

A program and instruments to collect tilt series automatically for ETM are now available. An accurate automatic alignment procedure, at least as accurate as the fiducial gold markers, would also be very useful. The two programs would constitute the all-automatic ETM kit. The system would use a supercomputer facility in connection with a computer-controlled TEM. Not too many obstacles are likely to be in the way of such a development. Whenever possible, the cryo methods would be used on unfixed and unstained specimens embedded in vitrified ice. The cryo-TEM techniques are fast and specimens are relatively easy to prepare and fully hydrated, i.e. close to the native state. However, the problem of noisy images in cryo preparations must be solved with filtration programs. We recently succeeded in eliminating at least partly the noise from some cryo-embedded preparations, with programs that we developed. The noise problem in cryo-specimens may thus be tractable.

A viewing room and facility, such as CAVE or VR (virtual reality) CUBE (Royal University of Technology, Stockholm, Sweden), will be indispensable for thorough analyses of high-resolution 3-D reconstructions.

With chromosome structure as an example, we realize that many structures are too complex for comprehension with present 3-D reconstruction methods, including hardware solid 3-D models. Thorough analysis with methods and programs may be useful, e.g. for tracking specified structures manually, semiautomatically or automatically, in 3-D form, from section to section. Such tracking studies on whole-mounted chromosomes were started as early as 1988[45] with an automatic algorithm included in the program IMAGIC. Semiautomatic programs for this purpose are now available, such as STERECON[71] and TINKERBELL[72] and in the framework of IMOD.[66]

A new exciting method (SITUS) (a public domain program),[73] available for testing, combines protein data from high resolution X-ray crystallography, available from the Protein Data Bank, with 3-D reconstructions with ETM. The program, which makes use of neural nets, searches the known protein configuration, as 'fingerprints', in the volume and, if found, displays its localization and orientation. This method would yield an identification method for proteins, without immunological techniques, and reveal their structure in the ETM-created volume at atomic resolution.

ACKNOWLEDGMENTS

For acquainting me with the practice of electron tomography, in 1989, I thank Lisa Borland, Marin van Heel

and Professor E. Zeitler, Head, Fritz Haber Institute, Berlin. Initial tips on MEM and further instruction on tomography methods, from Docent Ulf Skoglund, his tomography group (Lars-Göran Öfverstedt and Hans Mehlin) and Professor B. Daneholt, Head, Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, are gratefully acknowledged. The visit to Karolinska Institute in 1993 was supported by EMBO and Nordiska Ministerrådet. Financial support from the foundations Maud Kuistilan Säätiö and Oskar Öflunds Stiftelse is gratefully acknowledged. At CSC-Scientific Computing, Espoo, Finland, the professional skills of Juha Ruokolainen, in both programming efficiency and physical theories, have been indispensable. Jukka Heikkonen of the Laboratory of Computational Engineering and Jorma Laaksonen of the Laboratory of Computer and Information Science, Helsinki University of Technology, have cooperated in the development and refinement of programs. For working facilities, for advice and for comments on the manuscript I thank Professor Antti Vaheri, Head, and Professor Carl-Henrik von Bonsdorff, Haartman Institute, University of Helsinki. The manuscript was thoroughly discussed with and scrutinized by Kalevi Pusa, who made a number of useful suggestions.

ABBREVIATIONS AND ACRONYMS

AA	Acetic Acid
Ab	Antibody
AD	Air Drying
CAVE	Computer-aided Virtual Environment
CCD	Charge-coupled Device
CHO	Chinese Hamster Ovary
CPD	Critical-point Drying
DAPI	4′,6-Diamidino-2-phenylindole
EM	Electron Microscopy
EMISH	Electron Microscopy In Situ Hybridization
ETM	Electron Tomography Method
FD	Freeze-drying
$FddH_2O$	Filtered Double-distilled Water
FG	Fish Gelatin
FISH	Fluorescence In Situ Hybridization
FITC	Fluorescein Isothiocyanate
GA	Glutaraldehyde
GAM	Goat Anti-mouse
GAR	Goat Anti-rabbit
HC	10 mM HEPES, pH 7.4, 5 mM $CaCl_2$
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid)
HM	10 mM HEPES, pH 7.4, 5 mM $MgCl_2$
IEM	Immunoelectron Microscopy
IET	Immunoelectron Tomography
LM	Light Microscopy
MAR	Matrix-associated Region
MEM	Maximum Entropy Method
MM	Mounting Medium
NC	Sodium Cacodylate
NE	Nuclear Envelope
NPC	Nuclear Pore Complex
o/n	Overnight
OT	Osmium Tetraoxide
PL	Poly-L-lysine
RPT	Rapid Prototyping Technology
RR	Ruthenium Red
RS	Ringer Solution
RT	Room Temperature
SAR	Scaffold-associated Region
SC	Synaptonemal Complex
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SGI	Silicon Graphics Inc.
SMC	Structural Maintenance of Proteins
TA	Tannic Acid
TEM	Transmission Electron Microscopy
TXHS	0.05% Triton X-100, 30 mM HEPES, pH 7.4, 0.1–0.5 M NaCl, 20 mM KCl, 5 mM $MgCl_2$, 20 mM glycine, 0.02% NaN_3
UA	Uranyl Acetate
WBM	Weighted Back-projection Method
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Single Biomolecule Detection and Characterization

Clinical Chemistry **(Volume 2)**
Immunochemistry • Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Fluorescence In Situ Hybridization

Peptides and Proteins **(Volume 7)**
Protein–Oligonucleotide Interactions

REFERENCES

1. W. Waldeyer, 'Über Karyokinese und ihre Beziehung zu den Befruhtungsvorgängen', *Arch. Micr. Anat.*, **32**, 1–122 (1888).
2. J.D. Watson, F.H.C. Crick, 'A Study of Deoxyribose Nucleic Acid', *Nature (London)*, **171**, 737–738 (1953).
3. P. Pasero, S.M. Gasser, 'New System for Replication DNA *in Vitro*', *Curr. Opin. Cell Biol.*, **10**, 304–310 (1998).
4. C.L. Woodcock, 'The Organization of Chromosomes and Chromatin', in *Electron Tomography: Three-dimensional Imaging with the Transmission Electron Microscope*, ed. J. Frank, Plenum Press, New York, 313–357, 1992.
5. E.J. DuPraw, *DNA and Chromosomes*, Holt, Rinehart and Winston, New York, 1970.
6. C.P. Fussel, 'The Rabl Orientation: a Prelude to Synapsis', in *Meiosis*, ed. P.B. Moens, Academic Press, New York, 275–299, 1987.
7. W.F. Marshall, A. Straight, J.F. Marko, J. Swedlow, A. Dernburg, A. Belmont, A.W. Murray, D.A. Agard, J.W. Sedat, 'Interphase Chromosomes Undergo Constrained Diffusional Motion in Living Cells', *Curr. Biol.*, **7**, 930–939 (1997).
8. P. Engelhardt, K. Pusa, 'Nuclear Pore Complexes: ''Press-stud'' Elements of Chromosomes in Pairing and Control', *Nature New Biol.*, **240**, 163–166 (1972).
9. M.E. Dresser, 'The Synaptonemal Complex and Meiosis: an Immunocytochemical Approach', in *Meiosis*, ed. P.B. Moens, Academic Press, New York, 245–274, 1987.
10. S.M. Gasser, 'Chromosome Structure, Coiling Up Chromosomes, Dispatch', *Curr. Biol.*, **5**, 357–360 (1995).
11. K. Nasmyth, 'Separating Sister Chromatids', *Trends Biochem. Sci.*, **24**(3), 98–104 (1999).
12. K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, 'Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution', *Nature (London)*, **389**, 251–260 (1997).
13. H. Ris, 'Higher Order Structures in Chromosomes', in *Proceedings of 9th International Congress on Electron Microscopy*, Canadian Society for Electron Microscopy, Toronto, 545–556, Vol. 3, 1978.
14. J.O. Thomas, 'Histone H1: Location and Role', *Curr. Opin. Cell Biol.*, **11**, 312–317 (1999).
15. R.D. Drinkwater, P.J. Wilson, J.D. Skinner, L.A. Burgoune, 'Chromatin Structures: Dissecting Their Mixed Patterns in Nuclease Digests', *Nucleic Acids Res.*, **15**, 8087–8103 (1987).
16. P.R. Walker, M. Sikorska, 'Chromatin Structure. Evidence that the 30 nm Fiber is a Helical Coil with 12 Nucleosomes/Turn', *J. Biol. Chem.*, **262**, 12223–12227 (1987)
17. R.A. Horowitz, A.J. Koster, J. Waltz, C.L. Woodcock, 'Automated Electron Microscope Tomography of Frozen-hydrated Chromatin: the Irregular Three-dimensional Zigzag Architecture Persists in Compact, Isolated Fibers', *J. Struct. Biol.*, **120**, 353–362 (1997).
18. B. Rydberg, W.R. Holley, I.S. Mian, A. Chatterjee, 'Chromatin Conformation in Living Cells: Support for a Zig-zag Model of the 30 nm Chromatin Fiber', *J. Mol. Biol.*, **284**, 71–84 (1998).
19. S.M. Gasser, B.B. Amati, M.E. Cardenas, J.F.-X. Hofman, 'Studies on Scaffold Attachment Sites and their Relation to Genome Function', *Int. Rev. Cytol.*, **119**, 57–96 (1989).
20. M.M.S. Heck, 'Condensin, Cohesins, and Chromosome Architecture: How to Make and Break a Mitotic Chromosome', *Cell*, **91**, 5–8 (1997).
21. P. Engelhardt, U. Plagens, I.B. Zbarsky, L.S. Filatova, 'Granules 25–30 nm in Diameter: Basic Constituent of the Nuclear Matrix, Chromosome Scaffold, and Nuclear Envelope', *Proc. Natl. Acad. Sci. USA*, **79**, 6937–6940 (1982).
22. P. Engelhardt, 'Higher Order Structure of Eukaryotic Chromosomes – Scaffold Elements and DNA Folding', Doctoral Dissertation, University of Helsinki, 1988. ISBN 952-90032-1-8. URL: http://www.csc.fi/jpr/emt/engelhar/Doc/Diss.html
23. H. Zentgraf, W. Franke, 'Differences of Supranucleosomal Organization in Different Kinds of Chromatin: Cell Type-specific Globular Subunits Containing Different Numbers of Nucleosomes', *J. Cell Biol.*, **99**, 272–286 (1984).
24. I. Manton, 'The Spiral Structure of Chromosomes', *Biol. Rev.*, **25**, 496–508 (1950).
25. J.B. Rattner, C.C. Lin, 'Radial Loops and Helical Coils Coexist in Metaphase Chromosomes', *Cell*, **42**, 291–296 (1985).
26. O. Haapala, 'Metaphase and Chromosome Banding are Distinct Entities of Chromosome Substructure', *Hereditas*, **100**, 75–81 (1984).
27. E. Boy de la Tour, U.K. Laemmli, 'The Metaphase Scaffold is Helically Folded: Sister Chromatids Have Predominantly Opposite Helical Handedness', *Cell*, **55**, 937–944 (1988).
28. P. Engelhardt, J. Ruokolainen, A. Dolenc, 'Electron Tomography of Chromosomes and Viruses', *CSC News*, **6**(4), 14–20 (1994). URL: http://www.csc.fi/jpr/emt/engelhar/csc/CSC-94.html
29. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, *Molecular Biology of The Cell*, Garland Publishing, New York, 1994.
30. A.S. Belmont, 'Large-scale Chromosome Structure', in *Genome Structure and Function from Chromosomes Characterization to Genes Technology*, ed. C. Nicolini, Kluwer, Dordrecht, 261–278, 1996.
31. P.R. Cook, 'Commentary a Chromomeric Model for Nuclear and Chromosome Structure', *J. Cell Sci.*, **108**, 2927–2935 (1995).
32. P. Hozak, A.M.-J. Sasseville, Y. Raymond, P.R. Cook, 'Lamin Proteins form an Internal Nucleoskeleton as Well as a Peripheral Lamina in Human Cells', *J. Cell Sci.*, **108**, 635–644 (1995).

33. J. Newport, H. Yan, 'Organization of DNA into Foci During Replication', *Curr. Opin. Cell Biol.*, **8**, 365–368 (1996).
34. L. Reimer, *Transmission Electron Microscopy, Physics of Image Formation and Microanalysis*, Spinger Series in Optical Sciences, Springer, Berlin, 1997.
35. S. Hovmöller, 'CRISP: Cristallographic Image Processing on a Personal Computer', *Ultramicroscopy*, **41**, 121–135 (1992).
36. J. Frank (ed.), *Electron Tomography: Three-dimensional Imaging with the Transmision Electron Microscopy*, Plenum Press, New York, 1992.
37. D.J. De Rosier, A. Klug, 'Reconstruction of Three Dimensional Structures from Electron Micrographs', *Nature (London)*, **217**, 130 (1968).
38. R. Hart, 'Electron Microscopy of Unstained Biological Material: The Polytropic Montage', *Science*, **159**, 1464–1467 (1968).
39. W. Hoppe, R. Langer, G. Knesch, C. Poppe, 'Protein-Kristallstrukturanalyse mit Elektronenstrahlen', *Naturwissenschaften*, **55**(7), 333–336 (1968).
40. I. Daberkow, A.J. Koster, H.R. Tietz, D. Typke, J. Walz, 'A System for Automated Electron Tomography using Philips CM Series Transmission Electron Microscopes', *Philips Electron Opt. Bull.*, **134**, 27–31 (1996).
41. U. Skoglund, L.-G. Öfverstedt, R.M. Burnett, G. Bricogne, 'Maximum-entropy Three-dimensional Reconstruction with Deconvolution of the Contrast Transfer Function: a Test Application with Adenovirus', *J. Struct. Biol.*, **117**, 173–188 (1996).
42. J. Frank, *Three-dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, New York, 1996.
43. M.C. Lawrence, M.A. Jaffer, B.T. Sewell, 'The Application of the Maximum Entropy Method to Electron Microscopic Tomography', *Ultramicroscopy*, **31**, 285–302 (1978).
44. G. Harauz, L. Borland, G.F. Bahr, E. Zeitler, M. van Heel, 'Three Dimensional Reconstruction of a Human Metaphase Chromosome from Electron Micrographs', *Chromosoma*, **95**, 366–374 (1987).
45. L. Borland, G. Harauz, G. Bahr, M. van Heel, 'Packing of the 30 nm Chromatin Fiber in the Human Metaphase Chromosome', *Chromosoma*, **97**, 159–163 (1988).
46. A.J. Koster, D.A. Agard, 'Editorial: Electron Tomography Workshop', *J. Struct. Biol.*, **120**, 207–209 (1996).
47. P. Engelhardt, J. Ruokolainen, F. Zhao, O. Carpén, A. Vaheri, 'Immunoelecron Tomography (IET) using Maximum Entropy Method Reveals 1.4 nm Gold Conjugates in 3-D Reconstruction of MCF-7 Cell Microvilli', in *SCANDEM-98, Proceedings of the 50th Annual Meeting of the Scandinavian Society of Electron Microscopy*, Scandem, Espoo, Finland, 86–87, 1998. URL: http://www.csc.fi/jpr/emt/engelhar/IET/Scandem-98-WWW.html
48. C.V. Hanson, 'Techniques in Isolation and Fractionation of Eukaryotic Chromosomes', in *New Techniques in Biophysics and Cell Biology*, eds. R.H. Pain, B.J. Smith, John Wiley & Sons, New York, 243–283, 1975.
49. J. Mendelsohn, 'Studies of Isolated Mammalian Metaphase Chromosomes', in *The Cell Nucleus*, ed. H. Busch, Academic Press, New York, 123–147, Vol. II, 1974.
50. G.F. Bahr, 'Determination of the Dry Mass of Biological Objects by Quantitative EM', in *Micromethods in Molecular Biology*, ed. V. Neuhoff, Springer, Berlin, 257–284, 1973.
51. A.J. Solari, 'Structural Analysis of Meiotic Chromosomes and Synaptonemal Complex in Higher Verteberates', *Methods Cell Biol.*, **53**, 235–256 (1997).
52. B.F. McEwen, M. Marko, 'Three-dimensional Transmission Electron Microscopy and its Application to Mitosis Research', *Methods Cell Biol.*, **61**, 81–111 (1999).
53. O.L. Miller, Jr, A.H. Bakken, 'Morphological Studies of Transcription', in *Karolinska Symposia on Research Methods in Reproductive Endocrinology, 5th Symposium on Gene Transcription Reproductive Tissue*, Karolinksa Institute, Stockholm, 155–173, 1972.
54. P. Engelhardt, J. Ruokolainen, 'Electron Tomography of Acetic Acid Isolated Whole-mount Chromosomes and Chromosome Scaffolds Suggests a Chromatin–Scaffold–Fiber Complex', in *EUREM-11, Proceedings of the 11th European Congress on Electron Microscopy (Dublin, Ireland, 26–30 August, 1996), Biology*, CESM, Brussel, 32–33, Vol. 3, 1998. URL: http://www.csc.fi/jpr/emt/engelhar/EUREM-96.html
55. L. Wollweber, R. Stracke, U. Gothe, 'The use of Simple Method to Avoid Cell Shrinkage During SEM Preparation', *J. Microsc.*, **121**(2), 185–189 (1980).
56. H. Ris, 'The Cytoplasmic Filament System in Critical Point-dried Whole Mount and Plastic-embedded Sections', *J. Cell Biol.*, **100**, 1474–1487 (1985).
57. K.R. Porter, M.E. Stearns 'Stereomicroscopy of Whole Cells', *Methods Cell Biol.*, **22**, 53–75 (1981).
58. T. Inoué, H. Osatake, H. Takahashi, 'A New Freeze-drying Instrument using t-Butyl Alcohol for Scanning Electron Microscopy', *J. Electron Microsc.*, **38**(4), 246–249 (1989).
59. J. Heuser, 'Preparing Biological Samples for Stereomicroscopy by the Quick-freeze, Deep-etch, Rotary-replication Technique', *Methods Cell Biol.*, **22**, 97–122 (1981).
60. M.V. Nermut, A.C. Steven (eds.), *Animal Virus Structure*, Elsevier, Amsterdam, 1987.
61. M.A. Hayat, *Negative Staining*, McGraw-Hill, New York, 1990.
62. M.W. Clark, 'Immunogold Labeling of Yeast Ultrathin Sections, Appendix B', *Methods Enzymol.*, **194**, 608–626 (1991).
63. S. Penman, 'Rethinking Cell Structure', *Proc. Natl. Acad. Sci. USA*, **92**, 5251–5257 (1997).
64. J.N. Turner, U. Valdrè, 'Tilting Stages for Biological Applications', in *Electron Tomography: Three-dimensional Imaging with the Transmission Electron*

Microscope, ed. J. Frank, Plenum Press, New York, 167–194, 1992.
65. *Model 630 High Tilt Cryotransfer*, Gatan Inc., Pleasanton, CA. URL: http://www.gatan.com
66. J.R. Kremer, D.N. Mastronarde, J.R. MacIntosh, 'Computer Visualization of Three-dimensional Image Data using IMOD', *J. Struct. Biol.*, **116**, 71–76 (1996).
67. G. Grifiths, *Fine Structure Immunochemistry*, Spinger, Berlin, 1995.
68. J. Ruokolainen, P. Engelhardt, 'Maximum-entropy an Electron-tomography Program Kit for Silicon Graphics Computers', in *3-D Imaging Sciences in Microscopy, 15–17 April, Conference Proceedings*, Royal Microscopical Society, London, 1996.
69. P. Penczek, M. Marko, K. Buttle, J. Frank, 'Double-tilt Electron Tomography', *Ultramicroscopy*, **60**, 393–410 (1995).
70. D.N. Mastronarde, 'Dual-axis Tomography: an Approach with Alignment Methods that Preserve Resolution', *J. Struct. Biol.*, **120**, 343–352 (1997).
71. M. Marko, A. Leith, 'Sterecon – Three-dimensional Reconstruction from Stereoscopic Contouring', *J. Struct. Biol.*, **116**, 93–98 (1996).
72. Y.-H. Li, A. Leith, J. Frank, 'Tinkerbell – A Tool for Interactive Segmentation of 3-D Data', *J. Struct. Biol.*, **120**, 266–275 (1997).
73. W. Wriggers, R.A. Milligan, J.A. McCammon, 'Situs: a Package for Docking Crystal Structures into Low-resolution Maps from Electron Microscopy', *J. Struct. Biol.*, **125**, 185–195 (1999).

Fluorescence In Situ Hybridization

Thomas Haaf
Max-Planck-Institute of Molecular Genetics, Berlin, Germany

1 Introduction **2**
2 Equipment for Digital Imaging Microscopy **2**
3 DNA Probes **3**
3.1 Satellite and Localized Repeat DNAs 3
3.2 Chromosome-specific DNA Libraries 3
3.3 Region-specific Large-insert Clones 3
3.4 Total Genomic DNAs 3
3.5 cDNAs 5
4 Probe Labeling **5**
4.1 Nick Translation 5
4.2 Polymerase Chain Reaction 7
5 In Situ Suppression Hybridization (Standard Fluorescence In Situ Hybridization Protocol) **8**
5.1 Hybridization Mixture 8
5.2 Slide Preparation 9
5.3 Hybridization and Signal Detection 9
6 Multicolor Fluorescence In Situ Hybridization **10**
7 Comparative Genomic Hybridization **13**
8 (Ultra) High-resolution Fluorescence In Situ Hybridization **13**
8.1 Fiber Fluorescence In Situ Hybridization 15
8.2 Molecular Combing 16
9 Important Applications **17**
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **19**
Appendix 1 **19**
Buffers and Stock Solutions 19
References **21**

Fluorescence in situ hybridization (FISH) is a technique used to directly visualize specific DNA sequences on morphologically preserved cytological specimens such as metaphase chromosomes, interphase cell nuclei, and extended chromatin fibers or DNA molecules. The general usefulness of radioactive in situ hybridization was greatly hampered by the safety measures required when using radiolabeled probes and the rather long time (up to several weeks) needed for autoradiographic analysis. The development of powerful FISH protocols, in which unspecific hybridization of interspersed repetitive sequences contained within genomic DNA probes is suppressed, allowed for the first time the rapid visualization of single-copy sequences and whole chromosomes, and thus marked a turning point in the experimental and diagnostic application of in situ hybridization. The major advantages of FISH over radioactive methods are increased spatial resolution, speed, probe stability, and the ability to detect multiple chromosomal targets in different colors. Indirect FISH uses hapten-labeled DNA probes in combination with immunocytochemistry (secondary detection reagents). Direct methods use fluorochromes directly coupled to nucleotides in the probe DNA. Probe labeling is most easily accomplished by nick translation or other enzymatic labeling procedures. For hybridization of the modified probe DNA to chromosomes and cell nuclei, both probe and target DNA must be denatured. In the hybridization reaction, complementary single-stranded sequences in the probe and chromosomal target are allowed to reanneal. After posthybridization washing and immunocytochemical detection of in situ bound hapten molecules, a specific fluorescent signal is produced at the hybridization site, which can be viewed by eye through the epifluorescence microscope. Digital imaging and computerized storage of images have largely replaced photography and greatly facilitated data analysis and image handling. However, most importantly the success of FISH experiments depends on the quality of the chromosome preparations and the size of the chromosomal DNA target. When large-insert (>30 kb (kilobase)) clones or complex DNA libraries are used as FISH probes, between 80% and 100% of the target chromosomes carry visible hybridization signals. This percentage drops to only a few percent when 1 kb or less of a single-copy sequence is hybridized. In the past few years several technical advances have expanded FISH applications. Comparative genomic hybridization (CGH) of differentially labeled tester (tumor or patient) and normal reference DNA allows the analysis of all genetic imbalances (chromosomal gains and losses) in a tester genome within a single experiment and without the need to prepare chromosomes from the tester. Combinatorial probe labeling and multicolor FISH are used for the simultaneous detection of 24 (or more) chromosomal targets in different colors, and therefore can provide a comprehensive picture of extensively rearranged tumor karyotypes. The DNA resolution of FISH has been dramatically increased (from several megabases (Mbs) to several kilobases) by hybridization to

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

extended chromatin fibers and linearized DNA molecules. Fiber FISH and molecular combing allow one to order probes relative to each other, to orient probes within a contig, and to determine the degree of overlap or gap size between different probes. Overall, FISH has become an increasingly popular method with a broad spectrum of applications in cytogenetics, genomics, tumor biology, and many other research areas. It has a bright and colorful future.

1 INTRODUCTION

FISH has become an extremely powerful and widely used experimental method over the past 10–15 years, and has created an entirely new field, molecular cytogenetics. It can be used to localize specific DNA sequences on chromosomes, interphase nuclei, and experimentally extended chromatin fibers or DNA molecules. In situ hybridization techniques were originally developed in the late 1960s,[1,2] at a time when only radioactive labels were available for nucleic acids and cloned DNA probes did not yet exist. Enzymatic labeling of nucleic acids with biotin[3] and other haptens offered several important advantages over radioactive in situ hybridization. Nonisotopically labeled DNA sequences are stable for years and the hybridized probes can be evaluated directly at the microscope without autoradiography. This not only significantly shortens analysis time and avoids hazardous radioactive waste, but there is also something very satisfying about a fluorescing image. The advance of molecular cloning and, more recently, the launch of the human genome project have led to an explosion in the number of DNA probes that are available for in situ hybridization to visualize a chromosome or chromosome region of interest. In addition, FISH applications have been greatly expanded by the development of user-friendly microscope systems that use digital imaging for data acquisition, analysis, and storage.

The principle behind in situ hybridization (and many other molecular techniques) is the ability of complementary single-stranded DNA sequences to specifically reanneal, forming double-stranded DNA (hybrid) molecules. The probe DNA molecules are enzymatically labeled with modified nucleotides. The hapten- or fluorescent-labeled DNA molecules then bind to chromosomes and cell nuclei at the site of specific (homologous) DNA sequences by in situ hybridization. In contrast to the hybridization of nucleic acids in solution, the target DNA sequences are organized by histone and nonhistone proteins into the highly condensed (2000–10 000-fold) interphase or metaphase chromosome structure. This high degree of compaction involving a hierarchy on different levels hinders access of the probe DNA and the hybridization reaction. Therefore, at the theoretical level, many aspects of in situ hybridization procedures still have a largely empirical basis. Briefly, probe DNA and target chromosomes are denatured and the resulting single-stranded DNA molecules are then allowed to reanneal. After various washing and detection steps, a specific fluorescence signal is visible at the chromosomal site of probe hybridization.

Overall, FISH is a surprisingly simple and robust technique which has applications in many different research areas, including clinical cytogenetics, gene mapping and genomics, evolution, and tumor biology. Although FISH methodology is used predominantly for the study of the human genome, FISH techniques have been successfully adapted for use in the study of a wide variety of genomes, from those of yeast to man. In this light, the following protocols may be useful not only for human cytogeneticists but for a wide variety of chromosome researchers.

2 EQUIPMENT FOR DIGITAL IMAGING MICROSCOPY

A typical FISH workstation consists of the following components.[4]

1. A relatively simple upright epifluorescence microscope with objectives, epifluorescence filter sets, and excitation source. Switching between different fluorochromes is achieved by changing filter sets. In routine FISH work, cross-talk between blue, green, and red fluorochromes (i.e. 4′,6-diamidino-2-phenylindole (DAPI), fluorescein isothiocyanate (FITC), and Cy3) imaged with their specific filter sets should be less than 10%. Longpass filters produce brighter single-fluor images but poorer spectral discrimination between fluorochromes. The most commonly used light source that will excite fluorescent dyes from the ultraviolet (UV) into the red region of the spectrum is the high-pressure mercury arc. The projection optics form a real image at the focal plane of the camera.
2. Thermoelectronically cooled charge-coupled device (CCD) cameras with medium format CCD chips are recommended for FISH imaging. A host computer is needed for camera control, image acquisition, image processing, image display, and storage of data. Digital imaging benefits from a very fast computer with 40–60 Mb of RAM (random access memory). Large image files require a hard disk capacity of about 2 GB and a back-up system.
3. Software packages for image acquisition, processing, and data analysis are available from different

companies, i.e. Applied Imaging, Leica Microsystems Imaging, Meta Systems, Oncor, PSI, Vysis, and others. Since FISH software packages can be quite costly and usually not all of the functions are needed in a routine laboratory, it is strongly recommended that software is tested in the lab using typical chromosome preparations and experimental conditions. Typically a series of images is taken through different filters. These gray-scale source images have to be merged and pseudo-colored. Sometimes the image background needs to be cleaned up and/or the specific signal needs to be enhanced. For gene mapping, it is very useful to convert DAPI fluorescence of a chromosome into a G-banding pattern for chromosome identification. Furthermore, it should be possible to measure the chromosomal location (fractional chromosome length) of FISH signals and to (manually) arrange chromosomes into a karyotype.

3 DNA PROBES

The various applications of FISH would not be possible without the general availability of DNA probes for specific chromosomes or chromosome regions of interest. Today a large number of different types of probes for human and other important model genomes are available, either commercially or from various genome resource centers and research laboratories. The choice depends mainly on the specific requirements of the experiments and on the laboratory expertise in molecular cytogenetics. For example, satellite DNAs are ideal probes for beginners, because they do not require suppression hybridization, and therefore are much easier to use than genomic probes containing interspersed repetitive elements. The efficiency of hybridization and the fluorescence at the hybridization site largely depend on the size of the target DNA sequences on the chromosome. Hybridization sites spanning several kilobases should be detectable without sophisticated equipment.

3.1 Satellite and Localized Repeat DNAs

Satellite DNA families are based on tandemly repeated monomers or higher-order repeating structures consisting of diverged monomers. Since satellite-DNA-containing chromosomal regions are not interspersed with other cross-hybridizing sequences, and usually span several hundred kilobases to several megabases, they are easily detected by FISH without suppression hybridization. This means that the prehybridization step (see standard FISH protocol in section 5) can be omitted. Chromosome-specific satellite DNAs or rRNA gene clusters produce very intense and compact signals in both metaphase and interphase cells (Figure 1d, g). One of the most thoroughly studied repetitive DNA families is α-satellite DNA, which is found at the centromeric regions of all human and primate chromosomes.[5–7] Chromosome-specific α-satellite subsets have been described for the majority of human chromosomes and are widely used in routine cytogenetic laboratories, i.e. for aneuploidy studies.

3.2 Chromosome-specific DNA Libraries

DNA libraries represent complex mixtures of DNA sequences that are derived from a single chromosome type and, therefore, specifically paint the target chromosome by FISH (Figure 1c, e; Figure 2a). Whole-chromosome painting probes of high quality have been constructed from flow-sorted human and other mammalian chromosomes[8–10] and by chromosome microdissection.[11] In addition, microdissection has been used to generate region-specific DNA libraries.[12]

3.3 Region-specific Large-insert Clones

Large-insert probes, such as cosmids, λ phages, bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs), and YACs, produce high-intensity site-specific FISH signals on 80–100% of the hybridized chromosomes and cell nuclei. Our lab has created a large set of (to date) 2000 YAC probes with approximately one YAC every 3–5 cM (centimorgan) spaced evenly over the entire human genome covering all human chromosome bands.[13] These YACs have been well characterized by polymorphic genetic markers and FISH-mapped on human metaphase chromosomes. This standard set of cytogenetically and genetically anchored YACs allows the rapid integration of genetic and cytogenetic maps. It is an extremely flexible tool for use in a wide variety of FISH applications, such as the study of pathological and evolutionary chromosome rearrangements (Figure 1a–e; Figure 2b, c). Similar molecular cytogenetic resources have been generated for the mouse and other experimental animals.[14,15] For most FISH experiments, purification of the clone insert is not necessary. Labeled vector sequences in the DNA probe do not hybridize appreciably to human chromosomes. Nonspecifically formed (vector–chromosome) hybrid molecules are dissociated during the posthybridization washings (see section 5.3).

3.4 Total Genomic DNAs

In situ hybridization with total (genomic) nuclear DNA was first applied in plant cytogenetics for identifying parental chromosomes in interspecific plant hybrids and

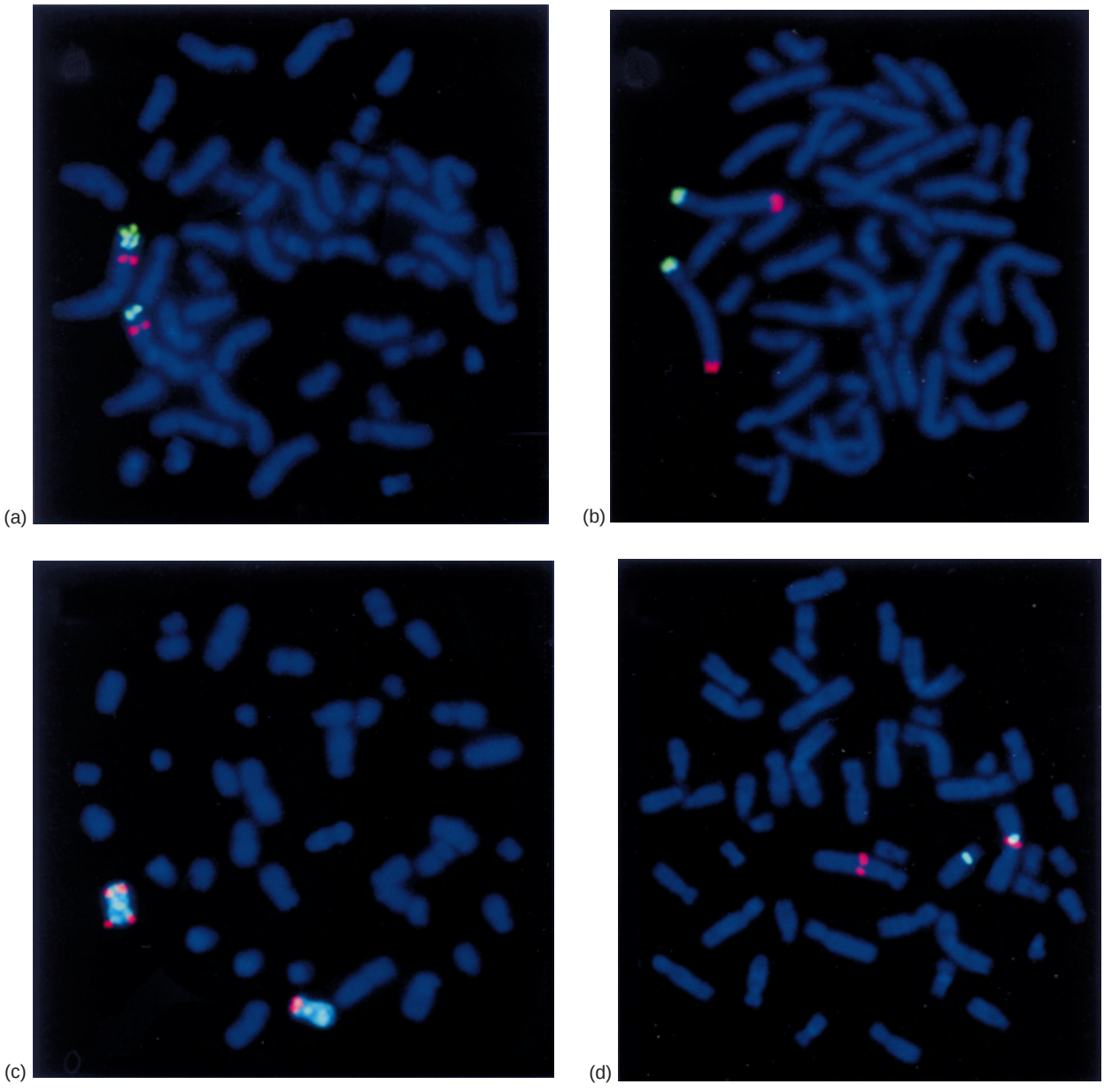

Figure 1 Clinical cytogenetic applications of FISH: detection of structural and numerical chromosomal aberrations. (a) Triplication of subtelomeric sequences in the long arm of chromosome 1. The red yeast artificial chromosome (YAC) produces similar hybridization signals on both chromosomes. The green YAC produces three double-dot hybridization signals on the aberrant chromosome. (b) Hybridization of YACs specific for the subtelomeric regions on the short (green) and long (red) arms of chromosome 6, respectively. (c) Paracentric inversion on the X chromosome of a female patient. Both X's are painted with a chromosome-specific DNA library (green). The red YAC produces a single hybridization signal on the short arm of the normal X and a split signal on the long and short arms of the inverted X. (d) Balanced translocation between chromosomes 2 and 12. The green α-satellite probe hybridizes to the centromere of the normal chromosome 2 and the translocated chromosome der(2). The red YAC hybridizes to the proximal long arm of the normal chromosome 2 and to the translocated chromosome der(12). The hybridized probes flank the breakpoint on chromosome 2. (e) The same X inversion as seen in (c) during interphase. Note the split YAC hybridization signal in the upper X-chromosomal domain. (f) Asynchronous replication of subtelomeric sequences. The green 4qter YAC produces two hybridization doublets, indicating that the delineated homologous segments are fully replicated. The red 10qter YAC produces one unreplicated singlet and one replicated doublet hybridization signal, consistent with asynchronously replicating homologous chromosome regions. (g) Hybridization of five differentially labeled α-satellite DNAs specific for chromosomes 5, 6, 8, 20, and X on a normal female interphase nucleus.

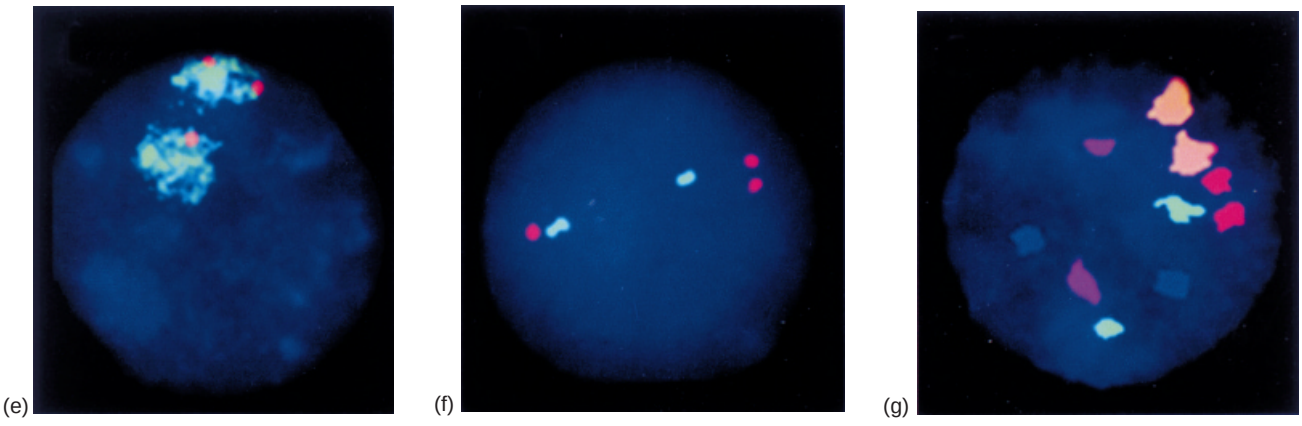

Figure 1 (*Continued*)

following the introgression of alien chromosomes.[16,17] More recently comparative hybridization of genomic tumor DNA versus normal reference DNA on normal human metaphase spreads has been used to assess chromosome gains and losses in tumor tissues.[18]

3.5 cDNAs

Complementary DNAs (cDNAs) represent only the transcribed (exonic) sequences of genes and contain as little as several hundred base pairs (bps) of target sequence. The hybridization efficiency is low and depends on both cDNA size and intron/exon structure of the targeted gene. FISH mapping of cDNAs and small-insert (<1 kb) plasmid clones requires special expertise and/or digital imaging microscopy.

4 PROBE LABELING

Two major alternatives exist for fluorescent hybridization methods: direct and indirect. In the direct method, fluorescent reporter molecules are incorporated directly into the DNA probe. A wide variety of fluorescent deoxynucleotide triphosphates (dNTPs) can be used for direct labeling, i.e. FITC-, Texas Red-, Rhodamine-, Cy3-dUTP, and many more. Since the probe–target DNA hybrids resulting from in situ hybridization are immediately visible under the microscope, direct methods are very fast. They offer other advantages, such as low background fluorescence and increased possibilities for multicolor analysis. On the other hand fluorescent probes are prone to photobleaching during preparation and hybridization. In the indirect method, hapten reporter molecules, i.e. biotin- or digoxigenin-dUTP, which are introduced into the DNA probe, are detected by affinity cytochemistry after the hybridization reaction. Fluorescent (strept)avidin molecules bind specifically to biotin-labeled probe–target hybrids. Fluorescent anti-digoxigenin antibodies are used to detect digoxigenated probes. The fluorescence signals of indirectly labeled probes may be up to ten times brighter than those produced by direct methods. In addition, small hapten molecules such as biotin are more efficiently incorporated into DNA than most fluorescent dNTPs. Thus, indirect labeling may be preferable for most FISH applications. Haptenized and, to a lesser extent, fluorescent probes can be stored at −20 °C for many years without compromising performance.

Different enzymatic labeling procedures can be used to introduce modified dNTPs into DNA probes, i.e. nick translation, random priming, and polymerase chain reaction (PCR). Nick translation[19] is recommended for labeling double-stranded DNA molecules because large amounts of DNA can be labeled and most FISH experiments require high probe concentrations. Random priming[20] has the advantage that small amounts of DNA and single-stranded probes can be efficiently labeled. PCR is a very convenient and fast method for labeling PCR-amplifiable large-insert probes and DNA libraries.

4.1 Nick Translation

The DNA template can be supercoiled or linear. After nicking the DNA with DNase I, the 5′–3′ exonuclease activity of DNA polymerase removes nucleotides and the DNA polymerase activity replaces the excised nucleotides with dNTPs from the reaction mixture, including the labeled nucleotide. The procedures for incorporating biotin, digoxigenin, or fluorochromes are nearly identical. The kinetics of the reaction reaches an optimum after 30–60 min. The final size of the nick-translated probe DNA fragments is very important. Labeled probe sequences which are larger than 500 bp have difficulty penetrating the specimen and a tendency

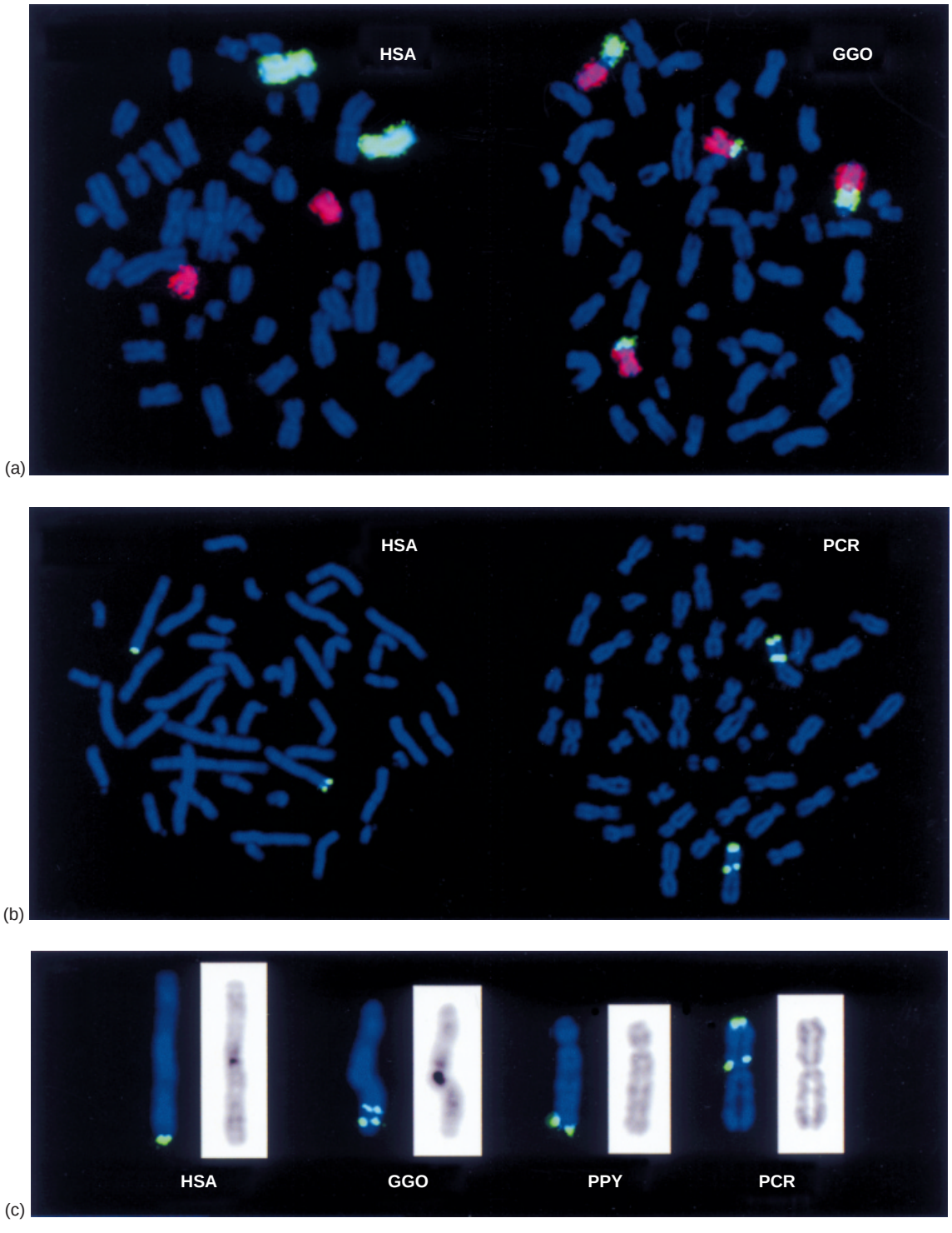

Figure 2 Comparative genome mapping by FISH. (a) Hybridization of painting probes specific for human chromosome 5 (green) and 17 (red) on human (HSA) and gorilla (GGO) metaphase spreads. Note the evolutionary rearrangement (translocation) involving the gorilla homologs of HSA 5 and 17. (b) Hybridization of a subtelomeric YAC from the long arm of human chromosome 3 on the Old World monkey *Presbytis cristata* (PCR). Note the split hybridization signal on the monkey homolog of HSA 3. The hybridized YAC spans an evolutionary chromosome breakpoint. (c) Hybridization of the same YAC on great ape and monkey homologs of HSA 3. Orangutans (PPY) and humans show the same hybridization patterns. In gorilla and Old World monkey the YAC delineates paracentric inversions involving the same distal but different proximal breakpoints. This indicates the occurrence of convergent chromosomal mutations. Insets show converted DAPI images (G bands).

to stick nonspecifically to both the glass and the cellular material, resulting in high background which obscures the signal. Sequences shorter than 100 bp do not hybridize efficiently under routine conditions of stringency.

Mix on ice, in a microcentrifuge tube:

DNA (in liquid format)	1 μg
10 × nick-translation buffer (see Appendix 1)	5 μL
0.1 M β-mercaptoethanol (BME) (see Appendix 1)	5 μL
10 × cold dNTP solution (0.5 mM each of dATP, dCTP, dGTP)	5 μL
10 × 1 : 3 labeled dUTP/cold dTTP solution (0.125 mM labeled dUTP, i.e. biotin-16-dUTP or digoxigenin-11-dUTP, and 0.375 mM dTTP) A labeling density of one modified nucleotide at approximately every 20–25th position in the nick-translated DNA is optimal for most FISH experiments.	5 μL
10 × DNase I (DNase stock with an activity of 10 U μL^{-1} is diluted 1 : 3000–1 : 4000 with ddH_2O (double-distilled water)) Optimal DNase concentration must be determined by titration, as the size distribution of the probe depends on the amount of enzyme added. Each new stock of DNase I should be tested to determine the appropriate working concentration.	5 μL
DNA polymerase I (5 U μL^{-1})	1.3 μL
ddH_2O	Add as needed to achieve final reaction volume of 50 μL

If larger amounts of DNA (up to 5 μg) need to be labeled, the reaction volumes can be scaled up to 250 μL.

Incubate the reaction mixture for 2 h at 15 °C. The time of incubation should not be shortened; at least 2 h are necessary for optimal incorporation, as the initial incorporation rate of modified nucleotides is less than that of unsubstituted dNTP.

Check the size of the nick-translated DNA fragments on a 1.5% agarose gel, along with suitable size markers (0.1–1 kb range). If the DNA fragments are still too large, add a second aliquot of DNase (optional) and incubate for another 30–60 min at 15 °C.

When the majority of the fragments are between 100 bp and 500 bp in size, terminate the reaction by adding stop buffer (1.25 μL 0.5 M EDTA (ethylenediamine tetraacetic acid) and 0.5 μL 10% SDS (sodium dodecyl sulfate) (see Appendix 1) to the 50 μL of nick-translation mixture. Heat tube to 68 °C for 10 min.

For most types of probes, i.e. satellite DNAs, DNA libraries or large-insert clones, purification of the labeled probe DNA is not necessary. Simple ethanol precipitation is sufficient. For small single-copy DNA probes, purification by centrifugation through Sephadex G-50 or commercially available columns is recommended to remove unincorporated haptenized or fluorescent nucleotides.

4.2 Polymerase Chain Reaction

Degenerate oligonucleotide primed (DOP)[21] and inter-Alu[22] PCR are widely used to amplify complex DNA probes from large-insert clones, i.e. PACs and YACs, and flow-sorted or microdissected DNA libaries. Alu PCR amplifies only sequences located between Alu repeats in the human insert, whereas DOP PCR amplifies vector (bacterial or yeast) sequences along with the entire insert. However, since DOP PCR products are much more complex than Alu PCR products, they usually generate better FISH signals. PCR-generated probes can be efficiently labeled by re-amplification of the DOP or Alu PCR product, in reactions where dTTP is partially replaced by modified dUTP. Since optimal PCR conditions are largely dependent on the DNA template and primer(s), the DOP PCR labeling protocol described here is, by necessity, generalized.

Mix on ice, in a centrifuge tube:

Template DNA (in liquid format)	200–400 ng
10 × *Taq* polymerase buffer (see Appendix 1)	5 μL
25 mM $MgCl_2$	4 μL
20 pM μL^{-1} DOP primer (5′-CCG ACT CGA GNN NNN NAT GTG G-3′)	20 μL
2 mM dATP	3.75 μL
2 mM dCTP	3.75 μL
2 mM dGTP	3.75 μL
2 mM dTTP	1.25 μL
1 mM biotin-16-dUTP (or other labeled dUTP)	0.5 μL
Taq polymerase	0.5 μL
ddH_2O	Add as needed to achieve final reaction volume of 50 μL

After an initial denaturation at 94 °C for 1 min, 30 cycles are carried out with denaturation at 94 °C for 1 min, annealing at 56 °C for 1.5 min, and extension at 72 °C for 2 min followed by incubation at 72 °C for 10 min.

5 IN SITU SUPPRESSION HYBRIDIZATION (STANDARD FLUORESCENCE IN SITU HYBRIDIZATION PROTOCOL)

Ubiquitously repeated sequences, i.e. of the Alu and L1 families, occur approximately 900 000 and 100 000 times in the human genome respectively.[23] Since interspersed repeats are present at least once in every 5 kb of genomic DNA, complex DNA libraries, large-insert clones, and most other DNA probes contain a significant number of repetitive elements that can cross-hybridize with closely related repeats throughout the entire genome. Consequently, these probes will not only hybridize to their specific target DNA sequences but to all chromosomes, yielding ambiguous FISH signals. To prevent such unspecific hybridization, repetitive DNA sequences in the probe DNA are blocked by prehybridization with unlabeled total genomic DNA or repetitive Cot-1 DNA.[24–26] The fast reassociating (Cot-1) fractions of genomic DNA probes reanneal under the appropriate (pre)hybridization conditions, and thus mainly single-copy DNA sequences in the hybridization mixture remain single-stranded and can hybridize to their target sequences on chromosomes. The following is a detailed protocol for suppression hybridization that has been optimized for the fluorescent detection of single-copy DNA sequences.

5.1 Hybridization Mixture

Nick-translated DNA probes have a final concentration of approximately 20 ng μL^{-1} (in nick translation buffer), PCR-labeled probes one of approximately 100 ng μL^{-1} (in *Taq* polymerase buffer). First, differentially labeled probes which are to be hybridized on the same specimen are precipitated together with an approximately 50-fold excess of Cot-1 competitor DNA (Gibco BRL) or a 500-fold excess of fragmented (100–500 bp) total genomic DNA, and a 50-fold excess of fragmented salmon sperm DNA (see Appendix 1) which is used as a carrier. The optimal DNA concentrations in the hybridization mixture (30 µL for a whole slide) depend mainly on the probe types. For large-insert PAC and YAC clones, 10–20 ng μL^{-1} in the hybridization mixture (or 300–600 ng per slide) are recommended. For small-insert plasmid clones and cDNAs, higher probe DNA concentrations of 20–50 ng μL^{-1} may be superior. For very complex (painting) probes, such as chromosome- or arm-specific DNA libaries, 5–10 ng μL^{-1} in the hybridization mixture (or 150–300 ng per slide) usually yield good results. The probe concentration in the hybridization mixture affects the rate at which the first few base pairs are formed (nucleation reaction). The adjacent base pairs are formed afterwards. Since the nucleation reaction is the rate-limiting step in hybridization, the higher the concentration of the probe, the higher the reannealing rate.

As a rule, the higher the proportion of repetitive elements present in the probe DNA, the more the competitor DNA that is needed. Probes containing mainly interspersed repeats and only few single-copy sequences are difficult to hybridize and yield inferior FISH signals. As might be expected, if too little competitor DNA is added, background signals from the interspersed repetitive elements will not be effectively suppressed. Vice versa, the addition of too much competitor DNA will reduce the signal intensity as well as the background.

In the following example, a YAC and a chromosome-specific DNA library will be hybridized together on the same slide.

	Stock	Amount required
Biotinylated (nick-translated) YAC DNA	20 ng μL^{-1}	20 µL (400 ng)
Digoxigenated PCR-labeled) DNA library	100 ng μL^{-1}	2 µL (200 ng)
Cot-1 DNA (Gibco BRL)	1 µg μL^{-1}	30 µL (30 µg)
Salmon sperm DNA	10 µg μL^{-1}	3 µL (30 µg)

Precipitate the DNA as usual with 1/20 volume 3 M NaAcetate (see Appendix 1) and 2 volumes 100% EtOH (p.a.) in a microcentrifuge tube. Mix well and keep the tube for at least 30 min at −70 °C or overnight at −20 °C.

Spin for 30 min at 15 000 rpm at 4 °C. Discard the EtOH and dry the pellet in a vacuum system or a 37 °C heating block.

Add 15 µL of deionized formamide (see Appendix 1) and shake at 37 °C for at least 15 min to resuspend the probe DNA. Then add an equal volume of 2 × hybridization mixture (20% dextran sulfate, 4 × SSC (standard saline citrate)) (see Appendix 1) and shake vigorously for at least 15 min. Dextran sulfate is very hydrophilic. The consequence is that macromolecules such as DNA have no access to the hydrating water, which leads to an apparent increase in probe concentration and consequently higher hybridization rates.

Routine conditions of stringency are 50% formamide, 10% dextran sulfate, 2 × SSC in the hybridization mixture. Occasionally, stringency conditions may need to be increased (e.g. for highly repetitive satellite DNAs or if a single-copy probe also cross-hybridizes weakly to a second site) or decreased (e.g. if trying to hybridize with a very short oligonucleotide probe or with a complex probe from a related species). Hybridization stringency can be increased by increasing the incubation temperature or the formamide concentration, or by decreasing the salt (SSC)

concentration in the hybridization mixture. Usually, stringency is increased by using 60–70% formamide in the hybridization mixture, or decreased by using only 20–30% formamide.

Denature the hybridization mixture (containing the probe DNA) at 80 °C for 10 min. Centrifuge the probes for 2–5 s in order to pellet the condensed water.

To allow reannealing of repetitive DNA fractions, incubate the tube with the denatured competitor and probe DNA at 37 °C for at least 15 min (but not longer than 1 h).

Place the preannealed DNA probe (30 μL) on the denatured slide and cover with a coverslip. Seal the edges with rubber cement and incubate overnight (or longer) at 37 °C in a moist chamber.

5.2 Slide Preparation

The quality of the chromosome preparations is a very critical parameter for the efficiency of FISH. Metaphases can be prepared from any (exponentially) growing cell culture, i.e. short-term blood cultures, bone marrow, fibroblasts, lymphoblastoid cell lines, and tumor cells.[27] The most widely used procedure for the preparation of high-quality chromosomes involves phytohemagglutinin stimulation of T lymphocytes of human peripheral blood in culture and metaphase arrest by adding the spindle poison colcemid shortly before harvesting. Harvested cells are swollen in hypotonic solution (50 mM KCl at 37 °C for 5–10 min) and fixed in a 3 : 1 mixture of ice-cold methanol–acetic acid. Chromosome suspensions can be stored at −20 °C for several months without compromising performance.

To prepare slides, put one drop of chromosome suspension on a very clean glass slide. Ambient temperature and humidity influence how quickly the fixative dries on the slide. Slides which dry too quickly have very low contrast and the chromosomes have a tendency to come off the slide when denatured. Slides which dry too slowly appear refractile under the microscope and may exhibit poor morphology after denaturation. Chromosome preparations for FISH should be between one day and one month old. Extra slides can be stored in a desiccating box for several months at −20 °C.

Rinse the slides in 2 × SSC, pH 7.0 for 5 min. Treat them for at least 30 min at 37 °C in 2 × SSC containing 100 μg mL^{-1} RNase A (see Appendix 1). Then wash slides 3 × 5 min in 2 × SSC, pH 7.0. RNase treatment serves to remove endogenous RNA which may hybridize with homologous probe DNA sequences, forming RNA–DNA hybrids. This improves the signal-to-noise ratio in hybridization to DNA targets.

Incubate the slides with pepsin for 10 min (but not much longer) at 37 °C in 0.05% (w/v) pepsin powder in 0.01 M HCl. Wash slides 2 × 5 min in 1 × PBS (phosphate-buffered saline) (see Appendix 1) and then 2 × 5 min in 1 × PBS containing 50 mM $MgCl_2$. Fix slides with 1% formaldehyde, 1 × PBS, 50 mM $MgCl_2$ for 10–60 min at room temperature. Wash the slides thoroughly in PBS. Pepsin and other proteases (i.e. proteinase K or pronase) increase accessibility by digesting chromosomal proteins that package the target DNA. Since metaphases and nuclei have the tendency to come off the slides after prolonged protease treatment, refixation of the slides with formaldehyde is recommended.

Dehydrate slides in an ethanol series (70%, 85%, 100%) for 5 min each and air-dry. Place 200 μL of denaturation solution, consisting of 70% deionized formamide and 2 × SSC (see Appendix 1), on each slide and cover with a coverslip. Place in a 90 °C oven for approximately 60 s. Shake off coverslip and put slide immediately in ice-cold 70% ethanol. Dehydrate again in an ice-cold ethanol series (70%, 85%, 100%). After air-drying, the slides are ready for in situ hybridization.

Alternatively, chromosomes may be denatured in a water bath at 70 °C by immersing the slide for 2–5 min in a Coplin jar with 70% formamide, 2 × SSC. This may be more accurate and easier to standardize than the procedure described above, but it also creates more waste. Formamide is a proven carcinogen and teratogen. Denaturation of chromosomal target DNA can also be achieved with extremes of alkaline pH; however, heat (formamide) denaturations have become popular because of their experimental reproducibility and greater effectiveness.

Poorly fixed chromosomes may over-denature, resulting in a hollow, puffy apperarance under the phase contrast microscope, and/or come off the slide during denaturation in formamide. Inferior grade formamide may also destroy much of the chromosomal morphology. Denaturation inevitably leads to loss of morphology, so in practice a compromise must be found between hybridization signal and chromosome morphology. Variations in time and temperature should be evaluated to find the best conditions for denaturation.

5.3 Hybridization and Signal Detection

Pipette 30 μL of denatured and pre-annealed probe mixture on the prewarmed slide (15 μL for half a slide). Cover the slide with a coverslip, being careful not to trap air bubbles, and seal the edges with rubber cement.

Hybridize overnight in a moist chamber at 37 °C. Longer hybridization times (2–7 days) may be required for probes which do not show 100% sequence identity with the chromosomal target.

Wash the slides for 3 × 5 min at 42 °C in 50% formamide, 2 × SSC (see Appendix 1) and then 2 × 5 min at 60 °C in 0.1 × SSC. To lower the stringency, increase the salt concentration (i.e. to 0.5–2 × SSC) of the last two

washes. To increase the stringency, use 65% formamide, 2 × SSC for the first three washes. However, hybridization rather than stringent washing is recommended. The washing solutions are pre-warmed in different water baths. The Coplin jar with slides and pre-warmed washing solution is agitated on a platform shaker.

The probe DNA hybridizes not only to the target DNA but also nonspecifically to sequences which bear partial homology to the probe sequence. Since such nonspecific hybrids are less stable than perfectly matched DNA hybrids, they can be dissociated by posthybridization washes of various stringencies.

Directly labeled probes can be viewed immediately after the posthybridization washes. Secondary incubations and washes are required for the visualization of haptenized DNA probes. Prior to incubation with secondary reagents, slides should be blocked by treating with 5% bovine serum albumin (BSA) in 4 × SSC, 0.1% Tween 20. Incubate slides in a Coplin jar with fresh blocking solution (see Appendix 1) for at least 30 min at 37 °C. Blocking prevents unspecific antibody binding and reduces background.

Dilute fluorescent (strept)avidin and anti-digoxigenin antibodies, i.e. FITC-avidin and Cy3-conjugated anti-digoxigenin, 1 : 800 and 1 : 200, respectively (or according to recommendations of the supplier) in 1% BSA, 0.1% Tween 20, 4 × SSC (see Appendix 1). Cover the slides with 200 µL of detection solution and a coverslip. Incubate in a moist chamber in the dark at 37 °C for 30 min.

Shake off the coverslips and wash the slides 3 × 5 min in a Coplin jar with 0.1% Tween 20, 4 × SSC (see Appendix 1) at 42 °C.

To visualize chromosomes and cell nuclei, counterstain the slides for 1–2 min with 1 µg mL^{-1} DAPI in 2 × SSC. Wash the slides thoroughly with distilled water, air-dry and then mount them with antifade mounting medium (see Appendix 1). 1,4-Diazobicyclo[2.2.2]octane (DABCO) retards fading of the fluorescence after exposure of the slides to fluorescence light.

Large-insert clones and DNA libraries should produce specific fluorescence signals, which are clearly visible by eye through the microscope. Small probes (<5 kb) and cDNAs may require immunological signal amplification to be seen by eye. The fluorescence of (strept)avidin–biotin complexes can be enhanced by using biotinylated goat anti-avidin antibody followed by another incubation with fluorescent (strept)avidin. This ''sandwich technique'' for signal amplification of biotinylated probes can be repeated several times.[28] Digoxigenated DNA probes binding fluorescent mouse anti-digoxigenin antibody can be amplified with a second layer of fluorescent rabbit anti-mouse antibodies and, if necessary, with a third layer of fluorescent goat anti-rabbit antibodies. FISH signal intensities can be even more dramatically enhanced by the tyramide method.[29] Peroxidase-conjugated (strept)avidin or anti-digoxigenin antibody is applied as a first layer. In a second step, this peroxidase uses fluorescent or biotinylated tyramide as a substrate to deposit many fluorochrome or biotin molecules in the immediate vicinity of the in situ bound enzyme. These can be evaluated either directly or after a third incubation with fluorochrome-labeled (strept)avidin. As a general rule, every amplification level not only enhances the signal, but also the background intensity. Because of the high sensitivity of digital cameras, use of electronic image enhancement is arguably preferred over immunocytochemical signal amplification.

Proper chromosome identification requires the simultaneous observation of chromosome bands and hybridization signals. A large number of widely different fluorescent banding procedures have been proposed, including FISH of interspersed Alu elements, which are enriched in R bands;[30] saline heat treatment followed by acridine orange staining;[31] and various counterstain-enhanced banding methods, i.e. actinomycin/DAPI staining for G bands or distamycin A/chromomycin A3 for R bands.[32] In general, the resulting chromosome banding patterns are just sufficient to identify chromosomes and to map probes at a low resolution (250–300 band level). Fluorescent replication banding of BrdU-substituted chromosomes tends to produce a more accurate banding after hybridization.[33]

Specialized FISH software allows the conversion of the plain DAPI fluorescence of a chromosome into a G-like banding pattern (Figure 2c; Figure 3c). In addition, the cytogenetic location of a hybridization signal can be expressed as the fractional length (FLpter) of the total chromosome relative to the short-arm telomere.[34] Many FISH programs infer the cytogenetic band location of a hybridization signal from the FLpter measurement. In this context it is important to note that varying degrees of chromosome condensation can lead to discordances between FLpter values and band assignments. Nevertheless, both digital banding and FLpter measurements are very straightforward and now used in most laboratories.

6 MULTICOLOR FLUORESCENCE IN SITU HYBRIDIZATION

One of the major advantages of FISH is the possibility of simultaneously detecting multiple DNA targets on the same metaphase in different colors. Using standard experimental conditions and FISH workstations, one probe may be detected with a green fluorescent dye, a second probe with a red dye, while the chromosomes

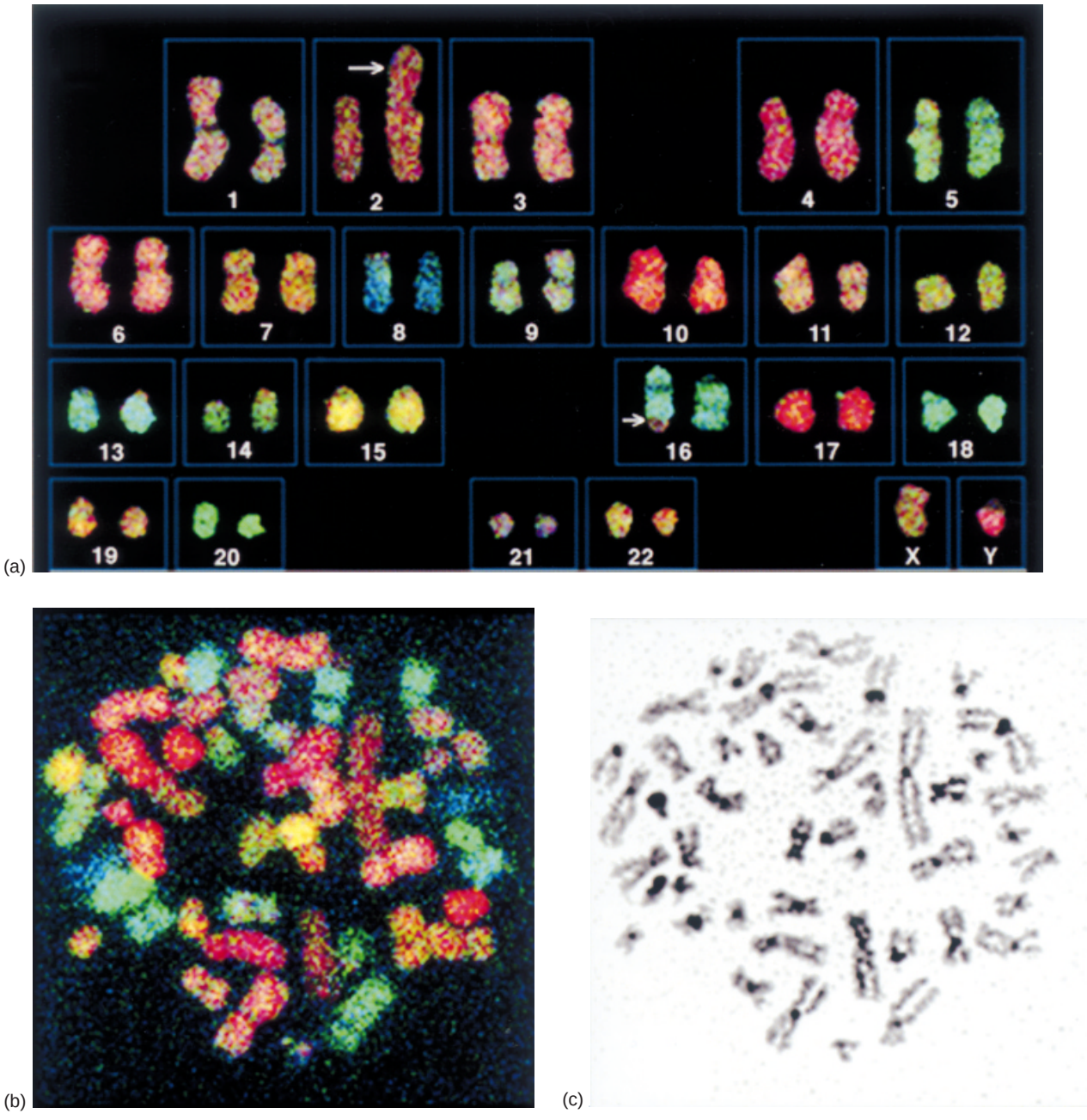

Figure 3 Multicolor FISH analysis of a tumor metaphase spread. (a) Spectral karyotype. Note the enlarged chromosome 2 (large arrow) and the additional chromosome 2 material (small arrow) on the long-arm tip of chromosome 16. (b) Spectral image and (c) converted DAPI image of the respective metaphase spread. Despite inferior tumor metaphase quality, the rearranged chromosomes are easily identified.

are counterstained in blue with DAPI. The three different colors are discriminated with the aid of specific filter sets. In the past few years, an increasing number of spectrally resolvable fluorochromes, ranging from the visible blue to the near infrared, has been developed for the fluorescent detection of DNA probes in multicolor FISH experiments. By combinatorial probe labeling, many more DNA targets can be discriminated than the number of distinguishable fluorochromes. Combinatorial labeling involves the introduction of two or more fluorochromes into a single probe in approximately equal amounts. Different probes are characterized by different combinations of fluorochromes. The maximum number of useful Boolean combinations of n fluors is $2^n - 1$. The 31 possible combinations for five fluorochromes allow one to uniquely identify all 24 human chromosome types.

There are two strategies for analyzing combinatorial multicolor FISH images. The multiplex fluorescence in situ hybridization (M-FISH) procedure is based on highly specific epifluorescence filter sets that have very narrow bandwidths, compared with standard filters, and, therefore, give a high degree of discrimination between all possible fluor combinations.[35] Images are acquired separately for each of the five fluorochromes and for DAPI counterstaining, then are overlaid, generating a composite image. Specialized software is used to pseudocolor each chromosome according to its fluorochrome composition. In contrast, spectral karyotype analysis (SKY) is significantly different from conventional epifluorescence microscopy and does not rely on fluorochrome-specific optical filters. It combines optical microscopy, Fourier spectroscopy, and CCD imaging. Discrimination of multiple fluorochromes is achieved by measuring a discrete fluorescence emission spectrum (representing a sequence of intensities at many different wavelengths) simultaneously at all image points after a single exposure. A filter set with broad emission bands allows the simultaneous excitement of all dyes and measurement of their emission spectra. Specialized software then classifies each chromosome, by assigning the same pseudocolor to all pixels with the same emission spectrum (Figure 3).[36]

PCR-amplifiable DNA libraries from microdissected or flow-sorted chromosomes are used as a starting material. Table 1 shows a combinatorial labeling scheme for 24-color FISH. Spectrum Orange, Texas Red, and Rhodamine are directly incorporated by PCR. Cy5-avidin and Cy5.5-conjugated antibodies are used as secondary reagents to detect biotinylated and digoxigenated probes. For each probe, one to four separate PCR labeling reactions are necessary, each with a fluorescent, biotinylated or digoxigenated dUTP.

To hybridize one slide, take 5 μL of PCR-labeled DNA for each DNA library and each color. For example, 5 μL of each Texas Red®-, rhodamine-, and biotin-labeled chromosome 1 library are taken. The final probe mixture consists of 57 aliquots or 285 μL. Then add 50 μL Cot-1 DNA, 1 μL salmon sperm DNA, 20 μL 3 M NaAcetate (see Appendix 1), 1 mL ice-cold EtOH, and precipitate the DNA at −20 °C overnight. Spin in a microcentrifuge for 45 min at 15 000 rpm and discard the supernatant. Wash the DNA pellet in 1 mL 70% EtOH and spin again. Then allow the precipitated DNA pellet to air-dry in the dark.

Resuspend the DNA pellet in 30 μL hybridization mixture (50% formamide, 10% dextran sulfate, 2 × SSC). In order to resuspend this large amount of probe DNA, it may be necessary to agitate the hybridization mixture on a shaker at 37 °C for several hours. Denature the hybridization mixture for 10 min at 80 °C and incubate it for approximately 30 min at 37 °C.

Table 1 Fluorescence labeling scheme of human chromosome-specific painting probes

Chromosome	Spectrum Orange	Texas Red®	Rhodamine 110	Biotin (Cy5)	Digoxigenin (Cy5.5)
1		×	×	×	
2					×
3	×		×	×	×
4			×	×	
5	×	×	×		×
6		×	×	×	×
7		×		×	
8			×		
9	×		×		×
10				×	×
11	×		×	×	
12		×			×
13	×		×		
14		×			
15	×	×		×	
16		×	×		
17				×	
18	×	×	×		
19	×			×	
20	×				
21			×		×
22	×	×		×	×
X	×				×
Y			×	×	×

After the pre-annealing step, put the probe mixture on the denatured chromosome preparation. Hybridization should be done for several days (up to one week) at 37 °C in a moist chamber in the dark.

Posthybridization washes should be done, using standard conditions of stringency (see section 5.3). Then incubate the slides with Cy5-avidin and mouse anti-digoxigenin antibody, diluted 1 : 500 and 1 : 200, respectively, in 1% BSA, 0.1% Tween, 4 × SSC. After washing the slides 3 × 5 min with 0.1% Tween, 4 × SSC, incubate slides with Cy5.5-conjugated anti-mouse IgG, diluted 1 : 200 with detection buffer (see Appendix 1). All incubation and washing steps must be performed in the dark. Finally, counterstain the slides with DAPI and mount them.

Spectral analysis requires a SpectraCube system (Applied Spectral Imaging). The imaging system is attached to an epifluorescence microscope. It consists of an optical head with a Sagnac common path interferometer. The interferogram of each pixel is analyzed by Fourier transformation, to create a spectrum. This spectrum is imaged with a CCD camera. A spectral-based classification algorithm identifies the multiple different spectra in the acquired image and highlights them in RGB (red/green/blue) classification colors.[36,37] By computer classification of spectra, the spectrally overlapping (combinatorially labeled) 24 DNA libraries can be resolved

and the entire genome can be analyzed simultaneously on the basis of the differential display of all human chromosomes.

Multicolor karyotyping is a very powerful screening method for complex interchromosomal rearrangements, i.e. in tumor cells (Figure 3). It can decipher extensively rearranged karyotypes that cannot be fully characterized by classical chromosome banding or conventional FISH. The limit of sensitivity for karyotype analysis with combinatorial probe labeling and currently available M-FISH or SKY systems is in the order of several megabases and, thus, lower than that with standard (three-color) FISH. An alternative approach to increase the number of detectable chromosomal targets beyond the number of spectrally resolvable fluorochromes is ratio labeling. For example, all 24 human chromosomes can be discrimated by a combination of only four fluorochromes at different ratios.[38]

7 COMPARATIVE GENOMIC HYBRIDIZATION

CGH is a method to screen a genome for DNA-sequence copy-number variation.[18,39] Tester DNA, isolated from a tumor or a patient, and normal (human) reference DNA are labeled with different fluorochromes, i.e. FITC for the tester DNA and Cy3 for the reference DNA, and hybridized together on normal metaphase spreads under suppression conditions. Hybridization of tester DNA from a chromosomally balanced genome results in a more or less uniform staining of the target chromosomes with both colors. Regions of gain and loss of DNA sequences in a tumor change the relative amounts of tester and reference DNA bound at the respective chromosomal loci and can be identified by measuring the ratio of intensities of the two fluorochromes (Figure 4). Amplifications and deletions in the test sample should be visible as higher and lower green fluorescence intensity relative to the red fluorescence intensity of the reference DNA. Large chromosomal imbalances, i.e. loss or gain of an entire chromosome, lead to marked differences in the hybridization profile of the target chromosome on the hybridized normal metaphase spread and should be visible by eye through the microscope. Specialized CGH software, which integrates the green-to-red fluorescence intensities and calculates the ratio along the hybridized chromosomes (Figure 4b), allows the detection of smaller regional changes involving gain or loss of one copy of approximately 10 Mb DNA. Of course, high-level amplifications, i.e. of certain oncogenes, yield a much higher DNA resolution of CGH. Most importantly, in contrast to classical tumor cytogenetics, CGH does not require tumor metaphase spreads. Karyotype analysis in solid and many other tumor types has been hampered by low mitotic index and poor metaphase quality. A single CGH experiment provides comprehensive information on genetic imbalances. However, it cannot detect balanced chromosome rearrangements such as translocations and inversions. In this light, multicolor FISH and CGH are complementary techniques.

If possible, total genomic tumor and reference DNA should be used for CGH experiments. High-molecular-weight DNA is prepared according to standard procedures.[40] One milligram of tissue yields approximately one microgram of genomic DNA. When only small amounts of fresh tumor tissue or archival tumor material are available, a DNA template is generated by DOP PCR amplification.[21,41]

Tumor DNA is labeled by standard nick translation with biotin and detected with FITC-avidin. Reference DNA is labeled with digoxigenin and detected with Cy3-conjugated anti-digoxigenin antibody. Direct labeling is also possible. The optimum probe-fragment-size distribution after nick translation ranges from about 500 bp to 2000 bp. For CGH, the probe DNA fragments should be somewhat larger than for standard FISH experiments (200–500 bp) in order to obtain uniform and intense hybridization on chromosomes.

Samples (100 ng) of each biotinylated tumor and digoxigenated reference DNA are combined with 50–100 μg unlabeled Cot-1 DNA. The precipitated probe DNA is resuspended in 50% formamide, 10% dextran sulfate, 2 × SSC.

Following denaturation of probe and slides, comparative hybridization is performed for 2–3 days at 37 °C in a moist chamber. The slides are washed and probes detected with secondary reagents (see section 5.3). The hybridization profiles are evaluated by specialized CGH software.

8 (ULTRA) HIGH-RESOLUTION FLUORESCENCE IN SITU HYBRIDIZATION

Standard FISH techniques are capable of resolving sequences separated by two or more megabases on metaphase chromosomes.[34] Hybridization on less condensed interphase nuclei further increases the resolution to around 100 kb. The average interphase distance between different probe signals correlates with physical distances up to 1–2 Mb.[42,43] Interphase mapping allows the order and relative proximity of sequences which are closely juxtaposed on the same chromosomal region, and therefore cannot be resolved on metaphase chromosomes, to be determined. Interphase

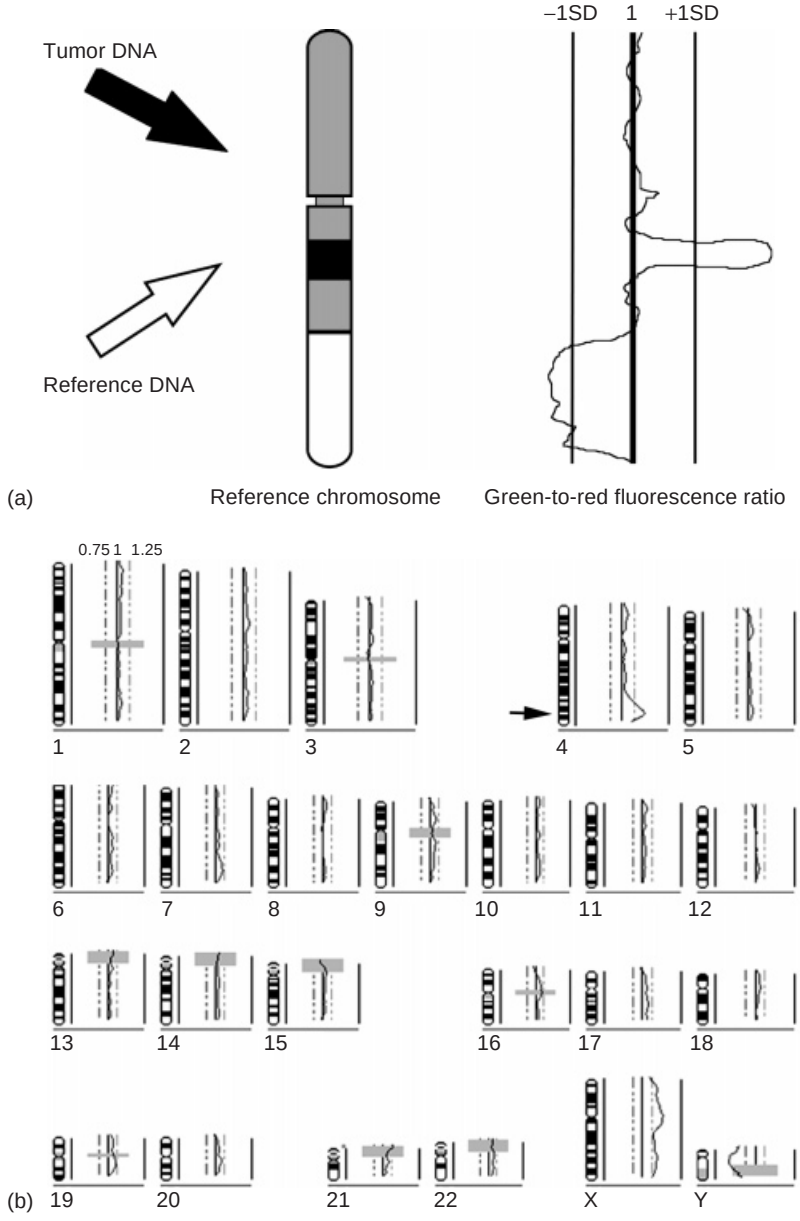

Figure 4 CGH. (a) Schematic representation of the CGH approach. Amplification of a DNA segment (in the proximal long arm) enhances the green-to-red fluorescence ratio of tumor DNA versus reference DNA on the respective chromosome and is shown in black. Conversely, deletion of the distal long arm in the tumor decreases the green-to-red fluorescence ratio of the target region and is shown in white on the reference chromosome. The mean ratio (thick line) plus/minus one standard deviation (thin lines) are shown for the entire chromosome from pter to qter. (b) Copy-number karyotype of a patient with constitutional chromosome aberration. The baseline value (1) representing the mean green-to-red ratio for the entire metaphase is shown in a solid line and ratios of 0.75 (loss bar) and 1.25 (gain bar) as dotted lines. Based on this analysis, the distal long arm of chromosome 4 (arrow) is over-represented in the patient. The X and Y ratios reflect the fact that DNA from a female patient was compared with a male reference DNA. The ratio data of regions that contain mainly repeat DNAs and may exhibit large copy-number polymorphisms are unreliable and, therefore, excluded from the analysis (horizontal gray bars). Hybridization signals in these regions are partially suppressed by Cot-1 DNA.

FISH has also proved valuable for determining replication order and timing of specific DNA sequences.[44] Replicated DNA segments consisting of two sister chromatids appear as doublet hybridization signals during interphase, whereas unreplicated chromosomal segments with one chromatid give only single hybridization dots (Figure 1f). Most sequences have a defined time zone of replication within the cell cycle and replicate in a fairly

synchronous manner. Asynchronous DNA replication, which generates one doublet and one singlet hybridization signal on homologous interphase chromosomes, defines regions containing imprinted genes[45] or repeat DNA families.[46]

The general usefulness of interphase mapping is greatly hampered by the three-dimensional organization of chromatin in the cell nucleus. The data become more difficult or even impossible to analyze when the distance between markers is larger than several megabases. Based on the observation that less condensed interphase chromatin provides greater DNA resolution than metaphase chromosomes, different strategies have been developed to push the limits of resolution of FISH further by artificially decondensing chromatin. Briefly, the cells are lysed with high salt and/or detergent and the chromatin is spread across the surface of a glass slide either by gravity (tilting the slide) or by mechanical spreading. This can cause complete detachment of the DNA loops from the nuclear matrix, producing highly elongated chromatin fibers.[46–48] The finest level of DNA resolution of the various fiber FISH protocols is 1–10 kb, thus closing the gap between (molecular) cytogenetic and molecular physical mapping techniques. Alternatively, high-molecular-weight DNA molecules in solution can be stretched on glass surfaces by molecular combing.[49–51]

8.1 Fiber Fluorescence In Situ Hybridization

The following protocol describes the preparation of DNA fibers from agarose-embedded cells.[48] This technique can be adapted in most laboratories, producing suitable targets for high-resolution mapping.

To improve retention of DNA fibers on slides, the slides are coated with poly-L-lysine as follows. Put slides through 0.2 M HCl, ddH_2O, and acetone for 30 s in each solution, then allow to air-dry at room temperature. Place the clean slides in 0.15% gelatin, 0.03% sodium azide solution for 5 min at room temperature, then rinse for 30 s with ddH_2O and air-dry at room temperature overnight. Use molecular biology grade gelatin and filter the solution to get rid of any gelatin not completely in solution.

Place slides in 0.02% poly-L-lysine for 10 min at room temperature, rinse for 30 s with ddH_2O and air-dry for 1 h. Then repeat poly-L-lysine incubation and ddH_2O rinse and dry overnight at room temperature. Slides stored at 4 °C can be kept for months before use.

Most cell types can be used to prepare blocks for fiber FISH. Most commonly used are fibroblasts or peripheral blood lymphocytes. Collect, wash, and resuspend cells in $1 \times$ PBS (see Appendix 1) to a concentration of 2×10^6 cells per 100 µL. Warm this cell suspension slightly at 37 °C. Have ready 1.9% low-melt agarose kept at 50 °C. Mix melted agarose and cell suspension 1 : 1 and pipette into molds (blocks 100 µL or less) and chill for 30 min at 4 °C. Final cell concentration in the blocks is 10^6 cells in 100 µL.

Incubate blocks in a solution of 2 mg mL^{-1} proteinase K, 0.5 M EDTA, pH 8.0, 1% *n*-lauroylsarcosine. Use at least 1 mL of this solution per 100-µL block. Incubate for 48–72 h at 50 °C (with gentle agitation, if possible).

Leaving the blocks in the proteinase K solution, chill them for 30 min at 4 °C to firm up the gel fragments, then transfer them into TE (10 mM Tris–HCl, pH 7.5, 1 mM EDTA) buffer (see Appendix 1), and wash for two days at room temperature, changing the TE several times over this time period. Blocks can be stored in 0.5 M EDTA at 4 °C for months. Dialyse blocks stored in EDTA with several changes of TE before using for fibers.

For fiber formation, place approximately 10 µL of block on one end of a poly-L-lysine coated slide. Pipette onto the block around 40 µL of ddH_2O. Preheat the microwave oven to get the rotating glass plate very hot, by running empty for 5 min. One at a time, place a slide, with agarose and ddH_2O as described above, into the microwave with the slide near the outer edge of the rotating plate, and microwave at full power for 20–40 s.

Using a plain glass slide, spread the melted gel down the slide. Be sure that the agarose is completely melted. The angle of pulling should be low (20° or so) and the plain slide should be pulled with mild pressure. Ideally, the fluid should boil a bit as it gets pulled down the slide and the slide should appear dry immediately after pulling, i.e. extending the fibers. Allow slides to dry for about 1 h at room temperature.

Denature by placing the slides in a Coplin jar containing 70% formamide, $2 \times$ SSC at 74 °C for 4 min. This step is likely to remove quite a bit of the agarose. Put through an ethanol series (70%, 90%, 100%) at −20 °C and then allow to air-dry. As soon as the slides are dry, they are ready for probe application and hybridization. Make the fibers the same day as the hybridization will be done.

Following probe hybridization for 1–3 days, wash slides for 3×5 min at 42 °C in 50% formamide, $2 \times$ SSC and then 2×5 min at 42 °C in $0.5 \times$ SSC. Immunocytochemical detection should be performed according to standard procedures.

The distance for fiber FISH mapping is 5–500 kb. Cosmids of 30–40 kb length produce a string of fluorescent signals up to several dozen micrometers long. YAC probes produce very elongated straight signals on a certain percentage of fibers. The longer the hybridized YAC, the more likely it is that its sequence may be disrupted during fiber preparation. It is important to note

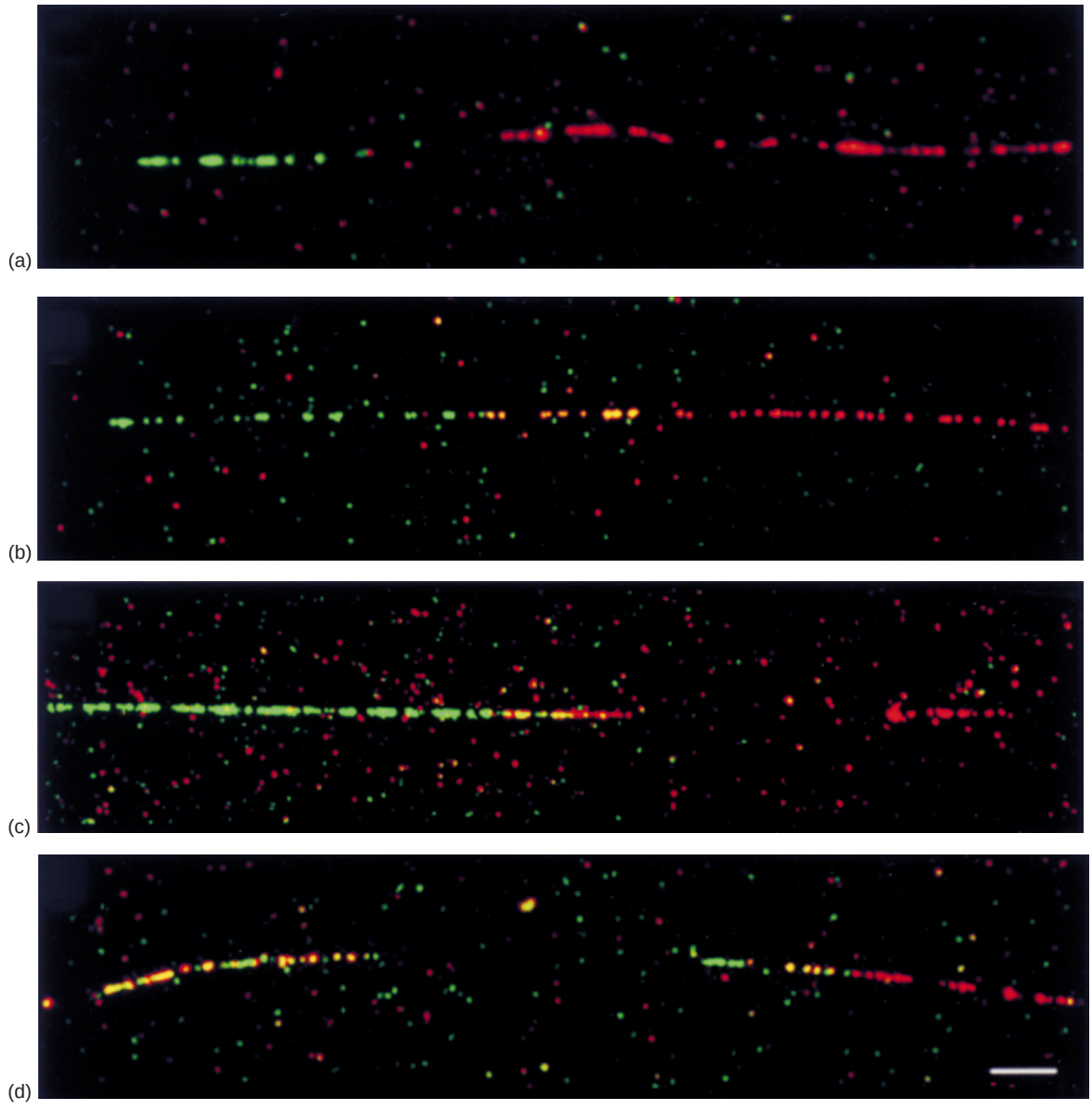

Figure 5 High-resolution FISH on extended chromatin fibers. (a) Hybridization of a green cosmid and a red PAC, which are closely adjacent but nonoverlapping. (b) Two partially overlapping PACs. (c) The red cosmid is partially overlapping the green PAC. The second red cosmid signal on the right-hand side is located at a distance of 20–30 kb and delineates a gene duplication. (d) The nonoverlapping PAC on the left-hand side is labeled simultaneously in green and red. The green cosmid on the right-hand side is partially overlapping with the red PAC. Scale bar, 10 μm.

that the fibers onto which clones are mapped are condensed to varying degrees and, therefore, a larger number of fibers (at least 20) should be analyzed for quantitative measurements.

Fiber FISH techniques can be used to map clones, i.e. cDNAs or cosmids, with a few kilobases resolution on individual straightened DNA molecules. It allows one to order probes relative to each other, to orient probes within a contig, and to determine the degree of overlap or gap size between different probes (Figure 5).

8.2 Molecular Combing

Molecular combing is a process in which a solution of DNA target molecules is placed in parallel orientation on a flat glass surface. In the most simple method, a drop of DNA solution is placed on a silanized coverslip and covered with another untreated coverslip, forcing the drop to spread. DNA molecules bind to the vinyl end groups ($-CH{=}CH_2$) of the silanized glass surface by their extremities only, presumably because of the

presence of a free protonated phosphate at the 5′-end of DNA molecules. The receding meniscus of the air–water interface during evaporation of the solution exerts a force on the DNA, resulting in a more or less uniform extension of the attached DNA molecules.[49,50] The highest density of linearized DNA molecules and constant stretching are achieved by using a DNA combing apparatus (Adil Instruments).[51] Silanized coverslips are dipped into a Teflon® reservoir with buffered DNA solution, resulting in random binding of the DNA molecules at one end (usually the 5′-end) or occasionally at both ends to the glass surface. When the slide is pulled out of the DNA solution by a mechanical device, the hydrodynamic action of the receding meniscus straightens the DNA molecules. Due to the silanized surface, the DNA molecules dry immediately as the coverslip comes out of the solution, permanently fixing them to the coverslip surface. A major advantage of molecular combing is that DNA molecules are aligned in a single direction and homogeneously stretched (approximately 2 kb μm^{-1}). Similarly to fiber FISH, most combed molecules are longer than several hundred kilobases and occasionally as long as 1 Mb.

Standard agarose blocks (see section 8.1) containing YAC DNA or total human genomic DNA are used for the preparation of DNA solution. The blocks will contain approximately 10^7 yeast cells, equivalent to 1 μg yeast DNA, or 10^6 diploid human cells, equivalent to 6.6 μg human genomic DNA, per 100 μL.

If blocks have been stored in 0.5 M EDTA, wash them in 15 mL TE (see Appendix 1) for at least 2 h. Stain each block with the fluorescent dye YOYO-1 (0.33 nmol μg^{-1} DNA) in 100 μL of 40 mM Tris–HCl, pH 8.0, 2 mM EDTA in 2-mL Eppendorf tubes for 1 h, protecting from light. YOYO-1 staining is optional. It allows visualization of combed DNA molecules on the glass surfaces, to judge the quality of preparations before hybridization.

Discard staining solution, add 500 μL TE and incubate for 20 min at 68 °C, then add 6 μL of 10 × NEB buffer (100 mM Bis Tris–HCl, pH 6.5, 10 mM EDTA, supplied with the enzyme) and 2 U of β-agarase and incubate for at least 2 h at 40 °C.

Dilute DNA with 0.5 M MES (morpholinoethanesulfonic acid) buffer, pH 5.5, very carefully in 15-mL round-bottom tubes. The final concentration of yeast DNA should be 0.25 μg mL^{-1} in 50 mM MES and of human genomic DNA 1.5 μg mL^{-1} in 150 mM MES. For human genomic DNA incubate for 30 min at 75 °C to assist re-suspension of the DNA.

Cool DNA solution to room temperature and pour into the Teflon® reservoir of the molecular combing apparatus. This solution can be stored at 4 °C for several weeks.

The molecular combing apparatus[51] consists of a vertically moving clip holding up to three coverslips. Silanized coverslips are dipped vertically into the Teflon® reservoir containing DNA solution and after 5 min of incubation are pulled out at a constant speed of 300 μm s^{-1}. During incubation DNA molecules in solution bind to the coverslip surface by their extremities. When the coverslips are pulled out of the DNA solution, the fixed horizontal meniscus (air–solution–coverslip interface) exerts a localized, constant, downward vertical force on the immersed part of attached DNA molecules, resulting in the DNA unwinding and straightening. The hydrophobic silanized glass surface dries immediately as it is pulled out of solution. The stretched, parallel DNA molecules are irreversibly fixed on the dry glass surface. Using a cyanoacrylate glue, mount coverslips on glass slides and dry slides overnight at 60 °C. Either use immediately for FISH or store in a desiccating box at −20 °C. Using this protocol, up to hundreds of human genomes or thousands of YAC molecules can be combed on a single coverslip, which allows accurate DNA mapping of probes on a single slide.

9 IMPORTANT APPLICATIONS

FISH is a very straightforward and powerful physical mapping technique which plays an increasingly important role in the study of genome structure and functional organization. Compared with other molecular mapping procedures, i.e. long- and short-range restriction analysis and PCR, the major advantage of FISH is the direct visualization of a sequence of interest on the homologous chromosome region and/or with respect to already mapped loci. FISH mapping on metaphase chromosomes allows the rapid localization of cloned DNA sequences to a specific chromosome region at a resolution of 1–3 Mb. FISH on interphase nuclei, extended chromatin fibers, and combed DNA molecules allows the ordering of probes, definition of the size of gaps and overlaps between clones, and the construction of contigs with much higher resolutions between 1 kb and 100 kb (Figure 5).[47–52] Metaphase and high-resolution FISH mapping have found widespread application in positional cloning projects, i.e. for the cytogenetic and molecular characterization of chromosome breakpoints (Figure 1c, d; Figure 2b, c). They greatly facilitate the construction of sequence-ready maps of well-ordered cosmid and PAC clones, which is crucial for the human genome project.

FISH has also added a new dimension to the cytogenetic analysis of structural and numerical chromosomal aberrations. It has become the most important adjunct technique to classical chromosome banding analysis, which is routinely being applied in prenatal and postnatal diagnostics and cancer cytogenetics. Because of

its higher sensitivity, specificity, and DNA resolution, FISH can detect cytogenetically cryptic chromosome rearrangements that could not be determined previously by conventional banding methods alone. For example, microdeletions and translocations involving the chromosome ends (Figure 1a, b) appear to be an important cause of disease in patients with unexplained mental retardation[53] and most likely also in other patient populations. Similarly, patients with seemingly balanced chromosome rearrangements and a recognizable clinical phenotype may show deletions of several megabases of DNA in the breakpoint region(s).[13] Cytogenetically and genetically anchored large-insert clones for any chromosome region of interest are available for routine cytogenetics and can be adapted with unlimited flexibility to the study of cytogenetically cryptic and visible chromosome abnormalities.

Chromosome-specific DNA libraries are particularly useful to paint chromosomes involved in structural or numerical aberrations in both metaphase and interphase cells and to identify the chromosomal origin of marker chromosomes. Recent advances in combinatorial probe labeling and multicolor FISH allow the simultaneous visualization of all 24 human chromosomes in different colors (Figure 3). This has wide utility in cancer cytogenetics for characterizing even extensively rearranged chromosomes and karyotypes that produce very complex banding patterns[35,36,38] In addition, CGH approaches allow one to rapidly screen tumor and patient genomes for complete and partial chromosome gains and losses and for gene amplifications (Figure 4).[18,39]

FISH has the added advantage over chromosome banding that it also detects chromosomal abnormalities, especially numerical aberrations, in interphase nuclei (Figure 1e, g). Chromosome-specific α-satellites produce very intense and discrete hybridization signals on interphase nuclei and, therefore, are frequently used to measure aneuploidy rates for individual chromosomes in tumors,[54] sperm,[55] pre-implantation embryos,[56] and other cell substrates, where metaphase chromosome preparation is difficult or even impossible. To a lesser extent, region-specific large-insert clones may also be used. Given the large contribution of chromosomes 13, 18, 21, X, and Y to constitutional chromosomal aneuploidies, interphase assays determining copy-numbers of these chromosomes have become an important complementary tool in pre-implantation and prenatal diagnostics.[57] Specific probes spanning or binding near chromosomal breakpoints, i.e. of the Philadelphia chromosome in chronic myelocytic leukemia, allow the interphase analysis of translocations and other structural rearrangements.[58]

In addition to physical mapping and diagnostic applications, FISH has proven tremendously useful for the study of genome evolution (Figure 2). Comparative mapping of individual gene loci has allowed the generation of homology maps in at least some important animal models.[59,60] More direct approaches to comparative genome mapping involve the delineation of contiguous conserved (syntenic) segments containing hundreds of homologous loci by Zoo-FISH using human (or mouse) DNA libraries on chromosomes of heterologous species.[61–63] The presence of orthologous genes on interspecific homologous chromosome segments (conservation of synteny) reflects the common phylogenetic origin of species and also the ancestral genomic organization. Comparative (chromosome homology) maps can be used to predict the locations of homologous genes in the human and animal genomes and to answer important questions about the forces guiding mammalian genome organization, evolution, and pathology. They will also be instrumental for integrating the genetic, cytogenetic, and molecular information in human and animal species that is accumulating through genome research.

ACKNOWLEDGMENTS

The author would like to thank Tanja Hardt, Lisa Riesselmann, Frank Grützner, Zhihong Shan, and Michael Speicher for providing images.

ABBREVIATIONS AND ACRONYMS

BAC	Bacterial Artificial Chromosome
BME	β-Mercaptoethanol
bp	Base Pair
BSA	Bovine Serum Albumin
CCD	Charge-coupled Device
cDNA	Complementary DNA
CGH	Comparative Genomic Hybridization
cM	Centimorgan
DABCO	1,4-Diazobicyclo[2.2.2]octane
DAPI	4′,6-Diamidino-2-phenylindole
ddH_2O	Double-distilled Water
dNTP	Deoxynucleotide Triphosphate
DOP	Degenerate Oligonucleotide Primed
EDTA	Ethylenediamine Tetraacetic Acid
FISH	Fluorescence In Situ Hybridization
FITC	Fluorescein Isothiocyanate
FLpter	Fractional Length
kb	Kilobase
Mb	Megabase
MES	Morpholinoethanesulfonic Acid
M-FISH	Multiplex Fluorescence In Situ Hybridization

PAC	P1-derived Artificial Chromosome
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
RAM	Random Access Memory
RGB	Red/Green/Blue
SDS	Sodium Dodecyl Sulfate
SKY	Spectral Karyotype Analysis
SSC	Standard Saline Citrate
UV	Ultraviolet
YAC	Yeast Artificial Chromosome

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Immunochemistry

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Probes • Electron Tomography of Chromosome Structure • Genome Physical Mapping Using BACs • Optical Mapping in Genomic Analysis • Polymerase Chain Reaction and Other Amplification Systems • Radiation Hybrid Mapping • Restriction Landmark Genomic and cDNA Scanning • Sequencing Strategies and Tactics in DNA and RNA Analysis

APPENDIX 1

Buffers and Stock Solutions

Antifade Mounting Solution

DABCO	2.33 g	(2.3%)
Tris–HCl (1 M) (pH 8.0)	2 mL	(20 mM)
Glycerol	90 mL	(90%)
Distilled water	to 100 mL	

Store at −20 °C in dark vials. Solution can be used for up to 6 months.

Blocking Solution

20 × SSC	20 mL	(4 × SSC)
BSA (inexpensive grade)	3 g	(3%)
Tween 20 (10%)	1 mL	(0.1%)
Distilled water	to 100 mL	

Stir for at least 10 min to completely dissolve the BSA. Always prepare freshly.

Bovine Serum Albumin (10%)

BSA (fraction V) (molecular biology grade)	1 g	(10%)
Distilled water	to 10 mL	

Dissolve in distilled water and sterilize by filtration. Store at −20 °C in aliquots.

Denaturation Solution

Formamide (deionized)	1.4 mL	(70%)
20 × SSC (pH 7.0)	0.2 mL	(2 × SSC)
Distilled water	0.4 mL	

Always prepare freshly.

Detection Buffer (for the Dilution of Streptavidin and Antibodies)

20 × SSC	200 µL	(4 × SSC)
BSA (10%)	100 µL	(1%)
Tween 20 (10%)	10 µL	(0.1%)
Distilled water	to 1 mL	

Dextran Sulfate (50%)

Dextran sulfate	50 g	(50%)
Distilled water	to 100 mL	

Dissolve in distilled water, autoclave, and store at −20 °C.

Disodium Ethylenediamine Tetraacetate (0.5 M) (pH 8.0)

EDTA · $2H_2O$	186.1 g	(0.5 M)
NaOH pellets	~20 g	
Distilled water	to 1 L	

Adjust the pH by addition of NaOH to approximately 8.0. Stir vigorously until the EDTA goes into solution.

Formamide (Deionized)

Formamide (molecular biology grade)	100 mL
Mixed-bed ion-exchange resin (i.e. Bio-Rad AG 501-X8)	10 g

Stir for at least 30 min and filter twice to remove the resin. The conductivity of deionized formamide should be >100 µS. Store in small aliquots at −20 °C.

Hybridization Mixture (2×)

20 × SSC	2 mL	(4 × SSC)
Dextran sulfate (50%)	4 mL	(20%)
Distilled water	4 mL	

Mix vigorously and store in small aliquots at −20 °C.

Magnesium Chloride (1 M)

$MgCl_2 \cdot 6H_2O$	203.3 g	(1 M)
Distilled water	to 1 L	

$MgCl_2$ is extremely hygroscopic. Do not store opened bottles for a long time.

β-Mercaptoethanol (0.1 M)

BME (14.4 M)	70 µL	(0.1 M)
Distilled water	to 10 mL	

Store in small aliquots in the dark at −20 °C.

Nick Translation Buffer (10×)

Tris–HCl (1 M) (pH 8.0)	5 mL	(0.5 M)
$MgCl_2$ (1 M)	0.5 mL	(50 mM)
BSA (10%)	50 µL	(0.5 mg mL^{-1})
Distilled water	to 10 mL	

Phosphate-buffered Saline (10 × PBS)

Sodium chloride (NaCl)	79.5 g	(1360 mM)
Potassium chloride (KCl)	1.5 g	(20 mM)
Sodium phosphate, dibasic (Na_2HPO_4)	15 g	(106 mM)
Potassium phosphate, monobasic (KH_2PO_4)	2 g	(15 mM)
Distilled water	to 1 L	

Adjust pH to 7.2–7.4 and autoclave.

Phosphate-buffered Saline (1×) (pH 7.3)

PBS (10×)	100 mL
Distilled water	900 mL

Postdetection Wash Solution

20 × SSC	100 mL	(4 × SSC)
Tween 20 (10%)	5 mL	(0.1%)
Distilled water	to 500 mL	

Posthybridization Wash Solution

Formamide (does not have to be deionized)	50 mL	(50%)
20 × SSC	10 mL	(2 × SSC)
Distilled water	to 100 mL	

RNase (100 mg mL^{-1})

RNase A (powder)	1 g	(100 mg mL^{-1})
NaAcetate ($C_2H_3O_2Na \cdot 3H_2O$) (3 M)	33 µL	(10 mM)
Distilled water	to 10 mL	

Boil for 10 min to inactivate contaminating DNase activity and store in small aliquots at −20 °C.

Saline Sodium Citrate (20 × SSC) (pH 7.0)

Sodium chloride (NaCl)	175.3 g	(3 M)
Sodium citrate ($Na_3C_6H_5O_7 \cdot 2H_2O$)	88.2 g	(0.3 M)

Adjust pH to 7.0 with a few drops of 10 M NaOH and sterilize by autoclaving.

Salmon Sperm DNA

Salmon sperm	1 g	(10 mg mL^{-1})
Distilled water	100 ml	

Dissolve salmon sperm DNA in distilled water, keep in boiling water for ~30 min so that the size range is between 100 and 500 bp, and store at −20 °C.

Sodium Acetate (3 M) (pH 5.2)

NaAcetate ($C_2H_3O_2Na \cdot 3H_2O$)	408.3 g	(3 M)
Distilled water	to 1 L	(~800 mL)

Adjust pH with glacial acetic acid to 5.2 and adjust volume to 1 L.

Sodium Dodecyl Sulfate (10%)

SDS	100 g	(10%)
Distilled water	to 1 L	(~900 mL)

Wear a mask when weighing SDS. Heat to 68 °C to facilitate dissolution of the SDS and adjust pH to 7.2 with HCl.

Taq *Polymerase Buffer (10×) (without $MgCl_2$)*

Tris–HCl (1 M) (pH 9.0)	1 mL	(100 mM)
KCl (1 M) (74.55 g L^{-1})	5 mL	(500 mM)
Triton X-100	100 µL	(1%)
Distilled water	to 10 mL	

Buffer is usually supplied with the enzyme.

TE

Tris–HCl (1 M) (pH 7.4)	1 mL	(10 mM)
EDTA (0.5 M)	0.2 mL	(1 mM)
Distilled water	to 100 mL	

Tris–HCl Buffer (1 M) (pH 7.4)

Tris base	121.1 g
Hydrochloric acid (HCl conc.)	~70 mL

Distilled water	to 1 L

Cool; adjust to pH 7.4 and add distilled water to 1 L; autoclave.

Tris–HCl (1 M) (pH 8.0)

Tris base	121.1 g
Hydrochloric acid (HCl conc.)	~42 mL
Distilled water	to 1 L

Tween 20 (10%)

Tween 20	10 mL	(10%)
Distilled water	90 mL	

Stir for at least 10 min to completely dissolve the Tween 20 and store at 4 °C.

REFERENCES

1. H. John, M. Birnstiel, K. Jones, 'RNA–DNA Hybrids at the Cytological Level', *Nature*, **223**, 582–587 (1969).
2. M.L. Pardue, J.G. Gall, 'Molecular Hybridization of Radioactive DNA to the DNA of Cytological Preparations', *Proc. Natl. Acad. Sci. USA*, **64**, 600–604 (1969).
3. P.R. Langer, A.A. Waldrop, D.C. Ward, 'Enzymatic Synthesis of Biotin-labeled Polynucleotides: Novel Nucleic Acid Affinity Probes', *Proc. Natl. Acad. Sci. USA*, **78**, 6633–6637 (1981).
4. S.G. Ballard, D.C. Ward, 'Fluorescence In Situ Hybridization Using Digital Imaging Microscopy', *J. Histochem. Cytochem.*, **43**, 1755–1759 (1993).
5. H.F. Willard, 'Evolution of Alpha Satellite', *Curr. Opin. Genet. Dev.*, **1**, 509–514 (1991).
6. I.A. Alexandrov, L.I. Medvedev, T.D. Mashkova, L.L. Kisselev, L.Y. Romanova, Y.B. Yurov, 'Definition of a New Alpha Satellite Suprachromosomal Family Characterized by Monomeric Organization', *Nucleic Acids Res.*, **21**, 2209–2215 (1993).
7. T. Haaf, H.F. Willard, 'Orangutan α-Satellite Monomers are Closely Related to the Human Consensus Sequence', *Mamm. Genome*, **9**, 440–447 (1998).
8. M. Vooijs, L.-C. Yu, D. Tkachuk, D. Pinkel, D. Johnson, J.W. Gray, 'Libraries for Each Human Chromosome, Constructed from Sorter-enriched Chromosomes by Using Linker-adaptor PCR', *Am. J. Hum. Genet.*, **52**, 586–597 (1993).
9. P. Rabbitts, H. Impey, A. Heppell-Parton, C. Langford, C. Tease, N. Lowe, D. Bailey, M.A. Ferguson-Smith, N.P. Carter, 'Chromosome Specific Paints from a High Resolution Flow Karyotype of the Mouse', *Nat. Genet.*, **9**, 369–375 (1995).
10. M.A. Ferguson-Smith, 'Genetic Analysis by Chromosome Sorting and Painting: Phylogenetic and Diagnostic Applications', *Eur. J. Hum. Genet.*, **5**, 253–265 (1997).
11. X.-Y. Guan, P.S. Meltzer, J.M. Trent, 'Rapid Generation of Whole Chromosome Painting Probes by Chromosome Microdissection', *Genomics*, **22**, 101–107 (1994).
12. X.-Y. Guan, P.S. Meltzer, A.C. Burgess, J.M. Trent, 'Coverage of Chromosome 6 by Chromosome Microdissection: Generation of 14 Subregion-specific Probes', *Hum. Genet.*, **95**, 637–640 (1995).
13. J. Wirth, H.-G. Nothwang, S. van der Maarel, C. Menzel, G. Borck, I. Lopez-Pajares, K. Brøndum-Nielsen, N. Tommerup, M. Bugge, H.-H. Ropers, T. Haaf, 'Systematic Characterization of Disease-associated Balanced Chromosome Rearrangements by FISH: Cytogenetically and Genetically Anchored YACs Identify Microdeletions and Candidate Regions for Mental Retardation Genes', *J. Med. Genet.*, **36**, 271–278 (1999).
14. J.R. Korenberg, X.-N. Chen, K.L. Devon, D. Noya, M.L. Oster-Granite, B.W. Birren, 'Mouse Molecular Cytogenetic Resource: 157 BACs Link the Chromosomal and Genetic Maps', *Genome Res.*, **9**, 514–523 (1999).
15. C. Szpirer, J. Szpirer, P. Van Vooren, F. Tissir, J.S. Simon, G. Koike, H.J. Jacob, E.S. Lander, K. Helou, K. Klinga-Levan, G. Levan, 'Gene-based Anchoring of the Rat Genetic Linkage and Cytogenetic Maps: New Regional Localizations, Orientation of the Linkage Groups, and Insights into Mammalian Chromosome Evolution', *Mamm. Genome*, **9**, 721–734 (1998).
16. T. Schwarzacher, A.R. Leitch, M.D. Bennett, J.S. Heslop-Harrison, 'In Situ Localization of Parental Genomes in a Wide Hybrid', *Ann. Bot.*, **64**, 315–324 (1989).
17. J. Jiang, B.S. Gill, 'Nonisotopic In Situ Hybridization and Plant Genome Mapping: The First 10 years', *Genome*, **37**, 717–725 (1994).
18. A. Kallioniemi, O.P. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman, D. Pinkel, 'Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors', *Science*, **258**, 818–821 (1992).
19. P.W.J. Rigby, M. Dieckmann, C. Rhodes, P. Berg, 'Labeling Deoxyribonucleic Acid to High Specific Activity In Vitro by Nick Translation with DNA Polymerase I', *J. Mol. Biol.*, **113**, 237–241 (1977).
20. P. Feinberg, B. Vogelstein, 'A Technique for Radiolabeling DNA Restriction Fragments to High Specific Activity', *Anal. Biochem.*, **137**, 266–267 (1984).
21. H. Telenius, N.P. Carter, C.E. Bebb, M. Nordenskjöld, B.A. Ponder, A. Tunnacliffe, 'Degenerate Oligonucleotide-primed PCR. General Amplification of Target DNA by a Single Degenerate Primer', *Genomics*, **13**, 718–725 (1992).
22. D.L. Nelson, S.A. Ledbetter, L. Corbo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbetter, C.T. Caskey, 'Alu Polymerase Chain Reaction: A Method for Rapid Isolation of Human-specific Sequences from Complex DNA Sources', *Proc. Natl. Acad. Sci. USA*, **86**, 6686–6690 (1989).
23. A.M. Weiner, P.L. Deininger, S. Efstratiadis, 'Nonviral Retroposons: Genes, Pseudogenes, and Transposable

Elements Generated by the Reverse Flow of Genetic Information', *Annu. Rev. Biochem.*, **55**, 631–661 (1986).
24. P. Lichter, T. Cremer, J. Borden, L. Manuelidis, D.C. Ward, 'Delineation of Individual Human Chromosomes in Metaphase and Interphase Cells by In Situ Suppression Hybridization Using Recombinant DNA Libraries', *Hum. Genet.*, **80**, 224–234 (1988).
25. D. Pinkel, J. Landegent, C. Collins, J. Fuscoe, R. Segraves, J. Lucas, J. Gray, 'Fluorescence In Situ Hybridization with Human Chromosome-specific Libraries: Detection of Trisomy 21 and Translocations of Human Chromosome 4', *Proc. Natl. Acad. Sci. USA*, **85**, 9138–9142 (1988).
26. P.E. Nisson, P.C. Watkins, A. Boyle, 'Efficient Mapping of Mouse Clones Using Mouse Cot-1 DNA', *Focus*, **14**, 119–122 (1992).
27. R.S. Verma, A. Babu, *Human Chromosomes. Principles and Techniques*, McGraw-Hill, New York, 1995.
28. D. Pinkel, T. Straume, J.W. Gray, 'Cytogenetic Analysis Using Quantitative, High-sensitivity Fluorescence Hybridization', *Proc. Natl. Acad. Sci. USA*, **83**, 2934–2938 (1986).
29. A.K. Raap, M.P.C. van de Corput, R.A.W. Vervenne, R.P.M. van Gijlswijk, H.J. Tanke, J. Wiegant, 'Ultrasensitive FISH Using Peroxidase-mediated Deposition of Biotin- or Fluorochrome Tyramides', *Hum. Mol. Genet.*, **4**, 529–534 (1995).
30. A. Baldini, D.C. Ward, 'In Situ Hybridization Banding of Human Chromosomes with Alu-PCR Products: A Simultaneous Karyotype for Gene Mapping Studies', *Genomics*, **4**, 770–774 (1991).
31. M. Baumgartner, E. Viegas-Péquignot, F. Hoffschir, M. Ricoul, A. Bravard, B. Dutrillaux, 'Detection of Translocations of 10p by Non-radioactive In Situ Hybridization of VIM Gene in SV40-Transformed Human Cell Lines', *Cancer Genet. Cytogenet.*, **56**, 23–29 (1991).
32. D. Schweizer, 'Reverse Fluorescent Chromosome Banding with Chromomycin and DAPI', *Chromosoma*, **58**, 307–324 (1976).
33. D. Cherif, C. Julier, O. Delattre, J. Derre, G.M. Lathrop, R. Berger, 'Simultaneous Localization of Cosmids and Chromosome R-Banding by Fluorescence Microscopy: Application to Regional Mapping of Human Chromosome 11', *Proc. Natl. Acad. Sci. USA*, **87**, 6639–6643 (1990).
34. P. Lichter, C.J. Tang, K. Call, G. Hermanson, G.A. Evans, D. Housman, D.C. Ward, 'High-resolution Mapping of Human Chromosome 11 by In Situ Hybridization with Cosmid Clones', *Science*, **247**, 64–69 (1990).
35. M.R. Speicher, G.S. Ballard, D.C. Ward, 'Karyotyping Human Chromosomes by Combinatorial Multi-color FISH', *Nat. Genet.*, **12**, 368–375 (1996).
36. E. Schröck, S. du Manoir, T. Veldman, B. Schoell, J. Wienberg, M.A. Ferguson-Smith, Y. Ning, D.H. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini, T. Ried, 'Multicolor Spectral Karyotyping of Human Chromosomes', *Science*, **273**, 494–497 (1996).
37. Z. Malik, D. Cabib, R.A. Buckwald, A. Talmi, Y. Garini, S.G. Lipson, 'Fourier Transform Multipixel Spectroscopy for Quantitative Cytology', *J. Microscopy*, **182**, 133–140 (1996).
38. H.J. Tanke, J. Wiegant, R.P.M. van Gijlswijk, V. Bezrookove, H. Pattenier, R.J. Heetebrij, E.G. Talman, A.K. Raap, J. Vrolijk, 'New Strategy for Multi-color Fluorescence In Situ Hybridisation: COBRA: Combined Binary Ratio Labeling', *Eur. J. Hum. Genet.*, **7**, 2–11 (1999).
39. S. Du Manoir, M.R. Speicher, S. Joos, E. Schröck, S. Popp, H. Döhner, G. Kovacs, M. Robert-Nicoud, P. Lichter, T. Cremer, 'Detection of Complete and Partial Chromosome Gains and Losses by Comparative Genomic In-situ Hybridization', *Hum. Genet.*, **90**, 590–610 (1993).
40. T. Maniatis, E.F. Fritsch, J. Sambrook, *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986.
41. M.R. Speicher, S. du Manoir, E. Schröck, H. Holtgreve-Grez, B. Schoell, C. Lengauer, T. Cremer, T. Ried, 'Molecular Cytogenetic Analysis of Formalin-embedded Solid Tumors by Comparative Genomic Hybridization after Universal DNA Amplification', *Hum. Mol. Genet.*, **11**, 1907–1914 (1993).
42. B. Trask, D. Pinkel, G. van den Engh, 'The Proximity of DNA Sequences in Interphase Cell Nuclei is Correlated to Genomic Distance and Permits Ordering of Cosmids Spanning 250 Kilobase Pairs', *Genomics*, **5**, 710–717 (1989).
43. G. van den Engh, R. Sachs, B. Trask, 'Estimating Genomic Distance from DNA Sequence Location in Cell Nuclei by a Random Walk Model', *Science*, **257**, 1410–1412 (1992).
44. S. Selig, K. Okumura, D.C. Ward, H. Cedar, 'Delineation of Replication Time Zones by Fluorescence In Situ Hybridization', *EMBO J.*, **11**, 1217–1225 (1992).
45. D. Kitsberg, S. Selig, M. Brandeis, I. Simon, I. Keshet, D.J. Driscoll, R.D. Nicholls, H. Cedar, 'Allele-specific Replication Timing of Imprinted Gene Regions', *Nature*, **264**, 459–463 (1993).
46. T. Haaf, D.C. Ward, 'Structural Analysis of α-Satellite DNA and Centromere Proteins Using Extended Chromatin and Chromosomes', *Hum. Mol. Genet.*, **3**, 697–709 (1994).
47. I. Parra, B. Windle, 'High Resolution Visual Mapping of Stretched DNA by Fluorescent Hybridization', *Nat. Genet.*, **5**, 17–21 (1993).
48. M. Heiskanen, R. Karhu, E. Hellsten, L. Peltonen, O.P. Kallioniemi, A. Palotie, 'High Resolution Mapping Using Fluorescence In Situ Hybridization to Extended DNA Fibers Prepared From Agarose-embedded Cells', *BioTechniques*, **17**, 928–933 (1994).
49. A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot, D. Bensimon, 'Alignment and Sensitive Detection of DNA by a Moving Interface', *Science*, **265**, 2096–2098 (1994).

50. H.-U.G. Weier, M. Wang, J.C. Mullikin, Y. Zhu, J.-F. Cheng, K.M. Greulich, A. Bensimon, J.W. Gray, 'Quantitative DNA Fiber Mapping', *Hum. Mol. Genet.*, **4**, 1903–1910 (1995).
51. X. Michalet, R. Ekong, F. Fougerousse, S. Rousseaux, C. Schurra, N. Hornigold, M. Vanslegtenhorst, J. Wolfe, S. Povey, J.S. Beckmann, A. Bensimon, 'Dynamic Molecular Combing: Stretching the Whole Human Genome for High-resolution Studies', *Science*, **277**, 1518–1523 (1997).
52. T. Haaf, D.C. Ward, 'High Resolution Ordering of YAC Contigs Using Extended Chromatin and Chromosomes', *Hum. Mol. Gen.*, **3**, 629–633 (1994).
53. J. Flint, A. Wilkie, V. Buckle, R. Winter, A. Holland, H. McDermid, 'The Detection of Subtelomeric Chromosomal Rearrangements in Idiopathic Mental Retardation', *Nat. Genet.*, **9**, 132–139 (1995).
54. P.M. Nederlof, S. van der Flier, A.K. Raap, H.J. Tanke, M. van der Ploeg, F. Kornips, J.P.M. Geraedts, 'Detection of Chromosome Aberrations in Interphase Nuclei by Nonradioactive In Situ Hybridization', *Cancer Genet. Cytogenet.*, **42**, 87–98 (1989).
55. R.H. Martin, A. Rademaker, 'The Frequency of Aneuploidy Among Individual Chromosomes in 6821 Human Sperm Complements', *Hum. Genet.*, **53**, 103–107 (1990).
56. S. Munné, A. Lee, Z. Rosenwaks, J. Grifo, J. Cohen, 'Diagnostics of Major Chromosome Aneuploidies in Human Preimplantation Embryos', *Hum. Reprod.*, **8**, 2185–2191 (1993).
57. B.E. Ward, S.L. Gersen, M.P. Carelli, N.M. McGuire, W.R. Dackowski, M. Weinstein, C. Sandlin, R. Warren, K.W. Klinger, 'Rapid Prenatal Diagnosis of Chromosomal Aneuploidies by Fluorescence In Situ Hybridization: Clinical Experience with 4500 Specimens', *Am. J. Hum. Genet.*, **52**, 854–865 (1993).
58. E.P.J. Arnoldus, J. Wiegant, I.A. Noordermeer, J.W. Wessels, G.C. Beerstock, G.C. Grosveld, M. van der Ploeg, A.K. Raap, 'Detection of the Philadelphia Chromosome in Interphase Nuclei', *Cytogenet. Cell Genet.*, **54**, 108–111 (1990).
59. S.J. O'Brien, J.E. Womack, L.A. Lyons, K.J. Moore, N.A. Jenkins, N.G. Copeland, 'Anchored Reference Loci for Comparative Genome Mapping in Mammals', *Nat. Genet.*, **3**, 103–112 (1993).
60. I. Nanda, Z. Shan, M. Schartl, D.W. Burt, M. Koehler, H.G. Nothwang, F. Grützner, I.R. Paton, D. Windsor, I. Dunn, W. Engel, P. Staeheli, S. Mizuno, T. Haaf, M. Schmid, '300 Million Years of Conserved Synteny Between Chicken Z and Human Chromosome 9', *Nat. Genet.*, **21**, 258–259 (1999).
61. J. Wienberg, R. Stanyon, 'Comparative Painting of Mammalian Chromosomes', *Curr. Opin. Genet. Dev.*, **7**, 784–791 (1997).
62. B.P. Chowdary, T. Raudsepp, L. Froenicke, H. Scherthan, 'Emerging Patterns of Comparative Genome Organization in Some Mammalian Species as Revealed by Zoo-FISH', *Genome Res.*, **8**, 577–589 (1998).
63. F. Grützner, H. Himmelbauer, M. Paulsen, H.-H. Ropers, T. Haaf, 'Comparative Mapping of Mouse and Rat Chromosomes by Fluorescence In Situ Hybridization', *Genomics*, **55**, 306–313 (1999).

Genome Physical Mapping Using BACs

Thomas Altmann
Max-Planck-Institut für Molekulare Pflanzenphysiologie, Golm, Germany

1 Introduction **1**
2 DNA Cloning Systems Applicable to Physical Mapping Experiments **2**
2.1 Cosmids 2
2.2 Bacteriophage P1 2
2.3 Yeast Artificial Chromosomes 3
2.4 Bacterial Artificial Chromosomes and P1-derived Artificial Chromosomes 3
3 Approaches to Physical Mapping **5**
3.1 Hybridization-based Procedures 5
3.2 Methods Based on Polymerase Chain Reaction 7
3.3 DNA Fingerprint Analysis 10
3.4 Clone (End-) Sequencing (Sequence-tagged Connector Approach) 10
4 BAC-based Physical Mapping of the *Arabidopsis thaliana* Genome **10**
Abbreviations and Acronyms **12**
Related Articles **12**
References **12**

A physical map of a region or the entirety of the genome of an organism is a representation of the genome providing information on the relative order and the arrangement of physically defined entities. For most physical maps, these entities are cloned genomic DNA fragments propagated in cosmids, bacteriophage P1, bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs), or yeast artificial chromosomes (YACs). The various cloning systems differ in the sizes of the cloned DNA fragments and the fidelity of DNA structure maintenance. Owing to its excellent properties, the BAC system is the premier system for this purpose. Several different, highly efficient procedures have been developed and applied successfully to identify overlaps between the DNA fragments represented in different clones. The procedures are based on different experimental techniques, including nucleic acid hybridization, DNA amplification through polymerase chain reaction (PCR), DNA restriction fragment analysis, and DNA sequencing. Applied separately or in combination, these procedures have been used successfully to generate extensive physical maps of human, animal, and plant genomes.

1 INTRODUCTION

Physical mapping refers to an experimental approach devoted to the establishment of a representation of the genome (the total complement of genetic information, represented by the cellular DNA) of a particular organism. The representation thus created is called a physical map. This map of the genome, which must not be mixed up with a genetic map (see below), is based on physically defined entities such as DNA restriction fragments or cloned DNA fragments. It provides information on the relative order and arrangement of these entities. Partial or complete physical maps may be constructed on the basis of DNA restriction fragment patterns or through in situ hybridization of isolated (cloned) DNA fragments to chromosomes. Most physical maps of genomes hitherto assembled, however, are based on cloned fragments of genomic DNA maintained in clones of genomic libraries (e.g. cosmid, YAC, or BAC libraries; see sections 2.1–2.3). The fully elucidated genomic DNA sequence currently available for yeast and several prokaryotic microorganisms may be regarded as the ultimate physical map. The quality of clone-based physical maps, which essentially cover the information about which clones contain overlapping genomic DNA fragments, is very strongly influenced by the properties of the cloning system used to propagate the DNA fragments. Thus, DNA cloning systems, which provide the highest degrees of maintenance of the original arrangements of the DNA molecules as they are present in the DNA of the organism under investigation and provide the highest degree of stability during propagation, are most suitable for these purposes. Owing to their specific properties, BACs are the premier system currently available for cloning and maintenance of large genomic DNA.

In contrast to a physical map, the genetic map of the genome of an organism is a rather imaginary map, which is based on the frequencies of recombination (crossing over) events occurring between genetic loci. These loci are defined either as mutations in genes that are identified through alterations in the properties of the organism (phenotypic changes) or as polymorphic DNA fragments, the sequence differences of which can be monitored directly or indirectly.

The establishment of a physical map of a genome is an intrinsic part of a detailed genome analysis. In particular, when linked to the genetic map, a high-resolution physical map is of high value for studies

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

devoted to genome organization and function, including the elucidation of the relations of physical to genetic distances and for the coordination and performance of sequencing projects. Its immediate value to the research community, however, may be even higher through the strong support it provides to map-based gene-cloning efforts. Through this procedure, genes could be isolated and studied at the molecular level that are only characterized by a corresponding mutant phenotype (alteration in morphological, developmental, metabolic, or behavioral properties of an organism caused by the genetic change). The availability of this procedure thus provides a means to gain entirely novel information on the molecular mechanisms underlying the multitude of processes occurring in living organisms.

2 DNA CLONING SYSTEMS APPLICABLE TO PHYSICAL MAPPING EXPERIMENTS

To be useful for the application of physical mapping procedures, a DNA cloning system has to meet several requirements. Higher eukaryotic organisms usually have genomes of enormous sizes, which may range from ca. 97 Mbp as in the case of the nematode *Caenorhabditis elegans*,[1] 130 Mbp in the case of the plant *Arabidopsis thaliana*,[2] ca. 165 Mbp in the case of the fruit fly *Drosophila melanogaster*,[3] ca. 430 Mbp for rice (*Oryza sativa*[4]), ca. 3 Gbp in the case of man (*Homo sapiens*[5]) and corn (*Zea mays*[6]), and up to 16 Gbp for wheat (*Triticum aestivum*[6]). Of major importance, therefore, is the size of DNA fragments, which can be cloned and stably maintained in a particular cloning system. The size of the DNA fragments determines the number of clones to be handled in order to represent (nearly) the entire genome. To provide sufficient coverage of the genome, a genomic library of a minimum of 10-fold genome representation (number of clones in the library × average insert size/genome size) should be used for a physical mapping project.[7] This extensive coverage is required to ensure a very high probability of representation of every region in the genome and to provide sufficient overlap between the individual clones to be detectable by the mapping procedure. Thus, in the case of a 100-Mbp genome, a library of ca. 100 000 clones would be necessary in the case of an average insert size of 10 kbp, a library of ca. 10 000 clones would be necessary in the case of an average insert size of 100 kbp, and a library of ca. 1000 clones would be sufficient in the case of an average insert size of 1000 kbp (= 1 Mbp). The number of clones would influence not only the effort required for ordered handling, but also the number of experiments to be performed for ordering them (see below). Although clone libraries with very large insert sizes are advantageous in terms of reducing the amount of labor for handling and processing, they are disadvantageous, on the other hand, in terms of the resolution achieved. Since robotic devices have been developed, which allow the handling of very large numbers of clones (in the range of 100 000),[8] the balance between the number of clones to be handled/number of experiments to be done and the resolution achieved may be shifted more and more toward the latter. A second important criterion is the quality of the DNA represented in the clones of the library. Various DNA cloning systems differ strongly in the fraction of aberrant clones, which may be chimeric (composed of DNA fragments derived from different, noncollinear regions of the genome that were fused upon cloning), may carry rearranged DNA fragments, or may have suffered from deletions. The latter two changes in the cloned DNA fragments may occur not only during the cloning procedure itself, but also during clone propagation. Finally, various cloning systems differ in their properties in terms of handling, including the requirements for culture and/or storage of the clones, or the procedures used for the isolation of the cloned DNA.

Several different DNA cloning systems are available; the major properties and advantages/disadvantages will be mentioned briefly in the following sections before the preferred BAC system is described in more detail.

2.1 Cosmids

Cosmids are vector systems for propagation of DNA fragments in the bacterial host *Escherichia coli*.[9] Owing to the specific cloning procedure, which involves packaging of the DNA molecules into bacteriophage λ particles, the size of inserted DNA fragments is limited to a maximum of ca. 50 kbp. Various different cosmid vectors are available, which may differ in the stability of the cloned DNA fragments. Routine procedures are available for the generation, handling and storage of cosmids.[9]

2.2 Bacteriophage P1

The bacteriophage P1-based DNA cloning system also relies on the bacterial host *E. coli* for propagation.[10] In analogy to the cosmid system, DNA molecules are packaged into phage P1 particles for delivery into the *E. coli* cells. Advantageous in comparison to the cosmids is the increased size of the packaged DNA molecules, which allows a maximum insert size of ca. 100 kbp. Although a variety of P1 libraries have been established in various laboratories, the P1 cloning procedure cannot be regarded generally as a routine.

2.3 Yeast Artificial Chromosomes

YACs are propagated in the bakers' yeast *Saccharomyces cerevisiae*[11,12] as linear molecules (chromosomes) equipped with centromeric and telomeric sequences, elements required for initiation of replication, and selectable marker genes. A variety of YAC libraries have been established for animals,[13–17] plants,[18–26] and humans,[27–30] which proved the suitability of the YAC system to clone very large DNA fragments (above 1 Mbp) and which displayed average insert sizes ranging from ca. 100 kbp up to 800 kbp. Although YACs are superior to all other DNA cloning systems in terms of the size of the inserted DNA fragments, this cloning system suffers from two major disadvantages. In many YAC libraries, very large fractions of clones (up to 40 or 50%) carry aberrant, nonauthentic DNA fragments (showing chimerism, rearrangements, deletions, and/or instability[31–33]). A very high proportion of the results obtained on YAC clones during the course of a physical mapping project may thus be erroneous and may require the investment of additional very strong efforts for their identification and elimination. The fraction of aberrant clones may, however, be reduced through the use of optimized yeast strains.[34–36] Other limitations of this system are the procedures for handling and processing of the clones, which in general are more laborious than for the *E. coli*-based systems. This is particularly true for the procedures to be applied for the isolation and purification of the cloned DNA fragments.[37]

2.4 Bacterial Artificial Chromosomes and P1-derived Artificial Chromosomes

2.4.1 Characteristics of the Systems

These two DNA cloning systems rely on the host system *E. coli*, which thus provides the opportunity to apply (almost) all routine procedures for clone propagation, handling, and storage that are usually established in molecular biologically oriented laboratories.[9] The BAC system[38] has been developed on the basis of an F-plasmid-derived replicon, which includes the factors that control the number of plasmid copies per bacterial cell to be 1–2. This system has been demonstrated to be capable of taking up and stably maintaining large DNA fragments (up to a size of ca. 350 kbp).[38] Apparently, the strict copy number control exerted by this system is the key to the highly faithful representation of the genomic DNA molecules and their very stable propagation. The developed BAC vectors, such as pBAC108L,[38] pBeloBAC11,[39] pECBAC1,[40] or pBeloBAC-Kan[41] (see Figure 1), display a variety of favorable features. These include multiple cloning sites flanked by recognition sites for rare cutting restriction enzymes (useful for the excision of the cloned DNA fragments), a lacZ-based reporter system to monitor recombinant clones, and the presence of bacteriophage T7 and Sp6 promoters flanking the cloning site (useful for the generation of labeled end-probes; see below). Specific BAC vector variants such as pBiBAC[42] or pTAC[43] have been created, which allow their use in *Agrobacterium tumefaciens*-mediated transfer of the inserted DNA into plants. This opens the attractive opportunity to transfer cloned large DNA fragments into plants, providing the means to identify rapidly novel genes or to identify and transfer quantitative trait loci. The PAC vector[44] resembles the BAC vector, but

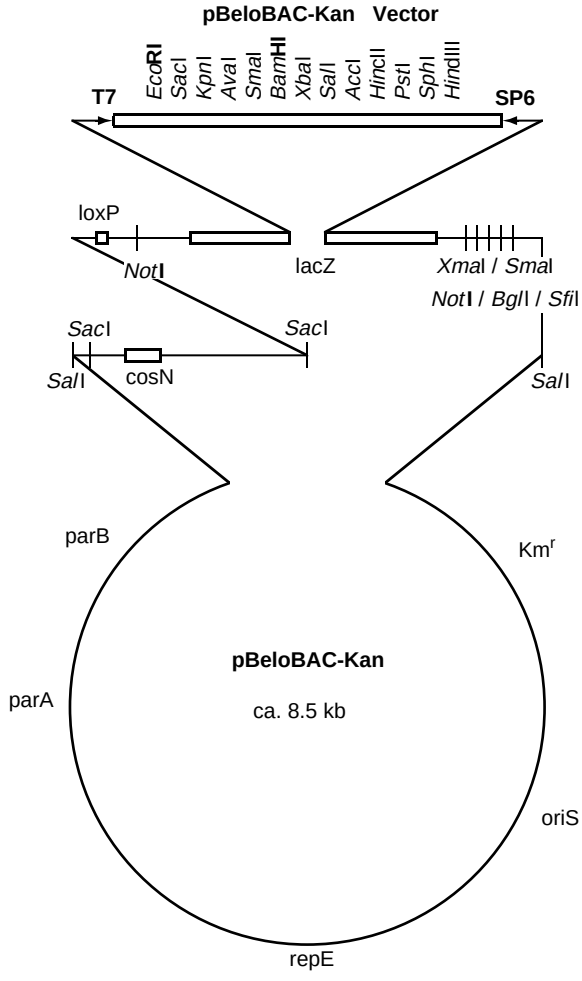

Figure 1 Schematic representation of the pBeloBAC-Kan vector. Indicated are the functional elements involved in replication and copy number control (oriS, repE, parA, parB), the selectable marker (Kanamycin resistance), and details of the multiple cloning site. The latter is flanked by the phage T7 and SP6 promoters suitable for the generation of labeled end-probes and is located within the lacZ gene that allows blue/white selection for recombinant clones. The cloning site is furthermore flanked by recognition sites for rare cutting restriction enzymes, and loxP and cosN sites are present, which may be used for the high-specificity cleavage.

is based on the P1 replicon and carries a marker for the selection for recombinant clones (sacB). In contrast to the aforementioned P1 cloning system (section 2.2), PACs, like BACs, are introduced into the *E. coli* host cells by electroporation. The sizes of the recombinant BAC or PAC molecules are therefore not limited by the capacity of phage particles for DNA uptake.

Several BAC and PAC libraries have been established recently for plants,[41,45–49] animals,[50–53] and humans,[39,44,51] with average insert sizes ranging from ca. 100 kbp up to 235 kbp. These libraries provide deep genomic coverages and easy access to high-quality genomic DNA suitable for manipulation, such as sequencing, exon trapping, complementary DNA selection, microsatellite marker isolation, and fluorescence in situ hybridization studies.[38,54–57]

2.4.2 Generation of a BAC Library

For the generation of a BAC library, several experimental steps are necessary (see Figure 2; for a detailed description of the BAC cloning procedure, see Shizuya et al.[38] and Mozo et al.[41]):

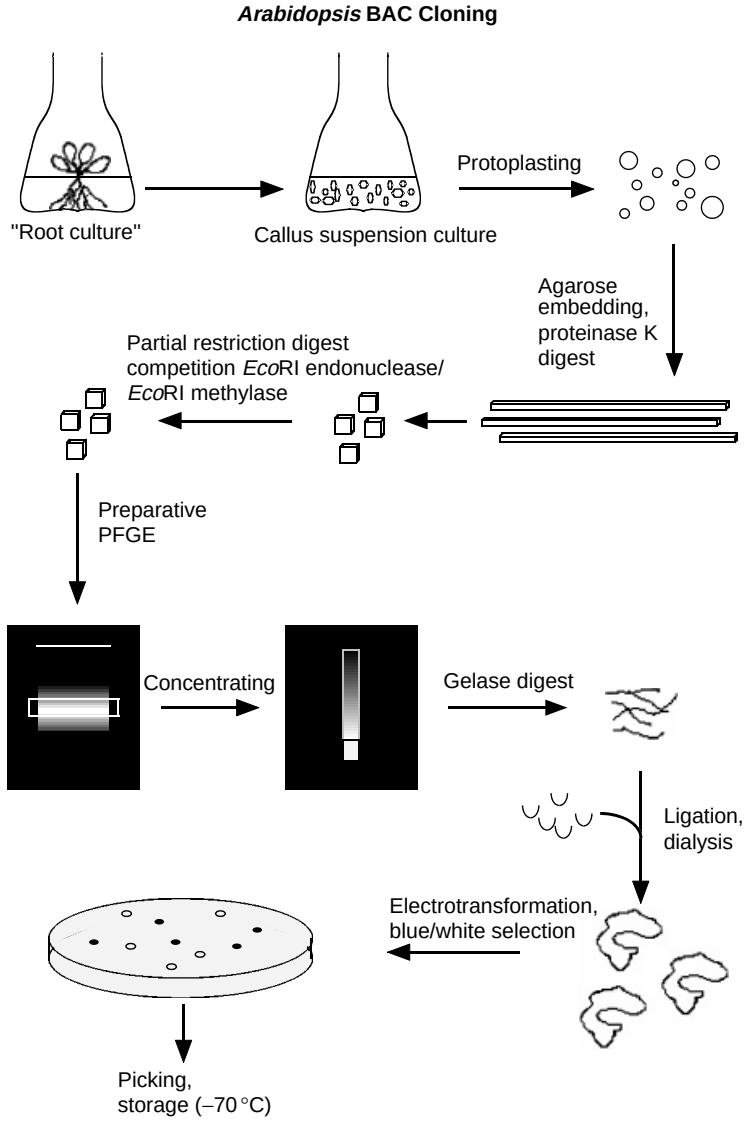

Figure 2 BAC cloning procedure applied for the generation of an *Arabidopsis thaliana* BAC library. In this case, protoplasts prepared from root-derived suspension cultures were used as a source of high-molecular-weight DNA prepared through proteinase K digestion after embedding in agarose. The high-molecular-weight DNA was partially digested by *Eco*RI restriction enzyme/methylase competition, fractionated by preparative pulsed-field gel electrophoresis (PFGE), concentrated and ligated into the pBeloBAC-Kan vector. Ligation products were transfected into *E. coli* by electroporation and white colonies grown in the presence of X-Gal/IPTG (substrate and inducer of the lacZ gene) were picked and stored at −70 °C in microtiter dishes.

1. High-quality high-molecular-weight DNA (≥1 Mbp) has to be prepared from the source organism. This may be achieved by agarose embedding of cells or protoplasts (plant cells without cell wall) or nuclei and liberation of the DNA by a proteinase digestion.
2. This DNA has to be partially digested with a suitable restriction enzyme (such as *Bam*HI, *Sau*3A, *Hin*dIII, *Eco*RI) to yield DNA fragments in the desired size range (150–250 kbp). The limited digestion can be achieved by limited incubation time, suboptimal (e.g. low temperature) incubation conditions, or by a competition reaction of restriction enzyme and a corresponding methylase.[58] The last possibility has the advantage over the first two of allowing the reaction to be run to completion, which is of relevance if the available high-molecular-weight DNA concentration is low.
3. The DNA fragments of the desired size range have to be separated from the molecules of smaller size that inevitably also occur and would be preferentially cloned subsequently. PFGE[59] is used to separate the large DNA molecules. Usually, at least two successive rounds of PFGE under different conditions are necessary to achieve sufficient purification. In the case of low DNA concentrations, a subsequent step for DNA concentration may be necessary/desirable through another electrophoresis involving an agarose gel of higher concentration.[60]
4. The insert DNA fragments thus generated are released from the agarose matrix by enzymatic digestion of the agarose and the liberated DNA molecules are ligated to dephosphorylated pBAC vector DNA.
5. The ligation products are transferred into *E.coli* cells by electroporation under optimized conditions[61] that are favorable for the uptake of large molecules. Depending on the BAC or PAC vector used, recombinant clones may be recognized through the expression of scorable (e.g. lacZ) or selectable marker systems (e.g. sacB). Recombinant clones are picked into liquid culture medium (preferably in microtiter plates) and transferred to −70 °C for long-term storage. For propagation (e.g. library replication), establishment of colony filters for DNA hybridization experiments (see below), or BAC DNA preparations, standard *E. coli* handling procedures can be applied.[9] To verify the presence of high-quality recombinant clones, a set of randomly selected clones should be tested for insert sizes by BAC DNA isolation, digestion with an appropriate rare cutting restriction enzyme (e.g. *Not*I) and PFGE.

3 APPROACHES TO PHYSICAL MAPPING

Various techniques are available for identification of overlapping clones, which essentially is the information to be gathered during a physical mapping program. To this end, experiments are designed to group efficiently as many clones as possible into contigs (contiguous sequences). Several procedures have been developed that rely on different types of information to detect the presence of common DNA regions in different clones.

3.1 Hybridization-based Procedures

The property of DNA molecules to be composed of two complementary strands is used in hybridization experiments. Thus, DNA molecules can be identified through hybridization to carry identical or highly similar sequences. Screening of a clone library by hybridization is usually done after fixation of DNA liberated from the clones to a solid support, e.g. a nylon membrane (see Figure 3). The immobilized DNA is then incubated with a labeled probe (a certain species of radioactively, fluorescently, or digoxygenin-labeled DNA molecules) under appropriate conditions.[9] Through the detection of the label at specific positions of the filter, the identity of the corresponding (overlapping) clone can be determined. It is particularly advantageous when applying such a procedure to use gridded libraries, the clones of which are fixed to defined positions on the filter.[8] In this case, large numbers of hybridizations can be performed efficiently and the results of the hybridizations can be evaluated rapidly.[62] This approach has gained strong support by the development of robotic devices, which provide the means to fix at high precision very large numbers of clones at high density to membranes that can be produced in large quantities.[8] The labeled probes used may be derived from individual selected clones (''chromosome walking''[63]), they may be oligonucleotides (''hybridization fingerprinting''[64,65]), or they may be genetically mapped markers (''anchoring''[66]). Relatively short stretches of sequence identity are sufficient to detect overlaps by hybridization, which is thus useful also for genomic regions of sparse coverage in the libraries used.

The information gathered by hybridization, however, does not provide any indication of the degree of overlap between clones. Such information may be obtained, for example, by application of the DNA fingerprinting technique (see section 3.3). Erroneous hybridization results may be generated through the appearance of highly related DNA sequences present at different regions in the genome. The specific properties of the genome investigated therefore strongly influence the viability of a hybridization-based mapping approach. Because all

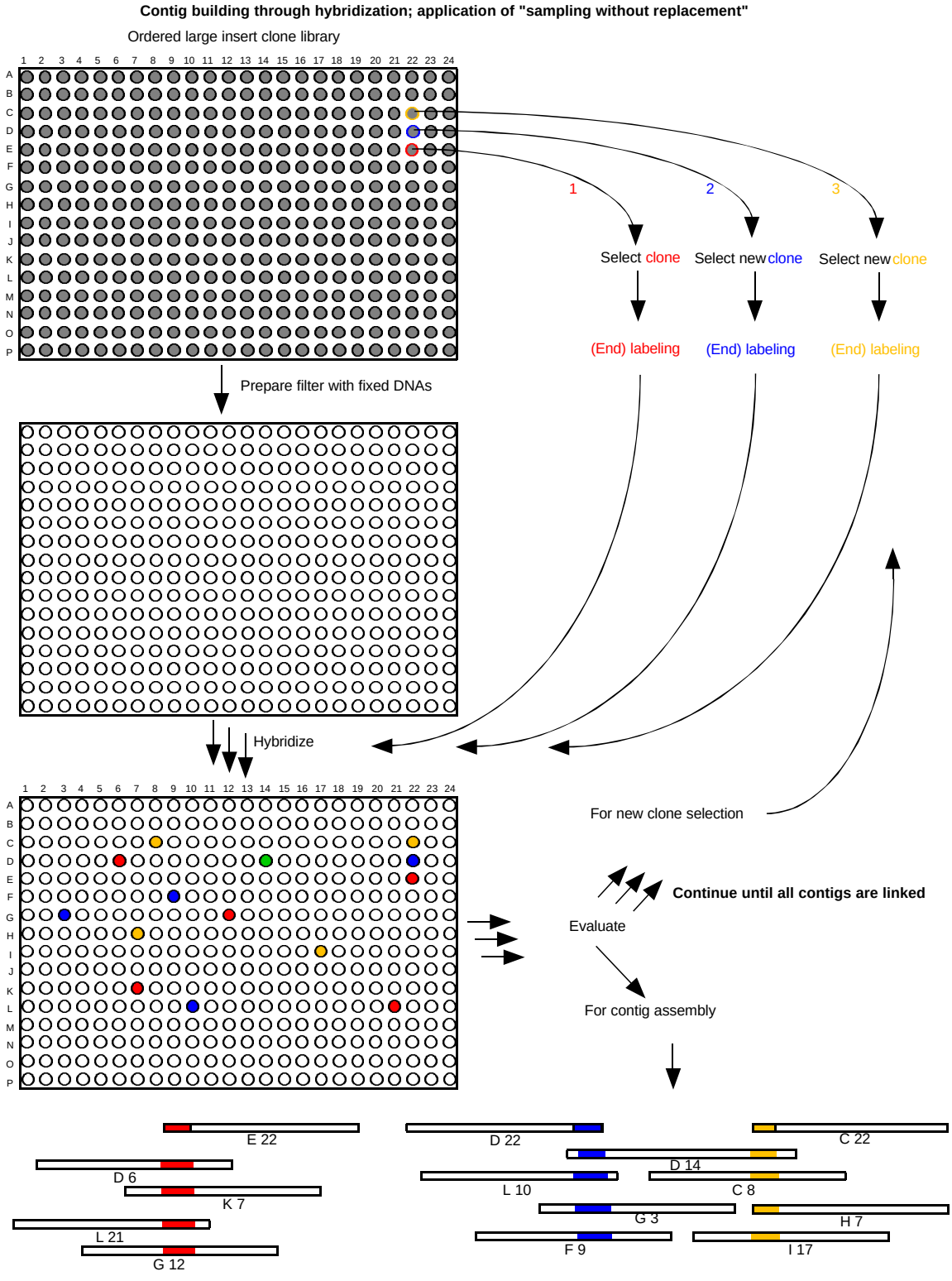

Figure 3 Physical mapping through hybridization. To identify overlap between clones of an ordered large insert library, labeled DNA probes (e.g. derived from clones of the library) are hybridized to the DNAs of the clones fixed to a solid support (e.g. a nylon membrane). The procedure is most efficiently performed in an iterative manner (‘‘sampling without replacement’’) by preparing new labeled probes only from clones that have not been hit by any previous hybridization. Established contigs eventually merge through clones hit by more than one probe. Here, D14 is a connecting clone; it is marked in green on the filter because it was identified by the probe derived from clone D22 (marked in blue) and by the probe derived from clone C22 (marked in yellow).

clones of the library are checked for overlap with the probe at the very same time, hybridization is intrinsically highly parallel and it has an extremely high throughput. In addition, it readily allows the integration of data generated by the application of different types of probes, such as clones from different libraries (e.g. cosmid, BAC, or YAC clones), total labeled clones or clone end-probes, or mapped markers. The use of the latter type of probe is highly desirable because it would result in an alignment of the physical map with the genetic map.[67]

An efficient procedure for the generation of contigs (physical mapping) through iterative hybridization has been developed and applied successfully to generate high-resolution cosmid and P1 maps of *Schizosaccharomyces pombe*.[7] The strategy of ''sampling without replacement'' (see Figure 3) was applied here, which involves the selection of clones to be labeled and hybridized to the complete genomic libraries from only the subset of clones never hit by any of the previous hybridizations. This procedure greatly reduces the number of hybridization reactions that have to be carried out in comparison to random probe selection. A modification of this procedure involves the generation of hybridization probes from both ends of the selected clones.[68] In addition to generating larger contigs than through the use of single end-probes, the double end-probe procedure provides information on the relative orientation of the identified clones. The latter approach recently was applied successfully to the generation of the first BAC-based physical map[89,70] (see section 4).

The efficiency of the ''sampling without replacement'' procedure is limited by the frequency of repetitive sequences present in the cloned DNA fragments, which may comprise very substantial fractions of large eukaryotic genomes. The presence of such repetitive sequences within the labeled probes would cause the generation of noninformative results. In this situation it is advisable to use oligonucleotides as probes or to select DNA probes from sources enriched for single/low-copy sequences such as complementary DNAs, identified as expressed sequence tags (ESTs).

3.2 Methods Based on Polymerase Chain Reaction

Sequence-specific DNA amplification via PCR provides another efficient means to test recombinant clones for the presence of a particular piece of DNA. Thus, specific pairs of oligonucleotide primers designed according to the sequences of known stretches of DNA are applied in large sets of amplification reactions containing DNAs isolated from the various different clones of a genomic library as templates. Because DNA amplification to detectable levels would only occur in those reactions supplied with template DNA derived from clones carrying the cognate DNA region, PCR can be used as a very sensitive and reliable assay for the presence/absence of specific DNA sequences, even in complex mixtures. The DNA regions of known sequence have been termed ''sequence-tagged sites'' (STSs)[71] and the corresponding mapping procedure is known as ''STS content mapping''[72] (see Figure 4).

The major advantage of STS content mapping over the use of cloned DNA fragments, e.g. as hybridization probes, is the immediate access to any STS by any interested laboratory as soon as it is made known. An STS can be defined completely as information (essentially a short stretch of DNA sequence plus instruction for the PCR assay) that can easily be deposited in a database. No access to specific biological material such as a cloned DNA fragment, which initially might have led to the identification of an STS, is required to perform an STS assay. This ''common language'' provides an ideal basis for collaborative approaches between many geographically dispersed groups, as may frequently be the case for mapping projects devoted to the analysis of very large and complex genomes. A further advantage of STS content mapping is its applicability even to genomic regions rich in repetitive sequences, provided that these sequences are interspersed with single-copy sequences that can be used for primer design.

Disadvantages of STS content mapping include the requirement for quite large amounts of sequence information to be available prior to the initiation of a mapping project or to be produced during the conductance of the project. The PCR method itself bears the risk of producing false negative (lack of DNA amplification in cases where it should occur) and false positive results (DNA amplification in cases where it should not occur). These need to be identified through proper controls and may require additional efforts for the optimization of PCR conditions. Owing to the need for performance and evaluation of very many DNA amplification reactions (see below), this procedure is very labor- and cost-intensive.

STS content mapping usually involves the isolation of DNA from a specific subset or, for all clones, from an ordered large insert library, and subjecting these DNAs to individual STS assays that include PCR amplification and evaluation of the PCR products. Because the clone libraries usually consist of several tens of thousands of clones and several thousand STSs have to be tested for the construction of genome-wide maps,[73,74] several million STS assays may have to be performed. It is therefore almost inevitable that appropriate clone pooling strategies are used, which provide an efficient means to reduce the number of STS assays required but still allow the identification of individual positive clones (see Figure 4). To be able to process the enormous numbers of STS assays ($\geq$15 million

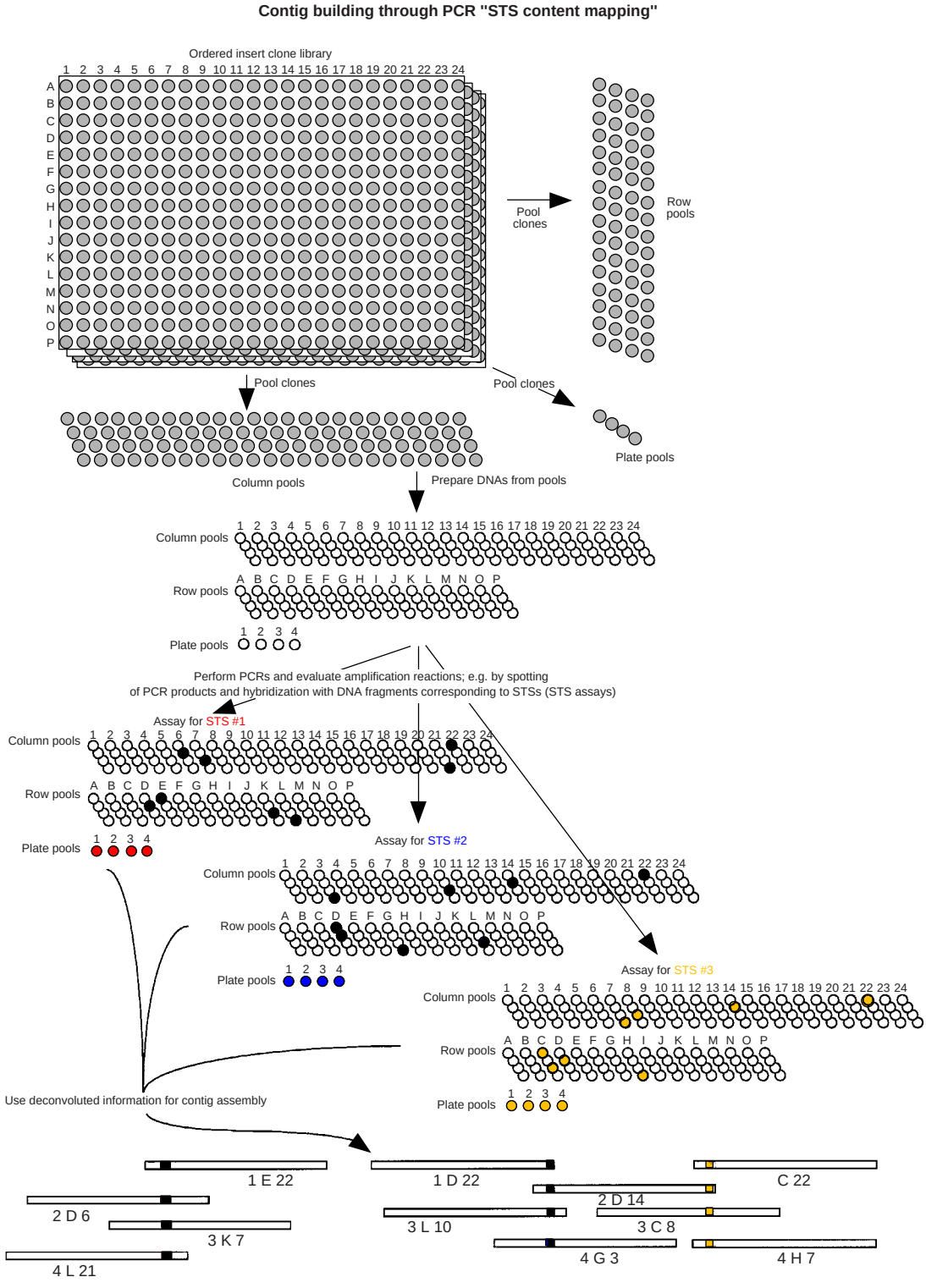

Figure 4 Physical mapping through application of PCR. After multidimensional pooling of clones of an ordered large insert clone library, DNA prepared from the clone pools is subjected to PCR using single-copy-sequence (STS)-specific oligonucleotide primers. Successful amplification highlights the presence of a clone within the positive pool that carries the particular DNA sequence. The information obtained on the various pools is deconvoluted to identify individual positive clones. In the case of STS 1 (marked red), a positive signal was detected in column pool 22 and row pool E of plate 1. The positive clone therefore is clone E 22 from plate 1 (1 E 22). Clone contigs are established according to the presence of common STSs; contigs eventually merge through clones hit by more than one STS.

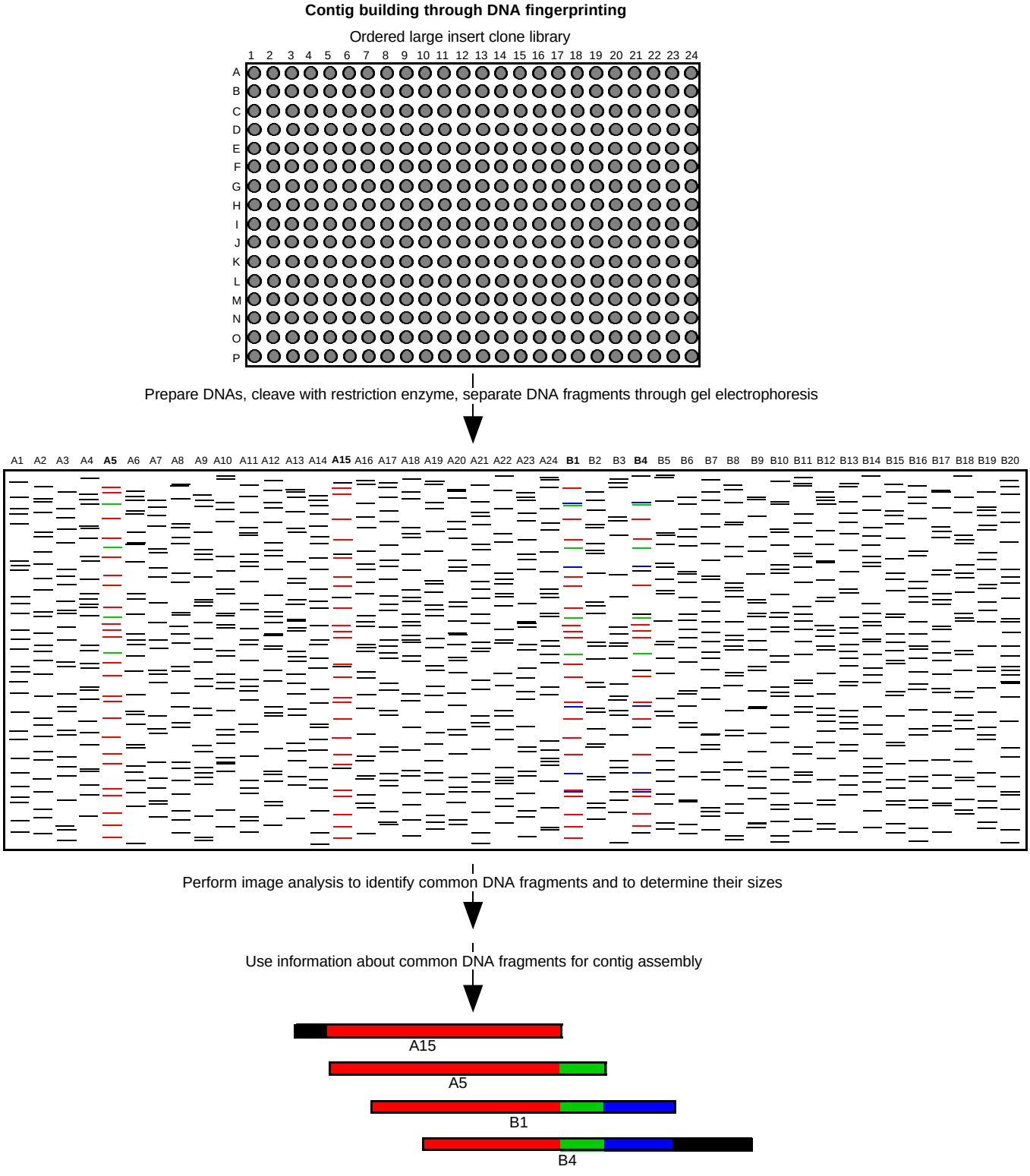

Figure 5 Physical mapping through DNA fingerprinting. DNA isolated individually from all clones of an ordered large insert clone library is treated with a restriction endonuclease and the resulting DNA fragments are separated according to their size by gel electrophoresis. Overlap between clones is detected through the presence of common bands in the gel image, which correspond to common DNA fragments. Contigs are established according to the information about common DNA fragments (marked in red, green, and blue) carried by the clones.

have been performed during the construction of a human STS-based map,[73] and ≥17 million for establishing a map of the mouse genome[74]), large-scale automation systems are required. Analogous to the hybridization-based approaches, specific optimized procedures for selection of the STSs[68] may be applied to limit the number of necessary STS assays. Nonrandom selection of STSs was applied successfully for the generation of a physical map of the *Drosophila* euchromatic genome.[75] In addition to the aforementioned physical maps of entire genomes, a large number of YAC-, P1-, PAC-, or BAC-based maps for specific chromosomes or chromosomal regions have been established through involvement of STS content mapping.[76–83]

3.3 DNA Fingerprint Analysis

The property of DNA restriction enzymes to cleave DNA molecules at specific recognition sequences provides the basis for another type of procedure for the recognition of the presence of common (identical) DNA stretches: when digested with an appropriate restriction enzyme, DNA molecules are cleaved into a specific pattern of DNA fragments that form a characteristic pattern of bands when separated by gel electrophoresis and visualized (DNA fingerprint[84]). In the case of two clones sharing a common stretch of DNA, this would give rise to a number of common DNA fragments that can thus be used to detect overlaps (see Figure 5). The probability of detection of true overlaps by this procedure is determined by the number of common fragments and the size of the analyzed genome. This procedure has been applied successfully for mapping all or parts of the genomes of *Caenorhabditis elegans*,[84,85] yeast,[86,87] *E. coli*,[88] *Drosophila melanogaster*,[89] and humans.[30,90–92]

In comparison to hybridization, much more extended regions of sequence identity are required to provide a sufficient number of common DNA fragments for the exclusion of similarly sized DNA fragments appearing by chance. Therefore, a high redundancy of the clone libraries is an essential prerequisite and very many clones have to be fingerprinted to generate a complete physical map.[30] Furthermore, analytical systems that permit precise and reproducible DNA fragment sizing and data acquisition in a high-throughput fashion are necessary. Automated methods are therefore required to pursue this approach, which has been further developed to be applicable to very-high-throughput analysis.[93] In contrast to hybridization, however, DNA fingerprinting is less prone to producing erroneous results caused by the appearance of highly related DNA sequences present at different regions in the genome. Furthermore, the results of a DNA fingerprint analysis provide information on the extent of overlap of clones derived from the number and the sizes of common DNA fragments. This information is of particular relevance for the selection of a minimal set of clones spanning a certain region, e.g. for sequencing.

3.4 Clone (End-) Sequencing (Sequence-tagged Connector Approach)

As a fourth type of approach and a complement to the DNA fingerprinting approach, Venter et al.[94] proposed a procedure for mapping clones primarily based on sequence analysis, which has come within reach through the enormous increases in DNA sequencing capacity. The principle underlying this approach is the identification of overlap between clones directly through the identification of a common DNA sequence (see Figure 6). To this end, a large collection of BAC clones, such as a BAC library representing an average 15-fold genome representation, is analyzed by sequencing both ends of the inserted DNA fragments and a corresponding end-sequence database is established. Starting with a ''seed-BAC'' that is completely sequenced, all overlapping BACs are recognized through a stretch of common DNA identified upon search of the end-sequence database. This approach is cost effective, especially for very large genome sequence analysis projects, and has proven to be highly viable, e.g. for the *Arabidopsis thaliana* sequencing project (S. Rounsley, personal communication). Because it relies rather exclusively on sequence analysis and the physical map is actually developed through the complete genomic sequencing, it is only applicable to mapping genomes that are intended to be fully sequenced.

4 BAC-BASED PHYSICAL MAPPING OF THE *ARABIDOPSIS THALIANA* GENOME

Arabidopsis is a small flowering plant that today serves as the major model system in plant molecular genetics[95] and that has been chosen as the subject of the first large-scale project intended to determine the full genome sequence of a plant.[96] This sequencing project, together with the increasing interest in map-based gene cloning, has highlighted the requirement for a complete and accurate physical map of this plant species. Two complementary approaches have been initiated to achieve this goal.[69,97] The final merging of the data produced by these two approaches recently led to the establishment of the first reported map of a complex organism based entirely on BAC clones. This map represents the most homogeneous and complete physical map established to date for any plant genome.[70]

In the first of the two approaches, the aforementioned ''sampling without replacement'' procedure of high-throughput hybridizations of BAC end-probes was

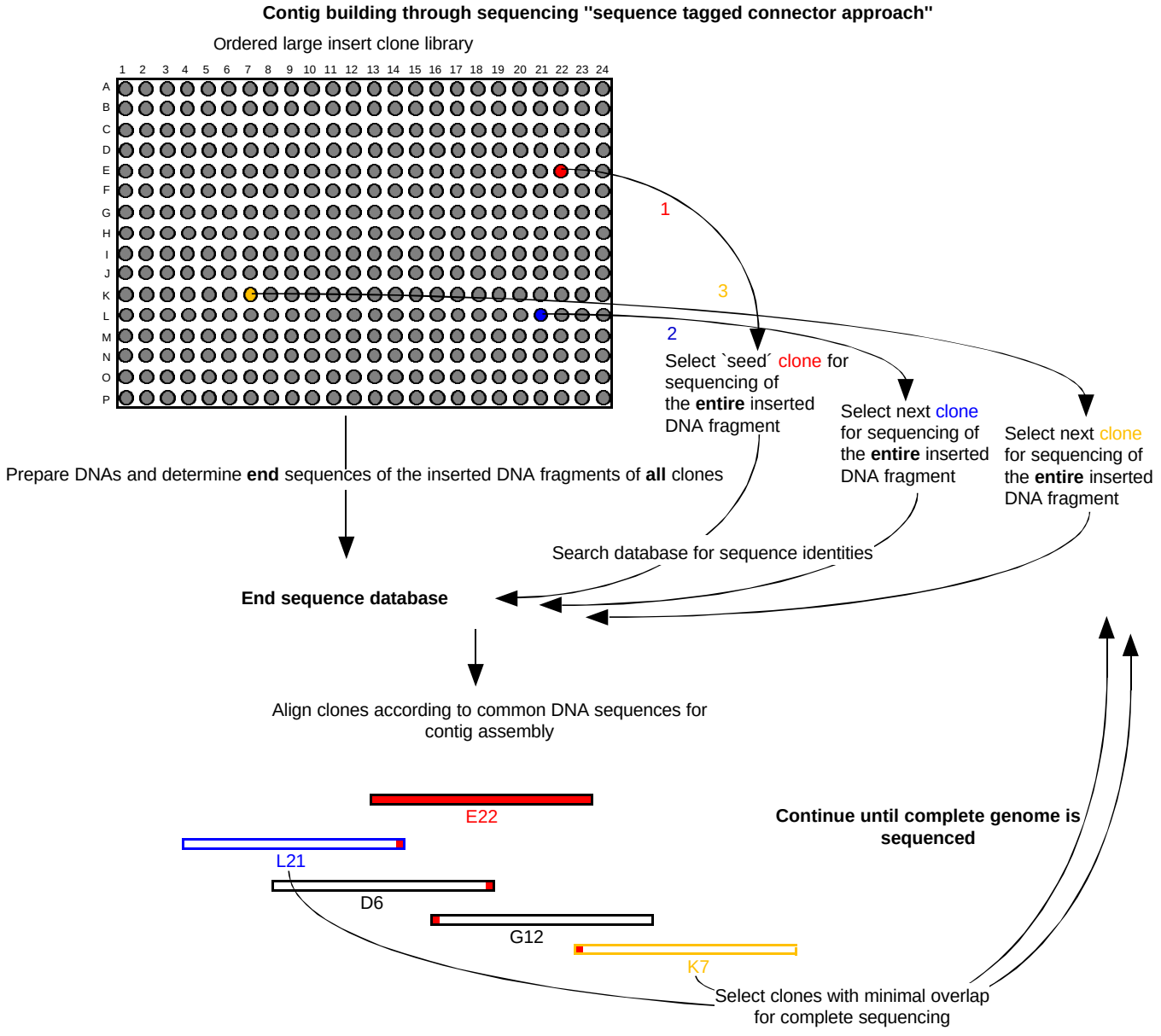

Figure 6 Physical mapping through DNA sequencing. The DNA sequences located at the ends of the cloned DNA fragments of every clone of an ordered large insert clone library are determined and deposited in a database. Thereafter, an individual clone (“seed clone”) is selected for complete sequencing. This sequence information is used to search the end sequence database for sequence identities. Matches between the sequence of the “seed clone” and end sequences of other clones demonstrate the presence of common DNA regions within the cloned DNA fragment of the various clones and provide the means to determine precisely the degrees of overlap. Clones with minimal overlap to the “seed clone” are selected for sequencing of the entire inserted DNA fragment and the elucidated sequence is used again to query the end sequence database. Through this procedure, a clone-based physical map of a genome is established simultaneously with the determination of the entire genome sequence.

applied because it was shown to work efficiently for this organism.[69] The nuclear DNA of *Arabidopsis thaliana*, which is estimated to cover ca. 130 Mbp, is characterized by a low content of interspersed repetitive sequences.[98] It is therefore particularly well suited for the application of a hybridization-based physical mapping approach. When evaluated on the basis of 1000 end-probe hybridization experiments, the applied “double-end-probe sampling without replacement” approach was found to be highly efficient, with about 50% of all mappable clones identified and assembled into a minimum of 220 contigs.[69] However, the results obtained deviated slightly from theoretical predictions[68] in that a lower than expected fraction of the genome was covered by the assembled

contigs. This result might have been caused by a non-random genome representation in the BAC library used. An intriguing 200-kb periodicity of over- and under-represented regions has been observed in an independent study.[99] A total of more than 3000 hybridization experiments was performed on a 10 752 clone BAC library[41] representing ca. 7 genome equivalents. The results of 2723 hybridization experiments were included in the final map, which were complemented by 272 genetic marker data.[70]

In the second approach, a high-throughput BAC DNA fingerprint analysis was performed on a set of 21 403 *Arabidopsis thaliana* BAC clones, of which 20 206 successfully yielded fingerprint data.[97] Using the FPC software[100] for automated assembly of contigs based on the fingerprint data and by manual editing, a total of 169 contigs containing at least 3 BAC clones and another 124 contigs containing 2 BAC clones were assembled. These 293 contigs contained 19 909 BAC clones.

To assemble the final map, the results obtained by the hybridization experiments were merged with the BAC fingerprint data by querying the latter for the presence of BAC clones not identified in the hybridization experiments and for clones closing gaps between the previously established contigs. Ninety-nine connections shown in the map were inferred exclusively from this information. In addition, the fingerprint data were particularly useful to solve conflicting hybridization data (due to cross-hybridization) and to assemble and extend contigs of clones containing repetitive sequences. The map (displayed at http://www.mpimp-golm.mpg.de/101/mpi_mp_map/access.html) consists of 27 contigs that cover the entire *Arabidopsis thaliana* nuclear genome, except for the presumptive centromeric regions, nucleolar organization regions, and telomeric areas. The 27 contigs are distributed over the 10 chromosome arms leaving 14 gaps, 12 of which are bridged by YAC, P1 or transformation-competent artificial chromosome clones according to other maps already established for chromosomes II–V.[101–106]

This report demonstrates the strong advantage of having available two large sets of data generated by the use of different techniques and approaches (which proved to be complementary rather than redundant) to achieve the goal of rapidly and efficiently establishing a complete physical map of a complex genome.

ABBREVIATIONS AND ACRONYMS

BAC	Bacterial Artificial Chromosome
EST	Expressed Sequence Tag
PAC	P1-derived Artificial Chromosome
PCR	Polymerase Chain Reaction
PFGE	Pulsed-field Gel Electrophoresis
STS	Sequence-tagged Site
YAC	Yeast Artificial Chromosome

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Probes • Optical Mapping in Genomic Analysis • Polymerase Chain Reaction and Other Amplification Systems • Restriction Landmark Genomic and cDNA Scanning • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

REFERENCES

1. R. Wilson, 'How the Worm was Won', *Trends Genet.*, **15**, 51–58 (1999).
2. F. Creusot, E. Fouilloux, M. Dron, J. Lafleuriel, G. Picard, A. Billault, D. Le Paslier, D. Cohen, M.-E. Chabouté, A. Durr, J. Fleck, C. Gigot, C. Camilleri, C. Bellini, M. Caboche, D. Bouchez, 'The CIC Library: a Large Insert YAC Library for Genome Mapping in *Arabidopsis thaliana*', *Plant J.*, **8**, 763–770 (1995).
3. G.M. Rubin, 'The Drosophila Genome Project: a Progress Report', *Trends Genet.*, **14**, 340–343 (1998).
4. T. Izawa, K. Shimamoto, 'Becoming a Model Plant: the Importance of Rice to Plant Science', *Trends Plant Sci.*, **1**, 95–99 (1996).
5. M. Olson, 'The Human Genome Project', *Proc. Natl. Acad. Sci. USA*, **90**, 4338–4344 (1993).
6. K. Arumuganathan, E.D. Earle, 'Nuclear DNA Content of Some Important Plant Species', *Plant Mol. Biol. Rep.*, **9**, 208–219 (1991).
7. J.D. Hoheisel, E. Maier, R. Mott, L. McCarthy, A.V. Grigoriev, L.C. Schalkwyk, D. Nizetic, F. Francis, H. Lehrach, 'High Resolution Cosmid and P1 Maps Spanning the 14 Mb Genome of the Fission Yeast *S. pombe*', *Cell*, **73**, 109–120 (1993).
8. E. Maier, D.R. Bancroft, H. Lehrach, 'Large-scale Library Characterization', in *Wiley–Interscience Series on Laboratory Automation: Automation Technologies for Genome Characterization*, ed. T.J. Beugelsdijk, John Wiley & Sons, New York, 65–88, 1997.
9. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Press, New York, 1989.
10. N. Sternberg, 'Bacteriophage P1 Cloning System for the Isolation, Amplification, and Recovery of DNA

Fragments as Large as 100 Kilobase Pairs', *Proc. Natl. Acad. Sci. USA*, **87**, 103–107 (1990).
11. D.T. Burke, G.F. Carle, M.V. Olson, 'Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors', *Science*, **236**, 806–812 (1987).
12. D. Schlessinger, 'Yeast Artificial Chromosomes: Tools for Mapping and Analysis of Complex Genomes', *Trends Genet.*, **6**, 248–258 (1990).
13. D. Garza, J.W. Ajioka, D.T. Burke, D.L. Hartl, 'Mapping the *Drosophila* Genome with Yeast Artificial Chromosomes', *Science*, **246**, 641–646 (1989).
14. F. Libert, A. Lefort, R. Okimoto, J. Womack, M. Georges, 'Construction of a Bovine Genomic Library of Large Yeast Artificial Chromosome Clones', *Genomics*, **18**, 270–276 (1993).
15. L. Cai, L.C. Schalkwyk, A. Schoeberlein-Stehli, R.Y.L. Zee, A. Smith, T. Haaf, M. Georges, H. Lehrach, K. Lindpaintner, 'Construction and Characterization of a 10-Genome Equivalent Yeast Artificial Chromosome Library for the Laboratory Rat, *Rattus norvegicus*', *Genomics*, **39**, 385–392 (1997).
16. H. Takeda, H. Yamakuchi, N. Ihara, K. Hara, T. Watanabe, Y. Sugimoto, T. Oshiro, H. Kishine, Y. Kano, K. Kohno, 'Construction of a Bovine Yeast Artificial Chromosome (YAC) Library', *Anim. Genet.*, **29**, 216–219 (1998).
17. T.P. Zhong, K. Kaphingst, U. Akella, M. Haldi, E.S. Lander, M.C. Fishman, 'Zebrafish Genomic Library in Yeast Artificial Chromosomes', *Genomics*, **48**, 136–138 (1998).
18. J. Ecker, 'PFGE and YAC Analysis of the *Arabidopsis* Genome', *Methods*, **1**, 186–194 (1990).
19. E.R. Ward, G.C. Jen, 'Isolation of Single-copy-sequence Clones from a Yeast Artificial Chromosome Library of Randomly-sheared *Arabidopsis thaliana* DNA', *Plant Mol. Biol.*, **14**, 561–568 (1990).
20. E. Grill, C. Somerville, 'Construction and Characterization of a Yeast Artificial Chromosome Library of *Arabidopsis* which is Suitable for Chromosome Walking', *Mol. Gen. Genet.*, **226**, 484–490 (1991).
21. G.B. Martin, M.W. Ganal, S.D. Tanksley, 'Construction of a Yeast Artificial Chromosome Library of Tomato and Identification of Cloned Segments Linked to Two Disease Resistance Loci', *Mol. Gen. Genet.*, **233**, 25–32 (1992).
22. K.J. Edwards, H. Thompson, D. Edwards, A. De Saizieu, C. Sparks, J.A. Thompson, A.J. Greenland, M. Eyers, W. Schuch, 'Construction and Characterization of a Yeast Artificial Chromosome Library Containing Three Haploid Maize Genome Equivalents', *Plant Mol. Biol.*, **19**, 299–308 (1992).
23. M. Eyers, K. Edwards, W. Schuch, 'Construction and Characterization of a Yeast Artificial Chromosome Library Containing Two Haploid *Beta vulgaris* L. Genome Equivalents', *Gene*, **121**, 195–201 (1992).
24. R.M. Klein-Lankhorst, E.M.J. Salentijn, W.G. Dirkse, M. Arens-De Reuver, W.J. Stiekema, 'Construction of a YAC Library from a *Beta vulgaris* Fragment Addition and Isolation of a Major Satellite DNA Cluster Linked to the Beet Cyst Nematode Resistance Locus *Hs1-pat-1*', *Theor. Appl. Genet.*, **89**, 426–434 (1994).
25. G. Bonnema, J. Hontelez, R. Verkerk, Y.Q. Zhang, R. Van Daelen, A. Van Kammen, P. Zabel, 'An Improved Method of Partially Digesting Plant Megabase DNA Suitable for YAC Cloning: Application to the Construction of a 5.5 Genome Equivalent YAC Library of Tomato', *Plant J.*, **9**, 125–133 (1996).
26. M. Kleine, W. Michalek, T. Diefenthal, H. Dargatz, C. Jung, 'Construction of a *Mlu*I-YAC Library from Barley (*Hordeum vulgare* L.) and Analysis of YAC Insert Terminal Regions', *Genome*, **40**, 896–902 (1997).
27. R. Anand, J.H. Riley, R. Butler, J.C. Smith, A.F. Markham, 'A 3.5 Genome Equivalent Multiaccess YAC Library: Construction, Characterization, Screening and Storage', *Nucleic Acids Res.*, **18**, 1951–1956 (1990).
28. H.M. Albertsen, H. Abderrahim, H.M. Cann, J. Dausset, D. Le Paslier, D. Cohen, 'Construction and Characterization of a Yeast Artificial Chromosome Library Containing Seven Haploid Human Genome Equivalents', *Proc. Natl. Acad. Sci. USA*, **87**, 4256–4260 (1990).
29. J. Dausset, P. Ougen, H. Abderrahim, A. Billault, J.L. Sambucy, D. Cohen, D. Le Paslier, 'The CEPH YAC Library', *Behring Inst. Mitt.*, **91**, 13–20 (1992).
30. C. Bellanné-Chantelot, B. Lacroix, P. Ougen, A. Billault, S. Beaufils, S. Bertrand, I. Georges, F. Glibert, I. Gros, G. Lucotte, 'Mapping the Whole Human Genome by Fingerprinting Yeast Artificial Chromosomes', *Cell*, **70**, 1059–1068 (1992).
31. E.D. Green, H.C. Riethman, J.E. Dutchik, M.V. Olson, 'Detection and Characterization of Chimeric Yeast Artificial-chromosome Clones', *Genomics*, **11**, 658–669 (1991).
32. D.L. Neil, A. Villasante, R.B. Fisher, D. Vetrie, B. Cox, C. Tyler-Smith, 'Structural Instability of Human Tandemly Repeated DNA Sequences Cloned in Yeast Artificial Chromosome Vectors', *Nucleic Acids Res.*, **18**, 1421–1428 (1990).
33. V. Larionov, N. Kouprina, N. Nikolaishvili, M.A. Resnik, 'Recombination during Transformation as a Source of Chimeric Mammalian Artificial Chromosomes in Yeast (YACs)', *Nucleic Acids Res.*, **22**, 4154–4162 (1994).
34. F.L. Chartier, J.T. Keer, M.J. Sutcliffe, D.A. Henriques, P. Mileham, S.D.M. Brown, 'Construction of a Mouse Yeast Artificial Chromosome Library in a Recombination-deficient Strain of Yeast', *Nat. Genet.*, **1**, 132–136 (1992).
35. G. Palmieri, M.G. Miano, A. Casamassimi, A. Lania, K. Kohno, D. Schlessinger, M. D'Urso, T. Featherstone, 'Construction of a Pilot Human YAC Library in

a Recombination-defective Yeast Strain', *Gene*, **188**, 169–174 (1997).
36. Y. Le, M.J. Dobson, 'Stabilization of Yeast Artificial Chromosome Clones in a *rad54-3* Recombination-deficient Host Strain', *Nucleic Acids Res.*, **25**, 1248–1253 (1997).
37. A.P. Monaco, Z. Larin, 'YACs, BACs, PACs, and MACs: Artificial Chromosomes as Research Tools', *Trends Biotechnol.*, **12**, 280–286 (1994).
38. H. Shizuya, B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, M. Simon, 'Cloning and Stable Maintenance of 300-kilobase-pair Fragments of Human DNA in *Escherichia coli* using an F-factor-based Vector', *Proc. Natl. Acad. Sci. USA*, **89**, 8794–8797 (1992).
39. U.J. Kim, B.W. Birren, T. Slepak, V. Mancino, C. Boysen, H.L. Kang, M.I. Simon, H. Shizuya, 'Construction and Characterization of a Human Bacterial Artificial Chromosome Library', *Genomics*, **34**, 213–218 (1996).
40. A.C.J. Frijters, Z. Zhang, M. van Damme, G.-L. Wang, P.C. Ronald, R.W. Michelmore, 'Construction of a Bacterial Artificial Chromosome Library Containing Large *Eco*RI and *Hin*dIII Genomic Fragments of Lettuce', *Theor. Appl. Genet.*, **94**, 390–399 (1997).
41. T. Mozo, S. Fischer, H. Shizuya, T. Altmann, 'Construction and Characterization of the IGF *Arabidopsis* BAC Library', *Mol. Gen. Genet.*, **258**, 562–570 (1998).
42. C.M. Hamilton, A. Frary, C. Lewis, S.D. Tanksley, 'Stable Transfer of Intact High Molecular Weight DNA into Plant Chromosomes', *Proc. Natl. Acad. Sci. USA*, **93**, 9975–9979 (1996).
43. Y.-G. Liu, Y. Shirano, H. Fukaki, Y. Yanai, M. Tassaka, S. Tabata, D. Shibata, 'Complementation of Plant Mutants with Large Genomic DNA Fragments by a Transformation-competent Artificial Chromosome Vector Accelerates Positional Cloning', *Proc. Natl. Acad. Sci. USA*, **96**, 6535–6540 (1999).
44. P.A. Ioannou, C.T. Amemiya, J. Garnes, P.M. Kroisel, H. Shizuya, C. Chen, M.A. Bather, P.J. de Jong, 'A New Bacteriophage P1-derived Vector for the Propagation of Large Human DNA Fragments', *Nat. Genet.*, **6**, 84–89 (1994).
45. S. Choi, R.A. Creelman, J.E. Mullet, R. Wing, 'Construction and Characterization of a Bacterial Artificial Chromosome Library of *Arabidopsis thaliana*', *Plant Mol. Biol. Rep.*, **13**, 124–128 (1995).
46. G.-L. Wang, T.E. Holsten, W.Y. Song, H.P. Wang, P.C. Ronald, 'Construction of a Rice Bacterial Artificial Chromosome Library and Identification of Clones Linked to the Xa-21 Disease-resistance Locus', *Plant J.*, **7**, 525–533 (1995).
47. S.-S. Woo, J. Jiang, B.S. Gill, A.H. Paterson, R.A. Wing, 'Construction and Characterization of a Bacterial Artificial Chromosome Library of *Sorghum Bicolor*', *Nuleic Acids Res.*, **22**, 4922–4931 (1995).
48. H.-B. Zhang, S. Choi, S.-S. Woo, Z. Li, R.A. Wing, 'Construction and Characterization of Two Rice Bacterial Artificial Chromosome Libraries from the Parents of a Permanent Recombinant Inbred Mapping Population', *Mol. Breed.*, **2**, 11–24 (1995).
49. S. Nakamura, S. Asakawa, N. Ohmido, K. Fukui, N. Shimizu, S. Kawasaki, 'Construction of an 800-kb Contig in the Near-centromeric Region of the Rice Blast Resistance Gene *Pi-ta2* using a Highly Representative Rice BAC Library', *Mol. Gen. Genet.*, **254**, 611–620 (1997).
50. L. Cai, J.F. Taylor, R.A. Wing, D.S. Gallagher, S.S. Woo, S.K. Davis, 'Construction and Characterization of a Bovine Bacterial Artificial Chromosome Library', *Genomics*, **29**, 413–425 (1995).
51. K. Oseogawa, P.Y. Woon, B. Zhao, E. Frengen, M. Tateno, J.J. Catanese, P.J. de Jong, 'An Improved Approach for Construction of Bacterial Artificial Chromosome Libraries', *Genomics*, **52**, 1–8 (1998).
52. P.Y. Woon, K. Oesogawa, P.J. Kaisaki, B. Zhao, J.J. Catanese, Gauguier, D.R. Cox, E.R. Levy, G.M. Lathrop, A.P. Monaco, P.J. de Jong', 'Construction and Characterization of a 10-fold Genome Equivalent Rat P1-derived Artificial Chromosome Library', *Genomics*, **50**, 306–316 (1998).
53. S. Godard, L. Schibler, A. Oustry, E.P. Cribiu, G. Guerin, 'Construction of a Horse Bac Library and Cytogenetical Assignment of 20 Type I and Type II Markers', *Mamm. Genome*, **9**, 633–637 (1998).
54. A. Marziali, N. Federspiel, R. Davis, 'Automation for the *Arabidopsis* Genome Sequencing Project', *Trends Plant Sci.*, **2**, 71–74 (1997).
55. M. Lovett, 'Fishing for Complements: Finding Genes by Direct Selection', *Trends Genet.*, **10**, 352–357 (1994).
56. A.J. Buckler, D.D. Chang, S.L. Graw, J.D. Brook, D.A. Haber, P.A. Sharp, D.E. Housman, 'Exon Amplification: a Strategy to Isolate Mammalian Genes Based on RNA Splicing', *Proc. Natl. Acad. Sci. USA*, **88**, 4005–4009 (1991).
57. M.I. Gomez, M. Nuru Islam Faridi, S.S. Woo, K.F. Schertz, D. Czeschin Jr, M.S. Zwick, R.A. Sing, D.M. Stelly, H.J. Price, 'FISH of a Maize sh2-selected Sorghum BAC to Chromosomes of *Sorghum Bicolor*', *Genome*, **40**, 475–478 (1997).
58. Z. Larin, A.P. Monaco, H. Lehrach, 'Yeast Artificial Chromosome Libraries Containing Large Inserts from Mouse and Human DNA', *Proc. Natl. Acad. Sci. USA*, **88**, 4123–4127 (1991).
59. S.M. Clark, E. Lai, B.W. Birren, L. Hood, 'A Novel Instrument for Separating Large DNA Molecules with Pulsed Homogeneous Electric Fields', *Science*, **241**, 1203–1205 (1988).
60. J.C. Maule, D.J. Porteous, A.J. Brookes, 'An Improved Method for Recovering Intact Pulsed Field Gel Purified DNA, of at Least 1.6 Megabases', *Nucleic Acids Res.*, **22**, 3245–3246 (1994).
61. Y.L. Sheng, V. Mancino, B. Birren, 'Transformation of *Escherichia coli* with Large DNA Molecules by

Electroporation', *Nucleic Acids Res.*, **23**, 1990–1996 (1995).
62. E. Maier, H.R. Crollius, H. Lehrach, 'Hybridization Techniques on Gridded High Density DNA and in situ Colony Filters based on Fluorescence Detection', *Nucleic Acids Res.*, **22**, 3423–3424 (1994).
63. W. Bender, P. Spierer, D.S. Hogness, 'Chromosomal Walking and Jumping to Isolate DNA from the *Ace* and *Rosy* Loci and the *Bithorax* Complex in *Drosophila melanogaster*', *J. Mol. Biol.*, **168**, 17–33 (1983).
64. H. Lehrach, R. Drmanac, J. Hoheisel, Z. Larin, G. Lennon, A.P. Monaco, D. Nizetic, G. Zehetner, A. Poustka, 'Hybridization Fingerprinting in Genome Mapping and Sequencing', in *Genome Analysis: Genetic and Physical Mapping*, eds. K.E. Davis, S.M. Tilghman, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 39–81, Vol. 1, 1990.
65. A. Craig, D. Nizetic, J.D. Hoheisel, G. Zehetner, H. Lehrach, 'Ordering of Cosmid Clones Covering the Herpes Simplex Virus Type I (HSV-I) Genome: a Test Case for Fingerprinting by Hybridization', *Nucl. Acids Res.*, **18**, 2653–2660 (1990).
66. E. Matallana, C.J. Bell, P.J. Dunn, M. Lu, J.R. Ecker, 'Genetic and Physical Linkage of the *Arabidopsis* Genome: Methods for Anchoring Yeast Artificial Chromosomes', in *Methods in Arabidopsis Research*, eds. C. Koncz, N.-H. Chua, J. Schell, World Scientific Publishing, Singapore, 144–169, 1992.
67. E. Maier, J.D. Hoheisel, L. McCarthy, R. Mott, A.V. Grigoriev, A.P. Monaco, Z. Larin, H. Lehrach, 'Complete Coverage of the *Schizosaccharomyces pombe* Genome in Yeast Artificial Chromosomes', *Nat. Genet.*, **1**, 273–277 (1992).
68. M.J. Palazzolo, S.A. Sawyer, C.H. Martin, D.A. Smoller, D.L. Hartl, 'Optimized Strategies for Sequence-tagged-site Selection in Genome Mapping', *Proc. Natl. Acad. Sci. USA*, **88**, 8034–8038 (1991).
69. T. Mozo, S. Fischer, S. Meier-Ewert, H. Lehrach, T. Altmann, 'Use of the IGF BAC Library for Physical Mapping of the *Arabidopsis thaliana* Genome', *Plant J.*, **16**, 377–384 (1998).
70. T. Mozo, K. Dewar, P. Dunn, J.R. Ecker, S. Fischer, S. Kloska, H. Lehrach, M. Marra, R. Martienssen, S. Meier-Ewert, T. Altmann, 'A Complete BAC-based Physical Map of the *Arabidopsis thaliana* Genome', *Nat. Genet.*, **22**, 271–275 (1999).
71. M. Olson, L. Hood, C. Cantor, D. Botstein, 'A Common Language for Physical Mapping of the Human Genome', *Science*, **245**, 1434–1435 (1989).
72. E.D. Green, M.V. Olson, 'Chromosomal Region of the Cystic Fibrosis Gene in Yeast Artificial Chromosomes: a Model for Human Genome Mapping', *Science*, **250**, 94–98 (1990).
73. T.J. Hudson, L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.-H. Xu, X. Hu, A.M.E. Colbert, C. Rosenberg, M.P. Reeve-Daly, S. Rozen, L. Hui, X. Wu, C. Vestergaard, K.M. Wilson, J.S. Bae, S. Maitra, S. Ganiatsas, C.A. Evans, M.M. DeAngelis, K.A. Ingalls, R.W. Nahf, L.T. Horton, Jr, M. Oskin Anderson, A.J. Collymore, W. Ye, V. Kouyoumijan, I.S. Zemsteva, J. Tam, R. Devine, D.F. Courtney, M. Turner Renaud, H. Nguyen, T.J. O'Connor, C. Fizames, S. Fauré, G. Gyapay, C. Dib, J. Morissette, J.B. Orlin, B.W. Birren, N. Goodman, J. Weissenbach, T.L. Hawkins, S. Foote, D.C. Page, E.S. Lander, 'An STS-based Map of the Human Genome', *Science*, **270**, 1945–1954 (1995).
74. C. Nusbaum, D.K. Slonim, K.L. Harris, B.W. Birren, R.G. Stehen, L.D. Stein, J. Miller, W.F. Dietrich, R. Nahf, V. Wang, O. Merport, A.B. Castle, Z. Husain, G. Farino, D. Gray, M.O. Anderson, R. Devine, L.T. Horton, Jr, W. Ye, X. Wu, V. Kouyoumijan, I.S. Zemsteva, Y. Wu, A.J. Collymore, D.F. Courtney, J. Tam, M. Cadman, A.R. Haynes, C. Heuston, T. Marsland, A. Southwell, P. Trickett, M.A. Strivens, M.T. Ross, W. Makalowski, Y. Xu, M.S. Boguski, N.P. Carter, P. Denny, S.D.M. Brown, T.J. Hudson, E.S. Lander, 'A YAC-based Physical Map of the Mouse Genome', *Nat. Genet.*, **22**, 388–393 (1999).
75. W. Kimmerly, K. Stultz, S. Lewis, V. Lustre, R. Romero, J. Benke, D. Sun, G. Shirley, C. Martin, M. Palazzolo, 'A P1-based Physical Map of the *Drosophila* Euchromatic Genome', *Genome Res.*, **6**, 414–430 (1996).
76. I. Chumakov, P. Rigault, S. Guillou, P. Ougen, A. Billaut, G. Guasconi, P. Gervy, I. LeGall, P. Soularue, L. Grinas, L. Bougueleret, C. Bellanné-Chantelot, B. Lacroix, E. Barillot, P. Gesnouin, S. Pook, G. Vaysseix, G. Frelat, A. Schmitz, J.-L. Sambucy, A. Bosch, X. Estivill, J. Weissenbach, A. Vignal, H. Riethman, D. Cox, D. Patterson, K. Gardiner, M. Hattori, Y. Sakaki, H. Ichikawa, M. Ohki, D. Le Paslier, R. Heilig, S. Antonarakis, D. Cohen, 'Continuum of Overlapping Clones Spanning the Entire Human Chromosome 21q', *Nature*, **359**, 380–387 (1992).
77. J. Quackenbush, C. Davies, J.M. Bailis, J.V. Khristich, K. Diggle, Y. Marchuck, J. Tobin, S.P. Clark, A. Rodkins, S. Marcano, A.C. Churukian, J.S. Hutchinson, S. Probst, L. Romberg, Y.H. Wie, N.J. Nowak, H.R. Garner, M.W. Smith, L. Selleri, G.A. Evans, 'An STS Content Map of Human Chromosome 11 – Localization of 910 YAC Clones and 109 Islands', *Genomics*, **29**, 512–525 (1995).
78. F. Hildebrandt, M. Cybulla, B. Strahm, H.G. Nothwang, I. Singhsawhney, K. Berz, M. Nicklin, O. Reiner, M. Brandis, 'Physical Mapping of the Gene for Juvenile Nephronophthisis (NPH1) by Construction of a Complete YAC Contig of 7 Mb on Chromosome 2q13', *Cytogenet. Cell Genet.*, **73**, 235–239 (1996).
79. G.G. Bouffard, J.R. Idol, V.V. Braden, L.M. Iyler, A.F. Cunningham, L.A. Weintraub, J.W. Touchman, R.M. Mohr-Tidwell, D.C. Peluso, R.S. Fulton, M.S. Ueltzen, J. Weissenbach, C.L. Magness, E.D. Green, 'A Physical

Map of Human Chromosome 7: an Integrated YAC Contig Map with Average STS Spacing of 79 kb', *Genome Res.*, **7**, 673–692 (1997).

80. B. Renault, A. Hovnanian, S. Bryce, J.J. Chang, S. Lau, A. Sakuntabhai, S. Mok, S. Carter, C.J.D. Ross, J. Pang, R. Twells, S. Chamberlain, A.P. Monaco, T. Strachan, R. Kucherlapati, 'A Sequence-ready Physical Map of a Region of 12q24.1', *Genomics*, **45**, 271–278 (1997).
81. R. Nagaraja, S. MacMillan, C. Jones, M. Masisi, G. Pengue, P. Giovanni, S. Miao, A. Casamassimi, M. D'Urso, B. Brownstein, D. Schlessinger, 'Integrated YAC/STS Physical and Genetic Map of 22.5 Mb of Human Xq24-q26 at 56-kb Inter-STS Resolution', *Genomics*, **52**, 247–266 (1998).
82. R.E. Ellsworth, V. Ionasescu, C. Searby, V.C. Sheffield, V.V. Braden, T.A. Kucaba, J.D. McPherson, M.A. Marra, E.D. Green, 'The CMT2D Locus: Refined Genetic Position and Construction of a Bacterial Clone-based Physical Map', *Genome Res.*, **9**, 568–574 (1999).
83. Y.C. Cao, H.L. Kang, X.Q. Xu, M. Wang, S.H. Dho, J.R. Huh, B.J. Lee, F. Kalush, D. Bocskai, Y. Ding, J.G. Tesmer, J. Lee, E. Moon, V. Jurecic, A. Baldini, H.U. Weier, N.A. Doggett, M.I. Simon, M.D. Adams, U.J. Kim, 'A 12 mb Complete Coverage BAC Contig Map in Human Chromosome 16p13.1-p11.2', *Genome Res.*, **9**, 763–774 (1999).
84. A.R. Coulson, J. Sulston, S. Brenner, J. Karn, 'Toward a Physical Map of the Genome of the Nematode *Caenorhabditis elegans*', *Proc. Natl. Acad. Sci. USA*, **83**, 7821–7825 (1986).
85. A.R. Coulson, Y. Kozono, B. Lutterbach, R. Shownkeen, J. Sulston, R. Waterston, 'YACs and the *C. elegans* Genome', *Bioessays*, **13**, 413–417 (1991).
86. M. Olson, J.E. Dutchik, M.Y. Graham, G.M. Brodeur, C. Helms, M. Frank, M. Mac Collin, R. Scheinman, T. Frank, 'Random Clone Strategy for Genomic Restriction Mapping in Yeast', *Proc. Natl. Acad. Sci. USA*, **83**, 7826–7830 (1986).
87. L. Riles, J.E. Dutchik, A. Baktha, B.K. McCauley, E.C. Thayer, M.P. Leckie, V.V. Braden, J.E. Depke, M.V. Olson, 'Physical Maps of the Six Smallest Chromosomes of *Saccharomyces cerevisiae* at a Resolution of 2.6 Kilobase Pairs', *Genetics*, **134**, 81–150 (1993).
88. Y. Kohara, K. Akiyama, K. Isono, 'The Physical Map of the Whole *E. coli* Chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library', *Cell*, **50**, 495–508 (1987).
89. I. Siden-Kiamos, R.D.C. Saunders, I. Spanos, T. Majerus, J. Trenear, C. Savakis, C. louis, D.M. Glover, M. Ashburner, F.C. Kafatos, 'Towards a Physical Map of the *Drosophila melanogaster* Genome: Mapping of Cosmid Clones within Defined Genomic Divisions', *Nucleic Acids Res.*, **18**, 6261–6270 (1990).
90. R.L. Stallings, D.C. Torney, C.E. Hildebrand, J.L. Longmire, L.L. Deaven, J.H. Jett, N.A. Dogget, R.K. Moysis, 'Physical Mapping of Human Chromosomes by Repetitive Sequence Fingerprinting', *Proc. Natl. Acad. Sci. USA*, **87**, 6218–6222 (1990).
91. K. Taylor, N. Hornigold, D. Convay, D. Williams, Z. Olinowski, M. Agochiya, P. Fattory, P. de Jong, P.F.R. Little, J. Wolfe, 'Mapping the Human Y Chromosome by Fingerprinting Cosmid Clones', *Genome Res.*, **6**, 235–248 (1996).
92. G.K. Wong, J. Yu, E.C. Thayer, M.V. Olson, 'Multiple-complete-digest Maps for Large-scale DNA Sequencing', *Proc. Natl. Acad, Sci. USA*, **94**, 5225–5230 (1997).
93. M.A. Marra, T.A. Kucaba, N.L. Dietrich, E.D. Green, B. Brownstein, R.K. Wilson, K.M. McDonald, L.W. Hillier, J. McPherson, R.H. Waterston, 'High Throughput Fingerprint Analysis of Large-insert Clones', *Genome Res.*, **7**, 1072–1084 (1997).
94. J.C. Venter, H.O. Smith, L. Hood, 'A New Strategy for Genome Sequencing', *Nature*, **381**, 364–366 (1996).
95. D.W. Meinke, J.M. Cherry, C. Dean, S.D. Rounsley, M. Koornneef, '*Arabidopsis thaliana*: a Model Plant for Genome Analysis', *Science*, **282**, 662–682 (1998).
96. M. Bevan, J. Ecker, S. Theologis, N. Federspiel, R. Davis, D. McCombie, R. Martienssen, E. Chen, B. Waterston, R. Wilson, S. Rounsley, C. Venter, S. Tabata, M. Salanoubat, F. Quetier, J.M. Cherry, D. Meinke, 'Objective: the Complete Sequence of a Plant Genome', *Plant Cell*, **9**, 476–478 (1997).
97. M.A. Marra, T. Kucaba, M. Sekhon, L. Hillier, R. Martienssen, A. Chinwalla, J. Crockett, J. Fedele, H. Grover, C. Gund, W.R. McCombie, K. McDonald, J. McPherson, N. Mudd, L. Parnell, J. Schein, R. Seim, P. Shelby, R. Waterston, R. Wilson, 'A Map for Sequence Analysis of the *Arabidopsis thaliana* Genome,' *Nat. Genet.*, **22**, 265–270 (1999).
98. E.M. Meyerowitz, 'Structure and Organization of the *Arabidopsis thaliana* Nuclear Genome', in *Arabidopsis*, eds. E.M. Meyerowitz, C.R. Somerville, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 21–36, 1994.
99. E. Bent, S. Johnson, I. Bancroft, 'BAC Representation of Two Low-copy Regions of the Genome of *Arabidopsis thaliana*', *Plant J.*, **13**, 849–855 (1998).
100. C. Soderlund, I. Longden, R. Mott, 'FPC: a System for Building Contigs from Restriction Fingerprinted Clones', *Comput. Appl. Biosci.*, **13**, 523–535 (1997).
101. R. Schmidt, J. West, K. Love, Z. Lenehan, C. Lister, H. Thompson, D. Bouchez, C. Dean, 'Physical Map and Organization of *Arabidopsis thaliana* Chromosome 4', *Science*, **270**, 480–483 (1995).
102. R. Schmidt, K. Love, J. West, Z. Lenehan, C. Dean, 'Description of 31 YAC Contigs Spanning the Majority of *Arabidopsis thaliana* Chromosome 5', *Plant J.*, **11**, 563–572 (1997).
103. H. Kotani, S. Sato, T. Hosouchi, N. Nakazaki, S. Okumura, T. Wada, Y.-G. Liu, D. Shibata, S. Tabata, 'A Fine Physical Map of *Arabidopsis thaliana* Chromosome 5:

Construction of a Sequence-ready Contig Map', *DNA Res.*, **4**, 371–378 (1997).
104. E.A. Zachgo, M.L. Wang, J. Dewney, D. Bouchez, C. Camilleri, S. Belmonte, L. Huang, M. Dolan, H.M. Goodman, 'A Physical Map of Chromosome 2 of *Arabidopsis thaliana*', *Genome Res.*, **6**, 19–25 (1996).
105. C. Camilleri, J. Lafleuriel, C. Macadré, F. Varoquaux, Y. Parmentier, G. Picard, M. Caboche, D. Bouchez, 'A YAC Contig Map of *Arabidopsis thaliana* Chromosome 3', *Plant J.*, **15**, 633–642 (1998).
106. S. Sato, H. Kotani, R. Hayashi, Y.G. Liu, D. Shibata, S. Tabata, 'A Physical Map of *Arabidopsis thaliana* Chromosome 3 Represented by Two Contigs of CIC YAC, P1, TAC and BAC Clones', *DNA Res.*, **5**, 163–168 (1998).

Mass Spectrometry of Nucleic Acids

Julia Gross and Franz Hillenkamp
University of Münster, Germany

1 Introduction **1**
1.1 Characteristics of Nucleic Acids Under Mass Spectrometric Conditions 2
1.2 General Characteristics of Mass Spectra 7
2 Applications **7**
2.1 Mass Range 7
2.2 Mixture Analysis and Combination of Mass Spectrometry with Separation Techniques 9
2.3 Noncovalent Complexes 10
2.4 Sequencing Approaches 12
2.5 Clinical Diagnostics with Mass Spectrometry 19
2.6 Conclusions 21
Acknowledgments **21**
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

Nucleic acids are prone to fragmentation in the ionization process of mass spectrometry (MS). The cause can be found in the polar nature of this analyte class. Following the introduction of the so-called soft-ionization techniques of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) at the end of the 1980s, mass spectrometric analysis of nucleic acids has gained broader applicability. In combination with further developments of the associated instrumentation and optimized preparation and purification methods, a dramatic enhancement of the mass range, detection sensitivity, and resolution became possible – in routine analysis oligonucleotides up to a length of 50 nucleotides are accessible. At present the necessary quantity of sample is in the subfemtomole range.

This article describes the fundamentals of mass spectrometric analyses of nucleic acids and common applications, such as the analysis of noncovalent complexes, mixture analysis, different sequencing strategies and clinical diagnostics. The focus is on MALDI/MS (matrix-assisted laser desorption ionization/mass spectrometry).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Characterization of nucleic acids by MS has always been both a challenging and promising endeavor. Since 1962, when the first mass spectra of nucleosides obtained with electron impact mass spectrometry (EI/MS) were reported[1] MS has gained importance in nucleic acid analysis.[2–4]

Although the classical ionization techniques (electron impact (EI) and chemical ionization (CI)) were limited in analyte size to dinucleotides, they were successfully employed for the structural analysis of nucleic acid constituents (i.e. nucleobases, nucleosides, nucleotides) and had their significance especially in the elucidation of posttranscriptional processings of RNA and DNA.[5] The limitation to dinucleotides is a result of the polar nature of the analyte and the elevated tendency to fragmentation (thermal decomposition predominates volatilization).

Chemical derivatization of the analyte was often necessary to enhance the volatility and to reduce fragmentation for MS examination.[6] Hunt et al.[7] intensively studied per(trimethylsilyl) derivatized dinucleoside monophosphates with EI/MS. In addition to the molecular ions they observed sequence-specific fragment ions. As dinucleotides are the smallest nucleic acid subunit that contains sequence information, these studies paved the way for direct MS sequencing. The investigation of all four underivatized di-2′-deoxyribonucleoside monophosphates showed no molecular ions, but fragment ions in terminus-related intensities.[8] However, diribonucleoside monophosphates were found to be much more stable, a feature attributed to the 2′-hydroxyl group.[9] Unprotected dinucleotide monophosphates could be analyzed with field desorption (FD) MS,[10] but it was found that this technique was experimentally difficult to handle with low run-to-run reproducibility.[11]

The analysis of oligonucleotides became feasible with the introduction of plasma desorption (PD), secondary ion (SI), and fast atom bombardment (FAB) MS.

Positive ion spectra of plasma-desorbed protected oligonucleotides showed mostly cationized quasimolecular ions[12,13] with a size limit of nine nucleotides.[14] In the negative ion mode distinctive fragment ions could be observed in addition to the deprotonated quasimolecular ions. The fragmentation pattern showed two complementary series retaining either of the two termini of the nucleotide strand. Beginning with the more intense series of fragment ions containing the 5′ terminus, these fragments bear the phosphate group at their 3′ end after cleavage of the respective 5′ C–O bond and, correspondingly, the 3′ terminus series carries the phosphate group at the 5′ end after cleavage of the 3′ C–O bond.[12,15] SIMS (secondary ion mass spectrometry) revealed the

attachment of Na^+ and Cs^+ to the molecule in the positive ion mode and intact quasimolecular ions in the negative ion mode.[16] Wu et al.[17] reported an enhanced ionization efficiency with SIMS if the analyte was mixed with a typical MALDI matrix (e.g. dihydroxybenzoic acid). This technique made it possible to detect a DNA 26-mer with only minor fragmentation. The pioneering work on MS oligonucleotide sequencing was performed by Grotjahn et al. who compared the mass spectra of more than 100 oligomers obtained by negative ion mode FAB/MS (fast atom bombardment/mass spectrometry). These workers deduced rules for deriving sequence information from fragment ion patterns.[18,19] Two major fragment ion series were detected: one more intense with the so-called 5′ sequence ions bearing a 5′ phosphate group, and a second one with the phosphate group at the 3′ end (3′ sequence ions). It is noteworthy that this description may be misleading: the 5′ sequence ions (3′ sequence ions) only carry the phosphate group of the cleaved phosphodiester bond at their 5′ end (3′ end), but contain the 3′ terminus (5′ terminus) of the oligonucleotide. Although additional fragmentation (−18 Da, explained as loss of water and +80 Da, described as $+HPO_3$ artifacts) was observed, bidirectional sequencing based on these two distinct fragment ion series was proposed. This concept for sequencing unknown oligonucleotides relied on the assignment of the more intense series to the so-called 5′ sequence ions. It was reported that this was indeed the more intense series for 8-mers but not for 2–4-mers. Other researchers also questioned the predominance of the 5′ series.[20,21] The complete bidirectional sequence information could be obtained for up to a 10-mer using about 10 nmol of sample.[19,22] Although efforts were made to increase the accessible mass range of oligonucleotides for FAB/MS the biggest detected intact oligonucleotide was an RNA 24-mer.[23] A detailed comparison of these ionization methods for short oligonucleotides with special consideration of the fragmentation is given by Crain.[2]

The introduction of ESI by Fenn et al.[24] and MALDI by Karas and Hillenkamp[25] in the late 1980s opened a new chapter in MS nucleic acid research. Further development of the associated instrumentation in combination with optimized preparation and purification methods led to dramatic increases in the accessible mass range, resolution, and sensitivity.[26–30]

1.1 Characteristics of Nucleic Acids Under Mass Spectrometric Conditions

ESI and MALDI exhibit similarities but also differences due to the respective ionization technique. In ESI, an analyte solution is sprayed through a needle into an electrical field and ionized via repetitive coulombic bursting of the generated droplets until the fully desolvated analyte ions finally enter the mass analyzer. For a MALDI analysis, the analyte solution is mixed with a matrix (organic substance in a high molar excess) and deposited on a metallic sample support. For solid matrices used predominantly in MALDI with ultraviolet (UV) laser wavelengths (UV/MALDI), the volatile solvent is evaporated leaving a polycrystalline sample behind. Vacuum-stable liquid matrices such as glycerol are used predominantly in MALDI with infrared (IR) wavelengths (IR/MALDI) and the matrix/analyte preparation is transferred directly into the spectrometer. Both analyte and matrix molecules are desorbed and ionized in an interaction of the sample with a high intensity laser light pulse of nanosecond duration.

In solution, the polyanionic backbone of nucleic acids is fully dissociated as a result of the pK_a value of the backbone phosphates being less than 1 (for a comparison, phosphoric acid has a pK_a of 2.1).[31] The pK_a values of all protonated nucleobases but the uracil (U) derivatives ($pK_a \approx 10$) vary from 3.5 to 5.[31] For ribonucleosides the pK_a values are 4.2 for cytidine, 3.5 for adenosine, and 2.4 for guanosine. These values are only slightly less than those for the deoxyribonucleosides. Due to the electrostatic interaction of the negatively charged phosphodiester groups with the protonated bases, the respective values are decreased in nucleotides by 0.2–0.5 pK units.[31] These variations may occur because of the more complex interactions possible within bigger molecules. It should be kept in mind that ionic strength[32] and temperature[33,34] both influence the ionization constants. Kochetov and Budovskij[35] evaluated the possible protonation sites of the nucleobases and stated that they are N1, N3 and N7 for adenine (A), N3 and N7 for guanine (G) and N3 for cytosine (C). The same authors reported that X-ray, spectroscopic, and NMR studies revealed N1 in A and N7 in G as favored protonation sites (Figure 1).[35]

As the molecules are transferred into the gas phase for mass spectrometric analysis, the gas-phase acidities and basicities must be considered for the interpretation of the mass spectra, especially to account for process-induced fragmentation. Greco et al. determined the following order of proton affinities for deoxyribonucleoside 5′ monophosphates (dNMPs): dGMP (234.4 kcal mol^{-1}) > dAMP (233.6 kcal mol^{-1}) ≈ dCMP (233.2 kcal mol^{-1}) ≫ dTMP (224.9 kcal mol^{-1}).[36] Similar values can be expected for the related ribocomponents. The proton affinity of the phosphodiester group has not been determined experimentally, but was deduced from the known value for triethylphosphate as 220–225 kcal mol^{-1}.[37] The following can be concluded for singly protonated or deprotonated ions (usually observed in MALDI/MS). In the positive ion mode (protonated quasimolecular ion), the most stable structure is the one in which all phosphodiester groups

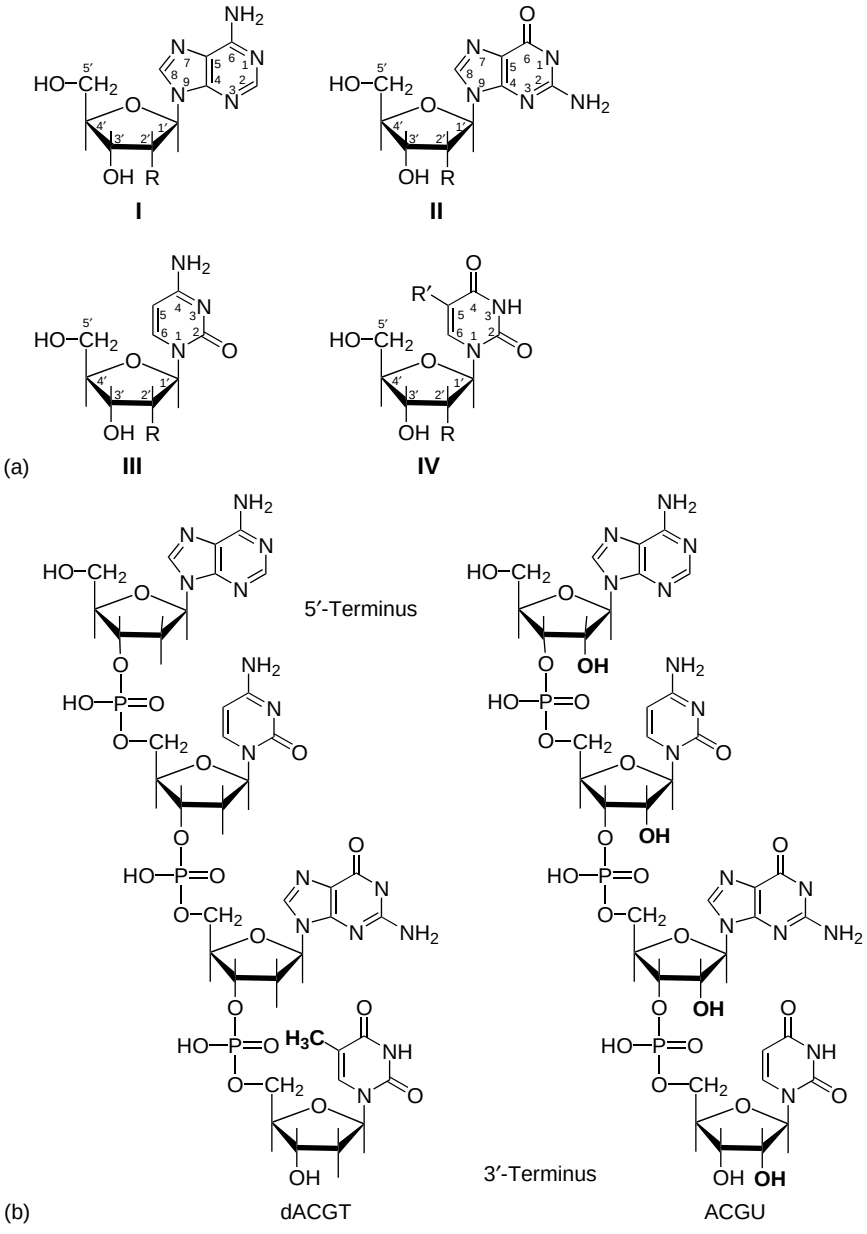

Figure 1 (a) Nucleoside structures with the numbering of the ring atoms in the bases and of the carbon atoms in the sugar moieties: I adenosine; II guanosine; III C; IV with R′ = CH_3 thymidine, and with R′ = H uridine. The sugar moiety is deoxyribose with R = H and ribose with R = CH_3. R = F indicates a 2′ fluorine modification. An exchange of N in position 7 for C in structures I and II leads to the respective 7-deaza purines. (b) Structures of the 4-mer are (left) DNA ACGT and (right) RNA ACGU. The differences between DNA and RNA are printed in bold letters.

are neutral and the proton is located on either G, A, or C. Only for oligothymidylic acids are the proton affinities of the phosphodiester group and the base too similar in value to predict a localization of the proton. It is noteworthy that interactions between phosphate groups and nucleobases are probable, resulting in the transient formation of zwitterions in dependence of the respective proton affinities for both the positive and negative ion mode. Rodgers et al. have calculated the gas-phase acidities of nucleobases in dideoxyribonucleotides and for the phosphodiester group to obtain G (333.4 kcal mol^{-1}) > thymine(T) (344.1 kcal mol^{-1}) ≈ C (345.6 kcal mol^{-1}) > A (349.0 kcal mol^{-1}) and 329 kcal mol^{-1} for the phosphodiester group.[38] Hence, in the negative ion mode the charge is expected to be located on one of the phosphodiester groups of the oligonucleotide.

Another characteristic of nucleic acids is the strong affinity towards metal ions. Binding sites are the oxygen

and nitrogen atoms in the nucleobases and, to a lesser extent, the phosphate and sugar moieties.[39] The importance of metal binding of nucleic acids in vivo lies in the specific interaction with proteins and in the stabilization of the secondary structure. However, this property affects the quality of the mass spectra. Heterogeneous cation attachment, such as the ubiquitous Na^+ and K^+, decreases sensitivity, resolution, and mass accuracy. Instead of just the protonated/deprotonated quasimolecular ion, the total ion current (TIC) is distributed over a multitude of ion species. As the mass differences between these different ions are relatively small, the separation of neighboring signals becomes increasingly difficult with increasing analyte mass until only the envelope can be resolved. Owing to the higher resolution of quadrupole analyzers (in the case of the detection of multiply charged larger analytes) when coupled to ESI sources, this effect is less pronounced there than it is for the singly charged MALDI ions in the time-of-flight (TOF) analyzer of MALDI instruments. However, impurities are often displaced from the crystalline rim of the sample to the microcrystalline center in MALDI sample preparations (purification by crystallization). ESI lacks this inherent purification mechanism. Moreover, ESI generates predominantly multiply charged ions. Therefore, cations present in the preparation give rise to ion series with the same charge state but a diversity of attached cations. This further complicates the mass spectrum. Therefore, careful sample handling and/or desalting procedures are vital for successful MS analyses.

Improvement in resolution and sensitivity of negative ion ESI mass spectra was first reported by Stults and Marsters.[40] Cation exchange with ammonium acetate via precipitation of the oligonucleotide led to a clear decrease in sodium salt formation for the tested DNA 30-, 48-, and 77-mers (Figure 2).

Desalting is also very beneficial for MALDI/MS. Exchanging metal cations with organic ammonium salts

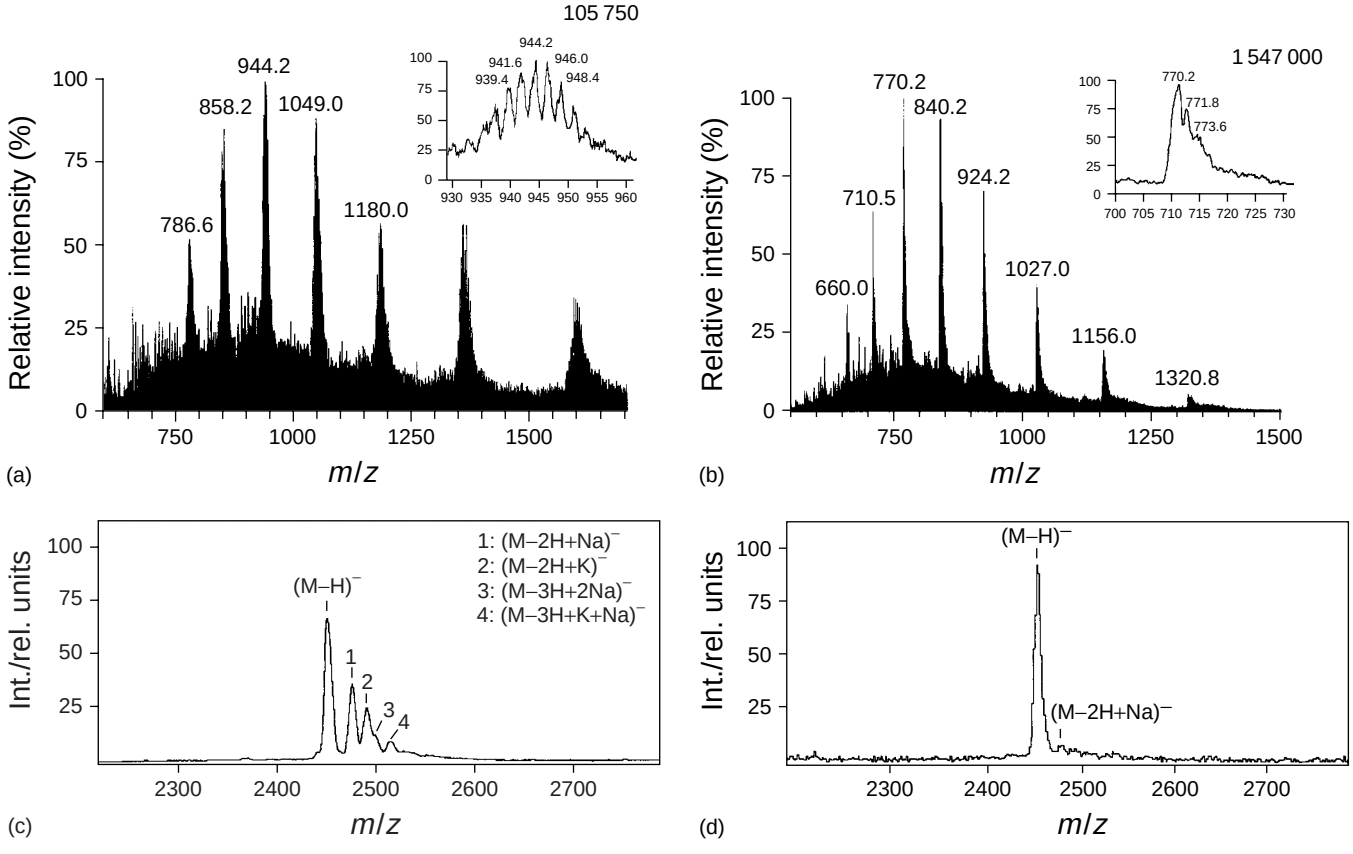

Figure 2 Electrospray mass spectra of the 30-mer of sequence $H_2N-(CH_2)_6-PO_2H-$ATC GTT CCT TTG GCA TTC CTG ACT CAT CTA (a) before and (b) after ammonium acetate precipitation. The number in the top right corner of each spectrum gives the absolute intensity of the most abundant ion in counts per second. (Adapted from J.T. Stults and J.C. Marsters, *Rapid Commun. Mass Spectrom.*, **5**, 359–363 (1991) with permission of John Wiley & Sons Limited, Copyright 1991.) (c) Negative-ion MALDI mass spectra of a poly-T 8-mer with succinic acid using an erbium/yttrium–aluminum–garnet (Er/YAG) laser. (d) The same analyte after addition of 5–10 cation exchange beads. (Adapted from E. Nordhoff et al., *Rapid Commun. Mass Spectrom.*, **6**, 771–776 (1992) with permission from John Wiley & Sons Limited, Copyright 1992.)

turned out to be an effective solution for both ionization techniques. In the gas phase, these ammonium salts release ammonia, and the oligonucleotides are detected as free acids. This principle has been applied to both ESI and MALDI and to various classes of nucleic acid analytes.[27–29,41,42]

The purification and preparation depends on the chemical environment and the heterogeneity of the sample. Salts present in only small amounts can be suppressed by additives to the analyte solution (ESI/MS) and also to the MALDI matrix. The addition of ammonium tartrate or citrate to matrices such as 6-aza-2-thiothymine[43] (ATT), and to a mixture of 2,3,4- and 2,4,6-trihydroxyacetophenone[44] (THAP) has been shown to increase the analyte ion intensity. It is possible that this effect is due to enhanced ionization of the analyte,[44] more efficient incorporation into the matrix crystals,[45] or a reduction of cation attachment.[46] An elegant way to remove monovalent metal cations is the addition of cation-exchange polymer beads (BioRad™, Hercules, CA, USA) loaded with ammonium.[47,48] Both analyte and matrix solutions can be desalted with these beads. An on-target addition of a small aliquot of cation-exchange beads to the MALDI preparation further eliminates cations introduced from the sample support or from pipette tips.[47]

It was reported that the addition of *trans*-1,2-diaminocyclohexane-*N*,*N*,*N*′,*N*′-tetraacetic acid leads to a successful elimination of divalent cations from tRNAs for ESI/MS[49] but has only a minor effect for MALDI/MS.[50,51] Reversed-phase high-performance liquid chromatography (RPHPLC) (section 2.2.) has been used to desalt analyte solutions prior to ESI/MS analysis.[52] This approach offers advantages for shorter oligonucleotides with their lower precipitation efficiencies. Kussmann et al.[53] introduced a miniaturized version for a fast sample clean-up. The analyte is concentrated in a column (GelLoader™ tip, Eppendorf), packed with only a few microliters of Poros™ material (PerSeptive Biosystems, Framingham, MA, USA), washed and then eluted directly onto the MALDI target. For DNA and RNA a triethylammonium acetate buffer, pH 7, with acetonitrile–water was employed.[54] Similar microcolumns in a pipette tip are now commercially available (Millipore, Bedford, MA, USA).

Whereas fragmentation is less pronounced in ESI/MS, it clearly limits the accessible mass range for UV/MALDI. Although fragmentation occurs at all times during the MALDI/TOF measurement (section 2.4.) it is the metastable decay in the field-free drift tube that impairs routine analyses by peak broadening, loss of signal intensity, and therefore reduced resolution (Figure 3). This is particularly marked in instruments with an ion

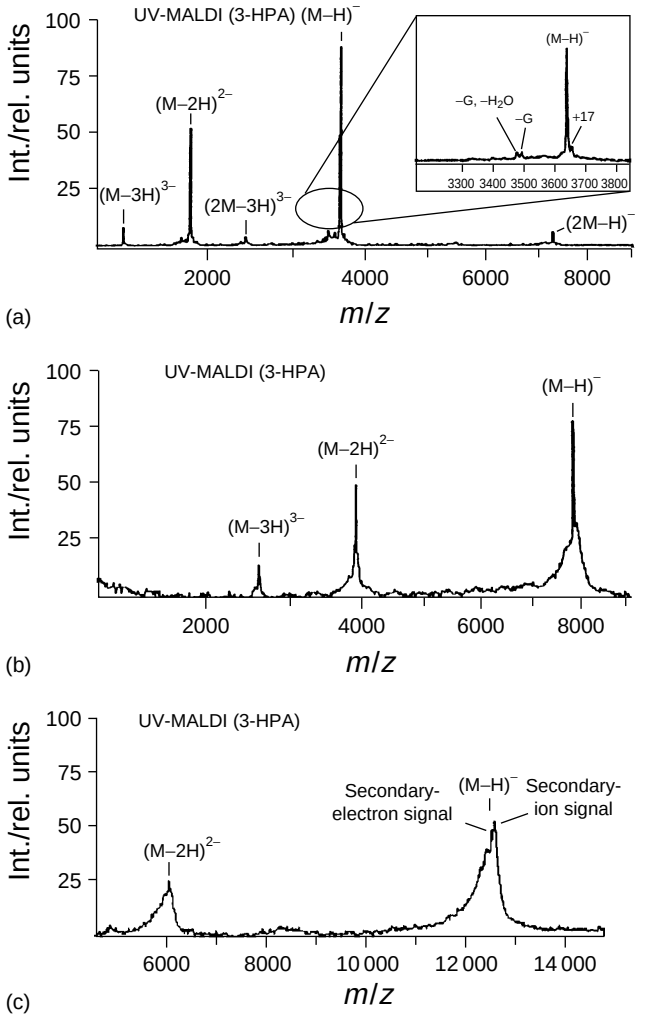

Figure 3 Comparison of MALDI negative-ion reflectron-TOF-mass spectra of oligodeoxynucleotides with increasing sequence length with 3-hydroxypicolinic acid (3-HPA) and a laser wavelength of 355 nm: (a) 12-mer 5′-dCGC GAT ATC GCG-3′; (b) 26-mer 5′-dCTA CTC TCA CTA CGC GCA AGC TAC CG-3′; (c) 40-mer 5′-dTAC AGA TTT TAT GTT TAT TTA TGT TTG CTT TTC AAA AGGC-3′. In spectrum (c) in addition to the secondary electron signal of the secondary electron multiplier, a second detector signal (SI signal) is observed and assigned, due to the detection process of higher masses. (Adapted from E. Nordhoff et al., *Nucleic Acids Res.*, **22**, 2460–2465 (1994) with permission from Oxford University Press, Copyright 1994.)

reflector which must be used for large oligonucleotides in IR/MALDI.

It was found that fragmentation increases with the mass of the analyte and is strongly dependent on base composition and applied matrix–laser wavelength combination.[48,55,56] Three groups determined the order of nucleobase stability T(U) ≫ C > A, G under MALDI conditions,[48,55,57] and one group found the order T(U) ≫ A > C, G with A and C reversed.[56] The following

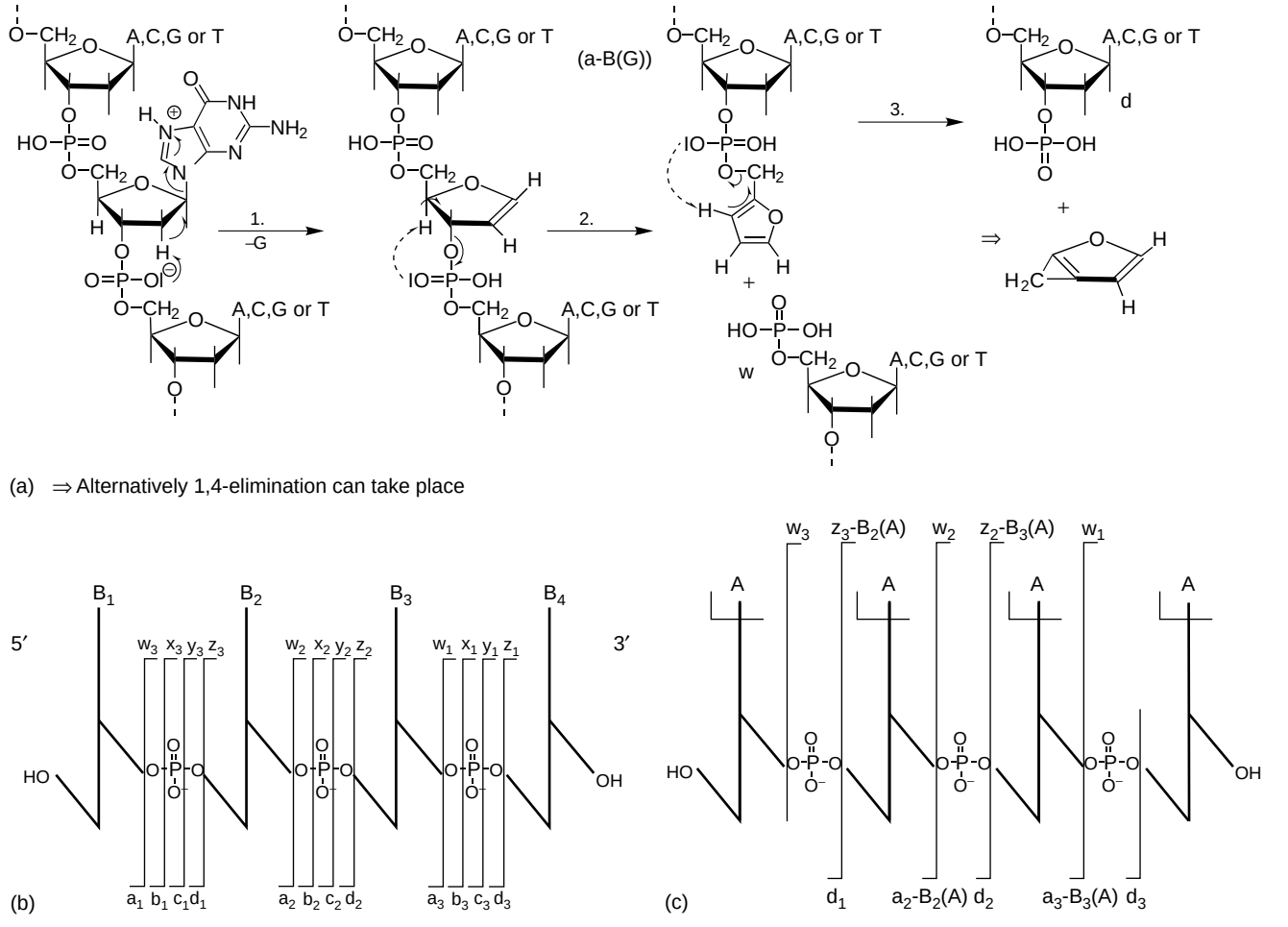

Figure 4 (a) Suggested reaction schemes for the fragmentation of oligonucleotides with the fragment ions assigned using McLuckey nomenclature. (Adapted from E. Nordhoff et al., *J. Mass Spectrom.*, **30**, 99–112 (1995) with permission of John Wiley & Sons Limited, Copyright 1995.) (b) Putative backbone cleavages and (c) the loss of base. (Reprinted from S.A. McLuckey et al., *J. Am. Soc. Mass Spectrom.*, **3**, 60–70 (1992) with permission from Elsevier Science, Copyright 1992 by American Society for Mass Spectrometry.)

oligonucleotide stability order was observed: oligouridylic (oligothymidylic) acid > RNA > DNA.[48] The higher stability of RNA relative to DNA can be explained by the balancing effect of the 2′-hydroxyl group on polarization of the *N*-glycosidic bond, induced by protonation of N7 of the purine bases and N3 of C. It was speculated that the 2′-hydroxyl group may also sterically hinder the base loss. This explanation can be ruled out because nucleic acids containing C-arabinose (with the hydroxyl group in the *cis* position) were found to exhibit a similar stability.[58,59] Base protonation was soon found to be the initial step of ion fragmentation in the positive ion mode, leading to cleavage of the *N*-glycosidic bond, often followed by backbone cleavage (Figure 4).[48,55,60–62] This figure also gives the fragment assignment by McLuckey et al.[63]

For the negative ions the details of the fragmentation mechanisms are still under discussion. A positional dependence of the protonation of base (charge remote fragmentation) has been suggested.[64]

The suppression of fragmentation can be accomplished by modifying either the nucleobase, the sugar, or the phosphodiester moieties. 7-Deaza-purine nucleotides, long employed in gel-electrophoresis, afforded a substantial improvement in ion stability.[61,65] However, due to the participation of N3 of C in base pairing, the corresponding deaza-C-analog is incompatible with enzymatic reactions. It was shown that, at least for mixed-sequence 4-mers, the loss of C is comparable in magnitude to the loss of A and G. Therefore, the stabilization of this base is required as well.[62] Exchange of the sugar moiety for arabinose and 2′-fluorine modification of ribose have also been reported.[58,59] Unfortunately, the combination of these modifications with enzymatic reactions turned out to be problematic.[66] Alkylation of the phosphate backbone

to reduce an interaction of the acidic protons with the bases was demonstrated to be effective for shorter oligonucleotides.[67] The efficiency of this approach to longer oligomers remains to be demonstrated. The choice of the matrix also has a strong influence on the fragmentation of nucleic acids under MALDI conditions. Since its introduction by Wu et al.,[68] 3-HPA is the matrix of choice for MALDI mass spectrometric measurements of DNA and RNA because of the limited prompt and metastable fragmentation observed with this matrix. Some prompt fragmentation was shown to take place for this matrix, but fragment ions are not detected because further serial cleavages occur leading finally to low-mass fragments.[59] A few other matrices, such as ATT and several di- and THAPs, have found to be useful for special cases of nucleic acid analysis. It is fair to state, however, that metastable decay remains the main problem for UV/MALDI/MS of nucleic acids.

1.2 General Characteristics of Mass Spectra

DNA and RNA are detected as protonated or deprotonated quasimolecular ions in MALDI/MS. Generally, singly charged ions $[M \pm H]^{+/-}$ constitute the base peak in MALDI mass spectra, although multiply charged ions $[M \pm nH]^{n+/-}$ and unspecific oligomers (singly and multiply charged, $[mM \pm nH]^{n+/-}$) are observed. The occurrence of multiply charged ions is more pronounced in the negative ion mode.[57] For most matrices the negative ion mode yields higher signal intensities, with the exception of 3-HPA where the signal intensities are comparable for both ion modes.[57,69] Matrix signals often dominate the low-mass range of the spectrum, increasing in mass and intensity with the applied laser fluence. A characteristic of ESI is the generation of multiply charged ions. As a rule of thumb, about one charge is present per 1000 Da of ion mass and, therefore, the mass spectrum of every single compound consists of several signals reflecting different charge states of the analyte. However, the charge distribution is strongly dependent on the molecular mass of the analyte, the solvents used, and the instrumental parameters. ESI mass spectra are commonly deconvoluted to determine the peak shape of the equivalent singly charged ion and the molecular weight. In combination with high-performance liquid chromatography (HPLC) the display of the TIC versus the retention time yields a chromatogram of the sample. Each signal of this total ion chromatogram usually consists of several mass spectra depending on the scanning time of the instrument (see Figure 7). In ESI/MS the negative ion mode generally provides higher signal intensities than the positive mode.

2 APPLICATIONS

The accessible mass range, sensitivity, and resolution determine the usefulness of MS in competition with other analytical techniques. The continuing improvements of these parameters have made MS of nucleic acids a widely used tool for a broad range of applications. In addition to accurate mass determination sequencing of polynucleotides, clinical and forensic diagnostics and the analysis of biologically relevant noncovalent complexes have become feasible. Sensitivity has been increased to the subfemtomole level using miniaturized sample preparation techniques.[70,71] Resolution has made a great leap forward with the implementation of delayed ion extraction[72] (DE) in MALDI/TOF/MS. For DNA up to about 50 nucleotides in length, enhanced resolution has been demonstrated.[73–76] In addition to better instrumental resolving power, DE leads to a reduced fragmentation in the initial field-free expansion of the generated plume. This further contributes to an increased mass accuracy[75] and sensitivity.[76]

2.1 Mass Range

In MALDI/MS, fragmentation of the analyte and cation attachment are the main factors limiting both mass range and resolution. Reduced efficiency of ionization[69] and detection may also hamper the analysis of large molecules. Nevertheless, the analysis of DNA up to 50 nucleotides in length by UV/MALDI/MS with a resolution of approximately 1000 FWHM (full-width at half-maximum) for this mass range has become routine. For DNA exceeding 80 nucleotides the resolution quickly decreases to 60–20 FWHM. The detection of a 461-mer in vitro transcript was reported by Kirpekar et al. after removal of the 5′ phosphate groups with calf intestine phosphatase to increase homogeneity and sensitivity.[77] The investigation of double-stranded DNA (dsDNA) exceeding 200 base pairs (bps), generated by restriction enzymes from plasmid DNA, was reported by Bai et al. The sample was prepared on a Nafion™ membrane with 3-HPA as the matrix.[78,79] Liu and colleagues coated the sample support with nitrocellulose (NC) and observed an improvement in the analysis of complex dsDNA mixtures with up to 622 bp using a 4 : 1 molar ratio of 3-HPA and PA (picolinic acid).[80,81] Furthermore, the detection of up to 500-mer ssDNA (single-stranded DNA) PCR (polymerase chain reaction) products was reported by Tang et al. with mixtures of 3-HPA and PA on a standard sample support.[82] Ross and Belgrader were able to detect a ssDNA 268-mer PCR product with a mixture of 3-HPA and PA in a solution of ammonium citrate and acetonitrile. The key to this successful analysis was the purification and isolation of a

single-stranded PCR product by affinity capture using streptavidin-coated magnetic beads.[83] In most cases only the dehybridized single strands are detected under the given MALDI conditions. Most of the individual strands of dsDNA exhibit only a small mass difference. For higher-mass analytes, this mass difference cannot be resolved further, leading to additional peak broadening. To avoid the severe fragmentation observed for larger oligonucleotides in UV/MALDI, Berkenkamp et al. used IR/MALDI with glycerol as the matrix and an Er/YAG laser with a wavelength of 2.94 μm for the analysis of restriction enzyme digests of plasmid DNA of more than 2000 nucleotides.[71] They observed a dramatically lower fragmentation under these conditions. However, so far mass resolution is limited to only 50–100 throughout the mass range representing about 100 to 1500 nucleotides (Figure 5).

Using ESI, the detection of intact highly charged DNA ions in the megadalton range has been demonstrated. Ions exceeding 1 MDa were detected by Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS)[84] and a TOF analyzer with a charge detector.[85] Extremely high values for the mass resolution have been obtained by FTICR/MS, but the heterogeneity of the charge states of the trapped ion population complicates the interpretation of the spectra.[86] Chen et al. succeeded in the detection of coliphage T4 DNA ions by FTICR/MS.[87] In this case the charge state of an individual highly charged anion could be determined by analyzing the decrease of the charge state over a given time interval. For the T4 DNA a charge state of 31 530 and a mass of 90.9×10^6 Da could be determined in this manner with a precision of 10%. In a subsequent work Cheng et al. reported the detection of pGEM-5S plasmid DNA. Determining a mass of 1.5 MDa (calculated 1.946 MDa) these workers reduced the uncertainty to ca. 3%.[88] Nonetheless, the uncertainty in determining the charge state and hence the molecular weight remains a problem in the analysis of large nucleic acids by ESI/FTICR/MS. Fuerstenau and Benner generated ions from plasmid and phage DNA ranging from 2.8 to 31 MDa by ESI.[85] The analyzer was a TOF mass spectrometer equipped with a novel charge-sensitive detector that allows for an independent determination of the m/z ratio and the charge of the ion. With this detector (a Faraday cage-like metal tube) the image charge signal of an individual ion is amplified by a low-noise charge-sensitive amplifier and the m/z corresponds to the time that the ion needs to pass the detector tube. Again, the uncertainty in determining the charge states decreases the mass accuracy. Refining this technique has led to the development of a gated electrostatic ion trap,[89] in which an individual ion is trapped between two plates and, therefore, traverses the detector tube

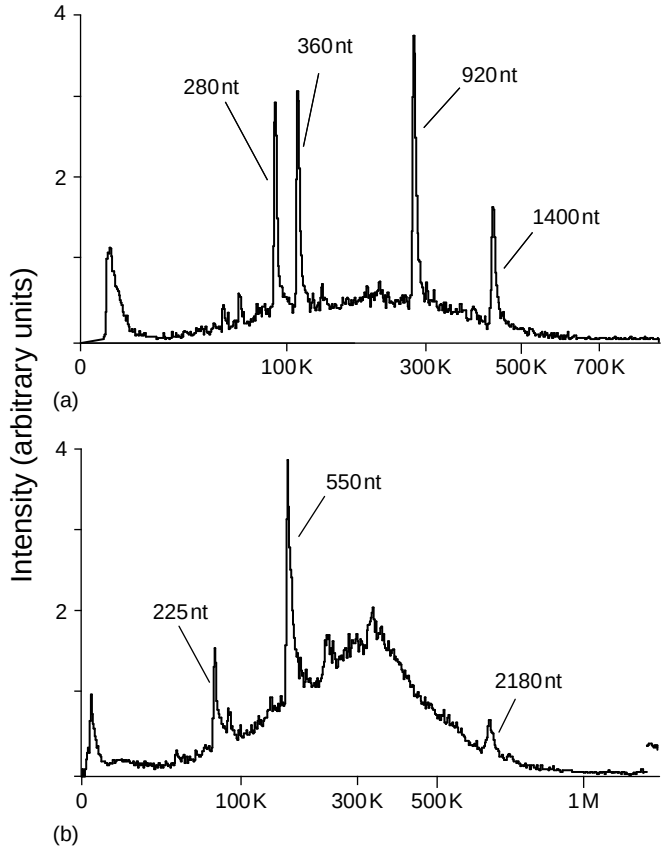

Figure 5 IR/MALDI reflectron TOF mass spectra of (a) plasmid DNA restriction enzyme digest (pBluescript-KS$^+$; *Bgl*I and *Rsa*I; 87, 112, 285, and 433 kDa) and (b) plasmid DNA restriction enzyme digest (pBluescript-KS$^-$; *Not*I and *Ssp*I; 70, 170, and 673 kDa). (Adapted from S. Berkenkamp et al., *Science*, **281**, 260–262 (1998) with permission from American Association for the Advancement of Science, Copyright 1998.)

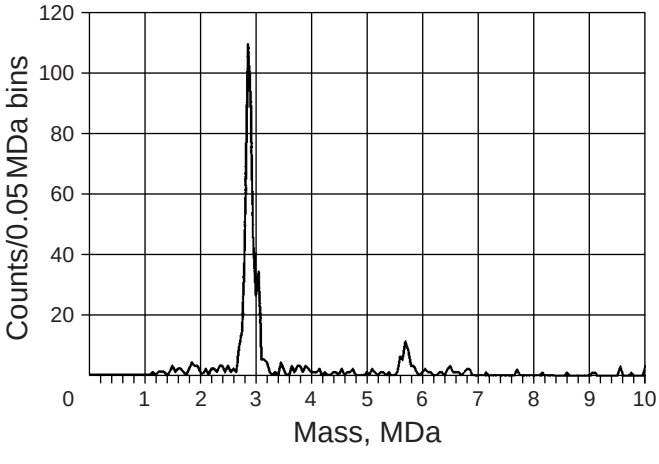

Figure 6 Histogram of the average mass obtained from each of 549 pBR322 trapped ions. (Reprinted from W.H. Benner, *Anal. Chem.*, **69**, 4162–4168 (1997) with permission from American Chemical Society, Copyright 1997.)

several times. The single measurements of charge state and m/z can be averaged and hence reduce the signal-to-noise ratio compared with a single measurement. Figure 6 shows the histogram of the average mass acquired from 549 trapped pBR322 plasmid ions. The signal at mass 2.88 Da corresponds to the sodiated (protonated) pBR322 ion and the signal at 5.85 Da to the dimer.

2.2 Mixture Analysis and Combination of Mass Spectrometry with Separation Techniques

The ability to analyze mixtures is an important feature because many samples, such as those from a Sanger reaction, contain a number of individual species of different mass. Furthermore, the impurities of real-life samples and the heterogeneity (e.g. different modifications) of the sample can be determined. One important point for the analysis of mixtures in which the analytes are present in different concentrations is the dynamic range. This parameter is dependent on the analyzer used and describes the concentration difference between two components when both are still detectable – one signal distinguishable from the background noise (e.g. a signal-to-noise ratio greater than 2) and the other signal near saturation. Another critical point in mixture analyses with MALDI is that higher-mass analytes often require higher desorption laser fluxes. At these fluxes, the resolution of lower-mass analyte signals may already be lower than that attainable for the pure low-mass compounds. As mentioned before, the difficulty in analyzing larger molecules may have its cause in a higher degree of fragmentation and a decreased detection efficiency.[69] Larger molecules may also be in competition with smaller molecules for the charges. For an easy interpretation of the spectra it is beneficial if every component has only one major charge state, i.e. one peak in the MALDI mass spectrum. Therefore, the positive ion mode should be preferred for MALDI/MS, particularly with 3-HPA as the matrix, as in the negative ion mode multiply charged ions are observed to a larger extent.[57] This is also the reason why ESI/MS with its multiply charged ion distributions is inferior to MALDI/MS for mixture analyses if the masses of the mixture components extend over a wide mass range. However, ESI is superior for the analysis of mixtures which vary only slightly in mass, because of its inherently higher-resolution mass analyzers. Therefore, the demand on resolution depends on the expected mass differences between adjacent peaks. An elegant way of performing mixture analysis is on-line coupling with different separation techniques such as chromatography and electrophoresis.[90] In addition to the separation of molecules which may differ only slightly from each other, the elimination of sample impurities is a great advantage of such combinations. Electrospray is especially suited to the combination with liquid chromatography because the analyte is ionized from solution.

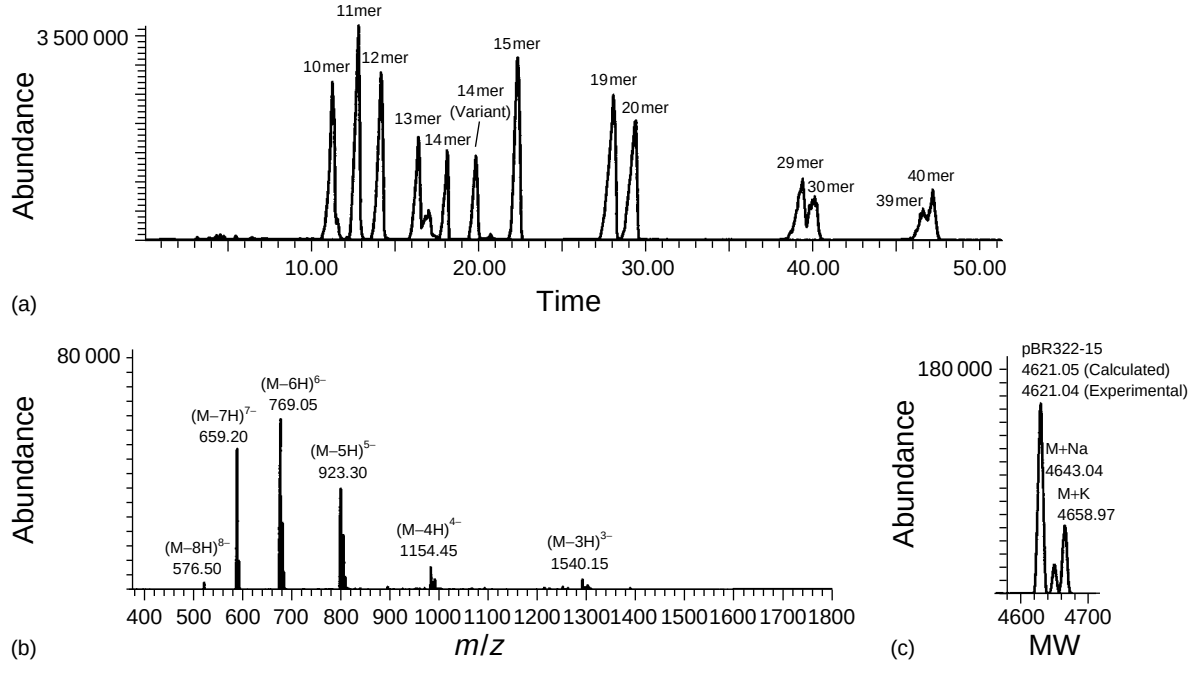

Figure 7 Analysis of synthetic pBR322 oligonucleotides by HPLC/ESI/MS. (a) Total ion chromatogram, (b) raw mass spectrum, (c) deconvoluted mass spectrum. (Reprinted from A. Apffel et al., *Anal. Chem.*, **69**, 1320–1325 (1997) with permission from American Chemical Society, Copyright 1997.)

However, the buffers have to be optimized with respect to both the chromatographic separation and the MS ionization steps. An ammonium acetate or triethylammonium acetate–acetonitrile gradient was used in a concentration of 5–10 mM by Bleicher and Bayer.[91] The good compatibility with the ionization was compromised by the low separation efficiency. Apffel et al.[92] introduced a gradient system with 1,1,1,3,3,3-hexafluoro-2-propanol adjusted to pH 7 with triethylamine, which yielded a superior separation resolution and also higher ionization efficiency (Figure 7).

This buffer system was also used by Griffey et al.[93] for the characterization of phosphorothioate (replacement of the nonbridging oxygen of the phosphate backbone by sulfur) oligodeoxynucleotides from in vivo metabolism in pig kidney. The coupling of HPLC to an ESI ion trap made it possible to determine the base compositions of the respective metabolites of the injected phosphorothioate DNA. Sequence information could also be obtained via fragmentation. Thus 3′ and 5′ exonuclease activity and, presumably, also endonuclease activity as well as the addition of single deoxyribo- and ribonucleotides were observed. For MALDI/MS both off-line and on-line coupling with chromatographic and electrophoretic techniques have been reported. On-line coupling is less common and requires modification of the preparation technique.[94,95] A detailed discussion on combining MALDI with liquid separation techniques is available.[96]

2.3 Noncovalent Complexes

An understanding of biomolecular interactions helps in comprehending life on the molecular level. Noncovalent interactions between biomolecules govern all cellular processes. Knowledge of the recognition between proteins and nucleic acids, such as between polymerase and DNA, is of great importance to the understanding of cell division and cell proliferation. The detection of noncovalent complexes using MALDI/MS is not possible with the preparation protocols mentioned so far. This indicates that the choice of matrix and preparation are of vital importance. A general characteristic of MALDI mass spectra is the frequent observation of the most probable gas-phase-generated unspecific analyte oligomers. These oligomer signals are generally much weaker than the quasimolecular ion peaks. Furthermore, their intensities decrease with increasing degree of oligomerization. Although these characteristics can aid in distinguishing specific complexes from unspecific adducts, great care has to be taken to verify the nature of noncovalent complexes, e.g. by competition experiments and careful attention to the relative signal intensities.

The first noncovalent nucleic acid complex observed by MALDI was reported by Lecchi and Pannell. They

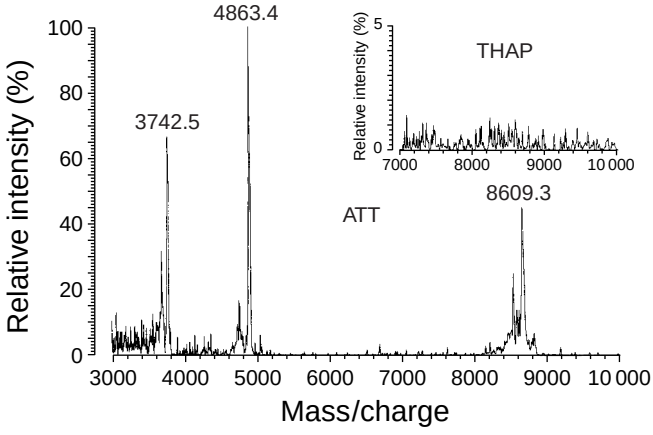

Figure 8 MALDI spectrum of EcoR1 12/16 adapter with ATT as the matrix. The DNA dimer region with 2,4,6-THAP as the matrix is shown for comparison at a 20-fold expansion in the inset. Only peaks from the single-stranded oligonucleotides were observed with 2,4,6-THAP. The ion at 8609.3 (calculated value 8606.6) in the ATT spectrum represents the detection of dsDNA. (Adapted from P. Lecchi and L.K. Pannell, *J. Am. Soc. Mass Spectrom.*, **6**, 972–975 (1995) with permission from Elsevier Science, Copyright 1995 by American Society for Mass Spectrometry.)

succeeded in detection of the EcoR1 adaptor double-stranded 16/12-mer in addition to the single strands in 1995 using ATT as the matrix.[43] No unspecific oligomerization was observed. Control experiments with 3-HPA and THAP showed no signal for the double strand (Figure 8).

Digestion with nuclease S1, which is much more active in cleaving ssDNA than dsDNA, was used to show that the single strands, as well as the duplex, originated from the original sample rather than being generated by the MALDI process. The same authors reported later that those chemicals having a stabilizing effect on oligonucleotide structure in solution also improve the detection of specific complexes. Thus, calcium as well as spermidine substantially enhanced the detection of a dsDNA 12/16-mer, whereas the denaturing agent formamide completely prevented detection of the complex.[97] It was also stated that the stabilizing effect of calcium on the duplex is somewhat counteracted by a loss of mass resolution due to cation attachment. Kirpekar et al. investigated the applicability of both UV/MALDI and IR/MALDI for the detection of dsDNA.[98] For double strands up to 70 bp, UV/MALDI and the matrix ATT were found to be superior. However, double strands exceeding 70 bp could be better analyzed by IR/MALDI because of the strongly decreased fragmentation and higher sensitivity. As pointed out by Lecchi and Pannell.[97] agents stabilizing the structure in solution (40 mM ammonium acetate or 40 mM tris(hydroxymethyl)aminomethane (Tris)–HCl) were found to have a positive effect on the detection of the

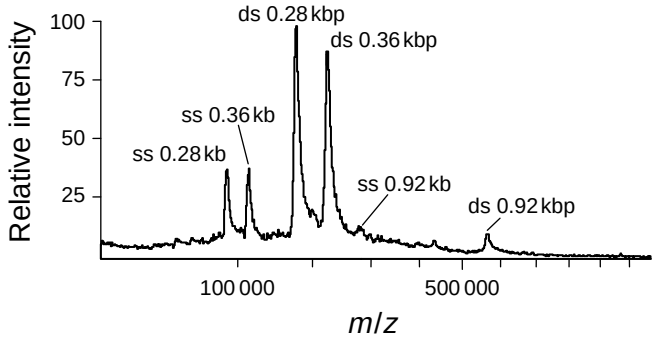

Figure 9 IR/MALDI reflectron TOF mass spectrum of a mixture of dsDNA fragments recorded from a glycerol/ammonium acetate matrix. The sample was generated by digestion of the plasmid pBluescript-KS$^+$ with the restriction enzymes *Bgl*I and *Rsa*I. (Reprinted from F. Kirpekar et al., *Anal. Chem.*, **71**(13), 2334–2339 (1999) with permission from American Chemical Society, Copyright 1999.)

complex. Further addition of a destabilizing agent such as succinic acid (acidic denaturation) hindered detection of the complex. The largest dsDNA detected was a 1.05 kbp fragment of about 645 kDa in mass. The authors stated that the quality of spectra with fragments exceeding 1 kbp of length is significantly reduced. Figure 9 shows the IR/MALDI reflection TOF mass spectrum of a mixture of dsDNA generated by the digestion of the plasmid pBluescript-KS+ with the enzymes *Bgl*I and *Rsa*I using glycerol/ammonium acetate matrix.

As an example of a biologically relevant protein–nucleic acid complex, Gruic-Sovulj et al. reported the specific complex between aminoacyl tRNA synthetases (serine (SerRS) and tyrosine (TyrRS)) and their cognate tRNAs ($tRNA^{Ser}$ and $tRNA^{Tyr}$).[99] In that case, the specificity of the complexes was verified by competition experiments testing both tRNAs against a given synthetase.

Because of the gentleness of the desorption ESI lends itself even more than MALDI to the detection of noncovalent associations. Instrumental parameters such as temperature and declustering voltage of the ion source as well as the chemical environment of the sample have a great influence on the success of the analysis. As nonspecific complexes have also been reported for this ionization technique,[100–102] care should be taken in spectra interpretation. Analysis of noncovalent interactions of oligonucleotides by ESI have been reported since 1993.[103–105] Schnier et al.[106] supplied evidence for the occurrence of Watson–Crick bps of the dsDNA in the gas phase by measuring the dissociation kinetics using blackbody IR radiative dissociation in a Fourier transform instrument. The activation energies for dissociation of specific noncovalent complexes were significantly higher than those obtained for nonspecific associations. Moreover, a correlation between the gas-phase dissociation activation energies and the dimerization enthalpies in solution was observed. Interaction between dsDNA and metal ions[107] as well as drugs[108–110] have also been investigated and collision-induced dissociation (CID) of the complexes was found to be a facile approach with which to verify the correctness of the noncovalent binding.[110–112] Cheng et al. determined the binding stoichiometry of gene V protein from bacteriophage f1 with oligonucleotides through competition experiments.[113] Oligonucleotides with 16 nucleotides showed a stoichiometry of 4 : 1 (protein monomer/oligonucleotide), whereas oligonucleotides with 15 nucleotides and lower had a stoichiometry of 2 : 1. Control experiments at low pH values (acid-induced denaturation) displayed no complex formation, whereas the complexes were unaffected by an increase of the interface capillary temperature of the instrument. These results agreed with those obtained by size-exclusion chromatography. Furthermore, competition experiments involving the transcription factor PU.1 DNA binding domain (PU.1-DBD) and dsDNA

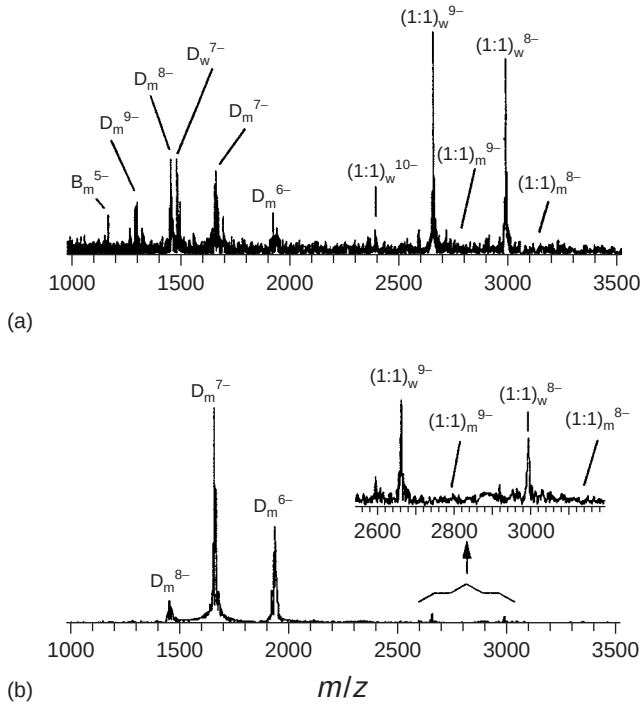

Figure 10 ESI/MS spectra of PU.1-DBD and mixtures of DNA in 10 μM ammonium acetate (pH 7.0) acquired on a fourier transform ion cyclotron resonance (FTICR) instrument. Mixtures contain (a) PU.1-DBD (5 μM), 17 bp DNA (15 μM), and 19 bp DNA (20 μM); (b) PU.1-DBD (3 μM), 17 bp DNA (10 μM); and 19 bp DNA (200 μM). Abbreviations in the spectra: D = dsDNA, B = heavy strand of dsDNA, and w = wild-type and m = mutant-type DNA sequences. (Reprinted from X. Cheng et al., *Anal. Biochem.*, **239**, 35–40 (1996) with permission from Academic Press, Copyright 1996.)

containing (17 bp, wild type) or missing (19 bp, mutant) the recognition motif revealed the specificity of the complex.[114] Even for a 20-fold excess of the dsDNA lacking the binding motif only the specific complex was detected (Figure 10).

The stability of a specific noncovalent tryptophan repressor-DNA-complex was shown by Potier et al.[115] with CID. Dissociation of this protein-dimer/operator DNA complex could be achieved only under conditions which also cleaved covalent bonds, probably depurinating the DNA.

These examples for the detection of noncovalent complexes illustrate that MS is a valuable method as long as the specificity of the complex can be verified and the complex reflects the true solution structure. Information on binding stoichiometry, dissociation constants,[116] base composition, and even structural information can be obtained in addition to proof of the existence of the complex itself.[117–119]

2.4 Sequencing Approaches

Because of the accessible mass range and the decrease in resolution with increasing length of nucleic acids, the available mass spectrometric sequencing techniques present no challenge to the conventional techniques used, e.g. in the human genome project. Nevertheless, MS is an accurate, rapid, and sensitive tool ideally suited for sequencing shorter nucleic acids. Consequently, MS sequencing holds great promise in the fields of clinical diagnostics, forensics, paternity analysis, livestock breeding, plant cultivation, and cell line typing. One of the advantages of MS over gel-based Sanger sequencing lies in the fact that the mass is an intrinsic property of the molecule rather than an extrinsic one such as electrophoretic mobility. Furthermore, the mass can provide evidence for the occurrence of modifications in the analyte and an accurate mass measurement allows for the determination of base composition, as was shown by Pomerantz et al.[120] Aaserud et al.[121] demonstrated that analyzing dsDNA with ESI/FTICR defines or restricts the bp composition up to 600 bp. Sequencing strategies can be divided in three categories: sequencing by process-induced fragmentation, sequencing via chemically generated fragments, and sequencing using enzymatic reactions.

2.4.1 Sequencing with Process-induced Fragmentation

Although fragmentation hampers MALDI analysis of nucleic acids to a great extent it can also be capitalized on to gain sequence information when the fragmentation is controlled. The emphasis lies on the term controlled. The desired cleavage pattern should be a consecutive disjunction of the nucleotide constituents at specific bonds. An additional, but less intense, loss of bases and phosphate groups can help to characterize and locate possible modification sites. However, multiple base losses and predominance of certain nucleotide cleavages over others will only disturb or even prevent sequence determination. It has to be kept in mind that the impact of fragmentation on the mass spectrum depends on the decay time in relation to the instrumentally determined time frame for the analysis. Depending on the mass, the flight time of an ion in a TOF mass spectrometer is in the microsecond to low millisecond range. This dependence of the decay time constant is particularly pronounced for TOF analyzers. Therefore, a short description of their relevant properties is given here. More detailed information about mass analyzers is given in the instrumentation section of this encyclopedia.

2.4.1.1 Decay Time in Regions where the Electric Field is Nonzero These are primarily the region of acceleration right in the source and in the ion reflector. As a rule of thumb, fragments generated in these regions are distributed and are therefore unpredictable in mass. For TOF instruments with static ion extraction a fragment is termed prompt if its decay time constant is comparable to the generation time of its parent ion. Such fragment ions will encounter the full acceleration voltage and are detected at their correct m/z value in the mass spectrum. Ions that fragment shortly after the generation of the parent ions (fast fragments) up to the time until the parent ion leaves the accelerating region have an energy deficit and are, therefore, not detected at their correct mass in the spectrum. Ions resulting from this type of fragmentation are not suited for sequencing, because the true fragment masses cannot be determined.

2.4.1.2 Decay in Regions where the Electrical Field is Zero With the implementation of DE, ions with longer decay time constants can, however, provide valuable sequence information.[122] Fragments, generated during the delay time of typically a few hundred nanoseconds are detected at their true masses in the spectrum. These ions are called in-source decay (ISD) fragments. Metastable fragmentation occurs in the field-free drift tube of the TOF analyzer, i.e. after the parent ion has left the acceleration region (post-source decay (PSD)). These fragments continue to fly with very nearly the same velocity as the parent ion and are detected together with the parent ions in instruments with a linear geometry. This is also true if the mass difference compared to the parent ion is small enough for the associated energy deficit to be compensated by a reflectron. (A reflectron functions as an ion mirror and is used for compensating kinetic energy distributions of the generated ions.) With increasing mass loss, metastable decay

fragments manifest themselves in the spectra recorded with a reflectron TOF as a shoulder on the low-mass side of the peak. This type of fragmentation can be used for sequencing, if the deceleration potential of the reflectron is lowered in steps to display a small mass window of fragments whose mass can be determined with sufficient accuracy (PSD spectrum[123]).

Process-induced fragmentation has the advantage that no additional steps of preparation and sample handling are required. Nordhoff et al. have reported direct mass spectrometric sequencing of DNA up to 21 bases by IR/MALDI (Er/YAG laser, 2.94 μm) in the negative ion mode.[60] With low-picomole sample amounts and the matrix succinic acid three series of sequence-specific prompt fragmentations were observed (Figure 11).

In contrast to the explanation given by Grotjahn et al.,[19] Nordhoff et al. found that the 80 Da mass shift is not an HPO_3^+ artifact, but originates from an additional sugar moiety of the 3′ terminal phosphate group of d. Loss of a base is followed by cleavage of the 3′ C–O bond and results in fragments w and (a-B). Further loss of the furan moiety from (a-B) leads to d (see Figure 4). This sort of analysis is impeded by a relatively low shot-to-shot reproducibility, the competing high rate of metastable decay of the DNA, and the lack of fragmentation at T sites. Prompt fragmentation providing sequence information for DNA was also investigated in the negative ion mode

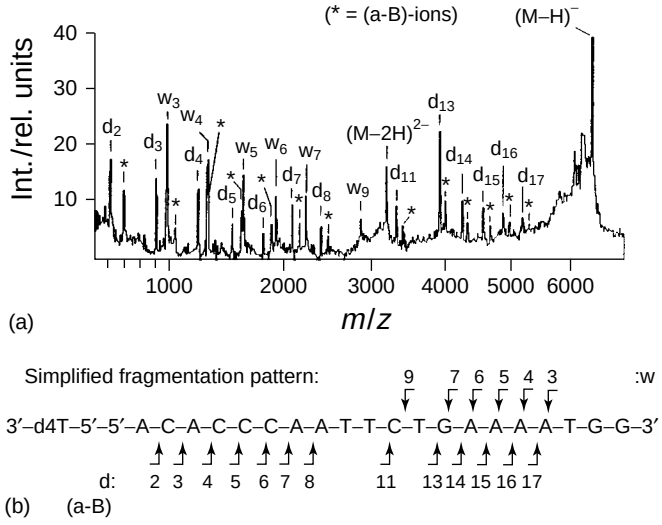

Figure 11 (a) Negative ion IR/MALDI reflectron TOF mass spectrum of a modified antisense-DNA 21-mer obtained with succinic acid as matrix. The defected d-, (a-B)-, and w-ions are assigned. The modified nucleoside d4T (2′,3′-dideoxy-2′, 3′-didehydrothymidine) has been added to the 5′ end of the antisense-DNA 20-mer by a 5′,5′ phosphate diester linkage. (b) Simplified fragmentation pattern derived from (a). (Adapted from E. Nordhoff et al., *J. Mass Spectrom.*, **30**, 99–112 (1995) with permission from John Wiley & Sons Limited, Copyright 1995.)

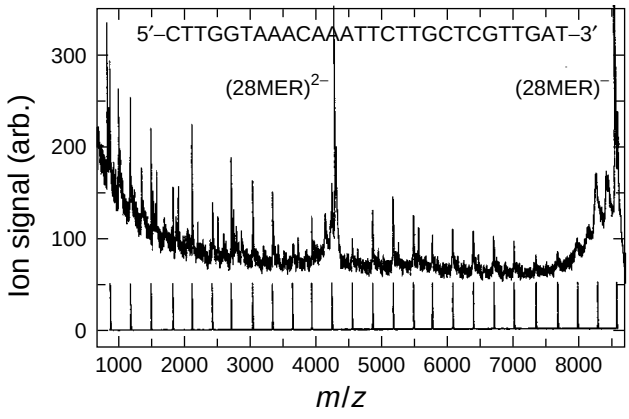

Figure 12 The MALDI spectrum of negative ions for a 28-mer 5′-CTT GGT AAA CAA ATT CTT GCT CGT TGA T-3′ oligonucleotide and its simulated spectrum of y-fragment ions. (Reprinted from Y.E. Zhu et al., *Rapid Commun. Mass Spectrom.*, **11**, 897–903 (1997) with permission from John Wiley & Sons Limited, Copyright 1997.)

with a UV laser (N_2) emitting at 337 nm.[124] Sequencing of DNA up to 35 nucleotides was demonstrated with the matrix systems 2-aminobenzoic acid and 2,4,6-THAP, both with dibasic ammonium citrate. Figure 12 shows the mass spectrum for a DNA 28-mer.

The full sequence can be read from the 5′ terminus, except for the last three nucleotides which are buried in the matrix background. Interestingly only the 5′ terminal w and y fragments appear in the spectrum. An additional loss of single bases was at least seen for a DNA 10-mer. Juhasz et al.[75] applied DE for sequencing DNA in the low-picomole range based on ISD in the negative ion mode. Whereas the combination of 3-HPA with an N_2 laser emitting at 337 nm showed no fragment ions of a DNA 11-mer, an ion series consisting of w, d, and also y ions was observed with PA at a laser wavelength of 266 nm (frequency quadruplet neodymium/yttrium–aluminum–garnet (Nd/YAG) laser). No a ions or base losses were detected. The w ions provided complete sequence information. Wang et al.[125] also applied ISD to sequence modified oligonucleotides with a combination matrix of 3-HPA and *N*-(3-indolylacetyl)-l-leucine at higher laser irradiances, significantly above the ion detection threshold. Sample amounts prepared were in the 30–160 pmol range. The fragmentation pattern of several modified DNAs composed of among others phosphorothioate and methylphosphonate (replacement of the hydroxyl group in the phosphate backbone by a methyl group) linkages, 2′-methoxygroups and DNA/RNA chimeras contain several different ion types for both ion polarities. The complete w series was observed accompanied by a, (a-B), and d ions and, at lower abundance, b, c, d, and z ions (Figure 13).

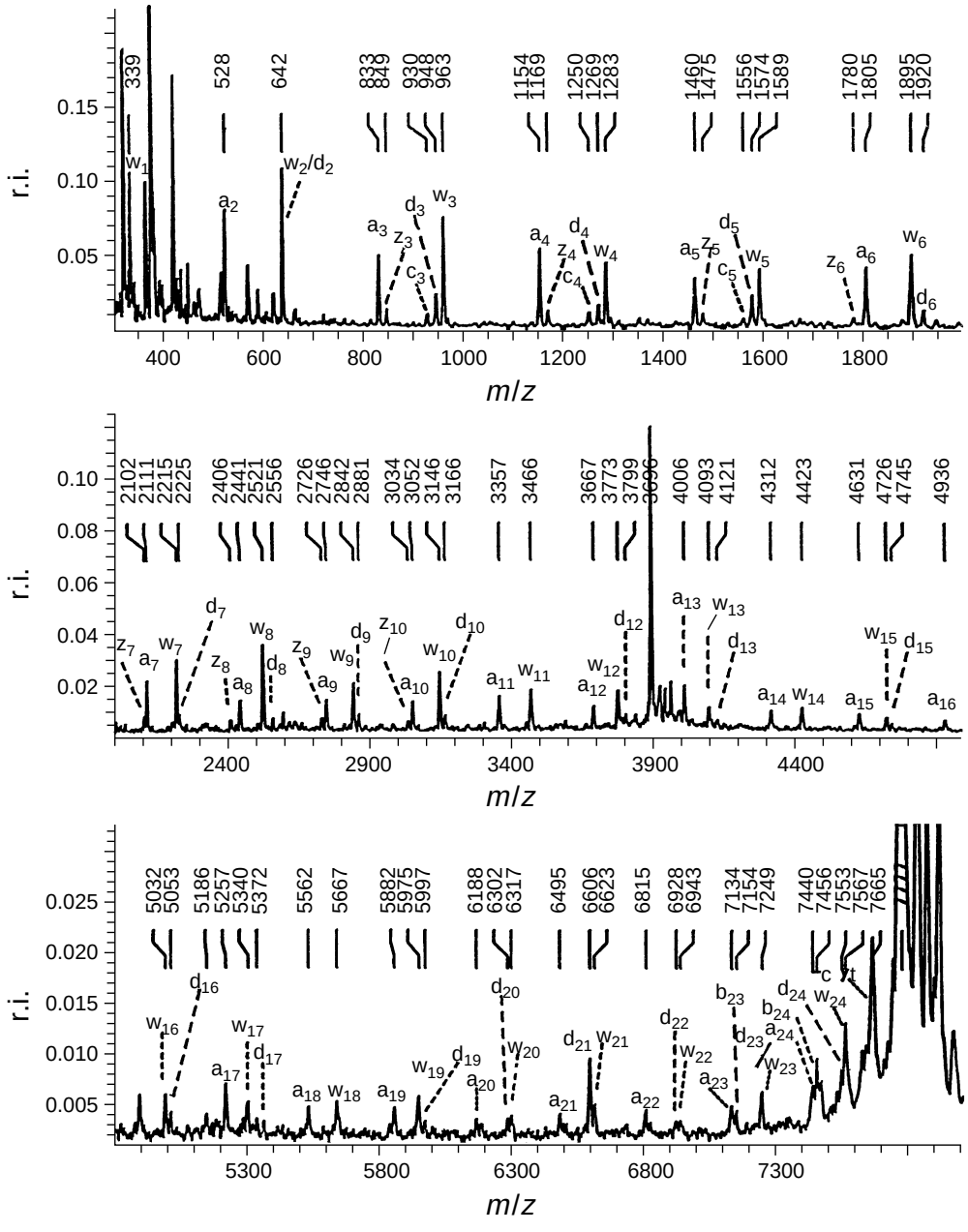

Figure 13 MALDI/TOF/MS of a 25-mer all-phosphorothioate oligonucleotide with the sequence 5′-CsTsCsTsCsGsCsAsCsCsCsAs TsCsTsCsTsCsTsCsCsTsTsCsT-3′ (with s indicating the phosphorothioate backbone). (Reprinted from B.H. Wang et al., *Int. J. Mass Spectrom. Ion Processes*, **169/170**, 331–350 (1997) with permission from Elsevier Science, Copyright 1997.)

For these samples no fragments were detected without DE, even at higher laser irradiances. An explanation given by the authors for the occurrence of a ions is the high density of matrix ions, radicals, neutrals, and analyte in the generated plume. The absence of this ion type in the measurements of Juhasz et al.[75] may be attributed to the fact that, in addition to the different matrix–wavelength combination, much lower sample amounts were prepared.

Sequencing by PSD can be achieved routinely for peptides and small proteins,[126] but this method is more difficult to apply to oligonucleotides. The first PSD spectra were reported by Kirsch et al.,[127] representing a DNA 5-mer analyzed in the positive ion mode. The spectrum is complicated by multiple base losses from the parent ion as well as from the fragments. As was observed for prompt fragmentation,[60,56] T sites in the molecule are only cleaved with very low intensities. Wang

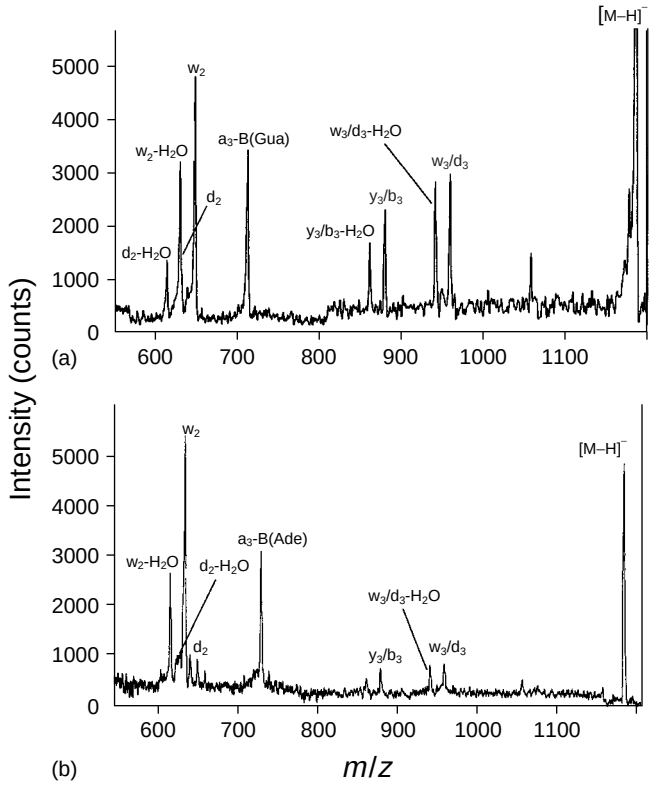

Figure 14 Comparison of MALDI/PSD spectra for the structural isomers: (a) dTAGT and (b) dTGAT. (Reprinted from Z. Wang et al., *J. Am. Soc. Mass Spectrom.*, **9**, 683–691 (1998) with permission from Elsevier Science, Copyright 1998 by American Society for Mass Spectrometry.)

et al.[64] demonstrated the feasibility of PSD/MALDI to differentiate between structural isomers in the negative ion mode. Figure 14 illustrates the distinction between dTAGT and dTGAT by different w_2 and (a_3-B_3) ions.

PSD analyses of modified oligonucleotides have also been reported.[128,129] The coupling of MALDI to FTICR to investigate process-induced fragmentation of oligodeoxynucleotides with and without modifications have been reported by Buchanan and colleagues.[130–133] Unfortunately the lower fragment masses dominate these mass spectra.[133]

Under standard ionization conditions fragmentation is less prominent for ESI. However, fragmentation can be induced in the atmospheric pressure region between the ESI spray needle and the skimmer orifice (nozzle skimmer (NS) dissociation),[134] by collisional activation with inert gas molecules (i.e. by CID)[135] or by irradiation of the ions with an IR blackbody (or laser), usually from a CO_2 laser source, in the FTICR cell (infrared multiphoton dissociation (IRMPD)).[136] Quadrupole ion traps,[63,137–139] FTICR cells[136,140–142] and triple quadrupole instruments[143–145] have been used for these investigations. Fragmentation depends on the amount of energy transferred to the analyte. Owing to the higher collision energy in quadrupoles as compared with ion traps, fragmentation is more pronounced in the former, leading to a more complete sequence information. In FTICR instruments, the combined application of NS dissociation, CID, and IRMPD can provide complete sequence information up to 50 nucleotides and extensive fragmentation up to 108 nucleotides.[142] Figure 15 shows the spectra obtained with (a) NS dissociation and (b) CID of a DNA 8-mer.

For process-induced fragmentation the sample amount needed is often elevated, which is of concern for biological samples, but usually not for synthetic nucleic acids. The complexity of the mass spectrum and possible isobaric fragments of partially modified oligonucleotides somewhat limits these applications for routine complete sequencing of unknown samples. However, sequence verification of oligonucleotides can easily be performed with these approaches.

2.4.2 Sequencing with Chemically and Enzymatically Generated Fragments

Bond cleavage can also be achieved with enzymes or chemical agents added to the sample solution prior to MS analysis. One advantage of chemical over enzymatic cleavage is the less pronounced dependence on the secondary structure of the analyte. Furthermore, modifications of the backbone which lead to resistance against enzymatic degradation can be capitalized on to gain sequence information. Thus methylphosphonate DNA (used in antisense oligonucleotides) is resistant to nuclease digestions but can be sequenced by chemical means. Keough et al.[146] demonstrated the use of synthesis failure sequences for MALDI/MS to verify the identity of the synthetic product. Owing to the nature of the synthesis and the yield of 96–99% per synthesis cycle, a ladder consisting of the specific failure sequences is generated. In a subsequent paper Keough et al. combined failure synthesis analysis with exonuclease digestion and limited alkaline hydrolysis of the methylphosphonate linkages, allowing the complete sequence of a modified 18-mer to be read.[147]

Failure sequences allow one to read the sequence from the 5′ terminus of the oligonucleotide, whereas snake venom exonuclease successively cleaves nucleotides from the 3′ end. Sequence information can be completed by alkaline hydrolysis because the methylphosphonate linkages are exclusively susceptible to base-catalyzed hydrolysis. Polo et al.[148] demonstrated that the approach to sequence phosphorothioate-containing oligonucleotides with alkylating agents, first presented by Nakamaye et al.,[149] can be adapted for mass spectrometric analysis – 2-iodoethanol specifically alkylates

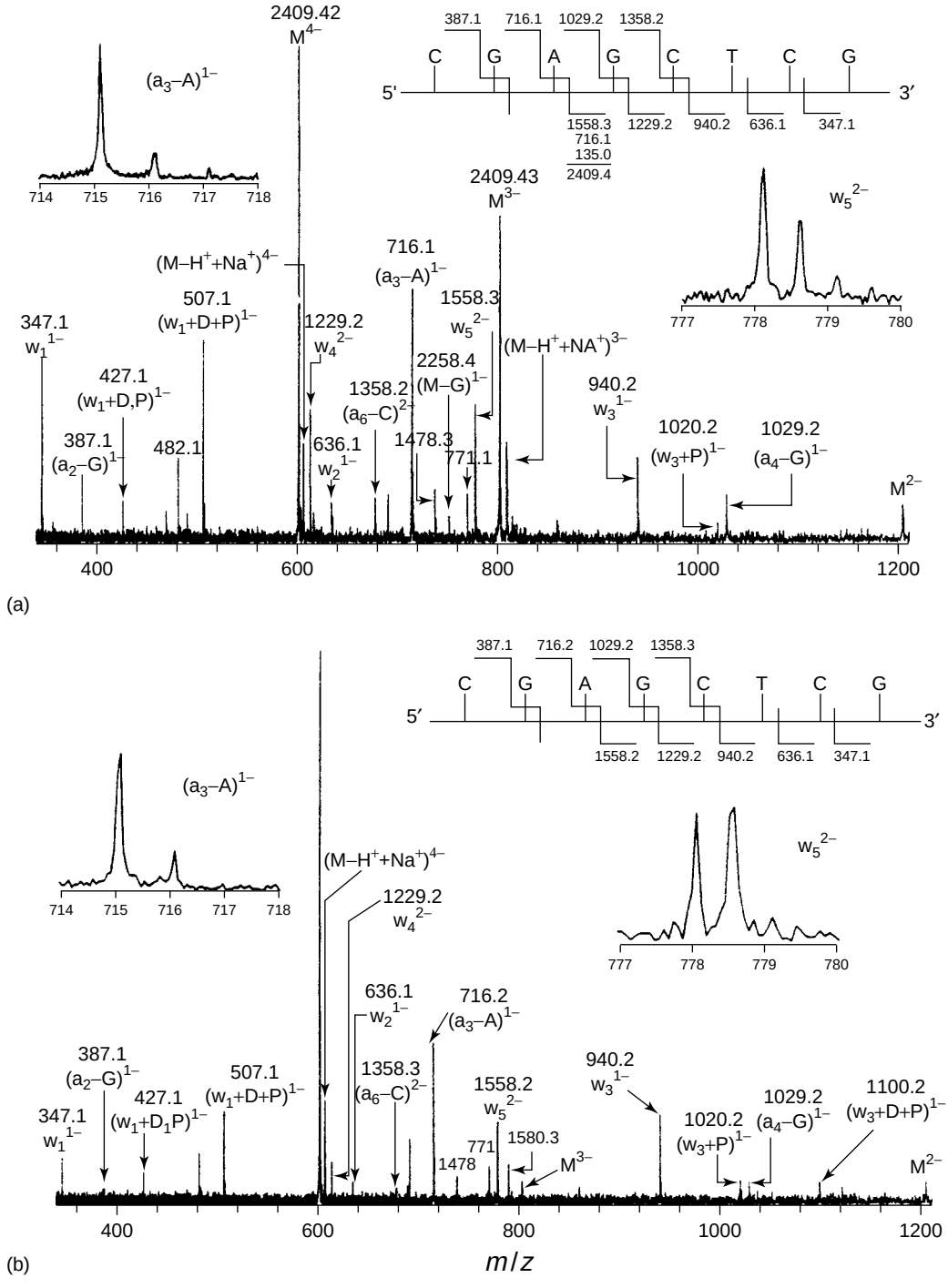

Figure 15 ESI mass spectra of 5′-CGA GCT CG-3′ with a calculated mass of 2409.45 Da. D and P refer to the deoxy and phosphate moieties, respectively. (a) NS and (b) CID spectra. (Adapted from D.P. Little et al., *J. Am. Chem. Soc.*, **116**, 4893–4897 (1994) with permission from American Chemical Society, Copyright 1994.)

phosphorothioate linkages, and cleavage occurs generating 3′ and 5′ fragment ladders. Although the complete sequence could be obtained for an all-phosphorothioate DNA 10-mer without being disturbed by the presence of internal fragments, the spectrum quality was poor, particularly the signal-to-noise ratio of the fragments. Limited alkaline hydrolysis was used by Hahner et al.[150] in combination with endonuclease cleavage of RNA to deduce sequence information. Alkaline hydrolysis alone provided nearly the complete sequence information for

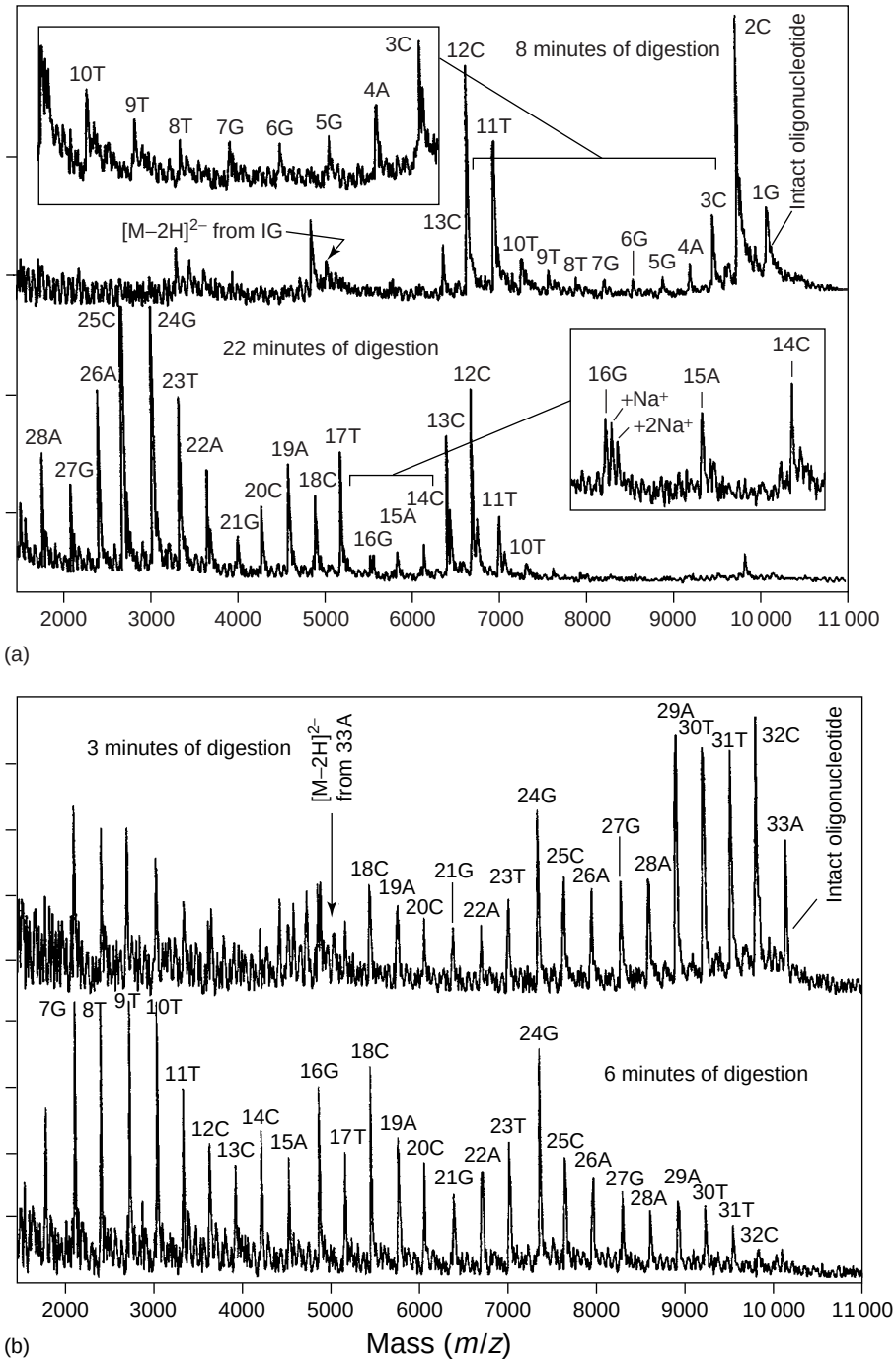

Figure 16 Comparison of partial 3′ → 5′ and 5′ → 3′ exonuclease digestions analyzed by DE/MALDI. The time of each digestion is indicated. Both peak 1G in (a) and 33A in (b) correspond to the undigested 33-mer of sequence 5′-GCC AGG GTT TCC CCA GTC ACG ATG CAG AAT TCA. (a) 5′ → 3′ digestion with calf spleen phosphodiesterase. (b) 3′ → 5′ digestion with snake venom phosphodiesterase. (Adapted from I.P. Smirnov et al., *Anal. Biochem.*, **238**, 19–25 (1996) with permission from Academic Press, Copyright 1996.)

a synthetic RNA 25-mer. Because of the mass difference of only 1 Da between U and C, differentiating between them is difficult and a high resolution is vital, especially for large specimens. The use of base-specific endonucleases helped to solve this problem, but enzymatic specificity and efficiency may be compromised under the conditions used for MALDI analysis.[52] Tolson and Nicholson[151] applied hydrazine and aniline

as chemical agents for the specific cleavage of uridine sites in RNA to verify the sequence information obtained by exonuclease digestion. McCloskey and colleagues developed an ESI/MS-based approach[52,152] that involves cleavage with base-specific endonucleases to locate and determine modifications in RNA. Major contributions to the localization and characterization of RNA modifications were given by the McCloskey and Crain group.[153–155] Nonspecific endonucleases were used by Linscheid and colleagues in combination with capillary zone electrophoresis ESI[156,157] to determine the sites of styrene adducts. The use of exonucleases to generate sequence ladders[158] with subsequent mass spectrometric analysis was first reported by Pieles et al.[46] involving the phosphodiesterase from snake venom and calf spleen (the latter cleaves specifically from the 5′ terminus) and was soon adapted by others. Smirnov et al.[159] optimized the cleavage conditions specifically for the MALDI preparation and presented excellent quality spectra containing the complete sequence information of a DNA 33-mer (Figure 16).

The analysis of exonuclease-generated sequence ladders with ESI was shown by Limbach et al.,[160] simplifying the interpretation of the complex mass spectrum by a continuous introduction of the digest solution into the ESI source.

Although gel-based Sanger sequencing can be performed routinely for DNA of up to 500 bp, it can be severely disturbed by gel-band compression or other mobility shifts of the generated fragments. In 1993 Fitzgerald et al.[161] demonstrated that the classical Sanger approach can be adapted for detection by MALDI/MS by analyzing a mixture of synthetic oligonucleotides which mimicked a sequence ladder. These and other similar experiments soon revealed that the removal of template, buffer, salts, and nucleotides prior to the MALDI analysis was crucial.[162,163] This can be achieved by immobilizing the PCR product on solid phases such as streptavidin-coated beads.[83,164] Cycle sequencing is another straightforward approach as the template DNA strand is copied several times by a thermostable DNA polymerase to generate increased amounts of the sequence fragments.[165] Fu et al.[166] reported the sequencing of exons 5 to 8 of the p53 gene (670 bases in total) by primer walking with 21 primers – the nature and site of the polymorphisms of heterozygous DNA templates were elucidated. Taranenko and colleagues[167,168] reported the sequencing of double-stranded templates up to 200 bp with a one-strand-specific primer. Kirpekar et al.[54] analyzed clinical heterozygously expressed samples and demonstrated the superiority of MALDI/MS in exploring the murine PPARδ exon 1 PCR fragment. A 10 bp long part of this fragment was not amenable to gel electrophoresis as a consequence of pronounced band compression and marked nonspecific termination. Figure 17 illustrates the de novo sequencing of this problematic sequence segment.

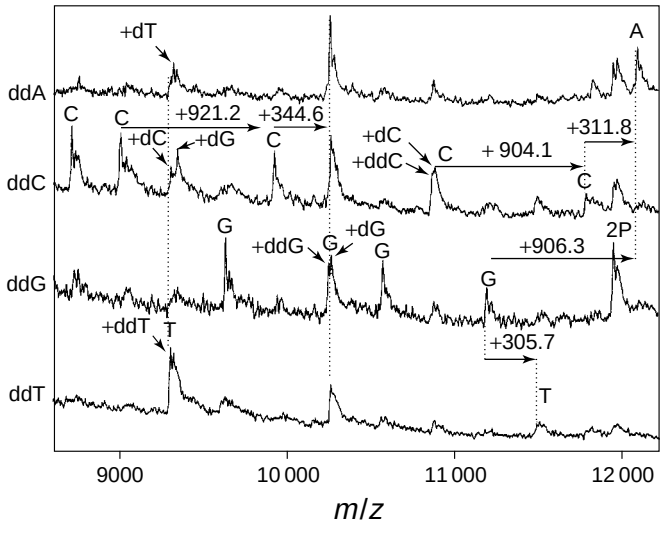

Figure 17 Generation of de novo sequence information by MADLI/MS product separation. The cycle sequencing was performed on a murine PPARδ, exon 1 PCR fragment. Horizontal arrows show the mass differences used for sequence determination of problematic positions in the sequence. 2P marks the signal of the dimer of the sequencing primer. (Reprinted from F. Kirpekar et al., *Nucleic Acids Res.*, **26**, 2554–2559 (1998) with permission from Oxford University Press, Copyright 1998.)

In this investigation an unexpected nucleotide polymerization event was detected at the third position. Misincorportion of dGMP and also dCMP – clearly seen in the ddC reaction – at this position leads to termination whereas the strong signal in the ddT spectrum reveals the presence of T at this position. False and specific termination can be distinguished from each other as the mass difference of two adjacent signals in the mass spectrum provides information about the type of nucleotide. Doublets (mass difference 16 Da) can be observed in the respective mass spectrum as the incorrect product carries a deoxynucleotide rather than the dideoxynucleotide at the terminus. Further advantages of coupling Sanger sequencing of rather short templates with MS are the ability to obtain sequence information in close proximity to the primer,[168] the high speed of analysis and the ability for automation in process lines. Köster et al. captured biotinylated DNA templates with streptavidin-coated magnetic beads and performed primer extension terminated with the respective ddNTPs in four reactions for sequencing up to 63 nucleotides.[165] Only the primer-extended products are detected as the ionization–desorption process does not disrupt the biotin–streptavidin linkage.[164]

2.5 Clinical Diagnostics with Mass Spectrometry

Clinical diagnostics do not necessarily depend on complete sequencing. Various rapid screening methods for mutations and polymorphisms have been developed which can be easily performed by MS. Multiplexing greatly enhances the amount of information that can be gained in a single measurement. A classical approach to screen for mutations in a given DNA strand is a restriction endonuclease digest followed by gel-electrophoretic separation of the generated fragments, in which differences of the fragment length and/or the pattern are indicators of mutations (restriction fragment length polymorphism RFLP). Lubman and colleagues have adapted restriction endonuclease cleavage of PCR-amplified DNA with subsequent detection by MALDI/MS to screen buccall cell DNA up to 426 bp for polymorphisms.[81,169] The cystic fibrosis gene was the target for Ch'ang et al.[170] who screened the DNA of 30 individuals for the most common ΔF508 deletion using MALDI/MS. Another strategy pursued by this group involves PCR amplification of the DNA of interest (cystic fibrosis transmembrane conductance regulator, CTFR gene) with two pairs of primers, one pair for the wild-type DNA and one pair for the mutant DNA, thus leading to different PCR products for the homozygote wild type, heterozygote, and homozygote mutant patients.[171] Jurinke et al.[172] screened blood bank samples for hepatitis B virus (HBV) DNA by amplifying its encoding core antigen (HBc) with PCR. Biotinylation of one of the primers enabled purification after immobilization on streptavidin-coated magnetic beads and detection of the complementary strand after hybridization with MALDI/MS. The sensitivity was enhanced in following work[173] with the aid of nested PCR, in which HBV DNA was detected in 11 out of 21 sera of patients supposedly cured (positive isolated anti-HBc). The same group reported a rapid screening method for known mutation sites by primer oligo base extension (PROBE) and MALDI/MS.[174] This strategy relies on the annealing of a specific primer, or in a biplex reaction of two primers, downstream of the suspected mutation site and primer extension, in which one of the four deoxynucleotides is completely substituted by its dideoxy analog. Mutations manifest themselves through the length of the PROBE product, and the number of signals are indicators of the heterozygosity of the allele. Performing PROBE as a temperature-cycled rather than an isothermal assay leads to an enhancement in sensitivity.[175] Braun et al.[176] demonstrated that the PROBE reaction is also suited for the analysis of microsatellites by MS if the repeat unit does not consist of all four nucleotides. Investigating the number of AluVpA microsatellite repeat units of the five members of a given family by digestion of the respective PCR product with a restriction endonuclease and subsequent separation by gel electrophoresis showed 13 and 12 repeat units for the mother, 11 and

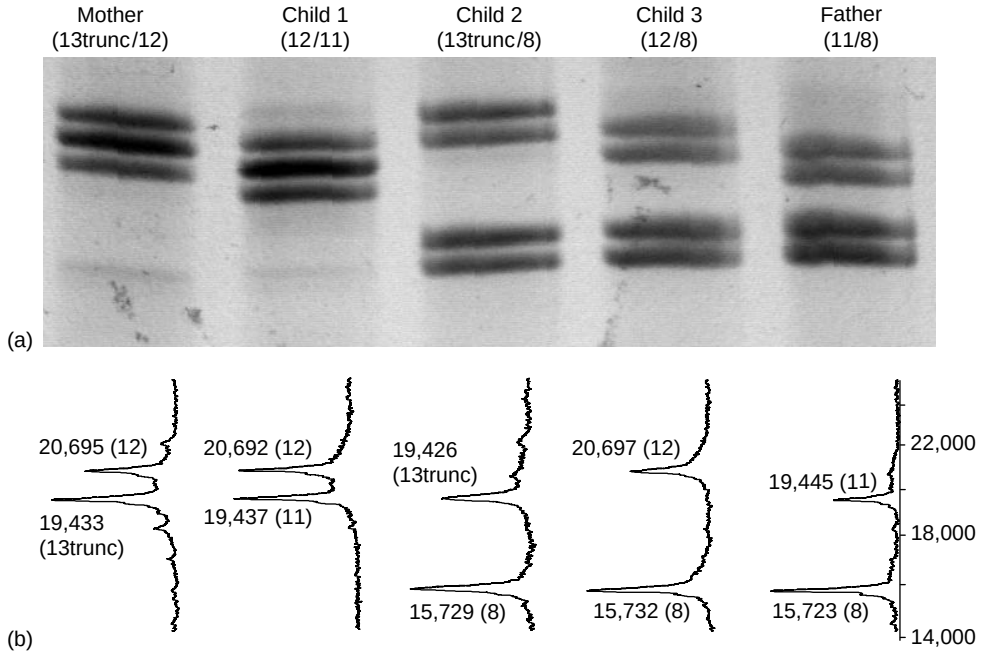

Figure 18 Comparison of an AluVpA microsatellite analysis using (a) conventional denaturing gel electrophoresis followed by silver staining of DNA fragments and (b) measurements of diagnostic PROBE products (ddGTP termination) by MS. (Reprinted from A. Braun et al., *Genomics*, **46**, 18–23 (1997) with permission from Academic Press, Copyright 1997.)

Table 1 Genotyping primers used in model multiplex PinPoint assay

Locus no.	Locus name/polymorphism	PinPoint primer sequence (5′-3′)	Primer M, (Da)	Extended primer mass window (Da)
1	Antithrombin III (AT3)	ctccatgggcccagc	4513.98	4787.17–4827.19
2	Cytochrome P450 2D6 (CP450)	aatgatgagaacctg	4625.09	4898.28–4938.30
3	Neurofibromatosis (NFI)	ttgtcaccatattaatt	5134.42	5407.61–5447.63
4	Complement component C6 (CC6)	ggggacagccatgcactg	5549.65	5822.84–5862.87
5	Alpha-2-macroglobulin (A2M)	acagcagcttactccagag	5781.83	6055.02–6095.04
6	Insulin growth factor (IGF)	ccagcaaagagaaaagaagg	6226.15	6499.34–6539.36
7	Triglyceride lipase (TGL)	ttctgtcttccaggaatctgt	6378.19	6651.38–6691.41
8	Aldolase B (ALDB)	cgggccaagaaggtatctacc	6440.25	6713.43–6753.46
9	Integrin B2 subunit (ISB2)	gggacatagtgaccgtgcaggt	6840.49	7113.68–7153.70
10	Low-density lipoprotein receptor (LDLR)	ttttatgacaccgtcatcagcag	6998.61	7271.80–7311.83
11	Interleukin 1 alpha (IL 1A)	gggaaatcatcaagcctaggtca	7081.68	7354.87–7394.89
12	Protein S alpha (PRSA)	atgatattagagctcactcatgtcc	7616.02	7889.21–7929.23

Molecular weight values of primers and extended primers are based on the following nucleotide exact masses: dC, 289.1847; dT, 304.1964; dA, 313.2097; dG, 329.2091; ddC, 273.1898; ddT, 288.2015; ddA, 297.2148; ddG, 313.2142.

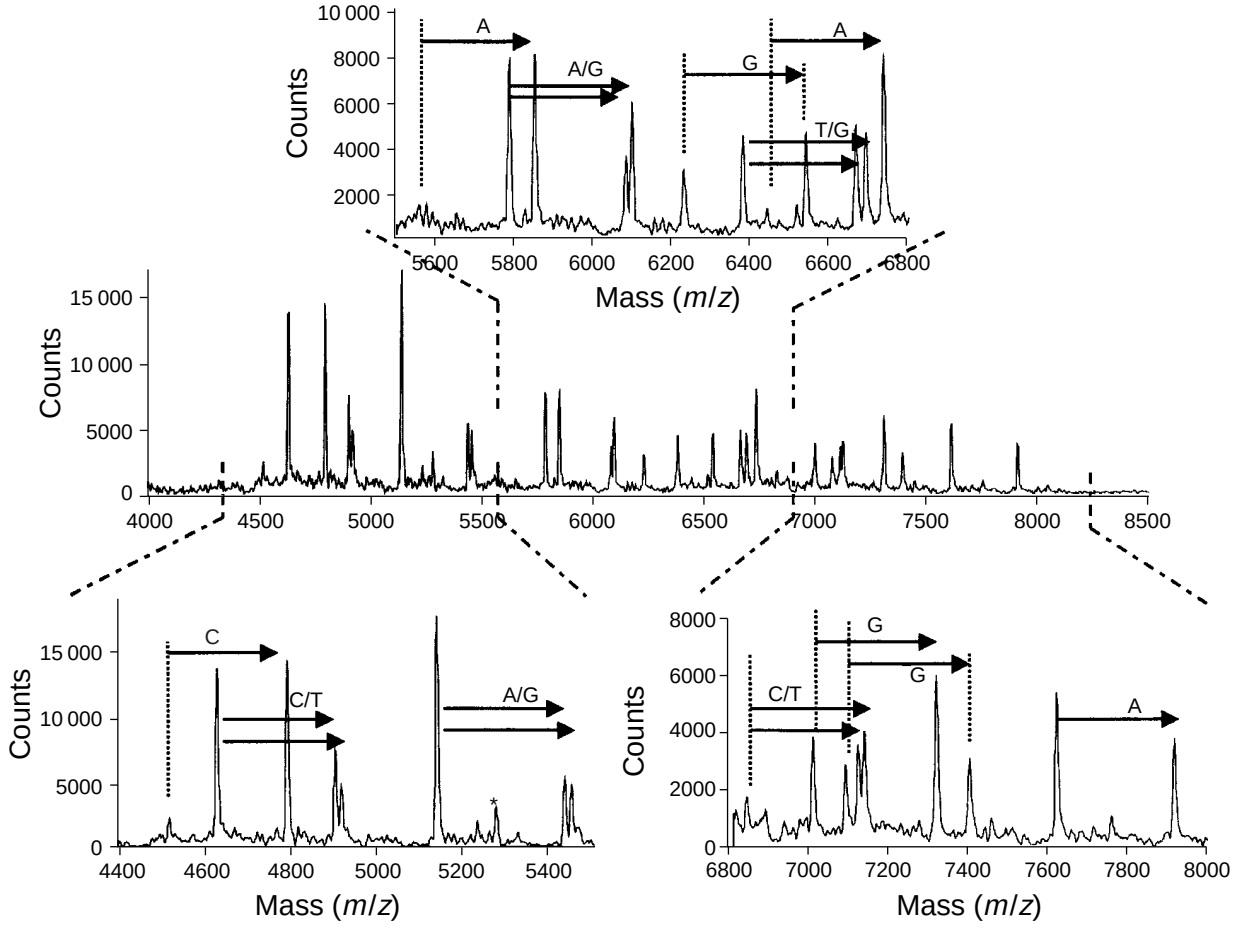

Figure 19 Twelvefold multiplex spectrum expanded in three regions, illustrating primer-extended primer pairs (horizontal arrows); $*$ = matrix adduct peaks. The table lists the used genotyping primers for the corresponding polymorphism. (Reprinted from P. Ross et al., *Nat. Biotechnol.*, **16**, 1347–1351 (1998) with permission from Nature America, Copyright 1998.)

8 for the father, and combinations thereof for the three children, whereas the PROBE reaction in combination with MALDI/MS revealed second-site mutations within the 13 repeat unit. It was shown by sequence analysis that transitions of T to C in the second and T to G in the eighth repeat lead to a truncated version of the 13 repeat when the PROBE reaction is performed with ddGTP (Figure 18)

Since its invention PROBE assay has found applications in the analysis of more than 100 mutations/polymorphisms of human genes.[70,177]

Another rapid screening method for the detection of single nucleotide polymorphisms, nucleotide deletion, insertion, and substitution is the pinpoint assay.[178] This approach uses the annealing of specific primers downstream of the putative mutation site/polymorphic region. The analytic information is contained in the mass shift of the single base-extended primer. The assay has a high potential for multiplexing. Figure 19 shows the 12-fold multiplex spectrum of loci listed in the table.

Wunschel et al.[179] used ESI/FTICR analysis for the detection of heterogeneity in the highly conserved 16/23S rDNA interspacing region from *Bacillus cereus* 6464. Interestingly, more PCR-amplified double-strands were detected than expected. The accurate mass measurement of the single-strands revealed a missing T in the coding and a missing A in the noncoding strand, in addition to the predicted masses. Laken et al.[180] analyzed genetic variants of PCR-amplified DNA with ESI ion trap MS coupled to HPLC. The key of this approach, called SOMA (short oligonucleotide mass analysis), is the implantation of an artificial restriction endonuclease type IIS recognition site within the employed primer pair, hence determining the length of the amplified DNA and the length of the digest fragments (7–20 nucleotides) which mirror the genetic variation. The cleavage of the PCR-amplified DNA with restriction endonucleases ensures that the fragment length is amenable for ion trap MS. Figure 20 describes the investigation of common polymorphisms in the APC gene at the codons 485, 545, and 1756.

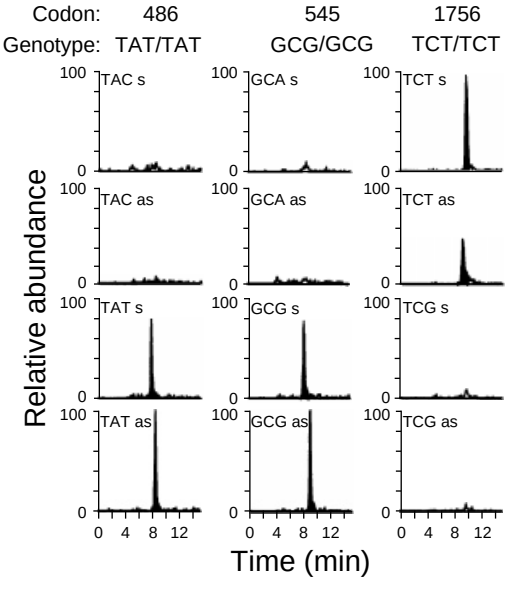

Figure 20 Simultaneous analysis of three different APC variants for two patients. For each patient, PCR products containing APC codons 486, 545, and 1756 were combined and introduced into the mass spectrometer via the HPLC. The sense (s) and antisense (as) signals are indicated for each genotype. (Reprinted from S.L. Laken et al., *Nat. Biotechnol.*, **16**, 1352–1356 (1998) with permission from Nature America, Copyright 1998.)

The simultaneous detection of the three different codons after combining the separate PCR reactions and digests illustrates the ability to analyze multiplexed reactions. Furthermore, the ion trap can be used to analyze induced fragmentation of the single DNA stretches, providing even more information in cases of isobaric single strands.

Performing diagnostics with MALDI/MS has the advantage that the analysis can be easily combined with computer technology, allowing large-scale diagnostics using parallel processing, whereas ESI is superior when analyzing larger analytes.

2.6 Conclusions

MS is only a recent addition to the plethora of analytical techniques in nucleic acid analysis. Its contributions to the fields of biology, biochemistry, biomedicine, and biotechnology will, no doubt, increase substantially in the near future as new applications and analytical procedures are developed.

ACKNOWLEDGMENTS

We would like to thank Hans C. Lüdemann, Finn Kirpekar and Kerstin Strupat for a critical reading of the manuscript and helpful discussions.

ABBREVIATIONS AND ACRONYMS

A	Adenine
ATT	6-Aza-2-thiothymine
bp	Base Pair
C	Cytosine
CI	Chemical Ionization
CID	Collision-induced Dissociation
DE	Delayed Ion Extraction
dNMP	deoxyribonucleoside 5′ monophosphate
dsDNA	Double-stranded DNA
EI	Electron Impact
EI/MS	Electron Impact Mass Spectrometry
Er/YAG	Erbium/Yttrium–Aluminum–Garnet
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FAB/MS	Fast Atom Bombardment/Mass Spectrometry

FD	Field Desorption
FTICR	Fourier Transform Ion Cyclotron Resonance
FTICR/MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
FWHM	Full-width at Half-maximum
G	Guanine
HBV	Hepatitis B Virus
HPLC	High-performance Liquid Chromatography
ISD	In-source Decay
IR	Infrared
IRMPD	Infrared Multiphoton Dissociation
MALDI	Matrix-assisted Laser Desorption Ionization
MALDI/MS	Matrix-assisted Laser Desorption Ionization/Mass Spectrometry
MS	Mass Spectrometry
NC	Nitrocellulose
Nd/YAG	Neodymium/Yttrium–Aluminum–Garnet
NS	Nozzle Skimmer
PA	Picolinic Acid
PCR	Polymerase Chain Reaction
PD	Plasma Desorption
PROBE	Primer Oligo Base Extension
PSD	Post-source Decay
RFLP	Restriction Fragment Length Polymorphism
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SI	Secondary Ion
SIMS	Secondary Ion Mass Spectrometry
SOMA	Short Oligonucleotide Mass Analysis
ssDNA	Single-stranded DNA
T	Thymine
TIC	Total Ion Current
TOF	Time-of-flight
THAP	Trihydroxyacetophenone
Tris	Tris(hydroxymethyl)aminomethane
U	Uracil
UV	Ultraviolet
3-HPA	3-Hydroxypicolinic Acid

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Clinical Chemistry **(Volume 2)**
DNA Arrays: Preparation and Application • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Molecules, Properties and Detection of Single • Polymerase Chain Reaction and Other Amplification Systems • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Sequencing Strategies and Tactics in DNA and RNA Analysis

Mass Spectrometry **(Volume 13)**
Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Quadrupole Ion Trap Mass Spectrometer • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

Forensic Science **(Volume 5)**
Polymerase Chain Reaction in the Forensic Analysis of DNA

REFERENCES

1. K. Biemann, J.A. McCloskey, 'Application of Mass Spectrometry to Structure Problems. VI. Nucleosides', *J. Am. Chem. Soc.*, **84**, 2005–2007 (1962).
2. P.F. Crain, 'Mass Spectrometric Techniques in Nucleic Acid Research', *Mass Spectrom. Rev.*, **9**, 505–554 (1990).
3. J.A. McCloskey, 'Constituents of Nucleic Acids: Overview and Strategy', *Methods Enzymol.*, **193**, 771–781 (1990).
4. J.A. McCloskey, P.F. Crain, 'Progress in Mass Spectrometry of Nucleic Acid Constituents: Analysis of Xenobiotic Modifications and Measurements at High Mass', *Int. J. Mass Spectrom. Ion Proc.*, **118/119**, 593–615 (1992).
5. J.A. McCloskey, 'Experimental Approaches to the Characterization of Nucleic Acid Constituents by Mass Spectrometry', in *Mass Spectrometry in Biomedical Research*, ed. S.J. Gaskell, John Wiley & Sons, Chichester, 75–95, 1986.
6. D.R. Knapp, 'Chemical Derivatization for Mass Spectrometry', *Methods Enzymol.*, **193**, 314–329 (1990).
7. D.F. Hunt, C.E. Hignite, K. Biemann, 'Structure Elucidation of Dinucleotides by Mass Spectrometry', *Biochem. Biophys. Res. Commun.*, **33**(3), 378–383 (1968).
8. J.L. Wiebers, J.A. Shapiro, 'Sequence Analysis of Oligonucleotides by Mass Spectrometry. 1. Dinucleoside Monophosphates', *Biochemistry*, **16**(6), 1044–1050 (1977).
9. D.R. Burgard, S.P. Perone, J.L. Wiebers, 'Sequence Analysis of Oligodeoxyribonucleotides by Mass Spectrometry. 2. Application of Computerized Pattern Recognition to Sequence Determination of Di-, Tri-,

and Tetranucleotides', *Biochemistry*, **16**(6), 1051–1057 (1977).
10. H.-R. Schulten, H.M. Schiebel, 'High Resolution Field Desorption Mass Spectrometry IV. Dinucleotides', *Z. Anal. Chem.*, **280**, 139–144 (1976).
11. H. Budzikiewicz, M. Linscheid, 'Field Ionization Mass Spectrometry I-Field Desorption Spectra of Nucleotides – Experimental Problems', *Biomed. Mass Spectrom.*, **4**(2), 103–106 (1977).
12. C.J. McNeal, S.A. Narang, R.D. Macfarlane, H.M. Hsiung, R. Brousseau, 'Sequence Determination of Protected Oligodeoxyribonucleotides Containing Phosphotriester Linkages by Californium-252 Plasma Desorption Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **77**(2), 735–739 (1980).
13. C.J. McNeal, K.K. Ogilvie, N.Y. Theriault, M.J. Nemer, 'A New Method for Sequencing Fully Protected Oligonucleotides using ^{252}Cf-Plasma Desorption Mass Spectrometry. 3. Positive Ions', *J. Am. Chem. Soc.*, **104**(4), 981–984 (1982).
14. A. Viari, J.-P. Ballini, P. Meleard, P. Vigny, P. Dousset, C. Blonski, D. Shire, 'Characterization and Sequencing of Normal and Modified Oligonucleotides by ^{252}Cf Plasma Desorption Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **16**, 225–228 (1988).
15. C.J. McNeal, K.K. Ogilvie, N.Y. Theriault, M.J. Nemer, 'A New Method For Sequencing Fully Protected Oligonucleotides using ^{252}Cf-Plasma Desorption Mass Spectrometry. 2. Negative Ions of Subunits in the Stepwise Synthesis of a Heptanucleotide', *J. Am. Chem. Soc.*, **104**(4), 976–980 (1982).
16. R. Beavis, W. Ens, M.J. Nemer, K.K. Ogilvie, K.G. Standing, J.B. Westmore, 'Secondary Ion Mass Spectrometry of Protected Oligonucleotides', *Int. J. Mass Spectrom. Ion Phys.*, **46**, 475–478 (1983).
17. K.J. Wu, R.W. Odom, 'Matrix-enhanced Secondary Ion Mass Spectrometry: A Method for Molecular Analysis of Solid Surfaces', *Anal. Chem.*, **68**(5), 873–882 (1996).
18. L. Grotjahn, R. Frank, H. Blöcker, 'Ultrafast Sequencing of Oligodeoxyribonucleotides by FAB-Mass Spectrometry', *Nucleic Acids Res.*, **10**(15), 4671–4678 (1982).
19. L. Grotjahn, H. Blöcker, R. Frank, 'Mass Spectroscopic Sequence Analysis of Oligonucleotides', *Biomed. Mass Spectrom.*, **12**, 514–524 (1985).
20. A.M. Hogg, J.G. Kelland, J.C. Vederas, C. Tamm, 'Nucleosides and Nucleotides Part 24: Investigation of Ribo- and Deoxyribonucleosides and -nucleotides by Fast-atom-bombardment Mass Spectrometry', *Helv. Chim. Acta*, **69**, 908–917 (1986).
21. M. Panico, G. Sindona, N. Uccella, 'Bioorganic Applications of Mass Spectrometry. 3. Fast-atom-bombardment-induced Zwitterionic Oligonucleotide Quasimolecular Ions Sequenced by MS/MS', *J. Am. Chem. Soc.*, **105**(17), 5607–5610 (1983).
22. L. Grotjahn, H. Steinert, 'Application of Fast Atom Bombardment-mass Spectrometry in Gene and Protein Design', *Biochem. Soc. Trans.*, **15**, 164–170 (1987).
23. T. Matsuo, T. Sakurai, H. Matsuda, M. Matsugi, M. Ikehara, T. Kobayashi, Y. Kammei, E. Kubota, 'Characterization of Oligoribonucleotides up to 24 mer (M/z 13726) by Sector Type Mass Spectrometer', Proceedings of the 34th Annual Conference on Mass Spectrometry and Allied Topics, ASMS, Cincinnati, Ohio, 329–330, 1986.
24. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, 'Electrospray Ionization for Mass Spectrometry of Large Biomolecules', *Science*, **246**, 64–71 (1989).
25. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**(20), 2299–2301 (1988).
26. P.A. Limbach, P.F. Crain, J.A. McCloskey, 'Characterization of Oligonucleotides and Nucleic Acids by Mass Spectrometry', *Curr. Opin. Biotechnol.*, **6**, 96–102 (1995).
27. E. Nordhoff, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry as a New Method for the Characterization of Nucleic Acids', *Trends Anal. Chem.*, **15**(6), 240–250 (1996).
28. E. Nordhoff, F. Kirpekar, P. Roepstorff, 'Mass Spectrometry of Nucleic Acids', *Mass Spectrom. Rev.*, **15**(2), 67–138 (1996).
29. P.F. Crain, 'Electrospray Ionization Mass Spectrometry of Nucleic Acids and their Constituents', in *Electrospray Ionization Mass Spectrometry*, ed. R.B. Cole, John Wiley & Sons, 421–457, 1997.
30. P.F. Crain, J.A. McCloskey, 'Applications of Mass Spectrometry to the Characterization of Oligonucleotides and Nucleic Acids', *Curr. Opin. Biotechnol.*, **9**, 25–34 (1998).
31. C.R. Cantor, P.R. Schimmel, 'The Conformation of Biological Macromolecules', in *Biophysical Chemistry*, Part I, ed. A.C. Bartlett, W.H. Freeman and Company, 1980.
32. J. Clauwaert, J. Stockx, 'Interactions of Polynucleotides and their Components. I. Dissociation Constants of the Bases and their Derivatives', *Z. Naturforsch. B*, **23**, 25– (1968).
33. S. Lewin, D.A. Humpfreys, 'Reactions of Nucleic Acids and their Components. Part IV. Thermodynamic Constants of the Ionizations of Cytosine and Cytidine', *J. Chem. Soc. (B)*, 210–211 (1966).
34. P.O.P. Ts'o, 'Bases, Nucleosides, and Nucleotides', in *Basic Principles in Nucleic Acid Chemistry*, ed. P.O.P. Ts'o, Academic Press, Vol. 1, 1974.
35. N.K. Kochetov, E.I. Budovskij, 'Electronic Structure and Reactivity of the Monomer Components of Nucleic Acids', in *Organic Chemistry of Nucleic Acids*, Part A, Plenum Press, 1971.
36. F. Greco, A. Lugiori, G. Sidona, N. Uccella, 'Gas-phase Proton Affinity of Deoxyribonucleosides and Related Nucleobases by Fast-atom-bombardment Tandem Mass

Spectrometry', *J. Am. Chem. Soc.*, **112**(25), 9092–9096 (1990).
37. M.T. Rodgers, S. Campbell, E.M. Marzluff, J.L. Beauchamp, 'Site-specific Protonation Directs Low-energy Dissociation Pathways of Dinucleotides in the Gas-phase', *Int. J. Mass Spectrom. Ion Proc.*, **148**, 1–23 (1995).
38. M.T. Rodgers, S. Campbell, E.M. Marzluff, J.L. Beauchamp, 'Low-energy Collision-induced Dissociation of Deprotonated Dinucleotides: Determination of the Energetically Favored Dissociation Pathways and the Relative Acidities of the Nucleic Acid Bases', *Int. J. Mass Spectrom. Ion Proc.*, **137**, 121–149 (1994).
39. W. Saenger, *Principles in Nucleic Acid Structure*, ed. C.R. Cantor, Springer Advanced Texts in Chemistry, 1984.
40. J.T. Stults, J.C. Marsters, 'Improved Electrospray Ionization of Synthetic Oligodeoxynucleotides', *Rapid Commun. Mass Spectrom.*, **5**(8), 359–363 (1991).
41. P.F. Crain, 'Characterization of Oligonucleotides by Electrospray Mass Spectrometry', in *Mass Spectrometry of Biological Materials*, ed. B.S. Larsen, C.N. McEwen, Marcel Dekker, New York, 389–404, 1998.
42. J. Gross, K. Strupat, 'Matrix-assisted Laser Desorption/Ionization-Mass Spectrometry (MALDI/MS) Applied to Biological Macromolecules', *Trends Anal. Chem.*, **17**(8 + 9), 470–484 (1998).
43. P. Lecchi, L.K. Pannell, 'The Detection of Intact Double-stranded DNA by MALDI', *J. Am. Soc. Mass Spectrom.*, **6**(10), 972–975 (1995).
44. Y.F. Zhu, C.N. Chung, N.I. Taranenko, S.L. Allman, S.A. Martin, L. Haff, C.H. Chen, 'The Study of 2,3,4-Trihydroxyacetophenone and 2,4,6-Trihydroxyacetophenone as Matrices for DNA Detection in Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**(3), 383–388 (1996).
45. G.J. Currie, J.R. Yates, 'Analysis of Oligodeoxynucleotides by Negative-ion Matrix-assisted Laser Desorption Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**(12), 955–963 (1993).
46. U. Pieles, W. Zürcher, M. Schär, H.W. Moser, 'Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry: a Powerful Tool for the Mass and Sequence Analysis of Natural and Modified Oligonucleotides', *Nucleic Acids Res.*, **21**(14), 3191–3196 (1993).
47. E. Nordhoff, A. Ingendoh, R. Cramer, A. Overberg, B. Stahl, M. Karas, F. Hillenkamp, P.F. Crain, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Nucleic Acids with Wavelengths in the Ultraviolet and Infrared', *Rapid Commun. Mass Spectrom.*, **6**(11), 771–776 (1992).
48. E. Nordhoff, R. Cramer, M. Karas, F. Hillenkamp, F. Kirpekar, K. Kristiansen, P. Roepstorff, 'Ion Stability of Nucleic Acids in Infrared Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Nucleic Acids Res.*, **21**(15), 3347–3357 (1993).
49. P.A. Limbach, P.F. Crain, J.A. McCloskey, 'Molecular Mass Measurement of Intact Ribonucleic Acids Via Electrospray Ionization Quadrupole Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **6**(1), 27–39 (1995).
50. I. Gruic-Sovulj, H.C. Lüdemann, F. Hillenkamp, I. Weygand-Durasevic, Z. Kucan, J. Peter-Katalinic, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Transfer Ribonucleic Acids Isolated from Yeast', *Nucleic Acids Res.*, **25**(9), 1859–1861 (1996).
51. T.A. Shaler, J.N. Wickham, K.A. Sannes, K.J. Wu, S.H. Becker, 'Effect of Impurities on the Matrix-assisted Laser Desorption Mass Spectra of Single-stranded Oligodeoxynucleotides', *Anal. Chem.*, **68**(3), 576–579 (1996).
52. J.A. Kowalak, S.C. Pomerantz, P.F. Crain, J.A. McCloskey, 'A Novel Method for the Determination of Post-transcriptional Modification in RNA by Mass Spectrometry', *Nucleic Acids Res.*, **21**(19), 4577–4585 (1993).
53. M. Kussmann, E. Nordhoff, H. Rahbek-Nielsen, S. Haebel, M. Rossel-Larsen, L. Jakobsen, J. Gobom, E. Mirgorodskaya, A. Kroll-Kristensen, L. Palm, P. Roepstorff, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry Sample Preparation Techniques Designed for Various Peptide and Protein Analytes', *J. Mass Spectrom.*, **32**(5), 593–601 (1997).
54. F. Kirpekar, E. Nordhoff, K. Larsen, K. Kristiansen, P. Roepstorff, F. Hillenkamp, 'DNA Sequence Analysis by MALDI Mass Spectrometry', *Nucleic Acids Res.*, **11**(26), 2554–2559 (1998).
55. K. Schneider, B.T. Chait, 'Matrix-assisted Laser Desorption Mass Spectrometry of Homopolymer Oligodeoxynucleotides. Influence of Base Composition on the Mass Spectrometric Response', *Org. Mass Spectrom.*, **28**(11), 1353–1361 (1993).
56. L. Zhu, G.R. Parr, M.C. Fitzgerald, C.M. Nelson, L.M. Smith, 'Oligodeoxynucleotide Fragmentation in MALDI/TOF Mass Spectrometry using 355-nm Radiation', *J. Am. Chem. Soc.*, **117**(22), 6048–6065 (1995).
57. E. Nordhoff, F. Kirpekar, M. Karas, R. Cramer, S. Hahner, F. Hillenkamp, K. Kristiansen, P. Roepstorff, A. Lezius, 'Comparison of IR- and UV-matrix Assisted Laser Desorption/Ionization Mass Spectrometry of Oligodeoxynucleotides', *Nucleic Acids Res.*, **22**(13), 2460–2465 (1994).
58. W. Tang, L. Zhu, L.M. Smith, 'Controlling DNA Fragmentation in MALDI/MS by Chemical Modification', *Anal. Chem.*, **69**(3), 302–312 (1997).
59. W. Tang, J. Krause, L. Zhu, L.M. Smith, 'Factors Influencing Oligonucleotide Stability in Matrix-assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **169/170**, 301–312 (1997).

60. E. Nordhoff, M. Karas, R. Cramer, S. Hahner, F. Hillenkamp, F. Kirpekar, A. Lezius, J. Muth, C. Meier, J.W. Engels, 'Direct Mass Spectrometric Sequencing of Low-picomole Amounts of Oligodeoxynucleotides with up to 21 Bases by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *J. Mass Spectrom.*, **30**(1), 99–112 (1995).
61. K. Schneider, B.T. Chait, 'Increased Stability of Nucleic Acids Containing 7-Deaza-guanosine and 7-Deaza-adenosine may enable Rapid DNA Sequencing by Matrix-assisted Laser Desorption Mass Spectrometry', *Nucleic Acids Res.*, **23**(9), 1570–1575 (1995).
62. J. Gross, A. Leisner, F. Hillenkamp, S. Hahner, M. Karas, J. Schäfer, F. Lützenkirchen, E. Nordhoff, 'Investigations of the Metastable Decay of DNA under Ultraviolet Matrix-assisted Laser Desorption/Ionization Conditions with Post-source-decay Analysis and Hydrogen/Deuterium Exchange', *J. Am. Soc. Mass Spectrom.*, **9**(9), 866–878 (1998).
63. S.A. McLuckey, G.J. Van Berkel, G.L. Glish, 'Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides', *J. Am. Soc. Mass Spectrom.*, **3**(1), 60–70 (1992).
64. Z. Wang, K.X. Wan, R. Ramanathan, J.S. Taylor, M.L. Gross, 'Structure and Fragmentation Mechanisms of Isomeric T-rich Oligodeoxynucleotides: A Comparison of Four Tandem Mass Spectrometric Methods', *J. Am. Soc. Mass Spectrom.*, **9**(7), 683–691 (1998).
65. F. Kirpekar, E. Nordhoff, K. Kristiansen, P. Roepstorff, S. Hahner, F. Hillenkamp, '7-Deaza Purine Bases Offer a Higher Ion Stability in the Analysis of DNA by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(6), 525–531 (1995).
66. T. Ono, M. Scalf, L.M. Smith, '2′-Fluoro Modified Nucleic Acids: Polymerase-directed Synthesis, Properties and Stability to Analysis by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Nucleic Acids Res.*, **25**(22), 4581–4588 (1997).
67. I.G. Gut, S. Beck, 'A Procedure for Selective DNA Alkylation and Detection by Mass Spectrometry', *Nucleic Acids Res.*, **23**(8), 1367–1373 (1995).
68. K.J. Wu, A. Steding, C.H. Becker, 'Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry of Oligonucleotides using 3-Hydroxypicolinic Acid as an Ultraviolet-sensitive Matrix', *Rapid Commun. Mass Spectrom.*, **7**(1), 142–146 (1993).
69. K.J. Wu, T.A. Shaler, C.H. Becker, 'Time-of-flight Mass Spectrometry of Underivatized Single-stranded DNA Oligomers by Matrix-assisted Laser Desorption', *Anal. Chem.*, **66**(10), 1637–1645 (1994).
70. D.P. Little, T.J. Cornish, M.J. O'Donnell, A. Braun, R.J. Cotter, H. Köster, 'MALDI on a Chip: Analysis of Arrays of Low-femtomole to Subfemtomole Quantities of Synthetic Oligonucleotides and DNA Diagnostic Products Dispensed by a Piezoelectric Pipet', *Anal. Chem.*, **69**(22), 2438–2443 (1997).
71. S. Berkenkamp, F. Kirpekar, F. Hillenkamp, 'Infrared MALDI Mass Spectrometry of Large Nucleic Acids', *Science*, **281**, 260–262 (1998).
72. W.C. Wiley, I.H. McLaren, 'Time-of-flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, **26**, 1150–1157 (1955).
73. M.L. Vestal, P. Juhasz, S.A. Martin, 'Delayed Extraction Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(11), 1044–1050 (1995).
74. N.P. Christian, S.M. Colby, L. Giver, C.T. Houston, R.J. Arnold, A.D. Ellington, J.P. Reilley, 'High Resolution Matrix-assisted Laser Desorption/Ionization Time-of-flight Analysis of Single-stranded DNA of 27 to 68 Nucleotides in Length', *Rapid Commun. Mass Spectrom.*, **9**(11), 1061–1066 (1995).
75. P. Juhasz, M.T. Roskey, I.P. Smirnov, L.A. Haff, M.L. Vestal, S.A. Martin, 'Applications of Delayed Extraction Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry to Oligonucleotide Analysis', *Anal. Chem.*, **68**(6), 941–946 (1996).
76. M.T. Roskey, P. Juhasz, I.P. Smirnov, E.J. Takach, S.A. Martin, L.A. Haff, 'DNA Sequencing by Delayed Extraction-matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **93**, 4724–4729 (1996).
77. F. Kirpekar, E. Nordhoff, K. Kristiansen, P. Roepstorff, A. Lezius, S. Hahner, M. Karas, F. Hillenkamp, 'Matrix Assisted Laser Desorption/Ionization Mass Spectrometry of Enzymatically Synthesized RNA up to 150 kDa', *Nucleic Acids Res.*, **22**(19), 3866–3870 (1994).
78. J. Bai, Y.H. Liu, D.M. Lubman, D. Siemieniak, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Restriction Enzyme-digested Plasmid DNA using an Active Nafion Substrate', *Rapid Commun. Mass Spectrom.*, **8**(9), 687–691 (1994).
79. J. Bai, Y.H. Liu, X. Liang, Y. Zhu, D.M. Lubman, 'Procedures for Detection of DNA by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry using a Modified Nafion Film Substrate', *Rapid Commun. Mass Spectrom.*, **9**(12), 1172–1176 (1995).
80. Y.H. Liu, J. Bai, D.M. Lubman, P.J. Venta, 'Use of a Nitrocellulose Film Substrate in Matrix-assisted Laser Desorption/Ionization Mass Spectrometry for DNA Mapping and Screening', *Anal. Chem.*, **67**(19), 3482–3490 (1995).
81. Y.H. Liu, J. Bai, Y. Zhu, X. Liang, D. Siemieniak, P.J. Venta, D.M. Lubman, 'Rapid Screening of Genetic Polymorphisms using Buccal Cell DNA with Detection by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(9), 735–743 (1995).
82. K. Tang, N.I. Taranenko, S.L. Allmann, L.Y. Ch'ang, C.H. Chen, 'Detection of 500-Nucleotide DNA by Laser

Desorption Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**(9), 727–730 (1994).
83. P.L. Ross, P. Belgrader, 'Analysis of Short Tandem Repeat Polymorphisms in Human DNA by Matrix-assisted Laser Desorption Mass Spectrometry', *Anal. Chem.*, **69**(19), 3966–3972 (1997).
84. R.D. Smith, X. Cheng, J.E. Bruce, S.A. Hofstadler, G.A. Anderson, 'Trapping, Detection and Reaction of Very Large Single Molecular Ions by Mass Spectrometry', *Nature*, **369**, 137–139 (1994).
85. S.D. Fuerstenau, W.H. Benner, 'Molecular Weight Determination of Megadalton DNA Electrospray Ions using Charge Detection Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(15), 1528–1538 (1995).
86. J.E. Bruce, X. Cheng, R. Bakhtiar, Q. Wu, S.A. Hofstadler, G.A. Anderson, R.D. Smith, 'Trapping, Detection, and Mass Measurement of Individual Ions in a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer', *J. Am. Chem. Soc.*, **116**(17), 7839–7847 (1994).
87. R. Chen, X. Cheng, D.W. Mitchell, S.A. Hafstadler, Q. Wu, A.L. Rockwood, M.G. Sherman, R.D. Smith, 'Trapping, Detection, and Mass Determination of Coliphage T4 DNA Ions of 10^8 Da by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **67**(7), 1159–1163 (1995).
88. X. Cheng, D.G. Camp, II, Q. Wu, R. Bakhtiar, D.L. Springer, B.J. Morris, J.E. Bruce, G.A. Anderson, C.G. Edmonds, R.D. Smith, 'Molecular Weight Determination of Plasmid DNA using Electrospray Ionization Mass Spectrometry', *Nucleic Acids Res.*, **24**(11), 2183–2189 (1996).
89. W.H. Benner, 'A Gated Electrostatic Ion Trap to Repetitiously Measure the Charge and m/z of Large Electrospray Ions', *Anal. Chem.*, **69**(20), 4162–4168 (1997).
90. R.D. Voyksner, 'Combining Liquid Chromatography with Electrospray Mass Spectrometry', in *Electrospray Ionization Mass Spectrometry*, ed. R.B. Cole, John Wiley & Sons, 323–382, 1997.
91. K. Bleicher, E. Bayer, 'Analysis of Oligonucleotides using Coupled High Performance Liquid Chromatography–Electrospray Mass Spectrometry', *Chromatographia*, **39**(7/8), 405–408 (1994).
92. A. Apffel, J.A. Chakel, S. Fischer, K. Lichtenwalter, W.S. Hancock, 'Analysis of Oligonucleotides by HPLC–Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **69**(7), 1320–1325 (1997).
93. R.H. Griffey, M.J. Greig, H.J. Gaus, K. Liu, D. Monteith, M. Winniman, L.L. Cummins, 'Characterization of Oligonucleotide Metabolism In Vivo via Liquid Chromatography/Electrospray Tandem Mass Spectrometry with a Quadrupole Ion Trap Mass Spectrometer', *J. Mass Spectrom.*, **32**(3), 305–313 (1997).
94. L. Li, A.P.L. Wang, L.D. Coulson, 'Continuous-flow Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **65**(4), 493–495 (1993).
95. K.K. Murray, D.H. Russel, 'Liquid Sample Introduction for Matrix-assisted Laser Desorption/Ionization', *Anal. Chem.*, **65**(18), 2534–2537 (1993).
96. K.K. Murray, 'Coupling Matrix-assisted Laser Desorption/Ionization to Liquid Separations', *Mass Spectrom. Rev.*, **16**(5), 283–299 (1997).
97. P. Lecchi, L.K. Pannell, 'The Detection of Noncovalent Interactions in Nucleic Acids using MALDI', in *New Methods for the Study of Biomolecular Complexes*, eds. W. Ens, K.G. Standing, I.V. Chernushevich, NATO-ASI series C, Kluwer Academic Publishers, 217–224, Vol. 510, 1998.
98. F. Kirpekar, S. Berkenkamp, F. Hillenkamp, 'Detection of Double-stranded DNA by IR- and UV/MALDI Mass Spectrometry', *Anal. Chem.*, **71**(13), 2334–2339 (1999).
99. I. Gruic-Sovulj, H.-C. Lüdemann, F. Hillenkamp, I. Weygand-Durasevic, Z. Kucan, J. Peter-Katalinic, 'Detection of Noncovalent tRNA•aminoacyl-tRNA Synthetase Complexes by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *J. Biol. Chem.*, **272**(51), 32084–32091 (1997).
100. R.D. Smith, K.J. Light-Wahl, 'The Observation of Noncovalent Interactions in Solution by Electrospray Ionization Mass Spectrometry: Promise, Pitfalls and Prognosis', *Biol. Mass Spectrom.*, **22**, 493–501 (1993).
101. E. Bayer, T. Bauer, K. Schmeer, K. Bleicher, M. Maler, H.-J. Gaus, 'Analysis of Double-stranded Oligonucleotides by Electrospray Mass Spectrometry', *Anal. Chem.*, **66**(22), 3858–3863 (1994).
102. J. Ding, R.J. Anderegg, 'Specific and Nonspecific Dimer Formation in the Electrospray Ionization Mass Spectrometry of Oligonucleotides', *J. Am. Soc. Mass Spectrom.*, **6**(3), 159–164 (1995).
103. D.R. Goodlett, D.G. Camp, II, C.C. Hardin, M. Corregan, R.D. Smith, 'Direct Observation of a DNA Quadruplex by Electrospray Ionization Mass Spectrometry', *Biol. Mass Spectrom.*, **22**, 181–183 (1993).
104. B. Ganem, Y.-T. Li, J.D. Henion, 'Detection of Oligonucleotide Duplex Forms by Ion-spray Mass-spectrometry', *Tetrahedron Lett.*, **34**(9), 1445–1448 (1993).
105. K.J. Light-Wahl, D.L. Springer, B.E. Winger, C.G. Edmonds, D.G. Camp, II, B.D. Thrall, R.D. Smith, 'Observation of a Small Oligonucleotide Duplex by Electrospray Ionization Mass Spectrometry', *J. Am. Chem. Soc.*, **115**, 803–804 (1993).
106. P.D. Schnier, J.S. Klassen, E.F. Strittmatter, E.R. Williams, 'Activation Energies for Dissociation of Double Strand Oligonucleotide Anions: Evidence for Watson–Crick Base Pairing in Vacuo', *J. Am. Chem. Soc.*, **120**(37), 9605–9613 (1998).
107. Q. Wu, X. Cheng, S.A. Hofstadler, R.D. Smith, 'Specific Metal–Oligonucleotide Binding Studied by High

Resolution Tandem Mass Spectrometry', *J. Mass Spectrom.*, **31**(6), 669–675 (1996).
108. D.C. Gale, D.R. Goodlett, K.J. Light-Wahl, R.D. Smith, 'Observation of Duplex DNA–Drug Noncovalent Complexes by Electrospray Ionization Mass Spectrometry', *J. Am. Chem. Soc.*, **116**(13), 6027–6028 (1994).
109. A. Triolo, F.M. Acramone, A. Raffaelli, R. Salvadori, 'Noncovalent Complexes Between DNA-binding Drugs and Double-stranded Deoxyoligonucleotides: A Study by Ionspray Mass Spectrometry', *J. Mass Spectrom.*, **32**(9), 1186–1194 (1997).
110. P. Iannitti, M.M. Sheil, G. Wickham, 'High Sensitivity and Fragmentation Specificity in the Analysis of Drug–DNA Adducts by Electrospray Tandem Mass Spectrometry', *J. Am. Chem. Soc.*, **119**(6), 1490–1491 (1997).
111. D.C. Gale, R.D. Smith, 'Characterization of Noncovalent Complexes Formed between Minor Groove Binding Molecules and Duplex DNA by Electrospray Ionization-Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **6**(12), 1154–1164 (1995).
112. G. Lowe, J.A. McCloskey, J. Ni, T. Vilaivan, 'A Mass Spectrometric Investigation of the Reaction Between 4,4′-Vinylenedipyridine Bis[2,2′:6′,2′-terpyridine Platinum(II)] and the Self-complementary Oligonucleotide d(CpGpTpApCpG)', *Bioorg. Med. Chem.*, **4**(7), 1007–1013 (1996).
113. X. Cheng, A.C. Harms, R.N. Goudreau, T.C. Terwilliger, R.D. Smith, 'Direct Measurement of Oligonucleotide Binding Stoichiometry of Gene V Protein by Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **93**, 7022–7027 (1996).
114. X. Cheng, P.E. Morin, A.C. Harm, J.E. Bruce, Y. Ben-David, R.D. Smith, 'Mass Spectrometric Characterization of Sequence-specific Complexes of DNA and Transcription Factor PU.1 DNA Binding Domain', *Anal. Biochem.*, **239**, 35–40 (1996).
115. N. Potier, L.-J. Donald, I. Chernushevich, A. Ayed, W. Ens, C.H. Arrowsmith, K.G. Standing, H.W. Duckworth, 'Study of a Noncovalent Trp Repressor: DNA Operator Complex by Electrospray Ionization Time-of-flight Mass Spectrometry', *Protein Sci.*, **7**, 1388–1395 (1998).
116. M.J. Greig, H. Gaus, L.L. Cummins, H. Sasmor, R. Griffey, 'Measurement of Macromolecular Binding using Electrospray Mass Spectrometry. Determination of Dissociation Constants for Oligonucleotide: Serum Albumin Complexes', *J. Am. Chem. Soc.*, **117**(43), 10765–10766 (1995).
117. F. Hillenkamp, 'Matrix-assisted Laser Desorption/Ionization of Noncovalent Complexes', in *New Methods for the Study of Biomolecular Complexes*, eds. W. Ens, K.G. Standing, I.V. Chernushevich, NATO-ASI Series C, Kluwer Academic Publishers, 181–191, Vol. 510, 1998.
118. T.B. Farmer, R.M. Caprioli, 'Determination of Protein – Protein Interactions by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *J. Mass Spectrom.*, **33**(8), 697–704 (1998).
119. J.A. Loo, 'Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **16**(1), 1–23 (1997).
120. S.C. Pomerantz, J.A. Kowalak, J.A. McCloskey, 'Determination of Oligonucleotide Composition from Mass Spectrometrically Measured Molecular Weight', *J. Am. Soc. Mass Spectrom.*, **4**(3), 204–209 (1993).
121. D.J. Aaserud, N.L. Kelleher, D.P. Little, F.W. McLafferty, 'Accurate Base Composition of Double-strand DNA by Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **7**(12), 1266–1269 (1996).
122. R.S. Brown, J.J. Lennon, 'Sequence-specific Fragmentation of Matrix-assisted Laser-desorbed Protein/Peptide Ions', *Anal. Chem.*, **67**(21), 3990–3999 (1995).
123. B. Spengler, D. Kirsch, R. Kaufmann, 'Metastable Decay of Peptides and Proteins in Matrix-assisted Laser-desorption Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **5**(4), 198–202 (1991).
124. Y.F. Zhu, N.I. Taranenko, S.L. Allman, N.V. Taranenko, S.A. Martin, L.A. Haff, C.H. Chen, 'Oligonucleotide Sequencing by Fragmentation in Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**(7), 897–903 (1997).
125. B.H. Wang, C.E. Hopkins, A.B. Belenky, A.S. Cohen, 'Sequencing of Modified Oligonucleotides using In-source Fragmentation and Delayed Pulsed Ion Extraction Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **169/170**, 331–350 (1997).
126. B. Spengler, 'Post-source Decay Analysis in Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biomolecules', *J. Mass Spectrom.*, **32**(9), 1019–1063 (1997).
127. D. Kirsch, B. Spengler, R. Kaufmann, 'Ion Stability of Deoxyoligonucleotides and its Implication for PSD/MALDI MS', Proceedings of the 42nd Annual Conference on Mass Spectrometry and Allied Topics, ASMS, Chicago, Illinois, 821, 1994.
128. G. Talbo, M. Mann, 'Aspects of the Sequencing of Carbohydrates and Oligonucleotides by Matrix-assisted Laser Desorption/Ionization Post-source Decay', *Rapid Commun. Mass Spectrom.*, **10**(1), 100–101 (1996).
129. J.K. Gooden, J. Byun, L. Chen, E.G. Rogan, E.L. Cavalieri, M.L. Gross, 'Application of MALDI and PSD to the Structure Determination of Adducts between DNA Bases and the Carcinogen 7H-Dibenzo[*c*,*g*]carbazole', *Int. J. Mass Spectrom. Ion Processes*, **169/170**, 241–249 (1997).

130. R. Hettich, M. Buchanan, 'Structural Characterization of Normal and Modified Oligonucleotides by Matrix-assisted Laser Desorption Fourier Transform Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **2**(5), 402–412 (1991).
131. R.L. Hettich, M.V. Buchanan, 'Matrix-assisted Laser Desorption Fourier Transform Mass Spectrometry for the Structural Examination of Modified Nucleic Acid Constituents', *Int. J. Mass Spectrom. Ion Processes*, **111**, 365–380 (1991).
132. E.A. Stemmler, R.L. Hettich, G.B. Hurst, M.V. Buchanan, 'Matrix-assisted Laser Desorption/Ionization Fourier Transform Mass Spectrometry of Oligodeoxyribonucleotides', *Rapid Commun. Mass Spectrom.*, **7**(9), 828–836 (1993).
133. E.A. Stemmler, M.V. Buchanan, G.B. Hurst, R.L. Hettich, 'Analysis of Modified Oligonucleotides by Matrix-assisted Laser Desorption/Ionization Fourier Transform Mass Spectrometry', *Anal. Chem.*, **67**(17), 2924–2930 (1995).
134. J.A. Loo, H.R. Udseth, R.D. Smith, 'Collisional Effects on the Charge Distribution of Ions from Large Molecules, Formed by Electrospray-ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **2**(10), 207–210 (1988).
135. T.J. Carlin, B.S. Freiser, 'Pulsed Valve Addition of Collision and Reagent Gases in Fourier Transform Mass Spectrometry', *Anal. Chem.*, **55**(3), 571–574 (1983).
136. D. Little, P. Speir, M.W. Senko, P.B. O'Connor, F.W. McLafferty, 'Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing', *Anal. Chem.*, **66**(18), 2809–2815 (1994).
137. S.A. McLuckey, S. Habibi-Goudarzi, 'Decompositions of Multiply Charged Oligonucleotide Anions', *J. Am. Chem. Soc.*, **115**(25), 12 085–12 095 (1993).
138. S.A. McLuckey, S. Habibi-Goudarzi, 'Ion Trap Tandem Mass Spectrometry Applied to Small Multiply Charged Oligonucleotides with a Modified Base', *J. Am. Soc. Mass Spectrom.*, **5**(8), 740–747 (1994).
139. S.A. McLuckey, V. Gopalakrishnan, S. Habibi-Goudarzi, 'Charged vs. Neutral Nucleobase Loss from Multiply Charged Oligonucleotide Anions', *J. Mass Spectrom.*, **30**(9), 1222–1229 (1995).
140. D.P. Little, A. Russel, J. Chorush, P. Speir, M.W. Senko, N.L. Kelleher, F.W. McLafferty, 'Rapid Sequencing of Oligonucleotides by High-resolution Mass Spectrometry', *J. Am. Chem. Soc.*, **116**(11), 4893–4897 (1994).
141. D.P. Little, F.W. McLafferty, 'Sequencing 50-mer DNAs using Electrospray Tandem Mass Spectrometry and Complementary Fragmentation Methods', *J. Am. Chem. Soc.*, **117**(25), 6783–6784 (1995).
142. D.P. Little, D.J. Aaserud, G.A. Valaskovic, F.W. McLafferty, 'Sequence Information from 42-108-mer DNAs (Complete for a 50-mer) by Tandem Mass Spectrometry', *J. Am. Chem. Soc.*, **118**(39), 9352–9359 (1996).
143. J.P. Barry, P. Vouros, A. Van Schepdael, S.-J. Law, 'Mass and Sequence Verification of Modified Oligonucleotides using Electrospray Tandem Mass Spectrometry', *J. Mass Spectrom.*, **30**(7), 993–1006 (1995).
144. P.F. Crain, J.M. Gregson, J.A. McCloskey, C.C. Nelson, J.M. Peltier, D.R. Phillips, S.C. Pomerantz, D.M. Reddy, 'Characterization of Posttranscriptional Modification in Nucleic Acids by Tandem Mass Spectrometry', in *Mass Spectrometry in the Biological Sciences*, eds. A.L. Burlingame, S.A. Carr, Humana Press, Totawa, NJ, 497–517, 1996.
145. M.G. Bartlett, J.A. McCloskey, S. Manalili, R.H. Griffey, 'The Effect of Backbone Charge on the Collision-induced Dissociation of Oligonucleotides', *J. Mass Spectrom.*, **31**(11), 1277–1283 (1996).
146. T. Keough, T.R. Baker, R.L.M. Dobson, M.P. Lacey, T.A. Riley, J.A. Hasselfield, P.E. Hesselberth, 'Antisense DNA Oligonucleotides II: The Use of Matrix-assisted Laser Desorption/Ionization Mass Spectrometry for the Sequence Verification of Methylphosphonate Oligodeoxyribonucleotides', *Rapid Commun. Mass Spectrom.*, **7**(1), 195–200 (1993).
147. T. Keough, J.D. Shaffer, M.P. Lacey, T.A. Riley, W.B. Marvin, M.A. Scurria, J.A. Hasselfield, E.P. Hellelberth, 'Detailed Characterization of Antisense DNA Oligonucleotides', *Anal. Chem.*, **68**(19), 3405–3412 (1996).
148. L.M. Polo, T.D. McCarley, P.A. Limbach, 'Chemical Sequencing of Phosphorothioate Oligonucleotides using Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **69**(6), 1107–1112 (1997).
149. K.L. Nakamaye, G. Gish, F. Eckstein, H.P. Vosberg, 'Direct Sequencing of Polymerase Chain Reaction Amplified DNA Fragments through the Incorporation of Deoxynucleoside α-Thiophosphates', *Nucleic Acids Res.*, **16**(21), 9947–9959 (1988).
150. S. Hahner, H.-C. Lüdemann, F. Kirpekar, E. Nordhoff, P. Roepstorff, H.-J. Galla, F. Hillenkamp, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry (MALDI) of Endonuclease Digests of RNA', *Nucleic Acids Res.*, **25**(19), 1957–1964 (1997).
151. D.A. Tolson, N.H. Nicholson, 'Sequencing RNA by a Combination of Exonuclease Digestion and Uridine Specific Chemical Cleavage using MALDI/TOF', *Nucleic Acids Res.*, **26**(2), 446–451 (1998).
152. J.A. Kowalak, E. Bruenger, J.A. McCloskey, 'Posttranscriptional Modification of the Central Loop of Domain V in *Escherichia coli* 23S Ribosomal RNA', *J. Biol. Chem.*, **270**, 17 758–17 764 (1995).
153. B. Felden, K. Hanawa, J.F. Atkins, H. Himeno, A. Muto, R.F. Gesteland, J.A. McCloskey, P.F. Crain, 'Presence and Location of Modified Nucleotides in *Escherichia coli* tmRNA: Structural Mimicry with rRNA Acceptor Branches', *EMBO J.*, **17**(11), 3188–3196 (1998).
154. K.R. Noon, E. Bruenger, J.A. McCloskey, 'Posttranscriptional Modifications in 16S and 23S rRNAs of the

Archaeal Hyperthermophile *Sulfolobus solfataricus*', *J. Bacteriol.*, **180**(11), 2883–2888 (1998).
155. J. Rozenski, J.A. McCloskey, 'Determination of Nearest Neighbors in Nucleic Acids by Mass Spectrometry', *Anal. Chem.*, **71**(7), 1454–1459 (1999).
156. P. Janning, W. Schrader, M. Linscheid, 'A New Mass Spectrometric Approach to Detect Modifications in DNA', *Rapid Commun. Mass Spectrom.*, **8**(12), 1035–1040 (1994).
157. M. Linscheid, C. Siethoff, B. Pongratz, M. Janik, 'In-source-CID Electrospray MS of Modified Oligonucleotides Combined with CZE', Proceedings of the 44th Conference of Mass Spectrometry and Allied Topics, ASMS, Portland, Oregon, 311, 1996.
158. R.W. Holley, J.T. Madison, A. Zamir, 'A New Method for Sequence Determination of Large Oligonucleotides', *Biochem. Biophys. Res. Commun.*, **17**(4), 389–394 (1964).
159. I.P. Smirnov, M.T. Roskey, P. Juhasz, E.J. Takach, S.A. Martin, L.A. Haff, 'Sequencing Oligonucleotides by Exonuclease Digestion and Delayed Extraction Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry', *Anal. Biochem.*, **238**, 19–25 (1996).
160. P.A. Limbach, P.F. Crain, J.A. McCloskey, 'Enzymatic Sequencing of Oligonucleotides with Electrospray Mass Spectrometry', *Nucleic Acids Res. Symp. Ser.*, **31**, 127–128 (1994).
161. M.C. Fitzgerald, L. Zhu, L.M. Smith, 'The Analysis of Mock DNA Sequencing Reactions using Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**(11), 895–897 (1993).
162. T.A. Shaler, Y. Tan, J.N. Wikham, K.J. Wu, C.H. Becker, 'Analysis of Enzymatic DNA Sequencing Reactions by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(10), 942–947 (1995).
163. S. Mouradian, D.R. Rank, L.M. Smith, 'Analyzing Sequencing Reactions from Bacteriophage M13 by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**(12), 1475–1478 (1996).
164. K. Tang, D.-J. Fu, S. Kötter, R.J. Cotter, H. Köster, 'Matrix Assisted Laser Desorption/Ionization Mass Spectrometry of Immobilized Duplex DNA Probes', *Nucleic Acids Res.*, **23**(16), 3126–3131 (1995).
165. H. Köster, K. Tang, D.-J. Fu, A. Braun, D. van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, 'A Strategy for Rapid and Efficient DNA Sequencing by Mass Spectrometry', *Nat. Biotechnol.*, **14**, 1123–1128 (1996).
166. D.-J. Fu, K. Tang, A. Braun, D. Reuter, B. Darnhofer-Demar, D.P. Little, M.J. O'Donnell, C.R. Cantor, H. Köster, 'Sequencing Exons 5 to 8 of the p53 Gene by MALDI/TOF Mass Spectrometry', *Nat. Biotechnol.*, **16**, 381–384 (1998).
167. N.I. Taranenko, C.N. Chung, Y.E. Zhu, S.L. Allmann, V.V. Golovlev, N.R. Isola, S.A. Martin, L.A. Haff, C.H. Chen, 'Matrix-assisted Laser Desorption/Ionization for Sequencing Single-stranded and Double-stranded DNA', *Rapid Commun. Mass Spectrom.*, **11**(4), 386–392 (1997).
168. N.I. Taranenko, S.L. Allmann, V.V. Golovlev, N.V. Taranenko, N.R. Isola, C.H. Chen, 'Sequencing DNA using Mass Spectrometry for Ladder Detection', *Nucleic Acids Res.*, **26**(10), 2488–2490 (1998).
169. J.R. Srinivasan, Y.-H. Liu, P.J. Venta, D. Siemieniak, A. Killeen, Y. Zhu, D.M. Lubman, 'Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry as a Rapid Screening Method to Detect Mutations Causing Tay–Sachs Disease', *Rapid Commun. Mass Spectrom.*, **11**(10), 1144–1150 (1997).
170. L.Y. Ch'ang, K. Tang, M. Schell, C. Ringelberg, K.J. Matteson, S.L. Allman, C.H. Chen, 'Detection of ΔFD508 Mutation of the Cystic Fibrosis Gene by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(9), 772–774 (1995).
171. N.I. Taranenko, K.J. Matteson, C.N. Chung, Y.F. Zhu, L.Y. Chang, S.L. Allman, L. Haff, S.A. Martin, C.H. Chen, 'Laser Desorption Mass Spectrometry for Point Mutation Detection', *Genet. Anal.*, **13**, 87–94 (1996).
172. C. Jurinke, B. Zöllner, H.-H. Feucht, A. Jacob, J. Kirchhübel, A. Lüchow, D. van den Boom, R. Laufs, H. Köster, 'Detection of Hepatitis B Virus DNA in Serum Samples via Nested PCR and MALDI/FOF Mass Spectrometry', *Genet. Anal.*, **13**, 67–71 (1996).
173. C. Jurinke, B. Zollner, H.H. Feucht, D. van den Boom, A. Jacob, S. Polywka, R. Laufs, H. Köster, 'Application of Nested PCR and Mass Spectrometry for DNA-based Virus Detection: HBV-DNA Detected in the Majority of Isolated Anti-BHc Positive Sera', *Genet. Anal.*, **14**(3), 97–102 (1998).
174. A. Braun, D.P. Little, H. Köster, 'Detecting CFTR Gene Mutations by using Primer Oligo Base Extension and Mass Spectrometry', *Clin. Chem.*, **43**(7), 1151–1158 (1997).
175. D.P. Little, A. Braun, B. Darnhofer-Demar, H. Köster, 'Identification of Apolipoprotein E Polymorphisms using Temperature Cycled Primer Oligo Base Extension and Mass Spectrometry', *Eur. J. Chem. Clin. Biochem.*, **35**(7), 545–548 (1997).
176. A. Braun, D.P. Little, D. Reuter, B. Müller-Mysok, H. Köster, 'Improved Analysis of Microsatellites using Mass Spectrometry', *Genomics*, **46**, 18–23 (1997).
177. D.P. Little, A. Braun, M.J. O'Donnell, H. Köster, 'Mass Spectrometry from Miniaturized Arrays for Full Comparative DNA Analysis', *Nat. Med.*, **3**(12), 1413–1416 (1997).
178. L.A. Haff, I.P. Smirnov, 'Single-Nucleotide Polymorphism Identification Assays using a Thermostable DNA

Polymerase and Delayed Extraction MALDI/TOF Mass Spectrometry', *Genome Res.*, **7**, 378–388 (1997).

179. D.S. Wunschel, D.C. Muddiman, K.F. Fox, A. Fox, R.D. Smith, 'Heterogeneity in *Bacillus cereus* PCR Products Detected by ESI/FTICR Mass Spectrometry', *Anal. Chem.*, **70**(6), 1203–1207 (1998).

180. S.J. Laken, P.E. Jackson, K.W. Kinzler, B. Volgelstein, P.T. Strickland, J.D. Groopman, M.D. Friesen, 'Genotyping by Mass Spectrometric Analysis of Short DNA Fragments', *Nat. Biotechnol.*, **16**, 1352–1356 (1998).

Nuclear Magnetic Resonance and Nucleic Acid Structures

Magdalena Eriksson
Chalmers University of Technology, Gothenburg, Sweden

1 Introduction **1**
2 Nuclear Magnetic Resonance and Nucleic Acids **1**
2.1 Spectral Properties 1
2.2 Techniques Used 2
2.3 Critical Comments 7
3 Structures Determined **8**
3.1 DNA Motifs 8
3.2 RNA 13
3.3 PNA and Other Variations 16
4 Future Outlook **16**
4.1 Transverse Relaxation 16
4.2 Oriented Samples 17
Acknowledgments **17**
Abbreviations and Acronyms **17**
Related Articles **17**
References **18**

Nuclear magnetic resonance (NMR) is one of the most important and widely used techniques for studying nucleic acid structure. With modern techniques, including isotope labeling and multidimensional spectroscopy, structures of large molecules – nucleic acids alone or in complexes, for example with proteins or ligands – can be determined. Here is given an overview of experimental NMR methods for nucleic acids studies, complemented by illustrative examples of DNA, RNA and PNA structures.

1 INTRODUCTION

NMR is one of the most important and widely used techniques for studying nucleic acid structure. NMR offers the advantage of permitting high-resolution studies of molecules in solution, and thus of manipulating conditions such as concentration of various ions and ligands, temperature and pH. Although the number of NMR spectra of nucleic acids acquired in solid phase is growing rapidly, the vast majority of experiments are still carried out in aqueous solution. This review will therefore focus on nucleic acid structure in solution as studied by NMR methods. This subject has previously been carefully reviewed, for example by Van de Ven and Hilbers[1] and by Patel et al.[2]

There has been a hand-in-hand development of increasingly more affordable methods of preparing isotope-labeled oligonucleotides, both DNA and RNA, and new pulse sequences taking advantage of their incorporation. Many of these pulse sequences are tailored to detect single correlation pathways that are instrumental in making correct assignments of a spectrum. To cover an inclusive menu of experiments now available would be formidable, task, such a list could not be condensed into the format of this article, and even if it could, would quickly become outdated. The aim here is instead to give a general introduction to available techniques complemented by illustrative examples of a few representative structures. These structures should give an idea of the range of problems now tractable by NMR and indicate strategies for solving them.

2 NUCLEAR MAGNETIC RESONANCE AND NUCLEIC ACIDS

Nucleic acids are structurally versatile molecules (for detailed descriptions of nucleic acids nomenclature and parameterization, please refer to Saenger[3] and Lu and Olson[4]). With DNA this is manifested as a multitude of tertiary structure motifs seen primarily as a variety of multistranded complexes. RNA is also highly variable at a secondary structure level, with alternate basepairing interactions and intricate folding topologies. For structural information about nucleic acids to be meaningful, a high degree of accuracy is therefore required and structural information derived from NMR experiments will often not be conclusive unless the spectra of a molecule can be unambiguously interpreted. Techniques have been developed that have considerably widened the scope of NMR studies. In particular, improved and more affordable isotope-labeling methods combined with higher field instrumentation and multidimensional pulse sequences have permitted structure studies of nucleic acids as long as 40 nucleotides, even in complexes with proteins, possible by NMR.[5]

2.1 Spectral Properties

2.1.1 DNA

The proton NMR spectrum of DNA (Figure 1) is usually nicely distributed between 0 and 15 ppm. For molecules with a relatively low degree of complexity, such as a duplex, it is often possible to carry out assignments

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

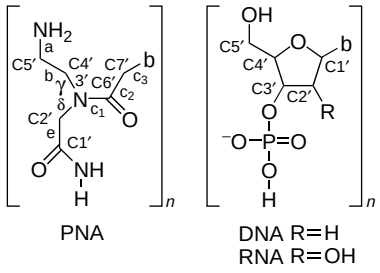

Figure 1 Schematic structure of PNA, DNA and RNA. b = base.

unambiguously for the majority of the protons in an unlabeled sample based solely on two-dimensional (2D) experiments. Typical proton spectra of a DNA duplex in D_2O and H_2O are shown in Figure 2. Protons that do not exchange at a significant rate with the solvent generally appear in the region between 0.5 and 8.5 ppm (Figure 2a). Labile protons, such as the amino and imino protons are usually observed between 6–10 ppm and 11–15 ppm, respectively (Figure 2b).

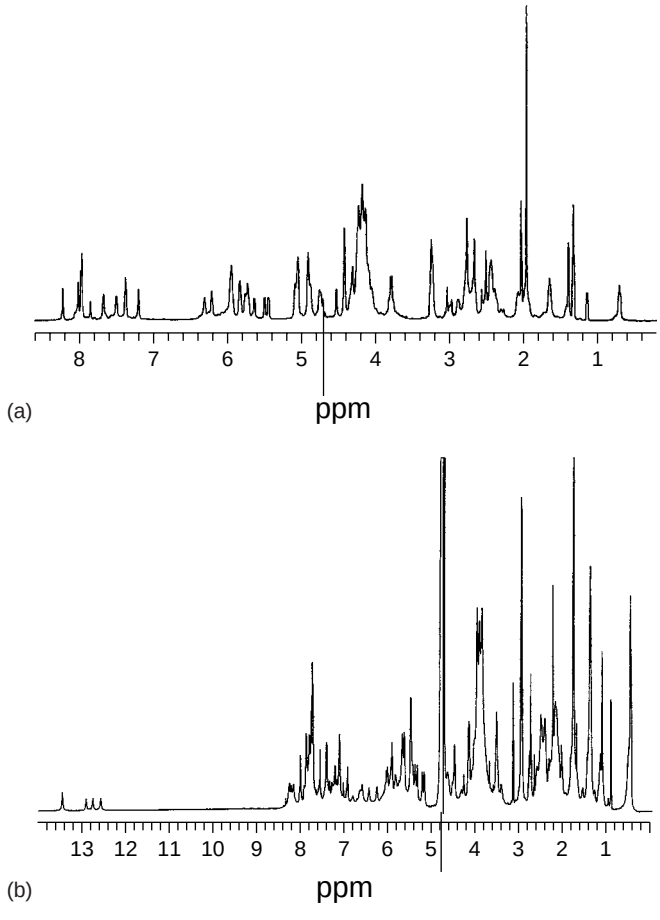

Figure 2 One-dimensional proton spectrum of the self-complementary DNA duplex $[d(CGCGATCGCG)]_2$ in D_2O solution at 38 °C (a), and in H_2O solution at 10 °C (b).

Assignment of the nonexchangeable protons – the aromatic H6/H8, H2, H5 and the methyl and sugar protons of a regular B-form DNA duplex – can usually be done from NOESY (nuclear Overhauser spectroscopy) and COSY (correlation spectroscopy) data recorded in D_2O, while the exchangeable protons can be assigned from H_2O NOESY spectra. Sometimes data acquired at a few different temperatures are helpful in overcoming assignment problems caused by spectral overlap.

2.1.2 RNA

In ribonucleotides the 2′-hydroxy group (Figure 1) causes the H2′ resonances to shift downfield so that they overlap with the H3′ signals. The H5 resonances of the uracil bases appear in the 5.0–6.0 ppm region along with their cytosine analogs. Because of overlap in the sugar proton region, complete assignment of an RNA oligonucleotide may be difficult without the use of isotope labeling and a number of multidimensional experiments. Strategies for this are described in more detail in section 2.2 below.

2.1.3 PNA

The nucleobases in PNA (peptide nucleic acid, Figure 1) are the same as in DNA and RNA, and their spectral features are thus similar. The chemical shifts of the aromatic protons tend to appear slightly upfield of the corresponding DNA resonances (Figure 3).[6] Each residue of the PNA backbone (*N*-(2-aminoethyl)-glycine) contains four pairs of nonexchangeable methylene protons and one exchangeable amide proton. Owing to the different environments of the methylene pairs in PNA, assignments of these can generally be carried out, but stereospecific discrimination cannot be performed unambiguously without specific labeling. Basepairing of a PNA strand to DNA of a complementary sequence results in the formation of a stable hybrid complex that exhibits an NMR spectrum that differs both from the PNA and DNA single strand spectra. A representative spectrum of a PNA strand hybridized to DNA is shown in Figure 3.

2.2 Techniques Used

Nucleic acids of biological significance are often, even as model systems, large enough to pose technical challenges when studied by NMR. Whereas small molecules of DNA are usually tractable solely using their proton resonances, NMR studies of RNA and larger or more complex DNA molecules would be prohibitively difficult without the use of additional paramagnetic nuclei, typically ^{13}C and ^{15}N, as labels. Important progress has been made in techniques for producing isotope-labeled DNA and RNA molecules. In addition, partial deuteration of large molecules can be

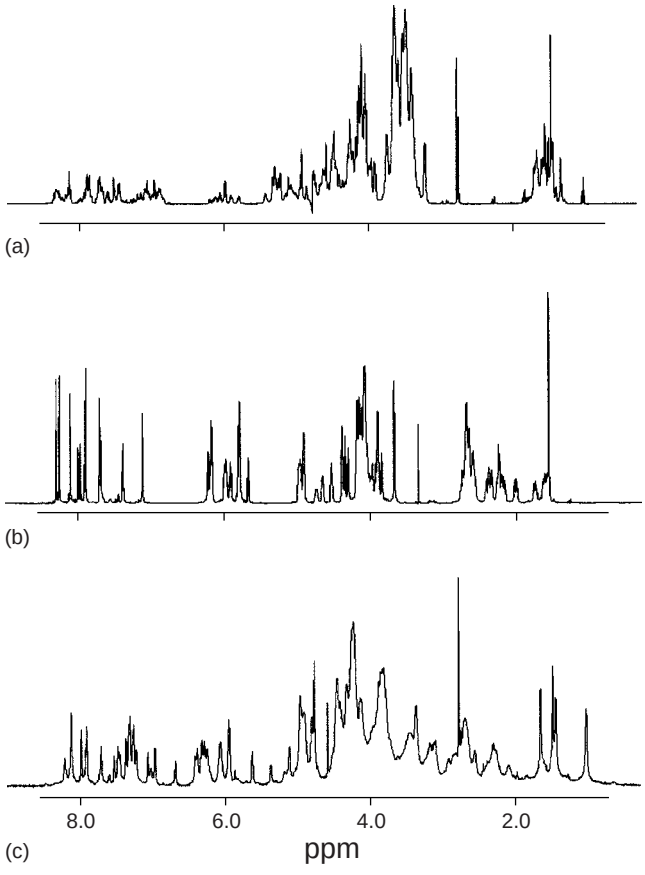

Figure 3 One-dimensional proton spectra of PNA strand H-GCTATGTC-NH_2 (a), DNA strand d(GACATAGC) (b) and the 1:1 PNA:DNA duplex H-GCTATGTC-NH_2·d(GACATAGC) (c) in D_2O solution at 24 °C.

used to alleviate a number of problems associated with large molecules in NMR experiments.

2.2.1 Isotope Labeling and Sample Preparation

2.2.1.1 RNA Labeling of RNA oligonucleotides with the NMR active isotopes ^{13}C and ^{15}N is efficiently done by enzymatic synthesis in vitro. Isotope-enriched nucleotide 5′-triphosphates (NTPs) can be prepared by growing bacterial cells in glucose–mineral salt containing the chosen isotope labels – typically *E. coli* cells are grown in ^{13}C glucose and ^{15}N ammonium sulfate.[7] Alternatively, ^{13}C methanol may serve as a less costly source of ^{13}C.[8] After isolation, the nucleic acid chains are degraded to mononucleotides by nuclease treatment. The RNA monomers are separated from DNA and enzymatically phosphorylated to NTPs. The labeled NTPs can then be used as substrate for in vitro transcription using a synthetic DNA template, typically using T7 polymerase, of RNA molecules of the desired length and sequence. Unincorporated nucleotides can be recovered and reused. The procedure can yield milligram quantities of labeled RNA for NMR experiments. Methods describing base selective isotope enrichment of RNA by residue type and the incorporation of these into large RNA fragments have also been reported.[9]

2.2.1.2 DNA A number of strategies have been demonstrated to produce high yields of labeled deoxyribonucleotides at low cost.[10] Today there are several strategies that provide uniformly ^{13}C and ^{15}N labeled DNA oligonucleotides by enzymatic synthesis.[11–13] Deoxynucleotides can be obtained from bacteria grown on ^{13}C and ^{15}N enriched nutrients, isolated and enzymatically phosphorylated to dNTPs. These can be used as building blocks in a polymerase chain reaction (PCR) using a DNA template and a DNA primer carrying a 3′-terminal ribonucleotide. The single 2′OH acts as a substrate for alkaline hydrolysis, which is used to cleave the product strand to be separated from the primer.

Labeled DNA oligomers can be made by cloning multiple copies of the desired oligonucleotide flanked by restriction sites into a plasmid which is transformed into bacteria that are grown with $^{15}NH_4Cl$ as nitrogen source. The use of ^{13}C for this purpose is at present excessively costly. Cleavage of the isolated plasmid DNA gives labeled oligonucleotides, with starting material for labeled dNTPs as a side product. Alternatively, endonuclease sensitive repeat amplification, or ESRA, can be performed by running many cycles of PCR to produce very large amounts of DNA fragments that will be ^{13}C and ^{15}N labeled if isotope enriched dNTPs are used. Procedures for site specific[14] and region specific incorporation of labeled DNA oligonucleotides[15] have been described as well. These particularly enable highlighting of portions of large DNA molecules or protein–DNA complexes that are of special interest.

2.2.1.3 Partial Deuteration Selective deuteration of oligonucleotides may be introduced in order to suppress part or parts of the proton spectrum. Conditions for efficient deuteration have been worked out for the exchange of pyrimidine H5/H6 and purine H8 protons, and applied to simplify the assignment of homologous DNA repeat sequences.[16] Synthesis protocols for incorporating deuterons stereospecifically in the 2′, 3′, 4′, 5′ and 5″ positions in RNA and DNA have been reported.[17,18] Aside from reduced spectral crowding, the advantages of partial deuteration of oligonucleotides include improved resolution and sensitivity. These effects are mainly due to the fact that both the T_1 and T_2 relaxation times of nuclei vicinal to the deuterons increase considerably.[18,19]

It is desirable that NMR samples of macromolecules (such as nucleic acids) are pure. In the case of nucleic

acids it is therefore necessary to include a purification step in the sample preparation in which the desired molecules are separated from fragments of similar size. The most common method for such purification is HPLC (high-performance liquid chromatography) on a reverse-phase column. This is typically followed by exchange of solvent and counter ions on an ion-exchange column or by dialysis.

2.2.2 *Experiments*

For structure studies of nucleic acids by NMR, the experiments can be divided into three stages. First, one-dimensional spectra are recorded. These are useful for indicating sample purity (both with respect to chemical and conformational composition), sensitivity to pH and temperature, and characteristics such as spectral distribution and the general quality of the spectrum and feasibility of further studies. In addition, the presence of stable basepairs or hydrogen bonds can be seen from one-dimensional spectra. Furthermore, titration with proteins or other ligands is monitored as changes in a one-dimensional spectrum. Second, various multidimensional experiments are recorded in order to establish assignments of as many as possible of the NMR active nuclei in the molecule. Stage three consists of recording quantitative measurements of structural parameters such as internuclear distances (NOE, nuclear Overhauser effect volumes) or dihedral angles (coupling constants) that can be used for structure determination.

In studying nucleic acids in aqueous solution by NMR one has to distinguish between spectra recorded in H_2O and D_2O solution. In H_2O solution all protons are visible (including solvent water), whereas in deuterated solvents only the protons that are covalently attached to the molecule can be detected beside the residual water peak. Because in nucleic acids the labile amino and imino protons generally exchange rapidly and appear mostly downfield of the stable ones, the spectral width of a D_2O experiment can often be kept relatively small, allowing for higher resolution without increasing the amount of data recorded. Disregarding the labile proton resonances, which often require low temperatures (below 10 °C) for stabilization and detection, permits spectra in D_2O solution to be recorded at the higher temperatures at which the peak widths tend to be smaller.

There are two particular types of effect that can be measured quantitatively from NMR data – coupling constants and interatom distances. Different varieties of the 2D COSY and NOESY experiments are primarily used for these purposes, respectively. In a COSY spectrum, crosspeaks that show splitting patterns contain information about dihedral angles between protons separated by no more than three bonds. The relation between the coupling constant J, measured as the splitting observed within one crosspeak in a phase-sensitive COSY spectrum, typically measured in Hertz, can be expressed by the semiempirical Karplus equation, Equation (1)

$$^3J = A\cos^2(\phi) + B\cos(\phi) + C \qquad (1)$$

The constants A, B, and C can be determined from measurements of ^3J in molecules with known values of the angle ϕ. Fitting of measured angular data, for example between the protons in a ribose sugar, has been worked out extensively (see Wüthrich[20] and Rinkel and Altona[21]). In a NOESY spectrum of a molecule the size of an oligonucleotide, the crosspeak buildup rate, μ_{ij}, measured as its integrated volume versus mixing time, is related to the interatom distance of the resonances involved, r_{ij}, as, Equation (2)

$$r_{ij} = c\mu_{ij}^{1/6} \qquad (2)$$

where c is a constant that can be determined from a crosspeak representing two covalently connected protons of known separation (e.g. H5–H6 of cytosine or uracil of 2.45 Å, or the intrasugar H2′–H2″ distance of 1.85 Å).

Small molecules with a regular conformation, such as double-stranded DNA oligonucleotides or short RNA stems, can often be successfully approached by homonuclear 2D NOESY and COSY experiments, relying on the regular helical conformation for which sequential assignment strategies are relatively straightforward.[2,20]

Sequential assignment of duplex DNA in a B-like conformation[3] is most easily performed in the aromatic–H1′ region of the NOESY spectrum which is characterized by intraresidue and sequential crosspeaks of similar intensity. Knowledge of the nucleotide sequence and the expected approximate chemical shifts of the different bases usually suffices to find a trace along the stack of residues. Good markers are the strong intraresidue crosspeaks arising from H5–H6 interactions in cytosines, seen in COSY as well as NOESY spectra. Strong intraresidue NOESY crosspeaks are also seen from the thymine H6–methyl interactions. These markers are helpful in identifying the H6 of DNA pyrimidines and their 5′ flanking residues. Alternating inter- and intraresidue crosspeaks between H6/H8 and H1′ protons (Figure 4) can typically be connected to a trace in the aromatic–H1′ region of a NOESY spectrum. A similar sequential trace can also be followed between the aromatic–H2′ and H2″ protons crosspeaks where the intraresidue H6/H8–H2″ interactions tend to be stronger and resonate slightly downfield compared with the corresponding H2′ protons. Furthermore, the H3′ and H4′ resonances may sometimes be assigned in

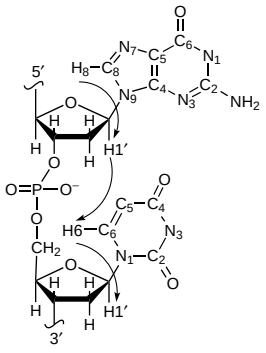

Figure 4 Schematic illustration of intra- and interresidue dipolar NOE interactions observable in a B-form DNA strand.

the same manner, although weaker sequential interactions and more extensive overlaps often make this more difficult, and additional spectral regions usually need to be analyzed for conclusive assignment of the H3′ and H4′ protons. Once tentative assignments of the H1′, H2′ and H2″ resonances have been made, these can be crosschecked by identifying the intraresidual cross-peaks in the H1′–H2′ and H1′–H2″ regions of NOESY and COSY spectra. The H2 protons of adenine bases can be distinguished from other aromatic resonances by their characteristically long spin–lattice relaxation times, which can be detected through inversion recovery experiments.[22]

As for the exchangeable protons, the amino resonances show strong NOEs to the H5 and H6 resonances on the same residue. The thymine imino resonances are most easily identified from their strong interactions with the H2 protons of their respective basepaired adenines. The guanine imino protons show strong NOE interactions with the amino protons of their respective basepaired cytosines. The imino resonances often exhibit sequential imino–imino NOE connectivities among themselves that can be readily traced. However, the imino protons of the terminal basepairs usually exchange too rapidly to be detected under normal experimental conditions.

The assignment of larger molecules that give rise to overcrowded spectra and of molecules of unknown secondary structure, for which sequential assignment strategies break down, requires combination of isotope-labeled samples and various 2D and 3D (three-dimensional) experiments. This need is particularly pronounced for RNA, where overlap in the sugar region is severe and the H1′ and H2′ coupling is very weak (<3 Hz) in the C3′-*endo* conformation which generally predominates in RNA sugars. Whereas in a 2D experiment, interactions involving two specific nuclei can be resolved, in a 3D spectrum the involvement of a third nucleus in the magnetization transfer pathway can be visualized. This not only adds resolution along the third dimension, but also offers the possibility of 'filtering' the magnetization transfer through a chosen type of nucleus.

When assigning an RNA molecule HMQC (homonuclear multiple quantum correlation spectroscopy) and HSQC (homonuclear single quantum correlation spectroscopy) experiments are useful for classification of nonexchangeable proton resonances through $^{1}H-^{13}C$ and $^{1}H-^{15}N$ correlations. Ribose protons can be efficiently assigned by a set of 3D correlation experiments as described by Pardi.[23] Assignment of all protons in a ribose spin system of a ^{13}C-labeled sample can be accomplished by performing a set of HCCH–COSY, HCCH–RELAY and HCCH–TOCSY (total correlation spectroscopy) experiments.[24] The individual ribose spin systems can then be connected by $H_sC_sN_bC_bH_b$ triple-resonance experiments (s and b refer to sugar and base, respectively) (Figure 5a).[25,26] A method for sequential through-bond correlation of ^{13}C-labeled nucleotides has been described,[27] and through a ^{1}H, ^{13}C, ^{31}P triple-resonance experiment (HCP-CCH-TOCSY) correlation of all ribose protons on both the 5′-side and 3′-side of a backbone phosphate nucleus can be made (Figure 5b). In addition to these experiments, usage of partial deuteration and selectively incorporated isotope labels may increase the resolution and sensitivity as discussed above.

Correlation of the exchangeable amino and imino protons in fully ^{13}C and ^{15}N labeled RNA can be done

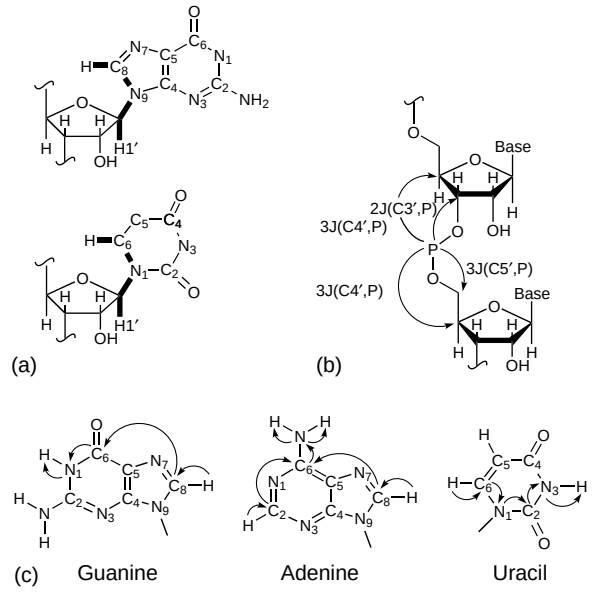

Figure 5 Schematic representation of coherence transfer pathways connecting base and sugar in the HCNCH experiment, indicated in bold gray (a), along the backbone in the HCP-CCH-TOCSY experiment (b), and within the guanine, adenine and uracil bases in the HCCNH-TOCSY experiment (c).

using HCC(C)NH-TOCSY experiments that are well suited to assigning exchangeable protons in nonregular portions of structures (Figure 5c).[28,29–31] To assign DNA structures containing ^{13}C and ^{15}N labels strategies that use the same NMR experiments that are used for the assignment of RNA molecules can be applied.

Following successful assignment of the resonances in a molecule, experiments tailored to generate structure information can be conducted. For internuclear distance measurements, various types of NOESY experiments, such as 3D NOESY/HMQC[32] and four-dimensional HMQC/NOESY/HMQC[7,32] may be used. In addition to homonuclear 2D NOESY spectra, from which interproton distances can be estimated with reasonable accuracy, heteronuclear 3D experiments may be a useful source of distance constraints, particularly for large molecules in which spectral overlap may prohibit the measurement of individual 2D crosspeak intensities.

The use of dihedral angles (derived from *J*-couplings) in the structure determination of nucleic acids can be of crucial importance, particularly for determining conformations of the backbone where dipolar proton interactions (NOESY crosspeaks) are sparse or difficult to resolve. For unlabeled molecules the determination of sugar conformations from *J*-coupling data has been described by graphical[21] and simulation[33] methods. For ^{13}C-labeled nucleic acids, special experiments have been developed to determine a complete set of coupling constants for oligonucleotides.[34] A novel method for quantitative measurements of the three-bond $^{1}H3'-^{31}P$ couplings in nucleic acids and nucleic acid–protein complexes based on constant time COSY experiments has been reported.[35]

The pioneering technique of measuring dipolar coupling in samples partially oriented by anisotropic cosolutes[36] has been applied to nucleic acids.[37] The cosolute, in this case filamentous phage, becomes aligned in the magnetic field, which is sufficient to extract net through-space $^{1}H-^{15}N$ dipolar coupling. Double quantum filtered (DQF) COSY experiments carried out with the addition of phage not only have the potential to give a large number of angular constraints that complement distance and *J*-coupling derived dihedral angle constraints, but also to yield long-range angular information of a type that has not previously been accessible using standard NMR techniques.

Biological molecules in solution always undergo dynamics on different timescales. Perhaps the most well-characterized motion in nucleic acids is the basepair opening dynamics. Imino protons in Watson–Crick basepairs are usually well resolved and thus can, after assignment, report on hydrogen bonding events specific to individual basepairs. The guanine H1 and thymine H3 protons, located in the center of the basepair, are only accessible to exchange after transient opening and lateral motion of one or both of the bases. Amino protons on basepaired cytosines rotate much less frequently than non-base-paired amino protons, resulting in the appearance of well-resolved resonances from hydrogen bonded and exposed protons at characteristic chemical shift intervals (8.0–8.5 ppm and 6.5–7.0 ppm, respectively). Amino protons of adenines and guanines, on the other hand, rotate at intermediate rates and give rise to broad peaks that are often difficult to detect. The exchange of imino protons is catalyzed by base.[38] When paired and open states are in rapid equilibrium the imino protons exchange occasionally and at a measurable rate, τ_{ex}, that depends on the concentration of catalyst [B] and can be expressed according to,[2] Equation (3)

$$\tau_{ex} = \tau_0 \left(\frac{1 + B_{eq}}{[B]} \right. \tag{3}$$

where B_{eq} is the buffer concentration at which $k_{cl} = k_{tr}$, i.e. the closing rate is equal to the proton transfer rate, and can be estimated from the base concentration for which $\tau_{ex} = 2\tau_0$. Extrapolation to infinite buffer concentration gives the basepair lifetime, τ_0. The exchange rate is most commonly measured as $\tau_{ex} = (\pi\Delta\nu)^{-1}$, where $\Delta\nu$ is the line width at half height of the resonance of interest. From studies of imino proton exchange it has become clear that the basepair opening rates are sequence dependent. For example, the exchange times of guanine iminos in G_n tracts are shorter (less than 5 ms) than in $(GC)_n$ tracts or isolated GC basepairs. In contrast the lifetimes of thymine iminos in AT tracts are very long.[39]

Water molecules have an important role in stabilizing nucleic acid structures. This role is partly played by specific association of water molecules to the surface of nucleic acid molecules. Detection of such bound water and estimation of residence times can be done by NMR techniques.[40,41] Interactions between water and nucleic acid protons, for example in one of the grooves of B-form DNA, can be seen as crosspeaks in a NOESY spectrum. Owing to efficient exchange of water molecules between the hydration sites and the bulk, all water peaks appear along the chemical shift of the bulk water signal. It is necessary to differentiate between crosspeaks arising from chemical exchange of water molecules, on the one hand, and exchange NOE crosspeaks arising from nonexchangeable DNA protons and labile DNA protons that exchange rapidly, on the other hand. If the nonexchangeable DNA proton of interest is located close to potentially labile DNA protons, these may appear as slowly exchanging water molecules visible at the chemical shift of bulk water, although they cannot be distinguished from DNA protons in rapid exchange with solvent. DNA protons that exchange rapidly give rise to crosspeaks with the water signal that are positive in a NOESY (NOE

spectroscopy in the rotating frame) and positive in a ROESY (nuclear Overhauser enhancement spectroscopy in the rotating frame). True water–DNA NOEs, in contrast, will give rise to negative ROESY crosspeaks.

Basepairing through hydrogen bonds primarily manifests itself as imino proton resonances, visible because of the slowed down imino proton exchange. In a reported experiment direct evidence of basepairing hydrogen bonds has been observed. In a J_{NN} COSY experiment of ^{15}N-labeled RNA two-bond couplings between the N1 of adenosine and the N3 of uracil and between the N1 of guanine and the N3 of cytosine, couplings of approximately 7 Hz have been measured.[42] Further experimental development has also made possible the observation of J_{NN} coupling between ^{15}N nuclei with large chemical shift separations, as shown between amino and N7 nuclei in a uniformly ^{15}N-labeled DNA molecule.[43]

2.2.3 Refinement

Completing a macromolecular structure based on NMR data in practice requires computational modeling into which experimental constraints consisting of internuclear distances and/or dihedral angles can be incorporated. It is generally agreed that the most efficient way of arriving at a molecular structure that satisfies all experimentally derived constraints is to generate a number of starting structures by embedding the molecule according to distance data using so-called distance geometry.[44] Thereafter, molecular dynamics calculations using one of the available forcefields are carried out, typically using a simulated annealing protocol to refine the structure.[44] Incorporation of any adduct or complex into the structure should be done at the structure generation stage. For molecules that can be anticipated to be close to a known conformation, suitable starting structures can be built, for example canonical B-form DNA. However, one has to verify that the final product does not suffer from being biased by the choice of starting conformation. If the structure relies heavily on interproton distance data and high accuracy, for example regarding sugar pucker conformation, it may be appropriate to include relaxation matrix refinement in the molecular dynamics protocol to account for spin diffusion effects adequately, as discussed below.

Care should be taken to ensure that chirality errors do not occur as a result of refinement processes, which has been observed to happen in numerous reported nucleic acid structures.[45]

2.3 Critical Comments

2.3.1 Accuracy and Validity

Several factors limit how accurately nucleic acid structures can be determined by NMR methods. In solution nucleic acids are not static molecules that adopt structures that can be characterized with infinite accuracy. Rather, RNA and DNA constantly undergo dynamic processes that should be taken into account when describing an RNA or a DNA solution structure. Since NOE intensities do not scale linearly with internuclear distances, it is often not possible to satisfy consistently all NOESY-derived distance constraints and angular constraints within one structure, even if a two-state model with alternate sugar conformations is assumed. However, from analysis of the NOE intensities in a spectrum of a duplex DNA it has been concluded that restricted rotational diffusion around the glycosidic bond, χ, combined with sugar repuckering can explain the observed spectral features satisfactorily.[46] Furthermore, from relaxation studies of ^{13}C-labeled purines in a duplex it has been found that in addition to overall isotropic tumbling, which occurs on the timescale of 3–4 ns, there is fast (20 ± 20 ps) internal motion of individual purines. This motion, if described as wobble-in-a-cone, is restricted to $\pm 22.5°$ for the 1′, 3′, 4′, 2 and 8 protons in the purine, with no apparent sequence dependence.[47]

2.3.2 Potential and Limitations

The effects of spin diffusion give rise to significant errors in NOE-derived internuclear distance measurements, particularly when long NOESY mixing times (>100 ms) are employed.[48] Experimental methods to overcome this through network editing techniques have been developed. In BDNOESY (band decoupled nuclear Overhauser spectroscopy)[49] and CBDNOESY (complementary band decoupled nuclear Overhauser spectroscopy)[50] experiments, pulses are applied during the mixing time that effectively suppress the effects of spin diffusion through non-H1′ ribose protons and all intraribose cross-relaxations, respectively. The resulting data allow tighter error limits during refinement which give improved precision and allow longer mixing times, thereby giving higher sensitivity and increasing the maximum distance that interactions can be detected. This method has been demonstrated on a ribozyme structure.[50]

It has been concluded that RNA molecules as large as 15 kDa should be amenable to structure determination by NMR methods with a precision and accuracy of 1–1.5 Å.[51] It is also noted that the most substantial improvements in structure quality were accomplished by including constraints derived from exchangeable protons rather than using narrower distance constraints. The molecular size limit is to a large degree set by the availability of experimental constraints – internuclear distances and dihedral angles. The number of constraints per residue is more likely to be large in molecules with

globular shapes, such as compactly folded RNAs. Thus, with instruments of higher sensitivity and with efficient labeling techniques and more robust refinement methods the future appears promising for structure studies of molecules of increasing size.

3 STRUCTURES DETERMINED

3.1 DNA Motifs

3.1.1 Duplex Structures

Among the many solution structures of DNA oligomers that have been determined by NMR methods, the structure of the TATA-box containing self-complementary octamer d(GGTATACC) will be discussed first.[48] The final structures of the octamer duplex were based solely on interproton distance constraints, with the effects of spin diffusion accounted for by inclusion of relaxation matrix calculations in the refinement protocol. The use of a 3D proton NOESY–NOESY experiment (120 ms mixing time in both indirect dimensions) clearly illustrated spin-diffusion pathways as crosspeaks appearing at chemical shifts that were different in all three dimensions. For example intraresidue H6 → H2′ → H3′ and H6 → H2″ → H1′ crosspeaks were seen, demonstrating that the H2′ and H2″ protons mediate spin diffusion efficiently. Similar diffusion pathways were also seen in sequential aromatic–sugar interactions. From the resulting structures it was concluded that the molecule in solution adopts a structure that relatively closely approximates the canonical B-conformation.

DNA displays structural irregularities – sequence and condition-dependent features that may be crucial to its biological function. A molecular structure of extended shape, such as a DNA duplex, determined from short-range distance constraints, has an intrinsic lack of accuracy attached to it. It should thus be pointed out here that the structural details that may be sought in a high-resolution structure of an oligonucleotide duplex alone in solution need to be most carefully established in order to merit significant conclusions to be based upon them.

3.1.2 Triplex Structures

DNA of appropriate sequences under favorable conditions will form three-stranded complexes. A number of different triplet basepairing schemes have been observed, as reviewed by Radhakrishnan and Patel.[52] The most commonly encountered triplex is based on the pyrimidine·purine–pyrimidine motif. In this complex, the Watson–Crick pairing pyrimidine-rich strand runs antiparallel to the purine-rich strand to which the second pyrimidine strand binds in a parallel alignment through Hoogsteen hydrogen bonds (Figure 6).

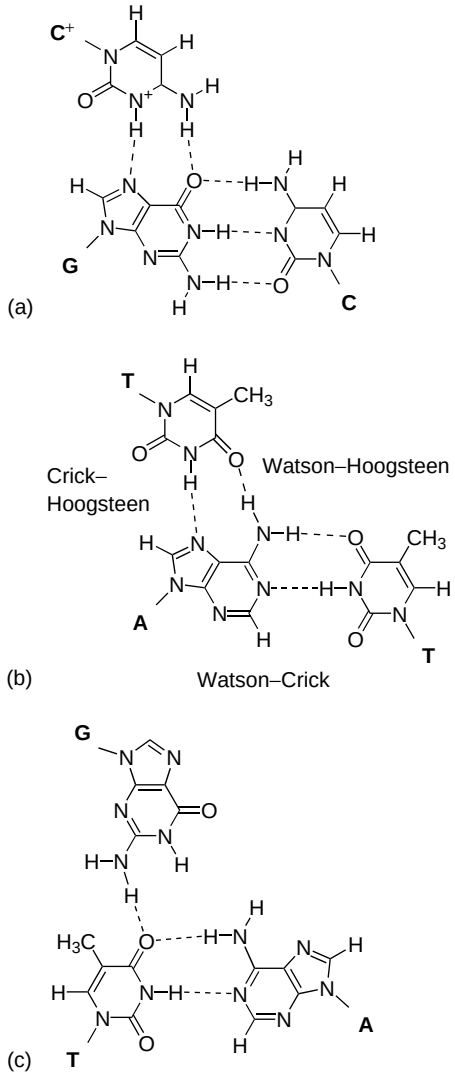

Figure 6 Alignment of bases in base triplets: C^+·G-C (a), T·A-T (b) and G·T-A (c).

The hydrogen bonds predicted for this motif were confirmed by NMR studies focusing on exchangeable resonances and strand orientation.[53,54] Stable formation of such triplexes requires protonation of the cytosines on the Hoogsteen strand, and is thus favored by low pH. The one-dimensional proton NMR spectrum of a unimolecular DNA triplex that includes a G·T-A triplet in the center is shown in Figure 7.

A few notable features will be pointed out. First, the cytosine imino protons appear downfield of the other imino proton peaks (14.5–16.0 ppm). Downfield shifting of the exchangeable as well as nonexchangeable protons in the protonated cytosine generally follows upon protonation. Second, the thymine iminos

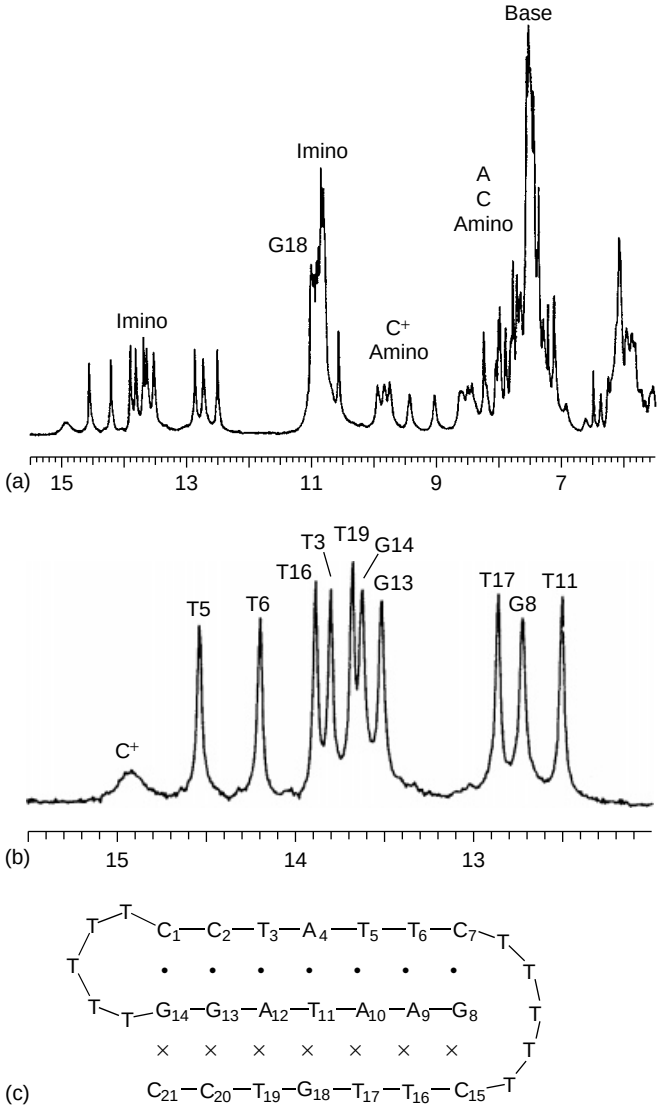

Figure 7 Proton spectrum, 5.5–15.5 ppm (a), with expanded imino proton region (b) of the triplex forming DNA sequence (c). (Reproduced by permission of Radhakrishnan et al., *Biochemistry*, 1991.[55] American Chemical Society.)

on the Hoogsteen strand resonate somewhat upfield (12.8–13.9 ppm) of the Watson–Crick thymine imino resonances (13.8–14.6 ppm). Third, both of the adenine amino protons are engaged in hydrogen bonding and appear in the 8.0–8.5 ppm region, whereas the hydrogen-bonded amino resonances of the protonated cytosine appear between 8.5 and 10.0 ppm. The spectrum of exchangeable protons is thus mostly well resolved, the imino resonances of the thymine loops seen around 11 ppm being an exception. The imino and amino protons in the Watson and Crick strands in this sequence were assigned by strategies similar to those described for a regular duplex (section 2.2 above) using 2D NOESY spectra recorded in H_2O solution with intermediate mixing times (120 ms).[54] In addition NOESY crosspeaks between thymine imino protons in T·A-T triplets and hydrogen bonded adenine amino protons are observed. The Hoogsteen pairing imino protons in the T·A-T triplets show NOEs to the H8 of the adenine and to the hydrogen-bonded adenine amino proton. The H8 of all adenines and guanines can be selectively deuterated in order to aid the identification of some of the imino crosspeaks. For the C^+·G-C triplet, strong NOEs are seen between the hydrogen-bonded and exposed amino protons on the protonated cytosine Hoogsteen-bound to the guanine.

The nonexchangeable protons of the molecule were assigned from a 3D NOESY/TOCSY experiment.[56] In this experiment the first step is a dipolar interaction NOE, whereas the second step involves a coherent transfer of magnetization between scalar coupled spin pairs. The experiment facilitated identification, for example of the various protons showing NOE interactions with H2′ and H2″ protons, which in turn exhibit strong couplings with the well-resolved H1′ resonances. The 3D structure of the triplex (sequence in Figure 7) is shown in Figure 8. The triplex was allowed to form in solution with Li^+ as counterions to avoid formation of four-stranded complexes which may be favored in G-rich sequences with Na^+ or K^+ present. The Watson–Crick paired strands do not deviate substantially from a regular duplex conformation, although some basepair displacement and helical unwinding is seen. The glycosidic angles are all in the *anti* domain and all sugars are in the S conformation, except for the guanine of the G·T-A triplet, which adopts the N

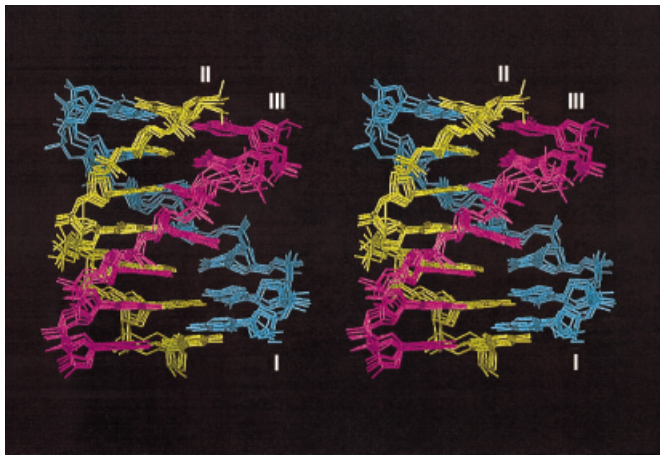

Figure 8 Stereo view of DNA triplex containing a G·T-A triplet, loops excluded. Strand I corresponds to bases C1–C7, strand II to G8–G14 and strand III to C15–C21 in the sequence shown in Figure 7. The strands are labeled at their 5′ ends. (Reproduced by permission of Dr. D.J. Patel of Sloan-Kettering Institute, New York.)

conformation. While the base triples are generally planar, the guanine of the G·T-A is tilted towards the 3′ direction. The grooves are all different, with the Watson–Crick groove being slightly narrower than in a duplex with a fairly smooth surface. The Watson–Hoogsteen groove in contrast appears largely wide, shallow and uneven. The Crick–Hoogsteen groove is narrow and has distinct cavities on its surface. Considering its lining of polar residues, this groove provides potential binding sites for water or positively charged ions. Ordered water molecules in the Crick–Hoogsteen groove have indeed been found, as shown from a combination of homonuclear NOESY and ROESY experiments carried out on the same triplex molecule.[57] Similarly, in the Watson–Hoogsteen groove, water molecules with residence times longer than 1 ns were found. The water molecules in the Watson–Crick groove of this triplex were found to be more mobile, in contrast to the spine of hydration found in the minor groove of duplex DNA.[41]

3.1.3 Quadruplex Structures

Four-stranded DNA complexes have received much attention, partly owing to their potential relevance to telomere sequences (guanine-rich multiply repeated short sequences at the ends of chromosomes). Understanding the structure and topology of these repeat sequences may lead to functional insights of biological or medical importance. Several types of intra- and intermolecular quadruplexes with different types of pairing alignments and combinations of strand directionalities have been studied (reviewed by Rhodes and Giraldo[58] and Darlow and Leach[59]). Both NMR and X-ray crystallography methods have been employed to determine structures of DNA quadruplexes.

Generally, G quartets (Figure 9) are highly stable, as has been evidenced, for example, by NMR studies of the human telomere sequence. Upon change of solvent from H_2O to D_2O the exchange of the imino protons with deuterons occurred at remarkably low rates that could be followed over days in the NMR spectrum.[60]

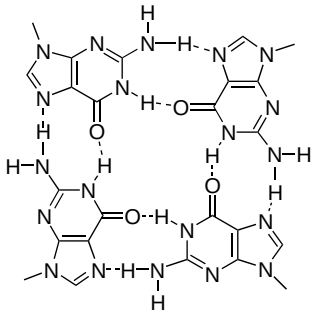

Figure 9 Alignment of guanine bases in G tetrad.

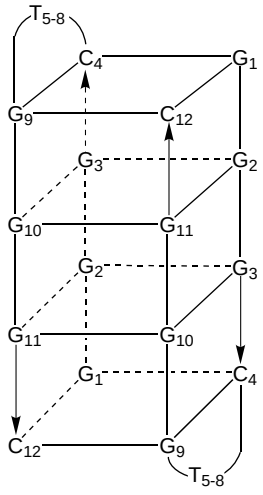

Figure 10 Topology of quadruplex structure formed by d($G_1G_2G_3C_4T_5T_6T_7T_8G_9G_{10}G_{11}C_{12}$).

Quadruplexes generally require the presence of low concentrations of monovalent cations for their stability, often discriminating between different types of ions. Typically potassium and sodium ions are found to coordinate the arrangement of bases in and between the tetrads, whereas lithium ions are too small to fulfill this function.

As an example of a four-stranded complex the structure formed by GGGC repeats, found in adeno-associated viral DNA, will be discussed.[61] In addition to forming G-tetrads this structure also forms G·C·G·C quartets (Figure 10). The four strands, combined by head-to-tail dimerization of two hairpins, run antiparallel to each other around the quadruplex, with lateral T_4 loops connecting the pair of arms on each hairpin. The interaction between the two intrastrand Watson–Crick GC basepairs is mediated by bifurcated Hoogsteen hydrogen bonds involving the major groove edges of the bases. In guanine tetrads the *syn* conformation around the glycosidic bond is commonly observed. In the present example the guanines of the G2 and G10 in the G·G·G·G·G tetrads are in the *syn* form, while the remainder of the bases are in the *anti* conformation.

The spectrum of this sequence (Figure 11a) in H_2O solution containing sodium ions shows two resonances between 13.0 and 13.4 ppm, characteristic of Watson–Crick interactions, and four additional peaks between 11.1 and 12.0 ppm, characteristic of G_4 tetrad formation.[62] The number of peaks seen in the spectrum shows that any four stranded complex formed must possess two-fold symmetry. All four cytosine amino protons of one symmetric unit are seen in the 8.7–9.4 ppm region, indicating that they all are engaged in hydrogen bonding.

Unambiguous assignment of the guanine imino proton resonances was reached through site-specific

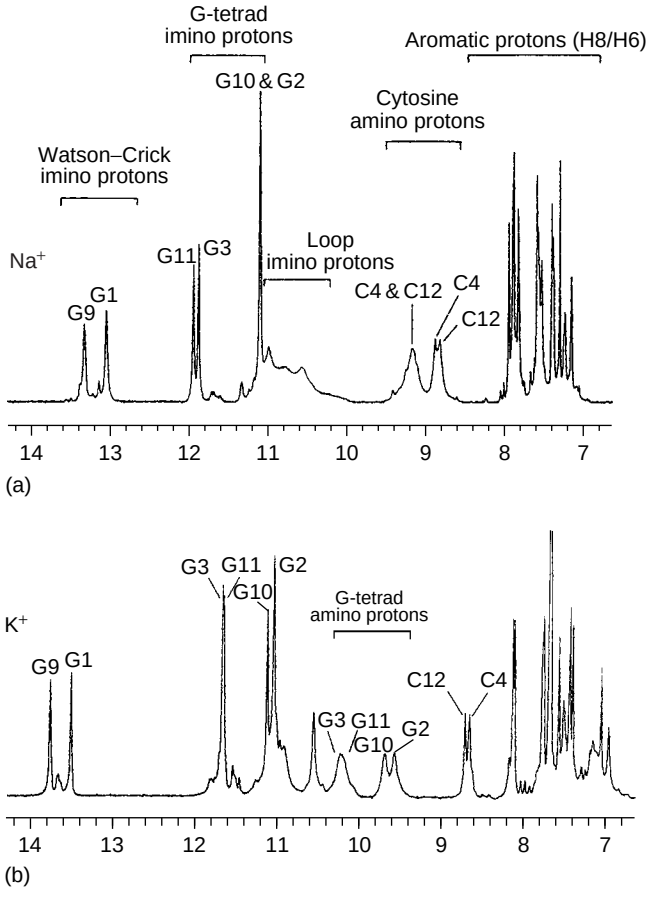

Figure 11 Proton spectra of d($G_1G_2G_3C_4T_5T_6T_7T_8G_9G_{10}$-$G_{11}C_{12}$) in H_2O solution at 10 °C stabilized by 100-mM Na^+ (a), and 100-mM K^+ (b). (Reproduced by permission of Bouaziz et al., *J. Mol. Biol.*, (1998) Academic Press.[63])

incorporation of ^{15}N isotopes in the N1, N2 and N7 positions of bases G1 and G10. Differences in the proton spectral pattern between ^{15}N decoupled and ^{15}N undecoupled spectra then enabled identification of the G1 and G10 imino proton resonances, which in turn permitted the assignment of the remaining imino protons of the molecule through analysis of NOESY spectra. The involvement of the G1 and G9 imino protons in G-C Watson–Crick base pairing was concluded from their showing NOEs to cytosine amino and H5 protons on the opposing base. Each of the G2, G3, G10 and G11 imino protons exhibits an NOE to the guanine H8 proton in the Hoogsteen aligned position. In the nonexchangeable proton NOESY spectrum strong intraresidue crosspeaks between H1′ and H8 are seen for G2 and G10 at short mixing times, evidence of a *syn* conformation around the glycosidic bond. The 3D structure of d(GGGC-TTTT-GGGC) stabilized by 100 mM Na^+ is shown in Figure 12(a). The quadruplex has four grooves, of which two are equivalent and wide-formed within each hairpin.

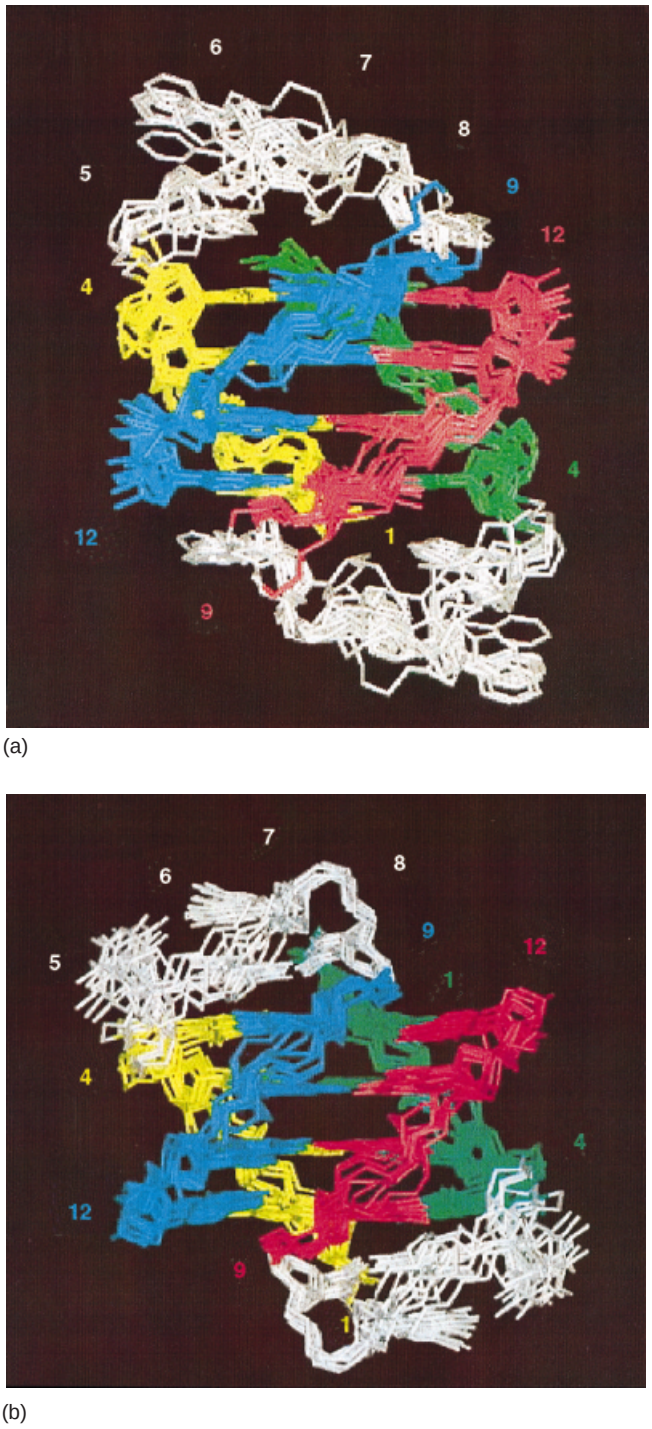

Figure 12 Structure of quadruplex formed by d($G_1G_2G_3C_4T_5$-$T_6T_7T_8G_9G_{10}G_{11}C_1$) stabilized by 100-mM Na^+ (a), and by 100-mM K^+ (b). (Reproduced by permission of Bouaziz et al., *J. Mol. Biol.*, (1998).[63] Academic Press.)

The other two grooves are nonequivalent and narrow and are formed by the G1-G2-G3-C4 and G9-G10-G11-C12 interstrand interactions.

An interesting isomerization occurs in the molecule upon shifting from Na^+ to K^+ as the predominant counterion in the sample solution.[63] The change is manifested as distinctly different features in the proton NMR spectrum (Figure 11b). The spectrum of exchangeable protons shows that the molecule remains in a structure of two-fold symmetry, but that the basepairing pattern is different with K^+ ions present. The cytosine amino protons no longer appear close together, instead two of them have shifted upfield, indicating that they are no longer hydrogen bonded. The NOESY crosspeaks indicative of G·C·G·C tetrads (i.e. between H8(G1) and H5(C4) and between H8(G9) and H5(G12)) are not observed in the K^+ form of the quadruplex. A basepairing scheme for the G·C·G·C planes of the quadruplex with intrastrand Watson–Crick hydrogen bonds, but without pairing interactions in between the two strands was concluded. The K^+ stabilized structure remains a quadruplex (Figure 12b), but one with different hydrogen bond patterns in the outer tetrads and with T4-loop structures that differ from the ones seen in Na^+ solution (Figure 13).

A novel method of directly localizing the cation binding sites in DNA quadruplexes has been reported.[64] The $^{15}NH_4^+$ ion was used as a probe of cation coordination of G-tetrads, replacing Na^+ or K^+. In ^{15}N filtered proton spectra of the *Oxytricha nova* telomere sequence $d(G_4T_4G_4)$, the ammonium ions showed three distinct peaks at low pH. These could be assigned to NH_4^+ ions bound to two different DNA sites, populated as 2:1, while one much larger peak arises from bulk NH_4^+. From ROESY and NOESY experiments the ammonium ions could be mapped to loci between the planes of the G-tetrads (cf. section 2.2). Two more ammonium ions, one on either end of the quadruplex, are bound between the central G4 tetrads and its flanking outer tetrad. From measurements of the resonance line widths of the coordinated NH_4^+, their lifetimes were estimated to over 50 ms for the outer ions [N.V. Hud, P. Schultze, V. Sklenář, J. Feigon, 'Binding Sites and Dynamics of Ammonium Ions in a Telomere Repeat DNA Quadruplex', *J. Mol. Biol.*, **285**, 233–243 (1999)]. From chemical exchange N_2-Ex-HSQZ experiments with varying mixing times it was estimated that ammonium ions binding to the inner sites have residue times as long as 250 ms. These experiments also gave evidence for a pathway along which the NH_4^+ ions enter the quadruplex through the end openings of the cavity running along the helix axis and not through the grooves.

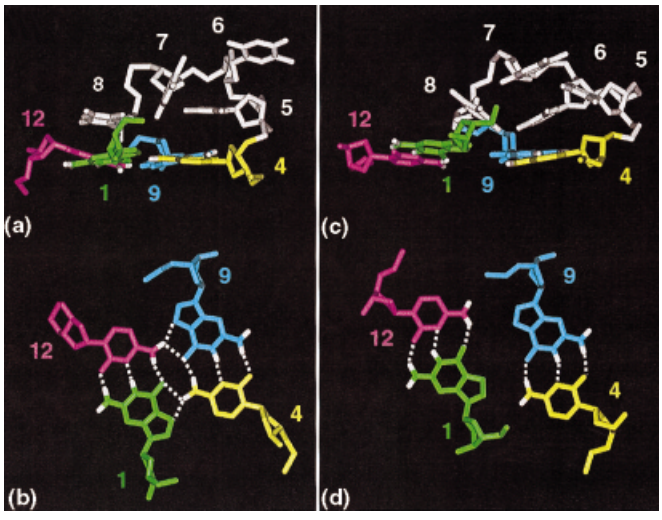

Figure 13 Details from structure of $d(G_1G_2G_3C_4T_5T_6T_7T_8G_9G_{10}G_{11}C_1)$ stabilized by 100-mM Na^+ (a) and (b) and 100-mM K^+ (c) and (d). Comparison of $T_5T_6T_7T_8$ loop structures and alignments of $G_1C_{12}G_9C_4$ tetrads with hydrogen bonds shown as broken lines. (Reproduced by permission of Dr. D.J. Patel of Sloan-Kettering Institute, New York.)

3.1.4 DNA–Drug Complexes

DNA interacting with small molecules will here be represented by the binding of the chiral transition metal complex $[Ru(phen)_3]^{2+}$ and the oligonucleotide $[d(CGCGATCGCG)_2]$.[65] The propeller shaped Ru-complex binds to the oligonucleotide with some sequence preference and with exchange kinetics that are rapid on the NMR timescale (milliseconds). While the fast kinetics precludes determination of a well-defined DNA–metal complex structure by NMR methods, conclusions about the type of interaction and its effects on the structure of the DNA oligomer can be reached.

Figure 14 shows the change in the aromatic region of the one-dimensional proton spectrum as the Ru/oligonucleotide duplex concentration ratio increases. Most notably a few resonances, both of the oligonucleotide and of the phenanthroline ligands, shift drastically as more Ru-complex is added. These chemical shift changes can be used to estimate binding constants for the metal complex, which were found to be in the range of 10^3–10^4, and also to provide evidence for sequence and groove specific interactions. In the sequence studied the most marked effects are seen around the central AT step where the chemical shift of the H2 proton of adenine reports drastic alterations in its local environment. Crosspeaks in NOESY spectra of the $[Ru(phen)_3]^{2+}$–oligonucleotide complex supplied strong evidence for binding in the AT region of the minor groove – many intermolecular NOEs of varying intensities were observed. The NOESY data also indicates that the DNA molecule does not undergo any significant structure changes but remains in a B-like conformation, and that the metal complex binds in a nonintercalative fashion.

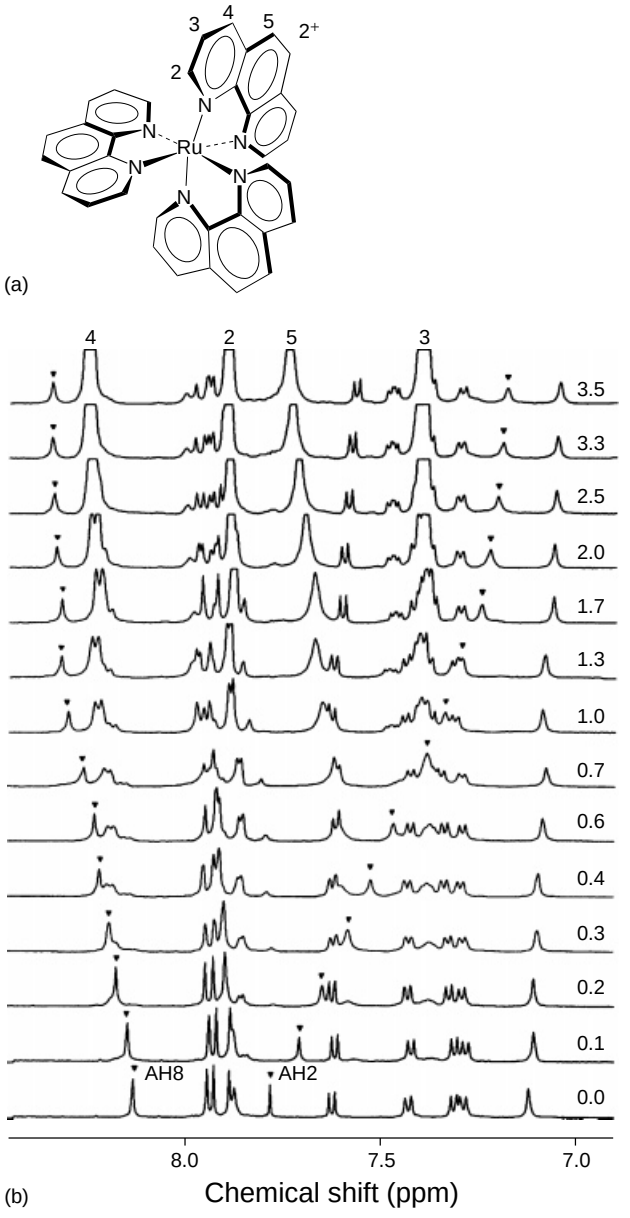

Figure 14 (a) Δ-$[Ru(phen)_3]^{2+}$, (b) aromatic region of one-dimensional proton spectrum $[d(CGCGATCGCG)]_2$ and Δ-$[Ru(phen)_3]^{2+}$ in D_2O solution at 41 °C. Ru/DNA–duplex concentration ratios are indicated to the right, and the phenanthroline resonances and the H2 and H8 resonances of A5 are indicated on each spectrum.

3.1.5 DNA–Protein Complexes

The structure of the lymphoid enhancer-binding factor (LEF-1) to a 15 basepair long DNA oligomer will serve here to exemplify DNA–protein complex structures. The high-mobility group (HMG) of LEF-1 from mouse, containing 84 amino acids, was bound to the oligonucleotide 5′-CAC**CCTTTGAAG**CTC·CAG**CTTCAAAGG**GTG-3′ (bold face representing the protein binding site) in a 1 : 1 stoichiometric ratio, and the structure was determined by NMR methods.[66] Uniformly ^{13}C,^{15}N labeled LEF-1 and an extensive array of 2D and 3D experiments were used. The DNA portion of the complex was assigned from 2D ^{13}C,^{15}N double half-filtered NOESY and ^{13}C half-filtered NOESY and ^{13}C-decoupled homonuclear proton NOESY spectra. NOE crosspeaks between LEF-1 and DNA were identified from a 3D ^{13}C-edited ^{12}C-selected NOESY experiment. The structure determination was based on nearly 1450 intraprotein, 325 intra-DNA and 332 intermolecular distance constraints, and 42 dihedral angle constraints for the protein. The HMG domain of LEF-1 is dominated by three α-helices arranged to give it an L-like shape. The protein is wrapped around the DNA, forcing it to bend into a compressed major groove. Intimate protein–DNA contacts are formed between a concave surface of LEF-1 and the minor groove of the oligonucleotide where a methionine side chain is partly intercalated, wedging into the DNA without disrupting any of its basepairs.

3.2 RNA

3.2.1 Choice of Domains

Many RNA molecules of interest are prohibitively large for structure determination using NMR methods. This problem can often be circumvented by reducing the size of the RNA molecule of interest to a minimum portion that retains its specific functions. In some cases the anticipated secondary structure of an RNA molecule can suggest how such reductions may be performed without excluding vital interactions. A few structure motifs have been observed to reoccur in several RNA structures. The UUGC and GNRA tetraloops (where N signifies any base and R signifies a purine) are examples of this. The stable GNRA loop has also been used to cap a shortened stem or to connect loose ends, for example in the U1A binding RNA polyadenylation inhibition element.[67] The stability of the GNRA loop has been found to form a network of heterogeneous hydrogen bonds and extensive base stacking within the loop, explaining its unusual stability.[68] It should also be kept in mind that an RNA molecule that specifically binds to a ligand may not be correctly folded, or may not even adopt a stable structure, in the absence of its cognate molecule. An example of this is given by the AMP (adenosine monophosphate) aptamer RNA.[69] Furthermore, magnesium ions often play important roles in RNA structure stability.

3.2.2 Examples of RNA Structures

3.2.2.1 Adenosine Monophosphate–RNA Aptamer

An RNA aptamer structure can be nicely illustrated by an ATP (adenosine triphosphate) specific sequence of which

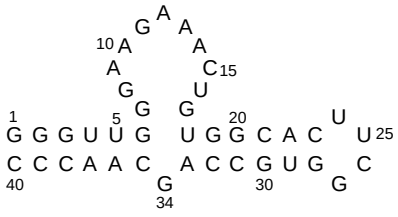

Figure 15 Sequence of AMP aptamer RNA.

the structure has been determined by NMR methods. In vitro selection performed on a column to which ATP was attached via a linker to its C8 position, combined with enrichment procedures, yielded a consensus sequence from a pool of RNAs 169 nucleotides long.[70] This sequence could be shortened to a 40-mer containing an 11 nucleotide consensus sequence (Figure 15) keeping its ATP affinity, which was characterized by a K_d (dissociation constant) of about 10 μM. The affinity was found to be the same for adenine and AMP as for ATP, and to increase upon raising the concentration of Mg^{2+}. The aptamer molecule was, based on its sequence, proposed to adopt a structure containing an internal loop opposite an invariant guanine, flanked by two basepaired stems. This prediction was confirmed by two independent structure studies. The 40-mer RNA aptamer, shown in Figure 15, in a complex with AMP was determined by Jiang et al.,[71,72] and Dieckmann et al.[73] have determined a closely related 36-mer RNA–AMP.

Structure determination by NMR methods revealed that the conserved internal loop and the bulged guanine fold around the ATP in a compact manner, embracing most of the purine portion of the AMP, leaving only its C8 edge exposed (Figure 16). The backbone of this portion of the RNA has been likened to a Greek zeta (ζ). The loop nucleotides G8, A9 and A10 combine with the AMP to stack in a way similar to a GNRA turn.[68] Two asymmetric G·G mismatch basepairs, one of them involving the bulged G34, cap the two helical stems that extend from the loop, giving the entire aptamer an L-like shape. Based on nearly 40 intermolecular NOEs between the AMP and several of the bases in the loop, the AMP could be located quite precisely. The AMP was found to stack between the A10 and G11 bases of the RNA loop. The RNA molecule was found to undergo an AMP-induced structural transition,[69] evidenced by the appearance of 10 new resonances from exchangeable protons upon addition of AMP.

Assignment of the ATP aptamer complex presented challenges, not only because of the size of the system, but also because of the unknown structure of the internal loop which, in combination with the large number of NOEs observed, makes it difficult to distinguish sequential from long-range interactions. For the unambiguous assignment

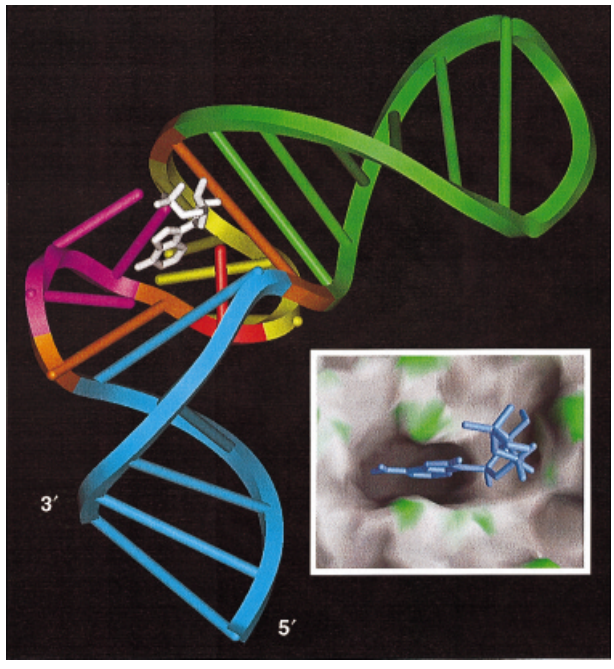

Figure 16 Solution structure of the AMP–RNA aptamer complex, sequence shown in Figure 15. RNA is represented by ribbons, the G1–G6·C35–C40 stem is colored blue, the G19–C32 hairpin portion is colored green, and the large G7–G17 loop is colored in red–yellow shades. The bound AMP molecule is shown in stick representation in white. Insert shows AMP in blue and RNA in surface representation. (Reproduced by permission of Dr. D.J. Patel of Sloan-Kettering Institute, New York.)

of the AMP aptamer, a variety of multidimensional experiments were carried out on samples of unlabeled and uniformly ^{13}C and ^{15}N labeled RNA and AMP. The use of labeled RNA in a complex with unlabeled AMP, or vice versa, permitted the distinction of intra- from intermolecular NOE interactions involving the RNA and AMP molecules.[72,74]

Identification of exchangeable proton resonances in the internal loop can be approached by different strategies.[74,75] Site-specific ^{15}N labeling of RNA purines (N1 imino and N2 amino positions in guanines and N6 amino positions in adenines) of most purines one at a time in the loop and bulge regions of the AMP–RNA aptamer made possible the assignment of the exchangeable protons in this region. The specifically isotope-labeled bases required for this strategy may, however, not be readily available. In an alternative approach, the guanines of the internal loop and the bulge guanine were systematically substituted for inosine. The replacement by inosines causes a downfield shift of nearly 2 ppm of the corresponding imino proton, facilitating its assignment. However, some of the inosine-for-guanine substitutions resulted in weakened binding of AMP, suggesting cautious use of

the inosine substitution strategy.[75] One assignment of special importance for this structure is the proton of the 2′-hydroxy group of G34, which shows a large number of NOEs to surrounding RNA and AMP protons. This hydroxy proton was identified from its appearance as a singlet peak in spectra of uniformly ^{13}C and ^{15}N labeled RNA recorded without decoupling, thus excluding it from being attached to a carbon or a nitrogen atom. It is notable that the OH-proton appeared extensively downfield shifted (9.34 ppm) compared with previously observed hydroxy proton resonances (ca. 6.8 ppm in the UUGC loop[76]).

The exchange rates of RNA imino protons have been measured in the AMP–RNA aptamer complex.[77] The study reported that the imino protons around the ATP-binding loop containing several non-Watson–Crick basepairing interactions are stabilized by ATP complexation, as well as those in its surrounding stem basepairs. Furthermore, the exchange time of the sugar 2′OH proton of G34 (Figure 15) was measured successfully and found to be about 20s, which is very long compared to the corresponding 2′OH in a ribonucleoside monomer (<1 ms), lending strong support to this hydroxy group participating in hydrogen-bond formation.

3.2.2.2 RNA–U1A Protein As a representative structure of RNA–protein complexes determined by NMR, the human U1A protein bound to an RNA molecule containing the polyadenylation regulatory element of the pre-mRNA 3′-UTR (untranslated region) will be discussed. The 30 nucleotide long RNA sequence, shown in Figure 17, contains a flexible internal loop, which becomes ordered when binding to the protein. The RNA binding domain of U1A comprises residues 2–102, and contain the amino terminal ribonucleoprotein domain. For reasons of sample stability the U1A protein studied contained two mutations of amino acids not directly involved in RNA recognition, which gave a K_d of less than 10^{-9} M with the RNA construct. The strategy chosen for assigning the RNA–protein complex was to make extensive use of previously obtained assignments of the free components.[78] This was found attractive partly because of the higher spectral quality of the individual components compared with the complex. The RNA component had been previously assigned both alone and in complex with U1A.[67] Resonances originating from the protein and the RNA resonances and their respective NOE crosspeak patterns were distinguished in spectra of the complex, based on their chemical shift differences – RNA only rarely gives rise to signals upfield of 3.5 ppm – and from spectra of complexes containing double-labeled protein. The helical portions of the protein were mostly unaffected by RNA binding, except for one of the helices for which large chemical shift changes occur. Only a minority of the amino acid residues could not be assigned from comparison with free protein. For these residues heteronuclear through-bond correlation data obtained with $^{13}C/^{15}N$ double-labeled protein was used.

Intermolecular interactions were observed selectively through 2D half-filtered experiments on isotope-labeled protein in complex with unlabeled RNA.[79] Most intermolecular NOEs were identified from 3D ^{13}C-edited spectra of labeled protein in complex with unlabeled RNA. In total more than 2600 unambiguous distance constraints were used for structure refinement. A majority of these, over 1700 constraints, came from intramolecular protein interactions, near 600 from the RNA internally and 123 constraints were intermolecular. In addition, more than one hundred dihedral angles and 25 hydrogen bonds were used to determine the RNA component.[78] The resulting structure, Figure 18, shows that the RNA molecule is severely kinked and that its helical parts interact with the variable loops of the protein. The β_2–β_3 loop

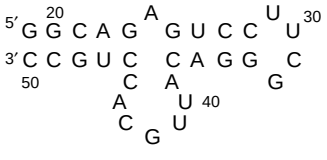

Figure 17 Sequence of U1A binding RNA.

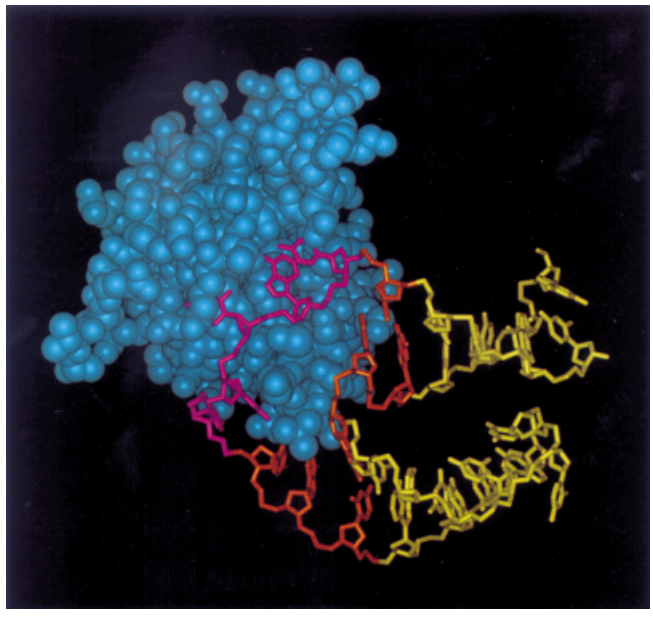

Figure 18 Structure of the RNA binding domain of human U1A bound to the RNA binding portion of the polyadenylation regulatory element of the U1A pre-mRNA 3′ UTR studied by Allain et al.[5] Protein shown in space filling (cyan), and RNA shown in stick representation. RNA residues interacting with the protein are shown in orange and magenta.

protrudes through a hole in the RNA formed between the internal loop and the helical stems. The single stranded residues of the RNA internal loop are splayed out across the surface of the β-sheet of U1A, with each of the nucleotides engaged in intra- or intermolecular stacking or hydrophobic interactions with other bases or aromatic protein side chains.

3.3 PNA and Other Variations

A growing interest in employing the basepairing specificity of nucleic acids to target specific genes has spurred the development of nucleic acid analogs tailored to combine improved function with desired properties (reviewed by Mesmaeker et al.).[80] Several modifications of the DNA backbone have been implemented in this pursuit, including substitutions within the phosphate groups and modification of the sugar moieties, as well as replacement of the entire sugar phosphate skeleton by a peptide-like chain, as in PNA (Figure 1). Insights into the structure of these DNA analogs may lead to an understanding of their function, as well as suggesting ways of modifying them for improved applicability. PNA forms stable hybrid duplexes with both DNA and RNA, and forms PNA·DNA–PNA triplexes with purine-rich DNA strands.[81]

3.3.1 PNA–DNA Hybrid Duplex

The structure of the PNA–DNA hybrid H-GCTATGTC-NH_2·d(GACATAGC) has been studied.[6] Generally assignment strategies of PNA protons need to be tailored to the specific system. In the instance of a PNA–DNA duplex the assignment of the PNA strand was complicated because of the lack of an easily observed continuous sequential crosspeak pattern along the exchangeable portion of the PNA protons. Instead, aromatic–methyl crosspeaks within thymines and their flanking residues were used to identify sets of three residues, which, along with similar patterns around H5–H6 interactions in cytosines, allowed assignment of all aromatic protons. The imino resonances, assigned via the DNA strand were, together with the adenine H2 crosspeaks, useful for confirming these assignments and for assigning additional exchangeable basepairing protons. The amide resonances were assigned from their NOEs to the H6/H8 protons on their own and N-terminal side flanking residues. The amide–aromatic NOEs observed in H_2O solution provide a pattern that may be amenable to sequential assignment. Scalar interactions, seen as TOCSY crosspeaks, were seen between backbone amide protons and the intraresidue H4′ and H5′ protons of the preceding residue. Each aromatic H6/H8 proton on the PNA strand shows several NOEs to the C7′, C4′ and C5′ methylene protons, which confirm the assignments of these protons. The pattern in the aromatic–sugar region of a NOESY spectrum shows typical B-form characteristics such as *anti* glycosidic bonds and near C2′-*endo* sugar conformations for the DNA strand.[82] Determination of the 3D structure of the hybrid duplex revealed a B-like conformation with a wide major groove and a narrower minor groove.[6] Somewhat irregular basepair stacking was observed and was supported by an unusually dispersed imino proton spectral region and by several interstrand NOE interactions. The structure of this PNA–DNA duplex is in many ways similar to the Watson–Crick portion of a PNA·DNA–PNA triplex structure that was determined by X-ray crystallography methods.[83]

3.3.2 PNA–RNA Hybrid Duplex

A PNA strand basepaired to an RNA oligonucleotide adopts a different conformation and requires different assignment strategies. Assignment of a PNA–RNA duplex was accomplished by incorporating a single fully $^{13}C-^{15}N$ labeled thymine residue in the PNA strand,[84] which was helpful in several ways. First, it helped to identify the labeled residue in double quantum or TOCSY spectra. Second, through NOESY experiments its neighboring residues could be identified. Third, 2D ^{13}C filtered NOESY experiments were useful for characterizing the local backbone conformation. Sequential NOE connectivities were observed throughout the PNA strand. The structure of a PNA–RNA hexamer duplex was determined and found to be in an A-like conformation with C3′ *endo* sugars in the RNA strand. The PNA strand backbone has many features in common with the one of the PNA–DNA hybrid.[81]

4 FUTURE OUTLOOK

New developments are under way that will move the horizon of the impossible further away. Most of the newly developed techniques work more efficiently at higher magnetic fields, and hence may be increasingly used as NMR instrumentation becomes further advanced.

4.1 Transverse Relaxation

Transverse relaxation optimized spectroscopy (TROSY) is a new technique used to measure scalar couplings in large molecules. TROSY experiments have been used to measure $^{15}N-^{15}N$ and $^{15}N-^{1}H$ couplings through hydrogen bonds in DNA and should be applicable to large molecules such as nucleic acids in complexes with proteins.[85]

4.2 Oriented Samples

Newly developed NMR methods that use the addition of matrixes to the sample in order to accomplish a net orientation of the sample, have proven powerful for protein structure determination.[36] By measuring the orientation of vectors in the molecule (derived from different coupling constants measured at different magnetic field strengths), the orientation of molecular vectors such as amide and carbonyl bonds can be determined independently relative to the magnetic susceptibility tensor of the molecule. The magnetic susceptibility is relatively large in aromatic molecules, particularly when ordered in a stacked manner. Thus nucleic acid molecules give efficient contributions to the measurability of dipolar couplings. The dipolar effects increase as the square of the magnetic field strength and can therefore be expected to become a more studied phenomenon as higher field NMR instruments become available. Dipolar couplings were measured and used to determine the structure of the DNA binding domain of the transcription factor GATA-1 bound to an oligonucleotide.[86] In a different study the addition of a filamentous bacteriophage was used to increase the degree of anisotropy in the sample and it was shown that in addition to heteronuclear dipolar couplings, $^{1}H-^{1}H$ dipolar couplings could also be measured.[37] These recent developments may present a novel approach to one of the holy grails of DNA structures – the overall shape and the extent of any sequence-dependent bend of a molecule – which traditional methods that rely on short-distance constraints that easily accumulate small errors have failed to answer.

ACKNOWLEDGMENTS

I am grateful to Dr. Asif Suri and Mr. Tommi Ratilainen for their assistance in figure preparation, and to Dr. Bengt Nordén for valuable comments on the manuscript.

ABBREVIATIONS AND ACRONYMS

AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
BDNOESY	Band Decoupled Nuclear Overhauser Spectroscopy
CBDNOESY	Complementary Band Decoupled Nuclear Overhauser Spectroscopy
COSY	Correlation Spectroscopy
DQF	Double Quantum Filtered
HMG	High-mobility Group
HMQC	Homonuclear Multiple Quantum Correlation Spectroscopy
HPLC	High-performance Liquid Chromatography
HSQC	Homonuclear Single Quantum Correlation Spectroscopy
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Spectroscopy
NTP	Nucleotide 5′-triphosphate
PCR	Polymerase Chain Reaction
ROESY	Nuclear Overhauser Enhancement Spectroscopy in the Laboratory Frame
TOCSY	Total Correlation Spectroscopy
TROSY	Transverse Relaxation Optimized Spectroscopy
UTR	Untranslated Region
2D	Two-dimensional
3D	Three-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance in Medicine, High Resolution Ex Vivo • Magnetic Resonance, General Medical • Multinuclear Magnetic Resonance Spectroscopic Imaging • Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Circular Dichroism in Analysis of Biomolecules • Nuclear Magnetic Resonance of Biomolecules

Chemical Weapons Chemicals Analysis **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Nucleic Acid Analysis in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring

Food **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Aptamers • DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Nucleic Acid

Structural Energetics • PNA and Its Applications • RNA Tertiary Structure • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Structural Analysis of Ribozymes • X-ray Structures of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Protein–Oligonucleotide Interactions • X-ray Crystallography of Biological Macromolecules

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis • Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance • High-performance Liquid Chromatography Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. F.J.M. Van de Ven, C.W. Hilbers, 'Nucleic Acids and Nuclear Magnetic Resonance', *Eur. J. Biochem.*, **178**, 1–38 (1988).
2. D.J. Patel, L. Shapiro, D. Hare, 'DNA and RNA: NMR Studies of Conformations and Dynamics in Solution', *Q. Rev. Biophys.*, **20**, 35–112 (1987).
3. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
4. X.J. Lu, W.K. Olson, 'Resolving the Discrepancies among Nucleic Acid Conformational Analyses', *J. Mol. Biol.*, **285**, 1563–1575 (1999).
5. F.H.-T. Allain, C. Gubser, P.W.A. Howe, K. Nagai, D. Neuhaus, G. Varani, 'Specificity of Ribonucleoprotein Interaction Determined by RNA Folding During Complex Formation', *Nature*, **380**, 646–650 (1996).
6. M. Eriksson, P.E. Nielsen, 'Solution Structure of a Peptide Nucleic Acid–DNA Duplex', *Nat. Struct. Biol.*, **3**, 410–413 (1996).
7. E.P. Nikonowicz, A. Sirr, P. Legault, F.M. Jucker, L.M. Baer, A. Pardi, 'Preparation of ^{13}C and ^{15}N Labelled RNAs for Heteronuclear Multi-dimensional NMR Studies', *Nucleic Acids Res.*, **20**, 4507–4513 (1992).
8. R.T. Batey, M. Inada, E. Kujawinski, J.D. Puglisi, J.R. Williamson, 'Preparation of Isotopically Labeled Ribonucleotides for Multidimensional NMR Spectroscopy of RNA', *Nucleic Acids Res.*, **20**, 4515–4523 (1992).
9. M.J. Michnicka, J.W. Harper, G.C. King, 'Selective Isotopic Enrichment of Synthetic RNA: Application to the HIV-1 TAR Element', *Biochemistry*, **32**, 395–400 (1993).
10. J.M. Louis, R.G. Martin, G.M. Clore, A.M. Gronenborn, 'Preparation of Uniformly Isotope-labeled DNA Oligonucleotides for NMR Spectroscopy', *J. Biol. Chem.*, **273**, 2374–2378 (1998).
11. D.P. Zimmer, D.M. Crothers, 'NMR of Enzymatically Synthesized Uniformly $^{13}C^{15}N$-labeled DNA Oligonucleotides', *Proc. Natl. Acad. Sci. USA*, **92**, 3091–3095 (1995).
12. J.E. Masse, P. Bortmann, T. Dieckman, J. Feigon, 'Simple, Efficient Protocol for Enzymatic Synthesis of Uniformly ^{13}C,^{15}N-labeled DNA for Heteronuclear NMR Studies', *Nucleic Acids Res.*, **26**, 2618–2624 (1998).
13. D.E. Smith, J.-Y. Su, F.M. Jucker, 'Efficient Enzymatic Synthesis of ^{13}C, ^{15}N-labeled DNA for NMR Studies', *J. Biomol. NMR*, **10**, 245–253 (1997).
14. A. Ono, S.-I. Tate, Y. Ishido, M. Kainosho, 'Preparation and Heteronuclear 2D NMR Spectroscopy of a DNA Dodecamer Containing a Thymidine Residue with a Uniformly ^{13}C-labeled Deoxyribose Ring', *J. Biomol. NMR*, **4**, 581–586 (1994).
15. G. Mer, W.J. Chazin, 'Enzymatic Synthesis of Region-specific Isotope-labeled DNA Oligomers for NMR Analysis', *J. Am. Chem. Soc.*, **120**, 607–608 (1998).
16. X. Huang, P. Yu, E. LeProust, X. Gao, 'An Efficient and Economic Site-specific Deuteration Strategy for NMR Studies of Homologous Oligonucleotide Repeat Sequences', *Nucleic Acids Res.*, **25**, 4758–4763 (1997).
17. A. Földesi, S.I. Yamakage, F.P.R. Nilsson, T.V. Maltseva, J. Chattopadhyaya, 'The Use of Non-uniform Deuterium Labelling ('NMR-window') to Study the NMR Structure of a 21mer RNA Hairpin', *Nucleic Acids Res.*, **24**, 1187–1194 (1996).
18. A. Földesi, T.V. Maltseva, Z. Dinya, J. Chattopadhyaya, 'Synthesis of Partially-deuterated 2′Deoxyribonucleoside Blocks and their Incorporation into an Oligo-DNA for Simplification of Overcrowding and Selective Enhancement of Resolution and Sensitivity in the ^{1}H-NMR Spectra', *Tetrahedron*, **54**, 14 487–14 514 (1998).

19. T.V. Maltseva, A. Földesi, J. Chattopadhyaya, 'T_1 and T_2 Relaxations of the ^{13}C Nuclei of Deuterium-labeled Nucleosides', *Magn. Reson. Chem.*, **36**, 227–239 (1998).
20. K. Wüthrich, *NMR of Protein and Nucleic Acids*, Wiley, New York, 1986.
21. L.J. Rinkel, C. Altona, 'Conformational Analysis of the Deoxyribofuranose Ring in DNA by Means of Sums of Proton–Proton Coupling Constants: A Graphical Method', *J. Biomol. Struct. Dyn.*, **4**, 621–649 (1987).
22. M.A. Weiss, D.J. Patel, R.T. Sauer, M. Karplus, '^{1}H-NMR Study of the Lambda Operator Site OL1: Assignment of the Imino and Adenine H2 Resonances', *Nucleic Acids Res.*, **12**, 4035–4047 (1984).
23. A. Pardi, 'Multidimensional Heteronuclear NMR Experiments for Structure Determination of Isotopically Labeled RNA', *Methods Enzymol.*, **261**, 350–380 (1995).
24. A. Pardi, E.P. Nikonowicz, 'Simple Procedure for Resonance Assignment of the Sugar Protons in ^{13}C-labeled RNAs', *J. Am. Chem. Soc.*, **114**, 9202–9203 (1992).
25. V. Sklenar, R.D. Peterson, M.R. Rejante, E. Wang, J. Feigon, 'Two-dimensional Triple-resonance HCNCH Experiment for Direct Correlation of Ribose H1′ and Base H8, H6 Protons in ^{13}C,^{15}N-labeled RNA Oligonucleotides', *J. Am. Chem. Soc.*, **115**, 12 181–12 182 (1993).
26. B.T. Farmer II, L. Müller, E.P. Nikonowicz, A. Pardi, 'Unambiguous Resonance Assignments in ^{13}C,^{15}N-labeled Nucleic Acids by 3D Triple-resonance NMR', *J. Am. Chem. Soc.*, **115**, 11 040–11 041 (1993).
27. J.P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D.M. Crothers, C. Griesinger, 'Sequential Correlation of Anomeric Ribose Protons and Intervening Phosphorus in RNA Oligonucleotides by a ^{1}H, ^{13}C, ^{31}P Triple Resonance Experiment: HCP-CCH-TOCSY', *J. Biomol. NMR*, **5**, 87–92 (1995).
28. R. Fiala, F. Jiang, D.J. Patel, 'Direct Correlation of Exchangeable and Nonexchangeable Protons on Purine Bases in ^{13}C,^{15}N-labeled RNA Using a HCCNH-TOCSY Experiment', *J. Am. Chem. Soc.*, **118**, 689–690 (1996).
29. J.-P. Simorre, G.R. Zimmermann, A. Pardi, B.T. Farmer II, L. Müller, 'Triple Resonance HNCCCH Experiments for Correlating Exchangeable and Nonexchangeable Cytidine and Uridine Base Protons in RNA', *J. Biomol. NMR*, **6**, 427–432 (1995).
30. J.-P. Simorre, G.R. Zimmermann, L. Müller, A. Pardi, 'Triple-resonance Experiments for Assignment of Adenine Base Resonances in ^{13}C/^{15}N-labeled RNA', *J. Am. Chem. Soc.*, **118**, 5316–5317 (1996).
31. J.-P. Simorre, G.R. Zimmermann, L. Müller, A. Pardi, 'Correlation of the Guanosine Exchangeable and Nonexchangeable Base Protons in ^{13}C/^{15}N-labeled RNA with an HNC-TOCSY-CH Experiment', *J. Biomol. NMR*, **7**, 153–156 (1996).
32. E.P. Nikonowicz, A. Pardi, 'An Efficient Procedure for Assignment of the Proton, Carbon and Nitrogen Resonances in ^{13}C/^{15}N-labeled Nucleic Acids', *J. Mol. Biol.*, **232**, 1141–1156 (1993).
33. U. Schmitz, G. Zon, T.L. James, 'Deoxyribose Conformation in [d(GTATATAC)]$_2$: Evaluation of Sugar Pucker by Simulation of Double-quantum-filtered COSY Cross-peaks', *Biochemistry*, **29**, 2357–2368 (1990).
34. H. Schwalbe, J.P. Marino, G.C. King, W.B. Weschelberger, C. Griesinger, 'Determination of a Complete Set of Coupling Constants in ^{13}C-labeled Oligonucleotides', *J. Biomol. NMR*, **4**, 631–644 (1994).
35. G.M. Clore, E.C. Murphy, A.M. Gronenborn, A. Bax, 'Determination of Three-bond ^{1}H3′–^{31}P Couplings in Nucleic Acids and Protein Nucleic Acid Complexes by Quantitative *J* Correlation Spectroscopy', *J. Magn. Reson.*, **134**, 164–167 (1998).
36. N. Tjandra, A. Bax, 'Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium', *Science*, **278**, 1111–1114 (1997).
37. M.R. Hansen, L. Mueller, A. Pardi, 'Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions', *Nat. Struct. Biol.*, **5**, 1065–1074 (1998).
38. J.L. Leroy, D. Broseta, M. Gueron, 'Proton Exchange and Base Pair Kinetics of Poly(rA)·Poly(rU) and Poly(rI)·Poly(rC)', *J. Mol. Biol.*, **184**, 165–178 (1985).
39. U. Dornberger, M. Leijon, H. Fritzche, 'High Base Pair Opening Rates in Tracts of GC Base Pairs', *J. Biol. Chem.*, **274**, 6957–6962 (1999).
40. E. Liepinsh, G. Otting, K. Wüthrich, 'NMR Observation of Individual Molecules of Hydration Water Bound to DNA Duplexes: Direct Evidence for a Spine of Hydration Water Present in Aqueous Solution', *Nucleic Acids Res.*, **20**, 6549–6553 (1992).
41. M.G. Kubinec, D.E. Wemmer, 'NMR Evidence for DNA Bound Water in Solution', *J. Am. Chem. Soc.*, **114**, 8739–8740 (1992).
42. A.J. Dingley, S. Grzesiek, 'Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide $^2J_{NN}$ Couplings', *J. Am. Chem. Soc.*, **120**, 8293–8297 (1998).
43. A. Majumdar, A. Kettani, E. Skripkin, 'Observation and Measurement of Internucleotide $^2J_{NN}$ Coupling Constants Between ^{15}N Nuclei with Widely Separated Chemical Shifts', *J. Biomol. NMR*, **14**, 67–70 (1999).
44. A.T. Brünger, *X-plor Manual V. 3.1*, The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.
45. P. Schultze, J. Feigon, 'Chirality Errors in Nucleic Acid Structures', *Nature*, **387**, 668 (1997).
46. M. Tonelli, T.L. James, 'Insights into the Dynamic Nature of DNA Duplex Structure via Analysis of Nuclear Overhauser Intensities', *Biochemistry*, **37**, 11 478–11 487 (1998).
47. C. Kojima, A. Ono, M. Kainosho, T.L. James, 'DNA Duplex Dynamics: NMR Relaxation Studies of a Decamer with Uniformly ^{13}C-labeled Purine Nucleotides', *J. Magn. Reson.*, **135**, 310–333 (1998).

48. L. Shapiro, M. Nilges, M. Eriksson, 'NMR Relaxation Matrix Refinement of a DNA Octamer Solution Structure', *Acta Chem. Scand.*, **47**, 43–56 (1993).
49. C.G. Hoogstraten, W.M. Westler, S. Macura, J.L. Markley, 'Improved Measurement of Longer Proton–Proton Distances in Proteins by Relaxation Network Editing', *J. Magn. Res., Series B*, **102**, 232 (1993).
50. C.G. Hoogstraten, A. Pardi, 'Improved Distance Analysis in RNA Using Network-editing Techniques for Overcoming Errors due to Spin Diffusion', *J. Biomol. NMR*, **11**, 85–95 (1998).
51. H.-T. Allain, G. Varani, 'How Accurately and Precisely Can RNA Structure be Determined by NMR?', *J. Mol. Biol.*, **267**, 338–351 (1997).
52. I. Radhakrishnan, D.J. Patel, 'DNA Triplexes: Solution Structures, Hydration Sites, Energetics, Interactions, and Function', *Biochemistry*, **33**, 11 405–11 416 (1994).
53. P. Rajagopal, J. Feigon, 'NMR Studies of Triple-strand Formation from Homopurine–Homopyrimidine Deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$', *Biochemistry*, **28**, 7859–7870 (1989).
54. C. de los Santos, M. Rosen, D.J. Patel, 'NMR Studies of DNA $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ Triple Helices in Solution: Imino and Amino Proton Markers of T·A·T and $C\cdot G\cdot C^+$ Base-triple Formation', *Biochemistry*, **28**, 7282–7289 (1989).
55. I. Radhakrishnan, X. Gao, C. de los Santos, D. Live, D.J. Patel, 'NMR Structural Studies of Intramolecular $(Y+)_n\cdot(R+)_n(Y-)_n$ DNA Triplexes in Solution: Imino and Amino Proton and Nitrogen Markers of G·TA Base Triple Formation', *Biochemistry*, **30**, 9022–9030 (1991).
56. I. Radhakrishnan, D.J. Patel, X. Gao, 'Three-dimensional Homonuclear NOESY–TOCSY of an Intermolecular Pyrimidine·Purine·Pyrimidine DNA Triplex Containing a Central G·TA Triple: Nonexchangeable Proton Assignments and Structural Implications', *Biochemistry*, **31**, 2514–2523 (1992).
57. I. Radhakrishnan, D.J. Patel, 'Hydration Sites in Purine·Purine·Pyrimidine and Pyrimidine·Purine·Pyrimidine DNA Triplexes in Aqueous Solution', *Structure*, **2**, 395–405 (1994).
58. D. Rhodes, R. Giraldo, 'Telomere Structure and Function', *Curr. Opin. Struct. Biol.*, **5**, 311–322 (1995).
59. J.M. Darlow, D.R. Leach, 'Secondary Structures in d(CGG) and d(CCG) Repeat Tracts', *J. Mol. Biol.*, **275**, 3–16 (1998).
60. Y. Wang, D.J. Patel, 'Guanine Residues in $d(T_2AG_3)$ and $d(T_2G_4)$ form Parallel-stranded Potassium Cation Stabilized G-quadruplexes with anti Glycosidic Torsion Angles in Solution', *Biochemistry*, **31**, 8112–8119 (1992).
61. A. Kettani, S. Bouaziz, A. Gorin, H. Zhao, R.A. Joners, D.J. Patel, 'Solution Structure of a Na Cation Stabilized DNA Quadruplex Containing G·G·G·G and G·C·G·C Tetrads formed by G-G-G-C Repeats Observed in Adeno-associated Viral DNA', *J. Mol. Biol.*, **282**, 619–636 (1998).
62. Y. Wang, C. de los Santos, X. Gao, K. Greene, D. Live, D.J. Patel, 'Multinuclear Nuclear Magnetic Resonance Studies of Na Cation-stabilized Complex Formed by d(G-G-T-T-T-T-C-G-G) in Solution. Implications for G-tetrad Structures', *J. Mol. Biol.*, **222**, 819–832 (1991).
63. S. Bouaziz, A. Kettani, D.J. Patel, 'A K Cation-induced Conformational Switch Within a Loop Spanning Segment of a DNA Quadruplex Containing G-G-G-G Repeats', *J. Mol. Biol.*, **282**, 637–652 (1998).
64. N.V. Hud, P. Schultze, J. Feigon, 'Ammonium Ion as an NMR Probe for Monovalent Cation Coordination Sites of DNA Quadruplexes', *J. Am. Chem. Soc.*, **120**, 6403–6404 (1998).
65. M. Eriksson, M. Leijon, C. Hiort, B. Nordén, A. Gräslund, 'Binding of Δ- and Λ-$[Ru(phen)_3]^{2+}$ to $[d(CGCGAT CGCG)]_2$ studied by NMR', *Biochemistry*, **33**, 5031–5040 (1994).
66. J.J. Love, X. Li, D.A. Case, K. Giese, R. Grosschedl, P.E. Wright, 'Structural Basis for DNA Bending by the Architectural Transcription Factor LEF-1', *Nature*, **376**, 791–795 (1995).
67. C.C. Gubser, G. Varani, 'Structure of the Polyadenylation Regulatory Element of the Human U1A pre-mRNA 3′-untranslated Region and Interaction with the U1A Protein', *Biochemistry*, **35**, 2253–2267 (1996).
68. F.M. Jucker, H.A. Heus, P.F. Yip, E.H.M. Moors, A. Pardi, 'A Network of Heterogeneous Hydrogen Bonds in GNRA Tetraloops', *J. Mol. Biol.*, **264**, 968–980 (1996).
69. D.J. Patel, A.K. Suri, F. Jiang, L. Jiang, P. Fan, R.A. Kumar, S. Nonin, 'Structure, Recognition and Adaptive Binding in RNA Aptamer Complexes', *J. Mol. Biol.*, **272**, 645–664 (1997).
70. M. Sassanfar, J.W. Szostak, 'An RNA Motif that Binds ATP', *Nature*, **364**, 550–553 (1993).
71. F. Jiang, A.R. Kumar, R.A. Jones, D.J. Patel, 'Structural Basis of RNA Folding and Recognition in an AMP–RNA Aptamer Complex', *Nature*, **382**, 183–186 (1996).
72. F. Jiang, R. Fiala, D. Live, R.R. Kumar, D.J. Patel, 'RNA Folding Topology and Intermolecular Contacts in the AMP–RNA Aptamer Complex', *Biochemistry*, **35**, 13 250–13 266 (1996).
73. T. Dieckman, E. Suzuki, G.K. Nakamura, J. Feigon, 'Solution Structure of an ATP-binding RNA Aptamer Reveals a Novel Fold', *RNA*, **2** (1996).
74. T. Dieckman, J. Feigon, 'Assignment Methodology for Larger RNA Oligonucleotides: Application to an ATP-binding RNA Aptamer', *J. Biomol. NMR*, **9**, 259–272 (1997).
75. F. Jiang, D.J. Patel, X. Zhang, H. Zhao, R.A. Jones, 'Specific Labeling Approaches to Guanine and Adenine Imino and Amino Proton Assignments in the AMP–RNA Aptamer Complex', *J. Biomol. NMR*, **9**, 55–62 (1997).
76. F.H.-T. Allain, G. Varani, 'Structure of the P1 Helix from Group I Self Splicing Introns', *J. Mol. Biol.*, **250**, 333–353 (1995).

77. S. Nonin, F. Jiang, D.J. Patel, 'Imino Proton Exchange and Base-pair Kinetics in the AMP–RNA Aptamer Complex', *J. Mol. Biol.*, **268**, 359–374 (1997).
78. P.W.A. Howe, F.H-T. Allain, G. Varani, D. Neuhaus, 'Determination of the NMR Structure of the Complex Between U1A Protein and its Polyadenylation Inhibition Element', *J. Biomol. NMR*, **11**, 59–84 (1998).
79. G. Otting, K. Wüthrich, 'Heteronuclear Filters in Two-dimensional [^{1}H,^{1}H]-NMR Spectroscopy: Combined Use with Isotope Labelling for Studies of Macromolecular Conformation and Intermolecular Interactions', *Q. Rev. Biophys.*, **23**, 39–96 (1990).
80. A.D. Mesmaeker, K.-H. Altmann, A. Waldner, S. Wendeborn, 'Backbone Modifications in Oligonucleotides and Peptide Nucleic Acid Systems', *Curr. Opin. Struct. Biol.*, **5** (1995).
81. M. Eriksson, P.E. Nielsen, 'PNA–Nucleic Acid Complexes. Structure, Stability and Dynamics', *Q. Rev. Biophys.*, **29**, 369–394 (1996).
82. M. Leijon, A. Gräslund, P.E. Nielsen, O. Buchardt, B. Nordén, S.M. Kristensen, M. Eriksson, 'Structural Characterization of PNA–DNA Duplexes by NMR. Evidence for DNA in a B-like Conformation', *Biochemistry*, **33**, 9820–9825 (1994).
83. L. Betts, J.A. Josey, J.M. Veal, S.R. Jordan, 'A Nucleic Acid Triple Helix Formed by a Peptide Nucleic Acid–DNA Complex', *Science*, **270**, 1838–1841 (1995).
84. S.C. Brown, S.A. Thomson, J.M. Veal, D.G. Davis, 'NMR Solution Structure of a Peptide Nucleic Acid Complexed with RNA', *Science*, **265**, 777–780 (1994).
85. K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, K. Wüthrich, 'NMR Scalar Coupling Across Watson–Crick Base Pair Hydrogen Bonds in DNA Observed by Transverse Relaxation-optimized Spectroscopy', *Proc. Natl. Acad. Sci. USA*, **95**, 14 147–14 151 (1998).
86. N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, 'Use of Dipolar ^{1}H–^{15}N and ^{1}H–^{13}C Couplings in the Structure Determination of Magnetically Oriented Macromolecules in Solution', *Nat. Struct. Biol.*, **4**, 732–738 (1997).

Nucleic Acid Structural Energetics

Charles C. Hardin and Andrea F. Moon[a]
North Carolina State University, Raleigh, North Carolina, USA

1 Introduction **1**
2 Organization of Structural Hierarchy **1**
2.1 Components and Linkages 2
2.2 Secondary Structure 5
2.3 Superstructural Considerations 6
3 DNA Structure **7**
3.1 Duplex Structures 7
3.2 Triple Helical Structures 8
3.3 Four-stranded Structures 9
3.4 Higher Order Structures 11
4 RNA Structure **12**
4.1 Duplex Structures 12
4.2 Substructural Components 12
4.3 Higher Order RNA Structures 14
5 Energetics **17**
5.1 Thermodynamic Characterization 17
5.2 Calorimetric Determination of Enthalpies 19
5.3 Spectroscopic Techniques 21
5.4 Electrophoretic Techniques 24
6 Condition-dependent Structural Transitions **25**
6.1 Cation Dependences 25
6.2 Ligand-induced Structural Transformations 26
Acknowledgment **27**
Abbreviations and Acronyms **27**
Related Articles **27**
Further Reading **29**
Books **29**
Primary Literature and Reviews **29**
References **32**

The general relations between structure and energetics of nucleic acids are described. A detailed, multileveled definition of the structures adopted by one, two, three and four strands is presented, with emphasis on the chemical functionalities of the structural subcomponents. Descriptions include results from studies of the thermodynamics of transformations and structures (e.g. double helix formation, superhelical coiling). Different aspects of this large array of problems have been determined using X-ray diffraction and spectroscopic techniques, such as ultraviolet (UV) and infrared (IR) absorption, circular dichroism (CD), Raman scattering and nuclear magnetic resonance (NMR). Interactions, within one strand or between two or more strands, with cations, buffers, aromatic chemicals and proteins are all described. Methods include spectroscopy, calorimetry and computational approaches (statistical mechanics, molecular mechanics, dynamics, quantum mechanics). A Further Reading section is included to provide an idea of the vast scope of this area.

1 INTRODUCTION

The foundation of life itself is encoded by a vast array of nucleic acids. The functional capability of a particular nucleic acid in its biological niche is determined by the viability of its physical structure. As with all chemical species, the assumed structure is a compromise between stabilizing and repulsive interactions between and among the functional subcomponents. The relations between structure, function and energetics are complicated, and therefore require a detailed, multileveled definition. The structures of nucleic acids are so varied and numerous that we can only summarize a broad selection of the key areas. A much less inclusive selection of the associated energetics is discussed, only providing an overview of general approaches. The cited literature expands upon the methodology, assumptions and interpretations, presenting a reasonably broad, but by no means exhaustive, perspective of the subject.

2 ORGANIZATION OF STRUCTURAL HIERARCHY

Nucleic acid structure are classified in a hierarchical manner, beginning with the simplest components and proceeding to progressively larger assemblies. Associations between two or more simple subcomponents produce secondary and tertiary structures. We use the term ''superstructure'' to define the even larger functional domains occupied by nucleic acid components in more complex assemblies (e.g. supercoiled chromatin, virus capsids).

[a] Present address: National Institutes of Environmental Health Sciences, Research Triangle Park, North Carolina, USA.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2.1 Components and Linkages

The basic structural components of nucleic acids include the sugars ribose in ribonucleic acid (RNA) and deoxyribose in deoxyribonucleic acid (DNA). Both polymers are composed of two or more nucleotides connected by phosphates, which are connected by sugars ribose and deoxyribose, respectively. Each sugar has one attached nucleobase (Figure 1a–d) which connects noncovalently with other strands through base pairing mediated by interbase hydrogen bonds (Figures 2a and b, 3 and 4a–c).[1] The 2′ carbon of deoxyribose does not contain

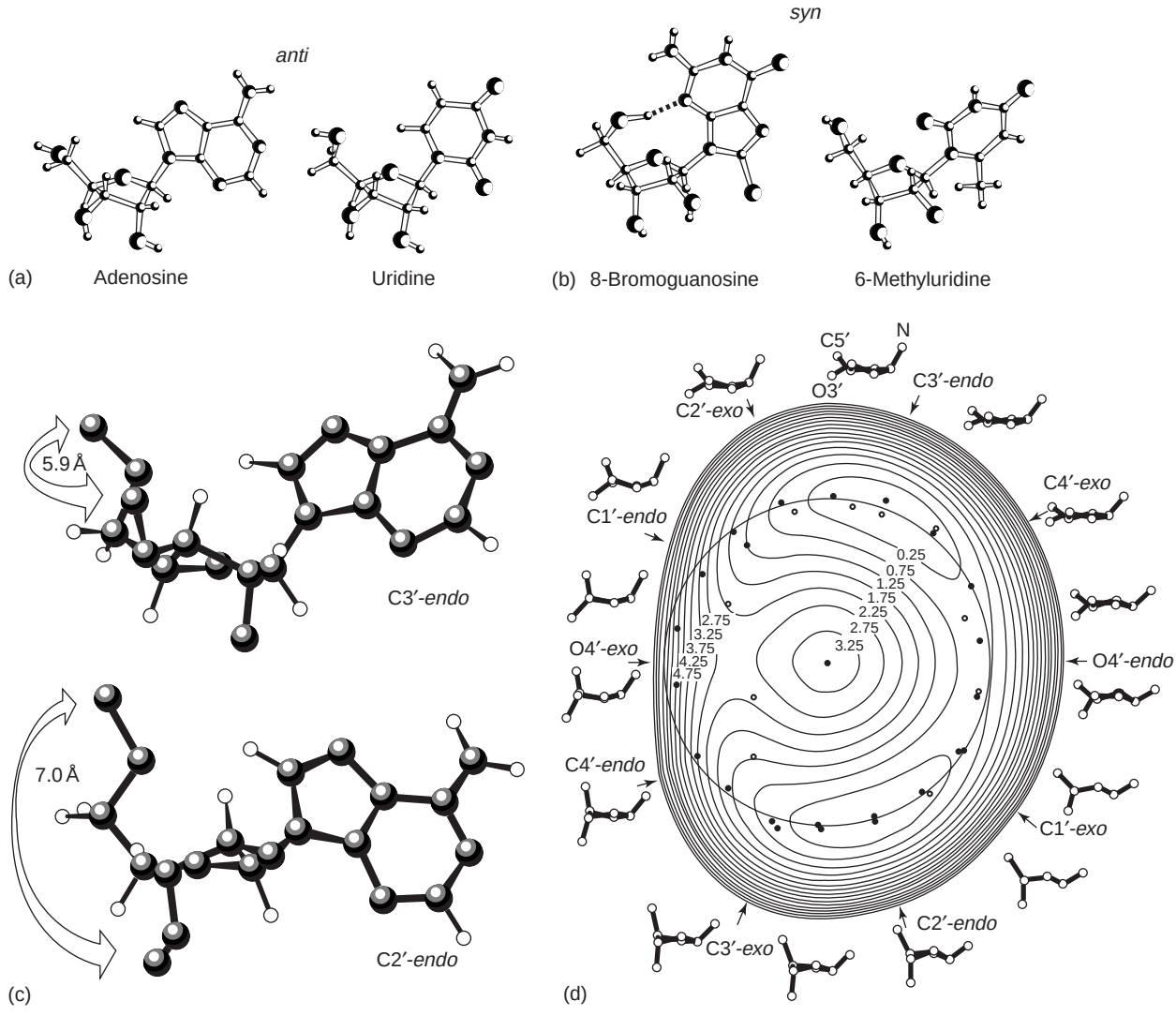

Figure 1 Geometries of a nucleosides in the *syn* and *anti* orientation. (a) Adenosine and uridine are in the *anti* conformation. (b) 8-Bromoguanosine (br^8G) and 6-methyluridine are *syn* owing to their bulky substituents *ortho* to the glycosidic bond. In br^8G, an intramolecular −O5′H•N3− hydrogen bond stabilizes the *syn* conformation. The C3′-*endo* sugar puckering is preferred by the *anti* conformation; C2′-*endo* is preferred by *syn* nucleotides. Spheres of increasing size represent H, C, N, O atoms; atomic positions obtained from crystal structure analysis. (c) Nucleotides in the C3′-*endo* and C2′-*endo* sugar conformations preferred in A- and B-form DNAs, respectively. The distances between adjacent P atoms in the sugar–phosphate backbone are indicated. (d) Energy map of furanose conformations as a function of pseudorotation phase angle (P). The latter is counted in a clockwise sense with $P = 0°$ in a vertical position. The puckering amplitude (q) corresponds to ν_{max} and increases radially from the central dot representing a planar furanose. The ring indicates the path of true pseudorotation with $q = 0.4$ Å; filled circles describe energies calculated when all five torsion angles ν_n are constrained to the path of pseudorotation; empty circles correspond to only ν_3 constrained and other torsion angles relaxed. Note large differences at $P \approx 270°$ (O4′-*exo* puckering). Drawn molecules show relaxed geometries and belong to nearest neighbor circles in the energy plot. The nomenclature and orientation are given on the molecule $P = 0$ (top of circle). Numbers on the diagram indicate energies (kcal mol^{-1}) above global minimum. (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 21, 63, 1984, Copyright Springer-Verlag 1984, and of John Wiley and Sons from D. Voet, J.G. Voet, *Biochemistry*, John Wiley & Sons, New York, 808, 1990.)

a hydroxyl group. This distinction separates the natural nucleic acids into the two major classes, RNA and DNA. The 2′-OH group in RNAs participates in alkali-catalyzed hydrolytic chain breakage, when the hydroxide attacks and binds the adjacent 3′ phosphodiester phosphorus and eliminates the 3′-terminal chain. In contrast, DNA, which has a hydrogen at the 2′ carbon, is relatively resistant to chain breakage. As a result, the transient functions of RNA rely on its natural tendency toward breakdown.

RNA functions to provide a source of encoded ''molecular information'' made from the genomic DNA via transcription. Messenger RNA (mRNA) is used to make protein during translation (''gene expression''). The tendency of RNAs to break down allows cells to reabsorb their used mRNAs, leading to genetic ''information turnover''. Many, many other functions of RNA have been described.[2,3] In contrast, DNA functions as a reservoir of relatively stable stored information (like computer permanent memory, ROM) which is decoded during transcription to produce RNA. DNA also participates in replication, to make a daughter genome from the copy provided by the maternal genome, recombination (to swap and reassociate DNA fragments) and many other aspects of genetic metabolism.

The term ''glycosidic bond'' refers to the covalent linkage between the C1′ carbon of the sugar and N1 nitrogen of pyrimidines or N9 of purines, the heterocyclic nucleobases (Figure 2a and b). Sugar ''pucker'' variants of the pentose rings include C2′-*endo* and -*exo*, and C3′-*endo* and -*exo* (Figure 1b and c). These conformations can interconvert by the pseudorotation process, in which one atom is viewed as being out of the plane formed by the other four atoms (similar to the appearance of the corner of an envelope) then another atom becomes the nonplanar one, and so on to create a cycle of stepwise conformational exchanges about the periphery of the sugar ring.

The C2′-*endo* pucker projects the atom out of the plane of the other ring carbons toward the nucleobase. This pucker variant is favored in the structure of B-DNA (Figure 4a) by a few kilojoules per mole over the other predominant pseudorotation variant, the C3′-*endo* sugar pucker. The C3′-*endo* sugar pucker projects the C3′ carbon out of the plane toward the nucleobase (Figure 1c and d). This conformation occurs in the A-form (Figure 4b) structures adopted by both DNA and RNA. Both types of pucker variants are found within the zig-zagged backbone of Z-DNA (Figure 4c).

The sugar residues of nucleic acids are often modified. For example, in 2′-*O*-methylated ribose, a methyl group forms an ether linkage, sharing the hydroxylate oxygen with C2′. Sugar modifications are found in transfer RNAs (tRNAs)[1] and in the C/D box of small nucleolar RNA (snoRNA).[2] The snoRNAs are used to specify sites for enzymatic methylation of nucleolar rRNAs to the methylase(s).

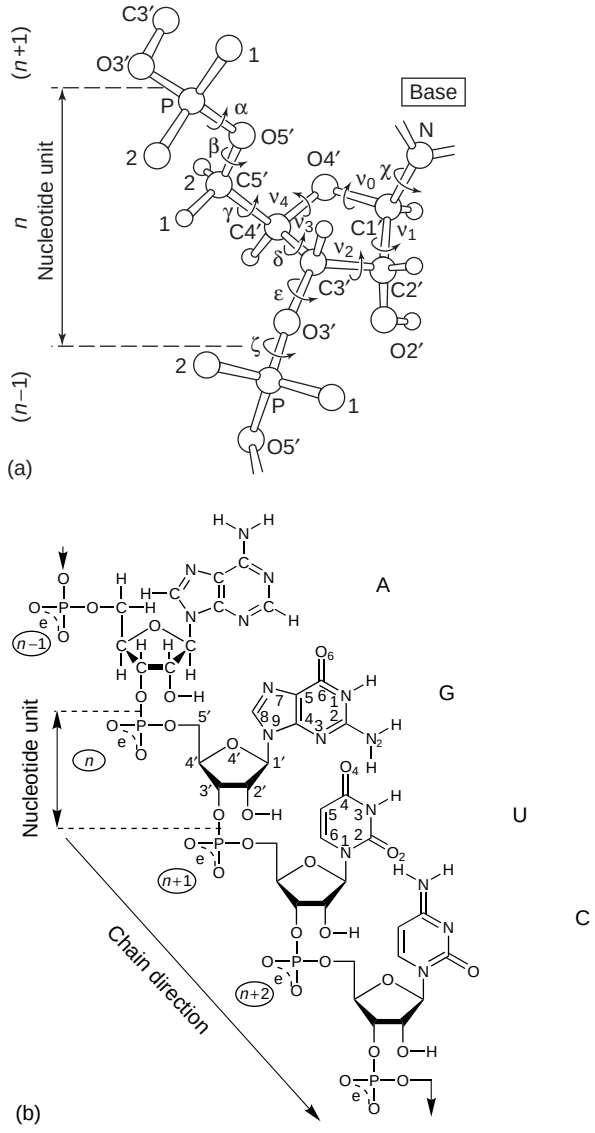

Figure 2 (a) Atomic numbering scheme and definition of torsion angles for a polyribonucleotide chain. Counting of nucleotide from (a) to (b), i.e. in the direction O5′ → O3′. Hydrogens at C5′ and oxygens at P are differentiated by 1 and 2 according to the rule given in the text. In deoxyribose, the hydrogen replacing O2′ is labeled 1 and the other is 2. (b) Fragment of RNA with sequence adenosine (A), guanosine (G), uridine (U) and cytidine (C) linked by 3′,5′-phosphodiester bonds. Chain direction is from 5′- to 3′-end as shown by arrow. Atom numbering scheme is indicated in one framed nucleotide unit, 5′-GMP. All hydrogen atoms drawn in A and only functional hydrogens in other nucleotides. In short notation, this fragment would be pApGpUpCp or pAGUCp. In DNA, the hydroxyl attached to C2′ is replaced by hydrogen and uracil (U) by thymidine. (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 10, 14, 1984. Copyright Springer-Verlag 1984.)

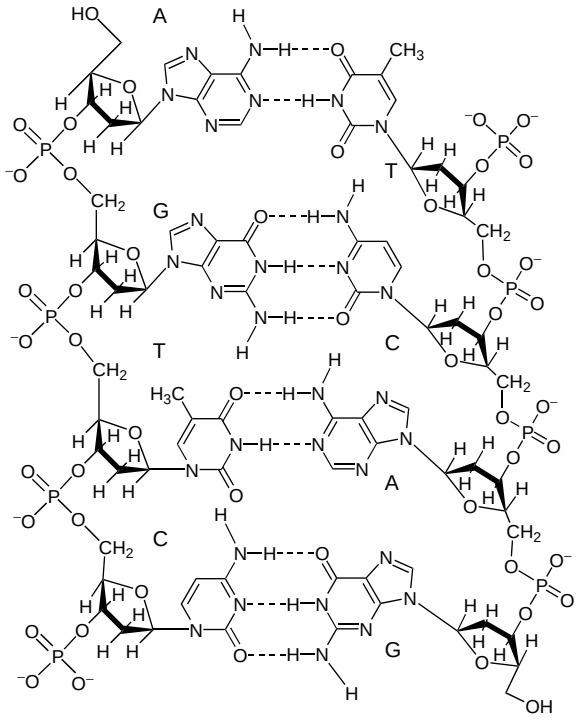

Figure 3 Base pairing in double-stranded DNA. The two strands run in opposite directions. Adenine (A) in one strand pairs with thymine (T) in the opposite strand and guanine (G) pairs with cytosine (C). In actuality, the base pairs are tilted so that the plane of the base pair is perpendicular to the page. (Adapted from Prentice-Hall from L. Moran, K. Scrimgeour, H. Horton, R. Ochs, J. Rawn, *Biochemistry*, Prentice-Hall, Upper Saddle River, NJ, 24–28, 1994. Copyright © 1994 Prentice-Hall, Inc.)

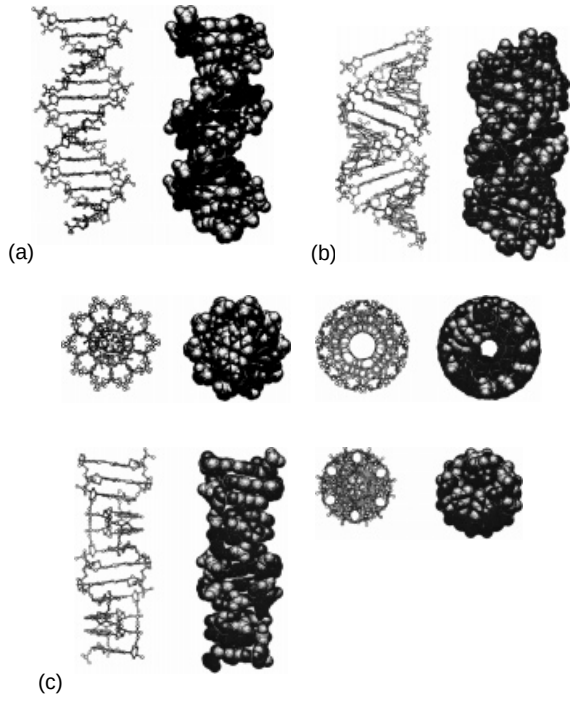

Figure 4 (a) Structure of B-DNA. Side and top view of B-DNA are shown in both in ball-and-stick and space filling representations. The helix is tilted 32° from the viewer to show minor (m) and major (M) grooves. (b) Side and top views of A-DNA in ball-and-stick and space filling representations. The helix is tilted about 30° from the viewer; atomic radii are C = 1.6 Å, N = 1.5 Å and O = 1.4 Å. Hydrogen atoms are omitted. A light source shines on the molecule under an angle of 45° from the (a); the deep, narrow major groove appears dark because it is in shadow. (c) Structure of left-handed, double-helical poly(dG–dC)•poly(dG–dC) drawn with coordinates for Z_I-DNA obtained with the hexanucleotide d(CGCGCG) (side and (a) views). (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 257, 262, 286, 1984. Copyright Springer-Verlag 1984.)

The second major structural component of nucleic acid structure is the phosphodiester backbone, which joins the nucleotides together (Figures 2–4). The 5′-hydroxylate of one nucleotide is covalently bound to the 3′-hydroxylate of the next ''upstream nucleotide'' in the sequence. By convention, sequences are always listed from the end with the free 5′-hydroxyl group (the 5′ terminus) to the 3′ terminus, the other end of the $(\text{phosphodiester})_i$ strand (where *i* indicates the number of nucleotide residues).

At physiological pH, the exposed hydroxyl groups of the phosphate backbone are deprotonated, producing a negatively charged backbone, with one anionic site per nucleotide. This polyanionic nature of nucleic acid structures plays a predominant role in the physical, chemical and functional properties of these biomacromolecules. Mono- and polyatomic cations bind to this feature of the external surfaces of nucleic acid structures, modulating their stabilities and sometimes inducing significant structural transformations.

Synthetic modifications of the phosphate backbone have been produced for biotechnological applications. For example, in phosphorothioates a sulfur atom is covalently linked by a double bond to the central phosphorus atom, replacing the carbonyl oxygen. As a result, the backbone is nuclease resistant and the nucleic acid has an extended functional lifetime. Methyl groups may also be covalently linked to the phosphate groups in place of the carbonyl oxygen. These methyl phosphate modifications also impart nuclease resistance.

The nucleobase sequence confers structural specificity and uniqueness to a given nucleic acid. There are two general classes of nucleobases, purines and pyrimidines. The bicyclic purines, adenine (A) and guanine (G), are found in both DNA and RNA. In canonical Watson–Crick (WC) base pairs, these purines form hydrogen bonds with the monocyclic pyrimidines – thymine (T) in DNA and uracil (U) in RNA – and cytosine (C) (Figure 3).

T and U differ only in the replacement of an exocyclic methyl group found in T by hydrogen.

Modified nucleobases occur in many nucleic acid species. The RNA nucleobase U is methylated to form the DNA base (T, really m^5U). The content of the developmentally important modified base 5-methylcytosine (m^5C) differs in the chromosomal DNAs of wild-type and diseased cells.[4] A more complicated modification, found in some bacterial virus DNAs, involves covalent attachment of a glucose ring to a C5-hydroxyl group on C. Another example involves enzymatic methylation of A N1, which is required to protect the duplex DNA from restriction enzyme cleavage.

Modified nucleobases are very common in natural RNAs, especially transfer RNA (tRNA).[1] For example, nucleobases are methylated at all ring positions. Two common RNA base variants, pseudouridine and dihydrouridine, occur in all tRNAs. Pseudouridine is a U that is attached to ribose at C5 instead of N1. Dihydrouridine is structurally similar to uridine, except the aromatic ring is saturated at the 5,6 π-bond, making dihydrouridine much more flexible than uridine. Dihydrouridine is always present in tRNA (at position 17) and common in small nuclear and nucleolar RNAs. More complex modifications, composed of isopropyl groups, amino acid side chains, and many others, have been described. A particularly useful example, the three-ringed G derivative wyebutine, allowed early researchers to isolate phenylalanyl transfer RNA ($tRNA^{Phe}$), by virtue of its hydrophobicity, and detect it, since the base is highly fluorescent. These complex functional groups can affect the kinetics of biological reactions by modifying the chemical and physical properties of the nucleic acids and, as a result, their propensity to move freely, bury themselves, become deprotonated, react with or shun other charged entities, etc.

2.2 Secondary Structure

The sequence of nucleobases is referred to as primary structure. Interactions between nucleotides within the same strand or another strand produce secondary structures. Natural DNA molecules most commonly consist of double-helical (duplex) strand variants (Figures 2–4). In contrast, RNAs usually form complex globular structures more like those found in proteins. RNA structures are typically comprised of stem–loop (hairpin) pairings, a common subdomain, formed when a strand pairs with itself and ''loops out'' a region (Figure 5a–d).

The primary energetic contributions to RNA folding result from the ''hydrophobic effect'', as found for globular proteins structures. However, ''despite the obvious analogies between the folding of proteins and RNA, key differences remain. An important factor in the folding of RNA is that condensation of a polyanion is necessarily sensitive to differences in ionic strength, as the electrostatic repulsion along the backbone provides a strong opposing force to the collapse of hydrophobic bases. Nearly all complex RNAs require Mg^{2+} for biological activity, both to stabilize the native structure and help mediate catalysis. Magnesium binding is cooperative and coincides with the formation of tertiary structure''.[5]

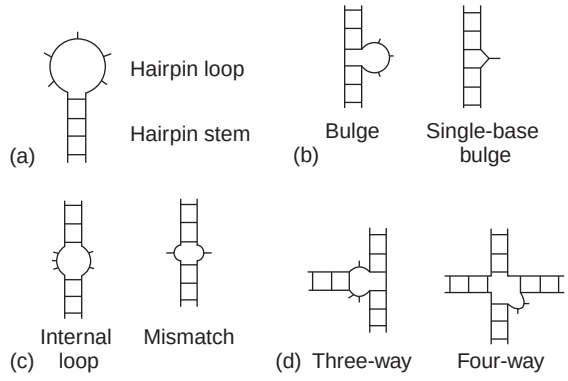

Figure 5 Secondary structure motifs. (a) Hairpin. A hairpin is formed when an RNA strand folds back on itself. The stem of the hairpin can consist of any number of WC base pairs. In principle, the loop may contain any number of unpaired nucleotides. The smallest loop thus far observed in RNA has three nucleotides. (b) Bulges. Unlike an internal loop, a bulge has unpaired nucleotides on only one strand. The other strand has uninterrupted base pairing. The base pairs on either side of a bulge may stack to form a regular helix. In a single-base bulge, only one unpaired nucleotide interrupts the paired region. (c) Internal loops. In an internal loop, at least one base is unpaired in each strand of the loop separating two unpaired regions. A mismatch is a special type of internal loop; only one nucleotide is unpaired on each strand. Any two bases may be in apposition. (d) Junctions. A junction occurs when double-stranded regions, separated by any number of unpaired nucleotides, come together. Three-stem junctions are the smallest; when two stems separated by unpaired nucleotides adjoin, either a bulge or internal loop is formed. Four-stem junctions are characteristic of tRNAs. Five stems are the most that adjoin in 16S ribosomal RNA. (Reproduced from J.R. Wyatt, J.D. Puglisi, I. Tinoco, Jr, *Bioessays*, **11**(4), 104 (1989). Copyright © 1989 John Wiley & Sons. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.)

Kinetic studies have suggested that RNA folding can occur sequentially, following a defined pattern. The ''folding funnel'' concept captures the idea that polymers fold by descending down an energy gradient.[5] Folding is depicted symbolically as one or more ''funnel spouts'' with different ''energy topologies''. The shape indicates the relative intrinsic propensities of chains to fold correctly, become trapped in an intermediate conformation(s), remain unfolded, etc.

Small proteins and other RNAs can enhance the rate of RNA folding processes and increase the net stability of

the folded domain. One very important class of stabilizing agents is the counterions, including mono- and divalent cations and organic nitrogen-based polycations, which offset the high anionic charges that would otherwise result in strand–strand repulsion.

2.3 Superstructural Considerations

The most common higher order structure assumed by DNA is the supercoil (Figure 6a and b). Supercoils are induced by and stabilized by negative superhelical torsional stress. This energy source helps, along with the histone proteins, to produce the high degree of compaction necessary to package the genomic DNA within the relatively small volume of the nucleus. Supercoiled structures are found in circular duplexes (i.e. plasmids) or in molecules that are rotationally constrained at both ends, as in the case of chromosomes in vivo, which are segregated into separate scaffold-bound subdomains. Another important class of contributors to the compacted

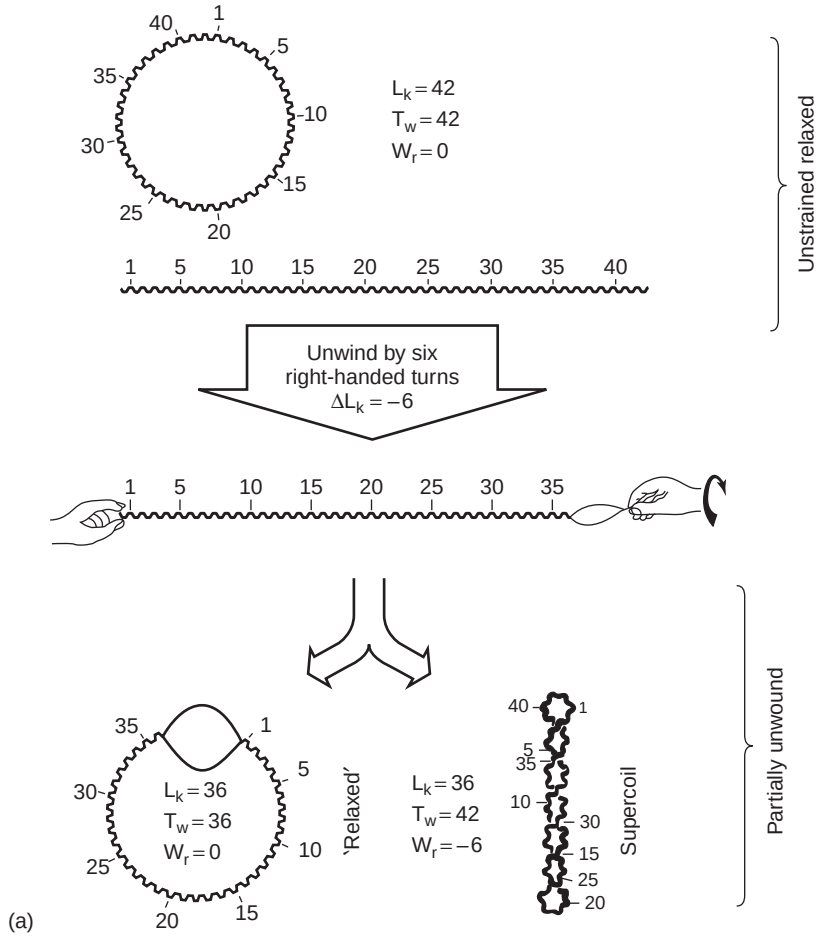

Figure 6 (a) Description of the interrelation between linking (L_k), twisting (T_w) and writhing (W_r). Unstrained, relaxed B-DNA with 420 base pairs has 42 turns of one strand wound about the other, $L_k = T_w = 42$ and $W_r = 0$ (top). If, now, the DNA is held tight at the left end, and the right end is rotated clockwise by six turns so the double helix is partially unwound, L_k and consequently T_w reduce to 36. This DNA can again be circularly closed with $L_k = T_w = 36$ and $W_r = 0$, i.e. compared with the above picture, base pairs of six turns are disrupted. Owing to the tendency of DNA to retain the B-DNA form, however, T_w increases to the original $T_w = 42$ with 420 base pairs. However, since L_k is a topological number, it is constant, $L_k = 36$ and [in order to satisfy Equation 19-1 in the original publication] the writhe must become $W_r = -6$, corresponding to a right-handed supercoil (superhelix) with six cross-overs. (b) Sedimentation rate of closed circular duplex DNA as a function of ethidium bromide concentration. Intercalation of ethidium between the base pairs unwinds the double helix. Since the linking number of the circle is constant, this is accompanied by an equivalent increase in the writhing number. As the superhelix unwinds, it becomes less compact and sediments more slowly. At the low point on the curve, the circular DNAs (plasmids) have bound sufficient ethidium to become fully relaxed. As the ethidium concentration is further increased, the DNA supercoils in the opposite direction. The supertwisted appearances in the depicted DNAs have been characterized by electron microscopy. (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 452, 1984. Copyright Springer-Verlag, 1984, and of John Wiley and Sons from D. Voet, J.G. Voet, *Biochemistry*, John Wiley & Sons, New York, 820, 1990.)

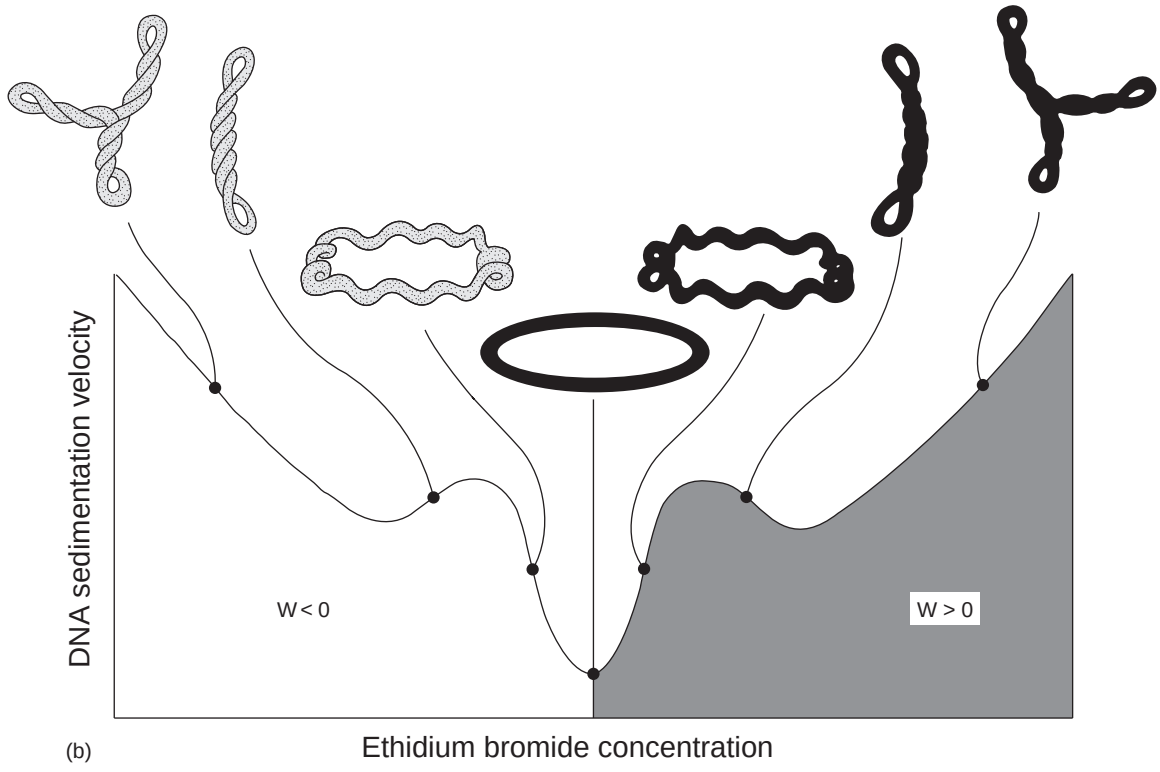

Figure 6 (*Continued*)

DNA structure found in the nucleus are the highly electropositive DNA-binding proteins, especially the histones around which the DNA wraps in approximately 200 base pair segments to form the ''beads on a string'' motif.[1–3] Thus, charge neutralization plays a crucial role in accommodating DNA in its natural niche.

3 DNA STRUCTURE

The delineation of the details of DNA and RNA structures has benefited tremendously from X-ray crystallographic techniques, in which highly ordered crystals are formed and illuminated with narrowly collimated X-ray light. The diffraction angles and intensities of spots produced by the light beam are used to calculate distances and angular relations between atoms in the molecular lattice. The crystals have been difficult to produce, so only limited information is available, especially for RNAs. Structures of a moderately large set of duplex DNAs have been determined best exemplified by the very well-resolved dodecamer fragment studied by Dickerson et al.[6] Structures of left-handed DNAs, four-stranded quadruplex DNAs, mismatched and bulged DNA duplexes, a few tRNAs,[1] Group I intron RNA fragments, hammerhead RNAs[7] and some non-WC (mismatched) base-pair-containing RNA duplexes are described below and in the Further Reading section.[8]

3.1 Duplex Structures

RNA and DNA molecules have very different intrinsic structural predispositions. The duplex structure occurs in most native DNAs and RNAs. DNAs are typically duplexes in long strands, whereas RNAs are usually formative subcomponents within more complex tertiary motifs.

The exact conformations adopted by strands in duplex DNA depend critically on sequence, salt and DNA concentrations, temperature and solution conditions (e.g. pH, presence and concentration of apolar or amphipathic solutes, ligands). The most common duplex conformation is the right-handed B-form (Figure 4a). The base pairs are centered on and rotate about the helical axis in a counterclockwise direction; the planes formed by the bases are perpendicular to this helical axis. B-form DNA is found in well-hydrated fiber preparations, as first discovered using X-ray diffraction results obtained by Franklin and interpreted by Watson and Crick.[1] This form predominates under physiological conditions. However, it is now known to occur in equilibrium with sub-helical domains consisting of the right-handed A-forms of DNA and conformations intermediate between

B- and A-forms (Figure 4a–c). Bases in the right-handed A-form helix are skewed slightly away from the helix axis and their planes are tilted about 18° away from it. A-form DNA is preferred in fibers that are less hydrated than those favoring the B-form.

At still lower relative humidity, alternating CG-block sequence DNAs readily adopt the left-handed Z conformation (Figure 4c).[9] This conformation is induced by torsional stress, rotating the nucleotides clockwise about the helical axis. Maintaining the antiparallel strand polarity twists the phosphodiester backbone, producing the characteristic zigzag pattern.

3.2 Triple Helical Structures

DNAs (and RNAs) also form triple-stranded (triplex) structures. Triplexes can be produced in supercoiled bacterial plasmid DNA of the correct sequence and in promoter sequences of eukaryotic genes.[10]

The triple helix is typically comprised of (pyrimidine)$_2$purine base triples in which the N7 face of the WC pyrimidine•purine base pair is associated with another pyrimidine residue (hydrogen bonds are symbolized using a centered bold dot, X•Y). The mode of pairing with the second pyrimidine residue is referred to as a "Hoogsteen interaction" (Figure 7a and b).[1,11] A number of compatible sequence variants have been described and the predisposition for triplex formation is very dependent on the length.

An interesting well-defined triplex structure is called H-DNA (Figures 8a and b and 9a).[11] This structure is characterized by mirror repeated oligopurine or oligopyrimidine sequences that loop a fragment of one strand out and coalesce together. One strand forms a stem–loop and the third of four segments binds to it; the fourth segment is the free loop. H-DNA only forms in extended regions of correct sequences and has been characterized by its susceptibility to S1 nuclease digestion. Although H-DNA maintains a triplex structure, it is comprised of two strands. When the mirror repeat segments coalesce, the axis of the

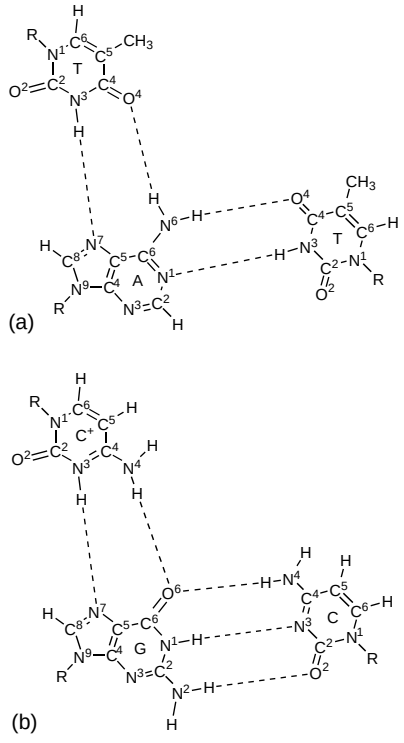

Figure 7 T•A•T and C•G•C_H^+ base triplets. (Reproduced with permission from S.M. Mirkin, M.D. Frank-Kamenetskii, *Annual Review of Biophysics and Biomolecular Structure*, Volume 23, 543, © 1994 by Annual Reviews, www.Annual Reviews.org.)

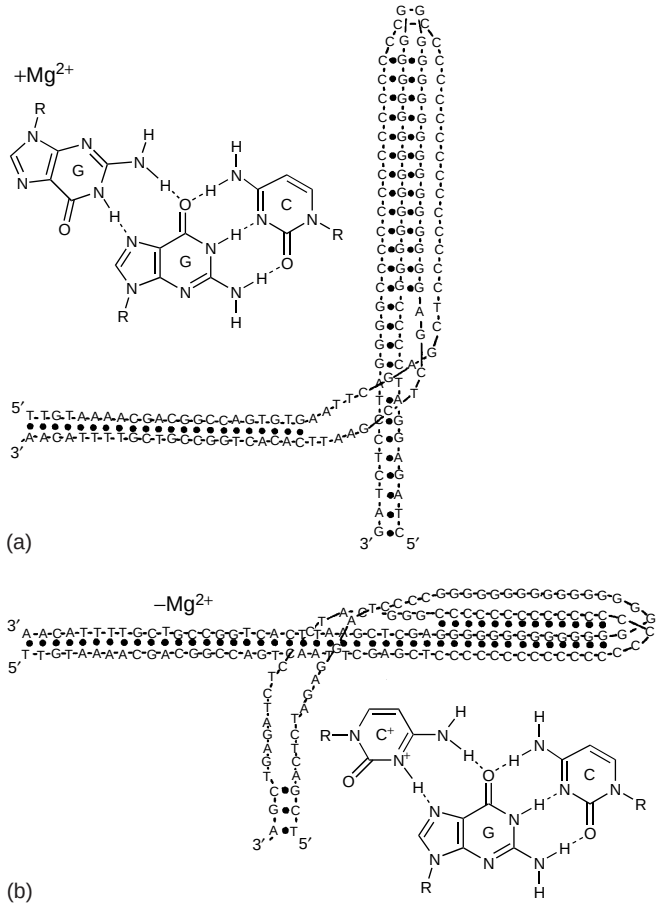

Figure 8 Probable structure adopted by the poly(dG)•poly(dC) sequence in supercoiled plasmid DNA and the corresponding hydrogen bonding scheme for the base triplet in the presence (a) or absence (b) of 2 mM Mg^{2+}. (•) Hydrogen bonds in the non-B-DNA structure; (- - -) hydrogen bonds in the triplets. (Reproduced by permission of the National Academy of Sciences from Y. Kohwi, T. Kohwi-Shigematsu, *Proc. Natl. Acad. Sci. USA*, **85**, 3785 (1988).)

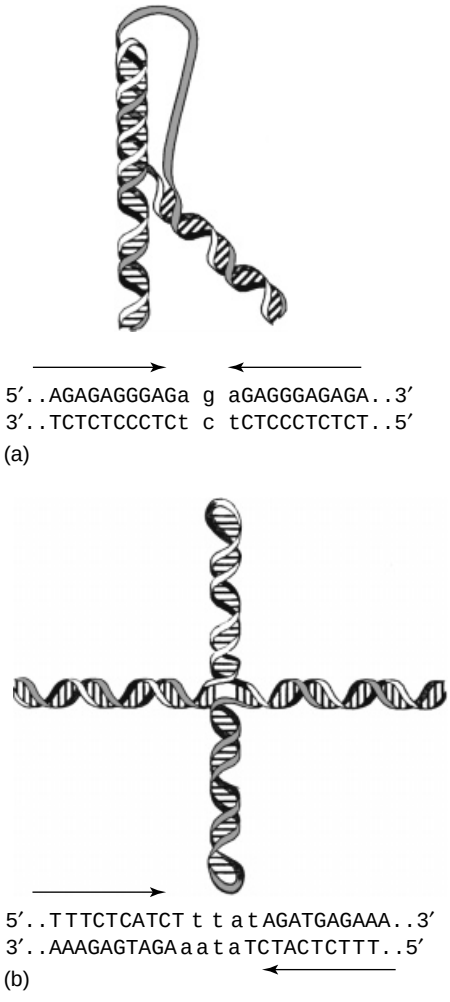

Figure 9 Triple-stranded H-form and cruciform DNAs. (a) Triple-strand H-DNA is stabilized by hydrogen bonding between the nucleobases to form base triplets. They are typically formed by sequences that are rich in purines in one strand and pyrimidines in the complementary strand. (b) Cruciform DNA is formed by inverted repeat sequences, with a dyad axis of symmetry between the two strands of the duplex. (Adapted from Prentice-Hall from K. Van Holde, W.C. Johnson, P.S. Ho, *Physical Biochemistry*, Prentice-Hall, Upper Saddle River, NJ, 59, 1998. Copyright © 1998 Prentice-Hall, Inc.)

DNA molecule bends. Consequently, structure formation requires input energy, which is supplied by superhelical stress.

Certain C-containing triplexes are stable at acidic pH and dissociate rapidly as the pH is increased above the pK_a for loss of the N3 proton.[11] Stabilization requires protonation of C N3, a prerequisite for hydrogen bond formation with unprotonated C to produce a hemiprotonated C•C_H^+ base pair. Methylating C can increase the tolerance to elevated pH, stabilizing the triplex by shifting the pK_a for protonation upward by ca. 0.2 unit.[12,13]

3.3 Four-stranded Structures

G-rich sequences common in telomeric and GC-rich promoter DNAs are highly predisposed to forming four-stranded quadruplexes. Unconfirmed proposals hypothesize that quadruplexes form and contribute to the mechanism of several processes, including the regulation of telomeric interactions, meiotic recombination, DNA methylation, gene expression and viral packaging. Quadruplexes form one-, two- and four-stranded inter- and intramolecular species, paired in both parallel and antiparallel conformations (Figure 10b and c).[14,15] The base-pairing motif, a G-quartet, is comprised of four coplanar bases that are hydrogen bonded to form a closed square (Figure 10a). They are right-handed twisted helical structures with *anti* G glycosidic conformations in most nucleotides, interspersed in some complexes with *syn* G residues.[16]

Quadruplexes are stabilized by monovalent cations C•C_H^+ base pair formation and m^5C inclusion.[13]

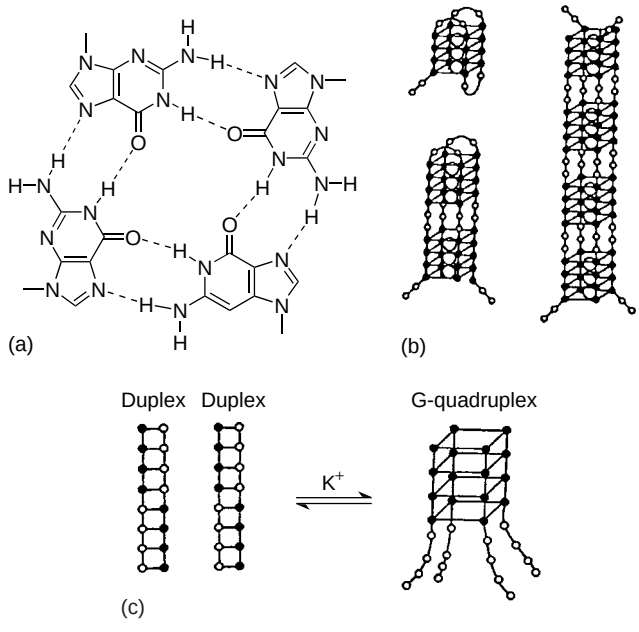

Figure 10 (a) "G-quartet" hydrogen bonding scheme. Phosphodiester backbone sugars are indicated by dR. (b) Quadruplexes formed from (i) one, (ii) two and (iii) four strands. Closed circles refer to guanines, small open circles indicate thymines and large open circles represent octahedral coordination complexes involving monovalent cations "caged" within two stacked G-quartets. (c) K^+-induced formation of G-quadruplex structure from the corresponding self-complementary WC duplex d(G_4C_4). (Reproduced from H. Deng, W.H. Branulin, *Biopolymers*, **35**, 680 (1995). Copyright © 1995 John Wiley & Sons, Inc., reprinted by permission of John Wiley & Sons, Inc., and with permission from R. Jin, K.J. Breslauer, R.A. Jones, B.L. Gaffney, *Science*, **250**, 544 (1990). Copyright 1990 American Association for the Advancement of Science.)

Among monovalent cations, potassium most stabilizes quadruplex formation, associating with the C6 carbonyl oxygens of guanosine residues to form cryptand complexes.[17,18] Physiological Mg^{2+} concentrations decrease the rate of formation of triplexes and increase the dissociation rate.[19] Competition between duplex, triplex and quadruplex depends sensitively on the nucleic acid concentration and type(s) and concentration(s) of mono- and divalent cations.[20] These folding and multimerization processes have been studied from many perspectives, demonstrating some intriguing properties.[21]

The four-stranded i-motif (i-tetrad) structure forms by interdigitation of two $C{\bullet}C_H^+$ paired parallel stranded duplexes (Figure 11a).[22] G-rich strands can form quadruplexes containing $C{\bullet}C_H^+$ and G-quartets.[10] The i-motif is an antiparallel tetrad with dyad symmetry that forms at acidic pH.[23] Cytidine protonation can occur near, at or even above neutral pH.[24] X-ray diffraction analyses showed that monoprotonated $C{\bullet}C_H^+$ base pairs form across the center of the helix, pairing the two diagonal pairs of strands, trapping the two duplexes by virtue of their interleaved base pairs, like two diagonally interlocked ladders. Interdigitation of the co-stacked duplexes brings the pairs in van der Waals contact.[23] Intercalation liberates a lot of free energy, producing a very stable structure.[25] Both types of four-stranded structures have two narrow minor grooves and two wider major grooves. The phosphate backbones are aligned in very close proximity, especially across the two minor grooves.[23]

The i-tetrad can form from a single telomeric strand. Interestingly, there are negligible differences in electrostatic interactions between the i-tetrad motif and the two identical duplexes from which it is comprised.[25] RNA cannot form the i-motif because the C2′ hydroxyl sterically hinders the intercalated geometry.

Similar to the competition between triplex and quadruplex conformation, duplexes may also change structure

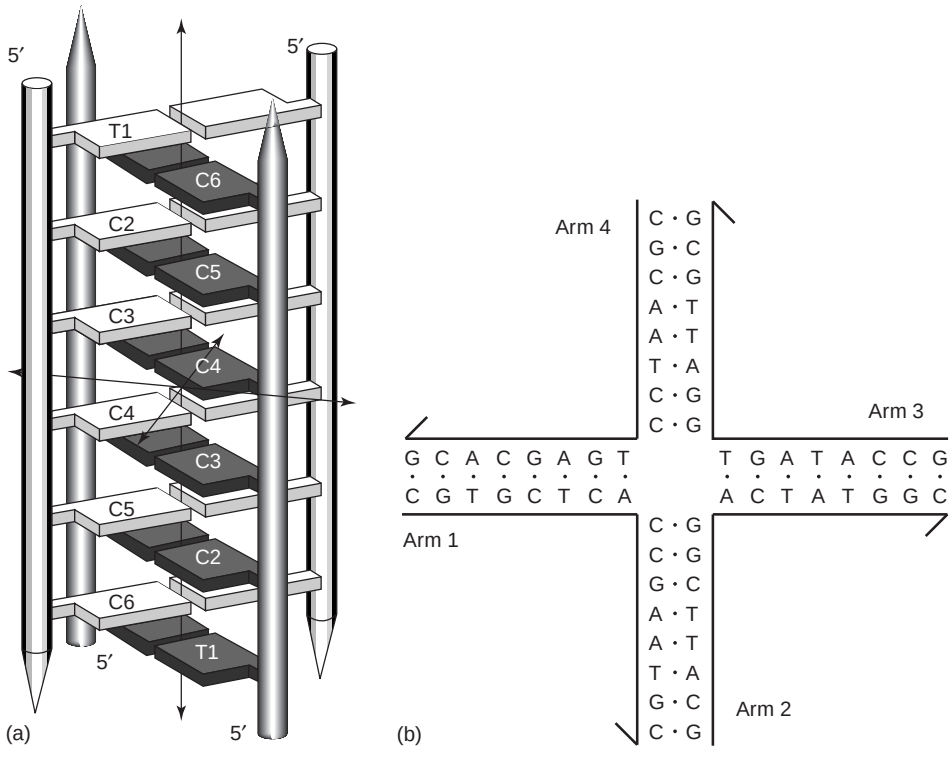

Figure 11 (a) Proposed model of the d(TC_5) tetrad. Intercalation scheme of the two parallel duplexes. The terminal thymidine residues may not be base paired. There are three axes of twofold symmetry (arrows), one along the helix axis through the center of each $C{\bullet}C_H^+$ base pair and two between the central C4–C4 base pairs. Many close van der Waals contacts between sugars give rise to the strong inter-residue H1′–H1′, H1′–H2″ and H1′–H4′ nuclear Overhauser effects (NOEs). The 5′ phosphates of C4–C4 are above the bases and the C3–C3 phosphates below. Protons C3 H1′ and C4 H2″ are separated by only 2.2 Å. (b) Sequence of the immobile junction, J1. The molecule consists of four 16-mer deoxynucleotide strands labeled with arrowheads at the 3′-OH termini. In this study, the 5′ termini are unphosphorylated. In the text the duplex branches (''arms'') are numbered counterclockwise with arm 1 at the left. Note that three arms (1, 2 and 4) contain three AT base pairs, whereas arm 3 contains four. The octameric duplexes correspond to each of the arms of this structure, and also lack 5′ phosphate termini. (Reproduced by permission of Nature, Inc. from K. Gehring, J. Leroy, M. Gueron, *Nature (London)*, **363**, 564 (1993) and of John Wiley & Sons from L.A. Markey, N.R. Kallenbach, K.A. McDonough, N.C. Seeman, K.J. Breslauer, *Biopolymers*, **26**, 1622 (1987). Copyright © 1987 John Wiley & Sons, Inc.)

to a quadruplex configuration (Figure 10c).[18–20] Long stretches of self-complementary G-rich DNAs form tetrad structures called ''eclectic DNA''.[11,18]

3.4 Higher Order Structures

The complexity of DNA structure increases in cruciform DNA (Figure 9b), immobile junctions (Figure 11b)[26] and Holliday junctions (Figure 12).[3] Seeman et al. have made and characterized a cubic DNA complex, built from segments of designed structural components comprised of rigid motifs (reported in[27]). This general field is being investigated intensively in the context of producing molecular machines, such as a nanodevice with rotating arms that interconverts between two structural states via B–Z and Z–B structural transitions.[27]

Cruciform DNA is composed of two DNA strands that have a special arrangement of complementary duplex DNA sequences called ''palindromes''. A cruciform forms when duplexes form at both ends while the two intervening strands become unpaired and loop out (Figure 9b). These strands then pair intramolecularly to form two separate hairpins (stem–loops), located external to the two flanking duplexes, forming a molecular cross.

Immobile junctions are composed of four strands which each form a duplex with two strands simultaneously, including pairing between the first and fourth strands to form a cross (Figure 11b). The junction residues form one or more mismatched base pairs.[26] An immobile junction is converted into a cruciform by covalently closing two of the four duplex segments to form hairpin loops. Immobile junctions were designed to understand Holliday junctions, a hypothetical structure proposed to form in the course of genetic recombination.[3] Immobile junctions are structurally constrained and do not ''slip'' to different pairings. The sum of stabilizing enthalpy contributions due to duplex formation in the arms approximately equals that of the junction.[26] In contrast, Holliday junctions must move freely to produce the paranemic junction, the structure formed when two duplexes intermingle during genetic recombination. In this case, enthalpies (ΔH) of both the arms and junction must be nearly zero to ensure that zipping is facile. This process is mediated by the recombinase proteins.

The recently discovered e-motif has been associated with triplet repeat expansion diseases such as Fragile-X syndrome and Huntington's disease, one of which causes approximately 20% of juvenile-onset retardation. The syndromes are attributed to overproduction (''expansion'') of triplet repeat sequences, yielding long fragile chromosomal domains. The problematic sequences were found to form hairpins containing distorted helices in vitro. This led to the proposal that such structures also form in the lagging strands of replication forks in vivo.[24] Polymerase slippage, promoted by hairpin formation, is thought to produce the structural instability and disrupt functions.[28,29] Cytosines flip out and away from the central axis of the helix, especially in sequences that

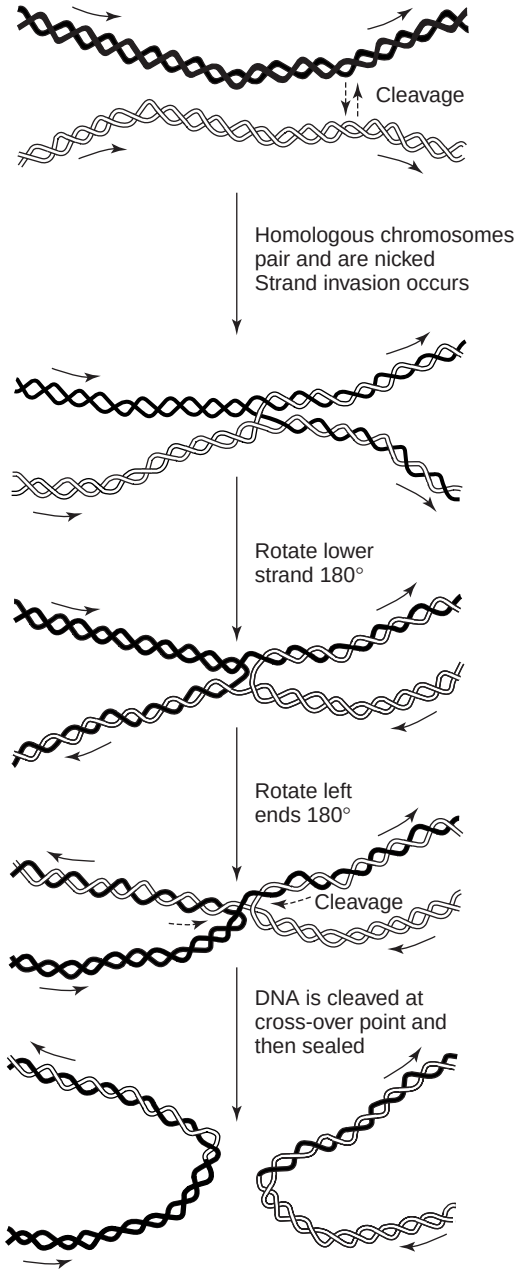

Figure 12 Holliday model of general recombination. First, nicks are introduced into a homologous region of each sequence. Subsequent strand invasion, DNA cleavage at the cross-over junction, and sealing of nicked strands results in exchange of the ends of the chromosomes. (Adapted from Prentice-Hall from L. Moran, K. Scrimgeour, H. Horton, R. Ochs, J. Rawn, *Biochemistry*, Prentice-Hall, Upper Saddle River, NJ, 26–25, 1994. Copyright © 1994 Prentice-Hall, Inc.)

accommodate C•C_H^+ base pairs. These extrahelical bases stress the phosphodiester backbone, distort the helix and induce hairpin formation. The two extrahelical bases fold symmetrically into the minor groove but do not form a base pair. [Note: an analogous ''base flipping'' mechanism leads to enzymatic methylation of C C5 atoms in duplex DNAs.[24]]

4 RNA STRUCTURE

Structures formed by RNAs and DNAs differ by the presence and absence of the C2′-hydroxyl, respectively. Hydroxide ions in basic solutions catalyze hydrolysis and cleavage of the phosphodiester backbone of RNA.[30] Hydroxide removes the hydrogen from the 2′-OH, producing an oxygen anion which then attacks the adjacent phosphate phosphorus, thereby eliminating the rest of the chain. Hydroxides are in solution and bound as inner- and outer-sphere ligands to divalent cations, a common source of catalytic activity in RNA reactions.

4.1 Duplex Structures

The most common duplex form maintained in RNA is similar to the right-handed A-form found in DNA. RNA can also form a left-handed conformation in high temperatures and salt concentrations, as well as in low hydration conditions.[31]

Both DNAs and RNAs form stable hairpins, but they are more prevalent in natural RNAs than in natural DNAs. They form when two sequential complementary regions of a strand fold together to produce a double-stranded region, terminated by one looped end and a duplex with either a blunt end, dangling strand or two extending strands (Figures 5a–d, 9b and 13a and b). In most biological examples, the free strands emerge from the duplex and continue into other strand domains. Lone hairpins have been extracted for study by researchers to assess their stabilities, structures and other properties. This general approach has its drawbacks but has led to the determination of many structural details that could not be obtained with their full-length counterparts. DNA cruciform structures contain two hairpins (Figure 9b).

RNAs can also form three- and four-stranded structures, with properties grossly similar to those of the analogous DNAs.[11,12,16,32]

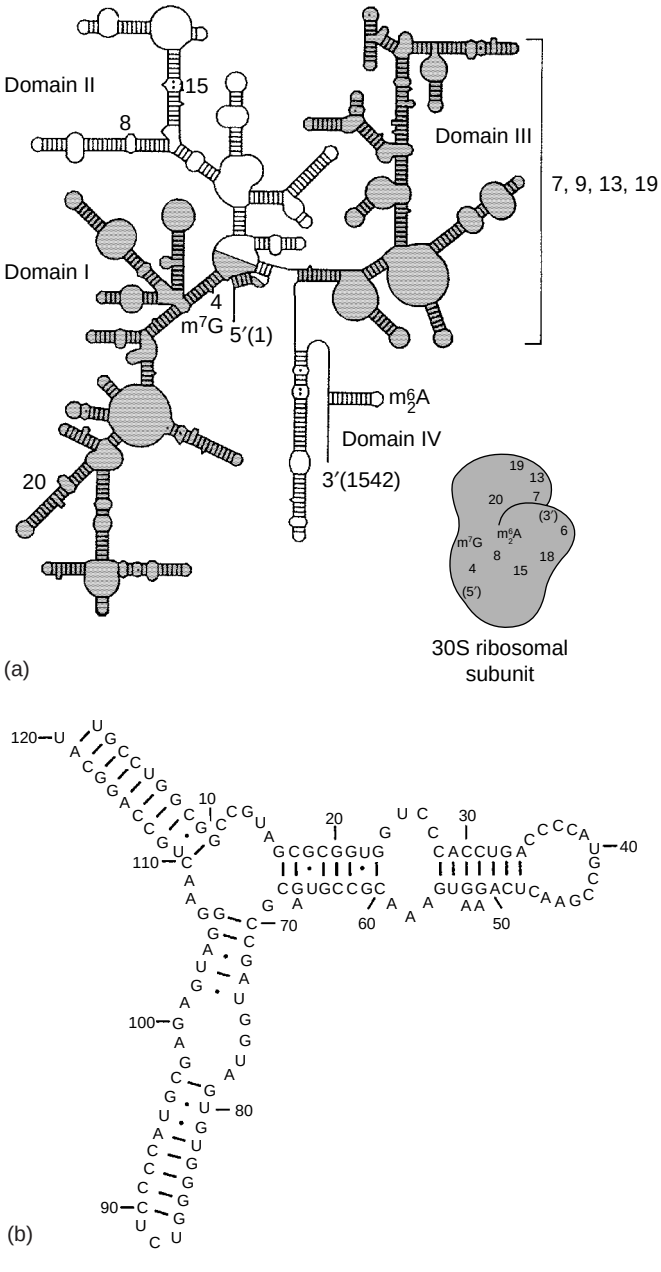

Figure 13 (a) The proposed secondary structure of the 1542 nucleotide *E. coli* 16S rRNA based on the comparison of sequences from different species under the assumption that this secondary structure is evolutionarily conserved. The flower-like series of stems and loops forms four domains that are indicated by Roman numerals. The placement of certain features with respect to specific ribosomal proteins and the entire 30S ribosomal subunit are indicated. (b) The sequence of *E. coli* 5S rRNA showing its proposed secondary structure. (Reproduced by permission of John Wiley & Sons from D. Voet, J.G. Voet, *Biochemistry*, John Wiley & Sons, New York, 916, 1990.)

4.2 Substructural Components

Complex RNA structures are described in reference to its substructural components (Figure 5a–d). These substructural components form interesting three-dimensional structures that are important for binding and protein recognition. One very common substructural component is the loop. The stability of hairpins and loops

depends on the size of the loop and number of bases included; those with more negative enthalpic (ΔH) and entropy-dependent contributions ($-T\Delta S$) to the overall Gibbs free energy (ΔG).[1,7,30] Large internal loops, composed of two strands with (ostensibly) unpaired nucleotides, are thought to be essential for protein recognition. The smallest stable internal loop could be composed of a single-base mismatch.

Base mismatches can also produce bulged RNA (Figure 5a–d). Molecules with mismatches are less stable than those without mismatches; however, some mismatches are more stable than others.[8] As expected, stabilities of mismatched base pairs are often dependent on the surrounding stacked bases. Pyrimidine•pyrimidine base pairs are generally less stable than purine•purine base pairs. For example, the G•A, U•U, and C•C_H^+ base pairs are much more stable than G•G, C•A, C•U, A•A or C•C pairs. The U•U mismatch is stabilized by hydrogen bonding between complementary imino form nucleobase functional groups.

In an internal bulge (Figure 5a–d), one or more unpaired bases exist on one strand, sometimes stacked between the two flanking base pairs; in others, the base is excluded from the helix to produce a moderate deformation (it ''flips out''). The fate of the unpaired base depends on the surrounding sequences; pyrimidines tend to move away from the axis, whereas purines can remain extrahelical or intercalate, depending on the context and solution conditions.[7,30] Phosphates in bulged strands are extremely close together, producing an anionic domain. In one example, the anionic charge of the closely grouped phosphate oxygens interacts with two Mg^{2+} ions to stabilize the bulge.[33] Bulges and loops are similar in that they are more constrained than single-stranded molecules. As hairpin loops increase in size, their constraints decrease progressively toward those of lone strands.[7]

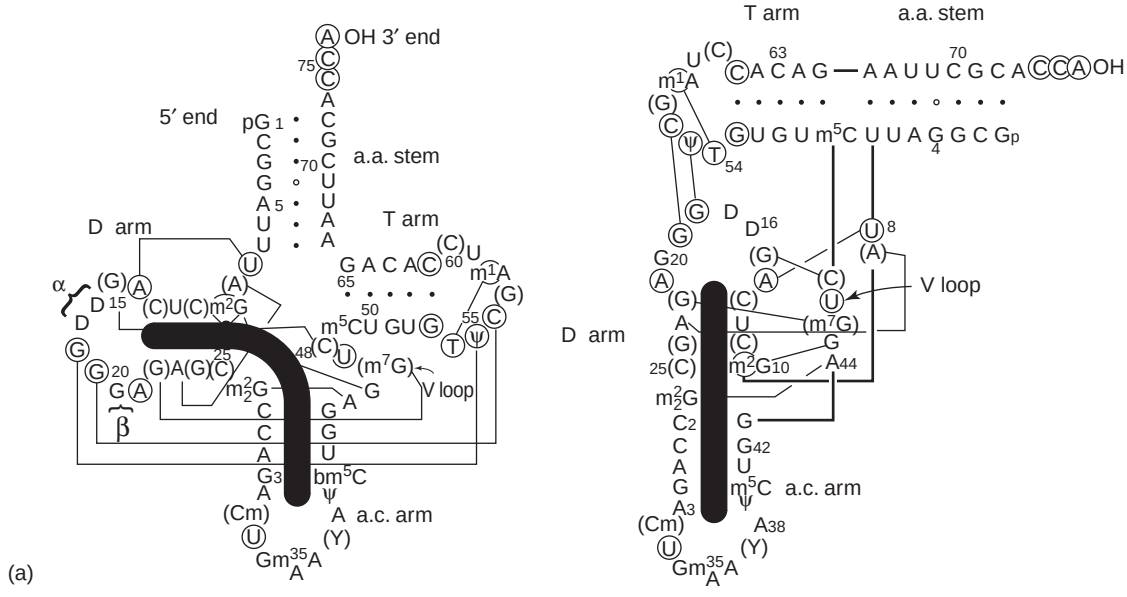

Figure 14 (a) Nucleotide sequence of yeast $tRNA^{Phe}$ folded (left) into secondary ''cloverleaf'' and (right) into tertiary ''L-shaped'' arrangements. Invariant bases in class I tRNA are indicated by circles; those which are semi-invariant are enclosed in parentheses. WC-type hydrogen bonds in stem regions are marked by dots and the G•U ''wobble'' pair (G4•U69) by a small circle. The number of nucleotides in variable loop V and in α and β regions can vary, depending on the tRNA species. Bases forming tertiary base–base hydrogen bonds are connected by thin lines. Note that most of the invariant and semi-invariant bases are positioned in the middle and hinge regions of the ''L'' and involved in tertiary hydrogen bonds which, together with base stacking, illustrates which parts of the ''cloverleaf'' join together to form the two legs of the L. (b) Illustration of base-pairing and stacking interactions in yeast $tRNA^{Phe}$. Base pairs stabilizing secondary and tertiary structure are drawn as bent or fused slabs or by connecting two slabs with rods. Invariant and semi-invariant bases cluster in anticodon loop, CCA end of acceptor stem, and at outer corner and hinge regions of the ''L''. Note stacking of bases 34–38 in the anticodon loop, consisting of the anticodon 2′-*O*-methylguanine (G_m)43–A35–A36. [Circled letters a–g indicate where base pairs and triplets displayed in Figure 15-4 in the original publication are located.] (c) Schematic representation of the π_3-turn situation in anticodon and T loops of yeast $tRNA^{Phe}$. Sugars and phosphate groups are simplified as large and small spheres, respectively; bases as slabs; base pair interactions by a bar; H represents hypermodified base. Semi-invariant and invariant bases are stacked with the base pair at the top end of the loop, and the bases at the bottom of the anticodon loop are again stacked with the bases to the right. Phosphates are actively engaged in the turn and form hydrogen bonds and stacking interactions. (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 332, 335, 339, 1984. Copyright Springer-Verlag 1984.)

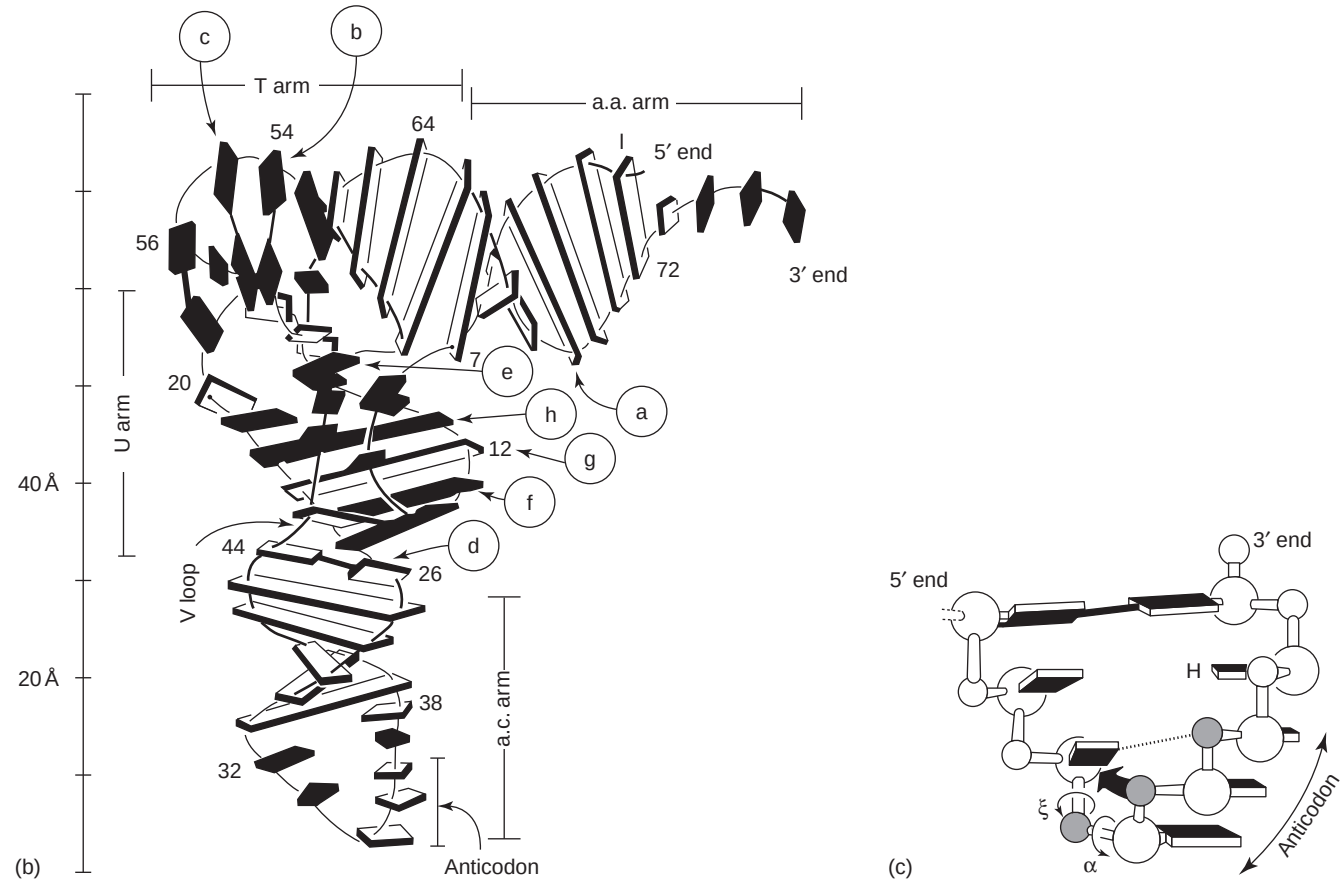

Figure 14 (*Continued*)

4.3 Higher Order RNA Structures

Substructural components interact to form higher order structures. For example, the amino acid acceptor stems and anticodon stem–loops of tRNAs form an L-shaped tertiary structure when the other two stem–loops fold together to form a kind of ''glue'' that stabilizes the ''joint'' region of the L (Figure 14a–c).[1] Structure is one of the defining components of function and is intimately related to sequence. Strands composed of different sequences can sometimes form similar secondary structures. In fact, in a few known cases secondary structures can adopt similar tertiary structures.[7]

The anticodon bases of tRNA jut away from the rest of the loop and can interact with ligands (Figure 14a–c). The anticodon loop bases can convert between the ''3′-stacked'' and ''5′-stacked'' structures, in which the nucleotides in the loop align above either the 3′- or 5′-terminal strand of the attached duplex stem. Although the anticodon region of tRNA is single-stranded, its bases stack and thus stabilize the extended exposed face of the trinucleotide, which facilitates codon recognition and decoding. Another stabilizing structural component of the anticodon is the U-turn, in which U33 interacts with the phosphate of residue U37 (Figure 14c). A wealth of different base assemblies occur in tRNAs, forming a large percentage of our lexicon of RNA substructural components.

Pseudoknots are a common higher order structural motif in natural RNAs. In these structures, the loop of a hairpin binds to a complementary sequence adjacent to the sequence that forms half of the stem of the hairpin (Figure 15a and b).[7,34] Unlike U-turns, pseudoknots do not occur in tRNAs. The ability of the pseudoknot structure to bend is affected by steric hindrance between nucleotides that comprise the two entangled loops. The structure is compact, not elongated, resembling a moderately extended globular protein. Because the pseudoknot is comprised of two smaller hairpin substructures, the RNA molecule can form two different hairpins with either a 5′- or 3′-terminal ''dangling end'', depending on their respective free energies (Figure 14a–c).[7] Transitions between these two structures depend on both base sequence and salt concentration.

Pseudoknots are thought to regulate transcription, possibly controlled by cation concentration fluxes, protein

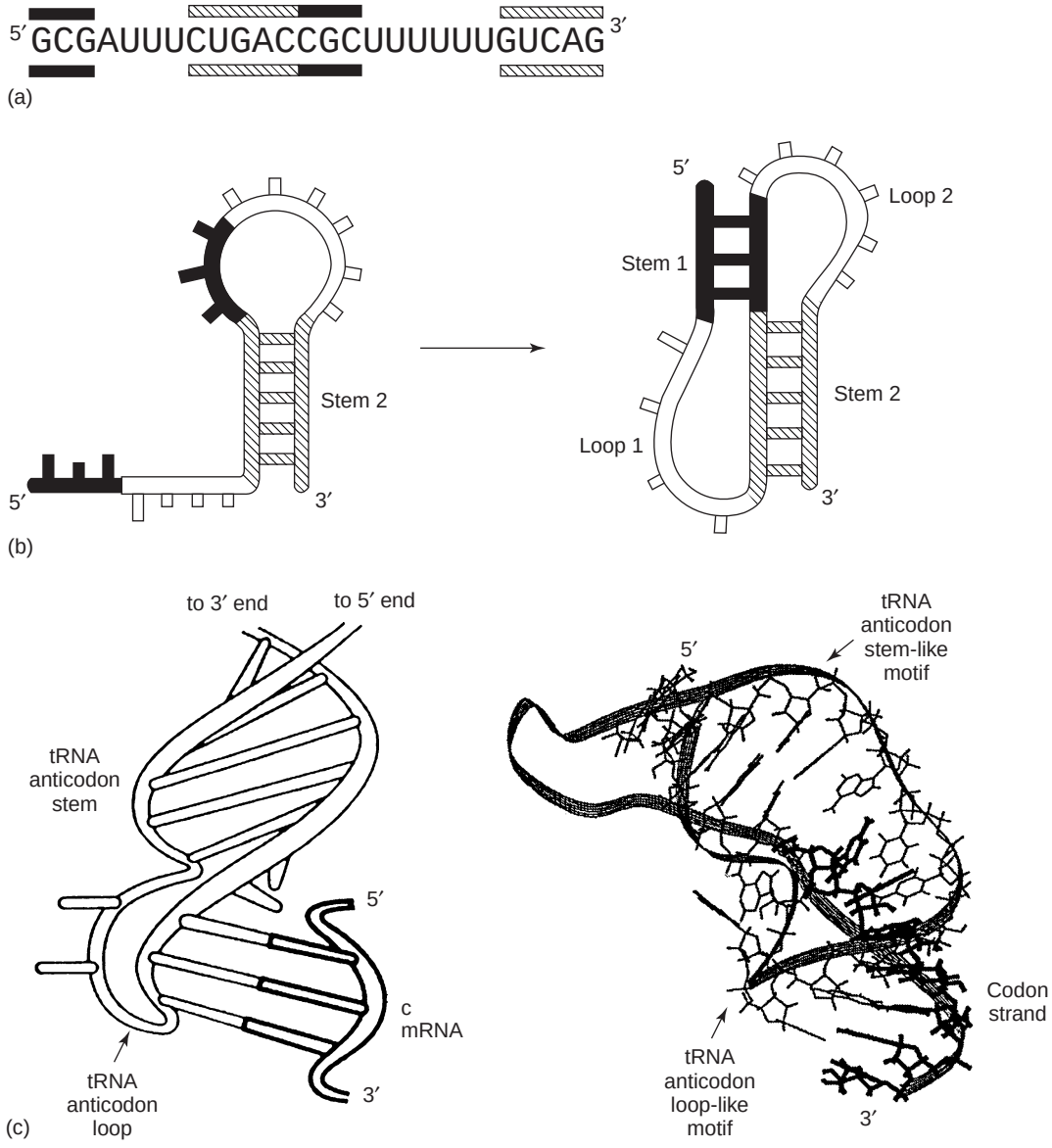

Figure 15 Pseudoknot formation. (a) Sequence of the oligonucleotide. (b) A pseudoknot is formed when a few bases outside the hairpin formed at the 3′ end of the oligonucleotide pair with a few bases within the loop of the hairpin to form a second stem and loop region. Stem I is denoted in black and consists of three base pairs. The five base pairs of stem 2 are hatched. Loop I bridges stem 2 across the major groove of the RNA helix. Loop 2 crosses stem 1; since this loop bridges the minor groove this distance is further than that crossed by loop I. (c) Similarity between the structures of the [tRNA anticodon•codon] complex and mouse mammary tumor virus RNA frameshift pseudoknot. (Reproduced from J.R. Wyatt, J.D. Puglisi, I. Tinoco, Jr, *Bioessays*, **11**(4), 104 (1989). Copyright © 1989 John Wiley & Sons, Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc., and by permission of Harcourt Brace Inc. from L.X. Shen, I. Tinoco, Jr, *J. Mol. Biol.*, **247**, 974 (1995).)

recognition or both.[7] They can mediate protein recognition by forming structures resembling tRNA acceptor stems. For example, this occurs at the 3′-termini of some plant viral genomic RNAs (e.g. tobacco yellow mosaic virus). In this case, two different secondary structures, viral pseudoknots and tRNA stems, form similar tertiary structures. The 3′-terminal hydroxyl group of the viral RNA is aminoacylated by the ''cognate'' aminoacyl tRNA synthetase, forming a protective genomic ''cap'' that prevents exonucleolytic digestion.

Ribosomal recognition depends on sequentially binding similar, but subtly distinctive, tRNA anticodon and mRNA codon structures formed as the mRNA is processed by the ribosome and read by tRNA. This normal process can be subverted in several known ways. For

example, a suppressor tRNA skips (reads through) a stop codon by inserting an amino acid (commonly tyrosine) into the nascent protein chain instead of terminating protein synthesis.

In another situation, translating ribosomes can pause when the mRNA is allowed enough time after its transcription to form a certain stem–loop structure. The time-frame of the pausing event depends on the stability of this hairpin. The phenomenon is the basis for translational attenuation, in which amino acid starvation results in the absence of that aminoacylated tRNA. During this duration, one type of folded stem–loop motif forms and translation continues.[2,3] To ensure that the anabolic enzyme needed to produce this amino acid is made, sufficient ''charged'' aminoacyl tRNAs must be available to translate the leader sequence codons. If stalling lasts long enough, an alternatively folded hairpin forms and translation is aborted. The amino acid concentration, titrated by the time-frame of transcription and movement of tRNAs through the ribosome, determines whether protein is made or not. This mechanism illustrates the tightly linked interplay between two different folded mRNA structures and the initial unstructured transcript. Their energetics and relative folding time-frames determine finally whether it is translated or abandoned.

Translational frameshifting occurs when ribosomes pause preferentially at internal sites on the decoding mRNA strand. When a translating mRNA forms a pseudoknot, the ribosome pauses much longer at that site. This increases the degree of ''slippage'' in base pair alignment between tRNAs and the reading frame of the mRNA, making a frameshift possible. One such example occurs during translation of the mouse mammary tumor virus genome, in which alternative viral proteins are encoded in different overlapping reading frames (Figure 15c).[22] Pseudoknot formation produces a (−1) nucleotide frameshift and, as a result, alternative proteins.

The recently identified A-platform structure was determined by X-ray crystallography (Figure 16a–c).[21,33] This large RNA fragment forms a catalytically active subregion which was extracted from the much larger *Tetrahymena* Group I intron structure. This fragment is important because the residues in the vicinity of the A-rich bulge are essential for ribozyme activity.[20] The A and flanking residues bind a string of Us in a second domain formed by distant primary sequence residues. This ''docking'' interaction links distantly separated sequences via base pairing. A ''GNRA tetraloop'' subfragment[35] binds the A-platform, termed its receptor, and as a result produces a tertiary structural motif composed of a stem fragment ''super-looped'' back on itself and adhered to a distant stem with an intervening bend region. The platform undergoes a large conformational change in the absence of this tertiary interaction. The A-platform is stabilized more by base stacking forces than hydrogen bonding, but both contribute.[21]

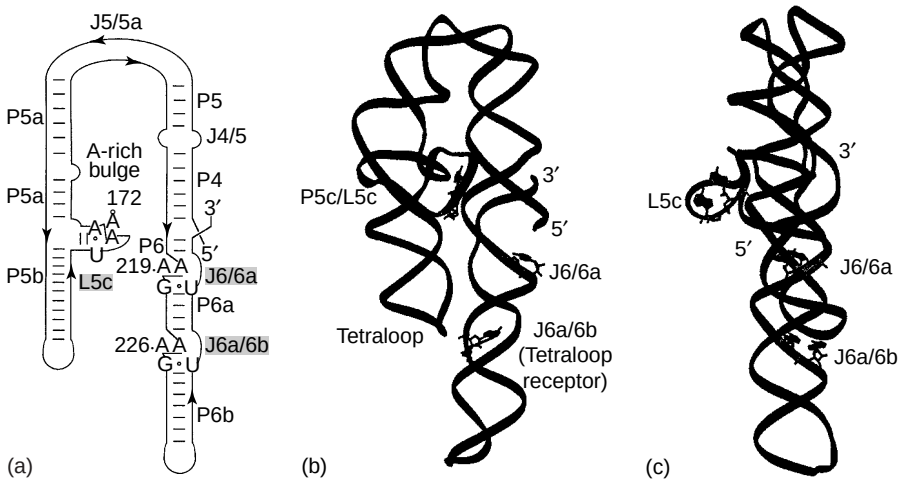

Figure 16 (a) Schematic secondary structure of the P4–P6 domain, with adenosine platform in bold. The noncanonical base pair below each platform is also shown. (b) Location of adenosine platforms in the crystal structure. The structure is shown in the same orientation as that of (a); adenosines of the platforms are highlighted in blue and green (in the original article), while the backbone positions of the wobble base pairs below the platforms are indicated in red. Blue adenosines are accessible to dimethyl sulfate (DMS) in the presence and absence of the docking partner associated with the platform; green adenosines (in the original article) are protected from DMS modification when the docking partner is present. (c) Another view of the crystal structure, rotated 90° about the long axis of the molecule. The conserved core region is to the right in (a) and (b) and the core faces the reader in (c). (Reprinted with permission from J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, A.A. Szewczak, C.E. Kundrot, T.R. Cech, J.A. Doudna, *Science*, **273**, 1697 (1996). Copyright 1996 American Association for the Advancement of Science.)

Many of the stem loop sequences in known RNA folds contain consensus GNRA tetraloop sequences.[2,3,35,36] Ribosomal RNA secondary structures have been predicted from sequences obtained from organisms spanning a broad range of phylogeny. Sequence requirements for the unique stabilization are very specific and clearly linked to binding interactions between the participating residues. The initial G base and final A phosphate form a hydrogen bond and the N7 atom of the third purine (R) forms a hydrogen bond with the sugar hydroxyl hydrogen at the N position, which can therefore be any nucleotide. The second nucleotide is not specifically involved in stabilizing the structure. These more stable tetraloops are thought to play a key role in the tertiary folding of large RNAs. Reverse transcriptase stop sites often contain this sequence.

Another unique example of RNA tertiary structure is formed when two hairpin loops dock to produce ''kissing hairpins'', the evolutionarily conserved motif formed by genomic human immunodeficiency virus (HIV) RNA.[37] The HIV genome is comprised of two identical positive sense RNA strands. Hairpin loops at the 5′ ends of each molecule promote genomic dimerization in vitro and are required for viral packaging in vivo. RNA quadruplex formation by two condensed HIV genomes has also been proposed to mediate viral packaging.[16]

The hammerhead RNA motif is functionally and structurally similar to the Group I intron fragment.[33,38] The three stems in the hammerhead ribozyme form a three-way junction (Figure 5d); the nucleotides all maintain the *anti* configuration. Hammerhead RNA is thought to mediate RNA processing, catalyzing specific cleavage and maturation of the initial RNA transcript. Structural formation and enzymatic activity require the presence of divalent cations, primarily Mg^{2+} in vitro, although other cations (and some of the transition and posttransition metals) function in a similar fashion. They are thought to act as Lewis bases, increasing the local OH^- concentration and possibly directing reactants into the precise tertiary structure(s) essential for cleavage.[38]

Hammerhead RNA cleavage has been characterized using Michaelis–Menten kinetics. Reaction rates depend on reactant and product concentrations and pH values.[38] Cleavage does not occur below pH 5, most likely because required nucleotides are protonated. The reaction rate increases until pH 7 is reached, then decreases rapidly.[7] Cleavage does not occur at higher temperatures. Since proper tertiary structure formation is so essential, the destabilization that occurs at higher temperatures probably short-circuits the reaction mechanism.

5 ENERGETICS

The numerous examples described so far clearly illustrate that the structures, functions and energetics of nucleic acids are intricately linked. Many techniques and strategies have been developed to study these relationships quantitatively, each emphasizing different aspects of detected properties and/or functional groups of the nucleic acid under investigation. These structural entities range from individual bonds to moderately sized complexes (e.g. duplexes) to large structures, such as superhelically stressed plasmid DNAs. One strategy used to study these problems is to ''divide and conquer''. Studies define facets of a problem in a piecewise fashion in the hopes of using the resulting information to solve problems involving the structure within its overall context. As previously outlined, the smallest complex substructure is the base pair. Adjacent base pairs affect the stability of each other. Nucleotides on both sides contribute when a base or base pair is embedded within a duplex, producing a sequence-dependent synergism. This additivity is called and modeled mathematically as the ''nearest neighbor effect''.

Interbase orbital overlap is optimized within the constraints otherwise available to the aggregated domain as a result of the hydrophobic effect. Dipole–dipole London dispersion forces also participate in stabilization, but knowledge regarding relative contributions is not precise. Theoretical calculations support the expectation that large subdomains form synergetically. Quantum mechanical energies and van der Waals forces require including structures up to at least a 15 Å distance away from a particular atom to achieve self-consistency. Individual base pairs affect those around them and vice versa.

5.1 Thermodynamic Characterization

The most common method to characterize the energetics of nucleic acids in the context of structure involves measuring equilibrium thermodynamics. The net spontaneity of a reaction is characterized by ΔG, but only if the mechanism is rapid compared with the scan rate of the melt temperature-ramping process. The familiar result is that more stable interactions have more negative free energies. In the usual vernacular, the process (e.g. structure formation) is more spontaneous. An important characteristic of ΔG is its logarithmic proportionality to the equilibrium constant and its dependence on temperature, characterized by the fundamental equations [Equations 1 and 2].[30,39–41] Reactants must attain enough energy to pass the transition state to form products.

$$\Delta G = \Delta H - T\Delta S \tag{1}$$

$$\Delta G = -RT \ln K \tag{2}$$

Both ΔH and ΔS contribute significantly to the net ΔG of many duplex denaturation–renaturation reactions. These parameters characterize sequence-dependent stabilities of intra- and intermolecular nucleic acid complexes. Their relative contributions demonstrate many subtleties regarding interdependences between molecular energies, structures and solvent components. In many cases, they give important insights into the relations between structures and functions. In this account, we generally only list qualitative comparisons or trends found, especially focusing on how the parameters depend on the relevant variable(s) (but many quantitative results are cited in the Further Reading).

Base stacking is one of the predominant driving forces behind the maintenance of nucleic acid structure. The partially hydrophobic bases stack with one another to minimize interaction with the surrounding solution and each other (Figure 17).[1,6,42] As expected, the more hydrophobic purines stack with more stability than pyrimidines. Stacking ability depends on the electron distribution of the rings and their exocyclic components. Disrupting the stacked structure is usually accomplished by raising the solution temperature, leading to thermal denaturation (melting). Less stacked or unstacked random coil structures are rarely found, except at elevated temperatures or under relatively destabilizing conditions (i.e. in lower polarity solvents or in 7 M urea).

Hydrogen bonds also play a major role in stabilizing nucleic acid structure.[1] In fact, helix structures literally revolve around them. Exocyclic protons and lone-pair electron components on opposing bases form these complementary interactions. To a reasonable approximation, relative stabilities and flexibilities of different structures are determined by hydrogen bonds. Their strength depends critically on interatomic distance and orientation, with linear alignment of the three atoms being optimal. Since G•C base pairs have three hydrogen bonds, they are stronger than A•T or A•U base pairs, which only form two. Relative stabilities correlate with the number of hydrogen bonds. Since both G and C are more hydrophobic than A and T (and less hydrophobic U), stabilization is also correlated with less polar components.

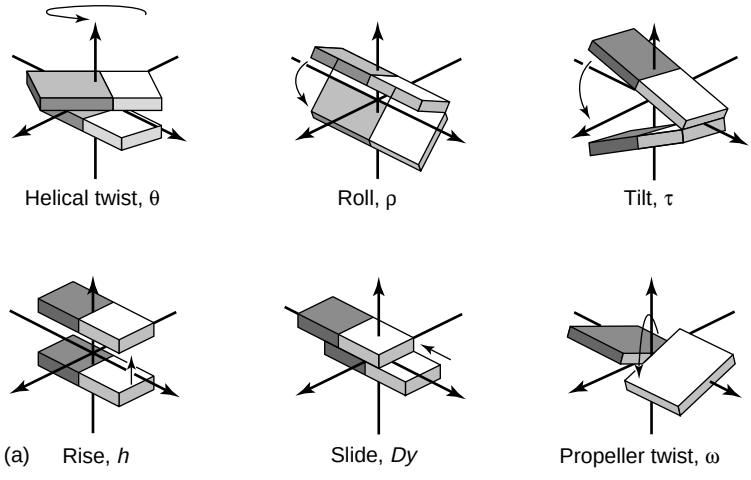

Figure 17 (a) Base-pair and base-step parameters of nuclei acid double helices. The structures of double-stranded nucleic acids are defined by the relative conformations of two adjacent base pairs in a base step (e.g. helical twist, roll, tilt, rise and slide) and the relative conformations of the bases in a base pair (e.g. the propeller twist). (b) Illustration of ''Calladine's rules''. (i) Adjacent G•C base pairs in a DNA double helix, with positive propeller twist, display sequence-dependent interactions between guanines on opposite polynucleotide chains. (Top) The sequence pyrimidine-(3′,5′)-purine leads to clashes in minor grooves (open arrow), whereas the purine-(3′,5′)-pyrimidine sequence (middle) causes clashes in the major groove. Clashes can be avoided by reduction of propeller twist, opening up the roll angle θ_R (bottom), reduction of helical twist (t) shown in (c), or shifting base pairs such that the purine (guanine) is pulled out of the helical stack [shown in (ii)]. (ii) The purine–purine clash displayed in (a) can be reduced by shifting the G•C base pair laterally, which pulls the G out of the helical stack (arrow). In order to maintain similar helical P–P distances, sugar puckers are changed, yielding a larger δ angle for purine and a smaller one for pyrimidine. Associated with these changes are χ rotations, as required for the ''autocorrelation principle'' [section 11-2 in the original publication]. Increase of helical twist from 36° to 45° (top) shortens the distance between N3 atoms in the minor groove (top), whereas reduction to 27° (bottom) widens this distance. If coupled with a positive propeller twist, N3 atoms in the top base pairs (heavy lines) move down (indicated by minus signs), those in the bottom base pairs (heavy lines) move down (indicated by minus signs) and those in the bottom base pairs (light lines) move up, leading to unfavorable N3–N3 clashes. These clashes are diminished by helical untwisting to 27° (bottom). (Adapted from Prentice-Hall from K. Van Holde, W.C. Johnson, P.S. Ho, *Physical Biochemistry*, Prentice-Hall, Upper Saddle River, NJ, 56, 1998. Copyright © 1998 Prentice-Hall, Inc., and of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 254, 255, 1984. Copyright Springer-Verlag, 1984.)

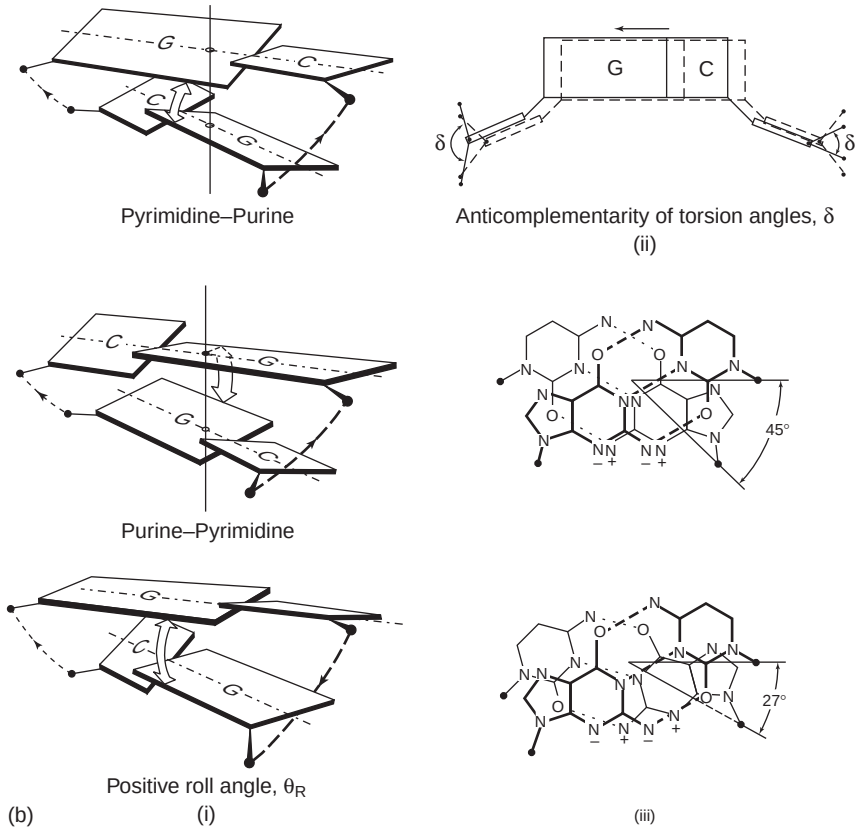

Figure 17 (*Continued*)

Enthalpy changes can be determined indirectly via Van't Hoff analysis or directly using calorimetric approaches. Van't Hoff approaches monitor the properties of a certain substructure as they change (Figures 18a and b, 19a and b and 20). One must find conditions suitable to measure the equilibrium constant of a reaction as a function of temperature. The melting behavior of nucleic acid complexes is sensitive to its strand number, its molecularity (n). The Van't Hoff enthalpy (ΔH_{VH}) is calculated using Equation (3):[40]

$$\Delta H_{\mathrm{VH}} = (2 + 2n)RT_{\mathrm{m}}^2 \left(\frac{\partial \alpha}{\partial T} \right)_{T - T_{\mathrm{m}}} \qquad (3)$$

The "melting temperature" (T_{m}) refers to the temperature at which the concentrations of single (melted) strands and paired structures are equal, the midpoint of the equilibrium, and α is the extent of the reaction. The term $T - T_{\mathrm{m}}$ indicates the temperature range studied to probe the reaction equilibrium. For species with $n > 1$, the equilibrium poise is strongly dependent on the strand concentration, adding more shifts the poise toward the products.[32,39,40] Increasing the temperature dissociates the strands. The breadth of the $(1 - \alpha)$ vs T curve is an indication of the degree of cooperativity of the transformation. Zipper models capture the image well; initial pairings facilitate subsequent pairings (positive cooperativity). Both duplex assembly and melting are relatively (positively) cooperative. Increased ionic strength and cationic polyamines and proteins increase the cooperativity.[39,43,44]

5.2 Calorimetric Determination of Enthalpies

Calorimetric methods also provide an approach to obtaining information about the enthalpy changes associated with structural transformations, ligand binding, etc. The change in heat capacity (ΔC_{p}) is determined calorimetrically by scanning across the same range of temperatures used for the UV spectral melt data acquisition, then integrating the resulting derivative plot [Equation 4]:[40]

$$\Delta H^{\circ} = \int \Delta C_{\mathrm{p}} \, \mathrm{d}T \qquad (4)$$

The differential scanning calorimetry (DSC) apparatus scans through the temperature range encompassing the structural transition slowly enough to ensure that the strands equilibrate at each intermediate temperature.[39] DSC enthalpies should be equal to those calculated by

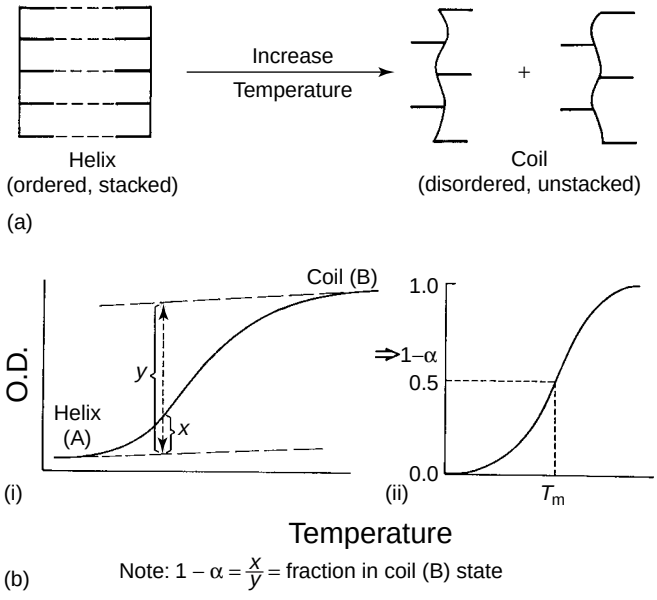

Figure 18 (a) Schematic representation of a thermally induced transition from a fully base-paired and base-stacked double helix to two unpaired and unstacked single strands. (b) (i) Schematic of an optical density (O.D.) vs temperature profile (a UV melting curve) showing the hyperchromic effect associated with a transition from a base stacked (helix) to an unstacked (coil) state. (ii) A derived melting curve showing the fraction in the coil state $(1-\alpha)$ as a function of temperature. The two figures together illustrate how the experimental absorbance vs temperature profile can be converted into a $(1-\alpha)$ versus temperature melting curve. (Reproduced by permission of Springer-Verlag from K.J. Breslauer, in *Thermodynamic Data for Biochemistry and Biotechnology*, ed. H.-J. Hinz, Springer, New York, 405, 1986. Copyright Springer-Verlag 1986.)

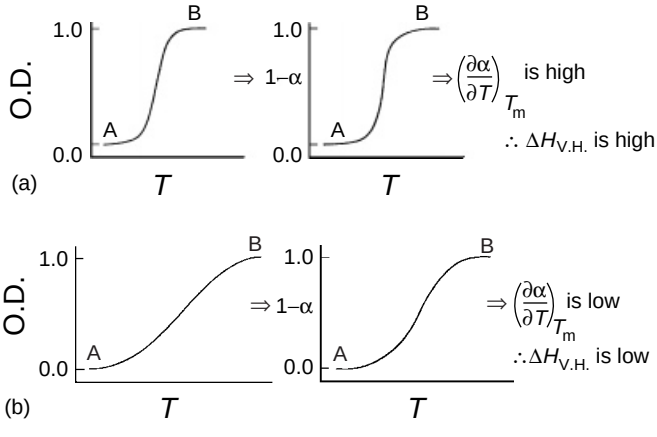

Figure 19 The shape of the derived $(1-\alpha)$ vs temperature melting profiles for (a) cooperative and (b) noncooperative transformations. Noncooperative transitions have broader melting profiles and $(1-\alpha)$ vs temperature curves with reduced $(\partial\alpha/\partial T)$ slopes at the T_m (below). This results in a lower value for the transition enthalpy. (Reproduced by permission of Springer-Verlag from K.J. Breslauer, in *Thermodynamic Data for Biochemistry and Biotechnology*, ed. H.-J. Hinz, Springer, New York, 408, 1986. Copyright Springer-Verlag 1986.)

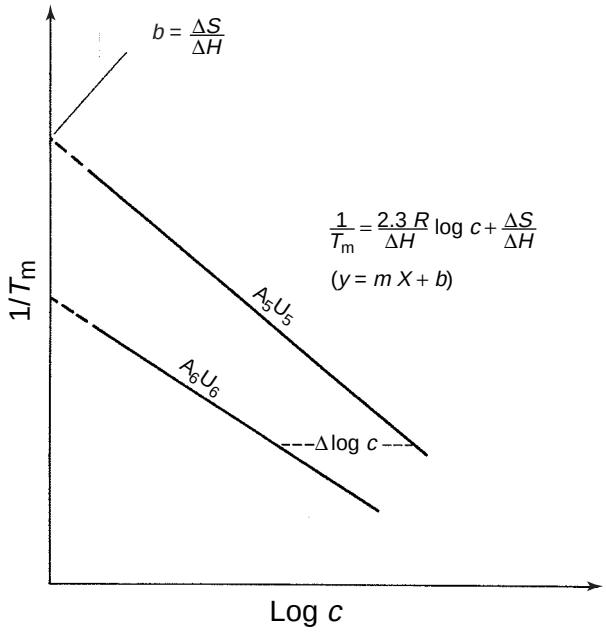

Figure 20 Typical plots of the concentration dependence of the melting temperature. When plotted as $1/T_m$ vs $\log C$, the slope is equal to $R/\Delta H$ and the intercept is equal to $\Delta S/\Delta H$. At any temperature, the horizontal distance between the two lines $(\Delta \log C)$ is proportional to the free energy difference between the two self-complementary oligomers as described in the text. (Reproduced by permission of Springer-Verlag from K.J. Breslauer, in *Thermodynamic Data for Biochemistry and Biotechnology*, ed. H.-J. Hinz, Springer, New York, 151, 1986. Copyright Springer-Verlag 1986.)

van't Hoff analysis if one applies the correct molecularity. For example, n was calculated from calorimetric and spectroscopic data obtained with the quadruplex DNA formed by $d(G_2T_5G_2)$ in 1 M NaCl.[17]

Entropy is related to the degree of disordering open to a molecule, which is modeled mathematically by enumerating the number of possible statistical variations of a "system" and stipulating the population distributions among them. These distributions could characterize a set of different molecular conformations, orientations or interactions between functional groups or more than one molecule. They could also include their interactions with ligands or solvent.

Water contributes in a particularly important way to entropy changes of biomolecules.[1,45,46] The stability of nucleic acid structural aggregates is strongly influenced by hydrophobic interactions involving buried and exposed residues and the surrounding solvent. The energy associated with the cage of water molecules surrounding the structure is governed by both entropy and enthalpy; the dominant contribution changes with increasing temperature. At temperatures ($\lesssim$60 °C), hydrogen bonds (enthalpic contributions) interconnect the surrounding

H_2O molecules and ''encage'' the more hydrophobic nucleic acid, forcing it to fold into the native structure. When the temperature is increased above ca. 60 °C, vibrational fluctuations overcome the hydrogen bond energies, the cage H_2O molecules dissipate and the nucleic acid denatures. Enthalpic contributions to folding energies dominate at lower temperatures; the ΔS associated with H_2O randomization and release dominates at higher temperatures. A similar situation governs equilibria associated with protein folding and membrane assembly.

5.3 Spectroscopic Techniques

As discussed previously, UV absorbance spectroscopy is commonly used to study the thermal denaturation behavior of nucleic acids (Figures 18–21).[1,30,39] This method allows one to study temperature-dependent dissociation (melting) and association (hybridization) behaviors, conformational changes and ligand binding. The large extinction coefficients of the nucleobases allow one to perform studies with relatively low sample concentrations (micromolar to millimolar). The molar extinction coefficients of individual bases decrease when the bases stack, a phenomenon called the hypochromic effect (Figures 18 and 19). UV spectroscopy allows one to monitor the course of temperature-dependent transitions by studying the ratio of stacked to unstacked bases. The total absorbance of the disrupted structure is larger than the sum of the absorbances of each individual nucleotide. Denaturation by melting the stacked bases produces an increase in the absorbance at 260 nm, allowing one to obtain quantitative energetics, usually by studying the concentration dependence of T_m (Figure 20).

The sequence of a nucleic acid influences the stability of the strand in a base-paired secondary structure. Contributions to stabilization by stacking interactions, resulting from different nearest neighbor free energies, determined for a set of DNA sequence variants, are shown in Table 1. Similar data obtained with RNA variants are shown in Tables 2 and 3. These nearest neighbor stacking energies provide a method to estimate the stabilities of RNA or DNA duplexes (Figure 21). This information has been used extensively to predict the most energetically likely secondary structures (e.g. Figures 13–16). While these predictions are useful, they are only approximations and must be interpreted cautiously. For example, slow reaction kinetics can lead to invalidation of the equilibrium assumption upon which the analysis is based.

Secondary structures are also predicted from evolutionary sequence conservation data. Biological sequences remain similar over time and large portions of genomes often exhibit cross-species similarities. Therefore, it is inferred that they maintain similar structures. Promoter region sequences of genes, GpC islands, centromeres, telomeres and other repetitive sequences (SINES, LINES, Alu sites) are often evolutionarily conserved.[4,47] Slight changes in sequence (and as a result structure) may produce changes in gene expression or RNA-mediated functions. Mutations do not typically affect the functions of DNAs, RNAs or their encoded proteins. Although usually quiescent, the mutant RNA or protein structure may contain interactions or adopt structures with severely compromised functions, resulting in dire health consequences. Many of the modern approaches to studying disease revolve around the protein aspect of this

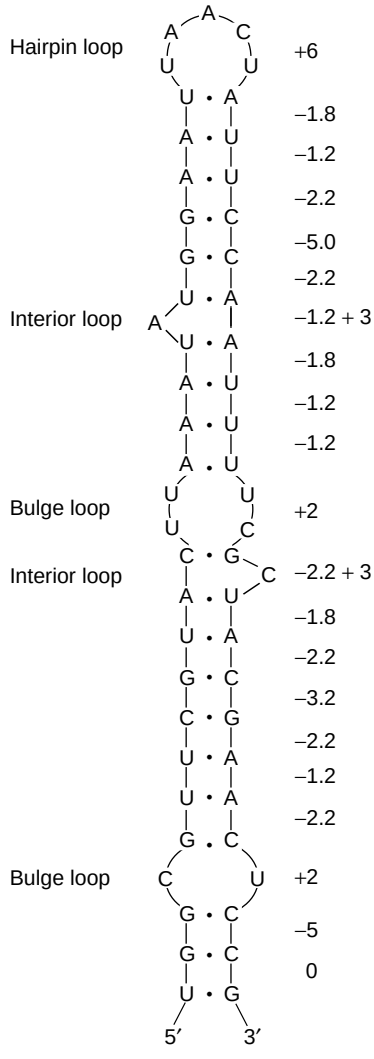

Figure 21 Application of data to evaluate a possible secondary structure for a 55-nucleotide fragment from R17 virus. Overall ΔG at 25 °C is -21.8 kcal mol^{-1}. (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 411, 1984. Copyright Springer-Verlag 1984.)

Table 1 Nearest neighbor stacking energies: ten possible dimer duplexes found in B-DNA (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 139, 1984. Copyright Springer-Verlag 1984.)

Stacked dimer (strand•strand)		Stacking energy (kcal mol^{-1})
$^{5'}$GC•CG$^{5'}$		−14.59
$^{5'}$AC•TG$^{5'}$	$^{5'}$GT•CA$^{5'}$	−10.51
$^{5'}$TC•AG$^{5'}$	$^{5'}$GA•CT$^{5'}$	−9.81
$^{5'}$CG•GC$^{5'}$		−9.69
$^{5'}$GG•CC$^{5'}$	$^{5'}$CC•GG$^{5'}$	−8.26
$^{5'}$AT•TA$^{5'}$	$^{5'}$CA•GT$^{5'}$	−6.57
$^{5'}$TG•AC$^{5'}$	$^{5'}$CT•GA$^{5'}$	−5.78
$^{5'}$AA•TT$^{5'}$	$^{5'}$TT•AA$^{5'}$	−5.37
$^{5'}$TA•AT$^{5'}$		−3.82

problem, in part because much is known about protein structures, functions and the effects produced by mutations. Much less is understood about nucleic acids, although the situation is improving rapidly. One cannot overemphasize the key linkage between mutations, structural consequences and altered functional viabilities. This is used to advantage by some error-prone viruses (e.g. HIV), which mutate rampantly, and in the process ''learn'' (adapt) to out-fox antiviral pharmaceutical therapies.

Nucleotides at specific positions of a structure are evolutionarily conserved in order to preserve the secondary and tertiary structures in which they participate. Covariant mutations occur when two paired bases in the helix of one wild-type secondary structure are replaced by the inverted base pair in another. This swaps the flanking bases, which can change the stability, but preserves the base pair. The second mutation occurs to compensate for the lost interaction produced when the first mutation occurs.

A more complicated absorption property of nucleobases, their interaction with circularly polarized UV light, has been used extensively to study nucleic acid structures and interconversions. The circular dichroism (CD) spectrum, obtained with a spectropolarimeter, is a plot of the difference in absorbance of negatively (right) and positively (left) circularly polarized light as a function of the wavelength range studied.[1,45,48] This difference is very sensitive to the detailed chirality of stacked and interacting base assemblies and their structural changes. Many structural transformations, ligand binding reactions and chemical modifications (e.g. with covalently reactive mutagens) have been characterized.[49] Nucleic acids become more chiral when a helix forms and less so when it denatures. When information can be acquired using other techniques that yield structures [i.e. molecular coordinates; e.g. X-ray crystallography or NMR] or about specific bond motions (e.g. NMR, Raman scattering spectroscopy), CD results can be used to great advantage to study changes between two or more previously characterized states. The more qualitative CD technique is sensitive and easier, allowing one to characterize a much more extensive range of variables.

IR and Raman spectra are sensitive to vibrational motions of specific functional groups and are used to study the stretching and bending frequency modes of molecules. IR and Raman methods allow one to study interactions between functional groups, e.g. to monitor the formation and loss of hydrogen bonds. Bond formation and dissociation change the frequency of the absorption (or scattering) peak, allowing detailed analysis of bonding changes in nucleic acid and between them and proteins, ligands, etc.

Raman and IR spectra provide complementary information about the active types of vibrational modes predominant in the absorptive functional groups being monitored.[1,45,48] The bond is the chromophore. For example, Raman spectroscopy was used to study the assembly and disassembly of DNA quadruplexes formed

Table 2 Experimentally determined free energies for RNA duplexes (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 150, 1984. Copyright Springer-Verlag 1984.)

Stacked dimer base pairs (strand•strand)				Stacking free energy (kcal mol^{-1} dimer)
$^{5'}$AA•UU$^{5'}$				−1.2
$^{5'}$UA•AU$^{5'}$				−1.8
$^{5'}$AC•UG$^{5'}$	$^{5'}$CA•GU$^{5'}$	$^{5'}$AG•UC$^{5'}$	$^{5'}$GA•CU$^{5'}$	−2.2
$^{5'}$CG•GC$^{5'}$				−3.2
$^{5'}$GC•CG$^{5'}$	$^{5'}$GG•CC$^{5'}$			−5.0
$^{5'}$GU•UG$^{5'}$				−0.3
$^{5'}$GX•UY$^{5'}$	$^{5'}$XG•UY$^{5'}$			0.0

Table 3 Experimentally determined free energies from RNA looped structures (Reproduced by permission of Springer-Verlag from W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 150, 1984. Copyright Springer-Verlag 1984.)

m (loop size, bases unbonded)	Stacking energy (kcal mol^{-1} dimer)	
Interior loops:		
2–6	2	
7–20	3	
$m > 20$	$1 + (2 \log m)$	
Bulge loops:		
1	3	
2–3	4	
4–7	5	
8–20	6	
$m > 20$	$4 + (2 \log m)$	
Hairpin loops:		
	Closed by G•C	Closed by A•U
3	8	>8
4–5	5	7
6–7	4	6
8–9	5	7
10–30	6	8
$m > 30$	$3.5 + (2 \log m)$	$5.5 + (2 \log m)$

Table 4 Thermodynamic parameters of loop formation (Reproduced by permission of the American Chemical Society from J. SantaLucia, R. Kierzek, D.H. Turner, *Biochemistry*, **30**, 8242 (1991).)

RNA sequence[a,b]	ΔG°_{37} (kcal mol^{-1})	ΔH°_{37} (kcal mol^{-1})	ΔS°_{37} (eu)
5′GCUGGCG3′	-3.94 ± 0.16[c]	-28.8 ± 1.8	-80.2 ± 5.3
5′CGCAGGCG3′	-0.64 ± 0.23	-13.9 ± 2.8	-42.6 ± 8.4
5′GGCGUGCC3′	-0.40 ± 0.14	-14.1 ± 1.6	-44.1 ± 4.7
5′GGCGAGCC3′	-0.36 ± 0.28	-6.3 ± 2.6	-19.2 ± 7.7
5′CGCUUGCG3′	-0.08 ± 0.12	-14.9 ± 1.5	-47.7 ± 4.5
5′UGCGGGCA3′	1.18 ± 0.36	-20.2 ± 6.3	-68.9 ± 18.3
5′CGCCAGCG3′	1.44 ± 0.08	-8.2 ± 1.0	30.9 ± 3.2
5′CGCAAGCG3′	1.69 ± 0.11	-1.6 ± 1.5	10.3 ± 4.5
5′CGCCUGCG3′	1.70 ± 0.13	2.0 ± 1.3	1.2 ± 4.0
5′CGCUCGCG3′	1.70 ± 0.25	-0.2 ± 2.2	-6.1 ± 6.8
5′CGCCCGCG3′	2.00 ± 0.11	0.8 ± 1.2	-3.7 ± 3.7
5′GGCACGCC3′	2.33 ± 0.11	0.9 ± 1.3	-4.7 ± 3.9

[a] Listed in order of decreasing loop free energy.
[b] Calculated from T_M^{-1} vs log C_T parameters.
[c] Error based on assuming error for rUGCGCA is 10% for ΔH° and ΔS° and 2% for ΔG°.[71]

by one or two strands, demonstrating that the mono- and bistranded complexes depend differently on Na^+ and K^+ concentrations.[15] Raman scattering works better than IR to characterize most aqueous samples. Atoms can move in a symmetric scissors motion (toward and away from each other within the plane formed by the three atoms). Asymmetric wagging fluctuations (of the two bonds with respect to each other) are also possible. In the symmetrical motion analogy, one scissor blade can either close when the other closes or opens simultaneously. In the asymmetric case, the motions of the bonds occur in two planes that are skewed with respect to each other and at frequencies that are not correlated in time. Only symmetrical shifts are active (observed) using IR spectroscopy. Molecules with complex conformational constraints rarely maintain perfectly symmetrical motions, so Raman scattering can be used to study functional groups that fluctuate predominantly via symmetrical motions, but also have these imperfect symmetrical stretches. Raman spectroscopy is relatively sensitive and has a key advantage relative to IR. Peaks normally observed using IR are obscured by the dominant symmetrical stretch motion of solvents such as water. Water is less Raman active and can usually be filtered out using deuterium oxide, allowing visualization of most functional group peaks. Structures and dynamics of nucleic acids have been characterized in systems ranging from nucleotides to duplexes to viral genomes in packaged virus particles. Interactions, motions and assembly–disassembly phenomena have been elucidated.

NMR has also provided a wealth of structural and dynamic information about nucleic acids. ^{1}H-NMR has been used to study protons; ^{31}P-NMR provides information about the backbone phosphorus atoms and connected structures.[49–53] ^{13}C-NMR has been used to study skeleton carbons; ^{15}N-NMR has yielded information about amino and imino NH groups. All of these isotopes, in addition to deuterium, are used in various multidimensional NMR experiments to assign peaks by characterizing through-bond interconnectivities. Spectroscopic relaxation studies with ^{1}H-, ^{13}C-, ^{31}P- and ^{15}N-labeled molecules have been used to probe fluctuational motions, their spatial magnitudes and timescales. The ^{1}H–^{1}H nuclear Overhauser effect (NOE) and scalar J-coupling methods allow one to determine the frequencies (chemical shifts) of peaks of specific hydrogen atoms and determine calibrated distances between these assigned atoms. Three-dimensional structures are determined by reconstructing spatial connectivities between neighboring hydrogen atoms.

Problems associated with simultaneously stipulating structure and the ranges and frequencies of molecular motions, when many different types of motion are occurring, complicate calculating structures. Full matrix data analysis techniques and molecular mechanics/dynamics simulation protocols have been developed to do so, but information is limited so far (some examples are described in the Further Reading section).

5.4 Electrophoretic Techniques

Electrophoresis separates molecules, structures and/or complexes based on different charges and size-dependent properties. Larger molecules and plasmids are separated on agarose gels; moderately sized and smaller molecules are resolved on polylacrylamide gels. The electrophoretic mobility decreases with the size (charge) of the negative nucleic acid backbone. High-percentage polylacrylamide gels (ca. 20%) can separate oligo- and polynucleotides ranging up to ca. 500 nucleotides in length, differing by only one nucleotide, termed single-base resolution in sequence determination. This has allowed the molecular characterization of many sequences, ligand binding sites, complex formation reactions, mutagen action scenarios and other reactions. The results are always based on how the process of interest alters the size and/or charge of the nucleic acid fragments under study. The chaotropic solute urea is typically used to denature nucleic acids in gels. Native structures and complexes have been studied extensively. For example, protein binding decreases the mobilities of bound nucleic acids, the basis of the mobility shift technique.[54]

Electrophoretic techniques have been used to understand the effects of superhelical stress on the supercoiling behavior of DNA. Supercoiled closed-circular plasmid DNAs are more mobile than relaxed plasmids because they are smaller hydrodynamic particles. Introducing negative supercoils while maintaining the B conformation introduces torsional strain; adding a single negative supercoil increases ΔG (at 25 °C) for duplex formation by about 38 kJ mol^{-1}. Thermodynamic stress is relieved by induction of Z-DNA and cruciform formation. Extremely large amounts of DNA are present in the small volume of the nucleus. The DNA packaging process requires a large amount of torsional stress.[55]

Such molecules have traditionally been modeled as ordered condensed double helices. Aged cells contain significant amounts of damaged aggregates. Nucleic acids are recycled when cells undergo apoptosis (programmed cell death), degraded by Ca^{2+}/Mg^{2+}-containing nucleases.

Electrophoresis has been used to determine how topoisomerases relax and induce torsional stress by catalyzing induction and/or relief of supercoiling. Topoisomerases are essential for DNA replication and transcription, since torsional stress must be relieved in order to unwind the double helix. Topoisomerase I relaxes torsional stress by cutting only one strand in duplex DNA and rotating the freely dangling ends in opposite directions around the axis of the uncut strand. Topoisomerase II enzymes (DNA gyrases) can introduce negative supercoils by introducing double-stranded breaks, then resealing the backbones. Since the number of rotations prior to resealing vary in a stepwise (integer) pattern, one obtains a family of discrete topoisomers. The individual topoisomers can be separated and the energies associated with their interconversions quantitated using a two-dimensional gel electrophoresis technique involving the use of the intercalative molecule chloroquine.[56,57]

Electrophoretic analyses of products generated using more traditional methods involving chemical reactivity and enzymatic recognition have also been used to determine how a broad range of intercalators affect the structure and energetics of DNAs and RNAs. Intercalation involves stacking of a planar aromatic molecule (e.g. ethidium bromide) between two base pairs, thereby increasing the length of double-helical DNA.[1] This distortion can result in mistakes in replication which may produce genetic mutations. Since intercalators relieve superhelical stress, the hydrodynamic volume of a plasmid increases, decreasing its electrophoretic mobility.

DNA can form higher order structured aggregates. One example of a noncytoplasmic, minimally soluble, DNA state forms when DNAs and RNAs packed inside viral particles to form condensed aggregates. Topological stresses in such situations are not well understood. Structures have been studied using low-resolution electron microscopy, low-angle X-ray scattering and Raman spectroscopy approaches. Model structures called P-DNA and ψ-type DNA have been developed based on CD, linear dichroism and CD scattering results. Left-handed Z-DNA duplexes can form multistranded, torus-shaped macroaggregates which have been characterized using electron microscopy.

Temperature gradient gel electrophoresis is another example of a method that can be used to determine the melting behavior of a nucleic acid structure. One determines the change in electrophoretic mobility of a structure produced by increasing the temperature. Samples are electrophoresed in two dimensions, through an increasing temperature gradient in the first, and through an electric field that separates molecules according to size in the perpendicular direction. For example, analysis of interactions between the P4 and P6 domains of the self-splicing Group I RNA from *Tetrahymena* demonstrated that tertiary structural folding is driven predominantly by ΔH, with a smaller $-T\Delta S$ contribution to ΔG.[58]

Continuous mixing analysis has been used to study equilibria between strands, a duplex and a triplex. Absorbance changes, due to continuous variations in strand number in populations containing two- and three-stranded complexes, were used to determine stoichiometries. Relative concentrations of oligo(A) relative to oligo(U) were varied from 1 : 10 to 10 : 1. The data led to the interpretation that two duplexes disproportionate to produce an [oligo(A)•oligo(U)] duplex at A/U $\approx$ 1 : 1 and an [oligo(A)•oligo$(U)_2$] triplex when the ratio is ca. 1 : 2.[32]

6 CONDITION-DEPENDENT STRUCTURAL TRANSITIONS

6.1 Cation Dependences

Nucleic acid structural transformations can be induced by changing the existing solution components, concentration, temperature, ligand binding and many other factors. Cation concentrations can strongly influence their stabilities by modulating melting rates and reactant–product equilibrium distributions. Atmospherically bound cations are relatively loosely bound to the anionic oxygens of the phosphodiester backbone [equilibrium dissociation constant (K_d) $\approx 10^3\,M^{-1}$]. Duplex dissociation is accompanied by cation (counterion) release. Counterion condensation theory[59] was developed to fit cation–nucleic acid binding isotherms (curves) measured using ^{23}Na-NMR. Fits require that only about 90% of the phosphodiester backbone charge (due to anionic oxygens) be neutralized to achieve thermodynamic equilibration. Duplexes have a higher negative radial charge density than individual strands. Because cations help neutralize the additional charge density, duplexes become favored and stabilized as the salt concentration is increased. The trend of increased charge density continues in proceeding to triplexes and quadruplexes. The surrounding cations are accompanied by H_2O molecules that are packed more densely than bulk H_2O, held within a Donnan sphere as a result of osmotic pressure (and other effects produced by dissolved osmolytes).

The right-handed B-form duplex poly[d(CG)$_2$] is converted to the left-handed Z-form with a transformation midpoint at ca. 2.2 M NaCl using strands in the micromolar range.[9] Placing bulky bromine atoms or methyl groups at the C8 atoms of guanines 8-bromoguanosine (br^{8}G) induces Z formation because the base swings about the glycosidic bond, relative to the sugar ring, from the B-form *anti* conformational isomer (conformer) to the *syn* structure. The B- to Z-DNA conversion also occurs upon adding 5 mM $MgCl_2$, but only if the Cs are methylated, poly[d(m^{5}CG)$_2$], illustrating a case of transformational induction mediated by two synergistic factors. The physiological $MgCl_2$ concentration is ca. 1–5 mM, but varies with organism and cell type. Cytidine methylation occurs naturally in many chromosomal domains in patterns that vary with cell-cycle phase and developmental stage dependent factors.[2,4,47] Correlations suggest that Mg^{2+} and m^{5}C may work to induce B to Z transitions in live cells under some circumstances.

The Z-form of RNA adopted by poly[r(CG)] is much more difficult to induce than Z-DNA. Z-RNA formation occurs in buffer at a transition midpoint of 6 M $NaClO_4$, a much more chaotropic salt than NaCl, and heating to ca. 40 °C. A solution of 4 M $MgCl_2$ in buffered 10% ethanol also works. Like DNA, the presence of br^{8}G induces Z-RNA formation; unlike DNA, 5-bromocytidine (br^{5}C) inhibits Z-RNA production. Molecular mechanics modeling analysis suggests that the difference can be attributed to steric repulsion between the C5 bromine atom and the 2′-hydroxyl, which is missing in DNA.[60]

Changing the type of cations and their concentrations and also the strand concentration(s) can profoundly affect the stability of quadruplex DNA. Guanine-tract-containing RNAs also readily form the quadruplex, e.g. a fragment from *E. coli* 5S rRNA (Figure 13b, ca. nucleotide 80).[16] Manipulating these factors can induce transitions from duplex to quadruplex and (under very specific circumstances) vice versa. Quadruplexes are stabilized by Group 1a monovalent cations in the order $K^+ > Na^+ > Rb^+ > Li^+$ and Cs^+ by binding in the center of the Gs, forming a cation spine down the central axis of the complex (Figure 10b). The distribution of products formed by two different quadruplex formation pathways, from two or four strands, can be varied by manipulating the total concentrations of strands, K^+ and Na^+. This phenomenon, the sodium–potassium switch, permits control over the predisposition to form alternative quadruplex structures from strand precursors by manipulating cation and strand concentrations.[17–21] Re-equilibration involves competition for unpaired strands, the common source of reactants used in the two competing assembly processes. At low Na^+ or K^+ concentrations (<240 mM Na^+ or <65 mM K^+), the antiparallel quadruplex formed by annealing *two* hairpin side-by-side forms. At higher salt concentrations, the parallel *four*-stranded quadruplex forms. As the DNA concentration is increased, less K^+ is required to induce the parallel-stranded quadruplex, which is preferred over the two-stranded antiparallel structure.

WC duplex formation is usually preferred relative to quadruplex formation because the duplex association rate is typically much faster than that for quadruplexes, but not always. For example, at elevated strand and K^+ concentrations, [d(CGC G_3 GCG)$_4$] quadruplex formation becomes preferred over the WC hairpin.[20] Quadruplex assembly competes successfully with the hairpin formation pathway for strands.

When sufficient monovalent cation is present, divalent cations typically stabilize duplexes, triplexes and quadruplexes. Concentrated nucleic acid/divalent cation solutions are prone to forming aggregates containing large strand stoichiometries.[61] Curiously, >20 mM Cd^{2+} or Zn^{2+} can dissociate examples of all three types of complex.[62] For example, the quadruplex [d(CGC G_4 GCG)$_4$] is dissociated with the following predispositions: $Cd^{2+} > Zn^{2+} > Ca^{2+} > Mg^{2+}$. Much higher concentrations of some divalent cations (Cd^{2+}, Zn^{2+}) are also known to dissociate duplexes and triplexes. The sensitivity

of duplexes, triplexes and quadruplexes to dissociate at comparatively low divalent cation concentrations may support significant transformation events among different forms of condensed nucleic acids in genomic chromatin, packaged virus particles or other niches.

A duplex DNA fragment in supercoiled plasmid DNA was found to form two different types of H-DNA triplex, with and without added Mg^{2+}. A triplex stabilized by G•G•C base triplet forms when the sequences shown in Figure 8(a) and (b) contains 2 mM Mg^{2+}.[10] In contrast, an alternative $C_H{}^+$•G•C-containing H-DNA triplex structure forms in the absence of Mg^{2+}.

B-form duplex DNA has between 15 and 20 surface-bound H_2O molecules per nucleotide.[63] In one structure, about 17 waters bind, five in the vicinity of each base pair, four associated with each phosphate and two bound to the sugar. All of the base pairs are associated with the same number of water molecules, water binding is approximately sequence independent. A spine of hydration can form in the minor groove of the crystal structures of AATT sequence-containing DNAs, reinforcing the B-form structure by participating in bidentate interactions with T O2 and A N3.[1,63] Energy studies suggest that these hydration interactions help stabilize the duplex. Osmotic pressure is a key energetic consideration; added osmolytes can dramatically affect the stabilities.

Certain repeating oligo(A•T) tract sequences can be placed in the proper position and appropriate phase with respect to the helix pitch, the length of one turn of the helix, to produce bent DNA.[64] Only properly phased sequences work because they ensure that subcontributions to the bend accumulate far enough to produce the fully realized change in net helix orientation. When the sequences are not properly phased, long-range bending does not occur.

This is a relatively extreme example of DNA structural microheterogeneity. Crystallography results demonstrated that there are sequence-specific differences in rotation between base pairs in B-form DNAs, lateral positioning of base pairs relative to the helix axis and propeller twisting of the two bases within a base pair with respect to each other, etc. (Figure 17a and b).[1,6,42] These specificities have been incorporated into rules that can be used to predict how sequence changes affect helical microstructures.

6.2 Ligand-induced Structural Transformations

Nucleic acids often undergo structural transformations when proteins bind. A vast canon of enzyme-catalyzed, sequence-specific covalent cleavages, modifications, ATP-requiring locomotion events, decoding and rearrangement reactions, etc., form the basis of modern molecular biology.[2,3,55] Protein structures must often adapt and mold to their ligand upon binding to accomplish these processes. Because the backbone of DNA is negatively charged, DNA-binding domains in proteins are typically composed of positively charged, often protonated, amino acids, arginine, lysine and histidine. Since the bases are prone to stacking interactions, the aromatic amino acids phenylalanine, tyrosine and tryptophan are also common in nucleic acid binding proteins. They function by stacking between base residues to form intercalative interaction in the [protein•DNA (or RNA)] co-complex.

The protonated, polycationic polyamines spermine and spermidine resemble the electropositive amino acids, providing a small molecular model to help understand how analogous protein fragments bind, neutralize and stabilize DNA and RNA structures in vivo. Spermidine is required to crystallize tRNAs and in many nucleic acid and [nucleic acid•protein] co-complex crystallization protocols.[1] It is thought to function by neutralizing the polyanionic tRNAs and reducing intermolecular repulsions in the crystal. Both of these polyamines induce the B- to Z-DNA transition, at micromolar concentrations when m^5Cs are present – almost three orders of magnitude less than the Mg^{2+} concentration required to do the same.

Specific DNA-binding elements in proteins mediate most known genetic control mechanisms. Two important duplex DNA-binding motifs, the helix–turn–helix and zinc finger, are very common in the DNA-interaction domains of repressor and transcription factor proteins.[55] Sequence-dependent differences in functional groups in the major and minor DNA grooves produce specific recognition sites, leading to protein-mediated recognition, specificity and regulation. Two DNA-binding proteins, e.g. transcription factors, often bind each other via two entwined α-helices, one contributed by each protein, called a leucine zipper.

Protein binding is usually fairly specific for one base sequence, the consensus sequence. In some cases, however, a broad range of sequences is recognized. An extreme example of broad RNA recognition specificity occurs with phage T4 late transcript regA mRNAs, which display a broad range of viable consensus binding sequences.[65,66] This range was studied quantitatively by assessing the propensities of randomized RNAs to evolve through a series of regA protein-binding selection steps using a technique called systematic evolution of ligands by exponential enhancement (SELEX). This process involves a series of consecutive cycles of substeps: nitrocellulose filter binding, reverse transcription, cloning to complete each cycle and sequence determinations following each round to follow the process. Tetraloop structures formed by the RNA strongly inhibit their

reverse transcription, being so stable they essentially stop protein transit.

A general model has been developed to understand specificity in protein–DNA binding interactions.[67] The researchers wanted to determine how repressor proteins specifically recognize their correct operator DNA and reject nonspecific DNAs. The findings highlight the subtlety of specific DNA recognition. Electropositive binding proteins all have an intrinsic propensity to bind any nucleic acid sequence by virtue of its polyanionic charge, the predominant externally recognizable property. The protein cannot bind too tightly or repression cannot be easily reversed; binding constants for repressor binding to specific DNAs are typically only about five-fold larger for the correct consensus sequence than for any sequence. All sequences are bound with comparable association rate constants (k_a). Specificity is maintained by having incorrect sites dissociate much more readily (quickly) than correct sites, i.e. the dissociation rate constant (k_d) is larger for correct than for incorrect sites. As a result, proteins stick to correct sites and are lost from incorrect sites. The high k_a ensures that the protein will rebind elsewhere, eventually sticking when the correct sequence is encountered.

Modular components, excised from their full molecular context for study, are being used to learn how they function in their more complicated niche. A few selected examples are described briefly below. In each of these cases, the nucleic acid structure interacts with a protein ligand that contributes significantly to the net energetics of the process.

The arginine fork motif mediates specificity in the HIV [tat protein•TAR RNA] binding interaction.[68] In this model, an internal bulge and a nearby six-nucleotide loop mediate specific binding of the tat protein, leading to transcriptional activation. Upon arginine binding, the bulge changes conformation, and essential nucleotides for binding form a U•A•U base–triple interaction that stabilizes arginine hydrogen bonding to a nearby G and phosphates.

The HIV Rev peptide binds stem–loop IIB of the Rev Response Element RNA. The Rev peptide has an α-helical segment that binds in the major groove of the RNA near a purine-rich internal loop, with several arginines making base-specific contacts, including a G•A base pair. NMR studies demonstrated that the phosphodiester backbone adjacent to a G•G base pair adopts an unusual structure that allows the peptide to access a widened major groove.[69] The major groove of A-RNA is normally deep and too narrow to accommodate an α-helix, thus requiring non-WC interactions to widen the groove.

The protein CBP2 binds RNA tertiary structure and is required for efficient splicing of a yeast mitochondrial Group I intron. CBP2 must wait for folding of the two RNA domains that make up the catalytic core before it can bind.[70] Subsequently, association of the 5′ domain of the RNA is stabilized by additional interactions with the protein. Thus, CBP2 functions primarily to capture otherwise transient RNA tertiary structures.

ACKNOWLEDGMENT

We thank Rebekah Gunn for obtaining copyright releases and Professor Dennis Brown for providing support. C.C.H. thanks his students, professional colleagues and mentors, Professors Jack Horowitz, Ignacio Tinoco, Jr and Paul Agris, for their contributions, insights and enthusiasm for this ''good fight''.

ABBREVIATIONS AND ACRONYMS

A	Adenine
br^{5}C	5-Bromocytidine
br^{8}G	8-Bromoguanosine
C	Cytosine
CD	Circular Dichroism
DSC	Differential Scanning Calorimetry
G	Guanine
HIV	Human Immunodeficiency Virus
IR	Infrared
m^{5}C	5-Methylcytosine
mRNA	Messenger RNA
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
SELEX	Systematic Evolution of Ligands by Exponential Enhancement
snoRNA	Small Nucleolar RNA
T	Thymine
tRNA	Transfer RNA
tRNA^{Phe}	Phenylalanyl Transfer RNA
U	Uracil
UV	Ultraviolet
WC	Watson–Crick

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance, General Medical • Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Circular Dichroism in Analysis of Biomolecules • High-performance

Liquid Chromatography of Biological Macromolecules • Infrared Spectroscopy of Biological Applications • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules • Raman Spectroscopy in Analysis of Biomolecules • Single Biomolecule Detection and Characterization • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory • Biosensor Design and Fabrication • Capillary Electrophoresis in Clinical Chemistry • Immunochemistry • Infrared Spectroscopy in Clinical Chemistry • Laboratory Instruments in Clinical Chemistry, Principles of • Molecular Biological Analyses and Molecular Pathology in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Forensic Science **(Volume 5)**
Forensic Science: Introduction • DNA Extraction Methods in Forensic Analysis • Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Capillary Electrophoresis of Nucleic Acids • Comparative Genomics: Differential Display and Subtractive Hybridization • DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Electron Tomography of Chromosome Structure • Fluorescence In Situ Hybridization • Genome Physical Mapping Using BACs • Nuclear Magnetic Resonance and Nucleic Acid Structures • Polymerase Chain Reaction and Other Amplification Systems • RNA Tertiary Structure • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Structural Analysis of Ribozymes • X-ray Structures of Nucleic Acids

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction

Peptides and Proteins **(Volume 7)**
Protein–Drug Interactions • Protein–Oligonucleotide Interactions • Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery • Robotics and Laboratory Automation in Pharmaceuticals Analysis

Polymers and Rubbers **(Volume 8)**
Atomic Force Microscopy in Analysis of Polymers

Polymers and Rubbers cont'd **(Volume 9)**
Dynamic Mechanical Analysis of Polymers and Rubbers • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Surfaces **(Volume 10)**
Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Circular Dichroism and Linear Dichroism

Infrared Spectroscopy **(Volume 12)**
Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General •

Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Raman Scattering, Fundamentals

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Structure Determination, X-ray Diffraction for

General Articles **(Volume 15)**
Archaeological Chemical Analysis

FURTHER READING

We have attempted to present an even-handed view of this vast field, but this account is limited. We included the following Further Reading to broaden the scope, yet minimize space. We have glossed over many important theoretical developments, biotechnological advances and medical applications of nucleic acids. Consult the cited articles for much more about the inner workings and uses of these fascinating molecules.

Our sentiments about nucleic acids track well with those of author Jim DeRogatis: ''Like any history, this one is subjective. Another account of psychedelic rock [Jim's subject] could be written emphasizing an entirely different set of artists (scientists in our case), and I would enjoy reading it. . . . My goal was to share my enthusiasm for music (nucleic acids studies) that I consider to be some of the finest that rock (biophysical chemistry) has produced''. [J. DeRogatis, *Kaleidoscope Eyes*, Citadel Press, Secaucus, NJ, 1996.]

Books

V.A. Bloomfield, D.M. Crothers, I. Tinoco, *Physical Chemistry of Nucleic Acids*, Harper and Row, New York, 1974.

C.R. Cantor, P.R. Schimmel, *Biophysical Chemistry*, Freeman, New York, Vols. 1–3, 1980.

S.T. Crooke, B. Lebleu, *Antisense Research and Applications*, CRC Press, Boca Raton, FL, 1993.

V.N. Soyfer, V.N. Potaman, *Triple-helical Nucleic Acids*, Springer, New York, 1996.

P.C. Turner (ed.), *Ribozyme Protocols*, Humana Press, Totowa, NJ, 1997.

P.O.P. Ts'o, *Basic Principles in Nucleic Acid Chemistry*, Academic Press, New York, Vols. 1 and 2, 1974.

Primary Literature and Reviews

Structural Microheterogeneity

R.E. Dickerson, 'Base Sequence and Helix Variation in B and A DNA', *J. Mol. Biol.*, **166**, 419–441 (1983).

Thermodynamic Measurements

G.A. Gelfand, G.E. Plum, S. Mielewczyk, D.P. Remeta, K.J. Breslauer, 'A Quantitative Method for Evaluating the Stabilities of Nucleic Acids', *Proc. Natl. Acad. Sci. USA*, **96**, 6113–6118 (1999).

C.W. Hilbers, C.A.G. Haasnoot, S.H. de Bruin, J.J.M. Joordens, G.A. Van der Marel, J.H. Van Boom, 'Hairpin Formation in Synthetic Oligonucleotides', *Biochemie*, **67**, 685–695 (1985).

J. SantaLucia, Jr, 'A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics', *Proc. Natl. Acad. Sci. USA*, **95**, 1460–1465 (1998).

R.D. Sheardy, D. Suh, R. Kurzinsky, M.J. Doktycz, A.S. Benight, J.B. Chaires, 'Sequence Dependence of the Free Energy of B–Z Junction Formation in Deoxyoligonucleotides', *J. Mol. Biol.*, **231**, 475–488 (1993).

N. Sugimoto, S. Nakano, K. Katoh, A. Matsumura, H. Nakamura, T. Ohmichi, M. Yoneyama, M. Sasaki, 'Thermodynamic Parameters to Predict Stability of RNA/DNA Hybrid Duplexes', *Biochemistry*, **34**, 11 211–11 216 (1995).

L.E. Xodo, G. Manzini, F. Quadrifoglio, G. Van der Marel, J.H. Van Boom, 'Hairpin Structures in Synthetic Oligodeoxynucleotides: Sequence Effects on the Duplex-to-hairpin Transition', *Biochimie*, **71**, 793–803 (1989).

Hydration, Activation Volume Determination

T.V. Chalikian, J. Volker, W.K. Olson, K.J. Breslauer, 'The Hydration of Nucleic Acid Duplexes as Assessed by a Combination of Volumetric and Structural Techniques', *Biopolymers*, **50**, 347–359 (1999).

T.V. Chalikian, J. Volker, G.E. Plum, K.J. Breslauer, 'A More Unified Picture for the Thermodynamics of Nucleic Acid Duplex Melting: a Characterization of Calorimetric and Volumetric Techniques', *Proc. Natl. Acad. Sci. USA*, **96**, 7853–7858 (1999).

M.-C. Lin, R.B. MacGregor, Jr, 'Activation Volume of DNA Duplex Formation', *Biochemistry*, **36**, 6539–6544 (1997).

Sequence-dependent DNA Curvature

S.S. Chan, K.J. Breslauer, M.E. Hogan, D.J. Kessler, 'Physical Studies of DNA Premelting Equilibria in Duplexes with and without Homo dA•dT Tracts: Correlations with DNA Bending', *Biochemistry*, **29**, 6161–6171 (1990).

P.J. Hagerman, 'Sequence Dependence of the Curvature of DNA: a Test of the Phasing Hypothesis', *Biochemistry*, **24**, 7033–7037 (1985).

I. Rouzina, V.A. Bloomfield, 'DNA Bending by Small, Mobile Multivalent Cations', *Biophys. J.*, **74**, 3152–3164 (1998).

Divalent Cations, DNA Packaging and Aggregation

J. Duguid, V.A. Bloomfield, J. Benevides, G.J. Thomas, Jr, 'Raman Spectroscopy of DNA–Metal Complexes. I. Interactions and Conformational Effects of the Divalent Cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd', *Biophys. J.*, **65**, 1916–1928 (1993).

J.H. Shibata, J.M. Schurr, 'A Theory of Aggregation in the Thermal Denaturation Region of Multistrand Biopolymers', *Biopolymers*, **20**, 525–549 (1981).

E.K. Wilson, 'A Meeting of Minds (How DNA Packages Itself)', *Chem. Eng. News*, Aug. 2, 24–25 (1999).

Structural Transformations

S.W. Blume, V. Guarcello, W. Zacharias, D.M. Miller, 'Divalent Transition Metal Cations Counteract Potassium-induced Quadruplex Assembly of Oligo(dG) Sequences', *Nucleic Acids Res.*, **25**, 617–625 (1997).

S. Fujii, A.H.-J. Wang, G. van der Marel, J.H. van Boom, A. Rich, 'Molecular Structure of $(m^5dC-dG)_3$: the Role of the Methyl Group on 5-Methyl Cytosine in Stabilizing Z-DNA', *Nucleic Acids Res.*, **10**, 7879–7892 (1982).

C.C. Hardin, M.J. Corregan, D.V. Lieberman, B.A. Brown, II, 'Allosteric Interactions Between DNA Strands and Monovalent Cations in DNA Quadruplex Assembly: Thermodynamic Evidence for Three Linked Association Pathways', *Biochemistry*, **36**, 15 428–15 458 (1997).

T.J. Povsic, P.B. Dervan, 'Triple-helix Formation by Oligonucleotides on DNA Extended to the Physiological pH Range', *J. Am. Chem. Soc.*, **111**, 3059–3061 (1989).

R.D. Sheardy, N. Levine, S. Marotta, D. Suh, J.B. Chaires, 'A Thermodynamic Investigation of the Melting of the B–Z Junction Forming DNA Oligomers', *Biochemistry*, **33**, 1385–1391 (1994).

S.F. Singleton, P.B. Dervan, 'Temperature Dependence of the Energetics of Oligonucleotide-directed Triple-helix Formation at a Single DNA Site', *J. Am. Chem. Soc.*, **116**, 10 376–10 382 (1994).

Plasmid Supercoiling, Transformations

M.J. Ellison, R.J. Kelleher, III, A.H.-J. Wang, J.F. Habener, A. Rich, 'Sequence-dependent Energetics of the B–Z Transition in Supercoiled DNA-containing Nonalternating Purine–Pyrimidine Sequences', *Proc. Natl. Acad. Sci. USA*, **82**, 8320–8324 (1985).

D.M.J. Lilley, K.M. Sullivan, A.I.H. Murchie, 'The Extrusion of Cruciform Structures in Supercoiled DNA – Kinetics and Mechanism', in *Nucleic Acids and Molecular Biology*, eds. F. Eckstein, D.M.J. Lilley, Springer, Heidelberg, Vol. 1, 1987.

J.M. Schurr, J.J. Delrow, B.S. Fujimoto, A.S. Benight, 'The Question of Long-range Allosteric Transitions in DNA', *Biopolym. Nucleic Acids Sci.*, **44**, 283–308 (1998).

J. Volker, R.D. Blake, S.G. Delacourt, K.J. Breslauer, 'High-resolution Calorimetric and Optical Melting Profiles of DNA Plasmids: Resolving Contributions from Intrinsic Melting Domains and Specifically Designed Inserts', *Biopolymers*, **50**, 303–318 (1999).

RNA Catalysis, Divalent Cations

J. Ciesiolka, M. Yarus, 'Small RNA–Divalent Domains', *RNA*, **2**, 785–793 (1996).

S.C. Dahm, W.B. Derrick, O.C. Uhlenbeck, 'Evidence for the Role of Solvated Metal Hydroxide in the Hammerhead Cleavage Mechanism', *Biochemistry*, **32**, 13 040–13 045 (1993).

Protein–DNA Recognition

J.M. Berg, Y. Shi, 'The Galvanization of Biology: a Growing Appreciation for the Roles of Zinc', *Science*, **271**, 1081–1085 (1996).

P. Howard-Flanders, S.C. West, A. Stasiak, 'Role of RecA Protein Spiral Filaments in Genetic Recombination', *Nature (London)*, **309**, 215–219 (1984).

P.B. McIntosh, T.A. Frenkiel, U. Wollborn, J.E. McCormick, K.-H. Klempnauer, J. Feeney, M.D. Carr, 'Solution Structure of the B-Myb DNA-binding Domain: a Possible Role for Conformational Instability of the Protein in DNA Binding and Control of Gene Expression', *Biochemistry*, **37**, 9619–9629 (1998).

K. Nadassy, S. Wodak, J. Janin, 'Structural Features of Protein–Nucleic Acid Recognition Sites', *Biochemistry*, **38**, 1999–2017 (1999).

M. Suzuki, M. Gerstein, N. Yagi, 'Stereochemical Basis of DNA Recognition by Zn Fingers', *Nucleic Acids Res.*, **22**, 3397–3405 (1994).

J.C. Wang, 'DNA Topoisomerases', *Annu. Rev. Biochem.*, **65**, 635–692 (1996).

Antisense-mediated Inhibition of DNA Functions

H.-G. Kim, J.F. Reddoch, C. Mayfield, S. Ebbinghaus, N. Vigneswaran, S. Thomas, D.E. Jones, Jr, D.M. Miller, 'Inhibition of Transcription of the Human C-myc Protooncogene by Intermolecular Triplex', *Biochemistry*, **37**, 2299–2304 (1998).

K. Nordstrom, E.G.H. Wagner, 'Kinetic Aspects of Control of Plasmid Replication by Antisense RNA', *Trends Biochem. Sci.*, **19**, 294–300 (1994).

Theories: Ligand–and Protein–Nucleic Acid Interactions, Locomotion, Stretching

M.T. Record, T.M. Lohman, P. De Haseth, 'Ion Effects on Ligand–Nucleic Acid Interaction', *J. Mol. Biol.*, **107**, 145–158 (1976).

S.B. Smith, Y. Cui, C. Bustamante, 'Overstretching B-DNA: the Elastic Response of Individual Double-stranded and Single-stranded DNA Molecules', *Science*, **271**, 795–799 (1996).

P.H. Von Hippel, O.G. Berg, 'Facilitated Target Location in Biological Systems', *J. Biol. Chem.*, **264**, 675–678 (1989).

H.-Y. Wang, T. Elston, A. Mogilner, G. Oster, 'Force Generation in RNA Polymerase', *Biophys. J.*, **74**, 1186–1202 (1998).

RNA Tertiary Folding, Protein Binding

P. Brion, E. Westhof, 'Hierarchy and Dynamics of RNA Folding', *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137 (1997).

E.A. Doherty, D. Hershlag, J.A. Doudna, 'Assembly of an Exceptionally Stable RNA Tertiary Interface in a Group I Ribozyme', *Biochemistry*, **38**, 2982–2990 (1999).

D.E. Draper, 'Protein–RNA Recognition', *Annu. Rev. Biochem.*, **64**, 593–620 (1995).

T.C. Gluick, N.M. Wills, R.F. Gesteland, D.E. Draper, 'Folding of an mRNA Pseudoknot Required for Stop Codon Readthrough: Effects of Mono- and Divalent Ions on Stability', *Biochemistry*, **36**, 16 173–16 186 (1997).

T.M. Lohman, K.P. Bjornson, 'Mechanisms of Helicase-catalyzed DNA Unwinding', *Annu. Rev. Biochem.*, **65**, 169–214 (1996).

E.S. Maxwell, M.J. Fournier, 'The Small Nucleolar RNAs', *Annu. Rev. Biochem.*, **64**, 897–934 (1995).

K.M. Weeks, D.M. Crothers, 'RNA Binding Assays for Tat-derived Peptides: Implications for Specificity', *Biochemistry*, **31**, 10 281–10 287 (1992).

G. Zacchai, S.-Y. Xian, 'Structure of Phenylalanine-accepting tRNA and its Environment in Aqueous Solvents with Different Salts', *Biochemistry*, **27**, 1313–1320 (1988).

P.P. Zarrinkar, J.R. Williamson, 'Kinetic Intermediates in RNA Folding', *Science*, **265**, 819–924 (1994).

Sequence Matrix Scanning and Detection, Functional Selection Strategies

S. Borman, 'DNA Chips Come of Age', *Chem. Eng. News*, Dec. 9, 42–43 (1996).

V. Chan, D.J. Graves, S.E. McKenzie, 'The Biophysics of DNA Hybridization with Immobilized Oligonucleotide Probes', *Biophys. J.*, **69**, 2243–2255 (1995).

J.G. Duguid, C. Li, M. Shi, M.J. Logan, H. Alila, A. Rolland, E. Tomlinson, J.T. Sparrow, L.C. Smith, 'A Physicochemical Approach for Predicting the Effectiveness of Peptide-based Gene Delivery Systems for Use in Plasmid-based Gene Therapy', *Biophys. J.*, **74**, 2802–2814 (1998).

Nucleic Acid Nanostructures and Machines

C. Doering, B. Ermentrout, G. Oster, 'Rotary DNA Motors', *Biophys. J.*, **69**, 2256–2267 (1995).

N.S. Halim, 'DNA: from Living Cells to Electronic Hardware', *Scientist*, Aug. 16, 14 (1999).

N.C. Seeman, 'DNA Nanotechnology', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 225–248 (1998).

E.K. Wilson, 'DNA Conductance Convergence?', *Chem. Eng. News*, Aug. 23, 43–48 (1999).

E.K. Wilson, 'Instant DNA Detection', *Chem. Eng. News*, May 25, 47–49 (1998).

Nuclear Magnetic Resonance Analysis of Nucleic Acid Structure and Dynamics

O.Y. Federoff, M. Salazar, B.R. Reid, 'Structural Variation Among Retroviral Primer–DNA Junctions: Solution Structure of the HIV-1 (–)-Strand Okazaki Fragment r(gcca) d(CTGC)•d(GCAGTGGC)', *Biochemistry*, **35**, 11 070–11 080 (1996).

A.N. Lane, 'Determination of Fast Dynamics of Nucleic Acids by NMR', *Methods Enzymol.*, **261**, 413–435 (1995).

P.B. Moore, 'Determination of RNA Conformation by NMR', *Acc. Chem. Res.*, **28**, 251–256 (1995).

B.H. Robinson, G.P. Drobney, 'Site-specific Dynamics in DNA', *Methods Enzymol.*, **261**, 451–509 (1995).

N.B. Ulyanov, V.I. Ivanov, E.E. Minbyat, E.B. Khomyakova, M.V. Petrova, K. Lesiak, T.L. James, 'A Pseudosquare Knot Structure of DNA in Solution', *Biochemistry*, **37**, 12 715–12 726 (1998).

G. Varani, K. Nagai, 'RNA Recognition by RNP Proteins', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 407–502 (1998).

Kinetic Studies: Codon–Anticodon RNA and DNA Duplex Associations, Anticodon Dynamics, Ribozyme Substrate Binding

P.C. Bevilacqua, R. Kierzek, K.A. Johnson, D.H. Turner, 'Dynamics of Ribozyme Binding of Substrate Revealed by Fluorescence-detected Stopped-flow Methods', *Science*, **258**, 1355–1358 (1992).

W.H. Braunlin, V.A. Bloomfield, '^{1}H NMR Study of the Base-pairing Reactions of d(GGAATTCC): Salt Effects on the Equilibria and Kinetics of Strand Association', *Biochemistry*, **30**, 754–758 (1991).

J.L. Bresloff, D.M. Crothers, 'DNA–Ethidium Reaction Kinetics: Demonstration of Direct Ligand Transfer between DNA Binding Sites', *J. Mol. Biol.*, **95**, 103–123 (1975).

M. Ehrenberg, R. Rigler, W. Wintermeyer, 'On the Structure and Conformational Dynamics of Yeast Phenylalanine-accepting Transfer Ribonucleic Acid in Solution', *Biochemistry*, **18**, 4588–4599 (1979).

A.P. Williams, C.E. Longfellow, S.M. Freier, R. Kierzek, D.H. Turner, 'Laser Temperature-jump, Spectroscopic, and Thermodynamic Study of Salt Effects on Duplex Formation by dGCATGC', *Biochemistry*, **28**, 4283–4291 (1989).

K. Yoon, D.H. Turner, I. Tinoco, Jr, 'The Kinetics of Codon–Anticodon Interaction in Yeast Phenylalanine Transfer RNA', *J. Mol. Biol.*, **99**, 507–518 (1975).

Theories: Quantum Mechanics, Electrostatic Potential and Superhelical Fluctuations

S.C. Harvey, D. Sprous, 'Action at a Distance in Supercoiled DNA', *Biophys. J.*, **70**, 1893–1903 (1996).

P. Hobza, C. Sandorfy, 'Nonempirical Calculations on All the 29 Possible DNA Base Pairs', *J. Am. Chem. Soc.*, **109**, 1302–1307 (1987).

T. Schlick, B. Li, W.K. Olson, 'The Influence of Salt on the Structure and Energetics of Supercoiled DNA', *Biophys. J.*, **67**, 2146–2166 (1994).

T. Schlick, W.K. Olson, 'Supercoiled DNA Energetics and Dynamics by Computer Simulation', *J. Mol. Biol.*, **223**, 1089–1119 (1992).

K.A. Sharp, B. Honig, S.C. Harvey, 'Electric Potential of Transfer RNAs: Codon–Anticodon Recognition', *Biochemistry*, **29**, 340–346 (1990).

Molecular Mechanics and Dynamics, Free Energy Perturbation Analyses: Theories and Implementation

C.I. Bayley, P. Cieplak, W.D. Cornell, P.A. Kollman, 'A Well Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model', *J. Chem. Phys.*, **97**, 10 269–10 280 (1993).

B.R. Brooks, R.E. Broccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, 'CHARMM: a Program for Macromolecular Energy Minimization and Dynamics Calculations', *J. Comput. Chem.*, **4**, 187–217 (1983).

D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, 'AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules', *Comput. Phys. Commun.*, **91**, 1–41 (1995).

S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr, P. Weiner, 'A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins', *J. Am. Chem. Soc.*, **106**, 765–784 (1984).

Free Energy Perturbation Analysis Applications

W.D. Cornell, P. Cieplak, C.I. Bayley, P.A. Kollman, 'Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation', *J. Am. Chem. Soc.*, **115**, 9620–9631 (1993).

D.A. Pearlman, P.A. Kollman, 'The Calculated Free Energy Effects of 5-Methyl Cytosine on the B to Z Transition in DNA', *Biopolymers*, **29**, 1093–1209 (1990).

W.S. Ross, C.C. Hardin, 'Ion-induced Stabilization of the G-DNA Quadruplex: Free Energy Perturbation Studies', *J. Am. Chem. Soc.*, **116**, 6070–6080 (1994).

S.B. Singh, P.A. Kollman, 'Understanding the Thermodynamic Stability of an RNA Hairpin and its Mutant', *Biophys. J.*, **70**, 1940–1948 (1996).

REFERENCES

1. W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, 1984.
2. R.F. Weaver, *Molecular Biology*, WCB/McGraw-Hill, Boston, MA, 1999.
3. L. Moran, K. Scrimgeour, H. Horton, R. Ochs, J. Rawn, *Biochemistry*, Prentice-Hall, Upper Saddle River, NJ, 1994.
4. G. Bernardi, 'The Isochore Organization of the Human Genome', *Annu. Rev. Genet.*, **23**, 637–661 (1989).
5. D. Thirumalai, S.A. Woodson, 'Kinetics of Folding of Proteins and RNA', *Acc. Chem. Res.*, **29**, 433–439 (1996).
6. A.V. Fratini, M.L. Kopka, H.R. Drew, R.E. Dickerson, 'Reversible Bending and Helix Geometry in a B-DNA Dodecamer: CGCGAATT^{Br}CGCG', *J. Biol. Chem.*, **257**, 14 686–14 707 (1982).
7. J.R. Wyatt, J.D. Puglisi, I. Tinoco, Jr, 'RNA Folding: Pseudoknots, Loops and Bulges', *Bioessays*, **11**(4), 100–106 (1989).
8. J. Santa Lucia, R. Kierzek, D.H. Turner, 'Stabilities of Consecutive A–C, C–C, G–G, U–C, and U–U Mismatches in RNA Internal Loops: Evidence for Stable Hydrogen-bonded U–U and C–C$^+$ Pairs', *Biochemistry*, **30**, 8242–8251 (1991).
9. A. Rich, A. Nordheim, A.-H. Wang, 'Z-DNA', *Annu. Rev. Biochem.*, **53**, 791–846 (1984).
10. Y. Kohwi, T. Kohwi-Shigematsu, 'Magnesium Ion-dependent Triple-helix Structure Formed by Homopurine–Homopyrimidine Sequences in Supercoiled Plasmid DNA', *Proc. Natl. Acad. Sci. USA*, **85**, 3781–3785 (1988).
11. S.M. Mirkin, M.D. Frank-Kamenetskii, 'H-DNA and Related Structures', *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 541–576 (1994).
12. S.F. Singleton, P.B. Dervan, 'Influence of pH on the Equilibrium Association Constants for Oligodeoxyribonucleotide-directed Triple Helix Formation at Single DNA Sites', *Biochemistry*, **31**, 10 995–11 003 (1992).
13. C.C. Hardin, M. Corregan, B.A. Brown, II, L. Frederick, 'Cytosine•Cytosine Base-pairing Stabilizes DNA Quadruplexes and Cytosine Methylation Greatly Enhances the Effect', *Biochemistry*, **32**, 5870–5880 (1993).
14. C.C. Hardin, E. Henderson, T. Watson, J.K. Prosser, 'Monovalent Cation Induced Structural Transitions in Telomeric DNAs: G-DNA Folding Intermediates', *Biochemistry*, **30**, 4460–4472 (1991).
15. T. Miura, J.M. Benevides, G.J. Thomas, 'A Phase Diagram for Sodium and Potassium Ion Control of Polymorphism in Telomeric DNA', *J. Mol. Biol.*, **248**, 233–238 (1995).
16. J.R. Williamson, 'G-Quartet Structures in Telomeric DNA', *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703–730 (1994).
17. R. Jin, K.J. Breslauer, R.A. Jones, B.L. Gaffney, 'Tetraplex Formation of a Guanine-containing Nonameric DNA Fragment', *Science*, **250**, 543–546 (1990).
18. H. Deng, W.H. Braunlin, 'Duplex to Quadruplex Equilibrium of the Self-complementary Oligonucleotide d(GGGGCCCC)', *Biopolymers*, **35**, 677–681 (1995).
19. W.M. Olivas, L.J. Maher, 'Competitive Triplex/Quadruplex Equilibria Involving Guanine-rich Oligonucleotides', *Biochemistry*, **34**, 278–284 (1995).

20. C.C. Hardin, T. Watson, M. Corregan, C. Bailey, 'Cation-dependent Transition Between the Quadruplex and Watson–Crick Hairpin Forms of d(CGCG$_3$GCG)', *Biochemistry*, **31**, 833–841 (1992).
21. J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, A.A. Szewczak, C.E. Kundrot, T.R. Cech, J.A. Doudna, 'RNA Tertiary Structure Mediation by Adenosine Platforms', *Science*, **273**, 1696–1699 (1996).
22. K. Gehring, J. Leroy, M. Gueron, 'A Telomeric DNA Structure with Protonated Cytosine•Cytosine Base Pairs', *Nature (London)*, **363**, 561–565 (1993).
23. D.J. Patel, 'Tetrads Through Interdigitation', *Nature (London)*, **363**, 499–500 (1993).
24. A. Yu, M.C. Barron, R.M. Romero, M. Christy, B. Gold, J. Dai, D. Gray, I. Haworth, M. Mitas, 'At Physiological pH, d(CCG)$_{15}$ Forms a Hairpin Containing Protonated Cytosines and a Distorted Helix', *Biochemistry*, **36**, 3687–3699 (1997).
25. J. Leroy, K. Gehring, A. Kettani, M. Gueron, 'Acid Multimers of Oligodeoxycytidine Strands: Stoichiometry, Base-pair Characterization, and Proton Exchange Properties', *Biochemistry*, **32**, 6019–6031 (1993).
26. L.A. Marky, N.R. Kallenbach, K.A. McDonough, N.C. Seeman, K.J. Breslauer, 'The Melting Behavior of a DNA Junction Structure: a Calorimetric and Spectroscopic Study', *Biopolymers*, **26**, 1621–1634 (1987).
27. M. Rouhi, 'Molecular Machines', *Chem. Eng. News*, Jan. 18, 13 (1999).
28. X. Gao, X. Huang, G.K. Smith, M. Zheng, H. Liu, 'New Antiparallel Duplex Motif or DNA CCG Repeats that is Stabilized by Extrahelical Bases Symmetrically Located in the Minor Groove', *J. Am. Chem. Soc.*, **117**, 8883–8884 (1995).
29. A.M. Gacy, C.T. McMurray, 'Influence of Hairpins on Template Reannealing at Trinucleotide Repeat Duplexes: a Model for Slipped DNA', *Biochemistry*, **37**, 9426–9434 (1998).
30. J.D. Puglisi, I. Tinoco, 'Absorbance Melting Curves of RNA', *Methods Enzymol.*, **180**, 304–325 (1989).
31. P. Cruz, K.B. Hall, J. Puglisi, P. Davis, C.C. Hardin, M.O. Trulson, R.A. Mathies, I. Tinoco, Jr, W.C. Johnson, Jr, T. Neilson, 'The Left-handed Z-form of Double-stranded RNA', in *Biomolecular Stereodynamics IV*, eds. R.H. Sarma, M.H. Sarma, Adenine Press, Albany, NY, 179–200, 1985.
32. C.L. Stevens, G. Felsenfeld, 'The Conversion of Two-stranded Poly(A + U) to Three-strand Poly(A + 2U) and Poly(A) by Heat', *Biopolymers*, **2**, 293–314 (1964).
33. J.H. Cate, A.R. Gooding, E. Podell, K.L. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, J.A. Doudna, 'Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing', *Science*, **273**, 1625–1764 (1996).
34. L.X. Shen, I. Tinoco, Jr, 'The Structure of an RNA Pseudoknot that Causes Efficient Frameshifting in Mouse Mammary Tumor Virus', *J. Mol. Biol.*, **247**, 963–978 (1995).
35. H.A. Heus, A. Pardi, 'Structural Features that Give Rise to the Unusual Stability of RNA Hairpins Containing GNRA Loops', *Science*, **253**, 191–193 (1991).
36. D. Voet, J.G. Voet, *Biochemistry*, John Wiley & Sons, New York, 1990.
37. M. Laughrea, L. Jette, 'HIV-α Genome Dimerization: Kissing-loop Hairpin Dictates Whether Nucleotides Downstream of the 5′ Splice Junction Contribute to Loose and Tight Dimerization of Human Immunodeficiency Virus RNA', *Biochemistry*, **36**, 9501–9508 (1997).
38. O.C. Uhlenbeck, 'A Small Catalytic Oligonucleotide', *Nature (London)*, **328**, 596–600 (1987).
39. (a) V.V. Filimonov, 'Thermodynamics of Conformation Transitions in Polynucleotides', in *Thermodynamic Data for Biochemistry and Biotechnology*, ed. H.-J. Hinze, Springer, New York, Chapter 14, 1986; (b) K.J. Breslauer, 'Methods for Obtaining Thermodynamic Data on Oligonucleotide Transitions', in *Thermodynamic Data for Biochemistry and Biotechnology*, ed. H.-J. Hinze, Springer, New York, Chapter 15, 1986; see also Chapters 4 and 5.
40. L.A. Marky, K.J. Breslauer, 'Calculating Thermodynamic Data for Transitions of Any Molecularity from Equilibrium Melting Curves', *Biopolymers*, **26**(9), 1601–1620 (1987).
41. K.J. Breslauer, R. Frank, H. Blocker, L.A. Marky, 'Predicting DNA Duplex Stability from the Base Sequence', *Proc. Natl. Acad. Sci. USA*, **83**, 3746–3750 (1986).
42. C.R. Calladine, H.R. Drew, 'Principles of Sequence-dependent Flexure of DNA', *J. Mol. Biol.*, **192**, 907–918 (1986).
43. C.F. Bernasconi, *Relaxation Kinetics*, Academic Press, New York, 1994.
44. J. Wyman, S.J. Gill, *Binding and Linkage*, University Science Books, Mill Valley, CA, 1990.
45. K.E. Van Holde, W.C. Johnson, P.S. Ho, *Physical Biochemistry*, Prentice-Hall, Upper Saddle River, NJ, 1998.
46. A. Ben-Naim, *Hydrophobic Interactions*, Plenum Press, New York, 1980.
47. M. Singer, P. Berg, *Genes and Genomes*, University Science Books, Mill Valley, CA, 1991.
48. I.D. Campbell, R.A. Dwek, *Biological Spectroscopy*, Benjamin/Cummings, Menlo Park, CA, 1984.
49. J.A. Jaeger, J. SantaLucia, Jr, I. Tinoco, Jr, 'Determination of RNA Structure and Thermodynamics', *Annu. Rev. Biochem.*, **62**, 255–287 (1993).
50. K. Wuthrich, *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York, 1986.
51. O. Jardetsky, G.C.K. Roberts, *NMR in Molecular Biology*, Academic Press, Orlando, FL, 1981.
52. G. Varani, I. Tinoco, Jr, 'RNA Structure and NMR Spectroscopy', *Q. Rev. Biophys.*, **24**(4), 479–532 (1991).
53. U. Schmitz, T.L. James, 'How to Generate Accurate Solution Structures of Double-helical Nucleic Acid Fragments Using NMR and Restrained Molecular Dynamics', *Methods Enzymol.*, **261**, 3–44 (1995).

54. K.M. Weeks, D.M. Crothers, 'RNA Binding Assays for Tat-derived Peptides: Implications for Specificity', *Biochemistry*, **31**, 10 281–10 287 (1992).
55. A. Travers, *DNA–Protein Interactions*, Chapman and Hall, New York, 1993.
56. D.M. Ribeiro, D.A. Carson, 'Ca^{2+}/Mg^{2+}-Dependent Endonuclease from Human Spleen: Purification, Properties and Role in Apoptosis', *Biochemistry*, **32**, 9129–9136 (1993).
57. A. Wyllie, 'An Endonuclease at Last', *Nature (London)*, **391**, 20–21 (1998).
58. A.A. Szewczak, E.R. Podell, P.C. Bevilacqua, T.R. Cech, 'Thermodynamic Stability of the P4–P6 Domain RNA Tertiary Structure Measured by Temperature Gradient Gel Electrophoresis', *Biochemistry*, **37**, 11 162–11 170 (1998).
59. M.T. Record, Jr, C.F. Anderson, T.M. Lohman, 'Thermodynamic Analysis of Ion Effects on the Binding and Conformational Equilibria of Proteins and Nucleic Acids: the Roles of Ion Association or Release, Screening, and Ion Effects on Water Activity', *Q. Rev. Biophys. II*, **2**, 103–178 (1978).
60. W.S. Ross, C.C. Hardin, I. Tinoco, Jr, S.N. Rao, D.A. Pearlman, P.A. Kollman, 'Effects of Nucleotide Bromination on the Stabilities of Z-RNA and Z-DNA: a Molecular Mechanics/Thermodynamic Perturbation Study', *Biopolymers*, **28**, 1939–1957 (1989).
61. J.G. Duguid, V.A. Bloomfield, 'Aggregation of Melted DNA by Divalent Metal Ion-mediated Cross-linking', *Biophys. J.*, **69**, 2642–2648 (1995).
62. C.C. Hardin, M.J. Corregan, unpublished results.
63. C.H. Spink, J.B. Chaires, 'Effects of Hydration, Ion Release, and Excluded Volume on the Melting of Triplex and Duplex DNA', *Biochemistry*, **38**, 496–508 (1999).
64. P.J. Hagerman, 'Sequence-directed Curvature of DNA', *Annu. Rev. Biochem.*, **59**, 755–781 (1990).
65. D. Irvine, C. Tuerk, L. Gold, 'Selexion: Systematic Evolution of Ligands by Exponential Enrichment with Integrated Optimization by Non-linear Analysis', *J. Mol. Biol.*, **222**, 739–761 (1991).
66. L. Gold, B. Polisky, O. Uhlenbeck, M. Yarus, 'Diversity of Oligonucleotide Functions', *Annu. Rev. Biochem.*, **64**, 763–797 (1995).
67. J.D. McGhee, P.H. von Hippel, 'Theoretical Aspects of DNA–Protein Interactions: Cooperative and Non-cooperative Binding of Large Ligands to a One-dimensional Homogeneous Lattice', *J. Mol. Biol.*, **86**, 469–489 (1974).
68. J.D. Puglisi, R. Tan, B.J. Calnan, A.D. Frankel, J.R. Williamson, 'Conformation of the TAR RNA–Arginine Complex by NMR Spectroscopy', *Science*, **257**, 76–80 (1992).
69. J.L. Battiste, H. Mao, N.S. Rao, R. Tan, D.R. Muhandiram, L.E. Kay, A.D. Frankel, J.R. Williamson, 'α Helix–RNA Major Groove Recognition in an HIV-1 Rev Peptide–RRE RNA Complex', *Science*, **273**, 1547–1551 (1996).
70. K.M. Weeks, T.R. Cech, 'Assembly of Ribonucleoprotein Catalyst by Tertiary Structure Capture', *Science*, **271**, 345–348 (1996).
71. S.M. Freier, N. Sugimoto, A. Sinclair, D. Alkema, T. Neilson, R. Kierzek, M.H. Caruthers, D.H. Turner, *Biochemistry*, **25**, 3214–3219 (1986b).

Optical Mapping in Genomic Analysis

Christopher Aston
Wyeth-Ayerst Research, Princeton, NJ, USA

David C. Schwartz
University of Wisconsin, Madison, W1, USA

1 Introduction 1
1.1 Principle 1
1.2 History 2
1.3 Optical Mapping of Cloned DNA 2
1.4 Optical Mapping of Genomic DNA 3
2 Methodology 3
2.1 Cleaning of Glass Surfaces 3
2.2 Derivatization of Glass Surfaces 3
2.3 Preparation of DNA Molecules 3
2.4 Mounting of DNA Molecules on Surfaces 4
2.5 Restriction Digestion of DNA Molecules on Surfaces 5
2.6 Choice of Restriction Enzyme 5
2.7 Sizing of DNA Molecules 5
2.8 Instrumentation 5
2.9 Microscope Automation Software and Collection of Images 6
2.10 Correction of Images 7
2.11 Analysis of Images 7
2.12 Map Construction Using Cloned DNA 8
2.13 Map Construction Using Genomic DNA 9
2.14 Visualization of Map Data 9
3 Optical Mapping to Characterize Genomes 9
3.1 Optical Mapping for Sequence Assembly and Verification 9
3.2 Very High-resolution Optical Mapping 10
3.3 Identification of Repeated Regions 11
3.4 Identification of Genic Regions 11
3.5 Optical Mapping for Population-based Genomic Studies 11
4 Comparison with Other Physical Mapping Approaches 11
4.1 Gel Electrophoresis 11
4.2 Fluorescent In Situ Hybridization 11
4.3 Hybridization-based Microarrays 12
4.4 Other Single Molecule Approaches 12
5 Recent Improvements and Future Developments 12
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 13
References 13

Optical mapping is a system for the construction of ordered restriction maps from single deoxyribonucleic acid (DNA) molecules. DNA molecules bound to derivatized glass surfaces are cleaved with restriction enzymes and imaged by fluorescence microscopy. Cut sites are visualized as gaps between cleaved DNA fragments, which retain their original order. Optical mapping offers advantages over electrophoretic methods of restriction enzyme analysis, which require separate techniques for separating cleaved DNA fragments and determining fragment order. Furthermore, small amounts of DNA are required for optical mapping analysis. By determining the existence of these sequence-specific cut sites and the distances between them, a precise landmark map of the DNA sequence can be constructed. Prokaryote and eukaryote DNA can be mapped. Maps can be constructed from both cloned DNA and genomic DNA. Restriction maps can be constructed for whole genomes by overlapping maps derived from individual molecules. Such maps provide a useful scaffold for accurate alignment of sequence contigs generated by a shotgun sequencing approach and a means of comparing populations and strains.

1 INTRODUCTION

1.1 Principle

Optical mapping is a single molecule approach for the rapid construction of ordered restriction maps.[1] It dispenses with electrophoretic approaches and uses fluorescence microscopy directly to image individual DNA molecules bound to specially derivatized glass surfaces which have been cleaved by restriction enzymes. Restriction enzymes are enzymes that recognize and then cleave DNA at a sequence specific site, usually a four, six or eight nucleotide motif. Different restriction enzymes recognize different motifs. A restriction map is a linear array of sites on DNA cleaved by one or more restriction enzymes showing the mass, in basepairs (bp), of the ordered fragments predicted from the digest.

Large DNA molecules, greater than 10 kilobases (kb) in size, are mounted on silanized glass surfaces. Fluid flows elongate the DNA molecules and electrostatic interactions hold the stretched DNA molecules on

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the surface. Molecules are digested with a restriction enzyme. Cleavage sites are visible as small gaps between DNA fragments, which retain their original order. DNA is stained with an intercalating dye, visualized by fluorescence microscopy and digital images are recorded. Fragments are sized by measuring integrated fluorescence intensity and ordered restriction maps of single molecules are constructed. A restriction map created from a single DNA molecule is limited in its accuracy by the resolution of the microscope, the imaging system, surface conditions and the biochemistries used. Statistical approaches are therefore used to construct finished maps from an ensemble of imperfect maps such as those derived from partially digested molecules. Consensus maps of cloned DNA are created using a Bayesian inference scheme.[2] Maps of genomic DNA are contiged using Gentig software.[3] The final maps are an average of many individual maps.

1.2 History

Optical mapping was first reported in 1992. In the original method, ensembles of yeast chromosomes fluorescently labeled with 4,6-diamidino-2 phenylindole dihydrochloride (DAPI) were elongated in a flow of molten agarose generated between a coverslip and a microscope slide.[4,5] The DNA was fixed in an elongated state when the agarose gelled. Restriction enzymes were present in the molten agarose and the DNA was cut when magnesium ions diffused into the gel. As the DNA relaxed, the cut sites in individual molecules were resolved by imaging the receding cut ends by time-lapse video microscopy. Ordered restriction maps were constructed by measuring the relative fluorescence intensity and apparent molecular contour length of a number of DNA fragments, followed by averaging. Fragment sizes determined by optical mapping correlated with sizes obtained by pulsed field gel electrophoresis (PFGE)[6,7] and were more informative since the fragment order was determined concurrently. The approach had other advantages over PFGE in that problems associated with disambiguation of co-migrating bands on pulsed field gels were obviated.

In an advanced approach, DNA was digested on open glass surfaces which had been coated with polylysine.[8] DNA was bound to the surface by electrostatic interaction. The interaction was strong enough to retain cut fragments on the surface but weak enough to allow cut ends to relax to form visible gaps. Since restriction enzyme activity can be inhibited by some DNA binding fluorochromes,[9] DNA fragments were stained with ethidium homodimer fluorochrome following digestion. DNA fragments were imaged with a charge-coupled device (CCD) camera and fragments as small as 800 bp[8] could be imaged and sized, representing a major improvement from the original agarose gel-based optical mapping approach which had a limit of resolution of 60 kb.[5] The current approach uses open silanized glass surfaces which reliably retain small fragments in the 1–10 kb range.[10,11]

The approach has been developed in a number of other areas and can now be used to map a wide variety of clone types including λ-bacteriophage,[8] bacterial artificial chromosomes (BACs),[10] yeast artificial chromosomes (YACs)[12] as well as genomic DNA from prokaryotes[13] and eukaryotes.[14] The range of effective restriction enzymes has increased,[10] image acquisition has accelerated and data analysis has been refined.[2,11] Samples can now be gridded onto the surface enabling parallel processing.[11] These developments have increased throughput and provide the foundation for automated approaches to mapping DNA.

1.3 Optical Mapping of Cloned DNA

Mapping of cloned DNA offers advantages in that multiple copies of the same sequence are present on the same surface (Figure 1). Statistical approaches are used to construct finished maps from an ensemble of imperfect maps derived from individual molecules. Imperfect maps are generated as a consequence of partial digestion, desorbtion of small fragments, inaccurate sizing, etc. The final optical map is averaged from 20–40 single molecule maps.[10] BAC maps can be contiged together by aligning restriction sites to cover large genomic regions and generate minimal tiling paths.[10]

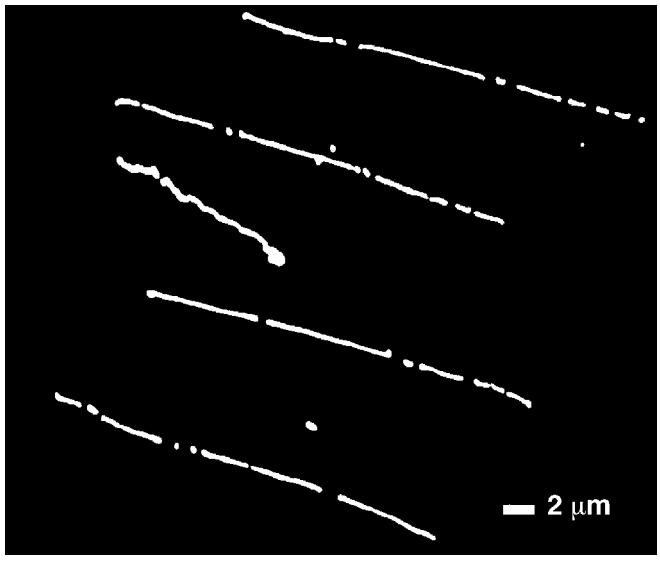

Figure 1 Digital fluorescence micrograph of some typical BAC molecules. This 100 kb BAC, derived from the human Y chromosome, was digested with *Nhe*I.[15] Ordered restriction fragment sizes are 26.6, 3.5, 2.9, 29.1, 3.1, 2.3, 12.4, 6.2, 4.5, 5.7 and 3.5 kb.

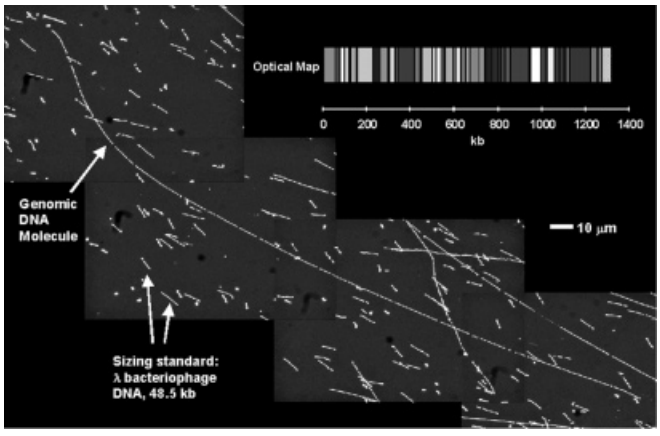

Figure 2 Digital fluorescence micrograph of a typical genomic DNA molecule. A *Plasmodium falciparum* molecule digested with *Nhe*I is shown with its corresponding optical map. Comparison with the consensus optical map shows this molecule to be an intact chromosome. Image composed by tiling a series of 63× (objective power) images using ''gencol''. Co-mounted λ-bacteriophage DNA is used as a sizing standard and to estimate cutting efficiencies.

1.4 Optical Mapping of Genomic DNA

Optical mapping of genomic DNA enables restriction maps of entire microbial genomes to be constructed.[13,14] The substrate is large randomly sheared genomic DNA molecules 0.4–3.0 Mb (megabases) in size (Figure 2). The resulting maps generated using such a ''shotgun optical mapping'' strategy have many overlapping restriction sites permitting contig formation either by visual alignment[14] or by ''Gentig'' software (described below).[3,13] Library construction is obviated enabling mapping of organisms with AT-rich DNA, which is difficult to clone. Also, cloning artifacts are precluded, enabling more accurate maps to be generated. Such maps have proved invaluable as a scaffold for assembling of sequence contigs and as a means of sequence verification.[16] Optical mapping also provides insight into the genomic structure of the organism, revealing the presence of large extrachromosomal elements, which can only be identified at the single molecule level.[13]

2 METHODOLOGY

2.1 Cleaning of Glass Surfaces

Optical mapping surfaces must be cleaned thoroughly before use to ensure uniform derivatization and optical transparency. For safety and cleanliness a self-contained acid boiling system, built from pyrex glass and Teflon™ tubing and assembled in a fume hood, works best. No vacuum grease is used to seal the joints. A sodium hydroxide scrubbing system neutralizes acid vapors. Coverslips are racked in custom-made Teflon™ racks, which hold the coverslips snugly on three edges, and cleaned by boiling in concentrated nitric acid (HNO_3) for 24 h. The system is rinsed extensively with deionized dust-free water until the effluent attains pH 6. The cleaning procedure is repeated with concentrated hydrochloric acid (HCl) which hydrolyzes the glass surface, preparing it for subsequent derivatization. The racked coverslips are rinsed extensively and then sonicated in deionized water to remove residual acid which collects between the Teflon™ parts in the racks. Residual water is then removed by sonicating the racks in distilled ethanol. The cleaned coverslips are stored at room temperature in the racks under distilled ethanol in polypropylene containers.

2.2 Derivatization of Glass Surfaces

Cleaned coverslips are derivatized in slotted Teflon™ racks which hold the coverslips by one edge. Surfaces are derivatized for 48 h at room temperature on a rotary shaker in 3-aminopropyltriethoxysilane (APTES), between 7 and 28 μM, in ethanol.[10,12,17] Macroscopic derivatization parameters such as silane concentration and duration of silane treatment alter the surface charge density which plays a critical role in surface-based optical mapping. The surfaces are checked for optimal performance by estimating the hydrophobicity by measuring the contact angle and assaying the degree of DNA stretching and fragment retention. Derivatized surfaces can be stored in treatment solution for a few weeks. Surfaces can also be derivatized with 3-aminopropyldiethoxymethylsilane (APDEMS)[14] or 3-(triethoxysilylpropyl)trimethylammonium chloride (TESP).[11]

2.3 Preparation of DNA Molecules

Circular BAC and cosmid molecules are prepared by standard alkaline lysis protocols[18] or commercial plasmid preparation columns. Molecules are linearized with λ-terminase prior to surface mounting.[10] λ-Terminase is an endonuclease which cleaves at the λ *cos* site. Occasionally, circular DNA is linearized by restriction digest utilizing a *Not* I site in the polylinker of the cloning vector when no internal *Not* I sites are present in the insert.[10]

Increasing the size of measurable genomic DNA molecules decreases the number required for mapping a complete genome.[19] To minimize shearing, approaches similar to those used to prepare large DNA molecules for PFGE are used.[6,7] Genomic DNA is isolated from cells embedded in low melting point agarose gel inserts which are treated with proteinase K and detergents. Essentially intact DNA is isolated following melting of

the agarose. Physical manipulations such as pipetting can shear large DNA molecules so these are kept to a minimum, and where necessary, wide bore pipetting devices with a slow delivery rate are used. Molecules over 2 Mb can be mounted.[13] Shearing of genomic DNA can also be minimized by condensation of DNA molecules with spermine, a polyamine which collapses DNA coils embedded within agarose.[12] The subsequently molten agarose can be subjected to manipulations as severe as vortexing which would ordinarily shear DNA. The DNA particles are decondensed by washing of gel inserts in spermine-free buffer supplemented with NaCl.

2.4 Mounting of DNA Molecules on Surfaces

DNA molecules exhibit rod-like behavior below ~3 kb, and show evidence of random coil behavior above 3 kb.[20] DNA molecules interact with derivatized surfaces via electrostatic interactions between the anionic DNA and cationic surface. Molecules attach to the surface at one end and then stretch out in the fluid flow.[11,21] Efficient binding and analysis of molecules relies on a balance between flow and electrostatic forces. High-fluid flow combined with a high surface charge density causes overstretching of the molecules, leading to breakage of molecules during and after mounting. Low-fluid flow combined with a high-surface charge density causes understretching of the molecules, leading to coils of DNA deposited on the surface.

By carefully controlling surface conditions, large DNA molecules can be mounted without significant breakage.[10,12] Too high a surface charge density yields poorly stretched, or stretched intact DNA molecules which fail to digest.[8,12] This may be due to occlusion of enzyme binding sites caused by severe adsorption of the molecule to the surface. Alternatively, cleavage may occur, but insufficient relaxation occurs to produce an observable gap.[22] Indeed, both mechanisms may apply. Insufficient charge density leads to desorbtion of DNA fragments smaller than about 1 kb from the charged surface following digestion. The larger the DNA molecules, the lower the required surface charge density. Genomic DNA requires lower charge density than BACs, which require lower charge density than cosmids.[17] Molecules mounted by the peeling method require lower surface charge density than those mounted by the spotting method (see below). Because optical mapping uses fluorescence intensity measurements for mass determination, complete elongation of the DNA molecules is not required for mapping, but it does facilitate the data analysis and provides more mapping information because more cutting sites are resolved.[17,23] DNA is mounted by either ''peeling'' or ''spotting'' protocols.

In the peeling protocol, a few microliters of DNA solution are pipetted onto a microscope slide and a derivatized coverslip is carefully placed on top of the sample to spread out the drop. The coverslip is removed carefully with forceps (''peeled'') and rinsed in Tris–ethylenediaminetetraacetic acid (TE) buffer.[17] The amount of DNA mounted on a surface is titrated on a surface-by-surface basis to give optimum numbers of DNA molecules per microscope field of view. A high density of mounted molecules gives more information but also leads to crossed molecules and an inability to resolve groups of fragments from different molecules. DNA molecules have also been mounted by motor-driven movement of a meniscus held at a 45° angle between a coverslip and a derivatized slide.[24] A procedure for binding DNA molecules at a gel/solution interface has also been developed.[25]

Gridding of multiple samples on a single surface and processing them in parallel is a strategy employed in hybridization-based approaches[26–29] and also in optical mapping.[11] The spotting protocol permits multiple DNA samples to be mounted on a single surface. This enables

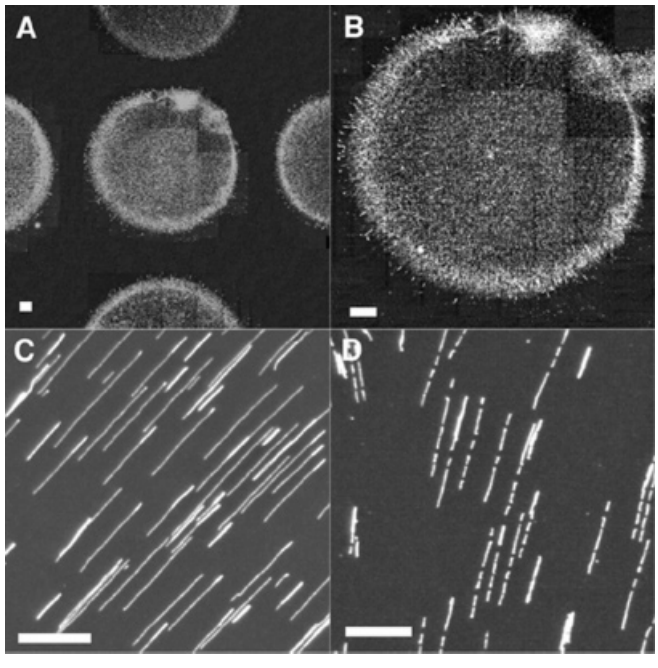

Figure 3 Digital fluorescence micrographs of gridded spots containing fluid fixed molecules. Droplets of λ-bacteriophage DNA onto APTES treated glass surfaces, dried and stained. (A) Section of a 10 × 10 spot grid on a derivatized surface. Image composed by tiling a series of 16× (objective power) images. (B) Close-up of a DNA spot within the grid. Image composed by tiling a series of 16× images. (C) Magnified image of elongated DNA molecules contained within the spot shown in (B) before restriction digestion, 100×. (D) DNA molecules in (B), different field, after digestion with *Bam*HI (100×). Bars: 20 μm (A, B); 5 μm (C, D).[11] (Reprinted from Jing et al., *Proc. Natl. Acad. Sci. USA*, **95**, 8046–8051 with permission. Copyright (1998) National Academy of Sciences, USA.)

digestion and analysis in parallel. Droplets of DNA samples are "spotted" in 10×10 grids using a modified commercially available Biomek laboratory automation robot (Figure 3). The fluid flows generated within drying droplets bring the DNA molecules closer to the surface where they bind in an elongated conformation. The robot is able to deposit one sample approximately every 10 s. Spotting tools include glass capillaries, pipette tips or solid tips which deposit volumes of 10–20 nL using several programs which alter spotting patterns. Such automated high-density deposition techniques make optical mapping an ideal high throughput system for genomic analysis of clones.

2.5 Restriction Digestion of DNA Molecules on Surfaces

The optimal restriction enzyme digestion conditions, in terms of digestion time and enzyme concentration, depend on the enzyme being used and the surface characteristics. Cutting efficiencies are routinely greater than 80%. Efficient cutting of genomic DNA is obligatory for confident contig formation by Gentig software (described below). Typically, 5–20 units of enzyme in 50–200 μL of buffer, are pipetted onto the surface. Digestion is carried out at room temperature for 10 min to 2 h in a humidified closed chamber. The digestion reaction is terminated by aspirating off the enzyme solution and washing the coverslip twice with TE buffer and then distilled water. DNA molecules are stained with oxazole yellow homodimer (YOYO-1) in 20% β-mercaptoethanol. YOYO-1 is a fluorescent dye which bis-intercalates between the DNA bases at one molecule (two fluorophores) per four bps.[30] YOYO-1 solution is pipetted onto a microscope slide and the sample coverslip is placed on top of it. The edges of the coverslip are sealed onto the slide with transparent nail polish. Alternatively the coverslip is covered with 100–200 μL YOYO-1 solution and rinsed with distilled water after 1 min to decrease background staining.[17]

2.6 Choice of Restriction Enzyme

Four, six, and eight bp cutter enzymes can digest surface-mounted DNA. The more frequently a restriction enzyme cuts, the smaller the fragments will be and the higher the resolution of the optical map. However, a compromise must be made since enzymes that cut too frequently generate a population of small fragments which desorb from the surface, generating less accurate maps.

The choice of restriction enzyme also depends on the GC/AT content of the sample DNA, which can vary greatly depending on the organism and the recognition sequence of the enzyme. The GC/AT content will determine the average fragment size for a given enzyme. Human DNA is roughly 50% GC but the content of microorganisms varies greatly. For example, *Plasmodium falciparum* DNA is only 18% GC. *Nhe*I and *Bam*HI were chosen to map this organism since they recognize sequences with a GC bias (GCTAGC and GGATCC) and would prevent too frequent cutting. Assuming a random distribution of nucleotides across the genome, the predicted cutting frequency with these enzymes would be every 91 kb. The average sizes of the fragments generated by optical mapping of the whole genome were 24 kb and 32 kb, respectively.[14] *Deinococcus radiodurans* DNA is 68% GC, predicting an *Nhe*I cutting frequency of 3 kb. When this organism was mapped with *Nhe*I, the average size of the fragments was 29 kb.[13] This demonstrates that the GC/AT content of these organisms is not randomly distributed and each genome must be evaluated empirically on a organism by organism basis. Consequently, before optical mapping begins, a digest is performed in solution with a variety of candidate enzymes and a distribution of fragment sizes is determined by PFGE.

Enzymes that work in a NaCl-containing buffer (50–150 mM) are required since salt-free buffer does not permit relaxation of the DNA on the surface following digestion. Higher salt concentrations cause more DNA relaxation and less retention of small fragments. The restriction enzyme buffer needs to be compatible with the surface charge density. Surfaces with high charge density are used with high salt buffers and surfaces with low charge density are used with low salt buffers. Buffers containing potassium salts in place of NaCl alter the integrity of the surface-mounted DNA, in a presently unknown manner, and therefore enzymes that work in these buffers cannot be used. Consequently, not all restriction enzymes can be used for optical mapping.

2.7 Sizing of DNA Molecules

Fluorescence intensity measurements are used for DNA fragment mass determination. An internal standard of known length is used to size fragments accurately. In the case of cloned DNA this can be the vector which has *Not*I restriction sites in the polylinker at either end.[10] The pBAC108L vector, for example, gives a 6.9-kb sizing standard. Circular molecules can therefore be mounted on surfaces and sized in a circular form. In the case of genomic DNA, λ-bacteriophage DNA (48.5 kb) or cosmid DNA (about 45 kb) can be co-mounted with the sample (Figure 2). Cutting efficiency can also be estimated if the co-mounted standard possesses a restriction site for the enzyme used for mapping.

2.8 Instrumentation

The optical mapping workstation (Figure 4) is built around a Zeiss™ 135 Axiovert inverted microscope

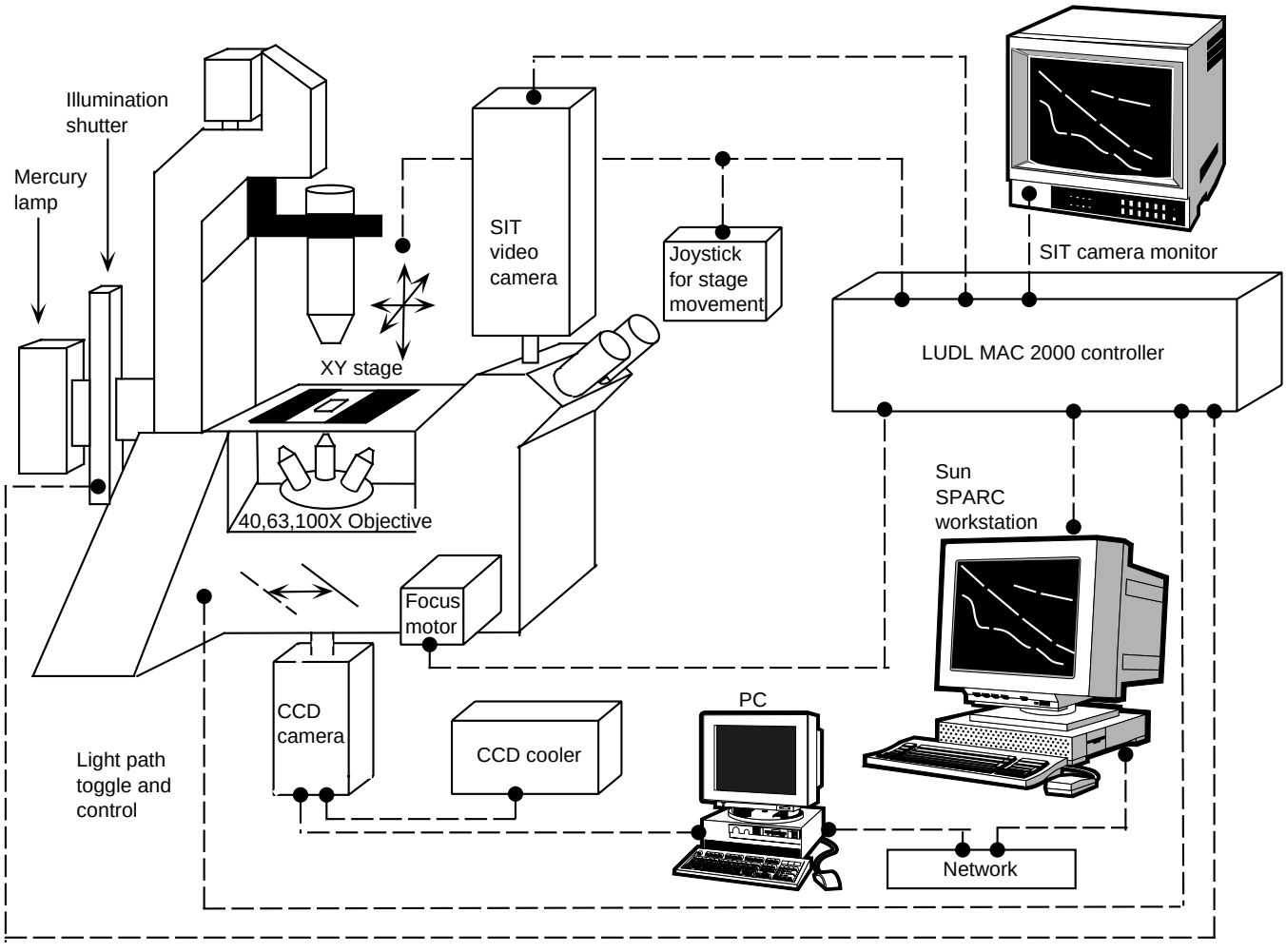

Figure 4 Optical mapping workstation. The schematic illustration shows how the microscope, silicon intensified target (SIT) video camera, cooled CCD camera, LUDL™ controller and Sun workstation are integrated.

equipped for epifluorescence, with several different objectives (40, 63 and 100× Zeiss™ plan-neofluor oil immersion) allowing for imaging of biomolecules of a wide range of sizes. The 100× objective is most commonly used, whereas the 63× objective is used for clones larger than 200 kb. Filter packs allow for imaging of biomolecules labeled with fluorescent dyes of different spectral properties. The microscope is equipped with a Dage™SIT 68GL low light-level video camera and accompanying monitor for automatic and manual focusing, and a Princeton Instruments cooled CCD camera (1317 × 1035 pixels, KAF 1400 chip, 12-bit digitization) for acquiring high-resolution digital images. Automatic features such as movement of the microscope stage (0.1-μm resolution), image collection, light shutter operation and video autofocusing are controlled by a LUDL™ Electronics box, which contains a LUDL™ Electronics MAC 2000 interface bus, housing all interface cards (PSSYST 200, MCMSE 500, MDMSP 503, AFCMS 801, FWSC 800, and RS232INT 400). The LUDL™ MAC 2000 is interfaced via a RS232 serial connection to a Sun Microsystems™ SPARC 20 dual processor computer workstation. The CCD camera is also interfaced, via a pentium-based microcomputer controller and a distributed network, to the Sun workstation. The Sun workstation is part of a distributed network on which interactive and automatic image processing is performed. Software for control of the above peripherals is written in the C programming language. The optical mapping workstation is mounted on a floating optical table.

2.9 Microscope Automation Software and Collection of Images

DNA molecules are imaged using optical map maker (OMM) software, operated through a collection of user-friendly interfaces, and the image collection process can be followed on the Sun workstation monitor.

OMM software includes full computer control of all the workstation functions such as focusing, sample translation (x–y position), switching between SIT and CCD cameras, and image collection. Manual and fully automatic modes are incorporated. Image acquisition parameters such as imaging time, sample grid position and size, and spot diameter can also be varied. A sample image is collected by the CCD camera and the final imaging time is then set by the operator so that the pixel value on any part of the molecules is below saturation (4095 gray levels). Digital images can be acquired by the workstation at the rate of 4 min^{-1} (using 5 s imaging time). The development and implementation of automated microscope operations has enabled operator-free imaging of optical mapping surfaces. An operator can enter the coordinates of a series of gridded spots and images are subsequently collected and tiled together automatically, only from the areas on the surface where DNA has been spotted.[11] As part of a developmental path for large clone mapping with automatic image collection, the program ''Mark and Collect'' has been developed. This software saves the operators time spent waiting for images to be collected and transferred to the file system, which normally takes about 10 s per image. Microscope fields of interest are selected by the operator and the coordinates stored. When enough areas of the optical mapping surface have been identified the program subsequently moves the stage and collects the images. The image acquisition output is linked to the automatic map-making routines, so that after the images have been acquired and stored on hard disk arrays they can be either directly submitted for automatic image processing and extraction of restriction map data or saved for later processing. The OMM system can run on a network of other workstations. Access to all aspects of the OMM data and processing is made through one shared directory hierarchy. This file system structure and the accompanying software libraries provide uniform controlled access to all collection and processing activities and data. Several specially designed ''viewers'' permit the user to follow the different steps of the image analysis visually, so that convenient assessment and troubleshooting can be performed. A distributed processing system allows all the available computational resources on the network to be shared.[17]

2.10 Correction of Images

The CCD camera produces images composed of 1317 × 1035 pixels of 12 bits each which are read out and stored as an array of unsigned 16-bit values. The intensity value for a pixel at no light is the dark value, and an image taken with the microscope shutter closed is a dark image. This dark image, which represents the electronic noise in the CCD chip, varies from camera to camera and with the parameters used for taking the image (exposure time, etc.). The first correction made to an optical mapping image is to subtract the corresponding dark image. The second correction, flatfield correction, is carried out to adjust for differences in the amount of light registered by a particular pixel that are not due to the difference in distribution of dye in the sample. Such variations can be due to nonuniformity of the light source, differences in the light path (smudges on optical components, etc.), and possibly other factors in the CCD and readout electronics. The correction involves dividing each pixel value in the image by the corresponding pixel value in a bright image, an image of a thin, in-focus, and uniformly fluorescent sample. Since the only fluorescence of interest is that emanating from dye bound to DNA molecules, other sources of fluorescence are considered background and the subtraction of this background contribution is the last correction made. Based on the assumption that the major source of background is unbound dye on the surface and that this does not vary rapidly, the background value for each pixel can be estimated. First, the median pixel values for pixels on concentric circles around the pixel in question are found. Then, a median of medians is taken, a median image calculated and a Gaussian smoothing operation applied to give the final background image. This image is subtracted from the flatfield corrected image.[17]

2.11 Analysis of Images

The purpose of the analysis of optical mapping images is to find the cognate fragments derived from a single DNA molecule, and for each fragment calculate a signal representing the intensity in the image along the length of the molecule. DNA molecules are found by looking for long thin bright objects which have a dominant direction.

In the first phase, an algorithm identifies these isolated regions in the image, using both the fluorescence intensity and local directionality properties at each pixel, where each region contains the fragments of a parental molecule not overlapping with other molecules. This is done by first applying a ridge filter[31] to the input corrected image which produces 16 new images. Each image corresponds to one of 16 different directions and the value of a pixel in one of these images represents a calculation of the degree to which the pixel appears to lie on a ridge in the particular direction. An image is then constructed which contains, at each pixel, the highest of the 16 values for that pixel. A threshold value is picked so that the peak on the histogram for an area containing the segment to be measured is the midpoint between the threshold value and the lowest background value. A threshold is calculated for the new image and objects are identified as contiguous collections of pixels whose values exceed the threshold. At this point the objects (collections of pixels

with high numerical values) consist of pixels which are on ridges in some of the 16 directions, but since all 16 directions are included, and since the ridge filter tends to swell up in regions which have no predominant direction (e.g. at the ends of molecules), this image has a tendency to merge molecules which happen to be close to each other. To separate these molecules, the ridge filtering process is repeated for nine, then five, and finally three directions so that, in the final image, the objects consist of pixels lying along a ridge in one of three directions. This process has been shown to find and separate DNA molecules.

In the second phase, a path down the center of the molecule (the ''backbone'') is computed using Dijkstra's ''all paths least cost'' algorithm.[32,33] A pixel in the middle of the molecule is chosen and the algorithm is applied to find the lowest cost path (traversing the molecule one step at a time from a pixel to one of its eight neighboring pixels) from the first pixel to each other pixel in the object. The cost of a path is the sum of the costs of the individual steps from pixel to pixel, and the cost function applied to each step favors moving to a pixel which is bright enough to be in the center of the molecule, but does not favor brighter pixels so strongly that the path wanders all over the middle of the molecule looking for the very brightest pixels despite the extra length. The paths from the first pixel to all boundary pixels are evaluated to identify an end-point pixel and the process is repeated using the end-point pixel instead of the first pixel. This finds the other end-point pixel and the path to it is the backbone. A smooth intensity signal along the backbone is then computed; for each position along the backbone, the intensity is calculated by summing the intensities for a set of pixels which are close to the backbone and lying along a line orthogonal to the backbone at that position. This intensity signal is then used to determine potential cut sites and estimated masses of all the potential fragments in that molecule. The estimated fragment mass data is normalized, and then, together with the information about the order of the fragments, collected for statistical analysis by the map-making program.

Images of single digested DNA molecules are taken using ''GenCol'', a semiautomatic image collection software which allows storage and seamless tiling of consecutive CCD images, covering the entire length of long genomic DNA molecules which span multiple microscope fields and is necessary for imaging of extraordinarily long DNA molecules.

The program ''visionade'' is an interactive graphical editing tool that allows a user to process a group of optical mapping images quickly, identifying fragments and molecules in the images with a few point-and-click actions. In the optical mapping system this step is done after a group of images has been collected and automatically flattened. Visionade is run on a group of images that contain molecules that are processed together to create a restriction map; e.g. a set of overlapping images containing a long genomic molecule and many copies of a smaller standard molecule which are used to size the fragments of the genomic molecule (see below). The user brings up an image, adjusts the color mapping to make the fragments clear, uses a threshold slider in the interface to color the fragments, clicks on the fragments to identify and number them, and then clicks again to group the current set of fragments into a molecule. The outputs of visionade are mask images locating the fragments and molecules. These outputs are combined with the flattened images to measure the sizes of the fragments in the molecules and this data is the input to the statistical analysis program ''lcluster'' (see below) which is a novel computational approach to map construction from optical data and finds the consensus map.

2.12 Map Construction Using Cloned DNA

The map making program ''lcluster'' gathers all of the per-molecule computations from all images and operates on them together. Here, the complete restriction map is computed by testing a set of plausible hypotheses about the restriction cleavage sites and choosing the most likely hypothesis supporting the previously accumulated data. The program starts by filtering out unsuitable molecules based on several criteria. (1) A molecule must have a total length in pixels between a specified minimum and maximum. (2) A molecule must not be within 10 pixels of the image boundary. (3) The maximum angle of curvature over the entire length must be below a specified angle (usually 30°). (4) The total intensity of the molecule due to pixels outside a specified distance from the backbone must be less than a specified threshold. These criteria tend to exclude most partial and overlapping molecules. Bayesian inference techniques enable the construction of maps from noisy data. For example, the map construction algorithms can produce maps from a population of partially digested clone molecules having a digestion rate as low as 15%.[2] Bayesian estimation is used to determine the optimum way of accounting for each of the fragments in the filtered data set. A hypothetical model of the map is formulated, which specifies the following parameters: cut locations, the variance of cut site measurements, the cutting efficiency and the sample purity. Suitable probability distributions are applied to each of the parameters in the model. The cut locations follow a Gaussian distribution, false cuts are assumed to be uniformly distributed over the length of the molecule, the number of false cuts follow a Poisson distribution, and the cutting frequency is a simple function of the number of cuts. The object of the analysis is to find the maximum-probability model of the map given the data. The program

considers about 8000 models for each model size, i.e. maps of 2, 3, 4... fragments, and picks the 32 best models of each size, whose parameters are then optimized locally to converge to the nearest maximum in probability space. The best model of any size is accepted as the solution map if its probability density is significantly better than that of any other model.

2.13 Map Construction Using Genomic DNA

Whole genome optical maps are constructed by mapping large randomly sheared genomic DNA molecules. Overlapping maps are contiged by aligning restriction sites. Map contigs can be laboriously assembled manually over a course of weeks,[14] or within minutes using ''Gentig'' software which automatically assembles map contigs and a consensus map from optical mapping data.[13] Gentig uses a greedy algorithm with limited backtrack to find an almost optimal scoring set of map contigs, while constraining the false positive error rate below a negligible value. This approach ensures that the computed contigs are largely correct, although it fails to minimize the number of contigs. On the other hand, Gentig avoids the high computational complexity that would be incurred in attempting to find the optimal placement, a problem shown to be computationally infeasible in the most general setting.[3] Molecular length and digestion efficiency control fragment alignment and are well-modeled by prior probabilistic (Bayesian) analysis.[3] The Bayesian approach requires careful choice of parameters and a known prior statistical distribution of error sources. The standard deviation, digestion rate, false cut rate and false match probability can be altered to change the number of molecules that Gentig contigs together. The consensus map computed by Gentig is free of errors due to partial digestion, sizing error and false cuts.

The scoring function is computed from a Bayesian probability density estimating the probability that the two distinct component maps could have been derived from the proposed placement while subject to various data errors. The data errors which are modeled include sizing errors, missing cut sites, and false ''optical'' cut sites. First the most likely placement is hypothesized, then the probability density is computed for the mismatch errors in the component maps given the hypothesized placement and the error model. The second step is similar to the Bayesian probability density computation;[2] one difference is that simplifying assumptions are made to improve computational efficiency. In particular, the hypothesized placement is computed by averaging of the contiged fragment sizes, rather than a true Bayesian probability density maximization with fragment sizes as parameters. Furthermore, good estimates of the error model parameters are assumed to be known a priori, but further improved by a re-estimation from the data using a limited number of iterations of a true Bayesian probability density maximization. The best offset and alignment between a pair of maps is finally computed by an efficient dynamic programming algorithm, thus avoiding a potentially exponential complexity.[15]

2.14 Visualization of Map Data

Whole genome maps composed of many restriction sites and deep contigs are viewed with ''ConVEx'' (Contig visualization and exploration tool), which is a multiscale zoomable interface for cognition of large high-resolution contiged restriction maps.[34] Users can examine the consensus maps together with the raw uncorrected data. Multiple representation of the graphical objects is presented in a scale-dependent manner and a powerful ''lens'' mechanism analyzes the underlying sequence data, enabling real time browsing of very large reference maps. The lens mechanism is augmented with a programming language (S-Lang language) allowing the user to create their own customized lenses. Such a lens (or a portfolio of lenses) provides a powerful tool for a user to examine the vast amount of data quickly stopping only infrequently to note a pattern, a point of interest or some editing of the data. Lenses facilitate examination of the global structure of the contig and provide annotation and editing mechanisms, allowing users to communicate interesting features such as STS (sequence tagged site) markers, even sequence reads.[15]

3 OPTICAL MAPPING TO CHARACTERIZE GENOMES

3.1 Optical Mapping for Sequence Assembly and Verification

A prerequisite for characterizing and sequencing complex genomes is the availability of high-resolution maps of cloned DNA molecules. Optical maps support whole genome sequencing in two ways. First, optical maps are used to generate minimal tiling paths of large insert clones.[10] Overlapping clones are contiged together by alignment of restriction sites (Figure 5). A minimal tiling path of clones is then chosen to minimize redundant sequencing. Second, optical maps of cloned or genomic DNA are extremely valuable to assist and ultimately confirm assembly of sequence generated by a shotgun approach (Figure 6).

A shotgun strategy has been employed to sequence the genomes of a number of microorganisms completely and sequencing of the human genome is an ongoing endeavor. DNA can only be sequenced in small portions

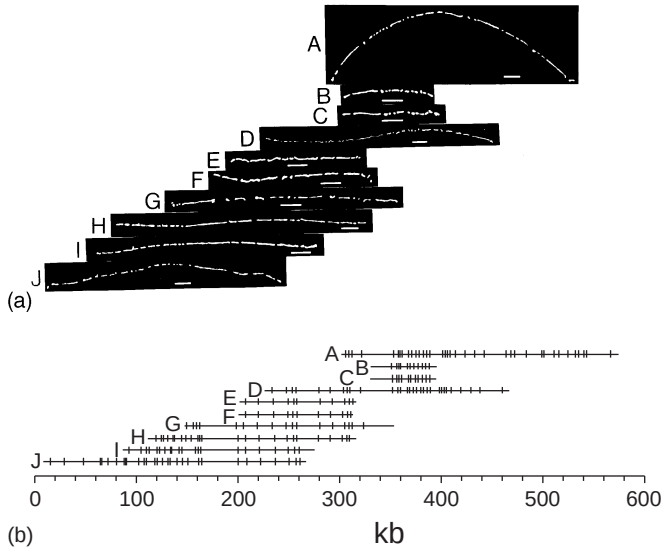

Figure 5 A BAC contig of human chromosome region 11p13. (a) images of individual clone molecules arranged in a contig. (Bar 5 μm) (b) *Bam*HI restriction maps of corresponding clones.[10] (Reprinted from Cai et al., *Proc. Natl. Acad. Sci. USA*, **95**, 3390–3395 with permission. Copyright (1998) National Academy of Sciences, USA.[10])

of about 500 bp. Shotgun sequencing essentially involves random fragmentation of the target DNA sequence, which can be an entire microbial genome or in the case of larger metazoan genomes a large insert clone such as a BAC. The resulting small fragments are cloned and sequenced. The assembly of sequence into contigs is accomplished using a number of software packages that read raw sequence from traces, search for homologous sequence to construct contigs, and allow significant editing of intermediate results. The entire target is redundantly sequenced to ensure maximal coverage and sequence accuracy. Nevertheless, some gaps still remain, often due to under representation in the library.

Optical maps can be used as a scaffold for accurate alignment of sequence contigs.[15] A restriction map is created for each sequence contig in silico and optimally aligned with the optical map. Optical maps can also characterize gaps in terms of breadth and location and may facilitate closure techniques such as long-range PCR.[35]

3.2 Very High-resolution Optical Mapping

The information content of a multiple restriction enzyme map is greater than the sum of its parts. Smaller sequence contigs can be placed on a multienzyme map. Multienzyme maps are generated by constructing single enzyme maps and then, by a series of double digestions, or use of available sequence information, these maps are combined into one dense construct. Optical maps of BACs digested with seven different enzymes have been overlaid to create a very high-resolution map with average fragment size 1.2 kb.[10] Although this gives a very high density of markers, this resolution is still insufficient to anchor single sequence reads or genes.

Optical maps of genomic DNA have been linked using available sequence information.[14] If a sequence contig contains restriction sites for two or more of the enzymes used to make the optical map, this contig can act as a linkage point to orient and align single enzyme optical maps. Rare cutter maps can be linked to existing frequent cutter maps by digesting the DNA into large fragments in a test tube with the rare cutter and then mounting these fragments on the optical mapping surface and mapping with the frequent cutter.[13] The resulting maps are contiged against a consensus map using just the frequent cutter. Blunt ended, abutting contigs indicate

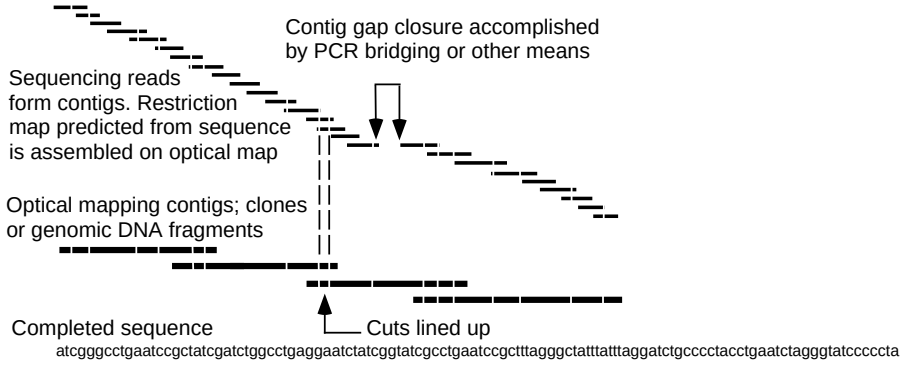

Figure 6 Optical maps provide a scaffold for assembling sequence contigs. The analysis scheme shows alignment of sequence contigs with a consensus optical map generated by alignment of overlapping restriction fragments generated from genomic DNA molecules or large insert clones. Restriction maps are generated in silico from sequence data. Restriction sites on maps and sequence can be used to anchor sequence contigs. They can also characterize gaps in terms of breadth and location and may facilitate closure techniques such as long-range polymerase chain reaction (PCR).

the rare cutter sites. A scheme for aligning optical maps using Southern blotting has also been employed.[13–15,17]

3.3 Identification of Repeated Regions

In the absence of restriction maps, repeated regions can confound sequence assembly efforts. The number of repeats and their orientation is often difficult to determine using sequence assembly software. Clones thought to contain repeats can be optically mapped using a restriction enzyme which cuts within the repeat, enabling the number of repeats and their organization to be revealed.[36]

3.4 Identification of Genic Regions

Identifying coding sequences within mammalian genomes has proven both technically challenging and time-consuming, primarily because only small regions are transcribed; in the case of the complex human genome, 0.06%. The figure drops to 0.007% if purely exonic regions are considered.[37] Optical mapping can be used to screen for genes in mammalian genomes.[17] Expressed mammalian genes are characterized by a CG motif (CpG islands) which normally occurs at a low frequency in noncoding regions. CpG islands constitute a distinctive fraction of the genome because, unlike bulk DNA, they are nonmethylated and contain the dinucleotide CpG at its expected frequency. The average length of CpG islands is 1 kb and the haploid human genome is estimated to contain about 45 000 islands.[38] All human housekeeping genes and 40% of tissue-restricted genes are associated with CpG islands.[39] In total, only 57% of human genes have CpG islands, so unfortunately not all genes can be identified by CpG islands. Putative genes can be identified by optical mapping with enzymes which all have CpG in their recognition sequences such as *Not*I or *Eag*I. However, CpG islands may not be of such value in locating genes in microbial genomes since these often possess strikingly different A + T and G + C content compared to human DNA.

3.5 Optical Mapping for Population-based Genomic Studies

If high-resolution maps can be rapidly generated, microbial populations can be analyzed at the whole genome level. Whole genome maps from related strains can be aligned and differences in the maps can show regions where the genomes differ. With the sequence of one strain in hand, maps generated from the unsequenced strain can be compared to identify regions unique to the other strain which could be targeted for sequencing.

4 COMPARISON WITH OTHER PHYSICAL MAPPING APPROACHES

4.1 Gel Electrophoresis

The invention of PFGE in 1984[7] enabled large DNA molecules to be separated based on their size, and opened the door to physical mapping of whole chromosomes and genomes. Large DNA molecules digested with rare cutter restriction enzymes could be separated, generating a fingerprint, but determination of fragment order was still dependent on Southern blotting, generation of probes and hybridization of filters. Optical mapping, on the other hand, provides unambiguous fragment order without further manipulations. Fragments can be sized more accurately by optical mapping versus PFGE. Sizing accuracy for restriction fragments generated from BACs is comparable if not superior to conventional and PFGE.[10,40] Physical maps prepared by gel electrophoresis often fail to disambiguate repeats. Since optical mapping characterizes complex genomic regions at the level of single molecules, repeat complexities are discerned using restriction enzymes that cut once within a single repeating unit.

With a few notable exceptions, which include an eight-enzyme map of the *Escherichia coli* K12 genome[41] and physical maps of the *Saccharomyces cerevisiae* genome,[42,43] very few whole genome restriction maps have been constructed using electrophoretic approaches. High-resolution whole genome maps can be rapidly prepared by optical mapping. *Deinococcus radiodurans* and *Plasmodium falciparum* have been mapped and the fidelity of the maps was confirmed by sequencing of these genomes.[13,14,16,44] Microgram quantities of DNA are required to analyze DNA by electrophoresis. Since optical mapping analyzes DNA at the single molecule level, small amounts of starting material are required. This enables mapping of microorganisms, such as parasites, which are problematic to culture. For example, an optical map of the whole 30-mb *P. falciparum* genome was prepared using no more than 2 ng of genomic DNA.

Electrophoretic approaches are not amenable to full automation to include sample preparation and analysis. Gridding of samples on a single surface and development of object recognition via computer, ''machine vision'', has made optical mapping of cloned DNA a semi-automated process with the expectation that this type of development path can be extended to mapping of genomic DNA.

4.2 Fluorescent In Situ Hybridization

Fluorescent in situ hybridization (FISH) has emerged as a powerful approach for gene and chromosome level analyses in clinical cytogenetics.[45,46] FISH can be performed on metaphase chromosome spreads, interphase

nuclei[47,48] and decondensed ''halo'' DNA.[49] FISH is suitable for visualization of genetic aberrations affecting a few specific loci for which probes are available. However, its dependence on prior knowledge of genomic sequence makes it unsuitable for screening unknown genetic aberrations. FISH probes typically span 40–200 kb of genomic DNA.[50] Aberrations smaller than 10 kb cannot be reliably detected. Interphase FISH cannot estimate distances between hybridization markers greater than 100 kb.[51] Whole chromosome painting, multifluor FISH[46,52] and spectral karyotyping (SKY)[53] have enabled simultaneous visualization of up to 27 combinatorially labeled probes and this approach has revealed complex chromosome rearrangements not evident by conventional means.[52] Development of fiber FISH,[54,55] where the hybridization template consists of highly deproteinized chromatin aggregated together to form multistranded DNA fibers, has enabled probes as small as 200 bp to be used to map exonic regions on cosmid clones.[56] Such approaches could potentially be used to estimate numbers of duplicated expressed sequence tags (ESTs), or STSs examine chromosomal breakpoints in genomic DNA, or generate minimal tiling paths of clones.[54,57] Restriction fragments have also been used as fiber FISH probes to generate rudimentary restriction maps of large insert clones.[58] Short padlock FISH probes[59] can be amplified at the site of hybridization by a rolling-circle mechanism[60] and can be used to detect mutations.

Related approaches such as comparative genomic hybridization (CGH), which is a two-color hybridization of differentially labeled test and reference DNAs to normal human metaphase spreads[61] or DNA fibers,[62] and primed in situ labeling (PRINS), which is chromosome labeling performed by annealing of specific oligonucleotide primers followed by primer elongation by a *Taq* polymerase in the presence of labeled nucleotides,[63] are used predominantly for studying gene copy number aberrations in tumors and are limited to gross chromosomal changes such as deletions, translocations and amplifications.

In situ hybridization-based approaches and optical mapping are similar in requiring surface mounting of large intact DNA molecules. Extended DNA molecules are mounted for fiber FISH analysis by ''molecular combing'', whereby molecules attach at one end of the derivatized surface and are extended and aligned by a receding meniscus.[64,65] Limitations of these other approaches include low throughput, DNA fragmentation which prevents subsequent or simultaneous multimethod analyses, and difficulties in automation. These molecular approaches generally yield low resolution maps and were developed primarily for characterization of single loci, not entire genomes.[66,67]

4.3 Hybridization-based Microarrays

DNA probes can be gridded onto charged membrane surfaces[26–29] and modification of technologies from the photolithography and semiconductor industries enable oligonucleotide probes to be gridded at extremely high density creating microarrays on ''biochips''.[68,69] DNA samples labeled with fluorescent tags hybridize to homologous probes and the fluorescence intensity for each segment measured. Biochips are used to detect mutations, quantitate nucleic acids, and create gene expression profiles from different cell populations.[70–72] Although such hybridization-based approaches and optical mapping are both performed on surfaces and are amenable to high throughput by gridding of samples, this is where the similarities end. There is no underlying biochemical process in hybridization to probes. Biochips require prior knowledge of sequence whereas optical mapping can characterize unknown sequence. The hybridization on biochips is to short oligonucleotide DNA sequences[68,69] and the means by which biochips cover large genomic regions is by gridding of hundreds of thousands of probes on a single chip.[73] The number of samples that can be gridded on an optical mapping surface is limited first by the length of the molecules, which in the case of BACs can be in the order of microns, and second by the area on the surface which is required to generate sufficient fluid flow to stretch DNA molecules in the sample.

4.4 Other Single Molecule Approaches

Other single DNA molecule approaches have included visualization of single DNA molecules by fluorescence microscopy in solution,[74–76] and during gel electrophoresis,[75,77–79] restriction mapping of cosmid clones by atomic force microscopy of restriction enzymes bound to DNA,[80] visualization of DNA by scanning tunneling microscopy,[81] scanning force microscopy of circular DNA and DNA bending in transcription complexes,[82,83] direct mechanical measurement of the elasticity[84] and supercoiling of single DNA molecules using magnetic beads.[85,86] These approaches have been useful in characterizing structural and functional properties of single DNA molecules but have not found application in generating physical maps of genomes since applicability to automation and throughput are obstacles.

5 RECENT IMPROVEMENTS AND FUTURE DEVELOPMENTS

In order to map arbitrary sequences and gene locations, RecA-assisted restriction endonuclease (RARE)

digestion has been combined with optical mapping.[87] In this approach, DNA samples containing *Eco*RI sites are embedded in agarose inserts. RecA protein/oligonucleotide complexes bind to specific targets which contain *Eco*RI sites spanned by additional specific sequence of interest. Binding of the complex prevents the action of *Eco*RI methylase activity, which is used to methylate *Eco*RI sites which lack the specific oligonucleotide sequence. After RecA protein removal, the protected sites become the only non-methylated sites and are thus uniquely cleavable with *Eco*RI. The molecules are then mounted and mapped optically.

Restriction digests on surfaces are not the only biochemical reactions that can be assayed at the single molecule level. DNA polymerase I incorporation of fluorescently labeled nucleotides has been observed to occur on surface-mounted DNA molecules.[11] The action of DNA polymerases is quite different from that of restriction endonucleases. Restriction endonucleases bind their target, often with little distortion to the DNA helix.[88] In contrast, DNA polymerase I has a more complicated action. The enzyme binds to a nick site and then essentially wraps around the helix, using 5′-3′ exonuclease activity to remove bases, and 5′-3′ template addition to add new nucleotides.[89,90] Surface mounted molecules are therefore biochemically active. Severe adsorption of the DNA helix to the surface would preclude any meaningful degree of template addition or exonuclease activity.

Since optical mapping of genomic DNA has been used to analyze entire microbial genomes, similar approaches are being used to make a restriction map of the entire human genome. A pilot project to map the entire human genome using megabase-sized DNA molecules was recently completed. Molecules, directly extracted from peripheral blood, were mapped with *Pac*I. 0.6 human genome equivalents were mapped at 40-kb resolution, using genomic fragments with average size of 2.1 Mb. Analysis of the contigs formed showed good correspondence with suitably modified Lander–Waterman[19] physical mapping criteria in terms of the number and depth of overlapped genomic fragments. Goals are simultaneously to complete the human reference map to include 10–15 × coverage and to link with other physical maps by the alignment of restriction mapped BAC contigs. The utility of this map will be to facilitate assembly of sequence from the Human Genome Initiative. Optical mapping systems could potentially be used for large scale genomic analysis and as a means to perform population-based genomic studies such as analysis of human disease genes or characterization of breakpoints in tumors.

ACKNOWLEDGMENTS

The authors' work was supported in part by grants from the National Institutes of Health (HG-00225-08), the Department of Energy and the Burroughs Wellcome Fund.

ABBREVIATIONS AND ACRONYMS

APDEMS	3-Aminopropyldiethoxymethylsilane
APTES	3-Aminopropyltriethoxysilane
BAC	Bacterial Artificial Chromosomes
bp	Basepairs
CCD	Charge-coupled Device
CGH	Comparative Genomic Hybridization
ConVEx	Contig Visualization and Exploration Tool
DAPI	4,6-Diamidino-2 Phenylindole Dihydrochloride
DNA	Deoxyribonucleic Acid
ESTs	Expressed Sequence Tags
FISH	Fluorescent In Situ Hybridization
kb	Kilobases
Mb	Megabases
OMM	Optical Map Maker
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
PRINS	Primed In Situ Labeling
RARE	RecA-assisted Restriction Endonuclease
SIT	Silicon Intensified Target
SKY	Spectral Karyotyping
STS	Sequence Tagged Site
TE	Tris–Ethylenediaminetetraacetic acid
TESP	3-(Triethoxysilylpropyl)trimethyl-ammonium Chloride
YAC	Yeast Artificial Chromosomes
YOYO-1	Oxazole Yellow Homodimer

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
Comparative Genomics: Differential Display and Subtractive Hybridization • Genome Physical Mapping Using BACs • Restriction Landmark Genomic and cDNA Scanning

REFERENCES

1. A.H. Samad, W.W. Cai, X. Hu, B. Irvin, J. Jing, J. Reed, X. Meng, J. Huang, E. Huff, B. Porter, A. Shenkar,

T. Anantharaman, B. Mishra, V. Clarke, E. Dimalanta, J. Edington, C. Hiort, R. Rabbah, J. Skiadas, D.C. Schwartz, 'Mapping the Genome One Molecule at a Time: Optical Mapping', *Nature*, **378**, 516–517 (1995).

2. T.S. Anantharaman, B. Mishra, D.C. Schwartz, 'Genomics via Optical Mapping II: Ordered Restriction Maps', *J. Comput. Bio.*, **4**, 91–118 (1997).
3. T.S. Anantharaman, B. Mishra, D.C. Schwartz, 'Genomics via Optical Mapping III: Contiging genomic DNA and Variations', Courant Technical Report #760, Courant Institute, New York University, NY, March, 1998.
4. X. Guo, E.J. Huff, D.C. Schwartz, 'Sizing Single DNA Molecules', *Nature*, **359**, 783–784 (1992).
5. D.C. Schwartz, X. Li, L. Hernandez, S. Ramnarain, E. Huff, Y. Wang, 'Ordered Restriction Maps of *Saccharomyces cerevisae* Chromosomes Constructed by Optical Mapping', *Science*, **262**, 110–114 (1993).
6. B. Birren, E. Lai, *Pulsed Field Gel Electrophoresis; A Practical Guide*, Academic Press, San Diego, CA, 1993.
7. D.C. Schwartz, C.R. Cantor, 'Separation of Yeast Chromosome-sized DNAs by Pulsed Field Gradient Gel Electrophoresis', *Cell*, **37**, 67–75 (1984).
8. X. Meng, K. Benson, K. Chada, E. Huff, D.C. Schwartz, 'Optical Mapping of Lambda Bacteriophage Clones Using Restriction Endonucleases', *Nat. Genet.*, **9**, 432–438 (1995).
9. X. Meng, W. Cai, D.C. Schwartz, 'Inhibition of Restriction Endonuclease Activity by DNA Binding Fluorochromes', *J. Biomol. Struct. Dyn.*, **13**, 945–9951 (1996).
10. W. Cai, J. Jing, B. Irvin, L. Ohler, E. Rose, U. Kim, M. Simon, D.C. Schwartz, 'High Resolution Restriction Maps of Bacterial Artificial Chromosomes Constructed by Optical Mapping', *Proc. Natl. Acad. Sci. USA*, **95**, 3390–3395 (1998).
11. J. Jing, J. Reed, J. Huang, X. Hu, V. Clarke, J. Edington, D. Housman, T.S. Anathraman, E.J. Huff, B. Mishra, B. Porter, A. Shenker, E. Wolfson, C. Hiort, R. Kantor, C. Aston, D.C. Schwartz, 'Automated High Resolution Optical Mapping Using Arrayed, Fluid Fixed, DNA Molecules', *Proc. Natl. Acad. Sci. USA*, **95**, 8046–8051 (1998).
12. W. Cai, H. Aburatani, D. Housman, Y. Wang, D.C. Schwartz, 'Ordered Restriction Endonuclease Maps of Yeast Artificial Chromosomes Created by Optical Mapping on Surfaces', *Proc. Natl. Acad. Sci. USA*, **92**, 5164–5168 (1995).
13. J. Lin, R. Qi, C. Aston, J. Jing, T.S. Anantharaman, B. Mishra, O. White, M.J. Daly, K.W. Minton, J.C. Venter, D.C. Schwartz, 'Whole Genome Shotgun Optical Mapping of *Deinococcus radiodurans', Science*, **285**, 1558–1562 (1999).
14. J. Jing, Z. Lai, C. Aston, D.J. Carucci, M.J. Gardner, B. Mishra, T.S. Anantharaman, H. Tettelin, L.M. Cummings, S.L. Hoffman, J.C. Venter, D.C. Schwartz, 'Optical Mapping of *Plasmodium falciparum* Chromosome 2', *Genome Res.*, **9**, 175–181 (1999).
15. C. Aston, B. Mishra, D.C. Schwartz, 'Optical Mapping and its Potential for Large-scale Sequencing Projects', *Trends Biotechnol.*, **7**, 297–302 (1999).
16. M.J. Gardner, H. Tettelin, D.J. Carucci, L.M. Cummings, L. Aravind, E.V. Koonin, S. Shallom, T. Mason, K. Yu, C. Fujii, J. Pederson, K. Shen, J. Jing, C. Aston, Z. Lai, D.C. Schwartz, M. Pertea, S. Salzberg, L. Zhou, G.G. Sutton, R. Clayton, O. White, H.O. Smith, C.M. Fraser, M.D. Adams, J.C. Venter, S.L. Hoffman, 'Chromosome 2 Sequence of the Human Malaria Parasite *Plasmodium falciparum', Science*, **282**, 126–132 (1998).
17. C. Aston, C. Hiort, D.C. Schwartz, 'Optical Mapping: An Approach for Fine Mapping', *Methods Enzymol.*, **303**, 55–73 (1999).
18. Y.T.J. Silhavy, M.L. Berman, L.W. Enquest, *Experiments with Gene Fusion*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1984.
19. E.S. Lander, M.S. Waterman, 'Genomic Mapping by Fingerprinting Random Clones: a Mathematical Analysis', *Genomics*, **2**, 231–239 (1988).
20. J. Eigner, P. Doty, 'The Native, Denatured and Renatured States of Deoxyribonucleic Acid', *J. Mol. Biol.*, **12**, 549–580 (1965).
21. W. Wang, J. Lin, D.C. Schwartz, 'Scanning Force Microscopy of DNA Molecules Elongated by Convective Fluid Flow in an Evaporating Droplet', *Biophys. J.*, **75**, 513–520 (1998).
22. J. Reed, E. Singer, G. Kresbach, D.C. Schwartz, 'A Quantitative Study of Optical Mapping Surfaces by Atomic Force Microscopy and Restriction Endonuclease Digestion Assays', *Anal. Biochem.*, **259**, 80–88 (1998).
23. A. Samad, E.J. Huff, W. Cai, D.C. Schwartz, 'Optical Mapping: A Novel, Single-molecule Approach to Genomic Analysis', *Genome Res.*, **5**, 1–4 (1995).
24. H. Yokoda, F. Johnson, H. Lu, R.M. Robinson, A.M. Belu, M.D. Garrison, B.D. Ratner, B.J. Trask, D.L. Miller, 'A New Method for Straightening DNA Molecules for Optical Restriction Mapping', *Nucl. Acids Res.*, **25**, 1064–1070 (1997).
25. V.N. Morozov, T.Y. Morozova, C. Hiort, D.C. Schwartz, 'New Polyacrylamide Gel-based Methods of Sample Preparation for Optical Microscopy: Immobilization of DNA Molecules for Optical Mapping', *J. Microsc.*, **183**, 205–214 (1996).
26. M. Schena, D. Shalon, R.W. Davis, P.O. Brown, 'Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray', *Science*, **270**, 467–470 (1995).
27. R.A. Heller, M. Schena, A. Chai, D. Shalon, T. Bedilion, J. Gilmore, D.E. Woolley, R.W. Davis, 'Discovery and Analysis of Inflammatory Disease-related Genes Using cDNA Microarrays', *Proc. Natl. Acad. Sci. USA*, **94**, 2150–2155 (1997).

28. G. Craig, D. Nizetic, J.D. Hoheisel, G. Zehetner, H. Lehrach, 'Ordering of Cosmid Clones Covering the Herpes Simplex Virus Type I (HSV-I) Genome: a Test Case for Fingerprinting by Hybridization', *Nucl. Acids Res.*, **18**, 2653–2660 (1990).
29. D. Nizetic, G. Zehetner, A.P. Monaco, L. Gellen, B.D. Young, H. Lehrach, 'Construction, Arraying, and High-density Screening of Large Insert Libraries of Human Chromosomes X and 21: their Potential Use as Reference Libraries', *Proc. Natl. Acad. Sci. USA*, **88**, 3233–3237 (1991).
30. H.S. Rye, S. Yue, D.E. Wemmer, M.A. Quesada, R.P. Haugland, R.A. Mathies, A.N. Glazer, 'Stable Fluorescent Complexes of Double-stranded DNA with Bis-intercalating Asymmetric Cyanine Dyes: Properties and Applications', *Nucl. Acids Res.*, **20**, 2803–2812 (1992).
31. I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 75–76, 1992.
32. W. Grimson, D. Huttenlocher, 'On the Verification of Hypothesized Matches in Model-based Recognition', *IEEE Trans. on Pattern Recognition and Machine Intelligence*, **13**, 1201–1213 (1991).
33. D.H. Ballard, C.M. Brown, *Computer Vision*, Prentice Hall, NJ, 1982.
34. S. Paxia, B. Mishra, (manuscript in preparation).
35. W. Bautsch, U. Romling, K. Schmidt, A. Samad, D.C. Schwartz, B. Tummler, 'Long Range Restriction Mapping of Genomic DNA', in *Genome Mapping, A Practical Approach*, ed. P.H. Dear, Oxford University Press, 281–313, 1997.
36. J. Giacalone, S. Delobette, V. Gibaja, L. Ni, J. Skiadas, R. Qi, J. Eddington, Z. Lai, D. Gebauer, H. Zhao, T.S. Anantharaman, B. Mishra, L.G. Brown, R. Saxena, D.L. Page, D.L. Schwartz, 'Optical Mapping of BAL Clones from the Human Y Chromosome DAZ Locus' (submitted).
37. S.B. Primrose, *Principles of Genome Analysis: A Guide to Mapping and Sequencing DNA from Different Organisms*, Blackwell Science, Oxford, UK, 1998.
38. F. Antequera, A. Bird, 'Number of CpG Islands and Genes in Human and Mouse', *Proc. Natl. Acad. Sci. USA*, **90**, 11 995–11 999 (1993).
39. F. Larsen, G. Gundersen, R. Lopez, H. Prydz, 'CpG Islands as Gene Markers in the Human Genome', *Genomics*, **13**, 1095–1107 (1992).
40. G.R. Cutting, S.E. Antonarakis, H. Youssoufian, H.H. Kazazian, Jr, 'Accuracy and Limitations of Pulsed Field Gel Electrophoresis in Sizing Partial Deletions of the Factor VIII Gene', *Mol. Biol. Med.*, **5**, 173–184 (1988).
41. Y. Kohara, K. Akiyama, K. Isono, 'The Physical Map of the Whole *E. coli* Chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library', *Cell*, **50**, 495–508 (1987).
42. A.J. Link, M.V. Olson, 'Physical Map of the *Saccharomyces cerevisiae* Genome at 110-Kilobase Resolution', *Genetics*, **127**, 681–698 (1991).
43. L. Riles, J.E. Dutchik, A. Baktha, B.K. McCauley, E.C. Thayer, M.P. Leckie, V.V. Braden, J.E. Depke, M.V. Olson, 'Physical Maps of the Six Smallest Chromosomes of *Saccharomyces cerevisiae* at a Resolution of 2.6 Kilobase Pairs', *Genetics*, **134**, 81–150 (1993).
44. J. Jing, (manuscript in preparation).
45. M. Speicher, S. Ballard, D. Ward, 'Karyotyping Human Chromosomes by Combinatorial Multi-fluor FISH', *Nat. Genet.*, **12**, 368–375 (1996).
46. E. Schrock, S. Manoir, J. Veldman, B. Schoell, J. Weinberg, M. Ferguson-Smith, Y. Ning, D. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini, T. Ried, 'Multicolor Spectral Karyotyping of Human Chromosomes', *Science*, **273**, 494–497 (1996).
47. S. Hayashi, N. Miharu, E. Okamoto, O. Samura, T. Hara, K. Ohama, 'Detection of Chromosomal Abnormalities in Uterine Leiomyoma Using Conventional Cytogenetic Method and Interphase Fluorescence In Situ Hybridization', *Cancer Genet. Cytogenet*, **89**, 98–104 (1996).
48. L. Karenko, E. Hyytinen, S. Sarna, A. Ranki, 'Chromosomal Abnormalities in Cutaneous T-cell Lymphoma and in its Premalignant Conditions as Detected by G-banding and Interphase Cytogenetic Methods', *J. Invest Dermatol.*, **108**, 22–29 (1997).
49. H.J. Tanke, R.J. Florijn, J. Vrolijk, A.K. Raap, 'Molecular Cytogenetics: Unraveling of the Genetic Composition of Individual Cells by Fluorescence In Situ Hybridization and Digital Imaging Microscopy', *World J. Urol.*, **13**, 138–142 (1995).
50. A. Kallioniemi, T. Visakorpi, R. Karhu, D. Pinkel, O.P. Kallioniemi, 'Gene Copy Number Analysis by Fluorescence In Situ Hybridization and Comparative Genomic Hybridization', *Methods: A Companion to Methods Enzymol.*, **9**, 113–121 (1996).
51. H. Yokota, G. van den Engh, J.E. Hearst, R.K. Sachs, B.J. Trask, 'Evidence for the Organization of Chromatin in Megabase Pair-sized Loops Arranged Along a Random Walk Path in the Human G0/G1 Interphase Nucleus', *J. Cell Biol.*, **130**, 1239–1249 (1995).
52. M. Speicher, S. Ballard, D. Ward, 'Karyotyping Human Chromosomes by Combinatorial Multi-fluor FISH', *Nat. Genet.*, **12**, 368–375 (1996).
53. T. Veldman, C. Vignon, E. Schrock, J.D. Rowley, T. Ried, 'Hidden Chromosome Abnormalities in Haematological Malignancies Detected by Multicolor Spectral Karyotyping', *Nat. Genet.*, **15**, 406–410 (1997).
54. R.J. Florijn, L.A. Blonden, H. Vrolijk, J. Wiegant, J.W. Vaandrager, F. Baas, J. Dunnen, H. Tanke, G.J. van Ommen, A.K. Raap, 'High Resolution DNA Fiber-fish for Genomic DNA Mapping and Color Bar-coding of Large Genes', *Hum. Mol. Genet.*, **4**, 831–836 (1995).

55. H.U. Weier, M. Wang, J.C. Mullikin, Y. Zhu, J.F. Cheng, K.M. Greulich, A. Bensimon, J.W. Gray, 'Quantitative DNA Fiber Mapping', *Hum. Mol. Genet.*, **4**, 1903–1910 (1995).
56. R.J. Florijn, F.M. van de Rijke, H. Vrolijk, L.A. Blonden, M.H. Hofker, J.T. den Dunnen, H.J. Tanke, G.J. van Ommen, A.K. Raap, 'Exon Mapping by Fiber-FISH or LR-PCR', *Genomics*, **38**, 277–282 (1996).
57. T. Duell, M. Wang, J. Wu, U.J. Kim, H.U. Weier, 'High-resolution Physical Map of the Immunoglobulin Lambda Variant Gene Cluster Assembled by Quantitative DNA Fiber Mapping', *Genomics*, **45**, 479–486 (1997).
58. T. Duell, L.B. Nielsen, A. Jones, S.G. Young, H.U. Weier, 'Construction of Two Near-kilobase Resolution Restriction Maps of the 5′ Regulatory Region of the Human Apolipoprotein B gene by Quantitative DNA Fiber Mapping (QDFM)', *Cytogenet. Cell Genet.*, **79**, 64–70 (1997).
59. M. Nilsson, K. Krejci, J. Koch, M. Kwiatkowski, P. Gustavsson, U. Landegren, 'Padlock Probes Reveal Single-nucleotide Differences, Parent of Origin and In Situ Distribution of Centromeric Sequences in Human Chromosomes 13 and 21', *Nat. Genet.*, **16**, 252–255 (1997).
60. P.M. Lizardi, X. Huang, Z. Zhu, P. Bray-Ward, D.C. Thomas, D.C. Ward, 'Mutation Detection and Single-molecule Counting Using Isothermal Rolling Circle Amplification', *Nat. Genet.*, **19**, 25–232 (1998).
61. A. Kallioniemi, R. Karhu, O.P. Kallioniemi, 'Comparative Genomic Hybridization', in *Laboratory Methods for the Detection of Mutations and Polymorphisms in DNA*, ed. G.R. Taylor, CRC Press, Boca Raton, 273–285, 1997.
62. J. Kraus, R.G. Weber, M. Cremer, T. Seebacher, C. Fischer, C. Schurra, A. Jauch, P. Lichter, A. Bensimon, T. Cremer, 'High-resolution Comparative Hybridization to Combed DNA Fibers', *Hum. Genet.*, **99**, 374–380 (1997).
63. F. Pellestor, A. Girardet, G. Lefort, B. Andreo, J.P. Charlieu, 'Rapid In Situ Detection of Chromosome 21 by PRINS Technique', *Am. J. Med. Genet.*, **56**, 393–397 (1995).
64. A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot, D. Bensimon, 'Alignment and Sensitive Detection of DNA by a Moving Interface', *Science*, **265**, 2096–2098 (1994).
65. X. Michalet, R. Ekong, F. Fougerousse, S. Rousseaux, C. Schurra, N. Hornigold, M. van Slegtenhorst, J. Wolfe, S. Povey, J.S. Beckmann, A. Bensimon, 'Dynamic Molecular Combing: Stretching the Whole Human Genome for High-resolution Studies', *Science*, **277**, 1518–1523 (1997).
66. D.C. Schwartz, A. Samad, 'Optical Mapping Approaches to Molecular Genomics', *Curr. Opin. Biotechnol.*, **8**, 70–74 (1997).
67. N. Risch, K. Merikangas, 'The Future of Genetic Studies of Complex Human Diseases', *Science*, **273**, 1516–1517 (1996).
68. S.A. Fodor, 'Massively Parallel Genomics', *Science*, **277**, 393–395 (1997).
69. R. Lipshutz, D. Morris, M. Chee, E. Hubbell, M. Kozal, N. Shah, R. Yang, S.P. Fodor, 'Using Oligonucleotide Probe Arrays to Access Genetic Diversity', *BioTechniques*, **19**, 442–447 (1995).
70. M. Chee, R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, S.P. Fodor, 'Accessing Genetic Information with High-density DNA Arrays', *Science*, **274**, 610–614 (1996).
71. D.G. Wang, J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, E.S. Lander, 'Large-scale Identification, Mapping, and Genotyping of Single-nucleotide Polymorphisms in the Human Genome', *Science*, **280**, 1077–1082 (1998).
72. R.J. Cho, M. Fromont-Racine, L. Wodicka, B. Feierbach, T. Stearns, P. Legrain, D.J. Lockhart, R. Davis, 'Parallel Analysis of Genetic Selections Using Whole Genome Oligonucleotide Arrays', *Proc. Natl. Acad. Sci. USA*, **95**, 3752–3757 (1998).
73. L. Wodicka, H. Dong, M. Mittmann, M.H. Ho, D.J. Lockhart, 'Genome-wide Expression Monitoring in *Saccharomyces cerevisiae*', *Nat. Biotechnol.*, **15**, 1359–1367 (1997).
74. M. Yanagida, Y. Morikawa, S. Hiraoka, T. Matsumoto, S. Uemora, S. Okada, in *Applications of Fluorescence in the Biomedical Sciences*, eds. D.L. Taylor, A. Waggoner, R. Murphy, F. Lanni, R. Ridge, Alan Liss, New York, NY, 321–345, 1986.
75. D.C. Schwartz, M. Koval, 'Conformational Dynamics of Individual DNA Molecules During Gel Electrophoresis', *Nature*, **338**, 520–522 (1989).
76. T.W. Houseal, C. Bustamante, R.F. Stump, M.F. Maestre, 'Real-time Imaging of Single DNA Molecules with Fluorescence Microscopy', *Biophys. J.*, **56**, 507–516 (1989).
77. X. Guo, E.J. Huff, D.C. Schwartz, 'Sizing of Large DNA Molecules by Hook Formation in a Loose Matrix', *J. Biomol. Struct. Dynam.*, **11**, 1–10 (1993).
78. S.B. Smith, C. Heller, C. Bustamante, 'Model and Computer Simulations of the Motion of DNA Molecules During Pulse Field Gel Electrophoresis', *Biochemistry*, **30**, 5264–74 (1991).
79. S. Gurrieri, E. Rizzarelli, D. Beach, C. Bustamente, 'Imaging of Kinked Configurations of DNA Molecules Undergoing Orthogonal Field Alternating Gel Electrophoresis by Fluorescence Microscopy', *Biochemistry*, **29**, 3396–3401 (1990).
80. D.P. Allison, P.S. Kerper, M.J. Doktycz, T. Thundat, P. Modrich, F.W. Larimer, D.K. Johnson, P.R. Hoyt, M.L. Mucenski, R.J. Warmack, 'Mapping Individual

Cosmid DNAs by Direct AFM Imaging', *Genomics*, **41**, 379–384 (1997).
81. D. Keller, C. Bustamante, R. Keller, 'Imaging of Single Uncoated DNA Molecules by Scanning Tunneling Microscopy', *Proc. Natl. Acad. Sci. USA*, **86**, 5356–5360 (1989).
82. C. Bustamante, J. Vesenka, C. Tang, W. Rees, M. Guthold, R. Keller, 'Circular DNA Molecules Imaged in Air by Scanning Force Microscopy', *Biochemistry*, **31**, 22–26 (1992).
83. W.A. Rees, R.W. Keller, J. Vesenka, G. Yang, C. Bustamante, 'Evidence of DNA Bending in Transcription Complexes Imaged by Scanning Force Microscopy', *Science*, **260**, 1646–1649 (1993).
84. S.B. Smith, L. Finzi, C. Bustamante, 'Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads', *Science*, **258**, 1122–1126 (1992).
85. T.R. Strick, J.F. Allemand, D. Bensimon, V. Croquette, 'Behavior of Supercoiled DNA', *Biophys. J.*, **74**, 2016–2028 (1998).
86. T.R. Strick, J.F. Allemand, D. Bensimon, A. Bensimon, V. Croquette, 'The Elasticity of a Single Supercoiled DNA Molecule', *Science*, **271**, 1835–1837 (1996).
87. Y.K. Wang, E.J. Huff, D.C. Schwartz, 'Optical Mapping of Site-directed Cleavages on Single DNA Molecules by the RecA-assisted Restriction Endonuclease Technique', *Proc. Natl. Acad. Sci. USA*, **92**, 165–169 (1995).
88. M. Newman, T. Strzelecka, L.F. Dorner, I. Schildkraut, A.K. Aggarwal, 'Structure of *Bam*H I Endonuclease Bound to DNA: Partial Folding and Unfolding on DNA Binding', *Science*, **269**, 656–663 (1995).
89. V. Mizrahi, S.J. Benkovic, 'The Dynamics of DNA Polymerase-catalyzed Reactions', *Adv. Enzymol. Relat. Areas Mol. Biol.*, **61**, 437–457 (1988).
90. M.P. Deutscher, A. Kornberg, 'Enzymatic Synthesis of Deoxyribonucleic Acid. XXIX. Hydrolysis of Deoxyribonucleic Acid from the 5' Terminus by an Exonuclease Function of Deoxyribonucleic Acid Polymerase', *J. Biol. Chem.*, **244**, 3029–3037 (1969).

PNA and Its Applications

Arghya Ray and Bengt Nordén
Chalmers University of Technology, Gothenburg, Sweden

1 Introduction 1
2 Chemical Methods – Chemistry and Synthesis 2
2.1 Synthesis 2
2.2 Purification 2
3 Electrophoretic and Hybridization Techniques 2
3.1 PNA as a Tool for Molecular Biology/Biotechnology Research 2
3.2 Possible Applications of PNA in Medicine 4
3.3 PNA as a Diagnostic Tool 7
4 Spectroscopic and Crystallographic Methods 12
4.1 Electronic Spectroscopy 12
4.2 Vibrational Spectroscopy 15
4.3 Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry 16
4.4 Crystallography and Its Applications 17
5 Conclusions 18
Abbreviations and Acronyms 18
Related Articles 19
References 19

In recent years, there has been considerable interest in developing gene-targeting drugs. Synthetic molecules that bind effectively with high sequence specificity to a desired target in a particular ''destined'' gene sequence are of major interest in medicine and molecular biotechnological research. They show hope for the development of gene therapeutic agents, high-precision diagnostic devices for genetic analysis and convenient tools for nucleic acid manipulations. Peptide nucleic acid (PNA) is a nucleic acid analog in which the sugar–phosphate backbone of natural nucleic acid has been replaced by a synthetic peptide backbone usually formed from N*-(2-aminoethyl)glycine units resulting in an achiral and uncharged mimic. It is chemically stable and, in contrast to nucleic acids and peptides, is naturally resistant to hydrolytic (enzymatic) cleavage and thereby not expected to be degraded by cell extracts or inside a living cell. It is capable of sequence-specific recognition of DNA and RNA obeying a Watson–Crick hydrogen bonding scheme and the hybrid complexes exhibit extraordinary thermal stability and unique ionic strength effects. It also recognizes duplex homopurine sequences of DNA to which it binds by strand invasion, forming a stable PNA·DNA–PNA triplex with a looped-out DNA strand. This potential synthetic mimic of DNA, since its discovery in 1991, has attracted great attention within medicinal chemistry and molecular biology and also various other fields such as organic and physical chemistry because of its interesting chemical and physical properties. It has been regarded to have great potential for both diagnostic and pharmaceutical applications. In vitro studies have shown successful attempts to use PNA as an inhibitor for both transcription and translation of specific genes. This holds promise for using PNA for antigene and antisense therapy. However, the most exciting use of PNA, as an antisense or antigene drug, requires an efficient and safe method to deliver PNA into living cells, a problem that still remains to be solved.*

1 INTRODUCTION

Historically, PNA originates from attempts during the 1980s in the Danish organic chemist Ole Buchardt's laboratory to develop new DNA sequence-specific reagents. Based on the observation with flow linear dichroism (LD) that α-helical poly-γ-benzylglutamate (PBG) forms stacking complexes with aromatic chromophores, it was suggested that PBG with alternating nucleobases and acridine moieties instead of phenyls might bind sequence selectively to duplex DNA by combined Hoogsteen base-pair (bp) formation and intercalation with the helix backbone in the major groove. The suggested compound was tentatively named *peptide nucleic acid, PNA*.

Today's PNAs (Figure 1a and b) are DNA analogs in which generally a 2-aminoethylglycine linkage replaces the normal phosphodiester backbone.[1,2] A methyl carbonyl linker connects standard (in some case unusual) nucleotide bases to this backbone at the amino nitrogens. The nonprototype, yet interesting, chemistry of this synthetic molecule has three important consequences: peptide nucleic acids are neutral molecules, they are achiral and they are not susceptible to any hydrolytic (enzymatic cleavage). Despite these great differences from DNA, PNA is capable of sequence-specific recognition of DNA and RNA obeying the Watson–Crick hydrogen bonding rules[3,4] and the hybrid complexes thus formed exhibit extraordinary thermal stability and unique ionic strength properties. Regarding applications of PNA, one obvious question is what the advantage is

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

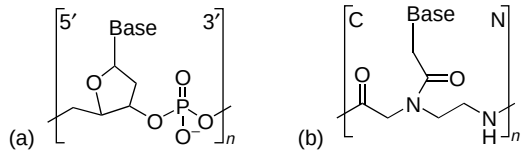

Figure 1 Chemical structures of (a) DNA and (b) PNA.

of using PNA over DNA. Scientists have been working towards the development of a wide range of applications for PNAs since 1991.[1] These can be generalized in four categories: first, its use as a tool for molecular biology and biotechnology;[5–9] second, towards the development of a gene-targeted drug using antigene or antisense strategy;[10–12] third, the use of PNA for diagnostics purpose and towards the development of biosensors;[13–16] and fourth, the study of basic chemistry to address the problems related to improvement of PNA.[17–20] As we review PNA applications, we shall focus on various analytical techniques necessary for the synthesis and characterization of novel PNA constructs, their development and application as gene therapeutic agents and detection systems for genetic analysis, and finally on understanding the physicochemical aspects of PNA and PNA–nucleic acid complexes.

2 CHEMICAL METHODS – CHEMISTRY AND SYNTHESIS

2.1 Synthesis

The backbone of PNA carries 2-aminoethylglycine linkages to replace the normal phosphodiester backbone of DNA.[1,2] The nucleotide bases are connected to this backbone at the amino nitrogens through a methylene carbonyl linker (Figure 1a and b). Since PNAs are achiral molecules, there is no need to develop a stereoselective pathway for their synthesis. PNA oligomers are prepared following the well-established solid-phase peptide synthesis protocols.[21,22] The protecting scheme for the amino groups of PNA monomers is based on either Boc or Fmoc chemistry.[23–25] In principle, PNA oligomers are synthesized using standard solid-phase peptide synthesis procedures with a (methylbenzhydryl)amine polystyrene resin as the solid support.[23,26,27] Several modifications of PNA, related to various functional requirements, have been reported.[27] It is possible to modify PNA oligomers either by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine group. The postsynthetic modification process employs coupling to an introduced lysine or cysteine residue in the PNA.[23,27] A bis-PNA is prepared in a continuous synthesis process by connecting two PNA segments via a flexible linker composed of multiple units of either 6-aminohexanoic acid or 8-amino-3,6-dioxaoctanoic acid.[28] It has been reported that changing cytosine to pseudoisocytosine in the Hoogsteen strand abolishes the pH dependence of a DNA·PNA–DNA triplex without affecting the stability of the triplex.[28]

2.2 Purification

Postsynthesis procedures involve cleaving PNA oligomers from the solid support by treatment with either anhydrous hydrogen fluoride or trifluoromethanesulfonic acid, as done in traditional peptide chemistry, followed by high-performance liquid chromatography (HPLC) purification.[23–30] In some cases, crude PNA products are cleaved from the solid matrix by treatment with trifluoroacetic acid–*m*-cresol (4 : 1) and precipitated in ice-cold diethyl ether and thereafter dissolved in 0.1% trifluoroacetic acid solution.[23,29,30] PNA molecules are further purified and characterized by reversed-phase HPLC followed by matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI/TOFMS).[23,28–30] We shall discuss this later.

3 ELECTROPHORETIC AND HYBRIDIZATION TECHNIQUES

PNAs can hybridize to complementary DNA and RNA sequences in a sequence-dependent manner obeying the Watson–Crick hydrogen bonding scheme. A unique multiple-stranded triplex structure containing both Watson–Crick and Hoogsteen base pairing is also possible from a 2 : 1 PNA–DNA complex.[31–34] The mechanistic pathway for the formation of the triplex is known as ''strand-invasion''. PNA–DNA hybridization is more affected by base mismatches than corresponding DNA–DNA hybridization. Surprisingly, PNA can maintain sequence discrimination up to the level of a single mismatch. High-affinity binding, a high level of mismatch discrimination and the novel strand-invasion property have allowed researchers to develop several PNA-based applications, employing the basic principles of electrophoresis and hybridization, in conjunction with molecular biology/biotechnology, medicine (as gene-targeted drug) and diagnostics.

3.1 PNA as a Tool for Molecular Biology/Biotechnology Research

The following applications evidence the potential of PNAs as a tool for molecular biology and biotechnology

and justify the development of new methodologies related to that.

3.1.1 PNA Hybridization as an Alternative to Southern Hybridization Technique

Southern hybridization is perhaps one of the most widely used techniques in molecular biology. It is used to analyze genetic loci, employing fragments of defined length obtained from either polymerase chain reaction (PCR) or restriction digestion. The great utility of this core molecular biological technique is derived from its ability to predict both size and sequence information and information regarding the genetic context. However, the disadvantages of this process include laborious multistep washing steps and poor sequence discrimination between closely related species.

PNA pregel hybridization simplifies the process of Southern hybridization and eliminates the cumbersome postseparation, probing and washing steps. It is possible to use labeled (fluoresceinated) PNA oligomers as probes and hybridize them to a denatured DNA sample at low ionic strength, and the mixture is thereafter directly subjected to electrophoresis for size separation. Single-stranded DNA fragments are separated on the basis of their length. The charge-neutral backbone of peptide nucleic acid allows for hybridization at low ionic strength and renders a much higher mobility to the complex compared with the excess unbound PNA. The hybrids are blotted (transferred) on to a nylon membrane, dried, ultraviolet (UV) cross-linked and detected using standard chemiluminescent techniques.[8] It is also possible to detect the bound PNA by direct fluorescence detection using capillary electrophoresis (CE) (to be discussed later). The strong binding of PNA to DNA allows the PNA–DNA duplex to remain intact under conditions where the normal DNA–DNA duplex tends to disrupt. This allows sequence-specific detection with simultaneous size separation of the target DNA following a simple and straightforward protocol. Consequently, the analysis is much faster than conventional Southern hybridization technique.

3.1.2 Nucleic Acid Purification

PNAs have also been used to purify target nucleic acids. The purification process is based on its unique hybridization property. PNAs carrying six histidine residues can be used to purify target nucleic acids by nickel affinity chromatography.[6] Biotinylated PNAs in combination with streptavidin-coated magnetic beads may be used to purify *Chlamydia trachomatis* genomic DNA directly from urine samples. Inspite of being simple, fast and straightforward, this ''purification by hybridization'' approach has a number of drawbacks. First, it requires knowledge of a target sequence and depending upon that a capture oligomer has to be synthesized for each different target nucleic acid. Since the target sequences for the short pyrimidine PNA, the most efficient probe for strand invasion, are prevalent in large nucleic acids, such short PNAs can be used as generic capture probes for purification of large nucleic acids. It has been shown that a biotin-tagged PNA–thymine heptamer could be used to purify efficiently human genomic DNA from whole blood by a simple and rapid procedure.

3.1.3 Enhanced Polymerase Chain Reaction Amplification

PCR is used to amplify unique regions of DNA so that they can be detected in large genomes. PCR, now widely used in research laboratories and hospitals, relies on the ability of DNA-copying enzymes to remain stable at high temperatures. The three steps in the PCR (carried out in the same vial), viz. the separation of the strands, annealing the primer to the template and the synthesis of new strands, take less than 2 min. At the end of a cycle, each piece of DNA in the vial has been duplicated. Each newly synthesized DNA piece can act as a new template, so after 30 cycles, 1 million copies of a single piece of DNA can be produced. Taking into account the time it takes to change the temperature of the reaction vial, 1 million copies can be ready in about 3 h.

PCR has been widely used for various applications including the amplification of variable number of tandem repeat (VNTR) locus for the purpose of genetic typing.[35,36] However, in some cases there are preferential amplifications of small allelic products relative to large allelic products. This results in an incorrect typing in a heterozygous sample.[37] PNA has been used to achieve an enhanced amplification of VNTR locus D1S80.[7] This renders a genetic pattern which is fairly easy to decipher. Small PNA oligomers are used to block the template. This makes the latter unavailable for interstrand and intrastrand interaction during reassociation. However, the primer extension is not hindered. During primer extension, polymerase displaces the PNA molecules from the template and extends the primer to completion. This application demonstrates the potential of PNA for PCR amplification where fragments of different sizes are more accurately and evenly amplified. As the probability of differential amplification is less, the risk of misclassification is greatly reduced.

A recent report by Misra et al.[38] has demonstrated that PNA–DNA chimera[39] lacking the true phosphate backbone can actually act as a primer for the polymerase reaction catalyzed by DNA polymerases. The analog, $(PNA)_{19}$-T_PG-OH, consists of 19 bases in the PNA part followed by a dinucleotide (T_PG-OH) with a single

phosphate and a free 3′-OH terminus. This chimera, when annealed with a complementary RNA or DNA template strand, can be used as an efficient primer to catalyze the addition of nucleotide by polymerase enzymes. The primer is recognized by two reverse transcriptases and also by the Klenow fragment of *Escherichia coli.* DNA polymerase I. From the results it has been proposed that the phosphate backbone of the template primer may not be essential, at least not in this case, for recognition and binding by the polymerases. The results also suggested that the diameter of the duplex region is a more critical factor than the phosphate backbone regarding the template-primer reaction and accommodating the same within the binding domain of the enzyme.

3.1.4 PNA and S1 Nuclease as Artificial Restriction Enzymes

S1 nuclease degrades single-stranded nucleic acids releasing 5′-phosphoryl mono- or oligonucleotides. The enzyme is basically a glycoprotein with a carbohydrate content of 18%. It is extensively used to remove the single-stranded overhangs of DNA fragments, RNA transcript mapping and construction of unidirectional deletions. PNA in combination with S1 nuclease can act as an ''artificial restriction enzyme'' system. Homopyrimidine PNA oligomers can hybridize with the complementary targets on dsDNA via a strand invasion mechanism, leading to the formation of looped-out noncomplementary DNA strands. Upon treatment by nuclease S1, the single-stranded DNA part is cut into well-defined fragments. The binding of two PNAs to two adjacent targets, on either the same or opposite DNA strands, essentially opens up the entire region making the substrate more accessible for the nuclease digestion and thereby increasing the cleavage efficiency.[5]

3.1.5 PNA-assisted Rare Cleavage

Peptide nucleic acids, in combination with methylases and other restriction endonucleases, can act as rare genome cutters.[9] The technique, called the PNA-assisted rare cleavage (PARC) technique (Figure 2), employs the strong and sequence-specific binding of PNAs, preferably bis-PNAs, to short homopyrimidine sites on large DNA molecules, e.g. yeast or λ DNA. The PNA binding site is selected in such a fashion that it overlaps with the methylation/restriction enzyme site on the DNA. When bound to DNA, this PNA molecule will efficiently shield the site from enzymatic methylation whereas the other so-called naked or unprotected methylation/restriction sites will be methylated. After the removal of bis-PNA, it is possible to cleave the whole DNA by restriction enzymes into a limited number of pieces. DNA is efficiently

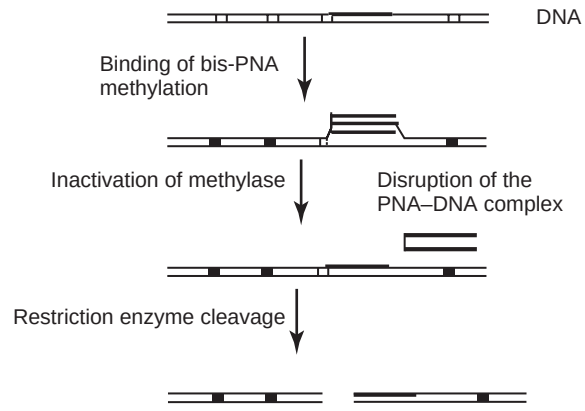

Figure 2 Schematic representation of the PARC technique. The symbols ——, and ■ represent PNA binding site, restriction enzyme site and methylated restriction enzyme site, respectively. Methylated sites are not subjected to a restriction digestion.

protected from enzymatic digestion due to methylation in most of the sites and is only cleaved through those sites which were previously bound to PNA. Thus, short PNA sequences, particularly positively charges bis-PNAs, combined with various methylation/restriction enzyme pairs, are capable of providing a new class of genome rare cutters.

3.1.6 Determination of Telomere Size

The standard procedure for the determination of telomere length involves Southern blot analysis of genomic DNA and provides a range for the telomere length of all chromosomes present. Another approach includes the use of fluorescein-labeled oligonucleotides and monitor the in situ hybridization of them to telomeric repeats. Much better and more quantitative results are obtained when DNA oligonucleotides are replaced by fluorescein-labeled PNAs, as shown by Lansdorp et al.[40] This PNA-mediated approach allows accurate estimates of the telomeric length to be made. In situ hybridization of fluorescein-labeled PNAs to telomeres is faster than to DNA hybridization and requires a lower concentration of the probe. Low photobleaching and an excellent signal-to-noise ratio make it possible to quantitate telomeric repeats on individual chromosomes in this way. It is likely that variations of this approach can be applied to other repetitive sequences to achieve similar results.

3.2 Possible Applications of PNA in Medicine

The potential of PNA to act as a genetic drug has been well appreciated. It is clearly a substantial lead in gene therapeutic drug design provided that we have well-identified targets for it. Two general strategies

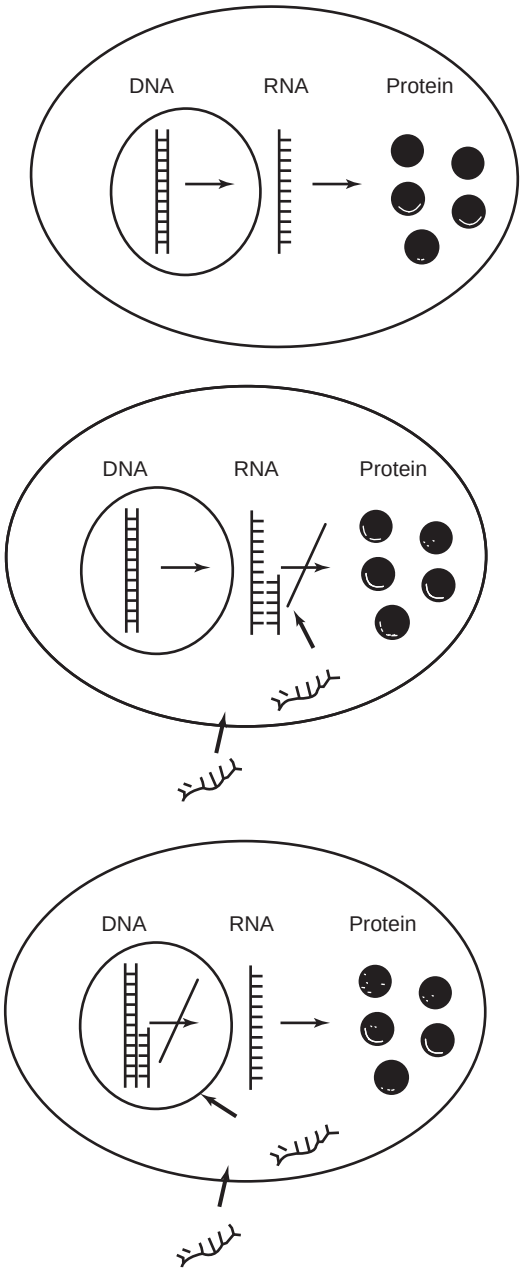

Figure 3 Antigene and antisense strategy: cells can be treated with an antigene oligonucleotide (such as PNA), which can bind to a (specific) complementary sequence in the DNA and inhibits transcription of the gene. On the other hand, an antisense oligomer can hybridize to a specific mRNA sequence to inhibit the expression of a protein at the level of translation.

are applicable towards gene therapeutic drug design (Figure 3). Oligonucleotides or their potential analogs can be designed to recognize and hybridize to complementary sequences in a particular gene and they can interfere with transcription of that particular gene (antigene strategy). Alternatively, analogs can be designed to recognize and hybridize to complementary sequences in mRNA and thereby inhibit its translation (the antisense strategy). Peptide nucleic acids are biologically and chemically stable and have significant effects on replication, transcription and translation, as revealed from in vitro measurements. Moreover, so far no signs of general toxicity of PNA have been observed. PNA binding to mRNA interferes with translation, PNA–dsDNA strand displacement complexes inhibit protein binding and block RNA polymerase elongation.

3.2.1 Transcription Arrest by PNA: the Antigene Effect

Homopyrimidine PNAs form extremely stable strand-displacement complexes upon binding to dsDNA. In principle, they should be capable of working as antigene agents. In reality, such complexes are indeed capable of affecting the process of transcription involving both prokaryotic and eucaryotic RNA polymerases in vitro. PNA bound to a promoter region is capable of blocking the DNA access of the corresponding polymerase protein. PNA strand-displacement complexes, located far downstream from the promoter, can also efficiently block polymerase progression and transcription elongation and thereby produce truncated RNA transcripts. This requires that the PNA target, i.e. the DNA polypurine target, must be present in the gene of interest. It has been found that even an 8-mer PNA (T_8) is able to block phage T_3 polymerase in action.[41] The effect of having a PNA target within a promoter region on gene expression has been demonstrated for IL-2Rα gene.[10,42] The PNA_2–DNA triplex thus can arrest transcription in vitro and can act as an antigene agent. However, one of the major obstacles to applying PNA as an antigene agent is that the strand invasion or the formation of strand displacement complex occurs very slowly at physiological salt concentration.[43,44] Several modifications of PNA have shown improvement in terms of binding. Modifications of PNA by the incorporation of intercalators,[45] positively charged lysine residues,[28,46] chemically linking the ends of the Watson–Crick and Hoogsteen PNA strands to each other[28] or pH-independent pseudoisocytosines into the Hoogsteen strand[28] can drastically increase the association rates with dsDNA. A recent report by Lee et al.[47] has demonstrated that PNA as well as the PNA–DNA chimera complementary to the primary site of the HIV-I genome can completely block priming by $tRNA_3^{lys}$ and consequently the in vitro initiation of the reverse transcription by HIV-1 RT. Their observations suggest that oligomeric PNAs targeted to various critical regions of the viral genome are likely to have a strong therapeutic potential for interrupting multiple steps involved in the replication of HIV-1.

It has been found that under physiological salt conditions, binding of PNA to supercoiled plasmid DNA is faster than linear DNA.[44,46] This indicates that the topology of DNA is an important factor or criterion for determining the binding rate. Since transcriptionally active chromosomal DNA is usually negatively supercoiled, this could be of relevance for PNA binding in vivo. It has also been shown that binding of PNA to dsDNA is enhanced when the DNA is being transcribed. Such transcription-mediated PNA binding occurs three times as efficiently when the PNA target is situated on the nontemplate strand than on the template strand. As transcription mediates template strand-associated complexes which arrest further elongation, the action of RNA polymerase results in repression of its own activity, i.e. *suicide transcription*.[48] These findings are highly relevant for the possible future use of PNA as an antigene agent.

It is important to mention that there are also reports regarding the ability of PNA to activate transcription which, however, is not related to its antigene effect. It has been demonstrated that the looped-out single-stranded structure formed as a result of strand invasion can also act as efficient initiation sites for *E. coli* and mammalian RNA polymerases in which the polymerase might start transcription using the single-stranded loop as a template.[49] This is consistent with the affinity of RNA polymerase for single-stranded DNA and its ability to transcribe single-stranded DNA.

3.2.2 Translation Arrest by PNA: the Antisense Effect

The principal mechanism of antisense effects by oligodeoxynucleotides is believed to be either a ribonuclease H (Rnase H)-mediated cleavage of the RNA strand in oligonucleotide–RNA heteroduplex or a steric blocking of the translation machinery at the oligonucleotide–RNA complex.[50] Oligodeoxynucleotide analogs, such as phosphorothioates, activate RNase H and thus hold the promise of working as efficient antisense agents. However, they have also been shown to exhibit nonspecific effects. PNA–RNA duplexes, on the other hand, cannot act as substrates for RNase H. Therefore, the antisense activity of PNA has to be dependent on other mechanisms. In fact, the antisense property of PNA is based on the steric blocking of either RNA processing, transport into cytoplasm or translation. In vitro translation experiments involving rabbit reticulocyte lysates have shown that both duplex- (mixed sequence) and triplex-forming (pyrimidine-rich) PNAs are capable of inhibiting translation at targets overlapping the AUG start codon.[50] Triplex-forming PNAs are able to hinder the translation machinery at targets in the coding region of mRNA. However, translation elongation arrest requires a $(PNA)_2$–RNA triplex and thus needs a homopurine target of 10–15 bases. However, duplex-forming PNAs are not capable of this. Triplex-forming PNAs can inhibit translation at initiation codon targets and ribosome elongation at codon region targets.

A recent report by Mologni et al.[51] has demonstrated the effects of three different types of antisense on the in vitro expression of PML/RARα gene. The first one was complementary to the first AUG (initiation) site. The second one was capable of binding to a sequence in the coding region that includes the second AUG, the starting site for the synthesis of an active protein. The third PNA was capable of binding to the 5′-untranslated region (5′-UTR) of the mRNA, the point of assembly of the translation machinery. These three PNAs provided an efficient inhibition when used together even at concentration much below the critical concentration used for each individual. This indicates that the PNA targeting of RNA molecules such as PML/RARα require effective blocking of different sequences on the 5′ part of the messenger. A 5′-UTR PNA target can be used as efficiently as an AUG target to achieve an antisense activity of PNA and a more effective translation inhibition can be achieved by combining PNA directed towards 5′-UTR and AUG regions.

3.2.3 Cellular Effects and Delivery of PNA

It is important to add here a few lines regarding the effect of PNA on intact cells and problems related to its delivery. Effects of PNAs on intact cells have only been demonstrated upon cellular microinjection and an antisense activity against a transfected gene has been established in this way.[10,18] However, the cellular uptake of PNA is very poor, which is still considered the major obstacle against the prospective use of PNA as a gene therapeutic drug. Efforts are being made to increase the cellular uptake of PNA, especially by modifying the molecule itself or by conjugating suitable ligands, which will result in an enhanced physical or receptor-mediated cellular uptake. Incorporation of ''guide'' sequence or some ''vector'' peptides might make it more attracted to the cell membrane and will eventually help in docking the desired PNA at the cell surface. A recent report by Aldrian-Herrada et al.[52] showed that a peptide nucleic acid is more rapidly internalized in culture neurons when coupled to a delivery peptide (retro-inverso peptide). A PNA antisense for the AUG translation initiation region of prepro-oxytocin mRNA was coupled to the vector retro-inverso peptide. The conjugated construct was internalized by cultured cerebral cortex neurons within minutes. Penetration of single PNA molecules was rather slow. It has been found that both PNA and the peptide–PNA construct lowered

the amounts of mRNA coding for prepro-oxytocin in these neurons. This shows that PNAs guided by suitable vector peptides can act as powerful antisense agents.

Another strategy which has been adapted to improvise the delivery of PNA is to incorporate it into some kind of delivery vehicles (vesicles), e.g. liposomes. However, the assessment of the latter method has to await tests in a real in vivo system involving animal models.

3.3 PNA as a Diagnostic Tool

Nucleic acid hybridization forms the basis for the diagnosis of inherited or infectious diseases. High-affinity binding of PNA oligomers to its complementary sequences provides a clear advantage for molecular recognition and leads to the development of new applications of PNA, especially as a diagnostic probe for detecting genetic mutations. Based on its unique hybridization property, a number of applications have been described utilizing PNA as a probe for the detection of genetic mutation and mismatch analysis.

In this section, we describe some of the developments related to the use of PNA as a probe to detect genetic mutation and corresponding mismatch analysis, confirming its potential as a diagnostic tool for different clinical applications. Fundamental principles related to each of these applications will also be discussed.

3.3.1 Single Base-pair Mutation Analysis Using PNA-directed Polymerase Chain Reaction Clamping

Amplification of the target nucleic acid to detectable quantities by the PCR technique is an important step regarding the detection of genetic diseases, especially those related to the single base pair mutations. The higher specificity of PNA binding to its complementary nucleic acid, higher thermal stability of a PNA–DNA duplex compared with the corresponding DNA–DNA duplex and its inefficiency to act as a primer for DNA polymerases are the basis for novel methodology. Figure 4 depicts the schematic representation of the PCR cycle and the strategy employed for this PNA-directed PCR clamping. This procedure includes a distinct annealing step involving the PNA targeted against one of the PCR primer sites. This takes place at a temperature higher than that for PCR primer annealing and thus PNA is selectively bound to the DNA molecule. The PNA–DNA complex formed at one of the primer sites can effectively block the formation of a PCR product. It was also found that PNA is able to discriminate between fully complementary and single mismatch targets in a mixed target PCR. Sequence-selective blockage by PNA

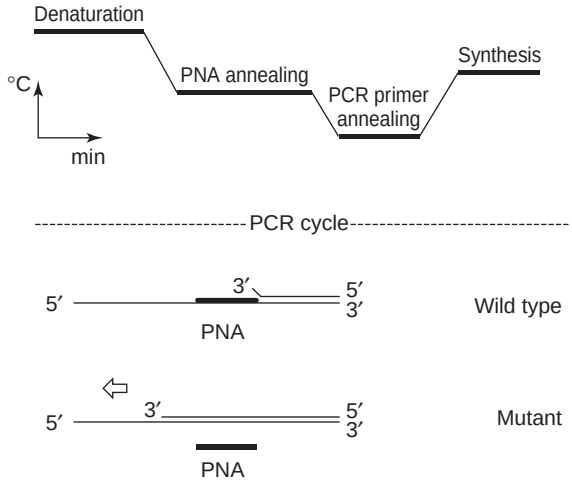

Figure 4 Mutation analysis using PNA-directed PCR clamping: schematic representation of the strategy for the PCR cycle involving PNA-directed clamping. In case of the normal (wild type) DNA, the bound PNA will sterically hinder annealing of a partially overlapping primer sequence, thus preventing the normal sequence from appropriate PCR amplification. In the case of mutant alleles, the melting temperature of the PNA/DNA is reduced and the primer can outcompete PNA annealing to carry on preferential amplification of mutant sequences.

allows suppression of target sequences that differ by only one base pair. Also, this PNA clamping was able to discriminate three different point mutations at a single position, as demonstrated in a model system by Örum et al.[13]

Based on the principle of PNA-mediated PCR clamping,[13] a study by Thiede et al.[15] has demonstrated a novel approach for simple and sensitive detection of mutations in the *ras* proto-oncogenes. The strategy behind this is schematically represented in Figure 4. Chromosomal DNA is hybridized to a 15-mer PNA complementary to a normal wild-type sequence surrounding codons 12 and 13. In the case of the normal Ki-*ras*, formation of a PNA–DNA duplex will be favored owing to its high melting temperature and thus it would sterically hinder the annealing of overlapping oligonucleotides. Hence normal Ki-*ras* sequences will not be sufficiently (PCR) amplified. On the other hand, the melting temperature of the mutated hybrid sequence is much lower than the corresponding normal one. Hence the binding of primer oligonucleotide will be relatively favored to outcompete PNA annealing. Hence the mutated sequence will be preferentially amplified.

One of the major advantages of this PNA mediated PCR clamping is that it allows the detection of mutations stretched over 4–6 bp regions in a single reaction and this could also be used to detect other hot-spot mutations.[13,15]

3.3.2 *Screening for Genetic Mutations by Capillary Electrophoresis*

Most of the powerful separation techniques for biomolecules are based on electrophoresis, i.e. the migration of electrically charged species in solutions under the influence of an electric field. In principle, to achieve the separation of two species, their electrophoretic mobilities, μ, must be different. The mobility, μ, is given by Equation (1):

$$\mu = \frac{v}{E} = \frac{Q}{f} \tag{1}$$

where v and Q are the electrophoretic velocity and the electrophoretic charge of the analyte, respectively, E is the electric field strength and f is the friction coefficient. For DNA in free solution, both Q and f are proportional to the contour length of the DNA, say L_c. In fact, it has been experimentally verified that in free solution the mobility is independent of the DNA size. A size-dependent mobility is generally obtained by using a separation medium such as a gel or a polymer solution. Instead of using a separation medium, a size-dependent mobility of DNA can also be obtained by breaking the scaling symmetry between the electrical and frictional forces either by changing the charge or the mass distribution. The electrophoresis in slabs of polyacrylamide or agarose gels has been the preferred method for separation of DNA for a long time. In recent years, especially owing to technological advances, there has been considerable interest in CE of DNA.

In CE,[53–55] separation takes place in a long, thin capillary (typically 50–80 cm long, inner diameter ~10–300 μm), generally made of fused silica. A portion of the coating is removed at the end of the capillary to detect the analyte. As the analyte passes the detection window it can be visualized by UV or laser-induced fluorescence (LIF) detection which is accomplished on-line and can be automated. CE is capable of analyzing minute amounts of sample (typically of the order of picograms to femtograms). However, the major disadvantage compared with slab-gel electrophoresis is that generally it is not possible to analyze more than one sample at a time.

The most important phenomenon described in CE is known as electroosmosis, an electrokinetic effect described as the motion of a liquid relative to a fixed charged surface caused by an electric field. Under the influence of an aqueous solution, the silica surface of the fused-silica capillaries forms silanol groups ($SiOH_2^+$, SiOH or SiO^-). This depends upon the pH values and for a pH value above 3–4, the silica surface will be negatively charged. These negative charges are counterbalanced by positive ions from the surrounding solution, forming an electrical double layer close to the surface (Figure 5).

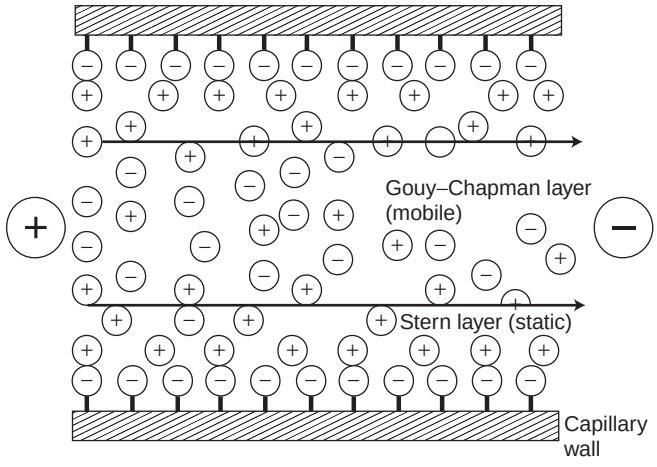

Figure 5 Illustration of the Stern–Gouy–Chapman model of the electrical double layer at the interface between the capillary wall and the electrolyte system. The arrows indicate the direction of electroosmotic flow through the capillary. *Note*: acidic silanol groups impart a negative charge on wall and counterions migrate towards cathode, dragging solvent along.

According to the model described by Stern, Gouy and Chapman,[56] the electric double layer can be described as consisting of one inner, static layer of adsorbed counterions, known as the Stern layer, and an outer, mobile layer in which ions diffuse due to thermal motion, known as the Gouy–Chapman layer. When an electric field is applied over the capillary, the positive ions migrate towards the negative electrode, dragging the mobile liquid layer with them. Although the electric double layer is only about 100 Å thick, the electroosmotic flow is transmitted throughout the whole cross-section of the capillary. In electrically driven systems, the flow caused by electroosmosis shows a uniform velocity across the capillary diameter except in the double layer region close to the wall where the flow velocity approaches zero. This so-called plug-flow appears as a result of the uniform distribution of the driving force for electroosmosis along the entire length of the capillary. For a particular field strength, the magnitude and direction of the electroosmotic flow depend on the composition of the capillary surface and on the pH and ionic strength of the electrolyte solution. Owing to the strong electroosmotic flow towards the cathode, all species, irrespective of their charge, move towards the negative electrode. Thus positively and negatively charged species can be analyzed in the same run. As the electrophoretic mobility of a free polyelectrolyte (DNA) is independent of molecular weight because the electric driving force and the frictional force are both proportional to the length (Equation 1), it may appear that electrophoresis in free solution is not a convenient tool for DNA separation. However, if the scaling symmetry between electrical and frictional

Table 1 Synthetic DNA fragments of the cystic fibrosis gene and corresponding PNA probe[14]

Nucleic acid fragment	Sequence	Mutation
Wild type (WT)	5′-TGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATA-3′	None
ΔF508	5′-TGGCACCATTAAAGAAAATATCAT---TGGTGTTTCCTATGATGAATATA-3′	Deletion of three bases
MU1	5′-TGGCACCATTAAAGAAAATATC**G**TCTTTGGTGTTTCCTATGATGAATATA-3′	One-base substitution
MU2	5′-TGGCACCATTAAAGAAAATAT**AC**TCTTTGGTGTTTCCTATGATGAATATA-3′	Two-base substitution
PNA (15-mer)	H-CACCAAAGATGATAT-lys-NH_2	–

forces can be disrupted, by changing either the charge or mass distribution or the hydrodynamics, a size-dependent mobility can be achieved for DNA.

Carlsson et al.[14] reported a novel, diagnostic method for the detection of genetic mutation using PNA as a probe for CE. The method is based on the unique hybridization of PNA and is sensitive enough to detect a single mismatch in the sample DNA. The model system consisted of four 50-mer single-stranded DNA fragments, one wild type and three mutant sequences (Table 1), representing a part of the cystic fibrosis transmembrane conductance regulator gene (CFTR). The probe was a 15-mer PNA oligomer with a sequence complementary to the wild-type sequence. The principle of the method is that PNA only binds to the DNA in the absence of any mutation in the particular region of the CFTR gene and it is possible to detect the presence of such a hybrid duplex using free solution CE. Separation of PNA–DNA duplexes was carried out at a high temperature (~70 °C) (Figure 6a–c) and 50 mM ionic strength. At this temperature and ionic strength, only the hybrid duplex carrying the wild-type DNA sequences will be detected. The hybrid complex carrying single-mismatch DNA will be melted. Free PNA is not detected because it binds to the capillary wall. In another experiment, the PNA probe was added to a PCR preparation of the 142-mer wild-type CF fragments and the 139-mer three-base deletion fragments. When studied at 50 °C, the PNA–DNA duplex was identified for the wild-type sequences but was not seen at all for the mutant.

In another study Nordén and co-workers. (unpublished observations) used free solution CE to detect point mutation in a 15-mer DNA fragment, which represents a vulnerable part of the p53 gene. The PNA probe used for this study was a 15-mer oligomer having sequence complementary to the wild-type p53 DNA (in this case the 15-mer DNA). The separation was carried out at a temperature below the thermal melting temperature, as determined from UV spectrophotometry, of the fully matched PNA–DNA hybrid, but above the thermal melting temperature of the PNA–DNA duplex carrying a single mismatch at the eighth position (corresponding to a mutation in the codon 175 of the p53 resulting in an altered P53 protein with an arginine replaced by a histidine). At this temperature, the fully matched duplex maintained its structural integrity and was detected by using UV or LIF

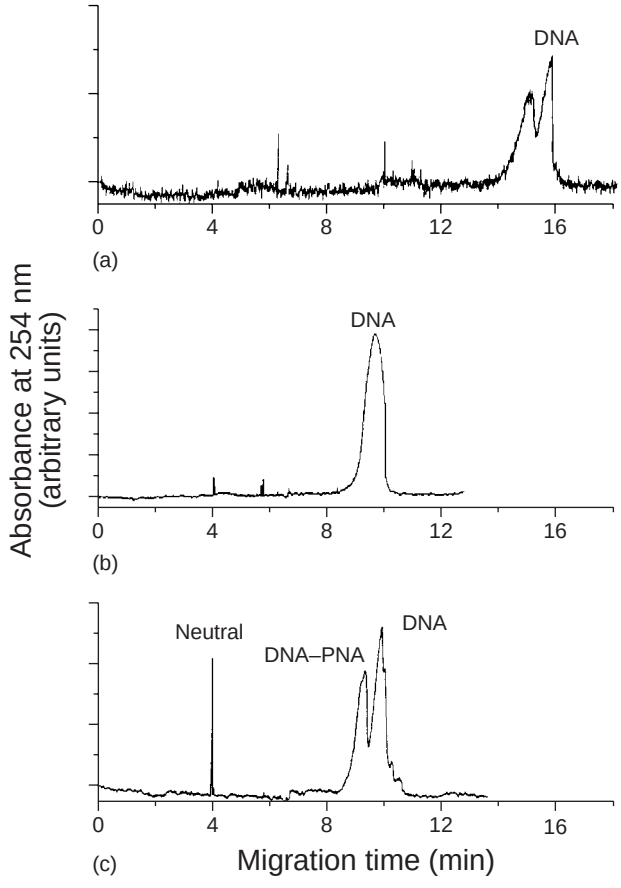

Figure 6 Free-solution CE of a mixture of PNA and DNA that has a single-base substitution in the recognition region (MU1, Table 1) at (a) 50 and (b) 70 °C. (c) Electropherogram at 70 °C of a mixture of PNA and the wild-type DNA sequence. At 70 °C, PNA cannot remain bound to the mutant DNA sequence and the peak corresponding to that hybrid duplex will not appear in the electropherogram. (Figure adapted by permission from Carlsson et al.[14])

detection. By contrast, as the PNA probe could not bind to the single-mismatch DNA sequence at this temperature, hence no peak corresponding to this duplex was noticed. In this way, the method has the capability to detect single-base substitutions. One great advantage of using PNA instead of DNA as a hybridization probe is that it gives a more efficient sequence discrimination and allows separation at high temperature, where DNA–DNA hairpin structures, responsible for band compression in gel electrophoresis, are eliminated.

More sensitive detection can be achieved by using PNA probes with a fluorescence tag, which requires only a very low concentration of the sample. If studied at high temperature, LIF detection will only produce a signal for the bound PNAs. In future, it might be possible to establish a universal screening strategy for any genetic disease with a known spectrum of mutations by developing a PNA probe library, possibly utilizing multiple fluorescence tags for multiplex testing of one or more exons.

Rose[57] reported the characterization of antisense binding properties of PNAs using capillary gel electrophoresis (CGE). Like free solution CE, only a minute amount (femtomoles) is required for the analysis. A known amount of PNA is mixed with an aliquot of an oligonucleotide stock solution and a portion of the mixture is introduced into the gel column of the capillary. The sample is then allowed to migrate through the column under the influence of an applied electric field. The area of the separated components is measured as a function of elapsed time after mixing of oligonucleotide-to-PNA ratio. Compared with free solution CE, this method shows a stronger dependence on mass. The ability of this method to resolve free and bound PNA allows the measurement of relative binding kinetics using gel-filled capillaries. This method is also useful for studying the interaction chemistry of antisense oligonucleotides as the separation mechanism of the CGE, i.e. the mobility of the experimental sample through a gel medium acting as a sieve, and permits the separation of the free oligomer from the bound species.

It is appropriate in this section to mention another gel-based technique, known as affinity electrophoresis. Affinity electrophoresis, depending on the noncovalent interaction of two species, requires the immobilization of one of the interacting species in a gel matrix.[58] It has been found that the entrapped PNA causes a retardation of the complementary oligonucleotides. PNA entrapped in a gel matrix allows specific hybridization during electrophoretic passage of a complementary DNA. Owing to the high affinity between PNA and DNA, the retardation of DNA is directly related to the binding of PNA through complementarity. This affinity electrophoresis permits the visualization of hybridization between an immobilized PNA and complementary DNA, migrating in an electric field, in real time. Moreover, a more sensitive detection of a reasonable number of samples in real time is possible by using LIF detection employing fluorescent samples.

3.3.3 PNA as a Probe for Nucleic Acid Biosensor

DNA biosensor technology, based on a very simple principle, holds promise for rapid and cost-effective detection of specific DNA sequences. A single-stranded probe oligonucleotide is immobilized onto optical, electrochemical or mass-sensitive transducers to detect the complementary (or mismatch) strand in a sample solution. The response from the hybridization event is converted into a useful electrical signal by the transducer. PNA, owing to its strong sequence-specific hybridization with DNA, can be used as the recognition layer in DNA biosensors. We describe here the use of PNA as a probe for sequence-specific biosensors.

3.3.3.1 BIAcore Technique: Kinetic Study and Mismatch Analysis It is possible to assess the kinetic parameters of PNA hybridization and make mismatch analysis using a BIAcore (biomolecular interaction analysis) instrument.[59] The system is generally used for studying real-time BIA using optical detection technology. The core part of the instrument is a sensor chip (surface) on which the real-time interactions are monitored. In BIA measurements, the probe molecule is directly attached to the surface and the analyte molecule is free in solution. The detection principle in BIA utilizes surface plasmon resonance (SPR). The response signal of the BIAcore apparatus is proportional to the change in the refractive index at the surface and is assumed to be proportional to the mass of substance bound to the chip. The important point is that this technique does not require any labeling of the interacting component and the response is primarily independent of the nature of the molecules. In a real-time BIA measurement, it is possible to monitor the changes in the surface concentration on a timescale of the order of 0.1 s, which, in principle, is sufficient enough to determine typical biomolecular kinetic rate constants in the range 10^3–$10^6\ M^{-1}\ s^{-1}$ (association) and 10^{-1}–$10^{-5}\ s^{-1}$ (dissociation).

Jensen et al.,[60] for the first time, reported an investigation of PNA–DNA/RNA hybridization using the BIAcore technique. The sensor chip used in this case was basically a thin gold surface covered with a layer of dextran that contained streptavidin chemically coupled to the dextran (Pharmacia sensor chip SA5). A biotinylated PNA [biotin-$(eg1)_3$-TGTACGTCACAACTA-NH_2; N- to C-terminal; note: PNA is generally written from N- to C-terminal] was used as a probe and it was immobilized to the streptavidin chip by employing the strong

coupling between biotin and streptavidin. By measuring the amount of bound substance (fully complementary and also various singly mismatched RNA and DNA oligonucleotides) as a function of time when a solution containing the complementary strands passes over the chip surface, the association kinetics were studied. The dissociation was correspondingly monitored by detecting the time dependence of the mass decrease as the surface was subsequently washed with buffer. It is also possible to regenerate the PNA surface by washing with HCl to remove the remaining hybridized products and perform consecutive studies with the same immobilized PNA. Assuming a two-state model, $A + B \rightleftharpoons AB$, the analysis of the hybridization kinetics was carried out. The dissociation kinetics, obtained from the dissociation part of the sensorgram (Figure 7), were analyzed using Equation (2):

$$[AB] = [AB]_0 e^{-k_d(t-t_0)} \tag{2}$$

where $[AB]_0$ represents the duplex concentration at time t_0 and k_d is the dissociation rate constant. The association kinetics were analyzed with respect to the concentration dependence of the association rate by varying the concentration of the analyzed agent, A, in the mobile phase (Equation 3):

$$[AB] = \left\{ \frac{k_a[A][B]_0}{k_a[A] + k_d} \right\} \quad 1 - e^{-(k_a[A]+k_d)t} \tag{3}$$

Using $k_a[A] + k_d = k_s$ and measuring k_s at different concentrations of A, the association rate constant can be obtained from the slope of the plot of k_s versus [A]. The results indicated that the introduction of a single mismatch at the center of 15-mer PNA–DNA and PNA–RNA duplexes gives rise to marked changes in the kinetics of both hybridization and dissociation.[60]

It is also possible to use a plain gold surface (e.g. BIAcore sensor chip J1) to immobilize PNA molecules

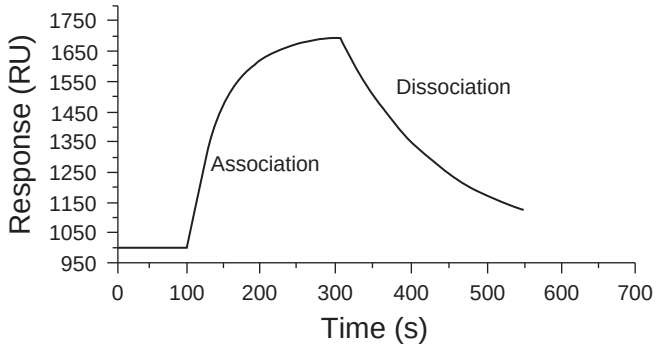

Figure 7 Typical response features (“sensorgram”) as measured by the BIAcore technique. The figure illustrates binding of an analyte to the sensor chip surface. The dissociation phase is generally followed by regeneration of the sensor chip surface.

carrying cysteine at the N-terminal side (Nordén and coworkers unpublished observations). The strong coupling between gold and sulfur was employed to immobilize the PNA molecule. This gives a better control over the surface coverage of the ligand and erroneous results due to nonspecific binding of ligands to the dextran layer can be eliminated. It is important to note that DNA has a very high affinity for gold and can be nonspecifically adsorbed on the surface without much effort. Direct addition of analyte DNA molecules to the sensor surface, to study its binding with PNA, might facilitate its adsorption to the gold surface. It is important to use short spacer molecules, e.g. mercaptohexanol, together with the ligand (probe) to form the PNA monolayer at the top of the sensor (gold) surface and prevent DNA molecules from being nonspecifically adsorbed to the surface.

3.3.3.2 Potentiometric Measurements Wang et al.[16] have reported the use of PNA as a recognition probe for the electrochemical detection of the hybridization event employing chronopotentiometric measurements. The method consisted of four steps: probe (PNA) immobilization on to the transducer surface, hybridization, indicator binding and chronopotentiometric transduction. The carbon paste electrode used in this process contained the immobilized DNA or PNA probe. The hybridization experiment was carried out by immersing the electrode in the stirred buffer solution containing a desired target. The strong adsorptive accumulation of the PNA was exploited for preparing the PNA modified surfaces and this adsorptive immobilization of short PNA oligomers leads to very effective hybridization biosensors. The potentials were sampled at a particular frequency and the derivative signal was recorded against the potential. The sensor was challenged with various noncomplementary oligomers and chromosomal DNA, including a 15-mer oligonucleotide containing a single-base mismatch. However, their results suggest that, in mixtures with the complementary strand, the noncomplementary oligomers have a profound effect upon the target signal at the DNA biosensor and a negligible effect upon that observed at the PNA-coated electrode.

3.3.3.3 Single-base Mismatch Analysis Using Quartz Crystal Microbalance Traditionally, the quartz crystal microbalance (QCM) has been used to monitor mass or thickness of thin films deposited on surfaces and to study gas adsorption and deposition on surfaces in the monolayer and submonolayer regimes.[61] The method has also been developed to use in liquid to apply it to the research areas related to electrochemistry or studying protein adsorption. The QCM is a very sensitive mass-measuring device. The resonant frequency of the crystal changes as a result of a minute weight increase on the surface. The

typical measurements include monitoring the shift in resonant frequency of the QCM due to mass changes. The technique also offers the possibility of obtaining additional information related to the dissipation factor, D, of the oscillator (the damping). Dissipation arises mainly owing to the internal friction and viscoelastic properties of the fluid. Recently, the method has developed further for use in liquids and simple techniques have been developed to measure dynamically changes in the dissipation of energy by the QCM operating in a liquid, e.g. due to adsorption on an electrode. Okahata et al.[62] reported the quantitative detection of M13 viral DNA containing a 10-base insert by hybridization to a complementary, 5′-thiol-modified, 10-base DNA probe adsorbed on a QCM. There are reports concerning the use of QCM to study sequence-specific binding of GCN4-bZIP peptides to immobilized DNA strands[63] and direct monitoring of DNA polymerase reactions on a QCM.[64] It is obvious that in situ detection of a hybridization event is possible by using QCM transducers and the method of detection is not related to the need of using an optical or redox indicator.

Since PNA hybridizes to complementary DNA strands with high affinity and sequence specificity, one might expect that immobilized PNA strands (or probes) would show an improved distinction between the closely related target sequences compared with an immobilized DNA probe. Wang et al.[65] reported a QCM biosensor based on peptide nucleic acid probes. The PNA probes, which formed the monolayer on the gold QCM surface, contained a cysteine attached to the PNA core with the help of an ethylene glycol unit. Immobilized PNA, in such a system, is capable of differentiating between a fully complementary and single mismatch oligonucleotide and a rapid detection of mismatch sequences is possible by monitoring the frequency–time response of the PNA–QCM biosensor upon addition of fully complementary and mismatch sequences. The remarkable specificity of the immobilized probe provides rapid hybridization and allows the use of low ionic strength solutions of short probes. Such a mismatch sensitivity of PNA immobilized QCM biosensors could be of great importance for diagnostic applications, particularly for genetic screening, and therapy.

4 SPECTROSCOPIC AND CRYSTALLOGRAPHIC METHODS

In this section we discuss briefly the applications of various spectroscopic methods to characterize the properties of PNA and the PNA–DNA complexes. We also discuss crystallography and its application to determining the structure of PNA and, finally, different computational techniques in connection with PNA applications.

4.1 Electronic Spectroscopy

In this section we consider different aspects of PNA applications related to ultraviolet/visible (UV/VIS) absorption spectroscopy and dichroic measurements.

4.1.1 Ultraviolet and Visible Absorption Spectroscopy

The UV absorption spectra of aqueous solutions of nucleic acids have been known since 1960. The absorption spectrum for polymeric nucleic acid(s) has been well characterized and is known to exhibit a characteristic maximum at 260 nm, a minimum around 230 nm and a second maximum below 200 nm.[66,67] Perhaps the simplest application of absorption spectroscopy to PNA is related to its concentration determination. In general, the relative mean molecular mass of a nucleotide is taken as 330 and the molar absorptivity of natural native double stranded DNA is considered as $\varepsilon_{260} = 6600\,\mathrm{M^{-1}\,cm^{-1}}$. In general, an aborbance at 260 nm (A_{260}) of 1.0 in a 1-cm cell corresponds to a DNA concentration of $50\,\mu\mathrm{g\,mL^{-1}}$. Molar absorptivities of single-strand DNA oligomers have been calculated from the base composition as a sum of the absorptivities of DNA monomers (mononucleotides) using Equation (4):

$$\varepsilon_{260} = 15.3n_{\mathrm{A}} + 7.4n_{\mathrm{C}} + 11.7n_{\mathrm{G}} + 9.3n_{\mathrm{T}} (\times 10^3\,\mathrm{M^{-1}\,cm^{-1}}) \tag{4}$$

where n_i denotes the number of bases of type I.[68–70] This equation has also been refined for a nearest neighbor model and the sequence dependence of the absorptivity of a ssDNA can be calculated using Equation (5):

$$\varepsilon_{260} = \sum_{i=1}^{N-1} \varepsilon(\mathrm{B}_i\mathrm{pB}_{i+1}) - \sum_{i=2}^{N-1} \varepsilon(\mathrm{B}_i\mathrm{p}) \tag{5}$$

where N is the total number of bases and $\varepsilon(\mathrm{B}_i\mathrm{pB}_{i+1})$ and $\varepsilon(\mathrm{B}_i\mathrm{p})$ are the molar absorptivities ($\mathrm{M^{-1}\,cm^{-1}}$) of the dinucleoside phosphates and mononucleotides, respectively, counted from the 5′-end.[71] The end nucleotides are not included in the second summation. However, the molar absorptivities for PNA monomers are not as well established as they are for DNA or RNA monomers. It is too much of a simplification to assume that a PNA oligomer should have the same molar absorptivity as its DNA or RNA counterpart. Since the backbone is different for this synthetic mimic, it is obvious that it could make a different kind of perturbation of the π-system of the nucleobases. Also, different stacking of the bases could give rise to a different extent of hypochromocity. The latter can be minimized if the absorption measurement is carried out at a high temperature (>75–80 °C). Once the molar absorptivity of a PNA sequence has been calculated, the concentration of the stock solution can be

determined by measuring the absorbance of the sample at 260 nm following the Lambert–Beer law as Equation (6):

$$A_\lambda = \log\left(\frac{I}{I_0}\right) = \varepsilon c l \qquad (6)$$

where A_λ is the absorbance at wavelength λ, I_0 and I are the intensities of light before and after the sample, respectively, l is the pathlength of the cell and c is the concentration. Since the molar absorptivities of different PNA monomers are not well characterized and contributions of the peptide backbone to the perturbation in the π-system of the nucleobases are not known, for all practical purposes the concentration of a PNA oligomer is determined in the following way.[43] Concentration determination is based on the absorption at 260 nm and 80 °C. At this temperature, the nucleobases are considered completely destacked and the contribution from the backbone is assumed to be negligible. The absorptivity is considered as the sum of the absorptivities of DNA monomers and effective ε_{260} per single strand is calculated by using Equation (4). A value of 3000 M^{-1} cm^{-1} can also be used[46] for the PNA oligomers containing pseudoisocytosines (J bases).

Absorption spectroscopy is also important in PNA applications to monitor the hybridization of appropriate single strands to form the corresponding duplex, both pure and hybrid, structures. In general, solutions of single strands are heated to 80 °C followed by adjustment of concentration (~2–3 mM per strand) from absorption measurements and the respective single-strand solutions are allowed to mix at that temperature. This is followed by a slow cooling step to ensure proper annealing of the strands.

The thermal melting temperature, T_m, defined as the temperature at which 50% of the complexes have been dissociated, gives an idea of the stability of the PNA–PNA or PNA–DNA duplex. In principle, the thermal melting temperature (and also entropy and enthalpy) of a PNA duplex, pure and hybrid, is estimated by monitoring the change in absorption at 260 nm at different temperatures and the corresponding plot of A_{260} versus temperature gives the thermal melting profile of a PNA duplex. Absorption melting provides information about different structural transitions. However, sometimes some information related to structural transitions is invisible in absorption melting owing to some associated factors, e.g. self-stacking and self-melting. It is often possible to overcome this problem by studying the circular dichroism (CD) melting, which can also provide supplementary information to complement the absorption melting study. The properties of PNA–PNA and PNA–DNA duplexes have been extensively investigated by employing absorption spectroscopy. The PNA oligomer H-TGTACGTCACAACTA-NH_2 forms an antiparallel duplex with complementary DNA with a T_m value of 70 °C. The corresponding DNA–DNA complex has a T_m of only 53 °C. However, the corresponding parallel duplex shows a transition temperature of 56 °C.[4] The thermal stability of PNA–RNA duplexes is higher than that for PNA–DNA duplexes.[60] The increased thermal stability of PNA–PNA duplexes relative to the corresponding DNA–DNA duplexes is fundamentally due to the absence of any significant electrostatic repulsion between the two strands in the former.

The stability of PNA–DNA hybrids is not very much affected by changes in ionic strength, except in the limit of low ionic strength where the stability increases. Owing to the higher sequence specificity of PNA upon binding to nucleic acids, incorporation of any mismatch in the duplex considerably affects the thermal melting temperature of the heteroduplex. For the heteroduplex corresponding to the PNA sequence H-egl-GGCAGTGCCTCACAA-NH_2 (C-terminal), and its full complementary DNA 5′-TTGTGAGGCACTGCC-3′, a mismatched incorporation of A in the eighth position of the DNA sequence, in place of G, reduces the T_m from 72.3 to 58.1 °C (Table 2) (Nordén and co-workers unpublished observations).

Homopyrimidine PNAs binds complementary DNA sequences to form $(PNA)_2$–DNA triplexes, and well-defined melting curves are obtained for these structures. The stability of the complexes depends on the length of

Table 2 Thermal stabilities of PNA–DNA duplexes

PNA sequence	DNA sequence	Melting temperature (°C)
H-egl-GGCAGTGCCTCACAA-NH_2[a]	5′-TTGTGAGGCACTGCC-3′[b]	72.3
	5′-TTGTGAG**A**CACTGCC-3′	58.1
	5′-TTGTGAG**A**CA**T**TGCC-3′	39.0
H-egl-GGCAGCGCCTCACAA-NH_2	5′-TTGTGAGGCGCTGCC-3′	>85
	5′-TTGTGAGGC**A**CTGCC-3′	69.9

[a] Self-melting of this PNA sequence is 44.7 °C.
[b] The thermal melting temperature of the full complementary DNA duplex (5′-TTGTGAGGCACTGCC-3′ and its complementary sequence) is 42.2 °C.

the oligomers in a regular manner with an increase in T_m of approximately 10 °C per base pair. For identical decamer sequences (antiparallel duplex), the T_m corresponding to PNA–PNA, PNA–DNA and DNA–DNA sequences are 67, 51 and 33.5 °C, respectively. However, the parallel PNA–PNA complex ($T_m = 47$ °C) is still considerably more stable than the corresponding parallel DNA–DNA duplex.[72]

4.1.2 Circular Dichroism and Linear Dichroism

CD and LD spectroscopic techniques early unveiled important structural aspects of PNA and PNA–nucleic acid hybrids.

4.1.2.1 Circular Dichroism CD is the differential absorption of left and right circularly polarized light ($CD = A_l - A_r$). Chiral molecules, i.e. molecules having nonsuperimposable mirror images, exhibit characteristic optical activity, i.e. they induce a rotation of the plane polarized light, and show an intrinsic CD. An achiral molecule has an identical mirror image and it does not show any CD. However, if an achiral molecule interacts with a chiral environment, the former may gain some optical activity through coupling of the two systems and may, in turn, show an induced CD. When a supermolecular structure includes a set of chromophores, assembled into a helical structure, the coupling between transitions may also generate a considerable amount of CD signal. The CD spectrum of a particular molecule is very sensitive to changes in the secondary structure and is useful for the detection of conformational changes of the molecule, e.g. the secondary structure of DNA and protein.

Since PNA contains an achiral backbone, it does not show any CD spectrum. However, certain modifications on the backbone structure and its conjugation with complementary DNA and RNA sequences to form hybrids render the potential for using it as a probe in CD studies (Figure 8a–c). A PNA duplex containing L-lysinylamide attached to the carboxy terminus of the PNAs demonstrates handedness resulting from the helicity of the structure. The source of CD can be attributed to the chiral orientation of the base pairs relative to each other. It has been demonstrated that the induced helicity, as measured by CD, is dependent on the nucleobase sequence proximal center.[73] CD spectroscopy of PNA–DNA and PNA–RNA duplexes provided the first information about the overall base-pair geometry in these complexes. The CD spectra of DNA–DNA and antiparallel PNA–DNA and PNA–RNA duplexes are similar, suggesting that PNA participates in the formation of right-handed helices with a base-pair geometry not much different from that found in a B- or an A-form DNA helix. In contrast, the spectra of parallel PNA–DNA and PNA–RNA duplexes deviate more from the DNA–DNA spectrum, suggesting a different kind of base stacking. CD spectroscopy was also early used to prove the formation of a right-handed PNA–DNA triplex structure.[34] Another important application of CD in this area is the study of the binding of different nucleic acid binding ligands

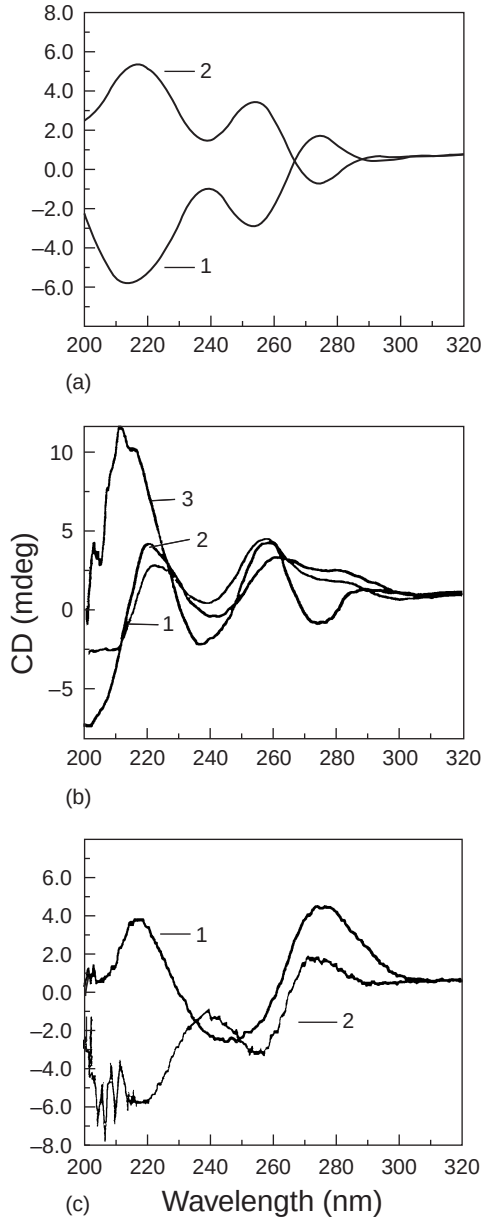

Figure 8 CD spectra of GTAGATCACT.AGTGATCTAC duplexes: (a) PNA–PNA with terminal L-lysine (curve 1) and D-lysine (curve 2) residue. (b) Antiparallel PNA–RNA (curve 1) and PNA–DNA (curve 2) duplexes, and parallel PNA–DNA (curve 3) duplex. (c) DNA–DNA (curve 1) and PNA–PNA (curve 2) duplexes. (Figure adapted by permission from Dr Pernilla Wittung.)

to PNA–DNA duplexes. It has been concluded that incorporation of a PNA strand/strands drastically reduces the binding capability of both minor groove binders and intercalators.[74]

4.1.2.2 Linear Dichroism LD is defined as the anisotropic absorption of linearly polarized light. LD depends mainly upon a vectorial quantity termed the ''transition moment'', which corresponds to one particular direction in each molecule.[75] Light polarized parallel to the transition moment has maximum probability of absorption, whereas if the polarization of the light is perpendicular to the transition moment, no absorption can take place. For solutions of macromolecules, such as DNA and other polyelectrolytes, flow fields or electric fields have been used for orienting the sample molecules. For a flow orientation experiment LD can be written as Equation (7):

$$\mathrm{LD}(\lambda) = A_{\parallel} - A_{\perp} \tag{7}$$

where $A_{\parallel}$ and $A_{\perp}$ are absorption of linearly polarized light in the parallel and perpendicular direction, respectively, with respect to the flow direction. A dimensionless quantity, often used in LD study, (LD^{r}) or reduced linear dichroism, is defined by Equation (8):

$$\mathrm{LD}^{\mathrm{r}}(\lambda) = \frac{\mathrm{LD}(\lambda)}{A_{\mathrm{iso}}(\lambda)} \tag{8}$$

where $A_{\mathrm{iso}}(\lambda)$ is obtained from the absorption spectrum of the isotropic sample, that is, in the absence of flow. LD^{r} is related to the orientation of the light-absorbing transition moments according to Equation (9):

$$\mathrm{LD}^{\mathrm{r}}(\lambda) = SO(\lambda) = \frac{3}{2} S \sum F_{\mathrm{I}}(3\cos^2\alpha_{\mathrm{I}} - 1) \tag{9}$$

where S is an orientation factor $0 \leq S \leq 1$ describing the degree of orientation of the principal axis, i.e. the helix axis in the case of DNA, relative to the flow direction, $O(\lambda)$ is an optical factor defined by the second sign of equality in the above equation, and is given by the angles α_{I} between the helix axis, as in case of DNA, and the transition moments responsible for the absorption of light, and the weight factors $F_i = \varepsilon(\lambda)_i/\quad \varepsilon(\lambda)_{\mathrm{I}}$ defining their respective contributions to the absorption at the particular wavelength λ. In principle, the π–π^* transitions of the DNA bases are responsible for the LD of polydeoxynucleotides around 260 nm and are polarized at various angles in the plane of the bases. If we consider that base planes are perfectly perpendicular to the helix axis of the DNA, the angle α_{I} would all be equal to 90° for the π–π^* transitions so that $O(\lambda) = -1.5$ and $\mathrm{LD}^{\mathrm{r}}(\lambda)$ becomes constant and independent of wavelength. In practice, the bases are generally somewhat inclined from perpendicularity and a certain wavelength dependence of the optical factor $O(\lambda)$ and of $\mathrm{LD}^{\mathrm{r}}(\lambda)$ can be anticipated.

From the measurements of LD it was shown early that PNA-T_8 forms a right-handed PNA_2–DNA triple helix with poly(dA) in which the base triplets are arranged with their planes approximately perpendicular to the helix axis, virtually similar to standard poly(dA)–$(\mathrm{poly(dT)})_2$ triplex. This PNA_2–DNA triplex structure is remarkably rigid, which indicates an efficient stacking interaction. The single-stranded DNA, owing to its high flexibility, cannot display any measurable (appreciable) LD signal under the influence of shear flow. If the PNA is added to this at a 2:1 (T:A) stoichiometry the LD signal shows a drastic increment around the 260-nm region. This indicates that the stiffness of the PNA–DNA triplex arises from base–base contacts, which prevent the bending of the chain. Also, the strong negative LD^{r} spectrum of the poly(dA)[PNA-T_8] triplex in the region 230–290 nm is more or less similar to that of the corresponding DNA triplex.[76] It was concluded from the above study that the base tilts in the two cases were similar to each other.

Interestingly, based on the LD measurements,[76] it was also concluded that no duplex stretches are present at subsaturation ratios but that segments of PNA_2–DNA triplexes are distributed evenly on the poly(dA) strand separated by flexible single-stranded regions. Only close to the saturation point, 2:1 PNA to DNA bases, does the complex become rigid owing to complete coverage by triplex and stacking between each PNA triplex segment. CD also indicated a complete absence of duplex at all mixing ratios.

4.2 Vibrational Spectroscopy

Vibrational spectroscopy is another important spectroscopic method which has been used to study biomolecules for many years. Vibrational spectroscopic studies include infrared (IR), infrared linear dichroism (IRLD) and Raman spectroscopy and, more recently, Raman optical activity (ROA). These techniques are capable of providing important structural information. So far, there has been no representative report related to vibrational spectroscopic studies regarding PNA applications. However, we believe that this particular area of spectroscopy is capable of providing new information related to peptide nucleic acid structure and function. Here, we discuss briefly how these different methods are applied to the study of nucleic acids and try to highlight how they can be made useful for PNA applications.

4.2.1 Infrared Spectroscopy

The IR spectrum of different forms of the same polynucleotides (DNA) shows large differences in vibrational

frequencies and intensities.[77,78] The whole spectrum of DNA covers the range 1800–700 cm^{-1}, which also includes vibrations involving the phosphate groups located around 1230 and 1090 cm^{-1}. However, for the bases, double-bond in-plane vibrations, in-plane ring vibrations and out-of plane ring bending vibrations are found in the regions 1800–1500, 1500–1100 and below 800 cm^{-1}, respectively. It is also known that bases contribute to the IR spectra of nucleic acids differentially and any structural transition is accompanied by a shift in the spectrum. This technique, particularly the most frequently employed Fourier transform infrared (FTIR) method, can be used for conformational studies of different PNA duplexes, pure, hybrid or chimeric. Moreover, a comparison with the corresponding DNA structure will be helpful in delineating the structure. Contributions from a different backbone will also be helpful in finding information regarding the influence of backbone structure in nucleic acid hybridization and of intermolecular interactions.

4.2.2 Infrared Linear Dichroism

This technique is capable of providing additional structural data regarding oriented samples. For nucleic acid studies, samples are generally oriented by depositing a droplet of a concentrated solution on an IR-transparent support and by stroking the drying sample unidirectionally, or by applying a wet spinning method utilizing a simultaneous precipitation and stretching of DNA fibers. In IRLD measurements, two IR spectra are recorded with the electric field of the incident light directed parallel and perpendicular to the orientation axis of the polynucleotide. The conformational transition related to the formation of PNA–PNA or PNA–DNA duplex and PNA–PNA–PNA or PNA–PNA–DNA triplex can be studied in this way.

4.2.3 Infrared Vibrational Circular Dichroism

Infrared vibrational circular dichroism (IRVCD) appears to have certain advantages over conventional CD for gaining structural information related to nucleic acids. The IRVCD spectrum can be interpreted in terms of the relative position of the carbonyl groups of the nucleic acid bases and information on base alignment can also be achieved. This technique, if applied to study different complexes of PNA, particularly those containing terminal lysinyl residues, might provide valuable information related to the structure and to study certain transitions which are otherwise invisible in electronic CD. IRVCD requires very high concentrations of sample (20–30 mg mL^{-1}), a major disadvantage of this application.

4.2.4 Raman Spectroscopy

Raman spectroscopy is also a powerful spectroscopic method suitable for use in the conformational study of nucleic acids.[79] The major advantage of this technique over conventional IR spectroscopy is the possibility of carrying out measurements in aqueous solutions. The Raman spectrum of nucleic acids is obtained within the range 600–1800 cm^{-1} and, in general, reflects the same vibrational states of various bonds associated with the conventional IR spectrum. Raman spectra can also be used to detect small (local) conformational changes associated with intermolecular interactions, particularly near the binding sites, using small model systems.

4.2.5 Resonance Raman Spectroscopy

Conformational changes in nucleic acids can also be studied using resonance Raman spectroscopy, where excitation in the wavelength range of the electronic absorption of nucleic acid bases is used. A concentration range about 20–100 times lower than in conventional Raman spectroscopy can be used for this kind of study. This method should be suitable for PNA applications, especially to gain information about conformational changes of the bases. Also, structural information related to pure, hybrid or chimeric duplexes or triplex structure of PNAs might be compared with corresponding pure deoxyriboucleotide or ribonucleotide structures to obtain an insight into the influence of the backbone on the individual conformation.

4.3 Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry

In this section we briefly discuss PNA applications of nuclear magnetic resonance (NMR) spectroscopic and mass spectrometric techniques. In mass spectrometry we mainly focus on the use of MALDI/TOFMS to characterize PNA and analyze genetic mutation.

4.3.1 Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy is probably the most powerful tool to analyze in detail the conformation of oligonucleotides in solution. ^{31}P-NMR has been widely used to elucidate phosphodiester bond conformation. In proton NMR, in most cases, the two-dimensional nuclear Overhauser effect[80] has been used to characterize the interproton distances of different parts of the nucleic acid molecule. Consequently, characterizations of different kinds of nucleic acid structures have been carried out. The majority of the work concerning the conformation of various DNA sequences has been done on synthetic oligonucleotides, as high-molecular-weight samples give broad peaks and less good resolution.

NMR spectroscopy has also been successfully applied to determine solution structures of different oligomers, pure and hybrid, of PNA. The monomer exists as both the *cis* and *trans* rotamers about the tertiary amide bond, slightly favoring the *trans* conformation[20] as revealed from the ^{1}H spectra. The two rotamers interconvert on the timescale of ~1 s at room temperature. For longer sequences, as determined from the ^{1}H-NMR spectrum of the PNA strand H-GCTATGTC-NH_2, a multitude of structural species can exist. If n is the total number of residues, then the possible number of structural species will be 2^n.

The solution structures of antiparallel PNA–DNA (H-GCTATGTC-NH_2-5′d[GACATCGC]) and PNA–RNA (GAACTC-5′r[GAGUUC]) duplexes have been determined by ^{1}H-NMR.[20,81–83] The PNA–DNA duplex shows a helical conformation with Watson–Crick base pairing scheme. The major groove is wide and deep while the minor groove is narrow and shallow. The helical rise is about 42 Å with approximately 13 base pairs per turn and the helical diameter is close to 23 Å. The carbonyl groups on the backbone base linkers of this structures are oriented along the backbone, pointing in the C-terminal direction. However, the sugar puckering is predominantly in the C2′-endo conformation with all *anti*-glycosidic torsion angles and *trans* primary amide bond conformations. The PNA–RNA duplex molecule has features similar to a right-handed A-helix. The carbonyl groups of the PNA backbone linkers are directed towards the C-terminus and primary amide bonds are also in a *trans* conformation. The nucleic acid strand possesses features very similar to an A-type structure, with C3′-endo sugar pucker and glycosidic torsion angles near -160°.

The structural properties of complexes of PNA are of great interest. However, the information available regarding the structural details has so far been rather limited.

4.3.2 Mass Spectrometry

Mass spectrometers measure the mass-to-charge ratios (m/z) of gas-phase ions. The gas phase is created by different ionization methods, e.g. electron ionization, chemical ionization, fast atom bombardment, electrospray ionization and matrix-assisted laser desorption ionization (MALDI). Mass spectrometry has become an increasingly important tool for the accurate, efficient and fast routine analysis of nucleic acids. It can instantly separate and detect DNA fragments on the basis of their mass-to-charge ratio and, thereby, can allow the determination of DNA length. The MALDI method is a comparatively new method and undoubtedly suitable for high-molecular-weight compounds such as proteins, peptides, DNA/RNA or PNA oligomers. A MALDI spectrum consists of singly charged quasimolecular ions, e.g. $[M + H]^+$, $[M + Na]^+$, $[M + K]^+$, for each sample component. The sample is mixed with a suitable MALDI matrix such as sinapinic acid in a ratio of 1 : 100. This is allowed to dry on a metal surface. A laser beam is then directed at this surface, causing desorption and ionization of sample components. The development of the MALDI method coupled with time-of-flight (TOF) mass spectrometry has provided an alternative approach to the analysis of nucleic acids.[84,85] It has also been successfully employed to study different PNA applications, particularly the characterization of PNA and genetic analysis.

It has been shown that typical matrices, e.g. sinapinic acid, useful for protein study are also suitable for use in PNA studies. PNAs survive the MALDI process intact and are easily ionized. The method holds potential for a rapid evaluation of the purity of a newly synthesized PNA oligomer and provides valuable sequence information regarding the synthesis by-products. It has also been found that dsPNA does not survive the MALDI process under certain experimental conditions.[29]

MALDI/TOFMS has also been successfully used to study discrimination of single-nucleotide polymorphisms (SNPs) in human DNA using PNA probes.[30] Human genomic and mitochondrial DNA contain a large number of SNPs that are linked to diseases. Rapid and accurate screening of important SNPs, based on high-affinity binding of PNA probes to DNA, is possible by using MALDI/TOFMS. The captured single-stranded DNA molecules are first prepared by PCR amplification and thereafter hybridized with PNA probes in an allele-specific fashion. MALDI/TOFMS is then used for the rapid and precise detection and identification of the hybridized PNA probes. This provides a very straightforward, rapid, accurate and specific detection of SNPs in amplified DNA. Another powerful feature of this approach is that the direct analysis is made in seconds.[30,86] Griffin et al.[87] reported an approach for the detection of multiple point mutations, using allele-specific, mass-labeled PNA hybridization probes following a direct analysis using MALDI/TOFMS. The resultant mass spectra depict peaks of distinct masses corresponding to each allele present. A mass spectral ''fingerprint'' of each DNA sample is obtained.

4.4 Crystallography and Its Applications

In this section we discuss crystallography and its application in determining the structure of PNA. We also include a few comments in connection with computational studies in PNA structures.

4.4.1 Structure of PNA

The crystal structure of a PNA duplex (H-CGTACG-$NH_2)_2$ obtained at 1.7 Å resolution and crystallized in space group $P_{1\ bar}$ shows both a right- and a left-handed helix in the unit cell.[88] The helical parameters estimated from this study reveal that the helix is wide (28 Å) and the pitch is large (18 bp) compared with the conventional A- or B-form. The twist, rise, base tilt and displacement values are 19.8, 3.2, 1.0 and 8.3 Å, respectively. One helical turn of PNA duplex is 58 Å long. However, the most important conclusion drawn from the results of the crystallographic study of Nielsen et al.[20,88] is that they confirm the discovery by Wittung et al.[72,73] of a PNA–PNA double helix, showing that a sugar phosphate backbone is not a prerequisite for the formation of a nucleobase double helix.

The crystal structure of the PNA_2–DNA triplex shows considerable deviations from those of nucleic acid triplexes.[89] The purine DNA strand undergoes a Watson–Crick base pairing with the homopyrimidine PNA strand in an antiparallel orientation and forms Hoogsteen hydrogen bonds to the second PNA strand of identical sequence running parallel with the DNA. The two PNA strands in the triplex structure resemble the PNA strand in the PNA–DNA duplex. The PNA_2–DNA triplex is wide (diameter $\sim$ 26 Å) and the pitch is approximately 16 (triplet). The sugar puckering in the DNA strand is C3′-endo and the backbone torsional angles are similar to those of an A-form structure. The structure is further stabilized through a network of water molecules specifically bound in the minor groove.

4.4.2 Computational Techniques

Molecular modeling and dynamics studies of nucleic acid duplex and triplex structures have complemented the structural information available from NMR and X-ray studies. They have provided potentially useful insights into the three-dimensional conformations of double and triple helical structures. Recent improvements in force field parametrizations and molecular dynamics (MD) algorithms have made a more accurate modeling of nucleic acid (and its mimic) structures possible. With the availability of NMR and X-ray data[20,81–83,88,89] regarding homo- and hetero-duplex and triplex structures, researchers have been taking interest in different computational studies of PNAs. Sen and Nilsson[90] reported molecular modeling and dynamics simulation studies on homo- and hetero-duplexes of PNA in aqueous solution under periodic boundary conditions and their results show very good agreement with the respective NMR and X-ray crystallographic data. A MD simulation study of a PNA·DNA–PNA triplex helix in aqueous solution by Shields et al.[91] gave a certain insight into the conformational flexibility of the structure. It was concluded from energetic analysis that the hybrid PNA–DNA triplex has helical characteristics which are different from those of its actual DNA counterpart. The reason is not simply the replacement of sugar–phosphate by PNA backbones. The definite conformational preferences of the PNA strands influence the driving force behind this conformational flexibility. Srinivasan and Olson[92] reported a molecular modeling study involving a PNA·DNA–PNA triplex. A total of 12 acyclic torsion angles determine the conformation of adjacent bases in PNA versus nine variables, including the sugar ring, in DNA. Moreover, it was also concluded that the replacement of the sugar ring by a linear bond sequence introduces torsional freedom not present in the nucleic acid. The 12 torsional angles are divided into two groups. The first group, the independent variables, includes torsional angles analogous to the glycosyl and ring torsions of the nucleic acids. The second or the dependent set arises in the successful closure of the PNA backbone.[92] The information available from these modeling studies, based on predicted structures and cross-examined in terms of the current spectroscopic, calorimetric and crystallographic data, explains why PNA cannot form a very B-like structure.

5 CONCLUSIONS

PNA is a very attractive lead compound owing to its considerable chemical and biological stability and its very specific interactions with DNA and RNA. The application of PNA as a genetic therapeutic agent has to await the development of efficient and safe methods for its cell penetration. However, its application as a recognition molecule has led to promising developments in many areas of chemistry, biology and biotechnology.[93,94] Continuing efforts in this area of research may yield novel base analogs[39,95–98] and provide further applications of this most exciting nucleic acid analog.

ABBREVIATIONS AND ACRONYMS

BIA	Biomolecular Interaction Analysis
CD	Circular Dichroism
CE	Capillary Electrophoresis
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator Gene
CGE	Capillary Gel Electrophoresis
FTIR	Fourier Transform Infrared

HPLC	High-performance Liquid Chromatography
IR	Infrared
IRLD	Infrared Linear Dichroism
IRVCD	Infrared Vibrational Circular Dichroism
LD	Linear Dichroism
LD^r	Reduced Linear Dichroism
LIF	Laser-induced Fluorescence
MALDI	Matrix-assisted Laser Desorption Ionization
MALDI/TOFMS	Matrix-assisted Laser Desorption Ionization/Time-of-Flight Mass Spectrometry
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
PARC	PNA-assisted Rare Cleavage
PBG	Poly-γ-benzylglutamate
PCR	Polymerase Chain Reaction
QCM	Quartz Crystal Microbalance
ROA	Raman Optical Activity
SNPs	Single-nucleotide Polymorphisms
SPR	Surface Plasmon Resonance
TOF	Time-of-flight
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VNTR	Variable Number of Tandem Repeat
5′-UTR	5′-Untranslated Region

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids • Mass Spectrometry of Nucleic Acids • Polymerase Chain Reaction and Other Amplification Systems

REFERENCES

1. P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, 'Sequence Selective Recognition of DNA by Strand Displacement with a Thymine Substituted Polyamide', *Science*, **254**(5037), 1497–1500 (1991).
2. M. Egholm, O. Buchardt, P.E. Nielsen, R.H. Berg, 'Peptide Nucleic Acids (PNA). Oligonucleotide Analogs with an Achiral Peptide Backbone', *J. Am. Chem. Soc.*, **114**(5), 1895–1897 (1992).
3. M. Egholm, P.E. Nielsen, O. Buchardt, R.H. Berg, 'Recognition of Guanine and Adenine in DNA by Thymine and Cytosine Containing Peptide Nucleic Acids', *J. Am. Chem. Soc.*, **114**(24), 9677–9678 (1992).
4. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S.M. Frier, D.A. Driver, R.H. Berg, S.K. Kim, B. Nordén, P.E. Nielsen, 'PNA Hybridizes to Complementary Oligonucleotides Obeying the Watson–Crick Hydrogen Bonding Rules', *Nature*, **365**(6446), 566–568 (1993).
5. V. Demidov, M.D. Frank-Kamenetskii, M. Egholm, O. Buchardt, P.E. Nielsen, 'Sequence Specific Double Strand DNA Cleavage by Peptide Nucleic Acid (PNA) Targeting Using Nuclease S1', *Nucleic Acids Res.*, **21**(9), 2103–2107 (1993).
6. H. Örum, P.E. Nielsen, M. Jorgensen, C. Larsson, C. Stanley, T. Koch, 'Sequence-specific Purification of Nucleic Acids by PNA-controlled Hybrid Selection', *BioTechniques*, **19**(3), 472–480 (1995).
7. D.B. Demers, E.T. Curry, M. Egholm, A.C. Sozer, 'Enhanced PCR Amplification of VNTR Locus D1S80 Using Peptide Nucleic Acid (PNA)', *Nucleic Acids Res.*, **23**(15), 3050–3055 (1995).
8. H. Perry-O'Keefe, X.-W. Yao, J.M. Coull, M. Fuchs, M. Egholm, 'Peptide Nucleic Acid Pre-gel Hybridization: an Alternative to Southern Hybridization', *Proc. Natl. Acad. Sci. USA*, **93**(25), 14 670–14 675 (1996).
9. A.G. Veselkov, Vadim V. Demidov, P.E. Nielsen, M.D. Frank-Kamenetskii, 'A New Class of Genome Rare Cutters', *Nucleic Acids Res.*, **24**(13), 2483–2487 (1996).
10. J.C. Hanvey, N.C. Peffer, J.E. Bisi, S.A. Thomson, R. Cadilla, J.A. Josey, D.J. Ricca, C.F. Hassman, M.A. Bonham, K.G. Au, S.G. Carter, D.A. Bruckenstein, A.L. Boyd, S.A. Noble, L.E. Babiss, 'Antisense and Antigene Properties of Peptide Nucleic Acids', *Science*, **258**(5087), 1481–1485 (1992).
11. P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, 'Peptide Nucleic Acids (PNA). Potential Antisense and Antigene Agents', *Anti Cancer Drug Des.*, **8**(1), 53–63 (1993).
12. L.C. Boffa, P.L. Morris, E.M. Carpeneto, M. Louissaint, V.G. Allfrey, 'Invasion of the CAG Triplet Repeats by a Complementary Peptide Nucleic Acid Inhibits Transcription of the Androgen Receptor and TATA-binding Protein Genes and Correlates with Refolding of an Active Nucleosome Containing a Unique AR Gene Sequence', *J. Biol. Chem.*, **271**(22), 13 228–13 233 (1996).
13. H. Örum, P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, C. Stanley, 'Single Base Pair Mutation Analysis by PNA Directed PCR Clamping', *Nucleic Acids Res.*, **21**(23), 5332–5336 (1993).
14. C. Carlsson, M. Jonsson, B. Nordén, M.T. Dulay, R.N. Zare, J. Noolandi, P.E. Nielsen, L.-C. Tsui, J. Zielenski, 'Screening for Genetic Mutations', *Nature*, **380**(6571), 207 (1996).
15. C. Thiede, E. Bayerdörffer, R. Blasczyk, B. Wittig, A. Neubauer, 'Simple and Sensitive Detection of Mutations in the ras Proto-oncogenes Using PNA-mediated PCR Clamping', *Nucleic Acids Res.*, **24**(5), 983–984 (1996).

16. J. Wang, E. Palecek, P.E. Nielsen, G. Rivas, X. Cai, H. Shirashi, N. Dontha, D. Luo, P.A.M. Farias, 'Peptide Nucleic Acid Probes for Sequence-specific DNA Biosensors', *J. Am. Chem. Soc.*, **118**(33), 7667–7670 (1996).
17. P.E. Nielsen, 'Structural and Biological Properties of Peptide Nucleic Acid (PNA)', *Pure Appl. Chem.*, **70**(1), 105–110 (1998).
18. P.E. Nielsen, 'Sequence-specific Recognition of Double-stranded DNA by Peptide Nucleic Acids', *Adv. DNA Sequence-specific Agents*, **3**, 267–278 (1998).
19. D.R. Corey, 'Peptide Nucleic Acids: Expanding the Scope of Nucleic Acid Recognition', *Trends Biotechnol.*, **15**(6), 224–229 (1997).
20. M. Eriksson, P.E. Nielsen, 'PNA–Nucleic Acid Complexes: Structure, Stability and Dynamics', *Q. Rev. Biophys.*, **29**(4), 369–394 (1996).
21. R.B. Merrifield, 'Solid-phase Synthesis. I. The Synthesis of a Tetrapeptide', *J. Am. Chem. Soc.*, **85**(14), 2149–2154 (1963).
22. B. Merrifield, 'Solid-phase Synthesis', *Science*, **232**(4748), 341–347 (1986).
23. L. Christensen, R. Fitzpatrick, B. Gildea, K.H. Petersen, H.F. Hansen, T. Koch, M. Egholm, O. Buchardt, P.E. Nielsen, J. Coull, R.H. Berg, 'Solid-phase Synthesis of Peptide Nucleic Acids', *J. Pept. Sci.*, **3**(1), 175–183 (1995).
24. K. Dueholm, M. Egholm, C. Behrens, L. Christensen, H.F. Hansen, T. Vulpius, K.H. Petersen, R.H. Berg, P.E. Nielsen, O. Buchardt, 'Synthesis of Peptide Nucleic–Acid Monomers Containing the 4 Natural Nucleobases – Thymine, Cytosine, Adenine, and Guanine and Their Oligomerization', *J. Org. Chem.*, **59**(19), 5767–5773 (1994).
25. S.A. Thomson, J.A. Josey, R. Cadilla, M.D. Gaul, C.F. Hassman, M.J. Luzzio, A.J. Pipe, K.L. Reed, D.J. Ricca, R.W. Wiethe, S.A. Noble, 'Fmoc Mediated Synthesis of Peptide Nucleic Acids', *Tetrahedron*, **51**(22), 6179–6194 (1995).
26. J.C. Norton, J.H. Waggenspack, E. Varnum, D.R. Corey, 'Targeting Peptide Nucleic Acid–Protein Conjugates to Structural Features within Duplex DNA, *Bioorg. Med. Chem.*, **3**(4), 437–445 (1995).
27. G. Haaima, A. Lohse, O. Buchardt, P.E. Nielsen, 'Peptide Nucleic Acids (PNAs) Containing Thymine Monomers Derived from Chiral Amino Acids: Hybridization and Solubility Properties of D-Lysine PNA', *Angew. Chem., Int. Ed. Engl.*, **35**(17), 1939–1942 (1996).
28. M. Egholm, L. Christensen, K.L. Dueholm, O. Buchardt, J. Coull, P.E. Nielsen, 'Efficient pH-independent Sequence-specific DNA-binding by Pseudoisocytosine-containing Bis-PNA', *Nucleic Acids Res.*, **23**(2), 217–222 (1995).
29. J.M. Butler, P. Jing-Baucom, M. Huang, P. Belgrader, J. Gairard, 'Peptide Nucleic Acid Characterization by MALDI-TOF Mass Spectrometry', *Anal. Chem.*, **68**(18), 3283–3287 (1996).
30. P.L. Ross, K. Lee, P. Belgrader, 'Discrimination of Single-nucleotide Polymorphisms in Human DNA Using Peptide Nucleic Acid Probes Detected by MALDI-TOF Mass Spectrometry', *Anal. Chem.*, **69**(20), 4197–4202 (1997).
31. P.E. Nielsen, M. Egholm, O. Buchardt, 'Evidence for $(PNA)_2$/DNA Triplex Structure upon Binding of PNA to dsDNA by Strand Displacement', *J. Mol. Recognit.*, **7**(3), 164–170 (1994).
32. D.Y. Cherney, B.P. Belotserkovskii, M.D. Frank-Kamenetskii, M. Egholm, O. Buchardt, R.H. Berg, P.E. Nielsen, 'DNA Unwinding Upon Strand-displacement Binding of a Thymine-substituted Polyamide to Double-stranded DNA', *Proc. Natl. Acad. Sci. USA*, **90**(5), 1667–1670 (1993).
33. N.J. Peffer, J.C. Hanvey, J.E. Bisi, S.A. Thomson, C.F. Hassman, S.A. Noble, L.E. Babiss, 'Strand-invasion of Duplex DNA by Peptide Nucleic Acid Oligomers', *Proc. Natl. Acad. Sci. USA*, **90**(22), 10 648–10 652 (1993).
34. P. Wittung, P. Nielsen, B. Nordén, 'Direct Observation of Strand Invasion by Peptide Nucleic Acid (PNA) into Double-stranded DNA', *J. Am. Chem. Soc.*, **118**(30), 7049–7054 (1996).
35. M.B. Petersen, E.P. Economou, S.A. Slaugenhaupt, A. Chakravarti, S.E. Antonarakis, 'Linkage Analysis of the Human HMG14 Gene on Chromosome 21 Using a GT Dinucleotide Repeat as Polymorphic Marker', *Genomics*, **7**(1), 136–138 (1990).
36. S.D. Pena, R. Chakraborty, 'Paternity Testing in the DNA Era', *Trends Genet.*, **10**(6), 204–209 (1994).
37. P.S. Walsh, H.A. Erlich, R. Higuchi, 'Preferential PCR Amplification of Alleles', *PCR Methods Appl.*, **1**, 241–250 (1992).
38. H.S. Misra, P.K. Pandey, M.J. Modak, R. Vinayak, V.N. Pandey, 'Polyamide Nucleic Acid–DNA Chimera Lacking the Phosphate Backbone are Novel Primers for Polymerase Reaction Catalyzed by DNA Polymerases', *Biochemistry*, **37**(7), 1917–1925 (1998).
39. M. Koppitz, P.E. Nielsen, L.E. Orgel, 'Formation of Oligonucleotide–PNA-chimeras by Template-directed Ligation', *J. Am Chem. Soc.*, **120**(19), 4563–4569 (1998).
40. P.M. Lansdorp, N.P. Verwoerd, F.M. van de Rijke, V. Dragowska, M.T. Little, R.W. Dirks, A.K. Raap, H.J. Tanke, 'Heterogeneity in Telomere Length of Human Chromosomes', *Hum. Mol. Genet.*, **5**(5), 685–691 (1996).
41. P.E. Nielsen, M. Egholm, O. Buchardt, 'Sequence-specific Transcription Arrest by Peptide Nucleic Acid Bound to the DNA Template Strand', *Gene*, **149**(1), 139–145 (1994).
42. D. Praseuth, M. Grigoriev, A.L. Guieysse, L.L. Pritchard, A. Harel-Bellan, P.E. Nielsen, C. Helene, 'Peptide Nucleic Acids Directed to the Promoter of the Alpha-chain of the Interleukin-2 Receptor', *Biochim. Biophys. Acta*, **1309**(3), 226–238 (1996).
43. S. Tomac, M. Sarkar, T. Ratilainen, P. Wittung, P.E. Nielsen, B. Nordén, A. Gräslund, 'Ionic Effects on the Stability and Conformation of Peptide Nucleic Acid Complexes', *J. Am. Chem. Soc.*, **118**(24), 5544–5552 (1996).

44. T. Bentin, P.E. Nielsen, 'Enhanced Peptide Nucleic Acid Binding to Supercoiled DNA: Possible Implications for DNA ''Breathing'' Dynamics', *Biochemistry*, **35**(27), 8863–8869 (1996).
45. B. Armitage, T. Koch, H. Frydenlund, H. Örum, G.B. Schuster, 'Peptide Nucleic Acid (PNA)/DNA Hybrid Duplexes: Intercalation by an Internally Linked Anthraquinone', *Nucleic Acids Res.*, **26**(3), 715–720 (1998).
46. H. Kuhn, V.V. Demidov, M.D. Frank-Kamenetskii, P.E. Nielsen, 'Kinetic Sequence Discrimination of Cationic Bis-PNAs Upon Targeting of Double-stranded DNA', *Nucleic Acids Res.*, **26**(2), 582–587 (1998).
47. R. Lee, N. Kaushik, M.J. Modak, R. Vinayak, V.N. Pandey, 'Polyamide Nucleic Acid Targeted to the Primer Binding Site of the HIV-1 RNA Genome Blocks In Vitro HIV-1 Reverse Transcription', *Biochemistry*, **37**(3), 900–910 (1998).
48. H.J. Larsen, P.E. Nielsen, 'Transcription-mediated Binding of Peptide Nucleic Acid (PNA) to Double-stranded DNA: Sequence-specific Suicide Transcription', *Nucleic Acids Res.*, **24**(3), 458–463 (1996).
49. N.E. Mollegaard, O. Buchardt, M. Egholm, P.E. Nielsen, 'Peptide Nucleic Acid. DNA Strand Displacement Loops as Artificial Transcription Promoters', *Proc. Natl. Acad. Sci. USA*, **91**(9), 3892–3895 (1994).
50. H. Knudsen, P.E. Nielsen, 'Antisense Properties of Duplex–Triplex Forming PNAs', *Nucleic Acids Res.*, **24**(3), 494–500 (1996).
51. L. Mologni, P. leCoutre, P.E. Nielsen, C. Gambacorti-Passerini, 'Additive Antisense Effects of Different PNAs on the In Vitro Translation of the PML/RARα Gene', *Nucleic Acids Res.*, **26**(8), 1934–1938 (1998).
52. G. Aldrian-Herrada, M.G. Desarménien, H. Orcel, L. Boissin-Agasse, J. Méry, J. Brugidou, A. Rabié, 'A Peptide Nucleic Acid (PNA) is More Rapidly Internalized in Culture Neurons when Coupled to a Retro-inverso Delivery Peptide. The Antisense Activity Depresses the Target mRNA and Protein in Magnocellular Oxytocin Neurons', *Nucleic Acids Res.*, **26**(21), 4910–4916 (1998).
53. P.G. Righetti, C. Gelfi, 'Capillary Electrophoresis of DNA', in *CRC Series in Analytical Biotechnology, Capillary Electrophoresis in Analytical Biotechnology*, ed. P.G. Righetti, CRC Press, Boca Raton, FL, 431–476, 1996.
54. H. Arakawa, M. Maeda, T. Hanai, 'Strategies for Electromigration Separations of Biologically Relevant Compounds', in *Advanced Chromatographic and Electromigration Methods in Biosciences*, eds. Z. Deyl, I. Miksik, F. Tagliaro, E. Tesarova, Elsevier Science, Amsterdam, 53–93, 1998.
55. M. Gilar, D.L. Smisek, A.S. Cohen, 'Nucleic Acids and Their Constituents', in *Advanced Chromatographic and Electromigration Methods in Biosciences*, eds. Z. Deyl, I. Miksik, F. Tagliaro, E. Tesarova, Elsevier Science, Amsterdam, 576–607, 1998.
56. D.J. Shaw, *Introduction to Colloid and Surface Chemistry*, 4th edition, Butterworth-Heinemann, Oxford, 1992.
57. D.J. Rose, 'Characterization of Antisense Binding Properties of Peptide Nucleic Acids by Capillary Gel Electrophoresis', *Anal. Chem.*, **65**(24), 3545–3549 (1993).
58. G.L. Igloi, 'Variability in the Stability of DNA–Peptide Nucleic Acid (PNA) Single-base Mismatched Duplexes: Real-time Hybridization During Affinity Electrophoresis in PNA Containing Gels', *Proc. Natl. Acad. Sci. USA*, **95**(15), 8562–8567 (1998).
59. *BIAcore X Instrument Handbook*, preliminary edition, Pharmacia Biosensor AB, Uppsala, 1996.
60. K.K. Jensen, H. Örum, P.E. Nielsen, B. Nordén, 'Kinetics for Hybridization of Peptide Nucleic Acids (PNA) with DNA and RNA Studied with the BIAcore Technique', *Biochemistry*, **36**(16), 5072–5077 (1997).
61. C. Lu, A.W. Czanderna (eds.), *Applications of Piezoelectric Quartz Crystal Microbalances*, Elsevier, Amsterdam, 1984.
62. Y. Okahata, Y. Matsunobo, K. Ijiro, M. Mukae, A. Murakami, K. Makino, 'Hybridization of Nucleic Acids Immobilized on a Quartz Crystal Microbalance', *J. Am. Chem. Soc.*, **114**(21), 8299–8300 (1992).
63. Y. Okahata, K. Niikura, Y. Sugiura, M. Sawada, T. Morii, 'Kinetic Studies of Sequence-specific Binding of GCN4-bZIP Peptides to DNA Strands Immobilized on a 27-MHz Quartz-crystal Microbalance', *Biochemistry*, **37**(16), 5666–5672 (1998).
64. K. Niikura, H. Matsuno, Y. Okahata, 'Direct Monitoring of DNA Polymerase Reactions on a Quartz-crystal Microbalance', *J. Am. Chem. Soc.*, **120**(33), 8537–8538 (1998).
65. J. Wang, P.E. Nielsen, M. Jiang, X. Cai, J.R. Fernandes, D.H. Grant, M. Ozsoz, A. Beglieter, M. Mowat, 'Mismatch-sensitive Hybridization Detection by Peptide Nucleic Acids Immobilized on a Quartz-crystal Microbalance', *Anal. Chem.*, **69**(24), 5200–5202 (1997).
66. B. Janik, *Physicochemical Characteristics of Oligonucleotides and Polynucleotides*, IFI/Plenum, New York, 1971.
67. W. Sanger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
68. G.D. Fasman, 'Nucleic Acids', in *Handbook of Biochemistry and Molecular Biology*, 3rd edition, ed. G.D. Fasman, CRC Press, Cleveland, OH, 589–590, 1975.
69. T. Brown, D.J.S. Brown, 'Modern Machine-aided Methods of Deoxyoligonucleotide Synthesis', in *Oligonucleotides and Analogs. A Practical Approach*, ed. F. Eckstein, Oxford University Press, Oxford, 20–35, 1991.
70. R.M.C. Dawson, D.C. Elliott, W.H. Elliott, K.M. Jones, 'Spectral Data and pK_a Values for Purines, Pyrimidines, Nucleosides and Nucleotides', in *Data for Biochemical Research*, 3rd edition, ed. R.M.C. Dawson, Clarendon Press, Oxford, 103–114, 1986.

71. C.S. Cantor, P.R. Schimmel, *Biophysical Chemistry*, Freeman, New York, 1980.
72. P. Wittung, P.E. Nielsen, O. Buchardt, M. Egholm, B. Nordén, 'DNA Like Double Helix Formed by Peptide Nucleic Acid', *Nature*, **368**(3461), 561–563 (1994).
73. P. Wittung, M. Eriksson, R. Lyng, P.E. Nielsen, B. Nordén, 'Induced Chirality in PNA–PNA Duplexes', *J. Am. Chem. Soc.*, **117**(41), 10 167–10 173 (1995).
74. P. Wittung, S.K. Kim, O. Buchardt, P.E. Nielsen, B. Nordén, 'Interactions of DNA Binding Ligands with PNA–DNA Hybrids', *Nucleic Acids Res.*, **22**(24), 5371–5377 (1994).
75. B. Nordén, M. Kubista, T. Kurucsev, 'Linear Dichroism Spectroscopy of Nucleic Acids', *Q. Rev. Biophys.*, **25**(1), 51–170 (1992).
76. S.K. Kim, P.E. Nielsen, M. Egholm, O. Buchardt, R.H. Berg, B. Nordén, 'Right-handed Triplex Formed between Peptide Nucleic Acid PNA-T_8 and Poly(dA) Shown by Linear and Circular Dichroism Spectroscopy', *J. Am. Chem. Soc.*, **115**(15), 6477–6481 (1993).
77. E. Taillandier, J. Liquier, J.A. Taboury, 'Infrared Spectral Studies of DNA Conformations', *Adv. Infrared Raman Spectrosc.*, **12**, 65–117 (1985).
78. E. Taillandier, J. Liquier, 'Infrared Spectroscopy of DNA', *Methods Enzymol.*, **211**, 307–335 (1992).
79. W.L. Peticolas, W.L. Kubasek, G.A. Thomas, M. Tsuboi, 'Nucleic Acids', in *Biological Applications of Raman Spectroscopy*, ed. T.G. Spiro, Wiley, New York, 81–133, Vol. 1, 1987.
80. K. Wuthrich, *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, 1984.
81. M. Leijon, A. Gräslund, P.E. Nielsen, O. Buchardt, B. Nordén, S.M. Kristensen, M. Eriksson, 'Structural Characterization of PNA–DNA Duplexes by NMR. Evidence for DNA in a B-like Conformation', *Biochemistry*, **33**(33), 9820–9825 (1994).
82. M. Eriksson, P.E. Nielsen, 'Solution Structure of a Peptide Nucleic Acid–DNA duplex', *Nat. Struct. Biol.*, **3**(5), 410–413 (1996).
83. S.C. Brown, S.A. Thomson, J.M. Veal, D.G. Davis, 'NMR Solution Structure of a Peptide Nucleic Acid Complexed with RNA', *Science*, **265**(5173), 777–780 (1994).
84. F. Kirpekar, E. Nordhoff, K.K. Larsen, K. Kristiansen, P. Roepstorff, F. Hillenkamp, 'DNA Sequence Analysis by MALDI Mass Spectrometry', *Nucleic Acids Res.*, **26**(11), 2554–2559 (1998).
85. S. Berkenkamp, F. Kirpekar, F. Hillenkamp, 'Infrared MALDI Mass Spectrometry of Large Nucleic Acids', *Science*, **281**(5374), 260–262 (1998).
86. M. Egholm, 'Spectrometry Senses More Than a Small Difference', *Nat. Biotechnol.*, **15**(13), 1346 (1997).
87. T.J. Griffin, W. Tang, L.M. Smith, 'Genetic Analysis by Peptide Nucleic Acid Affinity MALDI-TOF Mass Spectrometry', *Nat. Biotechnol.*, **15**(13), 1368–1372 (1997).
88. H. Rasmussen, J.S. Kastrup, J.N. Nielsen, J.M. Nielsen, P.E. Nielsen, 'Crystal Structure of a Peptide Nucleic Acid (PNA) Duplex at 1.7 Å Resolution', *Nat. Struct. Biol.*, **4**(2), 98–101 (1997).
89. L. Betts, J.A. Josey, J.M. Veal, S.R. Jordan, 'A Nucleic Acid Triple Helix Formed by a Peptide Nucleic Acid–DNA Complex', *Science*, **270**(5243), 1838–1841 (1995).
90. S. Sen, L. Nilsson, 'Molecular Dynamics of Duplex Systems Involving PNA: Structural and Dynamical Consequences of the Nucleic Acid Backbone', *J. Am. Chem. Soc.*, **120**(4), 619–631 (1998).
91. G.C. Shields, C.A. Laughton, M. Orozco, 'Molecular Dynamics Simulation of a PNA·DNA·PNA Triple Helix in Aqueous Solution', *J. Am. Chem. Soc.*, **120**(24), 5895–5904 (1998).
92. A.R. Srinivasan, W.K. Olson, 'Molecular Models of Nucleic Acid Triple Helixes. II. PNA and 2′–5′ Backbone Complexes', *J. Am. Chem. Soc.*, **120**(3), 492–499 (1998).
93. L. Good, P.E. Nielsen, 'Inhibition of Translation and Bacterial Growth by Peptide Nucleic Acid Targeted to Ribosomal RNA', *Proc. Natl. Acad. Sci. USA*, **95**(5), 2073–2076 (1998).
94. L. Good, P.E. Nielsen, 'Antisense Inhibition of Gene Expression in Bacteria by PNA Targeted to mRNA', *Nat. Biotechnol*, **16**(4), 355–358 (1998).
95. V.A. Efimov, M.V. Choob, A.A. Buryakova, A.L. Kalinkina, O.G. Chakhmakhcheva, 'Synthesis and Evaluation of Some Properties of Chimeric Oligomers Containing PNA and Phosphono-PNA Residues', *Nucleic Acids Res.*, **26**(2), 566–575 (1998).
96. B. Armitage, D. Ly, T. Koch, H. Frydenlund, H. Örum, G.B. Schuster, 'Hairpin-forming Peptide Nucleic Acid Oligomers', *Biochemistry*, **37**(26), 9417–9425 (1998).
97. L. Chirstensen, H.F. Hansen, T. Koch, P.E. Nielsen, 'Inhibition of PNA Triplex Formation by N^4-benzoylated Cytosine', *Nucleic Acids Res.*, **26**(11), 2735–2739 (1998).
98. E. Uhlmann, A. Peyman, G. Breipohl, D.W. Will, 'PNA: Synthetic Polyamide Nucleic Acids with Unusual Binding Properties', *Angew. Chem., Int. Ed. Engl.*, **37**(20), 2796–2823 (1998).

Polycyclic Aromatic Compounds Mapping

Dhrubajyoti Chakravarti
Eppley Institute for Research in Cancer and Allied Diseases, Omaha, USA

1 Introduction 1
2 Methods for Mapping Polycyclic Aromatic Compound – DNA Reaction Sites 3
2.1 Chemical Methods 3
2.2 Photochemical Methods 5
2.3 Enzymatic Methods 5
2.4 Analysis of Mutation Spectra 9
2.5 Mapping DNA Damage Induced In Vivo 9
Acknowledgments 11
Abbreviations and Acronyms 11
Related Articles 12
References 12

Polycyclic aromatic compounds (PACs) induce tumors by reacting with specific sequences of the initiating oncogene. The resulting DNA damage is converted into oncogenic mutation by in situ mutagenesis during repair or replication. The analysis of PAC/DNA reaction specificity is crucial for understanding the molecular biology of tumor initiation.

The sequence specificity of PAC/DNA reactions has been analyzed by various methods. These methods detect different types of DNA-damage lesions. For example, alkali-labile lesions can be mapped by a chemical procedure (piperidine treatment) and bulky stable adducts can be mapped by a photochemical method (laser-induced strand scission). Biochemical studies have identified several enzymes that act on damaged DNA and can be used for mapping the location of PAC/DNA reactions. For example, DNA lesions block the processive bypass of various enzymes (DNA polymerases, RNA polymerases and exonucleases), whereas some endonucleases (UvrABC exinuclease, S1 nuclease and apurinic/apyrimidinic (AP) endonuclease) make lesion-specific incisions. Therefore, the sites of PAC reactions with DNA are mapped by treating PAC-damaged DNA with these enzymes to generate lesion-specific DNA strand breaks that are resolved by denaturing polyacrylamide gel electrophoresis. More recently, various polymerase chain reactions (PCRs) have been developed for mapping the sequence preferences of PAC/DNA reactions in vivo. These techniques include ligation-mediated polymerase chain reactions (LMPCRs), single-strand ligation polymerase chain reactions (SSLPCRs) and terminal transferase-dependent polymerase chain reactions (TTDPCRs). In these methods, the PCR is used to amplify gene-specific DNA fragments generated by chemical or enzymatic DNA lesion processing.

1 INTRODUCTION

PACs, a group of potent environmental carcinogens found in organic combustion residues, are etiological agents of human skin and lung cancer.[1] These tumors often carry oncogenically mutated *ras* genes.[2] To facilitate the study of tumor formation by PACs, model mice suitable for carcinogenesis experiments have been developed.[3] A high number of the tumors that are induced by placing PACs on the backs of these mice are found to contain oncogenic H-*ras* mutations.[4,5]

Structure **(1)** is benzo[*a*]pyrene (BP); **(2)** is 7,12-dimethylbenz[*a*]anthracene (DMBA); and **(3)** is dibenzo-[*a*,*l*]pyrene (DB[*a*,*l*]P). All three are widely studied, potent PAC carcinogens. The tumor-forming potentials of PACs are determined by two types of experiments conducted in mouse skin: *initiation–promotion* and *repeated application*. In the initiation–promotion experiments, mice are treated once on the skin with a PAC (forming initiated cells) and then with multiple treatments with promoting agents, most commonly TPA (12-*O*-tetradecanoyl phorbol 13-acetate) to induce the initiated cells into forming tumors. This type of study determines the *tumor-initiating* potential of the PAC. In the repeated application experiments, mouse skin is repeatedly treated with a PAC. In these experiments, PACs act both as initiators and promoters. Tumors formed from this type of experiment reflect the *carcinogenic* potential of the PAC. Studies indicate that DB[*a*,*l*]P is the most potent carcinogen of all PACs tested so far.[6,7]

The reaction of PACs with biological macromolecules is necessary for cancer induction.[8] In their environmental forms, however, PACs do not react with biological macromolecules. PACs are oxidized in vivo by cytochrome P450-mediated pathways, a process that has been referred to as *metabolic activation*.[9] Two major pathways, commonly referred to as the *diol epoxide* pathway and the *radical cation* pathway, are involved in the metabolic activation.[9] A third possible pathway (quinone-forming) that may be active for some PACs (such as BP), has also been described.[10,11] Metabolic activation of the parental PAC generates electrophilic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

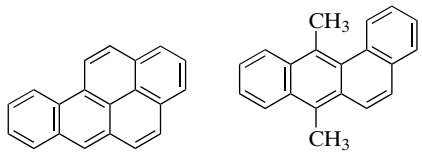

(1) (2)

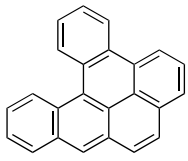

(3)

derivatives[12] that react with DNA, RNA and proteins. Studies have established that PAC/DNA reactions are important in tumor initiation.[3,9]

Various in vitro and in vivo studies have been conducted to investigate the nature of PAC/DNA reactions. These studies indicate that PACs react preferentially with purine bases. PAC reactions with certain sites in DNA bases, such as the N^2- of the guanine (Gua) or N^6- of adenine (Ade), lead to the formation of bulky adducts that are chemically stable and have therefore been described as *stable adducts*. On the other hand, when PACs react at the N3- or N7- of Ade or at the N3-, N7- or C8- of Gua, the adducts undergo spontaneous hydrolysis of the bond between the purine base and deoxyribose, and have therefore been described as *depurinating adducts*. Both PAC-diol epoxides and PAC-radical cations can form stable and depurinating adducts (for a review, see Cavalieri and Rogan[9]).

The stable adducts and abasic sites formed by the depurinating adducts are pre-mutagenic lesions and are involved in tumor initiation. Table 1 shows a partial summary of our studies that suggest that depurinating adduct formation may be a common characteristic of some potent PAC carcinogens. Specifically, these studies indicate that, for these strong carcinogens, depurinating DNA adducts are stochastically related to the oncogenic, clonal H-*ras* mutations found in mouse skin tumors.[5]

Studies indicate that stable adducts are converted into tumor-initiating H-*ras* oncogenic mutations when cells carrying adducted chromosomes enter the S phase. At this time, replication bypass of adducted bases in the template strand induces base-incorporation errors.[14,15] Research in our laboratory suggests that PAC-induced abasic sites may be converted into mutations by repair errors.[13] Figure 1 shows the proposed misreplication and misrepair models.

PAC treatment on mouse skin often causes local inflammation. This can result in the induction of indirect oxidative DNA damage, including single- and double-strand breaks, formation of abasic sites, and oxidation of DNA bases. Thus, anti-oxidants were found to reduce tumor induction without interfering with stable adduct formation.[17] Research suggests that in PAC-treated mice, oxidative DNA damage (oxidized DNA bases in particular) leads to tumor promotion.[18] Further evidence for a promoting role was obtained from observations that oxidative DNA damage shortens the latency period between adduct formation and tumor appearance.[19]

Experiments to map PAC-reactive sequences in DNA, especially in the H-*ras* gene, can help in the understanding of various early events of tumor initiation. For example, the idea that mutagenesis occurs at the damaged bases in situ would be supported if PAC-reactive hotspot sequences were found to be mutated in preneoplastic tissue. Evidence suggests that stable adducts are

Table 1 PAC carcinogenicity, adduct formation in mouse skin and clonal H-*ras* mutations in papillomas

PAC	Relative carcinogenicity[a]	Adducts in mouse skin[a]		H-*ras* mutations[b]	
		Depurinating adducts	Stable adducts	Mutation	Mutation/tumor
DB[*a*,*l*]P	+++++	N7&N3Ade (81%) N7&C8Gua (18%)	N^6dA (0.44%) N^2dG (0.52%)	Codon 61 C<u>A</u>A ⟶ C<u>T</u>A	4/5
DMBA	++++	N7Ade (79%) N7Gua (20%)	Unidentified (0.9%)	Codon 61 C<u>A</u>A ⟶ C<u>T</u>A	4/4
BP	+++	N7Ade (25%)		Codon 61 C<u>A</u>A ⟶ C<u>T</u>A	4/20
		N7&C8Gua (46%)	N^2dG (23%)	Codon 13 G<u>G</u>C ⟶ G<u>T</u>C	10/20

[a] Evaluation of carcinogenicity is based on initiation–promotion experiments conducted in mouse skin.[6,7] Carcinogenic activity: +++++, extremely active; ++++, very active; +++, active; ++, moderately active; +, weakly active; ±, very weakly active; –, inactive.

[b] Data from Chakravarti et al.[5] and from unpublished observations. The oncogenically mutated H-*ras* gene was found to be present in 14–47% of the cells in papillomas.[13]

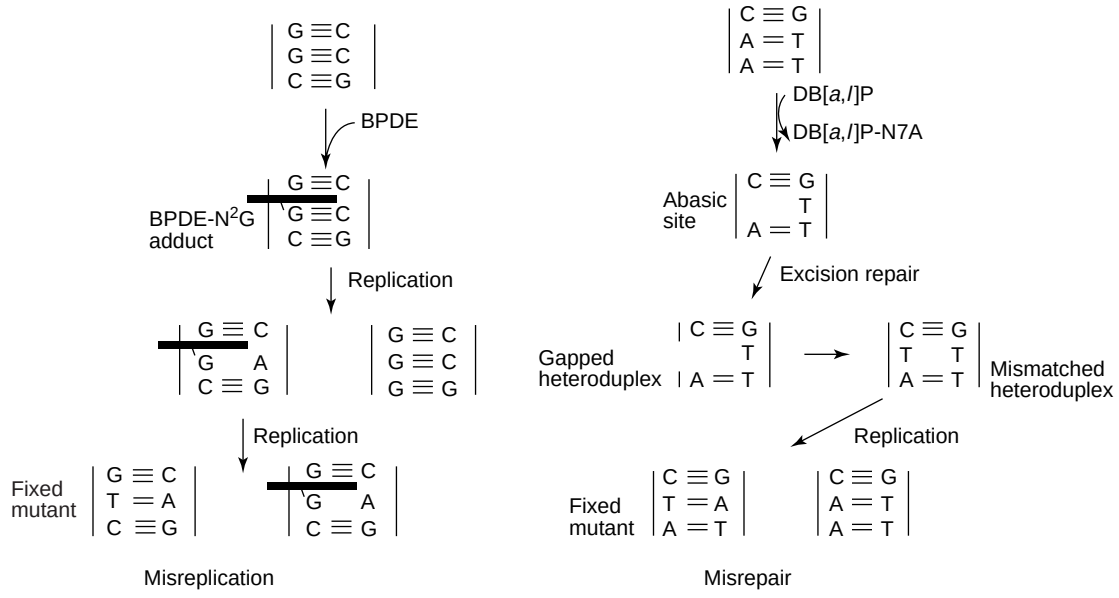

Figure 1 Proposed mechanisms for tumor-initiating mutagenesis from PAC-induced DNA damage. The misreplication model shows the formation of a GGC to GTC mutation at H-*ras* codon 13 from a benzo[*a*]pyrene-7,8-diol-9,10-epoxide (BPDE) stable adduct.[15] Two rounds of replication are necessary for generating the fixed (double-stranded) mutations. This pathway is thought to occur in replicating cells.[16] The misrepair model shows the formation of H-*ras* codon 61 CAA to CTA mutation from a DB[*a*,*l*]P depurinating adduct-induced abasic site through a single-strand mutation (a T · T mispair) intermediate.[13] The T · T heteroduplex is converted into a fixed mutation by one round of replication. This pathway may occur in resting cells. The fixed mutations are found in initiated cells.

repaired by nucleotide excision repair pathways in a sequence- and conformation-selective manner.[20,21] Mapping adduct sites at various times during DNA repair can help determine the effect of DNA sequence context and adduct conformation on the efficiency of repair. Some studies indicate that PAC-reactive hotspots (for example, benzo[*a*]pyrene-7,8-diol-9,10-epoxide (BPDE)-reactive gas chromatography (GC)-repeat sequences) frequently reside outside of oncogene-coding sequences, including the 5′-flanking sequences of H-*ras*.[22] The biological significance of such interactions is uncertain, but it is thought that the analysis of PAC interactions with non-coding sequences would be useful in antineoplastic drug design.[22] In this article we will review various approaches for mapping PAC/DNA reaction sites.

2 METHODS FOR MAPPING POLYCYCLIC AROMATIC COMPOUND – DNA REACTION SITES

2.1 Chemical Methods

The discovery that some of the BPDE-induced DNA damage is ''alkali-labile'',[23,24] led to the development of an approach to map the DNA sequence preference for the occurrence of these sites. Studies suggest that the alkali-labile sites may include at least two kinds of DNA damage. These include (a) the rapidly formed abasic sites from labile adducts and (b) the slowly hydrolyzing possible rearrangement derivative of N^2-Gua stable adducts.[24,25] Consequently, this approach has been a popular choice for PAC/DNA reaction site mapping. The alkali treatment often consists of heating PAC-reacted DNA with piperidine at 90 °C,[26–29] which results in DNA strand scission. The piperidine-induced strand scission reaction is an outgrowth of the Maxam–Gilbert sequencing methodology.[30] Investigation of N7-alkylguanines indicates that strand scission reaction occurs as the C8–N9 bond is ruptured and a formamido-pyrimidine structure is formed, which is displaced from the ribose moiety by piperidine.[31] The scission sites are electrophoretically analyzed in an appropriate denaturing polyacrylamide gel.

A majority of the mapping studies have been conducted with BPDE. In these experiments, linear DNA is radiolabeled at one end and reacted with the PAC and then treated with piperidine. These experiments indicate that PACs exhibit a specificity for three-nucleotide-long sequences. For example, Lobanenkov et al. reported that the central G of CGG, TGG, TGT and CGT is the preferred target for BPDE reaction.[27] Similar results have also been reported by other researchers, with Rill and Marsch reporting that the central G of AGG, CGG and TGN is the preferred target for

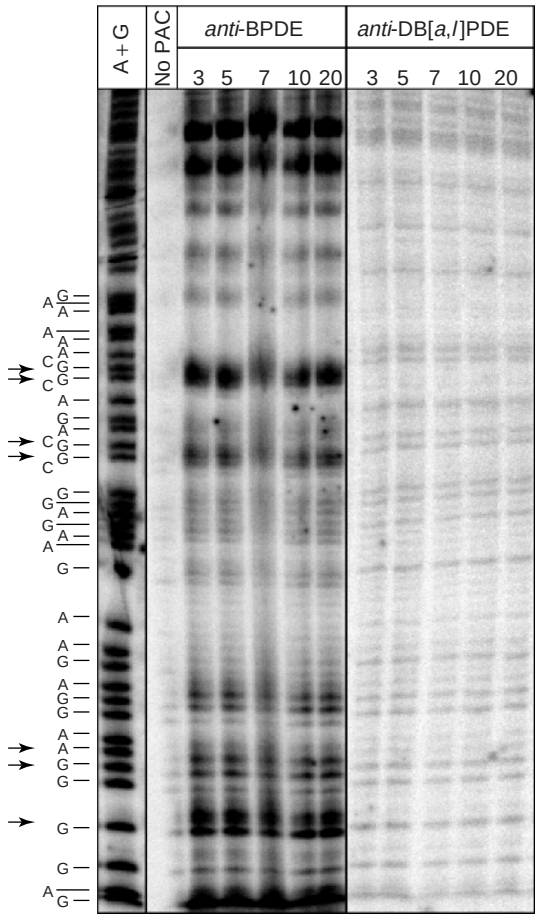

Figure 2 Piperidine mapping of "alkali-labile" sites induced by *anti*-BPDE and *anti*-DB[*a*,*l*]PDE in a segment of *p*BR322 DNA. DNA was linearized with *Nhe*I, and the 5′ end was dephosphorylated and then radiolabeled with T4 polynucleotide kinase and γ-^{32}P ATP. The resulting DNA was cleaved again with *Eco*RI and a 231 bp fragment gel purified to generate one-end-labeled linear DNA molecules. Then 25 ng of labeled DNA was reacted with 5 ng of a PAC diol epoxide in 30 µl reaction solutions containing 15% acetonitrile and 15% ethanol at room temperature in the dark for 3–20 min. Damaged DNA was extracted with chloroform/isoamyl alcohol, precipitated with ethanol and treated with 10% piperidine at 90 °C for 30 min. The fragmented DNA was resolved in an 8 M urea 6% polyacrylamide gel. A + G, Maxam–Gilbert sequencing ladder; no PAC, unadducted, piperidine-treated control.

BPDE reaction.[28] Using this technique, we have studied *anti*-BPDE- and *anti*-dibenzo[*a*,*l*]pyrene-9,10-diol-11,12-epoxide (*anti*-DB[*a*,*l*]PDE)-reactive sequences in a segment of an *Eco*RI-*Nhe*I fragment of *p*BR322 DNA[32] (Figure 2). In these experiments *anti*-BPDE was found to react with the central G of GGA sequences and both Gs of CGGC sequences. However, *anti*-DB[*a*,*l*]PDE showed comparatively weaker reactivity and no hotspots in these sequences. It is possible that *anti*-DB[*a*,*l*]PDE-induced DNA lesions are comparatively less alkali-labile.

It has been observed that BPDE lesions can be converted into strand breaks in the absence of alkali treatment, for example when DNA is incubated for extended periods at 37 °C.[26,27] Bigger et al. reported that when mapping BPDE- and benzo[*c*]phenanthrene diol epoxide-induced DNA damage, piperidine treatment could be replaced entirely by 24 h incubation at 37 °C.[29] These authors suggested that the strand breaks are formed by depurinating adduct-induced abasic site hydrolysis. However, BPDE induces a relatively small amount of depurinating adducts (1.2% N7Ade and 0.1% N7Gua) and a large amount of stable adducts (98% N^2Gua and 0.7% N^6Ade).[9] Since the BPDE-induced N7Gua depurinating adducts undergo rapid hydrolysis, the possibility that some of these strand breaks may have been induced by slow hydrolysis of stable adduct lesions cannot be ruled out. BPDE-induced strand nicking is also thought to occur from phosphate adducts.[33] We also observed that in this segment of DNA, the Gs in BPDE-reactive GGA sequences required piperidine treatment for strand breaking, whereas some of the BPDE-reactions with Gs in the CGGC sequences led to strand breaks without piperidine treatment (Figure 3). This suggests that the efficiency of DNA base loss could be determined either by the lability of particular adducts and/or by the sequence context of the adducted base.

Piperidine cleavage patterns may also help with the analysis of stable adduct-induced DNA ultrastructural changes. In these experiments, adducted DNA is first treated with a chemical probe for nucleic acid structure, such as diethyl pyrocarbonate (reacts with *syn*-purines in Z-DNA or single-stranded DNA), hydroxylamine (reacts with cytosine in single-stranded or distorted DNA),

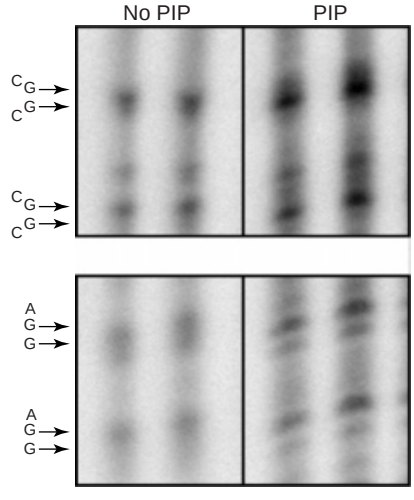

Figure 3 Liability of sequence-specific BPDE-adduct damage. Mapping experiment conducted as described in Figure 3. No PIP, strand cleavage patterns without piperidine treatment; PIP, strand cleavage with piperidine.

chloroacetaldehyde and bromoacetaldehyde (reacts with A and C in denatured regions and B–Z interfaces) and osmium tetroxide (reacts with thymine residues in distorted DNA such as B–Z junctions).[34] Next, the phosphate sugar backbone is cleaved at the level of the adducted bases with piperidine and the resulting fragments are separated in a denaturing polyacrylamide gel. Using this strategy, Belguise–Valladier and Fuchs analyzed the conformational changes induced by covalently bound N^2-acetylaminofluorene (AAF) residues to Gua bases in oligonucleotides.[34] These experiments revealed that adduct-induced structural deformations extend over 4–6 bases centered around the adduct in both strands, but the nonadducted strand is more extensively distorted.

2.2 Photochemical Methods

BPDE-N^2Gua stable adducts distort DNA secondary structure.[35,36] Irradiation of the adduct complex with laser light at 355 nm results in a specific scission at the adducted base, which results in one single-strand DNA cut at each BPDE-bound base.[37] This method has been used to map the sites of BPDE-induced stable adduct formation. In this procedure, the adducted/irradiated DNA is resolved by denaturing polyacrylamide gel electrophoresis.[38,39] As can be expected from adduct analysis (98% of BPDE adduct is BPDE-N^2Gua), mapping results indicate that BPDE primarily forms adducts at G residues, but 5–10% of the strand scissions were found at A and C residues.[38] In general, these experiments indicated that G residues with adjacent G neighbors were preferred as targets over isolated G residues. However, a recent investigation with BPDE-adducted oligonucleotides indicates that a small but significant amount of photochemical strand scission also occurs at neighboring bases, up to 4 bases upstream and 7 bases downstream from the adducted base.[40] The bases in the complementary strand were also cleaved, but at 20-fold weaker frequency than the adducted strand.[40] Therefore, it is possible that some of the strand breaks in BPDE-induced, irradiated DNA could include such secondary targets.

The analysis of BPDE reactivity to H-*ras* DNA is of special interest, because of the expected role of site-specific DNA adduct formation in inducing tumor-initiating mutations. Analysis of human H-*ras* gene exon 1 and 2 sequences indicated that the central G residues of the two oncogenic sites in exon 1 (codons 12 (GGC) and 13 (GGT)) were approximately five times more reactive to BPDE than isolated G residues were.[39] BPDE did not show any reactivity towards the central A of oncogenic codon 61 (CAG) sequence in exon 2, but reacted with the G with moderate reactivity.[39] If all tumor-initiating mutations induced by BPDE are from stable adducts, these results would predict that BPDE-induced tumors would not contain H-*ras* codon 61 mutations.

2.3 Enzymatic Methods

2.3.1 Processive Enzymes

PAC-induced DNA lesions inhibit the bypass of several types of processive enzymes. These enzymes include DNA and RNA polymerases, and exonucleases. This property has been widely used to construct precision maps of the locations of base lesions in DNA.

2.3.1.1 DNA Polymerases Studies indicate that both stable adducts and abasic sites primarily terminate the processive synthesis of DNA polymerases. Bypass synthesis over stable adduct and abasic site lesions occurs relatively infrequently, and in a subfraction of these bypassed templates DNA synthesis is erroneous, incorporating noncomplementary bases.[41–43] Studies indicate that adduct-induced DNA chain termination is not due to a failure of the 3′–5′ (proofreading) exonuclease activity of DNA polymerases to remove a misincorporated base, but due to the failure of lesion bypass.[41] DNA synthesis chain terminations occur opposite the base 3′ to the stable adducted base[44] or abasic sites[45,46] on the template strand. T7 DNA polymerase has been a popular choice for mapping PAC-reactive sites.[47,48] In this procedure, often described as *primer extension* mapping, the primer is radiolabeled at the 5′ end and DNA synthesis is conducted on adducted DNA. Experiments indicate that BPDE reacts mainly with G residues (identified as a synthesis block one base prior to G at the template strand).[48] In the rat H-*ras* gene, BPDE was found to react with moderate specificity with the oncogenic central A of rat codon 61 (CAA).[47]

If the adducted DNA can be repeatedly used as a template for DNA synthesis, finer resolution maps will be possible. To take advantage of this possibility, a *linear amplification* protocol was developed with *Taq* DNA polymerase.[49] While trying to implement this protocol to map PAC/DNA reaction sites, we encountered the problem of a large background from prematurely terminated DNA synthesis. *Taq* DNA polymerase generates large amounts of incompletely synthesized DNA strands.[50] Premature chain termination, however, has not been reported to significantly affect the primer extension mapping procedure. Such premature chain termination is unlikely to be due to differences in DNA polymerase processivity.[51] Studies suggest that spontaneous misincorporation resulting in an unextendable 3′ base,[51] as well as other factors including the presence of stable secondary structures in duplex DNA that have been noted as ''pause'' sites for DNA polymerases, could be

involved.[52] Our studies indicated that efficient proofreading reduces premature chain termination.[51]

The following considerations were helpful in developing an efficient linear amplification protocol that allowed full-length chain extension. First, studies suggest that during DNA synthesis, a 5′ endonuclease activity associated with the 5′–3′ (nick-translating) exonuclease activity of some DNA polymerases helps them to overcome stable secondary structure barriers of strand synthesis.[53,54] Therefore, conducting DNA synthesis with enzymes that contain this activity may be beneficial for full-length chain extension. Second, it has been suggested that the polymerase activity contained in a DNA polymerase outcompetes other activities associated with the enzyme during DNA synthesis.[55] We found that when DNA synthesis was carried out with a nick-translating DNA polymerase supplemented with a small amount of proofreading DNA polymerase, the fidelity of DNA synthesis was greater than when the enzymes were acting singly.[13] This suggested that such an enzyme combination would provide an effective way of enhancing proofreading activity during DNA synthesis. Accordingly, by using a 8:1 mixture of *Tth* and *Vent* DNA polymerase, we were able to obtain optimal full-length single-strand DNA synthesis.[51] We employed this strategy to construct a linear amplification map of *anti*-BPDE-, *syn*- and *anti*-DB[*a*,*l*]PDE-reactive sequences in the mouse H-*ras* exon 2 region. In these experiments, linearized target DNA is linearly amplified with a single 5′-end radiolabeled primer and the reaction products are analyzed in a denaturing polyacrylamide gel. The results indicate that the central A of codon 61 (CAA) reacts with these PACs with moderate specificity (*anti*-BPDE) and average reactivity (*syn*- and *anti*-DB[*a*,*l*]PDE) (Figure 4).[51]

2.3.1.2 RNA Polymerases RNA synthesis by DNA-dependent RNA polymerases is also terminated by DNA adduct lesions. Studies indicate that, like DNA polymerases, T7 RNA polymerase can also occasionally bypass DNA lesions and incorporate noncomplementary bases. One study suggests that when T7 RNA polymerase inserts an incorrect base across the BPDE–N^2-Gua adduct it leads to premature chain termination, and such a mechanism may contribute to the fidelity of transcription-coupled nucleotide excision repair.[56] Another study, however, suggests that T7 RNA polymerase can bypass BPDE-N^6Ade adducts with high efficiency and generate full-length runoff transcripts.[57] Eukaryotic RNA polymerase III is less able to bypass lesions than T7 RNA polymerase.[58] The presence of adducts in the template strand is expected, and they are found to be potent terminators of RNA chain growth. It was found that the presence of adducts in the displaced strand could slow down but not terminate RNA chain synthesis by T7 RNA polymerase.[59]

RNA polymerases have been used to map sites of DNA adduction. In these experiments the target DNA is positioned downstream from a suitable promoter sequence, is treated with PAC from adducts, and *runoff transcription* reactions are conducted with an appropriate RNA polymerase. T7 RNA polymerase[59–62] and SP6 RNA polymerase[48] have been commonly used in these studies.

Essentially similar results have been obtained from BPDE-induced DNA damage mapping experiments conducted either with primer extension reactions with DNA polymerase or runoff transcription reactions with RNA polymerase. For example, BPDE was found to react with G residues in GC boxes.[60] Since DNA and RNA fragments move differently in sequencing gels, it is difficult to compare the two types of mapping. Nevertheless, Thrall et al. were able to compare BPDE-reactive maps constructed by primer extension by T7 DNA polymerase and by runoff transcription by SP6 RNA polymerase.[48] From these studies they concluded that modified bases that pose relatively strong blocks to DNA polymerase do not always pose a similarly strong block for RNA polymerase. In addition, since different stable adducts were found to block T7 RNA polymerase differently,[56,57] runoff transcription experiments may yield poor resolution maps compared with DNA polymerase.

2.3.1.3 T4 Exonuclease (3′-Exonuclease Activity of T4 DNA Polymerase) T4 DNA polymerase contains a potent 3′-exonuclease activity that is optimal under nonpolymerizing conditions, such as in the absence of deoxyribonucleotides. The T4 exonuclease activity has been used to map the location of various stable lesions on DNA, including ultraviolet (UV) lesions[63,64] and BPDE-induced lesions.[48] In these studies, linear, 5′-end-labeled, adducted DNA is treated with T4 DNA polymerase in an appropriate buffer at 37 °C.[48] The reaction is terminated by heating to 65 °C for 10 min, and the resulting DNA is desalted by a spin column and resolved in a denaturing polyacrylamide gel.[48] In this procedure, however, it is difficult to decide on the exact position of the termination of T4 exonuclease, because the templates still contain the adducted base, which influences band migration in the gel. These studies indicate that T4 exonuclease can digest up to 2 or 3 nucleotides prior to the BPDE-adducted base. Unlike DNA polymerases, which progress in the 5′–3′ direction, T4 exonuclease progresses in the 3′–5′ direction. Therefore, the study of PAC reaction patterns by DNA polymerase and T4 exonuclease can provide important information about the spatial characteristics of adducts in various sequence contexts. Stable adducts are formed either in exocyclic

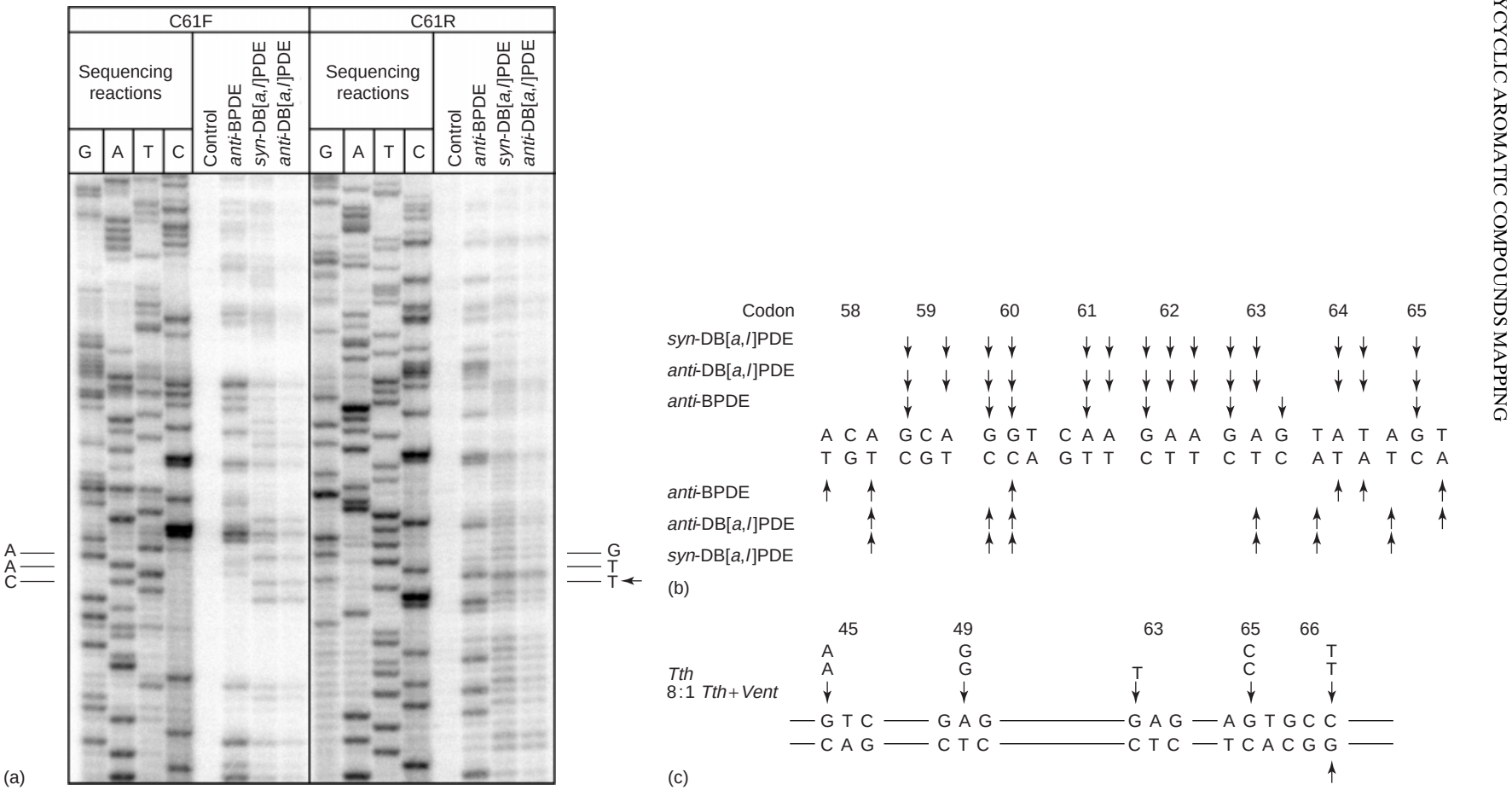

Figure 4 Linear amplification mapping of *anti*-BPDE-, *syn*- and *anti*-DB[*a*,*l*]PDE-reactive sites in mouse H-*ras* gene exon 2 codon 61 region. (a) A 2 µg sample of a plasmid-cloned H-*ras* gene (*p*WT) was reacted with 1 µg of the indicated PAC diol epoxides in the presence of 25% acetonitrile for 15 min at room temperature in the dark. The damaged DNA was extracted by phenol/chloroform and precipitated with ethanol. A 1 µg portion of the template was used in linear amplification reactions and a single 5′-^{32}P end-labeled primer C61F or C61R was used with a 8 : 1 combination of *Tth* and *Vent* DNA polymerase. In this technique, both stable adducts and abasic sites induced by depurinating adducts were mapped as DNA polymerase extension block sites, identified from the termination of extension (one nucleotide 3′ to the damaged base) of the complementary strand. The termination events were resolved in a 12% sequencing gel by comparison with the migration patterns of bands generated by enzymatic dideoxy sequencing (G, A, T and C). The gel was read in a PhosphorImager® (Molecular Dynamics). Linear amplification was conducted for both strands in this region. The arrow at T in the sequencing gel indicates DNA damage at the central A of codon 61 (CAA). (b) Map of PAC-reactive sites in H-*ras* exon 2. (c) Mutations identified in H-*ras* exon 2 (codons 45–66) region induced by *anti*-DB[*a*,*l*]PDE treatment of *p*WT DNA followed by PCR amplification with either *Tth* DNA polymerase or an 8 : 1 combination of *Tth* and *Vent* DNA polymerase, as described by Chakravarti et al.[13] *Anti*-DB[*a*,*l*]PDE-reactive bases are indicated by arrows. (Reproduced from Chakravarti et al. by permission of Mary Ann Liebert, Inc.)

or intercalated conformations,[20] which show differential repair activity.[21] Thrall et al. compared maps of BPDE-induced stable adducts by T7 DNA polymerase and T4 exonuclease and concluded that both procedures provide maps of comparable resolution.[48]

2.3.1.4 Exonuclease III Exonuclease III digests DNA processively from the 3′ end. It also contains AP endonuclease activity.[65] This enzyme stalls when it encounters adducted bases and is inhibited from further procession along the DNA strand. When modified DNA thus treated is separated in a sequencing gel, bands are observed corresponding to adduct-bonding sites. In these experiments, 5′-end-radiolabeled, adducted DNA is incubated with exonuclease III at 37 °C.[66] There are no reports of the successful use of this enzyme for mapping PAC/DNA reaction sites.

2.3.1.5 λ-Exonuclease λ-Exonuclease is a highly processive 5′–3′ exonuclease. It is double-strand specific and requires the presence of a 5′-phosphate, and the degradation product is a single-stranded DNA.[67] λ-Exonuclease is blocked by various forms of DNA damage, especially those which significantly alter the double helical structure of DNA. Using this enzyme, the sequence preference of BPDE/DNA,[68] *N*-acetoxy-*N*-acetyl-2-aminofluorene (AAAF)/DNA[69] adduct formation has been analyzed. These results indicate that λ-exonuclease is only transiently inhibited by adducts that make small structural perturbation, but is completely inhibited by lesions that induce large changes in duplex DNA, such as BPDE-induced stable adducts.[36,68] Since λ-exonuclease treatment generates truncated DNA that still contains the adducted base, the bands obtained from denaturing polyacrylamide gel are not sharp. The ladders produced by piperidine treatment and λ-exonuclease treatment on BPDE-adducted DNA, however, appeared to be similar, except that the λ-exonuclease ladder appeared to stop 6 nucleotides before the BPDE-induced stable adduct.[68]

Exonuclease III and λ-exonuclease maps of stable exocyclic N^2Gua adduct formation by two antitumor antibiotics, tomamycin and anthramycin, have been compared using a map constructed with an endonuclease (UvrABC exinuclease, see section 2.3.2.1).[66,70] The studies suggest that UvrABC exinuclease identifies more adducted sites than either exonuclease III or λ-exonuclease.

2.3.2 Endonucleases

PAC/DNA lesion-directed endonuclease cleavage provides another general approach to map the locations of these sites. For a recent review of the DNA structure-selective properties of nucleases, see Suck.[71]

2.3.2.1 UvrABC Exinuclease UvrABC exinuclease is an *Escherichia coli* excision repair enzyme involved in the recognition and the initial incision of a variety of bulky chemical-induced DNA damage reactions.[72–74] UvrABC exinuclease also removes O^6-methylguanine, thymine glycol and abasic sites.[75–77] UvrABC exinuclease has been commonly used for mapping experiments. This enzyme makes two incisions, one on each side of the DNA lesion, to initiate repair. For example, the exinuclease cuts seven nucleotides upstream and four nucleotides downstream to an AAF adduct.[78,79] Similar results were also obtained with BPDE-induced stable adducts (UvrABC exinuclease cuts 6–7 bases upstream and 4 bases downstream of these adducts).[80] In these experiments, radioactively end-labeled adducted DNA is incubated at 37 °C with a mixture of purified UvrA, UvrB and UvrC proteins in the presence of ATP and salts. Reactions are terminated by incubating the reaction mixture with proteinase K and DNA precipitated with ethanol (to remove salts). The resulting mixtures are resolved in denaturing polyacrylamide gels.[81] The DNA sequence context around AAF-adducted Gua residues was found to modulate the relative efficiency of UvrABC exinuclease activity. For example, of three targeted Gua residues in a *Nar*I restriction site sequence, one was excised at 77%, another at 45% and the third at 12% efficiency.[81] UvrABC exinuclease, however, was found to cut most BPDE-modified DNA sequences quantitatively, without regard for whether the adducts resided in linear or in supercoiled DNA.[80] Nevertheless, at some BPDE-modified sequences, UvrABC exinuclease could incise only at one side of the adduct.[80] Another possible limitation of the overall efficiency of the procedure is indicated by studies that suggest that UvrABC exinuclease excises only a fraction of the bulky stable adducts present in the templates.[82] Despite the demonstrated usefulness of UvrABC exinuclease in mapping the locations of bulky stable adducts in DNA, such observations undermine this approach.

2.3.2.2 S1 Nuclease The endonuclease activity of S1 nuclease has been tested on BPDE-modified DNA.[83,84] These experiments indicate that S1 cleaves at the local distortions induced by a stable adduct[83] in a non-sequence-specific manner,[85] but with a low efficiency (only 1 in 70 adduct sites are susceptible to S1 cleavage).[84] Studies with BPDE-damaged DNA indicate that S1 is not very active against single lesions but cleaves larger lesions efficiently.[86] S1 nuclease does not cleave discrete abasic sites, presumably because a single abasic site does not sufficiently denature the DNA secondary structure.[87] However, S1 can generate double-strand breaks by cutting through closely clustered abasic sites

straddling the two strands of DNA.[87] The single strand-specific endonuclease activity of S1 nuclease would be useful for mapping the locations of DNA damage, but an associated double-strand DNase activity makes the mapping experiments difficult to optimize.

2.3.2.3 Apurinic/Apyrimidinic Endonucleases AP endonucleases can be used to map the locations of PAC-induced abasic site formation (for a review, see Doetsch and Cunningham[88]). Two major classes of AP endonucleases are known. Class I AP endonucleases (AP lyase) catalyze β elimination reactions at AP sites, leaving 3′ residues. These enzymes also contain DNA glycosylase activity. Class II AP endonucleases, on the other hand, cleave 5′ to the AP site and do not contain an associated glycosylase activity. Mapping with AP endonuclease can provide a map of PAC adduct-induced base depurination sites. In addition, the comparison of maps generated by AP endonuclease and piperidine would be useful for the validation of PAC-induced alkali-labile sites. One study suggests that some alkali-labile sites are indeed AP sites.[65] Two commercially available AP endonucleases, exonuclease III and endonuclease IV (both class II enzymes), can be used for mapping PAC-induced abasic sites. In these experiments, 5′- or 3′-end-radiolabeled DNA is treated with PAC to form adduct lesions, then incubated with AP endonuclease and resolved in a denaturing polyacrylamide gel. It has been reported that the exonuclease activity of exonuclease III is inhibited in the presence of a citrate buffer,[89] but we did not obtain great success with this methodology. Treatment of exonuclease III either on circular DNA or on linear DNA with 4-bp 3′ end protrusions[90] (for exceptions, see Hoheisel[91]), appears to be a better strategy. On the other hand, endonuclease IV does not require any special DNA template preparation and can be used for cleaving the AP sites in a variety of templates.

2.4 Analysis of Mutation Spectra

A large body of evidence indicates that PAC-induced bulky stable adducts as well as abasic sites are premutagenic lesions. Studies indicate that these lesions primarily induce single point mutations. For example, in *uvr*$^-$ (excision repair deficient) bacteria, BPDE treatment (which induces bulky stable adducts almost exclusively, consisting of 98% N^2Gua and 0.7% N^6Ade adducts) has been found to induce equal proportions of frame-shift mutations and single base substitution mutations. Of the point mutations, 54% were G to T, 22% A to T and 9% each G to A and G to C changes.[92] In human cells, BPDE-induced DNA damage was found to induce 60% G to T and 15% each G to C and G to A mutations.[93] The mutational specificity is thought to be influenced at multiple levels, including the stereochemistry of the adducts, the bases flanking the lesion, and other host cell factors.[15]

AP sites are thought to be noninstructive. Studies indicate that as DNA polymerases bypass AP sites, purines are incorporated more readily than pyrimidines and dATP is more efficient than dGTP in sustaining chain elongation.[45] Some studies suggest that the affinity of dATP for template-growing chain-DNA polymerase complex, rather than the differences in the rate of proofreading excision, is responsible for the preferential A insertion across AP sites (A-rule).[94,95] The A-rule has been verified for both spontaneous and induced depurination. For example, the major form of spontaneous base loss in vivo involves Gua depurination.[96] Correspondingly, the most common mutation was G to T changes.[97] In addition, aflatoxin B1, which induces a major N7Gua depurinating adduct, also induces a high proportion of G to T mutations.[98] In *E. coli*, >50% of abasic site-induced mutations are A to T transversions.[43,97,99] In human cells, abasic sites were found to induce a majority (40%) of frame-shift deletions along with 30% G to C and approximately 15% each A to G, A to C and A to T mutations.[100]

The preponderance of single point mutations over multiple mutations induced by these types of PAC-induced DNA damage makes the analysis of mutational spectra a possible way of identifying PAC/DNA reaction sites. However, this approach is inefficient, and for reasons discussed previously would provide a low-resolution analysis of the PAC/DNA reaction. Nevertheless, mutational analysis can be used as complementary evidence for mapping experiments by more direct methods (Figure 4c).[51]

2.5 Mapping DNA Damage Induced In Vivo

One of the chief reasons for designing PAC/DNA lesion mapping experiments in vivo, is to understand whether the sequence context of a particular DNA damage contributes to the proficiency of its repair. This goal is conveniently accomplished by mapping DNA damage sites in target genes at different times. Three PCR-based methods, namely LMPCR (for a review, see Pfeifer et al.[101]), SSLPCR[102] and TTDPCR[103] have been developed for these purposes. Figure 5 shows flow diagrams of these methods.

2.5.1 Ligation-mediated Polymerase Chain Reaction

In this method, chromosomal DNA is extracted at various times after treating cells with DNA damaging agent. It is then digested with a restriction enzyme, followed by a mapping treatment to generate strand breaks at or near the damage site. The strand breaks create a ligatable 5′

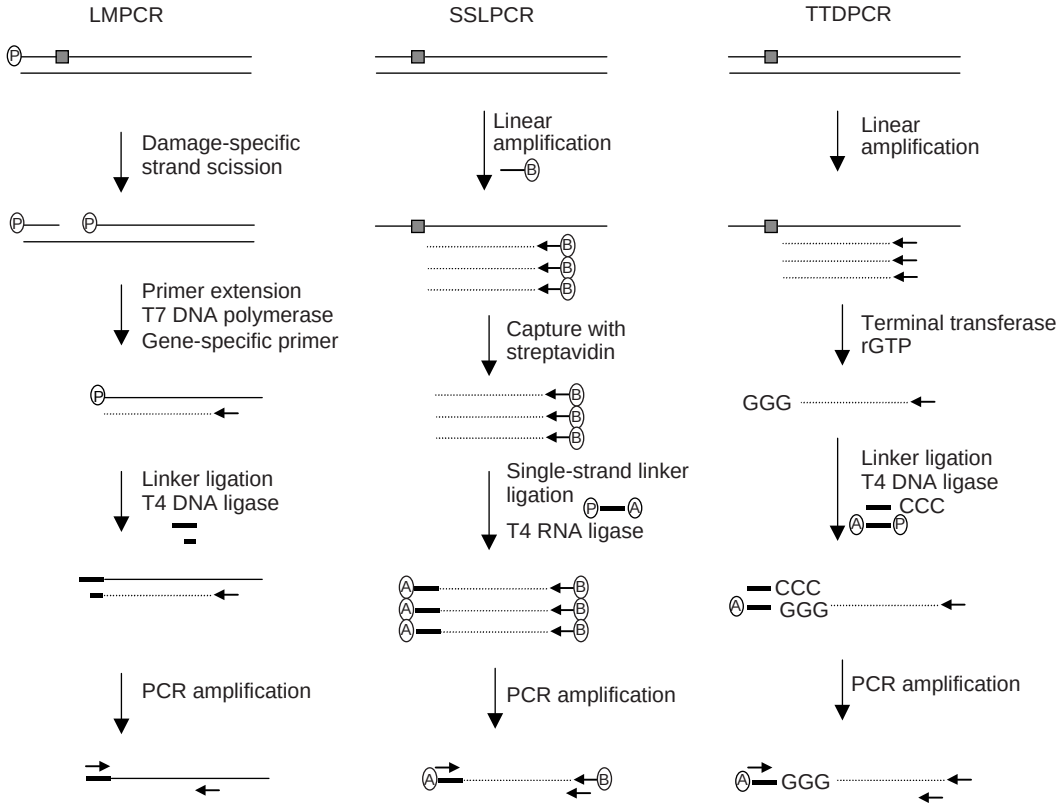

Figure 5 PCR methods for mapping PAC/DNA reactions in vivo. The PCR products are resolved in sequencing gels. When PCR is carried out with one of the primer's 5′ ends radiolabeled, the bands are detected by autoradiography. Alternatively, the electrophoresced DNA is blotted onto nylon membrane and bands visualized by hybridization with gene-specific probes. ■, DNA damage; P, phosphate, A, amine; B, biotin.

end. For example, to map UV (6–4) photoproduct lesions, which are alkali-labile, DNA is treated with piperidine[104] and, to map BPDE-induced stable adduct lesions, DNA is treated with UvrABC exinuclease.[105–107] Although the site of UvrABC exinuclease excision at the 5′ side of a BPDE adduct is variable (6–7 bases), this enzyme always cuts after 4 bases on the 3′ side.[80] The 3′ incision is relevant to LMPCR. Recently, a LMPCR method to map aflatoxin B1-induced DNA depurination has been reported, in which the strand breaks are induced by treating damaged DNA with piperidine.[108] Alternatively, abasic site-specific DNA strand breaks can be induced by treating DNA with an AP endonuclease. Next, a gene-specific primer is annealed and extended with T7 DNA polymerase (Sequenase). This creates a pool of variously sized, double-stranded products that are blunt-ended at the 3′ end.

An unphosphorylated, asymmetric double-stranded linker is then ligated to the blunt end of the extended products and a PCR amplification reaction is conducted with a linker-specific primer and a nested gene-specific primer. The PCR products are resolved in a denaturing polyacrylamide gel. The LMPCR amplified break positions are detected either by radio-densitometry in a PhosphorImager® (Molecular Dynamics, CA, USA)[105] or by electroblotting the gel onto nylon membranes and hybridizing with a gene-specific probe.[104] Comparison of BPDE/DNA reactive sites in vitro with corresponding LMPCR maps suggests that, in general, BPDE reacts with DNA with similar sequence specificity both in vivo and in vitro.[105]

LMPCRs can detect a relatively small population of damaged DNA. Studies suggest that a population of 100 molecules of template DNA[109] and a lesion population more than 1 in 10 000 nucleotides, is adequate for such analysis.[110] It has been reported that the LMPCR technique could be made more sensitive by electrophoretically enriching restriction enzyme-digested chromosomal DNA to the size of the target gene.[111] In a variation of the basic LMPCR protocol, biotinylated primers have been used for the initial gene-specific extension, and subsequently biotinylated, linker-ligated molecules have been purified with streptavidin.[105] However, it is not known whether such changes help the LMPCR protocol. Nevertheless, several steps of LMPCRs could contribute to inefficient band detection.

For example, the 5′ end generated by the strand scission reaction must be phosphorylated and also the primer extension reactions must be full-length in order for blunt-end linker ligation reaction to occur. In addition, the efficiency of blunt-end ligation may determine the effectiveness of the protocol. Although the ligation kinetics of the variously-sized double-strand fragments could possibly vary, it is unlikely that individual templates will be selectively lost in order to become under-represented in steps up to linker ligation. The final PCR amplification can only compensate for the loss of uniform signal strength.

2.5.2 Single-strand Ligation Polymerase Chain Reaction

SSLPCR[102] differs from LMPCR by not requiring DNA damage-specific strand cleavage. Instead, the method depends on the DNA polymerase-blocking properties of various DNA lesions. A second difference in this protocol is the use of single-strand ligation reaction with T4 RNA ligase. In this method, chromosomal DNA is isolated from damaged cells and subjected to linear amplification with a biotinylated primer specific for a target gene. Gene-specific linear amplification enhances the overall resolution of the procedure. This results in the formation of a population of single-stranded DNA molecules that are truncated by damage-induced termination of DNA synthesis. The biotinylated single-strand DNA is captured with streptavidin and separated from free genomic DNA by washing. Using T4 RNA ligase, these molecules are ligated to a single-strand linker DNA (containing a 5′ phosphate for ligation and a 3′ amine to prevent self-ligation). Using a two-step PCR with an oligonucleotide complementary to the linker sequence (to act as the forward primer) and a 5′-end-radiolabeled second primer (nested to the linear amplification primer), one-strand radiolabeled PCR products are generated. These products are denatured by heating to 95 °C in a formamide-containing sequencing gel loading buffer and resolved in a denaturing polyacrylamide gel. Although SSLPCRs are likely to be very useful, this method has yet to be used for mapping PAC-reactive sites in DNA.

The use of T4 RNA ligase may limit the sensitivity of SSLPCRs. DNA is not a preferred substrate for T4 RNA ligase and ligation proceeds in a slow and inefficient manner, using low amounts of single-stranded DNA.[112] It has been reported that SSLPCRs gave weaker intensity bands than LMPCRs.[103]

2.5.3 Terminal Transferase-dependent Polymerase Chain Reaction

TTDPCRs were designed to overcome the perceived deficiencies of LMPCRs.[103] Like SSLPCRs, TTDPCRs also use linear amplification with a gene-specific primer, relying on the polymerase-blocking property of the DNA lesion for synthesizing truncated single-strand molecules. Next the single-stranded DNA molecules are ribo-tailed with terminal deoxynucleotidyl transferase (TDT) using rGTP.[113] The use of ribonucleotide in this reaction limits the tailing to a few residues (an average of three nucleotides).[103] Next, a double-stranded linker containing a 3′ overhang of three complementary cytosines is ligated with T4 DNA ligase. The complementary oligonucleotide of the linker contains a 5′ phosphoryl and 3′ (aminopentyl) blocking group. This ligates the recessed 5′ end of the linker with the 3′ rG tail of the linear amplified molecules. The cohesive-end ligation in TTDPCRs may be more efficient than the blunt-end ligation of LMPCRs. Next, the linker-ligated templates are PCR amplified with a nested gene-specific primer and the top oligonucleotide of the linker. The products are resolved in a denaturing polyacrylamide gel, electroblotted and hybridized with an internal gene-specific probe as in LMPCRs. UV-induced DNA damage has been mapped by TTDPCRs and LMPCRs in parallel.[103] TTDPCRs were reported to yield approximately 4–7 times stronger bands than LMPCRs.[103]

LMPCRs are the older method and have found wide acceptance. In comparison, SSLPCRs or TTDPCRs are recently described methods. Both SSLPCRs and TTDPCRs are likely to be useful procedures for mapping the sequence preferences of various types of carcinogen-induced DNA damage.

ACKNOWLEDGMENTS

The author acknowledges the support of Drs Eleanor G. Rogan and Ercole L. Cavalieri of the Eppley Institute during the course of this work. The work was supported by US Public Health Service Grant P01 CA49210 and NRSA fellowship CA68761 from the National Cancer Institute. Core support was provided by National Cancer Institute Grant P30 CA36727.

ABBREVIATIONS AND ACRONYMS

AAAF	*N*-Acetoxy-*N*-acetyl-2-aminofluorene
AAF	N^2-Acetylaminofluorene
Ade	Adenine
AP	Apurinic/Apyrimidinic
BP	Benzo[*a*]pyrene
BPDE	Benzo[*a*]pyrene-7,8-diol-9,10-epoxide
DB[*a*,*l*]P	Dibenzo[*a*,*l*]pyrene
DB[*a*,*l*]PDE	Dibenzo[*a*,*l*]pyrene-9,10-diol-11,12-epoxide

DMBA	7,12-Dimethylbenz[*a*]anthracene
GC	Gas Chromatography
Gua	Guanine
LMPCR	Ligation-mediated Polymerase Chain Reaction
PAC	Polycyclic Aromatic Compound
PCR	Polymerase Chain Reaction
SSLPCR	Single-strand Ligation Polymerase Chain Reaction
TDT	Terminal Deoxynucleotidyl Transferase
TPA	12-*O*-tetradecanoyl phorbol 13-acetate
TTDPCR	Terminal Transferase-dependent Polymerase Chain Reaction
UV	Ultraviolet

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Optical Mapping in Genomic Analysis • Polymerase Chain Reaction and Other Amplification Systems • Sequencing Strategies and Tactics in DNA and RNA Analysis

REFERENCES

1. ACS Staff, 'Cancer Facts and Figures–1997', American Cancer Society, Atlanta, 97-300M-No. 5008.97, 1997.
2. J.G. van der Schroeff, L.M. Evers, A.J. Boot, J.L. Bos, 'Ras Oncogene Mutations in Basal Cell Carcinomas and Squamous Cell Carcinomas of Human Skin', *J. Invest. Dermatol.*, **94**, 423–425 (1990).
3. C.J. Kemp, P.A. Burns, K. Brown, H. Nagase, A. Balmain, 'Transgenic Approaches to the Analysis of Ras and p53 Function in Multistage Carcinogenesis', Cold Spring Harbor Symposia on Quantitative Biology, Cold Spring Harbor Laboratory Press, New York, 1994.
4. A.M. Colapietro, A.L. Goodell, R.C. Smart, 'Characterization of Benzo[*a*]pyrene-initiated Mouse Skin Papillomas for Ha-*ras* Mutations and Protein Kinase C Levels', *Carcinogenesis*, **14**, 2289–2295 (1993).
5. D. Chakravarti, J.C. Pelling, E.L. Cavalieri, E.G. Rogan, 'Relating Aromatic Hydrocarbon-induced DNA Adducts and *c*-H-*ras* Mutations in Mouse Skin Papillomas: The Role of Apurinic Sites', *Proc. Natl. Acad. Sci. USA*, **92**, 10 422–10 426 (1995).
6. S. Higginbotham, N.V. RamaKrishna, S.L. Johansson, E.G. Rogan, E.L. Cavalieri, 'Tumor-initiating Activity and Carcinogenicity of Dibenzo[*a*,*l*]pyrene Versus 7,12-Dimethylbenz[*a*]anthracene and Benzo[*a*]pyrene at Low Doses in Mouse Skin', *Carcinogenesis*, **14**, 875–878 (1993).
7. E.L. Cavalieri, S. Higginbotham, E.G. Rogan, 'Dibenzo-[*a*,*l*]pyrene: The Most Potent Carcinogenic Aromatic Hydrocarbon', *Polycyclic Aromatic Compounds*, **6**, 177–183 (1994).
8. E.C. Miller, J.A. Miller, 'Searches for Ultimate Chemical Carcinogens and their Reactions with Cellular Macromolecules', *Cancer*, **47**, 2327–2345 (1981).
9. E.L. Cavalieri, E.G. Rogan, 'Mechanisms of Tumor Initiation by Polycyclic Aromatic Hydrocarbons in Mammals', in *The Handbook of Environmental Chemistry*, 3J, ed. A.H. Nelson, Springer-Verlag, Heidelberg, 81–117, 1998.
10. L. Flowers, S.T. Ohnishi, T.M. Penning, 'DNA Strand Scission by Polycyclic Aromatic Hydrocarbon *o*-Quinones: Role of Reactive Oxygen Species, Cu(II)/Cu(I) Redox Cycling, and *o*-Semiquinone Anion Radicals', *Biochemistry*, **36**, 8640–8648 (1997).
11. T.M. Penning, M.E. Burczynski, C.F. Hung, K.D. McCoull, N.T. Palackal, L.S. Tsuruda, 'Dihydrodiol Dehydrogenases and Polycyclic Aromatic Hydrocarbon Activation: Generation of Reactive and Redox Active *o*-Quinones', *Chem. Res. Toxicol.*, **12**, 1–18 (1999).
12. E.L. Cavalieri, E.G. Rogan, 'Radical Cations in Aromatic Hydrocarbon Carcinogenesis', *Free Radic. Res. Commun.*, **11**, 77–87 (1990).
13. D. Chakravarti, P. Mailander, J. Franzen, S. Higginbotham, E.L. Cavalieri, E.G. Rogan, 'Detection of Dibenzo[*a*,*l*]pyrene-induced H-*ras* Codon 61 Mutant Genes in Preneoplastic SENCAR Mouse Skin Using a New PCR-RFLP Method', *Oncogene*, **16**, 3203–3210 (1998).
14. V.M. Maher, J.J. McCormick, 'Role of DNA Lesions and Repair in the Transformation of Human Cells', in *Mechanisms of Cellular Transformation by Carcinogenic Agents*, eds. D. Grunberger, S.P. Groff, Pergamon Press, New York, 135–149, 1987.
15. M. Moriya, S. Spiegel, A. Fernandes, S. Amin, T. Liu, N. Geacintov, A.P. Grollman, 'Fidelity of Translesional Synthesis Past Benzo[*a*]pyrene Diol Epoxide–2′-Deoxyguanosine DNA Adducts: Marked Effects of Host Cell, Sequence Context, and Chirality', *Biochemistry*, **35**, 16 646–16 651 (1996).
16. W.K. Kaufmann, 'Cell Cycle Checkpoints and DNA Repair Preserve the Stability of the Human Genome', *Cancer Metastasis Rev.*, **14**, 31–41 (1995).
17. A. Dipple, M.A. Pigott, C.A. Bigger, D.M. Blake, '7,12-Dimethylbenz[*a*]anthracene–DNA Binding in Mouse Skin: Response of Different Mouse Strains and Effects of Various Modifiers of Carcinogenesis', *Carcinogenesis*, **5**, 1087–1090 (1984).
18. K. Frenkel, L. Wei, H. Wei, J. Karkoszka, 'Polycyclic Aromatic Hydrocarbons (PAH) Induce Oxidative Stress and Oxidative DNA Modification–Characteristics of Tumor Promotion', *Polycyclic Aromatic Compounds*, **6**, 151–160 (1994).
19. W.H. Fischer, W.K. Lutz, 'Correlation of Individual Papilloma Latency Time with DNA Adducts,

8-Hydroxy-2′-deoxyguanosine, and the Rate of DNA Synthesis in the Epidermis of Mice Treated with 7,12-Dimethylbenz[*a*]anthracene', *Proc. Natl. Acad. Sci. USA*, **92**, 5900–5904 (1995).
20. R. Jankowiak, F. Ariese, A. Hewer, A. Luch, D. Zamzow, N.C. Hughes, D. Phillips, A. Seidel, K.L. Platt, F. Oesch, G.J. Small, 'Structure, Conformations, and Repair of DNA Adducts from Dibenzo[*a*,*l*]pyrene: 32P-Postlabeling and Fluorescence Studies', *Chem. Res. Toxicol.*, **11**, 674–685 (1998).
21. H. Naegeli, T. Buterin, M. Hess, N. Luneva, S. Amin, N.E. Geacintov, 'Variable Repair Efficiencies of B[*a*]-PDE-N2-dG and B[*a*]PDE-N6-dA Adducts by Human Nucleotide Excision Repair Enzymes', 90th Annual Meeting of the American Association for Cancer Research, Philadelphia, American Association for Cancer Research, Philadelphia, 1999.
22. W.B. Mattes, J.A. Hartley, K.W. Kohn, D.W. Matheson, 'GC-rich Regions in Genomes as Targets for DNA Alkylation', *Carcinogenesis*, **9**, 2065–2072 (1988).
23. N.R. Drinkwater, E.C. Miller, J.A. Miller, 'Estimation of Apurinic/Apyrimidinic Sites and Phosphotriesters in Deoxyribonucleic Acid Treated with Electrophilic Carcinogens and Mutagens', *Biochemistry*, **19**, 5087–5092 (1980).
24. H.B. Gamper, J.C. Bartholomew, M. Calvin, 'Mechanism of Benzo[*a*]pyrene Diol Epoxide Induced Deoxyribonucleic Acid Strand Scission', *Biochemistry*, **19**, 3948–3956 (1980).
25. M.F. Moran, K. Ebisuzaki, 'In Vivo Benzo[*a*]pyrene Diol Epoxide-induced Alkali-labile Sites are not Apurinic Sites', *Mutat. Res.*, **262**, 79–84 (1991).
26. E. Sage, W.A. Haseltine, 'High Ratio of Alkali-sensitive Lesions to Total DNA Modification Induced by Benzo[*a*]pyrene Diol Epoxide', *J. Biol. Chem.*, **259**, 11 098–11 102 (1984).
27. V.V. Lobanenkov, M. Plumb, G.H. Goodwin, P.L. Grover, 'The Effect of Neighbouring Bases on G-Specific DNA Cleavage Mediated by Treatment with the anti-Diol Epoxide of Benzo[*a*]pyrene In Vitro', *Carcinogenesis*, **7**, 1689–1695 (1986).
28. R.L. Rill, G.A. Marsch, 'Sequence Preferences of Covalent DNA Binding by anti-(+)- and anti-(−)-Benzo[*a*]pyrene Diol Epoxides', *Biochemistry*, **29**, 6050–6058 (1990).
29. C.A. Bigger, A. Cheh, F. Latif, R. Fishel, K.A. Canella, G.A. Stafford, H. Yagi, D.M. Jerina, A. Dipple, 'DNA Strand Breaks Induced by Configurationally Isomeric Hydrocarbon Diol Epoxides', *Drug Metab. Rev.*, **26**, 287–299 (1994).
30. A.M. Maxam, W. Gilbert, 'Sequencing End-labeled DNA with Base-specific Chemical Cleavages', *Methods Enzymol.*, **65**, 499–560 (1980).
31. W.B. Mattes, J.A. Hartley, K.W. Kohn, 'Mechanism of DNA Strand Breakage by Piperidine at Sites of N7-Alkylguanines', *Biochim. Biophys. Acta.*, **868**, 71–76 (1986).
32. D. Chakravarti, E.G. Rogan, E.L. Cavalieri, 'Mapping DNA Sequence-specific Aromatic Hydrocarbon Reaction Sites in a 231 bp EcoRI-NheI Fragment of pBR322', *Env. Mol. Mutagen.*, **25**, 7 (1995).
33. M. Koreeda, P.D. Moore, H. Yagi, H.J. Yeh, D.M. Jerina, 'Alkylation of Polyguanylic Acid at the 2-Amino Group and Phosphate by the Potent Mutagen (±)-7-beta,8-alpha-Dihydroxy-9-beta,10-beta-epoxy-7,8,9,10-Tetrahydrobenzo[*a*]pyrene', *J. Am. Chem. Soc.*, **98**, 6720–6722 (1976).
34. P. Belguise-Valladier, R.P. Fuchs, 'Strong Sequence-dependent Polymorphism in Adduct-induced DNA Structure: Analysis of Single N2-Acetylaminofluorene Residues Bound Within the *Nar*I Mutation Hot Spot', *Biochemistry*, **30**, 10 091–10 100 (1991).
35. T. Prusik, N.E. Geacintov, C. Tobiasz, V. Ivanovic, I.B. Weinstein, 'Fluorescence Study of the Physico-chemical Properties of a Benzo[*a*]pyrene 7,8-dihydrodiol 9,10-oxide Derivative Bound Covalently to DNA', *Photochem. Photobiol.*, **29**, 223–232 (1979).
36. M.E. Hogan, N. Dattagupta, J.P. Whitlock, Jr, 'Carcinogen-induced Alteration of DNA Structure', *J. Biol. Chem.*, **256**, 4504–4513 (1981).
37. T.C. Boles, M.E. Hogan, 'Site-specific Carcinogen Binding to DNA', *Proc. Natl. Acad. Sci. USA*, **81**, 5623–5627 (1984).
38. T.C. Boles, M.E. Hogan, 'High-resolution Mapping of Carcinogen Binding Sites on DNA', *Biochemistry*, **25**, 3039–3043 (1986).
39. K.A. Dittrich, T.R. Krugh, 'Mapping of (±)-anti-Benzo-[*a*]pyrene Diol Epoxide Adducts to Human *c*-Ha-*ras*1 Protooncogene', *Chem. Res. Toxicol.*, **4**, 277–281 (1991).
40. B. Li, B. Mao, T.M. Liu, J. Xu, A. Dourandin, S. Amin, N.E. Geacintov, 'Laser Pulse-induced Photochemical Strand Cleavage of Site-specifically and Covalently Modified (+)-anti-Benzo[*a*]pyrene Diol Epoxide–Oligonucleotide Adducts', *Chem. Res. Toxicol.*, **8**, 396–402 (1995).
41. P. Belguise-Valladier, H. Maki, M. Sekiguchi, R.P. Fuchs, 'Effect of Single DNA Lesions on In Vitro Replication with DNA Polymerase III Holoenzyme. Comparison with Other Polymerases', *J. Mol. Biol.*, **236**, 151–164 (1994).
42. D.C. Thomas, X. Veaute, R.P. Fuchs, T.A. Kunkel, 'Frequency and Fidelity of Translesion Synthesis of Site-specific N2-Acetylaminofluorene Adducts During DNA Replication in a Human Cell Extract', *J. Biol. Chem.*, **270**, 21 226–21 233 (1995).
43. C.W. Lawrence, A. Borden, S.K. Banerjee, J.E. LeClerc, 'Mutation Frequency and Spectrum Resulting from a Single Abasic Site in a Single-stranded Vector', *Nucleic Acids Res.*, **18**, 2153–2157 (1990).

44. A.M. Hruszkewycz, K.A. Canella, K. Peltonen, L. Kotrappa, A. Dipple, 'DNA Polymerase Action on Benzo[*a*]-pyrene–DNA Adducts', *Carcinogenesis*, **13**, 2347–2352 (1992).
45. D. Sagher, B. Strauss, 'Insertion of Nucleotides Opposite Apurinic/Apyrimidinic Sites in Deoxyribonucleic Acid During In Vitro Synthesis: Uniqueness of Adenine Nucleotides', *Biochemistry*, **22**, 4518–4526 (1983).
46. D. Hevroni, Z. Livneh, 'Bypass and Termination at Apurinic Sites During Replication of Single-stranded DNA *In Vitro*: A Model for Apurinic Site Mutagenesis', *Proc. Natl. Acad. Sci. USA*, **85**, 5046–5050 (1988).
47. D.B. Reardon, C.A. Bigger, J. Strandberg, H. Yagi, D.M. Jerina, A. Dipple, 'Sequence Selectivity in the Reaction of Optically Active Hydrocarbon Dihydrodiol Epoxides with Rat H-*ras* DNA', *Chem. Res. Toxicol.*, **2**, 12–14 (1989).
48. B.D. Thrall, D.B. Mann, M.J. Smerdon, D.L. Springer, 'DNA Polymerase, RNA Polymerase and Exonuclease Activities on a DNA Sequence Modified by Benzo[*a*]pyrene Diolepoxide', *Carcinogenesis*, **13**, 1529–1534 (1992).
49. M. Ponti, S.M. Forrow, R.L. Souhami, M. D'Incalci, J.A. Hartley, 'Measurement of the Sequence Specificity of Covalent DNA Modification by Antineoplastic Agents Using Taq DNA Polymerase', *Nucleic Acids Res.*, **19**, 2929–2933 (1991).
50. D.B. Olsen, F. Eckstein, 'Incomplete Primer Extension During In Vitro DNA Amplification Catalyzed by Taq Polymerase; Exploitation for DNA Sequencing', *Nucleic Acids Res.*, **17**, 9613–9620 (1989).
51. D. Chakravarti, E.L. Cavalieri, E.G. Rogan, 'Linear Amplification Mapping of Polycyclic Aromatic Hydrocarbon-reactive Sequences in H-*ras* Gene', *DNA Cell Biol.*, **17**, 529–539 (1998).
52. J.S. Hoffmann, M. Fry, J. Ji, K.J. Williams, L.A. Loeb, 'Codons 12 and 13 of H-ras Protooncogene Interrupt the Progression of DNA Synthesis Catalyzed by DNA Polymerase Alpha', *Cancer Res.*, **53**, 2895–2900 (1993).
53. V. Lyamichev, M.A. Brow, J.E. Dahlberg, 'Structure-specific Endonucleolytic Cleavage of Nucleic Acids by Eubacterial DNA Polymerases', *Science*, **260**, 778–783 (1993).
54. Y. Cao, T. Kogoma, 'Requirement for the Polymerization and $5' \rightarrow 3'$ Exonuclease Activities of DNA Polymerase I in Initiation of DNA Replication at *ori*K Sites in the Absence of RecA in *Escherichia coli* rnhA Mutants', *J. Bacteriol.*, **175**, 7254–7259 (1993).
55. S. Creighton, M.F. Goodman, 'Gel Kinetic Analysis of DNA Polymerase Fidelity in the Presence of Proofreading Using Bacteriophage T4 DNA Polymerase', *J. Biol. Chem.*, **270**, 4759–4774 (1995).
56. D.J. Choi, R.B. Roth, T. Liu, N.E. Geacintov, D.A. Scicchitano, 'Incorrect Base Insertion and Prematurely Terminated Transcripts During T7 RNA Polymerase Transcription Elongation Past Benzo[*a*]pyrenediol Epoxide-modified DNA', *J. Mol. Biol.*, **264**, 213–219 (1996).
57. K.M. Remington, S.E. Bennett, C.M. Harris, T.M. Harris, K. Bebenek, 'Highly Mutagenic Bypass Synthesis by T7 RNA Polymerase of Site-specific Benzo[*a*]pyrene Diol Epoxide-adducted Template DNA', *J. Biol. Chem.*, **273**, 13 170–13 176 (1998).
58. Y.H. Chen, Y. Matsumoto, S. Shibutani, D.F. Bogenhagen, 'Acetylaminofluorene and Aminofluorene Adducts Inhibit In Vitro Transcription of a Xenopus 5S RNA Gene Only When Located on the Coding Strand', *Proc. Natl. Acad. Sci. USA*, **88**, 9583–9587 (1991).
59. S.T. Nath, L.J. Romano, L.J. Romana, 'Transcription by T7 RNA Polymerase Using Benzo[*a*]pyrene-modified Templates', *Carcinogenesis*, **12**, 973–976 (1991). [Published erratum appears in *Carcinogenesis*, **12**, 1531 (1991).]
60. A. Kootstra, L.K. Lew, R.S. Nairn, M.C. MacLeod, 'Preferential Modification of GC Boxes by Benzo[*a*]pyrene-7,8-diol-9,10-epoxide', *Mol. Carcinog.*, **1**, 239–244 (1989).
61. Y.H. Chen, D.F. Bogenhagen, 'Effects of DNA Lesions on Transcription Elongation by T7 RNA Polymerase', *J. Biol. Chem.*, **268**, 5849–5855 (1993).
62. D.J. Choi, D.J. Marino-Alessandri, N.E. Geacintov, D.A. Scicchitano, 'Site-specific Benzo[*a*]pyrene Diol Epoxide–DNA Adducts Inhibit Transcription Elongation by Bacteriophage T7 RNA Polymerase', *Biochemistry*, **33**, 780–787 (1994).
63. P.W. Doetsch, G.L. Chan, W.A. Haseltine, 'T4 DNA Polymerase ($3'-5'$) Exonuclease, an Enzyme for the Detection and Quantitation of Stable DNA Lesions: The Ultraviolet Light Example', *Nucleic Acids Res.*, **13**, 3285–3304 (1985).
64. J.M. Gale, K.A. Nissen, M.J. Smerdon, 'UV-induced Formation of Pyrimidine Dimers in Nucleosome Core DNA is Strongly Modulated with a Period of 10.3 Bases', *Proc. Natl. Acad. Sci. USA*, **84**, 6644–6648 (1987). [Published erratum appears in *Proc. Natl. Acad. Sci. USA*, **86**, 8386 (1989).]
65. L.F. Povirk, I.H. Goldberg, 'Endonuclease-resistant Apyrimidinic Sites Formed by Neocarzinostatin at Cytosine Residues in DNA: Evidence for a Possible Role in Mutagenesis', *Proc. Natl. Acad. Sci. USA*, **82**, 3182–3186 (1985).
66. J.R. Pierce, M. Nazimiec, M.S. Tang, 'Comparison of Sequence Preference of Tomaymycin- and Anthramycin-DNA Bonding by Exonuclease III and Lambda Exonuclease Digestion and UvrABC Nuclease Incision Analysis', *Biochemistry*, **32**, 7069–7078 (1993).
67. R.G. Higuchi, H. Ochman, 'Production of Single-stranded DNA Templates by Exonuclease Digestion Following the Polymerase Chain Reaction', *Nucleic Acids Res.*, **17**, 5865 (1989).
68. W.B. Mattes, 'Lesion Selectivity in Blockage of Lambda Exonuclease by DNA Damage', *Nucleic Acids Res.*, **18**, 3723–3730 (1990).

69. N.J. Combates, S.A. Winkle, 'Using Lambda Exonuclease Inhibition Assays to Map Carcinogen Binding Sites', *J. Biomol. Struct. Dyn.*, **10**, 63–72 (1992).
70. R.B. Walter, J. Pierce, R. Case, M.S. Tang, 'Recognition of the DNA Helix Stabilizing Anthramycin-N^2 Guanine Adduct by UVRABC Nuclease', *J. Mol. Biol.*, **203**, 939–947 (1988).
71. D. Suck, 'DNA Recognition by Structure-selective Nucleases', *Biopolymers*, **44**, 405–421 (1997).
72. A. Sancar, G.B. Sancar, 'DNA Repair Enzymes', *Annu. Rev. Biochem.*, **57**, 29–67 (1988).
73. L. Grossman, A.T. Yeung, 'The UvrABC Endonuclease System of Escherichia coli – A View from Baltimore', *Mutat. Res.*, **236**, 213–221 (1990).
74. B. Van Houten, 'Nucleotide Excision Repair in *Escherichia coli*', *Microbiol. Rev.*, **54**, 18–51 (1990).
75. L. Samson, J. Thomale, M.F. Rajewsky, 'Alternative Pathways for the In Vivo Repair of O^6-Alkylguanine and O^4-Alkylthymine in Escherichia coli: The Adaptive Response and Nucleotide Excision Repair', *EMBO J.*, **7**, 2261–2267 (1988).
76. G.T. Pauly, S.H. Hughes, R.C. Moschel, 'Response of Repair-competent Repair-deficient Escherichia coli to Three O^6-Substituted Guanines and Involvement of Methyl-directed Mismatch Repair in the Processing of O^6-Methylguanine Residues', *Biochemistry*, **33**, 9169–9177 (1994).
77. J.J. Lin, A. Sancar, 'A New Mechanism for Repairing Oxidative Damage to DNA: (A)BC Excinuclease Removes AP Sites and Thymine Glycols from DNA', *Biochemistry*, **28**, 7979–7984 (1989).
78. R.P. Fuchs, E. Seeberg, 'pBR322 Plasmid DNA Modified with 2-Acetylaminofluorene Derivatives: Transforming Activity and In Vitro Strand Cleavage by the Escherichia coli uvrABC Endonuclease', *EMBO J.*, **3**, 757–760 (1984).
79. A. Sancar, K.A. Franklin, G. Sancar, M.S. Tang, 'Repair of Psoralen and Acetylaminofluorene DNA Adducts by ABC Excinuclease', *J. Mol. Biol.*, **184**, 725–734 (1985).
80. M.S. Tang, J.R. Pierce, R.P. Doisy, M.E. Nazimiec, M.C. MacLeod, 'Differences and Similarities in the Repair of Two Benzo[*a*]pyrene Diol Epoxide Isomers Induced DNA Adducts by uvrA, uvrB, and uvrC Gene Products', *Biochemistry*, **31**, 8429–8436 (1992).
81. E. Seeberg, R.P. Fuchs, 'Acetylaminofluorene Bound to Different Guanines of the Sequence –GGCGCC– is Excised with Different Efficiencies by the UvrABC Excision Nuclease in a Pattern Not Correlated to the Potency of Mutation Induction', *Proc. Natl. Acad. Sci. USA*, **87**, 191–194 (1990).
82. D.C. Thomas, A.G. Morton, V.A. Bohr, A. Sancar, 'General Method for Quantifying Base Adducts in Specific Mammalian Genes', *Proc. Natl. Acad. Sci. USA*, **85**, 3723–3727 (1988).
83. H. Yamasaki, T.W. Roush, I.B. Weinstein, 'Benzo[*a*]-pyrene 7,8-Dihydrodiol-9,10-oxide Modification of DNA: Relation to Chromatin Structure and Reconstitution', *Chem. Biol. Interact.*, **23**, 201–213 (1978).
84. E. Ben-Asher, B. Strauss, 'Reaction of T7 DNA with a Polycyclic Aromatic Hydrocarbon. Lack of Structural Perturbation', *Biochim. Biophys. Acta.*, **607**, 10–22 (1980).
85. P. Pulkrabek, S. Leffler, D. Grunberger, I.B. Weinstein, 'Modification of Deoxyribonucleic Acid by a Diol Epoxide of Benzo[*a*]pyrene. Relation to Deoxyribonucleic Acid Structure and Conformation and Effects on Transfectional Activity', *Biochemistry*, **18**, 5128–5134 (1979).
86. P. Pulkrabek, S. Leffler, I.B. Weinstein, D. Grunberger, 'Conformation of DNA Modified with a Dihydrodiol Epoxide Derivative of Benzo[*a*]pyrene', *Biochemistry*, **16**, 3127–3132 (1977).
87. M.A. Chaudhry, M. Weinfeld, 'Induction of Double-strand Breaks by S1 Nuclease, Mung Bean Nuclease and Nuclease P1 in DNA Containing Abasic Sites and Nicks', *Nucleic Acids Res.*, **23**, 3805–3809 (1995).
88. P.W. Doetsch, R.P. Cunningham, 'The Enzymology of Apurinic/Apyrimidinic Endonucleases', *Mutat. Res.*, **236**, 173–201 (1990).
89. S. Ljungquist, T. Lindahl, 'Relation Between *Escherichia coli* Endonucleases Specific for Apurinic Sites in DNA and Exonuclease III', *Nucleic Acids Res.*, **4**, 2871–2879 (1977).
90. S. Henikoff, 'Unidirectional Digestion with Exonuclease III Creates Targeted Breakpoints for DNA Sequencing', *Gene*, **28**, 351–359 (1984).
91. J.D. Hoheisel, 'On the Activities of *Escherichia coli* Exonuclease III', *Anal. Biochem.*, **209**, 238–246 (1993).
92. C. Bernelot-Moens, B.W. Glickman, A.J. Gordon, 'Induction of Specific Frameshift and Base Substitution Events by Benzo[*a*]pyrene Diol Epoxide in Excision-repair-deficient *Escherichia coli*', *Carcinogenesis*, **11**, 781–785 (1990).
93. J.L. Yang, R.H. Chen, V.M. Maher, J.J. McCormick, 'Kinds and Location of Mutations Induced by (±)-7 beta,8 alpha-Dihydroxy-9 alpha,10 alpha-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene in the Coding Region of the Hypoxanthine (Guanine) Phosphoribosyltransferase Gene in Diploid Human Fibroblasts', *Carcinogenesis*, **12**, 71–75 (1991).
94. H. Ide, H. Murayama, A. Murakami, T. Morii, K. Makino, 'Effects of Base Damages on DNA Replication-mechanism of Preferential Purine Nucleotide Insertion Opposite Abasic Site in Template DNA', *Nucleic Acids Symp. Ser.*, **27**, 167–168 (1992).
95. H. Ide, H. Murayama, S. Sakamoto, K. Makino, K. Honda, H. Nakamura, M. Sasaki, N. Sugimoto, 'On the Mechanism of Preferential Incorporation of dAMP at Abasic Sites in Translesional DNA Synthesis. Role of Proofreading Activity of DNA Polymerase and Thermodynamic Characterization of Model Template-primers Containing an Abasic Site', *Nucleic Acids Res.*, **23**, 123–129 (1995).

96. T. Lindahl, B. Nyberg, 'Rate of Depurination of Native Deoxyribonucleic Acid', *Biochemistry*, **11**, 3610–3618 (1972).
97. R.M. Schaaper, T.A. Kunkel, L.A. Loeb, 'Infidelity of DNA Synthesis Associated with Bypass of Apurinic Sites', *Proc. Natl. Acad. Sci. USA*, **80**, 487–491 (1983).
98. P.L. Foster, E. Eisenstadt, J.H. Miller, 'Base Substitution Mutations Induced by Metabolically Activated Aflatoxin B1', *Proc. Natl. Acad. Sci USA*, **80**, 2695–2698 (1983).
99. T.A. Kunkel, 'Mutational Specificity of Depurination', *Proc. Natl. Acad. Sci. USA*, **81**, 1494–1498 (1984).
100. D.K. Klinedinst, N.R. Drinkwater, 'Mutagenesis by Apurinic Sites in Normal and Ataxia Telangiectasia Human Lymphoblastoid Cells', *Mol. Carcinog.*, **6**, 32–42 (1992).
101. G.P. Pfeifer, M.F. Denissenko, M.S. Tang, 'PCR-based Approaches to Adduct Analysis', *Toxicol. Lett.*, **102–103**, 447–451 (1998).
102. K.A. Grimaldi, S.R. McAdam, R.L. Souhami, J.A. Hartley, 'DNA Damage by Anti-cancer Agents Resolved at the Nucleotide Level of a Single Copy Gene: Evidence for a Novel Binding Site for Cisplatin in Cells', *Nucleic Acids Res.*, **22**, 2311–2317 (1994).
103. J. Komura, A.D. Riggs, 'Terminal Transferase-dependent PCR: A Versatile and Sensitive Method for In Vivo Footprinting and Detection of DNA Adducts', *Nucleic Acids Res.*, **26**, 1807–1811 (1998).
104. G.P. Pfeifer, R. Drouin, A.D. Riggs, G.P. Holmquist, 'In Vivo Mapping of a DNA Adduct at Nucleotide Resolution: Detection of Pyrimidine (6–4) Pyrimidone Photoproducts by Ligation-mediated Polymerase Chain Reaction', *Proc. Natl. Acad. Sci. USA*, **88**, 1374–1378 (1991).
105. D. Wei, V.M. Maher, J.J. McCormick, 'Effect of Nuclear Environment on the Distribution of Benzo[*a*]pyrene Diol Epoxide-induced Adducts in the HPRT Gene of Human Fibroblasts', *Carcinogenesis*, **17**, 2695–2701 (1996).
106. M.F. Denissenko, A. Pao, G.P. Pfeifer, M. Tang, 'Slow Repair of Bulky DNA Adducts Along the Nontranscribed Strand of the Human p53 Gene May Explain the Strand Bias of Transversion Mutations in Cancers', *Oncogene*, **16**, 1241–1247 (1998).
107. M. Tang, J.B. Zheng, M.F. Denissenko, G.P. Pfeifer, Y. Zheng, 'Use of UvrABC Nuclease to Quantify Benzo[*a*]pyrene Diol Epoxide–DNA Adduct Formation at Methylated Versus Unmethylated CpG Sites in the p53 Gene', *Carcinogenesis*, **20**, 1085–1089 (1999).
108. M.F. Denissenko, J. Cahill, T.B. Koudriakova, N. Gerber, G.P. Pfeifer, 'Quantitation and Mapping of Aflatoxin B1-induced DNA Damage in Genomic DNA Using Aflatoxin B1-8,9-epoxide and Microsomal Activation Systems', *Mutat. Res.*, **425**, 205–211 (1999).
109. S.D. Steigerwald, G.P. Pfeifer, A.D. Riggs, 'Ligation-mediated PCR Improves the Sensitivity of Methylation Analysis by Restriction Enzymes and Detection of Specific DNA Strand Breaks', *Nucleic Acids Res.*, **18**, 1435–1439 (1990).
110. G.P. Pfeifer, R. Drouin, G.P. Holmquist, 'Detection of DNA Adducts at the DNA Sequence Level by Ligation-mediated PCR', *Mutat. Res.*, **288**, 39–46 (1993).
111. H. Rodriguez, S.A. Akman, 'Mapping Oxidative DNA Damage at Nucleotide Level', *Free Radic. Res.*, **29**, 499–510 (1998).
112. C.A. Brennan, A.E. Manthey, R.I. Gumport, 'Using T4 RNA Ligase with DNA Substrates', *Methods Enzymol.*, **100**, 38–52 (1983).
113. W.M. Schmidt, M.W. Mueller, 'Controlled Ribonucleotide Tailing of cDNA Ends (CRTC) by Terminal Deoxynucleotidyl Transferase: A New Approach in PCR-mediated Analysis of mRNA Sequences', *Nucleic Acids Res.*, **24**, 1789–1791 (1996).

Polymerase Chain Reaction and Other Amplification Systems

Theodore K. Christopoulos
University of Patras, Patras, Greece

1 Introduction 1
2 Polymerase Chain Reaction 2
2.1 Simple Polymerase Chain Reaction 2
2.2 Nested Polymerase Chain Reaction 4
2.3 Long Template Polymerase Chain Reaction 4
2.4 Multiplex Polymerase Chain Reaction 4
3 Reverse Transcription Polymerase Chain Reaction 4
3.1 One-step Reverse Transcription Polymerase Chain Reaction 5
4 Methods for Analysis of Polymerase Chain Reaction Products 5
5 Quantitative Polymerase Chain Reaction 8
6 Immuno-polymerase Chain Reaction 10
7 Ligase Chain Reaction 10
8 Self-sustained Sequence Replication and Nucleic Acid Sequence-based Amplification 10
9 Strand Displacement Amplification 11
Abbreviations and Acronyms 12
Related Articles 12
References 12

The discovery of restriction endonucleases is perhaps the most significant technical milestone in the development of recombinant DNA technology. For, together with ligation and transformation procedures, they enabled the insertion of genes of any organism in suitable vectors (e.g. plasmids), introduction of the vector in bacteria and subsequent replication of the DNA of interest during cell division (cloning). Other important technical milestones include the development of methods for DNA sequencing and the chemical synthesis of oligodeoxynucleotides.

The polymerase chain reaction (PCR) is another technique which had a major impact in every area of research that involves nucleic acid analysis. PCR is a method of exponentially increasing the concentration of a specific nucleic acid sequence in a reaction mixture. A great advantage of PCR is that the target nucleic acid does not need to be particularly pure. It can be a minute part of an extremely complex mixture of biological material. Crude cell lysates provide a very good starting material. The ability to propagate, in a few hours, minute amounts of DNA that are too small for standard amplification (i.e. cloning) techniques gives PCR an extraordinary power and sensitivity.

The advent of PCR stimulated the research activity in the development of alternative exponential amplification systems such as the ligase chain reaction (LCR), the self-sustained sequence replication (3SR), the nucleic acid sequence-based amplification (NASBA) and the strand displacement amplification system (SDA). An essential feature of target amplification techniques is that the amplification is an iterative process consisting of successive cycles. The products synthesized in one cycle act as templates for the next cycle.

Amplification techniques can be used to amplify trace amounts of nucleic acid sequences in blood, cells, water, food and other clinical and environmental samples. Fixed tissue specimens, buccal cells from mouth washes, human hair, dried blood at the scene of a crime, single lymphoid or sperm cells are examples of starting material that may be used for nucleic acid analysis.

1 INTRODUCTION

The PCR entails the in vitro enzymic exponential amplification of a specific DNA sequence from a complex mixture. Increasing the copy number of a target nucleic acid sequence greatly facilitates its detection. The method involves repetitive bidirectional synthesis of a segment of double-stranded DNA occurring in successive rounds (thermocycling) in such a way that the product of one round serves as a template for the next round. Exponential amplification of specific RNA sequences is achieved by the reverse transcription polymerase chain reaction (RTPCR). PCR can be adapted to allow quantification of a target DNA or RNA sequence originally present in the sample prior to amplification. PCR provides sensitivities which are order of magnitudes greater than Southern and Northern analysis and nuclease protection assays. Due to its extraordinary sensitivity, the PCR is the method of choice for detection of nucleic acids at very low levels. PCR has greatly facilitated the detection of inherited mutations associated with genetic disease. It has also found widespread application in the detection of infectious agents in a variety of clinical and environmental samples. Conventional means for identification

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of pathogens include immunoassays for viral antigens or antiviral antibodies in the serum of patients and, in many cases, tedious and time-consuming culturing techniques that often lack the sensitivity and specificity required for routine diagnosis. PCR-based detection assays are simpler and offer much higher sensitivity and specificity. LCR, 3SR, NASBA and SDA are alternative exponential amplification systems. LCR is especially useful in mutation detection. Contrary to PCR and LCR, in 3SR, NASBA and SDA the exponential amplification is achieved in an isothermal manner. The high sensitivity of the exponential amplification techniques calls for extreme caution to avoid contamination of the samples from amplified sequences.

2 POLYMERASE CHAIN REACTION

2.1 Simple Polymerase Chain Reaction

The PCR, one of the most powerful techniques used in nucleic acid analysis and molecular biology today, was invented by Kary Mullis[1,2] who won the 1993 Nobel Prize in Chemistry for this contribution.

The principle of PCR is illustrated in Figure 1. In PCR, successive rounds (cycles) of heat denaturation, primer annealing and primer extension by a DNA polymerase result in a repetitive bidirectional synthesis of a segment of double-stranded DNA, the end result being an exponential increase in the number of target DNA copies. The sequence amplified by PCR is determined by the position of the primers. In the first heating step (~95 °C), the double-stranded DNA is denatured to form two single strands (disruption of hydrogen bonds). The temperature is then reduced to between 40 and 72 °C, thereby allowing the binding of two oligodeoxynucleotide primers (upstream and downstream) to complementary sequences on opposite strands of the target DNA fragment. The primers are oriented so that their 3′ ends point toward one another along the intervening sequence. The final temperature change of the cycle (to 72 °C) allows the 5′ → 3′ extension of each primer catalyzed by DNA polymerase. Following one complete cycle, two copies containing the sequence of interest are generated. As the cycling continues, the primers bind to the original target as well as to the newly synthesized strands and the number of target DNA copies approximately doubles at every cycle. The short products, which accumulate exponentially during PCR, are of uniform size corresponding to the distance between the 5′ ends of the two primers. The theoretical abundance of the short products is 2^n where n is the number of cycles performed. This translates to approximately one million times amplification in 20 cycles of PCR. The strands that are synthesized as copies of the original template (long products) are bounded at their 5′ end by a primer while their 3′ end is determined by the position at which the DNA polymerase stops. Strands containing undetermined 3′ termini can be generated only by copying the original template and therefore their accumulation is linear ($2n$). As a consequence, the long products are rare compared to the short product. The size and homogeneity of the PCR product can be assessed by agarose gel electrophoresis and ethidium bromide staining or by Southern blot and hybridization to labeled probes.

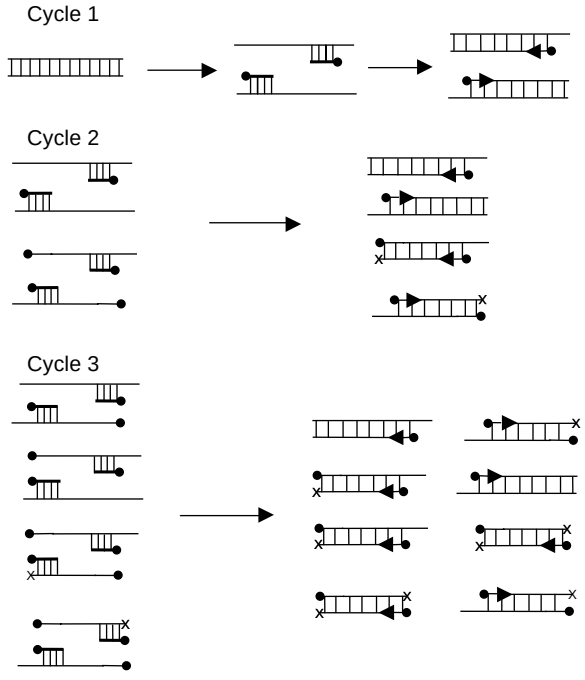

Figure 1 The principle of PCR. The denaturation, annealing and extension steps for the first three cycles are illustrated. The process results in exponential accumulation of the target sequence defined by the 5′ ends of the primers. At the end of the first cycle, the products are defined at their 5′ terminus (depicted as closed circles) by the primer but are not defined at the 3′ terminus. These products are termed long products. After the third cycle, the majority of the products will be defined at both termini (short products). The **x** represents a 3′ end that is defined by the primer. The arrows show the direction (5′ → 3′) of primer extension by the DNA polymerase.

The earliest PCR experiments used the Klenow fragment of *E. coli* DNA polymerase I at a temperature optimum of 37 °C, but this enzyme was easily inactivated at the high temperatures required for DNA denaturation and therefore fresh aliquots had to be added in each cycle. One of the most important developments in PCR was the introduction of Taq DNA polymerase, a thermostable DNA polymerase (temperature optimum at 72 °C) originally isolated from the thermophilic bacterium *Thermus aquaticus* (Taq) which lives in hot springs.[3] The

availability of Taq DNA polymerase has greatly simplified the automation of PCR since the enzyme is added only once, at the beginning of the reaction.[4] Furthermore, the use of Taq DNA polymerase has enhanced considerably the specificity and yield of the amplification process because it allowed the primer annealing and extension to occur at higher temperatures, thus minimizing competing mispriming events (annealing to nontarget sequences). The Taq DNA polymerase has no $3' \rightarrow 5'$ exonuclease activity (proofreading activity) but it possesses $5' \rightarrow 3'$ exonuclease activity during polymerization.[5] In situations where the PCR product is to be sequenced or sub-cloned for expression studies, the fidelity of the amplification is also crucial. Thermostable proofreading DNA polymerases (e.g. Pfu DNA polymerase) have become commercially available for these purposes.

The selection of primers is a key factor in determining the success of specific amplification. Primers are usually 20–28 bases long to permit high-stringency annealing, thus providing greater specificity. Shorter primers will yield nonspecific products and much longer sequences will result in low yield. Whenever possible, the two primers should have very similar melting temperatures to ensure that a given thermal profile is optimally efficient and specific for both primers. The G + C base composition should be about 40–60%. Intrastrand complementary regions resulting from runs of three or more G and C are inclined to form secondary structures (e.g. hairpin loops) and thus should be avoided especially at the 3′ end of the primer. Interstrand complementary regions at the 3′ ends of the primers should also be avoided in order to prevent the formation of primer dimers. This amplification artifact is a double-stranded DNA fragment whose length is equal to the sum of the two primers and appears to occur when one primer is extended by the polymerase over the other primer. Computer programs are available to aid in the process of primer selection. Primers are added in excess so that they do not become the rate-limiting factor in the reaction. The primer concentrations used range from 0.05 to 0.5 $\mu mol\,L^{-1}$. High concentrations may promote mispriming and the formation of primer dimers.

Since DNA polymerase extends the 3′ end of the primer, the 5′ end can be altered to create modified PCR products differing from the starting template. This provides a means of introducing restriction sites, specific mutations, promoters and regulatory elements (for subsequent expression). These sequences, although not complementary to the target DNA, become effectively incorporated into the final product. In a similar manner, various affinity tags or labels can be attached to the 5′ end of the primers and get incorporated into the newly synthesized strands to allow the capture and detection of the amplification products.[6]

Each of the deoxyribonucleoside triphosphates (dATP, dTTP, dCTP, dGTP) are usually present in the PCR mixtures at a concentration of 20–200 $\mu mol\,L^{-1}$. Higher concentrations are avoided because they promote mispriming at nontarget sequences (decreased PCR specificity) and misincorporation of nucleotides (decreased fidelity of DNA synthesis). The four dNTPs should be used at equal concentrations to avoid misincorporation. Like the primers, affinity tags or labels may be attached to the dNTPs and become incorporated into the amplification products for subsequent capture and detection. Mg^{2+} is bound to the dNTPs and is required for optimal activity of the polymerase. The Mg^{2+} concentration may affect the primer annealing, the denaturation temperature, PCR specificity and yield and therefore it should be optimized for each target DNA.

The inherent specificity of PCR lies in the fact that in order for exponential amplification to occur both primers must anneal, in the correct orientation, to a relatively short target DNA sequence. Moreover, after a few cycles of amplification the amount of target DNA has increased so much that less favourable competing events progressively become less frequent. The specificity of PCR may be improved further by increasing the annealing temperature and optimizing the incubation times for annealing and extension, to reduce the chances for mispriming and extension of nontarget sequences. The denaturation time and temperature should be long enough to allow complete separation of the DNA strands in each cycle, thus ensuring high yield.

A practical approach used widely to minimize the nonspecific amplification and the formation of primer dimers involves the addition of an essential reagent (e.g. DNA polymerase, primers, dNTPs or template) to the reaction mixture after it has reached high temperature, i.e. $>80\,^{\circ}C$ (''hot-start'' protocol).[7] Alternatively, the ''hot-start'' protocol may be performed by using a wax or jelly barrier that melts and permits mixing of aqueous components at a high temperature.[8] Another approach to ''hot-start'' PCR uses a monoclonal antibody that binds to and inactivates the Taq DNA polymerase at ambient temperature. This inhibition is reversed when the temperature is raised in the denaturation step of the first cycle and the enzyme functions normally during the course of PCR.[9]

The theoretical amplification of PCR (2^n) would be achieved only if the reaction could be performed with 100% efficiency (i.e. doubling of the DNA in each cycle). In practice, however, the system operates at a lower efficiency mainly as a result of an initial (lag) phase. This is attributed to a number of factors including the unavailability of the DNA target due to strand breaks, lack of its dissociation from other macromolecules, structural constraints, or a tendency of

the long parental strands to reanneal. A low efficiency during the first few cycles of PCR greatly compromises the overall yield of the reaction. Once the lag phase is overcome, the reaction enters the exponential phase. During late PCR cycles, however, the exponential rate of product accumulation decreases, a phenomenon commonly referred to as the ''plateau phase'' of PCR. The point at which the amplification reaches a plateau depends primarily on the amount of input target DNA and the number of cycles. The factors contributing to the plateau effect include competition between primer/template and template/template annealing, incomplete denaturation of the target strands and competition for reactants by any accumulated nonspecific products. Depletion and/or destruction at high temperatures of PCR reactants, such as primers, dNTPs and polymerase, as well as inhibition of PCR by end-products of the reaction, such as pyrophosphate, may also contribute to the plateau effect. The plateau phase is an unavoidable occurrence inherent to the technique.

The exquisite sensitivity resulting from the exponential amplification in PCR renders it highly vulnerable to the potential contamination of a sample by-product carried over from previous amplifications. Vigorous precautions must therefore be taken to avoid the setbacks associated with spurious amplifications.[10] Some of the common practices exercised in most PCR-performing laboratories include pre-aliquoting of reagents, use of dedicated pipettes and physical separation of the areas in which PCR samples are prepared, amplified and detected. Ultraviolet irradiation of reagent, buffers, pipettes, pipette tips and glassware is also common. Finally, the inclusion of negative (no target DNA) and positive controls in all experiments is essential for the accurate interpretation of results. The generation of false positives is a serious concern when PCR is used for diagnostic purposes.

An elegant approach used to minimize contamination involves the selective degradation of PCR product carryover.[11] Deoxyuridine triphosphate (dUTP) is added instead of dTTP in the PCR mixture and it is thus incorporated into the amplification products. If contamination is suspected, the enzyme uracil *N*-glycosylase (UNG) is added to the reaction mixture prior to thermal cycling. UNG catalyzes the excision of uracil from any potential single- or double-stranded carryover DNA (but not RNA) present in the reaction mixture prior to the first PCR cycle. The resulting abasic polynucleotides are susceptible to hydrolysis in slightly alkaline solutions (such as the PCR buffer) and elevated temperature and therefore cannot function as templates. Since UNG itself is inactivated at the high temperatures used in PCR, the amplification products generated during thermal cycling are not destroyed and can accumulate as usual. The amplified DNA can still be assayed by hybridization with complementary probes.

2.2 Nested Polymerase Chain Reaction

Nested primer PCR is a variation of PCR used to increase the specificity and sensitivity of amplification. This protocol uses two sets of primers, one set internal to the other, and amplification takes place in two PCR steps. The first PCR uses the outer set of primers to enrich the sample with the sequence of interest. The product serves as a template for a second PCR using the inner set of primers to generate a more homogeneous final product.[1] The disadvantage of this approach lies in the potential of contamination associated with opening the reaction tube, dilution (to decrease the concentration of the outer primers) and addition of the inner primer set. It is possible to perform nested PCR in a single reaction tube without dilution, in the presence of both primer pairs. This is achieved by designing the outer and inner primers to have different melting temperatures. During the first cycles only the outer primers are used and subsequently the denaturation and annealing temperatures are adjusted to allow only the inner primers to anneal.[12]

2.3 Long Template Polymerase Chain Reaction

The amplification of long fragments from genomic DNA has been a problem when using standard PCR conditions and Taq DNA polymerase. The size of PCR products generated with Taq DNA polymerase is limited to about 2 kbp. It was found[13,14] that combining the high processivity of Taq DNA polymerase with the proofreading accuracy of another thermostable DNA polymerase in a PCR allows amplification of much larger DNA templates (up to 50 kbp).

2.4 Multiplex Polymerase Chain Reaction

Multiplex PCR[15] entails the simultaneous amplification of several target DNA sequences (e.g. several regions of a large gene) in a single PCR. A unique pair of primers is designed for each target sequence. It is necessary that the primer pairs have close melting temperatures in order to avoid large variation in the amounts of amplified products. Multiplex PCR is more useful when the various target sequences exist in similar numbers in the sample. It has been applied to the diagnosis of genetic disease (e.g. muscular dystrophy), pathogen identification and forensic assays.

3 REVERSE TRANSCRIPTION POLYMERASE CHAIN REACTION

Amplification of specific RNA sequences can be accomplished by first generating a complementary copy of DNA

(cDNA) with reverse transcriptase. RTPCR has been an invaluable analytical tool for studying gene expression, monitoring various infectious agents and diagnosis and monitoring of disease.[16] Random hexamers, oligo-(dT) or the downstream PCR primer may be used as primers for the reverse transcription. The other PCR primer anneals to the generated complementary DNA strand and Taq polymerase synthesizes double-stranded cDNA which is further amplified in the usual way. The Moloney murine leukemia virus (MMLV) and the Amian myeloblastosis virus (AMV) reverse transcriptases are widely used in RTPCR. When amplifying RNA it is essential to remove the DNA present because it may serve as a template in PCR using the same primers. Undesirable target DNA should be degraded enzymically using deoxyribonuclease I. In the case of amplification of eukaryotic mRNA, the primers should be derived from separate exons so that PCR products arising from DNA are much longer, and in lower yield, because they contain intron sequences.

3.1 One-step Reverse Transcription Polymerase Chain Reaction

In this method both the reverse transcription and PCR are performed in one step (coupled) by using optimized buffer and one (or combination of) thermostable reverse transcriptase/DNA polymerase. RNA templates containing secondary structures are often difficult for viral reverse transcriptases to reverse-transcribe. The thermostable reverse transcriptases (e.g. Tth DNA polymerase) are able to synthesize cDNA at high temperatures that destabilize the secondary structures, thus providing increased sensitivity and specificity.[17]

4 METHODS FOR ANALYSIS OF POLYMERASE CHAIN REACTION PRODUCTS

The most commonly used method for analysis of PCR products is based on the electrophoretic separation of the DNA followed by staining with the intercalating fluorescent dye ethidium bromide. Quantification of the amplified DNA may then be carried out by taking the negative of a picture of the stained gel followed by scanning densitometry. Appropriate DNA markers are required in order to construct the calibration curve. The dependence of the amount of bound ethidium on the length of the DNA fragment should be taken into account when constructing a calibration curve. A major disadvantage of this method is its limited sensitivity (10 ng). Silver staining in conjunction with polyacrylamide gel electrophoresis provides higher sensitivity.

A more sensitive approach is based on the incorporation of radioactive dNTPs (labeled with ^{32}P) directly into the amplification products and detection by autoradiography of dried gels or membranes to which the products have been transferred (Southern blot). Alternatively, dNTP modified with nonradioactive labels (e.g. biotin or hapten) can be incorporated in the amplified DNA through PCR. The modified dNTP should be stable during thermal cycling. Following PCR, the mixture is electrophoresed and the fragments are transferred to a nylon membrane. The DNA is then detected by reacting with alkaline phosphatase or horseradish peroxidase conjugated to (strept)avidin. The substrates used give a color or fluorescent precipitate at the position at which the DNA is immobilized. The concentration of a modified (nonradioactively labeled) dNTP in the PCR mixture may have a significant influence on the activity of DNA polymerase. Thus, both the modified and unmodified dNTP (at an optimal ratio) should be included in the PCR mixture to ensure efficient incorporation and high yield.[18]

High-performance liquid chromatography (HPLC) with anion-exchange columns and absorbance monitoring at 260 nm has been used successfully for separation and determination of PCR products.[19] The reaction mixture can be injected onto the column directly without prior purification.

Capillary electrophoresis has been established rapidly as an efficient method for quantitative analysis of PCR products. The system consists of two electrolyte chambers connected with a separation capillary. Capillary electrophoresis is an easily automatable and rapid technique that provides qualitative and quantitative results with high resolution and low sample/reagent consumption.[20,21]

A fully automated instrument for PCR amplification and analysis of the products was developed.[39] The sample was automatically loaded, sealed, cycled in a rapid air thermal cycler and injected onto a gel filtration column for removal of salt and dNTPs followed by analysis of amplified DNA by capillary electrophoresis. A noncross-linked polymer-filled capillary was employed with automatic refilling. Detection was accomplished by laser-induced fluorescence.

The approaches described above detect all PCR products, specific and nonspecific. PCR products from clinical samples often appear on the slab gel as a number of bands, thus making difficult the identification of the correct size. Therefore, confirmation of the sequence of the specific product is usually required. This is accomplished by hybridization of the amplified DNA with complementary probes. Unlike ethidium bromide staining or direct labeling of PCR products, hybridization demonstrates the presence of the specific DNA band only.

The classical approach to DNA hybridization involves electrophoretic separation, Southern transfer and membrane hybridization using radioactive probes. Currently, however, the hybridization methods are undergoing a transition from radioactive labels to nonradioactive alternatives, which is driven by the need to enhance the sensitivity and facilitate automation while avoiding the health hazards and inconvenience associated with the use and disposal of radioactive reagents. The labels can be incorporated into the probes either enzymically (e.g. using DNA polymerase or deoxynucleotidyl transferase and modified deoxynucleoside triphosphates) or by chemical conjugation (e.g. introduction of $-NH_2$ groups into the probe via cytidine transamination and then conjugation to the reporter molecule). Generally, there are two strategies for the analysis of the hybrids. Either the reporter molecule is directly conjugated to the probe, or a ligand is attached to the probe and the hybrids are measured in a subsequent step by adding a specific, labeled binding protein. The ligand may be biotin or a hapten (e.g. digoxigenin). Labeled (strept)avidin or antihapten antibodies may then be employed for detection. Nonradioactive hybridization assays based on fluorescent, chemiluminescent or enzyme labels have been developed. Enzymes (such as alkaline phosphatase and horseradish peroxidase) are the most widely used nonradioactive reporters because they provide amplification through the high turnover of the substrate to the detectable product. The availability of a monoclonal antibody that selectively recognizes double-stranded DNA allowed the development of enzyme immunoassays for hybrids. PCR products are denatured and hybridized to immobilized probe. After washing, the hybrids are determined with an enzyme-labeled anti-double-stranded DNA antibody.[22]

Detection is usually based on fluorescence and chemiluminescence. However, the sensitivity of conventional fluorometry is restricted by the high background due to scattering, inherent fluorescence of the sample and luminescence from the cuvettes. Time-resolved fluorometry reduces the background in fluorescence measurements, thus improving the signal-to-noise ratio. This is achieved by using labels with long fluorescence lifetimes, such as the lanthanide ions Eu^{3+}, Tb^{3+}, Sm^{3+} and Dy^{3+} (Eu^{3+} is the most frequently used label). The sample is excited with a short pulse of light, from a laser or a flash lamp, and the measurement starts only after the short-lived background fluorescence has decayed.

The most successful strategy to attach Eu^{3+} to DNA is to use the isothiocyanate derivative of a complex of Eu^{3+} with an aminopolycarboxylic acid. These compounds react easily with primary amino groups that have been introduced in oligonucleotides and deoxynucleoside triphosphates. In one assay format, the PCR products are immobilized on a solid support and detected with a probe labeled with Eu^{3+}-chelate.[23] However, the complexes of Eu^{3+} and other lanthanides with aminopolycarboxylic acids fluoresce very weakly. For this reason, after the hybridization is completed, an enhancement solution is added which has a low pH (to ensure dissociation of Eu^{3+}) and contains a β-diketone (e.g. 2-naphthoyltrifluoroacetone) which is an excellent energy donor and forms highly fluorescent complexes with Eu^{3+}. The technique allows for multi-analyte assays because the four lanthanides have characteristic narrow emission bands.

A highly fluorescent Eu^{3+} trisbipyridine cryptate inclusion complex (Eu-TBP) has also been reported as a label for time-resolved fluorometric hybridization assays. In the Eu-TBP complex the bipyridine groups absorb the excitation light and, through an intramolecular energy transfer, the energy flows to the metal ion that subsequently fluoresces. Conjugates of Eu-TBP with antidinitrophenyl antibody have been prepared and used for determination of PCR products labeled with biotin and 2,4-dinitrophenol.[24] The advantage of this technique is that it allows direct determination of hybrids immobilized on a solid support without the need to dissociate the Eu^{3+}.

Another approach combines enzymic amplification with the background rejection capabilities of time-resolved fluorometry. This is accomplished by using alkaline phosphatase as a label and the phosphate ester of 5′-fluorosalicylate as a substrate. Enzymic cleavage of the phosphate group produces fluorosalicylate, which forms a long-lived fluorescent ternary complex with Tb^{3+}-EDTA at pH 12–13. The complex absorbs at a wavelength characteristic of the chelator (337 nm) and emits fluorescence characteristic of Tb^{3+}. PCR products are captured in microtiter wells and one strand is removed followed by hybridization to a digoxigenin-labeled probe. The hybrids are determined by using an antidigoxigenin antibody conjugated to alkaline phosphatase and time-resolved fluorometry.[25]

Chemiluminescence is the light emission occurring when chemically excited intermediates or products decay to the electronic ground state. Chemiluminescence provides higher sensitivity than conventional fluorometry because there is no background from excitation light.

Acridinium esters were used as chemiluminescent labels in hybridization assays. These molecules react rapidly (1–5 s) with hydrogen peroxide under alkaline conditions to produce light (flash-type reaction). The chemiluminescent decomposition of an acridinium ester involves attack of a peroxide anion at the C-9 ester to form a hydroperoxide, followed by loss of an alcoholate leaving group to produce an unstable dioxetanone. This species decomposes to CO_2 and *N*-methyl acridone in an electronically excited state. Decay of this molecule

to the ground state is accompanied by light emission at 430 nm. The *N*-hydroxysuccinimide derivative of acridinium ester reacts with primary amino groups introduced in oligonucleotide probes. Acridinium esters may be used in the analysis of PCR products either in a heterogeneous or a homogeneous assay configuration. Heterogeneous assays involve hybridization in solution and capture of the hybrids on a solid support (e.g. magnetic beads). The homogeneous assay is based on the alkaline hydrolysis of the ester bond of the acridinium molecule, rendering it permanently nonchemiluminescent. Conditions have been developed under which the hydrolysis is slow for the hybridized probe and fast for the free probe. Following hybridization, the reaction conditions are adjusted so that the excess probe is hydrolyzed and the chemiluminescence is a direct measure of the target sequence.[26]

The photoprotein aequorin has also been used as a reporter in chemiluminescence hybridization assays of amplified sequences. Aequorin is a complex of apoaequorin (a single polypeptide chain of 189 amino acids), the cofactor coelenterazine and molecular oxygen. Upon Ca^{2+} binding to aequorin, a flash-type luminescent reaction is triggered which yields apoaequorin coelenteramide, CO_2 and light at 469 nm with a quantum yield of 15%. Microtiter well-based sandwich-type hybridization assays have been developed in which the PCR product is denatured and hybridized simultaneously with an immobilized capture probe and a biotinylated detection probe. The hybrids are measured by using streptavidin–aequorin conjugates.[27]

Enzyme labels, such as alkaline phosphatase and horseradish peroxidase, can be combined with chemiluminogenic substrates for determination of amplified DNA. The adamantyl 1,2-dioxetane aryl phosphates, such as disodium 3-(4-methoxyspiro[1,2-dioxetane-3,2′-(5′-chloro)tricyclo[3.3.1.1$^{3.7}$]-decan]-4-yl)phenylphosphate (CSPD), are the most widely used substrates. Enzymic dephosphorylation of this molecule gives a phenoxide intermediate and this decomposes to produce light emission at 470 nm (glow type reaction).[28]

Electrochemiluminescence is the chemiluminescent reaction of species that are generated electrochemically at the surface of an electrode. The complex tris(2,2′-bipyridine)ruthenium (II) chelate ($[Ru(bpy)_3]^{2+}$) has been used as an electrochemiluminescent label in the analysis of PCR products.[29,30] When the voltage in an electrochemical cell becomes greater than a threshold value, the $[Ru(bpy)_3]^{2+}$ is oxidized at the surface of the anode to give $[Ru(bpy)_3]^{3+}$. Tripropylamine (TPA), which is present in large excess in the assay buffer, is also oxidized simultaneously at the electrode to $TPA^{+\bullet}$, which after deprotonation gives the strong reductant $TPA^{\bullet}$. $[Ru(bpy)_3]^{3+}$ undergoes reduction from $TPA^{\bullet}$ with production of the excited electronic state of $[Ru(bpy)_3]^{2+}$, which upon returning to the ground state emits a photon at 620 nm. An important feature of electrochemiluminescence is that the light-emitting reaction regenerates the original $[Ru(bpy)_3]^{2+}$ species which is then available to enter another cycle of oxidation and light emission. Thus, many photons are emitted from each $[Ru(bpy)_3]^{2+}$ label.

In a typical assay protocol, one of the PCR primers is biotinylated at the 5′ end. $[Ru(bpy)_3]^{2+}$ is attached to an oligonucleotide probe complementary to the biotinylated strand. Following PCR, the products are denatured, hybridized with the probe and captured to magnetic beads coated with streptavidin. A suspension of the beads is then introduced in the flow cell of the instrument (QPCR system 5000). A magnetic arm immobilizes the beads on the surface of the working electrode and allows washing with a TPA-containing buffer. Then voltage is applied and the luminescence is measured with a photomultiplier mounted closely to the flow cell. In an alternative protocol, $[Ru(bpy)_3]^{2+}$ is attached to the primer and incorporated into the amplified sequences, thus leading to double-labeled (biotin and $[Ru(bpy)_3]^{2+}$) DNA fragments that are captured to the beads and measured directly.

In recent years methods have been developed for continuous determination of the products generated during PCR, i.e. for real-time monitoring of DNA amplification.[31–33] One such approach uses the compound SYBR Green I, whose fluorescence increases upon binding to double-stranded DNA. A limitation of this method is that nonspecific products are co-determined. An alternative strategy employs an oligonucleotide probe that is labeled with a donor fluorescent molecule (6-carboxyfluorescein) at the 5′ end and an acceptor molecule (6-carboxytetramethylrhodamine) downstream a short distance. The probe is modified at the 3′ end to block its extension by the polymerase. The close proximity between donor and acceptor allows for energy transfer to occur and, therefore, no fluorescence from the donor is observed. During annealing, the probe hybridizes to the template DNA at a position flanked by the primers. However, because of the 5′ exonuclease activity of Taq DNA polymerase,[34] the hybridized probe is hydrolyzed during the extension phase of each cycle and the fluorescence increases.

A fluorometer with microvolume multisample capabilities and temperature control for real-time monitoring of PCR has been described.[35] Complete amplification and analysis requires 10–15 min (for 30 cycles). The samples are introduced in glass capillaries that facilitate heat transfer and also serve as cuvettes. The heating is by forced air.

Molecular beacons (oligonucleotide probes that consist of a stem-and-loop structure) have been used for monitoring of PCR product formation in a homogeneous

format during or after amplification. The loop part of the beacon is a single-stranded probe complementary to the target sequence. The stem is formed by extending each side of the probe with two oligonucleotide arms that are complementary to each other but unrelated to the target (the stem is a double helix). A donor fluorescent molecule is attached to the end of one arm and an acceptor molecule is attached to the other end. The close proximity between donor and acceptor in the stem structure allows for energy transfer to occur and, therefore, no fluorescence from the donor is observed. Upon hybridization of the probe region (loop) with the target sequence the double-stranded stem structure undergoes a conformational reorganization (strand separation) that places the donor and acceptor molecules far enough apart and the molecular beacon fluoresces. The dissociation of the stem strands is due to the fact that the probe–target hybrid is longer and more stable than the stem hybrid. It has been shown that the hairpin stem structure enhances the specificity of molecular beacons, enabling them to distinguish targets differing by a single nucleotide. Furthermore, several molecular beacons attached to different fluorophores can be used to monitor the accumulation of multiple targets in the same PCR.[36,37]

Oligonucleotide probes labeled with the fluorescent phthalocyanine dye La Jolla Blue were used for a homogeneous assay of PCR products by transient-state polarized fluorescence.[38] The dye has absorbance and emission maxima at 685 nm and 705 nm, respectively. Hybridization of the La Jolla Blue–oligonucleotide conjugate with the target results in longer rotational relaxation time of the hybrid and an increase in polarized fluorescence.

5 QUANTITATIVE POLYMERASE CHAIN REACTION

Perhaps the most challenging analytical problem associated with PCR is the determination of the amount of starting material, i.e. relating the analytical signal obtained from the amplification product to the initial number of target DNA or RNA molecules present in the sample prior to amplification.

The accumulation of product during the exponential phase of PCR can be described by Equation (1):

$$P = T(1 + E)^n \quad (1)$$

where P is the amount of amplification product, T is the initial amount of target DNA, n is the number of amplification cycles and E is the average efficiency of the reaction for each cycle. The theoretical value of E is 1 but usually it has a lower value. The above equation shows that, in the exponential phase, there is a linear relationship between the amount of product and the amount of input DNA. If E was constant, then determination of T could be accomplished simply by constructing a calibration curve with known amounts of a DNA standard (under constant n). The slope of the line would be equal to $(1 + E)^n$. However, the value of E depends on the sample and the reaction conditions. Moreover, small sample-to-sample variations in E lead to dramatic changes in the amount of product and large errors in estimating the initial quantity of target (especially when the cycle number is large). For instance, a 5% decrease in E (e.g. from 1 to 0.95) leads to an almost 50% decrease in the amount of product (when $n = 25$). In situations where quantitative analysis of RNA is being performed, the variation in the efficiency of reverse transcription should be taken into account.

One approach to quantitative PCR is to use an endogenous but unrelated sequence as an internal standard.[40,41] The standard is amplified with a different pair of primers either in a separate or in the same reaction tube. Thus, the ratio of the amplification products from target and internal standard is expected to be a linear function of the ratio of target DNA/internal standard molecules prior to amplification. This approach requires that the PCR for both the target DNA and the internal standard must be restricted to the exponential phase of product accumulation. If there is a large difference in the concentrations of target and internal standard, then the plateau phase will be reached much more rapidly for the one present at higher concentration and the ratio of the products will not reflect the ratio of the initial concentrations. This problem can be resolved by an appropriate dilution before PCR.

A DNA sequence from a single copy gene can be used as an internal standard for gene dosage studies. In the case of RNA determination, a cellular RNA whose expression level is constant between samples is employed. This process also corrects for inter-sample variation in RNA isolation and for any loss of template due to ribonucleases. The most widely used endogenous standards are RNAs from housekeeping genes such as β-actin and glyceraldehyde-3-phosphate dehydrogenase. Therefore, this approach provides the relative expression of the gene of interest with respect to a housekeeping gene but it does not give the absolute number of mRNA copies. A crucial assumption is that the housekeeping gene has an even rate of transcription among different tissues of an organism, at all stages of development and in all experimental treatments. However, it has been reported that β-actin mRNA concentrations increase with the malignant transformation of cells.[42]

Another approach to compensate for the variability in the efficiency is based on the co-amplification of the target

DNA or RNA with a known amount of recombinant DNA or RNA internal standard, respectively.[43,44] The internal standard is engineered to contain the same primer binding sites and to resemble as closely as possible the target sequence. Thus, any changes in the amplification efficiency are expected to affect both amplifications equally so that the ratio of the reaction products reflects the ratio of the two nucleic acid sequences in the starting mixture. Furthermore, the PCR products of target and internal standard should be distinguishable by size, hybridization or change in a restriction site. In the case of RNA analysis, RNA internal standards are advantageous compared to DNA standards because they allow compensation for any variation in the efficiency of both the reverse transcription and subsequent amplification. In the majority of reports the internal standard is designed to contain a deletion or insertion large enough to allow electrophoretic separation of the amplification products. Each DNA fragment is then quantified by ethidium bromide staining followed by scanning densitometry. In one widely used protocol, an increasing amount of internal standard is added separately to the same amount of target sequence followed by amplification. The goal is to find the amount of standard at which target and standard give an identical signal upon staining after agarose gel electrophoresis. Equal amounts of the two PCR products imply identical quantities prior to amplification. Thus, the ratios of the amplification products for target and standard are plotted as a function of the amount of internal standard and the point at which the ratio has a value of 1 is found by interpolation. The method provides absolute quantification of the target sequence. One of the main benefits of using a recombinant internal standard with the same primer binding sites as the target sequence is that it is not necessary to restrict the data acquisition in the exponential phase of amplification because the initial target to standard ratio remains the same even if the reaction proceeds to the plateau phase.

It has been shown that the amplification efficiency is inversely related to the size of the DNA.[45] Thus, the internal standard should have a similar size to the target sequence. For this reason, internal standards that differ from the target only in a restriction site have been constructed. Following amplification, the products are digested and separated electrophoretically. It has been observed however that co-amplification of DNA fragments that share considerable sequence homology leads to the formation of heteroduplexes during PCR even if their sizes are different.[45] During the annealing phase of a PCR cycle, a fraction of the single-stranded fragments of target DNA will hybridize with the complementary strand from the internal standard and vice versa. Heteroduplex formation poses serious problems with the analysis of the products by electrophoresis and HPLC. Upon electrophoresis the heteroduplexes may migrate between the target DNA and the internal standard, causing errors in the determination of the products, especially if they cannot be resolved from the homoduplexes. In the case of an internal standard having the same size as the target but differing only in a restriction site, the heteroduplexes interfere because they are resistant to digestion.

An internal standard that is distinguishable by hybridization has identical size to the target but contains a characteristic sequence to allow hybridization with a specific probe.[46,47] Analysis of the products includes a denaturation step prior to hybridization. Consequently, heteroduplex formation is not a concern for this method. Microtiter-well-based hybridization with nonradioactive detection is preferred because it facilitates automation and high-throughput analysis. One hybridization configuration involves capture of biotinylated products on streptavidin coated wells, removal of one strand with NaOH treatment and hybridization with probes specific for the target and the internal standard. Alternatively, the PCR products are denatured and allowed to hybridize with an immobilized capture probe and a labeled detection probe. Both configurations require splitting of each sample in different reaction vessels (wells). Recently, a dual-analyte chemiluminescence hybridization assay was developed for simultaneous determination of the amplified target and internal standard in the same well. The amplified target was hybridized with a digoxigenin-labeled probe and detected by using antidigoxigenin antibody labeled with aequorin. The amplified internal standard was hybridized, in the same well, with a fluorescein-labeled probe and the hybrids were determined by using antifluorescein antibody conjugated to alkaline phosphatase. Aequorin was quantified by adding a Ca^{2+}-containing triggering solution and alkaline phosphatase was measured by using a dioxetane chemiluminogenic substrate.[48]

Fluorometric methods that allow for homogeneous real-time monitoring of PCR products (as described in section 4) are also used in quantitative PCR or RTPCR. Specific probes labeled with different donor molecules and the same acceptor molecule may be used for determination of amplified target DNA and internal standard. In this approach, quantification is accomplished by measuring the number of cycles required to reach a certain fluorescent signal that corresponds to a concentration of PCR product that is clearly within the exponential phase of amplification. Indeed, by taking the logarithm of Equation (1) and solving for n, we have Equation (2):

$$n = \frac{\text{Log } P - \text{Log } T}{\log(1 + E)} \quad (2)$$

Equation (2) shows that if PCRs (performed with various amounts of input DNA) are stopped when a predetermined level of amplification product is generated, then the cycle number is linearly related to the logarithm of target DNA molecules. For Equation (2) to be valid, the PCR must be in the exponential phase of product accumulation.

Real-time quantitative PCR methods are advantageous in that they do not require post-PCR sample processing and they are easily automatable. However, the detectability of the fluorescent label is in the nanomole per liter range and the fluorescence value provides an indirect measure of the amplification product. On the other hand, post-PCR chemiluminescence and time-resolved fluorescence hybridization assays offer 1000 times higher sensitivity since they can detect amplification products in pmoles per liter.

6 IMMUNO-POLYMERASE CHAIN REACTION

Immuno-PCR is an immunoassay in which a DNA fragment is used as a label.[49] After completion of the immunoreaction, the label is subjected to PCR amplification and the products are detected by gel electrophoresis or enzyme-linked immunosorbent assay.[50]

7 LIGASE CHAIN REACTION

LCR[51] is an exponential in vitro amplification technique that employs a thermostable ligase and allows the discrimination of DNA sequences differing in only a single base pair. The principle of LCR is as follows (Figure 2): First, the DNA target is denatured at 95 °C. Then, two allele-specific oligonucleotide probes are allowed to hybridize at adjacent positions (no gap) of the sense strand of target DNA at 60–65 °C. In a similar manner, a second pair of probes hybridizes with the antisense strand of the target sequence. The junction of the primers is positioned so that the nucleotide at the 3′ end of the upstream primer coincides with a potential base difference in the target sequence. The probes are joined by ligase and the products can serve as templates in the next reaction cycle, leading to an exponential amplification process. If there is a mismatch at the primer junction, the primers will not be ligated and amplification product will not be formed. A second set of four oligonucleotides can be used in a separate reaction to detect the mutant allele.

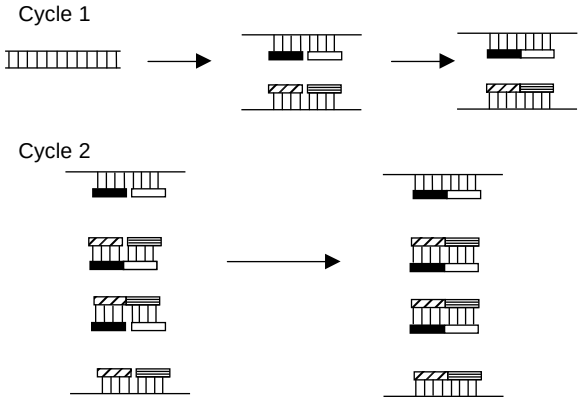

Figure 2 Schematic presentation of the principle of the LCR. Each cycle includes denaturation, annealing of four oligonucleotide probes (rectangular boxes) to the opposite strands and ligation of the nicks.

PLCR (a modification of LCR) is used for allele-specific DNA amplification and detection using the Stoffel fragment of Taq DNA polymerase and a thermostable ligase. After DNA denaturation, two pairs of complementary but not adjacent oligonucleotides are allowed to anneal with the sense and antisense strand of the target sequence, respectively. These primers anneal so that there is a 2–3 nucleotide gap between each other. The Stoffel fragment, lacking not only the $3' \rightarrow 5'$ proofreading exonuclease activity but also the $5' \rightarrow 3'$ exonuclease activity, can fill the gap with the appropriate dNTPs and subsequently the ligase will link the nick.

The classical approach to detection of LCR products involves labeling the upstream primer with ^{32}P at the 3′ end, separation of the LCR products by denaturing gel electrophoresis and autoradiography. Another approach employs fluorescently labeled probes that allow for analysis of LCR products by using a fluorometric DNA sequencer. Microtiter-well-based methods for determination of LCR products use one probe labeled at the 5′ end with biotin and the other probe conjugated to a nonradioactive reporter molecule (e.g. 5′-carboxyfluorescein or digoxigenin).[52,53]

LCR is often utilized in conjuction with an initial PCR amplification to achieve higher sensitivity.

8 SELF-SUSTAINED SEQUENCE REPLICATION AND NUCLEIC ACID SEQUENCE-BASED AMPLIFICATION

3SR[54] and NASBA[55] are isothermal techniques for in vitro amplification of target RNA or DNA by the simultaneous concerted action of AMV reverse transcriptase, ribonuclease H (RNase H) and T7 RNA

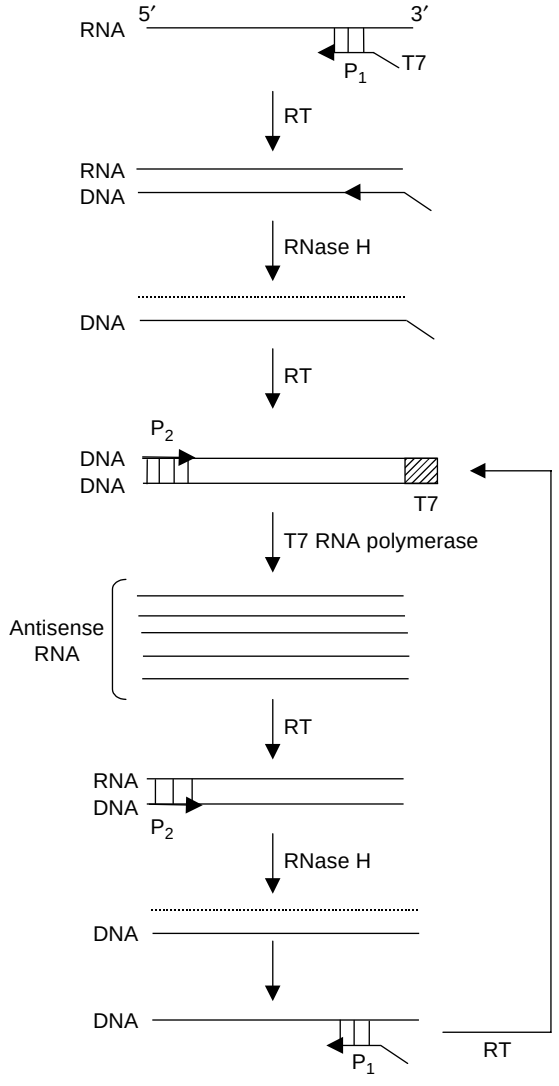

Figure 3 Schematic illustration of the principle of 3SR and NASBA systems. P_1, primer containing the T7 promoter sequence (T7) at the 5′ end; RT, reverse transcriptase; RNase H, ribonuclease H; P_2, second primer.

polymerase. The major amplification product is single-stranded RNA but RNA : DNA hybrids and double-stranded DNA are also synthesized, although to a minor extent. The amplification steps are shown in Figure 3. The reaction begins by annealing of the first primer on the RNA target. The primer carries a T7 promoter sequence at the 5′ end. Reverse transcriptase then synthesizes a complementary DNA strand resulting in a DNA : RNA hybrid. RNase H then hydrolyzes the RNA strand of DNA : RNA hybrids thus leaving single-stranded DNA. Subsequently a second primer anneals to the DNA strand followed by the synthesis of a complementary strand of DNA by reverse transcriptase. This double-stranded DNA carries the T7 promoter sequence and serves as a template for T7 RNA polymerase that synthesizes many copies of RNA transcripts. Each of the newly synthesized RNA molecules re-enters the cycle and serves as a template for a new round of amplification leading to an exponential accumulation of the product. An amplification of 10^9 fold is achieved in about 2 h.

It has been shown that with an initial denaturation step this system could be modified to amplify DNA but the product of the reaction is always RNA. It was found[56] that the presence of dimethylsulfoxide (DMSO) and sorbitol in the reaction mixture enhances the RNase activity of AMV reverse transcriptase sufficiently to allow the reaction to proceed without RNase H. Alternatively, human immunodeficiency virus (HIV) reverse transcriptase may be used instead of AMV reverse transcriptase to allow RNase H to be omitted.

Amplification products are determined by methods similar to those used in the analysis of PCR products.

9 STRAND DISPLACEMENT AMPLIFICATION

SDA is an isothermal in vitro method for amplifying DNA sequences.[57] The technique is based on the ability of a restriction enzyme to nick the unmodified strand of a hemimodified restriction site and the ability of a $5' \rightarrow 3'$ exonuclease-deficient DNA polymerase to extend the 3′ end at the nick and displace the downstream strand. The target DNA is heat-denatured in the presence of all reagents except the restriction enzyme and the polymerase. Two primers bind to opposite strands of the DNA. At their 5′ ends the primers contain a recognition site for a restriction endonuclease (e.g. Hinc II). DNA replication using dCTP, dGTP, dTTP and dATPαS produces a double-stranded hemiphosphorothioate recognition site. Hinc II nicks the unprotected primer strand leaving intact the modified complementary strand. DNA polymerase then extends the 3′ end at the nick and displaces the downstream fragment. The polymerization/displacement step regenerates a nickable recognition site. Target amplification by SDA is exponential because each displaced strand serves as target for the other primer (i.e. the primer that was extended for synthesis of the complementary strand).

Methods for analysis of SDA products include hybridization with ^{32}P-labeled probes or nonradioactive hybridization assay configurations. More recently, fluorescence polarization was employed for real-time monitoring of SDA products.[58]

ABBREVIATIONS AND ACRONYMS

AMV	Amian Myeloblastosis Virus
CSPD	Disodium 3-(4-methoxyspiro[1,2-dioxetane-3,2′-(5′-chloro)tricyclo[3.3.1.1$^{3.7}$]-decan]-4-yl)phenylphosphate
DMSO	Dimethylsulfoxide
HIV	Human Immunodeficiency Virus
HPLC	High-performance Liquid Chromatography
LCR	Ligase Chain Reaction
MMLV	Moloney Murine Leukemia Virus
NASBA	Nucleic Acid Sequence-based Amplification
PCR	Polymerase Chain Reaction
RTPCR	Reverse Transcription Polymerase Chain Reaction
SDA	Strand Displacement Amplification
TPA	Tripropylamine
UNG	Uracil *N*-glycosylase
3SR	Self-sustained Sequence Replication

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Nucleic Acid Analysis in Clinical Chemistry

Forensic Science **(Volume 5)**
Polymerase Chain Reaction in the Forensic Analysis of DNA

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Probes • Fluorescence In Situ Hybridization

REFERENCES

1. K.B. Mullis, F.A. Faloona, 'Specific Synthesis of DNA In Vitro via a Polymerase-catalyzed Chain Reaction', *Methods Enzymol.*, **155**, 335–350 (1987).
2. R.K. Saiki, S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, 'Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia', *Science*, **230**, 1350–1354 (1985).
3. F.C. Lawyer, S. Stoffel, R.K. Saiki, K. Myambo, R. Drummond, D.H. Gelfand, 'Isolation, Characterization, and Expression in *Escherichia coli* of the DNA Polymerase Gene from *Thermus aquaticus*', *J. Biol. Chem.*, **264**, 6427–6437 (1989).
4. R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, H.A. Erlich, 'Primer-directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase', *Science*, **239**, 487 (1988).
5. W. Bloch, 'A Biochemical Perspective of the Polymerase Chain Reaction', *Biochemistry*, **30**, 2735–2747 (1991).
6. S. Bortolin, T.K. Christopoulos, 'Time-resolved Immunofluorometric Determination of Specific mRNA Sequences Amplified by the Polymerase Chain Reaction', *Anal. Chem.*, **66**, 4302–4307 (1994).
7. R.T. D'Aquila, L.J. Bechtel, J.A. Videler, J.J. Eron, P. Gorczyca, J.C. Kaplan, 'Maximizing Sensitivity and Specificity of PCR by Pre-amplification Heating', *Nucleic Acids Res.*, **19**, 3749 (1991).
8. R.M. Horton, B.L. Hoppe, B.M. Conti-Tronconi, 'Ampligrease: Hot-start PCR using Petroleum Jelly', *BioTechniques*, **16**, 42–43 (1994).
9. D.E. Kellogg, I. Rybalkin, S. Chen, N. Mukhamedova, T. Vlasik, P.D. Siebert, A. Chenchik, 'TaqStart Antibody: Hot-start PCR Facilitated by a Neutralizing Monoclonal Antibody Directed Against Taq DNA Polymerase', *BioTechniques*, **16**, 1134–1137 (1994).
10. S. Kwok, R. Higuchi, 'Avoiding False Positives with PCR', *Nature*, **339**, 237–238 (1989).
11. M.C. Longo, M.S. Berninger, J.L. Hartley, 'Use of Uracil DNA Glycosylase to Control Carry-over Contamination in Polymerase Chain Reactions', *Gene*, **93**, 125–128 (1990).
12. H.A. Erlich, D. Gelfand, J.J. Sninsky, 'Recent Advances in the Polymerase Chain Reaction', *Science*, **252**, 1643–1651 (1991).
13. W.M. Barnes, 'PCR Amplification of up to 35-kb DNA with High Fidelity and High Yield from λ Bacteriophage Templates', *Proc. Natl. Acad. Sci. USA*, **91**, 2216–2220 (1994).
14. S. Cheng, C. Fockler, W.M. Barnes, R. Higuchi, 'Effective Amplification of Long Targets from Cloned Inserts and Human Genomic DNA', *Proc. Natl. Acad. Sci. USA*, **91**, 5695–5699 (1994).
15. J.S. Chamberlain, R.A. Gibbs, J. Ranier, P.N. Nguyen, C.T. Caskey, 'Deletions Screening of the Duchenne Muscular Dystrophy Locus via Multiplex DNA Amplification', *Nucleic Acids Res.*, **16**, 11 141 (1988).
16. E.S. Kawasaki, S.S. Clark, M.Y. Coyne, S.D. Smith, R. Champlin, O.N. Witte, F.P. McCormick, 'Diagnosis of Chronic Myeloid and Acute Lymphocytic Leukemias by Detection of Leukemia-specific mRNA Sequences Amplified In Vitro', *Proc. Natl. Acad. Sci. USA*, **85**, 5698–5702 (1988).
17. T.W. Myers, D.H. Gelfand, 'Reverse Transcription and DNA Amplification by a *Thermus thermophilus* DNA Polymerase', *Biochemistry*, **30**, 7661–7666 (1991).
18. S. Bortolin, T.K. Christopoulos, 'Determination of Polymerase Chain Reaction Products by Time-resolved Immunofluorometry', *Clin. Biochem.*, **29**, 179–182 (1996).
19. R. Zeillinger, C. Schneeberger, P. Speiser, F. Kury, 'Rapid Quantitative Analysis of Differential PCR Products by High-performance Liquid Chromatography', *BioTechniques*, **15**, 89–95 (1993).

20. J.M. Butler, B.R. McCord, J.M. Jung, M.R. Wilson, B. Budowle, R.O. Allen, 'Quantitation of Polymerase Reaction Products by Capillary Electrophoresis Using Laser Fluorescence', *J. Chromatogr. B*, **658**, 271–280 (1994).
21. P. Fortina, J. Cheng, M.A. Shoffner, S. Surrey, W.M. Hitchcock, L.J. Kricka, P. Wilding, 'Diagnosis of Duchenne/Becker Muscular Dystrophy and Quantitative Identification of Carrier Status by Use of Entangled Solution Capillary Electrophoresis', *Clin. Chem.*, **43**, 745–751 (1997).
22. G. Mantero, A. Zonaro, A. Albertini, P. Bertolo, D. Primi, 'DNA Enzyme Immunoassay: General Method for Detecting Products of Polymerase Chain Reaction', *Clin. Chem.*, **37**, 422–429 (1991).
23. A. Iitia, M. Mikola, N. Gregersen, P. Hurskainen, T. Lovgren, 'Detection of a Point Mutation Using Short Oligonucleotide Probes in Allele-specific Hybridization', *BioTechniques*, **17**, 566–573 (1994).
24. E. Lopez, C. Chypre, B. Alpha, G. Mathis, 'Europium (III) Trisbipyridine Cryptate Label for Time-resolved Fluorescence Detection of Polymerase Chain Reaction Products Fixed on a Solid Support', *Clin. Chem.*, **39**, 196–201 (1993).
25. S. Bortolin, T.K. Christopoulos, 'Detection of BCR-ABL Transcripts from the Philadelphia Translocation by Hybridization in Microtiter Wells and Time-resolved Immunofluorometry', *Clin. Chem.*, **41**, 693–699 (1995).
26. N.C. Nelson, M.A. Reynolds, L.J. Arnold, 'Detection of Acridinium Esters by Chemiluminescence', in *Nonisotopic Probing, Blotting and Sequencing*, ed. L.J. Kricka, 2nd edition, Academic Press, San Diego, 1995.
27. B. Galvan, T.K. Christopoulos, 'Bioluminescence Hybridization Assays Using Recombinant Aequorin. Application to the Detection of Prostate-specific Antigen mRNA', *Anal. Chem.*, **68**, 3545–3550 (1996).
28. C.S. Martin, L. Butler, I. Bronstein, 'Quantitation of PCR Products with Chemiluminescence', *BioTechniques*, **18**, 908–912 (1995).
29. G.F. Blackburn, H.P. Shah, J.H. Kenten, J. Leland, R.A. Kamin, J. Link, J. Peterman, M.J. Powell, A. Shah, D.B. Talley, S.K. Tyagi, E. Wilkins, T.-G. Wu, R.J. Massey, 'Electrochemiluminescence Detection for the Development of Immunoassays and DNA Probe Assays for Clinical Diagnostics', *Clin. Chem.*, **37**, 1534–1539 (1991).
30. J. DiCesare, B. Grossman, E. Katz, E. Picozza, R. Ragusa, T. Woudenberg, 'A High-sensitivity Electrochemiluminescence-based Detection System for Automated PCR Product Quantitation', *BioTechniques*, **15**, 152–157 (1993).
31. C.A. Heid, J. Stevens, K.J. Livak, P.M. Williams, 'Real Time Quantitative PCR', *Genome Res.*, **6**, 986–994 (1996).
32. U.E. Gibson, C.A. Heid, P.M. Williams, 'A Novel Method for Real Time Quantitative RT-PCR', *Genome Res.*, **6**, 995–1001 (1996).
33. C.T. Wittwer, M.G. Herrmann, A.A. Moss, R.P. Rasmussen, 'Continuous Fluorescence Monitoring of Rapid Cycle DNA Amplification', *BioTechniques*, **22**, 130–138 (1997).
34. P.M. Holland, R.D. Abramson, R. Watson, D.H. Gelfand, 'Detection of Specific Polymerase Chain Reaction Products by Utilizing the $5' \rightarrow 3'$ Exonuclease Activity of *Thermus aquaticus* DNA Polymerase', *Proc. Natl. Acad. Sci. USA*, **88**, 7276–7280 (1991).
35. C.T. Wittwer, K.M. Ririe, R.V. Andrew, D.A. David, R.A. Gundry, U.J. Balis, 'The Light Cycler: a Microvolume Multisample Fluorimeter with Rapid Temperature Control', *BioTechniques*, **22**, 176–181 (1997).
36. S. Tyagi, F.R. Kramer, 'Molecular Beacons: Probes that Fluoresce upon Hybridization', *Nat. Biotechnol.*, **14**, 303–308 (1996).
37. S. Tyagi, D.P. Bratu, F.R. Kramer, 'Multicolor Molecular Beacons for Allele Discrimination', *Nat. Biotechnol.*, **16**, 49–53 (1998).
38. R. Devlin, R.M. Studholme, W.B. Dandliker, E. Fahy, K. Blumeyer, S.S. Ghosh, 'Homogeneous Detection of Nucleic Acids by Transient-state Polarized Fluorescence', *Clin. Chem.*, **39**, 1939–1943 (1993).
39. H. Swerdlow, B.J. Jones, C.T. Wittwer, 'Fully Automated DNA Reaction and Analysis in a Fluidic Capillary Instrument', *Anal. Chem.*, **69**, 848–855 (1997).
40. K.E. Noonan, C. Beck, T.A. Holzmayer, J.E. Chin, J.S. Wunder, I.L. Andrulis, A.F. Gazdar, C.L. Willman, B. Griffith, D.D. Von Hoff, I.B. Roninson, 'Quantitative Analysis of MDR1 (Multidrug Resistance) Gene Expression in Human Tumors by Polymerase Chain Reaction', *Proc. Natl. Acad. Sci. USA*, **87**, 7160–7164 (1990).
41. L.D. Murphy, C.E. Herzog, J.B. Rudick, A.T. Fojo, S.E. Bates, 'Use of the Polymerase Chain Reaction in the Quantitation of mdr-1 Gene Expression', *Biochemistry*, **29**, 10 351–10 356 (1990).
42. L.E. Ostrowski, P. Krieg, J. Finch, A.E. Cress, R.B. Nagle, G.T. Bowden, 'Expression of β-Actin During Progression of Mouse Skin Tumors', *Carcinogenesis*, **10**, 1439–1444 (1989).
43. A.M. Wang, M.V. Doyle, D.F. Mark, 'Quantitation of mRNA by the Polymerase Chain Reaction', *Proc. Natl. Acad. Sci. USA*, **86**, 9717–9721 (1989).
44. G. Gilliland, S. Perrin, K. Blanchard, H.F. Bunn, 'Analysis of Cytokine mRNA and DNA: Detection and Quantitation by Competitive Polymerase Chain Reaction', *Proc. Natl. Acad. Sci. USA*, **87**, 2725–2729 (1990).
45. R.K. McCulloch, C.S. Choong, D.M. Hurley, 'An Evaluation of Competitor Type and Size for Use in the Determination of mRNA by Competitive PCR', *PCR Methods Appl.*, **4**, 219–226 (1995).

46. S. Bortolin, T.K. Christopoulos, M. Verhaegen, 'Quantitative Polymerase Chain Reaction Using a Recombinant DNA Internal Standard and Time-resolved Fluorometry', *Anal. Chem.*, **68**, 834–840 (1996).
47. M. Verhaegen, P.C. Ioannou, T.K. Christopoulos, 'Quantification of Prostate-specific Antigen mRNA by Coamplification with a Recombinant RNA Internal Standard and Microtiter Well-based Hybridization', *Clin. Chem.*, **44**, 1170–1176 (1998).
48. M. Verhaegen, T.K. Christopoulos, 'Quantitative Polymerase Chain Reaction Based on a Dual-analyte Chemiluminescence Hybridization Assay for Target DNA and Internal Standard', *Anal. Chem.*, **70**, 4120–4125 (1998).
49. T. Sano, C.L. Smith, C.R. Cantor, 'Immuno-PCR: Very Sensitive Antigen Detection by Means of Specific Antibody-DNA Conjugates', *Science*, **258**, 120–122 (1992).
50. C.M. Niemeyer, M. Adler, D. Blohm, 'Fluorometric Polymerase Chain Reaction (PCR) Enzyme-linked Immunosorbent Assay for Quantification of Immuno-PCR Products in Microplates', *Anal. Biochem.*, **246**, 140–145 (1997).
51. F. Barany, 'Genetic Disease Detection and DNA Amplification Using Cloned Thermostable Ligase', *Proc. Natl. Acad. Sci. USA*, **88**, 189–193 (1991).
52. E.S. Winn-Deen, C.A. Batt, M. Wiedmann, 'Non-radioactive Detection of Mycobacterium Tuberculosis LCR Products in a Microtitre Plate Format', *Mol. Cell. Probes*, **7**, 179–186 (1993).
53. M. Wiedmann, F. Barany, C.A. Batt, 'Detection of *Lysteria monocytogenes* with a Nonisotopic Polymerase Chain Reaction-coupled Legase Chain Reaction Assay', *Appl. Environ. Microbiol.*, **59**, 2743–2745 (1993).
54. J. Guatelli, K. Whitfield, D. Kwoh, K. Barringer, D. Richman, T. Gingeras, 'Isothermal In Vitro Amplification of Nucleic Acids by a Multienzyme Reaction Modeled After Retroviral Replication', *Proc. Natl. Acad. Sci. USA*, **87**, 1874–1878 (1990).
55. T. Kievits, B. von Gemen, D. van Strijp, R. Schukkink, M. Dircks, H. Adriaanse, L. Malek, R. Sooknanan, P. Lens, 'NASBA Isothermal In Vitro Nucleic Acid Amplification Optimized for the Diagnosis of HIV-1 Infection', *J. Virol. Methods*, **35**, 273–286 (1991).
56. E. Fahy, D.Y. Kwoh, T.R. Gingeras, 'Self-sustained Sequence Replication (3SR): an Isothermal Transcription-based Amplification System Alternative to PCR', *PCR Methods Appl.*, **1**, 25–33 (1991).
57. G.T. Walker, M.S. Fraiser, J.L. Schram, M.C. Little, J.G. Nadeau, D.P. Malinowski, 'Strand Displacement Amplification – an Isothermal In Vitro Amplification Technique', *Nucleic Acids Res.*, **20**, 1691–1696 (1992).
58. G.T. Walker, J.G. Nadeau, C.P. Linn, R.F. Devlin, W.B. Dandliker, 'Strand Displacement Amplification (SDA) and Transient-state Fluorescence Polarization Detection of *Mycobacterium tuberculosis* DNA', *Clin. Chem.*, **42**, 9–13 (1996).

Radiation Hybrid Mapping

William R. Newell
Aventis Cambridge Genomics Center, Cambridge, USA

1 Introduction **2**
1.1 Experimental Evidence for Linear Genetic Maps 2
1.2 Use of the Recombination Frequency to Construct Linear Maps 2
1.3 Somatic Cell (Goss–Harris) Mapping 3
1.4 Radiation Hybrid Mapping 3
1.5 Whole-genome Radiation Hybrid Mapping 3
1.6 Growth of the Field 4
2 Construction of a Radiation Hybrid Panel in Silico **4**
3 Analytical Methods **5**
3.1 Strength of Linkage in Meiotic Maps 5
3.2 Strength of Linkage in Radiation Hybrid Experiments 6
3.3 Estimation of θ by Goss and Harris (1975) 7
3.4 Estimation of θ by Cox et al. (1990) 8
3.5 Estimate of θ Directly from the Observed Marker Retention 9
3.6 Estimate of θ Using Maximum Likelihood 10
3.7 Methods for Determining Order 11
3.8 Genetic Analysis Workshop 7 11
3.9 Hidden Markov Models 11
3.10 Distance Methods 12
4 Main Software Available **12**
4.1 RHMAP 12
4.2 RHMAPPER 12
4.3 Z-RHMAPPER 12
4.4 SAMAPPER 12
4.5 MultiMap/RADMAP 13
4.6 Distance Geometry Map 13
5 Practical Methods **13**
5.1 General 13
5.2 Obtaining a Radiation Hybrid Panel 13
5.3 Typing Markers on the Panel 13
5.4 Finding Close Markers 13
5.5 The Map Location 13
6 Properties of Radiation Hybrid Panels **14**
6.1 Construction of Panels is under Experimental Control 14
6.2 Available Panels 14
6.3 Effect of Marker Retention Frequency 15
6.4 Effect of Number of Hybrids on Resolution and Range 15
6.5 Effect of Dose on Panel Range and Resolution 16
6.6 Effect of Marker Density on Significance of Linkage 17
7 Applications of Radiation Hybrid Mapping **17**
7.1 Disease Gene Mapping 17
7.2 Whole-genome Human Maps 18
7.3 Mapping Other Organisms 19
7.4 Integrating Data 19
8 Conclusions **19**
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

Radiation hybrid (RH) mapping is a convenient and widely used method for rapidly locating new DNA sequences on genomes in relation to other DNA markers. The aim of the technique is to determine the correct order and relative distances between a set of markers, or to locate a new marker to its most likely location on a genome. We review the history of the technique, its current status, and the main analytical techniques used to generate maps from raw data. The physical processes and assumptions underlying the technique are examined. These processes are used to derive models from which the required information, location of markers with respect to each other on a linear map of DNA, is obtained by a variety of computational means. The RH technique has proved indispensable in mapping the human genome, since, unlike genetic mapping, the conditions under which panels are made is completely under experimental control, so mapping at any degree of resolution is possible. The other major advantage of RH mapping over genetic mapping is that markers need not be polymorphic; it is presence or absence of a marker that is detected, not type. It has therefore been used to map the large numbers of expressed sequence tags that have been generated as part of the human genome project (HGP). This has given rise to the first human gene map and subsequent updates, so that currently 40 000 genes have been assigned locations on the human genome.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

A genetic linkage map comprises the locations and relative distances between genes and other unique DNA sequences on linear maps corresponding to different chromosomes. Genetic linkage maps are a fundamental tool in modern biological research. They enable identification, cloning and sequencing of new genes by localizing the area of the genome in which the gene is located. They also allow identification of conserved regions between genomes of different species, which can be used to identify previously undetected genes. They have been used extensively in the study of human genetic disease and economically important agricultural species (plants and livestock). High-resolution maps of the entire human genome have been constructed and are being used as the framework on which to place the complete DNA sequence at their correct positions on the genome as part of the HGP. It is expected that 95% of the human DNA sequence will have been measured by the middle of the year 2000.

The first genetic maps were constructed in the early part of this century by observing the inheritance of different traits in successive generations of natural populations or artificial crosses. More recently, this technique of ''genetic mapping'' has been superseded by the technique of ''RH mapping'', which enables many more markers to be mapped faster and using fewer resources. Many of the assumptions on which RH mapping is based derive from conclusions from genetic mapping. The most important of these are that genes are located relative to each other on linear maps, and the distances between different pairs of genes can be determined from experimental observations. The key experiments that led to these conclusions are now reviewed.

1.1 Experimental Evidence for Linear Genetic Maps

Mendel first demonstrated that individual traits of appearance could be inherited. He treated these traits (genes) as being independent of each other, and predicted the frequencies of different combinations of traits that should be observed given their independent inheritance. It was soon discovered that, rather than being independent, the genes occurred in a relatively small number of tightly linked groups. This was deduced by Bateson and Punnett, who observed significant deviations from the distributions expected from independent inheritance. Certain pairs of genes were inherited from either one parent or the other more frequently than expected. Morgan interpreted these ''linkage groups'' as corresponding to the physical chromosomes; genes on the same chromosome would be inherited more often together than independently. Drosophila was observed to have 4 linkage groups, maize 10, yeast 17, mouse 20, and human 23.

Within each linkage group, the strength of linkage between pairs of genes was observed to vary. It was quantified by observing how frequently a pair of genes were inherited from just one parent. The phenomenon of ''crossing-over'' of genetic material had been discovered to be the mechanism for sharing genes from both parents to the offspring, and was identified with the physical structures known as chiasmata. These occur during meiosis between the paired chromosomes of the parents, and function to exchange, or ''recombine'', genetic material between homologous chromosomes from the two parents, thereby achieving some mixing of genes on the corresponding chromosome in the offspring. The further apart are two genes, the more likely it is they have been exchanged by a crossover, and the weaker the observed linkage. The degree of linkage between a pair of genes was quantified by the ''recombination frequency'', θ, the fraction of offspring that contain one gene from one of the parents, and the second gene from the other parent. A genetic map unit (m.u.) was defined as that distance between genes for which the 1 product of meiosis out of 100 is recombinant, $\theta_{AB} = 0.01$. The map unit is also called a centimorgan (cM) which corresponds to a physical molecular distance of approximately 1 Mb (megabase or 10^6 base pairs) in humans. That is, 1 crossover occurs on average every 100 Mb in the human genome (see Analytical methods). Therefore, a measure of the association between two genes could be used as a measure of the physical distance separating them on a linear chromosome. It enabled estimates of physical distances between genes from observed measures of association.

1.2 Use of the Recombination Frequency to Construct Linear Maps

Sturtevant, a student of Morgan, examined whether linear maps could be constructed using the recombination frequency directly as a measure of distance between genes on the same chromosome. He found that for small values of θ, genes could be ordered on a linear map. For example, if three genes A, B, and C had recombination frequencies $\theta_{AB} = 0.01$, $\theta_{BC} = 0.03$, and $\theta_{AC} = 0.04$, the correct order and distances between adjacent markers is A–(0.01)–B–(0.03) C. However, for larger values of θ, it was observed that it underestimated distances, and genes could no longer be arranged linearly. For example, if $\theta_{AB} = \theta_{BC} = 0.2$, then θ_{AC} would typically be $\sim$0.3, and the markers could no longer be accurately represented on a linear map. θ reaches a maximum value of 0.5 when the genes are completely unlinked on separate chromosomes.

The nonlinearity of the recombination frequency for larger values was explained by the occurrence of more than one crossover in larger intervals. A double

crossover would be counted as a nonrecombinant, and the recombination frequency would therefore underestimate true physical distance. The relationship between recombination frequency and physical distance is examined further in Analytical methods. These discoveries made in genetic mapping with respect to the relationship between the observed association and physical distance were of fundamental importance for the interpretation of results from RH mapping.

Although many genes have been mapped using classical genetic mapping, a fundamental limitation of the technique is the requirement that the gene should be polymorphic, i.e. show measurably different forms such as in eye color or flower shape. Many genes in the genome are thought not to show detectably variant forms, and therefore cannot be located on the genome using genetic mapping.

1.3 Somatic Cell (Goss–Harris) Mapping

An entirely different method for determining linkage between genes and the corresponding distances between them was demonstrated in 1975 by Goss and Harris.[1] Rather than relying on observations of naturally occurring crossovers to quantify linkage between genes, they physically disrupted human chromosomes by irradiating whole cells with X-rays. Linkage could then be detected by how much the genes were randomized in the process. The result of irradiation was a series of chromosomal fragments, whose size was dependent on the X-ray dose; larger doses caused more fragmentation and smaller fragments. These fragments were then fused with unirradiated hamster cells to form so-called ''somatic cell hybrids'', i.e. nongermline cells which contained DNA from two different species. A number of individual hybrid cells were cultivated, each hybrid containing an independent and random sample of the fragments. In the original experiment, the initial cells of each hybrid colony were selected if they retained human fragments expressing the enzyme HGRT (hypoxanthine guanine phosphoribosyl transterase), by growing on HAT (hypoxanthine, aminopterin, thymidine) medium. This complemented the lack of HGRT, which is located on the long arm of the human X chromosome. Hamster cells that did not retain chromosomal fragments containing HGRT did not survive the selective medium. The frequency with which three other human genes on the X chromosome were co-transferred with the selected marker was measured, and interpreted in terms of how strongly ''linked'' they were. The data showed that one of the genes (PGK) segregates from the selected gene (HGRT) more frequently than do the other two genes (α-gal and G6PD). This was interpreted as due to PGK being further apart from HGRT than are the other two genes, since the closer that two markers are together, the less likely it is that they will be separated by X-rays. A quantitative treatment was developed[1,2] which allowed distance, in units of the mean number of breaks that occur between pairs of markers, to be estimated directly from the observed frequency of coretention in the hybrid panel. From these observations, Goss and Harris were able to determine the order and relative distances between the four genes.

In a second experiment, Goss and Harris[3] constructed a human–mouse panel, which retained fragments even if they did not contain a selected gene such as HGRT. The analysis became more involved, since a hybrid typed positive for two genes could be due to their dependent retention on the same fragment or independent retention of both genes on different fragments. Methods were developed by which the observed distribution of the number of different hybrid types could be used to estimate distances between pairs of markers. Mapping of any region of the genome could be accomplished with this panel, since the restriction of having a selectable marker had been removed. Their quantitative treatment of the observations is described in Analytical methods.

1.4 Radiation Hybrid Mapping

The method was underused until its revival in 1989/90 by Cox et al.[4,5] This was attributed to a lack of sufficient markers to exploit the technology.[6] Cox et al.[4] constructed a human–hamster hybrid panel which also contained human fragments without the need for specific selection. Their approach differed slightly from Goss and Harris's nonselective method by using a rodent–human somatic cell hybrid that contained a single chromosome as the donor cell, rather than a diploid human cell. The advantage of this technique was that all fragments retained in the RH were known to originate from a single known human chromosome. They demonstrated the technique by localizing a 10–20 Mb region on chromosome 4 in the region of the Huntington Disease gene.[4] It was subsequently used to map 20 genes on the proximal half of chromosome 21[5] and 21 genes on the distal half.[7] The authors also generated expressions relating retention frequencies to physical distances, along similar lines to Goss and Harris.[1] These are described in Analytical methods. Some differences between genetic mapping and RH (Goss–Harris) mapping are shown in Table 1.

1.5 Whole-genome Radiation Hybrid Mapping

Walter et al.[8] returned to the original methods of Goss and Harris, by using as a donor cell a diploid human cell (fibroblast) rather than a human–rodent hybrid cell

Table 1 Differences between genetic and RH mapping

Property	Genetic	RH
Measured variable	Allele	+/−
Means of segregation	Chiasmata	X-rays
Meaning of θ	''Recombination frequency''	''Breakage frequency''
What is sampled	No. of recombinant offspring	Coretention of marker pairs
Experimental control	No	Yes

containing a single human chromosome. This enabled mapping of all regions of the human genome, not just a single chromosome, using a single RH panel.

1.6 Growth of the Field

Following the publication of the 1990 Cox paper,[6] there was enormous renewed interest in the technique, stimulating many new studies and the development of many new analytical techniques. This was partly due to the large number of novel markers that had been generated in the early 1990s for the construction of high-resolution genetic maps of the human[9] and mouse[10] genomes. These unique DNA sequences were readily available for independent mapping using the complementary technique of RH mapping. Since the underlying order and physical distances between markers on chromosomes is thought to be largely constant, comparison of the maps generated by genetic mapping and RH mapping would reveal important differences between the techniques. An example would be the identification of regions of the genome that have a higher occurrence of crossovers (recombination ''hot-spots'') which would inflate distances on the genetic map with respect to the corresponding region on the RH map. Such ''hot-spots'' have often been implicated in disease. These studies continue as data accumulates.

The considerable theoretical background developed for genetic mapping was also used with some modifications in the development of new techniques for analyzing results from RH experiments. Therefore, although the physical processes underlying RH mapping are fundamentally different from those in genetic mapping, the analytical techniques used rely heavily on previous results from genetic mapping work.

2 CONSTRUCTION OF A RADIATION HYBRID PANEL IN SILICO

The processes occurring during construction of an RH panel can be visualized using Monte Carlo simulation techniques. We follow the steps of the experiment in silico as they are thought to occur in vitro.

1. We assume a 1–1 relationship between an irradiated cell and the recipient hybrid cell that will ''retain'' the fragments of the donor chromosomes. For a panel comprising 50 hybrids, we assume that each of the 50 recipient cells ''retains'' the fragments resulting from irradiation of 1 of 50 donor cells. The 50 donor chromosomes are represented as vertical lines in Figure 1(a).
2. Each of the donor chromosomes is independently irradiated, causing the donor chromosome to break at random and independent locations. The number of breaks occurring on a chromosome is proportional to the X-ray dose and the length of the chromosome. We chose a chromosome of arbitrary length 5, meaning that on average there are five breaks per chromosome. The actual number of breaks is chosen from a Poisson distribution with mean and standard deviation equal to 5, and this number of breaks is placed uniformly randomly on the chromosome. The positions of the breaks on each chromosome are represented as small horizontal lines.
3. Each recipient cell ''retains'' a certain fraction of the fragments of its donor chromosome by randomly selecting them from the fragment pool generated by breakage. This fraction is chosen to correspond to experimentally observed marker retention frequencies. Here we choose the retention frequency to 0.5, which maximizes the information provided by each marker. Retained fragments are drawn in black in Figure 1(a); fragments not retained are drawn in gray.
4. The positions of 50 randomly placed markers are shown at the right-hand edge of Figure 1(a).
5. Each of these markers is ''typed'' on each of the recipient hybrid cells for presence or absence, and the result is shown in Figure 1(b), in which each row corresponds to one marker, and each column is one recipient hybrid cell. Black indicates presence of a marker in a hybrid cell, and white absence.
6. In general, marker order is not known, so the order of markers in Figure 1(b) is randomized in Figure 1(c) (the rows are scrambled). Figure 1(c) is a typical starting point for construction of a chromosome map using RHs, with the rows in arbitrary order, i.e. the markers are in no particular order.

The goal of an RH panel, starting from Figure 1(c), is to determine the correct order of markers (Figure 1c → Figure 1b). Then the correct distances between adjacent markers should be obtained (Figure 1b → Figure 1a).

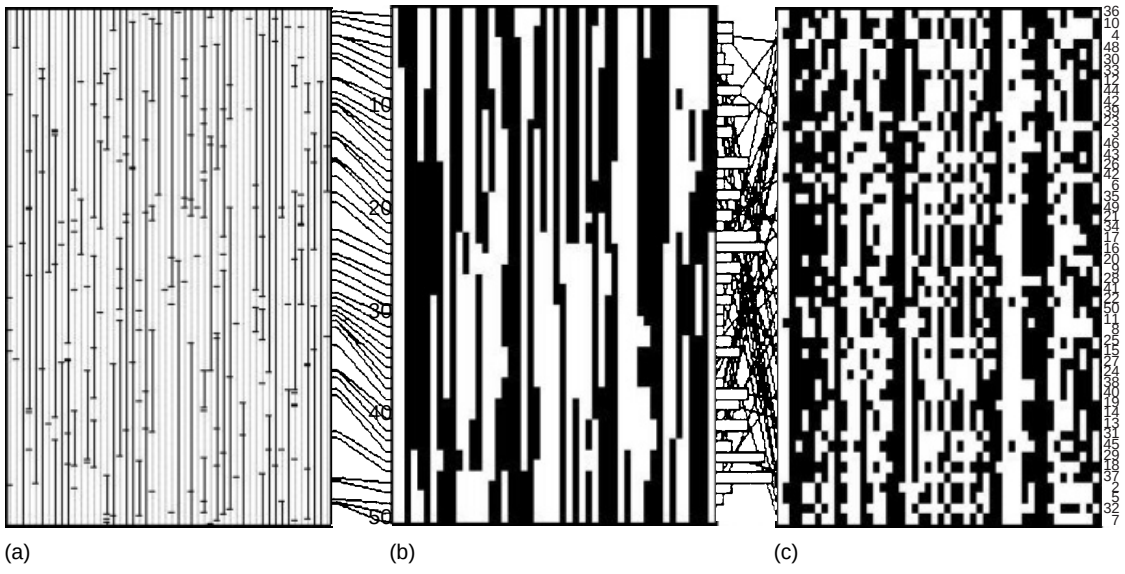

Figure 1 Simulation of an RH experiment. (a) Individual chromosomes are fragmented with X-rays, average of five breaks per chromosome, defining its length as 5R (5 Rays where 1 Ray is the distance interval in which 1 break occurs on average; 5 Rays is the distance in which 5 breaks occur on average). Positions of breaks are indicated on each column (chromosome) by small horizontal lines. Fragments are retained (black lines) or lost to the medium (gray lines). (b) Fifty markers are placed randomly on the chromosomes (at the same positions on each chromosome). These positions are indicated by short horizontal lines at the right-hand edge of (a). The lines joining (a) to (b) remove the interlocus distance information. The actual positions of the markers are on the left, their order on the right. The histogram to the right of (b) shows the number of chromosome breaks that have occurred between adjacent loci. Large numbers of breaks are clearly reflected as horizontal discontinuities in the marker retention pattern. (c) The order of the markers is randomized, and it is usually in this form that raw data exist before any mapping. There are 89 obligate breaks in the ordered set, and 759 in the randomized set.

3 ANALYTICAL METHODS

The process of constructing a map from experimental data involves two steps: detection and quantification of linkage in terms of physical distance between pairs of markers, and subsequently placing the markers on a linear map which best explains the physical distances. Many different methods have been developed, typically first for meiotic linkage mapping, since it has been in practice for much longer than RH mapping, then subsequently applied with modifications to RH data.

The experimental data from both meiotic mapping and RH mapping is used initially to determine whether there is any association between pairs of markers. Association can be quantified in two respects: the strength of the association; and its significance. The strength of association is greater for markers closer to each other, and can therefore be used as a measure of physical distance. It is measured as θ, the recombination frequency in genetic mapping or the breakage frequency in RH mapping, between a pair of markers. The significance of the association measures how likely the observed association would be to arise by chance, and is used to attach some ''confidence'' to the associated strength, and is usually expressed as a log-of-odds (lod) score, denoted $Z(\theta)$, which is the log of the likelihood of obtaining the observed value of θ if the markers were not linked. The significance of an observation increases with the number of samples taken (e.g. number of offspring in a genetic cross or number of hybrids in a RH panel), and allows weaker associations to be measured with more confidence that they are not due only to random fluctuations.

3.1 Strength of Linkage in Meiotic Maps

Results from genetic crosses indicated that certain pairs of genes appeared to be inherited together more frequently than would be expected if their inheritance were independent. The degree of linkage (or its complement, random inheritance) was quantified by counting the number of offspring that inherited genes from different parents (recombinants). These arise from the occurrence of ''crossovers'' during meiosis, which exchange genetic material between the parental chromosomes. Analysis of the structure of crossovers, which indicated a strictly reciprocal exchange of genetic material, shows that if two markers are separated by 1 *or more* crossovers, the frequency of recombinants in the offspring will be 0.5. If two markers are not separated by a crossover, the frequency of recombinants (recombination frequency) is 0.0. Intermediate values of the recombination frequency

are then determined by the probability that 1 or more crossovers have occurred between markers. The reciprocal exchange of genetic material from the two parents at each crossover enables θ to be estimated directly from the number of recombinant offspring.

Haldane[11] assumed that crossovers occurred randomly and independently of each other. The number of crossovers in a given interval is then given by the Poisson distribution. If the rate of occurrence of crossovers in a unit interval of distance is λ, then the probability of observing n crossovers in a given interval with length D is given by Equation (1)

$$p_n = \frac{1}{n!}\lambda^n e^{-\lambda D} \tag{1}$$

In particular, the probability that no crossovers have occurred between two markers separated by a distance D is given by Equation (2)

$$p_0 = e^{-\lambda D} \tag{2}$$

and its complement, the probability that one or more crossovers have occurred in the interval D, is given by Equation (3)

$$p_{>0} = 1 - e^{-\lambda D} \tag{3}$$

Since for 1 or more crossovers, the recombination frequency θ is always 0.5, it can be expressed by Equation (4) in terms of physical distance as

$$\theta = 0.5\left(1 - e^{-\lambda D}\right) \tag{4}$$

from which distance is obtained from the observed recombination frequency as shown in Equation (5)

$$D = -\frac{\ln(1 - 2\theta)}{\lambda} \tag{5}$$

We now have expressions relating the experimental observation (θ) to the underlying physical cause (the distance between the markers), by assuming a model in which crossovers occur randomly and independently of each other. The relation above between physical distance and θ is known as the ''Haldane mapping function''. The units of distance are the mean number of crossovers which have occurred between the markers and is not bounded, whereas θ is bounded in the interval $[0, 0.5]$, and is a good measure of distance only for small values. It is now known that crossovers occur on average every 100 Mb in the human genome, and every 200 Mb in the mouse genome, giving $\lambda = 0.01\,\text{Mb}^{-1}$ in human, and $\lambda = 0.005\,\text{Mb}^{-1}$ in mouse.

The two domains of θ and D are different. θ is a frequency, is dimensionless, and lies in the range $[0, 0.5]$. D is a distance measure, has units of mean number of crossovers, and is not bounded. Thus θ is not additive, whereas D is, and in general θ should not be used as a measure of distance, but first converted to the additive distance measure using Haldane's mapping function.

The assumption that crossovers occur randomly and independently is now known not to be strictly true, since the occurrence of a crossover at one location can inhibit formation of crossovers in its vicinity, a phenomenon known as ''interference''. Therefore physical distances between close markers are underestimated. Also, some regions of the genome are known to be more susceptible to crossovers. Recombination frequencies in these ''recombination hotspots'' overestimate true physical distance. However, the basic assumption does provide a means of relating recombination frequency to physical distance which is valid over a large range of distances, and relates directly to subsequent analytical methods developed for RH mapping.

3.2 Strength of Linkage in Radiation Hybrid Experiments

The RH experiment is another means of detecting ''linkage'' between genes. Here, the strength of linkage is detected by measuring the degree of disruption between genes induced by controlled radiation-induced chromosome breaks, and counting the number of ''offspring'', in this case human–rodent hybrid cells which retain a fraction of the resulting chromosomal fragments. The principle is that the closer two markers are together on the genome, the less likely it is that they will be separated by X-rays.[1] Therefore markers close to each other show highly correlated patterns of retention across the cells of the hybrid panel. Again, breaks are assumed to occur randomly and independently. This assumption appears to be valid in the case of RH mapping, although there are some reports that certain regions of the genome are more susceptible to breakage than others. However, we will not consider this. In contrast to determining which parent the genetic material in the offspring originates from, the experiment is to determine *presence* or *absence* of pairs of markers in a number of independent clones which retain fragments. The average retention frequency of a marker across the panel is the frequency of hybrids retaining the marker, lies in the range $[0, 1]$, and differs from panel to panel. This is in contrast to genetic mapping, in which the probability of retaining a gene from one particular parent is always 0.5. This adds an additional level of complexity to the analysis of an RH experiment.

We define formally here the ''breakage frequency'' as the probability that two markers have been separated by radiation, i.e. the probability that 1 or more breaks have occurred between them. If the physical distance between them is D and breaks occur at a rate λ, the breakage

frequency θ is determined by the Poisson distribution, and is given by Equation (6)

$$\theta = 1 - e^{-\lambda D} \tag{6}$$

from which Equation (7) gives the distance obtained

$$D = -\frac{\ln(1-\theta)}{\lambda} \tag{7}$$

The normalizing rate factor λ enables conversion of distances in terms of mean number of breaks to other units, e.g. Mb. As in meiotic mapping, the two domains of θ and D are different. Here, θ is the probability, lying in the range $[0,1]$, that a pair of markers have been separated by a break, whereas D is the number of breaks that occur on average between them, and its range is not bounded.

The problem then is to estimate θ between two markers from the results of an RH experiment, which comprise the numbers of hybrids typed AB, Ab, aB, and ab for all pairs of markers A and B (Figure 2). When markers are so close that they are never broken apart, A and B are always found on the same fragment which can either be retained by a hybrid cell, which will then be typed AB, or not retained (hybrid is typed ab), that is $p_{AB} = p_A = p_B$, and only concordant classes AB and ab are observed. If the breakage frequency is 1, A and B are always on different fragments, and assuming that fragments are retained independently, the retention of markers A and B will also be independent, i.e. $p_{AB} = p_A p_B$. For intermediate values of θ, the numbers of hybrids typed AB arise from dependent retention of A and B on the same fragment and independent retention on different fragments (Figure 3). The estimate of θ is therefore not as simple as for meiotic recombination-based mapping.

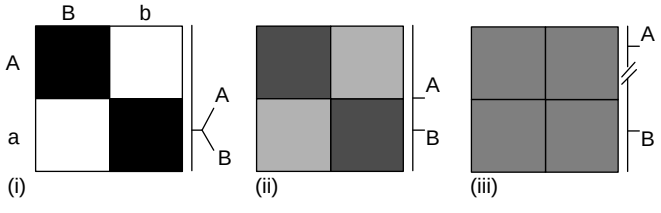

Figure 2 Relative frequencies of the types AB, Ab, aB, and ab expected for different values of θ (A,a–top, bottom row; B,b–left, right column): (i) $\theta = 0$; (ii) $\theta = 0.5$; (iii) $\theta = 1.0$. The retention frequencies of A and B are both 0.5. In (i), A and B are never broken apart, and markers are either retained or not retained on the same fragment, and only the concordant classes AB and ab are seen. In (iii), markers A and B are always on different fragments, retention of A and B is independent, and the frequency of all four classes is equal. Intermediate values of θ are apparent as a surplus of concordant types over discordant types.

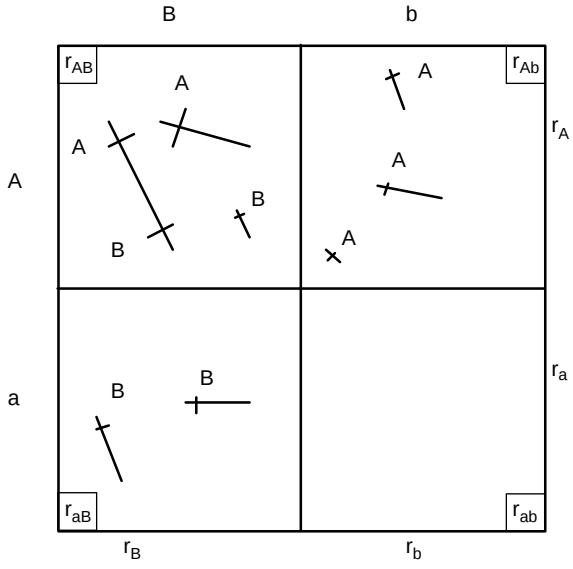

Figure 3 Composition of fragments in the four hybrid types AB, Ab, aB and ab. Type AB contains fragments containing A and B together, and fragments which contain either A or B. Type Ab contains fragments which contain only A, and type aB contains fragments which contain only B. The retention frequency of A is the sum of the frequencies of the classes AB and Ab. The retention frequency of B is the sum of the frequencies of the classes AB and aB.

In the following section it is useful to remember that the experimental observations, the numbers of hybrids typed AB, Ab, aB and ab for each pair of markers, are the result of two separate processes: (1) the fragmentation of chromosomes by X-rays; and (2) random retention of a subset of these fragments in a series of independent hybrid cells, which are then subsequently typed.

3.3 Estimation of θ by Goss and Harris (1975)

In the original Goss–Harris radiation-fusion experiment,[1] the co-retention of markers with one selected marker was measured by counting their co-occurrence in a series of independent samples (human–hamster hybrids). If the marker was separated from the selected marker by a "segregation event", i.e. chromosomal break, it was lost from the hybrid. Then θ is obtained directly from the frequency with which the selected marker occurred without the second marker as shown in Equation (8)

$$\theta = \frac{\mathrm{Ab}}{N} \tag{8}$$

The distances of all markers from the selected marker were measured as $D = -\ln(1-\theta)$, and the relative order of markers was inferred from these distances.

In a second a experiment,[3] the restriction that markers be linked to a selectable marker was removed, by using

human–mouse hybrids which were known to retain human chromosomal fragments in the absence of any selective pressure. This type of hybrid therefore allows mapping in any region of the human genome. The estimate of θ using this panel became more involved, since the observation of two markers in a single hybrid could be due to both dependent coretention on the same fragment and independent coretention on two different fragments. They examine the likely composition of fragments of different types (AB together, A and B on separate fragments) in a hybrid observed to contain both A and B. The frequency with which individual fragments of whichever type are retained in the hybrid was estimated from the data as Equation (9)

$$r = \frac{r_A + r_B}{2} \tag{9}$$

where r_A, r_B are the frequencies with which A and B are observed in the clones. The frequency of clones showing discordant expression of A and B (the frequency of clones expressing either A or B but not both) is then given by Equation (10)

$$r_{Ab} + r_{aB} = 2r(1-r)\theta \tag{10}$$

If A and B were originally on different chromosomes then $\theta = 1$, and Equation (11) gives

$$r_{\text{discordant}|\theta=1} = 2r(1-r) \tag{11}$$

It follows in Equation (12) that

$$\theta = \frac{r_{Ab} + r_{aB}}{r_{\text{discordant}|\theta=1}} \tag{12}$$

i.e.

$$\theta = \frac{r_{Ab} + r_{aB}}{2r(1-r)} \tag{13}$$

Equation (13) when r is expanded gives Equation (14)

$$\theta = \frac{r_{Ab} + r_{aB}}{r_A + r_B - (r_A r_B + (r_A^2 + r_B^2)/2)} \tag{14}$$

The authors express the relation between breakage frequency and distance as $D = -\ln(1-\theta)$ using the Poisson distribution, from the assumption of random and independent breaks. Central to this analysis is the estimate of frequencies of retention of different fragments, which is observed only indirectly through the retention of markers. The significance of linkage was measured from the χ^2 statistic with one degree of freedom.

3.4 Estimation of θ by Cox et al. (1990)

The Goss–Harris mapping technique was revisited and modified,[4,5] and the technique was named "RH mapping". Analytical expressions were introduced for estimating θ from the number of discordant clones, along similar lines to the formulation of Goss and Harris.[1–3] The authors also estimate the frequency of retention of different fragments. Their reasoning is as follows. The observed marker segregation could be defined in terms of four unknowns: θ, p_A, the probability of retaining a fragment containing A but not B, p_B, the probability of retaining a fragment containing B but not A, and p_{AB}, the probability of retaining a fragment that contains both A and B.

The fraction of all hybrids retaining marker A but not marker B is represented in Equation (15) in terms of these parameters as

$$\frac{A^+B^-}{N} = \theta p_A(1-p_B) \tag{15}$$

Similarly, Equations (16–18) give

$$\frac{A^-B^+}{N} = \theta(1-p_A)p_B \tag{16}$$

$$\frac{A^+B^+}{N} = (1-\theta)p_{AB} + \theta p_A p_B \tag{17}$$

$$\frac{A^-B^-}{N} = (1-\theta)(1-p_{AB}) + \theta(1-p_A)(1-p_B) \tag{18}$$

Using just the expressions for Ab and aB gives Equation (19)

$$\theta = \frac{p_{Ab} + p_{aB}}{p_A + p_B - 2p_A p_B} \tag{19}$$

But as the authors state, it is not possible to estimate p_A or p_B from the observed data, since the experiment types hybrids for presence or absence of markers, not for presence or absence of *fragments* of certain types (AB, A, or B). They suggested using the observed frequencies of markers A and B (r_A, r_B) as estimates of the unmeasurable quantities p_A and p_B. The expression for θ can then be given by Equation (20).

$$\theta = \frac{r_{Ab} + r_{aB}}{r_A + r_B - 2r_A r_B} \tag{20}$$

The use of r_A as an estimate of p_A was justified by observing that these numbers, when substituted into Equations (15) and (16), gave ratios of the discordant types Ab/aB that were not significantly different from the observed ratios. But r_A is the total retention of marker A whether or not it is on a separate fragment to B, and therefore overestimates p_A by the frequency with which A occurs together with B on the same fragment

(Figure 3). p_B is similarly overestimated. Therefore the second term of the right-hand side of Equation (15) over-contributes to the estimate of Ab, and the third term under-contributes. These effects might therefore cancel each other out, giving the observed approximate agreement with the data. However, their final expression can be derived by other means, and it can be seen that it is identical to that of Goss and Harris, if the equality $r_A r_B = (r_A^2 + r_B^2)/2$ holds. This is the case when $r_A = r_B$, since Goss and Harris estimated only one fragment retention frequency as the average of the retention frequencies of the two markers.

Cox et al. also use the assumption of random and independent occurrence of chromosome breaks to relate θ to physical distance, $D = -\ln(1-\theta)$, i.e. the mapping function derived above, but without the ''rate'' parameter λ. The units of distance are therefore ''mean number of breaks''. They define a new unit of distance, the centiray (cR) which is equivalent to $D = 0.01$, by analogy with the centimorgan in meiotic mapping. The rate parameter is introduced as a subscript corresponding to the dose; $1\,cR_{3000} = 0.01$ breaks at a dose of 3000 rad, $1\,cR_{8000} = 0.01$ breaks at a dose of 8000 rad. The use of the mapping function converts breakage frequency to number of breaks, but requires some calibration for panels constructed with different X-ray doses.

The significance of the estimated breakage frequency was expressed as the likelihood of obtaining the observed data, given the model of fragment retention:

$$L_\theta = [(1-\theta)r_{AB} + \theta r_A r_B]^{AB}[\theta r_A r_b]^{Ab} \times [\theta r_a r_B]^{aB}[(1-\theta)r_{ab} + \theta r_a r_b]^{ab}$$

where r_{AB} is an estimate of p_{AB}, the probability of retaining a fragment containing A and B together, estimated as $r_{AB} = \{[(A^+B^+)/T] - \theta r_A r_B\}/(1-\theta)$. This was compared with the likelihood expected if θ were 1 (markers completely unlinked) to give a likelihood ratio of linkage versus nonlinkage.

3.5 Estimate of θ Directly from the Observed Marker Retention

We tried a different approach, choosing not to model the unobserved probability of retaining different types of fragment, but to use only the observed data, i.e. the frequencies of retention and coretention of markers to estimate the expected frequencies of the discordant hybrid types Ab and aB.[12] We reasoned as follows, using only the two most basic assumptions of the RH experiment: random and independent occurrence of breaks, and random and independent retention of the resulting fragments.

For a hybrid typed Ab, marker A is necessarily on a different fragment to B. The probability of retaining a fragment that contains just A is the product of the probabilities of two independent events, retention of A and retention of A separate from B. The probability of retaining A is estimated from the data as r_A. The probability that the retained fragment contains A and not B is θ, since the fraction of all fragments containing A but not B is θ, by definition of θ as the frequency with which A is separated from B. The probability of retaining A on a separate fragment to B is therefore θr_A. The probability of not retaining B in any single hybrid in the panel is estimated from the observations as $1 - r_B = r_b$. Assuming independent retention of fragments, the expected frequency of the hybrids typed Ab is $p_{Ab} = \theta r_A r_b$, and similarly, $p_{aB} = \theta r_a r_B$. The total expected frequency of the discordant classes is then $\theta(r_A r_b + r_a r_B)$. Equating the expected frequencies with the observed frequencies r_{Ab} and r_{aB} gives an estimate of θ, as shown in Equation (21), which uses the frequency of the discordant types (r_{Ab}, r_{aB}), and individual marker retention frequencies (r_A, r_a, r_B, r_b),

$$\theta = \frac{r_{Ab} + r_{aB}}{r_A r_b + r_a r_B} \tag{21}$$

Here we observe that substituting $r_a = 1 - r_A$ and similarly for B in Equation (21) we obtain Cox's expression (Equation 20), but obtained using only the observed frequencies and not requiring any substitution of unobservable frequencies. It needs no model of fragment retention other than that fragments are retained randomly and independently. A more detailed formulation based on information theory, and using additionally the frequencies of the concordant types, was described,[12] and mapping tools based on these measures are available on the Web.

Distance was obtained from θ using the mapping function $D = -\ln(1-\theta)$, assuming that breaks occur randomly and independently. The variance of the distance estimate was calculated from the law of errors of functions[13] as Equation (22)

$$\sigma^2(D) = \left\{\frac{\partial}{\partial\theta}[-\ln(1-\theta)]\right\}^2 \sigma^2(\theta) \tag{22}$$

which simplifies to Equation (23)

$$\sigma^2(D) = \frac{\theta}{(1-\theta)N} \tag{23}$$

The variance is 0 when $\theta = 0$, $D = 0$, and ∞ when $\theta = 1$, $D = \infty$, as expected. This expression for the variance is useful for other analytical procedures such as weighted least squares fitting.

If the denominator of Equation (21) is replaced by the expression of Goss and Harris (Equation 11), for their estimate of the frequency of retention of *fragments*

containing A or B, we get Goss and Harris's expression for the breakage frequency. However, this requires that the retention frequencies of A and B are equal, which is not generally true.

The significance of linkage we measured as the probability of the χ^2 statistic for a 2×2 table

$$\chi^2 = \sum_{XY} \frac{(N_{xy} - n_{xy})^2}{n_{xy}}, \quad X \in \{A, a\},\ Y \in \{B, b\} \quad (24)$$

where N_{xy} are the observed numbers of the different classes, and $n_{xy} = N_x N_y / N$ are the expected numbers if the markers were retained independently. The statistic measures the probability that the observed numbers of the four hybrid types AB, Ab, aB, and ab could have been obtained if A and B were always retained independently on different fragments ($\theta = 1$). The probability of the observed χ^2 is obtained from the incomplete β function.[14] Small values for this probability indicate that it is unlikely that the observed distribution could be obtained by independent retention of A and B, and is therefore a likelihood ratio for nonlinkage versus linkage. Its negative logarithm is exactly equivalent to a lod score for linkage. We used the χ^2 statistic to obtain pairwise lod scores between markers.[12]

3.6 Estimate of θ Using Maximum Likelihood

The requirement for estimates of the frequencies of retention of different fragments containing A and B on the same fragment used by Cox et al.[5] prompted further work to estimate these frequencies by other means, since it now seemed that fragment retention had to be estimated in order to estimate θ. These other methods are now the basis for several software packages.

One means of estimating fragment retention frequencies was maximum likelihood (ML) used by Boehnke et al.[15] ML methods had previously been used in genetic mapping,[16–22] where map distances were optimized to maximize the probability of the observed data. They were again applied in the field of RH mapping.[15] Application of ML methods involves first constructing a model of the process assumed to be occurring, and estimate the values of the parameters by maximizing the likelihood of the data, given the model parameters. Parameters to ML models are breakage frequencies between all pairs of loci, and retention frequencies of different-sized fragments containing more or fewer markers on the same fragment depending on the fragment size. A breakage frequency θ_i is the probability that a break occurs between markers i and $i+1$. The retention probability p_{ij} is the probability that a fragment containing loci i, $i+1, \ldots,$ and j is retained in a RH. Trial interlocus distances are included in the calculation until a distribution of breaks across each chromosome is obtained that maximizes the likelihood of the observed distribution of markers across all hybrids in the panel. Likelihood methods therefore estimate distances between markers, and attach a likelihood to a given order of markers. ML methods require a means of scanning through the $N!/2$ possible orders to optimize the model. Boehnke et al. used Expectation maximization.[23]

The number of parameters that are adjusted is large: for N markers, there are $N(N-1)/2$ different probabilities of retaining each different pair of markers on the same fragment. This makes ML computations expensive. To reduce the computational burden, Boehnke et al.[15] restricted the ways in which fragment retention p_{ij} was parameterized in the model to a small number of classes. These were (1) the equal retention frequency model, which assumes that all fragments containing all different pairs of markers were retained with equal probability, (2) moving-average model, in which the retention probability is the average value of the retention frequencies of all markers on the fragment, (3) Markovian models, chosen for the availability of rapid computational treatments for Markov chains, (4) centromeric or telomeric model, in which a fragment containing the centromere is retained with one constant frequency which is higher than that for all other fragments that do not contain the centromere, and (5) the left-endpoint model, in which the retention frequency is chosen to be the retention frequency of the marker closest to the centromere. The last two models were introduced since it was observed that marker retention probabilities for the Cox et al.[4] 21q RH data decreased with increasing distance from the centromere.

The models therefore have the capacity for becoming extremely complex, and it is worth remembering here what is actually occurring. Retention frequencies of different fragments are chosen on the basis of their assumed location on the chromosome, and subsequently inter-marker distances are optimized to maximize the likelihood of the data. However, the normal case in mapping is that at the outset, there is no knowledge of chromosomal location. Determination of location is the aim of a mapping experiment, and yet assumed values are being used in ML methods prior to determining the location. Also, the restriction of fragment retention probabilities discards the experimentally observed frequencies of retention of individual markers. Despite these issues, ML methods form the basis of one of the most widely used software packages, RHMAP,[15,24] which has been used effectively in many mapping projects. Walter and Goodfellow[6] note that choice of different retention frequency models has little significant effect on the final maps generated by ML methods.

3.7 Methods for Determining Order

The methods described above are used to estimate distances between pairs of markers, but not the relative order of more than two markers. For N markers there are $N!/2$ possible orders, and the number of possibilities increases rapidly with the number of markers. Efficient means are needed for identifying the best order(s) from the vast number of possibilities.

One means of determining order, but with no estimate of distances between adjacent loci, or of the likelihood of the computed order, is to find the order which minimizes the obligate number of breaks. This can be seen in Figure 1. Looking down a column in the middle and right-hand panels, the number of obligate chromosome breaks implied by a particular order of markers is the number of times down a column when the hybrid type changes from present to absent or vice versa, since each such discontinuity must be due to a chromosomal break between the flanking markers. The number of breaks in the correct order in the middle panel is 89, and in the randomized order, 759. It therefore seemed natural to find an order which minimized the number of breaks, since a random order has more breaks than the true order. This method was used previously in genetic mapping[25] by permuting the order of markers to reduce the number of implied recombinants. Falk[26] and Boehnke et al.[15] describe the application of this method in RH mapping by permuting the linear order of markers. Thompson[25] observed that the method had the effect of moving spurious double recombinants to the ends of a linear map. A similar effect in the RH experiment is to move error-prone markers to the ends of the map. Since the number of different orders of N markers is large ($N!/2$), different methods of searching subsets of orders were suggested: stepwise locus ordering,[15] simulated annealing,[15,27] and branch-and-bound.[28] Only branch-and-bound is guaranteed to find the minimum. The statistical properties of the minimum breaks method were examined by Barrett.[29]

Another method of choosing likely marker orders from the vast number of possibilities was adopted in the program MultiMap/RADMAP.[30] This method uses CRIMAP[22] to calculate likelihoods for a given order, and a rule-based system using the AI language LISP to generate likely orders. The rules were constructed to replicate the process done manually by a human mapper, and were applied automatically to generate large numbers of likely orders. It increases the speed and efficiency of map construction. The best order chosen was the one with ML. Since the program reduced the time to generate maps, it also allowed exploration of different mapping approaches and heuristics used in map building. It was used in the construction of a map of human chromosome 13q containing 95 markers.[31]

A third method for determining order directly from a distance matrix using the techniques of distance geometry (DG) is discussed below (section 3.10).

3.8 Genetic Analysis Workshop 7

In 1992, the Genetic Analysis Workshop 7 focused on RH and somatic cell hybrid mapping of chromosome 21 and new computational methods.[32] This workshop demonstrated most of the methods currently used in the analysis of RH data, and provides a useful survey of the field. Nine groups analyzed the chromosome 21 RH data set using a variety of methods, mostly those recently developed by Boehnke et al.[15] Bishop and Crockford,[33] Boehnke,[34] and Weeks et al.[35] used different methods to find a marker order that minimized the number of obligate chromosome breaks. Bishop and Crockford[33] enumerated all possible orders of eight markers. Boehnke[34] used branch-and-bound to reduce the number of orders that need to be evaluated by eliminating groups of orders that had been shown to be suboptimal; Falk[36] used simulated annealing to the same end. Falk[36] also measured the coretention correlation (r) between all pairs of loci, and large values of the correlation indicated closeness of markers. The order of markers was permuted to maximize the sum of correlation between adjacent markers. The measure of distance was chosen to be $1 - r$. Weeks et al.[35] estimated the probabilities that pairs of markers be transferred to the hybrid on the same fragment. They then also iterated over different locus orders, using simulated annealing, to find the order that maximized these probabilities between adjacent markers.

Likelihood methods were also demonstrated, using the framework developed by Boehnke et al.[15] Chakravarti and Reefer[37] assumed that all markers were retained with equal frequency, and built locus orders one marker at a time, and obtained the same order as Cox et al.[5] Bishop and Crockford[33] used both equal retention and centromerically preferred retention, and found that distances were very similar for the two different models. Green used the likelihood method CRIMAP of Lander and Green[22] developed for the analysis of genetic mapping data, and obtained the same marker order as Cox et al.[5] Lawrence and Morton[38] used a multiple pairwise approach that had been used previously for genetic maps.[39] They also obtained Cox's marker order. Bayesian methods were used by Guerra et al.,[40] but it was found that the method was computationally very difficult, even for two or three loci.[32]

3.9 Hidden Markov Models

A novel application of the statistical models known as hidden Markov models (HMMs) was demonstrated

by Slonim et al.[41] HMMs were used to estimate the likelihood of a map, when there is some uncertainty about typing. Graph algorithms are used to find sparse but reliable initial maps. The method is resilient to typing error by assigning a priori false positive and negative rates, and allowing for such errors in the ML calculation. The method automatically detects markers that are likely to be error-prone.

3.10 Distance Methods

We adapted our previous method[42] for constructing genetic (meiotic) maps using DG to obtain maps from RH data.[12] This technique obtains coordinates from a measured distance matrix which gives distances on the derived map that optimally fit the observed distances. DG was developed in the field of nuclear magnetic resonance (NMR) to determine three-dimensional models of small proteins from observations of inter-proton distances obtained from NMR spectroscopy.[43–45] Applied to RH mapping, the procedure was as follows:

1. Obtain a full $N \times N$ distance matrix from the joint retention frequencies of all pairs of markers according to Equation (21), and their associated variance, Equation (23).
2. The distance matrix was first ''smoothed'' by replacing large imprecise distances with sums of smaller more precise distances. The variance estimate, Equation (23), was used as a measure of precision.
3. Obtain a ''metric matrix'' **G** from the smoothed distance matrix, which comprises the dot products of the vectors of each point to the origin[45] according to Equation (25)

$$g_{ij} = \frac{D_{i0}^2 + D_{j0}^2 - D_{ij}^2}{2} \tag{25}$$

where D_{i0}^2 is the squared distance of point i to the origin. This can be calculated without prior knowledge of the coordinates[43] according to Equation (26)

$$D_{i0}^2 = \frac{1}{N}\sum_{j=1}^{N} D_{ij}^2 - \frac{1}{N^2}\sum_{k>j=1}^{N} D_{jk}^2 \tag{26}$$

4. Coordinates are obtained for each of the points from the eigenvalues and eigenvectors of the metric matrix as Equation (27)

$$c_{ij} = \sqrt{\lambda_i} w_{ij} \tag{27}$$

where c_{ij} is the ith coordinate of the jth point, λ_i is the ith largest eigenvalue, and w_{ij} is the jth element of the corresponding eigenvector. Coordinates obtained in this way have been proved to be optimal[44] in the sense that the spatial configuration of points so obtained minimizes the rotation error against the original distance matrix.

The method was applied initially[42] to the chromosome 21 data,[5,7] and subsequently[12] to the RH data used to generate the human gene map.[46] It generates orders directly from the calculated coordinates in a method with complexity $O(N^3)$, i.e. the time taken for the computation to complete is proportional to N^3 where N is the number of markers being considered.

4 MAIN SOFTWARE AVAILABLE

Many of the methods described above have been incorporated into software packages that can either be downloaded from the Internet, or are implemented as methods on mapping World Wide Web (WWW) servers to which data can be submitted, and the results returned in a Web browser or by e-mail. The main packages available are listed below, with the methods that they use.

4.1 RHMAP

The minimum breaks method and the ML methods developed by Boehnke et al.[15] and further by Lange et al.,[24] have been incorporated into a package RHMAP. It is available at http://www.sph.umich.edu/group/statgen/software/.

4.2 RHMAPPER

RHMAPPER[41,47] is a software package for constructing RH maps, and was designed to map very large numbers of markers. It has other additional features: ML calculations using HMMs for analyzing diploid cells, such as those used in whole genomic mapping, and accommodation and detection of laboratory error. It is available at http://www-genome.wi.mit.edu/ftp/distribution/software/rhmapper/.

4.3 Z-RHMAPPER

Soderlund et al.[48] introduced modifications to RHMAPPER. The software is available at http://www.sanger.ac.uk/Users/cari/Zman.html.

4.4 SAMAPPER

Stewart et al.[49] used simulated annealing to order loci in the package SAMAPPER (http://www-shgc.stanford.edu/Mapping/SAMapper/). It first obtains lod scores between all pairs of markers. The sum of lod scores between adjacent markers is then maximized by permuting the linear

order of markers. The search through the $N!/2$ possible orders is directed by simulated annealing. Their model suggests that the solution with the maximum sum of lod scores is the most likely solution.

4.5 MultiMap/RADMAP

MultiMap/RADMAP[30] is located at http://linkage.rockefeller.edu/multimap/. It automates the generation of alternative but likely orders using the results of likelihood calculations at previous stages, using rules coded in the AI language LISP. The authors recommend its use for large-scale mapping projects (>20 markers), or for the construction of many smaller maps.

4.6 Distance Geometry Map

The distance methods developed for RH mapping[12,42] are available at http://www.oxmol.com/biolib/webmap.[50] There are three modes of use: (1) copy and paste RH data for a series of markers, which are mapped with respect to each other; (2) find the N nearest markers to a named marker, and construct a locally accurate map in the vicinity of the reference marker rather than a whole chromosome; (3) enter the RH data vector for a novel marker typed on a RH panel, retrieve the N nearest vectors to the marker vector from the specified RH panel, and map these markers relative to each other to build a locally accurate map including the novel marker.

5 PRACTICAL METHODS

Having reviewed the analytical processes involved in RH mapping, we now describe how data are obtained in the laboratory for subsequent analysis and determination of map location.

5.1 General

One common use of RH mapping is to determine the location of an unknown (previously unmapped) marker on the genome. This is typically done by typing the marker on a RH panel for its presence or absence in each cell in the panel, and comparing the retention pattern with those of previously typed markers to find the closest matching markers and hence chromosomal region. The other major use is to map a series of markers with respect to each other, which involves typing each marker against a panel and mapping them with respect to each other.

5.2 Obtaining a Radiation Hybrid Panel

RH panels are distributed by Research Genetics (http://www.resgen.com). There are three human panels (GB4, long range, low resolution; G3, mid range and resolution; TNG3, short range, high resolution), and single panels for mouse, rat, baboon, zebrafish and dog. For human mapping the choice of panel is determined by whether or not the approximate location of the marker is known previously. If not, it is best to use the long range, low resolution panel G4 to obtain an approximate location. If approximate location is known it is better to type on the higher resolution G3 or TNG4 panels, providing the marker density in the region is sufficiently high (see section 6).

5.3 Typing Markers on the Panel

''Typing'' a marker on a RH panel involves determining whether or not it is present in each of the individual component hybrids in the panel. Goss and Harris[1–3] used enzyme assays to determine presence or absence of genes. Southern blot hybridization was used in the later experiments of Cox et al.[5] Polymerase chain reaction (PCR) is now the preferred method, enabling small DNA samples to be detected in the hybrids. Several different protocols have been published on the WWW (Table 2). A typical RH typing gel from the Stanford Human Genome Center is shown in Figure 4. The presence of a marker in a hybrid is indicated by a PCR product which is visualized by agarose gel electrophoresis.

5.4 Finding Close Markers

The results of the PCR are entered into the computer either directly as a string of 0s (absence) or 1s (presence), or via a clickable ''typing frame'' such as the one in Figure 5. This second method of data entry reduces possible errors in data input. The retention pattern of the marker on the panel is compared against the typings of all other markers that have been typed previously on the panel. These are generally available in the public domain on various web servers. Close markers have similar patterns of retention on the panel. Several mapping servers are available on the WWW, to which individual marker typings, or sets of typings for more than one marker, can be submitted. The servers return lists of nearest neighboring markers. The nearest markers to ''marker_xyz'' from mouse are shown in Figure 6. The closest, RH74155, is on mouse chromosome 3.

5.5 The Map Location

The location of the new marker is then indicated by the known positions of the markers which it is close to. The

Table 2 WWW resources for RH mapping

Type	Name	Uniform resource locator
General	RH info page	http://linkage.rockefeller.edu/tara/rhmap
	UniGene	http://www.ncbi.nlm.nih.gov/UniGene
	OMIM	http://www.ncbi.nlm.nih.gov/Omim
	MedLine	http://www.ncbi.nlm.nih.gov/PubMed
	GeneMap	http://www.ncbi.nlm.nih.gov/genemap
Genome centers	WI	http://www-genome.wi.mit.edu
	SGHC	http://www-shgc.stanford.edu/
	SC	http://www.sanger.ac.uk
	WTCHG	http://www.well.ox.ac.uk/wEST/
Data	RHdb	http://www.ebi.ac.uk/RHdb
	WI	http://carbon.wi.mit.edu:8000/ftp/distribution/human_STS_releases/july97/rhmap/
	SHGC	ftp://shgc.stanford.edu/pub/hgmc/RH_data/
	Jax	http://www.jax.org/resources/documents/cmdata/rhmap/
Mapping servers	SC (GB4)	http://www.sanger.ac.uk/RHserver/
	SHGC	http://www-shgc.stanford.edu/RH/index.html
	WI	http://www.genome.wi.mit.edu/cgi-bin/mouse_rh/rhmap-auto/rhmapper.cgi
	HGMP	http://menu.hgmp.mrc.ac.uk/menu-bin/RHyME/RHyME.pl
	OML/GW	http://www.oxmol.com/biolib/webmap/
	OGENRI	http://www.otsuka.genome.ad.jp/menu/RH.html
	GeneMap	http://www.ncbi.nlm.nih.gov/genemap98
	Ratmap	http://ratmap.ims.u-tokyo.ac.jp/menu/RH.html
PCR protocols	SHGC	http://www-shgc.stanford.edu/Mapping/rh/procedure/rhassaynew.html
	RHdb	http://www.ebi.ac.uk/RHdb
Other organisms	Mouse	http://lena.jax.org/resources/documents/cmdata/
	Rat	http://ratmap.gen.gu.se http://www.nih.gov/niams/scientific/ratgbase http://www.ebi.ac.uk/RHdb/index.html http://goliath.ifrc.mcw.edu/LGR/research/rhp/ http://ratmap.ims.u-tokyo.ac.jp/
	Zebrafish	http://www.eb.tuebingen.mpg.de/abt.3/haffter_lab/rh_mapping.html http://dir.nichd.nih.gov/lmg/lmgdevb.htm
	Dog	http://www-recomgen.univ-rennes1.fr/doggy.html http://www-recomgen.univ-rennes1.fr/Dogs/outils.html
Panels	Research genetics	http://www.resgen.com

OMIM, online mendelian inheritance in man.

location of ''marker_xyz'' typed in Figure 5 is found to be close to a marker at 25.1 cM on the genetic map.

6 PROPERTIES OF RADIATION HYBRID PANELS

There are several RH panels available commercially, all of which differ in their means of construction, and therefore also in the properties they display. Here we discuss the variable properties of RH panels which must be considered for different mapping purposes.

6.1 Construction of Panels is under Experimental Control

Three parameters affect the results of a RH experiment: (1) the radiation dose used to construct the panel which determines the mean number of breaks in a unit physical interval (μ); (2) the density of typed markers in unit distance intervals (λ); (3) the number of hybrids in the panel, which determines the significance of any detected linkage and the precision of the distance estimates to other markers. The conditions of the RH experiment can be varied at will, and should be considered when constructing new hybrid panels, or choosing existing panels, for specific mapping projects. Specifically, if a whole-genome map of a previously unmapped organism is to be made and the number of markers is relatively low, then a lower dose should be used to construct the panel. If on the other hand a very high resolution map of a localized region of the genome characterized by a large number of unique markers is required, then a panel should be constructed with a very high dose, e.g. the human TNG3 panel.

6.2 Available Panels

Table 3 summarizes the main properties of the commercially available RH panels from Research Genetics

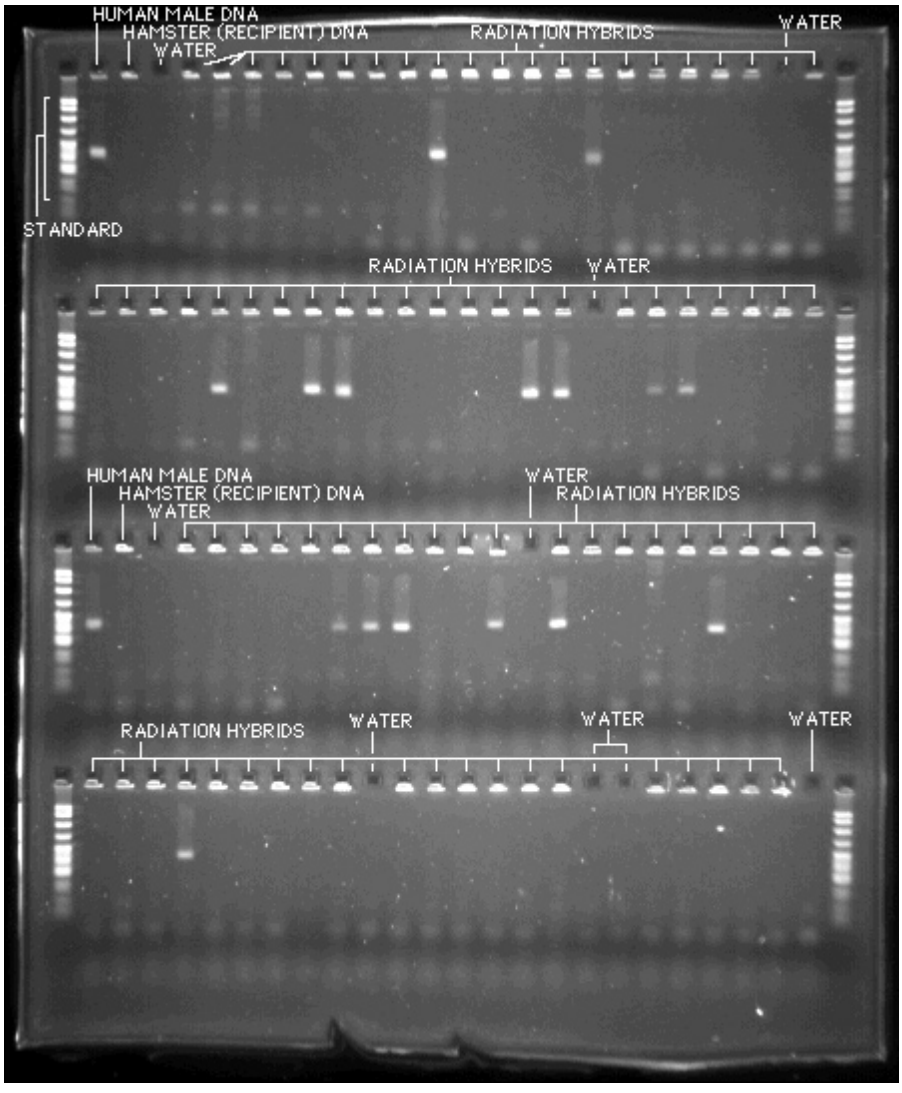

Figure 4 A typical RH typing gel from the Stanford Human Genome Center. Uniform resource locator: http://www-shgc.stanford.edu/Mapping/rh/procedure/image.html.

(http://www.resgen.com). The first panels to be constructed were for human (long-range G4,[51] medium-range G3,[49] short-range TNG3[52]). There are now panels available for mouse, rat, dog, zebrafish, and baboon.

6.3 Effect of Marker Retention Frequency

In order for a marker to be maximally informative about its coretention with another marker, it should be retained in 50% of the cells in the panel. The amount of information that a marker can provide about its linkage to other markers can be quantified by the statistical entropy, $H_{\mathrm{A}} = -\ {}_{X} p_x \log p_x$, $X \in \{\mathrm{A}, \mathrm{a}\}$. It is maximal when $p_{\mathrm{A}} = p_{\mathrm{a}} = 0.5$. The entropy is 0 for marker retention frequencies of 0 and 1.

6.4 Effect of Number of Hybrids on Resolution and Range

We define the resolution of a panel as the smallest breakage frequency that can be detected, and it is obtained when just one discordant hybrid is observed. As the number of hybrids in the panel is increased, the resolution increases. The range of a panel is the weakest linkage (largest value of the breakage frequency) that can be detected between adjacent markers with significance (e.g. with a lod score > 3). These values for simulated panels with the same radiation dose are summarized in Table 4. As the number of hybrids in the panel increases, so the resolution gets finer, and the range longer.

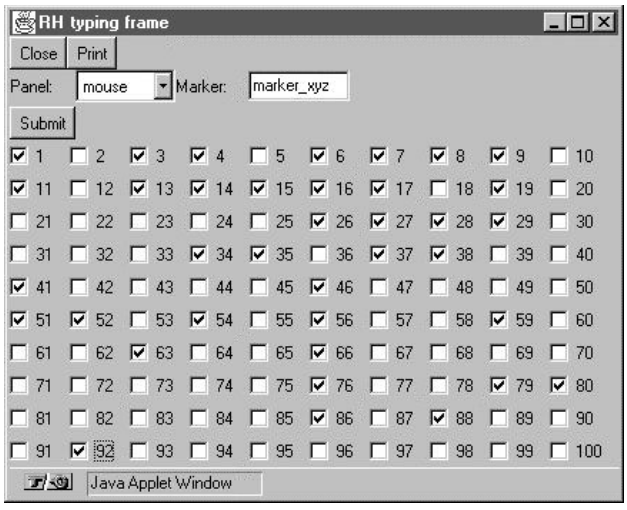

Figure 5 A ''typing frame'' originally developed for DG map at the BioLib site (http://www.oxmol.com/biolib/webmap). The user clicks the boxes which have been typed positive. ''Submit'' sends the data to the mapping server to find near neighbors.

6.5 Effect of Dose on Panel Range and Resolution

The X-ray dose used to fragment the donor chromosomes determines the size of the fragments generated: the higher the dose, the smaller the fragment size. Goss and Harris[1] determined empirically the relationship between rate of breakage and dose D as $D^{1.6}$. The exponent between 1 and 2 was interpreted as breakage involving one- and two-photon events. A compromise between two opposing properties of a panel must be struck when choosing the X-ray dose for construction of a particular panel.

Table 4 Predicted properties of RH panels with different numbers of hybrids. (Newell et al.[12])

Number of hybrids	Resolution			Range		
	θ	$Z(\theta)$	$D(R)$	θ	$Z(\theta)$	$D(R)$
10	0.156	2.01	0.170	0.40	1.24	0.51
20	0.080	4.30	0.084	0.40	3.83	0.51
50	0.033	11.0	0.034	0.48	3.63	0.65
100	0.017	22.0	0.017	0.64	3.50	1.02
200	0.009	43.8	0.009	0.76	3.16	1.43
500	0.004	109.2	0.004	0.85	3.17	1.88
1000	0.002	217.9	0.002	0.89	3.20	2.23

θ is the breakage frequency; $Z(\theta)$ is the lod score for the observed breakage frequency, where $Z(\theta) = -\log^{10}(P(\theta))$, and $P(\theta)$ is the probability that the observed value of chi-squared for a 2×2 (Equation 24) could have arisen by chance; $D(R)$ is the distance (in units of mean number of breaks) between a pair of markers corresponding to the breakage frequency θ, calculated from Equation (7), where $1 = 1$.

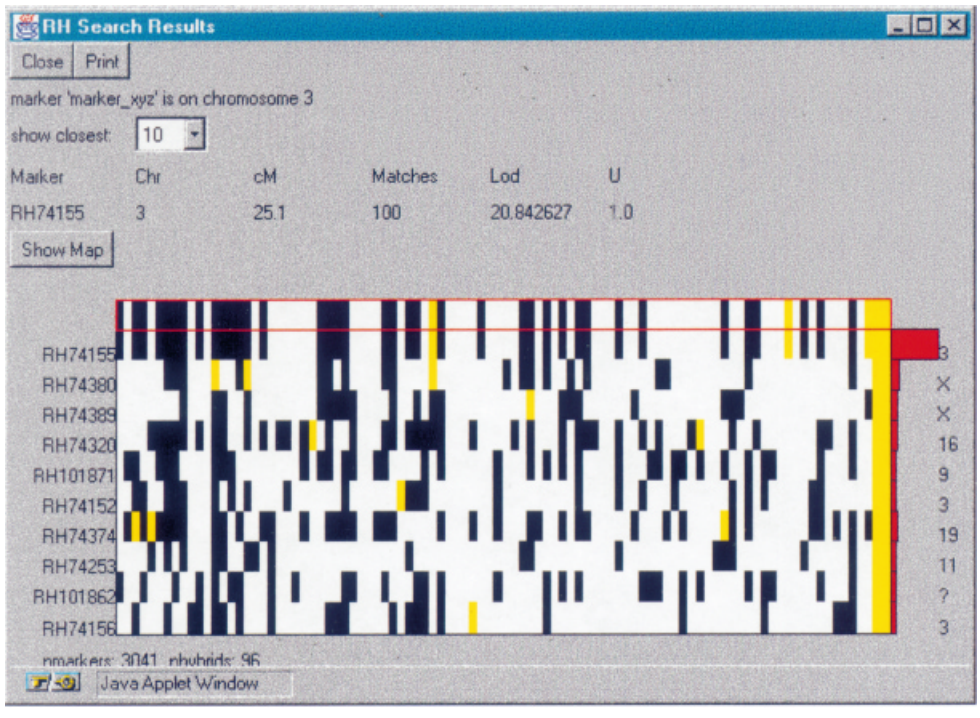

Figure 6 Results of the search for ''marker_xyz'' having the retention pattern on the mouse T51 panel shown in Figure 5, and boxed in red here. The nearest neighbors are ranked according to the number of type matches across the panel. The histogram to the right of the retention patterns is the lod score of linkage, calculated as the negative log of the χ^2 probability for the 2×2 joint retention frequency table. The chromosome to which the matching markers have been assigned is shown at the far right. The unknown marker is linked most strongly to RH74155 at 25.1 cM on chromosome 3 on the mouse genetic map with a lod score of 20.8.

A higher X-ray dose increases the potential resolution of the panel, since breaks occur closer to each other on average. Such panels can resolve markers that are closer to each other on a chromosome. But the significance of linkage over longer distances is severely reduced, since it is more likely that different markers segregate independently of each other, i.e. the range is shorter. This is manifest as ''holes'' in the map. Clusters of linked markers separated by holes cannot be ordered with respect to each other. This can be alleviated either by making an RH panel with a larger number of hybrids, increasing the number of observed concordant hybrids, or by typing a larger number of markers on the panel, thereby reducing the average distance between adjacent markers.

A lower dose reduces the resolution of the panel, since it is less likely that near markers will be separated by X-rays. However, longer distances between adjacent markers can be spanned.

An example of this is the G3 panel,[49] constructed with 8000 rad, which is useful for resolving close markers, but is not reliable for constructing whole genome maps, unless a very large number of markers is typed, and the GB4 panel,[51] made with a dose of 3000 rad, which is excellent for making maps of whole chromosomes, in which the distance between adjacent markers can be as large as 1 Mb. A new very high resolution panel, TNG4,[52] has been constructed with 50 000 rad, and will be used to construct ultra-high resolution maps of the human genome, possibly using single nucleotide polymorphisms as the mapping reagent.[53]

6.6 Effect of Marker Density on Significance of Linkage

The strength of a map can be defined as the average strength of linkage between adjacent (closest) markers on the map. This is determined by a combination of the mean number of breaks (μ) and the mean density of markers (λ) on the map. The mean density of breaks between adjacent markers is defined in terms of μ and λ as $\mu^* = \mu/\lambda$. The probability that adjacent markers are not disrupted by X-ray breaks is then $p_0 = e^{-\mu^*}$, and its complement is $\theta = 1 - e^{-\mu^*}$. Therefore, if the dose is high, or the density of markers is low, the linkage between adjacent markers is weaker and it is harder to build maps with confidence.

This was illustrated in the first human gene map,[46] which was constructed from the combined mapping on the GB4 and G3 panels. For the GB4 panel, $\mu_{GB4} = 0.04\,Mb^{-1}$, $\lambda_{GB4} = 0.36\,Mb^{-1}$, giving $\mu^*_{GB4} = 0.11$. The breakage frequency between adjacent markers on the GB4 panel is therefore $\theta = 1 - e^{-0.11} = 0.10$, which with an average retention frequency of 0.32 and 93 hybrids gives an average inter-marker lod score of 17.5. For the G3 panel, $\mu_{G3} = 0.27\,Mb^{-1}$, $\lambda_{G3} = 0.16\,Mb^{-1}$, giving $\mu^*_{G3} = 1.7$. The breakage frequency between adjacent markers on the G3 panel is therefore 0.82, with a lod score of 1.0. This very low average significance of linkage between adjacent markers on the G3 panels explains why the G3 map has many ''holes'', between which clusters of more strongly linked markers cannot be ordered with respect to each other. To avoid this problem on the even higher-resolution panel TNG3,[52] constructed with 50 000 rad, many new markers will have to be generated to maintain significant linkage across the panel with smaller chromosomal fragments.

7 APPLICATIONS OF RADIATION HYBRID MAPPING

The first applications of RH mapping were ordering of several markers in defined regions of the genome.[1–5] It was rapidly applied to data previously obtained from genetic studies of disease and of inheritance of polymorphic markers mapped by genetic linkage studies which had enabled construction of genetic maps of the human[9] and mouse[10] genomes.

7.1 Disease Gene Mapping

The technique of positional candidate cloning[54] involves first identifying an area implicated in disease, then identifying all genes that are thought to be located in a similar region. These genes are designated ''candidates'' due to their chromosomal location, and are categorized according to the likelihood that they might contribute to the disease phenotype. An example of the application of RH mapping to positional candidate cloning is the disease multiple endocrine neoplasia type 1, MEN-1.[55] The MEN-1 and bcl-1 disease loci were known to lie on chromosome 11. An RH panel was made from a hamster cell line fused with a hamster–human chromosome 11 hybrid cell line. Sixteen human genetic markers were mapped in this region, and ordered using the techniques of Cox et al.[5] This located the proto-oncogene SEA (S13 avian erythroblastosis oncogene homolog) between two other markers PYGM (muscle glycogen phosphorylase) and INT2 (murine mammary tumor virus integration site, v-int oncogene homolog) known to flank MEN-1; SEA was therefore designated a candidate gene for the MEN-1 locus. Many other disease regions identified by genetic studies have been mapped in a similar way: BRCA1,[56–58] neurofibromatosis type 2,[59] incontinentia pigmenti 1,[60] and Huntington's disease.[61]

7.2 Whole-genome Human Maps

In a similar way, the large number of genetic markers that had been mapped using meiotic techniques,[9,10] were subjected to the complementary technique of RH mapping. The markers had previously been shown to be unique and had been typed in naturally occurring populations. They were freely available for mapping on the new RH panels. Gyapay et al.[51] mapped 374 Genethon markers on the new GB4 panel, and found good agreement with the previous genetic maps. Stewart et al.[49] accomplished a similar task with 5994 markers generated at SHGC typed and mapped using the G3 panel. The largest map of sequence-tagged sites (STSs) contained 15 000 STSs.[62]

Earlier in the decade, it was proposed that sequencing of the entire human genome should be considered a practicality. Brenner[63] proposed that a reasonable intermediate stage should be first to sequence only those parts of the human genome that were actually expressed in tissue (1%), to maximize earlier the biologically functional information in the human genome. This proposal resulted in a vast effort to sequence all human expressed-sequence tags (ESTs).[64] It was soon apparent that the same underlying sequences were being measured independently by many different groups, and each many times. This prompted the construction of "UniGene",[46,65] whose aim was to cluster the individual EST sequences to the unique underlying gene that was being expressed. At the time, there were ~170 000 ESTs which were grouped into 49 625 clusters, of which 4563 (9%) corresponded to known genes. The current number of sequences in human UniGene is ~1.2 million, grouped in 75 000 clusters. Each gene is therefore represented by ~16 EST sequences on average, some genes being represented by many more.

Since most human genes are thought not to be polymorphic, it is not possible to map them using traditional genetic mapping techniques which detect different types of the same gene. RH mapping, in contrast, detects the presence or absence of sequences rather than sequence differences, and was therefore an obvious choice for mapping all human genes in a reasonable timeframe. An international consortium comprising 10 mapping centers was established,[66] with the intention of mapping all UniGene clusters. Representative sequences from each of the clusters (usually the longest) were distributed amongst the consortium members for typing on the G3 and GB4 panels. Map construction was coordinated at the Sanger Center, and the final maps were distributed from the NCBI web site. The first "gene map" of the human genome was published in 1996,[46] and showed the locations of more than 16 000 human genes. The latest release of this map[67] contains 30 000 genes.

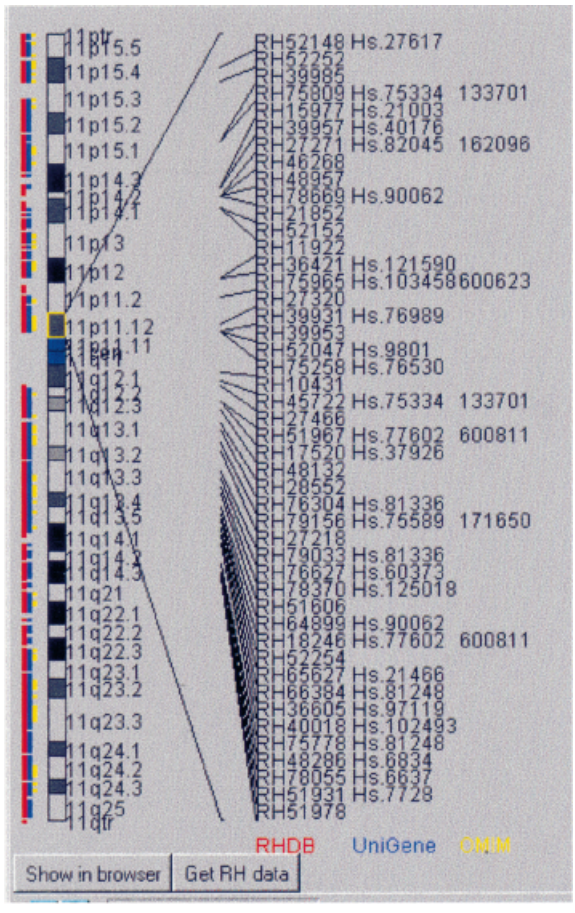

Figure 7 Transcripts mapped to human chromosome 11, band 11p11.12, in genemap98.[67]

Figure 7 shows the locations of 46 markers on band 11p11.12 estimated to span 4.46 Mb of the total 3000 Mb comprising the human genome (0.15%). It contains 46 mapped markers,[67] 29 of which are associated with UniGene clusters, and 7 of these clusters are associated via their gene symbols to the database of human genetic diseases and disorders, OMIM.[68,69] OMIM provides links to MedLine (Table 2). The loci corresponding to the OMIM entries are: **133701**–exostoses, multiple, type II; EXT2; **162096**–neurite growth-promoting factor 2; NEGF2; **600623**–prostate cancer antimetastasis gene KAII, leukocyte surface antigen R2, SAR2; antigen CD82, suppressor of antigenicity 6, prostate, ST6; R2 leukocyte antigen; **600811**–DNA damage-binding protein 2, DDB2; **171650**–acid phosphatase of tissues. The map has recently been updated[70,71] and contains 40 000 UniGene clusters.

The genomic localization of a unique DNA sequence can therefore potentially be linked to data and literature relating to human disease, and can provide a rapid means of devising novel strategies for the alleviation

or elimination of the symptoms of disease. This is one of the hopes of the new science of ''pharmacogenomics''.

7.3 Mapping Other Organisms

Genetic mapping studies of organisms other than human both predate and lag human studies. They predate human studies for two reasons: (1) genetic selection has been practiced for thousands of years on agricultural species (plants and livestock) to increase yield; and (2) the genetic basis for inheritance of characteristics was established primarily in nonhuman species. The results of genetic studies conducted on ''model'' organisms such as mouse and rat has also generated many variant phenotypes that resemble human disease phenotypes. They lag human studies since most attention following the development of modern molecular biology has been towards understanding malfunctions or disease phenotypes in man. However, the large-scale application of these techniques in nonhuman species is rapidly gaining pace.

The mouse provides many model systems for the study of human disease, so a dense genetic map is of great value. An RH panel for the mouse (T31) was constructed in 1996[72] and enabled construction of the first genome-wide RH anchor map of a model organism[73] containing 271 STSs. Markers were located on average 8.8 Mb apart, and contained some gaps which could not be bridged using the RH panel. Clusters of strongly linked markers were therefore ordered with respect to each other by anchoring on the genetic map. Mapping of ESTs on the mouse genome to a similar density as the human is in progress.[74] The current genetic map, containing ~7000 genes, is too sparse to allow successful application of the positional candidate cloning approach. The higher density RH map will allow mapping of many more genes. Centers mapping the mouse genome are the Jackson Laboratories, the Whitehead Institute and the Wellcome Trust Mouse Genome Center (see Table 2).

Many genetic experiments have also been conducted on the rat, and the phenotypes are useful as model systems for the study of similar conditions in humans. A RH panel of rat has been constructed and a dense map containing 5255 markers was recently published.[75] Another rat mapping project,[76] coordinated at the Whitehead Institute, has currently mapped 4736 simple sequence length polymorphism markers. These new maps of rat have revealed large regions of similarity between the rat and mouse and human genomes, providing valuable data for comparative mapping between the organisms.

The zebrafish has become a useful model organism for the study of the early stages of development, due to the ease with which its phenotypes can be determined by microscopy. A zebrafish RH panel has been constructed[77,78] and is being used for mapping cloned genes and ESTs which are candidates for developmental mutations.[79] The RH map is being anchored to the genetic map.[80]

An RH panel containing the entire dog genome has been constructed[81,82] on which 552 markers have been mapped. The dog species comprises 300 breeds and therefore contains a vast amount of polymorphism, manifest in many different phenotypes, e.g. physical appearance, behavior, and pathology. It is therefore an excellent species in which to study the genetic basis of these traits. The aim is to integrate the RH map with the third generation linkage map through the 207 common markers.

RH panels have also been constructed for cat,[83] pig,[84,85] and cow.[86]

7.4 Integrating Data

One of the major problems remaining is to integrate the vast amount of new mapping data with the existing knowledge of the genes themselves, their products, their functions and pathology. This was highlighted already in 1996[87,88] as a current problem. Eight years previously, it had been suggested that map and sequence data should be linked by STSs.[89] One locus would be represented by one STSs which has one DNA sequence. Weissenbach and Bentolila[87] regret that this has not been adopted in practice. They identify many sources of the problem, one being the proliferation and growth of databases such as dbEST, dbSTS, UniGene and OMIM. The problem is particularly acute in mouse and rat mapping, where different data are stored at many different sites. Databases easily and rapidly lose synchronization. Weissenbach and Bentolila suggest that mapping information should be simultaneously submitted to all relevant databases via a standard form. This would also prevent unlinked duplications. A universally acceptable mechanism is not yet implemented. When it is, the potential of the RH technique for comparative mapping and functional genomics will be greatly increased.

8 CONCLUSIONS

RH mapping is now an efficient and fast means of locating unique DNA sequences on a genome, and many methods exist for analyzing results of new RH experiments. The underlying physical processes are assumed to be simple, and have been of enormous importance in the HGP, providing a framework on which to place the DNA sequences of clones covering the entire human genome. The RH technique is now indispensable in rapidly generating genome-wide maps of different species. It was used to generate the high-resolution framework on which

to locate the large stretches of human DNA which are being measured in the HGP. It is almost certain that after completion of the HGP, RH mapping will be similarly used in the sequencing of other large vertebrate genomes, particularly mouse and rat. The high-resolution mapping of these organisms is already under way. It is also likely that similar effort will be directed at determining the gross genomic structure of any other organism that is considered important. It is already active in genomes of agricultural importance. It is further possible that any of the ~1 000 000 species on the planet will at some stage become abnormally hostile or abnormally beneficial. In these cases, RH mapping is a tool that can be used to investigate such species and characterize them at the genomic level, giving better understanding of their biological properties and effects.

ACKNOWLEDGMENTS

I would like to thank colleagues at the Imperial Cancer Research Fund, London, UK, Glaxo Wellcome Medicines Research Centre, Stevenage, UK, Oxford Molecular, Oxford, UK, and the Aventis Cambridge Genomics Center, Cambridge, USA, for valuable and informative discussions of different aspects of this subject: Stephan Beck, Kathy Call, Tim Connolly, Rainer Fuchs, Amanda Jackson, Hans Lehrach, Andrew Lyall, Brian Packard, John Riley, and Phillipe Sanseau.

ABBREVIATIONS AND ACRONYMS

DG	Distance Geometry
EST	Expressed-sequence Tag
HGP	Human Genome Project
HMMs	Hidden Markov Models
lod	Log-of-odds
ML	Maximum Likelihood
NMR	Nuclear Magnetic Resonance
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase Chain Reaction
RH	Radiation Hybrid
STSs	Sequence-tagged Sites
WWW	World Wide Web

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
Genome Physical Mapping Using BACs • Optical Mapping in Genomic Analysis • Polymerase Chain Reaction and Other Amplification Systems • Restriction Landmark Genomic and cDNA Scanning • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

REFERENCES

1. S.J. Goss, H. Harris, 'New Method for Mapping Genes in Human Chromosomes', *Nature*, **255**, 680–684 (1975).
2. S.J. Goss, H. Harris, 'Gene Transfer by Means of Cell Fusion. I. Statistical Mapping of the Human X-chromosome by Analysis of Radiation-induced Gene Segregation', *J. Cell. Sci.*, **25**, 17–37 (1977).
3. S.J. Goss, H. Harris, 'Gene Transfer by Means of Cell Fusion. II. The Mapping of 8 Loci on Human Chromosome 1 by Statistical Analysis of Gene Assortment in Somatic Cell Hybrids', *J. Cell. Sci.*, **25**, 39–57 (1977).
4. D.R. Cox, C.A. Pritchard, E. Uglum, D. Casher, J. Kobori, R.M. Myers, 'Segregation of the Huntington Disease Region of Human Chromosome 4 in a Somatic Cell Hybrid', *Genomics*, **4**, 397–407 (1989).
5. D.R. Cox, M. Burmeister, E.R. Price, S. Kim, R.M. Myers, 'Radiation Hybrid Mapping: A Somatic Cell Genetic Method for Constructing High-resolution Maps of Mammalian Chromosomes', *Science*, **250**, 245–250 (1990).
6. M.A. Walter, P.N. Goodfellow, 'Radiation Hybrids: Irradiation and Fusion Gene Transfer', *Trends Genet.*, **9**, 352–356 (1993).
7. M. Burmeister, S. Kim, E.R. Price, T. de Lange, U. Tantravahi, R.M. Myers, D.R. Cox, 'A Map of the Distal Region of the Long Arm of Human Chromosome 21 Constructed by Radiation Hybrid Mapping and Pulsed-field Gel Electrophoresis', *Genomics*, **9**, 19–30 (1991).
8. M.A. Walter, D.J. Spillett, P. Thomas, J. Weissenbach, P.N. Goodfellow, 'A Method for Constructing Radiation Hybrid Maps of Whole Genomes', *Nature Genet.*, **7**, 22–28 (1994).
9. C. Dib, S. Faure, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morissette, J. Weissenbach, 'A Comprehensive Genetic Map of the Human Genome Based on 5,264 Microsatellites', *Nature*, **380**, 152–154 (1996).
10. W. Dietrich, H. Katz, S.E. Lincoln, H.S. Shin, J. Friedman, N.C. Dracopoli, E.S. Lander, 'A Genetic Map of the Mouse Suitable for Typing Intraspecific Crosses', *Genetics*, **131**, 423–447 (1992).
11. J.B.S. Haldane, 'The Combination of Linkage Values and the Calculation of Distances between the Loci of Linked Factors', *J. Genet.*, **8**, 299–309 (1919).
12. W. Newell, S. Beck, H. Lehrach, A. Lyall, 'Estimation of Distances and Map Construction Using Radiation Hybrids', *Genome Res.*, **8**, 493–508 (1998).
13. J. Topping, *Errors of Observation and their Treatment*, 3rd edition, Chapman and Hall, London, 1962.

14. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1988.
15. M. Boehnke, K. Lange, D.R. Cox, 'Statistical Methods for Multipoint Radiation Hybrid Mapping', *Am. J. Hum. Genet.*, **49**, 1174–1188 (1991).
16. R.A. Fisher, 'Detection of Linkage with ''Dominant'' Abnormalities', *Ann. Eugen.*, **6**, 187–201 (1935).
17. J.B.S. Haldane, C.A.B. Smith, 'A New Estimate of the Linkage Between the Genes for Color-blindness and Haemophilia in Man', *Ann. Eugen.*, **14**, 10–31 (1947).
18. N.E. Morton, C.J. Maclean, R. Lew, S. Yell, 'Multipoint Linkage Analysis', *Am. J. Hum. Genet.*, **38**, 868–883 (1986).
19. N. Morton, V. Andrews, 'MAP, an Expert System for Multiple Pairwise Linkage Analysis', *Ann. Hum. Genet.*, **53**, 263–269 (1989).
20. J. Ott, 'Estimation of the Recombination in Human Pedigrees: Efficient Computation of the Likelihood for Human Linkage Studies', *Am. J. Hum. Genet.*, **26**, 588–597 (1974).
21. G.M. Lathrop, J.M. Lalouel, C. Julier, J. Ott, 'Strategies for Multilocus Linkage Analysis in Humans', *Proc. Natl Acad. Sci. USA*, **81**, 3443–3446 (1984).
22. E.S. Lander, P. Green, 'Construction of Multilocus Genetic Linkage Maps in Humans', *Proc. Natl Acad. Sci. USA*, **84**, 2363–2367 (1987).
23. A.P. Dempster, N.M. Laird, D.B. Rubin, 'Maximum-likelihood from Incomplete Data via the EM Algorithm', *J. R. Stat. Soc. B*, **39**, 1–22 (1977).
24. K. Lange, M. Boehnke, D.R. Cox, K.L. Lunetta, 'Statistical Methods for Polyploid Radiation Hybrid Mapping', *Genome Res.*, **5**, 136–150 (1995).
25. E.A. Thompson, 'Crossover Counts and Likelihood in Multipoint Linkage Analysis', *IMA J. Math. Appl. Med. Biol.*, **4**, 93–108 (1987).
26. C.T. Falk, 'A Simple Method for Ordering Loci Using Data from Radiation Hybrids', *Genomics*, **9**, 120–123 (1991).
27. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, 'Optimization by Simulated Annealing', *Science*, **220**, 671–680 (1983).
28. A. Nijenhuis, H.S. Wilf, *Combinatorial Algorithms*, 2nd edition, Academic Press, New York, 240–246, 1978.
29. J.H. Barrett, 'Genetic Mapping Based on Radiation Hybrid Data', *Genomics*, **13**, 95–103 (1992).
30. T.C. Matise, M. Perlin, A. Chakravarti, 'Automated Construction of Genetic Linkage Maps Using an Expert System (MultiMap): A Human Genome Linkage Map', *Nature Genet.*, **6**, 384–390 (1994).
31. S. Shaw, J. Farr, B. Thiel, T. Matise, J. Weissenbach, A. Chakravarti, C. Richard, 'A Radiation Hybrid Map of 95 STSs Spanning Human Chromosome 13q', *Genomics*, **27**, 502–510 (1995).
32. M. Boehnke, 'Genetic Analysis Workshop 7: Radiation Hybrid and Somatic Cell Hybrid Mapping of Chromosome 21', *Cytogenet. Cell Genet.*, **59**, 74–76 (1992).
33. D.T. Bishop, G.P. Crockford, 'Comparisons of Radiation Hybrid Mapping and Linkage Mapping', *Cytogenet. Cell Genet.*, **59**, 93–95 (1992).
34. M. Boehnke, 'Radiation Hybrid Mapping by Minimization of the Number of Obligate Chromosome Breaks', *Cytogenet. Cell Genet.*, **59**, 96–98 (1992).
35. D.E. Weeks, T. Lehner, J. Ott, 'Preliminary Ranking Procedures for Multilocus Ordering Based on Radiation Hybrid Data', *Cytogenet. Cell Genet.*, **59**, 125–127 (1992).
36. C.T. Falk, 'Multilocus Mapping Strategies on Chromosome 21 Data Sets: Comparison of Results from Family Data, Radiation Hybrids, and Somatic Cell Hybrids', *Cytogenet. Cell Genet.*, **59**, 119–121 (1992).
37. A. Chakravarti, J.E. Reefer, 'A Theory for Radiation Hybrid (Goss–Harris) Mapping: Application to Proximal 21q Markers', *Cytogenet. Cell Genet.*, **59**, 99–101 (1992).
38. S. Lawrence, N. Morton, 'Physical Mapping by Multiple Pairwise Analysis', *Cytogenet. Cell Genet.*, **59**, 107–109 (1992).
39. N.E. Morton, A. Collins, 'Standard Maps of Chromosome 10', *Ann. Hum. Genet.*, **54**, 233–251 (1990).
40. R. Guerra, M.S. McPeek, T.P. Speed, P.M. Stewart, 'A Bayesian Analysis for Mapping from Radiation Hybrid Data', *Cytogenet. Cell Genet.*, **59**, 104–106 (1992).
41. D. Slonim, L. Kruglyak, L. Stein, E. Lander, 'Building Human Genome Maps with Radiation Hybrids', *J. Comput. Biol.*, **4**, 487–504 (1997).
42. W.R. Newell, R. Mott, S. Beck, H. Lehrach, 'Construction of Genetic Maps Using Distance Geometry', *Genomics*, **30**, 59–70 (1995).
43. G.M. Crippen, 'Rapid Calculation of Coordinates from Distance Matrices', *J. Comput. Phys.*, **26**, 449–452 (1978).
44. G.M. Crippen, T.F. Havel, 'Stable Calculation of Coordinates from Distance Information', *Acta Crystallogr. A*, **34**, 282–284 (1978).
45. T.F. Havel, I.D. Kuntz, G.M. Crippen, 'Theory and Practice of Distance Geometry', *Bull. Math Biol.*, **45**, 665–720 (1983).
46. G.D. Schuler, M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tome, A. Aggarwal, B. Bajorek, S. Bentolila, B.B. Birren, A. Butler, A.B. Castle, N. Chiannilkulchai, A. Chu, C. Clee, S. Cowles, P.J. Day, T. Dibling, N. Drouot, I. Dunham, S. Duprat, C. East, T.J. Hudson, et al., 'A Gene Map of the Human Genome', *Science*, **274**, 540–546 (1996).
47. L. Stein, L. Kruglyak, D. Slonim, E. Lander, 'RHMAPPER, Installation and User's Guide', Whitehead Institute/MIT Center for Genome Research, http://www-genome.wi.mit.edu/ftp/distribution/software. 1995.
48. C. Soderlund, T. Lau, P. Deloukas, 'Z Extensions to the RHMAPPER Package', *Bioinformatics*, **14**, 538–539 (1998).
49. E.A. Stewart, K.B. McKusick, A. Aggarwal, E. Bajorek, S. Brady, A. Chu, N. Fang, D. Hadley, M. Harris, S. Hussain, R. Lee, A. Mratukulam, K. O'Connor, S. Perkins,

M. Piercy, F. Qin, T. Reif, C. Sanders, X. She, W.L. Sun, P. Tabar, S. Voyticky, S. Cowles, J.B. Fan, D.R. Cox, et al., 'An STS-based Radiation Hybrid Map of the Human Genome', *Genome Res.*, **7**, 422–433 (1997).

50. W. Newell, P. Sanseau, J. Riley, A. Lyall, 'WEBMAP: Radiation Hybrid Mapping on the WWW', *Bioinformatics*, **14**, 825–826 (1998).
51. G. Gyapay, K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillett, D. Muselet, J.F. Prud'Homme, C. Dib, C. Auffray, J. Morissette, J. Weissenbach, P.N. Goodfellow, 'A Radiation Hybrid Map of the Human Genome', *Hum. Mol. Genet.*, **5**, 339–346 (1996).
52. E.M. Beasley, E.A. Stewart, K.B. McKusick, A. Aggerwal, S. Brady-Hebert, N.Y. Fang, D. Hadley, M. Harris, S.C. Lewis, S.M. Perkins et al., 'TNG4 Radiation Hybrid Maps Improve the Resolution of the G3 RH Maps of the Human Genome', in *Genome Mapping and Sequencing 1997*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997.
53. D. Wang, J. Fan, C.-J. Siao, J. Spencer, P. Young, L. Kruglyak, R. Sapolsky, A. Berno, E. Hubbell, G. Ghandour, T. Topaloglou, J. Rioux, T. Hawkins, L. Stein, T. Hudson, R. Lipshutz, M. Chee, E. Lander, 'Toward a Third Generation Genetic Map of the Human Genome Based on Bi-allelic Polymorphisms', in *Genome Mapping and Sequencing 1997*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997.
54. F.S. Collins, 'Positional Cloning Moves from Perditional to Traditional', *Nature Genet.*, **9**, 347–350 (1995).
55. C.W. Richard, D.A. Withers, T.C. Meeker, S. Maurer, G.A. Evans, R.M. Myers, D.R. Cox, 'A Radiation Hybrid Map of Human Chromosome 11 Containing the Multiple Endocrine Neoplasia Type 1 (MEN-1) and bcl-1 Disease Loci', *Am. J. Hum. Genet.*, **49**, 1189–1196 (1991).
56. K.J. Abel, M. Boehnke, M. Prahalad, P. Ho, W.L. Fleiter, M. Watkins, J. VanderStoep, S.C. Chandrasekharappa, F.S. Collins, T.W. Glover, B.L. Weber, 'A Radiation Hybrid Map of the BRCA1 Region of Chromosome 17q12-q21', *Genomics*, **17**, 632–641 (1993).
57. D.M. Black, H. Nicolai, J. Borrow, E. Solomon, 'A Somatic Cell Hybrid Map of the Long Arm of Human Chromosome 17, Containing the Familial Breast Cancer Locus (BRCA1)', *Am. J. Hum. Genet.*, **52**, 702–710 (1993).
58. P. O'Connell, H. Albertson, N. Matsunami, T. Taylor, J.E. Hundley, T.L. Johnson-Pais, B. Reus, E. Lawrence, L. Ballard, R. White, R.J. Leach, 'A Radiation Hybrid Map of the BRCA1 Region', *Am. J. Hum. Genet.*, **54**, 526–534 (1994).
59. K.A. Frazer, M. Boehnke, M.L. Budarf, R.K. Wolf, B.S. Emanuel, R.M. Myers, D.R. Cox, 'A Radiation Hybrid Map of the Region on Human Chromosome 22 Containing the Neurofibromatosis Type 2 Locus', *Genomics*, **14**, 574–584 (1992).
60. J.L. Gorski, M. Boehnke, E.L. Reyner, E.M. Burright, 'A Radiation Hybrid Map of the Proximal Short Arm of the Human X Chromosome Spanning Incontinentia Pigmenti 1 (IP1) Translocation Breakpoints', *Genomics*, **14**, 657–665 (1992).
61. M.R. Altherr, S. Plummer, G. Bates, M. MacDonald, S. Taylor, H. Lehrach, A.-M. Frischauf, J.F. Gusella, M. Boehnke, J.J. Wasmuth, 'Radiation Hybrid Map Spanning the Huntington Disease Gene Region of Chromosome 4', *Genomics*, **13**, 1040–1046 (1992).
62. T.J. Hudson, L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.H. Xu, et al., 'An STS-based Map of the Human Genome', *Science*, **270**, 1945–1954 (1995).
63. S. Brenner, 'The Human Genome: The Nature of the Enterprise', *Ciba Found. Symp.*, **149**, 6–12 (1990).
64. M.D. Adams, J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno, et al., 'Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project', *Science*, **252**, 1651–1656 (1991).
65. UniGene, Unique Gene Sequence Collection for Human, Mouse and Rat, http://www.ncbi.nlm.nih.gov/UniGene/index.html.
66. Radiation Hybrid Mapping Consortium, http://www.sanger.ac.uk/HGP/Rhmap/rhconsortium.shtml.
67. P. Deloukas, G.D. Schuler, G. Gyapay, E.M. Beasley, C. Soderlund, P. Rodriguez-Tome, L. Hui, T.C. Matise, K.B. McKusick, J.S. Beckmann, S. Bentolila, M. Bihoreau, B.B. Birren, J. Browne, A. Butler, A.B. Castle, N. Chiannilkulchai, C. Clee, P.J. Day, A. Dehejia, T. Dibling, N. Drouot, S. Duprat, C. Fizames, D.R. Bentley, et al., 'A Physical Map of 30,000 Human Genes', *Science*, **282**, 744–746 (1998).
68. V.A. McKusick, *Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders*, 12th edition, Johns Hopkins University Press, Baltimore, 1998.
69. Online Mendelian Inheritance in Man, OMIM (TM), Center for Medical Genetics, Johns Hopkins University, Baltimore, MD and National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, 1999. World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/.
70. L. Green, The International RH Mapping Consortium, *Genome Sequencing and Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999.
71. The International RH Mapping Consortium, A New Gene Map of the Human Genome, 1999, http://www.ncbi.nlm.nih.gov/genemap/.
72. K. Schmitt, J.W. Foster, R.W. Feakes, C. Knights, M.E. Davis, D.J. Spillett, P.N. Goodfellow, 'Construction of a Mouse Whole-genome Radiation Panel and Application to MMU11', *Genomics*, **34**, 193–197 (1996).
73. L.C. McCarthy, J. Terrett, M.E. Davis, C.J. Knights, A.L. Smith, R. Critcher, K. Schmitt, J. Hudson, N.K. Spurr, P.N. Goodfellow, 'A First-generation Whole Genome Radiation Hybrid Map Spanning the Mouse Genome', *Genome Res.*, **7**, 1153–1161 (1997).

74. S. Greenaway, C. Heuston, P.J. Trickett, A. Southwell, M.A. Strivens, T. Matise, A.R. Haynes, H. Doi, J. Kitchen, M. Ko, S.D.M. Brown, P. Denny, 'Radiation Hybrid Mapping of Mouse Expressed Sequence Tags', in *Genome Sequencing and Biology 1999*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999.
75. T.K. Watanabe, M.T. Bihoreau, L.C. McCarthy, S.L. Kiguwa, H. Hishigaki, A. Tsuji, J. Browne, Y. Yamasaki, A. Mizoguchi-Miyakita, K. Oga, T. Ono, S. Okuno, N. Kanemoto, E. Takahashi, K. Tomita, H. Hayashi, M. Adachi, C. Webber, M. Davis, S. Kiel, C. Knights, A. Smith, R. Critcher, J. Miller, M.R. James, et al., 'A Radiation Hybrid Map of the Rat Genome Containing 5,255 Markers', *Nat. Genet.*, **22**, 27–36 (1999).
76. R.G. Steen, A.E. Kwitek-Black, C. Glenn, J. Gullings-Handley, W. Van Etten, O.S. Atkinson, D. Appel, S. Twigger, M. Muir, T. Mull, M. Granados, M. Kissebah, K. Russo, R. Crane, M. Popp, M. Peden, T. Matise, D.M. Brown, J. Lu, S. Kingsmore, P.J. Tonellato, S. Rozen, D. Slonim, P. Young, M. Knoblauch, A. Provoost, D. Ganten, S.D. Colman, J. Rothberg, E.S. Lander, H.J. Jacob, 'A High-density Integrated Genetic Linkage and Radiation Hybrid Map of the Laboratory Rat', *Genome Res.*, **9**, AP1–AP8 (1999).
77. C. Kwok, R.M. Korn, M.E. Davis, D.W. Burt, R. Critcher, L. McCarthy, B.H. Paw, L.I. Zon, P.N. Goodfellow, K. Schmitt, 'Characterization of Whole Genome Radiation Hybrid Mapping Resources for Non-mammalian Vertebrates', *Nucleic Acids Res.*, **26**, 3562–3566 (1998).
78. C. Kwok, R. Critcher, K. Schmitt, 'Construction and Characterization of Zebrafish Whole Genome Radiation Hybrids', *Methods Cell. Biol.*, **60**, 287–302 (1999).
79. F. van Bebber, R. Geisler, G.-J. Rauch, H. Schauerte, P. Haffter, 'Radiation Hybrid Mapping of cDNAs, ESTs and SSLPs in the Zebrafish', 1999, http://www.eb.tuebingen.mpg.de/abt.3/haffter_lab/rh_mapping.html.
80. E.W. Knapik, A. Goodman, M. Ekker, M. Chevrette, J. Delgado, S. Neuhauss, N. Shimoda, W. Driever, M.C. Fishman, H.J. Jacob, 'A Microsatellite Genetic Linkage Map for Zebrafish', *Nat. Genet.*, **18**, 338–343 (1999).
81. C. Priat, C. Hitte, F. Vignaux, C. Renier, Z. Jiang, S. Jouquand, A. Cheron, C. Andre, F. Galibert, 'A Whole-genome Radiation Hybrid Map of the Dog Genome', *Genomics*, **54**, 361–378 (1998).
82. F. Vignaux, C. Priat, S. Jouquand, C. Hitte, Z. Jiang, A. Cheron, C. Renier, C. Andre, F. Galibert, 'Toward a Dog Radiation Hybrid Map', *J. Hered.*, **90**, 62–67 (1999).
83. W.J. Murphy, M. Menotti-Raymond, L.A. Lyons, M.A. Thompson, S.J. O'Brien, 'Development of a Feline Whole Genome Radiation Hybrid Panel and Comparative Mapping of Human Chromosome 12 and 22 Loci', *Genomics*, **57**, 1–8 (1999).
84. M. Yerle, P. Pinton, A. Robic, A. Alfonso, Y. Palvadeau, C. Delcros, R. Hawken, L. Alexander, C. Beattie, L. Schook, D. Milan, J. Gellin, 'Construction of a Whole-genome Radiation Hybrid Panel for High-resolution Gene Mapping in Pigs', *Cytogenet. Cell Genet.*, **82**, 182–188 (1998).
85. A. Robic, V. Seroude, J.T. Jeon, M. Yerle, L. Wasungu, L. Andersson, J. Gellin, D. Milan, 'A Radiation Hybrid Map of the RN Region in Pigs Demonstrates Conserved Gene Order Compared with the Human and Mouse Genomes', *Mamm. Genome*, **10**, 565–568 (1999).
86. J.E. Womack, J.S. Johnson, E.K. Owens, C.E. Rexroad III, J. Schlapfer, Y.P. Yang, 'A Whole-genome Radiation Hybrid Panel for Bovine Gene Mapping', *Mamm. Genome*, **8**, 854–856 (1997).
87. J. Weissenbach, S. Bentolila, 'Integrating Maps Requires Integrated Data', *Nature Biotechnol.*, **14**, 678 (1996).
88. D.T. Kingsbury, 'Consensus, Common Entry, and Community Concern', *Nature Biotechnol.*, **14**, 678 (1996).
89. M. Olson, L. Hood, C. Cantor, D. Botstein, 'A Common Language for Physical Mapping of the Human Genome', *Science*, **245**, 1434–1435 (1989).

Restriction Landmark Genomic and cDNA Scanning

Jun Kawai and Yoshihide Hayashizaki
Institute of Physical and Chemical Research (RIKEN), Ibaraki, Japan and CREST, Japan Science and Technology Corporation (JST), Ibaraki, Japan

Yasushi Okazaki, Harukazu Suzuki, and Sachihiko Watanabe
Institute of Physical and Chemical Research (RIKEN), Ibaraki, Japan

1 Definitions 1
2 Introduction 1
3 General Principles 3
3.1 Restriction Landmark Genomic Scanning 3
3.2 Restriction Landmark cDNA Scanning 5
4 Protocols 8
4.1 Introduction 8
4.2 Preparation of Genomic DNA 10
4.3 Labeling of Restriction Landmarks 11
4.4 Two-dimensional Gel Electrophoresis 12
4.5 Methods for Spot Cloning 15
5 Application of Restriction Landmark Genomic Scanning 19
5.1 Potential Applications 19
5.2 Application of Restriction Landmark Genomic Scanning to Cancer Research 20
Acknowledgments 24
Abbreviations and Acronyms 24
Related Articles 24
References 24

Restriction landmark genomic scanning (RLGS) is a method for the high-speed survey of the presence or absence of restriction landmarks throughout a genome. RLGS also provides for the measurement of the copy number of restriction landmarks based on the concept that restriction enzyme sites can serve as landmarks. RLGS employs direct end-labeling of the genomic DNA digested with a restriction enzyme and high-resolution, two-dimensional electrophoresis. Compared with the conventional technologies for detecting the genes, such as Southern blot analysis and polymerase chain reaction (PCR), RLGS has the following advantages even though it needs specially designed instruments: (1) high-speed scanning ability – thousands of restriction landmarks can be scanned simultaneously; (2) applicability to any organism – RLGS employs direct labeling of restriction enzyme sites, not a hybridization procedure; (3) a spot intensity that reflects the copy number of the restriction landmarks; and (4) the ability, by using a methylation-sensitive enzyme, to screen the methylated state of genomic DNA.

RLGS is a very useful system not only for genome mapping, but also in a variety of biological and medical fields, such as cancer research and the growth and development of various organisms throughout all stages of the life cycle. RLGS can also be applied for the visualization of mRNAs [restriction landmark cDNA scanning (RLCS)]. In one procedure, RLCS can scan the expression of several hundred mRNAs.

1 DEFINITIONS

Landmark: a locus or site to be visualized, distributed across genomes. A landmark can be utilized as a genetic marker, which should represent the physical condition, such as copy number, of its flanking region.

Restriction landmark: a landmark which can be recognized by a restriction enzyme.

RLGS: a method for visualizing restriction landmarks as spots on a two-dimensional profile. RLGS visualizes more than 2000 restriction landmarks in a genome in one procedure. RLGS has many applications, such as making a whole genome linkage map and scanning the amplified or deleted loci in cancer.

RLCS: a cDNA visualization system in which RLGS is applied to display many cDNA species as two-dimensional gel spots. The intensity of each RLCS spot reflects the expression level of each corresponding gene.

DNA methylation: in higher organisms, methylation of cytosine residues is often observed in genomes, and DNA methylation generally associates with the gene expression.

Restriction landmark genomic scanning for methylation (RLGS-M): a method for scanning genome-wide methylation conditions. RLGS-M can survey both endogenously imprinted genes and potentially expressed genes, based on the correlation of DNA methylation with gene expression.

2 INTRODUCTION

In recent decades, international genome projects have been making great progress toward the determination

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of the sequence of genomic DNA and the full-length cDNA encoding its final protein products.[1–3] In terms of the analysis of the genomic structure, the full DNA sequence – including all information encoded by DNA other than DNA modification (methylation) – is the final goal. To achieve this, much effort has been devoted to the development of DNA sequencing technologies to facilitate sequencing speed. In fact, recent advances in these technologies should enable the entire genome-wide sequencing – or at least the rough sketching of the genomic sequence – to be completed within a few years, even for human, mouse and other higher organisms whose genome size exceeds 3×10^9 base pairs (bp). Once the genomic sequence and the full-length cDNA sequence have been determined, the diversity of genomes can then be analyzed by studying sequence variations, such as single nucleotide polymorphism (SNP).

The technologies for de novo mapping and sequencing of the genome are totally different from postsequencing technologies. In the initial sequencing project, the mapping and sequencing technologies should not require any pre-existing information concerning the genome-wide sequence. On the other hand, at the postsequencing stage, the existence of genome-wide sequencing information will allow us to design much faster and more accurate technologies to analyze the individual variations in a sequence which has already been determined. In other words, we can then use any region of the genomic DNA as a 'landmark', based on knowledge of the genome-wide sequence.

The concept of genomic scanning arose from the necessity to sketch out the framework of the genome, using appropriate technology to visualize landmarks distributed throughout the genome. In this context, a 'landmark' is defined as a locus or site to be visualized, distributed across genomes. However, different types of landmark are used for genomic scanning prior genome-wide sequencing. The landmarks used in the pregenome-wide sequencing stage (presequencing landmarks) should not require any a priori sequencing information. On the other hand, landmarks used subsequent to genome-wide sequencing (postsequencing landmarks) should be susceptible to much more rapid screening than those used in the presequencing stage, taking advantage of the fact that postsequencing technology can use landmarks already identified. Once the genomic sequence has been determined, oligonucleotides which hybridize at any given site on the genome and which amplify any region of the genome through PCR can be designed to single-base accuracy. In theory, oligonucleotide chips can be used to identify SNP at any site on the genome. However, even with the recent amazing improvements in sequencing technology, the sequencing of large genomes still requires enormous amounts of time, labor and financing, and at present the best that can be expected is that the human genome may be completely sequenced in the coming few years. Under the circumstances, with many genomes to be analyzed, a rapid genome scanning system for de novo mapping based on 'presequencing landmarks' remains essential, not only for mapping, but also for the sequencing itself.

The ideal landmark for genomic scanning should offer the following benefits: (1) robust landmark signals which can be detected under reproducible conditions; (2) high-speed (multiplex) scanning; (3) signals which facilitate measurement of landmark copy number; (4) simple procedures and robotization; (5) scanning field expandability; (6) applicability to any organism; and (7) exportability of both techniques and information to other laboratories. So far, three types of 'presequencing landmarks' have been used to scan the genome. The first, the DNA probe with unique sequence, was quantitatively and qualitatively visualized using the Southern blot method. Detection of landmarks results from the sequence-specific manner of DNA–DNA hybridization. PCR was developed as an alternative method for detecting a second sort of landmark. PCR is a technique for amplifying a region of several hundred base pairs, flanked by a pair of oligonucleotide primers. Here, too, the specificity of the signals depends on the specificity of hybridization of the primers to a unique genomic region.

Recently, our group introduced a new or third approach, termed the 'restriction landmark', in which each site recognized by a restriction enzyme can be utilized as a landmark. Based on this concept, we have developed a method of RLGS which can visualize restriction landmarks as spots on a two-dimensional profile.[4,5] This method employs (1) direct end-labeling of genomic DNA digested with a restriction enzyme and (2) high-resolution two-dimensional electrophoresis. As a method, it provides an alternative, multiplex approach to genomic analysis using robust landmarks, based on the strict specificity of sequence recognition of the restriction enzyme.

RLGS is a very important innovation for the following reasons. First, it provides us with a most powerful and speedy tool for genome-wide mapping for any organism.[6–8] RLGS can scan more than 2000 landmarks in a single procedure. Once DNA samples from the family tree of any organism have been prepared, all of the polymorphic spots can be used for very rapid genetic mapping. Even in sketching out the framework of a genome, a mapping approach based on presequence landmarks, such as restriction landmarks, is still generally much faster than genome-wide sequencing, and a map of these landmarks is very helpful for the genome-wide sequencing itself. Second, RLGS does not require any a priori sequence information, reducing the total cost of mapping. The

most salient characteristic of restriction landmarking is that the methylation state of the landmark can be reproducibly assayed if a methylation-sensitive enzyme is used as the landmarking enzyme. To scan the methylation state throughout the genome, RLGS is thus the most rapid method available. In particular, the methylation condition of CpG islands controlling the transcription of genes can be quickly screened using GC-rich restriction enzymes such as *Not*I as the landmarking enzyme.

To sum up, both pre- and postsequence landmarks are essential for genome mapping and scanning, complementing each other. Here, we describe the RLGS process which allows restriction landmarks to be surveyed throughout a genome, together with suggestions for its application.

3 GENERAL PRINCIPLES

3.1 Restriction Landmark Genomic Scanning

RLGS is based on a new approach to scanning efficiently and visibly many restriction enzyme sites on the entire genomic DNA. In essence, two issues must be resolved to achieve such scanning. The first is how to label specifically the particular restriction sites as landmarks without incurring nonspecific labeling. The second is how to detect through one procedure as many landmark loci as possible. Owing to the high complexity of genomic DNA in higher organisms, it has previously been considered difficult to achieve specific end-labeling at particular restriction sites and to detect the labeled fragments by conventional gel electrophoretic analysis. For example, when six-bp-recognizing enzymes were conventionally used as restriction enzymes to analyze the genomic DNA of higher organisms, it was very difficult to separate and detect the labeled restriction enzyme sites as discrete gel spots, because of the huge number of restriction enzyme sites and the nonspecific labeling of damaged sites such as nicks, gaps and/or double-stranded breaks in genomic DNA. However, these problems have been overcome by adopting new eight-bp-recognizing restriction enzymes known as 'rare-cutters' so as to give an appropriate number of DNA fragments for separation, in addition to new types of dideoxy nucleotide analogs to reduce nonspecific labeling. There has also been the development of a high-resolution two-dimensional electrophoresis system which enables us to detect simultaneously a multitude of loci as gel spots with enhanced sensitivity. For RLGS, it is essential, of course, to prepare genomic DNA of high quality. Average length and damage such as gaps and nicks are significant considerations. Since contamination by short fragments of less than 10 kbp in length results in increased background noise, the DNA must be prepared carefully without enzymatic degradation, especially at the lysis step at which various kinds of endonucleases leak from the lysosome. Treatment with proteinase K is effective in curbing such degradation in DNA preparation, but more care must be paid to avoiding enzymatic degradation by nuclease than to avoiding mechanical degradation such as shearing.

Figure 1 is a schematic representation of the principles and procedure of RLGS. The DNA is treated step by step as follows.

3.1.1 Blocking

When a fixed amount of genomic DNA is subjected to gel electrophoretic analysis, the copy number of the DNA decreases as the genome size increases. However, nonspecific labeling is independent of genome size. Hence the ratio of noise to meaningful signals increases with increased genome size. Such noise substantially hampers the detection of specific signals in the case of genome sizes in excess of 1×10^8 bp. Thus, damaged DNA sites such as nicks, gaps and/or double-stranded breaks must be blocked in advance to avoid nonspecific incorporation of the radioisotope. For this purpose, a blocking technique has been developed employing dideoxy nucleotide analogs such as dideoxy-[α-thio] nucleotides with one sulfur atom substituted for an oxygen atom (Figure 2). The incorporation of the analogs protects the damaged sites from elongation which would cause nonspecific labeling, as well as from 3′ exonuclease attack at the same sites in the labeling step. These analogs are incorporated enzymatically by a DNA polymerase such as DNA polymerase I.

3.1.2 Landmark Cleavage with Restriction Enzyme E_L

The sites for restriction enzyme E_L are isotopically labeled as landmarks. It is important to consider how many landmarks are desirable for the efficient detection of as many gel spots as possible. The appropriate number of restriction sites for the landmarks is around 3000 (each restriction site gives two landmarks). This means that restriction enzyme E_L must be chosen to give DNA fragments with an average length of 1 Mbp in the case of human and other mammalian genome DNA whose length is approximately 3×10^9 bp. Thus, enzymes that recognize eight-base or rare sequences are most convenient for use as restriction enzyme E_L. Table 1 lists representative restriction enzymes.

3.1.3 Labeling of Restriction Landmarks

Different labeling procedures are adopted depending on the nature of the cleaved sites generated by the restriction enzyme treatment. For a 5′-protruding end, a filling reaction of [α-^{32}P]deoxynucleotide is carried out

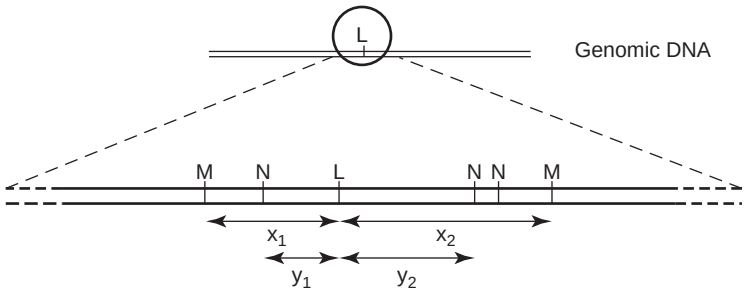

L: Site for restriction enzyme E_L (restriction landmark)

M: Site for restriction enzyme E_M

N: Site for restriction enzyme E_N

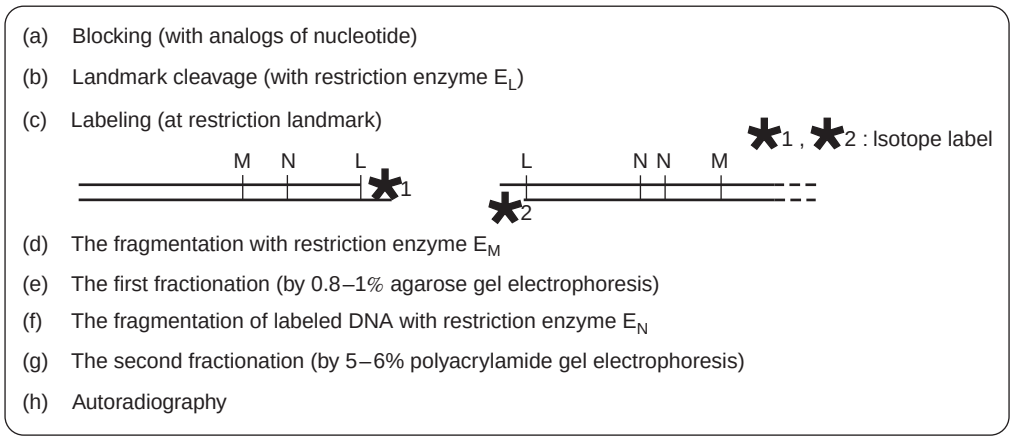

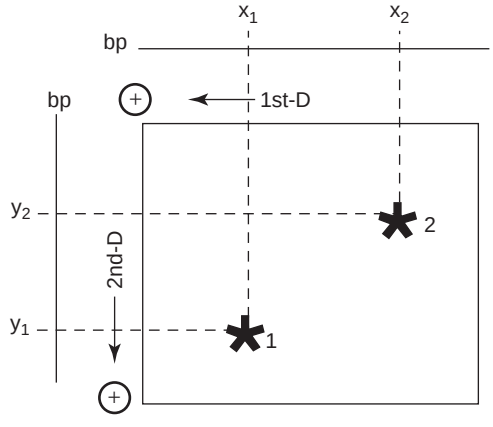

Figure 1 Schematic representation of the principles of RLGS using two-dimensional gel electrophoresis. ∗1 and ∗2 represent the distance from restriction landmark L (site for restriction enzyme E_L) to the nearest sites M for restriction enzyme E_M. y_1 and y_2 indicate the distance from the restriction landmarks L to the nearest sites N for restriction enzyme E_N. These end-labeled fragments appear as spots ∗1 and ∗2 in the two-dimensional gel electrophoresis profile.

with Sequenase® Version 2, whereas for a 3′-protruding (or blunt) end, labeling is done with terminal deoxynucleotidyl transferase (terminal transferase) which can add [α-^{32}P]dideoxynucleotides at the 3′ termini of DNA fragments.

3.1.4 Fragmentation with Restriction Enzyme E_M

Generally, for mammalian genomes, more than 3000 DNA fragments of 1 Mbp average length are generated by treatment with restriction enzyme E_L, even when the

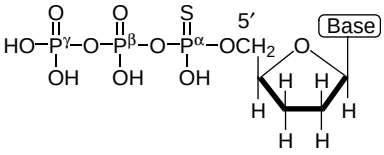

Figure 2 Dideoxy nucleotide analog, dideoxynucleoside [α-thio] triphosphate, used for the blocking procedure.

Table 1 Restriction enzymes usable for landmark cleavage (E_L) and sensitive to mammalian genome methylation

5′^{Me}CG-sensitive enzyme		5′ Non-^{Me}CG-sensitive enzyme	
*Not*I	GCGGCCGC	*Pac*I	AATTAATT
*Sfi*I	GGCCNNNNNGGCC	*Swa*I	ATTTAAAT
*Fse*I	GGCCGGCC	*Sse*8387I	CCTGCAGG
*Csp*I	CGG(A/T)CCG	*Pme*I	GTTTAAAC
*Mlu*I	ACGCGT		
*Asc*I	GGCGCGCC		
*Bss*HII	GCGCGC		
*Cla*I	ATCGAT		

chosen enzyme is *Not*I, the eight-bp-recognizing enzyme known as 'rare-cutter'. In this case, the DNA fragments are too large to separate by gel electrophoresis. To overcome the problem, treatment with restriction enzyme E_M, whose recognition sites occur more frequently in the genome than landmark E_L sites, is adopted. Restriction enzyme E_M must be selected to give a suitable number of gel spots allowing the use of a high-resolution electrophoretic thin-layer gel with a range extending to more than 40 cm.

3.1.5 First Fractionation by Agarose Gel Electrophoresis

In high-resolution electrophoresis, the thickness of the first dimension (1st-D) agarose gel is important, because its resolution affects that of second dimension (2nd-D) electrophoresis. Either horizontal or vertical gels are suitable if the gel is thin or fine enough to separate more than 3000 landmarks electrophoretically.

3.1.6 Fragmentation of Labeled DNA with Restriction Enzyme E_N

The agarose gel strip or gel cylinder used in 1st-D electrophoresis is treated in situ with restriction enzyme E_N, whose sites appear even more frequently than restriction enzyme E_M sites. Usually four-bp-recognizing enzymes are used as restriction enzyme E_N, but six-bp-recognizing enzymes are also usable.

3.1.7 Second Fractionation by Polyacrylamide Gel Electrophoresis

The thickness of the second gel is determined by that of the first gel, as already noted. The first cylindrical gel or 'noodle' is connected to the 2nd-D polyacrylamide gel, which is also usable either horizontally or vertically. DNA fragments between 70 bp and 2 kbp are separable by this electrophoresis.

3.1.8 Autoradiography

The gel sheets are dried and autoradiographed.

3.1.9 Results

Figure 3 shows a representative result of RLGS, when *Not*I, *Eco*RV and *Hin*fI are used as restriction enzymes E_L, E_M and E_N. More than 2000 spots are clearly visible.

3.2 Restriction Landmark cDNA Scanning

RLGS is a powerful and efficient method for analyzing an entire genome.[9] It enables us to visualize simultaneously many loci as gel spots. RLGS has proved to be powerful in various applications such as genetic mapping, detection of

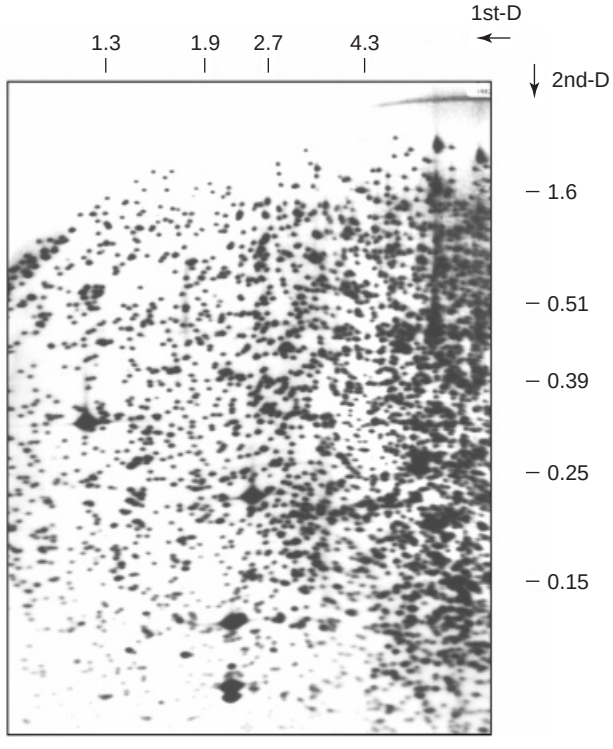

Figure 3 Typical RLGS profile of mouse liver DNA using *Not*I, *Eco*RV and *Hin*fI as restriction enzymes E_L, E_M and E_N, respectively. Approximate sizes are shown in kb.

methylatable loci including imprinted loci and detection of aberrated loci in cancer cells.

RLCS is a novel cDNA visualization system in which RLGS is applied to display many cDNA species as two-dimensional gel spots. The principle of RLCS is essentially the same as that of RLGS, with the cDNA cleaved by restriction enzymes being labeled isotopically at the restriction sites and separated by high-resolution two-dimensional gel electrophoresis, followed by autoradiography. The preparation of cDNA species of a uniform length for each mRNA species is crucial for RLCS, just as the quality of DNA is crucial for RLGS. For this purpose, a specially designed oligo(dT) anchor primer which facilitates anchoring to mRNA at the upstream end of the poly(A) tail, and easy recovery by its biotin residue at the 5′ end, is used for cDNA synthesis. The primer also carries multiple restriction enzyme sites such as *Not*I, *Nru*I, *Xba*I and *Spe*I. A representative primer is 3′-AMT15CCCGCCGGCGAGCGCTAGAT-CTTGATCAG-biotin (M = A, G or C). Alternatively, an anchor primer with G, C or T instead of A at the 3′ end is also usable. Figure 4(a) and (b) is a schematic representation of the principles and sequential procedures of RLCS.

cDNA synthesized using reverse transcriptase and the anchor primer is enzymatically blocked by 2′,3′-dideoxyribonucleoside triphosphate (ddNTP) to prevent nonspecific labeling in the labeling reaction (steps 1 and 2 in Figure 4a). The cDNA is then cleaved by the first restriction enzyme E_L whose sites (L) are successively labeled as landmarks (steps 3 and 4). An enzyme that generates protruding cohesive 5′ termini is recommended for use as restriction enzyme E_L because its restriction sites are easily radiolabeled by DNA polymerase I. Such

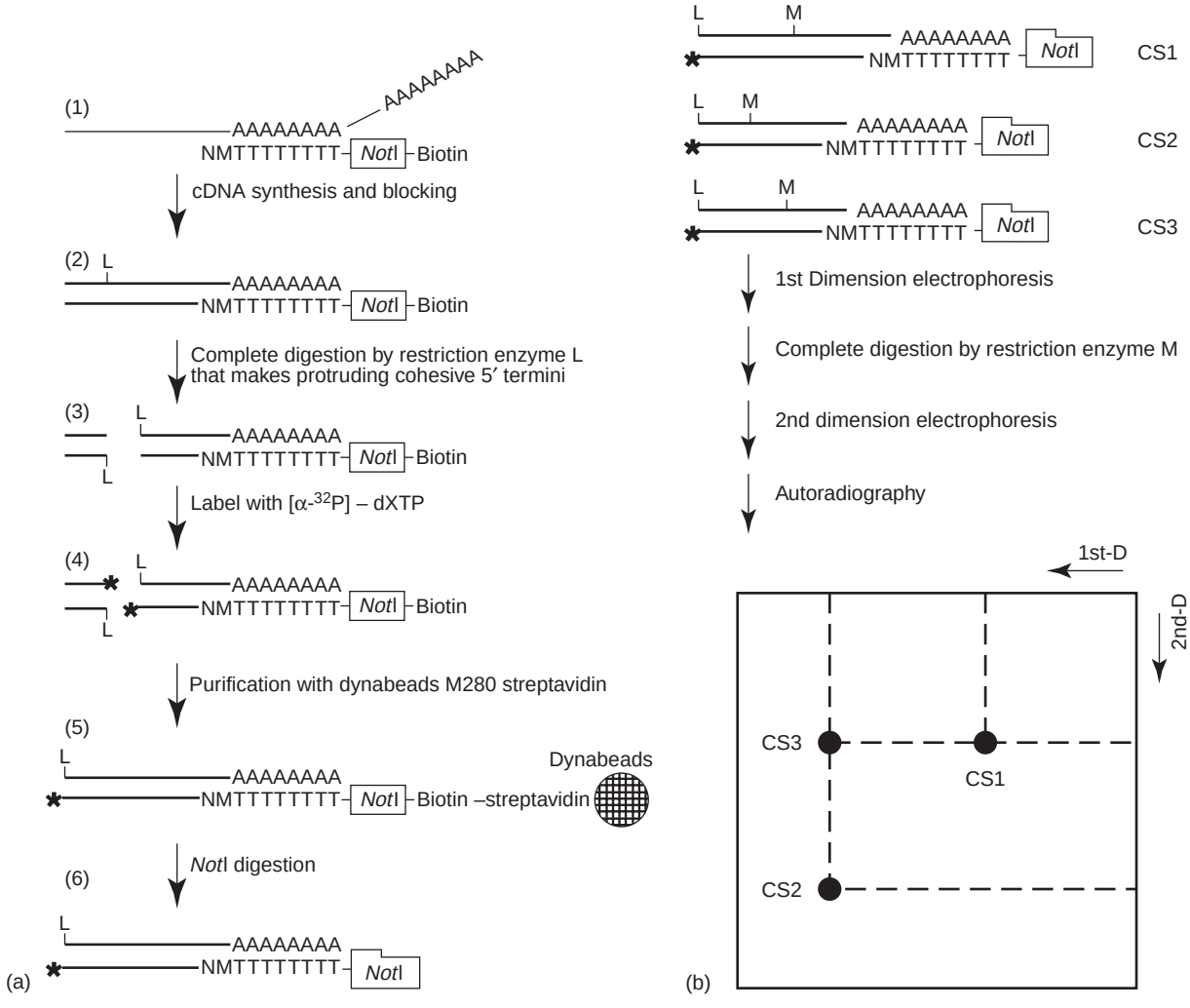

Figure 4 Schematic representation of the principles and procedure of RLCS. (a) Preparation of RLCS sample. The thin horizontal line and thick double horizontal lines indicate poly(A) RNA and double-stranded cDNA, respectively. (b) Separation of cDNA species by two-dimensional gel electrophoresis. cDNA fragments in RLCS samples (CS1, CS2 and CS3) are separated by two-dimensional gel electrophoresis and detected as spots CS1, CS2 and CS3, respectively.

an enzyme is also preferable for recovering target fragments from the gel spots when spot cloning is necessary. These labeled fragments are recovered using streptavidin-conjugated magnetic beads (step 5), and released from the beads by restriction enzyme *Not*I, whose site is included in the anchor primer (step 6). The recovered fragments, which must be uniform in length for individual cDNA species, are subjected to high-resolution two-dimensional gel electrophoresis as in the case of RLGS (Figure 4b). In practice, cDNA synthesis is carried out using several micrograms of poly(A)$^+$RNA and 1 μg of the anchor primer. Double-stranded cDNA (0.2–0.3 μg, 1000 bp

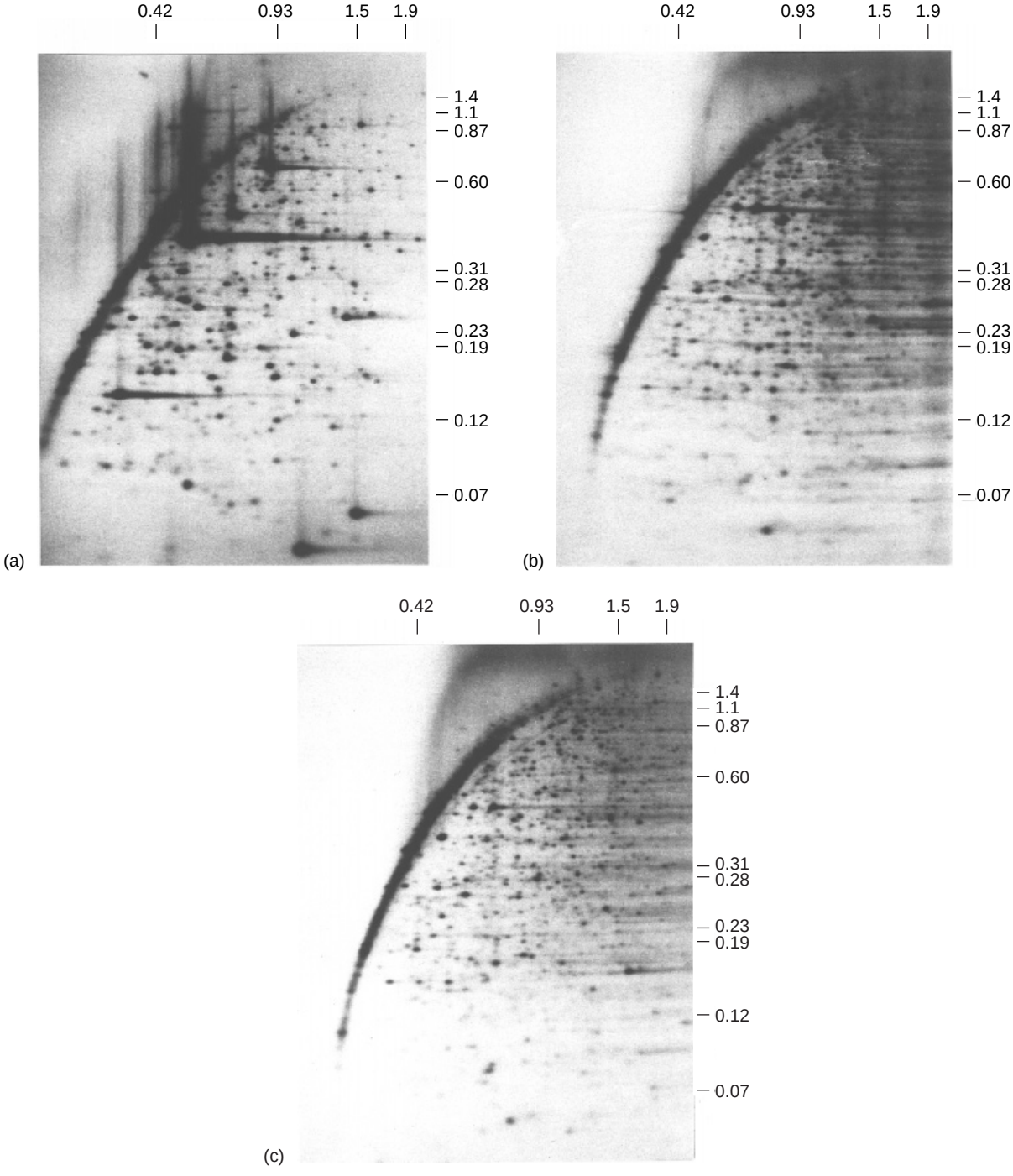

Figure 5 Representative RLCS patterns. Poly(A) RNAs were prepared from mouse liver (a), cerebral cortex (b) and cerebellum (c). For restriction enzyme cleavage both *Bam*HI and *Bgl*II were used together as enzyme E_L, and *Hin*fI was used as enzyme E_M.

average size) obtained from 1 μg of poly(A)$^+$RNA is usually treated with restriction enzyme E_L, followed by labeling, and subjected to 1st-D gel electrophoresis.

Figure 5(a–c) shows typical RLCS patterns resulting from gel electrophoresis of cDNA samples prepared from mouse liver and brain (cerebral cortex and cerebellum). Approximately 500 gel spots are detected for the liver, their pattern differing substantially from that of any brain region. More than 1000 gel spots are observed for the two brain regions (Figure 5b and c), their gel patterns being very similar but not identical. The differences in intensity have been confirmed (using conventional Northern analysis employing cloned spots as probes) to reflect the expression level of each transcript (Figure 6a and b).

Is this RLCS method sufficiently sensitive to detect even very rare mRNA species? As already described, in RLCS 0.2–0.3 μg of cDNA is applied to each gel. Since the average size of the cDNA is 1000 bp, the total copy number of the cDNA is calculated as $[(0.2–0.3) \times (10^{-6})/(660 \times 1000\,\text{bp})] \times (6 \times 10^{23}) = (1.8–2.7) \times 10^{11}$. If the population of very rare mRNA species corresponds to 0.0001% of total mRNA,[10] the cDNA copy number is calculated as $(1.8–2.7) \times 10^{11} \times 0.0001\% = (1.8–2.7) \times 10^5$. By comparison, in RLGS, 1 μg of genomic DNA, corresponding to 3×10^5 copies for mammalian genomes, is applied per gel, and is enough to detect a large number of gel spots. Even spots whose intensity is several times weaker than that of the major spot population are readily detectable. This means that even very rare mRNA species are detected by RLCS.

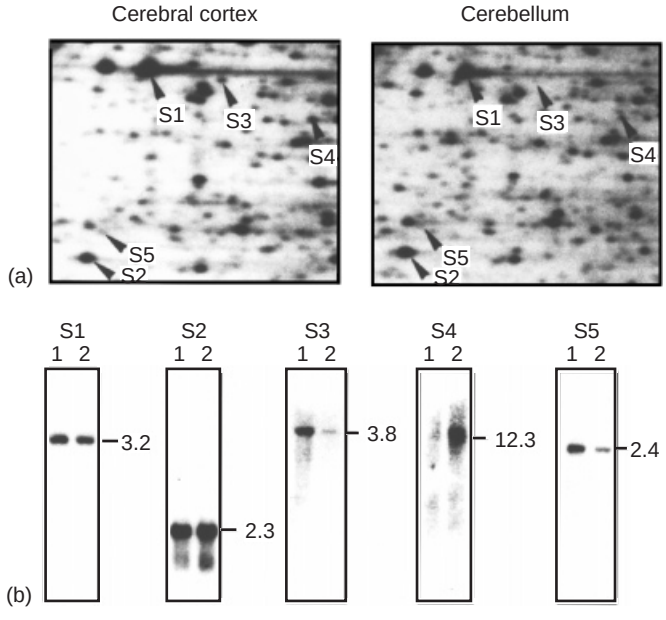

Figure 6 The intensities of the RLCS spots correlate with expression levels of the corresponding mRNA species. (a) Profiles of the cloned spots, indicated by arrowheads. [These RLCS profiles are magnified portions from patterns (b) and (c) in Figure 5.] (b) Northern analysis using cloned spot DNA fragments as probes. 2 μg of poly(A)$^+$RNA from the cerebral cortex (1) and cerebellum (2) were used for each lane. S1–S5 represent spots from which probes are made. Approximate size of each band is shown in kb.

Hence, RLCS is a sensitive quantitative method that enables us to visualize more than 1000 cDNA species at once. However, the display of even more cDNA species as gel spots requires enhancement of the technique.

Recently, RLCS has been further improved and refined as follows. (1) The display of an increased number of gel spots has been demonstrated by using more than two RLCS gels and different restriction enzymes. However, the appearance of redundant spots proved problematic (redundant spots are spots corresponding to the same gene but appearing in two or more loci). To overcome this problem, the combination and order of application of the restriction enzymes are being examined.[11] (2) Accurate priming of reverse transcription reactions using anchored oligo-dT primers has been examined under novel experimental conditions. Theoretically, cDNA species which have different 3′ end sequences flanking poly(A) tails could be separately displayed on the different gels by using two different kinds of anchor primers with MN at the 3′ end. Thus, cDNA having a GC sequence at the 3′ end just upstream from a poly(A) tail should appear only when a GC anchoring primer is used. However, this has proved problematic in practice, because mispriming of the anchor primers causes a single cDNA to be observed in two different 'windows'. To avoid mispriming, we successfully improved the results by performing the reverse transcriptase reaction at a higher temperature in the presence of trehalose and oligonucleotide blockers.[12] By applying these two modifications, many more cDNA species can be displayed.

4 PROTOCOLS

4.1 Introduction

RLGS analysis comprises several steps: DNA extraction, landmark labeling, two-dimensional electrophoresis, spot analysis and spot cloning as described in section 2. Here, we elaborate on several aspects of the RLGS technique.

4.1.1 Choice of Restriction Enzymes, E_L, E_M and E_N

With various combinations of restriction enzymes, e.g. *Bss*HII, *Eco*RV and *Hin*fI, or *Pac*I, *Bam*HI and *Hin*fI, RLGS can screen the physical conditions of the restriction landmarks of the entire genome. Here, we describe

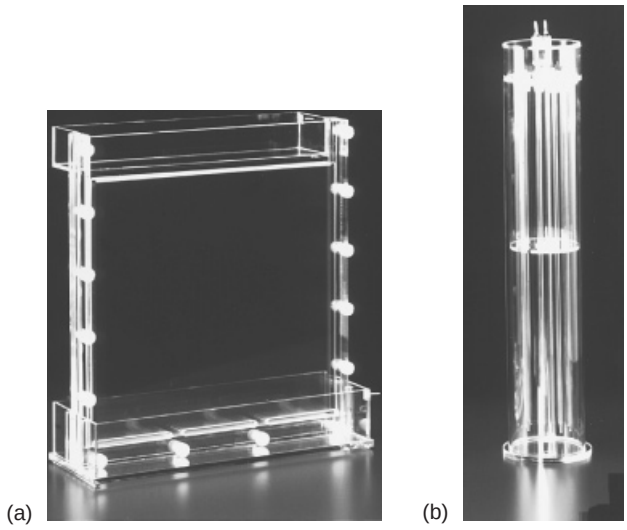

Figure 7 Apparatus for (a) 1st-D and (b) 2nd-D electrophoresis. (Photo: Biocraft Inc. catalog.)

the protocol for the set *Not*I, *Pvu*II and *Pst*I for screening deletion, amplification and methylation of the landmarks in the mammalian genome. The *Not*I recognition sequence, GCGGCCGC, is known to be generally located in the CpG islands,[13] enabling us to screen efficiently amplification and deletion of potential transcriptional units, because almost all CpG islands generally locate in the upstream region of transcriptional units.[14] Furthermore, since *Not*I is one of the methylation-sensitive enzymes, methylation conditions can also be screened.

4.1.2 DNA Extraction, Landmark Labeling and Two-dimensional Electrophoresis

As described in section 2, all procedures are optimized for producing a high-resolution profile with low background noise.

4.1.3 Apparatus for Two-dimensional Electrophoresis

RLGS uses specially designed apparatus manufactured by Biocraft Inc. (Tokyo, Japan) for both 1st-D and 2nd-D gel electrophoresis (Figure 7a and b). In 1st-D gel electrophoresis, an agarose gel noodle serves as the medium. Figure 8(a) and (b) shows the Teflon® tube used in the procedure, encased within an outer glass tube to maintain its shape. About 50 cm of agarose gel and slow electrophoresis facilitate the very detailed resolution of DNA fragments. The electrophoretic tank for the 2nd-D acrylamide gel electrophoresis is designed so that four gels can be processed at once (Figure 9a–d) The required acrylamide gels can also be made in this same electrophoresis apparatus, obviating the need to transfer gels.

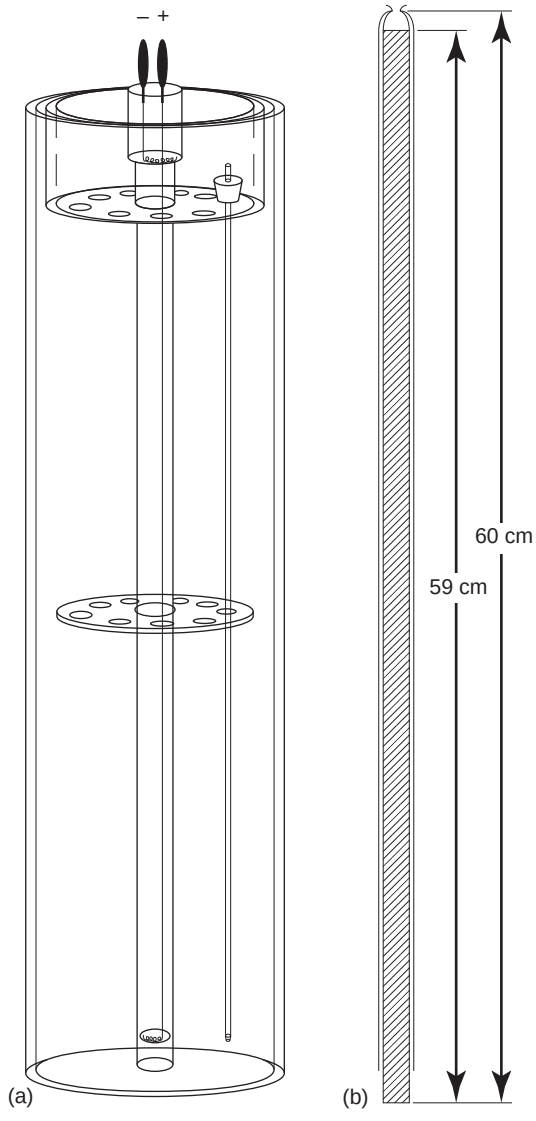

Figure 8 Schematic representation of 1st-D electrophoretic apparatus. (a) Overview. Nine 1st-D agarose disk gels can be subjected to electrophoresis in one run. The upper end of the tank serves as anode and the lower as cathode. (b) Holder for agarose gel noodle. The inner Teflon® tubing is encased in an outer glass support tube (60 cm length).

4.1.4 Spot Cloning

The method used to recover landmark spots on RLGS gels is essential for further studies such as identifying the genes that may be associated with the landmark sites and determining their functions. Target DNA fragments electro-eluted from an RLGS gel can be directly cloned by ligation with an appropriate cloning vector when the genomic DNA of microorganisms of relatively smaller genome size (10^6–10^7 bp) – such as *Staphylococcus aureus* – is subjected to the RLGS system.[15] However, in higher organisms like mice or humans,

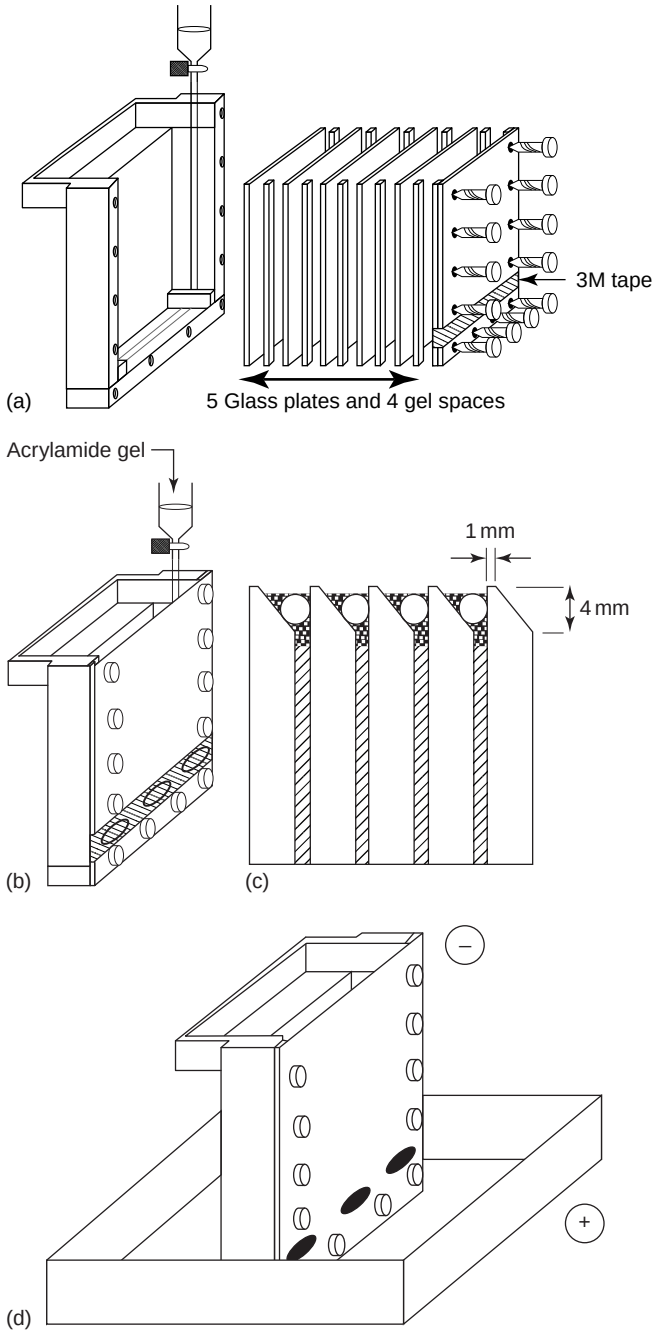

Figure 9 (a) Assembly of glass plates in 2nd-D electrophoretic apparatus. Four gel spaces, each 1 mm wide, are separated by five glass plates. (b) Adding acrylamide gel through Teflon® tubing at one end of the apparatus. Three holes, which serve to pass the electrophoretic current during the run, are temporarily sealed with 3 M Scotch tape. (c) Side view of the attachment of 1st-D agarose gel noodles to the 2nd-D polyacrylamide gel. (d) Fully assembled 2nd-D electrophoretic apparatus. The Scotch tape has been removed for electrophoresis.

which have a larger genome size of 3×10^9 bp, less than 0.75 attomoles (7.5×10^{-19}) of target DNA fragments are available from an RLGS gel spot for each 1.5 μg of sample application, making it necessary to enrich the DNA fragments before cloning. In section 3.4 below, we introduce two methods for spot cloning, a PCR-mediated method and a restriction trapper-mediated method.

4.2 Preparation of Genomic DNA

It is important to isolate genomic DNA with as little damage as possible, because nicks and gaps cause nonspecific incorporation of radioisotopes in the labeling step, resulting in background noise. To avoid any damage, the critical technical point is to suppress DNase activity.

4.2.1 Materials and Reagents

Tris–EDTA (TE) [10 mM Tris-Cl, pH 7.5–1 mM ethylenediaminetetraacetic acid (EDTA)]; phenol–chloroform–isoamyl alcohol (PCI) (50 : 49 : 1), 0.1% 8-hydroxyquinoline, buffered with TE; lysis buffer (150 mM EDTA, 1% sodium dodecyl sulfate (SDS), 10 mM Tris-Cl, pH 8.0); proteinase K solution (Merck, Darmstadt, Germany) [10 mg mL^{-1}, dissolved in distilled water (DW)]; RNase A solution (Boehringer, Mannheim, Germany) (1 mg mL^{-1} dissolved in DW, boiled for 15 min for inactivation of DNase); mortar; pestle; dialysis tube.

4.2.2 DNA Preparation

The mortar, pestle, spatula and aluminum foil are prechilled in liquid nitrogen (LN_2). A 2-mL volume of lysis buffer and 0.5 g of the tissue are frozen in LN_2 and wrapped in aluminum foil, after which the frozen tissue is crushed with a hammer. The tissue pieces are transferred to the mortar and ground into powder with the pestle. The ground tissue powder is transferred to a 50-mL Falcon tube, to which is added 25 mL of lysis buffer and 150 μL of proteinase K solution. Quick but gentle stirring with the spatula gives a viscous sample. This mixture is incubated for 20 min at 55 °C and, after chilling on ice, 25 mL of PCI are added, followed by gentle stirring with a rotator (25 rpm) for 30 min. The sample is then centrifuged at 3000 rpm for 30 min and the resultant aqueous phase transferred to a new Falcon tube. This PCI extraction should be repeated more than twice. After PCI extraction, the aqueous phase is transferred to a dialysis tube and dialyzed overnight against TE. Following dialysis, the sample is transferred to a Falcon tube, to which is added 40 μL of RNase A solution, and then incubated for 2 h at 37 °C. At this stage, the volume of the sample is about 30 mL. The sample is now divided into two Falcon tubes (about 15 mL each), to each of which 35 mL of ethanol are added, and the two are gently rotated (15 rpm) for 30–60 min to precipitate the DNA. After precipitation, the precipitates are retrieved and

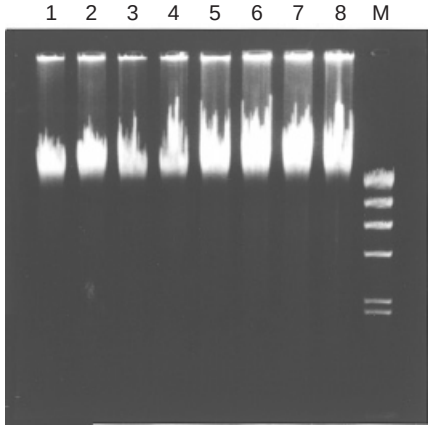

Figure 10 Quality checking of DNA by agarose electrophoresis. Eight samples of mouse genomic DNA, purified for RLGS, have been electrophoresed in 1% agarose gel. DNA fragments shorter than 20 kb are scarcely observed under ethidium bromide staining. M depicts lambda/*Hind*III DNA size marker.

transferred to an Eppendorf tube to which are added 300–500 μL of TE buffer (drying is not necessary) and then kept at 4 °C to dissolve the DNA. The concentration is adjusted to 1 mg mL^{-1}.

Genomic DNA over 200 kb long can be recovered using this protocol. For checking the quality of the DNA, an aliquot portion is subjected to electrophoresis in 1% agarose gel. DNA fragments shorter than 20 kb should scarcely be observed under ethidium bromide staining (Figure 10). If a smeared image of fragmented DNA is seen, the samples are significantly damaged and should not be used for RLGS analysis.

4.3 Labeling of Restriction Landmarks

The labeling protocol is designed so that all reactions can be performed sequentially in a single tube. Manipulation is very easy, involving only the adding of reagents and mixing.

4.3.1 Materials and Reagents

DNA polymerase I (Toyobo); deoxynucleoside [α-thio] triphosphates (dGTP[α]S, dCTP[α]S) (Toyobo); dideoxynucleoside triphosphates (ddGTP, ddATP, ddTTP, ddCTP [Toyobo]); [α-^{32}P]dGTP (3000 Ci $mmol^{-1}$) (Amersham); [α-^{32}P]dCTP (6000 Ci $mmol^{-1}$) (Amersham); *Not*I (Takara); *Pvu*II (Toyobo); *Pst*I (Takara); *Sca*I (Takara); bovine serum albumin (BSA) (Sigma fraction V); Triton X-100; Sequenase® Version 2.0 (USB, Cleveland, OH, USA); 10× high buffer (HB) [500 mM Tris-Cl (pH 7.4), 100 mM $MgCl_2$, 1 M NaCl, 10 mM dithiothreitol (DTT)]; 10× medium buffer (MB) [100 mM Tris-Cl (pH 7.5), 100 mM $MgCl_2$, 10 mM DTT, 500 mM NaCl]; 2.5× SH buffer (0.375 M NaCl, 0.25% Triton X-100, 0.25% BSA); pBSII solution [0.5 μg/8 μL pBluescript II (Stratagene, San Diego, CA, USA) in 1× HB]; pUC18 (0.5 μg μL^{-1} pUC18 in 1× MB); DTT.

4.3.2 Blocking of DNA

For blocking DNA, mix the following reagents:

DNA	10 μg
10× HB	1 μL
1 M DTT	0.1 μL
10 mM dGTP[α]S	0.4 μL
10 mM dCTP[α]S	0.2 μL
10 mM ddATP	0.4 μL
10 mM ddTTP	0.4 μL
DNA polymerase I	0.5 μL
DW	to 10 μL

Mix gently, but completely, with pipeting (since DNA solution is very viscous, solutions must be thoroughly mixed). Incubate for 20 min at 37 °C for digestion and for 30 min at 65 °C for inactivating DNA polymerase I. All solutions except the sample and DNA polymerase I can be premixed and stored at −20 °C for use as blocking buffer.

4.3.3 Landmark Cleavage

Mix the following reagents:

DNA sample (blocked)	10 μL
2.5× SH buffer	8 μL
*Not*I (10 U μL^{-1})	2 μL

Mix completely and incubate for 2 h at 37 °C for complete digestion. It is recommended that digestion be routinely checked for completion. For this purpose, transfer 2 μL of the mixture into another tube, add the control plasmid and incubate in parallel at 37 °C for 2 h.

DNA sample	2 μL
pBSII solution	8 μL

After incubation, electrophorese with 0.8% agarose gel. When digestion is complete, only two clear bands should be seen. Observation of a 2.9-kb band indicates that digestion is only partially complete (Figure 11).

4.3.4 Landmark Labeling

After digestion is complete, add the following reagents for labeling the *Not*I landmarks. (At restriction landmarks generated by *Not*I, a radioisotopic nucleotide is incorporated with DNA polymerase I. Sequenase®, one of the DNA polymerases, is better for suppressing background noise, because it lacks the 3′–5′ exonuclease activity.)

DNA sample (landmark-cleaved)	18 μL
1mM DTT	0.3 μL

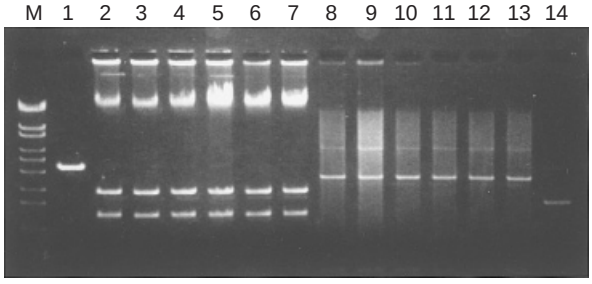

Figure 11 Checking for complete digestion with enzymes E_L and E_M. The test plasmid, linearized pBSII (lane 1), is added to an aliquot part of the E_L (*Not*I) digestion samples, incubated in parallel with the samples, and then electrophoresed in 1% agarose gel. If digestion is complete, only two bands can be seen (lanes 2–7). pUC18 circular plasmid (lane 14) is used in a similar manner to check E_M (*Pvu*II) digestion. Only linearized plasmids of 2.7 kb length are observed in lanes 8–13.

[α-^{32}P]dCTP	1 μL
[α-^{32}P]dGTP	1 μL
Sequenase® Version 2.0 (13 U mL^{-1})	0.1 μL

Mix well by pipeting, and incubate for 30 min at 37 °C.

In the case of 3′ protruding and blunt ends, landmarks can be labeled with [α-^{32}P]ddNTP by terminal deoxy transferase.[4]

4.3.5 Digestion with Restriction Enzyme E_M

The second digestion is performed with *Pvu*II. Add the following reagents to the sample tube:

DNA sample (landmark-labeled)	20.4 μL
1 mM ddGTP	1 μL
1 mM ddCTP	1 μL
100 mM $MgCl_2$	1.2 μL
DW	4.4 μL
*Pvu*II (10 U mL^{-1})	2 μL

Mix gently but completely by pipeting and incubate for 1 h at 37 °C. All solutions except for the sample and *Pvu*II can be premixed and stored at −20 °C for use as the second digestion buffer. As in the case of landmark cleavage, use of an aliquot portion is recommended to check for complete *Pvu*II digestion. For the purpose, 2 μL of the mixture are transferred to another Eppendorf tube and the circular pUC18 control DNA is added, followed by incubation for 1 h at 37 °C (the same as for the sample) and then electrophoresis in 0.8% agarose.

Sample	2 μL
0.1 μg of pUC18 circular plasmid	8 μL

After incubation, the sample is subjected to electrophoresis in 0.8% agarose and checked for digestion bands. If the digestion is complete, only linearized plasmids of 2.7 kb can be observed (Figure 11).

4.3.6 Adjustment of Sample Concentration

A 2-μL volume of the labeled sample is placed in an Eppendorf tube, to which is added 100 μL of salmon sperm DNA solution and 100 μL of ice-cold 20% trichloroacetic acid (TCA). After incubation for 5 min on ice, the precipitate is collected by filtering the solution through a glass microfiber filter. The filter is washed with 9 mL of ice-cold 20% TCA and the radioactivity remaining on the filter is measured with a liquid scintillation counter. The concentration of the sample is then adjusted to 9000 cpm μL^{-1} by the addition of 6× loading dye[10] and the sample is stored until 1st-D electrophoresis at 4 °C.

4.4 Two-dimensional Gel Electrophoresis

4.4.1 Materials and Reagents

All apparatus for both 1st-D and 2nd-D electrophoresis (Figure 7a and b) are manufactured by Biocraft (Tokyo, Japan), including the electrophoretic tanks, 1st-D gel holder (the glass support tube) and the Teflon® tubing.

10× 1st-D buffer (dissolve 242 g of Tris, 109 g of sodium acetate trihydrate, 42 g of NaCl and 23.4 g of EDTA disodium salt in 1.8 L of H_2O and adjust the pH to 8.15 with acetic acid; adjust the volume to 2 L and send through a 0.2-μm filter or autoclave; the pH of the 1× buffer should be 8.0); 10× Tris–boric acid–EDTA (TBE) buffer (dissolve 1210 g of Tris, 770 g of boric acid and 75 g of EDTA disodium salt in H_2O and adjust to 20 L); dye solution (0.25% bromophenol blue (BPB), 0.25% xylenecyanol in TE without glycerol); 0.8% agarose gel solution [dissolve 0.8 g of SeaKem® GTG agarose (FMC Inc.) and 5 g of sucrose in 1× 1st-D buffer]; 5% 2nd-D acrylamide gel solution (dissolve 96.7 g of acrylamide, 3.3 g of methylenebisacrylamide, 1.3 g of ammonium persulfate and 200 mL of 10× TBE in DW and adjust to 2 L with DW; add 800 μL of N,N,N′,N′-tetramethylethylenediamine (TEMED) just before use to solidify the gel).

4.4.2 Preparation for First-dimensional Electrophoresis

Gel holders for the agarose gel noodles are made from Teflon® tubing and glass support tubes, as shown in Figure 12. The hard glass outer cover to the Teflon® tubing not only protects the fragile agarose gel from damage but also makes handling easier. Because the inside diameter of the glass tube is slightly larger than the outside diameter of the Teflon® tubing that will be inserted, one end of the glass tube must be beveled and narrowed by heating and shaping to tightly fix the Teflon® tubing. As shown in Figure 12, the top of the Teflon® tubing is cut at an angle of about 15°, passed through the glass support tube from the unbeveled end and then pulled out about 2 cm at the top using pliers. It

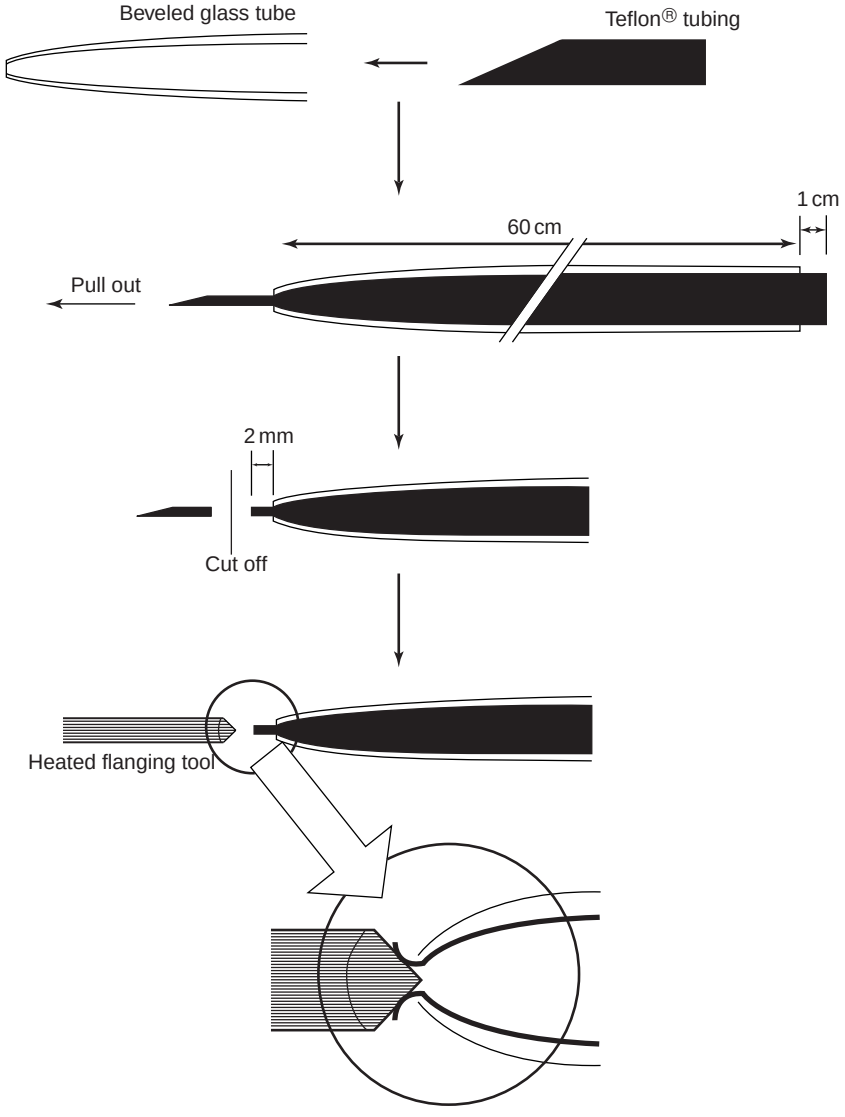

Figure 12 Preparation of Teflon® and glass gel holder for 1st-D gels.

is necessary to ensure that the upper part of the Teflon® tubing is stationary before it is trimmed to leave about 2 mm at the top and about 1 cm at the bottom. Next, the tip of a flanging tool is heated and the top of the tubing pushed on to the heated tip. After a few seconds, the tool is removed and the top of the tubing is pressed against a flat smooth surface such as a sheet of glass.

A 0.8% agarose gel is used to separate DNA fragments in 1st-D electrophoresis. As shown in Figure 13, the 0.8% agarose dissolved in the 1st-D buffer is gradually sucked up into the tube until it reaches a height of 59 cm, that is, about 1 cm from the top, using a syringe fitted with a three-way stopcock. The three-way stopcock is immediately closed and the tube left until the gel solidifies, after which the stopcock is opened and the syringe–stopcock unit carefully removed.

A 350-mL volume of 1× 1st-D buffer is added to the cathodal bottom of the tank and the gel holders are fitted into the anodal top, after which the tank is filled with another 250 mL of 1× 1st-D buffer. Nine samples can be run simultaneously on the described apparatus.

4.4.3 Preparation of Second-dimensional Acrylamide Gel

The 2nd-D electrophoretic tank is assembled as shown in Figure 9(a–d). After sealing the side holes with 3M Scotch tape, the 2nd-D acrylamide gel solution is poured through the side Teflon® tubing until it reaches the top of the glass plates. A 3-mL volume of 2-butanol is gently dropped on top of the gel, which is left to stand for about 2 h. Just before the 2nd-D electrophoresis, the 3M Scotch tape is peeled off and the 2-butanol washed out with 1× TBE.

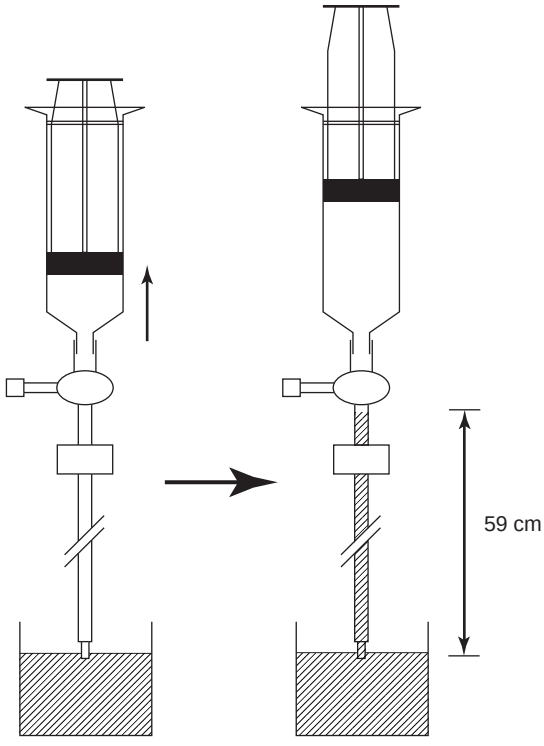

Figure 13 Schematic representation of gel casting. Dissolved agarose is sucked up using a syringe fitted with a three-way stopcock.

4.4.4 Running of First-dimensional Gels

After the gel noodles have been left at room temperature for about 1 h, they are ready for a run. A 15-μL volume of the labeled sample (containing about 140 000 cpm of labeled DNA) is applied to the gel. Electrophoresis is carried out at a constant voltage of 180 V for about 14 h, at which time the BPB dye reaches 50 cm.

4.4.5 Extruding and In Situ Restriction Enzyme Digestion

When the electrophoresis is completed, the anodal tank buffer is removed with an aspirator and the gel holders are taken out. The gel noodles can be expelled from the holders using a 1-mL syringe with a cut-off yellow Eppendorf tip. The gel noodles are cut 3 cm from the end, at the 500 bp point, and marked with BPB dye solution (0.01%) so as to make it easy to work with them. The noodles (about 57 cm in length) are soaked in 50-mL Falcon tubes containing 40 mL of 1× HB (Figure 14a and b) for 10 min. The buffer is changed and the noodles are soaked again for 10 min to reach complete equilibrium.

Digestion is carried out in situ in the Teflon® tubing. For this purpose, the equilibrated gel noodles are sucked up into the tubing, whose internal diameter is slightly bigger than that of the noodles (Figure 15a–c). After

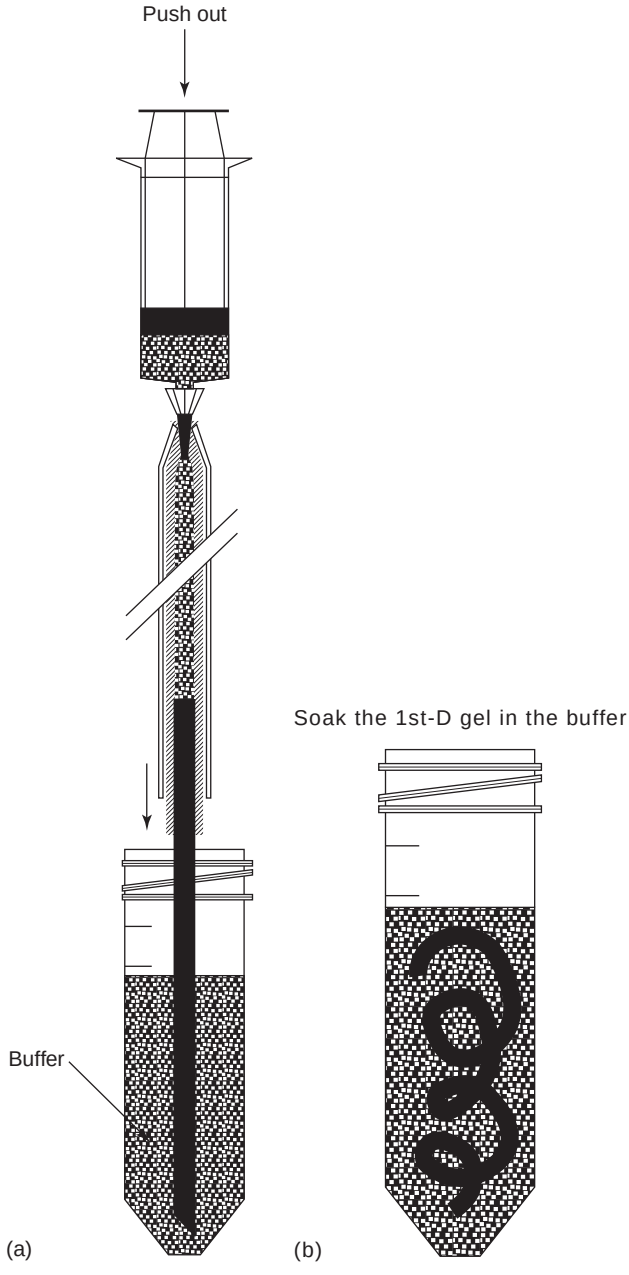

Figure 14 Dialysis of 1st-D agarose gel noodle.

removal of all traces of the buffer, the tubing is filled with *Pst*I solution (1500 units *Pst*I in 1500 μL of 1× HB), and the tubing looped, with one end connected to the other (Figure 16). The looped, Teflon® tubing-enclosed gels are incubated for 2 h at 37 °C.

4.4.6 Running Second-dimensional Gels

After digestion for 2 h with *Pst*I, the gel noodles are expelled from the tubing into 50-mL Falcon tubes containing 40 mL of 1× TBE buffer. After 10 min of equilibration, the noodles are transferred into the glass slits of the 2nd-D electrophoresis apparatus as shown

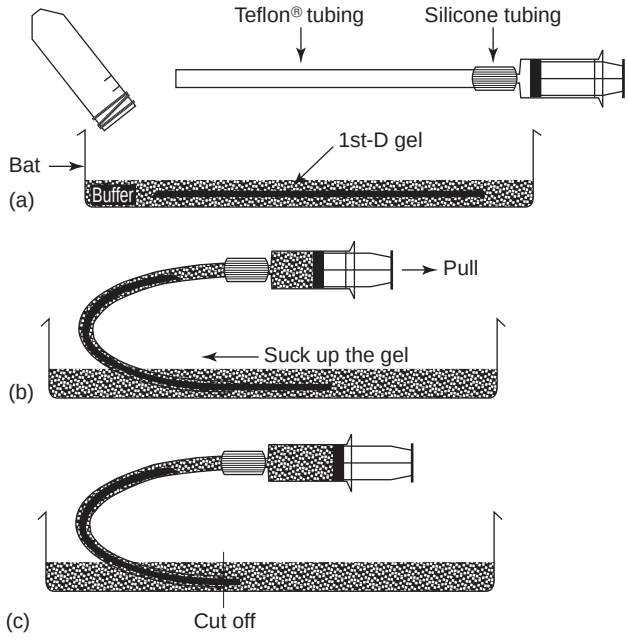

Figure 15 Method for sucking up a 1st-D gel noodle into the Teflon® digestion tube.

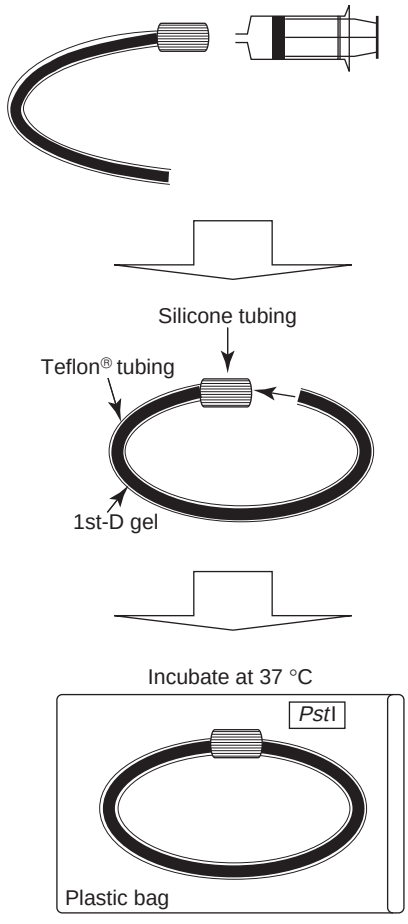

Figure 16 Schematic representation of the sealing of a digestion tube. The sealed tube is incubated for 2 h for complete digestion.

in Figure 17(a) and (b). The space between the agarose and acrylamide gels is filled with 0.8% SeaKem® agarose in 1× TBE containing 0.05% BPB. After the overlaid agarose has solidified, both reservoirs are filled with 1× TBE and electrophoresis carried out at 150 V for about 24 h, at which time the tracking dye BPB will have reached the end of the gels.

4.4.7 Autoradiography

The gels are dried and autoradiographed, usually for 3 days, using Kodak X-AR film.

4.4.8 Restriction Landmark Genomic Scanning Profile

A representative result is shown in Figure 18. In this particular autoradiogram, more than 3000 spots can be separately distinguished.

4.5 Methods for Spot Cloning

4.5.1 Polymerase Chain Reaction-mediated Method

The principles behind the method are illustrated in Figure 19(a) and (b). The method consists of PCR amplification of adapter-ligated spot DNA fragments without the exclusion of similar-sized co-localized DNA fragments, following selective ligation with the *Not*I–dT vector. It can be expected that the similar-sized DNA fragments are present in abundance and therefore will work well as DNA carriers to prevent loss of the target spot DNA through the cloning step. The method is highly advantageous when the total DNA sample prepared for cloning is limited, as only a few micrograms of DNA in all is sufficient for cloning. This method is applicable to any spot DNA derived through RLGS using different sets of restriction enzymes for analysis, simply by changing the adapters, PCR primers and vectors for the subcloning. The method is also applicable to RLCS spot cloning.

4.5.1.1 Apparatus and Reagents *E. coli* DNA ligase (Takara); 10× ligation buffer for *E. coli* DNA ligase [300 mM Tris-Cl (pH 8.0), 40 mM $MgCl_2$, 100 mM $(NH_4)_2SO_4$, 12 mM EDTA, 1 mM NAD, 0.05% BSA]; spin column (Chroma Spin-100™; Clontech); NACS™ column (BRL); 10 μM *Not*I and *Hin*fI adapters (preannealed double-stranded oligonucleotides; 5′-ACGCCAGGGTTTTCCCAGTCACGACGC-3′ and 5′-pGGCCGCGTCGTGACTGGGAAAACCCTGGCGT-3′ for *Not*I site and 5′-pANTCTGTACTGCACCAGCAAATCC-3′ and 5′-GGATTTGCTGGTGCAGTACAG-3′ for *Hin*fI site); forward primer P77 (10 μM; 5′-AGGGTTTTCCCAGTCACGACGCGG-3′); reverse primer Ad2-2 (10 μM; 5′-pTTGCTGGTGCAGTACAGANTC-3′); *Taq* DNA

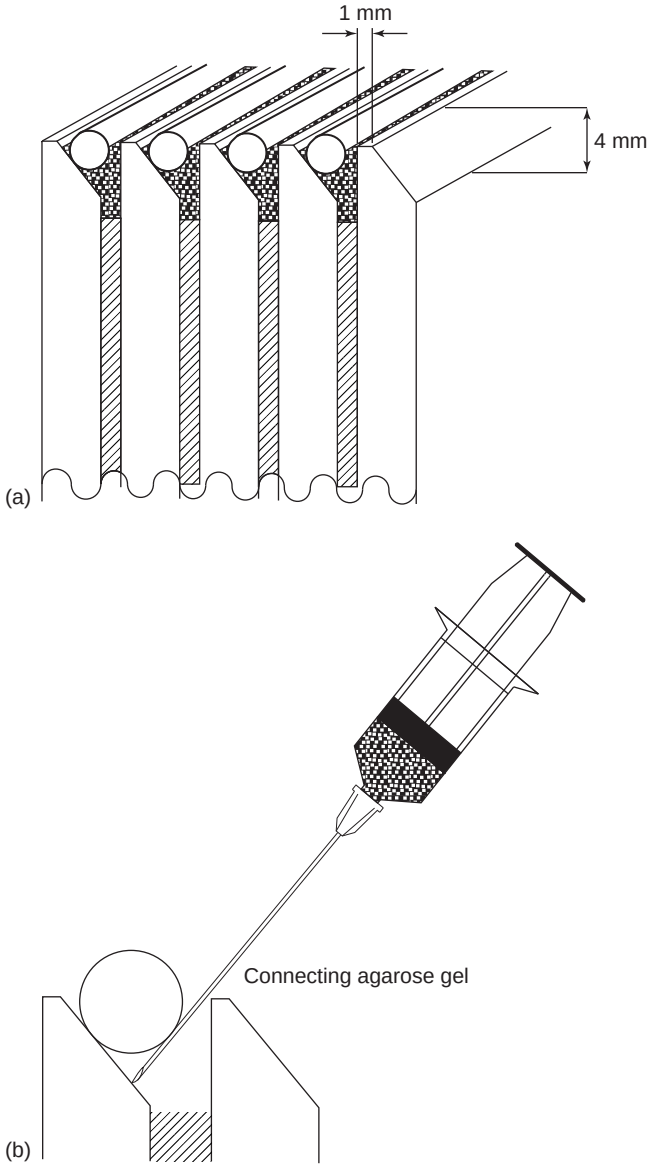

Figure 17 Schematic view of connection of 1st-D gel noodle to 2nd-D acrylamide gel. After laying the agarose gel noodle in one of the glass grooves, dissolved agarose (kept at 60 °C) is injected between the noodle and the acrylamide gel using a 5-mL syringe. Damage to the acrylamide gel surface by the syringe needle must be avoided.

polymerase (Takara); 10× PCR buffer [100 mM Tris-Cl (pH 8.3), 500 mM KCl, 15 mM $MgCl_2$]; dNTP mixture (2.5 mM each of dATP, dCTP, dGTP and dTTP); 50 ng μL^{-1} *Not*I–dT vector (a linearized dephosphorylated plasmid vector with both a *Not*I terminus and 3′-dT protruding terminus); linearly polymerized acrylamide (1% solution); 10× HB; 500 mM Tris-Cl (pH 7.4); T4 DNA ligase (Takara); 10× buffer for T4 DNA ligase [500 mM Tris-Cl (pH 7.5), 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP, 0.1% BSA]; 50% PEG6000.

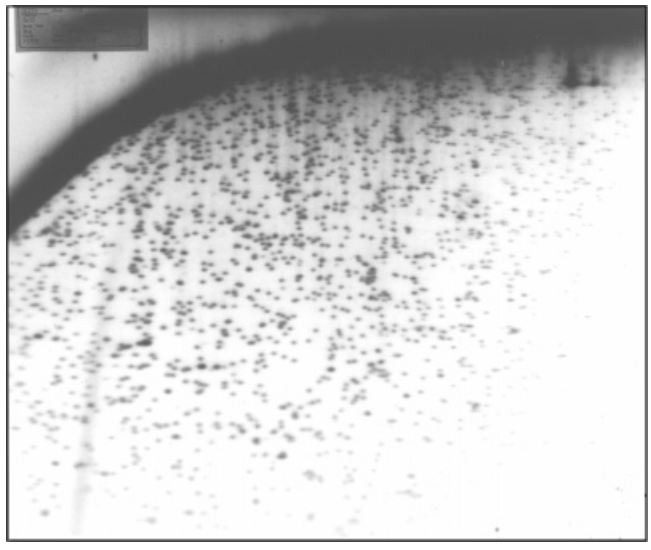

Figure 18 RLGS profile of rat brain tissue. *Not*I, *Pvu*II and *Pst*I were used as E_L, E_M and E_N, respectively.

4.5.1.2 Procedures For the basic operations, such as electroelution, PCI extraction, ethanol precipitation, enzyme treatment and PCR reaction, see the general protocol books.[10] The restriction enzymes *Not*I, *Bam*HI and *Hin*fI are used in this experiment as RLGS restriction enzymes E_L, E_M and E_N, respectively. A 1-μg amount of the sample to be cloned is mixed with the same quantity of the end-labeled sample and subjected to the RLGS procedure. Following autoradiography, the target spots are cut out and their DNA fragments electroeluted and extracted with PCI, followed by ethanol precipitation using 10 μg of linearly polymerized acrylamide as a carrier, and then dissolved in 3 μL of TE.

A third of each DNA fragment is ligated overnight at 16 °C with *Not*I and *Hin*fI adapters.

DNA solution	1 μL
10× buffer for *E. coli* DNA ligase	0.5 μL
Adapter for *Not*I	0.5 μL
Adapter for *Hin*fI	0.5 μL
E. coli DNA ligase	1 μL (30 U)
DW	1.5 μL

After incubation, the ligation mixture is diluted to 40 μL and subjected to a spin column, after which the eluate is PCR-amplified through 30 cycles using the primers P77 and Ad2-2 as follows:

DNA solution (adapter-ligated)	37.5 μL
dNTP mix	5 μL
10× PCR buffer	5 μL
P77 primer	1 μL
Ad2-2 primer	1 μL
Taq DNA polymerase	0.5 μL

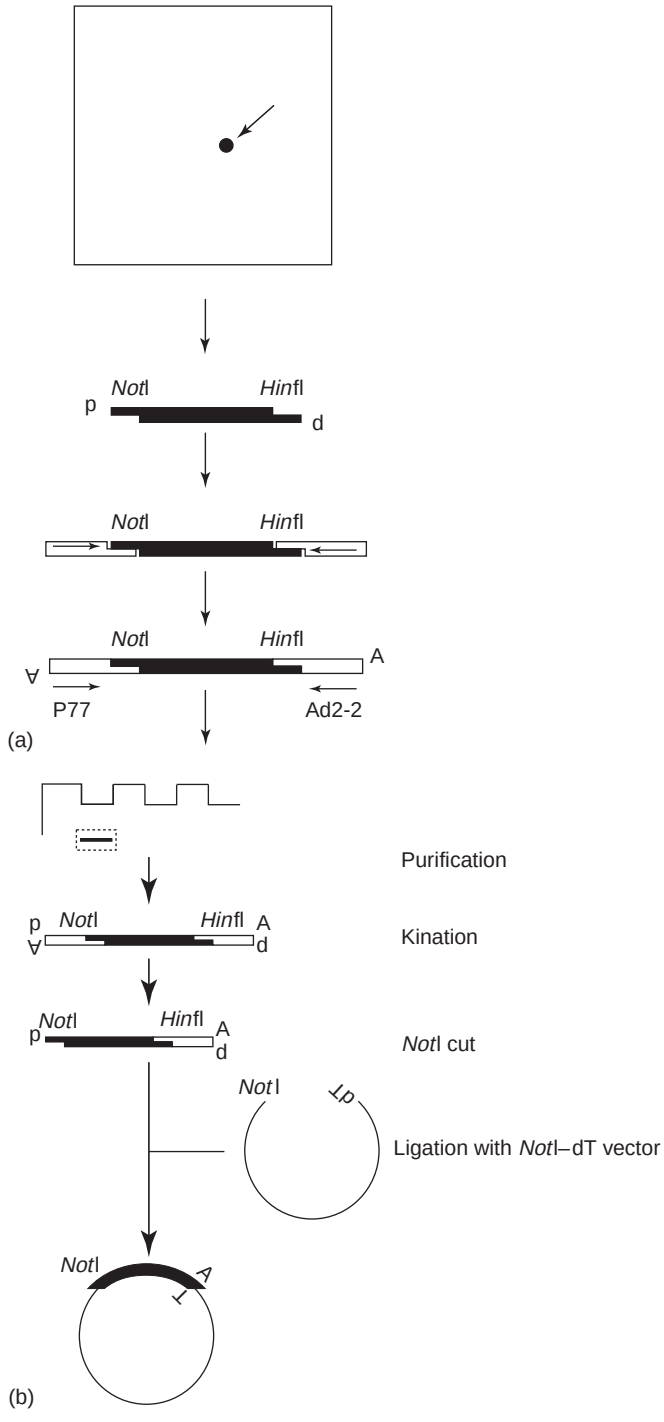

Figure 19 Schematic representation of PCR-mediated method of RLGS spot cloning.

The conditions for PCR amplification are 94 °C for 1 min, 60 °C for 1.5 min and 72 °C for 2 min. The PCR products are directly applied to the polyacrylamide gel electrophoresis and stained with ethidium bromide. The amplified bands are cut out and electroeluted, followed by purification through an NACS™ column. After ethanol precipitation, the resultant DNA is suspended in 10 μL of TE.

The amplified DNA is kinated and digested with *Not*I as follows:

DNA solution (amplified)	10 μL
10× kination buffer	5 μL
100 mM ATP	1 μL
DW	30 μL
T4 DNA polynucleotide kinase	4 μL

After incubation for 60 min at 37 °C, the DNA is extracted with PCI, precipitated with ethanol and suspended in 5 μL of TE, followed by *Not*I digestion.

DNA solution (kinated)	5 μL
10× HB	1 μL
*Not*I	1 μL (10 U)
DW	3 μL

Following *Not*I incubation for 60 min at 37 °C, the DNA is again extracted once with PCI, condensed through ethanol precipitation and suspended in 3.5 μL of TE. The resultant *Not*I-digested DNA is then ligated with *Not*I–dT vector:

DNA solution (*Not*I-digested)	3.5 μL
*Not*I–dT vector	2 μL
10× buffer for T4 DNA ligase	1 μL
T4 DNA ligase	1 μL (350 U μL^{-1})
50% PEG6000	1 μL
DW	1.4 μL

The product is incubated overnight at 16 °C and transformed into the appropriate *E. coli*, ex HB101.

4.5.1.3 Results We have succeeded in cloning much target spot DNA by the PCR-mediated method.[9,11,16,17] Almost 70% of the target DNA fragments have been successfully cloned. One of the disadvantages of the method is that it proved impossible to clone spot DNA fragments of more than 1000 bp in size. This may have been due to inefficient amplification of the target DNA fragments by PCR reaction. Nonetheless, the method has also been applied to cloning the spots from the RLCS system. Several target DNA fragments (genes) have already been cloned.

4.5.2 Restriction Trapper-mediated Method

The principles behind enrichment of target DNA fragments using a restriction trapper are illustrated in Figure 20. The restriction trapper, the key material for this process, consists of a hairpin-looped oligolinker covalently linked to the surface of latex beads (Figures 21 and 22). The method is based on purification of the target

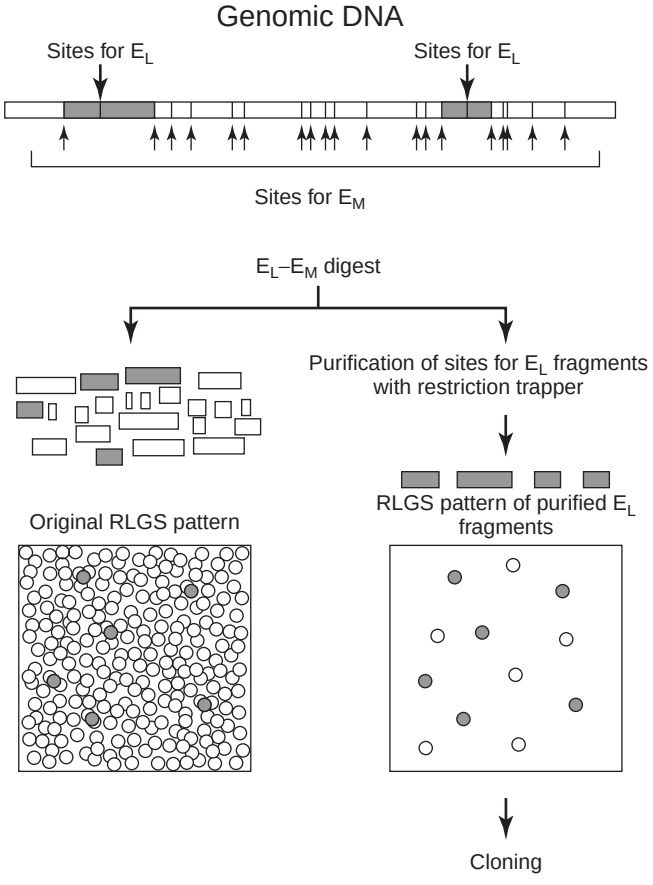

Figure 20 Procedure for RLGS spot cloning with restriction trapper mediation. The presence of many unlabeled invisible fragments in the target spot of conventional RLGS gels prevents specific cloning. Eluting DNA fragments from the RLGS gel with restriction trapper purifies the DNA, which can then be cloned.

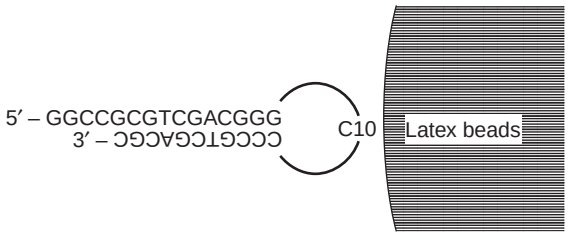

Figure 21 Structure of the restriction trapper for *Not*I fragments. A 36-base oligonucleotide is covalently linked to the surface of latex beads via a 10-base cytosine stretch. This oligonucleotide carries the *Not*I protruding end.

DNA fragments by the restriction trapper through ligation at their restriction landmark sites. Purified target DNA fragments sufficient for cloning can be obtained by increasing the initial amount of genomic DNA. It is therefore necessary to prepare several hundred micrograms of the genomic DNA and the rather specialized

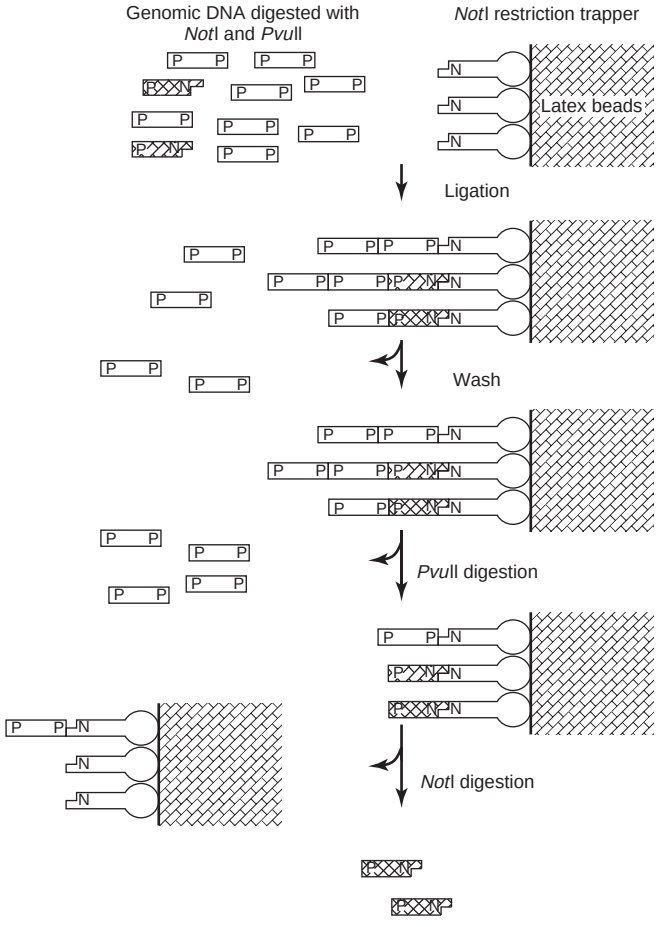

Figure 22 Principles behind the purification of a target DNA fragment with restriction trapper.

restriction trapper. The method is nevertheless useful because (1) the spot DNA fragments are evenly enriched with restriction trapper, independent of fragment size, and (2) DNAs of many spots can be enriched and cloned simultaneously.

4.5.2.1 Apparatus and Reagents 10× HB [500 mM Tris-Cl (pH 7.4), 100 mM $MgCl_2$, 1 M NaCl, 10 mM DTT]; 0.1% BSA; 0.1% Triton X-100; 10× low buffer (LB) [100 mM Tris-Cl (pH 7.5), 100 mM $MgCl_2$, 10 mM DTT]; *Not*I restriction trapper, pre-washed once with TE and adjusted to 5% solution with TE (Gene Trapper R-*Not*I; Japan Synthetic Rubber); T4 DNA ligase (Takara); 10× buffer for T4 DNA ligase [500 mM Tris-Cl (pH 7.5), 100 mM ATP, 50% PEG6000]; 10× MB [100 mM Tris-Cl (pH 7.5), $MgCl_2$, 10 mM DTT, 500 mM NaCl]; 3000 Ci mmol^{-1} [α-^{32}P]dGTP; 6000 Ci mmol^{-1} [α-^{32}P]dCTP; Sequenase® Version 2.0

4.5.2.2 Procedures The restriction enzymes *Not*I, *Pvu*II and *Pst*I are used in this experiment as RLGS restriction enzymes E_L, E_M and E_N, respectively. A 1-mg

amount of genomic DNA is prepared (10 reactions, each of 100 μg) and completely digested with *Not*I and *Pvu*II.

DNA	100 μg
10× HB	20 μL
0.1% BSA	20 μL
0.1% Triton X-100	20 μL
*Not*I (8 U μL^{-1})	10 μL
DW	to 200 μL

The *Not*I-digested DNA is incubated for 2 h at 37 °C and then further digested by *Pvu*II by addition of the following agents:

DNA (*Not*I-digested)	200 μL
10× LB	20 μL
*Pvu*II (12 U μL^{-1})	10 μL
DW	170 μL

The reaction mixture is incubated for 2 h at 37 °C, extracted with PCI followed by ethanol precipitation, and dissolved in 50 μL of TE.

Ligation to the restriction trapper and purification are then carried out.

DNA solution (*Not*I–*Pvu*II-digested)	50 μL
*Not*I restriction trapper	20 μL
10× buffer for T4 DNA ligase	15 μL
100 mM ATP	2 μL
50% PEG6000	30 μL
T4 DNA ligase (350 U μL^{-1})	4 μL
DW	14 μL

The mixture is incubated for 2 h at 18 °C with occasional mixing. After inactivation of the ligase at 65 °C for 15 min, the restriction trapper is recovered by centrifugation at 15 000 *g* for 5 min followed by washing with TE. The trapper is then treated with *Pvu*II.

DNA (restriction trapper)	
10× LB	45 μL
*Pvu*II (12 U μL^{-1})	10 μL
DW	395 μL

After incubation for 1 h and washing twice with TE, the trapper is then treated with *Not*I.

DNA (restriction trapper)	
10× HB	10 μL
DW	60 μL
0.1% BSA	10 μL
0.1% Triton X-100	10 μL
*Not*I (10 U μL^{-1})	10 μL

The mixture is incubated for 2 h, at 37 °C, mixing every 30 min. After centrifugation, the target DNA fragments (*Not*I–*Pvu*II fragments) released from the trapper in supernatant are extracted with PCI, precipitated with ethanol and dissolved in 16 μL of TE. A small portion of the DNA fragments purified by the above method is end-labeled in the presence of [α-^{32}P]dGTP and [α-^{32}P]dCTP with Sequenase® Version 2.0.

DNA (*Not*I–*Pvu*II fragments)	3 μL
10× MB	1 μL
[α-^{32}P]dGTP	1 μL
[α-^{32}P]dCTP	1 μL
Sequenase® Version 2.0	1 μL
DW	3 μL

The mixture is incubated for 30 min at 37 °C.

The now radiolabeled DNA fragments are mixed with the remaining unlabeled DNA fragments and the whole is subjected to RLGS. After autoradiography, the DNA fragments in the target spots are recovered by electroelution, ligated with *Not*I–*Pst*I-cleaved plasmid vector and transformed into appropriate *E. coli* by the conventional method.

4.5.2.3 Results Using this restriction trapper-mediated method, several target DNA fragments have been successfully cloned from RLGS gel spots.[18–22]

5 APPLICATION OF RESTRICTION LANDMARK GENOMIC SCANNING

5.1 Potential Applications

RLGS, utilizing the concept of restriction landmarks, has various potential applications in biological research. In general, two major categories of application can be listed, the first being scanning of copy numbers of restriction landmarks across an entire genome. Based on genetic polymorphism or sequence variations between individuals at the restriction landmarks, the entire genome map can be rapidly constructed by high-throughput RLGS Version 1.8.[6–8,22–25] Furthermore, since RLGS needs no DNA probes or STS primers for detecting polymorphisms, a genetic map of any organism, e.g. the Syrian hamster, can be determined.[8] These genetic maps facilitate the analysis of positional cloning of genes responsible for genetic diseases.[26,27]

The second category of application is scanning of genome-wide methylation conditions, so as to survey both endogenously imprinted genes and potentially expressed genes, based on the correlation of methylation with gene expression. This method of scanning for methylation is named 'RLGS for methylation' (RLGS-M). RLGS-M is very powerful, because more than 2000 loci can be surveyed for their methylation status in a

single procedure. Using this approach, our group has already found and reported some imprinted genes in mice.[21,28,29] By way of illustration, in the next section we describe the application of RLGS-M to the scanning of methylation status associated with liver cancer in mice.

5.2 Application of Restriction Landmark Genomic Scanning to Cancer Research

5.2.1 Scanning Methylation State with Restriction Landmark Genomic Scanning Using Methylation-sensitive Endonuclease

Restriction landmark sites across an entire genome can be visualized and screened using RLGS techniques. If methylation-sensitive endonucleases are used for landmark cleavage, the appearance of RLGS spots varies depending on the methylation status of the landmark sites. When a landmark site is methylated and thus resistant to digestion, the corresponding RLGS spot disappears. Hence the RLGS technique is applicable to the scanning of the methylation state of landmark sites. We have named this approach 'RLGS for methylation' (RLGS-M) and have used it to screen for methylated loci/genes[28,30] which could contribute to tumorigenesis.

In RLGS-M analysis, the intensity of a spot reflects the copy number of a corresponding landmark site, because the number of end-labeled molecules at landmark cleavage sites corresponds to the level of incorporated radioactivity (Figure 23). Uniform levels of spot intensity are observed in diploid tissues; this is consistent with the cleavage and end-labeling of both alleles for most landmark sites. However, when methylation-sensitive enzymes are used to identify the landmark cleavage sites, the spots are reduced to half- or zero-intensity, respectively, when one or both of the alleles are methylated (Figure 23). GC-rich enzymes such as *Not*I (GCGGCCGC) or *Bss*HII (GCGCGC) cleave sequences that tend to be located on CpG islands.[31] It has been estimated that approximately 89% of *Not*I sites and 74% of *Bss*HII sites are on CpG islands. Hence the methylation status of CpG islands[13,32] can be evaluated using *Not*I or *Bss*HII to identify landmarks. Moreover, since methylation has been correlated with gene expression, it has been possible to screen *Not*I and *Bss*HII sites for allele-specific patterns of methylation.

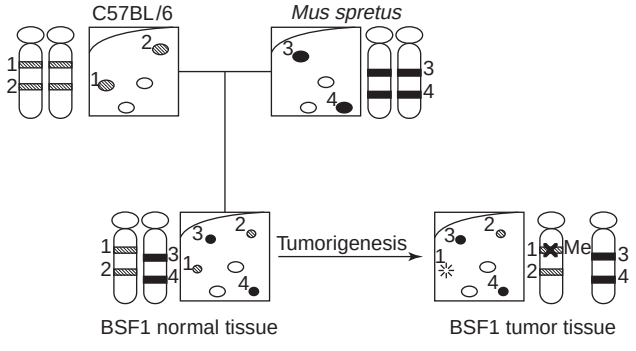

Figure 23 Principle of RLGS-M analysis to detect changes in methylation pattern in tumor DNAs. Spots 1 and 2 are C57BL/6 (B) specific, whereas spots 3 and 4 represent *Mus spretus* (S)-specific polymorphic alleles at the same loci. Empty circles indicate nonpolymorphic spots. In the F1 normal tissue of a cross between B and S (BSF1), polymorphic spots appear as haploid (half-intensity) spots and nonpolymorphic spots appear as diploid (full-intensity) spots. In tumor DNA from a BSF1 animal, DNA methylation at the landmark site of a polymorphic spot (B-specific spot 1) leads to the absence of the spot in the RLGS profile. Spot loss can be analyzed by a comparison of normal and tumor RLGS DNA profiles. [Reproduced from Akama et al.[30] by permission of the American Association for Cancer Research, Inc.]

5.2.2 Overview of Cancer

Tumorigenesis is a multi-step process involving both epigenetic and genetic alterations.[31,33] The identification of genetic mutations, which cause either increased functioning of proto-oncogenes or loss of functions of tumor suppressor genes, provided a rationale for understanding tumorigenesis. However, the mutation of a single proto-oncogene or tumor suppressor gene is usually not sufficient in itself to cause neoplastic growth, tumor progression depending on secondary events. The genetic targets for these secondary events could be expected to depend on the initiating event and also developmental and tissue-specific factors regulating cell proliferation which arise during cell proliferation. Additional 'steps' involving angiogenesis, invasive growth and metastasis generate more serious life-threatening[34] disease. Although the 'cast of characters' involved in these processes is large and growing, our understanding of them is complicated by the large number of genes involved, developmental and tissue-specific differences in growth regulation and the stochastic nature of the process.

Alterations in DNA methylation are consistently found associated with tumorigenesis and may play a variety of roles in tumor progression.[35,36] In general, tumor DNA is hypomethylated, while hypomethylation is associated with increased gene expression. Several proto-oncogenes have been shown to be hypomethylated in tumors.[35,37] By contrast, regional hypermethylation may promote tumorigenesis by silencing expression of tumor suppressor genes, the genes necessary for differentiation and cessation of growth,[35,38,39] or hormone receptors which have a role in the growth or properties of the tumor.[40,41] In addition, increased

methylation may be mutagenic since 5′-methylcytosine can deaminate spontaneously to T, resulting in a C to T transition.[35,36] Results indicating that expression of an exogenous DNA methyltransferase can induce transformation of NIH/3T3 cells supports the contention that DNA hypermethylation plays an important role in tumor progression.[42]

In recent years, a number of new technologies have been developed which can be used for rapidly screening tumors for genetic alteration on a genome-wide basis. Most of these methods depend on nucleic acid hybridization to detect gain or loss of DNA sequences. These methods include comparative genomic hybridization,[43] representational difference analysis,[44] PCR analysis of simple sequence length polymorphisms (SSLP)[45] and RLGS.[5] Although each method has its advantages and disadvantages, the RLGS method is potentially powerful in terms of the number of loci that it is capable of screening simultaneously, its facilitation of direct cloning of genetic regions which are altered during tumorigenesis[20] and its ability to detect and localize regional hypermethylation or hypomethylation changes.

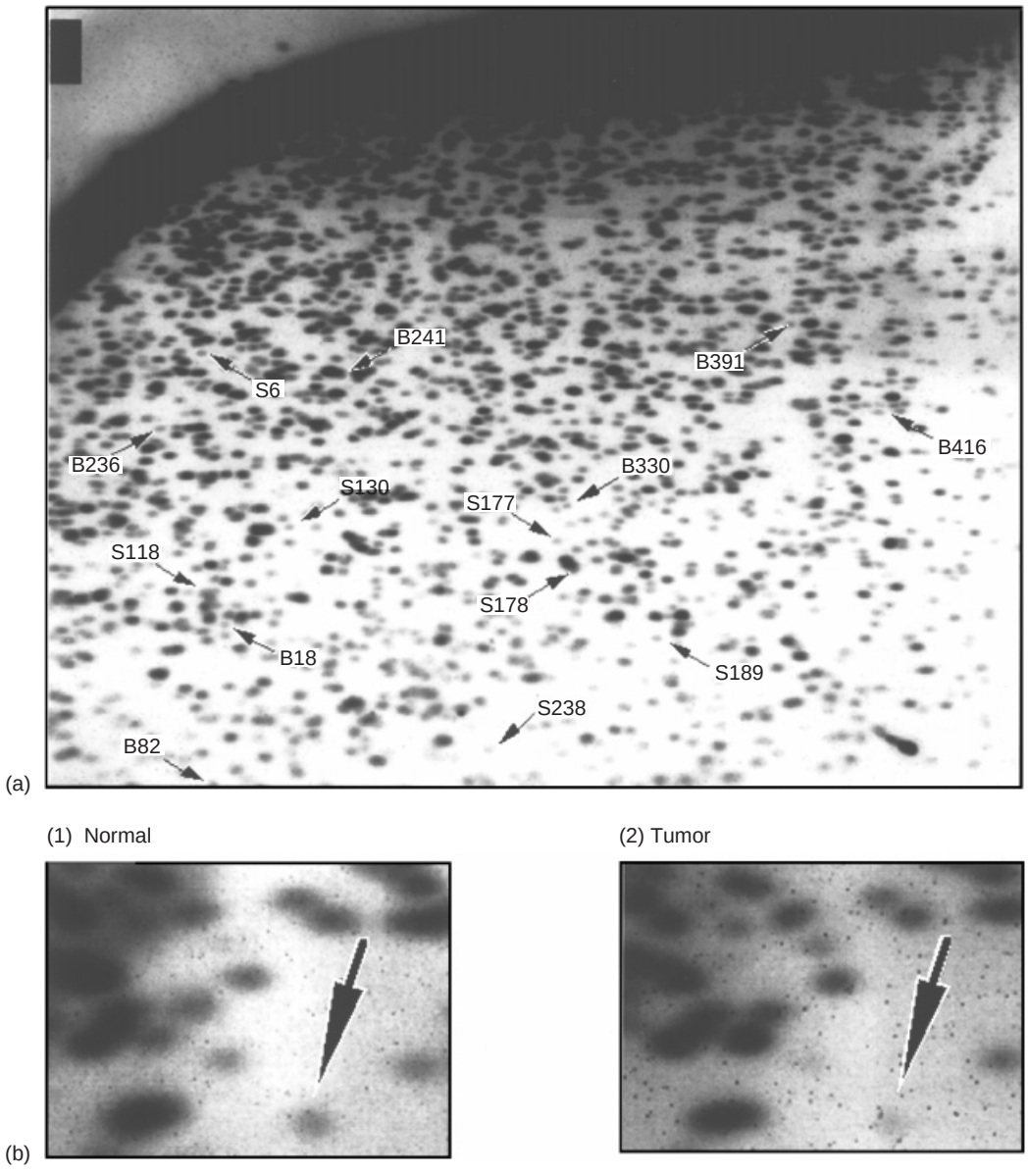

Figure 24 RLGS-M profile of normal BSF1 liver DNA produced by the enzyme combination *Not*I, *Pvu*II and *Pst*I. (a) Fourteen spots showing loss (methylation) in more than 75% of 30 hepatoma samples are indicated by arrows. (b) A magnified section of the normal RLGS DNA profile containing spot S238 – indicated by the arrow – (1) is compared with the same part from a profile of tumor DNA (2). [Reproduced from Akama et al.[30] by permission of the American Association for Cancer Research, Inc.]

5.2.3 *Analysis of Mouse Liver Tumors for Alterations in DNA Methylation Status Using Restriction Landmark Genomic Scanning Using Methylation-sensitive Endonuclease*

The principles by which RLGS-M analysis is used to detect DNA methylation status is shown in Figure 23 through a schematic representation of the RLGS patterns of normal tissues from parents and F1 progeny, as well as of F1 tumor tissue. Polymorphic paternal- (S) and maternal- (B) specific genes are transmitted to F1 progeny by Mendelian inheritance, and show up under general RLGS analysis as half-intensity spots. Using a methylation-sensitive enzyme, such as *Not*I, we can detect alterations of the DNA methylation status at such sites throughout the genome. When comparing tumor DNA relative to control DNA, methylation of a *Not*I site in the tumor DNA results in spot loss, whereas demethylation of a *Not*I site results in the appearance of a new spot. An example of the disappearance of the spot due to methylation in a tumor is illustrated (see Figures 23 and 24b). In order to help characterize the alterations of methylation status in tumor DNA, we utilized interspecific F1 hybrid, strain-specific haploid spots whose chromosomal locations are known through genetic analysis.[24] The 575 loci/spots (340 B- and 235 S-specific) were mapped by RLGS analysis (using *Not*I, *Pvu*II and *Pst*I) of BSS backcross progeny.[24] Using the new RLGS Version 1.8,[7,46] 507 of the mapped spots provided useful information for the rapid analysis and localization of genomic alterations such as changes in methylation at landmark sites.

RLGS-M analysis was performed on 30 liver tumor DNA samples from 14 interspecific F1 hybrid mice offspring of S and a transgenic B57BL/6 (B6) line containing the SV40 T antigen early region connected to a mouse major urinary protein (MUP) enhancer/promoter.[30] In cancer tissues, changes in the intensity of the RLGS spots may reflect DNA loss, rearrangement or amplification, and also alterations in methylation as described above. Random alterations would be expected to occur due to genomic instability in the course of tumorigenesis, but some alterations may be primary aberrations that promote tumorigenesis. These primary genomic aberrations would be expected to occur at a high frequency in particular tumor types. We believe that altered loci detected in 22 of the 30 tumor samples (75%) may be significantly associated with tumorigenesis. In all, 24 spots were detected which showed a loss or reduction in intensity in 75% of the tumor DNA samples (see Table 2). The loss of these spots could have resulted from either of two primary causes: loss of hetrozygosity (LOH) (i.e. DNA loss) or methylation of the landmark *Not*I site. To distinguish between these two possibilities, we combined the information on spot intensity with that on chromosomal location. We assumed that chromosomal regions which displayed more than two contiguous loci/spot losses were likely to reflect the deletion of DNA as we had found previously,[20] whereas solitary loci/spot losses probably reflected methylation at the landmark sites. Confirmation of this assumption required the cloning of the spots and analysis of tumor-specific alterations by means of Southern blotting, using the spot clones as probes. By this means, we determined that loss or reduction in intensity of 14 solitary spots resulted from landmark methylation and, furthermore, that this methylation occurred in more than 75% of the tumor samples. These 14 spots are shown with their spot identification numbers in Figure 24(a). Part of the RLGS gel containing a representative spot, S238, is shown (indicated by the arrow) under magnification in Figure 24(b). The intensity of S238 is clearly reduced in tumor tissue. Characterizations of the 14 spots which displayed such reduced intensity are presented in Table 3. Locus number is designated as D#Rik- or D#Ncvs-. Each of the 14 spots were classified as solitary losses. Although four of the loci, D5Rik119, D5Ncvs4, D5Rik122 and D5Rik124, are located on chromosome 5, losses at these loci are all solitary. The localization of these four spots, however, is restricted to the centromeric half of the chromosome, indicating that a high frequency of methylation sites associated with tumorigenesis may be localized in this area.

Table 2 Numbers of RLGS loci exhibiting spot loss or reduction of spot intensity in tumor DNA[a]

Tumor DNA exhibiting loss or reduction in intensity of spots (%)	Number of loci
76–100	24
51–75	53
26–50	137
0–25	293

[a] Approximately 20 tumors were analyzed for each spot. [Reproduced from Akama et al.[30] by permission of the American Association for Cancer Research, Inc.]

5.2.4 *Isolation of Spot Clones and DNA Sequence Analyses*

In order to characterize further the nature of the genomic alterations associated with spot loss at the loci in question, and potentially to identify a gene at a given locus, we cloned 13 of the 14 spots using the restriction trapper-based RLGS spot-cloning method described in section 3.4.2.[30] DNA sequence analysis and a homology search against existing DNA databases revealed that two

Table 3 Solitary RLGS loci/spots with spot loss or reduction (methylation) in more than 75% of the mouse liver tumor samples

Locus	Spot	Methylation frequency (methylated/sample)	Sequence homology
D4Rik141	S6	29/30 (97%)	
D4Rik136	S238	24/27 (89%)	p16/CDKN2
D5Rik119	S189	29/30 (97%)	
D5Ncvs4	B18	27/30 (90%)	
D5Rik122	S178	23/30 (77%)	
D5Rik124	S130	23/29 (90%)	mac25
D7Rik89	S177	26/27 (96%)	
D7Ncvs6	B247	26/30 (87%)	
D7Ncvs20	B82	23/29 (79%)	
D10Rik53	S118	20/23 (87%)	
D12Ncvs7	B236	26/30 (90%)	
D13Ncvs10	B416	26/30 (87%)	
D15Ncvs1	B330	27/30 (90%)	
D19Ncvs1	B391	29/30 (97%)	

of the cloned spots, S130 and S238, were identical with mouse mac25 and the mouse tumor suppressor gene p16/CDKN2, respectively.

5.2.5 *Southern Blot Analysis Using Spot Clone S238 as a Probe*

To examine the methylation status of p16/CDKN2 in tumor tissues, Southern blot analysis was performed using the cloned spot (S238) as a probe. Representative results are shown in Figure 25. The S238 probe used for hybridization is identical with the distal half of the first exon of p16/CDKN2. The S238 probe detects a *Pvu*II polymorphism between B6 and S in normal liver DNA (Figure 25, lanes 1 and 2). When these *Pvu*II–*Pvu*II fragments are digested with *Not*I, the S allele alone is cleaved into a 3.9-kb band (lanes 1 and 3), Thus, the *Not*I site from the spot clone S238 (*Not*I–*Pst*I fragment)

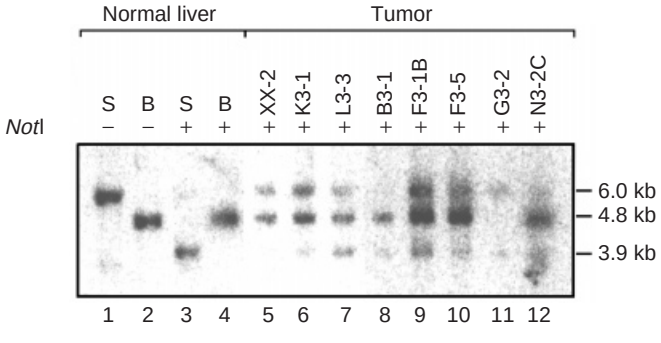

Figure 25 Southern blot analysis of normal liver DNA from B, S and mouse liver tumor DNAs from BSF1 animals. The DNAs were digested with *Pvu*II and *Not*I as indicated (+ or −). The size of each band is noted to the right and lane numbers are shown below. [Reproduced from Akama et al.[30] by permission of the American Association for Cancer Research, Inc.]

is present in the S allele but absent from the B6 allele. Lanes 5–12 are DNA from tumor samples digested with both *Not*I and *Pvu*II. In tumor XX-2 and K3-1 (lanes 5 and 6), the *Not*I–*Pvu*II 3.9-kb S allele is of reduced intensity whereas the 6.0-kb *Pvu*II–*Pvu*II S band is of increased intensity, indicating that the *Not*I site of the S allele has become methylated. Tumors L3-3, F3-1B and F3-5 show both methylated and unmethylated S bands of almost the same intensity (lanes 7, 9 and 10), suggesting that 50% of the S allele of the p16/CDKN2 gene has been methylated. Tumors B3-1 and N3-2C have no methylated 6.0-kb S band (lanes 8 and 12) but the intensity of the unmethylated 3.9-kb band is reduced, implying that the S allele of the p16/CDKN2 gene has been partially lost in this tumor sample. In tumor G3-2, both the methylated and unmethylated S bands are of reduced intensity and the B6 band is completely absent (lane 11), indicating that the S allele of the p16/CDKN2 gene is partially methylated and the B6 allele deleted.

5.2.6 *Methylation of p16/CDKN2 and Tumorigenesis*

Spot clone S238 was found to contain sequences identical to exon 1 of the p16/CDKN2 (INK4a, MTS1) gene. Homozygous and heterozygous deletion and also point mutations of this gene on 9p21 are commonly observed in human cancers.[47–49] Several reports also indicate that methylation of a 5′ CpG island which includes exon 1 is associated with transcriptional silencing of p16 in human tumors.[50–54] The product of the p16/CDKN2 gene, p16, binds to CDK4 and CDK6, preventing their interaction with cyclin D and subsequent cell cycle progression. The CDKN2 gene also encodes an unrelated protein, p19, which arises from an alternative reading frame. This protein (p19) is also capable of inducing arrested growth although the mechanism is not understood.[55]

Southern blot analysis of DNAs from liver tumors, using the S238 spot clone, indicated that the S *Not*I site within exon I was completely or almost completely methylated in 19 of 24 samples. Five samples showed partial methylation (Figure 25, lanes 8 and 12) or a combination of partial methylation and partial gene loss (lane 11). It is not known whether the tumors are heterogenous with respect to alterations at this locus. This region (exon 1) corresponds to the same region of the p16/CDKN2 gene which is hypermethylated in human tumors and results in transcriptional silencing.[54] Currently, it is not clear whether methylation silences p16 expression in the liver tumors analyzed in this study.

Since the liver tumors studied were induced as a result of SV40 T antigen expression, the Rb gene is presumably 'inactive'. P16 is considered to be upstream of pRb since it normally forms complexes with CDK4 and/or CDK6, inhibiting their interaction with cyclin D

and the subsequent phosphorylation of pRb necessary for cell cycle progression from G1 to S.[56] Hence it is difficult to perceive what the consequence of loss of p16 expression would be in an Rb-negative cell. Indeed, ectopic expression of p16 in Rb-negative cells does not result in G1 arrest.[56] Recent experiments indicate that p16 mRNA levels are normally elevated in Rb-negative cells or in SV40 T antigen-transformed cells, suggesting that Rb is normally a transcriptional repressor of p16.[57]

What, then, is the significance of p16 methylation and its presumed transcriptional silencing in the SV40 T antigen-induced liver tumors? Clearly, more work is necessary to elucidate the consequences of p16 methylation in this system. Recent studies indicate that D-type cyclins interact with a novel myb-like transcription factor, DMP1.[58] This suggests that cyclin D-dependent kinases may regulate gene expression through a mechanism independent of Rb. This pathway may function to link regulation of some genes with the cell cycle but may still be susceptible to regulation via p16–CDK4/CDK6 interaction. The methylation of p16 in the SV40 T antigen-induced liver tumors may perturb this pathway.

We conclude that RLGS-M can be used for a genome-wide search for tumor-related alterations in the DNA methylation status. RLGS-M can detect alterations in known tumor suppressor genes and provides a means for detecting and cloning novel genomic regions which have alterations in methylation status related to tumor progression.

ACKNOWLEDGMENTS

We are grateful to Drs M. Muramatsu and A. Wada for their encouragement, and also thank N. Kazuta and Y. Shigemoto for their secretarial assistance. We thank G. Chapman for his encouragement. This work was supported by a Research Grant for the Genome Exploration Research Project from the Science and Technology Agency to Y.H.

ABBREVIATIONS AND ACRONYMS

B	C57BL/6
B6	B57BL/6
BPB	Bromophenol Blue
BSA	Bovine Serum Albumin
ddNTP	2′,3′-dideoxyribonucleoside triphosphate
DTT	Dithiothreitol
DW	Distilled Water
EDTA	Ethylenediaminetetraacetic Acid
HB	High Buffer
LB	Low Buffer
LN_2	Liquid Nitrogen
LOH	Loss of Hetrozygosity
MB	Medium Buffer
MUP	Major Urinary Protein
PCI	Phenol–Chloroform–Isoamyl Alcohol
PCR	Polymerase Chain Reaction
RLCS	Restriction Landmark cDNA Scanning
RLGS	Restriction Landmark Genomic Scanning
RLGS-M	Restriction Landmark Genomic Scanning for Methylation
S	*Mus spretus*
SDS	Sodium Dodecyl Sulfate
SNP	Single Nucleotide Polymorphism
SSLP	Simple Sequence Length Polymorphisms
TBE	Tris–Boric Acid–EDTA
TCA	Trichloroacetic Acid
TE	Tris–EDTA
TEMED	N,N,N′,N′-tetramethylethylenediamine

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Comparative Genomics: Differential Display and Subtractive Hybridization • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Fluorescence In Situ Hybridization • Genome Physical Mapping Using BACs • Optical Mapping in Genomic Analysis • Polymerase Chain Reaction and Other Amplification Systems • Radiation Hybrid Mapping • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

REFERENCES

1. 'The *C. elegans* Sequencing Project. ''How the Worm was Won'' ', *Trends Genet.*, **15**(2), 51–58 (1999).
2. J. Battey, E. Jordan, D. Cox, W. Dove, 'An Action Plan for Mouse Genomics', *Nature Genet.*, 2173–2175 (1999).
3. The Sanger Centre, 'Toward a Complete Human Genome Sequence', *Genome Res.*, **8**(11), 1097–1108 (1998).
4. Y. Hayashizaki, S. Watanabe, *Restriction Landmark Genomic Scanning (RLGS)*, Springer, Berlin, 1997.
5. I. Hatada, Y. Hayashizaki, S. Hirotsune, H. Komatsubara, T. Mukai, 'A Genomic Scanning Method for Higher Organisms Using Restriction Sites as Landmarks', *Proc. Natl. Acad. Sci. USA*, **88**(21), 9523–9527 (1991).
6. Y. Hayashizaki, S. Hirotsune, Y. Okazaki, H. Shibata, A. Akasako, M. Muramatsu, J. Kawai, T. Hirasawa, S. Watanabe, T. Shiroishi, K. Moriwaki, B.A. Taylor, Y. Matsuda, R.W. Elliott, K.F. Manly, V.M. Chapman, 'A Genetic Linkage Map of the Mouse Using Restriction Landmark Genomic Scanning (RLGS)', *Genetics*, **138**(4), 1207–1238 (1994).

7. Y. Okazaki, H. Okuizumi, N. Sasaki, T. Ohsumi, J. Kuromitsu, H. Kataoka, M. Muramatsu, A. Iwadate, N. Hirota, M. Kitajima, C. Plass, V.M. Chapman, Y. Hayashizaki, 'A Genetic Linkage Map of the Mouse Using an Expanded Production System of Restriction Landmark Genomic Scanning (RLGS Ver. 1.8)', *Biochem. Biophys. Res. Commun.*, **205**, 1922–1929 (1994).
8. Y. Okazaki, H. Okuizumi, T. Ohsumi, O. Nomura, S. Takada, M. Kamiya, N. Sasaki, Y. Matsuda, M. Nishimura, O. Tagaya, M. Muramatsu, Y. Hayashizaki, 'A Genetic Linkage Map of the Syrian Hamster and Localization of Cardiomyopathy Locus on Chromosome 9qa2.1-b1 Using RLGS Spot-mapping', *Nature Genet.*, **13**(1), 87–90 (1996).
9. H. Suzuki, T. Yaoi, J. Kawai, A. Hara, G. Kuwajima, S. Watanabe, 'Restriction Landmark cDNA Scanning (RLCS): a Novel cDNA Display System Using Two-dimensional Gel Electrophoresis', *Nucleic Acids Res.*, **24**(2), 289–294 (1996).
10. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning. A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
11. K. Mayumi, T. Yaoi, J. Kawai, S. Kojima, S. Watanabe, H. Suzuki, 'Improved Restriction Landmark cDNA Scanning and Its Application to Global Analysis of Genes Regulated by Nerve Growth Factor in PC12 Cells', *Biochim. Biophys. Acta*, **139**, 910–918 (1998).
12. Y. Mizuno, P. Carninci, Y. Okazaki, M. Tateno, J. Kawai, H. Amanuma, M. Muramatsu, Y. Hayashizaki, 'Increased Specificity of Reverse Transcription Priming by Trehalose and Oligo-blockers Allows High-efficiency Window Separation of mRNA Display', *Nucleic Acids Res.*, **27**(5), 1345–1349 (1999).
13. S. Lindsay, A.P. Bird, 'Use of Restriction Enzymes to Detect Potential Gene Sequences in Mammalian DNA', *Nature*, **327**(6120), 336–338 (1987).
14. J. Kawai, K. Hirose, S. Fushiki, S. Hirotsune, N. Ozawa, A. Hara, Y. Hayashizaki, S. Watanabe, 'Comparison of DNA Methylation Patterns Among Mouse Cell Lines by Restriction Landmark Genomic Scanning', *Mol. Cell. Biol.*, **14**(11), 7421–7427 (1994).
15. T. Fujimura, K. Murakami, 'Increase of Methicillin Resistance in *Staphylococcus Aureus* Caused by Deletion of a Gene Whose Product is Homologous to Lytic Enzymes', *J. Bacteriol.*, **179**(20), 6294–6301 (1997).
16. J. Kawai, H. Suzuki, C. Taga, A. Hara, S. Watanabe, 'Correspondence of RLGS-M Spot Behavior with Tissue Expression on Mouse', *Biochem. Biophys. Res. Commun.*, **213**(3), 967–974 (1995).
17. H. Suzuki, J. Kawai, C. Taga, N. Ozawa, S. Watanabe, 'A PCR-mediated Method for Cloning Spot DNA on Restriction Landmark', *DNA Res.*, **1**(5), 245–250 (1994).
18. W. Miwa, J. Yasuda, Y. Murakami, K. Yashima, K. Sugano, T. Sekine, A. Kono, S. Egawa, K. Yamaguchi, Y. Hayashizaki, T. Sekiya, 'Isolation of DNA Sequences Amplified at Chromosome 19q13.1-q13.2 Including the AKT2 Locus in Human Pancreatic Cancer', *Biochem. Biophys. Res. Commun.*, **225**(3), 968–974 (1996).
19. T. Ohsumi, Y. Okazaki, S. Hirotsune, H. Shibata, M. Muramatsu, H. Suzuki, C. Taga, S. Watanabe, Y. Hayashizaki, 'A Spot Cloning Method for Restriction Landmark Genomic Scanning', *Electrophoresis*, **16**(2), 203–209 (1995).
20. T. Ohsumi, Y. Okazaki, H. Okuizumi, K. Shibata, T. Hanami, Y. Mizuno, T. Takahara, N. Sasaki, M. Ueda, M. Muramatsu, K.A. Kerns, V.M. Chapman, W.A. Held, Y. Hayashizaki, 'Loss of Heterozygosity in Chromosomes 1, 5, 7 and 14 in Mouse Hepatoma Detected by Systematic Genome-wide Scanning Using RLGS Genetic Map', *Biochem. Biophys. Res. Commun.*, **212**(2), 632–639 (1995).
21. C. Plass, H. Shibata, I. Kalcheva, L. Mullins, N. Kotelevtseva, J. Mullins, R. Kato, H. Sasaki, S. Hirotsune, Y. Okazaki, W.A. Held, Y. Hayashizaki, V.M. Chapman, 'Identification of *Grf1* on Mouse Chromosome 9 as an Imprinted Gene by RLGS-M', *Nature Genet.*, **14**(1), 106–109 (1996).
22. S. Hirotsune, K. Hirose, H. Kataoka, J. Kuromitsu, Y. Furuichi, M. Muramatsu, Y. Matsuda, Y. Hayashizaki, 'Spot Mapping on the Standard Profile of Restriction Landmark Genomic Scanning (RLGS) of Sorted Chromosome 20 Using Methylation-insensitive Enzyme', *Genomics*, **24**(3), 593–596 (1994).
23. S. Hirotsune, T. Takahara, N. Sasaki, H. Imoto, Y. Okazaki, T. Eki, Y. Murakami, M. Abe, K. Furuya, M. Muramatsu, Y. Eto, V.M. Chapman, Y. Hayashizaki, 'Construction of High-resolution Physical Maps from Yeast Artificial Chromosomes Using Restriction Landmark Genomic Scanning (RLGS)', *Genomics*, **37**(1), 87–95 (1996).
24. H. Okuizumi, Y. Okazaki, T. Ohsumi, T. Hanami, Y. Mizuno, M. Muramatsu, Y. Hayashizaki, C. Plass, V.M. Chapman, 'A Single Gel Analysis of 575 Dominant and Codominant Restriction Landmark Genomic Scanning Loci in Mice Interspecific Backcross Progeny', *Electrophoresis*, **16**(2), 253–260 (1995).
25. H. Okuizumi, Y. Okazaki, T. Ohsumi, Y. Hayashizaki, C. Plass, V.M. Chapman, 'Genetic Mapping of Restriction Landmark Genomic Scanning Loci in the Mouse', *Electrophoresis*, **16**(2), 233–240 (1995).
26. V. Nigro, Y. Okazaki, A. Belsito, G. Piluso, Y. Matsuda, L. Politano, G. Nigro, C. Ventura, C. Abbondanza, A.M. Molinari, D. Acampora, M. Nishimura, Y. Hayashizaki, G.A. Puca, 'Identification of the Syrian Hamster Cardiomyopathy Gene', *Hum. Mol. Genet.*, **6**(4), 601–607 (1997).
27. S. Hirotsune, T. Takahara, N. Sasaki, K. Hirose, A. Yoshiki, T. Ohashi, M. Kusakabe, Y. Murakami, M. Muramatsu, S. Watanabe, K. Nakao, M. Katsuki, Y. Hayashizaki, 'The Reeler Gene Encodes a Protein with an EGF-like Motif Expressed by Pioneer Neurons', *Nature Genet.*, **10**(1), 77–83 (1995).

28. Y. Hayashizaki, H. Shibata, S. Hirotsune, H. Sugino, Y. Okazaki, N. Sasaki, K. Hirose, H. Imoto, H. Okuizumi, M. Muramatsu, H. Komatsubara, T. Shiroishi, K. Moriwaki, M. Katsuki, N. Hatano, H. Sasaki, T. Ueda, N. Mise, N. Takagi, C. Plass, V.M. Chapman, 'Identification of an Imprinted U2af Binding Protein Related Sequence on Mouse Chromosome 11 Using the RLGS Method', *Nature Genet.*, **6**(1), 33–40 (1994).
29. H. Shibata, K. Yoshino, M. Muramatsu, C. Plass, V.M. Chapman, Y. Hayashizaki, 'The Use of Restriction Landmark Genomic Scanning to Scan the Mouse Genome for Endogenous Loci with Imprinted Patterns of Methylation', *Electrophoresis*, **16**(2), 210–217 (1995).
30. T.O. Akama, Y. Okazaki, M. Ito, H. Okuizumi, H. Konno, M. Muramatsu, C. Plass, W.A. Held, Y. Hayashizaki, 'Restriction Landmark Genomic Scanning (RLGS-M)-based Genome-wide Scanning of Mouse Liver Tumors for Alterations in DNA Methylation Status', *Cancer Res.*, **57**(15), 3294–3299 (1997).
31. S.A. Aaronson, 'Growth Factors and Cancer', *Science*, **254**(5035), 1146–1153 (1991).
32. S. Watanabe, J. Kawai, S. Hirotsune, H. Suzuki, K. Hirose, C. Taga, N. Ozawa, S. Fushiki, Y. Hayashizaki, 'Accessibility to Tissue-specific Genes from Methylation Profiles of Mouse Brain Genomic DNA', *Electrophoresis*, **16**(2), 218–226 (1995).
33. R.A. Weinberg, 'Tumor Suppressor Genes', *Science*, **254**(5035), 1138–1146 (1991).
34. J.M. Adams, S. Cory, 'Transgenic Models of Tumor Development', *Science*, **254**(5035), 1161–1167 (1991).
35. P.W. Laird, R. Jaenisch, 'DNA Methylation and Cancer', *Hum. Mol. Genet.*, **314**, 87–95 (1994).
36. J.L. Counts, J.I. Goodman, 'Alterations in DNA Methylation May Play a Variety of Roles in Carcinogenesis', *Cell*, **83**(1), 13–15 (1995).
37. J. Vachtenheim, I. Horakova, H. Novotna, 'Hypomethylation of CCGG Sites in the 3′ Region of H-ras Protooncogene', *Cancer Res.*, **54**(5), 1145–1148 (1994).
38. N. Ohtani-Fujita, T. Fujita, A. Aoike, N.E. Osifchin, P.D. Robbins, T. Sakai, 'CpG Methylation Inactivates the Promoter Activity of the Human Retinoblastoma Tumor-suppressor Gene', *Oncogene*, **8**(4), 1063–1067 (1993).
39. J.G. Herman, F. Latif, Y. Weng, M.I. Lerman, B. Zbar, S. Liu, D. Samid, D.S. Duan, J.R. Gnarra, W.M. Linehan, S.B. Baylin, 'Silencing of the VHL Tumor-suppressor Gene by DNA Methylation in Renal Carcinoma', *Proc. Natl. Acad. Sci. USA*, **91**(21), 9700–9704 (1994).
40. Y.L. Ottaviano, J.P. Issa, F.F. Parl, H.S. Smith, S.B. Baylin, N.E. Davidson, 'Methylation of the Estrogen Receptor Gene CpG Island Marks Loss of Estrogen Receptor Expression in Human Breast Cancer Cells', *Cancer Res.*, **54**(10), 2552–2555 (1994).
41. J.P. Issa, Y.L. Ottaviano, P. Celano, S.R. Hamilton, N.E. Davidson, S.B. Baylin, 'Methylation of the Oestrogen Receptor CpG Island Links Ageing and Neoplasia in Human Colon', *Nature Genet.*, **7**(4), 536–540 (1994).
42. J. Wu, J.P. Issa, J. Herman, D.E. Bassett, Jr, B.D. Nelkin, S.B. Baylin, 'Expression of an Exogenous Eukaryotic DNA Methyltransferase Gene Induces Transformation of NIH 3T3 Cells', *Proc. Natl. Acad. Sci. USA*, **90**(19), 8891–8895 (1993).
43. O.P. Kallioniemi, A. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman, D. Pinkel, 'Comparative Genomic Hybridization: a Rapid New Method for Detecting and Mapping DNA Amplification in Tumors', *Semin. Cancer Biol.*, **4**(1), 41–46 (1993).
44. N.A. Lisitsyn, N.M. Lisitsina, G. Dalbagni, P. Barker, C.A. Sanchez, J. Gnarra, W.M. Linehan, B.J. Reid, M.H. Wigler, 'Comparative Genomic Analysis of Tumors: Detection of DNA Losses and Amplification', *Proc. Natl. Acad. Sci. USA*, **92**(1), 151–155 (1995).
45. W. Dietrich, H. Katz, S.E. Lincoln, H.S. Shin, J. Friedman, N.C. Dracopoli, E.S. Lander, 'A Genetic Map of the Mouse Suitable for Typing Intraspecific Crosses', *Genetics*, **131**(2), 423–447 (1992).
46. Y. Okazaki, H. Okuizumi, N. Sasaki, T. Ohsumi, J. Kuromitsu, N. Hirota, M. Muramatsu, Y. Hayashizaki, 'An Expanded System of Restriction Landmark Genomic Scanning (RLGS Ver. 1.8)', *Electrophoresis*, **16**(2), 197–202 (1995).
47. G.I. Shapiro, J.E. Park, C.D. Edwards, L. Mao, A. Merlo, D. Sidransky, M.E. Ewen, B.J. Rollins, 'Multiple Mechanisms of p16INK4A Inactivation in Non-small Cell Lung', *Cancer Res.*, **55**(24), 6200–6209 (1995).
48. S.Y. Zhang, A.J. Klein-Szanto, E.R. Sauter, M. Shafarenko, S. Mitsunaga, T. Nobori, D.A. Carson, J.A. Ridge, T.L. Goodrow, 'Higher Frequency of Alterations in the p16/CDKN2 Gene in Squamous Cell', *Cancer Res.*, **54**(19), 5050–5053 (1994).
49. L. Xu, D. Sgroi, C.J. Sterner, R.L. Beauchamp, D.M. Pinney, S. Keel, K. Ueki, J.L. Rutter, A.J. Buckler, D.N. Louis, J.F. Gusella, V. Ramesh, 'Mutational Analysis of CDKN2 (MTS1/p16ink4) in Human Breast Carcinomas', *Cancer Res.*, **54**(20), 5262–5264 (1994).
50. K.W. Lo, S.T. Cheung, S.F. Leung, A. van Hasselt, Y.S. Tsang, K.F. Mak, Y.F. Chung, J.K. Woo, J.C. Lee, D.P. Huang, 'Hypermethylation of the p16 Gene in Nasopharyngeal Carcinoma', *Cancer Res.*, **56**(12), 2721–2725 (1996).
51. J.F. Costello, M.S. Berger, H.S. Huang, W.K. Cavenee, 'Silencing of p16/CDKN2 Expression in Human Gliomas by Methylation and Chromatin Condensation', *Cancer Res.*, **56**(10), 2405–2410 (1996).
52. M. Gonzalez-Zulueta, C.M. Bender, A.S. Yang, T. Nguyen, R.W. Beart, J.M. Van Tornout, P.A. Jones, 'Methylation of the 5′ CpG Island of the p16/CDKN2 Tumor Suppressor Gene', *Cancer Res.*, **55**(20), 4531–4535 (1995).
53. J.G. Herman, A. Merlo, L. Mao, R.G. Lapidus, J.P. Issa, N.E. Davidson, D. Sidransky, S.B. Baylin, 'Inactivation of the CDKN2/p16/MTS1 Gene is Frequently Associated with Aberrant DNA Methylation in all Common Human Cancers', *Cancer Res.*, **55**, 4525–4530 (1995).

54. A. Merlo, J.G. Herman, L. Mao, D.J. Lee, E. Gabrielson, P.C. Burger, S.B. Baylin, D. Sidransky, '5′ CpG Island Methylation is Associated with Transcriptional Silencing', *Nature Med.*, **1**(7), 686–692 (1995).
55. D.E. Quelle, F. Zindy, R.A. Ashmun, C.J. Sherr, 'Alternative Reading Frames of the INK4a Tumor Suppressor Gene Encode Two Unrelated Proteins Capable of Inducing Cell Cycle Arrest', *Cell*, **83**(6), 993–1000 (1995).
56. C.J. Sherr, 'Cancer Cell Cycles', *Science*, **274**(5293), 1672–1677 (1996).
57. Y. Li, M.A. Nichols, J.W. Shay, Y. Xiong, 'Transcriptional Repression of the D-type Cyclin-dependent Kinase', *Cancer Res.*, **54**(23), 6078–6082 (1994).
58. H. Hirai, C.J. Sherr, 'Interaction of D-type Cyclins with a Novel Myb-like Transcription', *Mol. Cell. Biol.*, **16**(11), 6457–6467 (1996).

RNA Tertiary Structure

Eric Westhof and Pascal Auffinger
Institute of Molecular and Cellular Biology, Strasbourg, France

1 Introduction 1
2 Chemical Structure of RNA 1
3 Definitions of Secondary and Tertiary Structures 5
4 RNA Tertiary Motifs 5
5 Instrumental Methods 5
5.1 Physical Methods 6
5.2 Phylogenetic Approach 7
5.3 Computer Modeling 8
6 Conclusions 8
Abbreviations and Acronyms 9
Related Articles 9
References 9

Ribonucleic acids are negatively charged polymers assembled from four different types of monomers. Each monomer is made of an invariant phosphorylated sugar to which is attached one of the four standard nucleic acid bases; the pyrimidines uracil and cytosine, and the purines guanine and adenine. The first level of organization is thus the sequence of bases attached to the sugar–phosphate backbone. In salty water, the RNA molecules fold back on themselves via Watson–Crick base pairing between the bases (A with U, G with C or U) leading to double-stranded helices interrupted by single-stranded regions in internal loops or hairpin loops. The enumeration of the base-paired regions or helices constitutes a description of the second level of organization, the secondary structure. The methods available to deduce the secondary structure of an RNA molecule are mainly of three types: the phylogenetic approach, the theoretical prediction, and chemical/enzymatic methods. The secondary structure of an RNA molecule is experimentally accessible and its content measurable. Under appropriate conditions, structured RNA molecules undergo a transition to a three-dimensional (3D) fold in which the helices and the unpaired regions are precisely organized in space. This folding process usually depends on the presence of divalent ions, such as magnesium ions, and on the temperature. The tertiary structure is the level of organization relevant for biological function of structured RNA molecules.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Ribonucleic acids are negatively charged polymers assembled from four different types of monomers. Each monomer is made of an invariant phosphorylated sugar to which is attached one of the four standard nucleic acid bases; the pyrimidines uracil and cytosine, and the purines guanine and adenine. The first level of organization is thus the sequence of bases attached to the sugar–phosphate backbone. In salty water, the RNA molecules fold back on themselves via Watson–Crick base pairing between the bases (A with U, G with C or U) leading to double-stranded helices interrupted by single-stranded regions in internal loops or hairpin loops. The enumeration of the base-paired regions or helices constitutes a description of the second level of organization, the secondary structure. The methods available to deduce the secondary structure of an RNA molecule are mainly of three types: the phylogenetic approach, the theoretical prediction, and chemical/enzymatic methods. The secondary structure of an RNA molecule is experimentally accessible and its content measurable. Under appropriate conditions, structured RNA molecules undergo a transition to a 3D fold in which the helices and the unpaired regions are precisely organized in space. This folding process usually depends on the presence of divalent ions like magnesium ions and on the temperature. The tertiary structure is the level of organization relevant for biological function of structured RNA molecules. Sections 2–4 discuss these structural properties of RNA.

The instrumental methods leading to the elucidation of the tertiary structure are then described. The first method relies on X-ray crystallography of single crystals of purified RNA molecules. The second most important structural method is nuclear magnetic resonance (NMR). Finally, the phylogenetic method coupled to computer modeling and experimental approaches is described. An overview of the RNA motifs underlying the assembly of RNA molecules in complex tertiary folds is given. These include loop–loop interactions, some of which constitute pseudoknots, tetraloops and their receptors, as well as structured internal loops such as the 5S loop E motif. The roles of non-Watson–Crick base pairing for tertiary structure is emphasized throughout.

2 CHEMICAL STRUCTURE OF RNA

Nucleic acid biopolymers comprise the DNA and RNA molecules. These two types of molecules possess very different functional roles. In brief, DNA molecules (deoxyribonucleic acids) contain the genetic code, whereas the more versatile RNA molecules (ribonucleic acids) are

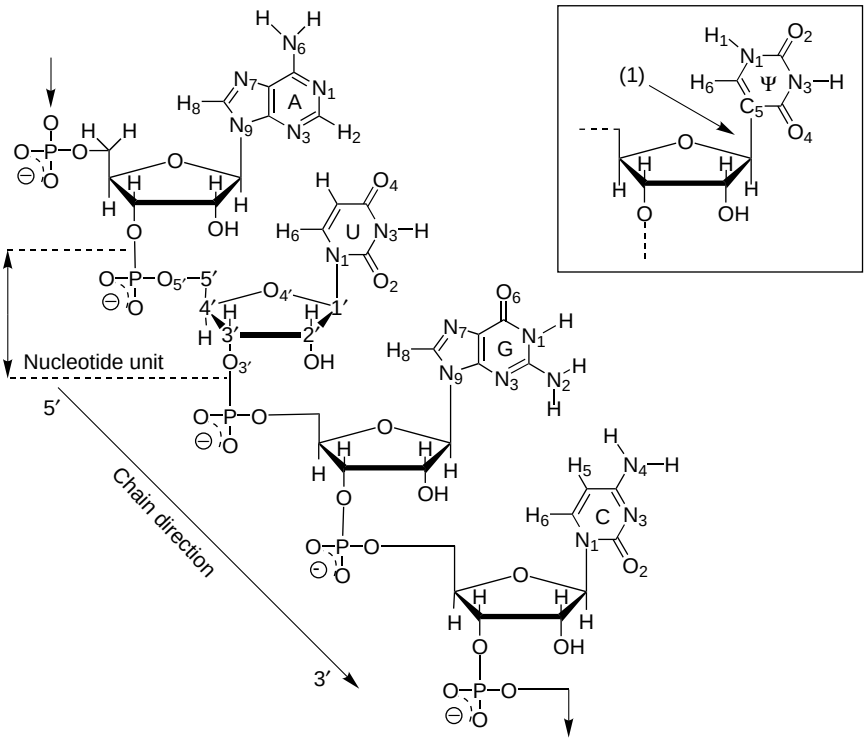

Figure 1 Chemical structure of a ···AUGC··· RNA sequence drawn from the 5′ to the 3′ end. The RNA bases are frequently modified, especially in transfer RNAs (tRNAs), where thymine (5-methyl U or T) or pseudouridine (Ψ, in which there is a C–C glycosyl bond between the base C5 and the sugar C1′ atoms) are often found.

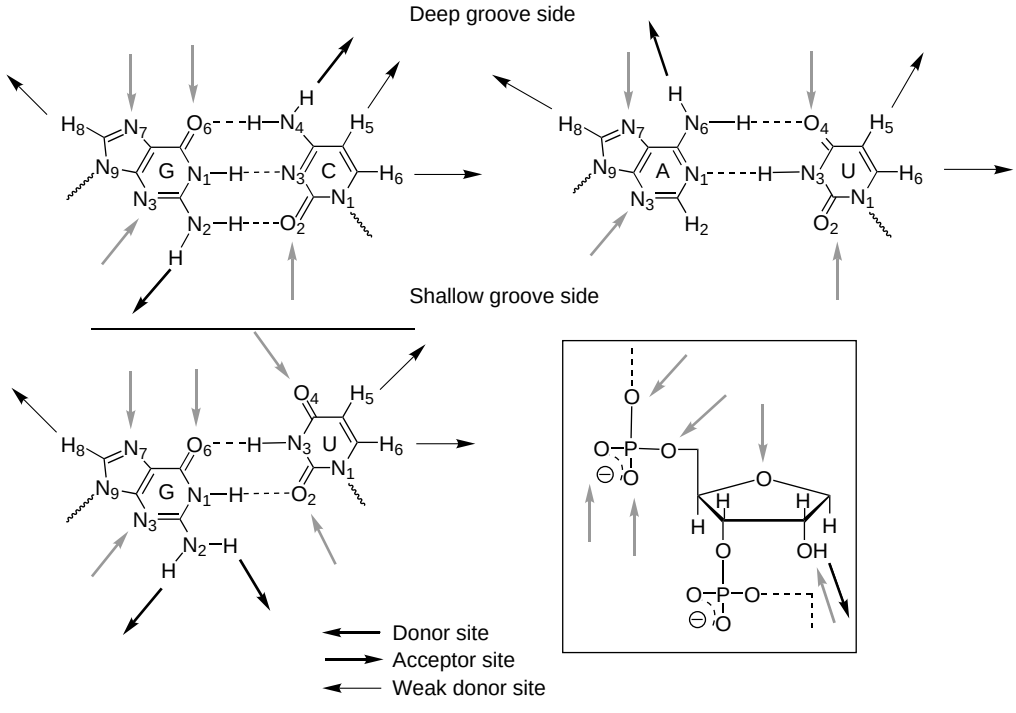

Figure 2 Possible interaction sites around the standard Watson–Crick base pairs G–C and A–U as well as the wobble G◦U pair and the sugar–phosphate backbone atoms. The edges facing the deep and shallow grooves of RNA helices are indicated. Hydrogen bond acceptor sites are marked by 'inward-pointing bold' arrows, hydrogen bond donor sites are marked by 'outward-pointing bold' arrows, weak hydrogen bond donor sites (C–H groups) are marked by thin arrows. The RNA 2′-hydroxyl group (see inset) displays simultaneously an hydrogen bond donor and acceptor potential.

involved in almost all crucial life processes and especially in the translation of the genetic code into proteins.

The four ribonucleosides that incorporate the purine bases, adenine (A) and guanine (G), and the pyrimidine bases, ucacil (U) and cytosine (C) constitute the basic building blocks of a RNA polymeric chain (Figure 1). These nucleosides comprise a ribose sugar ring and a purine (A, G) or a pyrimidine (C, U) base. They are connected together by a phosphodiester linkage. The nucleoside and its phosphodiester unit is called a nucleotide. The bases of an RNA polymeric chain associate with the complementary bases of the same chain or of another RNA chain by forming purine–pyrimidine A–U or G–C Watson–Crick base pairs (see Figure 2). The association of two self-complementary strands results in the formation of a right-handed double helical structure (Figure 3).

Although DNA and RNA molecules possess very distinct biological functions, chemically RNA differs from DNA only in two aspects: (a) the absence of a methyl group at position 5 of the uridine (U), and (b) the presence of a 2′-hydroxyl group of the RNA ribose sugar (Figure 1). These two small chemical modifications account for the profound functional and structural differences observed between DNA and RNA molecules. In contrast to DNA but similar to proteins, single-stranded polynucleotide RNA chains can fold in a variety of complex 3D structures.[1] This ability to form complex folds is exemplified by tRNA molecules which are constituted by a single chain of about 70 nucleotides (Figure 2).[2] The analogy with proteins includes also the fact that some RNA molecules, called ribozymes, are able to perform biologically crucial catalytic reactions.

In nucleic acids, there are six torsion angles in the sugar–phosphate backbone with an additional one between the base and the sugar (see Figure 5 in Berman's article **X-ray Structures of Nucleic Acids**). Sundaralingam reduced severely the variations in two torsion angles and constrained the other five to preferred domains of variations.[3] Thus, the torsion angles about the C–O bonds remain at about 180°, whereas the sugar rings adopt either the C3′-endo or the C2′-endo puckers and the base is either in an anti or a syn orientation with respect to the sugar. In proteins, the torsion angles on either side of the peptide bond constitute the main flexible links, whereas in nucleic acids the phosphodiester bonds themselves direct chain re-orientations.

The RNA molecules, in helical form (Figure 3), are not much affected by changes in their environment and adopt essentially the A-form in the helical parts of their structures under nearly all conditions.[4] However, many RNA molecules need specific divalent cations in order to fold properly into biologically active conformations.[5,6]

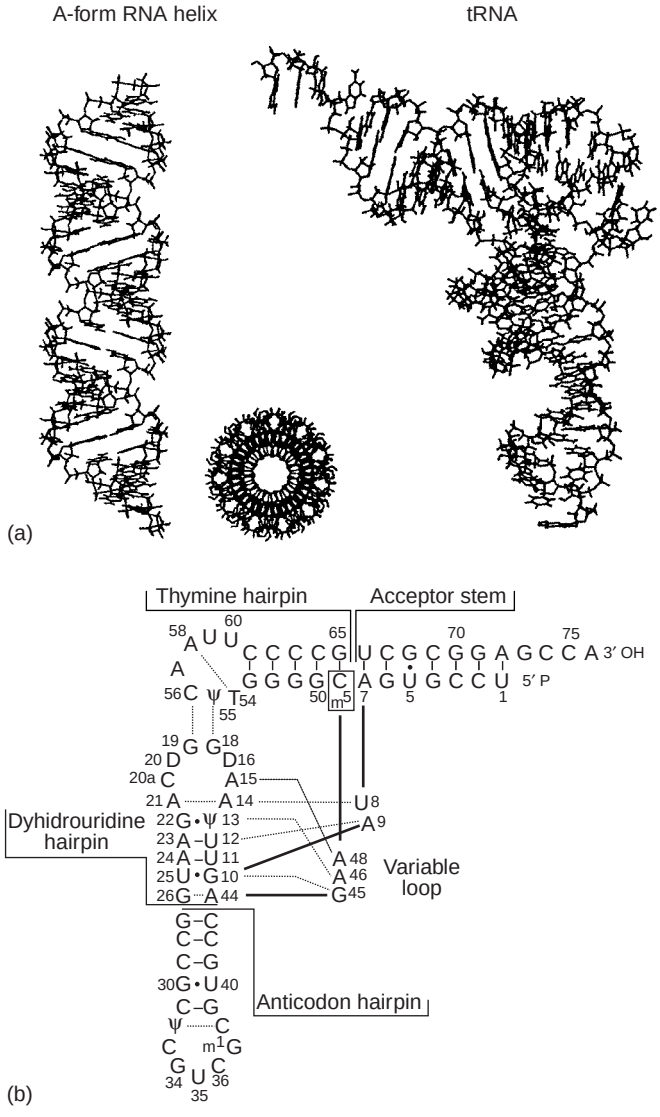

Figure 3 (a) Side and top views of the A-form adopted by RNA double helical structures. Notice the deep groove (corresponding to the major groove of B-DNA) and the shallow groove (corresponding to the minor groove of B-DNA). At the right, the 3D structure of a tRNA molecule resulting from the complex folding of a single polymer chain comprising 75 nucleotides is represented. (b) Schematic representation of the secondary and tertiary structures of yeast tRNA for aspartic acid. The dotted lines represent the tertiary pairs, which are except for one (G19–C56) of non-Watson–Crick nature. Notice the co-axial stacking between the acceptor and thymine helical stems as well as between the dihydrouridine (an uridine residue saturated at the C5=C6 bond) and the anticodon helical stems. At the interface between the dihydrouridine and anticodon stems there is usually a non-Watson–Crick pair. The thymine and dihydrouridine loops are connected via two important base pairs, a Watson–Crick G19–C56 and a bifurcated G18•Ψ55 base pairs. This spatial relationship between the secondary elements is further stabilized by triples between the single-stranded junctions (residues 8–9 and variable junction) and the deep groove of the dihydrouridine helix.

The absence of such ions may induce profound structural changes, such as loss of 3D interactions.[7]

Non-Watson–Crick base associations[8] are often found in RNA structures where they serve as specific recognition elements for proteins, nucleic acids, and ligands, or as ion-binding sites. Such noncanonical base pairs are linked by at least one interbase hydrogen bond and occasionally involve water-mediated base–base interactions. Wobble base pairs are typical of RNA molecules and can be inserted without great distortions into regular Watson–Crick helices.[9] The wobble base pairs display a characteristic shift of their Watson–Crick interaction sites. Two of these sites point, respectively, toward the shallow and deep grooves and are, thus, completely accessible to the solvent. Wobble base pairs comprise essentially the G◦U base pairs. For these base pairs, the guanine amino group protrudes into the shallow groove and the O4 atom of the uracil

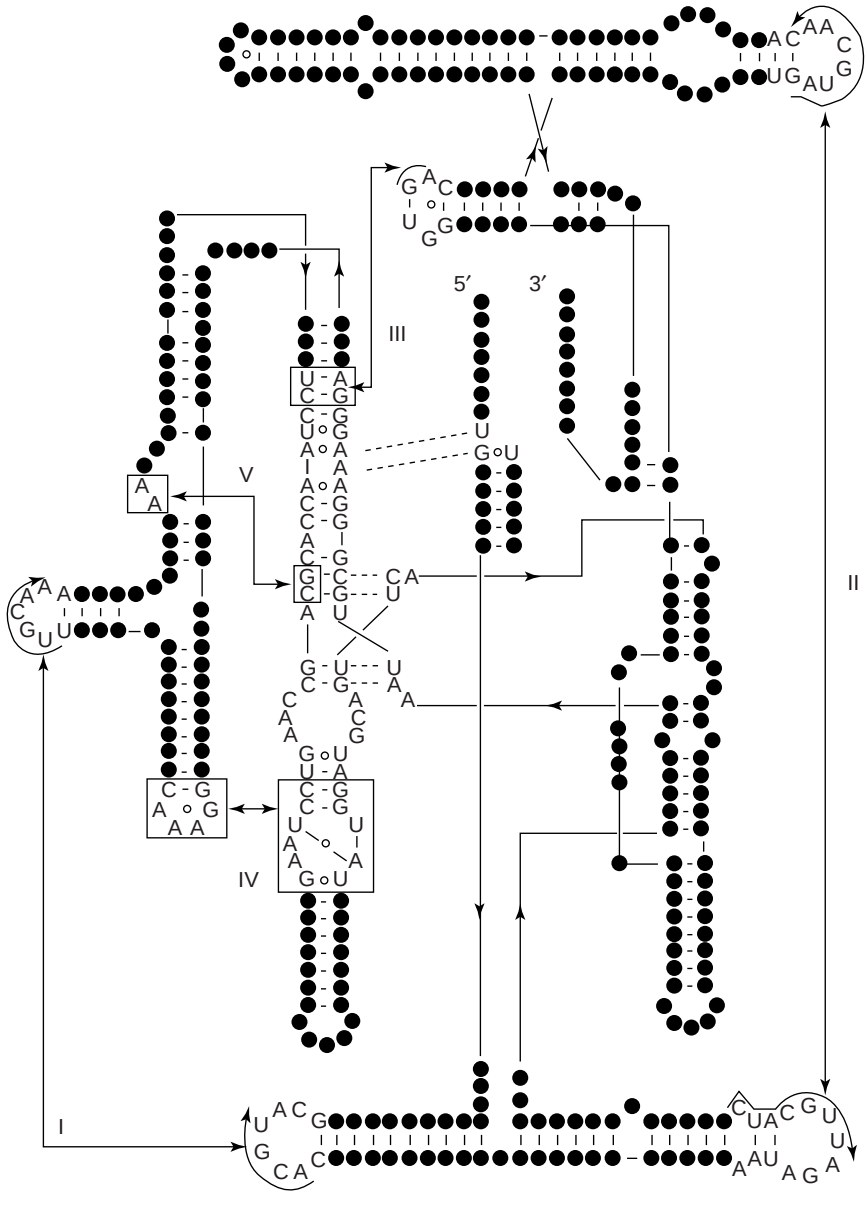

Figure 4 Representation of the secondary and tertiary structure of an artificial group I intron based on that of *Tetrahymena thermophyla*. Some tertiary motifs are highlighted. Motifs I and II are loop–loop interactions leading to a helical structure maintained via standard Watson–Crick pairings. Motif III is constituted of a GUGA tetraloop interacting with the shallow groove of two base pairs embedded within a helix. Motif IV is made of a GAAA tetraloop interacting with its receptor, the 11nt-motif, a complex structure maintained by an AA-platform (two consecutive As in the same plane[26]) and an A•U Hoogsteen base pair. In motif V, part of an A-rich bulge interacts with C–G pairs in the shallow groove of a helix.

base point towards the deep groove. Furthermore, G◦U base pairs display a deep groove side that is lined with three hydrophilic acceptor atoms (N7, O6, O4) and no hydrophilic donor atoms, whereas the shallow groove side presents a cavity wide enough for trapping a water molecule. Both grooves present, therefore, very unique recognition patterns.[10]

3 DEFINITIONS OF SECONDARY AND TERTIARY STRUCTURES

Probably the most straightforward definition of secondary structure would include any nucleotide such that both itself and one at least of its immediate neighbors in the 5′ or 3′ direction are involved in classical (Watson–Crick and G◦U) base pairing with a stretch of nucleotides in antiparallel orientation (Figures 3 and 4). Ideally, one would like an RNA secondary structure to be a planar graph which can be represented as a tree, i.e. the lines connecting the paired bases do not intersect.[11] The secondary structure will define various elements. A hairpin loop is formed when an RNA strand folds back on itself. In an internal loop, at least one base is unpaired on each strand of the loop separating two paired regions. A mismatch is a special type of internal loop for which only one nucleotide is unpaired on each strand. A bulge has unpaired nucleotides on only one strand. The other strand has uninterrupted base pairing. A multibranched loop occurs when double-stranded regions separated by any number of unpaired nucleotides, come together.

In the next level of organization, the tertiary structure, the secondary structure elements are associated through numerous van der Waals contacts, specific hydrogen bonds via the formation of a small number of additional Watson-Crick pairs and/or unusual pairs involving hairpin loops or internal bulges. The parsing of energy levels between secondary and tertiary structures is reasonable in large RNAs, considering the relative energies and the clear identification of the secondary structure elements. In some cases, it is even possible to cut RNA molecules into modular domains which can re-associate only through tertiary contacts.

4 RNA TERTIARY MOTIFS

RNA tertiary structure comprises those interactions involving (a) two helices, (b) two unpaired regions, or (c) one unpaired region and a double-stranded helix. The interactions between two helices are basically of two types: either two helices with a contiguous strand stack

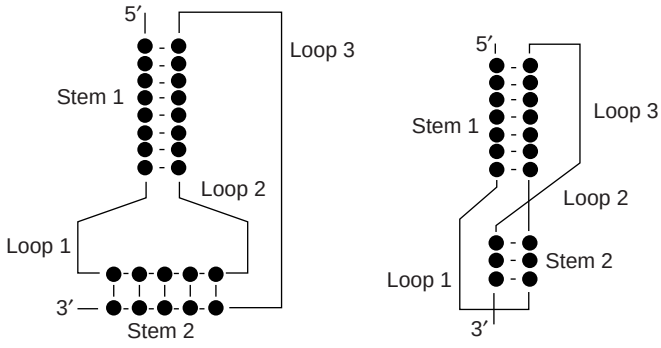

Figure 5 Two representations of a pseudoknot with a co-axial stacking of the two generated helices. Some consecutive bases (usually 6–8) of a hairpin loop are complementary, in the Watson–Crick sense, to bases either in the 5′- or 3′-strand leading to the hairpin. A second helix can, thus, be formed, once one hairpin is folded. In the classical type of pseudoknot represented (with segment L2 absent or with only one or two residues), the two helices can stack upon each other, leading to a stabilization of the tertiary motif. The single-stranded segments L1 and L3, which link stems 1 and 2, cross, respectively, the deep and shallow grooves of the contiguous helical stack to which they can make further base–base or base-sugar–phosphate backbone interactions.

on each other, or two distant helices position themselves so that their shallow grooves fit. An unpaired region belongs to either a single-stranded stretch (forming an internal loop or a bulge) or a hairpin loop closing a helix. Interactions between two unpaired regions lead to pseudoknots if a single loop is involved and to loop–loop motifs otherwise (Figure 5). Interactions between an unpaired region and a double-stranded helix can lead to various types of motifs. Pairing of a single-stranded stretch, either in the deep or the narrow groove of a double helix, yields a triple helix. One motif is known where the unpaired region constitutes a terminal loop, in which -GNRA- tetraloops bind the shallow groove of the helix (see Figure 6). Those motifs involving single-stranded stretches are especially rich in potential to form tertiary structure because they can lead to co-axiality between helices. Depending on their sequence, internal loops or bulges could constitute 3D motifs, but such motifs have not yet been characterized. The co-axial stacking of helices, together with specific helix–helix contacts or helix–loop interactions, lead to compact RNA assemblies, generally in the presence of divalent ions or polyamines (Figure 7).

5 INSTRUMENTAL METHODS

The experimental observations used for deriving a 3D structure, and thus the tertiary motifs, can be of quite different nature depending on the techniques employed.

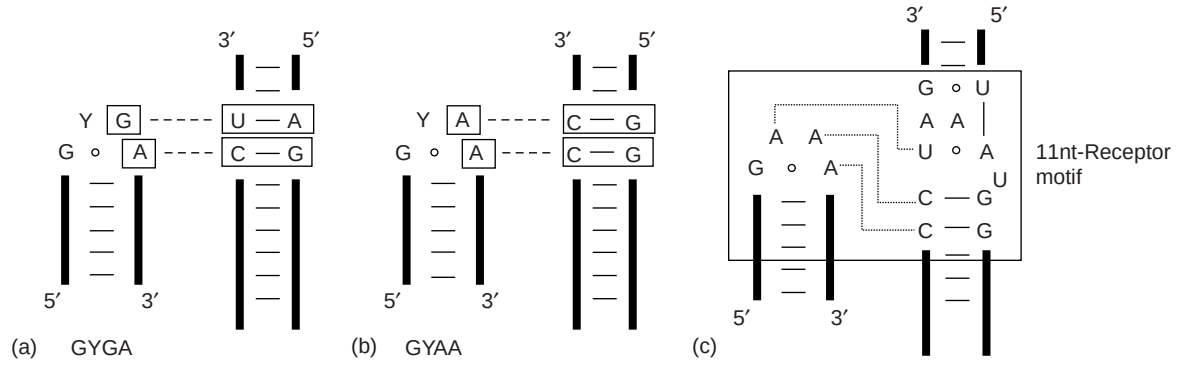

Figure 6 The three main tertiary motifs mediated by the tetraloops of the GNRA family. In the first two (a) and (b), the third and fourth residues of the GYRA loop forms triples in the shallow grooves of helices so that G binds a U−A pair and A a C−G pair. In the third type of contact (c), a GAAA tetraloop binds a complex 11nt-motif with a central U•A in the Hoogsteen configuration. Each A of the loop binds principally a different base pair in the receptor.

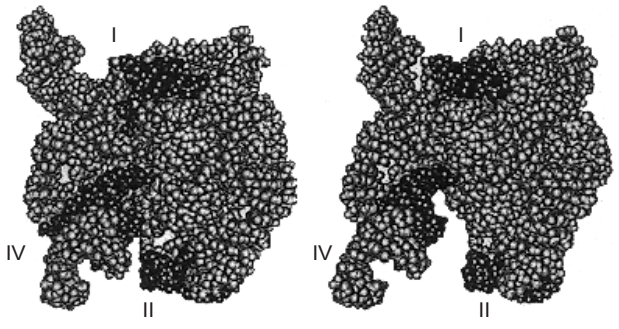

Figure 7 Stereoview of the modeled architecture of the intron of *Tetrahymena thermophyla*.[27] Some of the tertiary motifs, shown in black, are numbered as in Figure 4.

These range from biophysical methods (X-ray diffraction data, NMR couplings or nuclear Overhauser effects, and other spectroscopic methods such as ultraviolet, Raman, or circular dichroism), to biochemical approaches (chemical probing or enzymatic attack), and biological data (sequences, phylogenies, and in vitro selection). High-resolution X-ray crystallographic analysis (diffraction data at 1.5–0.9 Å resolution) yields a wealth of unequalled 3D information. Nucleic acids are difficult to crystallize because they are highly charged macromolecules which, in case of RNA molecules, can undergo spontaneous cleavages. In addition, large, nucleic acids and especially RNAs, often exchange between various base pairings and foldings.

Recently, several new RNA crystal structures have appeared, extending our structural knowledge enormously since the days of the tRNAs.[1,12] Those developments were possible first following the introduction of methods for preparing RNA molecules on a large scale with either chemical synthesis or using the T7 DNA-dependent RNA polymerase and, second, following progress in crystallographic techniques especially cryocrystallography and the availability of synchroton sources (see the article **X-ray Structures of Nucleic Acids**). NMR has proved useful in this area (reviewed in the chapter by Eriksson **Nuclear Magnetic Resonance and Nucleic Acid Structures**). Chemical and enzymatic probing of nucleic acids in solution yields important information on the stability of the structures and on those bases protected from chemical or enzymatic attack. However, such experimental approaches do not reveal the nature of the interacting partners. Cross-linking experiments have the potential to give that information, but the cross-linking reactions take place in an assembly of molecules generally not all in the same state, and it is difficult to prove that the reactions occurred solely on functional molecules. Sequence data are extremely rich in potential 3D information, as they result from adaptative evolution over millions of years. Thus, if the function is identical and the sequences are sufficiently diverse, the noise level (or covariations resulting from contingencies) will be decreased by sequence comparisons.

5.1 Physical Methods

The physical methods used for determining RNA tertiary structure are described in other articles. The most accurate method relies on X-ray diffraction of single crystals (Figure 8), as reviewed in the article **X-ray Structures of Nucleic Acids**. The development of NMR spectroscopy has progressed and allowed the determination of several solution structures of free RNAs as well as of RNAs complexes with peptides, proteins or antibiotics. This approach is reviewed in the chapter by Eriksson **Nuclear Magnetic Resonance and Nucleic Acid Structures**. Most published structures are accessible and can be retrieved from the nucleic acid database (http://ndbserver.rutgers.edu/). The RNA tertiary structure is subtended by the secondary structure and this

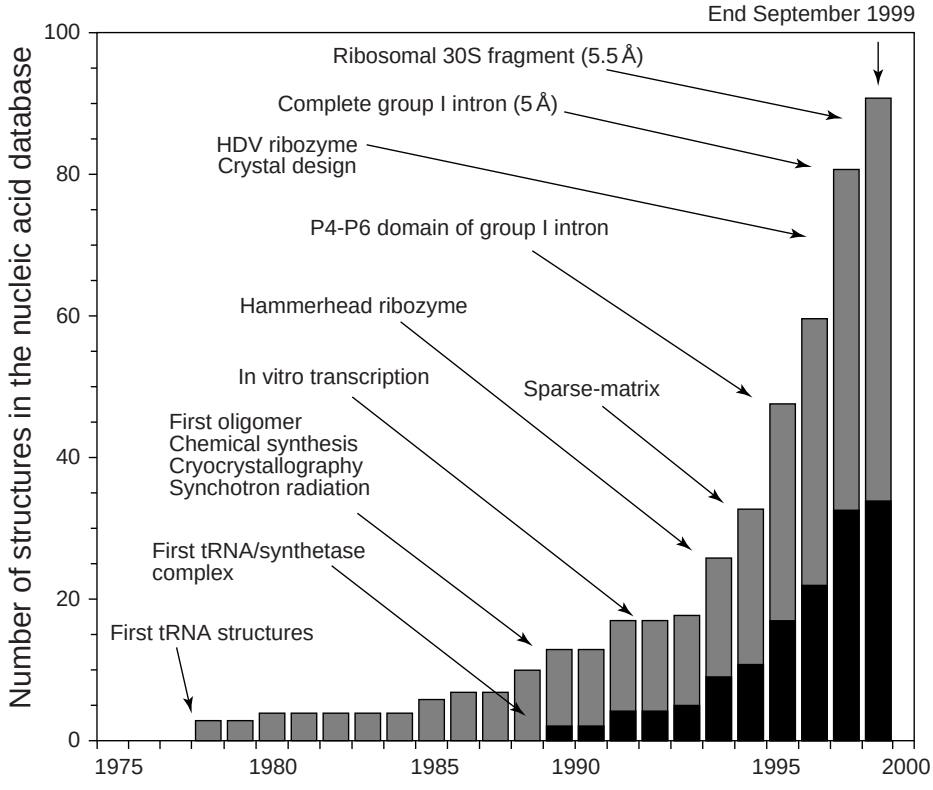

Figure 8 The cumulative number of RNA crystal structures alone (light shading) and in complexes (in dark) accessible in the nucleic acid database (http://ndbserver.rutgers.edu/).

needs to be established before large-scale synthesis is envisaged for crystallogenesis or NMR experiments. The best way to derive secondary structures is by using sequence comparisons.

5.2 Phylogenetic Approach

When a set of homologous sequences (homologous sequences have common ancestry and function) is available, one can search, with the help of a rough alignment, for compensatory base changes that maintain base-paired helices and, therefore, derive the secondary structure. When only one sequence is available or when RNAs are not conserved among a sufficiently diverse set of organisms, theoretical models of predictions, have to be associated with experiments. The related knowledge is based on a set of constraints, the thermodynamic model, and the available experimental data on the molecule.

Comparative analysis of nucleic acid sequences has been widely used for the detection and evaluation of similarities and evolutionary relationships. With RNA molecules, sequence alignments and RNA 2D prediction are intimately related. Comparative analysis is based on the biological paradigm that macromolecules are the product of their historical evolution and that functionally homologous sequences will adopt similar structures. The sequences are first aligned and then searched for compensatory base pair changes. If, during evolution, a base has been modified in a strand of a potential helix (mutation), then this modification must have been compensated on the complementary strand in order to maintain the structure. The presence of several compensatory changes (two or more) in a potential helix allows one to assert the existence of the helix in the structure. Several secondary structure models have been generated by using comparative analysis: tRNA,[13] 16S rRNA,[14] or group I and group II self-splicing introns.[15] The method requires that the molecules compared are sufficiently different to provide numerous instances of sequence variations with which to test pairing possibilities but that the molecules do not differ so much that homologous residues cannot be aligned with confidence.[16]

In an alignement, the objective is to juxtapose related sequences so that homologous residues in each sequence occupy the same column in the alignment. Standard alignment programs dedicated to molecular biology can be used for that purpose. More recent programs, dedicated to the alignment of RNA sequences, allow the user to manipulate interactively the proposed alignment (such as DCSE,[17] ALIGNOS[18]). They offer functions

dedicated to secondary structures as well as an interactive environment for manipulating the alignment.

Other recent and interesting programs automatically reconsider the alignment by taking into account new sequences and the pre-existing knowledge of the secondary structure.[19] Indeed, with the growing number of sequences, specific RNA databases are created and new sequences have to be added quickly in structured databases of homologous RNA molecules. In such databases, it is very desirable that sequences are aligned in accordance with the preservation of secondary structural features. Because, in an alignment, optimal structural elements can be misaligned, the program RNAlign[19] makes it possible to align a group of aligned sequences with a new sequence, using positions of high sequence conservation as a guide and a common secondary structure of the group as a guide for determining the secondary structure of the new sequence. Thus, RNAlign does not suppose that the related sequences are correctly aligned but instead reconsiders the alignment.

5.3 Computer Modeling

Molecular modeling attempts to assemble the 3D structure of a macromolecule on the basis of a mixture of theoretical and experimental data. Hence, prediction methods range from mathematically oriented approaches, relying solely on computer algorithms, to pragmatic and operational approaches in which insight comes from both theory and experiment. Modeling can be viewed as a heuristic tool which should help in the rationalization of experimental observations but also should suggest new relationships between the various components of the modeled molecule. Modeling relies on an existing body of knowledge and, therefore, necessitates an integration of that accumulated knowledge in a form which can be exploited. Hence, the importance of organized and annotated data banks.

The power of visualizing spatial relations is such that models need not be always detailed. In the end, the validity and the accuracy of the model obtained will depend on the nature of the experimental observations collected. However, a mathematical proof guaranteeing the correctness of the derived model is only possible with crystallographic methods (the Fourier theorem). Otherwise, the best that can be achieved is a network of evidence converging on the spatial contacts and relations embodied by a model.

Because of the strong base-pairing constraints and the Watson–Crick rules of complementarity, modeling has had some success in the field of nucleic acids, starting from the Watson–Crick double helix. In the RNA field, modeling first proposed the structure of a GNRA tetraloop closed by a sheared G•A pair with stacking of the remaining purine bases.[20] Also, the relative arrangements of helices in group I introns were correctly assembled in a 3D model[21] long before X-ray crystallography led to supportive evidence.[22] Although such models lack correctness in atomic detail, they allow the deduction of new spatial relationships. Those, together with sequence analysis, permit the localizing of recurrent motifs serving as RNA–RNA anchors responsible for recognition and stabilization of folding in structured RNAs. In this way, the recognition motif between the GNRA tetraloops and the shallow grooves of helices was correctly identified[21] before being observed in intermolecular crystal packing contacts.[23] Similarly, the recognition motif between the GAAA tetraloop and its 11nt-receptor was identified by in vitro experiments using selection methods[24] before being observed by crystallography.[25,26]

6 CONCLUSIONS

RNA molecules are often restricted to their roles as intermediate between the genomic DNA and the active proteins. This view has changed following the discovery of catalytic RNA by S. Altman and T. Cech:[27,28] RNA is nowadays the only molecule with the two properties of being a depository of genomic information with catalytic potential. Chemical catalysis requires a precise positioning of atoms in space and, therefore, RNA must achieve complex tertiary folds in order to reach transition states. Because RNA molecules carry one negative charge per residue, the compact assembly of large RNAs presents a formidable physical problem. The architecture of large structured RNAs is dominated by helix formation and co-axial stacking of helical stems with loop–loop interactions and GNRA/motifs prevalent as domain anchors.[29] Although the recurrent tertiary motifs can be identified by sequence analysis coupled with chemical probing and 3D modeling, the precise and subtle atomic details require X-ray crystallography or NMR techniques in order to be unveiled.

The developments in chemical and molecular biology techniques have allowed the circumvention of some of the difficulties in the crystallization of RNA molecules and one can hope to see a continuous progression in the number of crystal structures of RNA molecules. Similarly, NMR spectroscopy has become a decisive tool for the unravelling of RNA motifs and complexes between RNA fragments and other ligands, such as antibiotics. Not every single RNA molecule is as highly structured as a self-splicing group I intron. However, it is common to find structured regions in functionally important domains of eukaryotic messenger RNAs, such as the 5′- or 3′-untranslated regions, in eubacterial promoters or

Shine–Dalgarno regions, and naturally in control regions of viral genomes. In addition, new structured RNAs are still being discovered, like the tmRNAs or the RNA part of the ribonucleic particle telomerase. Structured RNAs are often part of complexes with proteins, sometimes a single one as in RNaseP, or with 21 different proteins such as 16S rRNA in the 30S eubacterial ribosomal particle.

ABBREVIATIONS AND ACRONYMS

NMR	Nuclear Magnetic Resonance
tRNA	Transfer RNA
3D	Three-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Circular Dichroism in Analysis of Biomolecules • Infrared Spectroscopy of Biological Applications • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules • Raman Spectroscopy in Analysis of Biomolecules • Single Biomolecule Detection and Characterization

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Aptamers • Mass Spectrometry of Nucleic Acids • Nuclear Magnetic Resonance and Nucleic Acid Structures • Nucleic Acid Structural Energetics • Sequencing Strategies and Tactics in DNA and RNA Analysis • Structural Analysis of Ribozymes • X-ray Structures of Nucleic Acids

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

Radiochemical Methods **(Volume 14)**
γ-Spectrometry, High-resolution, for Radionuclide Determination

REFERENCES

1. R.T. Batey, R.P. Rambo, J.A. Doudna, 'Tertiary Motifs in RNA Structure and Folding', *Angew. Chem., Int. Ed. Engl.*, **38**, 2326–2343 (1999).
2. T. Hermann, D. Patel, 'Stitching Together RNA Tertiary Architecture', *J. Mol. Biol.*, **294**, 829–849 (1999).
3. M. Sundaralingam, 'Stereochemistry of Nucleic Acids and their Constituents. IV. Allowed and Preferred Conformations of Nucleosides, Nucleoside, Mono-, Di-, Tri-, Tetraphosphates, Nucleic Acids and Polynucleotides', *Biopolymers*, **7**, 821–860 (1969).
4. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
5. J.H. Cate, R.L. Hanna, J.A. Doudna, 'A Magnesium Ion Core at the Heart of a Ribozyme Domain', *Nat. Struct. Biol.*, **4**, 553–558 (1997).
6. A.M. Pyle, 'Ribozymes: A Distinct Class of Metalloenzymes', *Science*, **261**, 709–714 (1993).
7. P. Brion, E. Westhof, 'Hierarchy and Dynamics of RNA Folding', *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137 (1997).
8. N.B. Leontis, E. Westhof, 'Conserved Geometrical Base Pairing Patterns in RNA', *Quart. Rev. Biophysics*, **31**, 399–455 (1998).
9. H.F.C. Crick, 'Codon-anticodon Pairing: The Wobble Hypothesis', *J. Mol. Biol.*, **19**, 548–555 (1966).
10. P. Auffinger, E. Westhof, 'RNA Base Pair Hydration', *J. Biomol. Struct. Dyn.*, **16**, 693–707 (1998).
11. E. Westhof, F. Michel, 'Prediction and Experimental Investigation of RNA Secondary and Tertiary Foldings', in *RNA–Protein Interactions* K. Nagai, I.W. Mattaj, IRL Press, Oxford, 25–51, 1994.
12. B. Masquida, E. Westhof, 'Crystallographic Structures of RNA Oligoribonucleotides and Ribozymes', in *Oxford Handbook of Nucleic Acid Structure*, ed. S. Neidle, London, 533–565, 1998.
13. R.W. Holley, J. Apgar, G.A. Everett, J.T. Madison, M. Marquisee, S.H. Merrill, S.H. Penswick, J.R. Zamir, 'Structure of a Ribonucleic Acid', *Science*, **147**, 1462–1465 (1965).
14. C.R. Woese, L.J. Magrum, R. Gupta, R.B. Siegel, D.A. Stahl, 'Secondary Structure Model for Bacterial 16S Ribosomal RNA: Phylogenetic, Enzymatic and Chemical Evidence', *Nucleic Acids Res.*, **8**, 2275–2293 (1980).
15. F. Michel, A. Jacquier, B. Dujon, 'Comparison of Fungal Mitochondrial Introns Reveals Extensive Homologies in RNA Secondary Structures', *Biochimie*, **64**, 867–881 (1982).
16. F. Michel, M. Costa, 'Inferring RNA Structure by Phylogenetic and Genetic Analysis', in *RNA Structure and Function*, eds. R.W. Simons, M. Grunberg-Manago, Cold Spring Harbor Laboratory Press, 175–202, 1996.
17. P. DeRijk, R. DeWachter, 'DCSE, an Interactive Tool for Sequence Alignment and Secondary Structure Research', *CABIOS*, **9**, 735–740 (1993).
18. H. Neurath, J. Wolters, 'The Alignment Editor ALIGNOS – Integrating Tool for a Cooperative Databanks for Structural RNAs', *Bioinformatics*, **1**, 22–25 (1992).
19. F. Corpet, B. Michot, 'RNAlign Program: Alignment of RNA Sequences Using both Primary and Secondary Structures', *Comput. Appl. Biosci.*, **10**, 389–395 (1994).

20. E. Westhof, P. Romby, P. Romaniuk, J.-P. Ebel, C. Ehresmann, B. Ehresmann, 'Computer Modeling from Solution Data of Spinach Chloroplast and of *Xenopus laevis* Somatic and Oocyte 5S rRNAs', *J. Mol. Biol.*, **207**, 417–431 (1989).
21. F. Michel, E. Westhof, 'Modelling of the Three-dimensional Architecture of Group I Catalytic Introns based on Comparative Sequence Analysis', *J. Mol. Biol.*, **216**, 585–610 (1990).
22. B.L. Golden, A.R. Gooding, E.R. Podell, T.R. Cech, 'A Preorganized Active Site in the Crytsal Structure of the Tetrahymena Ribozyme', *Science*, **282**, 259–264 (1998).
23. H.W. Pley, K.M. Flaherty, D.B. McKay, 'Model for a Tertiary Interaction from the Structure of an Intermolecular Complex Between a GAAA Tetraloop and an RNA Helix', *Nature*, **372**, 111–113 (1994).
24. M. Costa, F. Michel, 'Frequent Use of the Same Tertiary Motif by Self-folding RNAs', *EMBO J.*, **14**, 1276–1285 (1995).
25. J.H. Cate, A.R. Gooding, E. Podell, K.H. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, J.A. Doudna, 'Crystal Structure of a Group I Ribozyme Domain – Principles of RNA Packing', *Science*, **273**, 1678–1685 (1996).
26. J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, A.A. Szewczak, C.E. Kundrot, T.R. Cech, J.A. Doudna, 'RNA Tertiary Structure Mediation by Adenosine Platforms', *Science*, **273**, 1696–1699 (1996).
27. K. Kruger, P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gotschling, T.R. Cech, 'Self-splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of *Tetrahymena*', *Cell*, **31**, 147–157 (1982).
28. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, 'The RNA Moiety of Ribonuclease P is the Catalytic Subunit of the Enzyme', *Cell*, **35**, 849–857 (1983).
29. V. Lehnert, L. Jaeger, F. Michel, E. Westhof, 'New Loop–Loop Interactions in Self-splicing Introns of Subgroup IC and ID: A Complete 3D Model of the *Tetrahymena thermophila* Ribozyme', *Chem. Biol.*, **3**, 993–1009 (1996).

Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes

Radoje Drmanac, Snezana Drmanac, and Deane Little
Hyseq, Inc., Sunnyvale, USA

1 Introduction 1
2 Origins of Sequencing by Hybridization 2
3 Principles of Sequencing by Hybridization 3
3.1 Hybridization Search Principle 3
3.2 Probe Overlap Principle 4
4 Variations of the Sequencing by Hybridization Method 7
4.1 Format 1 Sequencing by Hybridization 7
4.2 Format 2 Sequencing by Hybridization 8
4.3 Format 3 Sequencing by Hybridization 8
5 Applications 8
5.1 Clone Contig Mapping Over a Wide Range of Clone Lengths 9
5.2 Clustering, Motif Searching, and Partial Sequencing of cDNAs 9
5.3 Detection of Known Polymorphisms and Mutations, Including Single Nucleotide Polymorphisms 10
5.4 Discovery of Novel Polymorphisms or Mutations by Individual DNA Sequencing 10
5.5 De Novo DNA Sequencing 11
6 Biochemistry 11
6.1 Sample Preparation 11
6.2 Probe Preparation 12
6.3 Array Chemistry 12
6.4 Labeling 12
6.5 Discriminative Hybridization 13
7 Sequencing by Hybridization Automation 15
8 Data Analysis 17
9 Selected Studies 19
9.1 De Novo Sequencing 19
9.2 Gene Discovery and Gene Expression 20
9.3 Individual Sequencing to Discover Novel Polymorphisms and Mutations 21
10 Conclusion 23
Acknowledgments 23
Abbreviations and Acronyms 24
Related Articles 24
References 24

Sequencing by hybridization (SBH) determines the sequences of DNA or RNA targets using hybridization data obtained from sets of overlapping oligonucleotide probes of known sequence. SBH shares interesting parallels with computer-based keyword searches of text files. In SBH, chemically synthesized short oligonucleotides (usually 4–25 bases in length) are hybridized to the target under conditions that predominantly allow formation of fully complementary probe/target hybrids. An oligonucleotide hybridization search is effectively a highly parallel molecular computation process with fully random access to the polynucleotide chain ''input data''. The target sequence is assembled by aligning sequences of the subset of positively hybridizing oligonucleotides. Probe sequence overlap allows indirect assignment of which of four bases corresponds to each position in the analyzed DNA chain, without performing any actual positional measurements on the sample. The fundamental characteristics of the SBH process confer unique opportunities for miniaturization and parallel analyses (currently in the form of high-density arrays), leading to rapid, cost-effective DNA sequencing. A unique advantage of SBH is its ability to analyze in a single reaction complex DNA or RNA samples that may be thousands of bases long. Hybridization of smaller selected subsets of probes provides a very efficient way of partially sequencing and fingerprinting samples. A very important advantage of SBH is that in a single reaction each base is ''read'' by multiple overlapping probes. To achieve these advantages, hybridization reactions are miniaturized and coupled with automated data acquisition and efficient computational analyses that can process data sets containing both false-positive and false-negative probe scores. Successful assembly of long target sequences involves synthesis and hybridization of a large number of long probes, due to the higher probability that given short oligonucleotide sequences will be repeated within the target sequence. Overall, SBH is a rapid and cost-effective method of obtaining large amounts of accurate DNA sequence information.

1 INTRODUCTION

Living organisms are self-encoded systems that possess a precise set of instructions, recorded in the sequence

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of their DNA or RNA, that uniquely specifies the organism's development and function. The nucleotide sequence of a DNA or RNA molecule is the specific order of nucleic acid bases in a polynucleotide chain that may be tens of millions of bases long. There are four bases commonly present in DNA: adenine, cytosine, guanine, and thymine (A, C, G, and T respectively). In its natural state, DNA is a double helical molecule, in which two complementary single-stranded DNA chains specifically intertwine with one another. Each purine DNA base in one chain recognizes and binds a specific pyrimidine partner in the other, A to T, and G to C. In this way, the sequence of each strand specifically and unambiguously defines the sequence of its complementary partner, which is of critical importance in the process of DNA replication and cell division.

The ability to efficiently read the sequences of genomic DNA or RNA molecules is essential to understanding and diagnosing genetic diseases, developing new medical treatments, engineering new traits in plants and animals, and unraveling developmental and evolutionary pathways of living organisms. Since the discovery of DNA's helical structure by Watson and Crick in the 1950s there have been intense efforts to develop biochemical, chemical, and physical methods to sequence individual genes or whole genomes. The efforts have culminated in the Human Genome Project, an international research project initiated in 1986, whose goal is to completely sequence the 3 billion nucleotides of the human genome.

The Human Genome Project, while of enormous importance, does not represent a complete answer to the many critical questions posed by human genetics. The maximum benefit of sequence information is obtained in comparative genome studies between various species and in correlation studies that relate individual sequence and phenotype differences. Such studies require sequencing the genomes of thousands of individuals and species. This situation has fueled a demand for dramatically more efficient ways of sequencing or fingerprinting DNA.

There are three principal ways to obtain partial or complete sequence information for a DNA fragment: (1) sequence-specific DNA degradation, synthesis, and/or separation; (2) sequence-specific DNA hybridization (strand pairing); and (3) nucleotide chain visualization. The first two approaches are currently used in many applications. Nucleotide chain visualization methods are still in the initial phases of development.

Within the first category of sequencing techniques, the automated Sanger gel electrophoresis method is currently most frequently used. This technique involves randomly terminating the synthesis of an unknown DNA at specific bases, followed by exact size (length) determination of the resulting DNA fragments.[1] The DNA fragments (all of which are synthesized from a single primer site) are then ordered by length to determine which bases are located at which sequence positions. Historically, RNA and DNA sequences were determined by chemical fragmentation techniques, such as the Maxam–Gilbert method,[2] or enzymatic fragmentation, such as various tRNA sequencing methods.[3] A number of variants of this basic scheme have been proposed, such as the cyclical removal or addition of single bases at the end of the sequenced chain.[4]

This article focuses on the second of the three principal sequencing approaches, the use of sequence-specific hybridization to determine primary DNA structure. The process takes advantage of one of the basic biochemical processes of life, the molecular specificity of pairing between two complementary DNA strands. Over the past decade, the use of oligonucleotide probes to sequence or fingerprint unknown DNAs by such hybridization methods has proved to be a highly potent and versatile way to obtain valuable genetic information.

Chemically synthesized oligonucleotides of known sequence allow a fast and informative hybridization search of a DNA sample, a process that is essentially similar to a keyword search of text by an Internet browser. Large sets of probes may be used to assemble an entire sequence, or smaller sets to obtain a ''fingerprint'' or ''sequence signature'' specific for a particular DNA. This article reviews the history, principles and experimental applications of DNA sequencing and fingerprinting by hybridization, and describes in detail several studies that demonstrate its tremendous potential as a method of analyzing individual genes or entire genomes.

2 ORIGINS OF SEQUENCING BY HYBRIDIZATION

Nucleic acid hybridization is the basis for all SBH methods. Hybridization is the sequence-specific process in which two polynucleotide chains recognize and bind one another at complementary sites. In 1960, Doty and colleagues at Harvard discovered that DNA molecules that have been denatured by heat or chemical treatment to form single-stranded chains can precisely pair or renature under the right reaction conditions.[5] Specificity of pairing is determined by base complementarity rules in which As from one chain bind to Ts in the other, and Gs bind to Cs.

Initial applications of hybridization in molecular biology (for example colony or Southern blot hybridizations) used nucleic acid probes longer than 50 bases that were derived from natural sources. The specificity and sensitivity of these reactions are such that sequences present at one part in a million can be readily detected. It is not

necessary for all bases in such a double-stranded hybrid to match perfectly; up to 1–2% mismatched bases may occur without affecting the strength of hybridization. The current widespread use of hybridization methods to identify samples containing previously isolated sequences is possible because natural sequences (i.e. genes) within a genome typically differ by far more than a few percentages in thousands of bases.

During the late 1970s and early 1980s powerful new research tools were developed to unravel the intricate sequences of DNA molecules. Caruthers and other researchers discovered ways to more efficiently synthesize specific DNA molecules on columns by stepwise addition of nucleotides.[6] These automated DNA synthesizers made possible the production of large numbers of specific oligonucleotides. In a parallel development, Bruce Wallace pioneered the use of such oligonucleotides as hybridization probes to confirm the existence of complementary sequences in target DNAs.[7]

As tools for detecting complementary sequences within a target DNA, oligonucleotide probes proved to be more sequence-specific than longer, natural probes. Wallace demonstrated that oligonucleotide probes 11–17 bases in length can be used to accurately detect single base mutations in defined DNA targets dot-blotted on a membrane.[7] He showed that such probes can be used to discriminate between perfect probe–target duplexes and those containing a single internal base mismatch.

Several years later, Saiki et al.[8] developed a ''reverse dot blot'' method of screening patient DNA samples for known mutations using oligonucleotide probe arrays. The method relies on an array of defined oligonucleotides designed to complement known mutations and corresponding wild-type sequences in a given gene, such as the human β-globin gene. Patient samples enriched in the particular gene sequence are then tested for hybridization against the array under conditions in which the formation of duplexes with an internal base mismatch is minimized. Positive hybridization of one probe (or two probes in the case of heterozygotes) confirms the existence of the mutation.

These early research developments provided basic research tools that were later used in developing methods for fingerprinting and sequencing DNA by oligonucleotide hybridization. They ensured essentially limitless supplies of oligonucleotides of any pre-specified sequence, and established the experimental conditions suitable for specific full-match hybridization using either target DNA or oligonucleotide probes bound to a support.

In 1986, Poustka and Lehrach[9] demonstrated that oligonucleotide hybridization can play a role in the field of DNA mapping. Their work showed that positive hybridization of an oligonucleotide to a cloned genomic fragment gives essentially the same mapping information as is provided by obtaining a single DNA band when cutting a cloned DNA with a restriction enzyme. In each case a specific short sequence is identified within the test DNA. With oligonucleotide hybridization, the identified sequence is identical in length and complementary in sequence to that of the oligonucleotide probe. The authors showed that hybridization ''fingerprints'' obtained with a set of short DNA probes can be used to identify overlapping cosmid clones sharing portions of their DNA.

In 1987, as a culmination of oligonucleotide applications in DNA analysis, Drmanac and Crkvenjakov[10] first proposed and later patented the method of SBH, in which the sequence of a DNA fragment is determined by hybridization with a sufficient number of short oligonucleotides of known sequence. The sequence is assembled from the subset of positively hybridizing probes detected by hybridizing a set of overlapping probes with the target DNA. The researchers presented an algorithm for reconstruction of DNA sequences as complex as the human genome that established the relationship between length and number of probes to length and number of clone fragments from the DNA being sequenced.[10,11] Hybridization of labeled probes to large arrays of clones was proposed as a method for sequencing complex genomes. Like the polymerase chain reaction (PCR) method, SBH radically expanded potential applications of the hybridization reaction. SBH is also thought to be more suitable for automation than classical sequencing methods.

Within a year of Drmanac and Crkvenjakov's patent application, other research groups presented variations of this new sequencing process. Bains and Smith[12] examined the process of sequence reconstruction from tetramers. In a 1988 patent application, Southern proposed a method for combinatorial in situ synthesis of complex oligonucleotide arrays on glass and hybridization of DNA to the array for mutation detection and complete sequencing.[13] Lysov et al.[14] also proposed SBH using an array of deposited oligonucleotides.

3 PRINCIPLES OF SEQUENCING BY HYBRIDIZATION

3.1 Hybridization Search Principle

SBH shares interesting parallels with conducting a computer search of a text file for a particular word or a phrase. In each case, a large string of characters is probed with a specific shorter string to detect matching sequences. In a computer text search, the search string or strings (key words) are used to browse a large Internet or local database to identify the subset of specific

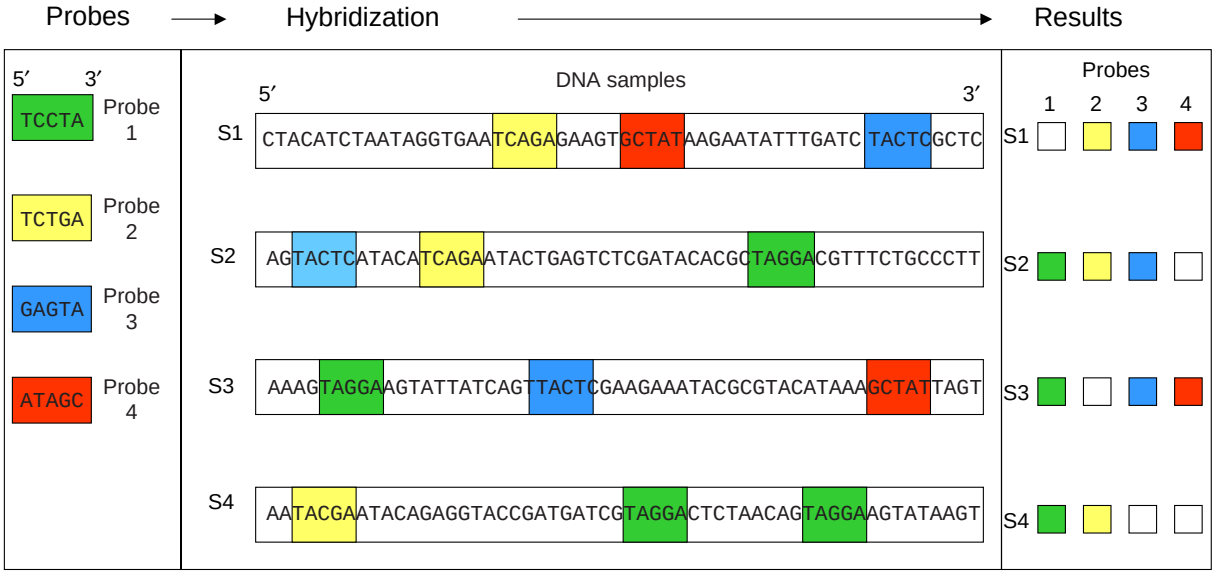

Figure 1 Hybridization search of a DNA library. The schema depicts hybridization sites for four 5-mer oligonucleotide probes (shown in color) in four target DNA samples, as well as the resulting data pattern. Each DNA has a specific pattern reflecting the distinct subset of positive probes that hybridize with it. However, a hybridization search identifies only which probes hybridize to each sample, without indicating the order in which probe binding sites occur. For example, the result for the first sample is ''missing green probe,'' and for the fourth sample ''only green and yellow probes positive.'' The presence of two green probe binding sites in DNA sample 4 would not be detected unless hybridization intensity scores were evaluated instead of simple presence/absence data, as discussed in section 8. These ''presence/absence patterns'' may be used to identify the DNA (oligonucleotide fingerprinting) or, in combination with similar data from many overlapping probes, to determine its complete sequence.

documents containing perfect sequence matches, which is then retrieved for further review or analysis. In SBH, oligonucleotide probes ranging from 4 to 25 characters in length are used to browse libraries of DNA segments to identify DNA molecules containing exact complementary sequences. These molecules may then be further analyzed by mapping or clustering, or by partial or full sequencing.

Figure 1 shows the results of a hybridization search of four simple DNA samples with four different 5-mer probes (key words, strings) indicated by different colors. In this case, each sample binds a different combination of probes, leading to a characteristic color hybridization pattern. Each positive binding event in a given DNA sample provides a discrete piece of information about its sequence. Neither frequency nor location of the string within the DNA molecule is obtained from a hybridization search, as is also the case in most computer text searches. For example, a positive search result for the word ''tag'' in a set of document titles does not identify whether the word is positioned at the beginning, middle, or end of the selected titles, nor whether it occurs once, twice, or many times in any of these titles. Similarly, the entire DNA is sampled by random probe-binding trials, without determination of exactly where in the chain particular probes bind.

Thus, a DNA hybridization search is effectively a highly parallel molecular computation process with fully random access to the ''input data'', in this case a polynucleotide chain that may be thousands of bases long. These fundamental characteristics of the SBH process confer unique opportunities for miniaturization and parallel analyses, leading to speed and cost efficiencies not available with other sequencing methods.

In a computer search, the complexity of the English alphabet (26 letters) generally allows a meaningful search of a given text to be done with one or a few specific words. With a DNA search, the simple four-letter genetic alphabet requires use of either more or longer ''words'' (strings) to precisely identify a specific DNA. A simple word like ''cat'' might yield useful results in a computer search of the Internet, but the genetic triplet ''CAT'' occurs far too frequently (about once in every 64 triplets) to be of much use in DNA identification. The lengths of the DNA string (sequence) and the probe (interrogating string) are important parameters in devising a successful SBH experiment. By choosing appropriate probe and sample lengths, a researcher can obtain useful sequence data.

3.2 Probe Overlap Principle

The first potential probe binding site in a DNA chain starts at the first base and extends for the length of probe. The second probe binding site starts at the second base and overlaps the first probe binding site, less one base.

This means that if a complete (or sufficient) set of probes is tested, the end of each positive probe overlaps with the beginning of another positive probe, except in the case of the last positive probe in the target. This situation is illustrated in Figure 2(a), which shows the results of a hybridization search of a short 16-base DNA sequence with all 1024 possible pentamers. Positive hybridizing probes are indicated in boxes. Using the OP, the positive binding of just 12 of the 1024 possible 5-mer probes unambiguously determines the sequence of the target DNA (see Figure 2b). In each sequence assembly cycle, four potential overlap probes are checked. Starting with a positive probe AAATC, the next positive overlapping probe to the right may be AATCA, AATCC, AATCG or AATCT. Of these probes, only AATCG is found to be positive and is used for further assembly. The cycles are repeated in both directions until all positive probes are incorporated and the complete sequence is assembled. By extension, the same process applies to a longer DNA target if enough probes of appropriate length are used to identify uniquely overlapped strings within it.

The use of overlapping positive probes is a key aspect of SBH methods. The OP allows the identification of sequences within a target DNA that are longer than any of the probes used in the assembly process. Probe overlap allows indirect assignment of one out of four bases for each position in the analyzed DNA chain without performing any actual positional measurements on the sample. The base/position information is in fact derived from the known sequences of the oligonucleotide probes obtained by accurate chemical synthesis.

Because the sequences of DNA molecules are nonrandom and irregular, statistical artifacts arise that must be addressed in SBH experiments. Even when the lengths of DNA targets and probes are selected to achieve a statistical expectation that each probe sequence occurs no more than once in the target, so-called ''branching ambiguities'' can occur.[10,11] Take the case of three probes that positively hybridize to a target DNA: TAGA, AGAC and AGAT. Both the second and the third probes overlap with the first probe, sharing the bases AGA and giving extended sequences TAGAC and TAGAT respectively. Owing to the occurrence of the sequence AGA in both the second and third probes (e.g. because of double AGA occurrence in the target), there is not enough information available to decide which of the two probes is actually the one that overlaps with the first probe in the sample. Sequence assembly can thus proceed along either of the two branches, only one of which is correct.

By using all probes of a given length (as in Figure 2a), a researcher can unambiguously determine a target sequence, provided the target DNA is short enough that most overlap sequences occur no more than once. The only other exception to this rule is tandem repeat regions (e.g. AAAAAAAAAA, ACACACACAC) that are longer than the probe length. In such cases, the exact length of these repeats may be determined by use of a special subset of longer probes. Longer targets may require longer probes for unambiguous sequence determination. Table 1 shows the exponential increase in the number of possible probes as a function of probe length. A variety of ways have been proposed to increase the read length with a given set of probes, or to reduce the number of experimental probe/target scores needed to sequence a target DNA. These include the use of pools of probes, competitive hybridization and overlapped clones,[10,11] gapped[12] and binary[15] probes, continuous stacking hybridization,[16] and the simultaneous sequencing of similar genomes.[17]

The information needed to solve branching ambiguities is directly provided by using a reference gene sequence when assembling gene sequences from individual patient samples. When a reference sequence is known, the maximal read length in bases may be extended to approximately one-tenth the number of probes used (Table 1). To sequence human and other complex genomes (over 3 billion base pairs) as a single sample, about 70 billion 18-mer

Table 1 Relationship of probe length, number of possible probes, and sequence read length

Number of nucleotides in probe (n)	Number of possible probes[a] (4^n)	Basic read length[b] (bases)	Enhanced read length[c] (bases)
3	64	6	6
4	256	12	25
5	1024	25	100
6	4096	50	500
7	16 384	100	1500
8	65 536	200	6500
9	262 144	400	25 000
10	1 048 576	800	100 000
12	16 777 216	3200	1 500 000
14	268 435 456	12 800	25 000 000
16	4 294 967 296	51 200	450 000 000
18	68 719 476 736	204 800	7 000 000 000
20	1 099 511 627 780	819 200	100 000 000 000

[a] The number of possible oligomer probes increases exponentially as a function of probe length. Short probe sequences (3- to 4-mers) occur so frequently in a target DNA that their use in SBH experiments is impractical except for the shortest DNAs. Probes longer than 12 bases are usually used in well-defined sets corresponding to a known target due to the huge numbers of possible probes.

[b] The ''basic read length'' is that target length which, for a given probe length, has a 95% chance of unique assembly. The length is very approximate for probes shorter than six bases.

[c] The ''enhanced read length'' is that target length which provides on average about 10% positive and 90% negative probes and less than one end-mismatched binding site per probe.

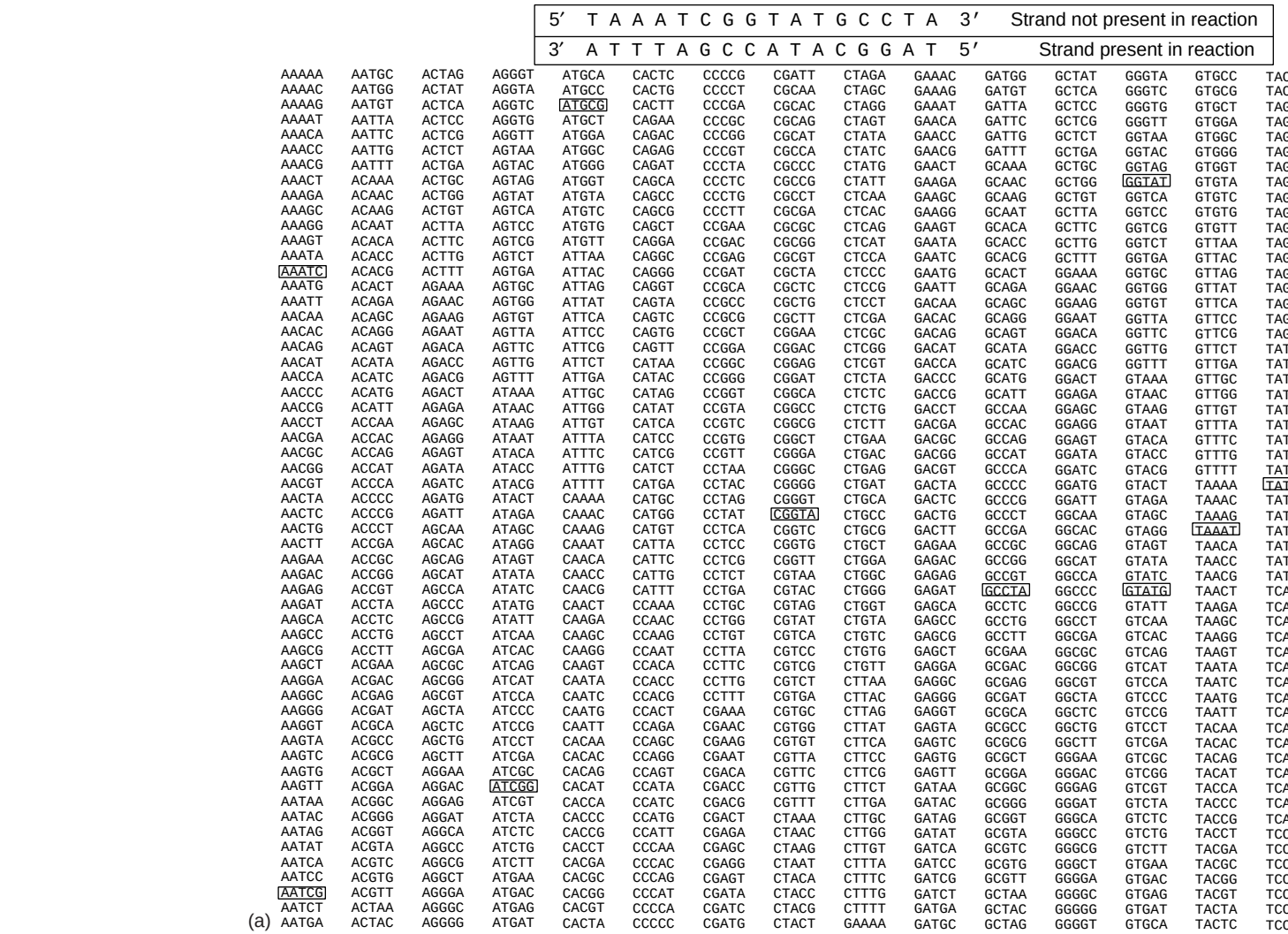

Figure 2 Probe overlap principle (OP). (a) Rare but overlapped positive probes. Hypothetical SBH experiment showing a 16-base DNA sample hybridized with the complete set of 1024 5-mer oligonucleotides. Probes are written in the 5′–3′ direction, and are sorted alphabetically in vertical columns. Positive-binding probes are indicated in boxes. The sequences of positive probes complement the strand present in the experiment and are identical to its complementary strand. The first binding site at the beginning of the target is for probe TAAAT and the second for probe AAATC. These two probes overlap over the four-base sequence AAAT. Every positive probe site (except at the target ends) has two probes overlapping four bases, one in each direction. Only about 1.2% of all probes are positive, with negative probes dominating. In most SBH experiments fewer than 10% of the probes used are positive. (b) (See overleaf) Sequence assembly. Positive overlapping probes from the above SBH experiment are used to assemble the target DNA sequence. In this case, assembly starts with the first positive probe in the alphabetical order (probe AAATC). The assembly process involves defining the end sequence of the initial, or last incorporated probe (AATC in this case), and then checking which of four probes that start with that defined sequence (AATCA, AATCC, AATCG, or AATCT) is positive (AATCG in this case). The new search sequence ATCG is thus defined for the next cycle. Ten probes are found to align to the right and only one to the left. Every base (excluding a few bases toward the target ends) is read with five overlapping probes (seen as vertical columns of five letters). In general, the number of overlapped probes reading each base equals the number of bases in the probe.

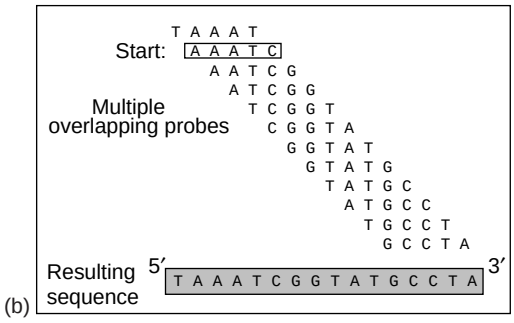

Figure 2 (*Continued*)

probes would be needed. With appropriate detection sensitivity and process miniaturization this astonishing task may someday be accomplished in a single hybridization experiment. SBH may be the only sequencing method that has the capacity for whole genome analysis in a single reaction.

4 VARIATIONS OF THE SEQUENCING BY HYBRIDIZATION METHOD

Three basic variations of the SBH method have been developed, as illustrated in Figure 3. These are discussed below.

4.1 Format 1 Sequencing by Hybridization

Format 1 SBH[10,11] involves using an array of target DNAs exposed to a single labeled oligomer probe or small probe pool in a hybridization step. After washing to remove unbound probe, arrays are scored for signal intensity. As shown in Figure 3, probes that do not match (or only partially match) the target are removed by washing. A subsequent more stringent wash is then used to remove hybridized probe, leaving the array ready for the next hybridization cycle. Probes are applied consecutively to a single array (or replica arrays) of

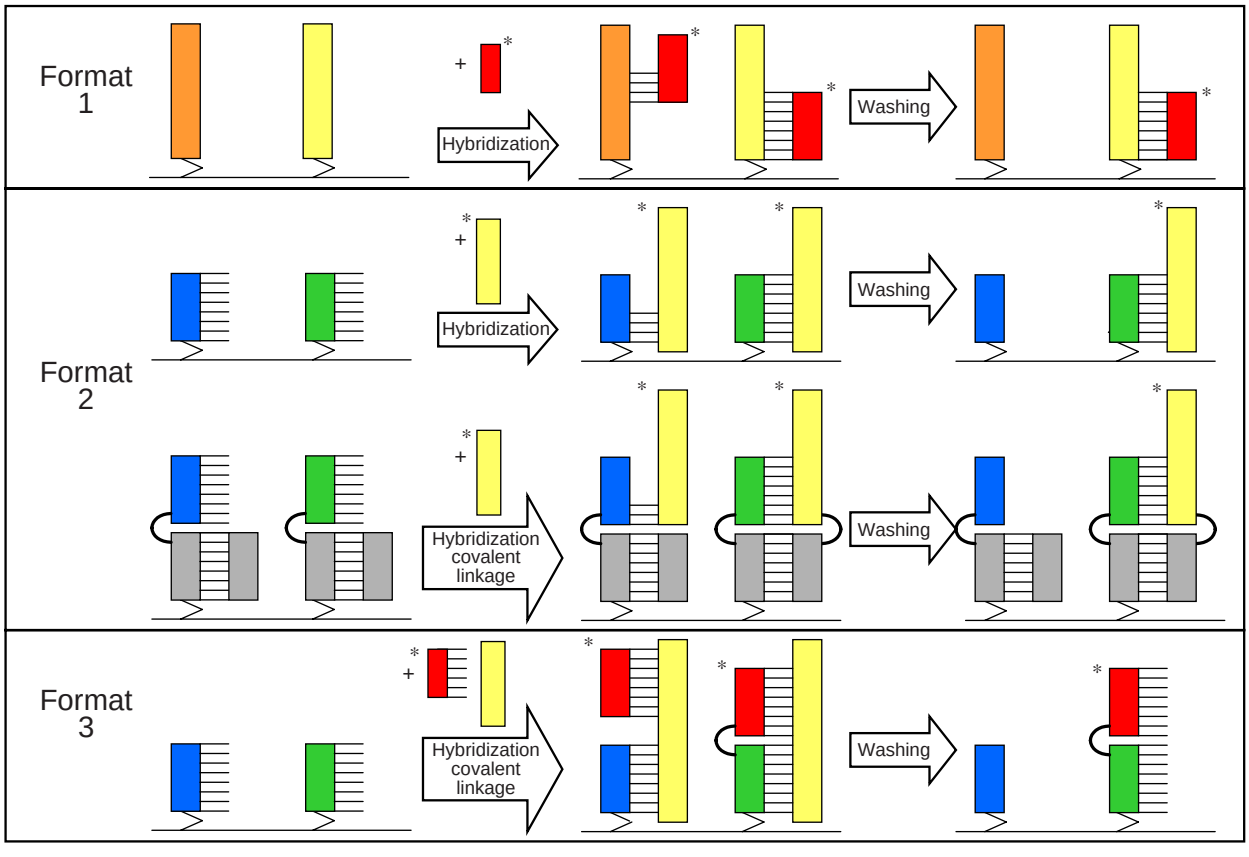

Figure 3 Schema of the three major SBH formats. In each case two probe or DNA spots are shown before and after hybridization and after a stringent wash to remove mismatch probes. In Format 1, target DNA samples (shown as orange and yellow bars) are bound to a solid support and tested for hybridization with labeled (red) probes in solution. A mismatched probe (left spot) binds poorly and is removed in the wash. In Format 2, blue and green probes (attached directly or through an adapter) are bound to a solid support and tested with a labeled DNA sample (yellow) in solution. In Format 3, bound and labeled probes hybridizing at adjacent positions on a target DNA are ligated to form a stable, covalently linked probe structure that identifies a longer complementary sequence in the target. Mismatched probes or those that bind to nonadjacent sites do not ligate and are removed in the wash step.

sample DNAs in a series of hybridization reactions. After all probes from a set have been tested, the subset of probes positively hybridizing to each sample is used to compile its fingerprint or exact sequence. An advantage of Format 1 SBH is that optimal hybridization conditions can be applied because only one probe (or small probe pool) is used at a time, minimizing the variety of different probe–DNA hybrids. This format is effective in analyzing large sets of samples.

4.2 Format 2 Sequencing by Hybridization

In Format 2 SBH,[13,14] a set of probes is attached in an array pattern on a support. A labeled, fragmented DNA sample is then hybridized to the probe arrays. The sample DNA sequence is assembled from the set of positive probe–DNA hybrids that remain after unhybridized DNAs are removed by washing. A disadvantage of Format 2 SBH is that a large variety of probe–DNA hybrids with different thermodynamic stabilities is formed and scored in a single hybridization reaction. A variant of Format 2 SBH (see Figure 3) was proposed by Broude et al.[18] in 1994. In this variation, probes are not end-attached single-stranded molecules, but consist of duplex anchored DNAs with single-stranded tails. The hybridization of target DNAs is facilitated by a stacking interaction with the abutting strand of duplex DNA, and each positive hybridization is converted into a stable covalent bond by the enzyme ligase. This method requires very precise sample fragmentation. To be able to score all probes a fraction of the target molecules has to be cut at each base.

4.3 Format 3 Sequencing by Hybridization

In the early 1990s, Drmanac[19] proposed Format 3 SBH, an advanced combinatorial method capable of scoring extremely large numbers of probes. Two sets of probes are used in Format 3 SBH (see Figure 3). One set of unlabeled probes is covalently bound to the support. An unlabeled target DNA sample is then hybridized to the bound probes, in the presence of a second set of labeled probes and DNA ligase or another coupling agent. When the labeled and unlabeled probes hybridize to the target DNA at precisely adjacent positions (see Figure 3) they are joined to form a labeled two-probe construct covalently bound to the support. Hot washing removes any labeled probes that are not ligated, as well as the target DNA, leaving only the two-probe constructs to be scored as positive signals. This format allows combinatorial scoring of a large number of probes by preparing two small sets of short probes. For example, over 1 million 10-mers can be scored using just two sets of 1024 5-mers. The combinatorial aspect arises from the arrangement of subarrays of bound probes in a larger superarray, each subarray being exposed to one labeled probe or probe pool in a single hybridization reaction. The superarray may contain as many subarrays and hybridization chambers as there are labeled probes or labeled probe pools to allow their simultaneous scoring. The use of two short probes, a unique feature of this SBH format, provides better discrimination than would be achieved with other formats using probes twice as long. In addition, there is no need to cut the target DNA at each base pair, as is required in the Format 2 ligation approach.[18]

In all SBH formats, a DNA polymerase elongation reaction may be used in addition to probe hybridization[18–20] in the scoring process. In this case, unlabeled probes act as primers, and are extended by DNA polymerase by at least one labeled base after hybridization. By using four different labels, one for each of the four dideoxy nucleotides, probes may be specifically extended by one base before final scoring.

5 APPLICATIONS

Oligonucleotide hybridization is a flexible method of screening and sequencing DNA samples. By selecting appropriate combinations of probes and samples, oligonucleotide hybridization can be used for a wide variety of applications ranging from DNA fingerprinting to complete sequencing. There are five general types of applications for oligonucleotide hybridization (Table 2):

Table 2 SBH applications in genomics research

Application	Principle[a]	No. of probes (in a typical test)[a]
Clone contig mapping	SP	>50
cDNA clone clustering/ partial sequencing	SP (OP)	>100
Known polymorphisms/ mutation detection	SP (OP)	>100
Novel polymorphisms/ mutation discovery	SP, OP	>250
De novo sequencing	SP, OP	>4000

[a] The numbers of probes used varies dramatically depending on the application, from ten to a few hundred in simple DNA fingerprinting to tens of thousands in some de novo sequencing applications. Each application takes advantage of one or both of the basic principles that underlie SBH methods: the search principle (SP), in which the presence of complementary oligonucleotide sequences within a target DNA is determined, and the OP, in which overlapping positive probes are used to assemble a target sequence.

- clone contig mapping over a wide range of clone lengths;
- clustering, motif searching, or partial sequencing of cDNAs;
- scoring known polymorphisms and mutations, including single nucleotide polymorphisms (SNPs);
- individual DNA sequencing for discovering novel polymorphisms or mutations;
- de novo sequencing.

The number of probes used and the informational principles behind each application vary widely depending on the experimental goals. Only about 100 probes are needed to obtain clone contigs and cDNA clusters, procedures that involve only the hybridization SP. For partial sequencing and detection of known mutations the number of probes used may range from a few oligonucleotides in some simple applications to thousands in others. When overlapped probes are used to sequence DNA, the probe OP is employed, in addition to obligatory use of the hybridization SP. For discovery of novel mutations or polymorphisms, both search and OPs are used, and the number of probes used may range from complete sets of 4-mers up to complete sets of 20-mers. Another possibility is to use incomplete sets of longer probes, ranging from 7 to 30 bases. Nucleic acid samples can vary from a few hundred bases in length, such as short PCR products, to lengthy cellular mRNAs or genomic DNAs. Each of these five basic oligonucleotide hybridization applications is discussed in detail below.

5.1 Clone Contig Mapping Over a Wide Range of Clone Lengths

Oligonucleotide hybridization can be used to identify overlapping clones that contain identical sequence strings.[9] Since overlapping cosmids share common sequence regions they should show significantly larger numbers of shared positive probes than would be statistically expected from a randomly selected set of oligonucleotides. The authors demonstrated that a set of 100 randomly chosen 12- to 16-mer oligonucleotides is generally sufficient to identify and order cosmid clones if the neighboring overlapped clone pairs share more than 30% of their sequences. The first real application of this idea was the construction of cosmid maps of a yeast genome in 1993.[21] This method, which can be used for long-range mapping of large cosmid or BAC (bacterial artificial chromosome) clones, relies on using a set of 100–200 12- to 16-mer oligonucleotides and additional longer repetitive DNA probes.

Long-range maps of such cosmids and BACs, obtained either by the above-described application of oligonucleotide hybridization or by classical restriction mapping techniques, are necessary for gene cloning and genomic sequencing projects. The BAC or cosmid clones chosen from the maps are generally subcloned into 1- to 5-kb fragments suitable for gel sequencing or SBH. These subclones can be either sequenced randomly in a shotgun approach by gel, or fingerprinted using Format 1 DNA arrays to map or group highly redundant clones. This fingerprinting procedure typically reduces the number of redundantly sequenced clones by 3- to 6-fold,[22,23] resulting in substantial cost and time savings. The map can be constructed by obtaining sequence signatures with a few hundred 7-mer probes from the sufficient number of 0.5- to 2-kb subclones (20 equivalents of the map length).

5.2 Clustering, Motif Searching, and Partial Sequencing of cDNAs

DNA sequence signatures provide a powerful tool for identifying new genes. Drmanac originally proposed obtaining sequence signatures in 1990, using Format 1 as an efficient, high throughput method for characterizing cDNA libraries.[24] Clones arising from the same genes are clustered together based on their similar hybridization signatures. A single representative of each cluster may then be sequenced, dramatically reducing the number of clones that must be sequenced compared to the randomly selected, repetitive DNA sets obtained by the expressed sequence tags (EST) method. This hybridization method has been used extensively[25–27] and is discussed in detail in section 9. At Hyseq, Inc., by processing over 1 million clones per month in various projects, hybridization signatures for over 15 million cDNA clones have been generated and tens of thousands of novel, rarely expressed genes discovered.

Using a large probe set, samples whose sequences are already present in databases can be identified without resorting to complete sequencing.[26,28] A more narrowly directed set of probes can be used to detect coding sequences for any and all variants of a protein motif present in cellular mRNAs. The use of incomplete sets of probes to interrogate only certain segments or types of sequences in the target is called partial SBH.[24]

The gene clustering data obtained from Format 1 SBH experiments are also used to accurately measure gene expression in a given cell or tissue. The size of a cluster for a specific gene is a direct measure of that gene's activity within the test tissue or cell line. This method provides new breadth and precision in monitoring gene expression in any tissue or cell line from which a cDNA library has been prepared. No previous knowledge of which genes are expressed or their sequences is necessary for the analysis. The method allows statistically verifiable results to be obtained by directly counting rarely expressed genes for the first time.[25–27] Comparative analysis of cDNA

libraries from diseased and normal tissues provides a comprehensive list of which genes are up- and down-regulated in the disease process.

Lockhart et al.[29] applied oligonucleotide arrays to measure the expression of previously discovered genes that had been sequenced. In this application, adjacent cells in an array contain a variety of 15- to 25-mer oligonucleotides. For example, a 4000-gene array might contain 80 000 different oligonucleotide spots, with roughly 20 spots corresponding to each mRNA. Manufacture, hybridization, and scoring of these miniaturized arrays are analogous to those used in the sequencing applications described below. When cellular mRNA mixtures are hybridized to the probe array, the signal strengths obtained are in theory proportional to the amounts of specific mRNAs in the mixture, and hence provide a measure of gene activity. Researchers at Affymetrix, Inc. (Santa Clara, CA) have recently reported oligonucleotide arrays representing 30 000 mouse and 40 000 human genes selected from public databases.

5.3 Detection of Known Polymorphisms and Mutations, Including Single Nucleotide Polymorphisms

Since 1979,[7] various hybridization methods have been developed to detect and identify known gene mutations and polymorphisms, including SNPs. An electronic variant of oligonucleotide probe arrays has been developed to detect the presence of known mutations or specific microorganisms in environmental samples.[30] In this method, DNA from lysed bacteria is concentrated and fragmented, then hybridized to a microarray of sequence-specific oligonucleotide probes.

The large arrays currently used in both Format 1 and Format 2 applications allow simultaneous scoring of many independent sequence sites in complex nucleic acid samples. The hybridization search process, with its random parallel access to all potential target sequences, determines which sequence variant or allele is present at each interrogated site. For instance, single arrays can be designed to interrogate 1000 potential mutation sites within 100 genes using complex cDNA mixtures obtained from a single patient. For the first time, such methods allow development of DNA diagnostic population screens.

One approach being vigorously pursued in human genetic mapping involves scoring thousands of SNPs from individual members of affected families or selected population samples. Format 1 arrays of 55 000 pools of 10 to 100 DNA amplicons prepared from individual genomic DNA samples allow extremely efficient population genotyping of known polymorphisms or mutations. Consecutive hybridization of a few probes per allele permits the scoring of millions of genotypes per day. In a complementary approach, Format 2 SBH arrays have been developed that allow simultaneous scoring of 500 SNPs for each individual patient using a set of overlapped probes for each allele.[31]

A comprehensive survey of defined functional SNPs might comprise about 1 million genomic sites that exhibit greater than 1% polymorphism frequency and are known or expected to influence phenotype. Such a survey could be used to obtain an almost complete diagnosis of the health of what is arguably the body's most important organ, the human genome. This type of patient genome description or "Genogram"[32] should prove to be of enormous medical importance. With proper miniaturization, for example using arrays of micron- or submicron-size beads,[32] Format 3 SBH is probably the best-suited technique for this application, which would require high-fidelity testing of several million probes more than 10 bases in length.

5.4 Discovery of Novel Polymorphisms or Mutations by Individual DNA Sequencing

This application usually requires large or complete sets of oligomer probes. A common procedure is to use all probes corresponding to a given reference sequence. In 1991, Strezoska et al.[33] demonstrated the potential of using Format 1 SBH to confirm the absence of base changes in an individual DNA sample. Using 100 octamers corresponding to a known 100-base sequence, the researchers were able to accurately assemble the test sequence based on the hybridization results obtained. The use of several overlapping probes to test each base assured high accuracy. In a more advanced application, Drmanac et al.[34] used a complete 7-mer probe set to accurately sequence over 10 000 bases of p53 exon DNA from a number of samples. The accuracy, turnaround time, and throughput of this Format 1 application make it suitable for the large sample volumes used to detect rare polymorphisms. The use of complete sets of probes allows the sequencing of any gene, and can determine virtually all types of mutations and polymorphisms. This work is discussed in detail in section 9.

Southern et al.[35] demonstrated the use of Format 2 SBH to sequence 24-base model targets using a 256-member array consisting of all possible 8-mer probes made up entirely of purines. Subsequently, researchers at Affymetrix, Inc., developed Format 2 SBH methods designed primarily for sequencing certain variants of individual DNA samples when a prototype sequence is known.[36] Using lithography, they produced miniaturized arrays consisting of sets of 18- to 25-mer oligonucleotide probes that are designed to detect unchanged sequences and those with single base substitutions. Probes are selected so that each base in the target sequence is covered

by four oligonucleotide probes that are identical except for one centrally placed base that varies in each of the four oligonucleotides. Large probe arrays may be able to analyze DNA segments several kilobases in length.[36] Since this method uses a very select, partial set of long probes, some mutations may be detected but cannot be precisely sequenced, particularly in the case of insertion mutants. Sequence-specific arrays have been constructed and used to sequence the HIV clade B protease, BRCA1, and ATM genes.[37–39]

Broude et al.[18] proposed the development of universal microarrays containing duplex DNA anchors with all possible single-stranded *n*-mer tails. This variant of the Format 2 SBH procedure involves ligation of the target DNA to matching probes. Gunderson et al.[40] used this method to sequence approximately 1 kb fragments of the HIV pol and cytochrome *b* genes in test DNA samples. The researchers used complete arrays of 8-mer and 9-mer probes.

Hyseq, Inc., has developed the HyChip™ system for accurate complete sequencing of any candidate gene (including heterozygotic sites) in individual genomic DNA samples and discovery of all types of polymorphisms and mutations. The HyChip™ system scores all possible 10-mers (1 048 576 combinations) using a Format 3 SBH ligation procedure involving two sets of 5-mer probes.[19] A sequence read length of 2000 bases has been demonstrated (S. Drmanac et al., paper in preparation).

The examples above demonstrate that SBH can provide accurate and efficient polymorphism discovery, information that is of critical scientific and medical importance. There is a growing need to sequence most of the genomic DNA of perhaps 1000 individuals to discover significant human polymorphisms (those with 1% frequency or greater) for correlation studies, and to develop Genogram tests for diagnosis.

5.5 De Novo DNA Sequencing

De novo sequencing of complex genomes is the most challenging SBH application.[11] Successful sequencing of different DNA fragments has been reported by some researchers. Drmanac et al.[41] demonstrated the accuracy of the SBH Format 1 technique by sequencing unknown short primate DNA sequences using a set of octamer oligonucleotides in a blind test. The probe set contained both exact match and mismatched probes and was tested against three 116 base pair homologous regions obtained from different primate T-cell receptor loci. Exact sequences were determined without a single error. False-positive probe hybridization was found to occur infrequently and was easily corrected due to the high data redundancy inherent in the SBH method.

In 1996, Hyseq, Inc., researchers sequenced nearly a million bases of cDNA clones using a combination of data obtained from single-pass gel sequencing and from a complete set of 7-mer probes scored by Format 1 SBH (unpublished data). Blind tests of Format 3 SBH demonstrate the reliability of using hybridization and ligation of adjacent probes to sequence unknown DNA samples (S. Drmanac et al., paper in preparation). By scoring all possible 10-mers (obtained by ligating two 5-mers) using HyChip™ 5-mer arrays, sequences as long as 800 bases may be assembled directly from the hybridization data (see Table 1). In the future, scoring longer probes (e.g. 14-mers) using pools of distinctly labeled 6- to 7-mer probes and miniaturized arrays of 7- to 8-mers may provide de novo read lengths of 10 kb or more, a task that is difficult if not impossible to achieve with nonhybridization sequencing methods.

6 BIOCHEMISTRY

6.1 Sample Preparation

Most current SBH applications, with the exception of Format 2 expression studies, use nucleic acid samples ranging in length from 200 to 2000 bases. Initial DNA samples are typically obtained by cloning or PCR. Some samples, such as multiplex PCR preparations, are mixtures of fragments, with a total length of up to several kilobases of DNA. PCR is the dominant preparation method, due to its ease of use, reproducibility, and high yield. Plasmid and phage preparations have also been used.[41,42] High-throughput PCR procedures have been developed to create the large DNA arrays used in Format 1 SBH applications.

Figure 4 shows the outcome of PCR preparations of 22 genomic DNA samples obtained from individual patients. A 10^8-fold amplification of as few as 100 copies of a ~1 kb ApoB gene segment was achieved, with only a two-fold yield variation between samples. Negative controls with no patient DNA have undetectable signals. Cross-contamination of up to 5% in the final arrayed samples can be tolerated due to the small contribution (below the level of mismatch signals) of such contaminants to the signal.

Throughputs of up to 50 000 PCR reactions per day are routinely achieved for current Format 1 SBH gene discovery projects.[43] Initial samples for PCR preparations may consist of cloned DNA in bacterial plasmids or purified genomic DNA. The small amounts of proteins and other contaminants present do not hinder the hybridization reaction, allowing simplified, automated sample preparation.

In order to prevent secondary structure formation within the target DNA prior to exposure to probes in Format 1 SBH, denatured target strands are anchored

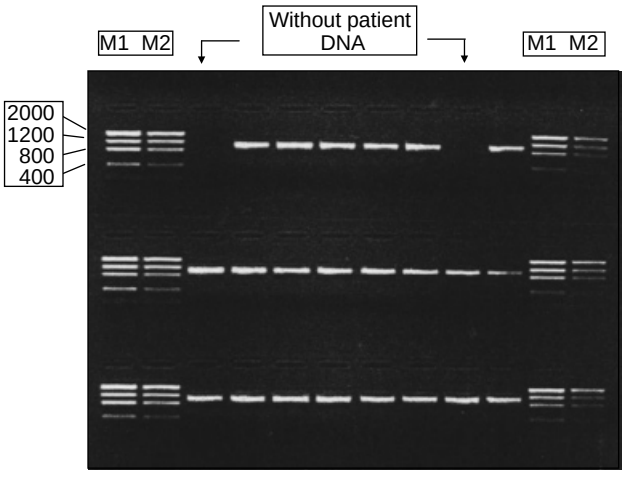

Figure 4 Sample preparation for an ApoB gene segment. PCR of a 967 base pair ApoB amplicon is performed on 22 patient DNA samples and run on an agarose gel with size markers (M1, M2). The 40 ng DNA bands are obtained by 10^8-fold amplification from about 100 initial chromosomes. There is no visible contamination in negative control reactions having no patient DNA.

to the substrate by a pseudo-covalent bond every 30–50 bases, thus preventing renaturation. In Formats 2 and 3 SBH, the target DNA is fragmented into small pieces about the size of the probes and then denatured, to assure that internal secondary structures within the target are less stable than probe–target duplexes. Under the excess probe conditions used, such target–target duplexes are simply out-competed kinetically. Owing to its amenability to fragmentation and labeling, the final target in most Format 2 SBH applications is RNA synthesized from clone inserts.[35–39]

6.2 Probe Preparation

Oligonucleotide probes may be synthesized quickly and accurately on modern DNA synthesis machines, with thousands of specific, high-quality probes produced in a few days. In some applications, additional purification of oligonucleotides by gel or column chromatography is necessary. No such purification is necessary for Format 1 procedures. In commercial batches of several thousand unique probes, less than 5% typically prove to be unsuitable for hybridization due to impurity or improper synthesis.[34] In Format 2 SBH (especially when ligase is used) and Format 3 SBH, probe purification is appropriate to avoid shorter products. In the case of in situ synthesis of probe arrays no purification procedure is possible, creating potential quality control (QC) and reproducibility problems between arrays.[35–40]

An additional concern with in situ synthesized miniaturized arrays is the density of probe molecules per spot. The density of usable probe strands per unit of surface area is a critical factor in determining signal strength. Too few or too many probes (which prevents target access) results in low or undetectable signals.

For a variety of reasons,[11,12] pools of probes are sometimes synthesized by incorporating more than one type of base in a single base-addition cycle. With such degenerate probes the goal is that equimolar incorporation of all bases occurs. To assure this condition, it is necessary to premix bases in proper ratios instead of depending on correct uptake by instrument pumps.

Most current SBH applications use natural bases. Modified bases may be used to equalize hybridization strength for G + C-rich and A + T-rich probes, and to modulate the overall stability and specificity of base pairing. Many initial studies are underway to test how these factors are affected by different chemical modifications of natural bases. The base modifications must be compatible with DNA ligase and polymerase or other enzymes if these enzymes are part of the probe-scoring process.

6.3 Array Chemistry

In preparing DNA arrays, attachment of strands to the solid substrate is maximized to reduce the quantity of soluble DNA or label needed to produce a strong signal. In addition, nonspecific binding of labeled stand to either the substrate or bound probes is minimized to reduce the background signal. This is achieved by choosing inert substrates or adding appropriate components of hybridization buffers that minimize or compete for nonspecific binding sites. In Format 1 SBH, the nylon membranes used for making arrays are typically a mesh of fibers, allowing three-dimensional binding via pseudo-covalent bonds. Subsequent ultraviolet cross-linking of T bases produces covalent interstrand bonding, leading to retention of a higher proportion of input strands. Approximately 60–90% of the initially bound DNA is retained, even after repeated cycles of hybridization and washing. In Format 2 SBH the substrate is usually flat, nonporous glass. The density of bound probes is a direct function of the number of activated sites per unit surface suitable for attachment of either pre-made oligonucleotides or initial monomers used in subsequent in situ synthesis. The probe density achieved is a function of surface activation and coupling chemistry. Other potential substrates include polyacrylamide gel-coated glass[15] or micro beads.[32]

6.4 Labeling

The soluble DNA strand in an SBH array experiment, whether probe or target DNA, is usually labeled to allow scoring of hybridization events. Any part of the strand

may be labeled, but typically the terminal base of a probe or target DNA fragment is chosen. The labeling procedure may be enzymatic or chemical. For instance, in Format 1 SBH applications, polynucleotide kinase adds a ^{33}P-labeled phosphate group on the 5′ end of oligonucleotides, while in Format 2 SBH applications, RNA polymerase is used to incorporate fluorescent-labeled RNA bases to create a multiple chromophore attached to the target DNA. Labeled groups can also be added during chemical synthesis of oligonucleotides. The most common labels are radioactive or fluorescent tags that are readily quantified by advanced detection devices such as Phosphoimagers (Molecular Dynamics, Sunnyvale, CA) or fluorescence scanners (GSI Lumonics, Inc., Watertown, MA).

Radioactive tagging is the traditional labeling method used in nucleic acid chemistry and is still widely used in Format 1 SBH applications. ^{33}P has several advantages as a radioactive label for nucleic acids. One is an optimal α particle energy, which provides the rapid, sensitive detection characteristics of higher-energy nuclides, but with the high resolution and added workplace safety of a particle that exhibits minimal air penetration (less than 1 cm). The other advantage is a short half-life, requiring only a 6-month decontamination period for any radiolabeled wastes. ^{33}P-labeled array spots need to be at least 200–300 μm apart to be successfully resolved, and the lower limit of detection is about 10 labeled duplexes per square micrometer.

Fluorescence labeling can be direct, in which case an organic chromophore is covalently attached to the DNA, or indirect, such as when a tag attached to DNA binds a chromophore-carrying molecule. In either case, chromophore-labeled duplexes are excited by light of the proper wavelength and the resulting emission signals are measured by a fluorescence detector. Although detection of a single molecule's fluorescence is theoretically possible, in reality both reader sensitivity and background fluorescence are limiting factors in using fluorescent labels for SBH. Signal detection limits with advanced commercially available readers are approximately 1–10 chromophores per square micrometer. Direct fluorescence has been the labeling method of choice in miniaturized SBH formats, although mass spectrometry is becoming increasingly popular.

6.5 Discriminative Hybridization

6.5.1 Hybrid Stability

SBH is possible because the specificity of binding between single-stranded DNAs and complementary oligonucleotide probes (expressed as binding strength or recognizable duplex structure) can be used to discriminate matched from mismatched probes. The experimental parameters used to influence binding/dissociation generally include hybridization and wash temperatures, wash time, and other solution conditions such as salt concentration (see Wetmur[44] for a more detailed treatment). In the case where one of the strands is attached to a solid support, the dissociation temperature, T_d, at which 50% of duplexes have dissociated is used as a measure of hybrid stability. Figure 5 shows DNA–probe dissociation curves for a typical full-match 8-mer probe and two different single-mismatch 8-mer probes as a function of temperature. The separation of the curves suggests that near T_d for a mismatched probe hybrid, most full-match probes remain bound to the target DNA, allowing discrimination of the two probes.

In general, both thermodynamic and kinetic factors influence the binding and dissociation of DNA–probe hybrids, but it is generally kinetic dissociation rates that are used to discriminate between matched and mismatched probes. In free solution, T_m (the temperature at which 50% of possible duplexes are melted) is a linear function of GC content. T_m may be converted to T_d by a simple equation.[44] In oligonucleotide hybridization, T_m is a more complex function of actual base sequence, and is dependent on ΔG, the standard free energy.[44] For an 8-mer hybrid, ΔG is dependent on three factors: a ΔG sum of nearest-neighbor contributions for seven (of 16 possible) dinucleotide interactions present in the probe, a temperature-independent initiation component, and a dangling end component. Measured values for the various

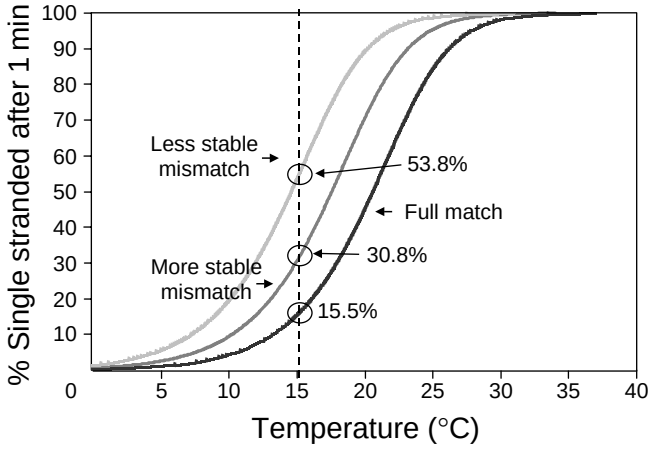

Figure 5 Dissociation curves of typical 8-mer probes. Graphs represent percentage of denatured hybrid after a 1-min wash at a given temperature. At 15 °C, the fraction of melted hybrids for mismatched probes is two to three times higher than that for full match probes. Probe-DNA hybrids formed with full match probes tolerate higher wash temperatures or longer wash times than mismatched probes, allowing experimental discrimination by manipulating wash temperature and/or wash time.

component contributions have been reported.[45,46] An internal mismatch leads to a decrease in two nearest-neighbor contributions, while an end mismatch results in a decrease of one nearest-neighbor contribution. The stacking interaction of two abutting but noncovalently linked probes hybridized to a longer complementary strand generally has the same effect on stability of the longer probe as lengthening it by one base.[44] Use of this stacking effect in SBH has been proposed by Khrapko et al.[16]

Dangling end contributions are dependent on the single-stranded DNA sequences that surround the DNA–probe duplex, leading to the observation that the same short duplex can have a different stability in different sequence settings.[41] While thermodynamic measurements on synthetic model duplexes generally agree with the theory, no precise prediction of hybridization signal strength has been possible with natural target DNAs. This is due to various sequence context effects on duplex stability or on physical access of probes to binding sites.

Duplex stability (as measured by T_d) decreases with shorter probe length. By manipulating target DNA and probe concentrations one can generally maximize the amounts of remaining duplex product at sub-room temperatures, even for very short probes. The sensitivity of detection provided by radioactive- and fluorescent-labeled probes creates a lower practical limit of approximately 5- to 6-base pair duplexes in SBH. In special conditions, such as ribosome binding sites, stable, base-specific duplex binding occurs in codons consisting of only three base pairs. To improve detection and reduce the number of probes needed for various applications, the use of probes with degenerate ends was proposed for fingerprinting[47] and SBH.[10,11] These are pools of probes that share a common core sequence but have one or more varying bases at the ends. For instance, a complete set of octamer probes with a single degenerate base at each end will exhibit the duplex stability of a decamer probe set. Discrimination is also improved, because all mismatches that occur in the core sequence are internal. Such an 8-mer probe set consists of 65 536 probe pools, as opposed to 1 048 576 possible decamers.

6.5.2 Discriminative Hybridization

Obtaining sufficient matching signal strength over detection background, and at the same time achieving significant discrimination between match and mismatch signals, are critical goals of the probe hybridization process. Drmanac et al.[48] presented a mathematical treatment of the discrimination kinetics of the SBH process that includes the influence of background signal. Using equations describing the association and dissociation of DNA–probe complexes, they showed that optimal discrimination, defined as the ratio of amounts of full-match over mismatched duplexes, is primarily determined by wash temperature and time. Based on their analysis and later experimental verification, they concluded that increases in temperature generally reduce discrimination values (Figure 6). Very low temperatures, on the other hand, require an exponential increase in wash time. The ideal conditions for discrimination of short probes (five–nine bases) were found to be an excess of initial hybrid, moderately low wash temperatures, and extended wash times, usually over 15 min.

Using optimized washing conditions, there are no limits on discrimination values if sufficiently large quantities of hybrid (relative to background) are obtained in the hybridization of large quantities of

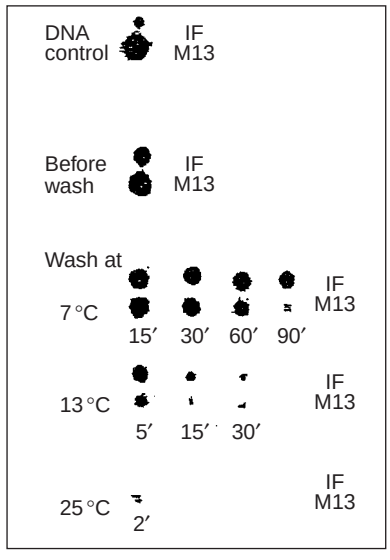

Figure 6 Improved discrimination with low-temperature wash. Experimental hybridization images show disproportionate removal of labeled, hybridized probe between full match and mismatch targets as a function of wash time and temperature. To make the phenomenon visually obvious, 50 times more mismatch DNA (M13) is used than full match sample (IF), as shown with control probe in the ''DNA control'' panel. After hybridizing test probe TGCTCATG, and before wash, the M13 signal is still stronger than IF full match signal (''Before wash'' panel). Both samples are then washed at the indicated temperatures and times. At 25 °C, the full match signal is invisible, and the mismatch signal is still detectable after a 2-min wash. The best discrimination is achieved with a long wash at the lowest temperature. At low temperatures the full match signal remains very high (compared to other temperatures), while the mismatch signal approaches background value. The two-fold stronger signal of full match over mismatch targets, combined with the fact that 50-fold more mismatch DNA is used, results in a 100-fold discrimination factor.

target and probe. Figure 6 illustrates 100-fold discrimination achieved in a specially designed test experiment.

There is a thermodynamic explanation for why better discrimination is achieved at lower temperatures and for shorter probes. Discrimination is a function of $\Delta\Delta G$, the difference of ΔGs for matched and single mismatched duplexes. For the same mismatch this parameter is independent of probe length and is inversely proportional to temperature.[45,46] For example, the TGTAC/gGTAC full-match/mismatch 5-mer pair has the same $\Delta\Delta G$ as the TGTACCC/gGTACCC 7-mer pair. Since longer mismatched probes require higher melting temperatures (which reduces the $\Delta\Delta G$ term) such long probes generally exhibit poorer mismatch discrimination than shorter oligomers. For example, T_d values for perfectly matched 14-mer and 20-mers are 41 and 57 °C, respectively,[44] and their dissociation curves would be displaced far to the right of those for the octamers shown in Figure 3. On the other hand, short probes can be hybridized and washed at much lower temperatures, since short mismatch duplexes can be denatured at such low temperatures. At these temperatures, the $\Delta\Delta G$ term is relatively larger, leading to enhanced mismatch discrimination. As a result of these factors the use of two shorter probes instead of a single longer one in Format 3 SBH results in greatly improved discrimination in scoring probes over 10 bases in length.

The basic low-temperature hybridization and washing conditions developed by Drmanac et al.[48] are routinely used in most current SBH applications. These conditions usually involve hybridization at 0–15 °C, using increased concentrations of both probe and target to accommodate the decreased stability of short or AT-rich duplexes. Although 15- to 25-mer oligonucleotides (commonly used in Format 2 SBH applications) can be hybridized at room temperature and above,[7] most Format 2 procedures use low-temperature conditions, primarily to reduce the amount of sample required. Many of the procedures and components used in SBH hybridization and washing are derived from traditional molecular biology hybridization applications.

Format 2 SBH arrays require the simultaneous formation of large numbers of duplexes of different stability. This poses the problem of choosing one optimal temperature (or other hybridization conditions) to maximize discrimination among the many duplex types. To address this situation, Khrapko[16] proposed to measure the differential melting of individual duplexes. This approach is experimentally demanding, and no simple solution to the problem has yet been found. The optimized low-temperature hybridization conditions developed by Drmanac et al.[48] maximize overall discrimination and the number of different duplexes that give measurable signals. Since different signals may be produced by different probe–DNA hybrids, sophisticated scoring systems that recognize more than a single threshold signal are generally used.

While hybrid stability is predominantly used in SBH experiments to detect the presence of complementary sequences in complex nucleic acid mixtures, other means may be applied to achieve mismatch discrimination. Coupling probes using either a chemical agent or ligase (as in Format 3 SBH) is one method of improving discrimination.[19] This procedure transforms short-lived complexes into stable, covalently bound products, producing stronger and more uniform signals even under very stringent hybridization conditions. DNA ligases with increased enzymatic specificity may also be used.[49] Another possibility is to use agents that cut nucleic acid strands at mismatched positions.[50]

7 SEQUENCING BY HYBRIDIZATION AUTOMATION

SBH procedures are generally large-scale analyses that may involve the transfer, storage, and processing of tens of thousands of DNA samples and probes each day. To scale up for such processing, multi-well microtiter plates are used to organize large numbers of dissolved DNA samples or probes. Standard 8 cm × 12 cm plates used in SBH may contain from 96 to 1536 wells. Transfer devices such as robotic pipettors and pin tools are used to create surface-bound arrays or to transfer samples/probes from one plate to another. These transfer devices are arranged in arrays of 8 to 1536 units and can be used to spot samples or probes in arrays on solid supports. The resulting arrays range in size from 100 to 100 000 nucleic acid spots. Electronic bar codes are used to track samples and probes with each transfer step, and to input information in a database. Bar coding eliminates the need for manual recording and tracking and greatly reduces the possibility of labeling error. Figure 7 summarizes the DNA handling and hybridization steps of a typical Format 1 SBH experiment.

Automatic handling of plates containing DNA samples and probes is convenient but not essential, especially when using plates with 384 or more wells. However, high-density plates provide progressively smaller sample volumes (50 μL for a 384-well plate down to 10 μL for a 1536-well plate), which pose potential evaporative loss problems during transfer operations. A variety of techniques have been developed to minimize evaporation and sublimation. These include layering mineral oil over water-based DNA solutions, covering plates with plastic wrap before freezing, and adding glycerol to increase

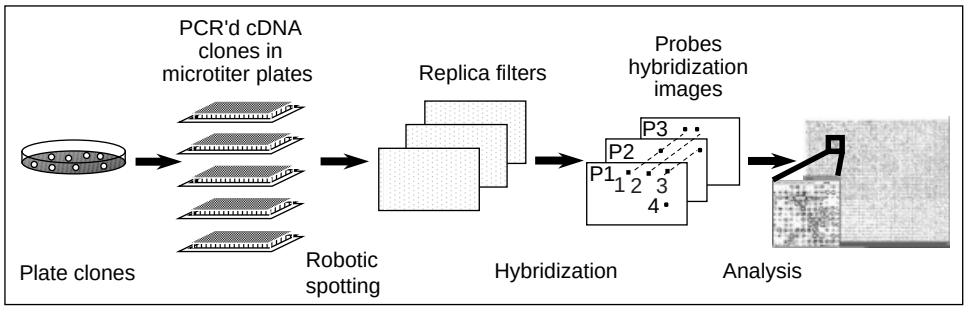

Figure 7 Format 1 SBH clone analysis procedures. DNA clones are first picked from agar plates and arranged in microtiter plates. Cloned DNAs are amplified by PCR in microtiter plates, then robotically spotted using a pin tool to replica filters, which may contain up to 55 000 DNA spots. Each filter is hybridized with a single probe, then washed and analyzed for signal intensity at all array positions. After scanning, replicate filters are stripped by a rigorous washing procedure to remove labeled probe and re-hybridized with new probes.

viscosity and minimize freezer damage to live cDNA clone cultures.

Arraying more than 100 DNA spots per square centimeter usually involves using tools with precise delivery characteristics mounted on an XYZ robotic station with a consistently repeatable resolution of 10–100 μm. Liquid transfer may be accomplished using a variety of physical principles, including active pipetting by hydrostatic or piezoelectric forces, capillary action, or simple surface tension on the tips of pins. Simple pin tools must be ''reloaded'' after each transfer, while other devices with a reservoir may transfer replicate aliquots of the same sample. Good reproducibility is essential in all arraying procedures, including spot-to-spot transfers from a single channel, channel-to-channel reproducibility within an array, and array-to-array reproducibility. At Hyseq, automated offset printing is used to create a batch of four replicate 55 000-spot arrays on nylon membranes. An 864-pin tool transfers cDNA samples from sixty-four 864-well plates, spotting a few nanoliters of sample with each transfer cycle.[43] This process has been in constant daily operation for over 2 years. Similar procedures are used in many laboratories to create arrays of samples or probes on glass microscope slides.

In situ synthesis of oligonucleotide probe arrays is a computer-controlled photolithographic process in which nucleotides are added sequentially in specific patterns. Specially designed photolithographic masks control the order of light-activated coupling reactions at each array position. This in situ synthesis process may be used to create dense arrays containing over 100 000 probes on a single 1.28 cm^2 substrate.[36–40] Significant amounts of shorter oligonucleotide products are produced as by-products of any long probe synthesis. QC procedures on such arrays are often not practical.

Oligonucleotide hybridization procedures are mechanically simple and are easily automated. Robotic equipment used for hybridization must be able to deliver and remove buffers and sample or probe solutions and a thermostatically controlled heating system is needed to control hybridization temperatures. A method of mixing liquid layers above surface-bound arrays may be used if the overall liquid depth is more than 2 mm. The larger the array, the more demanding the requirements for automation engineering. For example, small numbers of microarrays on glass slides can be processed in simple table-top hybridization devices, while Format 1 arrays containing tens of thousands of samples are processed in large robotic stations. The use of XYZ robotic stations for hybridization procedures allows flexible access to multiple hybridization chambers on a single platform. Drmanac et al.[34] used removable partitions on Format 1 SBH filters to create 384 individual hybridization chambers, each receiving a different labeled oligonucleotide probe for rapid parallel hybridization and washing.

The data readout of an SBH experiment is an array of signals corresponding to hybridization strength at each probe–DNA spot. The image is usually generated by scanning instruments and is composed of rows of pixels. Since each dot is represented by 10 or more pixels, a single large array may contain as many as 10^8 pixels. As described earlier (see section 6.4), owing to the practical limitations on probe and DNA concentrations, scanners must be able to detect between 0.1 and 10 labeled duplexes per square micrometer. Most large SBH applications require high image scanning speed for collection of multiple images per scanner per day. Confocal laser microscopy is generally used to achieve the sensitivity needed for array imaging. For more advanced applications, charge-coupled device (CCD) cameras with 1000×1000 or larger pixel arrays may provide the solution if detection sensitivity of less than 10 labeled duplexes per square micrometer is achieved. Such instruments can acquire multiple images per second, resulting in readouts of over 10^8 pixels per minute.

Future trends in SBH equipment development, particularly in the field of DNA diagnostics, will focus on the design and construction of efficient systems for sample preparation, hybridization, and data readout.

8 DATA ANALYSIS

All current hybridization-based fingerprinting and sequencing applications are based on arrays. The initial data obtained from each unit (DNA or probe spot) in the array is a hybridization signal intensity measured with a detector device. Since individual arrays or sets of arrays may range in size from 1000 up to 1 million array elements, large amounts of primary data must be collected and processed. Automated processing techniques have been developed to handle such large volumes of data. Although user-friendly interfaces have been developed that allow manual intervention in data analysis, the trend in SBH data analysis is to automate the entire process. Information-processing demands of the SBH method can be met by current computer technologies in a cost- and time-efficient manner.

Data analysis consists of four basic tasks, as outlined in Figure 8: (1) sample and probe tracking; (2) image analysis, to transform primary measurements into raw sample/probe responses; (3) normalization and QC of raw data; and (4) analysis of normalized data by application programs to obtain sequence or fingerprint results.

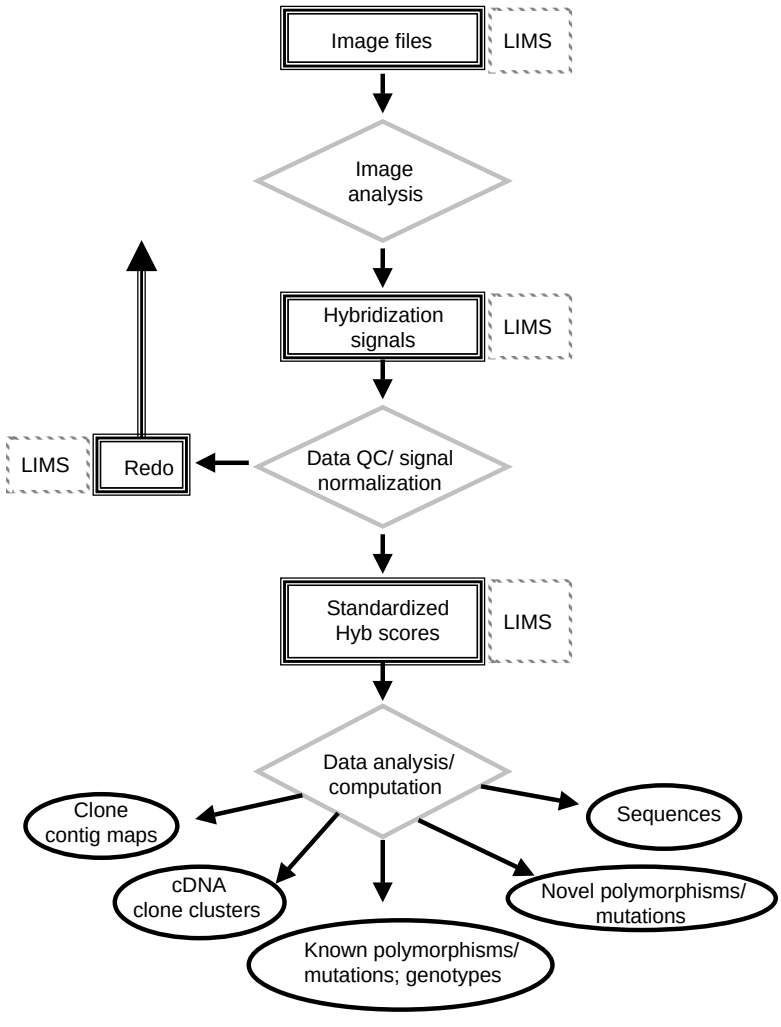

Figure 8 Flowchart of SBH data processing. Hybridization arrays are scanned to capture signal intensity at all array positions, then analyzed using specialized software to locate each array position (image analysis), and to evaluate and normalize signals. Appropriate sample/probe information is collected by a laboratory information management system (LIMS) and stored in the database. Using LIMS, unsatisfactory data are reported to the laboratory for repeated experiments. Normalized signals and LIMS are then used by other computational programs to assemble sequences, map clones, detect mutations and/or identify target DNAs and report results.

The need for precise sample and probe tracking is not specific to SBH applications. Automated laboratory information management systems (LIMSs) have been developed for other genomic and clinical laboratory applications to accurately track sample handling. In Format 1 SBH applications, all microtiter plates, arrays, exposure cassettes, robotic stations, freezers, etc. are bar-coded to allow careful tracking of all sample and probe handling procedures in a central LIMS database. Various programs and user interfaces are used to facilitate both ongoing experimentation as well as proper identification of samples for data analysis. In other SBH formats LIMS procedures are less extensive but still play an important role in data gathering and analysis.

Image analysis for SBH applications has involved the development of specific software packages, since accurate signal processing of spots in predefined arrays (Figure 9) is not a common image analysis application in other technical fields. A pixel image of each hybridized array is generated, with each dot in the array typically represented by 10 to 100 pixels. The first task of SBH image analysis is to define a line grid representing the rows and columns of sample spots in an array. Hybridization images typically do not have positive signals at every unit position in the array. The task of defining this grid becomes progressively more difficult as the percentage of positive spots in the array becomes smaller. Accurately assigning a small percentage of positive spots to an imperfect grid can be accomplished by advanced image analysis programs. Once the grid is accurately established, each spot's signal is quantified by averaging the signals from pixels covering that spot. Intensely positive neighboring spots may create occasional signal spillover, which may need to be corrected to achieve accurate data analysis.

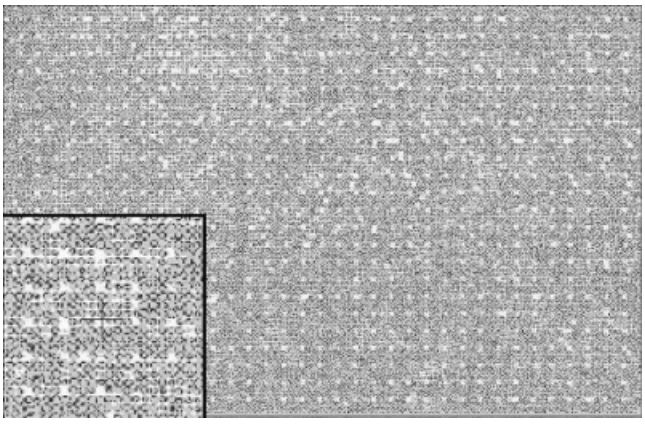

Figure 9 An SBH array image containing 55 000 cDNA spots. Inset shows enlarged region covering approximately 500 DNA samples. Signal intensity is proportional to the amount of bound DNA, the concentration of labeled probe used, numbers of full match and mismatch binding sites, and probe–DNA hybridization strength, which itself is a function of probe length, sequence, and hybridization conditions.

Normalization and QC procedures maintain the quality and statistical integrity of SBH data. Data normalization corrects for experimental variables such as differences in labeling efficiency of different probes or samples, variations in amount of spotted nucleic acids, and other experimental factors. Other normalization procedures measure internal consistency of duplicate data samples, or compare unknown samples and known controls. QC procedures eliminate data that cannot be adequately normalized, or that remain inconsistent after normalization. Also eliminated are data that do not meet empirically measured reproducibility and quality standards. These built-in QC controls help identify and eliminate problem data resulting from misidentified or poorly hybridized samples, probes, and arrays, allowing portions of experiments to be repeated (for more details see Drmanac and Drmanac[43]).

Application-specific data analysis software is developed for each specific combination of data formats and user needs. Owing to the large size of many SBH data sets, obtaining final experimental results is usually computationally intensive, particularly when data elements require pair-wise comparisons. Specific programs capable of extracting results from normalized hybridization data operate on one of two principles: (1) the normalized signal strength is used directly as a measure of the probability of a full match; or (2) the signal is converted into a discrete positive or negative outcome using a function designed for the specific experimental protocol. Owing to limitations associated with each of these approaches, application programs usually combine both methods. In addition, such programs are equipped with statistical routines designed to compensate for certain levels of false-positive and false-negative data.

The inherent redundancies in SBH data can compensate for surprisingly high error rates that may occur in some low-quality data sets. An ''expert system'' type of analysis is implemented to process such complex data sets. In this approach, a decision-making tree is developed to partition data based on signal strength or type of probes, and then to apply different criteria for each subset. For example, hybridization signals may be first divided into strong, weak, or insignificant signals, and then statistical tests used to determine positive and negative probes among the strong and weak signal groups. The nature of SBH data analysis makes it suitable for neural network applications, in which decision-making parameterization is learned by software analysis of data sets with known outcomes. The quality of processing software is measured by its ability to extract all useful information from complex hybridization data sets that contain noise

and error. With proper data analysis, accurate clustering, mapping, and sequencing results are obtained within tolerable error limits. Algorithms for some of these applications are described in section 9.

9 SELECTED STUDIES

In recent years, oligonucleotide hybridization technology has been applied in a variety of important research studies. These include de novo sequencing,[41] gene discovery and fingerprinting,[25] and genetic polymorphism studies.[34] These studies are described in detail below.

9.1 De Novo Sequencing

Drmanac et al.[41] were the first to report successful de novo SBH sequencing. In a blind test, the researchers used both traditional gel methods and SBH to sequence homologous primate T-cell DNA fragments cloned into an 8 kb M13 vector. The similarity of the DNA targets and the occurrence of many single-base mismatches between targets and probes provided a stringent test of the SBH method despite the reduced overall experimental complexity.

In the first stage of this study, one research team sequenced three 2-kb T-cell β loci inserts by traditional sequencing methods. These homologous inserts, derived from human and two rhesus monkey species, were shown to be 92–94% similar in sequence. A second research team used SBH to sequence homologous 116 base pair regions of these clones.

The gel sequencing team constructed a list of 272 octamer probes for determination of the sequences by SBH. The selected probe list contained about twice as many nonmatching as matching probes. The matching probes consisted of all 8-mers with more than three G + C bases or 9-mers with one to two G + C bases that occurred in the 116 base pair DNA segments. To increase the experimental complexity, similar probes corresponding to a fourth primate species (not included as a DNA target) were included in the probe list as well. The SBH research team was informed the list would allow determination of at least one of the sequences and that the other two sequences were similar.

Figure 10 shows representative hybridization images obtained using different 8-mer and 9-mer probes tested against a filter containing 96 DNA samples spotted in duplicate. The filter was prepared directly from M13 phage cultures containing the 116 base pair inserts. Positive and negative DNA controls consisting of M13 DNA were also included, to allow calibration of the hybridization signals for each probe. Pools of sixteen 10- and 11-mer probes containing two degenerate bases as well as the given 8- or 9-mer core were used in these hybridization studies. The longer probes increase hybrid stability and hence signal intensity.

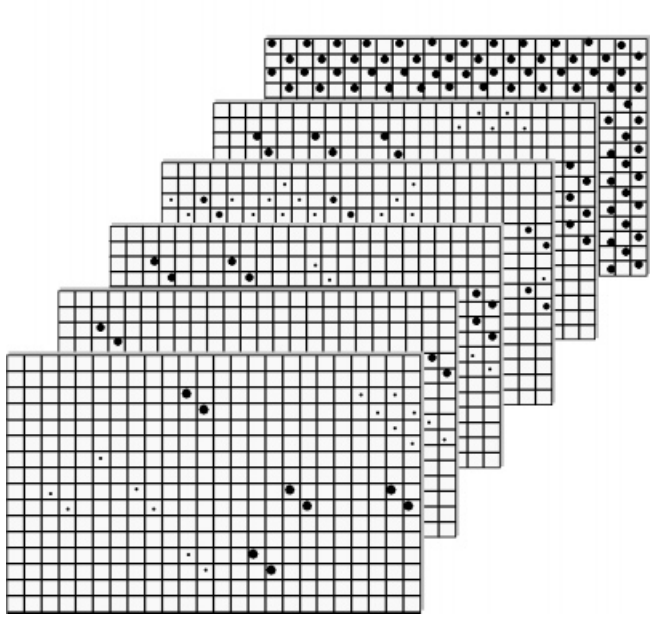

Figure 10 Representative probe hybridizations from a de novo SBH experiment. The same array is hybridized with a mass probe (GCCAGCTTTC, top image) to measure the amount of DNA in all dots, and five specific 8-mer probes. Three test samples are arrayed four times in diagonal pairs. The remaining dots (up to 96) are control samples. Different probes give different combinations of positive and negative results with test samples. Probe GGGCCCAA (second from top) is positive in all samples, and probe ATAAAGAGT (the front or bottom image) is positive only in the third sample. The other three probes in top to bottom order are: CGTGGCCT, GCTGGCCT, TGAGTGTC. The data show clear distinction between positive and negative samples; in most cases the negative samples correspond to a single-mismatch target for the given probe.

DNA filters were first hybridized with a ''mass probe'' specific to the M13 vector, to determine the amount of DNA present in each spot. Each filter was then successively hybridized to a series of test probes. The average discrimination value (defined as the average ratio of signals from hybridization with matched and mismatched targets) obtained for probes in this study was over 10. Figure 10 clearly shows the strong signal intensity and hybridization discrimination obtained from probe hybridizations in this study. The DNA sequences of the three unknown targets were precisely determined after hybridization with only 156 of the 272 probes.

Of the 468 probe hybridization scores obtained in this study, 22 positive signals were observed that were inconsistent with the final reconstructed sequences. In 11 of these cases, the positive signals resulted from full-match hybridizations with the M13 vector. Most of the remainder (seven cases) occurred because of the summation of

signals caused by two or more single-mismatch hybrids. In four cases, positive signals resulted from hybridization with a single mismatch target.

Both SBH and traditional gel methods gave identical results for all target DNAs sequenced in this study. Although the number of probes used was greatly reduced relative to the set of all possible octamers, the similarity of the DNA targets and the many single-base mismatches posed technical challenges comparable to those found in a much larger sequencing project.

9.2 Gene Discovery and Gene Expression

Oligonucleotide hybridization has great potential as an investigative tool in gene discovery studies. The ability to simultaneously process thousands of clones in parallel allows researchers to screen large gene libraries in search of rare or tissue-specific genes.

Gene discovery efforts are based on the concept of identifying gene-specific ''sequence signatures'' for each gene using the hybridization SP (see Figure 1). In gene discovery projects, thousands of cDNAs are spotted to a membrane and then hybridized sequentially to individual probes. Because each n-mer probe binds (on average) once in 4^n bases of target DNA, the set of probes that bind a given cDNA usually provides a unique, sequence-defining ''signature'' for that gene. Identical or closely related cDNAs can be ''clustered'' based on their similar signatures.

In an important early demonstration of gene discovery methods, Drmanac et al.[25] screened 73 536 cDNA clones derived from infant brain libraries. Using 200 to 320 oligomer probes (mainly 7-mers) to hybridize these clones, they identified 19 726 distinct genes and estimated that up to 20 000 additional genes are expressed in the infant brain. Clones from both ordinary and normalized gene libraries were used in the infant brain study. cDNA inserts were amplified directly from bacterial cultures in 96-well plates and then pipetted into 864-well plates. Samples from thirty-six 864-well plates were spotted onto a single 16 cm × 24 cm filter using a Beckman Biomek 1000® robot equipped with an 864-pin tool. Negative control spots (without DNA) were arrayed on each membrane to determine background binding. Selected clones of known sequence were used as positive controls.

The probes selected for the infant brain library study were chosen because they occur frequently in coding regions or otherwise helped distinguish unrelated clones. Probes were ^{33}P labeled and hybridized under short-wash, low-temperature conditions designed to generate a wide range of signal intensities and to allow maximum discrimination of unrelated clones. Signal intensities were measured for each clone–probe pair using a Molecular Dynamics PhosphorImager, and a raw intensity value calculated using a specialized analysis program.

A number of cDNAs were rejected from the data analysis due to insufficient DNA or problems with spotting. Using minimal acceptable standards for amount of DNA spotted and number of successful probe hybridizations per clone, a set of 57 419 cDNA clones was selected for further analysis. Of these, 38 900 were obtained from the ordinary library and 18 519 from the normalized library.

For each clone selected, the hybridization intensities of all probes tested against it are ranked from 1 to 100 and these rankings are used to calculate similarity values (SIMs) with other clones. The SIM for a pair of clones is defined as follows:

$$\mathrm{SIM} = \frac{\sum (S_i/L_i)^2}{N} \quad \text{for } L_i \geq T$$

where S_i and L_i are the smaller and the larger of the two rank scores of the ith probe for a pair of clones, T is a threshold used to eliminate probes that rank low with both clones, and N is the number of probes actually used in the pair-wise comparison. Using statistical considerations, a threshold SIM value of 0.57 was calculated to be the optimal level to distinguish clones within a shared cluster from those that are dissimilar. Observed SIM values range from an average of 0.34 for random, unrelated cDNA clones to 0.71 for cDNAs derived from different PCR preparations of a single clone. Sequences that differ randomly at more than 7% of base positions are likely to result in SIM values less than 0.57 and hence would group in different clusters.

Using the clustering criteria described above, the researchers discovered a number of interesting patterns in the ordinary and normalized gene libraries. As expected, the ordinary library contained much larger clusters than the normalized library. The three largest clusters from the ordinary library contained an average of 339 clones each, while the three largest clusters in the normalized library had an average of 56 members. Using various statistical criteria, the researchers were able to determine that most two-clone clusters represent real groupings rather than random associations. However, false negatives (failure to associate two related clones) were found to occur in about 5% of cases.

ESTs of several clones were compared with GenBank data to determine the likelihood that cluster pairings were inaccurate. In general it was found that clones from a single cluster whose sequences matched GenBank data were either identical or very similar to one another. Even in cases where ESTs from two clones within a single cluster seemed dissimilar the most plausible explanation is that they simply represented nonoverlapping regions of different length clones of a single gene.

The researchers also used a clustering process to examine the expression pattern of mRNAs within the

ordinary brain gene library. Only a few of the 14 010 genes that were expressed in the ordinary (non-normalized) gene library were shown to be responsible for most of the expressed mRNAs within the infant brain. Over 50% of all mRNAs were representatives of the largest 10.8% (1519) of clusters. Nearly 72% of infant brain genes were represented by only a single clone cluster. The average cluster contained 2.4 clones.

In the normalized library, clone abundances of the most commonly expressed genes were typically reduced by 10- to 50-fold relative to the ordinary library. Some genes, however, showed increased expression in the normalized library. Overall, the normalization process reduced the number of clones from the most common gene classes from 29.6% to 5.1% of the total.

One of the most valuable aspects of large-scale fingerprinting projects such as this one is the acquisition of large data sets that allow statistical inference about the population from which they are obtained. Using the number of observed one-, two- and three-clone clusters and assuming an expected binomial distribution, it is possible to predict that a total of at least 40 000 genes are expressed in the infant brain. The current study discovered roughly 19 726 of these, but a much larger clone set would be necessary to capture the complete set of expressed genes. It is predicted that up to 840 000 clones would be necessary to find 99.9% of the genes expressed.

The major advantages of using oligonucleotide hybridization in large gene discovery projects such as the infant brain library study are its ability to:

1. identify and assess genes and cDNA transcript variants that are expressed at low levels;
2. measure differential gene expression between various tissue or cell lines for all genes, without previous knowledge about gene sequence, expression, etc.

The use of very large clone sets greatly enhances the ability to detect rarely expressed cDNAs (due to increased sample sizes), and also improves clustering precision for alternatively spliced messenger variants. Given the low expression of many genes within the human genome and the expected continuum of similar genes within gene families, oligonucleotide technology is likely to be an enormously valuable tool in gene discovery and expression studies.

9.3 Individual Sequencing to Discover Novel Polymorphisms and Mutations

As disease prevention and personalized treatments such as gene therapy become increasingly important medical procedures, there is a growing need for methods that can detect mutations in patient DNA samples. Complete gene sequencing is the best approach to ensure that no mutation will be missed, but gel sequencing methods are slow and costly. On the other hand, SBH is an efficient, cost-effective method of screening large numbers of medically important DNA samples by sequencing for such individual variations.

Drmanac et al.[34] demonstrated the accuracy and speed of using SBH to detect individual mutations by sequencing the medically important p53 gene. In a blind experiment, they used 7-mer probe sets to accurately sequence 12 p53 samples that had a variety of individual mutations, including insertions, deletions, and substitutions. Test samples were generated by in situ mutagenesis of a single p53 clone, with a target mutation rate of one base change per exon. Exons 5–8 and the flanking intron regions were then amplified by PCR to obtain a high concentration of target DNA relative to background genomic DNA.

Amplified DNA was spotted directly to membranes without any further purification, a protocol that could be applied to any DNA sample. Samples were spotted in 8×8 replica arrays (each 6 mm $\times$ 6 mm) on GeneScreen membranes. Eight such membranes were prepared, containing a total of 3072 replica arrays. A metal grid matching the membrane array pattern was prepared, containing 384 isolated hybridization chambers.

The researchers used two probe sets in the p53 study. One set consisted of all 2186 7-mer probes that complement the reference p53 exon amplicons. The second set of 8192 probes consisted of all 7-mers that contain at least four G + T bases. The G – T-rich probes were chosen because empirical data have shown that they hybridize more efficiently than C – A-rich probes. Using only the G – T-rich probes does not result in loss of sequence information because each probe will bind either one DNA strand or its complement. Probes were ^{33}P-labeled and hybridized to DNA arrays under low-temperature short-wash conditions.

After hybridization and washing, images were scanned using a PhosphorImager scanner (Molecular Dynamics, Sunnyvale, CA) and then analyzed using a special software package that assigns a signal intensity score for each DNA dot. Figure 11(a) shows hybridization images for the target DNA samples analyzed in this experiment. Negative control spots were used to measure background signal intensity. The signal intensities for probes positive to the control (normal) samples were used to quantify the amount of DNA contained in each spot. The adjusted signal intensities were then used to form a sample/control (S/C) ratio for each base consisting of the median S/C values of the 7 or 14 probes that cover that base in sequence. The S/C value quantifies the relative binding strengths of probes for the test and control DNAs at that base position.

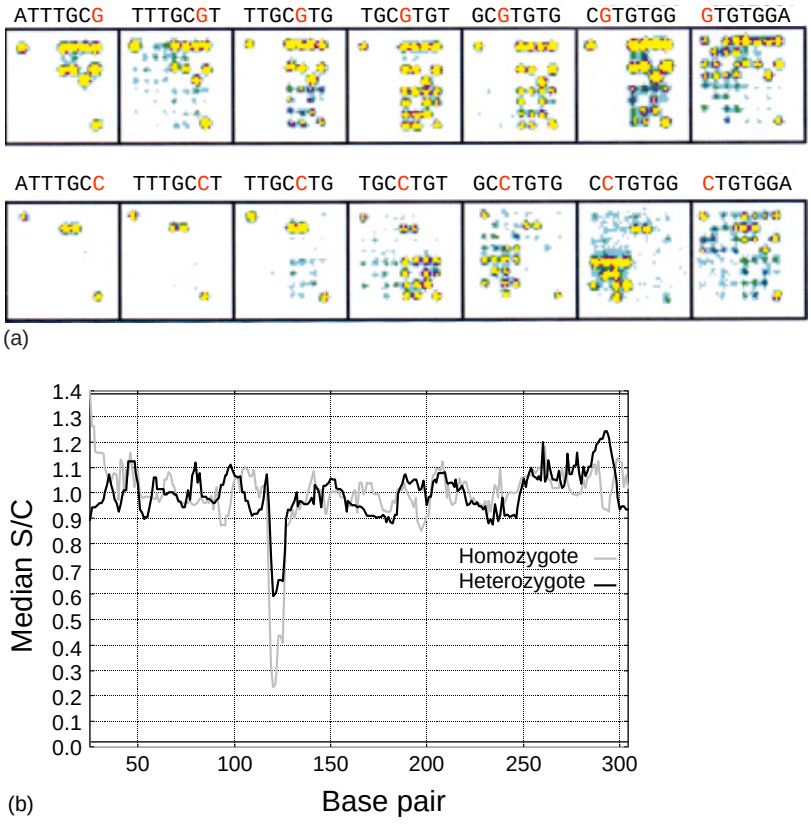

Figure 11 Mutation discovery. (a) Probe hybridization images used to detect G-to-C substitution in exon 6 of p53 gene mutants. Results of hybridization experiments involving two sets of seven probes covering this site are shown. Top panel: probes complementary to wild type sequence with base G; bottom panel: probes complementary to mutant sequence containing base C. DNA samples corresponding to exons 5, 6, 7, and 8 are spotted in an 8×8 square array. Exon 6 DNA dots are located in the top right quadrant of each square, with normal (wild type) reference DNA in the first row and mutant DNA in the second row. Top panel probes, positive to the reference DNA, show strong hybridization signals with dots 5 through 8 of the first row and negative signals for the mutant sample in dots 5 and 6 of the second row. Bottom panel (mutant) probes show exactly the reverse pattern. The other strong signals visible in various images are two corner markers, additional exon 6 test samples with no mutation, or the remaining three exons. (b) Mutation scanning of p53 gene samples. Median signal ratios of test and control DNA (S/C) for reference positive probes are plotted against base position for two exon 5 samples. For more details see text. Mutated samples contain an insertion of a C at position 122.

The sequencing of gene variants is a two-step process that involves plotting S/C ratios for both probe sets. A plot of S/C values for the probe set positive to the control DNA is called a ''mutation scanning graph'' and is used to detect mutations in test DNA samples. When test and control DNAs are identical in a given region, S/C ratios are expected to be unity. When a mutation occurs in a test DNA, the S/C ratio for the mutation scanning graph tends to drop dramatically due to failure of wild-type probes to hybridize the test DNA.

Figure 11(b) shows a typical mutation scanning graph for a p53 test sample mutated at base position 122 in exon 5. The two curves shown reflect homozygote and heterozygote mutations in test DNAs. When the mutation occurs in both chromosomes (homozygote) S/C ratios tend to drop to less than 0.5; when a mutation occurs in only one chromosome (heterozygote) S/C ratios typically drop to between 0.5 and 0.8. The clear ''opening'' in the curve at base position 122 confirms the occurrence of a mutation. Once a mutation has been detected, a sequence determination graph of median S/C ratios from the larger probe set is used to determine the mutated sequence. In this case, probes specific to the mutated test DNA but not the control provide high S/C signals.

A total of 13.2 kb of p53 DNA (12 samples $\times$ 1.1 kb each) were sequenced with 100% accuracy in this experiment. The most difficult sequence assignments are in regions that contain single-base repeats of six or more nucleotides. When a mutation occurs in one of these regions, only one or two of the seven overlapped probes are not shared by the two sequence variants. In one such case, a small number of longer (9-mer) probes were used to confirm the sequence. In longer direct repeat regions

Table 3 Types of mutations detected in p53 gene in blind SBH tests

Exon[a]	Position[b]	Sequence change[c]	Hetero[c]
e5	13 055	ctacagt(iT)tacycc	−
e5	13 106	gccctgt(iC)tgcagc	−
e5	13 157	cgcgccA to Gtggcca	−
e6	13 365	atttgcG to Ctgtgga	−
s6	13 389	acagaaA to Gcacttt	−
e6	13 395	acacttd(dTTCG)acatag	−
e6	13 507	cagtagG to Aggggct	+
e6	13 409	agtgtgG to Atggtgc	−
e6	14 365	atttgcG to Ctgtgga	+
e7	14 012	tctgacT to Cgtacca	−
e7	14 052	gttcctG to Ccatggg	+
e7	14 080	gcccatC to Tcycacc	+
e8	14 463	atctacT to Agggacg	+
e8	14 523	gcacagA to Gggaaga	−

[a] Amplicons corresponding to each of four exons are tested separately.
[b] The base position of the mutation in the p53 gene sequence is indicated using accession number X54156 from Gene Bank as a reference.
[c] The exact type of mutation and its heterozygote or homozygote status are also indicated, with "i" indicating an insertion and "d" a deletion.

it may be preferable to use gel sequencing to confirm the final sequence.

An important advantage of SBH is that multiple, independent reads of both DNA strands help ensure accurate, reproducible sequence determination. Five replicate samples of one p53 gene were amplified and sequenced independently to confirm the reproducibility of the SBH process between runs. The coefficient of variation for median S/C values at mutant sites was below 3% for all cases, which is considered well within standards for most clinical diagnostic studies. Overall, the standard deviation of median S/C values in this study was about 6%. It is estimated that such random variation will result in a median S/C value that falls below the 0.8 threshold roughly once in 2500 bases. In the few cases where such false positives occurred the absence of a corresponding peak in the sequence determination graph was used to rule out the possibility of a real sequence change. In conclusion, the SBH method provided 100% accurate sequence determination for over 13 kb of DNA containing a variety of substitution, deletion, and insertion mutants (Table 3).

10 CONCLUSION

The precise instructions for an organism's development and function are recorded and reproduced in the sequences of genomic DNA or RNA molecules through a self-encoding evolutionary process. This information is expressed in three distinct ways: directly as three-dimensional nucleic acid structures (for example, promoter regions that allow specific interactions with transcription factors) or, by transcription and translation, as RNA or protein molecules. Complex networks of interactions between these three basic types of molecules define spatial/temporal and quantitative aspects of self-control necessary for the organism's development or for responses to environmental changes. To be able to understand the enormous number of interactions that exist in these networks, we need to collect and analyze extremely large sequence data sets. We have argued, for example, that sequencing every human gene in about 1000 individuals and defining about 1 million functional DNA polymorphisms is of critical importance. We believe that SBH will become the technology of choice in such projects.

SBH is a versatile, cost-effective method for obtaining the accurate DNA sequence information that is needed to fully understand the implications of genetic variation. The oligonucleotide hybridization and SBH methods use either complete or partial sets of DNA oligomers to identify complementary sequences contained within a target DNA. A small set of oligomer probes may be used to identify an unknown DNA by obtaining its "fingerprint" or "sequence signature," greatly reducing the workload and cost of large-scale sequencing projects by minimizing the number of redundant clones that are sequenced. Larger probe sets may be used to completely sequence a target gene or to detect all mutations relative to a known reference DNA. The multiple overlapping nature of probe sets used in sequencing studies provides high data redundancy and helps ensure accurate sequence information. The use of robotics and microarray technologies allows the processing of hundreds of thousands of probe–DNA hybridization results per day. In its gene discovery program, Hyseq, Inc., routinely fingerprints over a million cDNA clones per month using a single production line. In the more than 10 years since its conception, SBH has become a critical research method, with applications in almost all aspects of genomics. With the explosive growth of genomics research, medical diagnostics, and pharmacogenomics, the coming decades promise even greater developments in SBH technology and its applications.

ACKNOWLEDGMENTS

We thank Radomir Crkvenjakov for analyses, discussions, and critical readings of the manuscript, and Elizabeth Garnett for her excellent organizational assistance throughout this project and for preparation of tables and figures.

ABBREVIATIONS AND ACRONYMS

A	Adenine
BAC	Bacterial Artificial Chromosome
C	Cytosine
CCD	Charge-coupled Device
EST	Expressed Sequence Tag
G	Guanine
LIMS	Laboratory Information Management System
OP	Overlap Principle
PCR	Polymerase Chain Reaction
QC	Quality Control
SBH	Sequencing by Hybridization
S/C	Sample/Control
SIM	Similarity Value
SNP	Single Nucleotide Polymorphism
SP	Search Principle
T	Thymine

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Sequencing Strategies and Tactics in DNA and RNA Analysis

Peptides and Proteins **(Volume 7)**
Protein–Oligonucleotide Interactions

REFERENCES

1. F. Sanger, S. Nicklen, A. Coulson, 'DNA Sequencing with Chain-terminating Inhibitors', *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467 (1977).
2. A.M. Maxam, W. Gilbert, 'A New Method for Sequencing DNA', *Proc. Natl. Acad. Sci. USA*, **74**, 560–564 (1977).
3. R.W. Holey, J. Apgar, G.A. Everett, J.T. Madison, M. Marquisee, S.H. Merrill, J.R. Penswick, A. Zamir, 'Structure of a Ribonucleic Acid', *Science*, **147**, 1462–1465 (1965).
4. T. Hunkapiller, R.J. Kaiser, B.F. Koop, L. Hood, 'Large-scale and Automated DNA Sequence Determination', *Science*, **254**, 59–63 (1991).
5. P. Doty, J. Marmur, J. Eigen, C.E. Schildkraut, 'Strand Separation and Specific Recombination in Deoxyribonucleic Acids: Physical Chemical Studies', *Proc. Natl. Acad. Sci. USA*, **46**, 461–466 (1960).
6. J.L. Beaucage, M.H. Caruthers, *Tetrahedron Lett.*, **22**, 1859–1862 (1981).
7. R.B. Wallace, J. Shaffer, R.E. Murphy, J. Bonner, T. Hirose, K. Itakura, 'Hybridization of Synthetic Oligodeoxyribonucleotides to X 174 DNA', *Nucleic Acids Res.*, **6**, 3543–3557 (1979).
8. R. Saiki, P.S. Walsh, C.H. Levenson, H.A. Ehrlich, 'Genetic Analysis of Amplified DNA with Immobilized Sequence-specific Oligonucleotide Probes', *Proc. Natl. Acad. Sci. USA*, **86**, 6230–6234 (1989).
9. A. Poustka, H. Lehrach, 'Jumping Libraries and Linking Libraries: The Next Generation of Molecular Tools in Mammalian Genetics', *Trends Genet.*, **2**, 174–179 (1986).
10. R. Drmanac, R. Crkvenjakov, Yugoslav Patent Application 570/87, 1987; issued as R. Drmanac, R. Crkvenjakov, 'Method of Sequencing of Genomes by Hybridization with Oligonucleotide Probes', US Patent 5 202 231, 1993.
11. R. Drmanac, I. Labat, I. Brukner, R. Crkvenjakov, 'Sequencing of Megabase Plus DNA by Hybridization: Theory of the Method', *Genomics*, **4**, 114–128 (1989).
12. W. Bains, G.C. Smith, 'A Novel Method for Nucleic Acid Sequencing', *J. Theor. Biol.*, **135**, 303–307 (1988).
13. E. Southern, 'Analyzing Polynucleotide Sequences', International Patent Application PCT GB 89/00460, 1988.
14. Y.P. Lysov, V.L. Florentiev, A.A. Khorlyn, K.R. Khrapko, V.V. Shick, A.D. Mirzabekov, 'Determination of the Nucleotide Sequence of DNA Using Hybridization with Oligonucleotides: A New Method', *Dokl. Akad. Nauk. SSSR*, **303**, 1508–1511 (1988).
15. P.A. Pevzner, R.J. Lipshutz, 'Towards DNA Sequencing Chips', in *Mathematical Foundations of Computer Science 1994*, eds. I. Privara, B. Rovan, P. Ruzicka, The Proceedings of 19th International Symposium, MFCS '94, Kosice, Slovakia, Springer-Verlag, Berlin, 143–158, 1995.
16. K.R. Khrapko, Y.P. Lysov, A.A. Khorlyn, V.V. Shick, V.L. Florentiev, A.D. Mirzabekov, 'An Oligonucleotide Hybridization Approach to DNA Sequencing', *FEBS Lett.*, **256**, 118–122 (1989).
17. R. Drmanac, R. Crkvenjakov, 'Sequencing by Hybridization (SBH) with Oligonucleotide Probes as an Integral Approach for the Analysis of Complex Genomes', *Int. J. Genome Res.*, **1**(1), 59–79 (1992).
18. N.E. Broude, T. Sano, C.L. Smith, C.R. Cantor, 'Enhanced DNA Sequencing by Hybridization', *Proc. Natl. Acad. Sci. USA*, **91**, 3072–3076 (1994).
19. R. Drmanac, US Patent Application PCT/US94/10945, 1993.
20. T.T. Nikiforov, R.B. Rendle, P. Goelet, Y.H. Rogers, M.L. Kotewicz, S. Anderson, G.L. Trainor, M.R. Knapp, 'Genetic Bit Analysis: A Solid Phase Method for Typing Single Nucleotide Polymorphisms', *Nucleic Acids Res.*, **22**, 4167–4175 (1994).
21. J.D. Hoheisel, E. Meier, R. Mott, L. McCarthy, A.V. Grigorov, L.A. Scalkwyk, D. Nizetic, F. Francis, H. Lehrach, 'High Resolution Cosmid and P1 Maps Spanning the

14 Mb Genome of the Fission Yeast *S. pombe*', *Cell*, **73**, 109–120 (1993).
22. I. Labat, R. Drmanac, 'Simulations of Ordering and Sequence Reconstruction of Random DNA Clones Hybridized with a Small Number of Oligomeric Probes', in *The Second International Conference on Electrophoresis, Supercomputing and the Human Genome*, World Scientific, Singapore, 555–565, 1992.
23. P. Scholler, A.E. Karger, S. Meir-Ewert, H. Lehrach, H. Delius, J.D. Hoheisel, 'Fine-mapping of Shotgun Template-libraries: An Efficient Strategy for Systematic Sequencing of Genomic DNA', *Nucleic Acids Res.*, **23**, 3842–3849 (1995).
24. R. Drmanac, G. Lennon, S. Drmanac, I. Labat, R. Crkvenjakov, H. Lehrach, 'Partial Sequencing by Hybridization: Concept and Applications in Genome Analysis', in *The First International Conference on Electrophoresis, Supercomputing and the Human Genome*, World Scientific, Singapore, 60–74, 1991.
25. S. Drmanac, N.A. Stravropoulos, I. Labat, J. Vonau, B. Hauser, M.B. Soares, R. Dramanac, 'Gene-representing cDNA Clusters Defined by Hybridization of 57 419 Clones from Infant Brain Libraries with Short Oligonucleotide Probes', *Genomics*, **37**, 29–40 (1996).
26. A. Milosavljevic, M. Zeremski, Z. Strezoska, D. Grujic, H. Dyanov, S. Batus, D. Salbego, T. Paunesku, M.B. Soares, R. Crkvenjakov, 'Discovering Distinct Genes Represented in 29 570 Clones from Infant Brain cDNA Libraries by Applying Sequencing by Hybridization Methodology', *Genome Res.*, **6**, 132–141 (1996).
27. E.S. Meier, J. Lange, H. Gerst, R. Herwig, A. Schmitt, J. Freund, T. Elge, R. Moss, B. Herrmann, H. Lehrach, 'Comprehensive Gene Profiling by Oligonucleotide Fingerprinting', *Nucleic Acids Res.*, **26**, 2216–2223 (1998).
28. A. Milosavljevic, S. Savkovic, R. Crkvenjakov, D. Salbego, H. Serrato, H. Kreutzer, A. Gemmell, S. Batus, D. Grujic, S. Carnahan, T. Paunesku, J. Tepavcevic, 'DNA Sequence Recognition by Hybridization to Short Oligomers: Experimental Verification of the Method on the *E. coli* Genome', *Genomics*, **37**, 77–86 (1996).
29. D.J. Lockhart, H. Dong, M.T. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, E.L. Brown, 'Expression Monitoring by Hybridization to High-density Oligonucleotide Arrays', *Nat. Biotechnol.*, **14**, 1675–1680 (1996).
30. J. Cheng, E.L. Sheldon, L. Wu, A. Uribe, L.O. Gerrue, J. Carrino, M. Heller, J. O'Connell, 'Electric Field Controlled Preparation and Hybridization Analysis of DNA/RNA from *E. coli* on Microfabricated Bioelectronic Chips', *Nature Biotechnol.*, **16**, 541–546 (1998).
31. D.G. Wang, J.B. Fan, C. Siao, A. Berno, P. Yang, R. Sapolsky, G. Ghanour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglu, E. Hubell, E. Robinson, M. Mittman, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T. Hudson, R. Lipshutz, M. Chee, E. Lander, 'Large Scale Identification, Mapping and Genotyping of Single-nucleotide Polymorphisms in the Human Genome', *Science*, **280**, 1077–1082 (1998).
32. R. Drmanac, R. Crkvenjakov, 'Prospects for a Simplified and Frugal Human Genome Project', *Scientia Yugoslav.*, **16**, 99–107 (1990).
33. Z. Strezoska, T. Paunesku, D. Radosavljevic, I. Labat, R. Drmanac, R. Crkvenjakov, 'DNA Sequencing by Hybridization: 100 Bases Read by a Non-gel-based Method', *Proc. Natl. Acad. Sci. USA*, **88**, 10089–10093 (1991).
34. S. Drmanac, D. Kita, I. Labat, B. Hauser, J. Burczak, R. Drmanac, 'Accurate Sequencing by Hybridization for DNA Diagnostics and Individual Genomics', *Nature Biotechnol.*, **16**, 54–58 (1998).
35. E.M. Southern, U. Maskos, J.K. Elder, 'Analyzing and Comparing Nucleic Acid Sequences by Hybridization to Arrays of Oligonucleotides: Evaluation Using Experimental Models', *Genomics*, **13**, 1008–1017 (1992).
36. M. Chee, R. Yang, E. Hubell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, S.P.A. Fodor, 'Accessing Genetic Information with High-density DNA Arrays', *Science*, **274**, 610–614 (1996).
37. M. Kozal, N. Shah, N. Shen, R. Fucini, R. Yang, T. Merigan, D.D. Richman, M.S. Morris, E. Hubell, M. Chee, T.R. Gingeras, 'Polymorphism Observed in HIV-1 Clade B Protease Genes Using High Density Oligonucleotide Arrays: Implications for Therapy', *Nature Medicine*, **7**, 753–759 (1996).
38. J.G. Hacia, L.C. Brody, M.S. Chee, S.P.A. Fodor, F.S. Collins, 'Detection of Heterozygous Mutations in BRCA 1 Gene Using High Density Oligonucleotide Arrays and Two-color Fluorescence Analysis', *Nature Genetics*, **14**, 441–447 (1996).
39. J.G. Hacia, B. Sun, N. Hunt, K. Edgemon, D. Mosbrook, C. Robbins, S.P.A. Fodor, D. Tagle, F.S. Collins, 'Strategies for Mutational Analysis of the Large Multiexon ATM Gene Using High Density Oligonucleotide Arrays', *Genome Res.*, **8**, 1245–1258 (1998).
40. K.L. Gunderson, X.C. Huang, M.S. Morris, R.J. Lipshutz, D.J. Lockhart, M.S. Chee, 'Mutation Detection by Ligation to Complete *N*-mer DNA Arrays', *Genome Res.*, **8**, 1142–1153 (1998).
41. R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, W.K. Funkhouser, B. Koop, L. Hood, R. Crkvenjakov, 'DNA Sequence Determination by Hybridization: A Strategy for Efficient Large-scale Sequencing', *Science*, **260**, 1649–1652 (1993).
42. R. Drmanac, S. Drmanac, I. Labat, R. Crkvenjakov, A. Vicentic, A. Gammell, 'Sequencing by Hybridization: Towards an Automated Sequencing of One Million M 13 Clones Arrayed on Membranes', *Electrophoresis*, **13**, 566–573 (1992).
43. R. Drmanac, S. Drmanac, 'cDNA Screening by Array Hybridization', *Methods Enzymol.*, **303**, 165–178 (1999).

44. J.G. Wetmur, 'DNA Probes: Applications of the Principles of Nucleic Acid Hybridization', *Crit. Rev. Biochem. Mol. Biol.*, **26**, 227–259 (1991).
45. K.J. Breslauer, R. Frank, H. Blocker, L.A. Marky, 'Predicting DNA Duplex Stability from the Base Sequence', *Proc. Natl. Acad. Sci. USA*, **83**, 3746–3750 (1986).
46. N. Sugimoto, S. Nakano, M. Yoreyama, K. Hanada, 'Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes', *Nucleic Acids Res.*, **24**, 4501–4505 (1996).
47. F. Michiels, A.G. Craig, G. Zehetner, G.P. Smith, H. Lehrach, 'Molecular Approaches to Genome Analysis: A Strategy for the Construction of Ordered Overlapping Clone Libraries', *CABIOS*, **3**, 203–210 (1987).
48. R. Drmanac, Z. Strezoska, I. Labat, S. Drmanac, R. Crkvenjakov, 'Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides', *DNA Cell Biol.*, **9**(7), 527–534 (1990).
49. J.N. Housby, E.M. Southern, 'Fidelity of DNA Ligation: A Novel Experimental Approach Based on Polymerization of Libraries of Oligonucleotides', *Nucleic Acids Res.*, **26**, 4259–4266 (1998).
50. I. Dianzani, C. Camaschella, G. Saglio, S.M. Forrest, S. Ramus, R.G. Cotton, 'Simultaneous Screening for beta-Thalassemia Mutations by Chemical Cleavage of Mismatch', *Genomics*, **11**, 48–53 (1991).

Sequencing Strategies and Tactics in DNA and RNA Analysis

Tatyana L. Azhikina
Russian Academy of Sciences, Moscow, Russia

1 Introduction **1**
2 Brief History of Nucleic Acid Sequencing **2**
3 Sequencing Principles and Technological Developments **2**
3.1 Methods Based on Hydrolysis of DNA (RNA) Molecules 2
3.2 Method Based on Specific Termination of Primer Extension 3
3.3 Delimiters Approach 3
4 Long-range Genomic DNA Sequencing Strategy **5**
4.1 Random (Shotgun) Strategy 5
4.2 Directed Sequencing 6
4.3 Combination of Random and Directed Sequencing as the Basic Strategy of the Human Genome Project 7
5 Short DNA/RNA Sequencing Approaches **8**
5.1 Complementary DNA Sequencing 8
5.2 RNA Sequencing 9
5.3 New Approaches 10
6 Comparative Sequencing **12**
6.1 DNA Chips 12
6.2 Bioinformatics 12
Acknowledgments **12**
Abbreviations and Acronyms **12**
Related Articles **13**
References **13**

Sequencing procedures are aimed at the determination of sequences of monomer units constituting linear biopolymer molecules. In the particular case of nucleic acids the monomer units are DNA or RNA residues. The present article reviews several sequencing strategies: different approaches (random or systematic sequencing) as well as their combinations giving the best results for long-range sequencing applicable to long and unique DNA fragments, including entire genomes, up to the Homo sapiens *genome size; massive sequencing of relatively short fragments such as RNAs, complementary DNAs (cDNAs), etc.; and comparative sequencing of related DNA fragments originating from various individuals or different species.*

Although particular sequencing approaches may involve different ways of DNA treatment (chemical, enzymatic, or physical), they are all based on the same idea: a labeled molecule is cut to obtain a set of fragments all having one common end, other ends being different owing to the statistical character of the nucleic acid chain cleavage. The fragments obtained are then separated according to their lengths.

The primary structure of nucleic acids is a key not only to physical and chemical properties of the molecules but also to the biological (genetic) information therein, including structures of genes and proteins, the ways of gene regulation and cell response to different signals. This is a great advantage over other physical, chemical, or biological methods of DNA/RNA analysis. However, sequencing in itself is usually insufficient to predict unambiguously the biological properties and functions of the encoded products.

1 INTRODUCTION

The organism stores its genetic information in linear molecules of the genomic DNA and this information can be read using DNA sequencing techniques. The primary structure of any natural DNA molecule harbors not only sequences of proteins and RNAs encoded by genes but also the information on gene regulation. Owing to recent advances in sequencing and cloning techniques, the nucleotide sequences of genes often become available long before the functions of these genes in the organism are established. The possibility to analyze entire genomes of various organisms greatly facilitates the finding and identification of new genes and promotes biological studies.

The Human Genome Project (HGP) was officially started in 1990 to map and sequence the whole human genome and the chromosomes of several model organisms. At present the genomes of several microorganisms and bacteria, as well as some model genomes, are already completely deciphered. HGP is a unique project that is supposed to revolutionize biology and medicine. The high-quality sequence of the human genome will have an unprecedented impact and will always be of great value for basic biology, biomedical research, biotechnology,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and health care. A new era in biology – sequence-based biology – will stimulate the progress in understanding gene–environmental interactions and in the development of highly accurate DNA-based medical diagnostics and therapeutics.

2 BRIEF HISTORY OF NUCLEIC ACID SEQUENCING

The first nucleotide sequence (alanine transport RNA of yeast) was determined in 1965 by Holley et al.[1] using an approach initially developed for proteins and successfully used for insulin sequencing.[2] Improvements of the approach made it possible to sequence RNA genomes, and the first RNA genome of MS-2 phage was deciphered in 1975. The development of DNA sequencing progressed substantially more slowly because DNA molecules are much larger, and at that time no deoxyribonucleases (DNAses) with single-base specificity suitable for enzymatic hydrolysis were known. Study of possible chemical modifications of heterocyclic bases resulted in the chemical cleavage method of Maxam and Gilbert (1977).[3] But only the discovery of restriction endonucleases capable of cutting DNA at specific sites, as well as DNA- and RNA-dependent polymerases, has made a revolutionary breakthrough in DNA sequencing. The development of the primer-extension idea reviewed in detail by Wu[4] has led to the dideoxynucleotide chain-termination method by Sanger et al.[5] who shared the Nobel prize with Gilbert for sequencing.

Bacteriophage ϕX 174 (5386 bp) was the first sequenced by Sanger et al.[6] After this, a number of whole genomes of several viruses and organelles were determined: among them the genome of bacteriophage λ,[7] the 229-kb genome of cytomegalovirus,[8] the 192-kb genome of vaccinia,[9] and others. The whole-genome random sequencing strategy was proposed by Venter et al.[10] and used for sequencing of some microbial genomes. The full-length genomes of *Saccharomyces cerevisiae* and *Escherichia coli* were reported in 1996 and 1997, respectively. Currently, active genome sequencing projects are being implemented on insects and more complicated organisms such as *Drosophila melanogaster* (9% done by the end of 1998), *Caenorhabditis elegans* (80% completed by the end of 1998), mouse (10 000 sequence-tagged sites (STSs) mapped), and *Homo sapiens*.[11]

3 SEQUENCING PRINCIPLES AND TECHNOLOGICAL DEVELOPMENTS

All modern sequencing approaches are based on one and the same idea: labeled DNA is treated chemically or enzymatically to obtain a set of DNA fragments all having one common (identical) end, other ends being different owing to the statistical character of the treatment. The fragments obtained can be separated according to their lengths. The methods of DNA treatment are described below.

3.1 Methods Based on Hydrolysis of DNA (RNA) Molecules

3.1.1 Chemical Modifications Followed by Hydrolysis

The idea proposed by Maxam and Gilbert[3] is illustrated in Figure 1. The DNA fragment is end-labeled and modified by a chemical agent specific for a certain

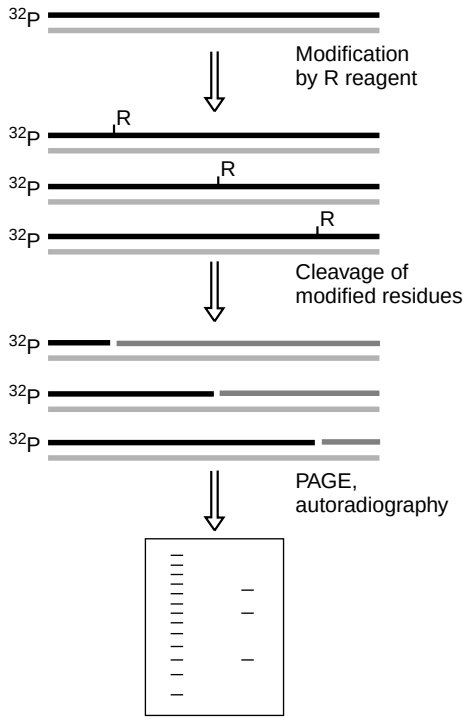

Figure 1 The principle of Maxam–Gilbert sequencing. PAGE = polyacrylamide gel electrophoresis.

Table 1 Chemical modifying agents

	Base			
	Guanine	Purine	Pyrimidine	Cytidine
DNA sequencing	Dimethyl sulfate	Formic acid	Hydrazine (low salt)	Hydrazine (high salt)
RNA sequencing	Dimethyl sulfate	Diethyl pyro-carbonate (only for adenines)	Hydrazine (low salt)	Hydrazine (high salt)

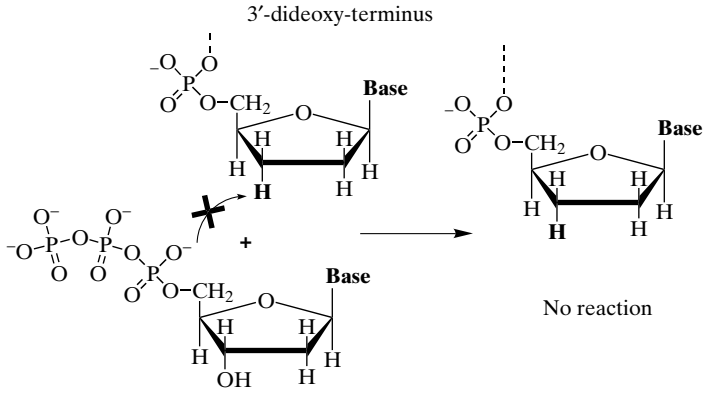

Scheme 1 Biochemistry of chain termination.

heterocyclic base or a definite type of base (listed in Table 1). The extent of modification is chosen so that finally an average of only one break is introduced in each polynucleotide chain. After this partial and specific modification, amine-catalyzed statistical degradation of the modified chain is performed, and the products of this degradation are separated in polyacrylamide gels. By using four different cleavage reactions, the entire sequence can be read. The Maxam–Gilbert procedure is complicated and difficult to automate, so its application to modern large-scale sequencing is limited. However, it can be used for sequencing of ''problem templates'', studies of DNA–protein interactions (''footprinting'') or DNA structures, and oligonucleotide sequencing. A similar approach was applied to RNA by Peattie.[12]

3.1.2 Digestion with Specific Ribonucleases: RNA Sequencing

Radioactively labeled fragments of RNA can be sequenced using partial splitting by specific ribonucleases (reviewed by Kuchino and Nishimura[13]). A comprehensive list of the ribonucleases used is given below (see section 5.2). The application of ribonucleases (RNAses) is restricted by their rather low specificity. Moreover, the RNA degradation rates differ depending on the type of heterocyclic bases.

3.2 Method Based on Specific Termination of Primer Extension

Another approach called the dideoxynucleotide method was proposed by Sanger et al. in 1977.[5] It is underlain by the ability of DNA polymerase to extend a primer annealed to the template in the presence of four deoxynucleoside triphosphates (dNTPs) – dATP, dCTP, dGTP, and dTTP – one of which is usually radioactively labeled to visualize a newly synthesized chain. The growing chain is terminated by a 2′,3′-dideoxynucleoside triphosphate (ddNTP) that blocks the running DNA synthesis (Scheme 1). In this way the synthesis from a unique primer in the presence of all four dNTPs and one ddNTP yields a population of molecules with common 5′-ends but different 3′-ends, depending on the site where the ddNTP was incorporated. Four separate reactions with different ddNTPs are performed, the size of each resulting fragment being determined by the sequence of the template.

The procedure was initially developed for the large fragment of *Escherichia coli* DNA polymerase I (Klenow fragment) and avian myeloblastosis virus reverse transcriptase.[14] Ideally, a sequencing enzyme should be characterized by high processivity and a high rate of nucleotide incorporation; it should also be free from exonuclease activity and should not discriminate against nucleotide analogs. Working in this direction, Tabor and Richardson[15] introduced a chemically modified derivative of phage T7 DNA polymerase (known also under the commercial name of Sequenase®), which met all these requirements. The method is very efficient, presently allowing up to 800 bp of the sequence to be read in one run. It became a routine procedure and, being easily robotized, it is now practically the only method used for large-scale sequencing (see below).

Wide use of polymerase chain reaction (PCR) amplification and new thermostable DNA polymerases allowed the PCR approach and dideoxy sequencing to be combined. These systems provide the option of using either end-labeled primers or primers with incorporated label (such as dNTP). The high temperature used for PCR helps to circumvent the problem of secondary structures and increases the stringency of primer hybridization. Moreover, the amount of template used for sequencing can be markedly reduced to as little as 4×10^{-15} mol.

3.3 Delimiters Approach

This approach combines the features of both the Sanger and Maxam–Gilbert procedures. It is based on the incorporation of so-called delimiters in the course of PCR

amplification that further determine the sites of chemical or enzymatic degradation. A version of this method using thiophosphate-modified nucleotides (2′-deoxynucleoside 5′-α-[P-thio]-triphosphates) as delimiters was proposed by Nakamaye et al.[16] PCR products in this version are cleaved chemically by 2-iodoethanol or 2,3-epoxy-1-propanol at the sites of thiophosphate incorporation. An essentially analogous but enzymatic approach[17] also uses thiophosphate delimiters incorporated into primer-extension products at specific sites subsequently digested by exonuclease III. This approach seemed promising but had a disadvantage of uneven band intensity due to chemical features of thiophosphates. Porter et al.[18] used boranophosphates (2′-deoxynucleoside 5′-α-[P-borano]-triphosphates) to perform both amplification and direct DNA sequencing of PCR products. The ability of boranophosphates to be incorporated into growing DNA chains during PCR amplification and to block exonuclease activity, combined with their thermostability, provides a tool for direct PCR sequencing, as demonstrated in Figure 2.

Technological developments, mostly aimed at improvement of the chain-termination Sanger sequencing, involve new ways of separating sequencing products, new DNA polymerases, and new nonradioactive labels. Development of more advantageous DNA polymerases through modification of natural enzymes can be illustrated with the following best-known examples: bacteriophage T7 DNA polymerase chemically modified to inactivate 3′-5′ exonuclease activity;[15] amino acid changes in the active site (phenylalanine into tyrosine) of several DNA polymerase I family members, to increase the rate of incorporation of dideoxynucleotides versus deoxynucleotides and to solve the problem of uniform band intensity;[19] truncation of the *Thermus aquaticus* (*Taq*) polymerase I N-terminus;[20] and the substitution of Asn^{543} for Ser^{543} in the large fragment of *Taq* DNA polymerase (Klentaq® DNA polymerase) to prevent pausing during DNA synthesis within template regions of complex structure.[21]

The use of radioactive labels and autoradiographic detection suffers from several limitations. Radioactive isotopes have a short half-life that requires their regular supply, and their storage is potentially hazardous. Moreover, autoradiography is a laborious procedure involving manual manipulations. Great progress was made in sequencing technology in 1987, when radioactive labels were replaced by fluorescent ones that allowed autoradiography to be substituted by laser-induced fluorescence detection, thus making possible the direct reading of electrophoregrams. In 1986 Smith et al.[22] reported the first automated DNA sequencer. They used primers labeled with different fluorescent dyes in four sequencing reactions. In one of the other simplest automated

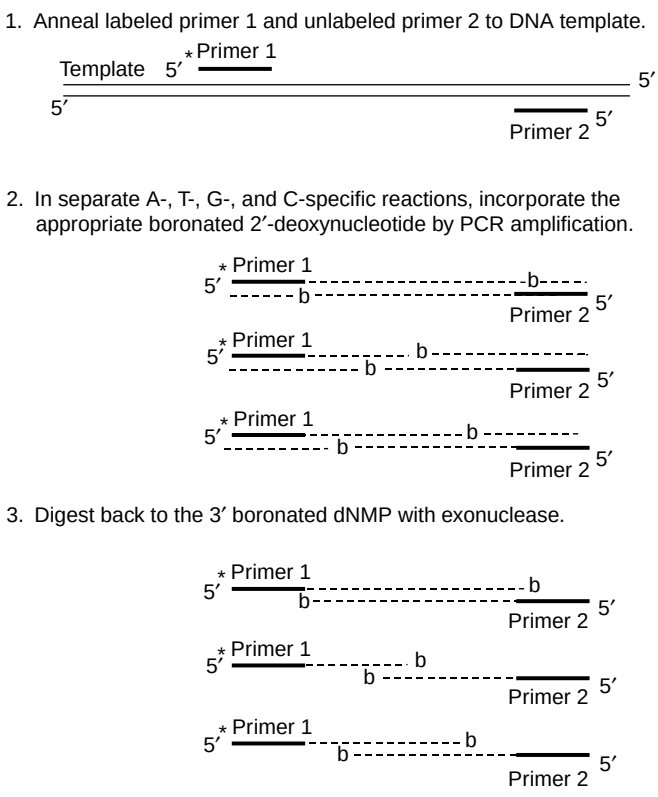

Figure 2 Directed sequencing with boronated nucleotides. A = adenine, T = thymidine, G = guanine, C = cytidine and dNMP = deoxynucleoside monophosphate. [Reproduced by permission of Oxford University Press from K.W. Porter, J.D. Briley, B.R. Shaw, *Nucleic Acids Res.*, **25**, 1611–1617 (1997).]

sequencers[23] the primer was labeled with tetramethylrhodamine iodoacetamide, the reaction mixture was separated into four parts, and sequence identification was similar to the Sanger technique. To avoid primer labeling, Prober et al.[24] proposed a four-colored sequencer based on the use of fluorescently labeled dideoxynucleotide chain terminators, allowing only one sequencing reaction to be performed instead of four. However, overlapping fluorescent spectra of the dyes used in this approach resulted in generally poor accuracy. Internal labeling with fluorescein–dNTPs was proposed by Voss et al.[25] Three automatic sequencers using this labeling and proposed by Smith et al. were commercialized by Applied Biosystems,

DuPont, and Pharmacia, respectively. The development of new fluorescent labels is in progress, including a new class of energy transfer dyes.[26,27]

Clearly being a step forward over manual sequencing, the automated sequencers also have some limitations. Modern sequencing procedures permit up to 600–800-bp-long sequences to be read in one gel. However, electrophoresis in the rather thick slab gels used earlier had to be run at high currents, with accompanying overheating. More recent slab-gel systems are adjusted to thinner gels, which reduces the separation time.[28] A new invention in this field is electrophoresis in capillary tubes filled with polyacrylamide gel. Efficient cooling of the capillaries allows higher electric fields (up to $400\,V\,cm^{-1}$ compared to $30\,V\,cm^{-1}$ in conventional gels) and provides better resolution of separated DNA fragments.[29] The DNA sequencing by capillary electrophoresis is described in detail in the review by Dovichi.[30] Although capillary electrophoresis does not have crucial advantages over ultrathin slab gels, it has some useful practical features, such as easy replacement of separation medium, easy connection to a microtiter plate, and economy in space at high-throughput sequencing. Moreover, low-viscosity polymers used as the separation medium [e.g. noncross-linked polyacrylamide,[30] polyethylene oxide,[31] or polyethylene glycol[32]] are handy for easy refilling of the capillary tubes.

4 LONG-RANGE GENOMIC DNA SEQUENCING STRATEGY

4.1 Random (Shotgun) Strategy

A random (shotgun) sequencing strategy was first employed by Sanger in the course of bacteriophage λ sequencing in 1982.[7] The strategy was based on sequencing of unselected (random) fragments of DNA and did not need any physical map. The theory of this strategy was put forward by Lander and Waterman,[33] and later was developed further and used by Venter's group to decipher successfully several microbial genomes. A detailed description of the shotgun strategy is given in a review by Fraser and Fleischmann[34] and in their work on sequencing of the *Haemophilus influenzae* genome.[10] The major steps of the strategy are listed in Table 2.

DNA of a genome is sheared, e.g. by extraction with phenol–chloroform, resulting in ~50-kb fragments, followed by sonication or nebulization up to fragment lengths of 1–2 kb. The mixture obtained is blunt-ended by BAL31 nuclease, size-fractionated, and cloned into plasmid (pUC), phagemid (pGEM), or phage (M13) vectors. The use of *E. coli* strains deficient in all re-combinant and restriction functions reduces possible rearrangements during the cloning. A library should not contain overrepresented clones, and it is usually characterized by an average representation of 6- to 8-fold genome coverage. Combining the sequence data obtained

Table 2 Whole-genome sequencing strategy[a]

Stage	Description
1. Random small-insert and large-insert library construction	Shear genomic DNA randomly to ~2 kb and 15–20 kb, respectively
2. Library plating	Verify random nature of library and maximize random selection of small-insert and large-insert clones for template production
3. High-throughput DNA sequencing	Sequence sufficient number of sequence fragments from both ends for 6 × coverage
4. Assembly	Assemble random sequence fragments and identify repeat regions
5. Gap closure	
Physical gaps	Order all contigs (fingerprints, peptide links, λ clones, PCR) and provide templates for closure
Sequence gaps	Complete the genome sequence by primer walking
6. Editing	Inspect the sequence visually and resolve sequence ambiguities, including frameshifts
7. Annotation	Identify and describe all predicted coding regions (putative identifications, starts and stops, role assignments, operons, regulatory regions

[a] Reproduced with permission from R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.I. Lui, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraiser, H.O. Smith, J.C. Venter, *Science*, **269**, 496–512 (1995). Copyright 1995 American Association for the Advancement of Science.

into one contiguous sequence is performed using software that simultaneously clusters and combines fragments of the genome using a best-matched strategy. For instance, well-developed TIGR Assembler software has been used for getting full-length microbial genomes.[10] A step of filling gaps is indispensable because, according to Lander and Waterman,[33] the distribution of the gaps within the genome is random. There are two types of gaps: sequence gaps with templates available; and physical gaps breaking links between contigs. Sequence gaps can be filled by primer walking along the template. Missing templates can be obtained either by PCR amplification using all possible primers designed for the ends of each contig or by construction of libraries with longer inserts (in bacteriophage λ or cosmids) covering these gaps. The full-length sequence obtained is checked for possible rearrangements and sequencing accuracy. The final step is identification and annotation of open reading frames (ORFs) by a set of algorithms such as BLAST or BLAZE.

The shotgun strategy is a well-developed, high-throughput, and cost-saving sequencing approach that does not need an ordered set of cosmids. However, it suffers from high (6–12) redundancy. To lower this value, Hoheisel et al.[35] fine-mapped shotgun templates before sequencing to order clones by hybridization with individual DNA fragments. A random library ordered by hybridization was then used for sequence analysis. Apart from lower redundancy, this preselection can add important structural information (e.g. positions and types of repeats). Radelof et al.[36] proposed the hybridization of short oligonucleotides with high-density clone filter grids and demonstrated a twofold decrease in redundancy.

The shotgun approach also demands additional efforts to close gaps. Repetitive elements widespread over the eukaryotic genomes aggravate the problem of mistaken contigs. This challenge will probably be eliminated by further progress in sequencing technologies.

4.2 Directed Sequencing

The final phase of any large-scale sequencing project includes closure of gaps, sequencing of repetitive motifs, etc. and is implemented by directed sequencing, mainly primer walking. In the primer walking procedure specific primers are generated for the sites close to the 3′-ends of a known sequence and used for the following elongation steps and sequencing. The steps of primer design and sequencing are repeated until contigs are fused and both strands are sequenced unambiguously. This scheme was proposed by Strauss et al. in 1986[37] and has been used in several genome projects.

The primer walking approach depends on the availability of inexpensive and reliable oligonucleotide primers. There are a lot of primer design computer programs that automatically suggest primers using a set of well-known criteria, such as features of primer–template interactions [e.g. stability of potential binding sites, especially important for sequencing large inserts in phage P1 artificial chromosome (PAC) or BAC vectors], test for self-complementarity and loop formation, stability of the 3′-terminus, and guanine/cytidine content. Several examples of these programs are OLIGO,[38] Primo,[39] Primer Master,[40] and PRIDE.[41]

The efficiency of the primer walking in manual sequencing is typically about 80% but under optimized conditions of automated DNA sequencing this value can be as high as 95%.[42] The primer walking is characterized by low redundancy (2–3 versus 8–10 in the shotgun approach) and low error level (less than 1% per 500 kb). The accuracy can be improved further by simultaneous walking along both strands using two oppositely directed primers, one to obtain new data and the other to confirm already made reads. The primer walking is of great help in the sequencing of repetitive elements and facilitates integration of individual sequences into contigs. All these data were obtained in the course of systematic sequencing of *S. cerevisiae* chromosome XI in the yeast genome sequencing project.[43]

To avoid the need to synthesize primers for each step of primer walking, Studier[44] proposed sequencing cosmids using libraries of 8-, 9-, and 10-mer oligonucleotide primers. The statistics of priming showed that primer libraries sufficient for determining the sequence of the entire human genome (100 000 cosmids) would be small enough to be prepared and managed. In 1990 Szybalski[45] suggested using a rapid assembly of long specific (modular) primers from short contiguous oligonucleotides taken from a relatively small library containing 4096 hexamers. This idea of a pre-made short oligonucleotide collection was developed further by using oligomers containing 5–7 monomers for constructing modular primers,[46–49] contiguous oligonucleotide components of the primers being either ligated on the template or used without ligation. A necessary pool of short oligonucleotide components additionally can be reduced, e.g. 1200–1400 hexanucleotides is sufficient for primer walking on natural templates, which was substantiated theoretically[50] and further confirmed[51] practically.

Yet another approach to systematic sequencing – the construction of ordered sets of subclones – was first described in 1980.[52] This idea was developed further in creating new methods to produce nested deletions in vitro. Henikoff[53] proposed the generation of nested deletions starting from any fixed point in a cloned DNA fragment using controlled exonuclease III digestion of the template. The resulting linear nested fragments

had common ends corresponding to the 5′-end of the sequencing primer and could be cloned and used for sequencing. There is also an analogous method in which nested deletions are formed by controlled BAL31 nuclease digestion (see Breault et al.[54]). Both procedures are manual and rather laborious; they were widely used before the era of automated sequencing but presently their application is very limited.

An idea to introduce sequencing priming sites throughout the target sequence by bacterial and yeast transposons was elaborated by several research groups. Transposons are inserted into many sites of target DNA cloned in plasmid or phage vectors. The integrated transposons are used to form unique priming sites for the sequencing of adjacent uncharacterized regions. The known DNA tracts within transposons are used as binding sites for sequencing primers. Transpositions are generally made in vivo in bacteria using a donor strain carrying the transposons either within a vector or integrated in the genome. A target of interest is introduced into a transposer host to be hit by transposons. The modified targets are then recovered from the host by genetic selection. To be efficient in sequencing, the transposon technique should meet the following requirements: the transposon should be mobilized easily; selection for transposon insertions into the plasmid and into the target sequence should be efficient, as opposed to the bacterial chromosome; and mapping of the transposition sites should be easy in order to minimize the number of required sequencing reactions. Here are some examples of the transposon systems: γδ;[55] Tn3;[56] Tn5;[57,58] Tn10;[59] AT-2;[60] Ty1.[61] The transposon sequencing approach was applied successfully to the sequencing of repetitive regions and filling the remaining gaps in the nematode *C. elegans* genome sequencing project.[62]

4.3 Combination of Random and Directed Sequencing as the Basic Strategy of the Human Genome Project

The HGP was officially started in 1990 as a 15-year collaborative program of mapping and sequencing of the complete set of human chromosomes (3 Gb). The most widely used approaches to human genome sequencing are currently based on preliminary genetic (genes ordered along a chromosome) and physical (overlapped DNA fragments covering a chromosome) mapping of chromosomes. Three clone libraries of different resolution levels were constructed by separating sheared DNA into fragments of different sizes and then inserting them into vectors capable of being propagated in appropriate hosts (Table 3).

The low-resolution physical map in Figure 3 is a set of overlapped YACs containing unique sites amplified by PCR [so-called STSs introduced by Olson et al.[63]]

Table 3 Clone libraries used for genome mapping and sequencing[a]

Vector	Human DNA insert size range (kb)	Number of clones required to cover the human genome
YAC	100–2000	3000 (1000 kb)
Bacterial artificial chromosome (BAC), PAC	80–350	20 000 (150 kb)
Cosmid	30–45	75 000 (40 kb)
Plasmid	3–10	600 000 (5 kb)
M13 phage	1	3 000 000 (1 kb)

[a] Reproduced with permission from J.C. Venter, H.O. Smith, L. Hood, *Nature*, **381**, 364–366 (1996). Copyright 1996 Macmillan Magazines Limited.
YACs, yeast artificial chromosomes.

or restriction enzyme recognition sites. Cohen et al.[64] described a YAC library of 33 000 clones spanning essentially all human chromosomes. A high-resolution map is made by subcloning YAC inserts into cosmid vectors. Sequencing a cosmid library from both ends provides medium-range linking (about 1 per 10 kb) that spans common repetitive elements. A partially overlapped set of cosmids is then subcloned into M13 vectors and sequenced by a combination of shotgun and primer walking methods.

The accuracy of this approach is high because of high redundancy. However, some limitations still exist:[65] e.g. YAC clones are structurally unstable, which results in rearrangements and deletions; and arrayed repeats present problems for high-resolution mapping.

A novel approach to genomic sequencing that eliminates the stage of prior physical mapping uses BAC clones as a starting point.[65,66] The BAC cloning system (Figure 3) was adapted to maintain genomic fragments 300-kb long, the clones being stable even after 100 generations of serial growth.[67] The BAC-end sequencing strategy originally proposed to accelerate genome sequencing provides markers every 5 kb throughout the genome. These so-called STCs link all BAC clones to each other and, being combined with the fingerprint of each BAC clone, supply BACs of interest with ''seeds''. These ''seed '' BACs can be sequenced by any method. Venter's group believes that such an approach might eliminate the problem of gaps in high-resolution physical mapping.

Moreover, Perkin-Elmer announced a new, fully automated capillary-based sequencer ABI PRISM 3700 characterized by high throughput (~1000 samples per day with 15 min of hands-on operator time). This allowed Venter's group to declare the completion of human genome sequencing in the year 2000.

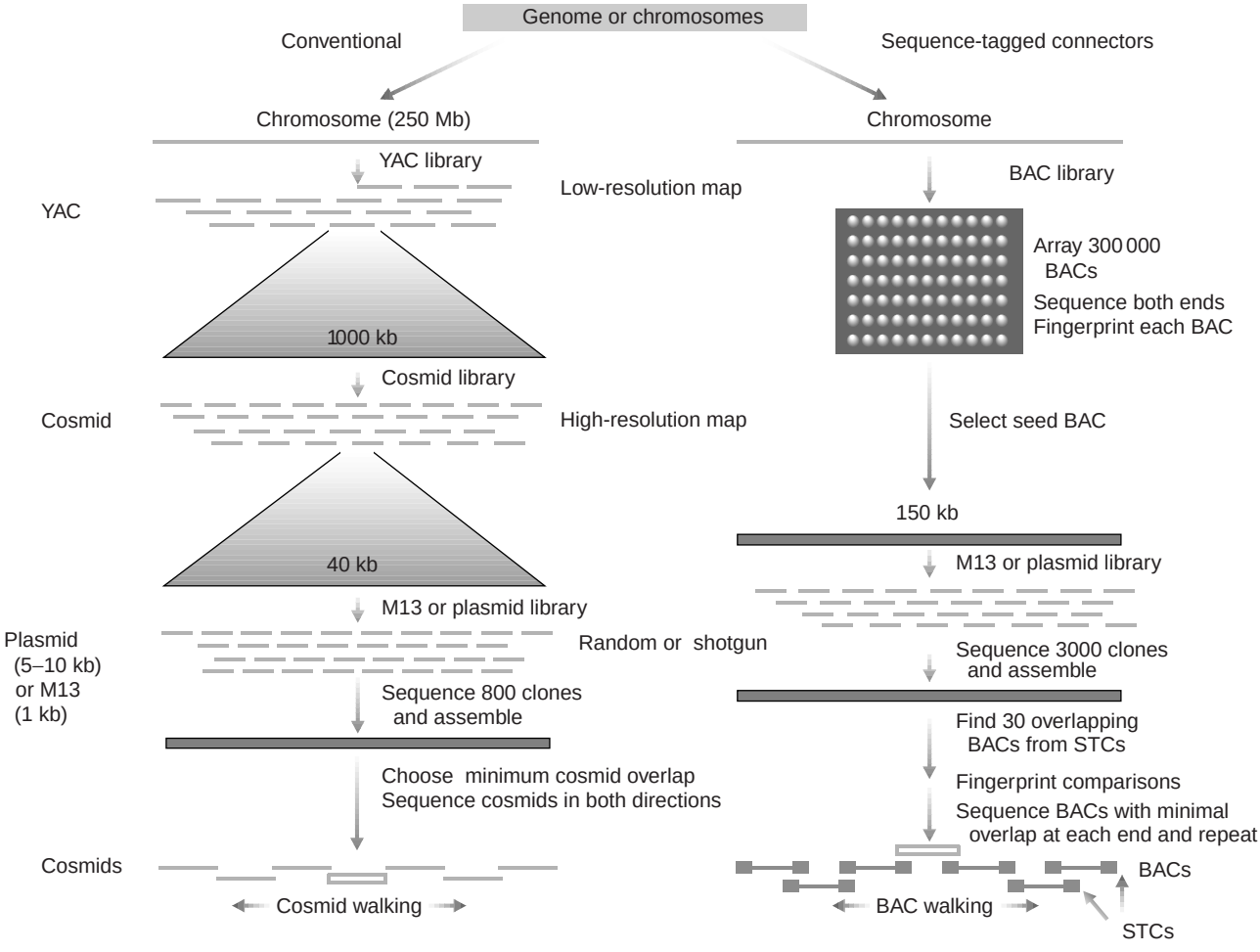

Figure 3 The conventional sequencing approach and the newly proposed STC (sequence-tagged connector) approach. [Reproduced with permission from J.V. Venter, H.O. Smith, L. Hood, *Nature*, **381**, 364–366 (1996). Copyright 1996 Macmillan Magazines Limited.]

5 SHORT DNA/RNA SEQUENCING APPROACHES

5.1 Complementary DNA Sequencing

Whole-genome sequencing is becoming a routine though laborious procedure, but gene identification and functional analysis of encoded genes remain problematic. From this point of view, cDNA sequencing is one of the main sources of information. Detailed analysis of new methodologies allowing expressed genes to be identified and differences in their expression levels to be quantitated is presented in reviews by Fraser and Fleischmann[34] and Carulli et al.[68]

5.1.1 Expressed Sequence Tags and Serial Analysis of Gene Expression

Expressed sequence tags (ESTs) and serial analysis of gene expression (SAGE) provide a rapid and efficient method of searching for new genes and identification of the coding regions within gene sequences, as well as studying differential gene expression in a variety of normal, developmental, and affected tissues. The EST approach[69] first appeared in the work of Venter's group on the identification of new human genes expressed in brain, and later was applied successfully to many other tissues (liver, heart, testis, pancreatic islet cells, etc.): a comprehensive list of references is given in a review by Fraser and Fleischmann.[34] The SAGE methodology was developed by Velculescu et al.,[70] who analyzed the messenger RNA (mRNA) from human pancreas. The technique allows one to measure the transcript abundance based on the position and content of small sequence tags. Recent examples of SAGE application are the cataloging of ESTs expressed in skeletal muscle,[71] the identification of genes differentially expressed after activation of human arterial endothelial cells,[72] and the analysis of transcripts in non-small cell lung cancer.[73]

The EST and SAGE methodologies complement each other in providing sequence tags for the identification of genes. Development of further modifications of genome-wide gene analysis combining both EST and SAGE[74] is under way. Being widely used, the EST and SAGE methodologies are not free from some limitations. In particular, both of them can be applied only to eukaryotic organisms with polyadenylated mRNAs. Accurate quantitative data of EST analysis depend on large-scale sequencing, which implies the availability of sequencing facilities. The classic SAGE requires relatively high amounts of input RNA, so it is inapplicable when RNA amounts are limited. However, a micro-SAGE modification of the method proposed by Datson et al.[75] works successfully with 500–5000-fold less starting material.

5.1.2 Differential Display

The differential display technique first developed by Liang and Pardee[76] is PCR-based and provides data on gene expression. This method uses a set of small random primers to amplify partial cDNAs from subsets of mRNAs. The resulting labeled short sequences are separated in sequencing gels; samples running in parallel show differences in mRNA compositions. The separated labeled tags are isolated from the gel and sequenced. This technique does not require knowledge of the genome context, but it requires extensive follow-up work and is relatively slow.

5.1.3 Complementary DNA Microarray Analysis and DNA Chips

These hybridization techniques, as alternatives to direct cDNA sequencing, are fast, accurate, and cheap. Providing that the genome context is known and using the recently developed high-density gridding technique, DNA microarrays can represent every individual ORF from a particular organism and provide a true global pattern of the gene expression under various conditions.

5.2 RNA Sequencing

RNA is sequenced according to the general principles given in section 3.

5.2.1 Chemical Degradation

Four different base-specific chemical reactions were proposed for partial and specific digestion of end-labeled RNA.[12] Strand breaks at the sites of the chemical attack generate a nested set of labeled fragments that are resolved by electrophoresis. The modifying reagents are presented in Table 1. Kochetkov et al.[77] introduced an improved uridine-specific modification, later used by Waldmann et al.[78] Solid-phase modifications of the chemical RNA sequencing were described. In this case the reactions are carried out on Whatman DE 81 anion-exchange paper[79] or DEAE-cellulose sheets[80] as support. The base-specific chemical cleavage for RNA sequencing is a powerful technique that is largely insensitive to secondary structure. Unfortunately, the procedure is time-consuming and its application is limited at present.

5.2.2 Enzymatic Degradation

Ribonucleases used for RNA sequencing are given in Table 4. All these endonucleases are single-strand-specific. A feature of RNA as compared to DNA is the more complex secondary structure, often hindering enzymatic hydrolysis. Kisselev et al.[86] proposed a complete and selective modification of cytidilic residues by methoxyamine–bisulfite mixtures that leads to RNA unfolding. This helps to generate a more uniform set of fragments after RNAse treatment. The $C^{modified}$–N bond is easily digested by T2 and A RNAses.

5.2.3 Primer Extension

The general principle of specific termination of newly synthesized polynucleotide chains proposed by Sanger[5] and adapted for RNA remains a dominant approach in RNA sequencing. The first chain-terminating RNA sequencing was demonstrated by Kramer and Mills,[87] who used RNA-dependent RNA polymerase (Qβ-replicase) and 3′-deoxyribonucleoside 5′[α-^{32}P]-triphosphates as specific chain-terminating ribonucleotide analogs. Another RNA polymerase[88] (SP6 polymerase) as well as other terminating ribonucleotide analogs[89] (3′-*O*-methylnucleoside 5′-triphosphates) were also tested for sequencing.

Nevertheless, RNA-dependent DNA polymerases (reverse transcriptases) introduced by Brownlee and Cartwright[90] are used preferentially (for a review, see Hahn et al.[91]). The procedures for making use of reverse transcriptases from avian myeloblastosis virus and Moloney murine leukemia virus are developed in detail.[92,93] These methods work equally well with all kinds of labeling, be it labeled α-dNTPs, labeled primers,[94] or fluorescently labeled ddNTPs.[95] The latter type of labeling minimizes unspecific chain-termination events and is adapted for automated sequencing. Other possible ways to avoid premature chain termination are treatment of RNA templates with methyl mercury[96] and reverse transcription at higher temperatures.[97]

A review of mass spectroscopic RNA sequencing methods can be found in section 5.3.

Table 4 Ribonucleases used for RNA sequencing

Ribonuclease	Specificity	Comments
T2	Unspecific	For statistical cleavage
T1[82]	Guanosine	
U2[82]	Adenine > guanosine	Proposed special conditions for specific cleavage adenosine-3′-phosphate ↓ N
A[82]	Uridine + cytidine	Cleavage rates differ greatly: pyrimidine ↓ purine ≫ pyrimidine ↓ pyrimidine
Phy1 (from *Physarum polycephalum*)[83]	Distinguishes between cytidine and uridine	Strong preference for N−cytidine over N−uridine uridine−N over cytidine−N
Bc (from *Bacillus cereus*)	Uridine + cytidine	Pyrimidine residues
Staphylococcus aureus nuclease[84]	Uridine + cytidine	Pyrimidine−N at pH 3.5 without Ca^{2+} N−uridine and N−adenine at pH 7.5 in Ca^{2+}
M1 nuclease (from *Cucumis melo*)[85]	N ↓ adenine, N ↓ uridine, N ↓ guanosine	Stable in 7 M urea
Ribonuclease CL3 (chicken liver)[86]	Cytidine	

5.3 New Approaches

5.3.1 Mass Spectrometry

Mass spectrometry (MS) is one of the unconventional approaches ranging between simple replacement of a fluorescent detector in gel electrophoresis with a mass spectrometric detector and attempts to sequence a single DNA molecule in an ion trap. Matrix-assisted laser desorption and ionization (MALDI) was first proposed for the sequencing of large proteins up to 300 000 Da,[98,99] but in a technically improved form it is now applied also to the sequencing of nucleic acids.[100,101] The most promising approach is to replace electrophoretic separation of nested DNA (RNA) fragments obtained by the traditional Sanger technique, chemical or enzymatic cleavage with mass spectrometric separation and detection.[102] Cantor et al.[103] suggested a new and potentially high-throughput DNA sequencing approach that uses immobilized duplex probes with five-base-long single-stranded overhangs to bind DNA templates serving as primers for Sanger sequencing. In early studies the sequencing of DNA by MALDI was limited to very short oligonucleotides owing to fragmentation of longer chains during a laser desorption step. Recently introduced matrices[100,104] have allowed the reliable sequencing of longer DNA fragments of approximately 100–130 nucleotides in length.[101]

Matrix-assisted laser desorption and ionization/mass spectrometry (MALDI/MS) relies on a sample preparation in which the analyzed substance (analyte) is co-crystallized with an excess of a matrix, usually an organic acid. The analyte–matrix mixture is desorbed/ionized by a short laser pulse; the generated ions are all accelerated to the same energy in the electric field and subsequently separated in a field-free drift region according to their mass-to-charge ratio. Sequencing data relying on the absolute molecular mass information rather than on relative gel electrophoretic mobility significantly increase the fidelity of sequencing. Other obvious advantages are fast sequence determination and use of unlabeled primers.

In spite of successful sequencing of rather long DNA stretches (about 500 bp), the main MALDI application is comparative (diagnostic) DNA sequencing and re-sequencing or detection of mutations, because the DNA to be re-sequenced is usually not very large. Unambiguous detection of polymorphic sequences has been demonstrated in both homozygous and heterozygous samples.[105]

5.3.2 Scanning Tunneling Microscopy

Scanning tunneling microscopy (STM) was used to visualize duplex DNA at atomic resolution.[106] Experimental STM profiles showed a double-helical structure, base pairs, atomic-scale substructure, and excellent correlation

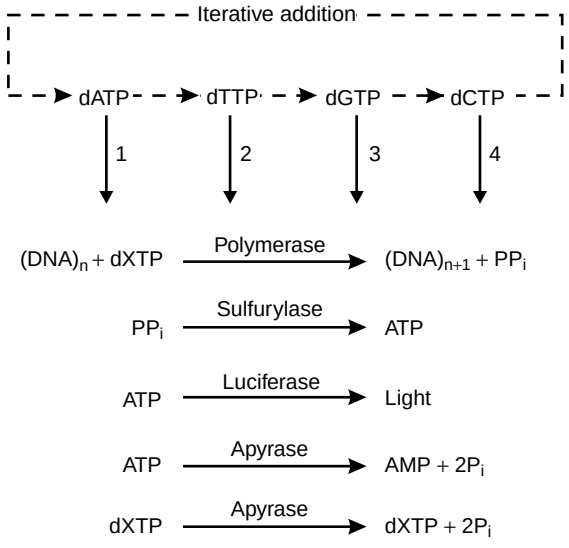

Figure 4 Pyrosequencing method. PP_i = inorganic pyrophosphate; dXTP = one of the four deoxynucleoside-5′-triphosphates. [Reproduced with permission from M. Ronaghi, M. Uhlen, P. Nyren, *Science*, **281**, 363–365 (1998). Copyright 1998 American Association for the Advancement of Science.]

with atomic contours of the Van der Waals surface of DNA. Since 1990, no additional information on this approach has appeared.

5.3.3 Pyrosequencing

A novel sequencing approach[107] involves coupled enzymatic reactions to monitor the release of inorganic pyrophosphate (PP_i) during the DNA polymerase reaction (Figure 4). An oligonucleotide primer is hybridized to a template, and sequencing is carried out by a stepwise elongation of the primer strand upon successive addition of four different dNTPs in the presence of DNA polymerase, ATP sulfurylase, firefly luciferase, and apyrase (nucleotide-degrading enzyme). The synthesis of DNA is accompanied by an immediately monitored release of PP_i in an amount equivalent to that of the incorporated dNTP. The intensity of light emitted in the luciferase-catalyzed reaction can be measured readily with a suitable light-sensitive device (e.g. a luminometer). dNTPs not incorporated are degraded by apyrase. This technique is advantageous in terms of accuracy, flexibility, parallel processing, and automation. Moreover, it makes the use of labeled primers, labeled nucleotides, and electrophoresis unnecessary. Another important advantage is the absence of the compression effect due to the secondary structure of the template. The problem of determining the number of incorporated nucleotides in homopolymer tracks, which stems from nonlinear light response, still exists but according to the authors can be compensated for by using certain software algorithms.

5.3.4 Sequencing by Hybridization

The sequence information is obtained here by the hybridization of small probes to a target to be sequenced. Two formats have been proposed. In format I DNA targets are bound to a substrate as a replica array and each replica is hybridized with every possible oligonucleotide probe added in solution. Up to 10 000 clones can be hybridized simultaneously on one membrane. This format was first successfully tested in a blind sequencing of three DNA fragments each of 343 bp in length.[108] In 1998, Drmanac's group reported sequencing by hybridization (SBH) of 1.1-kb DNA fragments.[109] However, in the case of octanucleotide probes as many as 65 000 hybridizations are required, which makes this format too laborious. In an alternative approach (format II) large arrays of oligonucleotides on a solid support are hybridized to a labeled target sequence. Details of format II are given in reviews by Southern[110] and O'Donnell-Maloney et al.,[111] where the techniques of oligonucleotide arrays preparation, possible support material (surface-modified glass, polypropylene, glass with small patches of activated polyacrylamide, etc.), and different schemes of data collection and processing are discussed. All the applications of oligonucleotide arrays are underlain by specific base pairing and include comparing sequences/detecting mutations (see section 6.1) and sequencing per se. The theory of the latter is based on the possibility of distinguishing between perfect and imperfect duplexes and reconstructing the sequence of a target nucleic acid from the corresponding set of oligonucleotides hybridized with it. The maximum length of the analyzed target is roughly equal to the square root of the number of oligonucleotides in the array (e.g. to sequence a 200-bp fragment, an array of 65 536 octanucleotides is needed). The length of oligonucleotides can also be a critical factor. The attempts to use hexanucleotide arrays[114] were of limited success because short duplexes are stable enough only under nonstringent conditions (at high salt and/or low temperatures), which can cause false hybridization signals or even prevent hybridization owing to internal base pairing in the target. Therefore, the use of longer oligonucleotides is preferable but the arrays become too large for existing production methods. Most of the suggested arrays are prepared from octa- and nanomers. The theory of SBH and the accompanying software is being developed.[115,116]

There are only a few known examples of successful direct SBH, but a combination of SBH and gel-sequencing techniques is promising for large-scale sequencing.

A major disadvantage of SBH is ambiguous reconstruction of DNA sequences due to the presence of various repeats within long DNA fragments. Mirzabekov et al.[117] supplemented the SBH procedure with the measurement of distances between certain sites (e.g.

restriction sites or PCR priming sites), which eliminated the problem of repetitive elements.

Another SBH feature is a strong dependence on base composition. An enhanced version of SBH, termed positional sequencing by hybridization (PSBH), was proposed by Broude et al.[118] It uses immobilized duplex probes with single-stranded 3′ overhangs, which pick their perfectly matched DNA targets out of a mixture of different single-stranded DNAs. Ligation of a single-stranded target hybridized to its complementary counterpart in a duplex probe improves the differentiation of perfectly matched targets from those containing mismatches. The use of PSBH in combination with standard solid-state Sanger sequencing without ligation was demonstrated.[119] As little as five bases of known terminal sequence is sufficient for priming.

6 COMPARATIVE SEQUENCING

6.1 DNA Chips

Possible applications of SBH include physical mapping of overlapped DNA clones, gene expression monitoring, identification of microorganisms, and thermodynamic analysis of DNA duplexes. Oligonucleotide arrays are characterized by a wide range of comparative (diagnostic) applications, such as screening of sequence polymorphism and detection of mutations, HLA allotyping, genetic identification, and revealing of genetic variations. Several examples are given in the work by Hacia et al.[120]

6.1.1 Detecting Mutations

High-density arrays have been used to screen 3.4-kb exon 11 of the human breast and ovarian cancer gene (BRCA 1) for all possible heterozygous polymorphisms and mutations. An evolutionary sequence comparison of human BRCA exon 11 and its orthologs in closely related species was done. A 2.3-Mb length of the human genomic DNA was examined by a combination of SBH and gel-based sequencing. Based on SBH pilot genotyping, chips were constructed to characterize human diversity.

6.1.2 Comparative Sequencing

As shown in model experiments, the use of arrays allows one to digitize hybridization patterns and thus to facilitate their comparison.

6.2 Bioinformatics

A scientific field called ''bioinformatics'' is focusing on the creation of computing systems for collection, refinement, analysis, and storage of information. Two areas of bioinformatics are most pertinent to sequencing:

1. Software that allows full-length sequences to be constructed from sequenced regions. An example is one of the most recent programs described by Parson et al.[121]
2. Comprehensive analysis and comparison of sequences available from databases. Overview of this application was summarized by Sander et al., who provided a creative computing system.[122]

Large-scale genome projects generate a rapidly increasing volume of sequence information. By 1999, the genomes of several model organisms had been sequenced, and the HGP is expected to be finished by the year 2001. However, most of the available sequences are functionally uncharacterized. The research in bioinformatics results in the development of methods for computational characterization of these sequences, and the number of computer programs specialized for various analytical tasks is growing rapidly. The most improved packages of information and software are distributed through web sites of the largest genome centers: Sanger Center, European Bioinformatics Institute (EBI), Centre National de la Recherche Scientifique (CNRS), Max-Plank-Institute in Europe; Washington University, National Center for Biotechnology Information (NCBI) and other institutions in the USA.

ACKNOWLEDGMENTS

The author thanks Professor Eugene Sverdlov and Drs Yuri Lebedev, Maria Kostina and Boris Glotov for fruitful discussions and essential help in manuscript preparation.

ABBREVIATIONS AND ACRONYMS

CNRS	Centre National de la Recherche Scientifique
ddNTP	Dideoxynucleoside Triphosphate
dNMP	Deoxynucleoside Monophosphate
dNTP	Deoxynucleoside Triphosphate
EBI	European Bioinformatics Institute
EST	Expressed Sequence Tag
HGP	Human Genome Project
MALDI	Matrix-assisted Laser Desorption and Ionization
MALDI/MS	Matrix-assisted Laser Desorption and Ionization/Mass Spectrometry
MS	Mass Spectrometry

NCBI	National Center for Biotechnology Information
ORF	Open Reading Frame
PAC	Phage P1 Artificial Chromosome
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PP_i	Inorganic Pyrophosphate
PSBH	Positional Sequencing by Hybridization
SAGE	Serial Analysis of Gene Expression
SBH	Sequencing by Hybridization
STC	Sequence-tagged Connector
STM	Scanning Tunneling Microscopy
STS	Sequence-tagged Site
YAC	Yeast Artificial Chromosome

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • Capillary Electrophoresis of Nucleic Acids • Genome Physical Mapping Using BACs • Mass Spectrometry of Nucleic Acids

REFERENCES

1. R.W. Holley, J. Apgar, G.A. Everett, J.T. Madison, M. Marquisee, S.H. Merrill, J.R. Penswick, A. Zamir, 'Structure of Ribonucleic Acid', *Science*, **147**(3664), 1462–1465 (1965).
2. F. Sanger, H. Tuppy, 'The Amino-acid Sequence in the Phenylalanyl Chain of Insulin', *Biochem. J.*, **49**(4), 463–481 (1951).
3. A.M. Maxam, W. Gilbert, 'A New Method for Sequencing DNA', *Proc. Natl. Acad. Sci. USA*, **74**(2), 560–564 (1977).
4. R. Wu, 'Development of the Primer-extension Approach: A Key Role in DNA Sequencing', *TIBS*, **19**, 429–433 (1994).
5. F. Sanger, S. Nicklen, A.R. Coulson, 'DNA Sequencing with Chain-terminating Inhibitors', *Proc. Natl. Acad. Sci. USA*, **74**(12), 5463–5467 (1977).
6. F. Sanger, G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, C.A. Fiddes, C.A. Hutchison, P.M. Slocombe, M. Smith, 'Nucleotide Sequence of Bacteriophage Phi X174 DNA', *Nature*, **246**(5596), 687–695 (1977).
7. F. Sanger, A.R. Coulson, G.F. Hong, D.F. Hill, G.B. Petersen, 'Nucleotide Sequence of Bacteriophage Lambda DNA', *J. Mol. Biol.*, **162**(4), 729–773 (1982).
8. A.T. Bankier, S. Beck, R. Bohni, C.M. Brown, R. Cerny, M.S. Chee, C.A. Hutchison 3d, T. Kouzaridis, J.A. Martignetti, E. Preddie, 'The DNA Sequence of the Human Cytomegalovirus Genome', *DNA Seq.*, **2**(1), 1–12 (1991).
9. S.J. Goebel, G.P. Johnson, M.E. Perkus, S.W. Davis, J.P. Winslow, E.S.J. Paoletti, 'The Complete DNA Sequence of Vaccinia Virus', *Virology*, **179**(1), 247–266 (1990).
10. R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.I. Lui, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraiser, H.O. Smith, J.C. Venter, 'Whole-genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd', *Science*, **269**(5223), 496–512 (1995).
11. F.S. Collins, A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland, L. Walters, 'New Goals for the US Human Genome Project: 1998–2003', *Science*, **282**(5389), 682–689 (1998).
12. D.A. Peattie, 'Direct Chemical Method for Sequencing RNA', *Proc. Natl. Acad. Sci. USA*, **76**(4), 1760–1764 (1979).
13. Y. Kuchino, S. Nishimura, 'Enzymatic RNA Sequencing', *Methods Enzymol.*, **180**, 154–163 (1989).
14. R.C. Mierensdorf, D. Pfeffer, 'Sequencing of RNA Transcripts Synthesized In Vitro from Plasmids Containing Bacteriophage Promoters', *Methods Enzymol.*, **152**, 563–566 (1987).
15. S. Tabor, C.C. Richardson, 'DNA Sequence Analysis with a Modified Bacteriophage T7 DNA Polymerase', *Proc. Natl. Acad. Sci. USA*, **84**(14), 4767–4771 (1987).
16. K.L. Nakamaye, G. Gish, F. Eckstein, H.P. Vosberg, 'Direct Sequencing of Polymerase Chain Reaction Amplified DNA Fragments Through the Incorporation of Deoxynucleoside Alpha-thiotriphosphates', *Nucleic Acids Res.*, **16**(21), 9947–9959 (1988).
17. S. Labeit, H. Lehrach, R.S. Goody, 'A New Method of DNA Sequencing Using Deoxynucleoside Alpha-thiotriphosphates', *DNA*, **5**(2), 173–177 (1986).
18. K.W. Porter, J.D. Briley, B.R. Shaw, 'Direct PCR Sequencing with Boronated Nucleotides', *Nucleic Acids Res.*, **25**(8), 1611–1617 (1997).
19. S. Tabor, C.C. Richardson, 'A Single Residue in DNA Polymerases of the *E. coli* DNA Polymerase I Family is Critical for Distinguishing between Deoxy- and Dideoxyribonucleotides', *Proc. Natl. Acad. Sci. USA*, **92**(14), 6339–6343 (1995).

20. W.M. Barnes, 'The Fidelity of *Taq* Polymerase Catalyzing PCR is Improved by an N-terminal Deletion', *Gene*, **112**(1), 29–35 (1992).
21. K.B. Ignatov, A.A. Bashirova, A.I. Miroshnikov, V.M. Kramarov, 'Mutation S543N in the Thumb Subdomain of the *Taq* DNA Polymerase Large Fragment Suppresses Pausing Associated with the Template Structure', *FEBS Lett.*, **448**(1), 145–148 (1999).
22. L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B.H. Kent, L.E. Hood, 'Fluorescence Detection in Automated DNA Sequence Analysis', *Nature*, **321**(6071), 674–679 (1986).
23. W. Ansorge, B.S. Sproat, J. Stegemann, C. Schwager, 'A Nonradioactive Automated Method for DNA Sequence Determination', *J. Biochem. Biophys. Methods*, **13**(6), 315–323 (1986).
24. J.M. Prober, G.L. Trainor, R.J. Dam, F.W. Hobbs, C.W. Robertson, R.J. Zagursky, A.J. Cocuzza, M.A. Jensen, K. Baumeister, 'A System for Rapid DNA Sequencing with Fluorescent Chain-terminating Dideoxynucleotides', *Science*, **238**(4825), 336–341 (1987).
25. H. Voss, C. Schwager, U. Wirkner, J. Zimmermann, H. Erfle, N.A. Hewitt, T. Rupee, J. Stegemann, W. Ansorge, 'New Procedure for Automated DNA Sequencing with Multiple Internal Labeling by Fluorescent dUTP', *Methods Mol. Cell. Biol.*, **3**, 30–34 (1992).
26. B.B. Rosenblum, L.G. Lee, S.L. Spurgeon, S.H. Khan, S.M. Menchen, C.R. Heiner, S.M. Chen, 'New Dye-labeled Terminators for Improved DNA Sequencing Patterns', *Nucleic Acids Res.*, **25**(22), 4500–4504 (1997).
27. L.G. Lee, S.L. Spurgeon, C.R. Heiner, S.C. Benson, B.B. Rosenblum, S.M. Menchen, R.J. Graham, A. Constantinescu, K.G. Upadhya, J.M. Cassel, 'New Energy Transfer Dyes for DNA Sequencing', *Nucleic Acids Res.*, **25**(14), 2816–2822 (1997).
28. L.M. Smith, R.L. Brumley, E.C. Buxton, M. Giddings, M. Marchbanks, X. Tong, 'High-speed Automated DNA Sequencing in Ultrathin Slab Gels', *Methods Enzymol.*, **271**, 219–237 (1996).
29. L.M. Smith, 'High-speed DNA Sequencing by Capillary Gel Electrophoresis', *Nature*, **349**(6312), 812–813 (1991).
30. N.J. Dovichi, 'DNA Sequencing by Capillary Electrophoresis', *Electrophoresis*, **18**(12–13), 2393–2399 (1997).
31. E.N. Fung, H.M. Pang, E.S. Yeung, 'Fast DNA Separations Using Poly(ethylene oxide) in Non-denaturing Medium with Temperature Programming', *J. Chromatogr. A*, **806**(1), 157–164 (1998).
32. S. Menchen, B. Johnson, M.A. Winnik, B. Xu, 'Flowable Networks as DNA Sequencing Media in Capillary Column', *Electrophoresis*, **17**(9), 1451–1459 (1996).
33. E.S. Lander, M.S. Waterman, 'Genomic Mapping by Fingerprinting Random Clones: A Mathematical Analysis', *Genomics*, **2**(3), 231–239 (1988).
34. C.M. Fraser, R.D. Fleischmann, 'Strategies for Whole Microbial Genome Sequencing and Analysis', *Electrophoresis*, **18**(8), 1207–1216 (1997).
35. P. Scholler, A.E. Karger, S. Meier-Ewert, H. Lehrach, H. Delius, J.D. Hoheisel, 'Fine-mapping of Shotgun Template-libraries; an Efficient Strategy for the Systematic Sequencing of Genomic DNA', *Nucleic Acids Res.*, **23**(19), 3842–3849 (1995).
36. U. Radelof, S. Hennig, P. Seranski, M. Steinfath, J. Ramser, R. Reinhardt, A. Poustka, F. Francis, H. Lehrach, 'Preselection of Shotgun Clones by Oligonucleotide Fingerprinting: An Efficient and High Throughput Strategy to Reduce Redundancy in Large-scale Sequencing Projects', *Nucleic Acids Res.*, **26**(23), 5358–5364 (1998).
37. E.C. Strauss, J.A. Kobori, G. Siu, L.H. Hood, 'Specific-primer-directed DNA Sequencing', *Anal. Biochem.*, **154**(1), 353–360 (1986).
38. W. Rychlik, R.E. Rhoads, 'A Computer Program for Choosing Optimal Oligonucleotides for Filter Hybridization, Sequencing and In Vitro Amplification of DNA', *Nucleic Acids Res.*, **17**(21), 8543–8551 (1989).
39. P. Li, K.C. Kupfer, C.J. Davies, D. Burbee, G.A. Evans, H.R. Garner, 'PRIMO: A Primer Design Program that Applies Base Quality Statistics for Automated Large-scale DNA Sequencing', *Genomics*, **40**(3), 476–485 (1997).
40. V. Proutski, E.C. Holmes, 'Primer Master: A New Program for the Design and Analysis of PCR Primers', *Comput. Appl. Biosci.*, **12**(3), 253–255 (1996).
41. S. Haas, M. Vingron, A. Poustka, S. Wiemann, 'Primer Design for Large Scale Sequencing', *Nucleic Acids Res.*, **26**(12), 3006–3012 (1998).
42. S. Wiemann, T. Rupp, J. Zimmermann, H. Voss, C. Schwager, W. Ansorge, 'Primer Design for Automated DNA Sequencing Utilising T7 DNA Polymerase and Internal Labeling with Fluorescein-15-dATP', *BioTechniques*, **18**(4), 688–696 (1995).
43. B. Dujon, D. Alexandraki, B. Andre, W. Ansorge, V. Baladron, J.P.G. Ballesta, A. Banrevi, P.A. Bollee, M. Bolotin-Fukuhara, P. Bossier, 'Complete DNA Sequence of Yeast Chromosome XI', *Nature*, **369**(6479), 371–378 (1994).
44. F.W. Studier, 'A Strategy for High-volume Sequencing of Cosmid DNAs: Random and Directed Priming with a Library of Oligonucleotides', *Proc. Natl. Acad. Sci. USA*, **86**(18), 6917–6921 (1989).
45. W. Szybalski, 'Proposal for Sequencing DNA Using Ligation of Hexamers to Generate Sequential Elongation Primers (SPEL-6)', *Gene*, **90**(1), 177–178 (1990).
46. J. Kieleszawa, J.J. Dunn, F.W. Studier, 'DNA Sequencing by Primer Walking with Strings of Contiguous Hexamers', *Science*, **258**(5089), 1787–1791 (1992).
47. L.E. Kotler, D. Levin-Sonkin, I.A. Sobolev, A.D. Beskin, L. Ulanovsky, 'DNA Sequencing: Modular Primers Assembled from a Library of Hexamers and

Pentamers', *Proc. Natl. Acad. Sci. USA*, **90**(9), 4241–4245 (1993).
48. T. Azhikina, S. Veselovskaya, V. Myasnikov, V. Potapov, O. Ermolayeva, E. Sverdlov, 'Strings of Contiguous Modified Pentamers with Increased DNA-binding Affinity can be Used for DNA Sequencing by Primer Walking', *Proc. Natl. Acad. Sci. USA*, **90**(24), 11 460–11 462 (1993).
49. W. Hou, L.M. Smith, 'DNA Sequencing with a Hexamer String Primer and Dye-labeled Terminators', *Anal. Biochem.*, **221**(1), 136–141 (1994).
50. N.S. Ghiso, H. Parekh, G.G. Lennon, 'A Subset of 1200 Hexamers is Sufficient to Sequence over 95% of cDNAs by Hexamer String Primer Walking', *Genomics*, **17**(3), 798–799 (1993).
51. T. Azhikina, M. Kostina, N. Skaptsova, V. Potapov, D. Berg, E. Sverdlov, 'Factors Affecting the Priming Efficiency of Short Contiguous Oligonucleotide Strings in the Primer Walking Strategy of DNA Sequencing', *DNA Seq.*, **6**(4), 211–216 (1996).
52. A.M. Frischauf, H. Garoff, H. Lehrach, 'A Subcloning Strategy for DNA Sequence Analysis', *Nucleic Acids Res.*, **8**(2), 5541–5549 (1980).
53. S. Henikoff, 'Ordered Deletions for DNA Sequencing and In Vitro Mutagenesis by Polymerase Extension and Exonuclease III Gapping of Circular Templates', *Nucleic Acids Res.*, **18**(10), 2961–2966 (1990).
54. D.T. Breault, M.L. Stover, M. Li, A.C. Lichtler, D.W. Rowe, 'Novel Use for BAL31 Nuclease-generated Nested Deletions: Sequencing from the Inside Out', *BioTechniques*, **18**(4), 614–616 (1995).
55. M. Strathmann, B.A. Hamilton, C.A. Mayeda, M.I. Simon, E.M. Meyerowitz, M.J. Pallazzolo, 'Transposon-facilitated DNA Sequencing', *Proc. Natl. Acad. Sci. USA*, **88**(4), 1247–1250 (1991).
56. M. Morita, A. Umemoto, Z.X. Li, N. Nakazono, Y. Sugino, 'Nested Deletions from a Fixed Site as an Aid to Nucleotide Sequencing: An In Vitro System Using Tn3 Transposase', *DNA Res.*, **3**(6), 431–433 (1996).
57. S.H. Phadnis, H.V. Huang, D.E. Berg, 'Tn5supF, a 264-base-pair Transposon Derived from Tn5 for Insertion Mutagenesis and Sequencing DNAs Cloned in Phage Lambda', *Proc. Natl. Acad. Sci. USA*, **86**(15), 5908–5912 (1989).
58. D. York, K. Welch, I.Y. Goryshin, W.S. Rezniikoff, 'Simple and Efficient Generation In Vitro of Nested Deletions and Inversions: Tn5 Intramolecular Transposition', *Nucleic Acids Res.*, **26**(8), 1927–1933 (1998).
59. N. Kleckner, J. Bender, S. Gottesman, 'Uses of Transposons with Emphasis on Tn10', *Methods Enzymol.*, **204**, 139–180 (1991).
60. S. Williams, L. Hayes, L. Elsenboss, A. Williams, C. Andre, R. Abramson, J.F. Thompson, P.M. Milos, 'Sequencing of the Cholesteryl Ester Transfer Protein 4′ Regulatory Region Using Artificial Trasposons', *Gene*, **197**(1–2), 101–107 (1997).
61. S.E. Devine, J.D. Boeke, 'Efficient Integration of Artificial Transposons into Plasmid Targets In Vitro: A Useful Tool for DNA Mapping, Sequencing and Genetic Analysis', *Nucleic Acids Res.*, **22**(18), 3765–3772 (1994).
62. S.E. Devine, S.L. Chissoe, Y. Eby, R.K. Wilson, J.D. Boeke, 'A Transposon-based Strategy for Sequencing Repetitive DNA in Eukaryotic Genomes', *Genome Res.*, **7**(5), 551–563 (1997).
63. M. Olsen, L. Hood, C. Cantor, D. Botstein, 'A Common Language for Physical Mapping of the Human Genome', *Science*, **245**(4925), 1434–1435 (1989).
64. D. Cohen, I. Chumakov, J. Weissenbach, 'A First-generation Physical Map of the Human Genome', *Nature*, **366**(6456), 698–701 (1993).
65. J.C. Venter, H.O. Smith, L. Hood, 'A New Strategy for Genome Sequencing', *Nature*, **381**(6581), 364–366 (1996).
66. J.C. Venter, M.D. Adams, G.G. Sutton, A.R. Kerlavage, H.O. Smith, M. Hunkapiller, 'Shotgun Sequencing of the Human Genome', *Science*, **280**(5369), 1540–1542 (1998).
67. H. Shizuya, B. Birren, U. Kim, V. Mancino, T. Slepak, Y. Tachiiri, M. Simon, 'Cloning and Stable Maintenance of 300-kilobase-pair Fragments of Human DNA in *E. coli* Using F-factor-based Vector', *Proc. Natl. Acad. Sci. USA*, **89**(18), 8794–8797 (1992).
68. J.P. Carulli, M. Artinger, P.M. Swain, C.D. Root, L. Chee, C. Tulig, J. Guerin, M. Osborne, G. Stein, J. Lian, P.T. Lomedico, 'High Throughput Analysis of Differential Gene Expression', *J. Cell Biochem. Suppl.*, **30–31**, 286–296 (1998).
69. M.D. Adams, J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno, A.R. Kerlavage, W.R. McCombie, J.C. Venter, 'Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project', *Science*, **252**(5013), 1651–1656 (1991).
70. V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, 'Serial Analysis of Gene Expression', *Science*, **270**(5235), 484–487 (1995).
71. S. Welle, K. Bhatt, C.A. Thornnton, 'Inventory of High-Abundance mRNAs in Skeletal Muscle of Normal Men', *Genome Res.*, **9**(5), 506–513 (1999).
72. V. de Waard, B.M. van den Berg, J. Veken, R. Schultz-Heienbrok, H. Pannekoek, A.J. van Zonneveld, 'Serial Analysis of Gene Expression to Assess the Endothelial Cell Response to an Atherogenic Stimulus', *Gene*, **226**(1), 1–8 (1999).
73. K. Hibi, Q. Liu, G.A. Beaudry, S.L. Madden, W.H. Westra, S.L. Wehage, S.C. Yang, R.F. Heitmiller, A.H. Bertelsen, D. Sidransky, J. Jen, 'Serial Analysis of Gene Expression in Non-small Cell Lung Cancer', *Cancer Res.*, **58**(24), 5690–5694 (1998).
74. S.M. Wang, J.D. Rowley, 'A Strategy for Genome-wide Gene Analysis: Integrated Procedure for Gene Identification', *Proc. Natl. Acad. Sci. USA*, **95**(20), 11 909–11 914 (1998).

75. N.A. Datson, J. van der Perk de Jong, M.P. van den Berg, E.R. de Kloet, E. Vrougdenhil, 'MicroSAGE: A Modified Procedure for Serial Analysis of Gene Expression in Limited Amounts of Tissue', *Nucleic Acids Res.*, **27**(5), 1300–1307 (1999).
76. P. Liang, A.B. Pardee, 'Differential Display of Eukaryotic Messenger RNA by Means of the Polymerase Chain Reaction', *Science*, **257**(5072), 967–971 (1992).
77. N.K. Kochetkov, E.I. Budovsky, V.P. Demushkin, M.F. Turchinsky, N.A. Simukova, E.D. Sverdlov, 'The Chemical Modification of Nucleic Acids. I. The Preparation of Deuridylic RNA's', *Biochim. Biophys. Acta*, **142**(1), 35–46 (1967).
78. R. Waldmann, H.J. Gross, G. Krupp, 'Protocol for Rapid Chemical RNA Sequencing', *Nucleic Acids Res.*, **15**(17), 7209 (1987).
79. B. Gross, A. Rosental, 'Simultaneous Sequencing of RNA by Solid-phase Chemical Degradation Using DE 81 Anion-exchange Paper', *Gene Anal. Tech.*, **4**(3), 57–61 (1987).
80. Y. Zhang, W.Y. Liu, Y.X. Feng, T.P. Wang, 'A Simple and Rapid Solid-phase RNA Sequencing Method', *Anal. Biochem.*, **163**(2), 513–516 (1987).
81. R.C. Gupta, K. Randerath, 'Use of Specific Endonuclease Cleavage in RNA Sequencing', *Nucleic Acids Res.*, **4**(6), 1957–1978 (1977).
82. R.C. Gupta, K. Randerath, 'Use of Specific Endonuclease Cleavage in RNA Sequencing – an Enzymic Method for Distinguishing Between Cytidine and Uridine Residues', *Nucleic Acids Res.*, **4**(10), 3441–3454 (1977).
83. R.E. Lockard, B. Alzner-Deweerd, J.E. Heckman, J. MacGee, M.W. Tabor, U.L. RajBhandary, 'Sequence Analysis of 5′[^{32}P]-Labeled mRNA and tRNA Using Polyacrylamide Gel Electrophoresis', *Nucleic Acids Res.*, **5**(1), 37–56 (1978).
84. H. Donnis-Keller, A.M. Maxam, W. Gilbert, 'Mapping Adenines, Guanines, and Pyrimidines in RNA', *Nucleic Acids Res.*, **4**(8), 2527–2538 (1977).
85. M.S. Boguski, P.A. Hieter, C.C. Levy, 'Identification of a Cytidine-specific Ribonuclease from Chicken Liver', *J. Biol. Chem.*, **255**(5), 2160–2163 (1980).
86. A.M. Mazo, T.D. Mashkova, T.A. Avdonina, N.S. Ambartsumyan, L.L. Kisselev, 'An Improved Rapid Enzymatic Method of RNA Sequencing Using Chemical Modification', *Nucleic Acids Res.*, **7**(8), 2469–2482 (1979).
87. F.R. Kramer, D.R. Mills, 'RNA Sequencing with Radioactive Chain-terminating Ribonucleotides', *Proc. Natl. Acad. Sci. USA*, **75**(11), 5334–5338 (1978).
88. J.D. Parvin, F.I. Smith, P. Palese, 'Rapid RNA Sequencing Using Double-stranded Template DNA, SP6 polymerase, and 3′-Deoxynucleotide Triphosphates', *DNA*, **5**(2), 167–171 (1986).
89. V.D. Axelrod, R.M. Vartikyan, V.A. Aivazashvili, R.S. Beabealashvili, 'Specific Termination of RNA Polymerase Synthesis as a Method of RNA and DNA Sequencing', *Nucleic Acids Res.*, **5**(10), 3549–3563 (1978).
90. G.G. Brownlee, E.M. Cartwright, 'Rapid Gel Sequencing of RNA by Primed Synthesis with Reverse Transcriptase', *J. Mol. Biol.*, **114**(1), 93–117 (1977).
91. C.S. Hahn, E.G. Strauss, J.H. Strauss, 'Dideoxy Sequencing of RNA Using Reverse Transcriptase', *Methods Enzymol.*, **180**, 121–130 (1989).
92. H. Agut, K. Kean, O. Fichot, J. Morasco, J.B. Flanegan, M. Girard, 'A Point Mutation in the Poliovirus Polymerase Gene Determines a Complementable Temperature-sensitive Defect of RNA Replication', *Virology*, **168**(2), 302–311 (1989).
93. O. Fichot, M. Girard, 'An Improved Method for Sequencing of RNA Templates', *Nucleic Acids Res.*, **18**(20), 6162 (1990).
94. F.J. Grant, S.D. Levin, T. Gilbert, W. Kindsvogel, 'Improved RNA Sequencing Method to Determine Immunoglobulin mRNA Sequence', *Nucleic Acids Res.*, **15**(13), 5496 (1987).
95. G.J. Bauer, 'RNA Sequencing Using Fluorescent-labeled Dideoxynucleotides and Automated Fluorescent Detection', *Nucleic Acids Res.*, **18**(4), 879–884 (1990).
96. J.M. Bailey, M. Davidson, 'Methylmercury as a Reversible Denaturing Agent for Agarose Gel Electrophoresis', *Anal. Biochem.*, **70**(1), 75–85 (1976).
97. E. Shimomaye, M. Salvato, 'Use of Avian Myeloblastosis Virus Reverse Transcriptase at High Temperature for Sequence Analysis of Highly Structured RNA', *Gene Anal. Tech.*, **6**(2), 25–28 (1989).
98. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**(20), 2299–2301 (1988).
99. F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**(24), 1193A–1203A (1991).
100. F. Kirpekar, E. Nordhoff, L.K. Larsen, K. Kristiansen, P. Roepstorff, F. Hillenkamp, 'DNA Sequence Analysis by MALDI Mass Spectrometry', *Nucleic Acids Res.*, **26**(11), 2554–2559 (1998).
101. N.I. Taranenko, S.L. Allman, V.V. Golovlev, N.V. Taranenko, N.R. Isola, C.H. Chen, 'Sequencing DNA Using Mass Spectrometry for Ladder Detection', *Nucleic Acids Res.*, **26**(10), 2488–2490 (1998).
102. D.A. Tolson, N.H. Nicholson, 'Sequencing RNA by a Combination of Exonuclease Digestion and Uridine Specific Chemical Cleavage Using MALDI-TOF', *Nucleic Acids Res.*, **26**(2), 446–451 (1998).
103. H. Koster, K. Tang, D.J. Fu, A. Braun, D. van den Boom, C.L. Smith, R.J. Cotter, C.R. Cantor, 'A Strategy for Rapid and Efficient DNA Sequencing by Mass Spectrometry', *Nat. Biotechnol.*, **14**(9), 1123–1128 (1996).
104. K.K. Murray, 'DNA Sequencing by Mass Spectrometry', *J. Mass Spectrom.*, **31**(11), 1203–1215 (1996).

105. J.H. Graber, C.L. Smith, C.R. Cantor, 'Differential Sequencing with Mass Spectrometry', *Genet. Anal.*, **14**(5–6), 215–219 (1999).
106. R.J. Driscoll, M.G. Youngquist, J.D. Baldeschwieler, 'Atomic-scale Imaging of DNA Using Scanning Tunnelling Microscopy', *Nature*, **346**(6281), 294–296 (1990).
107. M. Ronaghi, M. Uhlen, P. Nyren, 'A Sequencing Method Based on Real-time Pyrophosphate', *Science*, **281**(5375), 363–365 (1998).
108. R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, W.K. Funkhouser, B. Koop, L. Hood, 'DNA Sequence Determination by Hybridization: A Strategy for Efficient Large-scale Sequencing', *Science*, **260**(5114), 1649–1652 (1993).
109. S. Drmanac, D. Kita, I. Labat, B. Hauser, C. Schmidt, J.D. Burczak, R. Drmanac, 'Accurate Sequencing by Hybridization for DNA Diagnostics and Individual Genomics', *Nat. Biotechnol.*, **16**(1), 54–58 (1998).
110. E.M. Southern, 'DNA Chips: Analysing Sequence by Hybridization to Oligonucleotides on a Large Scale', *Trends Genet.*, **12**(3), 110–115 (1996).
111. M.J. O'Donnell-Maloney, C.L. Smith, C.R. Cantor, 'The Development of Microfabricated Arrays for DNA Sequencing and Analysis', *Trends Biotechnol.*, **14**(10), 401–407 (1996).
112. K.R. Khrabko, Yu.P. Lysov, A.A. Khorlin, I.B. Ivanov, G.M. Yershov, S.K. Vasilenko, V.L. Florentiev, A.D. Mirzabekov, 'A Method for DNA Sequencing by Hybridization with Oligonucleotide Matrix', *DNA Seq.*, **1**(6), 375–388 (1991).
113. S.P.A. Fodor, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, D. Solas, 'Light-directed, Spatially Addressable Parallel Chemical Synthesis', *Science*, **251**(4995), 767–773 (1991).
114. R. Drmanac, Z. Strezoska, I. Labat, S. Drmanac, R. Crkvenjakov, 'Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides', *DNA Cell Biol.*, **9**(7), 527–534 (1990).
115. A.V. Fotin, A.L. Drobyshev, D.V. Proudnikov, A.N. Perov, A.D. Mirzabekov, 'Parallel Thermodynamic Analysis of Duplexes on Oligodeoxynucleotide Microchips', *Nucleic Acids Res.*, **26**(6), 1515–1521 (1998).
116. M.A. Livshits, A.D. Mirzabekov, 'Theoretical Analysis of the Kinetics of DNA Hybridization with Gel-immobilized Oligonucleotides', *Biophys. J.*, **71**(5), 2795–2801 (1996).
117. Y.P. Lysov, F.N. Gnuchev, A.A. Mironov, A.A. Chernyi, K.L. Beattie, A.D. Mirzabekov, 'Efficiency of Sequencing by Hybridization on Oligonucleotide Matrix Supplemented by Measurement of the Distance Between DNA Segments', *DNA Seq.*, **6**(2), 65–73 (1996).
118. N.E. Broude, T. Sano, C.L. Smith, C.R. Cantor, 'Enhanced DNA Sequencing by Hybridization', *Proc. Natl. Acad. Sci. USA*, **91**(8), 3072–3076 (1994).
119. D.J. Fu, N.E. Broude, H. Koster, C.L. Smith, C.R. Cantor, 'A DNA Sequencing Strategy that Requires Only Five Bases of Known Terminal Sequence for Priming', *Proc. Natl. Acad. Sci. USA*, **92**(22), 10 162–10 166 (1995).
120. J.G. Hacia, W. Makalowski, K. Edgemon, M.R. Erdos, C.M. Robbins, S.P.A. Fodor, L.C. Brody, F.S. Collins, 'Evolutionary Sequence Comparisons Using High-density Oligonucleotide Arrays', *Nat. Genet.*, **18**(1), 155–158 (1998).
121. J.D. Parsons, E. Buehler, L. Hillier, 'DNA Sequence Chromatogram Browsing Using JAVA and COBRA', *Genome Res.*, **9**(3), 277–281 (1999).
122. M.A. Andrade, N.P. Brown, C. Leroy, S. Hoersch, A. de Daruvar, C. Reich, A. Franchini, J. Tamames, A. Valencia, C. Ouzounis, C. Sander, 'Automated Genome Sequence Analysis and Annotation', *Bioinformatics*, **15**(5), 391–412 (1999).

Structural Analysis of Ribozymes

Sirinart Ananvoranich and Jean-Pierre Perreault
Université de Sherbrooke, Sherbrooke, Canada

1 Introduction 1
2 History 2
3 Synthesis of RNA for Ribozyme Studies 2
3.1 In Vitro Transcription 2
3.2 Chemical Synthesis 4
3.3 Combination Method 4
4 Primary Structural Information 5
4.1 Sequence Mapping and Self-cleaving Activity 5
4.2 Construction of Ribozymes and Their Variants 6
5 Secondary Structure Analysis 6
5.1 Computer-aided Prediction 6
5.2 Nuclease Mapping or Enzymatic RNA Sequencing 6
5.3 Chemical Interference 7
6 Native or Tertiary Structure Analysis 8
6.1 Cleavage Assay 8
6.2 Metal Ion Dependence 9
6.3 Kinetic Determination 9
6.4 Three-dimensional Structural Information 10
6.5 Data Analysis 10
7 Conclusion 10
Acknowledgments 11
Abbreviations and Acronyms 11
Related Articles 11
References 11

We use delta *ribozymes as a model in illustrating the various steps involved in the structural analysis of ribozymes.* Delta *ribozymes were originally identified as a self-cleaving motif located on the single-stranded RNA genome of hepatitis* delta *virus (HDV). The methodologies are presented in the order of their requirement for determination of the structure/function relationship of a ribozyme, beginning with the preparation of the RNA molecules required for the studies. Primary structural information is needed in order to identify potential ribozymes, then secondary and tertiary structural information is required for ribozyme characterization. The manuscript was written to guide an investigator from the initial observation of RNA catalytic activity (a ribozyme discovery) to the deduction of a structural model of the ribozyme.*

1 INTRODUCTION

Ribozymes are a family of RNA molecules which possess various catalytic capabilities.[1–3] The best known ribozymes are the endoribonucleases which are capable of cleaving RNA molecules at specific sequences. This group of ribozymes has tremendous potential for the development of a novel approach for the selective inactivation of specific RNA molecules, including both those derived from pathogenic viruses and those associated with inherited diseases. The underlying theory of this inactivation is that the mRNA encoding a harmful protein would be intercepted and destroyed by the ribozymes before that mRNA is translated. In order to be able to fully exploit the potential of a ribozyme, it is crucial to have a complete understanding of the molecular mechanism of the ribozyme reaction. It has long been known that modification of RNA (ribozyme) structure often results in the alteration of its catalytic properties.

The various methodologies described in this chapter are presented in the order of their requirement for determination of the structure/function relationship of a ribozyme, beginning with the preparation of the RNA molecules required for the studies. Primary structural information is needed in order to identify potential ribozymes, then secondary and tertiary structural information are required for ribozyme characterization. The procedures described below are written so as to guide an investigator from the initial observation of RNA catalytic activity (a ribozyme discovery) to the deduction of a structural model of the ribozyme.

We use *delta* ribozymes as a model in illustrating the various steps involved in the structural analysis of ribozymes. *Delta* ribozymes were originally identified as a self-cleaving motif located on the single-stranded RNA genome of HDV.[4] The HDV genome is circular and replicates through a rolling circle mechanism involving only RNA intermediates.[5] Replication is initiated by the binding of the host RNA polymerase II to the parental RNA and results in the synthesis of a complementary multimeric copy of the HDV genome (Figure 1). The monomeric HDV genome is then released from this multimer by the catalytic activity of self-cleaving motif. This intramolecular cleavage is an RNA autocatalytic reaction. The RNA sequence responsible for the cleavage is known as the *delta*-cleaving RNA motif, or as the *cis*-acting ribozyme (Figure 1b). Based on subsequent

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

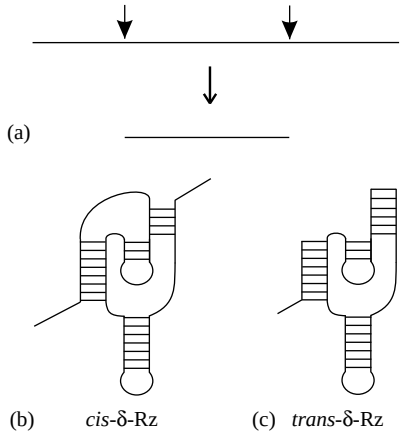

Figure 1 The discovery and development of *delta* ribozyme from HDV. (a) The single-stranded RNA genome of HDV is produced in a multimeric unit. Its self-cleaving motif is responsible for the release of multiple monomeric copies of the HDV genome, which are then re-circularized. (b) The *delta*-cleaving motif was subsequently identified. (c) From secondary structural analysis, a *trans*-acting *delta* ribozyme was derived by separating the junction between the P1 and P2 stems.

secondary structural information obtained by various investigators,[6–9] this motif has been modified into an intermolecular system (Figure 1c). This so called a *trans*-acting ribozyme system contains both a substrate molecule and an enzyme molecule.

2 HISTORY

RNA catalysis was first identified in the RNA components of both the group I intron by Cech[10] and RNaseP by Altman[11] in the early 1980s. These discoveries marked the first examples of enzymatic catalysis in the absence of any protein. These catalytically active ribonucleic acids were named ribozymes (RNA enzymes). During the past two decades several ribozymes have been identified that possess various catalytic capabilities which enable them to modify the phosphodiester bonds of their substrates. With the development of in vitro selection procedures, RNA molecules which catalyze a spectrum of reactions, for example phosphorylation and ligation, have been identified.[1–3] Without doubt in the decades to come, additional capabilities of ribozymes will be discovered.

3 SYNTHESIS OF RNA FOR RIBOZYME STUDIES

Ribozymes and their substrates can be produced either by in vitro transcription, or by chemical synthesis. The basic requirements and protocols for these two methods are described below.

3.1 In Vitro Transcription

This protocol makes use of enzymatic reactions catalyzed by purified bacteriophage T7 RNA polymerase, which uses DNA as a template. It has been demonstrated that the sequence immediately downstream of the T7 RNA promoter can affect the yield of transcript, and that the initiating nucleotides CCC or CCU give the best yield.[12] Large RNAs are routinely generated by this method. The model substrate of *delta* ribozyme, an oligomer 11 nt long, is also produced by this method (Figure 2a). Due to the propensity of T7 RNA polymerase to add one or two uncoded nucleotides to the 3′-end of the resultant transcripts,[1] the subsequent purification and verification of both the length and the sequence of the transcripts is required (section 5.2). The main limitation of this method is that it cannot be used when specific modifications are required.

3.1.1 Oligonucleotide Templates

A pair of synthetic DNA oligonucleotide templates can be designed so that one contains the complementary sequences of the T7 RNA promoter and the sequence coding for ribozyme or substrate, while the other contains the sequence of the T7 RNA promoter. Prior to the preparation of an in vitro transcription reaction mixture, the two oligonucleotides (500 pmol each) are mixed in diethylene pyrocarbonate (DEPC)-treated water (20 μL) containing 10 mM Tris(hydroxymethyl)aminomethane (Tris)-HCl pH 7.5, 10 mM $MgCl_2$ and 50 mM KCl, heated at 95 °C for 5 min and allowed to cool slowly to 37 °C. The partial duplexes formed then serve as templates for RNA synthesis by T7 RNA polymerase (Figure 2a).

3.1.2 Cloned Templates

The DNA template of *delta* ribozyme was cloned into the plasmid pUC19 using recombinant DNA techniques (section 4.2). The resultant recombinant plasmid containing the ribozyme sequence is then digested so as to either linearize or release the *delta* ribozyme insert. T7 RNA polymerase will use the resultant DNA duplex as a template and produce transcripts extending until the end of the duplex, so called run-off transcription reactions[13] (Figure 2b).

3.1.2.1 Materials and Methods Mix, in a final volume of 100 μL:

DNA template either as a partial duplex (500 pmol) or as a digested cloned template (5 μg)	20 μL

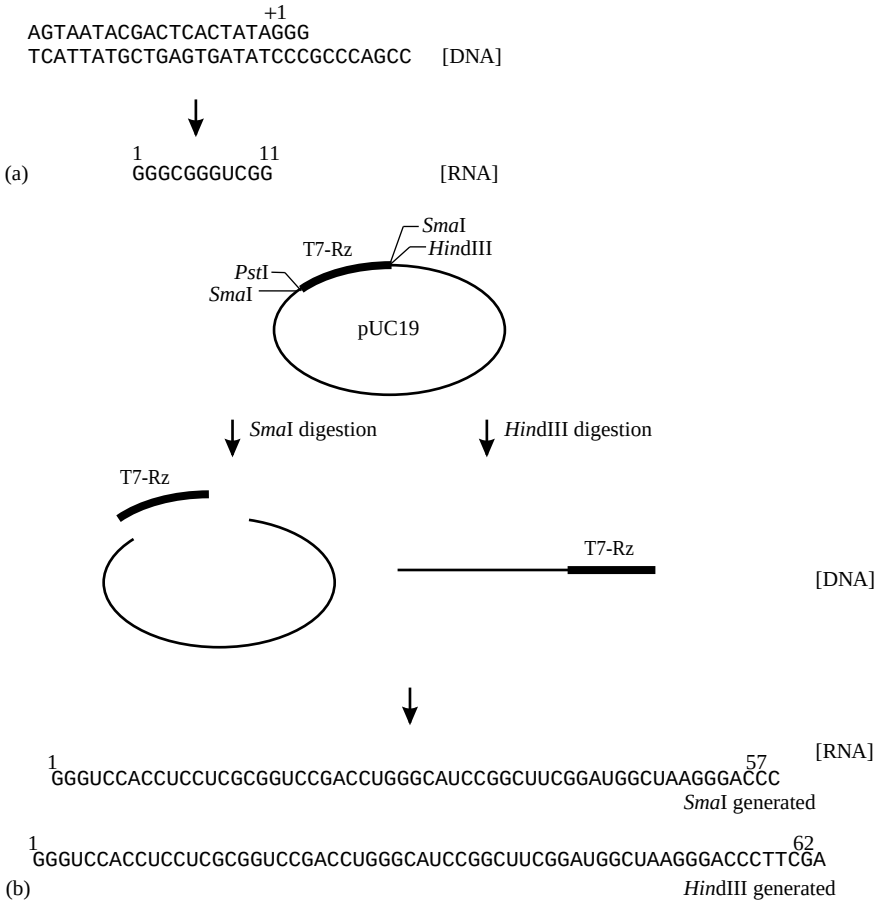

Figure 2 DNA templates for in vitro transcription reactions catalyzed by bateriophage T7 RNA polymerase. (a) The partial duplex formed by two oligonucleotides. The substrate (11 nt) of *delta* ribozyme is illustrated. (b) The double-stranded DNA. The plasmid pUC19 harboring the sequence of *trans* acting *delta* ribozyme is digested with either *Sma*I or *Hin*dIII and the resulting linear DNA is used as a template. RNA and DNA molecules are identified in brackets.

RNAGuard® RNase inhibitor (Pharmacia)	27 units
rNTP (Pharmacia) 10 mM each,	5 μL
Transcription buffer (5×)	20 μL
(400 mM *N*-(2-hydroxyethyl)piperazine-*N*′-ethanesulfonic acid (HEPES)-KOH pH 7.5, 120 mM $MgCl_2$, 10 mM spermidine) dithiothreitol (DTT) (100 mM)	2 μL
Pyrophosphatase (1 unit μL^{-1}, Boehringer Mannheim)	1 μL
Purified T7 RNA polymerase (2 μg μL^{-1})	2 μL

Incubate at 37 °C for 2–4 h. Add 5 μL RQ1 RNase-free DNase (1 unit μL^{-1}, Promega) to eliminate the DNA template, then extract twice with buffered phenol. Add 0.1 volume of 3 M sodium acetate pH 5.2 and 2 volume of ethanol to the aqueous phase. Chill for 15 min at −80 °C and centrifuge at 4 °C for 15 min in a microfuge. Discard the supernatant then add an equal volume of 70% ethanol and repeat the centrifugation step. Dissolve the pellet in 20 μL of DEPC-treated water and add 10 μL of 5× gel loading buffer (95% formamide, 10 mM ethylenediaminetetraactetic acid (EDTA), 0.05% bromophenol blue and 0.05% xylene cyanol). The mixtures are fractionated by denaturing 20% polyacrylamide gel electrophoresis (PAGE, 19 : 1 ratio of acrylamide to bisacrylamide) containing 45 mM Tris-borate pH 7.5, 7 M urea and 1 mM EDTA. The reaction products are visualized by ultraviolet (UV) shadowing, and the bands corresponding to the correct sizes of the ribozymes and substrates cut out, and the transcripts eluted overnight at 4 °C in a solution (200 μL) containing 0.1% SDS and 0.5 M ammonium acetate. The transcripts are then precipitated by the addition of 0.1 volume of 3 M sodium acetate pH 5.2 and 2.2 volume of ethanol. Transcript yield is determined by spectrophotometry at 260 nm. The sequence and size of the RNA products are then verified by RNase digestion (section 5.2).

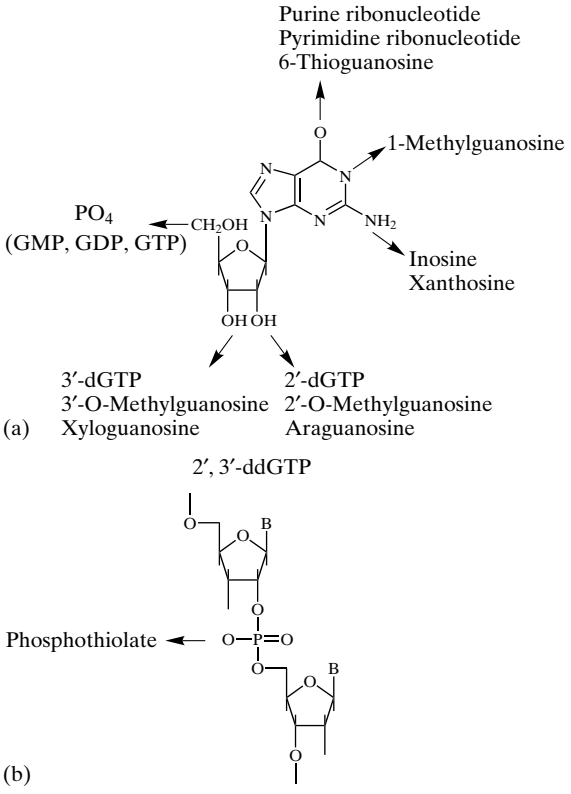

Scheme 1 Specific modifications of an RNA molecule.[32] (a) Some modifications on the mononucleotide can be on the sugar or the base residue. (b) The phosphodiester backbone can be modified to a phosphothiolate group.

3.2 Chemical Synthesis

RNA molecules can be chemically synthesized in the same fashion as DNA molecules from 3′ to 5′ by taking the advantage of the high chemical reactivity of the 5′-hydroxyl group. It has been reported that the enzymatic activity of the resultant RNA is equivalent to that of RNA derived by transcription methods.[14] Most importantly, chemical synthesis allows the introduction of single-atom substitutions at specific positions in the RNA molecules. There are three major targets for modification in RNAs: the exocyclic base, the sugar, and the internucleotide phosphodiester linkage (Scheme 1). Presently the cost of production is still prohibitive, especially when large RNA molecules are required. An improved method using the solid-phase chemistry is capable of production in the 200 μmol range; however the chemical synthesis of RNAs longer than 45 nt is not advised when 2′-*O*-*t*-butyldimethylsilyl-5′-*O*-DMT-ribonucleosides (tBDMS-amidites) are used, and of not more than 90 nt when 2′-*O*-triisopropylsilyl-oxymethyl-ribonucleosides (TOM-amidites) are used.[16] In some cases where longer RNA molecules are required, the investigator might combine T7 transcription and chemical synthesis (section 3.3) in order to produce the desired RNA.

An automated DNA/RNA synthesizer (for example, ABI Model 392 or 394 from Applied Biosystem, Foster City, CA) and high-performance liquid chromatography (HPLC) system are required for the synthesis of RNA oligonucleotides using tBDMS-amidites phosphoramidites or TOM-amidites, in addition to nucleoside-functionalized CPG or polystyrene supports (A, G, C, U). Due to the special amidites and deprotection reagents currently available, various modifications can be incorporated into the synthesized RNA. For example, inosine can be introduced at selective bases. 2′-deoxy or 2′-O-methyl RNA can be substituted for the 2′-OH of the ribose residue. The internucleotide phosphodiester bonds can be replaced by phosphothioate or phosphonate linkages. Moreover, other modifications, such as biotinylated RNA and halogenated deoxy and ribonucleotide incorporation, aimed at facilitating the downstream use of the resulting RNAs can also be incorporated.

Due to the high cost of equipment and materials in general, investigators are advised to order custom-made RNA oligonucleotides from facilities such as Keck Oligonucleotide Synthesis Facility (Yale University, CT), and Xeragon AG (Zurich, Switzerland). Following their synthesis RNA oligonucleotides are subjected to deprotection and desalting steps analogous to those for synthetic DNA oligonucleotides. The deprotection removes protecting groups from 2′-hydroxyl of the ribose residue using either TBAF or $Et_3N(HF)_3$ according to the manufacturer's instruction;[15] the subsequent desalting step removes inorganic salts, trace organic compounds, low-molecular-weight impurities and short failure sequences. The latter step can be performed using size-exclusion chromatographic columns such as G-25 Sephadex or reverse phase HPLC.[16] However, gel electrophoresis is the method of choice for removal of failure sequences. The sequence and size of the RNA products are verified by RNase digestion (section 5.2).

3.3 Combination Method

Both enzymatic and chemical techniques can be combined in order to obtain target RNA molecules with site-specific modifications at affordable prices. For example, target RNA molecules with a specific phosphothiolate linkage isomer can be generated by enzymatic ligation of two individual RNAs, one of which is a chemically synthesized RNA containing a phosphothiolate linkage at the desired position. The resultant Sp and Rp isomers are then separated by HPLC. The desired RNA isomer is then joined to the other RNA by T4 RNA ligase.[17] For more examples and details on using a combination of

chemical and enzymatic methods to generate the RNAs of interest, see Vinyak et al.[14]

4 PRIMARY STRUCTURAL INFORMATION

The observation of self-cleaving RNA molecules naturally lead to experiments aimed at identifying the responsible element whether its sequence or surroundings causes the cleavage. In the case of *delta* ribozymes, a linear dimer of the HDV genome was reported to be initially processed into a monomeric RNA when the responsible cDNA was transfected into a monkey kidney cell line.[4] These findings suggested that either HDV RNA has an unusual secondary structure that allows a specific attack by a cellular RNase, or that the HDV RNA has a specific self-cleaving activity. The self-cleaving motif of HDV was mapped on both strands of the HDV genome in order to locate the cleavage position.[6] The steps required for this identification are described below.

4.1 Sequence Mapping and Self-cleaving Activity

In general, cDNA coding for the RNA molecules of interest is generated and cloned into a plasmid for identification using standard recombinant DNA techniques. Exonuclease III, a 3′ to 5′ exonuclease specific for double-stranded DNA carrying either a blunt end, a 5′-overhang, or nick, is commonly used for the construction of unidirectional nested deletion sets from the plasmids. The nested deletion clones can be generated from either end using an appropriate restriction enodonuclease that leaves either a blunt end or 5′-overhang to linearize the DNA, followed by exonuclease III treatment. The sequence of each of the resulting deletions is determined by dideoxynucleotide chain-termination sequencing.[13]

RNA transcripts are then produced by in vitro transcription reactions in the presence of 0.05 mCi [α-^{32}P]GTP, and are purified and subjected to various buffered conditions so as to identify the responsible element. To prove that the self-cleaving activity is solely the result of the RNA transcripts, and not due to RNase or proteins present in the in vitro transcription reaction or cell extracts, the purified primary transcripts are incubated in 50 mM Tris-HCl, pH 8.0 and 10 mM $MgCl_2$ (standard conditions for many ribozyme cleavage assays, including *delta* ribozymes).[6,18] The newly formed products are resolved on denaturing polyacrylamide gels, and are visualized by exposure of the gels to either X-ray films or phosphorImaging™ screens.

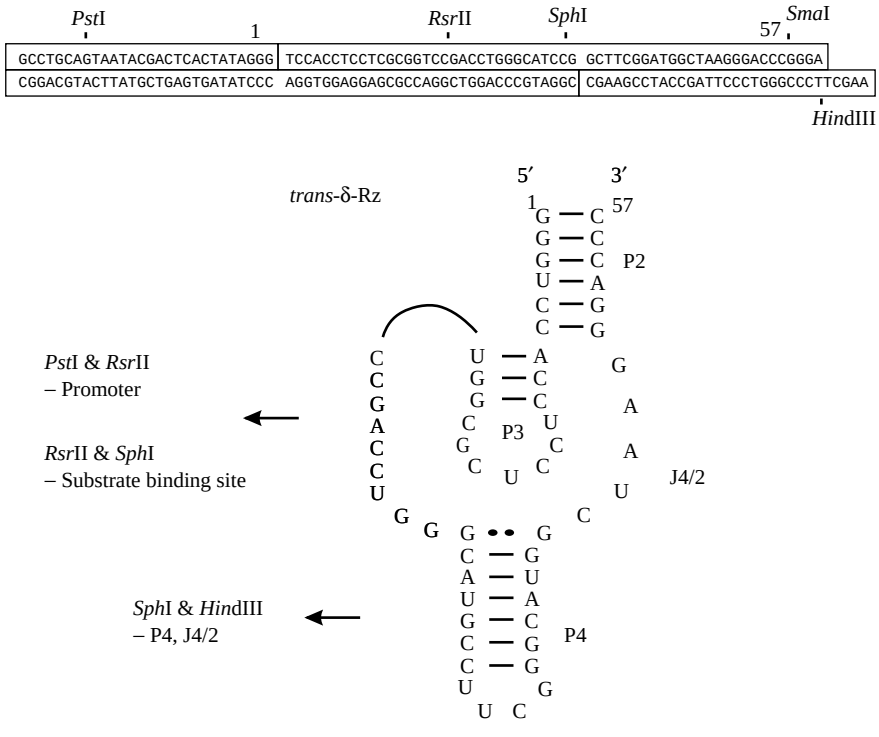

Figure 3 Cloning of the *trans*-acting *delta* ribozyme. Four overlapping oligonucleotides were designed so that they encode the entire sequence of *trans*-acting *delta* ribozyme. Following annealing of these four oligonucleotides, the resulting fragment was cloned into *Pst*I/*Hin*dIII-digested pUC19. The resultant clone was verified by dideoxynucleotide sequencing. The restriction endonuclease sites are identified on the double-stranded DNA.

To characterize the biochemical properties of the self-cleaving motif, the cleavage efficiency is determined in various buffer systems (pH 5.0–9.0) containing either monovalent (Na^+, K^+ or NH^{4+}) or divalent ions (Mg^{2+}, Ca^{2+} or Mn^{2+}). The *delta* ribozyme self-cleaving motif located on the HDV genome was found to be able to use either Mg^{2+} or Ca^{2+} as a cofactor for efficient cleavage in Tris-HCl pH 5.0–8.0.[6] The cleavage products were identified as a 3′ fragment with 5′-hydroxyl end, and a 5′ fragment with 2′,3′-cyclic phosphate terminus similar to the products of other ribozyme cleavage reactions.

4.2 Construction of Ribozymes and Their Variants

In order to explore enzymatic properties of a ribozyme, cDNA clones coding for the ribozymes are constructed using recombinant DNA techniques. From the primary structural information (i.e. the sequence and the cleavage site), deoxyoligonucleotides can be designed and synthesized for cloning purposes. The *delta* ribozymes used in our studies were initially constructed using four overlapping oligonucleotides (Figure 3). Several restriction sites were included in order to facilitate the creation of variant ribozymes. For example, ribozyme mutants carrying a single mutation in the P1 stem can be produced by digestion of the plasmid carrying the ribozyme with the restriction endonucleases, *Rsr*II and *Sph*I. Subsequent ligation of this predigested plasmid to new oligonucleotides having the altered sequence flanked by *Rsr*II and *Sph*I sites led to the production of the ribozyme variant.[19]

5 SECONDARY STRUCTURE ANALYSIS

Secondary structural information of RNA molecules is generally considered to be a simplification of what is in fact a three-dimensional complex. RNA secondary structure is predicted using either computer-aided alignment, or experimental data. Both approaches suggest putative base pairs in the three-dimensional structure of the RNA molecule.

5.1 Computer-aided Prediction

RNA secondary structure elucidation is similar to an alignment of protein or nucleic acid sequences, except that the RNA sequence folds back on to itself rather than on to identical or similar bases.[20] The complementary bases, G–C and A–U, form stable base pairs through hydrogen bonds between donor and acceptor sites on the bases, and are known as Watson–Crick base pairs. In addition, the weaker G–U wobble pair can be formed in a skewed fashion. These three types of base pairs are called canonical base pairs. Other base pairs (i.e. G–G or C–C) are called noncanonical base pairs.[21] The prediction of secondary structure can be made either from a single RNA sequence by minimizing the free energy of folding, or from a common folding pattern observed for a family of aligned, homologous RNAs. To calculate the free energy of RNA folding, arbitrary energy profiles are assigned to each individual base pair type and motif.[22] For the past decade, the formulae defined by Turner and his co-investigators have been widely used to define the free energy of stacking of canonical base pairs, hairpin loops, and both interior and bulge loops.[20,22] In general, the secondary structural information obtained using aligned RNA sequences is very valuable because the RNA structure is considered to be conserved to a greater degree than the sequence is. Over a period of sequence drift, the structural similarity might remain essentially the same through the phenomenon known as compensatory base changes, which conserve the base pairs,[22] as is observed in the structures of tRNAs or 5S RNA.

5.2 Nuclease Mapping or Enzymatic RNA Sequencing

Enzymatic RNA sequencing is generally used in both the mapping of RNA secondary structures and in the determination of RNA length. This method takes advantage of the cleavage specificity of a variety of ribonucleases and nucleases (Table 1) that are incubated with the substrate RNA in separate reactions. The reaction mixtures are then fractionated by denaturing gel electrophoresis. The resulting bands are compared to an RNA ladder generated from the same RNA. To determine the length of the RNA, several nucleases can be used (Table 1).

Table 1 Ribonucleases and nucleases that are commercially available, and are commonly used, are listed with their optimal buffer[23]

Nucleases	Cleavage	Buffers (5×)
CV or V1	Prefers double-stranded RNA	125 mM Tris-HCl, pH 7.2
Phy M	Ap↓N and Up↓N	250 mM sodium citrate, pH 5.0
S1	Single-stranded nucleic acid	200 mM sodium acetate, pH 4.5, 1 M NaCl, 50 mM $ZnSO_4$
T1	Single-stranded RNA and Gp↓N	250 mM sodium citrate, pH 5.0
T2	Prefers single-stranded RNA and Ap↓N	250 mM sodium acetate, pH 5.0 250 mM Tris-HCl, pH 7.5
U2	Single-stranded RNA and Ap↓N	50 mM sodium citrate, pH 4.5

5.2.1 Materials and Methods

5.2.1.1 5′-Dephosphorylation of RNAs Mix, in total volume of 10 μL:

RNA purified from in vitro transcription reaction (10 pmol μL^{-1})	2 μL
Calf intestine alkaline phosphatase (Pharmacia)	0.2 unit
1 M Tris-HCl, pH 8.0	5 μL
RNAGuard® RNase inhibitor (Pharmacia)	10 units
DEPC-treated water	

Incubate at 37 °C for 30 min. Extract twice with buffered phenol. Collect the aqueous phase, add 0.1 volume of 3 M sodium acetate pH 5.2 and 2 volume of ethanol. Chill the mixture for 15 min at −80 °C. Centrifuge at 4 °C for 15 min in a microfuge, add an equal volume of 70% ethanol and repeat the centrifugation step. Dissolve the pellet in 20 μL of DEPC-treated water to obtain 1 pmol μL^{-1} of dephosphorylated RNA.

5.2.1.2 5′-End-labeling of Transcripts Mix, in total of 10 μL:

Dephosphorylated RNA (1 pmol μL^{-1})	2 μL
10 × T4 polynucleotide kinase buffer (USB)	1 μL
RNAGuard® RNase inhibitor (Pharmacia)	10 units
γ-^{32}P ATP (Amersham, 10 μCi μL^{-1})	3.2 pmol
T4 polynucleotide kinase (USB)	3 units
DEPC-treated water	

Incubate at 37 °C for 30 min. Add 5 μL of loading buffer (95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol) and fractionate on a denaturing PAGE. The radioactively labeled RNA band is cut out following exposure of the gel to an X-ray film. The 5′-end-labeled RNA is eluted overnight at 4 °C in a solution (200 μL) containing 0.01% SDS and 0.5 M ammonium acetate.

5.2.1.3 Alkaline Hydrolysis Generation of an RNA Ladder Mix, in total of 5 μL, 5′-end-labeled RNA (50 000 cpm μL^{-1}) in solution containing 50 mM $NaHCO_3$ and 5 mM EDTA. Incubate at 95 °C for 5 min. Add 5 μL of loading buffer (95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol) and fractionate on a denaturing PAGE gel.

5.2.1.4 Nuclease Assay The specific cleavage by nucleases at the ribose-phosphate backbone are carried out using various buffering conditions as listed in Table 1. Frequently $MgCl_2$ or EDTA is also added in order to obtain conditions corresponding to either native or partially denatured folding. Note that the addition of heavy metal ions should be omitted in assays using RNase T2. Moreover, the optimal time required for the mapping is determined depending on the efficiency of each nuclease.[23] The cleavage reaction mixtures are fractionated on denaturing PAGE gels along with the corresponding RNA ladder.

5.3 Chemical Interference

Chemical reagents interact with the heterocyclic bases, the phosphodiester bonds and the ribose moieties resulting in a modified structure (Table 2). The modified residues are detected by primer extension,[2] except when using Fe (II)-EDTA and imidazoles where primer extension is not required in order to resolve the reaction products. Fe (II)-EDTA generates hydroxyl radicals which nonspecifically interact with nucleic acids. Consequently Fe (II)-EDTA is often used for elucidation of the surface residues of an RNA molecule. These interactions occur at the heterocyclic bases and ribose residues, the latter of which results in strand breaks which are detected following gel fractionation. Imidazole and its conjugates rapidly cleave the phosphodiester bonds located in single-stranded regions whereas those located in double-stranded regions are cleaved much more slowly.[23] For *delta* ribozymes, several research groups have used RNase mapping, chemical interference, and UV-cross-linking procedures to determine *delta* ribozyme structures.[24–26] Three versions of the secondary structure are illustrated in Figure 4. The differences result from the different assay conditions used and from the analysis of the mixed population of *delta* ribozymes present in aqueous solution.

Table 2 Chemical interference[23]

Chemicals	Attack at functional groups
Dimethyl sulfate	Nucleophilic centers of heterocyclic bases
Ethyl nitrourea	Internucleotide phosphates and nucleophilic centers of heterocyclic bases
CMCT	Uridine (N3), guanosine (N1)
Ketoxal	Guanosine in single-stranded regions
DEPC	Purine (N7)
Fe (II)-EDTA	Heterocyclic bases and ribose
Imidazole and conjugates	Phosphodiester bonds

CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metro-*p*-toluene sulfonate.

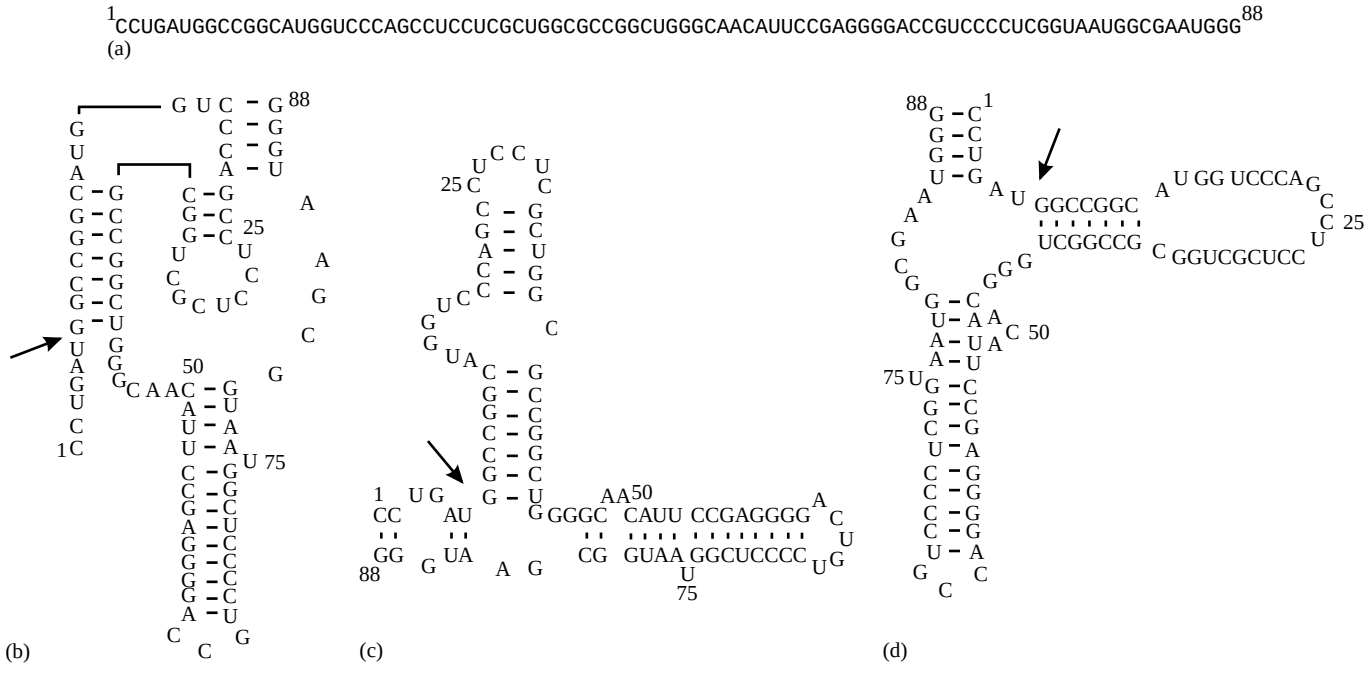

Figure 4 The primary and secondary structural information of genomic *delta* ribozymes. (a) The sequence of the genomic *cis*-acting *delta* ribozyme is 88 nt long.[33] (b–d) The secondary structures of *delta* ribozyme as determined by Perrotta and Been (b),[7] Wu et al. (c),[6] and Branch et al. (d).[8]

6 NATIVE OR TERTIARY STRUCTURE ANALYSIS

Using both intramolecular and intermolecular (i.e. with surrounding molecules) interactions, ribozymes adopt their native or tertiary structures and thereby gain their catalytic activity. The surrounding molecules used in the intermolecular interaction are metal ions, proteins and water. Over the last few years there has been considerable progress in our understanding of the kinetics of RNA folding and cleavage due to simultaneous advances in both experimental and theoretical methods.[27] Ribonuclease mapping, chemical interference and mutational analyses have been used to define the possible conformations of various RNA molecules. Most significantly, advanced techniques of crystallography, X-ray diffraction and NMR have determined precisely several RNA structures. Currently, the relationship between structure and function has been most thoroughly studied for the group I intron-derived ribozyme. Several of the approaches used in the study of this large ribozyme have also been applied to smaller ribozymes, the best studied of which are the hammerhead ribozymes. However, in some cases structural information from different methods resulted in discrepancies in the proposed structure–function relationship due to the heterogenous population of ribozyme present in aqueous solution. Kinetic characterization has been used widely in the survey of native structural analysis as it relates to enzymatic activity.

6.1 Cleavage Assay

A cleavage assay is the primary method of demonstrating that a ribozyme has adopted its native structure. Like protein enzymes, cleavage assays can be carried out in buffered solution in the presence of trace amounts of radioactive labeled substrate. The radioactive RNA molecules are either the *cis*- ribozyme (RNA carrying a self-cleaving motif), or the substrate for a *trans*-ribozyme. The cleavage reactions catalyzed by both *cis*- and *trans*-δ-Rz require metal ions as cofactors (i.e. $MgCl_2$, $CaCl_2$ and $MnCl_2$). Denaturing agents such as formamide and urea are sometimes included in the reaction mixture in order to disturb any misfolded molecules and thereby enhance the refolding of the ribozyme–substrate complexes.

6.1.1 Materials and Methods

Mix, in total of 20 μL, either radioactively labeled *cis*-acting ribozyme or radiactively labeled substrate of *trans*-acting ribozyme (ca. 50 000 cpm) in buffered solution containing 50 mM Tris-HCl, pH 7.5–8.0, and 5–50 mM $MgCl_2$. In the cases of *trans*-acting ribozyme, an approximate amount of ribozyme is added. Incubate the reaction for the time required (i.e. 10–30 min). Stop the reaction by adding 5 μL of loading buffer

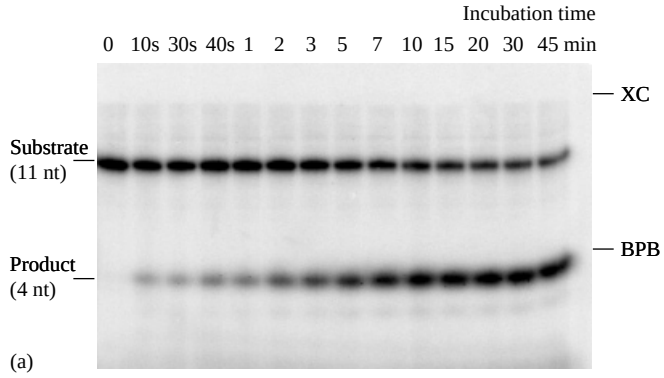

(b) Rz + S ↔ Rz–S ↔ Rz–P ↔ Rz + P

Figure 5 Kinetic analysis of *trans*-acting *delta* ribozyme. (a) An example of a time course experiment. Trace amounts of 5′-end-labeled substrate were incubated with 100 nM *delta* ribozyme in a solution containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$. (b) Simplified kinetic pathway of the cleavage reaction catalyzed by *delta* ribozyme.

(95% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol) and fractionate on a denaturing PAGE (Figure 5a).

6.2 Metal Ion Dependence

In general, ribozymes require the presence of metal ions for their folding and cleavage activity. These metal ions (Mg^{2+}, Ca^{2+}, Mn^{2+}, Sr^{2+} etc.) are added to the cleavage reactions in either the presence or absence of monovalent ions (Na^+, K^+ or NH_4^+) in order to determine the metal ion requirement.

6.3 Kinetic Determination

Time course experiments are performed at various substrate and ribozyme concentrations in order to determine kinetic parameters, such as a maximum rate of cleavage and substrate association constant.

6.3.1 Single Turnover Conditions

Various amounts of ribozyme are mixed with trace amounts of substrate (final concentration < 1 nM) in a 18-μL reaction mixture containing 50 mM Tris-HCl pH 7.5, and are then subjected to denaturation by heating at 95 °C for 2 min. The mixtures are quickly placed on ice for 2 min and equilibrated to 37 °C for 5 min prior to initiation of the reaction. Cleavage is initiated by the addition of $MgCl_2$ to 10 mM final concentration. The cleavage reactions are incubated at 37 °C, and followed for 3.5 h or until the end-point of cleavage is reached. The samples are quenced by the addition of 5 μL stop solution (95% formamide, 10 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol), and are analyzed by 20% PAGE as described above. Both the substrate (11 nt) and the reaction product (4 nt) bands are detected using a molecular dynamic radioanalytic scanner after exposition of the gels to a phosphorImaging™ screen (Figure 4a).

6.3.1.1 Measurement of Pseudo-first-order Rate Constant (k_{cat}, K_M and k_{cat}/K_M) Kinetic analyses are performed under single turnover conditions as described by Hertel et al.[28] with some modifications for *delta* ribozyme. Trace amounts of end-labeled substrate (<1 nM) are cleaved by various ribozyme concentrations (5–500 nM), and the fraction cleaved is determined. The rate of cleavage (k_{obs}) is obtained from the fitting of the data to the equation $A_t = A_\infty(1 - e^{-kt})$, where A_t is the percentage of cleavage at time t, A_∞ is the maximum percent cleavage (or the end-point of cleavage), and k is the rate constant (k_{obs}). Each rate constant should be calculated from at least two measurements. The values of k_{obs} obtained are then plotted as a function of ribozyme concentration in order to determine the other kinetic parameters: k_{cat}, K_M and k_{cat}/K_M.

6.3.2 Multiple Turnover Conditions

Trace amounts of labeled substrate are mixed with unlabeled substrate in order to obtain various substrate final concentrations. Fixed amounts of ribozyme (50 nM) are then added to the substrate mixtures in an 18-μL reaction mixture containing 50 mM Tris-HCl pH 7.5, and then subjected to denaturation by heating at 95 °C for 2 min. Again the mixtures are quickly placed on ice for 2 min and equilibrated to 37 °C for 5 min prior to the initiation of the reaction. Cleavage is again initiated by the addition of $MgCl_2$ to 10 mM.

6.3.3 Inhibition Analysis

The inhibitory effects of the substrate and the product can be kinetically tested under both single- and multiple-turnover conditions.

6.3.3.1 Single Turnover Conditions Conditions similar to those described above can be used with various amounts of potential inhibitors (either substrate, product or oligonucleotide). The reactions are initiated by mixing inhibitors (0.5–20 μM) with substrate (1 nM) prior to addition of the ribozyme (50 nM) in 20 μL enzyme assay buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$). Aliquots (2 μl) are withdrawn at various intervals during the 40-min incubation period, and are quenched by the addition of ice-cold stop solution (6 μL). The samples are fractionated on 10% denaturing PAGE gels, and the reaction products are quantified following exposure to phosphorImaging™ screens. Control reactions are carried out in the absence

of inhibitor. To evaluate the effect of an inhibitor on the intrinsic rate of *delta* ribozyme cleavage, the data are analyzed as described by Clouet-d'Orval et al.[29] in order to determine the fraction of inhibition (I) at each inhibitor concentration. Equation (1) is used:

$$I = 1 - \left(\frac{k_{2_{\text{inhibitor}}}}{k_2} \right. \quad (1)$$

where $k_{2_{\text{inhibitor}}}$ and k_2 are the rates of cleavage in the presence and absence of the inhibitor, respectively. The values of k_2 are obtained from fitting the experimental data to the pseudo-first-order rate equation (Equation 2):

$$A_t = A_\infty(1 - e^{k2t}) \quad (2)$$

where A_t is the percentage of product formed at time t, and A_∞ is the maximum amount of product formed. The fraction of inhibition (I) is plotted versus inhibitor concentration, and the data fitted to a hyperbolic equation in order to obtain N_I, the inhibitor concentration needed to reduce the rate of cleavage by one-half.

6.3.3.2 Multiple-turnover or Steady-state Conditions
Various concentrations of substrate are mixed with trace amounts of end-labeled substrate (<1 nM) so that final concentration is between 75 and 500 nM in the reaction mixtures. The reaction mixtures contain substrate, ribozyme (50 nM), and inhibitor (0.5–20 μM), and are performed as described for single-turnover conditions. The cleavage rates (v_i, μM min^{-1}) are determined at various substrate and inhibitor concentrations. Lineweaver-Burk, or reciprocal plots of $1/v_i$ and $1/[S]$ at all inhibitor (I) concentrations, are plotted, and the slopes and intercepts calculated by weighted linear regression analyses.

6.4 Three-dimensional Structural Information

6.4.1 Binding Shift Assay

Nondenaturing electrophoresis is commonly used in resolving the isomers that result from the folding of ribozymes. Moreover, kinetic parameters, such as equilibrium dissociation constant (K_d), the association and dissociation of substrate and product (Figure 5b), can be determined by this assay.[19]

6.4.1.1 Materials and Methods The equilibrium dissociation constants can be determined as follows. Various ribozyme concentrations, ranging from 5 to 600 nM, are individually mixed with trace amounts of end-labeled substrate (<1 nM) in a 9-μL solution containing 50 mM Tris-HCl pH 7.5. This reaction is then heated at 95 °C for 2 min and cooled to 37 °C for 5 min prior to addition of $MgCl_2$ to a final concentration of 10 mM in a manner similar to that of a regular cleavage reaction. The reactions are incubated at 37 °C for 1.5 h, at which point 2 μL of sample loading solution (50% glycerol, 0.025% of each bromophenol blue and xylene cyanol) is added and the resulting mixture electrophoresed through a nondenaturing polyacrylamide gel (20% acrylamide with a 19 : 1 ratio of acrylamide to bisacrylamide, in 45 mM Tris-borate pH 7.5 and 10 mM $MgCl_2$ buffer system). The gels are pre-run at 20 W for 1 h prior to sample loading, while the actual electrophorsis is carried out at 15 W for 4.5 h at room temperature. Quantification of bound and free substrate molecules is performed following exposure of the gels to a phosphorImaging™ screen.

6.4.2 In Vitro *Evolution*

An in vitro selection procedure can be used for the identification of important residues in the structure of *delta* ribozymes.[30] A pool of *trans*-acting ribozymes with sequences that have been randomized are produced from synthetic DNA templates using T7 RNA polymerase. In each selection cycle, the inactive ribozyme–substrate complexes can be separated from the active complexes using a biotinylated substrate against avidin. Using this procedure, the nucleotides essential for maintenance of the activity of *trans*-acting *delta* ribozymes were identified.[30]

6.5 Data Analysis

Delta ribozyme is able to catalyze the cleavage of an 11-mer RNA substrate under both single- and multiple-turnover conditions.[18–19,31] In both cases only small differences in the kinetic parameters are observed in the presence of either magnesium or calcium as cofactor. Under multiple-turnover conditions, the catalytic efficiency of the ribozyme (k_{cat}/K_M) is higher at 37 °C than at 56 °C.[17] The cleavage reaction seems to be limited by the product release step at 37 °C, and by the chemical cleavage step at 56 °C. Substrate inhibition is detected at high concentrations of the 11-mer substrate. The 2′-hydroxyl group adjacent to the scissile phosphate is involved in binding with the ribozyme, while the essential cytosine residue of the J4/2 junction has been shown to contribute to substrate association. The kinetic pathway of *delta* ribozyme is believed to involve a conformational transition step that is essential for the formation of the active ribozyme–substrate complex (Figure 5b).

7 CONCLUSION

In addition to fundamental methods for the ribozyme structural analysis, several advanced protocols are being

developed for determination of the structure–function relationship of ribozymes. For example, X-ray crystallography, nuclear magnetic resonance and in vitro selection of RNA molecules have provided the scientific community with considerable information. We believe that these topics should be reviewed by experts in those particular domains.

ACKNOWLEDGMENTS

The authors thank the Medical Research Council of Canada (MRC) and the Natural Sciences and Engineering Research Council of Canada (NSERC) for providing operational and supporting grants, respectively. S.A. was recipient of an NSERC postdoctoral fellowship, and J.P.P is an MRC scholar.

ABBREVIATIONS AND ACRONYMS

CMCT	1-cyclohexyl-3-(2-morpholino-ethyl)carbodiimide metro-*p*-toluene sulfonate
DEPC	Diethylene Pyrocarbonate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraactetic Acid
HDV	Hepatitis *Delta* Virus
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*′-ethanesulfonic Acid
HPLC	High-performance Liquid Chromatography
PAGE	Polyacrylamide Gel Electrophoresis
tBDMS-amidites	2′-*O*-*t*-butyldimethylsilyl-5′-*O*-DMT-ribonucleosides
TOM-amidites	2′-*O*-triisopropylsilyl-oxy-methyl-ribonucleosides
Tris	Tris(hydroxymethyl)amino-methane
UV	Ultraviolet

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
Aptamers • Mass Spectrometry of Nucleic Acids • Nuclear Magnetic Resonance and Nucleic Acid Structures • Nucleic Acid Structural Energetics • RNA Tertiary Structure • Sequencing Strategies and Tactics in DNA and RNA Analysis • X-ray Structures of Nucleic Acids

REFERENCES

1. F. Eckstein, D.M.J. Lilley, *10 Nucleic Acids and Molecular Biology: Catalytic RNA*, Springer, New York, 1996.
2. R.E. Gesteland, J.F. Atkins, *The RNA World*, Cold Spring Harbor Laboratory Press, Plainview, NY, 1993.
3. K.L. Scanlon, *Methods in Molecular Medicine: Therapeutic Applications of Ribozymes*, Humana Press, Totowa, NJ, 1998.
4. H.N. Wu, Y.J. Lin, F.P. Lin, S. Makino, M.F. Chang, M.M. Lai, 'Human Hepatitis *Delta* Virus RNA Subfragments Contain an Autocleavage Activity', *Proc. Natl. Acad. Sci. USA*, **86**(6), 1831–1835 (1989).
5. T.B. Fu, J. Taylor, 'The RNAs of Hepatitis *Delta* Virus are Copied by RNA Polymerase II in Nuclear Homogenates', *J. Virol.*, **67**(12), 6965–6972 (1993).
6. H.N. Wu, Y.J. Wang, C.F. Hung, H.J. Lee, M.M. Lai, 'Sequence and Structure of the Catalytic RNA of Hepatitis *Delta* Virus Genomic RNA', *J. Mol. Biol.*, **223**(1), 233–245 (1992).
7. A.T. Perrotta, M.D. Been, 'A Pseudoknot-like Structure Required for Efficient Self-cleavage of Hepatitis *Delta* Virus RNA' *Nature*, **350**(6317), 434–436 (1991).
8. A.D. Branch, H.D. Robertson, 'Efficient *trans* Cleavage and a Common Structural Motif for the Ribozymes of the Human Hepatitis *Delta* Agent', *Proc. Natl. Acad. Sci. USA*, **88**(22), 10 163–10 167 (1991).
9. M.G. Belinsky, G. Dinter-Gottlieb, 'Non-ribozyme Sequences Enhance Self-cleavage of Ribozymes Derived from Hepatitis *Delta* Virus', *Nucleic Acids Res.*, **19**(3), 559–564 (1991).
10. B.L. Bass, T.R. Cech, 'Specific Interaction Between the Self-splicing RNA of Tetrahymena and its Guanosine Substrate: Implications for Biological Catalysis by RNA', *Nature*, **308**(5962), 820–826 (1984).
11. C. Guerrier-Takada, S. Altman, 'Catalytic Activity of an RNA Molecule Prepared by Transcription In Vitro', *Science*, **223**(4633), 285–286 (1984).
12. J.F. Milligan, D.R. Groebe, G.W. Witherell, O.C. Uhlenbeck, 'Oligoribonucleotide Synthesis Using T7 RNA Polymerase and Synthetic DNA Templates', *Nucleic Acids Res.*, **15**(21), 8783–8798 (1987).
13. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Plainview, NY, 1989.
14. R. Vinayak, A. Andrus, A. Hampel, 'Rapid Desilylation of Oligoribonucleotides at Elevated Temperatures: Cleavage Activity in Ribozyme-substrate Assays', *Biomed. Pept. Proteins Nucleic Acids*, **1**(4), 227–230 (1995).
15. M.J. Gait, C. Pritchard, G. Slim, 'Oligonucleotide Synthesis', in *Oligonucleotides and Analogues, A Practical Approach*, IRL Press, Oxford, 1991.
16. R. Vinayak, 'Chemical Synthesis, Analysis, and Purification of Ribozymes', *Methods Mol. Biol.*, **74**, 51–58 (1997).

17. M.J. Moore, P.A. Sharp, 'Site-specific Modification of Pre-mRNA: the 2′-hydroxyl Groups at the Splice Sites', *Science*, **256**(5059), 992–997 (1992).
18. S. Ananvoranich, J.P. Perreault, 'Substrate Specificity of *Delta* Ribozyme Cleavage', *J. Biol. Chem.*, **273**(21), 13 182–13 188 (1998).
19. S. Ananvoranich, D.A. Lafontaine, J.P. Perreault, 'Mutational Analysis of the Antigenomic Trans-acting Delta Ribozyme: the Alterations of the Middle Nucleotides Located on the P1 Stem', *Nucleic Acids Res.*, **27**(6), 1473–1479 (1999).
20. J. SantaLucia, Jr, D.H. Turner, 'Measuring the Thermodynamics of RNA Secondary Structure Formation', *Biopolymers*, **44**(3), 309–319 (1997).
21. M.D. Been, A.T. Perrotta, 'Optimal Self-cleavage Activity of the Hepatitis Delta Virus RNA is Dependent on a Homopurine Base Pair in the Ribozyme Core', *RNA*, **1**(10), 1061–1070 (1995).
22. T. Xia, J.A. McDowell, D.H. Turner, 'Thermodynamics of Nonsymmetric Tandem Mismatches Adjacent to G.C Base Pairs in RNA', *Biochemistry*, **36**(41), 12 486–12 497 (1997).
23. N.A. Kolchanov, I.I. Titov, I.E. Vlassova, V.V. Vlassov, 'Chemical and Computer Probing of RNA Structure', *Prog. Nucleic Acid Res. Mol. Biol.*, **53**, 131–196 (1996).
24. T.L. Beattie, R.A. Collins, 'Identification of Functional Domains in the Self-cleaving Neurospora VS Ribozyme Using Damage Selection', *J. Mol. Biol.*, **267**(4), 830–840 (1997).
25. P.K. Kumar, K. Taira, S. Nishikawa, 'Chemical Probing Studies of Variants of the Genomic Hepatitis *Delta* Virus Ribozyme by Primer Extension Analysis', *Biochemistry*, **33**(2), 583–92 (1994).
26. S.P. Rosenstein, M.D. Been, 'Evidence that Genomic and Antigenomic RNA Self-cleaving Elements from Hepatitis *Delta* Virus have Similar Secondary Structures', *Nucleic Acids Res.*, **19**(19), 5409–5416 (1991).
27. P.C. Turner, *Methods in Molecular Biology: Ribozyme Protocols*, Vol 74, Humana Press, Totowa, NJ, 1997.
28. K.J. Hertel, D. Herschlag, O.C. Uhlenbeck, 'A Kinetic and Thermodynamic Framework for the Hammerhead Ribozyme Reaction', *Biochemistry*, **33**(11), 3374–3385 (1994).
29. B. Clouet-d'Orval, T.K. Stage, O.C. Uhlenbeck, 'Neomycin Inhibition of the Hammerhead Ribozyme Involves Ionic Interactions', *Biochemistry*, **34**, 1186–1190 (1995).
30. F. Nishikawa, H. Fauzi, S. Nishikawa, 'Detailed Analysis of Base Preferences at the Cleavage Site of a *trans*-acting HDV Ribozyme: a Mutation that Changes Cleavage Site Specificity', *Nucleic Acids Res.*, **25**(8), 1605–1610 (1997).
31. S. Mercure, D. Lafontaine, S. Ananvoranich, J.P. Perreault, 'Kinetic Analysis of *Delta* Ribozyme Cleavage', *Biochemistry*, **37**(48), 16 975–16 982 (1998).
32. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
33. S. Makino, M.F. Chang, C.K. Shieh, T. Kamahora, D.M. Vannier, S. Govindarajan, M.M. Lai, 'Molecular Cloning and Sequencing of a Human Hepatitis *Delta* (*Delta*) Virus RNA', *Nature*, **329**(6137), 343–346 (1987).

X-ray Structures of Nucleic Acids

Helen M. Berman
Department of Chemistry, Rutgers University, Piscataway, NJ, USA

1 Introduction and Background **1**
2 Description of Methods in Single-crystal Work **1**
2.1 Sample Preparation 1
2.2 Crystallization 2
2.3 Data Collection 2
2.4 Structure Determination 3
2.5 Refinement 4
2.6 Analysis of the Results 4
3 Survey of Key Results **7**
3.1 DNA 7
3.2 RNA 11
4 Challenges for the Future **11**
Abbreviations and Acronyms **11**
Related Articles **11**
References **11**

This article first summarizes the procedures used in the X-ray crystallographic determination of nucleic acids. The methods used to analyze the results of these determinations are presented, followed by a survey of the highlights of the more than 500 structures determined by these methods since the 1970s.

1 INTRODUCTION AND BACKGROUND

The landmark structure of double-helical DNA was discovered using X-ray diffraction data obtained from natural fibers.[1,2] In fibers, the helices are lined up in two-dimensional arrays. The third dimension is often indeterminate because the sample is heterogeneous in length, the helices do not line up in register if the length of every molecule is the same, or the helices have rotational disorder. Helices give characteristic X-ray diffraction patterns, such as the cross pattern of B-DNA, that make it possible to derive the pitch, rise, and number of residues per turn. The data to parameter ratio is low and therefore structures are determined by creating models and then calculating whether these models match the data. The inherent nature of the method is that one derives averaged structures. By using chemically synthesized samples with defined sequences, it has been possible to determine sequence–structure relationships for DNA and RNA. For example, it has been demonstrated that GC-rich sequences are in the A conformation whereas those that are AC-rich are more likely to assume the B conformation. It has also been possible to discover quaternary structures that have been validated using single-crystal methods. An excellent example is the quadruplex structure of poly G[3] that was first discovered in fibers, and then seen in single-crystals. This unusual structure is thought to be a feature of chromosomal telomeres.[4]

The first determinations of single-crystal structures of nucleic acids were published in the 1970s – significantly later than those of proteins. The main reason for this lag was the lack of pure material in sufficient quantity for crystallization. The crystal structure of tRNA was determined using X-ray methods in 1973[5] and was the first, and for a long time the only, biological RNA whose structure was determined. The ability to synthesize and purify oligonucleotides made it possible to determine their structures. The very first dinucleoside phosphate crystal structures provided a great deal of information about the conformational flexibility of nucleic acids.[6] Then in 1979, the first crystal structure determination of more than a full turn of DNA was published and gave rise to more than two decades of research on nucleic acid fine structure.[7]

In this article we outline the steps involved in sample preparation and structure determination of nucleic acids. At the end we summarize some of the highlights of the more than 500 structures determined using single-crystal X-ray methods.

2 DESCRIPTION OF METHODS IN SINGLE-CRYSTAL WORK

2.1 Sample Preparation

DNA samples are most often chemically synthesized using automated methods. Purification steps usually include the use of reversed-phase high-performance liquid chromatography followed by ion-exchange chromatography. RNA samples with fewer than 20–25 nucleotides can be also synthesized chemically. However, longer sequences require the use of enzymatic methods that involve transcribing the RNA from a DNA template using T7 polymerase.[8–10] These methods, however, do not always produce homogeneous samples and further processing of the sample using ribozymes has been employed.[11,12] Once the sample is made, it is

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

purified using both high-performance and fast protein liquid chromatographic methods. The preparation of RNA requires meticulous laboratory practices in order to ensure that the samples are not degraded by nucleases. This entails the use of autoclaved glassware and very high-quality chemicals. Thus, the preparation of RNA that is both pure and in large enough quantities for crystallization is significantly more difficult than for DNA.

2.2 Crystallization

The single-crystal method requires single, well-formed crystals that diffract well. The optimum size is about 0.2 mm on each side, although it is possible to use even smaller samples with the use of intense radiation sources. Oligonucleotide duplexes are sometimes difficult to crystallize in a form suitable for diffraction because the duplexes form quasi-infinite helices or even fibers. Sometimes an apparently beautiful crystal that looks well formed externally does not diffract at all, presumably because the molecules have not lined up in a lattice or have rotational disorder.

Generally, DNA crystals are grown using the vapor diffusion hanging drop method in which a drop of a solution of DNA is placed on a coverslip that is inverted over a reservoir containing the precipitating agent. The components of the drop usually include spermine, $MgCl_2$, sodium cacodylate, and a low percentage of methylpentanediol (MPD). The reservoir contains a higher percentage of MPD. Crystallization usually occurs within a few days. As shown in Figure 1(a) most DNA structures have been crystallized with MPD; attempts to use other agents have not been very successful.

Crystallization of oligonucleotides of RNA is done using procedures very similar to those used for DNA. Longer sequences of RNA including ribozymes and tRNA have been crystallized using a larger number of conditions than have been successful for DNA. Sparse matrix screens[14–17] have been developed that allow the exploration of a broader range of conditions for RNA crystallization in a manner similar to that which has been employed for proteins. Figure 1(a) and (b) shows that whereas MPD is the universal precipitating agent for DNA, a larger number of precipitating agents have been used for RNA.

2.3 Data Collection

Once a crystal has been obtained, it must be mounted for data collection. For data collection at temperatures between room temperature and freezing, the crystal is put into a sealed capillary with a small amount of mother liquor. Increasingly, however, crystals are flash frozen.[18]

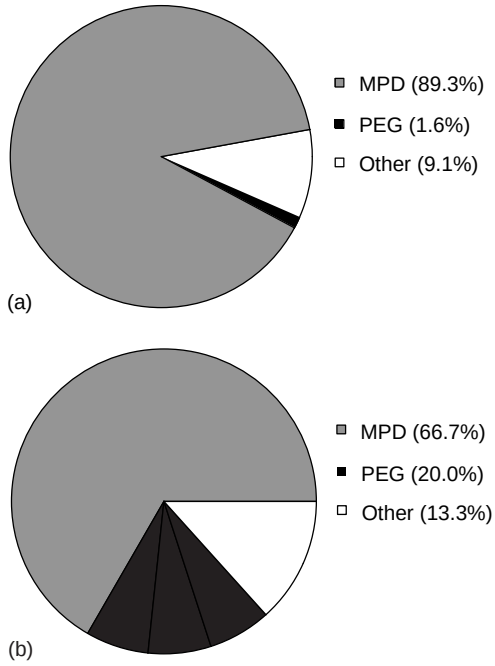

Figure 1 Precipitating agents for (a) DNA with length > 4 and (b) RNA with length > 4. (Data and figures were generated from NDBQuery.[13])

This makes it possible to collect an entire data set from a single-crystal since it will not decay in the X-ray beam.

Once the crystal has been mounted, data are collected. Over the years, the methods for data collection have changed considerably. Whereas in the past data were collected using a sealed tube and a diffractometer, nowadays almost all data are collected using a rotating anode which supplies a more intense source of X-rays and either an image plate or multiwire area detector (Figure 2a–d). Although most nucleic acid crystals do not usually require the use of synchrotron radiation, several structures have been determined using data from these sources. It is expected that as the demand for higher resolution information grows, the use of synchrotron data for these samples will increase. Data for protein–nucleic acid complexes are increasingly being obtained using synchrotron sources because of the improved resolution of the data and because of the possibility of using multiple anomalous dispersion (MAD) phasing for the structure determinations.

The resolution of the data determines how accurate the model of the structure will be. The resolution of DNA structures ranges from 0.9 to 3.25 Å with the average being around 2.2 Å (Figure 3a and b). Given the relatively small size of these molecules, these resolutions are relatively low. RNA structures tend to have even poorer resolutions, which may be because of the presence of many flexible loops.

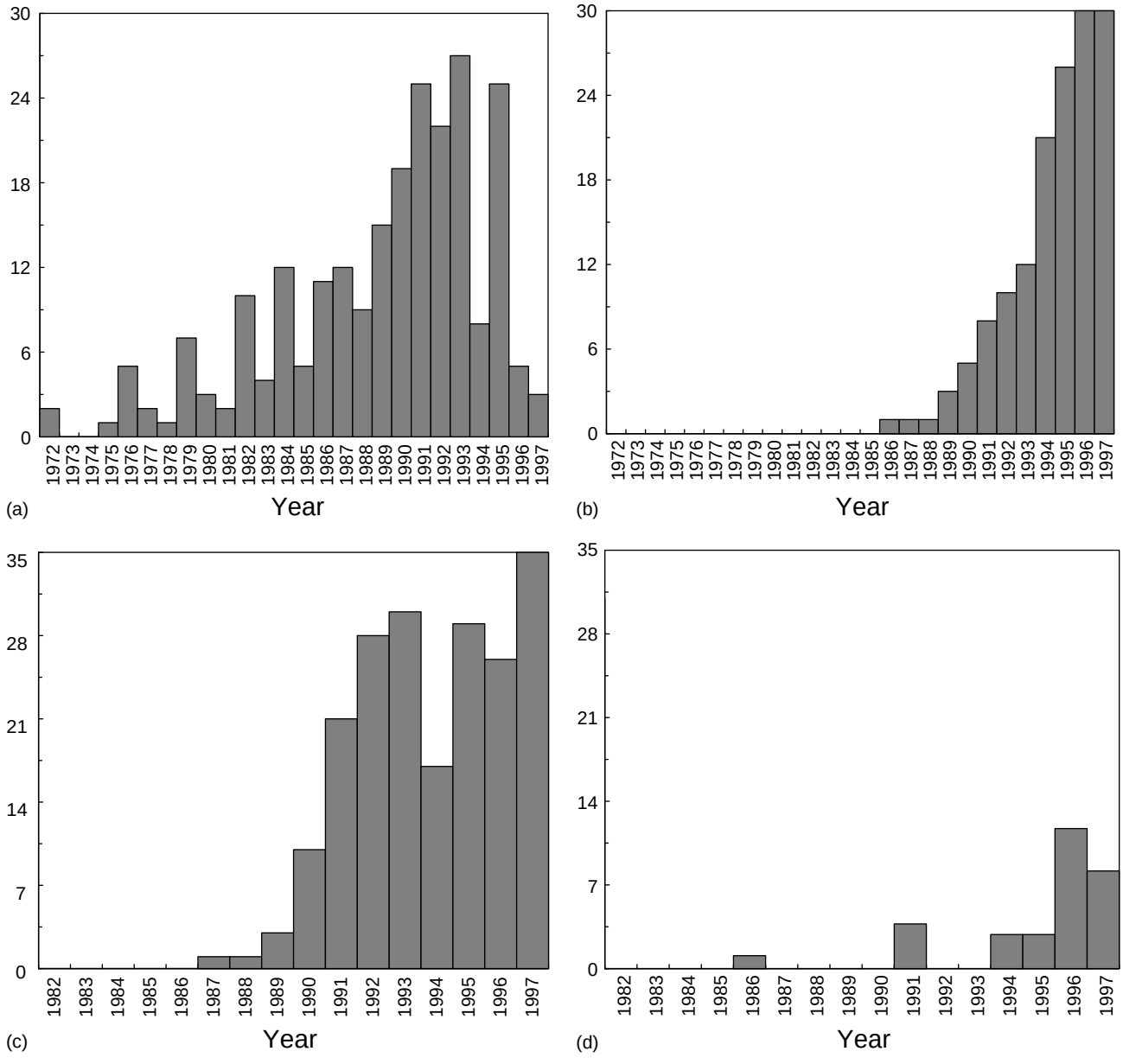

Figure 2 Data collection systems used in structure determination by year. (a) Diffractometer; (b) area detector/image plate; (c) rotating anode; (d) synchrotron. (Data and figures were generated from NDBQuery.[13])

2.4 Structure Determination

The challenge in X-ray crystallography is to obtain the phases of the reflections. Once the phases have been obtained, electron density maps can be calculated. From these maps one can trace the atomic structure of the model. Very small molecules can be phased using direct methods based on reflection statistics. Nucleic acids are in general too large and their resolution is too low for these methods to be used. The very first DNA and RNA structures were solved using isomorphous replacement methods in which heavy atoms are introduced into the lattice. A popular way to do this is to introduce brominated cytosines or uracils into the sequence. For larger and more globular molecules, heavy metals have been soaked into the crystal. Once these heavy atoms have been located, they are used to phase the reflections. This is the most unambiguous way of directly determining the positions of the atoms and is the least likely to lead to incorrect structures. Once a model has been determined for a particular DNA or RNA type, molecular replacement methods are often used for phase determination. In this method, a model of the structure is rotated and translated in the cell until

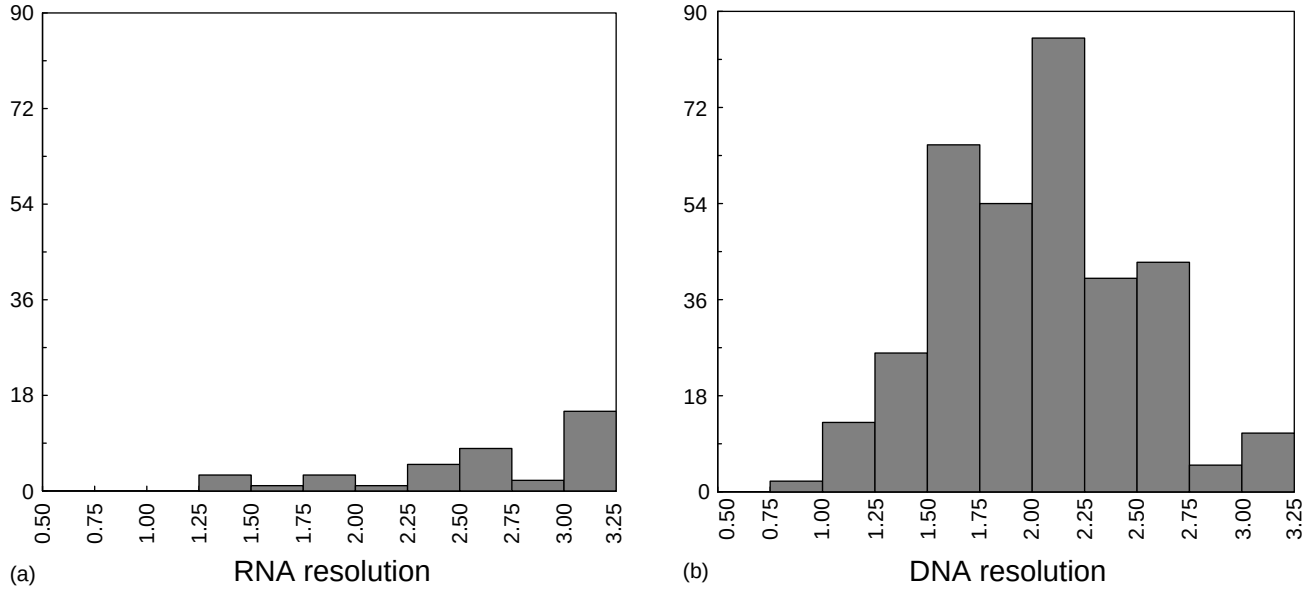

Figure 3 Distribution of resolution for oligonucleotides with length > 4. (a) RNA; (b) DNA. (Data and figures were generated from NDBQuery.[13])

the model fits the data. If the model is very good, this method works well and gives reliable results. Although molecular replacement has been used with great success for helical DNA, it is possible to obtain incorrect results because of the hypersymmetry of the helix. Hence it is important to carefully evaluate the results of such determinations.

2.5 Refinement

Nucleic acid structures are refined using programs that fit the model to the experimental reflection data; to reduce the number of variable parameters, geometrical restraints are applied by restraining the bond distances and angles to chemically reasonable values. Careful fitting of the electron density maps is also done with electronic visualization programs such as FRODO[19] or O.[20] Central to the use of these programs is the availability of dictionaries containing standard values for the restraints and descriptions of the topologies of the molecules; the popularity of certain programs for nucleic acid refinement matched the availability of these dictionaries. CORELS[21] which was developed for the refinement of tRNA, was superseded by NUCLSQ,[22] which was an adaptation of PROLSQ[23] for nucleic acids. For a long time, this program was the one of choice. X-PLOR,[24] which was enormously popular for proteins, did not begin to be used extensively for nucleic acids until an appropriate parameter file for the nucleic acid force constants was developed[25] (Figure 4a–c). The use of R_{free} in which the R factor for a subset of reflections is monitored as the model is changed has helped to improve the quality of nucleic acid structures. As the resolution of structures begins to improve to values below 0.8 Å, full matrix least-squares methods will begin to be used and the program SHELX[26] is likely to gain in usage for macromolecules (Figure 5).

For all the nucleic acid structures in the Nucleic Acid Database (NDB),[28] the measure of the fit of the model to data, known as the R-value, ranges from 0.07 to 0.25 and the mean value is about 0.18. Although it would be expected that there would be some correlation between resolution and R-value, there is not. By most criteria, the refinements of nucleic acid structures are very good with few anomalies present.

2.6 Analysis of the Results

Once the structure has been determined, there are a variety of analyses that are done to assess both the quality of the structure and the important structural features. In this section, we present a description of these analyses.

2.6.1 Structure Quality

Quality checks for nucleic acid crystal structures are done when the coordinates are deposited in the NDB, which is a searchable archive of nucleic acid structures. The average deviations of the valence geometry from standard values are calculated. In general, these values are <0.02 Å for the bond distances and $<1.5^\circ$ for the angles. These values are more of an indicator of the applicability of the dictionary that was used for refinement than of the actual quality of the structure. Individual values that deviate more than 4σ

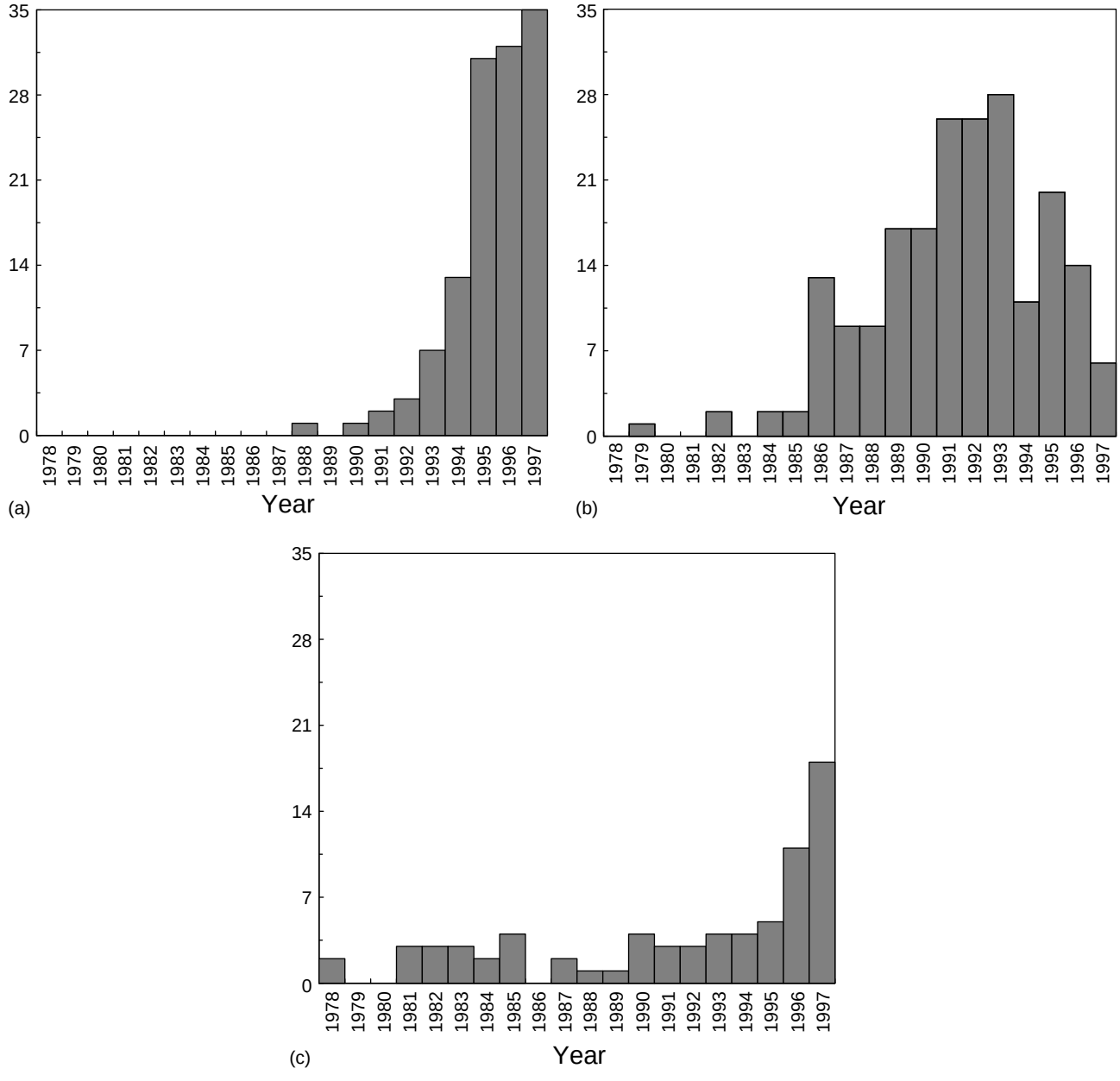

Figure 4 Refinement programs used in structure determination by year. (a) X-PLOR; (b) NUCLSQ; (c) other programs. (Data and figures were generated from NDBQuery.[13])

from these averages are also calculated. Intermolecular contacts within the asymmetric unit and in the unit cell are determined. If there are crystal contacts that are smaller than 2.2 Å, and there are no errors in the cell dimensions, this is an indication that there is a problem in the structure; in general, there are very few instances of these sorts of errors in nucleic acid structures. The chiralities of the asymmetric carbons are also calculated. These types of chirality errors are rarely seen in nucleic acid crystal structures although they have been observed in nuclear magnetic resonance nucleic acid structures.[29] Finally checks are made of the match of the model against the experimental reflection data.[30] The values of these *R*-factors are compared with those reported by the author and in most cases there is little difference between them.

2.6.2 Conformational Analysis

There are seven rotatable bonds in the backbone of nucleic acids (Figure 5) and five in the ribose sugar. The torsion angles for these bonds are different for the three classes of DNA double-helical duplexes: A, B, Z

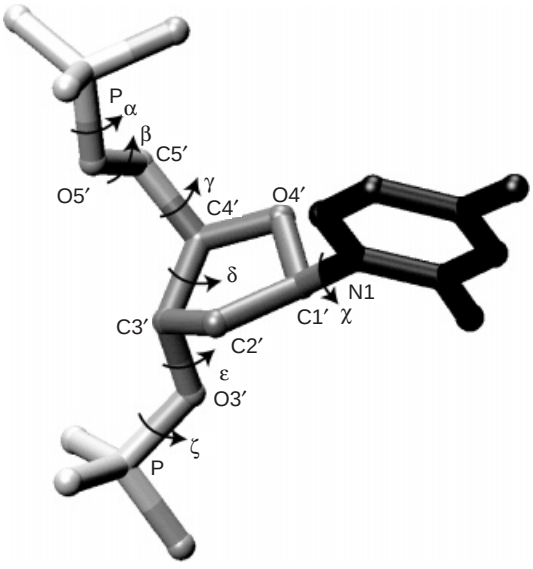

Figure 5 Rotatable bonds in the conformation of nucleic acids. (Reproduced from H.M. Berman, *Biopolymers*, **44**, 23–44 (1997).[27] Copyright 1997 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

(Figure 6a–c). The repeating unit for the A and B class of right-handed helices is a mononucleotide and there are subtle but distinct differences in the conformation angles for these building blocks. In A-DNA, the sugar pucker is C3′ endo with a δ value in the gauche + range and the χ angle low anti. In B-DNA, the sugars are closer to C2′ endo with the δ value trans and the χ angle high anti. In Z-DNA left-handed helices, the repeating unit is a dinucleotide having an alternation between syn and anti glycosidic angles as its key feature. The ranges for all the torsion angles in duplex DNA has been established (Figure 7a–c).[32] The ranges for the torsion angles for the other classes of DNA structures and for single stranded RNA molecules have not yet been established but it is likely that they will be significantly broader.

2.6.3 Base Morphology

The base pairs in duplex DNA and RNA are not planar (Figure 8). Indeed, it is common for some pairs to have a large propeller twist so that the base pair of one step can form hydrogen bonds with the base pair of the next step. In addition, there are parameters that describe the relationship of the base pairs in succeeding steps. The step parameters that are most useful in describing nucleic acid structure are roll, twist, and slide. For perfectly canonical B-DNA these values are approximately 0, 36, and 0°, respectively. The values for slide and twist are lower in A-DNA (Table 1). Individual structures can deviate widely from the ideal B- or A-DNA models. In the highly kinked DNA of CAP–DNA complexes, for example, the values of roll and twist deviate significantly. Several different algorithms have been developed to calculate these base morphology parameters.[37,38] All of these programs yield similar results, although some subtle differences do exist. Attempts to develop a sequence-specific code for these parameters have been partially successful; it will not be possible to have a reliable set of values for the base morphology parameters for each of the 10 unique base steps in DNA until high-resolution data are obtained.

2.6.4 Hydration

The very first studies of DNA highlighted the importance of the humidity in determining the molecular conformation. B-DNA occurs in high humidity and A-DNA at lower humidity. The crystal structures of nucleic acids have demonstrated certain hydration patterns including

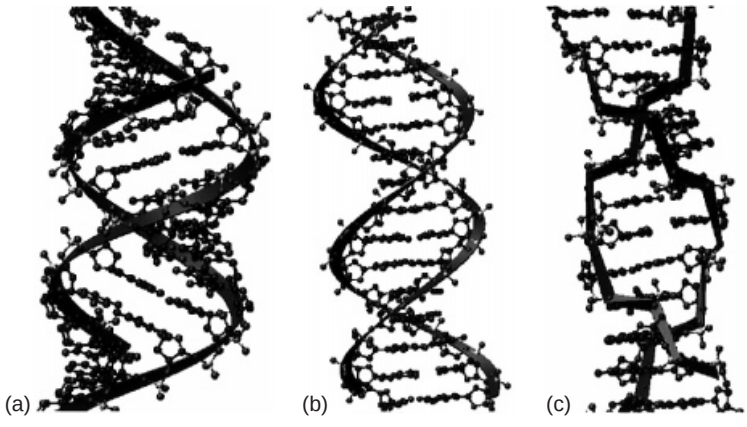

Figure 6 Canonical helical types of (a) A-DNA; (b) B-DNA; and (c) Z-DNA. (Reprinted from H.M. Berman, A. Gelbin, J. Westbrook, 'Nucleic Acid Crystallography: A View from the Nucleic Acid Database', *Prog. Biophys. Mol. Biol.*, **66**, 255–288 (1996).[31] Copyright 1996, with permission from Elsevier Science.)

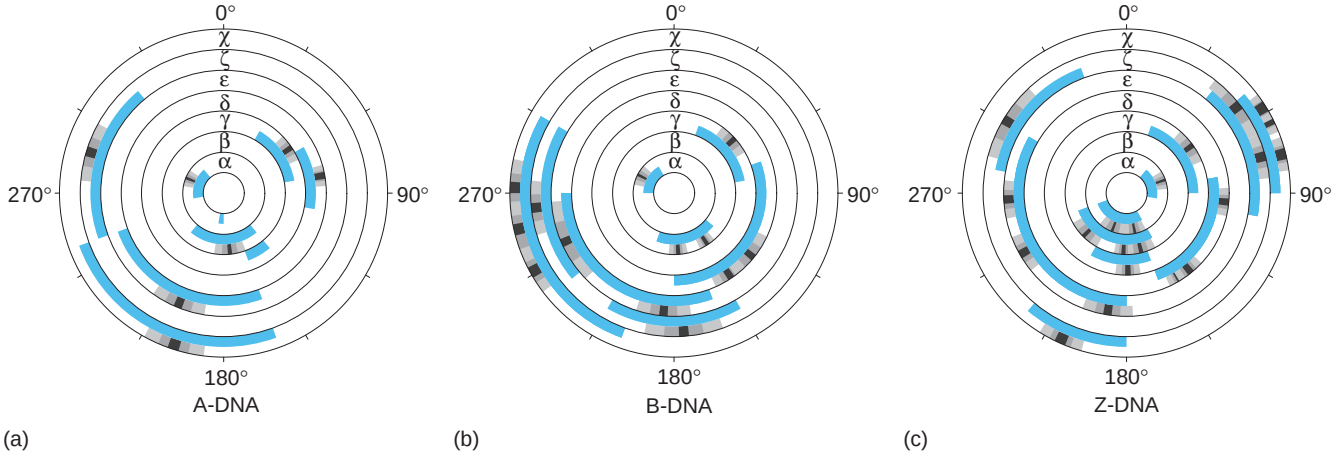

Figure 7 Conformation wheels for (a) A-DNA; (b) B-DNA; and (c) Z-DNA showing the ranges in blue. The average values are shown in black with the one and two estimated standard deviations of the experimental sample used to derive the averages in gray. (From B. Schneider, S. Neidle and H.M. Berman, *Biopolymers*, **42**, 113–124 (1997).[32] Copyright 1997 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

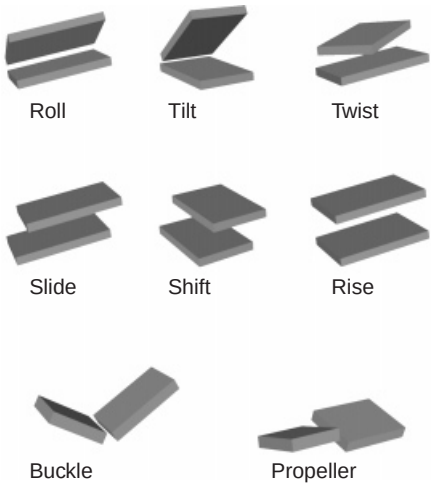

Figure 8 Definitions of base morphology parameters.[33] (From H.M. Berman, *Biopolymers*, **44**, 23–44 (1997).[27] Copyright 1997 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons.)

Table 1 Average values of the base morphology parameters[a] in DNA and RNA duplexes with resolution 2.0 Å or better and in selected nucleic acid fiber structures[34]

Type	Roll (°)	Slide (Å)	Twist (°)
B-DNA with resolution ≤ 2.0	1.0	−0.05	35.8
B-DNA fiber	1.7	0.5	36.0
A-DNA with resolution ≤ 2.0	7.0	−1.92	31.4
A-DNA fiber	12.4	−1.34	30.1
A-RNA with resolution ≤ 2.0	10.4	−2.02	31.8
A-RNA fiber	8.6	−1.43	31.5

[a] Parameters were computed using CURVES[35] for the crystal structures and CompDNA[36] for the fiber structures.

spines and ribbons of water in B-DNA (Figure 9a and b) and water-mediated bridges between neighboring phosphates in A-DNA. Systematic analyses of the hydration patterns around the duplexes have established that the hydration around the bases is local[41] and each base has a distinctive pattern of directly bound water. The hydration around phosphates is also well described.[32] By including the hydrated building blocks to model specific DNA conformations, the characteristic hydration geometries can be reproduced.

3 SURVEY OF KEY RESULTS

3.1 DNA

All the DNA structures determined thus far have been of synthetic oligonucleotides ranging in length from 2 to 16 nucleotides (Table 2). The largest set contains self-complementary dodecamers, which have been modeled after the first B-DNA structure whose sequence is CGCG-CAATTCGCG. Unlike the canonical structure derived from fiber work, this structure demonstrates unique characteristics for each step in the sequence. For example, the A–T base pairs have high propeller twist and the stretch of A sequence is very straight. The structure also shows an asymmetric bend. This template was used to explore the effects of modifications and mismatches and drugs on the structures of DNA; more than 50 structures of this type have been determined. Virtually all of the dodecamer structures fall into a single-crystal form. The advantage of this is that one can explore a variety of sequence–structure relationships while keeping the packing arrangement constant. A disadvantage is that one can

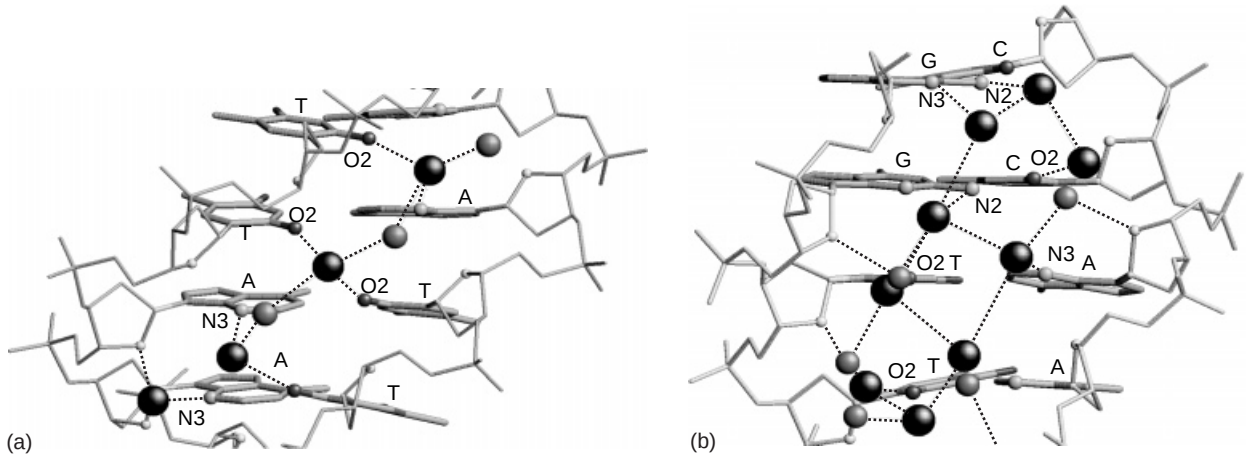

Figure 9 The hydration patterns observed in B-DNA. The hydrogen bonds are shown as dashed lines. (a) The spine of hydration in the minor groove of d[CGCGAATTCGCG]$_2$ (BDL001).[7] The first shell waters are hydrogen bonded to the purine N3 and pyrimidine O2 atoms and are shown as large spheres. The second shell waters bridge the first shell waters and are shown as small spheres. (b) The double row of waters in the minor groove of d[CCAACGTTGG]$_2$ (BDL019).[39] (Reproduced from H.M. Berman, B. Schneider, in *Oxford Handbook of Nucleic Acid Structure*, ed. S. Neidle, Oxford University Press, Oxford, 295–310, 1999,[40] by permission of Oxford University Press.)

Table 2 Types of nucleic acid structures

(a) Duplexes (as of September 1998)

Structure type	Ranges of length	Number with all Watson–Crick pairs	Number containing modifications	Number containing mismatches	Number containing special features
A-DNA	4–12	95	30	4	3
B-DNA	6–13	143	52	16	20
Z -DNA	2–16	51	28	8	7
A-RNA	2–17	23	3	10	5
DNA–RNA hybrid	6–11	25	4	1	6

(b) Structures with novel features (as of September 1998)

Type	Number
DNA tetramers: guanine quartets	3
DNA tetramers: intercalated cytosines	4
Non-classified DNA and RNA structures	78

(c) Biological RNA structures (as of September 1998)

Type	Number
Ribozymes	8
tRNA	11

(d) Protein–nucleic acid complexes (as of September 1998)

Type	Number
Protein–DNA	256
Protein–RNA	29

never be sure of how much this particular lattice is actually influencing the structure. For the most part, neither simple chemical modifications nor the presence of mismatches create global changes in the DNA. Although the effects are local, they could certainly have profound effects on recognition at the site of the lesion and thus aid in repair.

The more than 30 decamer structures crystallize in a variety of space groups. The crystal packing range from ones in which the helices line up one above the other forming crystals with almost idealized closest packing, to ones in which the major groove of one helix packs against the backbone of another (Figure 10a–d). These structures are in general higher resolution and again provided a fertile template for the understanding of DNA fine structure. The results of studies on B-DNA structure have been reviewed elsewhere.[27,46]

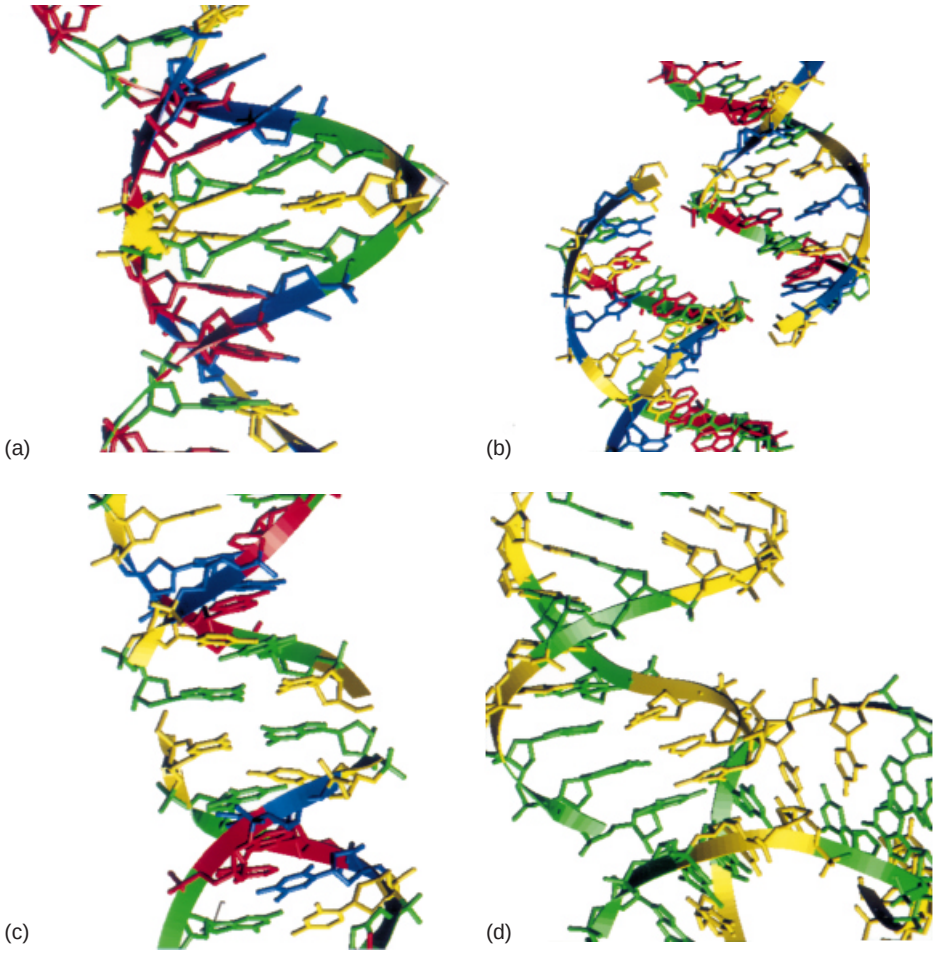

Figure 10 Packing motifs in DNA duplexes in B- and A-DNA. (a) Minor groove–minor groove interactions in BDL042;[42] (b) Major groove–backbone interactions in BDJ060;[43] (c) stacking interactions in BDJ025;[44] (d) terminal base pair–minor groove interaction in ADJ049.[45] The bases are color coded: (green = guanine; yellow = cytosine; red = adenine; blue = thymine). (Reprinted from H.M. Berman, A. Gelbin, J. Westbrook, 'Nucleic Acid Crystallography: A View from the Nucleic Acid Database', *Prog. Biophys. Mol. Biol.*, **66**, 255–288 (1996).[31] Copyright 1996, with permission from Elsevier Science.)

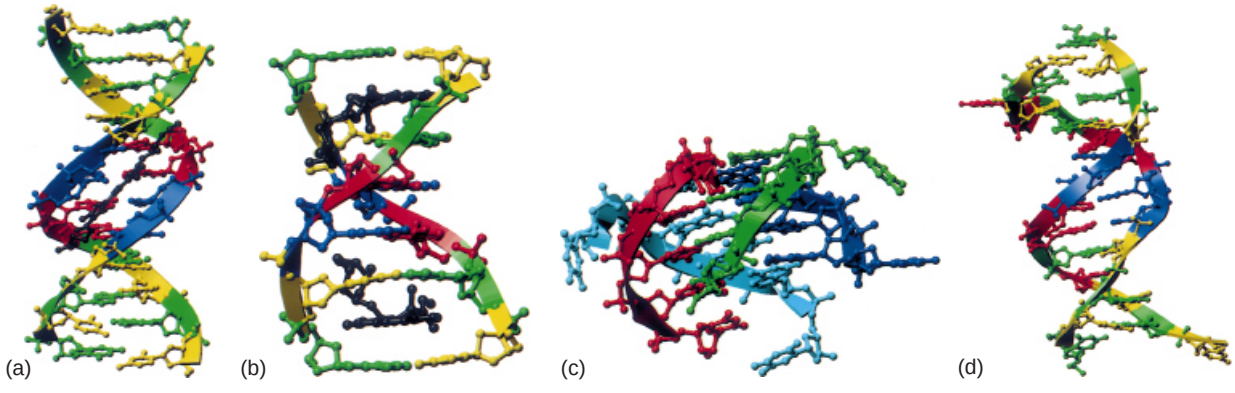

Figure 11 Some examples of crystalline DNA. (a) B-DNA with a drug bound in the minor groove, GDLB05;[49] (b) DNA with intercalated drugs, DDF019;[50] (c) tetraplex DNA with G quartets, UDF062;[51] (d) B-DNA containing a bulge, UDM010.[52] The bases are color coded: green = guanine; yellow = cytosine; red = adenine; blue = thymine; other residues and drugs are black [except in (c), which is colored by strand].

The dodecamer and decamer structures are amenable to the binding of drugs in the minor groove, making it possible to gain insight into molecular recognition. More than 40 structures have been determined in which the drug nestles into the minor groove. In some cases, the drugs bind in pairs. There are also several examples of intercalating drugs such as daunomycin and actinomycin whose planar groups fit between two base pairs and cause the DNA to unwind. Both of these drugs also have prosthetic groups that further interact with the DNA.

A-DNA structures have also been extensively studied and reviewed.[47] In general there are greater variations in the helices, probably because of packing effects. The packing of these molecules is very characteristic with the minor groove of one duplex butting up against the phosphate backbone of another.

One of the most surprising results of the explorations of DNA structure was the discovery of the left-handed Z-DNA [for a review, see Ho and Mooers[48]], which exists in solutions containing very high salt or alcohol

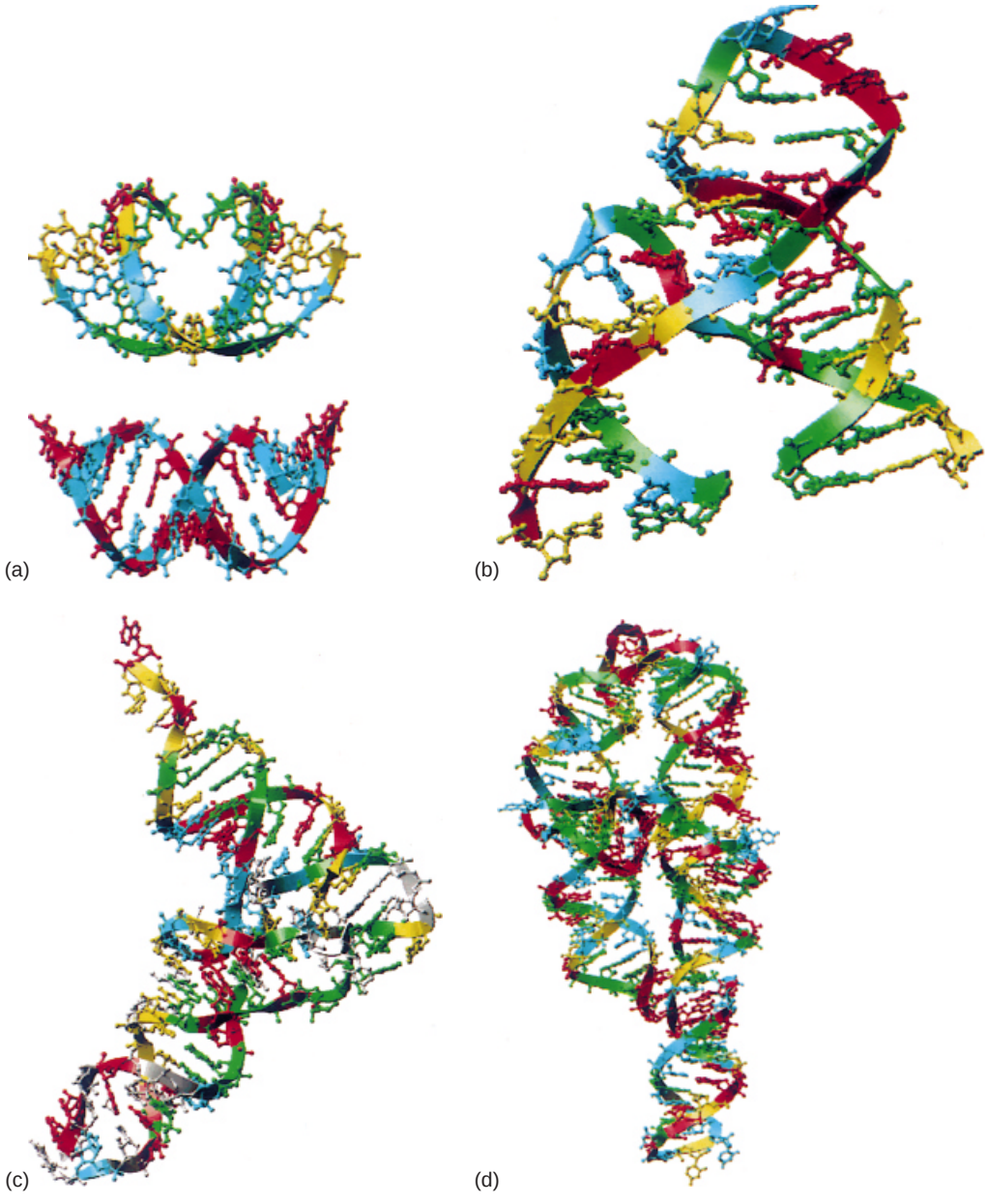

Figure 12 Some examples of RNA structures. (a) A-RNA duplex with internal loops, ARL037[55] (top) and Watson–Crick base-paired A-RNA, ARN035[56] (bottom); (b) a hammerhead ribozyme, URX035;[57] (c) yeast phe tRNA, TRNA04;[58] (d) a group I ribozyme domain, URX053.[59]

levels. The correlates of this conformation in crystal form are the presence of alternating pyrimidine and purine sequences and a strand length of six. Almost all the Z-DNA structures pack with the helices lined up one atop the other.

DNA can also form quaduplexes (Figure 11a–d). Two different types have been observed. In one type the guanines form quartets [e.g. UDF036[53]]. In the other cytosines from two strands intercalate between cytosines of the other two strands [e.g. UDD023[54]].

Within these major classes of DNA structures, many special features have been observed. These include bulges caused by extra bases in one strand, frayed ends or single base overhangs. These latter features enable the ends of the DNA molecules to associate in novel ways forming for example triplets. Interestingly, the one major structural type that has never been observed in crystals is the DNA triple helix.

3.2 RNA

The very first oligonucleotide structures to be determined were RNA dinucleoside phosphates and for a long time these were the only oligomers studied using X-ray methods. With the availability of methods to prepare, purify, and crystallize RNA, increasing numbers of RNA structures are appearing.

Duplexes of RNA with Watson–Crick base pairing have been crystallized; these have the expected A conformation (Figure 12a–d). In some of these cases, sequences were crystallized in order to model RNA hairpin stem-loops; instead, A-type helices were seen in the crystal.[55] To accomplish this, these helices have stretches of mismatches that form internal loops. Furthermore, the mismatches seen are not simply ones involving the bases themselves but also utilize water molecules. A summary of the types of mismatches and internal loops is given elsewhere.[60]

Two classes of biological RNA molecules have been crystallized: tRNAs[61] and ribozymes.[57] The tRNA molecules are single stranded molecules in which a cloverleaf type secondary structure folds into an L-shaped molecule with extensive base stacking and tertiary hydrogen bonds. These hydrogen bonds form between conserved bases and thus are key structural elements in all tRNA molecules. The structures of hammerhead ribozymes, like the structures of tRNA, have extensive base stacking and exhibit some of the same unusual conformation and hydrogen bonding features. The structures of both tRNA and the hammerhead ribozyme have provided testable models for understanding their biological functions. Base stacking, tertiary hydrogen bonding interactions, and more conformational motifs also dominate the spectacular structure of the P4–P6 domain of the group 1 intron ribozyme.[59] Although it is certain that there will be more RNA structural elements forthcoming, it should now be possible to classify systematically those that have been seen thus far in order to facilitate modeling of larger RNA structures.

4 CHALLENGES FOR THE FUTURE

Although many structures have been determined, there are still many unanswered questions. In the field of DNA crystallography, the issue of how much the sequence determines the structure is as yet unknown. To be able to understand this, it will be necessary to determine the structures of the same sequences in different crystalline environments. The effects of complexation to proteins and drugs can only be answered by studying sequences alone and in complexes. Further, of course, in dealing with the level of fine structure inherent in defining base morphology, there is a clear need for very high-resolution structures. RNA crystallography is still in its infant stage, and when fully developed it will allow us to understand more fully the complex rules that govern the folding of single strands into compact structures.

ABBREVIATIONS AND ACRONYMS

MAD	Multiple Anomalous Dispersion
MPD	Methylpentanediol
NDB	Nucleic Acid Database

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Molecules, Properties and Detection of Single • DNA Structures of Biological Relevance, Studies of Unusual Sequences • RNA Tertiary Structure • Sequencing Strategies and Tactics in DNA and RNA Analysis • Structural Analysis of Ribozymes

REFERENCES

1. J.D. Watson, F.H.C. Crick, 'A Structure for Deoxyribose Nucleic Acid', *Nature*, **171**, 737–738 (1953).
2. R.E. Franklin, R.G. Gosling, 'Molecular Configuration in Sodium Thymonucleate', *Nature*, **171**, 740–741 (1953).

3. S. Arnott, R. Chandrasekharan, C.M. Marttila, 'Structures of Polyinosinic Acid and Polyguanylic Acid', *Biochem. J.*, **141**, 537–538 (1974).
4. E.H. Blackburn, 'Structure and Function of Telomeres', *Nature*, **350**, 569–573 (1991).
5. S.-H. Kim, F.L. Suddath, G.J. Quigley, A. McPherson, J.L. Sussman, A.H.-J. Wang, N.C. Seeman, A. Rich, 'Three-dimensional Tertiary Structure of Yeast Phenylalanine Transfer RNA', *Science*, **185**, 435–440 (1974).
6. N.C. Seeman, J.L. Sussman, H.M. Berman, S.-H. Kim, 'Nucleic Acid Conformation: Crystal Structure of a Naturally Occurring Dinucleoside Phosphate (UpA)', *Nat. New Biol.*, **233**, 90–92 (1971).
7. H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson, 'Structure of a B-DNA Dodecamer: Conformation and Dynamics', *Proc. Natl. Acad. Sci. USA*, **78**, 2179–2183 (1981).
8. J. Wyatt, M. Christain, J. Puglisi, 'Synthesis and Purification of Large Amounts of RNA Oligonucleotides', *BioTechniques*, **11**, 764–769 (1991).
9. R. Kim, E.L. Holbrook, J. Jancarik, S.H. Kim, 'Synthesis and Purification of Milligram Quantities of Short RNA Transcripts', *BioTechniques*, **18**, 992–994 (1995).
10. J.F. Milligan, D.R. Groebe, G.W. Witherell, O.C. Uhlenbeck, 'Oligoribonucleotide Synthesis Using T7 RNA Polymerase and Synthetic DNA Templates', *Nucleic Acids Res.*, **15**, 8783–8798 (1987).
11. A.R. Ferre-D'Amare, J.A. Doudna, 'Use of *cis*- and *trans*-Ribozymes to Remove 5′ and 3′ Heterogeneities from Milligrams of in Vitro Transcribed RNA', *Nucleic Acids Res.*, **24**, 977–978 (1996).
12. S.R. Price, N. Ito, C. Oubridge, J.M. Avis, K. Nagai, 'Crystallization of RNA–Protein Complexes. I. Methods for the Large-scale Preparation of RNA Suitable for Crystallographic Studies', *J. Mol. Biol.*, **249**, 398–408 (1995).
13. J. Westbrook, T. Demeny, S.-H. Hsieh, *NDBQUERY, V.4.0. A Simplified User Interface to the Nucleic Acid Database*, NDB-99, Rutgers University, New Brunswick, NJ, 1996.
14. J.A. Doudna, C. Grosshans, A. Gooding, C.E. Kundrot, 'Crystallization of Ribozymes and Small RNA Motifs by a Sparse Matrix Approach.', *Proc. Natl. Acad. Sci. USA*, **90**, 7829–7833 (1993).
15. K.J. Baeyens, J. Jancarik, S.R. Holbrook, 'Use of Low-molecular-weight Polyethylene Glycol in the Crystallization of RNA Oligomers', *Acta Crystallogr., Sect. D*, **50**, 764–767 (1994).
16. W.G. Scott, J.T. Finch, R. Grenfell, J. Fogg, T. Smith, M.J. Gait, A. Klug, 'Rapid Crystallization of Chemically Synthesized Hammerhead RNAs Using a Double Screening Procedure', *J. Mol. Biol.*, **250**, 327–332 (1995).
17. I. Berger, C. Kang, N. Sinha, M. Wolters, A. Rich, 'A Highly Efficient 24-Condition Matrix for the Crystallization of Nucleic Acid Fragments', *Acta Crystallogr., Sect. D*, **52**, 465–468 (1996).
18. H. Hope, 'Cryocrystallography of Biological Macromolecules: a Generally Applicable Method', *Acta Crystallogr., Sect. B*, **44**, 22–26 (1988).
19. T.A. Jones, 'FRODO: a Graphic Model Building and Refinement System for Macromolecules', *J. Appl. Crystallogr.*, **11**, 268–272 (1978).
20. T.A. Jones, J.-Y. Zou, S.W. Cowan, M. Kjeldgaard, 'Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models', *Acta Crystallogr., Sect. A*, **47**, 110–119 (1991).
21. J.L. Sussman, S.R. Holbrook, G.M. Church, S.-H. Kim, 'CORELS', *Acta Crystallogr., Sect. A*, **33**, 800–804 (1977).
22. E. Westhof, P. Dumas, D. Moras, 'Crystallographic Refinement of Yeast Aspartic Acid Transfer RNA', *J. Mol. Biol.*, **184**, 119–145 (1985).
23. W.A. Hendrickson, J.H. Konnert, 'PROLSQ', in *Biomolecular Structure, Conformation, Function and Evolution, 1*, eds. R. Srinivasan, E. Subramanian, N. Yathindra, Pergamon Press, Oxford, 43–57, 1981.
24. A.T. Brünger, *X-PLOR, Version 3.1, a System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT, 1992.
25. G. Parkinson, J. Vojtechovsky, L. Clowney, A.T. Brünger, H.M. Berman, 'New Parameters for the Refinement of Nucleic Acid Containing Structures', *Acta Crystallogr., Sect. D*, **52**, 57–64 (1996).
26. G.M. Sheldrick, T.R. Schneider, 'SHELXL: High Resolution Refinement', In *Methods in Enzymology*, eds. R.M. Sweet, C.W. Carter, Jr, Academic Press, Orlando, **277**, 319–343 (1997).
27. H.M. Berman, 'Crystal Studies of B-DNA: the Answers and the Questions', *Biopolymers*, **44**, 23–44 (1997).
28. H.M. Berman, W.K. Olson, D.L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.H. Hsieh, A.R. Srinivasan, B. Schneider, 'The Nucleic Acid Database – a Comprehensive Relational Database of Three-dimensional Structures of Nucleic Acids', *Biophys. J.*, **63**, 751–759 (1992).
29. P. Schultze, J. Feigon, 'Chirality Errors in Nucleic Acid Structures', *Nature*, **387**, 668 (1997).
30. A.A. Vaguine, J. Richelle, S.J. Wodak, 'SFCHECK: a Unified Set of Procedures for Evaluating the Quality of Macromolecular Structure-factor Data and Their Agreement with the Atomic Model', *Acta Crystallogr., Sect. D*, **55**, 191–205 (1999).
31. H.M. Berman, A. Gelbin, J. Westbrook, 'Nucleic Acid Crystallography: a View from the Nucleic Acid Database', *Prog. Biophys. Mol. Biol.*, **66**, 255–288 (1996).
32. B. Schneider, S. Neidle, H.M. Berman, 'Conformations of the Sugar–Phosphate Backbone in Helical DNA Crystal Structures', *Biopolymers*, **42**, 113–124 (1997).
33. W.K. Olson, 'Simulating DNA at Low Resolution', *Curr. Opin. Cell. Biol.*, **6**, 242–256 (1996).
34. R. Chandrasekaran, S. Arnott, in *Landolt–Börnstein Numerical Data and Functional Relationships in*

Science and Technology, Group VII/1b, Nucleic acids, ed. W. Saenger, Springer, Berlin, 1989.
35. R. Lavery, H. Sklenar, 'Defining the Structure of Irregular Nucleic Acids: Conventions and Principles', *J. Biomol. Struct. Dyn.*, **6**, 655–667 (1989).
36. A.A. Gorin, V.B. Zhurkin, W.K. Olson, 'B-DNA Twisting Correlates with Base-pair Morphology', *J. Mol. Biol.*, **247**, 34–48 (1995).
37. R. Lavery, H. Sklenar, 'The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids', *J. Biomol. Struct. Dyn.*, **6**, 63–91 (1988).
38. R. Lavery, K. Zakrzewska, 'Base and Base Pair Morphologies, Helical Parameters, and Definitions', in *Oxford Handbook of Nucleic Acid Structure*, ed. S. Neidle, Oxford University Press, Oxford, 39–74, 1999.
39. G.G. Privé, K. Yanagi, R.E. Dickerson, 'Structure of the B-DNA Decamer C–C–A–A–C–G–T–T–G–G and comparison with isomorphous decamers C–C–A–A–G–A–T–T–G–G and C–C–A–G–G–C–C–T–G–G', *J. Mol. Biol.*, **217**, 177–199 (1991).
40. H.M. Berman, B. Schneider, 'Nucleic Acid Hydration', in *Oxford Handbook of Nucleic Acid Structure*, ed. S. Neidle, Oxford University Press, Oxford, 295–310, 1999.
41. B. Schneider, H.M. Berman, 'Hydration of the DNA Bases is Local', *Biophys. J.*, **69**, 2661–2669 (1995).
42. G.A. Leonard, W.N. Hunter, 'Crystal and Molecular Structure of d(CGTAGATCTACG) at 2.25 Å Resolution', *J. Mol. Biol.*, **234**, 198–208 (1993).
43. D.S. Goodsell, K. Grzeskowiak, R.E. Dickerson, 'Crystal structure of C–T–C–T–C–G–A–G–A–G. Implications for the Structure of the Holliday Junction.', *Biochemistry*, **34**, 1022–1029 (1995).
44. K. Grzeskowiak, K. Yanagi, G.G. Privé, R.E. Dickerson, 'The Structure of B-helical C–G–A–T–C–G–A–T–C–G and Comparison with C–C–A–A–C–G–T–T–G–G: the Effect of Base Pair Reversal', *J. Biol. Chem.*, **266**, 8861–8883 (1991).
45. B. Ramakrishnan, M. Sundaralingam, 'High-resolution Crystal Structure of the A-DNA Decamer d(CCCGGCCGGG) – Novel Intermolecular Base-paired G•(G•C) Triplets', *J. Mol. Biol.*, **231**, 431–444 (1993).
46. R. Dickerson, 'Helix Structure and Molecular Recognition by B-DNA', in *Oxford Handbook of Nucleic Acid Structure*, ed. S. Neidle, Oxford University Press, Oxford, 145–199, 1999.
47. M.C. Wahl, M. Sundaralingam, 'Crystal Structures of A-DNA Duplexes', *Biopolymers*, **44**, 45–63 (1997).
48. P.S. Ho, B.H.M. Mooers, 'Z-DNA Crystallography', *Biopolymers*, **44**, 65–90 (1997).
49. M.L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R.E. Dickerson, 'Binding of an Antitumor Drug to DNA. Netropsin and C–G–C–G–A–A–T–T–BrC–G–C–G', *J. Mol. Biol.*, **183**, 553–563 (1985).
50. C.A. Frederick, L.D. Williams, G. Ughetto, G.A. van der Marel, J.H. van Boom, A. Rich, A.H. -J. Wang, 'Structural Comparison of Anticancer Drug–DNA Complexes: Adriamycin and Daunomycin', *Biochemistry*, **29**, 2538–2549 (1990).
51. K. Phillips, Z. Dauter, A.I.H. Murchie, D.M.J. Lilley, B. Luisi, 'The Crystal Structure of a Parallel-stranded Guanine Tetraplex at 0.95 Å Resolution', *J. Mol. Biol.*, **273**, 171–182 (1997).
52. L. Joshua-Tor, F. Frolow, E. Appella, H. Hope, D. Rabinovich, J.L. Sussman, 'Three-dimensional Structures of Bulge-containing DNA Fragments', *J. Mol. Biol.*, **225**, 397–431 (1992).
53. G. Laughlan, A.I.H. Murchie, D.G. Norman, M.H. Moore, P.C.E. Moody, D.M.J. Lilley, B. Luisi, 'The High-resolution Crystal Structure of a Parallel-stranded Guanine Tetraplex', *Science*, **265**, 520–524 (1994).
54. C.-H. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzis, 'Crystal Structure of Intercalated Four-stranded d(C3T) at 1.4 Å Resolution', *Proc. Natl. Acad. Sci. USA*, **91**, 11 636–11 640 (1994).
55. S.R. Holbrook, C. Cheong, I. Tinoco, Jr, S.-H. Kim, 'Crystal Structure of an RNA Double-helix Incorporating a Track of non-Watson–Crick Base Pairs', *Nature*, **353**, 579–581 (1991).
56. A.C. Dock-Bregeon, B. Chevrier, A. Podjarny, J. Johnson, J.S. de Bear, G.R. Gough, P.T. Gilham, D. Moras, 'Crystallographic Structure of an RNA Helix: [U(UA)6A]2', *J. Mol. Biol.*, **209**, 459–474 (1989).
57. W.G. Scott, J.T. Finch, A. Klug, 'The Crystal Structure of an all-RNA Hammerhead Ribozyme: a Proposed Mechanism for RNA Catalytic Cleavage', *Cell*, **81**, 991–1002 (1995).
58. J.L. Sussman, S.R. Holbrook, R.W. Warrant, S.-H. Kim, 'Crystal Structure of Yeast Phenylalanine Transfer RNA. I. Crystallographic Refinement', *J. Mol. Biol.*, **123**, 607–630 (1978).
59. J.H. Cate, A.R. Gooding, E. Podell, K.H. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, J.A. Doudna, 'Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing', *Science*, **273**, 1678–1685 (1996).
60. S.R. Holbrook, S.H. Kim, 'RNA Crystallography', *Biopolymers*, **44**, 3–21 (1997).
61. S.R. Holbrook, J.L. Sussman, R.W. Warrant, S.-H. Kim, 'Crystal Structure of Yeast Phenylalanine T-RNA. II. Structural Features and Functional Implications', *J. Mol. Biol.*, **123**, 631–660 (1978).

Particle Size Analysis: Introduction

Nayland Stanley-Wood
St Peter Port, Guernsey, Channel Islands

1 Introduction **1**
1.1 Historical Background 2
1.2 Selection of an Instrument for Particle Characterization 4
2 Sampling of a Representative Aliquot of Powder **5**
2.1 The Classical Statistical Concept of Sampling 5
2.2 Sampling Reliability 6
2.3 Selection of Sampling Methods 7
2.4 Attainment of a Test Sample 8
3 Characterization of Irregularly Shaped Particles **9**
3.1 Graphical Representation 11
3.2 Particle Averages and Particle Size Distributions 13
3.3 Statistical Analytical Distribution Functions 13
4 The Metrology of Particle Characterization **15**
4.1 Off-line Metrology 15
4.2 In-line and In Situ Particle Characterization 20
5 Shape Characterization of Irregular Particles **20**
5.1 Direct Measurement of Single Particle Shape 21
5.2 Indirect Measurement of Particle Shape 26
6 National and International Standards and Reference Materials **27**
6.1 National and International Standards 27
6.2 Reference Materials 28
List of Symbols **33**
Abbreviations and Acronyms **34**
Related Articles **34**
References **35**

1 INTRODUCTION

Surveys have shown that in the 1980s the production of powdered materials had an annual value in excess of US$10 billion,[1] while more recent surveys showed that an estimated 25–30% of the total output value from the chemical industries was in the physical form of either powdered or agglomerated material.[2–4] There is now an awareness, within many sections of the chemical industry, to the increasing amount of bulk powdered materials – in the form of crystalline solids, agglomerated or granulated powders, and processed or comminuted powdered materials – that are being produced year by year. The characterization of the particulate material produced – together with the handling and behavior of powdered material in bulk quantities – is therefore of paramount importance in the optimization of many multiphasic and multiparticulate processes. The variations in the behavior and dynamics of various types of processing, due to the presence of particulate material and the formation and generation of intermediate or final product particles within manufacturing can, in turn, produce particles with many different chemical and physical characteristics. This can be illustrated using lactose as an example. Lactose can be crystallized at different temperatures to produce particles of either α-lactose (below a crystallization temperature of 93 °C) or β-lactose (above a crystallization temperature of 93 °C). These powders, although similar in their chemical make-up, have different physical properties, which show different compactability behavior due to the differences in surface forms.[5,6] In some circumstances slight changes in operational temperature or rate of mixing, or changes in concentrations of ingredients within the process, can produce particles which are better suited to one end effect than to others. Characterization of powder surfaces by inverse gas chromatography (GC), coupled with the use of computer models of single-crystal slip planes, has resulted in the thermodynamic discrimination of various powdered surfaces and their modification to improve efficacious drug release. With advances in processing and the measurement of a variety of particle characteristics, this should ultimately lead to a situation in which a chemical process could be manipulated to ''tailor-make'' and design particles for a range of desired functions (M. Ticehurst and I. Grimsey, private communication, 1998).

Throughout this section emphasis is placed on the various and diverse methods and techniques for measuring the sizes of irregularly shaped particles. In general, measurement of irregularly shaped particles can be classified into one of three groups (see points 1–3 in the list that follows). An additional group, although

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

not size-related, but still essential in the characterization of powdered materials and in the metrology of particle characterization, and which may be of more relevance when solid–liquid particle interfaces are considered is the parameter of surface area and porosity (see point 4).

1. Direct dimensional measurements, as seen with sieving and microscopy. These techniques are explained in more detail in **Sieving in Particle Size Analysis**.
2. Transport measurements. These techniques are explained in more detail in **Sedimentation in Particle Size Analysis**, and **Centrifugation in Particle Size Analysis**.
3. Rapid physical response measurements. These techniques are explained in more detail in
   (a) **Electrozone Sensing in Particle Size Analysis**
   (b) **Light Scattering, Classical: Size and Size Distribution Characterization**
   (c) **Turbidimetry in Particle Size Analysis**
   (d) **Photon Correlation Spectroscopy in Particle Sizing**
   (e) **Field-flow Fractionation in Particle Size Analysis**
   (f) **Ultrasonic Measurements in Particle Size Analysis**
   (g) **Diffraction in Particle Size Analysis**
   (h) **Optical Particle Counting**
4. Surface area and porosity measurements. These techniques are explained in more detail in **Surface Area and Pore Size Distributions**.

The surface area and porosity characteristics of powders may both be calculated from low-temperature adsorption isotherms obtained by gaseous adsorption. The amount of available surface area is generally measured from the lower relative pressure range of the isotherm, while the degree and magnitude of the porosity within powdered or agglomerated material, in terms of macro- and mesopores, is usually derived from the section of the isotherm that has a relative pressure greater than 0.4. For micropores the whole of the isotherm is commonly used for analysis.

In addition to those articles listed above, there are two other articles within this section; one illustrates the use of **Filtration in Particle Size Analysis** and the other addresses the characterization of **Velocimetry in Particle Size Analysis**.

1.1 Historical Background

The emergence of the modern-day discipline of powder technology and thus the measurement of irregularly shaped particles occurred at approximately the same time in North America, continental Europe and the UK during the late 1940s and early 1950s. However, it was believed by Heywood[7] that powder handling and classification of naturally occurring sands and mineral ores can be traced to Egyptian times.[8] The art of sieving in Elizabethan times, together with the settling of mined ores and the technical development of mineral dressings, has been vividly illustrated by Donald.[9] This book traces the formation of one of the first English companies to manufacture an article (in this case copper) as opposed to solely trading abroad. German scientists were brought over to England to extract and smelt the copper ore extracted from mines located in Cumberland and Cornwall. The techniques of metal dressing used to separate, grade and purify the Elizabethan copper had initially been laid down, a few years earlier, by Agricola.[10] A later publication by Ercker[11] showed how ore could be taken from a furnace, stamped (crushed by water-powered mills) into hazelnut- or acorn-sized pieces (2 cm), separated, refurnaced for further purification, stamped again and then graded into a powder which had particles ''liken to the size of meal'' (150–250 μm).

The separation or mineral dressing of ores probably involved the same method of three dry sieving stages followed by three wet sieving stages as given by Agricola and Ercker. The description of the smaller sieves used was ''a wooden tray with a wire bottom never occupied''.

To place into perspective the amount of effort needed to get mined ores into a suitable state for chemical processing, it was found that the cost of mining, stamping and grading the copper was similar to the furnacing, melting and carriage costs of the final product.

One of the first publications dedicated solely to outlining the various mechanico-physical techniques for particle measurement was written by DallaValle.[12]

This was closely followed by a book by Herdan,[13] which concentrated on the conceptualization of the statistics of particles by the combination of both number statistics and physical statistics, as seen in colloidal systems, to give rise to a new discipline called small-particle statistics. Rumpf[14] was concerned with process technology in terms of the transformation of matter by different mechanisms. In chemical process engineering transformation of matter is deemed to occur by chemical reactions, while in thermal process engineering it is regarded as the transformation of thermodynamically defined phases. In both cases the system parameters of pressure, temperature, and composition follow the laws of thermodynamics.[15] Rumpf,[16] however, suggested that there was a third engineering technology that could characterize changes in the state of disperse systems and of mixtures consisting of particulate matter and continuous fluid phases. Rumpf created an area of academic and industrial interest termed mechanical process technology which was a forerunner

Table 1 Chronological order of some of the early publications and milestones relating to particle sizing and characterization

Author	Title	Publisher	Date
R. Audubert	*The Action of Light upon Suspensions*	Chemical Catalog Company	1926
J.W. McBain	*The Sorption of Gases and Vapours by Solids*	George Routledge and Son Ltd, London	1932
J.M. DallaValle	*Micromeritics. The Technology of Fine Particles*	Pitman Publishing Co.	1943
S. Brunauer	*The Adsorption of Gases and Vapours 1. Physical Adsorption*	Princeton University Press	1943
–	*Symposium on Particle Size Analysis*	Institution of Chemical Engineers	1947
G. Herdan	*Small Particle Statistics*	Butterworth and Co. Ltd	1953
H.E. Rose	*The Measurement of Particle Size in Very Fine Powders*	Constable and Co. Ltd, London	1953
H.C. van de Hulst	*Light Scattering by Small Particles*	J. Wiley and Sons, New York	1957
C. Orr, Jr and J.M. DallaValle	*Fine Particle Measurement. Size, Surface and Pore Volume*	The Macmillan Company, London	1959
R.R. Irani and C.F. Callis	*Particle Size: Measurement, Interpretation and Application*	J. Wiley and Sons, New York	1963
W.E. Kuhn (ed.)	*Ultrafine Particles*	J. Wiley and Sons Inc., New York	1963
R.C. Cadle	*Particle Size: Theory and Industrial Applications*	Reinhold Publishing Corporation	1965
T. Allen	*Particle Size Measurement*	Chapman and Hall, London	1968
D.H. Everrett and R.H. Ottewell	*Surface Area Determination*	Butterworth and Co. Ltd,	1970
B. Kaye	*Direct Characterization of Fineparticles*	J. Wiley and Sons Inc., New York	1981
S. Lowell and J.E. Shields	*Powder Surface Area and Porosity*	Chapman and Hall, London	1984
M.E. Fayed and L. Otten	*Handbook of Powder Science and Technology*	Chapman and Hall Ltd	1991
I.C. Bernhardt	*Particle Size Analysis: Classification and Sedimentation Methods*	Chapman and Hall Ltd	1994
J.P.K. Seville, U. Tuzun and R. Clift	*Processing of Particulate Solids*	Chapman and Hall Ltd	1997

of the discipline now known as powder technology. In reality, although a mechanico-physical unit-processing operation may precede and at times terminate a chemical and thermodynamic unit-processing operation, all three technologies are inter-related and essential to the transformation of matter into a desirable end product. Table 1 highlights some of the publications that have advanced the knowledge and technology of particle size analysis and powder characterization.

One of the first symposia on particle size measurements, which was to supply an intellectual interest and stimulus to the subject of powders and solid particles, was held in London in 1947 under the auspices of the Society of Chemical Industries and the Institution of Chemical Engineers. At the opening of this symposium it was stated that, as far as the President of the Institution was aware, no previous attempts had been made to survey or appraise the difficult subject of particle size analysis.

From that symposium, and with the efforts of a few internationally active powder technologists, an expansion in the field of particle size measurement occurred. This was coupled with a greater understanding and comprehension of particle size distributions, particle shape, particle shape distributions and the surface area of powders.

The disciplines of powder characterization and powder technology, which initially emanated from the UK and continental Europe (Table 1), have now become global in their ramifications and acceptability. The forms of particulate matter generated, fabricated, produced and processed have increased enormously, along with the technologies of measurement, since that 1947 symposium. One of the reasons for the dramatic changes in the types of instruments now available, and the use of a wide range of different physical, optical, electromagnetic, and at times electrochemical measurement systems, is due to the advances in computer technology. The amount of data that can be measured, and the speed at which they can be collected, by new and novel measurement systems connected to plant items, has in recent years opened up new frontiers to the powder technologist.[17] Computer-assisted particle measurement instrumentation not only aids the collection of data but also gives the powder technologist the ability to carry out advanced statistical analysis of data gained from real-time measurement systems. The integration of computers to control and automate processing plants, be they chemical, mechanical or powder bulk handling, is extending the original role of off-line laboratory particle measurements to one in which measurements of particle sizes are made on-line and in-line in real time within a chemical processing plant. The ultimate measurement system, which may not be dependent solely on the measurement of particle

size, would be an in situ array of sensors, within a reactor or processing unit. Such a system could detect and characterize one or more particle parameters to control a dynamic particulate process and produce a specifically desired end product. Current research, which is now multidisciplinary as well as multi-authored, is adopting a wider selection of physical criteria than that initially used for off-line laboratory analysis. Through the use of ultrasonics and optical and tomographic sensing techniques, advances have been made to implement in-line particle characterization in the armoury of the particle/powder engineer.[18] Microelectrical resistance tomography has recently been used to find a solution to two problems that were affecting the advancement of particle characterization and the control of particulate processes. A novel in situ method, using resistance tomography, has enabled the measurement of both the size and the shape of a flowing particulate suspension to achieve what is termed an ''electrical texture fingerprint'' of the flowing particles. Use of similar tomographic sensors has also facilitated the access of multivariate data to enable the control and optimization of multiphase particulate and chemical processes.[19,20]

It should, however, be noted that particle size measurement by itself has little or no meaning, unless the reason for the measurement is also specified.

1.2 Selection of an Instrument for Particle Characterization

One of the most important factors that must be borne in mind when approaching any particle size analysis is that, although there are many ingenious, sensitive, reliable and repeatable techniques, no one method can be regarded as perfect. Tabular procedures and flow diagrams have been published[21] that help with the process of choosing an appropriate particle size measurement technique, but a compromise usually has to be made, in terms of cost and time available, to achieve the best particle characteristic or measurement of particle size.

To these ends several techniques are discussed in this encyclopaedia, to illustrate the diversity that exists and which can now be used by the powder technologist.

The choice of a particle size measurement technique is dependent upon the accuracy of the measurement required and the circumstances to which the measurements are to be applied. This rationale was introduced and emphasized by Heywood,[22,23] who is regarded as a founding father of powder technology in the UK, with the words[24]

> However, it must be realised that particle size analysis is not an objective in itself but is a means to an end, the end being the correlation of powder properties with some process of manufacture, usage or preparation.

This point was also made by Herdan,[13] who stated about fifty years ago that:

> Knowledge of the size characteristics of particulate matter is generally of little value in itself; but particle size measurements are often made to control the quality of the final product, because certain sizes may be correlated with certain desirable properties of the product.

This statement was made at a time when Herdan was interested in the statistics of small particles (particles that were generally classified as being either in the sieve or subsieve size range). In terms of physical dimensions, the lower size limit for a woven wire sieve was then customarily recognized as being 75 μm, while subsieve-sized particles had dimensions below this value.

A particle primarily refers to a object that has a physical boundary regardless of its finite size. Because a particle is a discrete portion of matter that is usually regarded as being small in relation to the space in which it is encompassed, the upper size limit of an individual particle is therefore arbitrary. With this definition, planets, rocks, boulders, aggregates, powdered particles, colloids, molecules, and submolecular and atomic entities may all be regarded as being technologically important to a powder technologist.

For practical purposes, however, an arbitrary definition of the size of powder particles has been chosen[25] so that ''a powder shall consist of discrete particles of any material with a maximum dimension of less than 1000 μm (=1 mm)''.

The lower limit these days, because of the introduction of ultracentrifugation, field flow fractionation, ultrasonics, light scattering and X-ray scattering, is commonly taken to be in the submicrometer to colloidal dimensional range (1 nm) (Table 2). With the advent of ''nanotechnology'' this lower limit is being superseded, and instrumentation is being developed to measure particles of this size.[26,27] Because a particle can be defined and described as a discrete entity of matter it is possible to obtain size distributions of populations of solid, liquid or gaseous particles in solid, liquid and gaseous phases by using an appropriate measurement technique.

Although particle size is an important factor in the determination of how a particle will behave, the measurement of that irregularly shaped particle characteristic may be achieved by several different physical techniques, depending on the final objective of the measurement. It is unusual to measure only one particle, thus particle sizing and powder characterization are generally described in terms of a particle size distribution. According to one glossary of powder technological terms,[25] a particle size distribution may be described in terms of the size and frequency of particles in a population, or alternatively as the functional relationship between a quantity of particles

Table 2 Range of particle sizes

Particle size (μm)	Type of particle
10^5 (100 mm)	Roll compacted or pelletized particles and powders, crushed ores
10^4 (10 mm)	Agglomerated or granulated materials, fertilizers, herbicides, fungicides, instant food products
10^3 (1 mm)	Detergents, crystalline sugars
10^2	Powdered materials, chemicals, cement, flour, pharmaceuticals
10^1	Powdered metals, toners, ceramics
10^0	Pigments, magnetic tape particles, air-borne environmental particles
10^{-1}	Organic pigments, micronized pharmaceuticals, contaminants in water and air
10^{-2} (10 nm)	Carbon blacks, fumed silica
10^{-3} (1 nm)	Diesel fuel pollutants, nanotechnology superconductors

and a dimensional measurement of the irregularly shaped particles.

2 SAMPLING OF A REPRESENTATIVE ALIQUOT OF POWDER[28]

Since off-line quality control of a particulate system is retrospective, there is always the possibility of inaccuracies in solid sampling and nonrepresentation of the bulk material.

The purpose of sampling is to obtain a sample (or samples) from a bulk material that is representative of the bulk material in some desired property, which can be particle size, surface area or chemical composition. A perfect sample from a bulk powder is one which contains all the constituents of the bulk powder in exactly the same proportions as the bulk powder. In other words the bulk powder is perfectly mixed and can be regarded as a homogeneously mixed powder.

If a material is believed to be homogeneously distributed in the bulk, then simple random sampling, using a thief sampler or a spoon to obtain samples from anywhere in the powder, is acceptable. Most materials, however, will not be homogeneous, and methods must be devised to achieve a sampling efficiency akin to that theoretically possible from the random variation to be expected under ideal conditions.

There is, however, always a problem with sampling and the taking of an aliquot from a powder, because any attack on a homogeneously mixed powder will cause particle segregation if the powder is free-flowing and noncohesive. There is also the possibility that even if the powder has a degree of cohesiveness there will be particle movement within the powder mass and segregation will again ensue.

Sampling is therefore an error-generating procedure and no sample provides more than an estimate of the truth or reality of a situation.

In the literature on solids mixing, the word ''sample'' has been given a different meaning from that usually intended in textbooks on classical statistics. In the literature on powder technology there is an established practice of using the term ''sample'' to refer to a number of contiguous particles removed from some position in a bulk powder, rather than to an integral unit as presupposed in classical statistics. A representative sample size must thus contain a certain number of particles. This ensures statistically that the sample taken is representative of the whole powder. For example, if 1 g of powder is taken from an industrial powder which has a particle density of $2\,g\,cm^{-3}$ ($2.0 \times 10^{-3}\,kg\,m^{-3}$) and the mean particle size in the powder is assumed to be 100 μm, then the number of particles present in each 1 g aliquot would be in excess of 10^6 particles.

2.1 The Classical Statistical Concept of Sampling

Because there are always many particles in a sample mass taken from a powder, a common approach is to regard powder sampling as being based on number statistics, as seen in many textbooks on classical number statistics. From the viewpoint of classical statistics the sample taken is visualized as an entire single entity and not as a collection of variously sized particles in that powder sample. In classical statistics, however, it is normal to obtain an estimate of some parameter of the single sample, and therefore obtain an interpretation of the deviation of that individual ''sample'' from a series of ''samples'' that all have a simple distribution of values. In addition an arithmetic average or mean, from all the samples, can be evaluated and deviation from that mean observed.

The range of deviation, degree of dispersion or spread of values around a mean can readily be obtained from number statistics, assuming the initial values are normally distributed about a mean value.

In the case of the probability of a number occurring on the throw of a die or dice the variation and distribution of the numbers could be expressed as a binary or Bernoulli distribution. As the number of events increases, for example when considering a million events or the thousands of particles in a gram of powder, a normal (Gaussian or bell shaped) distribution can be assumed. The representation of such Gaussian distributions can then be expressed in terms of

1. standard deviation, or a measure of the extent of a set of numbers dispersed about a mean;
2. variance, or the square of the standard deviation (regarded as a more common alternative measure of dispersion);

3. coefficient of variation, which is the ratio of the standard deviation to the mean;
4. correlation coefficient, which is the ratio of the standard deviations of two variables.

An additional useful estimator that provides an indication of the accuracy of the average estimate is the standard error, which is defined as the standard deviation of the population divided by the square root of the number of items in the sampled unit.

When these classical statistics are applied to sampling it is assumed that all units of the population have an equal chance of being in the sampled unit.

There are several factors that should be considered prior to subdivision or sampling of a powder.

1. Flowability: An assessment of whether the material is free-flowing or cohesive should be made visually or by pouring it through a funnel with an orifice about 1 cm in diameter. This may be affected by other properties listed here.
2. Maximum particle size: The maximum particle size should be assessed. This will not usually exceed 1000 μm, but sample examination and subdivision procedures can be applied for larger particles.
3. Uniformity: The material should be examined to determine whether it contains particles of widely differing chemical nature, density or size, and an assessment made of its likely behavior during sample division.
4. Friability: The tendency of particles to break or erode during handling or under pressure should be assessed.
5. Moisture: The presence of high levels of moisture should be determined. This might be detected by a tendency to coalesce under pressure. A moisture determination may be of value where hydroscopic or efflorescent materials are concerned.
6. Aggregation: The material may show signs of inherent aggregation either on standing or on handling. The presence of aggregates may be detected by a simple microscopic examination. This will also reveal the presence of very small particles which could affect the behavior of the sample.
7. Chemical nature: A knowledge of the chemical nature of the material is necessary to assess its possible reaction with the atmosphere and the working surfaces. The hazards associated with handling the material should be known.

2.2 Sampling Reliability

In the real world, the particle size distribution in a powder does not always show, or even approximate to, a normal or Gaussian model, so conclusions based on classical statistics with a simple mathematical model may not be valid. The objective of sampling is therefore to obtain one or more samples of the bulk material that are representative of that material in some chosen particular. To ensure accuracy, which can be regarded as a measure of the degree to which a value or estimate approaches a ''true'' value and which denotes an absence of bias,the sampling operation must be free of systematic errors.

To obtain precision, which is the measure of the ability of a procedure to yield the same or near identical results from successive treatments of the same sample material, the random errors should be narrowly dispersed about the mean. A well-designed sampling procedure is therefore one in which errors arising from (1) theoretical differences between the operation and the model, and (2) differences between the actual operation and the model, have been minimized.

These errors, which occur in every sampling procedure, can be separated into those that cannot be altered and those that can be minimized. An error that cannot be altered and is inherent in every sampling technique is a fundamental error (FE). This type of error results from the natural variance or inherent random nature of a powder because of the heterogeneity of the particle population. It possesses two important features:

1. it is the only sampling error that cannot be suppressed, as it occurs in ideal sampling;
2. it can be estimated beforehand.

FE can be mathematically expressed as shown in Equation (1).

$$s^2(\mathrm{FE}) = \frac{Cd^3}{m} \tag{1}$$

where $s^2(\mathrm{FE})$ is the variance of $(a_s - a_b)/a_b$ if a_s and a_b are content values of sample and bulk batch respectively; C is the sampling constant, a function of the material's properties, which can be determined from knowledge of the physical characteristics of the particle assembly – particle shape factor, size distribution and mass fraction of the size class under investigation – as well as the statistical variance of the sampling error desirable in the sampling procedure – m is the sample weight; and d is the size of the square mesh sieve required to retain 5% of the particles or a specified percentage of a specified size range of particles.

FE and its variance can be reduced by increasing the sample weight and reducing the value of d. The total model sampling error can be reduced to FE by taking individual particles at random or from part of a homogeneously distributed bulk.

The segregation error (SE) is the expression of the distribution heterogeneity of the bulk to be sampled. It depends on:

1. constitution heterogeneity (as for FE), because the larger the variation in particle size/mass, the greater the segregation;
2. segregation, or the demixing of differently sized particles into zones that then cause the powder to be nonhomogeneous or demixed;
3. weight of the increments (the greater the increment, the larger the group of particles taken each time).

Gy[29] found that the variance of the SE, s^2(SE), can be expressed as equivalent to a segregation factor times a group factor, where the segregation factor is zero for a homogeneous distribution and the group factor is zero for a single particle. To minimize SE, sampling should be made on a bulk material that is as homogeneously distributed as possible, with increments that are as small as possible (although a certain minimum weight is required to limit operational errors).

Another error that can significantly affect the representation of the bulk material by the sample material is the integration error (IE). IE is caused by the type of sampling strategy used to obtain a mass of sample. This error is introduced by the variability in the bulk and the spacing (in distance or in time) between the increments. Thus to reduce IE one should (1) reduce variability by homogenization of the bulk, and (2) sample in the whole space occupied by the batch (or during the whole time), taking increments as close together as possible.

Because IE is related to the type of sampling undertaken, IE can be minimized by using systematic sampling if the error does not vary periodically, or stratified random sampling if some periodicity is present, and by taking the greatest possible number of increments.

Analysis has shown that all theoretical errors, except one, can be reduced to small values by using homogeneous bulk material, by maintaining a constant rate of flow, and by taking a large number of small increments to form the sample. The exception is due to the heterogeneity of the particle population. This error exists even in ideal sampling, and although it cannot be removed, it can be estimated.

The three types of sample strategy in use are

1. random sampling
2. systematic sampling
3. random stratified sampling.

When the bulk material is not homogeneously distributed, random stratified sampling has been recommended to guard against the possibility of cyclic variations in bulk quality. In a continuous delivery operation, sample increments are taken at random times during successive standard time intervals (strata) spanning the whole process. For static bulks the strata could consist of wagonloads, or bags, or a number of bags.

2.3 Selection of Sampling Methods

The objective in the selection of a specific sampling technique prior to particle characterization is to obtain from the laboratory sample a test sample or test portion that represents the laboratory sample as closely as possible. The practice of sampling is based on the premise that there is a theoretical possibility of dividing a sample into smaller samples or portions without the introduction of biased errors. The collection of data from every sample or position in a sample is expected to embrace the mean value of a characteristic of the sample. It must also permit an estimation to be made of any variation of the measured characteristic throughout the sample.

In any homogeneous sample any aliquot or subdivision of the sample will precisely represent the sample as long as the sampling technique does not impart any systematic bias to the subdivision procedure.

When a laboratory sample is known to be a true homogenized bulk, conceptually, any volume selected, without demixing or the separation of particle sizes, will constitute a perfect sample. However, since any particle has a finite volume and real powders have a distribution of sizes, there has to be a minimum volume withdrawn from a homogeneous mixture to obtain both a representative and measurable sample. With such a homogeneously mixed material the technique of scoop sampling could be used to obtain a test sample or a test portion. To ascertain if the laboratory sample is homogeneously mixed with respect to a given characteristic it has to be established that the mean values of all subdivisions are identical in that characteristic.

To verify this homogeneity, several samples or portions must be chosen at random and tested, which in itself is a sampling technique. In reality no material can be deemed to be homogeneously mixed unless sampled.

To assume a homogeneous mix prior to sampling and then to use the method of scoop sampling as a sampling technique can only be justified if, from past experience, the assumption of a homogeneous material is correct and scoop sampling or coning and quartering gives results that are comparable to more sophisticated sampling methods.

When a heterogeneous mix occurs on subdivision, because particles are impeded in their movement throughout the particle assembly, a concentration of certain sizes or characteristics can occur in local positions. The sampling technique must therefore prevent or overcome segregation of these characteristics into local concentrations.

Heterogeneous material is material in which a spot sample, taken from the laboratory sample, will have a significantly different value of the characteristic under investigation from the mean value of the laboratory sample. Unless it is known otherwise, it is advisable to assume that all laboratory samples are heterogeneous and to use a sampling technique that subdivides the laboratory sample into subdivisions that are all representative of the laboratory sample.

Very rarely will the bulk material be homogeneous with regard to the characteristics to be tested. It is a frequent requirement in specifications that a measurement of this inhomogeneity be made. It is then necessary to take several different definitions and masses of samples; the number and size of which will depend on the quantity and particle characteristics of the material.

British Standard (BS) 3406[28] defines the following powder samples:

1. Gross sample: the final sample obtained or prepared from bulk material.
2. Laboratory sample: sample delivered to a laboratory. May be a gross sample.
3. Spot sample: a sample taken from a specified place in the bulk material. May be a gross or laboratory sample.
4. Test sample: a sample prepared from the laboratory sample and from which the test portions are withdrawn.
5. Test portion: that portion withdrawn from the test sample and entirely used in the test or observation.

The minimum size of sample to be taken to represent the bulk material can be derived by assuming that all errors are zero except that due to the spatial distribution of the differently sized particles themselves.

2.4 Attainment of a Test Sample

In general it is not possible to calculate the sample mass required for a given precision because, of all the sampling errors, only the FE can be determined. Assuming that an ideal sampling situation exists and Gaussian distributions predominate, a relation between the variance, sample size, and component proportions will, however, enable the minimum sample size to be determined (Table 3).[30–34]

Table 3 Mass of coarsest particle (particle diameter 1 mm and unit density)

Mass fraction	Minimum sample mass (g)			
	Coefficient of variation			
	0.01	0.05	0.1	0.15
0.01	523.81	20.95	5.24	2.33
0.02	261.90	10.48	2.62	1.16
0.05	104.76	4.19	1.05	0.47
0.08	65.48	2.62	0.65	0.29
0.1	52.38	2.10	0.52	0.23
0.2	26.19	1.05	0.26	0.12

The homogeneity definition requires specification of the size of the test sample or test portion because homogeneity implies that the variability between samples or portions is either negligible or too small to be measured. This should not be taken to imply that homogeneity persists as material is further subdivided. An optimum sample or portion volume must therefore be used which is capable of giving the necessary information about the material it represents.

2.4.1 Free-flowing Powders

With a free-flowing powder there is a greater tendency for segregation to occur than with non-free-flowing, cohesive powders. It is therefore advantageous to sample from a moving stream. This can be accomplished by using a rotary sampler. Various designs exist in this category of sample dividers, but most consist of one or more receivers placed on a turntable that rotates at a specific speed. This type of equipment is thus ideal for free-flowing nonfragile material, but at times can be used for non-free-flowing materials. The abrasion caused by the passage of particles through the sample divider may have an adverse effect on fragile material. A chute riffler can also be used for free-flowing powders.

2.4.2 Non-free-flowing Powders

Although in some cases it may be possible to sample non-free-flowing material with a rotary sample divider or a riffler, in most cases it will be necessary with non-homogeneous and non-free-flowing powders to premix the powder. With such powders it may be advantageous to use a wet sampling technique. The coning process of the coning and quartering technique[28] is to ensure an adequately mixed sample prior to subdivision. The degree of mixing obtained by this technique can in some cases (depending on the chemical and physical nature of the powder), be enhanced by the presence of a small amount of liquid. Continuous handling of the solid is not recommended for fragile particles.

The actual operation of sampling a material stream requires care in design, to ensure that the full cross-section of the stream is sampled on every occasion, that no mechanism is in operation that preferentially accepts or rejects pieces of certain sizes, and that no loss of sample or gain of adventitious material occurs (Figure 1 and Table 4).

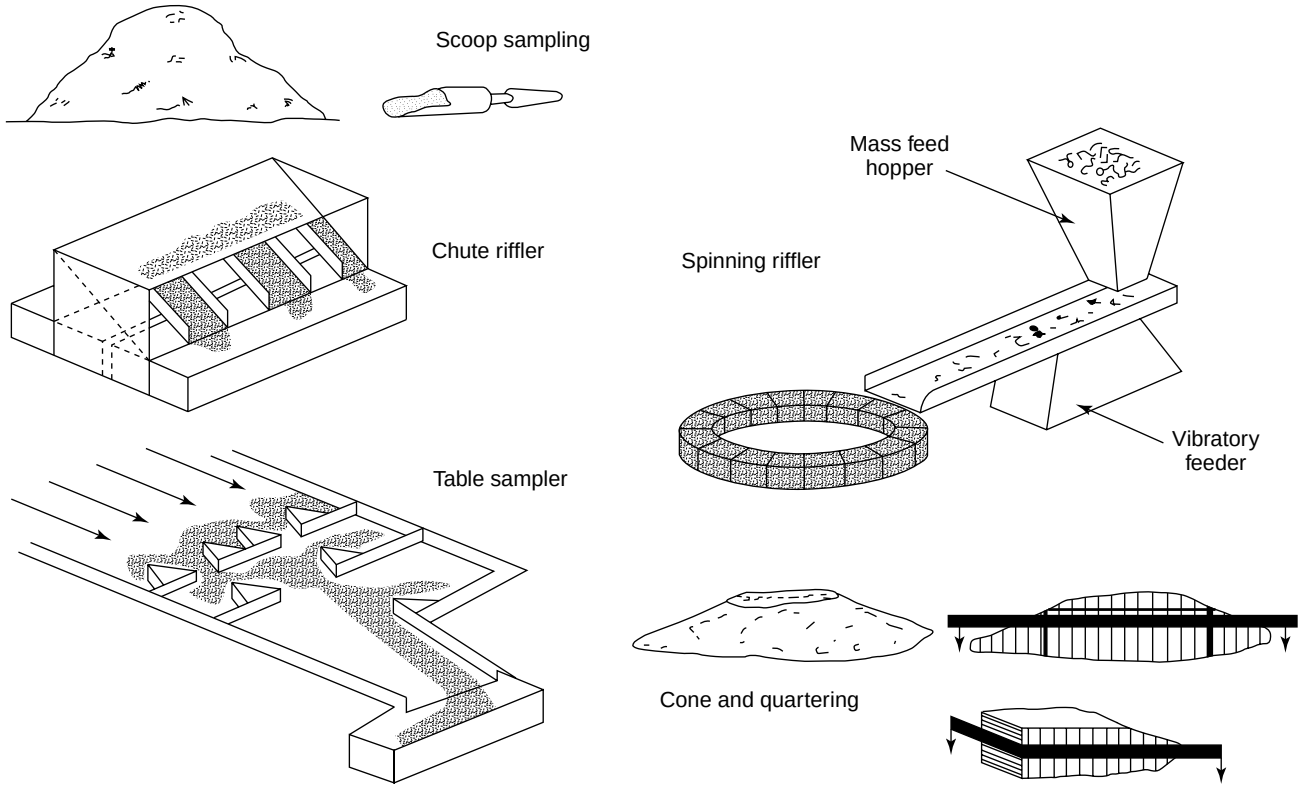

Figure 1 Selection of sampling techniques used to obtain a representative sample.

Table 4 Reliability of powder sampling methods

	Coarse/fine sand[a]		Sugar/sand[b]	
	Standard deviation (%)	Coefficient of variation[c] ($\%^2$)	Standard deviation (%)	Coefficient of variation[c] ($\%^2$)
Cone and quartering	6.81	11.4	5.76	9.6
Scoop sampling	5.14	8.6	6.31	6.5
Table sampler	2.09		2.11	
Chute riffler	1.01	1.7	1.10	1.8
Spinning riffler	0.125	0.21	0.27	0.45
Random variation	0.076	0.13	0.093	0.16

[a] Coarse sand, 420–500 μm; fine sand, 150–250 μm, density 2.65 g cm^{-3} for both.
[b] Fine sand as above, sugar, 420–500 μm, density 1.64 g cm^{-3}.
[c] Powder sampling according to BS 3406.[28] Coefficient of variation = standard deviation/mean value.

3 CHARACTERIZATION OF IRREGULARLY SHAPED PARTICLES

In many powder handling, powder production and powder related industries the sizing of particles is traditionally achieved by retrospective physical measurements of the sampled powdered material – off-line laboratory analysis. Off-line sizing of irregularly shaped particles and intermediate primary particles, from either batch or automated production, can only be confidently achieved when a representative sample has been taken from the process. The range of sizes measured off-line by modern instrumentation covers the majority of sizes encountered, from high-throughput, low-value products to low-throughput, high-value ''high tech'' products.

There is no single definition of particle size, as different methods of particle size analysis are based on a wide variety of different physical criteria.

In many textbooks (see Table 1) the particle size of an irregular solid is described in terms of the physical

criterion used to measure the particle's dimensions. Thus, in the past, if the method used to size particles was a mechanical sieving technique, the particle dimension was called the sieve size and had the symbol d_t, while if a sedimentation technique was used, the particle dimension obtained could be calculated and described in several ways, including as a free fall diameter, d_f, a drag diameter, d_d, or a Stokes diameter, d_{St}. ISO (International Standardization Organization) Standard 9276-1[35] now recommends that the symbol x be used for particle size and that a subscript denote the equivalent diameter measured by one of the physical measurement techniques. For example, the electrosensing zone method measures the volume of a particle and this can be identified by the symbol x_v, which is an equivalent volume diameter. Likewise x_s is equivalent surface diameter, x_w is equivalent settling diameter at either high or low Reynolds numbers, and x_n is equivalent mesh diameter.

This, as can be imagined when only the term ''particle size'' was used, has relived some of the ambiguities that could and did lead to confusion when any attempt was made to compare the particle size obtained by one technique with a size obtained by a different physical technique. There is no physical or theoretical justification why any two irregular particle sizes measured by two different techniques should be comparable. A sieve size is not equivalent to a drag or Stokes diameter. This nonequivalence can be advantageous when used to describe the shape of particles, but not when trying to achieve a consensus on the dimensions of an irregularly shaped particle.

Table 5 shows seven arbitrary physical classifications of particle sizing techniques that have been used to measure particle size. Also shown is the descriptive diameter ascribed to the particles measured by that technique. The symbols and nomenclature for particle size and the methods used to obtain a measurement of irregularly shaped particles have recently been re-symbolized. Table 6 shows the new ISO-approved nomenclature. To overcome the confusion that can occur when ''particle size'' is not fully notated, an Expert Group under the auspices of the Bureau Communautaire de Réference (BCR),[36] undertook a certification campaign to measure five powdered reference materials in the size range 0.1–1000 μm and to present the measured particle sizes graphically. It was consensually agreed that the classical treatment of size would be based not solely on physical criteria but on the concept of an equivalent sphere diameter and that the traceable fundamental units of length, mass and time would be used.

For the purpose of these BCR materials the size of the irregular particles was defined as the diameter of a sphere which is equivalent to the particle with respect to either its volume (volume diameter) or to its settling rate in a viscous fluid (Stokes diameter).

A sphere can be measured and defined by one single diameter, a cube by one length dimension and an oblongated body measured by in terms of length, breadth, and thickness. Since an irregularly shaped particle can have various lengths, breadths and thicknesses dependent upon the position of the measurement, it cannot therefore

Table 5 Off-line size-dependent physical criteria for the measurement of irregularly shaped particles

Physical criterion	Diameter measured	Size range applicability (μm)
Direct dimensional measurements		
Image 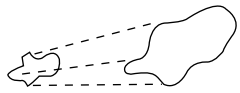	Projected area	Optical: 0.8–800 Electron: 0.002–15
Mechanical 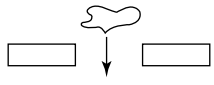	Sieve diameter	Dry: 40–1000 Wet: 1–40
Transport measurements		
Dynamic 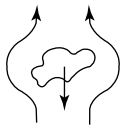	Free fall diameter Drag diameter Stokes diameter	Gravity: 1–100 Centrifugal: 0.05–25
Photo-zone	Projected area Stokes diameter	X-ray: 0.05–100 Centrifugal: 0.05–50
Rapid physical response measurements		
Light scattering 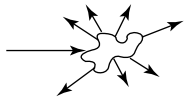	Surface volume diameter	
Electro-zone 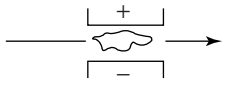	Volume diameter	
Topography measurements		
Surface 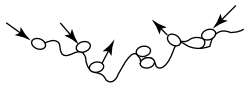	Surface diameter Surface volume diameter	Permeametry: 0.1–50 (mean) Adsorption: 0.005–50 (mean)

Table 6 International standard nomenclature for particle sizes

Symbol	Name	Definition	Formula
x_v	Volume diameter	Diameter of a sphere having the same volume as the particle	$V = \pi x_v^3/6$
x_s	Surface diameter	Diameter of a sphere having the same surface area as the particle	$S = \pi x_s^2$
x_{sv}	Surface volume diameter	Diameter of a sphere having the same external surface area to volume ratio as the particle	$x_{sv} = x_v^3/x_s^2$
x_d	Drag diameter	Diameter of a sphere having the same resistance to motion as the particle in a similar fluid (for small particles $x_d = x_s$)	$F_d = C_d A \rho_f v^2/2$
x_{St}	Stokes diameter	Diameter of a sphere having the same free-fall velocity as the particle in the laminar region	$F_d = 3\pi x_{St} \eta v\, x_{St}^2 = x_v^3/x_d$
x_a	Projected area diameter	Diameter of a circle having the same area as the projected area of the particle in a stable position	$A = \pi x_a^2/4$

be described or defined by a single or unique dimension and its size may be expressed in terms of the diameter of a circle or a sphere which is equivalent to the particle in some stated particle property (Figure 2). The particle size of an irregularly shaped particle has now been consensually denoted and defined globally[35] as "the diameter of a sphere having the same physical properties: this is known as the equivalent spherical diameter".

There are other definitions for particle size, such as those based on the openings of a sieve or on a statistical diameter such as a Feret's diameter which is measured by image analysis (Figure 3). It should, however, be mandatory that whenever the size of a particle is stated an indication is given of how that particle size was defined and measured. The symbol designated to represent the size of the particle diameter or equivalent spherical diameter in the past has always been "*d*". In many technological publications this is being superseded by the symbol "*x*" because ISO 9276-1[35] has recommended that "*d*" now be replaced by "*x*" to denote an equivalent spherical diameter or a diameter of a sphere. Thus $x_s = (d_s)$ = equivalent surface area diameter and

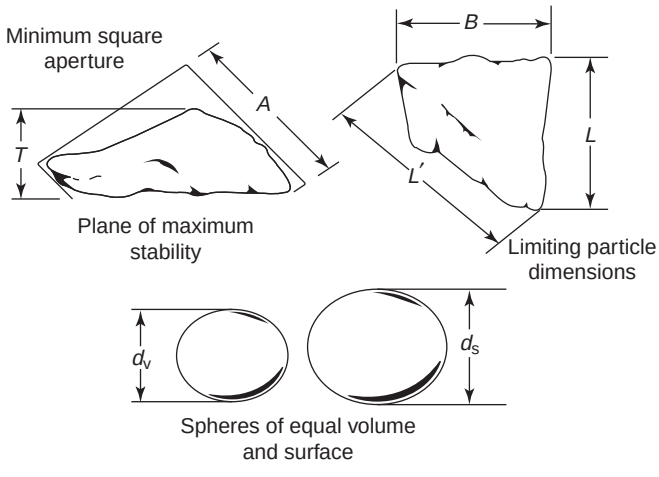

Figure 2 Equivalent particle diameters (after Heywood[23,94]).

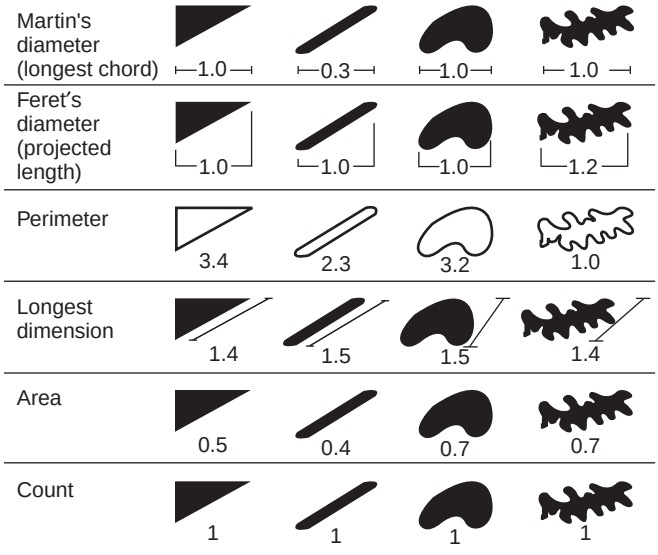

Figure 3 Alternative definitions of particle size by image analysis.

$x_v = (d_v)$ = equivalent volume area diameter, although this change in nomenclature is not mandatory.

3.1 Graphical Representation

Most modern instruments present the measurement results in graphical form, but at times it is necessary to express these results quantitatively so that only a single representative measurement is used and introduced into some particle system correlations and particulate models. Some of the simplest correlations reduce a particle size distribution to a single parameter. The statistical parameter calculated can be a mean size, a median size, or a modal size. These single averaged parameters can thus convey, by a single number, the order of the particle size involved within a distribution, but such a single parameter gives no indication of the spread of the distribution. The mean size is the average of the distribution. The median size, where half of the particles

are smaller and half are larger than the median size, is the midpoint of the distribution and is often denoted by the symbol x_{50}. Finally, the modal size is the size that occurs most frequently within the distribution and is thus the location of the maximum quantity within the distribution. Powders can, however, have more than one modal value and particle size distributions can be mono-, bi-, tri- or even multi-modal.

As powders are assemblies of particles, a powder can be characterized in terms of its distribution of particle sizes by a statistical parameter that is chosen to typify both the type of distribution and the average particle size within that distribution.

The distribution is usually determined either as a number distribution or as a distribution by weight, and the information may be displayed graphically either as a histogram or as a continuous distribution curve. ISO 9276-1[35] recommends that measurement of a distribution be graphically displayed as either a discrete density distribution, q, or a cumulative distribution, Q (Figure 4). A cumulative distribution curve is used most often because the curves can be more easily interpolated and normalized. The word ''density'' denotes a statistical parameter akin to the frequency of an event in number statistics and should not be confused with the physical property of mass per unit volume.

Since a powder consists of many particles which have a range of sizes it is possible to mathematically describe that range of sizes using Equation (2)

$$Q_r(x) = \int_{x_{\min}}^{x} q_r(x)\,\mathrm{d}x \tag{2}$$

where x represents a particle size, $Q_r(x)$ is the cumulative distribution, and $q_r(x)\,\mathrm{d}x$ the fraction of particles in the size range $(x - \mathrm{d}x)/2$ and $(x + \mathrm{d}x)/2$. The subscript r describes the type of distribution as well as the type of diameter measured. Thus when $r = 2$ the distribution is by area and the diameter measured is equivalent to a projected surface (x_s), while when $r = 3$ the diameter is a volume diameter (x_v), as determined by an electrosensing technique (Figure 5). It is, however, customary to normalize particle size distribution so that the total area of the distribution is unity, by imposing the condition shown in Equation (3)

$$Q_r(x) = \int_{x_{\min}}^{x_{\max}} q_r(x)\,\mathrm{d}x = 1 \tag{3}$$

where $x_{\max}$ and $x_{\min}$ are the largest and smallest particle sizes in the distribution.

The fraction of particles in the term $q_r(x)\,\mathrm{d}x$ can be expressed in several ways. If the particle size measurement technique involves sieving then the graphical representation would be by weight (for constant particle density (mass/volume) the distribution becomes one

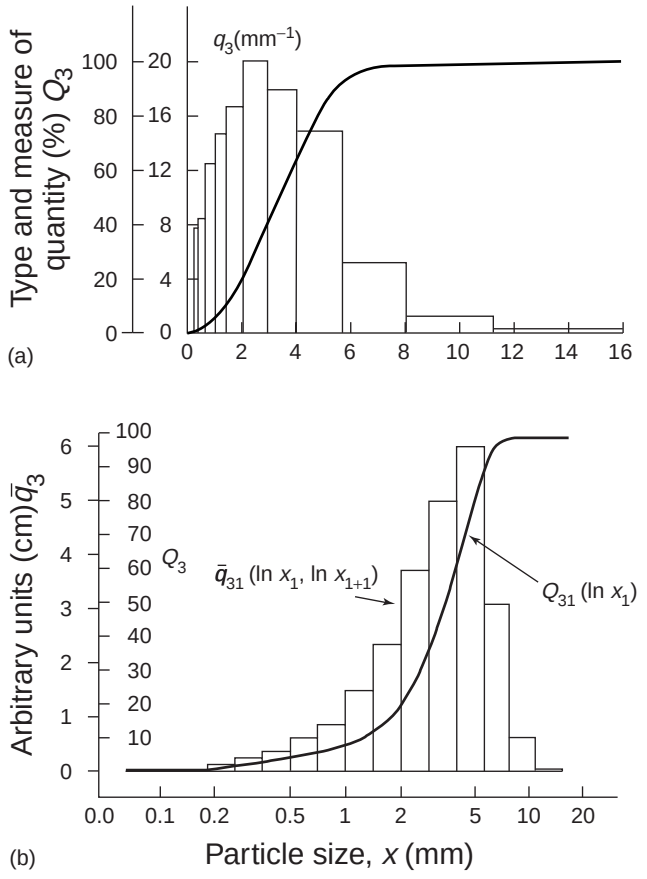

Figure 4 Graphical representation of particle sizes.[35] (a) Cumulative and density distributions; (b) cumulative and density logarithmic distributions. $\bar{q}_0, \bar{q}_1, \bar{q}_2, \bar{q}_3$, density distribution by number, length, surface or projected area, volume or weight; Q_0, Q_1, Q_2, Q_3, cumulative distribution by number, length, surface or projected area, volume or weight; $\bar{q}_0, \bar{q}_1, \bar{q}_2, \bar{q}_3$, density distribution by number, length, surface or projected area, volume or weight with logarithmic abscissa. (Reproduced with permission from BSI and ISO monographs.)

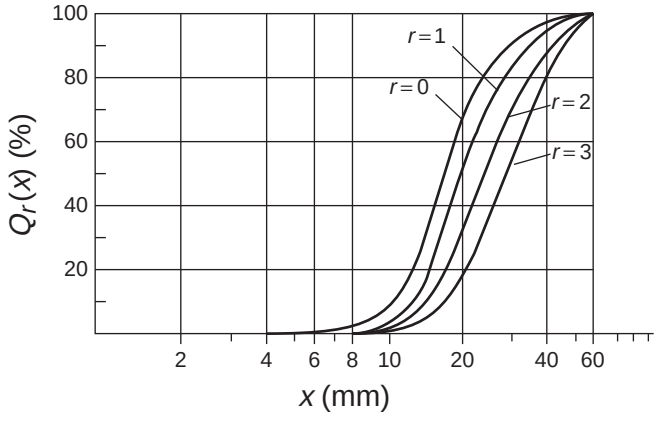

Figure 5 Statistically defined particle size distributions. Number, $r = 0$, length, $r = 1$, area, $r = 2$; volume or mass, $r = 3$.

Table 7 Statistically defined particle size distributions[35]

Type of distribution	Measurement of distribution of particle size, x	
	Density distribution	Cumulative distribution
General symbol, $r = r$	$q_r(x)$	$Q_r(x)$
Number, $r = 0$	$q_0(x)$	$Q_0(x)$
Length, $r = 1$	$q_1(x)$	$Q_1(x)$
Area, $r = 2$	$q_2(x)$	$Q_2(x)$
Volume or mass, $r = 3$	$q_3(x)$	$Q_3(x)$

which can be equated to a volume distribution, thus $r = 3$). If the particle size is measured by image analysis then the representation would be by number ($r = 0$) or area ($r = 2$). With a light-scattering particle size measurement technique an area distribution would be appropriate.

Table 7 shows the types of particle distribution that can be statistically defined (number, length, area and volume or mass) together with the terminology used to define density and cumulative distributions.

3.2 Particle Averages and Particle Size Distributions

To unambiguously describe the average size of the particles in an assembly or distribution of particles or powder, two statistical criteria are necessary. These criteria can be denoted by two subscripts attached to the symbol (x) for particle size. One subscript is related to the type of statistical distribution (r) and the other is related to the method used to calculate the average size of the particles (k) within the distribution.

The general description of an average particle size within a powder is thus dependent upon both the type of size distributions and the method of calculation from moments of the particle size mean. The moment of a distribution can be represented by the letter M qualified by two subscripts, where $M_{k,r}$ is the complete kth moment of a $q_r(x)$ density distribution,[37] as shown in Equation (4).

$$M_{k,r} = \int_{x_{\min}}^{x_{\max}} x_r q_r(x)\,\mathrm{d}x \qquad (4)$$

The first subscript denotes the power of the moment and the second denotes the power of the distribution function.[38,39]

This elegant nomenclature is not often used in practice. It is, however, extremely useful as it illustrates that a distribution can have an infinite number of ''average'' sizes.

In particle size analysis particulate matter is often characterized and based on a representative sample of the powder population, and the particle size is then related to some other important physical property such as strength, flowability or solubility. A more meaningful correlation may, at times, be obtained if a specific average particle size (from the range of averages that could be obtained using moments of a distribution) is derived or calculated from the measured distribution of particle sizes, and this specific average size can then be related to a process variable. Equations for the calculation of average particle sizes or average particle diameters from a given particle size distribution are defined in ISO 9276-2[39] to give a unique definition of average size, derived from the moments of a size distribution. As more than one moment can be calculated from a distribution, a range of averages can therefore be calculated from a measured particle size distribution. However, not all of these theoretically possible averages have a practical use in powder technology.

From the moments in a distribution, the average particle diameter can be expressed as $x_{k,r}$, where r is type or quantity of the distribution (Table 7) and k is the power of the moment of size x or the kth moment of the density distribution. Several combinations of the values of r (0 to 3) and of k (3 to -3) can therefore be obtained, which shows the wealth of averages that can be calculated. Table 8 illustrates the method of calculation of averages and weighted averages and shows some of the more common averages used in powder technology industries.

3.3 Statistical Analytical Distribution Functions

An alternative approach to describing a particle size distribution in terms of moment averages is to fit the distribution curve into an analytical function. Some of the most commonly used functions are the Gaudin–Schuman distribution, the log-normal distribution, and the Rosin–Rammler distributions. In each case

Table 8 Calculation of averages and weighted averages

Average particle diameter[a]	$x_{k,r}$
Arithmetic average particle diameter[b]	$x_{k,0}$
Weighted average particle diameter[c]	$x_{1,r}$

[a] Depending upon the chosen number of the subscript ($k = -3, -2, -1, 0, 1, 2, 3$ and $r = 0, 1, 2, 3$) different average particle diameters can exist. The arithmetic mean particle diameters are, however, always based on a number density distribution, so $r = 0$.

[b] Recommended arithmetic average particle diameters are $x_{1,0}$ for arithmetic average length, $x_{2,0}$ for arithmetic average surface area, $x_{3,0}$ for arithmetic average volume.

[c] k is the power of the dimension used to characterize or measure the particle size, x, which can be a volume (x^3), a surface area (x^2), a length (x^1), a number (x^0), a reciprocal length (x^{-1}), a reciprocal surface (x^{-2}) or a reciprocal volume (x^{-3}). The recommended weighted average particle diameters are $x_{1,0}$ from number density, $x_{1,1}$ from length density, $x_{1,2}$ from surface density and Sauter diameter, $x_{1,3}$ from volume density.

the equation contains two constants, one of which is a characteristic of the particle size while the other is dependent upon the width of the size distribution. Although these functions are sometimes convenient to use, they should not be extrapolated outside the range of the actual measurements.

3.3.1 Gaudin–Schuman Distribution

The Gaudin–Schuman distribution is shown in Equation (5).

$$Q_r(x) = Q_3(x) = \left(\frac{x}{x_{\max}}\right)^m \tag{5}$$

This expression is included because it is an example of a simple power law. Plotted on a logarithmic scale the distribution is a straight line. The characteristic size is that of the largest particle, $x_{\max}$. The width of the distribution or decrease of scatter is signified by the slope, m. The wider the distribution, the smaller the slope.

3.3.2 Rosin–Rammler–Sperling–Bennet Distribution

The Rosin–Rammler–Sperling–Bennet distribution is shown in Equation (6).

$$R(x) = 1 - Q_r(x) = 1 - Q_3(x) = \exp\left(-\left(\frac{x}{x'}\right)^n\right) \tag{6}$$

This distribution is based on an exponential function and thus tends to be applicable in a process where many random events occur consecutively. It is often used to describe the products of a comminution process. The function can be plotted as a straight line on a natural log of a natural log (ln ln) versus natural log (ln) scale. By taking natural logs twice, Equation (7) is obtained

$$\ln\ln\left(\frac{1}{R}\right) = n\ln x - n\ln x' + \ln\ln \mathrm{e} \tag{7}$$

where $\ln \mathrm{e} = 1$ and $\ln\ln \mathrm{e} = 0$.

The characteristic size x' is the size of x corresponding to a cumulative percentage $Q_3(x)$ equal to 63.2% undersize or an oversize of 36.8%. The width of the distribution or decrease of scatter is signified by the slope, n.

3.3.3 Log-normal Distribution (Gaussian/Laplace/De Moivre)

The normal distribution can be described by Equation (8).

$$q(z) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(\frac{-z^2}{2}\right) \tag{8}$$

This is the Gaussian normal distribution, which describes a completely random distribution of sizes around some mean. In a distribution which has a linear abscissa for particle size then z becomes equal to $(x - x_{50}, r)/2$ and for a logarithmic abscissa for log sizes z becomes equal to $\ln(x/x_{50}, r)$. The Gaussian curve can be useful for some narrow, naturally occurring distributions. Usually, the size distributions of powders are too wide for this function, and the log-normal distribution must be used. If this distribution is applicable a plot of probability versus ln size will yield a straight line. The width of the distribution or decrease of scatter can be signified by the slope, s. If the distribution has an arithmetic or normal distribution then $s = (x_{84} - x_{50}) = (x_{50} - x_{16})$, where x_{84}, x_{50} and x_{16} are the values of size at which $Q_r(x)$ is 84%, 50% and 16%, respectively.

With a log-normal distribution $s = \ln(x_{84}/x_{50}) = \ln(x_{50}/x_{16})$.

To obtain an evenly spaced graphical representation of a wide distribution of sizes, which will ensure the "noncrowding" of sizes together, the size axis can be changed from an arithmetic scale to a logarithmic scale. The degree of resolution of size, or space between points, is shown in Table 9. With the arithmetic scale the distance between the points decreases with an increase in particle

Table 9 Degree of resolution for arithmetic and geometric progressions

Arithmetic progression									
Size	1	2	3	4	5	6	7	8	9
Class interval	–	1	1	1	1	1	1	1	1
Mean class size		1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5
Degree of resolution	–	0.66	0.40	0.28	0.22	0.18	0.15	0.13	0.12
Geometric progression									
Size	1	$\sqrt{2}$	2	$2\sqrt{2}$	4	$4\sqrt{2}$	8	$8\sqrt{2}$	16
Class interval	–	$\sqrt{2}-1$	$2-\sqrt{2}$	–	–	–	$8-4\sqrt{2}$	$8\sqrt{2}-8$	$16-8\sqrt{2}$
Mean class size		$(\sqrt{2}\times 1)^{1/2}$	$(2\times\sqrt{2})^{1/2}$				$(8\times 4\sqrt{2})^{1/2}$	$(8\sqrt{2}\times 8)^{1/2}$	$(16\times 8\sqrt{2})^{1/2}$
Degree of resolution	–	0.347	0.348	–	–	–	0.348	0.348	0.348

size, while with a logarithmic scale the distance between points remains uniform.

In current particle and powder processing the required particle mixtures may be elaborate and thus far more complex than can be described by a two-parameter model. Modern computing techniques do, however, enable the easy solution of numerical equations and the computer can accept the distribution function as a series of numbers which can then be readily visualized graphically on a computer screen.

4 THE METROLOGY OF PARTICLE CHARACTERIZATION

In considering particle-measuring techniques, the first major distinction that should be made is between direct and indirect methods of measurement. If the particles can be directly observed by an optical microscope or an electron microscope then an indefinite amount of information can be obtained from the image of the particles. Frequently, there is a limitation on the number of particles that can be analyzed in a reasonable time, so image analysis techniques are usually used for more sophisticated experimental and research analyses. In principle, any physico-chemical phenomenon which is dependent upon particle size can be used to measure a particle dimension. In practice it is more convenient if the relationship between the phenomenal mechanism and the particle size is known. Currently there is a wide range of commercially available equipment, but it is based on relatively few physical principles. Nanotechnology has been used in the production of electrospray-dried biomolecules[40] and thermoelectrically and electronically configured nanofilms and nanocomposites.[41] The advent of this technique, together with concerns over the emission of nanosized particles and microsized particulate masses from low-emission diesel engines,[42] and the introduction of new legislation on air quality,[43,44] has meant that other size-dependent physical properties may have to be used for particle size characterization and incorporated into new and novel measurement systems.[45] Air pollution regulations do not apply only to the automotive industries since National Ambient Air Quality Standards are mandatory for most powder producers and users.[46]

New physical principles, such as those used in differential mobility analysis, small angle X-ray scattering and ultrafine condensation nucleus counting, may have to be considered for the future measurement of nanoparticles as well as the modification of accepted metrology to discriminate between, measured and monitored micrometer and nanometer particle sizes.

The methods that are commonly used these days for characterizing industrial powders, both off-line and in some cases on-line, are:

- sieving
- sedimentation
- electrical sensing zone techniques
- light scattering
- ultrasonic attenuation.

The first three are established off-line techniques that have been thoroughly tested and developed, and from which reproducible and replicatable results can be obtained. They are, however, less amenable to automation and on-line application. The last two techniques are at present generally off-line measurement techniques but are amenable to further technological modifications for use in automated powder processing plants.

4.1 Off-line Metrology

Off-line metrology, as seen in modern particle size instrumentation (Table 5), can be categorized as shown in the following sections.

4.1.1 Direct Dimensional Measurements

These measurements are typified by the microscopy and sieving techniques that are generally applied to particulate systems in the 10–25 000 μm size range (although the maximum particle size of a powder is generally accepted to be below 1000 μm). The lower size limit can be extended to the 1 μm level for both techniques.[47] Particles undergoing size analysis by sieving have to be robust to withstand the sieve action, which can cause attrition when a ''dry'' sieving technique is used. For fragile particles there is a ''wet'' technique, in which water or organic solvents are used to support delicate materials, and when either wire or electroformed sieves are used.[48–50] The particles must be insoluble in the suspending fluid.

4.1.2 Transport Measurements

Metrology based on the transport phenomena of particles, as seen with Stokesian sedimentation and elutriation, covers a particle size range from 10^{-3} to 10^{-6} m (1000 μm to 1 μm) when the force causing particle settling is gravity. The sizes measured by centrifugal separation are smaller and generally in the 0.05–25 μm range (Table 5). Recently it has been argued that the lower particle size limits for gravitational sedimentation can theoretically be specified in terms of height and time of fall, rather than relying on the vague statement that the Reynolds number should be less than 0.25. The lower particle size limits for gravity

sedimentation were found to be inversely proportional to the cube root of the measurement height and inversely proportional to the fifth root of the measurement time.[51] Convection currents will, however, always exist within a particulate gravitational suspension system with particle sizes of less than 1–2 μm. Thus a protracted measurement time, to achieve a supposed measurement of particles below 1 μm, tends to be suspect because of mass flow and possible temperature variations which cannot, with any degree of certainty, be measured. It is therefore essential that there is an increase in the degree of temperature control when the sizes of particles being analyzed by gravity sedimentation decrease below 50 μm.

4.1.3 Rapid Physical Response Measurements

Rapid response particle size analysis can be achieved with electrozone sensing, Fraunhofer diffraction and nephelogenic (light scattering) techniques. In diffraction and light scattering instruments the intensity of light scattered by the particles, usually in suspension, is recorded as a function of the scattering angle. The light scattering models used to determine particle size from rapid response instruments depend on Fraunhofer diffraction, Mie theory or ray optics.[52,55] The size range measured with this group of instruments, unlike the unlimited size range of the direct measurement group, tends to be between 10^{-3} m (1000 μm) and 10^{-7} m (0.1 μm). Colour absorbance coupled with radiation absorbency may cause some inaccuracies in photozone sensing. With light scattering instrumentation, which is dependent upon the computational models that deconvolute the measured scattered light intensities falling onto photodetectors, it is always assumed that the light scattered emanates from spherical particles. Thus either an under- or an overestimation of the finer particle sizes in a distribution may occur because of either the computer model or the insertion of incorrect particle and suspending fluid physical data into the computer program. The anomalies seen in laser light scattering instruments may at times be due to instrument misalignments, but the main source of error is due to incorrect choice and application of the computational model used for different types of particles, refractive indices or particle size distributions. The electrozone sensing technique can give size measurement distortions with highly conductive or very porous materials, because of incorrect volume displacement.

When the laser diffraction technique was first introduced, the scattering angles were less than 14°, but now instrumental technology has advanced and the ISO monographs (Table 10) have divided laser light scattering techniques into two or more categories. Low-angle laser light scattering (LALLS) instruments limit the minimum particle size that can be measured by this technique to 1 μm. The laser technique has now been broadened to capture light scattered over a much wider angular range. Other improvements include the placement of light detectors at various angles to the incident beam and the use of lasers with different wavelengths of light and polarization. Many recent instruments now have angles up to and at time greater than 150°. With the application of Mie theory[54] as well as the approximate theories of Fraunhofer and anomalous diffraction,[52] particle sizes down to the sub-micrometer range or less (<0.1 μm) can readily be analyzed with converging beams and more than one focusing lens. Since all the theories used are based on the scattering of light from spherical bodies, irregularly shaped particles in the nanometer size range can theoretically be measured. The measured/computed size for nonspherical particles is the equivalent sphere size, which is an equivalent cross-section generally termed the Sauter diameter (or in ISO terminology the $x_{1,2}$ (average length-surface)) or the average volume-surface diameter.

4.1.4 Topography, Surface Area and Porosity

In many powder technology industries, such as pharmaceuticals, foods, ceramics and cosmetics, information on a range of different particle and powder characteristics which can be used for discriminatory tests on resources – to measure the purity of solid materials and as aids to facilitate product improvement and development – as opposed to the particle and bulk characteristics

Table 10 Work groups (WG) within ISO committee TC 24/SC 4

Work group	Publications
WG 1:	Representation of analysis data
	Representation of results of particle size analysis – Part 1: Graphical representation (published as BS/ISO 9276-1: 1998 and ISO 9276-1: 1995)
	Representation of results of particle size analysis – Part 2: Calculation of average particle sizes/diameters and moments from particle size distributions (published as ISO 9276-2: 1999)
	Representation of results of particle size analysis – Part 3: Graphical representation of particle size plots
	Representation of results of particle size analysis – Part 4: Characterization of a classification process used for particle size analysis (at FDIS (Final Draft International Standard) stage)

Table 10 (*continued*)

Work group	Publications
	Representation of results of particle size analysis – Part 5: Validation of calculations relating to particle size analysis using the logarithmic probability distribution (at DIS (Draft International Standard) stage)
WG 2:	Sedimentation, classification
	Determination of particle size distribution by gravitational sedimentation methods – Part 1: General principles and guidelines (ISO 13317-1, at FDIS stage) Determination of particle size distribution by gravitational sedimentation methods – Part 2: Fixed pipette method (ISO 13317-2, at FDIS stage) Determination of particle size distribution by gravitational sedimentation methods – Part 3: X-ray method (ISO 13317-3, at FDIS stage) Determination of particle size distribution by gravitational sedimentation methods – Part 4: Photosedimentation method (ISO 13317-4, draft) Determination of particle size distribution by gravitational sedimentation methods – Part 5: Sedimentation balance method (ISO 13317-5, draft) Determination of particle size distribution by centrifugal liquid sedimentation methods – Part 1: General principles and guidelines (ISO 13318-1, at FDIS stage) Determination of particle size distribution by centrifugal liquid sedimentation methods – Part 2: Pipette method (ISO 13318-2, draft) Determination of particle size distribution by centrifugal liquid sedimentation methods – Part 3: X-ray method (ISO 13318-3, at DIS stage) Determination of particle size distribution by centrifugal liquid sedimentation methods – Part 4: Photocentrifuge method (ISO 13318-4, draft)
WG 3:	Pore size distribution, porosity
	Pore size distribution and porosity of materials – Evaluation by mercury porosimetry and gas sorption – Part 1: Mercury porosimetry (ISO 15901-1, draft) Pore size distribution and porosity of materials – Evaluation by mercury porosimetry and gas sorption – Part 2: Analysis of meso- and macropores by gas sorption (ISO 15901-2, draft) Pore size distribution and porosity of materials – Evaluation by mercury porosimetry and gas sorption – Part 3: Analysis of micropores by gas sorption (ISO 15901-3, draft)
WG 5:	Electrical sensing methods
	Determination of particle size distribution – Electrical sensing zone method (published as BS/ISO 13319: 1999 and ISO 13319: 1999)
WG 6:	Laser diffraction methods
	Particle size analysis – Laser diffraction methods – Part 1: General principles (to be published as BS ISO 13320-1: 1999) Particle size analysis – Laser diffraction methods – Part 2: Convergence algorithm
WG 7:	Photon correlation spectroscopy
	Particle size analysis – Photon correlation spectroscopy – Part 1: Guidance on laser diffraction (published as BS/ISO 13321: 1998 and ISO 13321: 1998) Particle size analysis – Photon correlation spectroscopy – Part 2: Validation of inversion procedures
WG 8:	Image analysis methods
	Particle size analysis – Image analysis methods (ISO 13322, at DIS stage)
WG 9:	Single particle light interaction methods
	Determination of particle size distribution – Single particle light interaction methods – Part 1: Light interaction considerations (ISO 13323-1, approved as a DIS (Draft International Standard)) Determination of particle size distribution – Single particle light interaction methods – Part 2: Light scattering single particle light interaction device design performance specifications and operation requirements (ISO 13323-2, committee draft) Determination of particle size distribution – Single particle light interaction methods – Part 3: Single particle light interaction device, design performance specifications and operation requirements (ISO 13323-2, draft)
WG 10:	Small angle X-ray scattering method
	Particle size analysis – small angle X-ray scattering method (ISO 13762, committee draft)
WG 11:	Sample preparation
	Sample preparation – Sample splitting of powders for the characterization of particle properties (ISO 14488, draft) Particle size analysis – Dispersing agents for powders in liquids (ISO 14887, approved as a DIS)

of powders to facilitate industrial production, has always been of paramount importance.

One of the alternatives to measuring particle size and shape, and one which is becoming of more and more importance mainly because of the ease with which it can now be measured, is measuring the surface area of a powdered solid.

The ease of agglomeration and compaction, together with many other powder technological processes, can vary greatly for materials that have chemically identical structures but possess differing physical natures. Expressing the physical nature of the materials in terms of the energy, enthalpy, solubility or dissolution potential per unit surface area has thus enabled optimization in the choice of preferred raw resources. Whenever a solid is in contact with a fluid the interface between these two phases is dependent upon the surface area available to interact with the surrounding fluid, be it gaseous or liquid.

4.1.4.1 Surface Area of Powdered Material Measurement of size, shape and size distribution is generally accepted as a means of characterizing simple end-user products, but with more complex particles, surface-modified granulates and ''high tech'' designer multifunctional powder systems, the parameter of surface area together with inter- and intraparticulate space within aggregated, agglomerated or compacted powders can give a more meaningful explanation of particle–particle behavior. The need to control particulate structures and to modify the surface of the resultant product is thus an emerging commercial demand. The morphology of a particulate surface can either be described as rugged or tortuous as opposed to a smooth surface. Fractal geometry, from image analysis, can be used to describe not only the particles' shape but also the texture or roughness of the agglomerated surface. Penetration into all the cracks and crannies which are generally present on the surface of a rugged particle is best accomplished by the adsorption of a fluid onto the surface of a solid.

Because of the wide use of powdered material in various fields of engineering and technology there is no unique method for analysis or characterization of the topography (surface) or the internal structure (pore/void distribution) in complex porous matrixes. The surface areas of size-enlarged particles or powders are generally determined from the monolayer or multilayer adsorption region of a low-temperature nitrogen adsorption isotherm using the Brunauer–Emmett–Teller (BET) equation.[55] The multilayer and capillary condensation region of the low-temperature nitrogen isotherm can also be used to evaluate the size and shape of pores within particles or the voids between particles. The maximum relative pressure achievable with low-temperature adsorption equipment usually limits the pore/void measured to a size not greater than 500 nm.[56–58]

The nitrogen BET method[55,59,60] is one accepted technique for the measurement of surface areas. Another, mainly used for low-surface-area powders ($<1.0\,m^2g^{-1}$), is krypton adsorption. The protocol of surface area measurement, together with the comparison of measured surface area values of powders against standards that have been approved by the powder technology community, is therefore a necessity in global validation of powdered materials. One of the most comprehensive descriptions for the determination of surface area and porosity by physisorption has been published by the International Union of Pure and Applied Chemistry (IUPAC).[61,67]

The degree of humidity is another important factor in the preparation and assessment of the mechanical properties of particles, powders, cakes and compacts. Most problems encountered in the formation and subsequent behavior of particulate material can be related to the influence of either a liquid or gaseous phase on the solid surface interface. Elbirli et al.[62] have shown that the degree of humidity controls the extent of capillary condensation of water between points of contact between particles.

The shear strength of packed beds of rounded, spherical type particles decreases with an increase in humidity. However, with irregularly shaped, angular particles with sharp contact points an increase in the adhesive force, and thus the shear strength of a powdered bed occurs with an increase in humidity. This contradiction is dependent upon the presence of liquid rings or the onset of the funicular state between particles. These phenomena are in turn dependent upon the surface area and surface irregularities of particles. Since the presence of moisture is generally influential in either the agglomeration or disintegration and dissolution of powdered materials, investigation into the effect of moisture, by the determination of water adsorption isotherms, is one of the more important characterization techniques used to measure the specifications and properties of compressed and uncompressed aggregates, granules and agglomerates of fine particles. The specific surface area of particulate and granulate material can be calculated from the BET equation at low temperature (77 K), but at room temperature it has been found,[63] because of the increase in the range of linearity, that it is preferable to use the Huttig equation[64] rather than the BET equation for surface area measurements by water adsorption.

Information obtained from either nitrogen or water physisorption measurements is at times limited to the surface topographical features of the solid, but recently analysis and interpretation of nitrogen and water adsorption–desorption isotherms has been extended using the

generalized Dubinin–Astakov equation[65] to determine the energetic heterogeneity of surfaces.

When the interparticulate sizes within compacts are greater than those readily measurable by low-temperature adsorption and application of the Kelvin equation (i.e. greater than 500 nm), the technique of mercury intrusion porosimetry (section 4.1.4.3) can evaluate intraparticle and interparticle space in the approximate size range from 3.5 nm to 1.0 μm.[58,66]

4.1.4.2 Porosity in Particles The techniques available for the physico-chemical characterization and measurement of either intraparticular and interparticular space are dependent upon the magnitude of the space dimensions to be measured. The material matrix within an agglomerated or compacted mass of powder may have pores or voids that are generally in the 1.0 nm to 0.1 mm size range. These pore/void spaces can be sized by the physical adsorption of various vapors or gases at low (nitrogen, krypton and carbon dioxide) and ambient (water, butane or organic vapors) temperatures. Since the techniques chosen for the characterization of porous materials are, to some extent, dependent upon the dimensions of the space to be measured and the mechanisms of pore space filling, it is expedient to classify pore space or porosity in terms of the dimensions of the space, with the caveat that the size limits are arbitrary and can be influenced by the interactions of the solid and fluid used in the measurement technique. Pores exceeding 50 nm in width are generally termed macropores, while pores with a width not exceeding 2 nm are micropores. Pores of intermediate sizes are termed mesopores.[67]

Mesopore Space. Although there is no single current theory of adsorption capable of providing a comprehensive mathematical description of physisorption isotherms over the entire range of relative pressures, it is possible to determine from an isotherm not only the surface area of a particulate solid, but also information on the texture, pore spaces or topography of solid materials from the multilayer and the capillary condensation regions of the isotherm. The demarcation between external and internal surface is dependent, however, on the accessibility, size and shape of the adsorbate molecules, as well as the pore space distribution in the solid. In the case of a solid which has mesoporosity, the adsorption isotherm tends to remain nearly horizontal as the relative pressure range approaches unity. The total pore volume can be evaluated in such isotherms – usually classified as type I or IV BDDT (Brunauer–Deming–Deming–Teller)[68] isotherms – from the Gurvitsch Rule.[69] When macropores are present the adsorption isotherm tends to rise rapidly near the relative pressure value of unity and reliable total pore volume measurements can only occur with carefully controlled experimental procedures. This type of isotherm is usually classified as a type II BDDT isotherm.

Micropore Space. When the dimensions of the adsorbate are similar in magnitude to the capillary dimensions within solids, the filling of these spaces (micropores) at low temperature results in a type I BDDT isotherm. Type I isotherms in most cases do not show hysteresis with microporous material, although narrow hysteresis has been found with some adsorbate–adsorbent systems.

Lippens et al.[70] showed that the specific surface size and shape of micropores can be obtained from a plot of experimental adsorbed volumes at specific relative pressure versus the statistical thickness at the same relative pressure of an adsorbed layer on a nonporous solid. Unhindered formation of multilayers on nonporous surfaces produces a linear relationship, passing through the origin between the volume adsorbed and the thickness of the multilayer adsorbed on the nonporous solid, a V_{a-t} curve. The slope of this line is a measure of the surface area of the porous or nonporous material. At higher relative pressures, and thus larger t values and pore space dimensions, deviations from linearity can occur. At certain relative pressures more adsorbate can be adsorbed into specifically shaped and sized pores than onto a corresponding nonporous solid. A decrease in adsorption will cause a decrease in the slope of the V_{a-t} relationship – the curve becoming convex to the x axis – and this is indicative of microporosity. An increase in the slope of the V_{a-t} equation occurs when the size of the pore space is larger or slit-shaped, and is indicative of mesoporosity.[71] Differentiation of the V_{a-t} curve or Sing's α_s plot,[72,73] at known relative pressures and thus pore radii, is the basis of the micropore analysis method of Brunauer et al.[74,75] It is essential, however, in the determination of micropore space distributions, that a nonporous reference solid t-curve be chosen which has an adsorption coefficient value similar to that of the unknown structure. Provided that the micropore size distribution is not too broad then excellent agreement between derived values of micropore volume and pore width can be obtained with Sing's ''alpha s'' (α_s) plot and the Dubin–Radushkevich equation.[76]

4.1.4.3 Mercury Intrusion Porosimetry Determination of the pore space distribution by mercury porosimetry has, because of the relative simplicity, versatility and speed of the technique, been used for many quality control characterizations in preference to other techniques. The range of porosimeters available and discussions on the accuracy and reliability of the mercury intrusion method have been covered by Orr[77] and van Brakel et al.[78] The principle of mercury pore space distribution is based on the fact that mercury does not wet most materials

(the contact angle is greater than 90°). One of the difficulties in interpretation of mercury porosimetry results is the inability to distinguish between penetration into inter- or intraparticle space. With nonporous material only interparticle space exists, but many compressed and uncompressed particles contain intraparticle space. Critical examination of porosimetry curves may, at times, show two sharply defined regions, one at relatively low pressures and the other at higher pressures. It is thus beneficial to measure an overall pore–void size distribution over the 1 nm to ca. 50 μm range using both the adsorption and mercury penetration techniques. This then aids the recognition and discrimination of the interparticle and intraparticle spaces, especially with microporous particles.

4.2 In-line and In Situ Particle Characterization

Orr[79] and later Leschonski[80] described the start of a new era involving rapid-response particle–process-control sizing instruments, while Scarlett[81] highlighted the statement that Heywood[24] had made many years before: ''a size analysis procedure which can form part of a fully automated production unit will be needed to feed back information to the process''.

On-line monitoring and process control does not, however, remove the problem of obtaining a representative powder sample. It only changes the essential sampling operation required to obtain a particle-process parameter from a manual laboratory collection of particles to a possibly automated particle collection process. These on-line sample collection processes, followed by size or particle characterization techniques, may have computer control, collect many more powder samples and possibly have more rapid-response times, but on-line particle characterization still does not overcome the problem of collecting, removing and transporting the particulate sample to the particle characterization instrument.

As time, effort and money can be wasted by analysis of a nonrepresentative sample, the emphasis for size measurement/characterization, with the advent of computerized information technology, has shifted from off-line batch techniques to on-line and in-line analysis. Ultimately the desired goal for many process and product engineers is in situ particle characterization and control. On-line measurements nowadays are largely devoted to modification and adaption of off-line methods.[79,83,84]

Even if the on-line analysis has a rapid response of only a few seconds or parts of a second this will eventually be superseded by in-line and/or in situ particle size and size distribution measurements.

In-line and in situ methods of characterizing particulate systems attract the attention of industry because:

1. The physical characteristics of the particles in a chemical process influence both the dynamics of the process and the final product specification.
2. In a dynamic continuous chemical process there should be a continuous dynamic measurement that can control the particle characteristics of the powdered material in terms of particle size, particle size distribution, shape, bulk density and other particle or powder characteristics.
3. Characterization of the mechanics and dynamics of the chemical process itself is also essential, initially to understand the process, then to control, improve and optimize the plant in terms of safety and cost effectiveness, and finally to tailor-design the final product from the chemical process.

Laser-based particle characterization instruments may be ideally positioned to achieve in-line or in situ particle size analysis because of the manner in which particles are detected and the ease with which computational manipulation of the collected information, needed to record a desired particle size or process parameters,[85] can be accomplished.

5 SHAPE CHARACTERIZATION OF IRREGULAR PARTICLES

The need nowadays is not only for the measurement of size and size distributions, but also for the measurement of shape and shape distributions. Although Rose[86] stated in 1961 that: ''perhaps one of the most abortive searches yet made is for a means of defining the shape of a particle'', many workers are now currently appraising past work on shape factors. Shape measurement, initially of academic interest only, began to find credibility when shape characteristics were incorporated into industrial and commercial powder and granulate specifications.

The complexity of shape analysis by analytical methods other than Mandelbrot[87,88] fractal analysis emphasizes the need to describe shape by more than one technique and parameter.

Most instruments for shape analysis compare the profile of an irregularly shaped geometric particle with that of a known geometric shape such as a sphere, cube or rectangle. The shape of the irregular particle is then described in terms of length and width or simply a ratio devised from computer image analysis. The main approaches to the development of techniques and instruments to characterize and measure the shape of irregularly shaped particles can be illustrated in several ways, as shown in Figure 6.[89,90]

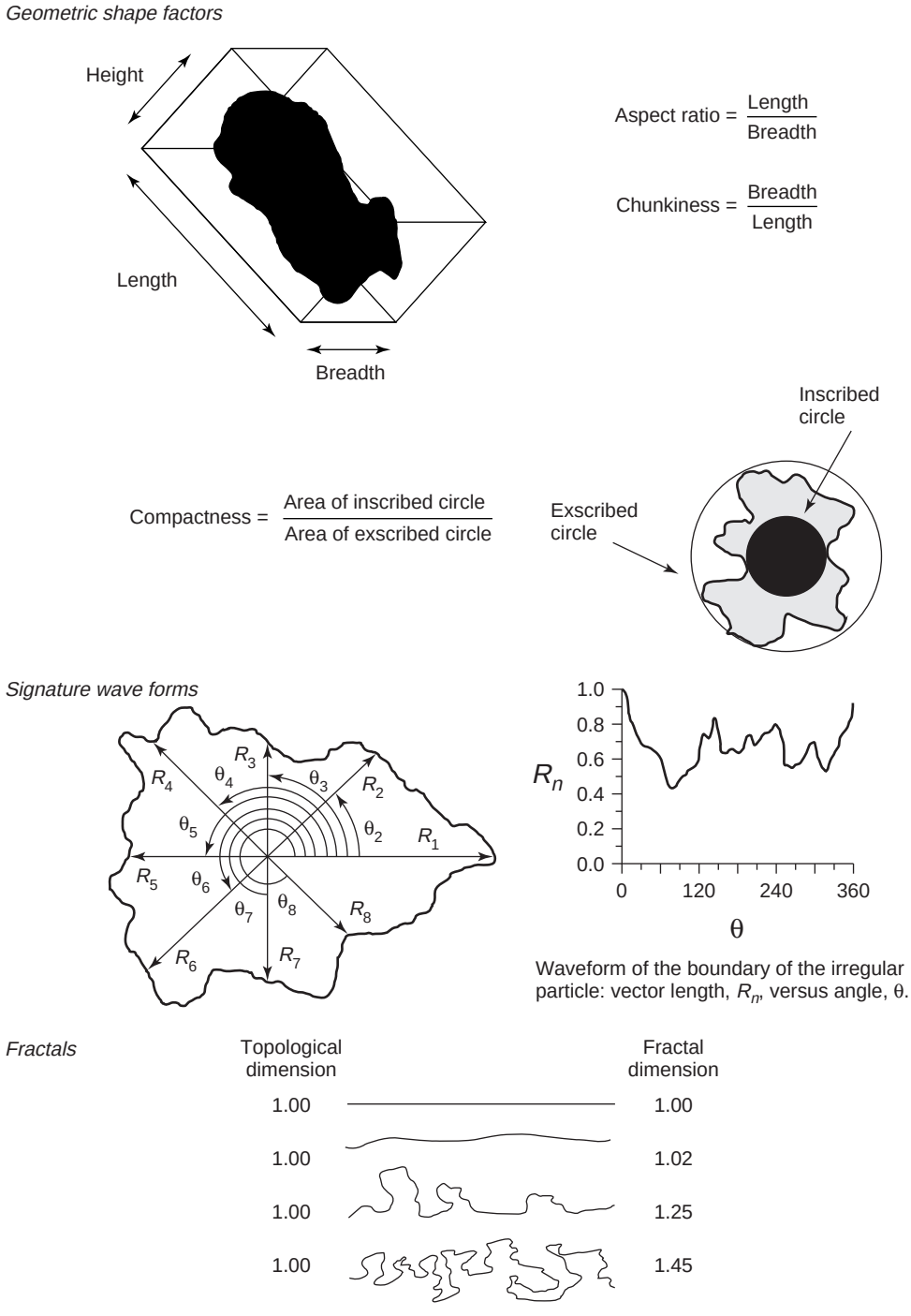

Figure 6 Various techniques for characterizing the shape of irregular particles.

5.1 Direct Measurement of Single Particle Shape

5.1.1 Dimensionless Geometric Shape Factors

Particle shape is a fundamental characteristic of a particle, in a manner similar to particle size. In BS 2955[25] various types of particle shape were described in words, and numerous general shapes have subsequently been defined (Table 11). These verbal descriptions are, however, inadequate for incorporation into equations and calculations to ascertain the effect of particle shape on various particle systems, particle properties or the design of bulk powder handling equipment. One of the first attempts to standardize particle shape measurement for individual particles was by measuring the length or

Table 11 Definitions of particle shape

Shape	Definition
Acicular	Needle-shaped
Angular	Sharp-edged or having a roughly polyhedral shape
Crystalline	Freely developed in a fluid medium of geometric shape
Dendritic	Having a branched crystalline shape
Fibrous	Regular or irregular thread-like
Flaky	Plate-like
Granular	Having an approximately equidimensional irregular shape
Irregular	Lacking any symmetry
Modular	Having rounded, irregular shape
Spherical	Global shape

diameter and the breadth or width of a particle and then, from the protocol outlined in BS 512 (1966), assigning different and arbitrary ratios of particle breadth to length, B/L, and particle thickness to breadth, T/B, to various three-dimensional particle shapes. Regular geometric shapes, such as disks, rods, blades and equidimensional particles, could then be assigned a numerical value relating to the shape of the body or particle rather than a literal description.

Herdan[13] had stated, before the publication of BS 512 (1966), that any ''shape factor'' should have a three-fold function:

1. It should be a factor of proportionality between the particle size determined by different methods of particle characterization, so that, for example, the particle size determined by microscopic techniques could be compared with that determined by sedimentation.
2. Shape factors should be conversion factors for expressing the results of different particle measurement techniques as an ''equivalent sphere'' parameter.
3. Shape factors should transform the second and third powers of a measured particle diameter into the parameters of particle surface area or particle volume.

The volume of a particle can thus be expressed as the volume coefficient (volume shape factor) multiplied by the cube of some characteristic dimension. Similarly, the surface area of a particle can be expressed as the surface coefficient (surface shape factor) multiplied by the square of some characteristic particle dimension.

Algebraically these definitions can be written as shown in Equations (9) and (10)

$$V = \alpha_v x^3 \tag{9}$$

$$S = \alpha_s x^2 \tag{10}$$

where x is the size characteristic of the particle and α is a shape factor. Using a microscope count method, the size characteristic as a one-dimensional quantity can be either a Feret's diameter or a projected area diameter, x_a. As the volume and surface area values for a given particle are fixed, the numerical values of the coefficients of shape α_v and α_s are thus dependent on the chosen size characteristic. Subscripts to α_s and α_v must be used to denote the method of particle size measurement. Hence for an irregular particle whose size is described in terms of the dimensions of a projected area diameter x_a, the geometric volume shape factor α_v can be symbolized as α_{va}, and the surface shape factor α_s as α_{sa}.

Measurement of the two-dimensional quantity of area can be obtained from the projected microscopic image of the particle, while the three-dimensional quantity (volume) can be measured from either a sedimentation or an electrical sensing zone technique. The electrical sensing zone method is, however, the preferred method for particle volume measurement nowadays.

When the characteristic dimension of an irregularly shaped particle is the projected area diameter x_a, the surface of that irregularly shaped particle can thus be symbolized as $\alpha_{sa}x_a^2$, where the subscript ''s'' in the shape factor α_{sa} denotes a surface shape factor and the subscript ''a'' is related to the projected area diameter x_a. This shape factor is multiplied by the square of the particle size measurement of the particle's projected area diameter, x_a, which is obtained from microscopy.

When the measured dimension is that of a sphere with the same surface area as an irregular particle then, by definition, the surface area of that sphere is equivalent to πd_s^2 or πx_s^2. Likewise, from the volume viewpoint, if the volume of an irregularly shaped particle is $\alpha_{va}x_a^3$, with x_a as the diameter of the irregularly shaped particle, then the volume of a sphere with diameter x_v is mathematically equivalent to $\pi x_v^3/6$. The surface area and volume shape factors of spherical particles α_s and α_v are thus numerically equal to 3.14 (π) and 0.52 ($\pi/6$) respectively. It has been found empirically[66] that for irregular and nonuniformly shaped particles the ratio V/x^3, where V is the average particle volume and x is the mean particle size, remains sensibly constant for various grades of particles. It has also been found empirically that the ratio S/x^2, where S is the average particle surface area and x is the mean particle size, remains sensibly constant for different sizes of the same material. It is thus possible, at times, to speak of an average particle shape within an assembly of particles, although variations in particle shape with size have been found for some processed and powdered materials.[91] However, such theoretical considerations, although of academic interest, had little or no application in particle processing for many years.

The practical application of shape analysis was initially addressed when Davies[92] stated that since particle shape information was needed to describe the behavior of particles in many industries in which particles were mixed, reacted, stored or transported, and although the main parameters still remained particle size, particle size distribution, density, and surface interactions within complex particle processes, for the purposes of clarity, particle shape should be represented by:

1. the general particle form, by an aspect ratio, a proportionality constant or a shape factor;
2. the surface topography, by roughness, angularity or roundness;
3. or a combination of the previous two categories.

Prior to Davies' practical recognition of the value of particle shape factors within industrial processing, Heywood[93] had derived an expression for the shape of a particle in terms of the general geometric form of an irregular particle from the ratio of length, L, breadth, B, and thickness, T (Figure 7). To separate the dimensional proportions of the particle, from the geometric shape of the particle, Heywood used the ratios of elongation, n (ratio L/B) and flatness, m (ratio B/T), together with a term defined as the equidimensional shape factor, α_0.

The volume shape factor of an irregular particle, α_v, could therefore be represented in part by the factor α_0 and also by the ratios of the elongation and flatness ratios, n and m respectively, which characterized the influence of the dimensions of the particle on the particle shape, as shown in Equation (11).

$$\alpha_v = \frac{\alpha_0}{m\sqrt{n}} \tag{11}$$

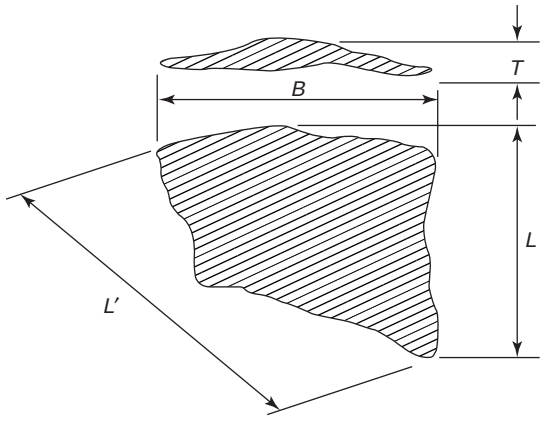

Figure 7 Heywood's shape characteristics for an irregular particle.

From a variety of known geometric shapes, and for irregularly shaped particles with known surface shape factors α_s, Heywood[23,94] deduced the relationships for particle shape as shown in Equation (12).

$$\alpha_s = 1.57 + C\left(\frac{\alpha_0}{m}\right)^{4/3}\left(\frac{n+1}{n}\right) \tag{12}$$

Combined with Equation (11), this could be written as shown in Equation (13)

$$\alpha_s = 1.57 + C\alpha_v^{4/3}\left(\frac{n+1}{n^{1/3}}\right) \tag{13}$$

where the shape coefficients α_0 and C can be determined experimentally (Tables 12 and 13). Values of L, B and T can easily be obtained using an automatic scanning microscope and software. This approach to shape recognition has been used by Church,[95] Dwyer et al.[96] and Mandelbrot[97] to obtain relationships between measured particle and linear diameters and other particle parameters, for a collection of powders.

A disadvantage of Heywood's elongation ratio is that as an irregularly shaped particle becomes more elongated and the breadth becomes smaller, the elongation ratio tends to an infinite value. Heywood's elongation ratio has fortunately also been defined, in different terminology, as an aspect ratio which is the ratio of the length of an irregular particle divided by the width of the particle

Table 12 Shape coefficients for some geometric and nongeometric shapes

Shape group	α_0	C	$\alpha_0 C^{4/3}$
Geometric form			
Tetrahedral	0.328	4.36	0.986
Cubical	0.696	2.55	1.571
Spherical	0.524	1.86	0.712
Approximate form			
Angular			
Tetrahedral	0.38	3.3	0.91
Prismoidal	0.47	3.0	1.10
Subangular	0.51	2.6	1.06
Rounded	0.54	2.1	0.0

Table 13 Typical values for the shape coefficients of irregular particles of various shapes

Material	α_0	α_s	ψ
Rounded particles, water-worn sand	0.32–0.41	2.7–3.4	0.817
Angular particles, pulverized minerals	0.20–0.28	2.5–3.2	0.655
Flaky particles, talc, gypsum	0.12–0.16	2.0–2.8	0.543
Thin flakes, mica, graphite	0.01–0.03	1.6–1.7	0.216

profile at right angles to the length measurement (L/W). Kaye and Clark[98] suggested that a more acceptable and possibly more practical ratio would be the chunkiness ratio, W/L or B/L, which is the reciprocal of Heywood's elongation ratio or the aspect ratio of the irregular particle. The chunkiness ratio thus gives a numerical range from zero to unity (instead of unity to infinity), as the breadth of the irregular particle decreases with length.

To distinguish between cubes and spheres, which have linear mono-dimensional values, an alternative technique is required, which involves the comparison of an irregularly shaped particle with a more symmetrical body. To these ends, comparison of the degree of asymmetry of an irregular shaped particle with that of a symmetrical body was proposed by Wadell for petrological and geological powdered soil samples. Wadell[99–101] defined the shape of an irregularly shaped particle as having either sphericity, ψ, or roundness, P. Sphericity was defined as the ratio of the surface area of a sphere that has the same volume as an irregular particle, to the actual surface area of the irregular particle. This can be expressed algebraically as shown in Equation (14)

$$\psi = \frac{\pi d_v^2}{\pi d_s^2} \tag{14}$$

but can also be expressed in a more practical manner as the working formula shown in Equation (15).

$$\psi = \frac{d_c}{D_c} = \frac{\text{Diameter of a circle equal in area to the projected area of a particle in a stable plane}}{\text{Diameter of the smaller circle circumscribing the projection of the particle}} \tag{15}$$

Treasure[102] showed the interdependence of shape factors using the relationship of the surface coefficient, α_s, the Heywood equidimensional shape factor, α_0, and Wadell's sphericity (Table 13), as shown in Equation (16).

$$\psi = 4.38\frac{\alpha_0^{2/3}}{\alpha_s} \tag{16}$$

Wadell's roundness can be obtained by measuring the radius of curvature (r) of each projection in the perimeter of an irregular particle. The sum of these radii can then be related to the radius of the maximum inscribed circle (R) by Equation (17)

$$P = \sum \frac{r/R}{N} \tag{17}$$

where N is the number of projections measured (Figure 8). As the corners of a particle are worn down, the radius of curvature, r, tends to R and thus P tends to unity.

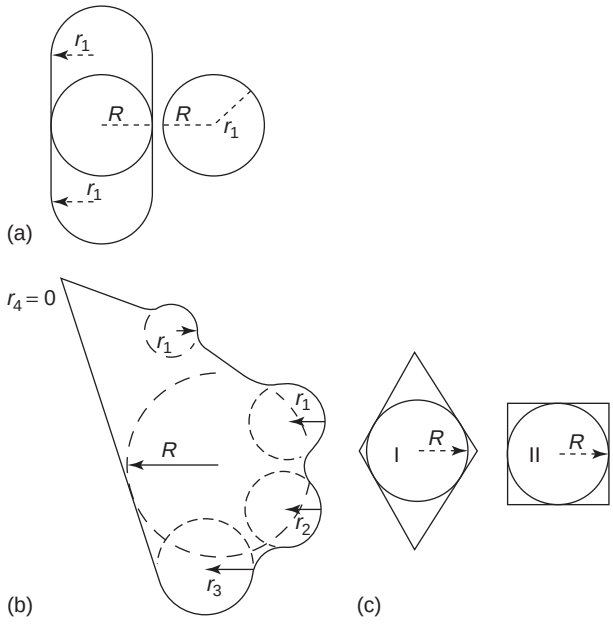

Figure 8 Wadell's shape characteristics for an irregular particle. (a) Two particles with identical roundness ($P = 1$) but different form; (b) Wadell's roundness for an irregularly shaped particle; (c) two particles with $P = 0$ but with different form and angularity.

Table 14 Shape coefficients for sections of a cube

Relative dimensions of cube	α_0	α_s	ψ
Thickness = breadth	0.696	4.71	0.806
Thickness = breadth/2	0.348	3.14	0.761
Thickness = breadth/10	0.070	1.88	0.434
Thickness = breadth/100	0.007	1.60	0.110
Thickness = 0	0.0	1.57	0.0

Wadell's roundness factor can be used to distinguish between the form of the particle and the roundness or surface perimeter, sometimes termed the roughness, of a particle. Thus Wadell's roundness ratio is, at the moment, favored instead of Wadell's sphericity. This is because of the insensitivity of Wadell's sphericity ratio, which requires large changes in relative particle proportional dimensions of length and breadth to thickness in order to produce significant changes in Wadell's sphericity, ψ (Table 14).

Any shape factor that attempts to visualize and enumerate the inherent roughness of the surface of most real powdered materials is dependent on the measured fineness or surface area of particles, because the evaluation of the surface area or perimeter of an irregularly shaped particle changes in accordance with the measurement technique used to characterize an assembly of irregularly shaped particles. This can be illustrated by the sphericity value of a porous spherical

particle, which is relatively low in numerical value and may therefore be indistinguishable from the sphericity of a smooth but irregular and highly convoluted particle. The area and perimeter of an irregular particle were used by Hausner,[103] who encompassed the complicated irregular shape of a particle image within a rectangle. The dimensional characteristics measured are the side lengths of the rectangle, a and b, the projected area of the particle's image, A, and the circumference of this area, C'. Since the cross-sectional area and circumference of a spherical particle are $A = d^2\pi/4$ and $C' = \mathrm{d}\pi$, respectively, the relationship $C'^2 = 4\pi A$ exists. Therefore for a spherical particle, which gives the minimum surface area for a given volume, a Hausner shape factor can be defined as shown in Equation (18), and has a value of unity for a spherical particle.

$$Z = \frac{C'^2}{12.6A} \tag{18}$$

Since automatic image analyzers with appropriate computer software can now rapidly measure particle surface area (with and without porosity), perimeter, Feret's diameter, longest chord length and number of particles (Figure 3), numerical values of the various shapes and forms of irregular particles can now be readily obtained from ratios of these various image parameters.[95,96]

5.1.2 *Signature Waveforms of Irregular Particles*

With the advent and development of automated microscope systems linked into and coupled with computers, the description of irregular particle shapes and the characterization of profiles by complex mathematical functions has now become available. Schwartz and Shane[104] were the first to describe a two-dimensional profile of an irregularly shaped particle by a geometric signature waveform that could undergo Fourier analysis.[105] The signature waveform was generated by locating, at a central position within the particle profile, a vector that touched the outline profile of the particle. The magnitude of the vector, at various angles when moved around the irregular profile, could either produce a graph of magnitude against angle or be treated as a continuous periodic harmonic waveform (Figure 6). The complex particle signature waveform could then be broken down by Fourier analysis[105–107] into a list that commonly used the first five harmonics in a Fourier series of the simple harmonic wave to uniquely characterize the shape of the irregular particles. A list of the Fourier harmonic series within the complex signature waveform gave a dimensionless shape description of the particle. The central location of the vector could be the center of the smallest circle inscribing the particle, as in the work of Schwartz and Shane, or could be at the center of gravity of the particle.[108,109] The higher harmonics (>5) were believed to give an indication of the texture of the irregular particle. Kaye[110] reviewed the work on Fourier shape analysis and found that the geometric signature method was, in reality, only of use for relatively rounded materials, such as geological fluvial sediments and spray-condensed metals or spray-dried products. Kaye, following the work of Austin et al.,[111] used the signature waveform methodology in combination with slip-chord analysis to separate the edges of an irregular particle image and therefore introduced a new concept to the armory of shape analysis. This concept, which is termed the angular facet signature waveform, is now used in the shape characterization of abrasives and comminuted materials.[113]

5.1.3 *Fractal Geometry of Rugged Irregular Particles*

Mandelbrot[87] used fractal geometry to describe the ability of a convoluted line to fill two-dimensional space. A straight line, on a flat piece of paper, without any ''ups and downs'' and without ''wiggles'' has a topological dimension of unity. As the ''straight'' line becomes more irregular in shape, wiggles up and down on the flat piece of paper, and has an ''irregular wavelike'' form, it occupies more two-dimensional space than a straight line. The fractal dimension thus becomes greater than one because of this irregular waveform or ruggedness of the line (Figure 6). The extended irregularities of the boundary of an irregular particle can be likened to a convoluted line and fractal geometry. The fractal dimension of the boundary of irregularly shaped profiles can thus be applied, as an alternative mathematical method, to obtain an index of ruggedness or texture of irregularly shaped particles. The length of a convoluted curve can be estimated from the number of steps required to traverse a convoluted curve. As the step size decreases, the zig-zag stepped curve follows the convoluted curve being measured more closely, until eventually a constant value for the length of the convoluted curve is reached (Figure 9).

Mandelbrot[97] showed that for a fractal curve a plot of the logarithm of the estimated length of a convoluted curve against the logarithm of the step size used gave a linear relationship which extrapolated to an infinite length. The slope of the log–log plot gives a quantity $(1 - D)$ which is defined as the fractal dimension of the rugged boundary curve of an irregular particle. The Mandelbrot fractal dimension is termed a fractal which has a number between 1 and 2 and which can be used to describe the structure or shape of an irregular profile. Flook[113] used fractal dimensions to characterize the texture and structure of geometrically constructed model particles and the profiles of carbon black aggregates. He found that for carbon black aggregates the plot of

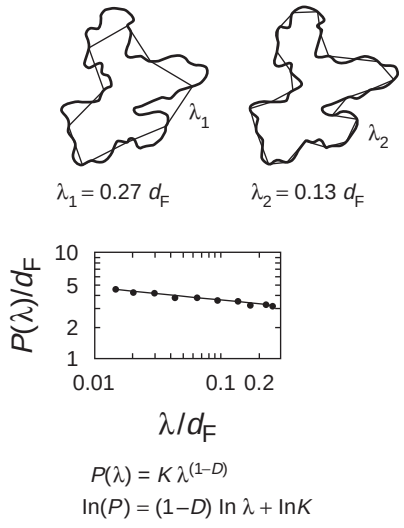

Figure 9 Fractal geometry of a rugged irregular particle. The two fractal step sizes, λ_1 and λ_2, are expressed in terms of the Feret's diameter, d_F, of the irregular particle as 0.27 d_F and 0.13 d_F, respectively. The log graph shows the normalized values of the perimeter, $P(\lambda)/d_F$, versus the normalized step length, λ/d_F, to give a slope of $(1 - D)$ which is the fractal dimension of the rugged irregular particle. $P(\lambda)$, perimeter; λ step length; d_F, Feret's diameter; K, constant.

logarithmic perimeter length against logarithmic step size showed a departure from linearity at large and small step sizes. At small step sizes the fractal dimension (from the slope of the log–log plot) was dominated by the structure of the subunit, while at large step sizes it was affected by the coarse structure of the agglomerate. Flook concluded that it was practical to use fractal dimensional measurements to characterize the textured and structured profiles of a wide variety of materials.

5.2 Indirect Measurement of Particle Shape

5.2.1 Off-line Shape Characterization

One of the inherent restrictions in the characterization of particle size is that the measurement of a particle dimension is dependent upon the physical criterion used in the measurement technique. An irregular particle will be found to have different diameters when characterized by sieve analysis, sedimentation and an electrical sensing zone technique. The difference between the measured sizes can therefore be advantageous in shape recognition, and the factor of proportionality between the particle sizes determined by different analysis techniques can be regarded as a shape factor. BS 3406 Parts 2, 3, and 4 (1963) states that a particle found by sieve analysis to have a diameter of 75 μm may have a mean projected diameter of 105 μm or a mean Stokes diameter of 70.5 μm. BS 4359 Part 3 (1970), Appendix B tabulates calculated and experimental values of surface, volume and specific-surface coefficients (shape factors) for various dimensional forms and irregular particle shapes. These shape factors were obtained from sieving, permeability, sedimentation, electrical sensing zone and light extinction methods of particle size and surface area measurements. In all cases the numerical values of 3.14, 0.52 and 6.0 for α_{sv}, α_{vv} and α_{svv}, respectively, applicable to spherical forms, are exceeded for nonspherical or irregularly shaped particles over a range of sizes from 2000 μm to 0.5 μm.

From Heywood's[93] flatness ratio ($m = B/T$) and elongation ratio ($n = L/B$), a relationship between the shape and sieve characteristics of a powder can be established, as shown in Equation (19)

$$x_a = 0.98 \left\{ \frac{2nm^2}{m^2 + 1} \right\} x_t \qquad (19)$$

where x_a and x_t are the projected area and sieve diameter respectively for various values of m and n.

Nystrom and Stanley-Wood[114,115] described a ring gap sizer, initially used for the characterization of fragile explosive materials, which could measure uniquely the thickness of particles, as opposed to an intermediate diameter between length and thickness as obtained by sieve classification. The geometric mean particle size obtained from the ring gap sizer was within 7% of the mean thickness of particles measured by microscopy. Comparison between the ring gap sizer diameter and the dry sieve diameter could thus be used to establish shape factors of free-flowing materials.

Pilpel[116] showed that the shape of small particles obtained by comminution is dependent upon the method of size reduction. Micronized pigments are often less irregular and more spherical in shape than particles which have been repeatedly fractured and sheared by attrition. Table 15 shows the shape factors for several micronized or fractured materials. As the shape factor – the ratio of the calculated surface area obtained from electron microscopy to the BET measured surface area – approaches unity, it indicates a particle with a more spherical form. The effect of size and shape on the flow and failure properties of penicillin powders, which have an angular prismoidal form, was shown by Walton and Pilpel[118] to affect the bulk powder characteristics of cohesion, tensile strength, internal friction and failure factors.

5.2.2 On-line Shape Characterization

In-line and on-line techniques for product engineering and the characterization of industrial powder processes continue to attract the attention and demands of industry because particles passing through any physico-chemical

Table 15 Effect of micronization on the shape of pigment particles

Pigment	Measured surface area ($m^2\ g^{-1}$)	Calculated surface area ($m^2\ g^{-1}$)	Shape factor
Prussian blue	61.3	2.3	0.04
Prussian blue, micronized	58.9	12.7	0.23
Red iron oxide	7.6	3.4	0.44
Red iron oxide, micronized	7.3	4.9	0.68
Burnt amber	138	1.2	0.01
Burnt amber, micronized	94	5.0	0.05
Yellow iron oxide	17.6	1.4	0.08
Yellow iron oxide, micronized	17.7	2.5	0.14

process affect the performance of the process and the specification of the product. For most operations the two most important parameters in process monitoring and control are particle size and morphology. With recent advances in sensor technology it is now possible to implement both size measurement and shape morphology characterization simultaneously. This can be done using two separate units,[118] one for size and the other for shape,[119] or a single unit[120–122] which can, by laser light illumination, produce both a particle size distribution from a flowing stream of particles, using a time-of-flight (TOF) or time-of-transition (TOT) technique, and also a morphological particle shape characteristic.

Although light-scattering techniques are not commonly used for particle shape classification there is, within the scattering signature from the scattered light intensities obtained from the particle size distribution, encoded information from which it should be possible to infer the shape of particles. By changing the geometric shape of the photodetectors from a series of semicircular concentric detectors to an array of wedge-shaped detectors in a segmented ring, the shape of particles may be measured.[123] To record the particle shape information from the scattered light, the classical laser diffraction detection system (which uses semi-ring-shaped detectors to collect the radial spatial frequencies or intensities of the particle field), has to be changed to include wedge-shaped photodetectors which collect encoded particle shape information in the form of azimuthal spatial intensity frequencies. The influence of shape on size measurement by light blockage and the characterization of plate-like particles has been appraised for rapid response instrumentation and in modified on-line light techniques.[124]

It is now currently accepted that a combination of both size and shape measurements will lead the way forward to the characterization of mixed particle species, complex designed particles and multifunctional agglomerated systems of varied size, shape, and mix.[125]

6 NATIONAL AND INTERNATIONAL STANDARDS AND REFERENCE MATERIALS

6.1 National and International Standards

For the purpose of scientific communication, and to avoid disputes in commerce, industry and legislation, it has long been recognized that great value can be derived from using measurements directly related to fundamental standards. Most physical, physico-chemical and chemical primary and fundamental units have been defined by scientific experts, and are maintained under specific conditions that permit little or no variation in the value of the unit. In the field of powder technology, with the proliferation of particle size measurement techniques, there are several ways that an irregularly shaped mass of material can be characterized and sized. The method generally used at present is known as the equivalent sphere diameter, but more complex particle parameters may be used in the future as advances occur in the in situ recognition of assemblies of particles within industrial particle processing operations. It may be beneficial one day to characterize particulate matter obtained from a production process in terms of the purpose for which the particle or powder has be produced rather than to undertake the arduous task of sizing. Scarlett[126] and Leschonski[127] both appraised the necessity to size particles using equivalent sphere diameters. The characterization and sizing of particles in industries dealing with powdered material is generally only used to specify and maintain product functionality, and not to generate products of a certain size. As long as the product functionality (be it flowability, solubility, taste or drug efficacy) is maintained within specifications, then size is a secondary parameter. With the diversity of particle-sizing techniques readily available it may therefore be advantageous to measure the appropriate size of the particle using a physical property as closely related to the product functionality property as possible. There is therefore a need to calibrate powder measurement techniques to measure reconcilable particle properties closely related to the product properties, which at the present time happens to be the reconcilable property of size. Reference materials are then required, together with the careful standardization of a few basic techniques, to achieve confident particle sizing.

Leschonski[127] has commented that the measurement of size and size distribution of irregularly shaped particles ranging from a tenth of a micrometer to

several centimeters is, and presumably always will be, a difficult task. Whichever method is used for particle size measurement, there are two prerequisites which have to be accepted by all powder analysts because of the truisms of powder technology. The first prerequisite is that whichever particle size measurement technique is selected there has to be a recognized and a commonly accepted protocol or procedure, and therefore ultimately an agreed national or international standardized technique, to determine the particle size metrology and to ensure that one analyst's results will tally with another analyst's results. Confidence in the measurement technique chosen and the instrument used, from the simplest to the most sophisticated, is enhanced when the measurement can be related to a recognizable standard. Traceability may not be essential for the simplest in-house routine quality control tests, provided that the instrument is verified to give consistent results in terms of reproducibility with a selected standard or possibly a certified material.

This consensual standard method of analysis need not give an absolute measure of size, but any technique of size measurement should (and ultimately legislation may make this mandatory) be based on measurements where all the physical criteria measured are traceable directly to the international standards of mass, length and time and not derived from a second-order effect, unless the alternative measurement technique has been calibrated by certified reference materials and therefore becomes a traceable secondary standard. The second prerequisite is to have available reference materials (section 6.2), which in the special case of powder technologists tend to be reference powders of known particle size and particle size distribution. These reference powders can then aid the verification of the first prerequisite.

Standardization of the measuring methods available to the powder technologist is becoming, because of the global advances in powder manufacturing and chemical processing, more and more international rather than national. However, a situation may arise where a more sophisticated national standard may be created by an advanced technological nation, but not accepted at an international level by other less technologically advanced nations. Nowadays national standards are being harmonized within the framework of the European Community and also at an international level. National standard organizations still update current standards and tend to initiate new working standards as the processes of powder handling, characterization and particle measurement undergo innovation. Table 16 lists some of the standardization institutes that actively promote drafts and renew existing standards for the powder technological community, while Table 17 gives details of the current published British and German standards relevant to powder technology.

Table 16 National and international institutes/organizations for standardization

Acronym	Contact information
AFNOR	Association Francaise de Normalisation, Tour Europe, Cedex 7, F-2080 Paris, France
AIA	Asbestos International Association, 68 Gloucester Place, London WIH 3HL, UK
ASTM	American Society for Testing and Materials, 1916 Race Street, Philadelphia, PA, USA
BSI	British Standards Institution, 389 Chiswick High Road, London W4 4AL, UK
DIN	Deutsches Institut für Normung eV, Postfach 1107, D-1000 Berlin 30, Germany
ISO	International Organization for Standardization, Case postale 56, CH-1211, Geneva 20, Switzerland (central@iso.ch).
IUPAC	International Union of Pure and Applied Chemistry, http://www.iupac.org
NIST	National Institute of Standards and Technology, Rm. B311 Chemistry Building, Gaithersburg, MD 20899, USA
VDI	Verein Deutscher Ingenieure eV, Postfach 1139, D-4000 Dusseldorf 1, Germany

Many national standards are, at present, rapidly being superseded at an international level. The ISO has established a technical committee (TC 24) to standardize equipment and methods used in the size classification of particulate materials in solid or liquid states. TC 24 is divided into several subcommittees interested in test sieves, test sieving (SC 1); industrial wire screens (SC 3); and sizing by methods other than sieving (SC 4).

The work groups (WG) within SC 4 are listed in Table 10. It can be seen from the titles of the work groups that the SC 4 group has a very large remit, with the responsibility of attempting to standardize a wide range of particle size, powder surface area and powdered material porosity measurement techniques. Some of the standardization committees working in the field of powder technology are also listed in Table 10. The members of TC 24/SC 4 are Australia, Belgium, China, Finland, France, Germany, Israel, Japan, Norway, Sweden, The Netherlands, the UK and the USA, who nominate delegates to work in various powder technological fields. Table 10 also shows some of the approved and published standards together with the progress of other work groups within TC 24/SC 4.

6.2 Reference Materials

In all particle size instruments available at present, except microscopes, an assessment is made of particle size distribution by recording values of a size-dependent physical property analogous to particle size. While this is not ideal it has long been accepted that if the limitations can be assessed by users and recipients of the data obtained

Table 17 National (BS and DIN) standards relevant to powder technology

BS 3406: Methods for the determination of particle size distribution	
Part 1: 1986:	Guide to powder sampling
Part 2: 1984:	Recommendations for gravitational liquid, sedimentation methods for powders and suspensions
Part 3: 1963: (reconfirmed 1983)	Air elutriation methods
Part 4: 1993:	Recommendations for electrical sensing zone method (the Coulter principle)
Part 5: 1983:	Guide to microscope and image analysis methods
Part 6: 1985:	Recommendations for centrifugal liquid sedimentation methods for powders and suspensions
Part 7: 1988:	Recommendations for single particle light interaction methods
Part 8:	Photon correlation spectroscopy
Part 9:	Recommendations for the filter blockage method (mesh obscuration)
BS 4359: Determination of the specific surface area of powders	
Part 1: 1984:	BET method of gas adsorption for solids (including porous materials)
Part 2: 1982:	Recommended air permeability methods
Part 3: 1970:	Calculation from the particle size distribution (withdrawn)
Part 4: 1995:	Recommendations for methods of determination of metal surface area using gas adsorption techniques
BS 2955: 1993: Glossary of terms relating to particle technology	
BS 7591: Porosity and pore size distribution of materials	
Part 1: 1992:	Method of evaluation by mercury porosimetry
Part 2: 1992:	Method of evaluation by gas adsorption
Part 4: 1993:	Method of evaluation by liquid expulsion
DIN Representation of size distributions	
DIN 66 141 (1974)	Basic standard on representation of size distributions
DIN 66 142	Separation of dispersive materials
Part 1 (1981)	Fundamentals
Part 2 (1981)	Analytical separations
Part 3 (1982)	Technical separations
DIN 66 143 (1974)	Graphical representation of size: Power-function grid
DIN 66 144 (1974)	Graphical representation of size: Log-normal grid
DIN 66 145 (1974)	Graphical representation of size: RRSB grid
DIN Particle size analysis	
DIN 66 111 (1973)	Sedimentation analysis in gravitational field: Fundamentals
DIN 66 115 (1983)	Sedimentation analysis in gravitational field: Pipette method
DIN 66 111 (1973)	Sedimentation analysis in gravitational field: Sedimentation balance
DIN 66 118 (1984)	Air classification: Fundamentals
DIN 66 119 (1983)	Air classification with gravitation counterflow elutriator
DIN 66 120 (1983)	Air classification with centrifugal counterflow classifier
DIN 66 165 (1987)	Sieving analysis
Part 1	Fundamentals
DIN Sieves	
DIN 4185	Screen bottoms: Definitions and symbols Part 1 (1963), Part 2 (1981) and Part 3 (1965)
DIN 4186	Screening surfaces: Round metal-wires Parts 1 and 2
DIN 4187 (1974)	Screening surfaces: Perforated plates for test sieves Parts 1, 2 and 3
DIN 4188 (1977)	Screening surfaces: Wire screens for test sieves Parts 1 and 2
DIN 4189 (1968)	Screening surfaces: Woven wire cloth made of steel Parts 1 and 2
DIN 4192 (1968)	Screening surfaces: Woven, precrimped or welded cloth of high tensile steel. Part 1. Dimensions
DIN 4195 (1976)	Screening surfaces: Woven filament cloth of silk or man-made fibres: Dimensions.
DIN 4196 (1984)	Screening media: Round monofilament yarns: Designation, dimensions, requirements, testing
DIN 4197 (1976)	Screening surfaces: Woven filament cloth of round man-made monofilaments: Dimensions
DIN 24 041 (1981)	Round hole perforated plates: Dimensions
DIN 24 042 (1981)	Square hole perforated plates: Dimensions
DIN 24 043 (1981)	Oblong hole perforated plates: Dimensions

from the range of size-dependent physical criteria used in present day instruments, and possible measurement techniques of the future, then a nonabsolute size measurement of irregularly shaped particles is justifiable. It is thus incumbent upon designers, vendors and users of instruments to identify the assumptions made in relating the results obtained from any instrumention to the fundamental units of length, mass or time, or other mathematically derived or correlated units. The mechanism by which the above relationship is achieved within the instrument is known as calibration.[25]

Leschonski, foreseeing the need for instrument calibration in terms of size and the measurement of reference materials of known size certified on the basis of traceability, advocated that the most desirable approach to describe size would be that of the volume or, if the density of the material is accurately known, the mass of the particle or particles. This has resulted in the certification, by BCR, of irregularly shaped solids (quartz) over a range of distributed sizes using the off-line techniques of sieving (160–630 μm) and sedimentation (0.35–90 μm) and a range of nitrogen BET specific surface area reference materials.[128] Certified mono-sized polymer microspheres in discrete sizes from 20 nm to 2000 μm are also commercially available from BCR and NIST.

6.2.1 Calibration and Traceability

Calibration can take many forms depending upon the nature of the measurements being made. In microscopy a graticule is examined concurrently with a stage micrometer that has been graduated in authenticated linear units, traceable to primary standards, by a verified standard procedure. Providing that the conditions of examination are unchanged, the graticule can then be used to make ''absolute'' measurements on the images of objects of unknown size.

In some other particle sizing instrumention, spherical latex particles of known and verified uniform size can be used, so that the results produced by instruments such as laser light scattering and electrozone sensing instruments coincide with the verified values. Calibration materials or calibrants used with such instrumentation tend to consist of populations of particles whose distributions have been determined by a method that has been traceably calibrated and validated. The efficacy of a spherical calibration procedure is shape sensitive. Since a large percentage of calibration methods depend upon spherical calibrants, few problems are likely to arise if unknown test samples comprising only spherical particles are analyzed. However, since only a small percentage of industrial powdered samples exclusively consist of spherical particles a judgment has to be made as to the acceptability of the calibration in relation to the shape, or more likely shapes, present in the samples analyzed.

Whenever a device is calibrated by a user or a manufacturer it should be calibrated using suitable internationally certified reference standards, or secondary standards which are directly traceable to those international standards. For good laboratory practice, the whole measurement and calibration process must also be validated. This requires that objective evidence is established to show that the measurement process consistently produces results to within required specifications, with consideration being given to all procedural and analytical variations. When certified material that has a traceable source and pedigree is presented to another instrument that does not have the same physical basis of particle size measurement or the same size-dependent physical property, then a comparison is taking place and the certified material is acting as a reference material (Table 18).

However, if the certified material is used to adjust the output of the instrument to the values consensually accepted from the material then the instrument has been calibrated. A reference material may be used for the calibration and checking of any instrument used for particle sizing, but certified material is usually only certified for one or two fundamental methods. Care must be taken in calibrating and checking other particle sizing instruments that do not use the same physical principles as those upon which the certified material was certified. In general, materials that have a particle size distribution tend to be regarded as reference materials, while mono-sized materials tend to be used to calibrate instruments. Traceable calibrants and control reference materials (Tables 19 and 20) are available from, for example, BCR. Evidence of traceability must always be supported by a trail of documentation.

6.2.2 Validation and Verification

It is likely that all particle-sizing instruments will require both the traceable calibrant and the reference material approaches, in which different proportions of a series of mono-sized particulate material are mixed together to assess the parameters of reproducibility, repeatability, accuracy and possibly sensitivity of the instrument.

The validity of the software that nowadays governs and controls many particle sizing instruments will also have to be addressed, so that there is now a need to have three levels to test particle-sizing instruments:

1. verification of the software routines used to interpret the measured size-dependent physical criterion used to size irregular particles;
2. physical tests to ascertain the correctness of the size-dependent physical criterion measured with traceability;

Table 18 Distributors of reference materials

Acronym	Contact Information
APPIE	The Association of Powder Process Industry and Engineering, Shibunkaku-kaikan, 2-7, Tanaka-sekiden-cho, Sakyo-ku, Kyoto, 606, Japan
BAM	Bundesanstalt für Materialforschung und -prüfung, Unter den Eichen 87, D-12 205 Berlin, Germany
BAS	Bureau of Analyse Samples Ltd, Newham Hall, Newby, Middlesbrough, Cleveland TS8 9EA, UK
CTIF	Centre Technique des Industries de la Fonderie, 44 Avenue de la Division Leclerc, F-92310 Sevres, France
Duke Standards Co.	445 Sherman Avenue, Palo Alto, CA 94306, USA, and Eidgenossische Materialprüfungs- und Versuchsanstalt für Industrie Bauwesen und Gewerbe, Unterstraße 11, CH-900 St. Gallen, Switzerland
BCR	Bureau Communautaire de Réference, 200 rue de la Loi, B-1049 Brussels, Belgium, and Joint Research Centre, Institute for Reference Materials and Measurements (IRMM), Retiesweg, B-2440 Geel, Belgium
IRSID	Institut de Recherches de la Siderurgie Francaise, BP 64, F-57210 Maizieres-les-Metz, France
MBH	Analytical Ltd., Holland House, Queens Road, Barnet, Herts, EN5 4DJ, UK
NBL	US Department of Energy, New Brunswick Laboratory, Reference Materials Sales, 9800 S. Cass Avenue, Building 350, Argonne, IL 60439, USA
NIST	Office of Standard Reference Materials, US Department of Commerce, National Institute of Standards and Technology, B311 Chemistry Building, Gaithersburg, MD 20899, USA
NPL	National Physical Laboratory. Distributor: Office of Reference Materials
LGC	Laboratory of the Government Chemist, Building 95 RMAI, Queen's Road, Teddington, Middlesex, TW11 OLY, UK
Particle Information Service, Inc.	PO Box 792, 222 Granite Hill Road, Grands Pass, Oregon 97526, USA
REMCO	Council Committee on Reference Materials, International Organization for Standardization, 1, rue de Varembe, BP 56 CH-1211 Geneva 20, Switzerland
RBS	Regine Brooks, Pariser straße 5, D-5300 Bonn 1, Germany
SMR-LNE	Service des Materiaux de Reference, Boissier, F-75015 Paris, France
Silikose-Forschungsinstitut	Hunscherdstraße 12, D-4630 Bochum, Germany
Staubforschungsinstitut des Hauptverbandes der gewerbliche Berufsgenossenschaften eV	Langwartweg 103, D-5300 Bonn 1, Germany
Steinkohlenbergbauverein	Hauptstelle fur Staub- und Silikose-bekampfung, Frillendorferstraße 351, D-4300 Essen-Kray, Germany
Testfabrics Inc.	55 Vandam Street, New York, NY, USA
Waschereiforschung Krefeld eV	Adlerstraße 44, D-4150 Krefeld, Germany
Wirtschaftsverband Asbestzement eV	Arbeits- und Umweltschutz, Kolnerstraße 102–104, D-4040 Neuss, Germany

Table 19 Certified reference materials for use with particle characterization and measurement

Material	Product number	Distributor	Particle diameter (μm)	Specific surface area measured by nitrogen sorption at 77 K ($m^2 g^{-1}$)
α-Alumina	CRM 169	BCR		0.104
α-Alumina	CRM 170	BCR		1.05
α-Alumina	M 11-05/09	NPL		2.1, 0.7[a]
α-Alumina	M 11-06/10	NPL		0.3, 0.1[a]
α-Alumina	M 11-07/11	NPL		0.1, 0.04[a]
α-Alumina	M 11-08/12	NPL		1.0, 0.3[a]
Alumina	CRM 171	BCR		2.95
Alumina	8571	NIST/ASTM		158.7
Bronze	CRM 174	BCR		0.06
Graphitized carbon black				
Vulcan FT-G	M 11-01	NPL		11
Sterling 3-G	M 11-02	NPL		71

(*continued overleaf*)

Table 19 (*continued*)

Material	Product number	Distributor	Particle diameter (μm)	Specific surface area measured by nitrogen sorption at 77 K ($m^2 g^{-1}$)
Glass spheres	1003a	NIST	8–58	
Glass spheres	1993b	NIST	10–60	
Glass spheres	1004a	NIST	40–170	
Glass spheres	1017a	NIST	100–310	
Glass spheres	1017b	NIST	100–400	
Glass spheres	1018a	NIST	225–780	
Glass spheres	1018b	NIST	200–750	
Glass spheres	1019a	NIST	760–2160	
Kaolin, calcined	8570	NIST/ASTM		10.89
Latex spheres	CRM 165	BCR	2.223	
Latex spheres	CRM 166	BCR	4.821	
Latex spheres	CRM 167	BCR	9.475	
Polystyrene spheres	1690	NIST	0.895	
Polystyrene spheres	1691	NIST	0.269	
Polystyrene spheres	1692	NIST	2.982	
Polystyrene spheres	1660	NIST	9.89	
Polystyrene spheres	1661	NIST	29.64	
Polystyrene spheres	1665	NIST	9.94	
Silica	M 12-01	NPL		Mesoporous
Silica		NPL		Mesoporous
Silica		NPL		152 nonporous
Silica Gasil™	M 11-04	NPL		260 mesoporous
Silica/alumina	8572	NIST/ASTM		291.2
Silica nitride	CRM 659	BCR	0.2–10	
Quartz	CRM 066	BCR	0.35–3.50	Stokes diameter
Quartz	CRM 067	BCR	2.4–32	Stokes diameter
Quartz	CRM 068	BCR	160–630	Volume diameter
Quartz	CRM 069	BCR	14–90	Stokes diameter
Quartz	CRM 070	BCR	1.2–20	Stokes diameter
Quartz	CRM 130	BCR	50–220	Volume diameter
Quartz	CRM 131	BCR	480–1800	Volume diameter
Quartz	CRM 132	BCR	1400–5000	Volume diameter
Quartz	M 13-02	NPL	0.1–3	
Quartz	M 13-03	NPL	3–40	
Quartz	M 13-04	NPL	40–1000	
Quartz	M 13-05	NPL	10–100	
Quartz	M 13-06	NPL	1–10	
Quartz	66	NPL	3–20	
Quartz	67	NPL	0.35–2.5	
Quartz	68	NPL	140–650	
Quartz	69	NPL	12–90	
Quartz	70	NPL	0.5–90	
Quartz	CRM 172	BCR		2.50
Titania	M 13-01	NPL	0.1–3	
Titania/rutile	CRM 173	BCR		8.23
Tungsten	CRM 175	BCR	0.181	
Portland/cement	14n	NIST		0.202,[b] 0.346[b]
Zirconia	M 13-07	NPL	1–2	
Zirconia	M 13-08	NPL	5	
Zirconia	M 13-09	NPL	15	
Zirconia	M 13-10	NPL	30	
Zirconia	M 13-11	NPL	60	
Test dust	Carborundum	Particle Technology Ltd, Hatton, Derbyshire DE 65 5D4, UK.		
Test dirt	Spark Plug	Waschereiforschung, Krefeld ev, D-4150, Krefeld, Germany.		

[a] Specific surface area measured using a permeation method.
[b] Specific surface area measured using a Wagner turbidimeter.

Table 20 Reference materials in terms of shape and size for aerosols

Product number	Distributor	Particle diameter (μm)		Shape
AEA 1001	AEA Technology	3.09	Length	Fibrous
		1.67	Width	
		0.96	Depth	
	Aerodynamic diameter	2.89	Minor axis	
	Aerodynamic diameter	3.14	Major axis	
AEA 1002	AEA Technology	7.51	Length	Fibrous
		1.72	Width	
		1.02	Depth	
	Aerodynamic diameter	3.54	Minor axis	
	Aerodynamic diameter	4.11	Major axis	
AEA 1003	AEA Technology	12.13	Length	Fibrous
		1.70	Width	
		1.00	Depth	
	Aerodynamic diameter	3.78	Minor axis	
	Aerodynamic diameter	4.51	Major axis	
AEA 1004[a]	AEA Technology	0.1, 0.2, 0.5		Cocktail of spherical polymer latex particles
AEA 1004[a]	AEA Technology	0.2, 0.5, 1.0		Cocktail of spherical polymer latex particles
AEA 1004[a]	AEA Technology	1.0, 2.0, 5.0		Cocktail of spherical polymer latex particles

[a] Number concentration: particles per mL in a 10 mL volume is 4×10^8.

3. certified powdered materials with known sizes and size distributions to verify and confirm that the particles displayed by the instrument are in reality the physical particle sizes introduced into the instrument.

One way of providing confidence and acceptability in the particle sizes measured from the diverse range of particle size instruments available is by verification, which can be defined as: ''the examination and provision of objective documentary evidence that specified requirements have been fulfilled''. If the verification can be extended to demonstrate that an instrument or process is suitable for fulfilling the requirement of a specific intended use, then that instrument or process is said to be validated.

LIST OF SYMBOLS

A	Projected area of an irregularly shaped particle image
B	Breadth of an irregularly shaped particle
C	Sampling constant, which is a function of a material's properties and which can be determined
C	Shape constant which can be determined experimentally (section 5.1)
C'	Circumference of the projected area of a particle's image
d	Size of square mesh sieve to retain 5% of the particles or a specified percentage of a specified size range of particles
d_d	Drag diameter of an irregularly shaped particle
d_f	Free-fall diameter of an irregularly shaped particle
d_s	Equivalent surface area diameter for an irregularly shaped particle
d_{St}	Stokes diameter of an irregularly shaped particle
d_t	Sieve size diameter of an irregularly shaped particle
d_v	Equivalent volume area diameter of an irregularly shaped particle
k	Power of the moment of particle size x or the kth moment of a density distribution, which is related to the method used to calculate the average size of the particle within the distribution
L	Length of an irregularly shaped particle
m	Sample weight used in sampling (section 2.2)
m	Heywood's flatness ratio (ratio B/T)

m	Slope of the Gaudin–Schuman distribution (section 3.3.1)
M	Moment of a distribution
n	Heywood's elongation ratio (ratio L/B)
n	Slope of the Rosin–Rammler–Sperling–Bennet distribution (section 3.3.2)
N	Number of projections measured in Wadell's roundness
P	Wadell's roundness
q	Particle density distribution
$q_r(x)\,dx$	Fraction of particles in the size range $(x - dx)/2$ and $(x + dx)/2$.
Q	Cumulative particle density distribution
$Q_r\,(x)$	Cumulative particle density distribution of size x
r	Quantity of a particle distribution or type of statistical distribution
$R\,(x)$	Rosin–Rammler–Sperling–Bennet distribution
s^2(FE)	Variance of fundamental error of $(a_s - a_b)/a_b$ where a_s and a_b are content values of sample and bulk batch respectively.
S	Average particle surface area
T	Thickness of an irregularly shaped particle
V	Average volume of an irregularly shaped particle
V_{a-t} curve	Linear relationship, passing through the origin, between the volume adsorbed (V_a) and the thickness of the multilayer adsorbed on the nonporous solid (t)
x	Size characteristic of an irregularly shaped particle
x_a, x_t	Projected area and sieve diameter of an irregularly shaped particle
$x_{k,r}$	Average particle size/diameter
x_{max}, x_{min}	Largest and smallest particle sizes in a distribution
x_n	Equivalent mesh diameter
x_s	Equivalent surface diameter
x_v	Equivalent volume diameter
x_w	Equivalent settling diameter at either high or low Reynolds numbers
x'	Particle size of x corresponding to a cumulative percentage $Q_3(x)$ equal to an undersize of 63.2% or an oversize of 36.8%
$x_{1,2}$	Average length-surface or average volume-surface diameter of an irregularly shaped particle
x_{84}, x_{50}, x_{16}	Values of size at which $Q_r\,(x)$ is 84%, 50% and 16% respectively
z	Equivalent to the percentage of an arithmetic particle size distribution at a specific size x and at a size which occurs at 50% of the distribution ($z = (x - x_{50,r})/2$). For logarithmic size distributions $z = \ln(x/x_{50,r})$.
Z	Hausner shape factor
$(1 - D)$	The fractal dimension of the rugged boundary curve of an irregular particle
α	Shape factor
α_0	Equidimensional shape factor
α_s	Sing's 'alpha s' plot
α_s, α_v	Shape factors which denote the method of particle size measurement
α_v	Volume shape factor of an irregular particle
ψ	Wadell's sphericity shape factor

ABBREVIATIONS AND ACRONYMS

BCR	Bureau Communautaire de Réference
BDDT	Brunauer–Deming–Deming–Teller
BET	Brunauer–Emmett–Teller
BS	British Standard
FE	Fundamental Error
GC	Gas Chromatography
IE	Integration Error
ISO	International Standardization Organization
IUPAC	International Union of Pure and Applied Chemistry
LALLS	Low-angle Laser Light Scattering
NIST	National Institute of Standards and Technology
SE	Segregation Error
TOF	Time-of-flight
TOT	Time-of-transition

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Filtration in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Sieving in Particle Size Analysis • Surface Area and Pore Size Distributions • Ultrasonic Measurements in Particle Size Analysis

REFERENCES

1. B. Waldie, 'Size Control in Particle Growth Process', Industrial Research Fellowship Report, Institution of Chemical Engineers, Rugby, 1980.
2. R. Oliver, L.J. Ford, *Proceedings of the Institute of Briquetting and Agglomeration*, IBA, Erie, PA, 20, 1987.
3. B. Waldie, *Proceedings of the 5th International Symposium on Agglomeration, Brighton*, Institution of Chemical Engineers, Rugby, 1989.
4. N. Stanley-Wood, M. Sarrafi, Z. Mavere, M. Schaefer, *Adv. Powder Technol.*, **4**(1), 33–40 (1993).
5. K. van der Voort Maarschalk, G.K. Bolhuis, *Pharm. Technol. Eur.*, **10**(9), 30–35 (1998).
6. K. van der Voort Maarschalk, G.K. Bolhuis, *Pharm. Technol. Eur.*, **10**(10), 28–36 (1998).
7. H. Heywood, 'The Origins and Development of Particle Size Analysis', in *Proceedings of Particle Size Analysis 1970*, eds. M.J. Groves, J.L. Wyatt-Sargent, Society for Analytical Chemistry, London, 1970.
8. C. Singer, E.J. Holmyard, A.R. Hall, T.T. Williams, *History of Technology*, Clarendon Press, Oxford, Vols. 1–5, 1954–1956.
9. M.B. Donald, *Elizabethan Copper, The History of the Company of Mines Royal 1568–1605*, reprinted by Michael Moon, Whitehaven, 1989.
10. G. Agricola, *De Re Metallica*, 1556 (in Latin). Translated by H.C. Hoover, L.H. Hoover, Dover Publications, New York, 1950.
11. L. Ercker, *Mineralische Ertzt*, Book 111, 1574 (in German).
12. J.M. DallaValle, *Micromeritics. The Technology of Fine Particles*, Pitman Publishing, 1943.
13. G. Herdan, *Small Particle Statistics*, Butterworth, Borough Green, Kent, UK, 1952.
14. H. Rumpf, *Winnacker-Küchler, Chemische Technology*, 3rd edition, Carl Haüser Verlag, München, Vol. 7, 1975.
15. W. Pietsch, *Size Enlargement by Agglomeration*, John Wiley, Chichester, 1991.
16. H. Rumpf, *Particle Technology*, Chapman and Hall, London, 1990.
17. A. Trachtler, F. Mesch, *Part. Part. Syst. Charact.*, **15**, 30–35 (1998).
18. D.M. Scott, A. Boxman, R. Davies, 'World Congress on Particle Technology 3', Brighton, UK, July 1998, Institution of Chemical Engineers, Rugby (CD-ROM).
19. P.J. Gregory, X. Jia, S.P. Luke, T.M. Shi, R.A. Williams, F.J. Dickin, L.F. Gate, S.E. Taylor, 'World Congress on Particle Technology 3', Brighton, UK, July 1998, Institution of Chemical Engineers, Rugby (CD-ROM).
20. H. Tabe, S.J. Simons, R.M. West, R.A. Williams, 'World Congress on Particle Technology 3', Brighton, UK, July 1998, Institution of Chemical Engineers, Rugby (CD-ROM).
21. R.R. Irani, C.F. Callis, *Particle Size: Measurement, Interpretation and Application*, John Wiley, New York, 1963.
22. H. Heywood, *Chemistry and Industry*, **15**, 149–154 (1937).
23. H. Heywood, *J. Imp. College, Chem. Engng. Soc.*, **8**, 25–33 (1954).
24. H. Heywood, *Proceedings of the 1st Particle Size Analysis Conference*, Heffer, London, 355–359, 1966.
25. 'Glossary of Terms Relating to Particle Technology', BS 2955, British Standards, London, 1993.
26. J. Meyer, M. Katzer, E. Schmidt, S. Cihlar, 'World Congress on Particle Technology 3', Brighton, July 1998, Institution of Chemical Engineers, Rugby, 1998 (CD-ROM).
27. M. Turk, R.W. O'Brien, W.N. Rowlands, S.E. Gibb, T.A. Wade, R.J. Hunter, 'World Congress on Particle Technology 3', Brighton, July 1998, Institution of Chemical Engineers, Rugby, 1998 (CD-ROM).
28. *Guide to Powder Sampling*, BS 3406, Part 1, British Standards, London, 1986.
29. P.M. Gy, *Sampling of Particulate Materials: Theory and Practice*, Elsevier Scientific, 1979.
30. A.B. Manning, *J. Inst. Fuel*, **9**, 132–136 (1935).
31. P. Gy, *Annales Mines*, **145**, 37–42 (1956).
32. P. Gy, *Trans. Inst. Min. Metall.*, **74**, 164–199 (1964/1965).
33. K. Stange, *Chem. Eng. Tech.*, **35**(8), 580–582 (1963).
34. H.G. Kirstensen, *Powder Technol.*, **7**, 249–297 (1973).
35. 'Representation of Results of Particle Size Analysis. Part 1: Graphic Representation', ISO 9276-1, 190, Geneva, 1995.
36. 'Certification Report on Reference Materials of Defined Particle Size: Quartz', BCR No. 66–70, BCR Information EUR 6825 EN, Commission of the European Communities, Luxembourg, 1980.
37. K. Leschonski, *Particle Characterization*, **1**, 89–95 (1984).
38. Basic Standard on Representation of Size Distributions, DIN, 66 141, 1974.
39. 'Calculation of Average Article Sizes/Diameter', ISO 9276-2, Deutsches Institut für Normung, Berlin, 1998.
40. S.L. Urman, *J. Aerosol Sci.*, **29**, 537–552 (1998).
41. F.E. Kruis, H. Fissan, A. Peled, *J. Aerosol Sci.*, **29**, 511–535 (1998).
42. D.B. Kittleson, *J. Aerosol Sci.*, **29**, 575–588 (1998).
43. 'Air Quality–Particle Size Fraction Definitions for Health-related Sampling', ISO 7708, ISO, Geneva, 1992.
44. Quality of Urban Air Review Group, 'Diesel Vehicle Emissions and Urban Air Quality', 3rd Report, UK Department of the Environment, HMSO, London.
45. TPL Inc., Advanced Technologies Division, 3921 Academy Parkway North NE, Albuquerque, New Mexico.
46. Environmental Protection Agency, 'Revisions to the National Ambient Air Quality Standards for Particulate Matter', *Federal Register*, **52**(126), 24 634–24 669 (1987).
47. N. Stanley-Wood, *Trends in Analytical Chemistry*, **6**, 100–105 (1987).
48. N. Stanley-Wood, *Powder Metallurgy International*, **9**, 27–29 (1977).

49. N. Stanley-Wood, *Powder Metallurgy International*, **9**, 84–87 (1977).
50. N. Stanley-Wood, *Powder Metallurgy International*, **9**, 138–142 (1977).
51. T. Allen, R. Nelson, *Powder Technol.*, **88**, 11 (1996).
52. H.C. van der Hulst, *Light Scattering by Small Particles*, John Wiley, New York, 1957.
53. M. Born, E. Wolfe, *Principles of Optics*, Pergamon Press, New York, 1965.
54. G. Mie, *Annalen der Physik*, **25**, 377–445 (1908).
55. S. Brunauer, P.H. Emmett, E.J. Teller, *J. Am. Chem. Soc.*, **60**, 309 (1935).
56. 'Porosity and Pore Size Distribution of Materials. Part 1: Method of Evaluation by Mercury Porosimetry', BS 7591, Part 1, British Standards, BSI, London, 1992.
57. 'Porosity and Pore Size Distribution of Materials. Part 2: Method of Evaluation by Mercury Porosimetry', BS 7591, Part 2, British Standards, BSI, London, 1992.
58. N. Stanley-Wood, *Advances in Mechanics and Flow of Granular Materials*, ed. H. Shaninpoor, Tran Tech Pub., Clausthal, Vol. 1, 1983.
59. 'Determination of the Specific Surface Area of Powders. Part 1. BET Method of Gas Adsorption for Solids (Including Porous Materials)', BS 4359, (ISO 9277: 1995), British Standards, London, 1996.
60. 'Determination of the Specific Surface Area of Solids by Gas Adsorption Using the BET Method', ISO 92777: 1995(E), Commission of the European Communities, Luxembourg, 1995.
61. K.S.W. Sing, D.H. Everett, R.A.W. Haul, L. Moscou, R.A. Pierotti, J. Rouquerol, T. Siemieniewska, 'Reporting Physisorption Data for Gas/Solid Systems with Special Reference to the Determination of Surface Area and Porosity', *Pure and Appl. Chem.*, **57**(4), 603–619 (1985).
62. B. Elbirli, L. Vergara-Edwards, R. Coughlin, *J. Colloid Interface Sci.*, **87**, 31–44 (1982).
63. N. Stanley-Wood, M. Sarrafi, S. Lagardes, *Proceedings of the Powder and Bulk Solid Conference*, Rosemont, Reed Exhibition Companies, Norwalk, CT, 1988.
64. G.F. Huttig, *Mschr. Chem.*, **78**, 177 (1948).
65. M.M. Dubinin, V.A. Astakhov, *Bull. Acad. Sci. USSR*, **20**, 3 (1971).
66. N.G. Stanley-Wood, *Enlargement and Compaction of Particulate Solids*, Butterworths, Cambridge, 1983.
67. *IUPAC Pure Appl. Chem.*, **57**, 603 (1985).
68. S. Brunauer, L.S. Deming, W.S. Deming, E.J. Teller, *J. Am. Chem. Soc.*, **62**, 1723 (1940).
69. L.G. Gurvitsch, *J. Russ. Phys. Chem. Soc.*, **47**, 805 (1915).
70. B.C. Lippens, B.G. Linsen, J.H. de Boer, *J. Catal.*, **3**, 32 (1964).
71. S.J. Gregg, K.S.W. Sing, *Adsorption, Surface Area and Porosity*, Academic Press, London, 1982.
72. K.S.W. Sing, in *Characterization of Powder Surfaces*, eds. G.D. Parfitt, K.S.W. Sing, Academic Press, London, 1976.
73. K.S.W. Sing, *J. Oil Colour. Chem. Assoc.*, **54**, 731 (1971).
74. S. Brunauer, R.S.L. Mikhail, E.E.J. Boder, *J. Colloid Interface Sci.*, **26**, 45–53 (1965).
75. S. Brunauer, R.S.L. Mikhail, E.E.J. Boder, *J. Colloid Interface Sci.*, **26**, 54–61 (1965).
76. M.M. Dubinin, L.V. Radushkevich, *Proc. Acad. Sci. USSR*, **55**, 331 (1947).
77. C. Orr, *Analytical Chemistry in Industry*, John Wiley, New York, Vol. 4, Part 3, 1977.
78. V. van Brakel, S. Modry, M. Svara, *Powder Technology*, **29**, 1–12 (1981).
79. C. Orr, *Proceedings of the PSA Conference, Salford, 1977*, ed. M.J. Groves, Heyden, London, 1977.
80. K.L. Leschonski, *I. Chem. E. Sym.*, **69**, 147–162 (1983).
81. B. Scarlett, *Proceedings of the 25th Anniversary Conference of the Particle Characterization Group of the Royal Society of Chemistry*, eds. N. Stanley-Wood, R. Lines, Royal Society of Chemistry, Cambridge, 1992.
82. N. Stanley-Wood, *Control and Instrument.*, **6**, 42–56 (1974).
83. N. Stanley-Wood, *Control and Instrument.*, **7**, 30–37 (1975).
84. K.L. Leschonski, *Proceedings of the PSA Conference, September 1981*, eds. N. Stanley-Wood, T. Allen, John Wiley, New York, 1982.
85. P.G.L. van de Wel, S.M. Lanbowity, C.J.M. van Wingerden, B. Scarlett, *Proceedings of the 25th Anniversary Conference of the Particle Characterization Group of the Royal Society of Chemistry*, eds. N. Stanley-Wood, R. Lines, Royal Society of Chemistry, Cambridge, 1992.
86. H.E. Rose, *Powders in Industry*, Society for Chemical Industry, London, 130, 1961.
87. B.P. Mandelbrot, *Fractals, Forms, Chance and Dimension*, W.H. Freeman, San Francisco, 1977.
88. H.-O. Peitgen, H. Jurgens, D. Saupe, *Chaos and Fractals*, Springer-Verlag, 1992.
89. B. Kaye, *Direct Characterization of Fineparticles*, John Wiley, New York, 1981.
90. A. Hawkins, *The Shape of Powder–Particle Outlines*, ed. G.W. Greenwood, Research Studies Press, Taunton, and John Wiley, UK, 1993.
91. T. Allen, *Proceedings of Second International Conference on Particle Size Analysis*, Society for Analytical Chemistry, London, 1970.
92. R. Davies, *Powder Technol.*, **12**, 111 (1975).
93. H. Heywood, *Symposium on Particle Size Analysis*, Institution of Chemical Engineers, London, Suppl. 25, 14, 1947.
94. H. Heywood, *J. Pharm. Pharmac. Suppl.*, **15**, 56T (1963).
95. T. Church, *Powder Technol.*, **2**(1), 27 (1968).
96. J.L. Dwyer, D.A. Manalon, R.R.A. Merbon, *Proceedings of Second International Conference on Particle Size Analysis*, Society for Analytical Chemistry, London, 1970.
97. B. Mandelbrot, *Science, N.Y.*, **156**, 636 (1957).

98. B. Kaye, G. Clark, *Part. Part. Syst. Charact.*, **9**(1), 1–82 (1992).
99. H. Wadell, *J. Geol.*, **43**, 250 (1935).
100. H. Wadell, *J. Geol.*, **41**, 310 (1933).
101. H. Wadell, *J. Geol.*, **40**, 443 (1932).
102. G. Treasure, *Storage and Recovery of Particulate Solids*, ed. A. Richards, Institution of Chemical Engineers, London, 1966.
103. H.H. Hausner, *Proceedings of Conference on Particle Size Analysis*, Society for Analytical Chemistry, London, 1966.
104. H.P. Schwartz, K.C. Shane, *Sedimentology*, **13**, 213–231 (1969).
105. A. Flook, *Proceedings of the 4th Particle Size Analysis Conference, Loughborough University of Technology, September, 1981*, eds. N. Stanley-Wood, T. Allen, John Wiley, Chichester, 1982.
106. J.K. Beddow, A.F. Vetter, K. Sissan, *Powder Metall. Int.*, **8**(2), 69 (1976).
107. A.G. Naylor, C.D. Wright, *Proceedings of Third International Conference on Particle Size Analysis*, ed. M.J. Groves, Heyden & Sons Ltd., London, 1978.
108. B. Erlich, B. Weinberg, *J. Sediment. Petrol.*, **40**, 205 (1970).
109. T.P. Meloy, *Powder Technol.*, **17**, 27 (1977).
110. B.H. Kaye, *Powder Technol.*, **21**, 1 (1978).
111. L. Austin, M. Yekeler, T. Dumm, R. Hogg, *Part. Part. Syst. Charact.*, **7**(4), 169–254 (1990).
112. B. Kaye, *Part. Part. Syst. Charact.*, **10**(3), 99–110 (1993).
113. A.G. Flook, *Powder Technol.*, **21**, 295 (1978).
114. C. Nystrom, N.G. Stanley-Wood, *Acta Pharm. Suecia*, **13**, 277 (1976).
115. C. Nystrom, N.G. Stanley-Wood, *Acta Pharm. Suecia*, **14**, 181 (1977).
116. N. Pilpel, *Paint Mfn.*, July, 23 (1969).
117. C.A. Walton, N. Pilpel, *J. Pharm. Pharmac.*, **24**, 10P–16P (1972).
118. D.M. Scott, A. Boxman, C. Jochen, *Part. Part. Syst. Charact.*, **15**, 47–50 (1998).
119. G. Bonifazi, F. Francini, V. Guamierri, G. Longobardi, P. Massacci, M. Recinella, *Part. Part. Syst. Charact.*, **12**, 318–323 (1995).
120. C. Castellini, F. Francini, G. Longobardi, E. Pampaloni, *Part. Part. Syst. Charact.*, **10**, 7–10 (1993).
121. Galai Production Ltd., Ramat Gabriel Ind. Park, 10 500 Migdal Haemek, Israel.
122. Focused Beam Reflectance Measurement, [FBRM] Lasentec NE 95th Street Redmond.
123. C.M. Heffels, D. Heitzmann, E. Dan Hirleman, B. Scarlett, *Part. Part. Syst. Charact.*, **11**, 194 (1994).
124. A.D. Ah Chin, P.B. Bulter, D.W. Luerkens, *Powder Technol.*, **54**, 99–105 (1988).
125. R. Davies, *Powder Technol.*, **88**, 1 (1996).
126. B. Scarlett, *Proceedings of the PSA Conference, September 1981*, eds. N. Stanley-Wood, T. Allen, John Wiley, Chichester, 1982.
127. K.L. Leschonski, *Proceedings of the PSA Conference, September 1981*, eds. N. Stanley-Wood, T. Allen, John Wiley, Chichester, 1982.
128. 'The Certification of Nitrogen BET Specific Surface Areas of Six Materials in the Range 0.1 to 8 $m^2 g^{-1}$, Report EUR 12025 EN, BCR, Brussels, 1989.

Centrifugation in Particle Size Analysis

Walter Mächtle
BASF Aktiengesellschaft, Ludwigshafen, Germany

1 Introduction 1
2 History of Ultracentrifugation Particle Size Analysis 2
3 Experimental Set-up 2
4 Physical Principles 4
5 Water/Deuterium Oxide Sedimentation Analysis 5
6 Coupling Particle Size Distribution Technique 7
7 Examples of Analytical Ultracentrifugation Particle Size Distribution Technique 8
8 Outlook – The New Two-laser-beam Technique and Very Small (1–20 nm) Particles 9
Abbreviations and Acronyms 12
Related Articles 12
References 12

*Millions of tons per year of submicron particles are produced in industrial processes around the world. They are also referred to as nanoparticles, colloids, latices, polymer dispersions, mini-emulsions or micro-emulsions. These solid or liquid particles, which are mostly dispersed in an aqueous medium, have a diameter (*D*) in the* $10 < D < 3000$ *nm range. They are used as dyes and pigments, in paints, adhesives and coatings applied to paper, as impact modifiers in plastics, as catalysts, in cosmetic and photographic emulsions, metal sols and electrodeposition coatings applied to automobiles, as excipients for pharmaceuticals, etc.*

Most of the valuable practical properties of these nanoparticle dispersions/emulsions result from their extremely small size and their high specific surface area. This implies that it is essential to be able to control and measure the particle size in order to produce nanoparticle systems of consistent quality. The particle size distribution is just as important as the average particle diameter. The practical consequences of changing from a very narrow to a very broad particle size distribution with the same average diameter can be quite dramatic in many applications. For example, the viscosity of a high-solid polymer latex will decrease, the impact strength of plastic material will increase, surfaces coated with emulsion paints will become more level, etc.

Many methods, such as light scattering, electron microscopy, field flow fractionation (FFF), capillary hydrodynamic chromatography and ultracentrifugation, are used to measure the average particle diameters and overall particle size distribution of nanoparticles. Analytical ultracentrifugation (AUC) is the most versatile method with the broadest diameter range (1–5000 nm) and the highest resolution. The nanoparticles are forced to move through a predominantly aqueous medium inside an AUC measuring cell under a centrifugal force of 500–500 000 × g *(rotor speed 100–60 000 rpm). Sedimentation causes fractionation to take place, because the larger particles move faster. The sedimentation velocities* s_i *of the different fractions* i *are measured with refractive index, ultraviolet (UV)-absorption or light scattering/turbidity detectors. The diameters* D_i *of the different fractions can be calculated from the difference in* s_i*, i.e. the complete particle size distribution, according to Stokes' law. The resolution of the AUC method is so high because, according to Stokes,* s_i *is proportional to the square of* D_i*. The AUC particle size distribution method can give spurious results if* D *is too high, i.e. if the particles form a sediment too quickly or if they display a broad chemical heterogeneity (i.e. no uniform particle density).*

1 INTRODUCTION

Submicron or nanoparticles with diameters in the $10 < D < 3000$ nm range are produced and consumed in large quantities around the world, mostly in the form of aqueous dispersions or emulsions. Important applications for polymer dispersions (latices) are in paints, papermaking and adhesives, for modifying the impact strength of plastics and in the production of synthetic rubber. Their particle size distribution is one of their most important characteristics, because it affects many of their practical properties. Many different techniques have been developed to measure particle size distribution, such as light scattering, hydrodynamic chromatography, electron microscopy, sedimentation and FFF and AUC. The AUC method is the most versatile, high-resolution, relatively fast technique,[1] and it also covers the greatest diameter range ($1 < D < 5000$ nm).

Three AUC methods are described in this paper,

1. standard particle size distribution,
2. H_2O/D_2O sedimentation analysis (which is also used to measure particle densities)
3. the coupling method (which is used to measure extremely broad particle size distributions).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

All three methods are based on Stokes' equation and on Mie's light scattering theory, which means that light scattering takes place inside the AUC apparatus. After briefly recounting the history of the AUC method, we will go on to look at the experimental set-up with an eight-cell rotor and multiplexer and its physical principles, H_2O/D_2O sedimentation analysis and the coupling particle size distribution method with the aid of some illustrative examples. We will then consider future prospects by looking at a new two-laser-beam technique and AUC particle size distribution measurements of extremely small inorganic colloids (1–20 nm) with angstrom resolution, using very high rotor speeds and refractive index or absorption optics detectors.

2 HISTORY OF ULTRACENTRIFUGATION PARTICLE SIZE ANALYSIS

The first AUC particle size distribution measurements on colloidal particles were performed in 1924 by Rinde and Svedberg.[2,3] (Svedberg invented AUC, and was awarded the 1926 Nobel Prize in Chemistry for his work.) They analyzed gold sols, measured their sedimentation velocity s, and used the results to calculate the diameter D of the gold particles using Stokes' law. Nichols, Kraemer and Bailey[4] later improved the initial AUC particle size distribution set-up. Cantow[5] was the first to measure polymer dispersions with a turbidity detector inside the AUC apparatus, using Mie's light scattering theory[6] of spherical-shaped colloidal particles to calculate the correct mass percentages m_i of the different particle fractions. Later, Scholtan and Lange[7] developed another, very simple AUC particle size distribution set-up which employed a small one-beam turbidity/light scattering detector. This set-up was further developed by Mächtle,[8] who introduced the following new features: (1) a change from a single-beam to a double-beam technique; (2) an increase in the measuring capacity by a factor of seven by the introduction of an eight-cell rotor and a multiplexer, which allows simultaneous measurement of seven samples and a reference cell; (3) a modification from the common operation mode using a constant rotor speed N to a new time-dependent operation mode $N(t)$ in which the rotor speed is increased exponentially during each run from $N = 0$ to the maximum speed $N = 40\,000$ rpm within 1 h, always at the same rate. This allows samples containing very small particles of about 10 nm and very large particles of about 3000 nm to be measured in the same run. The multiplexer and $N(t)$ technique were prerequisites for H_2O/D_2O sedimentation analysis[8] and the coupling particle size distribution method.[9,10] Our latest development,[11] the introduction of a parallel two-laser-beam set-up, has increased the resolution and the accuracy of AUC particle size distribution measurements. Extremely small inorganic colloids in the 1–20 nm range have become a new issue over the last three years. Their light scattering is so weak that the turbidity/light scattering detector inside the AUC apparatus cannot be used. Instead modern UV scanners or refractive index detectors (schlieren or interference optics) and constant high rotor speeds are used, as for dissolved macromolecules.[12]

3 EXPERIMENTAL SET-UP

The photograph in Figure 1 shows our AUC set-up outside the rotor chamber. It includes eight measuring cells, an eight-cell rotor and a U-shaped stage, with the laser diode light source above the rotor and a fast photo diode light detector below the rotor. This set-up also contains a preparative ultracentrifuge (Optima XL, Beckman-Coulter, Palo Alto, CA, USA) which we have modified into an analytical ultracentrifuge by installing an analytical eight-cell rotor (Beckman type AN-50 Ti) and simple turbidity/light scattering optics. We have two versions of this turbidity detector, an older one with an incandescent lamp and a photomultiplier as the light source and light-receiving detector, and a newer one with a 670 nm red light laser diode and a very fast photo diode

Figure 1 Laser optical set-up showing rotor and cells, with a laser diode above the rotor, and a photo diode below the rotor.

(see Figure 1). An AUC measuring cell is situated to the right-hand-side of this eight-cell rotor in Figure 1. The centre piece has a thickness of 3 mm and a sector-shaped cut-out. A flat glass window fitted with a gasket is placed on either side and the gap is filled with the nanoparticle dispersion/emulsion to be analyzed by injecting it through a small hole using a syringe.

A schematic diagram of our set-up is shown in Figure 2(a). The turbid dispersion to be analyzed is diluted to about 1 g L^{-1} and placed into the centre section of the 3-mm cell. The lower quartz window of the cell is covered by an apertured stop, in the centre of which is a 0.2-mm wide slit that is arranged perpendicular to the radius of the rotor. The slit picks out a beam from the entering monochromatic light.

The intensity I of this beam, which is reduced by light scattering of the latex particles according to Mie's theory, is registered by the photomultiplier and recorded as a function of the time of travel t. The concentration of the dispersions is selected so as to yield an initial light intensity I_0 of approximately 10% of I_{DM}, the intensity of the pure dispersant. In our newer version, we use a 0.2-mm laser beam, no apertured stop, and a fast photo diode as the light detector.

In a monodisperse latex, all of the particles settle out with exactly the same velocity. This causes the broken one-step curve shown in Figure 2(b) to be obtained, because the intensity jumps from I_0 to I_{DM} at the moment when the sharply defined, sedimenting latex front passes the measuring slit (or the laser beam). The diameter D of the monodisperse particles can be calculated from the measured jump time t by means of Stokes' law.

In the event of a latex having a broad particle size distribution, the particles are fractionated within the sedimenting latex front according to their size, but the latex front now is blurred by large particles running ahead and small particles lagging behind. This gives a broad continuous transition from I_0 to I_{DM} (thick line). If we approximate this transition by plotting a multiple-step curve (fine line), i.e. by superimposing many monodisperse latex fractions, the $I(t)$ curve renders a curve for the integral particle size distribution (Figure 2c). This is obtained by calculating (a) the diameters D_i of the different fractions from their time of travel t_i, by means of Stokes' law, and (b) their mass fractions m_i from the respective light intensity steps ΔI_i by means of Mie's theory. The sum of all m_i values is unity. Details for calculating m_i with Mie's theory are given in the literature.[8,9]

It should be noted that this particle size distribution analysis strictly only works for hard spheres at a standard concentration of about 1 g L^{-1} ($\approx$ case of infinite

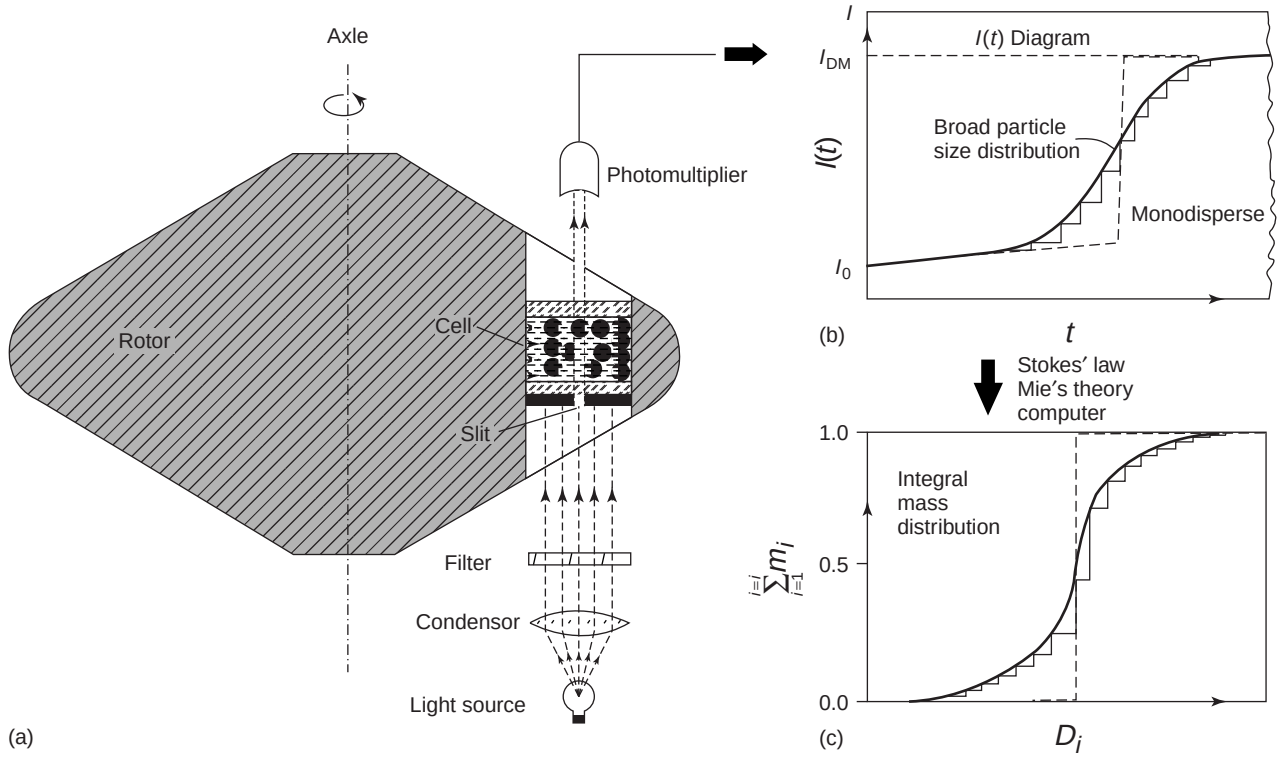

Figure 2 (a) Schematic diagram of the measuring set-up and (b) distribution curves. (Reproduced from Mächtle[9] by permission of Wiley-VCH Verlag.)

dilution). For rods or soft spheres, the dependency of the sedimentation velocity on the concentration might cause errors.

The process we have developed is much more complex than the basic procedure outlined here. In fact, we measure seven samples simultaneously, employing a quasi-two-light-beam technique. A schematic diagram of our method is shown in Figure 3. The process is triggered by reference cell 0 which is filled with the pure dispersant (mostly water). The amplitude of the photomultiplier (or photo diode) impulses from measuring cells 1–7 is measured and divided by the amplitude of the impulses from the cell containing only water using a fast electronic device, then digitized for computer processing, and recorded in analog form by an eight-channel recorder as apparently continuous $I(t)$ fractionation curves (for visual checks only). We no longer keep the rotor speed w constant: instead, we increase it exponentially from 0 to 40 000 rpm over 1 or 2 h during each run in order to cover the smallest particles ($D \approx 10$ nm) as well as the largest ones ($D \approx 5000$ nm). This procedure is fully automatic. The ability to measure seven cells simultaneously and to increase the rotor speed exponentially is a precondition for the two methods that will be described later, H_2O/D_2O analysis[8] and the coupling particle size distribution technique.[9]

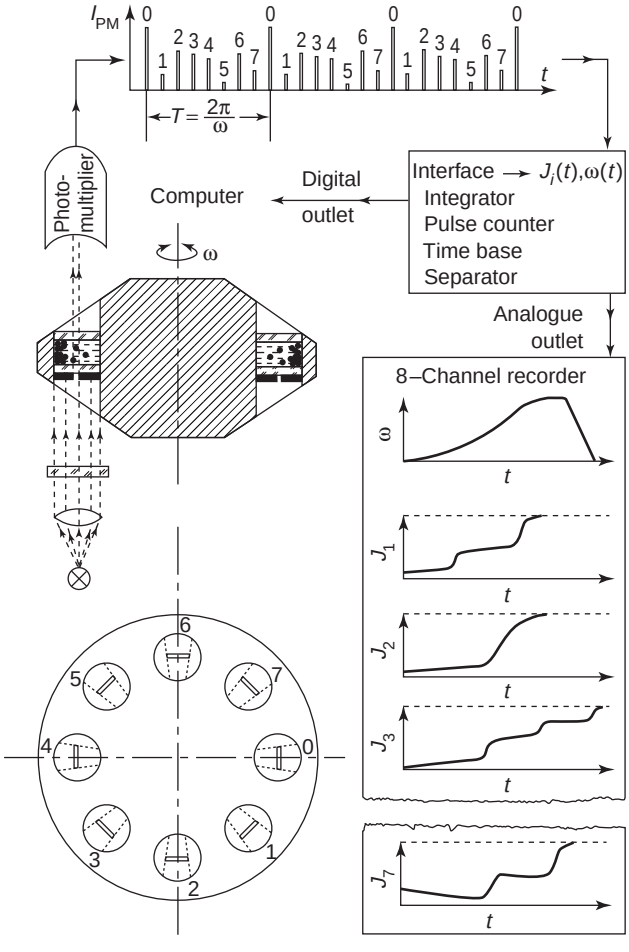

Figure 3 Schematic diagram of AUC method for determining particle size distribution, with an eight-cell rotor multiplexer set-up. (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

4 PHYSICAL PRINCIPLES

Figure 4 shows the basic equations for the determination of both the particle diameter D and the particle density ρ by Stokes' law. It also shows two "snapshots" of monodisperse latex particles in the cell, one of them taken at the start of the measurement ($t = 0$), the other one exactly at the point ($t = t_i$) at which the latex front passed across the measuring slit (or laser beam). A differential equation of motion can be calculated from the equilibrium of forces F acting on an arbitrary particle. The solution of the latter is Stokes' square root equation which allows D_i to be calculated from the measured time of travel t_i or, alternatively, from the sedimentation velocity $s_i = \ln((r_s/r)/w^2 t_i)$ if the particle density $\rho_i = \rho_{PM}$ is known. An unknown ρ_{PM} can be determined by measuring the time of travel in a second dispersant of different density ρ_{DM}. We employ heavy water, D_2O, which has a density of $1.10\,g\,cm^{-3}$, as the second medium. With the modified Stokes' equation (Figure 4, bottom), ρ_{PM} can be determined simply from measured times of travel in H_2O and D_2O, and this equation is the basis of the H_2O/D_2O analysis. There are also other methods of measuring the particle density ρ_{PM}: pycnometry, measurement on bulky materials, AUC density gradient techniques and the Paar density balance. It must be emphasized that the differential equation mentioned here can be solved for the case $w =$ constant and for our new mode of operation with $w = f(t)$. In the "classic" square root equation, $w^2 t$ changes to $\int w^2\,dt$, the integral of the time of travel, which is calculated from the measured curve for $w(t)$ using a computer.

The measurement of seven samples simultaneously is demonstrated in Figure 5. We made up a $1\,g\,L^{-1}$ mixture of three monodisperse Dow polystyrene calibration latices with diameters of 176, 312 and 794 nm at a ratio of 40 : 50 : 10% w/w. This mixture was filled into seven different cells and initially measured in the conventional operating mode at a constant rotor speed $N = 4000$ rpm. As a result, we obtained the seven full line $I(t)$ curves. The measurement was then repeated in the new operating mode with N rising exponentially from 0 to 40 000 rpm, which generated the seven broken line $I(t)$ curves. The respective seven three–step curves obtained in both operating modes agree well.

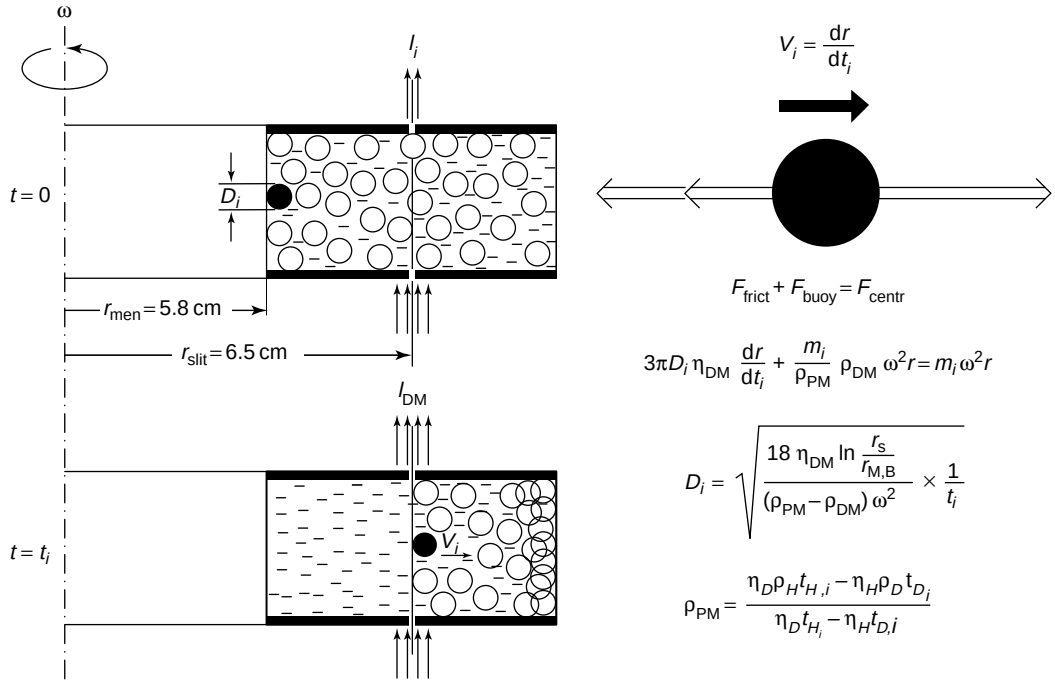

Figure 4 Determination of D and ρ according to Stokes' law. (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

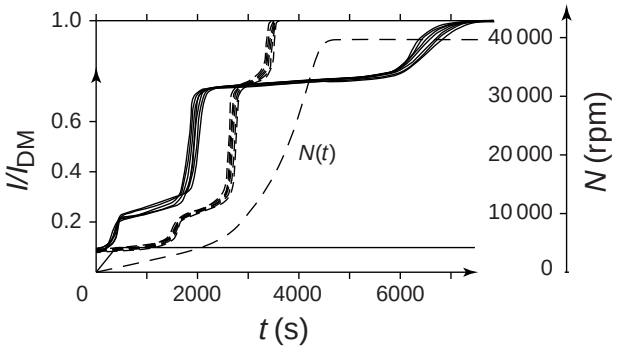

Figure 5 $I(t)$ curves of seven mixtures of three monodisperse polystyrene calibration latices, 176:312:794 nm (40:50:10% w/w). (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

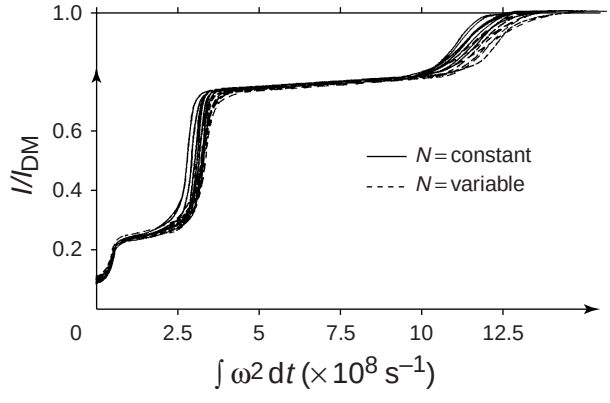

Figure 6 $I(\int w^2 \mathrm{d}t)$ curves of seven mixtures of three monodisperse polystyrene calibration latices, 176:312:794 nm (40:50:10% w/w). (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

If I/I_{DM} is plotted against $\int w^2 \mathrm{d}t$ instead of against t as shown in Figure 6, the seven curves obtained in the conventional operation mode (N = constant) coincide with those obtained in the new operating mode ($0 < N < 40\,000$ rpm). Because the latex particles are chemically homogeneous, we can convert these 14 $I(t)$ curves into 14 particle size distribution curves. The result is shown in Figure 7, where m_i is plotted against D_i. These 14 particle size distribution curves conform within 5% not only in terms of the known diameters D_i (176, 312 and 794 nm) but also in terms of the given mass portions m_i (40, 50 and 10%), thus proving that our design (seven cells measured simultaneously together with a variable N) does work.

5 WATER/DEUTERIUM OXIDE SEDIMENTATION ANALYSIS

However, the method discussed in the previous section only really works if the dispersion is chemically homogeneous (with the density and refractive index of the particles being uniform). We have therefore developed a

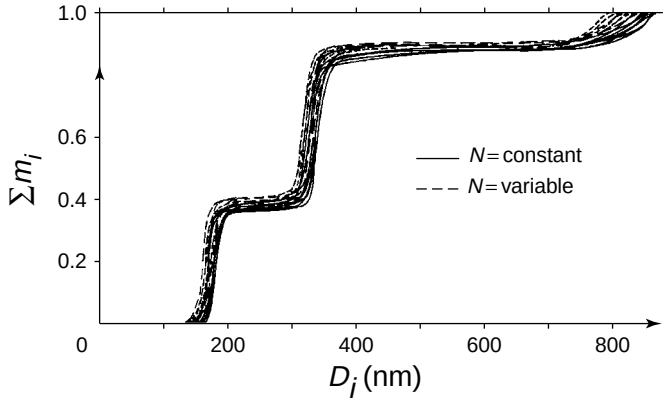

Figure 7 AUC particle size distribution curves of seven mixtures of three monodisperse polystyrene calibration latices, 176 : 312 : 794 nm (40 : 50 : 10% w/w). (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

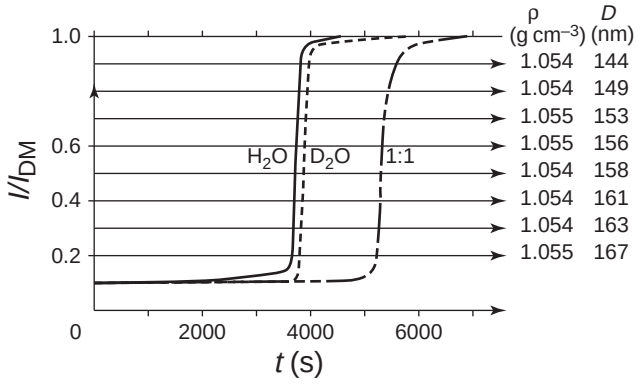

Figure 8 H_2O/D_2O sedimentation analysis of a polystyrene latex ($D = 155$ nm). (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

fast H_2O/D_2O sedimentation technique to analyze non-homogeneous dispersions. As an example, Figure 8 shows the results of the H_2O/D_2O analysis of a polystyrene latex in three dispersants with different densities. The unknown highly concentrated dispersion was diluted at a ratio of approximately 1 : 1000 with the three dispersants H_2O ($\rho = 1.00\,g\,cm^{-3}$), D_2O ($\rho = 1.10\,g\,cm^{-3}$), and a 1 : 1 mixture of H_2O and D_2O, and the three corresponding $I(t)$ curves of these diluted dispersions were measured simultaneously. The computer creates horizontal sections at intervals of 10% steps across the three $I(t)$ curves (2% intervals are possible too). It then calculates the particle density ρ and the particle diameter D of the corresponding fraction from the respective three coupled times of travel at each point of intersection with the curve by means of the two Stokes' equations shown in Figure 4. The table on the right-hand-side of Figure 8 shows the density distribution ρ and the particle size distribution D of the unknown dispersion. In this case, ρ is constant ($1.054\,g\,cm^{-3}$). The dispersion is a chemically homogeneous polystyrene latex which is nearly monodisperse, with its diameters varying around 155 nm.

The polymer dispersion analyzed in Figure 9 is markedly bimodal. After as little as 1 h the H_2O/D_2O analysis revealed that the dispersion was a chemically homogeneous latex with a particle density of $0.89\,g\,cm^{-3}$ (because the three coupled $I(t)$ curves can be forced to coincide by linear stretching along the x-axis). It therefore

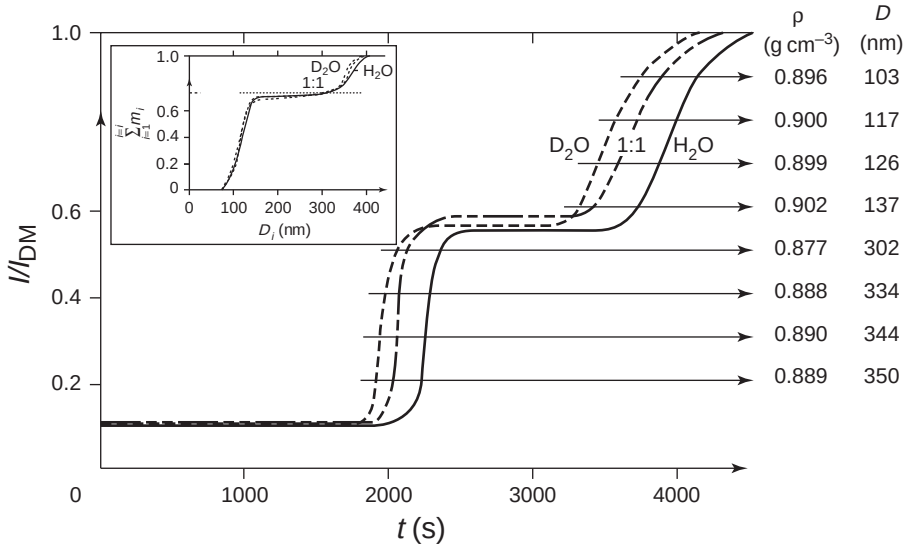

Figure 9 H_2O/D_2O sedimentation analysis of a polybutadiene latex mixture 115 nm : 350 nm (72 : 28% w/w). The insert shows the three resulting particle size distributions of this mixture in the three media $H_2O/1:1/D_2O$. (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

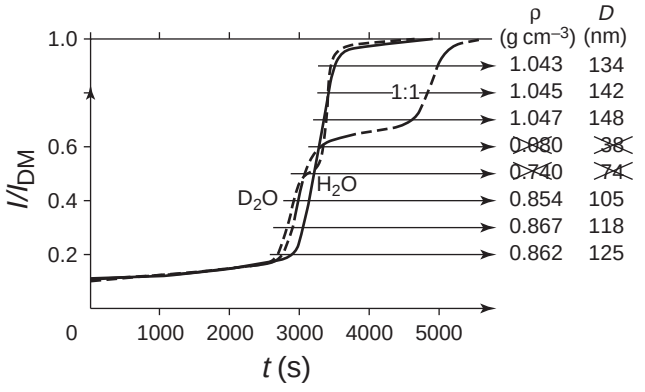

Figure 10 H_2O/D_2O sedimentation analysis of a polystyrene–polybutadiene latex mixture (Polystyrene 155 nm : Polybutadiene 115 nm, 12 : 88% w/w). (Reproduced from Mächtle[8] by permission of Wiley-VCH Verlag.)

had to be polybutadiene. The bimodality is caused by the two particle components having different diameters, 115 and 350 nm respectively. We had prepared this dispersion ourselves by mixing polybutadiene particles of diameters 115 nm and 350 nm in a ratio of 72 : 8% w/w. The insert in Figure 9 shows that the resulting three $H_2O/1:1/D_2O$ particle size distributions of this mixture reproduces very well the given mixing ratio.

Figure 10 shows another example. Here the three $I(t)$ curves of an unknown dispersion can not be forced to coincide by linear stretching along the x-axis, therefore it must be chemically inhomogeneous. Indeed we had prepared a mixture of the nearly matching 155-nm diameter polystyrene particles and the 115-nm diameter polybutadiene particles at a ratio of 12 : 88% w/w. Some of the values for ρ and D are reasonable, but some of them are physically absurd and must be disregarded (they are crossed out here). This indicates that we can confirm within 1 h whether an unknown dispersion is chemically homogeneous or inhomogeneous by means of fast H_2O/D_2O analysis, and we are able to decide whether or not it is possible to determine the particle size distribution by AUC. If H_2O/D_2O analysis states that the sample is chemically inhomogeneous, we then use the powerful AUC density gradient technique[8,13,14] to analyze the inhomogeneity in the particle density in a quantitative, detailed manner.

6 COUPLING PARTICLE SIZE DISTRIBUTION TECHNIQUE

Now we will turn our attention away from monodisperse particles and to extremely broad distributed dispersions, for which we use the coupling particle size distribution technique. We prepared a sample with a broad particle size distribution by mixing ten polystyrene dispersions with a narrow distribution, having $D_{50\%}$ diameters from 67 to 1220 nm, in equal portions of 10% w/w each as shown in Figure 11. Figure 11 also shows the $I(t)$ curve obtained by measuring this mixture of components at the standard concentration of 0.35 g L^{-1} (solid line) which gives an initial light intensity of about 5% of I_{DM}. Each component can clearly be distinguished as a single step with the exception of component 10, the smallest one. The $I(t)$ step for the latter is tiny because of its extremely poor light scattering, i.e. it is badly resolved.

Our coupling technique can solve the problem in cases such as these. We measured the ten-component mixture simultaneously in the same AUC run at a ten-times higher concentration, i.e. at 3.5 g L^{-1}. The additional $I(t)$ curve (broken line) that we obtained shows component 10 at high resolution as well as components 9–7, while component 6 can be partially seen. Components 1–5 cannot be seen because their measuring signal is zero. These two $I(t)$ curves can now be coupled mathematically at the two marked points by combining the lower section of the 0.35 g L^{-1} curve with the upper section of the 3.5 g L^{-1} curve (see Mächtle[9] for details).

If we convert the coupled $I(t)$ curve into a curve for the particle size distribution, we obtain the result shown in Figure 12. The bold line is the integral form and the fine line the differential form of the coupling particle size distribution. The differential form demonstrates the excellent resolution of our particle size distribution equipment especially well. All ten components of this extremely broadly distributed dispersion are separated down to the baseline, including the two components at 318 and 356 nm respectively, even though their diameters are only approximately 12% apart. The ''given'' values in the left table of Figure 12 were used to prepare the

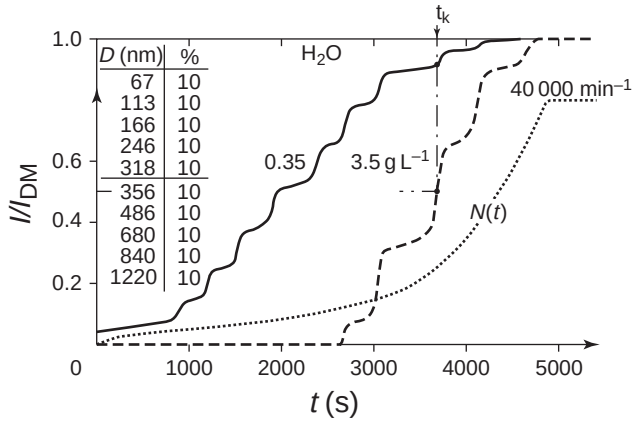

Figure 11 $I(t)$ fractionation curves for a mixture of 10 polystyrene latices with narrow particle size distributions. (Reproduced from Mächtle[10] by permission of Wiley-VCH Verlag.)

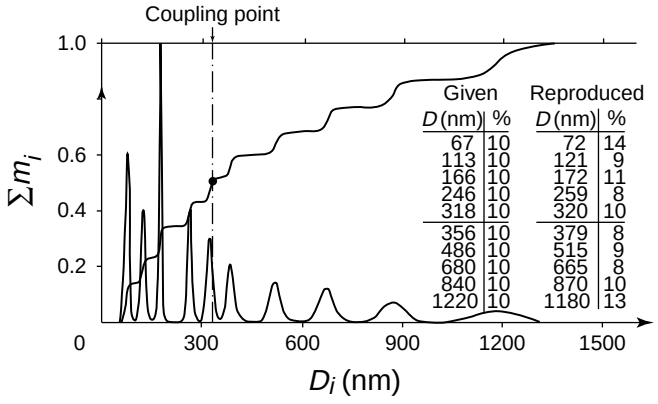

Figure 12 Coupling particle size distribution curve of a mixture of 10 polystyrene latices with narrow particle size distributions. (Reproduced from Mächtle[10] by permission of Wiley-VCH Verlag.)

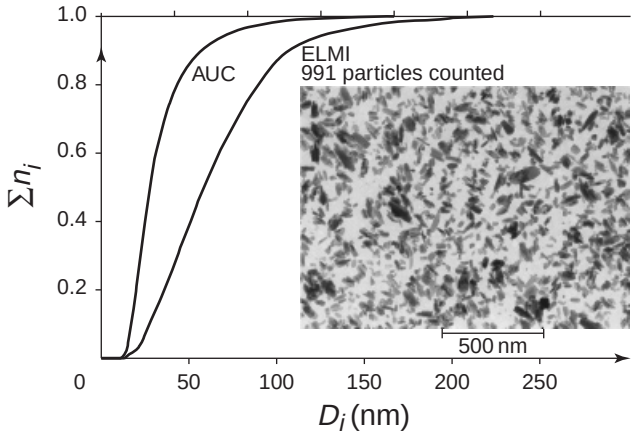

Figure 13 AUC particle size distribution of a copper phthalocyanine pigment.

dispersion, and the "reproduced" values in the right table were obtained empirically by measuring the mixture of ten components. We were able to reproduce the diameters to within 5% and the proportions of the mixture to within 15%. This shows that our coupling particle size distribution technique works for dispersions with an extremely broad distribution. We also tried to measure the particle size distribution of this ten-component latex mixture by sedimentation FFF,[15] but this method was not able to reproduce the correct mixing ratio. This demonstrates the superiority of the AUC method (see also Lange[1]).

7 EXAMPLES OF ANALYTICAL ULTRACENTRIFUGATION PARTICLE SIZE DISTRIBUTION TECHNIQUE

It should be emphasized that we can measure the size distribution of particles contained in pigment preparations, paints and emulsions as well as those contained in polymer dispersions, that is to say any kind of dispersed microparticles having diameters in the range from 10 to 3000 nm. Nonspherical particles such as ellipsoids, cubes, rods or needles can be also measured, but the measured "diameter" is that of a Stokes-equivalent sphere. The dispersant does not necessarily have to be water. Dispersed particles can be measured in any liquid medium, including organic solvents. This can be demonstrated by a few examples.

Figure 13 shows needle-shaped crystals of a copper phthalocyanine dye used in printing inks together with the particle size distribution curve that we obtained by AUC measurement, yielding Stokes-equivalent sphere diameters of 20–150 nm. We also analyzed the electron micrograph of the needle-shaped crystals in Figure 13 with an optical image processing computer program to measure the electron microscope (ELMI) particle size distribution of the needle length, which is also shown in Figure 13. The ELMI particle size distribution is not very accurate, because we counted only 991 particles and it was very difficult to distinguish between neighboring particles. Nevertheless the AUC particle size distribution agreed reasonably well with the ELMI particle size distribution. The ELMI particle size would be expected to be higher, because the needle length (ELMI) is being compared with the equivalent sphere diameter (AUC).

Figure 14 shows another example with nonspherical particles. Instead of being dispersed in water, the heavy ($\rho = 4.2\,\mathrm{g\,cm^{-3}}$) Mn/Zn ferrite nanoparticles were dispersed in the organic solvent triethylene glycol (TEG).

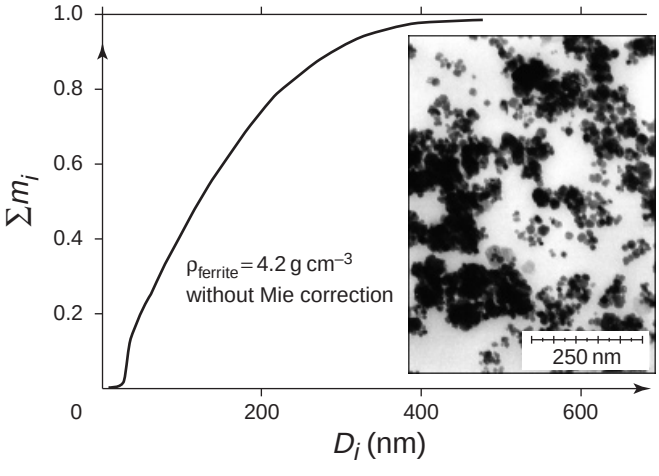

Figure 14 AUC particle size distribution of Mn/Zn ferrite particles in a magnetic fluid, dispersed in triethylene glycol, concentration $1\,\mathrm{g\,L^{-1}}$. (Reproduced from Mächtle[11] by permission of Biophysical Society.)

The dispersion is a magnetic fluid which is switchable in a magnetic field, and it may perhaps be used in the near future in coupling devices, seals and dampers. Our task was to study the agglomeration of the primary 20-nm particles under different conditions. In the special case shown in Figure 14, it was found that the dispersion consisted of 20% w/w of primary 20-nm particles and 80% w/w of highly agglomerated particles with diameters between 40 and 400 nm.

Figure 15 shows an example of historical importance, the original sample of Rinde and Svedberg:[3] pure colloidal gold particles with a very high density of $\rho = 19.3\,\mathrm{g\,cm^{-3}}$ (and a complex refractive index $n(\mathrm{Au}) = 0.706 + i \cdot 2.42$ at $\lambda = 546\,\mathrm{nm}$ and $T = 25\,^{\circ}\mathrm{C}$, which we use for the Mie theory). These gold particles have an extremely small diameter ($D = 5$ to 50 nm), which demonstrates the lower measuring limit for particle diameters of the AUC particle size distribution method with a light scattering/turbidity detector. The gold nanoparticles themselves are embedded in a transparent film of polyvinylpyrrolidone, as the electron micrograph in Figure 15 shows. For the AUC particle size distribution measurement we simply dissolved this film in water. There is a reasonable agreement between the ELMI and AUC results. The film itself has been tested as a fast nonlinear optics switch for optical computers.

Figure 16, our last example, demonstrates that the AUC particle size distribution method works very well with emulsions, including liquid nanodroplets.

We prepared an aqueous 10% w/w emulsion of two monomers by mixing 50% w/w styrene and 50% w/w stearyl acrylate, stirring them with an Ultra Turrax™ mixer and applying ultrasound. We describe this emulsion as a mini-emulsion, because the diameters of the monomer droplets are about 100–500 nm, as the AUC particle size distribution in Figure 16 shows. This mini-emulsion remains stable for 2 weeks, because the hydrophobic stearyl acrylate acts as a cosurfactant. We measured the density of these liquid droplets by H_2O/D_2O analysis, and obtained a result of $\rho = 0.882\,\mathrm{g\,cm^{-3}}$. We polymerized the droplets by increasing the temperature and adding a water-soluble initiator. We then again measured the particle density of the resulting polymer dispersion by H_2O/D_2O analysis, and obtained a result of $\rho = 0.984\,\mathrm{g\,cm^{-3}}$. The AUC particle size distribution is also shown in Figure 16. As expected, the droplet/particle density increased due to polymerization, but the particle size distributions of the monomer droplet emulsion and the polymer latex particles are identical.

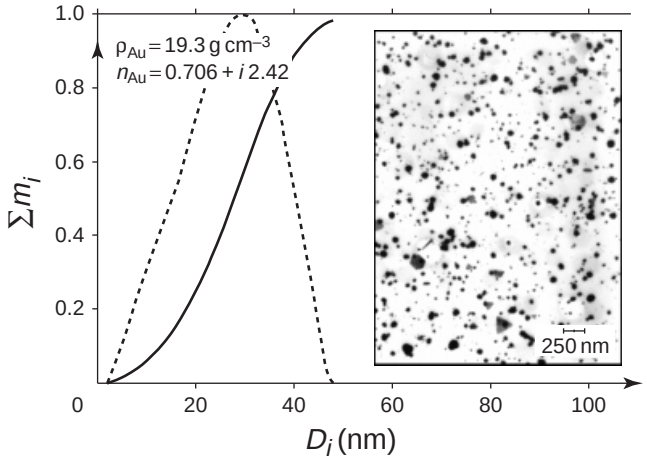

Figure 15 AUC particle size distribution of colloid gold particles dispersed in H_2O at a concentration of $0.1\,\mathrm{g\,L^{-1}}$. (Reproduced from Mächtle[11] by permission of Biophysical Society.)

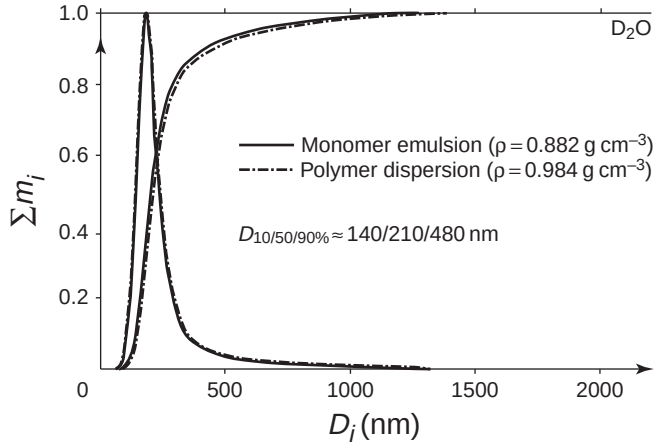

Figure 16 AUC particle size distribution of a 50% w/w styrene : 50% w/w stearyl acrylate monomer mini-emulsion and the resulting polymer dispersion.

8 OUTLOOK – THE NEW TWO-LASER-BEAM TECHNIQUE AND VERY SMALL (1–20 nm) PARTICLES

This section deals with two current topics. Firstly, how is it possible to increase the precision and resolution of the AUC particle size distribution method with the light scattering/turbidity detector for particles with a diameter of 10–5000 nm? Secondly, how can we measure extremely small particles (1–20 nm) which scatter hardly any light and cause hardly any turbidity?

The experimental set-up with one (laser) light beam, described in Figure 2, may offer a solution to the first problem, but it has three disadvantages:

1. It is not always possible to fill the measuring cell up to exactly the same height, because small air bubbles with different sizes are formed during the

filling process. This results in an error Δr_M in the radius of meniscus r_M (see Figure 4) and, according to Stokes' law, the particle diameter D as well. Thus if we could reduce Δr_M we could reduce the error in D too.

2. Because we never know if the particles of an unknown dispersion will form a sediment or float to the surface, we position the laser beam in the middle of the cell between the radius of the meniscus r_M and radius of the base r_B. Thus we always use only the half of the maximum possible running distance $(r_B - r_M)$ of the particles. Increasing the running distance would also increase the range of diameters which can be measured.
3. In H_2O/D_2O analysis, we always need a third cell with the dispersion medium 50% H_2O : 50% D_2O to allow the computer to recognize whether the unknown dispersion particles form a sediment or float to the surface, i.e. whether we have to use r_M or r_B in our calculations, according to Stokes' law in Figure 4. If we did not need to use this third cell, we could run three H_2O/D_2O analyses simultaneously in an eight-cell rotor instead of only two.

To circumvent these three disadvantages of the set-up shown in Figure 2, we have devised a new set-up, shown in Figure 17 by introducing a second parallel measuring beam.

We can change the radial position of the two laser beams, the inner and outer positions r_1 and r_2, continuously between r_M and r_B. It is still not clear what their optimum position is. The preliminary measurements presented in Figure 18 were made at two different positions, at one-third and two-thirds of the way along the distance $r_B - r_M$.

The right side of Figure 17 shows a measurement performed on a monodisperse latex. The two simultaneously measured $I(t)$ fractionation curves are steep one-step curves. Both of them and – very importantly as a third possibility – the *difference* between them deliver the particle size distribution of the measured latex according to Stokes and Mie, if the particle density and the refractive index are homogeneous and known. The computer calculates which of the three particle size distributions is the best one. If the particles form a sediment (and if r_M is precisely known) the outer particle size distribution is the best one. If the particles float to the surface ($\rho_P < \rho_{DM}$) the inner particle size distribution is best. If r_M is not precisely known (because of air bubbles of different size or a small leakage in the cell) the difference between the inner and outer curves gives the best particle size distribution because we do not need to know the value for r_M to be able to calculate it. Instead, we need to know the value for the new "running distance" $(r_2 - r_1)$, which has to be measured once only and is then constant for all cells and every new run. Figure 18 shows an example of a measurement made with this new two-laser-beam set-up.

The insert in Figure 18 shows the two simultaneously measured $I(t)$ curves (and the exponentially increasing rotor-speed curve $N(t)$) for a 1 : 1 : 1 mixture of three monodisperse polystyrene calibration latices with known diameters of 120, 550 and 1300 nm. These two $I(t)$ curves and the difference between them, $I(t)_{inner} - I(t)_{outer}$,

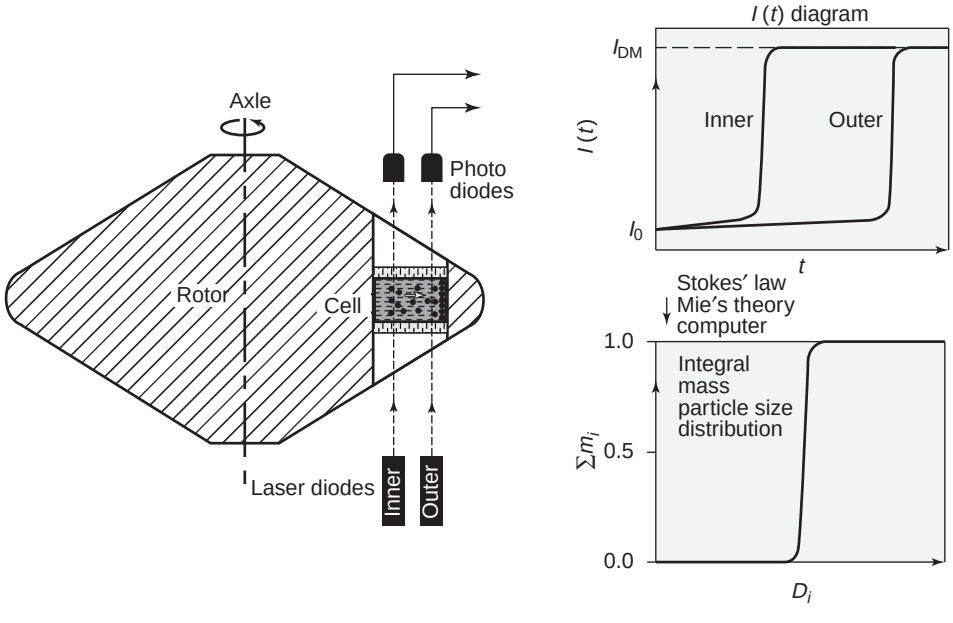

Figure 17 Schematic diagram of a two-laser-beam set-up in an ultracentrifuge used to measure particle size distribution. (Reproduced from Mächtle[11] by permission of Biophysical Society.)

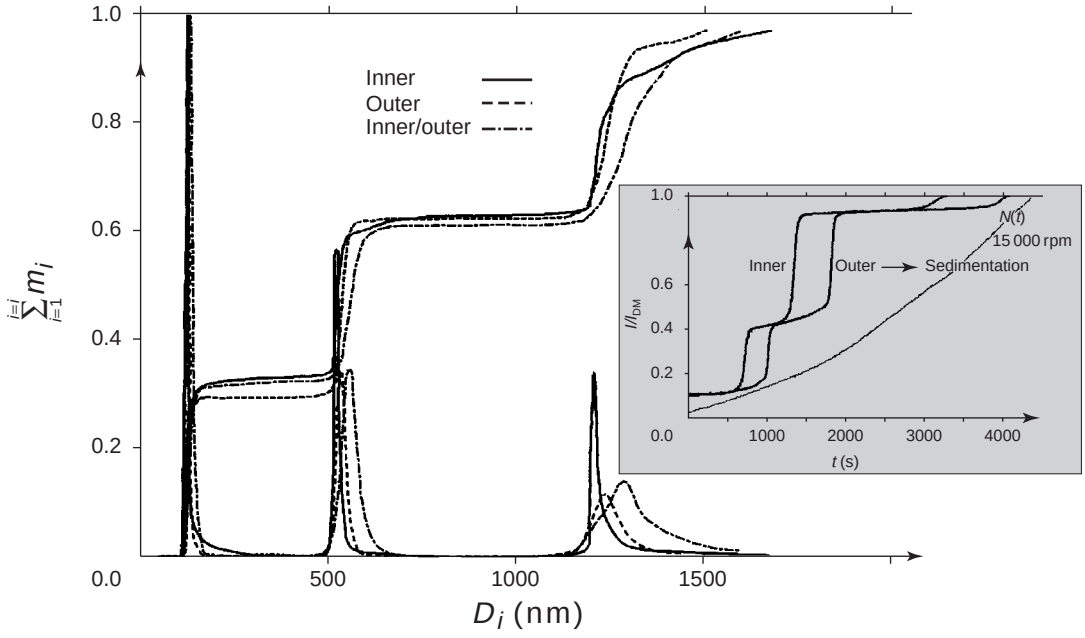

Figure 18 Three AUC particle size distribution curves of a mixture of three monodisperse polystyrene calibration latices (120 : 550 : 1300 nm (1 : 1 : 1 wt. parts), calculated from the two $I(t)$ curves shown in the insert and from the difference between the curves for $I(t)_{\text{inner}} - I(t)_{\text{outer}}$. (Reproduced from Mächtle[11] by permission of Biophysical Society.)

can be transformed according to Stokes and Mie into the three particle size distribution curves shown in Figure 18 (the differential particle size distribution is shown alongside the integral particle size distribution). There is fairly good agreement between these three particle size distribution curves. The mixing ratio and the diameters are reproduced to within ±5%.

There are some small differences between the three particle size distribution curves, but we will not discuss them here because these are the very first measurements that we have made with our new two-laser-beam set-up. These issues will be discussed in a forthcoming paper. However, there is evidence to suggest that the difference between the curves gives the most precise particle size distribution and also the highest resolution. The main lesson to be learnt here is that Figure 18 demonstrates that our new concept actually does work.

It can always be seen from the two $I(t)$ curves (as shown in the insert of Figure 18) whether sedimentation or flotation takes place within the latex. If the running time of the particles in the outer beam t_{outer} is higher than in the inner beam t_{inner}, the latex forms a sediment. If $t_{\text{inner}} > t_{\text{outer}}$ is valid, the latex particles float to the cell meniscus. This behavior can be distinguished very easily by a computer. This is important for the H_2O/D_2O sedimentation analysis, because it can then be performed without the 50 : 50 H_2O/D_2O mixture.

The answer to the second problem is simple: we have to use other detectors for nonturbid, very small (1–20 nm) particles instead of light scattering/turbidity detectors. We need to use refractive index detectors (schlieren or interference optics) or UV-absorption detectors, as in the case of dissolved molecules with high and low molar masses. Mächtle[13] presents an example of a measurement made of 5–25 nm SiO_2 particles, performed with the famous old Model E AUC device from Beckman fitted with schlieren optics. The best example[12] in the literature of this aspect of inorganic colloids, performed with the latest AUC instrument (the OPTIMA-XL-IA from Beckman, which has a 180–900 nm UV scanner and interference optics), is presented in Figure 19. It shows the AUC particle size distribution of extremely small (0.4–2.0 nm) particles of colloidal platinum (aqueous sol, $\rho_{\text{Pt}} = 24.45\,\text{g cm}^{-3}$) in angstrom resolution performed with the UV scanner at $\lambda = 380$ nm, at 25 °C and maximum constant rotor speed of 60 000 rpm. The resolution is nearly atomic. Figure 19(a) shows different radial UV scans at intervals of 2 min. The bold line is transformed by Stokes' law into the particle size distribution in Figure 19(b). Different scans yield nearly the same particle size distribution (if they are not too early or too late), indicating that diffusion effects do not have a disruptive effect. We believe that inorganic colloidal (1–20 nm) particles of this type will play an important role in the near future in metal sols, semiconductors and nonlinear optics. AUC makes it possible to study their growth mechanisms.[12]

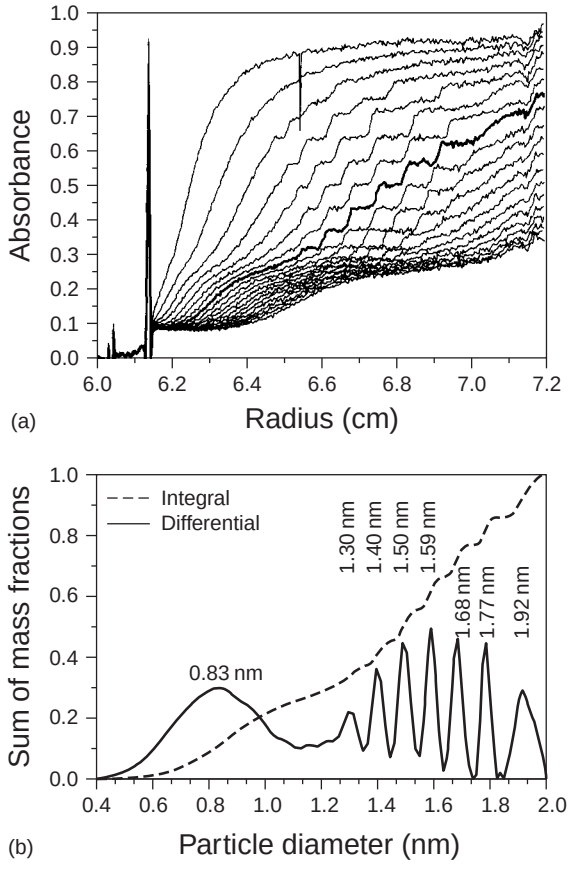

Figure 19 AUC particle size distribution of colloidal platinum. (a) Radial UV-scanner (380 nm) sedimentation profiles at different times of centrifugation (absorbance = $\log I_0/I$; here I is light intensity). (b) Integral and differential particle size distribution calculated from the bold scan shown above. (Reproduced from H. Cölfen, T. Pauck, 'Determination of Particle Size Distributions with Angström Resolution', *Colloid Polym. Sci.*, **275**, 175–180. Figures 1 and 2, 1997 © Springer-Verlag GmbH & Co., KG.)

ABBREVIATIONS AND ACRONYMS

AUC	Analytical Ultracentrifugation
ELMI	Electron Microscope
FFF	Field Flow Fractionation
UV	Ultraviolet

RELATED ARTICLES

Food **(Volume 5)**
Particle Size Analysis in Food

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Field-flow Fractionation in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Turbidimetry in Particle Size Analysis

Polymers and Rubbers **(Volume 8)**
Atomic Force Microscopy in Analysis of Polymers

Polymers and Rubbers cont'd **(Volume 9)**
Field Flow Fractionation in Analysis of Polymers and Rubbers

REFERENCES

1. H. Lange, 'Comparative Test of Methods to Determine Particle Size and Particle Size Distribution in the Submicron Range', *Part. Part. Syst. Charact.*, **12**, 148–157 (1995).
2. H. Rinde, 'The Distribution of the Sizes of Particles of Gold Sols', PhD Thesis, Uppsala, Sweden, 1928.
3. T. Svedberg, H. Rinde, 'The Ultra-centrifuge, a New Instrument for the Determination of Size and Distribution of Size of Particle in Microscopic Colloids', *J. Am. Chem. Soc.*, **46**, 2677–2693 (1924).
4. J.B. Nichols, E.O. Kraemer, E.D. Bailey, 'Particle Size and Constitution of Colloidal Ferric Oxide', *J. Phys. Chem.*, **36**, 505–514 (1932).
5. H.-J. Cantow, 'Zur Bestimmung der Teilchengrößenverteilung in der Ultrazentrifuge', *Makromol. Chem.*, **70**, 130–149 (1964).
6. G. Mie, 'Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen', *Ann. Phys.*, **25**, 377–445 (1908).
7. W. Scholtan, H. Lange, 'Bestimmung der Teilchengrößenverteilung von Latices mit der Ultrazentrifuge', *Kolloid Z. Z. Polym.*, **250**, 782–796 (1972).
8. W. Mächtle, 'Charakterisierung von Dispersionen durch gekoppelte H_2O/D_2O-Ultrazentrifugenmessungen', *Makromol. Chem.*, **185**, 1025–1039 (1984).
9. W. Mächtle, 'Coupling Particle Size Distribution Technique', *Angew. Makromol. Chem.*, **162**, 35–52 (1988).
10. W. Mächtle, 'Determination of Highly Resolved Particle Size Distributions in the Submicron Range by Ultracentrifugation', *Makromol. Chem., Macromol. Symp.*, **61**, 131–142 (1992).
11. W. Mächtle, 'High-resolution, Submicron Particle Size Distribution Analysis Using Gravitational-sweep Sedimentation', *Biophys. J.*, **76**, 1080–1091 (1999).
12. H. Cölfen, T. Pauck, 'Determination of Particle Size Distributions with Angström Resolution', *Colloid Polym. Sci.*, **275**, 175–180 (1997).
13. W. Mächtle, 'Schnelle dynamische H_2O/PERCOLL-Dichtegradienten für Mikropartikel in der analytischen Ultrazentrifuge', *Colloid Polym. Sci.*, **262**, 270–282 (1984).

14. W. Mächtle, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds. S.E. Harding et al., Royal Society of Chemistry, Cambridge, Chapter 10, 147–175, 1992.

15. J. Li, K.D. Caldwell, W. Mächtle, 'Particle Characterization in Centrifugal Fields – Comparison between Ultracentrifugation and Sedimentation Field-flow Fractionation', *J. Chromatogr.*, **517**, 361–376 (1990).

Diffraction in Particle Size Analysis

Bruno Novales
Institut National de la Recherche Agronomique, Nantes, France

1 Introduction **1**
2 Theory of Light Diffraction **2**
2.1 Mie Theory 2
2.2 Rayleigh Scattering 2
2.3 Fraunhofer Scattering 3
3 Description of a Laser Diffraction Particle Sizer **4**
4 Practical Measurements **4**
4.1 Particle Size Range 4
4.2 Sample Preparation 4
4.3 Sample Amount 5
4.4 Experimental Measurements 5
4.5 Results 5
5 Problems Related to Size Analysis by Light Diffraction **6**
5.1 Submicrometer Measurements 6
5.2 Particle Shape 6
5.3 Multiple Scattering 6
5.4 Calibration 6
5.5 Data Processing Artefacts 7
5.6 Intercomparison of Laser Diffraction Instruments 7
6 Other Light-scattering Methods for Particle Size Analysis **7**
6.1 Single Particle Scattering 7
6.2 Photon Correlation Spectroscopy 7
6.3 Laser Doppler Velocimetry 8
7 Conclusion **8**
Abbreviations and Acronyms **8**
Related Articles **8**
References **8**

The diffraction techniques for particle size analysis are based on the measurement of the angular distribution of light diffracted by particles situated in the path of a light beam. These techniques have been used for many years to measure the size of small particulates but the calculation to generate a particle size distribution is so complex that it is only in the last 15 years with the development of small, inexpensive, and powerful computers that these techniques have become so popular. The complete theory describing the diffraction of light by a spherical particle was described by Mie in 1908 and is able to predict the electromagnetic field at any point in space. This theory is adequate for particles with diameters of the same order as the wavelength of the incident light. Some other models have been presented. The most commonly used are based on the Fraunhofer diffraction theory. Light-scattering techniques have obvious advantages for particle size analysis as compared to other methods. They are faster than mechanical techniques, nonintrusive, nondestructive and do not require large amounts of sample. These techniques can be performed on dry powders or on suspended particles in a liquid or in a gas. The operating range of light diffraction techniques is from 20 μm to several millimeters. This article presents the main theoretical aspects of light diffraction by particles. Also, a brief description of a diffraction particle sizer is given with some practical recommendations. Problems related to particle size analysis by light diffraction are examined. Some other light-scattering methods (photon correlation spectroscopy, single particle scattering, Doppler techniques) are also presented.

1 INTRODUCTION

Particle size is an important physical parameter in many complex systems. In many industries, raw materials are processed to be converted into finished products. The processes include grinding, sieving, mixing, and heating operations. Each of these operations is affected by the particle size. So, it is not surprising that many techniques have been developed for particle size analysis. Current methods include techniques based upon sieve analysis, microscopic and image analysis, electrical and optical sensing zones, sedimentation, elutriation, and chromatography.[1–3]

In recent years, techniques involving light-scattering for particle size analysis have become increasingly popular due to laser light sources and the development of condensed but powerful microcomputers which are inexpensive. Techniques involving light–particle interactions have certain advantages for particle size analysis compared to other methods. Also, they are easy to manipulate as many scattering instruments are considered as black boxes from a user point of view.

In fact, light-scattering techniques cover several methods, amongst which light diffraction techniques are the most commonly used. The latter are based on the measurement of the angular distribution of light diffracted by particles in a light beam path. The intensity scattered in any particular direction depends on the size of the particles and their optical properties. If the scattered light

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

is measured as a function of the angle, the resulting scattering pattern contains information regarding the particle size, which can be assessed by appropriate calculations. The extraction of this information is rather mathematically complex and will ultimately result in a particle size distribution. Particle size analysis using diffraction techniques can be performed on dry or suspended particles in a carrier. They are particularly well adapted for particles ranging from 20 to 3500 μm.

Other techniques based on light properties also provide particle size evaluations. Amongst these, we consider single particle scattering,[4] photon correlation spectroscopy,[5] or Doppler techniques.[6] The first refers to the measurement of the intensity of the scattered light from each particle individually at a fixed angle. The amplitude of the scattered light pulse is used to compute the particle size. These instruments are also called single particle optical counters (SPOCs). Photon correlation spectroscopy is designed for assessing the size of particles from 10 to 500 nm. In effect, these instruments measure the fluctuations of scattered light generated by the particles due to their Brownian motion. Since large particles diffuse more slowly than small, fluctuations in scattered light are indicative of the particle size, assuming that some essential restrictions are fulfilled.

The application of these techniques grows steadily making them increasingly popular versus traditional techniques. A few years ago, light-scattering instruments were mostly found in research laboratories, whilst now some are already inserted for on-line production control. In that sense they cover fields such as food production,[7–9] pharmaceuticals,[10] and coatings.[11]

2 THEORY OF LIGHT DIFFRACTION

When a light beam interacts with a particle, it can be absorbed, transmitted, or scattered. Scattering of light concerns the dispersion of the light beam at any angle by the particle. The theory of light-scattering by a particle has been developed by many authors, especially Mie, Rayleigh, and Fraunhofer, and has been reviewed in many books and articles.[12–14]

2.1 Mie Theory

The general theory describing the scattering of light by a spherical particle has been formulated by Mie. It allows the energy of the scattered light to be calculated as a function of the particle size, its refractive index, the wavelength of the incident beam, and the angle of observation. It is derived by solving Maxwell's equations for the incidence of a plane wavefront on a particle and is mathematically complex and does not result in simple analytical solutions. Therefore it is often necessary to make many restrictions and simplifications to become usable. Even in these conditions, proper angular distribution of scattered light is only accessible for quite simple particle shapes, such as spheres which will behave ideally. Consequently, almost all light-scattering techniques assume the particles to be spherical and provide subsequently corresponding equivalent diameters. In order to be able to obtain the scattering pattern of a particle, it is necessary to specify the refractive index of the particle, the refractive index of the suspending medium, the particle size, and the wavelength of the light beam. If for some standard materials the refractive index is an established parameter, problems may arise when biological or food products are considered. Moreover, for very small particles, the angular intensity distribution is dependent on the polarization of the incident light. As a concluding remark, it can be stated that Mie theory is applicable for particles whose size is of a comparable order to the wavelength of the incident light.

For particles shorter than the wavelength of the incident beam, Rayleigh scattering theory is available. For particles larger than the wavelength of the incident beam, Fraunhofer theory can be applied.

2.2 Rayleigh Scattering

If the particles are very small compared to the wavelength of the incident light, they can be considered as opaque points wherein light is not able to penetrate. According to the Rayleigh theory, the scattering pattern is independent of the shape of the particle, as they are considered as points. This leads to a very simple scattering pattern identical to that obtained from the Mie theory for particles of very small size. According to this theory, the scattering intensity I at an angle θ is given by Equation (1):

$$I = I_0 \alpha^2 \left(\frac{2\pi}{\lambda} \right)^4 \frac{1 + \cos^2 \theta}{2r^2} \tag{1}$$

where I_0 is the incident intensity, r is the distance from the particle to the detector, λ is the wavelength of light, and α the particle polarizability which for spheres is given by Equation (2):

$$\alpha = \frac{3(m^2 - 1)}{m^2 + 2} \frac{V}{4} \tag{2}$$

with m the relative refractive index and V the volume of the particle. The scattering intensity is thus proportional to the sixth power of the particle diameter. It follows that small changes in size will greatly affect scattered intensities. However, small particles scatter a very small amount of light. It is also understood that the study of the scattering properties of these small particles is very limited.

2.3 Fraunhofer Scattering

Fraunhofer scattering is a diffraction theory which was developed for large opaque particles. It assumes that light that has penetrated into the particles does not contribute to the diffraction pattern, so the particles are considered as completely opaque. Consequently, this theory does not take into account the optical properties of the particles. It is very useful for the calculation of the size distribution of particles of several micrometers and even higher. Most of the earlier laser diffraction instruments are based on this theory.

The Fraunhofer diffraction theory is sometimes described as the light-scattering caused by the edges of an aperture. Babinet's theorem explains that the diffraction pattern obtained with an aperture of given size and shape is the same with an opaque object (particle) of equivalent size and shape. Two requirements are needed to obtain a diffraction according to the Fraunhofer theory: the size of the aperture (or the particle) must be small in relation to the distance between its position and the light source and the diffraction must be observed at infinite distance. In that case, the diffraction patterns of particles having the same size superimpose perfectly. The first requirement can be satisfied by considering a plane wave incident on the diffraction plane. The second requirement can be satisfied by placing the detector far away from the aperture or by using a lens in order to focus the diffracted light onto a detector at the focal plane of the lens.

For a given spherical particle, the Fraunhofer theory allows the intensity of the scattered light to be defined as shown in Equation (3):

$$I = I_0 \left(\frac{2J_1(x \sin\theta)}{x \sin\theta} \right)^2 \tag{3}$$

where I_0 is the intensity of the incident light, θ is the diffraction angle, J_1 is a Bessel's function of first order, and x is a parameter, often called the size parameter, given by Equation (4):

$$x = \frac{2\pi r}{\lambda} \tag{4}$$

where r is the particle radius and λ the wavelength of the incident light.

The scattering is observed for small angles and in this case it is possible to write Equation (5):

$$I = I_0 \left(\frac{2J_1(x\theta)}{x\theta} \right)^2 \tag{5}$$

The angles for which the diffracted intensity will cancel depend on the particle size and the ring of order k corresponding to beams diffracted at the angle given by

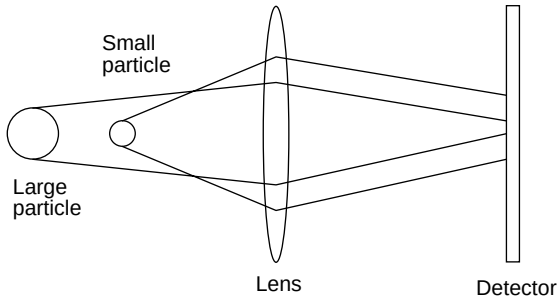

Figure 1 Effect of particle size on the diffraction angle.

Equation (6):

$$\theta_k = \frac{\lambda}{2\pi r} m_k \tag{6}$$

where m_k is the value for which the function J_1 is equal to 0. The radius of a diffraction ring is conversely proportional to the particle diameter. It follows that small particles will scatter light at larger angles, whereas large particles scatter at smaller angles (Figure 1).

The diffraction pattern of the light onto the detector will exhibit intensity variations as a sequence of minima and maxima (Figure 2a) which correspond to a series of concentric rings alternately dark and bright, with a more intense disk at the center (Figure 2b).

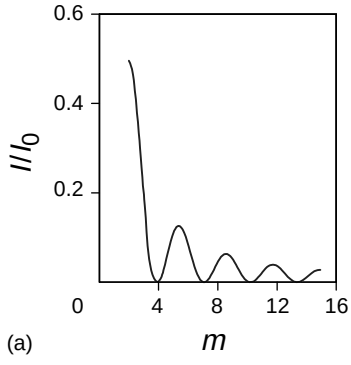

Figure 2 (a) Scattering intensities and (b) corresponding scattering pattern for a circular particle.

The particles are not analyzed individually. The intensity at the detector is the result of the light diffracted by the particles in the analyzed sample whatever their size as defined by Equation (7):

$$I = C \int_0^{\infty} N_{(r)} r^4 \left(\frac{J_1(x\theta)}{x\theta} \right)^2 \mathrm{d}r \qquad (7)$$

where $N_{(r)}$ represents the population of particles of size r and C is a constant. The size classes and the angles for which the diffraction is measured are fixed. The diffraction model is an additive model for which the constants are fixed by the instrument and the unknown parameters are the ratios of the distribution. Mathematical algorithms are required to transfer light intensities into particle sizes. For polydisperse systems the diffraction pattern of each particle size overlaps with those of the same order in the system. As a consequence, the mathematical algorithms and necessary assumptions required for the calculations are very complex. For the instruments available, these algorithms are strictly confidential and in general fit between observed and theoretical intensities which will be obtained by a population of spherical particles.

3 DESCRIPTION OF A LASER DIFFRACTION PARTICLE SIZER

In Figure 3, the principle of a diffraction particle sizer is illustrated. The light source is a helium–neon laser generating light at a wavelength of 633 nm. It provides a better stability and a higher signal-to-noise ratio than lasers with higher wavelength. The beam diameter is about 1 mm, so a beam expander is mounted in order to produce a uniform parallel beam of nearly 10 mm in diameter in order to monitor a useful sample volume. The expanded beam passes through the sample cell and the diffracted and transmitted light is focused by a lens onto the detector in the focal plane of the lens. The latter is often a Fourier lens providing a Fourier transform image of the scattered light. Every parallel light beam emitted from the sample at a given angle θ will be focused in the focal plane at the same azimuth and at a position $X = F(\theta)$ (Figure 4). This very interesting property means that the diffraction will be independent of the position of the particles in the light beam. The undiffracted light is focused at the center of the detector and evacuated through a pinhole. The detector generally consists of photosensitive silicon rings with increasing widths. Since the light scattering at large angles may be several orders of magnitude less than that near the center, the external rings should have a subsequent higher sensitivity or surface. The signal at a specific position on the detector corresponds to the light intensity scattered

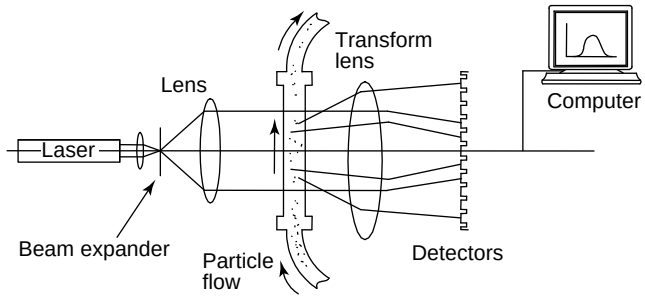

Figure 3 Schematic of a laser diffraction instrument.

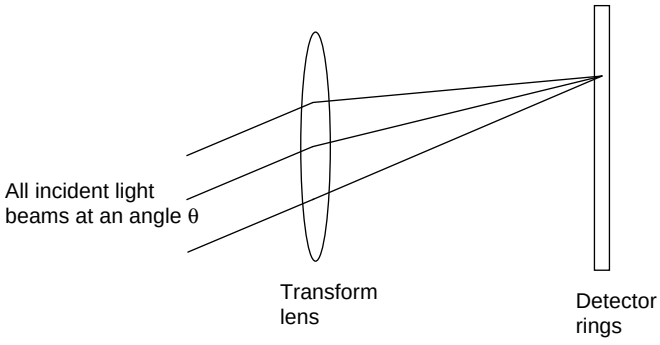

Figure 4 Schematic of Fourier transform lens.

at a definite angle, whatever the position of the particle in the sample cell. The smaller particles detected are estimated at 1 μm. The focal length sets the size of the diffraction field and then the range of measurements. A large focal length enables the resolution at the center of the diffraction rings to be improved, allowing larger particles to be analyzed.

4 PRACTICAL MEASUREMENTS

4.1 Particle Size Range

The lower diffraction limit will correspond to the limit of validity of the geometrical laws of optics. It will be influenced by the relative refractive index of the particles. For water, this value is set at 1.33. When the relative refractive index is closer to 1, this limit can increase. The upper limit of size analysis by diffraction is due to the lack of resolution near the central bright point on the detector.

4.2 Sample Preparation

The samples should have homogeneous physical properties (refractive index, volumetric mass, etc.). For measurements in the liquid phase, the choice of the suspending medium is very important. The dispersion

can be facilitated by ultrasound or with appropriate additives. However, the dispersion method should not modify the sample. The refractive index of the powder must differ from that of the suspending medium. It must be transparent to the incident laser beam and the refractive index must be stable during the experiment. The main cause of variations of the refractive index is temperature changes. These variations lead to a contribution of the small angles corresponding to an overestimation of the larger particle sizes. If large particles are to be measured in a suspending medium, the flow rate should be set to the minimum value that will allow them to be suspended without sedimentation. Sufficient time should be allowed for a perfect mixing of the suspension before a measurement. The most commonly used suspending media are water, methanol, ethylene glycol, and trichloroethylene.

Dry powders can be analyzed by sending the sample through the laser beam when free-flowing. The dry powders are blown through the light beam by compressed air and evacuated by a vacuum cleaner. If the particles are cohesive, the material may not be properly dispersed. The advantage of dry measurements is that the quantity of sample analyzed is not limited. It is then possible to perform a measurement from 100 mg to several kilograms with a continuous flow rate. This may avoid the preparation of a representative sample, which is often a critical step.

Laser diffraction is also particularly well adapted for aerosol sizing since many liquid droplets measure several micrometers. It is necessary to spray the aerosol into the laser beam during the measurement.

4.3 Sample Amount

The sample amount does not play a part in the measurement, except for the representativeness. However, for measurements in a liquid phase, an optimal concentration of the sample in the suspending medium is needed to avoid multiple scattering problems (see below). Most of the commercially available instruments set the concentration in the sample cell by measuring the light absorbed by the sample. The optimal concentration is obtained for an absorbance of 0.1–0.25. If the concentration is too low, too few particles are measured and the statistical errors predominate. Generally, the quantities required for an optimal measurement are about 1–2 g for liquid measurements, and 20–40 g for dry measurements.

4.4 Experimental Measurements

The focal lens is chosen according to the range of particle size studied. An analysis consists first of taking data for the background and then for the sample. The background is subtracted from the sample signal. Both measurements need to be done in the same operational conditions. This is easily done for a measurement of a suspension in a liquid. The background is acquired by injecting the suspending medium alone. Then the particles are added to the sample cell. A protocol is established for each kind of powder which determines the stirrer speed, the pump speed, the air pressure, the use of ultrasound, etc.

4.5 Results

We have seen above that it is possible to calculate the scattering pattern from a sample of known size distribution by using a particular theory (Mie, Rayleigh, Fraunhofer). It is also necessary to solve the inverse problem in order to obtain a particle size distribution from a scattering pattern. From a practical point of view, it is almost impossible to find the particle size distribution from the scattering pattern by applying only simple algebraic operations.

In fact, a first distribution is roughly calculated from the scattering pattern by examining any maximum of intensity in that pattern. Then the scattering model that would be obtained from that distribution is calculated by using the appropriate model. The difference observed between the calculated and the experimental pattern are used to correct the distribution. This procedure is applied until the difference between the two patterns is a minimum.

The transformation of light intensities into a size distribution can be done by assuming an a priori distribution shape (normal or log normal distribution for instance) or by fitting each class of particle size in the

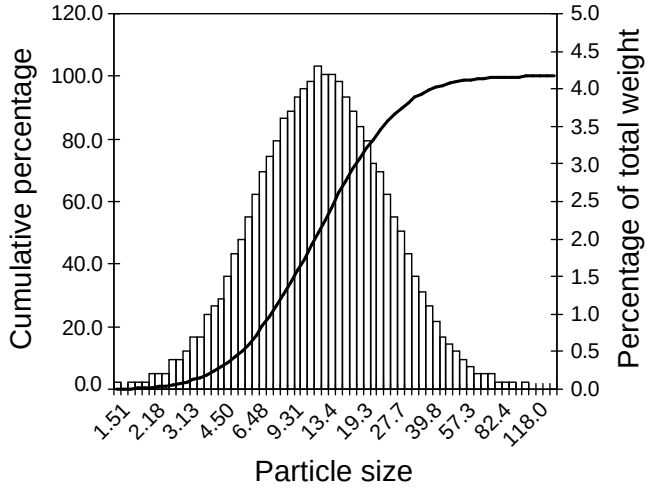

Figure 5 Particle size distribution obtained with a laser light diffraction apparatus. Sixty-four size classes are calculated. The line corresponds to the cumulative percentages. The histogram represents the percentage of particles in each size class.

distribution independently of each other. In many cases, the distribution shape is not known. This is particularly true when the particles are heterogeneous in size and the distributions are multimodal. The size distribution obtained by laser diffraction is expressed in percentage of mass or in volume (by assuming that the density of a particle is independent of its size).

Figure 5 shows an example of particle size distribution obtained with a laser light diffraction instrument. The results can be expressed either as the cumulative percentages of particles or by the percentage in each particle size class.

5 PROBLEMS RELATED TO SIZE ANALYSIS BY LIGHT DIFFRACTION

Even though diffraction techniques have numerous advantages over other techniques, they also suffer from a large number of potential sources of systematic errors, particularly when calculating the particle size distribution. The factors which may cause inaccuracies of particle size distribution obtained by diffraction techniques are the following:

- the Fraunhofer diffraction cannot be applied to small or transparent particles. The Mie theory requires some optical properties of the particles to be known (for example the refractive index) which may not be easily accessible for the user;
- influence of the particle shape on the diffraction pattern, which may lead to an overestimation of the larger particles caused by large-angle diffraction due to the presence of edges on the particles;
- overlapping of the size classes in polydisperse systems leading to misinterpretation of the diffraction pattern;
- unknown computer algorithms for transforming the diffraction pattern into a distribution, which could vary from one manufacturer to another.

5.1 Submicrometer Measurements

As the Fraunhofer theory does not apply to systems composed of particles whose sizes are not significantly larger than the wavelength of the incident light, the examination of submicrometer particles by Fraunhofer methods is questionable. Some authors have shown sizing error greater than 20% in the mean diameter for particles in the 2–7 μm range.[15] This is due to the fact that the scattering of the smallest particles is extremely weak. This is not such a problem if only small particles are present in the system. However, when small particles are mixed with larger ones, whose scattering is many times stronger, the smallest particles may not be detected on the scattering pattern. The use of the Mie theory to solve the problem of submicrometer particles, while a possible improvement, is not always satisfactory as the optical parameters required for the necessary calculations may be unavailable. Particularly, it is often impossible to know the refractive index of the particles in complex systems. To measure the size of very small particles, photon correlation spectroscopy can be used if the particles are in suspension and do not sediment.

5.2 Particle Shape

As indicated previously, the diffraction pattern can only be analyzed for simple shapes. For complex or irregular particles, it should be interpreted with caution.[16,17] Nathier-Dufour et al.[18] have shown that particles having complex shapes (nonspherical) exhibit diffraction patterns containing large-angle scattered beams. If Fraunhofer theory is used to calculate the particle size distribution then the proportion of small particles may be overestimated. If particles of a given size are very irregular, the size distribution may exhibit artefact peaks indicating that more than one particle size is present in the sample. A shape factor may be proposed for nonspherical particles to correct for the effect of the shape on the scattering pattern but this implies that all the particles in a given sample have the same shape.

5.3 Multiple Scattering

The light-scattering models assume that each photon encounters only one particle. When a photon is scattered by a series of particles, its incidence angle will be known only for the first encounter and then the scattering pattern could not be interpreted. This phenomenon is called multiple scattering and is highly undesirable in many cases. It results in larger diffraction angles, since light scattered once is no longer parallel when it encounters another particle. This leads to an overestimation of the small particles. To ensure that no multiple scattering occurs, it is important to keep the sample perfectly dilute.

5.4 Calibration

Laser diffraction instruments do not need calibration in the usual sense. The diffraction pattern can be converted into a size distribution without standard measurements since they are based only on fundamental physical properties. It is however important to regularly check the instrument performance in order to ensure that it gives a relevant particle size distribution. This can be done by measurements of the monodisperse standards distribution covering the range of the instrument. The most generally available standards are latex particles.

5.5 Data Processing Artefacts

Generally, the fitting models are accurate for monodisperse samples in the center of the measuring range. For polydisperse systems, particles having close sizes will lead to an overlapping of their diffraction pattern. For commercially available instruments, the algorithms used for the transformation of the diffraction pattern into a distribution are unknown to the user. The manufacturer's assumptions and their effect on the resultant particle distribution are not generally verifiable. Moreover, the calculations are so complex that approximate theories or iterative methods are used and each manufacturer provides its own algorithm which may lead to differences in the results from different instruments using the same sample. It is then always necessary to compare particle size distribution obtained by laser diffraction with other independent methods.

5.6 Intercomparison of Laser Diffraction Instruments

Some work has been done to compare results of various laser diffraction instruments. For instance, Etzler and Sanderson[19] have shown that different Fraunhofer instruments could produce largely different particle size distributions from the same material. This indicates that the results obtained from any given instrument may not accurately describe the powder and thus may lead to incorrect technical recommendations. These authors suggest the results obtained with one laser instrument should be compared with those of an other independent method to insure against the presence of instrumental artefacts.

6 OTHER LIGHT-SCATTERING METHODS FOR PARTICLE SIZE ANALYSIS

6.1 Single Particle Scattering

The principle of single particle scattering instruments is very simple. The particles are passed in an air jet through a light beam and scatter a light pulse whose duration is proportional to the time taken to pass through the beam and whose intensity depends on particle size. These instruments are called SPOCs. They are well adapted for particles with sizes up to about the wavelength of light. They offer the advantage of providing an absolute concentration in terms of the number of particles per unit volume. However, since each particle is analyzed separately these instruments are slower than Fraunhofer instruments which average the particle sizes over the particles present in the sample.

6.2 Photon Correlation Spectroscopy

Photon correlation spectroscopy, also known as quasi-elastic or dynamic light-scattering,[20] is a method of size measurement for suspensions of spherical particles in the 20–500 nm size range. In a typical photon correlation spectrometer, particles are illuminated by a laser beam. The scattered light is detected at an angle θ with respect to the incident direction. Since the particles are continuously displacing due to Brownian motion, the observed scattered intensity will fluctuate along the time axis. Therefore, the analysis of these fluctuations as a function of time provides information on the mobility of the particles. This mobility is related to the diffusion coefficient and the latter to the particle radius. Since large particles diffuse more slowly than small ones, the fluctuations in scattered light intensity from large particles occur over a longer timescale than fluctuations induced by small particles. The problem for photon correlation spectroscopy is to extract the quantitative information from the fluctuations. This can be done by an autocorrelation analysis. The autocorrelation function $G(\tau)$ of a fluctuating intensity is defined in Equation (8) as the product of the intensity at an instant t and the intensity at an instant $t + \tau$:

$$G(\tau) = \langle I(t)I(t+\tau)\rangle \tag{8}$$

This correlation function depends only on the time difference τ and is independent of the arbitrary time t. The symbol $\langle\cdots\rangle$ indicates that G is an average of $I(t)I(t+\tau)$ for various times t.

For monodisperse small spherical particles, this autocorrelation function is essentially an exponentially decaying function of the time difference τ and can be written as shown in Equation (9):

$$G(\tau) = A + B\exp(-2Dq^2\tau) \tag{9}$$

where A and B can be considered as instrumental factors and D is the diffusion coefficient, with q being given by Equation (10):

$$q = \frac{4\pi n}{\lambda}\sin\left(\frac{\theta}{2}\right. \tag{10}$$

with n the refractive index of the medium, λ the laser wavelength, and θ the scattering angle. The particle size is obtained by relating it to the diffusion coefficient. However, there is no general relation that applies in all situations. With the assumption that the particles are spherical and noninteracting, the particle diameter d is obtained from the Stokes–Einstein equation shown in Equation (11):

$$d = \frac{kT}{3\pi\eta D} \tag{11}$$

where k is the Boltzmann constant, T the absolute temperature, and η the shear viscosity of the solvent.

For a polydisperse system, the size distribution is obtained by deconvolution of the sum of all the single exponentials contributing to the measurement of the autocorrelation function.

6.3 Laser Doppler Velocimetry

Laser Doppler velocimetry is a velocity measurement technique based on the Doppler shifting properties of moving particles, as they interact with light. In laser Doppler velocimeters, the laser beam is divided into two and the focusing lens forces the two beams to intersect, creating interferences (Figure 6). The interferences of the light beams in the measurement volume create a set of equally spaced fringes (light and dark bands) that are parallel to the bisector of the beams. A measurement is made when a particle being carried by the flow passes through these fringes. The light scattered by the particle will have a frequency (the Doppler frequency) proportional to the fringe spaces and to the component of its velocity normal to the fringes.

Two main techniques can be used to relate the Doppler frequency to the particle size. The first method is based on the interpretation of the Doppler signal shape generated by a particle moving through the fringes, which depends of the ratio of the particle size upon the interfringe space. However, this method leads to good results only if the Doppler signal is of good quality. As a consequence, this method is often replaced by the phase Doppler technique.

In phase Doppler velocimeters, multiple detectors are used to collect the light scattered by a single particle. The principle is that if two light sources of the same frequency, but at different angles, are scattered from a single moving particle, the light reaching the detectors will have the same frequency but phase Doppler shifts. This is a result of the fact that when a particle passes through the beams, it does not uniformly illuminate the receiving lens. The diffraction pattern passes over it from one side to the other. As a consequence, the signal will be received in the detectors at slightly different times, i.e. the separate detector signals are out of phase. These phase differences can be related to the particle size if the relative refractive index of the particle is known. For particular positions of the detectors, the relation between the phase shift and the particle size can be considered as linear.[21] The phase Doppler method requires no calibration because the particle size and velocity are dependent only on the laser wavelength and on the optical configuration. Phase Doppler measurements are not based upon the scattered light intensity, and consequently, are not subject to errors from beam attenuation or deflection.

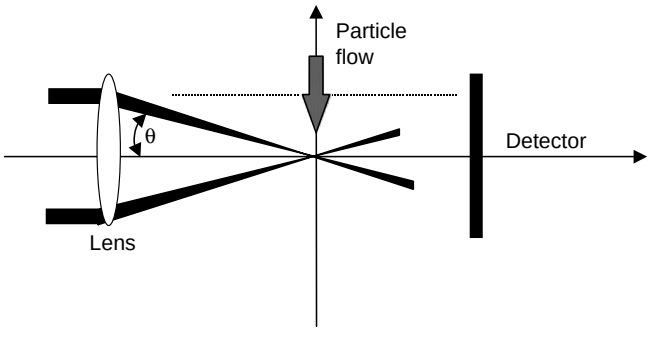

Figure 6 Schematic of a laser Doppler anemometer.

7 CONCLUSION

This article has attempted to show the main advantages of particle analysis by light diffraction in terms of usefulness, rapidity, and performance. This technique can be used for particles ranging from 1 μm to more than 3000 μm. It can be adapted for dry or suspended particles in a liquid phase, for aerosols and emulsion. It is a nonintrusive and nondestructive technique, very easy to use even if the mathematical theories to obtain a particle size from the diffraction pattern are very complex. The results are obtained very rapidly (less than 1 min to obtain a particle size distribution) with good repeatability. With all these advantages, it is not surprising that light diffraction techniques are being increasingly used for particle size analysis and will replace ''traditional'' techniques in many fields in the coming years.

ABBREVIATIONS AND ACRONYMS

SPOC Single Particle Optical Counter

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Light Scattering, Classical: Size and Size Distribution Characterization • Photon Correlation Spectroscopy in Particle Sizing

REFERENCES

1. H.G. Barth, *Modern Methods of Particle Size Analysis*, Wiley & Sons, New York, 1984.
2. H.G. Barth, R.B. Flippen, 'Particle Size Analysis', *Anal. Chem.*, **67**, 257R–272R (1995).

3. T. Allen, *Particle Size Measurements*, 4th edition, Chapman & Hall, New York, 1991.
4. C.F. Hess, C.P. Wood, 'The Pulse Displacement Technique. A Single Particle Counter with a Size Range Larger than 1000:1', *Part. Part. Syst. Charact.*, **11**, 107–113 (1994).
5. N. De Jaeger, H. Demeyere, R. Finsy, R. Sneyers, J. Vanderdeelen, P. Van der Meeren, M. van Laethem, 'Particle Sizing by Photon Correlation Spectroscopy. Part I: Monodisperse Latices: Influence of Scattering Angle and Concentration Dispersed Material', *Part. Part. Syst. Charact.*, **8**, 179–186 (1991).
6. H.E. Albrecht, M. Borys, K. Hubner, 'Generalized Theory for the Simultaneous Measurement of Particle Size and Velocity Using Laser Doppler and Laser Two-focus Methods', *Part. Part. Syst. Charact.*, **10**, 138–145 (1993).
7. M.F. Devaux, F. Le Deschaud de Monredon, D. Guibert, B. Novales, J. Abécassis, 'Particle Size Distribution of Break, Sizing and Middling Flours by Laser Diffraction', *J. Sci. Food Agric.*, **78**, 237–244 (1998).
8. W.L. Lammers, H.J. van der Stege, P. Walstra, 'Determination of the Particle Size Distribution of Non-agglomerated Milk Powders', *Neth. Milk Dairy J.*, **41**, 147–160 (1987).
9. F.W.C. Den Ouden, T. Van Vliet, 'Particle Size Distribution in Tomato Concentrate and Effects on Rheological Properties', *J. Food Sci.*, **62**, 565–567 (1997).
10. C. Washington, *Particle Size Analysis in Pharmaceutics and Other Industries*, Ellis Horwood, New York, 1992.
11. M. McGarvey, D. McGregor, R.B. McKay, 'Particle Size Analysis by Laser Diffraction in Organic Pigment', *Prog. Org. Coatings*, **31**, 223–228 (1997).
12. H.C. Van de Hulst, *Light Scattering by Small Particles*, Wiley & Sons, New York, 1957.
13. C.F. Bohren, D.R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley & Sons, New York, 1983.
14. B.B. Weiner, 'Particle and Droplet Sizing Using Fraunhofer Diffraction', in *Modern Methods for Particle Size Analysis*, Wiley, New York, 135–172, 1984.
15. P.J. Bayvel, A.R. Jones, *Electromagnetic Scattering and its Application*, Applied Science, London, 63, 1981.
16. J.W. Novak, J.R. Thompson, 'Extending the Use of Particle Sizing Instrumentation to Calculate Particle Shape Factors', *Powder Technol.*, **45**, 159–167 (1986).
17. F.M. Barreiros, P.J. Ferreira, M.M. Figueiredo, 'Calculating Shape Factor from Particle Sizing Data', *Part. Part. Syst. Charact.*, **13**, 368–373 (1996).
18. N. Nathier-Dufour, L. Bougeard, M.F. Devaux, D. Bertrand, F. Le Deschault de Monredon, 'Comparison of Sieving and Laser Diffraction for the Particle Size Measurements of Raw Materials in Foodstuff', *Powder Technol.*, **76**, 191–200 (1993).
19. F.M. Etzler, M.S. Sanderson, 'Particle Size Analysis: A Comparative Study of Various Methods', *Part. Part. Syst. Charact.*, **12**, 217–224 (1995).
20. F.R. Hallett, 'Particle Size Analysis by Dynamic Light Scattering', *Food Res. Int.*, **27**, 195–198 (1994).
21. K. Bauckhage, 'The Phase-Doppler-difference-method, a New Laser-Doppler-technique for Simultanuous Size and Velocity Measurements, Part 1: Description of the Method', *Part. Part. Syst. Charact.*, **5**, 16–22 (1988).

Electrozone Sensing in Particle Size Analysis

M. Margarida Figueiredo
Coimbra University, Portugal

1 Introduction **1**
2 Particle Detection Principle **2**
2.1 General Operation 2
2.2 Sizing Response 4
2.3 Correction for Coincident Passages 5
2.4 Precision, Accuracy, Resolution and Sensitivity 6
3 Analytical Considerations **7**
3.1 Sample Preparation 7
3.2 Choice of Electrolyte Solution 7
3.3 Orifice Tube Selection 8
3.4 Data Processing 9
4 Instrument Calibration and Validation **9**
4.1 Latex Calibration 9
4.2 Self-calibration 10
4.3 Instrument and Method Validation 10
5 Special Applications **11**
6 Commercial Equipment **12**
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

Particle sizing is a rapidly growing field, not only because of its importance in a great variety of technological applications but also due to its outstanding relevance to numerous areas of scientific research.

The currently available instruments to measure particle-size distributions are many and varied, the results often being dependent on the measuring technique used. It is thus essential to understand the physical principles behind each technique in order to make an adequate choice.

Electrozone sensing, a widely used method of particle-size analysis, constitutes the main topic of this paper. A general description of its operating principle is provided, followed by some practical considerations mainly concerning sample preparation and the selection of electrolyte solutions and aperture tubes. Being a technique that requires prior calibration, emphasis is placed on the calibration procedures. The section dedicated to the measurement of special materials also reviews some recent improvements of this technique. Finally, a list of commercially available equipment is supplied, together with its most relevant features.

1 INTRODUCTION

Almost every solid material either starts out as a powder, ends up as a powder, or goes through at least one powder stage during manufacture. It is thus not surprising that powder technology plays such an important role throughout process industries, including areas as diverse as chemicals, foods, agricultural products, mineral processing, ceramics, cements, advanced materials, defense, and the environment. In the chemical industry, for example, 60% of the products are manufactured in particle form and a further 20% use powders to impart specific properties to the final products.[1] Moreover, it has been widely recognized that the behavior and properties of the end products are highly dependent upon the characteristics of the particles from which they are formed, namely particle morphology (shape and texture), size, and size distribution.[1,2]

With regard to particle sizing, it should be kept in mind that, for irregular particles, there is not a unique size and that the equivalent-sphere approach is method dependent[3] – in effect, differences between sets of data obtained for irregular particles by different measuring techniques constitute relevant information on particle shape, rather than representing an error.[4] Therefore, a proper knowledge of the method of measurement, and in particular of the equivalent spherical diameter measured, is essential for the adequate use of the data.

In principle, any aspect of particle behavior which is dependent on particle size can be used to measure that size. This explains, at least partly, the great variety of instruments commercially available which, despite their differences in configuration and degrees of sophistication and automation, are based on only a few physical principles.[3]

Electrozone sensing (also named the electrical sensing zone method) is the most widely used of the so-called *stream scanning methods* where particles are measured one at a time. In these methods, particle size is determined by the interaction between the particles and an external field which can be an electrical field (electrozone sensing) or a light beam (optical sensing). Electrozone sensing was originally developed by Wallace Coulter (and so is frequently known as the 'Coulter principle') in the late 1940s for counting blood cells rapidly, but was only patented in 1953[5] and described in detail much later.[6,7] It has become very popular not only for counting and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

sizing blood and other cells and organisms, but also for measuring the particle-size distribution of a wide range of particulate materials as drugs, pigments, fillers, toners, foodstuffs, cosmetics, explosives, construction materials, minerals, coatings, metals, and polymeric powders.

Electrozone sensing is the only method that measures a particle volumetrically, i.e. in three dimensions, and thus is less sensitive to particle shape, in contrast to optical sensing – indeed it is reasonably independent of this parameter, except for extreme shapes such as flaky particles. Additionally, the analyses can be carried out rapidly and with good reproducibility by semiskilled operators. As a result, this method has been the subject of a large number of studies and applications (biological and nonbiological),[8,9] and is extensively described in textbooks.[10] The technique has been accepted as a standard method by the British Standards Institution,[11] as an AFNOR standard[12] and, more recently, by the International Organization for Standardization (ISO),[13] in addition to being recommended for many special applications.[8]

As with other particle sizers, electrozone sensors have advanced significantly and the modern apparatus exhibits a range of specific features, including the storage, handling, and presentation of data. In particular, the main advantages of this technique are:

- it measures particle volume (one of the most representative expressions of particle size), for which it is often used as a reference method;
- the instrument response is not influenced by particle composition, density and refractive index or light interaction effects (inherent to many light blockage/scatter and diffraction systems);
- it produces accurate counts and high-resolution particle-size distributions;
- it is able to measure absolute concentrations (number or volume of particles per total volume of suspension).

The basic limitations are:

- the need to suspend the particles in an electrically conductive fluid;
- the necessity of prior calibration which, depending upon the degree of accuracy, can be laborious and, consequently, time-consuming;
- the limited size range (about 30 : 1 by diameter) covered by a single aperture, which requires the use of more than one aperture for very broad size distributions.

In summary, electrozone sensing presents a good combination of high resolution, size range, sensitivity, and mildness of secondary effects, and has been extensively used to measure particles from about 0.5 μm to approximately 1 mm, the upper limit depending mainly on particle density.

2 PARTICLE DETECTION PRINCIPLE

2.1 General Operation

In the electrozone measurement principle, particles homogeneously suspended at a low concentration in an electrolyte solution are forced to flow, substantially singly, through a small aperture (or orifice) in a nonconductive material that separates two electrodes of opposite potential (Figure 1). A current passing through the aperture between the electrodes enables the particles to be sensed by the changes in electrical impedance as they traverse the aperture.

The passage of a number of particles produces a train of pulses whose amplitudes are essentially proportional to the particle volumes. These pulses are then amplified, sorted, and counted. After instrument calibration, a distribution of the number of particles against volume, or equivalent volume diameter (usually denoted by d_v), can be obtained. This distribution is in turn easily converted

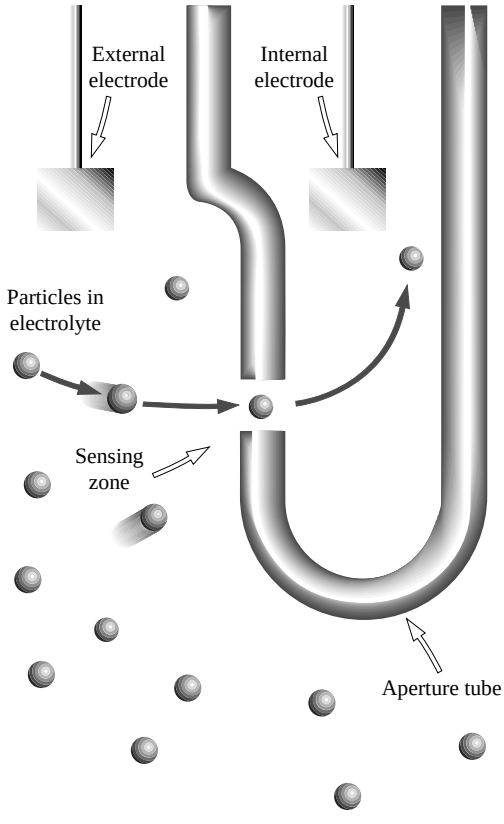

Figure 1 Electrozone sensing operating principle. (Reproduced by permission of Beckman Coulter.)

to a volume-based distribution of d_v. As the volume of a particle is directly related to its mass through the density, the latter distribution is identical to a mass distribution for particles of uniform density.

The sample to be analyzed is dispersed in an electrolyte solution which is placed in a beaker and stirred. An orifice tube (a closed-end glass tube containing a precision-drilled hole) is immersed in the suspension and the flow through the orifice, controlled by a vacuum source, unbalances a mercury manometer (Figure 2).

Originally, all analyses were performed in the *manometer* (or *siphon*) *mode*. In this mode, the system is isolated from the vacuum unit but the flow continues as the mercury returns to equilibrium. The advancing column of mercury activates start and stop contacts, placed so that the count is carried out in a precise and reproducible volume of suspension (usually 0.05, 0.5 or 2.0 mL). This mode is therefore most useful for measuring particle concentration values (number or volume per total volume). Most modern instrumentation also operate in the *manual mode*, where data are acquired by manually initiating and stopping the counting, or in the *time* or *count mode*, in which the counting ends when the time period or the count reaches a preselected value.

As particles pass sequentially through the aperture, each displaces a volume of electrolyte equal to its own immersed volume producing a change in the resistance between the electrodes. As the current is maintained constant, these changes produce voltage pulses of short duration whose heights are essentially proportional to the volume of the particles. The pulses are amplified and fed to a threshold circuit having an adjustable threshold level. When this level is reached or exceeded by a pulse, the pulse is counted. (In some apparatus, the threshold level can be visualized on a small oscilloscope screen, facilitating the selection of appropriate counting levels.) The number of pulses counted represents the number of particles larger than some determinable value – the threshold value is related to the equivalent particle diameter through a calibration constant. In the simple models (called single-channel models), the usual procedure is to reset the instrument and initiate another count at a different threshold. Other instruments have upper and lower threshold circuits which permit sizing between two predetermined volumes (i.e. relative frequency distribution). Normally, in single and dual models, data are manually converted to the volume (or mass) distribution. More recent models are capable of multichannel analyses, i.e. the results can be simultaneously accumulated at a number of channels (size levels), typically between 16 and 256, in a very short time. The main advantages of multichannel analyses are the saving of operator time by removing repetitive counts and the ability to store the data in an internal or external computer for further processing. However, some authors claim that in some of these models a loss of counts may occur.[10,14] This is because it takes a finite time to process each pulse and, especially for relatively high count

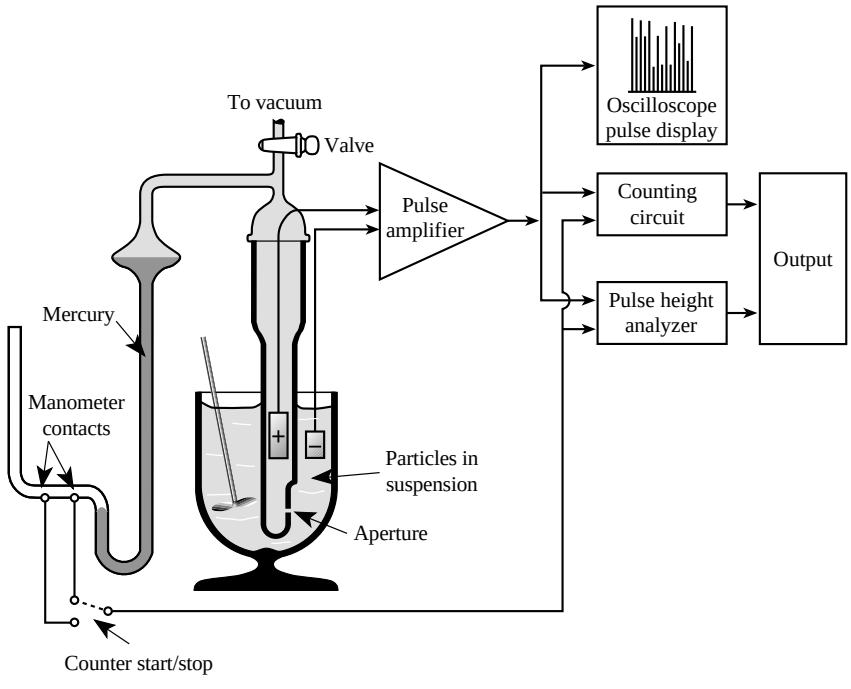

Figure 2 Diagram of the electrozone sensing analyzer.

rates, it is possible that a significant proportion of counts may be lost. In general this will not affect particle-size distributions.

2.2 Sizing Response

Electrozone sensing measures particles by volume, assuming that the height of the pulses generated is directly proportional to the volume of the particles producing them. The reliability of the instrument therefore depends on the accuracy of this assumption. The relationship between the instrument response and particle size has been the subject of several theoretical models,[15–20] as there is no precise and simple solution with which to determine the change in resistance caused by the insertion of a nonconductive sphere in a conductive circular orifice.

This increase in resistance will be determined here using a simplified theory for ease of understanding.[16] Figure 3 shows a particle passing through the orifice and an element (slice) of the particle and orifice.

The resistance of the element without the particle is (Equation 1)

$$\delta R_0 = \frac{\phi_f \delta l}{A} \tag{1}$$

and the resistance of the element including the particle can be seen as two resistors in parallel (Equation 2):

$$\delta R = \frac{1}{A - a/\phi_f \delta l + A/\phi_s \delta l} \tag{2}$$

where A is the cross-sectional area of the orifice, a is the cross-sectional area of the particle, ϕ_f is the resistivity of the fluid, ϕ_s is the resistivity of the particle, and δl is the element thickness. Thus the resistance change due to the presence of the particle is (Equation 3)

$$\delta(\Delta R) = \delta R_0 - \delta R = -\frac{\phi_f a \delta l}{A^2}\left(1 - \frac{\phi_f}{\phi_s}\right)\frac{1}{1 - (1 - \phi_f/\phi_s)a/A} \tag{3}$$

Assuming a constant current I across the aperture, the voltage pulse generated is $I\Delta R$.

Theoretically, the response depends on the resistivity of the particle; however, this is rarely observed due to particle surface phenomena which render the particle resistivity practically infinite.[16] The quotient of the resistivity in Equation (3) is therefore negligible compared to unity, giving (Equation 4)

$$\delta(\Delta R) = \frac{-\phi_f a \delta l / A^2}{1 - a/A} \tag{4}$$

For spherical particles the change in resistance can be derived by integrating Equation (4), which yields (Equation 5)[10]

$$\Delta R = \frac{8\phi_f d^3}{3\pi D^4}\left[1 + \frac{4}{5}\left(\frac{d}{D}\right)^2 + \frac{24}{35}\left(\frac{d}{D}\right)^4 + \cdots\right] \tag{5}$$

where d and D are the particle and the aperture diameters respectively. As can be seen in this equation, the response increasingly deviates from linearity as particle size approaches aperture diameter. Errors due to neglecting diameter powers higher than d^3 have been calculated for various shapes.[15,16]

Other theories have been proposed based on the Maxwell equation for the composite resistivity of a dilute suspension of insulating spheres in a conductive solution.[17–19] This can be written in the general form of Equation (6)

$$\Delta R = \frac{\phi_f v}{A^2} F \tag{6}$$

from which it is evident that the instrument response is proportional to the volume of the particle v, modified by the function F which depends on particle shape and size. Although containing unequal powers of d and D, most formulas proposed for F predict a constant value, i.e. a linear response for a particular aperture if the ratio d/D is much less than unity. However, significant deviations are indicated for large values of this ratio. Nonetheless, more recent experimental studies, carried out with constant-current instruments, showed linear particle responses up to 80% of the aperture diameter, for latex spheres.[21,22]

For grossly irregular shapes it has been recognized that this parameter has a considerable effect, and a self-calibration of the instrument (discussed later) is required which, nevertheless, will hold over the measuring range

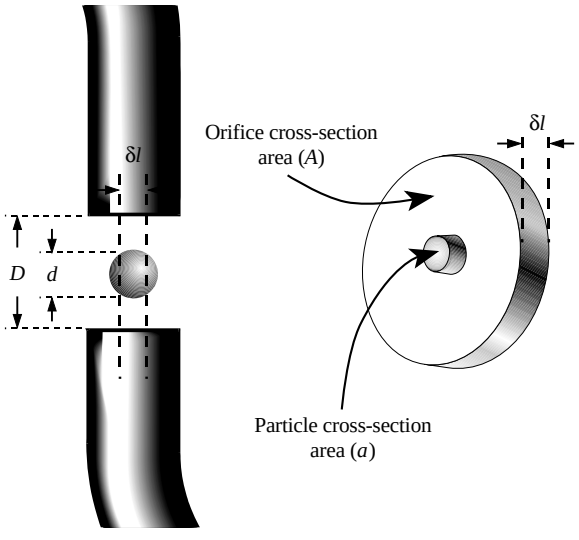

Figure 3 Passage of a particle through the aperture.

of the aperture.[21,23] In any case, in practice, ratios of d/D exceeding 0.6 should be avoided as aperture blockage becomes too troublesome.

In addition to particle size and shape, the instrument response is also influenced by the inhomogeneous electrical field, or current density, in the aperture.[24] In effect, the passage of the particles through regions where the field is not uniform will give rise to relationships between particle volume and resistance change that are more complex than that of Equation (6). A normalized pulse height topography in and near an orifice with a length (i.e. thickness) equal to its diameter is illustrated in Figure 4. As can be seen, pulse height and shape depend on the path of the particle through the orifice – for example, at the orifice inlet and outlet edges characteristic double-peak shapes are obtained, whereas those pulses from the center flow of the aperture possess a bell shape. Longer tubes generate nearly homogeneous fields in the central core of the aperture. Figure 4 also shows that the sensing zone of the orifice is not restricted to the orifice itself.

One way of eliminating these distorted pulses, and which is available in certain models, is to electronically edit them; that is, the pulses are examined and only those that conform to a reference pulse shape parameter (e.g. height to width ratio) are accepted.[25] In some circumstances, this can also lead to a considerable loss of pulses which, in turn, is compensated by an improvement in the distribution resolution (discussed later). An alternative, that avoids the rejection of so much information, is to reshape pulses using mathematical equations.

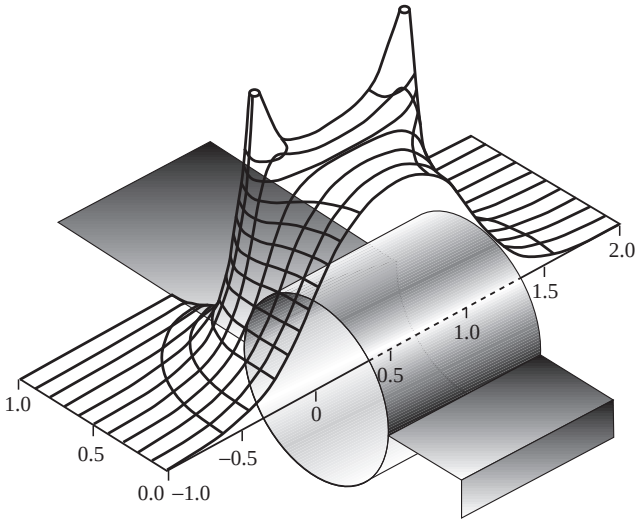

Figure 4 Pulse height distribution in and near an orifice of thickness equal to diameter. The maximum pulse height in the axial path is 96% of that in a homogeneous field. (Reproduced by permission of Wiley-VCH from V. Kachel, *Part. Charact.*, **3**, 45–55 (1986).)

2.3 Correction for Coincident Passages

Ideally, the particles should traverse the aperture singly. However, there are inevitable probabilities of multiple-particle passages, which can result in two types of errors (Figure 5).

- *Primary coincidence* – two (or more) particles simultaneously present in the sensing volume may be counted as only one, depending on their proximity and on the instrument electronic resolution.
- *Secondary coincidence* – two very close particles, which individually originate pulses below a given height level, may collectively be sized as one particle well above that level.

As particle-size distributions should not be a function of concentration, the coincidence effect can be tested by comparing a distribution at one concentration to that at a lower concentration. Primary coincidence can be corrected mathematically, as explained below. Nevertheless the use of extremely dilute suspensions is recommended, as these will reduce both types of coincidence errors.

Several coincidence theories varying in complexity have been used.[26–28] Most are based on the Poisson distribution of probabilities, in which the relationship between the true count N and the observed count n is (Equation 7)

$$n = \frac{V_m}{V_z}\ 1 - \exp\left(-\frac{V_z}{V_m}N\right) \tag{7}$$

where V_z is the sensing zone volume and V_m the metered volume for each count. It should be emphasized that the effective volume of the sensing zone exceeds the physical

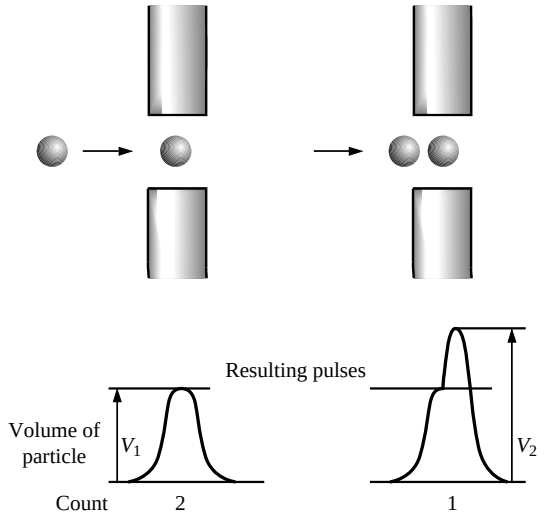

Figure 5 Effect of coincidence on count and size recorded.

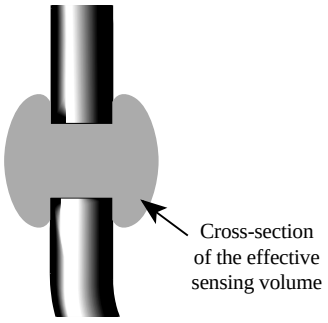

Figure 6 Cross-sectional representation of the aperture, showing the effective sensing volume.

edges of the aperture itself, as indicated in Figure 6 – in many cases V_z is approximately 2.5 times the aperture volume.[16,21]

Equation (7) can be rearranged (Equation 8) as

$$N = -\frac{1}{\alpha} \ln (1 - \alpha n) \tag{8}$$

where (Equation 9)

$$\alpha = \frac{V_z}{V_m} \tag{9}$$

The unknown α can be eliminated by solving simultaneous equations obtained for different dilutions.[28]

Usually simplifications of Equation (7) are utilized, as for example Equation (10)[27]

$$n = N - cN^2 \tag{10}$$

where c is a constant related to the aperture size and sampling volumes, being determined experimentally.

The use of the Poisson distribution model to predict coincident losses has been questioned, especially for higher concentrations, and a more convenient algorithm has been proposed[29] that does not need a quadratic solution, as required by Equation (10).

Published data on coincident correction are sometimes contradictory and frequently depend on the instrument and on the aperture tube. However, providing that the concentrations are kept quite low (below 5% coincidence) almost all coincidence theories yield similar corrections. Most modern instruments make these corrections automatically.

Electronic pulse editing can also remove coincident passage pulses, as these have similar width and shape characteristics to edge-effect pulses.

2.4 Precision, Accuracy, Resolution and Sensitivity

Being a method frequently used as a reference, electrozone sensing should be capable of high precision, both in terms of repeatability and reproducibility, and accurate measurements. Its resolution and sensitivity should also be adequate to detect slight changes in particle-size distributions.

Although there is not much information on this subject in the literature, it is recognized that, for many particle-sizing techniques, repeatability is highly dependent upon sampling and/or dispersion of the powder and on the number of particles. For electrozone sensing measurements, a minimum count of 100 000 particles is recommended for obtaining good repeatability in the volumetric size distribution[13] – e.g. standard deviations less than 1–2 μm are expected for median sizes of moderately wide distributions of particles smaller than 100 μm. As for reproducibility, and despite written standards being available for this method, different operating and/or calibrating procedures may result in significant errors.[30]

As these analyzers are usually calibrated with certified materials (as described later), the accuracy of the analysis is related to the precision of the calibration. This is governed essentially by the quality of the calibrating material and by the constancy of the calibration constant, both in time and over the measuring range – the accuracy of the calibration constant is usually about 0.5–1% for apertures below 100 μm.

Another important aspect of particle sizing is resolution, which is the ability to differentiate between two closely situated peaks or the ability to measure very narrow size distributions. The resolution of the early models was limited by the number of channels available. In modern multichannel analyzers, however, size distribution analyses may be performed using up to 256-channel resolution. In addition, data may be re-expressed as windowed data for even greater sizing resolution. Hence, the resolution now is normally limited by effects such as coincident particle passages and electronic field inhomogeneity. Several methods have been devised for improving the resolution of electrozone sensors, such as electronic pulse editing and the use of longer than conventional apertures. Pulse editing also enables particles to be resolved from baseline noise,[25] thus improving the quality of the measurements – in most situations where about 100 channels are used independently over the measuring range, a typical resolution of 0.1–0.5 μm is anticipated in the measurement range below 100 μm, which is high in comparison to many other techniques.

In the absolute sense, sensitivity can be defined as the number or volume of particles that differs significantly from blank electrolyte. For electrozone sensing the sensitivity is generally very good. Depending on the size of the particles, the sensitivity usually ranges from single particles (at larger sizes) to a few hundred (at small sizes) per size class and per milliliter of liquid sampled.

3 ANALYTICAL CONSIDERATIONS

As previously discussed, the amplitude of the pulses originated by the passage of the particles through the sensing zone is essentially proportional to particle volume, provided that certain conditions are fulfilled:

- particles should be adequately dispersed and homogeneously suspended in the beaker;
- the orifice diameter should be at least twice that of the largest particles to reduce blockage and nonlinear response;
- particle concentration should be low to minimize coincident passages through the orifice;
- particle diameters should give a signal above the noise level (which is normally obtained for $d/D > 0.02$).

The aim of this section is to supply some information regarding operational procedures so that these conditions are accomplished.

3.1 Sample Preparation

The quality of any particle-size analysis result is greatly influenced by the accuracy and reproducibility of the sampling and dispersing techniques. Generally, the quality of the measurement depends on how representative the sample is of the bulk from which it is drawn. In many cases the analyst is presented with laboratory material, the sampling of which he has no control over. Then there is often the need to reduce the laboratory sample to quantities suitable for particle-size analyses, it being therefore indispensable to ensure the homogeneity of the material throughout subdivision. Various devices have been developed for taking samples from the bulk of a powder and for splitting primary samples to smaller ones.[10,31] The spinning riffler has been widely used for this purpose. In this instrument, the powder delivered from a hopper by a small vibratory feeder falls on to a succession of receivers mounted on a rotating table. Sampling an agitated suspension with a syringe or pipette is also a good method for the preparation of very small measurement samples (like those required for electrozone sensing). However, for coarse powders (>20 μm), concentration gradients and particle segregation due to settling may lead to selective sampling. Suspension samplers with multiple capillary tubes have been designed to eliminate these errors. Whichever method is adopted, it is recommended that its efficiency is checked.

The material under test should be well dispersed because the presence of agglomerates generates untrue size distributions and is often connected with poor repeatability of the results. In some cases the test is related to the state of dispersion in an application, so that the quantity of agglomerates is the reason for the measurement. Ideally, the following requirements have to be met:

- the solid particles should exist in the suspension individually (i.e. free of agglomerates);
- the dispersion state should not be modified during the time of the analysis.

If a dry powder is to be used for preparation of a suspension, it should be spatulated with a few drops of a suitable dispersant and slowly diluted with the electrolyte solution. Further dispersion with ultrasonic energy is normally used, except for very fragile or biological materials. It is always important to indicate clearly the specific protocols adopted for both sampling and dispersion.

A microscopic examination of the suspension should be carried out prior to the analysis, to check the degree of dispersion and to estimate the size range of the particles. This is necessary to select the most adequate aperture tube.

The stability of the dispersion during the analysis time should be checked by comparing repeat analyses, using a multichannel counter if possible. This type of instrument is particularly useful to study this phenomenon, owing to the rapidity with which the analyses can be performed. Table 1 shows some symptoms of instability and its main causes.[11,32]

Suitable agitation should also be assured by adjusting the stirrer speed to maximum effect without entraining air bubbles.

3.2 Choice of Electrolyte Solution

Electrolytes for electrozone measurements must have suitable electrical conductivity and physicochemical neutrality towards the particles. A list of the most useful electrolytes as well as dispersants and their applications is usually supplied by the instrument manufacturers and

Table 1 Dispersion instability and possible causes

Change in count[a]		Symptom
Maximum size	Minimum size	
+	+	Crystallization; precipitation
−	−	Dissolution
−	+	Size reduction; deagglomeration
+	−	Agglomeration
−	0	Settling of larger sizes

[a] + denotes count increase, − denotes count decrease, 0 denotes no change.

is also available in the literature.[11,32] For a given material, the electrolyte used must not cause the particles to agglomerate, react, dissolve, disintegrate, shrink, or swell.

Aqueous media are widely employed because the great majority of materials are water insoluble or only slightly soluble. An electrolyte solution of about $10\,g\,L^{-1}$ sodium chloride in water or $50\,g\,L^{-1}$ hydrated trisodium orthophosphate solution is often used because the conductivity is adequate. In some cases, chloride ions can inhibit dispersion and alkaline phosphates present more advantages. In general, the addition of an effective wetting agent or dispersant (directly to the sample, as explained above) greatly helps in dispersing the sample. The optimum choice of dispersant is often a matter of trial and error, with microscope examination being usual – particular attention should be paid to any eventual negative effects (such as dissolution). When the particulate material is sparingly soluble in water, the electrolyte should be first saturated with that material and filtered for usage. The saturated solution should not be left standing and the analysis should be performed immediately.

Alternatively, organic media (such as alcohols, ketones and chlorinated hydrocarbons) are needed for water-soluble materials, including salts, sugars, and foodstuff. Other materials may already be suspended in media that are immiscible with water. Polar organic liquids having dielectric constants greater than 10 with readily ionized materials such as lithium chloride or ammonium thiocyanate are widely used in both cases.

To measure small particles ($\leq 1\,\mu m$) with small apertures, the use of an electrolyte with a very high conductivity is recommended to increase the signal-to-noise ratio – a possible solution would be to use $60–80\,g\,L^{-1}$ sodium chloride in water instead of $10\,g\,L^{-1}$. However, problems may then be encountered in suspending large and/or dense particles that tend to settle in the dispersion beaker and/or in the tube behind the aperture. The addition of a thickening agent to increase the electrolyte viscosity (e.g. glycerol or sucrose) together with the use of a round-bottomed beaker fitted with baffles and special stirrer paddles favor the homogeneity of the suspension.[8]

For highly accurate particle-size analyses, electrolytes must be virtually free of particles. The electrolyte solutions should therefore be carefully filtered before use (0.45 μm or 0.22 μm membrane filters are commonly employed) and the background counting should not exceed the values recommended in the instrument manuals (see section 4.3).

As part of the normal operating procedure, a background count in clear electrolyte is always carried out and recorded, and later subtracted from the sample-plus-background data to obtain the true particle counts.

3.3 Orifice Tube Selection

Most distributions of particulate materials sized by electrical sensing have equivalent diameters in the range between 0.5 μm and 150–250 μm. Some commercial instruments are capable of measuring particles up to 1200 μm and experimental designs have provided results down to 0.1 μm.[18,33] However, broad size ranges can only be covered by using multiaperture analyses, as described later.

The effective size range that a typical orifice can measure is from about 2% to 60% of its orifice diameter, which corresponds to a size span of 30 : 1 by diameter. The lower limit is normally determined by the electrical noise which prevents the pulses generated by the smaller particles from being discernible. This noise may come from the electronic circuitry of the instrument, from surrounding equipment, or from particles in the sample that pass through the sensitive zone of the aperture without entering it. A special case may occur with sedimented particles that accumulate behind the aperture during analysis. They can be diagnosed by performing a background analysis after sample measurement without clearing the aperture tube.

The upper limit is essentially dictated by the aperture blockage which, for nonspherical particles, is mainly governed by their largest size, and by the increasing nonlinearity of the response (discussed above). There is experimental evidence that some models produce accurate measurements up to 70–80% of the aperture sizes.[21]

Determination of the largest and smallest particles present in the sample is usually made by microscope inspection, by trial-and-error procedure with various aperture tubes, or by some previous knowledge about the sample.

The apertures available span from 15 μm to 2000 μm. The use of the large apertures is controlled by the ability to suspend the particles uniformly in the beaker. Analyses with apertures less than 50 μm are more sensitive to noise and require special care with the experimental conditions – the strength and cleanliness of the electrolyte solution is crucial and any atmospheric contamination or electrical interference (internal or external) should be avoided as they can result in false counts.

Particles larger than recommended for a given aperture, or that tend to agglomerate, may cause aperture blockage. An aperture projection system, normally incorporated in the instrument, allows the operator to visualize any blockage or partial blockage. Blockages can also be detected by changes in the pulse pattern displayed on the oscilloscope, when available. In most cases, the blockages are readily cleared by backflushing or by brushing the particles off the orifice with a special brush (bent to a 90°

angle at the metal end and with the hairs cut short). However, certain analyzers have automatic blockage detection and clearance of the sensing orifice.[32]

Although the majority of materials encountered fall within a size range that can be analyzed using only one orifice, samples with a very wide size distribution require the use of two (or more) apertures. For example, over 2% by volume of the particles fall in the smallest size interval, if a bimodal size distribution with one mode essentially below the threshold is expected or if a significant mass-balance deficit is encountered, it is then advisable to use the multiaperture method[11] to obtain the full size distribution. The ratio of the orifice diameters should not exceed 5:1 to ensure that the ranges covered by each tube are allowed to overlap by several micrometers. In general, multitube operations require experienced operators and extra time because larger particles have to be removed by sieving or partial sedimentation, before using smaller aperture tubes. Furthermore, if the particle-size curves (number versus diameter) obtained with the two (or more) apertures do not overlap (due to incorrect calibration or loss of particles during separation), the procedure has to be repeated. Data blending can be performed manually or by software.

3.4 Data Processing

The capabilities of particle sizers have been immensely improved by increases in computer power and the ability to record and handle huge amounts of data in very short times has made virtually all measurements possible. The problem now is that instruments which measure the same particle property, and use essentially the same hardware, may give different results because of disparities in software.[3] This is a matter of concern for particle technologists who should be aware of how software packages manipulate the data. This is not so critical in stream counters, especially in electrozone sensing, because here there is no need for data deconvolution and the relationship between instrument response and particle size is well established.

Most commercially available instruments, based on the electrical sensing zone method, include a dedicated microprocessor or may be linked to a personal computer for data acquisition, storage, and processing. The analytical settings with which the data was obtained can also be saved, allowing easy retrieval for future analyses. Numerous options are now available to present the data in tabular form or graphically (cumulative or differential curves, undersize or oversize), using logarithmic or linear scales. The operator is able to select the presentation form and sequence for a particular report from a vast list of preferences.

Statistical analysis of the size distribution as a function of diameter, volume or surface area, including mean, median, mode, standard deviation, skewness, and kurtosis, is normally accessible. In some cases users can electronically isolate subpopulations of interest and further expand the measurement resolution.

Data saving also enables the plotting of multiple files on the same graph for comparative studies together with the corresponding statistics. Extrapolations to estimate the percentage below the detection limit may also be possible automatically, if neither a second aperture is used to complete the analysis nor a mass balance is performed (as explained later).

4 INSTRUMENT CALIBRATION AND VALIDATION

As the response of the electrical sensing zone method is essentially proportional to the volume of the particles, the pulse height provides a relative scale of particle volumes. Calibration involves converting the arbitrary pulse height into spherical diameter, that is, to determine the calibration constant, according to Equation (11)

$$K_{\mathrm{d}} = \frac{d}{\sqrt[3]{T}} \tag{11}$$

where K_{d} is the calibration constant based on diameter, d is the equivalent volume diameter, and T is the pulse height (threshold value).

Calibration can be performed in two different ways:

- Latex calibration – using narrowly distributed spherical latex particles.
- Self-calibration – utilizing the particles of the material under test.

Self-calibration, also called *mass calibration* or *mass integration*, is recommended in the British Standard[11] as a primary calibration procedure. It is superior to latex calibration because it eliminates any possible errors caused by shape, porosity or conductivity effects of the material to be analyzed. This procedure is therefore more direct and traceable; however, it is not used routinely, so it requires a more elaborate and consequently more time-consuming procedure. In addition, it must be ensured that all particles fall well within the size range of the aperture and no count losses occur in the instrument.

4.1 Latex Calibration

Latex calibration is commonly performed with latex samples which have certified modal, median and/or mean

sizes. Poly(styrene divinyl benzene) latices are particularly recommended for this purpose as these particles are durable and will not change size upon immersion in most electrolyte solutions (special precautions should be taken when using organic electrolytes because the latices may swell or dissolve in some of these liquids). The latex particles should have a mean diameter between 5% and 20% of the aperture diameter, and are obtainable from several sources including manufacturers of electrical sensing zone equipment.

In single-channel models, the calibration is often performed using the so-called half-count method.[34] This consists basically of determining the counts corresponding to threshold levels equal to $0.5T$ and $1.5T$, where T is the average height of the pulses visualized on the oscilloscope. The threshold value corresponding to the count which equals the average of the above counts is the one to be used in Equation (11) to evaluate the calibration constant (K_d), knowing the singlet number median diameter of the latex. More sophisticated particle counters, allowing multichannel analyses, can be more rapidly calibrated by relating the mode of the generated size distribution to the latex number mode.

The calibration constant, which is valid for each aperture tube–electrolyte pair, should be checked periodically and the results recorded to monitor any significant long-term deviation.

4.2 Self-calibration

Although latex calibration is more widely used, the most accurate method is self-calibration.[11,35,36] In this method, a size fraction of the material being examined, preferably not exceeding 10:1 by particle diameter, is diluted to an exactly known mass concentration in electrolyte solution. The volume of the particles in a metered volume of suspension V_p will be given by Equation (12)

$$V_p = \frac{V_m}{V_T}\frac{W}{\rho_s} \tag{12}$$

where V_m is the metered suspension volume, V_T is the total suspension volume, W is the total mass of the powder, and ρ_s is the particle density.

This volume can also be obtained by measuring the particle-size distribution, resulting in Equation (13):

$$V_p = \frac{\pi}{6}\sum \Delta n_i \overline{d}_i^3 \tag{13}$$

where Δn_i is the number of particles in a size interval i and $\overline{d}_i$ is the mean diameter (based on the mean volume) for that size interval.

Substituting Equation (11) in Equation (13) results in Equation (14),

$$V_p = \frac{\pi}{6}K_d^3\sum \Delta n_i \overline{T}_i \tag{14}$$

where $\overline{T}_i$ is the average threshold (corresponding to $\overline{d}_i$).

If all the particles have been counted over the whole measuring range, Equations (13) and (14) can be compared to give Equation (15):

$$K_d = \left(\frac{6}{\pi}\frac{V_m}{V_T}\frac{W}{\rho_s}\frac{1}{\sum \Delta n_i \overline{T}_i}\right)^{1/3} \tag{15}$$

This equation is often presented in the literature[11,34] in the alternative form (Equation 16)

$$K_d = \left(\frac{6}{\pi}\frac{V_m}{V_T}\frac{W}{\rho_s}\frac{1}{\sum \Delta n_i \overline{v}_i}\right)^{1/3} \tag{16}$$

where $\overline{v}_i$ is not always well defined and can be incorrectly identified with a real volume. Real volumes can only be obtained if a prior calibration has been performed. A more suitable form of Equation (16), in terms of volume values, has been proposed (Equation 17) as[37]

$$K_d = K_{d_0}\frac{V_m}{V_T}\frac{W}{\rho_s}\frac{1}{\sum \Delta n_i \overline{v}_{0,i}} \tag{17}$$

where K_{d_0} is the initial estimate of the calibration constant (which could be the latex calibration constant) and $\Delta n_i \overline{v}_{0,i}$ is the total volume of particles computed by the instrument when using K_{d_0}.

The difference between the particle-size distribution obtained by using the latex and the self-calibration constants were found to be negligible for moderately aspherical particles.[8,35,38] However, the self-calibration method is more fundamental and should be used to remove any doubt, especially when dealing with highly irregular, porous, or conductive particles, for which appreciable differences have been reported.[23,35,39–41] Self-calibration requires great accuracy in determining the particle concentration in the suspension and a counting efficiency of 100% in order to obtain a precise value for the measured volume. The chosen method of calibration should be specified when results are reported.

4.3 Instrument and Method Validation

The first check on instrument, aperture, and electrolyte occurs in the blank (background) measurement. Background counts should be lower than the values given in Table 2.

Table 2 Maximum acceptable background counts

Orifice diameter (μm)	Manometer volume (cm^3)	Maximum background count
280	2.0	75
200	2.0	200
140	0.5	150
100	0.5	400
70	0.5	1200
70	0.05	120
50	0.05	300
30	0.05	1500

The second check should come from the monitoring exercise of the calibration constant – the value of the calibration constant should lie within the limits for 95% confidence level relative to the predetermined mean values.[13]

Finally, a *mass-balance* procedure[10,11,42] is also recommended. This procedure is similar to that used for self- (or mass) calibration. Basically, it balances the volume of the particles measured by the instrument in a metered volume of suspension (Equations 13 or 14) with that calculated from the particle concentration (Equation 12). The ratio of these two values gives the fraction of the powder that is accounted for in the measurement. If there are no other particles present in the sample, there should be less than 3% deviation from 100%. Alternatively, this method may also be used to quantify the fraction of particles missed in the measurement, for example because their size is below the lower limit of the aperture. It is evident that the measurement results should be as precise and accurate as possible. Potential error sources and procedures to eliminate them have been discussed elsewhere.[42]

5 SPECIAL APPLICATIONS

Electrozone sensing, despite being a long-established and widely used method, is still the subject of study and development. Several approaches have been investigated to improve the resolution of electrozone sensors to measure closely sized particles. Increased interest in narrow and ultranarrow (coefficient of variation (CV) less than 2% for diameter) size-range materials, such as biological cells, polymer latices (used as size calibration references), and man-made monodisperse materials (for use as chromatography column packings), has demonstrated the need to improve the resolution of standard orifices. The signal produced when a particle passes through a conventional aperture depends on its size and other physical characteristics, and also on its path near and through the aperture, mainly due to the aperture electric field inhomogeneity.

As depicted in Figure 7, particles passing close to the edge of the orifice generate pulses somewhat higher and considerably longer (frequently M-shaped) than they should be, thus reducing the accuracy of the method. This does not affect the analysis of powders having a wide size distribution (as for most powders of industrial interest) but causes overestimated distributions of nearly monosized materials, which are typically skewed towards the coarser sizes and often present an additional peak.[24,43,44] The two procedures that have been suggested to improve this situation are electronic pulse editing (already discussed) and *hydrodynamic focusing*, where particles are directed along the central axis of the aperture.[45–47]

Hydrodynamic focusing (Figure 8) has been the most successful because it is straightforward, prevents particles from passing through regions of inhomogeneous electrical field and also eliminates circulation of particles close to the orifice (which normally results in countable pulses).

A system similar to hydrodynamic focusing has also been applied to measuring fibers (which are particles with a very low aspect ratio) using electrozone sensing.[48] Fiber measurement is of paramount importance in the pulp and paper industry, as fiber length and diameter relate directly to paper properties. The Kajaani FS-200,

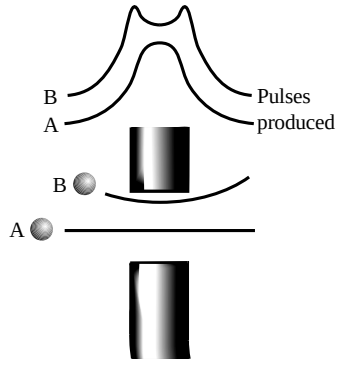

Figure 7 Influence of particle flow path on pulse shape.

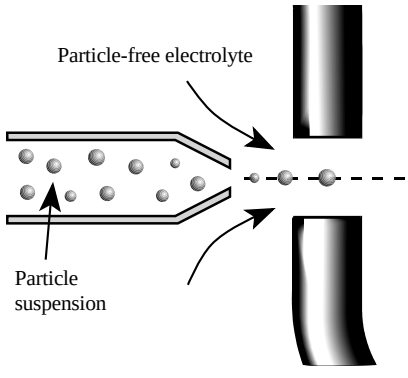

Figure 8 The hydrodynamic focusing principle.

an automated optical analyzer based on the ability of cellulosic fibers to change the direction of polarized light, has been extensively used for sizing pulp fibers.[49] More recently, improvements in computer-based image processing have also led to the development of fiber analyzers that provide rapid and operator-independent measurements.[50] All these automated analyzers comprise a configuration that allows fiber alignment in the measuring zone. As long as fibers flow axially through the aperture, it has been shown that if a prior calibration is performed with fibers of constant length, the instrument response is proportional to the fiber cross-sectional area, from which fiber diameter can be calculated. Fiber lengths can also be evaluated to some degree of accuracy by measuring the pulse duration instead of pulse amplitude.[8]

Other modifications relate to real-time analyses, that are not possible with the traditional off-line procedures and imply either on-line or in-line measurements.[3] The major difference between off-line and in-line measurement is that in the latter sampling and analyzing are closer both in time and space. In addition, on-line measurements reduce the sampling step even more by inserting a probe into the stream, thus allowing localized measurements. Although this increase in automation usually means a loss of accuracy, the benefits are considerable. For example, the data may be passed to a control program enabling process monitoring. It has been reported that electrozone sensors have been adapted for real-time processing.[51,52]

In the pharmaceutical industry electrozone analyzers are used not only to assess the drug particle size but also to study the disintegration of tablets, particle swelling, and dissolution rates[8,53–55] – the size of the drug particle determines its dissolution rate and, ultimately, the rate of absorption.

The limitations of the electrozone method mainly relate to particle characteristics. For example, particle porosity may influence the response of electrical sensing devices.[40,56,57] This is basically because when the pores get filled by the suspending medium they may not be counted as particle volume, resulting in measured volumes considerably smaller than the real particle envelope size. There are also some indications that the wetting properties of the electrolyte solution could also play an important role in the obtained data.[58] A number of researchers have investigated the effect of porosity by filling the pores with an inert material and have even suggested that the pore volume could be estimated from the differences in sizing raw and filled-pore samples.[40,56] However, by not assuming that the specific resistance of the particles (nonconductive as well as conductive) is practically infinite, it has been shown that the apparent resistance of some conductive particles (e.g. carbon, palladium, and nickel) are dependent on the voltage drop over the particle when it passes the orifice.[39,41,59] Critical values have been determined, above which particles may act as a conductor causing large deviations; therefore, adequate conditions must be established beforehand to ensure reliable measurements. Finally, there is evidence that the response of electrozone sensors, besides being a function of particle volume, is also affected by its shape.[8,21,23] Experiments carried out with large-scale versions of electrical sensing instruments indicated that shape has a constant effect which can be eliminated with self-calibration.[23]

6 COMMERCIAL EQUIPMENT

The majority of the electrical sensing zone-based instruments currently in use were manufactured by Coulter Electronics, and by Particle Data under the trade name of Elzone™. Coulter™ and Elzone™ are presently marketed by Beckman Coulter and by Micromeritics respectively. The most widely used models are the Coulter Model TAII, the Coulter ZM and the Elzone™ series 180.[10] Although many of these models are still in use, they are not in current production; the descriptions below are confined only to those models that are commercially available.

The Coulter models in current production are the Z series and the Coulter™ Multisizer IIe. The Z series was designed to be easy to use and consists of the Z1 single threshold, the Z1 dual threshold and the Z2. The Z1 is fundamentally a particle counter, whereas the Z2 is a particle counter and size analyzer that works in the 1–120 μm range. The latter records particle data in 64, 128, and 256 channels, enables the operator to view the data, allows statistical treatment of the results, and permits automatic run and an average of up to 10 replicates. The Coulter™ Multisizer operates in the overall size range 0.4–1200 μm at up to 5000 counts per second and also enables the selection of the number of size classes (channels), from 64 to 256. Built-in cursors permit the isolation of subpopulations, thus increasing the instrument resolution to 25 600 channels. Editing facilities are also available to remove the most distorted pulses which, as discussed before, improves the accuracy of the analysis of narrow-sized range materials. Dedicated personal computer software enables data processing and storage, size distribution statistics, run overlays and averages, multitube overlapping and data export (Figure 9).

At present, Micromeritics markets the Elzone™ 5380 and 5382 particle analyzers, both performing complete counting and sizing capability in the range 0.4–1200 μm. These analyzers are operated via an external personal

Figure 9 The Coulter™ Multisizer II. (Reproduced by permission of Beckman Coulter.)

Figure 10 The Elzone™ 5380. (Reproduced by permission of Micromeritics.)

computer control module and both perform high-speed and multichannel analyses. The main difference between them is that the model 5380 (illustrated in Figure 10) has one analysis station whereas the 5382 has two. In addition, the latter has automated valves which highly reduce the operator intervention. A set of macro commands allows repetitive analyses and the software enables sample information and collected data to be saved which can be presented in tabular and graphic form and plot overlays. These models are capable of autoblockage detection and clearance at the sensing orifice.

ABBREVIATIONS AND ACRONYMS

CV Coefficient of Variation

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction

REFERENCES

1. R. Davies, A. Boxman, B.J. Ennis, 'Conference Summary Paper: Control of Particulate Process III', *Powder Technol.*, **82**, 3–12 (1995).
2. B.J. Ennis, J. Green, R. Davies, 'Key Challenges in Particle Technology – The Legacy of Neglection in the US', *Chem. Eng. Prog.*, **90**, 32–43 (1994).
3. B. Scarlett, '25 Years of Particle Size Conferences', in *Particle Size Analysis*, eds. N.G. Stanley-Wood, R.W. Lines, Royal Society of Chemistry, Cambridge, 1–12, 1992.
4. F. Barreiros, P.J. Ferreira, M.M. Figueiredo, 'Calculating Shape Factors from Particle Sizing Data', *Part. Part. Syst. Charact.*, **13**, 368–373 (1996).
5. W.H. Coulter, 'Means for Counting Particles Suspended in a Fluid', US Patent 2656508, 1953.
6. W.H. Coulter, 'High Speed Automatic Blood Cell Counter and Cell Size Analyzer', *Proc. Natl. Electr. Conf.*, **12**, 1034–1040, 1956.
7. R.H. Berg, *Electronic Size Analysis of Subsieve Particles by Flowing Through a Small Liquid Resistor*, Special Technical Publication 234, ASTM, 1958.
8. R.W. Lines, *The Electrical Sensing Zone Method (the Coulter Principle), in Liquid-and-Surface-Borne Particle Measurement Handbook*, eds. J.Z. Knapp, T.A. Barber, A. Lieberman, Marcell Dekker, New York, 113–154, 1996.
9. 'Coulter Counter and other Coulter Scientific Instruments. Industrial and Scientific Bibliography', Coulter Electronics, Luton, UK, 1994.
10. T. Allen, *Particle Size Measurement*, 5th edition, Chapman & Hall, London, Vol. 1, 1997.
11. BS 3406, 'Determination of Particle Size of Powders. Part 5: Recommendations for Electrical Sensing Zone Method (The Coulter Principle) BSI, London, 1983.
12. AFNOR NF X 11-670, 'Analyse Granulométrique de Particules en Suspension dans un Electrolyte par Compteur à Variation de Résistance', Norme Française, 1989.

13. ISO/13319, 'Determination of Particle Size Distribution – Electrical Sensing Zone Method', International Standards Organization, 1999.
14. H.G. Merkus, H.L. Jansma, B. Scarlett, 'Counting Accuracy in the Coulter Counter Technique', in Proceedings 2nd World Congress on Particle Technology, Kyoto, Japan, 502–510, 1990.
15. B.A. Batch, 'The Application of an Electronic Particle Counter to Size Analysis of Pulverized Coal and Fly Ash', *J. Inst. Fuel*, **37**, 455–461 (1964).
16. T. Allen, 'A Critical Evaluation of the Coulter Counter', in Proceedings of Particle Size Analysis Conference, Loughborough (UK), Society Analytical Chemistry, London, 110–127, 1967.
17. N.B. Grover, J. Naaman, S. Ben-Sasson, F. Doljansk, 'Electrical Sizing of Particles in Suspensions. 1. Theory', *Biophys. J.*, **9**, 1398–1413 (1969).
18. R.W. DeBlois, C.P. Bean, 'Counting and Sizing of Submicron Particles by the Resistivity Pulse Technique', *Rev. Sci. Instr.*, **41**(7), 909–916 (1970).
19. N. Saranummit, 'On the Elimination of Some Factors Decreasing the Accuracy of a Resistivity Transducer When Sizing Blood Cells', Report No. 10, Technical Research Centre of Finland, 1975.
20. B. Scarlett, 'Theoretical Deviation of the Response of a Coulter Counter', 2nd European Symposium on Particle Characterization, PARTEC, Nürnberg, 681–693, 1979.
21. J.G. Harfield, P. Knight, 'Experimental Evidence of the Linear Response of the Coulter Counter', in *Particle Size Analysis 1981*, eds. N.G. Stanley-Wood, T. Allen, Wiley Heyden, Chichester, 151–163 (1982).
22. J.G. Harfield, R.T. Wharton, R.W. Lines, 'Response of the Coulter Counter® Model ZM to Spheres', *Part. Charact.*, **1**, 32–36 (1984).
23. P.J. Lloyd, 'Response of the Electrical Sensing Zone Method to Non-spherical Particles', in *Proc. Particle Size Analysis 1981*, eds. N.G. Stanley-Wood, T. Allen, Wiley Heyden, Chichester, 199–208, 1981.
24. V. Kachel, 'Investigations into Coulter Sizing of Biological Particles; Theoretical Background and Instrumental Improvements', *Part. Charact.*, **3**, 45–55 (1986).
25. J.G. Harfield, R.A. Wharton, 'Increased Resolution of Size Distributions with the Coulter Counter', *Part. Part. Syst. Charact.*, **5**, 29–37 (1988).
26. M. Wales, J.N. Wilson, 'Theory of Coincidence in Coulter Particle Counters', *Rev. Sci. Instr.*, **32**(10), 1132–1136 (1961).
27. L.H. Princen, W.F. Kwolek, 'Coincidence Corrections for Particle Size Determinations with the Coulter Counter', *Rev. Sci. Instr.*, **36**(5), 646–653 (1965).
28. H. Bader, H.R. Gordon, O.B. Brown, 'Theory of Coincidence Counts and Simple Practical Methods of Coincidence Count Correction for Optical and Resistive Pulse Particle Counters', *Rev. Sci. Instr.*, **43**(10), 1407–1412 (1972).
29. J.G. Harfield, P.J. Lloyd, M.P. Cowan, 'Relationship of Coincidence Correction to Coulter Counter® Aperture Dimensions and Instrument Calibration Constant', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, Chichester, 121–133, 1988.
30. H.G. Merkus, O. Bischof, S. Drescher, B. Scarlett, 'Precision and Accuracy in Particle Sizing. Round-robin Results for Sedimentation, Laser Diffraction and Electrical Sensing Zone Using BCR 67 and 69', 6th European Symposium Particle Characterization, PARTEC, Nürnberg, 427–436, 1995.
31. BS 3406, 'Determination of Particle Size Distribution. Part 1: Guide to Powder Sampling, BSI, London, 1986.
32. R.F. Karuhn, R.H. Berg, 'Practical Aspects of Electrozone Size Analysis', in *Particle Characterization in Technology*, ed. J.K. Beddow, CRC Press, Boca Raton, Florida, 157–181, Vol. 1, 1986.
33. R.W. DeBlois, C.P. Bean, R.K.A. Wesley, 'Electrokinetic Measurements with Submicron Particles and Pores by the Resistive Pulse Technique', *J. Coll. Int. Sci.*, **61**(2), 323–335 (1977).
34. 'Instruction Manual for the Coulter Counter Model ZM', Coulter Electronics Ltd, Issue H, 1985.
35. C.M.L. Atkinson, R. Wilson, 'The Mass Integration Method for the Calibration of the Electrical Sensing Zone Technique used for the Sizing and Counting of Fine Particles, in Particle Size Analysis 1981, eds. N.G. Stanley-Wood, T. Allen, Wiley Heyden, Chichester, 185–197, 1982.
36. J.G. Bernard, H.L. Jansma, H.G. Merkus, B. Scarlett, 'Mass Calibration of PFBC Ash Separated by Density with the Coulter Multisizer: Presentation of Cyclone Efficiency in Three Dimensions', in *Particle Size Analysis 1988*, ed. P.J. Lloyd, Wiley, Chichester, 135–149, 1988.
37. M.M. Figueiredo, M.G. Rasteiro, C. Santos, C. Monteiro, 'Mass Calibration of the Coulter Counter, Model ZM', *Part. Part. Syst. Charact.*, **8**, 294–296 (1991).
38. R.K. Eckhoff, 'Experimental Indication of the Volume Proportional Response of the Coulter Counter for Irregularly Shaped Particles', *J. Sci. Instrum.*, **44**, 648–649 (1967).
39. J.G. Harfield, 'Behavior of Conducting Particles in the Sensing Zone of the Coulter Counter', in *Particle Size Analysis 1981*, eds. N.G. Stanley-Wood, T. Allen, Wiley Heyden, Chichester, 165–183, 1982.
40. D. Horák, J. Peška, F. Švec, J. Štamberg, 'The Influence of Porosity of Discrete Particles upon their Apparent Dimensions as Measured by the Coulter Principle', *Powder Technol.*, **31**, 263–267 (1982).
41. G. van der Plaats, H. Herps, 'A Study on the Sizing Process of an Instrument Based on the Electrical Sensing Zone Principle. Part I. The Influence of Particle Material', *Powder Technol.*, **36**, 131–136 (1983).
42. H.G. Merkus, H.L. Jansma, B. Scarlett, M. Figueiredo, 'The Use of a Mass Balance in the Coulter Counter Technique', in *Particle Size Analysis*, eds. N.G. Stanley-Wood,

R.W. Lines, Royal Society of Chemistry, Cambridge, 400–408, 1992.
43. C.M.L. Atkinson, R. Wilson, 'Artefact Peaks in Particle Size Distributions Measured by the Electrical Sensing Zone (Coulter Counter) Method, *Powder Technol.*, **34**, 275–284 (1983).
44. D.A. Elkington, R. Wilson, 'The Effects of Artefact Peaks on the Calibration of Electrical Sensing Zone (Coulter Counter) Instruments with Reference Latices', in *Particle Size Analysis 1985*, ed. P.J. Lloyd, Wiley, Chichester, 509–523, 1987.
45. L. Spielman, S.L. Goren, 'Improving Resolution in Coulter Counting by Hydrodynamic Focusing', *J. Col. Int. Sci.*, **26**, 175–182 (1968).
46. R. Karuhn, R. Davies, B.H. Kaye, M.J. Clinch, 'Studies on the Coulter Counter. Part I. Investigations into the Effect of Orifice Geometry and Flow Direction on the Measurement of Particle Volume', *Powder Technol.*, **11**, 157–171 (1975).
47. H.G. Merkus, H. Liu, B. Scarlett, 'Improved Resolution and Accuracy in Electrical Sensing Zone Particle Counters through Hydrodynamic Focussing', *Part. Part. Syst. Charact.*, **7**, 11–15 (1990).
48. R. Davies, R. Karuhn, J. Graf, J. Stockham, 'Studies on the Coulter Counter. Part III. Applications of the IITRI Sensing System', *Powder Technol.*, **13**, 193–201 (1976).
49. F. Jackson, 'Fiber Length Measurement and its Application to Paper Machine Operation', *Appita*, **41**(3), 212–216 (1988).
50. M.G. Carvalho, P.J. Ferreira, A.A. Martins, M.M. Figueiredo, 'A Comparative Study of Two Automated Techniques for Measuring Fiber Length', *Tappi J.*, **80**(2), 137–142 (1997).
51. M.I. Barnett, E. Sims, R.W. Lines, 'Laboratory Evaluation of the Coulter On-line Particle Monitor', *Powder Technol.*, **14**, 125–130 (1976).
52. V. Degouys, J. Harfield, L. Fabry, F.D. Menozzi, A.O.A. Miller, 'Fast In Situ Determination of the Biomass of Anchorage-Dependent Cells', *Cytotechnology*, **19**, 107–110 (1996).
53. C. Nyström, J. Mazur, M.J. Barnett, M. Glazer, 'Dissolution Rate Measurements of Sparingly Soluble Compounds with the Coulter Counter Model TAII', *J. Pharm. Pharmacol.*, **37**, 217–221 (1985).
54. L.P. Almeida, S. Simões, P. Brito, A. Portugal, M.M. Figueiredo, 'Modeling Dissolution of Sparingly Soluble Multisized Powders', *J. Pharm. Sci.*, **86**, 726–732 (1997).
55. C. Caramella, F. Ferrari, M. Bonferoni, 'Characterization of Particle Swelling of Materials of Pharmaceutical Interest', *Part. Part. Syst. Charact.*, **7**, 131–135 (1990).
56. G. van der Plaats, H. Herps, 'A Study on the Sizing Process of an Instrument Based on the Electrical Sensing Zone Principle. Part 2. The Influence of Particle Porosity', *Powder Technol.*, **38**, 73–76 (1984).
57. A. Michoel, N. De Jaeger, R. Sneyers, W. De Wispelaere, E. Geladé, J. Kern, W. Gorter, P. van Amsterdam, Y.D. Tandt, E. Houtmeyers, L. van Cotthem, 'Influence of Porosity on the Electrical Sensing Zone and Laser Diffraction Sizing of Silicas. A Collaborative Study', *Part. Part. Syst. Charact.*, **11**, 391–397 (1994).
58. N.C. De Jaeger, A. Janes, 'The Sizing of Porous Silica Particles Using an Electrical Sensing Zone Instrument', in *Particulate and Multi-phase Processes, I: General Particulate Phenomena*, eds. T. Ariman, T.G. Veziroglu, Hemisphere, Washington, DC, 169–179, 1987.
59. G. van der Plaats, H. Herps, 'A Study on the Sizing Process of an Instrument Based on the Electrical Sensing Zone Principle. Part III. Size Determination of Activated Carbon', *Powder Technol.*, **39**, 129–132 (1984).

Field-flow Fractionation in Particle Size Analysis

Josef Janča
Université de La Rochelle, France

1 Introduction **1**
2 Principles and Theory **4**
2.1 Polarization Mechanism 5
2.2 Focusing Mechanism 5
2.3 Flow Velocity Profiles 6
2.4 Convolution of Transverse Concentration and Flow Nonuniformities 6
2.5 Retention 7
2.6 Retention Ratio in Polarization Field-flow Fractionation 7
2.7 Retention Ratio in Focusing Field-flow Fractionation 9
2.8 Zone Dispersion 9
2.9 Particle Size Distributions 10
3 Methods, Techniques and their Applications **15**
3.1 Polarization Field-flow Fractionation 15
3.2 Focusing Field-flow Fractionation 19
4 Concluding Remarks **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **23**

Field-flow fractionation (FFF) is an analytical methodology suitable for the separation and characterization of macromolecules within a wide range of molar masses, and for particles in submicrometer and micrometer ranges. The separation is based on the action of effective physical or chemical forces across the separation channel. The field forces interact with the separated species and concentrate them at the appropriate position inside the channel. The resulting concentration gradient induces an opposite diffusion flux. This process leads to a quasi-steady-state distribution of the affected sample components across the channel. The velocity of the longitudinal flow of the carrier liquid also varies across the channel. A parabolic flow velocity profile is usually established inside the channel. The components of the separated sample are transported in the longitudinal direction inside the channel at different velocities, depending on the transverse positions of the zones of the sample components within the flow of the carrier liquid.

The separation is performed in one liquid phase. This is of fundamental importance for the fractionation and particle size analysis of biological samples, which can be sensitive to the type and intensity of the interaction with the surfaces and can denaturate. Steady-state inside the channel is usually reached within a short time due to a small channel thickness. The simplicity of the channel construction permits an accurate description of the separation processes.

The field strength can be controlled over a wide range. Many operational variables in FFF can be manipulated during the experiment by suitable programming.

Accurate analytical results can be obtained from FFF by correct treatment and interpretation of the experimental data. If the relationship between the retention parameters and the characteristics of the separated particles is known, or can be predetermined by using an appropriate calibration procedure, the characteristics of an unknown analyzed sample can be evaluated quantitatively. The particle size distribution (PSD) of the analyzed sample then is determined from a fractogram.

1 INTRODUCTION

FFF is relatively new but important analytical methodology suitable for the separation and characterization of macromolecules over a wide range of molar masses, starting at several thousand grams per mole up to perhaps $10^{18}\,g\,mol^{-1}$, and especially for particles in the submicrometer and micrometer ranges. Very few, if any, of today's separation techniques possess such flexibility. Its development since the late 1960s into a variety of specific methods and techniques represents an important contribution to the science and technology of analytical separations of large molecules and particles.

The separation in FFF is based on the action of effective physical or chemical forces (e.g. temperature gradient, electric, magnetic, gravitational, or centrifugal forces, chemical potential gradient, etc.) across the separation channel. The field forces interact with the separated species and concentrate them at the appropriate position inside the channel. The resulting concentration gradient induces an opposite diffusion flux. This process leads to a quasi-steady-state distribution of the affected sample components across the channel. The velocity of the longitudinal flow of the carrier liquid also varies across the channel. A parabolic flow velocity profile is usually established under conditions of laminar isothermal flow of a Newtonian liquid inside an imaginary channel composed of two parallel infinite planes or, in reality, in a channel of rectangular cross-section. As a result, the components of the separated sample are transported in the longitudinal direction inside the channel at different velocities

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

depending on the transverse positions of the zones of the sample components within the flow of the carrier liquid.

The general principle of FFF is demonstrated in Figure 1. A carrier liquid is pumped through the sample injector to the fractionation channel and leaves the separation system after passing through the detector, which allows the monitoring of the eluting species and the recording of the fractograms.

Two specific mechanisms – *polarization*[1] and *focusing*[2] – lead to the formation of different concentration distributions across the fractionation channel. The components of the fractionated sample are thereby differently compressed onto the accumulating wall of the channel, or focused at different positions, as shown by details (A) and (B) in Figure 1. Despite their principal differences, polarization and focusing FFF have many common characteristics such as the experimental procedures, instrumentation, data treatment, and the range of potential applications. The separation is performed in a single liquid phase. A stationary phase of large surface area (typical for a chromatographic column of comparable separation capacity) is absent in all FFF separation channels, regardless of the particular method or technique applied. This feature is of fundamental importance for the fractionation and particle size analysis of biological samples, which can be sensitive to the type and intensity of the interaction with the surfaces and can denaturate. The inner surface of the FFF channel plays no active role in the separation process and it can be deactivated so that undesirable interactions are minimized. Especially in focusing FFF methods, only a very small part of the separated species comes into direct contact with the channel walls. Steady-state inside the channel is usually reached within a short time due to a small channel thickness. The simplicity of the channel construction permits an accurate description of the separation processes.

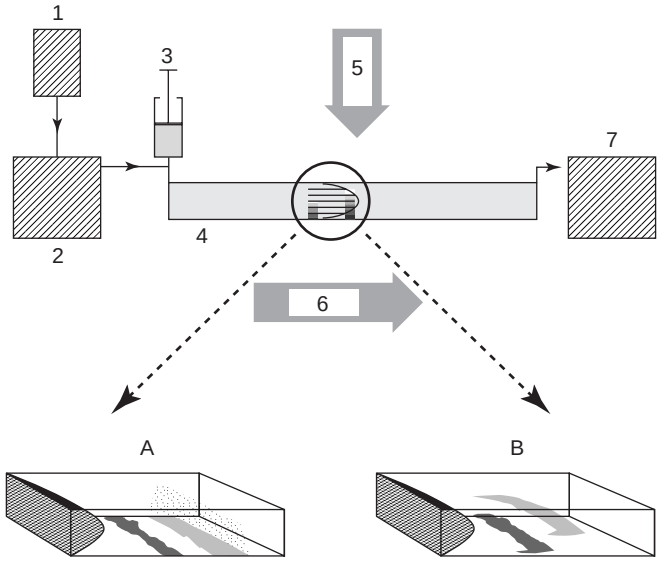

Figure 1 The general principle and experimental arrangement of FFF: (1) carrier liquid reservoir, (2) pump, (3) injector, (4) separation channel, (5) external field, (6) hydrodynamic flow, (7) detector. The three-dimensional details of the channel section show (A) the polarization FFF mechanism and (B) the focusing FFF mechanism.

The strength of the field can be controlled over a wide range in order to manipulate the retention. Consequently, in polarization FFF methods, the complete passage of a particular sample component through the channel represents the lower retention limit, whereas the retention is total when the given component is completely compressed against the channel wall. Retention ratio, as defined by Equation (26), from zero to approximately 0.02 comes into account in practice for polarization FFF methods. With focusing FFF methods, the retention is even more versatile. Many operational variables in FFF can be manipulated during the experiment by a suitable programming, although not all have been applied experimentally.

Any displacement of the separated species depends on the existence of the driving forces. The processes which cause the formation of concentration gradients vary with the character of the field. Accordingly, the applied field determines the individual polarization FFF methods. However, the focusing process is always related to the existence of the gradient of the effective forces. Only when the driving forces are dependent on the position across the channel, and vanish at a given position, can focusing occur.

As a result, *polarization FFF* methods are classified with regard to the nature of the main field applied, whereas *focusing FFF* methods are classified by considering the combination of various gradients and fields, thereby emphasizing the focusing processes.

The historical background of this modern separation methodology can be seen, for example, in the recovery of gold by panning, or of other metals and minerals as demonstrated in the middle-ages rendition in Figure 2.[3] The gravitation and the stream of running water were used to separate the gold particles from other particulate material (soil, clay, sand, etc.). The use of this principle dates back to some 5000 years BC.

Some formerly proposed separation methods are also based on the combined action of field forces and hydrodynamic flow. The fractionation of proteins by combined electrophoresis-convection described by Kirkwood[4] in 1941 and extended in 1952 by Kirkwood and Brown[5] to a combination of diffusion-convection, or thermal diffusion combined with free thermal convection applied in Clusius and Dickel[6] thermogravitational columns, can be also considered as precursors of FFF.

Figure 2 Middle-ages device for the recovery of gold particles and minerals from sand, clay, and soil blends by combining sedimentation and quasihorizontal movement of water, accompanied by vigorous manual stirring of the mud cake.

However, the true beginnings of FFF can be attributed to Giddings[1] who described the general concept as well as polarization FFF. The aerostat shown in Figure 3, created by Joseph and Etienne Montgolfier,[7] is a phenomenological predecessor of the generic invention of the focusing FFF.[2] More specific or generalized separation mechanisms were subsequently proposed and exploited such as steric FFF[8] for the polarization FFF mode or generalized isoperichoric focusing FFF.[9,10] The term 'isoperichoric' was introduced by Kolin[11] and is derived from the Greek language where *isos* means 'equal' and *perichoron* means the 'environment'. This term therefore means that the focusing forces are generated by the difference between the local effective property of the dispersing carrier liquid and the corresponding parameter of the focused sample component.

Polarization FFF methods utilize the exponential concentration distribution that forms for each sample component across the channel, with the maximum

Figure 3 Aerostat of Montgolfier's brothers invention in 1783.

concentration at the accumulation wall. This is the consequence of the constant and position-independent velocity of the transverse migration of the affected species due to the field forces. This concentration distribution is combined with the velocity profile formed in the flowing liquid.

Focusing FFF methods make use of the transverse migration of each sample component under the effect of the driving forces that vary across the channel. As a result, the sample components are focused to positions where the intensity of the effective forces is zero and are transported longitudinally with different velocities according to the established flow velocity profile. The concentration distribution within a zone of a focused sample component can be described by Gaussian or similar distribution functions.

Accurate analytical results can be obtained from FFF by applying the correct treatment and interpretation of the experimental data. If the relationship between the retention parameters and the characteristics of the separated particles is known, or can be predetermined by using an appropriate calibration procedure, the characteristics of an unknown analyzed sample can be evaluated quantitatively. The PSD of the analyzed sample then is determined from a fractogram.

2 PRINCIPLES AND THEORY

The FFF methodology is based on the simultaneous action of the effective field forces and of the liquid flow inside a separation channel operating on the fractionated species. The channel thickness is small, usually between 50 and 500 μm; the ratio of the channel width to its thickness is typically between 20 and 200; the usual length is from 10 to 100 cm. The mutual orientation of the field forces and of the carrier liquid flow is perpendicular. The liquid flows in the direction of the channel longitudinal axis and the field forces act across the channel (Figure 1). The sample is injected into the channel as a short pulse or as a continuous stream carried by the flow. The field generates the driving forces which induce a mass flux of the sample components. A concentration gradient of each sample component is formed across the channel. This concentration gradient induces an opposite flux due to the diffusion, Brownian motion, repulsive interactions, etc. As a result, a steady-state concentration distribution of each sample component across the channel is established. Simultaneously, a flow velocity profile is formed across the separation channel. The sample components are transported in the direction of the flow at different velocities, depending on the position of the centers of gravity of their concentration distributions in the flow velocity profile. Consequently, they leave the channel at different elution times and are thus separated. Two distinct cases, A and B in Figure 1, correspond to the two different separation mechanisms, polarization and focusing, resulting in zones either differentially compressed at the accumulation wall or focused at different levels in a free volume of the channel. Although the separation and accompanying processes are carried out in a three-dimensional channel, the following two-dimensional model is a fair approximation of the reality.

The driving forces can be generated by a single field as well as by the coupled action of two or more different fields. Polarizing and focusing forces can operate simultaneously, resulting in a complex mechanism of separation. The field force, F, and consequently the migration velocity, U, are independent of the position in the direction of the x axis in polarization FFF (Equation 1):

$$F \neq 0 \text{ and } U \neq 0 \text{ for } 0 < x < w \tag{1}$$

Here w is the distance between the main channel walls in the direction of the x axis, with $x = 0$ at one wall of the channel. Dispersion processes act against field induced flux until the steady state is reached. The steady-state concentration distributions of the affected sample components across the channel are exponential only if ideal polarization FFF is effective.

However, it holds for the x axis-dependent direction of the field force in focusing FFF (Equations 2–5):

$$F = F(x) \text{ and } U = U(x) \quad \text{within } 0 < x < w \tag{2}$$

$$F(x) = 0 \text{ and } U(x) = 0 \text{ for } x = x_{\max} \quad \text{with } 0 < x_{\max} < w \tag{3}$$

$$F(x) > 0 \text{ and } U(x) > 0 \quad \text{for } x < x_{\max} \tag{4}$$

$$F(x) < 0 \text{ and } U(x) < 0 \quad \text{for } x > x_{\max} \tag{5}$$

Here the coordinate $x_{\max}$ corresponds to the position at which the concentration distribution of a sample component across the channel attains its maximal value. Each sample component is focused in the direction of the x axis around its proper $x_{\max}$ position. The dispersive forces, in this case, act in both opposite directions from the $x_{\max}$ position until the steady state is reached. The resulting shape of the concentration distribution is close to the Gaussian distribution function.

Selective transport of the separated sample components and related dispersion phenomena are the main processes that characterize any separation method. A sample component undergoing the effect of the field forces in FFF is displaced in the moving carrier liquid by the combined action of the flow and field applied. The resulting velocity vector $\mathbf{u}$ can be decomposed into two perpendicular vectors: $\mathbf{U}$, induced by the applied field, and a flow-induced

component **v** (Equation 6).

$$\mathbf{u} = \mathbf{U} + \mathbf{v} \tag{6}$$

The total flux of a sample component **j** can also be written as a sum of transversal flux in the direction of the x axis and of axial flux in the direction of the longitudinal y axis (Equations 7–9):

$$\mathbf{j} = \mathbf{j}_x + \mathbf{j}_y \tag{7}$$

$$\mathbf{j}_x = -D_x \mathbf{r}_x \frac{\partial c}{\partial x} + \mathbf{U}c \tag{8}$$

$$\mathbf{j}_y = -D_y \mathbf{r}_y \frac{\partial c}{\partial y} + \mathbf{v}c \tag{9}$$

where $D_{x,y}$ and $\mathbf{r}_{x,y}$ are the diffusion coefficients and the unit vectors, respectively, and c is the concentration.

The transverse and longitudinal transports can be considered separately, in a first approximation. However, the longitudinal flow of the carrier liquid permanently perturbs the transverse concentration distribution. The deviation from the steady-state can be characterized by the equilibrium departure term ε (Equation 10):

$$c = c^*(1 + \varepsilon) \tag{10}$$

where c and c^* are the actual and steady-state concentration distributions, respectively.

2.1 Polarization Mechanism

The polarization mechanism is characterized by the transversal flux at equilibrium (Equation 11).

$$D_x \mathbf{r}_x \frac{\partial c}{\partial x} - \mathbf{U}c = 0 \tag{11}$$

The solution of Equation (11) is the exponential concentration distribution of the sample component across the channel (Equation 12):

$$c(x) = c(0) \exp\left(\frac{\mathbf{r}_x \mathbf{U}}{D}\right. \tag{12}$$

With respect to the positive direction of the x axis, the vector **U** is negative. Consequently, Equation (11) can be rewritten in the form of Equation (13):

$$-D_x \frac{\partial c}{\partial x} - Uc = 0 \tag{13}$$

As a result, Equation (12) can be rewritten in scalar form as (Equation 14):

$$c(x) = c(0) \exp\left(\frac{-xU}{D}\right. \tag{14}$$

Defining the mean layer thickness as $l = D/U$ gives Equation (15):

$$c(x) = c(0) \exp\left(\frac{-x}{l}\right. \tag{15}$$

For strongly polarized concentration distributions, the mean layer thickness measured from the accumulation wall, is practically equal to the position of the center of gravity of the concentration distribution in the direction of the x axis.

2.2 Focusing Mechanism

The flux of a sample component in the direction of the x axis is expressed by Equation (16):

$$\mathbf{j}_x = -D_x \mathbf{r}_x \frac{\partial c}{\partial x} + \mathbf{U}(x)c \tag{16}$$

Equation (16) differs from Equation (8) by the x axis-dependent velocity $\mathbf{U}(x)$. At the steady state Equation (17) holds:

$$D_x \mathbf{r}_x \frac{\partial c}{\partial x} - \mathbf{U}(x)c = 0 \tag{17}$$

The velocity vector $\mathbf{U}(x)$ is either positive or negative regarding the orientation of the x axis and relative to $x_{\max}$. The force $F(x)$, acting on one particle undergoing the focusing can be written in scalar form as Equation (18):

$$F(x) = U(x)f \tag{18}$$

The friction coefficient is defined by Equation (19):

$$f = \frac{kT}{D} \tag{19}$$

where k is the Boltzmann constant, T is the absolute temperature, and D is the diffusion coefficient. Then Equation (20) holds:

$$\frac{\mathrm{d}c}{\mathrm{d}x} = -\frac{F(x)c}{kT} \tag{20}$$

The focusing force can be approximated by Equation (21):

$$F(x) = \left(\frac{\mathrm{d}F(x)}{\mathrm{d}x}\right._{x \approx x_{\max}} (x - x_{\max}) \tag{21}$$

where $(\mathrm{d}F(x)/\mathrm{d}x)_{x \approx x_{\max}}$ is the driving force gradient, assumed constant near $x_{\max}$. The solution of Equation (20) is given by Equation (22),

$$c(x) = c_{\max} \exp\ -\frac{1}{2kT} \left(\frac{\mathrm{d}F(x)}{\mathrm{d}x}\right._{x \approx x_{\max}} (x - x_{\max})^2 \tag{22}$$

which is the Gaussian concentration profile of a single focused component.

A completely different approach,[12,13] which takes into account the actual shape of the transverse gradient of the focusing forces, leads to accurate solution without an a priori assumption of the constancy of the gradient in the domain of the focused zone. The details are given in the section on focusing methods and techniques.

2.3 Flow Velocity Profiles

The separation is performed inside a narrow channel, usually composed of two highly polished plane and parallel surfaces between which a spacer of given constant thickness is clamped. Such a design results in a rectangular cross-section channel. Other shapes of the channel cross-section result in different hydrodynamic conditions and various flow velocity profiles. Velocity distribution in the isothermal, isoviscous flow of a Newtonian carrier liquid flowing between two parallel infinite planes under laminar conditions and not affected by a field is parabolic and described by Equation (23),

$$v(x) = \frac{\Delta P x(w - x)}{2L\mu} \tag{23}$$

where $v(x)$ is the longitudinal velocity at the x coordinate, ΔP is the pressure drop along the channel of length L, and μ is the viscosity of the carrier liquid. It holds for the average velocity (Equation 24):

$$\langle v(x) \rangle = \frac{\Delta P w^2}{12L\mu} \tag{24}$$

which is obtained by integration of Equation (23) over the whole thickness of the channel. An axially symmetric parabolic flow velocity profile is the most frequent case found in practice. However, it might be advantageous to construct a channel in which an axially asymmetrical flow velocity profile is formed. This interesting possibility is discussed in sections dealing with specific FFF methods and techniques.

2.4 Convolution of Transverse Concentration and Flow Nonuniformities

Separated species undergo two fundamental transport processes, transverse flow caused by the imposed field and longitudinal hydrodynamic flow. The separation is due to interference of the established concentration and flow nonhomogeneities. Whenever the exponential concentration distribution across the channel of each sample component is established under static conditions without hydrodynamic flow, the zones of all components are superimposed. Only under dynamic conditions, with the liquid flow applied, can separation of the individual zones occur. The sample components are transported in the direction of the longitudinal axis of the channel at different velocities, that depend on the distance of their centers of gravity from the channel accumulation wall. Larger species are usually compressed closer to the accumulation wall and, consequently, the elution order is from the smallest particles to the largest ones. Polarization FFF usually exploits constant driving forces; however, this can be varied during elution by programming.

Two limiting cases in polarization FFF are related to the intensity of the field forces. When this intensity is zero, the transverse flux of the separated species may occur under laminar flow conditions due to the shear forces. As a result, axial separation may occur as in hydrodynamic chromatography.[14] If the driving forces are strong enough to concentrate all components as close as possible to the accumulation wall, the mean distance of Brownian motion is lower than any component radius and the mean distance of the species from the accumulation wall is equal to their radius. Consequently, larger species migrate into streamlines of higher velocities than smaller ones, and are eluted more rapidly. In principle, any effective field may be applied to this steric mode of FFF (Figure 4).[8]

The situation is quite different for the convolution of Gaussian concentration distribution of the separated sample components appearing in focusing FFF with an axially symmetrical flow velocity profile. In this case, partial or complete separation can occur of the individual zones of some or all sample components in the (x axis) direction of the field. The resolution already achieved by focusing should not be degraded by the additional hydrodynamic transport in the longitudinal direction. It is evident, that if some of the zones are focused at the same but opposite distances from the central longitudinal axis of the channel, they are transported by the flow at identical velocities. Therefore, no longitudinal separation occurs even when the focused zones are well separated in the direction of the focusing field action. A convenient solution is to exploit only one half of the channel or to establish the conditions under which the asymmetrical flow velocity profile is formed. In such a case, the combination of the axially asymmetrical

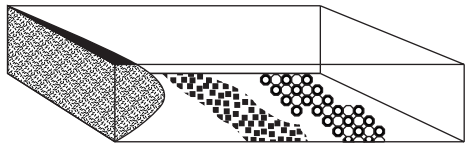

Figure 4 Principle of steric FFF. All sample components are in contact with the accumulation wall as a result of the high intensity of the applied field. As a result, steric hindrance causes larger particles to be eluted more rapidly than the smaller ones.

flow velocity profile with the Gaussian concentration distribution has the advantage that quasitotal volume of the channel can be used for the separation. It is demonstrated below that an important increase in the resolution of the focused zones can be achieved under dynamic conditions in comparison with the resolution obtained in the direction of the focusing processes under static conditions. This was an important finding, which demonstrated theoretically and experimentally the major advantage of dynamic focusing FFF over classical focusing separations performed under static conditions.[15]

2.5 Retention

Retention relates the elution velocity of the sample component in question with the average velocity of the carrier liquid. The average velocity of the zone migration of a retained sample component in the direction along the channel is given by Equation (25),

$$v_{\mathrm{r,ave}} = \frac{\langle c^*(x)v(x)\rangle}{\langle c^*(x)\rangle} \tag{25}$$

where the angle brackets denote average values. The retention ratio R is defined as the average velocity of a retained sample component to the average velocity of the carrier liquid which is equal to the average velocity of an unretained sample component (Equation 26):

$$R = \frac{v_{\mathrm{r,ave}}}{\langle v(x)\rangle} \tag{26}$$

The integral form of Equation (26) is Equation (27):

$$R = \frac{\int_0^w c(x)v(x)\,\mathrm{d}x \Big/ \int_0^w \mathrm{d}x}{\int_0^w c(x)\,\mathrm{d}x \Big/ \int_0^w v(x)\,\mathrm{d}x} \tag{27}$$

where $v(x)$, $c(x)$ are the local velocity and concentration, respectively, of the retained species. Both vary as a function of the transverse position in the direction of the x axis. The calculation of the retention ratio R gives different relationships for various concentration distributions of the retained species, corresponding to either polarization or focusing mechanisms, and various symmetrical or asymmetrical flow velocity profiles. From the practical point of view, the retention ratio R can be expressed as the ratio of the experimental retention time t_0 or the retention volume V_0 of an unretained sample component to the retention time t_{r} or the retention volume V_{r} of the retained sample component (Equation 28).

$$R = \frac{t_0}{t_{\mathrm{r}}} = \frac{V_0}{V_{\mathrm{r}}} \tag{28}$$

The retention volume of the unretained sample component should be equal to the void volume of the fractionation channel. The deviations from this ideal case can be caused by flow imperfections inside the channel and by extrachannel elements of the separation system. However, they are usually negligible for a carefully built separation system.

FFF is often performed in channels of simple geometry, which assists in establishing a rigorous analytical relationship between the retention ratio and the position of the center of gravity of the concentration distribution in the direction of the x axis. As in many cases, a known relationship between the position of the center of gravity of the zone and the molecular parameters of the sample component exists, the size parameters can be calculated from the experimental retention data without the need for a calibration procedure.

2.6 Retention Ratio in Polarization Field-flow Fractionation

Substitution for $c(x)$ and $v(x)$ in Equation (27) and integration give Equation (29):

$$R = 6\lambda\left(\coth\left(\frac{1}{2\lambda}\right. \right. \quad - 2\lambda \tag{29}$$

where $\lambda = l/w$. This equation is the basic relationship for the retention in polarization FFF performed in an ideal channel composed of two parallel and infinite planes.[16] If λ is small, $\coth(1/2\lambda)$ approaches unity and the approximation of Equation (30) holds:

$$\lim_{\lambda\to 0} R = 6(\lambda - 2\lambda^2) \tag{30}$$

Thus, for very small λ, Equation (31) is true:

$$\lim_{\lambda\to 0} R = 6\lambda \tag{31}$$

The graphical representations of Equations (29) to (31) are given in Figure 5. The retention ratio R can theoretically reach values from 1 (for an unretained sample component) to 0 (for a totally retained sample component). It can be seen that the above approximations are acceptable within the range of common experimental errors for $R < 0.25$.

Equation (29) cannot be used to calculate λ directly from the experimental retention ratio. A numerical iteration procedure must be applied. Equation (30) represents a good first iteration step in a computer iteration procedure. The physical meaning of λ can be best understood by taking into account that the force of the friction counteracting the movement of the particle due to the applied field is proportional to the product of the

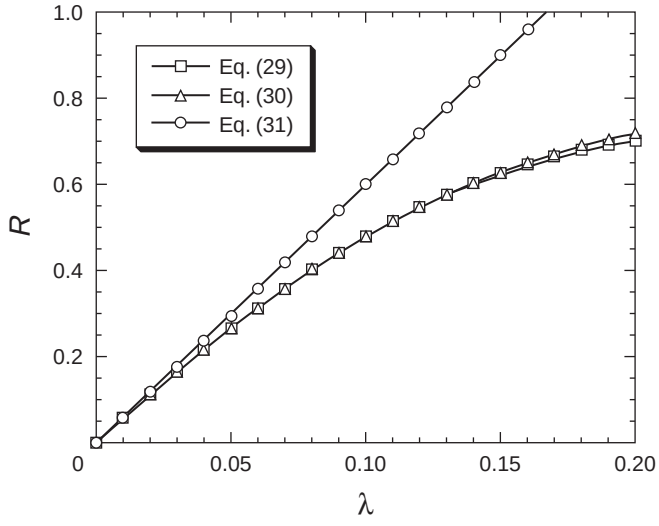

Figure 5 Dependence of the retention ratio R on the retention parameter λ calculated from rigorous and approximate relationships for polarization FFF.

velocity U and of the friction coefficient f (Equation 32):

$$= fU \tag{32}$$

At the steady state, the force of the friction is equal to the force acting on a sample component due to the effective field. Equation (33) holds for the friction coefficient:

$$f = \frac{\mathcal{R}T}{D} \tag{33}$$

where $\mathcal{R}$ is the gas constant and is Avogadro's number. Thus the velocity of the particle migration in the direction of the field action is determined by the ratio of the effective field force to the friction coefficient (Equation 34):

$$\lambda = \frac{kT}{Fw} \quad \text{or} \quad \lambda = \frac{\mathcal{R}T}{Fw} \tag{34}$$

The parameter λ is then the ratio of the thermal energy to the effect of the field on one particle. The above relationships enable evaluation of the size characteristics of the retained sample components from the experimental retention data. Whenever a straightforward relationship between the retention parameter λ and the size of the retained sample component exists, the size can be directly calculated from experimentally determined R values. Nevertheless, even in those cases when the relationship between λ and the size of the separated sample component is not known, it can be established by a convenient calibration procedure.

Most channels used for polarization FFF are constructed with a rectangular cross-section having a large width to thickness ratio that approaches an ideal two-infinite-parallel-planes configuration. In a few cases, a cylindrical capillary tube has been used as an FFF channel, so that the circular symmetry with respect to the central longitudinal axis has to be taken into account. The two distinct situations that can occur under these conditions are shown in Figure 6. In the first case the field-induced transport of the separated particles is generated in the radial direction from the central longitudinal axis of the cylindrical channel to the wall. The resulting radial concentration distribution is circularly homogeneous. This situation occurs when, for example, a semipermeable wall allows a radial uniform flow of a liquid due to the pressure difference between the interior and exterior of the cylindrical capillary, but retains larger size particles. The second situation is with an external field applied across the circular cross-section capillary.

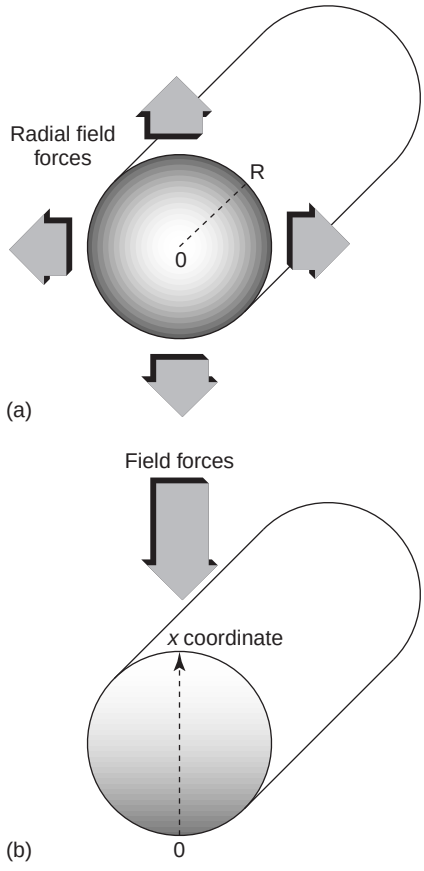

Figure 6 (a) Axially symmetric exponential concentration distribution and (b) axially asymmetric exponential concentration distribution formed in a cylindrical FFF channel.

An axially asymmetric flow velocity profile can be formed in a cross-sectional permeability separation channel whose cross-section is not rectangular but, for example, trapezoidal. This type of channel can exhibit important advantages when used for focusing FFF.[17]

However, the convolution of an axially asymmetrical flow velocity profile with an exponential concentration distribution does not improve the separation. This is because the central part of the velocity profile is often quasiparabolic, which is not advantageous when compared to a parabolic symmetrical flow velocity profile.

2.6.1 Steric Exclusion

In some cases, the size of the separated particles can be commensurable with the thickness of the channel, and the steric exclusion mechanism must be taken into account (namely in polarization FFF; Figure 4). The particles carried by the liquid flow are in permanent contact with the accumulation wall. Consequently, their size is crucial for the velocity of their longitudinal transport – larger particles move ahead of smaller ones. The limiting relationship for the retention ratio in steric FFF is given by Equation (35),[8]

$$\lim_{\alpha \to 0} R = 6\alpha \tag{35}$$

where $\alpha = r/w$, r being the particle radius. If both polarization and steric FFF participate on the separation, the limit value of R is given by Equation (36):

$$\lim_{\alpha,\lambda \to 0} R = 6\alpha + 6\lambda \tag{36}$$

More details on this mechanism can be found in the section on sedimentation FFF.

2.7 Retention Ratio in Focusing Field-flow Fractionation

General retention theory for focusing FFF[18] has been developed for basic shapes of fractionation channel cross-section, starting from the idea of sedimentation-flotation focusing field-flow fractionation (SFFFFF).[2] As mentioned above, the use of an axially symmetrical (e.g. parabolic) flow velocity profile results in being unable to separate two focused zones with the same, but opposite, distances from the central axis of the channel. To avoid this problem, it is possible to use only one half of the channel between one wall and the central axis, or to divide the channel at the end by a longitudinal splitting to conduct each half of the carrier liquid via a separate output. However, it is then necessary to use two independent detectors and this might be an experimental difficulty for analytical applications. Another solution is to form an axially asymmetrical flow velocity profile in the plane of the focused zones. These possibilities represent a challenge for fundamental methodological research, and have not been developed to a level suitable for analytical applications. Consequently only the simplest relationship, describing retention in infinite parallel planes geometry or in rectangular cross-section channels, for which the ratio of the channel width to its thickness is large enough, is given here.

The retention ratio R can be calculated by substituting from Equations (22) and (23) into Equation (27) and by integration. The approximate solution for infinite parallel planes configuration or for rectangular cross-section channel is (Equation 37)

$$R = 6(\Gamma_{\max} - \Gamma_{\max}^2) \tag{37}$$

where $\Gamma_{\max} = x_{\max}/w$ is the dimensionless coordinate of the maximal concentration of the focused zone. Numerical calculations have confirmed that Equation (37) represents a good approximation of the exact solution[18] provided that the focused zone is narrow. Relationships for retention in channels of other than rectangular cross-section can be found elsewhere.[15]

2.8 Zone Dispersion

Any separation process is accompanied by zone broadening due to the spreading processes that have a tendency to disperse the concentration distribution already achieved by separation. The efficiency of the separation increases with decreasing proportion of these dispersive processes. A conventional parameter that describes quantitatively the efficiency of any separation system is the height equivalent to a theoretical plate H (Equation 38):

$$H = L\left(\frac{\sigma}{V_r}\right)^2 \tag{38}$$

where V_r is the retention volume and σ is the standard deviation of the elution curve of a sample component. The width of the elution curve described by the standard deviation reflects several contributions to the zone broadening: longitudinal diffusion; nonequilibrium and relaxation processes; spreading due to the external parts of the whole separation system, such as the injector, detector, connecting capillaries, etc. A graphical representation of these contributions as a function of the average linear velocity of the carrier liquid is shown in Figure 7. The sum of all contributions results in a curve which exhibits a minimum. As the diffusion coefficients of the particles are very low, the longitudinal diffusion is practically negligible and the optimal efficiency (the minimum on the resulting curve) is at a very low flow velocity. Instrumental and relaxation spreading can be minimized by optimizing the experimental conditions. From the practical point of view, whenever the zone dispersion cannot be suppressed enough by optimization of the experimental conditions, then correction of the fractograms for zone broadening must be applied. These correction procedures are discussed below.

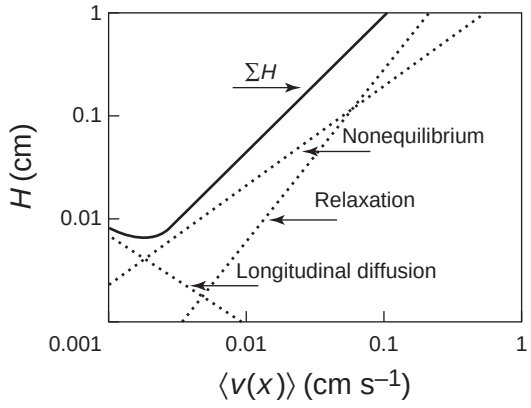

Figure 7 Dependence of the efficiency of FFF, expressed as the height equivalent to a theoretical plate H on the average linear velocity of the carrier liquid $\langle v(x)\rangle$, demonstrating various contributions to the total H.

2.8.1 Relaxation

An important contribution to zone broadening is from relaxation processes, but it can be reduced by a convenient experimental procedure. The concentration distribution of the fractionated sample across the channel immediately after the injection is far from equilibrium; the steady state is established progressively. If the flow is stopped after injection for a time necessary for the relaxation, the shift of the retention volume and the zone spreading can be suppressed.[19] It is even possible to continue the elution with a higher velocity after a stop-flow period without losing the efficiency and resolution. The effect of this relaxation on the height equivalent to a theoretical plate, and retention and its suppression by the stop-flow procedure, is demonstrated in Figure 8.[19]

If the intensity of the field varies during the elution (e.g. by programming), the additional, secondary relaxation processes occur. However, deviation from the steady state is minor if the variation of the field intensity is slow enough.

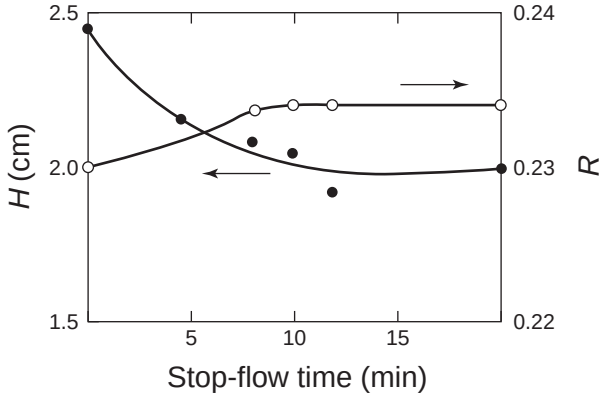

Figure 8 Effect of stop-flow time on the retention ratio R and the efficiency H.

2.9 Particle Size Distributions

Most of real particulate samples are not uniform in size but exhibit a heterogeneity (polydispersity) that can be described by a distribution function. The PSD function, or simply the PSD, is a numerical or graphical representation of the relative proportion of each particle size composing the whole analyzed sample. A minimal particle size difference in a polydisperse sample is given by an increment which corresponds to a construction unit of the particles in question. The increment can be a molecule of a given size, the monomer unit if the particle is a polymer, or the aggregate of many molecules, etc. As a result, each PSD has a discrete character. This is demonstrated in Figure 9 where the PSD of a model sample is composed of particles of six different diameters, d_1 to d_6. The heights h_i represent the corresponding amounts of the particles of the diameter d_i (e.g. h_2 for the particle of the diameter d_2). This amount can be expressed as the number of particles of a given diameter relative to the number of all particles in the sample (Equation 39),

$$N_i = \frac{n_i}{\sum_0^\infty n_i} \tag{39}$$

or as the mass of the particles of a given diameter relative to the total mass of the sample (Equation 40),

$$M_i = \frac{m_i}{\sum_0^\infty m_i} \tag{40}$$

Accordingly, the PSDs are called number or mass PSDs.

A separation method such as FFF provides the experimental data (a fractogram) that can be treated in order to obtain the required true PSD. The PSD can

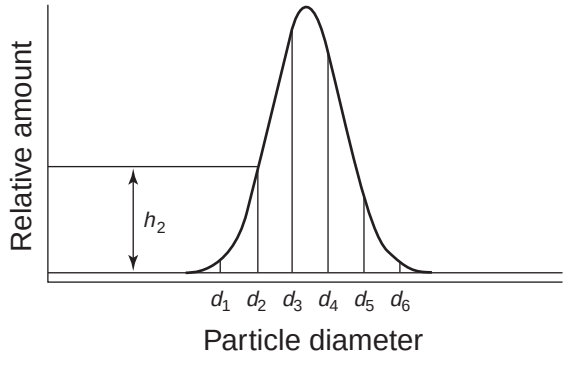

Figure 9 PSD of a model sample composed of six different particle diameters.

be used further to calculate various average particle sizes, polydispersity indexes (e.g. the standard deviation), and higher statistical parameters of the distribution function which represent a quantitative numerical information about the shape (asymmetry, etc.) of the experimentally measured PSD.

The increment between two closest particle sizes is usually too small to be distinguished by any separation method. The resulting fractogram is therefore a superposition of the individual particle sizes, and the physically discrete PSD is approximated by a continuous distribution function as shown graphically in Figure 9. The approximation can be performed by using standard methods of nonlinear regression. The raw experimental PSDs are very often and successfully approximated either by a continuous Gaussian distribution function, as defined by Equations (41) and (42), or less frequently by a discrete distribution, such as a Poisson distribution as defined by Equations (43) and (44).

The unnormalized Gaussian (differential) mass distribution function is

$$M(d_i) = \frac{\exp[-(d_i - d_{\max})^2/2\sigma^2]}{(2\pi\sigma^2)^{1/2}} \tag{41}$$

and the corresponding unnormalized Gaussian number distribution function is

$$N(d_i) = \frac{6}{\pi d_i^3 \rho_i} \frac{\exp[-(d_i - d_{\max})^2/2\sigma^2]}{(2\pi\sigma^2)^{1/2}} \tag{42}$$

The unnormalized discrete Poisson (differential) number distribution function is

$$N(P-1) = \frac{\exp(-r) r^{(P-1)}}{(P-1)!} \quad \text{with} \quad P = 1, 2, 3, \ldots \tag{43}$$

which can be easily converted into mass distribution function

$$M(P) = \frac{P \times N(P)}{\sum_0^\infty P \times N(P)} \tag{44}$$

The PSD can take the form of a differential distribution or a cumulative (integral) distribution. If the detector response is proportional to the mass of the particles, the differential mass distribution function $M(d_i)$ can be converted into integral mass distribution function $I(d_i)$ by Equation (45):

$$I_m(d_i) = \int_0^\infty M(d_i)\, \mathrm{d}d_i \tag{45}$$

Analogically, for integral number distribution function (Equation 46):

$$I_n(d_i) = \int_0^\infty N(d_i)\, \mathrm{d}d_i \tag{46}$$

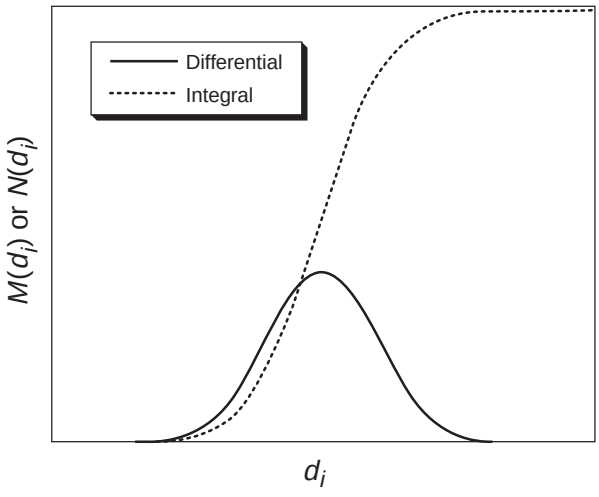

Figure 10 Model Gaussian differential and integral distribution functions.

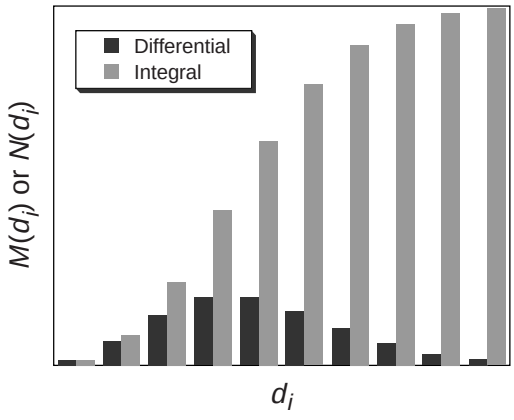

Figure 11 Model Poisson differential and integral distribution functions.

Graphical representations of the model Gaussian and Poisson differential and integral distribution functions are given in Figures 10 and 11.

2.9.1 Average Particle Sizes and Polydispersities

In addition to the PSDs, it is useful to be able to calculate various average particle sizes and polydispersity indexes. If the detector response is proportional to the mass of the particles (as already suggested above), which is the most frequent case in usual detection systems, the mass average particle diameter can be calculated from the fractogram by using Equation (47):

$$\overline{d}_m = \frac{\sum_0^\infty m_i d_i}{\sum_0^\infty m_i} = \frac{\sum_0^\infty h_i d_i}{\sum_0^\infty h_i} \tag{47}$$

The number average particle diameter can be calculated from the same fractogram by using Equation (48), provided that the density ρ_i is constant and independent of the particle diameter:

$$\overline{d}_n = \frac{\sum_0^\infty n_i d_i}{\sum_0^\infty n_i} = \frac{\sum_0^\infty h_i}{\sum_0^\infty h_i/d_i} \tag{48}$$

because Equation (49) holds true:

$$n_i = \frac{h_i}{m_i} = \frac{6h_i}{\pi \rho_i d_i^3} \tag{49}$$

The polydispersity (the width of the distribution curve) can be conveniently and most simply characterized by the index of polydispersity (Equation 50):

$$I = \frac{\overline{d}_m}{\overline{d}_n} \tag{50}$$

Another possibility is to calculate the standard deviation σ of the PSD from the fractogram, defined by Equation (51):

$$\sigma = \left[\frac{\sum_0^\infty (d_i - \overline{d}_n)^2 h_i}{\sum_0^\infty h_i} \right]^{1/2} \tag{51}$$

A graphical representation of the meaning of standard deviation of a model fractogram is given in Figure 12.

All above-mentioned PSD characteristics are derived from the defined statistical moments μ and central statistical moments relative to zero μ_0 of the jth order of a general distribution function $F(x)$ (Equations 52 and 53):

$$\mu_j = \frac{\int_{-\infty}^{+\infty} x^j F(x)\,\mathrm{d}x}{\int_{-\infty}^{+\infty} F(x)\,\mathrm{d}x} \tag{52}$$

$$\mu_{0,j} = \frac{\int_{-\infty}^{+\infty} (x - \mu_1)^j F(x)\,\mathrm{d}x}{\int_{-\infty}^{+\infty} F(x)\,\mathrm{d}x} \tag{53}$$

where Equation (54) must be satisfied for the normalized moments:

$$\int_{-\infty}^{+\infty} F(x)\,\mathrm{d}x = 1 \tag{54}$$

2.9.2 *Correction of Particle Size Distribution for Zone Broadening*

Section 2.8 demonstrated that a zone corresponding to a separation of a uniform particulate species is broadened due to dispersive processes. Consequently, the resulting fractogram of a polydisperse sample is a superposition of two processes: separation of discrete size particles and zone broadening. This is shown in Figure 13 where five different particle size fractions elute at different elution volumes, in which their zones are spread and overlapped, and the resulting fractogram is a convolution of all individual zones. The apparent PSD, calculated directly from such a raw experimental fractogram by using the above relationships, is broader than the true one.

To obtain an accurate PSD, corresponding average particle sizes, and polydispersity indexes, a correction procedure for the zone broadening should be used. An efficient correction method, applicable to FFF, was described by Jáhnová et al.[20] This method was derived from more general correction procedures, currently used

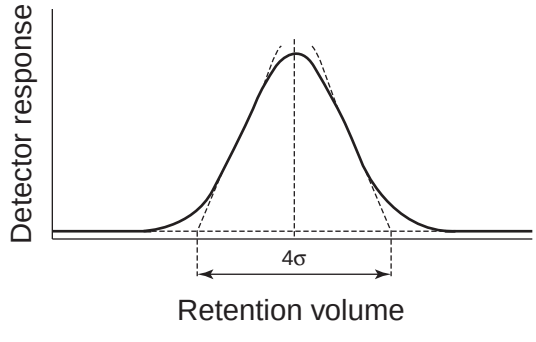

Figure 12 Graphical evaluation of the standard deviation σ from a fractogram. The distance between the sections of the tangents at inflection points with the baseline determines 4σ.

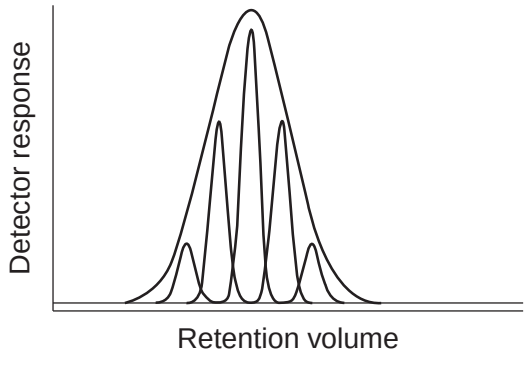

Figure 13 The fractogram of a sample of discrete PSDs is a superposition of the spread of zones of the individual discrete size species.

in size exlusion chromatography of polymers.[21] It is based on the above idea that a real experimental fractogram $h(V)$ of a polydisperse particulate sample is a convolution of two distribution functions, namely the fractogram corrected for zone broadening and which corresponds to the true PSD $g(Y)$, and the spreading function $G(V,Y)$ that represents the normalized detector response to a uniform particulate fraction having the elution volume Y (Equation 55):

$$h(V) = \int_0^\infty g(Y)G(V,Y)\,\mathrm{d}Y \tag{55}$$

where V and Y are then the elution volumes. Equation (55), called Tung integral equation, is the basis for all well-known correction methods.[21] It can be solved analytically only under certain conditions. It is, above all, the requirement of a uniform spreading function. In such a case, it becomes a convolution integral (Equation 56):

$$h(V) = \int_0^\infty g(Y)G(V - Y)\,\mathrm{d}Y \tag{56}$$

In a number of practical cases, the spreading function can be approximated by the normal Gaussian function (Equation 57):

$$G = \left(\frac{1}{2\pi\sigma_s^2}\right)^{1/2} \exp\left(\frac{-(V-Y)^2}{2\sigma_s^2}\right) \tag{57}$$

where σ_s is the standard deviation of the spreading function which is assumed to be independent of the elution volume. As this independence is valid in FFF only for a restricted range of elution volumes, Equation (57) can only be applied to fractograms of samples with a narrow PSD. However, if a nonuniform spreading function is used, the correction procedure can be applied to samples exhibiting wider PSDs (Equation 58):

$$h(V) = \int_0^\infty \left(g(Y)\frac{1}{2\pi\sigma_s^2(Y)}\right)^{1/2} \exp\left(\frac{-(V-Y)^2}{2\sigma_s^2(Y)}\right)\,\mathrm{d}Y \tag{58}$$

This function can be numerically solved. A graphical representation of a model result of the application of correction procedure is shown in Figure 14.

A practical application of the above-described treatment of an experimental fractogram is demonstrated by the real fractogram of a polymer latex obtained from sedimentation FFF, reproduced in Figure 15. The flowchart of the correction program is shown in Figure 16, and the resulting corrected PSD is compared with the uncorrected one in Figure 17.

Some authors have confirmed and extended the treatment of raw FFF experimental fractograms in order to correct the PSD for zone broadening.[22–24]

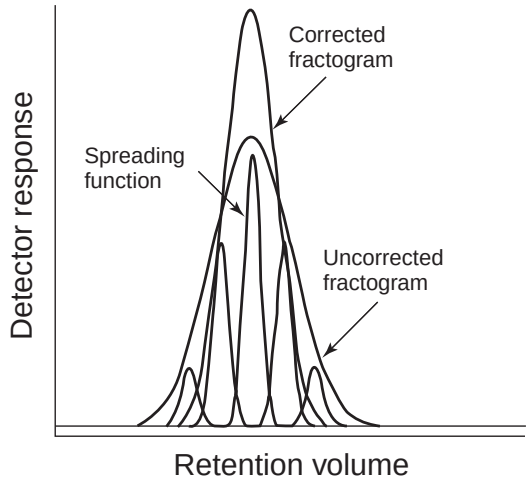

Figure 14 Effect of applying a zone-spreading correction to an uncorrected fractogram.

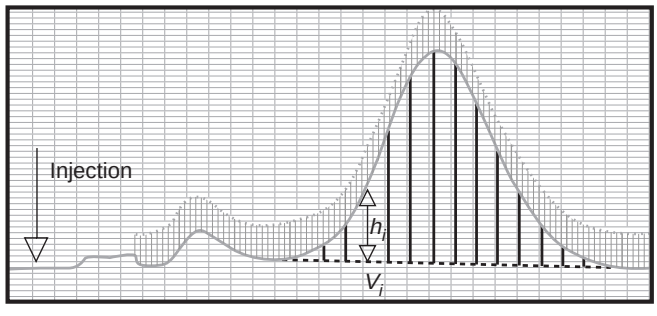

Figure 15 Treatment of an experimental sedimentation FFF fractogram of a polymer latex.

2.9.3 Practical Data Treatment

In following sections, the dependences of the retention ratio R on the diameter of the fractionated particles are discussed for various polarization FFF methods. If such a dependence is not linear but the segmentation of the experimental fractogram is regular (i.e. that ΔV_i between the h_i values is constant), the increment of the particle diameter Δd_i varies with corresponding h_i values. This fact must be taken into account in the experimental data treatment in order to obtain the accurate results.[25] The equations that follow give accurate relationships concerning various transformations of the integral to differential distributions and vice versa, and conversions of the raw experimental fractograms into PSDs. The differential mass PSD can be calculated by differentiating the integral PSD (or cumulative fractogram) according to Equation (59):

$$M(d_p) = -\frac{\mathrm{d}I(d_p)}{\mathrm{d}d_p} = -\frac{\mathrm{d}I(V)}{\mathrm{d}V}\frac{\mathrm{d}V}{\mathrm{d}V(d_p)}\frac{\mathrm{d}V(d_p)}{\mathrm{d}d_p} \tag{59}$$

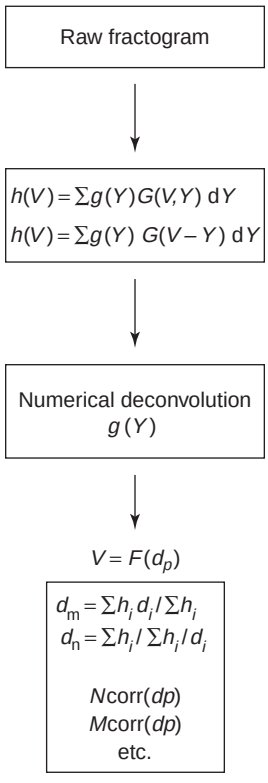

Figure 16 Flowchart of a program to correct for zone spreading.

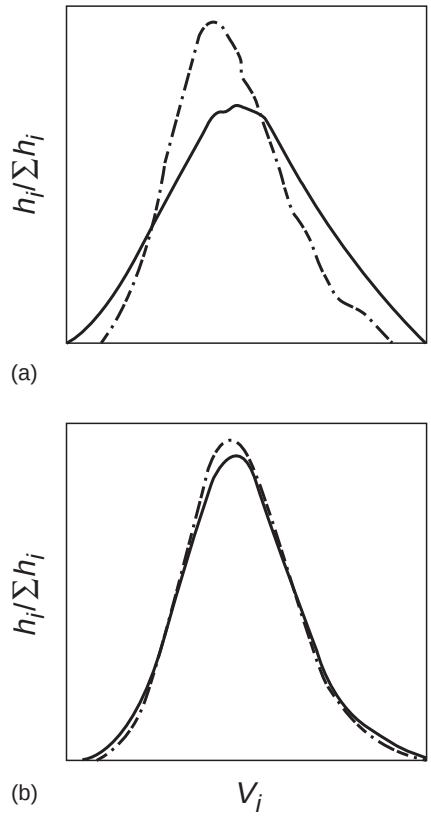

Figure 17 Effect of the correction procedure on two real fractograms. The difference between the corrected and uncorrected fractograms is very important for a fractogram obtained by (a) high-speed sedimentation FFF where the broadening of the zone due to the nonequilibrium processes cannot be neglected. (b) The difference between the corrected and uncorrected fractograms is minimal for a fractogram obtained by sedimentation FFF at reduced flow-rate where the broadening of the zone is lower. The correction for zone spreading need not necessarily be applied.

The cumulative particle mass values are calculated from Equation (60),

$$I(V) = \frac{1}{A_T} \sum_i A_i(V_i) \tag{60}$$

where A_T is the total area of the fractogram. If needed, the differential number PSD can be deduced from mass PSD by Equation (61):

$$N(d_p) = \frac{M(d_p)}{d_p} = \frac{h_i}{\sum_i h_i} \frac{\Delta V_i}{\Delta d_{p_i}} \tag{61}$$

If $M(d_p)$ and $F(V)$ are, respectively, the mass PSD and the experimental fractogram, the total mass of particles can be expressed by Equation (62):

$$\sum_i M^i(d_p)\Delta d_i = \sum_i F^i(V)\Delta V_i \tag{62}$$

and, by considering the relationship between the retention volume and the particle size, $V = g(d_p)$, and its derivative $g'(d_p)$ (Equation 63),

$$g'(d_p) = \frac{\Delta V_i}{\Delta d_i} \tag{63}$$

Equation (62) becomes Equation (64):

$$\sum_i \frac{M^i(d_p)}{g'_i(d_p)} \Delta V_i = \sum_i F^i(V)\Delta V_i \tag{64}$$

and, consequently (Equation 65),

$$F^i(V) = \frac{M^i(d_p)}{g'_i(d_p)} \tag{65}$$

For example, Equation (47) can be expressed as Equation (66):

$$\overline{d}_m = \frac{\sum_i d_i M^i(d_p)\Delta d_i}{\sum_i M^i(d_p)\Delta d_i} = \frac{\sum_i d_i F^i(V)\Delta V_i}{\sum_i F^i(V)\Delta V_i} \quad \text{etc.} \tag{66}$$

The acquisition of the experimental data and treatment of the fractogram are usually performed by a computer connected on line to the separation system.

3 METHODS, TECHNIQUES AND THEIR APPLICATIONS

The retention parameters R or λ are directly related to the effective diameters of the separated particles for some of the polarization techniques, such as sedimentation FFF. This relationship is not straightforward for some other polarization FFF methods. For example, λ is a complex function of the size and charge for electric FFF. It is a function of the thermal diffusion factor, which is a function of the diffusion coefficient and of the coefficient of the thermal diffusion, and the diffusion coefficient is inversely proportional to the size of the fractionated species, but a definite form of this dependence does not exist in thermal FFF.

The transverse positions of the zones and, as a result, the values of Γ_{max} in focusing FFF are related to the corresponding extensive properties of the fractionated samples, such as the density. The intensive properties often influence the width of the focused zones.

FFF is then used to characterize and to analyze quantitatively the properties related to the retention. Even if the determined characteristic of the separated sample is not explicitly related to the retention parameter, an empirical calibration can be established by first using known standards which are subsequently applied thereafter to determine this characteristic of the sample under investigation.

3.1 Polarization Field-flow Fractionation

The character of the applied field determines the particular method of polarization FFF. The most important of them, which have already found many practical applications in particle size analysis, are sedimentation FFF, flow FFF, electric FFF, and thermal FFF.

Other techniques have been proposed but their routine laboratory practice remains to be elaborated.

3.1.1 Sedimentation Field-flow Fractionation

This technique is based on the action of gravitational or centrifugal forces on the suspended particles. The sedimentation velocity is proportional to the product of the effective volume and density difference between the suspended particles and the carrier liquid. The channel is placed inside a centrifuge rotor, as shown in Figure 18. The technique can be applied to the analysis and characterization of polymer latex particles, inorganic particles, emulsions, biological cells, etc. The retention parameter λ depends on the effective mass of the particles according to Equation (67):

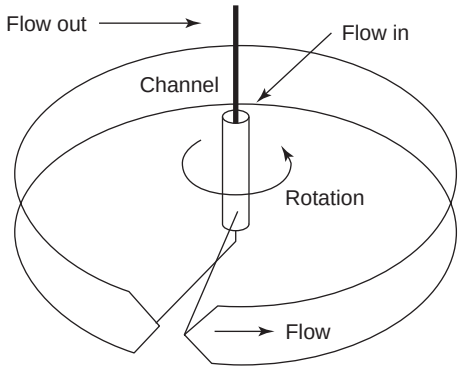

Figure 18 Schematic representation of sedimentation FFF channel.

$$\lambda = \frac{6kT}{\pi d^3 g w \Delta\rho} \tag{67}$$

where g is the gravitation or centrifugal acceleration and $\Delta\rho$ is the density difference between the particles and the carrier liquid. Calculation of the PSD is possible directly from the retention data by applying Equation (67) without any calibration.

The fractionation and analysis of colloidal particles in river water, the analysis of diesel exhaust soot, and of nuclear energy-related materials, are typical examples of the use of sedimentation FFF in the investigation of environmental samples.[26] Oil emulsions can also be analyzed by sedimentation FFF. Biopolymers and particles of biological origin are among the most interesting objects to have been fractionated.[15,27] The performance of sedimentation FFF is superior to or as good as other separation methods. A complication in interpreting experimental data is that the retention is proportional to the product of particle size and particle density. When performing fractionation in a single carrier liquid, the density must be assumed constant for all particles. However, it is possible to determine the size and density of the particles independently if fractionations are performed in more carrier liquids of various densities.

A typical application[28] of sedimentation FFF which allowed detection of a bimodal PSD in a sample of a polymer latex (supposed originally to have a very narrow and monomodal PSD) is shown in Figure 19(a). The order of the elution from the small to the large diameter particles corresponds to the polarization mechanism. Figure 19(b) shows a rapid, high resolution sedimentation FFF of the polymer latex particles.[29] The separation was carried out at high centrifugal force; the mechanism of steric FFF dominated, and the order of the elution was inverted.

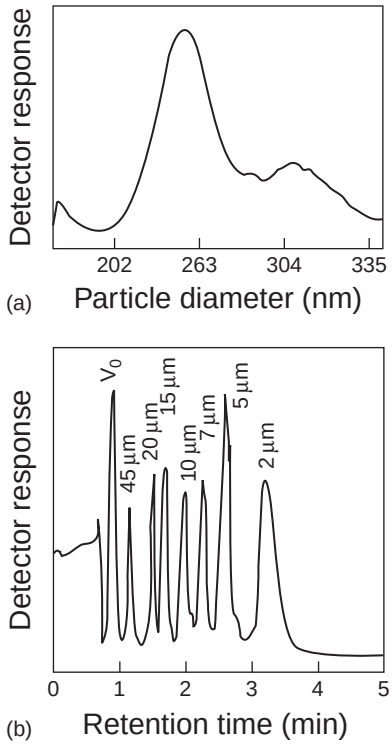

Figure 19 (a) Fractogram of poly(glycidyl methacrylate) latex showing the bimodal character of the PSD not recognised by electron microscopy. (b) Fractogram of high-speed high-resolution sedimentation FFF of latex beads obtained by steric separation.

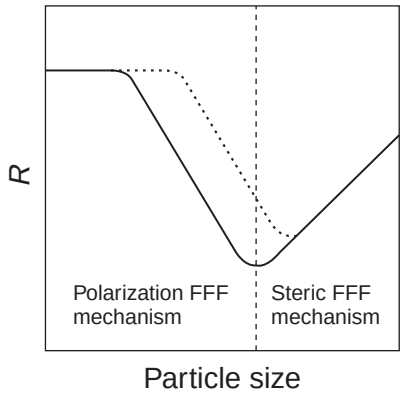

Figure 20 Dependence of the retention ratio on particle size for polarization and steric mode FFF at two different centrifugal forces. The dotted curve is for the lower force.

The domains of two competitive mechanisms are shown in Figure 20. The retention ratio decreases with increasing particle size in the range of the polarization mechanism and increases when the steric FFF mechanism prevails. When both mechanisms are in competition, the particles of different sizes can elute unseparated. However, if the field force is modified, the domain of the simultaneous action of both mechanisms can be displaced to avoid the deterioration of the separation.

Refinements to sedimentation FFF have generated new applications of this technique to the PSD analysis of samples of very diverse origin.[30–42]

3.1.2 Flow Field-flow Fractionation

Flow FFF is a universal method because the cross-flow, perpendicular to the flow of the carrier liquid along the channel, creates an external hydrodynamic field that acts on all particles uniformly,[43,44] and separation of particles of different sizes is caused by the differences in diffusion coefficients. The channel (Figure 21) is formed between two parallel semipermeable membranes fixed on porous supports, which maintain the uniform distribution of the cross-flow over the whole surface. The carrier liquid can permeate through the semipermeable membranes but the separated particles cannot. The retention parameter λ is given by Equation (68):

$$\lambda = \frac{kTV_0}{3\pi\mu V_C w^2 d_p} \tag{68}$$

where V_0 is the void volume of the channel, μ is the viscosity of the carrier liquid, and V_C is the volumetric velocity of the cross-flow. Flow FFF can also be carried out in a circular cross-section capillary. Radial forces are due to an internal over-pressure[45] which causes flow of the carrier liquid through the semipermeable cylindrical capillary. The advantage of this arrangement is lower dilution of the separated particles during elution, which is caused by a partial depletion of the liquid due to the radial cross-flow.

One example of an interesting application of flow FFF is shown in Figure 22: the retentions of two silica samples injected separately and in a mixture[46] indicated that the separation was not influenced by interactions between two particulate samples. The availability of the membranes and semipermeable capillaries compatible

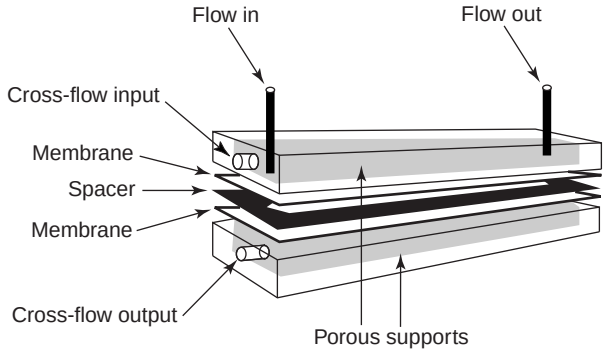

Figure 21 Diagram of a flow FFF channel.

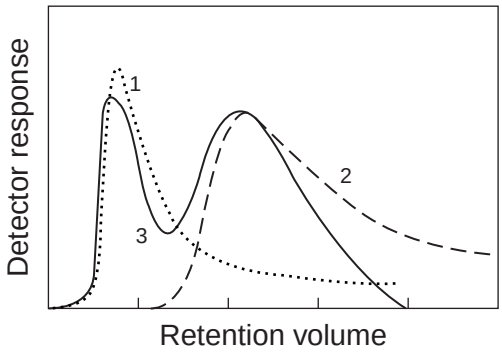

Figure 22 Flow FFF fractograms of two different samples of colloidal silica obtained separately (peaks 1 and 2) and in a mixture (peak 3).

with various solvents allows the operation of flow FFF with both water-based and organic solvents.

The original construction of the flow FFF channel was proposed and implemented experimentally.[47–53] Only the accumulation wall of the channel is semipermeable, so that a part of the carrier liquid passes through the membrane thus generating the hydrodynamic flow field across the channel. In order to optimize the linear velocity of the longitudinal flow, the channel width decreases from the inlet to the outlet.[53] However, the hollow fiber capillary used as a fractionation channel was found to be a simple and convenient solution for flow FFF.[54] Some selected review articles demonstrate a large field of practical applications of flow FFF in PSD analysis.[55–57]

3.1.3 Electric Field-flow Fractionation

Electric FFF uses an electric potential drop across the channel to generate a transverse flux of charged particles.[58] The walls of the channel are usually formed by semipermeable membranes that allow the passage of the small ions but not the permeation of the separated particles (Figure 23). The dependence of the retention parameter λ on the electrophoretic mobility μ_e, and the diffusion coefficient of the charged particles is given by Equation (69):

$$\lambda = \frac{D}{\mu_e E w} \tag{69}$$

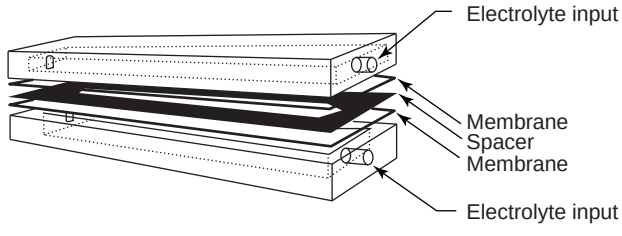

Figure 23 Diagram of an electric FFF channel.

where E is the electric field strength. The ratio of the diffusion coefficient to the electrophoretic mobility of the particles determines the retention in electric FFF. The particles exhibiting only small differences in electrophoretic mobilities but with an appreciably broad PSD and, consequently, important differences in diffusion coefficients, can be separated. The advantage of electric FFF compared with electrophoretic separations such as capillary electrophoresis, is that high electric field strength can be achieved at low absolute values of the electric potential due to a small distance between the walls of the channel.

Electric FFF is especially suited to separations of biological cells as well as charged polymer latexes and other colloidal particles. The separation is strongly influenced by the concentration of the injected sample, the concentration of the buffer solution, the pH, and by interparticle interactions. This can produce experimental difficulties for conventional fractionations; however, it can be exploited for investigation into the interactions of charged particles. As for flow FFF, electric FFF can be carried out by using a semipermeable tube of circular cross-section with the electric field applied transversally.[43,59] Fractionation of charged particles represents a wide field of application.[60–65]

3.1.4 Thermal Field-flow Fractionation

Thermal FFF was the first experimentally implemented technique.[16] Until recently it was used mostly for the fractionation of macromolecules; now there have been some attempts to apply this method to the fractionation of particles. The future potential of thermal FFF justifies a brief description of this technique, regardless of its recent limited use for the PSD analysis.

The temperature difference between two metallic bars, forming the channel walls with highly polished surfaces and separated by a spacer in which the channel proper is cut, produces the flux of the sample components, known as Soret effect, usually toward the cold wall. The construction of a channel for thermal FFF is shown in Figure 24.

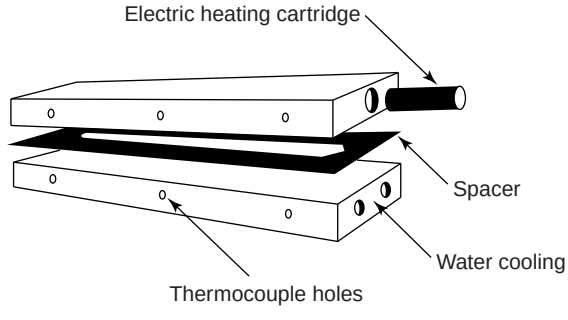

Figure 24 Diagram of a thermal FFF channel.

The relation between the λ, characteristics of the fractionated particles, and the operational variables is given by Equation (70):

$$\lambda = \frac{D}{wD_T(dT/dx)} \tag{70}$$

where D_T is the coefficient of the thermal diffusion which depends on the chemical composition and structure of the fractionated species but not on their size. As a result, the differences in thermal diffusion coefficients allow fractionation according to differences in chemical composition and structure. However, the diffusion coefficient D depends on the size which should allow fractionation according to PSD.

Although it might appear possible to evaluate molar masses or particle sizes from an experimental fractogram by using simply Equation (70), an empirical calibration curve (Figure 25[66]) is usually constructed by using a series of samples of known characteristics. The experimental retention ratios are used to plot the calibration curve. This calibration can be used to determine the characteristics of an unknown sample of the same chemical composition and structure with the same temperature gradient applied.

The most frequent applications of thermal FFF concern polymers. The pressurized separation systems permit operation above the normal boiling point of the solvent used. The fractionations can be finished in few minutes or seconds. Performance parameters favor thermal FFF over the competitive methods, at least for the fractionation of macromolecules. The fractogram in Figure 26(a) shows a wide range of the polymer molar masses which can be determined in a single experiment by using time-programmed temperature gradient.[67] Figure 26(b) demonstrates ultrarapid and high-resolution thermal FFF.[68]

As concerns the fractionation of the particles and consecutive analysis of the PSD, the actual potentials of thermal FFF have to be verified experimentally and theoretically. It has been shown that various particles

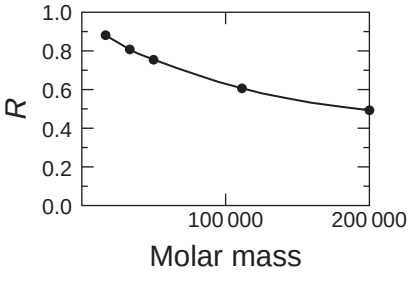

Figure 25 Calibration curve $R = f(M)$ for a thermal FFF channel for polystyrene standards at a given temperature gradient.

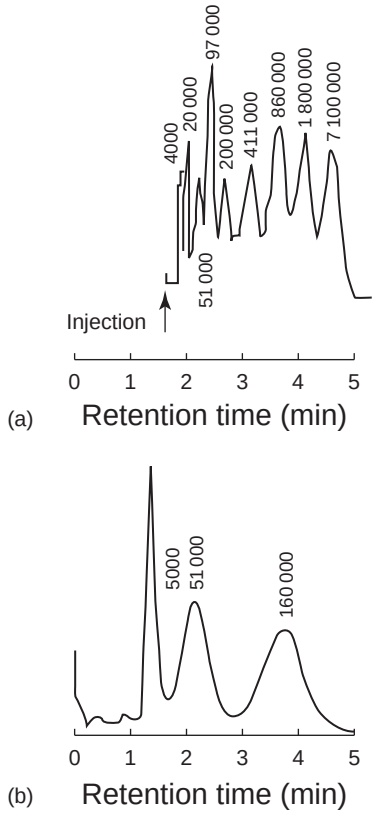

Figure 26 Fractograms of different molar mass polystyrenes: (a) obtained with temperature gradient programming and of polystyrene standards of different molar masses; (b) showing an ultrarapid separation.

exhibit retention in thermal FFF.[69,70] However, the retention phenomenon is driven by surface properties of the retained particles[71–75] and, consequently, the characterization of PSD of a studied sample from thermal FFF data is not yet easy.

3.1.5 Other Polarization Field-flow Fractionation Methods

Other polarization FFF methods include magnetic FFF, concentration FFF (making use of a concentration gradient of a mixed solvent to induce the chemical potential gradient), and shear FFF. Although these methods may become highly relevant for PSD analysis in the near future, their recent technical level does not allow their routine use.

Each of the above-described polarization FFF methods can apply the field, of constant or variable strength, during the fractionation. Steric FFF is applicable independently of the physical character of the field, but gravitation or centrifugal forces or the cross-flow of the carrier liquid are the most frequently used for the fractionation of practically important particles such as the coal particles,

human and animal cells, particles originating from the environment, etc. Precautions must be taken to prevent problematic situations that can occur when some particles may undergo steric exclusion, with the elution order from large to small, while other particles can experience polarization, with the elution order inverted.

3.2 Focusing Field-flow Fractionation

Focusing FFF is less developed than polarization FFF and, consequently, has not yet been exploited in routine laboratory work due to its substantially shorter history. Focusing FFF methods can be classified according to various combinations of the driving field forces and gradients:

- effective property gradient (perichoric gradient) of the carrier liquid;
- preformed perichoric gradient;
- cross-flow velocity gradient;
- lift forces;
- shear stress;
- gradient of nonhomogeneous field action.

These principles are promising for PSD analysis and merit brief discussion here.

3.2.1 Perichoric Gradient of the Carrier Liquid Combined with a Field

Focusing can occur due to the effective property gradient of the carrier liquid in the direction across the channel, combined with a primary or secondary transverse field. The density gradient as in SFFFFF or the pH gradient as in isoelectric focusing field-flow fractionation (IEFFFF) have already been implemented.[9,10,12,13,76,77] Two simple channels are demonstrated in Figures 27(a) and 28(a) as well as two examples of practical applications showing the separation of two types of the polystyrene latex particles by SFFFFF (Figure 27b) and the separation of two components of horse myoglobin by IEFFFF (Figure 28b). The separation by SFFFFF was carried out according to the density difference of the latex particles, whereas the electric field was applied to generate the density gradient in a suspension of charged silica particles. The separation by IEFFFF was carried out according to the isoelectric point differences by using the electric field to generate the pH gradient and to focus the sample components.

Retention theory for the special case of focusing FFF in a density gradient is developed here in more detail because a similar approach can be applied to all the methods and techniques described below. The transverse gradient can be generated, for example, in a binary or multicomponent carrier liquid whose two or more components are affected unevenly by the primary

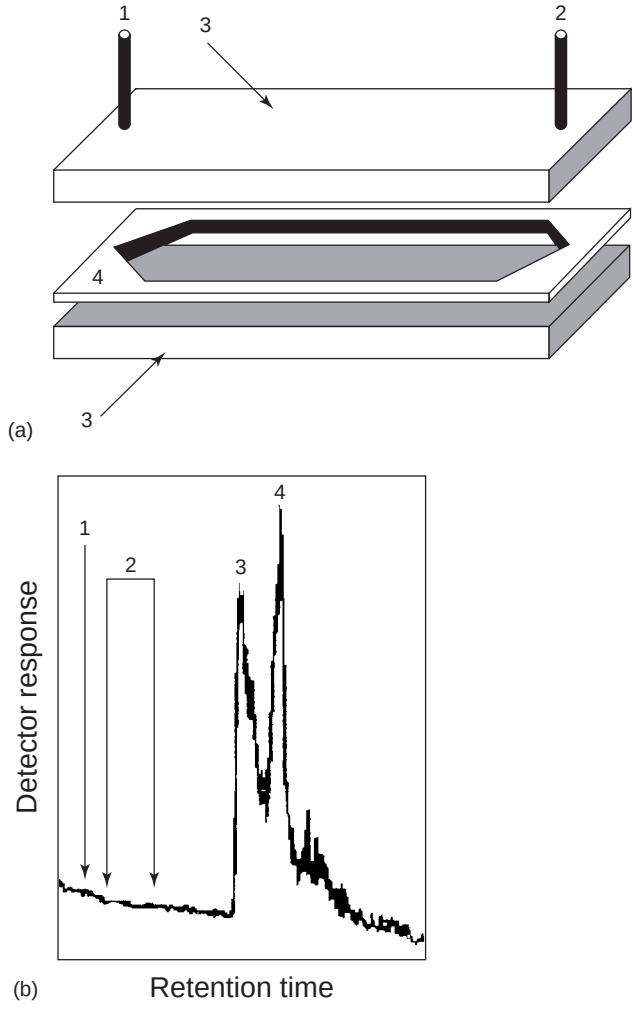

Figure 27 (a) Diagram of a channel for focusing FFF in coupled electric and gravitational fields: (1) input, (2) output, (3) electrodes and (4) spacer. (b) Fractogram of two samples of polystyrene latex particles showing a good resolution obtained by focusing FFF, whereas no detectable resolution was achieved under static conditions: (1) injection, (2) stop-flow period, peaks corresponding to particle diameters of (3) 9.87 μm and (4) 40.1 μm.

homogeneous field. The primary field can also act to form the focused zones of the separated sample components. Another possibility is to superimpose a secondary field of different nature on the gradient generated by the primary field.[9,10,12,13] The density gradient can be generated by centrifugal or electric field forces acting on colloidal particles, suspended in a liquid. The concentration distribution of the perichoric modifier particles due to the primary field is described by identical exponential distribution function as in the case of polarization mechanism (Equation 71):

$$c_{\mathrm{m}}(x) = c_{\mathrm{m}}(0) \exp\left(\frac{-x|U|}{D_{\mathrm{m}}}\right. \qquad (71)$$

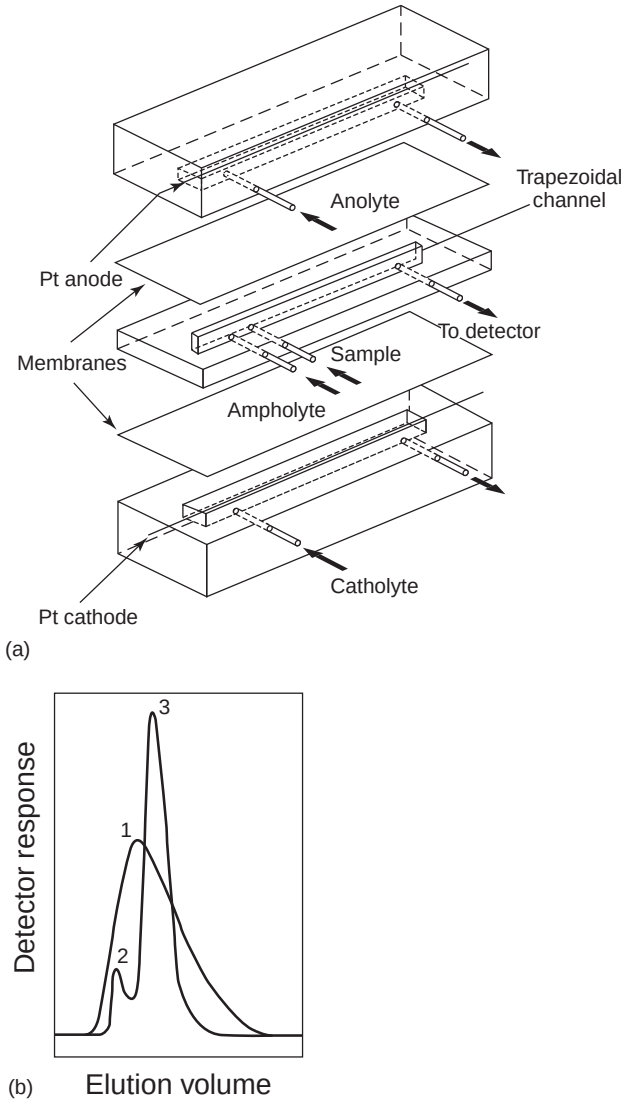

Figure 28 (a) Diagram of a channel for IEFFFF. (b) Fractograms: (1) without the electric field applied; and with the active field showing the separation of (2) acidic and (3) basic components of horse myoglobin.

The subscript m specifies the parameters concerning the modifier. The concentration $c_m(0)$ in Equation (71) can be calculated from the relationship for the average concentration $c_{m,ave}$ which is equal to the known initial concentration (Equation 72):

$$c_{m,ave} = \frac{\int_0^w c_m(x)\,dx}{\int_0^w dx} = c_m(0)\frac{D_m}{wU_m}\left(1 - \exp\left(\frac{-wU_m}{D_m}\right.\right. \tag{72}$$

Substituting into Equation (71) gives Equation (73):

$$c_m(x) = \frac{c_{m,ave} wU_m}{D_m(1 - \exp(-wU_m/D_m))}\exp\left(\frac{-xU_m}{D_m}\right. \tag{73}$$

The focusing force proper can be generated by combining the transverse gradient and the primary or secondary field action. Primary electric field can be applied to generate the density gradient due to the migration of the charged colloidal particles. The secondary gravitation coupled simultaneously with the established density gradient causes the focusing of a particulate sample component.

A modified version of Equation (20) for the focusing force gradient can be used to describe the focusing effect under these specific conditions (Equation 74):

$$\frac{dc_i}{c_i} = -\frac{U_i(x)}{D}\,dx \tag{74}$$

where c_i is the concentration of the focused sample component. The subscript i is used to distinguish between the concentration distribution, $c_i(x)$, and the related parameters of the ith focused sample component and the concentration distribution, $c_m(x)$, and the related parameters of the gradient forming component of the carrier liquid. Equation (73) can be transformed into density terms instead of concentrations (Equation 75):

$$\rho(x) = \rho_l + \frac{\phi_{m,ave}\Delta\rho_m wU_{m,e}}{D_m(1 - \exp(-wU_{m,e}/D_m))}\exp\left(\frac{-xU_{m,e}}{D_m}\right. \tag{75}$$

The macroscopic focusing force acting on a single focused species can be calculated from Equation (76):

$$F_i(x) = (\rho(x) - \rho_i)v_i g_f \tag{76}$$

The concentration distribution of the focused particles in a density gradient formed simultaneously due to the action of primary electric field on two component colloidal carrier liquid is obtained by substituting into Equation (74) and by integration (Equation 77):

$$c_i(x) = c_i(x_{max})\exp\left\{-\left(\frac{v_i g_f \phi_{m,ave}\Delta\rho_m w}{kT(1 - \exp(-wU_{m,e}/D_m))}\right.\right.$$
$$\times\left(\exp\left(-\frac{xU_{m,e}}{D_m}\right.\right. - \exp\left(-\frac{x_{max}U_{m,e}}{D_m}\right.\right.$$
$$\times\left.\left.\left(1 + \frac{U_{m,e}(x_{max} - x)}{D_m}\right.\right.\right\} \tag{77}$$

where v_i is the volume of the focused particle, g_f is the acceleration of the focusing field, $\phi_{m,ave}$ is the average volume fraction of the modifier, $\Delta\rho_m = \rho_m - \rho_l$ is the density difference between the modifier component and the suspending liquid and $U_{m,e}$ is the velocity of the modifier species due to the primary electric field. Equation (77) was obtained without assuming the constancy of the density gradient. The Gaussian distribution, Equation (22), deviates substantially from Equation (77).[9,10,12,13]

3.2.2 *Preformed Gradient Combined with a Field*

The effective property gradient of the carrier liquid, such as the density gradient, can be preformed at the beginning of the channel and combined with the primary or secondary field forces. The step density gradient is formed in such a case. The preforming is not limited to density gradient. A scheme of an experimental arrangement for this method and a typical experimental result[78,79] is shown in Figure 29.

3.2.3 *Cross-flow Velocity Gradient Combined with a Field*

The focusing occurs in the transverse flow velocity gradient of the carrier liquid which opposes the action of the field. The longitudinal flow of the liquid is imposed simultaneously. This elutriation focusing field-flow fractionation (EFFFF) method has been investigated experimentally by using the trapezoidal cross-section channel[80,81] but use of the rectangular cross-section channel is possible. The principle of the method and the result of fractionation of micrometer-size polystyrene latex particles are demonstrated in Figure 30.

3.2.4 *Lift Forces Combined with a Field*

The hydrodynamic lift forces that occur at high flow rates of the carrier liquid combined with the primary field, are able to concentrate the suspended particles into the focused layers.[82] The retention of the particles under the simultaneous effect of the primary field and lift forces generated by the high longitudinal flow rate can vary with the nature of various applied primary field forces.

3.2.5 *Shear Stress Combined with a Field*

A high shear gradient in a carrier liquid can lead to the deformation of soft particles. The established entropy gradient generates the driving forces that displace the particles into a low shear zone.[83] At a position where all the driving forces are balanced, the focusing of the sample components can occur.[84] Although this method was originally proposed for a temperature gradient acting as a primary field, generating the thermal diffusion flux of the macromolecules which opposes the flux due to the entropy changes generated motion, in principle it should be applicable to soft particles as well.

3.2.6 *Gradient of a Nonhomogeneous Field*

A nonhomogeneous high-gradient magnetic field can be used to separate various paramagnetic and diamagnetic particles of biological origin by a mechanism of focusing FFF.[85] A concentration of paramagnetic particles near the center of the cylindrical capillary and the focusing

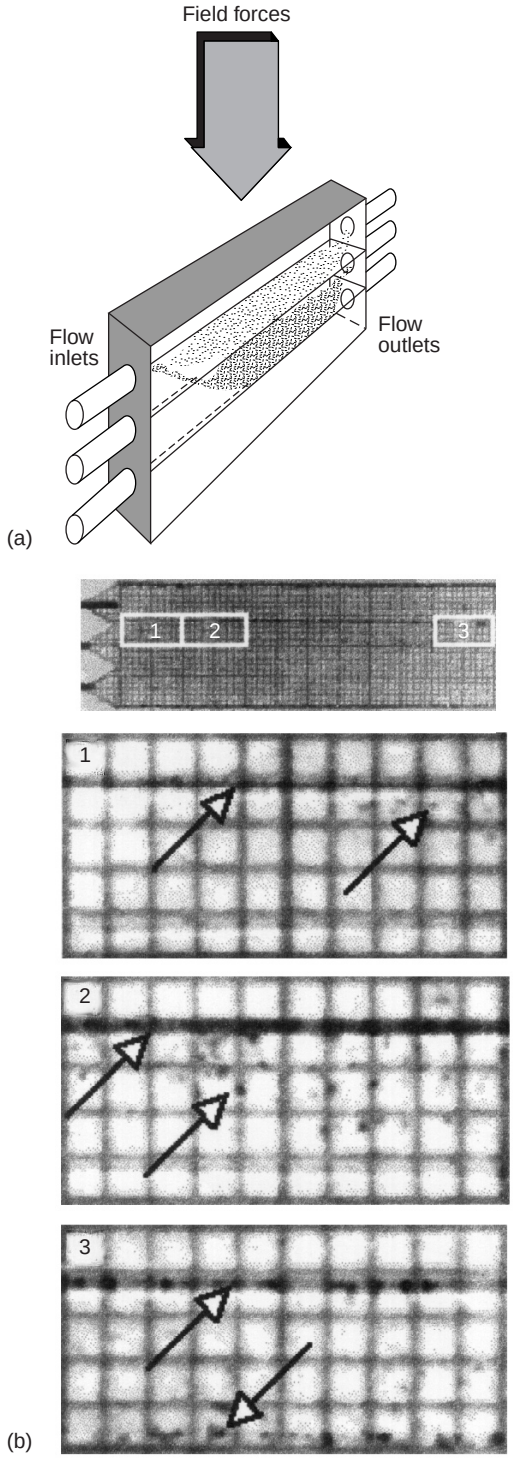

Figure 29 (a) Diagram of a separation channel for isopycnic focusing FFF in preformed stepwise density gradient. (b) Evolution of two focused layers of the particles at three different longitudinal positions inside the channel: (1) a layer of the particle mixture positioned at the interface between two different density liquids at the beginning of the separation; (2) evolution of the separation of different density particles; (3) two zones focused at the interfaces of different density liquids.

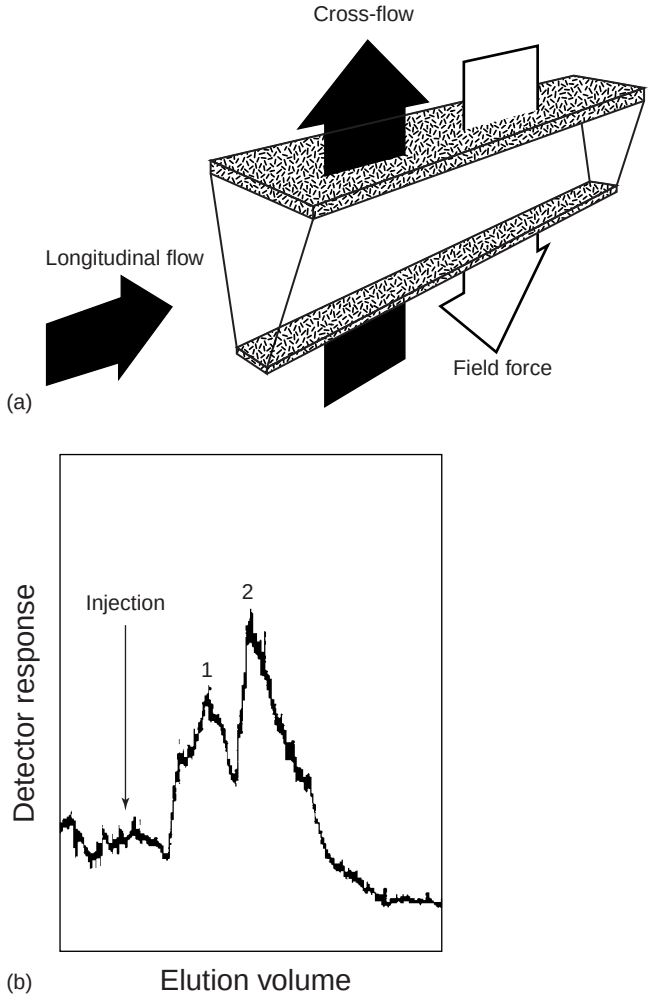

Figure 30 (a) Principle of EFFFF in a trapezoidal cross-section channel. Focused zones are formed due to the field forces and the opposite velocity gradient of the carrier liquid flow across the semipermeable lower and upper walls of the channel. Separation of the focused zones in the direction of the longitudinal flow in trapezoidal cross-section channel is facilitated by the formation of axially asymmetrical flow velocity profile. (b) Fractogram showing the separation of two polystyrene latexes of (1) 1.6 μm diameter and of (2) 5 μm in diameter particles.

of diamagnetic particles in a free volume of the capillary should occur. To our knowledge, no experimental results have yet been published.

3.2.7 Other Possibilities

Other gradients and a variety of fields can probably be combined to produce focusing, making these phenomena suitable for PSD analysis. The above discussions on the mechanisms exploited in focusing FFF give an idea as to their potential. More information on the state of the art can be found in a review.[86]

4 CONCLUDING REMARKS

Scientific progress passes through discoveries, inventions, and innovations. As a specific scientific discipline, analytical chemistry follows this process but also in the context of the whole area of scientific research. The polarization and focusing FFF methodologies represent an important contribution to the science and technology of separation, and the analysis of particles of synthetic or natural origin over a wide range of sizes, and of organized structures such as cells and microorganisms. Research and technology related to the life sciences and to macromolecular chemistry and technology, the analytical problems appearing in context with the protection of the environment, and many other scientific and technological activities, have stimulated the development of new analytical separation methods. The article shows that important contributions to experimental implementations have been obtained with model systems. The availability of commercial instrumentation systems is a crucial factor with regard to the progressive introduction of FFF methodologies into current laboratory practice. Moreover, it is important to recognise that all components of the apparatus are absolutely identical to those for liquid chromatography, with one exception – the separation channel. A liquid chromatograph can be modular, being composed of individual commercially available parts, such as the pump, injector, detector, etc., with a column that can be changed for individual applications. The FFF apparatus can be assembled similarly, with various channels which, in most cases, are easy to construct.

ABBREVIATIONS AND ACRONYMS

EFFFF	Elutriation Focusing Field-flow Fractionation
FFF	Field-flow Fractionation
IEFFFF	Isoelectric Focusing Field-flow Fractionation
PSD	Particle Size Distribution
SFFFFF	Sedimentation-flotation Focusing Field-flow Fractionation

RELATED ARTICLES

Food **(Volume 5)**
Particle Size Analysis in Food

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Filtration in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Sieving in Particle Size Analysis • Turbidimetry in Particle Size Analysis • Ultrasonic Measurements in Particle Size Analysis • Velocimetry in Particle Size Analysis

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization • Field Flow Fractionation in Analysis of Polymers and Rubbers • Size-exclusion Chromatography of Polymers • Supercritical Fluid Chromatography of Polymers • Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation

REFERENCES

1. J.C. Giddings, 'A New Separation Concept Based on a Coupling of Concentration and Flow Nonuniformities', *Sep. Sci.*, **1**, 123–125 (1966).
2. J. Janča, 'Sedimentation-flotation Focusing Field-flow Fractionation', *Makromol. Chem., Rapid Commun.*, **3**, 887–890 (1982); 'Correction', ibid, **4**, 267 (1983).
3. G. Agricolae, *De Re Metallica Libri XII*, Basileae, 1556.
4. J.G. Kirkwood, 'A Suggestion for a New Method of Fractionation of Proteins by Electrophoresis Convection', *J. Chem. Phys.*, **9**, 878–883 (1941).
5. J.G. Kirkwood, R.A. Brown, 'Diffusion-convection. A New Method for the Fractionation of Macromolecules', *J. Am. Chem. Soc.*, **74**, 1056–1061 (1952).
6. K. Clusius, G. Dickel, 'Neues Verfahren zur Gasentmischung und Isotopentrennung', *Naturwiss.*, **26**, 546–550 (1938).
7. J. Montgolfier, E. Montgolfier, *Description des expériences aérostatiques de MM. de Montgolfier, et de celles auxquelles cette découverte a donné lieu*, Cuchet, Paris, 1784.
8. J.C. Giddings, M.N. Myers, 'Steric Field-flow Fractionation: A New Method for Separating 1–100 μm Particles', *Sep. Sci. Technol.*, **13**, 637–645 (1978).
9. J. Janča, 'Isoperichoric Focusing Field-flow Fractionation and Thin Layer Isoperichoric Focusing Based on Coupling of Gradient Generated by Primary Field with Secondary Field Action: General Principles and Performances', *Mikrochim. Acta*, **112**, 197–215 (1994).
10. J. Janča, 'Isoperichoric Focusing Field-flow Fractionation Based on Coupling of Primary and Secondary Field Action', in *Chromatographic Characterization of Polymers, Hyphenated and Multidimensional Techniques*, eds. T. Provder, H.G. Barth, M.W. Urban, Advances in Chemistry Series 247, ACS, Washington, DC, 1995.
11. A. Kolin, 'Reminiscences about the Genesis of Isoelectric Focusing and Generalization of the Idea', in *Electrofocusing and Isotachophoresis*, eds. B.J. Radola, D. Graesslin, de Gruyter, Berlin, 1977.
12. J. Janča, R. Audebert, 'Focusing Field-flow Fractionation: Coupling of Density Gradient Generated by Primary Field with Secondary Field Focusing Forces', 5th International Symposium on Polymer Analysis and Characterization, Inuyama, Japan, 1992.
13. J. Janča, R. Audebert, 'Focusing Field-flow Fractionation: Coupling of Density Gradient Generated by Primary Field with Secondary Field Focusing Forces', *J. Appl. Polym. Sci.*, **52**, 63–78 (1993).
14. H. Small, 'Hydrodynamic Chromatography: A Technique for Size Analysis of Colloidal Particles', *J. Colloid Interface Sci.*, **48**, 147–150 (1974).
15. J. Janča, *Field-flow Fractionation: Analysis of Macromolecules and Particles*, M. Dekker, New York, 1987.
16. M.E. Hovingh, G.H. Thompson, J.C. Giddings, 'Column Parameters in Thermal Field-flow Fractionation', *Anal. Chem.*, **42**, 195–203 (1970).
17. J. Janča, V. Jáhnová, 'Sedimentation-flotation Focusing Field-flow Fractionation in Channels with Modulated Cross-sectional Permeability', *J. Liq. Chromatogr.*, **6**, 1559–1576 (1983).
18. J. Janča, J. Chmelík, 'Focusing in Field-flow Fractionation', *Anal. Chem.*, **56**, 2481–2484 (1984).
19. J. Janča, J. Chmelík, D. Přibylová, 'Optimization of Field-flow Fractionation with Respect to Relaxation and Use of Stop Flow Technique at Constant Field Operation', *J. Liq. Chromatogr.*, **8**, 2343–2368 (1985).
20. V. Jáhnová, F. Matulík, J. Janča, 'Correction for Zone Spreading in Sedimentation Field-flow Fractionation', *Anal. Chem.*, **59**, 1039–1043 (1987).
21. A.E. Hamielec, 'Correction for Axial Dispersion', in *Steric Exclusion Chromatography of Polymers*, ed. J. Janča, M. Dekker, New York, 1984.
22. M.R. Schure, B.N. Barman, J.C. Giddings, 'Deconvolution of Nonequilibrium Band Broadening Effects for Accurate Particle Size Distributions by Sedimentation Field-flow Fractionation', *Anal. Chem.*, **61**, 2735–2743 (1989).
23. Y. Mori, K. Kimura, M. Tanigaki, 'Influence of Zone Broadening on Particle Size Analysis by Sedimentation Field-flow Fractionation', *Spec. Publ. – R. Soc. Chem.*, **102**, 290–299 (1992).

24. J.-C. Vauthier, P.S. Williams, 'Numerical Simulation of Band-broadening During Hydrodynamic Relaxation in Frit-inlet Field-flow Fractionation Channels', *J. Chromatogr.*, **805**, 149–166 (1998).
25. C. Quivoron, 'Use for Polymer Analysis', in *Steric Exclusion Chromatography of Polymers*, ed. J. Janča, M. Dekker, New York, 1984.
26. R. Beckett, 'The Application of Field-flow Fractionation to the Characterization of Complex Environmental Samples', *Environ. Technol. Lett.*, **8**, 339–354 (1987).
27. G.B. Levy, A. Fox, 'Sedimentation Field-flow Fractionation', *Am. Biotechnol. Lab.*, **6**, 14–21 (1988).
28. J. Janča, D. Přibylová, K. Bouchal, V. Tyráčková, E. Žůrková, 'Characterization of Polystyrene and Poly(styrene-glycidyl methacrylate) Copolymer Latexes by Sedimentation Field-flow Fractionation', *J. Liq. Chromatogr.*, **9**, 2059–2072 (1986).
29. T. Koch, J.C. Giddings, 'High-speed Separation of Large (>1 μm) Particles by Steric Field-flow Fractionation', *Anal. Chem.*, **58**, 994–997 (1986).
30. J. Li, K.D. Caldwell, W. Machtle, 'Particle Characterization in Centrifugal Fields: Comparison Between Ultracentrifugation and Sedimentation Field-flow Fractionation', *J. Chromatogr.*, **517**, 361–376 (1990).
31. J.C. Giddings, M.N. Myers, M.H. Moon, B.N. Barman, 'Particle Separation and Size Characterization by Sedimentation Field-flow Fractionation', in *Particle Size Distribution*, ed. T. Provder, ACS Symposium Series No. 472, ACS, Washington, DC, 1991.
32. J.J. Kirkland, W.W. Yau, 'Quantitative Particle-size Distributions by Sedimentation Field-flow Fractionation with Densimeter Detector', *J. Chromatogr.*, **550**, 799–809 (1991).
33. R. Beckett, 'Particle Characterization by Field-flow Fractionation', in *Particle Characterization and Its Applications*, ed. J.A. Raper, The University of New South Wales, Sydney, 1991.
34. R. Finsy, M. Van Laethem, N. De Jaeger, R. Sneyers, J. Vanderdeelen, P. Van der Meeren, H. Demeyere, E. Gelade, 'Particle Size Distribution of Bimodals: A Comparative Study of Submicron Sizing Instrumentation', in *Advances in Measurement and Control of Colloidal Processes*, eds. R.A. Williams, N. De Jaeger, Butterworth-Heinemann, Oxford, 1991.
35. M.R. Schure, 'Optimization of Channel Dimensions used in Sedimentation Field-flow Fractionation Through Computer Simulations', *Polym. Mat. Sci. Eng.*, **65**, 59–60 (1991).
36. E.C. Arner, J.J. Kirkland, 'Sedimentation Field Flow Fractionation', in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds. S.E. Harding, A.J. Rowe, J.C. Horton, Royal Society of Chemistry, Cambridge, 1992.
37. P.S. Williams, M.H. Moon, J.C. Giddings, 'Fast Separation and Characterization of Micron Size Particles by Sedimentation/Steric Field-flow Fractionation: Role of Lift Forces', in *Particle Size Analysis*, eds. N.G. Stanley-Wood, R.W. Lines, Royal Society of Chemistry, Cambridge, 1992.
38. B.N. Barman, J.C. Giddings, 'Separation and Characterization of Polymeric Latex Beads and Aggregates by Sedimentation Field-flow Fractionation', in *Chromatography of Polymers: Characterization by SEC and FFF*, ed. T. Provder, ACS Symposium Series 521, ACS, Washington, DC, 1993.
39. Y. Mori, 'Retention Behavior of Colloidal Dispersions in Sedimentation Field-flow Fractionation', *Adv. Colloid Interface Sci.*, **53**, 129–140 (1994).
40. A. Athanasopoulou, G. Karaiskakis, A. Travlos, 'Colloidal Interactions Studied by Sedimentation Field-flow Fractionation', *J. Liq. Chromatogr. & Rel. Tech.*, **20**, 2525–2541 (1997).
41. C. Contado, G. Blo, F. Fagioli, F. Dondi, R. Beckett, 'Characterization of River Po Particles by Sedimentation Field-flow Fractionation Coupled with GFAAS and ICP-MS', *Colloids Surf., A*, **120**, 47–59 (1997).
42. Y. Mori, 'Sedimentation Field-flow Fractionation of Polystyrene Latex Particles in Alcohol Solvent', *J. Microcolumn Sep.*, **10**, 57–63 (1998).
43. H.L. Lee, J.F.G. Reis, J. Dohner, E.N. Lightfoot, 'Single-phase Chromatography: Solute Retardation by Ultrafiltration and Electrophoresis', *AICHE J.*, **20**, 776–784 (1974).
44. J.C. Giddings, F.J. Yang, M.N. Myers, 'Flow Field-flow Fractionation: A Versatile New Separation Method', *Science*, **193**, 1244–1245 (1976).
45. J.F.G. Reis, E.N. Lightfoot, 'Electropolarization Chromatography', *AICHE J.*, **22**, 779–785 (1976).
46. J.C. Giddings, G.C. Lin, M.N. Myers, 'Fractionation and Size Analysis of Colloidal Silica by Flow-field-flow Fractionation', *J. Colloid. Interface Sci.*, **65**, 67–78 (1978).
47. K.-G. Wahlund, A. Litzen, 'Application of an Asymmetric Flow Field-flow Fractionation Channel to the Separation and Characterization of Proteins, Plasmids, Plasmid Fragments, Polysaccharides, and Unicellular Algae', *J. Chromatogr.*, **461**, 73–87 (1989).
48. J.J. Kirkland, C.H. Dilks, Jr, S.W. Rementer, W.W. Yau, 'Asymmetric Channel Flow Field-flow Fractionation with Exponential Force-field Programming', *J. Chromatogr.*, **593**, 339–355 (1992).
49. A. Litzen, J.K. Walter, H. Krischollek, K.-G. Wahlund, 'Separation and Quantitation of Monoclonal Antibody Aggregates by Asymmetrical Flow Field-flow Fractionation and Comparison to Gel Permeation Chromatography', *Anal. Biochem.*, **212**, 469–480 (1993).
50. C. Tank, M. Antonietti, 'Characterization of Water-soluble Polymers and Aqueous Colloids with Asymmetrical Flow Field-flow Fractionation', *Macromol. Chem., Phys.*, **197**, 2943–2959 (1996).
51. B. Wittgren, K.-G. Wahlund, 'Fast Molecular Mass and Size Characterization of Polysaccharides Using Asymmetrical Flow Field-flow Fractionation-multiangle

Light Scattering', *J. Chromatogr., A*, **760**, 205–218 (1997).
52. M.E. Schimpf, K.-G. Wahlund, 'Asymmetrical Flow Field-flow Fractionation as a Method to Study the Behavior of Humic Acids in Solution', *J. Microcolumn Sep.*, **9**, 535–543 (1997).
53. P.S. Williams, 'Design of an Asymmetrical Flow Field-flow Fractionation Channel for Uniform Channel Flow Velocity', *J. Microcolumn Sep.*, **9**, 459–467 (1997).
54. A. Carlshaf, J.A. Jonsson, 'Perturbations of the Retention Parameter due to Sample Overloading in Hollow-fiber Flow Field-flow Fractionation', *Sep. Sci. Technol.*, **28**, 1191–1201 (1993).
55. S.K. Ratanathanawongs, J.C. Giddings, 'Particle Size Analysis Using Flow Field-flow Fractionation', in *Chromatography of Polymers: Characterization by SEC and FFF*, ed. T. Provder, ACS Symposium Series 521, ACS, Washington, DC, 1993.
56. B. Wittgren, K.-G. Wahlund, H. Derand, B. Wesslen, 'Size Characterization of a Charged Amphiphilic Copolymer in Solutions of Different Salts and Salt Concentrations Using Flow Field-flow Fractionation', *Langmuir*, **12**, 5999–6005 (1996).
57. T. Pauck, H. Coelfen, 'Hydrodynamic Analysis of Macromolecular Conformation. A Comparative Study of Flow Field-flow Fractionation and Analytical Ultracentrifugation', *Anal. Chem.*, **70**, 3886–3891 (1998).
58. K.D. Caldwell, L.F. Kesner, M.N. Myers, J.C. Giddings, 'Electrical Field-flow Fractionation of Proteins', *Science*, **176**, 296–298 (1972).
59. J.M. Davis, F.R.F. Fan, J. Bard, 'Retention by Electrical Field-flow Fractionation of Anions in a New Apparatus with Annular Porous Glass Channels', *Anal. Chem.*, **59**, 1339–1348 (1987).
60. K.D. Caldwell, Y.-S. Gao, 'Electrical Field-flow Fractionation in Particle Separation. 1. Monodisperse Standards', *Anal. Chem.*, **65**, 1764–1772 (1993).
61. M.E. Schimpf, D.D. Russell, J.K. Lewis, 'Separation of Charged Latex Particles by Electrical Field-flow Fractionation', *J. Liq. Chromatogr.*, **17**, 3221–3228 (1994).
62. M.E. Schimpf, K.D. Caldwell, 'Electrical Field-flow Fractionation for Colloid and Particle Analysis', *Am. Lab.*, **27**, 64–68 (1995).
63. S.A. Palkar, R.E. Murphy, M.R. Schure, 'Charge/Hydrophobicity Fractionation with Hyphenated Electrical FFF/Chromatography', *Polym. Mater. Sci. Eng.*, **75**, 6–7 (1996).
64. M. Dunkel, N. Tri, R. Beckett, K.D. Caldwell, 'Electrical Field-flow Fractionation: A Tool for Characterization of Colloidal Adsorption Complexes', *J. Microcolumn Sep.*, **9**, 177–183 (1997).
65. G.H. Markx, J. Rousselet, R. Pethig, 'DEP-FFF: Field-flow Fractionation Using Non-uniform Electric Fields', *J. Liq. Chromatogr. & Rel. Tech.*, **20**, 2857–2872 (1997).
66. J.C. Giddings, M.N. Myers, J. Janča, 'Retention Characteristics of Various Polymers in Thermal Field-flow Fractionation', *J. Chromatogr.*, **186**, 37–44 (1979).
67. J.C. Giddings, L.K. Smith, M.N. Myers, 'Programmed Thermal Field-flow Fractionation', *Anal. Chem.*, **48**, 1587–1592 (1976).
68. J.C. Giddings, M. Martin, M.N. Myers, 'High-speed Polymer Separations by Thermal Field-flow Fractionation', *J. Chromatogr.*, **158**, 419–435 (1978).
69. G. Liu, J.C. Giddings, 'Separation of Particles in Non-aqueous Suspensions by Thermal-electrical Field-flow Fractionation', *Anal. Chem.*, **63**, 296–299 (1991).
70. G. Liu, J.C. Giddings, 'Separation of Particles in Aqueous Suspensions by Thermal Field-flow Fractionation. Measurement of Thermal Diffusion Coefficients', *Chromatographia*, **34**, 483–492 (1992).
71. P.M. Shiundu, J.C. Giddings, 'Influence of Bulk and Surface Composition on the Retention of Colloidal Particles in Thermal Field-flow Fractionation', *J. Chromatogr.*, **715**, 117–126 (1995).
72. J.C. Giddings, P.M. Shiundu, S.N. Semenov, 'Thermophoresis of Metal Particles in a Liquid', *J. Colloid Interface Sci.*, **176**, 454–458 (1995).
73. S.J. Jeon, M.E. Schimpf, 'Separation of Particles Based on Surface Composition by Thermal Field-flow Fractionation', *Polym. Mater. Sci. Eng.*, **75**, 4–5 (1996).
74. S.N. Semenov, 'Surface Phenomena in Thermal FFF of Particles', *J. Liq. Chromatogr. & Rel. Tech.*, **20**, 2669–2685 (1997).
75. S.J. Jeon, M.E. Schimpf, 'Thermal Field-flow Fractionation of Colloidal Particles', in *Particle Size Distribution III: Assessment and Characterization*, ed. T. Provder, ACS, Washington, DC, 1998.
76. J. Janča, R. Audebert, 'Experimental Study of Isopycnic Focusing Generated by Coupled Electric and Gravitational Field Force: Use in Thin Layer Focusing and Focusing Field-flow Fractionation', *Mikrochim. Acta*, **113**, 299–311 (1994).
77. J. Chmelík, M. Deml, J. Janča, 'Separation of Two Components of Horse Myoglobin by Isoelectric Focusing Field-flow Fractionation', *Anal. Chem.*, **61**, 912–914 (1989).
78. J. Janča, 'Flow Velocity Profile in Channels with Modulated Cross-sectional Permeability for Sedimentation-flotation Focusing Field-flow Fractionation Using Step Density Gradients', *J. Chromatogr.*, **404**, 23–32 (1987).
79. J. Janča, N. Nováková, 'Sedimentation-flotation Focusing Field-flow Fractionation Using Step Density Gradient', *J. Liq. Chromatogr.*, **10**, 2869–2876 (1987).
80. J. Janča, 'Elutriation Focusing Field-flow Fractionation', *Makromol. Chem., Rapid Commun.*, **8**, 233–236 (1987).
81. E. Urbánková, J. Janča, 'An Attempt at Experimental Elutriation Focusing Field-flow Fractionation', *J. Liq. Chromatogr.*, **13**, 1877–1895 (1990).

82. K. Chen, K.G. Wahlund, J.C. Giddings, 'Gravity-augmented High-speed Flow/Steric Field-flow Fractionation: Simultaneous Use of Two Fields', *Anal. Chem.*, **60**, 362–364 (1988).
83. J.C. Giddings, S. Li, P.S. Williams, M.E. Schimpf, 'High Speed Separation of Ultra-high-molecular Weight Polymers by Thermal/Hyperlayer Field-flow Fractionation', *Makromol. Chem., Rapid Commun.*, **9**, 817–823 (1988).
84. J. Janča, M. Martin, 'Influence of Operational Parameters on Retention of Ultra-high Molecular Weight Polystyrenes in Thermal Field-flow Fractionation', *Chromatographia*, **34**, 125–131 (1992).
85. S.N. Semyonov, A.A. Kuznetsov, P.P. Zolotaryov, 'Theoretical Examination of Focusing Field-flow Fractionation', *J. Chromatogr.*, **364**, 389–395 (1986).
86. J. Janča, 'Isoperichoric Focusing Field-flow Fractionation for Characterization of Particles and Macromolecules', *J. Liq. Chromatogr. & Rel. Tech.*, **20**, 2555–2577 (1997).

Filtration in Particle Size Analysis

Ian G. Droppo
National Water Research Institute, Burlington, Canada

1 Introduction **1**
2 Definitions **3**
2.1 Particle 3
2.2 Particle Size 3
2.3 Filtration 5
3 Equipment **6**
3.1 Filter Types 6
3.2 Filtration Apparatus 7
3.3 Sequential (Cascade) Filtration Apparatus 8
3.4 Plankton Chamber 9
4 Particle Sizing by Filtration **9**
4.1 Effect of Physical Aspects of Filter 10
4.2 Effect of Nature of the Particles 10
4.3 Effect of Procedure Used for Filtration 11
5 Quality Assurance of Particle Separation Results **14**
6 Alternative Particle Sizing by Optical Image Analysis **14**
Disclaimer **14**
Acknowledgments **14**
Related Articles **15**
References **15**

Filtration is defined as the separation of solid particles from a liquid or gas by passing the suspension through a porous, fibrous or granular substance. In science and industry, filtration is commonly used for the concentrating and/or removal of particulate matter or for the separation of particles into size classes for subsequent physical, chemical or biological analysis. Separation of particles into multiple size classes is often required for the assessment of particle size. There are many different types of filters available (e.g. fibrous, membrane, polycarbonate) for particle characterization through filtration. The type of filter to be used for particle size characterization is dependent on a number of considerations such as sample (particle) characteristics, collection efficiency, pore size and structure and analytical compatibility. Filtration has the advantage of being one of the few techniques which allow for the physical separation of particles into different size classes for distribution or chemical–biological analysis. If appropriate precautions are taken to minimize sample perturbations (e.g. select appropriate filter for the application, filter in the field, avoid overloading a filter, ensure a low flow rate and stir if possible), a reasonably accurate estimate of particle size can be obtained. Generally, errors associated with filtration arise from filter clogging and when attempting to characterize flocculated/aggregated particles. The basic nature of a suspended aggregate (i.e. fragile, plastic, high water content) means that it is likely to be broken up or created owing to the often tortuous nature of filtration, or be deformed on the surface or within the filter owing to pressure effects. This has implications if optical image analysis of the filter is to be used for particle size analysis as the true particle morphology and size will have been altered. In such an instance, if only a grain size distribution is required, it is suggested that optical image analysis be employed in conjunction with a plankton chamber settling technique rather than with filtration.

1 INTRODUCTION

Particles, and their interactions with their surrounding environments, have both theoretical and practical relevance within science and industry. Physical, chemical and biological processes are at play with or within particles and as such are known to play an important role in the biological and chemical dynamics of any aquatic and terrestrial environment and within engineered systems such as wastewater treatment. The definition of a ''particle'' in science is wide and varied depending on the discipline, and may incorporate a wide range of sizes, compositions and structures. Particles are considered to be dynamic and, in most cases, unstable relative to their surrounding environment, resulting in the majority of cohesive particles ($<63\,\mu m$) existing primarily as flocculated or aggregated particles[1,2] with a high propensity to assimilate and transport various nutrients and contaminants.[3] As such, an accurate determination of particulate characteristics is of vital importance when dealing with issues of sediment and contaminant interactions. Common physical measurements of particle characteristics important to industry and the environment include grain size, surface area, specific gravity, surface charge, bulk density, sheer stress, porosity and permeability.[4] Grain size is the most significant of these characteristics in terms of a particle's capacity to concentrate and retain contaminants owing to its controlling influence over many of the above characteristics.[4–11] Hence, if only one physical characteristic of a particulate matter is to be determined for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

industrial or environmental purposes as related to contaminants, grain size is by far the characteristic of choice.

Investigations into the chemically active fraction (<63 μm[12]), the effects of particle size on the adsorption of contaminants[11,13,14] and the association of microbes with particles,[15] among others, often require that (a) the sediment be separated into a number of size fractions for subsequent analysis and/or (b) an accurate grain size distribution be obtained of the original sample. Such information and the associated chemical and/or biological analyses are often driven by environmental concerns, process control and monitoring demands and the development of new materials requiring evaluation. Accurate monitoring of particulate matter requires a knowledge of particle size distributions to assess fully health hazards, environmental conditions and industrial process efficiencies/deficiencies or performance. Particle fractionation and distributions are often used within models for sediment and contaminant source, fate and effect predictions in support of, or development of, environmental policies and management of water resources including wastewater treatment.

While there are a number of reviews of direct and indirect techniques for the fractionation of particulate matter into grain size ranges or for the determination of a particle size distribution,[4,16–20] there is no standard technique, and as such, each method can produce significantly different results (this is particularly true when dealing with particles in aggregate form, as the majority of methods do not maintain the integrity of the particles). Grain size analysis is therefore an operationally defined property.

Direct optical image analysis with plankton chambers or filters provides the least destructive (to aggregates), most accurate, direct method of determining grain size distributions. However, if particle separation is required there are four other categories of techniques which allow for the separation of particles into size classes with a minimum of particle perturbation:[20]

1. sieving, filtration, ultrafiltration and dialysis
2. sedimentation and centrifugation
3. field flow fractionation
4. hydrodynamic chromatography.

To date, categories 1 and 2 have been the most widely used methods for particle fractionation. Within these two categories, filtration, in spite of its drawbacks, is probably the most widely used, as its operational size range spans from the colloidal fraction up to and past what is considered the chemically active fraction (Figure 1). In addition, the chemical conditions in solution can be well controlled and particle size separation is generally well defined given appropriate controls. The primary chemical drawbacks are (a) possible contamination by metals and organic matter from the filter materials, (b) adsorption of ''dissolved'' compounds at the filter surface and (c) reactions induced by the filter, such as

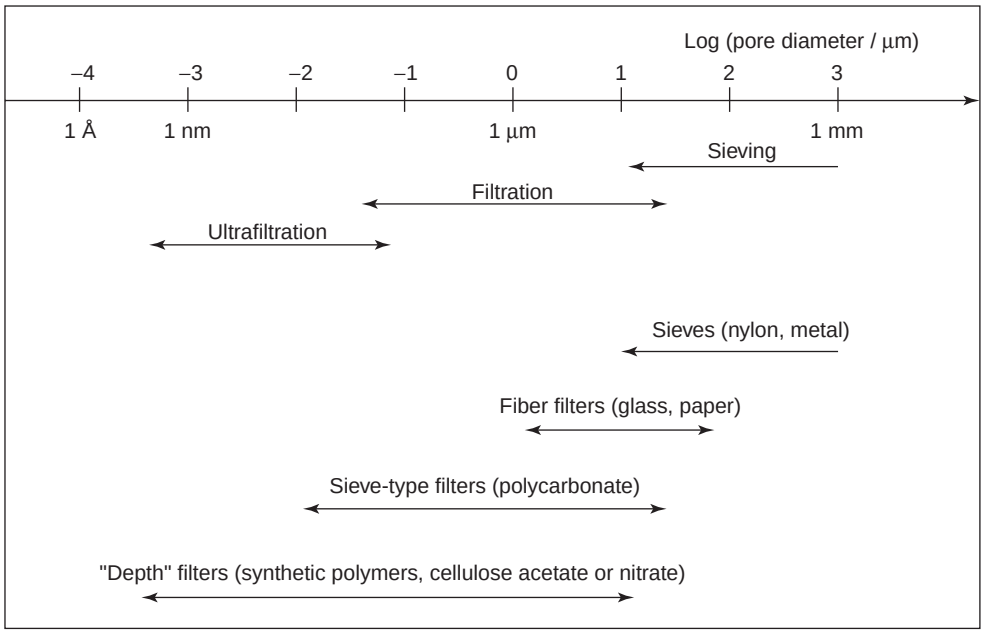

Figure 1 Size ranges for separation techniques and for the general filter types. (Modified with permission from J. Buffle, D. Perret, M. Newman, 'The Use of Filtration and Ultrafiltration for Size Fractionation of Aquatic Particles, Colloids and Macromolecules', in *Environmental Particles*, eds. J. Buffle, H. van Leeuwen, Lewis Publishers, Chelsea, MI, 1992. Copyright, CRC Press, Boca Raton, FL.)

aggregation of adsorbed natural organic material and adsorption of inorganic ions by particles retained on or within the filter.[21] An understanding of the chemical interaction between particles, solute and filters is an important analytical requirement for any chemical or biological focused study. This relationship is beyond the scope of this article and only the physical aspects will be addressed. The primary physical drawbacks are related to the particle (aggregate) morphological changes and filter clogging resulting in a poorly defined size threshold. While sieving can be used for the fractionation of larger particles (most often used in the soil sciences) and ultrafiltration can be used for the fractionation of colloidal particles (although with great cost and time consumption), this article will focus primarily on the use of pressure or vacuum filtration in particle sizing and the appropriate size ranges associated with this method (Figure 1). In addition, the use of filters or plankton chambers with image analysis will also be discussed as a viable way of obtaining a grain size distribution. Although filtration is widely used for the filtration of aerosols from gases,[22,23] this article will focus specifically on the sizing of particles from natural or engineered aqueous suspensions.

2 DEFINITIONS

2.1 Particle

It is universally accepted, and fundamental to science, that all matter is composed of particles. Figure 2 provides a breakdown of what in environmental science generally is considered a particle. As with most definitions within science, however, the definition of a ''particle'' is operationally defined. Its definition in terms of composition, size range and structure is dependent on the objectives of the investigation, the process and environment under study and the operational restriction around the available tools or techniques for analysis. While there is no standard definition of a particle, the general standard size cutoff as to what would be considered a particle versus the dissolved phase is arbitrarily set at 0.45 μm.[11,21]

The difficulty in providing a standard definition of a particle is further complicated by the knowledge that particles rarely exist in a homogeneous state, with the exception of larger particles such as quartz (even this is debatable depending on the scale of observation, as such particles are likely to have bacterial, organic or ionic associations or coating of some sort) or smaller organisms such as bacteria (these too can often be incorporated within aggregated particles). These aggregates and flocs generally consist of an interrelated network of organic and inorganic particles resulting in complex physical, chemical and biological interactions not only within themselves but also with the medium in which they exist as a whole. These complex interactions, which will affect aggregate or floc function and structure, are illustrated in the conceptual model of floc form and function provided in Figure 3. A common problem within analytical chemistry is that all too often particle evaluations are performed on the constituent or absolute particles of the aggregates and not on the aggregates themselves. Such disregard for the integrity of the true particle structure and distribution can lead to erroneous results in evaluation or implementation of such important issues as environmental management/policy, health risk assessment and industrial performance.

Given that environmental analytical chemistry will focus mostly on the chemically active fraction (<63 μm), and that this fraction exists extensively as composite particles,[1,2] this article will focus only on the sizing and fractionation of individual particles and flocculated or aggregated particles and their constituent particles down to a size of 0.45 μm. Therefore, for the purposes of this article, the term ''particle'' will be limited by structure and size and be defined as follows:

> any physical entity or combination (aggregate/floc) of entities (inorganic or organic, viable or non-viable) within the solid phase having a minimum diameter within the range of 0.45 μm to the aggregate/floc size that a given environment will support (to a maximum 1000 μm).

This article will focus mostly on particles immersed or suspended within natural waters or within hydrous-engineered media and will not include soil-type aggregates or aerosols. The colloidal fraction will not be addressed here because of the extremely different physicochemical characteristics of these particles.[24,25] A complete discussion on colloidal characterization can be found in Leppard and Buffle.[26] The size range stated above is that which can be easily viewed by standard optical microscopy. For the purposes of this article, a floc is defined as a particle which has resulted from aggregation within an aquatic medium (suspended or bed sediment) and an aggregate is a particle which has resulted from aggregation outside an aquatic environment (e.g. soil particle). Aggregates can be observed and transported within the aquatic environment owing to their delivery to the system by land-surface processes (e.g. stormwater runoff). Flocculation and aggregation will be used interchangeably within this article.

2.2 Particle Size

While the definition of a particle above is based in part on size, a brief discussion of what defines the size of a particle is required based on the methods to be evaluated in this article.

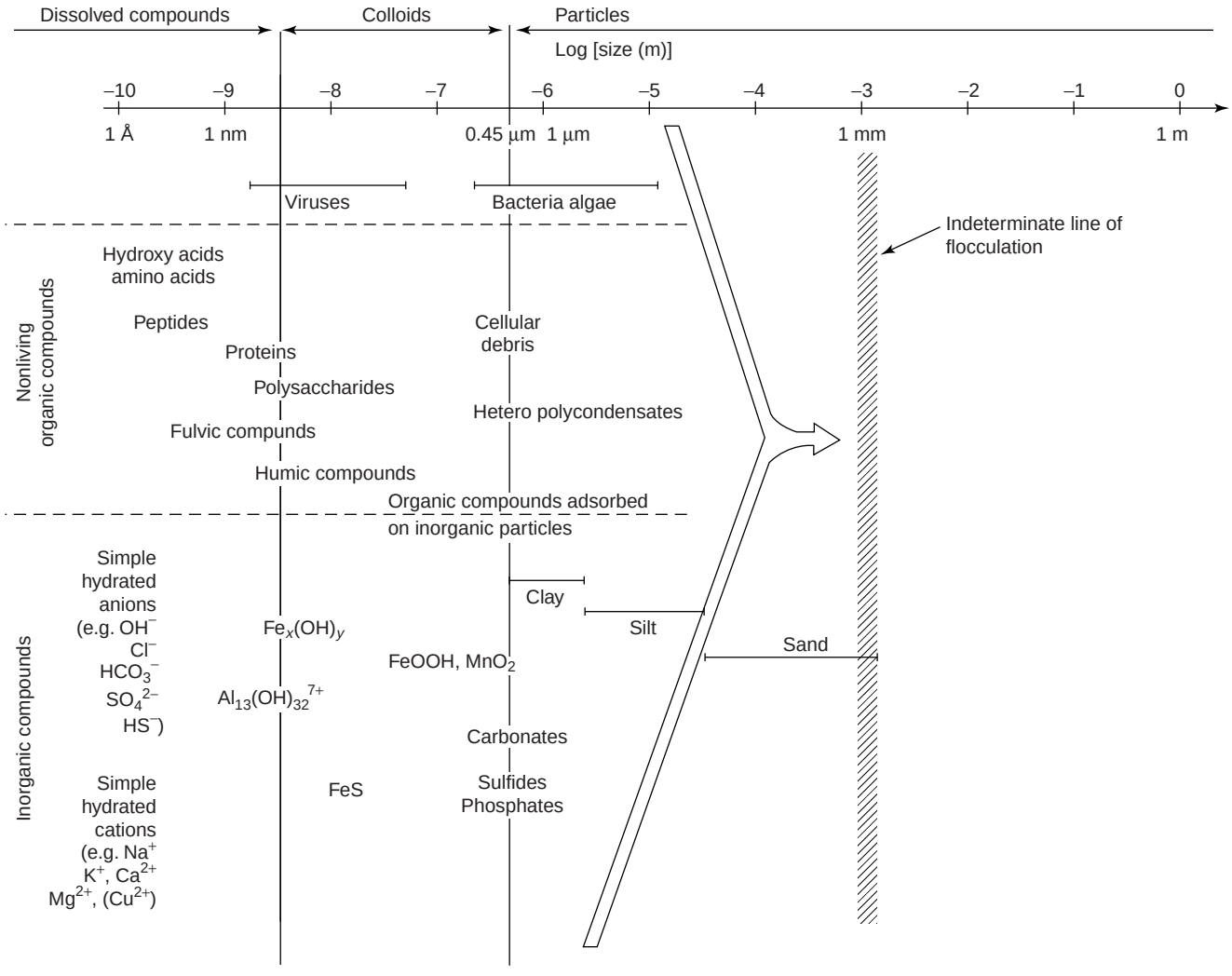

Figure 2 Schematic classification of what environmental science generally considers as dissolved, colloidal and particulate compounds as defined by size and organic and inorganic components. All of the components to the left of the flocculation wedge can be incorporated into flocculated or aggregated particles with a subsequent increase in effective size. No upper size range for floc size can be determined as it is dependent on a number of physical, chemical and biological factors, although marine snow has been observed in the order of centimeters. (Modified with permission from J. Buffle, D. Perret, M. Newman, 'The Use of Filtration and Ultrafiltration for Size Fractionation of Aquatic Particles, Colloids and Macromolecules', in *Environmental Particles*, eds. J. Buffle, H. van Leeuwen, Lewis Publishers, Chelsea, MI, 1992. Copyright, CRC Press, Boca Raton, FL.)

Convention dictates that most methods of particle sizing provide results in terms of particle diameter (or a range of possible diameters). In nature and in most industrial processes, particles are, however, not geometrically perfect spheres, squares, or the like, but rather constitute irregularly shaped particles. This is particularly true for flocs where the particles are not only irregular in shape but also in composition and density. Such complexity requires that an equivalent diameter be measured or derived from a given technique. Allen[27] provides 13 different measures of particle diameter which are based on one or all of the following physical characteristics of (1) the particle (e.g. volume, surface area), (2) the suspension (density and viscosity of the fluid relative to the particle during settling) and (3) the filter (pore size). As this article focuses only on filtration and optical size determination, particle size will be defined as follows:

Filter Particle Diameter: the width of the minimum aperture through which a particle will pass. In reality, however, filtration only provides a particle size range and not an estimate of discrete particle size (unless assessed optically under a microscope). A given filter will, therefore in theory, retain all particles on or within a filter whose particles are larger than the nominal or average pore size of the filter.

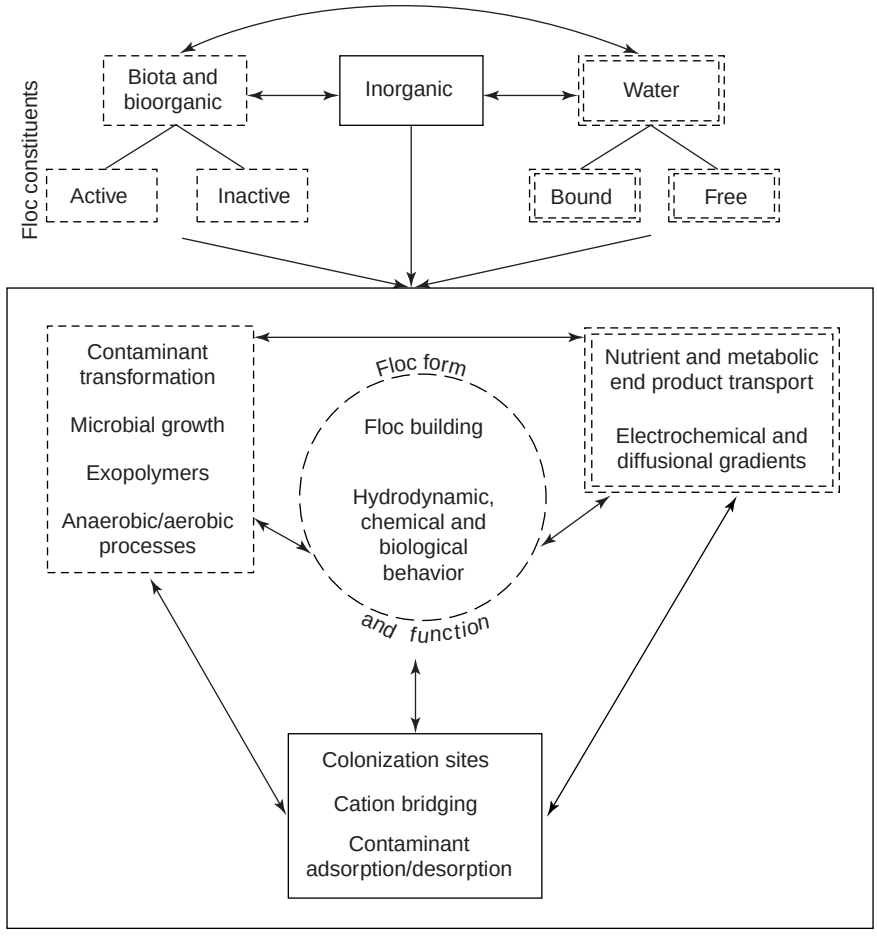

Figure 3 Conceptual model of floc form and function. (Reproduced with kind permission of Kluwer Academic Publishers from I.G. Droppo, G.G. Leppard, D.T. Flannigan, S.N. Liss, 'The Freshwater Floc: a Functional Relationship of Water and Organic and Inorganic Floc Constituents Affecting Suspended Sediment Properties', *Water Air Soil Pollut.*, **99**, 43–53 (1997).)

Optically Derived Equivalent Spherical Diameter: optical size determination of sediment sizes retained on a filter or in a plankton chamber is represented as a distribution of equivalent diameters. That is, the diameter of a sphere whose cross-sectional area is equal to the projected area of the particle.

2.3 Filtration

Filtration can be defined as the separation of solid particles from a liquid or gas by passing the mixture through a porous, fibrous or granular substance. In science and most industrial processes, filtration is most commonly used for the concentrating of particulate matter or for the separation of particles into size classes through sequential or cascade filtration for subsequent physical, chemical or biological analysis. Separation of particles into multiple size classes is also commonly used for the determination of grain size distributions (derived through a mass-balance approach). There are several mechanisms by which particles are removed from a given medium during filtration.[23,28–30] These are as follows:

Interception: occurs when a particle moving within a flow streamline comes in contact with the surface of the filter. Interception is most important when the ratio of particle diameter to the pore size is relatively large. The importance of this mechanism increases as filter clogging or caking occurs. Modeling efforts[31] have shown that for particles larger than 1 μm, interception is the main mechanism of particle removal suggesting that retention is a function of filter geometry and particle size and is independent of particle density, diffusivity, particle settling velocity and fluid velocity.

Inertial Collection: occurs when the flow streamline bends around a filter fiber or within a pore, but the inertia of the particle keeps it on its original track so that it impacts the filter surface. This mechanism is most important when the flow velocities are high and when fiber filters are employed.

Brownian Diffusion: occurs primarily with colloidal particles (<0.5 μm). There is an inverse relationship between the importance of this mechanism and particle size. Particles follow the flow stream to the surface of the filter where the concentration is initially zero. This mechanism works best with low flow rates as the time spent by a colloidal particle in the vicinity of the filter surface increases, resulting in enhanced diffusional removal.

Electrostatic Deposition: occurs through electrochemical attractions between the filter and the particle. This can be important for colloidal particles; however, it is generally neglected unless the particles or filters have been charged in some quantifiable way.

Gravitational Settling: generally gravitational settling is insignificant during filtration in aqueous media as the ratio of particle terminal settling velocity to the free stream velocity is small. This mechanism will only be important for very large particles which are above the range discussed in this article.

These mechanisms (excluding electrostatic deposition and gravitational settling) have been reviewed as related to modeling of filtration processes by Logan[30] and Logan et al.[31] Filtration models can be used to determine whether samples should be filtered in series or in parallel and to determine the appropriate filter types and pore sizes for different applications related to particle sampling and sizing. The actual mathematical modeling of filtration is beyond the scope of this article.

3 EQUIPMENT

3.1 Filter Types

There are numerous reviews of filter types, structures, compositions and applications.[29,31–35] Filters consist of either tightly woven fibrous mats of natural or synthetic materials or plastic porous membranes. Consideration as to what type of filter to use for a specific application in particle characterization requires a knowledge of a number of factors including cost, availability, collection efficiency, compatibility with the analytical and sampling procedures, compatibility of the filter pore size with the sample particle size range, the ability to remove particles for analysis following filtration, chemical inertness and sterilization considerations. Many filters (particularly membrane), depending on their composition and the digestion procedure used, can be digested along with the sediment for chemical analysis. If the filtrate is to be analyzed for the ''dissolved'' phase, care must be taken to choose the best filter type as this can have a profound effect on the filtrate chemical characteristics.[36] Filters come in a number of different sizes and formats. Cartridge filters, commonly used in industry, will not be discussed here as they are not appropriate for the sizing of particles. Commonly used filter media are listed below, with the general physical characteristics of each filter type given in Table 1 and environmental scanning electron micrographs of the surfaces of these filter types in Figure 4(a–d). For the purposes of this article, all separation media described will be referred to as filters although many can and are also called screens, nets or sieves.

Screens, nets or sieves are generally used for larger size particle separation and are basically a weave of usually nylon fibers with pore shape dictated by the nature of the weave (e.g. Nitex™). A number of different weaves are available, the square weave (square pores) being one of the most common.[37]

Fibrous filters (depth filters) are tightly woven mats of materials such as glass, cellulose (paper), quartz, silk, polyester, rayon or plastic fibers. These filters are relatively inexpensive, widely available (e.g. Whatman™ paper filters) and have high flow rates and high collection efficiencies for a wide range of particle sizes, with the cellulose filters appropriate for high-temperature ashing. They are not appropriate if the retained sediment is required to be removed from the filter prior to analysis, particularly if the sample is dominated by cohesive sediments. This is primarily due to the entrapment of sediments within the filter mat and the large surface roughness relative to the diameter of the particles. These filters are primarily used for clarification and removal of large particles. Often depth filters are used as prefilters in order to reduce clogging of membrane filters. Chemical inertness is questionable with regard to glass and quartz fiber filters as metal ions are strongly adsorbed, while the affinity of organics to quartz fibers can result in an overestimation of organic carbon concentration and mass.[21,23,35]

Table 1 General physical characteristics of filter types

Filter type	Fiber diameter (μm)	Pore diameter (μm)	Porosity (%)	Filter thickness
Fibrous	0.1–100	<1–100	60–>90	0.15–0.5 mm
Screen	28–1000	10–5500	2–72	1.9–0.045 mm
Membrane	N/A	0.01–10	<85	0.05–0.2 mm
Polycarbonate	N/A	0.1–8	5–10	6–10 μm

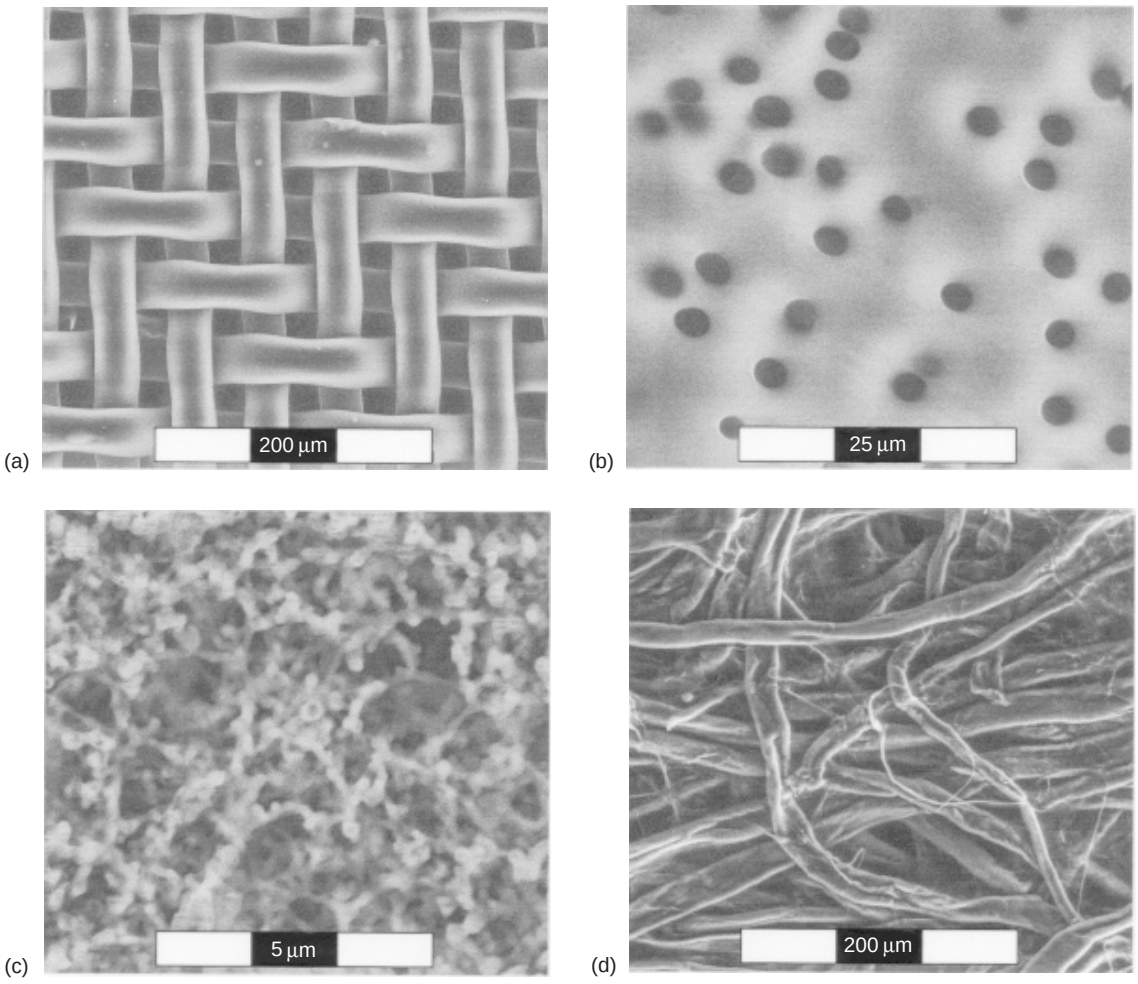

Figure 4 Micrographs of general filter-type surfaces: (a) woven mesh; (b) Nuclepore™ filters (polycarbonate); (c) membrane filters (cellulose nitrate or acetate); and (d) fiber filters (e.g. glass, paper fibers).

Membrane filters are porous membranes, such as Millipore™, composed of cellulose nitrate and cellulose triacetate (less common compositions are PVC, nylon, Teflon™ and silver). Teflon™ should be used for chemical studies as it is the most chemically inert although it cannot be used for analyzing organic species. Filtration on membrane filters results in the particles concentrating at or close to the surface of the filter. Unlike the fibrous filters, a membrane filter's loading capacity is limited. High pressure gradients are possible with resultant particle deformation.[21,23,35]

Polycarbonate pore filters (Nuclepore™) are smooth, sheet-like filters with cylindrical pores perpendicular to the surface of the filter.[21,23,35] Particles are collected at the surface of these filters, making them appropriate for particle wash-off for subsequent analysis (provided enough sample volume can be obtained) or microscopic observations and sizing of particles in the least disturbed state (albeit some particle flattening may occur of non-rigid particles such as flocs). These filters have a much smaller total void volume (porosity) and hence a lower flow rate than the other types of filters. Like other membrane filters, Nuclepore™ filters have a small load capacity before clogging occurs and high pressure gradients can result in particle deformation.[21,23,35]

3.2 Filtration Apparatus

While filtration apparatus is available in a number of different formats, generally at a minimal cost relative to other particle separation and sizing methods, to support a number of different filter types and sizes, they all generally consist of standard components. The most commonly used apparatus is a Millipore™-type filtration funnel that supports a 47-mm or 25-mm diameter filter. While larger apparatus are available to support filter sizes up to 30 cm in diameter, these types of apparatus are generally used for particle removal and solute analysis rather than for particle size analysis.[38] The type of apparatus used is

dependent on the requirements of the study and may be dictated by such factors as volume to be filtered, chemical requirements [stainless steel or plastic (generally polyethylene)], biological requirements (autoclavable or sealed units) and pressure restrictions. Often samples will be pretreated by processing the sample through a sieve (often 100 mesh, 0.15 mm) to define the particle size upper limit for filtration. A filtration system generally consists of the following components as illustrated in Figure 5 (only vacuum filtration is illustrated):

1. *funnel*: reservoir above the filter to hold the sample to be filtered, to allow for measurement of volume filtered and to provide a positive seal around the perimeter of the filter (can be plastic, Teflon™, glass or stainless steel);
2. *filter holder*: to hold the filter securely and to provide a positive seal around the perimeter of the filter (can be plastic, Teflon™, stainless steel or fritted glass);
3. *filtrate collection flask*: composition irrelevant for particle sizing; however, if the solute is to be chemically analyzed then consideration needs to be given to the makeup of the flask;
4. *source of vacuum/pressure*: electric, hand pump or water flow suction;
5. *supporting materials*: sieve for upper size class definition (composition important if chemical characterization is planned), spring clamp, stoppers, filtration hosing, filter tweezers (generally Teflon™ tipped), squirt bottle and overflow protection flask (optional).

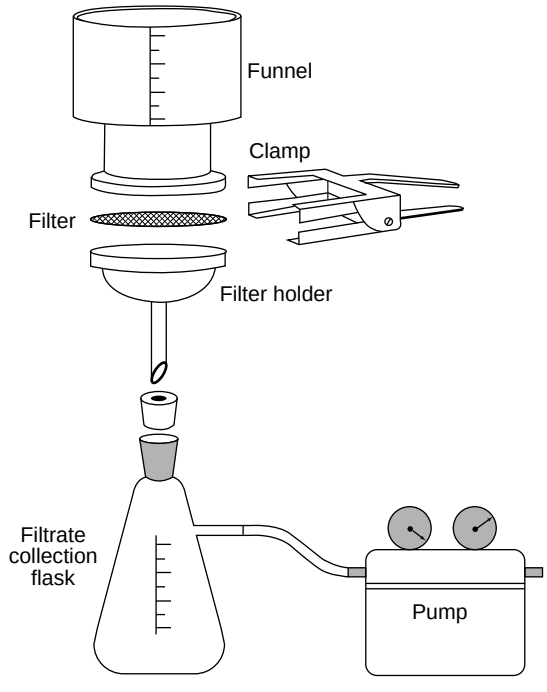

Figure 5 Generic filtration apparatus components. Note that component size, shape, material and configuration will vary.

3.3 Sequential (Cascade) Filtration Apparatus

As the term implies, sequential filtration is simply the method whereby the sample is passed through a series of filters in descending order of nominal pore size. This process may be performed in a discontinuous fashion (i.e. one filter at a time) or in a continuous fashion with the filters stacked. Such methodologies have been described by Laxen and Harrison,[39] Tessier et al.,[40] de Mora and Harrison,[41] Rao and Kwan,[42] Buffle et al.[20] and Droppo et al.[43] and are depicted in Figure 6 (pore sizes indicated from Droppo et al.)[43] and would be considered to be filtration in series. This method may be modified by performing particle separation in parallel, that is, splitting a sample into the number of filters used, running a sample through its respective filter, and then determining grain size distributions by subtracting cumulative masses from consecutive filter retained masses (in descending order).

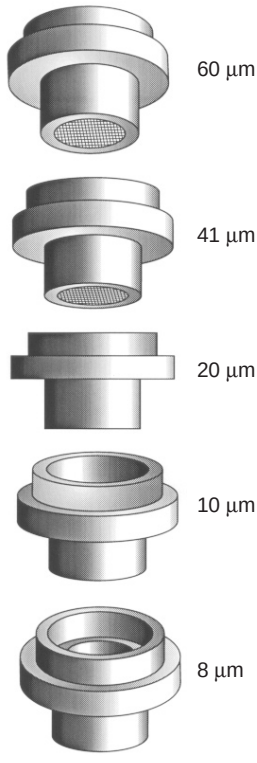

Figure 6 Generalized sequential filtration set-up. (Reproduced by permission of John Wiley & Sons Inc. from S.S. Rao, K.K. Kwan, 'Method for Measuring Toxicity of Suspended Particles in Waters', *Toxicity Assess.*, **5**, 91–101 (1990). Copyright © 1990, John Wiley & Sons Inc.)

3.4 Plankton Chamber

If the investigation is concerned only with obtaining a particle size distribution which is representative of the ''true'' size of the particles within a natural or engineered system, then the use of plankton chambers in conjunction with image analysis is recommended over filtration.

The plankton chamber (Hydro-Bios Kiel™) consists of four components (Figure 7): (a) a 3 mL reservoir with removable circular microscope slides in which the specimen comes to rest after settling, (b) a column available in various volumes (10, 25, 50 and 100 mL), (c) a top cap used to hydrostatically seal the plankton chamber and (d) a square piece of glass used to cover the reservoir after the column has been removed. In this technique the particulate sample may be introduced to the column section of the plankton chamber, either by a wide-mouth pipette (>3 mm) or the suspension can be sampled directly within the plankton chamber set up as described by Droppo and Ongley.[45] In general, within the field the appropriate column (Figure 7) volume is placed under the water and parallel to the direction of flow. The column is then capped under water and inverted upright for floc settling and subsequent observations such as that illustrated in the upper part of Figure 7. The size (volume) of the column used is dependent on the ambient suspended solids concentration (Droppo and Ongley,[45] provide a curve relating recommended column volumes to ambient concentration). Once sampled, the particles/aggregates/flocs settle under gravity until they come to rest on the microscope slide portion of the plankton chamber. Once at rest, the sample can be sized using an inverted microscope interfaced with any image analysis system capable of measuring particle size distributions. (There are many commercially available computer image analysis systems on the market spanning a wide range of price. These systems will not be discussed here.) Problems associated with the use of plankton chambers is that they are relatively expensive compared with standard filtration methods and the technique is concentration limited (also true for the optical analysis of particles on filters).[1,45]

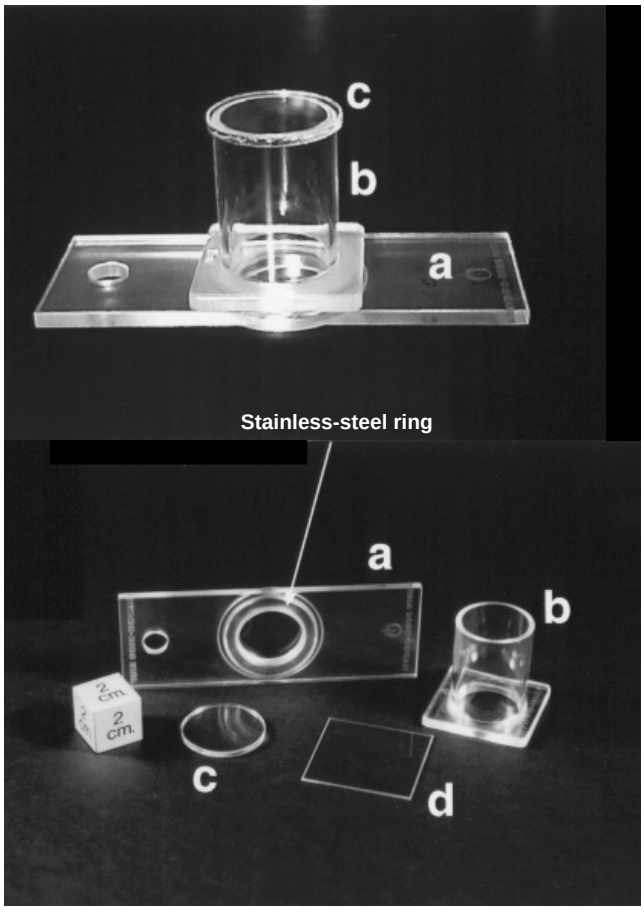

Figure 7 Plankton chamber components: (a) 3-mL reservoir; (b) column; (c) top cap; and (d) square glass plate. (Reproduced by permission of the American Society of Microbiology from Droppo et al.[44])

4 PARTICLE SIZING BY FILTRATION

For the effective separation of particles into size classes by filtration, three criteria must be met:[41]

1. the stated filter nominal porosity must approximate the effective or experimentally observed porosity;
2. the filter pores should have a relatively uniform size;
3. the effective pore size should remain relatively constant throughout the filtration procedure.

Very rarely, if at all, are all of these criteria met during filtration. Several studies have concluded that cellulose acetate membrane filters (Millipore™), glass-fiber filters and silver filters do not satisfy the above requirements.[20,46,47] Generally, it is found that Nuclepore™ filters provide the best separation, although there are still material and methodological problems, as discussed below. In general, sizing of particles by filtration is poor in relation to the nominal pore size specifications. The poor performance of filters used in experimental settings has been reported by Logan[30] and Droppo et al.[43] Logan[30] has shown that as many as 8% of the number of particles and 50% of the mass can be retained on a 210 μm (pore diameter) polypropylene filter even though all of the particles were <100 μm. Likewise, Droppo et al.[43] have shown, for sequential filtration, errors as high as 60% of the sediment (by volume) isolated by filtration being below the targeted nominal pore size of the filters (60–10 μm nylon mesh). As such, the stated

nominal pore size cannot be used as a definitive cutoff for targeted size classes; however, further experimentation or modeling[30] may provide the user with an acceptable pore size and filter type for a given targeted size range.

As implied above, of all the methods of particle sizing, filtration is likely to be one of the most problematic and potentially inaccurate methods and yet if precautions are taken, and potential errors are realized, it can also be one of the most useful as it is one of the few methods in which particles can be physically separated into designated size classes for subsequent analysis. In addition, it lends itself to optical applications which can be very accurate and provide visual confirmation of particle size, composition and texture. The problems around filtration are based primarily on (a) the physical aspects of the filters, (b) the nature of the particles to be separated and (c) the procedure used for filtration. These problems and possible error controlling measures are discussed below.

4.1 Effect of Physical Aspects of Filter

The efficiency by which various filter types separate particles into targeted size classes based on their nominal pore size has most often been assessed with retention curves (Figure 8a–d). The size range covered by the plots represents the width of the pore size distribution, while the median pore size (the effective pore size) is represented by the 50% value. The steeper the slope of the line, the more selective the filter's performance is. Figure 8(a–d) suggests that Millipore™ and Whatman™ glass-fiber filters have the greatest selectivity with Nuclepore™ and Flotronics™ filters having poorer selectivity. Figure 9 shows that there are great differences between the effective pore size and the nominal pore size of the various filters as stated by the manufacturers. The Flotronics™ and Nuclepore™ membrane filters show a reasonably close 1:1 relationship (i.e. sharpest separation in the median pore size) whereas the Millipore™ (cellulose) filters showed similar effective pore sizes irrespective of the nominal pore size (i.e. poor separation at the median pore size) (the Whatman™ glass-fiber filters had effective pore sizes all below 1.0 μm). Given these results, glass-fiber and cellulose membrane filters are not recommended for separating particles by size.[30,46–48] This is primarily because fiber filters have a wide array of pore sizes and tortuous flow paths which result in the trapping of particles within the filter bed. Because Nuclepore™ filters act more like sieves, these appear to be the best filters for particle fractionation and grain size distribution analysis below 8 μm.[20,33,49] Nylon mesh such as Nitex™ can also be effective for larger particle size separation (see Table 1), although errors were also found by Droppo et al.[43] for these and Nuclepore™ filters. However, it should be realized, given that most retention curves for calibration purposes are based on measurements made with particle size standards or beads of known size, shape and density (to avoid the issue of aggregation), or on low-concentration highly defined natural samples, it is unlikely that the retention curves in Figure 8(a–d) bear any resemblance to a filter's performance when environmental/industrial flocculated cohesive particles are filtered. This is primarily related to increased pore clogging by such particles[20,30,43] (see below).

Filter clogging[20,30,43,49,50] is primarily a function of the nature of the filter (specifically material and pore structure) and of the nature of the particle (see below) and is probably the most significant problem associated with filtration in causing the greatest associated errors in particle sizing. As a filter becomes overloaded, fiber width increases and pore size decreases. For polycarbonate filters, clogging is initiated through surface coagulation and pore clogging by aggregates larger than the pore size. This has the effect of reducing the flow rate and increasing the removal efficiency by widening the band of particle selectivity (i.e. smaller particles are trapped). This problem can be to the extreme as "dissolved" and colloidal materials or elements can be removed by filtration.[48] Polycarbonate filters will clog faster than a cellulose membrane filter owing to their more defined singular pore structure. The problems of filter clogging can be avoided by stopping filtration at signs of reduced flow rate, by not reusing filters, using a larger diameter filter (larger surface area per unit volume filtered), by minimizing the volume of sample fractionated per filter, by diluting the sample to minimize particle–particle interactions (if chemically acceptable), by insuring a low flow rate and, if instrumentation allows, by keeping the sample above the filter stirred. (If bulk removal is the goal of the procedure, then filter clogging is not an issue.)

4.2 Effect of Nature of the Particles

Different particles will have different sticking coefficients[30] and the higher the coefficient the greater is the potential for surface coagulation and pore and fiber clogging. The importance of the sticking coefficient relative to filtration will vary depending on the filter material, the particle type (cohesive or non-cohesive) and solution characteristics such as ionic and organic concentrations.[30] The influence of particle type on filter performance is compounded by the majority of cohesive sediment generally existing in a flocculated state.[1,2] The rheology of flocculated sediment has been described as plastic, and, as such, flocs will deform (flatten) on

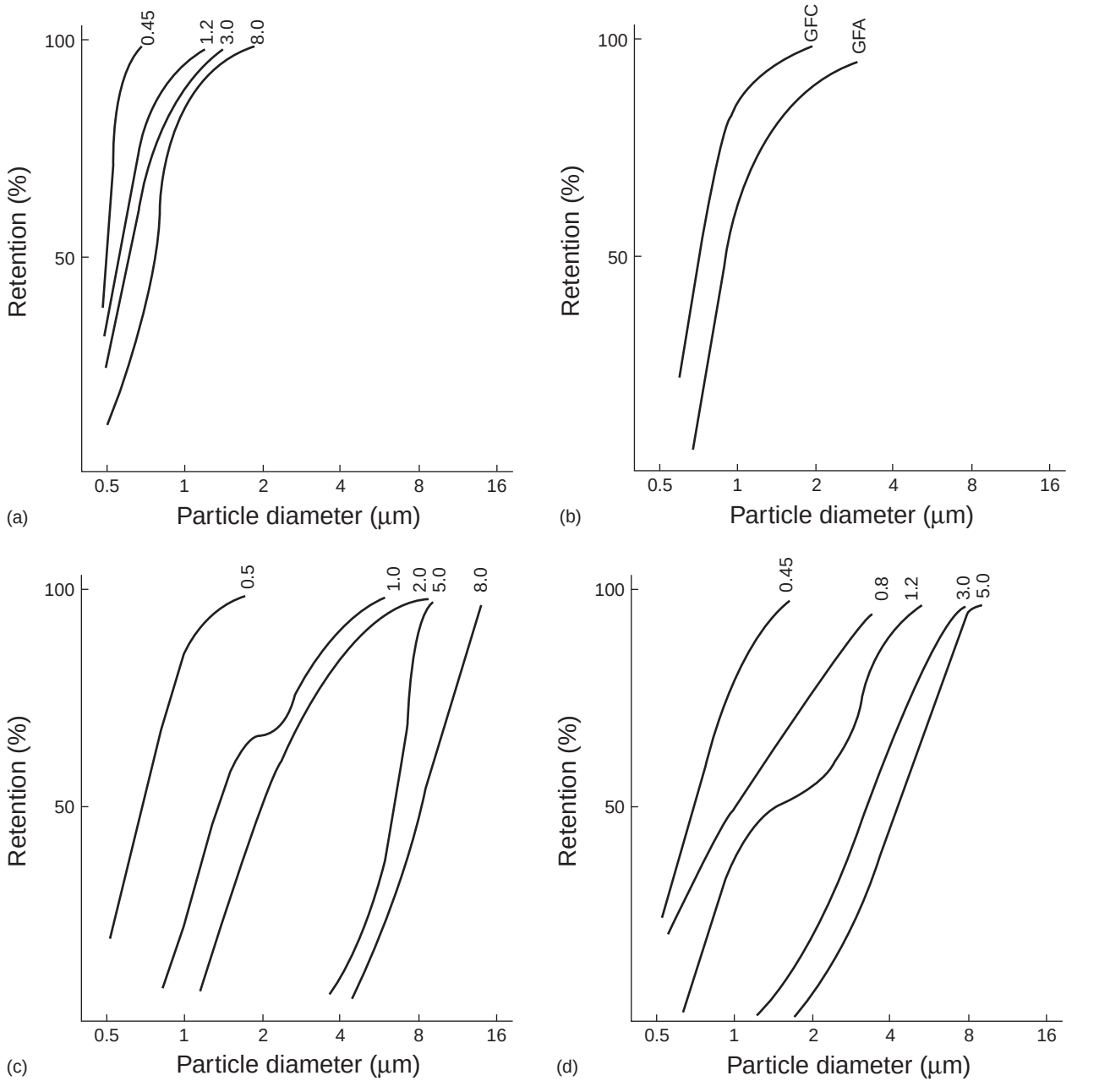

Figure 8 Retention curves for different types of filters. Each curve is labeled with the nominal pore size (μm) indicated by the manufacturer: (a) Millipore™ (cellulose ester depth filters); (b) Whatman™ (GFC GFA); (c) Nuclepore™ (polycarbonate); and (d) Flotronics™ (metal membrane). (Reproduced by permission of the American Society of Limnology and Oceanography from Sheldon.[47])

to the filter surface or into pores or fiber spacing, increasing the potential of filter clogging. If the filter is to be examined optically for the determination of particle size, then any deformation of the particles is undesirable. The use of plankton chambers, discussed in section 3.4, is recommended for providing an estimation of particle size distribution with a minimum of particle size distortion.

4.3 Effect of Procedure Used for Filtration

4.3.1 Effects of Sequential Filtration Procedure

Although there appears to be no significant difference in the resultant grain-size distribution derived between serial or parallel sample processing,[30] there are serious grain-size errors associated with the sequential filtering

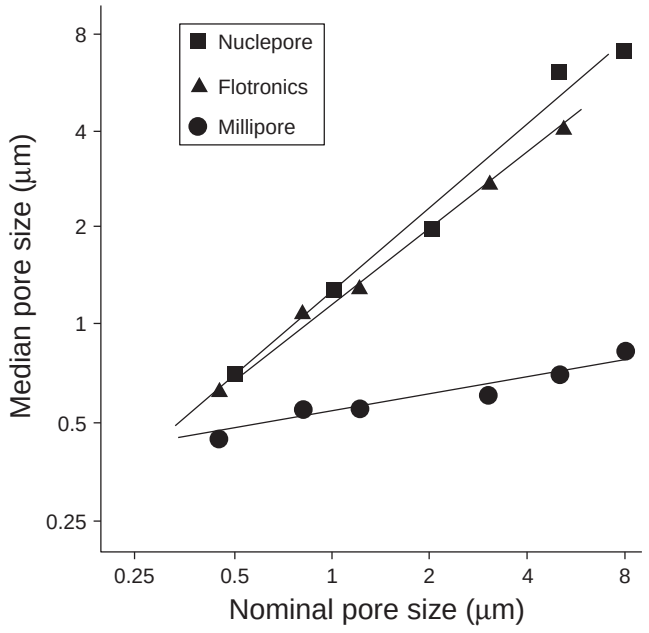

Figure 9 Relationship between the effective pore size and the nominal pore size indicated by the manufacturer. (Reproduced by permission of the American Society of Limnology and Oceanography from Sheldon.[47])

of samples.[43] Droppo et al.[43] concluded from fractionation experiments with primary particle inorganic sediment that using the sequential filtration technique (irrespective of filter type) is very inaccurate as filters retain a spectrum of particles both larger and smaller than the targeted size classes (Figure 10a–f). In addition, the distribution errors observed are cumulative owing to the sequential nature of the process, thus further magnifying the error for successively smaller size classes. This is exemplified in Figure 10(a–f) where every size fractionation retained possesses a portion of particles which are $<8\,\mu m$. Such inaccuracies result in broader size distributions for each fractionated compartment and will have an erroneous influence on the mass in each size category, thus providing an erroneous size distribution. Modelling results by Logan[30] have shown that for sequential filtration with a wide range of screens and a glass-fiber filter <90% of particles (carboxylate microspheres) in any size class contributed to the actual size class defined by the nominal pore size of the screens/filters. For example, Logan[30] calculated that a $210\,\mu m$ nylon mesh would retain 50.1% of the total mass of particles even though none of the particles in the size distribution exceeded $100\,\mu m$. These errors are probably due to a number of possible factors, including sediment size, sediment and pore geometry, pore clogging, particle sticking coefficients, the flocculation and disaggregation process and instrumental assumptions of spherical particles. The error induced by pore clogging is accentuated as pore size decreases[20,43] and sticking coefficients increase.[30] Targeted size classes cannot be derived from the nominal pore sizes of the filters.

4.3.2 Effects of Sample Volume

While the amount of sediment to be filtered out of solution will be dictated by the analysis to be performed on the sediment, the smallest sample volume should be used in order to avoid overloading the filter, resulting in clogging and resultant inaccurate chemical and particle sizing as discussed in section 4.1.[20,47,48] When large-volume filtrations are needed, frequent filter changes and/or large area filters should be used.

4.3.3 Effects of Filter Pressure and Flow Rate

Although Sheldon[47] found that the degree of vacuum had no effect on the retention characteristics of various types of filters, a working rule when using filtration for particle size analysis is that, when possible, only gravity filtration should be performed. However, when positive or negative pressure is required, the lowest pressure or vacuum possible is preferred (below 0.17–0.3 atm).[20,47,49] In so doing, there is a minimization of the erroneous effects of particle breakup and deformation (flattening) at the filter surface[43,47] and filter clogging is also minimized by reducing the rate of particle trapping (especially by inertial collection).

The degree of vacuum or pressure is also related to the flow rate and, as such, a low flow rate is also desirable. This is particularly true for small particle sizes and colloids. Buffle et al.[20] have shown that for colloidal particles, there is a narrow window flow rate and storage time at which separation is optimized. If the flow rate is too fast (i.e. the residence time of the solution above the filter is too short), then variability in the particle sizing result is greater owing to filtration conditions and the nature of the membrane. If, on the other hand, the flow rate is too slow and the residence time too long, then variability is a function of storage conditions and particle nature.

4.3.4 Effect of Stirring

Whereas Laxen and Chandler[49] found that sample stirring had no benefits for Sartorius™ or Nuclepore™ filters, many have suggested that stirring the sample in the filtration reservoir while filtering helps to keep an even distribution of particles in suspension above the filter and reduces any concentration polarization that may occur between particles and the filter surface.[20,21] Stirring will have the greatest positive effect for the colloidal fraction of particles. If stirring is not applied then particles concentrate close to the filter surface and coagulation and associated filter clogging can occur at a faster rate.[48] Too

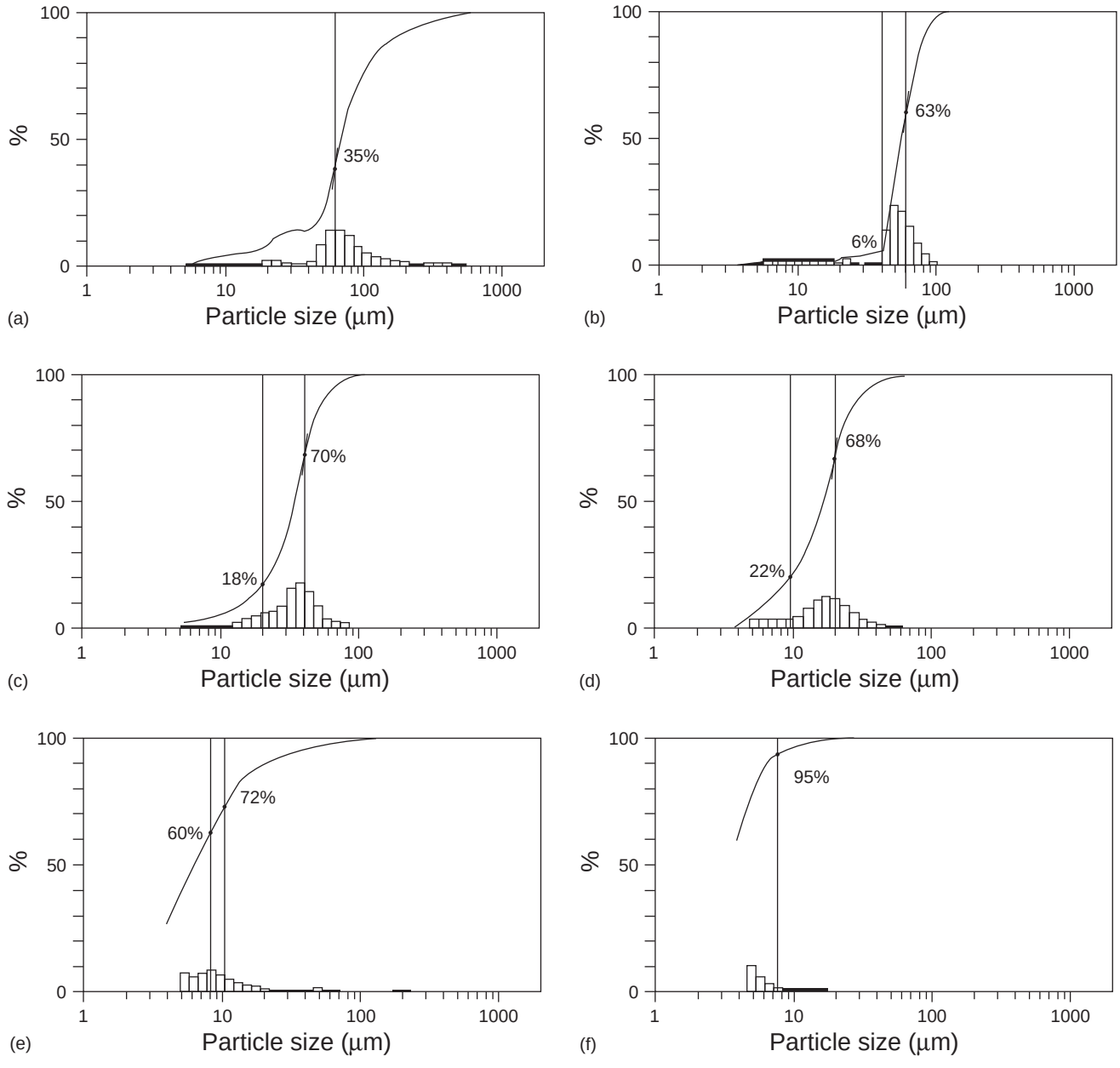

Figure 10 Distribution of particles for each targeted size class of a sequential filtration procedure: (a) >60; (b) 60–41; (c) 41–20; (d) 20–10; (e) 10–8; and (f) <8 μm. The vertical lines represent the nominal pore size of the applicable filters. The percentage of sediment finer than the nominal pore sizes is also indicated. (Reprinted with permission from I.G. Droppo, B.G. Krishnappan, S.S. Rao, E.D. Ongley, 'Investigation of a Sequential Filtration Technique for Particle Fractionation', *Environ. Sci. Technol.*, **29**, 546–550 (1995). Copyright 1995, American Chemical Society.)

much stirring can result in a change in particle size by promoting aggregation and/or disaggregation. Stirring, however, depending on the apparatus used, is not always possible.

4.3.5 Effect of Field Sampling

Filtering in the field, as opposed to filtering in the laboratory following sample transport/storage, will generally provide more accurate grain size results.[20,49] This is related to (a) minimizing aggregation within sample containers during storage and (b) minimizing chemical reactions brought about by changes in oxygen content, pH, ionic interactions and particle–particle interactions, which can affect the particle size distribution and chemical concentrations. This is primarily a result of precipitation, coprecipitation (often through oxidation) and adsorption

or desorption of ions. Sampling in the field is ideally suited to the direct observation of sediment by using plankton chambers as no sample manipulation is required (see section 3.4). Once the particles are at rest on the microscope slide, they do not move or distort and can then be transported to the laboratory for subsequent direct observation, microscopic analysis and imaging.

5 QUALITY ASSURANCE OF PARTICLE SEPARATION RESULTS

The quality assurance or validation of the accuracy of particle separation may be assessed through a number of technical procedures or through some simple approaches. Which approach to use is dependent on cost, availability of equipment, time of filtration and the nature of the experiment. As most artifacts from sample preparation, handling or filtration will result in an increase in the retention of particles on a filter, conditions which benefit the passage of solute and particles below the targeted size range should be considered as promoting the most realistic results. Buffle et al.[20] provide an excellent summation of the approaches that can be used to evaluate the reliability of filtration results. These are as follows:

1. *Flow-rate and Filtration Pressure*: flow rate and therefore filtration pressure should remain constant and independent of retention. Flow rate can be maintained at constant pressure by using a valve at the output of the cell.
2. *Stirring*: retention must be independent of stirring rate.
3. *Filtrate Concentration*: filtrate concentration should be constant during filtration as an indication of no filter clogging. This is one of the most effective methods of evaluating filtration as it is more qualitative than subjective visual inspection of filter caking.
4. *Microscopy*: microscopic examination of filters can be a very powerful tool in providing evidence of selectivity of filters relative to the grain size retained.
5. *Electrolyte Nature and Concentration*: although not discussed in this article [see Buffle et al.[20]] this may influence diverse processes relevant to filtration and therefore, separation results. Generalization of this effect is difficult and its effects should be tested for each application.

6 ALTERNATIVE PARTICLE SIZING BY OPTICAL IMAGE ANALYSIS

If the investigation only requires an accurate determination of grain size distribution, then filters can be observed through a standard microscope interfaced with an image analysis system. Such a method of particle sizing, however, has the potential to impart a number of artifacts within the results (although it is likely to be more accurate than that derived from mass-balance analysis using sequential filtration). These distribution artifacts arise from the following: (a) pressure filtration resulting in the deformation of flocculated particles either by disaggregation, aggregation, flattening or burying within pores; (b) particle morphological changes due to filter drying (drying of the particles (flocs) will result in morphological changes as many of these flocculated particles are over 90% water; as such, significant shrinkage can occur with concomitant influences on particle sizing results); and (c) particle morphological changes due to the need to make the filters transparent by the application of an oil such as paraffin or microscope immersion oil. Such an application of highly viscous material will in itself promote morphological changes and also a rehydration of particles with unknown effects. It is for these reasons that, if only a grain size distribution is required, the plankton chamber approach to particle sizing is preferred over filtration. The use of gravity settling on to the inverted microscope slide embedded within the plankton chamber allows the examination of particles in their natural state (hydrated) without the destructive manipulations possibly resulting from filtration. Furthermore, this technique allows the stabilization of flocs in agarose with minimal structural disturbance for their examination (single plankton chamber sample) by multiple microscopic techniques (conventional optical microscopy, scanning confocal laser microscopy and transmission electron microscopy).[44,51] Modification of particle structure in any way should always be avoided where possible as it will no doubt have a profound effect on particle size distributions or on other floc characteristics such as fractal dimensions.

DISCLAIMER

Information regarding equipment and the mention of trade names or commercial products are meant as a guide and should not be construed as an endorsement of any particular device or product.

ACKNOWLEDGMENTS

The author would like to thank Dr Gary Leppard for his insightful review of the manuscript. M. West performed the microscopy for the micrographs in Figure 4.

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges)

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Sieving in Particle Size Analysis • Surface Area and Pore Size Distributions • Turbidimetry in Particle Size Analysis • Ultrasonic Measurements in Particle Size Analysis • Velocimetry in Particle Size Analysis

Polymers and Rubbers **(Volume 9)**
Field Flow Fractionation in Analysis of Polymers and Rubbers • Size-exclusion Chromatography of Polymers

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. I.G. Droppo, E.D. Ongley, 'Flocculation of Suspended Sediment in Rivers of Southeastern Canada', *Water Res.*, **28**, 1799–1809 (1994).
2. I.G. Droppo, G.G. Leppard, D.T. Flannigan, S.N. Liss, 'The Freshwater Floc: a Functional Relationship of Water and Organic and Inorganic Floc Constituents Affecting Suspended Sediment Properties', *Water Air Soil Pollut.*, **99**, 43–53 (1997).
3. P.M. Chapman, G.P. Romberg, G.A. Vigers, 'Design of Monitoring Studies for Priority Pollutants', *J. Water Pollut. Control Fed.*, **54**, 292–297 (1982).
4. A. Horowitz, *A Primer on Sediment–Trace Element Chemistry*, 2nd edition, Lewis Publishers, Chelsea, MI, 1991.
5. R.J. Gibbs, 'Transport Phases of Transition Metals in the Amazon and Yukon Rivers', *Geol. Soc. Am. Bull.*, **88**, 829–843 (1977).
6. L. Filipek, R. Owen, 'Geochemical Associations and Grain Size Partitioning of Heavy Metals in Lacustrine Sediments', *Chem. Geol.*, **26**, 105–117 (1979).
7. L. Thorne, G. Nickless, 'The Relation Between Heavy Metals and Particle Size Fractions within the Severn Estuary (UK) Inter-tidal Sediments', *Sci. Total Environ.*, **19**, 207–213 (1981).
8. U. Forstner, G. Wittmann, *Metal Pollution in the Aquatic Environment*, 2nd edition, Springer-Verlag, New York, 1981.
9. W. Salomons, U. Forstner, *Metals in the Hydrocycle*, Springer-Verlag, New York, 1984.
10. A. Horowitz, K. Elrick, 'The Relation of Stream Sediment Surface Area, Grain Size, and Composition to Trace Element Chemistry', *Appl. Geochem.*, **2**, 437–451 (1987).
11. I.G. Droppo, C. Jaskot, 'Impact of River Transport Characteristics on Contaminant Sampling Error and Design', *Environ. Sci. Technol.*, **29**, 161–170 (1995).
12. E.D. Ongley, B.G. Krishnappan, I.G. Droppo, S.S. Rao, R.J. Maguire, 'Cohesive Sediment Transport: Emerging Issues for Toxic Chemical Management', *Hydrobiologia*, **235/236**, 177–187 (1992).
13. A. Mudroch, G. Duncan, 'Distribution of Metals in Different Size Fractions of Sediment from the Niagara River', *J. Great Lakes Res.*, **12**, 117–126 (1986).
14. A. Mudroch, J.M. Azcue, 'Aquatic Sediment Sampling', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R. Meyers, John Wiley and Sons, Chichester, 377–402, 1998.
15. S.S. Rao, I.G. Droppo, C.M. Taylor, B.K. Burnison, 'Freshwater Bacterial Aggregate Development: Effect of Dissolved Organic Matter', *Water Pollut. Res. J. Can.*, **26**, 163–171 (1991).
16. H.G. Barth, S.-T. Sun, 'Particle Size Analysis', *Anal. Chem.*, **57**, 151R–175R (1985).
17. H.G. Barth, S.-T. Sun, 'Particle Size Analysis', *Anal. Chem.*, **65**, 55R–66R (1993).
18. H.G. Barth, R.B. Flippen, 'Particle Size Analysis', *Anal. Chem.*, **67**, 257R–272R (1995).
19. H.G. Barth, S.-T. Sun, R.M. Nickol, 'Particle Size Analysis', *Anal. Chem.*, **59**, 142R–162R (1987).
20. J. Buffle, D. Perret, M. Newman, 'The Use of Filtration and Ultrafiltration for Size Fractionation of Aquatic Particles, Colloids and Macromolecules', in *Environmental Particles*, ed. J. Buffle, H.P. van Leeuwen, Lewis Publishers, Chelsea, MI, 171–230, Vol. 1, 1992.
21. J. Buffle, *Complexation Reactions in Aquatic Systems: an Analytical Approach*, Ellis Horwood, Chichester, 1988.
22. B.B. Crocker, K.B. Schnelle, 'Air Pollution Control for Stationary Sources', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R. Meyers, John Wiley and Sons, Chichester, 150–213, 1998.
23. A.S. Kao, 'Air Sampling Instruments', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R. Meyers, John Wiley and Sons, Chichester, 270–285, 1998.
24. J. Buffle, G.G. Leppard, 'Characterization of Aquatic Colloids and Macromolecules. 1. Structure and Behavior of Colloidal Material', *Environ. Sci. Technol.*, **29**, 2169–2175 (1995).
25. J. Buffle, G.G. Leppard, 'Characterization of Aquatic Colloids and Macromolecules. 2. Key Role of Physical

Structures on Analytical Results', *Environ. Sci. Technol.*, **29**, 2176–2184 (1995).
26. G.G. Leppard, J. Buffle, 'Aquatic Colloids and Macromolecules: Effects on Analysis', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R. Meyers, John Wiley and Sons, Chichester, 349–377, 1998.
27. T. Allen, *Particle Size Measurement*, Chapman and Hall, London, 1990.
28. R.D. Cadle, *Particle Size: Theory and Industrial Applications*, Reinhold Publishing, New York, 1965.
29. W.C. Hinds, *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particulates*, John Wiley and Sons, New York, 1982.
30. B.E. Logan, 'Theoretical Analysis of Size Distributions Determined with Screens and Filters', *Limnol. Oceanogr.*, **38**, 372–381 (1993).
31. B.E. Logan, T.A. Hilbert, R.G. Arnold, 'Removal of Bacteria in Laboratory Filters: Models and Experiments', *Water Res.*, **27**, 955–962 (1993).
32. S.B. Tuwiner, L.P. Miller, W.E. Brown, *Diffusion and Membrane Technology*, Reinhold Publishing, London, 1962.
33. R.E. Cranston, D.E. Buckley, 'The Application and Performance of Microfilters in Analysis of Suspended Particulate Matter', Report Series/BI-R-72-7, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada, 1972.
34. S. Sourirajan, *Reverse Osmosis and Synthetic Membranes*, National Research Council Canada, Ottawa, 1977.
35. T.D. Brock, *Membrane Filtration*, Science Tech., Madison, WI, 1983.
36. G.E.M. Hall, G.F. Bonham-Carter, A.J. Horowitz, K. Lum, C. Lemieux, B. Quemerais, J.R. Garbarino, 'The Effect of Using Different 0.45 μm Filter Membranes in ''Dissolved'' Element Concentrations in Natural Waters', *Appl. Geochem.*, **11**, 243–249 (1996).
37. *Synthetic Screening and Filtration Media*, Sefar Canada, Ville Mont-Royal, Quebec, 1997.
38. A. Horowitz, K. Lum, C. Lemieux, J. Garbarino, G. Hall, C. Hall, 'Problems Associated with Using Filtration to Define Dissolved Trace Element Concentrations in Natural Water Samples', *Environ. Sci. Technol.*, **30**, 954–963 (1996).
39. D.P.H. Laxen, R.M. Harrison, 'A Scheme for the Physicochemical Speciation of Trace Metals in Freshwater Samples', *Sci. Total Environ.*, **19**, 59–82 (1981).
40. A. Tessier, P.G.C. Campbell, M. Bisson, 'Particulate Trace Metal Speciation in Stream Sediments and Relationships with Grain Size: Implications for Geochemical Exploration', *J. Geochem. Explor.*, **16**, 77–104 (1982).
41. S.J. de Mora, R.M. Harrison, 'The Use of Physical Separation Techniques in Trace Metal Speciation Studies', *Water Res.*, **7**, 723–733 (1983).
42. S.S. Rao, K.K. Kwan, 'Method for Measuring Toxicity of Suspended Particles in Waters', *Toxicity Assess.*, **5**, 91–101 (1990).
43. I.G. Droppo, B.G. Krishnappan, S.S. Rao, E.D. Ongley, 'Investigation of a Sequential Filtration Technique for Particle Fractionation', *Environ. Sci. Technol.*, **29**, 546–550 (1995).
44. I.G. Droppo, D.T. Flannigan, G.G. Leppard, C. Jaskot, S.N. Liss, 'Floc Stabilization for Multiple Microscopic Techniques', *Appl. Environ. Microbiol.*, **62**, 3508–3515 (1996).
45. I.G. Droppo, E.D. Ongley, 'The State of Suspended Sediment in the Freshwater Fluvial Environment: a Method of Analysis', *Water Res.*, **26**, 65–72 (1992).
46. R.W. Sheldon, W.H. Sutcliffe, 'Retention of Marine Particles by Screens and Filters', *Limnol. Oceanogr.*, **14**, 441–444 (1969).
47. R.W. Sheldon, 'Size Separation of Marine Seston by Membrane and Glass-fiber Filters', *Limnol. Oceanogr.*, **17**, 494–498 (1972).
48. L.G. Danielsson, 'On the Use of Filters for Distinguishing Between Dissolved and Particulate Fractions in Natural Waters', *Water Res.*, **16**, 179–182 (1982).
49. D.P.H. Laxen, I.M. Chandler, 'Comparison of Filtration Techniques for Size Distribution in Freshwater', *Anal. Chem.*, **54**, 1350–1355 (1982).
50. I.G. Droppo, B.G. Krishnappan, E.D. Ongley, 'Concentrating Suspended Sediment Samples by Filtration: Effect on Primary Grain-size Distribution', *Environ. Sci.Technol.*, **26**, 1655–1658 (1992).
51. I.G. Droppo, D.T. Flannigan, G.G. Leppard, S.N. Liss, 'Microbial Floc Stabilization and Preparation for Structural Analysis by Correlative Microscopy', *Water Sci. Technol.*, **34**, 155–162 (1996).

Light Scattering, Classical: Size and Size Distribution Characterization

Guy C. Berry
Carnegie Mellon University, Pittsburgh, USA

1 Introduction **1**

2 Scattering Relations **1**
2.1 General Remarks 1
2.2 Scattering at Zero Angle and Infinite Dilution 4
2.3 Scattering at Small Angle and Infinite Dilution 7
2.4 Scattering at Arbitrary Angle and Infinite Dilution 9
2.5 The Size Distribution from Scattering Data at Infinite Dilution 15
2.6 Extrapolation to Infinite Dilution 17

3 Experimental Methods **18**
3.1 Instrumentation 18
3.2 Methods 19

4 Examples **20**
4.1 Static Scattering and Size Separation Chromatography 20
4.2 Light Scattering from Vesicles and Stratified Spheres 22
4.3 Scattering from Very Large Particles 23
4.4 Intermolecular Association 25
4.5 Scattering with Charged Species 28
4.6 Scattering from Optically Anisotropic Solute 29

List of Symbols **30**

Abbreviations and Acronyms **31**

Related Articles **31**

References **31**

The use of classical, or time-averaged, light scattering methods to characterize the size and size distribution of macromolecules in dilute solutions or particles in dilute dispersions is discussed. The necessary scattering relations are presented systematically, starting with three cases at infinite dilution: the scattering extrapolated to zero angle, the scattering at small angle, and the scattering for arbitrary angle, including the inversion of the scattering data to estimate the size distribution. The relationships needed to effect an extrapolation to infinite dilution from data on dilute solutions are also discussed. These are followed by remarks on light scattering methods, and the concluding sections give examples for several applications. The Rayleigh–Gans–Debye (RGD) approximation is usually appropriate in the scattering from dilute polymer solutions and is also adequate for the scattering from dilute dispersions of small particles. The approximation is assumed when appropriate, but more complete theories are introduced where necessary, as in the use of the Mie–Lorentz theory for large spherical particles.

1 INTRODUCTION

Classical, or time-averaged, light scattering methods are well suited to characterizing certain properties of macromolecules in dilute solutions or particles in dilute dispersions. This technique goes under a variety of names, including static, elastic, integrated and absolute scattering. Information on the size and size distribution may be obtained, and in some cases, it may be possible to elucidate the shape of the macromolecule or particle. The theory and practice of such measurements is well represented in reviews and monographs,[1–24] providing the basis for this article. In addition to these, very useful compilations of original papers are available on the scattering from macromolecules[25] and particles.[26] In the following, the scattering relations are presented systematically, starting with three cases at infinite dilution: the scattering extrapolated to zero angle, the scattering at small angle, and the scattering for arbitrary angle. The relationships needed to effect an extrapolation to infinite dilution from data on dilute solutions are then discussed. Remarks on light scattering methods follow, and the concluding sections give examples of several applications. As discussed below, the RGD approximation is usually appropriate in scattering from dilute polymer solutions and is also adequate for scattering from dilute dispersions of small particles. It will be assumed for much of the following, but more complete theories will be introduced where necessary, as in the use of the Mie–Lorentz theory for large spherical particles.

2 SCATTERING RELATIONS

2.1 General Remarks

Classical, or static, light scattering refers to an experiment in which an intensity $I(\vartheta)$ is determined at a scattering angle ϑ by averaging the fluctuating intensity of the light scattered from a material over a time long compared with the timescale of the fluctuations.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The temporal character of the fluctuations is studied in dynamic light scattering,[1,6,10,17,27] (also see Chu,[7] p. 243) a subject outside the purview here. For dilute solutions of macromolecules, or dispersions of particles, the averaged intensity carries information on the intramolecular properties of the molecular weight M, mean-square radius of gyration R_G^2, and mean-square molecular optical anisotropy δ^2, as well as the intermolecular property of the second virial coefficient A_2, and possibly higher order virial coefficients and intermolecular interference phenomena. Furthermore, for polydisperse systems, the data may also permit assessment of any structural distribution (e.g. molecular weight, size, or possibly shape, of a dissolved molecule or dispersed particle). The reader should be forewarned that owing to its history and variety of applications, the nomenclature in light scattering varies among authors (e.g. see Kerker,[21] p. 60). In the interests of consistency, the nomenclature used herein will follow that used previously by the author,[1,4,10,18] and where appropriate these sources will be cited for further detail in addition to, or in lieu of, citations to the original literature referenced therein.

In general, it is not necessary to distinguish between solutions of macromolecules or dispersions of particles in discussing the fundamental relations for scattering, and the following applies to either case. For convenience, both macromolecules and particles will be referred to as the *solute*. In most experimental arrangements of interest here, the scattered intensity in the scattering plane, the plane containing the incident and scattered beams, is determined as a function of ϑ, or modulus $q = (4\pi/\lambda)\sin(\vartheta/2)$ of the scattering vector, with $\lambda = \lambda_0/n_{\text{medium}}$, where n_{medium} is the refractive index of the solution, λ_0 is the wavelength of the incident light in vacuo, and ϑ is the angle between the incident and scattered beams. Often, the electric vector of the incident light is plane polarized perpendicular to the scattering plane, and this arrangement will be presumed in the following, unless otherwise specified. The scattered light intensity is monitored through a photomultiplier or other photon-sensitive device, either as an analog signal or a digital photon count rate. In either case, an instrument response $G(\vartheta)$ at angle ϑ is determined, with $G(\vartheta)$ proportional to the time-averaged intensity $I(\vartheta)$ of the scattered light. The latter is used to compute the Rayleigh ratio $\mathbf{R}(\vartheta)$, given by $r^2 I(\vartheta)/VI_{\text{INC}}$, with r the distance between the scattering centers and the detector, V the scattering volume, and I_{INC} the intensity of the incident light; $\mathbf{R}(\vartheta)$ is often called the differential scattering cross-section in the physics literature (see Bohren and Huffman,[13] p. 383). The ratio $\mathbf{R}(\vartheta)/G(\vartheta)$ may be determined by an appropriate experiment for a given arrangement (also see Chu,[7] p. 243).[1,10,18] The following section will discuss the information that may be obtained in order to characterize the solute in a solution (or dispersion) by analysis of $\mathbf{R}(\vartheta)$ under various conditions.

The excess Rayleigh ratio for the solution with concentration c of solute, less that due to the solvent, is usually the property of interest. This is normally calculated as $\mathbf{R}_{\text{xs}}(\vartheta, c) = \mathbf{R}_{\text{SOLN}}(\vartheta, c) - \mathbf{R}_{\text{SOLVENT}}(\vartheta)$, but alternative estimates of $\mathbf{R}_{\text{xs}}(\vartheta, c)$ may be required for very low-molecular-weight solute.[28] The subscripts 'xs', etc. will be suppressed throughout, and $\mathbf{R}(\vartheta, c)$ will imply the excess quantity. Here, c is the weight of the solute per unit volume of the solution. Alternatively, the number of molecules ν per unit volume of the solution is sometimes used, especially in theoretical work; $\nu = cN_A/M_n$, with M_n the number average molecular weight determined, for example, from colligative properties, and N_A Avogadro's constant (see Flory,[29] p. 273). It may be noted that the scattered intensity in discussions of the scattering from particles is sometimes expressed in intensity units per particle per volume, with an implied factor ν omitted, for example these units are used in two well-known monographs focussed on the scattering from particles.[21,22]

The notation $\mathbf{R}_{\text{Si}}(\vartheta, c)$ will be used in the following to designate the polarization states of the scattered and incident light rays, where subscripts S and i will designate the polarization of the electric vectors of the scattered and incident light, respectively, relative to the scattering plane. Thus, for vertically polarized incident light, and horizontally or vertically polarized scattered light, respectively, the observed scattering may be expressed as Equations (1) and (2),[4,10]

$$\mathbf{R}_{\text{Hv}}(\vartheta, c) = \mathbf{R}_{\text{aniso}}(\vartheta, c) \tag{1}$$

$$\mathbf{R}_{\text{Vv}}(\vartheta, c) = \mathbf{R}_{\text{iso}}(\vartheta, c) + \left(\frac{4}{3}\right. \mathbf{R}_{\text{aniso}}(\vartheta, c) + \mathbf{R}_{\text{cross}}(\vartheta, c) \tag{2}$$

where $\mathbf{R}_{\text{Vv}}(\vartheta, c) = \mathbf{R}_{\text{iso}}(\vartheta, c)$ for an optically isotropic scatterer, and $\mathbf{R}_{\text{cross}}(\vartheta, c)$ is a cross-term that vanishes for an optically isotropic solute or for $\vartheta = 0$. The functions $\mathbf{R}_{\text{iso}}(\vartheta, c)$ and $\mathbf{R}_{\text{aniso}}(\vartheta, c)$ are discussed in the following sections. The Rayleigh ratio $\mathbf{R}_{\text{Hh}}(\vartheta, c)$ is discussed below, but in the RGD regime, $\mathbf{R}_{\text{Hh}}(\vartheta, c) = \mathbf{R}_{\text{Vv}}(\vartheta, c)\cos^2(\vartheta)$. In much of the older work, written prior to the advent of lasers as a commonly used source of the incident light, an unpolarized (or natural) incident beam is assumed, e.g. comprising equal components of plane-polarized light with electric vectors in the vertical and horizontal directions. Consequently, the Rayleigh ratio for light with electric vector in the vertical direction might be denoted $\mathbf{R}_{\text{Vu}}(\vartheta, c)$ in such a case, with $\mathbf{R}_{\text{Vu}}(\vartheta, c) = \mathbf{R}_{\text{Vv}}(\vartheta, c)/2$ for an isotropic solute (with $\mathbf{R}_{\text{Vh}}(\vartheta, c) = \mathbf{R}_{\text{Hv}}(\vartheta, c) = 0$) if the total intensity of the incident beam is used in the calculation of each.

In most of the discussion the RGD approximation will be assumed, but this constraint will be relaxed where necessary, e.g. in the use of the complete Mie–Lorentz scattering theory for large spheres (Bohren and Huffman,[13] p. 158; Kerker,[21] p. 414; van de Hulst,[22] p. 85). In the RGD approximation, the scatterer is treated as an assembly of scattering elements which scatter radiation independently of all other elements. The elements are taken to represent the smallest unit that both gives a unique refractive index tensor and for which interference effects are negligible for rays scattered from a single element, e.g. the elements are independent Rayleigh scatterers (see Chu,[7] p. 14; Bohren and Huffman,[13] p. 7; Kerker,[21] p. 414; van de Hulst,[22] p. 86).[16,18] An important property in the RGD approximation that is used in much of the following is that the optical contrast factor can usually be considered to be independent of the size or shape of the scatterer, depending only on its composition relative to that of the medium. This allows a factorization to be assumed in much of the following for which the RGD approximation is utilized, but which must be abandoned if that approximation cannot be used. Use of the RGD approximation greatly facilitates the discussion of the effects of heterogeneity of the scatterers, permitting analytical representations of certain averages for properties determined by analysis of $\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/c$ and $\mathbf{R}_{\mathrm{Hv}}(\vartheta, c)/c$, as discussed in the following. The RGD approximation fails as the phase shift of the light within an individual scatterer domain becomes too large, see below.

In most cases, the discussion will be couched in terms of the weight fraction $w_\mu = c_\mu/c$ of scattering components distinguished by some characteristic, e.g. molecular weight, shape, optical contrast relative to the supporting medium, etc. with overall concentration $c = \sum_\mu c_\mu$. As discussed below, $[\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/Mc]_\mu$ reduces to a constant under certain conditions, but in general, this ratio depends on M, c and ϑ, reflecting interference effects from the scattered rays from different scatterers, as well as the particular characteristics of the individual scatterer. Of course, similar expressions may be written for $\mathbf{R}_{\mathrm{Hv}}(\vartheta, c)$.

It is convenient to express $\mathbf{R}_{\mathrm{iso}}(\vartheta, c)$ in the form of Equations (3) and (4)[4,10]

$$\mathbf{R}_{\mathrm{iso}}(\vartheta, c) = K'\psi^2_{\mathrm{solute}} M_{\mathrm{LS}} S_{\mathrm{iso}}(\vartheta, c) c \tag{3}$$

$$S_{\mathrm{iso}}(\vartheta, c) = P_{\mathrm{iso}}(\vartheta, c) F_{\mathrm{iso}}(\vartheta, c) \tag{4}$$

where M_{LS} is the light scattering averaged molecular weight, $S_{\mathrm{iso}}(\vartheta, c)$ is the structure factor (see more below), $P_{\mathrm{iso}}(\vartheta, c)$ and $F_{\mathrm{iso}}(\vartheta, c)$ are intramolecular and intermolecular structure factors, respectively, with $P_{\mathrm{iso}}(0, c) = F_{\mathrm{iso}}(0, 0) = 1$, $K' = 4\pi^2 n^2_{\mathrm{medium}}/N_{\mathrm{A}}\lambda_0^4$, and ψ_{solute} a contrast factor; it should be noted that some authors use a notation in which $F_{\mathrm{iso}}(\vartheta, c)$ would be denoted $S_{\mathrm{iso}}(\vartheta, c)$.[30] For many cases discussed below, ψ_{solute} is equal to the refractive index increment $(\partial n/\partial c)_\Pi$ at osmotic equilibrium of solvent components.[10,16,18] For a solute in a single component solvent, $(\partial n/\partial c)_\Pi \approx n_{\mathrm{medium}}(\tilde{n} - 1)/\rho_{\mathrm{solute}}$, with ρ_{solute} the density of the solute, and $\tilde{n} = n_{\mathrm{solute}}/n_{\mathrm{medium}}$, where n_{solute} and n_{medium} are the refractive indices of the solute and medium, respectively.[31] This form is also used for a suspension of particles in a gas or vacuum for small $\tilde{n} - 1$; a revised expression for large $\tilde{n} - 1$ is discussed below (Bohren and Huffman,[13] p. 158; Kerker,[21] p. 414; van de Hulst,[22] p. 85). Here, and in the following, the subscript 'LS' designates the average obtained in light scattering for solute polydisperse in its properties. The nature of that average for particular types of heterogeneity is delineated in the following sections.

In general, for a nonabsorbing isotropic solution, expansion of $P_{\mathrm{iso}}(\vartheta, c)$ at infinite dilution gives Equation (5),[10,18] (see also Kerker,[21] p. 424; van de Hulst,[22] p. 98; Guinier and Fournet,[24] p. 24)

$$P_{\mathrm{iso}}(\vartheta, 0)^{-1} = 1 + \frac{(R^2_{\mathrm{G}})_{\mathrm{LS}} q^2}{3} + \cdots \tag{5}$$

where $(R^2_{\mathrm{G}})_{\mathrm{LS}}$ is the light scattering average mean-square radius of gyration of the scatterers and the notation $P_{\mathrm{iso}}(\vartheta, 0)$ denotes the function at infinite dilution. In most cases, for optically isotropic scatters, it is useful to express $F_{\mathrm{iso}}(\vartheta, c)$ as Equation (6)[4,10,31] (see also Yamakawa,[20] p. 186)

$$F_{\mathrm{iso}}(\vartheta, c)^{-1} = 1 + c\Gamma_{\mathrm{iso}}(c) P_{\mathrm{iso}}(\vartheta, c) H_{\mathrm{iso}}(\vartheta, c) \tag{6}$$

since in many cases, $H_{\mathrm{iso}}(\vartheta, c) \approx 1$. With this representation (Equation 7)

$$\begin{aligned} KcM/\mathbf{R}_{\mathrm{iso}}(\vartheta, c) &= S_{\mathrm{iso}}(\vartheta, c)^{-1} \\ &= P_{\mathrm{iso}}(\vartheta, c)^{-1} + c\Gamma_{\mathrm{iso}}(c) H_{\mathrm{iso}}(\vartheta, c) \end{aligned} \tag{7}$$

where $K = K'\psi^2_{\mathrm{solute}}$.

As elaborated below, in most cases, a virial expansion of $\Gamma_{\mathrm{iso}}(c)$ suffices for the scattering from dilute solutions (i.e. solutions with $cN_{\mathrm{A}}(R^2_{\mathrm{G}})^{3/2}_{\mathrm{LS}}/M_{\mathrm{LS}} \ll 1$), Equation (8)[4,10,31] (see also Yamakawa,[20] p. 186)

$$\Gamma_{\mathrm{iso}}(c) = 2(A_2)_{\mathrm{LS}} M_{\mathrm{LS}} + 3(A_3)_{\mathrm{LS}} M_{\mathrm{LS}} c + \cdots \tag{8}$$

with $(A_2)_{\mathrm{LS}}$ the light scattering averaged second virial coefficient, etc. The virial expansion may fail for charged scatterers in a solvent with very low ionic strength, so that electrostatic interactions are not shielded (see section 4 for examples),[10,16,18,32] and other forms for $\Gamma_{\mathrm{iso}}(c)$ may be required for moderately concentrated (or semidilute) solutions.[3,4,33] The general interpretation of the functions $P_{\mathrm{iso}}(\vartheta, c)$, $H_{\mathrm{iso}}(\vartheta, c)$ and the parameters M_{LS},

$(A_2)_{LS}$, and $(R_G^2)_{LS}$ may be complex for scatterers that are optically anisotropic or exhibit intramolecular and/or intermolecular heterogeneity.[10,18] These cases will be the subject of much of the following discussion. It should be noted that these relations and those to follow apply either to polymeric or colloidal solute, and that similar relations also obtain for the scattering of X-rays and neutrons (with revision of the definition for K),[34,35] with the principal experimental difference being the range of wavelength normally encompassed.

Similarly, $\mathbf{R}_{aniso}(\vartheta, c)$ may be expressed as Equation (9)[4,10]

$$\mathbf{R}_{aniso}(\vartheta, c) = \left(\frac{3}{5}\right) K'\gamma^2_{solute} \left(\frac{\delta}{\delta_0}\right)^2_{LS} M_{LS} S_{aniso}(\vartheta, c) c \quad (9)$$

where $(\delta/\delta_0)^2_{LS}$ is the light scattering averaged mean-square molecular optical anisotropy form factor, $S_{aniso}(\vartheta, c)$ is a structure factor, and γ_{solute} is an anisotropic contrast factor, conveniently expressed as $\psi_{solute}\delta_0$ if $\psi_{solute} \neq 0$, with δ_0 the optical anisotropy of a scattering element, see the section 4, ''Examples'' below.[10] As above, the factor $K'\gamma^2_{solute}$ must be modified if the RGD approximation is not valid. A similar set of expressions to those for $S_{iso}(\vartheta, c)$ may be written for $S_{aniso}(\vartheta, c)$, with $P_{iso}(\vartheta, c)$ replaced by $P_{aniso}(\vartheta, c)$, etc. In most cases of interest here, $\Gamma_{aniso}(c)$ is considerably smaller than $\Gamma_{iso}(c)$, $H_{aniso}(\vartheta, c) \approx 1$, and $P_{aniso}(\vartheta, c) \neq P_{iso}(\vartheta, c)$, except for $\vartheta = 0$,[10] see below.

2.2 Scattering at Zero Angle and Infinite Dilution

2.2.1 *Isotropic Solute in the Rayleigh–Gans–Debye Regime*

In this limit, attention is focused on $[\mathbf{R}_{Vv}(0, c)/c]^0$ (or $[\mathbf{R}_{Hv}(0, c)/c]^0$ below), with the implicit assumption that the extrapolations to zero angle and infinite dilution can be accomplished; these extrapolations are discussed in section 2.6. The superscript ''0'' indicates the extrapolation to infinite dilution throughout, and unless stated otherwise, it is assumed that $\psi_{solute} \neq 0$, as would usually be the case in applications of interest in light scattering. In this regime, $S_{iso}(\vartheta, c) = S_{aniso}(\vartheta, c) = 1$, simplifying the analysis. For example, $[\mathbf{R}_{Vv}(0, c)/K'\psi^2_{solute}c]^0$ is equal to the parameter M_{LS}, which as seen below, may be equal to the weight-average molecular weight M_w under certain conditions, but will often be a more complex function of the properties of the scatterers. Beginning within the RGD approximation, for an optically isotropic solute comprising a number of components, possibly differing in chemical composition, Equations (10) and (11) apply[10,18]

$$M_{LS} = \left(\frac{\mathbf{R}_{Vv}(0, c)}{K'\psi^2_{solute}c}\right)^0 \quad (10)$$

$$M_{LS} = \frac{\sum_\mu w_\mu M_\mu^{-1} \left(\sum_i \psi_i m_i\right)^2_\mu}{\psi^2_{solute}} \quad (11)$$

with ψ_i and m_i the refractive index increment and the molar mass of element i, respectively, and w_μ the weight fraction of components with molecular weight $M_\mu = (\quad m_i)_\mu$ and identical values of $(\quad \psi_i m_i)_\mu$, with $\psi_{solute} = \quad w_\mu\psi_\mu$. Here, the sums over i extend over all of the scattering elements in molecule (particle) μ.

If each component is compositionally homogeneous, i.e. has only one type of scattering element, then $(\quad \psi_i m_i)_\mu = \psi_\mu M_\mu$, and Equation (12) follows

$$M_{LS} = \frac{\sum_\mu w_\mu M_\mu \psi^2_\mu}{\psi^2_{solute}} \quad (12)$$

In the simplest case, for an optically isotropic, compositionally homogeneous solute, $\psi_\mu = \psi_{solute}$, and this reduces to the well-known relation, Equation (13)[10,18,31]

$$M_{LS} = M_w = \sum_\mu w_\mu M_\mu \quad (13)$$

where w_μ is the weight fraction of component μ with molecular weight M_μ. This simple relation applies to solutions or dispersions in a mixed solvent, provided that $(\partial n/\partial c)_\Pi$ is used, as stipulated above. If the refractive index $(\partial n/\partial c)_w$ determined at constant composition of the mixed solvent is used instead, then, to a good approximation, Equation (14) holds[10,16,18]

$$M_{LS} = \left(\frac{(\partial n/\partial c)_\Pi}{(\partial n/\partial c)_w}\right)^2 M_w \quad (14)$$

where $(\partial n/\partial c)_\Pi = (\partial n/\partial c)_w$ for a single-component solvent. The ratio $(\partial n/\partial c)_\Pi/(\partial n/\partial c)_w$ may be interpreted in terms of preferential solvation of the solute by a component of the solvent.[10,16,18] Deviations of $(\partial n/\partial c)_\Pi/(\partial n/\partial c)_w$ from unity are usually most pronounced when the preferentially solvating component is present at low concentration in the mixed solvent.

The more complex relation given above must be used if the solute is optically heterogeneous, as with a copolymer or particles with a spatially varying refractive index. For the special case with only two types of element, characterized by refractive index increments ψ_A and ψ_B, with $\psi_{solute} = w_A\psi_A + (1 - w_A)\psi_B \neq 0$, Equation (15) applies[10,36,37]

$$M_{LS} = M_w + 2Y \sum w_\mu M_\mu \Delta w_\mu + Y^2 \sum w_\mu M_\mu (\Delta w_\mu)^2 \quad (15)$$

where $\Delta w_\mu = (w_A)_\mu - w_A = w_B - (w_B)_\mu$ and $Y = (\psi_A - \psi_B)/\psi_{solute}$. Thus, M_{LS} is expected to be parabolic in

Y, with $M_{LS} = M_w$ for $Y = 0$. Since $\Delta w_\mu = 0$ for either a strictly alternating copolymer, or a copolymer with random placement of the monomers, $M_{LS} = M_w$ for these important cases. Similarly, $M_{LS} = M_w$ for a collection of particles with two kinds of scattering element, provided that the composition of each particle is identical to the average composition.

2.2.2 *Isotropic Solute Beyond the Rayleigh–Gans–Debye Regime*

It should be recognized at the outset that the RGD approximation will almost always be adequate for (nonabsorbing) threadlike macromolecules (e.g. flexible or semiflexible coils, rod-like or helical chains, etc.). This is the case not only at zero angle and infinite dilution, but also for all angles and dilute solutions, owing to the sparse density of the macromolecule in the intramolecular "domain". Thus, the RGD approximation may be applied, even for a large macromolecule, with an appreciably different refractive index to that of the solvent. The failure of the RGD approximation arises for large particles, under certain conditions elaborated in the following. If the RGD approximation cannot be used, the $[\mathbf{R}_{Vv}(0, c)/K'\psi^2_{solute}]^0_\mu$ may still be computed using an appropriate model, but because the scattering may no longer be taken as a sum over independent Rayleigh scattering elements, $\left(\psi_i m_i\right)_\mu$ for the μ-th scatterer must be replaced by a more complex function. For the discussion of M_{LS}, this can be done for compositionally homogeneous scatterers by including a factor $h(\tilde{n}_\mu)$ in ψ_μ for the μ-th scattering component, and multiplication of the modified ψ_μ by a function $m(\tilde{n}_\mu, \lambda, M_\mu)$. Thus, even for chemically homogeneous scatterers differing only in size, the contrast factor may depend significantly on λ, $\tilde{n}$, M_μ, and the shape of the scatterer. With this modification, $M_{LS} = [m(\tilde{n}, \lambda, M)]^2 M$ for a monodisperse solute (with M_{LS} calculated using ψ_μ including the factor $h(\tilde{n}_\mu)$). For example, with the Mie theory for a homogeneous sphere $h(\tilde{n})$ is given by $h_{sph}(\tilde{n}) = 3(\tilde{n} + 1)/2(\tilde{n}^2 + 2)$ (see Bohren and Huffman,[13] p. 158; Kerker,[21] p. 414; van de Hulst,[22] p. 70) and $m(\tilde{n}, \lambda, M)$ may be expressed as $m_{sph}(\tilde{n}, \alpha)$, and calculated as a function of $\tilde{n}$ and $\alpha = 2\pi R/\lambda$; $m_{sph}(\tilde{n}, \alpha)$ reduces to unity as the magnitude of the maximum phase shift $2\alpha|\tilde{n} - 1|$ in the particle becomes small (see Bohren and Huffman,[13] p. 113; Kerker,[21] p. 91; van de Hulst,[22] p. 127). Examples of $M_{LS}/M = [m_{sph}(\tilde{n}, \alpha)]^2$ for monodisperse spheres given in Figure 1 over a range of $\tilde{n}$ and α reveal the complex character of this function, with the possibility that M_{LS}/M may be either larger or smaller than unity. In the use of this figure, consistency is required between M obtained from the determined M_{LS} and the ratio M_{LS}/M and $R = (3M/4\pi\rho N_A)^{1/3}$ used to select α, where ρ is the density of the particle. Deviation of M_{LS}/M from unity is important, for example, in the analysis of the eluent from chromatographic columns by low angle light scattering for large particles, see section 4 below.

It may be noted that the Rayleigh ratio is often not mentioned explicitly in discussions involving the Mie theory, but rather, the theory is developed in terms of scattering amplitude functions $S_i(\vartheta)$, which are the components of the scattering tensor for the scattering from an individual scatterer (see Bohren and Huffman,[13] p. 65; Kerker,[21] p. 47; van de Hulst,[22] p. 47); $S_i(\vartheta)$ is not to be confused with the structure factor $S(\vartheta, c)$ defined above. Reduced intensity functions are computed from the $S_i(\vartheta)$, e.g. for nonabsorbing, optically isotropic scatterers, $i_{Vv}(\vartheta) = |S_1(\vartheta)|^2$ and $i_{Hh}(\vartheta) = |S_2(\vartheta)|^2$, with $[\mathbf{R}_{Vv}(\vartheta, c)/\nu]^0 = N_A i_{Vv}(\vartheta)/k^2$, etc. where $k = 2\pi/\lambda$. For example, $i_{Vv}(\vartheta) = i_{Vv}(0)P_{Vv}(\vartheta, 0)$, with $i_{Vv}(0) = \alpha^6\{2(\tilde{n} - 1)h_{sph}(\tilde{n})/3\}^2[m_{sph}(\tilde{n}, \alpha)]^2$ for spherical particles in the nomenclature used here, which reduces to the expected result $[\mathbf{R}_{Vv}(\vartheta, c)/c]^0 = K'\psi^2_{solute}M_{LS}$ for this case (including the factor $h_{sph}(\tilde{n})$ in ψ_{solute}).

For isotropic spheres, functions in the Mie theory may be expanded for $\tilde{n}\alpha \leq 1$ to give a result (see van de Hulst,[22] p. 45) that may be cast in the form of Equations (16) and (17)

$$m_{sph}(\tilde{n}, \alpha) = 1 + j_{sph}(\tilde{n})\alpha^2|\tilde{n} - 1| + \cdots \tag{16}$$

$$j_{sph}(\tilde{n}) = \frac{(\tilde{n} + 1)(\tilde{n}^4 + 27\tilde{n}^2 + 38)}{[15(\tilde{n}^2 + 2)(2\tilde{n}^2 + 3)]} \tag{17}$$

for nonabsorbing monodisperse spheres, where $j_{sph}(\tilde{n})$ approaches 44/75 as $\tilde{n}$ tends to unity. As may be seen in Figure 1, this expression provides a useful fit to $m_{sph}(\tilde{n}, \alpha)$ for small $\alpha^2|\tilde{n} - 1|$, but is generally limited to $\alpha^2|\tilde{n} - 1| < 0.1$ and $\alpha < 1$. In principle, in this range, the deviation from the RGD approximation could be taken into account using the above in expressions for M_{LS} discussed below for the scattering from spheres heterogeneous in M.

For nonabsorbing spheres with $\alpha \gg 1$, $i_{Vv}(0) = k^2[\mathbf{R}_{Vv}(0, c)/\nu]^0/N_A$ is closely approximated by $\{(\alpha^2/4)Q_{sca}(\tilde{n}, \alpha)\}^2$,[38] (see also Kerker,[21] p. 91) where $Q_{sca}(\tilde{n}, \alpha)$ is the extinction efficiency for scattering for spheres at infinite dilution, i.e. $\exp(-\tau\nu b)$ is the fraction of the incident light transmitted through a sample with turbidity $\tau = Q_{sca}(\tilde{n}, \alpha)\pi R^2$, thickness b and concentration ν. Numerical calculations show that the approximation is quite reasonable for $\tilde{n} > 1.3$ and $\alpha > 4$, with the minimum value of α for which the approximation is useful decreasing with increasing $\tilde{n}$. The abundant literature on the extinction efficiency provides a means to compute $m_{sph}(\tilde{n}, \alpha)$ in this range and has itself been discussed as a means of characterizing size heterogeneity among spherical solutes (see Kerker,[21] p. 351).[39] For example,

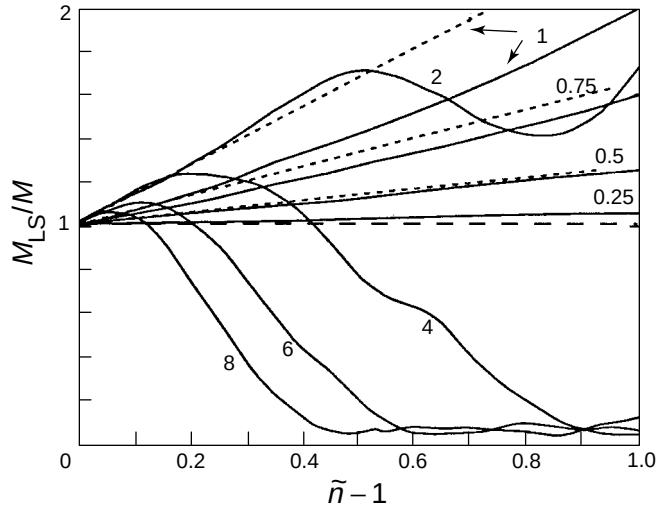

Figure 1 Ratio M_{LS}/M ($= m_{sph}(\tilde{n}, \alpha)^2$) of the light scattering averaged molecular weight M_{LS} for monodisperse spheres of radius R to the molecular weight M as a function of the relative refractive index $\tilde{n}$ for the indicated values of the size parameter $\alpha = 2\pi R/\lambda$. The dashed lines give the limiting behavior for small $\tilde{n} - 1$ (see Equations 16 and 17).

$Q_{sca}(\tilde{n}, \alpha) = (32/27)\alpha^4|\tilde{n} - 1|^2$ for Rayleigh scattering, increasing to $Q_{sca}(\tilde{n}, \alpha) = 2$ in the Fraunhofer scattering regime ($\alpha \gg 1$ and $\tilde{n} > 1.1$) discussed more extensively in the following (see van de Hulst,[22] p. 131).

For a polydisperse nonabsorbing solute, with all components having identical $\tilde{n}$, Equation (18) applies

$$M_{LS} = \sum_{\mu} w_\mu M_\mu [m(\tilde{n}, \lambda, M_\mu)]^2 \qquad (18)$$

Similar expressions could be obtained for particles of other shape for which the RGD approximation cannot be used. For a spherical solute, use of $m(\tilde{n}, \lambda, M_\mu) = m_{sph}(\tilde{n}, \alpha_\mu)$ gives $M_{LS} \geq M_w$, but M_{LS} cannot generally be reduced to an expression involving standard average molecular weights. In the limited range with $\tilde{n}\alpha_\mu < 1$ discussed above, the approximate expression for $m_{sph}(\tilde{n}, \alpha)$ may be applied, with the result (for particles of homogeneous refractive index, Equation 19)

$$M_{LS} \approx M_w + 2\left(\frac{2\pi R}{\lambda M^{1/3}}\right)^2 j(\tilde{n})|\tilde{n} - 1|M_{(5/3)}^{5/3} \qquad (19)$$

where $M_{(5/3)}^{5/3} \approx M_z^{2/3} M_w$ is defined in Table 1 (for $\varepsilon = 2/3$). In principle, the dependence on λ embodied in the parameters α_μ provides a means of interpreting M_{LS} (or equivalently, Q_{sca}) in terms of w_μ from the variation of M_{LS} with λ using the Mie theory to compute $m_{sph}(\tilde{n}, \alpha)$, after accounting for any dispersion in the refractive indices of solute or solvent. Such a procedure has been the basis of some methods proposed in the literature for characterizing the size of spheres by transmission measurements (see Bohren and Huffman,[13] p. 287; Kerker,[21] p. 325),[39] but these have been largely superseded by methods discussed in the following involving measurements over a range of scattering angles.

2.2.3 *Anisotropic Solute*

For a solute with anisotropic scattering elements, in the RGD regime $[\mathbf{R}_{Hv}(0, c)/K'\psi_{solute}^2 c]^0$ gives the light scattering averaged mean-square molecular anisotropy δ_{LS}^2 (with $\psi_{solute} \neq 0$), Equation (20):[10,40,41]

$$\frac{\mathbf{R}_{Hv}(0, c)}{K'\psi_{solute}^2 c}^0 = \left(\frac{3}{5}\right) M_{LS}\delta_{LS}^2 \qquad (20)$$

Table 1 Light scattering average mean-square radius of gyration and hydrodynamic radius (RGD approximation)

	$(R_G^2)_{LS}$	$(R_H)_{LS}$
Exact relation[a]	$(1/M_w) \quad w_\mu M_\mu (R_G^2)_\mu$	$M_w \quad w_\mu M_\mu (R_H^{-1})_\mu$
Approximation for[b] $R_H \propto R_G \propto M^{\varepsilon/2}$	$(R_G^2/M^\varepsilon)M_{(\varepsilon+1)}^{\varepsilon+1}/M_w$	$(R_H/M^{\varepsilon/2})M_w/M_{(1-\varepsilon/2)}^{1-\varepsilon/2}$
Random-flight coil;[c] $\varepsilon = 1$	$(R_G^2/M)M_z$	$(R_H/M^{1/2})M_w/M_{(1/2)}^{1/2} \approx (R_H/M^{1/2})M_w^{1/2}(M_w/M_n)^{0.10}$
Rod-like chain;[c] $\varepsilon \approx 2$	$(R_G^2/M^2)M_z M_{z+1}$	$(R_H/M)M_w$
Sphere;[c] $\varepsilon = 2/3$	$(R_G^2/M^{2/3})M_{(5/3)}^{5/3}/M_w \approx (R_G^2/M^{2/3})M_z^{2/3}(M_w/M_z)^{0.10}$	$(R_H/M^{1/3})M_w/M_{(2/3)}^{2/3} \approx (R_H/M^{1/3})M_w^{1/3}(M_w/M_n)^{0.10}$

[a] For optically isotropic solute, and with $\partial n/\partial c$ the same for all scattering elements.

[b] $M_{(\alpha)} = \left(\sum_\mu w_\mu M_\mu^\alpha\right)^{1/\alpha}$ and $M_n = 1 \Big/ \sum_\mu w_\mu M_\mu^{-1}$; $M_w = \sum_\mu w_\mu M_\mu$; $M_z = \sum_\mu w_\mu M_\mu^2 \Big/ \sum_\mu w_\mu M_\mu$.

[c] Approximations are for a solute with a Schulz–Zimm (two-parameter exponential) distribution of M, for which $M_{(\alpha)} \approx M_w\{\Gamma(1 + h + \alpha)/\Gamma(1 + h)\}^{1/\alpha}/(1 + h)$, where $h^{-1} = (M_w/M_n)^{-1}$, see Berry.[42]

For solute with identical, but optically anisotropic scattering elements, $M_{LS} = M_w$, and Equation (21) applies

$$\delta^2_{LS} = \sum_\mu \frac{w_\mu M_\mu \delta^2_\mu}{M_w} \tag{21}$$

Since δ^2_{LS} arises from orientational correlations among the scattering elements, it cannot be interpreted in the absence of a structural model for the scatterer. An example is included below. The expression for $[\mathbf{R}_{Vv}(0,c)/K'\psi^2_{solute}c]^0$ for this case becomes (Equation 22)[10]

$$\left[\frac{\mathbf{R}_{Vv}(0,c)}{K'\psi^2_{solute}c}\right]^0 = M_{LS}\left\{1 + \frac{4\delta^2_{LS}}{5}\right\} \tag{22}$$

2.3 Scattering at Small Angle and Infinite Dilution

2.3.1 Isotropic Solute in the Rayleigh–Gans–Debye Regime

The ambiguous term *small angle* indicates that ϑ is small enough that the parameter $(R^2_G)_{LS}$ defined by the equivalent expressions, Equation (23)

$$\frac{(R^2_G)_{LS}}{3} = \left[\frac{c}{\mathbf{R}_{Vv}(0,c)}\right]^0 \Bigg\}^{-1} \frac{\partial[c/\mathbf{R}_{Vv}(\vartheta,c)]^0}{\partial q^2} = \frac{\partial \ln[\mathbf{R}_{Vv}(\vartheta,c)/c]^0}{\partial q^2} \tag{23}$$

may be determined in the limit as ϑ goes to zero, where the superscript "0" indicates that the data are extrapolated to infinite dilution. Thus, for optically isotropic scattering elements, the expansion give above for $P_{iso}(\vartheta,0)^{-1}$ may be applied, so that at infinite dilution (averaged over all orientations of the scatterers), in the RGD approximation, Equation (24) follows[1,10,18]

$$(R^2_G)_{LS} = \sum_\mu w_\mu M_\mu^{-1} \frac{\sum\sum_{ij} \psi_i m_i \psi_j m_j \langle r^2_{ij}\rangle}{2M_{LS}\psi^2_{solute}}{}_\mu \tag{24}$$

where $\langle r^2_{ij}\rangle$ is the mean-square separation of scattering elements i and j, and $\psi_{solute} \neq 0$. In this expression, w_μ is the weight fraction of a component with molecular weight M_μ and common ($\psi_i m_i \psi_j m_j \langle r^2_{ij}\rangle)_\mu$, as scatterers with a given molecular weight may differ in composition and/or shape.

For a solute comprising optically identical scattering elements in all components (in the RGD approximation), Equation (25)

$$(R^2_G)_{LS} = \sum_\mu w_\mu M_\mu \frac{\sum\sum_{ij} \langle r^2_{ij}\rangle}{2M_w}{}_\mu = \sum_\mu w_\mu M_\mu \frac{(R^2_G)_\mu}{M_w} \tag{25}$$

where again, the ensemble may include chains with identical M_μ but different $(R^2_G)_\mu$, as for a collection of branched and linear chains, all with the same molecular weight. The similarity between this and the expression for the z-average molecular weight (obtained by replacing $(R^2_G)_\mu$ by M_μ) often motivates the notation $(R^2_G)_z$ for $(R^2_G)_{LS}$ (or Φ_z for any property Φ_μ similarly averaged), but that notation may be misleading if $(R^2_G)_\mu$ is not proportional to M_μ, and is not employed here. The parameter R^2_G may be related to the characteristics of the scatterer in this RGD regime, such as the contour length L and persistence length $\hat{a}$ of polymer chains, the radius R of spheres, etc. (see Table 1), or, in some cases, expressed as a power of the molecular weight of the scatterer. Thus, for a solute with the special property that $(R^2_G)_\mu = rM^\varepsilon_\mu$ with r a constant, the summation gives Equations (26) and (27)[1,42]

$$(R^2_G)_{LS} = \frac{rM_{(\varepsilon+1)}^{\varepsilon+1}}{M_w} \tag{26}$$

$$M_{(\alpha)} = \left(\sum_\mu w_\mu M^\alpha_\mu\right)^{1/\alpha} \tag{27}$$

where expressions for $M_{(\alpha)}$ are given in Berry[42] for a number of commonly used molecular-weight distributions (e.g. although $M_{(-1)} = M_n$, $M_{(1)} = M_w$, $M_{(2)} = (M_w M_z)^{1/2}$ such simple expressions are not possible if α is not an integer). Expressions for $(R^2_G)_{LS}$ are given in Table 1 for a few cases of interest.

Values of the light scattering averaged hydrodynamic radius $(R_H)_{LS}$ determined by dynamic light scattering are included in Table 1 for comparison. For example, for a random-flight chain ($\varepsilon = 1$), $(R^2_G)_{LS} = rM_z$, with $r = \hat{a}L/3M$, and for a rod ($\varepsilon = 2$), $(R^2_G)_{LS} = rM_zM_{z+1}$, with $r = (L/M)^2/12$, where M_z and M_{z+1} are the z- and $z+1$-average molecular weights. It may be noted that in general, the averages in $(R^2_G)_{LS}$ and $(R_H)_{LS}$ differ.

For an optically isotropic solute discussed above comprising two different types of scattering elements (with $\psi_{solute} \neq 0$, and in the RGD approximation, Equation 28),[10,36,37]

$$(R^2_G)_{LS} = x(R^2_G)_A + (1-x)(R^2_G)_B + x(1-x)\Delta^2_{AB} \tag{28}$$

for a sample monodisperse in molecular weight, with $x = w_A\psi_A/\psi_{solute}$, $(R^2_G)_A$ and $(R^2_G)_B$ the mean-square

radii of gyration of the portions of the solute molecules (particles) comprising only type A and type B scattering elements, respectively, and Δ^2_{AB} the mean-square separation of the center of mass for these portions. The possible dependence of the parameters $(R^2_G)_A$, $(R^2_G)_B$ and Δ^2_{AB} on solvent is neglected herein; this would be expected to be less of an issue with particles than with polymer chains. Expressions are available for Δ^2_{AB} for certain graft copolymers, and for a solute polydisperse in molecular weight.[36] In some cases Δ^2_{AB} is zero, simplifying the expression for $(R^2_G)_{LS}$, i.e. alternating copolymers or copolymers with a random placement of monomers, or spheres with a uniform coating (or shell). For block copolymers comprising n blocks each of A and B chains (e.g. $n = 1$ for an AB-diblock copolymer), $\Delta^2_{AB} = 2\{(R^2_G)_A + (R^2_G)_B\}/n$, showing that Δ^2_{AB} decreases with increasing n, as expected since $(R^2_G)_{LS} = (R^2_G)_A = (R^2_G)_B$ for an alternating copolymer.[36] Inspection of these expressions shows that $(R^2_G)_{LS}$ may bear little resemblance to the geometric mean-square radius of gyration $(R^2_G)_{geo}$ of the chain, e.g. $(R^2_G)_{LS}$ may be negative, whereas$(R^2_G)_{geo} > 0$. For example, one could construct a diblock polymer or a coated sphere for which $(R^2_G)_{LS}$ is zero, or even negative, though $(R^2_G)_{geo}$ must be positive, see section 4 examples below.

2.3.2 *Isotropic Solute Beyond the Rayleigh–Gans–Debye Regime*

The preceding expressions must be modified if the RGD approximation fails, as the basic assumption in the RGD model that the scattering from the elements may be added independently for a particle is no longer valid (the reader is reminded that as discussed above, the RGD approximation is almost always adequate for (nonabsorbing) thread-like macromolecules). For a compositionally homogeneous scatterer (all particles with the same $\tilde{n}$), Equation (29) applies

$$(R^2_G)_{LS} = \frac{\sum_\mu w_\mu M_\mu y(\tilde{n}, \lambda, M_\mu)[m(\tilde{n}, \lambda, M_\mu)]^2 (R^2_G)^{(RGD)}_\mu}{\sum_\mu w_\mu M_\mu [m(\tilde{n}, \lambda, M_\mu)]^2} \tag{29}$$

where $(R^2_G)^{(RGD)}_\mu$ is the value in the RGD regime, $m(\tilde{n}, \lambda, M_\mu)$ is defined above, and $y(\tilde{n}, \lambda, M_\mu)$ tends to unity as $\alpha|\tilde{n} - 1|$ tends to zero.

With the Mie theory for a compositionally homogeneous sphere, for which $(R^2_G)^{(RGD)} = 3R^2/5$, $y(\tilde{n}, \lambda, M)$ may be expressed as $y_{sph}(\tilde{n}, \alpha)$, and calculated as a function of $\tilde{n}$ and $\alpha = 2\pi R/\lambda$ (see Bohren and Huffman,[13] p. 113; Kerker,[21] p. 91; van de Hulst,[22] p. 127). Consequently, for monodisperse spheres, Equation (30) applies

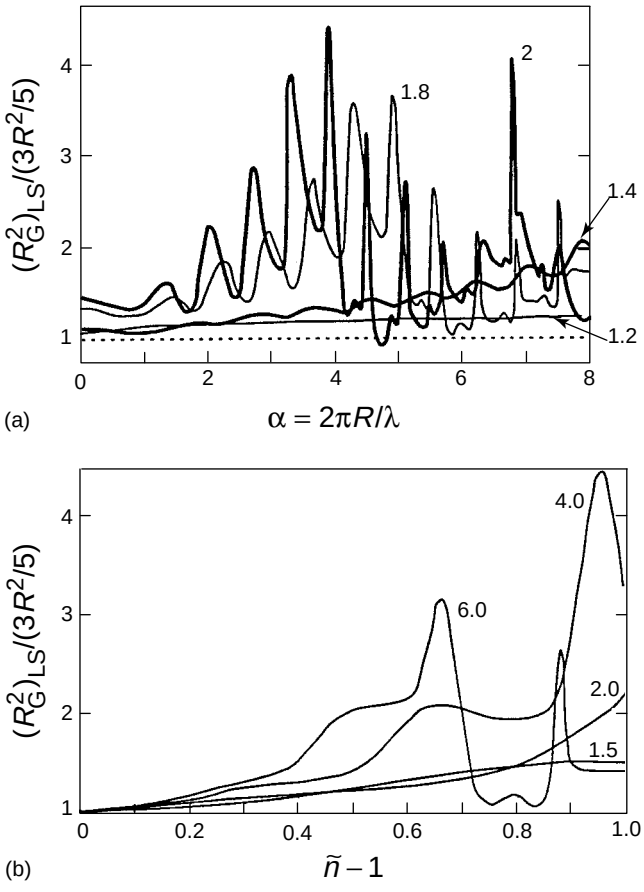

Figure 2 Ratio $(R^2_G)_{LS}/(3R^2/5)(=y_{sph}(\tilde{n}, \alpha))$ of the light scattering averaged mean-square radius of gyration $(R^2_G)_{LS}$ for monodisperse spheres of radius R to the geometric square radius of gyration as a function of (a) the size parameter $\alpha = 2\pi R/\lambda$ for the indicated values of the relative refractive index $\tilde{n}$, and (b) the relative refractive index $\tilde{n}$ for the indicated values of the size parameter $\alpha = 2\pi R/\lambda$.

$$(R^2_G)_{LS} = \left(\frac{3}{5}\right. \quad y_{sph}(\tilde{n}, \alpha)R^2 \tag{30}$$

Examples of $(R^2_G)_{LS}/\{3R^2/5\} = y_{sph}(\tilde{n}, \alpha)$ are shown in Figure 2 over a range of $\tilde{n}$ and α, showing the complex character of this function. Expansion of $y_{sph}(\tilde{n}, \alpha)$ for $\tilde{n}\alpha \leq 1$ gives a result (van de Hulst,[22] p. 45) that may be cast in the form, $y_{sph}(\tilde{n}, \alpha) \approx p_{sph}(\tilde{n}) = (\tilde{n}^2 + 2)(\tilde{n}^2 + 4)/[3(2\tilde{n}^2 + 3)]$, where $p_{sph}(\tilde{n})$ approaches unity as $\tilde{n}$ tends to unity, so that $(R^2_G)_{LS} = 3R^2/5$ as $\tilde{n}$ tends to unity and $\tilde{n}\alpha \ll 1$, as expected in the RGD regime. It may be noted that even if $\alpha \ll 1$, so that $m_{sph}(\tilde{n}, \alpha) \approx 1$ and $y_{sph}(\tilde{n}, \alpha) \approx p_{sph}(\tilde{n})$, $(R^2_G)_{LS}$ tends to $\tilde{n}^2R^2/10$ for $\tilde{n} \gg 1$ (but $\tilde{n}\alpha \leq 1$). The dependence on $\tilde{n}$ for small α is most clearly seen in Figure 2(a); the fit of this approximate relation is good for $\alpha < 1$. Since $(R^2_G)_{LS} = 3R^2/4$, independent of $\tilde{n}$, in the Fraunhofer scattering regime ($\alpha \gg 1$ and $\tilde{n} \gg 1$) discussed in the next

section (see Kerker,[21] p. 325),[43,44] $y_{\mathrm{sph}}(\tilde{n}, \alpha)$ tends to 5/4 in that limit; the values of α used in Figure 2 are too small to exhibit such behavior. Similar expressions could be developed for other particles to which the Mie theory has been applied, e.g. stratified spheres and related structures mentioned in the preceding (see Bohren and Huffman,[13] p. 475).[45,46]

As may be seen in Figure 2, the transition from the behavior for $\tilde{n} \approx 1$ and $\tilde{n}\alpha \ll 1$ to that for $\alpha \gg 1$, for any $\tilde{n}$, is complex, prohibiting an analytic representation of $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ for a sample of unknown size distribution. Thus, although Equation (31)

$$(R_{\mathrm{G}}^2)_{\mathrm{LS}} = \left(\frac{3}{5}\right) \frac{\sum_{\mu} w_\mu M_\mu y_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)[m_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)]^2 R_\mu^2}{\sum_{\mu} w_\mu M_\mu [m_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)]^2} \tag{31}$$

holds for a collection of spheres of different size, but identical $\tilde{n}$, this cannot generally be reduced to an expression involving standard average molecular weights. In principle, assuming that both $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ and M_{LS} can be determined (which may be difficult for large particles), one could compare these with values of $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ and M_{LS} calculated with an assumed two-parameter distribution function of M as a means to characterize the size distribution, e.g. for spheres, using the Mie theory to compute $y(\tilde{n}, \lambda, M_\mu) = y_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)$ and $m(\tilde{n}, \lambda, M_\mu) = m_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)$. In essence, a similar strategy is adopted by methods that attempt to fit the scattering determined at two small scattering angles (see Kerker,[21] p. 351).[44] In addition, in principle, the dependence on λ embodied in the parameters α_μ provides a means of interpreting $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ in terms of w_μ from the variation of $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ with λ using the Mie theory to compute $y_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)$ and $m_{\mathrm{sph}}(\tilde{n}, \alpha_\mu)$, after accounting for any dispersion in the refractive indices of solute or solvent.

2.3.3 Anisotropic Solute

An estimate for $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ may be obtained from the dependence of either $[K'c/\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)]^0$ or $[K'c/\mathbf{R}_{\mathrm{Hv}}(\vartheta, c)]^0$ for the scattering from anisotropic solutes in the RGD regime. Thus, for a solute with identical scattering elements, expressions for semiflexible chains give Equations (32) and (33),[10,41,47–49]

$$P_{\mathrm{Vv}}(\vartheta, 0)^{-1} = 1 + \frac{[J_{\mathrm{Vv}}(\delta) R_{\mathrm{G}}^2]_{\mathrm{LS}} q^2}{3 + \cdots} \tag{32}$$

$$P_{\mathrm{Hv}}(\vartheta, 0)^{-1} = 1 + \frac{3[J_{\mathrm{Hv}}(\delta) R_{\mathrm{G}}^2]_{\mathrm{LS}} q^2}{7 + \cdots} \tag{33}$$

Expressions for $[J_{\mathrm{Vv}}(\delta) R_{\mathrm{G}}^2]_{\mathrm{LS}}$ and $[J_{\mathrm{Hv}}(\delta) R_{\mathrm{G}}^2]_{\mathrm{LS}}$ are available for semiflexible chains.[10] In practice, the expression for rod-like chains (i.e. $\hat{a}/L > 1$) may be used as a first approximation for any $\hat{a}/L$, since $[J_{\mathrm{Vv}}(\delta) R_{\mathrm{G}}^2]_{\mathrm{LS}} \approx (R_{\mathrm{G}}^2)_{\mathrm{LS}}$ and $[J_{\mathrm{Hv}}(\delta) R_{\mathrm{G}}^2]_{\mathrm{LS}} \propto (\delta^2 R_{\mathrm{G}}^2)_{\mathrm{LS}} \approx 0$ as $\hat{a}/L$ becomes small and δ tends to zero for flexible chains;[10] see section 4, "Examples".

2.4 Scattering at Arbitrary Angle and Infinite Dilution

2.4.1 Isotropic Solute in the Rayleigh–Gans–Debye Regime

The angular dependence of the scattering over a wide angular range provides additional information on solute structure and dispersity, at least in principle. In practice, the effects of these two attributes may frustrate analysis, e.g. even in the preceding case for small scattering angle, the dependence of $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ on average molecular weights depends on the solute structure, see Table 1. For an optically isotropic solute (averaged over all orientations) in the RGD approximation,[1,10,18] Equations (34) and (35) apply

$$P_{\mathrm{Vv}}(\vartheta, 0) = \frac{[\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/K'c]^0}{M_{\mathrm{LS}} \psi^2_{\mathrm{solute}}} \tag{34}$$

$$P_{\mathrm{Vv}}(\vartheta, 0) = \sum_{\mu} w_\mu M_\mu^{-1} \quad \sum\sum_{ij} \psi_i m_i \psi_j m_j \times \langle \sin(q|r_{ij}|)/q|r_{ij}| \rangle_{\mu} \quad M_{\mathrm{LS}} \psi^2_{\mathrm{solute}} \tag{35}$$

where each sum runs over the M/m scattering elements on a component, and the second form is written in the RGD approximation. Expansion of the factor in $\langle \cdots \rangle$ brackets yields the relation in terms of $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ given above. Equations (36) and (37) apply to optically identical scattering elements with molar mass m,

$$P_{\mathrm{Vv}}(\vartheta, 0) = \frac{\sum_{\mu} w_\mu M_\mu [P_{\mathrm{Vv}}(\vartheta, 0)]_\mu}{M_{\mathrm{w}}} \tag{36}$$

$$[P_{\mathrm{Vv}}(\vartheta, 0)]_\mu = \left(\frac{m}{M}\right)_\mu^2 \sum\sum_{ij} \frac{\langle \sin(q|r_{ij}|)}{q|r_{ij}|\rangle_\mu} \tag{37}$$

where as above, the ensemble may include chains with identical M_μ but different $[P_{\mathrm{Vv}}(\vartheta, 0)]_\mu$, with w_μ the weight fraction of each distinct component.

Expressions for $P_{\mathrm{Vv}}(\vartheta, 0)$ for a number of widely used models in the RGD approximation are given in Table 2, and a selection are illustrated in Figure 3. The exponential form $P_{\mathrm{Vv}}(\vartheta, 0) = \exp(-R_{\mathrm{G}}^2 q^2/3)$ is included for comparison, as this relation is sometimes used as

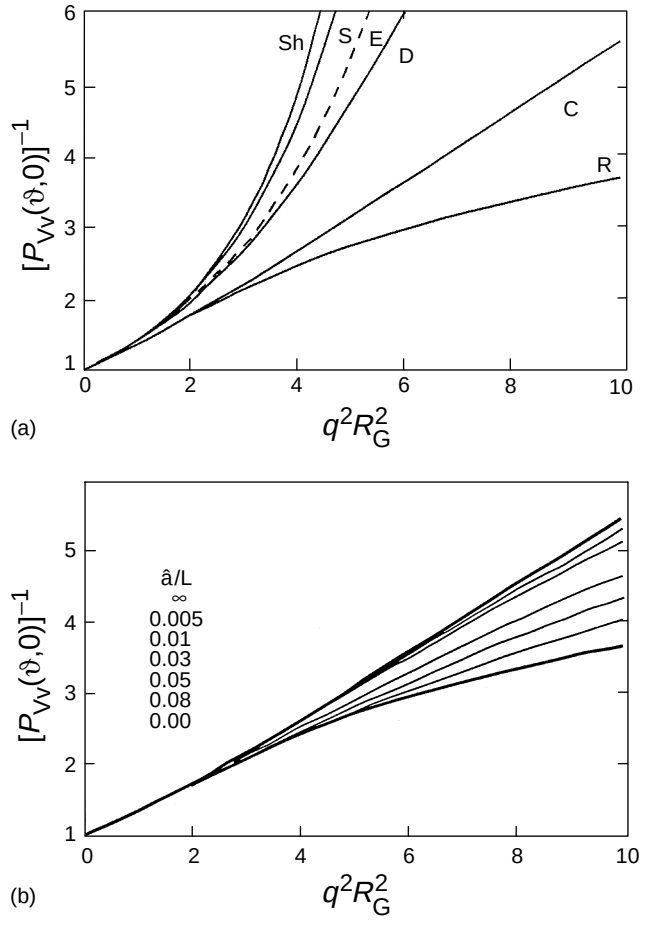

Figure 3 Examples of $P_{Vv}(\vartheta, 0)^{-1}$ versus $q^2R_G^2$ for monodisperse model structures (see Table 1). (a) Various-models: Rod-like chains (R), random-flight linear chains (C), disks (D), spheres (S), shells (Sh) and the exponential function $\exp(-q^2R_G^2/3)$ (E), included for comparison; (b) the persistent or worm-like chain, for the indicated values of the ratio of the persistence length $\hat{a}$ to the contour length L.

a convenient approximation; inspection shows that it is not usually a good approximation unless $R_G^2q^2$ is small, providing the best fit for spheres among the examples in Figure 3. Given the availability of computational equipment, it is usually not necessary to resort to tabulated numerical values of $P_{Vv}(\vartheta, 0)$ for scatterers in the RGD approximation; an extensive compilation of formulae for $P_{Vv}(\vartheta, 0)$ for a variety of macromolecular and particle shapes to supplement the entries in Table 2 may be found in Kerker (p. 482).[21] In most cases, $P_{Vv}(\vartheta, 0)$ may be expressed in terms of q in the RGD approximation, but with some structures, $P_{Vv}(\vartheta, 0)$ depends explicitly on ϑ, as well as on q, e.g. for a rod-like chain (see Table 2).

In some models, the summations in $P_{Vv}(\vartheta, 0)$ may be completed in terms of standard molecular-weight averages (M_n, etc.) without specification of a particular expression for the distribution w_μ. For example, for random-flight linear chains (e.g. at the Flory θ temperature), using $P_{Vv}(\vartheta, 0)$ in Table 2,[1,42] Equation (38)

$$P_{Vv}(\vartheta, 0) = \frac{2M_n}{M_w u_n^2} \quad u_n - 1 + \sum_\mu w_\mu \left(\frac{M_n}{M_\mu} \times \exp\left(-u_n \frac{M_\mu}{M_n} \right) \right\} \tag{38}$$

where $u_n = (R_G^2/M)M_nq^2$, with R_G^2/M a constant for this model. Further analysis requires an expression for w_μ, and this would be the normal situation with most forms for $P_{Vv}(\vartheta, 0)$. For example, in this case, using the Schulz–Zimm (two-parameter exponential) distribution function,[1,31,42] the remaining sum is equal to $(1 + u_n/h)^{-h}$, where $h^{-1} = (M_w/M_n) - 1$. For a most-probable distribution of molecular weight ($h = 1$), this reduces the expression to a simple form linear in q^2, Equation (39):[1,31,42]

$$P_{Vv}(\vartheta, 0)^{-1} = 1 + \frac{(R_G^2)_{LS}q^2}{3} \tag{39}$$

exactly, with $(R_G^2)_{LS} \propto M_z$ for this model for a linear polymer.[42,50] The same result obtains for a randomly branched flexible chain polymer.[12] An expression for $P_{Vv}(\vartheta, 0)$ for rod-like chains with the same distribution of molecular weight is available in the form of a series to be summed over $h - 1$ terms.[10] Examples of $P_{Vv}(\vartheta, 0)^{-1}$ computed for the random-flight linear chain and the sphere models using the Schulz–Zimm distribution function are shown in Figure 4.

The expression for $P_{Vv}(\vartheta, 0)$ given in Table 2 for linear flexible coil chains assumes random-flight chain statistics, and therefore is not strictly applicable to polymer chains in so-called "good solvents" (i.e. solvents for which $A_2 \gg 0$), for which excluded volume effects introduce a nonlinear dependence of $|r_{ij}|$ on the separation of scattering elements i and j, resulting in a dependence of R_G^2/L on L (see Yamakawa,[20] p. 302).[51] In practice, the expression given in Table 2 may usually be used in good solvents for the range of $(R_G^2)_{LS}q^2$ of usual interest, provided that the observed $(R_G^2)_{LS}$ is used, as opposed to the value of $(R_G^2)_{LS} = \hat{a}M_z/3$ obtaining under Flory θ conditions ($A_2 = 0$) called for by the model. Treatments attempting to include excluded volume effects by calculation of $P_{Vv}(\vartheta, 0)$ using a model for which $|r_{ij}| = |i - j|^{1+\beta}$, with $0 \leq \beta \leq 1$, have been developed as an alternative.[18,52] Similarly, although $[P_{Vv}(\vartheta, 0)]_{BR}$ has been computed for a variety of branch chain configurations,[12,53,54] the expression $[P_{Vv}(\vartheta, 0)]_{LIN}$ for linear flexible chains provides a first approximation to $[P_{Vv}(\vartheta, 0)]_{BR}$, provided $(R_G^2)_{BR}$ observed for the branched chain is used in place of $(R_G^2)_{LIN}$ appearing in $[P_{Vv}(\vartheta, 0)]_{LIN}$. Examples

Table 2 Intramolecular structure factor for several models[a] (RGD approximation)

Model	R_G^2 [b]		$P_{Vv}(\vartheta, 0)$
Random-flight linear coil	$\hat{a}L/3$	$u = \hat{a}Lq^2/3$	$p_c(u) = (2/u^2)[u - 1 + \exp(-u)]$
Persistent (worm-like) linear chain[c]	$(\hat{a}L/3)S(\hat{a}/L)$	$v = u\hat{a}/L$	$\exp(v) \quad_{i=0}(-v/i!)^{-1}p_c[(v+i)L/\hat{a}]$
Disk "infinitely thin"[d]	$R^2/2$	$y = Rq$	$(2/y^2)[1 - J_1(2y)/y]$
Sphere	$3R^2/5$	$y = Rq$	$(9/y^6)[\sin(y) - y\cos(y)]^2$
Shell "infinitely thin"	R	$y = Rq$	$[\sin(y)/y]^2$
Rod "infinitely thin"[e]	$L^2/12$	$x = Lq$	$p_1(x) = (2/x^2)[xSi(x) - 1 + \cos(x)]$

[a] With one exception, the original citations for the entries for $P_{Vv}(\vartheta, 0)$, along with expressions for a number of additional models may be found in Kerker[21] p. 482. All functions are for full orientational averaging.
[b] $\hat{a}$ is the persistence length; $L = M/M_L$ is the chain contour length (mass per unit length M_L), R is the radius for a disk or shell.
[c] $S(Z) = 1 - 3Z + 6Z^2 - 6Z^3[1 - \exp(-Z^{-1})] \approx (1 + 4Z)^{-1}$, the expression for $P_{Vv}(\vartheta, 0)$ is limited to $\hat{a}/L < 0.1$; a more complete representation may be found in Peterlin.[52]
[d] $J_1(\ldots)$ is the Bessel function of first order and kind.
[e] $Si(x) = \int_0^x ds\{\sin(s)/s\}$ is the sine integral.
For rods with optically anisotropic elements,[10,18,41]

$$\left(1 + \frac{4\delta^2}{5} \quad P_{Vv}(\vartheta, 0) = p_1(x) + \delta^2\left\{\left(\frac{4}{5} \quad p_3(x) - (2 - \delta^{-1})m_1(x) + \left(\frac{9}{8} \quad m_2(x) + m_3(x)\right\}\right.\right.$$

$$P_{Hv}(\vartheta, 0) = p_3(x) + \left(\frac{5}{8} \quad \sin^2\left(\frac{\vartheta}{2} \quad m_2(x)\right.\right.$$

$$p_2(x) = \left(\frac{6}{x^3} \quad [x - \sin(x)]\right.$$

$$p_3(x) = \left(\frac{10}{x^5} \quad [x^3 + 3x\cos(x) - 3\sin(x)]\right.$$

$$m_1 = p_1 - p_2$$

$$m_2 = 3p_1 - p_2 - 2p_3$$

$$m_3 = p_3 - p_2$$

of this are shown in Figure 5 for a range of comb-shaped branched chains with different fractions φ of monomers in the backbone and different numbers of branches f per molecule (e.g. φ is zero and unity for linear and star-shaped structures, respectively[53]); substantial deviations are noted between the approximation and the expected behavior for $(R_G^2)_{BR}q^2 > 3$. It may be noted that the ratio $g = (R_G^2)_{BR}/(R_G^2)_{LIN}$ of the R_G^2 for branched and linear chains of the same M is frequently used in discussions of branched chains, but that is a sufficient descriptor only for self-similar structures, such as star-shaped branched chains. For example, comb-shaped branched chains with different φ and f may have the same g, as exemplified by Equation (40)

$$g \approx \varphi + \frac{(1 - \varphi)^{7/3}(3f - 2)}{f^2} \tag{40}$$

which provides a useful approximation for this model.[53] A similar form applies for randomly branched polymers,[55] with $g_w \approx \varphi_w$, where φ_w is M_w for the longest linear chain in the branched structure divided by the total molecular weight, and $g_w = \left(\quad g_\mu w_\mu M_\mu\right)/M_w$. See further discussion in section 4.

In the limit of very large $(R_G^2)_{LS}q^2$, for thread-like chains in the RGD approximation[10,18,54], Equation (41) applies for a solute with $(R_G^2)_\mu = rM_\mu^\varepsilon$ (e.g. see Table 1),

$$\lim_{(R_G^2)_{LS}q^2 \gg 1}[P_{Vv}(\vartheta, 0)]_\mu^{-1} \approx s_\mu(r_\mu q^2)^{1/\varepsilon} + i_\mu + O(q^{-2}) \tag{41}$$

where i_μ and s_μ are numbers that depend on the solute structure. Thus, for a rod-like solute ($\varepsilon = 2$ and $r = (L/M)^2/12$, with L the chain contour length,[10,18,56] Equations 42 and 43)

$$\lim_{(R_G^2)_{LS}q^2 \gg 1}[P_{Vv}(\vartheta, 0)]_\mu^{-1} \approx \left(\frac{L_\mu}{\pi} \quad q + \left(\frac{2}{\pi^2} \quad + \cdots\right.\right. \tag{42}$$

$$\lim_{(R_G^2)_{LS}q^2 \gg 1} P_{Vv}(\vartheta, 0)^{-1} \approx \left(\frac{L}{\pi M} \quad M_w q + \cdots\right. \tag{43}$$

showing that data on $\mathbf{R}_{Vv}(\vartheta, 0)$ in this limit will provide the length per unit mass, but not the molecular weight; this situation may be realized with certain large rod-like structures.[56] For linear or branched flexible chain polymers ($\varepsilon = 1$ and $r = g\hat{a}L/3M$, with g the ratio of R_G^2 for branched and linear chains with the same M,[10,18,54] Equations 44 and 45)

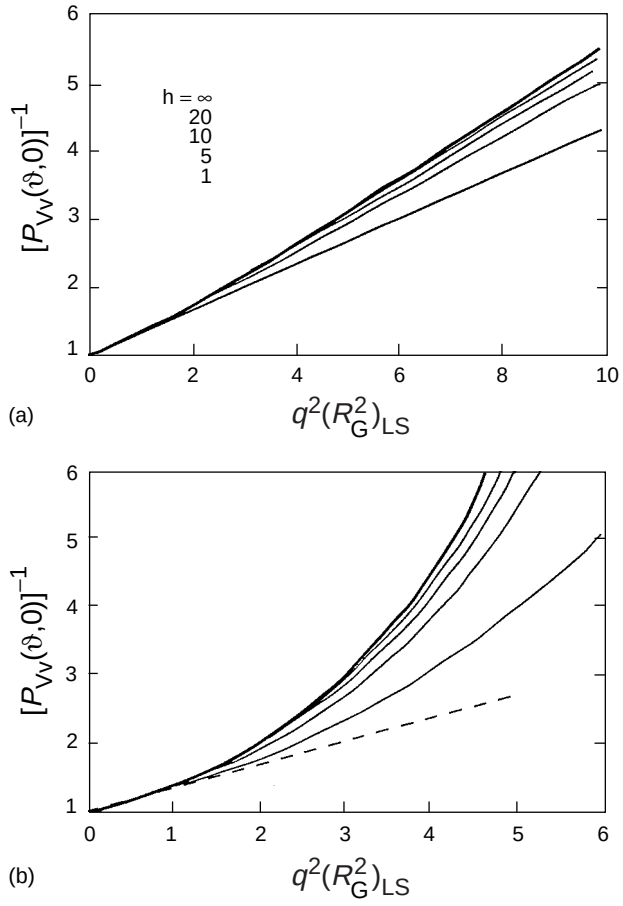

Figure 4 Examples of $P_{Vv}(\vartheta, 0)^{-1}$ versus $q^2(R_G^2)_{LS}$ for model structures with a weight distribution given by the two-parameter Zimm–Schulz distribution function, with $h = 1/[(M_w/M_n) - 1]$ as per the discussion in the text, for random-flight linear chains (a) and spheres (b), for the indicated values of h, with h decreasing from top to bottom in each set of curves (the uppermost curve for the monodisperse case is bold in each set). The dashed line in (b) is the initial tangent.

$$\lim_{(R_G^2)_{LS}q^2 \gg 1} P_{Vv}(\vartheta, 0)^{-1} \approx \frac{rM_n q^2}{2} + \left(\frac{M_w}{M_n} \quad C + \cdots \right) \tag{44}$$

$$\approx \left(\frac{M_n}{M_z} \quad (R_G^2)_{LS}\frac{q^2}{2} + \left(\frac{M_w}{M_n} \quad C + \cdots \right) \tag{45}$$

where C is 1/2 for linear chains and dependent on chain structure for branched chains. Consequently, in principle, investigation of the limiting behavior at large and small q can provide values for M_n, M_w and M_z for linear flexible chain polymers (or for branched flexible chains for which C is known). In practice, the limiting behavior at large q is seldom observed in light-scattering experiments with flexible chain macromolecules, but it is possible to observe such behavior with low-angle X-ray or neutron scattering – it may, however, be difficult to observe the limiting behavior for large and small q in the same experimental arrangement. Moreover, short-scale chain features not included in the models used to compute $P_{Vv}(\vartheta, 0)$ will introduce behavior not included in these asymptotic expressions for large $(R_G^2)_{LS}q^2$. A related behavior obtains with spheres, in which the value of $P_{Vv}(\vartheta, 0)$ for successive maxima at large qR decay as $(qR)^{-4}$ in the so-called Porod regime (see Kerker,[21] p. 471; Guinier and Fournet,[24] p. 17).[57,58]

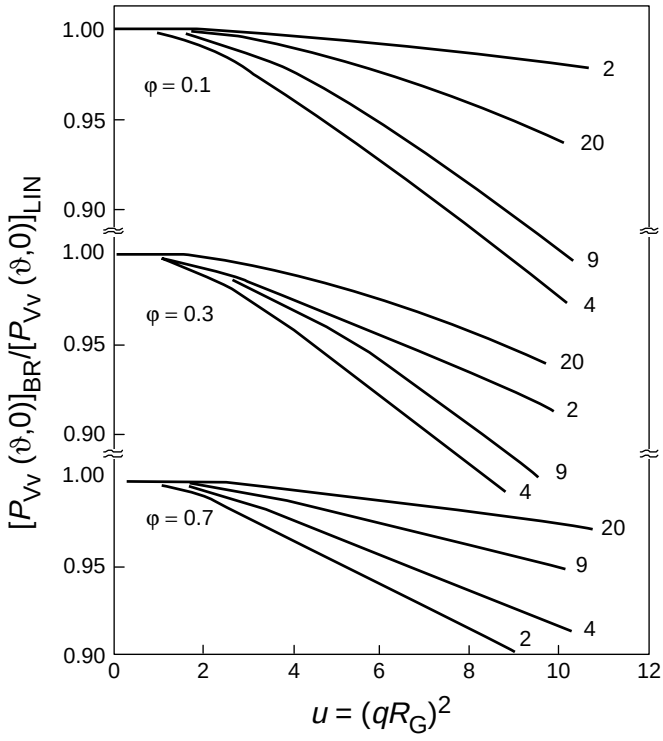

Figure 5 Examples of $[P_{Vv}(\vartheta, 0)]_{BR}$ versus $q^2R_G^2$ for comb-shaped branched chain polymers divided by $[P_{Vv}(\vartheta, 0)]_{LIN}$ for linear chains with the same R_G^2 (not the same molecular weight). The number of branches is indicated, along with the fraction φ of mass in the backbone of the branched chain. (From Casassa and Berry.[53])

For a solute monodisperse in molecular weight, but comprising two types of scattering elements,[10,36,37] Equations (46) and (47) apply

$$P_{Vv}(\vartheta, 0) = x[P_{Vv}(\vartheta, 0)]_A + (1 - x)[P_{Vv}(\vartheta, 0)]_B + x(1 - x)[Q_{Vv}(\vartheta, 0)]_{AB} \tag{46}$$

$$[Q_{Vv}(\vartheta, 0)]_{AB} = 2[P_{Vv}(\vartheta, 0)]_{AB} - \{[P_{Vv}(\vartheta, 0)]_A + [P_{Vv}(\vartheta, 0)]_B\} \tag{47}$$

where $x = w_A\psi_A/\psi_{solute}$ as above, and $\psi_{solute} \neq 0$. Here, $[P_{Vv}(\vartheta, 0)]_A$ and $[P_{Vv}(\vartheta, 0)]_B$ are calculated by the

expression given above, and $[P_{Vv}(\vartheta, 0)]_{AB}$ by a variation of this expression with one sum each over scattering elements of type A and B. The expressions for $P_{Vv}(\vartheta, 0)$ in Table 2 apply for $[P_{Vv}(\vartheta, 0)]_A$ and $[P_{Vv}(\vartheta, 0)]_B$, but $[P_{Vv}(\vartheta, 0)]_{AB}$ must be computed for each case. Expansion of $[Q_{Vv}(\vartheta, 0)]_{AB}$ gives the leading term $\Delta_{AB}^2 q^2/3$, with Δ_{AB}^2 as defined above, i.e. $[Q_{Vv}(0, 0)]_{AB} = 0$. Examples are available for random-flight linear chains.[36] Inspection shows that the initial tangent $\partial[P_{Vv}(\vartheta, 0)]^{-1}/\partial q^2$ may be negative, zero, or positive, reflecting the possibility mentioned above that $(R_G^2)_{LS}$ may be positive, zero, or negative, respectively, for such a solute. For a sphere with radius R_A and scattering elements of type A coated by a shell with outer radius $R_B > R_A$ and scattering elements of type B (Kerker,[21] p. 419) (Equation 48)

$$P_{Vv}(\vartheta, 0) = \left(x[P_{Vv}(\vartheta, 0)]_A^{1/2} + (1-x) \times \frac{R_B^3[P_{Vv}(\vartheta, 0)]_B^{1/2} - R_A^3[P_{Vv}(\vartheta, 0)]_A^{1/2}}{R_B^3 - R_A^3} \right)^2 \quad (48)$$

where the expressions for $[P_{Vv}(\vartheta, 0)]_B$ and $[P_{Vv}(\vartheta, 0)]_A$ are the functions for spheres with radii R_B and R_A, respectively (given in Table 2). This expression reduces to $P_{Vv}(\vartheta, 0)$ for a shell of thickness $\Delta = R_B - R_A$ filled with the solvent for $x = 0$, or to $P_{Vv}(\vartheta, 0)$ for a sphere of radius R_A if $x = 1$. In using this relation, it must be remembered that $w_A \propto \rho_A R_A^3$ and $w_B = 1 - w_A \propto \rho_B(R_B^3 - R_A^3)$. If $\psi_{solute} = 0$, then $[\mathbf{R}_{Vv}(0, c)/Kc]^0 = 0$, but $[\mathbf{R}_{Vv}(\vartheta, c)/Kc]^0$ for $\vartheta > 0$ may be computed using Equation (48) with x and $1 - x$ replaced by $w_A\psi_A$ and $w_B\psi_B$, respectively. The result will exhibit a series of maxima for $\vartheta > 0$, with a superficial similarity to $\mathbf{R}_{Vv}(\vartheta, c)$ observed with charged spheres, arising from interactions among the scatterers, see section 4.

2.4.2 *Isotropic Solute Beyond the Rayleigh–Gans–Debye Regime*

Although the accuracy of the RGD approximation varies with ϑ and $\tilde{n}$, deviations from the Mie theory are usually small if $\alpha|\tilde{n} - 1| < 0.25$ (but this depends on ϑ and both α and $\tilde{n}$ individually in detail) (Kerker,[21] p. 427; van de Hulst,[22] p. 131). With the RGD approximation, $P_{Hh}(\vartheta, 0) = P_{Vv}(\vartheta, 0)\cos^2(\vartheta)$, but this relation does not usually hold if the phase shift is not small. The Mie theory provides expressions for $P_{Vv}(\vartheta, 0)$, $P_{Hh}(\vartheta, 0)$, and $P_{Hv}(\vartheta, 0)$ for optically isotropic nonabsorbing spheres as the square of sums over functions of ϑ and the relevant refractive indices (see Bohren and Huffman,[13] p. 113; Kerker,[21] p. 97; van de Hulst,[22] p. 131).[8] Examples shown in Figure 6 give $P_{Vv}(\vartheta, 0)$ as a function of $q^2(R_G^2)_{LS}$ for a particular value of $\alpha = 2\pi R/\lambda$, over a range of $\tilde{n}$, with

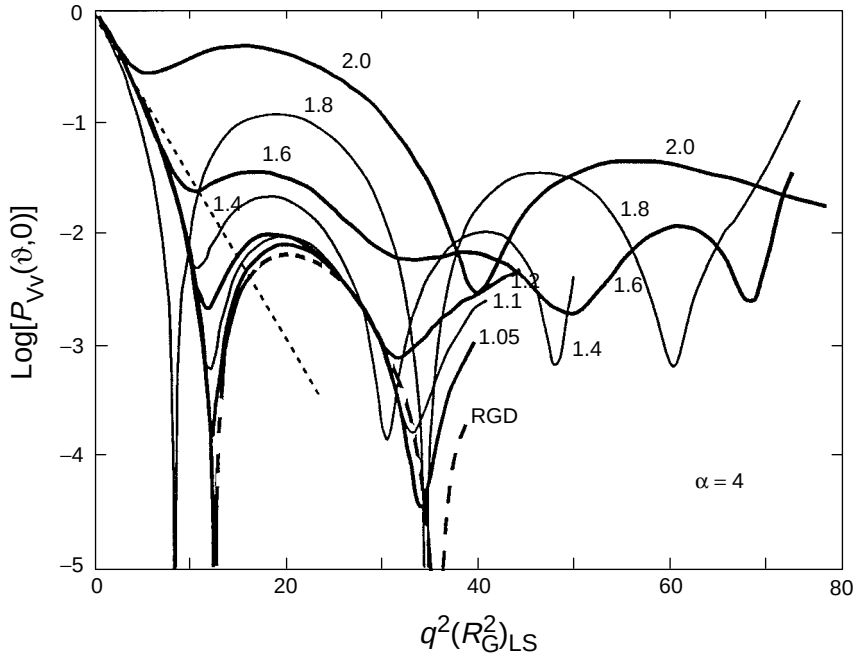

Figure 6 Examples of $\log[P_{Vv}(\vartheta, 0)]$ versus $q^2(R_G^2)_{LS}$ for spheres with size parameter $\alpha = 2\pi R/\lambda = 4$ for the indicated values of the relative refractive index $\tilde{n}$. The angular range is 0 to 180 degrees in all cases except for $\tilde{n} = 2$. The RGD limiting case for very small $\tilde{n} - 1$ is given by the bold dashed curve. The dashed line gives the initial tangent. Values of $(R_G^2)_{LS}/(3R^2/5)$ may be seen for this α in Figure 2.

$\lambda = \lambda_0/n_{\text{medium}}$ (e.g. the α value chosen in the example would correspond to aqueous suspensions of spheres with $R \approx 300\,\text{nm}$, with $\lambda_0 = 633\,\text{nm}$). The extreme dependence on $\tilde{n}$ may be observed, emphasizing the need to know this parameter accurately in this regime, and showing that it would be impossible to analyze the size distribution for an ensemble of spheres with heterogeneity of both $\tilde{n}$ and R. The data in Figure 6 are shown as $\log[P_{\text{Vv}}(\vartheta, 0)]$ versus $q^2(R_{\text{G}}^2)_{\text{LS}}$ to emphasize the correspondence at small q, but the systematic variation of $q^2(R_{\text{G}}^2)_{\text{LS}}$ for the first minimum in $P_{\text{Vv}}(\vartheta, 0)$ with $\tilde{n}$. In this format, the minima in $P_{\text{Vv}}(\vartheta, 0)$ tend to occur near the values of $q^2(R_{\text{G}}^2)_{\text{LS}}$ for the minima in the RGD approximation. Similar behavior may be observed for other values of α, including α large enough to be in the Fraunhofer diffraction limit discussed below.

Numerical methods to compute $P_{\text{Vv}}(\vartheta, 0)$ are available for a number of particle shapes (see Bohren and Huffman,[13] p. 475).[8,59,60] Fortunately, electronic computation at the desktop can facilitate the use of these, although extensive tables of the functions appearing in the Mie theory for the scattering from spherical particles are available (Kerker,[21] p. 78) (the functions reported herein were calculated using a MathCad file, kindly provided by D. C. Prieve, of the Department of Chemical Engineering of Carnegie Mellon University, with the number of terms in the relevant sums terminated at the next integer larger than $\alpha + 4\alpha^{1/3} + 4$ (slightly larger than an expression advocated in the literature (see Bohren and Huffman,[13] p. 477) for accessible α); data in published tables[61–63] are useful to confirm the methods used in such calculations). Comparable expressions are available in the Mie approximation for shells, stratified spheres, and examples with a continuous radially symmetric variation (see Bohren and Huffman,[13] p. 475; Kerker,[21] p. 189)[26,64–66] cylinders and ellipsoids (Kerker,[21] p. 255).[67–69] The application of these for polydisperse systems requires the use of Equation (36) to compute the observable $P_{\text{Vv}}(\vartheta, 0)$, usually by a numerical analysis.

For both $\tilde{n} \gg 1$ and phase shift magnitude $\alpha|\tilde{n} - 1| > 10$, the angular dependence reduces to a regime termed the Fraunhofer diffraction limit (see Bohren and Huffman,[13] p. 482; Kerker,[21] p. 159; van de Hulst,[22] p. 108).[11] In that regime, the angular dependence is independent of $\tilde{n}$, and the same for absorbing or nonabsorbing particles. For example, for a spherical solute, the angular dependence of the scattering (for a monodisperse solute at infinite dilution) is that for Fraunhofer diffraction from a circular aperture, Equation (49)

$$\lim_{R/\lambda \gg 1} P_{\text{Vv}}(\vartheta, 0) = \left\{ \frac{2J_1(\alpha \sin(\vartheta))}{\alpha \sin(\vartheta)} \right\}^2 \tag{49}$$

where $\alpha = 2\pi R/\lambda$, $J_1(\cdots)$ is the Bessel function of the first kind and first order. For a polydisperse spherical solute in this regime, Equation (50)

$$\lim_{R/\lambda \gg 1} \frac{\mathbf{R}_{\text{Vv}}(\vartheta, c)}{c}^{0} \propto \sum_{\mu} n_\mu \alpha_\mu^6 \left\{ \frac{2J_1(\alpha_\mu \sin(\vartheta))}{\alpha_\mu^2 \sin(\vartheta)} \right\}^2 \tag{50}$$

where n_μ is the number fraction of spheres with radius R_μ. As discussed below, the simplicity of this result can be exploited in methods to determine the size distribution from $[\mathbf{R}_{\text{Vv}}(\vartheta, c)/c]_0$. Similar results obtain for other particle shapes. For example, for cylinders with length L_{cyl} and radius R, in the same limit, and with $L_{\text{cyl}} \gg R \gg \lambda$ (van de Hulst,[22] p. 108), Equation (51)

$$\lim_{R/\lambda \gg 1} P_{\text{Vv}}(\vartheta, 0) = \frac{\sin^2(\alpha\vartheta)}{(\alpha\vartheta)^3} \tag{51}$$

Since the patterns of the extrema are similar for these two functions, it would not be possible to differentiate between solutes with these two shapes from the scattering in this limit. Nor can the expression for cylinders with $L_{\text{cyl}}/R \gg 1$ be used to determine the distribution of cylinder lengths.

It has been observed that in the regime for which Mie scattering theory applies for optically isotropic spheres, the angular dependence for small angles may be approximated by the expression for the Fraunhofer regime given above, even if R is not much larger than λ (see Kerker,[21] p. 351);[44] Although the discussion of $(R_{\text{G}}^2)_{\text{LS}}$ in the preceding shows that this cannot be accurate for small angles (and is inaccurate for the RGD regime), it may, nevertheless, be a useful approximation over most of the regime for which the Mie theory is needed, offering a substantial simplification in the analysis of the size distribution in such cases.

If $\alpha \gg 1$, but $\tilde{n} \approx 1$ (e.g. for a particle immersed in a solvent with closely matching refractive index), so that the phase shift $\alpha|\tilde{n} - 1|$ is small, the angular dependence differs considerably from that for Fraunhofer diffraction for larger ϑ, in a regime called anomalous diffraction (see van de Hulst,[22] p. 183)[21] giving what has been characterized as "a set of queer scattering diagrams" (van de Hulst,[22] p. 183). So-called anomalous scattering approximations to the Mie theory have been obtained in this regime for spheres (van de Hulst,[22] p. 172), reported to be accurate for computation of $P_{\text{Vv}}(\vartheta, 0)$ in the regime $\tilde{n} - 1 < 0.1$, $2\alpha|\tilde{n} - 1| < 0.1$, and $\alpha > 1$.[70] Use of the Fraunhofer diffraction approximation would, for example, give an erroneous estimate of the size distribution, see section 4. It is reported that depolarized scattering may result from isotropic spheres in this regime.[71] The preceding discussion has been limited to scattering with polarization of the incident and scattered beams either perpendicular or parallel to the scattering plane, the plane defined by incident and scattered beams. A more general notation is needed to discuss the anisotropic scattering for spheres since $P_{\text{Hv}}(\vartheta, 0)$ as

defined above is zero for isotropic spheres. In a more general description, the scattering plane may contain the incident beam and be at an azimuthal angle φ relative to the polarization of the incident light for vertically polarized light. Thus, $\varphi = \pi/2$ for the scattering plane considered thus far. An alternative nomenclature lets Vv and Hv indicate parallel and crossed polars, respectively, as in the preceding, but supplements the description by adding the azimuthal angle, e.g. $P_{\mathrm{Hv}}(\vartheta, \varphi, c)$, so that $P_{\mathrm{Hv}}(\vartheta, 0)$, $P_{\mathrm{Vv}}(\vartheta, 0)$, and $P_{\mathrm{Hh}}(\vartheta, 0)$ discussed above become $P_{\mathrm{Hv}}(\vartheta, \pi/2, 0)$, $P_{\mathrm{Vv}}(\vartheta, \pi/2, 0)$, and $P_{\mathrm{Vv}}(\vartheta, 0, 0)$, respectively. It has long been known that $P_{\mathrm{Hv}}(\vartheta, \pi/4, 0)$ is not zero for optically anisotropic spheres (see below), with a first maximum in $P_{\mathrm{Hv}}(\vartheta, \pi/4, 0)$ occurring at an angle ϑ that may be used to estimate the size of the sphere, see below.[5] Calculations in the RGD approximation and using the Mie theory reveal that a similar maximum occurs with optically isotropic spheres in the anomalous diffraction regime,[71] in agreement with observations on spheres thought to be optically isotropic.[70]

2.4.3 Anisotropic Solute

With the exception of semiflexible chains with relatively large $\hat{a}/L$, anisotropic scattering may normally be neglected in the excess scattering for polymer solutions (note that many solvents will exhibit anisotropic scattering, which must be accurately determined to study the excess anisotropic scattering due to the solute). The functions $P_{\mathrm{Hv}}(\vartheta, 0)$ and $P_{\mathrm{Vv}}(\vartheta, 0)$ for rod-like chains are given in Table 2, and are discussed in more detail below in an example, along with the anisotropic scattering from particles.

Expressions for $P_{\mathrm{Vv}}(\vartheta, 0)$ and $P_{\mathrm{Hv}}(\vartheta, 0)$ for optically anisotropic rod-like molecules in the RGD regime are given in Table 2, and discussed below in section 4. The functions $P_{\mathrm{Hv}}(\vartheta, \varphi, 0)$ and $P_{\mathrm{Vv}}(\vartheta, \varphi, 0)$ defined above have been computed for optically anisotropic spheres and other particle shapes.[5,26,72] A principal result for anisotropic spheres is that $P_{\mathrm{Hv}}(\vartheta, \pi/4, 0)$ exhibits extrema as a function of qR for monodisperse spheres (see Kerker,[21] p. 674),[5,45,70,73–78] see section 4.

2.5 The Size Distribution from Scattering Data at Infinite Dilution

In the preceding, the characterization of the size distribution is through averages of M and R_{G}^2. Thus, with a model relating R_{G}^2 to M, the different averages in M_{LS} and $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ can permit an estimate of the breadth of the size distribution. Two principal methods are used to estimate more detailed information on the size distribution: (i) comparison of experimental $P_{\mathrm{Vv}}(\vartheta, 0)$ with that calculated with an assumed distribution of M (or some other equivalent size parameter, such as R for spheres), and (ii) inversion of $P_{\mathrm{Vv}}(\vartheta, 0)$ to obtain the size distribution function. Neither is entirely satisfactory, each with its own set of limitations. In either case, it is necessary to have a model for $P_{\mathrm{Vv}}(\vartheta, 0)$, and to assume that the model applies to all scatterers in the system. Further, as a practical matter, it must be assumed that all components are identical in composition. In practice, data on $\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/Kc$ are often used to estimate a size distribution, rather than $P_{\mathrm{Vv}}(\vartheta, 0)$, or even $[\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/Kc]^0$. As discussed in the next section, this practice may introduce distortions into the size estimate unless $2(A_2)_{\mathrm{LS}}M_{\mathrm{w}}c \ll 1$. Multiple scattering is an additional practical issue sometimes encountered, especially in the scattering from dispersions of particles for which $\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/Kc$ is not extrapolated to infinite dilution. Some schemes have been introduced to account for this effect, especially in the scattering from large spheres,[79–83] but in general it should be avoided by reducing the solute concentration, though this can have a deleterious effect on the estimate of the population of very small scattering components in the presence of large particles.

In the first method, a model is used to compute $P_{\mathrm{Vv}}(\vartheta, 0)$ for comparison with the experimental data, based on an (assumed) solute structure and an assumed two or three parameter molecular-weight distribution function, e.g. one of the several functions given by Berry[42] with parameters M_{w}/M_n, M_z/M_{w}, etc. or alternatively, for spheres, these functions are expressed as a distribution in R with parameters R_{w}/R_n, R_z/R_{w}, etc. Estimates of M_{w} and $(R_{\mathrm{G}}^2)_{\mathrm{LS}}$ may be used to guide estimation of the parameters in the distribution function if these are available; in some cases, for a solute with very large M_{w} it may not be possible to obtain the reliable extrapolation to zero scattering angle needed to estimate these parameters. For example, experimental data on $P_{\mathrm{Vv}}(\vartheta, 0)$ for linear flexible chains or spheres (in the RGD regime) could be compared with Figure 4, calculated using the two-parameter exponential function mentioned above, to estimate the molecular-weight dispersion. Comparison with $P_{\mathrm{Vv}}(\vartheta, 0)$ for semiflexible chains provides an example of the ambiguity in matching experimental and calculated $P_{\mathrm{Vv}}(\vartheta, 0)$ unless a reliable model is available. A similar situation obtains with particles, e.g. deviations from spherical symmetry will result in distortions of the scattering function similar to those caused by a distribution of size in spheres.

There is a long history of attempts to implement a direct inversion of $P_{\mathrm{Vv}}(\vartheta, 0)$ (or $[\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)/c]^0$) for polydisperse scatterers (usually particles with spherical symmetry) to obtain a size distribution of the scattering species, much of it demonstrating the difficulty of obtaining reliable results. Present-day treatments are facilitated by the availability of computational facilities at the desk

top to implement the computations, with almost all applications being to the size distribution of particles owing to the nature and limited accessible span of $P_{Vv}(\vartheta, 0)$ for macromolecules. Two principal methods are used, depending on the size of the particles: numerical inversion of the data via iterative techniques,[84–90] and evaluation of an integral expression based on Fraunhofer diffraction for particles with dimensions much greater than the wavelength, $\tilde{n} \gg 1$.[81,91–101] The numerical inversions required are similar in kind to those applied to estimate size distributions from dynamic light scattering data, which involves a similar inversion relation, with similar limitations. For example, inversion is frequently attempted for the normalized electric-field correlation function $g^{(1)}_{LS}(\tau; \vartheta, c)$ of dynamic light scattering, given by[1,10,17,88,102,103] Equation (52) (see also Chu,[7] p. 243)

$$g^{(1)}_{LS}(\tau; \vartheta) = \frac{\sum_{\mu} w_\mu M_\mu [P_{Vv}(\vartheta, 0)]_\mu \exp[-\tau \gamma_\mu(q)]}{M_w P_{Vv}(\vartheta, 0)} \qquad (52)$$

in the limit of infinite dilution, with $\gamma_\mu(q)$ the relaxation rate of the component μ. The similarity of this to the expression, Equation (53)

$$P_{Vv}(\vartheta, 0) = \frac{\sum_{\mu} w_\mu M_\mu [P_{Vv}(\vartheta, 0)]_\mu}{M_w} \qquad (53)$$

of interest here is evident, with a principal distinction being that the range of τ may be made much larger than the corresponding range in ϑ (or q), to the advantage of accuracy in the inversion. On the other hand, data acquisition is usually more time-consuming in dynamic scattering. In some cases, it is convenient to replace w_μ by the number fraction n_μ of species with molecular weight M_μ, with $w_\mu = n_\mu M_\mu / M_n$, but in either case, implementation of an inversion protocol requires a relation between M_μ and a size parameter appearing in $[P_{Vv}(\vartheta, 0)]_\mu$, e.g. as in Table 1. Furthermore, the inversions are ill-posed for both static and dynamic scattering functions and constraints must be imposed to obtain a stable inversion.

It may be noted that expressions of this kind have been used for many years in the analysis of X-ray scattering data on particles, often with $P_{Vv}(\vartheta, 0) \approx \exp(-R^2_G q^2/3)$, along with a model to express R^2_G in terms of M, e.g. $R^2_G \propto M^{2/3}$ for a spherical particle (Guinier and Fournet,[24] p. 151). The several difficulties that hamper the determination of the molecular-weight (or size) distribution by an inversion of $P_{Vv}(\vartheta, 0)$, include:

1. the ill-posed character of the inversion
2. the need to have a theoretical $P_{Vv}(\vartheta, 0)$
3. noise in the data
4. the physical range of ϑ, and hence the limited accessible q
5. uncertainty in the relative refractive index $\tilde{n} = n_{solute}/n_{medium}$
6. the effects of solute aggregation
7. the effects of multiple scattering.

Various procedures have been investigated to effect the inversion, including versions of the methods used to invert dynamic scattering data, such as the constrained regularization method called CONTIN,[102,103] as well as other methods (see Chu,[7] p. 277).[85–89,98,99,104–108] As discussed above, the RGD approximation is generally valid for a nonabsorbing solute if the phase shift within the molecule or particle is small enough. Although this is generally met for a polymeric solute, a direct inversion of $P_{Vv}(\vartheta, 0)$ is seldom of interest in such cases owing to the limited range of $(R^2_G)_{LS} q^2$ typically accessible for the scattering for a polymeric solute, as well as the limitations due to noise in the experimental data, e.g. see the comments above on the difficulties of obtaining reliable behavior at large q. By contrast, with larger particulate solutes, $P_{Vv}(\vartheta, 0)$ may typically cover a wider range, and perhaps exhibit maxima and minima, for the accessible range of $(R^2_G)_{LS} q^2$. As discussed above, the RGD approximation may be used for a range of particle sizes, but in such cases, the range of $(R^2_G)_{LS} q^2$ may not be large, in which case the inversion may not be accurate, depending on the particle shape. For larger phase shift, M_w must be replaced by M_{LS}, and the Mie theory may be used for spheres to compute the $[P_{Vv}(\vartheta, 0)]_\mu$ needed in the inversion for particles with spherical symmetry, albeit with an increase in the computational time required. Similar treatments are available for other shapes, including spherical shells (Bohren and Huffman,[13] p. 109),[45] spheroids of revolution[46] and rods,[67] but it is not possible to permit both the particle shape and the size distribution to be unknowns in an inversion of the scattering from a polydisperse sample. Inversions in this regime may be reasonably stable and accurate, but usually require an accurate estimate of n_{solute}/n_{medium}, along with the attention of one skilled in the art to avoid erroneous results, see section 4.

The expression for $[\mathbf{R}_{Vv}(\vartheta, c)/c]^0$ given above in the Fraunhofer limit, with $R \gg \lambda$, may be used to obtain an integral equation for the number fraction of spheres with a given radius, eliminating the interactive methods described above. Experimental instruments to acquire data for this purpose differ in their design and in the form of the integral expression used.[43,81,91–95,98–101,109–111] A systematic, consistent analysis of these given by Koo and Hirleman[92] is discussed below in section 4.

2.6 Extrapolation to Infinite Dilution

All of the preceding has presumed that $\mathbf{R}_{\text{iso}}(\vartheta, c)/c$ and $\mathbf{R}_{\text{aniso}}(\vartheta, c)/c$ may be reliably extrapolated to infinite dilution. For dilute solutions, as discussed in section 2.1, $\mathbf{R}_{\text{iso}}(\vartheta, c)/c$ may be expressed in the form of Equation (54)

$$\frac{\mathbf{R}_{\text{iso}}(\vartheta, c)}{KMc} = S_{\text{iso}}(\vartheta, c) = \frac{P_{\text{iso}}(\vartheta, c)}{\{1 + c\Gamma_{\text{iso}}(c)P_{\text{iso}}(\vartheta, c)H_{\text{iso}}(\vartheta, c)\}} \tag{54}$$

where $K = K'\psi^2_{\text{solute}}$, with a similar expression for $\mathbf{R}_{\text{aniso}}(\vartheta, c)/c$. The function $H_{\text{iso}}(\vartheta, c)$, which is usually close to unity, is discussed below, as is the possible dependence of $P_{\text{iso}}(\vartheta, c)$ on c, the latter principally reflecting changes in the mean-square radius of gyration with c.[4] Since $P_{\text{iso}}(\vartheta, c) \ll 1$ for $q^2(R^2_{\text{G}})_{\text{LS}} \gg 1$, one can anticipate that $\mathbf{R}_{\text{iso}}(\vartheta, c)/KMc \approx P_{\text{iso}}(\vartheta, c)$, for any concentration in that limit. The range of $(R^2_{\text{G}})_{\text{LS}}q^2$ required to observe this convergence depends strongly on the solute structure.

For $\vartheta = 0$, $P_{\text{iso}}(0, c) = H_{\text{iso}}(0, c) = 1$, and for dilute solutions, a virial expansion may usually be used to represent $\Gamma_{\text{iso}}(c)$, with the result (see Yamakawa,[20] p. 186)[4,10,31] (Equation 55)

$$\frac{Kc}{\mathbf{R}_{\text{iso}}(0, c)} = M^{-1}_{\text{LS}}\{1 + 2(A_2)_{\text{LS}}M_{\text{LS}}c + 3(A_3)_{\text{LS}}M_{\text{LS}}c^2 + \cdots\} \tag{55}$$

For systems with large A_2, "good solvents", the term in c^2 may introduce appreciable curvature, motivating the use of the relation[4,10] Equation (56)

$$\left(\frac{Kc}{\mathbf{R}_{\text{iso}}(0, c)}\right)^{1/2} = M^{-1/2}_{\text{LS}}\{1 + (A_2)_{\text{LS}}M_{\text{LS}}c + \alpha_3((A_2)_{\text{LS}}M_{\text{LS}}c)^2 + \cdots\} \tag{56}$$

where $\alpha_3 = \{3(A_3)_{\text{LS}}M_{\text{LS}}/((A_2)_{\text{LS}}M_{\text{LS}})^2 - 1\}/2$; α_3 tends to be small for systems with large A_2, making the extrapolation more nearly linear with larger $(A_2)_{\text{LS}}M_{\text{LS}}c$ than with the first form. Possible exceptions to the behavior given by these relations may obtain if the solute is charged and the solvent is of very low ionic strength.[16,32]

Although the behavior for more concentrated solutions is outside the scope of this article, it may be noted that expressions are available for moderately concentrated solutions, both in good solvents, and under Flory θ conditions, for which A_2, but not A_3, is zero. In the latter case, the expression given above terminated at the term in c^2 provides a good fit to the data, whereas with good solvents, power-law expressions in A_2Mc may be used to replace the factor in $\{\cdots\}$ in Equation (55)[4,33]

The expression for $(A_2)_{\text{LS}}$ involves a double summation over components, and need not be elaborated here (however, see the example on association, section 4), except to note that for a polydisperse solute, $(A_2)_{\text{LS}}$ is not equal to the corresponding value $(A_2)_{\Pi}$ obtained by osmotic pressure.[4] Further, it may be noted that for polymer chains, for monodisperse solute, A_2 may be expressed in the form of Equation (57) (Yamakawa,[2] p. 289; Yamakawa,[20] p. 134),[4,10,51]

$$A_2 = (A_2)_{\text{ROD}}F\left(\frac{L}{\hat{a}}, z\right) \tag{57}$$

where $(A_2)_{\text{ROD}} = (\pi N_{\text{A}}L^2/4M^2)d_{\text{thermo}}$ is the second virial coefficient for a rod-like chain with thermodynamic diameter d_{thermo}, $z = (3d_{\text{thermo}}/16\hat{a})(3L/\pi\hat{a})^{1/2}$ is a thermodynamic interaction parameter, and $F(L/\hat{a}, z)$ is a function that depends on chain conformation, with $F(L/\hat{a}, 0) = F(0, z) = 1$. For uncharged solute, d_{thermo} is zero at the Flory θ temperature, and increases to about twice the geometric diameter d_{geo} for uncharged solute in so-called good solvents. In good solvents, $A_2M/[\eta] \approx 1$ for linear flexible chains, with $[\eta]$ the intrinsic viscosity.[51] By comparison, for particles interacting through a hard-core potential, (Yamakawa,[20] p. 181), Equation (58) applies

$$A_2 = \frac{4N_{\text{A}}fV_{\text{particle}}}{M^2} \tag{58}$$

where V_{particle} is the particle volume and f is a parameter that depends on the particle shape, e.g. $f = 1$ for a sphere, and $f = L/4d_{\text{geo}}$ for a rod of diameter $d_{\text{geo}} \ll L$; $A_2M/[\eta] \approx 16/5$ for spheres. The behavior with charged solutes or particles is discussed in section 4.

Similar relations obtain for optically anisotropic polymers, but the coefficients are generally smaller, reflecting the origin of the scattering from fluctuations in the orientation of the scattering elements. Thus, for rod-like chains, (Yamakawa,[20] p. 227)[4,10,112] (Equations 59 and 60)

$$\Gamma_{\text{Vv}}(c) = \{2(A_2)_{\text{LS}}M_{\text{LS}} + \cdots\} + \left(\frac{4\delta^2}{5}\right)\Gamma_{\text{Vh}}(c) \tag{59}$$

$$\Gamma_{\text{Vh}}(c) = \frac{(A_2)_{\text{LS}}M_{\text{LS}}}{4} + \cdots \tag{60}$$

or an eight-fold smaller leading term for depolarized scattering than obtains with polarized scattering.

The possible dependence of $P_{\text{iso}}(\vartheta, c)$ and $H_{\text{iso}}(\vartheta, c)$ on c may complicate the extrapolation of data to infinite dilution, or the interpretation of data at a single concentration. For example, the initial tangent, given by Equations (61) and (62)

$$\frac{\partial[Kc/\mathbf{R}_{\text{iso}}(\vartheta, c)]}{\partial q^2} = M^{-1}_{\text{LS}}\left\{\frac{\partial P_{\text{iso}}(\vartheta, c)^{-1}}{\partial q^2} + c\Gamma_{\text{iso}}(c)\frac{\partial H_{\text{iso}}(\vartheta, c)}{\partial q^2}\right\} \tag{61}$$

$$\frac{\partial[Kc/\mathbf{R}_{\rm iso}(\vartheta, c)]}{\partial q^2} = M_{\rm LS}^{-1}\left\{\frac{\partial P_{\rm iso}(\vartheta, 0)^{-1}}{\partial q^2} + c\Gamma_{\rm iso}(c)\,\frac{\partial W_{\rm iso}(\vartheta, c)}{\partial q^2}\right\} \quad (62)$$

may depend on c (the potential dependence of $P_{\rm iso}(\vartheta, c)$ on c is subsumed in the function $W_{\rm iso}(\vartheta, c)$ in the second form); $\Gamma_{\rm iso}$ may be expressed as a virial expansion in dilute solutions. For the range of $(R_{\rm G}^2)_{\rm LS}q^2$ normally accessible with polymer chains, plots of $KMc/\mathbf{R}_{\rm iso}(\vartheta, c)$ versus q^2 are usually observed to be essentially parallel for dilute solutions over a range of c, suggesting that $\partial W_{\rm iso}(\vartheta, c)/\partial q^2$ is small – exceptions may occur for chains with very large M, and the corresponding larger range of $(R_{\rm G}^2)_{\rm LS}q^2$. Note that with the expression for $H_{\rm iso}(\vartheta, c)$ for spheres given below, plots of $KMc/\mathbf{R}_{\rm iso}(\vartheta, c)$ versus q^2 would not be expected to be parallel, unlike the experience with flexible chain polymers. For convenience, the ratio of the initial tangent for small ϑ to the intercept $Kc/\mathbf{R}_{\rm iso}(0, c)$ at concentration c is denoted $b(c)^2 = [c/\mathbf{R}_{\rm iso}(0, c)]^{-1}\partial[c/\mathbf{R}_{\rm iso}(\vartheta, c)]/\partial q^2$, so that, Equation (63)

$$b(c)^2 = \frac{(1/3)(R_{\rm G}^2)_{\rm LS} + c\Gamma_{\rm iso}(c)\partial W_{\rm iso}(\vartheta, c)/\partial q^2}{1 + c\Gamma_{\rm iso}(c)} \quad (63)$$

Consequently, even if $\partial W_{\rm iso}(\vartheta, c)/\partial q^2$ tends to zero (leading to parallel plots of $KMc/\mathbf{R}_{\rm iso}(\vartheta, c)$ versus q^2 over a range of c), $b(c)^2$ will not equal $(R_{\rm G}^2)_{\rm LS}$ unless $c\Gamma_{\rm iso}(c) \approx 2(A_2)_{\rm LS}M_{\rm LS}c \approx 0$, as at very high dilution, or for a dilute solution with $(A_2)_{\rm LS} = 0$.

For $\vartheta > 0$, the function $H_{\rm iso}(\vartheta, c)$ may introduce an additional dependence on ϑ, and for polymer chains, $P_{\rm iso}(\vartheta, c)$ may vary with c, reflecting principally changes in the mean-square radius of gyration with c.[4,33] The behavior for spherical particles interacting through a hard-core potential provides an example of this behavior. In that case, (Guinier and Fournet,[24] p. 48),[4] Equation (64) holds

$$H_{\rm iso}(\vartheta, c) = \frac{[P_{\rm iso}(\vartheta', c)]^{1/2}}{P_{\rm iso}(\vartheta, c)} \quad (64)$$

where $\sin(\vartheta'/2) = 2\sin(\vartheta/2)$, and $P_{\rm iso}(\vartheta, c) = P_{\rm iso}(\vartheta, 0)$ for this model. Consequently, $P_{\rm iso}(\vartheta, c)H(\vartheta, c) = [P_{\rm iso}(\vartheta', 0)]^{1/2}$, and $\mathbf{R}_{\rm iso}(\vartheta, c)/KMc \approx P_{\rm iso}(\vartheta, 0)$ for large $(R_{\rm G}^2)_{\rm LS}q^2$ as expected.

For flexible chain polymers, the mean-square radius of gyration $(R_{\rm G}^2)_c$ at concentration c tends to its value $\hat{a}L/3$ obtaining under Flory θ conditions with increasing c, even in good solvents for which $(R_{\rm G}^2)_c$ exceeds this value at infinite dilution.[4,33] A first approximation to $P_{\rm iso}(\vartheta, c)$ and $H_{\rm iso}(\vartheta, c)$ for flexible chain polymers in this situation is to use the relation for $P_{\rm iso}(\vartheta, c)$ for a random-flight chain with the observed $(R_{\rm G}^2)_c$ at concentration c (as opposed to the value at infinite dilution), along with the above approximation for $H_{\rm iso}(\vartheta, c)$;[4,113,114] Although this provides a qualitatively useful approximation, it may not be quantitatively correct, and should not be relied on if accuracy is critical; in one alternative, the Flory–Krigbaum pair potential (Flory,[29] p. 519) has been used to estimate $H_{\rm iso}(\vartheta, c)$.[115] The effects of electrostatic interactions among charged spheres are discussed in section 4.

3 EXPERIMENTAL METHODS

3.1 Instrumentation

Discussion of the instrumentation, alignment and calibration methods, and methods of treating data may be found in a number of places[1,10,18,116,117] (see also Chu,[7] p. 155) and only a brief account will be given here. Light scattering photometers with a variety of designs and special purposes are commercially available. These comprise several components, including:

1. the incident light source and its associated optics;
2. the light scattering cell;
3. the detector and its associated optics;
4. the electronic components associated with the detector.

In addition, the experiment could require specialized apparatus in special cases, e.g. filters to remove (most of) any fluorescence emission, means of controlling the temperature, etc.

Several factors are necessary in converting the response $G(\vartheta)$ from the detector to useful parameters in the analysis of static light scattering, including:[1,7,10,18,116,118]

1. the dependence of the scattering volume on angle and solvent;
2. the effects of attenuation of the incident or scattered rays;
3. the polarization state of the incident and scattered rays;
4. the relation of the detected signal to the Rayleigh ratio.

Most light scattering photometers make use of laser light sources, ranging from a few milliwatts to a hundred milliwatts or so in output power. Usually this is a plane-polarized state with the electric vector in the direction vertical to the scattering plane (the plane containing the incident and scattered rays), see below. The plane of polarization may then be rotated by rotation of a half-wave plate placed in the incident beam, providing a convenient tool for optical alignment. The detector optics are designed to control the angular acceptance

of the scattered light and the horizontal and vertical dimensions of the scattering volume; the depth of the scattering volume is controlled by the width of the incident beam. One of a few optical designs is usually employed in the detector optics,[1,7,10,18,116,117] and one of two general arrangements to make use of these relations may be found in practice: (i) the use of a single detector mounted on a goniometer for use over a range of scattering angles, and (ii) the use of a separate detector for each angle. The angular resolution and scattering volume must be known at each scattering volume to permit assessment of the Rayleigh ratio. In some designs this is facilitated by insuring that the scattering volume has a vertical dimension smaller than that of the incident beam,[1,18,117] but other procedures may be adopted. It is usually advisable to place a band-pass filter with a high transmission for the scattered light before the detector to minimize the effects of any fluorescence from the sample or stray light from sources other than the incident beam. Similarly, a band-reject filter may be used to assess the presence of fluorescence. It is also advisable to place a polarizing filter of some kind before the detector, and it is essential to do this if the anisotropic scattering is to be studied. Unfortunately, not all commercially available instruments provide the option of placing filters in the scattered beam.

Specialized instruments for examining the scattering profile in the Fraunhofer diffraction limit (very small scattering angles) have been constructed and are commercially available, e.g. see Weiner[119] and Plantz[120] for a discussion of some of these. The instruments differ principally in the manner of data collection: measurement of the intensity at an angle ϑ (within a small span $\delta\vartheta$), or collection over an appreciable angular range $\Delta\vartheta = \vartheta_{max} - \vartheta_{min}$, over a range of values for ϑ_{max}, see section 4 for further discussion. In some cases, these small angle measurements are augmented by scattering measurements at a larger angle (or angles), sometimes using light from a laser with excitation at more than one wavelength to provide scattering over a range of q.

3.2 Methods

The optical clarification of the scattering sample is of paramount importance in obtaining reliable results in light scattering. The clarification should remove extraneous motes or ''dust'', undissolved polymer, gas bubbles, and any other scattering source not intrinsic to the sample of interest. At the same time, the clarification process should not cause unwanted removal of a portion of the sample. Often, in the final stages of clarification, the number of extraneous motes is very small, giving the appearance of a few widely separated bright spots in the sample when the scattering is viewed at a small scattering angle. In some instruments the scattering volume is made small in a strategy to reduce the number of extraneous particles in the scattering volume to zero or one for the most part, owing to the wide separation of any remaining extraneous motes. Since the scattering will then normally be rather different for these two states, it is possible to discriminate against the scattering from the extraneous matter. A similar strategy is sometimes employed in computer-aided data acquisition, by collecting data over a time interval that is long in comparison with the timescale for normal intensity fluctuations, but shorter than the timescale for fluctuations due to diffusion of the extraneous matter. The stronger than normal scattered intensity due to the motes may then be discriminated against in the data collection.

In some cases, the experimental data encompass a limited range of angles (e.g. two or three), leading to the use of ''dissymmetry methods'' to analyze scattering data in the RGD regime, following a procedure introduced early in the use of light scattering to characterize solute dimensions. Often, one of the scattering angles is taken to be 90°. The dissymmetry $Z(\vartheta_1, \vartheta_2) = G(\vartheta_1)V(\vartheta_2)/G(\vartheta_2)V(\vartheta_1)$, with instrument response $G(\vartheta)$ and scattering volume $V(\vartheta)$ for $\vartheta_1 < \vartheta_2$, may be used to estimate $(R_G^2)_{LS}/\lambda^2$ (with $\lambda = \lambda_0/n_{medium}$) provided a model is available, including the effects of molecular-weight heterogeneity. The dissymmetry method may not be applied beyond the RGD regime, because then $Z(\vartheta_1, \vartheta_2)$ is not generally a single-valued function of $(R_G^2)_{LS}/\lambda^2$.[23] The use of supplementary angles ($\vartheta_2 = \pi - \vartheta_1$) will usually make the scattering volumes equal at the two scattering angles, simplifying the analysis. Note that in general, $[\mathbf{R}_{Vv}(\vartheta_1, c)/c]^0/[\mathbf{R}_{Vv}(\vartheta_2, c)/c]^0 \neq \mathbf{R}_{Vv}(\vartheta_1, c)/\mathbf{R}_{Vv}(\vartheta_2, c)$, requiring extrapolation to infinite dilution to determine $Z(\vartheta_1, \vartheta_2)$. In the original dissymmetry method, ϑ_1 and ϑ_2 were equal to 45° and 135°, respectively, so that $V(\vartheta_1) = V(\vartheta_2)$ in a convenient cell geometry (e.g. cylindrical), and $Z(\vartheta_1, \vartheta_2) = G(\vartheta_1)/G(\vartheta_2)$ (see Kerker,[21] p. 432).[23] The value of $(R_G^2)_{LS}/\lambda^2$ may then be used to estimate for $P_{Vv}(\vartheta_3, 0)$ for the assumed model with the correction needed to deduce $[\mathbf{R}_{Vv}(0, c)/c]^0$ from $[\mathbf{R}_{Vv}(\vartheta_3, c)/c]^0$ determined at some particular angle, e.g. $\vartheta_3 = 90°$ in the original method. In some cases, only two scattering angles are used, one being fairly small, and the other being 90°, with due attention to the ratio $V(\vartheta_2)/V(\vartheta_1)$.[119–122] Tables to assist this treatment are available for a number of models,[125] or the relations needed may be readily calculated as required using a desktop computer. The use of $P_{Vv}(\vartheta, 0)$ for a linear chain to estimate $(R_G^2)_{LS}/\lambda^2$ by the dissymmetry ratio for a chain that is branched may introduce an error in the estimate for $(R_G^2)_{LS}$ if $(R_G^2)_{LS}q^2$ exceeds about 1. For example, the deviations shown in Figure 5 would

lead to increasingly serious overestimation of $(R_G^2)_{LS}$ with increasing $(R_G^2)_{LS}q^2$. An example of this effect may be seen for $Z(45, 135)$ versus $(R_G^2)_{LS}/\lambda^2$ for monodisperse linear chains and regular four-arm star branched molecules for the random-flight model.[123]

Finally, other methods of data treatment may be found in the literature. Examples include Guinier plots of $\ln(\mathbf{R}_{Vv}(\vartheta, c)/Kc)$ versus q^2, based on the approximation $P_{Vv}(\vartheta, 0) = \exp(-(R_G^2)_{LS}q^2/3)$ (Guinier and Fournet,[24] p. 21), Casassa plots of $Kc/\mathbf{R}_{Vv}(\vartheta, c)$ versus q for rod-like scatterers, based on Equations (42) and (43),[18,56] Porod plots of $q^2\mathbf{R}_{Vv}(\vartheta, c)/Kc$ versus q, based on $P_{Vv}(\vartheta, 0)$ for worm-like chains, see below (Yamakawa,[2] p. 122),[18,52,124] and plots of $(Kc/\mathbf{R}_{Vv}(\vartheta, c))^{1/2}$ versus q^2, introduced by the author.[10] These are useful for analysis of data on dilute solutions under special circumstances. Thus, the Guinier plot was devised for use with the scattering from particles polydisperse in size and shape, for which a more detailed treatment was not feasible; as seen in Figure 3, the actual scattering behavior is usually not represented by the exponential function over a range of q (unless $(R_G^2)_{LS}$ is very small); with the exponential function providing a useful approximation to larger $q^2R_G^2$ for spheres than for the other examples in Figure 3. The Casassa plot is only useful for very long rod-like solutes for which the asymptotic behavior for large q may dominate the scattering for all accessible q, in which case the tangent provides length per unit mass. Similarly, the Porod plot is designed to elucidate the behavior at large q, and may, for example reveal the persistence length of a semiflexible (worm-like) chain if large enough q can be attained; the Porod plot is usually of most interest with the scattering from X-ray or neutron radiation owing to the need to reach large q. The ''square-root'' plot is based on the observation that for a linear monodisperse random-flight linear chain model, for which R_G^2/M is independent of M, Equation (65)

$$P_{iso}(\vartheta, 0)^{-1/2} = 1 + \frac{R_G^2 q^2}{6} + \cdots \tag{65}$$

facilitates evaluation of the initial tangent[125] and hence R_G^2, because in this case the coefficient of q^4 tends to be smaller in this expansion than in the expansion for $P_{iso}(\vartheta, 0)$. Clearly, the use of $P_{Vv}(\vartheta, 0)^{-1/2}$ to facilitate data analysis should be applied only under appropriate conditions, e.g. it would be inappropriate for the polydisperse solute with a most-probable distribution of molecular weight, see below. Similarly, for monodisperse spheres, $P_{iso}(\vartheta, 0)$ is closely approximated by $\exp(-R_G^2q^2/3)$ for $R_G^2q^2 < 2$, with the deviations growing for larger $R_G^2q^2$, see Figures 3 and 6; this fit is less satisfactory for spheres heterogeneous in size, as may be deduced from Figure 4.

4 EXAMPLES

4.1 Static Scattering and Size Separation Chromatography

The application of light scattering and differential refractometry in on-line analysis of the eluent in size exclusion chromatography (SEC) which separates the eluent by size provides a powerful tool for analysis and characterization and has found wide application, e.g. see Lloyd et al.[126] and Barth and Flippen[127]. Examples include SEC and flow field-flow chromatography (FFFC). Since the solute concentration is usually very low in the eluent, it is normally assumed that $c\Gamma_{Vv}(c) \approx 0$ for all elution volumes, or a correction is made based on an estimate for A_2Mc. With optically isotropic homopolymers, the response h_e of the differential refractometer at elution volume V_e permits evaluation of the weight fraction $w_e = h_e/$ h_e for each V_e (excepting possible end-group effects low M). Then, the static (excess) scattering data $[\mathbf{R}_{Vv}(\vartheta, c)]_e$ at each elution volume may be used to compute the molecular weight as $M_e \approx [\mathbf{R}_{Vv}(\vartheta, c)]_e/Kcw_e[P_{Vv}(\vartheta, c)]_e$, given the concentration $c \propto$ h_e and $\partial n/\partial c$. The factor $[P_{Vv}(\vartheta, c)]_e$ may be evaluated using either an experimental estimate for $[P_{Vv}(\vartheta, c)]_e$ from on-line measurements of $[\mathbf{R}_{Vv}(\vartheta, c)]_e$ as a function of ϑ, or as $[P_{Vv}(\vartheta, c)]_e \approx 1$ if ϑ is small enough. In addition, on-line measurements over a range of scattering angles allow determination of $(R_G^2)_e$ as a function of V_e. In some instruments, $[\mathbf{R}_{Vv}(\vartheta, c)]_e$ is measured only at a single angle, e.g. 90°, with the resulting need to estimate $[P_{Vv}(\vartheta, c)]_e$ at that angle being a possible source of inaccuracy. In other instruments, it is possible to measure the scattering at two or three angles, permitting the use of dissymmetry methods to estimate $[P_{Vv}(\vartheta, c)]_e$ based on an assumed functional form, see for example the discussion of dissymmetry methods above. Comparison of $M_w \approx$ w_eM_e calculated from the data on the eluent with M_w determined directly on the unfractionated polymer provides a self-consistency check.

An example of the use of multiangle light scattering with SEC to study the dimensions of poly(di-*n*-hexylsilane), (PDHS), and poly(phenyl-*n*-hexylsilane), (PPHS), is given in Figure 7 (from Berry and Cotts[1] and Cotts[128]). Figure 7(a) gives M_e and $(R_G^2)_e$ resulting from analysis of the multiangle scattering data, and Figure 7(b) gives the response from the differential refractive index detector, normalized to give the same peak response. As expected from the mechanism of SEC, the values of $(R_G^2)_e$ are the same at a given elution volume for the two polymers, but the values of M_e differ. If the elution volume increments are small enough, then $(R_G^2)_e/M_e$ is expected to approximate the value for a monodisperse chain. Analysis of these data permitted assessment of the difference in R_G^2/M for these polymers, revealing the substantially

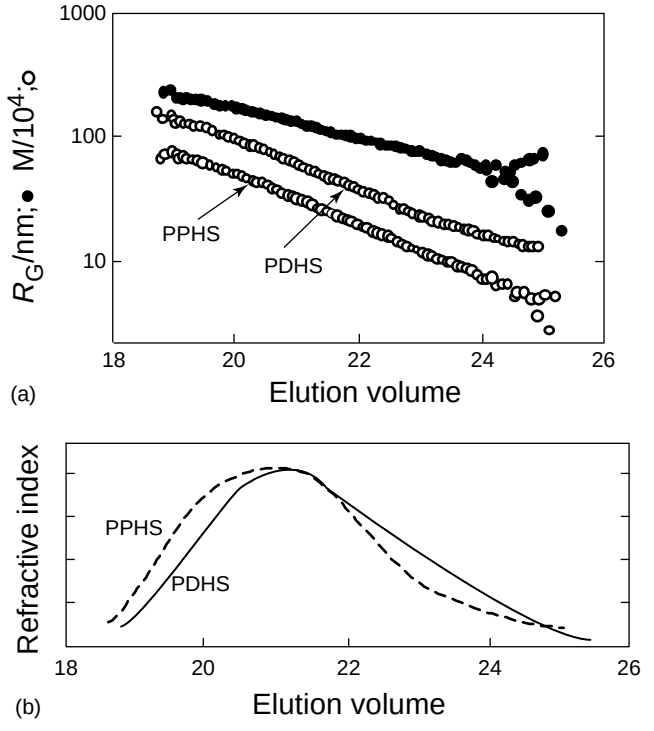

Figure 7 Example of the use of multiangle light scattering as an SEC detector in the analysis of PDHS, and PPHS. (a) M_w and $(R_G^2)_{LS}$ resulting from analysis of the multiangle scattering data. (b) The response from the differential refractive index, normalized to give the same peak response. (From Berry and Cotts.[1])

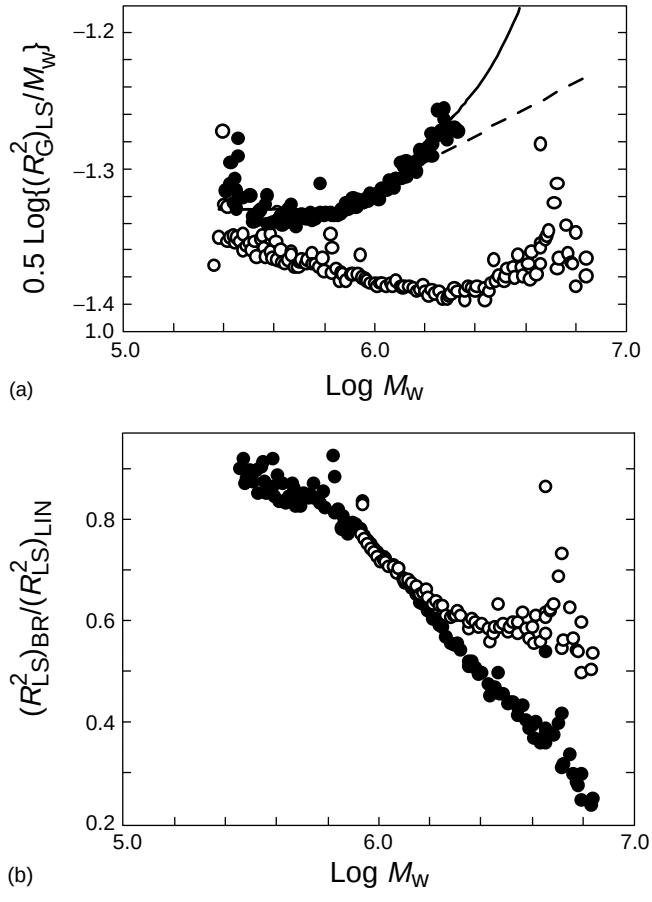

Figure 8 Example of the use of multiangle light scattering as an SEC detector in the analysis of randomly branched poly(methyl methacrylate) (PMMA). (a) M_w and $(R_G^2)_{LS}/M_w$ for the eluent for heterodisperse linear (●) and branched polymers (○) resulting from analysis of the multiangle scattering data, with the dashed line and solid curves giving power law and polynomial extrapolations, as discussed in the text. (b) The ratio $[(R_G^2)_{LS}]_{BR}/[(R_G^2)_{LS}]_{LIN}$ from the data in part (a), using the powerlaw (○) and polynomial (●) extrapolations to estimate $[(R_G^2)_{LS}]_{LIN}$. (Adapted from figures in Jackson et al.[131])

larger persistence length for PPHS, owing to the effects of the phenyl substituent.

Another example is provided by the analysis and characterization of branched macromolecules. Certain polymerizations are expected to produce a randomly branched chain structure, in which the number of branch nodes and their placements are randomly distributed among the chains in the ensemble (Flory,[29] p. 347) In such a case, it is not possible to achieve a simultaneous separation by the number m of branch nodes and overall molecular weight M, as each molecule in the ensemble is distinct in this respect. One could conceive of a separation by either p or M, with each separated component then narrow in one feature and polydisperse in the other. Analysis by SEC of such a polymer suggests that although the polymer at each elution volume V_e is polydisperse in both p and M, the distribution in M is much narrower than that in p.[129] Consequently, on-line determination of M_w and $(R_G^2)_{LS}$ of the eluent as functions of V_e offers the possibility that an average g_e may be estimated for an ensemble that approximates monodispersity in M, with a random distribution in p, with $g_e = (R_G^2)_{LS}/[(R_G^2)_{LS}]_{LIN}$, where $[(R_G^2)_{LS}]_{LIN}$ is for a linear polymer with same molecular weight as the branched chain. In this approximation, theoretical expressions[130] for g for a polymer monodisperse in M, with a random distribution in p, would be used to estimate p; the formula, involving summations, is readily calculated with a desk-top computer. An example of data of this type is given in Figure 8 for samples of linear and branched PMMAs.[131] According to the preceding, these data may be used to estimate g_e and hence the branching frequency, as a function of molecular weight, provided a reliable estimate for $[(R_G^2)_{LS}]_{LIN}$ is available. As demonstrated in Figure 8, the latter may not be readily available at the higher molecular weights encountered with branched chains. For the data in Figure 8, the dependence of $[(R_G^2)_{LS}/M_w]_{LIN}$ on M_w is stronger than anticipated for the larger V_e,

perhaps reflecting the effects of increased molecular weight dispersity with increasing V_e with the column set utilized. Two extrapolations have been presented and used to estimate g_e, with the true value likely to be somewhere between these estimates.

The size distributions in vesicle preparations similar to those described in the next section have been examined using FFFC.[132] Although a procedure involving simultaneous determination of the scattering and the concentration of the eluent could have been followed, only an (excess) instrument response $G(\vartheta, c)$ proportional to $\mathbf{R}_{Vv}(\vartheta, c)$ was determined as a function of elution volume. Owing to the low concentrations used, such that $2A_2Mc \ll 1$, it was assumed that $c\Gamma_{Vv}(c) \approx 0$ in calculations of the number fraction distribution n_μ. Then, since $[\mathbf{R}_{Vv}(\vartheta, c)]_\mu = KM_\mu^2 P_\mu(\vartheta, 0)\nu_\mu N_A$ for each fraction in this approximation, with ν_μ the number concentration, the number fraction was computed from Equation (66)

$$n_\mu = \frac{G_\mu(\vartheta, c)/M_\mu^2 P_\mu(\vartheta, 0)}{\sum G_\mu(\vartheta, c)/M_\mu^2 P_\mu(\vartheta, 0)} \tag{66}$$

In the analysis, $P_\mu(\vartheta, 0)$ was fitted to the expression for a hollow shell (the RGD approximation being adequate), to give the outer radius R_μ for each eluent fraction, on the assumption that the shell thickness Δ_{sh} was invariant with R_μ, and could be evaluated for the unfractionated polymer, as described in the next section. The expression given above was then used to compute M_μ, using R_μ, Δ_{sh}, and the solute specific volume v_2, to permit evaluation of n_μ for each R_μ. As anticipated from the mechanism for FFFC, R_μ increased linearly with elution volume, except for early elution times, for which the weak scattering may have introduced error in the evaluation of R_μ.

4.2 Light Scattering from Vesicles and Stratified Spheres

Many unilamellar vesicles comprise a single closed shell bilayer.[133] They may be prepared with a spherical shape, and that shape may be further stabilized by a slight imbalance in the compositions of the vesicle interior and exterior, to create a (small) positive osmotic pressure on the interior. Such vesicles provide an example of a shell-like structure, with a shell thickness Δ_{sh} generally much smaller than the outer radius R of the vesicle. The thickness Δ_{planar} of an analogous planar bilayer may be computed as $\Delta_{planar} = 2V_{lipid}/A_{lipid}$, where V_{lipid} and A_{lipid} are the volume and cross-sectional area of the molecules in the bilayer, respectively. It is expected that $\Delta_{sh} \approx \Delta_{planar}$ with increasing R. For a shell, Δ_{sh} may be calculated from the ratio of its volume v_2M/N_A to its surface area $4\pi R^2$, where v_2 is the specific volume of the shell, Equation (67):

$$\frac{v_2M}{N_A} = 4\pi R^2 \Delta_{sh}\left\{1 - \left(\frac{\Delta_{sh}}{R} \quad + \frac{(\Delta_{sh}/R)^2}{3}\right\}\right. \tag{67}$$

Consequently, the ratio of v_2M to R_G^2 may be used to approximate Δ_{sh} for spherical, monodisperse vesicles. The relation $P_{Vv}(\vartheta, 0)$ given above for a shell may be used to compute R_G^2 from Equations (68) and (69)

$$R_G^2 = R^2\left(\frac{3}{5}\frac{1-(1-(\Delta_{sh}/R)^5)}{1-(1-(\Delta_{sh}/R)^3)}\right. \tag{68}$$

$$R = \frac{(R_G^2)^{1/2}}{\{1 + (1/2)\beta + (1/20)\beta^2 + \cdots\}} \tag{69}$$

where $\beta = v_2M/4\pi N_A(R_G^2)^{1/2}$. Equation (70) follows

$$\Delta_{sh} = \frac{v_2M}{4\pi N_A R_G^2}\{1 + \left(\frac{1}{3} \quad \beta^2 + 0.06\beta^3 + \cdots\}^{-1}\right. \tag{70}$$

and inspection shows that $\Delta_{sh} \approx v_2M/4\pi N_A R_G^2$ and $R \approx (R_G^2)^{1/2}$ to a good approximation for $\beta < 1$, with correction factors readily calculated using experimental data for β. Light scattering data on unilamellar vesicles of phosphatidylcholine prepared over a range of size were analyzed to obtain R and Δ_{sh}, after extrapolation to infinite dilution, with results which gave Δ_{sh}, a weakly decreasing function of increasing R, tending to Δ_{planar} ($\Delta_{sh} \approx 3.5$ nm and $R \approx 34$–40 nm for the several samples examined).[134,135]

Size distributions for vesicles with characteristics similar to those discussed above have been reported from inversion of $P_{Vv}(\vartheta, c)$, using a nonnegative least squares fit over a predetermined selection of radii, using $P_{Vv}(\vartheta, c)$ for a vesicle, and the assumption that $c\Gamma_{Vv}(c) \approx 0$ in the analysis.[136] The authors report no significant difference between the use of the RGD approximation and the Mie theory for hollow vesicles, owing to the small radius (≈50 nm), and although the distribution functions obtained were reproducible, they did differ significantly from those obtained by inversion of dynamic light scattering data. The latter gave a broader distribution, especially for larger particle radii. Among possible reasons cited for this discrepancy, it was noted that the experimental time required for dynamic data acquisition was much longer than that for static data (hours compared with minutes), and that interparticle aggregation may have intervened during that longer time to introduce errors in the dynamic data. Another example of an analysis on hollow polystyrene spheres dispersed in water[107] produced the results shown in Figure 9, again under the assumption that $c\Gamma_{Vv}(c) \approx 0$. The spheres had a radius of about 0.23 μm, large enough to require the use of Mie theory in the analysis of the data at the larger q; the difference between the use of Mie theory and the RGD

approximation is shown in Figure 9. The size distribution obtained from dynamic light scattering was shifted to smaller sizes, see Figure 9(b), for reasons not definitively understood.

Block copolymers may form a core-shell micelle in solution, with a resemblance to a vesicle or hollow shell, but with scattering from the core as well as the shell.[137] In these cases, for a system with micelles nearly monodisperse in size and structure, $P_{Vv}(\vartheta, c)$ would be given by Equation (48) for a shell in the RGD regime or an appropriate relation for the Mie regime (Kerker,[21] p. 189), with scattering similar to that for a vesicle, but with a core with a refractive index different from that of the solvent. For example, in this case, since $\Delta^2_{AB} = 0$, for the centrosymmetric micelle (Equation 71)

$$(R^2_G)_{LS} = \left(\frac{3}{5}\right)\left\{(xR^2_A) + (1-x)\frac{[R^5_B - R^5_A]}{[R^3_B - R^3_A]}\right\} \quad (71)$$

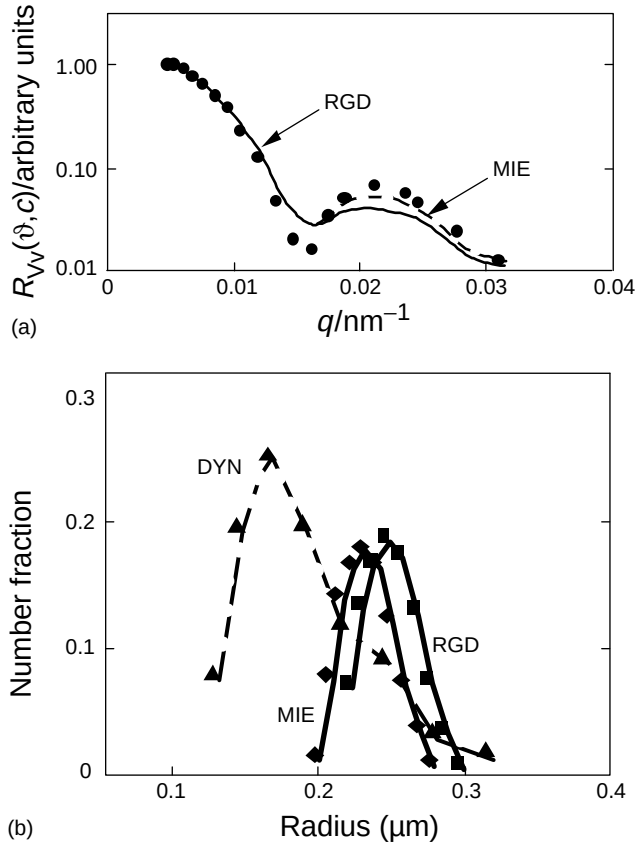

Figure 9 Example of the scattering from an aqueous dispersion of heterodisperse hollow polystyrene spheres. (a) Scattering function and fits thereto using the RGD approximation and the Mie theory. (b) Number fraction distribution of particles with a given size deduced from the inversion of the scattering function using the RGD approximation and the Mie theory; an estimate determined from analysis of the dynamic light scattering is included for comparison. (From figures in Strawbridge and Hallett.[107])

with the relation discussed above for a sphere of radius R_A coated by a shell of (outer) radius R_B and thickness $\Delta = R_B - R_A$, and $x = w_A\psi_A/[w_A\psi_A + (1 - w_A)\psi_B]$ Since the refractive increments ψ_A and ψ_B may be either negative, positive or zero, as may $(R^2_G)_{LS}$, which consequently does not bear any simple geometric significance. Owing to symmetry, $(R^2_G)_{LS}$ for a sphere with more complex but spherically symmetric distribution $\psi(r)$ of refractive index increment and density $\rho(r)$ about the center of mass may be expressed by converting the summations used above in the RGD approximation to integrals, Equation (72):[21]

$$(R^2_G)_{LS} = \frac{\int_0^R r^4\psi(r)\rho(r)\,dr}{\int_0^R r^2\psi(r)\rho(r)\,dr} \quad (72)$$

For example, the preceding expression results from separating the integrals into two parts, from 0 to R_A, and R_A to R_B.

4.3 Scattering from Very Large Particles

As noted above, for very large spheres, with both $\tilde{n} \gg 1$ and phase shift magnitude $\alpha|\tilde{n} - 1| > 10$, $P_{Vv}(\vartheta, c)$ reduces to the form for Fraunhofer diffraction from a circular aperture (or opaque disk), with the result that an exact integral expression may be written for the inversion of $P_{Vv}(\vartheta, c)$ to obtain the size distribution in terms of the scattered intensities. Thus, in this treatment, use is sometimes made of the expression,[119,132] Equation (73)

$$L(x_1, x_2) = \int_{x_1}^{x_2} s\left\{\frac{2J_1(s)}{s}\right\}^2 ds$$
$$= J^2_0(x_2) + J^2_1(x_2) - [J^2_0(x_1) + J^2_1(x_1)] \quad (73)$$

In this respect, an apparatus is used with a series of concentric-ring photodiodes, with each photodiode spanning successive angular increments. Thus, the forward scattering intensity is determined over a set of annular rings in this apparatus, rather than as a function of ϑ, with the function $L(\vartheta_1, \vartheta_2)$ representing the fraction of the intensity incident on a particle scattered between angles ϑ_1 and ϑ_2. The function $L(0, \vartheta)$ exhibits a series of plateaus for values of ϑ for which $J_1(\alpha\vartheta)/\alpha\vartheta = 0$, the first three of which are given by $\alpha\vartheta$ equal to 3.83, 7.02 and 10.17, e.g. see Bohren and Huffman,[13] p. 111; van de Hulst,[22] p. 110; Weiner,[119] Swithenbank et al.[138] In practice, the available angular range is adjusted for the set of ring photodiodes with fixed geometry by using a series of lenses of differing focal length f to manipulate the angular range, with $\vartheta \approx r/f$, where r is the radial position of the light on the photodiode. The maximum angle used should be large enough to encompass essentially

all of the scattering, for example to avoid skewing the measurement to the larger components in a polydisperse mixture, a problem called *vignetting* in some of the specialized literature in this area;[138] however, see below for an example of a related problem if the minimum angle is too large. Equation (74) applies to the analysis of a polydisperse sample (with $\sin(\vartheta) \approx \vartheta$),[119,138]

$$L(\vartheta_1, \vartheta_2) \propto \sum_{\mu} n_\mu \alpha_\mu^2 \{J_0^2(\alpha_\mu \vartheta_2) + J_1^2(\alpha_\mu \vartheta_2) - [(J_0^2(\alpha_\mu \vartheta_1) + J_1^2(\alpha_\mu \vartheta_1))]\} \tag{74}$$

where n_μ is the number fraction of spheres with radius R_μ. In practice, the sum over μ is approximated by a small number of discrete ''components'', each of these actually representing an average over components in a ''size class'' determined by the mean radius of the detector ring and the focal length of the lens in use.[119]

As mentioned above, a systematic consistent discussion of measurements in the Fraunhofer diffraction regime may be found in Koo and Hirleman[92] for the representative optics and inversion methods. Instruments collecting the scattered light over a small angular increment at a given ϑ are considered, along with instruments collecting light from some small angle $\vartheta_{\min}$ through angles up to ϑ; both types are used in practice. The forms of the integral expressions used in the inversion depend not only on the optical arrangement, but also on the number of times an integration by parts is used to develop the integral equation used, with consequent change in the way data are processed in the inversion. In the ideal case, the accuracy of the inversion would not depend on these factors, but with real data, containing noise, and limited to a range of ϑ by instrument design, these factors usually will affect the outcome. In addition, the method used to present the scattering medium (either as a dispersion in a fluid, or an aerosol) may affect the results. In this analysis, the expression for $\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)$ in the Fraunhofer limit is recast as an integral, Equation (75):[92]

$$\vartheta^a \mathbf{R}_{\mathrm{Vv}}(\vartheta, c) \propto \int_0^\infty k_{ab}(\alpha, \vartheta) \alpha^b n(\alpha)\, \mathrm{d}\alpha \tag{75}$$

where $\alpha = 2\pi R/\lambda$, $n(\alpha)$ is the (number) fraction of spheres with radius between R and $R + \mathrm{d}R$ (for fixed λ), the factors ϑ^a and α^b account for possible variation of the aperture over which the scattering is collected and the conversion to distributions in terms of particle volume ($b = 2$) or area ($b = 1$), respectively, and $k_{\mathrm{ab}}(\alpha, \vartheta)$ is the appropriate scattering function for the particle. For spheres, (with $\sin(\vartheta) \approx \vartheta$), Equation (76) applies

$$k_{\mathrm{ab}}(\alpha, \vartheta) = \frac{J_1^2(\alpha\vartheta)}{\vartheta^{2-a}\alpha^{b-2}} \tag{76}$$

This expression may then be inverted to obtain $\alpha^b n(\alpha)$ for each α as an integral over ϑ (from 0 to ∞) of a kernel comprising functions of $\alpha\vartheta$ and derivatives of $\vartheta^a \mathbf{R}_{\mathrm{Vv}}(\vartheta, c)$. The result may be cast in several forms, depending on a, b, and whether use is made of integration by parts in obtaining the relation used in the inversion. The required range of ϑ is, of course, nonphysical, but as the kernel involved reduces to essentially zero with increasing ϑ, the effect of the nonphysical upper limit on ϑ is usually nil. The inversion of noise-free scattering data over a span in ϑ typical of commercially available photometers is considered by Koo and Hirleman[92] for five different arrangements and a range of particle sizes (mean diameters of 7–506 μm) over a certain distribution of sizes given by a log-normal function in the particle area. Several statistical measures of the fit of the inversion to the starting size distribution were evaluated in order to arrive at the following principal conclusions in this idealized comparison: (i) the various methods are essentially equivalent over a relatively narrow range of particle size, but differ markedly for either small or large particles, (ii) none of the methods provide a satisfactory representation of the distribution of the smaller particles, and (iii) the methods vary considerably in their performance for distribution of larger particles. These results suggest that whereas the response may be reproducible, and therefore of use as a quality control protocol, the absolute distribution should be accepted with caution.

An example of the comparison of the size distribution deduced for some aerosols using an apparatus of the type described above with $L(\vartheta_1, \vartheta_2)$ determined over a set of angles is shown in Figure 10, including results obtained by inversion of $\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)$, and by fitting $\mathbf{R}_{\mathrm{Vv}}(\vartheta, c)$ with an assumed size distribution.[139] An example of the effect of the choice of the focal length f of the lens used on the analysis is illustrated in Figure 11 for the scattering from polydisperse glass beads with a size distribution (20–170 μm) calibrated by microscopic measurements on a large sample of the beads (National Institute of Standards and Technology (NIST) Standard Reference Material 1004).[138,140] Data were taken in the Fraunhofer regime ($\tilde{n} \approx 1.146$ and phase shift magnitude $\alpha|\tilde{n} - 1| > 40$) with both 100 and 300-mm focal length lenses. Whereas the latter fitted the reported size distribution well, except for the largest size, the data with the smaller focal length lens did not afford a measurement at small enough ϑ to permit a satisfactory assessment of the population of the larger beads.

As mentioned above, the scattering from large particles immersed in a solvent such that $\tilde{n} \approx 1$ may not be in the Fraunhofer regime, but rather may exhibit anomalous diffraction. The assumption that Fraunhofer diffraction

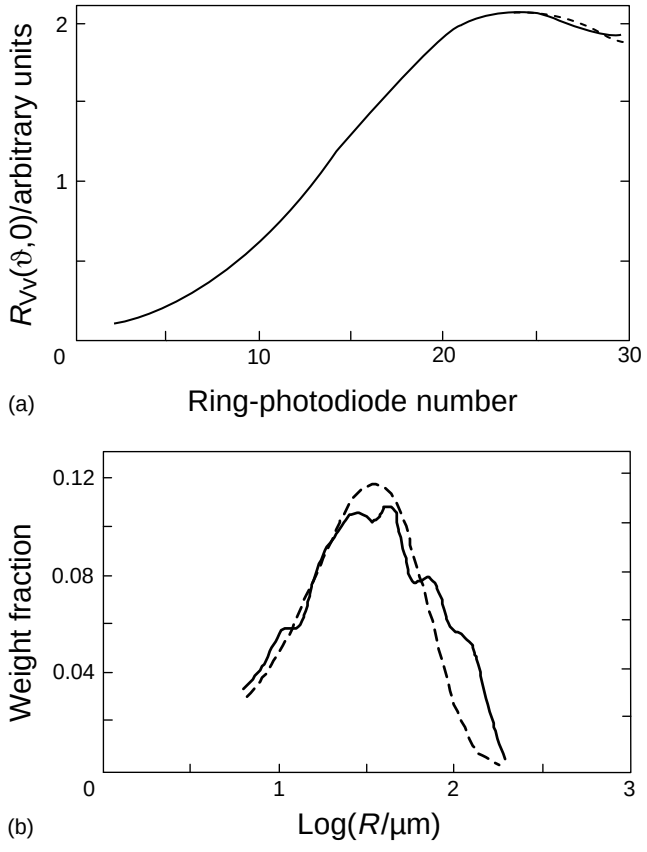

Figure 10 Example of the scattering from a heterodisperse aerosol. (a) Scattering function (solid curve) and fits thereto using the Fraunhofer diffraction approximation and an assumed size distribution function. (b) Weight fraction distribution of particles with a given size deduced from the inversion of the scattering function using the Fraunhofer diffraction approximation (solid curve) and the function used in part (a) (dashed curve). (Adapted from figures from Meyer and Chigier.[139])

applies may lead to serious error in the size distribution determined from the angular profile of the scattered light. An example of this is given in Figure 11, showing data obtained with the 100-mm focal length lens and dispersions in different liquids to provide two values of $\tilde{n}$ close to unity ($\tilde{n} \approx 1.02$ and 1.08).[138,140] The data on $L(\vartheta_1, \vartheta_2)$ show the profound effect of $\tilde{n}$. The expressions for the anomalous scattering regime were used to compute $L(\vartheta_1, \vartheta_2)$ using the size distribution, with the results being a reasonable fit to the data for both values of $\tilde{n}$. Further, the data for $\tilde{n} \approx 1.02$ were inverted under two assumptions: the use of the relations for the Fraunhofer and anomalous scattering regimes. Whereas the latter was in reasonable agreement with the result obtained with data in the Fraunhofer regime ($\tilde{n} \approx 1.146$), the use of the expressions for the Fraunhofer regime gave erroneous results, as expected.

4.4 Intermolecular Association

Intermolecular association is not uncommon in macromolecular solutions or dispersions of particles, especially in aqueous solvent. In general, two forms may be encountered in the extreme: association involving two or more components at equilibrium at any given concentration; and metastable association, in which the components present depend on processing history, but do not change sensibly with concentration in the range of interest for light scattering. Of course, intermediate situations may also occur.

Aside from any criterion imposed by the anticipated nature of the solute, the experimenter may have little reason to suspect association from the dependence of the scattering on concentration in the case of a metastable association. In some cases of metastable association, the molecular weight deduced from extrapolation of $Kc/\mathbf{R}_{Vv}(\vartheta, c)$ to zero q and infinite dilution will depend on temperature or solvent, revealing the association. An example of this sort in which the nature of intermolecular association of a solute with a helical conformation was elucidated by the use of static and dynamic light scattering as a function of temperature is given by Yue et al.[141] In the cited case, the scattering at any given temperature exhibited "normal" behavior and could not have been analyzed to reveal association if taken alone. In a different and somewhat unusual, but not unique, example, it has been reported that $Kc/\mathbf{R}_{Vv}(\vartheta, c)$ is linear in q^2, albeit giving a molecular weight that is much larger than the true value of M_w for the solute.[142] This was observed with a system that formed a gel at a higher solute concentration, suggesting that the observed scattering behavior reflects the anticipated $P_{Vv}(\vartheta, c)$ for a randomly branched polymer.[6] More frequently, with intermolecular association involving flexible chain polymers, $Kc/\mathbf{R}_{Vv}(\vartheta, c)$ exhibits enhanced scattering at small q. This is often taken as evidence for the presence of an aggregated species mixed with solute that is either fully dissociated, or much less aggregated. Although reasonable, it should be realized that such an interpretation is not unique.

A well-defined analysis is possible, at least in principle, for a case with equilibria among otherwise monodisperse monomers, dimers, etc. e.g. the equilibria obtaining among monomers, dimers, and tetramers for hemoglobin in solution.[143,144] Equilibrium association can lead to nonparallel plots of $Kc/\mathbf{R}_{Vv}(\vartheta, c)$ versus q^2 if the species are large enough. In the ideal situation, the ratio of R_G^2 for the aggregates to the unimer would be precisely known, as would be the effect of association on A_2, permitting an assessment of the equilibrium constant for the association for a given model.[144] An illustrative example of the

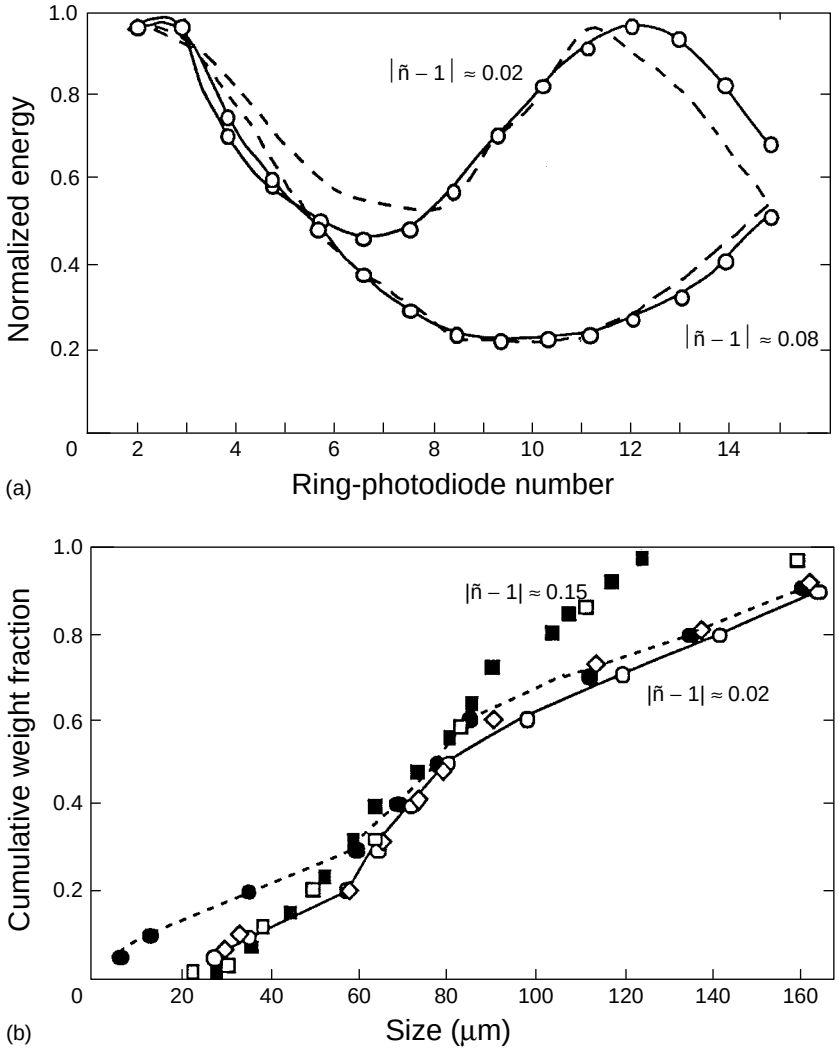

Figure 11 Example of the scattering from heterodisperse glass particles dispersed in different solvents to span a range of $\tilde{n}$. (a) Scattering function (points and solid curve) obtained using a lens with a 100-mm focal length and solvents to give the indicated $|\tilde{n} - 1|$, and fits thereto using the Mie theory and the known size distribution function (dashed curves). (b) Cumulative weight distribution of particles with a given size deduced by several methods: solid squares, from microscopic measurements (the ''known'' distribution function), as reported by NIST (see text), open squares and open diamonds, from inversion of inversion of the scattering function using the Mie theory for $|\tilde{n} - 1| \approx 0.15$ and two lenses, with focal lengths 300 mm (squares) and 100 mm (diamonds), and circles, for solvents to give $|\tilde{n} - 1| \approx 0.02$ calculated with the assumption of Fraunhofer diffraction (solid circles and dashed line) or anomalous diffraction (open circles and solid line). (Adapted from figures in Brown et al.[140])

effects of association is given in Figure 12. The example was calculated on the assumption that a monodisperse linear flexible coil chain of molecular weight M may undergo end-to-end dimerization to create a linear chain of molecular weight $2M$, with equilibrium constant K_{eq}. Further, it was assumed that $\partial \ln R_G^2 = \partial \ln M = \varepsilon$, and $\partial \ln A_2 = \partial \ln M = \gamma$, with $\varepsilon = 2(2 - \gamma)/3$, following the usual approximation for flexible chain polymers (see Yamakawa,[20] p. 146; Flory,[29], p. 536).[51] Finally, in addition to the expressions for $M_{LS} = M_w$, $(R_G^2)_{LS} = (R_G^2)_z$, and $P_{Vv}(\vartheta, c)$ given above, $(A_2)_{LS}$ was calculated as,[10,18] (Equation 77)

$$(A_2)_{LS} = \frac{\sum_{\mu}\sum_{\nu} w_\mu M_\mu w_\nu M_\nu (A_2)_{\mu\nu}}{M_w} \tag{77}$$

where $(A_2)_{\mu\mu}$ is A_2 for monodisperse component μ, and $(A_2)_{\mu\nu}$ is approximated by $\{(A_2)_{\mu\mu}(A_2)_{\nu\nu}\}^{1/2}$.[10,18] With the use of these relations, Equation (78–82)

$$M_w = M(2 - \zeta) \tag{78}$$

$$(R_G^2)_z = (R_G^2)_M \frac{[\zeta + 2^{1+\varepsilon}(1 - \zeta)]}{(2 - \zeta)} \tag{79}$$

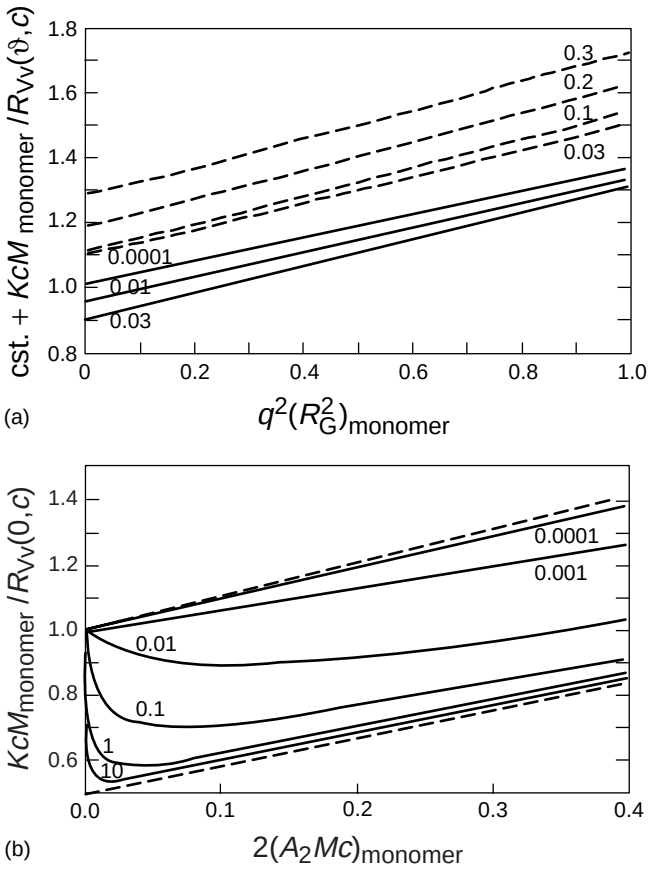

Figure 12 Scattering functions for an illustrative example of a flexible chain polymer undergoing end-to-end dimerization. (a) Dependence on angle, calculated as discussed in the text for a reduced equilibrium constant $K_{eq} = 0.1$ and the indicated values of $(A_2)_M Mc$, with the constant equal to zero or 0.2 for the solid and dashed curves, respectively. (b) Scattering extrapolated to zero angle as a function of $(A_2)_M Mc$, for the indicated values of K_{eq}.

$$P_{Vv}(\vartheta, c) = \left\{ \alpha[P_{Vv}(\vartheta, c)]_M + \frac{2(1-\zeta)[P_{Vv}(\vartheta, c)]_{2M}}{(2-\zeta)} \right\} \tag{80}$$

$$(A_2)_{LS} M_w = \frac{(A_2)_M M\{\zeta + 2^{1+\gamma/2}(1-\zeta)]^2\}}{(2-\zeta)} \tag{81}$$

$$\zeta = \left(\frac{\tilde{K}_{eq}}{2(A_2)_M Mc} \right) \left(\frac{1 + 4(A_2)_M Mc}{\tilde{K}_{eq}} \right)^{1/2} - 1 \tag{82}$$

where ζ is the degree of association, and $P_{Vv}(\vartheta, c)$ is given by the expression in Table 2, with u replaced by $2^{\varepsilon}u$ for $[P_{Vv}(\vartheta, c)]_{2M}$. The calculations were completed for $\varepsilon = 7/6$ ($\gamma = 1/4$), over a range of $(A_2)_M Mc$ and $\tilde{K}_{eq} = (A_2)_M M^2 K_{eq}$. It may be noted that for large $\tilde{K}_{eq}$, the extrapolation to obtain the true molecular weight M at infinite dilution ($\zeta = 0$) may not be possible and the experimenter may erroneously assume that the molecular weight of the polymer at infinite dilution is $2M$. The effect of the association in producing nonparallel $Kc/\mathbf{R}_{Vv}(\vartheta, c)$ versus q^2 for data at different concentration is illustrated in Figure 12.

The situation is usually more complex than the idealized equilibrium association, and analysis of such behavior is sometimes facilitated by an approximate representation with a few "pseudo" components (often two), each of which dominates the scattering over a limited range of q, with M, A_2 and $P_{Vv}(\vartheta, c)$ replaced by their light scattering averages for each pseudocomponent. That is (Equation 83),

$$\mathbf{R}_{Vv}(\vartheta, c) \approx \sum_{\mu} [\mathbf{R}_{Vv}(\vartheta, c)]_{\mu} = K \sum_{\mu} \left(\frac{Mc}{P_{Vv}(\vartheta, c)^{-1} + 2A_2 Mc} \right)_{\mu,c} \tag{83}$$

where the subscript "c" indicates that the parameters M and A_2, and the character of the function $P_{Vv}(\vartheta, c)$ may all depend on c through the dependence of the state of aggregation on c; note that this form does not properly account for the averaging among scattering elements, but it can provide a useful approximation if the components are few and widely separated in size. The analysis of the suspected association by this relation is similar to the determination of the size distribution discussed above in the absence of association, but is made more complex because it involves data over a range of c, with concentration-dependent parameters, and the function $P_{Vv}(\vartheta, c)$ may not be the same, or even known, for all of the aggregates present, and may depend on c. In some of the literature on small-angle X-ray scattering, it has been common to assume that $P_{Vv}(\vartheta, c) \approx \exp[-(R^2_G)_{LS} q^2/3]$ for each pseudocomponent, and frequently to assume that $2A_2 c \ll 1$ in an analysis used to estimate $(M_w)_{\mu} w_{\mu}$ and $[(R^2_G)_{LS}]_{\mu}$ for the assumed pseudocomponents as a function of the overall concentration c (Guinier and Fournet,[24] p. 151). Similar treatments are applied with light scattering, often restricted to small enough q that the expansion of $P_{Vv}(\vartheta, c)$ may be applied. The advent of dynamic light scattering has allowed some improvement in analyses of this type, since a similar expression (Equation 84) for the dynamic scattering can be applied in terms of pseudocomponents (see Equation 52):

$$g^{(1)}(\tau; \vartheta, c) \approx \sum_{\mu} r_{\mu}(\vartheta, c) \exp[-\tau \gamma_{\mu}(\vartheta, c)] \tag{84}$$

where $r_{\mu}(\vartheta, c) = [\mathbf{R}_{Vv}(\vartheta, c)]_{\mu} / \mathbf{R}_{Vv}(\vartheta, c)$. Analysis of $g^{(1)}(\tau; \vartheta, c)$ then provides information on $(M_w)_{\mu} w_{\mu}$ and a hydrodynamic length $[a_{LS,}]_{\mu} = kTq^2/6\pi\eta_s\gamma_{\mu}(\vartheta, c)$ for

each component, and some degree of consistency is expected among the estimates for $(M_w)_\mu w_\mu$ obtained in the two experiments. Further, comparison of $[a_{LS}]_\mu$ and $[(R_G^2)_{LS}]_\mu$ can provide insight into the nature of the component. In some cases, it may be reasonable to estimate $(M_w)_\mu$ for the component with smallest $(M_w)_\mu$ with the assumption that $w_\mu \approx 1$ for that component, i.e. the scattering at small q reflects a small fraction of a large component. An example of a treatment of this kind may be found in Einaga and Berry.[145] In some cases, the depolarized scattering can be particularly useful if the association induces order in the aggregated species.[146,147]

4.5 Scattering with Charged Species

In some cases, the polymeric and colloidal scattering species may bear electric charge. Whereas synthetic polyelectrolytes usually bear either anionic or cationic charge, biological macromolecules are often amphoteric, with a chain bearing both positive and negative charges. Electrostatic interactions among the scatterers may have two effects on macromolecules: (i) they may cause an expansion of the chain dimensions of macromolecules by intramolecular interactions, and (ii) they may alter the scattering through the effects of intermolecular interactions. Only the latter effect is relevant to charged particles. In the extreme, these interactions may lead to interactions with a very long coherence length, making it impossible to apply the relations discussed above.[16] Fortunately, for the purposes of size characterization, the effects of electrostatic interactions on $\mathbf{R}_{Vv}(\vartheta, c)$ and $\mathbf{R}_{Hv}(\vartheta, c)$ may be suppressed or screened by the presence of supporting simple electrolyte in the solution, thereby permitting application of the relations presented above, with one stipulation: the refractive index increment $(\partial n/\partial c)_\Pi$ must be determined for the solution in osmotic equilibrium with the solvent containing the supporting electrolyte.[16,18] Measurement of $(\partial n/\partial c)_\Pi$ can be achieved by equilibrating the solution with the mixed solvent through a membrane of the kind used in measurements of the osmotic pressure. The same procedure must be made with any mixed solvent with components of different refractive index, e.g. a mixed solvent comprising two organic liquids.

The Debye electrostatic length κ^{-1} provides an important measure of the spatial range of electrostatic interactions, with electrostatic interactions being suppressed between charged species that are much further apart than distance κ^{-1}. Here (in the electrostatic unit system most often employed by physical chemists, Equation 85)

$$\kappa = (8\pi N_A L_B I_0)^{1/2} \tag{85}$$

where $I_0 = \quad z_\mu^2 m_\mu/2$ the ionic strength derived from small molecule ions in the solution, m_μ being the molar concentration of species μ with charge z_μ, and $L_B = e^2/\varepsilon kT$ the Bjerrum length, with ε the dielectric constant (relative permittivity) of the solvent and e the charge on an electron. (It should be noted that variations of this expression with a factor $4\pi\varepsilon_0$ multiplying ε appear if SI units are employed, with ε_0 the permittivity of free space). Thus, at 25 °C, $L_B/\text{nm} \approx 57/\varepsilon$ and $\kappa^{-1}/\text{nm} \approx (4.57\pi N_A \quad z_\mu^2 m_\mu/\varepsilon)^{-1/2}$, or $L_B \approx 0.7$ nm and $\kappa^{-1}/\text{nm} \approx 0.42/\quad \tilde{m}_\mu/\text{mol L}^{-1}\ ^{1/2}$ for an aqueous solution of univalent charged simple electrolyte at 25 °C. With polyelectrolyte macromolecules, it is often convenient to adjust the supporting electrolyte composition to a level such that κ^{-1} is comparable to the geometric dimensions of a chain element. In this case, the thermodynamic diameter d_{thermo} appearing, for example in expressions for A_2, will not be dominated by the electrostatic interactions among chains, and the effects of intramolecular electrostatic interactions on the chain dimensions will be largely suppressed.[16,32,51,148] If this is not done, reliable extrapolation of $Kc/\mathbf{R}(0, c)$ to infinite dilution may be impossible if $\kappa^{-1} > (M/cN_A)^{1/3}$; an example of this effect for a rod-like chain is shown in Figure 13(a).[149] A similar effect will obtain with charged particles in solutions, e.g. as is well known, interparticle electrostatic repulsion among spheres can be strong enough to lead to an ordered mesophase with increasing c if the average separation of the spheres is less than κ^{-1}.[150] In addition, for macromolecules, the chain dimensions may expand with increasing κ^{-1} if $\kappa^{-1} > d_{geo}$, where d_{geo} is the geometric diameter of the chain repeat unit; an example of this is shown in Figure 13(b).

It should be realized that an aqueous solution of an organic solute is often close to intermolecular association, and that the addition of salt and consequent suppression of electrostatic interactions may induce association. Thus, with amphoteric proteins, it is often found that association will occur if the pH is adjusted to the isoelectric point, a condition for which appreciable numbers of anionic and cationic sites coexist on the chain.[16] For a pH far from the isoelectric point, the amphoteric macromolecule behaves as either an anionic or a cationic polyelectrolyte and the net charge can help stabilize the solution against association.

The effects of electrostatic interactions between charged spheres dispersed in a medium of low ionic strength can lead to a striking effect on $\mathbf{R}_{iso}(\vartheta, c)/KMc$, resulting from large values of $c\Gamma_{iso}(c)$ from electrostatic repulsion between the spheres. Thus, in this regime, Equation (86)

$$\frac{\mathbf{R}_{iso}(\vartheta, c)}{KMc} = \frac{P_{iso}(\vartheta, c)}{\{1 + c\Gamma_{iso}(c)P_{iso}(\vartheta, c)H_{iso}(\vartheta, c)\}} \approx \{c\Gamma_{iso}(c)H_{iso}(\vartheta, c)\}^{-1} \tag{86}$$

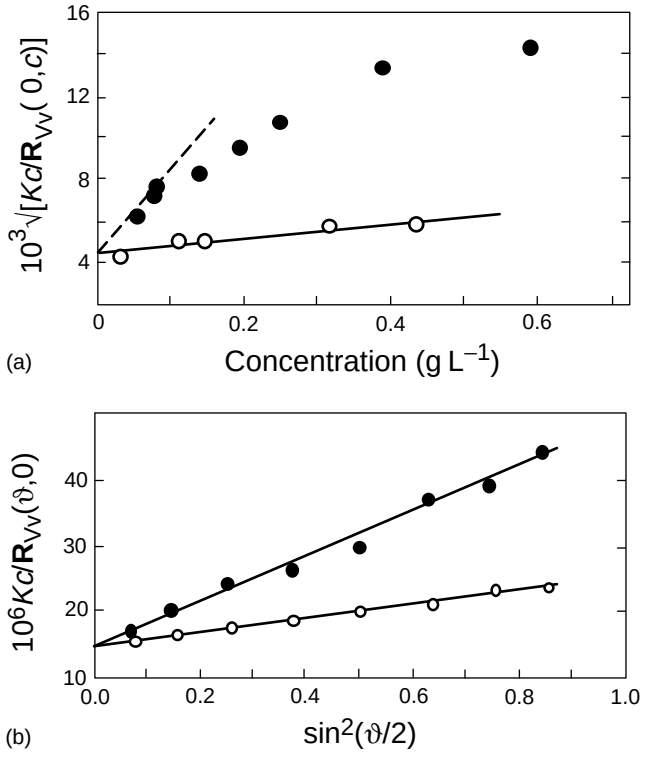

Figure 13 Scattering from a semiflexible polyelectrolyte chain in solvents with low and high ionic strengths (open and solid circles respectively). (a) $10^3\sqrt{Kc}/\mathbf{R}_{Vv}(0, c)$ versus concentration c; the dashed line is to emphasize the extreme nonlinear behavior for the case with a low ionic strength. (b) $10^6 Kc/\mathbf{R}_{Vv}(\vartheta, c)$ versus $\sin^2(\vartheta/2)$. (Adapted from figures in Sullivan and Berry.[149])

where $H_{iso}(\vartheta, c)^{-1}$ is expected to exhibit a maximum associated with the distance of average closest approach of the spheres, e.g. the expression for $H_{iso}(\vartheta, c)$ given above for uncharged spheres. Data on several dilute dispersions of charged polystyrene spheres ($R = 45$ nm) are given in Figure 14.[151] Similar results are reported for solutions of polyelectrolytes and for other charged particles.[148] As mentioned above, these curves bear a qualitative similarity to those obtaining for coated spheres under some conditions with the zero average refractive index increment, but their origin is very different, as is their shape in detail. A first-approximation to $H_{iso}(\vartheta, c)$ might be obtained by using $\sin(\vartheta'/2) \approx (2L_\kappa/R)\sin(\vartheta/2)$ in the expression for hard spheres if $L_\kappa \gg R$, where L_κ is an electrostatic length, expected to be related to κ^{-1}.[148,152] This approximation gives far too sharp a maximum, and values of L_κ to match the position of the maximum that are far larger than κ^{-1}.[151] In addition to the weakness of the ad hoc model, at least a part of this discrepancy may reflect heterogeneity or fluctuation of the charge density, which may broaden the peak.

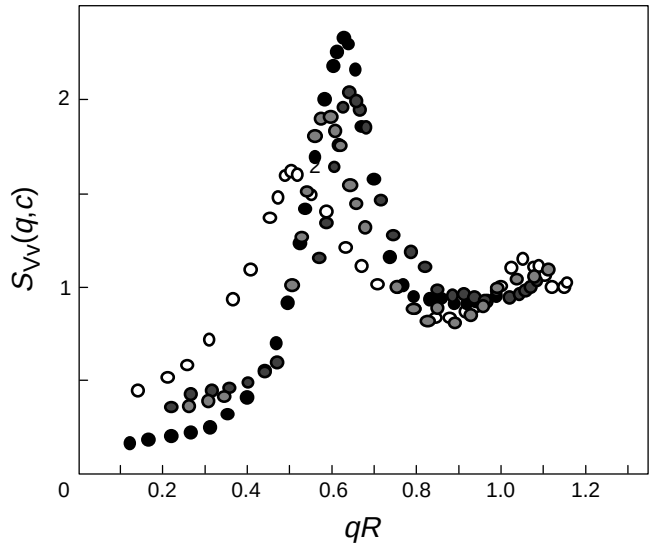

Figure 14 Dependence of the structure factor on qR for polystyrene spheres ($R = 45$ nm) immersed in deionized water, with the number concentration $\nu = 2.53$, 5.06, 7.59 and 10.12 particles μm^{-3} for the circles with increasing depth of the shading, respectively. (Adapted from figures in Grüner and Lehman.[151])

4.6 Scattering from Optically Anisotropic Solute

Although the (excess) scattering from polymer solutions is usually too small to be of much interest, exceptions can arise for chains with a semiflexible, or rod-like, conformation, and with spheres comprising optically anisotropic scattering elements. It is often assumed that the refractive index tensor of the scattering elements has cylindrical symmetry, with refractive index $n_\parallel$ and $n_\perp$ parallel and perpendicular to the cylinder axis, respectively.[10,49] For this model, the orientationally averaged refractive index is $\tilde{n}_{solute} = (n_\parallel + 2n_\perp)/3$, $\psi_{solute} \approx (\tilde{n}_{solute} - n_{medium})/\rho_{solute}$, and $\gamma_{solute} \approx (n_\parallel - n_\perp)/\rho_{solute}$, with ρ_{solute} the density of the solute (see Equation 9). If $\psi_{solute} \neq 0$, it is convenient to define the optical anisotropy of the scattering element as $\delta_0 = \gamma_{solute}/\psi_{solute}$.

A macromolecular model of interest is that of the persistent, or worm-like chain of contour length L and persistence length $\hat{a}$, with a cylindrical polarizability tensor for the scattering elements, each having an optical anisotropy δ_0. For this model, the mean-squared molecular anisotropy form factor is given by Equation (87),[10,40,41,48]

$$(\delta/\delta_0)^2 = \left(\frac{2}{3Z}\right)\{1 - (3Z)^{-1}[1 - \exp(-3Z)]\} \quad (87)$$

where $Z = L/\hat{a}$; the corresponding expression for R_G^2 is given in Table 2. In the limit where $\hat{a} \ll L$, as for a flexible chain polymer (e.g. vinyl polymers and most

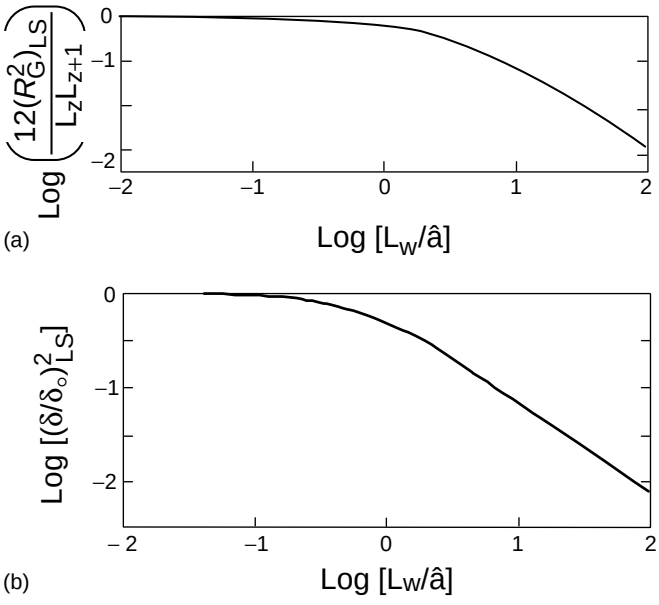

Figure 15 Dependence of $(R_G^2)_{LS}$ and $(\delta/\delta_0)^2_{LS}$ on the ratio of the weight-average contour length to the persistence length for rod-like molecules with a distribution of contour lengths in parts (a) and (b) respectively. (Adapted from figures in Berry.[10])

condensation polymers), $(\delta/\delta_0)^2$ tends to proportionality where $\hat{a}/L = \hat{a}M_L/M$, and M_L is the mass per unit chain contour length. As a consequence $\mathbf{R}_{Hv}(0, c)$ is proportional to the molecular weight of a chain element of length $\hat{a} \ll L$, is much smaller than $\mathbf{R}_{Vv}(0, c)$ for such chains, and may be neglected when considering $\mathbf{R}_{Vv}(0, c)$. In the opposite extreme where $L \ll \hat{a}$, as for a rod-like chain, $\delta \approx \delta_0$, and the anisotropic scattering must be taken into account to determine M_w from $\mathbf{R}_{Vv}(0, c)$. For chains with an overall rod-like character arising from helical structure, δ_0 may be so small that the anisotropic scattering may still be ignored when determining M_w from $\mathbf{R}_{Vv}(0, c)$. Nevertheless, $\mathbf{R}_{Hv}(0, c)$ may be of interest itself to evaluate δ^2 for use in an estimation of the persistence length. An example of a comparison of δ^2 as a function of L used to estimate δ_0^2 and $\hat{a}$ for a rod-like polymer is given in Figure 15.

As mentioned above, $P_{Hv}(\vartheta, \pi/4, 0)$ exhibits extrema as a function of qR for monodisperse spheres, for calculations in the RGD regime[72] and in the anomalous diffraction approximation,[70] the values of qR for these first maxima and following minima may be used to estimate R. In addition, the effects on the scattering pattern of the distortion of the anisotropic sphere, as might occur under a deformation, have been studied, see for example the discussion and literature citations in Samuels.[153]

LIST OF SYMBOLS

$\hat{a}$	Persistence length for semiflexible chains
A_2, A_3	Second virial coefficient third virial coefficient, etc.; see Equation (8)
$b(c)$	Correlation length obtained from the dependence of $\mathbf{R}_{Vv}(\vartheta)$ on ϑ; see Equation (63)
c, c_μ	Solute concentration (wt/vol); concentration of solute component μ
$F_{iso}(\vartheta, c)$	Intermolecular structure factor $\mathbf{R}_{iso}(\vartheta, c)/KcMP_{iso}(\vartheta, c)$; see Equation (4)
$g^{(1)}(\tau; \vartheta, c)$	Normalized electric field autocorrelation function, $g^{(1)}(\tau; \vartheta)$; see Equation (52)
$H_{iso}(\vartheta, c)$	Function $\dfrac{\{F_{iso}(\vartheta, c)^{-1} - 1\}}{c\Gamma_{iso}(c)P_{iso}(\vartheta, c)}$; see Equation (6)
k	Boltzmann's constant
K	Optical constant relating intensities to the Rayleigh ratio
L	Chain contour length
LS	Subscript to indicate the average of a function or parameter obtained in light scattering
M	Molecular weight
M_L	Mass per unit length, M/L
n_{medium}	Refractive index of the medium
n_{solute}	Refractive index of the solute
n_μ	Number fraction of solute component μ
$\tilde{n}$	Ratio n_{solute}/n_{medium}
N_A	Avogadro's constant
$P_{iso}(\vartheta, c)$	Intramolecular structure factor; see Equation (4)
$P_{Hv}(\vartheta, c)$	Intramolecular structure factor for the horizontally polarized component of the scattering with vertically polarized incident light
$P_{Vv}(\vartheta, c)$	Intramolecular structure factor for the vertically polarized component of the scattering with vertically polarized incident light
q	The modulus $q = (4\pi/\lambda)\sin(\vartheta/2)$ wave vector for an isotropic medium
$\mathbf{R}_{aniso}(\vartheta, c)$	Anisotropic component of $\mathbf{R}(\vartheta, c)$
$\mathbf{R}_{iso}(\vartheta, c)$	Isotropic component of $\mathbf{R}(\vartheta, c)$
$\mathbf{R}_{Hv}(\vartheta, c)$	Horizontally polarized component of $\mathbf{R}(\vartheta, c)$ for vertically polarized incident light
$\mathbf{R}_{Vv}(\vartheta, c)$	Vertically polarized component of $\mathbf{R}(\vartheta, c)$ for vertically polarized incident light

$\mathbf{R}(\vartheta, c)$	Excess Rayleigh ratio
R_G^2	Mean-square radius of gyration
R_H	Hydrodynamic radius, defined as $kT/6\pi\eta_s D_T$, with D_T the translational diffusion constant and η_s the solvent viscosity
$S_{iso}(\vartheta, c)$	Total structure factor $\mathbf{R}_{iso}(\vartheta, c)/KcM$; see Equation (7)
w, w_μ	Solute weight fraction; weight fraction of solute component μ
I_{INC}	Intensity of incident light
$I(\vartheta)$	Intensity at scattering angle ϑ
α	Parameter $2\pi R/\lambda$ for spheres of radius R
$\Gamma_{iso}(c)$	Function $\{F_{iso}(0, c)^{-1} - 1\}/c$; see Equation (6)
δ^2	Mean-square molecular optical anisotropy; see Equation (9)
δ_0	Optical anisotropy of a scattering element with molecular weight m_0
ϑ	Scattering angle
λ	Wavelength of light in the scattering medium; λ_0 the same in vacuo
ν	Number concentration of the solute, equal to cN_A/M_n
ρ	Density (wt/vol)
ψ_{solute}	Contrast factor for optically isotropic solute
ψ_μ	Contrast factor for component μ for optically isotropic media

ABBREVIATIONS AND ACRONYMS

FFFC	Flow Field-flow Chromatography
NIST	National Institute of Standards and Technology
PDHS	Poly(di-*n*-hexylsilane)
PMMA	Poly(methyl methacrylate)
PPHS	Poly(phenyl-*n*-hexylsilane)
RGD	Rayleigh–Gans–Debye
SEC	Size Exclusion Chromatography

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Diffraction in Particle Size Analysis • Optical Particle Counting • Photon Correlation Spectroscopy in Particle Sizing • Turbidimetry in Particle Size Analysis

REFERENCES

1. G.C. Berry, P.M. Cotts, 'Static and Dynamic Light Scattering', in *Experimental Methods in Polymer Characterization*, eds. R.A. Pethrick, R.S. Stein, J. Wiley & Sons, London, 1998.
2. H. Yamakawa, *Helical Wormlike Chains in Polymer Solutions*, Springer-Verlag, New York, 122, 289, 1997.
3. W. Brown (ed.), *Light Scattering: Principles and Development*, Monographs on the Physics and Chemistry of Materials, Oxford University Press, New York, Vol. 53, 1996.
4. G.C. Berry, 'Static and Dynamic Light Scattering on Moderately Concentrated Solutions: Isotropic Solutions of Flexible and Rod-like Chains and Nematic Solutions of Rod-like Chains', *Adv. Polym. Sci.*, **114**, 233–290 (1994).
5. R.S. Stein, M. Srinivasarao, 'Fifty Years of Light Scattering: a Perspective', *J. Polym. Sci.: Part B, Polym. Phys.*, **31**, 2003–2010 (1993).
6. W. Burchard, 'Static and Dynamic Light Scattering Approaches to Structure Determination in Biopolymers', in *Laser Light Scattering in Biochemistry*, eds. S.E. Harding, D.B. Sattelle, V.A. Bloomfield, Royal Society Chemistry, Cambridge, 3–22, 1992.
7. B. Chu, *Laser Light Scattering*, 2nd edition, Academic Press, Boston, MA, 1991.
8. P.W. Barber, S.C. Hill, *Light Scattering by Particles: Computational Methods*, World Scientific, Singapore, 1990.
9. G. Gouesbet, G. Grehan (eds.), *Optical Particle Sizing: Theory and Practice*, Plenum Press, New York, 1988.
10. G.C. Berry, 'Light Scattering', in *Encyclopedia of Polymer Science and Engineering*, eds. H. Mark et al, John Wiley & Sons, New York, 721–794, Vol. 8, 1987.
11. E. Gulari, A. Annapragada, E. Gulari, B. Jawad, 'Determination of Particle Size Distributions Using Light-scattering Techniques', *ACS Symp. Ser.*, **332**, 133–145 (1987).
12. W. Burchard, 'Static and Dynamic Light Scattering from Branched Polymers and Biopolymers', *Adv. Polym. Sci.*, **48**, 1–124 (1983).
13. C.F. Bohren, D.R. Huffman, *Absorption and Scattering of Light by Small Particles*, John Wiley & Sons, New York, 1983.
14. P. Kratochvil, 'Advances in Classical Light Scattering from Polymer Solutions', *Pure Appl. Chem.*, **54**, 379–393 (1982).
15. O. Glatter, O. Kratky (eds.), *Small Angle X-ray Scattering*, Academic Press, New York, 510, 1982.
16. H. Eisenberg, *Biological Macromolecules and Polyelectrolytes in Solution: Monographs on Physical Biochemistry*, eds. W.F. Harrington, A.R. Peacocke, Oxford University Press, London, 64, 1976.
17. B.J. Berne, R. Pecora, *Dynamic Light Scattering*, John Wiley & Sons, New York, 1976.

18. E.F. Casassa, G.C. Berry, 'Light Scattering from Solutions of Macromolecules', in *Polymer Molecular Weights, Part. 1*, ed. P.E. Slade, Jr, Marcel Dekker, New York, 161–286, Vol. 4, 1975.
19. M.B. Huglin (ed.), *Light Scattering from Polymer Solutions*, Academic Press, London, 1972.
20. H. Yamakawa, *Modern Theory of Polymer Solutions*, Harper and Row, New York, 1971.
21. M. Kerker, *The Scattering of Light, and Other Electromagnetic Radiation: Physical Chemistry, a Series of Monographs*, Academic Press, New York, Vol. 16, 1969.
22. H.C. van de Hulst, *Light Scattering by Small Particles*, John Wiley & Sons, New York, 1957.
23. K.A. Stacey, *Light Scattering in Physical Chemistry*, Academic Press, New York, 58, 1956.
24. A. Guinier, G. Fournet, *Small-angle Scattering of X-rays*, John Wiley & Sons, New York, 1955.
25. D. McIntyre, F. Gornick, *Light Scattering from Dilute Polymer Solutions*, International Science Review Series, ed. L. Klein, Gordon and Breach, New York, Vol. 3, 1964.
26. M. Kerker (ed.), *Selected Papers on Light Scattering*, SPIE Milestone Series, SPIE Int. Soc. Opt. Eng., Bellingham, WA, Vol. 951, 1988.
27. W. Brown (ed.), 'Dynamic Light Scattering', *Monographs on the Physics and Chemistry of Materials*, Clarendon Press, Oxford, No. 49, 1993.
28. Y. Einaga, F. Abe, H. Yamakawa, 'Light Scattering Method of Determining the Second Virial Coefficient for Simple Molecules and Oligomers', *J. Phys. Chem.*, **96**, 3948–3953 (1992).
29. P.J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
30. F.R. Hallet, 'Size Distributions from Static Light Scattering', in *Light Scattering: Principles and Development*, ed. W. Brown, Clarendon Press, Oxford, 477–493, 1996.
31. B.H. Zimm, 'Apparatus and Methods for Measurement and Interpretation of the Angular Variation of Light Scattering; Preliminary Results on Polystyrene Solutions', *J. Chem. Phys.*, **16**, 157–174 (1948).
32. M. Nagasawa, A. Takahashi, 'Light Scattering from Polyelectrolyte Solutions', in *Light Scattering from Polymer Solutions*, ed. M.B. Huglin, Academic Press, New York, 671–723, 1972.
33. P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY, 1979.
34. G.R. Mitchell, 'X-ray Scattering from Non-crystalline and Liquid Crystalline Polymers', in *Comprehensive Polymer Science*, ed. G. Allen, Pergamon Press, New York, Vol. 1, 1988.
35. D.M. Sadler, 'Neutron Scattering from Solid Polymers', in *Comprehensive Polymer Science*, ed. G. Allen, Pergamon Press, New York, Vol. 1, 1988.
36. H. Benoit, D. Froelich, 'Application of Light Scattering to Copolymers', in *Light Scattering from Polymer Solutions*, ed. M.B. Huglin, Academic Press, New York, 467–501, 1972.
37. H. Benoit, 'Light Scattering by Dilute Solutions of High Polymers', in *Electromagnetic Scattering*, ed. M. Kerker, Pergamon Press, Elmworth, NY, 285–301, 1963.
38. R.B. Penndorf, 'Approximation Formula for Forward Scattering', *J. Opt. Soc. Am.*, **52**, 797–800 (1962).
39. J.R. Urwin, 'Molecular Weight Distribution by Turbidimetric Titration', in *Light Scattering from Polymer Solutions*, ed. M.B. Huglin, Academic Press, New York, 789–824, 1972.
40. H. Benoit, 'Étude de la Dépolarization de la Lumière Diffusée par les Chaînes Moléculaires', *C.R. Acad. Sci.*, **236**, 687–689 (1953).
41. P. Horn, 'Light Scattering in Solutions of Anisotropic Macromolecules', *Ann. Phys.*, **10**, 386–434 (1955).
42. G.C. Berry, 'Molecular Weight Distribution', in *Encyclopedia of Materials Science and Engineering*, ed. M.B. Bever, Pergamon Press, Oxford, 3759–3768, 1986.
43. S. Boron, B. Waldie, 'Particle Sizing by Forward Lobe Scattered Intensity Ratio Technique: Errors Introduced by Applying Diffraction in the Mie Regime', *Appl. Opt.*, **17**, 1644–1648 (1978).
44. J.R. Hodkinson, 'Particle Sizing by Means of the Forward Scattering Lobe', *Appl. Opt.*, **5**, 839–844 (1966).
45. D.K. Hahn, S.R. Aragon, 'Mie Scattering from Anisotropic Thick Spherical Shells', *J. Chem. Phys.*, **101**, 8409–8417 (1994).
46. S. Asano, M. Sato, 'Light Scattering by Randomly Oriented Spheroidal Particles', *Appl. Opt.*, **19**, 962–974 (1980).
47. G.C. Berry, 'Properties of an Optically Anisotropic Heterocyclic Ladder Polymer (BBL) in Solution', *J. Polym. Sci., Polym. Symp.*, **65**, 143–172 (1978).
48. K. Nagai, 'Theory of Light Scattering by an Isotropic System Composed of Anisotropic Units with Application to the Porod–Kratky Chains', *Polym. J.*, **3**, 67–83 (1972).
49. H. Benoit, 'Determination of the Dimensions of Anisotropic Macromolecules by Means of Light Scattering', *Makromol. Chem.*, **18/19**, 397–405 (1956).
50. B.H. Zimm, 'The Scattering of Light and the Radial Distribution Function of High Polymer Solutions', *J. Chem. Phys.*, **16**, 1093–1099 (1948).
51. E.F. Casassa, G.C. Berry, 'Polymer Solutions', in *Comprehensive Polymer Science*, ed. G. Allen, Pergamon Press, New York, Vol. 2, Chapter 3, 1988.
52. A. Peterlin, 'Light Scattering by Non-Gaussian Macromolecular Coils', in *Electromagnetic Scattering*, ed. M. Kerker, Pergamon Press, Elmworth, NY, 357–375, 1963.
53. E.F. Casassa, G.C. Berry, 'Angular Distribution of Intensity of Rayleigh Scattering from Comblike Branched Molecules', *J. Polym. Sci., Part A-2*, **4**, 881 (1966).
54. H. Benoit, 'On the Effect of Branching and Polydispersity on the Angular Distribution of Light Scattered by Gaussian Coils', *J. Polym. Sci.*, **11**, 507–511 (1953).

55. D.R. Miller, E.M. Valles, C.W. Macosko, 'Calculation of Molecular Parameters for Stepwise Polyfunctional Polymerization', *Polym. Eng. Sci.*, **19**, 272–283 (1979).
56. E.F. Casassa, 'Light Scattering from very Long Rod-like Particles and an Application to Polymerized Fibrinogen', *J. Chem. Phys.*, **23**, 596–597 (1955).
57. F. Mallamace, N. Micali, 'Low Angle Light Scattering and its Applications', in *Light Scattering, Principles and Development*, ed. W. Brown, Clarendon Press, Oxford, 381–438, 1996.
58. C.M. Sorensen, 'Scattering and Absorption of Light by Particles and Aggregates', in *Handbook of Surface and Colloid Chemistry*, ed. K.S. Birdi, CRC Press, Boca Raton, FL, 533–558, 1997.
59. C.F.A. Bohren, 'Recurrence Relations for the Mie Scattering Coefficients', *J. Opt. Soc. Am. A*, **4**, 612–613 (1986).
60. W.J. Wiscombe, 'Improved Mie Scattering Algorithms', *Appl. Opt.*, **19**, 1505–1509 (1980).
61. R.H. Boll, J.A. Leacock, G.C. Clark, S.W. Churchill, *Tables of Light Scattering Functions, Relative Indices of Less than Unity, and Infinity*, University of Michigan Press, Ann Arbor, MI, 1958.
62. W.J. Pangonis, W. Heller, A. Jacobson, *Tables of Light Scattering Functions for Spherical Particles*, Wayne State University Press, Detroit, MI, 1957.
63. R.O. Gumprecht, C.M. Sliepcevich, *Tables of Light-scattering Functions for Spherical Particles*, University of Michigan Press, Ann Arbor, MI, 1951.
64. R. Bhandari, 'Scattering Coefficients for a Multilayered Sphere: Analytic Expressions and Algorithms', *Appl. Opt.*, **24**, 1960–1967 (1985).
65. S. Levine, 'Scattering of Electromagnetic Waves from Two Concentric Spheres, When Outer Shell has a Variable Refractive Index', in *Electromagnetic Scattering, ICES*, ed. M. Kerker, Pergamon Press, New York, 205–212, 1963.
66. P.J. Wyatt, 'Scattering of Electromagnetic Plane Waves from Inhomogeneous Spherically Symmetric Objects', *Phys. Rev.*, **127**, 1837–1843 (1962).
67. J. Buitenhus, J.K.G. Dhont, H.N.W. Lekkerkerker, 'Scattering of Light from Cylindrical Particles: Coupled Dipole Method Calculations and the Range of Validity of the Rayleigh–Gans–Debye Approximation', *J. Colloid Interface Sci.*, **162**, 19–24 (1994).
68. C.S. Buehler, J.M. Caruthers, E.I. Franses, 'Light Scattering Theory from Monodisperse Spheroidal Particles in the Rayleigh–Debye–Gans Regime', *J. Chem. Phys.*, **92**, 140–156 (1990).
69. S. Asano, 'Light Scattering Properties of Spheroidal Particles', *Appl. Opt.*, **18**, 712–723 (1979).
70. G.H. Meeten, P. Navard, 'Small-angle Scattering of Polarized Light. I. Comparison of Theoretical Predictions for Isotropic and Anisotropic Spheres', *J. Polym. Sci., Part B: Polym. Phys.*, **27**, 2023–2035 (1989).
71. G.H. Meeten, 'Small-angle Light Scattering by Spherulites in the Anomalous Diffraction Approximation', *Opt. Acta*, **29**, 759–770 (1982).
72. M.B. Rhodes, R.S. Stein, 'Scattering of Light from Assemblies of Oriented Rods', *J. Polym. Sci., Part A-2*, **7**, 1539–1558 (1969).
73. J. Holoubek, 'Small-angle Light Scattering from an Anisotropic Sphere, Anisotropy and Size Effects', *J. Polym. Sci., Part B: Polym. Phys.*, **32**, 351–357 (1994).
74. J. Holoubek, 'Small-angle Light Scattering from an Anisotropic Sphere, the Mueller Matrix Approach', *J. Polym. Sci., Part B: Polym. Phys.*, **29**, 1057–1068 (1991).
75. C. Arvieu, P. Navard, 'Small-angle Scattering of Polarized Light. III. Isotropic Sphere in an Anisotropic Medium (Large Refractive Index Mismatch)', *J. Polym. Sci., Part B: Polym. Phys.*, **27**, 2157–2158 (1989).
76. J.V. Champion, A. Killey, G.H. Meeten, 'Small-angle Polarized Light Scattering by Spherulites', *J. Polym. Sci., Polym. Phys. Ed.*, **23**, 1467–1476 (1985).
77. S. Asano, 'Light Scattering by Horizontally Oriented Spheroidal Particles', *Appl. Opt.*, **22**, 1390–1396 (1983).
78. S. Clough, J.J. van Aartesn, R.S. Stein, 'Low-angle Light-scattering Equations for Polymer Spherulites', *J. Polym. Sci., Part A-2*, **9**, 1147–1148 (1971) and citations therein.
79. H. Schnablegger, O. Glatter, 'Sizing of Colloidal Particles with Light Scattering, Corrections for Beginning Multiple Scattering', *Appl. Opt.*, **34**, 3489–3501 (1995).
80. H. Schnablegger, O. Glatter, 'Particle Sizing of Turbid Colloidal Systems, a Multiple Scattering Correction', *Progr. Colloid Polym. Sci.*, **93**, 352–354 (1993).
81. E.D. Hirleman, 'General Solution to the Inverse Near-forward-scattering Particle-sizing Problem in Multiple-scattering Environments: Theory', *Appl. Opt.*, **30**, 4832–4838 (1991).
82. E.D. Hirleman, 'Modeling of Multiple Scattering Effects in Fraunhofer Diffraction Particle Size Analysis', *Part. Part. Syst. Charact.*, **5**, 57–65 (1988).
83. E.D. Hirleman, 'Modeling of Multiple Scattering Effects in Fraunhofer Diffraction Particle Analysis', in *Optical Particle Sizing, Theory and Practice*, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 159–175, 1988.
84. A.K. Roy, S.K. Sharma, 'Inverse Scattering Problem Involving Soft Mie Particles', *Appl. Opt.*, **36**, 9487–9495 (1997).
85. R. Finsy, L. Deriemaeker, E. Geladé, J. Joosten, 'Inversion of Static Light Scattering Measurements for Particle Size Distributions', *J. Colloid Interface Sci.*, **153**, 337–354 (1992).
86. H. Schnablegger, O. Glatter, 'Optical Sizing of Small Colloidal Particles: an Optimized Regularization Technique', *Appl. Opt.*, **30**, 4889–4896 (1991).
87. O. Glatter, M. Hofer, 'Particle Sizing of Polydisperse Samples by Mie-scattering', in *Optical Particle Sizing:*

Theory and Practice, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 121–133, 1988.
88. N. Ostrowsky, D. Sornette, R. Parker, E.R. Pike, 'Exponential Sampling Method for Light Scattering Polydispersity Analysis', *Opt. Acta*, **28**, 1059–1070 (1981).
89. O. Glatter, 'Determination of Particle-size Distribution Functions from Small-angle Scattering Data by Means of the Indirect Transformation Method', *J. Appl. Cryst.*, **13**, 7–11 (1980).
90. P.J. Wyatt, 'Differential Light Scattering: A Physical Method for Identifying Living Bacterial Cells', *Appl. Opt.*, **7**, 1879–1896 (1968).
91. J.-J. Liu, 'Essential Parameters in Particle Sizing by Integral Transform Inversions', *Appl. Opt.*, **36**, 5535–5545 (1997).
92. J.H. Koo, E.D. Hirleman, 'Synthesis of Integral Transform Solutions for the Reconstruction of Particle-size Distributions from Forward-scattered Light', *Appl. Opt.*, **31**, 2130–2140 (1992).
93. E.D. Hirleman, P.A. Dellenback, 'Adaptive Fraunhofer Diffraction Particle Sizing Instrument Using a Spatial Light Modulator', *Appl. Opt.*, **28**, 4870–4878 (1989).
94. L.P. Bayvel, J.C. Knight, G.N. Robertson, 'Application of the Shifrin Inversion to the Malvern', in *Optical Particle Sizing: Theory and Practice*, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 311–319, 1988.
95. M. Bertero, P. Boccacci, C. De Mol, E.R. Pike, 'Particle-size Distributions from Fraunhofer Diffraction', in *Optical Particle Sizing, Theory and Practice*, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 99–105, 1988.
96. D.J. Brown, E.J. Weatherby, K. Alexander, 'Shape, Concentration and Anomalous Diffraction Effects in Sizing Solids in Liquids', in *Optical Particle Sizing, Theory and Practice*, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 351–362, 1988.
97. D. Kouzelis, S.M. Cande, E. Esposito, S. Zikikout, 'Particle Sizing by Laser Light Diffraction: Improvements in Optics and Algorithms', in *Optical Particle Sizing: Theory and Practice*, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 335–349, 1988.
98. E.D. Hirleman, 'Optimal Scaling of the Inverse Fraunhofer Diffraction Particle Sizing Problem: The Linear System Produced by Quadrature', in *Optical Particle Sizing: Theory and Practice*, eds. G. Gouesbet, G. Gréhan, Plenum Press, New York, 135–146, 1988.
99. E.D. Hirleman, 'Optimal Scaling of the Inverse Fraunhofer Diffraction Particle Sizing Problem: The Linear System Produced by Quadrature', *Part. Charact.*, **4**, 128–133 (1987).
100. J.H. Koo, E.D. Hirleman, 'Investigation of Integral Transform Techniques for Laser Diffraction Particle Size Analysis', *Chem. Phys. Processes Combust.*, 1–77 (1985).
101. E.D. Hirleman, V. Oechsle, N.A. Chigier, 'Response Characteristics of Laser Diffraction Particle Size Analyzers: Optical Sample Volume Extent and Lens Effects', *Opt. Eng.*, **23**, 610–619 (1984).
102. S.W. Provencher, 'A Constrained Regularization Method for Inverting Data Represented by Linear Algebraic or Integral Equations', *Computer Phys. Comm.*, **27**, 213–227 (1982).
103. S.W. Provencher, 'CONTIN: a General Purpose Constrained Regularization Program for Inverting Noisy Linear Algebraic and Integral Equations', *Computer Phys. Comm.*, **27**, 229–242 (1982).
104. J. Wang, F.R. Hallett, 'Spherical Particle Size Determination by Analytical Inversion of the UV-visible-NIR Extinction Spectrum', *Appl. Opt.*, **35**, 193–197 (1996).
105. M.R. Jones, B.P. Curry, M. Quinn Brewster, K.H. Leong, 'Inversion of Light-scattering Measurements for Particle Size and Optical Constants: Theoretical Study', *Appl. Opt.*, **33**, 4025–4034 (1994).
106. M.R. Jones, K.H. Leong, M. Quinn Brewster, B.P. Curry, 'Inversion of Light-scattering Measurements for Particle Size and Optical Constants: Experimental Study', *Appl. Opt.*, **33**, 4035–4041 (1994).
107. K.B. Strawbridge, F.R. Hallett, 'Size Distributions Obtained from the Inversion of $I(Q)$ Using Integrated Light Scattering Spectroscopy', *Macromolecules*, **27**, 2283–2290 (1994).
108. K.B. Strawbridge, F.R. Hallett, 'Polydisperse Mie Theory Applied to Hollow Latex Spheres: an Integrated Light-scattering Study', *Can. J. Phys.*, **70**, 401–406 (1992).
109. K.S. Shifrin, I.B. Kolmakov, 'Calculation of a Certain Class of Definite Integrals Containing the Square of a First Order Bessel Function (English Translation)', *Izv. USSR Acad. Sci. Atmos. Oceanic Phys.*, **3**, 749–753 (1967).
110. J.H. Chin, C.M. Sliepsevich, M. Tribus, 'Particle Size Distributions from Angular Variation of Intensity of Forward-scattered Light at very Small Angle', *J. Phys. Chem.*, **59**, 841–844 (1955).
111. J.H. Chin, C.M. Sliepsevich, M. Tribus, 'Determination of Particle Size Distribution in Polydisperse Systems by Means of Measurement of Angular Variation of Intensity of Forward-scattered Light at Very Small Angle', *J. Phys. Chem.*, **59**, 845–848 (1955).
112. H. Benoit, W.H. Stockmayer, 'A Study of the Influence of Interactions on Light Scattered by a Collection of Particles', *J. Phys. Radium*, **17**, 21–26 (1956).
113. S. Yue, G.C. Berry, 'Moderately Concentrated Solution of Polystyrene. Part 5. Static and Dynamic Light Scattering in bis(2-ethylhexyl) Phthalate', *Macromol. Symp.*, **98**, 1005–1027 (1995).
114. S.J. Chen, G.C. Berry, 'Moderately Concentrated Solutions of Polystyrene. 4. Elastic and Quasi-elastic Light Scattering at the Flory Theta Temperature', *Polymer*, **31**, 793–804 (1990).

115. S.H. Kim, D.J. Ramsay, G.D. Patterson, J.C. Selser, 'Static and Dynamic Light Scattering of Poly(α-methylstyrene) in Toluene in the Dilute Region', *J. Polym. Sci., Part B, Polym. Phys.*, **28**, 2023–2056 (1990).
116. N.C. Ford, Jr, 'Light Scattering Apparatus', in *Dynamic Light Scattering*, ed. R. Pecora, Plenum Press, New York, 7–58, 1985.
117. H. Utiyama, 'Light Scattering Instruments', in *Light Scattering from Polymer Solutions*, ed. M.B. Huglin, Academic Press, New York, 41–60, 1972.
118. H. Utiyama, 'Calibration and Correction Factors', in *Light Scattering from Polymer Solutions*, ed. M.B. Huglin, Academic Press, New York, 62–88, 1972.
119. B.B. Weiner, 'Particle and Droplet Sizing Using Fraunhofer Diffraction', in *Modern Methods of Particle Size Analysis*, ed. H.G. Barth, John Wiley & Sons, New York, 135–172, 1984.
120. P.E. Plantz, 'Particle Size Measurements from 0.1 to 1000 μM Based on Light Scattering and Diffraction', in *Modern Methods of Particle Size Analysis*, ed. H.G. Barth, John Wiley & Sons, New York, 173–209, 1984.
121. R. Frank, L. Frank, N.C. Ford, 'Molecular Characterization Using a Unified Refractive-index-light-scattering Intensity Detector', in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques*, eds. T. Provder, H.G. Barth, M.W. Urban, American Chemical Society, Washington, DC, 109–121, 1995.
122. T.H. Mourey, H. Coll, 'Size-exclusion Chromatography with Light-scattering Detection at two Angles', in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques*, eds. T. Provder, H.G. Barth, M.W. Urban, American Chemical Society, Washington, DC, 123–140, 1995.
123. E.F. Casassa, 'Particle Scattering Factors in Rayleigh Scattering', in *Polymer Handbook*, eds. E.H. Immergut, J. Branderup, John Wiley & Sons, New York, 485–491, Vol. VII, 1989.
124. G. Porod, 'X-ray and Light Scattering by Chain Molecules in Solution', *J. Polym. Sci.*, **10**, 157–166 (1953).
125. G.C. Berry, 'Thermodynamic and Conformational Properties of Polystyrene. I. Light-scattering Studies on Dilute Solutions of Linear Polystyrenes', *J. Chem. Phys.*, **44**, 4550 (1966).
126. L.L. Lloyd, J.F. Kennedy, C.J. Knill, 'Light Scattering and Chromatography in Combination', in *Light Scattering: Principles and Development*, ed. W. Brown, Clarendon Press, Oxford, 494–523, 1996.
127. H.G. Barth, R.B. Flippen, 'Particle Size Analysis', *Anal. Chem.*, **67**, 257–272 (1995) and prior reviews cited therein.
128. P.M. Cotts, 'Comparison of the Molecular Dimensions of Di-alkyl and Aryl-alkyl-substituted Polysilanes in Solution Using SEC/LS', *J. Polym. Sci., Pt. B, Polym. Phys.*, **32**, 771–778 (1994).
129. S. Shiga, E. Kato, 'Characterization of Polymers by GPC-LALLS.I. Computer Simulation of Fractionation of Branched Polymers by GPC', *Rubber Chem. Tech.*, **59**, 693–708 (1986).
130. M. Kurata, M. Fukatsu, 'Unperturbed Dimension and Translational Friction Constant of Branched Polymers', *J. Chem. Phys.*, **41**, 2934–2944 (1964).
131. C. Jackson, Y.-J. Chen, J.W. Mays, 'Dilute Solution Properties of Randomly Branched Poly(Methyl Methacrylate)', *J. Appl. Polym. Sci.*, **59**, 179–188 (1996).
132. B.A. Korgel, J.H. van Zanten, H.G. Monbouquette, 'Vesicle Size Distributions Measured by Flow Field-flow Fractionation Coupled with Multiangle Light Scattering', *Biophysical J.*, **74**, 3264–3272 (1998).
133. J.H. van Zanten, 'Characterization of Vesicles and Vesicular Dispersions via Scattering Techniques', in *Vesicles*, ed. M. Rosoff, Marcel Dekker, New York, 1996.
134. J.H. van Zanten, 'The Zimm Plot and its Analogs as Indicators of Vesicle and Micelle Size Polydispersity', *J. Chem. Phys.*, **102**, 9121–9128 (1995).
135. J.H. van Zanten, 'Unilamellar Vesicle Diameter and Wall Thickness Determined by Zimm's Light Scattering Technique', *Langmuir*, **10**, 4391–4393 (1994).
136. K.B. Strawbridge, L.R. Palmer, A.R. Merrill, F.R. Hallett, 'Integrated Light-scattering Spectroscopy, a Sensitive Probe for Peptide–Vesicle Binding, Application to the Membrane-bound Colicin E1 Channel Peptide', *Biophys. J.*, **68**, 131–136 (1995).
137. Z. Tuzar, P. Kratochvil, 'Scattering from Block Copolymer Micellar Systems', in *Light Scattering, Principles and Development*, ed. W. Brown, Clarendon Press, Oxford, 327–342, 1996.
138. J. Swithenbank, J. Cao, A.A. Hamidi, 'Spray Diagnostics by Laser Diffraction', in *Combustion Measurements*, ed. N. Chigier, Hemisphere Publishing, New York, 179–227, 1991.
139. P. Meyer, N. Chigier, 'Dropsize Measurements Using a Malvern 2000 Particle Sizer', *Atomization Spray Tech.*, **2**, 261–298 (1986).
140. D.J. Brown, K. Alexander, J. Cao, 'Anomalous Diffraction Effects in the Sizing of Solid Particles in Liquids', *Part. Part. Syst. Charact.*, **8**, 175–178 (1991).
141. S. Yue, G.C. Berry, M.S. Green, 'Intermolecular Association and Supramolecular Organization in Dilute Solution. 2. Light Scattering and Optical Activity of Poly(*p*-biphenylmethyl-L-glutamate)', *Macromolecules*, **29**, 6175–6182 (1996).
142. D.W. Tanner, G.C. Berry, 'Properties of Cellulose Acetate in Solution. I. Light Scattering, Osmometry and Viscometry on Dilute Solutions', *J. Polym. Sci., Polym. Phys. Ed.*, **12**, 941 (1974).
143. T. Nicolai, D. Durand, J.-C. Gimel, 'Scattering Properties and Modeling of Aggregating and Gelling Systems',

in *Light Scattering: Principles and Development*, ed. W. Brown, Clarendon Press, Oxford, 201–231, 1996.

144. H.G. Elias, 'The Study of Association and Aggregation Via Light Scattering', in *Light Scattering from Polymer Solutions*, ed. M.B. Huglin, Academic Press, New York, 397–457, 1972.
145. Y. Einaga, G.C. Berry, 'Studies on Dilute Solutions of Rod-like Macroions: III. Integrated Intensity and Photon Correlation Light Scattering Investigation of Association', in *Microdomains in Polymer Solutions*, ed. P. Dubin, Plenum Press, New York, 191–210, 1985.
146. S. Yue, G.C. Berry, R.D. McCullough, 'Intermolecular Association and Supramolecular Organization in Dilute Solution. 1. Regioregular Poly(3-dodecylthiophene)', *Macromolecules*, **29**, 933–939 (1996).
147. R. Furukawa, G.C. Berry, 'Studies on Dilute Solutions of Rod-like Macroions. 4. Aggregation with Enhanced Orientational Correlation', *Pure Appl. Chem.*, **57**, 913–920 (1985).
148. M. Sedlak, 'Polyelectrolytes in Solution', in *Light Scattering, Principles and Development*, ed. W. Brown, Clarendon Press, Oxford, 120–165, 1996.
149. V.J. Sullivan, G.C. Berry, 'Light Scattering Studies on Dilute Solutions of Semiflexible Polyelectrolytes', *Int. J. Polym. Anal. Charact.*, **2**, 55–69 (1995).
150. R.B. Russel, D.A. Saville, W.R. Schowalter, *Colloidal Dispersions*, Cambridge University Press, Cambridge, 343, 1989.
151. F. Grüner, W. Lehmann, 'On the Long Time Diffusion of Interacting Brownian Particles', in *Light Scattering in Liquids and Macromolecular Solutions*, eds. V. Degiorgio, M. Corti, M. Giglio, Plenum Press, New York, 51–69, 1980.
152. P. Doty, R.F. Steiner, 'Macro-ions. I. Light Scattering Theory and Experiments with Bovine Serum Albumin', *J. Chem. Phys.*, **20**, 85–94 (1952).
153. R.S. Samuels, *Structured Polymer Properties*, John Wiley & Sons, New York, 89–109, 1974.

Optical Particle Counting

Alvin Lieberman
Particle Measuring Systems, Inc., Alameda, USA

1 Introduction **1**
2 History **2**
3 Application Areas **2**
3.1 Determination of Product or Process Particle Size and Quantity 2
3.2 Determination of Fluid Cleanliness or Contamination Level 2
4 Operating Principles of Optical Particle Counters **3**
4.1 Light Scattering Optical Particle Counters 3
4.2 Light Extinction Optical Particle Counters 6
5 Performance Specifications for Optical Particle Counters and Factors Affecting Performance **7**
5.1 Sample Fluid Flow Rate 7
5.2 Background Noise Count Rate 8
5.3 Particle Sizing Accuracy and Size Measurement Range 8
5.4 Particle Size Resolution 8
5.5 Particle Counting Accuracy 8
5.6 Particle Concentration Limits 8
5.7 Instrument Stability and Verification 9
6 Standard Documents for Performance and Operation **9**
6.1 Gasborne Particle Counter Performance Documents 9
6.2 Gasborne Particle Counter Operation Documents 9
6.3 Liquidborne Particle Counter Performance Documents 10
6.4 Liquidborne Particle Counter Operation Documents 10
6.5 Document Sources 10
7 Operational Requirements for Particle Counting **10**
7.1 Sample Acquisition, Handling, and Storage Requirements 10
7.2 Calibration Specifications and Procedures 13
8 Performance Capabilities of Present Day Optical Particle Counters and Possible Future Developments **18**
8.1 Particle Size Capabilities 18
8.2 Particle Concentration Capability 19
8.3 Optical Particle Counter Configuration and Size 19
8.4 Capital and Operating Costs 20
Abbreviations and Acronyms **21**
Related Articles **21**
References **21**

This discussion of optical particle counting (OPC) points out application areas where the procedure is used to collect data that would not be available otherwise. These areas include determination of particle size distributions from the size data of individual particles, identification of particles that do not fit within a specified size distribution function, and characterizing the particles present in a process of product fluid. The designs, performance, and operation of OPCs for both gasborne and liquidborne particles are discussed. Capabilities and limitations in terms of particle size ranges, counting accuracy, concentration, and so on are pointed out. The primary limitation is related to the fact that OPCs respond not only to the size of the particles, but also to the optical nature of the particles and of the fluid. Data are also affected by the counter design. All respond to a physical and/or chemical property of the particles, as well as to their physical size and morphology. The limitations of OPCs in relation to those factors are considered.

Operational requirements for OPC are pointed out. These include good sample acquisition and handling procedures, consideration of the effects of differences between particle counter designs on reported data, requirements for careful calibration before use and the effects of differences between calibration particles and sampled particles. A brief summary of present national and international standard documents for calibration and operation of these instruments is included. Performance capabilities of current OPCs are summarized and possible future developments are discussed. Performance is considered in comparison with that of other particle measuring procedures.

1 INTRODUCTION

This discussion of OPC describes application areas for the process, the operating principles that are used for both light scattering and light extinction instruments, and the parameters affecting response of these instruments. The performance specifications and capabilities of present day instruments are summarized for measurement of particles suspended in gas or in liquid. Operational requirements in terms of sample acquisition and handling are pointed

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

out; effects of incorrect procedures are discussed. Calibration procedures and calibration materials are considered as they affect response and differences in response for instruments with design differences. The effects of such design differences are discussed. National and international procedure documents for calibration and operation of OPCs are summarized. Performance capabilities of present day instruments and possible future developments are summarized.

2 HISTORY

The first OPCs were modeled after optical devices developed for detection of airborne chemical and biological warfare agents in the late 1940s. These were based on light scattering systems used for testing high efficiency air filters. A laboratory instrument was produced for the US army and was modified for field use in 1954.[1] At that time, the airborne particle counter was capable of measuring particles in the "diameter" size range of 0.5–20 μm in concentrations up to approximately $3.5 \times 10^6\,m^{-3}$. In 1965, a light extinction counter was developed for measurement of contaminant particles in hydraulic fluids, fuels, mineral oils, and other lubricants. The liquidborne particle counter could measure particles in sizes from 5 to 200 μm in concentrations up to approximately $3.5 \times 10^5\,m^{-3}$.[2] Sensitivities and concentration capabilities have increased significantly for present day instruments and will probably continue to increase in the future. At the same time, the overall bulk and weight of instruments continues to decrease, while new electronic component developments permit more data acquisition, storage, and processing.

3 APPLICATION AREAS

3.1 Determination of Product or Process Particle Size and Quantity

Production and processing of a wide variety of materials are affected by the physical and chemical properties of finely divided materials that are either used in the production operation or are the end product. For example, reaction rates are affected by the available surface area of process materials. Reaction time requirements can be reduced by extending the reactant powder surface by reducing particle size. In handling bulk powders, power requirements to maintain flow through ducts can be reduced by including a range of particle sizes rather than using a single size batch of materials. Process effluents, when exhausted through a stack or in a gas stream, are often controlled by ordinances that specify allowable effluent opacity. The particle size of the material in that effluent will materially affect the visibility of the exhaust plume. This is due to the sensitivity to particle size of the extinction coefficient when particle size approaches the wavelength of visible light.

Many pharmaceutical products are applied by injection of a solids suspension in liquid or by the inhalation of aerosolized dry or liquid particles. The deposition location of such materials is frequently controlled by particle size. Correct operation of much material processing equipment includes verification that the product material particle size is within the specified size range limits. Particle size control may involve dry particle comminution or agglomeration or control of liquid droplet size. Essentially all of the particle size measurement in these application areas can be made either during or after production, using OPCs.

When the required information on particle size and/or quantity is only restricted to the mass or volume of particulate material or to the particle size distribution, OPC may not be the best method to obtain such information. Scattered light measurement of an array of particles, rather than of individual particles, can provide particle size distribution parameters directly. Collection and weighing a sample of the particles can provide particle mass data. However, when information is needed on whether or not all particles in the material being measured fit within the required size distribution limits, single particle sizing and counting by an OPC is frequently the only on-line method that can be used.

3.2 Determination of Fluid Cleanliness or Contamination Level

Ambient atmosphere contamination conditions are of importance in many production areas. Particulate air pollution in residential areas has been a concern for many years, but typical measurement methods did not involve detailed measurement of particle size. Environmental air pollution data for particulate materials has been based on either visibility of emission sources or on mass per unit volume of ambient air. During the 1990s, ambient air pollution in the USA became concerned with control of particles in the respirable size range. Specifications for concentrations of particles ≥ 2.5 μm and ≥ 10 μm are appearing. Further consideration of the physical and chemical effects of particles that are deposited and retained in the lungs can be expected. At this time, the procedures for measuring airborne particulate air pollution still require data in the form of particle mass per unit volume of sampled air. Measurement of particles by filtering an air sample and weighing the collected particles is now in use and is quite effective for particles larger than approximately 5 μm in diameter. However, the smaller particles require much larger air sample volumes

because of their much smaller mass. For this reason, OPCs can generate useful data much faster than the particle collection and weighing method.

In the industries producing electronic, pharmaceutical, optical, and precise mechanical components, the cleanroom production facilities require control of airborne particles to levels where no more than a specified number of submicrometer particles per unit volume of air can be present. Some semiconductor facilities allow no more than ten particles $\geq 0.1\,\mu m$ per cubic meter of air. Deionized water used for cleaning during production is filtered to cleanliness levels of no more than ten particles with diameter $\geq 0.05\,\mu m$ per liter of water. There are no formal specifications for these levels, but production history dictates such cleanliness requirements. That same history indicates that even cleaner production facilities will be required in the future. For operation of precise mechanical or hydraulic control systems, the fluid cleanliness requirements are also based on operational history. The cleanliness level requirements for those fluids are several orders of magnitude less severe than required by the semiconductor industry. In pharmaceutical and medical device production, the Food and Drug Administration (FDA) in the USA and similar agencies in other countries require that both air and water used in production be controlled to appropriate particulate cleanliness levels.

As a result of constant changes in product performance requirements in the semiconductor industry, the need to monitor continuously the cleanliness of the process fluids and of the air environment has resulted in development of some new OPC systems. These devices are used for in situ measurements within production tools during critical production stages. These in situ counters observe the cleanliness of the gases and liquids during such actions as chemical vapor deposition system pressure adjustments, operations, and movement of product elements prior to, during, and following those operations. These gasborne particle counters may be required to operate under extreme temperature and pressure conditions. In some cases, the particle counter sensing element is contained within the tool so that gas flow passes through that element. In other cases, the element is contained within a flow line containing fluids either to or from a process action. An in situ sample, consisting of a portion of the fluid flowing through that line may be measured. Care is required to ensure that the sample portion is sufficient to permit valid characterization of the particles in the total flow.

In a similar development that became important during the 1980s, continuous monitoring of a process or facility cleanliness, has been used to verify that product quality is not affected by contamination generation incidents. An OPC is used to observe fluid flow towards an in-process component or from a potential contamination source. Measurements can be made on a continuous basis or at specified time intervals. Procedures for measurement at a large number of sample points can be established. These procedures can use a single particle counter that receives sequential samples from a series of sample collection points. Each selected sample point feeds fluid into a line that is connected to a central source selection controller that sequentially selects one of the lines and feeds sample to the particle counter for a preselected time interval. The particle counter electronic system and/or external control system can report particle concentration for each of the selected sample points. Data from these measurements can be used to verify that the sample point or an area containing many sample points is in compliance with a fluid cleanliness specification.

4 OPERATING PRINCIPLES OF OPTICAL PARTICLE COUNTERS

4.1 Light Scattering Optical Particle Counters

The basic operation of a light scattering particle counter is shown in Figure 1. The operational part of the counter contains a light source, an optical system for beam shaping to define a sample view volume, a light collection optical system consisting of either lenses or mirrors to collect the scattered light, a photodetector that will convert optical signals to electrical signals, and a signal processing electronic system to convert the electrical signals to information on the size and number of detected particles. Ancillary flow control systems are required to handle the sampled fluid. Light scattering particle counters for liquidborne particles contain the fluid sample stream in transparent tubing, but the optical and electronic systems are essentially identical to those for the airborne particle counters. There is one other difference between some

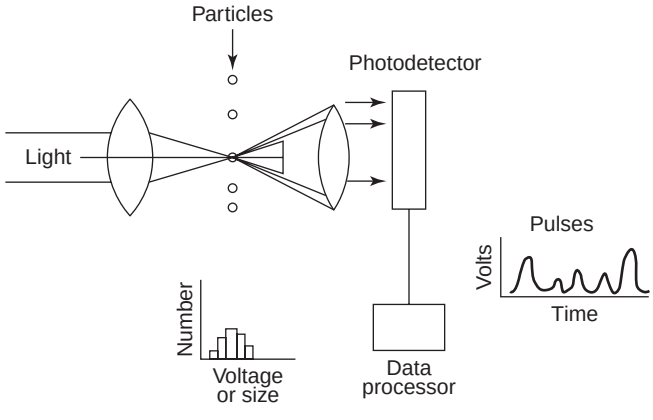

Figure 1 Light scattering OPC operation. (Adapted from A. Lieberman, *Contamination Control and Cleanrooms: Problems, Engineering Solutions, and Applications*, Van Nostrand Reinhold, New York, 291, 1992.)

airborne particle counters and the liquidborne particle counters. As will be seen later, the capability to detect very small particles increases as the illumination source intensity increases. When a gas laser is used as the light source, the beam intensity in the view volume is normally a very small fraction of the intensity of the beam present in the active cavity between the two mirrors where the gas is located. By replacing one of the containment mirrors with a Brewster's angle window, one of the mirrors can be moved sufficiently far from that panel that a space large enough to pass a gas stream containing the particles to be measured can be exposed to the very high light level present within the active cavity of the laser. This cannot be done when the fluid being sampled is a liquid. The refractive index of the liquid differs sufficiently from that of the gas in the laser and in the external active cavity so that insertion of the liquid degrades the beam in the ''external'' active cavity to the point where signals from particles are decreased significantly. Use of a solid-state diode laser that can emit light at high levels is used when very small particles ($\leq$0.1 μm) in liquid are to be measured.

When no particles are in the sensing volume, only the near-constant level light scattered by fluid molecules enters the collection optical system; the signal output from that source is below the signal level for particle data. As particles pass through the sensing volume, scattered light pulses are generated by each particle. The scattered light pulse pattern depends on the particle size and the ratio of the refractive index of the particle and of the substrate fluid. The refractive index ratio is shown in Equation (1)

$$m = \frac{n_p - ik_p}{n_m} \tag{1}$$

where m is the refractive index ratio, n_p is the real part of the particle refractive index, ik_p is the imaginary (absorption) part of the particle refractive index, and n_m is the refractive index of the fluid medium in which the particle is suspended.

Figure 2 illustrates scattering patterns for particles in several size ranges. The photodetector collects each light pulse that is scattered over the angle in which the collection optical system is operating. The photodetector

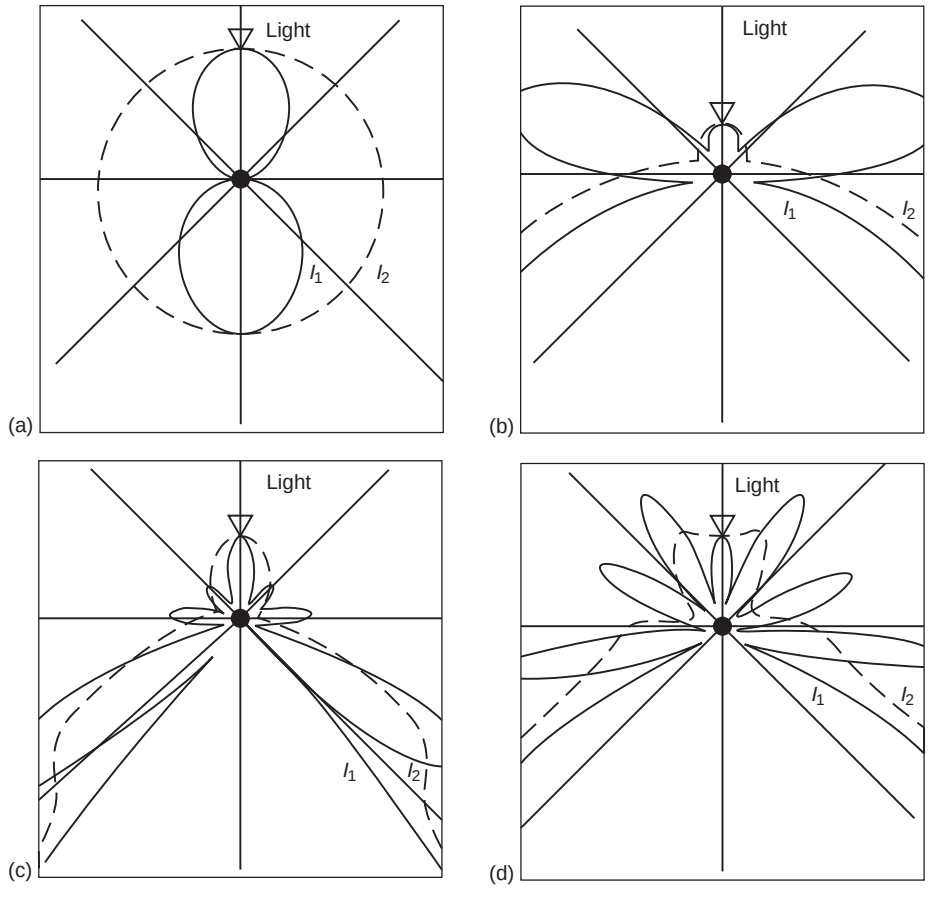

Figure 2 Radial intensity of light scattered from particles (spherical water droplets) of various sizes $\alpha = \pi d/\lambda$, scattered light flux polarized I_1 perpendicular and I_2 parallel to the incident light. (a) $\alpha = 0.5$, droplet diameter $d = 0.15$ μm; (b) $\alpha = 3.0$, $d = 1.0$ μm; (c) $\alpha = 4.8$, $d = 1.5$ μm; (d) $\alpha = 6.0$, $d = 2.0$ μm. (Reproduced by special permission from A. Lieberman, *Chem. Eng.*, **73**(2), 97–102 (1967). Copyright (1967) by McGraw-Hill Inc., New York.)

converts the light flux to an electrical signal and feeds that signal to the electronic system of the particle counter. Each pulse amplitude is measured and the pulses within defined amplitude ranges are counted. Pulse amplitude varies as a function of particle size. Since the sampled fluid flow rate is controlled, particle concentration in each size range can be determined.

Operation of the light scattering particle counter is neither as simple as the quick summary above may indicate, nor is it as complicated as might be expected from light scattering theory. Light scattering theory is complex. However, there are some fundamental parameters that affect particle counter response. These include (1) configuration of the illumination and collecting optical systems, summarized as scattering angle; (2) light intensity and wavelength; (3) size of the particle in the sensing volume; (4) the particle refractive index relative to that of the surrounding fluid. The scattering angle controls the quantity of light collected for any particle. For a larger solid angle, more light will be collected. Care is needed in optical design, since counters with larger solid angle collection optical systems can also collect background light reflected from components in the optical system. That background light is essentially optical noise signals that result in false data.

The basic optical system designs used in every light scattering particle counter are shown in Figure 3.[3] Minor differences exist in some instruments. For example, the right angle scattering optical system in the topmost design in Figure 3 has been modified to use a scattering angle

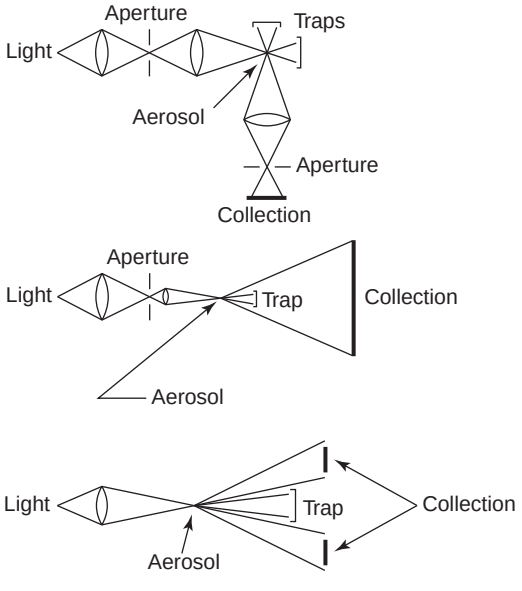

Figure 3 Optical systems for light scattering particle counters. (Adapted from J.R. Hodkinson, J.R. Greenfield, 'Response Calculations for Light Scattering Aerosol Counters', *Appl. Optics*, **4**(11), 1463–1474 (1965).[3])

base that is approximately 120°, rather than 90°. Dual detectors can be used to improve signal levels for better small particle measurement capability. The wide-angle collection system in the middle design uses a set of mirrors to collect the scattered light and to focus that light upon the photodetector.

The derivation of equations defining light scattering is available in the literature and will not be repeated here. Without going into detail, the important parameters that define scattering from a particle include the particle size, the refractive index ratio (including both real and imaginary components) of the particle and of the suspension medium, the illumination wavelength and the solid angle over which the light is collected. The theoretical relationship for light scattering by small particles is shown by Equations (2) and (3), the Rayleigh scattering equations for small particles. Equation (2), for perpendicularly polarized light, shows nearly equal scattering at all angles, while Equation (3), for parallel polarized light, shows scattering equal in intensity to that for perpendicular polarization at $\theta = 0$ rad and $\theta = \pi$ rad, while decreasing to zero at $\theta = 0.5\pi$ rad, as shown for the scattering from a 0.15 μm particle in Figure 2.

$$I_1 = \frac{\alpha^4}{2\pi}\left(\frac{m^2-1}{m^2+2}\right)^2 \tag{2}$$

$$I_2 = \frac{\alpha^4}{2\pi}\left(\frac{m^2-1}{m^2+2}\right)^2 \cos^2\theta \tag{3}$$

where I_1 is the scattered light flux polarized perpendicularly to the incident light, I_2 is the scattered light flux polarized parallel to the incident light, α is the particle circumference divided by the illumination wavelength, m is the particle refractive index, relative to that of the suspension medium, and θ is the scattering angle with respect to the forward direction in degrees.

Scattering theory states that the scattered light signal is a function of the wavelength of the incident light on the particle, the state of polarization of that light, the scattering angle, the refractive index ratio of the particle and the substrate fluid, the particle size, shape, and orientation. Light scattering from particles with dimensions at and below the wavelength of the incident light is identified as Rayleigh scattering. The scattered light signal level is proportional to d^6/λ^4, where d is the particle diameter and λ is the wavelength of the incident light. A small change in particle size changes the scattered light signal significantly. Light scattering from particles with dimensions at and above the wavelength of the incident light is identified as Mie scattering. Figure 2 shows scattered light patterns from particles in both size ranges. Equation (4) is the Mie scattering equation describing scattering for particles of larger size

with vertically polarized light.

$$I_1 = \sum_{m=1}^{\infty} \frac{2m+1}{m(m+1)} [a_m \tau_m(\cos\theta) + b_m \pi_m(\cos\theta)]^2 \quad (4)$$

where a_m and b_m are complex amplitude functions dependent on m and the particle diameter, and π_m and τ_m are angular functions depending on θ.

If a well-collimated beam of light is passed through a fluid, any particle with a light extinction coefficient other than zero that is in the light path will remove light by a combination of scattering out of the direct beam and by absorption. If the absorption part of the particle refractive index (k_p) is zero, light is removed from the beam solely by scattering. The particle extinction coefficient defines the combined effects of scattering and absorption. Beer's law, relating the transmitted and incident light flux, may be written as shown in Equation (5).

$$I = \exp(-Ac_nL) \quad (5)$$

where I is the ratio of light flux after interaction with the particle to incident light flux, A is the projected area of particle(s) illuminated by the incident light, c_n is the numerical particle concentration, and L is the beam pathlength through the effective sensing volume.

At this time, optical particle counters using scattered light signals for particle counting and sizing are in use with illumination sources using incandescent filaments, gas lasers, and solid state lasers. Incandescent filament sources (with their wide range of illumination wavelengths) are used mainly to reduce multivalue response effects. Laser sources are used for measuring particles significantly smaller than 1 μm in diameter. HeNe laser sources (illumination wavelength of some 633 nm) are used in the active or passive cavity mode to permit passing the observed particles through the sensing volume where illumination flux is at a maximum. Argon ion lasers (illumination wavelength of some 400 nm) are used as illumination sources for work with particles somewhat smaller than 0.1 μm in diameter. Solid-state laser sources (illumination wavelength in the range of 650–900 nm) are used as conventional illumination sources supplying high illumination levels. The small size and high reliability of high-power solid-state lasers, that operate without the need for high-voltage power supplies, has resulted in their increasingly widespread application. Particle counters using polychromatic light and large solid angle light collection systems produce the desired single-valued response, but sensitivity for small particle measurement is decreased and background noise must be minimized.

As of 1996, both airborne and liquidborne particles of 50 nm and larger could be counted and sized efficiently. Most light scattering optical designs use wide-angle light collection systems to maximize particle sizing resolution capability and to minimize multivalued response where the signal both increases and decreases as measured particle size increases. Right angle optical systems are widely used with Mangin mirrors (concave spherical mirrors that can collect and focus light to the photodetector) collecting the light over a total angle of $90° \pm 45°$. This system is now used for counting and sizing particles in the size range from 50 nm and larger. There are many optical designs used for particle counting and sizing instruments at this time. A more detailed discussion on design and operation of light scattering OPCs is given in a recent ISO (International Organization for Standardization) standard.[5] This discussion also covers air flow system design and operation to ensure that the fluid sample stream is completely within the sensing zone so that all particles are adequately illuminated.

The variety of designs and components result in significant differences in response when different instruments are used to characterize the same particle samples. Data variability is reduced by the following procedures:

1. Minimize differences in the optical designs of the instruments used.
2. Calibrate all instruments, preferably with certified reference particles with traceable dimensions, with all instruments using the same calibration procedures and materials.
3. Do not use instruments beyond their capabilities of particle size range, fluid flow rate, or environmental conditions.

4.2 Light Extinction Optical Particle Counters

The optical systems used for light extinction particle counters illuminate a small volume of flowing fluid and characterize particle content of that fluid by the reduction in the amount of light passing through that volume due to the particle(s) within the flowing fluid. The light beam is directed to the photodetector element. The fluid stream flows through a region of uniform illumination. When there are no particles in the illuminated volume of fluid, a constant signal is received by the photodetector. As a particle passes through the sensing volume, some light is removed by the particle as a result of scattering out of the direct beam and/or by absorption by the particle. Thus, a signal proportional to a function of particle size is generated. For particles smaller than approximately 5 μm, that signal varies with the refractive index of the particle. For larger particles the extinction coefficient stabilizes at a value of two, regardless of the refractive index of the particle. For this reason, differences in optical properties of the calibration particles and

the measured particles do not produce as great signal differences as occur for particles smaller than 5 μm. The scattering coefficient is strongly dependent on the particle/fluid refractive index ratio, especially for particles smaller than 5 μm in diameter. These instruments are used mainly for measurement of particles suspended in liquid. Some extinction instruments are also used for sizing and counting dry particles that are large enough so that they can be passed through the sensing volume by dropping them from a vibratory feeder and allowing gravity to move them through the light beam at a velocity sufficient to keep the residence time to no more than 50 μs or so. Figure 4 shows the optical system for an extinction optical counter that observes the full liquid sample stream. A transparent rectangular or circular flow tube defines the sensing volume cross-sectional area; the height of the sensing volume is defined by the configuration of the illumination beam. The beam height is normally a few tens of micrometers and the beam width extends across the full liquid sample stream; many systems have areas in the range of 0.5–1.0 cm^2.

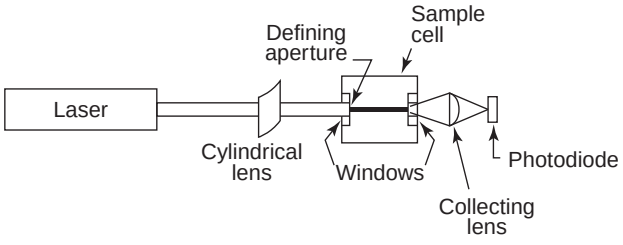

Figure 4 Optical systems for light extinction particle counter. (Reproduced by permission of Particle Measuring Systems, Inc.)

Essentially the same parameters affect the response of the OPC using light extinction as the counter using light scattering. The optical properties of the particle/fluid system affect the extinction system response markedly, particularly for small particles. This is due mainly to the fact that the refractive indices of many process and product liquids are frequently similar to the refractive index of the particles being measured. However, as the measured particle size increases above 5 μm, the extinction coefficient stabilizes at a value of 2.0 for essentially all refractive indices, reducing that cause of error. This effect is shown in Figure 5.[6] The projected area of the particle affects the amount of light removed from the direct beam and its orientation as it passes through the sensing volume in a manner similar to the effects pointed out for light scattering systems. The response curve for the extinction optical counter is a power function that varies from approximately the fourth power of particle size for particles in the micrometer size range and decreases to approximately the second power for particles larger than approximately 5 μm in diameter. Most light extinction OPCs use an optical system that observes light that is nearly a totally parallel beam. The beam intensity through the sensing volume is close to uniform, with minimum differences due to the normally

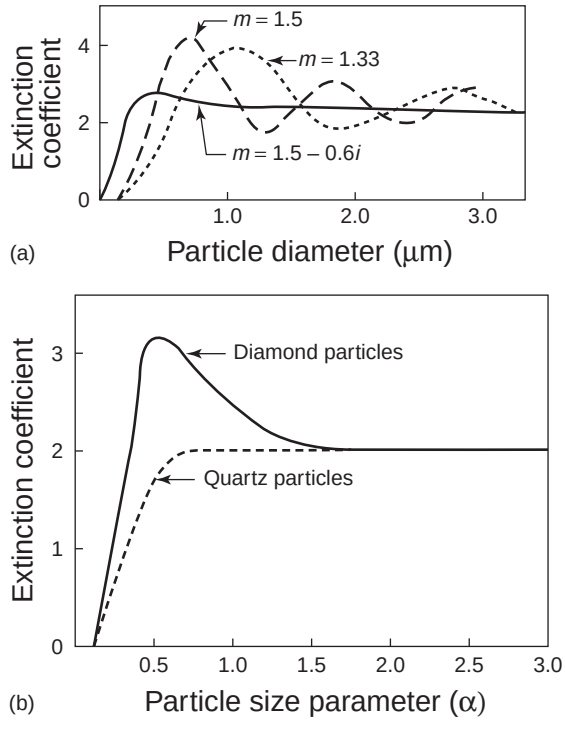

Figure 5 Theoretical and experimental light extinction coefficients. (a) Mie theory, (b) measured data. (Adapted from J.R. Hodkinson, 'The Optical Measurement of Aerosols', in *Aerosol Science*, ed. C.N. Davies, Academic Press, New York, 1966.[6])

Gaussian intensity distribution in a laser beam or the intensity variations from a ribbon filament incandescent light source. For this reason, particle sizing resolution for light extinction OPCs is excellent unless assembly or alignment problems cause excessive variation of light intensity in the sensing volume. This characteristic of extinction systems improves data reliability, particularly when measuring particles with wide variations in shape factor or composition.

5 PERFORMANCE SPECIFICATIONS FOR OPTICAL PARTICLE COUNTERS AND FACTORS AFFECTING PERFORMANCE

5.1 Sample Fluid Flow Rate

Since much of the data desired from OPC applications involves particle concentrations, accurate sampled fluid

flow rate data are a requirement. Specifications for such data require measurement of fluid flow with an external flow-measuring device that will not generate excessive pressure drop for gasborne particle counters. For liquidborne particle counters, a flow measuring device that can be operated to avoid indication of multiphase (gas and liquid) fluids as single phase fluids is necessary.

5.2 Background Noise Count Rate

When seeking maximum size sensitivity, OPCs may be operated with electronic systems set at maximum gain. In this situation, random electronic or optical noise may generate pulses large enough so that they can be reported as signals from small particles. In this case, erroneous data are generated that indicate more particles than are actually present in the measured sample. Operating the particle counter with particle-free fluid entering the sensing volume can identify this situation. Use of a filter with pore size no more than half the size of the smallest measurable particle on the fluid flow line can satisfy this requirement. The counter can be operated in this manner for time periods up to 30 min for two or three measurements and the observed noise count rate can be stated.

5.3 Particle Sizing Accuracy and Size Measurement Range

The need for sufficient data to provide statistical validity in characterizing particles has resulted in development of particle counters that can count and size particles of the smallest possible size with good accuracy and can portray the size distribution over as much as possible of the actual size range of the sample being measured. For light scattering particle counters used for either gasborne or liquidborne particles, size measurement of 0.05-μm particles is possible with several commercially available instruments. It is possible to obtain size data up to as much as 100 times the smallest possible size measurement. For light extinction OPCs, calibration particles $\geq 1\,\mu m$ can be measured. The other end of the dynamic range capability is at 100 times the smallest size. The particle sizing accuracy should be no less than 95%. These data are based on the modal diameter stated for a batch of calibration particles that have been measured by an accepted particle sizing procedure.

5.4 Particle Size Resolution

Size resolution is the capability to differentiate between particles of nearly the same size. Accepted sizing resolution is the ability to differentiate between particles that differ by more than 5% in diameter. This value is determined by measuring a batch of near-monodisperse calibration particles that have been measured by a reference method that has provided data on the modal size and the relative standard deviation of sizes. The reported relative standard deviation should be no more than 5% greater than the specified standard deviation from the reference method data.

5.5 Particle Counting Accuracy

Although particle counting accuracy is considered a requirement mainly for determination of particle concentration when the quantity of particles is of concern, this is also a problem when establishing particle size distribution if any of the size ranges being measured do not report the particle population accurately. The particle counting accuracy is verified by counting aliquot samples of a particle suspension by the particle counter under test and by a referee method. The referee method can be an instrument other than the OPC or it can be an OPC that has been verified as producing accurate particle count data over the size range of concern. Counting accuracy requirements are typically stated as $50\% \pm 5\%$ at the most sensitive size capability and $100\% \pm 5\%$ for all particle sizes at least 1.5 times larger than that most sensitive size.

5.6 Particle Concentration Limits

The maximum concentration that can be accurately measured by any particle counter is limited by the accepted coincidence error. This error is due to the probability that more than one particle is present in the sensing volume at any time. The error level is a function of the physical dimensions of the sensing volume, the actual particle concentration within the sample being measured, and the particle size distribution of the particles in that sample. For samples where all particles are so small as to be insignificant in dimensions in comparison to

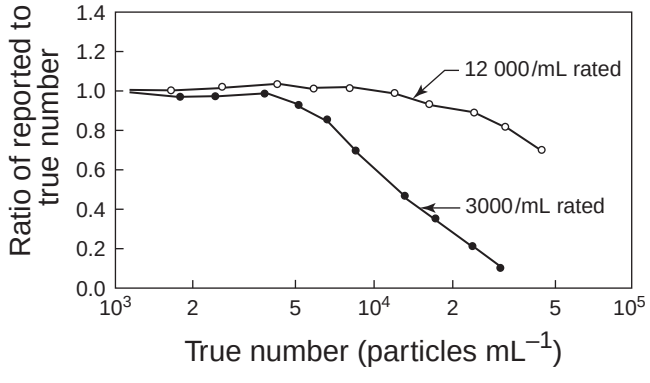

Figure 6 Coincidence effect on reported concentration accuracy. (Adapted from A. Lieberman, *Contamination Control and Cleanrooms: Problems, Engineering Solutions, and Applications*, Van Nostrand Reinhold, New York, 357, 1992.)

those of the sensing volume, the probability of more than one particle being present within a defined volume can be calculated directly from statistical theory,[7] as shown in Figure 6.[8] Figure 6 shows the ratio of reported concentrations to true concentrations for two optical liquidborne particle counters; one had a rated concentration capability of 12 000 particles per mL and the other a capability of 3000 particles per mL. However, the electronic circuitry in any particle counter cannot respond to the presence of a particle entering the sensing volume until the signal from a previous particle in the sensing volume has decreased to the system reset level. When particles are so large that their size is significant in comparison with a sensing volume dimension, the presence of such a particle causes an effective increase in the sensing volume dimension, as shown in Figure 7.[9] The presence of large particles in the sample decreases the reported particle concentration limit more than expected from the calculation of the probability of coincident small particles in the sensing volume alone.

5.7 Instrument Stability and Verification

This all-inclusive performance requirement is handled by ensuring that the calibration procedure fits the application requirements and that calibration is carried out at no more than the maximum time interval specified for the particle counter in use or the regulatory document specified for the application where the measurements are being made. Details of calibration procedures are provided below where a summary of some national and international standard documents and their sources are summarized. Not all of the existing standard documents are included, but those indicated cover essentially all of the information required to carry out the procedures and operations that will aid in generating valid data for particle counting.

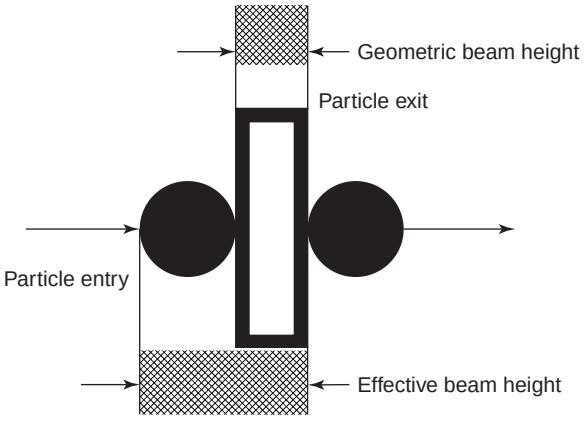

Figure 7 Large particle effect on sensing volume effective dimensions. (Reproduced by permission of Marcel Dekker, Inc. from J.Z. Knapp, L.R. Abramson, 'Validation of Counting Accuracy', in *Liquid- and Surface-borne Particle Measurement Handbook*, eds. J.Z. Knapp, T.A. Barber, A. Lieberman, Marcel Dekker, Inc., New York, 1996.)

6 STANDARD DOCUMENTS FOR PERFORMANCE AND OPERATION

6.1 Gasborne Particle Counter Performance Documents

1. ASTM F328: Determining Counting and Sizing Accuracy of an Airborne Particle Counter Using Near-monodisperse Spherical Particles (ASTM, American Society for Testing and Materials)
2. ASTM F649: Standard Practice for Secondary Calibration of Airborne Particle Counters Using Comparison Procedures (ASTM)
3. BS 3406: Part 7. Determination of Particle Size Distribution: Recommendations for Single Particle Light Interaction Methods (BSI, British Standards Institution)
4. ISO 13323-2: Determination of Particle Size Distribution–Single Particle Light Interaction Methods–Part 2: Light Scattering Single Particle Light Interaction Devices Design, Performance Specifications, and Operation Requirements (ISO)
5. JIS B9921: Light Scattering Automatic Particle Counter (JIS, Japanese Industrial Standards)
6. VDI 3489(3): Particulate Matter Measurement: Methods for Characterization and Monitoring of Test Aerosols: Optical Particle Counter (VDI, Verein Deutscher Ingenieure)

6.2 Gasborne Particle Counter Operation Documents

1. ASTM F50: Continuous Sizing and Counting of Airborne Particles in Dust-controlled Areas and Clean Rooms Using Instruments Capable of Detecting Single Submicrometer and Larger Particles (ASTM)
2. BS 5295(4): Environmental Cleanliness in Enclosed Spaces. 4. Specification for Monitoring Clean Rooms and Clean Air Devices to Prove Continued Compliance with BS 5295: Part 1 (BSI)
3. Federal Standard 209E: Airborne Particulate Cleanliness Classes in Cleanrooms and Clean Zones, Appendix B (IEST, Institute of Environmental Sciences and Technology)
4. ISO 14644-3: Cleanroom Measurements and Metrology (ISO)

5. JIS B9920: Measuring Methods for Airborne Particles in Clean Rooms and Evaluating Methods for Air Cleanliness of Clean Rooms (JIS)

6.3 Liquidborne Particle Counter Performance Documents

1. ASTM F658: Defining Size Calibration, Resolution, and Counting Accuracy of a Liquid-borne Particle Counter (ASTM)
2. ANSI/NFPA T2.9.6R2: NIST (National Institute of Standards and Technology) Traceable Calibration Method for Liquid Automatic Particle Counters (ANSI, American National Standards Institute)
3. DIN 50432: Part 2, Test Method for Particle Analysis in Liquids: Determination of Particles with Automatic Particle Counter (DIN, Deutsches Institut fur Normung)
4. JIS B 9925: Light Scattering Automatic Particle Counter for Liquid (JIS)
5. ISO 13323-3: Determination of Particle Size Distribution – Single Particle Light Interaction Methods – Part 3: Light Extinction Single Particle Light Interaction Devices Design, Performance Specifications, and Operation Requirements (ISO)
6. ISO 11171: Hydraulic Fluid Power – Calibration of Liquid Automatic Particle Counters (ISO)

6.4 Liquidborne Particle Counter Operation Documents

1. JIS K 0554: Testing Method for Concentration of Fine Particles in Highly Purified Water (JIS)
2. SEMI C1-88: Calibration & Measurement Methods for Particles in Liquids (SEMI, Semiconductor Equipment Manufacturers International)
3. ANSI B93-73M: In-line Liquid Automatic Particle Counting Systems – Method of validation (ANSI)

6.5 Document Sources

- ANSI (American National Standards Institute), 41 West 42nd Street, 13th floor, New York, NY 10036, USA
- ASTM (American Society for Testing and Materials), 100 Barr Harbor Drive, W. Conshohocken, PA 19438, USA
- BSI (British Standards Institution), 389 Chiswick High Road, London, W4 4AL, UK
- DIN (Deutsches Institut fur Normung e. V), Burggrafenstr. 6: Postfach 11 07, D-1000 Berlin 30, Germany
- IEST (Institute of Environmental Sciences and Technology), 940 E. Northwest Highway, Mt. Prospect, IL 60056, USA
- ISO (International Organization for Standardization), CP-56, 1 Rue de Verembe, 1211, Geneve 20, Switzerland
- JIS (Japanese Industrial Standards), 4-1-24 Akasaka, Minato-ku, Tokyo 107, Japan
- VDI (Verein Deutscher Ingenieure VDI-TGA), Graf-Reck-Strasse 84 D-4000, Dusseldorf, Germany

7 OPERATIONAL REQUIREMENTS FOR PARTICLE COUNTING

7.1 Sample Acquisition, Handling, and Storage Requirements

Particle counting is used to carry out three types of particle measurement. The particle size and size distribution of samples from bulk powders are defined; the particulate content of ambient air or liquid is measured; the particulate content of process or product gas or liquid is measured. The three application areas have one common requirement that must be satisfied before any useful data can be collected. This is the requirement that a valid sample be acquired and that the sample be handled so that no data are collected that have been generated by improper sample handling procedures. The procedures and the reasons for these procedures will be discussed for the three application areas.

7.1.1 Bulk Powder Sampling Procedures

When particle counting is used to make measurements of bulk powders, the objective is to determine a particle size distribution and to define outlier particles that are either too small or too large to fit into the size distribution correctly. During transport or storage, the particles in the powder will frequently pack in such a way that portions of the powder bulk will contain particles with a size distribution that differs from the original distribution. As an example, vibration may cause small particles to settle towards the bottom of the powder container by falling through the interstices between large particles. Even so, it is still necessary to procure a representative sample of the total material. This may require special care in taking samples from a container where polydisperse material is stored. To sample correctly, an accepted procedure should be followed including subsample preparation via spinning riffler or other accepted powder subdivision procedure.[10] This procedure may require more than one repetition to produce samples small enough for measurement by OPC. The sample is then dispersed into

the fluid in which it will be measured. It should be followed immediately by measurement of the particles suspended in that fluid. If samples are dispersed into liquid, stirring or agitation is required before measurement of the suspended particles is carried out. This dispersion should be carried out in such a way that a minimum quantity of air is aspirated into the suspension. If any gas is present, gas bubbles will form and may be interpreted as particles. Transport of fluid suspension in ducts or tubes should be carried out with tubing that does not contain any sharp bends or short radius elbows. These can cause deposition of particles due to centrifugal force during transit through the tubing. Excessive residence time (more than 3 min in gas or 5 min in liquid) in the tubing should be avoided. Large particles may be lost to tube walls via gravitational settling or turbulent flow impaction losses. Small particles may be lost via diffusion to the tubing walls or by agglomeration with other particles. Small dry powder samples can be stored in closed containers for long time periods with no significant changes in the samples. Powders dispersed in gas cannot be stored longer than a few minutes unless the particles are so small that diffusion by the gas molecules will keep the particles dispersed until size measurements are made. For particles in liquids, samples stored for more than 30 min or so will require stirring or agitation before particle size or concentration measurements are made unless no particles larger than 5 μm are in the powder.

7.1.2 Ambient Air Sampling Procedures

Ambient air sampling procedures are often specified by regional air pollution control agencies. Until the 1980–1990 period, the requirements for air pollution control of inert particulate materials were based only on mass loading of total airborne particles with no specifications as to the size of those particles. Most requirements for acceptable particulate levels in the ambient air are expressed in terms of milligrams per cubic meter. The particle population for those levels is beyond the detection capability of all but a few research type OPCs. In the late 1970s, information from industrial hygiene studies and research on particulate material retention in the respiratory tract generated concern about the particle size of inert airborne particles. From that time, the EPA (Environmental Protection Agency) in the USA, along with similar agencies in other countries, began specifying that data on the mass loading of particles $\geq$2.5 μm and $\geq$10 μm should be collected. For this reason, it can be expected that ambient air measurement using OPC and sizing may also become a requirement. These measurements will require sample acquisition procedures that collect particles with sizes $\geq$2.5 μm and $\geq$10 μm in such a way that the concentration of those particles can be defined correctly. Isokinetic sampling or an equivalent procedure will be required for airborne particles larger than approximately 5 μm. Sample storage for particles to be measured using an OPC is not anticipated. The optical system will be incorporated with the sampling system for immediate measurement with data either stored or transmitted electronically to a central data processing facility.

For air flow sampling isokinetic sampling is required for all particles larger than approximately 5 μm; for liquid flow sampling, isokinetic sampling is required for particles larger than approximately 10 μm. Isokinetic sampling requires that the sample probe inlet flow is at the same velocity as that of the sampled fluid flow. In areas where the sampled fluid flow rate may vary, then isoaxial sampling is recommended. In areas where sampled fluid flow direction and three-dimensional velocity may vary in a random manner, then a sample probe that faces upward is recommended so that ''large'' particles will be sampled with maximum accuracy. In essentially all areas where processor product fluid cleanliness data are of concern, the OPC has been widely used. In production areas for the electronics industry and the pharmaceutical industry, the standard procedures specify OPC instruments for particle size and concentration data production. For areas where particles smaller than 0.1 μm are of concern, scanning electron microscopy measurement of particles that have been collected from fluid samples is also used. However, the higher capital costs, labor time, and delay between sample acquisition and data availability connected with that method has resulted in use of OPCs in many areas where measurement of particles $\geq$0.05 μm is required.

7.1.3 Process and Product Fluid Sampling Procedures

Many products presently in common use are based upon complex technological components. Most of the operating parts of computers, automobiles, communication systems, and optical devices are components that must be produced in clean areas where particulate contamination in the air in production areas would cause immediate or later failures. Pharmaceutical and other health care products such as injectable liquids and medical devices must be produced in areas where both viable and inert particle content of the environmental air and the process liquids are carefully controlled. Hydraulic fluid power control systems contain components with moving parts operating with very small clearances. Particulate contaminants in the hydraulic fluid must be controlled to minimize control action errors. Performance of optical components, such as high-accuracy mirrors, long-range lens systems, film for use in medical applications of electron microscopy, and so on, can be degraded if contaminant particles in micrometer sizes are deposited upon those critical surfaces

during production. Most automobile painting facilities are now operating with ventilation air systems containing high-efficiency filters to prevent any atmospheric particulate debris from degrading those surfaces by deposition upon the still-wet paint. Air cleanliness is of paramount importance in essentially all areas where particle contaminants may degrade product quality. The facilities where particle-sensitive production occurs, known as ''cleanrooms'' are kept as clean as possible by controlling the ventilation air cleanliness with a filter which has a particle removal efficiency from 99.97% for particles $\geq$0.3 μm (for routine operations) to removal efficiencies of 99.9999% for particles $\geq$0.1 μm for semiconductor production. Unidirectional airflow moving at a velocity of 0.5 m s^{-1} is used in very clean areas. The concentration of particles $\geq$ 0.1 μm in some work spaces is kept at levels of no more than 100 m^{-3}. Air flow in the most sensitive areas is controlled so that air movement is unidirectional with minimum recirculation of potentially contaminated air in the cleanroom. The objective is to verify that the particle concentration in that air is less than 10–100 particles per cubic meter. Air sampling in critical areas is continuous, with particle counters sampling air at 0.47 m^3 s^{-1} and sizing and counting all particles $\geq$0.1 μm in diameter. The particle counter sample line inlet faces into the air-flow path and the inlet velocity is controlled so that isokinetic sampling is assured. In areas where products are not as sensitive as the semiconductor products, the air flow direction and three-dimensional velocity may be random; a sample probe that faces upward is recommended so that ''large'' (>5 μm) particles will be sampled with minimum loss. If the particle counter is not located directly in the sampled air flow, a sample probe may be located at the end of a sample transport tube. Experience has shown that residence time in the tube must be minimized and that turbulent flow is desired to minimize gravitational losses. For valid measurement of particles smaller than 2 μm, tubing lengths up to 30 m can be used with the Reynolds number for the gas flow maintained between 5000 and 25 000. For valid measurement of particles of 5 μm and larger, tubing lengths should not exceed 2 m. The same limitations on the Reynolds number for air flow apply for the larger particles. These limitations ensure that laminar air flow is not present in the tubes and that turbulence is not so great that excessive particle impaction occurs upon the tubing walls.

When product or process liquids are to be examined for particle content, the sampled liquid will be in a process or product flow line or batch samples already in containers will be examined. Acquisition of samples from a line containing flowing liquid requires that two factors be considered. First, it is necessary that samples be procured from a line that is completely filled with liquid. Sampling from zones of mixed gas and liquid should be avoided. The bubble content of liquid from such areas cannot be anticipated and the quantity of liquid removed for each sample is difficult to measure with any accuracy. The second requirement is that the sample should be removed so that the removal does not cause loss of any particles due to centrifugal force produced by excessive curvature of the sample flow path. When process or product liquids are to be measured for particle content, the sampled liquid will either be in a flow line, or batch samples will be examined. Acquisition of samples from a flowing line requires that a sample probe line be inserted into the sampled line. The sample probe line should not be configured so that the sampled liquid is moved through a short radius-of-curvature elbow or a line that imposes centrifugal force upon the particles in the sampled liquid. Figure 8 illustrates two recommended configurations and one less desirable configuration for liquid flow line sampling. The same configuration should be considered for possible use in gas flow sampling from lines where high velocity or high pressure gas is to be sampled. As previously mentioned, sampling for measurement of large particles is best accomplished with isokinetic sample flow control. After a flow line sample has been passed through the sensing volume of the OPC, disposal or return of the measured sample to the process line may be required. Care is required in disposal of liquids that may be hazardous. Consideration of the Material Safety Data Sheet (MSDS) requirements for handling as well as local and other municipal safety requirements in

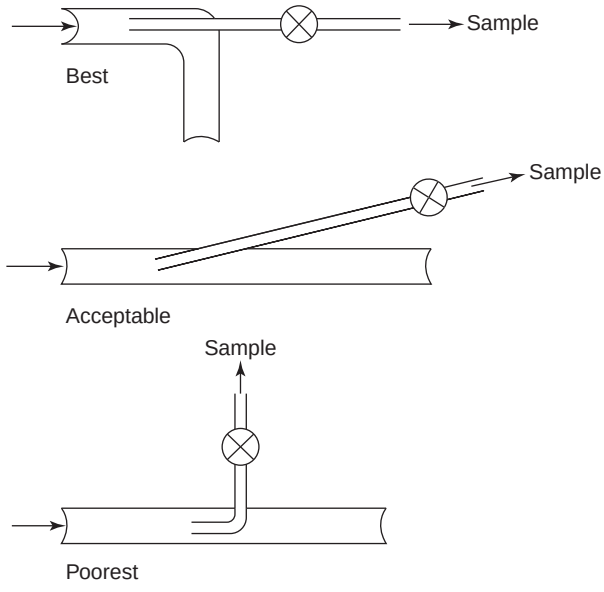

Figure 8 Sidestream sample probe configurations. (Adapted from A. Lieberman, *Contamination Control and Cleanrooms: Problems, Engineering Solutions, and Applications*, Van Nostrand Reinhold, New York, 398, 1992.)

handling process and product liquids must be part of the sample handling and disposal protocol.

If batch samples are to be measured, these samples should be in containers that can be placed in a sample feeder attached or attachable to the OPC. Use of ''round-bottom'' sample containers is suggested, whenever possible, so that any necessary stirring does not result in particles being deposited upon vertical container walls. Before samples are measured, the sample container should be shaken, or otherwise agitated, to ensure that particles are randomly distributed in the container, rather than settled at the container bottom after long periods of storage, or deposited upon container walls by centrifugal force from stirring. Before any samples are measured by removal from a container, dissolved gases should be removed by exposing the liquid to low-level ultrasonic vibration and to vacuum just before samples are withdrawn from the container. A sample withdrawal tube should be located with its inlet located in the lower third of the container. Sample volumes sufficient to produce a particle count of 1000 to 10 000 are suggested. The concentration of particles should be at a level where three replicate aliquot samples can be acquired from the container and no more than approximately 75% of the total container volume is removed from the container during this process. Note the particle count for each of the aliquot samples. The count for any sample should not differ by more than 20% from the average for all three samples. The same concern for proper disposal of measured samples and residual liquids in the sample containers and treatment of the containers is required as was used for liquids from process of product line sample disposal.

7.1.4 Inter- and Intrainstrument Variability and Means of Minimizing Differences

As was shown in Figure 2, the scattered light intensity varies markedly with scattering angle for all particles larger than approximately 0.3 μm. In addition, Figure 3 points out the differing optical system bases for light scattering particle counters. Considering both situations, it is apparent that the same size and type of particle measured by two dissimilar light scattering OPCs will produce different data. Light extinction OPCs use optical systems with collection angles that may be less than 5°, but scattered light directed out of that small angle will vary with the wavelength of the illumination source for the counter. Extinction counters that are in use now and probably for another decade or so use incandescent filament sources, gas lasers, and solid-state laser illumination sources with significant differences in light wavelengths. For this reason alone, when the calibration particles differ from the particle being measured, instruments with different light sources will generate different data from the measured particles. Although the differences in data are primarily in the reported particle sizes, the effect of reported size differences also affects the reported particle concentration when data on particle size distribution or on particle concentration in a specified size range is required. There are also differences in data from same model OPCs produced by any manufacturer. These arise from the fact that tolerances in manufacturing drawings may be enough to result in scattering angles that can differ slightly between instruments of the same model from any manufacturer. These phenomena may result in production of particle size data differing by no more than a few percent. However, if incorrect calibration materials or procedures are used, or if the instrument calibration is not current, then significant data differences can be expected between instruments. Procedures to minimize interinstrument variability can be summarized as:

1. Use instruments that are similar in design and operation.
2. Make sure that the operators are familiar with instrument operation.
3. Specify a calibration procedure and use calibration particles similar to the particles to be measured.
4. Verify that every phase of the calibration is current.
5. Operate the instrument within its specified capabilities for particle size and concentration.
6. Establish a consistent sample handling protocol

7.2 Calibration Specifications and Procedures

7.2.1 Calibration Specifications

The parameters to be calibrated for any OPC include:

1. Maximum optical and electronic noise pulse generation rate. This rate should be no more than 3–5% of the rate anticipated for the minimum particle counting rate.
2. Fluid sample flow rate. This should be established with minimum accuracy of 95% of the reported inlet sample flow rate.
3. Particle sizing accuracy. This parameter should be established with minimum accuracy of 95% of the smallest reported particle size, based on measuring monodisperse calibration particles with diameters defined by a recognized documented method.
4. Particle sizing resolution. This parameter is the increase in reported relative standard deviation of a monodisperse calibration particle suspension and shall be no more than 5% greater than the stated relative standard deviation of the calibration particles.

5. Particle counting efficiency. This parameter defines the accuracy with which the OPC counts individual particles at the minimum detectable size and at all larger defined sizes. The counting accuracy for monodisperse particles at the minimum detectable size should be 50% ± 10% and the counting accuracy for particles at least 1.5 times larger than the minimum detectable size should be 100% ± 10%.
6. Particle concentration limit. This parameter defines the maximum concentration where the probability of multiple particles within the sensing volume is no more than 10%.

7.2.2 Calibration Procedures

Brief descriptions of calibration procedures for OPCs will be provided here. For further details, the reader is referred to the documents listed in section 6 of this discussion. Section 6.1 lists some calibration procedures for gasborne particle counter performance verification. Section 6.2 lists some operation procedures for gasborne particle counters. Section 6.3 lists some calibration procedures for liquidborne particle counter performance verification. Section 6.4 lists procedures for liquidborne particle counter operation. Complete descriptions of these procedures will not be provided here; the reader is directed to the documented procedures of section 6. Sources for the procedures are also provided in section 6.

7.2.2.1 Performance Calibration Procedures for Optical Particle Counters

Sample Flow Rate. To report suspended particle concentration accurately, the measured fluid sample flow rate into the counter must be defined accurately. Flow is measured by a volumetric flow measurement device that operates with low-pressure drop that will not degrade performance of the counter flow control system. If a mass flowmeter is used for calibration of the sample flow rate for a gasborne particle counter, then correction to the specified ambient or standard temperature and pressure conditions must be calculated and used in the concentration determination. Sample flow rate should be defined with a minimum accuracy of 95%.

Noise Count Rate. Noise count rate calibration is carried out to ensure that data, particularly in the smallest size range capability of the counter, are generated by particles rather than by background noise pulses from the electronics or optical systems. The noise count rate measurement is made by operating the particle counter with particle-free fluid moving through the sensing volume. Recording the noise count rate for a specified time period provides valid data. The noise count rate should not exceed 5% of the count rate anticipated for the anticipated particle count rate.

Particle Sizing Accuracy. Although the particle counter may be used to characterize particles containing materials that vary in shape and composition, a consistent response to selected calibration particles is required. For many OPC sizing accuracy calibrations, isotropic spherical latex particles are most often used for this purpose. A suspension is generated and the modal pulse amplitude of the pulses produced from a batch of particles of known size and relative standard deviation is defined as the signal level for particles of the same size as that batch of calibration particles. Since the OPC response varies significantly with refractive index of the particles being measured, calibration with latex spheres with a refractive index of 1.6 and with no absorption may cause a problem when the counter is to be used for measurement of particles with different optical properties. In this situation, a solution of a known concentration of material with properties similar to those of the material to be measured can be prepared in a liquid with appreciable vapor pressure. Monodisperse droplets can be generated using a vibrating orifice droplet generator; short exposure to air will cause the liquid to evaporate from the droplet leaving a residual particle of the desired physical characteristics. Knowing the size of the monodisperse droplets and the solution concentration, the size of the residual droplet can be calculated and used for calibration purposes. Alternately, polydisperse particles with optical properties similar to the particles to be measured can be used if the particle size distribution for that suspension is known.

Particle Sizing Resolution. This parameter is determined after particle size calibration is carried out for the full size range of the counter being calibrated. If a suitable pulse height analyzer is available, either as an external instrument, or incorporated in the electronic section of the counter being calibrated, data obtained during size calibration can be used to define the relative standard deviation of the pulse height distribution when the monodisperse latex calibration particles are measured. Pulse size data for particles at least twice the size of the lower sizing limit of the counter can be used. These data are converted to values for the average particle size and standard deviation values. The increase (if any) in relative standard deviation, compared to the labeled value of the calibration particles, is reported as the particle size resolution of the counter being calibrated. A maximum reported relative standard deviation no more than 10% greater than that of the calibration particles is acceptable, but resolution values in the 5% range are preferred.

Particle Counting Efficiency. Particle counting efficiency is affected by several factors. If sample flow is not completely contained within the defined sensing volume, some portion of the particles in the sample will not be seen. If sizing resolution is poor, some particles smaller than the lower size boundary of any size range will be reported as being within that range, resulting in excessive particle count data in that range. Counting efficiency for an OPC is determined by producing a well-mixed suspension of particles in a closed chamber. Samples are withdrawn from a single location in the chamber by the counter being calibrated and by a reference particle counting device. That reference device is one that is known to have 100% counting efficiency for the smallest particles in the test suspension. Sample transport to both the counter being calibrated and the reference system uses sample handling systems that are designed so that particle losses during transit from the chamber to both instruments are either negligible or identical. Samples are counted by both instruments and counting efficiency of the calibrated counter is defined as the ratio of the data reported by that counter to the data reported by the reference instrument. Primary counting efficiency data are obtained using suspensions of monodisperse particles. Counting efficiency for specific materials can be procured with a suspension of polydisperse test particles. The counting efficiency so obtained may vary from the primary counting efficiency data.

Particle Concentration Limit. If a high particle concentration may be present in the sampled fluid, more than one particle may be present in the sensing volume at any time. Even if the concentration is known to be no more than the concentration equivalent to one particle per sensing volume, then the fact that particle distribution in the sample is random, rather than uniform, will result in more than one particle being present in the sensing volume at some time. Several small particles may be reported as a single larger particle, resulting in a report of larger and fewer particles than are actually present. As sample concentration increases, the probability of coincidence error also increases. The maximum recommended concentration accepted by many organizations is that where the coincidence error is no more than 10%. The particle concentration limit is determined by producing a series of particle suspensions carefully diluted and sampled by the instrument being calibrated. The initial sample concentration should be in the range between the maximum recommended concentration and twice that level. The diluted suspensions are prepared so that the concentration of each succeeding suspension is reduced by a constant known factor from that of the previous suspension. Dilution ratios between 0.1 and 2 can be selected. The reported concentration of each succeeding suspension is recorded and the reported concentration of that suspension is compared to that of the previous suspension. At concentrations above the maximum recommended concentration capability, the ratio of counts in successive suspensions will be less than the dilution ratio. This indicates that coincident particles were present in the more concentrated suspension. When both suspension concentrations are below the recommended maximum concentration limit, the reported concentration ratio for two succeeding suspensions will be equivalent (95% to 105%) to the dilution ratio. The concentration of the more concentrated suspension at this point is the recommended maximum particle concentration for that particle counter.

7.2.2.2 Operational Procedures for Airborne Particle Counters After the particle counter has been calibrated, it is ready to size and count particles in the sample that is to be measured. However, there are some precautions required to maintain the sample with no artifact introduction or particle loss. The requirement for isokinetic sampling in gases has been discussed previously and will not be repeated here. The procedures for isokinetic sampling involve measurement of the sampled gas stream velocity, knowledge of the particle counter inlet flow rate and selection of an inlet probe dimension that will establish the sample probe inlet velocity at the sampled gas velocity. Even when the particles to be counted are so small so that anisokinetic sampling will not cause particle count errors, isokinetic sampling may still be required to ensure that the sampled gas source can be identified exactly. When air to be sampled is not moving at a constant velocity or is static, isokinetic sampling cannot be used. The primary concern in this situation is collection of particles $\geq 5\,\mu m$ with minimum losses. Figure 9 describes the particle size and sampled air-flow parameters that allow accurate sampling for a vertical thin-wall sample probe with an upward-facing inlet.[11] Sampler inlet efficiency is shown as affected by the Stokes number (defined here as the ratio of the stopping distance of a particle to the diameter of the sample probe inlet) and relative settling velocity of the sampled particle (the ratio of the particle settling velocity in still air to the sample probe inlet velocity).

In many air sampling operations for cleanroom performance verification, intermittent sample collection from a number of locations may be required to verify that tool and processing operations are not producing contaminant particles. In this situation, economic and space limitations may dictate a sampling procedure wherein a single OPC is connected to a sampling manifold system. This system consists of sample tubing from each sample point to the manifold system that directs samples to the particle

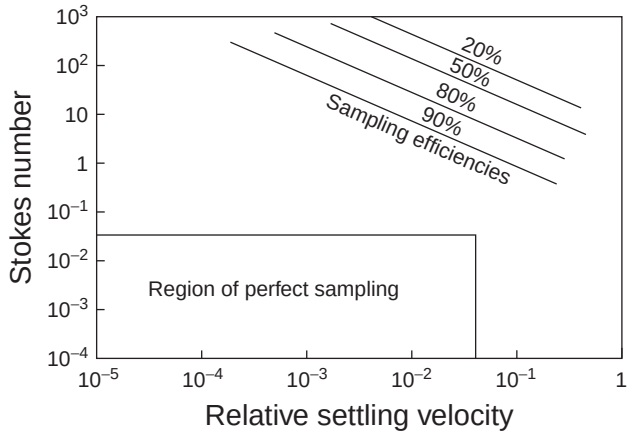

Figure 9 Sample tube efficiency in turbulent or still air. (Reproduced by permission of American Industrial Hygiene Association from J.K. Agarwal, B.Y.H. Liu, *Am. Ind. Hyg. Assoc. J.*, **41**, 191–197 (1980).[11])

counter inlet with minimum loss in the transfer process. However, the sample tubes must be configured correctly so there is minimum particle loss in the tubing during transit to the manifold and particle counter. This means that the tubing should not be so long that residence time in the tube will cause excessive loss of particles larger than 1–2 μm due to deposition caused by gravitational or turbulence effects. The tubing should not have an excessively small diameter (<6 mm) or pressure drop in the tubing will reduce the flow excessively. The tubing should have no bends with radius of curvature less than 15 cm to avoid centrifugal effects at the bend. The tubing material should be selected to minimize electrostatic charge by triboelectric effects. Stainless steel tubing is best if permanent installations are considered and cost is acceptable. Reactive metal (copper or aluminum) should not be used. Conductive polymer tubing with conductivity equivalent to that of plasticized polyvinyl chloride is recommended when sample line location or configuration may be modified as product or process changes dictate. Plasticized polyvinyl chloride is quite useful for handling particulate suspensions in many fluids, but even small quantities of the dioctyl phthalate plasticizer vapor may cause problems in product performance. Fluorocarbon or glass tubing is not recommended because of the electrostatic charge generated and retained by this tubing from aerosol flow.

7.2.2.3 Operational Calibration Procedures for Liquidborne Particle Counters The calibration procedures are summarized briefly here. For full details refer to the documents listed in section 6.

Sample Flow Rate and Volume. Liquidborne particle counters measure particles in liquid samples in several possible modes. The sample may be aspirated into the counter sensing volume by placing a batch sample container under the sensing volume, with a tube leading from the container to the sensing volume, reducing the pressure in the sensing volume, and drawing a sample through the sensing volume with subsequent disposal of the tested liquid. The sample may be forced through the sensing volume by increasing the pressure in the sample container to force the liquid upward through the sensing volume. A flowing stream of sample liquid may be forced through the sensing volume by placing a sample line inlet in that flowing stream with the sample line outlet connected to the counter sensing volume.

When aspirating samples from a sample container, the contents of that container must be well mixed before samples are measured. The aspiration process must not reduce the static pressure on the sensing volume to the point where dissolved gas in the liquid will form bubbles or to the point where the pressure in the sensing volume is below the vapor pressure of the liquid at the measurement temperature. When the samples are moved through the sensing volume by pressurizing the sample container, the pressure must be maintained at a constant level or the flow rate will change; this may cause the signal pulses to become so short that erroneous sizes are reported from those pulses. If the flow rate decreases excessively, then the liquid velocity upward through the sample tube to the sensing volume may decrease to the point where the gravitational settling velocity of large particles in that sample may exceed the upward velocity, with the result that no large particles move through the sensing volume. Figure 10 illustrates a pressure feeder for batch liquid samples in a container.

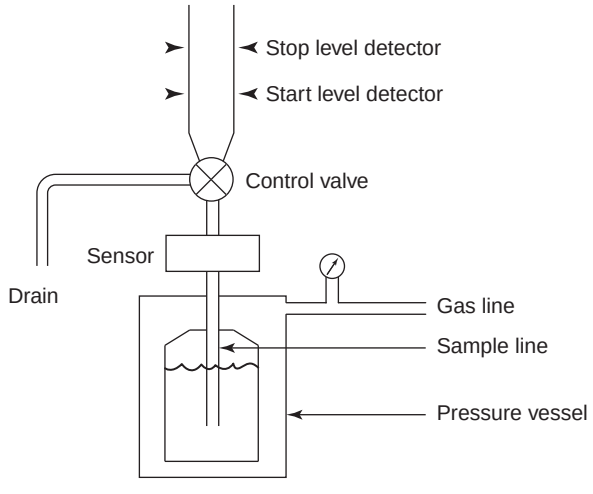

Figure 10 Pressure feeder for liquid samples in a container. (Adapted from A. Lieberman, *Contamination Control and Cleanrooms: Problems, Engineering Solutions, and Applications*, Van Nostrand Reinhold, New York, 401, 1992.)

When measuring samples taken from a process or product stream, isokinetic sampling is encouraged, but is seldom required for particles smaller than 10 μm in diameter. Sample removal from a process or product line by a sidestream sample line should be at least 12 line diameters before or after a change in line orientation or size. A sidestream can be removed from the sampled main line and the entire sidestream passed through the sensing volume. The sidestream volumetric flow rate should be controlled so that the flow is equal to the flow at which the particle counter was calibrated. Flow rate through the sidestream line can be measured with an internal flowmeter, preferably a nonintrusive type.

Background Noise Pulse Rate. With no liquid flowing through the sensing volume, the electronic system is set to report data in the cumulative mode. The first size range is set to the manufacturer's level for the lowest size sensitivity of the instrument. Approximately 25 mL of clean liquid is added to a carefully cleaned sample container and the sensor and sample line are flushed with three 10-mL batches of clean liquid without collecting any data. The sample feeder is set to feed a sample of at least 20 mL to the sensing volume for particle measurement. The previous sample passage of three replicate samples is repeated and the particle count data recorded. The average count reported per milliliter of liquid is determined. This value will represent the noise count rate of the instrument. Noise counts should not be generated at a rate greater than 5% of the counting rate anticipated for the samples to be measured.

Particle Sizing Accuracy. Two procedures are available for determining particle sizing accuracy. Either monodisperse particles of verified sizes can be used to set the sizing response of the instrument or a suspension of polydisperse particles with a verified size distribution and concentration can be used for that purpose. The choice is based on the degree of similarity of either of these calibration particles to the particulate materials to be measured. When the monodisperse particles are chosen, the procedure for setting the particle counter response is essentially identical to the procedure used for the gasborne particle counter calibration. Use of monodisperse latex particles for sizing calibration is required to calibrate particle counters used for measurement of pharmaceutical and health care products, to verify potable water cleanliness, to measure particles with little absorption, and in areas where measurement of submicrometer particles is required. The polydisperse particles are used to calibrate instruments used for measurement of particle content of hydraulic fluids, fuels, and lubricants. The photomicrographs of Figure 11 illustrate the difference between the two calibration particle types. The monodisperse particles are transparent and identical in size and configuration. The polydisperse particles are irregular in configuration and do not all have the same optical properties, as noted by the differences in shade between different particles.

The cumulative particle concentration per milliliter of suspension can be calculated from data, similar to those shown in Table 1. This table provides an example of the cumulative concentration of polydisperse particles as threshold size increases. Batch-to-batch data will be within the limits given in Table 1 for materials fitting within the specification for ISO ultrafine test dust (UFTD). Similar tabulated information is available for ISO fine, ISO medium, and ISO coarse test dusts to permit calibration over a wider size range.

There is one problem found when using the polydisperse calibration particles that is present with all the ISO test dusts. This problem is caused by the wide range of particle concentrations from the smallest to the largest particles in the batch. The result is that statistically valid particle count data cannot be obtained for the large particle sizes unless the suspension concentration is so high

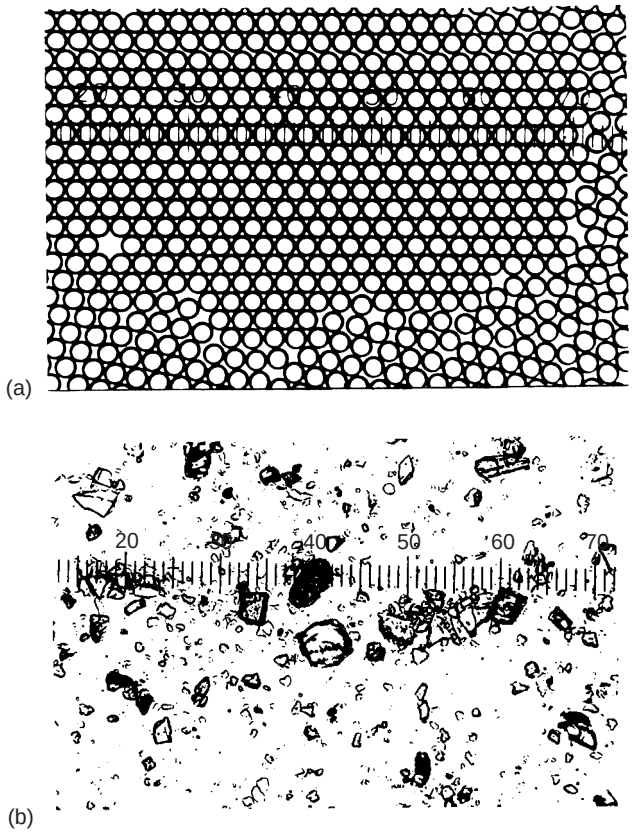

Figure 11 (a) Monodisperse 10-μm polystyrene latex particles and (b) ISO fine test dust. The scale in both photomicrographs defines 2.2 μm. (Reproduced by permission of Duke Scientific Corp.)

Table 1 ISO UFTD particle size distribution, number of particles per microgram

Particle size (μm)	Cumulative concentration, ISO UFTD particles ≥ stated size must be between:
1	70 000 and 100 000
2	19 000 and 29 000
3	8600 and 13 000
4	5400 and 8100
5	3300 and 4500
6	1600 and 2500
7	660 and 1400
8	280 and 760
9	120 and 410
10	58 and 220
11	28 and 120
12	14 and 63
13	7.4 and 34
14	4.1 and 19
15	2.3 and 11

that the small particle concentration exceeds the maximum concentration capability of the particle counter. In this situation, calibration for large particle sizing takes advantage of the situation shown in Figure 5. Monodisperse calibration particles can be obtained in the ''large'' sizes where there are insufficient large particles in the polydisperse calibration suspension for statistically valid data. As seen in Figure 5, the refractive index does not affect the extinction coefficient. Therefore, use of the large monodisperse particles permits definition of the counter response in the sizes where the polydisperse suspension will not provide adequate data.

Particle Sizing Resolution. Particle sizing resolution is determined using monodisperse calibration particles at least twice as large as the lower sizing limit of the instrument. The increase in relative standard deviation for the reported particle size range for the monodisperse particles selected for this measurement should be determined. The same procedures used to determine sizing resolution for the airborne particle counter could be used here. Most particle counter specifications for sizing resolution state that the counter will not increase the relative standard deviation shown for the monodisperse calibration particles by more than 10%. If the data fit that level, this requirement is met.

Particle Counting Accuracy. A batch of monodisperse particles with diameter in the size range from 5 μm to 25 μm is selected. These limits are chosen to permit ease of viewing at the small size end of the range and to minimize problems of loss due to deposition, settling, or inadequate dispersion at the large particle size end of the range. A suspension of these particles is prepared and the total particle concentration in the suspension is determined. Three aliquot samples of the suspension are passed through the sensing volume of the particle counter being validated and through the sensing volume of a particle counter that has been confirmed to have 100% counting efficiency for the particles in the size range selected. Both counter sensing volumes should be connected in series with a short length of tubing. The counter being calibrated should report count data equivalent to the data reported by the verified counter.

Particle Concentration Limit. The maximum concentration that can be accurately measured is limited by the accepted coincidence error. This error is due to the probability that more than one particle is present within the sensing volume at any time. The procedure for determining the maximum concentration limit is to pass a series of diluted suspensions with known dilution ratios through the sensing volume and to note the point at which the reported concentration ratio from one suspension to the previous suspension in the series is equivalent to the dilution ratio for the two suspensions. Since many liquidborne particle counters are used for measurement of a wide range of particle sizes, polydisperse particle suspensions are used to determine the maximum recommended particle concentration for liquidborne OPCs.

8 PERFORMANCE CAPABILITIES OF PRESENT DAY OPTICAL PARTICLE COUNTERS AND POSSIBLE FUTURE DEVELOPMENTS

8.1 Particle Size Capabilities

The smallest discrete particle that can be reliably counted and sized by commercially available OPCs for either gasborne or liquidborne particles is 0.05 μm in diameter. This capability applies for liquidborne particle counters used for measurement of particles in liquids that can be handled in glass and stainless steel with no attack on the counter materials of construction. If liquids as aggressive as hydrofluoric acid are to be used, then the sensitivity is decreased so that the smallest particle reliably counted and sized in those liquids is 0.075 μm in diameter. It is anticipated that the minimum size sensitivity will be 0.03 μm for gasborne particle counters and for liquidborne particle counters handling inert liquids before the end of the year 2000. No estimate of the maximum sensitivity for these instruments can be made at this time, but the need to define particles as small as possible is a continuing demand on the development capabilities of OPC producers. As part of the sizing capability of OPCs, particle sizing resolution is also a performance parameter

of concern. As an example, the calibration procedures and specifications in the section 6 documents for health care product testing specify that sizing resolution shall be no worse than 10%, particularly for liquidborne particle counters used to characterize large-volume and small-volume injectable liquids. Better control of illumination uniformity from laser light sources is expected to improve sizing resolution for both gasborne and liquidborne particle counters to the point where resolution worse than 5% will not be accepted in the health care product area after the year 2001. The other parameter connected with particle size measurement is the dynamic range of the instrument. This parameter defines the range of particle sizes that can be measured accurately. Particle counters with a 100 : 1 dynamic range are available. For application to measurement of particles in areas where a wide range of materials are used, a dynamic range of 1000 : 1 would be required. Use of advanced electronic systems with built-in autoranging amplifier gain control systems will permit operation of the particle counter over larger dynamic ranges. The capability of reporting particle size distribution in a very large number of size ranges has been seen in some listings of particle counter performance specifications. The capability to report accurate size data in more than 16 size ranges requires particle sizing resolution of 5% or better.

8.2 Particle Concentration Capability

In making measurements to characterize the particle content of fluids, the need for statistically valid data imposes a requirement to count sufficient particles to define particle content with a specified confidence limit (often 95%) as rapidly as possible. For gasborne OPCs, the capability of present day counters is adequate for most applications. Counters with gas flow rates of $47\,cm^3\,s^{-1}$ or $470\,cm^3\,s^{-1}$ can normally handle particle concentrations of $35\,cm^{-3}$ or $3.5\,cm^{-3}$. In situations where the particle concentration to be defined is beyond the capability of the counter, gas dilution can be used to decrease the measured concentration to the capability of the particle counter in use. The dilution system can be selected for a dilution ratio that will permit the true sample concentration to be defined correctly. For consistent measurement mainly of high concentrations of gasborne particles, as may be expected for many atmospheric research studies, optical gasborne particle counters with very small sensing volumes can be used to measure particle concentrations up to $2 \times 10^4\,cm^{-3}$. For valid measurements at such concentrations, sample gas flow rates may be in the range of $0.1\,cm^3\,s^{-1}$.

Particle concentrations in liquids are frequently higher than those in gases. Thus, the sample flow rates used in liquidborne OPCs are usually lower than those used for gasborne counters. For measurement of particle concentrations up to $10^4\,mL^{-3}$, a liquid sample flow rate of $0.0017\,mL\,min^{-1}$ is used. For measurement of particle concentrations up to approximately $35\,mL^{-3}$, a liquid flow rate of $1.67\,mL\,min^{-1}$ is used. Under these conditions, the liquids of concern can be characterized with present day instrument capability. As requirements for product quality increase, it is anticipated that there will be only a minor requirement for instruments that can measure particles in concentrations higher than present day instruments can handle. One example of the area where such a need may exist is in characterization of high-efficiency filter systems where the challenge concentration must be high enough so that even the few particles that penetrate the filter will be sufficient to permit reliable definition of the filter efficiency for particles at the most penetrating particle size.

8.3 Optical Particle Counter Configuration and Size

The particle counter described by Fisher et al.[1] was built in 1954. It operated with an electronic system using vacuum tubes. The optical system used an incandescent filament lamp for illumination and a photomultiplier tube photodetector. The electronic system required two 1.8-m relay racks to contain those components. The optical system was a right angle scattering system using lenses to direct the illumination beam and to collect the scattered light. The optical system weight was approximately 67 kg and the electronic system weight was approximately 135 kg. The instrument counted and sized particles from 1 to 64 μm in 12 size ranges. The sample air flow rate was $47\,mL\,s^{-1}$ and the maximum concentration capability was 1.1 particle cm^{-3}. This instrument can be compared with a recently described high-sensitivity airborne particle counter.[12] The counter is capable of counting and sizing airborne particles $\geq 0.05\,\mu m$ in an air sample flow of $47\,mL\,s^{-1}$. The illumination source is a solid-state laser. Parallel processor array detection permits background noise from air molecule scattering to be minimized. Light scattering over a scattering angle of $90° \pm 45°$ is collected by Mangin mirrors (concave spherical mirrors that collect and focus light to the photodetector) and focused upon a solid-state photodetector array. This instrument has a dynamic range capability of 100 : 1, allowing measurement of particles from 0.05 to 5 μm in diameter with sizing resolution of 3%. Data are reported in eight size ranges in either tabular or graphical format. The maximum recommended particle concentration is $35\,cm^{-3}$. Instrument operating parameters and data processing are controlled by an internal microprocessor. The entire instrument is in a package with dimensions of 100 cm width, 175 cm depth, and 25 cm height. The weight is approximately 15 kg. Data from this instrument can be

printed on a small on-board printer, stored internally on diskette, or transmitted to a central computer.

The comparison between the counter described in 1955 and that described in 1988 indicates that developments in performance have continued and that improvements in the performance of these instruments have not stopped. It is anticipated that future development will result in sensitivity to smaller particles – at a relatively slow development rate – and better sizing resolution as illumination source uniformity improves. The detection capability will also be modified. By adding a photodetector capability at wavelengths other than that of the illumination source, identification of some particle chemical composition becomes possible. This feature is being developed to identify microbiological materials and will probably be extended to other materials as new lasers are developed with more emission wavelengths. This type of activity is mainly aimed at identification of material type (organic, inorganic, etc.). Laboratory instrumentation for detailed analysis of single micrometer size particles has been used for several years.[13] Aerosol particles were collected and were analyzed by electron probe X-ray microanalysis and laser microprobe mass analysis. These techniques were used for trace element measurements by particle-induced X-ray emission. As more powerful lasers become available, the possibility of real-time analysis of individual airborne particles drawn into an instrument and analyzed in situ becomes a distinct possibility. This will permit identification of sources when ambient air pollutants are measured, as well as controlling product uniformity for powders.

Other areas where changes can be expected are in the counter size, functions, cost, and maintenance requirements. Better portability of OPCs has been a need expressed by users where modifications in process or product require measurements at different locations on an unscheduled basis as required by changing process conditions. This is a particular need when personnel must have a particle counter at a location where access is difficult or adequate space is a problem. The need for portable OPCs has resulted in development of battery-operated counters that are handheld, weigh 1 kg or so, and are capable of counting and sizing particles $\geq 0.3\,\mu m$ in sample air flows of $47\,cm^3\,s^{-1}$. These counters can report data in two to six particle size ranges. The electronic system can include a data storage capability of several hundred measurements with records of sample point location, time of measurement, and particle concentration data in the specified size ranges. Internal components are easily and quickly removable and replaceable with calibrated components that are kept in stock and can be inserted by any technician with minimum training required. A routine calibration procedure can be carried out in a central metrology facility for these components. The procedure is compatible with the procedures in the documents shown in section 6. Both the instrument producers and users should keep current issues of the calibration procedure documents at hand and be aware of any changes that have been made in the procedures. A ''calibration certificate'' should be prepared and filed with the cleanroom records for each instrument. The counter calibration status should be indicated on a tag attached to its case.

Particle measurement is the primary function of OPCs. However, particle morphology is also of interest in many areas. As digital photography technology continues to advance, the use of an image analyzer to report magnified image data on a two-dimensional pixel array can provide two-dimensional image information. This would be similar to the data based on photomicrograph images. By adding a second image detector to observe the particle in a direction orthogonal to that used by the first detector, a three-dimensional image may also be definable.

8.4 Capital and Operating Costs

Actual prices for OPCs will not be given here. During the interval between preparation of this article and distribution of the printed material, both technical developments and production costs will change pricing of OPCs. However, certain considerations will still affect the prices. An increased quantity of data to be reported will increase costs. Counters with the capability to measure very small particles will be more expensive than those with sensitivity for particles larger than 0.3–0.5 μm. A further increase in price by a factor of two to five can be expected when measurement of particles smaller than 0.1 μm is required.

Liquidborne particle counters require sample feeders that may increase system cost by a factor of twice the particle counter cost, depending upon requirements imposed by the nature of the liquid and the sample handling requirements for that liquid. Hazardous liquids of all types require costly containment and safety systems. Even with this significant capital cost, the savings in labor and facility costs when individual samples must be handled, measured and data processed for each one may result in savings that could pay for the sample feeder in a reasonable time. Uniformity of sample handling operations can also result in data that are more reliable than data produced by manual sample handling procedures.

Eventually the OPC will become a commodity product, similar to the way in which a voltmeter or oscilloscope is. For organizations large enough to incorporate a metrology facility, repairs and calibration can be carried out in that facility. For smaller organizations, express

shipments to and from a central repair and calibration facility may become the maintenance procedure for these and many other instruments. If this situation occurs, then capital costs for the OPC will drop to some extent and the labor costs for service will decrease even more.

ABBREVIATIONS AND ACRONYMS

ANSI	American National Standards Institute
ASTM	American Society for Testing and Materials
BSI	British Standards Institution
DIN	Deutsches Institut fur Normung
EPA	Environmental Protection Agency
FDA	Food and Drug Administration
IEST	Institute of Environmental Sciences and Technology
ISO	International Organization for Standardization
JIS	Japanese Industrial Standards
MSDS	Material Safety Data Sheet
NIST	National Institute of Standards and Technology
OPC	Optical Particle Counting
SEMI	Semiconductor Equipment Manufacturers International
UFTD	Ultrafine Test Dust
VDI	Verein Deutscher Ingenieure

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Optical Coherence Tomography

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory • Laboratory Instruments in Clinical Chemistry, Principles of

Coatings **(Volume 2)**
Microscopy of Coatings

Environment: Trace Gas Monitoring **(Volume 3)**
Airborne Instrumentation for Aerosol Measurements

Environment: Water and Waste **(Volume 3)**
Asbestos Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Sampling Considerations for Biomonitoring • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses

Food **(Volume 5)**
Particle Size Analysis in Food

Industrial Hygiene **(Volume 5)**
Industrial Hygiene: Introduction • Aerosols and Particulates Analysis: Indoor Air • Direct Reading Instruments for the Determination of Aerosols and Particulates

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Centrifugation in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Filtration in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Sedimentation in Particle Size Analysis • Turbidimetry in Particle Size Analysis • Velocimetry in Particle Size Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Diesel Fuels Analysis • Lube Products. Molecular Characterization of Base Oils • Lubricant Base Oils: Analysis and Characterization of • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams

REFERENCES

1. M.A. Fisher, S. Katz, A. Lieberman, N.E. Alexander, 'The Aerosoloscope: An Instrument for the Automatic Counting and Sizing of Aerosol Particles', *Proceedings of the Third National Air Pollution Symposium*, April 18–20, Pasadena, CA, USA, 1955.
2. L.D. Carver, 'Particle Size Analysis', *Ind. Res.*, June 1964.
3. J.R. Hodkinson, J.R. Greenfield, 'Response Calculations for Light Scattering Aerosol Counters', *Appl. Optics*, **4**(11), 1463–1474 (1965).
4. C.F. Bohren, D.R. Huffman, *Absorption and Scattering of Light by Small Particles*, John Wiley & Sons, New York, 1983.
5. ISO 13323-3, 'Determination of Particle Size Distribution – Single Particle Light Interaction Methods – Part 2: Light Scattering Single Particle Light Interaction Devices Design, Performance Specifications, and Operation Requirements, 1999.
6. J.R. Hodkinson, 'The Optical Measurement of Aerosols', in *Aerosol Science*, ed. C.N. Davies, Academic Press, London, 289, 292, 1965.
7. R. Jaenicke, 'The Optical Particle Counter: Cross-sensitivity and Coincidence', *J. Aerosol Sci.*, **30**, 95–111 (1972).

8. A. Lieberman, 'Flow Rate and Concentration Effects in Automatic Particle Counters', *Proceedings National Conference on Fluid Power*, Chicago, IL, 1975.
9. J.Z. Knapp, L.R. Abramson, 'A New Coincidence Model for Single Particle Counters. III: Realization of Single Particle Counter Counting Accuracy', *J. Pharm. Sci. Tech.*, **50**(2), 99–122 (1996).
10. B.H. Kaye, 'Efficient Sample Reduction of Powders by Means of a Riffler Sampler', *Soc. Chem. Ind. Monogr.*, **18**, 159–163 (1964).
11. J.K. Agarwal, B.Y.H. Liu, 'A Criterion for Sampling in Still Air', *Am. Ind. Hyg. Assoc. J.*, **41**(3), 191–197 (1980).
12. A. Lieberman, 'A New 0.05 μm, 0.1 CFM Optical Particle Counter', 1137–1143, *Proceedings 20th DOE/NRC Nuclear Air Cleaning Conference*, Boston, MA, Aug 22–24, 1988.
13. P. Artaxo, R. van Grieken, F. Watt, M. Jaksic, 'The Microanalysis of Individual Atmospheric Aerosol Particles by Electron, Proton, and Laser Microprobe', *Proceedings of Second World Congress on Particle Technology*, Kyoto, Japan, Sept. 19–22, 421–426, 1990.

Photon Correlation Spectroscopy in Particle Sizing

Walther Tscharnuter
Brookhaven Instruments Corporation, Holtsville, NY, USA

1 Introduction 1
2 History 1
3 Principles of Dynamic Light Scattering 2
3.1 Origin of Intensity Fluctuations 2
3.2 Detection Limits and Ranges 3
3.3 Time Domain Data Acquisition of a Random Signal (Correlation) 4
3.4 Data Reduction and Transformation 5
4 Sample Preparation 7
4.1 Concentration 7
4.2 Stability of Samples 8
4.3 Dust in the Sample 8
5 Instrument Performance 8
5.1 Efficiency of Data Acquisition 8
5.2 Result Verification 9
5.3 Standardization 9
6 Instrument Design for Dynamic Light Scattering 9
6.1 Light Source 9
6.2 Optical System 9
6.3 Detector System 9
6.4 The Digital Correlator 10
7 Examples of Dynamic Light Scattering Measurements 11
7.1 Particle Size Distributions 11
8 Comparison with Other Particle Sizing Methods 11
9 New Developments 12
9.1 New Correlator Developments 12
9.2 Solid-state Lasers 12
9.3 Fiber Optics 12
9.4 New Applications 12
10 Conclusion 13
10.1 Disadvantages 13
10.2 Advantages 13
List of Symbols 13
Abbreviations and Acronyms 14
Related Articles 14
References 14

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Photon correlation spectroscopy (PCS) has become a powerful light-scattering technique for studying the properties of suspensions and solutions of colloids, macromolecules and polymers, that is absolute, non-invasive and non-destructive. This text explains the principles of the technique and describes the required instrumentation. It also discusses new developments in instrumentation, which may substantially impact on future applications of this technique. Proper sample preparation is of major importance for the accuracy and precision of the mean size and size distribution results and thus one section is dedicated to sample handling.

1 INTRODUCTION

The acronym PCS is only one of several different names that have been used historically for this technique. The first name given to the technique was quasi-elastic light scattering (QELS) because, when photons are scattered by mobile particles, the process is quasi-elastic. QELS measurements yield information on the *dynamics* of the scatterer, which gave rise to the acronym DLS (dynamic light scattering). Another name, IFS (intensity fluctuation spectroscopy) was used by several authors in the past. Throughout this text the acronym DLS will be employed because its use has become more prevalent and presents a logical juxtaposition to SLS (static light scattering). Application of the DLS technique to particle sizing and commercial availability occurred only about seven years after the first size measurements[1] were made in 1972 merely to check the alignment of a multi-angle research light scattering system. Slowly, through the 1970s, DLS gained wide acceptance among experts in light scattering. This text is not meant to be exhaustive and emphasizes the application of DLS for particle sizing which relates to the *translational* motion (diffusion) of particles in liquids. However, DLS, by its nature, is not limited to this mode: any *time variable* parameter, such as the rotation or vibration of scattering centers, or fluctuations in the refractive index, entropy, thermal diffusivity etc. can be measured. The reader is referred to the extensive literature that exists on these subjects.

2 HISTORY

All materials scatter and absorb light and since the first light-scattering experiments described by Tyndall,[2] SLS experienced major developments in the first half of the 19th century and is well described in the publications by Kerker.[3] DLS theory is built upon the earlier foundation of classical light-scattering theory, which is usually dated

by Rayleigh's papers[4] in 1871 on the scattering from a single particle small compared to the wavelength of light. Scattering from larger particles was added later and is known as Mie scattering, which gives the complete solution for spherical particles of any size. As early as 1908 the temporal fluctuations about the average scattered light intensity were identified with the motion of the particles and their diffusion coefficients. Einstein[5] had already published the relationship between diffusion and size, so in principle the way was open for the birth of a new particle sizing technique. The fundamentals have been written about very often and these references should be consulted for a more complete treatment.[6–26] The history of experimental DLS begins with the advent of the laser. In the early 1960s Pecora[27] pioneered a new kind of light scattering: *time dependent* light scattering. He showed that, by analyzing the frequency distribution of the intensity fluctuations of light scattered from suspensions of macromolecules, information can be obtained about the translational and rotational diffusion coefficients of the macromolecules. Initially DLS was used to measure the diffusion coefficient of macromolecules, from which a hydrodynamic size was calculated. A few industrial users tried this technique for submicron particle sizing, mostly to replace transmission electron microscope (TEM) measurements in quality control (QC) applications. During the second half of the 1970s improvement of digital correlators and the introduction of several algorithms for analyzing decay time distributions were seen. Measurements that had taken months to set up and hours to make, were now reduced to minutes. Bertero et al.[28,29] published a fundamental paper in 1984 that derived the limiting conditions for the resolution of the "noisy" sum of an unknown number of exponentials based on information theory. The advent of highly efficient nonlinearly spaced correlators ushered in a steep ascent of the utility of DLS particle sizing in the mid-1980s, and the emerging fast computers speeded up data handling. Unfortunately, the pretty pictures and rapidly achieved size distribution results often mask the real underlying limitations that are imposed by the physics of the technique.

3 PRINCIPLES OF DYNAMIC LIGHT SCATTERING

3.1 Origin of Intensity Fluctuations

Colloidal sized particles in a liquid undergo random ("Brownian") motion owing to multiple collisions with the thermally driven molecules of the liquid. The scattered light intensity from these diffusing particles will fluctuate in time, thus carrying information about the diffusion

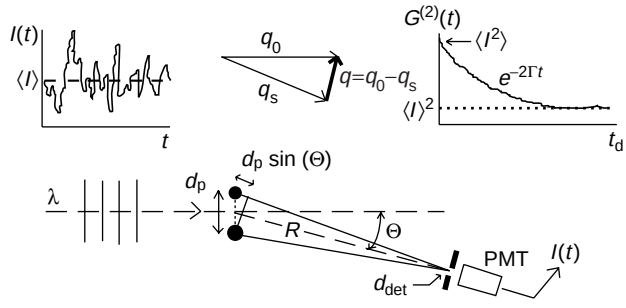

Figure 1 The basics of DLS.

coefficient of the particles. Cummins, Knable and Yeh[30] demonstrated a method to determine the spectrum of light scattered from a dilute suspension of polystyrene spheres by employing novel optical techniques. Following is a description of the basic physics of the DLS technique and of the instrumentation employed in these measurements. The approach used is not intended to be exhaustive; for more details, the literature presented at the end of this article should be consulted. The NATO proceedings edited by Cummins and Pike[7,8] are particularly valuable. The volume edited by Chu[9] conveniently collects many of the major papers together. The article by Ford[25] provides an excellent introduction to many of the practical aspects of the technique and the book by Wyn Brown[26] contains many valuable contributions to various data transformation methods. Equation (1) describes the time averaged scattered intensity $\langle I(q)\rangle$ in terms of particle parameters (also see Figure 1)

$$\langle I(q)\rangle = KNM^2P(\Theta)B(c) \tag{1}$$

$$q = \frac{4\pi n}{\lambda_0}\sin\left(\frac{\Theta}{2}\right. \tag{2}$$

where q is the magnitude of the scattering vector, Θ is the scattering angle, λ_0 is the wavelength of the laser in vacuo and n is the refractive index of the suspending liquid; K is an optical constant, N is the number of scattering particles, M is the mass of a particle, $P(\Theta)$ is the particle form factor, and $B(c)$ is the concentration factor. The terms M^2 and $P(\Theta)$ are of particular importance for the determination of size distributions. The central concept in DLS is as follows: the diffusion of the scatterers (Brownian motion) causes the phases of the fields scattered from each of them to change with time so that the *total* scattered intensity will fluctuate with time owing to constructive and destructive interference. The source of the intensity fluctuations can be understood with reference to Figure 1. A laser beam of wavelength λ is incident on two identical particles. Light scattered by the particles is received by a photo multiplier tube (PMT) and it can be readily seen that the pathlength

difference between the scattered waves is $d_p \sin(\Theta)$. When this pathlength difference is equal to an integer multiple of wavelengths $d_p \sin(\Theta) = m\lambda$, the wavelets will arrive at the detector *in phase* (constructive interference), i.e. the total intensity will be twice that from a single particle. When the pathlength difference is a half integer multiple of wavelengths $d_p \sin(\Theta) = \lambda/2, 3\lambda/2, 5\lambda/2, \ldots$, then the two wavelets will be exactly *out of phase* (destructive interference), i.e. the resulting intensity will be zero. The relative positions and orientations of the particles undergo Gaussian random changes in time (Brownian motion). The result is a total intensity which fluctuates in time from zero to double the single particle scattered intensity. In a real experiment there are perhaps $\sim 10^8$ particles in the scattering volume, and the total intensity is the result from the interference between the scattered fields from all of these. As a consequence, the intensity $I(t, q)$, as seen by the detector, will be a randomly fluctuating signal as shown. Its well-defined mean value, $\langle I(q) \rangle$, is the quantity measured in a SLS experiment. The intensity fluctuations are not readily observed because they occur on a rapid timescale (10^{-6}–10^{-3} s) and because they take place only at a point (Ford[31]). The first condition necessitates a high-speed detection and recording system. The second condition is of fundamental importance in DLS measurements because it determines the limiting properties of the optical system. When the light scattered from a suspension of colloidal particles undergoing Brownian motion is projected onto a screen or wall, a granular pattern (''speckles'', bright and dark patches) is observed. The integral intensity appears constant. However, the intensity of each individual speckle changes randomly in time from very dark to very bright. The intensity of adjacent spots also fluctuates, but independently. Thus the largest spread in intensity (the fluctuations) will be obtained by looking at exactly only one spot. As the number of spots observed increases, the magnitude of the fractional intensity *fluctuations* will decrease, because the spots' intensities fluctuate independently and the high probability of one spot's intensity increasing and another's decreasing yield small integral fluctuations, even though the *total time averaged* intensity increases by observing several speckles simultaneously. The intensity fluctuations will be in phase, or coherent, if the light is detected over one ''coherence area'' (one speckle) the size of the coherence area. For the configuration shown in Figure 1, where a scattering volume of diameter d_p is a distance R away from a detector, the size of the coherence area A_{coh}[15] is given by Equation (3):

$$A_{coh} = \frac{4\lambda^2 R^2}{\pi d_p^2} \tag{3}$$

The area of the detector is given by Equation (4)

$$A_{det} = \frac{\pi d_{det}^2}{4} \tag{4}$$

and if this is set equal to A_{coh}, the size of the detector that is sufficient to see only one coherence area may be calculated. For a typical arrangement with $\lambda = 475$ nm (HeNe laser wavelength in water), $R = 15$ cm and d_p (the diameter of the scattering region) $= 0.2$ mm, the diameter of the detector d_{det} will be ~ 0.9 mm for one coherence area, i.e. a detector aperture of ~ 0.9 mm is required to see one spot and achieve the maximum signal-to-noise ratio (S/N). Since there is no advantage for most experimental conditions in collecting more than one coherence area, it is important to have the maximum intensity scattered into a small area. This is achieved by tightly focusing the incident beam to decrease d_p and increase the size of the coherence area (Equation 3).

3.2 Detection Limits and Ranges

Several experimental conditions must be met for the applicability of the relevant physics for SLS and DLS techniques in general, and particle sizing in particular (also see Table 1 in section 4):

3.2.1 The Optical Mixing Mode

Any measurement must use exclusively one of the following three optical mixing modes:

Table 1 Summary of measurement problems

Condition	Cause	Effect
Multiple scattering present	Concentration too high	Measured sizes *too low*
Stray light present	Light reflections	Measured sizes *too large*
Dust present	Sample preparation	Incorrect size distribution
Duration too short	Operator impatience	Unrepeatable results
Ergodicity not assured	Nature of the sample	Unrepeatable results
Size range exceeded	Operator ignorance	Unrepeatable results
Sample not stationary	Aggregation, sedimentation	Unrepeatable, averaged results
Sample preparation	Impatience or ignorance	Incorrect or unrepeatable results

- Self beating: the scattered light mixes on the photo multiplier cathode with itself.
- Homodyning: the scattered light mixes with a strong signal from the same laser. This signal, a *local oscillator*, must have the *same* frequency, must be at least an order of magnitude higher than the scattered light, and must be in phase with it. If the auto correlation function (ACF) contains a mix of self beating and homodyning, the calculated sizes are invalid. Unfortunately there is no easy way to recognize this problem when it is present. Only one commercially available instrument employs the *homodyning* mode for particle sizing.[32]
- Heterodyning: the scattered light mixes with a strong signal of a *different* frequency (a local oscillator), which is generated either by frequency shifting the same laser or by using a laser beam with a different frequency.

It should be noted that in the literature the terms self-beating and homodyning are frequently used interchangeably and homodyning is confused with *heterodyning*. Here these terms are used according to their true definitions: DLS-based particle sizing applications operate almost exclusively in the self-beating mode, *homodyning* is used rarely and *heterodyning* is not employed at all. However, DLS Doppler (velocimetry) experiments are conducted almost exclusively in the heterodyne mode.

3.2.2 Concentration

The concentration must be sufficiently low so that in Equation (1) $B(c) = 1$ and *multiple scattering* (see below) is avoided, but also must be sufficiently high to prevent signal distortions due to number fluctuations. The number of the particles in the scattering region must always be sufficiently high to maintain a constant $\langle I(q)\rangle$ (Equation 1).

3.2.3 Multiple Scattering

Unless the concentration is vanishingly small, any scattered photon has a non-zero probability to be scattered again while traveling through the sample. The ACF of such multiple scattered light decays faster, yielding an apparent smaller particle size. Thus the sample concentration must be kept sufficiently low and/or the optical path must be sufficiently short to avoid multiple scattering.

3.2.4 Ergodicity

All particles must have an equal opportunity to be measured. This condition is sometimes difficult to meet if the diffusion coefficient of the scatterers is very low, see Pusey.[19,33]

3.2.5 Stationary Process

Scattering conditions must not change during the measurement.

3.2.6 Stray Light

Stray light must be avoided to prevent inadvertent homodyning (see section 3.2.1) and commonly determines the low limit of the angular measuring range.

3.2.7 Sample Preparation

The term M^2 in Equation (1) predicts the scattered intensity to be $\propto d^6$. Consequently a few large particles (commonly called ''dust''), will overwhelm the desired signal (see section 4).

3.2.8 Size Range

The DLS technique is commonly employed in the range of 0.002 to 2 microns. The low limit is usually determined by the available laser power. The high limit results from sedimentation and number fluctuations due to the low number of large particles that fit into the small scattering volume.

3.2.9 Measurement Duration

The typical range is from 1 to 10 minutes, depending on $\langle I(q)\rangle$.

3.3 Time Domain Data Acquisition of a Random Signal (Correlation)

To see how the signal $\langle I(q)\rangle$ changes with time it is convenient to compute its ACF. The *intensity* ACF is described by Equation (5):

$$G^{(2)}(t_d) = \frac{1}{N}\sum_{i=1}^{N} I(t_i)I(t_i - t_d) = \langle I(t)I(t - t_d)\rangle \quad (5)$$

The angle brackets indicate a time average, N is the number of samples and t_d is the time delay between the samples. Experimentally, $G^{(2)}(t_d)$ is determined by recording $I(t)$ at time intervals much shorter than the timescale of typical fluctuations and accumulating the products of the intensities as a function of (t_d). For a dilute suspension of monodisperse spheres (e.g. latex particles), $G^{(2)}(t_d)$ can be written as Equation (6):

$$G^{(2)}(t_d) = B + f \times e^{-2\Gamma t_d} \quad (6)$$

where $B = \langle I \rangle^2$ (Figure 1) is the baseline, f is an instrumental constant and 2Γ is the reciprocal of the decay time τ. $G^{(2)}(t_d)$ is normalized by dividing Equation (6) into the base B, yielding in Equation (7) the normalized intensity ACF $g^{(2)}(t_d)$. For Gaussian statistics Glauber[34] derived Equation (7) via the Siegert relation[35]

$$g^{(2)}(t_d) = 1 + a[g^{(1)}(t_d)]^2 \tag{7}$$

relating the electric field ACF $g^{(1)}(t_d)$ to the measured intensity ACF $g^{(2)}(t_d)$, and a is an instrumental constant that is now between 0 and 1. It should be noted that any photoelectric detector is sensitive to the intensity of light, which is the square of the field amplitude. The normalized electric field ACF $g^{(1)}(t_d)$ may be written as Equations (8) and (9):

$$g^{(1)}(t_d) = G_0 e^{-\frac{t_d}{\tau}} \tag{8}$$

$$\Gamma = \tfrac{1}{\tau} = D_T q^2 \tag{9}$$

where $\tau = 1/\Gamma$ is the decay constant and D_T is the translational diffusion coefficient of the spheres, and G_0 is the intercept of the exponential function, $g^{(1)}(t_d = 0)$. In the case of measurements on dilute monodisperse spheres, the intensity ACF will be a simple exponential function as shown in Figure 2. The significant parameter in Equation (8) is τ, which is invariant to the scattered intensity $\langle I(q) \rangle$. Thus, for DLS-based measurements, the absolute intensity is of secondary importance, as long as it is sufficiently large to complete the experiment within the lifetime of the operator or the stability of the sample, whichever comes first. Measurements in the frequency domain yield the same information; in Equation (10) the power spectral density $S(w)$ is the Fourier transform of $g^{(1)}(t_d)$ and has the shape of a Lorentzian.

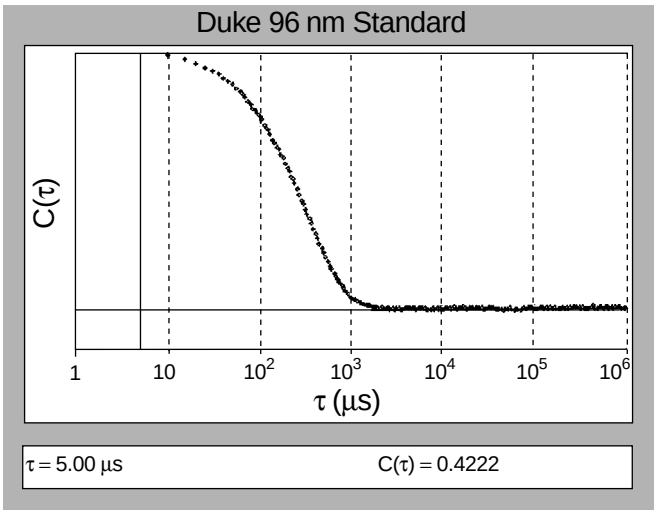

Figure 2 Semi-log plot of a single mode ACF.

$$S(w) = \langle I(t) \rangle \frac{(2\Gamma)^2}{w^2 + (2\Gamma)^2} \tag{10}$$

The relationship between Γ and τ is defined in Equation (9), from which the particle diffusion coefficient D_T can be calculated. Finally in Equation (11), for spheres, D_T is related directly to their hydrodynamic radius r_h by the Stokes–Einstein relationship

$$D_T = \frac{k_B T}{6\pi\eta(T) r_h} = \frac{k_B T}{3\pi\eta(T) d_h} \tag{11}$$

Here k_B is the Boltzmann constant, T is the absolute temperature (Kelvin) and $\eta(T)$ is the viscosity of the suspending liquid (the viscosity is strongly temperature-dependent, requiring a constant temperature throughout a measurement); r_h and d_h are the hydrodynamic radius and diameter respectively; these values are almost always larger than the dry particle diameter owing to the ''double layer'' that forms around charged particle surfaces. The counter-ions around the particle form a diffuse region which moves with the particle. An important parameter for colloidal stability, the ''zeta potential'', is defined at approximately the location of the shear plane (the outer ''edge'' of the diffuse double layer).

3.4 Data Reduction and Transformation

The simplest case involves a single exponential decay for a monodisperse sample as shown in Figure 2 (owing to the logarithmic time axis, the ACF is 'S' shaped). After dividing the correlation coefficients into the baseline and subtracting 1, the remaining part, $G_0 e^{-\Gamma t}$, is fit using a least-squares technique. The value of G_0 typically varies from 0.1 to 0.9 depending on the optical configuration. G_0 is a maximum when the coherence criterion is met. The particle diameter is calculated from D_T in Equation (11). Typical results on latex spheres show the technique is as reliable as an electron microscope on monodisperse samples or latex spheres. The size calculated from Equation (11) is a hydrodynamic diameter.

3.4.1 Monodisperse Distribution

The results for monodisperse samples can be compared directly with other techniques since there is only one ''average'' diameter. For nonspherical particles either the results can be interpreted in terms of an equivalent sphere diameter equal to the Stokes–Einstein hydrodynamic diameter, or the diffusion coefficient can be interpreted in terms of ellipsoids of revolution.[36] This requires either independent knowledge of one dimension or their ratio,

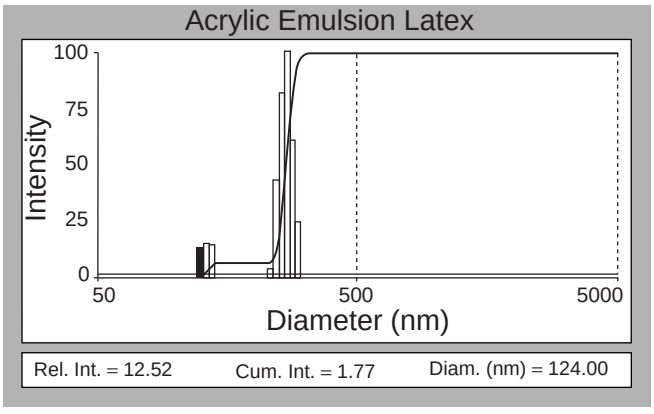

Figure 3 Bimodal distribution of an emulsion.

or angular measurements from which, in favorable cases, both dimensions can be calculated.[37]

3.4.2 Polydisperse Distribution

The interpretation of data from polydisperse samples (Figure 3) is considerably more difficult. Since the technique does not count single particles, size distribution information must be obtained from the deconvolution of the sum over all the single exponentials contributing to the measured ACF. The general deconvolution of a sum of single exponentials is difficult. The problem may be summarized by Equations (12) and (13) which define the normalized field ACF $g^{(1)}(t_d)$ as the sum of m particle sizes, each of which contributes $G(\Gamma_i)$ for the i-th size class.

$$\begin{aligned} g^{(1)}(t_d) &= G(\Gamma_1)\, e^{(-\Gamma_1 t_d)} + G(\Gamma_2)\, e^{(-\Gamma_2 t_d)} \\ &\quad + G(\Gamma_3)\, e^{(-\Gamma_3 t_d)} + \cdots \\ &= \sum_{i=1}^{m} G(\Gamma_i)\, e^{(-\Gamma_i t_d)} \end{aligned} \tag{12}$$

Here Γ_i is the value as defined in Equation (9). For the limit of $\Gamma_{i+1} - \Gamma_i = \Delta\Gamma = 0$, $g^{(1)}(t_d)$ is defined in Equation (13) by integrating over the entire range of particle sizes

$$g^{(1)}(t_d) = \int_{\Gamma_{min}}^{\Gamma_{max}} G(\Gamma)\, e^{-\Gamma t_d}\, d\Gamma \tag{13}$$

where $G(\Gamma)$ represents the distribution of line widths due to the distribution of particle sizes. Γ_{min} and Γ_{max} define the range of line width; in principle, $\Gamma_{min} = 0$ and $\Gamma_{max} = \infty$. It is easy to calculate $g^{(1)}(t_d)$ from a known size distribution $G(\Gamma)$, but exceedingly difficult to perform the inverse transformation to extract $G(\Gamma)$ from $g^{(1)}(t_d)$. Equation (13) belongs to a class of linear transformations (Laplace) that are known to be ill-conditioned. In the presence of any noise, many different solutions for the functional form of $G(\Gamma)$ are possible. A repeat measurement on the same sample will yield a slightly different $g^{(1)}(t_d)$ with a slightly different noise contribution, resulting in a different $G(\Gamma)$ and hence in a different size distribution $G(d)$. This behavior can render the results of a DLS instrument apparently "not repeatable". For these fundamental reasons of physics DLS will not become a high-resolution technique for particle sizing. Numerical solutions of Equation (13) were attempted by Chu and Gulari,[38] McWhirter and Pike,[39] McWhirter[40] and Ostrowsky and Pike[41] have demonstrated that $G(\Gamma)$ can be calculated at exponentially spaced delay times. The method, called *exponential sampling* introduces ϖ_0 as a parameter that limits the resolution of $G(\Gamma)$ for a given noise level in Equation (14).

$$\Gamma_{i+1} = \Gamma_i\, e^{\frac{\pi}{\varpi_0}} \tag{14}$$

Here ϖ_0 is the largest value for a fixed noise content of $g^{(1)}$ and therefore defines the best possible resolution for Γ_i and Γ_{i+1}. Example: if $\varpi_0 = 3.14$, $\Gamma_{i+1}/\Gamma_i = 2.71$; if Γ_{i+1} corresponds to a size of, e.g. 500 nm, the next neighboring peak cannot be closer than 184 nm (500/2.71) or 1050 nm (500*2.71). If the inversion routine asks for a higher resolution than can be supported by the ACF, the Gamma distributions will exhibit more and more artifacts such as additional maxima and negative values for $G(\Gamma)$. The practical limit for the ratio Γ_{i+1}/Γ_i is 2. It is important to note, that this resolution cannot be exceeded by any transformation routine without imposing additional, a priori imposed, constraints. The most commonly applied constraints are non-negativity (only zero or positive $G(\Gamma)$ are allowed), range limitations and force fits to specific distribution models (log-normal, bimodal). Provencher[42,43] has developed an effective inversion software package, CONTIN, which is a generalized Inverse Laplace Transform with constraints and parsimony. CONTIN contains a non-negative least square routine (NNLS), that may be used effectively as a fast executing stand-alone program to obtain rapid size distributions in real time during data acquisition.[44] Several other inversion programs are in use by practitioners, e.g. "Maximum entropy",[45] "REPES"[46] and "single value decomposition".[47,48] H. Ruf[49] investigated the errors that are introduced into the inversion process owing to normalization errors. Aside from the noise content of the ACF, baseline errors contribute a significant amount of ambiguities to any of the transformation routines and repeated, iterative, normalizations of the ACF can improve the stability of the results. The above approaches show that only a few parameters of the distribution can be calculated from typical experimental measurements of the correlation function. Present work on these advanced techniques is focusing on the separation of only two peaks

in a size distribution, the size ratio of which is at least 2 : 1. Thus DLS measurements must be considered as relatively insensitive to the particular size distribution and only a few moments of the distribution can be obtained, except for monodisperse samples, for which the technique works extremely well.

3.4.3 Cumulants

The most widely used and simplest data analysis technique to apply is the *method of cumulants*.[50,51] The method proceeds by expanding Equation (13) about an average line width $\langle\Gamma\rangle$, to give Equation (15):

$$\ln[a_0 g^{(1)}(t_d)] = a_0 - \langle\Gamma\rangle t_d + \frac{\mu_2}{2}t_d^2 - \frac{\mu_3}{6}t_d^3 + -\cdots \quad (15)$$

where a_0 is a constant and $\langle\Gamma\rangle$, μ_2, etc. are moments of the line width distribution $G(\Gamma)$. In practice only the first two moments are obtained with certainty, and care must be taken to limit the range of t_d such that higher order terms are negligible. Before the advent of non-linearly spaced correlator channels this used to be difficult for broad distributions. This technique has the advantage that no assumption about the form of the distribution is necessary, and, under the correct experimental conditions described earlier, the moments are well defined and useful parameters. It has been shown by Brown, Pusey and Dietz[50] that (Equation 16):

$$\langle\Gamma\rangle = \langle D_T\rangle q^2 \quad (16)$$

where $\langle D_T\rangle$ is given by Equation (17),

$$\langle D_T\rangle = \frac{\sum N_i M_i^2 P_i(\Theta) D_i}{\sum N_i M_i^2 P_i(\Theta)} \quad (17)$$

NM^2P is the time-averaged intensity weighting factor (Equation 1) for the contribution of D in the i-th size class. The sum is over all the particles contributing to the scattering, and it has been assumed that the measurements have been made or will be extrapolated zero concentration. If not, the interparticle interference term $B(c)$ must be included in Equation (17). It has also been assumed that the optical constant K in Equation (1) is the same for all particles independent of size. K includes the refractive index increment (change of solution refractive index with particle concentration). So the assumption of a constant K means assuming a sample of homogeneous composition independent of size. For Rayleigh scatterers ($d \ll \lambda$) and for all particles where measurements have been extrapolated to zero angle, where $P(\theta) = 1$, and

$$\langle D\rangle = \langle D\rangle_z = \frac{\sum N_i M_i^2 D_i}{\sum N_i M_i^2} \quad (18)$$

Here $\langle D\rangle_z$ is the well-defined z-average diffusion coefficient (z-weighting means the next higher weighting after volume). Thus, for spheres, where $M^2 \propto d^6$, the average particle size obtained is (Equation 19),

$$\frac{1}{d_z} = \frac{\sum N_i d_i^5}{\sum N_i d_i^6} \quad (19)$$

Here d_z is the inverse, z-average diameter and N_i is the number of particles in the i-th size class. The second moment in the cumulant analysis, μ_2 in Equation (20) yields (after extrapolation to zero concentration and zero angle)

$$\mu_2 = \langle D_z^2\rangle - \langle D_z\rangle^2 \quad (20)$$

which is the *variance* of the z-average diffusion coefficient distribution. For spherical particles the first term can be related to the inverse square, z-average diameter, $\langle(1/d^2)_z\rangle$. This technique has the advantage that no assumption about the form of the distribution is necessary, and the moments are well defined and useful parameters. A convenient polydispersity index is defined as Q in Equation (21)

$$Q = \frac{\mu_2}{\langle\Gamma\rangle^2} = \frac{\langle D_z^2\rangle - \langle D_z\rangle^2}{\langle D_z\rangle^2} \quad (21)$$

where μ_2, $\langle\Gamma\rangle$ and $\langle D_z\rangle$ are defined in Equations (20), (16) and (17) respectively. For QC measurements the exact interpretation of data is not usually of primary importance. In this case, measurement at one angle, usually 90°, is sufficient to establish an apparent average size and apparent measure of the width of the size distribution. Monitoring changes in these two parameters may be all that is required. In many cases the distribution is known to be log-normal or reasonably close to one; Thomas[52] showed that in this case the two parameters μ_2 and $\langle\Gamma\rangle$ may be mathematically transformed to geometric standard deviation and mass mean diameter, the two parameters that define a log-normal function. Such transformations are not limited to log-normal functions but can be performed for any other two parameter distribution. It is important, however, not to forget that this function may or may not resemble the real particle size distribution, although it definitely will produce a "nice" looking display on the computer screen.

4 SAMPLE PREPARATION

4.1 Concentration

Concentration effects manifest themselves in two fundamentally different ways:

4.1.1 Multiple Scattering

Even if D_T is indeed independent of c, the ''particle size'' will be reported as too small, because the correlation is lost on a shorter timescale for multiple scattered photons. A comparatively new set of experiments has been successful in analyzing the dynamics of multiple scattered photons from concentrated samples (see section 9).

4.1.2 Particle–Particle Interaction

The apparent viscosity depends on particle–particle interaction. In sufficiently high concentrations where particles experience significant interactions, the calculated particle size is different from those obtained under infinite dilution conditions. There is a large amount of physics to be learned from such experiments, as long as the instrument is capable of eliminating multiple scattering effects, so that the true functional form of $D_T(c)$ can be obtained and the theories for particle–particle interaction can be investigated (see section 9).

4.2 Stability of Samples

The scattering environment must not change during the measurement. Sometimes this is difficult or impossible to accomplish owing to the presence of changing chemical and/or physical parameters in the sample (chemical reactions, aggregation, temperature). If those changes are the objects of investigation, the experiment duration must be kept sufficiently small compared to the characteristic times of the reactions. Under most operating conditions, however, an unstable sample is the result of inappropriate sample preparation. The following describes a general approach to solving the sample preparation problem as it relates to particle sizing.

4.2.1 Coagulation

Physics predicts (van de Waals forces) that, allowing for a sufficiently long time, any colloidal dispersion will coagulate, be it seconds, minutes or years. Thus the problem to be solved can be stated very simply: how long is stability required? For particle sizing typically 30 minutes are sufficient, but for a commercial product a long shelf life is imperative. It will exceed the scope of this text to dwell on any details, but fortunately even only superficial attention to some basic surface chemistry will result in meaningful particle size measurements. Under most operating conditions an unstable sample is the product of inappropriate sample preparation. The following describes a general approach to sample preparation as it relates to particle sizing.

4.2.2 Hydrophobic or Hydrophilic Sample Surfaces?

Sprinkling a few particles on the surface of the water will rapidly reveal their properties: hydrophilic particles will rapidly disperse in the water, hydrophobic types will stay on the surface. The former will generally merely require a stabilizing surfactant, whereas the latter often can be converted into a hydrophilic system by creating a paste of the particles and methyl alcohol.

4.2.3 Zeta Potential

The ''zeta potential'' of particles (a measurement of the electrostatic repulsion between particles) is a good, but certainly not the only, predictor of the stability of a suspension. Unless the sample is sterically stabilized, a low zeta potential increases the probability for aggregation within a short time; thus the determination of this parameter is very important. Frequently a simple change of the pH value is all that is required to achieve a stable suspension. Further information about electrophoretic mobility (zeta potential) and the stability of colloidal systems can be found in a book by Hunter.[53]

4.3 Dust in the Sample

The appropriate use of surfactants to avoid aggregation, filtering, centrifugation, ultrasonication, cleaning of sample cells and the use of deionized and filtered water are of utmost importance for any serious size measurement. Dust will not only cause huge intensity spikes and distort the raw data but can also introduce unintentional homodyning to the self-beating mode by actually creating a local oscillator. Nonaqueous (nonpolar) liquids do not attract ''dust'' as much and are easier to clean, but sample preparation is often more complicated. Filtering of all liquids, especially water, and thorough cleaning of all sample cells as well as pipets, syringes, etc. are always required. DLS instruments, by their nature, will collect at least the dark counts from a detector, always resulting in some nonzero ACF. The computer will happily calculate a ''result'', regardless of the quality of the signal, thereby creating insidious and serious problems if any of the above listed conditions is violated. Table 1 shows a compilation of the most common error sources in DLS particle size measurements.

5 INSTRUMENT PERFORMANCE

5.1 Efficiency of Data Acquisition

The ACF consists of many product terms which asymptotically converge to a steady function. The overall

optimization of the statistics begins with the efficient launching of the laser beam, the collection of the scattered photons by bulk or fiber optics and the quantum efficiency (QE) of the detector. All available photon pulses should be utilized by the optics and the electronics in order to achieve the best possible statistics in the least amount of time. Here the design of correlators that do not require any prescaling presents a significant advance in this technology. Modern correlators do not depend on raw computer speed for data collection, but in QC applications the ''throughput'', which includes data reduction, display, storage and possible statistical process control (SPC) depends on the performance of the attached data processor.

5.2 Result Verification

DLS results are sometimes not easy to interpret since there always will be some result displayed. Several conditions can be stipulated for the correctness of the results:

- *repeatability*: this is a necessary, but not a sufficient condition;
- *invariance of the size distribution* to the delay time range, as long as the smallest range covers the set of decay times;
- *invariance of the peak positions* to the chosen range of particle sizes;
- *no peaks at the extremes* of the chosen size range;
- *consistency of the results* with the physical and chemical properties of the particles;
- *consistency* of the results with other sizing techniques. If data transformations are required they must be evaluated very carefully (see section 3).

5.3 Standardization

The International Organization for Standardization (ISO) published a draft paper that proposes to establish ''standard'' methods for the application of the DLS technique to the measurement of particle size distribution of samples with specified properties. Details of the proposal can be found in the publications of the subcommittee.[54] It should be noted that this publication is just a proposal subject to amendments. It presents a framework of design rules for DLS-based particle sizing that should yield results which are invariant to specific instruments or operators around the world.

6 INSTRUMENT DESIGN FOR DYNAMIC LIGHT SCATTERING

A typical DLS system may be broken down into several functional components which will be discussed separately. These are a light source, an optical system, a detector system and a digital correlator.

6.1 Light Source

Practical requirements for a sufficiently intense light source demand a narrow-band, polarized, monochromatic, CW laser. Table 2 summarizes the popular options available.

DPSS have become widely available within the last few years. Their wavelength is typically 532 nm (frequency doubled from powerful 1064 nm diode lasers) and their other properties match well with the requirements of DLS instruments.

6.2 Optical System

A lens focuses the laser beam down into the sample which is enclosed in a temperature-controlled scattering cell surrounded by a refractive index matching liquid. The scattered light is focused onto a PMT at an angle θ by another lens. Systems like this are constructed on a precision turntable with a stepper motor, and typically allow experiments to be conducted over a 10°–160° angular range.

6.3 Detector System

PMTs are almost universally used as detectors in DLS experiments (Table 3). These should have a low dark count and a high gain since most work is done in

Table 2 Summary of lasers commonly used in DLS installations

Type	Wavelength	Power	Size	Cost
HeNe	632.8 nm	5–35 mW	0.40–1.5 m	$500–$7000
Laser diodes	635–780 nm	5–100 mW	0.05–0.15 m	$300–$1200
Ar^+ (air cooled)	488–514.5 nm	~100 mW	1 m	$8000
Ar^+ (water cooled)	488–514.5 nm	~1.7 W	1.5–2 m	$16 000–$40 000
DPSS (Frequ. Doubled)	532 nm	10 mW–4 W	0.2–0.5 m	$3000–$40 000

DPSS, diode pumped solid-state lasers.

Table 3 Summary of PMTs commonly used in DLS

Type	Photo cathode	Dynodes (gain)	Dark count/Max. count	Comment
EMI 9863B	S20	14 (10^7)	40 cps/5 Mc	Red sensitive, QE: 3–4%
EMI 9130	S20	11 (10^6)	40 cps/1 Mc	Red sensitive small tube
Ham R464	Bialkali	12 (6×10^6)	5 cps/2 Mc	Quartz window for UV
Ham R585	Bialkali	12 (6×10^6)	5 cps/2 Mc	Blue-green sensitive only
Ham R649	S20	12 (6×10^6)	200 cps/2 Mc	Red sensitive, QE: 3–4%
Ham HC120	S20	10 (10^6)	100 cps/2 Mc	Red sensitive, QE: 3–4%
SPCM	APD	NA (10^4)	500 cps/1 Mc	QE (635 nm): 35–70%

NA, not applicable

the single photon counting regime. A comparatively new development is the single photon counting mode (SPCM) by EG&G[55] which incorporates an avalanche photo diode (APD), active reset and quenching electronics and a Peltier-type temperature controller in a small package. This author investigated the properties of the APD (supplied then by RCA Canada) in 1985 and found that they were not well suited for DLS applications owing to the lack of active quenching and the lack of an integrated temperature control for the semiconductor junction (unpublished). Brown[56] developed the required circuits and successfully implemented APDs into small fiber-optic-based DLS devices. New SPCMs can replace PMTs in many applications, particularly for wavelengths of more than 680 nm (near infrared).

6.4 The Digital Correlator

The correlator has become the device of choice to generate the raw data in a DLS experiment, although the early experiments mostly employed ''wave analyzers'' or ''spectrum analyzers''. (Today only one commercial instrument[32] still performs DLS particle sizing in the frequency domain by employing a real-time spectrum analyzer.) The basic design of a correlator is really very simple and was originally used at the Royal Radar Establishment, UK[57] (now Defence Evaluation and Research Agency) to recover radar signals that are deeply buried in noise. This instrument was commercialized by (then) Precision Devices, (now) Malvern Instruments,[58] in the early 1970s. The ACF is formed by recording the number of photons arriving in each sample time, maintaining a history of this signal over a large range of sample times (time series), multiplying the instantaneous and the delayed signal for a range of time delays t_d (the ''channels'') and accumulating these products. In the past most designs employed a multibit shift register to maintain a record of photon counts, as shown in Figure 4. In modern nonlinear correlators shift registers have been replaced by fast memories, allowing very flexible channel configurations with dynamic ranges that exceed 10 decades of time. Individual design details may be gathered from the manufacturers' literature.

Figure 5 shows an example of a complete standalone DLS-based, commercially available instrument that incorporates all of the above mentioned optics, electronics and data processing.

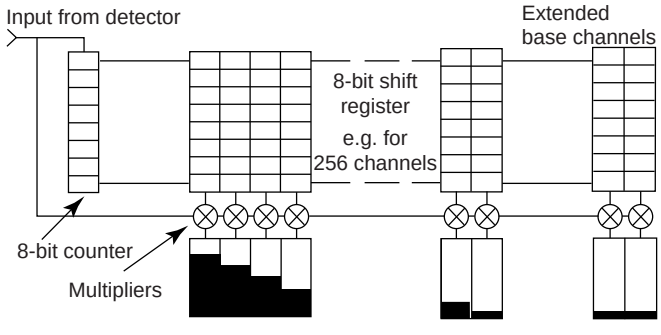

Figure 4 Block diagram of an 8-bit shift register digital correlator.

Figure 5 A typical DLS-based particle sizer. (Courtesy of Brookhaven Corp.)

7 EXAMPLES OF DYNAMIC LIGHT SCATTERING MEASUREMENTS

7.1 Particle Size Distributions

The majority of applications for DLS in particle sizing are the rapid routine measurements of mean sizes in QC work. Manufacturers of latexes, pigments, emulsions, micelles, liposomes, vesicles, sils and silica can track the consistency of the desired particle sizes rapidly and accurately, independent of different operators and different instruments anywhere in the world. Often a bimodal distribution characterizes a sample better and is actually expected from the known chemistry of the sample. In such cases $G(\Gamma)$ will contain two modes that can be separated if their size ratio is at least 1:2 and the relative intensities are similar. Samples that tend to coagulate are easily tracked by DLS. Although the exact size distribution cannot be determined, the first and second moments are very sensitive to any changes in the distribution and give an immediate and accurate response. DLS-based particle sizing is also widely used for biological samples such as bacteria, viruses, proteins, DNA, etc. Many applications are in crystal growth and polymer research. The following few examples illustrate the breadth of DLS-based size measurements.

- Spontaneous vesicle formation in a biological surfactant (ganglioside GM3) was investigated by L. Cantu and M. Corti.[59]
- Studies of BSA and lysozyme, very low molecular weight proteins, were conducted by H. Dhadwal et al.[60]
- Protein (lysozyme) crystallization was investigated by Mikol et al.[61]
- Submicron emulsion systems were measured by Herb et al.[62]
- Liposome production is monitored by DLS in QC and research by the Avestin Corp.[63]
- B. Weiner[44] et al. used a fiber optic backscattering device to observe particle–particle interaction in highly concentrated latex suspension.
- Particle sizing related research on non-ergodic systems have been published by van Megen.[64] DLS was applied to the glass phase of nonaqueous suspensions of sterically stabilized colloid spheres.
- A widely used industrial process, hydro metallurgical solvent extraction, was investigated by Neuman[65] et al., applying the DLS technique to very small particles in the region of 2 nm.
- Caldwell[66] investigated emulsions by employing DLS and sedimentation field flow fractionation (SFFF) methods in a complementary mode. SFFF provides for a high-resolution fractionation of the sample and DLS measures the sizes without the need to know the density of the sample.
- Multiangle particle sizing and its associated data transformation have been investigated by Cummins.[67] If the sample is very clean, the multiangle constraint can be a powerful conditioning for the inversion matrix to stabilize the set of $G(\Gamma)$.
- Auweter et al.[68] published on-line measurements with a one-fiber backscattering device as early as 1985. Their experiments demonstrated the feasibility of such fiber optics in a hostile environment.

8 COMPARISON WITH OTHER PARTICLE SIZING METHODS

Different techniques yield different ''sizes'' for the same particles simply because they employ different physics for the measurement. DLS, for example, will easily detect a few large particles among many small ones. Also, the DLS technique is *absolute* and there are no adjustable parameters available for a ''calibration''. Often the term

Table 4 Comparison of particle sizing techniques

Technique	Resolution	Dyn. range	Calibration	Mass balance	Time/Sec
DLS (PCS)	Low	1:1000	Not required	Intrinsic	2–15
Particle counters	Very high	1:10	Required	Must be verified	8–20
Microscopes	Very high	1:10	Required	Not verifiable	>30
Laser diffraction	Low	1:300	Not required	Intrinsic	2–5
SLS (Mie)	Low	1:10	Required	Intrinsic	2–5
Time of transition	High	1:10	Required	Must be verified	8–30
Sedimentation	Very high	1:30	Not required	Intrinsic	5–30
Centrifugation	Very high	1:30	Not required	Intrinsic	5–30
SFFF	Very high	1:100	Not required	Must be verified	8–20
HDC	Very high	1:10	Required	Must be verified	8–20

HDC, hydrodynamic chromatography.

''calibration'' is confused with the ''validation''. As with any instrument, its proper operation must be verified with known samples, but no adjustments for calibration can be utilized to achieve the ''correct'' result. This is in contrast to microscopes or ''particle counters'' for which a calibration must be performed (Table 4).

Dynamic ranges are valid for one measurement without changing optics, capillaries, etc. All the techniques cover at least part of the DLS range of particle sizes: ca. 0.001–3 μm. Measurement times are typical and depend on the breadth of the distribution, absolute size and physical properties of the particles (e.g. density), desired precision, etc. The widely used Fraunhofer diffraction instruments are advertised as covering a size range of as low as 20 nm to several millimeters. These devices actually combine two distinct instruments: one that is based on diffraction theory (low-angle scattering) with d^2 weighting and one based on Mie scattering with d^6 weighting. This explains the large dynamic range that is advertised for these instruments but because of the different weighting and low resolution a true merging of the mass distribution is impossible, regardless of the claims in the manufacturers' literature. Thus comparisons between DLS and ''Fraunhofer'' sizing results ought to be performed with great caution by carefully evaluating the experimental conditions and examinating the mathematical transformations that are required.

9 NEW DEVELOPMENTS

9.1 New Correlator Developments

In the last several years new correlators have become available, which are remarkably compact, have a large dynamic range in delay time and can accommodate many more counts in a given sample time. These developments are a direct result of the availability of more versatile, higher speed and higher density integrated circuits, and novel approaches to the implementation of the autocorrelation operation. Based on designs by K. Schätzel[22,23] the ALV company[69] is developing a correlator with 16 inputs and a minimum sample time of 800 nsec (50 nsec for a single input). Brookhaven[70] is developing a new correlator that incorporates sophisticated multiple digital signal processing (DSP) in real time. Flexible Instruments[71] offers a range of application-specific integrated circuits based correlator boards. Correlators are really nothing more than specialized DSP instruments, regardless of their specific designs. As such the design of these devices benefits from the relentless improvement in the semiconductor industry: higher speed, more integration, lower power, physically smaller and higher gate densities. These new correlators will extend the application of DLS and improve existing applications.

9.2 Solid-state Lasers

These diodes have become reliable and stable enough to be used in routine DLS based measurements. Higher power DPSS lasers are beginning to replace Ar^+ in many installations. Near infrared lasers (>780 nm) at 100 mW and more can be used in combination with APDs[55] and are particularly well suited for efficient coupling into single-mode fiber optics (large wavelength).

9.3 Fiber Optics

Bulk optics are being replaced more and more with fiber-optic devices. Ricka[72] has shown that single-mode fibers have the unique property of acting as spatial filters yielding high ACF intercepts near the theoretical maximum without sacrificing high intensities that are received by the detector. Ongoing developments will improve the efficiency of launching laser beams into single-mode fibers (currently about 40%) and make polarization-preserving filters more widely available.

9.4 New Applications

- Diffusing wave spectroscopy (DWS): in these new experiments by D. Pine and D. Weitz[73] a ''length parameter'' is calculated from the initial fast decay of the ACF from multiple scattered light, extending DLS applications to samples with higher particle concentrations.
- W. Meyer et al.[74,75] at the National Aeronautics and Space Administration (NASA) designed a promising set-up that employs two detectors and two single mode fibers that are located a tiny distance apart such that the scattered light originates most likely from one speckle. The cross correlation function (CCF) of the signals from the two detectors yields the proper D_T and hence particle size.
- Based on original work by Dhadwal et al.,[76] the commercially available FOQELS instrument from Brookhaven Instruments Corp.[70] utilizes two single-mode fibers at a 155° backscattering angle.
- ALV[69] is currently developing a dual-fiber instrument for backscattering at a 173° angle.

Continuing development of fiber-optic technology will make it possible to construct compact, multiangle light-scattering spectrometers for DLS (and SLS) with no moving parts.

10 CONCLUSION

As any other particle sizing technique, DLS has advantages and disadvantages and it is particularly important to use it strictly within the framework of physical laws, if meaningful results are to be obtained. The following is a short compilation of the advantages and disadvantages of DLS-based particle sizing.

10.1 Disadvantages

- It does not produce a high-resolution histogram of the size distribution.
- Like other nonimaging techniques an equivalent sphere diameter is usually, although not always, assumed. Shape information is not easily obtained.
- When proper measurements are made, the parameters which are most often obtained are the inverse z-average moments of the size distribution, not the usually reported parameters of a size distribution.
- Dust can make measurement and interpretation difficult.

10.2 Advantages

- Measurements are fast, from seconds to minutes.
- The technique is absolute, from first principles. Calibration with a known size distribution is not necessary to get answers.
- Very small quantities of sample can be measured.
- Any suitable suspending liquid can be used provided it is nonabsorbing, relatively clear and not too viscous.
- The technique is applicable from about 0.001 to several microns.
- Instrumentation is commercially available for both research and QC measurements with automation including data analysis.
- Although the interpretation of particle size is least ambiguous with a narrow distribution, an effective diameter and polydispersity index are measurable even with broad distributions.

In order to provide an introduction to the DLS technique, the physical basis of the technique and the associated instrumentation has been discussed. Several new developments in the instrumentation have been described which have the potential for fundamentally changing the way DLS measurements are made and lead to many new applications of the technique. The commercial importance of DLS as a submicrometer particle sizing tool ensures that technological developments will occur rapidly in this field.[32,55,58,63,69,70,71,77–79]

LIST OF SYMBOLS

a	Instrument constant
A_{coh}	Coherence area
A_{det}	Detector area
A/D	Amplifier/Discriminator
Ar^+	Argon Ion laser
B	Baseline of the intensity ACF $G^{(2)}(t_d)$ before normalization
$B(c)$	Concentration factor
c	Particle concentration
cps	Photon counts per second
$g^{(1)}(t_d)$	Normalized correlation coefficient at delay t_d (Field ACF)
d	Particle diameter
d_{det}	Diameter of the detector area
$\langle d_n \rangle$	Mean diameter by number
d_p	Distance between particles
d_{sr}	Diameter of scattering area
d_Z	z-average diameter (the next higher weighting after mass: intensity)
D_T	Translational diffusion coefficient
$g^{(1)}(t_d)$	First-order correlation coefficient at delay t_d
$G(\Gamma)$	Line width distribution function
HeNe	Helium Neon laser
$\langle I(q) \rangle$	Time averaged intensity at wave vector q
k_B	Boltzman constant
K	Optical constant
M	Particle mass
n	Refractive index
N	Number of particles
$P(\Theta)$	Particle form factor
q	Wave vector
Q	Polydispersity index
R	Distance between scattering volume and detector
t	Time
t_d	Delay time between multiplication terms for $C(t_d)$
T	Absolute temperature (Kelvin)
Γ	Line width of scattered light (radians/sec)
$\langle \Gamma \rangle$	Average line width
$\eta(T)$	Viscosity of the medium in which particles are suspended (Poise)
Θ	Scattering angle
λ	Wavelength of light in the suspension medium (λ_0/n)
λ_0	Wavelength of light in vacuo
μ_i	Moments of a distribution ($i = 1, 2, \ldots$)
τ	Decay constant for exponential forms of $C(t_d)$; $C(\tau) = C(0)/e$

ABBREVIATIONS AND ACRONYMS

ACF	Auto Correlation Function
APD	Avalanche Photo Diode
CCF	Cross Correlation Function
DLS	Dynamic Light Scattering
DPSS	Diode Pumped Solid-state Lasers
DSP	Digital Signal Processing
DWS	Diffusing Wave Spectroscopy
HDC	Hydrodynamic Chromatography
IFS	Intensity Fluctuation Spectroscopy
ISO	International Organization for Standardization
NASA	National Aeronautics and Space Administration
NNLS	Non-negative Least Square
PCS	Photon Correlation Spectroscopy
PMT	Photo Multiplier Tube
QC	Quality Control
QE	Quantum Efficiency
QELS	Quasi-elastic Light Scattering
SFFF	Sedimentation Field Flow Fractionation
SLS	Static Light Scattering
S/N	Signal-to-noise Ratio
SPC	Statistical Process Control
SPCM	Single Photon Counting Mode
TEM	Transmission Electron Microscope

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Sedimentation in Particle Size Analysis • Ultrasonic Measurements in Particle Size Analysis • Velocimetry in Particle Size Analysis

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Field Flow Fractionation in Analysis of Polymers and Rubbers • Neutron Scattering in Analysis of Polymers and Rubbers

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods

REFERENCES

1. S.P. Lee, W. Tscharnuter, B. Chu, 'Calibration of an Optical Self-beating Spectrometer by Polystyrene Latex Spheres, and Confirmation of the Stokes–Einstein Formula', *J. Polym. Sci.*, **10**, 2453 (1972).
2. J. Tyndall, 'On the Blue Color of the Sky, the Polarization of the Skylight and on the Polarization of Light by Cloudy Matter Generally', *Phil. Mag.*, **37**, 384–394 (1869).
3. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, Academic Press, New York, 1969.
4. Lord Rayleigh, 'On the Electromagnetic Theory of Light', *Philos. Mag.*, **12**, 81–101 (1881).
5. A. Einstein, 'Theorie der Opaleszenz von homogenen Flüssigkeitsgemischen in der Nähe des kritischen Zustandes', *Annalen der Physik*, **33**, 1275 (1910).
6. B.J. Berne, R. Pecora, *Dynamic Light Scattering*, Wiley-Interscience, New York, 1976.
7. H.Z. Cummins, E.R. Pike (eds.), *Photon Correlation and Light Beating Spectroscopy*, Plenum Press, New York, 1974.
8. H.Z. Cummins, E.R. Pike (eds.), *Photon Correlation Spectroscopy and Velocimetry*, Plenum Press, New York, 1977.
9. B. Chu, *Laser Light Scattering*, Academic Press, New York, 1974.
10. S.H. Chen, B. Chu, R. Nossal (eds.), *Scattering Techniques Applied to Supramolecular Systems*, Plenum Press, New York, 1981.
11. R. Pecora (ed.), *Dynamic Light Scattering – Applications of Photon Correlation Spectroscopy*, Plenum Press, New York, 1985.
12. B.E. Dahneke (ed.), *Measurement of Suspended Particles by Quasi-elastic Light Scattering*, Wiley-Interscience, New York, 1983.
13. K.S. Schmitz, *An Introduction to Dynamic Light Scattering by Macromolecules*, Academic Press, New York, 1990.
14. G.D.J. Phillies, 'Quasielastic Light Scattering', *Anal. Chem.*, **62**(20), 1049A–1057A (1990).
15. B. Chu, 'Advances in Methods of Light Scattering Spectroscopy,', *Pure and Appl. Chem.*, **49**, 941–962 (1977).
16. V.A. Bloomfield, 'Quasielastic Light Scattering Applications in Biochemistry and Biology', *Ann. Rev. Biophys. Bioeng.*, **10**, 421–450 (1981).
17. F.D. Carlson, 'The Application of Intensity Fluctuation Spectroscopy to Molecular Biology', *Ann. Rev. Biophys. Bioeng.*, **4**, 243–264 (1975).
18. W.H. Flygare, 'Light Scattering in Pure Liquids and Solutions', *Chem. Soc. Rev.*, **6**, 109–137 (1977).
19. P.N. Pusey, J.M. Vaughan, 'Light Scattering and Intensity Fluctuation Spectroscopy', in *Dielectric and Related Processes*, ed. M. Davies, Chemical Society, London, Vol. 2, 1975.

20. J.M. Schurr, 'Dynamic Light Scattering of Biopolymers and Biocolloids', *CRC Crit. Rev. Biochem.*, **4**, 371–431 (1977).
21. I.W. Shepherd, 'Inelastic Laser Light Scattering from Synthetic and Biological Polymers', *Rep. Prog. Phys.*, **38**, 565–630 (1975).
22. K. Schätzel, 'Correlation Techniques in Dynamic Light Scattering', *Appl. Phys. B.*, **42**, 193–213 (1987).
23. K. Schätzel, M. Drewel, Sven Stimac, 'Photon correlation measurements at large lag times: improving statistical accuracy', *J. Modern Optics*, **35**(4), 711–718 (1988).
24. B. Chu (ed.), 'Selected Papers on Light Scattering by Macromolecular, Supramolecular and Fluid Systems', SPIE Optical Engineering Press, Bellingham, WA, 1990.
25. N.C. Ford, 'Theory and Practice of Photon Correlation Spectroscopy', in *Measurement of Suspended Particles by Quasi-elastic Light Scattering*, ed. B.E. Dahneke, Wiley-Interscience, New York, 1983.
26. W. Brown, *Dynamic Light Scattering*, Clarendon Press, Oxford, 210–211, 1993.
27. R. Pecora, 'Doppler Shifts in Light Scattering from Pure Liquids and Polymer Solutions', *J. Chem. Phys.*, **40**, 1604–1614 (1964).
28. M. Bertero, P. Boccacci, E.R. Pike, 'On the Recovery and Resolution of Exponential Relaxation Rates from Experimental Data: A Singular-value Analysis of the Laplace Transform Inversion in the Presence of Noise', *Proc. Royal Soc. London*, **A383**, 15–29 (1982).
29. M. Bertero, P. Boccacci, E.R. Pike, 'On the Recovery and Resolution of Exponential Relaxation Rates from Experimental Data. II. The Optimum Choice of the Sampling Points', *Proc. Royal Soc. London*, **A393**, 51–65 (1984).
30. H.Z. Cummins, N. Knable, Y. Yeh, 'Observation of Diffusion Broadening of Raleigh Scattered Light', *Phys. Rev. Lett.*, **12**, 150–153 (1964).
31. N.C. Ford, 'Theory and Practice of Photon Correlation Spectroscopy', in *Measurement of Suspended Particles by Quasi-elastic Light Scattering*, ed. B.E. Dahneke, Wiley-Interscience, New York, 1983.
32. Honeywell IAC (Microtrac), 1100 Virginia Dr., Fort Washington, PA 19034, USA.
33. P.N. Pusey, in *Photon Correlation Spectroscopy and Velocimetry*, ed. H.Z. Cummins, E.R. Pike, Plenum, New York, 1977.
34. R.J. Glauber, 'Coherent and Incoherent States of the Radiation Field', *Phys. Rev.*, **131**, 2766–2788 (1963).
35. A.J.F. Siegert, 'On the Fluctuations in Signals Returned by many Independently Moving Scatterers', Massachusetts Institute of Technology, Boston, MA, M.I.T. Rad. Lab. Rep., No. 465, 1943.
36. C. Tanford, *Physical Chemistry of Macromolecules*, J. Wiley and Sons, New York, 1957.
37. H.Z. Cummins, F.D. Carlson, T.J. Herbert, G. Woods, 'Translational and Rotational Diffusion Constants of Tobacco Mosaic Virus from Rayleigh Line Widths', *Biophys. J.*, **9**, 518 (1969).
38. B. Chu, Es. Gulari, Er. Gulari, 'Details of Histogram Approach and Comparison of Methods of Data Analysis', *Phys. Scr.*, **19**, 476 (1979).
39. J.G. McWhirter, E.R. Pike, 'Numerical Inversion of Laplace Transform and Similar Fredholm Integral Equation of 1st Kind', *J. Phys. A: Math. Gen.*, **11**, 1729 (1978).
40. J.G. McWhirter, 'A Stabilized Model-fitting Approach to the Processing of Laser Anemometry and other Photon-correlation Data', *Opt. Acta*, **27**, 83 (1980).
41. N. Ostrowsky, D. Sornett, P. Parker, E.R. Pike, 'Exponential Sampling Method for Light-scattering Polydispersity', *Opt. Acta*, **28**, 1059 (1981).
42. S.W. Provencher, J. Hendrix, L. De Maeyer, N. Paulussen, 'Direct Determination of Molecular-Weight Distributions of Polystyrene in Cyclohexane with Photon Correlation Spectroscopy', *J. Chem. Phys.*, **69**, 4273 (1978).
43. S.W. Provencher, 'Inverse Problems in Polymer Characterization: Direct Analysis of Polydispersity with Photon Correlation Spectroscopy', *Macromol. Chem.*, **180**, 201 (1979).
44. B. Weiner, W. Tscharnuter, A. Banerjee, 'Developing Fiber Optic Probes for Noninvasive Particle Size Measurements in Concentrated Suspensions using DLS. Or, How to Make a Zero Shear Rheometer without Really Trying', Brookhaven Corp. Internal Memorandum, 1996.
45. S.F. Gull, *Developments in Maximum Entropy Data Analysis*, Kluwer Academic, Norwell, MA, 53–71, 1989.
46. J. Jakes, 'Testing of the Constrained Regularization Method of Inverting Laplace Transform on Simulated very wide Quasi-electric Light Scattering Auto-Correlation Functions', *Czech. J. Phys.*, **38**, 1305 (1988).
47. Chris De Vos, Luc Deriemaeker, Robert Finsy, 'Quantitative Assessment of the Conditioning of the Inversion of Quasi-elastic and Static Light Scattering Data for Particle Size Distributions', *Langmuir*, **12**, 2630–2636 (1996).
48. Robert Finsy, 'Particle Sizing by Quasi-elastic Light Scattering', *Adv. Coll. Interf. Sci.*, **52**, 79–143 (1994).
49. H. Ruf, E. Grell, E.H.K. Stelzer, 'Size Distribution of Submicron Particles by Dynamic Light Scattering Measurements: Analyses Considering Normalization Errors', *Eur. Biophys. J.*, **21**, 21–28 (1992).
50. J.C. Brown, P.N. Pusey, R. Dietz, 'Photon Correlation Study of Polydisperse Samples of Polystyrene in Cyclohexane', *J. Chem. Phys.*, **62**, 1136 (1975).
51. D.E. Koppel, 'Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants', *J. Chem. Phys.*, **57**, 4814 (1972).
52. John C. Thomas, 'The Determination of Log Normal Particle Size Distributions by Dynamic Light Scattering', *J. Coll. Interf. Sci.*, **117**(1) (1987).
53. R.J. Hunter, *Zeta Potential in Colloid Science, Principles and Applications*, Academic Press, London, 17–21, 1981.
54. ISO document, 'Particle Size Analysis – Photon Correlation Spectroscopy', Draft International Standard ISO/DIS 13321, ICS 19.120.00, 1994.

55. EG&G Electro Optics, P.O.Box 900, Vaudreuil J7V 7X3, Canada.
56. R.G.W. Brown, J.G. Burnett, J. Mansbridge, C.I. Moir, 'Miniature Laser Light Scattering Instrumentation for Particle Size Analysis', *Appl. Opt.*, **29**, 4159–4169 (1990).
57. E. Jakeman, R. Jones, C.J. Oliver, E.R. Pike, 'Optical Signal Processing', Royal Radar Establishment, Ministry of Defence, Malvern, Worcs, UK, RRE Memorandum, No. 2621, 1970.
58. Malvern Instruments Ltd., Malvern, Worcestershire, UK.
59. L. Cantu, M. Corti, M. Musolino, P. Salina, 'Spontaneous Vesicle Formation from a One-component Solution of a Biological Surfactant', *Progr. Colloid. Polym. Sci.*, **84**, 21–23 (1991).
60. H.S. Dhadwal, William W. Wilson, Rafat R. Ansari, William V. Meyer, 'Dynamic Light Scattering Studies of BSA and Lysozyme Using a Backscatter Fiber Optic System', Proceedings Static and Dynamic Scattering in Medicine and Biology, Society of Photo-optical Instrumentation Engineers, Bellingham, WA 98227, SPIE Vol. 1884, 1993.
61. V. Mikol, E. Hirsch, R. Giege, 'Monitoring Protein Crystallization by Dynamic Light Scattering', *FEBS Lett.*, **258**(1), 63–66 (1989).
62. C.A. Herb, E.J. Berger, K. Chang, I.D. Morrison, E.F. Grabowski, 'Using Quasi-elastic Light Scattering to Study Particle Size Distributions in Submicrometer Emulsion Systems', ACS Symposium Series No. 332, 1987.
63. Avestin Inc., 2450 Don Reid Drive, Ottawa K1H 8P5, Canada.
64. W. van Megen, S.M. Underwood, 'Dynamic-light-scattering Study of Glasses of Hard Colloid Spheres', *Phys. Rev. E*, **47**(1), 248–261 (1993).
65. Ronald D. Neuman, M.A. Jones, Nai-Fu Zhou, 'Photon Correlation Spectroscopy Applied to Hydro Metallurgical Solvent Extraction Systems', *Coll. Surf.*, **46**, 45–64 (1990).
66. Karin D. Caldwell, Jianmin Li, 'Emulsion Characterization by the Combined Sedimentation Field-flow Fractionation–Photon Correlation Spectroscopy Methods', *J. Coll. Interf. Sci.*, **132**(1), 256–268 (1989).
67. P.G. Cummins, E.J. Staples, 'Particle Size Distributions Determined by a ''Multi angle'' Analysis of Photon Correlation Spectroscopy Data', *Langmuir*, **3**, 1109–1113 (1987).
68. H. Auweter, D. Horn, 'Fiber-optical Quasi-elastic Light Scattering of Concentrated Dispersions', *J. Coll. Interf. Sci.*, **105**(2), 399–409 (1985).
69. Allgemeine Laser Vertriebsgesellschaft mbH, Robert-Bosch-Str. 9, D-63225 Langen, Germany.
70. Brookhaven Instruments Corp., 750 Blue Point Road, Holtsville, NY 11766, USA.
71. Flexible Instruments, 15 Colmart Way, Bridgewater, NJ 08807, USA.
72. J. Ricka, 'Brownian Dynamics in Strongly Scattering Porous Media – Dynamic Light Scattering with Single-mode Matching', *Macromol.*, **79**, 45–55 (1994).
73. D.J. Pine, D.A. Weitz, P.M. Chaikin, E. Herbholzheimer, 'Diffusing-wave Spectroscopy', *Phys. Rev. Let.*, **60**(12), 1134–1137 (1988).
74. William V. Meyer, David S. Cannell, Anthony E. Smart, Thomas W. Taylor, Padetha Tin, 'Multiple Scattering Suppression by Cross-correlation', Applied Optics, L.P.E. Div., Photon Correlation & Scattering, 1997.
75. W.V. Meyer, W.W. Tscharnuter, A.D. MacGregor, H. Dauter, P. Deschamps, F. Boucher, J. Zhu, P. Tin, R.B. Rogers, R.R. Ansari, 'Laser Light Scattering, From an Advanced Technology Development Program to Experiments in a Reduced Gravity Environment', International Symposium on Space Optics, Garmisch-Partenkirchen, Germany, EOS/SPIE Joint Venture, paper 2210-2221, April 18–22, 1994.
76. Harbans S. Dhadwal, Rafat R. Ansari, William V. Meyer, 'A Fiber-optic Probe for Particle Sizing in Concentrated Suspensions', *Rev. Sci. Instrum.*, **62**(12), 2963–2968 (1991).
77. Coulter Beckman Corp., P.O. Box 169015, Miami, FL 33116-9015, USA.
78. Particle Sizing Systems, 75 Aero Camino, Santa Barbara, CA 93117, USA.
79. Protein Solutions, Inc., 1224 West Main Street, Suite 777, Charlottesville, VA 22903.

Sedimentation in Particle Size Analysis

Claus Bernhardt
Technische Universität Bergakademie Freiberg, Germany

1 Introduction **1**
2 Particle Movement in a Fluid at Rest **2**
2.1 Undisturbed Particle Movement 2
2.2 Disturbances and Influences on Particle Movement 8
3 Dispersion of Solids in Liquids **15**
3.1 Introduction and Problems 15
3.2 Stabilization of Suspensions 16
3.3 Desagglomeration Methods 18
3.4 Characterization of the Dispersion State and the Dispersion Capacity 19
4 Sedimentation Methods **22**
4.1 Brief Historical View 22
4.2 Gravitational Methods 22
4.3 Methods Used in the Centrifugal Field 32
Acknowledgments **36**
List of Symbols **36**
Related Articles **38**
Appendix **38**
List of Liquids and Dispersing Aids 38
Liquids 40
Dispersing Aids 40
References **40**

Particle size analysis by sedimentation is one of the most important techniques of analysis. The principle is based on the fact that it is possible to draw conclusions about the geometrical dimensions (size) of the particles from the speed of their movement in a fluid. In the past many different apparatuses were developed. Nowadays, methods that allow quick, automatic measurement are most popular.

This article consists of two general sections valid for all methods: the first concerns the undisturbed and disturbed movement of particles in a fluid at rest, and the second concerns the dispersion of solids in liquids.

The methods described are divided in gravitational methods and methods used in a centrifugal field, respectively. Finally, the appendix gives a list of liquids and dispersing agents valid for measurement.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

It is well known that the measuring principle of sedimentation methods is partially based on the fact that it is possible to make conclusions about the geometrical dimensions (size) of particles from knowledge of the speed of their movement in a fluid. Therefore, the temporal and spatial changes in the particle concentration in a sedimenting system provide substantial information on the quantitative distribution of particle sizes. The measurement of such distributions is called sedimentation analysis; it requires the theoretical understanding and experimental handling of the following two phenomena:

- the connection between sedimentation velocity and particle size depending on the material properties of the system and the other conditions in the sedimentation vessel,
- the connection between the concentration distribution of the solid observed after time t and its size distribution. This requires methods that both indicate concentration changes and allow calculation of the distribution from the measured data.

The methods of sedimentation analysis are grouped according to the following fundamental criteria:[1–13]

- according to the position of particles at the beginning of measurement
  - homogeneous methods: the particles are uniformly distributed over the whole sedimentation space
  - line start methods: the particles in the beginning are concentrated in a thin layer on top of the solid-free medium
- according to the place of quantity measurement
  - incremental methods: the amount (concentration) of solids in a thin suspension layer (measurement level) is measured
  - cumulative methods: the total amount of solids above or below a measurement level is measured
- according to the force field
  - gravitational methods
  - centrifugal methods.

In effect, there are eight groups of sedimentation methods which, however, are not of equal practical significance. Methods widely and frequently used at the moment include the incremental and partially the cumulative methods in a gravitational or centrifugal field.

For incremental methods, the mass or volume distribution are directly measured in the one-layer (suspension) method, and the corresponding frequency distribution is measured in two-layer (line start) methods. In contrast, the mass or volume distribution are obtained for the two-layer scheme in the case of cumulative methods. For a cumulative one-layer scheme, the distribution has to be determined by differentiation on the basis of the cumulative measurement result.

A further subdivision of the sedimentation methods is based on the type of principle for measuring the solids concentration:

- Gravimetric determination (mass)
  - weighing of the solid mass contained in the suspension sample (pipette methods, decanting)
  - weighing of the solid mass sedimented on a balance pan (sedimentation balances)
- Measurement of the absorption of electromagnetic radiation
  - light attenuation (photometric extinction methods)
  - attenuation of γ-radiation
- Measurement of suspension density
  - manometer methods
  - diver methods (areometers, divers)
- Measurement of the backscattering of β-radiation.

Again, these principles of measurement are not all of the same importance for the different types of sedimentation analysis. A large number of known methods (several reviews can be found in the literature[1–15]) can be found in granulometric measurement technology. The following advantages may have played a decisive role in this:

1. Like all classification methods (and unlike the nonfractionating methods), sedimentation analysis gives the proportion of particles that are smaller than the smallest measurable particle size.
2. Very simple and inexpensive devices can be used which can even be built by the user.
3. Frequently measurement is possible without the disturbance of interference to the system.
4. At a higher cost, fast automatic devices can be employed.
5. The range of measurement can be extended below 0.1 µm if a centrifugal field is used.

Disadvantages particularly in comparison with modern nonfractionating methods are:

1. There are numerous subjective and objective sources of error.
2. The maximum particle size is usually approx. 50 µm; larger particles can be measured only in special cases.
3. The time of analysis during which numerous conditions have to be kept constant is relatively long in comparison with modern nonfractionating methods.

The latter problem has been most responsible for the decline in use of a considerable number of previously widespread devices and methods. This is especially the case for methods that do not allow for a decisive reduction in the sedimentation height, such as sedimentation balance, diver, and so on.

In the following chapters we will deal with particle movement in a fluid, dispersion problems and then with some of the techniques of sedimentation analysis.

2 PARTICLE MOVEMENT IN A FLUID AT REST

2.1 Undisturbed Particle Movement

Undisturbed sedimentation takes place if the particles in a fluid of infinite expansion are separated by such distances that any mutual influence can be excluded.

2.1.1 Stationary Movement of Spherical Particles

In the stationary case, a particle moves with constant velocity U relative to the surrounding medium. There is an equilibrium between the weight G_P minus buoyancy A_P and the drag force W_P exerted on the particle by the medium (fluid), Equation (1):

$$W_P = G_P - A_P = V_P(\rho_s - \rho_m)g \tag{1}$$

Furthermore, according to Newton, Equation (2):

$$W_P = \tfrac{1}{2} c_W F_P \rho_m U^2 \tag{2}$$

From Equations (1) and (2), the fundamental equation of stationary particle movement in a fluid is obtained, Equation (3):

$$U^2 = \frac{2\Delta\rho g}{c_W \rho_m} \frac{V_P}{F_P} \tag{3}$$

For spherical particles (Equation 4) with

$$V_P = \frac{\pi}{6}x^3; \quad F_P = \frac{\pi}{4}x^2 \tag{4}$$

Equations (2) and (3) become Equations (5) and (6)

$$W_P = \frac{\pi}{8}\rho_m c_W x^2 U^2 \tag{5}$$

$$U^2 = \frac{4}{3}\frac{\Delta\rho g}{c_W \rho_m} x \tag{6}$$

The particles are usually assumed to be spherical in shape. This, however, leads to a fundamental discrepancy with respect to the real conditions in most cases, because the size of a particle of any shape is described by the dimensions of a solid sphere which has the same hydrodynamic behavior as the particle. It is clear that such an approach is acceptable only as long as the shape of the particle does not change between measurements; otherwise, grain-shape characteristics have to be included in the description of the hydrodynamics of the particle.

For the entire size range, the dependence of the drag coefficient on the sedimentation rate of spheres can be determined only empirically. Some values are collected in Table 1.

The curves are shown in Figure 1, with the Reynolds number, Equation (7)

$$Re = \frac{\rho_m U x}{\eta} \tag{7}$$

on the axis of abscissa.

An exact physical description of the dependence $c_W(Re)$ has so far been impossible because the Navier–Stokes differential equations (see Happel and Brenner,[17] and Lamb[18]) can be integrated only if certain terms are neglected, so that the solutions are always true only for a limited range of *Re* numbers.

By neglecting inertia terms in the Navier–Stokes equations, Stokes[19] in the middle of 19th century found

Table 1 Empirical drag coefficients for spheres according to Allen[16]

Re	c_W	$c_W Re^2$	c_W/Re	$c_W = 24/Re$ (Stokes)
0.01	2400	0.24	240 000	2400
0.02	1200	0.48	60 000	1200
0.05	484	1.21	9680	480
0.1	244	2.44	2440	240
0.2	123	4.92	615	120
0.5	51.4	12.85	102.8	48
1	27.2	27.2	27.2	24
2	15.0	60.0	7.5	12
4	8.33	133.3	2.083	6.0
5	7.12	178	1.424	4.8
10	4.35	435	0.435	2.4
20	2.74	1096	0.137	1.2
50	1.56	3900	0.0312	0.48
100	1.10	11 000	0.011	0.24

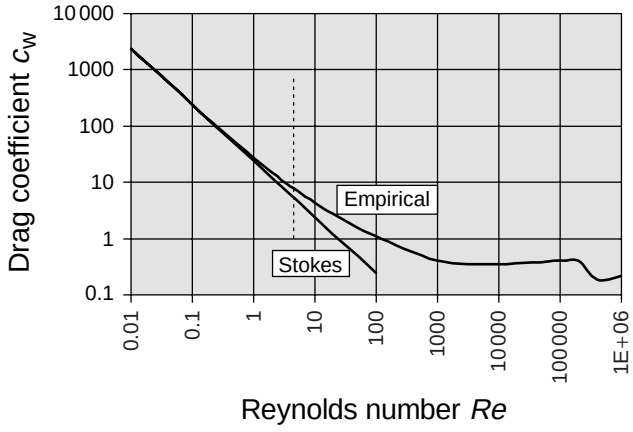

Figure 1 Drag coefficient plotted against Reynolds number.

the formula, Equation (8)

$$W_{St} = 3\pi\eta x U \tag{8}$$

for the drag force. This is equivalent to, Equation (9), (see Figure 1)

$$c_W = \frac{24}{Re} \tag{9}$$

With Equation (6), this gives the well-known "Stokes equation", Equation (10)

$$U = \frac{\Delta\rho g}{18\eta} x^2 \tag{10}$$

However, there are lower and upper limits for using the Stokes equation. If the Reynolds number and the particle size are very small and the latter is of the same order of magnitude as the mean free path of the fluid molecules, Equation (11)

$$L = \frac{3\eta}{\rho_m}\left(\frac{M_M}{3kTN_L}\right)^{1/2} \tag{11}$$

one can no longer assume that the fluid is a continuum. These conditions lead to a lower drag force and, consequently, to an increase in the stationary sedimentation velocity.

For liquids, the mean free path L of the molecules is about 10^{-9} cm. On the other hand, particles whose size is determined by the Stokes velocity are no smaller than approx. 10^{-6} cm (0.01 μm) so that the effect can be neglected in liquids. The situation is different with gases. Their mean free path is approx. 10^{-5} cm under normal conditions, that is the drag force must be corrected. The general type of this so-called "Cunningham correction" is given by Equation (12)

$$W = 3\pi\eta\frac{U}{C_L} \tag{12}$$

(terms for C_L see Table 2).

Table 2 Relationships for C_L in Equation (12)

$C_L =$	Ref.	Valid in the range
$\frac{(1+6AL/x)}{(1+4AL/x)}$	17, 20	
$1+\ 2.514+0.8\exp\ -0.55\frac{x}{L}\ \ \frac{L}{x}$	21	$2\frac{L}{x}<1$
$1+2A\frac{L}{x}$	16, 22	$A=0.7\ldots1.4$
$\frac{1}{(1-2AL/x)}$	17, 23	
$4.49\frac{L}{Bx}$	16	$2\frac{L}{x}>1$
$3\frac{L}{Bx}$	24	$B=1\ldots1.4$

Generally, correction is necessary for particles $< 10\,\mu m$ in air. Without any correction, the measured values $Q_3(x)$ are shifted in the direction of the coarser range with respect to the actual particle size distribution.

For large Reynolds numbers the empirical drag coefficient is larger than that predicted by Stokes, therefore the settling velocities in sedimentation will be smaller than expected. This means that the measurement values $Q_3(x)$ are shifted in the direction of smaller sizes. In general, a size error of approx. 1–2% (equivalent to an error in the drag coefficient of approx. 3–5%) can be accepted even for very narrow distributions. Then the upper limit of the acceptable Reynolds numbers is between 0.25 and 0.5 for these values.[13]

A considerable number of equations have been proposed in the literature to enable further extension towards higher Reynolds numbers.[13,25] On the whole, they all have the fundamental structure, Equation (13)

$$c_{W,i} = \frac{24}{Re} B_i \tag{13}$$

For examples of B_i see in Table 3.

The value of the maximum permitted Reynolds number Re_{max} applies to maximum particle size errors of $\pm 1\%$ and $\pm 2\%$ respectively. They were calculated on the basis of the difference between c_W according to Equation (13) and empirical c_W values (Allen,[16] Bernhardt;[31] see Table 1). The maximum particle size x_{max} can be calculated by, Equation (14)

$$x_{max}^3 = \frac{18\eta^2}{\rho_m \Delta\rho g} B_i Re_{max} \tag{14}$$

Table 3 also shows the case of quartz in water.

Unfortunately, some of the equations (Proudman and Pearson,[28] Schiller and Naumann[30]) do not allow for an analytical solution of the sedimentation velocity U or the particle size x, so that an iteration has to be programmed before they can be used.

On the other hand, in such cases the empirical calculation can possibly be done without an analytical approximation equation.

In 1933 Schiller and Naumann[30] developed a graphical method which later was improved and modified by several authors.[32,33] It is based on the calculation of either Equation (15)

$$c_W Re^2 = \frac{4\Delta\rho g \rho_m x^3}{3\eta^2} = \frac{4}{3} Ar \tag{15}$$

or Equation (16)

$$\frac{c_W}{Re} = \frac{4\Delta\rho g \eta}{3\rho_m^2 U^3} = \frac{4}{3}\frac{1}{Lj} \tag{16}$$

depending on whether the particle size x or the sedimentation velocity U is given. These values are points on the empirical curve c_W/Re ($c_W Re^2$) which are represented in Figure 2 (for further values, see e.g. Clift[25]).

Because of the large error in the graphical method, the following approximation, Equation (17) is useful for a numerical calculation (see Davies[21]):

$$Re = 4.1667\times10^{-2}K - 2.336\times10^{-4}K^2 + 2.0154\times10^{-6}K^3 - 6.9105\times10^{-9}K^4 \tag{17}$$

where $K = c_W Re^2$

Table 3 Equations of $c_W(Re)$

Author	Ref.	B_i in Equation (13)	Re_{max}		x_{max} (μm)	
			$f_x = 1\%$	$f_x = 2\%$	$f_x = 1\%$	$f_x = 2\%$
Stokes	19	1	0.25	0.5	65	82
Oseen	26	$1+0.1875\,Re$	0.75	1.5	104	129
Gumz	27	$1+0.0278\,Re+0.00364\,Re^2$	0.3	0.6	70	88
Proudman and Pearson	28	$1+0.01875\,Re+0.1125\,Re^2\ln Re$	0.75	1.2	94	112
Möller	29	$1+0.139\,Re$	2.5	3.5	155	180
Bernhardt	13	$1+0.0917\,Re$	0.7	10	94	277
Schiller and Naumann	30	$1+0.15\,Re^{0.687}$	5	1000	201	2728

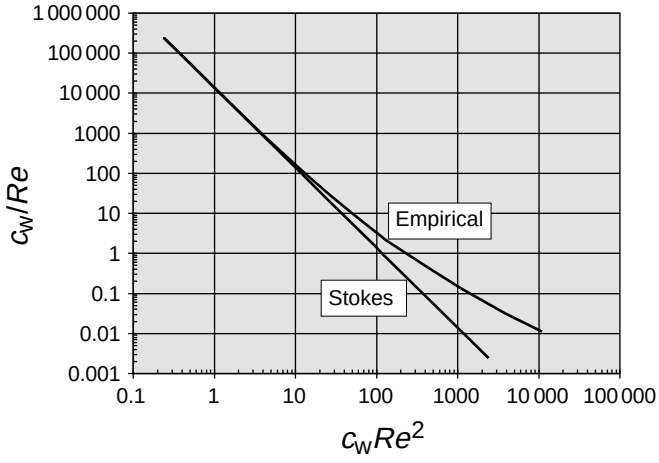

Figure 2 Empirical curve c_W/Re plotted against $c_W Re^2$.

It fits very well in the range $Re \leq 4$ ($c_W Re^2 \leq 133$): the drag error remains $\leq 0.6\%$.[31]

To calculate U for a given x (outside the Stokes range, i.e. for large Re numbers):

1. Calculate $c_W Re^2$ according to Equation (15), take out the c_W/Re value from Figure 2 and calculate U by means of transformed Equation (16) or
2. Calculate $c_W Re^2$ according to Equation (15), calculate Re by means of Equation (17) and then U by means of Equation (7).

2.1.2 Nonstationary Movement of Spherical Particles

If a particle in a fluid is subject to an acceleration, the process is called nonstationary movement. Such a case is given when, in a fluid at rest, the particle is brought from its rest position ($t = 0$; $v = 0$) to its terminal velocity of settling ($t = \infty$; $v = U$). In contrast to the stationary case, the (inertial) force acting on the particle is now equal to the difference between its weight minus buoyancy and the drag force of the medium, Equation (18)

$$m_P \frac{dv}{dt} = G_P - A_P - W_P \tag{18}$$

In the creeping flow range the instantaneous drag depends not only on the instantaneous velocity and acceleration but also on the conditions prevailing during the formation of the flow. For the drag force, the simple Equation (8) is no longer true and has to be replaced by Equation (19)

$$W_P = 3\pi\eta x v + \frac{1}{2} V_P \rho_m \frac{dv}{dt} + \frac{3}{2} x^2 \sqrt{\pi\eta\rho_m} \int_0^t \frac{dv/ds}{\sqrt{t-s}}\, ds \tag{19}$$

according to Clift et al.[25]

The first term is the Stokes drag for stationary motion at instantaneous velocity. The second term is the ''added mass'' or ''virtual mass'' caused because the acceleration of the particles requires acceleration of the fluid. The third term includes the ''Basset history integral'' representing past acceleration, where $(t - s)$ is the time elapsed since the past acceleration.[25] The substitution of Equation (19) into Equation (18) produces a differential equation with some difficulties for its practical application.[34] Therefore the solution without the Basset term is used in most cases:

$$\frac{v}{U} = 1 - \exp\left\{ -\frac{18\eta}{\rho_s + \rho_m/2} \frac{t}{x^2} \right\} \tag{20}$$

The time t_{99} for which $v/U = 0.99$ is then (without the Basset term):

$$t_{99} = 0.2558 \frac{\rho_s + \rho_m/2}{\eta} x^2 \tag{21}$$

This approximation is sufficient for solids in gases ($\rho_s/\rho_m > 1000$) only. In the case of large solid particles in liquids ($\rho_s/\rho_m < 100$) the Basset term must not be neglected; according to the calculations of Clift et al.[25] Equation (22) holds (including the Basset term):

$$t_{99} \approx 800 \frac{\rho_m x^2}{\eta} \tag{22}$$

Large particles with sizes near the upper border of the Stokes range need a few seconds to come close to the terminal velocity from the initial quiescent state. As an example, quartz particles with $x = 0.1$ mm in water give $t_{99} = 8$ s according to Equation (22).

2.1.3 Stationary Movement of Nonspherical Particles

2.1.3.1 Fundamentals Hydrodynamic concepts not related to a sphere have to be related to a different defined geometric shape, which should apply flexibly to all existing shapes from a very thin plate to a very long rod. Such a shape is the ellipsoid with the half axes a, b, c for which the equations of motion have long been known.[17,18,35,36] For the important borderline cases of the ellipsoid, that is, plates ($a = b > c$) and rods ($b = c < a$) Gans[36] gives complete analytical solutions for the sedimentation of the particles in the direction of either their a-axis or their c-axis. For these shape types Equations (23) and (24) hold

$$U_P = \frac{2}{9} \frac{\Delta\rho g a c}{\eta} R_P \quad \text{(for plates)} \tag{23}$$

$$U_R = \frac{2}{9} \frac{\Delta\rho g c^2}{\eta} R_R \quad \text{(for rods)} \tag{24}$$

Table 4 Factors R_P and R_R in Equations (23) and (24)

Shape		R	Equation
Plates $a = b > c$	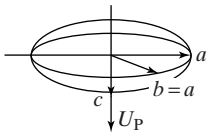	c-direction: $R_{P,c} = 0.750\ (A_1 \arccos p + B_1)$ a-direction: $R_{P,a} = 0.375\ (A_2 \arccos p - B_1)$	(25) (26)
Rods $b = c < a$	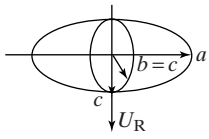	c-direction: $R_{R,c} = 0.375\ (A_3 \ln C - B_2)$ a-direction: $R_{R,a} = 0.1875\ (A_4 \ln C - B_2)$	(27) (28)

$A_1 = D(1-2p^2)$; $A_2 = D(3-p^2)$; $A_3 = D(2-p^2)$; $A_4 = D(2-3p^2)$. $B_1 = p/E$; $B_2 = 2/E$; $C = (1+E^{1/2})/(1-E^{1/2})$; $D = E^{-3/2}$; $E = (1-p^2)$; $p = c/a$.

Table 5 Values of R_P and R_R according to Gans[36] and Happel and Brenner[17]

$p = c/a$	Plates		Rods	
	$R_{P,a}$	$R_{P,c}$	$R_{R,a}$	$R_{R,c}$
0	1.767	1.178	∞	∞
0.01	1.752	1.178	7.197	4.348
0.1	1.631	1.174	3.764	2.617
0.2	1.516	1.160	2.799	2.108
0.4	1.335	1.125	1.914	1.606
0.6	1.196	1.084	1.467	1.326
0.8	1.139	1.041	1.191	1.138
0.9	1.041	1.019	1.091	1.063
1	1	1	1	1

The terms for the factors R_P and R_R for both directions c and a can be taken from Table 4.

Table 5 also contains some values of R_P and R_R for the two special sedimentation directions.

For spherical shape ($p = 1$), all R values assume the value 1 so that Equations (23) and (24) become identical with the Stokes formula.

According to Gans,[36] for very thin plates $p \ll 0.2$ there is an acceptable approximation, Equation (29)

$$R_{P,c} = \frac{3\pi}{8} = 1.178 \tag{29}$$

from which the well-known formula of the drag force, Equation (30)

$$W_{P,c} = 16\eta a U \tag{30}$$

is derived.[18,37] This, however, is true only for thin plates whose large half axis a is in the horizontal plane, and which sediment in the direction of their short half axis c (vertically). Experimental investigations concerning the validity and applicability of these theoretical concepts have frequently been published in the literature.[32,33,37–43]

The sedimentation directions discussed so far relate to the position of the a-axis or c-axis of the spheroids in the direction of sedimentation. In reality, however, every particle orientation has the same probability at the beginning of sedimentation. Theoretically, this situation remains unchanged in the range of the creep flow for the whole sedimentation path.[17]

Figure 3 shows schematically a spheroid, whose a-axis is inclined towards the horizontal x-axis by the angle α.

Dependent on its inclination with respect to the horizontal, forces (not represented) act on the particle in such a way that a resulting velocity U with the horizontal and vertical components U_x and U_z is formed. Equations (31–33) are valid:[36,44]

$$U^2 = U_a^2 \sin^2\alpha + U_c^2 \cos^2\alpha \tag{31}$$

$$U_x = (U_a - U_c) \sin\alpha \cos\alpha \tag{32}$$

$$U_z = U_a \sin^2\alpha + U_c \cos^2\alpha \tag{33}$$

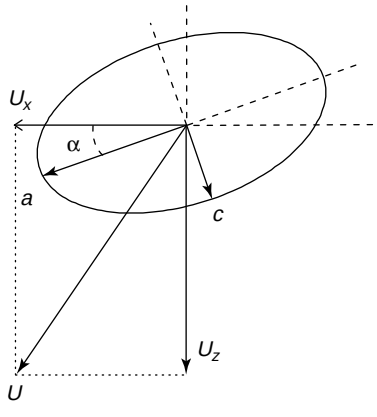

Figure 3 Velocities for an arbitrary position of the particle.

Table 6 Velocity ratios U/U_z and U_x/U_z of extremely thin plates and rods dependent on their angular position

α (°)	Plates $p = 0, U_c/U_a = 0.667$		Rods $p = 0, U_c/U_a = 0.5$	
	U/U_z	U_x/U_z	U/U_z	U_x/U_z
0	1.000	0	1.000	0
10	1.004	0.086	1.013	0.167
20	1.011	0.153	1.041	0.289
30	1.019	0.195	1.059	0.347
40	1.021	0.206	1.059	0.348
50	1.019	0.193	1.047	0.310
60	1.012	0.159	1.031	0.248
70	1.006	0.112	1.015	0.171
80	1.002	0.058	1.004	0.087
90	1.000	0	1.000	0

U_a and U_c are the velocities in the extreme positions $\alpha = 90°$ and $\alpha = 0°$. Table 6 gives some values of U/U_z and U_x/U_z dependent on the angle α.

It is evident that the horizontal component U_x can assume considerable values for extreme particle shapes. In most cases, however, the differences between the resultant U and its vertical component U_z are negligible. Experimental investigations of plates[43] and rods[44] fully confirm this theory.

2.1.3.2 Shape Factors Knowledge of the relationship between the geometrically defined shape of a particle and its hydrodynamic behavior suggests that it must be possible to obtain characteristics of the shape of real particles from their sedimentation velocity. There are two basic definitions of such characteristics; either the drag force or the sedimentation velocity of a particle of the same volume is corrected.

Dynamic particle shape factor κ, Equations (34) and (35)

$$W = 3\pi\eta U x_v \kappa \tag{34}$$

and

$$\kappa = \frac{W}{W_v} = \frac{U_v}{U} \tag{35}$$

where x_v is the diameter of a sphere which has the same volume as the particle and U_v is the stationary sedimentation velocity of a sphere of the same volume.

Shape correction factor K, Equations (36) and (37)

$$U = \frac{\Delta\rho g}{18\eta} x_v^2 K \tag{36}$$

and

$$K = \frac{U}{U_v} = \frac{1}{\kappa} \tag{37}$$

x_v has to be determined if the equations of motion are to be used for particle shape definition. Again, the approximation to an ellipsoid has proved highly useful since the equations of motion are well known for this geometrical body. The basis for this approach are the work of Heiss and Coull[38] for factor K and later of Walkenhorst[39,40] for κ.

Taking into account the equality of the volume of sphere and ellipsoid, for any orientation of the particle, K_P and K_R are obtained,[44] Equations (38) and (39)
for plates

$$K_P = \frac{U_z}{U_v} = p^{1/3}[R_{P,a}\sin^2\alpha + R_{P,c}\cos^2\alpha] \tag{38}$$

for rods

$$K_R = p^{2/3}[R_{R,a}\sin^2\alpha + R_{R,c}\cos^2\alpha] \tag{39}$$

There have been many attempts to correlate the shape factors defined by the particle geometry and the hydrodynamic behavior (see the overview by Pahl et al.[213]). Wadell[46] and Heywood[32,47] reported this in great detail. Wadell[46] holds the view that for sedimentation processes, the proportion of the surface area to the volume is the decisive factor. This leads him to the shape factor called sphericity, Equation (40)

$$\psi_{Wa} = \frac{x_V^2}{x_A^2} \tag{40}$$

where x_A is the diameter of the sphere of equal surface area.

Also by comprehensive experiments with isometric particles, Pettyjohn and Christiansen[41] found the empirical relationship between ψ_{Wa} and the shape factor K, Equation (41)

$$K = 0.843 \log\left(\frac{\psi_{Wa}}{0.065}\right. \tag{41}$$

Another shape factor, also defined by Heywood[32] was called the volume coefficient, Equation (42)

$$k = \frac{V_P}{x_F^3} \tag{42}$$

where x_F is the equivalent projected area diameter.

Heiss and Coull[38] eventually found empirical equations for the shape factor K; they also included, in addition to ψ_{Wa}, a new factor which was called circularity, Equation (43)

$$\psi_{HC} = \frac{x_V}{x_F} \tag{43}$$

2.2 Disturbances and Influences on Particle Movement

2.2.1 Apparent Distributions

Numerous influences on the sedimentation motion mean that the true distribution is not measured, only an apparent one. Consequently, a particle system presumed to be monodisperse is then measured as an apparently polydisperse system. Factors contributing to this effect are:

- Factors caused by the particles
  - deviation from the spherical shape
  - Brownian motion of the particles
  - heterogeneity of the solid
- Factors caused by the method
  - limitations of the vessel
  - extension of the measuring gap.

The question of great practical interest is, what distributions are to be measured finally if the starting distribution is already polydisperse? This problem can be solved by the strict application of the ''concept of apparent distributions'' developed by Bernhardt.[48] It allows passage from the monodisperse case to the polydisperse one.

What is the principle of the concept mentioned? The explanation is given in Figure 4.

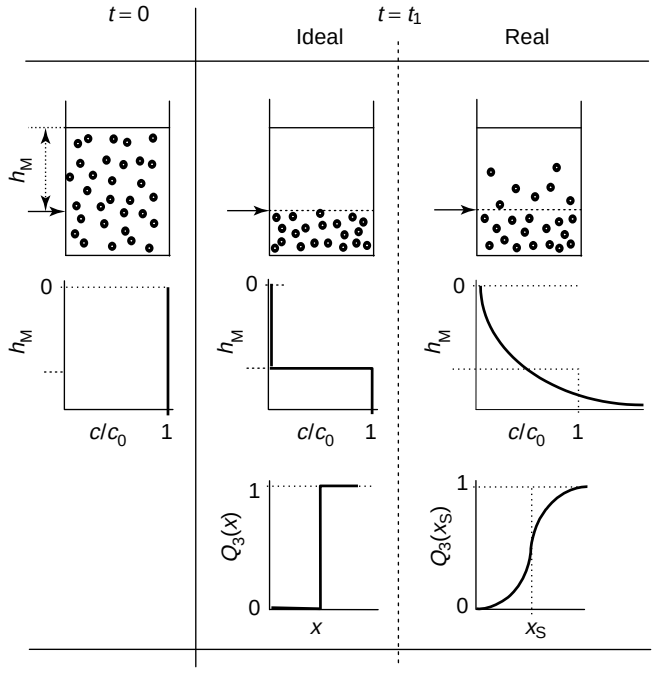

Figure 4 Ideal and real sedimentation.

The upper part of the figure shows the principle of the incremental sedimentation analysis: the solid particles are homogeneously distributed in a vessel at the time $t = 0$; the solid concentration is constant at every depth h as shown below. A sensor, situated at a depth h_M, measures the solid concentration there as a function of time. In the ideal case, the highest particle layer of a monodisperse system just reaches the measuring depth h_M at the time $t = t_1$. This leads to a concentration jump at this point and the cumulative distribution represents monodispersity exactly. In the real case, however, after t_1 there is a more or less diffuse border near the measuring plane. Consequently, the measured course of concentration is as shown in Figure 4, which results in an apparent polydisperse distribution.

2.2.2 Particle Shape

To estimate the influence of the particle shape on the results of the sedimentation analysis, we assume:

1. The particle system is monodisperse with constant particle shape.
2. It can be described by spheroids as explained above in section 2.1.3 (c or a and p = const.).
3. The actual particle orientation does not change during the sedimentation (α = const.).

In Figure 5(a) such a particle system is schematically represented in its initial state ($t = 0$); for the sake of simplicity, particles of the same orientation were arranged within one column.

After a certain period of time $t = t_1$, the picture becomes similar to that in Figure 5(b): the particles sedimenting in a-direction have traversed a longer distance than those with a smaller angle ($\alpha < 90°$). Therefore, unlike the case of monodisperse spheres, there will be no distinct concentration jump in the depth h_M (on the measurement level) but there will be a concentration change whose course will be reflected as an apparent particle size distribution $Q_3(x_S)$ as a result of the measurement. Owing to the equal probability of all

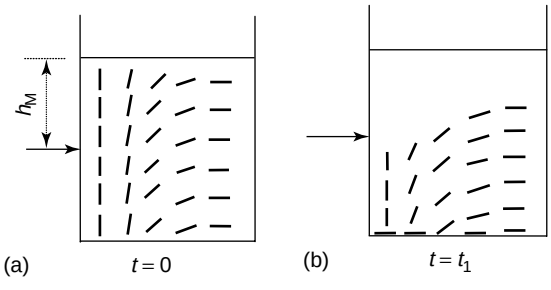

Figure 5 Sedimentation of nonspherical particles.

angular positions, Equation (44) holds

$$Q_3(x_S) = \frac{c}{c_0} = \frac{\alpha}{\alpha_{max}} = \frac{2\alpha}{\pi} \tag{44}$$

The apparent Stokes particle size x_S is calculated by equating U_z (Equation 33) with Stokes equation. Then, for the two spheroids (plates and rods) an apparent distribution is obtained, Equation (45)

$$Q_3(x_S) = \frac{1}{\pi} \arccos S_{P,R} \tag{45}$$

where Equations (46–48) apply

$$S_P = \frac{2(x_S^2/x_a x_c) - R_{P,c} - R_{P,a}}{R_{P,c} - R_{P,a}} \tag{46}$$

$$S_R = \frac{2(x_S^2/x_c^2) - R_{R,c} - R_{R,a}}{R_{R,c} - R_{R,a}} \tag{47}$$

$$x_a = 2a;\ x_c = 2c \tag{48}$$

If x_S is related to the median $x_{S,50}$ of the distribution, Equations (49) and (50) follows

$$Q_3\left(\frac{x_S}{x_{S,50}}\right) = \frac{1}{\pi} \arccos T_{P,R} \tag{49}$$

where

$$T_P = T_R = \frac{R_c + R_a}{R_c - R_a} \quad \frac{x_S^2}{x_{S,50}^2} - 1 \tag{50}$$

where R_a and R_c are according to Equations (25–28).

This distribution is graphically represented in Figure 6 for some values of $p = x_c/x_a$, starting with the smallest particle size $x_{S,c}$ (sedimentation in c-direction; $\alpha = 0°$) and ending with a maximum $x_{S,a}$ (sedimentation in a-direction; $\alpha = 90°$).

For thin plates ($p < 0.1$) the width of the apparent distribution can, to a rough approximation, be assumed to be independent of p for all practical cases. For thin rods (needles, fibers) the width of the distribution becomes larger and larger with decreasing p.

Which effect arises in the case of a polydisperse particle system is of practical interest, meaning for example that x_a is variable, and p is constant. Then the apparent distribution is, Equation (51)

$$Q_3(x_a, x_S) = \frac{1}{\pi} \int_{x_{a,min}}^{x_{a,max}} q_3(x_a) \arccos S_{P,R}\, dx_a \tag{51}$$

With Equation (51) we are able to estimate the modifications to apparent distributions determined by sedimentation analysis as a function of the shape factor p and of the distribution of one of the main axes. In Figure 7 the results of a calculation for plates are given where Equation (51) was numerically integrated using the Gates–Gaudin–Schumann frequency distribution, Equation (52)

$$q_3(x_a) = \frac{m}{x_{max}} \left(\frac{x_a}{x_{max}}\right)^{m-1} \tag{52}$$

It can be seen that the distribution parameter m (slope of the distribution of x_a) is maintained in practice for the distributions of x_S and that with decreasing p the distributions shift to smaller particle sizes, x_S, by a nearly constant factor. For plates this factor is approximately $1.21 \times p^{1/2}$. This behavior means that it is possible to calculate a mean shape factor from a measured apparent size distribution $Q_3(x_S)$ (by means of sedimentation analysis) and another measured distribution of one main axis (e.g. $Q_3(x_a)$ by means of image analysis).

The principle shown is also applicable for other kinds of polydispersity (see Bernhardt[13]).

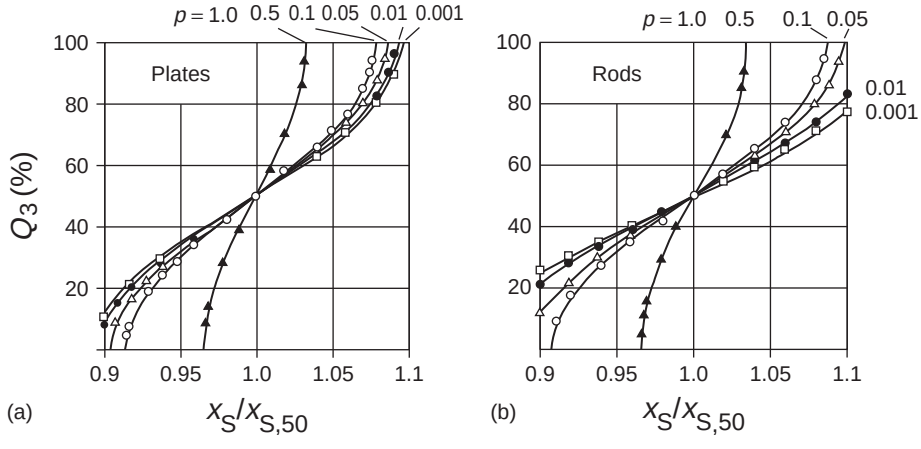

Figure 6 Apparent size distribution by shape influence (monodisperse: x_a or x_c = constant; constant particle shape: p = constant). (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

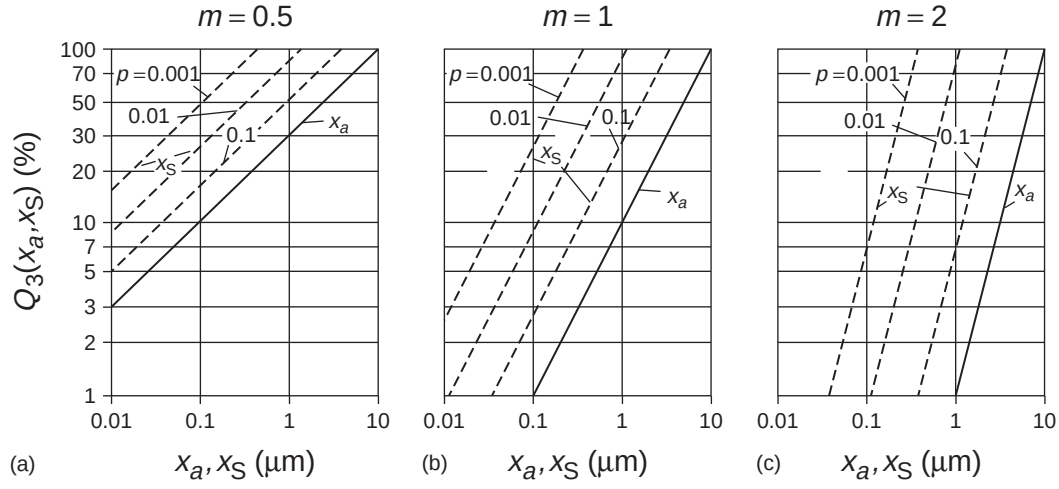

Figure 7 Apparent distributions of polydisperse plates. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

2.2.3 Brownian Motion

It is well known that the fluid molecules of the dispersion medium are subject to thermal motion whose kinetic energy is transferred to the suspended particles. The impact processes are of a stochastic nature; their amounts and directions do not compensate each other at the particle surface at any given time. This is why the particles themselves are subject to stochastic motion, the so-called Brownian motion, which is reflected both by shifts in the three spatial directions and rotation. The macroscopic appearance of this is described by the concepts of translation and rotation diffusion.

In this section, we shall attempt to assess the effects of translating diffusion taking place during sedimentation on the results of the incremental sedimentation analysis. The rotation diffusion will be neglected since we are also restricting ourselves to spherical particles for which the effects of this type of diffusion on the sedimentation movement can be neglected.[21]

In mathematical terms, the problem is equivalent to the solution of the well-known differential Equation (53)

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial h^2} - U\frac{\partial c}{\partial h} \tag{53}$$

with the diffusion coefficient, Equation (54)

$$D = \frac{kT}{3\pi\eta x} \tag{54}$$

where the Boltzmann constant $k = 1.38 \times 10^{-16}\,\mathrm{g\,cm\,s^{-2}\,K^{-1}}$ and c is the particle concentration at time t in depth h.

Solutions for different initial conditions have been found by Mason and Weaver[49] and later by Fürth[50,51] in the mid-1920s. New considerations were made using their results (see Bernhardt[13,31]). The equations for a vessel of finite dimensions (i.e. with a suspension level and with a bottom) are complicated and difficult to treat. Therefore, frequently the more simple case is found in the literature for an infinitely long vessel with a sharp borderline of particles at $t = 0$. Large differences between the two cases occur if the level of measurement is situated near the limitations (bottom and upper levels, respectively).[13] But if the measurement level is placed close to the middle of the vessel then the simple solution is sufficiently accurate. In this case the apparent distribution of a monodisperse system is given by, Equations (55–57)

$$Q_3(x_S) = \tfrac{1}{2} - \phi_0(z) \tag{55}$$

where

$$z = \frac{x^2 - x_S^2}{x}\; xh_M\frac{\pi}{12}\frac{\Delta\rho g}{kT}^{1/2} \tag{56}$$

$$\phi_0(z) = \frac{1}{\sqrt{2\pi}}\int_0^z \exp\left\{-\frac{\zeta^2}{2}\right\}\mathrm{d}\zeta \tag{57}$$

The larger the argument z, the smaller are the effects of the diffusion phenomena on the results of incremental sedimentation analysis, which means that a reduced influence occurs by:

- larger particle size x,
- larger sedimentation height h_M,
- higher acceleration g,
- higher density difference $\Delta\rho = \rho_s - \rho_m$,
- lower temperature T.

This is illustrated by two examples (see Figure 8).

Figure 8(a) demonstrates the extraordinary importance of the sedimentation path; in spite of a shorter measurement interval, measurements with small h_M lead to much

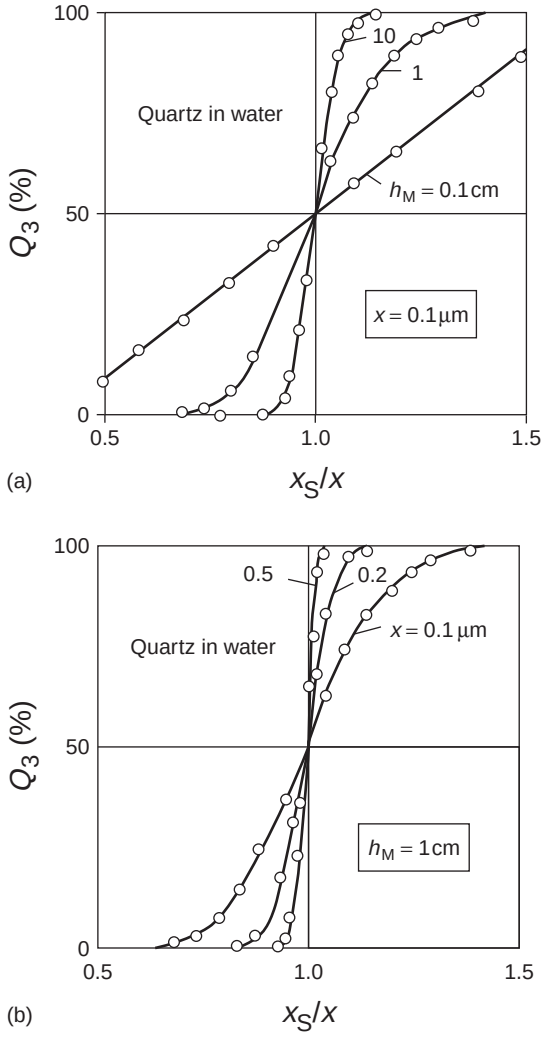

Figure 8 Apparent distributions of monodisperse particles caused by Brownian motion in the gravity field (quartz in water), according to Equation (57). (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

wider apparent distributions than are found with a large h_M. For modern sedimentation analyzers which measure very narrow distributions in the gravity field, this is a decisive restriction in use. In contrast, working in the centrifugal field is much easier. According to Equation (57), an acceleration of $b = 100$ g constant over the sedimentation path produces the same apparent distribution as a one hundred times greater height h_M or a five times greater size x in the gravity field (see also Svedberg and Rinde[52]).

For polydisperse solids with a true frequency distribution $q_3(x) = dQ_3(x)/dx$ the total concentration of all particles at depth h_M and time $t(x_S)$ is the integral of the concentrations of the individual particle sizes weighted

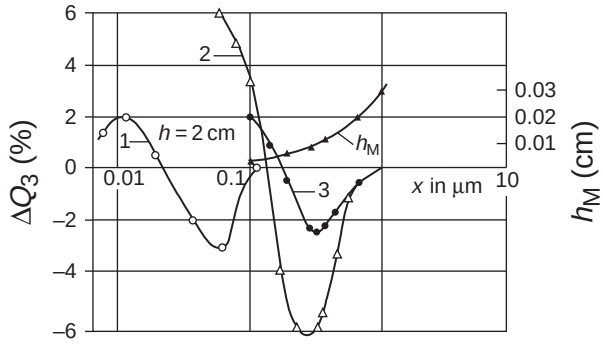

Figure 9 Measuring errors caused by diffusion/Brownian motion, 1 – according to Chung and Hogg,[54] 2, 3 – according to Allen[55] with simultaneous decrease in height h_M by scanning. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

by the frequency distribution, Equation (58)

$$Q_3(x_S) = \int_0^{x_{max}} \left(\frac{1}{2} - \phi_0(z)\right) \quad q_3(x)\,dx \qquad (58)$$

In order to estimate the deviations existing between the true and the measured distribution, it has been repeatedly proposed that one of the known distribution functions for the frequency distribution be used and that Equation (58) be integrated with it.[53–55] Several of the results are presented in Figure 9.

For the true distribution $Q_3(x)$, logarithmic normal distributions of different median x_{50} were used: 1, $x_{50} = 0.14\,\mu$m; 2, $x_{50} = 0.5\,\mu$m; 3, $x_{50} = 1\,\mu$m. In the range $x < x_{50}$ the error represented $\Delta Q_3 = Q_3(x_S) - Q_3(x)$ is positive and negative and remains small. On the whole, it has to be said, that polydispersity strongly reduces the effect of diffusion. The reason for this is that there is a certain symmetric concentration distribution of the individual particle sizes, and therefore, because of the integration the effect is only remarkable in so far as the symmetry is incomplete. This also explains why shifts occur in both the fine and the coarse range.

From this point of view, incremental measurements in the gravitational field down to 0.1 µm and in the centrifugal field down to 0.01 µm are the limits down to which measurements of polydisperse solids can be performed with an admissible error.

Further details about the diffusion effect in cumulative suspension methods and in incremental line start methods are summarized by Bernhardt[13] and Mason and Weaver.[49]

2.2.4 Geometry of the Sedimentation Vessel

The influence of the vessel geometry is due to the fact that the particles sediment more slowly in the neighborhood

of the suspension boundaries (walls, bottom, lid) than at an infinite distance from them.

2.2.4.1 Influence of Walls According to several authors[56–59] Equation (59) holds for a sphere situated in a cylindrical sedimentation vessel near the wall:

$$\frac{U}{U_0} = 1 - \frac{9}{32}\frac{x}{l} \tag{59}$$

where U_0 is the undisturbed sedimentation velocity and l is the distance of the sphere center from the wall.

The same relationship was derived by Lorentz[60] for a sphere sedimenting in parallel to a plane wall. The most modern sedimentation devices are equipped with measuring cells which have two parallel walls a small distance apart (some millimeters). Therefore, the theory of Faxén [cited in Happel and Brenner[17]] for a sphere between two parallel walls is of particular interest. An useful approximation developed by Oseen [cited in Happel and Brenner[17]] for a random position of the sphere is given by Equation (60)

$$\frac{U}{U_0} = 1 - \frac{9}{32}\frac{x}{l_1}\frac{1}{1 - l_1/L} \tag{60}$$

where L is the distance between the walls and l_1 is the distance of the particle center from one wall.

The question arises how the influence of the wall affects the measured result of the sedimentation analysis. Figure 10 shows the change in sedimentation velocity over the relative distance from the wall. For monodisperse material subjected to an incremental method, the curves also characterize the position at which the solid concentration jumps from 0 to 1.

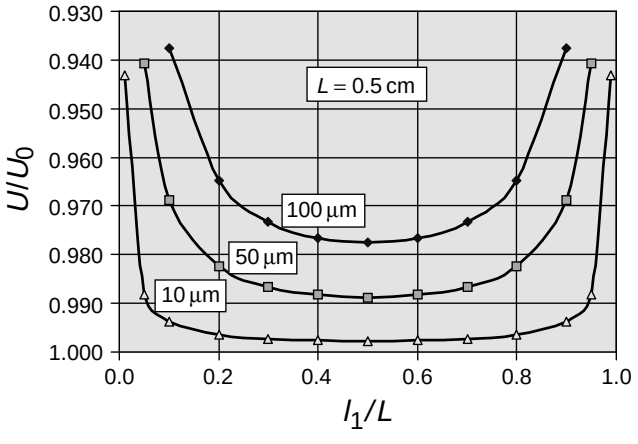

Figure 10 Change in the sedimentation velocity by the influence of the wall for different particle sizes and a wall distance of 0.5 cm.

This concentration distribution leads to an apparent distribution Equation (61):[13]

$$Q_3(x_S) = 1 - \left(1 - \frac{9}{8L}\frac{x^3}{x^2 - x_S^2}\right)^{1/2} \tag{61}$$

In the case of polydisperse materials, Equation (61) has to be integrated for all x. Taking into account that $c/c_0 = 1$ is already true for particles $x_P < x_M$ (see Bernhardt[13]) we obtain

$$Q_3(x_S) = 1 - \int_{x_P}^{x_{max}} \left(1 - \frac{9}{8L}\frac{x^3}{x^2 - x_S^2}\right)^{1/2} q_3(x)\,dx \tag{62}$$

The numerical integration of Equation (62) and the application of a Gates–Gaudin–Schumann distribution (Equation 52) leads to differences with respect to the true distribution, shown in Figure 11.

The difference $\Delta Q_3 = Q_3(x_S) - Q_3(x)$ usually reaches a measurement error of approx. 2% with sedimentation analyses only for narrow particle size distributions ($m \approx 2$), and even then only in the narrow and steep range close to the rear end of the distribution.

2.2.4.2 Influence of Horizontal Limitations Horizontal limitations of suspensions can be either closed/fixed (lid, bottom) or open (free surface). Theories comprising both cases were developed by Brenner, Wakiya and Faxén [cited in Happel and Brenner[17]]. For practical

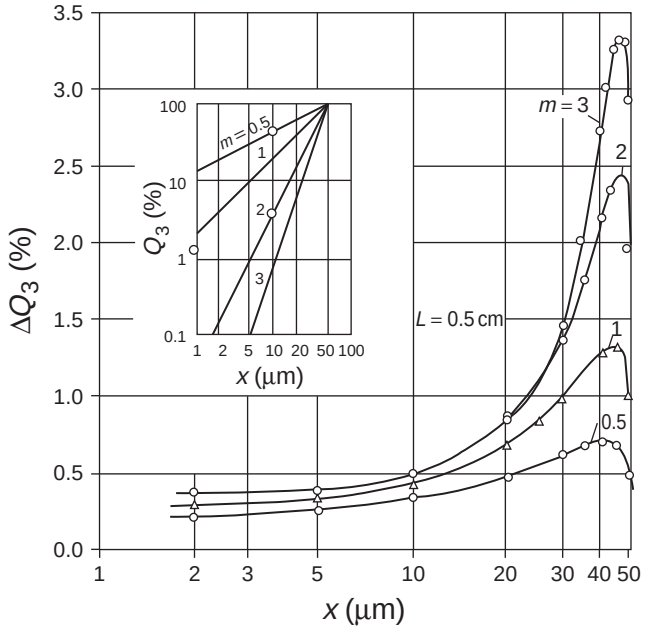

Figure 11 Errors due to influence of wall (for polydisperse material, wall distance 0.5 cm). (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

purposes ($x/h < 0.8$), the half-empirical approximations can be employed, Equations (63) and (64):
for a closed/fixed limitation:

$$\frac{U}{U_0} = 1 - 0.55\frac{x}{h} \tag{63}$$

for an open limitation (suspension level):

$$\frac{U}{U_0} = 1 - 0.375\frac{x}{h} \tag{64}$$

where h is the distance of the particle from the limitation.

The changes remain very small and can be neglected for practical purposes, because like other modern instruments no values of $x/h > 0.01$ are reached. The same is true for the apparent distribution measured by incremental analysis of a monodisperse material in a closed sedimentation vessel, Equation (65):

$$Q_3(x_S) = \frac{0.25x/h}{1 + 0.25x/h - x_S^2/x^2} \tag{65}$$

2.2.5 Other Parameters of Influence

2.2.5.1 Concentration of Solids For technical reasons, the sedimentation methods require suspensions, whose solids concentration must not fall below a certain minimum value. The influence on the sedimentation velocity is an extraordinarily complex problem essentially based on the following effects:

1. The density and viscosity of the suspension are greater than those of the pure liquid (not demonstrated here).
2. For reasons of continuity, the downward volume flow of solid particles is met by a liquid flow of the same magnitude (back flow).
3. In the vicinity of the sedimenting particle, the velocity field is disturbed leading to an increased exchange of momentum (cluster turbulence).
4. The drag force on the individual particle is reduced in the Stokes range by hydrodynamic interaction with neighboring particles.
5. By physicochemical interactions, the solvate sheath carried by the particle (change of effective particle density and suspension viscosity) and the coagulation during collisions (suspension stability) are influenced decisively.

Table 7 gives a survey of the necessary solids concentration for several sedimentation methods.

The relative particle distances h/x (h, the distance between the surfaces of the particles which are assumed to be spherical) were estimated by, Equation (66):

$$\frac{h}{x} = \left(\frac{0.68}{c_V}\right)^{1/3} - 1 \tag{66}$$

Table 7 Essential volume concentrations of solids and estimated particle distances of several sedimentation methods

Method	Essential solids concentration c_V (%)	Particle/particle distance h/x random arrangement	Particle/particle distance h_{12}/x median of distribution
Light-optical	0.005	22.9	11.0
Sedimentation balance	0.1	7.8	3.5
Pipette	0.5	4.1	1.7
X-ray absorption	3	1.8	0.3

However, the mean distance h is not a useful characteristic because in reality there is a distribution in the distances, and the smallest one is decisive for the suspension property. With solids concentrations c_V that are not too great, the number distribution of h_{12} (the distance between the surface of a particle and that of its nearest neighbor) is given by, Equation (67).[61,62]

$$Q_0\left(\frac{h_{12}}{x}\right) = 1 - \exp\left\{-8c_V\left[\left(\frac{h_{12}}{x} + 1\right)^3 - 1\right]\right\} \tag{67}$$

Some examples of the distribution are represented in Figure 12.

The median of this distribution is also given in Table 7. The very small distances below the median for larger solids concentration are the reason for most of the effects described in the following.

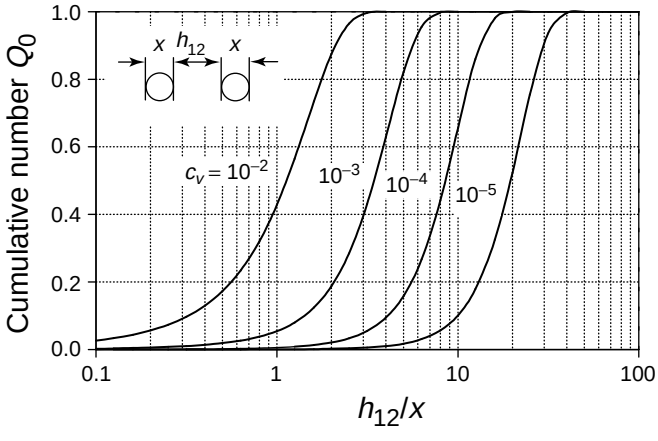

Figure 12 Number distribution of the distance h_{12}; parameter: volume concentration of solid.

Back Flow. Owing to the fluid flow counteracting the downward particle flow, the sedimentation velocity of particles of the same size reduced by the back flow velocity is, Equation (68)[63]

$$\frac{U}{U_0} = \frac{(1 - c_V)^2}{\eta/\eta_0} \tag{68}$$

where η_0 is the viscosity of the fluid without particles. As a result, the error in particle size $f_x = 1 - (U/U_0)^{1/2}$ has already reached approx. <7% in the concentration range $c_v < 3\%$.

Swarm Inhibition. Among all the numerous half-empirical and physically founded models describing the whole phenomenon of swarm inhibition (including back flow and swarm turbulence), only that proposed by Bauer and Thiele[64] will be mentioned here. For the case of a monodisperse solid, Equation (69) holds

$$\frac{U}{U_0} = (1 - c_V) \quad 1 + \frac{c_V}{(1 - c_V)^2} \times \quad 1 + \frac{1.05}{\sqrt{1 + (\pi/12c_V)^2} - 0.5} \Bigg\}^{-1} \tag{69}$$

Now, $f_x = 1 - (U/U_0)^{1/2}$ becomes <9% for $c_V < 3\%$. The influence is much stronger if the solid is polydisperse and if the analyzed particles are very small.

Hydrodynamic Interaction. The hydrodynamic interactions between two sedimenting spheres of the same size lead to a reduction in the drag force, for which a theory was developed by Happel and Brenner.[17]

The velocity ratio U_2/U is approximately given by Equation (70)

$$\frac{U_2}{U} = 1 + \frac{3}{8}\frac{x}{l}(1 + \cos^2\beta) = 1 + \frac{3}{8}\frac{1 + \cos^2\beta}{1 + h/x} \tag{70}$$

β is the angle between the sedimentation direction and the line connecting the sphere centers. According to Equation (70) the sedimentation velocity continuously increases with increasing solids concentration. Kaye and Boardman[65] and Johne,[66] however, have empirically found that this is true only up to a certain solids concentration, above which the swarm inhibition prevails. Johne's result, the average velocity of individual spheres related to their Stokes velocity, is shown in Figure 13.

Below $c_V = 10^{-4}$ the solids concentration is practically without influence on the particle movement. In the immediate vicinity of $c_V = 10^{-2}$, a velocity maximum forms which can be three times higher than the Stokes velocity. As shown by Koglin,[58] the phenomenon cannot be explained simply by the statistical arrangement of the particles and their hydrodynamic interaction plus swarm inhibition. In addition, the formation of complexes of particles (clusters), whose dimensions and sedimentation velocities grow with increasing sedimentation path and time, has to be assumed.[67] Furthermore, the size of these complexes is statistically distributed in the suspension so that the sedimentation velocity itself has a distribution. The mean value of this distribution (see Figure 13) strongly depends on both the particle size distribution and the particle shape distribution as well as on other physical parameters of the sedimenting solid.[58,68] The velocity maximum is lower, the wider the size distribution. Deviations from sphericity reduce the height of the maximum and simultaneously shift it towards lower solids concentrations.[1–12]

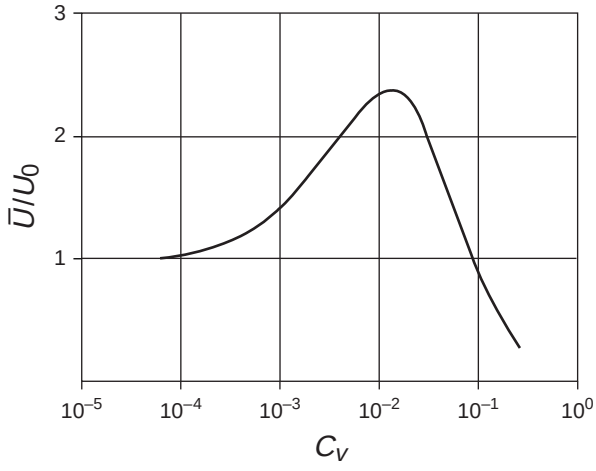

Figure 13 Empirical curve of the average sedimentation velocity of individual spheres in a suspension of monodisperse spheres. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

For light optical methods, sedimentation balances and usual pipette methods (with restrictions), the concentration of solids can normally be kept sufficiently low in practice to exclude the influence described. Other methods such as X-ray absorption reach the critical concentration range around $c_V = 0.01$ for various substances (silicates etc.). In this case, the measurement of narrowly classified material, for example, can yield a clear shift to coarser results in comparison to other methods.[13]

2.2.5.2 Convection Flows Convection flows occur whenever a suspension of higher density is situated beside or above a suspension of lower density. According to Koglin, Alex and Leschonski[1–12] the following are at least some of the reasons:

1. The sedimentation vessel does not have exactly vertical walls.

2. In the sedimentation vessel there are inserts (e.g. pipette or other bodies).
3. There are local temperature differences in the suspension caused by fast or inhomogeneous heating (e.g. solar heat radiation) or cooling (e.g. by evaporation at the suspension surface).

While the causes 1 and 2 can usually be overcome by improved design,[69] this is not so easy for cause 3. According to Rumpf[70] and Alex[71] a change of temperature of ± 1 K per analysis and 0.01 K min^{-1} should not be exceeded in a pipette device if the lower limit of measurement is 1 µm. It is difficult always to comply with the latter requirement even with good thermostats. Moreover, particularly in this field the commercially available measurement equipment very often does not meet these requirements.

Convection flows of a different type occur at the beginning of the analysis when a uniform distribution of the solid has to be achieved for suspension methods by stirring or shaking. The decay times observed with water were between 20 and 120 s dependent on the vessel geometry.

In line start methods, a suspension layer is positioned on top of a solid-free liquid at the beginning. When the solids concentration exceeds a certain value, density convection flows occur in the form of streaks, a phenomenon to which Alex et al.[1–12] have drawn to attention so vividly particularly in the case of centrifugal methods. The critical solids concentration at which this process begins is approx. 0.01 vol% according to investigations by Alex et al. This is an amount which is normally not sufficient to measure the mass proportions with satisfactory reliability (except light optical methods).

2.2.5.3 Hydrate Sheath For most particle systems it has to be assumed that a solvate (or in the special case of water, hydrate) sheath is formed, caused by the interaction between the electrically charged particle surface and the molecules of the liquid. This hydrate sheath is carried with the moving particle, with its thickness D_H depending on the equilibrium between the attraction forces and the flow forces. To estimate the influence of this phenomenon on particle motion, only the change in the effective particle density during sedimentation is taken into account. For spherical particles Equation (71) holds

$$\frac{\Delta\rho^*}{\Delta\rho} = \frac{x^3}{(x+2D_H)^3} = \left(\frac{1+2D_H}{x}\right)^{-3} \qquad (71)$$

Then, the error in the particle size is, Equation (72)

$$f_x = 1 - \left(\frac{\Delta\rho^*}{\Delta\rho}\right)^{1/2} = 1 - \left(\frac{1+2D_H}{x}\right)^{-3/2} \qquad (72)$$

The difficulty in such an estimation has to do with the fact that the exact density of the hydrate sheath ρ_H is not known. As a first approximation we have to assume that $\rho_H = \rho_m$.

For water a thickness of 3 nm equals approx. 10 molecular layers. Using this value it is evident that for particles >1 µm, the error in the particle size does not exceed 1%, and for particles > 0.2 µm does not exceed 5%. Owing to the lower density of the particles, the distribution curve $Q_3(x)$ is shifted towards greater fineness.

3 DISPERSION OF SOLIDS IN LIQUIDS

3.1 Introduction and Problems

For particle size analysis by sedimentation a suspension with a defined and constant dispersion state is an essential precondition. In many cases it is very difficult if not impossible to fulfil this precondition in a simple way, since the available information on the complex physicochemical conditions and their mutual interaction is not complete and can be extended only by extensive and partly empirical investigations. Furthermore, in many cases it is impossible to find a satisfactory answer to the question of the desirable dispersion state, which is dependent on the aim of measurement. Ideally, the following requirements would have to be met:

1. The solid particles exist in the suspension individually, i.e. free of agglomerates.
2. The dispersion state shall not be modified in any measurable way during the analysis.

Of course, the transition from separate particles to agglomerates is fluid for different solids, but even in spite of this, the analysis of real powder-type solids will always be at a greater or lesser distance from these ideal conditions.

Incomplete dispersion is often connected with a poor reproducibility of results, so that the decisive factor in sample preparation is to meet the above requirements as closely as possible and simultaneously to measure the state of the system investigated.[72]

If a dry powder is to be used for the preparation of a suspension for granulometric analysis, an agglomeration of the primary particles usually has to be expected dependent on the material properties and the conditions during manufacture and storage. The following binding mechanisms play a role:[73]

- binding by adhesion forces between the particles
- binding by wetting liquids of low viscosity
- binding by highly viscous liquids

- binding by solid bridges
- positive bridges.

Dispersion has the task of counteracting these forces by physicochemical and mechanical means to such a degree that the particles will eventually exist individually (i.e. agglomerate-free) and will not change this state in the course of the analysis. For this, certain basic mechanisms are required which can be reduced to the following three fundamental processes:

1. wetting of the solid particles by the liquid
2. desagglomeration of the particle agglomerates
3. stabilization of the suspension or prevention of renewed agglomeration.

The control of the wetting processes only allows the adhesion forces to be modified between the particles and the binding forces produced by liquids in the intermediate capillaries to be partially modified. The methods of desagglomeration should have an effect on all binding forces. Finally, most stabilization is achieved by modifying the interaction forces between the particles.

In spite of their extraordinarily wide range, the methods used to influence and control these processes can be reduced to the following tasks:

1. selection of a suitable liquid (suspending agent)
2. selection of the type and concentration of a suitable (interface active) aid (dispersant),
3. selection of the type and duration of mechanical preparation for dispersion.

In addition to its role in creating favorable conditions for wetting, desagglomeration and stabilization, the liquid medium has to meet several other requirements; these are (according to Irani and Callis[74] and Joglekar and Marathe[75]):

1. The solid must not dissolve or chemically be altered under the influence of the liquid, it must neither swell nor shrink.
2. Density and viscosity of the liquid have to be adapted to the solid density, maximum particle size and other parameters of the solid in such a way that measurement takes place at sedimentation velocities which are neither too high nor too low.
3. It has to be possible to render the liquid bubble-free.
4. The liquid should have a vapor pressure as low as possible.
5. The liquid has to be nontoxic, inexpensive and easily available.

The fundamentals of the mentioned physicochemical and physical problems (wetting, particle/particle attraction and repulsion etc.) cannot be treated here in detail. For further information the reader may refer to some well-known monographs[24,76–81] and overviews.[13,82–84]

3.2 Stabilization of Suspensions

In many cases the so-called DLVO theory (Deryaguin, Landau, Verwey and Overbeck) is a useful tool for solving dispersion problems, particularly the stabilization of suspensions. The total interaction energy between two solid particles can be written as the sum of the electrostatic repulsion, the van der Waals attraction and the steric repulsion, Equation (73)

$$V(h) = V_{el}(h) + V_A(h) + V_{Ster} \tag{73}$$

The theoretical fundamentals were developed by Deryaguin, Landau, Verwey and Overbeck; originally, the theory referred only to the combination of electrostatic and van der Waals interaction. Under these conditions they found a principal curve of $V(h)$, represented in Figure 14.

When two particles approach each other from a large distance, they proceed through the range of relatively small attraction energies called secondary minima. If these are greater than the collision energy of the particles,

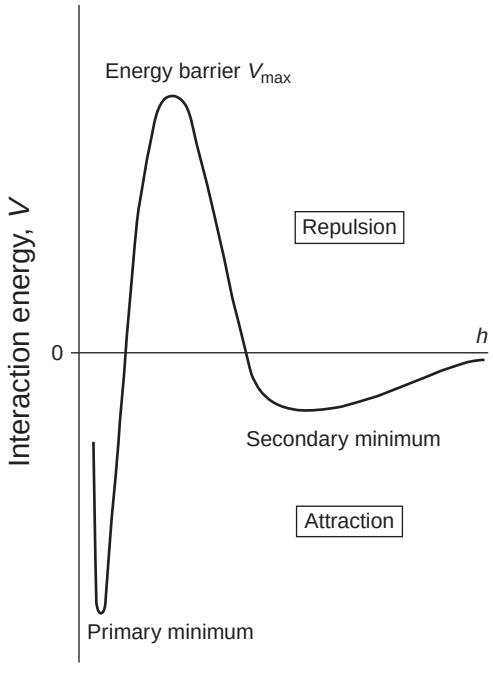

Figure 14 Total interaction energy between two particles (principal scheme) according to DLVO theory. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

agglomeration takes place then. Since it is impossible to destroy the agglomerates mechanically during sedimentation analysis, conditions have to be found under which the secondary minimum is as small as possible or disappears altogether. This is the case for high interfacial potentials or (preferably) large thicknesses of the double layer and small van der Waals attraction energies.

In these cases, the repulsion energy barrier is so high that agglomeration is possible only for very high-energy collisions of the particles. In this case, the particles adhere to each other in the primary minimum, forming relatively stable agglomerates. However, if the maximum repulsion energy is always greater than the collision energy, the suspension will be stable.

The opposite case, that is, the complete absence of repulsion energies (e.g. for small potentials and small double layer thickness) leads to agglomeration in every case of collision (fast coagulation).

Therefore, in addition to wetting and desagglomeration, the stability of the suspension over the time required for analysis is a decisive problem in the granulometric analysis of suspended particles.

Now, it has to be asked which changes in the number of particles caused by coagulation during sedimentation are within the error range of the analysis data. The results of Bernhardt[85] provide some information on this. Thus, it is fully possible for slowly coagulating systems to meet sedimentation conditions which limit the resulting error. An effective way is the use of minimum solids concentrations or sedimentation times that are as short as possible (low-sedimentation heights).

Modern devices sometimes have sedimentation heights under 0.1 mm; it can be expected that these can be used even for very unfavorable stability conditions, provided, of course, that the time for the other stages of suspension preparation can be kept short.

To give the reader a better overview, some conclusions will be made on the basis of the above statements (see also Bernhardt[13]).

Dispersion in liquids has the following objectives:

1. Spontaneous wetting as complete as possible
   - General aim:
     - high-wetting energy,
     - low-contact angle;
   - Solution:
     - low-interfacial tension liquid/gas by wetting agents,
     - low-interfacial tension solid/liquid by hydrophilizing agents.
2. High electrostatic repulsion energy
   - General aim:
     - high-interfacial potential,
     - great thickness of double layer at low-ionic strength;
   - Solution:
     - polar liquids (high-dielectric constant),
     - use of potential-controlling ions,
     - use of polyelectrolytes for adsorption in the Stern layer.
3. Low van der Waals attraction energy
   - General aim:
     - small difference in Hamaker constants of solid and liquid,
     - large particle distance.
   - Solution:
     - selection of suitable liquids,
     - masking of solid surface by adsorption layer with Hamaker constant similar to that of liquid,
     - adsorption of big molecules or small solid particles for steric repulsion,
     - low-solids concentration.
4. High-repulsion barrier and low-secondary energy minimum
   - same as 2 and 3.
5. Low probability of agglomeration at collision
   - General aim:
     - same as 2 and 3,
     - low-kinetic energy of particles;
   - Solution:
     - same as 2 and 3,
     - high viscosity of the fluid.
6. Low collision probability
   - General aim and solution:
     - low kinetic energy,
     - low solids concentration,
     - short duration of measurement.

It is obvious that the above recommendations are not a recipe for all cases; in many cases, they are not compatible with each other (this has not been described in detail), and they do not guarantee successful dispersion.

The overview shows that good dispersion depends on a very large complex of influencing factors acting together as well as against each other. For many solids, particularly if they are contaminated or heterogeneous, an empirical approach cannot be avoided. There are many examples of this in the literature. Appendix 1 gives an overview of liquids and dispersion aids for sedimentation analysis on the basis of a previous work.[13,84]

Of course, all these data are just recommendations which do not exempt the user from the obligation of informing him or herself on the effectiveness of dispersion techniques, and to test further materials and techniques.

3.3 Desagglomeration Methods

Mechanical processes have to be used in most cases of suspension pretreatment, to support or initiate the physicochemical phenomena or to destroy the agglomerates originally present in the system to allow stabilization measures to be come effective. The most important methods will be described briefly.

3.3.1 Dry Sieving

The dry powder is sieved through a test sieve of suitable mesh (e.g. 40 μm) using a brush or similar mechanical aid. The coarse agglomerates are destroyed by this treatment to a degree sufficient for the successful application of other methods (such as ultrasound) for complete desagglomeration. At the same time, the maximum particle size of the system to be analyzed is reduced to a value that can still be reliably measured. This method is preferable in comparison with comminution in a mortar or between two plates,[74] because the latter is susceptible to the danger of destroying large primary particles.

3.3.2 Treatment in Plastic State

In the method described by Allen[16] and Irani and Callis,[74] the powder sample is put on a glass plate or a very flat vessel after which a small amount of water is added, producing a plastic mixture. This is kneaded for some minutes with a suitable spatula. Then, a proper suspension is produced from this material. This method is used if coarser agglomerates must be destroyed with caution or if the material cannot be wetted with ease.

3.3.3 Deaeration of the Solid

For a very fine and/or barely wettable solid, it often makes sense to remove the included air. The following methods are recommended for the production of a sufficiently deaerated suspension:

- addition of liquid to powder in vacuum
- stirring of suspension in vacuum
- boiling of suspension
- treatment of suspension in a laboratory ball mill (without pebbles).

These measures also destroy agglomerates.

3.3.4 Treatment in a Turbulent Shear Field

The simple manual or machine shaking of a suspension in a closed vessel has only a slight desagglomerating effect, because the agglomerate strengths exceed the stress during shaking in most cases. Another possible treatment is stirring. Its use is recommended primarily for cases where wetting is difficult; however, it also contributes to agglomerate destruction to a limited extent.

In this context, the investigations by Nagel and Kürten[86] are of special interest; here, the agglomerate strengths estimated by Rumpf and Herrmann[73] are compared with stress calculated for the shear field. For possible technical values of power dissipation density between 10^3 and 10^4 kW m^{-3}, only particle agglomerates >10 μm can be destroyed if there is only van der Waals binding; in the case of sinter bridges the particles must be >100 μm. This was confirmed experimentally[86] and is in agreement with practical experience.[87]

Finally, another effective method of desagglomeration is the turbulent flow of an annular shear slot. The advantage of this method is the possibility of estimating the forces attacking the particle agglomerates in good approximation.[88,89]

3.3.5 Ultrasonic Treatment

In view of the limited desagglomeration by stirring for small agglomerates, treatment in the ultrasonic field is of special importance for suspension preparation for granulometric analysis. The mechanism of the processes involved has not yet been completely clarified; however, there is general agreement that cavitation is a prerequisite for desagglomeration.

In the ultrasonic field, cavitation is provoked by tensile stresses produced by the acoustical pressure. At first, small cavities (cavitation bubbles) form in the liquid when the so-called cavitation threshold is reached, which grow to a maximum size under oscillations and then collapse very rapidly. In real liquids containing solid particles the cavitation produces complex changes in diverse physical properties which lead to the damping of the ultrasound in the cavitation zone (see also Šutilov[90]) and, thus, a reduction of the cavitation

pressure. With increasing sound frequency, the sound energy producing the necessary cavitation increases. For dispersion purposes, frequencies between 20 and 40 Hz for which the necessary acoustical pressures are guaranteed as well are generally used. An increase in the external pressure and liquid temperature improves the effectiveness of ultrasound.[87,90] It is often indicated that, unlike other mechanical destruction methods, ultrasound has an eroding effect,[91] that is, the primary particles are removed from the agglomerates starting at the surface.

Practical ultrasonic treatment of a suspension for granulometric analysis is usually carried out using two principles:

Principle (a) uses a commercial ultrasound washing device. The suspension vessel is put or suspended in a water bath and irradiated by an ultrasound source fastened to the bottom of the bath. Because of the low-ultrasound intensity, the treatment usually requires a long time to attain the required state of dispersion. In spite of this, the method is widely used in view of its simplicity.[13]

In principle (b), the suspension in the vessel is irradiated directly by a sonotrode. Modern commercial devices usually have a fixed frequency; their intensity can be selected (specific power up to approx. 500 W cm^{-2}). This allows the duration of treatment to be reduced to a few minutes. Intensities between 100 and 200 W cm^{-2} are usually sufficient.

The dispersing effect of ultrasound decreases with increasing solids concentration, it can be increased in the presence of surfactants. In the most unfavorable cases, even coagulation can occur.[92–94]

In some studies the combination of ultrasound with another type of mechanical pretreatment reduces the essential time or provides a higher degree of dispersion.[45,95] Davies et al.[87] found that ultrasound is most effective for the destruction of small agglomerates while high-speed stirring is best for destroying large agglomerates.

A question always asked in the framework of such investigations is whether ultrasonic treatment destroys the primary particles forming the agglomerates. It is known[91] that ultrasound can destroy structures such as the solid bridges between the particles. Davies et al.[87] observed changes in the surface roughness of particles $> 5\,\mu m$.

According to Gärtner,[96] particle destruction at the weak connections also has to be expected for precipitation products whose particles consist of several primary crystallites. The newly formed primary crystallites, however, do not show a tendency towards further decomposition.

A remarkable reduction in the particle size was observed for solids with a pronounced anisotropy of cleavability; examples are chrysotile asbestos[97] or talcum.[47] Thus, it can be assumed that most substances (with the exception of those with extreme properties) are not comminuted by ultrasound.

3.4 Characterization of the Dispersion State and the Dispersion Capacity

If the task is to produce a suspension of maximum dispersion for a particle size analysis, it must first be possible to characterize the current degree of dispersion. At the same time, information is needed on possible changes in the physicochemical and mechanical methods with the aim of improving the current degree of dispersion. For this reason, apart from methods for the characterization of the dispersion degree this section will include methods in which the measured parameter has a connection with the dispersion capacity of the solid and from which further information on dispersion measures can be derived (liquid, dispersion aid, parameters of mechanical treatment).

3.4.1 Measurement of Wettability

For the measurement of the wettability the following methods are known:

- measurement of wetting heat
- measurement of adsorption isotherms
- measurement of the contact angle.

Their application to powder-type substances presupposes the availability of the necessary measurement equipment. On the other hand, extensive wetting problems are instantly noticed as an obstacle to suspension preparation that cannot be circumvented. In the case of an insufficient degree of dispersion, insufficient wetting should always be looked upon as a possible reason.

3.4.2 Rheologic Tests

Flow characteristics and dispersion degree are connected in a way that in most cases is specific to the given material. In general, the flowability is best at maximum dispersion. The exact determination of the rheologic characteristics requires measurement equipment (such as rotation viscosimeters) which is not always available. Therefore, such measurements will be done only in special cases; Joglekar and Marathe[75] recommend them for sedimentation analysis of clay suspensions.

In the case of a fast method also mentioned by Joglekar and Marathe[75] a small amount of suspension is placed on an inclined glass plate. Well-dispersed systems form a smooth uniform layer when flowing down, whereas flocculated suspensions produce a layer in stripes with a heterogeneous lumpy appearance.

Good characterization of the material behavior involves determination of the smear point (beginning of plastic behavior) and the flow point which can be done by relatively simple means.[16,74,75] For this, the dry powder is put into a small vessel (a few grams), and the liquid is added slowly from a burette while the powder is stirred with a spatula or glass rod. The amount of liquid consumed up to the smear point and the flow point is read. The smaller the difference between these two values, the better the dispersion capacity.

3.4.3 Microscopic Analysis

A simple microscopic assessment normally provides little more than qualitative information on the type of liquid and additive and the effect of the mechanical pretreatment.[74,75] A droplet from the ready suspension is brought to the microscope slide. This provides an immediate impression of the analyzed system, its dispersion state and its tendency towards coagulation.

3.4.4 Sedimentation Test

Sedimentation tests in suitable standard laboratory vessels (test tubes, small gas jars) provide a simple means of assessing the dispersion state prior to the proper analysis.[74,75] The suspension is prepared with the same solids concentration as for the analysis; only the type and concentration of the dispersion aid is normally varied. All vessels are installed for the observation of the sedimentation process after uniform pretreatment of the suspension. Good dispersion is reflected by stronger turbidity after a certain time. Even small differences can be recognized by comparison with the neighboring vessels.

After a sufficient time it is possible to determine the volume of solids sedimented on the bottom of the suspension vessel. This sediment volume is smaller, the higher the degree of dispersion. Agglomerated systems form loose sediments with a high content of liquid equivalent to a large sediment volume.

For example a silicon nitride suspension (see Figure 15) shows the dependence of the sediment volume V_S on the pH-value of the suspension compared with electrophoretically determined ζ-potentials. As expected, high ζ-potentials lead to small sediment volumes and low ζ-potentials lead to high sediment volumes. Therefore, granulometric analysis without problems is expected either for high or low pH values, since these are equivalent to maximum desagglomeration and maximum suspension stability. For a further improvement of this method see Hirosue et al.[98]

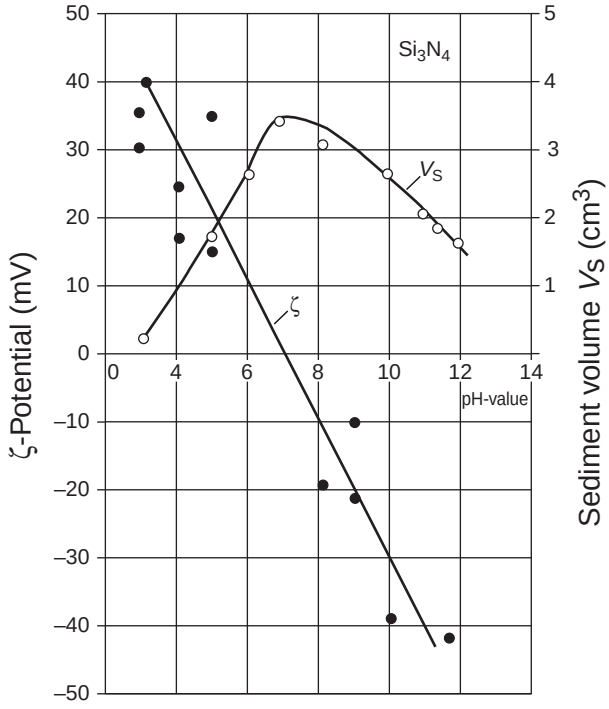

Figure 15 Sediment volume and ζ-potential of Si_3N_4 suspensions. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

3.4.5 Photometric Measurement

According to several authors,[16,45,75,92,95,99] photometric turbidity measurement is carried out dependent on time on a suspension kept in motion (without sedimentation) and with simultaneous mechanical treatment (ultrasound, stirring) and variation of the type and concentration of the dispersion aid. Koglin[95] recommends an optical small-angle scheme measuring the attenuation of white light traversing the suspension in a forward direction. If the Lambert–Beer law is valid he proposed that the agglomeration degree β of a suspension be defined as follows, Equation (74):

$$\beta = 1 - \frac{\ln T}{\ln T_{\min}} = \frac{\ln(I_{\min}/I)}{\ln(I_{\min}/I_{\infty})} \qquad (74)$$

where I is the light intensity, I_{∞} is I for a solid free liquid, $I_{\min}$ is I for an agglomerated system.

For the interpretation of the effect of measurement, the fact that both the particle size distribution and the value of the extinction function $K(x)$ (see section 4.2.1) change must be taken into account. Koglin[95] therefore restricts the use of the method to particles $x > 1\,\mu m$. However, his investigations have shown that the influence of all means of dispersion are very well reflected, and procedures for the further treatment of the suspension can be selected very well.

For very high solids concentrations ($c_V > 1\%$) it is possible that measurable transmissions are achieved only for very thin cuvets. In this case, however, the scattered light intensity occurring under a certain angle can be measured and its changes can be followed. This can be done easily for some commercial types of equipment.

Another version of this method is the approach described by Joglekar and Marathe[75] and Süß and Hantke[92] according to which the photoextinction method is employed for surface determination. By comparison with surface values determined by other methods, conclusions are made with respect to the degree of dispersion.

3.4.6 Zeta Potential Measurement

A zeta potential measurement provides information on the charging state at the solid/liquid phase boundary. If the ζ-potential and its dependence on pH or other charge generating ions are known then in most cases it is possible to find conditions with maximum repulsion energies between the particles. The first application of ζ-potential measurements for the purposes of particle size analysis is described by Bernhardt.[85]

Experimental determination of the ζ-potential requires special measurement equipment which cannot be described here in detail (see Hiemenz[77] and Stauff[24]). Comprehensive measurements of a wide range of materials are documented by Hunter[78] and Ney[79] which are available if needed. In such a case, however, it appears to make sense to analyze the dispersion conditions of the material system in question at least on the basis of a second independent quantity (see also Figure 15).

3.4.7 Particle Size Analysis

When investigating the influence of the degree of agglomeration on the granulometric result it is recommended that the granulometric method itself be used for the preliminary assessment and to change the conditions of pretreatment until the results of measurement remain constant. However, this requires sufficiently fast measurement, which only modern automatic devices are capable of in sedimentation analysis. On the other hand, it is of great advantage to have available the complete particle size distribution for a comprehensive assessment of the agglomeration state.

In the literature there are two main characteristics for this:

1. degree of agglomeration α (=1 for complete agglomeration, =0 for complete dispersion)
2. degree of dispersion/desagglomeration β (=1 for complete desagglomeration, =0 for complete agglomeration).

The essential problem of all the different degrees of agglomeration is that the extreme points (0 and 1) are unknown. Instead we have to use the state in which we start the dispersion experiment or in which we finish it. Table 8 gives some examples of definitions for α and β, in which only ''complete dispersion'' is used (characterized by the primary particle).

Which characteristic is suitable for a certain purpose must be decided by the user in each case.

3.4.8 Comparison of Methods for Dispersion Control

Table 9 gives an overview of all the above statements about the suitability of the described methods for dispersion control. It must not be forgotten that the classification as suitable, conditional or barely suitable and not suitable is subjective and may depend on the accuracy of the information required.

Nonetheless, the overview provides the correct impression that in the normal case only a few methods are suitable for the solution of all four problems of suspension

Table 8 Degrees of desagglomeration and agglomeration

Basic characteristic	Degree of desagglomeration β	Degree of agglomeration α	Ref.
Median x_{50}	$\frac{x_{50,\mathrm{PP}}}{x_{50}}$		100
Mean particle size $x_\mathrm{m} = M_{1,3}$	$\frac{M_{1,3,\mathrm{PP}}}{M_{1,3}}$		
Second moment $M_{2,3}$	$\frac{M_{2,3,\mathrm{PP}}}{M_{2,3}}$		
Third moment $M_{3,3}$	$\frac{M_{3,3,\mathrm{PP}}}{M_{3,3}}$		101
Specific surface area $S_V = 6\,M_{-1,3}$	$\frac{M_{-1,3,\mathrm{PP}}}{M_{-1,3}}$		101, 95
Specific particle number by volume $M_{-3,3} = \frac{1}{M_{3,0}}$		$1 - \frac{M_{-3,3,\mathrm{PP}}}{M_{-3,3}}$	102
Arithmetic mean volume $M_{3,0}$ (third moment of number distribution)		$1 - \frac{M_{3,0,\mathrm{PP}}}{M_{3,0}}$	102
Standard deviation $s = ({}_\mathrm{m}M_{2,3})^{1/2}$ (second moment related to x_m)	$\frac{s_\mathrm{PP}}{s}$		13

$M_{k,r} = \int_0^\infty x^k q_r(x)\,\mathrm{d}x$; ${}_\mathrm{m}M_{k,r} = \int_0^\infty (x - x_\mathrm{m})^k q_r(x)\,\mathrm{d}x$; PP = primary particle.

Table 9 Methods of dispersion control

Method	Determination of			
	Type of liquid	Type of dispersant	Concentration of dispersant	Type and duration of mechanical treatment
Wettability	+	+	(+)	−
Rheologic test	+	+	(+)	−
Microscopic assessment	+	+	(+)	(+)
Sedimentation test	(+)	+	+	+
Photometric measurement	+	+	+	+
Zeta potential measurement	(+)	+	+	−
Particle size analysis	+	+	+	+

+ suitable, (+) conditional or barely suitable, − not suitable.

preparation and that there are only a few methods that allow characterization of the type and duration of mechanical pretreatment.

In conclusion, it follows that no absolute rules can be given for the use of the methods and that it makes sense to consider a combination of several methods on the basis of an assessment of costs and desired effects.

4 SEDIMENTATION METHODS

4.1 Brief Historical View

The development of sedimentation methods began with the soil analyses which have been extensively performed since the middle of the 19th century in which fractions of a certain fineness had to be determined. Because the sample material was elutriated in water, this was called elutriation analysis (Geßner, 1931[103]), a name which was used for most diverse methods for a long time.

A further line of development originates from colloid chemistry which developed rapidly at the beginning of the 20th century (Ostwald 1911[104] and von Hahn 1928[105]).

Apart from appropriate elutriation methods, cumulative methods in all their versions played a decisive role up to the first third of the 20th century. The application of Stokes' equation to sedimentation analysis was a decisive development at that time.

Decanting was abandoned at that time, and various automation or continuous methods of measurement were introduced (sedimentation balance according to Odén (1915);[106] pressure difference method according to Wiegner (1918)[107]). Centrifuges were first used around 1923 by Svedberg and Nichols[108] leading to both time reduction and an extension in the range of the finer fractions.

In the decade between 1920 and 1930, incremental (pipette) methods were invented almost simultaneously and independently by Robinson (1922),[109] Jennings et al.,[110] Krauss (1923),[111] Köhn (1928)[112] and, finally, Andreasen (1928).[113] Other incremental methods were those using areometers (Bouyoucos (1927)[114]) and later diving apparatus (Berg (1941)[115]).

After World War II, incremental methods with contact-free measurement of reasonable quantities by the attenuation of light or γ-radiation were used, and centrifuges were employed to this end. Only this allowed a drastic reduction of the sedimentation height down to fractions of 1 mm which also decreased the time requirements, so that sedimentation analysis remained an alternative and, in many cases, an indispensable granulometric method.

4.2 Gravitational Methods

4.2.1 Incremental Suspension Methods

4.2.1.1 Principle and Fundamentals For this measurement principle, there is a homogeneously mixed suspension at the beginning of the analysis ($t = 0$) so that the overall concentration c_0 prevails in every part of the sedimentation vessel. In Figure 16(a) the initial situation is represented so that the individual particle sizes are arranged beside each other.

After time $t > 0$, there will be a state which is described by Figure 16(b). One portion of the particles has sedimented on the bottom – more large ones than small ones in view of the former's higher sedimentation velocity. Thus, the upper concentration limits of the particles still sedimenting are at different heights in the sedimentation vessel in the order of their size. Size x has only just reached height h. Here, in the ideal case we measure the solids concentration $c(h, t)$ in a differential

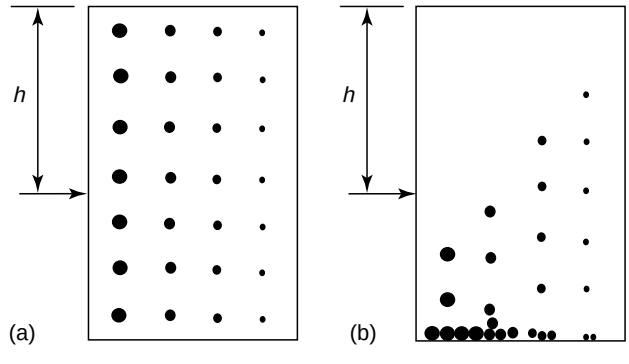

Figure 16 Principle of incremental suspension methods. (a) $t = 0$, (b) $t > 0$.

thin layer after time t. For this, Equation (75) evidently holds

$$c(h,t) = c_0 \int_0^{x_t} q_3(x)\,\mathrm{d}x \tag{75}$$

from which follows Equation (76)

$$Q_3(x_t) = \frac{c(h,t)}{c_0} \tag{76}$$

This is the fundamental relationship on which the method is based.

Deviations from the ideal case were described in section 2.2 as far as they influence the settling velocity. Moreover two further problems appear:

1. the finite extension of the measuring gap
2. the heterogeneity of the solid material.

The simple Equation (76) is valid only if the solids concentration is measured as a function of time in a measuring plane. In practice, however, there is always a measuring volume, since both the entrance slit and the exit slit of the measuring radiation have a finite extension Δh (see also Figures 4 and 16). The apparent distribution of a monodisperse material can be calculated by Equation (77):[13]

$$Q_3(x_S) = \frac{1}{2} - \frac{h_M}{\Delta h}\left(\frac{x^2}{x_S^2} - 1\right. \tag{77}$$

and those of polydisperse material by Equation (78)

$$Q_3(x_S) = \int_{x_{t,\min}}^{x_{t,\max}} \frac{1}{2} - \frac{h_M}{\Delta h}\left(\frac{x^2}{x_S^2} - 1\right. \quad q_3(x)\,\mathrm{d}x + \int_0^{x_{t,\min}} q_3(x)\,\mathrm{d}x \tag{78}$$

with the integration limits: $x_{t,\max} = x_S(1 + \Delta h/2)^{1/2}$ and $x_{t,\min} = x_S(1 - \Delta h/2)^{1/2}$.

Using a Gates–Gaudin–Schumann distribution for $q_3(x)$ (Equation 52) it is possible to integrate Equation (78) analytically and calculate the differences between the true distribution and the apparent one.[13] In Figure 17 some examples of the error, Equation (79)

$$f(x_S) = \frac{Q_3(x) - Q_3(x_S)}{Q_3(x)} \text{ in \%} \tag{79}$$

are represented for different distribution parameters m.

The error attains values considerably greater than 1% only for very narrow distributions ($m \geq 3$), and even this only if $\Delta h/h_M$ is close to 1. Usually, $\Delta h/h_M = 0.2$ is not exceeded even by modern sedimentation devices. In this case the error can be neglected as a rule. The same is valid for pipette analysis, where a sink flow occurs at the pipette tip.[1–12]

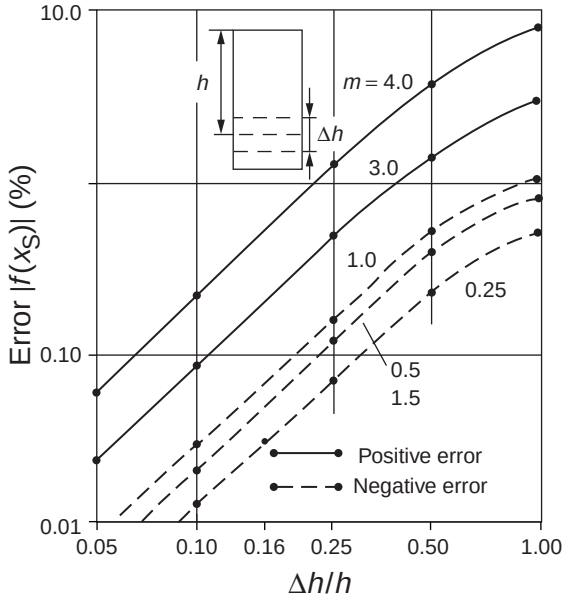

Figure 17 Influence of the measuring gap width on the measured results. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

Further, it is assumed that the solid is a uniform system so that two particles differ only in size and no other property is relevant for measurement. This is not always so in practice. Frequently, the system consists of two or more components which may differ in some of the following properties affecting the sedimentation analysis:

1. particle size distribution $Q(x)$
2. particle shape distribution $Q(\xi)$
3. density ρ_s of the solid
4. refractive index m of the solid
5. mass attenuation coefficient μ of the solid.

Bernhardt[116] has investigated the problems resulting for incremental suspension methods. He found that a distinction must be made between properties affecting the sedimentation velocity and thus the particle sizes present on measurement level (2 and 3) and properties affecting quantity measurement (4 and 5).

The difference between components with respect to particle size distribution is insignificant and they do not produce incorrect measurements. In practice, however, components normally differ in all three properties 1 to 3.

For two-component systems in which the components with mass fractions p_1 and p_2 differ only in size distribution ($Q_{3,1}(x)$ and $Q_{3,2}(x)$) and solid density (ρ_1 and ρ_2), Equation (80) holds

$$\Delta Q_3 = Q_{3,\mathrm{mea}}(x) - Q_{3,\mathrm{act}}(x) = p_1[Q_{3,1}(x_\mathrm{M} f_1) - Q_{3,1}(x_\mathrm{M})] + p_2[Q_{3,2}(x_\mathrm{M} f_2) - Q_{3,2}(x_\mathrm{M})] \quad (80)$$

where x_M is the particle size, calculated using the mean solid density, $Q_{3,\mathrm{mea}}(x)$ is the measured (apparent) particle size distribution, $Q_{3,\mathrm{act}}(x)$ is actual particle size distribution and, Equation (81)

$$f_1 = \left[\frac{\rho_1\rho_2 - \rho_\mathrm{m}(\rho_1 p_2 + \rho_2 p_1)}{(\rho_1 - \rho_\mathrm{m})(\rho_1 p_2 + \rho_2 p_1)}\right]^{1/2}$$
$$f_2 = \left[\frac{\rho_1\rho_2 - \rho_\mathrm{m}(\rho_1 p_2 + \rho_2 p_1)}{(\rho_2 - \rho_\mathrm{m})(\rho_1 p_2 + \rho_2 p_1)}\right]^{1/2} \quad (81)$$

In rough approximation, Equation (82)

$$\Delta Q_3 \approx p_1(f_1 - 1)x_\mathrm{M}[q_{3,1}(x_\mathrm{M}) - q_{3,2}(x_\mathrm{M})] \quad (82)$$

where $q_{3,1}(x_\mathrm{M})$, $q_{3,2}(x_\mathrm{M})$ are the frequency distributions of the two components at x_M.

An analysis of this relationship shows:[116]

- Because the product $p_1(f_1 - 1)$ assumes an extreme value in the vicinity of $p_2 = 0.5$, large errors in Q_3 have to be expected here; the extreme value itself is determined by the density difference of the components.
- The error in Q_3 furthermore depends on the particle size x and the difference of the frequency distributions of the two components at this position.

It is mostly the latter factor which makes the error relatively small even for large density differences. The examples of mixtures of SiO_2 ($\rho_1 = 2.65\,\mathrm{g\,cm^{-3}}$) and Fe_2O_3 ($\rho_2 = 5.55\,\mathrm{g\,cm^{-2}}$) calculated by Bernhardt[116] have errors $\Delta Q_3 < 3\%$.

4.2.1.2 Gravimetric Pipette Methods In pipette analysis, the solids concentration at the measurement level (depth h) is determined by taking a sample of volume V (after time t) and subsequently determining the solid mass m by weighing (after drying). Then, Equation (83)

$$Q_3(x) = \frac{c(h,t)}{c_0} = \frac{m(t)}{m_0}\frac{V_0}{V(t)} \quad (83)$$

Pipette methods can be subdivided into the following groups:

- methods introducing the pipette from the top
  - pipette with fixed immersion depth
  - pipette with variable immersion depth
- methods with the pipette introduced from below
- methods with lateral introduction of the pipette.

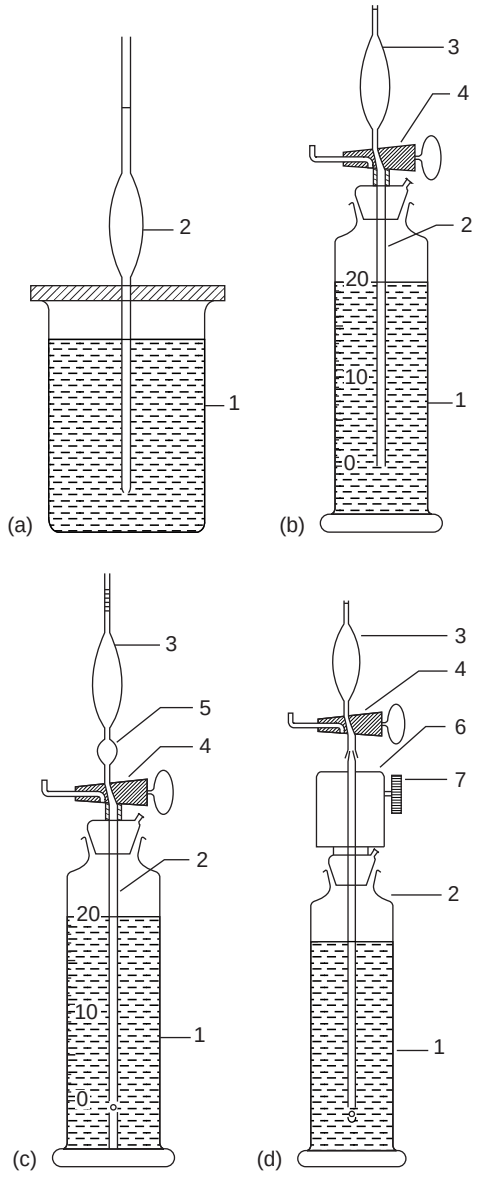

Figure 18 Pipette devices with pipette introduced from the top. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

Figure 18 shows some examples of the methods involving introducing the pipette from the top. Figure 18(a) reflects the arrangement proposed by Robinson[109] as early as 1922 in which normal laboratory devices are used as sedimentation vessel (1) and pipette (2). After each sampling, the pipette had to be removed from the suspension and then had to be reintroduced. This disadvantage was overcome by a development by Andreasen[113] whose improved version[1–12] is represented in Figure 18(b). Between pipette tip (2) and sample vessel (3) a special two-way stopcock (4) is installed allowing the transfer of the sample volume ($10\,\mathrm{cm^3}$) to an evaporating and

weighing pan. Later developments avoided pipettes which were open at the tip and these were replaced by closed pipettes with lateral suction openings or annular slit-type arrangements.

Error analyses of the scheme in Figure 18(b) were done by Rumpf,[70] Kast[69] and especially Leschonski[1–12] on the basis of which the device in Figure 18(c) was developed. The pipette (2) now extends to the bottom of the sedimentation vessel (to exclude density convection flows) and has four lateral openings. To avoid falsification of the solids concentration of the following sample by the suspension residue between suction openings and stopcock (4), an equal preliminary volume is removed by means of the pipette element (5) and rejected. To avoid the necessity of slowing down the process in the case of suction close to the pipette marking or even to have to send back excess suspension to the vessel, a calibrated scale enabling the drawn volume to be read is attached to the upper end of the pipette.

A detailed description of the entire procedure (sample preparation, thermostatting, time measurement, weighing etc.) was given by Alex.[71] Rationalization measures such as automatic suction are recommended by Alex.[71] Nevertheless, pipette analysis with fixed immersion depth does not have the same importance today that it had in the decades between 1930 and 1970, the main reason being the excessively long analysis time.

In early work it was tried to limit the required time by a reduction in sedimentation height; Krauss[111] used this principle in the system he proposed in 1923. Unlike the pipette arrangement of Robinson,[109] three tubes with adjustable height, sealed bottom and lateral suction openings were used, and the suspension samples were transferred into a weighing vessel. Köhn[112] and Andreasen[113] again reduced this method to just one pipette and introduced improvements in handling, so that this principle eventually was further developed to a standard technique. A further simple system using this principle but not including guiding of the pipette by an external structure was proposed by Börner (see Figure 18d). Here, a pipette with adjustable height (2) is guided by a gear (6) consisting of rubber rolls which can be actuated by the handwheel (7) [cited by Bernhardt[13]].

Many fewer versions are known of systems which use a pipette introduced into the sedimentation vessel from below. These have the decisive advantage that the suspension sample is taken spontaneously so that no convection flows can form at the pipette tip. The method originally developed by Jennings et al.[110] does not include height reduction during analysis. Later, Dallendörfer and Langhammer,[117] Bernhardt and Lippek[118] and Zecha[119] used this principle again with the aim of reducing the sedimentation height during the analysis. As their investigations have shown, it is possible to work with heights down to 20 mm.

Sedimentation systems with a lateral pipette are mentioned here only for completeness sake; the systems developed by Esenwein, Berg and Stairmand [cited in Bernhardt[13] and Berg[53]] were only of limited or local importance.

4.2.1.3 Methods of Density Measurement The fundamental equation for the application of density measurement for incremental sedimentation analyses is given by Equation (84)

$$Q_3(x) = \frac{\varphi(h, t) - \rho_m}{\varphi_0 - \rho_m} \tag{84}$$

where the suspension density $\varphi = c_V(\rho_s - \rho_m)$.

Areometer Method. Bouyoucos[114] was the first to report on the measurement of the suspension density $\varphi(h, t)$ by means of areometers. It was a fast method much used particularly for soil analysis between 1930 and 1960.[74,120–122]

Measurement is performed in such a way that a suitable areometer is introduced into the suspension at certain intervals, and the density is read dependent on time. According to Rumpf[70] the method is cumulative in the strict sense; however, there are fundamental obstacles to an evaluation in this sense. This is why since Casagrande[123] the approach has been simplified by the concept of the areometer indicating the mean density at the depth of the center of the areometer head.

The method has grave disadvantages [see Rumpf,[70] Conley[124] and Allen[16]].

Diver Method. To overcome certain disadvantages of the areometer method, Berg[115] used miniature areometers without a shaft (divers) of accurately staggered density that remained in the sedimentation vessel during analysis. Such a body is always in the concentration range equivalent to its density. The measurement requires a certain number of divers with carefully adjusted density values. The depth of the diver is usually determined by means of a magnet drawing the divers to the glass wall of the sedimentation vessel. There are also some other designs.[16,122]

4.2.1.4 Photoelectric Methods

Fundamentals. The photoelectric methods make use of the fact that a light beam of intensity I_E is attenuated to intensity I_A when crossing a suspension. The attenuation law (Lambert–Beer law after its discoverers) is, Equation (85)

$$\ln T = \ln\left(\frac{I_A}{I_E}\right. = -A_V c_V L \tag{85}$$

A_V is the so-called scattering section related to particle volume ($=3K(x)/2x$) so that, Equation (86)

$$\ln \frac{I_A}{I_E} = \frac{3}{2}\frac{K(x)}{x} c_V L \tag{86}$$

where $K(x)$ is the extinction function. If the trans-illuminated suspension contains a polydisperse solid of frequency distribution $q_3(x)$, the corresponding use of Equation (86) leads to Equation (87):[13]

$$Q_3(x) = 1 - \frac{\int_{\ln T(\infty)}^{\ln T(t)} \frac{x}{K(x)}\, d\ln T}{\int_{\ln T(0)}^{\ln T(\infty)} \frac{x}{K(x)}\, d\ln T} \tag{87}$$

where $T = I/I_\infty$ ($I_\infty = I_A$) at time $t = \infty$ (solid-free). The difficulty in applying this relationship to sedimentation analysis has mainly to do with the extinction function $K(x)$; it depends on a total of four quantities:

- particle size x
- wavelength λ of the light used
- the optical constants of solid and liquid
- the aperture angle Θ of the optical system used.

There are three possibilities with respect to the extinction function:

- calculation from material data
- assumption of constancy
- experimental determination.

Calculation can in principle be done on the basis of the theory of Mie[125] (see also van de Hulst,[126] and Kerker[127]). Its use requires knowledge of the optical properties (ratio of refractive indexes) of solid and liquid, their absorption capacity (imaginary refraction index) and, furthermore, the well-founded assumption that the solid particles can be characterized by a spherical shape. Tables of refraction indexes and approximations of the extinction function are given in the literature (see Alex et al.,[1–12] Bernhardt[13]).

Theoretically, the extinction function must have the trend described in the following (see Figure 19). In the case of small aperture angles the function starts close to one for large particles ($\alpha = \pi x/\lambda > 100$). With decreasing particle size, its value grows to a maximum under strong oscillations and then rapidly decreases and turns into the curve of the Raleigh theory in the range $\alpha < 1$.

Practical cases often are far from these ideal conditions. Frequently, there is neither a spherical shape nor a homogeneous solid whose optical properties are known. Examples include most rocks, ores, coal, silicatic raw materials, and so on.

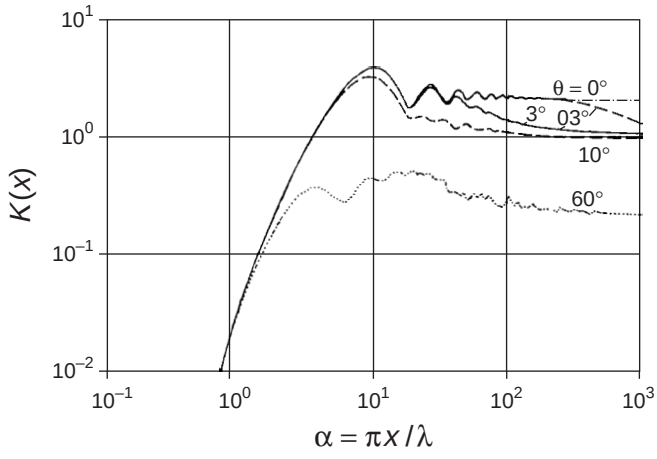

Figure 19 Extinction function dependent on Mie parameter $\alpha = \pi x/\lambda$ for different aperture angles θ; $m = 1.2$ (according to Weichert[128]). (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

The influence of the aperture angle θ on the extinction function has been studied by Weichert.[128] Some of his results are represented in Figure 19.

It is evident that:

- The aperture angle θ controls the position of the final value to which the extinction function tends for large particles;
- for a large aperture angle θ, the assumption of a constant value of the extinction function for a wide range of fineness is more justified than for small θ.

The latter fact led Allen[129] to improve the wide angle photosedimentometer according to Harner and Musgrave[130] and Telle.[131] For $K(x) =$ constant, Equation (88) is obtained from Equation (87)

$$Q_3(x) = \frac{\int_{\ln T(t)}^{\ln T(\infty)} x\, d\ln T}{\int_{\ln T(0)}^{\ln T(\infty)} x\, d\ln T} \tag{88}$$

If the direct determination of the volume distribution $Q_3(x)$ is replaced by the surface distribution $Q_2(x)$, the very simple and well known relationship, Equation (89), is obtained:

$$Q_2(x) = \frac{\int_{\ln T(t)}^{\ln T(\infty)} d\ln T}{\int_{\ln T(0)}^{\ln T(\infty)} d\ln T} = \frac{\ln(I_\infty/I_t)}{\ln(I_\infty/I_0)} \tag{89}$$

where I_0 is I_A at time $t = 0$; I_∞ is I_A at time $t = \infty$ (solid-free); I_t is I_A at time t.

When using this approach it must be taken into account that while it is possible to measure up to the given range, strictly speaking there must not be any solid fractions below the smallest size indicated; the greater the actual fines fraction, the higher the error of measurement.

The problems mentioned describe the whole approach as unsatisfactory. Therefore, there have been attempts to determine experimentally the extinction functions for given materials and devices since the introduction of this sedimentation method.[129,130,132,133] While these essentially relate to measurements of narrowly classified or monodisperse materials, Johne and Ramanujam[134] describe a method based on the studies of Locher[135] for which the frequency distribution $q_3(x)$ of a calibration material has to be known.

Weichert[128] solved the problem by measuring the transmission of a sedimenting system dependent not only on the time t (or size x) but also the wavelength λ. He was able to show that even simultaneous measurement at three different wavelengths would be sufficient if the dependence of the refractive index on the wavelength could be neglected for the selected conditions. This method has the advantage that the simple procedure of modifying the photometric sedimentation analysis simultaneously provides the information required for the calculation of the size distribution, with the material and device dependent influences being taken into account.

Photosedimentometer. Photosedimentometers are subdivided into wide angle and small angle systems. Figure 20 gives a comparison of the two principles.

In the small angle photometer (aperture angle $\Theta \leq 3^\circ$), the light emitted by a lamp (1) is formed into a parallel bundle by a lens system (2); a part is transmitted to the suspension in a measurement cuvet (4) through a diaphragm (3). By means of diaphragm (5), the arriving radiation is transmitted to the detector (6) under a small aperture angle.

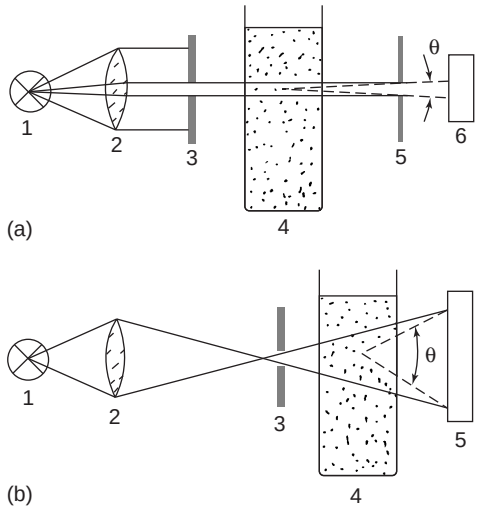

Figure 20 Photometer principles, (a) small angle photometer, (b) wide angle photometer. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

In contrast, no parallel light is used in the wide angle configuration; lamp (1), lens system (2) and diaphragm (3) are designed so that the suspension is transilluminated by a light bundle of the type shown. Since there is no diaphragm between cuvet (4) and detector (5), the aperture angle is very wide ($\Theta > 40^\circ$).

The two systems have occasionally been the subject of comparative analysis;[129,134] neither was revealed to be superior with respect to the other. This is why both systems are still in use.

Turbidity measurements for the characterization of colloid–chemical processes and conditions have been known for a long time (e.g. see von Hahn[105]). First studies on the use of this principle in sedimentation analysis for the determination of particle size distributions were done by Skinner and Boas-Traube[136] and later by Rose;[132] the latter developed a small-angle device which became known by the name of the manufacturer, Evans Electroselenium Ltd., London (EEL sedimentometer).

The simultaneous development of the wide angle technique can be traced back to investigations by Harner and Musgrave[130] and Telle,[131] the device described by the latter was built by Leitz, Wetzlar. Allen used the wide angle principle, initially without height adjustment.[129] Later he reported a new development that allowed the sedimentation vessel to be lower[16] and then one that allowed the optical system to shift upwards. The latter is known under the name of Ladal wide angle scanning photosedimentometer (WASP). A similar version, however, with downward shifting of the sedimentation vessel is distributed by Microsal, London, under the name Computerised WASP.

The Fritsch company, Idar-Oberstein, manufactures a wide angle device under the designation Scanning Fotosedimentograf (Analysette 20) in which the sedimentation vessel is fixed and the optical system is moved upwards at a constant velocity.

The small angle device Sedigraph 5500 L of Micromeritics, Norcross provides for the lowering of the measurement cuvet, albeit not at a constant velocity but following a scanning program in which the velocity diminishes with increasing sedimentation time. As a result of measurement, a surface distribution according to Equation (89) is recorded.

The spectral photosedimentometer described by Weichert[128] uses light of three different wavelengths. Both the sedimentation vessel and the small angle

optical system are at rest during measurement; however, the latter is a multiple system distributed along the sedimentation path.

A multiple optical system distributed along the sedimentation path is also used in the device called Lumosed which was developed by Staudinger et al.[137] A commercially available device from Retsch uses white light at three different sedimentation heights; a special procedure was developed for the combination of the results, which are evaluated under the assumption of a constant extinction function. At the same time, a number of values of the extinction function can be calculated by means of a calibration measurement with a substance of known distribution; these values can then be used in all measurements of the same material system.

In conclusion, another group of devices shall be mentioned which can be used in both gravitational and centrifugal fields. These are products of several important Japanese firms (Shimadzu Corp., Tokyo, Horiba, Kyoto, Seishin Enterprise Ltd., Tokyo); their main use, however, is as centrifuges. Therefore, they will be dealt with in section 4.3.

4.2.1.5 Methods Based on Gamma Attenuation

Fundamentals. Gamma radiation or X-rays of intensity I_E impinge on a suspension of thickness L, the intensity I_A after transillumination is calculated on the basis of the Lambert–Beer law, Equation (90):

$$\ln \frac{I_E}{I_A} = \mu_{m,eff} \rho_{eff} L \tag{90}$$

For a homogeneous solid in a liquid the very simple Equation (91) holds:

$$Q_3(x) = \frac{c_V(t)}{c_0} = \frac{\ln(I_\infty / I_t)}{\ln(I_\infty / I_0)} \tag{91}$$

For two solids (with different particle size distributions, different densities and mass attenuation coefficients) the measured distribution is produced by a more complicated relationship which was developed by Bernhardt.[116] He found that the differences between the measured and the true particle size distributions are only considerable for large mass portions near the maximum of the frequency distribution. The theory was confirmed experimentally for the measurement of mixtures of SiO_2 and Fe_2O_3 by using a sedigraph 5000 D (see Figure 21).

In general, it can be said that the particle size distribution of the components essentially determines the shape of the error curve and the solid mass proportions essentially control the value of the errors. It is difficult to give a general assessment of the effects of the other parameters, density, attenuation coefficient and solids concentration.

Gamma-ray and X-ray Sedimentometer. The first to propose the use of X-rays for incremental sedimentation analysis were Brown et al.[138] Their commercial system (ICI Stevenston, Scotland) is described in Allen.[16] Later improvements used different X-ray sources[139–142] and radionuclides.[143,144] Rumpf et al.[68] deal with the criteria of selection for different types and energies of sources as well as the required activity. Among more recent developments, two deserve to be mentioned: the LADAL X-ray sedimentometer[16,144] has a fixed

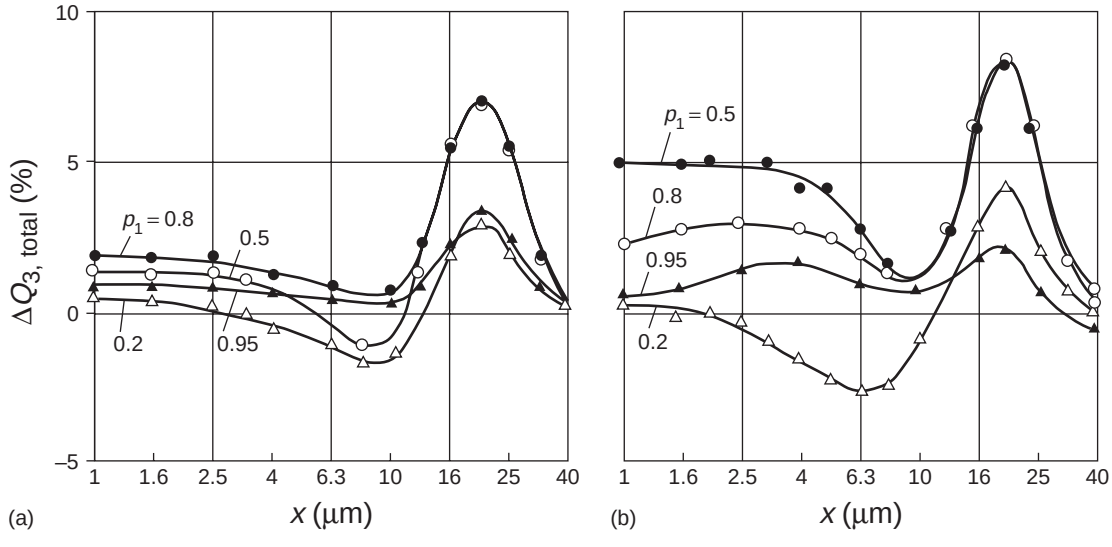

Figure 21 Total systematic error due to the density and attenuation coefficient of heterogeneous solids ($SiO_2 + Fe_2O_3$) (p_1 mass proportion of SiO_2): (a) calculated; (b) experimental (Sedigraph 5000 D). (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

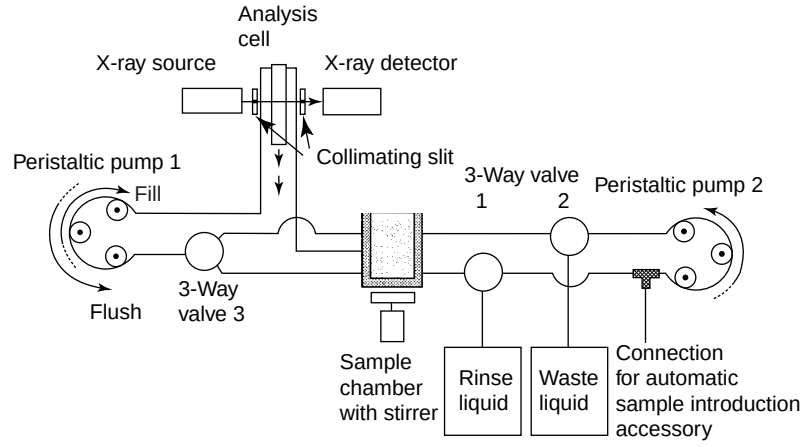

Figure 22 Schematic diagram of Sedigraph 5100.

measurement cell. After a certain time, which can be selected within defined limits, the radiation source and the proportional counter are moved upwards in the direction of the liquid surface at a constant velocity (approx. 1.5 cm min^{-1}). Maximum sedimentation height is 18 cm, the minimum value is 1 cm. The recommended range determined by these geometrical conditions is between 2 and 100 µm. The necessary solids concentration is $c_V \leq 1\%$ for materials with an atomic number ≥ 12.

Sedigraph 5000 D, 5000 ET or 5100[140,145] have a measurement cell moving downwards at a reducing velocity between the radiator (air-cooled X-ray tube with tungsten spot) and the scintillation counter. A scheme for the computer controlled Sedigraph 5100 is given in Figure 22.

A well-dispersed suspension is introduced into a sample chamber from which it circulates through the analysis cell by means of a peristaltic pump. At the start point of the analysis the pump stops and the particles begin to settle in the analysis cell. The measurement of the X-ray attenuation begins near the cell bottom, and because of the shift down the cell the distances between the particles become smaller and smaller. In this way acceptable analysis times are obtained even for very small particle sizes under 1 µm. After the analysis the suspension is pumped to a waste reservoir and using a second peristaltic pump, fresh rinse liquid flows through the system.

The Sedigraph device family has established itself well since its invention.[146–150]

4.2.1.6 Repeated Decanting The method of repeated decanting goes back to the elutriation analyses that were usual in soil analysis until some time ago (see older standard works[103,105,151]). This technique is related to names such as Appiani, Atterberg, Kühn, Wagner, Wahnschaffe, and so on; the principle is shown in Figure 23.

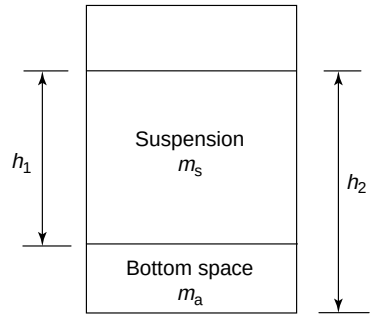

Figure 23 Principle of repeated decanting.

Even though this method is probably not used nowadays for sedimentation analysis, it is a simple method for preparative production of particle fractions below the sieving limit. This is why it will be briefly considered here.

If the suspension which was homogeneous at time $t = 0$ is allowed to sediment in a vessel (see Figure 23), the particle mass m_a in the bottom space after time t is given by Equation (92)

$$\frac{m_a}{M} = 1 - \frac{h_1}{h_2} \int_0^{x_t} 1 - \left(\frac{x}{x_t}\right)^2 q_3(x)\,dx \qquad (92)$$

Now in decanting, the volume of the suspension above the bottom space is removed after a given time t and then replaced by solid-free liquid up to level h_2. After mixing, the supernatant suspension is again removed after time t and so forth, until the amount of solids $m_s/M = 1 - m_a/M$ tends to zero and $m_a/M = 1 - Q_3(x)$ can be assumed with sufficient accuracy. This, of course, is why repeated decanting can be looked upon as an incremental method.

Rumpf and Alex[152] analyzed the process of decanting by calculating the grade efficiency curve after n times. Decanting has to be repeated very often to obtain a high cut sharpness, that is, it requires much time. For

instance, at least 32 decanting steps are required for a proportion <0.1% for $x = 0.9\,x_t$, but even then approx. 5% of the particles with $0.95\,x_t$ that were in the feed are still present in the bottom space. An interesting result is that a drastic reduction in the bottom space (h_1/h_2 near 1) does not give any improvement. Suggestions for the suitable design of decanting vessels are given in the literature.[16,103,105,151–153]

4.2.2 Cumulative Suspension Methods

4.2.2.1 Fundamentals There are two types of cumulative suspension methods:

- methods based on the measurement of m_a (the amount of solids that have sedimented from the suspension volume)
- methods based on the measurement of m_s (the amount of solids still in the suspension volume).

Consequently, there are two different basic relationships for the determination of the particle size distribution, Equations (93) and (94)

$$1 - Q_3(x_t) = \frac{1}{M}\left(m_a - t\frac{\mathrm{d}m_a}{\mathrm{d}t}\right. \quad = \frac{1}{M}\left(m_a - \frac{\mathrm{d}m_a}{\mathrm{d}\ln t}\right. \tag{93}$$

$$Q_3(x_t) = \frac{1}{M}\left(m_s - t\frac{\mathrm{d}m_s}{\mathrm{d}t}\right. \tag{94}$$

where M is the total solid mass in the suspension.

Equations (93) and (94) provide the basis of all cumulative principles. Their derivation originated with Odén[106] but is also found in later works, sometimes in a modified form.[103,154,155]

4.2.2.2 Measurement of Sedimented Amount in Solids All known versions of sediment measurement differing only in the physical measuring principle are shown in Figure 24.

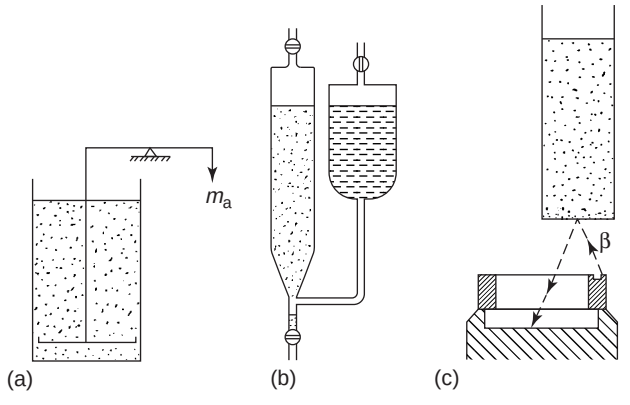

Figure 24 Cumulative suspension methods; measurement of sedimented solid mass m_a. (a) Sedimentation balance, (b) sedimentation column, (c) β-backscattering. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

Sedimentation Balance. The principle of the sedimentation balance (Figure 24a) was technically realized by Odén as early as 1915/1916[106] and in 1920 was available as an automatic device which registered the sedimented mass m_a dependent on time (see Geßner[103]). Since then, there have been many successors that have maintained the basic principle, only adapting the device to the state of the art.[130,154,156–162]

The way the balance pan is arranged has an influence on the measurement range. An arrangement like that schematically represented in Figure 24(a) leads to density convection flows at the pan brim[1–12,69] so that parts of the solids < 3–4 µm are transported under the pan and are thus lost for the measurement. The effect is avoided by a design according to Alex et al.[1–12] and Bürkholz.[163]

In view of their automatic mode of operation and their low solids concentration (<0.1 vol%), sedimentation balances were widely used in the past. This is demonstrated by the multitude of commercial designs (e.g. Gallenkamp (England), Cahn (USA), Mettler (Switzerland), Sartorius (Germany), Shimadzu (Japan)).[164] Because of their relatively long analysis time, nowadays these balances are used only where the advantages outweigh this disadvantage or where other systems cannot be used (e.g. Kowalski[165]).

For the older designs, the measured curve $m_a(t)$ was evaluated graphically according to Equation (93) followed by further improvements (see in von Hahn[105]).

Sedimentation Column. Particularly among English publications, there is information concerning the use of so-called sedimentation columns whose basic principle is traced back to Stairmand;[166] similar designs were already mentioned by Geßner[103] and von Hahn.[105] The principle is represented in Figure 24(b). The solid mass sedimented from a very thin suspension ($c_V < 0.1\%$) is released from the vessel at staggered time intervals, and the sample volume (10 cm^3) is automatically replaced by solid-free liquid from a second vessel. There are even more simple solutions (see Allen[16]) which do not include a liquid reservoir. Evaluation is done either by using Equation (93) or an approximation method proposed by Stairmand.[166]

Beta Backscattering. The principle represented in Figure 24(c) is based upon beta radiation being directed towards the solids sedimented on the bottom of the suspension vessel and the subsequent measurement of the backscattered radiation intensity. According to Hardwick and Laundy[167] the backscattered intensity depends only

on the so-called effective atomic number z_{eff} of the investigated material.

As an example: for water, z_{eff} is approximately 7, for ZrO_2 it is approximately 32. This difference produces the measurement effect. By the solid sedimenting to the bottom, water is displaced so that z_{eff} and thus the backscattered intensity change with the solid mass within a certain range of layer thickness.

An alternative device using β-radiation is the transmission scheme developed by Curzio.[168] Here, the radiation source is fastened above the bottom in such a way that the radiation is attenuated by the solid on the bottom dependent on the layer thickness before it reaches the detector.

4.2.2.3 Measurement of Amount of Solids in Suspension
The principles based on this method are represented in Figure 25.

The manometer method (Figure 25a) goes back to works by Wiegner,[107] Geßner[103] and von Hahn.[105] Lange and Gast[169] used a highly sensitive pressure measurement device which allowed a drastic reduction of the sedimentation height. This method was later been automated.[170] The particle size distribution is calculated from the measured pressure curve $\Delta p(t)$ by Equation (95)

$$Q_3(x) = \frac{\Delta p(t)}{\Delta p_0} - \frac{t}{\Delta p_0}\frac{\mathrm{d}\Delta p}{\mathrm{d}t} \tag{95}$$

In the case of cumulative diving methods, a distinction has to be made between those with constant and those with variable immersion depth (see Figure 25b and c). The use of these methods is simplest if the areometers have a constant cross-section over their entire length; other shapes, however, are possible as well.[103,105,153,171]

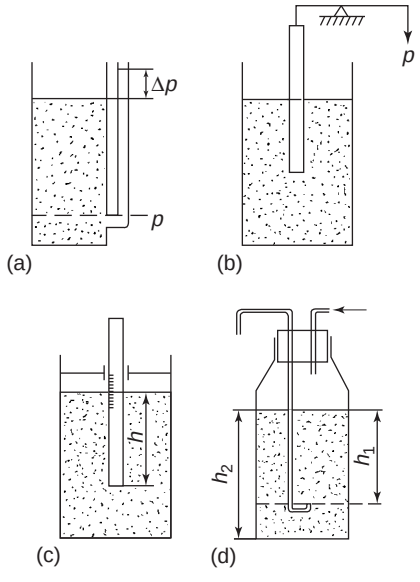

Figure 25 Cumulative suspension methods; measurement of solid mass in suspension. (a) Manometer method, (b) diving balance, (c) swimming balance, (d) single decanting. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

The diving balance (Figure 25b) allows the weight force $P(t)$ of the diving body to be determined dependent on the time t and the particle size distribution to be calculated from Equation (96)

$$Q_3(x) = \Pi - t\frac{\mathrm{d}\Pi}{\mathrm{d}t} \tag{96}$$

where

$$\Pi = \frac{P(t) - P_\infty}{P_0 - P_\infty} \tag{97}$$

where $P_0 = P$ for $t = 0$, $P_\infty = P$ for solid-free liquid.

The swimming balance (Figure 25c) indicates the variable immersion depth $h(t)$; if h_0 (at $t = 0$) and h_∞ (solid-free liquid) are known as well, the particle size distribution is calculated according to Equation (98):[153,171]

$$Q_3(x) = \frac{h_0}{h(t) - h_0}\frac{h_\infty - h(t) + t(\mathrm{d}h/\mathrm{d}t)}{h(t) - t(\mathrm{d}h/\mathrm{d}t)} \tag{98}$$

Single decanting (Figure 25d) as a cumulative method was investigated by Bennert[172] and Schultz;[173] the experimental side of decanting is described in the literature.[16,103,105,151–153,172] Measuring the mass m_S above the bottom space in several vessels at different times t or even in one vessel with different h_1 and t valves, the particle size distribution is determined by the modified Equation (94), Equation (99)

$$Q_3(x) = \frac{h_2}{h_1}\frac{1}{M}\left(m_s - t\frac{\mathrm{d}m_s}{\mathrm{d}t}\right. \tag{99}$$

4.2.3 Line Start Methods

If a thin suspension layer of thickness $\Delta h \ll h$ is arranged on top of a solid-free liquid column and disturbance-free sedimentation is secured, we will immediately obtain in depth h:

- for incremental measurement: the frequency distribution $q_3(x)$,
- for cumulative measurement: the distribution $Q_3(x)$.

The fact that this is a very attractive possibility was the reason that the range of application of this method was constantly investigated or attempts were made to expand it by the development of new devices; this, however, proved possible only to a certain limit. The main reason for this is the difficulty of ensuring disturbance-free sedimentation. Mainly at the beginning, instability occurs caused by the difference in density between the suspension and the solid-free liquid. The

consequence is streak-type vortexing leading to solid particles entering the liquid with high speed. The only reliable method of solving this problem is the density gradient method known from centrifugal equipment (see Allen et al.[1–12]). In gravity sedimentation, however, elimination of this effect has been tried by reduction of the solids concentration[174] or extreme expansion of the sedimentation path.[146]

The quantity measurement is exclusively cumulative and usually restricted to manometric measurement or gravimetric measurement in the form of a sedimentation balance.[146,174] A further device, the so-called US sedimentometer[149] uses the sediment volume as a measured variable; its preferential use in the manufacture of abrasives, however, obviously is due to the fact that the line start method offers a good possibility of determining even very tiny fractions of large grains, which is decisive in this area.

A device proposed by Muschelknautz[175] deserves to be mentioned; it imitates cumulative layer formation by weighing in two sedimentation vessels of different height. This removes all the above problems.

The use of air as a sedimentation medium offers the possibility of shifting the lower application limit to a few micrometers while maintaining long sedimentation paths.[176] This method was used by several authors.[177–179] The Micromerograph[177] uses a sensitive torsion balance for quantity measurement. The solid is fed by pressurized air; the accompanying dispersion effect is assumed to be sufficient for most cases.[180,181] A device described by Kouzov[178] uses a conventional air gun; at the same time, this is one of the rare cases of an incremental quantity measurement (photometric evaluation of plates exposed for different periods of time which are installed at the bottom of the sedimentation vessel). Another application of the line start method is the determination of the drop size of sprays using a sedimentation balance.[179]

4.3 Methods Used in the Centrifugal Field

4.3.1 Fundamentals

4.3.1.1 Modified Stokes Law In order to overcome the disadvantages of gravity sedimentation in the range of small particles $< 1\,\mu m$ (e.g. diffusion, long measurement periods), a centrifugal field can be used instead. However, the acceleration b is a function of radius r ($b = w^2 r$) so that Stokes velocity becomes, Equation (100)

$$\frac{dr}{dt} = \frac{\Delta\rho}{18\eta} w^2 r x^2 \tag{100}$$

For constant angular velocity w, an integration for any sedimentation between an inner radius r_i and an outer radius r_a yields, Equation (101)

$$\ln \frac{r_a}{r_i} = \frac{\Delta\rho}{18\eta} w^2 x^2 t \tag{101}$$

In practice, acceleration and deceleration phases will always occur in addition to phases of constant angular velocity; therefore, the curve of $w(t)$ has to be taken into account in the integration, Equation (102):

$$\ln \frac{r_a}{r_i} = \frac{\Delta\rho}{18\eta} x^2 \int_0^{t_{max}} w^2(t)\, dt \tag{102}$$

To be able to calculate the particle size x for a given running time t_{max} on this basis, $w^2(t)$ can be plotted against t, and the area under the curve can be determined. Rumpf et al.[68] propose the use of an effective time t_{eff} for the case of a constant acceleration and deceleration, Equation (103):

$$t_{eff} = \frac{t_1}{3} + (t_2 - t_1) + \frac{t_3 - t_2}{3} \tag{103}$$

where t_1 is the end of acceleration phase, t_2 is the start of deceleration phase, and t_3 is the end of deceleration phase.

The incremental and cumulative principles used for quantity measurement in centrifuges are nearly the same as in the gravitational field. In the following, the theoretical basis will be briefly described (see also several overviews[13,16,53,182–184]).

4.3.1.2 Quantity Measurement for Suspension Methods; Incremental Principle The suspension level in the centrifuge vessel reaches radius r_i. There is an incremental measurement device at r for determining the solids volume concentration $c(r, t)$. The bottom of the suspension vessel is at radius r_a. Then the integral Equation (104) holds

$$\frac{c(r,t)}{c_0} = \int_0^{x_t} \exp\left\{-2\frac{\Delta\rho}{18\eta} w^2 x^2 t\right\}^2 q_3(x)\, dx \tag{104}$$

and for a constant angle velocity, Equation (105)

$$\frac{c(r,t)}{c_0} = \int_0^{x_t} \exp\left\{\left(\frac{x}{x_t}\right)^2 \ln \frac{r_i}{r}\right\}^2 q_3(x)\, dx \tag{105}$$

Equations (104) and (105) are fundamental equations for incremental suspension methods in the centrifugal field. Their use in the evaluation of measured values depends on the experimental approach. The following procedures are possible:

1. The dependence of the solids concentration on time t is determined at distance r. In this case, there is no general solution to Equation (104) for the particle

size distribution that is being sought, so there are various approximations. The simplest and most common was proposed by Norton and Speil[185] assuming that r/r_1 tends to 1, which is true only for centrifuges with big rotors and short sedimentation paths (long-arm approximation). Evaluation is done as in the gravity field but taking into account the increased acceleration. The error of this approximation in comparison with an exact evaluation was investigated by Alex.[186]

There is also the theory developed by Kamack[187] which reaches an approximate solution by assuming that the dependence of the solids concentration is linear between two measurement points t_1 and t_2. Then, Equation (106) can be written:

$$\frac{c(x_t, k')}{c_0} = \sum_{l=1}^{k} q_3(x_{t,l}) B_{k,l}$$

$$B_{k,l} = \int_{x_{t,l-1}}^{x_{t,l}} \exp\left\{ \left(\frac{x}{x_{t,k}}\right)^2 \ln \left(\frac{r_i}{r}\right)^2 \right\} \mathrm{d}x \quad (106)$$

obtaining a linear set of equations for the discrete values of $q_3(x_{t,l})$.

2. The solids concentration is measured dependent on the radius r at a fixed time t. Equation (105) then can be differentiated with respect to x_t, and we obtain Equation (107):[53,68]

$$Q_3(x_t) = \int_0^{c(r)/c_0} \left(\frac{r}{r_i}\right)^2 \mathrm{d}\,\frac{c(r)}{c_0} \quad (107)$$

Hence, if $c(r)/c_0$ is plotted against $(r/r_i)^2$, the distribution is determined by integration.

3. Svarovsky and Svarovska[188] use a solution of Equation (104) which was also proposed by Kamack, obtained for the distribution after time t (when the particle size x_t passes the measurement level). Their equation is an exact solution but the contained convolution integral can be solved only numerically. Therefore, Svarovsky and Svarovska[188] replaced it by an analytical approximation easily produced by an analogous calculation scheme, enabling the continuous evaluation of the sedimentation curve. A further solution was suggested by Fautz.[184]
4. Finally, the fundamental equations can also be solved by using analytical distribution functions for $q_3(x)$ and determining their parameters.[188]

4.3.1.3 Quantity Measurement for Suspension Methods; Cumulative Principle The mass m_a/M which is in the bottom space $r_{a2} - r_{a1}$ after time t is given by Equation (108)

$$\frac{m_a}{M} = 1 - \int_0^{x_t} \frac{r_{a1}^2 \exp\left\{(x/x_t)^2 \ln (r_i/r_{a1})^2\right\} - r_i^2}{r_{a2}^2 - r_i^2} q_3(x)\,\mathrm{d}x \quad (108)$$

If the cumulative measurement is organized so that the mass that sedimented at the bottom is determined (sedimentation balance), that is $r_{a1} = r_{a2} = r_a$, Equation (108) is simplified to Equation (109)

$$\frac{m_a}{M} = 1 - Q_3(x_t) + \int_0^{x_t} \frac{1}{1 - (r_i/r_a)^2} \times \left[1 - \exp\left\{\left(\frac{x}{x_t}\right)^2 \ln \left(\frac{r_i}{r_a}\right)^2\right\}\right] q_3(x)\,\mathrm{d}x \quad (109)$$

This is the relationship which was indicated by Romwalter and Vendl.[189] For the case where m_a/M is measured dependent on time there is a well-known solution originally proposed by Robinson and Martin[190] [cited by Sullivan and Jacobsen[183]]. An approximation method was suggested by Lloyd et al.[191]

As with the incremental methods, there are exact solutions for the case where measurement is dependent on the radius at a constant time. The first possibility is to measure the sedimented mass m_a/M dependent on the inner radius r_i. According to Brown[192] the solution is, Equation (110)

$$1 - Q_3(x_t) = \mu(r_i) - \frac{r_a^2 - r_i^2}{2r_i} \frac{\mathrm{d}\mu}{\mathrm{d}r_i} \quad (110)$$

where $\mu = m_a/M$.

A second possibility is to vary the outer radius r_a according to a proposal made by Donoghue and Bostock[193] so that, Equation (111)

$$1 - Q_3(x_t) = \mu(r_a) + r_a \frac{r_a^2 - r_i^2}{2r_i^2} \frac{\mathrm{d}\mu}{\mathrm{d}r_i} \quad (111)$$

If the solid mass in suspension m_S is responsible for the measured effect, Equation (112) holds:

$$\frac{m_S}{M} = \int_0^{x_t} \frac{1}{1 - (r_i/r_a)^2} \exp\left\{\left(\frac{x}{x_t}\right)^2 \ln \left(\frac{r_i}{r_a}\right)^2\right\} q_3(x)\,\mathrm{d}x \quad (112)$$

For repeated decanting until the solid mass m_S/M in the supernatant liquid tends to zero (this is absolutely usual for preparation purposes) $m_a/M = 1 - Q_3(x_t)$ for the solid mass in the bottom space. Rumpf and Alex.[152] have again developed the grade efficiency function dependent on the number of decanting steps to approximate this state; see examples in Bernhardt.[13]

In single decanting, the solid mass m_S/M in the volume above the bottom space is usually determined. To obtain the distribution $Q_3(x_t)$, Weichert[194] has proposed a

similar method to that of incremental analysis: single decanting operations are carried out at different times and m_S/M is determined dependent on time t. Linearity is assumed between the measurement points so that we can write Equations (113) and (114):

$$\frac{m_S(x_{t,k})}{M} = \sum_{l=1}^{k} q_3(x_{t,l}) A_{k,l} \tag{113}$$

$$A_{k,l} = \int_0^{x_t} \frac{\exp\left\{(x/x_t)^2 \ln (r_i/r_{a1})^2\right\} - r_i^2/r_{a1}^2}{1 - r_i^2/r_{a2}^2} q_3(x)\,\mathrm{d}x \tag{114}$$

For $q_3(x_{t,l})$, a linear set of equations is again obtained by analogy with Kamack's theory.[187]

4.3.1.4 Quantity Measurement for Line Start Methods The mathematical difficulties of the suspension methods have repeatedly led to the use of line start methods in sedimentation centrifuges. As in the gravity method, a thin suspension layer has to be brought to the top of a solid-free liquid. Then, the frequency distribution $q_3(x)$ (incremental measurement) or the distribution $Q_3(x)$ or $1 - Q_3(x)$ (cumulative measurement) is obtained directly (with restrictions for light optical incremental methods[16]). The disadvantage of this simplification, however, is the considerable experimental difficulties caused by the formation of streak-type mixing phenomena at the interface between suspension and pure liquid which falsify the result. A detailed description and wide bibliography are given by several authors.[1–12,68,195,196] Accordingly, the cause of this problem is mainly the instability of the superposed layers which leads to a special form of density convection flow. According to Alex et al.,[1–12] two conditions have to be met to prevent such streak formation:

- The density of the superpositioned suspension has to be smaller than that of the pure sedimentation liquid. Both liquids have to be miscible to avoid additional problems caused by wetting.
- The solids concentration has to be small enough that the particles travelling from the suspension positioned on top of the pure sedimentation liquid cannot produce density convection flows by their presence. This is guaranteed for solids concentrations under 0.01%, a range for which, however, none of the usual methods of quantity measurement works.

Therefore, the so-called density gradient method has to be used as the only existing way of avoiding streak formation; its successful application is described by Alex et al.,[1–12] Brugger[197] and Coll and Searles.[198] Its content is to introduce into the centrifuge, in a suitable way, a sedimentation liquid whose density increases in the radial direction. This requires some experimental effort which will probably not be spent on routine analyses, since the problems of suspension methods can be solved by modern computer equipment.

In contrast, Kaye[199] holds the opinion that the above difficulties in the line start methods in many cases are caused by the lack of stability of the centrifuge structures at high rotational speed.

4.3.2 Sedimentation Centrifuges

Since the first work by Svedberg and co-workers[52,108] on the use of centrifuges for particle size analysis, there have been a large number of publications on methodological and technological solutions. An overview and historical development are given in the monographs by Allen,[16] Chodakov and Judkin,[182] Batel,[153] Bernhardt,[13] Köster,[151] Geßner[103] and von Hahn[105] as well as by the Analytical Methods Committee.[14]

On the whole, the experience seems to have been accepted according to how much methodological and technological effort is needed for accurate measurement. This applies both to the case of sedimentation in orbiting vessels (bucket centrifuges) and to the rotation of the suspension in a cylindrical vessel like a concentric ring (disk centrifuges).

4.3.2.1 Centrifuges with Incremental Measurement The three main principles are shown in Figure 26.

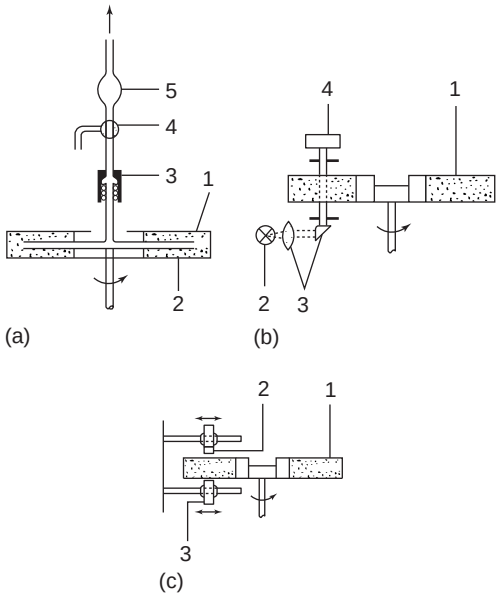

Figure 26 Basic principles of sedimentation centrifuges with incremental quantity measurement; (a) pipette centrifuge, (b) photosedimentation centrifuge, (c) X-ray or γ-ray centrifuge. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

Pipette Centrifuge. This device (Figure 26a) is usually a disk centrifuge. The suspension rotates like a ring in the disk rotor (1) into which between four and eight pipettes stand and participate in the rotation. It has to be ensured that a tight connection is created between the rotating pipette shaft and the part of the sample which is at rest (stopcock (4), glass vessel (5)), for instance by device (3). When a vacuum is applied, the suspension sample flows into the sample vessel (5) from where it can be taken for further treatment (drying, weighing). The evaluation is done either in the gravity field (long-arm approximation), as for example in the centrifuges by Slater and Cohen[200] or Gupta,[201] or by one of the approximations of Equation (105) as for example in the centrifuge according to Kamack[187] or the modern Ladal pipette centrifuge[202] and the Andreasen pipette centrifuge of Fritsch.

Photosedimentation Centrifuge. In the older types of centrifuge that use this principle, the centrifuge rotor (1) is a disk (see Figure 26b). Through the ring-shaped suspension, a light beam which is produced by the lamp (2) and optical device (3) (lens, prism) is emitted at a certain point and falls on the detector (4) on the opposite side. Light source and detector can also be shifted in the direction of the axis to shorten the measurement time (e.g. Allen and Khalili[203]).

The first photosedimentation centrifuge was developed by Kaye and Jackson;[204] it was followed by the version which became known as the Joyce–Loebl centrifuge.[1–13] In both cases, a line start technique was used, with the usual problems in spite of special prescriptions (buffer layers etc.).[195,196] Allen tried to circumvent the problems by developing a centrifuge with suspension technique (Ladal photo centrifuge[203]) which involved the evaluation problems discussed by Lloyd et al.[191] and Svarovsky and Svarovska.[188]

Several modern Japanese sedimentation centrifuges have abandoned the disk rotor and use rotating measurement cells with a rectangular cross-section instead. Two of these cells are mounted on the rotor, one being filled with suspension and the other opposite being used for balancing or reference purposes. When the rotor is at rest, the devices can be employed as gravity photosedimentometers.

As discussed already, incremental photoelectric quantity measurement requires knowledge of the extinction function, and particularly so in the measurement range of the centrifuges (0.01–10 μm). In the centrifuge devices of Shimadzu and Horiba, a solution to this problem has been attempted by the possible use of a fixed standard extinction function or an extinction function obtained in a different way. This approach is not satisfactory either, since the effective extinction function of the investigated material is not known in many cases and the user does not know how great is its deviation from standard.

Weichert[128] solved this problem by an experimental determination of the extinction function, however, the expense of the measurement and calculation is very high. This can be the reason for the fact that there has been no commercial use of the method.

X-ray or γ-ray Centrifuges. As is well known, the problem of the extinction coefficient does not exist for X-rays or γ-radiation. The principle represented in Figure 26(c) is completely analogous to the light optical method. The disk-shaped rotor (1) is subject to radiation from a suitable radiation source (2) at a certain point in the suspension ring; the intensity of the radiation incident on the opposite side is measured by the detector (3). Source (2) and detector (3) can be shifted in a radial direction to reduce the measurement time. The Ladal γ-ray centrifuge[202,205] uses a radioactive isotope (Pm 147/Al) and a scintillation counter. To obtain a measurable attenuation effect, the solids concentration has to be considerably higher than with photoelectric methods. This may lead to restrictions on use since the stability of the suspension strongly depends on the solids concentration for very small particles.

The development of a disk centrifuge (Du Pont/Brockhaven Scanning X-ray Disc Centrifuge) is described by Allen.[206] This device can be used both in the gravity field and the centrifugal field, that is, in a measurement range between 0.01 and 100 μm. Finely focussed X-rays are directed to the sedimentation path by a scanning program so that only a few minutes are required for the whole measurement.

4.3.2.2 Centrifuges with Cumulative Measurement

Some of the best known principles of cumulative sedimentation centrifuges are represented in Figure 27.

Manometer Centrifuge. In this scheme borrowed from gravity sedimentation (Figure 27a), on the centrifuge rotor (1) there is only one measurement cell (2) at which the pressure difference between upper and lower end is measured. The counterweight (4) ensures an equilibrium. By means of cables (3), the pressure is supplied from the rotating centrifuge shaft to the data transmitter (5) and then to the differential pressure converter (6) producing an electrical signal for further evaluation. The development of the device originated with Böwing and Gast;[207] Lauer[208] reported on further developments and the version marketed by Alpine. This centrifuge is designed so that the long-arm approximation and the related evaluation algorithm can be employed. The principle has been improved by Jimbo et al.[209] who used a highly sensitive pressure measurement technique which

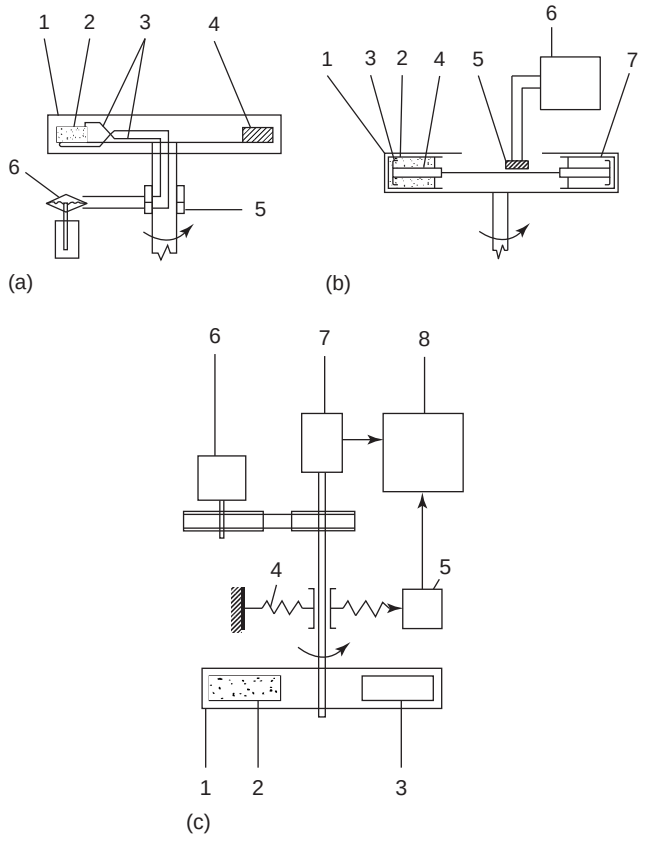

Figure 27 Basic principles of sedimentation centrifuges with cumulative quantity measurement. (a) Manometer centrifuge, (b) centrifugal sedimentation balance, (c) eccentric centrifuge. (Reproduced from Bernhardt,[13] Copyright 1994, with kind permission from Kluwer Academic Publishers.)

allowed the required solids concentration to be reduced further.

Centrifugal Sedimentation Balance. Solids sediment are placed on balance pans either in measurement cells of suitable design or in the disk rotor; the first principle is represented in Figure 27(b). On the rotor (1) there is a measurement cell (2) filled with suspension, in which there is a balance pan (3) and into which a diving body (4) is immersed. The cell (7) installed on the opposite side contains solid-free liquid. Thus, the measurement effect for the sedimenting mass is a change of pathlength measured by device (5) and supplied to the evaluation unit (6). Bürkholz et al.[210] and Muschelknautz[175] have reported on the development of such a balance. Similar devices with light optical scanning were introduced by Tarasov and Chodakov.[211]

Eccentric Centrifuge. An original solution for the cumulative measurement in a sedimentation centrifuge was found by Kaya and Yokoyama;[212] the principle is represented in Figure 27(c). In the rotor (1) two measurement cells are arranged opposite each other, one of which (2) contains the suspension (solids concentration approx. 3 wt%) and the second of which (3) contains solid-free liquid. Owing to the shift of the gravity center in cell (2) caused by the sedimentation process, an eccentricity variable in time is generated leading to a deflection of the rotor shaft. The relationships between sedimenting mass and eccentricity, on the one hand, and between eccentricity and deflection, on the other hand can be described by formula.[212] For evaluation, the rotational speed of the rotor shaft has to be known (tachometer (7)) in addition to the deflection measured by parts (4) and (5). The drive (6) propels the centrifuge; the evaluation unit (8) processes the information obtained (rotational speed and deflection) to determine the searched distribution function.

In conclusion, it should be noted for the sake of completeness that single and repeated decanting is also possible in centrifuges. A typical example is the Joyce–Loebl disc centrifuge.

ACKNOWLEDGMENTS

The author thanks the Kluwer Academic Publishers for kind permission to use the Figures 6–9, 11, 13–15, 17–21, 24–27 and corresponding explanations from C. Bernhardt, *Particle Size Analysis. Classification and Sedimentation Analysis*, Copyright, 1994.[13]

LIST OF SYMBOLS

a	Half-axis; radius
A	Factor; area; feed material (index); buoyancy
A_P	Buoyancy of a particle
Ar	Archimedes number
A_V	Scattering section of individual particle related to particle volume
B	Constant; factor
$B_{k,l}$	Integral function
c	Half-axis; concentration
c_0	Concentration at time $t = 0$
c_{eff}	Effective solids concentration
c_m	Solid mass concentration
$c_{m,0}$	Solid mass concentration at time $t = 0$
c_V	Solid volume concentration
c_w	Drag coefficient
$c_{w,st}$	c_w according to Stokes
C_L	Factor
D	Function; diffusion coefficient
D_H	Thickness of hydrate sheath
f	Systematic error

f_x	Error in particle sizes
F_P	Projection area of a particle vertical to flow direction
g	Gravitational acceleration
G	Weight
G_P	Weight of a particle
h	Height; depth; distance from the particle surface
h_M	Height of measurement level
I	Radiation/light intensity
I_A	Intensity after transmission
I_E	Intensity prior to transmission
I_t	I_A at time t
I_∞	I_A at time $t = \infty$ (solid-free)
I_0	I_A at time $t = 0$
k	Boltzmann constant ($=1.38 \times 10^{-13}\,J\,K^{-1} = 1.38 \times 10^{-16}\,g\,cm^2\,s^{-2}\,K^{-1}$); exponent
K	Shape correction factor
$K(x)$	Extinction function; corrected cumulative distribution
l	Distance between the centers of two spheres
l_1	Distance of sphere center from a wall
L	Distance between two walls; transilluminated length; mean free path
Lj	Ljaščenko number
m	Parameter of Gates–Gaudin–Schumann distribution; refractive index; coefficient
$m(t)$	Solid mass at time t
m_0	Solid mass at time $t = 0$
m_a	Particle mass in bottom space of a sedimentation vessel
mea	Measured (index)
m_p	Particle mass
m_S	Particle mass above bottom space of a sedimentation vessel
M	Mass (total); moment
M_M	Molecular mass
$M_{k,r}$	k th moment related to zero, quantity r
n	number
n_F	Refractive index (fluid)
N	Number
N_L	Avogadro number ($N_L = 6.022 \times 10^{23}\,mol^{-1}$)
p	Fraction, shape factor $= c/a$, pressure
P	Weight; pressure
P_0	P at time $t = 0$
P_∞	P at time $t = \infty$ (solid-free)
q_3	Density distribution with quantity type volume, mass
Q	Cumulative distribution
Q_0	Q with quantity type number
Q_2	Q with quantity type surface
Q_3	Q cumulative distribution with quantity type volume, mass
r	Radius
r_a	Outer radius
r_i	Inner radius
R	Radius; factor
Re	Reynolds number
R_P; R_R	Functions for plates or rods
S_P, S_R	Functions for plates or rods
t	Time
t_{99}	Time for $v/U = 0.99$
t_{eff}	Effective time
T	Temperature; transmission
$T(x)$	Cut function, grade efficiency function
T_P, T_R	Functions for plates or rods
U	Terminal velocity of a particle
U_0	U in ideal case (without external influence)
U_a	U in a-direction
U_c	U in c-direction
U_P, U_R	U for plates or rods
U_{rel}	Relative velocity
U_{st}	U according to Stokes
U_x	U in x-direction
U_v	Stationary sedimentation velocity
U_z	U in z-direction
v	Velocity
V	Volume; energy of interaction; particle volume (index)
V_A	Attraction energy
V_{el}	Energy of electrostatic repulsion
V_P	Volume of particle
V_{ster}	Steric repulsion energy
W_{El}	Drag force to an ellipsoid
W_P	Drag force of a particle, a plate
W_{St}	Drag force according to Stokes
x	Particle size; coordinate axis
x_{50}	Median of distribution
x_a, x_c	Axis of a spheroid
x_A	Diameter of a sphere of equal surface area
x_F	Equivalent projected diameter
x_S	Apparent particle size
x_{st}	Particle size according to Stokes
x_t	Particle size which has passed through distance h after time t
x_V	Diameter of a sphere of equal volume
z	Coordinate axis; valency
z_{eff}	Effective atomic number
α	Mie parameter; angle in degrees or rad; degree of desagglomeration
β	Angle; coefficient; function; degree of agglomeration; quantity type (index)
Δh	Measurement gap height
Δp	Pressure difference
Δp_0	Pressure difference at time $t = 0$
$\Delta\rho$	Density difference $\rho_s - \rho_m$

ΔQ	Quantity proportion in a particle size class
ζ	Flow (zeta) potential
η	Viscosity of medium
Θ	Angle
κ	Dynamic shape factor
λ	Wavelength
μ_m	Mass attenuation coefficient
$\mu_{m,eff}$	Mass attenuation coefficient of a mixed material
Π	Relative pressure difference
ρ	Density
ρ_{eff}	Effective density of a mixed material
ρ_H	Density of hydrate sheath
ρ_m	Density of medium
$\rho_{m,0}$	Density of pure liquid
ρ_S	Density of solid
φ	Suspension density
φ_0	Suspension density at time $t = 0$
$\varphi(h, t)$	Suspension density at time t in depth h
ϕ_0	Gaussian normal distribution
ψ_W	Shape factor according to Wadell (sphericity)
ψ_{HC}	Shape factor according to Heiss and Coull (circularity)
w	Angular velocity

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Sieving in Particle Size Analysis • Turbidimetry in Particle Size Analysis

APPENDIX

List of Liquids and Dispersing Aids

Abrasives	W + sp4, W + sp3
Activated charcoal	W + am, IB, IP, W + sp4, W + sl, W + so
Alumina	W, W + hc, W + sk6, W + sp6, W + st, W + sp4, CN, TC
Aluminum	TC, W + EG, W + sp6, W + sp3, W + st, W + so, GL, EG, PI
Aluminum silicate	W + sp4, W + sp3
Anhydrite	ML
Apatite	W + sp3, W + ac
Ash, Flyash	W + sp4, W + Daxad 23, W + sp6
Barium carbonate	W + sp6, W + sp4, ML, CN
Barium sulfate, Baryte	W + sp6, W + Lissapol NX, W + Dispersol T, W + sp6, W + sci
Bauxite	W + sp4
Bentonite	W + sca, W + am, W + sp4, W + so, W + sk6
Blast-furnace slag	W + sp6, W + sp4, CL, CN, PI
Boron carbide	W + sp4, BL + EL
Boron nitride	BL + EL
Brown coal	BI, DP, CN + ML, CL + ML, W + GC
Calcium carbonate	W + spa, W + am, W + Dispersol T, W + Triton X-100, W + GC, W + am, W + sp3, W + sp4, W + sci + sp4, AC, XL, EG, PI, CN + AA
Calcium fluorite	W + sp6, W + am, W + pc, ML + pc, CL, CN, AC, EG
Calcium hydroxide	CL, EL, PI
Calcium oxide	EG, AC, CL, CN, PM, EG + cac, EG + stc, EG + coc
Calcium phosphate	W + sp4, W + sp6, W + sp3, W + EL, EL, BL
Carbon black	CH + Aerosol OT, AC, CL, ML, EL, W + Aerosol OT, W + Lomar PW, W + sl, PI, W + ta, W + Succipon K3, BI, GC
Cement	PM + oc, PI, ME + sp4, BL, BI, CL, CN,CL + AA, EL + cac, EL + stc, EG + cac, EG + stc, EG + coc, PL, PL + oa, GC, ME + GC
Chromium	BI, BL, PM
Clay	W + sca, W + so, W + sp6, W + sp4, W + sp3, W + sk6, PI, BP + EL, W + spa
Coal	W + sl, W + cac, W + oa, W + OLOA 1200, EL + cac, EG + cac, EG + stc, PI, EG + coc, ML, CH, CL, CN, W + Lomar PW, XL, AC, PM + oa, BI, W + Tween 20, CL + ML, W + EL, W + CL, W + Aerosol OT, W + ta
Cocoa	DP, AA, BI, BL, DP + BI, AC, CN
Coke	BI, PI, W + Perminal BX, W + sl, W + ta + am, W + Lomar PW, EL + cac, EL + EG + cac
Diamant	W + sca, W + sp6, W + sp3, W + Aerosol OT, EL

Earth, soil	W + sp4, W + so, W + EG, BP + EL
Feldspar	W + sp6, W + sp3, W + sp4, W + so
Flour	PM + oa, BI, DP + BI, DP, AA, EL, CN, CL
Garnet	W + sp6
Glass	W + sp4, W + sp3, W + sk6, W + Dispersol T, W + Lissapol NX, W + Triton X-100, W + sp6, BL, CL, W + EG, W + GC, ML, EG
Graphite	W + ta, W + sl, W + am + sl, W + Teepol, W + sp3, W + ls, EL, BL, W + Aerosol OT
Gypsum	EL + cac, EG + cci, EL + EG + cci, ML + cac, AA, ML + EG + EL + cac, EG + cac, EG + stc, EG + coc, BI, EG + EL + sci
Gypsum (raw)	W + pci, W + Aerosol OT, EL + cac, EG + cci, ML
Iron	W + GC, CN, BL, W + EG, EL, W + Lissapol NX, W + EG + sp3, CL, CH, W + Triton X-100, W + sp4
Kaolin	W + am, W + ls, W + sp3, W + hc, W + sp4, W + sk6 + sca, W + so, W + sh, W + scaW + spa, W + sp6
Magnesium carbonate	W + sp4, W + sp6, ML, EL, W + am, CN, EG + cac, EG + stc, EG + coc
Magnesium oxide	BI, EG, ML
Magnesium silicates	W + sp3, W + sk6, W + sca, W + sp4, ML
Marl	W + sp4, W + sk6, W + sk6 + sca
Mica	W + sp4, W + sp6
Milk powder	BI + DP, ON, BI, AA
Pigments	W + Lissapol NX, W + pci, W + EG, W + sp4, W + Aerosol OT, PI, CL, CN
Polyvinylchloride	W + ta, W + sp3, W + sl, EG, PI + sl, BI
Portland cement	OI, BL, CL, CN, CN + AA, EG, EL + cac, EL + stc, EG + cac, EL, BI, ML, PI
Potato starch	BI, CN, BI + DP
Puzzolana	W + sp4
Quartz	W + Lissapol NX, W + sp4, EL + sp3, W + sp6, XL, W + so, W + EL, W + sp3, W + sp4 + Sterox, W + sh, W + Triton X-100
Raw meal	CL, CL + AA, BI, W + sp4
Sand, sandstone	W + CL, W + CN, EL + BP, W + sp3, W + sp4
Silicates	W + sp4, W + EL, W + GL, W + EG + sp3, W + sp4 + sp6 + so
Silicon	W + Cetrimide, W + Dispersol T, W + sp6, EL
Silicon carbide	BL, ML + Chelaplex III, W + Tween 20, W + sk6, W + sk6 + sca, W + EG + sp3
Silicon dioxide	W + sp6, W + Triton X-100, W + XL, W + EL
Slag	W, PI, W + sp4
Soil	W + sp4, W + EG, W + so, BP + EL
Starch	W + sp6, BI, PI, DP, BP + EL, BI + DP, CL, EL, ML, GC, AA
Steatite	W + sk6, W + sk6 + pca, W + sp4
Steel	W + EG, W + EG + coc, W + Lissapol NX
Sugar	BI, BI + DP, DP, AA, CN, PI
Talcum	W + sp4, W + sp6
Tin	BL, BI, BI + BL
Tin dioxide	W + sp4, W + sci, W + sk6, W + sk6 + sca
Titanium	W + Lissapol NX, W + sp4, W + sp6
Titanium carbide	W + sp4, W + Daxad 30
Titanium dioxide	W + EL, EL, W + sk6 + ph, W + EG, W + Daxad 30, W + Dispersol T, W + sp4, W + sp6, CN, CL + AA, ML, EG, W + GC, XL, W + Lissapol NX
Tuff	W + so, W + am
Tungsten	W + Renex 648, GC, EL, AC, ML, W + EG, W + GC, BL, W + GC + sp6, W + Triton X-100, EG + cac, EG + stc, EG + coc
Tungsten carbide	W + sp6, W + EG, EG + cac, EG + stc, EG + coc
Tungsten oxide	W + sp4, W + sp6, W + EL, CN, BL, W + GC, BL + EL, CL

Wheat flour	CN, EL, DP, BI, BI + DP, PM
Zinc	W + sp6, W + sp3, EL, BL, AC, CN, CL, EG + cac, BI
Zinc oxide	W + sp3, W + sp4, W + sp6, W + sk6, W + Tamol SN, W + GC, BL, EL
Zirconium	ML, W + sp4, W + sk6, ML, BI, W + ML, W + EL
Zirconium dioxide	W + GC, Woa, W + sp4, W + sk6, W + sp6
Zirconium silicate	W + sp4, W + sk6 + sca, W + EL, W + Triton X-100

Liquids

AA	Isoamylalcohol
AC	Acetone
BA	*n*-Butylamine
BI	Isobutanol
BL	*n*-Butanol
BN	*n*-Butane
BP	Butylphthalate
CH	Cyclohexane
CL	Cyclohexanol
CN	Cyclohexanone
DP	Diethylphthalate
EG	Ethyleneglycol
EL	Ethanol
ET	Ethylene
GC	Glycerol
GL	Glycol
HN	*n*-Hexane
ML	Methanol
OI	Isooctane
ON	*n*-Octanol
PI	Isopropanol
PL	Petrol
PM	Petroleum
TC	Carbon tetrachloride
W	Water
XL	Xylene

For more details, see in Bernhardt.[13]

Dispersing Aids

ac	Aluminum chloride
am	Ammonia
as	Aerosol OT
cac	Calcium chloride
cci	Cobalt citrate
coc	Cobalt chloride
hc	HCl
ls	Lignin sulfonate
ns	Naphthalene stearosulfonic acid
oa	Oleic acid
pc	Potassium chloride
pca	Potassium carbonate
pci	Potassium citrate
ph	Potassium hydroxide
sas	Sodium alkylsulfonate
sca	Sodium carbonate
sci	Sodium citrate
sh	Sodium hydroxide
sk6	Sodium/potassium hexametaphosphate
sl	Sodium linoleate
so	Sodium oxalate
sp3	Sodium triphosphate
sp4	Sodium tetrapyrophosphate
sp6	Sodium hexametaphosphate (Calgon)
spa	Sodium polyacrylate
st	Sodium tartrate
stc	Strontium chloride
ta	Tannic acid

REFERENCES

1. K. Leschonski, W. Alex, B. Koglin, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(1), 23–26 (1974).
2. K. Leschonski, W. Alex, B. Koglin, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(3), 101–106 (1974).
3. B. Koglin, K. Leschonski, W. Alex, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(7), 289–292 (1974).
4. W. Alex, B. Koglin, K. Leschonski, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(9), 387–390 (1974).
5. W. Alex, B. Koglin, K. Leschonski, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(11), 477–480 (1974).
6. B. Koglin, K. Leschonski, W. Alex, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(13), 563–566 (1974).
7. K. Leschonski, W. Alex, B. Koglin, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(15), 641–644 (1974).
8. W. Alex, B. Koglin, K. Leschonski, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(17), 729–732 (1974).
9. K. Leschonski, W. Alex, B. Koglin, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(19), 821–824 (1974).
10. K. Leschonski, W. Alex, B. Koglin, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(21), 901–904 (1974).
11. B. Koglin, K. Leschonski, W. Alex, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **46**(23), 984–987 (1974).
12. B. Koglin, K. Leschonski, W. Alex, 'Teilchengrößenanalyse', *Chem. Ing. Tech.*, **47**(1), 21–24 (1975).
13. C. Bernhardt, *Particle Size Analysis. Classification and Sedimentation Methods*, Chapman & Hall, London, 1994.
14. Analytical Methods Committee, Particle Size Analysis Subcommittee, 'Classification of Methods for Determining Particle Size, a Review', *Analyst*, **88**(1044), 156–187 (1963).

15. T. Allen, 'A Review of Sedimentation Methods of Particle Size Analysis', Proceedings Conference on PSA, Loughborough, Sept., 454–476, 1991.
16. T. Allen, *Particle Size Measurement*, 4th edition, Chapman & Hall, London, 1994.
17. J. Happel, H. Brenner, *Low Reynolds Number Hydrodynamics*, Kluwer Academic Publishers, Leyden, 1983.
18. H. Lamb, *Lehrbuch der Hydrodynamik*, Leipzig, 1931.
19. G.G. Stokes, *Cambridge Phil. Trans.*, **8**, 287 (1845).
20. A.B. Basset, 'A Treatise on Hydrodynamics', New York, Dover, Vol. 2, 1961.
21. C.N. Davies, 'The Sedimentation of Small Suspended Particles', Proceedings Symposium on PSA, London, February 1974, Suppl. to Trans. Inst. Chem. Eng., **25**, 25–39, 1947.
22. E. Cunningham, *Proc. R. Soc. London, Ser A*, **83**, 357 (1910).
23. P.S. Epstein, *Phys. Rev.*, **23**, 710 (1924).
24. J. Stauff, *Kolloidchemie.*, Springer-Verlag, Berlin, 1960.
25. R. Clift, J.R. Grace, M.E. Weber, *Bubbles, Drops and Particles*, Academic Press, 1978.
26. C.W. Oseen, 'Über die Stokessche Formel und über eine verwandte Aufgabe in der Hydrodynamik', *Arkiv. Mat., Astr. och Fysik*, **6**(29), 75 (1910).
27. W. Gumz, 'Uber die Fallgeschwindigkeit von Kugeln unter besonderer Berücksichtigung des für die Staubtechnik wichtigen Bereichs', *Feuerungstechnik*, **26**(8), 253, 255 (1938).
28. I. Proudman, J.R.A. Pearson, *J. Fluid Mech.*, **2**, 237 (1957).
29. W. Möller, 'Experimentelle Untersuchungen zur Hydrodynamik der Kugel', *Phys. Z.*, **39**(2), 57–80 (1938).
30. L. Schiller, A. Naumann, 'Über die Grundlegenden Berechnungen bei der Schwerkraftaufbereitung', *VDI-Z.*, **77**(12), 318–320 (1933).
31. C. Bernhardt, *Granulometrie, Klassier-und Sedimentationsmethoden*, Deutscher Verlag für Grundstoffindustrie, Leipzig, 1990.
32. H. Heywood, 'Uniform and Non-uniform Motion of Particles in Fluids', Proceedings Symposium on the Interaction between Fluids and Particles, Inst. Chem. Eng., London, 1–8, 1962.
33. A.I. Medalia, 'Dynamic Shape Factors of Particles', *Powder Technol.*, **4**, 117, 138 (1970/71).
34. A.T. Hjelmfeld, L.F. Mockros, 'Stokes Flow Behavior of an Accelerating Sphere', *J. Eng. Mech. Div., Proc. Am. Soc. Civil Eng.*, **93**, 87–102 (1967).
35. H.A. Oberbeck, 'Über Stationäre Flüssigkeitsbewegunngen mit Berücksichtigung der inneren Reibung', *Crelles J.*, **81**, 62–87 (1876).
36. R. Gans, 'Wie fallen Stäbe und Scheiben in einer reibenden Flüssigkeit?' *Sitzungsber. d. metall-physikal Klasse d. Königl. Bayer. Akademie. d. Wiss.*, **41**, 191–203 (1911).
37. J. Schmiedel, 'Experimentelle Untersuchungen über die Fallbewegung von Kugeln und Scheiben in einer reibenden Flüssigkeit', *Phys. Z.*, **29**(17), 593–610 (1928).
38. J.F. Heiss, I. Coull, 'The Effect of Orientation and Shape on the Settling Velocity of Non-isometric Particles in a Viscous Medium', *Chem. Eng. Progr.*, **48**(3), 133–140 (1952).
39. W. Walkenhorst, 'Untersuchungen über den Zusammenhang Zwischen Kornform und Sedimentationsgeschwindigkeit', *Staub*, **24**(8), 305–309 (1964).
40. W. Walkenhorst, 'Modellversuche zur Bestimmung des Dynamischen Formfaktors Nichtisometrischer Teilchen', *Staub*, **36**(4), 149–155 (1976).
41. E.S. Pettyjohn, E.B. Christiansen, 'Effect of Particle Shape on Free-settling Rates of Isometric Particles', *Chem. Eng. Progr.*, **44**(2), 157–172 (1948).
42. K.C.R. Chowdhury, W. Fritz, 'Sinkversuche mit isometrischen Teilchen in Flüssigkeiten', *Chem. Eng. Sci.*, **11**, 92–98 (1959).
43. C. Bernhardt, 'Sedimentation of Nonspherical Particles', *Particle & Particle Syst. Character.*, **8**, 209–214 (1991).
44. H. Horvath, 'Das Fallverhalten nicht-kugelförmiger Teilchen', *Staub*, **34**(7), 251–256 (1974).
45. R. Polke, R. Rieger, 'Partikelgrößenanalysen $<10\,\mu m$, Meßmethoden und Schwierigkeiten', *Chem.-Ing.-Tech.*, **50**(3), 149–154 (1978).
46. H. Wadell, 'The Coefficient of Resistance as a Function of Reynolds Number for Solids of Various Shapes', *J. Franklin Inst.*, **217**, 459–490 (1934).
47. H. Heywood, 'The Scope of Particle Size Analysis and Standardization', Proceedings, Symposium on PSA, London, February 1947, Suppl. to Trans. Inst. Chem. Eng., **25**, 14–24, 1947.
48. C. Bernhardt, 'Factors Causing Apparent Size Distributions in Sedimentation Analysis', Proceedings, Conference on PSA, Loughborough, 477–487, 1991.
49. M. Mason, W. Weaver, 'The Settling of Small Particles in a Fluid', *Phys. Rev. Ser.*, **23**(2), 412–426 (1924).
50. R. Fürth, 'Über ein Problem der Diffusion im Schwerefeld 1', *Z. Phys.*, **40**, 351–363 (1926).
51. R. Fürth, 'Über ein Problem der Diffusion im Schwerefeld 2', *Z. Phys.*, **45**, 83–85 (1927).
52. T. Svedberg, H. Rinde, 'The Ultracentrifuge, a New Instrument for the Determination of Size and Distribution of Size of Particles in Amicroscopic Colloids', *J. Am. Chem. Soc.*, **46**, 2677 (1924).
53. S. Berg, 'Determination of Particle Size Distribution by Examining Gravitational and Centrifugal Sedimentation According to the Pipet Method and with Divers', Symposium on Particle Size Measurement, June 1958, Boston, ASTM Special Techn. Publ. 234, 143–171, Vol. 2, 1959.
54. H.S. Chung, R. Hogg, 'The Effect of Brownian Motion on Particle Size Analysis by Sedimentation', *Powder Technol.*, **41**(3), 211–216 (1985).

55. T. Allen, 'Sedimentation Techniques of Particle Size Measurement', Proceedings, Conference on Particle Size Analysis, Bradford, Sept., 24–45, 1985.
56. H. Brenner, J. Happel, 'Slow Viscous Flow Past a Sphere in a Cylindrical Tube', *J. Fluid Mech.*, **4**, 195–213 (1958).
57. T. Greenstein, J. Happel, 'Theoretical Study of the Slow Motion of a Sphere and a Fluid in a Cylindrical Tube', *J. Fluid Mech.*, **34**(4), 705–710 (1938).
58. B. Koglin, 'Experimentelle Untersuchungen zur Sedimentation von Teilchenkomplexen in Suspensionen', *Chem.-Ing.-Tech.*, **44**(8), 515–521 (1972).
59. R. Ladenburg, 'Über den Einfluß von Wänden auf die Bewegung einer Kugel in einer reibenden Flüssigkeit', *Annal. Phys.*, **23**(4), 447–458 (1907).
60. H.A. Lorentz, *Abhandl. über theoret Physik*, **1**, 23 (1906).
61. C. Bernhardt, R. Herrmann, *Die Partikelgrößenanalyse und ihre Apparaturen*, Ceramic Forum International, Das Keramikerjahrbuch, Wiesbaden, 54–72, 1998.
62. J. Raasch, H. Umhauer, 'Computation of the Frequency Distribution of Distances between Particles Randomly Distributed in a Fluid Flow', *Particle & Particle Syst. Character.*, **6**(1), 13–16 (1989).
63. N. Ekdawi, R.J. Hunter, 'Sedimentation of Disperse and Coagulated Suspensions at High Particle Concentrations', *Colloids Surfaces*, **15**(1/2), 147–159 (1985).
64. H. Bauer, H. Thiele, 'Bewegung von Partikelschwärmen', *Chem.-Ing.-Tech.*, **45**(13), 909–912 (1973).
65. B.H. Kaye, R.P. Boardman, 'Cluster Formation in Dilute Suspensions', Proceedings, Symposium on the Interactions between Fluids and Particles, London, Jun 17–21, 1962.
66. R. Johne, 'Einfluß der Konzentration einer monodispersen Suspension auf die Sinkgeschwindigkeit ihrer Teilchen', *Chem.-Ing.-Tech.*, **38**, 428–430 (1966).
67. A.L. Graham, R.B. Bird, 'Particle Clusters in Concentrated Suspensions', *Ind. Eng. Chem. Fundam.*, **23**(4), 406–425 (1984).
68. H. Rumpf, W. Alex, R. Johne, K. Leschonski, 'Korngrößenanalyse feiner Teilchen, eine kritische Betrachtung der Methoden', *Ber. Bunsenges. Phys. Chem.*, **71**(3), 253–270 (1967).
69. W. Kast, 'Die Störung der Sedimentationsanalyse durch das Umströmen von Körpern im Sedimentationsgefäß', *Staub*, **20**(7), 205–211 (1960).
70. H. Rumpf, 'Untersuchungen zur Genauigkeit der Kornanalyse', *Staub*, **20**(8), 253–266 (1960).
71. W. Alex, 'Was gehört zu einer Pipette-Apparatur für die Teilchengrößenanalyse', *GIT-Fachzeitschr. für das Laboratorium*, **14**(6), 637–646 (1970).
72. R. Polke, M. Schäfer, N. Scholz, 'Preparation Technology for Fine Particle Measurement', *Particle & Particle Syst. Charact.*, **8**(1), 1–7 (1991).
73. H. Rumpf, W. Herrmann, 'Eigenschaften, Bindungsmechanismen und Festigkeit von Agglomeraten', *Aufbereitungstechn.*, **11**, 117–127 (1970).
74. R.R. Irani, F.C. Callis, *Particle Size; Measurement: Interpretation and Application*, John Wiley & Sons, New York, 1963.
75. G.D. Joglekar, B.R. Marathe, 'Liquids for Particle Size Analysis by Sedimentation Method', *J. Sci. Ind. Res.*, **17A**, 197–203 (1958).
76. B.V. Derjagin, N.V. Čuraev, V.M. Muller, *Poverchnostnye Sily*, Nauka, Moscow, 1985.
77. P.C. Hiemenz, *Principles of Colloid and Surface Chemistry*, Marcel Dekker, New York, 1997.
78. R.J. Hunter, *Zeta Potential in Colloid Science; Principles and Applications*, Academic Press, 1981.
79. G. Ney, *Zeta-Potentiale und Flotierbarkeit von Mineralen*, Springer-Verlag, Berlin, 1973.
80. G.D. Parfitt, *Dispersion of Powders in Liquids*, Elsevier Science, Amsterdam, 1969.
81. H. Sonntag, K. Strenge, *Koagulation und Stabilität disperser Systeme*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1970.
82. H. Krupp, 'Particle Adhesion – Theory and Experiment', *Adv. Colloid Interface Sci.*, **1**, 111–239 (1967).
83. B. Koglin, 'Systematik der Dispergiermittel', *Chem.-Ing.-Tech.*, **46**(17), 720–726 (1974).
84. C. Bernhardt, 'Preparation of Suspensions for Particle Size Analysis. Methodical Recommendations, Liquids and Dispersing Agents', *Adv. Colloid Interface Sci.*, **29**(12), 79–139 (1988).
85. C. Bernhardt, 'Koagulation als Fehlerursache bei der Sedimentationsanalyse', Proceedings, 1. European Symposium Partikelmeßtechnik, Nürnberg, Sept. 1975, DECHEMA Monographien Bd., **79**, 251–266, 1976.
86. O. Nagel, H. Kürten, 'Untersuchungen zum Dispergieren im turbulenten Scherfeld', *Chem.-Ing.-Tech.*, **48**(6), 513–519 (1976).
87. R. Davies, M.B. Ranade, J. Puretz, D.C. Freshwater, 'Factors Affecting the Ultrasonic Dispersion of Submicron Particles for Size Analysis', Proceedings, 2. European Symposium Partikelmeßtechnik, Nürnberg, Sept., 508–531, 1979.
88. Th. Neeße, K. Domasch, K. Mühle, A. Ivanauskas, 'Zur Modellierung des Flockungsprozesses', *Freiberger Forschungshefte A*, **720**, 1–106 (1985).
89. H. Reichert, 'Desagglomeration in Scherströmungen bei hohen Feststoffkonzentrationen', *Chem.-Ing.-Tech.*, **47**, 313 (1975).
90. V. Šutilov, *Physik des Ultraschalls*, Berlin, 1984.
91. C. Bernhardt, J. Gottschalk, 'On the Disintegration of Agglomerates by Ultrasonics', Proceedings, 5th European Symposium Particle Characterization, Nürnberg, March, 643–654, 1992.

92. W. Süß, M. Hantke, 'Über die Dispergierung suspendierter Feststoffe durch Ultraschall', *Pharmazie*, **24**, 211–220 (1969).
93. P.A. Egorov, ''OMechanisme dejstvija ul'trazvuka na gidrofobnye kolloidy', *Kolloidnyj žurn.*, (4), 834–836 (1971).
94. V.M. Fridman, 'The Interaction Mechanism between Cavitation Bubbles and Particles of Solid and Liquid Phases', *Ultrasonics*, **10**, 156–161 (1972).
95. B. Koglin, 'Methods for the Determination of the Degree of Agglomeration in Suspensions', Proceedings, 1st International Conference on Particle Technology, Chicago, Aug., 272–278, 1973.
96. U. Gärtner, 'Über die Möglichkeit der Zerkleinerung suspendierter Stoffe durch Ultraschall', *Acustica*, (3), 124–128 (1953).
97. B.A. Agranat, *Ul'trazvukovaja Technologija*, Metallurgija, Moscow, 1981.
98. H. Hirosue, M. Tanii et al., 'A Convenient Method for Determining the Optimum Concentration of a Dispersant', *J. Soc. Powder Technol.*, **20**(1), 28–31 (1983).
99. B. Koglin, 'Assessment of the Degree of Agglomeration in Suspensions', *Powder Technol.*, **17**, 219–227 (1977).
100. A. Zahradnicek, 'Methoden zur Aerosolherstellung aus vorgegebenen Feststoffhaufwerken', *Staub*, **35**(6), 226–231 (1975).
101. K. Leschonski, S. Röthele, U. Menzel, 'A Special Feeder for Diffraction Pattern Analysis of Dry Powders', *Particle Character.*, **1**, 161–166 (1984).
102. A. Zahradnicek, F. Löffler, 'Desagglomeration von feinkörnigen Feststoffen in Gasströmungen', *Verfahrenstechnik*, **11**(4), 229–232 (1977).
103. H. Geßner, *Die Schlämmanalyse*, Akademische Verlagsgesellschaft, Leipzig, 1931.
104. W. Ostwald, *Grundriß der Kolloidchemie*, 2nd edition, Dresden, 1911.
105. F.-V. von Hahn, *Dispersoidanalyse; die Methoden der Teilchengrößenbestimmung und ihre theoretischen Grundlagen*, Theodor Steinkopff, Dresden u. Leipzig, 1928.
106. S. Odén, 'Eine Neue Methode zur Bestimmung der Körnerverteilung in Suspensionen', *Kolloid-Z.*, **18**, 33–47 (1916).
107. G.A. Wiegner, 'Über eine neue Methode der Schlämmanalyse', *Landwirtschaftl. Versuchsstationen Bd.*, **91**, 41, 79 (1919).
108. T. Svedberg, J.B. Nichols, 'Determination of Size and Distribution of Size of Particles by Centrifugal Methods', *J. Amer. Chem. Soc.*, **45**, 2910 (1923).
109. G.W. Robinson, 'A New Method for a Mechanical Analysis of Soils and other Dispersions', *J. Agric. Sci.*, **12**, 306–321 (1922).
110. D.S. Jennings, M.D. Thomas, W. Gardner, 'A New Method of Mechanical Analysis of Soils', *Soil Sci.*, **14**, 485–499 (1922).
111. G. Krauss, 'Ergänzender Bericht über eine dem Prager Bodenkundlichen Kongreß vorgetragene neue Methode der mechanischen Bodenanalyse', *Int. Mitt. f. Bodenkunde*, (13), 146–160 (1923).
112. M. Köhn, 'Beiträge zur Theorie und Praxis der mechanischen Bodenanalyse', *Landwirtschaftl. Jahrbuch Bd.*, **67**(1), 485–546 (1928).
113. A.H.M. Andreasen, 'Zur Kenntnis des Mahlgutes: Theoretische und experimentelle Untersuchungen über die Verteilung der Stoffmenge auf die verschiedenen Korngrößen in zerkleinerten Produkten', *Kolloidchemische Beihefte, Bd.*, **27**, 349–458 (1928).
114. G.J. Bouyoucos, 'The Hydrometer as a New and Rapid Method for Determining the Colloidal Content of Soils', *Soil Sci.*, **23**(4/5), 319 (1927).
115. S. Berg, 'Untersuchungen über Korngrößenverteilung', *Kolloidchemische Beihefte*, **53**, 149 (1941).
116. C. Bernhardt, 'Particle Size Analysis of Heterogeneous Powders by Sedimentation', *Particle Character.*, **1**, 121–126 (1984).
117. R. Dallendörfer, L. Langhammer, 'Eine Variation des Verfahrens nach Andreasen zur beschleunigten Bestimmung der Korngrößenverteilung keramischer Rohstoffe', *Hermsdorfer Techn. Mitt.*, (1), 4–7 (1960).
118. C. Bernhardt, E. Lippek, 'Einige Verbesserungen an herkömmlichen Sedimentationsanalysegeräten', *Silikattechnik*, **244**, 134–137 (1973).
119. H. Zecha, 'Photometrische Dispersitäts- und Mengenbewertung in der Teilchengrößenanalyse', Proceedings, Fachtagg. Granulometrie, Dresden, Dez., 209–239, 1979.
120. C. Orr, J.M. Dallavalle, *Fine Particle Measurement; Size, Surface and Pore Volume*, Macmillan, New York, 1959.
121. E.E. Bauer, 'Recent Developments in the Hydrometer Method as Applied to Soils', Proceedings, Symposium on Particle Size Measurement, June 1958, Boston, ASTM Special Technical Publication 234, 143–171, 1959.
122. B.A. Jarrett, H. Heywood, 'A Comparison of Methods for Particle Size Analysis', *Br. J. Appl. Phys.*, Suppl. 3, 21–28 (1954).
123. A. Casagrande, 'Die Aräometermethode zur Bestimmung der Kornverteilung von Böden u. anderen Materialien', Berlin, 1934.
124. R.F. Conley, 'Some Inherent Errors in Hydrometer Sedimentation Analysis', *Powder Technol.*, **3**(2), 102–106 (1969).
125. G. Mie, 'Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen', *Annal. Phys.*, **25**(3), 377–445 (1908).
126. H.C. van de Hulst, *Light Scattering by Small Particles*, J. Wiley & Sons, New York, 1957.

127. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, Academic Press, New York, 1969.
128. R. Weichert, 'Determination of Extinction Efficiency and Particle Size Distribution by Photosedimentation using Light of Different Wavelengths', Proceedings, 4th Particle Size Analysis Conference, Loughborough, Sept., 28, 35, 1981.
129. T. Allen, 'Determination of Size Distribution and Specific Surface of Fine Powders by Photoextinction Method', *Powder Technol.*, **2**, 133–140, 141–153 (1968/69).
130. H.R. Harner, J.R. Musgrave, 'A Pholoelectric Sedimentation Method for Particle Size Determination in the Subsieve Range', Proceedings, Symposium on Particle Size Measurement, June 1958, Boston, ASTM Special Tech. Publ., **234**, 172–179, 1959.
131. O. Telle, 'Kornverteilungsanalysen an Pulvern und Stäuben mit einem lichtelektrischen Verfahren', *Tonind.-Ztg.*, **76**(23/24), 369–373 (1952).
132. H.E. Rose, 'The Design of Photo-extinction Sedimentometers', *Engineering*, **169**, 350–351, 405–408 (1950).
133. A.I. Michaels, 'Turbidimetric Particle Size Distribution Theory: Applications to Refractory Metal and Oxide Powders', Proceedings, Symposium on Particle Size Measurement, Boston, June 1958, ASTM Special Techn. Publ. 234, 207–244, 1959.
134. R. Johne, M. Ramanujam, 'Genauigkeit der Kornanalyse mit dem Photosedimentometer', *Staub*, **23**(5), 269–278 (1963).
135. F.W. Locher, 'Korngrößenbestimmung in Suspensionen durch Messung der Lichtdurchlässigkeit', *Zement-Kalk-Gips*, **8**(9), 301–308 (1955).
136. D.G. Skinner, S. Boas-Traube, 'The Light Extinction Method of Particle Size Estimation', Proceedings, Symposium on Particle Size Analysis, London 1947, Supplement to Trans. Inst. Chem. Eng., **25**, 57–63, 1947.
137. G. Staudinger, M. Hangl, P. Pechtl, 'Quick Optical Measurement of the Sedimentation Characteristics of Particles', Proceedings, Ist World Congress Particle Technology, Part 1: Particle Characterization, Nürnberg, Apr., 311–322, 1986.
138. J.F. Brown, J.K. Skrebowski, Comments on B.A. Jarrett, H. Heywood, 'A Comparison of Methods for Particle Size Analysis', *Br. J. Appl. Phys.*, Suppl. 3, 21–28 (1954).
139. A.M. Gaudin, M.C. Fuerstenau, 'The Transviewer in X-rays to Measure Suspended Solids Concentration', *Eng. Mining J.*, **159**(9), 11 (1958).
140. W.P. Hendrix, C. Orr, 'Automatic Sedimentation Size Analysis Instrument', Proceedings, Conference on PSA, Bradford, Sept., 133–146, 1970.
141. J. Kalshoven, 'Fast and Automatic Sedimentation Analysis', Proceedings, Conference on PSA, Loughborough, Sept., 197–205, 1966.
142. V.P. Šašilov, L.I. Vladimirova, G.V. Pavlinskij, 'Pogloščenie rentgenovskogo izlučenija geterogennymi probami s nepreryvnym raspredeleniem zeren po krupnosti', *Zavod. Labor.*, **45**(11), 1003, 1008 (1979).
143. C.P. Ross, 'Particle Size Analysis by Gamma-ray Absorption', *Anal. Chem.*, **31**(3), 337 (1959).
144. L. Svarovsky, T. Allen, 'Performance of a New X-ray Sedimentometer', Proceedings Conference on Particle Size Analysis, Bradford, Sept., 147–157, 1970.
145. P.A. Webb, C. Orr, 'Analytical Methods in Fine Particle Technology', *Micromeritics*, Norcross, USA, 1997.
146. J. Brezina, 'Stratified Sedimentation Above Stokes' Range and its use for Particle Size Analysis', Proceedings, Conference on PSA, Bradford, Sept., 255–266, 1970.
147. B. Hoffmann, C. Bernhardt, 'Particle Size Analysis of Some Difficult Measurable Systems', *Particle Character.*, **3**(4), 163–167 (1986).
148. R.K. Herrmann, 'Erfahrungen mit dem Sedigraph 5000', *Keram. Z.*, **31**(5), 275–277 (1979).
149. A. Matuschka, G. Schönbrunn, 'Progress Report on Microgrit Size Analysis of Aluminium Oxide and Silicon Carbide', *Sprechsaal*, **115**(2), 145–153 (1982).
150. H. Nieswand, 'Probleme bei der Korngrößenanalyse von keramischen Rohstoffen mit einem automatischen Korngrößenanalysator', *Freiberger Forschungshefte A*, **633**, 157–168 (1980).
151. E. Köster, *Granulometrische und morphometrische Meßmethoden an Mineralkörnern, Steinen und sonstigen Stoffen*, Ferdinand Enke Verlag, Stuttgart, 1964.
152. H. Rumpf, W. Alex, 'Die Trennkurven des Dekantierens', *Powder Technol.*, **2**, 52–57 (1968/69).
153. W. Batel, *Einführung in die Korngrößenmeßtechnik*, 3rd edition, Springer-Verlag, Berlin, 1971.
154. W. Bostock, 'A Sedimentation Balance for Particle Size Analysis in the Sub-sieve Range', *J. Sci. Instr.*, **29**, 209–211 (1952).
155. H.E. Rose, 'Calculation of Particle-size Distribution from Sedimentation Observations', *Nature*, **179**, 774–775 (1957).
156. D. Bachmann, 'Die Sedimentationswaage, ein neues schreibendes Gerät zur Feinheitsanalyse', *DECHEMA-Monographien*, **31**, 23–51 (1959).
157. H. Gerstenberg, 'Ermittlung der Korngrößenverteilung fester Stoffe mit der Sedimentationswaage', *DECHEMA Monographien*, **31**, 52–60 (1959).
158. J.G. Rabatin, R.H. Gale, 'Determination of Particle Size with a Simple Recording Sedimentation Balance', *Anal. Chem.*, **28**(8), 1314–1316 (1956).
159. M. Hirota, H. Kitakaze, K. Seki, 'Automatisches Gerät zur Aufzeichnung der im Sedimentationsverfahren bestimmten Kornverteilung' (in Japanese), *J. Ceram. Assoc., (Japan)*, **70**(2), 49–52 (1962).
160. A.M. Mironov, N.V. Čuraev, 'Dispersionnyj analiz pri pomošči vesovogo avtomatičeskogo sedimentometra', *Kolloidn. žurn.*, **29**(1), 132–136 (1967).
161. A.P. Tichonov, N.N. Cjurupa, 'Pribor dlja Sedimentacionnogo Analiza s Avtomaticeskoj Zapis'ju', *Zavod. Labor.*, **37**(1), 117, 118 (1971).

162. K. Matsumoto, Y. Mori, 'A Trial Production of Electric Auto-recording Type Small Amount Sedimentation Balance', *J. Res. Assoc. Powder Technol.*, **9**(5), 312–315 (1972).
163. A. Bürkholz, 'Korngrößenbestimmung auf der Schwerkraftwaage mit herkömmlichen und verbesserten Sedimentationsgefäßen', *Staub*, **32**(1 1), 442–446 (1972).
164. K. Leschonski, W. Alex, 'Theoretical and Experimental Study of Sedimentation Balances', Proceedings, Conference on Particle Size Analysis, Bradford, Sept., 236, 254, 1970.
165. W. Kowalski, 'Sedimentation Balance–Some Aspects of Measurement Result Approximation', Proceedings, 4th European Symposium Particle Characterization, Nürnberg, April, 271–286, 1989.
166. C.J. Stairmand, 'A New Sedimentation Apparatus for Particle Size Analysis in the Sub-sieve Range', Proceedings, Symposium on Particle Size Analysis, London, Febr. 1947, Supplement to Trans. Inst. Chem. Eng., **25**, 128–134, 1947.
167. W.H. Hardwick, B.J. Laundy, 'Particle Size by Beta-backscattering', Proceedings, Conference on PSA, Loughborough, Sept., 206, 211, 1966.
168. G. Curzio, 'Particle Size Determination by Beta-ray Absorption', *Powder Technol.*, **9**, 121–124 (1974).
169. M. Lange, Th. Gast, 'Sedimentationsanalyse mit Manometrischer Messung im Schwerefeld', *Staub*, **37**(4), 141–147 (1977).
170. P. Krischker, Th. Gast, 'Automatische Teilchengrößenanalyse mittels manometrischer Messung im Schwerefeld', Proceedings, 2. European Symposium Partikelmeßtechnik, Nürnberg, Sept., 532–543, 1979.
171. H. Rumpf, 'Über die Feinheitsbestimmung von Technischen Stäuben', *Staub,* **25**(1), 15–21 (1965).
172. W. Bennert, 'Eine einfache Methode der Korngrößenanalyse', Freiberger Forschungshefte A, **553**, Leipzig, 247–258, 1976.
173. G. Schultz, 'Korngrößenanalyse der Komponenten von Haufwerken durch einmalige Dekantationen', *Chem. Tech.*, **29**(12), 683–684 (1977).
174. R. Davies, B.H. Kaye, 'The Effect of Cluster Instability in Suspensions on Particle Size Analysis Measurements with the Cahn Balance', Proceedings, Conference on PSA, Bradford, Sept., 207, 222, 1970.
175. E. Muschelknautz, 'Über die Entwicklung einer Fliehkraftsedimentationswaage', Proceedings, 1. European Symposium Partikelmeßtechnik, Nürnberg, Sept. 1975, DECHEMA Monographien, **79**, 267–277, 1976.
176. O. Lauer, *Feinheitsmessungen an technischen Stäuben*, Alpine AG, Augsburg, 1963.
177. F.S. Eadie, R.E. Payne, 'Particle Size Distribution Analysed Quickly, Accurately', *Iron Age*, **174**(10), 99–103 (1954).
178. P.A. Kouzov, *Osnovy analiza dispersnogo sostava promyslennych pylej i izmel'čennych materialov*, 3rd edition, Chimija, Leningrad, 1987.
179. C.G. Samartzopoulos, 'Droplet Size Analyzer for Liquid Sprays by Sedimentation', Proceedings, Conference on Particle Size Analysis, Loughborough, Sept., 166–187, 1966.
180. J.R. May, 'Particle Size Measurement with a Micromerograph', *Min. Chem. Eng. Rev.*, **53**(12), 46–49 (1961).
181. A.C. Bryant, D.S. Freeman, F.L. Tye, 'An Appraisal of the Micromerograph for Particle Size Analysis of Manganese Dioxide', Proceedings, Conference on PSA, Loughborough, Sept., 154–165, 1966.
182. G.S. Chodakov, Ju.P. Judkin, 'Sedimentacionnyj analiz vysokodispersnych sistem', Chimija, Moscow, 1980.
183. W.F. Sullivan, A.E. Jacobsen, 'Sedimentation Procedures for Determining Particle Size Distribution', Proceedings Symposium on Particle Size Measurement, Boston, Jun 1958, ASTM Spec. Techn. Publ., 234, 98–116, 1959.
184. M. Fautz, 'Analytische Lösungen für Teilchengrößenanalysen mit Sedimentation im Zentrifugalfeld', *Chem.-Ing. Tech.*, **50**(1), 49 (1978).
185. F.H. Norton, S. Speil, 'The Measurement of Particle Sizes in Clays', *J. Am. Ceram. Soc.*, **21**, 89 (1938).
186. W. Alex, 'Korngrößenanalyse durch Sedimentation im Zentrifugalfeld', *Freiberger Forschungshefte A*, Leipzig, 27–34, Vol. 484, 1971.
187. H.J. Kamack, 'Particle Size Determination by Centrifugal Pipet Sedimentation', *Anal. Chem.*, **23**(6), 844–850 (1951).
188. L. Svarovsky, J. Svarovska, 'Particle Size Distribution Measurements by Homogeneous Sedimentation in a Disc Centrifuge', Proceedings, European Symposium Partikelmeßtechnik, Nürnberg, Sept. 1975, DECHEMA Monographien Bd, **79**, 293, 308, 1976.
189. A. Romwalter, M. Vendl, 'Die Ermittlung der Körnungskurve aus Schleuderversuchen', *Kolloid-Z.*, **72**, 1 (1935).
190. H.E. Robinson, S.W. Martin, 'Beaker-type Centrifugal Sedimentation of Subsieve Solid–Liquid Dispersions', *J. Phys. Chem.*, **52**, 848 (1948); **53**, 860 (1949).
191. P.J. Lloyd, B. Scarlett, I. Sinclair, 'Particle Size Analysis by Homogeneous Centrifugal Sedimentation', Proceedings Conference on Particle Size Analysis, Bradford, Sept., 267–275, 1970.
192. C. Brown, 'Particle Size Distribution by Centrifugal Sedimentation', *J. Phys. Chem.*, **48**, 246 (1944).
193. J.K. Donoghue, W. Bostock, 'A New Technique for Particle Size Analysis by Centrifugal Sedimentation', *Trans. Inst. Chem. Eng.*, **33**, 72–77 (1955).
194. R. Weichert, 'Auswerteverfahren für die Sedimentationsanalyse im Zentrifugalfeld', Proceedings, Fachtagg. Granulometrie, Dresden. Dez., 164–177, 1979.
195. R.J. Bailey, R.W. Lines, W.M. Wood, 'Applications of the Disc Centrifugal Photosedimentometer', Proceedings, Conference on Particle Size Analysis, Loughborough, Sept., 222–241, 1966.

196. B. Scarlett, M. Rippon, P.J. Lloyd, 'The Two-layer Method of Centrifugal Particle Size Analysis', Proceedings, Conference on Particle Size Analysis, Loughborough, Sept., 242–257, 1966.
197. K. Brugger, 'The Particle Size Determination of Pigments with the Disc Centrifuge', *Powder Technol.*, **13**, 215–221 (1976).
198. H. Coll, C.G. Searles, 'Particle Size Analysis with the Joyce–Loebl Disc Centrifuge: A Comparison of the Line-start with the Homogeneous-start Method', *J. Colloid Interface Sci.*, **115**(1), 121–129 (1987).
199. B.H. Kaye, N.I. Robb, 'Rapid Response Characterization of Fine Particle Suspensions using the Disc Centrifuge in the Homogeneous Start Mode', Proceedings, 4th European Symposium Particle Characterisation, Nürnberg, April, 129–136, 1989.
200. C. Slater, L. Cohen, 'A Centrifugal Particle Size Analyzer', *J. Sci. Instrum.*, **39**(12), 614–617 (1962).
201. A.K. Gupta, 'Particle Size Analysis by Sedimentation in a Long arm Centrifuge', *J. Appl. Chem.*, **99**, 487–489 (1959).
202. T. Allen, 'Review of Sedimentation Processes and Techniques', Proceedings, 2. European Symposion Partikelmeßtechnik, Nürnberg, Sept., 479–507, 1979.
203. T. Allen, M.A. Khalili, 'An Investigation of the Photocentrifugal Method of Particle Size Measurement', Proceedings, Conference on PSA, Loughborough, Sept., 15–27, 1981.
204. B.H. Kaye, M.R. Jackson, 'Size Analysis of Latex Emulsions using a Centrifugal Disc Photosedimentometer', *Powder Technol.*, **1**(2), 81–88 (1967).
205. T. Allen, L. Svarovsky, 'The Ladal X-ray Centrifugal Sedimentometer', *Powder Technol.*, **10**, 23–28 (1974).
206. T. Allen, 'The Du Point/Brookhaven Scanning X-ray Disc Centrifuge (B1-XDC)', Proceedings, Conference on PSA, Loughborough, Sept., 498–513, 1991.
207. H.A. Böwing, Th. Gast, 'Korngrößenanalyse im Bereich kleiner Durchmesser mit einer neuen Sedimentationszentrifuge', *Chem.-Ing.-Tech.*, **43**(8), 523–528 (1971).
208. O. Lauer, 'Experience Obtained on a Granulometric Centrifuge, Working on the Manometric Principle', Proceedings, Conference on Particle Size Analysis, Loughborough, Sept., 1–14, 1981.
209. G. Jimbo, J. Tsubaki et al., 'Development of a New Measuring Method for Size Distribution of Fine Powders using Pressure Difference Sensor', *J. Soc. Powder Technol.*, **23**(6), 430–436 (1986).
210. A. Bürkholz, W. Richter, E. Muschelknautz, 'Neue Zentrifuge zur Korngrößenbestimmung im Submikronbereich', *Chem.-Ing.-Tech.*, **45**, 680–684 (1973).
211. V.L. Tarasov, G.S. Chodakov, 'Centrobežnyj sedimennometr dlja granulometričeskogo analiza vysokodispersnych Sistem', *Kolloid. Z.*, **34**(6), 957, 959 (1972).
212. N. Kaya, T. Yokoyama, 'Particle Size Measurement Based on Detection of Unbalance by Centrifugal Sedimentation', Proceedings, Ist World Congress Particle Technology, Part 1: Particle characterization, Nürnberg, April, 301–309, 1986.
213. M.H. Pahl, G. Schädel, H. Rumpf, 'Zusammenstellung von Teilchenform-beschreibungsmethoden', *Part 1: Aufbereitungstechnik*, **14**(5), 257–264 (1973); *Part 2: Aufbereitungstechnik*, **14**(10), 672–683 (1973); *Part 3: Aufbereitungstechnik*, **14**(11), 759–764 (1973).

Sieving in Particle Size Analysis

Karla L. Dishman
P.O. Box 475, Kenova, USA

1 Introduction 1
2 Theory and Operating Principles 1
3 Method Description 1
3.1 Procedures 1
3.2 Sample Preparation 2
3.3 Apparatus 2
4 Interpretation of Results 3
4.1 Calculations 3
4.2 Data Representation 3
5 Effect of Particle Shape 4
Abbreviations and Acronyms 4
Related Articles 4
References 5

Sieve analysis is used to obtain the particle size distribution of a solid material by determining the amount of powder retained on a series of sieves with different sized apertures. A sample is added to the top of a nest of sieves arranged in decreasing size from top to bottom. As the sieves are vibrated, the sample is segregated onto the different sized sieves. The weight of sample retained on each sieve is then used to determine the particle size distribution as well as the mean diameter of the sample.

1 INTRODUCTION

The use of sieves is the oldest and most common technique used for particle size analysis. It is generally considered to be one of the most simple, reproducible, and inexpensive methods available. Sieving can be performed on particles ranging in size from 5 μm to 125 mm. One of the main disadvantages of sieve analysis is the effect of particle shape on results. Sieving measures the smallest diameter that can pass through the opening. With irregular-shaped particles, comparison between particle size methods can be difficult. Even though comparisons of methods are not always necessary, it is often useful to obtain information about particles using several techniques based on different principles.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 THEORY AND OPERATING PRINCIPLES

Sieve analysis is performed by segregating particles into different sized fractions on a series of sieves. The sample is placed on a nest of sieves with mesh openings decreasing in size from top to bottom. The sieves are then vibrated causing the sample to fall through the openings to a series of fractions on the individual sieves. The amount of material retained on each sieve is used to calculate the particle size distribution of the sample.[1–3]

3 METHOD DESCRIPTION

3.1 Procedures

Because sieving is used for many types of materials, it is impossible to specify a single procedure. Materials vary in size, shape, density, moisture content, cohesiveness, attritability, and so on. Therefore, general procedures for dry and wet sieving are given as guidelines. American Society for Testing and Materials (ASTM) standard methods can be referenced for information on a specific type of sample.[2]

3.1.1 Dry Sieving

The sieving procedure involves four basic steps: (1) sieve selection and set-up, (2) addition of sample, (3) sample fractionation, and (4) weighing retained fractions.[2]

3.1.1.1 Sieve Selection and Set-up Sieves are selected based on the size of the material to be characterized. The number of sieves should be minimized but give an accurate representation of the particle size distribution of the material. The sieves should be nested in a vertical stack with the coarsest sieve at the top and the solid pan at the bottom.

3.1.1.2 Addition of Sample The representative sample is dried to constant weight, when possible, and weighed to an accuracy of 0.1%. The weighed sample is then placed on the top sieve and covered.

3.1.1.3 Sample Fractionation The nest of sieves is shaken by vibration, manually, or with air pulses of sonic frequency for a specified time. A satisfactory ending time is reached when an additional one minute of sieving does not change the weight on the sieves by more than 1%.

3.1.1.4 Weighing Retained Fractions The material on each sieve is weighed to determine the amount in each particle size fraction. For small sample sizes, the material should not be removed from the sieve. The weight in each particle size fraction can be determined by difference

between the ending weight of the sieve with the sample fraction and the tare weight of the sieve.

3.1.2 Wet Sieving

Wet sieving is similar to dry, however, liquid which does not dissolve the material is continually added. The wet sieving process is particularly beneficial for separation of fine particles by eliminating electrostatic and surface forces hindering particle passage under dry conditions. Typically, wet sieving is used for materials originally suspended in a liquid or for powders that agglomerate during dry sieving. Several disadvantages of wet sieving include difficulty in obtaining accurate weight measurements, error due to swelling of particles, and increased time required for drying before weighing.

Several techniques for wet sieving have been found to be effective. In general, the sample is dispersed in a liquid and placed on the sieves. Additional liquid is then added to the residue on the sieves. Acceleration of the rate of the particles moving through the sieves is possible by placing the sieves on a vibrating apparatus and/or applying vacuum to the sieves.[1,2]

An automated wet sieving instrument has also been developed. Particles are initially placed on the finest sieves and fluidized by water. The particles that pass through the sieve are moved to a filter and weighed. After the weight remaining on the sieve has stabilized, the next larger sieve is brought into use. The procedure is continued for a series of sieves. The individual weights are then converted to weight percent data and printed.[4]

3.2 Sample Preparation

Accurate sampling is a prerequisite for reliable sieve analysis. The primary cause of inconsistencies in test results is in not obtaining a sample that is representative of the entire material. Because of the wide range of materials characterized by the sieve method, it is not possible to specify a single method of sampling. Sampling should be based upon the type and form of material. One problem that occurs is segregation of particles within the sample. During transfer of material, either in a small container or railroad car, the particles tend to segregate. In ideal circumstances, the material should be well mixed and a core sample taken. If the sample quantity is small enough, it may be reduced to test sample size using a sample splitter or riffler.[2]

3.3 Apparatus

Sieve measurements are performed with minimal equipment. The apparatus is made up of a balance, set of sieves, and possibly a mechanical shaker.[1,2] The balance should be accurate to 0.1% of the sample weight.

Table 1 Standard sieve sizes

Mesh number	Sieve aperture (mm)	Mesh number	Sieve aperture (μm)
3.5	5.6	40	425
4	4.75	45	355
5	4.00	50	300
6	3.35	60	250
7	2.80	70	212
8	2.36	80	180
10	2.00	100	150
12	1.70	120	125
14	1.40	140	106
16	1.18	170	90
18	1.00	200	75
20	850	230	63
25	710	270	53
30	600	325	45
35	500	400	38

The sieves should be chosen based upon the particle size distribution and sample quantity.

A typical sieve is a shallow pan with a wire mesh bottom or an electroformed grid. Sieves are available ranging in size from 5 μm to 125 mm. However, sieves are sometimes referred to by mesh size. Mesh is the nominal number of apertures. Table 1 gives details of the relationship between mesh size and aperture. Sieves used for sieve analysis are standardized according to ASTM E11 for woven wire mesh sieves and ASTM E161 for electroformed sieves.[1,2]

Woven wire sieves have square openings and are sturdier and less expensive than electroformed sieves. Electroformed or precision sieves have special mesh made by electrodeposition of nickel rather than weaving of wire. The conical hole structure reduces particle entrapment allowing for sharper cuts in sizing. Typically, all openings in an electroformed mesh will be identical in size. These openings are either round, square or rectangular. Electroformed sieves are typically used in the very fine particle range (5–40 μm).

Several types of mechanical sieve shaker are available. One type simulates hand sieving using a circular motion combined with a tapping motion. Another type uses electromagnetically induced high speed short stroke vibrations to agitate the sample causing it to flow through the sieve. A third type uses air pulses of sonic frequency to shake the sieves.

For wet sieving, the dry sieving apparatus can be modified by adding a vacuum pump.[1,2] In addition, companies, such as Micromeritics, have developed automated wet sieving methods. Particle size separation is achieved in an oscillating and periodically advancing rotor containing sieves of various meshes.[4]

4 INTERPRETATION OF RESULTS

4.1 Calculations

The weight of material retained on each sieve and the weight of the original test sample are used to calculate percentages. Results are normally reported in the form of percentages, not actual weights. These data are then used to calculate mean diameters, commonly referred to as average particle size. Midpoint diameters cannot be calculated for the smallest and largest fractions since the smallest and largest particle diameters are not known.

The arithmetic mean is defined as the sum of particles in a particular interval (n_i) with midpoint diameter (d_i) divided by the total number of particles (Equation 1):[3]

$$\text{Arithmetic mean diameter} = \frac{\sum n_i d_i}{\sum n_i} \qquad (1)$$

The geometric mean diameter (Equation 2) is the logarithmic equivalent of the arithmetic average:[3]

$$\log \text{geometric mean diameter} = \frac{\sum n_i \log d_i}{\sum n_i} \qquad (2)$$

In sieving, these data can also be reported as mass mean diameter or geometric mass mean diameter. The mass diameters are calculated based upon the weight of the sample rather than number of particles.

4.2 Data Representation

Particle size distribution data are presented in both tabular and graphical formats. The tabular data show the amount of material in each of the measured fractions. In addition, cumulative percent smaller than (and larger than) are often reported. An example of tabular data is presented in Table 2. These data are then graphed and used to determine the mean particle size of the material. Other fractiles, such as 10% or 90% can also be determined using this method.

The weight percent in each fraction is shown graphically in Figure 1. The weight percent is shown as a function of the midpoint diameter of the fraction given in Table 2. For example, 18% of the material is in the 40–80 μm fraction, meaning 18% has been retained on the 40 μm sieve. In the graph, this 18% value is plotted versus a midpoint particle diameter of 60 μm. The data shown in Figure 1 indicate that the particles are skewed toward the smaller sized particles.

A second type of graph commonly used in particle size testing is shown in Figure 2. This graph shows the amount of material less than a given particle diameter. For a normal distribution, a straight line can be obtained when the particle size on a linear axis is presented on a cumulative probability scale. The cumulative distribution

Table 2 Tabular representation of sieve particle size data

Diameter range, (μm)	Weight in fraction	Percent in fraction	Cumulative percent smaller than (largest size)
0–20	0.40	1.00	1.00
20–40	6.40	16.00	17.00
40–80	7.20	18.00	35.00
80–120	6.00	15.00	50.00
120–160	5.00	12.50	62.50
160–200	4.00	10.00	72.50
200–220	3.20	8.00	80.50
220–240	2.70	6.75	87.25
240–280	1.80	4.50	91.75
280–320	1.40	3.50	95.25
320–360	1.00	2.50	97.75
360–400	0.50	1.25	99.00
400–420	0.40	1.00	100.00
>420	0.00	0.00	100.00

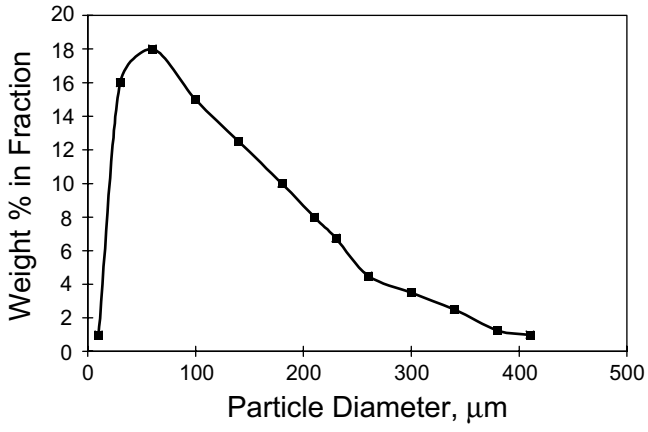

Figure 1 Amount of material as a function of particle size fraction.

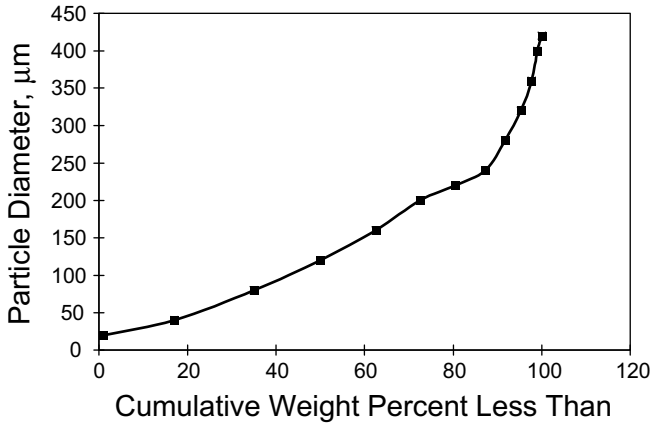

Figure 2 Cumulative percent less than graph.

data would then be symmetric at the 50% point. The arithmetic mean corresponds to the diameter at the 50% point.

The skewed data can be treated as a straight line if the arithmetic diameter scale is replaced with a logarithmic scale. Figure 3 illustrates the skewed data plotted on a log scale. To determine the geometric mean diameter, the diameter can be plotted as a function of cumulative weight percent less than using log-probability paper. The particle diameter extrapolated at the 50% value on the probability scale is the geometric mean diameter.

Graphical data representation is also useful in comparing samples. For example, Figure 4 shows three samples with the same mean diameter, but different types of distributions. The slope of the line is a measure of the breadth of the particle size distribution.[3]

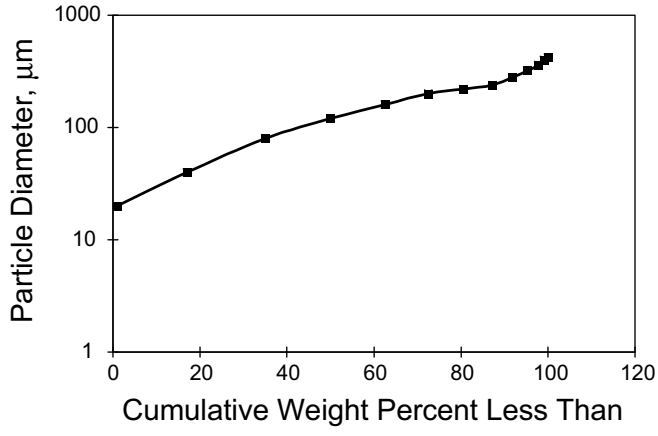

Figure 3 Use of logarithmic scale to linearize skewed data.

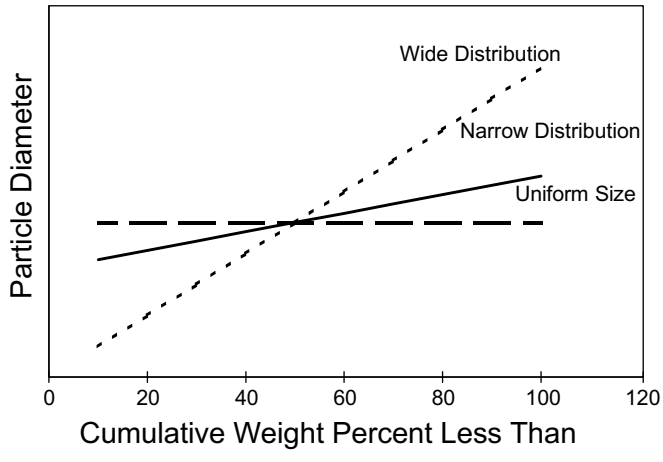

Figure 4 Types of particle size distribution.

5 EFFECT OF PARTICLE SHAPE

Particle size data are based on the assumption that the particles being analyzed are spheres with uniform density. However, many samples analyzed for particle size distribution are not spherical in particle shape. Particle shape features cause different dimensional response characteristics depending upon the measuring instrument. Because of the complexity of the problem, only limited success has been made in quantifying the effect of particle shape on particle size analyzes. Since the diameter of a nonspherical particle is based upon a measured geometric parameter, the distribution of particle sizes will shift depending upon the parameter measured. The effect of shape on the particle size distribution is minimized if the shape distribution is small relative to the particle size distribution. If the particles are close to the same size, any change in shape will broaden the measured size distribution.

Particle shape is an important parameter in describing differences in results from various particle size methods. For example, comparisons have shown differences in results from sieving and laser light diffraction to be a function of particle shape.[5–8] One of the most significant differences between laser light diffraction and sieving for particle size determination is the dimension of the particle measured. Light diffraction measures an average of all dimensions of the particle. With sieving, the measurement is based on the smallest diameter that can pass through the opening. With irregularly shaped particles this difference results in higher particle size measurements with laser light scattering than sieving.

Because of differences in particle size measurements as a function of particle shape, instruments are now being used to calculate particle shape factors.[9] Sieves are also being used to determine particle shape distribution by measuring the rate at which particles pass through a stack of identical sieves.[10]

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Light Scattering, Classical: Size and Size Distribution Characterization • Sedimentation in Particle Size Analysis

REFERENCES

1. R.R. Irani, C.F. Callis, *Particle Size: Measurement, Interpretation, and Application*, John Wiley & Sons, New York, 1963.
2. W.C. Mahlig, R.H. Smith, C.W. Ward, *Test Sieving Methods*, ASTM, Conshohocken, PA, 1985.
3. F.A. Zenz, D.F. Othmer, 'Particle and Powder Physical Characteristics, Chapter 3' in *Fluidization and Fluid-Particle Systems*, Reinhold, New York, 94–135, 1960.
4. C. Orr, Jr, 'Particle Size Characterization – A Review of Recent Technology', in *Particle Size Analysis*, ed. M.J. Groves, Heyden and Son, Philadelphia, 1978.
5. L.G. Austin, I. Shah, 'A Method for Inter-Conversion of Microtrac and Sieve Size Distributions', *Powder Technol.*, **35**, 271–278 (1983).
6. K.L. Dishman, P.K. Doolin, J.F. Hoffman, 'Comparison of Particle Size of Cracking Catalyst Determined by Laser Light Scattering and Dry Sieve Methods', *Ind. Eng. Chem. Res.*, **32**, 1457–1463 (1993).
7. N. Nathier-Dufour, L. Bougeard, M. Francoise-Devaux, D. Bertrand, F. Le Deschault de Monredon, 'Comparison of Sieving and Laser Diffraction for the Particle Size Measurements of Raw Materials Used in Foodstuff', *Powder Technol.*, **76**, 191–200 (1993).
8. W.H. Flank, 'Comparison of ASTM Round-Robin Data on Particle Size Using Three Different Methods', *Ind. Eng. Chem. Res.*, **26**, 1750–1753 (1987).
9. J.W. Novak, Jr, J.R. Thompson, 'Extending the Use of Particle Sizing Instrumentation to Calculate Particle Shape Factors', *Powder Technol.*, **45**, 159–167 (1986).
10. N.N. Clark, T.P. Meloy, 'Particle Sample Shape Description with an Automated Cascadograph Particle Analyzer', *Powder Technol.*, **54**, 271–277 (1988).

Surface Area and Pore Size Distributions

Takashi Takei
Tokyo Metropolitan University, Tokyo, Japan

1 Introduction **1**
2 Determination of Surface Area **1**
2.1 Gas Adsorption Method 1
2.2 Liquid Adsorption Method 3
2.3 Permeametry 4
2.4 Other Methods 5
3 Classification of Pore Sizes and Pore Structure **5**
4 Determination of Macropore and Mesopore Size **6**
4.1 Gas Adsorption Method 6
4.2 Mercury Porosimetry 8
4.3 Thermoporometry 9
4.4 Other Methods 9
5 Determination of Micropore Size **10**
5.1 Dubinin–Radushkevich Method 10
5.2 Horvath–Kawazoe Method 10
5.3 Other Methods 11
Abbreviations and Acronyms **11**
Related Articles **12**
References **12**

The determination methods of the surface area and pore size distribution (PSD) of a solid have been extensively described. The surface area and PSD are important as the basic properties for a powder and porous material. Nowadays the surface area and PSD are easily obtained by an automatic measuring instrument. However, there is no method which covers the complete range of a specific surface area and a pore size. All of the methods have their limits of measurement. In many cases the definitions of the surface area and pore size in each method differ from one another. The features and principles of the methods have been discussed mainly in order to select an appropriate method and to analyze data exactly.

1 INTRODUCTION

Determination of the surface area is important for characterization of solids, especially finely divided particles and porous materials. These materials have high specific surface areas, so the surface and interface phenomena (adsorption, adhesion, aggregation, chemical reaction and dispersion in liquids) are governed by the surface area. The determination of the surface area is necessary to understand such surface and interface phenomena on solids, and to control surface properties of solids. The measurement of the surface area is performed in various industrial fields: ceramics, pigments, fillers, catalysts, electronic materials, building materials, adsorbents, food materials, cosmetics, medicine, and so on. Several methods for determining the surface area are described in this article; however it is noticed that the definition of surface area differs from one method to another. Therefore the researcher should understand the principles of the determination methods and select an appropriate method.

Determination of the PSD of porous solids is important for porous structure analyses. Porous structures affect adsorption property, permeability, diffusion of gases and ions, thermal conductivity, strength, and so on. Many determination methods of the PSD have been proposed, and commercial apparatus is available. However, the researcher should note the following: most methods assume the simple pore model, for example the cylindrical pore model, but this is quite different from real pore structures; and all the methods have a lower and upper limit of detection for the pore size. In analogy with surface area determination, the selection or combination of methods and the understanding of methodology are important. Studies about the determination of the PSD are still being carried out. The current topics are available in the *Proceedings of the Characterization of Porous Solids*.[1–4]

2 DETERMINATION OF SURFACE AREA

2.1 Gas Adsorption Method

The determination of the surface area by gas adsorption is one of the most reliable methods, and is widely used for powders and porous materials. It is necessary for the adsorption method to estimate monolayer capacity (V_m) of an adsorbate and cross-sectional area (σ) of an adsorbed molecule. The specific surface area (S; the surface area of 1 g of the solid) is given by Equation (1):

$$S = V_m \sigma N_A \quad (1)$$

where V_m is expressed in moles per gram of the solid and N_A is the Avogadro number. Monolayer capacity is usually estimated from the adsorption isotherm by the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Brunauer–Emmet–Teller (BET) method.[5] The BET equation is shown in Equation (2):

$$V = V_m \frac{CX}{(1-X)(1-X+CX)} \tag{2}$$

where V is the amount of gas adsorbed, C is a constant value and X is the relative pressure: P/P_0 (P, equilibrium pressure and P_0, saturated vapor pressure). Usually Equation (2) can be rewritten for plotting experimental data as Equation (3):

$$\frac{X}{V(1-X)} = \frac{1}{V_m C} + \frac{(C-1)X}{V_m C} \tag{3}$$

The plot of $X/V(1-X)$ against X is called the BET plot, which is usually linear in the range of $0.05 < X < 0.35$. It should be noted that the relative pressure range in which the BET plot shows linearity depends on the kind of adsorbates and solids. The value of V_m and C are calculated from the slope $(C-1)/V_m C$ and intercept $1/V_m C$ of the BET plot. The adsorption isotherm of nitrogen and the BET plot are shown in Figure 1. The constant C is represented by Equation (4):

$$C \cong \exp \frac{Q_1 - Q_L}{RT} \tag{4}$$

where Q_1 is the isosteric heat of adsorption of molecules in the first layer and Q_L is the heat of condensation of the liquid adsorbate. The constant C represents the interaction between the adsorptive molecule and solid surface qualitatively. Thus the shape of the adsorption isotherm depends on the constant C. The adsorption isotherms deduced by the BET equation are shown in Figure 2.

The cross-sectional areas (σ) of adsorbed molecules have been calculated by various methods. Usually the σ-value is estimated from the liquid density (d) of the adsorbate at the working temperature, using Equation (5):[6]

$$\sigma = F \left(\frac{M}{N_A d} \right)^{\frac{2}{3}} \tag{5}$$

where F is the packing factor and M is the molecular weight of the adsorbate. The packing factor (F) is determined by the coordination number of adsorbed molecules on solid surfaces and in the adsorbed layer.[7] If the liquid structure of the adsorbate is the closest packing structure with 12 nearest neighbors, and the adsorptive molecules are adsorbed on the surface with 6 nearest neighbors in the close-packed hexagonal arrangement, the value of F is 1.081. Usually the adsorption gas used for the surface area determination is nitrogen, because nitrogen molecules can form a monolayer at 77 K on various solids. The σ-value for the nitrogen molecule is 0.162 nm^2. In the case of nitrogen adsorption, if V_m is

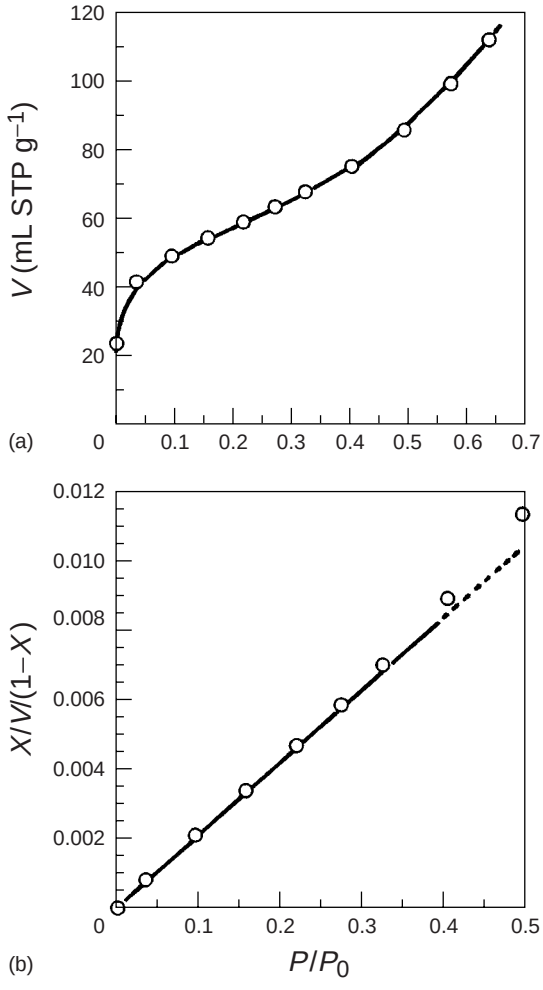

Figure 1 Adsorption of nitrogen on porous silica glass at 77 K: (a) adsorption isotherm; (b) BET plot.

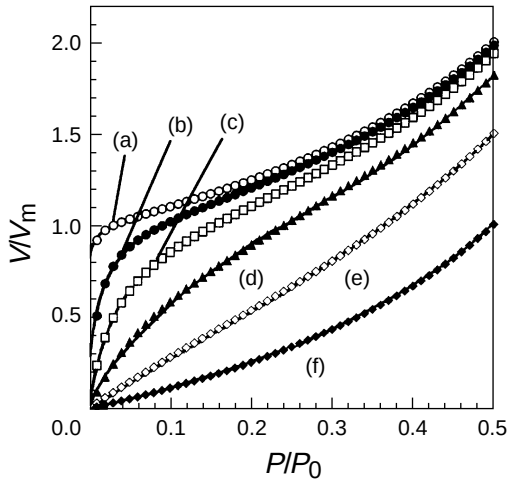

Figure 2 Dependence of the C-value in the BET equation on adsorption isotherms: (a) $C = 1000$; (b) $C = 100$; (c) $C = 30$; (d) $C = 10$; (e) $C = 3$; (f) $C = 1$.

Table 1 Average σ-values of adsorbed molecules

Formula	σ_{av} (nm^2)
N_2	0.162
Ar	0.147
Kr	0.202
Xe	0.232
CO_2	0.218
NH_3	0.140
n-C_4H_{10}	0.448
n-C_5H_{12}	0.492
n-C_6H_{14}	0.562

in milliliters at standard temperature and pressure per gram, the specific surface area can be estimated using Equation (6):

$$S(m^2\,g^{-1}) = 4.35\,V_m \tag{6}$$

McClellan and Harnsberger[8] have produced a table of the σ-values for about 106 compounds. The average σ-values of several molecules calculated from adsorption data are summarized in Table 1. Generally the surface areas determined by nitrogen adsorption differ from those estimated from adsorption of other gases. Several reasons for these differences are given: change in the σ-value due to orientation of the adsorbate molecule, the validity of the BET equation, and so on. One of the reasons is attributed to the difference in the accessibility of the adsorptive molecule to surface geometry. Consequently the surface areas depend on the size of probe molecules. Avnir et al.[9] have derived the fractal dimension (D) of solid surfaces from the adsorption of various gases on solid surfaces.

Usually nitrogen gas is used to measure the surface area. However, for the measurement of very small surface areas, adsorption of krypton[10–13] or xenon[14–16] gases at low temperature (77 K or 83 K) are used. The vapor pressure of these gases at 77 K is very low, so the correction of the vapor pressure is required for the thermal transpiration effect.[17]

A lot of commercial apparatus is available for surface area determination by gas adsorption measurement. This apparatus is classified into three categories: volumetric method, carrier gas method (dynamic method) and gravimetric method. Adsorption measurement by quartz crystal microbalance (QCM)[18,19] or surface acoustic wave (SAW)[20–22] devices are some of the new techniques for surface area determination. The principle of these techniques is based on the change in oscillation frequency of the quartz or the SAW oscillator, according to the change in surface mass density. The relationship between the change in oscillation frequency (Δf) of the oscillator and the change in surface mass density (ΔW) using the quartz oscillator is shown in Equation (7):

$$\Delta f = -f_0^2 \frac{\Delta W}{NA\rho} \tag{7}$$

where f_0 is the parent frequency, N is the frequency constant for AT-cut quartz (1.67×10^6 Hz mm), A is the area of the electrode, and ρ is the density of quartz ($2.648\,g\,cm^{-3}$). The sensitivity of a 10 MHz QCM is $5\,ng\,cm^{-2}$. The QCMs or SAW devices are capable of measuring nitrogen adsorption on thin films. Commercial volumetric or gravimetric adsorption apparatus cannot obtain nitrogen adsorption data of such low surface area materials. These techniques are applicable to estimate the pore structure analysis by gas adsorption (see section 4.1).

2.2 Liquid Adsorption Method

The principle of the determination of surface area by liquid adsorption is the same as that by gas adsorption. Dyes[23–28] or fatty acids[29–31] in solution are frequently used as a probe molecule for the surface area determination. The liquid adsorption method is easy to operate and does not need special apparatus compared to the gas adsorption method. However it is necessary to estimate the cross-sectional area of an adsorptive molecule and the effect of solvent. Because a cross-sectional area of an adsorptive molecule depends on the configuration of an adsorbed molecule, competitive adsorption of adsorptive molecules between solvent molecules occurs. So the solutes (adsorptives) and solvents should be carefully selected for the sample solid. The liquid adsorption isotherm is represented by a plot of the amount of adsorbed vs the equilibrium concentration of the adsorptive. Several equations for the liquid adsorption isotherm have been proposed: the Henry equation, the Langmuir equation, the Frumukin equation and the Freundlich equation. The Langmuir equation is frequently used to determine the monolayer capacity. The equation is expressed by Equation (8):

$$V = \frac{AV_mC}{1 + AC} \tag{8}$$

where A is a constant. The expression for plotting experimental data is given by Equation (9):

$$\frac{C}{V} = \frac{1}{V_mA} + \frac{C}{V_m} \tag{9}$$

The monolayer capacity is obtained from the slope of the plot of C/V against C. Figure 3 shows adsorption isotherms of several dyes on charcoal.[24] The cross-sectional areas of dye molecules[32] are shown in Table 2. In the case of dye adsorption, the formation of molecular aggregates in high concentration and the effect of

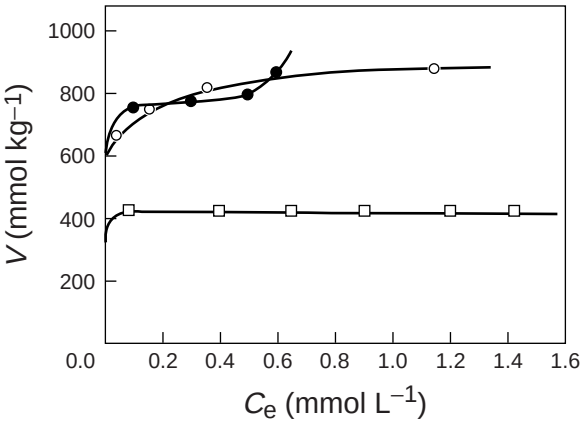

Figure 3 Adsorption isotherms for dyes on activated charcoal at 21 °C: (●) Methylene Blue; (○) Brilliant Basic Red; ( ) Victoria Pure Lake Blue BO. [Reproduced by permission of The Royal Society of Chemistry from C.H. Giles, A.P. D'Silva, *Trans. Faraday. Soc.*, **65**, 1943–1951 (1969).]

Table 2 Cross-sectional area and λ_{max} of dyes

Dye	Cross-sectional area (nm^2)	λ_{max} (nm)
Methylene Blue	1.20	661
Crystal Violet	2.25	588
Victoria Pure Lake Blue BO	2.75	619
Orange II	1.20	483

pH on adsorption should be noted. The concentration of dye solution is determined from the measurement of absorbance of the solution by ultraviolet/visible spectrometer. The dye adsorption method is applicable to measure low surface area ($\sim 0.01\,m^2\,g^{-1}$). However, accessibility of dye molecules to internal surfaces should be considered because the size of dye molecules is large compared to that of gas molecules, which are usually used as probe molecules for surface area determination.

In the case of fatty acid adsorption, steric acid in organic solution (ethanol, hexane, and benzene) is frequently used. The organic solvents should be dehydrated. The cross-sectional area of steric acid is assumed to be $0.208\,nm^2$ if molecules of steric acid are adsorbed perpendicular to the solid surfaces. The concentration of steric acid is estimated by the titration method.

The direct adsorption of ethylene glycol is used to determine the surface area of clay minerals.[28,33–35] The cross-sectional area of an ethylene glycol molecule is assumed to be $0.224\,nm^2$.

2.3 Permeametry

Surface area determination by permeametry is used for powder materials. This method is characterized by simple apparatus, easy operation and rapid measurement compared to the other method. Permeametry is based on fluid flow through a packed powder bed (porous media). The specific surface area is calculated from the Kozeny–Carman equation, shown in Equation (10):

$$S_w = \frac{1}{\rho}\left\{\left(\frac{\Delta P t A}{k\eta L V}\right)\left(\frac{\varepsilon^3}{(1-\varepsilon)^2}\right)\right\}^{\frac{1}{2}} \tag{10}$$

where S_w is the specific surface area of the sample powder, ρ is the true density of the sample powder, ΔP is pressure drop, t is time, A is the cross-sectional area of the powder bed, k is the Kozeny constant, which is usually assumed to be 5, η is the viscosity coefficient of the fluid, L is the length of the powder bed, V is the volume of fluid passed through powder bed during time t, and ε is the porosity of the powder bed. The surface area of the powder bed can be calculated from measurement of the flow rate V/t and pressure drop ΔP. Gases and liquids can be used as the flowing fluids, however, gases (air or He gas) are usually used, rather than liquids, in order to avoid a dissolution of samples and reconstruction of the powder bed. The many equations for permeametry based on the Kozeny–Carman equation have been derived to fit the experimental procedure, and many types of apparatus have been proposed.[36–41] The Blaine apparatus and the Fischer Sub-Sieve Sizer apparatus[42] are typical for air permeametry. This apparatus is commercially available. The surface area thus obtained means the external surface area and does not include internal surface: micropores and cracks. Equation (10) is applicable only in the viscous flow condition, so this analysis is limited to coarse powder particles ($\sim 2\,\mu m$).

When the particle size of the powder bed becomes smaller ($<\sim 1\,\mu m$), the gas flow mechanism changes from viscous flow to Knudsen flow, which is controlled by the rate of molecular diffusion. Carman modified Equation (10) by considering the Knudsen flow and derived the general equations shown in Equations (11) and (12):

$$\frac{V'L}{\Delta PA} = \left(\frac{P'}{5\eta S_v^2 RT}\right)\left(\frac{\varepsilon^3}{(1-\varepsilon)^2}\right) + \left(\frac{4Z}{3}\right)\left(\frac{v}{S_v}\right)\left(\frac{\varepsilon^2}{1-\varepsilon}\right) \tag{11}$$

$$\rho S_w = S_v \tag{12}$$

where V' is the velocity (moles per second) of fluid gas, P' is the average pressure in the powder bed, Z is a constant, and v is the rate of molecular diffusion. The first term of the right-hand side is the contribution of viscous flow and the second term is that of the Knudsen flow. The plot of $V'L/\Delta PA$ vs P' should be linear. The surface area calculated from the slope of the linear line corresponds to the external surface area. However, the surface area estimated from the intercept represents the total surface area of the powder sample. The constant Z depends on

the shape of the path. The Z value is experimentally determined to be 0.45 or 0.47. Several other studies on permeametry in the Knudsen flow condition have been reported.[37,38,43,44]

2.4 Other Methods

2.4.1 Microscopic Method

If the powder samples have no pores, the specific surface area of solids is estimated from the size of the primary particle as shown in Equation (13):

$$S = \frac{K}{\rho l} \tag{13}$$

where S is the specific surface area of the solid, K is the shape factor, ρ is true density of the particle, and l is the size of the particle. If the particle shape is cubic or spherical, the constant K is 6 and l represents the edge length of the cubic particle or the diameter of the spherical particle. Usually the size of the primary particle, l, is determined using optical or electron microscopy.

2.4.2 Heat of Immersion Method

When a solid surface is immersed in liquid which does not dissolve the solid, heat which is termed the heat of immersion is evolved. Harkins and Jura[45] developed an ''absolute method'' to estimate the surface areas of powders using the measurement of heat of immersion. When the sample powder with the liquid adsorbed layer is immersed in the adsorptive liquid, the heat of immersion is attributed to the total surface energy of the adsorptive liquid. The surface area (S) is represented by Equation (14):

$$S = \frac{Q}{h_{\mathrm{L}}} \tag{14}$$

where Q is the heat of immersion and h_{L} is the enthalpy per unit surface area of the adsorption liquid. In the case of water, a two-molecule-thick adsorbed layer on a solid surface is sufficient to measure[46] and h_{L} is 118 erg m^{-2}.

3 CLASSIFICATION OF PORE SIZES AND PORE STRUCTURE

Pores are classified into three types, based on their sizes, according to the IUPAC recommendation:[47]

Macropores	$d(w) > 50\,\mathrm{nm}$
Mesopores	$2\,\mathrm{nm} < d(w) < 50\,\mathrm{nm}$
Micropores	$d(w) < 2\,\mathrm{nm}$

where d and w mean a pore diameter and a pore width, respectively. Many methods for determing the PSD of porous materials have been proposed. However, there is no experimental technique covering the whole range of pore size. Hence, it is important to choose methods appropriate for the pore size. It should be noticed that each method has a detection limit for the pore size. The applicable range of pore size for PSD analysis methods described in this chapter is shown in Figure 4.

Figure 5 shows a model of pores in a solid. The pore structure is mainly classified into two categories: closed pores (a) and open pores (b–h). The open pores consist of several types: through pores (b) are open at two ends, blind pores (c) are open only at one end, independent pores (b), (c) do not connect with other pores and crossing pores (d) connect with other pores. From the viewpoint of pore shape, pores are classified as follows: cylindrical pores (e), conical pores (f), ink-bottle pores (g), slit pores (h), and so on. The consideration of the origin of pores is useful for the understanding of pore structure. The pores in zeolite, which is well-known as a

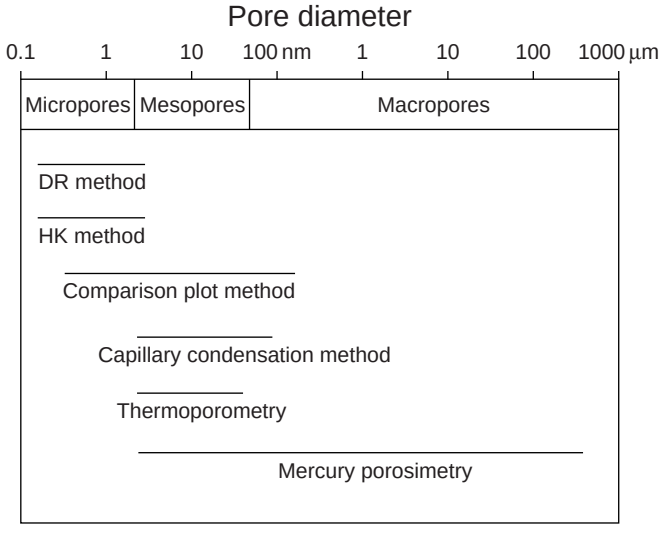

Figure 4 Applicable range of pore size to various pore analysis methods.

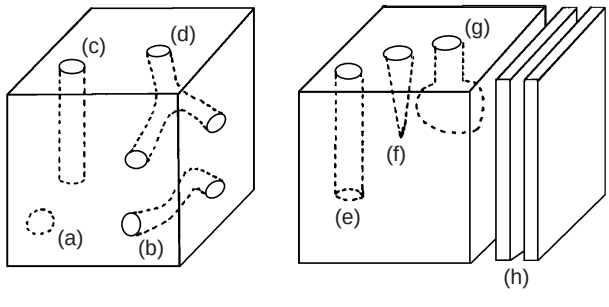

Figure 5 Models of a pore.

molecular sieve and catalyst, originate from the crystal structure. The slit pore in layered materials such as clay minerals and graphite is in this category. These pores are characterized by a regular structure. Some pores can be prepared by the removal of substances dispersed in a solid beforehand. Porous silica glass is prepared by the dissolution of a separated $Na_2O-B_2O_3$ phase in the SiO_2 matrix. The mesoporous silica MCM-41 is prepared by the micelle template.[48] Silica gel containing organic polymer can form pores by calcination. The size and structure of pores originated from the removal of additives in a solid depend on the dispersion condition and size of additives. The aggregation of small particles forms voids (pores) between the particles. These pore structures depend on the particle size and its distribution, particle shape and packing structure.[49] The sintered body and silica gel are attributed to this structure.

4 DETERMINATION OF MACROPORE AND MESOPORE SIZE

4.1 Gas Adsorption Method

4.1.1 Capillary Condensation Method

Generally the adsorption isotherm of gas on a mesoporous solid shows the Type IV isotherm in the IUPAC classification of adsorption isotherms.[47] The Type IV isotherm is shown in Figure 6. The adsorption isotherms on a mesoporous solid show a steep increase in the amount adsorbed due to capillary condensation, and a steep decrease in the adsorbed amount due to capillary evaporation. Consequently, the adsorption isotherm has a hysteresis loop. The capillary condensation method for the determination of PSDs is based on the relationship between the pressure at which capillary condensation occurs and the pore radius. This relationship is represented by the well-known Kelvin equation, shown in Equation (15):

$$r_k = \frac{-2\gamma V_m \cos\theta}{RT\ln(P/P_0)} \tag{15}$$

where r_k is the Kelvin radius, and γ and V_m are surface tension and molar volume of the condensate, respectively, T is a measuring temperature, θ is a contact angle between the condensate and the pore wall, and P/P_0 is the relative pressure at which capillary condensation occurs. Usually θ is assumed to be zero, so Equation (15) is simplified to Equation (16):

$$r_k = \frac{-2\gamma V_m}{RT\ln(P/P_0)} \tag{16}$$

When capillary condensation occurs, the thickness of a layer already adsorbed on the pore walls is given by t. Consequently the pore radius is given by Equation (17):

$$r_p = r_k + t \tag{17}$$

In the case of nitrogen adsorption, many studies on the thickness of the adsorbed layer have been reported.[50–57] Several equations for the relationship between the thickness of the adsorbed layer and relative pressure have been proposed for computation. Dollimore and Heal[58] recommended the use of Equation (18):

$$t\,(\text{nm}) = 0.35\left(\frac{-5}{\ln(P/P_0)}\right)^{\frac{1}{3}} \tag{18}$$

A number of calculation methods for the determination of PSDs based on the Kelvin equation (e.g. Barret–Joyner–Halenda (BJH) method, Dollimore–Heal (DH) method) have been proposed;[50,59–64] however, the principle of the computation procedure is almost the same. The calculation of PSDs is performed according to the following concept. The amount (ΔV) decreased in a desorption step is represented by the sum of the amount (ΔV_c) of capillary evaporation, and the amount (ΔV_m) of multilayer desorption from the pores in which capillary evaporation has already occurred, as shown in Equation (19):

$$\Delta V = \Delta V_c + \Delta V_m \tag{19}$$

The capillary condensation method is widely used for the determination of PSDs of mesoporous solids, and automatic measuring apparatus is commercially available. However, it should be noticed that this method has several

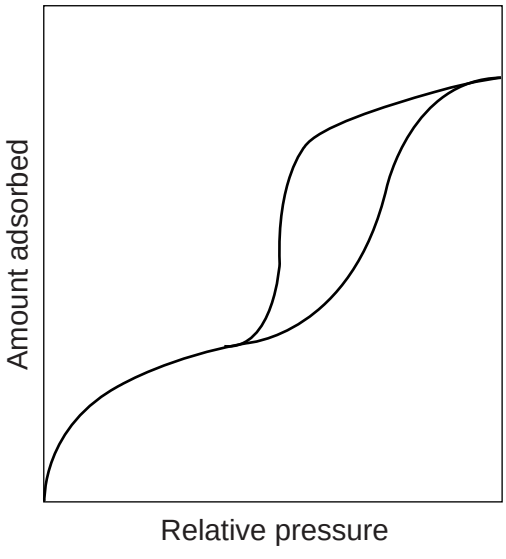

Figure 6 Model of Type IV adsorption isotherm.

problems. First, the simplified pore model (cylindrical pore or slit-shaped pore) is assumed. It is clear that this model is quite different from real porous structures. Second, as mentioned above, the adsorption isotherm on mesoporous solid shows a hysteresis loop. It should be noticed whether the calculation is performed on the adsorption or desorption branch of the hysteresis loop. Equation (15) is adopted for analysis of the desorption branch. When the adsorption branch is used for analysis, the Kelvin equation shown in Equation (20) is used:

$$r_k = \frac{-\gamma V_m}{RT \ln(P/P_0)} \tag{20}$$

This problem is very important, because the PSD depends on which branch of the hysteresis loop is used for the calculation. Further study, including the origin of the hysteresis loop, is required. The lower and upper pore size limits of the capillary condensation method are obscure. In a small pore it is expected that the physical properties of capillary condensed liquids are different from those of bulk liquids. Hence, the calculation of the pore radius by the Kelvin equation using the bulk values of surface tension and molar volume of condensates is inadequate for small pores. The criticisms on the application of the Kelvin equation in small pores have been reported.[65–72] The development of idealized mesoporous solids such as MCM-41,[48] and the progress in computer chemistry of adsorption simulation,[73,74] will be expected to refine the capillary condensation method in the small pore size region. However, the upper pore size limit depends on the accuracy of the adsorption measurement. The relationship between relative pressure and the pore radius is shown in Table 3. Clearly it is difficult to analyze macropore sizes with high precision compared with mesopore sizes.

4.1.2 Comparison Plot Method

The comparison plot method is one of the pore-structure analyses and reveals the existence of micropores and mesopores. Two kinds of comparison plots have been proposed: the t-plot developed by Lippens and de Boer[75] and the α_s-plot by Sing.[55,76] Both plots are obtained by comparison of the nitrogen adsorption isotherm on the sample with that on a standard solid. The standard solid should be equivalent to the sample solid and have no pores. In the case of the t-plot, the relationship between the relative pressure and the thickness of the adsorbed layer for the standard solid is required. The thickness (t) of the adsorbed layer of nitrogen is calculated from Equation (21):

$$t\,(\text{nm}) = \frac{0.354\,V}{V_m} \tag{21}$$

where V is the amount adsorbed, and V_m is the monolayer capacity. Standard data are shown in Table 4.[55] The t-plot of the sample is obtained by converting the relative pressure of the adsorption isotherm on the sample into the thickness of the adsorbed layer of nitrogen, according to the standard data. Models of t-plots are shown in Figure 7. When the sample has no pores, the t-plot gives a straight line. The specific surface area (S_t) is calculated from the slope of the t-plot using Equation (22):

$$S_t = \frac{1.547\,V}{t} \tag{22}$$

If the sample has mesopores, the t-plot deviates upwards from a straight line, due to capillary condensation. When

Table 3 Relationship between relative pressure and pore radius

P/P_0	P_e (torr)	r_p (nm)
0.999	759.2	945
0.995	756.2	191
0.990	752.4	96.3
0.950	722.0	19.9
0.900	684.0	10.2
0.800	608.0	5.2
0.700	532.0	3.5
0.600	456.0	2.6
0.500	380.0	2.0

Table 4 Standard adsorption data for nitrogen at −196 °C on non-porous hydroxylated silica. [Reproduced by permission of the Society of Chemical Industry from J.D. Carruthers, P.A. Cutting, R.E. Day, M.R. Harris, S.A. Mitchell, K.S.W. Sing, *Chem. Ind.*, 1772–1773 (1968).]

P/P_0	t (nm)	α_s	P/P_0	t (nm)	α_s
0.0010	–	0.26	0.28	0.471	0.88
0.005	–	0.35	0.30	0.482	0.90
0.01	–	0.40	0.32	0.492	0.92
0.02	–	0.50	0.34	0.503	0.94
0.03	–	0.55	0.36	0.514	0.96
0.04	–	0.58	0.38	0.524	0.98
0.05	–	0.60	0.40	0.535	1.00
0.06	–	0.61	0.42	0.540	1.01
0.07	–	0.63	0.44	0.556	1.04
0.08	–	0.65	0.46	0.567	1.06
0.09	0.353	0.66	0.50	0.589	1.10
0.10	0.364	0.68	0.55	0.615	1.14
0.12	0.375	0.70	0.60	0.653	1.22
0.14	0.391	0.73	0.65	0.690	1.29
0.16	0.401	0.75	0.70	0.738	1.38
0.18	0.412	0.77	0.75	0.787	1.47
0.20	0.428	0.80	0.80	0.867	1.62
0.22	0.439	0.82	0.85	0.967	1.81
0.24	0.449	0.84	0.90	1.28	2.4
0.26	0.460	0.86	0.95	1.98	3.7

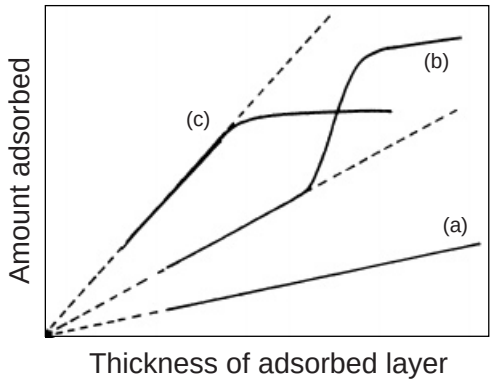

Figure 7 Model of t-plots: (a) non-porous solid; (b) mesoporous solid; (c) microporous solid.

the pore fills with the adsorbate, the slope of the t-plot is gradual. The external surface area of the sample is estimated from the slope of this part. If the sample has micropores, the t-plot again deviates downwards from a straight line, due to micropore filling. The disadvantage of the t-plot is the uncertainty of the thickness below the monolayer (<0.354 nm). The α_s-plot was developed to improve this disadvantage. The α_s-value is defined instead of the t-value, as shown in Equation (23):

$$\alpha_s = \frac{V}{V_{0.4}} \tag{23}$$

where $V_{0.4}$ is the amount adsorbed at $P/P_0 = 0.4$. Standard data are shown in Table 4.[55] The plot of V against α_s is called the α_s-plot. Basically, the interpretation of the α_s-plot is similar to that of the t-plot. The specific surface area (S_{α_s}) is represented by the slope of the α_s-plot, as shown in Equation (24):

$$S_{\alpha_s} = \frac{2.89\,V}{\alpha_s} \tag{24}$$

The α_s-plot is especially useful to determine the surface area of microporous materials.[77,78]

4.2 Mercury Porosimetry

Mercury porosimetry is based on the fact that excess pressure is required to cause the intrusion of mercury, which is the non-wetting liquid for most of the materials into the pore. The relationship between the pressure and the pore radius is expressed by the Washburn equation shown in Equation (25):

$$r = \frac{-2\gamma\cos\theta}{P} \tag{25}$$

where r is the pore radius, γ is the surface tension of mercury, θ is the contact angle between mercury and the pore wall, and P is the excess pressure. The surface tension of mercury is 485 dyn cm^{-2} and θ is usually assumed to be 140°. So Equation (25) can be expressed by Equation (26):

$$r\,(\mu\text{m}) = \frac{107.7}{P\,(\text{psi})} \tag{26}$$

One feature of mercury porosimetry is that it is capable of estimating the size of a wide range of pores (from mesopores to macropores). It is especially useful for determining the macropore size range, because there are few methods applicable to analysis of macropores. The lower-limit pore size and upper-limit pore size depend on the pressures which the apparatus can achieve. Nowadays commercially available apparatus can work at 60 000 psi, so the measurable pore size range is between 1.8 nm and 108 μm (if the minimum pressure is 1 psi). Usually a value of 140° as the contact angle is used for computation; however, it is predicted that the contact angle is affected by the surface chemical composition and the surface roughness.[79] The contact angle at atmospheric pressure will differ from that at high pressures. A change in contact angle by 10° from 140° causes a change in pore size of 16%. The problem of contact angle should be discussed in the future. Destruction or deformation of the sample is caused by compression. Reversible deformation should be corrected. However, irreversible deformation or destruction is seriously affected by the PSDs. The sample which amalgamates with mercury is unsuitable for this method. Contamination by amalgam causes a change in contact angle and surface tension of mercury. Intrusion–extrusion curves of mercury porosimetry are shown in Figure 8.[80] Intrusion–extrusion curves frequently display a hysteresis loop. Several reasons for the occurrence of the hysteresis loop have been proposed: the difference in contact angles at intrusion and

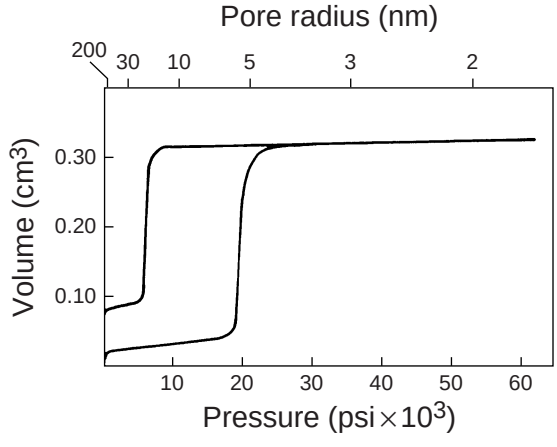

Figure 8 Intrusion–extrusion curves for a porous glass sample. [Reproduced by permission of Academic Press from S. Lowell, J.E. Shields, *J. Colloid Interface Sci.*, **83**, 273–298 (1981).]

extrusion;[80,81] the geometrical shape of individual pores (e.g. ''ink-bottle'' type pore); and the network effect of pores. The morphology[82] and pore connectivity[83] have been estimated from the hysteresis loop.

If a non-wetting liquid is used for the sample solids, the principle of mercury porosimetry can be applied to other liquids. Recently, water porosimetry for hydrophobic materials has been attempted.[84]

4.3 Thermoporometry

Thermoporometry is based on calorimetry of the liquid–solid transition of a condensate confined in pores. It is well known that the freezing and melting points of liquids confined in pores decrease with decrease in the pore size. Brun et al.[85] developed a method of determining PSDs from a solidification thermogram by means of theoretical relationships. The relationship between the depression of the freezing or melting point (ΔT) of water and pore radius (r_p) is given by Equation (27) for freezing and Equation (28) for melting:

$$r_p(\text{nm}) = -\left(\frac{64.67}{\Delta T} + 0.57\right) \quad (27)$$

$$r_p(\text{nm}) = -\left(\frac{32.33}{\Delta T} + 0.68\right) \quad (28)$$

The pore volume is estimated from the apparent energy (W_a) of freezing or melting of water. These values are obtained for freezing from Equation (29) and for melting from Equation (30):

$$W_a(\text{J g}^{-1}) = -5.56 \times 10^{-2}\Delta T^2 - 7.43\Delta T - 332 \quad (29)$$

$$W_a(\text{J g}^{-1}) = -0.155\Delta T^2 - 11.39\Delta T - 332 \quad (30)$$

Usually water is used as the condensate in pores, because water shows a larger latent heat at the liquid–solid transition as compared to other condensates. However, some organic liquids used as condensate for thermoporometry have been reported: benzene,[85] undecane,[86,87] decane,[88] and paranitrophenol.[89] These organic liquids are useful for the measurement of the hydrophobic materials and overambient measurements.[89] The dependence of melting points of water and benzene on pore radius is shown in Figure 9.[85] The advantage of thermoporometry is that pore analysis can be performed in wet environments. Thermoporometry is appropriate for the analysis of wet gels[88,90] and polymeric ultrafiltration (UF) membranes.[91–93] The pore shape or structure of aggregates of these materials changes with the drying process. Thermoporometry can be used for rigid porous materials.[94,95] Several methods of determining PSDs based on the liquid–solid transition have been proposed.[96–99] The thermograms are obtained using a dilatometer or differential scanning calorimeter, and are measured at a slow scanning rate (0.1–1 K min^{-1}) in order to maintain equilibrium.

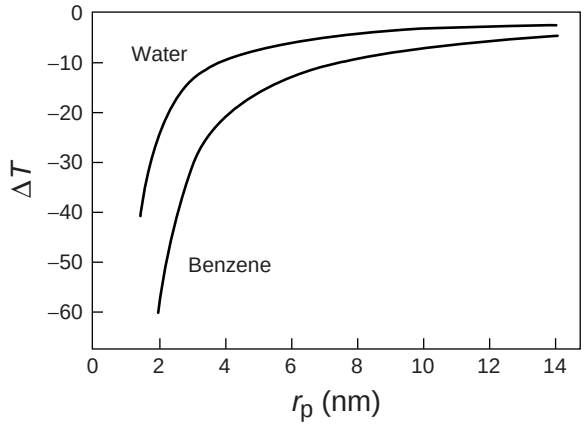

Figure 9 Dependence of melting points of water and benzene on pore radius. [Reproduced by permission of Elsevier Science from M. Brun, A. Lallemand, J.-F. Quinson, C. Eyraud, *Thermochim. Acta*, **21**, 59–88 (1977).]

The freezing/melting transition of water, benzene and cyclohexane confined in porous materials has been measured by ^{1}H-NMR (nuclear magnetic resonance).[100] The PSD has been derived from ^{1}H-NMR signal intensity vs temperature curves of the confined liquids in the pores. The liquid–solid transition in pores is estimated from the ^{1}H-NMR signal intensity.

4.4 Other Methods

4.4.1 X-ray Scattering Method

Porous materials have differences in electron density (solids and voids), so small-angle X-ray scattering (SAXS) occurs. The Guinier equation is represented by Equations (31) and (32):

$$I(Q) = I_0 \exp\left(\frac{-R_G^2 Q^2}{3}\right. \quad (31)$$

$$Q = \frac{4\pi \sin\theta}{\lambda} \quad (32)$$

where $I(Q)$ is the strength of scattering, I_0 is a constant, R_G is the gyration radius of the pore, Q is the scattering parameter, and λ is the wavelength of the X-ray. If the pore is spherical, the pore radius (r) is related to R_G by Equation (33):

$$R_G = \left(\frac{3}{5}\right)^{\frac{1}{2}} r \quad (33)$$

According to this method the size of closed pores can be estimated.

5 DETERMINATION OF MICROPORE SIZE

5.1 Dubinin–Radushkevich Method

The Dubinin–Radushkevich (DR) method is based on the analysis of the gas adsorption isotherm. The gas adsorption isotherm on microporous solids shows a Type I isotherm (Figure 10). The Type I isotherm is called the Langmuir type, which is frequently observed during chemisorption. However, in the case of physisorption on microporous solids, the origin of the shape of the adsorption isotherm is due to micropore filling. The micropores tend to fill with adsorbate at low relative pressures. The adsorption isotherm on the microporous solids is described by the DR equation shown in Equations (34) and (35):[101,102]

$$W = W_0 \exp\ -B\left(\frac{T}{\beta}\right)^2 \log^2\left(\frac{P_0}{P}\right) \tag{34}$$

$$E_0 = \frac{0.01915}{B^{1/2}} \tag{35}$$

where W is the volume of micropores filled at relative pressure P/P_0 and temperature T, W_0 is the total volume of micropores, B is the so-called structural constant, the parameter β is the affinity coefficient, which depends on the adsorptive [$\beta(C_6H_6) = 1$], and E_0 is the characteristic energy of adsorption. Equation (34) is rewritten for experimental use as Equation (36):

$$\log W = \log W_0 - D\log^2\left(\frac{P_0}{P}\right) \tag{36}$$

A plot of $\log W$ against $\log^2(P_0/P)$ is called a DR plot and the total volume of micropores is calculated from the intercept. A DR plot is shown in Figure 11.[103] The characteristic dimension of the micropore (x),

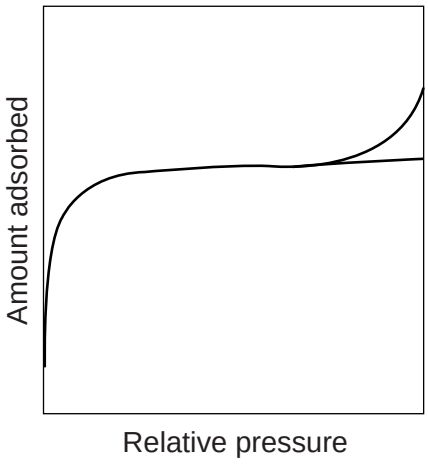

Figure 10 Model of Type I adsorption isotherm.

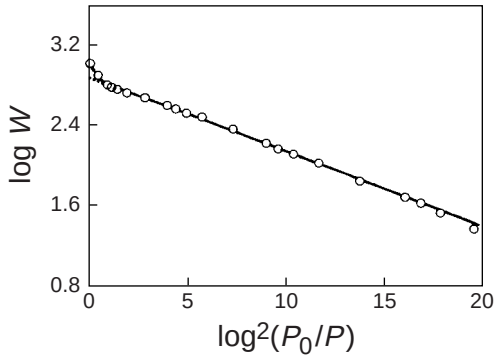

Figure 11 DR plot of adsorption of argon at 77 K on activated carbon. [Reproduced by permission of Academic Press from H. Marsh, B. Rand, *J. Colloid Interface Sci.*, **33**, 101–116 (1970).]

corresponding to the half width of the slit pore, is related to the characteristic energy of adsorption (E_0) by Equation (37):

$$x = \frac{k}{E_0} \tag{37}$$

where the value of the parameter k depends on E_0. When the value of E_0 is 22–14 kJ mol^{-1}, the value of k is 12.5–12.9 kJ nm mol^{-1}.[104] The pore size of the micropore can be calculated from Equation (37) using the DR equation.[104,105]

5.2 Horvath–Kawazoe Method

Horvath and Kawazoe[106] theoretically determined an average potential function inside the slit-shaped pore model. Equation (38) is the general expression of the Horvath–Kawazoe (HK) equation:

$$RT\ln\left(\frac{P}{P_0}\right) = \frac{K}{\sigma^4(l-2\delta)}\ \frac{\sigma^4}{3(l-\delta)^3} - \frac{\sigma^{10}}{9(l-\delta)^9} - \frac{\sigma^4}{3\delta^3} + \frac{\sigma^{10}}{9\delta^9} \tag{38}$$

where P/P_0 is the relative pressure, parameter K depends on the properties of adsorbent and adsorbate, σ is the distance between a gas atom and the surface at zero interaction energy, 2δ is the sum of the molecular diameter of adsorbed molecules and the atomic diameter of solid atoms, and l is the distance between the nuclei of the parallel layers (Figure 12). Consequently, the effective pore size (w) is given by Equation (39):

$$w = l - \delta_s \tag{39}$$

where δ_s is the diameter of solid atoms. In the case of the nitrogen adsorption in slit-shaped pore (graphite),[106] Equations (38) and (39) are given by Equations (40)

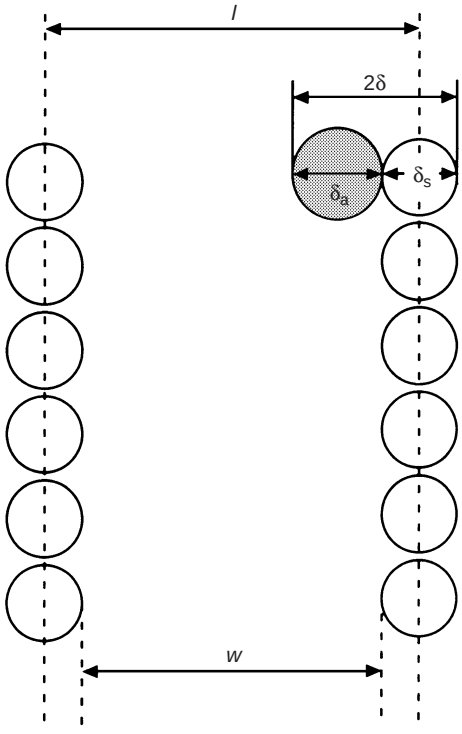

Figure 12 Model of a slit-shaped pore.

and (41):

$$\ln\left(\frac{P}{P_0}\right) = \frac{62.38}{l-0.64}\,\frac{1.859\times10^{-3}}{(l-0.32)^3} - \frac{2.7087\times10^{-7}}{(l-0.32)^9} - 0.05014 \qquad (40)$$

$$w = l - 0.34 \qquad (41)$$

where l and w are in nanometers. The effective pore sizes corresponding to the relative pressures are shown in Table 5. The HK equation has extended the other systems. Swiatkowski et al.[107] deduced the equations for benzene– and argon–carbon systems. These are shown in Equations (42) and (43), respectively:

$$\ln\left(\frac{P}{P_0}\right) = \frac{72.619}{l-2\delta}\,\frac{2.874\times10^{-3}}{(l-\delta)^3} - \frac{7.67\times10^{-7}}{(l-\delta)^9} - 0.05567 \qquad (42)$$

$$\ln\left(\frac{P}{P_0}\right) = \frac{62.38}{l-2\delta}\,\frac{1.859\times10^{-3}}{(l-\delta)^3} - \frac{2.7087\times10^{-6}}{(l-\delta)^9} - 0.05014 \qquad (43)$$

The HK equation has been extended to argon adsorption in cylindrical[108] and spherical pores[109] of zeolites.

Table 5 Relationship between relative pressures and effective pore sizes calculated from Equations (40) and (41). [Reproduced by permission of the Copyright Society of Chemical Engineers, Japan from G. Horvath, K. Kawazoe, *J. Chem. Eng. Jpn.*, **16**, 470–475 (1983).]

P/P_0	w (nm)
1.46×10^{-7}	0.4
1.05×10^{-5}	0.5
1.54×10^{-4}	0.6
8.86×10^{-4}	0.7
2.95×10^{-3}	0.8
2.22×10^{-2}	1.1
7.59×10^{-2}	1.5
3.15×10^{-1}	3
7.24×10^{-1}	10

5.3 Other Methods

5.3.1 Comparison Plot Method

The comparison plot method is described in section 4.1.2.

5.3.2 Molecular Probe Method

The molecular probe method is based on the molecular sieving abilities of micropores. The effective pore size can be estimated from the adsorption capacities of a series of molecules varying in size and shape.[110,111]

5.3.3 Nuclear Magnetic Resonance Method

The chemical shift of NMR spectra reflects the local environment of the observed nucleus. ^{129}Xe NMR of adsorbed Xe in micropores has been developed to study the local environment in pores.[112–114]

ABBREVIATIONS AND ACRONYMS

BET	Brunauer–Emmet–Teller
BJH	Barret–Joyner–Halenda
DH	Dollimore–Heal
DR	Dubinin–Radushkevich
HK	Horvath–Kawazoe
NMR	Nuclear Magnetic Resonance
PSD	Pore Size Distribution
QCM	Quartz Crystal Microbalance
SAW	Surface Acoustic Wave
SAXS	Small-angle X-ray Scattering
UF	Ultrafiltration

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction

Surfaces **(Volume 10)**
Surfaces: Introduction • Electron Microscopy and Scanning Microanalysis • Scanning Electron Microscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Microbalance, Electrochemical Quartz Crystal

Thermal Analysis **(Volume 15)**
Differential Scanning Calorimetry and Differential Thermal Analysis

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview

REFERENCES

1. K.K. Unger, J. Rouquerol, K.S.W. Sing, H. Kral (eds.), *Characterization of Porous Solids I*, Elsevier, Amsterdam, 1988.
2. F. Rodriquez-Reinoso, J. Roquerol, K.S.W. Sing, K.K. Unger (eds.), *Characterization of Porous Solids II*, Elsevier, Amsterdam, 1991.
3. J. Rouquerol, F. Rodriguez-Reinoso, K.S.W. Sing, K.K. Unger (eds.), *Characterization of Porous Solids III*, Elsevier, Amsterdam, 1994.
4. B. McEnaney, T.J. Mays, J. Roquerol, F. Rodriguez-Reinso, K.S.W. Sing, K.K. Unger, *Characterization of Porous Solids*, The Royal Society of Chemistry, Cambridge, IV, 1997.
5. S. Brunauer, P.H. Emmett, E. Teller, 'Adsorption of Gases in Multimolecular Layers', *J. Am. Chem. Soc.*, **60**, 309–319 (1938).
6. P.H. Emmet, S. Brunauer, 'The Use of Low Temperature van der Waals Adsorption Isotherms in Determing the Surface Area Iron Synthetic Ammonia Catalysts', *J. Am. Chem. Soc.*, **59**, 1553–1564 (1937).
7. H.K. Livingston, 'Cross-sectional Areas of Molecules Adsorbed on Solid Surfaces', *J. Am. Chem. Soc.*, **66**, 569–573 (1944).
8. A.L. McClellan, H.F. Harnsberger, 'Cross-sectional Area of Molecules Adsorbed on Solid Surfaces', *J. Colloid Interface Sci.*, **23**, 577–599 (1967).
9. D. Avnir, D. Farin, P. Pfeifer, 'Chemistry in Noninteger Dimensions Between Two and Three. II. Fractal Surfaces of Adsorbents', *J. Chem. Phys.*, **79**, 3566–3571 (1983).
10. R.A. Beebe, J.B. Beckwith, J.M. Honig, 'The Determination of Small Surface Areas by Krypton Adsorption at Low Temperature', *J. Am. Chem. Soc.*, **67**, 1554–1558 (1945).
11. D.E. Meyer, J.E. Wells, 'Surface Areas and C Values for Chemically Etched and Thermally Oxidized Silicon Single-crystal Surfaces by Krypton Adsorption', *J. Colloid Interface Sci.*, **22**, 503–512 (1966).
12. J.L. Carden, R.A. Pierotti, 'The Adsorption of Krypton on the (1,1,1) Face of Copper Single Crystals', *J. Colloid Interface Sci.*, **47**, 379–394 (1974).
13. K. Yasuda, S. Ohsawa, Y. Matsuo, S. Kimura, 'The Evaluation of the Actual Fracture Energy of the Carbon Fiber Reinforced Silicon Nitride Composite by the Kr-BET Adsorption Technique', *J. Ceram. Soc. Jpn.*, **100**, 577–580 (1992).
14. W.A. Cannon, 'Use of Xenon Adsorption for Measurement of Surface Area', *Nature*, **197**, 1000–1001 (1963).
15. T. Kabe, T. Mizuno, I. Yasumori, 'The Surface Area of the Annealed Palladium Foil Catalyst as Determined by a Novel Method of Measuring a Small Surface Area', *Bull. Chem. Soc. Japan*, **40**, 2047–2051 (1967).
16. T. Takaishi, A. Yusa, T. Ohgusi, 'Determination of the Surface Roughness of Passive Film by Gas Adsorption Techniques', *J. Electrochem. Soc.*, **126**, 1720–1722 (1979).
17. T. Takaishi, Y. Sensui, 'Thermal Transpiration Effect of Hydrogen, Rare Gases and Methane', *Trans. Faraday. Soc.*, **59**, 2503–2514 (1963).
18. J.H. Thomas, S.P. Sharma, 'Adsorption and Desorption of Water on Au by the Quartz-crystal-oscillator Method', *J. Vac. Sci. Technol.*, **13**, 549–551 (1976).
19. Y. Yan, T. Bein, K.D. Brown, R. Forrister, C.J. Brinker, 'Molecular Recognition on Acoustic Wave Devices: Modified Zeolite–Silica Thin Films with Tailored Adsorption Properties', *Mater. Res. Soc. Symp. Proc.*, **271**, 435–441 (1992).
20. H. Wohltjen, 'Mechanism of Operation and Design Considerations for Surface Acoustic Wave Device Vapour Sensors', *Sens. Actuators*, **5**, 307–325 (1984).
21. A.J. Ricco, G.C. Frye, S.J. Martin, 'Determination of BET Surface Areas of Porous Thin Films Using Surface Acoustic Wave Devices', *Langmuir*, **5**, 273–276 (1989).
22. S.L. Hietala, D.M. Smith, V.M. Hietala, G.C. Frye, S.J. Martin, 'Pore Structure Characterization of Thin Films Using a Surface Acoustic Wave/Volumetric Adsorption Technique', *Langmuir*, **9**, 249–251 (1993).
23. J.J. Kipling, R.B. Wilson, 'Adsorption of Methylene Blue in the Determination of Surface Areas', *J. Appl. Chem.*, **10**, 109–113 (1960).
24. C.H. Giles, A.P. D'Silva, 'Molecular Sieve Effects of Powders Towards Dyes: Measurement of Porosity by Dye Adsorption', *Trans. Faraday. Soc.*, **65**, 1943–1951 (1969).
25. R. Brina, A.D. Battisti, 'Determinations of the Specific Surface Area of Solids by Means of Adsorption Data', *J. Chem. Edu.*, **64**, 175–176 (1987).

26. S. Ozeki, 'Estimation of Pore and Surface Structures of Solids by Dye Adsorption: The Preadsorption Method and Fractal Analysis', *Langmuir*, **5**, 186–191 (1989).
27. G. Kahr, F.T. Madsen, 'Determination of the Cation Exchange Capacity and the Surface Area of Bentonite, Illite and Kaolinite by Methylene Blue Adsorption', *Appl. Clay Sci.*, **9**, 327–336 (1995).
28. A.K. Helmy, E.A. Ferreiro, S.G. Bussetti, 'Surface Area Evaluation of Montmorillonite', *J. Colloid Interface Sci.*, **210**, 167–171 (1999).
29. A.S. Hilton, J.F. Fuzek, 'The Adsorption of Fatty Acids on Nickel and Platinum Catalysts', *J. Am. Chem. Soc.*, **68**, 229–231 (1946).
30. A.S. Russel, C.N. Cochran, 'Alumina Surface Area Measurements', *Ind. Eng. Chem.*, **42**, 1332–1335 (1950).
31. M.A. Rahman, S.K. Basu, A. Roychoudhury, A.K. Ghosh, 'Comparison of Methods of Determination of Specific Surface Area Using Samples of Cupric Oxide, Ferric Oxide, and Alumina Powders', *J. Colloid Interface Sci.*, **85**, 452–456 (1982).
32. C.H. Giles, A.P. D'Silva, 'Electro-kinetic and Diffusion Measurements of Association of Dyes in Solution and Adsorbed at Solid Surfaces', *Trans. Faraday. Soc.*, **65**, 2516–2528 (1969).
33. C.A. Bower, J.O. Goertzen, 'Surface Area of Solids and Clays by an Equilibrium Ethylene Glycol Method', *Soil Sci.*, **87**, 289–292 (1959).
34. R.A.G. Rawson, 'A Rapid Method for Determining the Surface Area of Aluminosilicates from the Adsorption Dynamics of Ethylene Glycol Vapor', *J. Soil Sci.*, **20**, 325–335 (1969).
35. I.M. Eltantawy, P.W. Arnold, 'Ethylene Glycol Sorption by Homoionic Montmorillonites', *J. Soil Sci.*, **25**, 99–110 (1974).
36. S. Ergun, 'Determination of Geometric Surface Area of Crushed Porous Solids: Gas Flow Method', *Anal. Chem.*, **24**, 388–393 (1952).
37. D.T. Wasan, W. Wnek, R. Davies, M. Jackson, B.H. Kaye, 'Analysis and Evaluation of Permeability Techniques for Characterizing Fine Particles. Part I. Diffusion and Flow Through Porous Media', *Powder Technol.*, **14**, 209–228 (1976).
38. D.T. Wasan, M.B. Ranade, S.K. Sood, R. Davies, M. Jackson, B.H. Kaye, W. Wnek, 'Analysis and Evaluation of Permeability Techniques for Characterizing Fine Particles. Part II. Measurement of Specific Surface Area', *Powder Technol.*, **14**, 229–244 (1976).
39. C. Nyström, K. Malmqvist, W. Alex, 'Comparison of Permeametry and Photometry for the Measurement of Surface Area of Powders', *Acta Pharm. Suec.*, **14**, 497–504 (1977).
40. J. Casal, J. Arnaldos, N. Pellicer, 'Measurement of External Surface Area of Powders by a Permeametric Method', *Powder Technol.*, **58**, 93–97 (1989).
41. M. Eriksson, C. Nyström, G. Alderborn, 'Evaluation of a Permeametry Technique for the Surface Area Measurements of Coarse Particulate Materials', *Int. J. Pharm.*, **63**, 189–199 (1990).
42. E.L. Gooden, C.M. Smith, 'Measuring Average Particle Diameter of Powders – An Air-permeation Apparatus', *Ind. Eng. Chem.*, **12**, 479–482 (1940).
43. G.J.I. Igwe, 'Improved Apparatus for Measurement of Specific Surface Areas of Powders', *Ind. Eng. Chem. Res.*, **26**, 1259–1262 (1987).
44. G.J.I. Igwe, 'Porosity and Permeametry at Sub-atmospheric Pressures', *Chem. Eng. Technol.*, **12**, 255–261 (1989).
45. W.D. Harkins, G. Jura, 'Surfaces of Solid. XII. An Absolute Method for the Determination of the Area of a Finely Divided Crystalline Solid', *J. Am. Chem. Soc.*, **66**, 1362–1366 (1944).
46. S. Partyka, F. Rouquerol, J. Rouquerol, 'Calorimetric Determination of Surface Areas: Possibilities of a Modified Harkins and Jura Procedure', *J. Colloid Interface Sci.*, **68**, 21–31 (1979).
47. K.S.W. Sing, D.H. Everett, R.A.W. Haul, L. Moscou, R.A. Pierotti, J. Rouqérol, T. Siemieniewska, 'Reporting Physisorption Data for Gas/Solid Systems with Special Reference to the Determination of Surface Area and Porosity', *Pure Appl. Chem.*, **57**, 603–619 (1985).
48. J.S. Beck, J.C. Vartuli, W.J. Roth, M.E. Leonowicz, C.T. Kresge, K.D. Schmitt, C.T.-W. Chu, D.H. Olson, E.W. Sheppard, S.B. McCullen, J.B. Higgins, J.L. Schlenker, 'A New Family of Mesoporous Molecular Sieves Prepared with Liquid Crystal Templates', *J. Am. Chem. Soc.*, **114**, 10 834–10 843 (1992).
49. R.G. Avery, J.D.F. Ramsay, 'The Sorption of Nitrogen in Porous Compacts of Silica and Zirconia Powders', *J. Colloid Interface Sci.*, **42**, 597–606 (1973).
50. C.G. Shull, 'The Determination of Pore Size Distribution from Gas Adsorption Data', *J. Am. Chem. Soc.*, **70**, 1405–1410 (1948).
51. B.C. Lippens, B.G. Linsen, J.H. de Boer, 'Studies on Pore Systems in Catalysts I. The Adsorption of Nitrogen; Apparatus and Calculation', *J. Catal.*, **3**, 32–37 (1964).
52. J.B. Butt, 'Multilayer Nitrogen Adsorption in Porous Structures', *J. Catal.*, **4**, 685–690 (1965).
53. J.H. de Boer, B.G. Linsen, T.H. Osinga, 'Studies on Pore Systems in Catalysts VI. The Universal t Curve', *J. Catal.*, **4**, 643–648 (1965).
54. J.H. de Boer, B.G. Linsen, J.C.P. Broekhoff, T.H. Osinga, 'Studies on Pore Systems in Catalysts XV. The Influence of the Geometrical Factor on the t Curve Measured on Aerosil', *J. Catal.*, **11**, 46–53 (1968).
55. J.D. Carruthers, P.A. Cutting, R.E. Day, M.R. Harris, S.A. Mitchell, K.S.W. Sing, 'Standard Data for the Adsorption of Nitrogen at −196 °C on Non-porous Hydroxylated Silica', *Chem. Ind.*, 1772–1773 (1968).
56. J.R. Blachere, 'Modification of t Curve for N_2 Adsorption on Porous Glass', *J. Am. Ceram. Soc.*, **57**, 268–269 (1974).

57. A. Lecloux, J.P. Pirard, 'The Importance of Standard Isotherms in the Analysis of Adsorption Isotherms for Determining the Porous Texture of Solids', *J. Colloid Interface Sci.*, **70**, 265–281 (1979).
58. D. Dollimore, G.R. Heal, 'Pore-size Distribution in Typical Adsorbent Systems', *J. Colloid Interface Sci.*, **33**, 508–519 (1970).
59. E.P. Barrett, L.G. Joyner, P.P. Halenda, 'The Determination of Pore Volume and Area Distributions in Porous Substances. I. Computations from Nitrogen Isotherms', *J. Am. Chem. Soc.*, **73**, 373–380 (1951).
60. W.B. Innes, 'Use of a Parallel Plate Model in Calculation of Pore Size Distribution', *Anal. Chem.*, **29**, 1069–1073 (1957).
61. R.B. Anderson, 'Pore Distributions from Desorption Isotherms', *J. Catal.*, **3**, 50–56 (1964).
62. D. Dollimore, G.R. Heal, 'An Improved Method for the Calculation of Pore Size Distribution from Adsorption Data', *J. Appl. Chem.*, **14**, 109–114 (1964).
63. J.C.P. Broekhoff, J.H. de Boer, 'Studies on Pore Systems in Catalysts. IX. Calculation of Pore Distributions from the Adsorption Branch of Nitrogen Sorption Isotherms in the Case of Open Cylindrical Pores A. Fundamental Equations', *J. Catal.*, **9**, 8–14 (1967).
64. J.C.P. Broekhoff, J.H. de Boer, 'Studies on Pore System in Catalysts XI. Pore Distribution Calculations from the Adsorption Branch of a Nitrogen Adsorption Isotherm in the Case of ''Ink-bottle'' Type Pores', *J. Catal.*, **10**, 153–165 (1968).
65. M.R. Harris, 'A Source of Error in the Computation of Pore Size Data from Nitrogen Sorption Isotherms', *Chem. Ind.*, 268–269 (1965).
66. L.A. de Wit, J.J.F. Scholten, 'Studies on Pore Structure of Adsorbents and Catalysts II. Pore Size Distribution in Synthetic Chrysotile: The Classical and Corrected Kelvin Equation Applied to Nitrogen Capillary Condensation', *J. Catal.*, **36**, 30–35 (1975).
67. L.R. Fisher, J.N. Israelachvili, 'Direct Experimental Verification of the Kelvin Equation for Capillary Condensation', *Nature*, **277**, 548–549 (1979).
68. L.R. Fisher, J.N. Israelachvili, 'Experimental Studies on the Applicability of the Kelvin Equation to Highly Curved Concave Menisci', *J. Colloid Interface Sci.*, **80**, 528–541 (1981).
69. B.K. Peterson, K.E. Gubbins, 'Phase Transitions in a Cylindrical Pore: Grand Canonical Monte Carlo, Mean Field Theory and the Kelvin Equation', *Mol. Phys.*, **62**, 215–226 (1987).
70. A. Elkamel, R.D. Noble, 'Prediction of Capillary Condensation in Small Cylindrical Pores Using the Local Density Approximation and a Full Lennard–Jones 6–12 Potential', *J. Phys. Chem.*, **95**, 10076–10080 (1991).
71. M.W. Maddox, K.E. Gubbins, 'Molecular Simulation of Fluid Adsorption in Buckytubes and MCM-41', *Int. J. Thermophys.*, **15**, 1115–1123 (1994).
72. T. Takei, M. Chikazawa, T. Kanazawa, 'The Validity of the Kelvin Equation in Estimation of Small Pore Size by Nitrogen Adsorption', *Colloid Polym. Sci.*, **275**, 1156–1161 (1997).
73. M.W. Maddox, J.P. Olivier, K.E. Gubbins, 'Characterization of MCM-41 Using Molecular Simulation: Heterogeneity Effects', *Langmuir*, **13**, 1737–1745 (1997).
74. L.D. Gelb, K.E. Gubbins, 'Characterization of Porous Silica Glasses: Simulation Models, Adsorption Isotherms, and the Brunauer–Emmet–Teller Analysis Method', *Langmuir*, **14**, 2087–2111 (1998).
75. B.C. Lippens, J.H. de Boer, 'Studies of Pore Systems in Catalysts V. The t Method', *J. Catal.*, **4**, 319–323 (1965).
76. M.R. Bhambhani, P.A. Cutting, K.S.W. Sing, D.H. Turk, 'Analysis of Nitrogen Adsorption Isotherms on Porous and Nonporous Silicas by the BET and α_s-Method', *J. Colloid Interface Sci.*, **38**, 109–117 (1972).
77. K. Kaneko, C. Ishii, M. Ruike, H. Kuwabara, 'Origin of Superhigh Surface Area and Microcrystalline Graphitic Structures of Activated Carbons', *Carbon*, **30**, 1075–1088 (1991).
78. K. Kaneko, C. Ishii, 'Superhigh Surface Area Determination of Microporous Solids', *Colloids Surf.*, **67**, 203–212 (1992).
79. A.W. Adamson, *Physical Chemistry of Surfaces*, 5th edition, J. Wiley & Sons, New York, 387, 1990.
80. S. Lowell, J.E. Shields, 'Influence of Contact Angle on Hysteresis in Mercury Porosimetry', *J. Colloid Interface Sci.*, **80**, 192–196 (1981).
81. S. Lowell, J.E. Shields, 'Hysteresis, Entrapment, and Wetting Angle in Mercury Porosimetry', *J. Colloid Interface Sci.*, **83**, 273–278 (1981).
82. W.C. Conner, C. Blanco, K. Coyne, J. Neil, J. Pajares, 'Analysis of the Morphology of High Surface Area Solids: Studies of Agglomeration and the Determination of Shape', *J. Catal.*, **106**, 202–209 (1987).
83. R.L. Portsmouth, L.F. Gladden, 'Determination of Pore Connectivity by Mercury Porosimetry', *Chem. Eng. Sci.*, **46**, 3023–3036 (1991).
84. A.Y. Fadeev, V.A. Eroshenko, 'Study of Penetration of Water into Hydrophobized Porous Silicas', *J. Colloid Interface Sci.*, **187**, 275–282 (1997).
85. M. Brun, A. Lallemand, J.-F. Quinson, C. Eyraud, 'A New Method for the Simultaneous Determination of the Size and the Shape of Pores: The Thermoporometry', *Thermochim. Acta*, **21**, 59–88 (1977).
86. F. Ehrburger, J. Lahaye, 'Thermoporometric Study of the Intra-aggregate Porosity of Pyrogenic Silica Dispersed in Aqueous and Non-aqueous Liquids', *Colloids Surf.*, **23**, 105–117 (1987).
87. F. Ehrburger, S. Misono, J. Lahaye, 'Characterization of the Aggregate Void Structure of Carbon Blacks by Thermoporometry', *J. Colloid Interface Sci.*, **135**, 468–485 (1990).

88. J.F. Quinson, N. Tchipkam, J. Dumas, C. Bovier, J. Serughetti, C. Guizard, A. Larbot, L. Cot, 'Swelling of Titania Gels in Decane', *J. Non-cryst. Solids*, **99**, 151–159 (1988).
89. J.F. Quinson, M. Brun, 'Porous Material Characterization by Paranitrophenol Thermoporometry', *High Temperatures–High Pressures*, **30**, 677–682 (1998).
90. J.F. Quinson, J. Dumas, J. Serughetti, 'Alkoxide Silica Gel: Porous Structure by Thermoporometry', *J. Non-cryst. Solids*, **79**, 397–404 (1986).
91. J.F. Quinson, N. Mameri, L. Guihard, B. Bariou, 'The Study of the Swelling of an Ultrafiltration Membrane Under the Influence of Solvents by Thermoporometry and Measurement of Permeability', *J. Membrane Sci.*, **58**, 191–200 (1991).
92. F.P. Cuperus, D. Bargeman, C.A. Smolders, 'Critical Points in the Analysis of Membrane Pore Structures by Thermoporometry', *J. Membrane Sci.*, **66**, 45–53 (1992).
93. K.J. Kim, A.G. Fane, R.B. Aim, M.G. Liu, G. Jonsson, I.C. Tessaro, A.P. Broek, D. Bargeman, 'A Comparative Study of Techniques Used for Porous Membrane Characterization: Pore Characterization', *J. Membrane Sci.*, **87**, 35–46 (1994).
94. K.R. Kloetstra, H.W. Zandbergen, M.A. Koten, H. Bekkum, 'Thermoporometry as a New Tool in Analyzing Mesoporous MCM-41 Materials', *Catal. Lett.*, **33**, 145–156 (1995).
95. K. Ishikiriyama, M. Todoki, K.H. Min, S. Yonemori, M. Noshiro, 'Thermoporometry: Pore Size Distribution Measurements for Microporous Glass Using Differential Scanning Calorimetry', *J. Therm. Anal.*, **46**, 1177–1189 (1996).
96. L.G. Homshaw, 'High Resolution Heat Flow DSC: Application to Study of Phase Transitions, and Pore Size Distribution in Saturated Porous Materials', *J. Therm. Anal.*, **19**, 215–234 (1980).
97. B.V. Enüstün, B.W. Gunnink, T. Demirel, 'Phase Transition Porosimetry and Surface Area Determination', *J. Colloid Interface Sci.*, **134**, 264–274 (1990).
98. K. Ishikiriyama, M. Todoki, K. Motomura, 'Pore Size Distribution (PSD) Measurements of Silica Gels by Means of Differential Scanning Calorimetry I. Optimization for Determination of PSD', *J. Colloid Interface Sci.*, **171**, 92–102 (1995).
99. K. Ishikiriyama, M. Todoki, K. Motomura, 'Pore Size Distribution Measurements of Silica Gels by Means of Differential Scanning Calorimetry II. Thermoporometry', *J. Colloid Interface Sci.*, **171**, 103–111 (1995).
100. E.W. Hansen, R. Schmidt, M. Stöcker, 'Pore Structure Characterization of Porous Silica by ^{1}H-NMR Using Water, Benzene, and Cyclohexane as Probe Molecules', *J. Phys. Chem.*, **100**, 11 396–11 401 (1996).
101. M.M. Dubinin, 'The Potential Theory of Adsorption of Gases and Vapors for Adsorbents with Energetically Nonuniform Surfaces', *Chem. Rev.*, **60**, 235–241 (1960).
102. M.M. Dubinin, 'Adsorption in Micropores', *J. Colloid Interface Sci.*, **23**, 487–499 (1967).
103. H. Marsh, B. Rand, 'The Characterization of Microporous Carbons by Means of the Dubinin–Radushkevich Equation', *J. Colloid Interface Sci.*, **33**, 101–116 (1970).
104. M.M. Dubinin, H.F. Stoeckli, 'Homogeneous and Heterogeneous Micropore Structures in Carbonaceous Adsorbents', *J. Colloid Interface Sci.*, **75**, 34–42 (1980).
105. B. McEnaney, 'Estimation of the Dimensions of Micropores in Active Carbons Using the Dubinin–Radushkevich Equation', *Carbon*, **25**, 69–75 (1987).
106. G. Horvath, K. Kawazoe, 'Method for the Calculation of Effective Pore Size Distribution in Molecular Sieve Carbon', *J. Chem. Eng. Jpn.*, **16**, 470–475 (1983).
107. A. Swiatkowski, B.J. Trznadel, S. Zietek, 'Description of an Active Carbon Micropore Size Distribution Based on the Horvath–Kawazoe Equation Adapted to Benzene Adsorption Isotherms', *Adsorpt. Sci. Technol.*, **14**, 59–67 (1996).
108. A. Saito, H.C. Foley, 'Curvature and Parametric Sensitivity in Models for Adsorption in Micropores', *AIChE J.*, **37**, 429–436 (1991).
109. L.S. Cheng, R.T. Yang, 'Improved Horvath–Kawazoe Equations Including Spherical Pore Models for Calculating Micropore Size Distribution', *Chem. Eng. Sci.*, **49**, 2599–2609 (1994).
110. Z. Hu, N. Maes, E.F. Vansant, 'Molecular Probe Technique for the Assessment of the Carbon Molecular Sieve Structure', *J. Porous Mater.*, **2**, 19–23 (1995).
111. G. Horvath, 'Energetic Interactions in Phase and Molecular Level Pore Characterization in Nano-range', *Colloids Surf.*, **141**, 295–304 (1998).
112. P.J. Barrie, J. Klinowski, '^{129}Xe NMR as a Probe for the Study of Microporous Solids: A Critical Review', *Prog. NMR Spectrosc.*, **24**, 91–108 (1992).
113. N. Bansal, H.C. Foley, D.S. Lafyatis, C. Dybowski, 'Analysis of Porous Carbon with ^{129}Xe, ^{1}H-NMR and N_2 Porosimetry', *Catal. Today*, **14**, 305–316 (1992).
114. M. Mansfeld, W.S. Veeman, 'The Use of ^{129}Xe NMR Exchange Spectroscopy for Probing the Microstructure of Porous Materials', *Chem. Phys. Lett.*, **213**, 153–157 (1993).

Turbidimetry in Particle Size Analysis

Theodora Kourti
Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4L7

1 Introduction **2**
2 Theoretical Background **5**
2.1 Turbidity and Particle Size 5
2.2 The Scattering Coefficient 6
2.3 Turbidity and Moments of the Particle Size Distribution 7
3 Particle Size Determination in Monodisperse Suspensions **7**
3.1 Specific Turbidity 8
3.2 Turbidity Ratio 8
4 Particle Size Distribution Determination in Polydisperse Suspensions **9**
4.1 Specific Turbidity 9
4.2 Turbidity Ratio 11
4.3 The Wavelength Exponent Technique 12
4.4 Resolution of a Conflict in the Literature 13
4.5 Solving for the Particle Size Distribution 13
5 Estimation of Average Diameters **16**
5.1 Turbidity-average Diameters 17
5.2 Volume–Surface-average Diameters 17
5.3 Weight-average Diameters 18
5.4 Other Applications of Turbidimetry 18
6 Procedural Details **19**
6.1 Instrumentation and Forward Scattering 19
6.2 Refractive Indices of the Particles and the Medium 20
6.3 The Choice of the Wavelengths 22
6.4 Preparation of Samples 22
6.5 Particle Volume Fraction for Specific Turbidity 23
7 Effect of Errors on the Estimated Particle Size and Particle Size Distribution **23**
7.1 Error in Measurements 23
7.2 Error in the Value of the Refractive Index Ratio, *m* 23
7.3 Error in the Assumed Distributional Form 24
8 On-line Particle Size Determination Using Turbidimetry **26**
9 Other Considerations **27**
9.1 Choice of the Methods 27
9.2 Corroboration of Results 28
10 Conclusion **28**
Abbreviations and Acronyms **29**
Related Articles **29**
References **29**

The size of particles in a suspension can be estimated by measuring the turbidity of the suspension. Turbidity measures the attenuation of a beam of light traveling through the suspension caused by the scattering and absorption of light by the particles. The amount of scattering and absorption depends on the size of the particles and their concentration in the suspension. Most of the reported applications in the literature use turbidimetry for the estimation of the particle size distribution (PSD) or an average size for particles with diameters ranging from 0.035 to 50 μm. Turbidimetric techniques are low cost, experimentally simple, fast, nondestructive and require no calibration. They are robust methods with very good reproducibility and can be used either off- or on-line, in a laboratory or a plant environment. They have been used for particle or droplet size determination in a variety of diverse systems such as latex particles, silver bromide sols and cow's milk. Most of the time common ultraviolet/visible (UV/VIS) spectrophotometers are utilized for the turbidity measurements. Certain modifications may be necessary on these instruments depending on the size of the particles. The turbidity of the suspension is measured and the size of the particles or the PSD is estimated utilizing light scattering theory. For spherical isotropic homogeneous particles and absence of multiple scattering this theory is well established. The extension to nonspherical and inhomogeneous particles is a field of active research yielding numerous publications.

A controversy existed in the literature regarding the capability of turbidimetry to provide the full PSD but has now been resolved. For polydisperse suspensions, for certain combinations of particle sizes and optical properties, the method cannot provide the full PSD, but only an average of the PSD. This limitation is simply a result of the light scattering patterns and it is expected theoretically. In some cases it can be overcome by simply altering the optical properties of the suspension (i.e. choosing the proper wavelengths for the turbidity measurements or a medium with a different refractive index to suspend the particles). To make the best use of the method and interpret the results in a meaningful way

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

one should be aware of this limitation and work with it. When this is done, excellent results can be achieved. The advantages of the method and a good knowledge of its capabilities make it a very robust choice for many applications. The numerous reports from the successful application of the method are a testimony to this.

1 INTRODUCTION

Turbidimetry refers to a group of techniques that utilize either turbidity measurements at different wavelengths, or a combination of measurements of turbidity and of the particle volume fraction, to estimate the size of particles in suspensions. Turbidity measures the attenuation of a beam of light when it passes through a suspension of particles. Using light scattering theory this attenuation can be theoretically related to the size and concentration of particles in the suspension if the optical properties (refractive indices of particles and medium) are known. For spherical particles and absence of multiple scattering this theory is well established. Extensions to nonspherical particles have been reported. Simple UV/VIS spectrophotometers can be modified (and in some cases used as bought) for the turbidity measurements. Wavelengths in the near-infrared (NIR) range can also be utilized. Particle sizes from the lower submicrometer range to many micrometers can be measured with the proper choice of wavelength for the turbidity measurements. Applications of turbidimetry for particle size determination have been reported mainly for particles with diameters between 0.035 and 50 μm. The simplicity of this fast, noninvasive method, its low cost, good reproducibility and the wide range of particle sizes that it covers have made it very attractive for particle size determination. Mainly owing to its simplicity, the method is robust and suitable for industrial production environments as an on-line particle size determination method.

The first attempts to use turbidimetry for particle size determination started in the 1950s. At the time workers tried to use approximation theories[1] or tabulate turbidity functions[2,3] mainly owing to a lack of powerful computing machines. Ever since, there has been an abundance of reports on the use of turbidimetric techniques for the determination of particle size in a variety of systems, such as silver bromide sols in water,[4–6] polystyrene latexes,[7,8] poly(vinyl acetate) latexes,[9–15] homogenized milk and oil-in-water emulsions,[16–18] coupler dispersions[19] and minerals.[20] A wide range of wavelengths in the UV/VIS region have been utilized for the turbidity measurements but also work in the NIR region has been reported.[6,16,17,19] The determination of particle size using on-line turbidimetry has also attracted a lot of attention.[6,9–11,15,21–24] Some of the work reported in the literature for particle size determination in polydisperse systems (full PSD or an average diameter) for different materials is given in Table 1.

A controversy in the literature[11,12,31] regarding the capability of turbidimetry to provide the full PSD has been resolved by a theoretical analysis and experimental verification reported in Kourti et al.[15,32] The results can be summarized as follows.

The capability of turbidimetric techniques to provide an average particle size or the PSD of a suspension and their sensitivity to experimental error cannot be generalized across refractive indices and particle size ranges. It depends on the optical properties of the suspension and the size of the particles relative to the wavelength and conclusions cannot be extrapolated from one optical regime and particle size to another and from one technique to another. For monodisperse suspensions the particle size can always be estimated using turbidimetry, provided that two turbidity measurements at widely separated wavelengths together with a measurement of the particle volume fraction in the suspension is used. This is irrespective of the optical properties of the suspension (refractive indices of particles and medium, wavelength of light) and the size of the particles. In some simpler cases only two turbidity measurements or measurement of one turbidity and the volume fraction are sufficient. For polydisperse suspensions, turbidity measurements at sufficient wavelengths (the number depends on the detail of the PSD one wishes to resolve) together with a measure of the particle volume fraction are required for the determination of the full PSD. However, for certain combinations of particle sizes and optical properties, the method cannot provide the full PSD, but only an average of the PSD. This limitation is simply a result of the light scattering patterns and it is expected theoretically. More specifically, it stems from the fact that for some combinations of optical properties and particle size, a light scattering function (the wavelength exponent) changes very slowly with the relative size of the particles to the wavelength. When this happens, in the presence of experimental noise, estimation of the full PSD is not possible. In some cases this limitation can be overcome by simply altering the optical properties of the suspension (i.e. choosing proper wavelengths for the turbidity measurements or a medium different than water to suspend the particles). Therefore, to make the best use of the method and interpret the results in a meaningful way, one should be aware of the regimes where this limitation occurs.

Because of the behavior of the wavelength exponent, turbidimetric techniques that utilize the measurement of the particle volume fraction (which is a measurement of

Table 1 Examples of applications of turbidimetry in particle sizing[a]

System	m	Particle size, D (μm)	Wavelengths (nm)	Turbidimetric method	Assumed PSD[b]	Comments[b]	Other methods[b]
Minerals[20]: TiO_2 $Al(OH)_3$ Al_2O_3 SiO_2	1.9 1.86 1.33 1.163	$0.1 < D < 10$	436, 546, 650, 800	Specific turbidity	Log-normal on weight basis	Estimated PSD agrees very well with measured PSD from centrifugal sedimentation	Centrifugal sedimentation
TiO_2 pigments.[25] Oval-shaped crystals. Estimate crystal size-area equivalent diameter	–	Mean crystal sizes: 0.086–0.237	190–1100	44 turbidities and volume fraction	No shape assumed. Estimate 26 heights of histogram in the range 0.012–0.6375 nm	Study involves empirical modeling with latent variables–PLS	TEM
Stable silver bromide sols in water[5]	1.7	$D < 0.25$	(1) 436, 546 (2) 450, 800	Specific turbidity	(1) Apparent diameters (2) Log-normal (ZOLD)	Important discussions on apparent diameters. See section 5	EM
Ostwald ripening process: silver bromide grains[6]	1.7	$D < 0.8$	2.22 μm	Specific turbidity	Apparent diameter (D_τ)	On-line monitoring of grain growth. See section 8	EM
Variety of materials[26] PVC, sand, protein, KCl	–	Weight average: 138 116, 196 20–30 92–284	900	Specific turbidity	Assumed γ PSD with known spread. Reported average size	Some particles too large. Only D_{32} expected to be correct.	Sieve Coulter counter
Polystyrene latexes[7]	1.2	$0.65 < D < 1.3$	440–610 basis: 546	Turbidity ratio	Positively skewed with maximum polydispersity 1.5	For polystyrene see also Haseler.[19] See sections 4.2 and 7.3	EM
Polystyrene[8]	1.2	Synthesized bimodal and trimodal PSDs	300–800	Turbidities only	No assumption on shape	Solve directly for the heights of the PSD	
Styrene–butadiene rubber and acrylonitrile–butadiene rubber[27]	1.16, 1.14	$0.04 < D < 0.14$	300–600 (101 points)	Specific turbidity	(1) Apparent D_τ (2) Log-normal (3) No shape assumption	Only D_τ correct	SEM DLS
Copolymer latexes: poly(butadiene–styrene)[28]	1.17, 1.15	$D < 0.4$	350–1000	Specific turbidity	Apparent diameters	See section 5.1	EM
Photographic dye-forming coupler dispersions[29]	1.166, 1.21	$0.05 < D < 1.0$	400–800	Specific turbidity and wavelength exponent	Log-normal on weight basis (verified by EM)	Conclude that success of method depends on values of α	EM

(*continued overleaf*)

Table 1 (*continued*)

System	m	Particle size, D (μm)	Wavelengths (nm)	Turbidimetric method	Assumed PSD[b]	Comments[b]	Other methods[b]
Ektacolor coupler dispersions[19] – droplet size distribution	1.15	$D < 1.0$	400, 700 400, 2.22	Specific turbidity	Log-normal	See sections 4.1, 4.2, 5.2 and 7.3	–
Cis-polyisoprene Latexes[58]	1.136	$0.6 < D < 1.1$	650, 800, 950, 1100	Specific turbidity	Log-normal on weight basis	Broad PSDs[15] See also section 9.2	Coulter counter EM
Copolymer latexes: divinyl–benzene/ styrene[66]	–	$12 < D_N < 50$	400–600	Specific turbidity	Apparent diameter (D_{32})	See section 5.2. Some coincidental observations are discussed by Kubota et al.[68]	EM
Latexes[68]	–	$2 < D < 40$	400, 700	Specific turbidity	Apparent diameter (D_{32})	See section 5.2	EM
Octacosane suspensions[48]	1.15	$D < 3.0$	436, 546	Turbidity ratio and wavelength exponent	1. Log-normal 2. One parameter PSD, positively skewed	Approach has serious limitations.[50] See section 4.3.	EM
Poly(vinyl acetate)[16]	1.1	$0.1 < D < 1.0$	320–720	Turbidity ratio	Positively skewed	See sections 4.2 and 4.4	EM
Poly(vinyl acetate) latex[11,12]	1.1	1. D_w : 0.20 2. D_τ : 0.11	300, 600	Specific turbidity	Log-normal	See sections 4.4 and 8.0	Disc centri-fugation Angular light scattering
Poly(vinyl acetate)[15,44]	1.095	$0.035 < D < 0.4$	400–600	Specific turbidity	Log-normal ($D_{ap} = D_w$)	Experimental & theoretical analysis demonstrates the limitations of turbidimetry in some (m, α) regions	HDC DLS
Poly(vinyl acetate)[23,24]	1.095	$D < 0.4$	400	Specific turbidity	Apparent diameter, D_w	First successful on-line application	DLS
Poly(vinyl acetate) latex[10]	1.1	$D < 0.4$	350–700	Method of moments	Moments of PSD	See section 4.5	HDC
Paraffin oil in (water + glycerol) + milk[16]	1.1, 1.05	$0.1 < D < 10.0$ $0.1 < D < 40.0$	300–1700	Specific turbidity	(1) Log-normal (2) An upper limit function (3) Exponential type	See also other work[17,30] for similar systems. Successful application	Coulter counter

[a] The systems have been ordered based on the optical properties of the suspension (value of m).
[b] Abbreviations (other than those already defined in the text): PLS, partial least squares; TEM, transmission electron microscopy; ZOLD, zeroth-order logarithmic distribution; EM, electron microscopy; SEM, scanning electron microscopy; DLS, dynamic light scattering; HDC, hydrodynamic chromatography.

the particle size and number independent of the scattering properties) provide reliable estimates for the particle size (the PSD or an average diameter) in a wider range of sizes and optical properties than techniques that utilize turbidity measurements only.

In this article, a brief overview of light scattering theory is first presented followed by a description of the turbidimetric techniques that are most commonly used for particle size determination. For each technique we include its theoretical background, comments on its advantages and limitations and references to its successful applications in the literature. We also include discussions regarding the regions of particle sizes and optical properties where the technique may present difficulties in providing reliable estimates of the particle size or the PSD. We then present cases where turbidimetry was used to provide estimates of the average diameters of the PSD and discuss the advantages and pitfalls of such uses. We follow with details on the procedure such as choice of wavelengths, calculation of refractive index and instrumentation. The sensitivity of the turbidimetric methods to experimental error is extensively discussed. Applications on the use of the method for on-line particle size measurements follow. We conclude with a discussion on how users should proceed with choosing the appropriate method for their application and with comparing results from other particle sizing techniques. Throughout the article several applications reported in the literature are critically discussed in order to point out some approaches with serious limitations and some coincidental observations that may result in misleading conclusions. In this way the readers can judge if the approaches reported in the literature are suitable for their use.

2 THEORETICAL BACKGROUND

2.1 Turbidity and Particle Size

Turbidity, τ_{λ_0} (cm^{-1}), gives a measure of the attenuation of the intensity of a beam of light traversing through a suspension of particles (Equation 1):

$$\tau_{\lambda_0} = \frac{1}{l} \ln \left(\frac{I_0}{I} \right. \tag{1}$$

where I_0 and I represent intensities of the incident and transmitted beams, respectively, and l is the length of the optical path (i.e. cell length, cm). The value of the turbidity changes with the wavelength of the incident light beam, λ_0. I is measured at 180° to I_0.

The attenuation of the intensity of light traversing through a particle suspension is due to scattering of light by the particles and in some cases (depending on the wavelength) absorption of light by the particles and/or the medium. The amount of light scattered and absorbed by the particles depends on their size, their concentration in the suspension, the size of the particles relative to the wavelength of the incident beam and the refractive indices of the particles and the medium. Therefore, theoretically, turbidity can be related to the particle size. For a monodisperse suspension (particles of equal diameter, D) of spherical, isotropic particles, in the absence of multiple scattering (Equation 2):

$$\tau_{\lambda_0} = N \frac{\pi D^2}{4} K_{\text{ext}} \tag{2}$$

where N is the number of particles per cm^3 and K_{ext} is the extinction coefficient at wavelength λ_0. K_{ext} is equal to the sum of the absorption K_{abs} and the scattering K_{scat} coefficients. The extinction coefficient can be calculated from the rigorous Mie theory and the equations for this calculation can be found in several sources.[3,33–36] In the work presented in this article we assume no absorption ($K_{\text{ext}} \equiv K_{\text{scat}}$) and this is the case for most of the work reported in the literature. In most of the reported applications workers have chosen to avoid wavelengths where the medium or the particles absorb, for practical reasons, explained in section 6.3. However, should readers decide to work with absorbing particles, they should use K_{ext} for Equations (3), (4), (12), (13), (15) and (18). The validity of the discussions in this article does not change with the nonabsorption assumption.

For a monodisperse suspension of spherical, nonabsorbing, isotropic particles, in the absence of multiple scattering (Equation 3):

$$\tau_{\lambda_0} = N R_{\text{scat}} = N \frac{\pi D^2}{4} K_{\text{scat}} = N \frac{\pi D^2}{4} K \left(\frac{D}{\lambda_m}, \frac{n_p}{n_m} \right. \tag{3}$$

where R_{scat}, the scattering cross-section, is the total energy scattered by one sphere when the incident light is of unit intensity and either plane polarized or unpolarized. The scattering coefficient, K_{scat}, or $K(D/\lambda_m, n_p/n_m)$, is dimensionless and it is an involved function of two parameters, α and m. $\alpha = \pi(D/\lambda_m)$ is defined as the relative size of the particle diameter to the wavelength of the light in the medium (λ_m) and $m = n_p/n_m$ is the ratio of the refractive index of the particles to the refractive index of the medium. $\lambda_m = \lambda_0/n_m$, where λ_0 is the wavelength of the incident beam in vacuo. Both n_p and n_m are evaluated at λ_0. The estimation of the refractive indices for homogeneous and inhomogeneous particles is discussed in section 6.2. In cases where the particles absorb, the refractive index has both real and imaginary parts.

For a polydisperse suspension of spherical nonabsorbing, isotropic particles, in the absence of multiple

scattering turbidity is given by Equation (4):

$$\tau_{\lambda_0} = N \int_0^{\infty} \frac{\pi D^2}{4} K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right) f(D)\, d(D) \qquad (4)$$

where $f(D)$ is the normalized PSD.

In general, because K_{scat} is an involved nonlinear function of α (i.e. D/λ_m), it is possible from turbidity measurements at a sufficient number of wavelengths (or a combination of turbidity measurements and other properties of the particle suspension) to estimate the size of the particles in a monodisperse suspension or the PSD in a polydisperse suspension. However, it has been shown[32] that for certain combinations of (m, α) values the estimation of the particle size or the PSD of the suspension may present difficulties with some turbidimetric techniques. Being aware of these (m, α) regimes will help the user choose the proper turbidimetric technique and proper wavelengths to work with in order to obtain the most reliable estimate of particle size of a suspension.

2.2 The Scattering Coefficient

Certain features in the behavior of the scattering coefficient are important in explaining the capability of turbidimetric techniques to provide the particle size or the PSD of a suspension. These features will be discussed here in some detail. The scattering coefficient is defined as the ratio of the scattering to the geometrical cross-section. The scattering coefficient exhibits an oscillatory dependence on α, with a series of successive minima and maxima (Figure 1). The first maximum occurs at lower α values the larger is the value of m; the numerical value of K_{scat} at the first maximum is larger the larger is m. At large values of α, m has virtually no effect on K_{scat}, in agreement with the theory of diffraction for objects very large compared with the wavelength used, and K_{scat} approaches the value of 2. Values of K_{scat} for various large α values are shown in Table 2. Note that for $m = 1.2$ and $200 < \alpha < 220$ the value of the oscillating K_{scat} can be found between 2.12 and 1.99, whereas for $2500 < \alpha < 2700$, K_{scat} oscillates between 2.016 and 2.006. When plotted against α, the scattering coefficient also exhibits secondary fluctuations (ripples) that become increasingly irregular for higher values of m. Another parameter which is often used in light scattering is $\rho = 2\alpha|m - 1|$; ρ is the phase shift suffered by a wave passing through the center of a sphere.[37] When K_{scat} is plotted against ρ, the ρ position of the extremes (minima and maxima) of the main oscillation is independent of the value of m.[33,34] Recall that this discussion is for spherical, isotropic spheres. Scattering for nonspherical particles is an area of active research[38] and publications are available.[39–41]

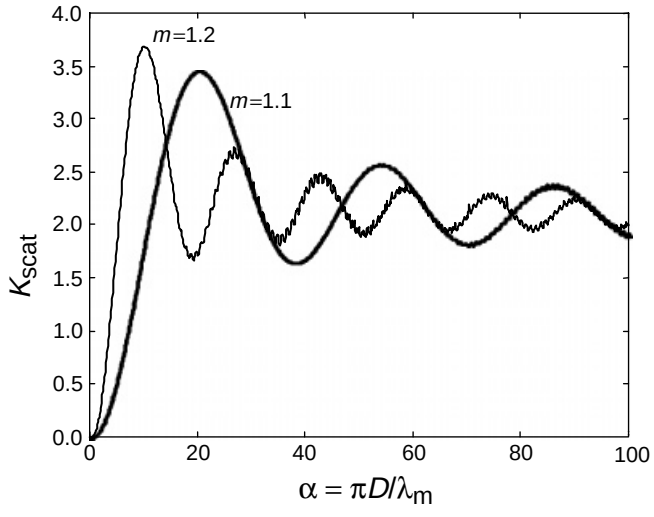

Figure 1 Scattering coefficient, K_{scat}, plotted against α for $m = 1.1$ and 1.2.

As mentioned earlier, in the general case the scattering coefficient can be calculated from the rigorous Mie theory.[3,33–36] In certain (m, α) regimes the equations for K_{scat} can be approximated with simple expressions which give K_{scat} within 1% accuracy from its true value.[37] These expressions are discussed here because they help in elucidating some of the problems encountered by researchers when converting turbidity measurements to particle size.

1. When the suspended particles are very small compared with the wavelength both outside ($\alpha \ll 1$) and inside ($|m\alpha| \ll 1$), the particle,[34,37,42] the Rayleigh scattering theory gives Equation (5):

$$K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right) = \frac{8}{3}\left(\frac{m^2 - 1}{m^2 + 2}\right)^2 \alpha^4 \qquad (5)$$

2. For particles very large compared with the wavelength, the scattering coefficient approaches a

Table 2 Range of values (maximum–minimum) observed for the oscillating K_{scat} in different ranges of α values for $m = 1.1$ and 1.2

m	$20 < \alpha < 40$	$80 < \alpha < 100$	$140 < \alpha < 160$	$200 < \alpha < 220$	$800 < \alpha < 1000$	$1500 < \alpha < 1700$	$2500 < \alpha < 2700$
1.1	3.45–1.64	2.39–1.88	2.24–1.99	2.175–1.98	2.051–1.998	2.03–2.002	2.0195–2.002
1.2	2.7–1.7	2.26–1.94	2.16–1.98	2.12–1.99	2.0365–2.0065	2.023–2.006	2.016–2.006

constant value independent of m[3,33,34,37] (Equation 6):

$$K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right) = 2 \quad (6)$$

Walstra[37] has reported that this is true for $\rho = 2\alpha|m-1| > 300$ and $\alpha > 2000$. Table 2 gives the range of K_{scat} values for different α values and $m = 1.1$ and 1.2. Note that for $m = 1.2$ and α values between 200 and 220 the values of the oscillating K_{scat} can be found between 2.12 and 1.99. Assuming a constant value of 2.0 may introduce a maximum of 6% error in K_{scat} in this range. For very large values of α, assuming $K_{scat} = 2.0$ gives K_{scat} within 1% accuracy from the true value.

Equations (5) and (6) are the two extremes of a general expression where for monodisperse suspensions of nonabsorbing spheres, the dependence of the scattering coefficient on the diameter at a given wavelength is approximated by Equation (7):[4,5]

$$K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right) = k'\left(\frac{D}{\lambda_m}\right)^z = k'_c\alpha^z \quad (7)$$

where k' and k'_c are proportionality constants. Note that $z = 4$ in the Rayleigh regime and $z = 0$ for large particles (compare Equation 7 with Equations 5 and 6, respectively). When plotting K_{scat} against the wavelength for constant D, we have $K_{scat} = k''\lambda_0^{-z}$, and therefore z has been termed "the wavelength exponent".[43] When plotting K_{scat} against α at constant λ_m, we can write Equation (8):

$$\ln(K_{scat}) = z\ln a + \ln k'_c \quad (8)$$

The value of z can be obtained for any α at any m value from the slope of $\ln K_{scat}$ vs $\ln\alpha$.

2.3 Turbidity and Moments of the Particle Size Distribution

In regimes where K_{scat} can be approximated by Equation (7), the turbidity of a monodisperse suspension at a given wavelength λ_0 can be written as $\tau = NkD^y$, where $y = z + 2$ and k is a constant. Several authors[5,43] have tabulated values of y or z for several pairs of (m, α) values (in the notation of Heller et al.,[43] $z \equiv n$). Values of y and z are plotted in Figure 2 for several m values and $\alpha \leq 12.0$. From Equations (4) and (7), for a polydisperse system, at a given λ_0, we can write Equation (9):

$$\tau_{\lambda_0} = N\int_0^\infty k(D)D^{y(D)}f(D)\,d(D) \quad (9)$$

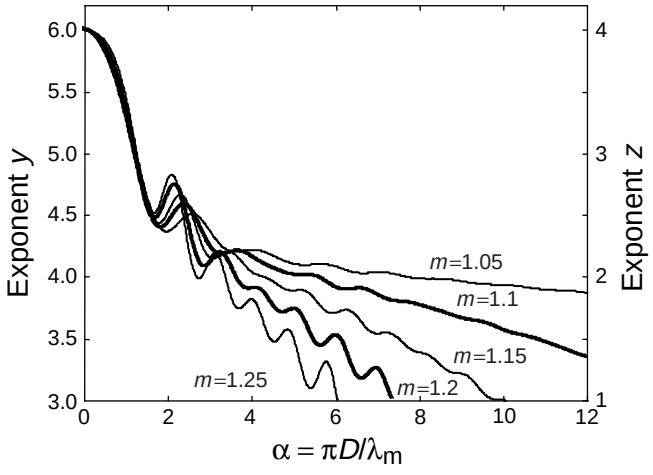

Figure 2 Exponents y and z for $m = 1.05, 1.1, 1.15, 1.2$ and 1.25.

where both k and y are functions of D. When the particle sizes in the suspension are such that k and y do not change significantly with D, then (Equation 10):

$$\tau_{\lambda_0} = Nk\int_0^\infty D^y f(D)\,d(D) = NkJ_y \quad (10)$$

where k and y are constants. In these cases, turbidity is proportional to the yth moment, J_y, of the PSD. The value of y (i.e. the order of the moment) changes with α and m (Figure 2). For $\alpha < 0.5$, y approaches the limiting value of 6 ($z = 4$, Rayleigh regime) and therefore turbidity is proportional to the sixth moment of the PSD. For values of m between 1.1 and 1.2, and for $\alpha < 1.6$, the value of y is almost independent of the value of m. Kourti et al.[15,32,44] discussed that for small values of m ($m = 1.05, 1.1, 1.15$) the value of y is approximately 4.0 ($4.2 \leq y \leq 3.8$) for a very wide range of α values (for $m = 1.05$, $3 < \alpha < 14$; for $m = 1.15$, $3 < \alpha < 6$). For suspensions with PSDs in the above (m, α) regions, turbidity would be approximately proportional to the fourth moment of the PSD. Finally, for very large particles, $y = 2$ and turbidity is proportional to the second moment of the PSD.

3 PARTICLE SIZE DETERMINATION IN MONODISPERSE SUSPENSIONS

The turbidity of a suspension of spherical particles as given in Equation (3) is a function of both the number concentration of particles (N) and the particle diameter (D). The estimation of these two unknowns requires sufficient measurements of turbidity at different wavelengths or measurements of turbidity and another property that is also related to particle concentration and diameter. Two approaches have been used in the literature:[33] the turbidity at a given wavelength is divided

either by the particle volume fraction or by the turbidity value at another wavelength, such that the unknown number of particles is eliminated, and the new quantity is related to particle size only. Once the particle size is determined, substitution into Equation (3) will give N.

3.1 Specific Turbidity

The volume fraction of the spherical particles in a monodisperse suspension is given by Equation (11):

$$\phi = N\frac{\pi D^3}{6} \tag{11}$$

The particle volume fraction can be measured (see section 6.5) using gravimetric analysis or densitometry.[21,23,45] The specific turbidity, defined as the ratio of the turbidity of the suspension to the volume fraction of the suspended particles, is independent of the particle number concentration (Equation 12):

$$\frac{\tau_{\lambda_0}}{\phi} = \frac{3}{2}\frac{K\ \ \frac{D}{\lambda_m}, \frac{n_p}{n_m}}{D} \tag{12}$$

Specific turbidity is plotted against diameter in Figure 3 for two different types of particles suspended in water: poly(vinyl acetate) with $m \approx 1.1$ and polystyrene with $m \approx 1.2$. The particle diameter D can be calculated from one turbidity measurement and the particle volume fraction when the optical properties of the suspension are known and the expected particle size range is known. For example, if it is known that we have submicrometer particles then one specific turbidity measurement for poly(vinyl acetate) at 600 or 400 nm would give the particle diameter. Note that owing to the oscillatory behavior of the specific turbidity, more diameters may correspond to one specific turbidity value. For example, for polystyrene a specific turbidity reading of 15 000 cm^{-1} at 400 nm in the range shown in Figure 3 corresponds to six different diameters (0.15, 1.7075, 1.855, 1.865, 1.9225 and 2.56 μm; four due to the main oscillating pattern of K_{scat} and two due to a secondary rippling effect). Therefore, if the size range is not a priori known, measurements at two (preferably widely spaced) wavelengths are required to extend the applicability of the method to a wide range of particle sizes.[15] Once the particle size has been estimated, the concentration of the particles in the suspension (N) can be calculated from Equation (3). Note that in the Rayleigh regime, where K_{scat} is given by Equation (5), the specific turbidity at given λ_m is directly proportional to the volume of the particles ($\tau/\phi \propto D^3$), and that for very large particles ($K_{scat} = 2.0$) it is proportional to the inverse of the diameter ($\tau/\phi = 3/D$).

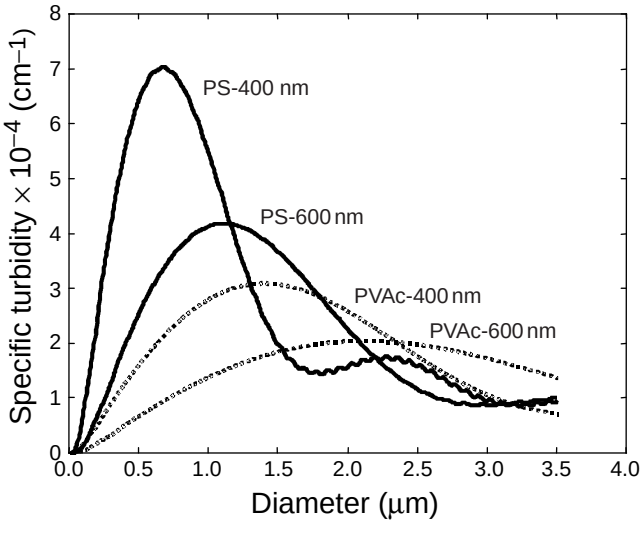

Figure 3 Specific turbidity plotted against particle diameter for monodisperse poly(vinyl acetate) and polystyrene latexes at 400 and 600 nm.

3.2 Turbidity Ratio

The ratio of two turbidity readings taken at two wavelengths is a quantity independent of the particle number concentration (Equation 13):

$$\frac{\tau_{\lambda_{01}}}{\tau_{\lambda_{02}}} = \frac{K\ \ \frac{D}{\lambda_m}, \frac{n_p}{n_m}\ \ _{\lambda_{01}}}{K\ \ \frac{D}{\lambda_m}, \frac{n_p}{n_m}\ \ _{\lambda_{02}}} \tag{13}$$

Subscripts λ_{01} and λ_{02} denote that the quantities τ, n_p, n_m and λ_m are evaluated at these wavelengths; λ_{01}, λ_{02} are the wavelengths in vacuo. Figure 4(a) shows the turbidity ratio (at 400/600 nm) for polystyrene and poly(vinyl acetate) particles suspended in water. The particle diameter could be estimated from one turbidity ratio (two turbidity measurements) if the ratio was a monotonic function of particle size or α. However, this is not the case, as can be seen in Figure 4(a). An indentation exists in the turbidity ratio vs α curve, at the lower α regime, owing to the first inflection point of the K_{scat} vs α curve (in polydisperse suspensions, this indentation vanishes). Also, for larger particles, the turbidity ratio vs α curve exhibits an oscillatory behavior. Therefore, unless the size range is a priori known, a reading at a third wavelength (two turbidity ratios) is advised, to define uniquely the particle size. Again once the particle size is estimated the number of particles can be calculated from Equation (3).

An example where only one turbidity ratio is not sufficient to determine the particle size was reported by Kourti.[15] The turbidity ratio measurement at $\lambda_{01} = 400$ nm and $\lambda_{02} = 500$ nm for a 60 : 40 poly(styrene–butadiene) latex was repeated several times

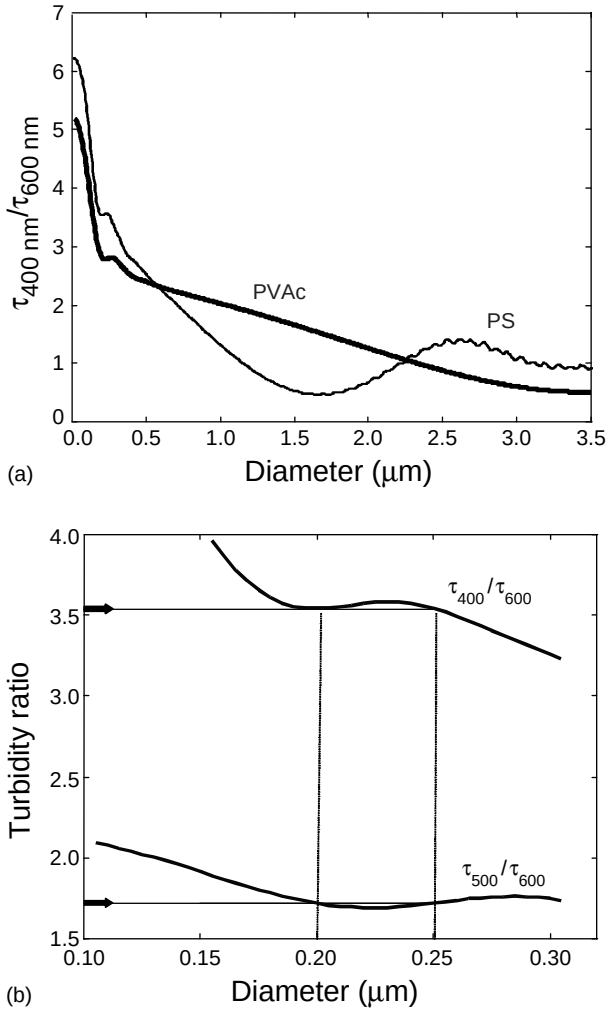

Figure 4 (a) Turbidity ratio at 400/600 nm plotted against particle diameter for polystyrene with $m \approx 1.2$ and poly(vinyl acetate) with $m \approx 1.1$ latexes. (b) Turbidity ratios for polystyrene, $m \approx 1.2$. Note that owing to the oscillation in the curves, particle suspensions with two different diameters (0.2 and 0.25 μm) give the same turbidity ratio values at two different wavelengths.

(for four different dilutions of the same sample). Each time, two estimates (two solutions) of particle size were obtained. In one of the solutions the estimates of the diameter were between 209 and 228 nm and in the other between 157 and 165 nm. Results from other methods had shown that the true diameter range was around 200 nm [specific turbidity (208–221 nm), HDC (228 nm) and DLS (194 nm)] and therefore the first solution could be picked as the correct one.

Caution should be exercised when diameters are in a range around the first indentation in the turbidity ratio–diameter curve when using the turbidity ratio method. Figure 4(b) shows turbidity ratios (τ_{400}/τ_{600} and τ_{500}/τ_{600}) plotted against diameter for monodisperse polystyrene suspensions. The diameters are in a range around the first indentation in the turbidity ratio–diameter curve. Note that suspensions with two very different particle sizes (0.2 and 0.25 μm) will give the same turbidity ratio at two different wavelengths (1.72 for τ_{500}/τ_{600} and 3.54 for τ_{400}/τ_{600}). In fact, for this example, with an experimental error as small as 2.0% in the turbidity ratios, any solution between 0.2 and 0.25 μm is possible. It would be preferable if the three wavelengths at which the measurements were taken were more widely separated.

The turbidity ratio method practically gives no information on the particle size for (i) very large particles, where for all the wavelengths used the scattering coefficient is constant (substitution of $K_{scat} = 2$ in Equation 13 gives a turbidity ratio equal to unity), and (ii) very small particles, where for all the wavelengths used the Rayleigh Equation (5) applies and turbidity ratio becomes independent of the particle size.

4 PARTICLE SIZE DISTRIBUTION DETERMINATION IN POLYDISPERSE SUSPENSIONS

Specific turbidity and turbidity ratio measurements can also be used to estimate the PSD in polydisperse suspensions. Numerical approaches that can be used to solve for the PSD from turbidity and specific turbidity measurements are discussed in detail in section 4.5. However, the estimation of the full PSD is not always possible with turbidimetry.[15,32] Some turbidimetric techniques are not theoretically expected to provide the full PSD of a suspension if, because of its optical properties (refractive indices of medium and particles) and the sizes of the particles, the (m, α) values of the suspension are within certain ranges. In such cases the specific turbidity technique will provide an estimate of an average diameter while techniques similar to the turbidity ratio provide no information on the particle size. The (m, α) regimes where the estimation of the full PSD is practically possible are discussed for each one of the turbidimetric methods presented in this section.

4.1 Specific Turbidity

Examples of successful applications of the method for particle and droplet size distribution determination can be found in Meehan and Beattie,[5] Walstra[16] and Yang and Hogg.[20] For a polydisperse suspension of spherical particles, the volume fraction is (Equation 14):

$$\phi = N \int_0^\infty \frac{\pi D^3}{6} f(D)\, \mathrm{d}(D) = N \frac{\pi}{6} J_3 \qquad (14)$$

where J_3 is the third moment of the PSD. The specific turbidity is given by Equation (15a):

$$\frac{\tau_{\lambda_0}}{\phi} = \frac{3}{2} \frac{\int_0^\infty D^2 K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right) f(D)\,\mathrm{d}(D)}{\int_0^\infty D^3 f(D)\,\mathrm{d}(D)} \quad (15a)$$

Alternatively, the specific turbidity of a suspension can be written as Equation (15b):

$$\frac{\tau_{\lambda_0}}{\phi} = \int_0^\infty \frac{\tau}{\phi}(D, \lambda_0) F(D)\,\mathrm{d}(D) \quad (15b)$$

where $\tau/\phi(D, \lambda_0)$ is the specific turbidity corresponding to particles with diameter D and wavelength λ_0 and $F(D)$ is the normalized PSD on a weight basis.

For a suspension of spheres, when the optical properties of the suspension are known, a minimum of P specific turbidity measurements are required for the estimation of a P parameter PSD from Equation (15a) or (15b), as discussed in the next section. For example, if we assume that the PSD is known to be log-normal described by the mean diameter and the variance, we need at least two specific turbidity measurements. If the form of the PSD is not assumed and we attempt to recover the frequency distribution at P given diameters (i.e. P discrete points), then at least P specific turbidity measurements are required.

In the presence of experimental error, the estimation of the full PSD from Equation (15a) or (15b) is impossible if all the turbidities (i.e. at all the wavelengths required for the estimation) are approximately proportional to the same moment of the PSD, J_y. Given that ϕ is always proportional to the third moment J_3 of the PSD, in these cases specific turbidity becomes proportional to J_y/J_3, that is proportional to the $(y-3)$rd power of the D_{y3} average diameter.[32] Kourti et al.[32] showed that when this happens for all of the P wavelengths used, many alternative solutions (PSDs) correspond to the same specific turbidity measurements and it is impossible to isolate the true solution unless there is some a priori knowledge. However, all these alternative solutions have their D_{y3} average diameters numerically very close to each other. In that case it was suggested that only this average diameter D_{y3} can be correctly estimated from specific turbidities[15,32] in this range of wavelengths. The (m, α) regimes where the full PSD is practically impossible to estimate for nonabsorbing particles are as follows:

1. For any m value and very small α (Rayleigh regime). If for all the wavelengths used for the turbidity measurements we are in the Rayleigh regime, then substitution of Equation (5) in Equation (15a) gives Equation (16):

$$\frac{\tau_{\lambda_0}}{\phi} = k_{\lambda_0} \frac{\int_0^\infty D^6 f(D)\,\mathrm{d}(D)}{\int_0^\infty D^3 f(D)\,\mathrm{d}(D)} = k_{\lambda_0} \frac{J_6}{J_3} = k_{\lambda_0} D_\tau^3 \quad (16)$$

where D_τ (or D_{63}) is the turbidity-average diameter and $k_{\lambda_0} = 4\pi^4 \lambda_m^{-4}[(m^2-1)/(m^2+2)]^2$ and the subscript λ_0 denotes that τ, λ_m and m are evaluated at λ_0. Llosent et al.[27] tried to estimate the full PSD in the Rayleigh regime using either an assumed shape of the PSD or using a regularization method (discussed later) to solve at discrete points for the PSD. Their results confirmed the suggestions of Kourti et al.[32] that only the turbidity-average diameter can be correctly estimated in these regimes.

2. For any m value and very large diameters (where $K_{scat} \approx 2$), we have Equation (17):

$$\frac{\tau_{\lambda_0}}{\phi} = 3 \frac{\int_0^\infty D^2 f(D)\,\mathrm{d}(D)}{\int_0^\infty D^3 f(D)\,\mathrm{d}(D)} = 3\frac{J_2}{J_3} = \frac{3}{D_{32}} \quad (17)$$

D_{32}, the volume–surface average, is also sometimes referred to as the Sauter mean diameter.

3. (m, α) regimes where the exponent y is close to 4.0. In an extensive study with theoretical analysis and experimental validation, Kourti et al.[15,32] showed that for suspensions with PSDs in (m, α) regimes where the exponent y is very close to 4.0 ($3.8 < y < 4.2$), specific turbidities result in apparent diameters numerically very close to the weight average (D_w or D_{43}) of the true PSD. The regions of (m, α) values where this happens can be read from a y vs α plot, such as that in Figure 2. From such a plot we can see that $y \approx 4$, for $m = 1.05$ at $3.0 < \alpha < 14$ and for $m = 1.15$ at $3.0 < \alpha < 6.0$. If the whole distribution covers α values in this regime for all the wavelengths used for the turbidity measurements, then within experimental error many alternative solutions are possible and only the weight-average of the distribution can be correctly estimated. This is true for any type of distribution (multimodal or unimodal continuous distribution, or simply the mixture of two populations). Table 3 gives an illustration of the effect of small measurement errors on the estimation of the PSD in these regimes, for $m = 1.1$. In case A, the true distribution is log-normal (see Equation 21) with $D_g = 175\,\mathrm{nm}$ and $\sigma = 0.2$ and weight-average diameter $D_w = 201\,\mathrm{nm}$. The turbidities for such a suspension were theoretically calculated at three

Table 3 Illustration of effect of small random measurement errors on the estimated PSD for $m = 1.1$[a]

Case	True PSD				Different random error for each test	Estimated PSD			
	D_g (nm)	σ	D_N (nm)	D_W (nm)		D_g (nm)	σ	D_N (nm)	D_W (nm)
A	175	0.2	178	201	A1	116	0.39	125	199
					A2	198	0.08	198	202
B	455	0.2	464	523	B1	520	0.002	520	520
					B2	336	0.362	359	533

[a] Two cases with different true PSDs are examined: (A) $D_g = 175$, $\sigma = 0.2$ and (B) $D_g = 455$, $\sigma = 0.2$. A maximum error of $\pm 1.2\%$ was introduced randomly in the specific turbidities at several wavelengths. Then the PSD was re-estimated. For each case two different random errors were tried, and two very different PSDs were obtained, both far from the true solution, but with D_w close to the true solution.

wavelengths. An error of less than 1.0% was then introduced in the three turbidity values and the parameters of the PSD were re-estimated (case A1). Note that the estimated $D_w = 199$ is within 1% of the value of the true weight average of the PSD whereas D_g and σ are significantly different. A different random error less than 1% was introduced in the three turbidities for case A2. Again, only the weight average of the estimated distribution $D_w = 202$ is within 0.5% of the true value and the other parameters are significantly different from the true distribution but also from case A1. Similar results were obtained for case B. For different random errors of $\pm 1.2\%$ on the turbidity values, cases B1 and B2 were obtained with weight averages within 2% of the true value but with all the other parameters significantly different from the true distribution.

Reports in the literature confirm that the PSD cannot be estimated with specific turbidity in some cases. Haseler[19] working with Ektacolor suspensions ($m = 1.15$) reported that experimental errors in specific turbidities (measured at 400 and 700 nm) would lead to large errors in the calculated distribution. These results were observed because of the small (m, α) values with which he was working. Based on the above discussion, only D_w could be correctly estimated in those (m, α) regimes. Walstra,[16] trying to estimate the globule size distribution in cow's milk ($m = 1.1$) from specific turbidity measurements, reported that for very fine emulsions ($D_{32} < 200$ nm) with rather high polydispersity the reliability of the method decreases and showed that a set of experimental data could be fitted to many different types of distributions with the same D_w.

The observations of Kourti et al.[15,32] revealed a very useful property of the specific turbidity: the weight-average diameter of suspensions can be correctly estimated even though the determination of the PSD is impossible. This property of specific turbidity can be utilized for practical purposes where a fast estimate of the weight-average diameter is required and is sufficient to monitor the progress of a process. Gossen[23,24,46] utilized this information and used specific turbidity to monitor the progress of polymerization reactions with on-line measurements. The α range where specific turbidity measurements give D_w may be extended further when the distribution is log-normal. For example, for suspensions with $m = 1.1$ and log-normal distributions of submicrometer particles, it has been observed that the mean value of the apparent diameters, estimated from specific turbidities at 400 and 600 nm, is within 5% the true weight average diameter of the PSD.[15,32] Monitoring these two apparent diameters gives information on the weight average of the PSD.

It should be noted that in all the regimes where the estimation of the full PSD is impossible, specific turbidity gives good estimates of meaningful average diameters. Compared with other turbidimetric techniques, where information only from turbidity measurements is used, the specific turbidity has the advantage that an additional measurement is used, the solids volume fraction, which is always proportional to the third moment of the PSD. Therefore, whenever turbidities at two or more wavelengths are proportional to the yth moment of the distribution, although the estimation of the true PSD is impossible, a meaningful average of the PSD, the D_{y3} average, can always be obtained from specific turbidity measurements.

4.2 Turbidity Ratio

The ratio of two turbidity readings taken at different wavelengths for a polydisperse suspension is given by Equation (18):

$$\frac{\tau_{\lambda_{01}}}{\tau_{\lambda_{02}}} = \frac{\int_0^\infty D^2 K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right)_{\lambda_{01}} f(D)\,\mathrm{d}(D)}{\int_0^\infty D^2 K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right)_{\lambda_{02}} f(D)\,\mathrm{d}(D)} \tag{18}$$

Again, subscripts λ_{01} and λ_{02} denote that the quantities τ, n_p, n_m and λ_m are evaluated at the corresponding wavelengths; λ_{01}, λ_{02} are wavelengths in vacuo. For a PSD described by P parameters, a minimum of $(P + 1)$ turbidity measurements at different wavelengths (one of

which is chosen as reference) are required, in principle, for the estimation of these parameters. The method gives no information on the particle size in the following cases, for nonabsorbing particles:

1. For any value of m and very large particles where $K_{scat} = 2$, the turbidity ratio equals 1.0.
2. For any value of m and very small particles, where the Rayleigh Equation (5) applies, the turbidity ratio equals a constant value, independent of the particle size.
3. When, for all the wavelengths used, turbidities are approximately proportional to the same moment of the PSD. This case is observed for very small m values ($m \approx 1.15, 1.1, 1.05$) and for a very wide range of α values below the first K_{scat} maximum, where the exponent y changes very slowly with α. This in turn means that even for broad distributions with sizes falling in these α regimes, turbidities at widely separated wavelengths will be proportional to approximately the same moment of the PSD, and their ratio will be almost independent of the particle size. Kourti et al.[15,32] showed that for $m = 1.1$ and submicrometer particles, the estimation of the true PSD from turbidity measurements at 350 and 700 nm (with reference $\lambda_0 = 546$ nm) is impossible for measurement errors as small as 0.5%. One example is reported in which the effect of error on the estimated PSD is studied for a turbidity ratio error as small as $\pm 0.35\%$ in the above three wavelengths. The PSD was re-estimated for different errors (less than $\pm 0.35\%$) each time. Several very different log-normal distributions were obtained; for example, one of the estimated PSDs had a geometric mean $D_g = 500$ nm and standard deviation $\sigma = 0.15$, $D_w = 541$ nm and polydispersity 1.05, while another had $D_g = 285$ nm, $\sigma = 0.4$, $D_w = 499$ nm and polydispersity 1.6.

The turbidity ratio method should therefore be avoided for the estimation of the PSD for systems of very small or very large particles and for systems with small m values ($m < 1.15$) and α values below approximately 10 (for a short wavelength of 400 nm and water medium, these α values cover all the submicrometer size range). In these regimes the method is so sensitive to experimental error that it offers practically no information on the size of the suspended particles (i.e. neither the PSD nor an average can be estimated).

Owing to the behavior of the turbidity ratio in the ranges of (m, α) values described above, controversial reports on the capabilities of the turbidity ratio to provide the full PSD have appeared in the literature. Wallach and Heller[7] obtained estimates very close to the true distribution for polystyrene suspensions ($m = 1.2$) with particle diameters between 650 nm and 1.3 μm from turbidity measurements at $\lambda_{01} = 440$ nm, $\lambda_{02} = 610$ nm and $\lambda_{03} = 546$ nm. The values of m and α in their system were such as to allow for successful PSD determination. However, when Haseler[19] tried the turbidity ratio method for a system with smaller m and α values ($m = 1.15$, λ_0 in the range 400–700 nm and submicrometer particles), he reported that within experimental error a large number of different distributions gave the same turbidity spectra. Similarly, Maxim et al.[31] reported that the "turbidity ratio leads to multivalued solutions and unless prior estimates of the answer by some other method are available there are no criteria for choosing between alternative solutions". They were working with small m values ($m = 1.1$), submicrometer particles and wavelengths of 320–720 nm with a reference wavelength of 546 nm, that is an (m, α) area where Kourti[15] showed that even a 0.5% error in the turbidity ratio measurements means that the PSD estimation is practically impossible. The results of both Haseler and Maxim et al. were expected based on the (m, α) values of their systems. It is clear from the above examples that one cannot generalize conclusions on the capability of turbidimetric techniques to provide the PSD across (m, α) values.

4.2.1 Methods Similar to Turbidity Ratio

Alternatively to Equation (18), some authors[47,48] use the ratio given in Equation (19):

$$\frac{(\tau\lambda/\phi)_{\lambda_{01}}}{(\tau\lambda/\phi)_{\lambda_{02}}} \tag{19}$$

Notice that, numerically, ϕ is eliminated from Equation (19), hence there is no reason to measure it. Other authors use the DQ (dispersion-quotient) method (Equation 20):[33]

$$DQ = \frac{\tau_{\lambda_{01}}\lambda_{02}^2}{\tau_{\lambda_{02}}\lambda_{01}^2} \tag{20}$$

Both of the above approaches have the same limitations as the turbidity ratio method for the estimation of the PSD.

4.3 The Wavelength Exponent Technique

In this approach, the particle size can be estimated only from turbidity measurements as in the turbidity ratio technique (i.e. the particle volume fraction does not need to be known). The change of the turbidity of a suspension with the wavelength is recorded and related to the size of the suspended particles. The technique is described in detail in Bateman et al.,[49] Heller et al.[43] and Kourti.[15] It has been shown[15,50] that the turbidity ratio and the wavelength exponent are theoretically equivalent

methods for the determination of the parameters of the PSD. The functions used in the formulations of the two techniques to convert the turbidity measurements to PSDs can be derived from each other, and in case of no measurement error the use of either method gives identical results. Consequently, in the (m, α) regions where the estimation of the particle size in a suspension is not possible using the turbidity ratio method, the wavelength exponent method also fails. It should also be noted that although the two methods are theoretically equivalent for the determination of the PSD, the formulation of the turbidity ratio method is less complex than that of the wavelength exponent. Consequently, the turbidity ratio method requires less computation to calculate the PSD and the error propagation is smaller.

Melik and Fogler[48] suggested an approach in which a combination of the wavelength exponent and the turbidity ratio method is used and claimed that with this approach the two parameters of the PSD can be estimated from only two turbidity measurements. However, since these are not two different methods but are theoretically equivalent, at least three turbidity measurements are required for the determination of a two-parameter PSD using either of the methods or their combination. Otherwise (i.e. using two turbidity measurements only), an infinite number of solutions will be obtained. The claims of Melik and Fogler[48] were shown[50] to be unfounded.

4.4 Resolution of a Conflict in the Literature

The analysis presented by Kourti et al.[15,32] which showed that the capability of a turbidimetric technique to provide the PSD is strongly related to the (m, α) values of the suspension helped to resolve a controversy that existed in the literature. This conflict, on whether the true PSD of a suspension can be estimated from turbidimetric techniques, arose between Maxim et al.[31] and Zollars.[11,12] As discussed earlier, Maxim et al. used the turbidity ratio method to estimate the full PSD in commercial suspensions, with $m = 1.1$ and particles with diameters smaller than 1.0 μm. A high correlation between the parameters of the distribution was observed. The authors stated that a valid estimate of the true distribution is impossible in the presence of experimental error. According to the analysis presented by Kourti et al.[32] these observations are justified (and were expected) for the (m, α) values of the suspension and the turbidity ratio method. Zollars used a simulation study with no experimental error to re-estimate the parameters of log-normal PSDs from theoretically calculated specific turbidities. Of course, the parameters were correctly re-estimated. The author stated that ''turbidimetry does not exhibit multiple solutions as reported by other investigators''. His observations[11,12] are also justified for a specific turbidity method and re-estimation of the parameters with no experimental error. Both the workers were correct for the types of cases they were studying. However, their findings cannot be extrapolated over turbidimetric methods and (m, α) regimes. This extrapolation caused the controversy.

4.5 Solving for the Particle Size Distribution

Equations (15a) and (18) give the value of the specific turbidity or turbidity ratio, respectively, at given wavelengths if the PSD, $f(D)$, is known. The inverse problem is to determine the PSD given the turbidity or the specific turbidity measurements at several wavelengths, λ_{0i}. If a continuous set of measurements of τ_{λ_0} vs λ_0 was available, one could theoretically deconvolute these equations to yield $f(D)$. However, in practice, one usually has turbidity measurements at a finite number of wavelengths, so we have to reconstruct $f(D)$ from only a number of discrete values of $\tau_{\lambda_{0i}}$. This is a typical problem in science.[51] We routinely measure ''enough points'' and then ''draw a curve through them''. In doing so for this problem, we are making some assumptions either about the underlying function $f(D)$ or about the nature of K_{scat}, or both.

Two approaches have been taken in the literature to solve for $f(D)$. When a small number of turbidity measurements are available, a type of distribution that can be described by a few parameters is assumed for $f(D)$ (i.e. Gaussian, log-normal, etc.) and we solve for these parameters. When many turbidity measurements are available it is possible in some cases to solve for f_j [the discretized $f(D)$] at a number of selected diameters, D_j. In general, the second approach requires many more turbidity measurements than the first.

4.5.1 Approximate the Particle Size Distribution with a Known Distributional Form

When a small number of turbidity measurements are available, a convenient approach is to approximate $f(D)$ with a distributional form that can be described by a few parameters. This form is then substituted into Equation (15a) or (18) and its parameters are estimated from the turbidity measurements using non-linear parameter estimation procedures such as the Levenberg–Marquardt algorithm.[52] The assumed distributional form is one that is known to describe similar systems (e.g. for latex particles the log-normal distribution described by two parameters is frequently used).

Extensive reviews on the distributional forms that have been used to describe the PSD in various suspensions can be found in several sources.[15,33,53] Some workers[1] calculated turbidity for a Gaussian distribution of spheres. However, most naturally occurring

populations are skewed and not symmetrical. The logarithmic normal distribution[11,15,20,33] has been used to described such populations. In this distribution $\ln D$ is normally distributed. The normalized PSD, $f(D)$ is given by Equation (21):

$$f(D) = \frac{1}{\sigma D\sqrt{2\pi}} \exp\ -\frac{(\ln D - \ln D_g)^2}{2\sigma^2} \qquad (21)$$

where $\ln D_g$ is the mean of $\ln D$ and D_g is both the median and the geometric mean diameter; σ is the standard deviation of $\ln D$ and it is called the geometric mean standard deviation. For small values of σ ($\sigma < 0.14$) the frequency curve can be closely approximated by a normal distribution.[33] The nth moment of the log-normal distribution is given by Equation (22):[54]

$$J_n = E(D^n) = \exp\ n(\ln D_g) + \frac{(n\sigma)^2}{2} \qquad (22)$$

The number- (D_N), volume–surface- (D_{32}), weight- (D_w) and turbidity- (D_τ) average diameters can therefore be calculated as (Equation 23):[15]

$$D_N = \frac{J_1}{J_0}; \quad D_{32} = \frac{J_3}{J_2}; \quad D_w = \frac{J_4}{J_3}; \quad D_\tau = \left(\frac{J_6}{J_3}\right.^{1/3} \qquad (23)$$

The calculation of the moments of the estimated PSD from Equation (22) are correct only if the integration is done from $D_{min} = 0$ to $D_{max} = \infty$ or, in practice, if the upper limit of the PSD, D'_{max}, is such that when using $f(D)$ from Equation (21), we have Equation (24):

$$\int_0^{D'_{max}} f(D)\,\mathrm{d}(D) = 1.0 \qquad (24)$$

Sometimes it may be known a priori (either from another method such as EM, DLS or particle chromatography, or by filtering of the suspension) that particle sizes smaller than D_{min} and/or larger than D_{max} do not exist. In these cases if we wish to estimate a log-normal envelope we can do so. In the parameter estimation all the integrations should be done between D_{min} and D_{max}. A log-normal envelope, characterized by D_g and σ, can be estimated; however, the PSD should be normalized before each iteration in the parameter estimation. Furthermore, the calculation of the various moments should be done using the explicit equation for the moments (Equation 25):

$$J_n = \int_{D_{min}}^{D_{max}} D^n f(D)\,\mathrm{d}D \qquad (25)$$

and not Equation (22). The log-normal distribution estimated with previous knowledge of D_{min} and D_{max} is expected to be a more accurate estimate of the true PSD, since the contribution of tails of very large particles to the theoretically calculated specific turbidity is eliminated.

Note that if Equation (22) can be directly used, then the polydispersity of a suspension with a log-normal distribution is given by $D_w/D_N = \exp(3\sigma^2)$ and the particle volume fraction is given by Equation (26):

$$\phi = ND_g^3 \exp(4.5\sigma^2) \qquad (26)$$

Some workers[4,5] had named as log-normal the distribution with the frequency function in Equation (27):

$$f'(D) = Df(D) \qquad (27)$$

The two distribution functions are related by Equation (28):[55–57]

$$f(D)\,\mathrm{d}D = f'(D)\,\mathrm{d}\ln D \qquad (28)$$

Therefore, if $f'(D)$ is used, then Equation (4) should be integrated with respect to $\ln D$[29] and not with respect to D. This distribution described by $f'(D)$, has been called a ZOLD.

When applying Equation (15b), $f(D)$ should be expressed on a weight basis. A log-normal distribution on a weight basis can be used.[20,29,58] If the number distribution is log-normal, then the weight distribution is also log-normal with the same geometric standard deviation σ (Equation 29):[54,59]

$$F(D) = \frac{1}{\sigma D\sqrt{2\pi}} \exp\ -\frac{(\ln D - \ln D_{wm})^2}{2\sigma^2} \qquad (29)$$

where $F(D)$ is the corresponding weight distribution for $f(D)$ of Equation (21). D_{wm} is both the median and the geometric mean diameter on a weight basis. D_{wm} and D_g are related by Equation (30):

$$D_{wm} = D_g \exp(3\sigma^2) \qquad (30)$$

Heller et al.[7,47,53] used another type of distribution that can be described by two parameters. However, the maximum polydispersity for that distribution is $D_W/D_N = 1.5$.[15] Therefore, even though the distributional form can describe positively skewed distributions, it cannot successfully describe highly polydisperse systems.

4.5.2 No Assumptions on Shape of Particle Size Distribution

If no distributional form is assumed for $f(D)$, then an alternative is to solve for $f(D)$ at a finite number of points.[8,60–64] In this approach, in the space of diameters from D_{min} to D_{max} we define P points (i.e. P diameters D_j, $j = 1:P$) and we try to estimate f_j at the corresponding diameter D_j. The number of P points must be less than

or equal to the number of specific turbidity or turbidity ratio measurements made.

Note that if the integrand of Equation (4) is discretized at $P-1$ intervals, then turbidity at a given wavelength λ_{0i} can be approximated by Equation (31):

$$\tau_{\lambda_{0i}} \approx \sum_{j=1=D_{\min}}^{j=P=D_{\max}} f_j \gamma_{ij} \tag{31}$$

where $f_j \approx Nf(D_j) = N_j$ is the discretized PSD and γ_{ij} depend on the kernel and the quadrature equation used.[27,61] γ_{ij} is evaluated at given D_j from the Mie theory. Note that for the simple case that the suspension is a mixture of P monodisperse suspensions with diameters $D_1, D_2, \ldots, D_P$, with relative number of particles $N_1, N_2 \ldots, N_P$, we have Equations (32) and (33):

$$\gamma_{ij} = \frac{\pi}{4} D_j^2 K(\lambda_{0i}, D_j) \tag{32}$$

$$\sum_{j=1}^{P} N_j = N \tag{33}$$

When turbidity is calculated for L wavelengths then Equation (31) can be written as Equation (34):

$$\boldsymbol{\tau} \approx \boldsymbol{\Gamma f} \tag{34}$$

where $\boldsymbol{\tau} = [\tau_{\lambda_{01}}, \tau_{\lambda_{02}}, \ldots, \tau_{\lambda_{0L}}]^{\mathrm{T}}$, $\boldsymbol{\Gamma} = \{\gamma_{ij}\}$, $\boldsymbol{f} = [f_1, f_2, \ldots, f_P]^{\mathrm{T}}$ and T denotes transpose.

In the inverse problem, we have to estimate $\boldsymbol{f}$ from turbidity measurements $\boldsymbol{\tau}^*$ (Equation 35):

$$\boldsymbol{\tau}^* = \boldsymbol{\Gamma f} + \boldsymbol{\varepsilon} \tag{35}$$

where $\boldsymbol{\varepsilon}$ is the vector of errors (measurement errors plus errors due to discretization of the true distribution). Solving Equation (35) is a linear algebra problem.[65] The least-squares solution is Equation (36):

$$\boldsymbol{f}_{\mathrm{ls}} = (\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\tau}^* \tag{36}$$

However, because the rank of $\boldsymbol{\Gamma}$ is incomplete, small errors in $\boldsymbol{\tau}^*$ result in large errors in the estimated parameters. For this reason, regularization methods[65] should be used. A penalty function is introduced to the original least-squares problem (Equation 37):

$$\min_{\hat{f}} \{|\boldsymbol{\Gamma}\hat{\boldsymbol{f}} - \boldsymbol{\tau}^*|^2 + \xi q(\hat{\boldsymbol{f}})\} \tag{37}$$

where ξ is a nonnegative parameter and $|\ |$ indicates the modulus. $q(\hat{\boldsymbol{f}})$ could be the sum of squares, or the first or second differences, of the estimates. Eliçabe and Garcia-Rubio[61,62] added another term to the sum of squares of the second differences in order to force f_1 and f_{p} to zero (Equation 38):

$$q(\hat{f}) = \beta^2(\hat{f}_1^2 + \hat{f}_2^2) + \sum_{j=2}^{P-1}(2\hat{f}_j - \hat{f}_{j-1} - \hat{f}_{j+1})^2 \tag{38}$$

with $\beta \gg 1$. The solution now is:

$$\boldsymbol{f} = (\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma} + \xi\boldsymbol{H})^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\tau}^* \tag{39}$$

where (Equation 40):

$$\boldsymbol{H} = \begin{matrix} 1+\beta^2 & -2 & 1 & 0 & 0 & 0 & \ldots & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 & \ldots & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & \ldots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1+\beta^2 \end{matrix} \tag{40}$$

Note that in this method one has to choose the location of $D_{\min}$ and $D_{\max}$ and the number of intervals but also the value of ξ. The choice of these parameters was discussed in detail by Eliçabe and Garcia-Rubio.[61,62]

This technique was tested with simulations[61,62] for the estimation of unimodal and bimodal PSDs of varying breadth and mean diameter. Satisfactory results were reported[61] for broad distributions with particle diameters in the micrometer range (distributions with $D_{\max} = 1.45$ or $2.25\,\mu\mathrm{m}$). For $m = 1.2$ the results were not satisfactory for distributions covering sizes from 50 to 325 nm, that is, a regime where a large number of latex particle diameters are expected to be found.

More attempts to resolve the PSD with no assumptions were reported later by other workers.[63,64] Finally, an empirical method that relates turbidity measurements to particle size using latent variable methods has also been reported.[25]

Gulari et al.[8,60] tried to estimate PSDs in polystyrene suspensions ($m \approx 1.2$), from a large number of turbidity measurements at the wavelength range between 300 and 800 nm. In each case, the unknown PSD was assumed to be the sum of 20 equally spaced delta functions (γ_{ij} from Equation 32). The height (mass fraction) and position of each diameter was optimized using a nonlinear least-squares minimization procedure. During the analysis, the insignificant parameters are zeroed and the parameter matrix collapses until convergence is reached. The suspensions tested in that study were synthesized by mixing two or three monodisperse latexes, with particle diameters approximately 1 μm separated from each other. The distributions were resolved on a weight basis. In each distribution the weight fraction of the small particles was significant. Users should be careful when using this approach (i) when a very small weight fraction of the small particles is present (see discussion on bimodal

distributions, section 7.3) and (ii) for suspensions with (m, α) values where y does not change significantly with α. Readers should also keep in mind that the resolution of the method depends on the location of the assumed delta functions (number and width of ΔD intervals) and of course on the number of measurements (observations) available.

In general, the estimation of $f(D)$ at a finite number of points is expected to give satisfactory results only for certain suspensions (i.e. corresponding to certain m and α values). The results will be questionable in (m, α) regimes where y does not change significantly with α. Kourti and MacGregor[44] emphasized that more sophisticated deconvolution methods cannot overcome the limitations that turbidity ratio and specific turbidity have for particles in the (m, α) regimes where turbidity is simply proportional to a moment of the PSD. Several years later, Llosent et al.[27] confirmed this by trying several deconvolution approaches to estimate the PSD in the Rayleigh regime; only turbidity-average diameters could be correctly estimated and not the PSD.

4.5.3 The Method of Moments

This method was presented[9,10] as an alternative technique to solving for the PSD, but has some limitations for the particular application it was used. In this approach the term $K(D/\lambda_m, n_p/n_m)\pi D^2/4$ in Equation (4) is approximated with a polynomial (Equation 41):

$$K\left(\frac{D}{\lambda_m}, \frac{n_p}{n_m}\right)\frac{\pi D^2}{4} \approx \sum_{i=0}^{P} C_{i\lambda} D^i \tag{41}$$

Turbidity can thus be expressed as a sum of P moments of the unknown PSD. Using P turbidity ratios ($P+1$ turbidity measurements) one can estimate the P moments (J_0 of the normalized distribution is unity). The PSD is then represented analytically using a set of its moments, by modifying a gamma distribution which serves as a basis. The method was applied for particle size determination in poly(vinyl acetate) latexes.[9,10] First, the $C_{i\lambda}$ coefficients at each wavelength were estimated from Equation (41) by calculating the value of $K_{scat}\,\pi D^2/4$ exactly from the Mie theory for a set of 20 diameters (between 0.04 and 0.8 μm) and setting the value of P equal to 6. Six $C_{i\lambda}$ were estimated for each wavelength. The procedure was repeated for seven wavelengths (350–800 nm).

As discussed in a critical review,[15] the expansion of the scattering coefficient to a power series of α (or for constant λ_m, to a power series of D) is an approximation applicable for small α and $m\alpha$ values. In these cases the coefficients of α^i (or D^i) are known functions of the optical constants[37] (i.e. no estimation of $C_{i\lambda}$ is necessary). The α values corresponding to the diameters and wavelengths reported in the particular application[9,10] were very large (α as large as 9.5). For larger m and α values the approximating equations for the calculation of K_{scat} include more involved functions[37] of α [or $\rho = 2\alpha(m-1)$] than the simple power series expansion attempted in the method of moments. As discussed earlier, the term $K_{scat}\,D^2$ can be set proportional to D^y for larger values of α, and the value of y changes as a function of the particle size. For poly(vinyl acetate) latexes in particular, and for a very wide range of α values, the value of y is very close to 4.0. For most of the diameters used for the calculation of the above term in that study[9,10] at each wavelength, the α values were such that y was close to 4.0. It is clear that a reasonable value from the parameter estimation is expected only for $C_{4\lambda}$, since all the other coefficients are practically equal to zero. From the results reported by Kiparissides,[9] only $C_{4\lambda}$ seems to have a consistent value for all the wavelengths; the estimated values of all the other coefficients change, both in sign and value, randomly.[15] It should be clear that with meaningless estimates of the $C_{i\lambda}$ coefficients the correct estimation of the moments of the PSD is impossible.

5 ESTIMATION OF AVERAGE DIAMETERS

Several workers have tried to use turbidimetry to estimate only an average of the PSD of a polydisperse system. To do this they treat the system as monodisperse. That is, for specific turbidity, instead of using Equation (15a) they use Equation (12) and estimate a single, average diameter which is termed an ''apparent diameter'' (Equation 42):

$$\frac{\tau_{\lambda_0}}{\phi} = \frac{3}{2}\frac{K(\alpha_{av}, m)}{D_{av}} \tag{42}$$

where $\alpha_{av} = \pi D_{av}/\lambda_m$. D_{av} is the apparent diameter that always corresponds to the D_{y3} average of the PSD.

The concept of the ''apparent'' diameter was introduced and theoretically explained in very careful work by Meehan and Beattie.[4,5] They commented that the apparent diameter obtained by treating a polydisperse system as monodisperse corresponds to an average of the unknown PSD. They also concluded that different kinds of averages are yielded by different light scattering methods, and that the kind of average obtained by a light scattering method depends on the value of m of the suspension, the type of the PSD and the actual size of the particles in the suspension. Their comments have been corroborated by numerous reports[3,4,7,15,16,29,66] where it is either discussed and/or shown experimentally that the type of average obtained from the apparent diameter depends on the turbidimetric technique used and the m and α values of the suspension.

It has been suggested[7] that the ratio of the apparent diameters estimated at two wavelengths can provide a rough idea of the polydispersity of the system. However, this is not always true. It has already been shown that, sometimes, especially for small m values ($m \approx 1.1$), the apparent diameters corresponding to very broad distributions or to bimodal distributions do not change significantly with the wavelength, leaving the impression that the system is monodisperse. Note that for monodisperse systems, the apparent diameters obtained from two different turbidimetric methods (i.e. specific turbidity, turbidity ratio) should be numerically very close to each other (i.e. equal, within experimental error).

Specific turbidity has been the only turbidimetric method used in the literature for the determination of averages of the PSD. This is because, as already discussed, it is the only turbidimetric method that can provide estimates of meaningful averages of the PSD (D_{y3} averages) at some well-defined (m, α) regimes. The following averages may be obtained from specific turbidity: (i) the turbidity-average for very small particles ($\alpha < 0.5$) and any value of m, (ii) the volume–surface average for very large particles and any m value and (iii) weight-average for a combination of (m, α) values and $m < 1.2$ which can be identified very easily for the system under investigation. In this section we discuss the use of specific turbidity for the estimation of meaningful average diameters. A critical discussion of several applications reported in the literature is presented, so that the readers can identify with situations closer to their needs.

5.1 Turbidity-average Diameters

These averages can be obtained from specific turbidity measurements, for suspensions with any value of m, provided that all the particles in the suspension are small compared with the wavelength at which the turbidity measurements are taken (Rayleigh regime).

A practical application, where the growth of silver bromide grains is monitored on-line by following turbidity-average diameters, was reported by Haseler and Parkin[6] for an Ostwald ripening process. They were careful to choose long wavelengths [infrared (IR) regime, 2.22 μm] so that the Rayleigh scattering theory could be applied even for the largest particles in their system.[15]

Note that in the Rayleigh regime, if two particle suspensions have the same value of m, the ratio of their specific turbidities measured at the same wavelength (see Equation 16) gives the ratio of their turbidity-average diameters, independent of the value of m (Equation 43):

$$\frac{\tau_1/\phi_1}{\tau_2/\phi_2} = \left(\frac{D_{\tau_1}}{D_{\tau_2}}\right)^3 \qquad (43)$$

This characteristic can be exploited to provide a fast and inexpensive tool to monitor particle growth in a suspension where the value of m is not accurately known. By simply measuring specific turbidity (turbidity and solids fraction) at each time interval, the relative increase in the turbidity-average diameter as a function of time can be recorded. From a measurement of the diameter or the PSD with another technique (which is independent of m) at the end of the process, we can back-calculate the turbidity-average diameter at each instant in time.

A report[28] that a correct turbidity-average can be obtained for α values as large as 4.74 from specific turbidity measurements should be treated with caution. These observations were rather coincidental with the polydispersities of the systems investigated, as explained in detail elsewhere.[15] In the same work[28] the term ''weight'' average was used for what in fact is the ''turbidity'' average, as can be deduced from the definition in the original publication,[67] where EM results for the latex samples used in the later study[28] were presented.

5.2 Volume–Surface-average Diameters

Volume–surface-average diameters have been estimated in the literature from specific turbidity measurements of suspended particles that were very large compared with the wavelength ($K_{scat} \approx 2$). It should be noted that, although the value of the oscillating K_{scat} stays very close to 2.0 for large α values, the error by assuming it to be equal to 2.0 is small only for very large α values (i.e. the error is within 3% for $\alpha > 800$ and within 1% for $\alpha > 2000$) (see Table 2). For α values around 200–260, where some of the cases reported in the literature are, the error of assuming $K_{scat} = 2.0$, depending on the spread of the distribution, could be as large as 6% for $m = 1.2$ and 9% for $m = 1.1$ (at these α values K_{scat} still depends on m). Therefore, depending on the range of α values covered by the true distribution, the estimated D_{32} average diameter could be within 10% error of the true value.

Bagchi and Vold[66] showed experimentally that for very large particles, the specific turbidity varied linearly with the inverse of the volume–surface-average diameter ($1/D_{32}$). However, a linearity observed with respect to the inverse of other averages was coincidental and attributed to the fact that the degrees of polydispersity of the latexes investigated were very similar.[15] Kubota et al.[68] discussed this coincidental linearity and presented more experimental results from suspensions of various polydispersities, verifying that only the specific turbidity vs $(1/D_{32})$ curves are linear, as was theoretically expected.

Kourti[15] showed that a claim by Dobbins[69] and Dobbins and Jizmagian[70] that the volume–surface-average diameter can be obtained from one specific turbidity measurement in any particle size regime cannot

be supported. Their observations, that ''the specific turbidity is primarily dependent on the D_{32} diameter and only weakly dependent on the type of the distribution'', were shown to be coincidental with the types of distributions that they used for their investigation. These distributions had also their weight averages close to each other and all of them included particle sizes in the same range. Similar claims by Haseler[19] were also discussed in Kourti.[15] It was shown that in the examples reported by Haseler,[19] the estimated D_{32} was close to the true value because the location of the whole distribution was correctly estimated, and not because specific turbidity is always expected to give a correct D_{32} as claimed.[69,70]

It has been suggested[71] that the specific turbidity of a polydisperse system can be expressed as Equation (44):

$$\frac{\tau_{\lambda_0}}{\phi} = \frac{3}{2}\frac{K(\alpha_{\mathrm{av}}, m)}{D_{32}} \tag{44}$$

where (Equation 45):

$$K(\alpha_{\mathrm{av}}, m)\int_0^\infty D^2 f(D)\,\mathrm{d}(D) = \int_0^\infty D^2 K(\alpha_{\mathrm{av}}, m) f(D)\,\mathrm{d}(D) \tag{45}$$

where α_{av} corresponds to an average diameter D_{av}. It was suggested that the two parameters α_{av} and D_{32} can be used to interpret turbidity data over a wavelength range. The D_{32} diameter will correspond to the true PSD whereas α_{av} will correspond to the average diameter ''seen'' by the measurements ''averaged over the wavelength range''. However, by definition, the only correct form for the calculation of an average diameter from specific turbidity measurements is that of Equation (42). Equation (44) is valid only for very large diameters, where $K(\alpha_{\mathrm{av}}, m) = 2$, independent of particle size. In any other case if we try to estimate the two parameters (α_{av} and D_{32}) in Equation (44), ''over a range of wavelengths'', these estimates will be highly correlated. Their numerical values will depend on the wavelength range used and the D_{32} estimate will not, in general, correspond to the volume–surface-average of the true PSD.

5.3 Weight-average Diameters

For suspensions with small values of m (usually $m < 1.15$) and particle sizes corresponding to certain α values (i.e. for $m = 1.05, 3 < \alpha < 14$; for $m = 1.15, 3 < \alpha < 6$), the apparent diameters obtained from specific turbidity measurements are numerically close to the weight-average diameter, D_{w}. This average is essentially the only correct estimate of the particle size that can be obtained for polydisperse systems in these (m, α) regimes. If users know the value of m of their suspensions they can devise simple ways to obtain this average. For example, it has been observed[15] that the mean value of the apparent diameters from specific turbidities at 600 and 400 nm, obtained for a variety of polydisperse poly(vinyl acetate) latexes ($m \approx 1.1$) with different distributions (unimodal, multimodal) in the submicrometer regime, is numerically very close to the true weight average of the distribution. That means that for $m = 1.1$ the weight average from submicrometer particle suspensions can be estimated very simply, from two specific turbidity measurements at 400 and 600 nm. Similar heuristic rules could be derived for other suspensions with small m values.

Cheesman,[72] for example, suggested a technique in which the specific turbidity is measured at several wavelengths and the apparent diameter (D_{av}) is estimated for each wavelength. Then the corresponding α_{av} values are calculated ($\alpha_{\mathrm{av}} = \pi D_{\mathrm{av}}/\lambda_{\mathrm{m}}$) at each wavelength. The apparent diameter that gives α_{av} near the α value of the first point of inflection of the K_{scat} vs α curve is chosen to be the weight-average diameter (for $m = 1.1$, this α value was chosen 2.7 and for $m = 1.2$, it was 2.2). The apparent diameter selected this way in the polydisperse systems he studied was within 4% of the true weight-average diameter.

The observations of Cheesman can be explained theoretically. At the α value at the inflection point (see peaks of the first ripple of the y vs α curve in Figure 2) the value of y, for almost any value of m, is close to 4.5. If the weight-average diameter corresponds to this α value, then the α values for the individual particle diameters of the distribution are located around it and the value of y varies between 4 and 4.5. Hence in this technique, the wavelength is chosen such that the entire distribution is brought to an α regime where y is close to 4.0 and the apparent diameter is the weight-average diameter.

As has been pointed out,[44] the fact that the weight-average diameter of suspensions can be correctly estimated even though the PSD determination is impossible is a very useful property of the specific turbidity. It was suggested[44] that this property could be utilized for practical applications, in situations where following an average diameter is sufficient to monitor the progress of a process. An example would be on-line particle size measurements to monitor the progress of polymerization reactions. Gossen[23,46] and Gossen and MacGregor[24] used this suggestion and applied specific turbidity on-line to monitor weight-average diameters of poly(vinyl acetate) particles during latex production. The results reported[23,24,46,73] show good agreement between the weight-average diameters obtained by turbidimetry and those obtained by on-line DLS.

5.4 Other Applications of Turbidimetry

The dependence of turbidity on particle size can also be exploited in other ways than its conventional use in particle size determination. Kourti[15] suggested that simple

turbidity measurements at one wavelength can be used to monitor polymerization processes in order to detect the onset of several events, as for example the beginning of the reaction in latex systems (i.e. the end of induction period) or the onset of agglomeration. By following turbidity measurements at one wavelength during the emulsion polymerization of vinyl acetate in a continuous laboratory-scale reactor, Kiparissides et al.[9,10] could detect whether the reactor was operating under steady-state conditions or under oscillations. Turbidity measurements at 400 nm were used by Kourti[15] to follow qualitatively the effect of several variables on particle size during vinyl acetate emulsion polymerizations. Turbidity measurements at one or more wavelengths (and reduced to the same degree of dilution) can also be used for routine operations to monitor the progress of the reaction and detect any deviations from a known desirable path. Finally, Crawely et al.[74] discussed the use of turbidity measurements at several wavelengths in studies of crystallization, agglomeration and attrition.

6 PROCEDURAL DETAILS

6.1 Instrumentation and Forward Scattering

The turbidity of a suspension can be measured very easily in a laboratory or a plant environment. Very often spectrophotometers in the UV/VIS range are utilized for this purpose. A dilute suspension is inserted in the cell of the spectrophotometer. We have (Equation 46):

$$\tau_{\lambda_0} = 2.303\frac{A}{l} = \frac{1}{l}\ln\left(\frac{100}{T}\right. \tag{46}$$

where A is the absorbance reading, T is the percent transmittance at wavelength λ_0 and l is the optical path (i.e. width of cell). There are also reported applications where turbidity measurements at longer wavelengths, in NIR regime,[6,16,17,19,26] were used for particle size determination. Sometimes, certain modifications may be necessary[19] in the instruments used to measure turbidity so that forward scattering is rejected and theoretical calculations of turbidity can be legitimately compared with the measured values. This is because for the determination of particle size from turbidity measurements, it is assumed that the detector measures the attenuation of the beam of the source light. However, all the spectrophotometers measure the attenuated light from the source plus a certain amount of light scattered at small angles. Forward scattering will add error to the value of the true turbidity. Reports in the literature[34] mention that the larger the particles the greater is the possible discrepancy between the true and the measured (or apparent) turbidity, but for small particles and particles comparable to the wavelength the difference between the true and apparent turbidity should be minute. Bohren and Huffman[75] suggested that although conventional spectrophotometers are not usually designed to reject scattered light at small angles, such instruments may be adequate for scattering measurements if the particles are not too large. For monodisperse polystyrene latexes ($m = 1.2$) with diameters of 109, 312 and 1050 nm and turbidity measurements at wavelengths between 300 and 600 nm, they used a conventional spectrophotometer, suggesting that the forward scattering should not be significant. Walstra[76] suggested that sometimes errors attributed to forward scattering[18] were in fact due to extremely high optical densities that the investigators had included in their results.

The brief analysis that follows illustrates the consequences of the error added to the value of the true turbidity due to forward scattering. The effect that an error in the measured turbidity value has on the estimated particle size depends on the turbidimetric technique used and the (m, α) values of the system and it cannot be generalized. The turbidity ratio utilizes two turbidity measurements at different wavelengths, therefore the error due to forward scattering is introduced twice in this method, but only once in the specific turbidity. In Figure 5 we plot[15] as a function of α, for two different m values, the percentage error in the estimated diameter in monodisperse suspensions for a 3% error in the measured specific turbidity. Note that the effect changes with the (m, α) combination. For $m = 1.2$ and α values near the specific turbidity maximum (which for $m = 1.2$ occurs at $\alpha = 7.33$), the error becomes very large. However, for α

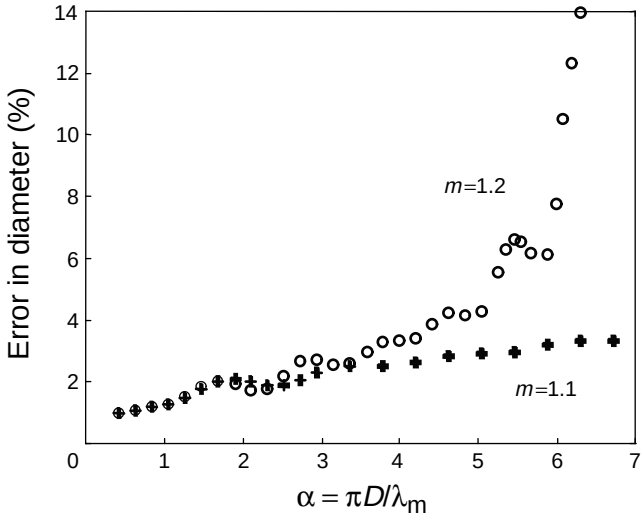

Figure 5 Percentage error in the estimated particle diameter of monodisperse suspensions plotted as a function of α for $m = 1.1$ and 1.2 when there is a 3% error in the specific turbidity measurements.

values less than 4.0 the error in the diameter is the same for $m = 1.1$ and 1.2, and it is approximately 4.0%. For water medium ($n_m \approx 1.33$), a value of $\alpha = 4$ corresponds to $D/\lambda_0 \approx 1$ ($\alpha = 4 = \pi D n_m/\lambda_0$), which means that for particles smaller and comparable to the wavelength the error is small. Therefore, if the error due to forward scattering introduces a 3–4% error in specific turbidity, then for particles comparable to λ_0 the effect on particle size should not be significant, and this would explain the observations reported by various workers. A final note on Figure 5: it demonstrates that a 3% error in the specific turbidity measurements may cause a very large error in the estimated size for α values near the peaks or the valleys of the K_{scat} vs α curve. This is due to the light scattering behavior and has nothing to do with the source of the error (i.e. this 3% could be due to either a wrong solids fraction measurement or to forward scattering, and would have the same effect on the estimated size).

Errors in turbidity measurements due to forward scattering are discussed in a number of references, where suggestions are given for the modification of the spectrophotometers.[6,18,20,30,49,76–78] Also, several empirical approaches have been suggested to account for the difference between true and apparent turbidity.[20,66] For example, a correction factor (R) can be estimated[20] for several suspensions with known particle diameters; R is equal to the ratio of the measured to the theoretically expected turbidity. This factor can then be used for unknown suspensions, but with (m, α) values in the same range for which it was estimated. In another approach, a calibration was used[66] to relate apparent specific turbidity with the true particle size for large latex particles where forward scattering is significant.

Users can assess the accuracy of the turbidity measurements of their spectrophotometers by comparing the experimental values with theoretically calculated values for monodisperse suspensions with known diameters, covering the range of (m, α) values that they plan to analyze. Before proceeding to any modifications one should judge whether the accuracy is satisfactory for the purpose of the study to be carried out using the turbidimetric results. For example, if the purpose of the study is to use the method as a reproducible and inexpensive way to monitor particle size during a reaction, then the accuracy of the spectrophotometer with no modifications may be satisfactory, and similarly if the objective is to detect the qualitative effect of various parameters (emulsifier, impurities) on the particle size of the latex. For other applications that require more accurate determinations of larger particles, one may consider modifications of the spectrophotometer and the use of a correction equation[76] that accounts for the angle of acceptance of the detector for the calculation of the scattering coefficient. Kourti[15] used two spectrophotometers without modifications for diameters up to 400 nm and suspensions with $m = 1.1$ and 1.2: a Hewlett-Packard Model 8450A at wavelengths of 350–700 nm and a bandwidth of 2 nm and a Bausch and Lomb (now Spectronic Instruments) Spectronic 20 at wavelengths of 380–580 nm and a bandwidth of 20 nm. The accuracy of the method was tested by estimating the size of monodisperse suspensions with known diameters and also by comparing the size estimated by turbidimetry with that measured by EM and DLS. For numerous unknown samples, the diameters estimated by turbidimetry were in very good agreement with those measured by DLS and EM. Forward scattering did not introduce any problems.

Gossen and MacGregor[24] used for on-line measurements in a pilot-plant environment a Bausch and Lomb single-beam Spectronic 20 equipped with a cylindrical Pyrex flow-through cell with a 1 cm pathlength, without any modifications for forward scattering. The incident beam had a bandwidth of 20 nm and the wavelength was 400 nm. Weight-average diameters up to 200 nm obtained during latex production were in very good agreement with weight-average diameters obtained by on-line DLS,[73] and a change in D_w of the order of 10 nm was detectable. The accuracy of the results was considered satisfactory for their objective to monitor changes in the particle size due to process disturbances.

6.2 Refractive Indices of the Particles and the Medium

Turbidity is a very strong function of the optical properties of the particle suspension, that is, the ratio of the refractive index of the particles to that of the medium, m, and the refractive index of the medium, n_m. The latter is used for the calculation of λ_m and consequently the calculation of α. For the correct determination of the particle size the values of these properties must be accurately known. The values of the refractive indices for different materials can be found in relevant handbooks.[79] Refractive indices can also be directly measured, or estimated using a combination of techniques as described below.

Llosent et al.[27] used a Bellingham Abbé 60 refractometer to measure the refractive index of latex particles, n_p, required to calculate m for their work. The refractive index of the polymer film was measured at several wavelengths. They then used the measured values to estimate the constants for a Cauchy[80] equation that expresses n_p as a function of the wavelength λ_{0i} (Equation 47):

$$n_p(\lambda_{0i}) = A + \frac{B}{\lambda_{0i}^2} + \frac{C}{\lambda_{0i}^4} \tag{47}$$

This equation was subsequently used to calculate n_p at other wavelengths.

If the value of the refractive index of the medium n_m is known but n_p and therefore m are not known, it may be possible to estimate m directly using a combination of

techniques. Kourti[15] suggested that, since turbidimetry is very sensitive to the value of m, it is probably the best method to ''estimate'' the values of m as a function of wavelength for a specific suspension. Suppose we wish to estimate the value of m for suspensions of polybutadiene particles in water. Using a monodisperse suspension of polybutadiene particles in water with known diameters (estimated by EM or DLS), the scattering coefficient and therefore m can be calculated from the specific turbidity measurements at one wavelength (provided n_m is known). By measuring specific turbidities at more wavelengths one can estimate the dependence of m on wavelength. The advantage of this approach is that the value of m for a particular suspension is measured in the environment in which it will be used. This may be more meaningful than, say, using a latex film to measure the refractive index of particles and dividing by the refractive index of medium. Once m has been measured it can be used for suspensions of the same material (i.e. polybutadiene in this example), but of course for any particle sizes or PSD. DLS is an excellent method for this purpose because the particle sizes can be measured in suspension and for monodisperse particles DLS is independent of the refractive index. Another method to determine optical constants would be angular light scattering.

Sometimes the medium may not be a clear substance but may have additives. In latex samples, for example, emulsifier may be added to the water medium, or monomer may be dissolved in it if the samples are withdrawn at low monomer conversions. In these cases the refractive index of clear water cannot be used. The refractive index of the medium for a given concentration of water-soluble monomer (or other substance present in it) can be read directly from a calibration curve of refractive index as a function of the concentration of monomer in water. Such calibration curves can be easily constructed from refractive index–concentration increments obtained from differential refractometers. In high dilutions the effect of monomer on the refractive index of the medium is usually negligible.

6.2.1 *Inhomogeneous Particles/Latex Particle Applications*

Sometimes the particles in the suspension may not be homogeneous. Examples from latex systems are carboxylated particles that swell in water, particles swollen with monomer in latex samples withdrawn at low conversions and particles produced by core and shell polymerizations. There are several issues related to determining the optical properties of such particles and research is active on light scattering by inhomogeneous particles.[81–83]

For monomer swollen particles, or particles produced by core and shell polymerizations, one approach is to calculate the value of m by treating the particles as coated spheres. For coated spheres the refractive index ratio can be calculated as $m = m_1 v_1 + m_2 v_2$, where m_1 is the ratio of the refractive index of the core to the medium and v_1 is the volume fraction of the core; m_2 and v_2 are the corresponding values for the coating. ($v_1 + v_2 = 1.0$). This approximation is not very accurate for thin coatings. In the above approach, for monomer swollen polymers, the volume fractions of the monomer and polymer are treated as core and coating respectively. It should be emphasized that this is an approximation, and the true morphology of monomer swollen particles may be different than a core and coating approximation. Bohren and Hoffman[75] also presented an algorithm to calculate the scattering functions of coated spheres (consisted of homogeneous core and homogeneous coating of uniform thickness) and they discuss the limitations of this algorithm, in general, for anisotropic spheres. For very accurate calculations, these approximations to estimating the value of m for core and shell particles must be checked experimentally. This can be done as mentioned earlier by using another method (i.e. DLS) to estimate the particle size and then using turbidimetry to estimate the value of m for particles with known polymer composition. The above calculations should be performed on monodisperse suspensions. A calibration curve could then be constructed to give the value of m as a function of the polymer composition in the particle. If DLS is used to verify the particle size of monomer swollen particles, the same degree of dilution must be used to make sure that the same particle size is measured (because monomer is dissolved in water and extracted from the particle with high dilutions).

For monomer swollen particles and carboxylated particles that swell in water, both the density and the refractive index depend on the degree of swelling and appropriate corrections should be made. In monomer swollen particles, the amount of monomer associated with the particles (and therefore the refractive index of the particles and the medium) depends on the conversion, the solubility of monomer and the degree of dilution of the original sample.

Latex produced by multicomponent polymerization is also a case where the calculation of the refractive index of the particles may present difficulties. The refractive index depends on the polymer composition and may change during the reaction due to composition drift. Therefore, calibration may be needed to relate the refractive index of the particles to their composition (by taking into account the conversion of the monomers at the time of sampling). When monitoring particle growth with on-line turbidimetry during multicomponent polymerizations, on-line gas chromatography can be utilized to provide

information on the composition of the particles. The refractive index can then be calculated from the particle composition utilizing some type of calibration. On-line densitometry combined with on-line gas chromatography can provide information on the particle volume fraction, ϕ. More details on this topic can be found in Kourti.[15]

6.3 The Choice of the Wavelengths

As discussed earlier, the successful determination of the PSD or of an average particle size of a suspension depends on the m and α values of the suspension and the turbidimetric technique used. The refractive index of the medium and therefore the value of m of a suspension can be altered by adding other substances to the medium [e.g. glycerol[5,16]] provided that these substances do not affect the particle size. The range of α values covered by the PSD can also be manipulated by changing the wavelengths at which the turbidity measurements are taken. As a general rule, wavelengths widely separated from each other should be used when the estimation of the full PSD is required. In this way the range of α values covered by the PSD changes significantly from one wavelength to another (the y exponent changes) and the probability of the turbidity being proportional to the same moment of the PSD at both wavelengths is reduced. It should be noted, however, that when wavelengths widely separated from each other are used, the value of m at each wavelength must be accurately known; using a constant m may result in false estimates.[15] If only one value for m, estimated at, say, λ_0', is available and must be kept constant for all the wavelengths, then wavelengths on either side of λ_0' (i.e. both shorter and longer) and as close to λ_0' as possible must be used.

To keep wavelengths widely separated, several workers chose one wavelength in the UV/VIS range and one in the IR region. The first (1.7 μm)[16,17] and the second (2.22 μm)[19] IR transmission windows for water have been the choices for long wavelengths. Most workers have avoided wavelengths below 400 nm where some particles may absorb light and wavelengths where the medium absorbs light.

Eliçabe and Garcia-Rubio[62] used wavelengths in regimes where particles absorb to resolve the PSD for a computer-simulated polystyrene latex example, where clear water was assumed to be the medium. The resolution of the PSD was improved. However, in practical situations for latex particle size analysis, it is preferable to avoid working at short wavelengths where absorption occurs (e.g. the styrene ring absorbs at 269.5 nm). In latex samples withdrawn at low monomer conversions, residual monomer which absorbs light at certain short wavelengths may be present in the particles and extracted owing to dilution in the medium. As a consequence, if particle size determination is attempted at these short wavelengths without accounting for the distribution of the absorbing monomer between the medium and the particles, the results may be questionable. However, calculating the monomer distribution is tedious and requires a knowledge of conversion, partition coefficient of the monomer between the water and the polymer phase and the degree of dilution. It is clear that using short wavelengths is not practical if the objective is to follow particle growth during the reaction.

6.4 Preparation of Samples

In particle size determination by light scattering methods, it is usually assumed that the scattering by an array of particles is incoherent (the particles are randomly positioned in space to avoid mutual polarization) so that the scattering functions corresponding to an isolated particle may be used, and also that there is no significant multiple scattering. These conditions are practically met[33] when turbidity measurements are taken with sufficiently dilute suspensions. In practice this is achieved by diluting the suspension and working at a concentration range where the turbidity, at the shortest wavelength to be used, varies linearly with the particle concentration. Figure 6 shows how turbidity changes with concentration. The shorter the wavelength, the lower is the concentration of suspended solids required to avoid multiple scattering.

The fact that the sample needs to be diluted may have an effect on the particle size in certain cases. For example, in latex samples withdrawn at low conversion, polymer particles swollen with monomer may change size by changing the degree of dilution, and this should be taken into account. Also, for latex particles stabilized

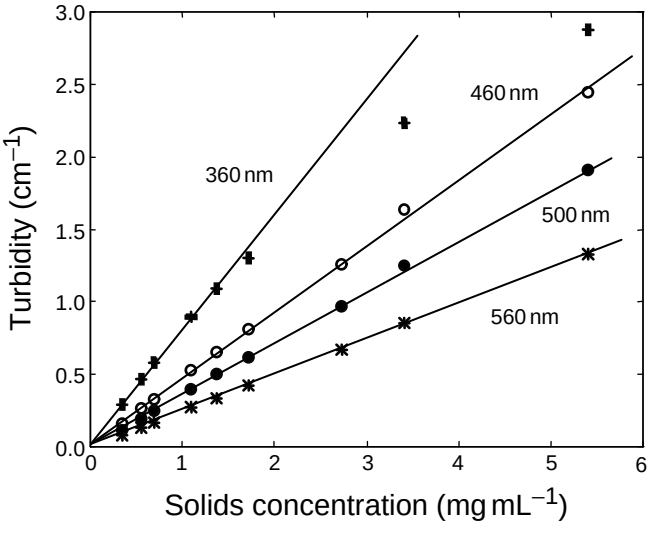

Figure 6 Turbidity plotted against solids concentration in the sample: linearity check.

by ionic or nonionic emulsifiers or other groups, the user must be careful to dilute the sample without risking destabilization. Some emulsifier (the same as the one used in the latex production) must be added to the diluting medium to prevent destabilization. Samples prepared in this way remained stable for a long time;[15] turbidity measurements taken 3 years after the samples were prepared showed no difference within experimental error with measurements taken for those samples when they were freshly prepared. Kourti[15] discussed in detail other considerations for latex particle size determination, for both off-line and on-line applications.

6.5 Particle Volume Fraction for Specific Turbidity

Particle volume fraction is measured by gravimetric analysis. The weight of the dry solids per cubic centimeter (c, g cm^{-3}) is measured and converted to volume fraction using the density of the particles (ρ, g cm^{-3}); $\phi = c/\rho$. Particle volume fraction is usually measured on undiluted samples. The particle volume fraction corresponding to the diluted sample on which the turbidity measurement is taken is calculated by taking into account the degree of dilution. Therefore, the degree of dilution must be known very accurately. Particle volume fraction for on-line applications can be measured by densitometry.[21,23,24]

7 EFFECT OF ERRORS ON THE ESTIMATED PARTICLE SIZE AND PARTICLE SIZE DISTRIBUTION

The determination of the particle size or PSD of a suspension using turbidimetry involves measurements and assumptions. Errors in the measurements and uncertainty in the assumptions will have a minor or a major effect on the estimated size. The sources of errors can be grouped as follows:

- Experimental errors in measurements: these involve random measurement errors and errors due to forward scattering (bias) in the turbidity measurements. For specific turbidity methods there are also errors in the calculation of the solids volume fraction due to dilution, error in gravimetric analysis and uncertainty in the value of density of the particles.
- Uncertainty in the m values and the dependence of the m on the wavelength: m is required for the theoretical calculation of the scattering functions.
- For polydisperse samples, error in the assumptions related to the PSD: either the form of the PSD has to be assumed or the number of discrete points, their range and the value of the smoothing parameters.

The sensitivity of the estimated particle size to such errors has been investigated by several workers. The results are briefly summarized in the following sections.

7.1 Error in Measurements

The effect of measurement errors on the estimated particle size depends on the technique used (specific turbidity or turbidity ratio) and on the (m, α) values of the system and cannot be generalized, as illustrated in the following. Kourti[15] reported an average error of 0.4% in the repeatability of transmittance readings. Such an error is translated to an average 2% error in specific turbidity. For the (m, α) values used, with $m = 1.1$ and $\alpha < 3.5$ this error would translate to a maximum error of 2% in the particle diameter for monodisperse suspensions. If a different technique, say turbidity ratio, was used for the same (m, α) regime, then from a small bias of +0.5% on the transmittance readings at both wavelengths (the same bias would be the case due to forward scattering) the error in the particle size for monodisperse particles can be 2–3%. For the same size regime, if instead of the bias the error was random ($\pm 0.5\%$), then the error in the particle size can be as large as 9%. Because the effects of turbidity measurements on particle size cannot be generalized over (m, α) values, a good practice for users would be to construct a curve such as that in Figure 5 for the m value of the suspensions with which they plan to work. The effect of a certain percent error in specific turbidity (or turbidity ratio) on the estimated particle diameter can be calculated with simulations for a wide α range. Then, by changing the value of α used (by changing the wavelength), they can move the system to a regime where measurement errors would not affect the particle size significantly.

The effect of measurement errors on the estimated PSD in polydisperse systems was the subject of a detailed study[15,32] and it has been shown to depend on the technique used (specific turbidity or turbidity ratio) and on the (m, α) values of the suspension. As discussed in section 4, the estimation of full PSD is impossible within experimental error in some (m, α) regimes. However, in these regimes, if specific turbidity is used, an average diameter can be correctly estimated and the effect of measurement error on this average diameter is not large. In regimes where only the weight-average diameter can be estimated, for a random error of $\pm 3\%$ in specific turbidity at three wavelengths, the maximum error was for a weight average 5%. As a comparison the maximum error for the number average in these regimes was 40%.[15]

7.2 Error in the Value of the Refractive Index Ratio, m

Kourti[15] studied the effect that errors in the value of m have on the estimated PSD. It was concluded that

turbidimetry is sensitive to such errors but the magnitude of the effect on the estimated PSD depends on the (m, α) values of the suspension. In certain (m, α) regimes, the effect that the error in m has on the estimated parameters of the PSD is comparable to the effect of turbidity measurement errors, whereas in others it is much higher. In regimes where the y exponent does not change significantly with the particle size, the errors in the value of m had a larger effect on the estimated size than turbidity measurement errors. Therefore, observations reported in the literature on the magnitude of this effect cannot be generalized for any (m, α) values. For example, the observation of Wallach and Heller[7] that "the error of the disregard of the dispersion of m is less that experimental error" was simply a coincidence of the (m, α) values of the system with which they were working. Whenever there is uncertainty in the value of m of the suspension to be analyzed, a sensitivity analysis should accompany the results, to indicate the degree of confidence. Alternatively, the results should be corroborated with another method.

Sometimes, the dependence of the refractive index of the medium, n_m on the wavelength is known, but only one measurement for the refractive index of the particles, n_p, at one wavelength λ_0' is available. If this is the case, it is recommended that a constant m (calculated at λ_0') is used at all the wavelengths, and not a variable m calculated with constant n_p and wavelength-dependent n_m. For water medium, a wavelength-dependent n_m decreases with longer wavelengths and using constant n_p will force m to increase with longer wavelengths (when it is expected to decrease); this in turn means significant errors when analyzing real data since the estimated particle size moves in the opposite direction to the true value. Finally, when a constant value of m is used, the error is smaller if the turbidity measurements are taken at wavelengths both shorter and longer and as close as possible to the one at which m was measured.

In all cases, however, if the wavelength-dependent n_m is known, it should be used to calculate the value of λ_m and therefore α for the scattering coefficient calculation.

7.3 Error in the Assumed Distributional Form

The main criticism in the literature about approaches that attempt to recover the PSD by assuming a type of the distribution (i.e. log-normal, Gaussian, etc.) is that the assumed distribution may be very different from the true one. When a distributional form is assumed, a constraint (i.e. the shape of the distribution) is forced upon the solution. The expected result is a PSD with the assumed shape that will envelope the true distribution; the details of the true PSD are not expected to be resolved. The estimation of the location of the main body of the distribution is therefore the most satisfactory result expected in this case.

The error introduced in the estimated PSD when the assumed distributional form is different from the true one has been studied either experimentally or with simulated cases by several workers. Wallach and Heller[7] tried to estimate the PSD of three suspensions with positively skewed, negatively skewed and symmetrical distributions, using the turbidity ratio method and assuming a distributional form which can describe only positively skewed distributions. In the cases of negatively skewed and symmetrical PSDs, the estimates of the parameters chosen to describe the distribution had a significant error; however, Kourti[15] reported that a careful examination of those results showed that the estimated PSD enveloped (i.e. recovered correctly) the main body of the true distribution. This behavior was expected since the polydispersities of the three synthesized distributions and of the assumed distribution were very small; no distributions with very long tails to small or large particles were studied.

In an extensive simulation study, Kourti[15] investigated the error introduced in the estimated PSD when the shape is assumed to be log-normal. Specific turbidities were calculated (at 400, 500 and 600 nm) for several polystyrene suspensions of submicrometer particles with several types of continuous distributions: positively skewed, negatively skewed, symmetrical, bimodals with broad peaks and overlapping tails. The polydispersities were from 1.04 to 1.9. Then, from these calculated specific turbidities and assuming a log-normal form, the PSD of the suspension was re-estimated. The following conclusions were drawn that are valid for the estimation of the PSD (using the specific turbidity method) of suspensions of submicrometer particles, excluding the Rayleigh regime where only D_τ can be estimated and the (m, α) regimes where specific turbidity can only give D_w:

- When the true distribution is continuous with no long tails to either small or large particles, the estimated distribution with a log-normal assumption will cover the main body of the true PSD when plotted on a number basis.
- On a weight basis the main body of the true distribution will be covered by the estimated log-normal distribution, in all cases, even when the true PSD has long tails.
- The weight-average diameter of the distribution estimated with a log-normal assumption is very close to the true weight average of the suspension, even in cases that the true distribution is significantly different from a log-normal shape. For the cases studied the weight-average diameter of the PSD estimated when assuming a log-normal form (on a

number or weight basis) was, in general, within 1.0% of the weight average of the true PSD.

These conclusions are corroborated by experimental results reported by Haseler[19] for submicrometer polystyrene suspensions. They also agree with the observations of Yang and Hogg.[20] These authors also studied experimentally the error on the estimated PSD when the shape of the true distribution is different from the assumed log-normal form. The estimated distribution was compared with the true one on a weight basis; the distributions studied were broad, covering particle sizes from 0.1 to 10.0 μm. They found that "the estimated distribution conformed closely to the central portion of the true distribution (on a weight basis) and was not strongly biased towards either of the tails".

It should be noted that when using Equation (31) to estimate f_j for diameter $D_j (j = 1, 2, \ldots, P)$, instead of assuming a form for the PSD, there are also assumptions involved. These include the choice of the parameter in the regularization algorithm and the number of points to which the PSD will be discretized and the location of the D_js. The number of points depends on the available number of turbidity measurements at different wavelengths. For a study on the effect that the number of discrete points at which f_j is estimated has on the estimated PSD, readers are referred to Eliçabe and Garcia-Rubio.[62] In that study the true distributions were all simulated continuous distributions.

There is also the question of whether there is an advantage in proceeding with collecting more wavelengths and trying to estimate the distribution using Equation (31). In the case of continuous unimodal distributions, the resolutions of the two techniques under experimental error seem to be comparable; an assumption of a log-normal shape that will provide the main body of the true PSD and the true D_w may be sufficient for most routine applications in the submicrometer and lower micrometer region.

7.3.1 The Resolution of Bimodal Distributions

Regularization methods have the potential to recover the PSD without assuming any shape, if the true distribution is a bimodal one with two broad peaks, such as that shown by the histogram in Figure 7. Eliçabe and Garcia-Rubio[62] resolved such a simulated distribution where the particle diameters covered a very wide range from 50 to 2550 nm; in that study wavelengths in the absorption range were included for the turbidity measurements. For such bimodal distributions, the regularization methods have the capability (provided that there are enough turbidity measurements) to uncover the true shape whereas the log-normal assumption will always produce an envelope around the main body, as shown in Figure 7. Distribution I

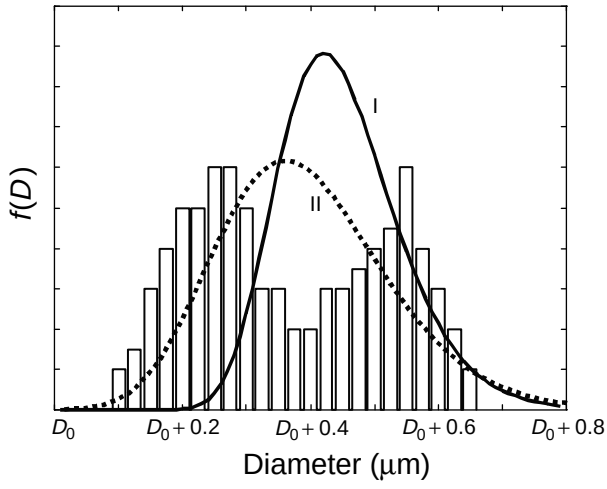

Figure 7 Bimodal distribution as resolved by a log-normal assumption. Case I: the estimated log-normal distribution (continuous line) and the true distribution (histogram) are located between 0 and 0.8 μm ($D_0 = 0.0$). Case II: the estimated log-normal distribution (dashed line) and the true distribution (histogram) are located between 0.4 and 1.2 μm ($D_0 = 0.4$).

in Figure 7 shows the distribution (on number basis) estimated with a log-normal assumption for polystyrene suspensions ($m \approx 1.2$) where the true distribution (shown by the histogram) was in the size range between 0.0 and 0.8 μm ($D_0 = 0.0$). Distribution II is the log-normal PSD estimated when the size range for the true distribution was between 4.0 and 1.2 μm ($D_0 = 0.4$). Note that although the two histogram distributions have the same breadth (0.8 μm), in case II the distribution has a smaller polydispersity (since the average size is larger). Note that in case I the estimated log-normal distribution tends to envelope only the larger particles of the true distribution, whereas in case II it uncovers the correct range but of course cannot resolve the bimodal shape.

When dealing with bimodal distributions one should keep in mind that the contribution of small particles to the value of the specific turbidity could be totally obscured by the contribution of the large ones, depending on their α values (relative to the α value location where the peak of K_{scat} occurs). As an example, assume a polystyrene suspension ($m \approx 1.2$) with a bimodal distribution which consists of two monodisperse populations with $D_1 = 0.6$ μm and $D_2 = 0.1$ μm, and where the number of small particles is 10 times the number of the large ones ($N_2/N_1 = 10$). The contribution of the small particles to the total specific turbidity at 400 nm is only 0.46%. Hence the presence of small particles will cause less disturbance than the experimental error even though there are 10 times more small than large particles. In such a case it is not expected that the regularization methods will produce the correct solution within experimental error.

Wallach and Heller[7] suggested using combinations of two distributions with the same form to describe distributions that are negatively skewed, or bimodals. A combination of two-parameter distributions (say two log-normal distributions) requires the estimation of five parameters ($D_{1g}, D_{2g}, \sigma_1, \sigma_2, N_2/N_1$). At least five specific turbidity measurements at wavelengths far from each other are required for this estimation if the particles are in the submicrometer range. These parameters are expected to be very highly correlated and therefore the solution to be sensitive to experimental error. A good solution may be obtained in cases where the contribution of the distribution of the smaller particles to the overall specific turbidity is large; otherwise the large particles will always obscure the small particles.

Readers should keep in mind that in general, bimodal distributions (except those where the distance between the modal diameters of the two populations is large compared with the breadth of the two populations) are very difficult to resolve with most particle sizing methods. Using EM, for example, a large number of particles must be counted, sometimes at different magnifications in order to resolve the distributions successfully. DLS will usually result in distributions weighted towards larger particles. For bimodals where the distance of the two modal diameters is large compared with the breadth of the distributions, one can usually determine very successfully the particle sizes involved in the two populations (by EM and sometimes by DLS), but it is difficult to estimate the relative number of particles in the two populations. In such cases, if the diameters involved are known very accurately from EM ($D_1, D_2, \ldots, D_n$), then using a sufficient number of turbidity measurements one could estimate the relative number of particles ($f_1, f_2, \ldots, f_n$) (from Equations 31 and 39). Again, in order to obtain reasonable estimates, the small particles should weigh significantly and the distribution must be in an (m, α) regime where specific turbidity measurements do not simply result in the same apparent diameters (otherwise all combinations resulting in the same average are possible solutions).

8 ON-LINE PARTICLE SIZE DETERMINATION USING TURBIDIMETRY

As early as 1973, Hamielec and Wright[84] suggested using turbidimetry for on-line latex particle size determination. The configuration of a sampling device for on-line measurements of the PSD of poly(vinyl acetate) latexes using turbidimetry was suggested from that laboratory in 1976.[85] On-line turbidity measurements in fully automated systems have been reported for poly(vinyl acetate) latexes by Kiparissides et al.[9,10] and Zollars[11] but no quantitative results from on-line particle size determination were reported by these workers. Haseler and Parkin[6] used on-line turbidity measurements to monitor the growth of silver bromide in a ripening process. The ripening suspension was circulated continuously in an undiluted state through a thin cell.

In all of the applications mentioned above only turbidity was measured on-line. An experimental set-up to combine on-line determination of the particle volume fraction with on-line turbidity measurements, required for on-line specific turbidity measurements, was first suggested by Pollock.[86] However, no results were reported on particle sizing, probably because he faced some problems with that set-up. A simulation for on-line latex PSD determination in continuous latex reactors was presented by Bradolin and Garcia-Rubio.[22]

Raphael and Rohani[26] described a device that could be used on-line to measure the mean particle size in continuous precipitators or crystallizers. The solids concentration for this set-up is assumed to be known and no measurements on the particle volume fraction are taken on-line. This device was not tested on-line, however, for continuous production; the measurements reported were from spot samples. These samples were diluted in a vessel from which the diluted sample is pumped through the device and turbidity measurements are taken.

A serious disadvantage of some of the set-ups suggested for on-line latex particle size determination[9,10,11,22,86] is that they require a continuous flow of diluted latex through the spectrophotometer. For specific turbidity measurements this means that a known amount of latex is continuously mixed with a known amount of diluent. Both the latex and diluent have to be pumped at constant flow rates. This in turn requires that a very small amount of latex be pumped at the exit of the reactor, at flow rates as low as $10\,\mathrm{mL\,min^{-1}}$, for long periods of time. A peristaltic pump tried by Gossen[21] for this set-up failed owing to coagulation of the latex due to high shear and chemical attack of the tube by the monomer. Other types of pumps tried in the same laboratory also failed.[23] Another problem with continuous dilution of the latex is that an enormous amount of waste (very dilute latex) is created that requires treatment before disposal. This would cause problems in large-scale operations. The set-up suggested earlier from the same laboratory[85] utilizes discrete sampling, but it has serious disadvantages.[23]

A system suggested by Gossen[23,24] measures both solids fraction and turbidity on-line, utilizes discrete sampling, is mechanically simple, has a short cycle time and high accuracy and it is set up such that thorough cleaning of the sampling element is performed after each sample. This sampling device was used for the on-line

application of specific turbidity to monitor particle growth during poly(vinyl acetate) production in continuous pilot-plant reactors. Samples were taken at discrete intervals of 5 min. Results were reported for a period of 10 h. The weight-average diameters estimated by turbidimetry showed excellent agreement with results from on-line DLS.[73] These results, obtained by Gossen, represent the first published successful effort[24,46,73] to combine on-line turbidity and density measurements and to monitor, on-line, weight-average diameters during latex production in pilot-scale reactors, in a fully automated manner. Details on the set-up and an error propagation analysis study can be found in Gossen.[23,24]

9 OTHER CONSIDERATIONS

9.1 Choice of the Methods

It is a well-established perception in particle size analysis that there is no method which can be successfully utilized for every application. The effectiveness of each particle sizing method depends on the special needs applicable to each case. The selection of the method is a compromise between accuracy, cost, speed of analysis and robustness of the instrument in the plant or the laboratory environment. This section gives some examples that might help users decide if turbidimetry is suitable for their applications.

There are cases where an accurate estimate of the full PSD is required. In such cases the time required for analysis and the cost of the method are not usually an issue. Examples of such cases are the evaluation of a new process and the PSD of the new product, the study of the effects of various parameters on particle size in a set of designed experiments and the detailed study of a nucleation process where the PSD reflects the mechanisms that take place. For such cases more time and effort are usually spent than for routine particle size analysis and more than one particle sizing methods and/or alternative approaches to the same method are used. EM will provide measurements of the sizes involved in the distribution, and will give an indication of the type of the distribution (i.e. detect the presence of two populations, etc.). However, resolving the full $f(D)$ with EM is a tedious task because a very large number of particles need to be photographed and counted from representative parts of the sample. Specific turbidity can be used to obtain a correct estimate of the weight-average diameter of submicrometer particles if the (m, α) regimes do not allow for the full PSD determination. If the (m, α) values allow for the full PSD determination, then regularization methods, where no fixed shape of the distribution is assumed, should be tried first. If these methods show that the shape is smooth and the log-normal assumption is sufficient, then this assumption can be used for subsequent routine analysis. DLS could also be used to corroborate the results for submicrometer particles. Alternatively, sedimentation field fractionation[87–89] could be used; this method would be especially useful in cases where EM cannot be used either because the particles change size when they are measured dry or they need special treatment for EM measurements.

Sometimes only routine-type checks are needed and an average of the PSD or the log-normal assumption that will provide the main body of the distribution may be sufficient. Turbidimetry could be used as a fast, inexpensive technique for such applications.

In some cases only qualitative information is required, for example when studying qualitatively the effect of certain process parameters on the particle size. In such cases almost any technique can be utilized provided that the same quantity is followed and that this quantity is sensitive to small changes in the particle size. For example, specific turbidity can be used to follow a weight-average diameter or DLS to follow the intensity-average diameter. It is important that the same method be used to follow the same average throughout the study because, as discussed in the next section, different methods are theoretically directly related (and therefore are more sensitive) to different averages of the PSD. Other applications where the dependence of turbidity on particle size can be qualitatively exploited is to use turbidimetry as a detection system (as discussed in section 5.4) to detect the onset of agglomeration or reactor oscillations in latex production.

When the product is suspected to contain two populations of particles with different sizes (e.g. large and small particles due to a secondary nucleation), the small particle population will be clearly detected by EM or chromatographic (i.e. size separation) techniques. Size separation techniques are very powerful in detecting small-particle populations. If the contribution of the small particles to the overall turbidity is significant then the bimodal distribution can be resolved by turbidimetry using regularization algorithms (i.e. solving for PSD without assuming a fixed shape). Otherwise, specific turbidity will only provide the weight-average diameter for a bimodal suspension of submicrometer particles.

When choosing a method for on-line measurements, one must consider several factors: the type of process (batch or continuous), the purpose of the measurement (is it routine analysis where we are satisfied with an average size, or do we need accurate PSD determination), the required sampling frequency and analysis time of the method, the cost relative to the expected accuracy and the reliability and robustness of the method in the

plant environment. Turbidimetric methods utilize the fastest measurements (light is used as a probe). As a comparison, fractionation (separation) techniques cannot provide any estimate of the particle size or the PSD until the fractionation is completed.

It is clear that users must consider several factors before choosing the technique or a combination of techniques that are most suitable for their needs. This requires a good knowledge of the process, the product and the capabilities and limitations of the technique. Finally, quoting Groves,[90] "the validity of the data need have no direct relation to the cost of instrumentation".

9.2 Corroboration of Results

When another particle sizing method is used to corroborate results from turbidimetry, this second method must be chosen carefully. In general, when comparing results on the full PSD from different methods, the size range to which each of the methods is applicable, their resolution, the effects of errors on the estimated size and any assumptions introduced in each technique should be considered. This holds even when comparing results from two different turbidimetric techniques (i.e. turbidity ratio and specific turbidity). Erroneous conclusions may be obtained, as illustrated with the following examples.

The size range to which each of methods is applicable should be considered. Wales[58] used particle size measurements from a Coulter Counter instrument to corroborate results from turbidimetry. Some discrepancies were observed between the two methods. From a detailed discussion[15] of his work it became clear that in particle size regions where both Coulter Counter and turbidimetry are capable of providing a valid estimate of the PSD the two methods compared favorably. The discrepancies observed for suspensions with PSDs in the lower submicrometer range were not due to forward scattering as originally claimed.[58] Rather, they were due to the fact that the Coulter Counter instrument could not detect very small particles (in the lower submicrometer region) and specific turbidity for the (m, α) values of those suspensions could only provide a correct D_w and not the full PSD. Discrepancies between turbidimetry and Coulter Counter results were also observed by Walstra.[16] The Coulter Counter instrument should not be used to corroborate results from other techniques when dealing with distributions of small particle sizes (i.e. lower submicrometer particles) or PSD with tails of small particles.

Any assumptions involved in the methods should be considered. Some average diameters obtained by various techniques can be directly derived from the underlying theory whereas others depend on the validity of certain assumptions. Only quantities estimated directly from theory should be compared. For example, DLS measures an "average diameter" that corresponds to the mean diffusion coefficient of the suspension. With the assumption that the distribution of the diffusion coefficients is Gaussian, a PSD and its corresponding weight-average diameter can be calculated. However, although the estimated mean diffusion coefficient is correct, this weight-average diameter is not always equal to the true weight average of the suspension. In contrast, the weight average measured by turbidimetry is numerically close to the true weight average for most submicrometer suspensions. In another example, some DLS instruments utilize the Rayleigh–Gans approximation to convert distributions from an intensity basis to a weight basis. This approximation is not valid for very large particles and may cause serious errors in the calculation of the PSD on a weight basis for broad distributions or widely spaced bimodal distributions, even if the distribution estimated on an intensity basis is correct.

10 CONCLUSION

Turbidimetric particle sizing methods are experimentally simple, robust and have very good reproducibility and potential for on-line applications. They have been applied for the estimation of the PSD or an average size to a variety of particle suspensions (Table 1) with diameters ranging from 0.035 to 50 μm.

Specific turbidity can be used for reliable estimates of the average particle size and of the PSD of a suspension for a wider range of particle sizes than the turbidity ratio or the wavelength exponent techniques. This is because specific turbidity involves two independent measurements of the particle size via two different properties of the suspension (turbidity and solids volume fraction), whereas the other turbidimetric techniques are based on turbidity measurements only. Therefore, if the volume fraction of particles in the suspension can be measured, specific turbidity should be the turbidimetric technique of choice. The turbidity ratio or the wavelength exponent method should be used only if the particle volume fraction is unknown.

Because of the light scattering behavior, turbidimetric methods are not theoretically expected to provide information on the full PSD in many situations. In order to decide the type of information about the particle size (average size or full PSD) that can be extracted from any turbidimetric method, one must consider both the value of the refractive index ratio (m) of the suspension and the α values covered by the distribution. The same PSD that can be successfully estimated for one system may be impossible to estimate for another system with a different m value.

There are a few critical combinations of (m, α) values in turbidimetric analysis for nonabsorbing particles:

- suspensions of submicron particles and values of m smaller than 1.15;
- suspensions with any value of m and very small particles ($\alpha < 0.5$);
- suspensions of very large particles ($\alpha > 150$).

In these regimes, the turbidity ratio and the wavelength exponent method should be avoided and only specific turbidity should be used. If the particle volume fraction is not known, then for suspensions in the above regimes it is better to choose a sizing method other than turbidimetry. Furthermore, reliable information on the full PSD may be very difficult to obtain in the above regimes, by any turbidimetric technique, within experimental error. Specific turbidity, however, will give a very good estimate of meaningful averages of the PSD: the weight-average diameter in the first case (from the mean value of the apparent diameters at 400 and 600 nm), the turbidity-average diameter for suspensions of very small particles and the surface–volume-average diameter for suspensions of very large particles.

ABBREVIATIONS AND ACRONYMS

DLS	Dynamic Light Scattering
EM	Electron Microscopy
HDC	Hydrodynamic Chromatography
IR	Infrared
NIR	Near-Infrared
PLS	Partial Least Squares
PSD	Particle Size Distribution
SEM	Scanning Electron Microscopy
TEM	Transmission Electron Microscopy
UV/VIS	Ultraviolet/Visible
ZOLD	Zeroth-order Logarithmic Distribution

RELATED ARTICLES

Food **(Volume 5)**
Particle Size Analysis in Food

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Filtration in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Sieving in Particle Size Analysis • Surface Area and Pore Size Distributions • Ultrasonic Measurements in Particle Size Analysis • Velocimetry in Particle Size Analysis

REFERENCES

1. E. Atherton, R.H. Peters, 'Some Aspects of Light Scattering from Polydisperse Systems of Spherical Particles', *Br. J. Appl. Phys.*, **4**, 344–349, 366–369 (1953).
2. W. Heller, W.J. Pangonis, 'Theoretical Investigations on the Light Scattering of Colloidal Spheres. I. The Specific Turbidity', *J. Chem. Phys.*, **26**, 498–506 (1957).
3. W. Heller, 'Theoretical Investigations on the Light Scattering of Colloidal Spheres. III. Analytical Expressions for the Turbidity Approximating the Performance of the Mie Equations Prior to the First Maximum', *J. Chem. Phys.*, **26**, 1258–1264 (1957).
4. W.H. Beattie, 'Determination of Particle Size in Silver Bromide Sols by Light Scattering', PhD Thesis, University of Minnesota, 1958.
5. E.J. Meehan, W.H. Beattie, 'Determination of Particle Size in Silver Bromide Sols by Light Scattering', *J. Phys. Chem.*, **64**, 1006–1016 (1960).
6. S.C. Haseler, E.M.L. Parkin, 'Silver Halide Particle Size Measurement by Infrared Thin Cell Turbidimetry', in *Particle Size Analysis 1981*, eds. N.G. Stanley Wood, T. Allen, Wiley Heyden, Chichester, 403–413, 1982.
7. M.L. Wallach, W. Heller, 'Experimental Investigations on the Light Scattering of Colloidal Spheres. VI. Determination of Size Distribution Curves by Means of Turbidity Spectra', *J. Phys. Chem.*, **68**, 924–930 (1964).
8. Es. Gulari, G. Bazzi, Er. Gulari, A. Annapragada, 'Latex Particle Size Distributions from Multiwavelength Turbidity Spectra', *Part. Charact.*, **4**, 96–100 (1987).
9. C. Kiparissides, 'Modeling and Experimental Studies of a Continuous Emulsion Polymerization Reactor', PhD Thesis, McMaster University, Hamilton, Ontario, 1978.
10. C. Kiparissides, J.F. MacGregor, S. Singh, A.E. Hamielec, 'Continuous Emulsion Polymerization of Vinyl Acetate, Part III. Detection of Reactor Performance by Turbidity Spectra and Liquid Exclusion Chromatography', *CJChE*, **58**, 65–71 (1980).
11. R.L. Zollars, 'Turbidimetric Method for On-line Determination of Latex Particle Number and Particle Size Distribution', *J. Colloid Interface Sci.*, **74**, 163–172 (1980).
12. R.L. Zollars, 'Turbidimetric Method for Determination of Latex Particle Size Distribution', *J. Dispersion Sci. Technol.*, **2**(2–3), 331–344 (1981).
13. T. Kourti, A. Penlidis, A.E. Hamielec, J.F. MacGregor, 'Measurement of Particle Size Distribution of Latex Particles in the Submicron Range Using Size Exclusion Chromatography and Turbidity Spectra', in *Am. Chem. Soc. PMSE Div. Prepr.*, **53**, 147 (1985).
14. T. Kourti, A. Penlidis, J.F. MacGregor, A.E. Hamielec, 'Measuring Particle Size Distribution of Latex Particles

in the Submicrometer Range Using Size Exclusion Chromatography and Turbidity Spectra', *Am. Chem. Soc. Symp. Ser.*, **332**, 242–255 (1987).
15. T. Kourti, 'Polymer Latexes: Production by Homogeneous Nucleation and Methods for Particle Size Determination', PhD Thesis, McMaster University, Hamilton, Ontario, 1989.
16. P. Walstra, 'Estimating Globule-size Distribution of Oil in Water Emulsions by Spectroturbidimetry', *J. Colloid Interface Sci.*, **27**, 493–500 (1968).
17. A.T. Shaw, D.F. Darling, 'Fat Droplet Size Analysis: the Application of Spectroturbidimetry', in *Particle Size Analysis 1981*, eds. N.G. Stanley Wood, T. Allen, Wiley Heyden, Chichester, 355, 1982.
18. W. Heller, R.M. Tabibian, 'Experimental Investigation of the Light Scattering of Colloidal Spheres. II. Sources of Error in Turbidity Measurements', *J. Colloid Sci.*, **12**, 25–39 (1957).
19. S.C. Haseler, 'Particle Size Distribution Measurement by Automated Dual Wavelength Turbidimetry', *Part. Charact.*, **2**, 14–19 (1985).
20. K.C. Yang, R. Hogg, 'Estimation of the Particle Size Distributions from Turbidimetric Measurements', *Anal. Chem.*, **51**, 758 (1979).
21. P.D. Gossen, 'On-line Particle Size Measurement and Mathematical Modeling of Continuous Poly(vinyl acetate) Emulsion Polymerization', Master Thesis, McMaster University, Hamilton, Ontario, 1988.
22. A. Bradolin, L.H. Garcia-Rubio, 'On-line Particle Size Distribution Measurements for Latex Reactors', *Am. Chem. Soc. Symp. Ser.*, **472**, 64–85 (1991).
23. P.D. Gossen, 'On-line Instrumentation for Latex Production', PhD Thesis, McMaster University, Hamilton, Ontario, 1991.
24. P.D. Gossen, J.F. MacGregor, 'On-line Particle Diameter for Poly-vinyl Acetate Latex Using Specific Turbidity Method', *J. Colloid Interface Sci.*, **160**, 24–38 (1993).
25. V.-M. Taavitsainen, J.-P. Jalava, 'Soft and Harder Multivariate Modelling in Developing the Properties of Titanium Dioxide Pigments', *Chemom. Intell. Lab. Syst.*, **29**, 307–319 (1995).
26. N. Raphael, S. Rohani, 'On-line Estimation of the Solids Concentrations and Mean Particle Size Using a Turbidimetry Method', *Powder Technol.*, **89**, 157–163 (1996).
27. M.A. Llosent, L.M. Gugliotta, G.R. Meira, 'Particle Size Distribution of SBR and NBR Latexes by UV/VIS Turbidimetry Near the Rayleigh Region', *Rubber Chem. Technol.*, **69**(4), 696–712 (1996).
28. S.H. Maron, P.E. Pierce, I.N. Ulevitch, 'Determination of Latex Particle Size by Light Scattering', *J. Colloid Sci.*, **18**, 470–482 (1963).
29. R.J. Gledhill, 'Particle Size Distribution Determination by Turbidimetry', *J. Phys. Chem.*, **66**, 458–463 (1962).
30. J.D.S. Goulden, 'Spectroturbidimetry of Emulsions', *Br. J. Appl. Phys.*, **12**, 456–460 (1961).
31. L.D. Maxim, A. Klein, M.E. Meyer, C.H. Kuist, 'Particle Size Determination by Turbidimetry', *J. Polym. Sci. Part C*, **27**, 195–205 (1969).
32. T. Kourti, J.F. MacGregor, A.E. Hamielec, 'Turbidimetric Techniques: Capability to Provide the Full Particle Size Distribution', *Am. Chem. Soc. Symp. Ser.*, **472**, 2–19 (1991).
33. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, Academic Press, New York, 1969.
34. H.C. Van de Hulst, *Light Scattering by Small Particles*, John Wiley & Sons, New York, 1957.
35. W.J. Wiscobe, 'Mie Scattering Calculations. Advances in Technique and Fast Vector Speed Computer Codes', National Center for Atmospheric Research, Boulder, CO, NCAR Technical Note Tn-140_str, 1979.
36. W. Wiscobe, 'Improved Mie Scattering Algorithms', *Appl. Opt.*, 1505–1509 (1980).
37. P. Walstra, 'Approximation Formulae for the Light Scattering Coefficient of Dielectric Spheres', *Br. J. Appl. Phys.*, **15**, 1545–1552 (1964).
38. 'Conference on Light Scattering by Nonspherical Particles: Theory, Measurements and Applications', Goddard Institute for Space Studies, New York, 29 September–1 October 1998.
39. M.I. Mishchenko, 'Light Scattering by Randomly Oriented Axially Symmetric Particles', *J. Opt. Soc. Am., A: Opt. Image Sci.*, **8**(6), 871–882 (1991).
40. N.G. Khlebtsov, S. Yu Shchegolev, 'Allowance for Nonspherical Particles in Determining the Parameters of Dispersed Systems by Turbidity Spectrum Method. II. Characteristic Scattering Functions for Systems of Nonspherical Particles with Random Orientation in the Van de Hulst Approximation', *Opt. Spektrosk.*, **42**(6), 1152–1157 (1977).
41. M.I. Mishchenko, L.D. Travis, D.W. Mackowski, *J. Quant. Spectrosc. Radiat. Transfer*, **55**, 535–575 (1996).
42. W. Heller, 'Theoretical Investigations on the Light Scattering of Spheres. XVI. Range of Practical Validity of the Rayleigh Theory', *J. Chem. Phys.*, **42**, 1609–1615 (1965).
43. W. Heller, H.L. Bhatnagar, M. Nakagaki, 'Theoretical Investigations on the Light Scattering of Spheres. XIII', *J. Chem. Phys.*, **36**, 1163–1170 (1962).
44. T. Kourti, J.F. MacGregor, 'Particle Size Determination Using Turbidimetry: Capabilities, Limitations and Evaluation for On-line Applications', *Am. Chem. Soc. Symp. Ser.*, **472**, 34–63 (1991).
45. P.D. Gossen, T. Kourti, A. Penlidis, J.F. MacGregor, A.E. Hamielec, 'On-line Particle Size and Conversion Determination in Continuous Latex Reactor Trains', in *Preprints, 6th International IFAC/IFIP/IMEKO Conference on Instrumentation and Automation in the Paper, Rubber, Plastics and Polymerization Industries*, Akron, OH, 1986.
46. P.D. Gossen, Presented at the Symposium on Particle Size Analysis in Polymer Science, Organized by the PMSE Division, 191st ACS Meeting, Boston, MA, April 1990.

47. M.L. Wallach, W. Heller, A.F. Stevenson, 'Theoretical Investigations on the Light Scattering of Colloidal Spheres. XII. The Determination of Size Distribution Curves from Turbidity Spectra', *J. Chem. Phys.*, **34**, 1796–1802 (1961).
48. D.H. Melik, H.S. Fogler, 'Turbidimetric Determination of Particle Size Distributions of Colloidal Systems', *J. Colloid Interface Sci.*, **92**, 161–180 (1983).
49. J.B. Bateman, E.J. Weneck, D.C. Eshler, 'Determination of Particle Size and Concentration from Spectrophotometric Transmission', *J. Colloid Sci.*, **14**, 308–329 (1959).
50. T. Kourti, J.F. MacGregor, A.E. Hamielec, 'Minimum Number of Turbidity Measurements Required for the Determination of Particle Size Distributions', *J. Colloid Interface Sci.*, **120**, 292–295 (1987).
51. W.H. Press, S.A. Teukolsky, W.T. Vettering, B.P. Flanney, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1992.
52. D.W. Marquardt, 'An Algorithm for Least Squares Estimation of Non-linear Problems', *J. Soc. Ind. Appl. Math.*, **11**(2), 431–441 (1963).
53. A.F. Stevenson, W. Heller, M.L. Wallach, 'Theoretical Investigation on the Light Scattering of Colloidal Spheres. XI. Determination of Size Distribution Curves from Spectra of the Scattering Ratio or from Depolarization Spectra', *J. Chem. Phys.*, **34**, 1789–1795 (1961).
54. R.R. Irani, C.F. Callis, *Particle Size: Measurement, Interpretation and Application*, John Wiley & Sons, New York, 1963.
55. E.P. Honig, 'Logarithmic Distribution Functions for Colloidal Particles', *J. Phys. Chem.*, **69**, 4418–4419 (1965).
56. J.F. Yan, 'Mass Log-normal Distribution Function for Particle Size Analysis', *J. Colloid Interface Sci.*, **49**, 152–153 (1974).
57. W.F. Espenscheid, M. Kerker, E. Matijevic, 'Logarithmic Distribution Functions for Colloidal Particles', *J. Phys. Chem.*, **68**, 3093–3096 (1964).
58. M. Wales, 'Particle Size Distribution in Rubber Latex', *J. Phys. Chem.*, **66**, 1768–1772 (1962).
59. G. Herdan, *Small Particle Statistics*, 2nd edition, Butterworths, London, 1960.
60. Er. Gulari, A. Annapragada, Es. Gulari, B. Jawad, 'Determination of Particle Size Distribution using Light Scattering Techniques', *Am. Chem. Soc. Symp. Ser.*, **332**, 133–145 (1987).
61. G.E. Eliçabe, L.H. Garcia-Rubio, 'Latex Particle Size Distribution from Turbidimetry Using a Combination of Regularization Techniques and Generalized Cross-validation', *Adv. Chem. Ser.*, **227**, 83–104 (1990).
62. G.E. Eliçabe, L.H. Garcia-Rubio, 'Latex Particle Size Distribution from Turbidimetry Using Inversion Techniques', *J. Colloid Interface Sci.*, **129**(1), 192–200 (1989).
63. J. He, S.-M. Wang, J.-C. Cheng, S.-Y. Zhang, 'Inversion of Particle Size Distribution from Light Scattering Spectrum', *Inverse Problems*, **12**, 633–639 (1996).
64. M. Bertero, C. De Mol, E.R. Pike, 'Particle Size Distributions from Spectral Turbidity: a Singular System Analysis', *Inverse Problems*, **2**, 247–258 (1986).
65. L.C. Lawson, R.J. Hanson, *Solving Least Squares Problems*, SIAM – Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995.
66. P. Bagchi, R.D. Vold, 'A Simple Method for Determination of the Average Particle Size of Coarse Suspensions from Measurements of Apparent Specific Turbidity', *J. Colloid Interface Sci.*, **53**, 194 (1975).
67. S.H. Maron, C. Moore, A.S. Powell, 'Electron Microscopy of Synthetic Latices', *J. Appl. Phys.*, **23**, 900–905 (1952).
68. K. Kubota, S. Hayashi, H. Mandai, 'A Few Comments on Simple Optical Measurements for Determining Sizes of Latex Particles Suspended in Water', *J. Colloid. Interface Sci.*, **98**, 581 (1984).
69. R.A. Dobbins, 'Measurement of a Mean Particle Size in a Gas Particle Flow', *AIAA J.*, **1**, 1940–1942 (1963).
70. R.A. Dobbins, G.S. Jizmagian, 'Optical Scattering Cross Sections for Polydispersions of Dielectric Spheres', *J. Opt. Soc. Am.*, **56**, 1345–1350 (1966).
71. L.H. Garcia-Rubio, 'Averages from Turbidity Measurements', *Am. Chem. Soc. Symp. Ser.*, **332**, 161–178 (1987).
72. G.C.N. Cheesman, 'A Rapid Method for the Measurement of the Average Particle Size of Non-uniform Latices and Dispersions', in *Particle Size Analysis*, ed. M.J. Groves, Heyden, London, 160–166, 1978.
73. D.F. Nicoli, T. Kourti, P.D. Gossen, J.-S. Wu, J.F. MacGregor, Y.-J. Chang, 'On-line Latex Particle Size Determination by Dynamic Light Scattering: Design for an Industrial Environment', *Am. Chem. Soc. Symp. Ser.*, **472**, 86–97 (1991).
74. G. Crawley, M. Cournil, D. Benedetto, 'Size Analysis of Fine Particle Suspensions by Spectral Turbidimetry: Potential and Limits', *Powder Technol.*, **91**, 197–208 (1997).
75. C.F. Bohren, D.R. Huffman, *Absorption and Scattering of Light by Small Particles*, John Wiley & Sons, New York, 1983.
76. P. Walstra, 'Discussion of Errors in Turbidimetry', *Br. J. Appl. Phys.*, **16**, 1187–1192 (1965).
77. R.O. Cumprecht, C.M. Sliepcevich, 'Scattering of Light by Large Spherical Particles', *J. Phys. Chem.*, **57**, 90–95 (1953).
78. R.M. Tabibian, W. Heller, J.N. Eppel, 'Light Scattering of Colloidal Spheres. (I) Specific Turbidity', *J. Colloid Sci.*, **11**, 195–213 (1956).
79. J. Brandrup, E.H. Immergut, *Polymer Handbook*, 2nd edition, John Wiley & Sons, New York, 1975.
80. J. Morgan, *Introduction to Geometric and Physical Optics*', McGraw-Hill, New York, 1953.
81. U. Comberg, T. Wriedt, 'Scattering by Inhomogeneous Particles, Exact Theories and Effective Medium Theories', in *Electromagnetic and Light Scattering – Theory and*

Applications III, eds. T. Wriedt, Yu. Eremin, University of Bremen, Bremen, 1998.

82. V.I. Ivakhnenko, J.C. Stover, Yu.A. Eremin, 'Modeling of Light Scattering by Nonspherical Inhomogeneous Particles', in *Electromagnetic and Light Scattering – Theory and Applications III*, eds. T. Wriedt, Yu. Eremin, University of Bremen, Bremen, 1998.
83. J. Sinzig, M. Quinten, 'Scattering and Absorption by Spherical Multilayer Particles', *Appl. Phys. A: Solids Surf.*, **A58**, 157–162 (1994).
84. A.E. Hamielec, J.D. Wright, *Chem. Can.*, December, 27–29 (1973).
85. S.W. Ray, 'Minicomputer Control of a Latex Reactor', M. Eng. Thesis, McMaster University, Hamilton, Ontario, 1976.
86. M.J. Pollock, 'Modeling and Control of Sustained Oscillations in the Continuous Emulsion Polymerization of Vinyl Acetate', PhD Thesis, McMaster University, Hamilton, Ontario, 1984.
87. J.C. Giddings, G. Karaiskakis, K.D. Caldwell, M.N. Myers, 'Colloid Characterization by Sedimentation Field-flow Fractionation. I. Monodisperse Populations', *J. Colloid Interface Sci.*, **92**, 66–80 (1983).
88. F.-S. Yang, K.D. Caldwell, J.C. Giddings, 'Colloid Characterization by Sedimentation Field-flow Fractionation. II. Particle Size Distribution', *J. Colloid Interface Sci.*, **92**, 81–91 (1983).
89. F.-S. Yang, K.D. Caldwell, M.N. Myers, J.C. Giddings, 'Colloid Characterization by Sedimentation Field-flow Fractionation. III. Emulsions,' *J. Colloid Interface Sci.*, **93**, 115–125 (1983).
90. M.J. Groves, 'Particle Size Characterization in Dispersions', *J. Dispersion Sci. Technol.*, **1**, 97–124 (1980).

Ultrasonic Measurements in Particle Size Analysis

David Julian McClements
University of Massachusetts, Amherst, USA

1 Introduction 1
2 Measurement Techniques 1
2.1 Continuous Wave Techniques 2
2.2 Pulse Techniques 3
3 Particle-ultrasonic Wave Interactions 4
3.1 Intrinsic Absorption 4
3.2 Visco-inertial Dissipation Losses 4
3.3 Thermal Dissipation Losses 4
3.4 Scattering Losses 4
4 Ultrasonic Scattering Theory 5
5 Interpretation of Ultrasonic Spectra 5
6 Applications 6
7 Perspective and Future Developments 6
Abbreviations and Acronyms 7
Related Articles 7
References 7

The particle size distribution and concentration of a colloidal dispersion can be determined by measuring its ultrasonic velocity and/or attenuation coefficient as a function of frequency and then using a suitable mathematical model to interpret the spectra. Ultrasonic spectroscopy can be used to analyze particle sizes between about 10 nm and 1000 μm, and is suitable for application to concentrated systems (often up to 50 wt%). This technique has considerable advantages over many alternative technologies because it can be applied to systems that are optically opaque without the need of any sample preparation.

1 INTRODUCTION

Ultrasonic spectroscopy utilizes measurements of the frequency dependence of the ultrasonic velocity and/or attenuation coefficient of colloidal dispersions to obtain information about the concentration and size distribution of colloidal particles. The velocity (c) at which an ultrasonic wave propagates through a particulate suspension, and the amount by which it is attenuated (α), are governed by interactions between the ultrasonic wave and the particles, e.g. transmission, reflection, absorption, and scattering. Ultrasonic attenuation spectroscopy (UAS) is more commonly used than ultrasonic velocity spectroscopy (UVS), although both methods can provide similar information. Ultrasonic spectroscopy has advantages over many existing particle-sizing technologies because it is non-destructive and non-invasive, is capable of rapid measurements and can be used to characterize systems which are concentrated and optically opaque. Ultrasonic spectroscopy is sensitive to particles with radius between about 10 nm and 1000 μm. The maximum particle concentration which can be analyzed varies between about 10 and 50 wt% depending on the nature of the system. On the other hand, the technique is unsuitable for analyzing dilute suspensions, i.e. particle concentrations below about 1 wt%. The other major limitations of the technique are that air bubbles can interfere with measurements and that a large amount of information about the thermophysical properties of the component phases are needed to interpret ultrasonic spectra.

There are two major steps in the determination of particle size using ultrasonic spectroscopy: (i) measurement of the ultrasonic velocity and/or attenuation spectra of the colloidal dispersion and (ii) interpretation of the resulting spectra using a suitable theory. To obtain accurate measurements it is necessary to carry out both steps carefully.

2 MEASUREMENT TECHNIQUES

Determination of the particle size distribution of a colloidal dispersion relies on being able to measure the frequency dependence of the ultrasonic velocity and/or attenuation coefficient.[1–3] The ultrasonic velocity is the distance the ultrasonic wave moves through the sample per unit time, whereas the attenuation coefficient is a measure of the decrease in the amplitude of the ultrasonic wave per unit distance traveled. The ultrasonic velocity can be determined by two different methods: (i) measuring the time, t taken for a pulse of ultrasound to travel a known distance, x: $c = x/t$, or (ii) measuring the wavelength of an ultrasonic wave, λ, at a known frequency, f: $c = \lambda f$. The attenuation coefficient is determined by measuring the reduction in amplitude of an ultrasonic wave which has traveled a known distance through a material (Equation 1):

$$\alpha = -\frac{1}{x} \ln \frac{A_x}{A_0} \tag{1}$$

where A_0 is the initial amplitude of the ultrasonic wave ($x = 0$), and A_x is the amplitude after it has traveled a distance x.

To cover a wide range of particle sizes it is necessary to measure the ultrasonic properties of a colloidal

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

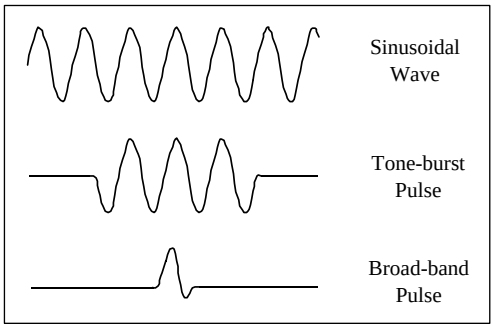

Figure 1 Different types of signal used in the ultrasonic analysis of colloidal dispersions.

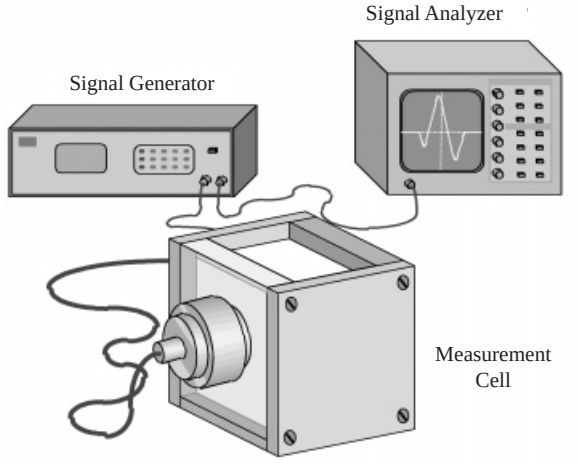

Figure 2 A typical ultrasonic spectrometer consists of a signal generator, a signal analyzer and a measurement cell.

dispersion over a wide range of frequencies.[2] Commercial ultrasonic particle-sizing instruments typically make measurements over the range 1 to 200 MHz, which enables them to analyze particles between about 10 nm and 1000 μm in radius. This range can be extended by using custom-built techniques to carry out measurements at higher or lower frequencies.[4–7]

The frequency dependence of the ultrasonic properties of colloidal dispersions can be determined using a number of different measurement principles.[4–8] The major difference between them is the form in which the ultrasonic signal is applied to the sample, i.e. continuous wave, tone-burst or pulse (Figure 1). Nevertheless, the experimental apparatus used in the different techniques is usually fairly similar, consisting of a signal generator, a signal analyzer and a measurement cell (Figure 2).

2.1 Continuous Wave Techniques

Ultrasonic spectrometers which utilize continuous waves are usually referred to as *interferometers*.[9] Interferometry of liquids can be carried out using two different types of experimental arrangement: fixed path length and variable path length (Figure 3).

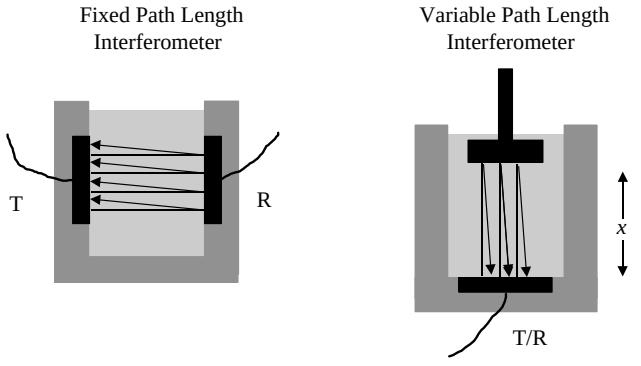

Figure 3 An interferometer used to measure the frequency-dependent ultrasonic properties of liquids may be of the fixed path length or moveable path length type.

2.1.1 Fixed Path Length Devices

In a fixed path length (or resonator) device, the sample to be analyzed is placed in a thermostated measurement cell between two parallel piezoelectric ultrasonic transducers: a transmitter and a receiver. A signal generator produces an electrical sinusoidal wave of appropriate frequency and amplitude. This wave is applied to the transmitter where it is converted into a sinusoidal ultrasonic wave that propagates into the sample. The ultrasonic wave undergoes multiple reflections between the two transducers which results in the formation of a stationary wave. The signal detected by the receiving transducer is amplified and relayed to a signal analyzer. The amplitude of the received signal goes through a series of maxima and minima, due to constructive and destructive interference, when the frequency of the ultrasonic wave is varied. The ultrasonic velocity and attenuation coefficient are determined by analyzing the resulting amplitude versus frequency spectra.

To a first approximation the attenuation per cycle ($\alpha\lambda$) is given by Equation (2):

$$\alpha\lambda = \pi \frac{\Delta f_N}{f_N} \tag{2}$$

where f_N is the central frequency of a maxima and Δf_N is the width of the maxima at -3 dB, whilst the ultrasonic velocity is given by Equation (3):

$$c = 2d\Delta f \tag{3}$$

where d is the path length of the measurement cell and Δf is the frequency difference between the central frequencies of two successive maxima.

2.1.2 *Variable Path Length Devices*

In a variable path length device the sample to be analyzed is placed in a thermostated measurement cell which usually contains an ultrasonic transducer and a moveable reflector plate. A signal generator produces a sinusoidal electrical wave of the appropriate frequency and amplitude which is applied to the ultrasonic transducer where it is converted into a sinusoidal ultrasonic wave. The ultrasonic wave propagates into the sample and undergoes multiple reflections between the transducer and reflector plate which results in the formation of a standing wave. When the distance between the transducer and reflector plate is varied the amplitude of the signal received by the transducer goes through a series of maxima and minima because of constructive and destructive interference. The distance between successive maxima (Δd) is equal to half the ultrasonic wavelength of the sample, and so the ultrasonic velocity can be calculated: $c = f\lambda = 2f\Delta d$. The amplitudes of the maxima decrease as the distance between the reflector plate and transducer is increased because of attenuation by the sample. The attenuation coefficient is determined from Equation (1) by measuring the amplitude of the maxima as a function of the separation between the reflector plate and transducer.

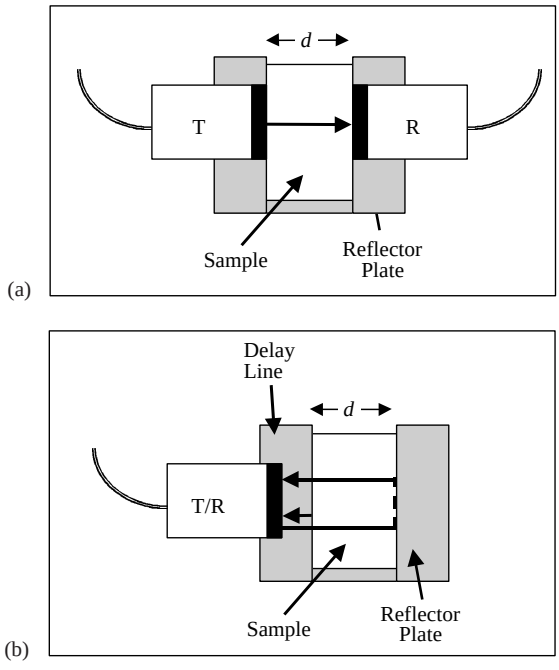

Figure 4 The ultrasonic properties of a colloidal dispersion can be determined using pulsed techniques by (a) the through-transmission or (b) the pulse-echo methods.

2.2 Pulse Techniques

Ultrasonic spectrometers that utilize pulse techniques may be operated in a *pulse-echo* or *through-transmission* mode.[8,9] In the pulse-echo mode a single transducer is used to both transmit and receive the ultrasonic signal, whereas in the through-transmission mode separate transducers are used to transmit and receive the signal (Figure 4).

2.2.1 *Through-transmission Techniques*

The sample to be analyzed is placed in a thermostated measurement cell between two ultrasonic transducers: a transmitter and a receiver (Figure 4). The transmitting transducer produces a pulse of ultrasound which travels across the sample and is detected by the receiving transducer. The ultrasonic velocity and attenuation coefficient of the sample are determined by measuring the time-of-flight (Δt) and amplitude (A) of the ultrasonic pulse which has traveled across the sample. The ultrasonic velocity is equal to the length of the sample, d, divided by the time required to travel this distance: $c = d/\Delta t$. The attenuation coefficient is calculated by comparing the reduction in the amplitude of the pulse that has traveled through the sample with that of a pulse which has traveled through a calibration material: $\alpha_s = \alpha_c - \ln(A_s/A_c)/d$, where the subscripts c and s refer to the properties of the calibrant and sample, respectively. To obtain accurate attenuation measurements it is necessary to ensure that the measurement cell is well designed to minimize temperature fluctuations, reverberation of ultrasonic pulses in cell walls and transducers, diffraction effects and phase cancellation due to nonparallel walls.[8]

The frequency dependence of the ultrasonic properties of a sample can be obtained using two different approaches: (i) the Fourier Transform (FT) method, and (ii) the tone-burst method. For both approaches, a pair of ultrasonic transducers is used which is capable of generating ultrasonic energy over a wide range of frequencies. These *broad-band* transducers contain piezoelectric crystals whose oscillation is highly damped by attaching a backing material to the crystal.

In the FT method, a signal generator is used which is capable of producing a broad-band electrical pulse which contains a wide range of frequencies.[9–11] This pulse is applied to the transmitting transducer where it is converted into a broad-band ultrasonic pulse. After this pulse has traveled through the sample it is detected by the receiving transducer, and then digitized and stored. The frequency dependence of the magnitude and phase of the pulse are determined using a FT technique. The ultrasonic velocity is then determined from the phase, whilst the attenuation coefficient is determined from the magnitude. To cover the whole frequency range it is often necessary to use two or three pairs of broad-band transducers

with different central frequencies. Ultrasonic velocity and attenuation spectra are acquired much more rapidly using FT pulse techniques than continuous-wave techniques. This is because the whole of the frequency range is measured by propagating a pulse through the sample, rather than having to make measurements separately at each individual frequency.

In the tone-burst technique, a signal generator is used which is capable of generating tone-burst pulses over the required frequency range. A ''tone-burst'' is a single pulse which contains a number of cycles of ultrasound at a fixed frequency (Figure 1). A tone-burst pulse of a particular frequency is applied to the transmitter, and then the magnitude and phase of the tone-burst pulse that travels through the sample and is detected by the receiver are measured.[12] The ultrasonic attenuation coefficient is determined by measuring the amplitude of the received signal, whilst the ultrasonic velocity is determined by measuring the change in the phase of the received signal. An ultrasonic spectrum is acquired by repeating this process for a number of frequencies. Because measurements are carried out separately at each frequency, this approach is more time-consuming and laborious than the FT method.

2.2.2 Pulse-echo Techniques

The frequency-dependent ultrasonic properties of a sample are measured in a very similar fashion as in the through-transmission technique, except that a single transducer is used to both transmit and receive the ultrasonic pulses (Figure 4). The ultrasonic transducer generates a pulse of ultrasound which travels across the sample, is reflected from the back wall of the measurement cell, travels back through the sample, and is then detected by the same transducer. The velocity and attenuation coefficient are calculated in exactly the same manner as described for the through-transmission technique, expect that the pulse has now traveled a distance $2d$ rather than d.

To obtain accurate and reliable measurements it is important to take extreme care in designing and manufacturing measurement cells, for all types of ultrasonic technique. Some of the most important factors to consider are nonparallelism of surfaces, imperfect reflection at boundaries, reverberations in cell walls, temperature fluctuations and diffraction losses.[8]

3 PARTICLE-ULTRASONIC WAVE INTERACTIONS

A knowledge of the physical basis of the various forms of interaction between ultrasonic waves and particles is necessary to understand how ultrasound can be used to measure particle size.[1–3,8] The four most important types of interaction between an ultrasonic wave and a colloidal dispersion are listed below.

3.1 Intrinsic Absorption

Each of the individual component phases in a colloidal suspension absorbs ultrasound as a result of classical (viscous and thermal) and relaxation loss mechanisms. These losses are determined by the composition of the suspension, rather than by its microstructure. The magnitude of these losses can be calculated by measuring the ultrasonic absorption coefficients of the individual components.

3.2 Visco-inertial Dissipation Losses

In the presence of an ultrasonic wave a particle oscillates backwards and forwards because it has a different density to that of the surrounding liquid. This oscillation is damped by the viscosity of the surrounding liquid, which causes some of the ultrasonic energy to be converted to heat, and therefore leads to a reduction in the amplitude of the ultrasonic wave. The magnitude of these visco-inertial losses increases as the density difference between the droplet and surrounding liquid increases. Visco-inertial dissipation losses depend on the size of the droplets in a colloidal suspension.

3.3 Thermal Dissipation Losses

In the presence of an ultrasonic wave the temperature of a particle periodically increases and decreases relative to that of the surrounding liquid because of differences in the thermal properties of the continuous and disperse phases. The temperature gradient which arises at the particle surface causes a net flow of heat out of the droplet, which leads to a reduction in the amplitude of the ultrasonic wave, i.e. some of the energy initially stored in the ultrasonic wave is converted to heat. The magnitude of these losses is governed by the difference in thermal properties between the droplet and the surrounding liquid, e.g. thermal expansion coefficient, specific heat capacity and thermal conductivity. Thermal dissipation losses depend on the size of the droplets in a colloidal suspension.

3.4 Scattering Losses

The pulsation and oscillation of a particle in the presence of an ultrasonic wave causes the generation of secondary ultrasonic waves by the particle. Thus some of the ultrasonic energy associated with the incident wave is redirected into different directions, and an increase in

the attenuation coefficient may be detected. Scattering losses are usually negligible in the long wavelength limit ($r \ll \lambda$), but become dominant when the particle size is of the same order of magnitude as the ultrasonic wavelength. Scattering losses depend on the size of the droplets in a colloidal suspension.

To a first approximation, the overall attenuation coefficient of a colloidal suspension can be considered to be the sum of these various contributions (although in reality some of these mechanisms are coupled to each another). In most suspensions, one or two of the above mechanisms usually dominate the overall attenuation in a particular frequency range. At relatively low frequencies, the visco-inertial and thermal loss mechanisms usually dominate, but at higher frequencies the intrinsic absorption and scattering losses usually dominate.

4 ULTRASONIC SCATTERING THEORY

Once the ultrasonic velocity and/or attenuation spectra of an emulsion have been measured it is necessary to convert them into a particle size distribution using an appropriate theory. Theories are based on a mathematical treatment of the physical processes that occur when an ultrasonic wave propagates through an ensemble of particles suspended in a fluid.[1–3] One of the most comprehensive models (Equation 4) is based on multiple scattering theory:[13]

$$\left(\frac{K}{k_1}\right)^2 = 1 + \frac{4\pi N f(0)}{k_1^2} + \frac{4\pi^2 N^2}{k_1^4}(f^2(0) - f^2(\pi)) \quad (4)$$

where $f(0)$ and $f(\pi)$ are the *scattering amplitudes* of the individual droplets (Equations 5 and 6):

$$f(0) = \frac{1}{ik_1}\sum_{n=0}^{\infty}(2n+1)A_n \quad (5)$$

$$f(\pi) = \frac{1}{ik_1}\sum_{n=0}^{\infty}(-1)^n(2n+1)A_n \quad (6)$$

Here $K(=w/c_S + i\alpha_S)$ is the complex propagation constant, c_S is the ultrasonic velocity and α_S is the attenuation coefficient of the colloidal suspension. k_1 is the complex propagation constant of the continuous phase $(= w/c_1 + i\alpha_1)$, ϕ is the disperse phase volume fraction, $w(= 2\pi f)$ is the angular frequency f is the frequency, $i = (-1)^{1/2}$, and r is the droplet radius. The A_n terms are the *scattering coefficients* of the various types of waves scattered from the individual droplets, e.g. monopole (A_0), dipole (A_1), quadrupole (A_2) etc.

Approaches for calculating the scattering coefficients of both fluid and solid particles are available in the literature.[14,15] The most rigorous approach calculates the A_n terms by solving a series of 6×6 complex linear simultaneous equations at each value of n, although simpler analytical expressions are available in the long wavelength limit.[14,15] The values of the scattering coefficients depend on the relative thermophysical properties of the component phases, the ultrasonic frequency used and the size of the emulsion droplets. The terms containing ϕ in Equation (4) describe single scattering effects, whilst the terms containing ϕ^2 describe multiple scattering effects. Multiple scattering becomes increasingly important as the concentration of droplets in a colloidal suspension increases.

For a colloidal suspension containing polydisperse particles the above equation must be modified (Equation 7):

$$\left(\frac{K}{k_1}\right)^2 = 1 + \frac{4\pi}{k_1^2}\sum_j N_j f_j(0) + \frac{4\pi^2}{k_1^4}\sum_j N_j^2 \quad f_j^2(0) - f_j^2(\pi) \quad (7)$$

where the subscript j refers to the property with droplet size r_j. This equation can be used to relate the ultrasonic properties of a colloidal suspension (velocity and attenuation coefficient) to its thermophysical properties, composition (ϕ) and particle size distribution.

Recently, it has been shown that the above theory must be modified to take into account interactions due to overlap of viscous[16] and thermal waves[17] generated by the particles. These interactions cause large deviations between the classical multiple scattering theory and experimental measurements at low ultrasonic frequencies, small droplet sizes and high droplet concentrations.

5 INTERPRETATION OF ULTRASONIC SPECTRA

To determine the particle size distribution of a colloidal dispersion one usually measures its ultrasonic velocity and/or attenuation coefficient as a function of frequency. The particle size distribution is then determined by finding the droplet size distribution that gives the best fit between the predictions of the ultrasonic scattering theory and the experimental ultrasonic spectra.[1,2] There are two approaches to solving this inverse scattering problem: *model-independent* inversion and *model-dependent* inversion. In the model-dependent inversion method it is assumed that the particle size distribution follows some common form which can simply be modeled mathematically, e.g. log-normal (Equation 8).

$$P(r) = \frac{1}{x_g \ln \sigma_g \sqrt{2\pi}} \exp - \left(\frac{\ln^2 \sigma_g}{2}\right. \quad \exp\left(\frac{[\ln r - \ln x_g]^2}{2\ln^2 \sigma_g}\right. \quad (8)$$

where $P(r)$ is the probability of having a particle of radius r, x_g is the geometric mean of the radius and σ_g is the standard deviation of the geometric mean. The droplet size distribution can then be characterized by only two parameters: x_g and σ_g. The geometric mean and standard deviation which give the best agreement between the measured and predicted ultrasonic spectra are found by a least squares analysis. Most commercial ultrasonic instruments assume that the particle size distribution follows either a monomodal or bimodal log-normal distribution.

In the model-independent analysis no a priori assumption is made about the form of the particle size distribution. Although this method is scientifically more satisfactory, because many suspensions do not have particle size distributions which can be described by simple mathematical expressions, it is extremely time-consuming to compute. In addition, there are sometimes situations where a number of different particle size distributions give an equally good agreement between theory and experiment and it is not possible to decide which is correct.

6 APPLICATIONS

The ability of ultrasonic spectroscopy to determine the particle size distribution of colloidal suspensions has been demonstrated by many workers.[3,6,16,18–21] Fully automated particle sizing instruments which utilize ultrasonic spectroscopy are available commercially. These instruments are capable of measuring both the disperse phase volume fraction and particle size distribution of colloidal suspensions in a few minutes. For certain systems, these instruments are capable of characterizing suspensions containing high particle concentrations (sometimes as high as 60%).

The attenuation spectra of a series of corn oil-in-water emulsions with different disperse phase volume fractions (1–50 vol%) are shown in Figure 5. The particle size distribution and droplet concentration measured by the ultrasonic technique are tabulated in Table 1. There is excellent agreement between the measured and actual volume fractions of the emulsions up to the highest droplet concentration. In addition, the mean particle diameter is relatively insensitive to droplet concentration, indicating that the ultrasonic technique is capable of analyzing this system over a wide range of concentrations without any sample preparation. Nevertheless, it should be noted that in systems where thermal or viscous overlap effects become important there may be quite large disagreements between the experimental data and the theories used to interpret them.[16,17]

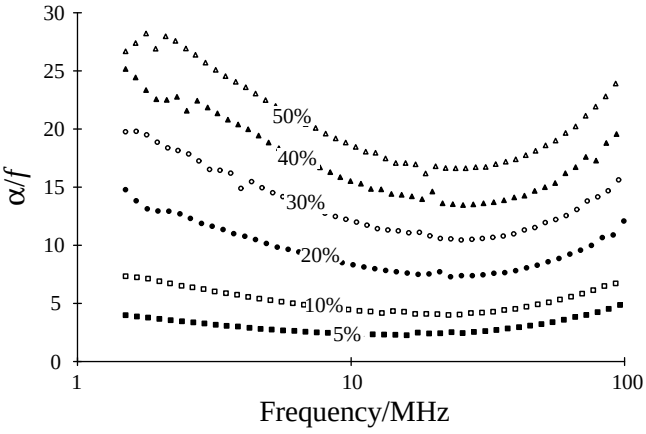

Figure 5 Ultrasonic attenuation spectra of corn oil-in-water emulsions containing different droplet concentrations. The particle size distribution and disperse phase volume fraction are determined by finding the values which give the best-fit between the measurements and ultrasonic scattering theory.

Table 1 Disperse phase volume fraction (ϕ), geometric mean (x_g) and standard deviation of the geometric mean (σ) of various corn oil-in-water emulsions determined by finding the best fit between ultrasonic scattering theory and the experimental measurements shown in Figure 5

Actual ϕ	Measured ϕ	x_g (μm)	σ
50	47.9	1.32	2.50
40	37.9	1.17	2.24
30	30.1	1.20	2.07
20	20.6	1.18	1.93
10	10.5	1.14	1.85
5	5.5	1.16	1.74
1	1.5	1.30	1.14

7 PERSPECTIVE AND FUTURE DEVELOPMENTS

The ability of ultrasonic spectroscopy to characterize colloidal suspensions which are concentrated and optically opaque without the need for any form of sample preparation is extremely important for many applications. There are few other particle sizing technologies which are capable of analyzing this type of system and therefore it seems likely that the ultrasonic technique will find increasing applications in the future. One of the areas where the technique is most likely to be used is for the on-line determination of particle size distributions of colloidal suspensions during processing. The ultrasonic technique could be used to monitor the efficiency of a processing operation in real time that could lead to a major improvement in the manufacture of many colloidal-based materials, e.g. foods, pharmaceuticals, petrochemicals, agrochemicals and cosmetics.

Despite its many advantages over alternative technologies there are a number of technical issues that need to be addressed before the technique finds more widespread use:

1. The classical multiple scattering theory used to interpret ultrasonic measurements gives excellent agreement with experimental data for semi-dilute emulsions containing noninteracting droplets. Nevertheless, it needs to be extended so that it can be applied to more concentrated emulsions and to take into account the effects of droplet flocculation.
2. The interpretation of ultrasonic measurements requires knowledge of a number of the thermophysical properties of the disperse and continuous phases, which are often not readily available in the literature or are difficult to measure. It is therefore important to develop a database of the relevant thermophysical properties for the materials commonly found in colloidal suspensions.
3. The presence of small gas bubbles can obscure the signal from colloidal particles because of their ability to strongly scatter ultrasound. Strategies for overcoming this problem need to be developed, e.g. degassing the sample or making measurements at frequencies where the gas bubbles do not scatter significantly.

ABBREVIATIONS AND ACRONYMS

FT	Fourier Transform
UAS	Ultrasonic Attenuation Spectroscopy
UVS	Ultrasonic Velocity Spectroscopy

RELATED ARTICLES

Food **(Volume 5)**
Particle Size Analysis in Food

Particle Size Analysis **(Volume 6)**
Centrifugation in Particle Size Analysis • Diffraction in Particle Size Analysis • Electrozone Sensing in Particle Size Analysis • Field-flow Fractionation in Particle Size Analysis • Filtration in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Photon Correlation Spectroscopy in Particle Sizing • Sedimentation in Particle Size Analysis • Sieving in Particle Size Analysis • Surface Area and Pore Size Distributions • Turbidimetry in Particle Size Analysis • Velocimetry in Particle Size Analysis

REFERENCES

1. U. Riebel, F. Loffler, 'The Fundamentals of Particle Size Analysis by means of Ultrasonic Spectrometry', *Part. Part. Syst. Charact.*, **6**, 135–143 (1989).
2. D.J. McClements, 'Principles of Ultrasonic Droplet Size Determination', *Langmuir*, **12**, 3454–3461 (1996).
3. M.J.W. Povey, *Ultrasonic Techniques for Fluid Characterization*, Academic Press, San Diego, 1997.
4. U. Kaatze, B. Wehrmann, R. Pottel, 'Acoustic Absorption Spectroscopy of Liquids between 0.15 and 3000 MHz. I. High Resolution Ultrasonic Resonantor Method', *J. Phys. E.*, **20**, 1035–1040 (1987).
5. U. Kaatze, K. Lautscham, M. Brai, 'Acoustic Absorption Spectroscopy of Liquids between 0.15 and 3000 MHz. II. Ultrasonic Pulse Transmission Methods', *J. Phys. E.*, **21**, 98–103 (1988).
6. U. Kaatze, B.K. Lautscham, 'Acoustic Absorption Spectroscopy of Liquids between 0.15 and 3000 MHz: II. Hypersonic Comparator Technique', *J. Phys. E.*, **21**, 402–406 (1988).
7. U. Kaatze, C. Trachimow, R. Pottel, M. Brai, 'Broadband Study of the Scattering of Ultrasound by Polystyrene-Latex-in-water Suspensions', *Ann. Physik*, **5**, 13–33 (1996).
8. D.J. McClements, 'Ultrasonic Characterization of Emulsions and Suspensions', *Adv. Coll. Int. Sci.*, **37**, 33–72 (1991).
9. F.G. Eggers, U. Kaatze, 'Broadband Ultrasonic Measurement Techniques for Liquids', *Meas. Sci. Tech.*, **7**, 1–19 (1996).
10. R.E. Challis, J.A. Harrison, A.K. Holmes, R.P. Cocker, 'A Wide Bandwidth Spectrometer for Rapid Ultrasonic Absorption Measurements in Liquids', *J. Acoust. Soc. Am.*, **90**, 730–740 (1991).
11. D.J. McClements, P. Fairley, 'Frequency Scanning Ultrasonic Pulse Echo Reflectometer', *Ultrasonics*, **30**, 403–406, (1992).
12. D.J. McClements, P. Fairley, 'Ultrasonic Pulse Echo Reflectometer', *Ultrasonics*, **29**, 58–60 (1991).
13. P.C. Waterman, R. Truell 'Multiple Scattering of Waves', *J. Math. Phys.*, **2**, 512–537 (1962).
14. P.S. Epstein, R.R. Carhart, 'The Absorption of Sound in Suspensions and Emulsions', *J. Acoust. Soc. Am.*, **25**, 553–565 (1953).
15. J.R. Allegra, S.A. Hawley, 'Attenuation of Sound in Suspensions and Emulsions: Theory and Experiments', *J. Acoust. Soc. Am.*, **51**, 1545–1564 (1972).
16. A.S. Dukhin, P.J. Goetz, 'Acoustic and Electroacoustic Spectroscopy', *Langmuir*, **12**, 4998–5003 (1996).
17. Y. Hemar, N. Herrmann, P. Lemarechal, R. Hocquart, F. Lequeux, 'Effective Medium Model for Ultrasonic Attenuation due to the Thermo-elastic Effect in Concentrated Emulsions', *J. Phys. II France*, **7**, 637–642 (1997).
18. D.J. McClements M.J.W. Povey. 'Scattering of Ultrasound by Emulsions', *J. Phys. D.*, **22**, 38–47 (1989).

19. D.J. McClements, M.J.W. Povey, M. Jury, E. Betsanis, 'Ultrasonic Characterization of a Food Emulsion', *Ultrasonics*, **28**, 266–272 (1990).
20. N. Herrmann, 'Application of Ultrasonic Techniques to the Study of Dispersions', Ph.D thesis, University of Louis Pasteur, Strasbourg, France, 1997.
21. A.S. Dukhin, P.J. Goetz, 'Acoustic Spectroscopy for Concentrated Polydisperse Colloids with High Density Contrast', *Langmuir*, **12**, 4987–4997 (1996).

Velocimetry in Particle Size Analysis

Patrick V. Farrell
University of Wisconsin-Madison, Madison, USA

1 Introduction **1**
2 Classification of Flow Types **2**
2.1 Characteristics of Particle Flows for Velocity Measurements 2
2.2 Characteristics of Continuous Phase for Particle Velocity Measurements 3
3 Fundamental Characteristics of Measurement Methods **5**
3.1 Particle Characteristics Useful for Velocity Measurements 5
3.2 Temporal and Spatial Resolution Requirements and Limitations 7
3.3 Coupling Velocity and Particle Size Measurements 8
3.4 Biases and Uncertainties in Particle Velocity Measurements 9
4 Applications of Measurement Approaches **9**
4.1 Frequency Shift (Doppler) Methods 9
4.2 Direct Imaging Methods 15
4.3 Recent Developments in Particle Image Velocimetry 20
5 Summary **21**
Acknowledgments **21**
Abbreviations and Acronyms **21**
Related Articles **21**
References **21**

Velocimetry in particle size analysis encompasses the range of issues and methods used to measure particle velocities consistent with particle size measurement systems. Particle size may be measured simultaneously or consecutively with the velocity measurements described in this article. This article will present some issues that should be considered in selection of a velocity measurement technique. Three examples of commonly used particle velocity methods will be discussed as illustrations of possible approaches. For each of these examples, a brief discussion of advantages and limitations of the methods will also be presented.

1 INTRODUCTION

A wide variety of physical situations call for measurement of particle or droplet velocities. In many cases, these measurements accompany size measurements, but in some cases particle size measurements are not necessary. This section will focus primarily on the velocity measurements, with some mention of systems which can acquire both size and velocity measurements. The intent is to introduce some of the more common methods in use for measurement of particle and droplet velocities. It is likely that some useful techniques will not be included, but this article may serve as a useful introduction to some of the techniques. In many cases, the measurement methods described are available as partial or complete systems from commercial vendors.

Measurement of particle or droplet velocity can be a challenging task both in acquiring data with appropriate characteristics and in interpreting the measurements so that accurate and appropriate conclusions can be drawn from the experiment. This article will outline some of the attributes of the particles (distributed phase) and the background fluid or continuous phase that should be considered in making measurements of particle or droplet velocities. I will also describe some of the most common current methods used to acquire such measurements. While some example applications will be included, inevitably many areas of interest will not be fully explored. Hopefully, readers will be able to use this introduction to aid in thinking about such measurements, and will feel capable of searching for applications in the open literature that are directly applicable to their areas of interest.

In the context of velocity measurements for particles and droplets, ''particles'' may strictly include a wide range of materials, with the common understanding that a ''particle'' is a solid or apparently solid-phase material suspended in a different phase (gas or liquid) field. Solid particles may exist in solid suspensions, but this combination will not be explicitly discussed in this article. Droplet velocity measurements, however, will be included. A ''droplet'' is considered to be a small amount of a liquid-phase substance suspended in a different liquid or a gas phase. Many of the methods and considerations in this article also apply for bubbles–gas phase suspended in liquid phase. In subsequent parts of this section, ''particle'' velocity measurements will refer to measurements of solid particles, droplets, or bubbles, unless otherwise specified.

Since particle or droplet motion usually involves a continuous phase different from the particle or droplet phase, two-phase or multiphase flow considerations are important in developing measurement systems for these types

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of flows.[1] One of the most important considerations in a multiphase flow is the lack of continuity for the distributed phase (droplet or particle) of the flow. Whether particles or droplets, as discrete entities, they have a distinguishable, Lagrangian character. In many cases, this means that the appearance of particles or droplets at a specific place and time in a flow (aside from any vector or scalar characteristics) seems to be a random event with some statistical character. In contrast, the continuous phase (usually liquid or gas) is almost always present at any location at any time in the flow field (except when displaced by particle or droplet). While some measured characteristics of the continuous phase may appear to be random (like instantaneous velocity in a turbulent field) the presence of a velocity value is not. The discontinuous sampling for particles and droplets presents some difficulties in measurement and interpretation of measurements, which will be discussed in later sections of this article.

The type, quantity, and quality of velocity information sought for particles or droplets will vary with the application and the flow of interest. One useful initial distinction is whether the particle velocity field is of interest in its own right, or is intended to be a marker for another flow field, typically the continuous phase flow field. While particle measurement methods may be similar for these two, the type of information sought may differ considerably.

The kind of information sought about a particle or droplet velocity may include the velocity magnitude, direction, and some marker of time and space to indicate where and when the velocity sample was acquired. In many cases, the detailed spatial and temporal markers are not retained; instead, spatial information is recorded from the known sample volume location or the location of the particle velocity sampling, and temporal marking may be eliminated in favor of some statistical temporal moments (like mean velocity and root mean square (RMS) velocity). Since measurement methods and details may change depending on which of these quantities is needed, initial selection of the type of information sought can be an important step in developing or selecting a measurement system.

While the range of possible measurement methods is broad, the most commonly used particle velocity measurement techniques seem to fall into three major categories: optical LIDAR (Light Detection and Ranging) type methods which rely on Doppler shift of laser light; laser Doppler velocimetry (LDV) systems which also rely on Doppler effects but in a different configuration; and direct particle imaging systems. These three types of measurements, and some variations, will be emphasized throughout this article.

2 CLASSIFICATION OF FLOW TYPES

2.1 Characteristics of Particle Flows for Velocity Measurements

In order to select velocity measurement methods that are well suited to the situation of interest, it may be useful to identify and describe some relevant parameters of the particles or droplets and some parameters of the flow field. In this section, some relevant particle parameters will be identified and typical ranges described. In the following section, relevant parameters of the continuous phase or mean flow field will be described.

Some particle characteristics that impact selection and use of velocity measurement methods are shown in Table 1, and will be discussed below. The range of ''typical'' values shown in Table 1 is primarily for illustration. It is quite likely that successful particle velocity measurements can be made beyond the cited range of each characteristic, but a range is cited to give readers an idea of the range encountered in experiments.

Particle size is one of the most important of the particle characteristics that affect what kind of velocity measurement could be made and the quality of the

Table 1 Some relevant particle characteristics

Particle characteristic	''Typical'' range of values
Size (diameter)	$10\,\text{nm} < d_p < 1.0\,\text{mm}$
Number density (no. of particles/volume of distributed phase)	$1\,\text{cm}^{-3} < n < 10\,000\,\text{cm}^{-3}$ (or more)
Material density (mass of particle/particle volume)	$1.2\,\text{kg m}^{-3} < \rho_m < 1000\,\text{kg m}^{-3}$
Shape (aspect ratios, e.g. d_1/d_2, d_1/d_3)	Spherical ($d_1 = d_2 = d_3$) to $d_1/d_2 \gg 1$ or $d_1/d_3 \gg 1$
Complex index of refraction ($n = m + ik$)	Nonabsorbing, near vacuum ($n = 1 + i0$) to highly absorbing ($k > 0$), high density ($m > 1$)
Steady or transient particle velocities	$1\,\mu\text{s} < \tau < 10\,\text{s}$ (or more)
Surface finish	No clear definition, in this context
Chemical characteristics (toxicity, reactivity, ...)	Nontoxic, nonreactive ... to ... quite toxic and/or reactive
Phase or chemical mutability (vaporization of droplets, reaction for liquid $TiCl_4$ with water vapor $\Rightarrow TiO_2$, ...)	Depends on physical constants (vapor pressure), particle chemistry, and environment (T, P, composition)

resulting measurement. The size range of particles whose velocity can be measured is relatively large. In particular, if the velocity measurement does not require a parallel size measurement, quite small particle sizes can have velocities measured. As will be discussed in the following sections, particle size (along with some other parameters) affects the detectability of the particle, its ability to accelerate and decelerate in concert with the continuous phase, and even whether and how particles will distribute themselves throughout a flow. Some of these topics will be discussed in later sections of this article; for others, readers will be referred to resources addressing these issues.

Perhaps the second most limiting characteristic for particle velocity measurement is the *number density* of particles in the region of interest. In many applications, identifying the velocity of a single particle, not the average velocity of an aggregate of particles, is the goal. To the extent that a single particle cannot be ''isolated'' mechanically or optically, it may be difficult to tag a particular particle with a specific velocity value. In other flows, an average velocity for an aggregate (spatial average) of particles may be sought; if the number density is too low, some methods designed for this type of measurement may not work properly. Most of the particle velocity techniques that will be discussed have ranges of particle number density in which they will work best. Outside of these ranges, the results of the method may not prove satisfactory.

Particle material density is related to index of refraction of the particle as well as its ability (or lack thereof) to be neutrally buoyant in the distributed phase. For marker particles, an expectation is that particles follow the continuous phase to some limit, which usually includes minimum and maximum velocities and acceleration. For example, to follow low velocity continuous-phase flows, particles need to be nearly neutrally buoyant in the fluid, or else they will have gravitationally induced velocities unrelated to the continuous phase velocity. Even in high-velocity flows, particles and droplets may have slip velocities relative to the continuous phase for transient flows.[2]

Particle shape affects the way in which a particle can or will follow a continuous-phase flow, as well as the way in which a particle will scatter light. As mentioned above, many of the particle velocity techniques currently in use depend on light scattering from particles. Nonspherical particles may scatter adequately for some velocity measurements, but present significant difficulties when accompanying sizing methods are used. Further, since many size and velocity methods, like phase Doppler interferometry (PDI) expect spherical particles, the velocities associated with nonspherical particles may be discarded, leading to shape-biased velocity statistics. Identifying scattered light signatures, if that is important, from randomly oriented nonspherical particles is likely to be very difficult.

The particle *complex index of refraction* (and its relation to the continuous-phase index of refraction) combines the refractive effect of change of index of refraction at the particle–fluid interface and the absorption characteristics of the particle. In general, index of refraction for a particle is temperature dependent, and for many materials is only known approximately. Since many particle velocity measurement methods are optically based, this optical parameter is frequently an important parameter affecting scattered signal strength and system signal-to-noise.

Surface finish and the *chemical characteristics* of the particles are also characteristics that can affect choice and implementation of methods. Surface finish is presumably an issue only for solid particles (or perhaps very large droplets and bubbles where the local surface tension and radius of curvature of the interface are not sufficient to produce a ''smooth'' interface). Chemical characteristics may include mostly practical considerations for material handling and disposal, although some characteristics (like fluorescence) can be used as part of the measurement method.

Mutability of the particle refers to whether the particle composition or phase will change under the conditions of the experiment. In some cases, this feature may be a disadvantage, for example when oil droplets are used as markers for a heated gas flow, and the markers vaporize leaving no particles to detect in the flow. In other cases, this mutability can be used to advantage. For example, a liquid like $TiCl_4$ could be atomized into a fine spray in a gas flow with some water vapor in the gas. A chemical reaction of the $TiCl_4$ with the water vapor produces TiO_2 powder and HCl vapor. The TiO_2 provides small markers for the flow; dispersing the $TiCl_4$ as a liquid spray may provide better particle size control than dispersing the TiO_2 powder itself.

In summary, in order to focus on a set of methods likely to be successful for measuring the velocity of a class of particles or droplets, identifying the range of expected particle characteristics is an important first step. When particles are intentionally seeded into a flow, the selection of seed may be revisited once a measurement method has been selected to optimize the combination of particle and method. For naturally occurring particles, careful characterization of the particle characteristics can help make a successful measurement with a minimum of trial-and-error.

2.2 Characteristics of Continuous Phase for Particle Velocity Measurements

As was the case for particle characteristics, certain characteristics of the continuous phase have a

Table 2 Some relevant flow (continuous phase) characteristics

Flow characteristic	"Typical" range of values
Size of flow field of interest ($l \times w \times h$)	1 mm × 1 mm × 1 mm to 10 m × 10 m × 10 m
Access (locations, characteristics of access locations)	Limited (e.g. single location) to everywhere in flow field
Temporal variation of velocity and scale of variation	Steady laminar to unsteady (mean) turbulent
Spatial variation of flow field velocity	Uniform or not
Repeatable?	Yes or no
Complex index of refraction ($n = m + ik$)	Nonabsorbing, near vacuum ($n = 1 + i0$) to highly absorbing ($k > 0$), high density ($m > 1$)
Fluid density	$1.2\,\mathrm{kg\,m^{-3}} < \rho_f < 1000\,\mathrm{kg\,m^{-3}}$
Chemical characteristics	Nontoxic, nonreactive ... to ... quite toxic and/or reactive

significant impact on the success of particle measurement. Table 2 outlines some of the more common flow parameters that affect the choice of particle velocity measurement methods and the measurements themselves.

Many of the general features of a particular experiment are embodied in the description of the continuous phase flow field characteristics. These features include the size of the region of interest, access to the region, and whether the general flow is expected to be steady or transient. Taken together, these features start to suggest whether some sort of instantaneous "snapshot" of the particle velocities everywhere in the flow field is needed, or sampling particle velocities at one location over time is sufficient, or if something in between is needed.

In Table 2, the *size of the flow field* represents the total spatial region of interest. For imaging systems, this may indicate the size of the object field that needs to be imaged. *Access*, in Table 2, refers to the range of locations in the flow field where a sample of the continuous and distributed phase velocity might be available. As most current particle velocity measurement systems are optical, *access* refers primarily to optical access. For some flows, the flow may be accessed or sampled at only a few locations, with a limited "field of view". Other experiments may have fewer limitations, perhaps allowing optical access to a large portion of the flow field for much of the time of interest. These access constraints may be important in selecting and applying a velocity measurement system because the systems themselves require different levels of access.

Temporal variation of velocity and the *scale and frequency of variation* are indications of the flow characteristics of the continuous phase. Many flows have an unsteady or even turbulent character. Depending on the magnitude and frequency of the behavior, for many of these flows, a statistical representation of particle velocity may be most useful. Typical flow statistics of interest may include a mean velocity, a RMS velocity, and perhaps higher moments of the velocity. Flows may exhibit very little variation from a steady mean velocity or slowly varying mean velocity (i.e. small magnitude or very low-frequency variations), indicating a laminar-type flow. Alternatively, the flow may be a transient turbulent flow with time-varying statistics (mean, RMS, moments). Estimating the temporal characteristic of the flow is quite useful in assessing what kind of measurement, what data rate, and what resulting statistics are desired. Flow repeatability is related to this transient character, since this characteristic, when present, makes a repeated measurement on different realizations of the flow possible. A nonrepeatable flow would not permit such a data combination without significant degradation to the resulting data set.

Spatial variation of the velocity is the estimate of whether data sampled in one region of the flow are representative of that region alone, or of some broader range of locations. In parallel with temporal variations, spatial variations of flow velocity can be very small to quite large, responding to the flow geometry as well as to the overall flow type and structure.

The *complex index of refraction* and *fluid density* are of interest on their own, but are even more important when compared with the distributed phase of particles or droplets. Similar particle and continuous phase densities may make particles more likely to follow the flow (neutrally buoyant), while similar indices of refraction may make particles more difficult to identify because of the refractive characteristics for optical measurement systems. While most particle velocity measurement methods do not require exact knowledge of the particle or fluid (continuous phase) index of refraction, signal levels are affected by the difference in index between the particle and fluid.

As for the particles, *chemical characteristics* may be of interest in the measurement itself, or may be part of the measurement system (as for fluorescence) or may be issues to handle for chemical handling and disposal.

3 FUNDAMENTAL CHARACTERISTICS OF MEASUREMENT METHODS

3.1 Particle Characteristics Useful for Velocity Measurements

Particle velocity measurements are based on a variety of measurable quantities. Currently, the most common particle velocity measurement methods employ optical or ultrasonic scattering of incident waves. Typically, a portion of a signal, scattered by the particle field, is collected and analyzed to estimate particle velocity. Of the many techniques available, optical techniques seem to be most widely used, and will be emphasized throughout this discussion. The methods used in extracting the information from the scattered or unscattered signal distinguish one particle velocity method from another. As a result of the information extraction and analysis methods, the resulting measurements acquire specific characteristics and limitations.

In some situations, it is useful to use some other particle-related signal source besides scattered light. One type of particle-related signal source, which has been used successfully, is particle or droplet fluorescence, in which a particle absorbs light of a certain wavelength and immediately emits light of a different, usually longer wavelength. Since the particle acts as a virtual light source for the emission wavelength, the variation of the signal with particle size, number density, and other characteristics is a little different than for signals from light scattering. Nevertheless since scattering is the most common approach in use, we will look at light scattering in some detail.

Among many different specific particle features which are usable for velocity measurement, there are two that seem to be the most widely used. One of these is the particle location, provided by viewing or recording an image of the particle at a sequence of well-known times. The second quite different particle characteristic that is frequently used is the frequency of light scattered from a particle of interest. A common requirement for either of these optical methods is the ability to collect a sufficient amount of light from a particle at some time and location. Among the issues of interest in enabling collection to occur are the object signal level (actually signal-to-noise) of the particle, and the resolution of the illumination and imaging system. Issues affecting the signal level will be described briefly in this section, while resolution issues will be presented in the following section.

For optical particle velocity methods, the intensity of light scattered from a particle is usually calculated from Lorenz–Mie theory (LMT) for scattering of light from spherical particles. Van de Hulst[3] presents the general LMT equations as well as some special cases (e.g. $d < \lambda$, $d = \lambda$, or $d > \lambda$, where d is the particle diameter), although all examples assume uniform illumination. Other work[4] has extended LMT to include nonuniform illumination, as might occur for illumination by a Gaussian cross-section intensity laser beam. To illustrate the effect of the important parameters in particle light scattering, the basic LMT will be briefly outlined. A more complete presentation is given in Van de Hulst.[3] The formulation used by Van de Hulst for a spherical particle gives the description of scattering as a generalized solution to a wave equation as shown in Equations (1) and (2).

$$\Delta\psi + k^2 m^2 \psi = 0 \tag{1}$$

$$\psi_{\ln} = \left\{ \begin{matrix} \cos l\phi \\ \sin l\phi \end{matrix} \right\} P_n^l(\cos\theta) z_n(mkr) \tag{2}$$

where $n \geq l \geq 0$, $k = 2\pi/\lambda$, r is the radial distance from the center of the particle, z is the propagation direction of the beam, $P_n^l(\cos\theta)$ is a Legendre polynomial, and $z_n(mkr)$ is a spherical Bessel function. The incident light is assumed to be linearly polarized.

The incident beams can be written (for the two solutions to the wave equation) as in Equations (3) and (4):

$$u = e^{iwt} \cos\phi \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} P_n^l(\cos\theta) j_n(kr) \tag{3}$$

$$v = e^{iwt} \sin\phi \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} P_n^l(\cos\theta) j_n(kr) \tag{4}$$

The field outside a sphere resulting from scattering from the sphere and the incident wave, is given by Equations (5) and (6)

$$u = e^{iwt} \cos\phi \sum -a_n (-i)^n \frac{2n+1}{n(n+1)} P_n^l(\cos\theta) h_n^{(2)}(kr) \tag{5}$$

$$v = e^{iwt} \sin\phi \sum -b_n (-i)^n \frac{2n+1}{n(n+1)} P_n^l(\cos\theta) h_n^{(2)}(kr) \tag{6}$$

where $h_n^{(2)}(kr)$ is a spherical Bessel function of the second kind.

The coefficients a_n and b_n can be determined using Riccati–Bessel functions as in Equations (7–9)

$$\begin{aligned} \psi_n(z) &= z j_n(z) \\ \chi_n(z) &= -z n_n(z) \\ \zeta_n(z) &= z h_n^{(2)}(z) \end{aligned} \tag{7}$$

$$a_n = \frac{\psi_n'(y)\psi_n(x) - m\psi_n(y)\psi_n'(x)}{\psi_n'(y)\zeta_n(x) - m\psi_n(y)\zeta_n'(x)} \tag{8}$$

$$b_n = \frac{m\psi_n'(y)\psi_n(x) - \psi_n(y)\psi_n'(x)}{m\psi_n'(y)\zeta_n(x) - \psi_n(y)\zeta_n'(x)} \tag{9}$$

where $x = ka = 2\pi a/\lambda$ and $y = mka$.

For a specific index of refraction particle, calculations of the relative scattered intensity as a function of particle size and scattering direction can be calculated from

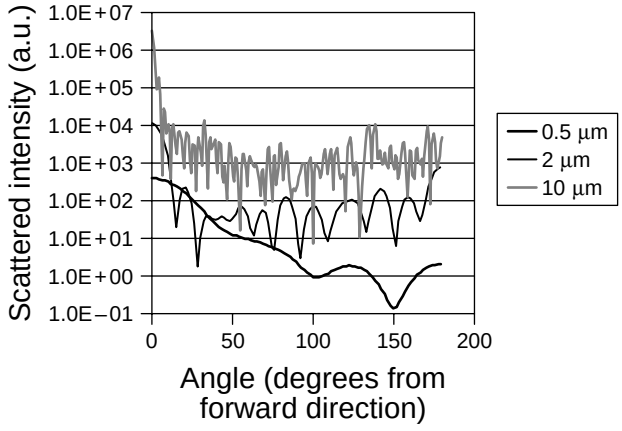

Figure 1 Calculated Mie scatter for 0.5, 2, and 10 μm spherical particles ($m = 1.5 + 0i$) as a function of collection angle. The calculation uses the methods of Bohren and Huffman.[5]

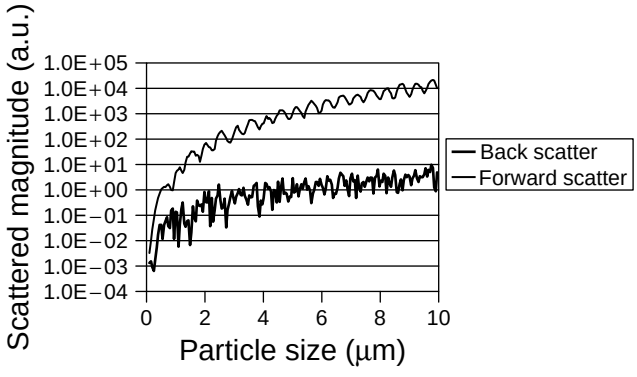

Figure 2 Calculated Mie scatter for forward scatter (collection from 0 to 4°) and back scatter (collection from 176 to 180°) as a function of particle size. The calculation uses the methods of Bohren and Huffman.[5]

Equations (5–9). Sample results for such calculations are shown in Figures 1 and 2 for several particle sizes at a range of scattering angles. Figure 1 shows the variation of scattered intensity for uniformly illuminated 0.5-μm, 2-μm, and 10-μm particles, as a function of collection angle, calculated using the method described by Bohren and Huffman.[5] Even with the apparently small difference in size between 0.5 μm and 2 μm, the scattered intensity values are very different at most of the scattering angles shown. At the forward scatter collection direction (0°) the 10-μm particle shows over 10^2 greater scattering intensity than the 2-μm particle, and about 10^4 greater than the 0.5-μm particle scattering intensity at this collection angle. In this calculation, individual angles are calculated separately; actual collected light signals may differ somewhat from the scattered intensities shown because the optical collection system collects over a finite solid angle and will integrate over some of the irregularities of the single particle scattering curves. Figure 2 shows the variation of scattered light intensity with particle size for "forward scatter" and "back scatter" collection angles. For these calculations, scattered intensity was integrated over 4° about 178° (176° to 180°) for forward scatter, and over 4° about 2° (0° to 4°) for back scatter. The changes in signal level with particle size for either collection configuration are significant. In addition, the difference between forward and back scatter for a given particle size is evident. The calculations shown neglect some real behavior likely to be present in scattering situations, like nonuniform illumination, but they give some order-of-magnitude estimates for the impact of particle size and collection angle on signal strength.

Experimental measurements of particle scattering efficiency for different particles are available.[6,7] An example of these measurements is shown in Figures 3 and 4. Figure 3 shows the minimum detectable particle diameter (defined as allowing >70% of the particles to be detected on film) in air as a function of the f number of the collection lens and the laser power used. Figure 4 shows the same basic results in a water flow. These measurements have focused on optical arrangements typical of LDV systems or particle imaging systems.[7] In general, the experimental results are consistent with the scattered light calculations for dilute particle systems in which single-particle scatter of nearly spherical particles dominates.

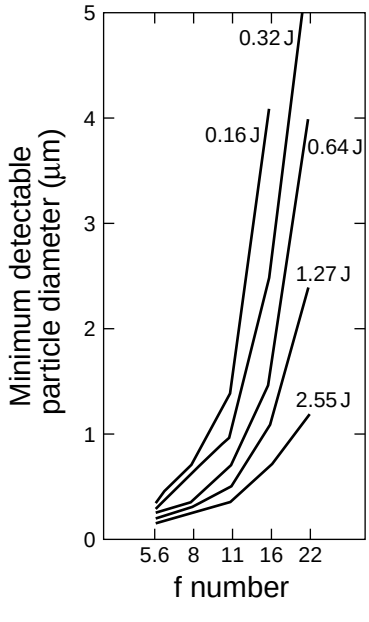

Figure 3 Experimental minimum detectable particle diameter in air as a function of collection system f number and illuminating laser power. (Figure from Adrian and Yao[7] and reproduced by permission of the Optical Society of America.)

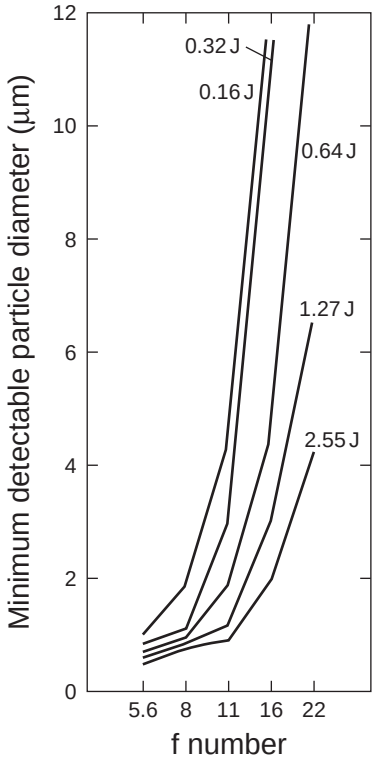

Figure 4 Experimental minimum detectable particle diameter in water as a function of collection system f number and illuminating laser power. (Figure from Adrian and Yao[7] and reproduced by permission of the Optical Society of America.)

For some dense particle fields, the single particle scatter predicted by Van de Hulst and others is more the exception than the rule. In these systems the actual collected signal may differ considerably from the predictions due to multiple scattering of photons from not one, but several particles. There does not seem to be the same kind of detailed analysis of the high number density particle fields as for the low number density fields. The issue has been, and continues to be one of great interest because so many practical particle or droplet fields are, in fact, quite dense. These past studies include analysis of dense fields themselves,[8] their effect on common measurement methods,[9] and possible resulting biases in measurements made in these kinds of fields.[10] More specific comments on the difficulties inherent in high number density particle fields will be mentioned in the discussions of detailed techniques. The problems associated with these high particle number density fields vary somewhat depending on the choice of measurement method.

The results of either calculations or measurement for single particle scattering provide guidance in selecting parameters for experiments. Scattering from large particles is clearly much stronger than from smaller particles; "forward" scatter directions result in much higher signal levels than for other scattering directions. In general, diffraction is not included in these calculations, and for forward scatter, diffraction may dominate the refraction and reflection signals modeled by the LMT calculations.

Depending on the measurement scheme selected, different specific attributes of the scattered particle signal may be sought. Understanding of the general behavior of particle scattering may be a great help in designing a successful experimental method. Under some circumstances, calculating the relative or absolute scattering in a given direction for a particle size range may be useful, using the equations given by Van de Hulst or some more recent work suggesting how to speed up the recursive calculations involved.[11] Some of this more recent work has also included nonuniform illumination (i.e. Gaussian beam profiles – a common occurrence in laser illumination of particles).[12,13]

3.2 Temporal and Spatial Resolution Requirements and Limitations

A second major issue for most particle velocity measurement systems is accommodating the temporal and spatial resolution required to accurately evaluate the physical system. The range of spatial and temporal resolutions available from measurements is closely tied to the ability to capture a scattered signal with adequate signal-to-noise in a limited time over a prescribed volume. Currently, hardware limitations have a significant effect on the range of temporal and spatial resolutions achievable. Hardware and data "bandwidth" limitations often cause spatial and temporal resolution capabilities to be intertwined, not independent of one another.

There are many common choices for spatial and temporal resolution. For example:

- low spatial and low temporal resolution – as for measurement of a steady homogeneous particle field; particle velocity measurement might be spatially averaged over a volume of 10 m × 10 m × 10 m with a temporal resolution of 1 s (update rate);
- moderate spatial resolution, low temporal resolution – as for the measurement of a steady particle flow with spatial values desired every 1 cm in a 1 m × 2 m × 10 m volume; temporal resolution might be about 1 s;
- high spatial and temporal resolution – as for measurement of an unsteady, inhomogeneous flow with spatial values desired every 1 mm in a 100 mm × 100 mm × 100 mm volume; time constant of the flow is expected to be about 50 μs, requiring data rates of 40 kHz (2 × 20 kHz to satisfy Nyquist criterion) or greater.

Table 3 Examples of temporal and spatial resolution capabilities for typical particle velocity measurement systems

Technique	Spatial sample volume size range	Temporal sample range
Doppler shift (LIDAR-type) methods	10 cm × 10 cm × 300 m (typ)	10 ns "snapshot" at 10 Hz or more (laser pulse rate) update rate
LDV	300 μm × 300 μm × 1 mm	10 μs at 200 kHz (signal processing rate) update rate
Direct imaging methods	1 mm × 1 mm × 1 mm	10 ns "snapshot" at 10 Hz or more (laser pulse rate or camera framing rate) update rate

Each of these measurements presents quite different spatial and temporal resolution requirements and will result in different selection of experimental parameters, if not in different experimental procedures altogether. Within any technique, a range of capabilities is possible with different experimental strategies or hardware. For the techniques most commonly used in particle velocity measurements, typical ranges are shown in Table 3.

In this table, "Spatial sample volume size" is intended to indicate the spatial volume over which measured particle values are averaged, producing one velocity vector for each specified volume of particles regardless of how many particles may be present in the volume. Clearly other values, both larger and smaller are possible with different selection of experimental parameters. The major point of Table 3 is to illustrate the differences among three common particle velocity measurement techniques in terms of spatial and temporal velocity value resolution. For example, a LDV (sometimes known as a laser Doppler anemometer or LDA) system usually has a relatively small sample volume and has the ability to sample particle velocities within that volume at quite high rates. However, LDV or LDA systems have no innate capability of providing multiple simultaneous samples at many different locations – although multiple systems can provide multiple sample volumes. In contrast, some direct imaging systems can sample at a great many locations within a flow simultaneously (via a two-dimensional image), but due to the difficulties of rapid, pulsed, illumination and rapid storage and transfer of images, most are not able to provide a temporal succession of images at a rate anywhere close to the sampling rate of the LDV system. Depending on the flow and type of information sought, either or both of these systems may produce useful information, but the choice of measurement system needs to be matched to the kind of information sought, bearing in mind the current limitations of each of the methods.

It should be noted that these methods need not be used exclusively, but can be used in complimentary ways. For example, a direct imaging method might be used to provide large-scale spatial data from a flow field and to provide data on spatial flow patterns in the flow field. An LDV system could be used to probe locations of particular interest to provide a temporal record of velocity values at a number of discrete locations.

New methods of particle velocity measurement are constantly being developed to provide capabilities not available with current systems or to adapt some current systems to new applications. For example, a recent combination of full-field particle imaging with Doppler shift measurement produces "global" Doppler velocimetry[14] which looks like a high spatial resolution version of the single detector LIDAR type systems.

One unmentioned, to this point, but quite important issue in comparing different approaches to particle velocity measurement is the inherent dynamic range of velocity which can be measured, the accuracy with which a velocity can be measured, and the uncertainty in the measurement. These attributes are also significant in selecting an approach for velocity measurement, but the actual values will vary with details of the application, including hardware and software used. Some discussion of these issues, and the ranges that may be possible, will appear in the sections specifically describing the most common implementations of the approaches outlined here.

3.3 Coupling Velocity and Particle Size Measurements

Throughout this discussion of particle velocity measurements, the possibility of obtaining simultaneous particle size and velocity measurements has been ignored. There are methods that combine one of the presented velocity measurement methods with a particle size measurement to produce simultaneous size and velocity measurement, and consequently size–velocity correlations. By far the most common commercial method for this kind of measurement is the phase Doppler method, here generically called PDI.[15,16] This method relies on an LDV type system for velocity measurement and a phase-shift method for size analysis. Since particle size measurements are covered in a different entry in this volume, they will not be presented here. The velocity measurement portion of these systems is very similar to the LDV systems used in velocity-only measurements, so discussion of velocity-only measurements may be extended to the velocity part of PDI size and velocity measurements fairly easily.

A second set of methods for combined size and velocity measurements involves some of the direct imaging methods discussed below, combined with particle size evaluation from the images. As for the PDI systems, the velocity portion of these measurements is similar to direct imaging, velocity-only measurements. A major exception to this similarity is the requirement for accurate particle image sizing. For velocity and sizing, the imaging system must produce accurately sized particle images. As this is not strictly necessary for velocity-only measurements, in most cases this presents a much stricter requirement on the imaging and recording system than velocity-only systems require. These additional requirements will be discussed in the portion of this volume that focuses on particle sizing.

3.4 Biases and Uncertainties in Particle Velocity Measurements

Any measurement is likely to have some bias and uncertainty. For the methods described in this short overview, many of these biases have been studied extensively. Rather than repeat those evaluations, readers will be referred to reference works on the subject.[2,17]

In general, biases for particle velocity measurement arise for a variety of reasons. The most notable biases are due to the limited ability of a sample of particle velocities to represent the entire ensemble of particle velocities in the flow, and the instrument characteristics that do not sample all particle sizes and velocities "fairly". For example, in many flows particles exhibit a distribution of velocity magnitudes and directions. For a sample of particles streaming through a fixed control volume over a fixed period of time, the probability of high-velocity particles passing through the volume is higher than the probability of low-velocity particles. The resulting ensemble of particle velocities may not accurately represent the true distribution of velocities as high-velocity particles may be overrepresented. This particular "velocity bias" is well known in the LDV field and some corrections for the bias have been suggested.[2,17]

Because of the complexity of obtaining any particle velocity values in some situations, it is tempting to assign any values obtained as representative of the entire particle field. Any method known to the author has some bias associated with it, whether that bias can be quantified and "corrected" or not. Users should try to recognize the range of biases likely in any application and, if it is not possible to reduce or correct for these characteristics, acknowledge their presence and likely influence on the resulting data and conclusions.

For approaches outlined in this article which rely on the ability to detect Doppler frequency shift in a light beam for measurement, the major difficulties arise in the separation of the shifted light frequency from the unshifted light frequency, and the association of the intensity of shifted light at a shift frequency with particle velocity. This topic has been presented in several papers discussing LIDAR applications and similar Doppler shift measurements.[18]

For LDV systems, measurement bias has also been a consideration almost since the technique was developed. While corrections for some of the known biases inherent in the method have been developed,[19] for other biases (like a nonrepresentative sample), no real correction is available without significant assumptions about the overall character of the particle and continuous phase flow.

Direct imaging approaches have also been examined in some cases for measurement biases, and a series of guidelines have been suggested to minimize the impact of some of these biases.[20,21] Even so, because of the complexity of the experimental technique, direct imaging approaches remain susceptible to biases in the measurement, which are not always recognized.

4 APPLICATIONS OF MEASUREMENT APPROACHES

4.1 Frequency Shift (Doppler) Methods

Frequency shift or Doppler shift methods rely on the frequency shift of light scattered from a particle for measurement of particle velocity. Successful implementation implies a well-characterized linewidth for the incident light, and some knowledge of the scattering characteristics of the particle or particles. Two implementations of Doppler shift methods will be outlined in this section. The first of these is a LIDAR system designed for measurement of atmospheric velocities by sampling particle (aerosol) or molecular velocities. The second implementation of a Doppler shift approach is LDV. This method is commonly used to measure particle or droplet velocities and particle size in a PDI system or separately in a velocity-only system such as an LDV or LDA system.

In Doppler shift measurement systems, a well-characterized, narrow linewidth light source (laser) is used to illuminate one or more particles. Scattered light is collected from the particles in the field. Equation (10) shows that the frequency of the scattered light is shifted by the vector velocity of the particles in a direction related to the illumination direction and the collection direction,

$$f_{\mathrm{D}} = \frac{\vec{u}}{\lambda}(\vec{r} - \vec{S}) \tag{10}$$

where is $\vec{r}$ a unit vector in the direction of the collection optics, and $\vec{S}$ is the Poynting vector of the incident light.

Equation (11) shows that for typical values of $\vec{u}$ (50 m s^{-1}) and λ (532 nm), particles traveling toward the incoming beam, and the scattered light collected back along the direction of illumination,

$$f_D = \frac{\vec{u}}{\lambda}(\vec{r} - \vec{S}) = \frac{50\,\text{m s}^{-1}}{532\,\text{nm}} = 93.9\,\text{MHz} \qquad (11)$$

This represents a large absolute frequency, but it is relatively small compared with the base frequency of light at 532 nm ($\sim 5.6 \times 10^{14}$ Hz) or 1.67×10^{-5}% of the mean frequency. This Doppler shift frequency is clearly very small relative to the light frequency, which can make the shift difficult to detect directly. LIDAR systems typically make this measurement, using several different kinds of set-ups and filters to measure the frequency of the Doppler shift. LDV systems usually use a dual-beam arrangement, discussed below, which makes the electronic signals corresponding to the Doppler shift easier to detect.

4.1.1 Light Detection and Ranging Systems

For measurement of particle velocity, frequently as a marker for wind velocity, LIDAR systems have been developed which are capable of providing velocity data at long ranges (>10 km) from the source/detector. In general, these systems aim to gather the velocity of particles in some region illuminated by a laser beam, and measure an average velocity for the particles in the region. Depending on the specific technique, the particles may be droplets, dust, or molecules themselves. Figure 5 shows a comparison of the relative Doppler shift frequency widths from aerosol (particle) scatter and from molecular scatter.

Most of these systems rely on a Doppler shift of the laser frequency and subsequent determination of the magnitude and direction (relative to the laser line) of the shift to determine particle velocity. These systems are usually divided into coherent[22] and incoherent[23–28] systems, depending on the strategy used for detecting the Doppler shift. A schematic of a typical LIDAR system for velocity measurement is shown in Figure 6. In this method, a laser, usually a pulsed laser like a Nd:YAG (neodymium:yttrium aluminum garnet) (10–1000 pps, $\lambda = 1064$ or 532 nm), is the light source. A pulse of light leaves the laser and optical system, pointing at a region of interest. Scattering from airborne particles, like dust or water droplets, scatters laser light back along the path of the laser. An optical system collects the scattered light and focuses it onto a detector. The optical filtering of the collected scattered light is the key to this method. The frequency of the light has been Doppler shifted and perhaps broadened by the motion of the scatterers. The frequency of the Doppler shift is directly related to the particle velocity (towards or away from the source and detector).

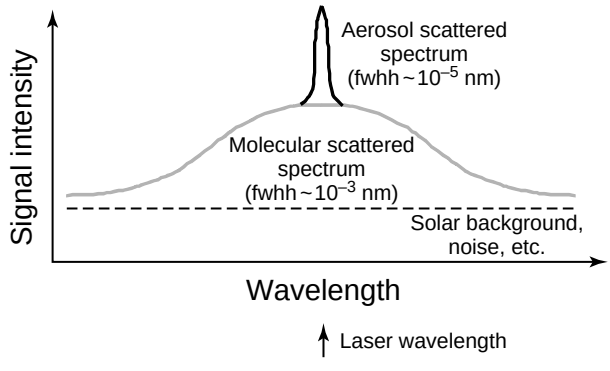

Figure 5 Characterization of the scattered signal intensity for laser light scattered from aerosols and molecules in air for back scatter LIDAR systems. (Figure from McGill and Spirhine[22] and reproduced by permission of SPIE.) fwhh, full width at half-height.

Detection of the Doppler shift in these systems differs depending on whether the system is designed as "coherent" or "incoherent". For the coherent case, where the collected light is presumably sufficiently coherent that optical interference is possible, optical heterodyne or homodyne processes using the backscattered light and a reference beam are used. Typical examples[22] use a CO_2 (10.6 μm) or Nd:YAG (1.064 μm) laser as a probe beam, and a stabilized version of the same laser frequency (homodyne) or slightly shifted (heterodyne) mixed with the scattered signal at the detector. The mixing of the two lasers at the detector allows the slightly shifted light returning from the sample volume to interfere with the unshifted light from the reference laser. The resulting signal at the detector will exhibit, among other frequencies, a frequency of oscillation corresponding to the difference between the return beam frequency and the reference beam frequency. The pulse delay from the outgoing beam to the return beam gives the range of the sample (distance out and back) and the frequency detected is related to the velocity parallel to the sample and return beams.

Incoherent systems rely on frequency separation of the laser frequency from the Doppler shift frequency by wavelength filtering. Two different approaches use either a high-resolution interferometer (Fabry-Perot) or a steep cut-off wavelength filter.[23–28] Normally colored glass filters or interference filters are not narrow enough, although argon ion laser sources (514 nm) have used an iodine cell to filter out the 514 nm line very effectively, producing a narrow bandwidth line filter which attenuates the laser line significantly.[29] These types of detection do not require mixing at the detector and can use less strictly controlled and coherent laser sources.

Other filtering strategies include "edge" filtering in which the cut-off edge of an optical filter is characterized

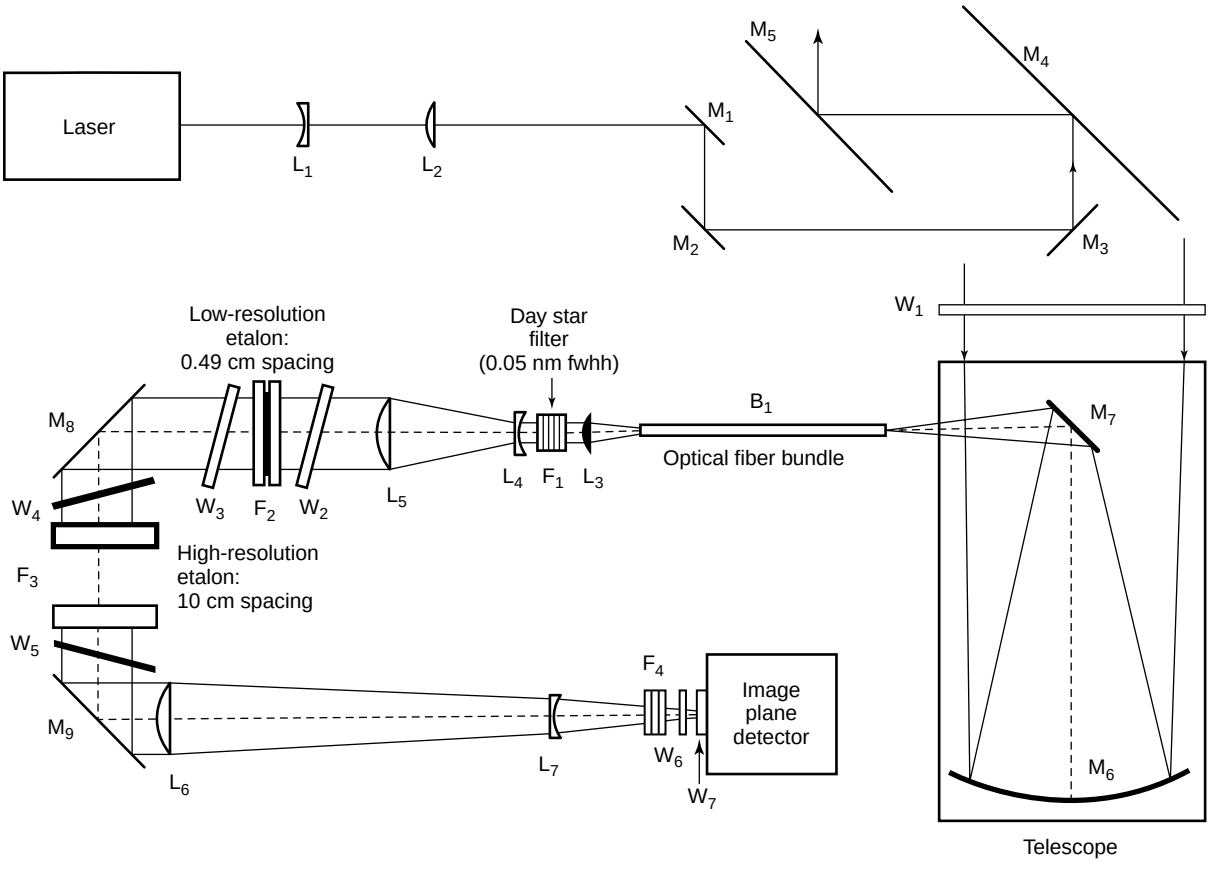

Figure 6 Schematic of incoherent LIDAR system for airborne particle velocity measurement. (Figure from Fischer et al.[25] and reproduced by permission of SPIE.)

and used in the return signal from the LIDAR system with the calibration information to produce intensity values proportional to frequency, and therefore particle velocity.[23,24] For each of these cases, the actual measurement is usually of particle velocity for airborne aerosols ($d_p > 1\,\mu m$). A sample of a system and results from an incoherent system are shown in Figures 6–8. In these figures, individual locations provided measurements of wind speed over time to produce a space–time profile (Figure 7) or a vertical profile (Figure 8).

Alternatively, LIDAR methods can be used with molecular scattering ($d_p < \lambda$ or $d_p < 1\,\mu m$) and get similar types of results, although the scattering intensities and linewidths are quite different for the case of molecular scattering as noted in Figure 4. An example of molecular scattering using an incoherent system is shown in the work by Abreu et al.[24] It may be noted that for molecular scattering, since the particle size is much less than the wavelength of light, the type of scattering is called Rayleigh scattering. One evident characteristic of Rayleigh scattering is that the scattering cross-section varies as $1/\lambda^4$. Shorter wavelength sources therefore appear to be advantageous if molecules are the scattering target, and many newer LIDAR systems use second harmonic generated wavelengths of an Nd:YAG laser (at $\lambda = 532\,nm$) or the third harmonic of an Nd:YAG

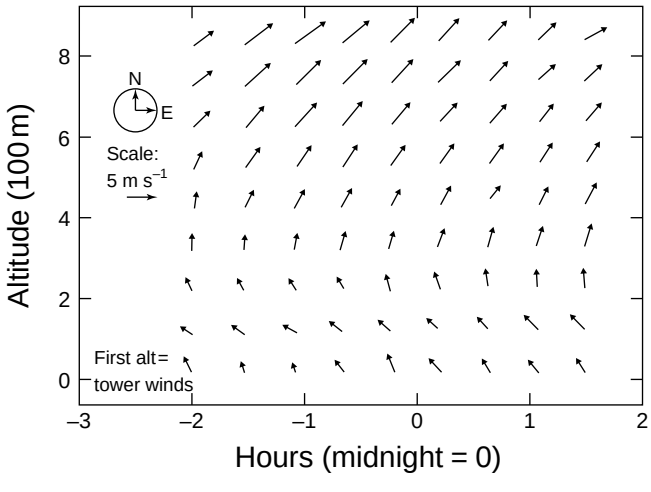

Figure 7 Sample results for LIDAR systems for airborne particle velocity measurement. (Figure from Abreu et al.[24] and reproduced by permission of the Optical Society of America.)

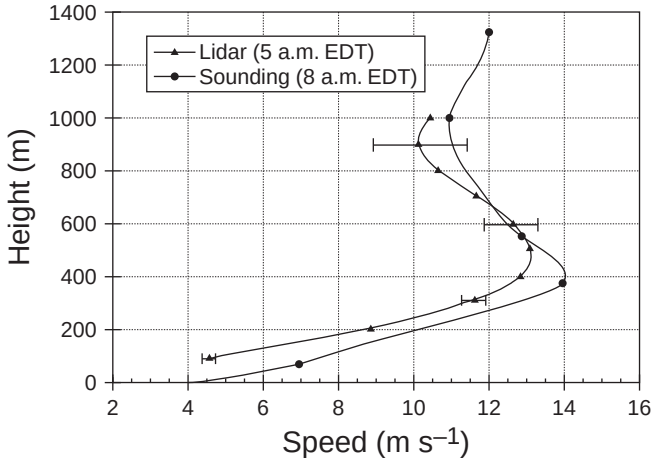

Figure 8 Sample results for LIDAR systems for airborne particle velocity measurement. (Figure from Abreu et al.[24] and reproduced by permission of the Optical Society of America.)

($\lambda = 355\,\text{nm}$) to take advantage of the larger scattering cross-section at these shorter wavelengths.

4.1.2 Laser Doppler Velocimetry

LDV or LDA is among the most widely used particle or droplet velocity measurement techniques for single point measurement of particles. LDV has found applications in a wide range of situations from liquid to gaseous flows, for relatively dilute to fairly dense particle number densities, and for a wide variety of other relevant parameters.

A typical simple LDV system is shown in Figure 9. A laser source (usually a continuous-wave helium–neon, argon ion, or laser diode) is split into two parallel beams. The beams are focused to a single point, where particles are expected to appear. As a particle passes through the crossing volume in the plane of the crossing beams, photons are scattered simultaneously from each beam. From the geometry of the system, it may be evident that the photons scattered from beam 1, are frequency shifted (from Equation 11) as shown in Equation (12)

$$f_{D1} = \frac{\vec{u}}{\lambda}(\vec{r}_1 - \vec{S}_1) \tag{12}$$

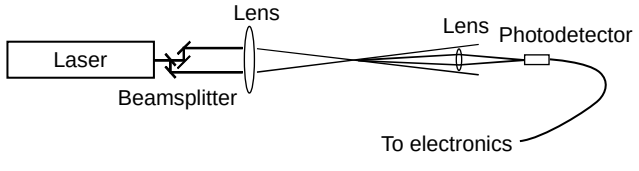

Figure 9 Schematic of a simple LDV system.

Similarly, the photons from beam 2, are scattered by the particle with a frequency given by Equation (13)

$$f_{D2} = \frac{\vec{u}}{\lambda}(\vec{r}_2 - \vec{S}_2) \tag{13}$$

Scattered photons are collected on to a single point detector, typically a photomultiplier tube (PMT), and the resulting signal is processed to determine the frequencies present in the signal. Signal mixing of the coherent scattered light (amplitude addition) at the photocathode of the PMT results in frequencies given in Equation (14)

$$f_{net} = (\nu + f_{D1}) \pm (\nu + f_{D2}) = 0$$

or

$$f_{net} = 2\nu + f_{D1} + f_{D2} \tag{14}$$

Since the base laser frequency ν is well beyond the high frequency limit of the PMT, only signals proportional to the Doppler shift, f_D fall within the bandwidth of the detector and are recognized. Using the previous definition of Doppler shift, Equation (15) gives

$$f_{D1} = \frac{\vec{u}}{\lambda}(\vec{r}_1 - \vec{S}_1)$$

and

$$f_{D2} = \frac{\vec{u}}{\lambda}(\vec{r}_2 - \vec{S}_2) \tag{15}$$

the detected signal should look like

$$f_D = f_{D1} + f_{D2} = -\frac{\vec{u}}{\lambda}(\vec{S}_2 - \vec{S}_1)$$
$$= \frac{2u_x \sin(\kappa/2)}{\lambda} = \frac{u_x}{\delta} \tag{16}$$

and

$$\delta = \frac{\lambda}{2\sin(\kappa/2)} \tag{17}$$

where u_x indicates the velocity component in the plane containing the two laser beams, in a direction normal to the bisector of κ.

Combining Equations (16) and (17) gives Equation (18)

$$u_x = \frac{f_D \lambda}{2\sin(\kappa/2)} \tag{18}$$

where κ is the angle between the two incoming beams.

An alternative explanation for the operation of an LDV system, which is often offered, revolves around the formation of interference fringes at the point where the laser beams cross. A particle passing through the crossing region will scatter photons in the region of the crossing. The interference pattern fringe spacing is determined by the geometry of the LDV system. For initial parallel beam

spacing d_s, with a lens focal length f, the crossing angle is given by Equation (19)

$$\tan\frac{\kappa}{2} = \frac{d_s}{2f} \approx \sin\frac{\kappa}{2} \tag{19}$$

where the assumption of small κ is made.

The resulting crossing volume, assuming Gaussian laser beam cross-sections is estimated to have a diameter $d = 4f\lambda/(\pi d_{e^{-2}})$, and a length of $l = d/\tan(\kappa/2)$. The fringe spacing in the crossing region is given by Equation (20)

$$d_f = \frac{\lambda}{2\sin(\kappa/2)} \tag{20}$$

Since the time for a particle to pass between fringes is known from the optical signal, velocity is calculated from Equation (21)

$$u_x = \frac{d_f}{t_f} = \frac{f_D\lambda}{2\sin(\kappa/2)} \tag{21}$$

The results of these two perspectives on LDV signals are the same, as seen in Equations (18) and (21). A typical LDV set-up might have conditions shown in Table 4.

Table 4 Typical LDV system conditions

LDV system characteristic	Value
Laser	Argon ion (514 nm)
Beam separation	50 mm
Beam diameter	1 mm
Focusing lens focal length	250 mm
Sample volume dimensions (diameter and length)	$d = 0.16$ mm, $l = 1.64$ mm

While there are many different configurations for LDV systems, two important ones will be mentioned. For other configurations, some specialized to certain applications, the reader is referred to the references at the end of this section.

The first major modification to the simple geometry of the system described above, is the addition of some frequency-shifting to one or both of the laser beams. The goal of this shifting is to allow determination of velocity direction. In the simple system described above, particles with velocity components normal to the fringe pattern in the crossing volume are all evaluated as ''positive'' velocities, because they present an oscillating frequency at the output detector. Doppler shifts in frequency can be positive or negative about the laser frequency. Since the laser frequency appears as a time invariant signal, because its frequency is beyond the range of detectors, the Doppler shift frequency appears as positive or negative frequency shifts about the signal. Under these circumstances, positive and negative shifts appear the same. By frequency shifting one or both of the beams, an apparent motion of the crossing volume fringes at the frequency difference between the two beams superimposes a ''bias'' frequency on the fringe frequency from the scattered particles. The result is shown in Equation (22)

$$f_{deff} = f_d + f_s \tag{22}$$

where f_s is the frequency shift imposed on the beams. For $f_s > f_d$ the resulting f_{deff} can now indicate positive Doppler shift at frequencies above f_s and negative Doppler shift at frequencies below f_s. This frequency shift is usually accomplished using an acousto-optic modulator (also called a Bragg cell). A typical shifted and high pass

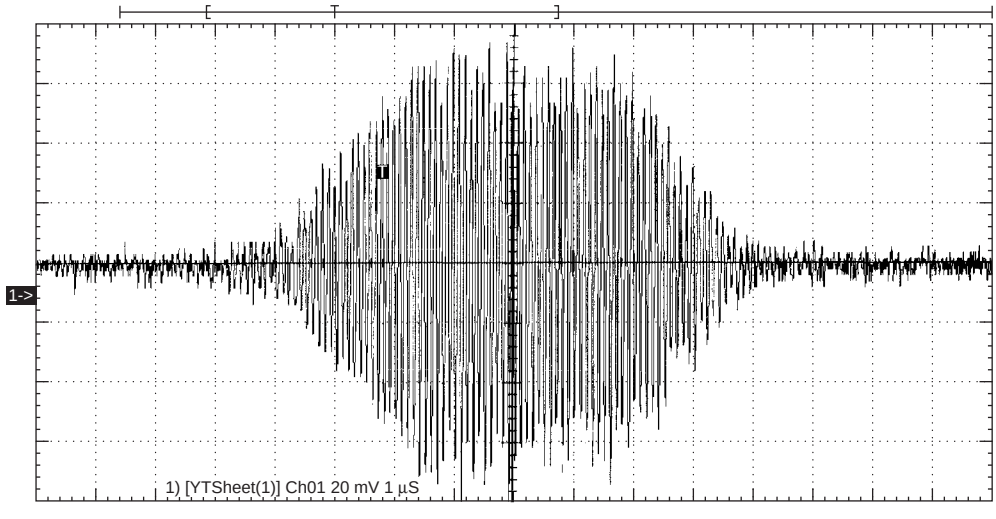

Figure 10 Typical LDV burst (with frequency shifting). This is a direct copy of an oscilloscope trace with a horizontal time base of 1 µs per division and a vertical gain of 20 mV per division.

filtered LDV signal is shown in Figure 10. This figure is taken directly from a digital oscilloscope. The vertical gain was set to 20 mV per division, and the sweep speed was 1 μs per division.

The second major change in configuration from the simple system outlined above is the use of fiber optics for delivery of the laser beams and transmission of the collected light from near the measurement volume to the detector. The use of optical fibers allows the laser and detector system to be mounted remotely from the experiment, with only a small number of optical elements near the actual sampling volume. With careful alignment, these systems can be made to operate as effectively as conventional delivery systems, and provide some system mounting and motion flexibility not easily available for conventional systems.[2,30]

More recently, many other innovations in LDV configurations have been successfully demonstrated, including laser diodes for the illumination source, solid-state detectors for collection, and advanced electronics for signal detection and analysis.

Analyzing the scattered light from particles in the crossing volume quickly and accurately extracting the frequency information is a challenging problem in signal processing and analysis. For low noise signals (as in Figure 9), estimating the frequency of the high-frequency portion to calculate velocity is relatively straightforward. Historically, this has been handled by special purpose electronics.

In more challenging environments, such as ones with significant background noise from laser light scattering off windows, surfaces, etc. these systems struggle to find particle signals with peaks sufficiently above a filtered noise floor to provide accurate velocity values. Newer processors rely on fast Fourier transforms (FFTs) or digital Fourier transforms, or related autocorrelation methods to extract burst frequencies in the presence of higher levels of background noise. These systems are all commercially available, and details of their operation are left to their respective manufacturers.

The LDV system described provides one component of the particle velocity vector. To obtain the other two components, additional optics may be included essentially replicating the one-dimensional system in three dimensions. Usually, to allow for simultaneous measurement of two or three velocity components, different laser wavelengths are used for each component allowing easy separation of the different scattering signals.

One common application for LDV systems or other velocity measurement systems is to make useful measurements of particle turbulence characteristics. The particles may be serving either as a marker for the continuous flow or may be of interest in their own right. A velocity measurement system like an LDV system provides a series of "instantaneous" velocity samples over some collection time window. While turbulence data are contained in the instantaneous samples, the volume and character of this kind of data makes it awkward to use in describing the flow in a concise way. As a result, instantaneous velocity values are often decomposed using "Reynolds averaging" as in Equation (23)

$$u_i = \bar{u}_i + u_i' \tag{23}$$

where u_i is the instantaneous velocity, $\bar{u}_i$ is an average, either time average, as in Equation (24),

$$\bar{u}_i = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} u_i \, \mathrm{d}t \tag{24}$$

or an ensemble average, as in Equation (25)

$$\bar{u}_i = \frac{1}{N} \sum_{j=1}^{N} u_{ij} \tag{25}$$

Resulting data are often given as mean velocities ($\bar{u}_i$) and RMS ($\sqrt{\overline{u_i'^2}}$) values, normalized by an appropriate value.

A common extension to these basic velocity moments is the calculation of a timescale, called the "integral scale", which is considered to be an indication of the timescales for large-scale turbulence in the flow. The integral timescale, I_t, shown in Equation (27), is calculated from the autocorrelation coefficient, $\rho(\tau)$, shown in Equation (26)

$$\rho(\tau) = \frac{\overline{u_i(t)u_i(t+\tau)}}{\overline{u_i(t)^2}} \tag{26}$$

$$I_t = \int_0^{\infty} \rho(\tau) \, \mathrm{d}\tau \tag{27}$$

Results of mean velocities, RMS velocities, and integral timescales can be used for comparison with other flow conditions or other flows. It should be noted that the simple decomposition suggested by the Reynolds decomposition of Equation (22) is most useful for turbulent flows in which the mean flow is presumed to be steady, the flow is incompressible, and a "typical" turbulent energy cascade might be expected. For flows in which any of these assumptions is not valid, other more complex ways of defining a mean flow and the "turbulent" portion of the flow (and related turbulent characteristics) may be needed.[31]

A sample of LDV results for a turbulent flow downstream of a rearward-facing step is shown in Figures 11, 12, and 13.[33] The flow is a steady water channel flow over a step (an abrupt expansion). This is a common test problem for instrumentation and modeling systems as it involves a largely two-dimensional shear

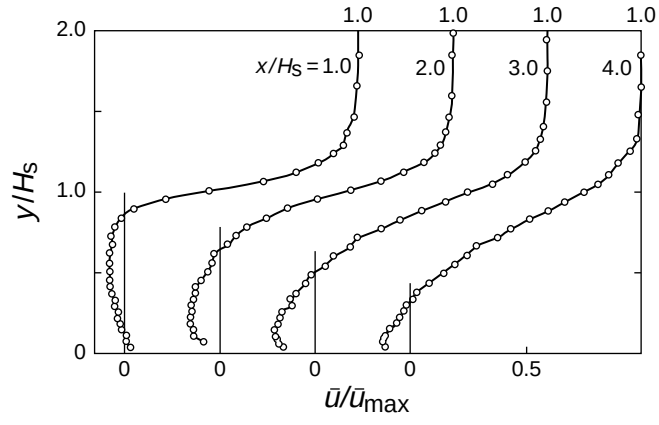

Figure 11 Sample LDV results for water flow over a rearward-facing step. This figure shows mean axial velocity profiles at various distances downstream of the step. (Figure from Etheridge and Kemp[31] and reprinted by permission of Cambridge University Press.)

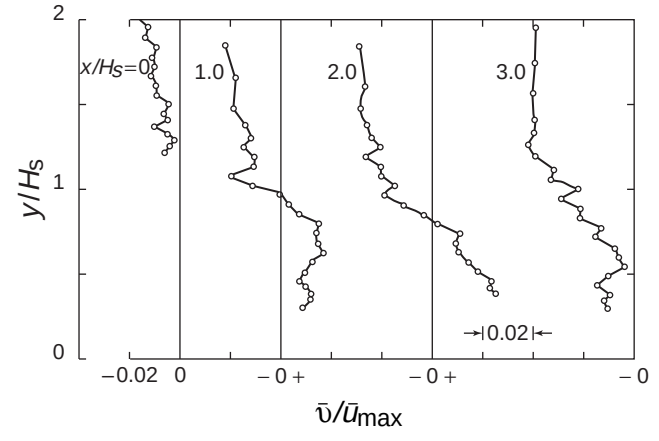

Figure 12 Sample LDV results for water flow over a rearward-facing step. This figure shows mean transverse velocity profiles at various distances downstream of the step. (Figure from Etheridge and Kemp[31] and reprinted by permission of Cambridge University Press.)

flow with separation and recirculation downstream of the step. Figure 11 shows several mean (time average) profiles at different locations downstream of the step, where x is the axial dimension, y is the vertical dimension, and H_s is the step height. In this experiment, cross-stream velocities were also measured separately by rotating the LDV system, and sample mean results are shown in Figure 12. Figure 13 shows normalized RMS values for the turbulence for both x and y velocity components.

There are many other examples of LDV results in journal literature. Since application details may vary depending on specific application, further examples will be left to journals in the respective field of interest.

Before leaving this overview of LDV systems, some cautions regarding the constraints and biases in LDV measurements should be mentioned. As may be surmised, the LDV system expects only one particle to be in the crossing volume at any one time. If more than one particle is present, the actual frequency sampled is not likely to be related to the particle velocities in the form of Equation (21) unless the particles happen to have the same vector velocity in the plane of the illumination beams normal to the center line of the beams. Most signal processors go to great lengths to identify and discard data from multiple particles in the control volume. Clearly this presents a problem for very dense particle fields in which the likelihood of multiple scatterers in the control volume is high.

Since the actual time of appearance of a particle in the measurement volume is somewhat random, evaluation of LDV results is usually approached statistically. The underlying assumption (hope?) is that the particles which did pass through the control volume, and whose signals passed all the processor tests, are representative of all "similar" particles in the flow field.

For LDV measurement systems, there are several major expected biases in the velocity signal. One of these arises from relative oversampling of the higher velocity particles, since they are more likely to pass through the control volume than lower velocity particles in a given sample time. A second bias is size biasing, in that larger particles can have an effectively larger detectable control volume than do smaller particles because of their larger scattering signal even in the weaker portions of the Gaussian probe beams. Note that simultaneous measurement of particle size and velocity allows users to access velocity–size correlations and assess the significance of this size bias. A third bias is velocity average biasing, in that standard time averages of particle velocities may overrepresent some samples because of the irregular arrival times (hence integration times) for particles in the system. There are methods to minimize the effects of these biases, and interested readers should consult appropriate references to understand the implications of these biases for the experiments of interest.[2,17]

4.2 Direct Imaging Methods

Probably the most straightforward way of measuring particle velocity is to record an image of the particle of interest at a known time, and record a second image at some later time. An estimate of the displacement of the particle between the two images is available from the images, and combined with the known time between the two recordings, an average velocity can be calculated. Implementing this method in realistic particle or droplet

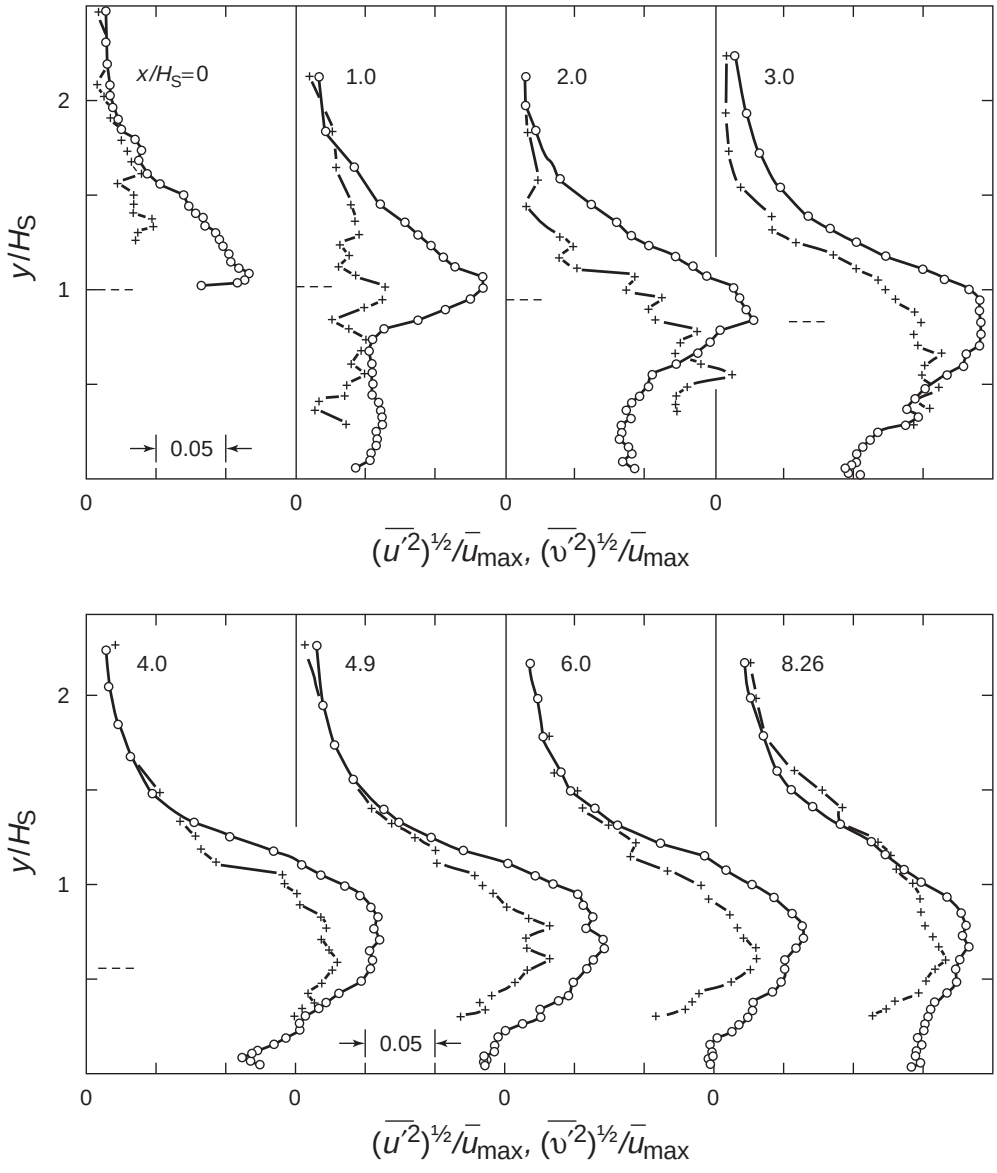

Figure 13 Sample LDV results for water flow over a rearward-facing step. This figure shows fluctuating axial and transverse velocity profiles at various distances downstream of the step. (Figure from Etheridge and Kemp[31] and reprinted by permission of Cambridge University Press.)

flows is frequently far from straightforward. Depending on many of the parameters of the flow and particles, different strategies may be needed to successfully acquire velocity data. In the end, direct imaging of particles for velocity measurements is not merely one technique, but a whole family of techniques for the range of situations of interest.

Many of the techniques used for direct imaging are relatively new, and several recent review papers provide a thorough background and reference list of recent fundamental and applied work in this field.[10,34–36] In this section, some of the different approaches to direct imaging will be described and some of the choices that users need to make in selecting a specific technique will be indicated. In many cases, constraints on the techniques available for direct imaging are related to current hardware and software limitations. As new hardware and software with greater capabilities become available, some of the current limitations on direct imaging will change and the optimal system for a given application will likely change, too.

In the following paragraphs, I will outline some of the major features of a particle field which impact how a direct imaging experiment might be performed, and some of the

choices available to experimenters which also impact the type and quality of the results.

4.2.1 Particle Visibility

Recording a particle image requires an adequate number of photons to be scattered from a particle into the collection optics and system detector. The scattered intensity, as noted in an earlier section, is a function of particle size, index of refraction, shape, and collection direction and solid angle, among other things. As was discussed in the previous section, for particles near or greater than the wavelength of light to be used, LMT provides useful estimates of the spatial variation of scattered light intensity for various particle sizes and indexes for spherical particles. With a few exceptions, there are not many predictive results on scatter for nonspherical particles.

The range of particle visibilities is very large; for near-forward scatter directions, Mie scatter intensities vary approximately as d^2 (recall Figures 1–3, for example). In many polydisperse particle fields, a size distribution of an order of magnitude in particle diameter is likely (say 10–100 μm), giving a range of scattering intensities of at least 1 : 100. This range is not particularly large for a size range, but may present some difficulties for image recording for limited dynamic range detectors like 8-bit (256 gray-level) charge-coupled device (CCD) cameras. For velocity measurement, accurate image size may not be essential or necessary as long as the image size is constant and relatively small. A wide range of particle visibilities presents the likelihood of size biasing the velocity statistics, in that the more visible larger particles are more likely to be seen and therefore will represent a higher proportion of the distribution of velocity samples than they represent in the actual particle system. If larger particles have different velocities from smaller particles (as is likely) the velocity mean and standard deviation will be affected by the sampling.

4.2.2 Particle Size

While particle size affects visibility, it also affects the minimum displacement that can be measured accurately for most of the analysis schemes, as this minimum displacement is some significant fraction of the particle size. In addition, for situations in which the particle is seeded into the flow as a marker, the size affects the ability of the particle to follow the continuous phase of the flow and provide an appropriate velocity value. This issue is the same one commonly considered for LDV measurements in flows.

4.2.3 Particle Density

Depending on the number of particles in an imaging volume and their distinguishability, different methods may be required to extract velocity data from a set of images. Particle density is clearly a function of the particle loading in the region of interest (i.e. particles cm^{-3}) but for imaging, the number of particles in the image volume may be more relevant. For dilute particle fields, individual particles can be identified and ''tracked'' from image to image, making a particle tracking velocimetry (PTV) system. For dense particle fields, obtaining an undistorted image of the particle region of interest may be difficult. In addition, as particles are unlikely to be distinguishable from one another in a sequence of images, an area average analysis for the particles falling in a region of the flow is usually used. Dense particle fields may cause significant attenuation of the illumination beam, leading to stronger scattering signals from some regions of the field than from others. Imaging through dense particle fields can also be somewhat problematic.[37]

4.2.4 Particle Image Density

For imaging of particles, we normally assume that there will be little or no particle image overlap; actual experiments may impose a more severe requirement in that identifying particles in different images and ''matching'' the same particle between two images may be difficult if the particle image density is high. Actual particle density as well as the total particle volume collected in the image affect particle image density.

4.2.5 Particle Velocity

Since particle displacements are the actual measurement, expected particle velocities and the range of velocities will affect the choice of experimental parameters, like the time between images. This is because displacements that are quite small (less than a particle diameter) or quite large (>1/3 the ''interrogation spot size'') may cause problems in measuring displacement.

4.2.6 Imaging System Characteristics

These characteristics include system magnification, resolution, aberrations, field size, working distance, and framing rate among others. Clearly, many of these characteristics are interrelated, but they affect the image quality, and the apparent particle size, displacement, and dynamic range. Commercial 35 mm format or medium format lenses of very high quality and moderate price are available for the optical imaging system. The specific characteristics of the selected imaging system along

with the detector characteristics will determine the system characteristics and aberrations. Typical choices of detectors range from conventional 35 mm cameras using high-contrast silver-halide film, to solid-state CCD detectors. Very high spatial resolution and/or large dynamic range systems typically operate at relatively slow speeds (imaging rates), while some lower resolution systems can operate at high imaging rates. These rates can range from less than one image s^{-1} (cooled 1000 pixel by 1000 pixel or more, by 16 bit, CCD slow scan) to 4500 images s^{-1} (256 pixel by 256 pixel CCD) to over 200 000 frames s^{-1} (drum camera with effective 16-mm image size on film). As new hardware is developed, particularly in the solid-state sensor area (CCD cameras), higher resolutions (e.g. 4 K pixel by 4 K pixel is now available), with large dynamic ranges (12 or 16 bit), at faster framing rates (>30 fps) will become available. Because of the trade-offs in the imaging system, clear system requirements are needed before specific hardware is selected. Conversely, access to state-of-the-art hardware may make the trade-offs in temporal and spatial resolution less limiting.

4.2.7 Illumination System Characteristics

Many issues regarding spatial and temporal resolution in the direct imaging system are affected by choice of illumination system and the associated illumination optics. A common choice for high-velocity particle imaging (say particle velocities $> 10\,m\,s^{-1}$) is a pulsed laser, like a Nd:YAG (10 ns pulse length, 10–1000 Hz, $\lambda = 532$ nm or 355 nm) which is formed into a thin ($\sim$500 μm) sheet to illuminate one plane of particles. For many experiments, the desired time between pulses is frequently so short that the standard repetition rate for the laser is too slow. This may require more than one laser synchronized together, or some clever modifications to the Q-switch circuitry to obtain two or more pulses with a time between pulses of the order of 10–100 μs. For lower-velocity particles, a continuous-wave laser source may be satisfactory, since continuous-wave lasers can be shuttered mechanically or acousto-optically with microsecond precision.

Other requirements for the illumination source may include specific wavelengths for fluorescent particles (pumping an absorption band for the particle, then imaging a fluorescence wavelength to improve particle visibility),[38] different pulse lengths (e.g. longer pulses for lower power density to conserve optics and windows), and polarization selection.

4.2.8 Interrogation or Analysis System

Once a pair or more of particle field images have been captured, evaluating the particle displacement between the two images can be done a variety of ways. As there is extensive literature on many aspects of this ''interrogation'' problem, readers are referred to the references, and only outlines of some strategies will be suggested here.

For relatively low particle density images, PTV analysis schemes can be used.[36,39] These schemes usually rely on large particle-to-particle spacing allowing individual particles to be easily identified in each image and from image to image. Finding the particle velocity is a matter of tracking the particle images and calculating the displacements from image to image of the identified, ''paired'' particles. The result of this kind of interrogation of a set of images is a set of vectors for each particle pair, indicating velocity direction and magnitude. Since the method requires that there be relatively few particles in each image, only a few vectors result from each set of images. To build a spatially ''complete'' set of vectors that covers a significant portion of a volume of interest, a large number of replicates of the experiment may be required.

Data analysis methods for PTV focus on following particles from image to image, and ensuring that a single particle is ''matched'' with its corresponding image in other image frames. Two- and three-dimensional velocity vectors can be obtained with these methods.

For relatively high particle density images, PTV is difficult since the particles in each image cannot be easily ''paired'' with their images (i.e. all particles look alike). To overcome this identification problem, the solution sought is changed slightly. Rather than seeking a vector for each particle, we may be content to seek an average (or most probable) velocity or particle displacement in a small region of the flow. There are two major approaches to finding this local average particle displacement. They differ primarily in terms of what kind of image information is available. One approach, known as an autocorrelation approach, begins with both particle images (separated by some Δt) superimposed, as might be the case for a double exposure on film or a solid-state detector. The second approach, known as the cross-correlation approach, begins with independent first and second images, as for consecutive images of the particle field.

Briefly, the autocorrelation approach selects a small region of the double exposure image to interrogate (interrogation region). Typical interrogation region sizes might be 1 mm in diameter or 1 mm by 1 mm square. The result of this data analysis will be an average particle velocity and direction over this interrogation spot size for all particles that lie within the spot and have a second image also lying within the spot. The autocorrelation of the interrogation spot with itself should provide information on the most probable displacement

of particles in the region. The strongest correlation should be all particles with themselves, at zero displacement. The next strongest correlation of particle images in the spot should, we expect, correspond to the most frequent spacing between particles. Assuming the initial particle spacing in either of the two exposures is random, the most probable spacing (which is not random) should be the displacement that all particles underwent between the exposures of the first image and the second image. In general, the autocorrelation is given by Equation (28)

$$R(\vec{s}) = \int I(\vec{x})I(\vec{x}+\vec{s})\,\mathrm{d}\vec{x} \tag{28}$$

While this can be analytically calculated for each interrogation spot, it is usually more convenient to invoke the fact that for a function $f(x)$ with Fourier transform $F(s)$, $\int_{-\infty}^{\infty} f^*(x)f(x+u)\,\mathrm{d}x$ has a Fourier transform of $|F(s)|^2$.[40] The utility of this theorem is that while calculating autocorrelations for many interrogation spots in an image may be rather slow, calculating Fourier transforms using the FFT may be significantly faster. As a result, most autocorrelation schemes for particle image velocimetry (PIV) calculate the autocorrelation using the following sequence:

Select interrogation spot

$\Rightarrow$ "snap" image of spot

$\Rightarrow$ two-dimensional FFT of image

$\Rightarrow |F_{2D}(s)|^2$

$\Rightarrow F_{2D}^{-1}|F_{2D}(s)|^2$(inverse two-dimensional FFT)

$\Rightarrow R(\vec{s})$

Once the two-dimensional autocorrelation is acquired, the largest peak outside of the $\vec{s} \approx 0$ region should give the value corresponding to the average particle displacement. Suggestions for optimizing imaging, interrogation, and "peak finding" are given by Keane and Adrian.[20]

Autocorrelations have been widely used in PIV. They are typically calculated digitally,[41,42] but with an appropriate set-up may be "calculated" using an optical correlator.[43] A major drawback of autocorrelation methods is that while velocity magnitude and direction can be obtained, the sense (positive or negative value of velocity) of the velocity cannot, because with double exposure images, it is not known which set of particle images corresponds to the "first" image, and which the "second". When the two images are recorded separately, this information is available and can be used with cross-correlation to remove the 180° ambiguity of direction. The image separation may be accomplished by recording each of two images on separate pieces of film or separate solid-state detectors, or by tagging one set of images. The most common tagging method is by color. For example, if two different color pulsed lasers are available, they may be synchronized such that one provides the illumination for the first set of particle images, and the second laser provides the illumination for the second set. If the images are recorded on a color sensitive detector or film, they may be later separated into a "first" and "second" image.[44]

Performing a cross-correlation of particle images in a region of the total image with a similar region in the second image should produce a maximum cross-correlation value when the first image subset and the second subset are offset by the most probable displacement of the particles in the region. The cross-correlation typically used is defined[45] in Equation (29)

$$\phi_{fg}(m,n) = \sum_{k=k_{\min}}^{k_{\max}} \sum_{l=l_{\min}}^{l_{\max}} f(k,l)g(k+m,l+n) \tag{29}$$

where $f(k,l)$ represents the intensity values at pixels locations x_k and y_l in the first of the two images, and $g(k,l)$ represents the intensity values at pixels locations x_k and y_l in the second of the two images. The correlation essentially searches for the offset (m,n) of $g(k,l)$ with respect to $f(k,l)$ which produces the largest sum over the region. The offset corresponding to this peak is taken to be the average displacement of particle images in the region.

This process of cross-correlation continues with successive subregions of the image, sometimes with overlap of the image subsets, until the entire image has been sampled. As with autocorrelation methods, this calculation is usually performed using FFTs because the numerics are much faster than direct calculation of the cross-correlation. The sequence of operations is similar to that for autocorrelation, except two-dimensional-FFTs of each of $f(k,l)$ and $g(k,l)$ are needed at the beginning to produce the cross-correlation in Equation (28) after the last two-dimensional inverse fast Fourier transform (IFFT).

4.2.9 Particle Image Velocimetry Example

One example of direct imaging systems using particles as markers for flows will be shown. This example relates to the flow inside an internal combustion engine which is expected to have relatively high mean flow velocities ($50\,\mathrm{m\,s^{-1}}$) and high turbulence levels (>50%).[46] The spatial scales of interest range from the full size of the flow chamber (100 mm × 100 mm × 100 mm) to smaller scales of turbulence (<1 mm). The mean flows are unsteady, varying at frequencies of up to about 1 kHz, with small-scale variations at 10 kHz or greater. Additionally, the

flow is not precisely repeatable (cycle-to-cycle variation), so replicates from different engine cycles have neither the same mean value nor the same turbulent statistics. The interest is to acquire velocity data with a reasonably high spatial resolution (1 mm × 1 mm × 1 mm) and temporal resolution (100 μs).

This application of PIV methods to engine flows is in general similar in concept to the previous PIV work by Adrian[10] and engine work by Reuss et al.[41] and Reeves et al.[42] but differs in implementation and detail. The PIV method implemented here includes a data recording step with a double pulsed Nd:YAG laser as a light source, and an analysis step using an optical Fourier transform method to estimate particle displacement.

The particular flow of interest in this example was the gas flow near the periphery of a spray plume from a fuel injector. The spray chamber is a cylindrical chamber, about 150 mm in inner diameter, with the injector mounted at one end of the cylinder. The opposite end of the cylinder had a 100-mm fused silica window, and a small window was available at the periphery for introducing the laser light. The chamber is operated at room temperature, and is typically pressurized with argon to about 20 bar. Two frequency doubled Q-switched Nd:YAG lasers were used as illumination sources for the particle image recording. Balanced double pulses were obtained with a 6 ns pulse width and 20 μs pulse separation time. The laser beam was formed into a sheet and passed into the test section through the side so that the flow region of interest could be photographed through the large front window. The sheet thickness was about 0.5 mm and the beam height was about 25 mm. A 35 mm format camera was used to take the PIV photographs on Kodak Technical Pan 2415 film. TiO_2 was mixed with air in a particle generator and was mixed into the chamber with the argon.

A sample of the particle images and resulting auto-correlation based velocity vectors are shown in Figure 14. In this figure, the spray plume is clearly visible. Near the spray plume, small droplets may mix with the seed particles and the velocity in those regions is likely to be an average of the small droplet velocity and small seed particle velocities.

For turbulent flows, PIV or other direct imaging methods can provide detailed time-series data and resulting statistics, but with considerable difficulty. To produce a time series of velocity vectors at a point in a flow requires a statistically significant number of realizations (image pairs) at that location in the flow. Direct imaging systems do provide spatial information in each image, so spatial turbulence information is readily available, as opposed to temporal turbulent data single point systems like LDV provide. With spatially

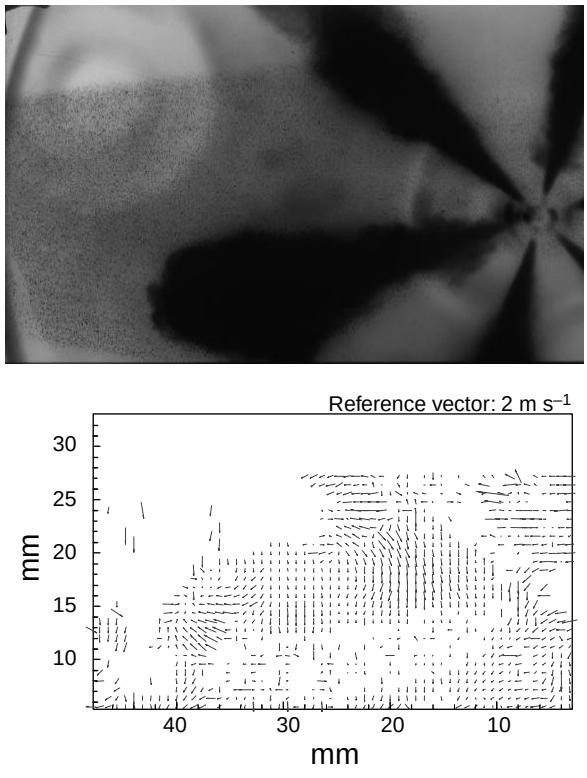

Figure 14 Sample PIV image from TiO_2 seeded gas in a closed pressure vessel. The flow is generated by a high-pressure fuel spray (seen in photograph). Velocities of interest are of entrained gas flow. A vector plot shows the measured particle (gas) flow velocities around the liquid jet.[46]

varying velocity vectors, instantaneous, spatially varying vorticity and strain rate values can be calculated.[41] Using the spatial autocorrelation, in analogy to the temporal autocorrelation of Equations (25) and (26), an integral length scale could be calculated for the region of velocity vectors. For replicates of an experiment, these spatially varying turbulent quantities can be compiled into spatial and temporal values, using spatial and temporal averaging as appropriate to the application.

4.3 Recent Developments in Particle Image Velocimetry

Recently, direct imaging methods, and, in particular, PIV methods have expanded into three-dimensional velocity vector measurements. These systems usually employ two or more imaging systems to produce stereoscopic images or in a quite different approach, use holographic recording to allow for three-dimensional reconstruction of the image field.

The stereoscopic methods rely on the same basic imaging approaches outlined for the two-dimensional approaches discussed above, with some additional

complexity in imaging system set-up and subsequent data analysis to extract three-dimensional vectors from nonorthogonal data.[47]

Holographic recording of particle images for reconstruction and evaluation of three-dimensional particle velocities is more complex than the stereoscopic approaches, but promises more than a single plane of three-dimensional vectors. These systems are still under development in research laboratories. Interested readers may find some discussion of holographic PIV in recent literature.[48–50]

5 SUMMARY

The field of particle velocity measurement is very large, with a wide range of techniques available. In many cases, the most useful techniques turn out to be optical methods because of their accuracy, nonintrusive character, and ability to be applied remotely to a flow. In this article several characteristics of the particles (distributed phase) and the background fluid (continuous phase), which may be relevant in selecting a particle measurement method, have been outlined. Three of the most commonly used particle velocity measurement methods have been discussed. Readers interested in applying any of these are encouraged to take this section as an introduction and to dig deeper into the method and applications specific to their field of interest. Hopefully the introduction provided here and the short list of references will provide a good starting point from which potential users can further explore these and other techniques.

ACKNOWLEDGMENTS

The writing of this article and the results generated at the University of Wisconsin-Madison Engine Research Center were supported by the Army Research Office under Grant No. DAAH04-94-G-0328.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
FFT	Fast Fourier Transform
fwhh	Full Width at Half-height
IFFT	Inverse Fast Fourier Transform
LDA	Laser Doppler Anemometer
LDV	Laser Doppler Velocimetry
LIDAR	Light Detection and Ranging
LMT	Lorenz–Mie Theory
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
PDI	Phase Doppler Interferometry
PIV	Particle Image Velocimetry
PMT	Photomultiplier Tube
PTV	Particle Tracking Velocimetry
RMS	Root Mean Square

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Diffraction in Particle Size Analysis • Light Scattering, Classical: Size and Size Distribution Characterization • Optical Particle Counting • Turbidimetry in Particle Size Analysis • Ultrasonic Measurements in Particle Size Analysis

REFERENCES

1. W.D. Bachalo, 'Experimental Methods in Multiphase Flows', *Int. J. Multiphase Flow*, **20**, 261–295 (1994).
2. R.J. Adrian, 'Laser Velocimetry' in *Fluid Mechanics Measurements*, 2nd edition, ed. R.J. Goldstein, Taylor and Francis, Washington, DC, 175–299, 1996.
3. H.C. Van de Hulst, *Light Scattering by Small Particles*, Dover Press, New York, 1981.
4. J.T. Hodges, G. Grehan, G. Grouesbet, C. Presser, 'Forward Scattering of a Gaussian Beam by a Nonabsorbing Sphere', *Appl. Opt.*, **34**, 2120–2132 (1995).
5. C.F. Bohren, D.R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, New York, 477–482, 1983.
6. J.K. Agarwal, L.M. Fingerson, 'Evaluation of Various Particles for their Suitability as Seeds in Laser Velocimetry', in *Laser Velocimetry and Particle Sizing*, eds. H.D. Thompson, W.H. Stevenson, Hemispheric Press, New York, 50–66, 1978.
7. R.J. Adrian, C.-S. Yao, 'Pulsed Laser Technique Application to Liquid Gases and Gaseous Flows and the Scattering Power of Seed Materials', *Appl. Opt.*, **24**, 44–52 (1985).
8. F. Borghese, P. Denti, G. Toscano, O.I. Sindoni, 'Electromagnetic Scattering by a Cluster of Spheres', *Appl. Opt.*, **18**, 116–120 (1979).
9. E.D. Hirleman, 'General Solution to the Inverse Near-forward Scattering Particle-sizing Problem in Multiple-scattering Environments: Theory', *Appl. Opt.*, **30**, 4832–4838 (1991).
10. R.J. Adrian, 'Particle Imaging Techniques for Experimental Fluid Mechanics', *Ann. Rev. Fluid Mech.*, **23**, 261–304 (1991).
11. W.J. Wiscombe, 'Improved Mie Scattering Algorithms', *Appl. Opt.*, **19**, 1505–1509 (1980).

12. G. Grouesbet, B. Maheu, G. Grehan, 'Light Scattering from a Sphere Arbitrarily Located in a Gaussian Beam, Using a Bromwich Formulation', *J. Opt. Soc. Am. A*, **5**, 1427–1443 (1988).
13. J.T. Hodges, G. Grehan, G. Gouesbet, C. Presser, 'Forward Scatter of a Gaussian Beam by a Nonabsorbing Sphere', *Appl. Opt.*, **34**, 2120–2132 (1995).
14. V.S.S. Chan, D.I. Robinson, J.T. Turner, A.L. Heyes, 'A Simplified Doppler Global Velocimeter', in *Laser Anemometry 1995*, ed. T.T. Huang, ASME, New York, 9–16, Vol. 229, 1995.
15. W.D. Bachalo, 'The Phase Doppler Method: Analysis, Performance Evaluations and Applications', *Par. Part. Syst. Charact.*, **11**, 73–83 (1994).
16. Y. Aizu, J. Domnick, F. Durst, G. Grehan, F. Onofri, H. Que, M. Sommerfeld, T. Hu, M. Ziema, 'New Generation of Phase-Doppler Instruments for Particle Velocity, Size, and Concentration Measurements', *Par. Part. Syst. Charact.*, **11**, 43–54 (1994).
17. F. Durst, A. Melling, J.H. Whitelaw, *Principles and Practice of Laser-Doppler Anemometry*, Academic Press, New York, 1976.
18. J.N. Forkey, W.R. Lempert, R.B. Miles, 'Accuracy Limits for Planar Measurements of Flow Field Velocity. Temperature, and Pressure Using Filtered Rayleigh Scattering', *Exp. Fluids*, **24**, 151–162 (1998).
19. P. Buchave, W.K. George, 'Bias Correction in Turbulence Measurements by the Laser Doppler Anemometer', in *Laser Velocimetry and Particle Sizing*, eds. H.D. Thompson, W.H. Stevenson, Hemispheric Press, New York, 110–122, 1978.
20. R.D. Keane, R.J. Adrian, 'Optimization of Particle Image Velocimeters. Part 1: Double Pulsed Systems', *Meas. Sci. Technol*, **1**, 1202–1215 (1990).
21. J.M. Coupland, C.D.J. Pickering, 'Particle Image Velocimetry: Estimation of Measurement Confidence at Low Seeding Densities', *Opt. Lasers Eng.*, **9**, 201–210 (1988).
22. M.J. McGill, J.D. Spirhine, 'Comparison of Two Direct-detection Doppler Lidar Techniques', *Opt. Eng.*, **37**, 2675–2686 (1998).
23. M.J. Post, R.E. Cupp, 'Optimizing a Pulsed Doppler Lidar', *Appl. Opt.*, **29**, 4145–4158 (1990).
24. V.J. Abreu, J.E. Barnes, P.B. Hays, 'Observations of Winds with an Incoherent Lidar Detector', *Appl. Opt.*, **31**, 4509–4514 (1992).
25. K.W. Fischer, V.J. Abreu, W.R. Skinner, J.E. Barnes, M.J. McGill, T.D. Irgang, 'Visible Wavelength Doppler Lidar for Measurement of Wind and Aerosol Profiles During Day and Night', *Opt. Eng.*, **34**, 499–511 (1995).
26. D. Rees, I.S. McDermid, 'Doppler Lidar Atmospheric Wind Sensor: Reevaluation of a 355-nm Incoherent Doppler Lidar', *Appl. Opt.*, **29**, 4133–4144 (1990).
27. C.L. Korb, B.M. Gentry, S.X. Li, 'Edge Techniques For Doppler Lidar Wind Measurements with High Vertical Resolution', *Appl. Opt.*, **36**, 5976–5983 (1997).
28. C.L. Korb, B.M. Gentry, S.X. Li, C. Flesia, 'Theory of Double-edge Technique for Doppler LIDAR Wind Measurement', *Appl. Opt.*, **37**, 3097–3104 (1998).
29. B. Hiller, R.K. Hanson, 'Properties of the Iodine Molecule Relevant to Laser-induced Fluorescence Experiments in Gas Flows', *Exp. Fluids*, **10**, 1–11 (1990).
30. T. Obokata, S. Bopp, C. Tropea, 'Two-point Adapter for LDA Probe Using Multimode Fibers', in *Applications of Laser Techniques to Fluid Mechanics*, ed. R.J. Adrian, Springer-Verlag, Berlin, 347–363, 1990.
31. T.D. Fansler, 'Turbulent Production and Relaxation in Bowl-in-piston Engines', SAE Paper 930479, SAE International, Warrendale, PA, 1993.
32. A.O. zue Loye, D.L. Siebers, T. McKinley, H.K. Ng, R.J. Primus, 'Cycle Resolved LDV Measurements in a Motored Diesel Engine and Comparison with k-e Model Predictions', SAE Paper 890618, SAE International, Warrendale, PA, 1989.
33. D.W. Etheridge, P.H. Kemp, 'Measurements of Turbulent Flow Downstream of a Rearward-facing Step', *J. Fluid Mech.*, **86**, 545–566 (1978).
34. L.M. Lourenco, A. Kothrapalli, C.A. Smith, 'Particle Image Velocimetry' in *Advances in Fluid Mechanics Measurements*, ed. M. Gad El Hak, Springer-Verlag, Berlin, 127–199, 1989.
35. I. Grant (ed.), *Selected Papers on Particle Image Velocimetry*, SPIE Milestone Series Volume MS 99, SPIE, Bellingham, WA, 1994.
36. I. Grant, 'Particle Image Velocimetry: a Review', *Proc. Inst. Mech. Eng. C*, **211**, 55–76 (1997).
37. C. Konak, J. Jakes, P. Stepanek, F. Petras, M. Karska, J. Krepelka, J. Perina, 'Effect of Multiple Light Scattering on Transmitted and Scattered Light', *Appl. Opt.*, **30**, 4865–4871 (1991).
38. M. Megahed, P. Roosen, 'First Results of the Delayed Fluorescence Velocimetry as Applied to Diesel Spray Diagnostics', in *Laser Anemometry Advances and Applications*, SPIE, Bellingham, WA, 95–100, Vol. 2052, 1993.
39. S. Ushijima, N. Tanaka, 'Three Dimensional Particle Tracking Velocimetry with Laser-light Sheet Scanning', *J. Fluids Eng.*, **118**, 352–357 (1996).
40. R.N. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1978.
41. D.L. Reuss, R.J. Adrian, C.C. Landreth, D.T. French, T.D. Fansler, 'Instantaneous Planar Measurements of Velocity and Large-scale Vorticity and Strain Rate in an Engine Using Particle-image Velocimetry', SAE Paper 890616, SAE International, Warrendale, PA, 1989.
42. M. Reeves, C.P. Garner, J.C. Dent, N. Halliwell, 'Full-field IC Engine Flow Measurement Using Particle Image Velocimetry', *Opt. Eng.*, **35**, 579–587 (1996).

43. P.V. Farrell, 'Applications of Numerical and Optical Evaluation Schemes for Particle Image Velocimetry', *Exp. Fluids*, **14**, 433 (1993).
44. G.J. Smallwood, 'A Technique for Two-color Particle Image Velocimetry', MS thesis, University of Ottawa, 1992.
45. C.E. Willert, M. Gharib, 'Digital Particle Image Velocimetry', *Exp. Fluids*, **10**, 181–193 (1991).
46. B. Rajalingam, P.V. Farrell, 'The Effect of Injection Pressure on Air Entrainment into Transient Diesel Sprays', SAE Paper 1999-01-0523, SAE International, Warrendale, PA, 1999.
47. A.K. Prasad, R.J. Adrian, 'Stereoscopic Particle Image Velocimetry Applied to Liquid Flows', *Exp. Fluids*, **15**, 49–60 (1993).
48. D.H. Barnhart, R.J. Adrian, G.C. Papen, 'Phase-conjugate Holographic System for High-resolution Particle-image Velocimetry', *Appl. Opt.*, **33**, 7159–7170 (1994).
49. J.M. Barker, J.A. Liburdy, 'Directionally Sensitive Double-pulsed Holographic Particle Velocimetry', in *Laser Anemometry 1995*, ed. T.T. Huang, ASME, New York, 45–50, Vol. 229, 1995.
50. K. Sholes, 'Holographic PIV for Engine Flows', PhD thesis, University of Wisconsin-Madison, 1999.

Separation and Analysis of Peptides and Proteins: Introduction

Christian Schöneich
University of Kansas, Lawrence, USA

Rapid advances in molecular biology, medicine, and the various genome projects will continue to demand new analytical methods and the improvement of existing techniques for the rapid and reliable analysis of peptides and proteins.[1] Complex mixtures of peptides and proteins will be derived from cell lysates and body fluids. Several or all components of these mixtures shall be simultaneously characterized and quantified avoiding lengthy purification schemes for the individual components. Protein samples containing only one gene product may be quite heterogeneous depending on the degree and manifold of potential posttranslational modifications. Such modifications may be predictable based on known target sites for phosphorylation, sulfation, glycosylation, ubiquitinylation etc. However, proteins may also suffer additional, non-enzymatic, chemical modifications such as the deamidation of Asn (to L- and D-Asp, and L- and D-isoAsp),[2] oxidation, β-elimination, and racemization.[3] The latter processes are usually encountered during protein purification or long-time storage in pharmaceutical formulations. Some parameters controlling these non-enzymatic pathways have been established, according to which the most sensitive amino acids can be predicted. For example, 'deamidation hot spots' are located in flexible protein regions, predominantly when Asn is followed by a Gly residue.[4] Nevertheless, there is considerable uncertainty with regard to the most sensitive positions affected by the other degradation pathways. For example, solvent exposure is an important parameter controlling Met oxidation in proteins by peroxides but, in some cases, the metal-catalyzed oxidation of these proteins targeted especially the least accessible regions which, however, contained several metal-binding residues in the vicinity of the affected Met residues.[5] Diastereoselective product formation can add complexity to the analytical task. Asn deamidation can yield two enantiomers of each product, Asp and isoAsp[2] (diastereomers, if these products are components of polypeptides). Met oxidation yields two MetSO diastereomers, Met-d-SO and Met-l-SO.[6] γ-Irradiation of protein samples, important for pharmaceutical or food sterilization processes, can essentially racemize every amino acid at Cα.[7] Finally, D-amino acids are an important component of peptide drugs designed for longer biological half lives. In the majority of the cases, the resolution of posttranslational modifications, including the respective enantiomers or diastereomers, can be achieved by one of several modes of high-performance liquid chromatography (HPLC) or capillary electrophoresis. Therefore, the present section includes the following articles on various modes of HPLC and capillary electrophoresis, important for the separation of peptides and proteins (see also the separate section on *Liquid Chromatography*): **Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis; Hydrophilic-interaction Chromatography in Peptide and Protein Analysis; High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of; Capillary Electrophoresis of Peptides; Capillary Electrophoresis of Proteins and Glycoproteins; Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for; Posttranslational Oxidative Modifications of Proteins**. Often, proteins are subjected to proteolytic digestion, so that individual peptides and their potential modifications can be resolved ('peptide mapping'). An important detail of the peptide mapping procedure is that, depending on the nature of the protein substrate, proteases may show nonspecific cleavage sites. Moreover, posttranslational modifications may lead to conformational changes of the affected proteins, which can affect the accessibility of cleavage sites. Thus, proteolytic digestion shall be carefully optimized and validated, and a chapter in this section introduces the reader to the various aspects of peptide mapping, **Proteolytic Mapping**. A separate article has been devoted entirely to the **Peptide Diastereomers, Separation of**. Mass spectrometry has advanced to become one of the most powerful tools in the analysis of peptides and proteins, largely due to the availability of soft ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). The potential of mass spectrometry in the analysis of peptides and proteins is reviewed in three chapters in this volume: **Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis; High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis; Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis** (for more details, see also the section *Mass Spectrometry*).

A fundamental task in protein biochemistry is the analysis of protein function, structure, and dynamics. A clean analysis of these parameters requires homogeneous protein samples, conveniently monitored by gel electrophoresis. Hence, we have included an article on **Gel Electrophoresis in Protein and Peptide Analysis**, together with articles on **Protein Purification: Theoretical and Methodological Considerations** [see also **Proteins and Peptides Purification in Pharmaceuticals Analysis** in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the section *Pharmaceuticals and Drugs*] and the **Chromatography of Membrane Proteins and Lipoproteins**. The structural analysis of proteins requires one of several possible diffraction techniques, and the article **X-ray Crystallography of Biological Macromolecules**, introduces the reader to various aspects of protein crystallization and X-ray crystallography. Certain structural characteristics of peptides and protein such as, e.g. the extent of α-helical and β-sheet conformation, can be quantitatively measured in solution by circular dichroism (CD) or Fourier transform infrared spectroscopy (FTIR). In this section, we have included an article on **Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis** [see also the section *Infrared Spectroscopy*] while an article on **Circular Dichroism in Analysis of Biomolecules** can be found in the section on *Biomolecules Analysis*. In particular, the FTIR experiments are also possible with solids, permitting the assessment of protein structural changes as a result of processing, e.g. freeze-drying. Protein structure and dynamics can be investigated by means of nuclear magnetic resonance (NMR), electron spin resonance (ESR) and fluorescence spectroscopy (where the latter two techniques require the covalent attachment of a spin or fluorescent label, respectively). In this section, the reader will find an article on **Electron Spin Resonance Spectroscopy Labeling in Peptide and Protein Analysis** and **Fluorescence Spectroscopy in Peptide and Protein Analysis**. The **Nuclear Magnetic Resonance of Biomolecules** and **Glycoprotein Analysis: Using Nuclear Magnetic Resonance** are covered in the sections on *Biomolecules Analysis* and *Carbohydrate Analysis*, respectively. More information on the NMR and ESR techniques can be found in the section *Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy*. Many proteins function through the interaction with biomolecules such as other proteins, lipids, and/or (poly)nucleotides. Moreover, proteins play an active role in the binding and transport of many drugs. Therefore, we have included articles on **Protein–Oligonucleotide Interactions**, and **Protein–Drug Interactions**. A relatively novel technique to measure binding equilibria of proteins is surface plasmon resonance spectroscopy, covered in the article **Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis**, which has the advantage that it can specifically measure on- and off-rates of the binding process. Finally, more insight into the dynamic processes of proteins may be obtained by molecular modeling techniques, which, however, strongly benefit from experimental data which can be used as constraints in order to speed up or bias the modeling process. This topic is covered in the article **Molecular Modeling in Peptide and Protein Analysis**.

ABBREVIATIONS AND ACRONYMS

CD	Circular Dichroism
ESR	Electron Spin Resonance
ESI	Electrospray Ionization
FTIR	Fourier Transform Infrared Spectroscopy
HPLC	High-performance Liquid Chromatography
MALDI	Matrix-assisted Laser Desorption Ionization
NMR	Nuclear Magnetic Resonance

REFERENCES

1. C.K. Larive, S. Lunte, M. Zhong, M. Perkins, G.S. Wilson, G. Gokulrangan, T. Williams, F. Afroz, Ch. Schöneich, T.S. Derrick, R. Middaugh, S. Bogdanowich-Knipp, 'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **71**, 389R–423R (1999).
2. T.V. Brennan, J.W. Anderson, Z. Jia, E.B. Waygood, S. Clarke, 'Repair of Spontaneously Deamidated HPr Phosphocarrier Protein Catalyzed by the L-Isoaspartate-(D-aspartate) *O*-Methyltransferase', *J. Biol. Chem.*, **269**, 24 586–24 595 (1994).
3. M.C. Manning, K. Patel, R.T. Borchardt, 'Stability of Protein Pharmaceuticals', *Pharm. Res.*, **6**, 903–917 (1989).
4. S.M. Potter, W.J. Henzel, D.W. Aswad, 'In Vitro Aging of Calmodulin Generates Isoaspartate at Multiple Asn–Gly and Asp–Gly Sites in Calcium-binding Domains II, III, and IV', *Protein Sci.*, **2**, 1648–1663 (1993).
5. J.L. Jensen, C. Kolvenbach, S. Roy, Ch. Schöneich, 'Metal-catalyzed Oxidation of Brain-derived Neurotrophic Factor (BDNF): Analytical Challenges for the Identification of Modified Sites', *Pharm. Res.*, (2000), in press.
6. V.S. Sharov, D.A. Ferrington, T.C. Squier, Ch. Schöneich, 'Diastereoselective Reduction of Protein-bound Methionine Sulfoxide by Methionine Sulfoxide Reductase', *FEBS Lett.*, **455**, 247–250 (1999).
7. W.A. Bonner, 'The Radiolysis and Radioracemization of Poly-L-leucines', *Radiat. Res.*, **152**, 83–87 (1999).

Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for

Paul L. Weber
Briar Cliff College, Sioux City, USA

1 Introduction 1
2 Direct Detection 1
2.1 Ultraviolet Detection 1
2.2 Fluorescence Detection 2
2.3 Electrochemical Detection 2
2.4 Mass Spectrometric Detection 3
3 Fluorescence Detection Methods for Derivatized Samples 8
4 Other Detection Methods 10
5 Comparison of Detection Methods 11
Acknowledgments 11
Abbreviations and Acronyms 11
Related Articles 12
References 12

Capillary electrophoresis (CE) is a highly efficient separation technique which achieves high resolution with low sample consumption. Crucial to any separation technique are detection modes capable of giving useful data about separated analytes, including both quantitative and qualitative information. A variety of general detection modes for CE have been developed, most based on similar systems used in high-performance liquid chromatography (HPLC), but only some are useful in the analysis of peptides and proteins. Modes which are selective for certain spectral or chemical properties of these analytes are often employed. The absorption of ultraviolet/visible (UV/VIS) light as well as fluorescence (Fl) fit into the former category while electrochemical (EC) detection falls under the latter. The development of mass spectrometry (MS) as a detection mode for peptide and protein analysis may be the ideal choice since a great deal of qualitative information can be obtained and used for identification and structural characterization. By definition, CE involves separation in a ''micro'' environment, a capillary, and this places certain demands on the performance of the detection system and perhaps the greatest challenge is that of sensitivity.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

There are a variety of detection modes for CE and most of these are at least theoretically applicable to the detection of peptides and proteins. Of these detection methods, those which measure some universal property, such as refractive index or conductivity, are not generally useful. On the other hand, detection methods which depend on some property of the peptide or protein which allows differentiation from the background electrolyte and other sample components are of great utility. Thus, detectors sensitive to the absorption of UV/VIS electromagnetic radiation, Fl emission or EC activity are all used. Detection by MS is extremely powerful and can also function selectively by monitoring ions of a particular mass.

Some detection methods deserve special consideration since they can supply additional information about an analyte useful in its identification and characterization. MS and multiwavelength or scanning UV/VIS detectors provide such information. Obviously, the availability of commercial systems with such detectors and ease of use are of concern when selecting a method, prompting careful thought about the use of MS or EC detection.

The sensitivity provided by the detection method can be of paramount importance if the peptide or protein is present in low concentrations and cannot be concentrated sufficiently prior to analysis by off- or on-column techniques. This is particularly important for UV/VIS absorbance detection owing to the small optical pathlength encountered in CE. Derivatization of peptides or proteins to provide fluorescent tags provides increased sensitivity and, often, selectivity.

2 DIRECT DETECTION

2.1 Ultraviolet Detection

For underivatized samples, the most prevalent method of detection is ultraviolet (UV) absorbance detection in the 200–230 nm range, which is primarily due to the presence of amide groups necessarily present in all peptides and proteins. This is evidenced by an inspection of data presented in a recent handbook on the application of CE to protein and peptide analysis. Of the 80 applications cited[1,2] for protein analysis by capillary zone electrophoresis (CZE), micellar electrokinetic chromatography (MEKC) and capillary gel electrophoresis, 75 used detection wavelengths in the 200–230 nm range. Of the remainder, four used UV absorbance detection at other wavelengths while one used Fl detection. However, when capillary isoelectric focusing (CIEF) is used as the mode of CE separation, 10 out of 12 applications monitored absorption at 280 nm. Proteins

will show an absorption maximum at this wavelength owing to the presence of aromatic amino acids, and use of 280 nm detection avoids the problem of a high background absorption in the low-wavelength UV range due to the ampholytes utilized in the separation medium. Unfortunately, 280 nm detection results in a loss of sensitivity compared with detection near 215 nm and this requires higher concentrations of proteins to be used. This, in turn, may result in precipitation of focused proteins.[3] If CIEF is used to analyze peptides that do not contain aromatic amino acid residues, an alternative, low-cost method which uses a universal detector that senses refractive index gradients can be employed and has been demonstrated to be effective in the analysis of tryptic digests of bovine and chicken cytochrome c.[4] Of the 59 applications cited[5] for the analysis of peptide hormones and model peptides, 47 used detection wavelengths in the 200–230 nm range. Of the remainder, four used UV absorption at other wavelengths, four used Fl, two EC, one MS and one refractive index detection methods.

UV detection systems, including multiwavelength detectors, are commonly found in commercial CE systems and are relatively rugged and easy to use. Multiwavelength detectors such as scanning and photodiode-array detectors are used routinely because of certain advantages that they provide. For example, in a report on the refinement of a protein assay by CE for apolipoprotein A-I, a major indicator of arteriosclerosis, multiwavelength detection (Figure 1) indicated that a significant increase in the sensitivity of the method could be achieved by utilizing 195 nm detection rather than the commonly used 200–220 nm range.[6]

The main problem associated with UV detection is the relatively low concentration sensitivity, typically exhibiting a limit of detection in the low-micromolar region for small peptides.[7] Although the use of extended optical pathlength capillaries, available from commercial vendors, can increase sensitivity by an order of magnitude, further increases in sensitivity can be obtained most readily by using a different detection method.

2.2 Fluorescence Detection

Detection of fluorescent emission offers an inherently more sensitive method than UV absorption detection. Direct Fl detection of peptides and proteins is dependent on the presence of two amino acid residues, tryptophan and, to a lesser extent, tyrosine. Unfortunately, these are not particularly strong fluorophores and an even greater limitation is that their excitation range is in the 210–290 nm wavelength range, which is well outside the 488-nm argon-ion laser or the 594-nm helium–neon lasers used in commercial instruments. Nonetheless, in-house detection systems have been used for analysis

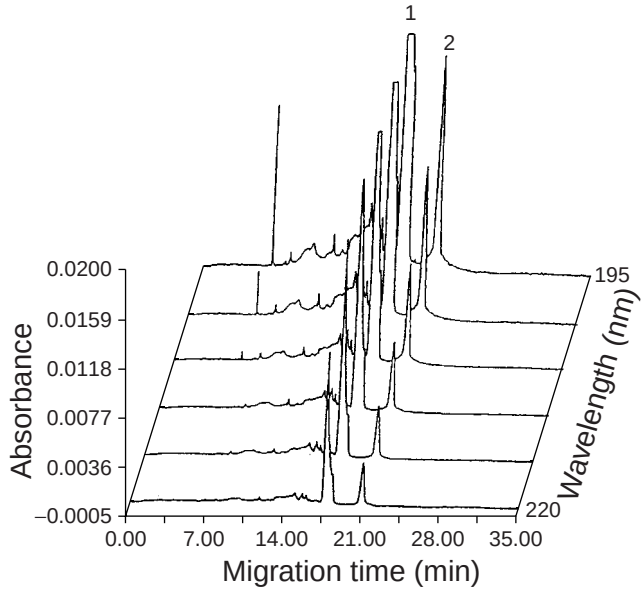

Figure 1 Multiwavelength electropherogram of a normal human serum sample (dilution 1 : 30) spiked with 0.25 mg mL^{-1} apolipoprotein A-I. Peak 1 is albumin and peak 2 is apolipoprotein A-I. (Reproduced from W. Voelter, J. Schutz, O.E. Tsisiloni, A. Weiler, G. Grubler, G. Paulus, S. Stoeva, R. Lehman, 'CE in Biochemical and Clinical Laboratories: Selected Attractive Examples', *J. Chromatogr., A*, **807**(1), 135–149 (1998), Copyright 1998, with permission from Elsevier Science.)

with detection limits in the low-nanomolar range[8,9] for peptide analysis. Even lower limits (200 pM) have been reached[10] for selected peptides with high tryptophan content. Additionally, this detection method permits the acquisition of emission spectra, affording discrimination between peptides containing tryptophan and/or tyrosine.

In general, derivatization of peptide or protein with a fluorescent tag conducive to detection with commercial detectors is the preferred method for sensitive detection. Direct Fl detection is probably confined to use for samples containing tryptophan or tyrosine residues where the additional demands inherent in the derivatization process, such as sample handling, dilution of sample and background Fl introduced by the derivatizing reagent, are undesirable. Such an example is found in the analysis of hemoglobin variants in single cells.[11]

2.3 Electrochemical Detection

The potential of EC detection as an alternative method to UV absorbance detection to enhance sensitivity is indicated by the success of EC detection in HPLC. One EC detection method is based on a traditional method used for determining protein concentrations in biochemical analysis – the biuret reaction. In this method,[12] the wall of the capillary is loaded with Cu^{2+}

ions by a rinse procedure and peptides migrating through the column form complexes with Cu^{2+} ions released by the interaction. The complexes are then detected at carbon fiber microelectrodes yielding detection limits as low as 700 nM. Of interest is the fact that this method can be applied to the detection of cyclic peptides. Sensitive detection of this class of peptides is important because they often cannot be detected by Fl methods since they lack a free amino group which is the functionality required for labeling by most fluorescent reagents. An alternative method for the detection of these molecules was demonstrated using a mixture of two cyclic pentapeptides obtained from solid-phase synthesis which were separated by CE with EC detection at a copper electrode.[13]

Other EC detection methods reported include the analysis of bovine serum albumin using a carbon fiber microdisk array electrode[14] and cytochrome c using a gold microdisk array electrode modified with cysteine.[15] The potential of pulsed amperometric detection (PAD) at a high pH was demonstrated on a mixture of peptides released from recombinant coagulation factor VII by treatment with a glycopeptidase.[16] By using integrated pulsed amperometric detection (IPAD) under mildly basic conditions (pH 9), mass detection limits in the picomole range were observed.[17] EC methods based on the detection of thiols present in many peptides, such as glutathione, have been developed using modified electrodes[18,19] or IPAD at various pH levels.[20]

In general, to date, EC detection has not offered significant increases in sensitivity for peptide or protein analysis. However, this detection method relies on the EC behavior of the analyte rather than on its spectrophotometric activity monitored by UV or Fl detection. This could have advantages in cases where species with high spectrophotometric activity are not separated from the peptidic analyte of interest during electrophoresis or when they are encountered in the background electrolyte. Necessarily, a peptide's response to EC detection will depend on structural features different from those giving a spectrophotometric response, and thus the two methods can work in an orthogonal manner to give structural information. An example is given in Figure 2 in which comparison of PAD and 215 nm absorption detection methods indicate the presence of seven peaks which give significantly larger PAD responses relative to other peaks, owing to a higher content of specific oxidizable groups, such as oligosaccharide chains.[21] Still, the development of more applications in peptide and protein analysis has been hampered by the lack of a commercial CE system available with an EC detector. Lack of commercial development is in turn probably tied to the recent developments in MS as a powerful and informative detection system.

2.4 Mass Spectrometric Detection

Among the various direct methods of detection for peptides and proteins, MS has the potential to become one of the most important of all, certainly *the* most important when qualitative information about the identity of sample components is desired. Thus it is that much effort has been expended in developing capillary electrophoresis/mass spectrometry (CE/MS) as a ''user-friendly'' technique since the first report of the coupling of the MS to CE in 1987.[22] Since both CE and MS instrumentation have each been well developed independently, the key component is the interface. Advances in this area are to the point where commercial vendors have made CE/MS interface kits available for their instruments. For those CE instruments not designed for use with MS, there is an article which provides an experimental foundation on such interfacing with specific reference to application toward the separation of bioactive peptides and of a complex protein mixture.[23] Before examining the various modes of interfacing and their suitability for peptide and protein analysis, it should be noted that the power of MS has been utilized extensively off-line on samples collected from CE. In this procedure, the matrix-assisted laser desorption/ionization (MALDI) technique[24] of sample ionization has been utilized with great success for protein analysis. MALDI can provide good qualitative information on peptide identity but it is not a quantitative technique. Also, since it is technically not a detection technique but rather an off-line method, it will not be discussed further.

Although the potential of a variety of CE/MS ionization techniques has been explored, there appear to be two methods that have been used most often – fast atom bombardment (FAB) and electrospray ionization (ESI). An excellent chapter with references to these techniques and other interface methods has been published.[25] Of these two methods, FAB is the more difficult to implement since it requires a high vacuum for ionization whereas ESI can produce ions at atmospheric pressure. Also, ESI offers the advantage of being able to produce molecular ions with greater than one positive charge. Mass spectrometers measure the intensities of species separated according to their mass-to-charge ratio, m/z, and molecules such as large proteins, if only singly charged, would produce very high m/z ions, often beyond the range of the mass spectrometer. Multiple charges on the protein would bring the m/z values back down into the spectrometer's range. Thus, a protein of M_r 25 000 possessing 25 positive charges can be detected on a spectrometer with a 1000 m/z range. In general, the ESI interface has found widespread use in current peptide and protein applications owing to its combination of versatility, as demonstrated above for producing multiply charged ions,

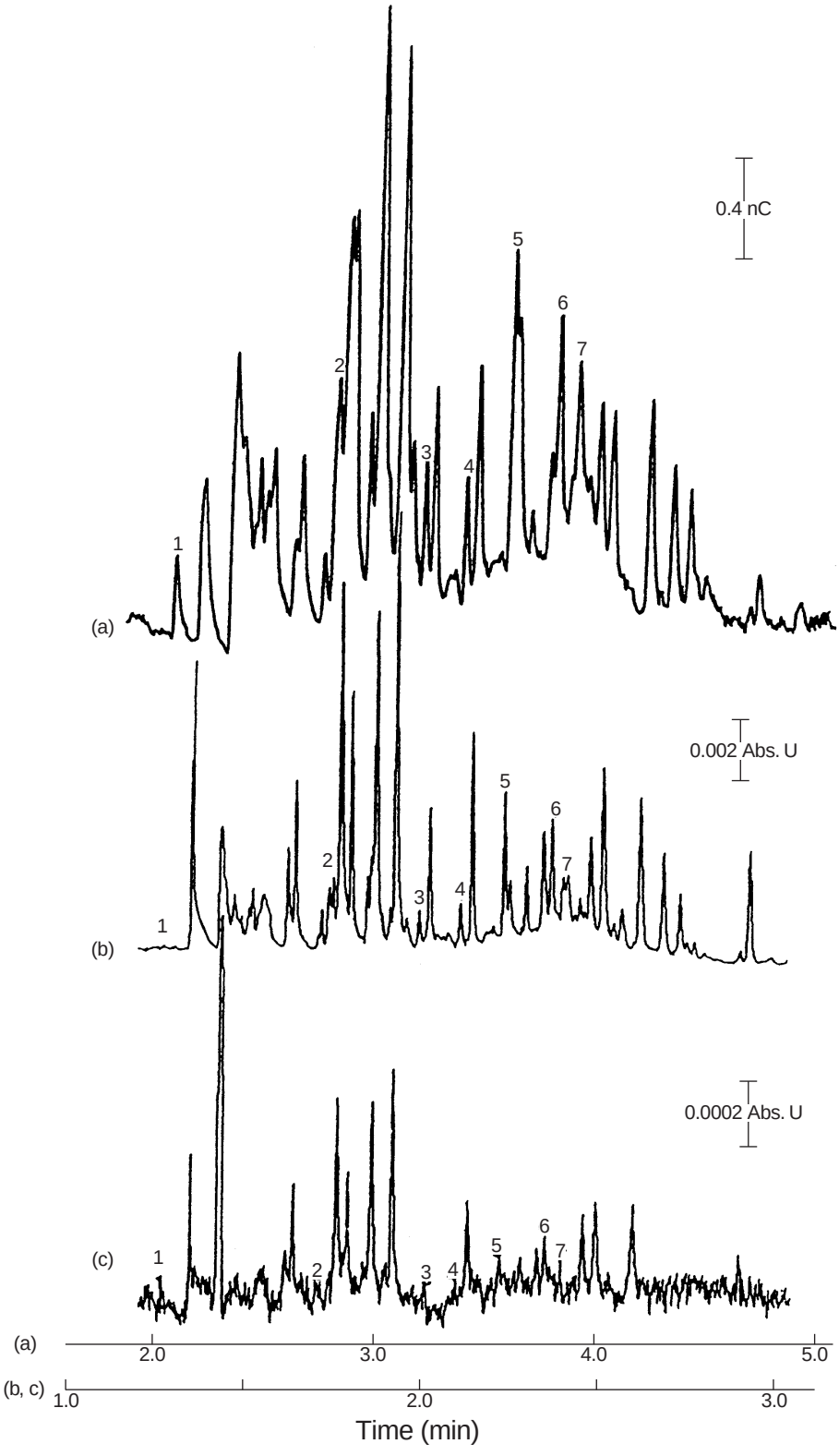

Figure 2 Comparison of CE analysis using PAD *vs* UV detection of a tryptic digest of recombinant tissue plasminogen activator. Detection: (a) PAD; (b) 215 nm; (c) 280 nm. Peaks 1–7 give increased response in PAD relative to other peaks. (Reproduced from Weber and Lunte,[21] by permission of Wiley-VCH.)

facility in implementation, high mass accuracy (0.1%) and high sensitivity.[3,25]

In ESI, a high electric field is generated between the tip of the capillary and the source of the mass spectrometer, causing charged liquid droplets to form. The droplets undergo a loss of solvent, resulting in an accumulation of charge until coulombic explosion occurs. This explosive process repeats to produce ultimately analyte ions in the gas phase. Of the different types of ESI interfaces that have been developed, all can be categorized as either sheath flow or sheathless designs. Figure 3(a) and (b) depict a sheath flow design typical of commercial interfaces and a new sheathless design, respectively. The sheath flow design utilizes the coaxial flow of a liquid to maintain the electrical contact necessary for the electrophoretic process. Also, the composition of the coaxial solution can be adjusted to permit the use of CE buffers that would otherwise perform inadequately when electrosprayed. The general performance in terms of the mass spectral considerations, such as spectral quality, are similar for all designs, whether sheathed or not. The major disadvantage of the sheath flow method is that it suffers from an inherent loss of sensitivity due to the mixing of the coaxial solution with the CE effluent. This loss is due not only to analyte dilution by the incoming solution but also to ionization of solutes present in the coaxial solution. Hence there exists an interest in developing sheathless designs similar to Figure 3(b) that possess the versatility and reliability of the sheathed design.[25]

Although the use of sheath flow ESI results in a loss in sensitivity, the dilution of the CE buffer by the flow does offer some advantages in the terms of the variety of usable CE buffers. In order to be soluble, certain proteins may require relatively high salt concentrations and buffers up 0.1 M salt concentration are acceptable in ESI. Still, maximum sensitivity is obtained by using volatile buffers, such as acetic acid, formic acid or ammonium acetate, at low concentration. Unfortunately, surfactants such as sodium dodecyl sulfate (SDS) in the buffer generate intense background signals, thus limiting the use of MEKC as an electrophoretic separation mode.

The real power of CE/MS utilizing an ESI interface lies in its ability to give qualitative information about an analyte, but to what extent can that power be exploited for analysis of peptides and proteins? To answer this question, a number of different factors must be considered. These factors can be divided into (a) those characteristic of the analyte and its matrix, (b) those of the CE separation and (c) those of the MS performance in conjunction with the ESI interface.

Concerning analyte and matrix considerations, obviously any complexity in the matrix will diminish the usefulness to the extent that it interferes with the CE separation or the performance of the mass spectrometer.

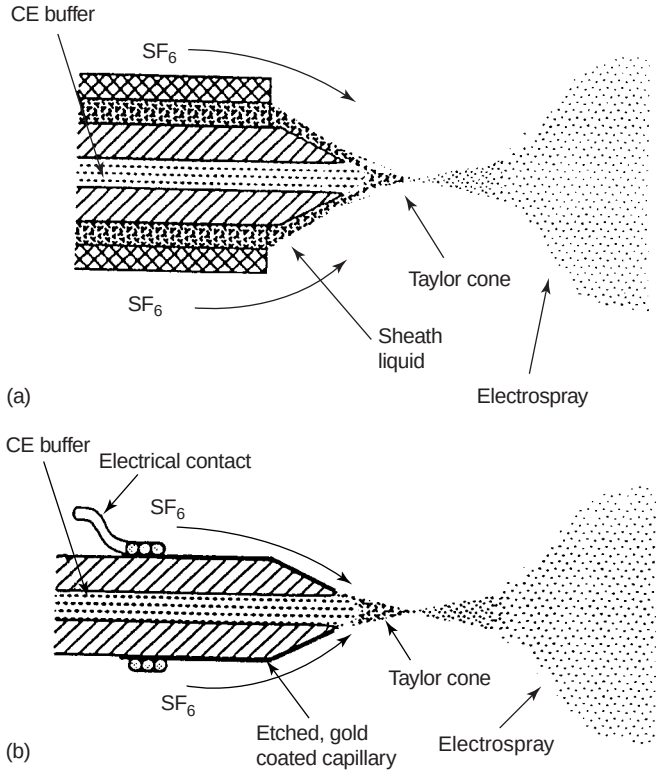

Figure 3 Schematic illustration of CE/electrospray interfaces: (a) sheath flow (coaxial) and (b) exploratory sheathless design. (Reproduced from Smith et al.[25] by permission of Academic Press, Inc.)

More importantly, the higher the molecular weight of the analyte, the more complex is the structure and, thus, the greater the range of masses that must be scanned in order to provide enough information to be useful in structural determinations. That is, a larger scan range would be required for an octapeptide than a tripeptide in order to determine its amino acid sequence.

The CE separation conditions will determine, in part, characteristics of the electropherogram such as resolution and, related to resolution, peak widths at half-heights ($w_{1/2}$). Of course, two well-resolved peaks will place less demand on mass spectral performance than would be otherwise, but also important is the $w_{1/2}$ of a particular peak. Depending on the analyte, this is generally in the range 1–5 s. Peaks with low $w_{1/2}$ put a demand on the performance of the mass spectrometer by requiring a high scan rate in order to sample the peak precisely.

Currently, quadrupole MS instruments are most often used in CE/MS systems using ESI as they give the best overall performance relative to expense. The main impediment they suffer is a lack of sufficient sensitivity due to the maximum number of ions actually introduced and detected per second ($\sim 6 \times 10^7$). It has been determined that if 10 scans are required to profile adequately a peak

with a $w_{1/2}$ of 1–3 s (a scan speed of ~0.2 s) while scanning from m/z 100 to 2000 in 0.2 m/z steps, a dynamic range of only 5 would be achieved. If the scan speed is increased to 1.4 s while scanning only a range of m/z 500 in 1 m/z steps, the dynamic range increases to ~400.[26] In general, as one lengthens the scan speed while lowering the scan range and the m/z resolution, one gains in sensitivity and dynamic range. Thus, it has been stated that a scan range of a few hundred daltons at a scan speed of about 1 s is about all that can be expected with conventional instrumentation.[27] The performance is limited in part to the fact that the CE buffer ions in addition to the analyte will be ionized in the process, thus reducing the percentage of total ions in the spectrometer due to the analyte. The use of a volatile buffers will decrease their contribution to the total ion current and improve performance.

Thus, in order to introduce the needed mass (in the range from high picomoles to low femtomoles) for MS analysis one generally needs relatively high analyte concentrations (in the micromolar range). A number of general methods have been developed for concentrating samples in CE regardless of the detection method. These procedures, which include sample stacking, transient isotachophoretic focusing and using packed precolumns, have been applied successfully to peptide analysis with MS detection.[28,29]

Another simple way to increase capillary electrophoresis/electrospray ionization/mass spectrometry (CE/ESI/MS) performance is to use reduced elution speed detection (RES).[30] In this method, as exemplified in Figure 4(a) and (b), the sample mixture is electrophoresed under the usual high field strength (300 V cm^{-1}, in this case) until just before the first peak is ready to elute from the CE column. At that time, the field strength is dropped (to 30 V cm^{-1}, in this case) to allow the separated component to elute more slowly from the column. This permits more efficient ionization of the solute in the electrospray source, resulting in a greater fraction of the analyte actually sampled. This can be utilized to do one or more of the following: (a) increase the scan speed and enhance sensitivity, (b) increase the scan range, (c) increase the m/z resolution or (d) maintain the same scan speed and thus increase the number of scans for a particular peak. In the example cited, molecular weights were obtained with high precision and detection limits were lowered.

Since conventional CE/ESI/MS suffers from the restrictions listed above, it is often useful to use the system in an MS mode referred to as selective ion monitoring (SIM). This can be used when molecular weights and the m/z values likely to be generated are known and one can select a particular m/z value to monitor. The increase in sensitivity of SIM over the scanning mode is remarkable.

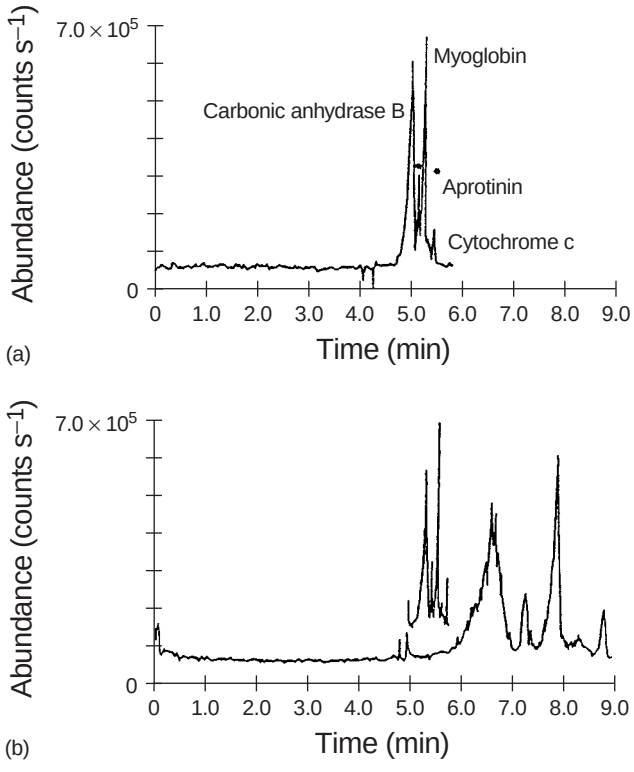

Figure 4 Comparison of (a) constant field strength and (b) RES in the CE/ESI/MS analysis of a mixture of four proteins. Fields strengths: (a) constant at 300 V cm^{-1}; (b) 300 V cm^{-1} until 1 min prior to elution of the first protein, when the field strength was reduced to 30 V cm^{-1}. (Reproduced from D.R. Goodlett, J.H. Wahl, H.R. Udseth, R.D. Smith, 'RES Detection for CE/MS', *J. Microcol. Sep.*, **5**(1), 57–62 (1993), Copyright 1993 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

For a test pentapeptide, the detection limits in the low femtomole range in the scanning mode are reduced to the low attomole range.[25]

Alternative types of MS detectors are in various stages of development. One of the most promising techniques seems to be time-of-flight mass spectrometry (TOFMS), with its parallel detection and high repetition rates. A published report utilized a TOFMS system with a sheathless ESI interface to analyze test proteins, yielding low femtomole sensitivity and, more importantly, high mass resolution and accuracy (resolution greater than 6000 at 10 ppm).[27] Interestingly, the amount of data generated by such detection systems is often beyond the storage and software capabilities of the computer systems used and only selected data acquired are actually recorded.[31]

In terms of actual applications to peptide and protein analysis, CE/MS has been used primarily for the examination of standards[32] and not ''real-world'' samples.[28] A review indicated that 29 different peptides have been

analyzed by various researchers using CE/MS. While 27 of these were studied using a system with an ESI interface, only 12 used a FAB design.[28] The same review also indicated that 21 kinds of proteins have also been examined by CE/MS. The majority of these were ''developmental'' studies not designed to answer biological or protein-synthesis problems. Examples of some of the exceptions include the use of CE/MS to determine protein purity. In one example, impurities in a preparation of a synthetic monoclonal antibody against herpes virus were detected.[33] An example of a biological application can be found in the analysis of protein fragments in urine by on-line preconcentration followed by CE/MS.[34]

One of the most promising areas of CE/MS application is that of peptide mapping and sequencing. Protein digests usually contain a large number of different peptide fragments as a result of proteolytic action. The high separating efficiency of CE affords good separation of the fragments while the power of MS lies in its ability to identify separated fragments. The electropherograms shown in Figure 5(a) and (b) are instructive in that both UV and MS detection were used in this analysis of a tryptic digest of tuna cytochrome[35] and, as expected, the general profiles of both electropherograms are similar. As discussed earlier, the relatively high-molecular-weight fragments were doubly charged and fell within the m/z range of the mass spectrometer, which can be used to identify peaks. Generally, most fragments detectable by UV absorption are observed by MS. Exceptions may be small entities such as amino acids and dipeptides owing to the difficulty in operating MS over a wide m/z range.[25] It should be noted that the equipment used to obtain the results in Figure 5(a) and (b) was a commercial CE system that had been modified to allow UV detection at half the MS detection time. Since the time of this publication, commercial equipment has become available that permits simultaneous UV/VIS and MS detection.

Newer developments in the CE/MS analysis of protein digest have focused on aspects necessary in identifying components in small amounts of sample. One thrust of this effort is to develop CE capabilities to not only provide molecular weight values for the peptides in a CE peptide map, but also to sequence peptides by tandem mass spectrometry (MS/MS). A second aim is to be able to apply these techniques to the relatively low concentrations often encountered in analysis. A good example of the current research directed towards these ends is found in a report detailing the use of a solid-phase extraction (SPE) device coupled to a CE system which, in turn, was coupled to a triple quadrupole mass spectrometer and operated in the SIM mode.[36] SPE led to picomolar detection limits while MS/MS analysis permitted the identification of peptides containing 9–20 amino acid residues. The procedure was

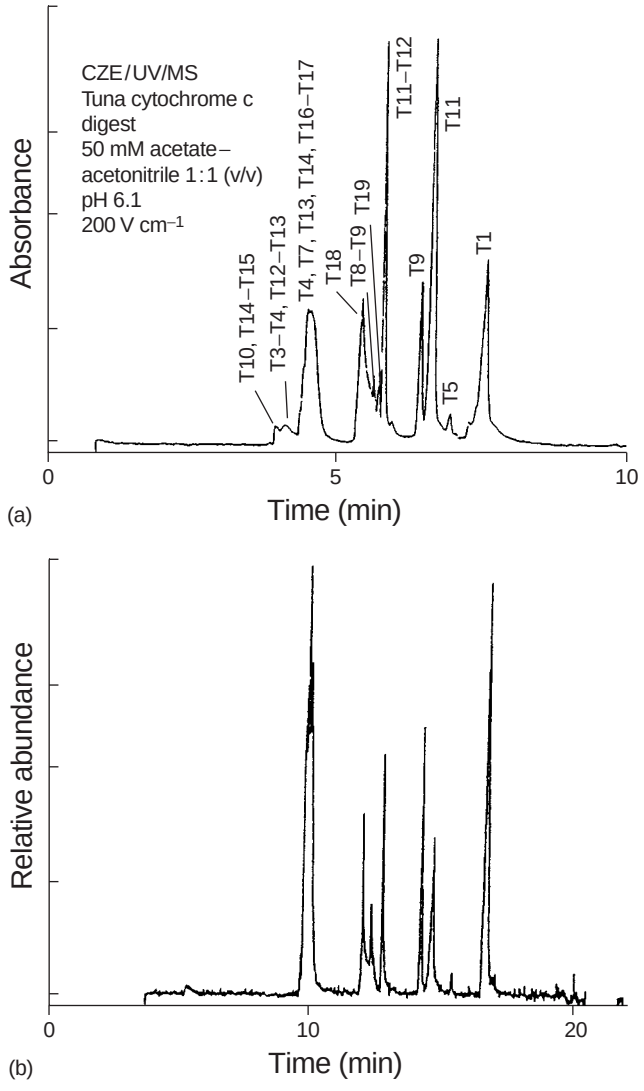

Figure 5 Comparison of (a) UV absorbance (214 nm) detection and (b) MS reconstructed ion electropherograms, with time axes adjusted for corresponding peaks, for a CE/MS separation of a tryptic digest of tuna cytochrome c. (Reproduced from Smith et al.[25] by permission of Academic Press, Inc.)

applied successfully to the analysis of minute amounts of peptides that had been separated and isolated by slab gel electrophoresis. The enthusiasm and optimism of researchers in the field is evident in the following quotation: ''It appears that continued development in the chemical and instrumental aspects of capillary electrophoresis/tandem mass spectrometry (CE/MS/MS) will push this technique to the forefront of the field of bioanalytical chemistry.''[3]

Other research in the CE/MS area has focused on the development of alternative MS detection systems. Alternative detection systems can give superior performance but the systems are not readily available, may be difficult to interface, and are often expensive. Also

explored has been the application of MS detection to modes of CE other than conventional CZE. Since capillary isotachophoresis (CITP) can accommodate much larger sample volumes than CZE and, in addition, analyte bands are concentrated during separation, improvement in concentration detection limits over CZE should be possible. CITP has been coupled to MS and applied to the analysis of proteins,[37] but subsequent reports using the method have been sparse. MEKC coupled to MS would normally be prohibitive owing to interference of the surfactant, but one application used the method for the analysis of peptides by causing analytes to migrate through a micellar solution for separation and then into buffer containing no surfactant before entry into the mass spectrometer.[38] Coupling CIEF to MS has also been accomplished and applied to separations of hemoglobin variants differing in p*I* by ~0.05 pH unit.[39]

Overall, developments in MS detection systems have been proceeding at rapid pace and resources such as the biannual reviews on CE and MS that appear in June 15th issues of *Analytical Chemistry* are a helpful tool to help keep abreast of the field. One general trend that is evident is advancement of microscale MS detection methods. Various research groups are utilizing microfabricated devices for CE/MS with an ESI interface[40,41] and evaluating the performance of such systems for biomolecule analysis.[42] Microchip devices will be an important part of analytical instrumentation of the future and the high informational content available through MS detection would certainly make such systems extremely useful.

3 FLUORESCENCE DETECTION METHODS FOR DERIVATIZED SAMPLES

For the sensitive analysis of peptides and proteins using relatively affordable, commercially available equipment, derivatization of the analyte with a fluorophore and use of Fl detection, particularly laser-induced fluorescence (LIF), is the current method of choice. Sensitivity enhancements over UV absorbance detection are typically two to three orders of magnitude, with concentration detection limits usually in the low-nanomolar range. However, the method is not without its own problems.

Since Fl detection had seen widespread use in liquid chromatography (LC) prior to the advent of CE, a variety of derivatizing reagents for the introduction of the fluorophore have been developed for LC and are useful for CE analysis. Since the introduction of CE, new reagents have been introduced and often find applications in both LC and CE. Indeed, an extensive review of labeling reactions for small amounts of protein in CE and LC analysis has appeared,[43] while another review has focused on derivatization in CE of all types of analytes, not just peptides and proteins.[44]

Some of the more widely used derivatizing agents in CE are given in Table 1. No one reagent can be said to be the best for use in all situations, for a variety of reasons. An initial consideration would be the type of analyte.[43] Large proteins contain multiple sites for a tagging reaction while a small peptide may contain few. From a cursory examination of Table 1, one can see that most reagents are designed to react with amino groups of the peptide or protein. If such a reagent is used with a large protein, multiple amino groups may be labeled, depending on the reaction conditions, while a small peptide may be singly labeled owing to the presence of only one available reaction site. The problem with having many potential labeling sites is that not all sites may react to the same extent, thus introducing greater heterogeneity into the analysis. Large variations in reactivity within proteins are a very real concern since each of the sites exists in a microenvironment with distinct reactivities such as the nucleophilicities of amino groups. Labeling of undesired sites, such as the cross-reactivity of thiols with amine

Table 1 Properties of fluorescent derivatization reagents for peptide and protein analysis by CE

Derivatization reagent	Abbreviation	$\lambda_{ex}/\lambda_{em}$ (nm)	Reacts with	Ref.
4-Phenylspiro[furan-2(3*H*), 1-phthalan]-3,3-dione	Fluorescamine	390/450	Primary amines	46, 52
Fluorescein isothiocyanate	FITC	494/525	Primary and secondary amines	46
9-Fluorenylmethyl chloroformate	FMOC	260/305	Primary and secondary amines	46
o-Phthalaldehyde	OPA	340/450	Primary amines	46, 52, 53
3-(4-Carboxybenzoyl)-2-quinoline carboxaldehyde	CBQCA	442/550	Primary amines	47
Naphthalene-2,3-dicarboxaldehyde	NDA	440/490	Primary amines	48, 54, 55
4-Chloro-7-nitrobenzofuran	NBD-Cl	468/540	Primary and secondary amines	45
Tetramethylrhodamine-5-isothiocyanate	TRITC	548/578	Primary and secondary amines	50
Benzoin	–	325/440	Arginine-containing peptides	49
4-Methoxy-1,2-phenylenediamine	–	330/438	Tyrosine-containing peptides	49

labeling reagents, is also a problem. One might consider adjusting the reaction conditions so that only one reactive site per protein or peptide molecule is labeled but, even if this could be achieved, the relatively small number of labels per molecule would result in low sensitivities, which defeats the purpose of labeling. Hence complete labeling of all desired functionalities is usually the goal of derivatization but can be difficult to achieve, especially for large proteins. In general, the reagents listed in Table 1 are more useful for small peptides than large proteins.

If heterogeneity results from labeling, then separation modes and conditions must be used to cause co-migration of a heterogeneously labeled analyte. An example of such a situation is found in a report in which four proteins of M_r 20 000–77 000 were fluorescently labeled and then separated by SDS capillary gel electrophoresis.[45] CE of labeled proteins exhibited the same separation efficiencies as for unlabeled analytes whereas the limits of detection were enhanced 20-fold over UV absorbance detection. In general, a derivatizing agent should not adversely affect the separation properties of the analytes, a requirement which is perhaps not too difficult to achieve for small peptides which are singly labeled but represents a challenge for large proteins.

Of course, it is also desirable to have a labeling reagent which reacts quickly in an efficient manner while producing a high yield of tagged analyte. These are general features one looks for in any reaction, but they can be particularly important in labeling reactions, especially when postcolumn derivatization is used. Most derivatization procedures are applied prior to injection in an off-line mode for convenience. This is often acceptable as long as one is careful to avoid excessive dilution of the sample with reagent, which diminishes the increase in sensitivity that one is striving for. Table 1 lists examples of a number of useful reagents. Precolumn derivatizations of peptides for CE analysis have been accomplished using some of the ''classical reagents'' such as fluorescamine, fluorescein isothiocyanate (FITC), 9-fluorenylmethyl chloroformate (FMOC) and *o*-phthalaldehyde (OPA).[46] Unfortunately, some of these reagents do not exhibit an excitation maximum coherent with the wavelength of common, commercial lasers such as the He–Cd or Ar-ion lasers and thus exhibit diminished Fl with such sources, resulting in decreased sensitivity. The development of specific reagents, such as 3-(4-carboxybenzoyl)-2-quinoline carboxaldehyde (CBQCA), with an excitation maximum close to the 442 nm light of commercial He–Cd laser, has been realized and applied to the detection of peptides.[47] Prior to the advent of CBQCA, naphthalene-2,3-dicarboxaldehyde (NDA) had been used to generate tagged peptides with a similar excitation maximum, and it has seen widespread use, including the analysis of single vesicles from a gastropod mollusk.[48] Its popularity stems in part from the fact that the reagent is not inherently fluorescent but rather fluorogenic. Thus, unreacted reagent will not interfere in the Fl detection of an analyte. Indeed, such interferences can make reagents such as FITC, which contains a fluorescent fluorescein group, unsuitable for protein analysis even though the labeled protein would have an excitation maximum near 448 nm, the output from commercial Ar-ion lasers. However, this laser's light is close in wavelength to the excitation maximum of analytes tagged with 4-chloro-7-nitrobenzofuran (NBD-Cl) and proteins so labeled have been used for molecular mass studies.[45] Certain reagents have been developed to react with functionalities other than amino groups. These include benzoin, which labels arginine, and 4-methoxy-1,2-phenylenediamine which labels formylated tyrosine.[49] An application of precolumn derivatization utilized tetramethylrhodamine-5-isothiocyanate (TRITC) to label an enkephalin in rat serum with detection limits in the low attomole range, suggesting that capillary electrophoresis/laser-induced fluorescence (CE/LIF) may be useful in the analysis of therapeutic peptides in physiological matrices.[50]

While peptides can be fluorescently labeled by on-column derivatization,[51] the method has seen little use since some of the reagents are present in the background electrolyte. Additionally, conditions must be optimized concurrently for derivatization of the analyte and the separation by CE. However, postcolumn on-line derivatization of peptides is an active area of research that offers some significant advantages over precolumn derivatization. There are fewer manipulations of the sample prior to injection, facilitating the ease with which the sample is analyzed. Also, a peak due to labeling reagent will not be present to hide analyte peaks. Most importantly, peptides are separated by their inherent electrophoretic mobilities unencumbered by the fluorescent tag and the anomalies this may produce, such as the previously discussed heterogeneity of multiply tagged proteins. Reaction kinetics must be fast, resulting in a restriction on the variety of useful reagents. The most popular reagents for this purpose are fluorescamine,[46,52] OPA[52,53] and NDA.[54,55] Detection levels in the high nanomolar range were obtained in the analysis of neuropeptides and proteins using NDA,[54] while another method used fluorescamine to attain detection limits below nanomolar levels and the method was applied to the analysis of tryptic digests.[52] An interesting application of this postcolumn derivatization method used NDA and β-mercaptoethanol to label the physiologically important peptide substance P (SP) and its metabolites.[55] In this study, microdialysis samples were obtained periodically after perfusion of rat striatum with SP and subjected to

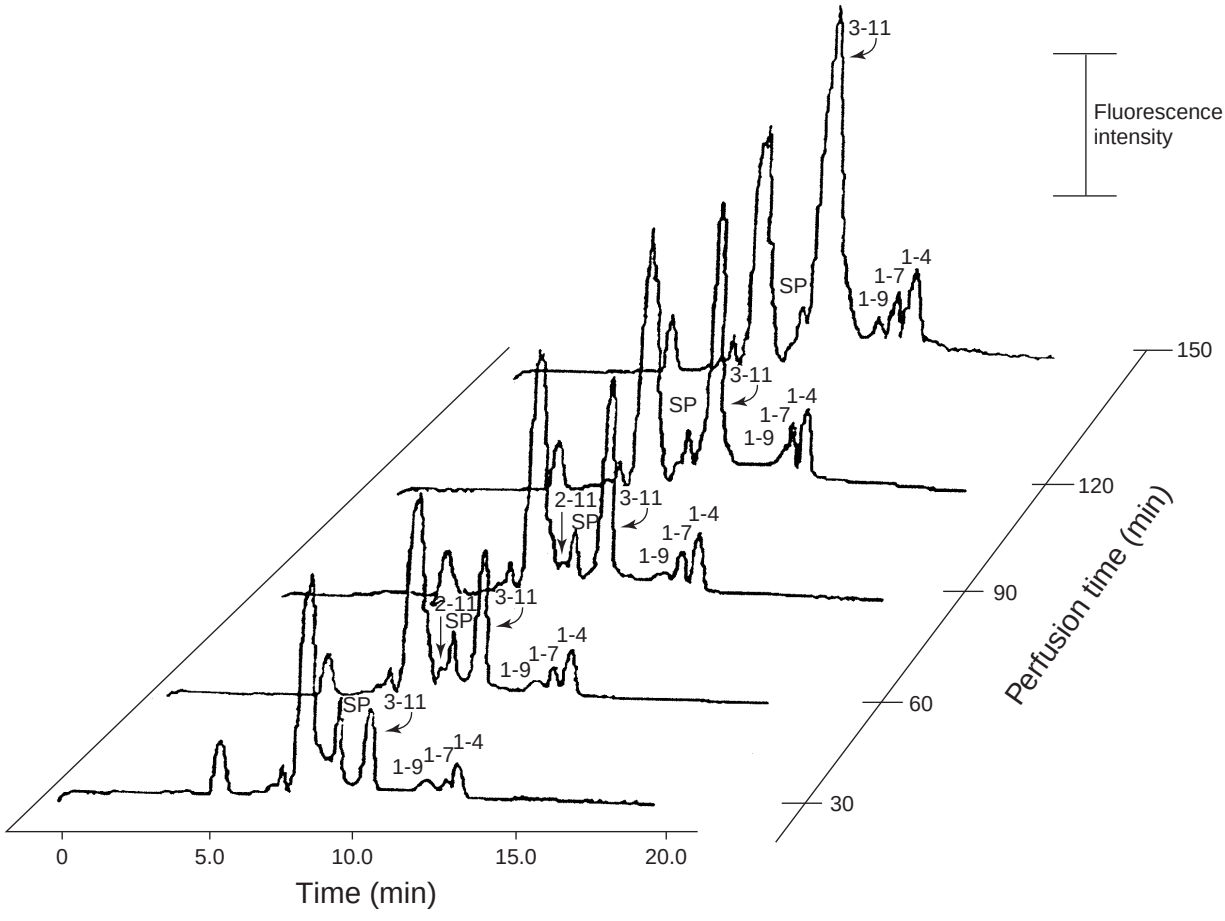

Figure 6 Electropherograms of microdialysis samples obtained after perfusion of rat striatum with 300 μM SP at 0.3 μL min^{-1}. (Reproduced from K.L. Kostel, S.M. Lunte, 'Evaluation of Capillary Electrophoresis with Post-column Derivatization and LIF Detection for the Determination of Substance P and Its Metabolites', *J. Chromatogr., B: Biomed. Appl.*, **695**(1), 27–38 (1997), Copyright 1997, with permission from Elsevier Science.)

CE/LIF analysis to generate the electropherograms in Figure 6.

A variation of postcolumn derivatization that enhances the specificity of the fluorescent tag is found in the use of an affinity-labeling technique. In this method, a fluorophore is bonded to a protein which will form an immune complex with the protein to be analyzed. One recent example of this system employs fluorescein-labeled fragment B of protein A to detect human IgG FC variants.[56]

Overall, off-line precolumn derivatization is most commonly used but no reagent meets the ideal of selectively tagging an analyte at trace levels in an efficient, homogeneous manner. The matter is further exacerbated when analytes are found in complex matrices such as serum or urine. Although on-line precolumn tagging in CE has not been developed, on-line postcolumn methods seem promising, but the lack of commercially available systems has slowed application developments.

4 OTHER DETECTION METHODS

Peptides and proteins can be detected using other methods but reports using these techniques are rather isolated and have not been developed. These include the use of indirect detection in which a fluorescent species is added to the separation buffer and the analytes generate negative peaks as they displace the fluorophore. Although some increase in sensitivity over UV absorbance detection is gained in protein analysis using this method,[57] it suffers from difficulties encountered in quantitation and also from the fact the method is non-selective. Another universal detection method applied to protein studies is thermo-optical absorbance (TOA) detection.[58] In TOA, absorption of incident laser light by an analyte causes a temperature rise in the surrounding solution. A probe beam monitors this solution and changes in intensity proportional to analyte levels. This technique, though technically demanding, exhibits sensitivity similar to that

of LIF. Chemiluminescence has also been used for peptide detection in a peroxide–luminol-based system. As peptides bind to Cu^{2+} in the system, the ability of the copper to catalyze the reaction is diminished and less light is generated. This indirect method provides a rather limited method of analysis.[59]

The use of a biological system, such as a cell, as a detection method has also been introduced. Bradykinin peptides were separated by CE and the effluent containing the peptides was allowed to pass over the surface of cells containing receptors for the bradykinins. Binding to receptors caused release of Ca^{2+} which, in turn, caused an increase in the Fl of Fluo-3-AM, a calcium indicator.[60]

5 COMPARISON OF DETECTION METHODS

In summary, there are a variety of CE detection modes for peptides and proteins and the suitability of any one method for a particular application depends on the conditions both of the sample itself and, when considering the cost of the method, the facility performing the analysis. For applications where sensitivity is not of paramount importance, that is, the analyte is found in concentrations in the micromolar range, UV absorbance detection is well suited to the task. It is robust, straightforward and relatively inexpensive. With minimal increase in expenses, a diode-array or multiwavelength version can yield useful information for identification purposes. It should be noted that preconcentration of a sample either off-line or on-line can effectively increase the ability to analyze lower concentrations to some extent.

EC detection is a method orthogonal to UV absorbance detection since it relies on the EC activity of the analyte. It has not been shown to be significantly more sensitive than UV detection but could offer a mode that may be less expensive than UV detection if commercialized. Miniaturization of this mode is being explored and EC detection may find its real future in microscale applications.

For CE analysis of peptides at nanomolar levels, derivatization with a fluorophore and LIF detection is usually recommended. The manipulations involved and the need for reaction efficiency can be concerns, but a variety of reagents exist which often suit the task. Use of this method on large proteins can be difficult, however, owing to problems arising from multiple, potential reaction sites.

Finally, MS detection may be the ideal detector in terms of the quality of information provided for peptide and protein analysis, particularly in the area of peptide mapping. However, the method is not a ''mature'' one. In the scanning mode, the sensitivity is lower than for LIF of derivatized samples and, at normal scanning rates, the scanning range is limited. Considering the relatively high expense of the system and the need for the user to interface the CE and MS units, perhaps with a commercial interface, one must consider carefully the value of such a detection method. Still, there is probably more method development occurring in this area of CE detection than in any other. Perhaps in the future, commercial CE/MS systems analogous to the ubiquitous gas chromatography/mass spectrometry (GC/MS) units may be found as part of the instrumentation in laboratories performing peptide and protein analysis.

ACKNOWLEDGMENTS

The author would like to thank Sr Mary Jane Koenigs for her help in obtaining literature materials used to prepare this work.

ABBREVIATIONS AND ACRONYMS

CBQCA	3-(4-Carboxybenzoyl)-2-quinoline Carboxaldehyde
CE	Capillary Electrophoresis
CE/ESI/MS	Capillary Electrophoresis/Electrospray Ionization/Mass Spectrometry
CE/LIF	Capillary Electrophoresis/Laser-induced Fluorescence
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CE/MS/MS	Capillary Electrophoresis/Tandem Mass Spectrometry
CIEF	Capillary Isoelectric Focusing
CITP	Capillary Isotachophoresis
CZE	Capillary Zone Electrophoresis
EC	Electrochemical
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FITC	Fluorescein Isothiocyanate
Fl	Fluorescence
FMOC	9-Fluorenylmethyl Chloroformate
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
IPAD	Integrated Pulsed Amperometric Detection
LC	Liquid Chromatography
LIF	Laser-induced Fluorescence
MALDI	Matrix-assisted Laser Desorption/Ionization
MEKC	Micellar Electrokinetic Chromatography

MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NBD-Cl	4-Chloro-7-nitrobenzofuran
NDA	Naphthalene-2,3-dicarboxaldehyde
OPA	*o*-Phthalaldehyde
PAD	Pulsed Amperometric Detection
RES	Reduced Elution Speed
SDS	Sodium Dodecyl Sulfate
SIM	Selective Ion Monitoring
SP	Substance P
SPE	Solid-phase Extraction
TOA	Thermo-optical Absorbance
TOFMS	Time-of-flight Mass Spectrometry
TRITC	Tetramethylrhodamine-5-isothiocyanate
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Fluorescence Spectroscopy in Peptide and Protein Analysis • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Detectors, Absorption and Luminescence • Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Micellar Electrokinetic Chromatography

Mass Spectrometry **(Volume 13)**
Electron Ionization Mass Spectrometry

REFERENCES

1. Y. Kurosu, M. Saito, 'Sodium Dodecyl Sulfate (SDS)-Non-Gel Molecular Sieving Capillary Electrophoresis for Proteins', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonsky, Chapman & Hall, London, 158–172, 1997.
2. F.-T.A. Chen, R.A. Evangelista, 'Protein Analysis by Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonsky, Chapman & Hall, London, 173–197, 1997.
3. M.A. Strege, A.L. Lagu, 'Capillary Electrophoresis of Biotechnology-derived Proteins', *Electrophoresis*, **18** (12–13), 2343–2352 (1997).
4. L. Vonguyen, J. Wu, J. Pawliszyn, 'Peptide Mapping of Bovine and Chicken Cytochrome c by Capillary Isoelectric Focusing with Universal Concentration Gradient Imaging', *J. Chromatogr., B: Biomed. Appl.*, **657**(2), 333–338 (1994).
5. J.C. Osborne, 'Analysis of Peptide Hormones and Model Peptides by Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis Applications*, eds. H. Shintani, J. Polonsky, Chapman & Hall, London, 255–269, 1997.
6. W. Voelter, J. Schutz, O.E. Tsisiloni, A. Weiler, G. Grubler, G. Paulus, S. Stoeva, R. Lehman, 'Capillary Electrophoresis in Biochemical and Clinical Laboratories: Selected Attractive Examples', *J. Chromatogr., A*, **807**(1), 135–149 (1998).
7. I. Messana, D.V. Rossetti, L. Cassiano, F. Misiti, B. Giardina, M. Castagnola, 'Peptide Analysis by Capillary (Zone) Electrophoresis', *J. Chromatogr., B: Biomed. Appl.*, **699**(1–2), 149–171 (1997).
8. D.F. Swaile, M.J. Sepaniak, 'Laser-based Fluorometric Schemes for the Analysis of Proteins by Capillary Zone Electrophoresis', *J. Liq. Chromatogr.*, **14**(5), 869–893 (1991).
9. T.T. Lee, E.S. Yeung, 'High-sensitivity Laser-induced Fluorescence Detection of Native Proteins in Capillary Electrophoresis', *J. Chromatogr.*, **595**(1–2), 319–325 (1992).
10. A.T. Tipperman, K.E. Oldenberg, J.V. Sweedler, 'Native Fluorescence Detection and Spectral Differentiation of Peptides Containing Tryptophan and Tyrosine in Capillary Electrophoresis', *Anal. Chem.*, **67**(19), 3421–3426 (1995).
11. S.J. Lillard, E.S. Yeung, R.M. Lautamo, D.T. Mao, 'Separation of Hemoglobin Variants in Single Human Erythrocytes by Capillary Electrophoresis with Laser-induced Native Fluorescence Detection', *J. Chromatogr., A*, **718**(2), 397–404 (1995).
12. M. Deacon, T.J. O'Shea, S.M. Lunte, M.R. Smyth, 'Determination of Peptides by Capillary Electrophoresis–Electrochemical Detection Using On-column Cu(II) Complexation', *J. Chromatogr., A*, **652**(2), 377–383 (1993).
13. J. Ye, R.P. Baldwin, 'Determination of Amino Acids and Peptides by Capillary Electrophoresis and Electrochemical Detection at a Copper Electrode', *Anal. Chem.*, **66**(17), 2669–2674 (1994).

14. W. Jin, Q. Weng, J. Wu, 'Determination of Bovine Serum Albumin by Capillary Zone Electrophoresis with End-column Amperometric Detection at the Carbon Fiber Microdisk Array Electrode', *Anal. Chim. Acta*, **342**(1), 67–74 (1997).
15. W. Jin, Q. Weng, J. Wu, 'Selective Determination of Cytochrome c by Capillary Zone Electrophoresis with Amperometric Detection at an L-Cysteine-modified Gold Microdisk Array Electrode', *Anal. Lett.*, **30**(4), 753–769 (1997).
16. P.L. Weber, T. Kornfelt, N.K. Klausen, S.M. Lunte, 'Characterization of Glycopeptides from Recombinant Coagulation Factor VIIa by High-performance Liquid Chromatography and Capillary Zone Electrophoresis Using Ultraviolet and Pulsed Electrochemical Detection', *Anal. Biochem.*, **225**(1), 135–142 (1995).
17. W.R. LaCourse, G.S. Owens, 'Pulsed Electrochemical Detection of Nonchromophoric Compounds Following Capillary Electrophoresis', *Electrophoresis*, **17**(2), 310–318 (1996).
18. T.J. O'Shea, S.M. Lunte, 'Selective Detection of Free Thiols by Capillary Electrophoresis–Electrochemistry Using a Gold/Mercury Amalgam Microelectrode', *Anal. Chem.*, **65**(3), 247–250 (1993).
19. T.J. O'Shea, S.M. Lunte, 'Chemically Modified Microelectrodes for Capillary Electrophoresis/Electrochemistry', *Anal. Chem.*, **66**(2), 307–311 (1994).
20. G.S. Owens, W.R. LaCourse, 'Pulsed Electrochemical Detection of Thiols and Disulfides Following Capillary Electrophoresis', *J. Chromatogr., B: Biomed. Appl.*, **695**(1), 15–25 (1997).
21. P.L. Weber, S.M. Lunte, 'Capillary Electrophoresis with Pulsed Amperometric Detection of Carbohydrates and Glycopeptides', *Electrophoresis*, **17**(2), 302–309 (1996).
22. J.A. Olivares, N.T. Nguyen, C.R. Yonker, R.D. Smith, 'On-line Mass Spectrometric Detection for Capillary Zone Electrophoresis', *Anal. Chem.*, **59**(8), 1230–1232 (1987).
23. K.B. Tomer, L.J. Deterding, C.E. Parker, 'Capillary Electrophoresis/Mass Spectrometry for the Determination of Bioactive Proteins and Peptides', *Prog. HPLC–HPCE*, **5**, 3–18 (1997).
24. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**(20), 2299–2301 (1989).
25. R.D. Smith, H.R. Udseth, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, *Methods Enzymol.*, **271**, 448–486 (1996).
26. R.D. Smith, H.R. Udseth, 'Capillary Electrophoresis/Mass Spectrometry', in *Pharmaceutical and Biomedical Applications of Capillary Electrophoresis*, eds. S.M. Lunte, D.M. Radzik, Elsevier Science, Oxford, 229–275, 1996.
27. M.E. McComb, A.N. Krutchinsky, W. Ens, K.G. Standing, H. Perreault, 'Sensitive High-resolution Analysis of Biological Molecules by Capillary Zone Electrophoresis with Reflecting Time-of-flight Mass Spectrometry', *J. Chromatogr., A*, **800**(1), 1–11 (1998).
28. K.B. Tomer, C.E. Parker, L.J. Deterding, 'Capillary Electrophoresis Interfaced with Mass Spectrometry: Electrospray Ionization and Continuous Flow Fast Atom Bombardment', in *Capillary Electrophoresis in Analytical Biotechnology*, ed. P.G. Righetti, CRC Press, Boca Raton, FL, 123–275, 1996.
29. A.J. Tomlinson, S. Jameson, S. Naylor, 'Strategy for Isolating and Sequencing Biologically Derived MHC Class I Peptides', *J. Chromatogr., A*, **744**(1–2), 273–278 (1996).
30. D.R. Goodlett, J.H. Wahl, H.R. Udseth, R.D. Smith, 'Reduced Elution Speed Detection for Capillary Electrophoresis/Mass Spectrometry', *J. Microcol. Sep.*, **5**(1), 57–62 (1993).
31. D.C. Muddiman, A.L. Rockwood, Q. Gao, J.C. Severs, H.R. Udseth, R.D. Smith, 'Application of Sequential Paired Covariance to Capillary Electrophoresis Electrospray Ionization Time-of-flight Mass Spectrometry: Unraveling the Signal from the Noise in the Electropherogram', *Anal. Chem.*, **67**(23), 4371–4375 (1995).
32. D. Figeys, R. Aebersold, 'High Sensitivity Analysis of Proteins and Peptides by Capillary Electrophoresis–Tandem Mass Spectrometry: Recent Developments in Technology and Applications', *Electrophoresis*, **19**(2), 885–892 (1998).
33. R. Kostiainen, E. Lasonder, W. Bloemhoff, P.A. van Veelen, G.W. Welling, A.P. Bruins, 'Characterization of a Synthetic 37-Residue Fragment of a Monoclonal Antibody against Herpes Virus by Capillary Electrophoresis/Electrospray (Ionspray) Mass Spectrometry and ^{252}Cf Plasma Desorption Mass Spectrometry', *Biol. Mass Spectrom.*, **23**(6), 346–352 (1994).
34. A.J. Tomlinson, L.M. Benson, R.P. Oda, W.D. Braddock, B.L. Riggs, J.A. Katzmann, S. Naylor, 'Novel Modifications and Clinical Applications of Preconcentration–Capillary Electrophoresis–Mass Spectrometry', *J. Capillary Electrophor.*, **2**(2), 97–104 (1995).
35. R.D. Smith, H.R. Udseth, C.J. Barinaga, C.G. Edmonds, 'Instrumentation for High-performance Capillary Electrophoresis–Mass Spectrometry', *J. Chromatogr.*, **559**(1–2), 197–208 (1991).
36. D. Figeys, A. Ducret, R. Aebersold, 'Identification of Proteins by Capillary Electrophoresis–Tandem Mass Spectrometry: Evaluation of an On-line Solid-phase Extraction Device', *J. Chromatogr., A*, **763**(1–2), 295–306 (1997).
37. R.D. Smith, J.A. Loo, C.J. Barinaga, C.G. Edmonds, H.R. Udseth, 'Capillary Zone Electrophoresis and Isotachophoresis–Mass Spectrometry of Polypeptides and Proteins Based Upon an Electrospray Ionization Interface', *J. Chromatogr.*, **480**, 211–232 (1989).
38. K. Koezuka, H. Ozaki, N. Matsubara, S. Terabe, 'Separation and Detection of Closely Related Peptides by Micellar Electrokinetic Chromatography Coupled with Electrospray Ionization Mass Spectrometry Using the

Partial Filling Technique', *J. Chromatogr., B: Biomed. Appl.*, **689**(1), 3–11 (1997).
39. Q. Tang, A.K. Harrata, C.S. Lee, 'High-resolution Capillary Isoelectric Focusing–Electrospray Ionization Mass Spectrometry for Hemoglobin Variants Analysis', *Anal. Chem.*, **68**(15), 2482–2487 (1996).
40. R.S. Ramsey, J.M. Ramsey, 'Generating Electrospray from Microchip Devices Using Electroosmotic Pumping', *Anal. Chem.*, **69**(6), 1174–1178 (1997).
41. F. Foret, B. Zhang, H. Liu, B. Karger, 'Hybrid Microfabricated Devices for Electrospray Mass Spectrometry', presented at the 12th International Symposium on High Performance Capillary Electrophoresis and Related Microscale Techniques, Palm Springs, CA, January 25, 1999.
42. P. Thibault, J. Li, N.H. Bings, C.D. Skinner, C. Wang, D.J. Harrison, 'Development of Microfabricated Devices with Low Dead Volume Connection for Capillary Electrophoresis–Electrospray Mass Spectrometry; Evaluation of Analytical Performances for Biomolecule Analyses', presented at the 12th International Symposium on High Performance Capillary Electrophoresis and Related Microscale Techniques, Palm Springs, CA, January 25, 1999.
43. I.S. Krull, R. Strong, Z. Sosic, B.Y. Cho, S.C. Beale, C.C. Wang, S. Cohen, 'Labeling Reactions Applicable to Chromatography and Electrophoresis of Minute Amounts of Proteins', *J. Chromatogr., B: Biomed. Appl.*, **699**(1–2), 173–208 (1997).
44. H.A. Bardelmeijer, H. Lingeman, C. de Ruiter, W.J. Underberg, 'Derivatization in Capillary Electrophoresis', *J. Chromatogr., A*, **807**(1), 3–26 (1998).
45. E.T. Wise, N. Singh, B.L. Hogan, 'Argon-laser-induced Fluorescence Detection in Sodium Dodecyl Sulfate–Capillary Gel Electrophoretic Separations of Proteins', *J. Chromatogr., A*, **746**(1–2), 109–121 (1996).
46. M. Albin, R. Weinberger, E. Sapp, S. Moring, 'Fluorescence Detection in Capillary Electrophoresis: Evaluation of Derivatizing Reagents and Techniques', *Anal. Chem.*, **63**(5), 417–422 (1991).
47. J. Liu, Y. Hsieh, D. Wiesler, M. Novotny, 'Design of 3-(4-Carboxybenzoyl)-2-quinolinecarboxaldehyde as a Reagent for Ultrasensitive Determination of Primary Amines by Capillary Electrophoresis Using Laser Fluorescence Detection', *Anal. Chem.*, **63**(5), 408–412 (1991).
48. D.T. Liu, S.J. Lillard, R.H. Scheller, R.N. Zare, S.E. Rodriquez-Cruz, E.R. Williams, O. Owar, M. Sandberg, J.A. Lundqvist, 'Probing Single Secretory Vesicles with Capillary Electrophoresis', *Science*, **279**(5354), 1190–1193 (1998).
49. K.A. Cobb, M.V. Novotny, 'Selective Determination of Arginine-containing and Tyrosine-containing Peptides Using Capillary Electrophoresis and Laser-induced Fluorescence Detection', *Anal. Biochem.*, **200**(1), 149–155 (1992).
50. C. Chen, D. Jeffrey, J.W. Joregenson, M.A. Moseley, G.M. Pollack, 'Sensitive Analysis of [D-Pen2,5]Enkephalin in Rat Serum by Capillary Electrophoresis and Laser-induced Fluorescence Detection', *J. Chromatogr., B: Biomed. Appl.*, **697**(1–2), 149–162 (1997).
51. S.D. Gilman, A.G. Ewing, 'Analysis of Single Cells by Capillary Electrophoresis with On-column Derivatization and Laser-induced Fluorescence Detection', *Anal. Chem.*, **67**(1), 58–64 (1995).
52. R. Zhu, W.T. Kok, 'Postcolumn Derivatization of Peptides with Fluorescamine in Capillary Electrophoresis', *J. Chromatogr., A*, **814**(1–2), 213–221 (1998).
53. S.D. Gilman, J.J. Pietron, A.G. Ewing, 'Post-column Derivatization in Narrow-bore Capillaries for the Analysis of Amino Acids and Proteins by Capillary Electrophoresis with Fluorescence Detection', *J. Microcol. Sep.*, **6**(4), 373–384 (1994).
54. S.D. Gilman, A.G. Ewing, 'Post-column Derivatization for Capillary Electrophoresis Using Naphthalene-2,3-dicarboxaldehyde and 2-Mercaptoethanol', *Anal. Methods Instrum.*, **2**(3), 133–141 (1995).
55. K.L. Kostel, S.M. Lunte, 'Evaluation of Capillary Electrophoresis with Post-column Derivatization and Laser-induced Fluorescence Detection for the Determination of Substance P and Its Metabolites', *J. Chromatogr., B: Biomed. Appl.*, **695**(1), 27–38 (1997).
56. J.K. Abler, K.R. Reddy, C.S. Lee, 'Post-capillary Affinity Detection of Protein Microheterogeneity in Capillary Zone Electrophoresis', *J. Chromatogr., A*, **759**(1–2), 139–147 (1997).
57. B.L. Hogen, E.S. Yeung, 'Indirect Fluorometric Detection of Tryptic Digests Separated by Capillary Zone Electrophoresis', *J. Chromatogr., Sci.*, **28**(1), 15–18 (1990).
58. J.M. Saz, B. Krattiger, A.E. Bruno, J.C. Diez-Masa, H.M. Widmer, 'Thermo-optical Absorbance Detection of Native Proteins Separated by Capillary Electrophoresis in 10 μm I.D. Tubes', *J. Chromatogr., A*, **699**(1–2), 315–322 (1995).
59. S.-Y. Liao, Y.-C. Chao, C.-W. Whang, 'Indirect Chemiluminescence Detection in Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **18**(10), 667–669 (1995).
60. H.A. Fishman, O. Owar, N.L. Allbritton, B.P. Modi, J.B. Shear, R.H. Scheller, R.N. Zare, 'Cell-to-Cell Scanning in Capillary Electrophoresis', *Anal. Chem.*, **68**(7), 1181–1186 (1996).

Capillary Electrophoresis of Peptides

Robert P. Oda and Michael Strausbauch
Mayo Clinic, Rochester, USA

James P. Landers
University of Virginia and University of Virginia Health Science Center, Charlottesville, USA

1 Introduction **1**
1.1 Instrumentation 1
1.2 Detection 2
1.3 Modes 2
1.4 Introducing Peptides into the Capillary 3
1.5 Electrophoresis-based Capillary Sample Concentration Techniques 3
2 Theory of Electrophoresis **4**
2.1 Mobility of the Analyte 4
2.2 Resolution 5
2.3 Efficiency 5
2.4 Sources of Variance 6
3 Applications **6**
3.1 Predicted Separation: Simulated Versus Observed Migration 6
3.2 Interrogating Peptide Structure 7
4 Posttranslationally-modified Peptides **13**
4.1 Glycopeptides 13
4.2 Micellar Capillary Electrophoresis of Glycopeptides 14
5 Concluding Remarks **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

The ability to resolve effectively peptides obtained by chemical or proteolytic cleavage is important in the structural investigation of proteins, whether it be minute amounts of protein isolated from a biological source or protein produced recombinantly for pharmaceutical use. Capillary electrophoresis (CE) can be used rapidly to resolve peptides electrophoretically in free solution with a selectivity that differs from that of other separation approaches. In this article, we provide a description of the instrumentation for CE and the fundamental principles driving peptide separation at low and high pH. Guidelines for method optimization of peptide injection and separation by CE are given along with a number of examples from the literature demonstrating the utility of CE for interrogating protein structure via CE-based peptide and glycopeptide analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Electrophoresis in capillaries (CE) has gained popularity as a separation technique applicable to both low molecular mass and macromolecular analytes.[1] Separation in low ionic strength buffer is primarily due to differences in analyte charge-to-size ratio. The inability to separate neutral or uncharged analytes by capillary zone electrophoresis (CZE), where they migrate at the electroosmotic boundary, led to the development of micellar capillary electrophoresis (MCE) by Terabe et al.[2] With this mode, an appropriate micelle-forming detergent is added to the separation medium, and separation of the analytes occurs as a result of both electrophoretic mobility (due to charge/mass) and chromatography (due to partitioning of the analyte into the micelles through hydrophobic interaction). Capillary isoelectric focusing (CIEF) involves filling a coated capillary [to eliminate or severely reduce electroosmotic flow (EOF)] with ampholines in an appropriate buffer, injecting the sample, focusing the analytes for a short time, and then mobilizing the focused analyte zones past the detector. This can be accomplished utilizing either a change in electrolyte media to effect EOF or by alteration of hydrostatic pressure; most commercial instruments allow the use of a combination of the two mobilization techniques.[3] While capillary isotachophoresis (CITP) is a separate analytical mode of CE and one of the first to be explored in CE, it is currently used mostly as an on-capillary preconcentration technique.[4] Finally, capillary gel electrophoresis (CGE) effects separation in either an immobilized polymerized matrix within the capillary (e.g. polyacrylamide) or in a free-flowing sieving matrix (e.g. methylcellulose).[5] Sieving matrices are necessary with analytes, such as sodium dodecyl sulfate (SDS)–protein complexes or DNA, where the charge-to-mass ratio is identical regardless of size; here the separation is proportional to size. The primary modes used for the separation of peptides are CZE and MCE, although CIEF and CITP have their potential for specialized applications. CGE could be used for small SDS–peptide complexes with the appropriate matrix.

1.1 Instrumentation

A generalized configuration for CE instrumentation is given in Figure 1. With CZE, sample is introduced into

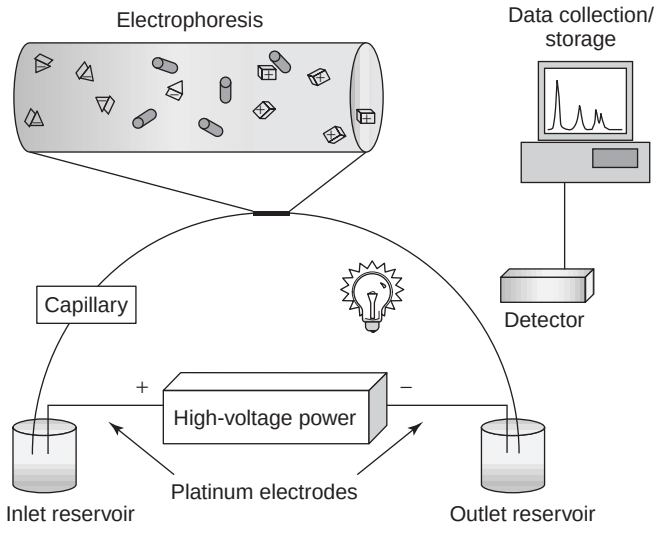

Figure 1 General configuration for CE instrumentation.

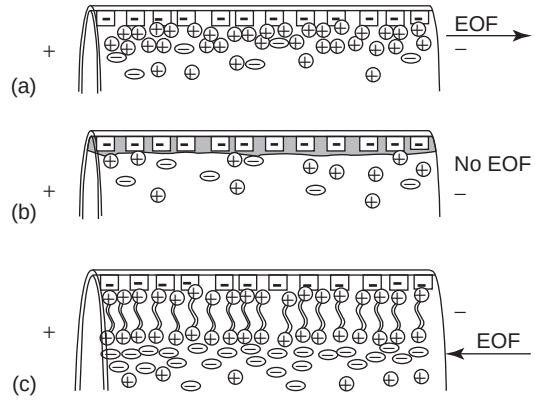

Figure 2 Deprotonated silanols on the capillary surface generate an EOF. (a) A fused-silica capillary has a negative surface charge in most pH ranges, resulting in the formation of a positive double ion layer (and hence a ζ potential) at the surface. Under normal polarity (inlet, anode; outlet, cathode), the positively charged ions will migrate toward the outlet inducing an EOF towards the detector. (b) If the surface is coated with a nonionic layer, no surface charge (ζ potential) will be established and, therefore, there is no EOF. Because there is reduced or no EOF, the electrode polarity must be chosen appropriately so that the molecules of interest have the opposite charge to the outlet electrode. (c) Capillary coatings imparting a positive charge to the inner capillary surface will attract a layer of negative ions. Here the EOF will be anodic in direction.

a buffer-filled capillary either electrokinetically (with low voltage) or hydrodynamically (with pressure or vacuum). Both ends of the capillary are then placed into buffer solution that also contains the electrodes and a high voltage applied to the system. As a result of the applied voltage, the analytes migrate through the capillary and past a detector window where information is collected and stored by an appropriate data acquisition system. Two electrically driven phenomena contribute to the mobility of the analyte: electrophoretic mobility of the analyte and EOF of bulk solution. The charged analyte is mobilized electrophoretically in a manner that is a function of its charge and size, but is also carried by the net movement of buffer toward the cathode due to EOF. EOF is the result of the negative charge imparted to the inner surface of the capillary by the bare silica where cations form an organized double layer along the wall. Under an applied field, the double layer is propelled along the wall in the direction of the cathode. Unless altered through the addition of buffer additives, the use of a coated capillary or with low-pH buffers, the EOF is usually adequate to force the movement of all molecules toward the cathode and past the detector (Figure 2). This includes those with a net negative charge which are electrophoretically drawn toward the anode [assuming a polarity of (anode) → detector → (cathode)]. Subsequently, the analyte order in any given electropherogram has the fastest migrating species passing the detector window first and the slowest passing last (i.e. peaks from left to right represent the fastest to slowest migration). The capillary used for electrophoresis is made from fused silica and typically has an inner diameter in the range 20–200 μm (outer diameter of 200 or 375 μm) and an effective length of 7–100 cm. An external coating of polyimide makes a narrow-bore capillary surprisingly flexible.

1.2 Detection

There exist a number of on-line detection modes for CE, many of which have been developed based on high-performance liquid chromatography (HPLC) detection systems. These include ultraviolet/visible (UV/VIS) absorbance, fluorescence and conductivity. With some instrumentation, diode-array detection capabilities provide additional information. The absorbance spectra obtained through the use of scanning diode-array detection are often useful for component identification; however, sensitivity is somewhat compromised. Mass spectrometry (MS) shows promise as a versatile, powerful and sensitive detector for certain CE separations.[4] There are defined limitations associated with the use of MS as a CE detector when interfaced via electrospray; however, low-pH separations in low ionic strength buffers are easily accomplished and very useful for peptide analysis.

1.3 Modes

Similar to the diversity associated with both conventional gel electrophoresis and HPLC, several modes of CE have been developed. The family of specialized techniques that

have resulted include CZE, MCE, CIEF and CGE. Of these, only CZE and MCE are pertinent to a discussion of CE of peptides. CZE, the most common CE technique, is utilized for both low molecular mass analytes and macromolecules. In contrast to conventional slab gels, separation using CZE is performed in free solution. This means that a gel matrix support is not required, although one can be implemented if desired. A necessity for CZE is that the analytes must be charged or chargeable as separation is achieved based on the species mass-to-charge ratio. CZE may be a misnomer because it implies that CZE is the only CE technique in which the solute forms a zone and migrates at constant velocity. However, CGE is also a zonal technique; thus, CZE is often referred to as free solution CE.

The inability to separate neutral or uncharged analytes with CZE led Terabe's group[6] to develop MCE.[7] The use of an appropriate micelle-forming detergent in the separation buffer allows for separation to occur based on both electrophoretic mobility (due to charge) and partitioning of the solutes into the micelles. While this technique has mainly been exploited for the analysis of small, hydrophobic solutes, it has some applications in the analysis of peptides.

1.4 Introducing Peptides into the Capillary

As a result of the small dimensions of capillaries used in CE, the total capillary volume is typically in the microliter range. For example, a 47 cm × 50 μm capillary has a total capillary volume of 920 nl. Adhering to a commonly accepted ''rule of thumb'' restricting the injected sample volume to ~1% of the total capillary volume, sample volumes are limited to the nanoliter range to avoid overloading.[8] The small sample volume injection into a capillary is usually accomplished by one of three methods. With hydrostatic injection, sample is introduced by pressurizing the inlet vial containing the sample or applying a vacuum to the outlet vial (opposite end) while immersing the capillary inlet into a vial containing the sample. Although much less common, sample introduction can also be accomplished through siphoning, which relies on gravity to introduce the sample into the capillary. Elevation of the inlet end of the capillary to a predetermined height while positioned in the sample vial allows for reasonably accurate delivery of the sample into the capillary and is used routinely with noncommercial (home-made) instruments.

With electrokinetic injection, the most popular method of sample introduction for peptides, the inlet is immersed in the sample and the outlet end placed in the separation buffer with a low voltage (1–10 kV) applied for a short period (usually several seconds) while the sample peptides migrate electrophoretically into the capillary. With peptides, one must be aware of the dependence of analyte injection on a number of parameters. One of the most important is the electrophoretic mobility of the sample components as they migrate into the capillary, in combination with the magnitude and direction of the EOF (the ''bulk'' flow though the capillary when current is applied). Other parameters which can effect this mode of loading include the sample buffer (particularly the ionic strength and pH) and the composition of the sample matrix. The strict dependence of this injection mode on electrophoretic mobility ''biases'' sample introduction; sample components with the highest electrophoretic mobility will be preferentially introduced over those with lower mobilities. Therefore, the results obtained with electrokinetic loading may not be a quantitative representation of the sample peptides. However, electrokinetic loading may be advantageous if the analyte of interest is only a small percentage of the total sample but has a much higher electrophoretic mobility than other constituents. Because of the unique charge characteristics of proteins and peptides (i.e. the presence of charged residues), the charge of the peptides of interest can be manipulated by conditions of pH to enhance their electrophoretic mobilities and provide a positive sample loading bias that may enhance detection by virtue of the selective injection of peptides and proteins preferentially over uncharged sample species. Another situation where electrokinetic introduction is advantageous involves electrophoretic separation in the presence of a polymerized or cross-linked matrix, where pressure injection cannot be used without physically damaging the separation matrix (e.g. SDS–protein complexes separated in a gel-filled capillary). It is noteworthy that sample matrices containing significant concentrations of electrolyte or salts are not efficiently electrokinetically introduced and must be either diluted or desalted before injection.

1.5 Electrophoresis-based Capillary Sample Concentration Techniques

One of the main drawbacks of ultraviolet (UV) detection in CE is the low sensitivity resulting from the inherently small dimensions of the flow cell (most commonly the inner diameter of the capillary) and the injected sample volume. As mentioned previously, introduction of sample volumes substantially larger than ~1% of the total capillary volume can result in poor resolution. While CE is a good mass detector, its poor performance as a concentration detector severely limits its use as an analytical technique to samples having nominally high concentrations (10 $\mu g\,mL^{-1}$ or greater). For this reason, the development of several approaches for ''on-capillary sample concentration'' have been pivotal to the acceptance of CE as a microanalytical technique. These

techniques are only briefly discussed here but may be investigated in greater detail from other sources.[9]

1.5.1 Sample Stacking

One of the simplest methods for sample preconcentration is to induce sample ''stacking'' in a manner comparable to that developed by Hjerten for slab gels.[10] This is easily accomplished in CE by exploiting ionic strength differences between the sample matrix and separation buffer.[11] Sample components in an environment lower in ionic strength than the separation buffer experience an amplified field (and an enhanced electrophoretic mobility) and accelerate to the sample/separation buffer interface where they are slowed by the higher ionic strength of the separation buffer. Within a short time, the ionic strength gradient dissipates and the charged analyte molecules begin to move from the ''stacked'' sample zone toward the cathode. The subsequent reduction in the length of the sample zone as a result of ''stacking'' is roughly estimated to be the ratio of the ionic strength of separation buffer and sample matrix. Stacking can be utilized with either hydrostatic or electrokinetic injection and can typically yield a 10-fold enhancement in sample concentration and, hence, sensitivity (Figure 3). When possible, dissolving the sample in dilute separation buffer may be more advisable than in water since the dramatic differences in EOF between the sample plug and separation buffer may cause laminar flow within the capillary and, hence, lead to peak broadening.[13–16] Since excessive heat produced in the sample plug may denature sample components,[17–19] it may be advisable to ''stack'' at lower applied voltages or ''ramp'' to the separation voltage over several minutes.

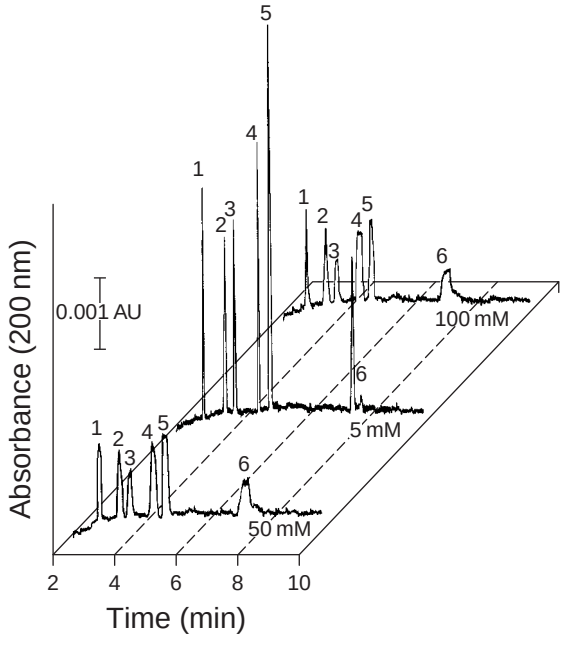

Figure 3 Field amplified stacking of peptides. Separation of a standard peptide mixture (peptide calibration kit, Bio-Rad, Richmond, CA, USA) in a phosphate buffer (pH 2.5) sample matrix at varying concentrations: bottom, 50 mM; middle, 5 mM; top, 100 mM. Separation conditions: 5 s hydrostatic injection; 10 kV running voltage; 50 μm × 27 cm (20 cm to detector) capillary. Peaks: 1 = bradykinin; 2 = angiotensin; 3 = α-melanin-stimulating hormone; 4 = thyrotropin-releasing hormone; 5 = luteinizing hormone-releasing hormone; 6 = leucine enkephalin; 50 μg mL^{-1} each. Separation buffer was 50 mM phosphate buffer (pH 2.5). (Reproduced by permission from Castognola et al.[12])

1.5.2 Sample Focusing

An alternative approach identified for on-capillary stacking is sample ''focusing'' and is based on pH differences between the sample plug and separation buffer. This has been shown to be very useful for the analysis of peptides, mainly owing to their relative stability over a wide pH range.[8] Focusing is accomplished by altering the pH of the sample to a point above the net sample pI (isoelectric point). The high-pH sample plug is flanked between low-pH separation buffer zones (i.e. an equivalent volume of low-pH separation buffer following introduction of the sample plug) and, upon applying a voltage, the negatively charged peptides in the initial sample zone migrate toward the anode. Upon entering the lower pH separation buffer, a pH-induced change in their charge state causes a reversal in the direction of mobility resulting in a ''focusing'' of the peptides at the interface of the sample (high pH) and low-pH buffer plugs. After the pH gradient has dissipated, the peptides, again positively charged in a low-pH separation buffer, migrate toward the cathode as a sharp zone. This approach, limited to samples that can withstand the inherent changes in pH without substantial denaturation, may yield as much as a fivefold enhancement in sample concentration and, hence, sensitivity. Other approaches, including isotachophoresis (ITP) sample enrichment,[20] have been utilized successfully for peptide analysis and will not be discussed here.

2 THEORY OF ELECTROPHORESIS

2.1 Mobility of the Analyte

The electrophoretic mobility (μ) of a charged molecular species can be approximated from the Debye–Hückel–Henry theory (Equation 1):

$$\mu = \frac{q}{6\pi\eta r} \tag{1}$$

where q is the charge on the particle, η is the viscosity of the medium and r is the Stokes radius of the particle. The mass of the particle may be related to the Stokes radius by $M = (4/3)\pi r^3 V$, where V is the partial specific volume of the solute. While one might infer the direct proportionality of mass and radius of the particle, empirical data suggest modifications of Equation (1) to allow for the nonspherical shape, the counter-ion effects and nonideal behavior of proteins and biological molecules.[21]

2.2 Resolution

The simplest way to characterize the separation of two components is to divide the difference in migration distance by the average peak width to obtain the resolution (*Res*) (Equation 2):

$$Res = \frac{2(x_{i2} - x_{i1})}{w_1 + w_2} \tag{2}$$

where x_i is the migration distance of analyte i, the subscript 2 denotes the slower moving component and w is the width of the peak at the baseline.[22] The position of a peak, x_i, is determined by its electrophoretic mobility. The peak width, w, is determined by diffusion and other dispersive phenomena (see below). For two neighboring peaks, $w_1 \simeq w_2$, and (Equation 3):

$$Res = \frac{x_{i2} - x_{i1}}{w_2} \tag{3}$$

from the equation describing a Gaussian curve, the two peaks touch at the baseline when $\Delta x_i = w_2 = 4\sigma$, and $Res = 1$, or (Equation 4):

$$Res = \frac{\Delta x_i}{4\sigma} \tag{4}$$

Substituting for distance ($x_i = v_i t$) and $\sigma = (2D_i t)^{1/2}$ (the width of a Gaussian curve, the theoretical shape of the analyte plug) in Equation (4), we obtain Equation (5):

$$Res = \frac{(\Delta\mu_{app} E)t}{4(2D_{i,avg}t)^{1/2}} \tag{5}$$

where $\Delta\mu_{app}$ is the difference in apparent electrophoretic mobility of the two solutes and $D_{i,avg}$ is the average diffusion of the two solutes.

2.3 Efficiency

Probability theory allows us to assess the efficiency of the transport process.[23] For a random walk process of length L, made of n steps, the variance is given by Equation (6):

$$\sigma = l(n)^{1/2} \tag{6}$$

where l is the length of each step. If each step is independent of any other step, each contributes to the total variance of the process[24] (Equation 7):

$$\sigma_{tot}^2 = \sum \sigma_i^2 \tag{7}$$

Substituting from Equation (6) and rearranging, we obtain Equation (8):

$$\frac{1}{n} = \frac{L^2}{\sigma_{tot}^2} = N \tag{8}$$

The number of steps in the random process, n, is inversely related to the number of theoretical plates, N, a measure of the efficiency of the process.

We may substitute for L and σ in Equation (8) (as we did in Equation 5) to give Equation (9):

$$N = \frac{(\mu_{avg} E)^2 t^2}{2D_i t} \tag{9}$$

where μ_{avg} is the average mobility of the two solutes. Comparing the expression for *Res* (Equation 5) with that for N (Equation 9), we may combine the two to obtain an expression relating resolution to the number of theoretical plates (Equation 10):

$$Res = \left(\frac{1}{4}\right)\left(\frac{\Delta\mu_{app}}{\mu_{avg}}\right) N^{1/2} \tag{10}$$

Equation (10) permits one to assess independently the two factors which affect resolution: selectivity, reflected in the difference of the mobility of the analyte(s); and efficiency, indicated by N.

If $Res = 1$, then (Equation 11)

$$N = \frac{16}{(\Delta\mu_{app}/\mu_{avg})^2} \tag{11}$$

Another expression for N may be derived from Equation (8), using the width at half-height of a Gaussian peak (Equation 12):

$$N = 5.54\left(\frac{L}{w_{1/2}}\right)^2 \tag{12}$$

where $5.54 = 8\ln 2$ and $w_{1/2}$ is the peak width at half-height.[25]

Note that it is misleading to discuss theoretical plates in electrophoresis. The concept comes from chromatographic theory, where a true partition equilibrium between a mobile solute and a stationary phase is the physical basis of separation. In electrophoresis, separation of the components of a mixture is determined by their relative mobilities in the applied electric field, which is a function of their charge, size and shape. The theoretical plate is merely a convenient concept to describe the analyte peak shape and to assess the factors which affect separation.

2.4 Sources of Variance

While N is a useful concept to compare the efficiency of separation among columns, or between laboratories, it refers to the behavior of a single component during the separation process, and hence is unsuited for describing the separation of two components or the resolving power of a capillary. A more useful parameter is the height equivalent to a theoretical plate (HETP)[23] (Equation 13):

$$\text{HETP} = \frac{L}{N} = \frac{\sigma_{\text{tot}}^2}{L} \tag{13}$$

HETP might be thought of as the fraction of the capillary volume occupied by the analyte. It is more practical to measure HETP as an index of separation efficiency, rather than N, as the individual components which contribute to HETP may be individually evaluated and combined to determine an overall value. The variance of multiple dispersive phenomena on the analyte may be summed (Equation 14):

$$\sigma_{\text{tot}}^2 = \sigma_{\text{diff}}^2 + \sigma_T^2 + \sigma_{\text{int}}^2 + \sigma_{\text{wall}}^2 \tag{14}$$

Ideally, electrophoresis in a noninteracting medium should result in symmetrical zones of analytes. The length of time for separation would be determined by the differences in mobility of the two analytes and thus, in theory, given a long enough separation time, one should be able to resolve any two analytes. However, in the real world, materials do not behave ideally, and analytes frequently interact with the capillary or components of the separation medium. Therefore, we must consider the sources that create nonideal behavior.

A consideration of all the factors influencing σ_{tot}^2 includes not only diffusion, but also differences in mobility or diffusion caused by Joule heating (i.e. thermal gradients generated within the capillary during the electrophoretic process), the fact that the sample is not introduced as an infinitely thin disk but as a plug of finite length, and interaction of analytes (which may not be the analyte of interest, but might affect electrophoretic mobility of the analyte of interest) with the capillary wall. While important in understanding separation in CE, this will not be discussed here and the reader is referred to literature which addresses this issue in detail.[13,26–30]

3 APPLICATIONS

3.1 Predicted Separation: Simulated Versus Observed Migration

Two approaches have been used to evaluate the relationship between electrophoretic mobility and peptide charge and mass. The first modifies Equation (1) to Equation (15):

$$\mu = kqM^{-x} \tag{15}$$

where $x = 1/3$ assuming the spherical shape of Stokes law, $x = 1/2$ assuming the shape is related to the radius of gyration and proportional to the length of the peptide chain or $x = 2/3$ assuming the frictional forces are proportional to the surface area of the peptide.

Offord[31] first proposed $x = 2/3$ based on studies of a large number of peptides by paper electrophoresis; this value has since been experimentally verified by numerous studies. Nyberg et al.[32] found a good correlation with Offord's model for a series of synthetic substance P (SP) peptides. Likewise, Deyl et al.[33,34] found a linear relationship between the mobility of cyanogen bromide fragments of collagen between pH 6.9 and 10.5 and $M^{-2/3}$, surmising that the effect of molecular size was smaller than the effect of charge, at least for small- to medium-sized peptides.

A comparison of the three models was carried out by Rickard et al.[21] using the mobilities of 33 peptides obtained at pH 2.35 and 8.15. The strongest correlation for peptides derived from the tryptic digestion of human growth hormone was with $x = 2/3$. In a study of 35 peptides with buffers at pH 2.0, 4.0, 7.0 and 10.0, Kurosu et al.[35] found the best correlation with low-pH buffers and Offord's parameter. Landers et al.[36] obtained similar results for two peptides and their homo- and heterodimers using peptide migration time (proportional to μ^{-1}), and Florance et al.[37] arrived at a similar conclusion for motilin peptides at acidic, neutral or basic pH. Offord's model has been substantiated by van de Goor et al.[38] for modified adrenocorticotropic hormone peptides over a wide pH range and by Adamson and Reynolds[39] with casein phosphopeptides.

A second, more empirical approach to define the relationship between electrophoretic mobility and peptide charge and mass was proposed by Grossman et al.[40] Using the data from 40 peptides, ranging in length from 3 to 39 amino acids and in charge from 0.33 to 13, they fitted their mobility data to Equation (16):

$$\mu = \frac{k_1[\ln(q+1) + k_2]}{n^{0.43}} \tag{16}$$

where k_1 and k_2 were buffer constants and n is the number of amino acid residues in the peptide. This modification of the basic equation includes contributions to size by shape and length of the peptide chain. A plot of mobility versus hydrophobicity suggested a causal relationship with increased molecular mass. However, because the size parameter is based on number of residues and not molecular mass of the amino acids,

these authors could not directly account for an increase in peptide molecular mass. A further difficulty lies in the inability to estimate the effects of buffer ions and ionic strength on the double layer of the peptide chain; however, one would expect an increase in the value of the exponent for a compact molecule as ionic strength increased. Issaq et al.,[41] comparing Grossman et al.'s model with Offord's, concluded that $kq/M^{2/3}$ gave the best correlation for small polyalanine peptides in 50 mM ionic strength buffer. However, Chen et al.,[42] carrying out a study similar to that of Rickard et al.,[21] found the best correlation with $M^{-1/2}$; likewise, Survay et al.[43] obtained the best fit of their data to Grossman et al.'s model, rather than Offord's model. Castagnola et al.[44] found a good correlation with Grossman et al.'s model with $n^{0.5}$ for tryptic peptides from horse myoglobin.

Hilser et al.[45] found that Grossman et al.'s model gave a better fit with tryptic peptides derived from recombinant human insulin-like growth factor I than Offord's model. Further inspection indicated that both models inadequately treated peptide size, and a more complex treatment including thermodynamic functions of enthalpic and entropic contributions of polar and apolar moieties was required to improve the ability to predict peptide migration.

Compton[46,47] attempted to reconcile the differences in these two approaches by developing a model based on Debye–Hückel–Henry theory, demonstrating that ionic strength would influence the mobility–mass relationship depending on the magnitude of the polymer molecular mass and the ionic strength of the buffer. In fact, mobility is a continuous and complex function of molecular mass between the limits of $M^{-1/3}$ and $M^{-2/3}$. For compact polymers in low ionic strength solutions one would expect mobility to reflect a dependence on $M^{-1/3}$; for larger molecules in high ionic strength buffers, $M^{-2/3}$ would dominate. Hence the dependence observed by Grossman and others would be indicative of moderate polymers in a medium ionic strength buffer.

In a comparison of the models of Offord, Grossman et al., and Compton, Cifuentes and Poppe[48,49] found good agreement with any of the three models when constants were chosen appropriately. They attempted to account for the electrostatic interactions of the charges within the polymer, deriving Equation (17):

$$\mu = \frac{A[\log(1 + Bq)]}{M^C} \tag{17}$$

From a set of 48 peptides, these investigators obtained values of $A = 1758$, $B = 0.297$ and $C = 0.411$.

Application of Equation (15), (16) or (17) to the prediction of mobility of peptides is relatively straightforward. The charge on the peptide may be calculated by the Henderson–Hasselbach equation at the buffer pH, using the ionization constants of the component amino acids. While the p*K*s of the side groups are typically assumed to be identical to those of the individual free amino acids, the C- and N-terminal ionization constants may differ by one or two orders of magnitude. Within the peptide, the ionization constants of individual residues may be affected by neighboring side chains, the local dielectric constant due to solvation and the ionic strength of the medium. Tabulated values for amino acid functions are available from Skoog and Wichmann,[50] Rickard et al.,[21] Pennio[51] and van de Goor.[38] Several computer programs have also been developed to simplify the process by simply entering the sequence, if it is known (CHARGEPRO™ and PHYSCHEM™, IntelliGenetics, Mountain View, CA; Compudrug Chemistry, Budapest, Hungary). The predicted migration may then be adjusted (most typically by modifying the pH of the buffer) to obtain the desired separation of the analytes of interest.

With respect to buffer pH, a number of reports have evaluated the importance of the separation pH in predicting peptide mobility based on charge and size. Castognola et al.[52] interrogated the importance of the Stokes radius as a function of the peptide pH. Since the Stokes radius of the peptide changes with pH, one would predict that the accuracy of peptide migration would change as solution pH is altered – this is not accounted for by the standard equation. Applying ''best-fit'' procedures to CE mobility data, the authors found the best predicative pH region to be 2.25–2.5 where the polymer Stokes radius relates to mass by the coefficient 0.68, close to Offord's 2/3, and suggests that the Grossman et al.[53] approach using the p*K*s of free amino acids is insufficient. This same topic was further explored by Castognola et al.[12] using 2,2,2-trifluoroethanol (TFE) as an organic solvent in the separation buffer. This is a clever approach to interrogating predictive models since TFE modifies both the charge and Stokes radius. They found that predictions of mobility, size and charge were best in the pH range 2.0–2.5. This is likely due to the C-terminus being protonated, which the authors suggest eliminates an intramolecular ion pair causing a more compact form and exposing the apolar side chains; as a result the predication becomes less accurate.

3.2 Interrogating Peptide Structure

As indicated above, selectivity (resolution or separation) in CE is determined by differences in analyte electrophoretic mobility. Electrophoretic mobility is a function of analyte charge and size (more specifically, hydrodynamic volume). The charge on the molecule is most readily influenced by the pH of the separation medium; hence the buffer and pH are typically the first operating parameters optimized. Additives, such as

organic solvents which modify microenvironmental pH and solvation of the analyte, or salts, organic or inorganic, which alter the ionic double layer and the protein microenvironment, can also be used. Compounds that complex or interact with specific functional groups may alter the size or shape of the peptide and thus affect mobility.

3.2.1 Buffer Type and pH

The choice of separation medium (buffer) is critical to obtaining successful separations of analytes. Once the optimum wavelength for detection of the analyte(s) has been established, a buffer is required that does not interfere with the ability to detect the analytes of interest but maintains the solubility of the analytes, maintains buffering capacity through the analysis and permits the desired separation. It is important to prepare buffers with reagents of the highest quality, including Milli-Q (or equivalent) purified water, and filtering the buffers through a 0.22- or 0.45-μm filter to remove any particulate matter. Buffers stored at 4 °C should be brought to room temperature and degassed before use.

A variety of electrolytes can be used to prepare buffers for CZE separations. With absorbance detection, a major requirement of all components of the buffer is a low UV absorbance at the wavelength used for detection. This restriction substantially limits the choices to a moderate number of nonUV-absorbing electrolytes. However, a number of UV-absorbing organic buffers have been used with success at low concentrations to minimize background absorbance. At low pH, phosphate, acetate, formate and citrate have been used effectively; in the basic pH range, Tris, tricine, borate and 3-(cyclohexylamine)-1-propane sulfonic acid (CAPS) are acceptable electrolytes. For detection modes other than UV absorbance, a number of other electrolytes can be utilized. In fact, for indirect detection, a buffer with high background absorbance to allow lower detection limits of analytes is desired; thus chromate or phthalate are often utilized. For electrochemical detection, the buffer must be compatible with the analytes and also present a stable background conductivity against which the analytes can be detected.

When information about the sample is available, it is advisable to choose a buffer pH that approximates the pK of the solute mixture. This choice is relatively simple with pure or partially purified preparations; however, with crude biological mixtures, the pK_{avg} will typically be close to neutrality. Note that increasing the pH between 4 and 9 will result in an increase in the EOF. The buffer pH may be altered in a secondary manner by other parameters such as temperature, ion depletion of the buffer (caused by repetitive use of the same separation buffer) and organic additives. Examples of the effect of buffer pH on CE separation is given in Figure 4(a) and (b) for a series of peptides. The separation of a mixture of five peptide "isomers" representing residues 101–112 from the thyroid stimulating hormone (TSH) over a wide range of pH values is illustrated. At pH 2.5, all five peptides (differing only in sequence, not composition), which are identical in amino acid composition but vary in sequence, comigrate as a broad peak (not shown). At pH 8.61, the peptides migrate as two groups, while at pH 11.64 the resolution of four of the five peptides is observed.[54]

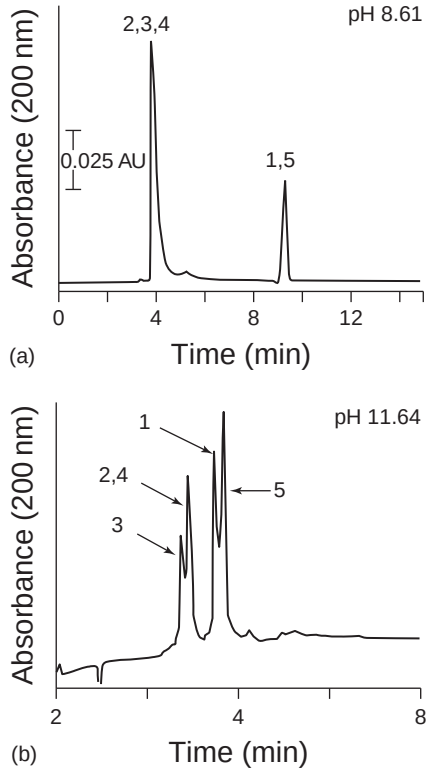

Figure 4 Dependence of peptide separation on buffer pH. Five peptides containing the same 12 amino acids but in varied sequence were analyzed at (a) pH 8.61 and (b) pH 11.64. Sequence of native peptide: (1) KTNYCTKPQKSY; (5) reverse sequence. Experimental conditions: 100 mM borate buffer with 10 mM diaminopentane, adjusted to the required pH with NaOH; 25 kV running voltage; 50 μm × 27 cm (20 cm to detector) capillary. (Reproduced by permission from Oda et al.[61])

3.2.2 Buffer Concentration and Ionic Strength Effects

Several parameters define the useful buffer concentration, including the capillary length and internal diameter, the applied electrical field strength and the efficiency of the capillary thermostating system. Generally, the use of moderately high ionic strength buffers is desirable for suppression of ion-exchange effects between the charged analyte ions and the ionized silanol groups on the capillary

wall. However, the current (Joule heat) associated with buffer concentrations greater than 100 mM may overcome the capillary thermostating capability of the system at higher applied voltages; excessive Joule heating may affect both resolution and analyte stability. Buffers that may be problematic include those containing high-mobility electrolytes such as chloride, citrate, phosphate and sulfate. The excessive Joule heating associated with high-concentration buffers can be circumvented in two ways. The easiest solution is to lengthen the capillary so that a tolerable current is maintained (decrease the field strength). However, this will increase the effective capillary length and may compromise resolution or detection if excessive diffusion or analyte–wall interactions occur. A reduction in cross-sectional area of the capillary will also reduce heating by decreasing the current flux density, allowing for more effective dissipation of heat due to a greater surface-to-volume ratio. Alternatively, buffers that run at relatively low current (and Joule heat) can be used. One such buffer in the pH 7–9 range is borate buffer, which has been shown to be an excellent CE buffer at concentrations as high as 500 mM.[55] This may be useful for the CE separation of fluorescently derivatized peptides in the neutral pH range. An added advantage of using a higher ionic strength separation buffer is the increased sample loading capacity that results from the on-capillary stacking effect, as described in an earlier section.

The method described by Nelson et al.,[19] termed the ''Ohm's law plot'', allows for facile determination of the 'functional' buffer concentration and the maximum voltage that can be utilized with the particular buffer system (i.e. the functional limit for capillary thermostating), and allowing for adequate equilibration, voltage is applied to a buffer-filled capillary (Figure 5) for short time intervals (e.g. 1 min) and the current recorded as the voltage is incrementally increased. When current is plotted as a function of voltage, loss of linearity indicates that the thermostating capacity of the system has been exceeded (i.e. the generated Joule heat is not effectively being dissipated). One should strive for heat generation of $<1\,W\,m^{-1}$ for optimum separation and it should not exceed a power of $5\,W\,m^{-1}$.[56]

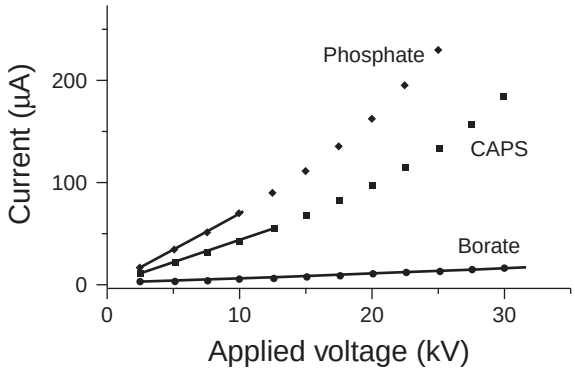

Figure 5 Determination of optimum separation parameters using the Ohm's law plot. Current is plotted against applied voltage for each of three buffers as the applied field is incremented at $2.5\,kV\,min^{-1}$. Buffers were: 100 mM phosphate (pH 2.5), made by dilution of phosphoric acid and titration with NaOH; 100 mM borate (pH 8.3), made by titrating 25 mM sodium tetraborate with 100 mM boric acid; and 100 mM CAPS (pH 11.0), made by titration of the appropriate concentration dissolved in water with NaOH. The lines indicate the linear portions of the plots. The departure from linearity indicates the excessive increase in current at the applied voltage, and is a reflection of the increase in the capillary temperature.

Increasing the ionic strength of the separation buffer increases the thickness of the ionic double layer, and has the effect of decreasing the EOF, hence increasing the separation time. An advantage of increased ionic strength, in addition to the obvious improvement in buffer capacity, will be to decrease analyte–wall interactions.[57] The net effect on the separation, therefore, will be to increase the resolution, provided that the capillary thermostating capability is not exceeded and that unwanted analyte dissociative processes (e.g. ligand–peptide/protein dissociation) do not occur. On the other hand, an increase in ionic strength might improve resolution in mixtures by decreasing nonspecific analyte interactions. As mentioned above, increasing the ionic strength will also increase the current at a constant voltage, hence adequate thermostating of the capillary may become a concern. Also, use of buffers with poor transparency in the low-UV region (e.g. citrate, $\lambda_{max} = 230\,nm$) may have adverse affects on detection sensitivity.

3.2.3 Augmenting the Separation Buffer with Additives

The selectivity in peptide separations can be enhanced with the use of buffer additives. Several classes of additives have been identified as applicable to CE and, as seen in Table 1, the additives are multifunctional, not only suppressing analyte–wall interactions, but also affecting analyte solubility and in some cases selectivity.

3.2.3.1 Organic Salts A number of organic modifiers have been used successfully in peptide CE with the effect being dependent on the nature of the additive. Positive effects on separation result from alteration of the solvation properties of the modified buffer on the analyte or diminished EOF due to changes in the thickness of the double-ionic layer (and hence the zeta potential) or reduction of the viscosity of the buffer. For example, the addition of 1,4-diaminobutane to the buffer has been proposed to enhance resolution by slowing the

Table 1 Common buffer additives in CE and their effects

Additive	Example	Function
Inorganic salts	NaCl, $CaCl_2$, K_2SO_4	Modification of EOF; protein conformational changes; protein hydration
Organic solvents	Methanol, acetonitrile, ethylene glycol	Modification of EOF; analyte solubilization; analyte solvation
Organic additives	Urea	Modification of EOF; protein solubilization; denaturation of oligonucleotides
Sulfonic acids	Hexane, heptane, octane or nonane analogs	Analyte ion pairing; hydrophobic interaction
Divalent amines	Diaminoalkanes, hexamethonium bromide, decamethonium bromide	Modification of EOF; charge neutralization; analyte interaction
Cationic surfactants	Dodecyltrimethylammonium bromide, cetyltrimethylammonium bromide, tetradecyltrimethylammonium chloride	Charge reversal on capillary wall; hydrophobic interaction
Cellulose derivatives	Hydroxyethylcellulose, methylcellulose, hydroxypropylmethylcellulose	Reduce EOF; provide sieving medium

EOF through a dynamic modification of the capillary wall.[58–60] Ion-pairing agents, such as alkylsulfonic acid salts, have been used with success to resolve peptides.[61] Figure 6 demonstrates the influence of HSA on the separation of a series of peptides of identical amino acid composition but differing sequence. Here the resolution appears to be due to the differential solvation and/or the micro-environment of nearest neighbor effects on pK values of the charged amino acids.

3.2.3.2 Organic Solvents Organic solvents, such as methanol or acetonitrile, decrease both the conductivity of the buffer and EOF through their ability to disrupt the ordered structure of the water molecules. Enhancement of resolution may result from a combination of the decreased EOF (i.e. increased on-capillary time), decreased thermal diffusion and improved analyte solubility. In Figure 6(a–c) the effect of acetonitrile concentration on the solvation and resolution of the same series of peptides is shown.[61]

3.2.3.3 Chiral Agents In the pharmaceutical industry, it is important to have the ability to resolve closely related impurities and degradation products in drug formulations. Skanchy et al.[62] explored the separation of potential degradation products of a dipeptide using a novel anionic cyclodextrin, $SBE[4]_{ds}(1–10)_{sr}$-β-CD, as a pseudostationary phase for the CE separation. Evaluating a variety of neutral and anionic cyclodextrins as well as pH and the cyclodextrin concentration, they found that the stereoisomers of an Asp–Phe dipeptide (α-D,L, α-L,D, α-D,D) could be optimally resolved from the α-L,L parent compound using low concentrations of sulfobutyl ether-derivatized β-cyclodextrin at an acidic pH. This is significant in light of the fact that, to date, dipeptide stereoisomers have only been separated with chiral buffer additives or neutral cyclodextrins.

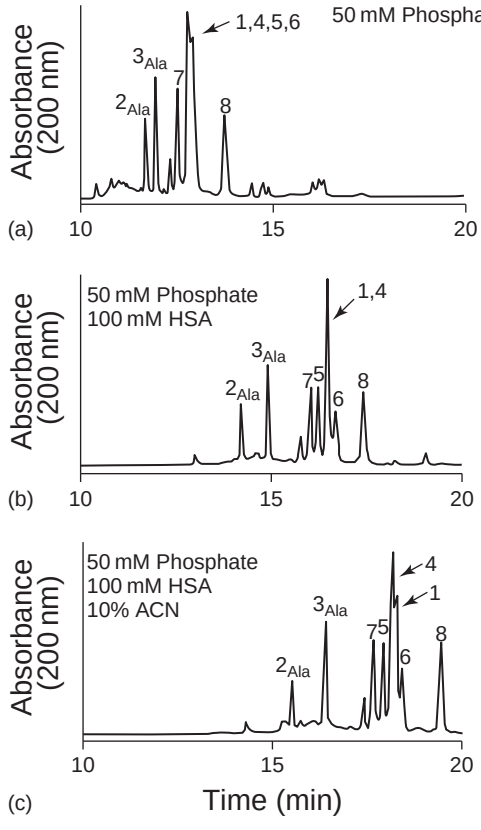

Figure 6 Enhanced peptide resolution afforded by the use of organic additives. The combined effect of acetonitrile and hexanesulfonic acid (HSA) on the separation of the peptides in 50 mM phosphoric acid (pH 2.0). Separation was carried out in a bare fused-silica capillary, 50 μm ID × 57 cm long (50 cm to the detector) at 15 kV and 28 °C with detection at 200 nm. (Reproduced by permission from Oda et al.[61])

3.2.4 Other Modes

3.2.4.1 Micellar Capillary Electrophoresis of Nonglycosylated Peptides MCE is a mode of CE which exploits

the micelle-forming properties of surfactants to partition neutral molecules according to their hydrophobic character. While this mode has typically been used for the analysis of small, drug-like molecules, it has also been used with some success for the resolution of peptides. Grimm et al.[63] improved upon a CZE method used for the separation of urinary imidodipeptides of prolidase-deficient patients exploiting MCE for disease screening. Using 65 mM SDS as the micellar phase in a pH 8.3 borate buffer, MCE separation allowed the collection and sequencing of almost all peaks for definition of the imidodipeptide composition of the urine. Once accomplished, imidodipeptide excretion patterns based on MCE separation may be used for disease diagnosis as well as determining severity. A follow-up paper from this group[64] evaluated different surfactants in an attempt to optimize the separation of the urinary imidodipeptides. After trying three nonionic and four anionic detergents and also varying detergent concentration, pH and temperature, with and without organic solvents, they determined that the urinary imidodipeptides were optimally resolved in a pH 9.3 borate buffer containing 50 mM pentanesulfonate and 10% methanol at 10 °C.

3.2.4.2 Capillary Isoelectric Focusing Several researchers have begun to explore CIEF as a powerful alternative to CZE because it allows for large sample volumes to be loaded on the capillary without adverse effects on resolution. While more typically associated with protein analysis, recent studies have begun to show its utility for peptide analysis. Shimura et al.[65] showed the effectiveness of peptide CIEF when coupled with laser-induced fluorescence (LIF) for sensitive detection. Using a nine-amino-acid peptide containing a cysteine, rhodamine labeling (at the thiol) was followed by trypsin digestion, which generated a six-amino-acid product. CIEF of the tryptic digest allowed the detection of the peptide product in the low picomolar range. Unfortunately, the authors did not demonstrate the application of CIEF with LIF to more complex samples. Taking a different approach, Wu et al.[66] exploited imaged CIEF for the analysis of peptides as a means of circumventing the drawbacks associated with EOF or pressure mobilization of the focused zones in conventional CIEF. Whole-column imaging can reveal these methodological problems and their ramifications, including increased analysis time, uneven resolution and uneven mobilization speed. In this study, they demonstrated the successful imaged focusing of a synthetic peptide, although the significance of this sequence was not defined by the authors. The presence of 20% glycerol appears to be required for effective focusing and while both fluorocarbon- and polyacrylamide-coated capillaries were used in this study, it is not clear which was used in this particular experiment.

3.2.5 Peptide Dimerization

Landers et al.[36] studied cysteine-mediated peptide homo- and heterodimer formation using CZE. Using a low-pH citrate buffer, they were able to monitor the kinetics of homodimer formation with a 28-amino-acid N-terminal cysteine peptide and a 13-amino-acid C-terminal cysteine peptide. Low-pH CZE allowed the deciphering of the time course for formation of the homo- and heterodimers when both peptides were incubated with H_2O_2 with a 12-min separation. LeTourneau and Allen[67] carried out similar studies using glycopeptides where homodimerization correlated directly with the antibacterial activity, e.g. vancomycin and ristocetin A. Using CZE in a neutral phosphate buffer, they were able to evaluate the dimerization of the three antibiotics (vancomycin, ristocetin A and LY264826) in the presence of a bacterial cell wall analog. Based on the mobilities, the glycopeptide antibiotic dimerization constants determined by CE were consistent in trend with those obtained by nuclear magnetic resonance (NMR) spectroscopy and sedimentation equilibrium, but generally smaller in magnitude.

3.2.6 Detection

3.2.6.1 Derivatization In order for CE to be effective for many of the applications involving peptide analysis in biological systems, sensitive detection schemes are needed. This often requires the employment of fluorescence detection and, in particular, LIF detection. Despite the intrinsic fluorescence of the aromatic amino acids (Tyr, Trp, Phe), fluorescent derivatization is often required for sensitive and unbiased detection. Common reagents for derivatization include fluorescein isothiocyanate (FITC), *o*-phthaldialdehyde (OPA), naphthalene-2,3-dicarboxyaldehyde (NDA), fluorescamine and 3-(4-carboxybenzoyl)quinoline-2-carboxaldehyde (CBQCA), while less common derivatization reagents such as fluorescein succinimidyl ester (FSE) have also been tested.

Zhu and Kok[68] studied the reaction kinetics for the postcolumn derivatization of peptides using fluorescamine and compared the labeling, both in effectiveness and detection sensitivity, with OPA. Using an uncoated capillary and a borate buffer at pH 9.5 for peptide separation, fluorescamine was found to be superior to OPA, providing limits of detection of ~0.1 μM. This was approximately 10 times better than the detection limits attainable with UV detection and, consequently, nanomolar to micromolar quantities of peptides are needed.

SP, a basic undecapeptide with a hydrophobic C-terminus, is a member of the tachykinin family, present in the central and peripheral nervous systems, and has a wide range of biological effects. The presence of this protein in brain fluid at picomolar concentrations has been

associated with schizophrenia and Huntington's, Parkinson's and Alzheimer's diseases. In order to elucidate the function of SP and the contribution to its modulatory effects made by its metabolites, Whelpton et al.[69] used CE in combination with HPLC to study the in vivo metabolism of SP in rat striatum. CE was used to separate the fractions from gradient elution HPLC and to avoid the high background signals when detecting at a low wavelength (214 nm) in HPLC. Based on the migration times, SP and significant amounts of its N-terminal fragments, SP (1–7) and SP (1–4), were detected but not the major C-terminal fragments at the concentration studied (10 nmol in the rat by intrastriatal injection). The CE results showed that the metabolism of SP has zero-order elimination kinetics with a rate of decay of 0.2 nmol min^{-1}. The N-terminal metabolites may contribute to the overall pharmacological pattern of activity of the parent peptide as SP (1–4) and SP (1–7) are produced in vivo in the striatum in relatively large amounts. Kostel et al.[70] also studied SP via direct sampling from brain tissue using microdialysis. Pseudo-real-time monitoring of tissue using this technique involves continuous sampling of brain fluid at approximately 1 μL min^{-1} and provides volumes compatible with CE. Using a postcolumn derivatization involving naphthalene-2,3-dicarboxaldehyde-β-mercaptoethanol (NDA-β-ME) and an etched reactor in the membrane, they examined the detection enhancement in response to varying pH and reaction time. They observed a limit of detection of ~75–100 nM for lysine-containing peptides, but only 20–25 μM for non-lysine-containing peptides with 12-s reaction times.

Chen et al.[71] studied the fluorescent labeling of [D-$Pen^{2,5}$]enkephalin (DPDPE), a synthetic opioid peptide used as a highly selective delta-receptor agonist in studies of receptor–ligand binding. Following solid-phase extraction (SPE) of serum samples on C_{18} and fluorescent derivatization with tetramethylrhodamine-5-isothiocyanate isomer G, a CZE/LIF method was used to quantitate the DPDPE in rat serum. The on-column detection limit for this method was 1.3 amol (10^{-18} mol) of DPDPE, much higher than in HPLC (the analysis range is 50–100 μM) and CZE/UV detection (the detection limit is 0.4 μM). Their results show that this CZE/LIF method could be used for the sensitive and specific quantitation of therapeutic peptides in biological matrices.

3.2.6.2 Solid-phase Extraction/Capillary Electrophoresis for On-line Concentration and Analysis of Peptides

An alternative to derivatizing peptides for low-level detection using LIF is to concentrate the solutes of interest from large volumes on-capillary. This can be accomplished using solid-phase extraction/capillary electrophoresis (SPE/CE), which involves nothing more than creating a nanoliquid chromatographic column at the inlet of the capillary. Using as little as 100 μg of C_8 or C_{18} resin packed into a microscale device as illustrated in Figure 7, a bed volume between 50 and 500 nL can be created. Guzman et al.[72,73] led the way in this area, along with contributions from Debets et al.[74] and Merion and co-workers.[75,76] These devices, when fastened to the capillary inlet, can be used to adsorb solutes from 1 to 1000 μL with good efficiency and have been demonstrated to be effective for a variety of analytes.[77] More importantly, SPE/CE has been shown to be extremely effective for peptide analysis at concentrations below the detection limit with standard CE by our group[78,79] and others.[80] Figure 8(a–c) shows the ability of the on-line SPE device for concentrating peptides. A mixture of standard peptides, when diluted below the limit of CE/UV detection (200 nm), can be detected when 20 μL of standard are concentrated on the SPE device and eluted in a ~50 nL volume. Figure 9(a) and (b) shows how this approach is effective for real peptide samples. Reversed-phase high-performance liquid chromatography (RPHPLC) fractions of 100 μL each are analyzed directly by CZE (once the acetonitrile has evaporated, leaving 55 μL). Interestingly, these fractions, containing peptides generated by an in-gel trypsinization of bovine serum albumin (BSA), including fraction 3, would all be expected to contain peptides based on absorbance with RPHPLC. However, two-dimensional analysis with CZE clearly shows peptides in all fractions except 3. This fraction was thought to contain a contaminant extracted from the acrylamide gel.

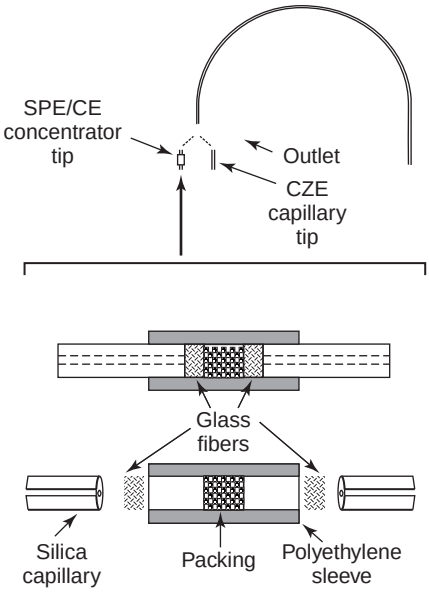

Figure 7 Diagrammatic representation of the construction of a SPE tip.

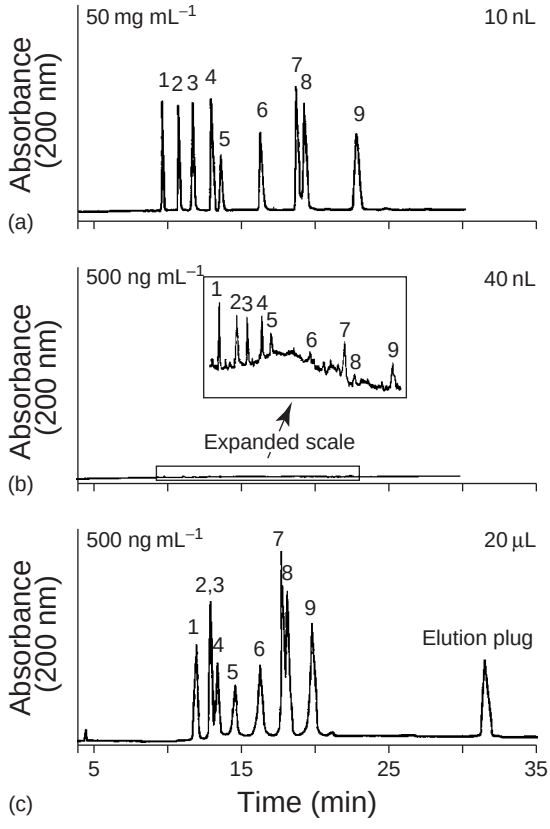

Figure 8 Comparison of standard CE and SPE/CE for the analysis of peptides. (a) Separation of a standard peptide mixture containing (1) bradykinin, (2) angiotensin, (3) α-melanin-stimulating hormone, (4) thyrotropin-releasing hormone, (5) luteinizing hormone-releasing hormone, (6) leucine enkephalin, (7) bombeoin, (8) methionine enkaphalin and (9) oxytocin, all at individual concentrations of 50 mg mL^{-1}. Separation was carried out in a 57-cm capillary containing 50 mM phosphate buffer (pH 2.50) at 20 kV with the capillary thermostated at 25 °C. Sample was pressure injected for 10 s at 0.5 psi. Detection was at 200 nm. (b) Separation of the same mixture of peptides diluted 1 : 99 with 5 mM phosphate buffer (pH 2.50). Separation conditions as in (a). Sample was electrokinetic for 32 s at 8 kV. (c) Separation of the diluted peptide standards with SPE/CE where 20 mL of sample was loaded onto the packing by pressure (20 psi) for 1 min. Separation conditions as in (a). The bar represents 0.020 AU. (Reproduced by permission from Strausbach et al.[78])

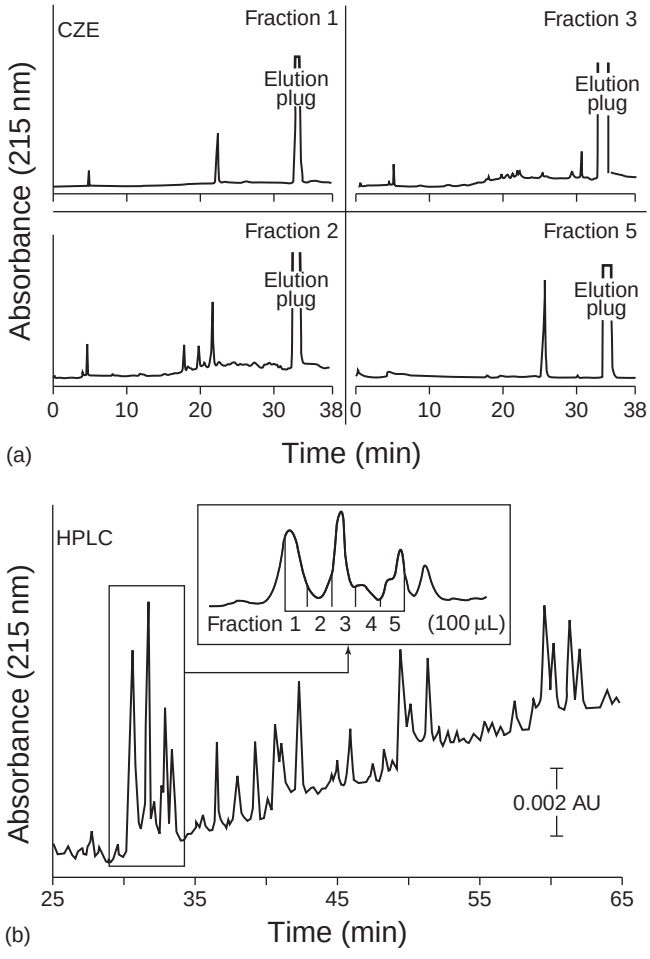

Figure 9 Analysis of entire HPLC fractions in a single pass by CZE. (b) RPHPLC analysis of tryptic peptides from BSA. The peptides released from the tryptic digestion (0.4 µg) of 2 mg of BSA in a polyacrylamide gel slice were separated by HPLC and collected for two-dimensional SPE/CE analysis. The expanded figure identifies the peaks associated with the fractions collected from 31.5 to 34.0 min. (a) SPE/CE analysis of 55-µL HPLC fractions (100 µL prior to removal of acetonitrile) which were loaded onto the SPE capillary at 20 µL min^{-1}. Conditions for the separation were identical to those in Figure 8. (Reproduced by permission from Strausbach et al.[78])

4 POSTTRANSLATIONALLY-MODIFIED PEPTIDES

4.1 Glycopeptides

The biological role of oligosaccharides present in proteins has been subject to investigation since their discovery in the 1950s. These covalent components, originally thought to be artifacts, are now recognized to be important to the proper biological activities of these macromolecules, which include protein stability and solubility, biological recognition and function, antigenicity, transport and clearance. It has become increasingly clear that the intermolecular reactions associated with the function of glycoproteins involve both the carbohydrate and polypeptide portions of the molecule.[81] One aspect of glycoproteins that makes characterization difficult is their tendency to be ''microheterogeneous''. Cells do not synthesize a single glycoprotein, but instead produce a family of glycosylation variants differing in their oligosaccharide content. It is thought that, while each member of this cellular pool of glycoforms may have

a unique biological activity, the molecular characteristics necessary for structural integrity are maintained. As a result, the combined functional activities of the glycoforms provide a continuum of biological activity which the cell can control with high fidelity by varying the relative amount of each glycoform.

From an analytical perspective, glycoproteins form a separate class of proteins whose structure is complicated by the presence of the surface carbohydrates. The oligosaccharides add a second dimension of complexity to the glycoprotein structure, as a result of the types of linkages possible between the oligosaccharide and the protein, as well as the variety of monosaccharides which compose the glycan. Oligosaccharides may be covalently bound to the protein via linkage to the amide nitrogen of the asparagine side chain (N-linked) or through the oxygen of serine, threonine and, occasionally, hydroxyproline (O-linked). In addition, oligosaccharides can be further classified based on the monosaccharides which constitute the glycan. The N-linked oligosaccharides are typically (1) ''high-mannose'', composed of only mannose and *N*-acetylglucosamine, (2) ''complex'', containing a combination of galactose, fucose, neuraminic acid, mannose and glucosamine, or (3) ''hybrid'', composed of the elements of both the high-mannose and complex oligosaccharides. O-Linked oligosaccharides generally have fewer branches than their N-linked counterparts, but have more diversity with respect to the monosaccharide units which constitute them. These include glucose, xylose, *N*-acetylgalactosamine and arabinose in addition to the monosaccharides found in the N-linked structures.

The chemical structure of oligosaccharides presents some unique analytical problems in comparison with other biologically active molecules, whether the intact glycoprotein (and its glycoforms) or the glycopeptides (formed from selective enzymatic digestion of the protein) are analyzed, or whether just the oligosaccharides are analyzed following cleavage from the protein. The molar absorptivities for carbohydrates are remarkably low in the UV range and, therefore, preclude them from detection by fluorescence and absorbance methods, the most universal detection modes. Furthermore, saccharides are generally similar in structure, being polar and uncharged, and thus are difficult to separate in an unaltered form by RPHPLC or supercritical fluid chromatography (SFC). Oligosaccharides can be effectively resolved with a number of standard methods once appropriately derivatized: HPLC[82–94] and gas chromatography[95–99] using absorbance,[100,101] refractive index,[91] fluorescence[102,103] and electrochemical[86–94] detection schemes have been effectively used. The high resolving power of CE was recognized as a means to resolve individual glycoforms, providing the glycobiologist or carbohydrate chemist with the ability to answer many of the structural and biological questions that surround glycoprotein science. These studies made use of the extensive structural work done over the past several decades by glycobiologists. However, the difficulty in resolving the glycoforms of well-characterized glycoproteins has highlighted the idiosyncratic nature of some separations.

Taverna et al.[58] analyzed the glycopeptides obtained by tryptic digestion of rt-PA and isolated two glycopeptide fractions by RPHPLC. CZE of glycoprotein fraction 2 (GP-2) in a bare silica capillary gave the same elution pattern whether the buffer was 150 mM borate (pH 9.1) or 100 mM phosphate (pH 6.6), suggesting that minimal borate complexation occurred due to terminal sialic acids. The glycopeptide fraction 3 (GP-3) was best resolved in a 50 mM tricine buffer (pH 8.5) containing 2.5 mM diaminobutane (DAB); the authors suggested that the improved resolution was due to the slowing of EOF by the divalent DAB.

Wu[104] used CZE as a second dimension for the two-dimensional analysis of complex peptides resulting from the tryptic digest of recombinant tissue plasminogen activator. The first dimension, RPHPLC, lacked the resolution required to resolve the microheterogeneous glycopeptides, a task perfectly suited to CZE. Although high-performance anion-exchange chromatography (HPAEC) is capable of such separations, it suffers limitations in that only N-linked carbohydrates are separated and relatively large amounts of sample are required.

4.2 Micellar Capillary Electrophoresis of Glycopeptides

As mentioned earlier, MCE can be exploited for the separation of neutral molecules according to their hydrophobic character. This mode has also seen use for the resolution of glycopeptides, owing to the chromatographic partitioning of the neutral glycans into the micelle.

In the analysis of interferon-γ glycopeptides by CE, James et al.[105] separated the glycoforms in a bare silica capillary using 400 mM borate–100 mM SDS (pH 8.5). The proteins separated into three groups: PG1, containing the diglycosylated forms (Asn 25 and Asn 97); PG2, containing the monogylcosylated protein (Asn 25); and PG3, the nonglycosylated protein. Borate buffers in the pH range 8.5–9 gave the best peak resolution and shape, although the glycoforms were not resolved without the addition of SDS. Further characterization of the protein by PNGase F digestion resulted in a loss of complexity in the PG1 region, with an increase in complexity of pattern in the PG2 area. Longer digestion led to loss of peaks in PG2 with increase in peaks at PG3, and a final collapse of PG3 into a single, broad band, suggesting denaturation or proteolytic degradation. Some lower molecular mass

components were detected by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE), suggesting the presence of a contaminating protease in the PNGase F preparation. Neuraminidase altered the complexity within a PG group, reducing charge heterogeneity without changing the basic three groups, indicating maintenance of structural variants between the fractions.

Attempts to separate glycoforms of other glycoproteins again indicate the idiosyncratic nature of glycoprotein resolution. James and co-workers[104] were able to separate the glycoforms of ribonuclease (RNAse) B and horseradish peroxidase, but were unable to resolve those of fetuin or α_1-acid glycoprotein (AGP) in 400 mM borate–100 mM SDS (pH 8.5), suggesting that terminal sialic acid residues might prevent partitioning into the micelle, thereby precluding resolution with this additive.

Glycopeptides have been relatively easier to separate. Nashabeh and El Rassi[106] investigated the CE separation of the glycopeptides obtained from AGP released by tryptic digestion. A complex pattern of peaks was observed following electrophoresis in a polyethylene glycol (PEG)-coated capillary containing 100 mM phosphate (pH 5.0). Parts of the electropherogram were correlated with fractions obtained by chromatography of the digest on a concanavalin A–Sepharose column. The complex pattern was attributed to glycosyl heterogeneity without identification of structures.

The retardation of electrophoretic migration of peptides containing glycosyl groups is highlighted in an elegant study by Rush et al.,[107] who produced a tryptic map of the glycopeptides of recombinant human erythropoietin. They were able to resolve 30 peaks in 80 min using 40 mM phosphate (pH 2.5), containing 100 mM heptanesulfonic acid as an ion-pairing agent (Figure 10). The electrophoretic pattern divided the peptides into nonglycosylated peptides migrating within 35 min and a glycosylated fraction from 40 to 80 min. Sialydase treatment resulted in peaks within the glyco window migrating faster, but still beyond 40 min, indicating that the neutral glycan retards the mobility of the glycan-containing peptides. *N*-Glycanase treatment resulted in a shift of peptides into the aglycone window (<40 min), as would be expected. Combining high-performance CE and HPLC as two-dimensional techniques, the individual glycosylation sites, Asn 83, Asn 24 and Asn 38, and an O-linked site were mapped. The authors suggested a strategy for the identification of glycosylation sites by a combination of RPHPLC and CE and possible assignment of structural identity of glycoforms based on the alterations in migration resulting from glycosidase digestion. This is certainly feasible based on the development of instrumentation for rapid two-dimensional HPLC/CE analysis by Larmann et al.[108] Sampling directly from a one-dimensional RPHPLC separation at regular intervals with CE analysis completed on the low seconds timescale in the second dimension, analysis of fluorescently labeled peptides from a tryptic digestion of horse heart cytochrome *c* was possible. Clearly, this type of analysis would be extremely useful for glycoprotein characterization.

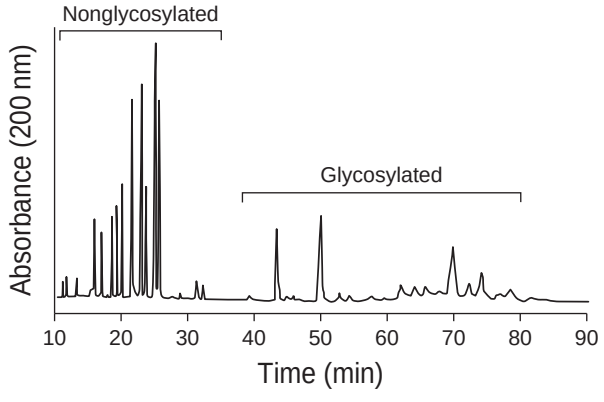

Figure 10 CE separation of peptides and glycopeptides from recombinant human erythropoietin. The group of peaks with the faster mobility represent nonglycosylated peptides, and those with retarded migration are glycosylated. Experimental conditions: 40 mM sodium phosphate buffer with 100 mM heptanesulfonic acid at pH 2.5; 16 kV running voltage, 110 μA; 50 μm × 75 cm (50 cm to the detector). (Reproduced by permission from Rush et al.[107])

Rudd et al.[109] isolated five peaks, corresponding to the individual manno-5- through manno-9-containing glycoforms, from RNAse B. Analysis of oligosaccharides obtained by hydrazinolysis gave a pattern of five peaks, corresponding to the manno-5-through manno-9-glycomers. Using isolated fractions of RNAse B, Rudd et al.[110] investigated the structure–function relation of the glycosylation. Enzymatic digestion by *A. satoi* α-mannosidase reduced manno-5 to -9 (RNAse B5–9) glycomers to manno-5 (RNAse B5). RNAse B manno-1 (RNAse B1) was obtained by digestion with jack bean α-mannosidase. Mannosidase from *Helix pomatia* was used to produce a manno-0 (RNAse B0) isoform containing just the core *N*-acetylglucosamine residues. All glycoforms were separated on 50 μm × 100 cm bare silica capillary in 20 mM phosphate–50 mM SDS–5 mM borate buffer (pH 7.2) using a discontinuous voltage ramp, and collected preparatively. The isolated protein fractions showed no gross alterations in structure as reflected by H/D exchange rates monitored by NMR spectroscopy. Pronase proteolysis suggested no gross structural changes, as all glycoforms were more resistant to proteolytic digestion with a $t_{1/2}$ of 450 min for the glycosylated RNAses compared with 300 min for RNAse A. Comparing RNAse activity on whole-cell RNA, the glycoforms were more active: RNAse A (1.0, relative activity) : RNAse B0 (1.6) = RNAse B1

(2.2) : RNAse B5 (3.6) = RNAse B5–9 (3.8). These studies support the hypothesis that there might be a range of functional activity in the cell provided by glycoforms of an enzyme. Rudd et al.[110] noted that the individual glycomers might have differential activity against purified fractions of RNA; this hypothesis was not tested.

5 CONCLUDING REMARKS

As the Proteome Project enters our scope as the latest challenge for researchers in a quest to understand cellular processes in normal and pathological states, the ability to perform rapid and information-rich analysis of peptides will become critical. CE is a key part of the analytical arsenal that will be used to interrogate the complex array of components constituting the proteome. While CE has been established as a bona fide technique for peptide analysis, it is clear that we have not yet fully exploited this technology for application in the vast arena defined as peptide and protein analysis. The potential clearly exists for CE to provide a rapid means for not only interrogating peptide primary structure, but also for characterizing the myriad of post-translation modifications (e.g. phosphorylation, ribosylation, myristolation, methylation and hydroxylation). These capabilities will surely be enhanced by the effective interfacing of CE with other techniques. For example, Severs and Smith[112] have already demonstrated the power of MS as a detection method for CE in peptide analysis, and Moore et al.[111] have demonstrated, convincingly, the power of CE as part of a two-dimensional analysis of peptide mixtures. As these and other technological advances mature, as method development and instrumentation improve and as approaches for sensitive detection both in hardware and chemistry improve, CE-based peptide analysis will evolve into a robust, turnkey operation applied to a diverse array of applications.

ABBREVIATIONS AND ACRONYMS

AGP	α_1-Acid Glycoprotein
BSA	Bovine Serum Albumin
CBQCA	3-(4-Carboxybenzoyl)quinoline-2-carboxaldehyde
CE	Capillary Electrophoresis
CGE	Capillary Gel Electrophoresis
CIEF	Capillary Isoelectric Focusing
CITP	Capillary Isotachophoresis
CZE	Capillary Zone Electrophoresis
DAB	Diaminobutane
DPDPE	[D-Pen^{2,5}]enkephalin
EOF	Electroosmotic Flow
FITC	Fluorescein Isothiocyanate
FSE	Fluorescein Succinimidyl Ester
HETP	Height Equivalent to a Theoretical Plate
HPAEC	High-performance Anion-exchange Chromatography
HPLC	High-performance Liquid Chromatography
HSA	Hexanesulfonic Acid
ITP	Isotachophoresis
LIF	Laser-induced Fluorescence
MCE	Micellar Capillary Electrophoresis
MS	Mass Spectrometry
NDA	Naphthalene-2,3-dicarboxyaldehyde
NDA-β-ME	Naphthalene-2,3-dicarboxaldehyde-β-mercaptoethanol
NMR	Nuclear Magnetic Resonance
OPA	*o*-Phthaldialdehyde
PEG	Polyethylene Glycol
RNAse	Ribonuclease
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SFC	Supercritical Fluid Chromatography
SP	Substance P
SPE	Solid-phase Extraction
SPE/CE	Solid-phase Extraction/Capillary Electrophoresis
TFE	2,2,2-Trifluoroethanol
TSH	Thyroid Stimulating Hormone
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography

REFERENCES

1. J.J. Jorgenson, K.D. Lukacs, 'Zone Electrophoresis in Open Tubular Glass Capillaries', *Anal. Chem.*, **53**, 1298–1302 (1981).
2. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations With Micellar Solution and Open-tubular Capillaries', *Anal. Chem.*, **56**, 111–113 (1984).
3. R. Rodriguez-Diaz, T. Wehr, M. Zhu, V. Levi, 'Capillary Isoelectric Focusing' in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 101–138, 1997.
4. B.J. Wanders, F.M. Everaerts, 'Isotachophoresis in Capillary Electrophoresis' in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 111–128, 1993.
5. K.J. Ulfelder, B.R. McCord, 'Separation of DNA by Capillary Electrophoresis' in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 347–378, 1997.
6. S. Terabe, K. Otsuka, T. Ando, 'Electrokinetic Chromatography with Micellar Solution and Open-tubular Capillaries', *Anal. Chem.*, **57**, 834–841 (1985).
7. K.Otsuka, S. Terabe, T.Ando, 'Electrokinetic Chromatography with Micellar Solutions: Retention Behavior and Separation of Chlorinated Phenols', *J. Chromatogr.*, **348**, 39–47 (1985).
8. R. Abersold, H.D. Morrison, 'Analysis of Dilute Peptide Samples by Capillary Zone Electrophoresis', *J. Chromatogr.*, **516**, 79–88 (1990).
9. D.S. Burgi, R.-L. Chien, 'On-line Sample Preconcentration for Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 479–493, 1997.
10. S. Hjerten, 'Zone Sharpening in Paper Electrophoresis – A Method Allowing Application of Dilute Protein Solutions', *Biochim. Biophys. Acta*, **32**, 531–534 (1959).
11. R.-L. Chien, D.S. Burgi, 'On-column Sample Concentration Using Field Amplification in CZE', *Anal. Chem.*, **64**, 489A–496A (1992).
12. M. Castognola, D.V. Rossetti, M. Corda, M. Pellegrini, F. Misiti, A. Olianas, B. Giardina, I. Messana, 'Predictive Model of CE Peptide Mobility in 2,2,2-Trifluoroethanol–Water Solution', *Electrophoresis*, **19**, 1728–1732 (1998).
13. E. Grushka, R.M. McCormick, J.J. Kirkland, 'Effect of Temperature Gradients on the Efficiency of Capillary Zone Electrophoresis Separations', *Anal. Chem.*, **61**, 241–246 (1989).
14. W.A. Gobie, C.F. Ivory, 'Thermal Model of Capillary Electrophoresis and a Method for Counteracting Thermal Band Broadening', *J. Chromatogr.*, **516**, 191–210 (1990).
15. J.M. Davis, 'Influence of Thermal Variations of Diffusion Coefficient on Non-equilibrium Plate Height in Capillary Zone Electrophoresis', *J. Chromatogr.*, **517**, 521–547 (1990).
16. R.-L. Chien, J.C. Helmer, 'Electroosmotic Properties and Peak Broadening in Field-amplified Capillary Electrophoresis', *Anal. Chem.*, **63**, 1354–1361 (1990).
17. Y. Kurosu, K. Hibi, T. Sasaki, M. Saito, 'Influence of Temperature Control in Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **14**, 200–203 (1991).
18. A. Guttman, J. Horvath, N. Cooke, 'Influence of Temperature on the Sieving Effect of Different Polymer Matrices in Capillary SDS Gel Electrophoresis of Proteins', *Anal. Chem.*, **65**, 199–203 (1993).
19. R.J. Nelson, A. Paulus, A.S. Cohen, G.A. Guttman, B.L. Karger, 'Use of Peltier Thermoelectric Devices to Control Column Temperature in High-performance Capillary Electrophoresis', *J. Chromatogr.*, **480**, 111–127 (1989).
20. F. Foret, E. Szoko, B.L. Karger, 'Trace Analysis of Proteins by Capillary Zone Electrophoresis with In-column Transient Isotachophoresis Preconcentration', *Electrophoresis*, **14**, 417–428 (1993).
21. E.C. Rickard, M.M. Strohl, R.G. Nielsen, 'Correlation of Electrophoretic Mobilities From Capillary Electrophoresis with Physicochemical Properties of Proteins and Peptides', *Anal. Biochem.*, **197**, 197–207 (1991).
22. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, 2nd edition, Wiley-Interscience, New York, 1979.
23. J.C. Giddings, 'Generation of Variance, '"Theoretical Plates'', Resolution, and Peak Capacity in Electrophoresis and Sedimentation', *Sep. Sci.*, **4**, 181–189 (1969).
24. J.C. Sternberg, 'Extracolumn Contributions To Chromatographic Band Broadening', *Adv. Chromatogr.*, **2**, 205–270 (1966).
25. L. Szepesy, *Gas Chromatography*, CRC Press, Boca Raton, FL, 1970.
26. X. Huang, W.F. Coleman, R.N. Zare, 'Analysis of Factors Causing Peak Broadening in Capillary Zone Electrophoresis', *J. Chromatogr.*, **480**, 95–110 (1989).
27. H.K. Jones, N.T. Nguyen, R.D. Smith, 'Variance Contributions to Band Spreading in Capillary Zone Electrophoresis', *J. Chromatogr.*, **504**, 1–19 (1990).
28. R. Weber, 'Concerning Theories of Electrophoretic Separation in Porous Tubes', *Helv. Chim. Acta*, **36**, 424 (1953).
29. J.P. Landers, R.P. Oda, B. Madden, T.P. Sismelich, T.C. Spelsberg, 'Reproducibility of Sample Separation Using Liquid or Forced Air Convection Themostated High Performance Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **15**, 517–525 (1992).
30. D. McManigill, S.A. Swedberg, 'Factors Affecting Plate Height in High Performance Zonal Capillary Electrophoresis (Hpzce)', in *Techniques in Protein Chemistry*, ed. T. Hugli, Academic Press, San Diego, CA, 468–478 (1989).

31. R.E. Offord, 'Electrophoretic Mobilities of Peptides on Paper and Their Use in the Determination of Amide Groups', *Nature (London)*, **211**, 591–593 (1966).
32. F. Nyberg, M.-D. Zhu, J.-L. Liao, S. Hjerten, 'High Performance Electrophoresis in Studies of Substance P Degradation', in *Electrophoresis '88*, ed. C. Schafer-Neilsen, VCH, Weinheim, 141–150 (1988).
33. Z. Deyl, V. Rohlicek, R. Struzinsky, 'Some Rules Applicable to Capillary Zone Electrophoresis of Peptides and Proteins', *J. Liq. Chromatogr.*, **12**(13), 2515–2526 (1989).
34. Z. Deyl, V. Rohlicek, M. Adam, 'Separation of Collagens by Capillary Zone Electrophoresis', *J. Chromatogr.*, **480**, 371–378 (1989).
35. Y. Kurosu, Y. Sato, M. Senda, 'Investigation of Migration Behaviors of Peptides and Proteins by Capillary Zone Electrophoresis', *Anal. Sci.*, **7**(Suppl. Proc. Int. Congr. Anal. Sci.), 273–276 (1991).
36. J.P. Landers, R.P. Oda, J.A. Liebenow, T.C. Spelsberg, 'The Utility of High Resolution Capillary Electrophoresis for Monitoring Peptide Homo- and Hetero-dimer Formation', *J. Chromatogr.*, **652**(1), 109–117 (1993).
37. J.R. Florance, Z.D. Konteatis, M.J. Macielag, R.A. Lessor, A. Galdes, 'Capillary Zone Electrophoresis Studies of Motilin Peptides. Effect of Charge, Hydrophobicity, Secondary Structure and Length', *J. Chromatogr.*, **559**(1–2), 391–399 (1991).
38. T.A.A.M. Van de Goor, P.S. Janssen, J.W. van Nispen, M.J. van Zeeland, F.M. Everarts, 'Capillary Electrophoresis of Peptides: Analysis of Adrenocorticotropic Hormone Related Fragments', *J. Chromatogr.*, **545**(2), 379–389 (1991).
39. N.J. Adamson, E.C. Reynolds, 'High Performance Capillary Electrophoresis of Casein Phosphopeptides Containing 2–5 Phosphoserine Residues: Relationship Between Absolute Electrophoretic Mobility and Peptide Charge and Size', *Electrophoresis*, **16**(4), 525–528 (1995).
40. P.D. Grossman, J.C. Colburn, H.H. Lauer, 'A Semiempirical Model for the Electrophoretic Mobilities of Peptides in Free Solution Capillary Electrophoresis', *Anal. Biochem.*, **179**(1), 28–33 (1989).
41. H.J. Issaq, G.M. Janini, I.Z. Atamna, G.M. Muschik, J. Lukszo, 'Capillary Electrophoresis Separation of Small Peptides: Effect of pH, Buffer Additives and Temperature', *J. Liq. Chromatogr.*, **15**(6–7), 1129–1142 (1992).
42. N. Chen., L. Wang, Y. Zhang, 'Electrophoretic Selectivity as a Function of Operating Parameters in Free-solution Capillary Electrophoresis Separation of Dipeptides', *J. Liq. Chromatogr.*, **16**(17), 3609–3622 (1993).
43. M.A. Survay, D.M. Goodall, S.A.C. Wren, R.C. Rowe, 'Oligoglycines and Oligoalanines as Tests for Modelling Mobility of Peptides in Capillary Electrophoresis', *J. Chromatogr.*, **636**(1), 81–86 (1993).
44. M. Castagnola, L. Cassiano, R. Rabino, D.V. Rossetti, F.A. Bassi, 'Peptide Mapping Through the Coupling of Capillary Electrophoresis and High-performance Liquid Chromatography: Map Prediction of the Tryptic Digest of Myoglobin', *J. Chromatogr.*, **572**(1–2), 51–58 (1991).
45. V.J.J. Hilser, G.D. Worosila, S.E. Rudnick, 'Protein and Peptide Mobility in Capillary Zone Electrophoresis. A Comparison of Existing Models and Further Analysis', *J. Chromatogr.*, **630**(1–2), 329–336 (1993).
46. B.J. Compton, 'Electrophoretic Mobility Modeling of Proteins in Free Zone Capillary Electrophoresis and Its Application to Monoclonal Antibody Microheterogeneity Analysis', *J. Chromatogr.*, **559**, 357–366 (1991).
47. B.J. Compton, E.A. O'Grady, 'The Role of Charge Suppression and Ionic Strength in Free Zone Electrophoresis of Proteins', *Anal. Chem.*, **63**, 2597–2602 (1991).
48. A. Cifuentes, H. Poppe, 'Simulation and Optimization of Peptide Separation by Capillary Electrophoresis', *J. Chromatogr. A.*, **680**(1), 321–340 (1994).
49. A. Cifuentes, H. Poppe, 'Effect of pH and Ionic Strength of Running Buffer on Peptide Behavior in Capillary Electrophoresis: Theoretical Calculation and Experimental Evaluation', *Electrophoresis*, **16**(4), 516–524 (1995).
50. B. Skoog, A. Wichmann, 'Calculation of the Isoelectric Points of Polypeptides from the Amino Acid Composition Trends', *Anal. Chem.*, **5**(4), 82–85 (1986).
51. J.D. Pennio, 'Free Zone Capillary Electrophoresis. A Simple Computer Program for Predicting Peptide Migration Times', *Biopharmacology*, **2**(8), 41–44 (1989).
52. M. Castognola, D.V. Rossetti, M. Corda, M. Pellegrini, F. Misiti, A. Olianas, B. Giardina, I. Messana, 'The pH Dependence of Predictive Models Relating Electrophoretic Mobility to Peptide Chemico-physical Properties in CZE', *Electrophoresis*, **19**, 2273–2277 (1998).
53. P.D. Grossman, J.C. Colburn, H.H. Lauer, R.G. Nielsen, R.M. Riggin, G.S. Sittampalam, E.C. Rickard, 'Application of Free-solution Capillary Electrophoresis to the Analytical Scale Separation of Proteins and Peptides', *Anal. Chem.*, **61**(11), 1186–1194 (1989).
54. R.P. Oda, J.P. Landers, 'Introduction to Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 1–48, 1997.
55. F.A. Chen, L. Kelly, R. Palmieri, R. Biehler, H. Schwartz, 'Use of High Ionic Strength Buffers for High-performance Capillary Electrophoresis', *J. Liq. Chromatogr.*, **15**, 1143–1161 (1992).
56. M.J. Sepaniak, R.O. Cole, 'Column Efficiencies in Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **59**, 472–476 (1987).
57. J.P. Landers, R.P. Oda, T.C. Spelsberg, J.A. Nolan, K.J. Ulfelder, 'Capillary Electrophoresis: a Powerful Microanalytical Technique for Biologically Active Molecules', *Biotechniques*, **14**, 98–111 (1993).

58. M. Taverna, A. Baillet, D. Biou, M. Schluter, R. Werner, D. Ferrier, 'Analysis of Carbohydrate-mediated Heterogeneity and Characterization of N-Linked Oligosaccharides of Glycoproteins by High-performance Capillary Electrophoresis', *Electrophoresis*, **13**, 359–366 (1992).
59. J.P. Landers, R.P. Oda, B.J. Madden, T.C. Spelsberg, 'High-performance Capillary Electrophoresis of Glycoproteins: the Use of Modifiers of Electroosmotic Flow for Analysis of Microheterogeneity', *Anal. Biochem.*, **205**, 115–124 (1992).
60. R.P. Oda, B.J. Madden, T.C. Spelsberg, J.P. Landers, 'α,w-Bis-quaternary Ammonium Alkanes as Effective Buffer Additives for Enhanced Capillary Electrophoretic Separation of Glycoproteins', *J. Chromatogr.*, **680**, 85–92 (1994).
61. R.P. Oda, B.J. Madden, J.C. Morris, T.C. Spelsberg, J.P. Landers, 'Multiple-buffer-additive Strategies for Enhanced Electrophoretic Separation of Peptides', *J. Chromatogr.*, **680**, 341–351 (1994).
62. D.J. Skanchy, R. Wilson, T. Poh, G.H. Xie, C.W. Demarest, J.F. Stobaugh, 'Resolution of Acylated Dipeptide Stereoisomers by CE Using Sulfobutylether Derivatized Beta-cyclodextrin', *Electrophoresis*, **18**(98), 85–95 (1997).
63. R. Grimm, G. Zanaboni, S. Viglio, K. Dyne, M. Valli, G. Cetta, P. Iadarola, 'Complete Resolution of Imidodipeptide Mixtures in Urine of Prolidase-deficient Patients Using Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **768**, 57–66 (1997).
64. R. Grimm, G. Zanaboni, S. Viglio, K.M. Dyne, G. Cetta, P. Iadarola, 'Effect of Different Surfactants on the Separation by Micellar Electrokinetic Chromatography of a Complex Mixture of Dipeptides in Urine of Prolidase-deficient Patients', *J. Chromatogr. B*, **698**(1–2), 47–57 (1997).
65. K. Shimura, H. Matsumoto, K. Kasai, 'Assay of Trypsin Activity by Capillary Isoelectric Focusing with Laser-induced Fluorescence Detection', *Electrophoresis*, **19**(13), 2296–2300 (1998).
66. J. Wu, S.C. Li, A. Watson, 'Optimizing Separation Conditions for Proteins and Peptides Using Imaged Capillary Isoelectric Focusing', *J. Chromatogr. A*, **817**, 163–171 (1998).
67. D.L. LeTourneau, N.E. Allen, 'Use of Capillary Electrophoresis to Measure Dimerization of Glycopeptide Antibiotics', *Anal. Biochem.*, **246**(1), 62–66 (1997).
68. R. Zhu, W.T. Kok, 'Post-column Derivatization for Fluorescence and Chemiluminescence Detection in Capillary Electrophores', *J. Pharm. Biomed. Anal.*, **17**, 985–999 (1998).
69. R. Whelpton, A.T. Michael-Titus, S.M. Stephens, K.W. Yau, D. Fengas, 'Identification of Substance P Metabolites Using a Combination of Reversed-phase High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr. B*, **716**(1–2), 95–106 (1998).
70. K.L. Kostel, S.M. Lunte, 'Evaluation of Capillary Electrophoresis with Post-column Derivatization and Laser-induced Fluorescence Detection for the Determination of Substance P and its Metabolites', *J. Chromatogr. B*, **695**(1), 27–38 (1997).
71. C. Chen, D. Jeffery, J.W. Jorgenson, M.A. Moseley, G.M. Pollack, 'Sensitive Analysis of [D-Pen^{2,5}]enkephalin in Rat Serum by Capillary Electrophoresis and Laser-induced Fluorescence Detection', *J. Chromatogr. B*, **697**(1–2), 149–162 (1997).
72. N.A. Guzman, M.A. Trebilcok, J.P. Advis, *J. Liq. Chromatogr.*, **14**(5), 997–1015 (1991).
73. N.A. Guzman, *US Pat.*, 5 045 172, 1991.
74. A.J.J. Debets, M. Mazereeuw, W.H. Voogt, D.J. Van Iperen, H. Lingeman, K.-P. Hupe, U.A.Th. Brinkman, *J. Chromatogr.*, **608**, 151–158 (1992).
75. M. Merion, M.E. Swartz, 'On-line Sample Preconcentration on a Packed-inlet Capillary for Improving the Sensitivity of Capillary Electrophoretic Analysis of Pharmaceuticals', *J. Chromatogr.*, **632**, 209–213 (1993).
76. M. Fuchs, M. Merion, 'Apparatus for Effecting Capillary Electrophoresis', *US Pat.*, 5 246 577, 1993.
77. M.A. Strausbauch, S.Z. Xu, J.E. Ferguson, M. Nunez, D. Machacek, G.M. Lawson, P.J. Wettstein, J.P. Landers, 'Concentration and Analysis of Hypoglycemic Drugs Using Solid Phase Extraction–Capillary Electrophoresis', *J. Chromatogr.*, **717**, 279–291 (1995).
78. M.A. Strausbauch, B.J. Madden, P.J. Wettstein, J.P. Landers, 'Sensitivity Enhancement and Second Dimensional Information from the SPE/CE Analysis of Entire HPLC Fractions', *Electrophoresis*, **16**, 541–548 (1995).
79. M. Strausbauch, J.P. Landers, P.J. Wettstein, 'Mechanism of Peptide Separations by Solid Phase Extraction/Capillary Electrophoresis (SPE/CE) at Low pH', *Anal. Chem.*, **68**(2), 306–314 (1996).
80. A.J. Tomlinson, W.D. Braddock, L.M. Benson, R.P. Oda, S. Naylor, 'Preliminary Investigations of Preconcentration–Capillary Electrophoresis/Mass Spectrometry', *J. Chromatogr. B*, **669**(1), 67–73 (1995).
81. T.W. Rademacher, R.B. Parekh, R.A. Dwek, 'Glycobiology' *Annu. Rev. Biochem.*, **57**, 785–832 (1988).
82. N. Takemoto, S. Hase, T. Ikenaka, 'Microquantitative Analysis of Neutral and Amino Sugars as Fluorescent Pyridylamino Derivatives by High-performance Liquid Chromatography', *Anal. Biochem.*, **145**, 245–250 (1985).
83. N. Jenthoft, 'Analysis of Sugars in Glycoproteins by High-pressure Liquid Chromatography', *Anal. Biochem.*, **148**, 424–433 (1987).
84. N. Tomiya, M. Kurono, H. Ishihara, S. Tejima, S. Endo, Y. Arata, N. Takahashi, 'Structural Analysis of N-linked Oligosaccharides by a Combination of Glycopeptidase, Exoglycosidases, and High-performance Liquid Chromatography', *Anal. Biochem.*, **163**, 489–499 (1987).

85. K. Kaheki, S. Honda, 'Analysis of Carbohydrates in Glycoproteins by High-performance Liquid Chromatography and High-performance Capillary Electrophoresis', in *Methods in Molecular Biology*, Vol. 14, *Glycoprotein Analysis in Biomedicine*, ed. E.F. Hounsel, Humana Press, Totowa, NJ, 1993.
86. L.J. Basa, M.W. Spellman, 'Analysis of Glycoprotein-derived Oligosaccharides by High-pH Anion-exchange Chromatography', *J. Chromatogr.*, **499**, 205–220 (1990).
87. M.R. Hardy, R.R. Townsend, Y.C. Lee, 'Monosaccharide Analysis of Glycoconjugates by Anion Exchange Chromatography with Pulsed Amperometric Detection', *Anal. Biochem.*, **170**, 54–62 (1988).
88. R.R. Townsend, M.R. Hardy, D.A. Cunningham, J.P. Carver, B. Bendiak, 'Separation of Branched Sialylated Oligosaccharides Using High-pH Anion-exchange Chromatography with Pulsed Amperometric Detection', *Anal. Biochem.*, **182**, 1–8 (1989).
89. K.R. Anumula, P.B. Taylor, 'Rapid Characterization of Asparagine-linked Oligosaccharides Isolated from Glycoproteins Using a Carbohydrate Analyzer', *Eur. J. Biochem.*, **195**, 269–280 (1991).
90. P. Hermentin, R. Witzel, R. Doenges, R. Bauer, H. Haupt, T. Patel, R.B. Parekh, D. Brazel, 'The Mapping by High-pH Anion-exchange Chromatography with Pulsed Amperometric Detection and Capillary Electrophoresis of the Carbohydrate Moieties of Human Plasma α_1-Acid Glycoprotein', *Anal. Biochem.*, **206**, 419–429 (1992).
91. T.J. Paskach, H.-P. Lieker, P.J. Reilly, K. Thielecke, 'High-performance Anion-exchange Chromatography of Sugar and Sugar Alcohols on Quaternary Ammonium Resins Under Alkaline Conditions', *Carbohydr. Res.*, **215**, 1–14 (1991).
92. J.S. Rohrer, J.D. Olechno, 'Secondary Isotope Effect: the Resolution of Deuterated Glucoses by Anion-exchange Chromatography', *Anal. Chem.*, **64**, 914–916 (1992).
93. P. Lee, M.T. Bunker, 'Carbohydrate Analysis by Ion Chromatography', *J. Chromatogr. Sci.*, **27**, 496–503 (1989).
94. R. Hermentin, J.F.G. Witzel, J.P. Liegenthart, M. Kamerling, H. Nimtz, S. Conradt, 'A Strategy for the Mapping of *N*-Glycans by High-pH Anion-exchange Chromatography with Pulsed Amperometric Detection', *Anal. Biochem.*, **203**, 281–289 (1992).
95. M.F. Chaplin, 'A Rapid and Sensitive Method for the Analysis of Carbohydrate Components in Glycoproteins by Gas–Liquid Chromatography', *Anal. Biochem.*, **123**, 336–341 (1982).
96. D.M. Sheeley, V.N. Reinhold, 'Characterization of N-Linked Glycans by Supercritical Fluid Chromatography', *Anal. Biochem.*, **193**, 240–247 (1991).
97. Y. Leroy, J. Lemoine, G. Ricart, J.-C. Michalski, J. Montreuil, B. Fournet, 'Separation of Oligosaccharides by Capillary Supercritical Fluid Chromatography and Analysis by Direct Coupling High Resolution Mass Spectrometer: Application to Analysis of Oligomannosidic *N*-Glycans', *Anal. Biochem.*, **184**, 235–243 (1990).
98. V.N. Reinhold, D.M. Sheeley, J. Kuei, G.-R. Her, 'Analysis of High Molecular Weight Samples on a Double-focusing Magnetic Sector Instrument by Supercritical Fluid Chromatography/Mass Spectrometry', *Anal. Chem.*, **60**, 2719–2722 (1988).
99. T.L. Chester, D.P. Innis, 'Separation of Oligo- and Polysaccharides by Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **9**, 209–212 (1986).
100. S. Honda, E. Akao, S. Suzuki, M. Okuda, K. Kaheki, J. Nakamura, 'HPLC of Reducing Carbohydrates as Strongly Ultraviolet-absorbing and Electrochemically Sensitive 1-Phenyl-3-methyl-5-pyrazolone Derivatives' *Anal. Biochem.*, **180**, 351–357 (1989).
101. K. Kaheki, S. Suzuki, S. Honda, Y.C. Lee, 'Precolumn Labeling of Reducing Carbohydrates with 1-(*p*-Methoxy)phenyl-3-methyl-5-pyrazolone: Analysis of Neutral and Sialic Acid-containing Oligosaccharides Found in Glycoproteins', *Anal. Biochem.*, **199**, 256–268 (1991).
102. J. Suzuki, 'Methods for Analysis of Component Sugars by Fluorescent Labeling', *Trends Glycosci. Glycobiol.*, **3**, 48–50 (1991).
103. S. Hase, S. Hara, Y. Matsushima, 'Tagging of Sugars with a Fluorescent Compound, 2-Aminopyridine', *J. Biochem.*, **85**, 217–220 (1979).
104. S.L. Wu, 'The Use of Sequential HPLC and CZE to Separate the Glycosylated Peptides from Recombinant Tissue Plasminogen Activator to a Detailed Level of Microheterogeneity', *Anal. Biochem.*, **253**, 85–97 (1997).
105. D.C. James, R.B. Freedman, M. Hoare, N. Jenkins, 'High-resolution Separation of Recombinant Human Interferon-gamma Glycoforms by Micellar Electrokinetic Capillary Chromatography', *Anal. Biochem.*, **222**, 315–322 (1994).
106. W. Nashabeh, Z. El Rassi, 'Capillary Zone Electrophoresis of α_1-Acid Glycoprotein Fragments from Trypsin and Endoglycosidase Digestion', *J. Chromatogr.*, **536**, 31–42 (1991).
107. R.S. Rush, P.L. Derby, T.W. Strickland, M.F. Rohde, 'Peptide Mapping and Evaluation of Glycopeptide Microheterogeneity Derived from Endoproteinase Digestion of Erythropoietin by Affinity High-performance Capillary Electrophoresis', *Anal. Chem.*, **65**, 1834–1842 (1993).
108. J.P. Larmann, A.V. Lemmo, A.W. Moore, J.W. Jorgenson, 'Two-dimensional Separations of Peptides and Proteins by Comprehensive Liquid Chromatography/Capillary Electrophoresis', *Electrophoresis*, **14**, 439–447 (1993).
109. P.M. Rudd, I.G. Scragg, E.C. Coghill, R.A. Dwek, 'Separation and Analysis of the Glycoform Populations of

Ribonuclease B Using Capillary Electrophoresis', *Glycoconj. J.*, **9**, 86–91 (1992).
110. P.M. Rudd, H.C. Joao, E. Coghill, P. Fiten, M.R. Saunders, G. Opendakker, R.A. Dwek, *Biochemistry*, **33**, 17–22 (1994).
111. A.W. Moore, Jr, J.P. Larmann, Jr, A.V. Lemmo, J.W. Jorgenson, 'Two-dimensional Liquid Chromatography/Capillary Electrophoresis Techniques for Analysis of Proteins and Peptides', *Methods Enzymol.*, **270**, 401–419 (1996).
112. J.C. Severs, R.D. Smith, 'Capillary Electrophoresis/Mass Spectrometry', in *Handbook of Capillary Electrophoresis*, 2nd edition, ed. J.P. Landers, CRC Press, Boca Raton, FL, 791–826, 1997.

Capillary Electrophoresis of Proteins and Glycoproteins

Amy Trout and Michelle M. Muza
University of Pittsburgh, USA

James P. Landers
University of Virginia and University of Virginia Health Science Center, Charlottesville, USA

1 Introduction 1
1.1 Electrophoresis in a Capillary 1
1.2 Instrumentation 2
1.3 Modes of Capillary Electrophoresis 3
2 Capillary Electrophoresis Analysis of Proteins and Glycoproteins 3
3 Factors Affecting Separation 4
3.1 Capillary Wall Interactions 4
3.2 Buffer 5
3.3 Capillary Coatings 5
3.4 Protein Structure 6
4 Practical Aspects of Capillary Electrophoresis 6
4.1 Theory 6
4.2 Separation Parameters 6
4.3 Electrode Polarity 7
4.4 Injection 7
4.5 Capillary 7
4.6 Method Development 8
5 Separation Methods 8
5.1 Capillary Zone Electrophoresis 8
5.2 Capillary Gel Electrophoresis 12
5.3 Capillary Isoelectric Focusing 15
5.4 Affinity Electrophoresis 17
6 Concluding Remarks 19
Abbreviations and Acronyms 20
Related Articles 20
References 21

As a result of their complex nature, the separation of proteins and glycoproteins by capillary electrophoresis (CE) is a challenge. Fortunately, a number of approaches allow for effective analysis of these biomolecules. Of particular concern is the proclivity of proteins and their glycoconjugates to adsorb to the capillary wall. Various covalent and dynamic capillary coatings are described which minimize this effect. We also address general concerns including electrode polarity, capillary conditioning and method development that are important for achieving optimal separation. Finally, each separation mode, including capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE), and capillary isoelectric focusing (CIEF), is discussed with particular attention to its niche in the realm of protein and glycoprotein analysis.

1 INTRODUCTION

Electrophoresis is a process that involves the movement of charged particles in an applied electric field. This concept, first defined by the moving boundary electrophoresis experiments of Tiselius[1] in 1937, was eventually adapted to slab gels in the early 1960s.[2,3] In this format, electrophoretic separation is performed through a cross-linked, porous gel polymerized between two plates. Slab gel electrophoresis, as well as adaptations in the tube format, provided the basis for a variety of electrophoretic modes to evolve and has been and remains a dominant force in the field of separations. In the early 1980s, Jorgenson and Lukacs[4] revitalized a technique that Hjérten[5] had pioneered more than a decade earlier and, thus, introduced CE to the scientific community.

1.1 Electrophoresis in a Capillary

CE involves the separation of analytes in narrow-bore capillary tubes which are typically composed of fused silica. The attractive aspects of CE stem from the fact that the instrumental control common to chromatography is now possible with electrophoresis. CE circumvents the labor-intensive and time-consuming processes associated with conventional electrophoresis (casting, loading, running, staining, and destaining) with little or no sample preparation, decreased analysis time, and on-column detection. Capillaries are typically less than 100 μm inner diameter by 375 μm outer diameter by 20–100 cm in length and provide a high surface-to-volume ratio permitting rapid dissipation of Joule heat.[4] As a result of the microscale dimensions, applied electric fields as much as 50 times that associated with slab gels can be routinely employed, enabling separations on the order of minutes or even seconds.[6]

A direct consequence of the microscalar nature of the capillary dimensions is the minute quantities of buffer and sample required for analysis. The total volume of the capillary, routinely 0.4–5.0 μL, translates to only microliters of reagent consumed per run. More importantly, the sample volume injected is restricted to approximately 1% of the total capillary volume (as low as 0.2 nL) requiring as little as 2–5 μL of sample for repetitive separations. The use of small volumes allows for the capillary to be easily rinsed and regenerated between

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

analyses with moderate pressure (10–50 psi), vacuum, or even gravity. The dramatic decrease in the amount of reagents used in CE, compared with slab gel methods, is directly linked to another advantage with the capillary-based electrophoretic approach – cost. Although the capital investment for commercial CE instrumentation or a homemade system is initially more expensive than slab gel equipment, the reduced cost for reagents, and small sample volumes, combined with dispensing of staining procedural costs, and the added benefit of automation, make CE an economical alternative.[7]

In addition to providing an instrumental approach to electrophoretic protein separations, CE allows for an on-line detection system, eliminating the need for staining and destaining procedures characteristic of the slab gel format. An electrophoresis technology that allows for evolution of detection from the semiquantitative staining procedure to direct wavelength-specific on-line detection has obvious merit. The problems associated with gel-based staining methods for detection of proteins are well-established and dependent heavily on a variety of parameters (stain source, stain age, temperature, specific protein, etc.), all of which affect staining efficiency and, thus, quantitation. Staining procedures are particularly problematic with glycoproteins as a result of stain binding interference by surface carbohydrate.

Although both optical and electrochemical detection methods have been demonstrated with CE, the optical techniques have been implemented more universally than the others. Ultraviolet/visible (UV/VIS) detection is the simplest to employ and applicable to a wide range of compounds including many biopolymers. However, the short pathlength of the optical cell (which is the internal diameter capillary itself) leads to poor concentration sensitivity and detection limits in the 10^{-6}–10^{-8} M range. It is in this respect that fluorescence detection is important, providing a more selective detection scheme and, in comparison with UV/VIS detection, sensitivity that is 100–1000-fold greater. Unfortunately, only the amino acids tryptophan, tyrosine, and phenylalanine fluoresce naturally and, therefore, most proteins must be tagged prior to analysis for laser-induced fluorescence (LIF) detection using simple (e.g. argon ion lasers) systems.

Yet another problem encountered with gel-based separations is poor reproducibility, particularly with two-dimensional systems. The quantitative reproducibility of CE is very good with typical migration time relative standard deviation (%RSD) values between 0.1% and 1.0%. In addition, for well-characterized CE systems, protein separation efficiencies reach 10^4–10^5 theoretical plates.[8] Thus, a technique surpassing the performance of its gel counterpart emerges, providing rapid, high efficiency, and reproducible separations with the added value of on-line detection and automation.

1.2 Instrumentation

A diagram of the instrumental set-up for CE is presented in Figure 1. The necessary components for this system include a capillary (outer surface polyimide-coated), a high voltage supply, two buffer reservoirs capable of accommodating the capillary and platinum electrodes, and an on-line detector. To perform a separation, the preconditioned capillary is filled with buffer and then sample is introduced at the inlet by pressure, vacuum, gravity or an applied potential. Using an uncoated (internal) fused silica capillary, under normal polarity, the inlet will be positive (anode) and the outlet negative (cathode). Both ends of the capillary are transferred to their respective buffer reservoirs where a separation voltage (up to 30 000 V) is applied. The charged species present in the sample plug migrate with an electrophoretic mobility deemed by their charge and mass (under nondenaturing conditions), and pass through the on-line detector, where information is collected and stored electronically.

Conventional separation techniques (such as gas chromatography (GC) and high-performance liquid chromatography (HPLC)) employ pressure gradients to pump the mobile phase and solutes from the column inlet to the detector. CE is unique in this respect in that it does not utilize a mechanical pumping mechanism for bulk flow. Instead, the physical configuration and chemical make-up of the capillary allow for an electric field-driven pumping of bulk solution during the electrophoretic process. Using an uncoated fused silica capillary and a normal polarity configuration (inlet → anode, outlet → cathode), ionized silanol groups on the capillary wall induce the formation of a double layer of cations (and their hydration spheres) which produce an electroosmotic flow (EOF) in the presence of an applied field (Figure 2a). The magnitude of this flow becomes significant ($\simeq 10^{-4}$ cm^2 V^{-1} s^{-1}) above pH 5–6 where the silanol groups will be largely deprotonated. EOF is responsible for the flow of analytes, regardless

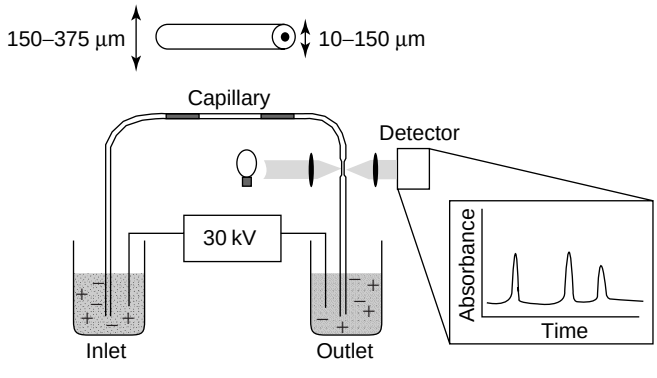

Figure 1 Diagram of a CE instrument. An enlarged section of the capillary is shown to highlight the dimensions.

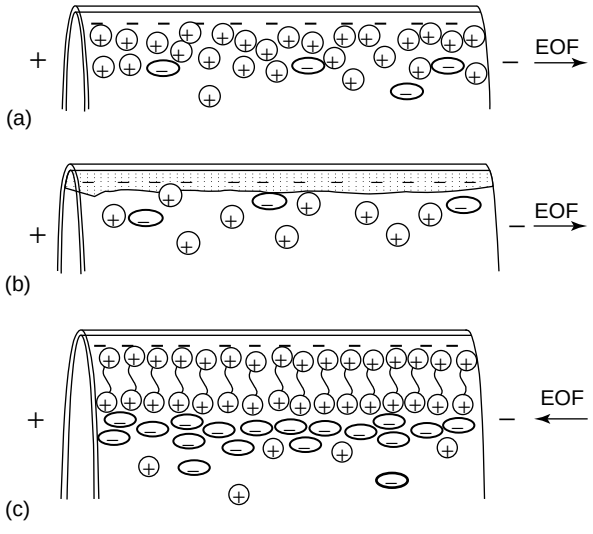

Figure 2 EOF and effects of wall coatings. (a) In most pH ranges (pH 3–11) a bare fused silica capillary has a negative surface charge. Consequently, cations in solution align, forming a positively charged double layer. Under an applied field and normal polarity (inlet → anode, outlet → cathode), the positive ions will move toward the cathode resulting in bulk flow toward the detector. (b) Coating with a nonionic layer shields the surface charge contributed by the silanols and thereby reduces the EOF. As the pumping mechanism has been minimized, the electrode polarity must be chosen so that molecules of interest migrate toward the detector. (c) Capillary coatings changing the effective charge of the wall from negative to positive will result in the formation of an anionic double layer. Consequently, the EOF is reversed and will be in the direction of the inlet (anode). Thus, only species with an electrophoretic mobility greater than the EOF will reach the detector under normal polarity. Reverse polarity may be employed for the detection of analytes with low electrophoretic mobility.

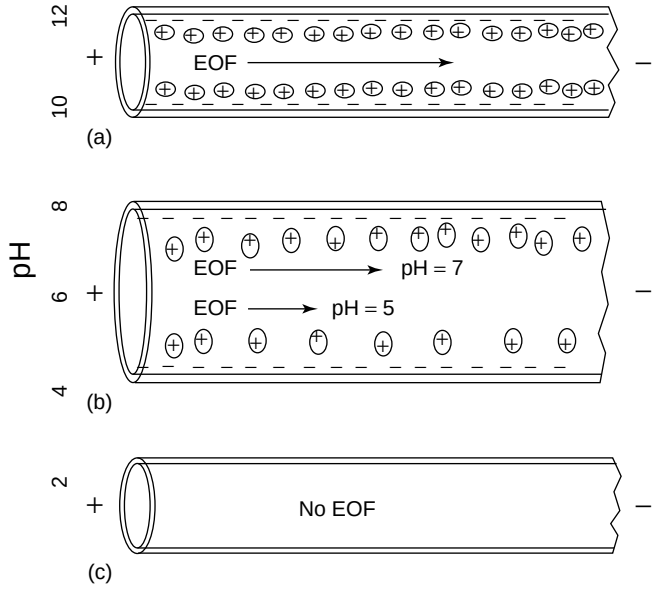

Figure 3 Diagram correlating EOF with pH. (a) At high pH, the capillary is largely deprotonated and therefore highly negatively charged. Consequently, a large EOF is present resulting in bulk movement of the solution toward the cathode. (b) The silanol surface of the capillary wall has a pK_a between pH 5 and pH 7. Around pH 5 a few silanol groups will be deprotonated and therefore a small EOF will exist in the capillary. As the pH is raised to basic, a larger fraction of the silanols will become deprotonated and therefore the EOF will increase within this range. (c) At low pH, the capillary wall will be mostly nonionic due to the protonated silanols. Here, the EOF is significantly diminished. Thus, the charge of the analyte and polarity of the electrodes determine which analytes migrate toward the detector.

of charge or size, to the detector. Thus, even neutral and negatively charged species will pass the detector as a result of the EOF, in most cases, being greater than the electrophoretic mobility of the anodically mobile species. EOF is largely (but not solely) governed by the buffer pH; a low-pH buffer produces relatively low EOF as a result of the decreased number of deprotonated silanols. Figure 3 illustrates the pH dependence of EOF. It is important to note that EOF results in the bulk movement of all analytes past the detector, therefore, a decrease in EOF will result in a situation where only solutes of like charge can be separated.[9]

1.3 Modes of Capillary Electrophoresis

Similar to the diversity found with standard gel electrophoretic techniques, several modes of CE have been established. These include CZE, micellar electrokinetic chromatography (MEKC), CIEF, and CGE. Although capillary isotachophoresis (CITP) has already been developed as a stand-alone mode for the analysis of proteins, it is primarily used as a preconcentration technique in CE. A basic knowledge of the existing similarities and defining differences between the modes is often necessary for optimizing a given separation. For instance, this may involve nothing more than use of a buffer additive to switch from one mode to another. Of the four main modes (CZE, MEKC, CGE, CIEF) only CZE, CGE, and CIEF play a major role in the CE analysis of proteins. These are described in the sections where they are discussed.

2 CAPILLARY ELECTROPHORESIS ANALYSIS OF PROTEINS AND GLYCOPROTEINS

Proteins and glycoproteins present a unique challenge to the separation scientist as a result of their complex nature. Crude or partially purified protein preparations are not trivial to analyze by CE due to the diverse nature of molecular masses and pH- and temperature-dependent stabilities that may be represented in the mixture. With

traditional denaturing slab gel electrophoresis, these inherent variations do not present a problem but, within the realm of CE where nondenaturing analysis is easily executed, these become important parameters.

Additional considerations associated with CE stem from the fact that separation is carried out in a capillary composed of fused silica, a substance with poor biocompatibility, especially for proteins. Under nondenaturing conditions, proteins have a propensity for adsorption to the capillary wall. This interaction must be minimized or negated if CE separation of proteins based on both charge and size is to be successful. A number of approaches are addressed in this review to deal with problematic areas of protein analysis. Once solved, CE evolves into an extremely powerful technique for biomolecule analysis.

Finally, glycoproteins are a group of macromolecules with complexity beyond that of nonglycosylated proteins as a result of the carbohydrate structures present on the protein surface. The carbohydrate moieties on the protein have been shown to be associated with numerous biological functions, including solubility, antigenicity, protein folding, transport, and clearance from the cell.[10] Additionally, abnormalities in glycosylation have been found to lead to or be the result of a number of pathological states.[11,12] To further complicate the issue, even purified preparations of certain glycoproteins are not trivial to analyze electrophoretically as a result of the inherent microheterogeneity that results from the natural variation in the carbohydrate content of the protein in biological systems. These glycoforms arise from the cellular production of a set of proteins that are identical in their polypeptide structure, but differ in their glycosylation with each serving a specific function within the cell. This microheterogeneity is thought to play a critical role in the high fidelity control of cellular processes.[10] It is clear that in order to understand the role of glycoprotein microheterogeneity in regulating complex cellular processes, their separation is a necessary component of biochemical research. This is often difficult if not impossible to accomplish with conventional electrophoretic formats. However, as will become clear in the sections that follow, nondenaturing separations by CE provide a new tool for interrogating glycoprotein structure and microheterogeneity.

3 FACTORS AFFECTING SEPARATION

3.1 Capillary Wall Interactions

One of the major problems that had to be addressed with early attempts to resolve proteins under nondenaturing conditions in fused silica capillaries was their nonspecific adsorption to the capillary wall. In the earliest studies by Jorgenson and Lukacs[4] the equation governing separation efficiency (derived by Giddings for chromatography)[13] was simplified based on the fact that, as this was electrophoresis and not chromatography, the partition coefficient k' in the radial diffusion term was essentially zero, giving Equation (1), where D is the diffusional coefficient

$$N = \frac{\mu V}{2D_{n,T}} \tag{1}$$

This made CE, in a sense, ''antichromatography''; with the assumption that $k' = 0$, band broadening (number of theoretical plates) would be primarily dependent on axial diffusion contributions. Due to the large molecular mass of proteins, diffusion would be relatively small and, therefore, calculated efficiencies of the order of 10^6 theoretical plates would be expected. However, this was not observed, even under the most ideal separation conditions. The inability to achieve predicted efficiencies was attributed to adsorption of analytes onto the highly negatively charged silanol surface of the capillary wall,[14] indicating that $k' \neq 0$ and, therefore, cannot be ignored. Swedberg's group at Hewlett-Packard has provided insight into this and shown that even small k' values (as low as 0.001) result in a substantial decrease in the achieved theoretical plate number.[15] Accordingly, attention must be paid to define CE conditions that strive to meet the assumption that $k' = 0$ (i.e. no wall interactions) for both peptides and proteins.

This can be accomplished for CE-based peptide separations with relative ease by working at pH extremes. At a basic pH of 10–11, both the capillary silanol groups and the peptides are fully deprotonated and, hence, negatively charged allowing for repulsion to prevent adsorption. At acidic pH of 1.9–2.5, the capillary silanol groups are fully protonated and, therefore, neutral in charge, while the peptides are cationic with no predominant proclivity for adsorption to the wall. Consequently, acidic peptides, difficult to resolve by reversed-phase HPLC methods, are readily analyzed in simple, low-pH buffer systems, such as 100 mM sodium phosphate, pH 2.5. This approach is extremely effective for peptide separations due, in part, to the fact that the peptides can easily withstand the extreme nature of the separation conditions. This is not the case with proteins where their stability is questionable outside the physiological pH range (pH 7.5). Although there have been some examples demonstrating that working at extremes of pH can be effective for protein analysis,[16] this is more the exception than the rule and precludes analysis in the most relevant pH range (physiological).[17] Furthermore, adsorption is still frequently observed even at high pH values (e.g. pH 10.0) above the isoelectric

point (pI) of many proteins. One must be cognizant that the pI is a net charge for the molecule and does not account for localized surface basic amino acids which may retain a positive charge even at extreme pH ranges. This phenomenon is exacerbated with proteins due to their increased surface area.

Thus, a challenge inherent in performing protein separations in silica capillaries, which is not encountered in HPLC or conventional electrophoresis, involves determining conditions that will minimize or even eliminate protein–wall interaction (i.e. $k' = 0$). Approaches for doing so have been evaluated extensively and involve either covalent or dynamic modification of the capillary wall or augmentation of the separation buffer composition. These are addressed in the sections that follow.

3.2 Buffer

Buffers play a major role in achieving a successful separation in CE. Ideally, CE buffers must maintain the solubility of the analyte, have good buffering capacity in the pH range of interest, and have low absorbance at the detection wavelength desired (typically 200 or 214 nm for proteins). There are a number of considerations for selecting an appropriate separation buffer. Most important is choosing an electrolyte. A number of effective electrolytes have been identified for protein separations at slightly acidic and basic pH ranges with borate and phosphate often employed. Another consideration is the ionic strength of the buffer. Higher buffer concentrations are obviously beneficial for increasing the buffering capacity of the solution, but they also influence the EOF by directly affecting the thickness of the ionic double layer (increased ionic strength → decreased EOF), and are associated with more current at high applied fields. As detailed in the article in this encyclopedia on **Capillary Electrophoresis of Peptides**, an Ohm's law plot (OLP) can aid in defining the functional concentration for a particular separation buffer.[18]

3.3 Capillary Coatings

In 1986, Lauer and McManigill[19] published a seminal paper demonstrating that protein separations could be achieved successfully in a bare silica capillary if wall adsorption was addressed. Over the past decade, research into the development of novel technologies for masking or eliminating the negative charge of the capillary wall has been productive. A number of unique covalent coatings have emerged to manipulate the EOF and minimize protein adsorption. Of equal importance, noncovalent (dynamic) approaches have been developed which also show promise in effecting protein separations in capillaries that exhibit degradation of efficiency due to adsorption.

3.3.1 Covalent Coatings

Capillaries can be functionalized (covalently coated) so that the inner walls are permanently modified. For example, a static wall coating usually consists of two chemical layers – the first reacts covalently with the silica, the outer layer being a hydrophilic neutral polymer. The polymer serves to shield residual silanol groups and increase the viscosity of the electric double layer, thus resulting in a substantial decrease, if not elimination, of EOF (Figure 2b). These coatings should be inert and resistant to degradation over a reasonably large pH range. The first material to be used as a covalent coating was polyacrylamide, a logical candidate stemming from its use in traditional slab gel electrophoresis techniques. Hjérten[20] was able to achieve near-zero EOF with polyacrylamide as the capillary coating for initial application to isoelectric focusing (IEF). Although effective, this coating suffered from stability and longevity problems.[21] Fortunately, improvements in coating procedures have minimized the problems associated with polyacrylamide coatings and commercial capillaries are now available.[9] Other, more traditional, polymer coatings including poly(vinyl pyrrolidone)[16] and poly(ethylene glycol)[22] have been employed for protein separation. In addition, a number of commercial capillaries have proved to be successful for CE-based protein separation including capillaries from the DB-series, where the coating is some dimethylpolysiloxane (originally developed for GC; J & W Scientific), fluorocarbon-coated capillaries (J & W Scientific), capillaries with an uncharged hydrophilic coating (Beckman eCAP Neutral™), anionic coated capillaries (CE-100-SA; Scientific Resources), and capillaries with a cationic coating (Celect-Amine™; Supelco).[9] Also, as a consequence of the necessity for stable, relatively inert coatings, a number of homemade capillaries have also been found to be effective for unique protein and glycoprotein separation systems.[23–28]

3.3.2 Dynamically Coated Capillaries

As an alternative to covalent passivation of the capillary surface as described in the previous section, dynamic coating of the capillary can also be carried out. This involves specific adsorption of solutes to the capillary wall, resulting in a temporary, yet stable, masking of the surface silanol groups. Depending on the nature of the dynamic coat, the negative charge imparted by the silanols is significantly reduced, eliminated, or in some cases reversed. The advantage of this approach lies in the simplicity, as it often involves nothing more

than augmenting the separation buffer with the dynamic coating additive at the appropriate concentration. As the dynamic coating is pumpable, it can be easily removed and regenerated.

3.3.2.1 Cationic Additives One approach to dynamically coating the capillary employs cationic additives some of which actually reverse EOF by changing the charge of the capillary wall from negative to positive (Figure 2c). In particular, long-chain alkyl quaternary amines (e.g. polybrene and cetyltrimethylammonium bromide (CTAB)), have been employed effectively for this purpose.[29–33] The cationic head of the amine aligns with the anionic silanols on the capillary wall so that the hydrophobic tail extends into the capillary core. Unbound surfactant molecules align with their hydrophobic tails binding hydrophobically, resulting in a pseudodouble layer and a cationic wall, thus resulting in a reversal of EOF. Employing a cationic additive, separation is carried out in a pH below the pI of the protein so the positively charged proteins will be repelled by the capillary wall. This differs from dicationic long-chain alkyl quaternary amines (e.g. hexamethonium or decamethonium bromide) where the surface is adequately passivated without the reversal of EOF.[34] Numerous reports in the literature demonstrate the success of diaminoalkane cationic additives for investigating proteins and their glycoconjugates.[19,35,36]

3.3.2.2 High-salt Buffers Increasing the ionic strength of the separation buffer can also be utilized to reduce protein–wall interactions. Lauer and McManigill[19] first illustrated this approach by adding potassium sulfate to their separation buffer. This prevented adsorption of standard proteins, with pI values ranging from 4.5 to 11, to the bare silica wall. Similar experiments conducted by Green and Jorgenson[37] employed alkali metal salts to reduce protein adsorption. Although effective, this alternative is associated with two main disadvantages: high molar absorptivity in the low-UV (ultraviolet) region and high current resulting in increased Joule heating.[17] The latter of these problems may be circumvented with the use of zwitterionic additives which can be added at molar concentrations without invoking high current with the use of high electric fields (their electroneutrality does not increase the conductivity of the buffer and creates a boundary layer of ions between the silanol layer and the proteins).[38,39] The effectiveness of this approach has been illustrated by Bushey and Jorgenson[40] for the resolution of basic proteins, and has since led to the use of many zwitterionic additives, including betaine,[41] CAPS (3-cyclohexylamino-1-propanesulfonic acid),[19] CHES [2-(N-cyclohexylamino)ethanesulfonic acid],[19] and tricine[19] for the successful separation of proteins. More detail on approaches for minimizing protein–wall interactions in CE is given in a number of excellent reviews on the subject.[9,17,42,43]

3.4 Protein Structure

An important factor involved in optimizing protein separation by CE is the protein itself. Many proteins are unstable in simple buffer solutions (precipitation, aggregation, etc.) and, consequently, solubilizing (detergents) or stabilizing agents (glycerol) must be present to maintain integrity. Where these might have a negligible effect with conventional slab gel electrophoresis, they can have significant impact on the CE separation. It is reasonable to assume that, once solubilized, proteins have an electrophoretic mobility solely dependent on charge and size. However, it is now apparent that a number of factors including, but not limited to, mass and charge contribute to a protein's mobility. For example, the electrophoretic mobility of an analyte is affected by the surrounding environment – in particular the buffer, any additives (organic modifiers), other analytes, and the capillary wall. In addition, one must be cognizant of the fact that the net charge of the protein is the total charge contribution from amino acids, metal ions, and cofactors. However, this charge is shielded by a number of environmental factors, including attraction to counter-ions in solution. Thus, it is the effective charge of a protein which contributes to mobility in an applied field.[44] The hydrodynamic radius of a protein is another important factor affecting mobility in an applied field. Kálmán et al.[45] demonstrated that mutants of nuclease A with a larger hydrodynamic radius had longer migration times than those proteins with smaller radii. Thus, for protein separations it may be more accurate to describe the mechanism of electrophoresis as a function of effective net charge and hydrodynamic radius instead of simply charge and size.

4 PRACTICAL ASPECTS OF CAPILLARY ELECTROPHORESIS

4.1 Theory

The equations governing mobility, shape of the analyte zone, resolution, efficiency, and sources of variance in electrophoresis are discussed in detail in **Capillary Electrophoresis of Peptides** and are therefore not addressed here. An in-depth treatment of these topics is also given by Giddings.[46]

4.2 Separation Parameters

With the numerous parameters that must be addressed when developing a method or optimizing a CE separation, it is imperative to have an understanding of the

interdependence of one parameter with another. For example, a protein separation can be profoundly affected by slight changes in the capillary dimensions. Specifically, if closely migrating species (e.g. protein glycoforms) are not adequately resolved, increasing the capillary length will increase resolution by affording a longer on-capillary time. However, a longer capillary is associated with increased analysis times which may become unfavorable with rapid separation schemes. Fortunately, increasing the capillary length decreases the electric field strength at constant voltage, thus allowing the application of higher electric fields (usually characterized by decreased migration time). Increasing the capillary diameter has advantages and disadvantages. An increase in capillary inner diameter will improve detection sensitivity by increasing the pathlength. However, a decreased surface-to-volume ratio accompanies an increased diameter, ultimately resulting in less efficient heat dissipation which is characterized by band broadening. It is important to note that these and other parameters (such as pH, EOF, field strength, etc.) can be effectively optimized to achieve separation conditions. An evaluation of the relationship of CE variables is given in a review by Oda and Landers.[47] Discussion is restricted in this text to only fundamental separation parameters.

4.3 Electrode Polarity

Determining the electrode polarity is of paramount importance to achieving a desired separation. Manipulation of the EOF via electrode polarity or capillary wall charge modification (through the use of coatings) enables selection of the analytes migrating toward the detector. As discussed earlier, normal polarity in a bare silica capillary is associated with an anodic inlet and a cathodic outlet. Here, positively charged species will move with the EOF and a characteristic mobility (based on charge and size) in the direction of the detector. Neutral analytes having no electrophoretic mobility will be carried with the EOF. In this system, negatively charged species will have an intrinsic mobility toward the inlet, therefore, away from the detector. However, the EOF is usually the dominating force (greater than electrophoretic mobility), thus allowing for separation and detection of anodically mobile analytes. In comparison, if the reverse polarity is set, the EOF will be in the opposite direction, away from the detector. In this case, positively charged analytes have a characteristic mobility toward the inlet (cathode) while neutral species are carried with the EOF in this same direction. Consequently, only negatively charged analytes with a mobility greater than the EOF will reach the detector. Reverse polarity is often employed when a coated capillary is used to separate negatively charged analytes or with substances that reverse the net charge of the capillary wall.

4.4 Injection

Regardless of the mode of CE utilized for separation, there are three primary methods for sample injection into the capillary: hydrodynamic, hydrostatic (by siphoning), or electrokinetic. Hydrodynamic injection is accomplished by immersing the capillary inlet into the sample and either pressurizing the sample vial or applying a vacuum to the outlet. Hydrostatic sample injection (by gravity), more common in noncommercial instruments, relies on siphoning of the sample into the capillary by elevating the sample vial relative to the outlet vial. Electrokinetic injection is accomplished by immersing the capillary inlet into the sample and, with the outlet in buffer, applying a low voltage to the system. This injection method is discriminating in that movement of the sample into the capillary depends on electrophoretic mobility and EOF. Consequently, components of the sample matrix with high electrophoretic mobility will be more likely than those with a lower mobility to be introduced into the capillary. The advantages of biased injection are discussed in detail in **Capillary Electrophoresis of Peptides**.

4.5 Capillary

The condition of the capillary has a substantial impact on the efficiency and reproducibility of a CE separation. This is particularly true when dealing with coated capillaries. Both the fused silica and coated capillaries can be purchased from a number of companies; however, treatment of the capillary prior to, during, and after an electrophoretic separation is crucial.

4.5.1 Conditioning

When using a new capillary or changing to a new buffer system the capillary must be conditioned. A typical conditioning procedure for a bare silica capillary is a 5–10 column volume rinse with 100 mM NaOH, followed by a 5–10 column volume rinse with water and, finally, a 3–5 column volume rinse with separation buffer. If the capillary is coated, conditioning should be performed as per the manufacturer's protocol.

4.5.2 Regeneration

Maintenance of the capillary surface is necessary for achieving reproducible results which are largely dependent on consistent EOF. However, during the course of a separation and, particularly after multiple separations, the surface may become fouled by the sample. For this reason, the capillary surface must be cleansed of any adsorbed species (i.e. regenerated) after each separation. For a bare silica capillary and neutral or basic buffer system, regeneration is typically achieved by rinsing 3–5 column volumes with 100 mM NaOH followed

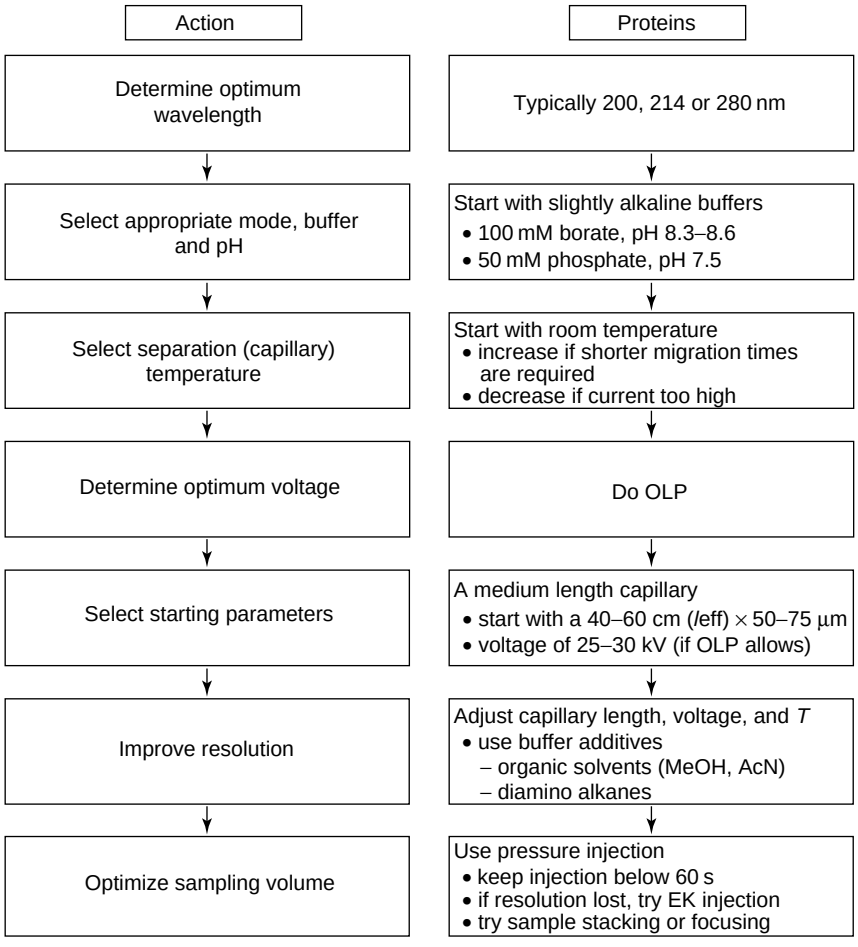

Figure 4 A step-wise approach to method development for the analysis of proteins by CE; *l*eff is the effective length, T = temperature and EK = electrokinetic.

by a 5–8 column volume rinse with water and then separation buffer. When using an acidic buffer, NaOH regeneration may be disadvantageous as the drastic pH change may require extensive rinsing to reequilibrate the capillary. In this situation, following the water wash with a brief rinse with concentrated buffer (10× or more) and then separation buffer is the advised approach. Depending on the sample, the concentrated buffer wash may be effective for regenerating the surface. In this situation, the NaOH rinse can be completely eliminated. After extensive use, NaOH regeneration may not be effective for completely removing adhered species. Here, rinsing with sodium ethoxide (blended from equal volumes of 1 M NaOH and 95% ethanol) for 5 min is usually sufficient. The sodium ethoxide must be followed with 10–20 column volumes of water and then finally 5–8 column volumes of separation buffer. Regenerating a coated capillary is more complicated. When possible, it is best to follow the manufacturer's instructions for regeneration.

4.6 Method Development

The development of CE methods is as idiosyncratic in nature as with HPLC. Even slight heterology in the primary structure or differences in the secondary or tertiary structure of proteins (single amino acid substitution) can require dramatic changes in methodology. Consequently, separation conditions must be unique to the system of interest. Figure 4 illustrates a step-wise approach to designing a method for protein separation by CE.

5 SEPARATION METHODS

5.1 Capillary Zone Electrophoresis

CZE is the simplest form of CE and is the mode described in most of the previous discussion. With this technique, the capillary is filled with buffer, the sample injected, the separation voltage applied to the capillary reservoirs, and the charged species within the field move toward

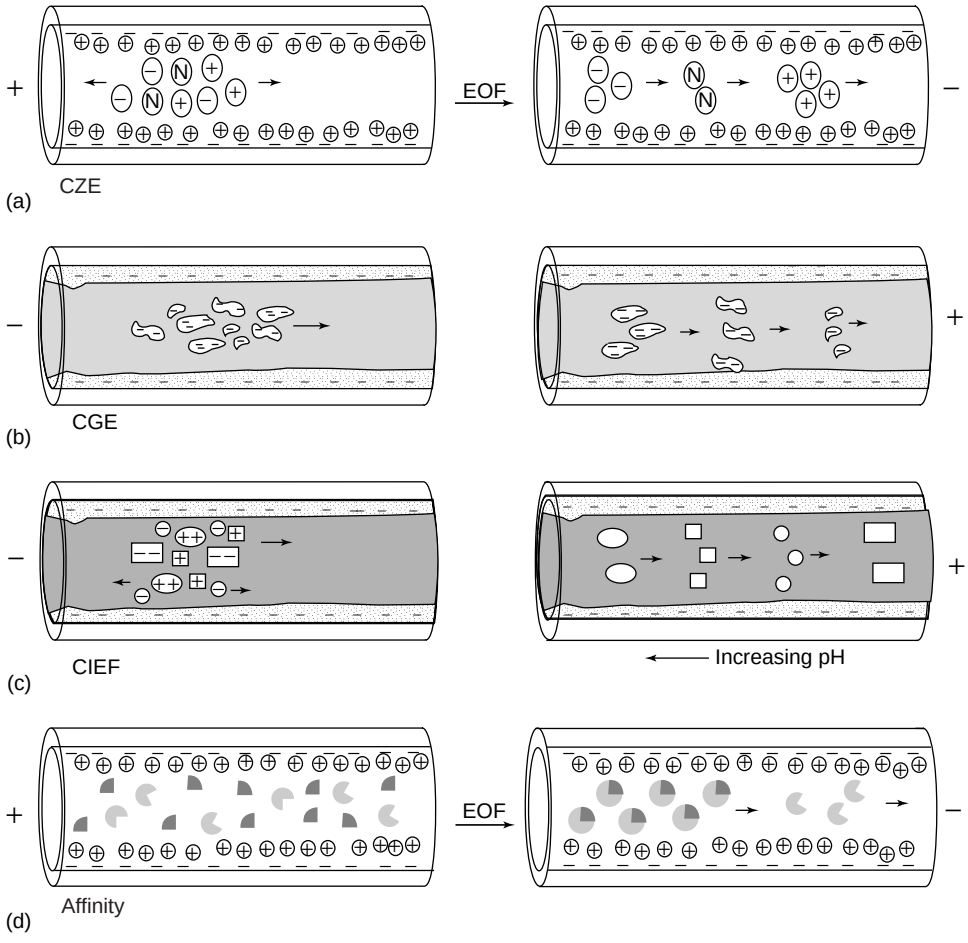

Figure 5 CE separation modes. (a) CZE separates analytes based on charge and size. Once injected, the sample components migrate into zones determined by their electrophoretic mobility. Under normal polarity (inlet → anode, outlet → cathode) positively charged proteins will have a mobility greater than the EOF and reach the detector first. Neutral proteins having zero electrophoretic mobility will migrate with the EOF. Negatively charged proteins will have mobility in a direction opposite to the EOF. Fortunately, the EOF is often the dominating force and negatively charged species reach the detector. (b) In CGE mode, a static polymer coating is used to shield the silanols of the capillary wall, thus eliminating the EOF. The capillary is then filled with a chemical (cross-linked) gel or a physical (entangled) gel. Under denaturing conditions, the proteins are negatively charged. As a result, separation must be carried out under reverse polarity (inlet → cathode, outlet → anode) for the proteins to reach the detector. Under nondenaturing conditions, the protein's charge will be influenced by the pH of the buffer system. Consequently, electrode polarity may be normal or reverse depending on the conditions of the separation. (c) In CIEF ampholytes are introduced into the capillary to set up a pH gradient. Under an applied field charged species migrate (solely based on charge and size because a coated capillary is used to minimize the EOF) to the region where the pH is equal to their pI at which point they become electrically net neutral. Analytes remain focused into characteristic zones because any movement into the surrounding pH medium results in a net charge difference, an opposing mobility, and consequently a return to the region where the pH is equal to the analyte's pI. (d) Affinity electrophoresis is utilized to determine binding constants between ligands and receptors. In particular, the electrophoretic mobility of a protein varies with the occupancy of the binding site ligand and therefore changes in migration time are used to estimate interaction.

the opposite electrodes with their movement dependent on both charge and size (Figure 5a). Here, EOF is fundamental, allowing for the separation of positively charged, negatively charged, and neutral or uncharged species. However, situations may arise when EOF is undesirable and the inherent electrophoretic mobility of the protein of interest is of sufficient magnitude to reach the detector.

5.1.1 Proteins and Glycoproteins

The separating power of CZE has been demonstrated for a number of proteins. Early separations by Nielson et al.[48] highlight the utility of the technique for separation of the desamido and didesamido forms of human growth hormone (hGH). To limit protein–wall interactions, the pH was raised to 8.0, well above the pI of

the hGH which is 5.2. More recently, CZE with LIF detection has been used to separate and detect green fluorescent protein (GFP) and GFP antibody complexes. Using a bare silica capillary and a borate buffer (pH 8.5), Korf et al.[49] successfully separated the two forms of GFP (GFP-1 and GFP-2). Implementing CE/LIF, a six order of magnitude increase in detection sensitivity (over conventional methods) was achieved. This allowed for detection of GFP in the attomolar range. In addition, GFP was separated from GFP-complexed antibody. This presents the possibility of using CE for rapid screening of anti-GFP antibodies. Hiraoka et al.[50] demonstrated the utility of CZE for the rapid separation of β-trace protein (βTP), a low-molecular-weight protein synthesized in the central nervous system (CNS). βTP was clearly determined in all of the samples, and internal standard quantitation was possible using a allylamine. The results indicate that βTP is nonspecifically increased in organic diseases of the CNS, especially those associated with severe physical damage to the brain tissues (such as multiple sclerosis and cerebral infarction).

It was previously not appreciated that many proteins contained glycan structures, making them difficult to separate. Tiselius[1] offered proof of this complex structure showing that some proteins migrated as broad zones. Since then, the resolution of individual glycoforms has been achieved in several laboratories. Kilar and Hjérten[51] were among the first to report successful separation of glycoforms differing by a single charge. Using a static polyacrylamide coating and a buffer system composed of Tris/borate/EDTA (ethylenediaminetetraacetic acid) the glycoforms of purified transferrin (Tf) were resolved. Not long to follow were CE-based separations of the glycoprotein human erythropoietin (rHuEPO), whose primary function is to promote and stimulate the production of red blood cells. Tran et al.[52] experimented with various buffer systems, organic modifiers, and a variety of pH regions revealing the effects of optimal buffer selection for glycoform analysis using CE. This concept of condition optimization using additives or pH modification applied to the separation of a number of glycoproteins. Almost simultaneously, Landers et al.[53] and Traverna et al.[35] examined the utility of borate complexation with a diaminoalkane for the resolution of glycoprotein isoforms. Using 50 mM phosphate (pH 9.0) or 100 mM borate (pH 6–10) Landers et al.[54] partially separated five glycoforms of ovalbumin. Addition of diaminobutane (DAB) to the borate buffer (pH 8–9) appeared to slow the EOF and allowed for the resolution of 9–12 glycoforms in a purified preparation. Traverna et al.[35] found that recombinant tissue plasminogen activator (rtPA) glycoforms were poorly resolved in coated or bare silica capillaries using a 100 mM phosphate buffer (pH 3.6). In an effort to enhance the separation, DAB was added to a number of buffers at varying pHs to examine effects on glycoform resolution. The addition of DAB to a borate buffer (pH 8.5–10) was not effective for achieving improved resolution of the rtPA; however, a tricine–DAB buffer resulted in an improved separation. DAB was not efficient with the borate buffer because the highly sialyated glycoprotein cannot effectively complex with borate. This prevents the ionic borate–DAB interaction which appears to be necessary for the resolution of glycoforms. More recently, Watson and Yao[55] employed an uncoated capillary and a buffer containing 10 mM tricine/10 mM NaCl/2.5 mM DAB to partially resolve five peaks of erythropoietin. Addition of urea to the separation buffer allowed for the resolution of six peaks. The isoforms were identified by individually spiking with characterized sialoforms. Similar conditions to those employed by Landers et al.[53] were employed by Morbeck et al.[56] to separate the glycoforms of the hormone human chorionic gonadotropin (hCG). A 25 mM borate buffer containing 5 mM diaminopropane was used to separate the eight glycoforms of native hCG. The distribution of glycoforms was found to vary when hCG was purified from patients with choriocarcinoma, thus highlighting the potential of capillary electrophoretic separations for diagnostic and screening purposes. Although effective for resolving the complex heterogeneity of a number of glycoproteins, the mechanistic role of diaminoalkanes in improving the resolution of CE separations is still under investigation. Preliminary results indicate that the mechanism through which DAB and other similar additives effect resolution of glycoforms may involve more than slowing the EOF.[57,58]

5.1.2 Clinical Applications

As a result of its speed, efficiency and miserly use of reagents and sample, CE has begun to gain momentum as an acceptable analytical technique in the clinical laboratory for both routine and esoteric diagnostic analyses. For example, hemoglobin (Hb) analysis is important because mutations in the protein chains induce variations in structure and impair their oxygen carrying ability. Variants of Hb, typically detected by agarose and IEF gel electrophoresis, have been shown to be analyzed effectively with a single CE analysis. Ong et al.[59] were able to separate the components in hemolysates from normal individuals and those thought to have thalassemia. A high pH phosphate buffer was used to achieve separation of these proteins in less than 8 min. Klein and Jolliff[60] used a near-neutral buffer to separate globin variants A_2, A_1, A, S, and C. Under these conditions, dramatically different profiles were found for a number of Hbs. Others have used CE technology to analyze

purified globin chains in a substantially decreased time frame from that obtained using paper chromatography or HPLC. Both Ishioka et al.[61] and Ross et al.[62] have applied this approach to rapid screening of globin chain and Hb variants.

CE has also been shown to be applicable to the analysis of serum proteins. Since Chen et al.[63] proposed CZE as a viable alternative to agarose gel electrophoresis (AGE), a number of studies have been carried out showing the superiority of CE to conventional cellulose acetate electrophoresis (CAE) or AGE methods.[64–66] Clark et al.[67] and Katzmann et al.[68] compared the effectiveness of CAE, AGE, and CE for the detection of monoclonal proteins in serum. Using a bare silica capillary and borate buffer, separation was achieved in 120 s and clear differences in abnormal and normal monoclonal profiles were observed. Clark et al.[67] concluded that qualitative results obtained by CZE were comparable or better than those obtained using the gel electrophoresis counterpart. In particular, CE provided a clear advantage over gel techniques for determining the presence of a monoclonal protein for two reasons: first, precipitation is not a problem in CE and, second, in CE, the point of detection is unaffected by the point of injection. In comparison, Katzmann et al.[68] focused more on the quantitative aspects between AGE and CE. In this study, details concerning sensitivity, specificity, and quantitation were evaluated and discussed. Jolliff and Blessum[69] recently conducted a study comparing results of serum protein analysis by AGE and CE. It was determined that CE exhibited superior reproducibility over AGE methodology. This was attributed to protein detection using direct UV detection compared to conventional staining techniques. In a more comprehensive study conducted by Katzmann et al.[70] sera from more than 1300 patients was examined using both conventional AGE and CE (Figure 6). CE was superior to AGE for the detection of serum abnormalities including polyclonal increases in immunoglobulin, point of application artifacts, β-region abnormalities, and, finally, the presence of free light chains. In many instances, sera identified as normal using conventional techniques was shown by CE to be abnormal. More comprehensive overviews of serum protein analysis via CZE are available.[64,65,71–73]

In retrospect, serum protein analysis was an interesting challenge for CE in light of the complex nature of this matrix, the diverse character of the proteins, and their high concentration in serum. The high-pH approach (pH 9.8–10.2) works extremely well with serum where the serum sample can be injected neat (undiluted) without any sign of fouling the capillary. This is, no doubt, a function of the basic separation buffer as well as the between-run capillary regeneration with 1.0 M NaOH. Although there is a long-standing concern about the longevity and integrity of capillaries repeatedly exposed to these extreme conditions, the reproducible use of the same capillary for analysis of 20–30 serum samples daily for months without any indication of performance degradation has been observed. This speaks volumes for the ruggedness and robust nature of CE for the analysis of crude mixtures.

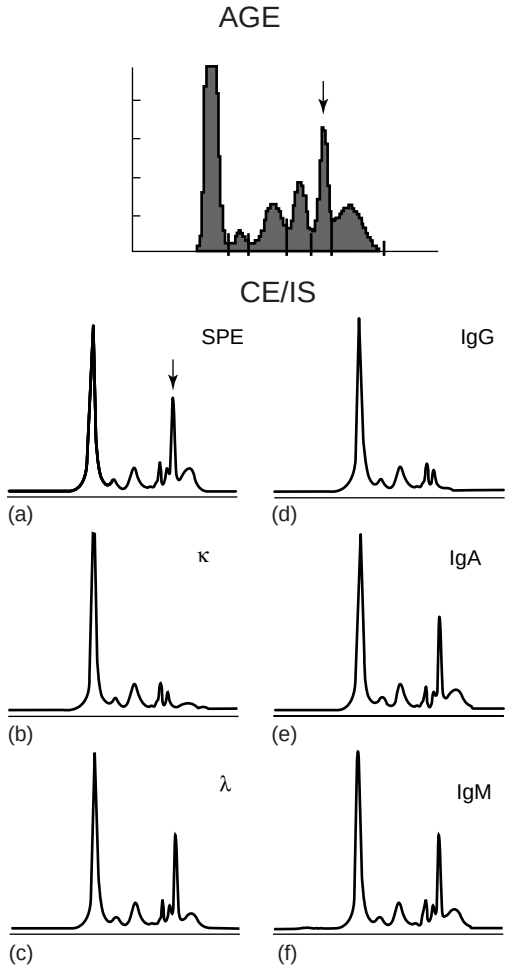

Figure 6 Detection and identification of monoclonal proteins by AGE and CE. An abnormal serum sample with a monoclonal peak in the γ-region (indicated by the arrow) was separated and analyzed using AGE and CE. AGE was performed on a Helena REP system using the REP SPE-30 agarose gels visualized with Ponceau S Stain. Peaks were calculated using the Helena ECD scanner and software designed for the multilane agarose clinical gels (Helena Laboratories, Beaumont, Texas). CE was performed on a Beckman Paragon CZE 2000 system using Beckman system buffers, the automated sample handling module and Beckman software designed for the seven-capillary clinical instrument (Beckman Instruments, Brea, California). Immunosubtraction (IS) was performed to determine the identity of the spike. Subtraction with anti-κ, anti-λ, and anti-IgG reagents removed the M-spike (panels b, c and d respectively). The anti-IgA and IgM (panels e and f) reagents had little effect. (Taken with permission from ref. 70.)

In addition, CE has been applied to the analysis of urine proteins. Urinalysis has wide clinical significance, particularly because abnormal patterns are seen in many pathological conditions including various renal disorders and light chain disease.[74–77] It was initially thought that urinary proteins, which originate from serum protein and have a similar pattern and distribution, could easily be analyzed by CE. However, a number of problems were encountered including interference from other compounds at the wavelength of detection, capillary overloading due to variability in sample protein concentration, and heterogeneous migration times. Early attempts by Chen et al.[78] to separate and identify urine proteins by CE involved the use of a dialysis membrane with a molecular mass cut-off at 14 kDa. The dialysate was electrophoresed using UV detection at 214 or 254 nm. However, a large number of peaks from small molecules and protein breakdown products were present. Jenkins et al.[79] evaluated the correlation between CE and high-resolution agarose gel electrophoresis (HRAGE) for the quantitation of Bence Jones protein (BJ) and albumin, both important in assessing patients with multiple myeloma. Using CE, identification and determination of migration times for a number of urine proteins was achieved. In addition, good correlation (0.93 for albumin and 0.95 for BJ) was found between CE and HRAGE. Shihabi[80] used CE to separate and quantitate myoglobin, a protein found in patients with disorders reflecting muscle tissue breakdown including burns and trauma infections. The analysis of myoglobin in urine appears simple but is difficult for two reasons: Hb, the main pigment protein in urine, can interfere with myoglobin detection; and myoglobin denatures rapidly (within a few hours) in urine. Using an uncoated capillary and a borate buffer, separation of myoglobin and Hb was achieved in less than 7 min. Similar to many CE analyses, addition of the buffer additive poly(ethylene glycol) improved the separation by providing a sieving matrix and limiting protein–wall interactions. Friedberg and Shihabi[81] also showed detection of urine proteins at 214 nm can be achieved by using either a wash step or cold ethanol precipitation to eliminate many interferences, whereas Oda et al.[82] demonstrated the utility of CE for identifying the presence of monoclonal proteins in urine. In a study of more than 200 patients, diagnostic information obtained by CE was comparable to AGE. More recently, Jenkins[83] demonstrated CE analysis of unconcentrated urine using a bare fused silica capillary where protein concentrations as low as 0.004 g L^{-1} could be detected (Figure 7). Under these conditions, BJ, intact immunoglobulin, and Tamm Horsfall protein, were separated and identified. In addition, electropherograms similar to those obtained using concentrated urine were obtained, thus eliminating the expense and time consumed for concentration procedures.

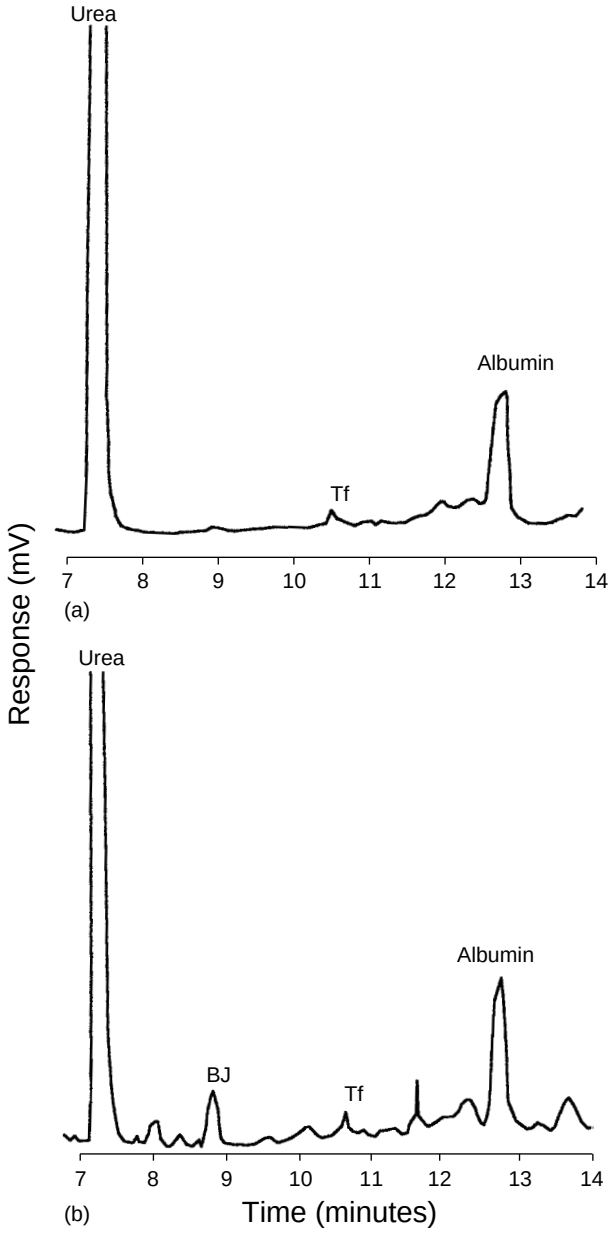

Figure 7 CZE separation of unconcentrated urine. Electrophoretic separation was carried out in a 72 cm × 50 µm fused silica capillary with buffer consisting of 150 mM boric acid (pH 7.5) and 1 mM calcium lactate. 40 µL of urine were diluted in 60 µL of buffer and vacuum injected for 5 s. (a) Detection of normal urine proteins urea, Tf and albumin. (b) Detection of four major urine proteins (0.04 g L^{-1}) including BJ. (Taken with permission from ref. 83.)

5.2 Capillary Gel Electrophoresis

Although one might expect that denaturing separations of proteins in sieving matrices should be easily extrapolated

from the gel format to the capillary, this has not been trivial. In fact, generating cross-linked polymer matrices or chemical gels (such as cross-linked polyacrylamide) in capillaries was difficult and suffered from technical and longevity problems. This inadequacy was solved with physical gels or pumpable polymer networks using polymers that were soluble, had low viscosity and were noncross-linked. The successful implementation of physical gels for use in DNA separations provided the format for denaturing protein analysis in capillaries. This involves standard heat denaturation of the protein sample in the presence of surfactant, such as sodium dodecyl sulfate (SDS), and separating in a sieving matrix such that separation is dependent on the molecular mass. Typically, the capillary is coated to eliminate EOF and is filled with a polymer sieving matrix buffer. The sample is then introduced either electrokinetically or via pressure injection under an applied electric field. As the proteins are SDS coated and hence negatively charged, they have anodic mobility and the separation must be carried out using reverse polarity (Figure 5b). CGE has begun to gain popularity for molecular weight determination of proteins but still possesses some limitations.

5.2.1 Denaturing Analysis of Proteins

SDS/PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) separation of proteins was first reported by Hjérten[84] who was able to separate human serum proteins using polyacrylamide as a sieving medium. Early work in this field was also carried out by Cohen and Karger[85] who demonstrated the utility of SDS/PAGE for separation of myoglobin fragments and a number standard proteins. These and other CGE separations were performed using cross-linked acrylamide as the chemical gel, which has been found to provide high resolution separation of low-molecular-weight biopolymers. However, several problems with the chemical gel-filled capillaries were realized. For instance, polyacrylamide absorbs at low-UV wavelengths, increasing the difficulty of high-sensitivity detection. Also, as a result of the rigidity of the gel, the lifetime of the capillary suffered. For example, Tsuji[86] found that, with a linear polyacrylamide gel-filled capillary, only a few runs were possible before the capillary had to be replaced. Although longevity was improved by the addition of ethylene glycol, the short capillary lifetime remains a major obstacle in implementing chemical gel-filled polymers for use in CE. Consequently, a search for a replacement for these permanent gels was undertaken.

Low-viscosity physical gels which are replaced after each run have become increasingly popular for CGE and can be used as noncross-linked polymer networks for the separation of high-molecular-weight SDS–protein complexes. Ganzler et al.[87] showed successful separation of SDS–protein complexes, ranging in molecular weight from 17 to 206 kDa using polymer networks and low-UV detection. With the polymer gel replaced after every run, a minimum of 300 separations were performed in the same capillary without a detectable loss in resolution. Shieh et al.[88] demonstrated the reproducibility and stability of capillary SDS–protein separations with seven protein mixtures that were separated in 15 min with less than 2% RSD (corrected migration time) in run-to-run and day-to-day reproducibility. Vecchione et al.[89] used denaturing CE conditions to separate platelet membranes from normal individuals and those with Glanzmann Thrombasthenia (GT), a hemorrhagic defect. In this study, the glycoproteins IIb and IIIa, which form heterodimers allowing for platelet–platelet cohesion, were analyzed. Using an eCAP SDS 14–200 fused silica capillary (Beckman) and reverse polarity, separation was achieved in less than 30 min. Healthy controls consistently showed the presence of glycoprotein IIb and IIIa, whereas these peaks were absent from electropherograms obtained from infected patients. Manabe et al.[90] employed linear polyacrylamide as a sieving matrix for separation of SDS-complexed plasma proteins. Here the effects of the buffer composition, concentration of the sieving matrix, and sample solution composition were examined. In addition, high-resolution size separation of SDS-complexed proteins was achieved. Major plasma proteins, including albumin, IgG, and Tf, were identified from their known molecular masses. Also, linear polyacrylamide capillary electrophoresis (LPA/CE) analysis of IgG myeloma sera revealed extremely high values of IgG, suggesting the methodology may be useful in screening serum samples. SDS/CGE is useful for the analysis of proteins with a large affinity for adsorption to surfaces. This has been demonstrated for the separation of fibronectin (FN), a cellular adhesion protein with a relative molecular mass (M_r) of about 400 kDa. Employing SDS denaturing conditions, the separation of FN from a series of protein standards was achieved (Figure 8).

The pharmaceutical/biotechnology industry was quick to recognize the potential for CE as a rapid quality control process. CE is especially attractive because it combines the quantification benefits of HPLC with the separating power of conventional gel techniques.[91,92] SDS/CGE has been compared to conventional gel SDS techniques (SDS/PAGE) for in-process control of antithrombin III (ATIII) and human clotting factor IX (FIX). In the case of ATIII, SDS/CE showed results comparable with conventional techniques with the advantage of a shorter analysis time. SDS/CGE also gave similar results for FIX but was less sensitive for detection of low-concentration proteins.

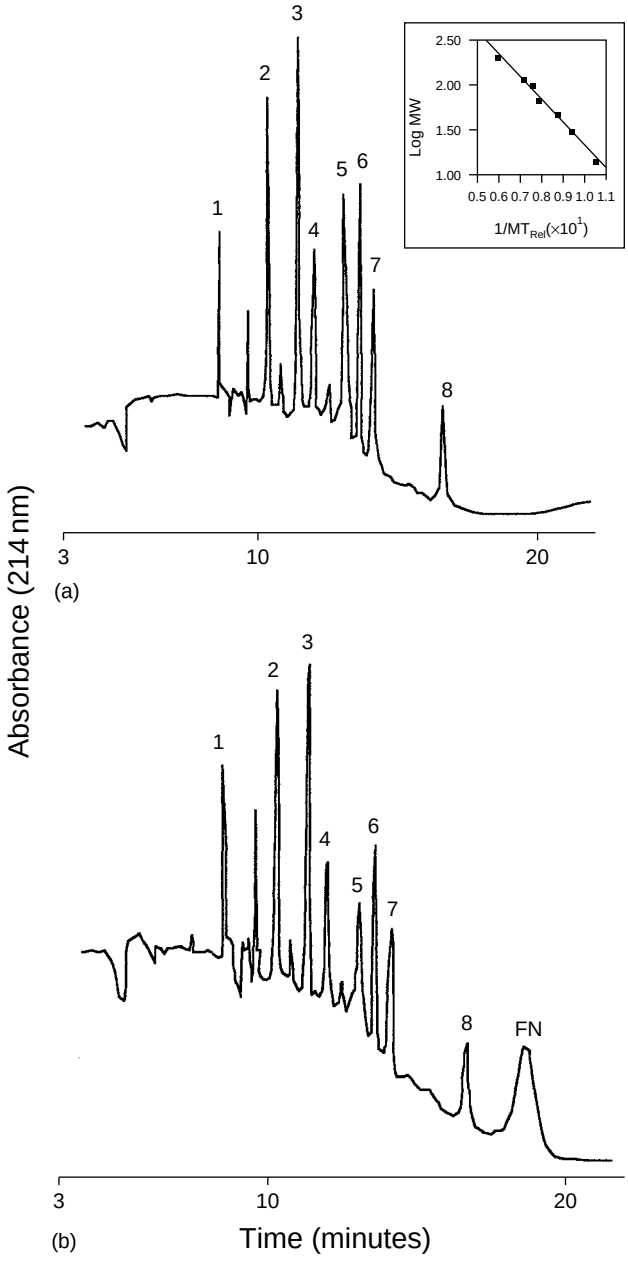

Figure 8 Denaturing SDS/CE. Electrophoretic separation was carried out in a 37 cm × 100 μm acrylamide coated capillary with gel buffer consisting of 1% SDS and 5% ethylene glycol at 14.1 kV with reverse polarity. Denatured FN in 1% SDS and 1% Trizma [tris(hydroxymethyl)aminomethane] was injected for 30 s. (a) Separation of molecular weight standards: 1 = Orange G, reference marker; 2 = α-lactalbumin, 14.2 kDa; 3 = carbonic anhydrase, 29 kDa; 4 = ovalbumin, 45 kDa; 5 = bovine serum albumin, 66 kDa; 6 = phosphorylase b, 97.4 kDa; 7 = β-galactosidase, 116 kDa; 8 = myosin, 210 kDa. The inset is a plot of log molecular weight versus inverse of relative migration time. The regression equation is $y = -26.035x + 3.908$, with $r^2 = 0.994$. (b) The separation of 1 mg mL^{-1} of FN (19.5 min) and dual injected standards. The calculated molecular weight for plasma FN based on the standard curve given in the inset of (a) is 374 kDa.

5.2.2 Nondenaturing Gel-based Analyses

In addition to denaturing protein separations, CGE can also be utilized for nondenaturing protein separations. For instance, Oda et al.[93] and Prasad et al.[94] have developed a separation method using borate buffer with a cellulose-based sieving matrix to allow for baseline resolution of Tf glycoforms. Of particular interest to these researchers is a region known to contain carbohydrate deficient transferrin (CDTf), Tf containing less sialic acid (and carbohydrate) than the tetrasialotransferrin found as a dominant form in normal individuals. The serum Tf glycoform distribution is affected by a number of pathological states. Included in these are carbohydrate deficient glycoprotein syndrome (CDGS), a pathological condition where the cellular machinery for protein glycosylation is defective. This disease manifests itself as a neurological disorder and is characterized by an increase in the CDTf (which is composed of the di-, mono-, and asialoforms) level. Chronic, excessive alcohol abuse has also been linked to an increase in CDTf levels in serum and, therefore, presents the possibility of monitoring excessive alcohol abuse using this methodology.[95] More recently, Trout et al.[96] have improved this assay for the analysis of CDTf in neat serum, circumventing the need for immunopurification. Using an FC-coated capillary to slow EOF and a borate buffer system, Tf sialoforms are resolved with the direct injection of diluted serum, allowing for resolution and identification of both normal and abnormal (alcoholic) Tf glycoform patterns (Figure 9). This presents the possibility for the rapid testing of compliance in both alcohol treatment and liver transplant clinics.

It is important to bear in mind that results obtained with the use of physical gel-filled capillaries can be influenced by capillary temperature. Guttman et al.[97] found dramatically different temperature effects with two different sieving matrices, poly(ethylene oxide) (PEO) and dextran. At elevated temperatures dextran chains are able to orient, through polymer–polymer interaction, into channel-like structures which yielded better resolution compared to that achieved at room temperature. The opposite effect was observed with PEO.

5.2.3 Ferguson Plots

Proteins undergoing post-translational modification, such as the addition of carbohydrates or lipids, do not bind SDS in the same ratio observed with unmodified proteins (1.4 g SDS per gram of protein). Consequently, gel-based molecular weight determination is often inaccurate due the lower charge to mass ratios, i.e. the molecular mass is overestimated. Ferguson[98] developed a method to correct for the variability in SDS binding which involves determining the mobilities of standard proteins

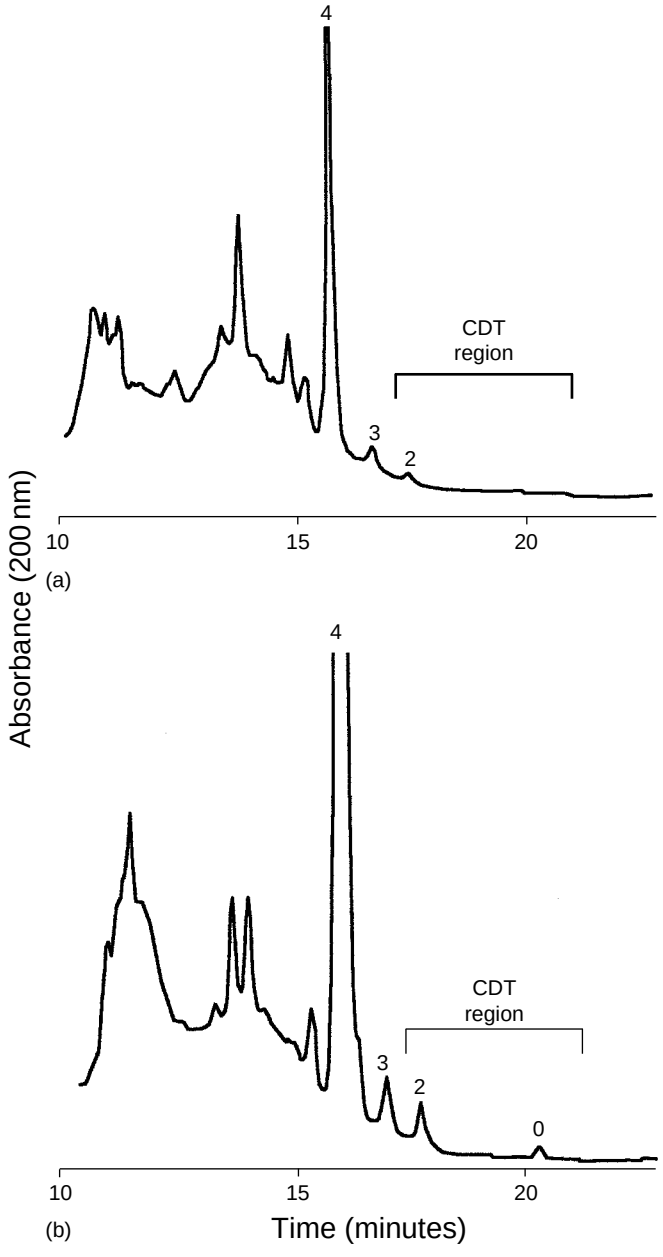

Figure 9 Separation of (a) normal and (b) alcoholic human serum Tf using CZE. Using an FC-coated capillary (J & W Scientific) and a buffer containing a polymer sieving matrix (0.5% hydroxyethyl cellulose) separation of Tf glycoforms is achieved in less than 25 min. The nonalcoholic serum profile has normal concentrations of the sialoforms in the CDTf region. In comparison, elevated concentrations of the disialoform and the presence of asialo-Tf are seen in the alcoholic profile. The peaks are 0, asialoform; 2, disialoform; 3, trisialoform; 4, tetrasialoform.

at several different gel concentrations. The slope of an initial plot of the log of reciprocal migration time versus gel concentration gives the negative of the retardation coefficient k_r. A second plot (log molecular weight as a function of k_r) yields a more universal calibration curve. Performing Ferguson analysis using multiple slab gels at different acrylamide concentrations has obvious disadvantages – the advantage of translating this analysis to CE with replaceable gels and an automated format is clear. Werner et al.[99] demonstrated this with a variety of proteins including α-lactalbumin. In most cases, Ferguson plots yielded better molecular weight estimates than SDS/PAGE. It is important to remember, however, that although abnormal charge-to-mass ratios are accounted for, abnormal protein shapes are not.

5.3 Capillary Isoelectric Focusing

IEF is an equilibrium protein separation technique that relies on the establishment of a pH gradient in the gel. Unlike density gradient centrifugation and gradient gel electrophoresis, which separate based on size or density, IEF separates molecules based on their pI. By establishing a pH gradient, molecules will migrate to the zone at which the pH is equal to their pI, i.e. their net charge is zero (Figure 5c). If diffusion occurs to any extent, the molecule enters a region with a different pH resulting in the analyte acquiring a charge and electromigrating back into the focused zone. As a result of a mechanism where electrophoretic migration counters diffusional effects (as long as voltage is applied to the system), IEF has enormous resolving power.

The analysis of polypeptides has been the most prominent application of this focusing technique, particularly for determining pIs of proteins, characterizing biological fluids and monitoring protein purification. IEF has also been exploited for the analysis of glycoform microheterogeneity and resolution of charge variants.

5.3.1 Technical Considerations

Consideration of the sample matrix is an integral component in developing CIEF methods. Attention must then be paid to selection of appropriate ampholytes, adjustment of salt concentration, and dilution or concentration of the protein sample if necessary.

Broad-range ampholytes (pH 3–10) are generally used for protein mixtures with a wide spectrum of pIs, whereas narrow-range ampholytes (pH 5–8) are utilized to achieve increased resolution. It is unfortunate that the very mechanism of separation (an ampholyte-based pH gradient) masquerades as one of the obstacles in CIEF protein separations. Specifically, ampholytes (essentially polypeptides) absorb UV radiation at 200 nm which, for protein separations, is the ideal wavelength of detection. Consequently, CIEF requires detection at 254 or 280 nm which results in a substantial decrease in sensitivity for protein analyses.

With respect to salt concentration, several problems are encountered when the sample salt concentration is too high. For example, increased current due to high ionic strength, loss of resolution and gradient compression during focusing are all complications associated with high salt concentration. Consequently, samples with salt concentrations higher than 50 mM should be desalted by dialysis, gel filtration or ultrafiltration prior to focusing. A companion problem associated with high sample concentrations is protein precipitation, which is a combined function of the high protein concentrations used in CIEF and the decreased solubility of individual proteins at their pI. For example, in addition to capillary occlusion, protein precipitation can result in irreproducible migration times, current instability and slow mobilization. To counteract these consequences, detergents such as urea are often employed as additives to increase the solubility of proteins. Nonionic detergents including Triton® X-100 and Nonidet P-40, and zwitterionic detergents such as CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate) are the most commonly employed solubilizing agents.[100]

To detect the protein zones, mobilization past the detector must be initiated. This can be achieved in a number of ways. Single-step CIEF uses EOF to transport the focused zones past the detector. Although EOF mobilization allows for more rapid analysis times than the two-step method, the loss of resolution and complications in relating the migration time to the pI have resulted in the two-step method remaining the most common approach to CIEF. In two-step CIEF, the protein bands are focused to steady state as described earlier. Coated capillaries are used to reduce the magnitude of the EOF so that the zones are not mobilized prior to the completion of focusing. Transport of the focused zones by the detector can then be performed by several different mechanisms including chemical and hydraulic mobilization. In chemical mobilization, the application of voltage is terminated after focusing is complete and the analyte or catholyte is replaced with mobilization reagent. An electric field is reapplied to facilitate mobilization. Neutral salts, such as sodium chloride, are the most common means of achieving chemical mobilization. Sodium serves as the nonproton cation in anodic mobilization and chloride as the nonhydroxyl anion in cathodic mobilization.[101]

An alternative is hydraulic mobilization which utilizes the application of pressure[102] or vacuum[103] to mobilize the focused protein zones past the detector. It is necessary to apply an electric field to the capillary simultaneously to decrease the diffusion of the protein zones and, hence, loss of resolution during hydraulic mobilization. The disadvantage of using this method of mobilization is that the parabolic shape of the hydrodynamic flow can decrease resolution. To minimize this effect, only weak forces are used. Chemical mobilization generally produces sharper peaks for the basic proteins whereas hydraulic mobilization increases the efficiency of mobilizing acidic proteins. In addition, chemical mobilization provides much better linearity when plotting migration time versus pI, as well as better reproducibility than its hydraulic counterpart.

5.3.2 Protein Isoelectric Focusing in Capillaries

CIEF is useful for analyzing proteins that, due to amino acid mutation, post-translational modification, or the binding of metals, have extremely small differences in pI. For example, CIEF has been utilized to resolve HbA (adult hemoglobin) and HbF (fetal hemoglobin) as well as abnormal Hb variants differing only by a single point mutation in the Hb gene. Zhu et al.[104] as well as other groups[105,106] demonstrated the separation of Hb A and F, and variants S and C (associated with various thalassemias). Although analysis times were somewhat lengthy (approximately 13 min), remarkable resolution of the globin chains was achieved. With an unorthodox approach, Yowell et al.[105] showed promising results focusing protein at the detector end of the capillary, significantly reducing the analysis time. In addition, Zhu et al.[107] also showed the utility of CIEF for separation of Hb G Philadelphia and Hb C, two Hb variants associated with Bart's disease, another potentially lethal hematological disorder. Hempe and Carver[106] examined CIEF for use in Hb-variant screening. Using conditions that provided separation in less than 15 min for rapid turnaround times in a clinical setting, they obtained quantitative data on Hb variants and identified several rare Hb types using CIEF. The quantitative results agreed well with conventional methods, thus confirming the utility of CIEF for routine assays. Hempe et al.[108] recently improved this assay by limiting capillary exposure to NaOH, consequently increasing capillary lifetime. In addition, improvements in the between-run reproducibility (9.6% RSD using the original method) enhance the quantitation precision. Using this novel methodology, diagnosis of HB s/β^+ thalassemia is possible in one analysis (Figure 10). Conventional clinical analysis methods would require at least three separate assays to make the same interpretation.

CIEF has also has been exploited for the separation and characterization of monoclonal antibodies. These biomolecules are challenging to separate as a consequence of their large size, tendency to aggregate, and heterogeneity due to post-translational modifications. Analysis of monoclonal antibodies including recombinant

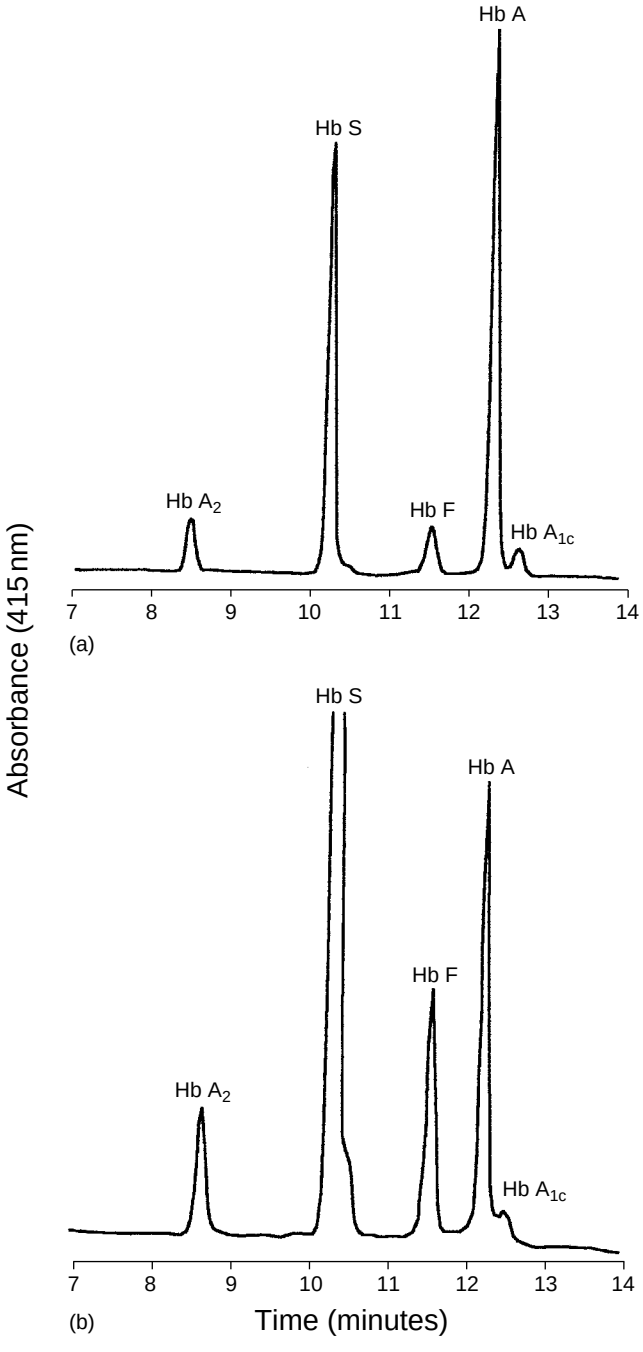

Figure 10 Separation of Hb variants using CIEF. Electrophoretic separation was carried out in a 27 cm × 50 μm DB-1 capillary (J & W Scientific) at 30 kV. Hydraulic mobilization was employed. Hemolysate was introduced by low pressure injection (3.5 kPa) for 10–30 s. (a) Abnormal control formulated to resemble blood with the sickle cell trait (HbS) and HbA_2 (observed in patients with β-thalassemia). (b) A blood sample from a patient with HB S/β^+ thalassemia. High-efficiency separation is necessary for identification of the underlying double heterozygous. Here elevated HbA_2 helps differentiate potential causes of elevated red blood cell count (RBC), low Hb concentration and low mean corpuscular volume (MCV). (Taken with permission from ref. 108.)

ATIII[109] and recombinant humanized monoclonal antibody HER2,[110] which are significant in the diagnosis of certain types of cancer, has become possible using CIEF techniques.

Kilár et al.[111] have developed a unique set-up for CIEF in uncoated capillaries. With this approach, a sandwich technique is employed where the ampholytes and analytes are injected in three steps: ampholyte solution, sample, and ampholyte solution. This allows for the use of homolog or heterolog (containing both broad and narrow range) ampholytes to be used. Also possible use of the sandwich approach is the detection of proteins with pI values outside the ampholyte range. The methodology was used for separation of Hb variants from normal and diabetic patients (Figure 11). The results obtained were similar to those using other conventional techniques.

5.3.3 Glycoproteins

Another important application of CIEF methodology is its utility for interrogating glycoform microheterogeneity. Glycoproteins often display microheterogeneity as a result of qualitative and quantitative differences in the location, length and content of the carbohydrate moieties linked to the polypeptide. This becomes important not only for understanding the function of microheterogeneity in biological systems, but also in monitoring the glycosylation pattern in the development and production of protein therapeutics. CIEF has been shown to be an excellent automated and quantitative method for analysis of glycoforms when detection sensitivity is not a critical issue. An example of the potential of this application is the analysis of human rtPA, a protein that degrades blood clots and is manufactured for treatment of myocardial infarction.[112–115] Although purification of rtPA is required for high-purity IEF, in some cases as many as 20 peaks persist. However, neuraminidase treatment simplifies the pattern, suggesting that the heterogeneity may be attributed to varied sialylation.[112] In addition, analysis of human Tf isoforms is possible using IEF. Kilár and Hjérten[116] successfully resolved the Tf glycoforms from normal individuals showing the expected di-, tri-, tetra-, penta-, and hexasialoforms. Also, Wu and Pawliszyn[117] implemented IEF methodology to examine changes in the pI of Tf due to iron binding.

5.4 Affinity Electrophoresis

Affinity capillary electrophoresis (ACE) refers to the analysis of general ligand–receptor interactions and can be used to (1) determine which sample components might undergo specific complexation, (2) investigate conditions surrounding this formation, and (3) quantitatively study

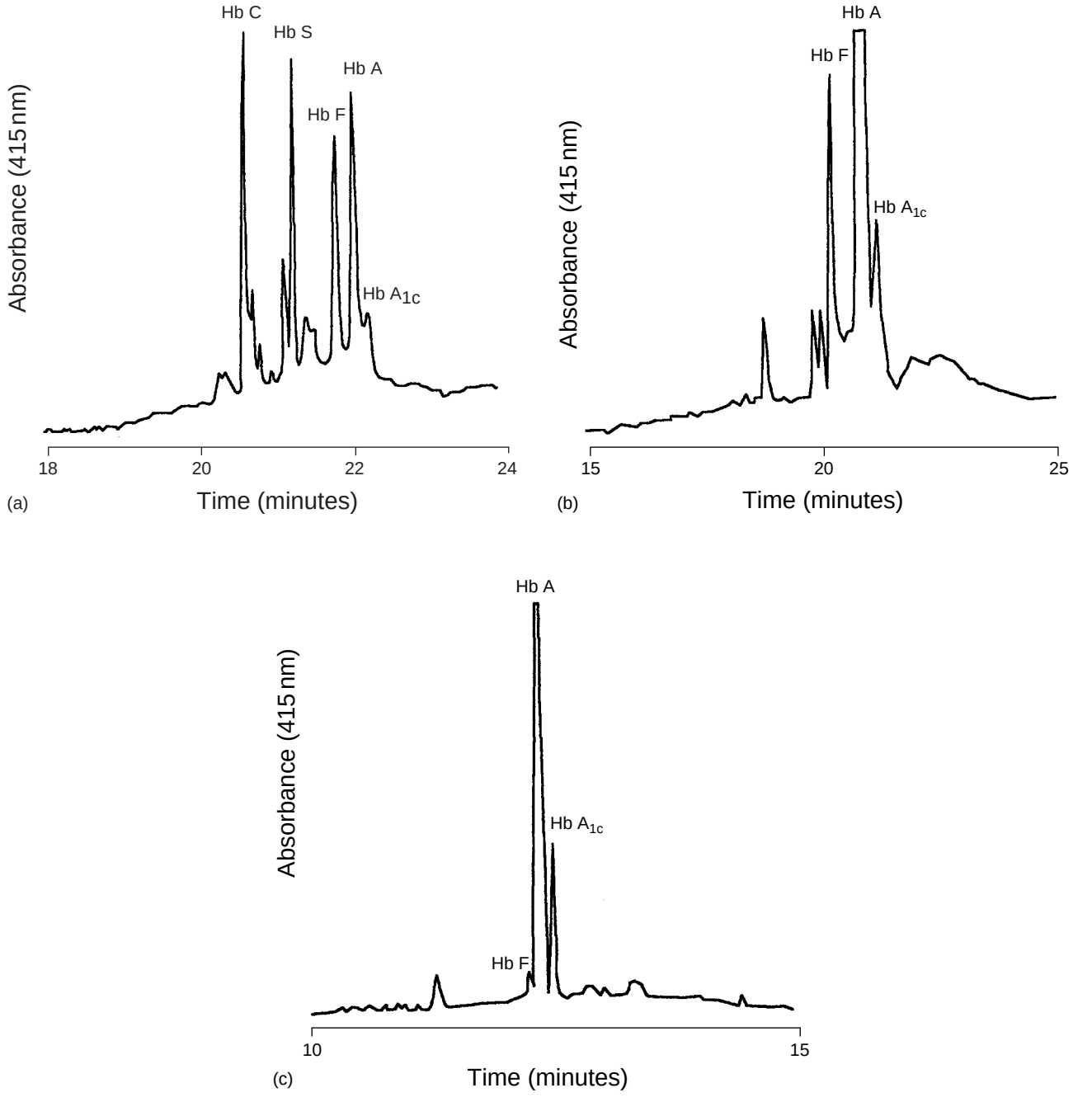

Figure 11 Novel CIEF methodology for the analysis of Hb variants. Electrophoretic separation was carried out in a 67 cm × 50 μm bare silica capillary at 20 kV. Single-step mobilization was employed with a 10 mM phosphoric acid analyte and a 20 mM sodium hydroxide catholyte. A heterologue ampholyte solution was employed using both a narrow (pH 6.7–7.7) and broad (pH 3.5–10) pH range. (a) Hb variant control (Isolab) solution (1 mg mL^{-1}). (b) Hb from a normal adult (1 mg mL^{-1}). (c) Hb from a diabetic patient (1 mg mL^{-1}). (Modified with permission from ref. 111.)

the complex. ACE can therefore be used to provide both qualitative and quantitative information about molecular interactions. This is possible because the electrophoretic mobility of a free protein changes upon binding with a ligand present in the electrophoresis buffer (Figure 5d). The difference in mobility is attributed to changes in charge, conformation, or both. Thus, by measuring the migration time of the complex as a function of ligand concentration the equilibrium constant may be calculated.

Classical gel methods may be employed to study these characteristics. However, they often require the

use of radio-labeled, fluorescently labeled, or secondary reagents that can be costly and difficult to use. In contrast, the capillary format does not demand purification of receptors and requires only small amounts of sample for analysis. As a result, the binding constants for the interaction of several receptors with their respective ligands can be determined simultaneously. Other advantages of the capillary format include temperature control, free solution separations and high-speed analysis.

5.4.1 Proteins and Glycoproteins

The utility of CE for determination of binding constants was realized when the first paper describing the use of CE for protein analysis appeared in 1986. As mentioned previously, Lauer and McManigill[118] found that protein migration times were retarded due to ionic interaction with the capillary wall. Thus, as many of the binding forces between proteins and ligands are ionic in nature, it was obvious these interactions could like-wise be examined. Consequently, two approaches for measuring binding constants using CE emerged – nonequilibrium and equilibrium methods. The latter of the two, commonly termed ACE, will be discussed briefly here. A comprehensive treatment of the subject is given by a number of published reviews on the subject.[119–121]

ACE is an ideal technique for examining the binding of metal ions to metal-binding proteins. This results from the fact that minute changes in the charge of the protein will effect the electrophoretic mobility and, consequently, the migration time. Heegard and Robey[122] examined the effects of Ca^{2+} concentration on the migration time of human C-reactive protein (CRP). They observed the expected decrease in mobility with increased ion concentration. Similar results were obtained by Kajiwara et al.[123] using the proteins calmodulin and parvalbumin.

Protein–protein interactions can also be studied using CE techniques. Early work by Grossman et al.[124] showed separation of antibody–antigen complexes using CE. Nielsen et al.[125] extended these findings by separating free hGH and antibody from the hGH–antibody complex. Although protein–protein measurements (as a result of their size) can be accomplished by a host of other techniques which are equal or superior to ACE, this technology offers advantages when the proteins of interest are present in limited quantities.

A diagnostic test for the scrapie prion protein (PrP^{SC}), a transmissible spongiform encephalopathy (TSE) found in sheep and goats which causes progressive degenerative disorders to the CNS, has been designed using CE competitive immunoassay.[126] Here CZE with LIF was used to detect the fluorescein-labeled peptides from PrP^{SC}. Antibodies for each peptide were used and antibody-bound peptides were separated from unbound peptides. Using PrP^{SC} extracted from infected sheep, detection of 135 pg of PrP^{SC} was possible. However, protein extracted from normal sheep did not compete for bound peptides, indicating absence of the antigen. Another clinical application of ACE involves the investigation of the epitope on the human immunodeficiency virus (HIV) recognized by a monoclonal antibody.[127] Antibodies to the protein p24 are among the first to appear in individuals infected with HIV.[128–131] In fact, the presence of this antibody is used as a screen for HIV in blood samples.[132–134] A correlation between a decreased level of p24 antibodies and the progression of HIV to full blown AIDS has been determined.[135–137] Consequently, an investigation of elements eliciting antibodies to p24 may be useful in developing an HIV vaccine. Previous excision-mass spectrometry (MS) experiments determined the epitope sequence of HIV p24 to be ^{102}HPVHAGPIAPG^{112}.[138] Using the ACE methodology, Qian and Tomer[127] identified two other sequences which exhibited weaker binding. These interactions were later characterized as nonspecific. However, a strong affinity of the antibody was found with the peptide VHPVHAGPIAP.

6 CONCLUDING REMARKS

Although only a fraction of the literature associated with the use of CE for protein analysis has been covered here, the growing acceptance of capillary-based analysis as an alternative to the conventional slab and tube gel formats, and as a complementary technique to HPLC, is clear. The advantages driving the interest in this technology include the simplicity of a system that can provide electrophoretic protein analysis in free solution, with multimode capabilities, minimal reagent and sample requirements, high efficiencies, on-line detection, and all in an automated format. This allows for rapid definition of optimal conditions as well as experimentation with numerous modes in search of a suitable method for a particular protein(s).

Unlike other analytes, the CE-based analysis of proteins is complicated by the predisposition of these solutes for adsorption to the capillary wall. One must, therefore, be prepared to seek a compromise between conditions ideal for the stability of the protein(s) of interest and those which minimize adsorption, so that separation efficiency is not destroyed. Capillary wall deactivation can be accomplished using a battery of approaches including covalent modification of the surface as well as dynamic deactivation. The latter is a largely

unexploited area that warrants exploration based on the simplicity of surface passivation via simple modification of the separation buffer.

It is clear that the same multimode flexibility associated with protein analysis by gel electrophoresis and HPLC is applicable to CE. Consequently, the simplicity of CE will allow for accelerated method development for applications ranging from purity characterization and microheterogeneity analysis to pI determination and clinical marker detection. As methods for protein analysis evolve, one can expect to see the development of protein CE buffer systems with improved compatibility for detectors such as MS. This will facilitate the use of hyphenated techniques such as (CE/MS),[139] where CE is used as the first dimension with MS as the detector, and improve capabilities for structural identification and, ultimately, the establishment of functional properties of individual proteins and glycoproteins. The future will also see other hyphenated techniques furthered in their development where the speed of CE allows it to function as the second dimension. The pioneering work of Jorgenson's group with the development of liquid chromatography/capillary electrophoresis (LC/CE)[140] presents the possibility of revolutionizing protein analysis, especially if the two-dimensional system can be interfaced with MS as the detector. Technological advances such as these, combined with advances in method development, including capillary coating and detection, are sure to make protein analysis by CE commonplace in the future.

ABBREVIATIONS AND ACRONYMS

ACE	Affinity Capillary Electrophoresis
AGE	Agarose Gel Electrophoresis
ATIII	Antithrombin III
BJ	Bence Jones Protein
CAE	Cellulose Acetate Electrophoresis
CDGS	Carbohydrate Deficient Glycoprotein Syndrome
CDTf	Carbohydrate Deficient Transferrin
CE	Capillary Electrophoresis
CGE	Capillary Gel Electrophoresis
CIEF	Capillary Isoelectric Focusing
CITP	Capillary Isotachophoresis
CNS	Central Nervous System
CRP	C-reactive Protein
CTAB	Cetyltrimethylammonium Bromide
CZE	Capillary Zone Electrophoresis
DAB	Diaminobutane
EDTA	Ethylenediaminetetraacetic Acid
EOF	Electroosmotic Flow
FIX	Human Clotting Factor IX
FN	Fibronectin
GC	Gas Chromatography
GFP	Green Fluorescent Protein
GT	Glanzmann Thrombasthenia
Hb	Hemoglobin
HbA	Adult Hemoglobin
HbF	Fetal Hemoglobin
hCG	Human Chorionic Gonadotropin
hGH	Human Growth Hormone
HIV	Human Immunodeficiency Virus
HPLC	High-performance Liquid Chromatography
HRAGE	High-resolution Agarose Gel Electrophoresis
IEF	Isoelectric Focusing
IS	Immunosubtraction
LC/CE	Liquid Chromatography/Capillary Electrophoresis
LIF	Laser-induced Fluorescence
LPA/CE	Linear Polyacrylamide Capillary Electrophoresis
MCV	Mean Corpuscular Volume
MEKC	Micellar Electrokinetic Chromatography
M_r	Relative Molecular Mass
MS	Mass Spectrometry
OLP	Ohm's Law Plot
PEO	Poly(ethylene oxide)
pI	Isoelectric Point
PrP^{SC}	Scrapie Prion Protein
RBC	Red Blood cell Count
rHuEPO	Human Erythropoietin
RSD	Relative Standard Deviation
rtPA	Recombinant Tissue Plasminogen Activator
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Tf	Transferrin
βTP	β-trace Protein
TSE	Transmissible Spongiform Encephalopathy
UV	Ultraviolet
UV/VIS	Ultraviolet/visible

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Glycoprotein Analysis: General Methods

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry • Serum Proteins

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction

REFERENCES

1. A. Tiselius, 'A New Apparatus for Electrophoretic Analysis of Colloidal Mixtures', *Trans. Faraday Soc.*, **33**, 524 (1937).
2. S. Hjérten, 'Agarose as an Anticonvection Agent in Zone Electrophoresis', *Biochem. Biophys. Acta*, **53**, 514–517 (1961).
3. S. Raymond, L. Weintraub, 'Acrylamide Gel as a Supporting Medium for Zone Electrophoresis', *Science*, **130**, 711 (1959).
4. J.W. Jorgenson, K.D. Lukacs, 'Zone Electrophoresis in Open-tubular Glass Capillaries', *Anal. Chem.*, **53**, 1298–1302 (1981).
5. S. Hjérten, 'Free Zone Electrophoresis', *Chromatogr. Rev.*, **9**, 122–219 (1967).
6. B.L. Karger, Y.H. Chu, F. Foret, 'Capillary Electrophoresis of Proteins and Nucleic Acids', *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 579–610 (1995).
7. P. Pancholi, R.P. Oda, P.S. Mitchell, J.P. Landers, D.H. Persomg, 'Diagnostic Detection of Herpes Simplex and Hepatitis C Viral Amplicons by Capillary Electrophoresis: Comparison with Southern Blot Detection', *Mol. Diagnostics*, **2**, 27–37 (1997).
8. R.P. Oda, B.J. Madden, J.P. Landers, 'Capillary Electrophoretic Analysis of Glycoproteins and Glycoprotein-derived Oligosaccharides', in *Advances in Chromatography*, eds. P.R. Brown, E. Grushka, Marcel Dekker, New York, 163–196, 1996.
9. A.M. Dougherty, N. Cooke, P. Shieh, 'Capillary Surface Modification in Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, 675–715, 1997.
10. T.W. Rademacher, R.B. Parekh, R.A. Dwek, 'Glycobiology', *Ann. Rev. Biochem.*, **57**, 785–838 (1988).
11. S. Thompson, E. Dargan, I.D. Griffiths, C.A. Kelly, G.A. Turner, 'The Glycosylation of Haptoglobin in Rheumatoid Arthritis', *Clin. Chim. Acta*, **220**, 107–114 (1993).
12. G. DeJong, R. Feelders, W.L. Van Noort, H.G. Van Eijk, 'Transferrin Microheterogeneity as a Probe in Normal and Disease States', *Glycoconjugate J.*, **12**, 219–226 (1995).
13. J.C. Giddings, *Dynamics of Chromatography, Part I, Principles and Theory*, Marcel Dekker, New York, 1965.
14. X. Huang, W.F. Coleman, R.N. Zare, 'Analysis of Factors Causing Peak Broadening in Capillary Zone Electrophoresis', *J. Chromatogr.*, **480**, 95–110 (1989).
15. S. Swedberg, personal communication.
16. R.M. McCormick, 'Capillary Zone Electrophoretic Separations of Peptides and Proteins Using Low pH Buffers in Modified Silica Capillaries', *Anal. Chem.*, **60**, 2322–2328 (1988).
17. R.P. Oda, J.P. Landers, 'Application of Capillary Electrophoresis to the Analysis of Proteins', *Bioseparation*, **5**, 315–328 (1995).
18. R.J. Nelson, A. Paulus, A.S. Cohen, A. Guttman, B.L. Karger, 'Use of Peltier Thermoelectric Devices to Control Column Temperature in High-performance Capillary Electrophoresis', *J. Chromatogr.*, **480**, 111–127 (1989).
19. H.H. Lauer, D. McManigill, 'Capillary Zone Electrophoresis of Proteins in Untreated Fused Silica Tubing', *Anal. Chem.*, **58**, 166–170 (1986).
20. S. Hjérten, 'High Performance Electrophoresis Elimination of Electroendoosmosis and Solute Adsorption', *J. Chromatogr.*, **347**, 191–198 (1985).
21. W.J. Lambert, D.L. Middleton, 'pH Hysteresis Effect with Silica Capillaries in Capillary Zone Electrophoresis', *Anal. Chem.*, **62**, 1585–1587 (1990).
22. G.M. Bruin, J.P. Chang, R.H. Kulhmann, K. Zegers, J.C. Kraak, H. Poppe, 'Capillary Zone Electrophoretic Separations of Proteins in Polyethylene Glycol-modified Capillaries', *J. Chromatogr.*, **471**, 429–436 (1989).
23. W. Nashabeh, Z.E. Rassi, 'Capillary Zone Electrophoresis of Proteins with Hydrophilic Fused-silica Capillaries', *J. Chromatogr.*, **559**, 367–383 (1991).
24. S. Hjérten, K. Kubo, 'A New Type of pH- and Detergent-stable Coating for Elimination of Electroendoosmosis and Adsorption in (Capillary) Electrophoresis', *Electrophoresis*, **14**, 390–395 (1993).
25. M. Huang, J. Plocek, M. Novotny, 'Hydrolytically Stable Cellulose-derivative Coatings for Capillary Electrophoresis of Peptides, Proteins and Glycoconjugates', *Electrophoresis*, **16**, 396–401 (1995).
26. J.K. Towns, F.E. Regnier, 'Capillary Electrophoretic Separations of Proteins using Nonionic Surfactant Coatings', *Anal. Chem.*, **63**, 1126–1132 (1991).
27. J.T. Smith, Z.E. Rassi, 'Capillary Zone Electrophoresis of Biological Substances with Fused Silica Capillaries having Zero or Constant Electroosmotic Flow', *Electrophoresis*, **14**, 396–406 (1993).
28. Y. Liu, F. Ruonong, J. Gu, 'Epoxy Resin Coating for Capillary Zone Electrophoretic Separation of Basic Proteins', *J. Chromatogr. A*, **723**, 157–167 (1996).
29. J.E. Wiktorowicz, J.C. Colburn, 'Separation of Cationic Proteins Via Charge Reversal in Capillary Electrophoresis', *Electrophoresis*, **11**, 769–773 (1991).
30. T. Kaneta, T. Shunitz, H. Yoshida, 'Improvement of Resolution in Capillary Electrophoretic Separation of Catecholamines by Complex Formation with Boric Acid and Control of Electroosmsis with a Cationic Surfactant', *J. Chromatogr.*, **538**, 385–391 (1991).
31. E. Cordova, J. Gao, G.M. Whitesides, 'Noncovalent Polycationic Coating for Capillaries in Capillary Electrophoresis of Proteins', *Anal. Chem.*, **69**, 1370–1379 (1997).

32. M.X. Li, L. Liu, J. Wu, D.M. Lubman, 'Use of Polybrene Capillary Coating in Capillary Electrophoresis for Rapid Analysis of Hemoglobin Variants with On-line Detection via an Ion Trap Storage/Reflectron Time-of Flight Mass Spectrometer', *Anal. Chem.*, **69**, 2451–2456 (1997).
33. K. Tsuji, R.J. Little, 'Charge-reversed, Polymer-coated Capillary Column for the Analysis of a Recombinant Chimeric Glycoprotein', *J. Chromatogr.*, **594**, 317–324 (1992).
34. R.P. Oda, B.J. Madden, T.C. Spelsberg, J.P. Landers, 'α,w-Bis-quaternary Ammonium Alkanes as Effective Buffer Additives for Enhanced Capillary Electrophoretic Separation of Glycoproteins', *J. Chromatogr.*, **680**, 85–92 (1994).
35. M. Traverna, A. Balliet, D. Biou, M. Schulter, R. Werner, D. Ferrier, 'Analysis of Carbohydrate-mediated Heterogeneity and Characterization of N-linked Oligosaccharides of Glycoproteins by High-performance Capillary Electrophoresis', *Electrophoresis*, **13**, 359–366 (1992).
36. D.E. Morbeck, B.J. Madden, D.J. McCormick, 'Analysis of the Microheterogeneity of the Glycoprotein Chorionic Gonadotropin with High-performance Capillary Electrophoresis', *J. Chromatogr.*, **680**, 217–224 (1994).
37. J.S. Green, J.W. Jorgenson, 'Minimizing Adsorption of Proteins on Fused Silica in Capillary Zone Electrophoresis by the Addition of Alkali Metal Salts to the Buffers', *J. Chromatogr.*, **478**, 63–70 (1989).
38. N.E. Good, G.D. Winget, W. Winter, T.N. Conolly, S. Jzawa, R.M.M. Singh, 'Hydrogen Ion Buffers for Biological Research', *Biochemistry*, **5**, 467–477 (1966).
39. N.E. Good, S. Jzawa, 'Hydrogen Ion Buffers', *Methods Enzymol.*, **24**, 53–68 (1972).
40. M.M. Bushey, J.W. Jorgenson, 'Performance of Carbohydrate-modified Fused-silica Capillaries for Separation of Proteins by Zone Electrophoresis', *J. Chromatogr.*, **480**, 339–349 (1989).
41. M.M. Bushey, J.W. Jorgenson, 'Capillary Electrophoresis of Proteins in Buffers containing High Concentrations of Zwitterionic Salts', *J. Chromatogr.*, **480**, 301–310 (1989).
42. S. Beale, 'Capillary Electrophoresis', *Anal. Chem.*, **70**, 279R–300R (1998).
43. G. Schomburg, 'Coated Capillaries in High-performance Capillary Electrophoresis', in *High-performance Capillary Electrophoresis*, ed. M.G. Khaledi, John Wiley and Sons, New York, 481–523, 1998.
44. J. Gao, F.A. Gomez, R. Harter, G.M. Whitesides, 'Determination of the Effective Charge of a Protein in Solution by Capillary Electrophoresis', *Proc. Natl. Acad. Sci. USA*, **91**, 12 027–12 030 (1994).
45. F. Kálmán, S. Ma, A. Hodel, R.O. Fox, C. Harvath, 'Charge and Size Effects in the Capillary Zone Electrophoresis of Nuclease A and its Variants', *Electrophoresis*, **16**, 595–603 (1995).
46. J.C. Giddings, *Unified Separation Science*, John Wiley and Sons, New York, 1991.
47. R.P. Oda, J.P. Landers, 'Introduction to Capillary Electrophoresis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, 1–47, 1997.
48. R.G. Nielson, G.S. Sittampalam, E.C. Rickard, 'Capillary Zone Electrophoresis of Insulin and Growth Hormone', *Anal. Biochem.*, **177**, 20–26 (1989).
49. G.M. Korf, J.P. Landers, D.J. O'Kane, 'Capillary Electrophoresis with Laser-induced Fluorescence Detection for the Analysis of Free and Immune-complexed Green Fluorescent Protein', *Anal. Biochem.*, **251**, 210–218 (1997).
50. A. Hiraoka, T. Arato, I. Tominaga, A. Anjyo, 'Capillary Electrophoretic Analyses of β-trace Protein and Other Low-molecular-weight Proteins in Cerebrospinal Fluid from Patients with Central Nervous System Diseases', *J. Pharm. Biomed. Anal.*, **15**, 1257–1263 (1997).
51. F. Kilar, S. Hjérten, 'Separation of Human Tf Isoforms by Carrier-free High Performance Zone Electrophoresis and Isoelectric Focusing', *J. Chromatogr.*, **480**, 351–357 (1989).
52. A. Tran, S. Park, P.J. Lisi, O. Huynh, R.R. Ryall, P.A. Lane, 'Separation of Carbohydrate-mediated Microheterogeneity of Recombinant Human Erythropoietin by Free Solution Capillary Electrophoresis', *J. Chromatogr.*, **542**, 459–471 (1991).
53. J.P. Landers, R.P. Oda, B.J. Madden, T.C. Spelsberg, 'High-performance Capillary Electrophoresis of Glycoproteins: The Use of Modifiers of Electroosmotic Flow for Analysis of Microheterogeneity', *Anal. Biochem.*, **205**, 115–124 (1992).
54. J.P. Landers, R.P. Oda, B.J. Madden, T.C. Spelsberg, 'High-performance Capillary Electrophoresis of Glycoproteins: The Use of Modifiers of Electroosmotic Flow for Analysis of Microheterogeneity', *Anal. Biochem.*, **205**, 115–124 (1992).
55. E. Watson, F. Yao, 'Capillary Electrophoretic Separation of Human Recombinant Erythropoietin (r-HuEPO) Glycoforms', *Anal. Biochem.*, **210**, 389–393 (1993).
56. D.E. Morbeck, B.J. Madden, D.J. McCormick, 'Analysis of the Microheterogeneity of the Glycoprotein Chorionic Gonadotropin with High-performance Capillary Electrophoresis', *J. Chromatogr.*, **680**, 217–224 (1994).
57. R.P. Oda, B. Madden, S. Swedberg, T.C. Spelsberg, J.P. Landers, 'Mechanistic Aspects of Ovalbumin Analysis by Capillary Electrophoresis: Part I: Impact of α,w-Diaminoalkanes and Capillary Surface Characterization – Dynamic Deactivation', *Anal. Chem.*, (submitted).
58. S. Swedberg, R.P. Oda, B. Madden, J.P. Landers, 'Mechanistic Aspects of Ovalbumin Analysis by Capillary Electrophoresis Part II: Impact of Buffer Constituents and Capillary Surface Characteristics – Bonded Phase Deactivation', *Anal. Chem.*, (submitted).

59. C.N. Ong, L.S. Liau, Y.H. Ong, 'Separation of Globins using Free Zone Capillary Electrophoresis', *J. Chromatogr.*, **576**, 346–350 (1992).
60. G. Klein, C. Jolliff, 'Capillary Electrophoresis for the Routine Clinical Laboratory', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, 419–457, 1997.
61. N. Ishioka, N. Lyori, J. Noji, S. Kurioka, 'Detection of Abnormal Hemoglobin by Capillary Electrophoresis and Structural Identification', *Biomed. Chromatogr.*, **6**, 224–226 (1992).
62. G.A. Ross, P. Lorkin, D. Perrett, 'Separation of Tryptic Digest Mapping of Normal and Variant Hemoglobins by Capillary Electrophoresis', *J. Chromatogr.*, **636**, 69–79 (1993).
63. F.-T.A. Chen, C.-M. Liu, Y.-Z. Hsieh, J.C. Sternberg, 'Capillary Electrophoresis – A New Clinical Tool', *Clin. Chem.*, **37**, 14–19 (1991).
64. J.W. Kim, J.H. Park, J.W. Park, H.J. Doh, K.-J. Lee, 'Quantitative Analysis of Serum Proteins Separated by Capillary Electrophoresis', *Clin. Chem.*, **39**, 689–692 (1991).
65. F.-T.A. Chen, 'Rapid Protein Analysis by Capillary Electrophoresis', *J. Chromatogr.*, **559**, 445–453 (1991).
66. M.J. Gordon, K.J. Lee, A.A. Arias, R.N. Zare, 'Protocol for Resolving Protein Mixtures in Capillary Zone Electrophoresis', *Anal. Chem.*, **63**, 69–72 (1991).
67. R. Clark, J. Katzmann, E. Wiegert, C. Namyst-Goldberg, L. Sanders, R. Oda, R. Kyle, J.P. Landers, 'Rapid Capillary Electrophoretic Analysis of Human Serum Proteins: Qualitative Comparison with High-throughput Agarose Gel Electrophoresis', *J. Chromatogr.*, **744**, 205–213 (1996).
68. J. Katzmann, R. Clark, E. Wiegert, E. Sanders, R. Oda, R. Kyle, C. Goldberg, J.P. Landers, 'Identification of Monoclonal Proteins in Serum: A Quantitative Comparison of Acetate, Agarose Gel and Capillary Electrophoresis', *Electrophoresis*, **18**, 1775–1780 (1997).
69. C.R. Jolliff, C.R. Blessum, 'Comparison of Serum Protein Electrophoresis by Agarose Gel and Capillary Zone Electrophoresis in a Clinical Setting', *Electrophoresis*, **18**, 1781–1784 (1997).
70. J.A. Katzmann, R. Clark, E. Sanders, J.P. Landers, R.A. Kyle, 'Prospective Study of Serum Protein Capillary Zone Electrophoresis and Immunotyping of Monoclonal proteins by Immunosubtraction', *Am. J. Clin. Pathol.*, **110**, 503–509 (1998).
71. M.A. Jenkins, M.D. Guerin, 'Quantification of Serum Proteins using Capillary Electrophoresis', *Ann. Clin. Biochem.*, **32**, 493–497 (1995).
72. V. Dolnik, 'Capillary Zone Electrophoresis of Serum Proteins: Study of Separation Variables', *J. Chromatogr.*, **709**, 99–110 (1995).
73. R. Lehmann, H. Liebich, G. Grubler, W. Voetler, 'Capillary Electrophoresis of Human Serum Proteins and Apolipoproteins', *Electrophoresis*, **16**, 998–1001 (1995).
74. L.M. Silverman, R.H. Christenson, 'Amino Acids and Proteins', in *Tiez Textbook of Clinical Chemistry*, eds. C.A. Burtis, E.R. Ashwood, W.B. Saunders (publisher), Philadelphia, 717–712, 1994.
75. G.B. Schuann, S.C. Schweitzer, 'Examination of Urine', in *Clinical Diagnosis and Management by Laboratory Methods*, ed. J.B. Henry, W.B. Saunders (publisher), Philadelphia, 381–444, 1991.
76. R.S. Coltran, V. Kuman, S.L. Robbins, *Robbins Pathological Basis of Disease*, W.B. Saunders (publisher), Philadelphia, 1994.
77. L.J. Silverman, R.H. Christenson, G.H. Grant, 'Amino Acids and Proteins', in *Textbook of Clinical Chemistry*, ed. N.W. Tiez, W.B. Saunders (publisher), Philadelphia, 519–618, 1986.
78. F.-T.A. Chen, C.-M. Liu, Y.-Z. Hsieh, J.C. Sternberg, 'Capillary Electrophoresis – A New Clinical Tool', *Clin. Chem.*, **37**, 14–19 (1991).
79. M.A. Jenkins, T.D. O'Leary, M.D. Guerin, 'Identification and Quantitation of Human Urine Proteins by Capillary Electrophoresis', *J. Chromatogr. B*, **662**, 108–112 (1994).
80. Z.K. Shihabi, 'Myoglobinuria Detection by Capillary Electrophoresis', *J. Chromatogr. B*, **669**, 53–58 (1995).
81. M.A. Friedberg, Z.K. Shihabi, 'Urine Protein Analysis by Capillary Electrophoresis', *Electrophoresis*, **18**, 1836–1841 (1997).
82. R.P. Oda, R. Clark, J.A. Katzmann, J.P. Landers, 'Capillary Electrophoresis as a Clinical Tool for the Analysis of Protein in Serum and Other Body Fluids', *Electrophoresis*, **18**, 1715–1723 (1997).
83. M.A. Jenkins, 'Clinical Application of Capillary Electrophoresis to Unconcentrated Human Urine Proteins', *Electrophoresis*, **18**, 1842–1846 (1997).
84. S. Hjérten, 'High-performance Electrophoresis: The Electrophoretic Counterpart of High-performance Liquid Chromatography', *J. Chromatogr.*, **270**, 1–6 (1983).
85. A.S. Cohen, B.L. Karger, 'High-performance Sodium Dodecyl Sulfate Polyacrylamide Gel Capillary Electrophoresis of Peptides and Proteins. *J. Chromatogr.*, **397**, 409–417 (1987).
86. K. Tsuji, 'Sodium Dodecyl Sulfate–Polyacrylamide Gel-filled Capillary Column for the Detection of Recombinant Biotechnology-derived Proteins', *J. Chromatogr.*, **550**, 823–830 (1991).
87. K. Ganzler, R.S. Greve, A.S. Cohen, B.L. Karger, A. Guttman, N.C. Cooke, 'High-performance Capillary Electrophoresis of SDS–Protein Complexes using UV-transparent Polymer Networks', *Anal. Chem.*, **64**, 2665–2670 (1992).
88. P. Shieh, D. Hoang, A. Guttman, N. Cooke, 'Capillary Sodium Dodecyl Sulfate Gel Electrophoresis of Proteins – I. Reproducibility and Stability', *J. Chromatogr.*, **676**, 219–226 (1994).

89. G. Vecchione, M. Margaglione, E. Grandone, D. Colaizzo, G. Cappucci, N. Giuliani, M. d'Addedda, G. D'Andrea, M. Nobile, A. Amoriello, G. Di Minno, 'Identifying Human Platelet Glycoproteins IIb and IIIa by Capillary Electrophoresis', *Electrophoresis*, **19**, 1468–1474 (1998).
90. T. Manabe, H. Oota, J. Mukai, 'Size Separation of Sodium Dodecyl Sulfate Complexes of Human Plasma Proteins by Capillary Electrophoresis Employing Linear Polyacrylamide as a Sieving Polymer', *Electrophoresis*, **19**, 2308–2316 (1998).
91. K.G. Moorehouse, C.A. Eusebio, G. Hunt, A.B. Chen, 'Rapid One-step Capillary Isoelectric Focusing Method to Monitor Charged Glycoforms of Recombinant Human Tissue-type Plasminogen Activator', *J. Chromatogr. A*, **717**, 61–69 (1995).
92. H.P. Bietlot, M. Girard, 'Analysis of Recombinant Human Erythropoietin in Drug Formulations by High-performance Capillary Electrophoresis', *J. Chromatogr. A*, **759**, 177–184 (1997).
93. R.P. Oda, R. Prasad, R.L. Stout, D. Coffin, W.P. Patton, D.L. Kraft, J.F. O'Brien, J.P. Landers, 'Capillary Electrophoresis-based Separation of Transferrin Sialoforms in Patients with Carbohydrate-deficient Glycoprotein Syndrome', *Electrophoresis*, **18**, 1819–1826 (1997).
94. R. Prasad, R.L. Stout, D. Coffin, J. Smith, 'Analysis of Carbohydrate Deficient Transferrin by Capillary Zone Electrophoresis', *Electrophoresis*, **18**, 1814–1818 (1997).
95. H. Stibler, O. Sydow, S. Van Den Berg, 'Quantitative Estimation of Abnormal Microheterogeneity of Serum Transferrin in Alcoholics', *Pharmacol. Biochem. Behav.*, **13**, 47–51 (1980).
96. A.L. Trout, A. DiMartini, T. Lane, R. Prasad, D. Coffin, A.F. Hühmer, J.P. Landers, 'Direct Capillary Electrophoresis-based Detection of Carbohydrate Deficient Transferrin (CDTf) in Neat Serum', *Electrophoresis* (submitted).
97. A. Guttman, J. Horvath, N. Cooke, 'Influence of Temperature on the Sieving Effect of Different Polymer Matrices in Capillary SDS Gel Electrophoresis of Proteins', *Anal. Chem.*, **65**, 199–203 (1993).
98. K.A. Ferguson, 'Starch-Gel Electrophoresis – Application to the Classification of Pituitary Proteins and Polypeptides', *Metabolism*, **13**, 985–1002 (1964).
99. W.E. Werner, D.M. Demorest, J. Stevens, J.E. Wiktorowicz, 'Size-dependent Separation of Proteins Denatured in SDS by Capillary Electrophoresis using a Replaceable Sieving Matrix', *Anal. Biochem.*, **212**, 253–258 (1993).
100. R. Rodriguez-Diaz, T. Wehr, M. Zhu, V. Levi, 'Capillary Isoelectric Focusing', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, 101–138, 1997.
101. S. Hjérten, J.L. Liao, K. Yao, 'Theoretical and Experimental Study of High-performance Electrophoretic Mobilization of Isoelectrically Focused Protein Zones', *J. Chromatogr.*, **387**, 127–138 (1987).
102. S. Hjérten, M. Zhu, 'Adaptation of the Equipment for High Performance Electrophoresis to Isoelectric Focusing', *J. Chromatogr.*, **347**, 265–270 (1985).
103. S. Chen, J.E. Wiktorowicz, 'Isoelectric Focusing by Free Solution Capillary Electrophoresis', *Anal. Biochem.*, **206**, 84–90 (1992).
104. M. Zhu, R. Rodriguez, T. Wehr, C. Siebert, 'Capillary Electrophoresis of Hemoglobins and Globin Charins', *J. Chromatogr.*, **608**, 225–237 (1992).
105. G.G. Yowell, S.D. Fazio, R.V. Vivilecchia, 'Analysis of a Recombinant Granulocyte-macrophage Colony-stimulating Factor Dosage Form by Capillary Electrophoresis, Capillary Isoelectric-focusing and High Performance Liquid Chromatography', *J. Chromatogr. A*, **652**, 215–224 (1993).
106. J.M. Hempe, R.D. Carver, 'Quantitation of Hemoglobin Variants by Capillary Isoelectric Focusing', *Clin. Chem.*, **40**, 2288–2295 (1994).
107. M. Zhu, T. Wehr, V. Levi, R. Rodriguez, K. Shiffer, C. Zhu, 'Capillary Electrophoresis of Abnormal Hemoglobins Associated with α-Thalassemias', *J. Chromatogr. A*, **652**, 119–129 (1993).
108. J.M. Hempe, J.N. Granger, R.D. Craver, 'Capillary Isoelectric Focusing of Hemoglobin Variants in the Pediatric Clinical Laboratory', *Electrophoresis*, **18**, 1785–1795 (1997).
109. O.S. Reif, R. Freitag, 'Control of the Cultivation Process of Antithrombin III and its Characterization by Capillary Electrophoresis', *J. Chromatogr.*, **680**, 383–394 (1994).
110. G. Hunt, K.G. Moorhouse, A.B. Chen, B.L. Karger, 'Capillary Isoelectric Focusing and Sodium Dodecyl Sulfate Capillary Gel Electrophoresis of Recombinant Humanized Monoclonal Antibody HER2', *J. Chromatogr. A*, **744**, 295–301 (1996).
111. F. Kilár, A. Végvári, A. Mód, 'New Set-up for Capillary Isoelectric Focusing in Uncoated Capillaries', *J. Chromatogr. A*, **813**, 349–360 (1998).
112. K.W. Yim, 'Fractionation of the Human Recombinant Tissue Plasminogen Activator (rtPA) Glycoforms by High-performance Capillary Zone Electrophoresis and Capillary Isoelectric Focusing', *J. Chromatogr.*, **559**, 401–410 (1991).
113. J.M. Thorne, W.K. Goetzinger, A.B. Chen, B.L. Karger, 'Examination of Capillary Zone Electrophoresis, Capillary Isoelectric Focusing and Sodium Dodecyl Sulfate Capillary Electrophoresis for the Analysis of Recombinant Tissue Plasma Activator', *J. Chromatogr. A*, **744**, 155–165 (1996).
114. K.G. Moorhouse, C.A. Eusebio, G. Hunt, A.B. Chen, 'Rapid One-step Capillary Isoelectric Focusing Method to Monitor Charged Glycoforms of Recombinant Tissue-type Plasminogen Activator', *J. Chromatogr. A*, **717**, 61–169 (1995).
115. J. Kubach, R. Grimm, 'Non-native Capillary Isoelectric Focusing for the Analysis of the Microheterogeneity of Glycoproteins', *J. Chromatogr. A*, **737**, 281–289 (1996).

116. F. Kilár, S. Hjérten, 'Fast and High Resolution Analysis of Human Serum Transferrin by High Performance Isoelectric Focusing in Capillaries', *Electrophoresis*, **10**, 23–29 (1989).
117. J. Wu, J. Pawliszyn, 'In Vitro Observation of Interactions of Iron and Transferrin by Capillary Isoelectric Focusing with a Concentration Gradient Imaging Detection System', *J. Chromatogr. A*, **652**, 295–299 (1993).
118. H.H. Lauer, D. McManigill, 'Capillary Zone Electrophoresis of Proteins in Untreated Fused Silica Tubing', *Anal. Chem.*, **58**, 166–170 (1986).
119. F.A. Robey, 'Use of Capillary Electrophoresis for Binding Studies', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, 591–609, 1997.
120. I.J. Colton, J.D. Carbeck, J. Rao, G.M. Whitesides, 'Affinity Capillary Electrophoresis: A Physical-organic Tool for Studying Interactions in Biomolecular Recognition', *Electrophoresis*, **19**, 367–382 (1998).
121. N.H.H. Heegard, S. Nilsson, N.A. Guzman, 'Affinity Capillary Electrophoresis: Important Application Areas and Some Recent Developments', *J. Chromatogr. B*, **175**, 29–54 (1998).
122. N.H.H. Heegard, F.A. Robey, 'A Capillary Electrophoresis-based Assay for the Binding of Ca^{+2} and Phosphorylcholine to Human C-Reactive Proteins', *J. Immunol. Methods*, **166**, 103–110 (1993).
123. H. Kajiwara, H. Hirano, K. Oono, 'Binding Shift Assay of Parvalbumin, Calmodulin, and Carbonic Anhydrase by High-performance Capillary Electrophoresis', *J. Biochem. Biophy. Methods*, **22**, 263–268 (1991).
124. P.D. Grossman, J.C. Colburn, H.H. Lauer, R.G. Neilsen, R.M. Riggin, G.S. Sittampalam, E.C. Rickard, 'Application of Free-solution Capillary Electrophoresis to the Analytical Scale Separation of Proteins and Peptides', *Anal. Chem.*, **61**, 1186–1194 (1989).
125. R.G. Nielsen, E.C. Rickard, P.F. Santa, D.A. Sharknas, G.S. Sittampalam, 'Separation of Antibody–Antigen Complexes by Capillary Zone Electrophoresis, Isoelectric Focusing and High-performance Size-exclusion Chromatography', *J. Chromatogr.*, **539**, 177–185 (1991).
126. M.J. Schmerr, A. Jenny, 'A Diagnostic Test for Scrapie-infected Sheep Using a Capillary Electrophoresis Immunoassay with Fluorescent-labeled Peptides', *Electrophoresis*, **19**, 409–414 (1998).
127. X.-H. Qian, K. Tomer, 'Affinity Capillary Electrophoresis Investigation of an Epitope on Human Immunodeficiency Virus Recognized by a Monoclonal Antibody', *Electrophoresis*, **19**, 415–419 (1998).
128. J.M. Lange, D.A. Paul, H.G. Huisman, F. De Wolf, H. Van Der Berg, R.A. Coutinho, S.A. Danner, J. Van Der Noodaa, J. Goudsmit, 'Persistent HIV Antigenaemia and Decline of HIV Core Antibodies Associated with Transition to AIDS', *Br. Med. J.*, **293**, 1459–1462 (1986).
129. J.N. Weber, P.R. Clapham, R.A. Weiss, D. Parker, C. Roberts, J. Duncan, I. Weller, C. Carne, R.S. Tedder, A.J. Pinching, R. Cheingsong-Popov, 'Human Immunodeficiency Virus Infection in Two Cohorts of Homosexual Men: Neutralizing Sera and Association of Anti-gag Antibody with Prognosis', *Lancet*, **1**, 119–122 (1987).
130. M. Niedrig, J. Hinkula, W. Weigelt, J. L'Age-Stehr, G. Pauli, J. Rosen, B. Wahren, 'Epitope Mapping of Monoclonal Antibodies Against Human Immunodeficiency Virus Type 1 Structural Proteins by Using Peptides', *J. Virol.*, **63**, 3525–3528 (1989).
131. S. Graham, E.A. Follett, L. Wallace, U. Desselberger, H.S. Marsden, 'Immunodominant Epitopes of HIV-1 p17 and p24', *AIDS Res. Hum. Retroviruses*, **8**, 1781–1788 (1992).
132. L.R. Haaheim, J.P. Maskell, R. Mascagni, A.R. Coates, 'Fine Molecular Specificity of Linear Assembled Antibody Binding Sites in HIV-1 p24', *Scand. J. Immunol.*, **34**, 341–350 (1991).
133. J.P. Langedijk, J.J. Schalken, M. Tersmette, J.G. Huisman, R.H. Meloen, 'Location of Epitopes on the Major Core Protein p24 of Human Immunodeficiency Virus', *J. Gen. Virol.*, **71**, 2609–2614 (1990).
134. M. Niedrig, J.P. Rabanus, S.J. L'Age, H.R. Gelderblom, G. Pauli, 'Monoclonal Antibodies Directed Against Human Immunodeficiency Virus (HIV) Gag Proteins with Specificity for Conserved Epitopes in HIV-1, HIV-2, and Simian Immunodeficiency Virus', *J. Gen. Virol.*, **69**, 2109–2114 (1988).
135. D.H. Schwartz, A. Mazumdar, S. Winston, S. Harkonen, 'Utility of Various Commercially Available Human Immunodeficiency Virus (HIV) Antibody Diagnostic Kits for use in Conjunction with Efficacy Trials of HIV-1 Vaccines', *Clin. Diagn. Lab. Immunol.*, **2**, 268–271 (1995).
136. K.H. Mills, B.P. Kitchin, A.L. Mahon, S.E. Barnard, S.M. Adams, S.M. Kingsman, A.J. Kingsman, 'HIV p24-Specific Helper T Cell Clones from Immunized Primates Recognize Highly Conserved Regions of HIV-1', *J. Immunol.*, **144**, 1677–1683 (1990).
137. A.R. Coates, J. Cookson, G.J. Barton, M.J. Zvelebil, M.J.E. Sternberg, 'AIDS Vaccine Predictions (Letter)', *Nature*, **326**, 549–550 (1987).
138. C.E. Parker, D.I. Papac, S.K. Trojak, K.B. Tomer, 'Epitope Mapping by Mass Spectrometry: Determination of an Epitope on HIV-1 IIIB p26 Recognized by a Monoclonal Antibody', *J. Immunol.*, **157**, 198–206 (1996).
139. J.C. Severs, R.D. Smith, 'Capillary Electrophoresis–Mass Spectrometry', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, 791–826, 1997.
140. J.P. Larmann, J.W. Jorgenson, 'Two-dimensional Liquid Chromatography CE Techniques for Analysis of Proteins and Peptides', *Methods Enzymol.*, **270**, 401–419 (1996).

Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Andy J. Tomlinson
Pangaea Pharmaceuticals Inc., ZYCOS Inc., (formerly) Cambridge, USA

Stephen Naylor
Mayo Clinic, Rochester, USA

1 Introduction 1
2 History 2
3 Considerations for Capillary Electrophoresis Separations of Peptides and Proteins 3
3.1 Sample Injection 3
3.2 Concentration Sensitivity 3
3.3 Capillary Surface Technology 8
3.4 Capillary Electrophoresis/Mass Spectrometry Compatible Buffers 9
4 Capillary Electrophoresis/Mass Spectrometry Interfaces for Peptide and Protein Analysis 10
4.1 Off-line Coupling of Capillary Electrophoresis with Mass Spectrometry 10
4.2 Sheath Liquid Interface 10
4.3 Liquid Junction Interface 12
4.4 Sheathless Interfaces 12
4.5 Sample Preparation 14
5 Method Development 14
6 Sequencing Peptides by Capillary Electrophoresis/Tandem Mass Spectrometry 15
7 Future Directions 17
Acknowledgments 18
Abbreviations and Acronyms 18
Related Articles 19
References 19

The focus of this article is to describe the developments of capillary electrophoresis (CE) on-line with detection by mass spectrometry (MS) for the analysis of peptides and proteins. Following a brief introduction in which we discuss the increasing role of the mass spectrometer for characterizing these important biomolecules, we present a historical overview of capillary electrophoresis/mass spectrometry (CE/MS). The latter discussion describes the experiments that allowed the connection of CE with MS, through developments to increase the volumes of analyte solutions that can be analyzed, to the current investigations of microfabricated chips that bring CE/MS towards a nanoscale technology. We also consider the technical aspects of CE/MS, which include how to inject analyte mixtures into the capillary, ways to improve the concentration sensitivity of the technique, the optimal capillary surface technology to use, and considerations of buffer compositions that are compatible with on-line MS detection. Other practical issues, including the physical connection between the CE and MS components, along with sample preparation and sequencing peptides by this technology are also reviewed. In the final section, the future of CE/MS is contemplated, with particular emphasis on the current developments of nanotechnologies that include the interfacing of microchip-based CE with MS.

1 INTRODUCTION

The use of MS in the biomedical sciences has experienced rapid growth in recent years. This can be attributed to the development of technology [such as electrospray ionization (ESI)] that allows the rapid and cost-effective analysis of many thermally labile, hydrophilic molecules. Protein analysis, peptide mapping, and peptide sequencing at concentrations compatible with extracted amounts from biological systems by MS is a current area of continuous improvement.[1] This has accelerated developments in proteomic research, in which proteinaceous products of expressed genes are rapidly mapped.

Many MS strategies are evolving for sequencing of peptides and proteins. These include the use of micro- and nanospray technologies that allow low-flow infusion of analyte mixtures, and a more efficient utilization of precious samples.[2,3] This overcomes one of the main limitations of the scanning mass spectrometer, which is the acquisition of acceptable ion statistics for compounds of low transience in the instrument. High scan speeds usually correlate with reduced MS sensitivity. Prolonged persistence of molecules of interest within the mass spectrometer permits the use of slower scan speeds and a concomitant increase in sensitivity is observed. However, these techniques are not without limitations. In particular, ion suppression results in the detection of one species in preference to another. An inability to detect and characterize an analyte of interest (e.g. a phosphorylated peptide) may result when such molecule is a minor component of a complex mixture. This situation can often be overcome by separation of the components of the analyte mixture using chromatography on-line

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

with the mass spectrometer. In this regard, the coupling of condensed phase chromatography [high-performance liquid chromatography (HPLC) and CE] to the mass spectrometer has been thoroughly investigated and developments have occurred to provide technologies for the analysis of many widely different complex samples.[4,5]

In this article we present and review the techniques of on-line CE/MS for the analysis of peptides and proteins. Particular emphasis will be placed upon the developments that improve the robustness and sensitivity of this technology. We also review strategies that facilitate on-line manipulation of samples to effect isolation, concentration, and analysis of components of interest.

2 HISTORY

Characterized by fast analysis times, unrivalled analyte resolution and ultrahigh separation efficiencies, CE was expected to revolutionize the analysis of complex solutions of peptides and proteins.[6] Indeed, these characteristics along with low sample consumption and in theory improved analyte recovery (and thus high sensitivity) combine to yield a very powerful analytical technology, which was further enhanced through its coupling to the mass spectrometer.[7–10] Furthermore, CE offers a number of different separation mechanisms that complement traditional technologies such as HPLC. In addition to free solution separations, CE affords analyte resolution by isoelectric focusing,[11–15] isotachophoresis,[16] molecular weight sieving,[17] and even separation according to analyte hydrophobicity by interactions with micelles [as in micellar electrokinetic chromatography (MEKC)].[18] Affinity capillary electrophoresis (ACE)[19] and affinity capillary electrophoresis/mass spectrometry (ACE/MS)[20–24] separations have also been developed to investigate ligand–receptor interactions, and the potential for such technology for screening combinatorial drug candidate libraries has been demonstrated. However, the development of CE and on-line CE/MS for the analysis of peptides and proteins has not been without complex challenges, and these technologies have been shown to suffer some severe limitations. First, highest analyte resolution is usually only achieved when the sample injection volume is less than 2% of the total capillary volume.[25,26] Hence sample injection volumes are usually in the range of tens of nanoliters for regularly used capillaries [50 μm internal diameter (ID)][27] to low picoliters for small internal diameter capillaries (<10 μm ID)[28] that have been shown to yield the best CE/MS sensitivity. The concentration sensitivity of CE (and CE/MS) is therefore significantly inferior to that of other techniques such as HPLC, which can accommodate sample injection volumes of tens to hundreds of microliters.[4] Sample concentrations for CE and CE/MS analysis are thus often required to be exceptionally high. Typically, this is not ideal for mixtures of biologically important peptides and proteins. At high concentration, these molecules tend to denature, aggregate, and precipitate. Peptides and proteins also tend to adhere to surfaces of vials and pipet tips, which can lead to further substantial losses. Therefore, various creative approaches have been developed to enable the injection of larger sample volumes into the CE capillary. These include in-capillary electrophoretic analyte stacking[29–34] and focusing,[35–46] coupled capillary isotachophoresis (cITP) with CE,[47–49] and various designs of analyte concentrator[25,26,50–82] and membrane preconcentration (mPC) preprocessors,[83–96] and all of these approaches will be discussed later.

A second significant problem with CE (and CE/MS) arises from the fact that the silica surface of CE capillaries is not chemically inert. When filled with an aqueous solution, the silica surface takes on a negative charge that is dependent upon the pH of this solution.[97] This is one of the primary contributors to the development of an electroosmotic flow (EOF) from anode to cathode. However, such anionic character also provides an active surface for, in particular, basic peptides and proteins to adsorb. In addition to these electrostatic interactions, hydrophobic interactions and the formation of hydrogen bonds between the peptide or protein and the silica capillary are also prevalent. Hence, while there are no solid particles for analytes to adsorb irreversibly, peptides and proteins are often lost to the silica surface of the CE capillary.[98–100] Consequently, method sensitivity is reduced, as are analyte resolution and separation efficiency. Again, creative solutions have been developed to overcome this limitation. These include the use of buffers of extremes of pH, to suppress the ionization of the silica surface (low pH),[101] which has been successfully used for the analysis of peptides,[102] or to ensure that analytes are anionic (high pH).[103] Various capillary coatings have also been developed and several polymeric capillaries have been evaluated.[104] All of these approaches will be reviewed later.

The coupling of CE on-line with MS has also not been without challenge. Here, the technology had to be considered as a whole, since the predominantly aqueous buffer used for the CE separation has been shown to influence directly the MS performance and overall method sensitivity.[105,106] In addition, the interface between the CE and MS systems requires careful thought since this device has to provide both the counter electrode for the CE separation and the voltage (or ground) for the electrospray interface. While there have been a number of CE/MS interface configurations, one of the most popular is the sheath liquid interface that was first described by Smith et al.[7,8] This interface is readily assembled,

and is reasonably robust.[107] One limitation of this approach is that the sheath liquid can introduce a sizable chemical background. Miniaturization of the sheath liquid interface, so that the coaxially delivered sheath liquid is delivered at a flow rate of less than $1\,\mu L\,min^{-1}$, reduces such a background, and leads to a significantly improved signal-to-noise ratio (S/N).[108] Subsequently, development of a sheathless interface and the coupling of CE to micro- and nanospray MS has further improved the analyte sensitivity and S/N.[82,96,109–113] The coupling of micro-fabricated CE chips to MS in which the interface is a ''pin-hole'' in the side of the CE chip has also recently been evaluated.[109,114,115] These devices show much promise for the further development of this technology for high sensitivity analysis of peptides and proteins, and they are reviewed in a later section.

3 CONSIDERATIONS FOR CAPILLARY ELECTROPHORESIS SEPARATIONS OF PEPTIDES AND PROTEINS

3.1 Sample Injection

Three modes of sample injection into the CE capillary are typically used for analyte analysis by CE with on-line MS detection. These include electrokinetic injection, during which high voltage is applied to the sample solution; analytes migrate into the capillary according to their electrophoretic mobility, and may be transported by the migration of solvated ions.[116] This mode of injection is dramatically affected by the salt concentration of the sample matrix. The amount of analyte injected into the capillary can vary from sample to sample unless much care is taken to ensure that each sample is isotonic. Furthermore, analytes of differing charge will migrate into the capillary at different rates. Hence, this technique introduces selectivity into the analysis. For these reasons, electrokinetic injection is not usually particularly useful for the analysis of biologically derived mixtures of peptides and proteins.

The two other sample injection modes for CE/MS are based upon hydrodynamic flow. For one of these methods, the inlet of the CE capillary is immersed in the sample vial and raised, for a specific period, to a defined height above its outlet.[116] The volume of sample injected is dependent upon the back-pressure of the capillary, the height differential between the capillary inlet and outlet, and the time of the injection. This technique is often most commonly used when a home-made CE system is connected to the mass spectrometer. For commercially available units, the CE capillary is inserted into a sealed sample vial, to which a nitrogen head pressure is applied for a specific length of time. Sample is forced from its vial into the CE capillary, in preparation for subsequent electrophoretic separation. Since both of these techniques are based upon hydrodynamic flow, all components of a complex mixture are injected into the CE capillary and there is less selectivity than observed with electrokinetic injections. However, salt from the matrix of biologically derived samples will be introduced along with analytes of interest into the capillary. This will significantly affect the performance characteristics of the separation, and this subject will be addressed in the sections that follow.

3.2 Concentration Sensitivity

Many factors affect the overall sensitivity of CE/MS. One such factor is the design of the interface and the actual coupling of the two techniques. There is also an incompatibility of the peak widths of CE responses and the scan speed of the mass spectrometer. As mentioned above, faster mass spectrometer scan speeds typically result in reduced sensitivity of the scanning mass spectrometer. This problem has been overcome using one of several methodologies. First, if the analyte is known, selected ion monitoring (SIM) can be used to continuously monitor the compound(s) of interest.[5] Second, the type of MS detector can also play a role. For example, an array detector for a sector mass spectrometer can yield 50–100-fold more signal than a standard electron (or photon) multiplier.[117] Reduced elution speed (RES), in which the CE voltage is substantially reduced prior to elution of the first analyte, has been developed by Smith's group to widen CE peaks, thereby allowing the use of slower scan speeds.[118] Separations of proteins by cITP have also been coupled with MS.[46] This separation technique produces zones of constant protein concentration that are eluted into the mass spectrometer. This compares with a peaked response for most other separation techniques. Improved ion statistics may result. However, the development of cITP separation conditions is not as simple as free solution conditions, since buffer zones of different mobility are required to effect analyte resolution by the cITP mechanism.[16,49] Ion trap mass spectrometry (ITMS) (quadrupole traps)[112] and Fourier transform mass spectrometry (FTMS) (ion cyclotron resonance)[119] instruments and fast scanning time-of-flight (TOF) systems[120] may also afford improved sensitivity of CE/MS analyses (see later discussions).

The composition of the separation buffer (or electrolyte) and the ID of the CE capillary also affect the overall sensitivity of the CE/MS method.[104,105] A highly conductive background electrolyte (BGE) may be used for efficient CE separations but often interferes with electrospray processes, which can lead to reduced mass spectrometer sensitivity and ion beam instabilities. In this regard, it has been shown that the smaller bore capillaries

(<10 μm ID) lead to the highest CE/MS sensitivity.[28,121] The lower flow of BGE from these capillaries enables a wider variety of aqueous solutions to be used to effect analyte separations.[28] However, the use of small-bore capillaries also exaggerates the problem of poor CE and CE/MS concentration sensitivity. Such capillaries accommodate CE analyses of only very low volumes (often low picoliters) of sample.[25–27]

The low internal volume of CE capillaries leads to a requirement for high analyte concentration for all CE and CE/MS studies. Analyte preconcentration can be achieved by off-line sample preparation, using lyophilization, or adsorption on a solid phase to permit both sample cleanup and analyte concentration. However, off-line manipulations of dilute solutions of peptides and proteins should be minimized, since losses to exposed surfaces (e.g. vials, pipet tips, solid phases) can be substantial. In addition the tendency for these molecules to denature, aggregate, and, ultimately, precipitate causes further analyte losses. Therefore, many investigators have attempted to improve the sample loading capacity of the CE capillary while maintaining optimal analyte resolution and separation efficiency.

Initially, a variety of electrophoretic concentration techniques, including analyte stacking,[29–34] field amplification[35–45] and transient isotachophoresis (tITP)[46–49] were developed to preconcentrate analytes following injection of the sample into the capillary. All of these techniques are implemented as voltage is applied across the CE capillary, and cause analyte zones to concentrate owing to different field strengths or chemical microenvironments that form within the CE capillary.[116] Consequently, larger analyte volumes can be analyzed with minimal loss of component resolution and separation efficiency. In most favorable cases, the sample can constitute up to approximately 90% of the total CE capillary volume with minimal loss of CE performance. However, since the total capillary volume is small, maximum sample loading is usually limited to less than 1 μL. Furthermore, as mentioned above, salts in the sample matrix can disrupt such analyte stacking and focusing processes. Such matrix components introduce zones of low electric field within the capillary upon application of high voltage.[27,116] Analyte ion migration velocities are slower, and this usually prevents the analytes focusing into discrete zones. Therefore, preprocessing of biologically derived mixtures may be required to reduce their salt concentration prior to use of these electrophoretic stacking and focusing techniques. Nevertheless, use of in particular capillary electrophoresis/transient isotachophoresis/mass spectrometry (CE/tITP/MS) has provided a means for analyte preconcentration to a sufficient level to allow its use for the analysis of peptide and protein mixtures.[46]

A technique that emerged from single-column tITP was the use of coupled capillary isotachophoresis/capillary electrophoresis/mass spectrometry (cITP/CE/MS).[47–49] For this approach, cITP separations are performed in wide-bore (100–200 μm ID), larger volume capillaries. Following the cITP step, analytes are transferred into the CE capillary for further separation and subsequent MS detection. This technique permits the analysis of sample volumes as large as 20 μL, and as such it provides a practical means of enhancing the concentration sensitivity of CE/MS. However, this technique is significantly more complex than other strategies for concentration sensitivity enhancement, and is not particularly well suited to routine operation.

In another approach, capillary isoelectric focusing/mass spectrometry (cIEF/MS) allows the analysis of a full capillary volume of sample.[11–15,108,122,123] This technique has proved to be particularly useful for the analysis of complex mixtures of proteins. Separations are usually performed in a coated capillary that exhibits negligible or zero EOF.[15,108,122,123] In a focusing step, ampholites, which are premixed with the sample, migrate through the capillary to form a pH gradient. Simultaneously, proteins migrate until they reach the pH zone that corresponds to their isoelectric point (p*I*), where they become neutral, and no longer migrate. Here, the proteins focus into discrete zones, and are preconcentrated. Ultimately, a stationary state is achieved when all proteins have migrated into the pH zones that correspond to their p*I*.[11] Mobilization of the proteins into the mass spectrometer is achieved using an electrophoretic approach, hydrodynamic flow, or a combination of these techniques.[11–15,108,122,123] While cIEF/MS has been shown to be useful for the analysis of proteins, the processes of pH gradient formation and analyte focusing are severely perturbed by salt. High salt concentrations lead to excessive currents during the focusing phase of the experiment. This causes ampholyte zone broadening and a less well-defined pH gradient. Recently, we have found that use of a 5–10 min gradient application of voltage facilitates on-line removal of salt.[15] The current profiles observed in these experiments led us to conclude that the relatively slow application of the focusing voltage allows the ejection of highly mobile salt, as the pH gradient is developed and proteins migrate to zones that correspond to their p*I*. This is shown in Figure 1(a) and (b). Here, two protein solutions were analyzed using the same zero EOF, PVA-coated capillary. First, a solution containing lysozyme, cytochrome *c*, myoglobin, carbonic anhydrase, and trypsin inhibitor in water was analyzed using a standard cIEF approach, and the results of this analysis are shown in Figure 1(a). Subsequently, the same proteins were dissolved in water that contained 200 mM sodium chloride. Analysis of the latter solution by a

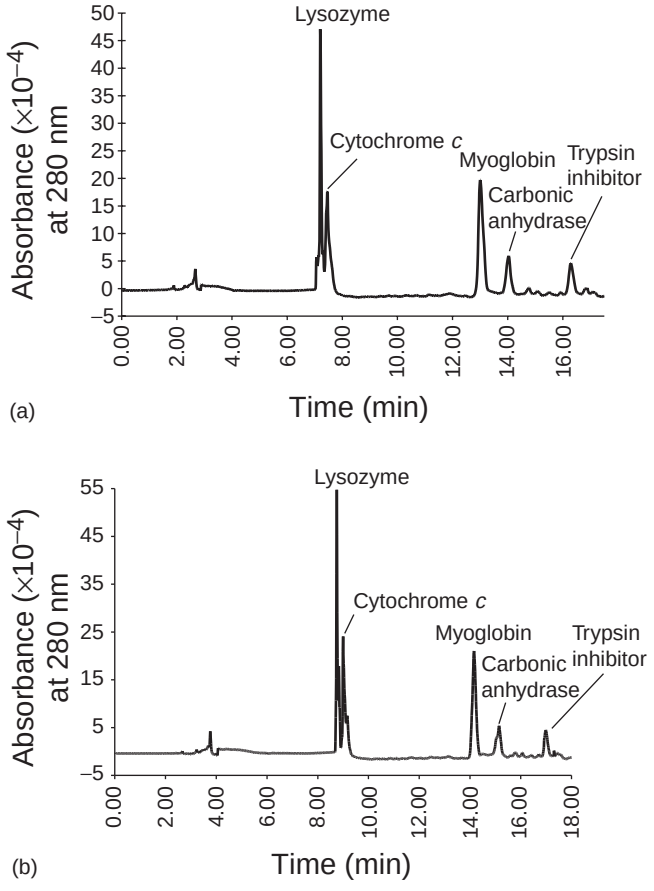

Figure 1 Capillary isoelectric focusing (cIEF) electropherograms of a five-component protein mixture (a) in the absence of salt and (b) from a solution that was prepared in 200 mM NaCl. Both samples were diluted 1 : 1 with ampholyte solution. Focusing and mobilization solutions were anolyte 50 : 49 : 1 (v/v/v) methanol–water–acetic acid and catholyte 50 : 49 : 1 (v/v/v) methanol–water–ammonia solution. The desalting protocol used for (b) was a linear gradient to 10 kV in 6 min. Protein focusing was at 20 kV for ~20 min. Mobilization was at 20 kV until the experiment was complete and all of the proteins were detected. Protein detection was by ultraviolet (UV) absorption at 280 nm. The CE capillary was 50 μm ID × 37 cm (effective length was 30 cm to the UV detector), and coated in-house with poly(vinyl alcohol) (PVA). (Reprinted with permission from N.J. Clarke, A.J. Tomlinson, G. Schomberg, S. Naylor, *Anal. Chem.*, **69**, 2786–2792 (1997). Copyright 1997, American Chemical Society.)

standard cIEF approach failed. As a voltage of 20 kV was applied to the capillary (over 30 s) to affect analyte focusing, the current rose to ~45 μA and became unstable. It was impossible to obtain satisfactory results from this system without first removing the sodium chloride by means of a slow voltage ramp, and this is demonstrated in Figure 1(b). Clearly, this technique worked well, and was, in our hands, reproducible. As expected, a small shrinkage of the pH gradient was observed. This occurred during the desalting process when cations and anions are mobilized from the capillary and are replaced by counterions of the BGE. While a UV detector was used to acquire the data shown in Figure 1(a) and (b), we have subsequently demonstrated the efficacy of this approach for analysis of physiologically derived protein mixtures by cIEF/MS. It the latter study we demonstrated the detection of glycated hemoglobin chains in diabetic blood and the direct analysis of cerebral spinal fluid by cIEF/MS.[124,125]

Analyte concentrators and mPC devices have also been developed to overcome the relatively poor concentration limits of detection of CE and CE/MS. This technology has been the subject of three extensive reviews.[27,126,127] Briefly, such technology is based upon the insertion of a small bed of adsorptive phase, multiple capillary bundle, immunoaffinity capture support or impregnated membrane at the inlet of the CE capillary (see Figure 2). Principally analytes are adsorbed on a phase to effect their concentration. Injection volumes in excess of 200 μL are routine with this technology, which corresponds to an increase in concentration sensitivity often upwards of 100–1000 times that of a conventional CE injection method. In addition, since analytes are adsorbed on a solid phase or impregnated membrane, on-line sample cleanup can be effected by washing the phase with a suitable solvent.[87,88] For biologically derived peptide and protein samples, this process permits the removal of salt and other hydrophilic contaminants that interfere with the electrophoretic separation of these analytes. This eliminates the need for excessive off-line pretreatment, which improves analyte recovery, which ultimately further enhances concentration sensitivity. Some of these techniques (e.g. mPC) have also been shown to be compatible with contemporary CE stacking and focusing chemistries. Indeed, membrane preconcentration/transient isotachophoresis/capillary electrophoresis/mass spectrometry (mPC/tITP/CE/MS) has been shown to be an optimal technique for the separation and sequencing of low concentrations of peptides.[86–88,96] A major consideration for optimal membrane preconcentration/capillary electrophoresis/mass spectrometry

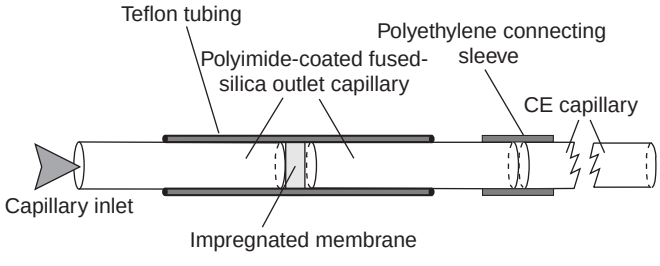

Figure 2 Schematic diagram of a mPC cartridge. Note the gaps in the cartridge; these are to prevent perturbation of the EOF cause by a high back-pressure in the cartridge if the membrane is compressed. Also, push-fit joints construct the cartridge.

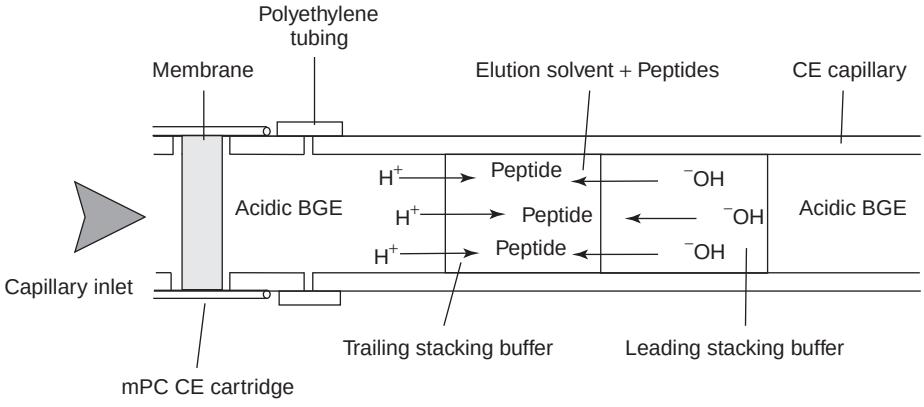

Figure 3 Schematic diagram of the tITP conditions that have been developed for peptide analysis by mPC/CE/MS. Typical conditions are 60 nL of each of the reagents.

(mPC/CE/MS) performance is the use of a sufficient volume of an elution solvent (e.g. 80% methanol or acetonitrile in water) to ensure maximized peptide recovery. However, the use of an excessive volume of elution solvent severely degrades the separation performance of the analysis. Use of a tITP focusing strategy in conjunction with mPC tends to overcome this problem, since analytes are focused into discrete zones as voltage is applied across the capillary.[86–88,96]

The tITP condition that we developed for use in conjunction with mPC/CE/MS is shown schematically in Figure 3. In a systematic study we have shown that mechanistically this stacking/focusing regime is a complex combination of transient cIEF, tITP and analyte stacking, that needs to be fully optimized for each membrane preconcentration/capillary electrophoresis (mPC/CE) capillary used. However, once optimized this system works extremely well. This is shown in Figure 4. Here, a single reversed-phase HPLC fraction that was derived from a biological mixture of rat major histocompatibility complex (MHC) class I peptides were analyzed by mPC/tITP/CE/MS. The power of the technique is demonstrated by the resolution of isobaric peptides (see Figure 4).

We have also used this technique to identify a mouse Kb peptide that promotes T-cell maturation in the thymus. More recently, Hunt's group also demonstrated the value of mPC/tITP/CE/MS for sequencing MHC class I peptides (see Figures 5 and 6). In this example, the mPC/CE capillary was coupled to an ITMS instrument via a novel microspray interface (see later discussions for a review of interface technology).[96] Here, peptide sequencing with attomole sensitivity was demonstrated by the analysis of an immunologically active MHC class I peptide fraction that was isolated from an extract of 10^{10} cells from the human melanoma tumor (DM6) cell line. Full sequence data was achieved

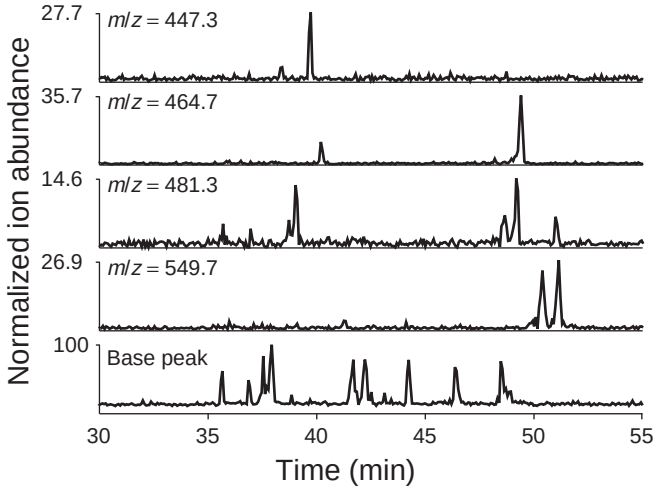

Figure 4 Partial mPC/CE/MS electropherograms of rat RT1^{a}–TAP^{u} MHC class I peptides. The mPC cartridge contained a polymeric styrene–divinylbenzene-impregnated membrane and the CE capillary was uncoated fused silica, 25 μm ID × 70 cm. Injection volume was 50 μL. BGE was 2 mM ammonium acetate in 1% aqueous acetic acid, leading stacking buffer was 1% ammonia solution in water (60 nL), elution solvent was 80% methanol in water (60 nL), trailing stacking buffer was BGE (60 nL). Applied separation voltage was 25 kV. ESI voltage was 2.5 kV referenced to the accelerating voltage (5 kV) of a Finnigan MAT 95Q mass spectrometer. A sheath liquid interface of 60% 2-propanol in 1% aqueous acetic acid was used at 3 μL min^{-1}. Peptide detection was by an electron multiplier, instrument resolution was ~300. Scan range was 300–1300 u at 3 s per decade.

for a known active peptide by membrane preconcentration/transient isotachophoresis/capillary electrophoresis/tandem mass spectrometry (mPC/tITP/CE/MS/MS) (see Figure 6).

While peptide analysis by mPC/tITP/CE/MS can be achieved using an uncoated fused-silica capillary and an acidic BGE, it is our experience that peptides

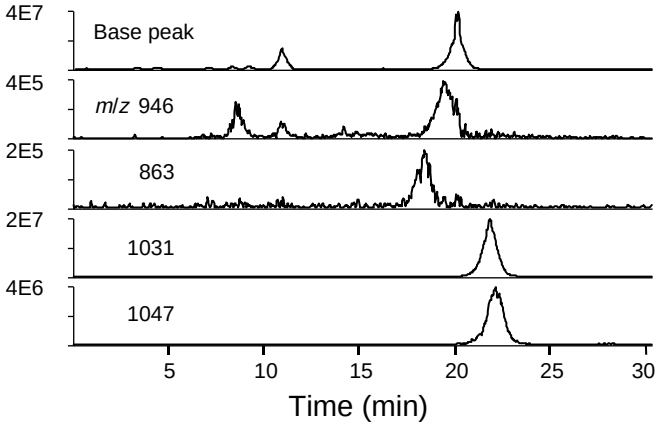

Figure 5 mPC/CE/MS analysis of human MHC class I peptides in HPLC fraction 17 from a melanoma DM6 cell line. (Reproduced from R.E. Settlage, P.S. Russo, J. Shabanowitcz, D.F. Hunt, *J. Microcol. Sep.*, **10**, 281–285 (1998). Copyright 1998 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

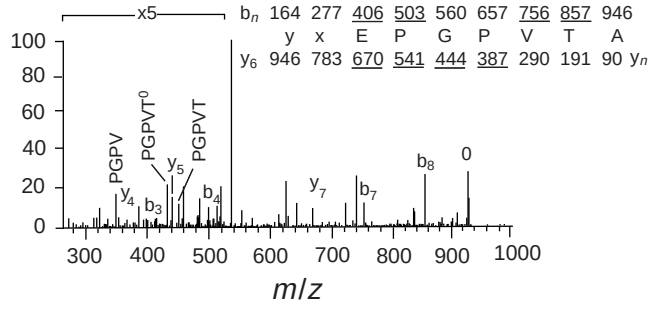

Figure 6 Membrane preconcentration/capillary electrophoresis/tandem mass spectrometry (mPC/CE/MS/MS) analysis of an active melanoma peptide of molecular weight 946. (Reproduced from R.E. Settlage, P.S. Russo, J. Shabanowitcz, D.F. Hunt, *J. Microcol. Sep.*, **10**, 281–285 (1998). Copyright 1998 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

and proteins are more optimally analyzed using a coated CE capillary.[128] In our studies of MHC class I peptides, tear proteins, and the composition of aqueous humor, we have found a polybrene-coated capillary is most suitable.[128–130] Again, the use of a tITP strategy in conjunction with mPC/CE/MS ensures that a larger volume of an elution solvent can be used with little detriment to the separation performance. Indeed, it has been our experience that unless a relatively large volume of elution solvent is used, proteins are not efficiently recovered from currently available impregnated membranes. In a systematic study, we found that the composition of the elution solvent was critical for elution of proteins from a C_2-impregnated membrane.[129] Increasing acetonitrile concentration from 40 to 80% in water significantly improved protein recovery from the membrane, and overcame a selectivity problem at the lower acetonitrile concentrations. Furthermore, we demonstrated that there was no requirement for the addition of an acid to this elution solvent. Indeed, addition of 0.1% trifluoroacetic acid (TFA) to the 80% acetonitrile elution solvent had only detrimental effects on protein separation.[129] We attributed this result to localized interaction of the TFA, an efficient ion-pair reagent, with the positively charged polybrene-coated capillary. In addition, we found that use of a less efficient ion-pair reagent (such as acetic acid) did not influence protein recovery. We attributed this result to the fact that gaps are deliberately introduced in mPC cartridges as they are prepared. This prevents perturbation of EOF by the cartridge. However, turbulent mixing of injected solvent will also occur. Thus while an elution solvent of 80% acetonitrile in water is injected into the mPC/CE capillary, proteins will be eluted from the membrane by the formation of a solvent gradient. As acetic acid is a constituent of a typical BGE for this application, the front edge of the elution solvent will be modified with this reagent. We postulate that this would tend to be beneficial for the recovery of analytes from the membrane.[129]

Subsequent to the systematic study, we have successfully applied mPC/CE and mPC/CE/MS to the analysis of several physiological fluids, including cerebrospinal fluid (CSF), tears, and aqueous humor [see Figure 7(a) and (b)].[130] In this example we demonstrate the use of mPC/CE/MS to allow a comparison of the composition of ''normal'' aqueous humor (taken from a patient undergoing cataract surgery) and that which was obtained from an individual with PEX syndrome. Several proteins have been tentatively identified from these studies, and we have found few qualitative differences in the composition of the syndrome sample and the control. One lead that we are currently pursuing is the absence of a protein of molecular weight 13 343, and its putative oxidized form (molecular weight 13 361) in the PEX sample, and these studies are on-going. During these studies, we found that protein recoveries were often poor (<25%) and irreproducible when we analyzed standard mixtures of these analytes. In contrast, the recovery of components of physiologically derived protein mixtures was much higher (>90%). We attributed this result to the presence of high salt concentrations in these later solutions that tended to modify the adsorptive properties of the C_2-impregnated membrane.

From our studies it is clear that mPC/CE/MS analysis of protein mixtures is still in its infancy, being limited by the current choice of impregnated membranes. This approach will undoubtedly improve in the future as new impregnated membranes of more appropriate

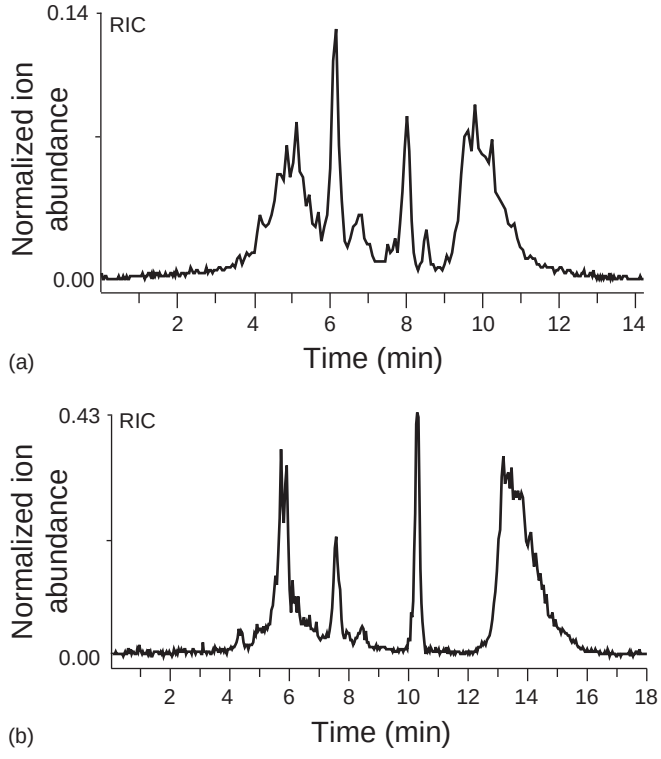

Figure 7 mPC/CE/MS analysis of human aqueous humor. (a) 1 μL of ''normal'' aqueous humor and (b) the same volume of aqueous humor taken from a patient with pseudoexfoliation (PEX) syndrome. The mPC cartridge contained a C_2-impregnated membrane, and the CE capillary was 50 μm ID by 70 cm polybrene-coated. BGE was 2 mM ammonium acetate in 5% aqueous acetic acid, the elution solvent was 80% acetonitrile in water (60 nL), and the trailing stacking buffer was 0.5% ammonia solution in water (60 nL). The applied separation voltage was 25 kV. ESI voltage was 2.5 kV referenced to the accelerating voltage (5 kV) of a Finnigan MAT 900 mass spectrometer. A sheath liquid interface of 60% 2-propanol in 1% aqueous acetic acid was used at 3 μL min^{-1}. Protein detection was by the PATRIC™ detector, instrument resolution was ~300. Scan range was 300–2500 u at 3 s per decade.

physicochemical properties (e.g. wide-pore polymeric phases such as PLRP-S 4000 Å) become available. In addition, further development of affinity capture analyte concentrators will prove to be very valuable for the isolation and analysis of target proteins.

3.3 Capillary Surface Technology

As mentioned previously, CE was initially thought to be a new approach that would revolutionize the analysis of peptides and proteins.[6] For some applications (e.g. the sequencing of MHC class I peptides) CE has been proved to be extremely valuable. However, capillary surface technology has proved to be a major challenge for the analysis of these important biopolymers. Indeed, manufacture of fused-silica capillaries that exhibit constant surface properties has been the subject of much industrial research by companies that supply this consumable to the users of CE. In our own hands, we have found that while a CE capillary that is prepared from a bulk supply may provide high performance, the next length of fused-silica may not condition at all. In addition, there is no way of predicting in advance which piece of fused-silica will work, and this can be both frustrating and time-consuming. For peptides and proteins, capillary preparation is further complicated by potential analyte–wall interactions. As noted above, the internal surface of a fused-silica capillary is ionized when filled with an aqueous BGE. While the degree of ionization is dependent upon the pH of the BGE, the ionized silanol groups of the CE capillary provide an active anionic surface to which peptides and proteins can adsorb.[98–100] This problem is particularly evident for proteins. These analytes are amphiphilic, such that their structure and chemical environment determines their characteristics (e.g. charge, polarity, and hydrophobicity).

The interactions of proteins at the liquid–solid boundary of fused-silica CE capillaries have been thoroughly investigated.[98–101] Forces that are invoked during these interactions include electrostatic, ion pairing, hydrophobic, and the formation of hydrogen bonds. In CE separations adsorption equilibria between proteins and the capillary wall cause severe zone broadening. Peaks also become asymmetric, which leads to decreased analyte resolution and separation efficiency. Furthermore, a changing capillary surface also affects the magnitude of the EOF, and analyte separations become irreproducible.

To negate analyte–wall interactions, major efforts have been directed toward shielding or inverting the negative charge on the capillary wall. Various capillary coatings have been described [for a review of this subject, see Schomburg[104]]. These include polymers, such as polyacrylamide (PAA), polyethylene glycol (PEG), polyethylenimine (PEI), PVA, aminopropyltrimethoxysilane (APS), and C_8 and C_{18} phases that contain siloxane-bound anchor groups for covalent attachment to the capillary surface. Alternatively, static coatings such as PEI and polybrene have been adsorbed on the capillary surface prior to the analysis of peptides or proteins. Similarly, dynamic coatings such as PVA can be added to the BGE in an attempt to eliminate the interaction of analytes with the capillary wall. There have also been some investigations of the use of polymeric capillaries.[131–133] However, these were difficult to make with a consistent small ID, and did not readily overcome the issue of protein interaction with the capillary wall, since even polymers can take on a charge during electrophoresis.

By design, all capillary coatings change the chemistry of the silica surface. Characteristics of the polymer or chemical coating determine the nature of the capillary surface. For example, hydrophilic neutral coatings (such as PVA, PEG, and PAA) impart no charge to the capillary wall and effectively eliminate EOF.[104] In these capillaries analyte resolution is based solely upon the differences of electrophoretic mobility of each analyte. Other coatings such as PEI, APS, polybrene, or surfactants such as cetyltrimethylammonium bromide reverse the charge on the capillary wall, and EOF flows from the cathode to the anode.[104] CE capillaries coated with such positively charged materials have been shown to be especially advantageous for the analysis of basic proteins. More recently, we have reported that a polybrene-coated capillary is very useful for the analysis of the protein composition of physiologically derived fluids (such as aqueous humor). Indeed, separations in a polybrene-coated capillary were found to be less affected by the salty matrix of this fluid than other tested capillary coatings. In fact, in capillaries that were coated with a neutral polymer, we found that the high salt concentration of these samples prevented the separation of the proteins that were present in physiologically derived fluids. Often only one peak was observed during separations under these conditions.

For CE/MS experiments, the use of a coated capillary further complicates this methodology. If the coating is not sufficiently anchored to the capillary wall, it can bleed into the mass spectrometer. Here it be may ionized, thereby generating a large chemical background or interfering with electrospray processes to impair stability or reduce analyte sensitivity. Hence, in addition to preventing analyte adsorption, the coating has to be stable and remain in the capillary. Most of those described above, with the exception of the dynamic coating strategy, have been successfully used for protein analysis by CE/MS. In general, coatings that invert the charge on the capillary wall have been preferred for the CE/MS analysis of proteins.[109,119,129,130,134–139] This can be attributed to the stability of such coating in acidic BGEs, which is a regime that also promotes the formation of positively charged analytes, thereby enabling the use of positive ion MS conditions.

3.4 Capillary Electrophoresis/Mass Spectrometry Compatible Buffers

As alluded to above, the BGE used in the CE capillary to separate analytes can be detrimental to electrospray MS performance.[105,106] Buffers of relatively high ionic strength are often used for CE separations, since these help to prevent analyte–analyte and analyte–wall interactions. Involatile salts are also a favorite choice for CE separations in conjunction with detection methods other than MS. By contrast, volatile buffers (e.g. ammonium salts) of low concentration, ionic strength, and conductivity are typically used for most CE/MS experiments. This is to ensure stable electrospray conditions and help prevent contamination of the mass spectrometer by salts and ultimately instrument break down. For peptides and proteins, Moseley et al. found that acidic buffers of low ionic strength provided the best CE/MS sensitivity when using a sheath liquid interface.[105] More recently, Wahl and Smith compared the effect of BGE composition on CE/MS using both sheath liquid and sheathless electrospray interfaces.[106] The results of this study were in good agreement with theory, and showed that CE/MS sensitivity was reduced with increasing buffer concentration and ionic strength. In addition, comparison of ammonium acetate–acetic acid and sodium phosphate buffer systems indicated that at a 1 mM concentration the latter buffer provided a seven-fold less signal-to-background ratio than the volatile acetate system. This was attributed to the difference in ionic strength and volatility of the examined BGEs. Other results of these studies demonstrated that the sheathless interface often provided better analyte detectability than a sheath liquid interface for most buffer systems. Likewise, a smaller bore capillary allowed improved tolerance of the buffer system, and this was attributed to the lower BGE flow rate. The BGE flow rate in a 10 μm ID CE capillary is approximately 0.8 nL min^{-1} for a 10 mM ammonium acetate–acetic acid BGE.[106]

In other studies, organic modifiers have been added to the BGE, to aid analyte solubility.[140] The use of organic modifiers also changes the physical properties of the separation solution, which can change the magnitude of the EOF.[141] The organic solvent can alter the thickness of the electrical double-layer at the capillary wall, and/or change the viscosity of the BGE.[140,141] As demonstrated in Figure 8, a nonaqueous BGE can also offer advantages for CE/MS analyses of hydrophobic peptides. Here, an ammonium acetate–formic acid BGE system dissolved in a mixture containing only acetonitrile and methanol (75 : 25 v/v) was used to separate two hydrophobic peptides, namely gramicidin S and bacitracin. Both of these analytes are only sparingly soluble in aqueous solution, and poor analyte sensitivity was demonstrated using a conventional aqueous CE/MS BGE (data not shown). However, a nonaqueous BGE permitted efficient analysis of these analytes by CE/MS, when using a sheath liquid interface (consisting of 5 mM ammonium acetate in 80 : 20 v/v 2-propanol–water). Indeed, we were able to detect a number of minor contaminants in the samples of both peptides by CE/MS using the nonaqueous BGE system (see Figure 8).[140]

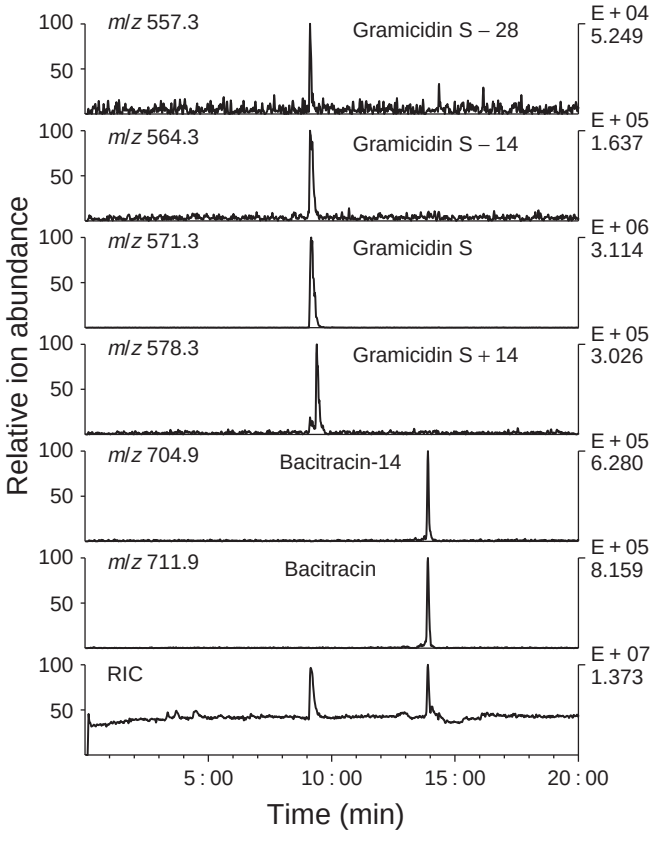

Figure 8 CE/MS analysis of a mixture of gramicidin S and bacitracin. CE capillary was 50 μm ID by 70 cm uncoated fused silica. BGE was 20 mM ammonium acetate, 0.5% formic acid in a mixture of 75% acetonitrile in methanol. Applied separation voltage was 25 kV. ESI voltage was 3.45 kV referenced to the accelerating voltage (5 kV) of a Finnigan MAT 900 mass spectrometer. A sheath liquid interface of 5 mM ammonium acetate and 80% 2-propanol in water was used at 2 μL min^{-1}. Peptide detection was by an electron multiplier, MS resolution was 1500 (10% valley definition). Scan range was 300–1500 u at 1 s per decade.

4 CAPILLARY ELECTROPHORESIS/MASS SPECTROMETRY INTERFACES FOR PEPTIDE AND PROTEIN ANALYSIS

4.1 Off-line Coupling of Capillary Electrophoresis with Mass Spectrometry

One of the first approaches for connection of CE to MS was via the collection of fractions from the CE capillary followed by independent analysis by MS. By this approach, probe MS techniques such as plasma desorption[142–144] and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOFMS)[145–151] techniques may be used to measure the molecular weight of isolated analytes. In addition, more recently, nanoelectrospray MS has been used to analyze peptide and protein fractions that were collected following analyte separation by CE.[152] In this area of research, one of the more attractive off-line couplings of CE with MS has been described by Zhang and Caprioli[149] They built a moving target system on which components eluting from the CE capillary were deposited. Specifically, the eluent of the CE capillary was collected onto a cellulose membrane that was fixed on the MALDI/TOFMS instrument target, and preloaded with matrix-assisted laser desorption/ionization (MALDI) matrix (e.g. α-cyano-4-hydroxycinnamic acid). The process of depositing small amounts of aqueous solution on the membrane caused the matrix to partially dissolve, which in turn stimulated cocrystallization of analytes and matrix in preparation for subsequent MALDI/TOFMS analysis of the separated analytes. The target was mounted on the mobile block of a syringe pump to allow a single sample track across the membrane, with a time window of ~3.3 min. Detection limits (DLs) (S/N > 3) were found to be ~50 amol for neurotensin and ~250 amol for cytochrome *c* and apomyoglobin. Such DLs are likely to be attributable to the fact that these analytes were not subjected to further manipulation following their collection from the CE capillary. Clearly, this ingenious method overcame some of the early challenges that were experienced as CE was coupled on-line with MS. However, other methods for the off-line coupling of CE with MS, that advocate manipulation of collected analyte solutions may be less favorable. As mentioned above, manipulation of dilute solutions of peptides and proteins can lead to significant losses. Therefore, in our opinion, many of the reported off-line strategies for coupling CE with MS are less attractive techniques for the analysis of peptides and proteins than on-line CE/MS.

The coupling of CE on-line with MS has been the subject of many reports and several reviews. Therefore, the following is a brief discussion of this area [see elsewhere[5,109,153–158] for more comprehensive review of this subject matter].

4.2 Sheath Liquid Interface

The sheath liquid interface has become one of the most preferred methods for connecting the CE capillary to the mass spectrometer. Briefly, the CE capillary is extended though the ion source, and is surrounded by a coaxially delivered sheath liquid.[7,8] This approach has been used for both electrospray and continuous-flow fast atom bombardment (CFFAB) interfaces.[155] Conceptually the sheath liquid interface is the same for both MS devices with the coaxially delivered solution providing sufficient solvent flow through the interface. The sheath liquid also ensures electrical connectivity across the CE capillary,

and in the ESI interface provides the electrical contact for the ESI voltage.

Some challenges have been reported for both ESI and CFFAB interfaces. For the CFFAB design, the ID of the CE capillary must be small (typically $<13\,\mu m$ ID). This interface is inserted into the high vacuum of the mass spectrometer, and vacuum-induced flow becomes a problem for wider bore capillaries, resulting in a dramatic loss of separation efficiency.[155] The ESI interface does not suffer this problem since ionization occurs at atmospheric pressure, and capillaries as large as $100\,\mu m$ ID have been successfully coupled to this interface. However, as mentioned above, use of such wide-bore capillaries establish further constraints on buffer concentration and ionic strength, which can impair both analyte resolution and CE separation efficiency.[106]

Karger et al. investigated ionic mobility in CE/MS using a sheath liquid interface.[159] Since ions migrate towards both anode and cathode, ionic components of the sheath liquid can migrate into the separation capillary. Both sharp and diffuse ionic boundaries were detected, and resulted in migration delays, inversion of migration order, and often loss of analyte resolution. This effect is shown in Figure 9(a) and (b). Here, while the BGE was the same in the two experiments, replacement of a sheath liquid containing 1% acetic acid in 50% methanol–water by 20 mM TFA in 50% methanol–water had a dramatic effect on the analysis of a mixture of proteins. The time of the separation was compressed and the protein resolution was significantly different. Hence the sheath liquid is a further parameter that requires close attention. In addition to avoiding the possible detrimental effects on analyte separation (described above), careful selection of the sheath liquid can also lead to the use of BGEs that contain involatile buffers. Ions present in the sheath liquid can penetrate the CE capillary to prevent such involatile salts from entering the interface. Furthermore, this process can effectively remove adducts (e.g. phosphates) from proteins prior to entering the mass spectrometer. Similarly, we have noted on occasion that proteins may not be detected at the mass spectrometer (even though they have been seen to migrate passed a UV window that is close to the inlet of the CE capillary) when using a mostly organic sheath liquid interface. In addition to the above explanation, we have postulated that the mostly organic sheath liquid causes the proteins to precipitate as they elute from the CE capillary. As a result, proteins are neither ionized nor detected.

One of the major disadvantages of the sheath liquid interface is that it can introduce a large chemical background into the mass spectrometer. This tends to compromise analyte sensitivity, which is a problem that has been addressed by the use of a sheathless

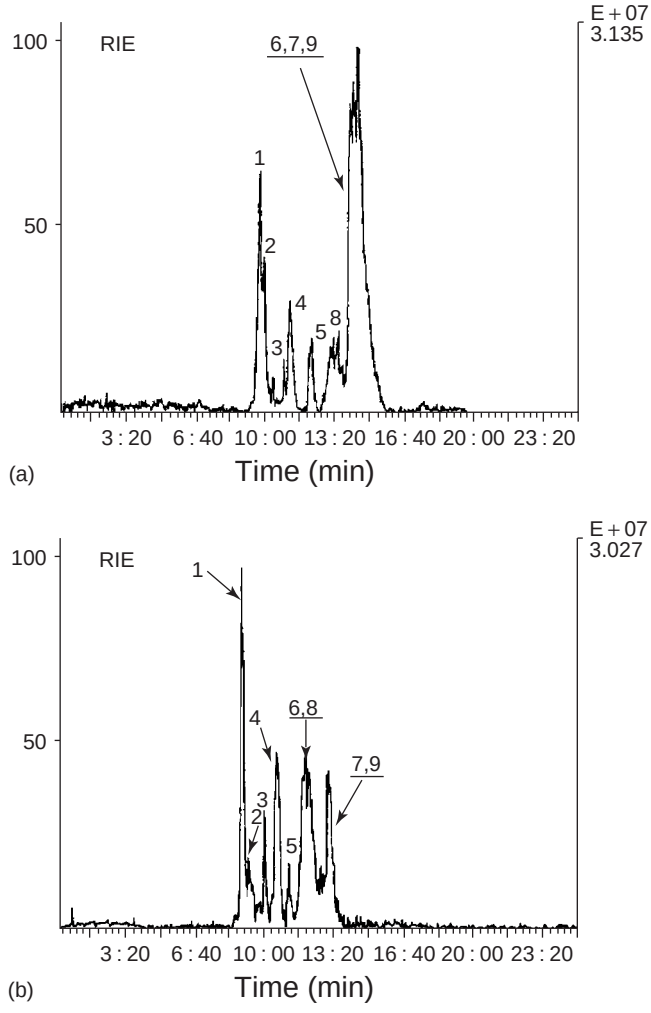

Figure 9 CE/MS analysis of a model mixture of proteins: 1, cytochrome *c*; 2, lysozyme; 3, aprotinin; 4, myoglobin; 5, RNase A; 6, β-lactoglobulin B; 7 and 8, β-lactoglobulin A and carbonic anhydrase; 9, α-chymotrypsinogen A. CE/MS conditions were BGE 20 mM β-alanine–formic acid at pH 3.4, CE capillary was $75\,\mu m$ ID by 60 cm. Separation voltage was 25 kV. Liquid sheath interface was (a) 1% acetic acid in 50% methanol–water and (b) 20 mM TFA in 50% methanol–water. (Reprinted with permission from F. Foret, T.J. Thompson, P. Vouros, B.L. Karger, *Anal. Chem.*, **66**, 4450–4458 (1994). Copyright 1994, American Chemical Society.)

interface,[106] or low sheath flow design. Kirby et al. advocated this latter approach.[108] Reduction of the annular volume between the sheath liquid needle and the CE capillary allows the use of lower sheath liquid flow rates ($<100\,nL\,min^{-1}$ has been reported). As a result, less chemical noise is detected and the S/Ns of analytes of interest are improved. A particularly useful application of this interface is its use in conjunction with CE capillaries that have negligible or zero EOF.[108] Such capillaries require the makeup flow of the sheath liquid to maintain stable spray conditions.

Most recent reports of the connection of CE with MS have focused upon the use of an ESI interface. This is due primarily to the fact that the outlet of the CE capillary is at atmosphere, hence the MS vacuum does not influence analyte separations. In addition, a major limitation of the CFFAB source is that the ions produced typically exhibit a small number of charges. For proteins, this requires an instrument with an extended mass range. By contrast, the ESI produces multiply charged gaseous ions from solution. Since all mass spectrometers measure mass-to-charge ratio (m/z) values, this effectively extends the mass range of the instrument and permits the analysis of large proteins using relatively inexpensive mass analyzers (such as standard mass range quadrupole instruments).

4.3 Liquid Junction Interface

The liquid junction interface has also been used to couple CE on-line with both ESI and CFFAB interfaces.[155] In this device the CE capillary is placed in a buffer reservoir. Analytes are transported to the mass spectrometer using a transfer line that is positioned in the buffer reservoir in close proximity (~10–25 μm) to the CE capillary. Alignment of these capillaries has been shown to be critical, and Henion et al. have described a modified liquid junction interface in which the CE and transfer capillaries are self-aligning.[156] One of the advantages of the liquid junction interface is its tolerance of involatile BGEs.[5] In addition, this interface was used to successfully couple gel-filled CE capillaries to the mass spectrometer.[160] However, the ease of set-up and use of the sheath liquid interface have ensured that this interface is usually the one of choice, especially by those new to CE/MS.[107]

4.4 Sheathless Interfaces

A sheathless ESI interface has several significant advantages over sheath and liquid junction designs.[109] First, there is a reduced chemical background. Also, since there is no makeup solution that can compete with analyte for the available charge, sensitivity is enhanced. For peptides and proteins, low femtomole sensitivity has been reported. In addition, more recently, peptide sequencing with attomole sensitivity has been reported for a combination of mPC/CE with a sheathless ESI interface, using an ion trap mass spectrometer (see Figure 6).[96]

Sheathless ESI interfaces typically conform to one of six major designs, shown in Figure 10(a–f). The

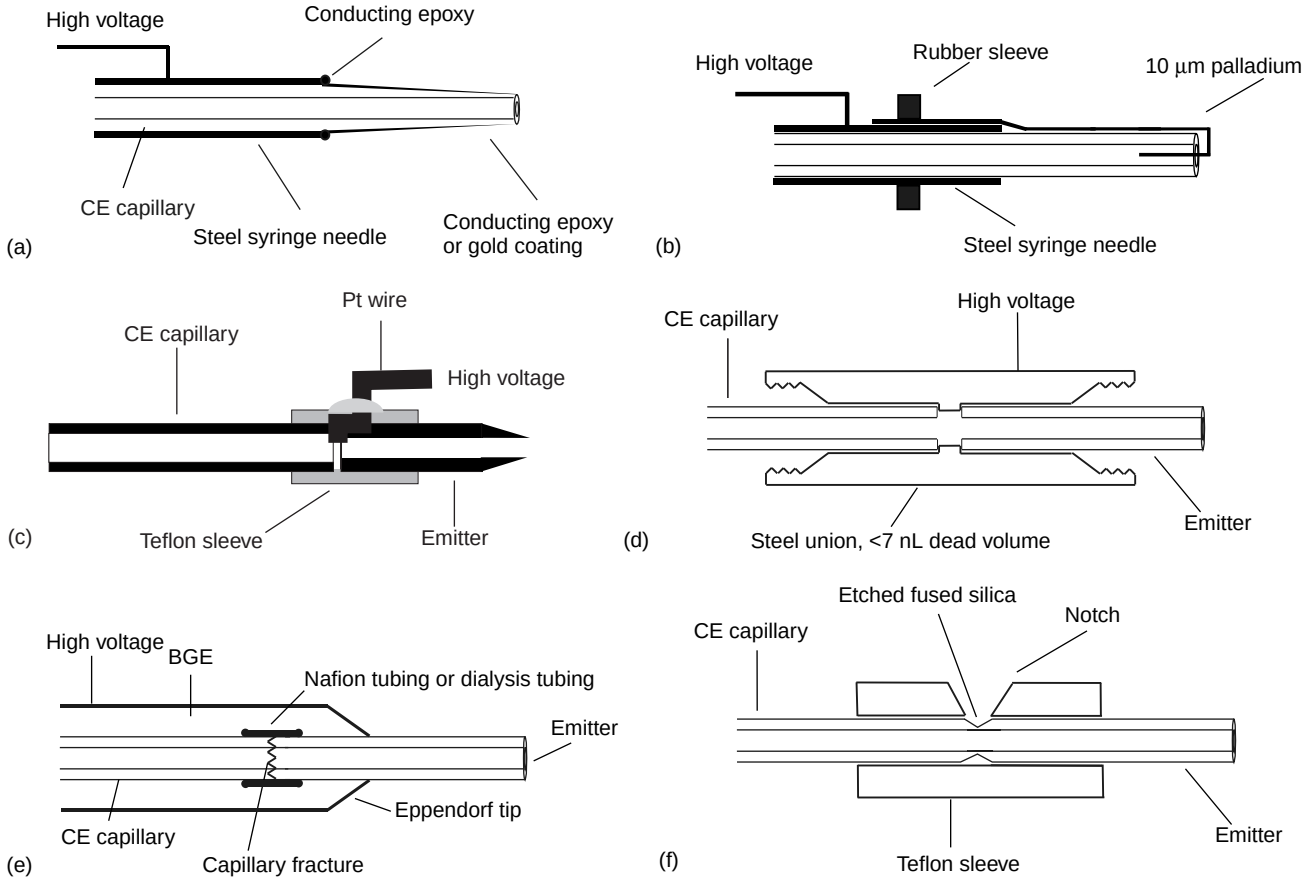

Figure 10 Schematic diagram of a series of microspray CE/MS interfaces.

simplest sheathless interface (Figure 10a) is usually made by etching and tapering the fused-silica, which is then coated with gold[161] or conducting epoxy.[109] Other designs include the insertion of a wire into the outlet of the CE capillary (Figure 10b)[82] or the use of a thin wire between the CE capillary outlet and a transfer capillary (Figure 10c).[111,128] A low-volume metal union has also been used to connect a transfer line to the CE capillary (Figure 10d).[82] Smith et al. described the use of a dialysis tube to cover a fracture close to the outlet of the capillary (Figure 10e),[162] and more recently, Hunt et al. described a further sheathless CE/MS interface (Figure 10f).[96]

Each of these designs has its merits and challenges, and these have been evaluated (for some of the above-mentioned interfaces) by Herring and Qin.[82] They evaluated the most appropriate CE/MS methodology for the analysis of phosphorylated tryptic peptides of myosin I heavy chain (MIHCK). Again, a method of analyte preconcentration on-line with the CE capillary was preferred to allow efficient sample handling. The findings of Qin and Herring are summarized in Table 1. One of the most common problems encountered during these studies was electrolysis and the generation of hydrogen bubbles at the junction of the CE capillary and the ESI emitter.[82] This invariably stopped the EOF, thereby terminating the experiment. To overcome this significant challenge, Qin and Herring inserted a 10-μm palladium wire into the outlet of the CE capillary (Figure 10b), which also acted as the ESI emitter. In their hands, this interface was reported to be easy to construct. In addition, a very useful property of palladium for this application is that it can absorb >900 times more hydrogen than its own volume. Therefore, unlike some of the other interfaces tested in their study, this simple sheathless CE/MS interface design did not suffer problems associated with hydrogen bubbles halting EOF. However, this interface did tend to induce the oxidation of some peptides.[82] For example, multiple oxidation of bombesin (EQRLGNQWAVGHLM) was demonstrated (see Figure 11). The placement of the palladium wire within the CE capillary also caused turbulence, and some loss of analyte resolution. However, these investigators were able to obtain good MS sensitivity for the two phosphopeptides expected in a tryptic digest of MIHCK, using this interface in conjunction with their adaptation of on-line solid-phase extraction/capillary electrophoresis/mass spectrometry (SPE/CE/MS).[82]

Clearly, miniaturized sheathless CE/MS designs will dominate the future of this technology. It is hoped

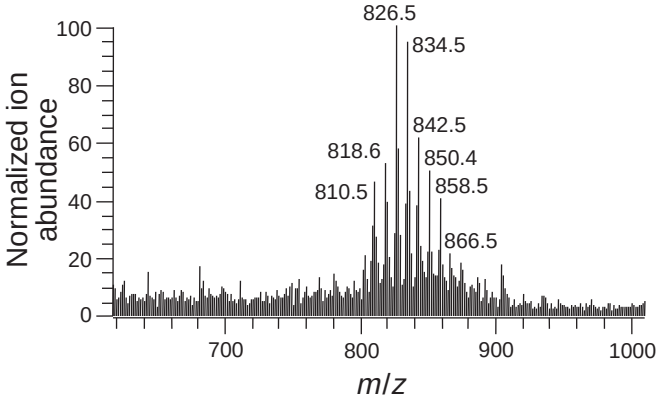

Figure 11 CE/microspray mass spectrum of bombesin using a palladium wire interface (Figure 10b). Note the mass difference of 8 u, corresponding to multiple sites of oxidation of this peptide. (Reproduced with permission of Dr Jun Qin and Christopher Herring.)

Table 1 Comparison of four microspray CE/MS interfaces (results are published with permission of Jun Qin and Christopher Herring)

Interface	Figure	Potential advantage	Observation by Qin and Herring	Reported limitations
Tapered coated tip	10(a)	Zero dead volume	Interface is not robust, conductive epoxy is lost within 2 days of operation	Interface needs to be rebuilt following loss of epoxy
			Interface works well for short periods	Oxidation of peptides can occur
Palladium wire	10(b)	Easy to construct Palladium adsorbs hydrogen	Easy to construct No loss of flow due to hydrogen bubbles	Some disruption to analyte resolution due to wire-induced turbulence Some peptides are prone to oxidation
Metal junction	10(d)	Very easy to construct Should be reliable	Electrolysis stopped flow within minutes of electrophoresis	EOF was not fast enough to elute hydrogen bubbles from the interface
Dialysis junction	10(e)	Small dead volume Relatively easy to construct	Tubing was prone to swelling, creating a large dead volume	Capillary alignment was lost owing to the swelling of tubing

that a consensus will be reached by the community as to the most appropriate way to achieve a sensitive and robust design for this approach of interfacing CE with MS.

4.5 Sample Preparation

As mentioned above, extreme caution is required during preparation of dilute solutions of peptides and proteins. These analytes readily adhere to every surface that they contact. Therefore, manipulations need to be minimized. However, biologically relevant peptides and proteins are often most soluble in solutions of high ionic strength. Such solutions are usually not optimal for separation by CE. Typically, these solutions will reduce the effect of stacking mechanisms, or compress the pH gradient in cIEF separations.[15,116] Clearly, the salt concentrations need to be reduced. However, this should not be at the expense of analyte concentration. In this regard, use of an on-line sample preparation (e.g. SPE/CE/MS or mPC/CE/MS) appears to be most appropriate.[27,50–96]

As described above, these techniques enable analyte preconcentration and sample cleanup with minimal intervention by the operator. Improved analyte recovery results. Furthermore, as we have found in our studies, the salt in the sample matrix may aid protein recovery from the solid phase or impregnated membrane.[129,130] Also, traces of salt that remain after such on-line cleanup may improve analyte stacking, provided a positively charged capillary is used for analyte separation.[129,130] For peptides the combination of reversed phase HPLC off-line with mPC/CE/MS has been proven to be a sensitive method of analysis, since subsequent sample preparation involves merely removing the organic solvent from fractions and diluting the residue in a suitable aqueous solvent.[87,88,96] This method could be further improved by on-line coupling of these techniques, as has been demonstrated by Jorgenson's group.[163] However, this tandem technique requires that the mass spectrometer is capable of high sensitivity at fast scan rates, since the CE separation is usually complete in just a few seconds. As such the development of instruments for fast scanning, highly sensitive electrospray ionization/time-of-flight mass spectrometry (ESI/TOFMS) will make this technology more viable.[120]

An area that has currently received only scant attention is that of on-line digestion CE/MS.[164] Such methodology, demonstrated by Kuhr and his group, appears to be a useful way of generating peptide maps of small amounts of biologically relevant proteins.[56,57,164] In conjunction with on-line SPE/CE/MS or mPC/CE/MS, digestion in an open-tubular enzyme capillary or an enzyme-modified solid support would appear to be a powerful methodology for proteomic research. Furthermore, up-front separation by an appropriate chromatographic step (either HPLC or CE) could provide an attractive method for characterizing the protein composition of a biological system, that once optimized would need little operator intervention. Hence analyte losses and sample contamination should be minimized. On-line, automated peptide/protein sequencing using tandem mass spectrometry (MS/MS) could also be achieved by this integrated approach.

5 METHOD DEVELOPMENT

Capillary preparation is the first consideration during the development of a new method. The approach taken is dependent upon the separation mechanism that is considered to be appropriate to yield analyte resolution. The simplest approach is to use a solution of sodium or potassium hydroxide (or methoxide) to etch and clean the silica surface. Following such cleaning, the capillary is usually washed with water and BGE. The capillary is then considered to be ready for analyte separation. However, for peptide analysis it is our experience that the capillary needs to be conditioned before it yields optimal performance. Therefore, in our hands, we usually analyze a standard solution of peptides until reproducible performance is obtained.[165] Typically, 2–5 analyses of the standard peptide solution are sufficient to condition the capillary, but, as noted above, there are some capillaries that do not yield acceptable performance. During our early encounters with CE and CE/MS, we would replace such capillaries and restart the laborious, time-consuming conditioning procedure. As we became more sophisticated users of CE, we found that flushing those capillaries that would not condition (in the conventional manner described above) with a solution of 70% formic acid in 1-propanol would restore CE separation performance.[166] This solution has been reported to remove hydrophobic proteins from HPLC columns, and may be flushing hydrophobic contaminants (that were introduced during manufacture) from the capillary. However, it was also apparent that the silica surface was modified, since subsequent washing of the CE capillary with potassium hydroxide immediately degraded the CE performance. In addition, we proved that this effect was not merely due to the use of a concentrated acid solution, since rinsing the capillary with 1 M hydrochloric acid was not able to restore the separation performance.[166]

It has been shown that in free solution peptide migration follows a linear relationship with $m^{2/3}/z$, where m is the molecular weight of the peptide and z is its

charge or valency [see McCormick[167] for a review of this subject]. Consequently, several computer programs have been developed to correlate the p*I* of a peptide with the charge on the molecule at a given pH. Therefore, if the peptide composition of a mixture is known, the free solution CE conditions that would effect separation can be predicted. From these considerations, the most important parameter that effects the selectivity of the separation is the pH of the BGE. Other parameters such as the ionic strength of the BGE, capillary dimensions, temperature, and electrical field strength influence analyte separation in free solution. However, for peptides, the most direct route to method optimization is to change the pH of the BGE. As mentioned above, BGEs prepared with volatile salts have been reported to be most suitable for all CE/MS studies. Acidic BGEs are usually prepared with mixtures of ammonium acetate and acetic acid solutions, and those at higher pH are typically composed of mixtures of ammonium acetate and ammonia solutions. One issue with these solutions is that they do not exhibit high buffer capacity, and this can lead to irreproducible analyte migration times, as ions are depleted from the BGE. Frequent replenishment of the BGE with fresh solution can overcome this limitation and lead to improved reproducibility of analyte separation and migration.[165] In our studies, we have found for small peptides (molecular weight < 1500) that a solution of 2 mM ammonium acetate in 1% aqueous acetic acid provides a reasonable BGE for analysis by CE/MS.[86–88]

In addition, increasing the acetic acid concentration to 5% tends to enhance peptide separation.[128] Since the pH of the 5% acetic acid solution is very close to that of the 2 mM ammonium acetate in 1% aqueous acetic acid BGE, we attribute the increased peptide resolution to an increased BGE viscosity.[128] More recently, while such ammonium acetate BGEs can be used effectively in uncoated fused-silica capillaries, we have tended to favor the use of the polybrene coating for CE/MS analysis of peptides.[128] As mentioned above, this coating reverses the charge on the capillary wall and tends to prevent peptide adsorption. However, it is our experience that this coating requires conditioning, with the reproducibility of peptide migration times and resolution improving with each analysis. Typically, a polybrene-coated capillary is conditioned within five analyses. This capillary then yields optimal performance for 15–20 separations, after which it has to be completely stripped and recoated (a cycle that can be repeated about 20 times before replacing the capillary).[128] Adding polybrene to the BGE to coat the capillary dynamically during use would keep it functional longer. However, this is not an option for CE/MS studies, since this reagent would enter the mass spectrometer and interfere with ESI processes and method sensitivity. As noted above, the issue of limited CE and CE/MS concentration sensitivity has been addressed by several groups through the development of analyte concentrators and membrane preconcentrators. These have proved to be extremely useful for the analysis of low concentrations of biologically active peptides and enzymatic protein digests. Indeed, we now use mPC/CE/MS exclusively for this application, since it minimizes sample handling, which maximizes analyte recoveries.

The development of conditions for the separation of proteins is considerably more complex than the optimization of CE parameters for peptide resolution. Protein denaturation, aggregation, and precipitation and adsorption on the capillary wall are all factors that can degrade CE performance.[129,130] For most protein separations, a coated capillary is usually essential. Thereby, the complexity of analysis is already increased. In addition, there are a number of separation mechanisms (including free solution, cIEF, cITP, molecular weight sieving, etc.) that can be used to effect the separation of the analytes.[11–24] All of these options and the usual CE parameters need to be considered prior to attempting to separate proteins by CE. We recently evaluated a number of coated capillaries for the analysis of aqueous humor.[168] As noted above, we found that the salt matrix of this physiologically derived fluid interfered with the separation of the proteins that are present in this fluid. In most cases, only a single peak was detected.[168] However, use of a polybrene capillary permitted good resolution of the components of this important physiological fluid. We have also found that a BGE composed of ammonium acetate and acetic acid is very suitable for protein analysis by CE/MS, with increasing acetic acid concentration often improving analyte resolution.[168] Again, we are strong advocates of the use of mPC/CE/MS for this application, since this technique enables both analyte concentration and sample cleanup without excessive sample handling (see Figure 7).[129,130]

6 SEQUENCING PEPTIDES BY CAPILLARY ELECTROPHORESIS/TANDEM MASS SPECTROMETRY

Peptide sequencing by MS/MS, with two mass analyzers coupled via a collision cell, was pioneered by Hunt et al.[169] It is a technique that has provided the sequence of peptides at concentrations that are below the typical DLs of conventional Edman chemistries. The development of the ESI interface, along with on-line condensed-phase chromatography [high-performance liquid chromatography/mass spectrometry (HPLC/MS) and CE/MS] has further extended the use of this approach for peptide sequencing. DLs are often in the sub-picomole range of material. Typical peptide cleavages are shown

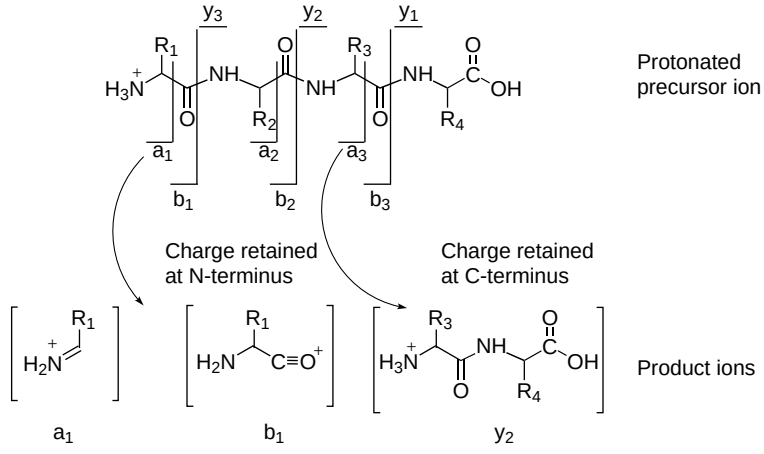

Figure 12 Low-energy collision-induced peptide fragmentation pathways in a tandem mass spectrometer.

in Figure 12. Random cleavage of the amide bonds is desired, and often results from low-energy collisions (as are typical in triple-quadrupole, hybrid, or ion trap instruments). It is often possible to read the peptide sequence from both the N- and C-termini from these data. However, there is ambiguity since isobaric amino acids (e.g. leucine and isoleucine) give rise to the generation of identical fragment ions. On occasion, these may be resolved by high-energy collisions, which induce side-chain fragmentation. More recently, the use of protein database searching algorithms has also tended to overcome the problem of identifying the isobaric amino acids.[1] By this approach, peptide sequences are derived by matching the acquired data with peptides that are produced from the known proteins in the database, hence there is often little ambiguity in the amino acid sequence.

In recent years, with the exception of a few isolated cases, high-performance liquid chromatography/tandem mass spectrometry (HPLC/MS/MS) has dominated this area of research.[170] This can be attributed, at least in part, to a better concentration sensitivity of HPLC as compared with CE.[27] Furthermore, chromatographic peak widths may be greater in HPLC separations and more compatible with scan rates of the scanning mass spectrometer. In addition, coupling of the HPLC column to the electrospray interface has been much less challenging to achieve than interfacing CE to the same ionization source. However, CE does afford a second complementary dimension of separation, and we have exploited this to sequence MHC class I peptides.[92,94,171] In a first dimension, these immunologically important peptides are isolated by reversed-phase HPLC. Each fraction is then concentrated to remove the acetonitrile, and analyzed by mPC/CE/MS. Indeed, the development of SPE/CE/MS and mPC/CE/MS has provided technology that has a comparable concentration sensitivity to that of HPLC. However, analyte peak widths of SPE/CE/MS and mPC/CE/MS separations are a challenge to the scanning mass spectrometer (e.g. triple-quadrupole or sector-quadrupole hybrid instrument). As mentioned above, MS sensitivity is reduced at high scan rates. Slowing the rate of data acquisition is often not an option, particularly for CE separations. Often, mass resolution is sacrificed in both mass spectrometers MS1 and MS2 to attempt to regain some of the sensitivity that is lost to the fast scan.[87,88,170] Such a compromise usually leads to a more difficult spectrum interpretation, since isobaric peptides may be introduced simultaneously into the collision cell. Fragments of both are then detected in the resultant spectrum. A further compromise of this approach is that mass accuracy may be sacrificed, with an error in the m/z measurement of fragment ions no better than ± 1 Da. Since this is the mass difference between some amino acids (e.g. I/L and N) ambiguity is introduced into the peptide sequence. However, this situation is often independent of the mode of separation, when sensitivity is pushed to the limits of the scanning mass spectrometer. Reported peptide sequencing sensitivities using scanning MS are comparable to those for HPLC/MS/MS (5 fmol on-column)[170] and mPC/CE/MS/MS (10–50 fmol on-column).[87,88] More recently, ion trap devices (ITMS)[96] and FTMS instruments[172] have been used to acquire peptide sequence data. Indeed, as described above, Hunt et al. have reported 50 amol sensitivity for sequencing MHC class I peptides by mPC/tITP/CE/MS with an ion trap instrument (see Figure 6).[96] Such increase in sensitivity can be attributed to the manner in which these instruments are operated. In these devices ions are trapped and accumulated for a specific length of time. Furthermore, in MS/MS experiments all but the ions of interest are ejected from the trap prior to

peptide fragmentation and subsequent analysis of the generated fragment ions. This approach often leads to a less complex tandem mass spectrum, and if the trap is not overloaded with ions, mass accuracy can be better than is achieved by a scanning instrument. In many areas of research such an increase in sensitivity can reduce the need for growing huge cell colonies, which is both costly and time-consuming. In studies of MHC class I peptide cell colonies, in excess of 10^{10} cells are extracted in contemporary methodologies to yield sufficient peptide for sequencing.[96,170,171] As MS sensitivity is improved this number should fall significantly, and sequencing peptides from a colony of 10^8 cells may become routine in the very near future.

7 FUTURE DIRECTIONS

Developing practices in the area of proteomics will inevitably be more demanding on the sensitivity of analytical methods.[1] Current practices demand that the sensitivity of protein sequencing methods is comparable to that of slab gel silver staining techniques.[110] While these approaches are yielding results, it is likely that many biologically significant proteins will not be detected or characterized. However, at the sensitivity levels that will be required in this area of research, sample manipulations become a major challenge. Both analyte losses and sample contamination by ubiquitous human proteins (such as keratin) become significant issues at picomolar concentrations.[110] Automated preparation and analysis of low-volume, low-concentration samples of high complexity will be required to overcome these issues. Clean-room technology is also likely to become a major requirement of these investigations. Figeys et al. have proposed a solution to these issues.[110] These investigators have suggested the use of coupled microfabricated, integrated analytical modules with MS for this application, and they have proposed microfabricated device/mass spectrometry (MFD/MS) as the acronym for this technology.[110] While research on these devices is in its infancy, such a miniaturized approach coupled with a sensitive mass spectrometer could provide the DLs required in the next generation of proteomic studies. Furthermore, with careful design of the array, on-chip protein isolation, digestion, and analysis could be accommodated in this nanoenvironment.[110] However, the use of nanotechnologies will provide new challenges. First, creative coupling of the macro world of biochemistry to the nanoanalytical environment will be required. In this regard, the lessons learned from the development of mPC/CE/MS and SPE/CE/MS technologies should aid these studies, particularly in the area of reducing potentially long sample injection times, caused by slow liquid flow rates in these nanodevices. The sensitivity of protein sequencing will also require further creative input. In this area, interfacing the microfabricated array to the mass spectrometer will need particular attention.

So far, since MFD/MS research is in its infancy, there have been few reports in the scientific literature.[110,114,115,173] Most devices are prepared using technology that was first described in the computer microchip industry. Typically, reservoirs and microchannels are etched into glass, which is bonded to a second piece of glass, and electrodes are inserted to complete the device. There have also been limited reports of the use of polymers to prepare for this application, and this is an area that may require much more attention. It is known from conventional CE that proteins adsorb to glass capillaries, leading to the development of a large number of capillary coatings to prevent protein adsorption to the capillary wall (see earlier discussions).[104] It is envisaged that similar coatings will need to be applied to a microfabricated device (MFD) to prevent protein adsorption on these devices, and improve separation or manipulation of these important analytes in the nanoenvironment.

Currently, there is no consensus in the literature as to the most appropriate method for interfacing an MFD with a mass spectrometer. It is agreed that the micro- or nanoelectrospray interface is the most efficient ionization source for this application.[110,114,115,173] In one of the approaches described, an array of nine channels was continued to the edge of the device, which served as the ESI/mass spectrometer interface.[173] Analytes were mobilized by the application of pressure to the appropriate reservoir and ionized directly from the edge of the unit. The different channels were moved sequentially in front of the inlet orifice of the mass spectrometer to permit the characterization of individual analytes. Ramsey and Ramsey have described a similar approach, and demonstrated an MFD/MS separation with their device.[114] In contrast, Figeys et al. preferred to use a transfer capillary and a liquid junction microspray (see Figure 10c) to interface an MFD to the mass spectrometer.[110] Enhanced MS sensitivity was given as the reason for their preference for this interface. However, it does significantly extend analysis times. This technique may be most suitable for enhancing separations, by a combined microfabricated device/capillary electrophoresis/mass spectrometry (MFD/CE/MS) approach. More recently, Foret and Karger described the use of ionspray directly from the edge of the MFD.[174] This device includes the ability to apply a nebulizing gas to aid the electrospray processes. However, the use of the nebulizing gas induces hydrodynamic flow within the device, hence a makeup solvent is

required. The use of a makeup solvent inevitably dilutes the separated analytes prior to their transport into the mass spectrometer, and may reduce the ultimate sensitivity of this approach. However, careful selection of this solvent may aid ionization efficiency and thus tend to improve method sensitivity.[174]

A further technology that is expected to improve MS sensitivity is the development of the highly sensitive ESI/TOFMS,[120] and electrospray ionization quadrupole time-of-flight mass spectrometry (ESI/QTOFMS) and a quadrupole ion trap coupled with a reflecting TOF instrument.[175,176] The development of such instruments is well under way. In addition to high sensitivity, these instruments have a short duty cycle, which allows very fast data acquisition, with no loss of sensitivity. These instruments also provide a constant medium resolution (>5000 full width at half-maximum definition) and good mass accuracy. The use of isotopic labeling (as is achieved when enzymatic protein digests are carried out in ^{18}O-labeled water) can aid peptide identification with MS/MS data since only Y ions will contain this label. The mass resolution of the TOF instruments allows ^{18}O-enriched ions to be readily identified in MS/MS spectra, to facilitate peptide sequence identification.[177]

In summary, it is clear that there will need to be a significant pursuit of more sensitivity for CE/MS. In this regard, MFD/MS may, with further development and possible coupling to a highly sensitive mass spectrometer (such as ESI/TOFMS), provide the technology that is needed for future generations of proteomic research. However, lessons learnt in the macro world of CE/MS will need to be heeded as new MFD and MS interfacing techniques are developed. In particular, input by materials scientists may provide improved surfaces to prevent interaction and loss of peptides and proteins within the channels of MFDs. Assuming that these challenges will be overcome, technology for the isolation, manipulation, digestion, and sequencing of extremely low levels of biologically active proteins will become a reality within the next few years.

ACKNOWLEDGMENTS

The authors thank Mrs Diana Ayerhart for her help in preparing this article. They are also grateful to Christopher Herring and Dr Jun Qin (Laboratory of Biophysical Chemistry, National Heart, Lung, and Blood Institute) for sharing the results of their studies of microspray CE/MS interfaces and Professor Hunt (University of Virginia) and Professor Karger (Barnet Institute Northeastern University) for graciously sharing their work. The support of funding by Mayo Foundation, Finnigan MAT, Beckman Instruments, AHAF, and NIH is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

ACE	Affinity Capillary Electrophoresis
ACE/MS	Affinity Capillary Electrophoresis/Mass Spectrometry
APS	Aminopropyltrimethoxy-silane
BGE	Background Electrolyte
CE	Capillary Electrophoresis
CE/MS	Capillary Electrophoresis/ Mass Spectrometry
CE/tITP/MS	Capillary Electrophoresis/ Transient Isotachophoresis/ Mass Spectrometry
CFFAB	Continuous-flow Fast Atom Bombardment
cIEF	Capillary Isoelectric Focusing
cIEF/MS	Capillary Isoelectric Focusing/Mass Spectrometry
cITP	Capillary Isotachophoresis
cITP/CE/MS	Capillary Isotachophoresis/ Capillary Electrophoresis/ Mass Spectrometry
CSF	Cerebrospinal Fluid
DL	Detection Limit
EOF	Electroosmotic Flow
ESI	Electrospray Ionization
ESI/QTOFMS	Electrospray Ionization Quadrupole Time-of-flight Mass Spectrometry
ESI/TOFMS	Electrospray Ionization/ Time-of-flight Mass Spectrometry
FTMS	Fourier Transform Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
HPLC/MS/MS	High-performance Liquid Chromatography/Tandem Mass Spectrometry
ID	Internal Diameter
ITMS	Ion Trap Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/Ionization
MALDI/TOFMS	Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry

MEKC	Micellar Electrokinetic Chromatography
MFD	Microfabricated Device
MFD/CE/MS	Microfabricated Device/ Capillary Electrophoresis/ Mass Spectrometry
MFD/MS	Microfabricated Device/ Mass Spectrometry
MHC	Major Histocompatibility Complex
MIHCK	Myosin I Heavy Chain
mPC	Membrane Preconcentration
mPC/CE	Membrane Preconcentration/ Capillary Electrophoresis
mPC/CE/MS	Membrane Preconcentration/ Capillary Electrophoresis/ Mass Spectrometry
mPC/CE/MS/MS	Membrane Preconcentration/ Capillary Electrophoresis/ Tandem Mass Spectrometry
mPC/tITP/CE/MS	Membrane Preconcentration/ Transient Isotachophoresis/ Capillary Electrophoresis/ Mass Spectrometry
mPC/tITP/CE/MS/MS	Membrane Preconcentration/ Transient Isotachophoresis/ Capillary Electrophoresis/ Tandem Mass Spectrometry
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass-to-charge Ratio
PAA	Polyacrylamide
PEG	Polyethylene Glycol
PEI	Polyethylenimine
PEX	Pseudoexfoliation
PVA	Poly(vinyl alcohol)
RES	Reduced Elution Speed
SIM	Selected Ion Monitoring
S/N	Signal-to-noise Ratio
SPE/CE/MS	Solid-phase Extraction/ Capillary Electrophoresis/Mass Spectrometry
TFA	Trifluoroacetic Acid
tITP	Transient Isotachophoresis
TOF	Time-of-flight
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Clinical Chemistry **(Volume 2)**
Biochemical Markers of Acute Coronary Syndromes • Capillary Electrophoresis in Clinical Chemistry • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Serum Proteins

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. J. Yates III, 'Mass Spectrometry and the Age of the Proteome', *J. Mass Spectrom.*, **33**, 1–19 (1998).
2. M. Mann, M. Wilm, 'Error-tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags', *Anal. Chem.*, **66**, 4390–4399 (1994).
3. M. Wilm, M. Mann, 'Analytical Properties of the Nanoelectrospray Ion Source', *Anal. Chem.*, **68**, 1–8 (1996).
4. R.M. Caprioli, 'On-line LC/MS and CE/MS', in *Biological Mass Spectrometry: Present and Future*, eds. T. Matsuo, R.M. Caprioli, M.L. Gross, Y. Seyama, J. Wiley & Sons, Chichester, 1994.
5. J. Cai, J.D. Henion, 'Capillary Electrophoresis/Mass Spectrometry', *J. Chromatogr.*, **703**, 667–692 (1995).
6. J.P. Landers, R.P. Oda, T.C. Spelsberg, J.A. Nolan, K.J. Ulfelder, 'Capillary Electrophoresis: A Powerful Microanalytical Technique for Biologically Active Molecules', *Biotechniques*, **14**, 98–111 (1993).
7. R.D. Smith, J.A. Olivares, N.T. Nguyen, H.R. Udseth, 'Capillary Zone Electrophoresis/Mass Spectrometry using an Electrospray Ionization interface', *Anal. Chem.*, **60**, 436–441 (1988).
8. R.D. Smith, J.A. Loo, C.J. Barinaga, C.G. Edmonds, H.R. Udseth, 'Capillary Zone Electrophoresis and Isotachophoresis/Mass Spectrometry of Polypeptides and

Proteins Based Upon an Electrospray Ionization Interface', *J. Chromatogr.*, **480**, 211–232 (1989).
9. E.D. Lee, W. Mueck, J.D. Henion, T.R. Covey, 'On-line Capillary Zone Electrophoresis in Spray Tandem Mass Spectrometry for the Determination of Dynorphons', *J. Chromatogr.*, **458**, 313–321 (1988).
10. I.M. Johannson, E.C. Huang, J.D. Henion, J. Zweigenbaum, 'Capillary Electrophoresis for the Characterization of Peptides. Instrumental Considerations for Mass Spectrometric Detection', *J. Chromatogr.*, **554**, 311–327 (1991).
11. S. Hjerten, M.-D. Zhu, 'Adaptation of the Equipment for High Performance Electrophoresis to Isoelectric Focusing', *J. Chromatogr.*, **346**, 255–270 (1985).
12. J. Wu, J. Pawliszyn, 'Application of Capillary Isoelectric Focusing with Absorption Imaging Detection to the Analysis of Proteins', *J. Chromatogr. B*, **657**, 327–332 (1994).
13. X. Liu, Z. Sosic, I.S. Krull, 'Capillary Isoelectric Focusing as a Tool in the Examination of Antibodies, Peptides and Proteins of Pharmaceutical interest', *J. Chromatogr. A*, **735**, 165–190 (1996).
14. T.J. Pritchett, 'Capillary Isoelectric Focusing of Proteins', *Electrophoresis*, **17**, 1195–1201 (1996).
15. N.J. Clarke, A.J. Tomlinson, G. Schomberg, S. Naylor, 'Capillary Isoelectric Focusing of Physiologically Derived Proteins with On-line Desalting of Isotonic Salt Concentrations', *Anal. Chem.*, **69**, 2786–2792 (1997).
16. L. Krivankova, P. Bocek, 'Capillary Isotachophoresis', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 251–275, 1998.
17. P. Shieh, N. Cooke, A. Guttman, 'Capillary Gel Electrophoresis', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 185–221, 1998.
18. M.G. Khaledi, 'Micellar Electrokinetic Chromatography', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 77–140, 1998.
19. Y.-H. Chu, L.Z. Avila, J. Goa, G.M. Whitesides, 'Affinity Capillary Electrophoresis', *Acc. Chem. Res.*, **28**, 461–468 (1995).
20. Y.-L. Hsieh, J. Cai, Y.-T. Li, J.D. Henion, 'Detection of Noncovalent FKBP-FK506 and FKBP–Rapamycin Complexes by Capillary Electrophoresis–Mass Spectrometry and Capillary Electrophoresis–Tandem Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **6**, 85–90 (1995).
21. Y.-H. Chu, D.P. Kirby, B.L. Karger, 'Free Solution Identification of Candidate Peptides from Combinatorial Libraries by Affinity Capillary Electrophoresis/Mass Spectrometry', *J. Am. Chem. Soc.*, **117**, 5419–5420 (1995).
22. Y.-H. Chu, Y.M. Dunayevskiy, D.P. Kirby, P. Vouros, B.L. Karger, 'Affinity Capillary Electrophoresis–Mass Spectrometry for Screening Combinatorial Libraries', *J. Am. Chem. Soc.*, **118**, 7822–7835 (1996).
23. Y.V. Lyubarskaya, Y.M. Dunayevskiy, P. Vouros, B.L. Karger, 'Microscale Epitope Mapping by Affinity Capillary Electrophoresis–Mass Spectrometry', *Anal. Chem.*, **69**, 3008–3014 (1997).
24. C.C. Cheng, Y.-H. Chu, 'Affinity Capillary Electrophoresis in Combinatorial Library Screening', *Am. Lab.*, **30**, 79–81 (1998).
25. J. Cai, Z. El Rassi, 'On-line Preconcentration of Triazine Herbicides with Tandem Octadecyl Capillaries–Capillary Zone Electrophoresis', *J. Liq. Chromatogr.*, **15**, 1179–1192 (1992).
26. J. Cai, Z. El Rassi, 'Selective On-line Preconcentration of Proteins by Tandem Metal Chelate Capillaries–Capillary Zone Electrophoresis', *J. Liq. Chromatogr.*, **16**, 2007–2024 (1993).
27. A.J. Tomlinson, N.A. Guzman, S. Naylor, 'Enhancement of Concentration Limits of Detection in CE and CE/MS: A Review of On-line Sample Extraction, Cleanup, Analyte Preconcentration, and Microreactor Technology', *J. Capillary Electrophor.*, **2**(6), 247–266 (1995).
28. J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, 'Attomole Level Capillary Electrophoresis–Mass Spectrometric Protein Analysis Using 5-μm ID Capillaries', *Anal. Chem.*, **64**, 3194–3196 (1992).
29. F.E.P. Mikkers, F.M. Everaerts, T.P.E.M. Verheggen, 'Concentration Distributions in Free Zone Electrophoresis', *J. Chromatogr.*, **169**, 1–10 (1979).
30. S.E. Moring, J.C. Colburn, P.D. Grossman, H.H. Lauer, 'Analytical Aspects of an Automated Capillary Electrophoresis System', *LC/GC*, **8**, 34–36 (1989).
31. R. Abersold, H.D. Morrison, 'Analysis of Dilute Peptide Samples by Capillary Zone Electrophoresis', *J. Chromatogr.*, **516**, 79–88 (1990).
32. P. Gebauer, W. Thormann, P. Bocek, 'Sample Self-stacking in Zone Electrophoresis: Theoretical Description of the Zone Electrophoretic Separation of Minor Compounds in the Presence of Bulk Amounts of a Sample Component with High Mobility and Like Charge', *J. Chromatogr.*, **608**, 47–57 (1992).
33. C. Schwer, F. Lottspeich, 'Analytical and Micropreparative Separation of Peptides by Capillary Zone Electrophoresis using Discontinuous Buffer Systems', *J. Chromatogr.*, **623**, 345–355 (1992).
34. J.L. Beckers, M.T. Ackermans, 'Effect of Sample Stacking on Resolution, Calibration Graphs and pH in Capillary Zone Electrophoresis', *J. Chromatogr.*, **629**, 371–378 (1993).

35. R. Chien, D.S. Burgi, 'Field Amplified Sample injection in High Performance Capillary Electrophoresis', *J. Chromatogr.*, **559**, 141–152 (1991).
36. D.S. Burgi, R. Chien, 'Optimization in Sample Stacking for High Performance Capillary Electrophoresis', *Anal. Chem.*, **63**, 2042–2047 (1991).
37. R. Chien, D.S. Burgi, 'Sample Stacking of an Extremely Large Injection Volume in High Performance Capillary Electrophoresis', *Anal. Chem.*, **64**, 1046–1050 (1992).
38. R.L. Chien, D.S. Burgi, 'On-line Sample Concentration Using Field Amplification in CZE', *Anal. Chem.*, **64**, 489A–496A (1992).
39. F. Foret, V. Sustacek, P. Bocek, 'On-line Isotachophoretic Sample Preconcentration for Enhancement of Zone Detectability in Capillary Zone Electrophoresis', *J. Microcol. Sep.*, **2**, 229–233 (1990).
40. R.D. Smith, S.M. Fields, J.A. Loo, C.J. Barinaga, H.R. Udseth, C.G. Edmonds, 'Capillary Isotachophoresis with UV and Tandem Mass Spectrometric Detection for Peptides and Proteins', *Electrophoresis*, **11**, 709–717 (1990).
41. J.L. Beckers, F.M. Everaerts, 'Isotachophoresis with Two Leading Ions and Migration Behaviour in Capillary Zone Electrophoresis', *J. Chromatogr.*, **508**, 19–26 (1990).
42. M.H. Lamoree, N.J. Reinhoud, U.R. Tjaden, W.M.A. Niessen, J. van der Greef, 'On-capillary Isotachophoresis for Loadability Enhancement in Capillary Zone Electrophoresis/Mass Spectrometry of β-Agonists', *Biol. Mass Spectrom.*, **23**, 339–345 (1994).
43. S.J. Locke, P. Thibault, 'Improvement in Detection Limits for the Determination of Paralytic Shellfish Poisoning Toxins in Shellfish Tissues Using Capillary Electrophoresis/Electrospray Mass Spectrometry and Discontinuous Buffer Systems', *Anal. Chem.*, **66**, 3436–3446 (1994).
44. N.J. Reinhoud, U.R. Tjaden, J. van der Greef, 'Automated Isotachophoretic Analyte Focusing for Capillary Zone Electrophoresis in a Single Capillary Using Hydrodynamic Back-pressure Programming', *J. Chromatogr.*, **641**, 155–162 (1993).
45. L.M. Benson, A.J. Tomlinson, J.M. Reid, D.L. Walker, M.M. Ames, S. Naylor, 'Study of In Vivo Pyrazoloacridine Metabolism by Capillary Electrophoresis Using Isotachophoresis Preconcentration in a Non-aqueous Separation Buffer', *J. High Resolut. Chromatogr.*, **16**, 324–326 (1993).
46. T.J. Thompson, F. Foret, P. Vouros, B.L. Karger, 'Capillary Electrophoresis/Electrospray Ionization Mass Spectrometry: Improvement of Protein Detection Limits Using On-column Transient Isotachophoretic Sample Preconcentration', *Anal. Chem.*, **65**, 900–906 (1993).
47. D.S. Stegehuis, H. Irth, U.R. Tjaden, J. van der Greef, 'Isotachophoresis as an On-line Concentration Pretreatment Technique in Capillary Electrophoresis', *J. Chromatogr.*, **538**, 393–402 (1991).
48. A.P. Tinke, N.J. Reinhoud, W.M.A. Niessen, U.R. Tjaden, J. van der Greef, 'On-line Isotachophoretic Analyte Focusing for Improved Detection Limits in Capillary Electrophoresis/Electrospray Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 560–563 (1992).
49. F. Foret, E. Szoko, B.L. Karger, 'On-line Transient and Coupled Column Isotachophoretic Preconcentration of Protein Samples in Capillary Zone Electrophoresis', *J. Chromatogr.*, **608**, 3–12 (1992).
50. V. Kasicka, Z. Prusík, 'Isotachophoretic Electrodesorption of Proteins from an Affinity Adsorbent on a Microscale', *J. Chromatogr.*, **273**, 117–128 (1983).
51. N.A. Guzman, M.A. Trebilcock, J.P. Advis, 'The Use of a Concentration Step to Collect Urinary Components Separated by Capillary Electrophoresis and Further Characterization of Collected Analytes by Mass Spectrometry', *J. Liq. Chromatogr.*, **14**, 997–1015 (1991).
52. N.A. Guzman, 'Automated Capillary Electrophoresis Apparatus', *US Pat.*, 5 202 0101, April 13, 1993.
53. N.A. Guzman, 'Automated Capillary Electrophoresis Apparatus', *World Pat.*, WO93/05390, March 18, 1993.
54. N.A. Guzman, C.L. Gonzalez, M.A. Trebilcock, L. Her nandez, C.M. Berck, J.P. Advis, 'The Use of Capillary Electrophoresis in Clinical Diagnosis', in *Capillary Electrophoresis Technology*, ed. N.A. Guzman, Marcel Dekker, New York, 643–672, 1993.
55. A.J. Debets, M. Mazereeuw, W.H. Voogt, D.J. van Iperen, H. Lingeman, K.-P. Hupe, U.A.T. Brinkman, 'Switching Valve with internal Micro Precolumn for On-line Sample Enrichment in Capillary Zone Electrophoresis', *J. Chromatogr.*, **608**, 151–158 (1992).
56. L.N. Amankwa, W.G. Kuhr, 'Trypsin-modified Fused Silica Capillary Microreactor for Peptide Mapping by Capillary Zone Electrophoresis', *Anal. Chem.*, **64**, 1610–1613 (1992).
57. L.N. Amankwa, W.G. Kuhr, 'On-line Peptide Mapping by Capillary Zone Electrophoresis', *Anal. Chem.*, **65**, 2693–2697 (1993).
58. M. Fuchs, M. Merion, 'Apparatus for Effecting Capillary Electrophoresis', *US Pat.*, 5 246 577, September 21, 1993.
59. M.E. Swartz, M. Merion, 'On-line Sample Preconcentration on a Packed-inlet Capillary for Improving the Sensitivity of Capillary Electrophoretic Analysis of Pharmaceuticals', *J. Chromatogr.*, **632**, 209–213 (1993).
60. I. Morita, J. Sawada, 'Capillary Electrophoresis with On-line Sample Pretreatment for the Analysis of Biological Samples with Direct Injection', *J. Chromatogr.*, **641**, 375–381 (1993).
61. A.M. Hoyt, Jr, S.C. Beale, J.P. Larmann, Jr, J.W. Jorgenson, 'Preparation and Evaluation of an On-line Preconcentrator for Capillary Electrophoresis', *J. Microcol. Sep.*, **5**, 325–330 (1993).
62. A.J. Tomlinson, L.M. Benson, R.P. Oda, W.D. Braddock, M.A. Strausbauch, P.J. Wettstein, S. Naylor, 'Modification of Electroosmotic Flow in

Preconcentration–Capillary Electrophoresis (PC/CE)', *J. High Resolut. Chromatogr.*, **17**, 669–671 (1994).
63. L.M. Benson, A.J. Tomlinson, S. Naylor, 'Time Course Analysis of a Microsomal Incubation of a Therapeutic Drug Using Preconcentration Capillary Electrophoresis (PC/CE)', *J. High Resolut. Chromatogr.*, **17**, 671–673 (1994).
64. A.J. Tomlinson, L.M. Benson, W.D. Braddock, R.P. Oda, S. Naylor, 'On-line Preconcentration–Capillary Electrophoresis–Mass Spectrometry (PC/CE/MS)', *J. High Resolut. Chromatogr.*, **17**, 729–732 (1994).
65. S. Naylor, L.M. Benson, A.J. Tomlinson, 'Capillary Zone Electrophoresis for the Analysis of Small Molecules: Practical Considerations', *J. Capillary Electrophor.*, **1**, 181–189 (1994).
66. N. Brenner, R.H. Palmieri, 'On-column Preconcentration of Samples in Capillary Electrophoresis', *US Pat.*, 5 340 452, August 23, 1994.
67. H.A. Fishman, J.B. Shear, L.A. Colon, R.N. Zare, J.V. Sweedler, 'On-column Derivatization in Capillary Electrophoresis', *US Pat.*, 5 318 680, June 7, 1994.
68. M. Fuchs, M. Merion, 'Process and Apparatus for Effecting Capillary Electrophoresis', *Eur. Pat.*, EP0 459 241 B1, October 26, 1994.
69. S. Pálmarsdóttir, L.-E. Edholm, 'Enhancement of Selectivity and Concentration Sensitivity in Capillary Zone Electrophoresis by On-line Coupling with Column Liquid Chromatography and Utilizing a Double Stacking Procedure Allowing for Microliter injections', *J. Chromatogr. A*, **693**, 131–143 (1995).
70. A.J. Tomlinson, W.D. Braddock, L.M. Benson, R.P. Oda, S. Naylor, 'Preliminary Investigations of Preconcentration/Capillary Electrophoresis–Mass Spectrometry', *J. Chromatogr. B*, **669**, 67–73 (1995).
71. J. Wiktorowicz, 'On-line Ion Contaminant Removal Apparatus and Method for CZE', *US Pat.*, 5 423 966, June 13, 1995.
72. M.A. Strausbauch, B.J. Madden, P.J. Wettstein, J.P. Landers, 'Sensitivity Enhancement and Second-dimensional Information from Solid Phase Extraction–Capillary Electrophoresis of Entire High Performance Liquid Chromatography Fractions', *Electrophoresis*, **16**, 541–548 (1995).
73. L.J. Cole, R.T. Kennedy, 'Selective Preconcentration for Capillary Zone Electrophoresis Using Protein G Immunoaffinity Capillary Chromatography', *Electrophoresis*, **16**, 549–556, (1995).
74. J.H. Beattie, R. Self, M.P. Richards, 'The Use of Solid Phase Concentrators for On-line Preconcentration of Metallothionein Prior to Isoform Separation by Capillary Zone Electrophoresis', *Electrophoresis*, **16**, 322–328 (1995).
75. N.A. Guzman, 'Consecutive Protein Digestion and Peptide Derivatization Employing an On-line Analyte Concentrator to Map Proteins Using Capillary Electrophoresis', in *Capillary Electrophoresis: An Analytical Tool in Biotechnology*, ed. P.G. Rigetti, CRC Press, Boca Raton, FL, 101–121, 1995.
76. N.A. Guzman, 'Biomedical Applications of On-line Preconcentration–Capillary Electrophoresis Using an Analyte Concentrator: Investigation of Design Options', *J. Liq. Chromatogr.*, **18**, 3751–3768 (1995).
77. M.A. Strausbauch, S.J. Xu, J.E. Ferguson, M.E. Nunez, D. Machacek, G.M. Lawson, P.J. Wettstein, J.P. Landers, 'Concentration and Separation of Hypoglycemic Drugs Using Solid Phase Extraction–Capillary Electrophoresis', *J. Chromatogr. A*, **717**, 279–291 (1995).
78. M.A. Strausbauch, J.P. Landers, P.J. Wettstein, 'Mechanism of Peptide Separations by Solid Phase Extraction Capillary Electrophoresis at Low pH', *Anal. Chem.*, **68**, 306–314, (1996).
79. D. Figeys, A. Ducret, J.R. Yates III, R. Aebersold, 'Protein Identification by Solid Phase Microextraction–Capillary Zone Electrophoresis–Microspray–Tandem Mass Spectrometry', *Nature Biotechnol.*, **14**, 1579–1583 (1996).
80. D. Figeys, A. Ducret, R. Aebersold, 'Identification of Proteins by Capillary Electrophoresis–Tandem Mass Spectrometry: Evaluation of an On-line Solid-phase Extraction Device', *J. Chromatogr. A*, **763**, 295–306 (1997).
81. D. Figeys, R. Aebersold, 'High Sensitivity Identification of Proteins by Electrospray Ionization Tandem Mass Spectrometry: Initial Comparison Between an Ion Trap Mass Spectrometer and a Triple Quadrupole Mass Spectrometer', *Electrophoresis*, **18**, 360–368 (1997).
82. C.J. Herring, J. Qin, 'Sheathless CE/MS with On-line Preconcentration', presented at the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, June 1–4, 1998.
83. S. Naylor, A.J. Tomlinson, L.M. Benson, W.D. Braddock, R.P. Oda, 'Preseparation Processor for Use in Capillary Electrophoresis', *US Pat.*, 8/423 220, 1995.
84. A.J. Tomlinson, L.M. Benson, R.P. Oda, W.D. Braddock, B.L. Riggs, J.A. Katzmann, S. Naylor, 'Novel Modifications and Clinical Applications of Preconcentration–Capillary Electrophoresis–Mass Spectrometry', *J. Capillary Electrophor.*, **2**, 97–104 (1995).
85. A.J. Tomlinson, L.M. Benson, W.D. Braddock, R.P. Oda, S. Naylor, 'Improved On-line Membrane Preconcentration–Capillary Electrophoresis (mPC/CE)', *J. High Resolut. Chromatogr.*, **18**, 381–383 (1995).
86. A.J. Tomlinson, S. Naylor, 'Enhanced Performance Membrane Preconcentration–Capillary Electrophoresis–Mass Spectrometry (mPC/CE/MS) in Conjunction with Transient Isotachophoresis to Analyze Peptide Mixtures', *J. High Resolut. Chromatogr.*, **18**, 384–386 (1995).
87. A.J. Tomlinson, S. Naylor, 'Systematic Development of On-line Membrane Preconcentration–Capillary Electro phoresis–Mass Spectrometry (mPC/CE/MS) for the Analysis of Peptide Mixtures', *J. Capillary Electrophor.*, **2**, 225–233 (1995).

88. A.J. Tomlinson, S. Naylor, 'A Strategy for Sequencing Peptides from Dilute Mixtures at the Low Femtomole Level Using Membrane Preconcentration Capillary Electrophoresis Tandem Mass Spectrometry', *J. Liq. Chromatogr.*, **18**, 3591–3615 (1995).
89. S. Naylor, A.J. Tomlinson, 'Membrane Preconcentration–Capillary Electrophoresis–Mass Spectrometry in the Analysis of Biologically Derived Metabolites and Biopolymers', *Biomed. Chromatogr.*, **10**, 325–330 (1996).
90. L.M. Benson, A.J. Tomlinson, A.N. Mayeno, D. Wells, S. Naylor, 'Membrane Preconcentration–Capillary Electrophoresis–Mass Spectrometry (mPC/CE/MS) Analysis of 3-Phenylamino-1,2-propanediol (PAP) Metabolites', *J. High Resolut. Chromatogr.*, **19**, 291–294 (1996).
91. E. Kurian, F.G. Prendergast, A.J. Tomlinson, M.W. Holmes, S. Naylor, 'Identifying Sites of Protein Modification by Using Matrix Assisted Laser Desorption/Ionization–Time-of-flight–Mass Spectrometry and On-line Membrane Preconcentration–Capillary Electrophoresis–Tandem Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **8**, 8–14 (1997).
92. A.J. Tomlinson, L.M. Benson, S. Jameson, D.H. Johnson, S. Naylor, 'Utility of Membrane Preconcentration–Capillary Electrophoresis–Mass spectrometry in Overcoming Limited Sample Loading for Analysis of Biologically Derived Drug Metabolites, Peptides, and Proteins', *J. Am. Soc. Mass Spectrom.*, **8**, 15–24 (1997).
93. A.J. Tomlinson, L.M. Benson, N.A. Guzman, S. Naylor, 'Enhancement of Concentration Limits of Detection in Capillary Electrophoresis: Examples of On-line Sample Preconcentration, Cleanup, and Microreactor Technology in Protein Characterization', in *Techniques in Protein Chemistry VIII*, ed. D.R. Marshall, Academic Press, San Diego, CA, 15–24, 1997.
94. A.J. Tomlinson, S. Jameson, S. Naylor, 'Sequencing MHC Class I Peptides Using Membrane Preconcentration–Capillary Electrophoresis–Tandem Mass Spectrometry (mPC/CE/MS/MS)', in *Techniques in Protein Chemistry VIII*, ed. D.R. Marshall, Academic Press, San Diego, CA, 25–35, 1997.
95. S. Naylor, A.J. Tomlinson, 'Membrane Preconcentration–Capillary Electrophoresis–Tandem Mass Spectrometry (mPC/CE/MS/MS) in the Sequence Analysis of Biologically Derived Peptides', *Talanta*, **45**, 603–612 (1998).
96. R.E. Settlage, P.S. Russo, J. Shabanowitcz, D.F. Hunt, 'A Novel μ-ESI Source for Coupling Capillary Electrophoresis and Mass Spectrometry: Sequence Determination of Tumor Peptides at the Attomole Level', *J. Microcol. Sep.*, **10**, 281–285 (1998).
97. R. Weinberger, *Practical Capillary Electrophoresis*, Academic Press, San Diego, CA, 1993.
98. J. Kohr, H. Engelhardt, 'Capillary Electrophoresis with Coated Capillaries', *Chromatogr. Sci. Ser.*, **64**, 357–381 (1993).
99. S. Madgassi, A. Kamyshny, 'Surface Activity and Functional Properties of Proteins', in *Surface Activity of Proteins*, ed. S. Madgassi, Marcel Dekker, New York, 2, 1996.
100. M. Chiari, M. Nesi, P.G. Righetti, 'Surface Modifications of Silica Walls: A Review of Different Methodologies', in *Capillary Electrophoresis in Biotechnology*, ed. P.G. Righetti, CRC Press, Boca Raton, FL, 1–36, 1996.
101. J. Kohr, H. Engelhardt, 'Characterization of Quartz Capillaries for Capillary Electrophoresis', *J. Chromatogr. A*, **652**, 309–316 (1993).
102. P. Thibault, C. Paris, S. Pleasance, 'Analysis of Peptides and Proteins by Capillary Electrophoresis/Mass Spectrometry Using Acidic Buffers and Coated Capillaries', *Rapid Commun. Mass Spectrom.*, **5**, 484–490 (1991).
103. R. Clark, J.A. Katzmann, E. Wiegert, C. Namyst-Goldberg, L. Sanders, R.P. Oda, R.A. Kyle, J.P. Landers, 'Rapid Capillary Electrophoretic Analysis of Human Serum Proteins: Qualitative Comparison with High-throughput Agarose Gel Electrophoresis', *J. Chromatogr. A*, **744**, 205–213 (1996).
104. G. Schomburg, 'Coated Capillaries in High Performance Capillary Electrophoresis', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 481–523, 1998.
105. M.A. Moseley, J.W. Jorgenson, J. Shabanowitz, D.F. Hunt, K.B. Tomer, 'Optimization of Capillary Zone Electrophoresis Electrospray Parameters for the Mass Spectrometry and Tandem Mass Spectrometry of Proteins', *J. Am. Soc. Mass Spectrom.*, **3**, 289–300 (1992).
106. J.H. Wahl, R.D. Smith, 'Comparison of Buffer Systems and Interface Designs for Capillary Electrophoresis–Mass spectrometry', *J. Capillary Electrophor.*, **1**, 62–71 (1994).
107. S. Pleasance, P. Thibault, J. Kelly, 'A Comparison of Liquid Junction and Coaxial Interfaces for Capillary Electrophoresis with Application to Compounds of Concern to the Aquaculture Industry', *J. Chromatogr.*, **591**, 325–329 (1992).
108. D.P. Kirby, J.M. Thorne, W.K. Gotzinger, B.L. Karger, 'A CE/ESI/MS interface for Stable, Low-flow Operation', *Anal. Chem.*, **68**, 4451–4457 (1996).
109. R.D. Smith, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, 'Capillary Electrophoresis/Mass Spectrometry', *Anal. Chem.*, **65**, 574A–584A (1993).
110. D. Figeys, I. van Oostveen, A. Ducret, R. Aebersold, 'Protein Identification by Capillary Zone Electrophoresis/Microelectrospray Ionization–Tandem Mass Spectrometry at the Subfemtomole Level', *Anal. Chem.*, **68**, 1822–1828 (1996).
111. D. Figeys, R. Aebersold, 'High Sensitivity Analysis of Proteins and Peptides by Capillary Electrophoresis–Tandem Mass Spectrometry: Recent Developments in Technology and Application', *Electrophoresis*, **19**, 885–892 (1998).

112. R.S. Ramsey, S.A. McLuckey, 'Capillary Electrophoresis/Electrospray Ionization Ion Trap Mass Spectrometry Using a Sheathless Interface', *J. Microcol. Sep.*, **7**, 461–469 (1995).
113. P. Cao, M. Moini, 'A Novel Sheathless Interface for Capillary Electrophoresis/Electrospray Ionization Mass Spectrometry Using an In-capillary Electrode', *J. Am. Soc. Mass Spectrom.*, **8**, 561–564 (1997).
114. R.S. Ramsey, J.M. Ramsey, 'Generating Electrospray from Microchip Devices Using Electroosmotic Pumping', *Anal. Chem.*, **69**, 1174–1178 (1997).
115. B.L. Karger, 'Capillary Electrophoresis: Overview and Perspective', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 1–24, 1998.
116. R.-L. Chien, 'Sample Introduction and Stacking', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 449–479, 1998.
117. N.J. Reinhold, E. Schroder, U.R. Tjaden, W.M.A. Niessen, M.C. Ten Noever De Brauw, J. van der Greef, 'Static and Scanning Array Detection in Capillary Electrophoresis–Mass Spectrometry', *J. Chromatogr.*, **516**, 147–155 (1990).
118. D.R. Goodlett, J.H. Wahl, H.R. Udseth, R.D. Smith, 'Reduced Elution Speed Detection for Capillary Electrophoresis/Mass Spectrometry', *J. Microcol. Sep.*, **5**, 57–62 (1993).
119. G.A. Valaskovic, N.L. Kelleher, F.W. McLafferty, 'Attomole Protein Characterization by Capillary Electrophoresis/Mass Spectrometry', *Science*, **273**, 1199–1202 (1996).
120. M.E. Mccomb, A.N. Krutchinsky, W. Ens, K.G. Standing, H. Perreault, 'Sensitive High Resolution Analysis of Biological Molecules by Capillary Zone Electrophoresis Coupled with Reflecting Time-of-flight Mass Spectrometry', *J. Chromatogr.*, **800**, 1–11 (1998).
121. J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, 'Use of Small-diameter Capillaries for Increasing Peptide and Protein Detection Sensitivity in Capillary Electrophoresis–Mass Spectrometry', *Electrophoresis*, **14**, 448–457 (1993).
122. Q. Tang, A.K. Harrata, C.S. Lee, 'Capillary Isoelectric Focusing–Electrospray Mass Spectrometry for Protein Analysis', *Anal. Chem.*, **67**, 3515–3519 (1995).
123. M.H. Lamoree, U.R. Tjaden, J. van der Greef, 'Use of Microdialysis for the On-line Coupling of Capillary Isoelectric Focusing with Electrospray Mass Spectrometry', *J. Chromatogr. A*, **777**, 31–39 (1997).
124. N.J. Clark, A.J. Tomlinson, S. Naylor, 'On-line Desalting of Physiologically Derived Fluids in Conjunction with Capillary Isoelectric Focusing–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **8**, 743–748 (1997).
125. N.J. Clark, C.L. Croy, A.J. Tomlinson, S. Naylor, 'Capillary Isoelectric Focusing (cIEF/MS) as an Efficient Method for the Analysis of Biologically Derived Protein Mixtures', *Anal. Biochem.*, submitted.
126. A.J. Tomlinson, L.M. Benson, N.A. Guzman, S. Naylor, 'Preconcentration and Microreaction Technology On-line with Capillary Electrophoresis', *J. Chromatogr.*, **744**, 3–15 (1996).
127. N.A. Guzman, S.S. Park, D. Schaufelberger, L. Hernandez, X. Paez, P. Rada, A.J. Tomlinson, S. Naylor, 'New Approaches in Clinical Chemistry: On-line Analyte Concentration and Microreaction Capillary Electrophoresis for the Determination of Drugs, Metabolic Intermediates and Biopolymers in Biological Fluids', *J. Chromatogr. B*, **697**, 37–66 (1997).
128. S. Naylor, Q. Ji, K.L. Johnson, A.J. Tomlinson, W.C. Kieper, S.C. Jameson, 'Enhanced Sensitivity for Sequence Determination of MHC Class I Peptides by Membrane Preconcentration Capillary Electrophoresis–Microspray–Mass Spectrometry', *Electrophoresis*, **19**, 2207–2212 (1998).
129. E. Rohde, A.J. Tomlinson, D.H. Johnson, S. Naylor, 'Protein Analysis by Membrane Preconcentration–Capillary Electrophoresis: Systematic Evaluation of Parameters Affecting Preconcentration and Separation', *J. Chromatogr. B*, **713**, 301 (1998).
130. E. Rohde, A.J. Tomlinson, D.H. Johnson, S. Naylor, 'Comparison of Protein Mixtures in Aqueous Humor by Membrane Preconcentration–Capillary Electrophoresis–Mass Spectrometry (mPC/CE/MS)', *Electrophoresis*, **19**, 2361–2379 (1998).
131. X. Ren, P.Z. Liu, M.L. Lee, 'Cellulose Modified Polypropylene Hollow Fibers for Capillary Electrophoresis', *J. Microcol. Sep.*, **8**, 529–534 (1996).
132. X. Ren, P.Z. Liu, A. Malik, M.L. Lee, 'Hydrophilic Polymer-modified Polypropylene Hollow Fibers with Controllable Electroosmotic Flow for Capillary Electrophoresis', *J. Microcol. Sep.*, **8**, 535–540 (1996).
133. A. Fridstrom, N. Lundell, B. Ekstrom, K.E. Markides, 'Dextran-modified Polypropylene Hollow Fibers for Use in Capillary Electrophoresis', *J. Microcol. Sep.*, **9**, 1–7 (1997).
134. S.A. Hofstadler, J.H. Wahl, J.E. Bruce, R.D. Smith, 'On-line Capillary Electrophoresis with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Chem Soc.*, **115**, 6983–6984 (1993).
135. J.R. Perkins, C.E. Parker, K.B. Tomer, 'The Characterization of Snake Venoms Using Capillary Electrophoresis in Conjunction with Electrospray Mass Spectrometry: Black Mambas', *Electrophoresis*, **14**, 458–468 (1993).
136. R.B. Cole, J. Varghese, R.M. McCormick, D. Kadlecek, 'Evaluation of a Novel Hydrophilic Derivatized Capillary for Protein Analysis by Capillary Electrophoresis–Electrospray Mass Spectrometry', *J. Chromatogr. A*, **680**, 363–373 (1994).

137. S.A. Hofstadler, F.D. Swanek, D.C. Gale, A.G. Ewing, R.D. Smith, 'Capillary Electrophoresis–Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Direct Analysis of Cellular Proteins', *Anal. Chem.*, **67**, 1477–1480 (1995).
138. J.F. Kelly, S.J. Locke, L. Ramaley, P. Tibault, 'Development of Electrophoretic Conditions for the Characterization of Protein Glycoforms by Capillary Electrophoresis–Electrospray Mass Spectrometry', *J. Chromatogr. A*, **720**, 409–427 (1996).
139. D. Figeys, R. Aebersold, 'Capillary Electrophoresis of Peptides and Proteins at Neutral pH in Capillaries Covalently Coated with Polyethyleneimine', *J. Chromatogr. B*, **695**, 163–168 (1997).
140. K.L. Johnson, A.J. Tomlinson, L.M. Benson, S. Naylor, 'Capillary Electrophoresis Using Nonaqueous Media Interface with Electrospray Ionization Mass Spectrometer: Analysis of Hydrophobic Peptide on Uncoated CE Capillaries', Finnigan MAT Application Note, Finnigan MAT, San Jose, CA, March 1997.
141. J.L. Miller, M.G. Khaledi, 'Nonaqueous Capillary Electrophoresis', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 525–555, 1998.
142. R. Takigiku, T. Keough, M.P. Lacey, R.E. Schneider, 'Capillary-zone Electrophoresis with Fraction Collection for Desorption Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **4**, 24–29 (1990).
143. W. Weinmann, K. Baumeister, I. Kaufmann, M. Przybylski, 'Structural Characterization of Polypeptides and Proteins by Combination of Capillary Electrophoresis and ^{252}Cf Plasma Desorption Mass Spectrometry', *J. Chromatogr.*, **628**, 111–121 (1993).
144. W. Weinmann, C.E. Parker, K. Baumeister, C. Maier, K.B. Tomer, M. Przybylski, 'Capillary Electrophoresis Combined with ^{252}Cf Plasma Desorption and Electrospray Mass Spectrometry for the Structural Characterization of Hydrophobic Polypeptides Using Organic Solvents', *Electrophoresis*, **15**, 228–233 (1994).
145. K.L.Walker, R.W. Chiu, C.A. Monnig, C.L. Wilkins, 'Off-line Coupling of Capillary Electrophoresis and Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **67**, 4197–4204 (1995).
146. H.J. Hollis, M.F. Rohde, R.S. Rush, 'Multiple Sequential Fraction Collection of Peptides and Glycopeptides by High Performance Capillary Electrophoresis', *Anal. Biochem.*, **230**, 123–129 (1995).
147. N.H.H. Heegaard, P. Roepstorff, 'Preparative Capillary Electrophoresis and Mass Spectrometry for the Identification of a Putative Heparin-binding Site in Amyloid P Component', *J. Capillary Electrophor.*, **2**, 219–223 (1995).
148. L. Cruz, R.W. Garden, H.J. Kaiser, J.V. Sweedler, 'Studies of the Degradation Products of Nisin, a Peptide Antibiotic, Using Capillary Electrophoresis with Off-line Mass Spectrometry', *J. Chromatogr. A*, **735**, 375–386 (1996).
149. H. Zhang, R.M. Caprioli, 'Capillary Electrophoresis Combined with Matrix-assisted Laser Desorption/Ionization Mass Spectrometry; Continuous Sample Deposition on a Matrix-precoated Membrane Target', *J. Mass Spectrom.*, **31**, 1039–1046 (1996).
150. A.-C. Bergman, T. Bergman, 'Micropreparation of Peptides by Capillary Electrophoresis for Matrix Assisted Laser Desorption Mass Spectrometry', *FEBS Lett.*, **397**, 45–49 (1996).
151. D. Belder, H. Husmann, J. Schoppenthau, G. Paulus, 'CE/MALDI/MS of Peptides and Low Molecular Weight Drugs Using Non-volatile Buffers', *J. High Resolut. Chromatogr.*, **21**, 59–62 (1998).
152. M.D. Bauer, Y. Sun, F. Wang, 'Nano-electrospray Mass Spectrometry and Edman Sequencing of Peptides and Proteins Collected From Capillary Electrophoresis', in *Techniques in Protein Chemistry VIII*, ed. D.R. Marshall, Academic Press, San Diego, CA, 37–46, 1997.
153. K.B. Tomer, L.J. Deterding, C.E. Parker, 'Capillary Electrophoresis Coupled with Mass Spectrometry', in *Capillary Electrophoresis Technology*, ed. N.A. Guzman, Marcel Dekker, New York, 53–99, 1993.
154. R.D. Smith, D.R. Goodlett, J.H. Wahl, 'Capillary Electrophoresis–Mass Spectrometry', in *Handbook of Capillary Electrophoresis: A Practical Approach*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 185–206, 1994.
155. K.B. Tomer, C.E. Parker, L.J. Deterding, 'Capillary Electrophoresis Interfaced with Mass Spectrometry: Electrospray Ionization and Continuous Flow Fast Atom Bombardment', in *Capillary Electrophoresis: An Analytical Tool in Biotechnology*, ed. P.G. Righetti, CRC Press, Boca Raton, FL, 121–153, 1995.
156. J.D. Henion, 'Current Status of CE/MS Coupling', presented at the ASMS Fall Workshop, Alexandria, VA, 29–57, 1995.
157. R.D. Smith, H.R. Udseth, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, 'Capillary Electrophoresis–Mass Spectrometry', *Methods Enzymol.*, **271**, 448–486 (1996).
158. K.B. Tomer, C.E. Parker, L.J. Deterding, 'High Performance Capillary Electrophoresis–Mass Spectrometry', in *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, ed. M.G. Khaledi, ACS, Washington, DC, ACS Monograph Series, Vol. 146, 405–446, 1998.
159. F. Foret, T.J. Thompson, P. Vouros, B.L. Karger, 'Liquid Sheath Effects on the Separation of Proteins in Capillary Electrophoresis/Electrospray Mass Spectrometry', *Anal. Chem.*, **66**, 4450–4458 (1994).
160. F. Garcia, J.D. Henion, 'Gel-filled Capillary Electrophoresis/Mass Spectrometry Using a Liquid Junction–Ion Spray Interface', *Anal. Chem.*, **64**, 985–990 (1992).

161. M.S. Kriger, K.D. Cook, R.S. Rasmey, 'Durable Gold-coated Fused-silica Capillaries for Use in Electrospray Mass Spectrometry', *Anal. Chem.*, **67**, 385–389 (1995).
162. J. Severs, A.C. Harms, R.D. Smith, 'A New High Performance Interface for Capillary Electrophoresis/Electrospray Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 1175–1178 (1996).
163. K.C. Lewis, G.J. Opiteck, J.W. Jorgenson, D.M. Sheeley, 'Comprehensive On-line RPLC/CZE/MS of Peptides', *J. Am. Soc. Mass Spectrom.*, **8**, 495–500 (1997).
164. L. Licklider, W.G. Kuhr, M.P. Lacy, T. Keough, M.P. Purdon, R. Takigiku, 'On-line Microreactors/Capillary Electrophoresis/Mass Spectrometry for the Analysis of Proteins and Peptides', *Anal. Chem.*, **67**, 4170–4177 (1995).
165. A.J. Tomlinson, S. Naylor, unpublished results.
166. K.L. Johnson, A.J. Tomlinson, S. Naylor, 'Capillary Conditioning and Electrospray Ionization for Optimal Capillary Electrophoresis/Mass Spectrometry Performance', *Rapid Commun. Mass Spectrom.*, **10**, 1159–1160 (1996).
167. R.M. McCormick, 'Capillary Zone Electrophoresis of Peptides', in *Handbook of Capillary Electrophoresis: A Practical Approach*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 287–323, 1994.
168. E. Rohde, A.J. Tomlinson, D.H. Johnson, S. Naylor, 'Evaluation of Capillary Coating for the Capillary Electrophoresis Separation of Proteins in Aqueous Humor', *Biomed. Chromatogr.*, submitted.
169. D.F. Hunt, J.R. Yates III, J. Shabanowitz, S. Winston, C.R. Hauer, 'Protein Sequencing by Tandem Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **83**, 6233–6237 (1986).
170. A.L. Cox, J. Skipper, R.A. Henderson, T.L. Darrow, J. Shabanowitz, V.H. Engelhard, D.F. Hunt, C.L. Slingluff, 'Identification of a Peptide Recognized by Five Melanoma-specific Human Cytotoxic T Cell Lines', *Science*, **264**, 716 (1994).
171. K.A. Hoguist, A.J. Tomlinson, W.C. Kieper, M.A. McGargill, M.C. Hart, S. Naylor, S.C. Jameson, 'Identification of a Naturally Occurring Ligand for Thymic Positive Selection', *Immunity*, **6**, 389–399 (1997).
172. S.A. Hofstadler, J.H. Wahl, R. Bakhtiar, G.A. Anderson, J.E. Bruce, R.D. Smith, 'Capillary Electrophoresis Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Sustained Off-resonance Irradiation for the Characterization of Protein and Peptide Mixtures', *J. Am. Soc. Mass Spectrom*, **5**, 894–899 (1995).
173. Q.F. Xue, F. Foret, Y.M. Dunayevskiy, P.M. Zavracky, N.E. McGruer, B.L. Karger, 'Multichannel Microchip Electrospray Mass Spectrometry', *Anal. Chem.*, **69**, 426–430 (1997).
174. F. Foret, B.L. Karger, 'Microfabricated Devices for Rapid Capillary Electrophoresis–Electrospray–Mass Spectrometry', presented at the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, June 1–4, 1998.
175. J.-T. Wu, L. He, M.X. Li, S. Parus, D.M. Lubman, 'On-line Capillary Separations/Tandem Mass Spectrometry for Protein Digest Analysis by Using an Ion Trap Storage/Reflectron Time-of-flight Mass Detector', *J. Am. Soc. Mass Spectrom.*, **8**, 1237–1246 (1997).
176. J.-T. Wu, M.G. Qian, M.X. Li, K. Zheng, P.Huang, D.M. Lubman, 'On-line Analysis by Capillary Separations Interfaced to an Ion Trap Storage/Reflectron Time-of-flight Mass Spectrometer', *J. Chromatogr. A*, **794**, 377–389 (1998).
177. A. Shevchenko, I. Chernushevich, W. Ens, K.G. Standing, B. Thomson, M. Wilm, M. Mann, 'Rapid 'De Novo' Peptide Sequencing by a Combination of Nanoelectrospray, Isotopic Labeling and a Quadrupole/Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **11**, 1015–1024 (1997).

Chromatography of Membrane Proteins and Lipoproteins

Lello Zolla
University of Tuscia, Viterbo, Italy

1 Introduction **1**
2 Chromatography of Membrane Proteins **1**
2.1 Methods for Protein Separation and Characterization 2
2.2 Separation Depending on the Membrane Protein Category 4
3 Chromatography of Lipoproteins **5**
3.1 Methods for Lipoprotein Separation 5
4 Application Examples **6**
4.1 Experimental Considerations 7
4.2 Membrane Proteins 7
4.3 Separation of Lipoproteins 15
5 Current Trends **18**
5.1 Membrane Proteins 18
5.2 Lipoproteins 20
6 Comparison with Other Methods **20**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **23**

The available methods for the separation of membrane proteins and lipoproteins are sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE), followed by immunoblotting, isoelectric focusing (IEF), capillary electrophoresis (CE) and high-performance liquid chromatography (HPLC). In this article it is shown that HPLC techniques, given their wide versatility, relative ease of use, and high resolution, may be considered the most valuable tool for the characterization of virtually any hydrophobic protein. Moreover, HPLC is not a destructive technique and therefore proteins, once separated, are available for further analytical investigations. Application examples are described and comparisons with other methods are discussed.

1 INTRODUCTION

Owing to their lipophilic character, the solubilization and separation of membrane proteins and lipoproteins normally require the use of detergents. Consequently, classical protein purification strategies, designed for water-soluble proteins, are of limited value and the number of reports dealing with the separation and characterization of such proteins is limited. Moreover, their hydrophobic nature induces self-association into noncovalent multimers, and therefore all separative methods require a prior procedure such as tedious sequential gradient ultracentrifugation for sample preparation, with the risk of affecting the results. Although the available methods for the final separation of these hydrophobic proteins are numerous, theirs high hydrophobicity and the presence of detergent makes most of these available methods expensive and technically demanding. The traditional approaches by SDS/PAGE are not only cumbersome but also rather ineffective for evaluating differences in small molecular masses and the run times required are very long, typically more than 20–30 h. CE is emerging as a powerful separation technique for lipoproteins but the number of reports on membrane proteins is still limited. It is not surprising that HPLC techniques, given their wide versatility, relative ease of use, and high resolution, may be considered the most valuable tool for the characterization of virtually any hydrophobic protein. Moreover, HPLC is not a destructive technique and therefore proteins, once separated, are available for other analytical investigations, such as by SDS/PAGE. In any case, before loading a sample on any HPLC system, a preseparation of membrane proteins or lipoproteins, according to their hydrophobic characteristics, must be achieved by selective extractions. The pretreatment of these proteins facilitates subsequent separation and provides a first guideline for the choice of detergent. It is a general rule, in fact, that when choosing the detergent for the running buffers, one should use the one with which the protein was solubilized, if possible. It is also advisable to use less denaturing detergents and nonionic detergents for solubilization and likewise as an additive to the running buffers, in order to retain the biological activity of proteins. The presence of detergent in all steps preserves the subsequent tendency toward association and aggregation and the possibility of nonspecific interactions with the support used for chromatographic separation.

Although lipoproteins and membrane proteins show common features, specific protocols for the separation of each group of proteins have recently been proposed, and therefore they will be presented separately.

2 CHROMATOGRAPHY OF MEMBRANE PROTEINS

Methods for the separation and characterization of membrane proteins differ depending on the type of interaction

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of the protein with the membrane: proteins embedded in a lipid bilayer or associated with membrane structures. In the latter case, usually ionic interactions or hydrogen bonds are involved in the anchorage, and therefore the protein may be removed by gentle solubilization with buffered salt solution and then analyzed. Dilute sodium hydroxide and sodium hydrogencarbonate (between pH 10 and 12), concentrated salt solutions, and complex-forming substances such as ethylenediaminetetraacetic acid (EDTA) and ethylene glycol bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA) are used satisfactorily for their solubilization. In the case of embedded proteins, they must first be extracted from the membrane structures and then isolated. Repeated freezing and thawing can be applied to dissolve the structures mechanically, thereby allowing most membrane-associated proteins to be removed. However, solubilization of the membrane is the most suitable approach. Chaotropic reagents such as urea and guanidine hydrochloride are now used less frequently, while solubilization using detergents is still in principle more practical. It is also advisable to use less denaturing detergents such as 3-[(3-cholamidopropyl)dimethylammonio]-l-propanesulfonate (CHAPS), or nonionic detergents such as n-octyl β-D-glucopyranoside (OD) or *n*-dodecyl β-D-maltoside (β-DM) for solubilization.

Through the use of various reagents in different steps, the membrane proteins can be prepared according to their respective solubility and hydrophobic characteristics. The effort necessary for solubilization of the proteins increases with growing complexity of the membrane structure. In section 4.2.3 is reported a complete solubilization scheme for the thylakoid membrane of chloroplasts which contains the photosynthetic apparatus consisting of a large number of proteins, namely 40 different proteins assembled into two main complexes: photosystem I (PSI) and photosystem II (PSII). It will be shown that by selective extraction, a preseparation of membrane proteins according to their hydrophobic characteristics is achieved. This in turn allows further separation by use of different, mainly chromatographic and electrophoretic, methods.

Once the membrane proteins have been pretreated, several chromatographic methods may be used.

2.1 Methods for Protein Separation and Characterization

With regard to the membrane protein separation, strategies having general validity are reversed-phase high-performance liquid chromatography (RPHPLC), affinity chromatography, size-exclusion high-performance liquid chromatography (SEHPLC) or ion-exchange HPLC and high-performance membrane chromatography. Usually membrane proteins are organized as multimeric units, therefore no single technique is likely to be sufficient for the complete characterization of different proteins. It is important, therefore, to select a battery of complementary techniques that will ensure that the necessary separation is obtained. Specific considerations for selection of an appropriate battery of techniques are described in section 4.

Conventional gel permeation and ion-exchange chromatography on soft gel supports, although time-consuming, are used for the preparation of large quantities (10–50 mg) of pure proteins, while the preparation of smaller quantities is usually performed at low pressure [fast protein liquid chromatography (FPLC)]. HPLC, usually carried out at high pressure (300–400 psi), has been an essential tool for the quantitative and qualitative analysis of proteins. Its success increased when rigid supports, which are able to withstand high pressures and show chemical stability within the working range of biological separations, were introduced. The separation time decreased dramatically.

The first consideration in defining an HPLC method development strategy is to establish if we need proteins with the biological properties preserved or we simply need their characterization. Proteins vary widely in their physicochemical properties and these properties can have a great impact on the conditions necessary for HPLC analysis. Solution stability can differ widely from one protein to another, and for a given protein the solution stability can be highly dependent on pH, temperature, presence of denaturants or detergents, and other factors. Such factors must be explored as part of the separation development process in order to select appropriate conditions and an appropriate detergent under which artifactual denaturation is minimized. Other physicochemical properties to be considered include the presence or absence of free thiol groups or bound metals and the effect of the organic solvent used in RPHPLC on the secondary or tertiary structure. Finally, the presence of detergent must be taken into account along with the possibility of nonspecific interactions with the support used for chromatographic separation. Furthermore, membrane proteins readily self-associate into noncovalent multimers. Hence in some instances one may wish to determine only covalent related substances and it is therefore appropriate to conduct the HPLC separation using conditions that dissociate the noncovalent multimers, whereas in other instances measurement of the extent of self-association may be an objective of the separation.

2.1.1 Reversed-phase High-performance Liquid Chromatography

RPHPLC offers the best resolution of all chromatographic methods, although up to now it has been rarely

used for the separation of membrane proteins. This is because under the separation conditions usually required in RPHPLC [elution with acetonitrile or 2-propanol gradient, in the presence of 0.1% (v/v) trifluoroacetic acid (TFA)], the membrane proteins can be recovered only partly and sometimes not at all. The advent of supports having reduced hydrophobicity, such as a C-4 column, has allowed the use of milder elution conditions with a consequent higher recovery of proteins (see section 4). Clearly, the application of organic solvents and acids causes the loss of the protein biological activity. Thus RPHPLC is a powerful method for the isolation of membrane proteins, whose primary structure is to be investigated subsequently or whose identification is the main aim.

2.1.2 High-performance Affinity Chromatography

Analytical affinity provides a powerful means to purify proteins and other biomolecules with a basic two-step retention–chaotropic elution procedure with minimal nonspecific interactions and their subsequent elution in a highly purified state and in the native state. Successes of affinity chromatography for preparative purposes arose when researchers improved several key features of immobilized ligand interactions with eluting macromolecules, namely accessibility of immobilized ligand, selectivity of ligand interaction with soluble macromolecule, and reversibility of macromolecule binding, which allows their elution without denaturation. In this regard, important parameters for a successful affinity sorbent for bioseparations include mechanical, chemical and biological stability, and also the potential for nonspecific binding. Rigid matrices such as polymeric supports, silica and controlled-pore glass may suffer from nonspecific binding and low recovery. Cross-linked beaded agarose or cellulose offers a good compromise between mechanical stability and nonspecific binding. These matrices also have good chemical stability within the working range of biological separations and the product recovery is excellent. The interested reader is referred to other literature for a discussion of core chromatographic matrices themselves.[1]

Once a matrix has been selected, the ligand is immobilized via a stable covalent bond to avoid progressive leaching and consequent capacity loss. The most common procedure to activate agarose is cyanogen bromide (CNBr), which results in the coupling of an amine group of the ligand or spacer through an isourea bond. Matrices prepared by this procedure can suffer from the serious drawback of high ligand leaching, while more stable linkages are produced by organic sulfonyl chlorides or epihalohydrins.[2]

Lectin affinity chromatography is largely used for lipoproteins, while two steps with protein-specific ligands, namely immunoaffinity (IA) and transition-state ligands, are commonly used for the separation of receptors and enzyme proteins, respectively.[3]

2.1.3 Size-exclusion High-performance Liquid Chromatography

The separation of membrane proteins may be achieved satisfactorily under nondenaturing conditions by SEHPLC. In this case, the size of the molecules or multimeric complexes is the basis of separation, although differences of 5–10 kDa in molecular weight are not sufficient for a significant separation. Since nonspecific interactions between the sample and support, and among different sample components, must be suppressed, sometimes the addition of denaturing agents, such as chaotropic reagents or sodium dodecyl sulfate (SDS), is required. Obviously, under these denaturing conditions, the resolution and yield are optimized in SEHPLC, along with good reproducibility of the results, but proteins may be denatured.

2.1.4 Ion-exchange High-performance Liquid Chromatography

Ion-exchange HPLC is based on the different ionic interactions between proteins and anionic or cationic charges of the support. Unfortunately, in this type of chromatography, protein interaction with the support is usually strong and the conditions used for elution are often denaturing. The advent of supports based on agarose, which owing to its hydrophilic characteristic has a lower level of nonspecific interactions with the sample, has renewed the use of ion-exchange HPLC for protein separations. However, in the particular case of membrane proteins, although the degree of resolution of ion-exchange HPLC is surpassed only by RPHPLC, interaction with the support is still too strong and other methods must be chosen, or detergents must be added to the separation buffers.

2.1.5 Hydrophobic Interaction High-performance Liquid Chromatography

Hydrophobic interaction HPLC can also be used for the isolation of membrane proteins, even if they are so strong that the sample can no longer be recovered from the column, even with the use of detergents or, in extreme cases, organic solvents.

2.1.6 High-performance Membrane Chromatography

Red cells, biomembrane vesicles, proteoliposomes, and liposomes noncovalently immobilized in gel particles or beads have been used as stationary phases

for biomembrane affinity analyses and ion-exchange chromatographic separations. Lipid monolayers coupled to silica beads have been utilized for membrane protein purification in detergent solution. Proteins are adsorbed on the liposome surfaces and subsequently separated by salt gradient elution on charged liposomes formed and entrapped in gel beads upon detergent depletion by dialysis.[4] Hence both protein size and charge affect the separation and the elution time. These techniques, in the particular case of membrane proteins, have considerable potential because of their high resolution, short running times, and low nonspecific interactions. For more details the reader is referred to a review.[5]

2.2 Separation Depending on the Membrane Protein Category

From a didactic point of view, proteins embedded in membranes may be divided on the basis of their biological role into receptors, glycosylated proteins, channel or carrier proteins, and structural proteins or those without a particular function.

The first consideration in defining the development strategy of an HPLC method is to establish whether the main interest is to isolate the protein functionally active or its characterization. In the former case nondenaturanting conditions must be used, whereas in the latter case any type of detergent or organic eluent may be used for the optimum separation.

2.2.1 Cellular Receptors

In the case of receptorial proteins, they rarely exist in aggregated form and they are specific for recognizing a particular ligand. Thus nondenaturating size-exclusion chromatography (SEC) may not be used, and affinity chromatography is more indicated, because immobilized ligands proteins can be bound specifically and eluted selectively. The natural ligand might be an ideal choice for binding selectivity but may suffer from a high intrinsic binding affinity, which could lead to inactivation of the receptor or ligand due to the harsh conditions needed for release from the affinity support. Lectin affinity chromatography is largely used for glycoproteins and IA chromatography is the most widely used method for receptors.

A more practical ligand may be a monoclonal antibody (MAb) to the receptor, which can be preselected for modest affinity and appropriate binding kinetics (on and off rates). In general, to prepare an IA sorbent, MAb is preferred over polyclonal antibody, since the MAb can be obtained once a hybridoma clone has been isolated reproducibly, and second because the appropriate MAb can be selected with the desired binding properties to optimize biomolecular adsorption and elution. The desired elution conditions can be incorporated into the screening procedure to identify the most advantageous MAb. Since the binding constant varies from clone to clone, selection of a clone producing a MAb with a desirable binding constant is necessary. An antibody that gives a good response in either Western blot or enzyme-linked immunosorbent assay (ELISA) is not necessarily the best ligand for the IA sorbent. An example of this procedure is described in section 4.2.2. Interaction analysis using such current tools as optical biosensors can be used to screen MAbs for those with a good balance of sufficiently high affinity and finite off rate (see section 5). Many receptors have been separated by IA.[6] A milestone in this field is the separation of transferrin receptor from plasma membranes of various mammalian cells.[7] It is known that this protein binds with high affinity to immobilized diferric transferrin, whereas its affinity for apotransferrin is low. Consequently, the complex of transferrin and transferrin receptor can be dissociated by chelation of ferric ions after the addition of chelating reagents. In this way transferrin receptor can be eluted from the column under mild conditions.

2.2.2 Carriers or Channel Proteins Translocating Ions

Membrane proteins with the function of carriers or containing channels translocating ions across membranes may be separated and identified by the use of the perfusion planar lipid membrane (BLM) technique coupled with an HPLC system. This technique has been demonstrated to be efficient for the fast identification, isolation and characterization of transport proteins (porins) in the outer membrane of bacteria and other organisms,[8] but it seems to be of general validity.[9] The method is based on solubilization of the membrane, separation of solubilized membrane proteins into fractions (100–150) eluted from an HPLC column followed by immediate screening on the BLM for channel-forming activity, allowing precise localization of the porin-containing protein peak. The principle of the BLM method is simple. A planar lipid membrane interposed between two electrodes is predictably dull, showing little permeability to ions, no voltage dependence, and no interesting transport behavior. However, the introduction of a channel-forming protein, previously separated by HPLC, that is spontaneously incorporated into the bilayer dramatically alters the situation, allowing chemical substances to cross the lipid bilayer and giving rise to an ionic current. Furthermore, the opening and closing of this single ionic channel can be easily detected and modification of the channel properties by voltage, pH, ionic composition, blockers, mutation and chemical reagents can also be quantified. Although the BLM is an in vitro system, its validity has been demonstrated repeatedly through the more recent technique

of patch clamping, giving confidence that what is measured using a BLM corresponds qualitatively and often quantitatively to what is observed in whole cells. BLMs have been intensively used in the identification and reconstruction of channel-forming proteins,[10] to determine the deposition of amyloid proteins observed in Alzheimer's disease pathology,[11] and in the study of membrane proteins of other disease-related bacteria, whose porins or pathogenic toxins might be labile.[9,10]

2.2.3 Glycoproteins

Determination of the carbohydrate composition, type, and branching pattern is an important step in understanding the biological function of glycoproteins and in the development of a recombinant DNA-derived glycoprotein as a pharmaceutical. Unfortunately, glycoproteins exist in a variety of biologically active forms, owing to the nature of the diversity of monosaccharides and the variety of possible linkages. It may be estimated that the linkage possibilities of a hexasaccharide yield a possible 4.76×10^9 structures.[12]

The complexity of carbohydrate structures mandates that a variety of analytical methods be used for the study of these forms.

Lectin affinity chromatography is used to separate membrane glycoproteins, but a promising new procedure is described in section 4 that uses HPLC in conjunction with electrospray ionization mass spectrometry (ESIMS) as a tool to identify the sites of glycosylation and the general nature of the glycosylation. ESIMS can detect whether an oligosaccharide is O- or N-linked and can also differentiate between complex, high-mannose, and hybrid forms.

2.2.4 Others

When the membrane proteins that are to be analyzed do not play any functional role and their identification and characterization are the main aim, RPHPLC offers the best resolution of all chromatographic methods, especially when coupled on-line with a mass spectrometer. In sections 4 and 5 some aspects of this application will be presented and discussed.

3 CHROMATOGRAPHY OF LIPOPROTEINS

The term lipoprotein refers to particles which are heterogeneous with respect to size, hydrated density, and composition. The main function of the lipoprotein system is to transport lipids to surrounding tissues. All are made up of cholesterol, cholesterol esters, triglycerides, phospholipids, and apoproteins in varying proportions. Lipoproteins have a globular structure containing hydrophobic lipids such as triglycerides and cholesterol esters within the interior core, and a periphery containing more polar lipids, phospholipids, cholesterol, and proteins.

Owing to their strong heterogeneity, classification of lipoproteins remains undefined. Hydrated density is the more common form of lipoprotein classification. Thus lipoproteins are separated into very low-density lipoprotein (VLDL), low-density lipoprotein (LDL) and high-density lipoprotein (HDL) fractions. In this case, lipoproteins are viewed as particles which are heterogeneous with respect to their physical properties but homogeneous with respect to apoliprotein composition. The HDL, LDL and VLDL groups show different functions, which results in their having different properties with respect to atherosclerosis, and therefore it is very important to have methods available for their separation.[13] On the other hand, due to the growing evidence that apolipoproteins are better markers of coronary heart disease (CHD) than are serum cholesterol levels,[14] and that the apoliprotein distribution is important in diagnosing lipoprotein abnormalities, the separation and characterization of lipoproteins with respect to protein composition is of increasing interest. In Table 1 are summarized the physical and chemical properties of the main lipoprotein classes and their subfractions and also the different types of apoproteins and their isomers.

3.1 Methods for Lipoprotein Separation

As an alternative to ultracentrifuge separation, gel permeation separation does not require expensive ultracentrifugation instrumentation and has the advantage of being relatively mild and nondestructive in the separation

Table 1 Physical and chemical characteristics of lipoproteins: subfraction classes and apolipoproteins

Lipoprotein class	Diameter (nm)	Subfractions	Major apolipoprotein	Apolipoprotein isomers
VLDL	30–70		ApoC, apoB, apoE	ApoC-I, C-II and C-$III_{(0,1,2)}$ ApoE-1, E-2, E-3, E-4, E-5 ApoD
LDL	17–26	5	ApoB	Apo-B_{100} Apo-B_{48}
HDL	7–9	5	ApoA-I, apoA-II, apoE	ApoA-$I_{1,2,3,4}$ ApoA-II and A-IV

of lipoprotein particles. Gel permeation and affinity chromatography are used satisfactorily for preparative purposes and to separate lipoproteins according to size and apolipoprotein content, respectively. However, these chromatographic methods have not gained wide acceptance as routine techniques, requiring long analysis times and excessive dilution of samples during separation. In this respect, the development of rigid supports for gel permeation has led to large improvements in speed and resolution. Gel permeation chromatography (GPC) of whole plasma is now used primarily for analytical rather than preparative separation of plasma lipoproteins.

The advent of RPHPLC resulted in renewed interest in the chromatographic separation of lipoproteins, offering the best performance in terms of speed and resolution of structural variants. In particular, the use of rigid supports, which are able to withstand high pressures, decreased the separation time dramatically to 1 h compared with 20 h by the conventional gel permeation method. Furthermore, the separation may be easily automated, allowing unattended analysis of multiple samples. Automation also allows on-line monitoring of cholesterol in the separated LDL and HDL fractions (see later).

Two new emergent chromatographic techniques for lipoprotein separations are hydroxyapatite and countercurrent chromatography (CCC).[15] Their combined use allows a better separation and analysis of HDLs, LDLs and VLDLs from human serum without prior ultracentrifugation (see section 5).

3.1.1 Separation by Particle Size and Subfractions

Gel permeation is normally used to separate lipoproteins on the basis of differences in particle size. It was initially performed on cross-linked dextran supports, which, owing to their relatively small pore size, yielded poor separations, whereas by agarose gel, which has a larger pore size, has been found to yield better separations.[16] Usually the eluting lipoproteins are derivatized postcolumn with an enzymatic cholesterol reagent, and therefore the mixture is passed through a reaction capillary and detected at 500 nm. Complete separation is usually achieved in less than 80 min.[17]

Separation of the main classes of lipoproteins into their subfractions is of growing interest since each subfraction has a different specific physiological function. By GPC, HDL particles may be divided into five subfractions,[18] whereas by heparin affinity chromatography, HDL may be subfractionated according to the apoE content.[19,20] By ion-exchange chromatography, five different fractions of oxidatively modified LDL were identified.[21] The human serum HDL fraction may be divided into five or six subclasses by using hydroxyapatite chromatography.[15]

3.1.2 Separation of Apolipoprotein and Isomers

Affinity chromatography, which employs specific and reversible interactions between lipoprotein and bound ligands, is capable of separating lipoproteins on the basis of their apolipoprotein content. Affinity chromatography of lipoproteins can be performed using group-specific or biospecific ligands. Group-specific ligands include concanavalin A, which binds to apoB-containing lipoproteins,[22] and heparin, which interacts specifically with apoB and apoE,[23] while antibodies to specific apolipoproteins serve as biospecific ligands in IA chromatography. This technique provides the highest specificity for the separation and isolation of lipoproteins on the basis of their apolipoprotein content, while group-specific ligands are more economically convenient but the separation is less accurate.

In addition to affinity chromatography, the high-performance gel permeation, ion-exchange or reversed-phase mode may be used to separate apolipoproteins into their isomers. Unfortunally, in these HPLC systems only small amounts of sample (<5 mg) can be applied to the column. However, the speed (<60 min) and ease of operation make these techniques ideal for the analytical separation and characterization of apolipoproteins. GPC is often used as a preparative technique for isolating apolipoproteins derived from HDL or VLDL, while separation of isoforms can be achieved by either ion-exchange or RPHPLC. Apolipoproteins from HDL and VLDL can be separated in less than 60 min on high-performance TSK gel permeation columns.[24] Separations of apolipoproteins by ion-exchange chromatography are based on differences in charge. Therefore, this mode of chromatography can separate apolipoproteins with different isoelectric points.[21] In contrast, differences in hydrophobicity are the basis of separation by RPHPLC. The reversed-phase columns usually have a lower sample capacity (<100 mg) than those used in gel permeation HPLC. Moreover, the column lifetime is relatively short, owing to the instability of silica-based stationary phases at pH extremes. However, there is growing interest in the use of RPHPLC as a rapid and highly efficient technique for monitoring heterogenicity in apolipoprotein structure.[25]

4 APPLICATION EXAMPLES

As mentioned above, the separation of membrane proteins and lipoproteins will be presented separately owing to the specificity of the methods set up for the two main protein categories. However, in both cases HPLC is the most convenient method used, and therefore a few considerations specifically relevant to the instrumentation required are presented.

4.1 Experimental Considerations

In the case of proteins, more than for small organic compounds, HPLC requires extremely precise control of the solvent composition, because the RPHPLC elution time is much more dependent on organic solvent composition. Single-pump, low-pressure mixing HPLC gradient systems are widely used for protein analysis. However, when using solvents that are difficult to mix (e.g. 1-propanol), dual-pump, high-pressure mixing is advantageous because the two solvent streams are delivered continuously rather than delivered in a segmented fashion. Moreover, with single-pump operation the continuous sparging of solvents with helium, for degassing purposes, results in changes to the organic solvent composition. This problem may be minimized by presaturation of the helium with organic vapor by passing the sparge gas through a solvent reservoir containing the mobile phase.

With respect to detection, generally either 280- or 214-nm wavelengths are employed for protein analysis. Detection at 214 nm is more sensitive and more generally applicable because absorption at this wavelength is due to the peptide backbone. However, some HPLC mobile-phase constituents may interfere at this wavelength (e.g. acetate salts). Detection at 280 nm is somewhat less sensitive, but can be used to detect most proteins (absorption at 280 nm is primarily due to the aromatic amino acid side chains of tryptophan, phenylalanine and tyrosine). In some cases, dual-wavelength or diode-array detectors are recommended in order to provide more detailed spectral information for each peak observed.

In general, the sample solution should be similar in composition (e.g. pH) to the HPLC mobile phase so as to minimize the effect of the sample injection on the separation efficiency. One exception to this rule is RPHPLC, where the organic solvent composition of the sample should be less than that of the initial mobile phase in order to focus the injected proteins at the head of the column. Clearly, the protein must have adequate stability in the selected solvent and in many cases stability must be enhanced by use of a refrigerated autoinjector. When conducting size-exclusion HPLC for determining noncovalent multimers, one must avoid the use of denaturants (e.g. organic solvents) or extreme pH conditions that might disturb the equilibrium between monomer and multimer forms.

Clearly the column is the heart of the HPLC separation system. Unfortunately, this is the one component over which the analyst has the least direct control. Consequently, it is important to work closely with the column manufacturer to obtain several independent column manufacturing batches, and to investigate whether a particular method performs consistently on a number of such batches. Whenever possible, an HPLC method to be used on a routine basis should be validated using two alternative sources (brands) of columns, in order to minimize the impact of column manufacture problems.

The advent of new reversed phases such as C-4 allowed excellent separations of hydrophobic proteins with a high protein recovery (see later). The use of microbore HPLC and packed capillary HPLC for characterizing biosynthetic proteins is becoming more widespread. However, such techniques have not been widely adopted for routine use or method development within the laboratory owing to limitations on hardware and the relatively large amounts of assay material available. As hardware advances occur, this situation may change because the advantage of lower solvent consumption is clearly attractive. Furthermore, such techniques are more compatible with high-performance liquid chromatography/mass spectrometry (HPLC/MS) systems, which are now widely available and offer profound advantages for method development (see section 5).

4.2 Membrane Proteins

Here three examples are presented.

4.2.1 Cellular Receptor

CD4 is an integral membrane glycoprotein of T cells that acts as the cellular receptor for human immunodeficiency virus (HIV). Many protocols have been set up to recognize and separate CD4. All are based on IA, using monoclonal antibodies and autoantibody[26] directed against conformational epitopes of CD4.

4.2.1.1 Selecting a Monoclonal Antibody Soluble CD4 (sCD4) has been considered as a possible therapeutic agent for acquired immune deficiency syndrome (AIDS) by acting as a molecular decoy, i.e. by binding to the gpl20 coat protein of HIV and thereby preventing cellular binding of HIV. Consequently, there is a growing interest in searching CD4 variants from either mammalian cell culture or microbial extracts, showing improved pharmacokinetic properties. A rapid generic purification scheme for sCD4 constructs has been developed using IA separation. To prepare a robust IA sorbent for the purification of sCD4, a number of sCD4 mutants and a series of MAbs were examined.[3] Five different commercial anti-CD4 monoclonal antibodies were immobilized through protein amino groups on to a Sepharose matrix containing an 11-atom spacer using active ester chemistry and all of the sorbents were evaluated individually on a small test column. Figure 1 shows representative chromatograms to illustrate that different antibodies bind CD4 differently.

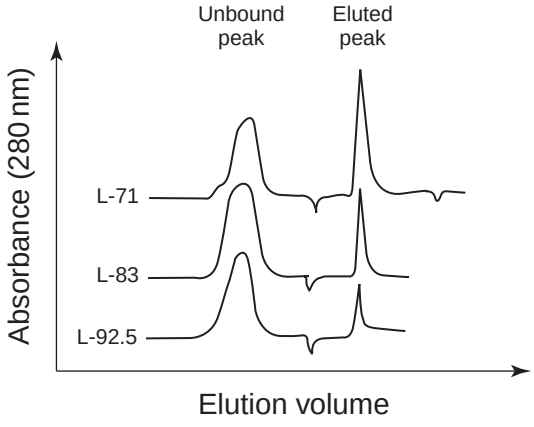

Ligand	L-71	L-83	L-92.5
Load (μg)	1400	1400	1400
Unbound	1062	1306	820
Eluted	336	55	61

Figure 1 Absorbance (A_{280}) profiles of sCD4 binding to IA column. Soluble CD4 (sCD4), at 0.1 mg mL^{-1} in *N*-(2-hydroxyethyl)piperazine-*N*-(2-ethanesulfonic acid) (HEPES) buffer at pH 7.5, was loaded at 60 cm h^{-1} at 4 °C on to a 1 × 6.4 cm IA column, preequilibrated with 0.05 M HEPES, 0.15 M NaCl, 0.01% PEG 3400 (pH 7.5). To remove nonspecifically bound proteins, the column buffer was washed with two column volumes of HEPES containing 0.5 M NaCl at the same flow rate. Bound sCD4 was eluted with 0.1 M acetic acid, 0.15 M NaCl, 0.01% PEG 3400 at 60 cm h^{-1}. The inset shows the quantities of sCD4 in various fractions determined by Bio-Rad protein determination. (Reproduced by permission of Elsevier Science from C. Jones, A. Patel, S. Griffin, J. Martin, P. Young, K. O'Donnell, C. Silverman, T. Porter, I. Chaiken, *J. Chromatogr. A*, **707**, 3–22 (1995).)

In the cases of L-92.5 and L-83 clones, the binding was restrictively tight, whereas L-71, L-77, and L-104.5 showed moderate and more tractable binding affinity. Although L-92.5 sorbent bound the highest amount of CD4, the recovery was the least of all the sorbents. On the other hand, the L-71 sorbent showed moderate binding but the recovery of sCD4 was quantitative. Thus, MAb L-71 was judged to be the most suitable candidate for immobilization to prepare an affinity sorbent to purify CD4 and a scale-up IA sorbent to purify CD4 congeners. Using this antibody as IA sorbent, recombinant proteins loaded on the column were recovered with very high yield using elution with 0.1 M acetic acid. The proteins recovered formed an electrophoretically homogeneous product and after neutralization were highly active with respect to gpl20 binding, as judged by a radioligand-bead binding assay. The sorbent was also used successfully to purify full-length CD4 in highly active form.

A large-scale IA purification process was developed for the isolation of recombinant sCD4 from *Escherichia coli* fermentations.[27] The MAb used for IA purification of sCD4 recognized a conformation-dependent epitope on the surface of domain 1 of CD4. IA chromatography was used to purify both sCD4–183, consisting of the N-terminal 183 amino acids of human CD4, and sCD4–PE40, a fusion protein consisting of the N-terminal 178 amino acids of CD4 and amino acids 1–3 and 253–613 of *Pseudomonas* exotoxin A (PE40). Relatively high recoveries of sCD4–183 and sCD4–PE40 were observed in the IA step of the purification process (71 and 79% recovery, respectively). sCD4–183 was purified from *E. coli* cell pellets using cell disruption, protein solubilization, oxidation, Q-Sepharose anion-exchange and IA chromatography steps. sCD4–PE40 was purified from cell pellets using cell disruption, protein solubilization, oxidation, Cu^{2+} immobilized metal-affinity chromatography, anion-exchange and IA chromatography steps. The immobilized MAb appeared to select for correctly folded CD4 protein, since sCD4–183 and sCD4–PE40 purified by the IA method bound HIV glycoprotein gp 120 (HIV gp 120) in vitro. The results demonstrate that immobilized monoclonal antibodies directed against conformational epitopes may be used for the rapid purification of gram amounts of correctly folded protein from mixtures of oxidized *E. coli* proteins.

4.2.2 Red Blood Cell Membrane

Molecular masses of the proteins contained in red blood cells membrane range between 20 and 200 kDa, and consequently the most common HPLC techniques used for the purification of these proteins are based on SEC (which may be used in the presence of either ionic or nonionic detergents) and, to a lesser extent, on ion-exchange or hydrophobic interaction chromatography (compatible with nonionic detergents). We have used a reversed-phase C-4 column for the separation of

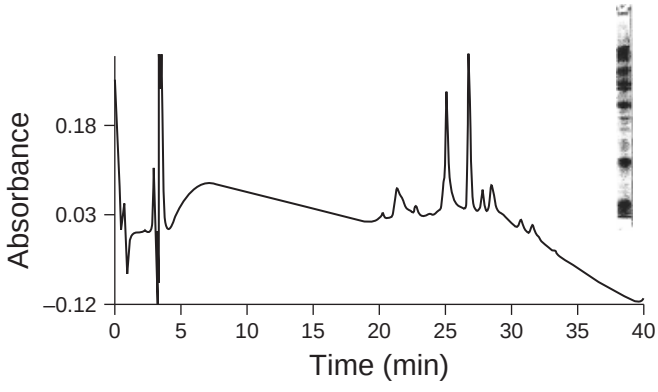

Figure 2 Separation of erythrocyte membrane proteins extracted in pure acetonitrile. Inset shows SDS/PAGE of proteins loaded into the column. Vydac C-4 column (250 × 4.6 mm ID), eluted at 1.0 mL min^{-1} with an acetonitrile gradient (20–100%) in 0.1% TFA over 60 min.

membrane proteins from erythrocyte ghosts using mobile phases containing acetonitrile and 0.1% TFA. The erythrocytes were swollen by 5 mM Na_2HPO_4 isotonic solution and after several centrifugations at 3000 *g* the pellet was solubilized in two volumes of pure acetonitrile. Samples were loaded directly on the column without any other manipulation. Figure 2 shows the chromatogram obtained using a gradient of acetonitrile.

A silica-based C-4 column eluted with TFA–acetonitrile is a perfect combination for several polypeptide separations, but was considerably less suitable for membrane protein separation than a resin-based phenyl column eluted with the same mobile phase. In our case the separation obtained with Vydac C-4 may be considered excellent, better than that reported previously with a Nucleosil C-4 column[28] where only one main peak was observed. In our opinion, the better separation is related to the absence of detergent during the extraction and elution run and probably also to the better quality of resin in the column.

4.2.3 Separation of Photosynthetic Proteins from the Thylakoid Membrane of Chloroplasts

In this section we present the results of a study performed to develop a rapid and straightforward HPLC method to resolve and identify the protein components of PSI and PSII. These two large complexes, of at least 40 proteins, are present in the thylakoid membrane of chloroplasts (Figure 3) and have the main role of capturing the light and giving the necessary energy to transfer electrons from water to nicotinamide adenine dinucleotide phosphate (NADP) via the two photosystems, PSI and PSII, and a number of electron carriers.

PSII is embedded in the thylakoid membrane and contains a reaction center (core) of 21 large and smaller proteins, surrounded by a specific light-harvesting system, which is the major protein–chlorophyll *a*–*b* complex [light-harvesting complex of photosystem II (LHCII)] and minor antennas called CP29, CP26 and CP24.[29] In contrast, structural information on PSI at the atomic level is lacking.

The proteins associated with PSI and PSII have been traditionally resolved by SDS/PAGE into several closely related hydrophobic membrane proteins, typically in the range 21–40 kDa.[30] The uncertainty of the number of proteins in the complex is mainly due to the different procedures and detergents used to solubilize PSI and PSII from the thylakoid membranes and to the different methods employed in isolating and characterizing the individual membrane proteins in the complex. The antenna proteins show amino acid compositions similar in length and sequence,[31] and therefore the characterization of these proteins is performed by tedious electrophoresis and immunoblotting, which require run times of 20–30 h.[32] Traditional approaches by SDS/PAGE are not only cumbersome but also rather ineffective for evaluating differences in the relative quantity of each LHCII component. We have set up a high-performance separation technique for the resolution of thylakoid membrane proteins employing a reversed-phase Vydac Protein C-4 column containing 5-μm pores.[33]

4.2.3.1 Preparation of Photosystem I and Photosystem II

In Figure 4 are summarized the main steps to separate the PSI and PSII present in the thylakoid membrane.

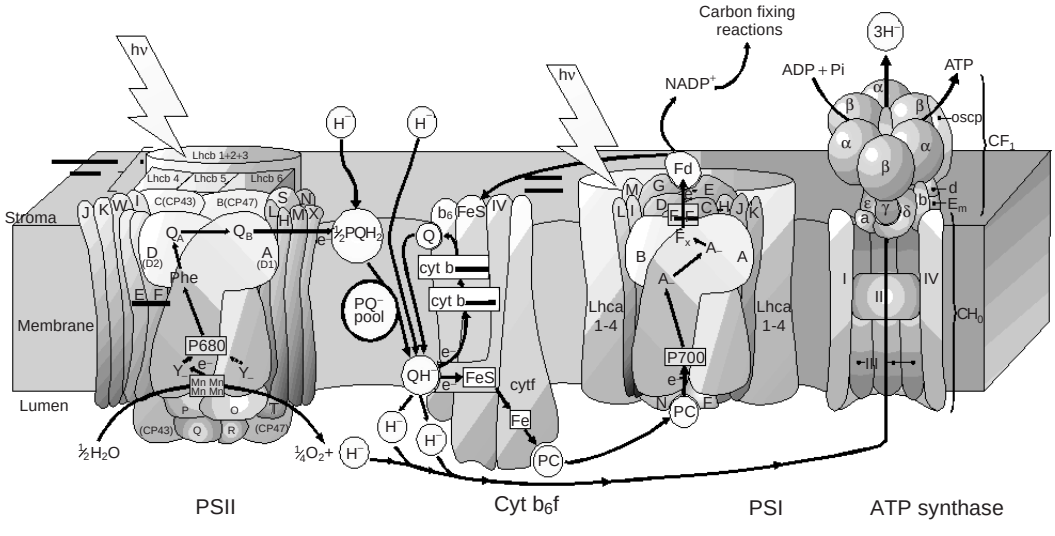

Figure 3 Organization of proteins contained in the thylakoid membrane and involved in photosynthesis. (Reproduced by permission of Professor J. Barber, Imperial College, London.)

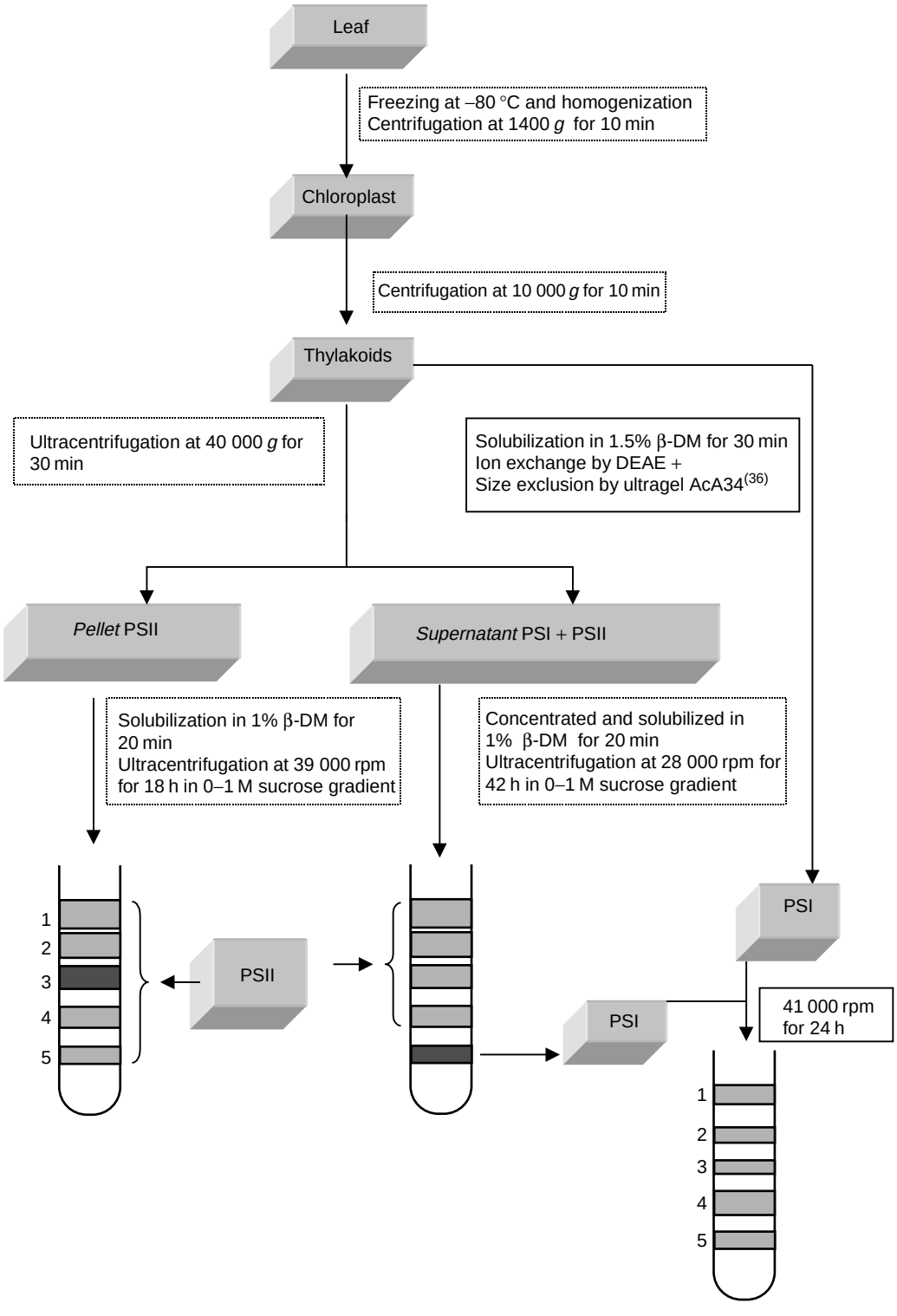

Figure 4 Separation scheme for PSII and PSI.

Chloroplasts were isolated from spinach leaves according to the method of Berthold et al.[34] with the following modifications. Leaves were powdered in liquid nitrogen and subsequently homogenized in an ice-cold 20 mM tricine buffer (pH 7.8) containing 0.3 M sucrose and 5.0 mM magnesium chloride (buffer B1). The homogenization was followed by filtration through one layer of Miracloth (Calbiochem, San Diego, CA, USA)

and centrifugation at 4000 *g* for 10 min at 4 °C. Pellets were suspended in buffer B1 and centrifuged as above. These second pellets were resuspended in 20 mM tricine buffer (pH 7.8) containing 70 mM sucrose and 5.0 mM magnesium chloride (buffer B2) and centrifuged at 4500 *g* for 10 min. Pellets containing the thylakoid membranes were then resuspended in 50 mM morpholinoethanesulfonic acid (MES) buffer (pH 6.3) containing 15 mM sodium chloride and 5 mM magnesium chloride (buffer B3) at 2.0 mg mL^{-1} chlorophyll for 15 min after adding Triton X-100 at a final ratio of 20 mg mg^{-1} chlorophyll. The concentration of chlorophyll was determined according to the method described by Porra et al.[35] The incubation was terminated by centrifugation at 40 000 *g* for 30 min at 4 °C. This pellet containing the PSII complex and corresponding to the BBY preparation described by Berthold et al.[34] was resuspended in buffer B3 containing 20% (v/v) glycerol and stored at −80 °C.

4.2.3.2 Subfraction of Photosystem II by Sucrose-gradient Ultracentrifugation PSII membranes, once isolated from thylakoid membranes, were then subjected to sucrose-gradient ultracentrifugation in order to isolate the protein components of the minor antenna system (band 2) from the major (band 3) and also from the reaction-center complexes (bands 4–5). With this aim in mind, PSII membranes were pelleted by centrifugation at 10 000 *g* for 5.0 min at 4 °C, suspended in buffer B3 at 1.0 mg ml^{-1} chlorophyll and then solubilized by adding 1% (w/v) β-DM. Unsolubilized material was removed by centrifugation at 10 000 *g* for 10 min. The supernatant was rapidly loaded on to a 0.1–1.0 M sucrose gradient containing buffer B3 and 5.0 mM β-DM. The gradient was then spun on a Kontron (Milan, Italy) Centricon T-1080 ultracentrifuge equipped with a TST 41.14 rotor at 39 000 rpm for 18 h at 4 °C. Green bands were harvested with a syringe. The SDS/PAGE analysis of these green bands revealed that band 3 contained essentially LHCII proteins, as reported previously by Bassi and Dainese.[37] Consequently, this was the purified starting material used for the HPLC study.

4.2.3.3 Separation of Photosystem II Proteins by Reversed-phase High-performance Liquid Chromatography In this study we searched for the conditions to resolve the protein components of the major and the minor antenna system of PSII either as the complex isolated by sucrose-gradient ultracentrifugation and as assembled in the grana membrane (BBY particles), using either an analytical (250 × 4.6 mm ID) or a semipreparative (250 × 10 mm ID) sized column, both packed with the same 5-μm spherical Vydac C-4 stationary phase. The use of a semipreparative sized column was needed in order to obtain an amount of purified polypeptide sufficient for peak identification by SDS/PAGE, immunoblotting, and amino acid microsequencing. The optimum separation of the protein components of the PSII antenna system was obtained by the following procedure.

The Vydac C-4 columns were preequilibrated with 38% (v/v) aqueous acetonitrile solution containing 0.1% (v/v) TFA and samples were eluted using either gradient I or II, depending on the HPLC unit employed for the separation. Gradient I consisted of a first linear gradient from 38 to 55.4% (v/v) acetonitrile in 22 min, followed by 3 min of isocratic elution with the eluent containing 55.4% acetonitrile, followed by a second gradient segment from 55.4 to 61.8% (v/v) acetonitrile in 8 min and by a third gradient segment from 61.8 to 95% acetonitrile in 1 min. Gradient II consisted of a first linear gradient from 38.0 to 61.8% (v/v) acetonitrile in 40 min, followed by a second gradient segment from 61.8 to 95% (v/v) acetonitrile in 1 min. For both gradients the last gradient segment up to 95% acetonitrile was used for washing out hydrophobic contaminants of the PSII antenna system from the column. Gradient I was used to elute either the analytical or the semipreparative column. The flow rate was 1.0 mL min^{-1} with the analytical column and 4.7 mL min^{-1} with the semipreparative column. These conditions were selected in order to maintain the same gradient shape with both columns by keeping the ratio of the gradient volume to the column volume constant.

The chromatogram displayed in Figure 5(a) shows that the material harvested from the second band of the sucrose-gradient ultracentrifugation, containing a mixture of the protein components of both the major and the minor PSII antenna system, was resolved into eight main peaks. Five of these peaks, with corresponding retention times, were also obtained on separating, under the same conditions, the material harvested from the third band of the sucrose-gradient ultracentrifugation (Figure 5b), containing essentially the protein components of the major PSII antenna system. From these data it can be inferred that the peaks labeled 1, 2, 5, 6 and 7 correspond to components present in both the sucrose bands 2 and 3, whereas peaks 3, 4 and 8 are essentially due to the component present only in the sucrose band 2, which can be tentatively identified with the minor antenna complexes CP29, CP26, and CP24.[38] Figure 5(c) shows the chromatogram of BBY injected on the column directly, avoiding the separation step by sucrose-gradient ultracentrifugation. It can be observed that upon injection of BBY directly, the protein components of the PSII major and minor antenna system are well resolved without interference from the other protein components of PSII. Therefore, the use of the crude PSII membrane preparation does not affect the resolution and retention times, while new and smaller peaks appear at longer elution times. The

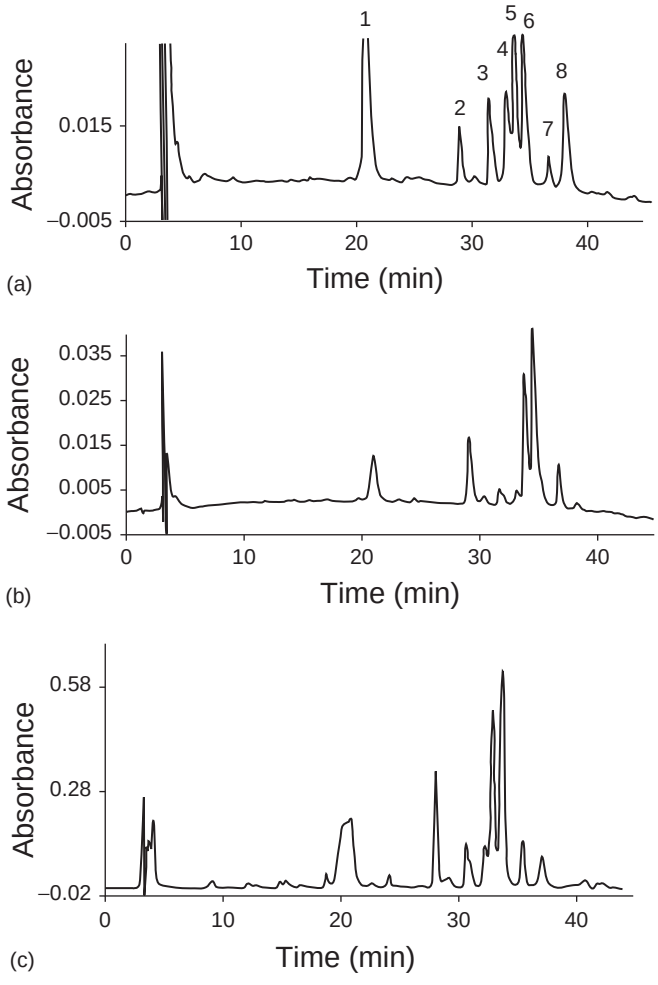

Figure 5 RPHPLC separation of spinach chlorophyll *a*/*b* binding proteins contained in (a) band 2 and (b) band 3 of spinach BBY fractionated by sucrose-gradient ultracentrifugation and (c) in whole BBY, directly injected into the column. Column, Vydac Protein C-4 (250 × 10 mm ID), eluted with an increasing acetonitrile concentration gradient as described in the text. Flow rate, 4.7 mL min^{-1}; detection at 214 nm.

new peaks, observed in Figure 5(c), represent the core proteins, which are more hydrophobic and elute at higher concentrations of acetonitrile (data not shown).

In order to assign to each HPLC peak a name of the well known antenna proteins, SDS/PAGE systems employing different buffer systems and either Coomassie Brilliant Blue or silver staining followed by immunoblotting detection with antisera directed to the individual antenna proteins were performed, in addition to amino acid microsequence analysis of the material collected throughout the chromatogram.

In the case of SDS/PAGE and immunoblotting, each fraction collected from the semipreparative chromatographic separation lyophilized, were dissolved in 120 mM tris(hydroxymethyl)aminomethane (TRIS)–HCl buffer (pH 8.45), containing 120 mM dithiothreitol (DTT), 5 M urea and 4% (w/v) SDS, and then analyzed by SDS/PAGE, according to the method reported by Shaegger and von Jagon[39] (Figure 6a).

Following electrophoresis, the gels were either silver stained or transferred to nitrocellulose. Replicates were assayed with antisera directed to LHCII, CP29, CP26, and CP24[40] (Figure 6b–d). Because of the high degree of homology shared by all of the antenna polypeptides, it was essential for identification of the fractions to take into account both the immunoreactivity and the electrophoretic mobility of the SDS/PAGE bands. In TRIS/Tricine electrophoresis, antenna proteins migrate in the following order of increasing mobility: CP29 > CP26 > LHCII > CP24 and, within LHCII, Lhcb1 > Lhcb2 > Lhcb3.[41] Accordingly, the slower migrating band was detected by the anti-CP29 in peak 4 (Figure 6b), while peak 8, containing a slightly more mobile band, was recognized by anti-CP26 (Figure 6c). The anti-CP24 detected several bands (Figure 6d) but the strongest signal, with respect to the intensity of the band in the silver stained gel, was obtained with the most mobile band in peak 3. Therefore, peaks 3, 4 and 8 seem to contain CP24, CP29, and CP26, respectively. The higher hydrophobicity of CP26 agrees with the longer elution time required.

Thus, according to the migration order of increasing mobility reported above, the fastest migrating band was recognized as Lhcb3, the second fastest migrating band as Lhcb2, and the two newly resolved bands as Lhcb1 components of LHCII.

From the above data, we conclude that peak 1 does not contain protein components; peak 2, the Lhcb2 component of LHCII; peak 3, the CP24 (Lhcb6); peak 4, CP29 (Lhcb4); peaks 5 and 6, two Lhcb1 components of LHCII; peak 7, Lhcb3; and peak 8, CP26 (Lhcb5).

In order to support the identification previously assigned, the protein contained in each HPLC peak was subjected to amino acid microsequencing and compared with amino acid sequences reported in the literature for other species (data not shown). Good agrrement with the above assignment was found.

Finally, the assignment of each peak resolved by RPHPLC performed by electrophoresis, immunoblotting, and amino acid sequencing is corroborated from the values of molecular masses determined by the combined use of microbore HPLC coupled on-line with a mass spectrometer[42] equipped with an electrospray ion source (see section 5).

From the results presented, it can be inferred that the major antenna system of PSII isolated from spinach leaves contains two different Lhcb1 proteins which can be resolved by the RPHPLC system employing a 250 mm long Vydac C-4 column eluted with a multisegment acetonitrile gradient, as well as by the high-resolution

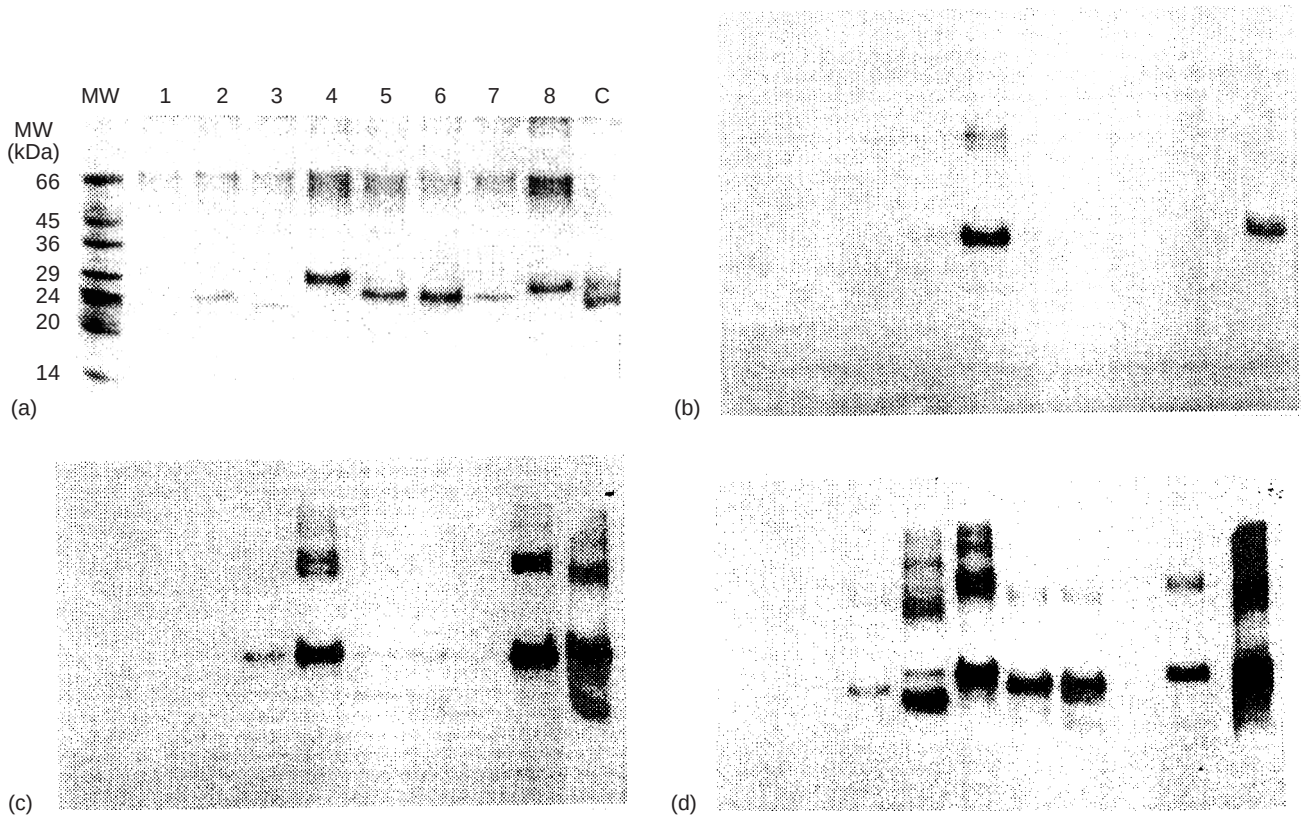

Figure 6 Gel electrophoresis and immunoblotting of each protein collected throughout the chromatographic run reported in (a). Lanes correspond to the peak labels in (a); the last lane (c) represents the LHCII from *Zea mais* used as a control. (a) SDS/PAGE of uniform polyacrylamide concentration (13%) with TRIS–Tricine system of each collected peak, which were dried and than solubilized in 4% LDS, 120 mM DTT and 120 mM TRIS–HCl (pH 8.45), containing 5 M urea. (b) Immunoblotting of the gel reported in (a) detected by using antisera directed to CP29. (c) Immunoblotting of the gel reported in (a) detected by using antisera directed to CP26. (d) Immunoblotting of the gel reported in (a) detected by using antisera directed to CP24. (Reproduced by permission of Kluwer Academic Publishers from L. Zolla, A.M. Timperio, M.G. Testi, M. Bianchetti, R. Bassi, F. Manera, D. Corradini, *Photosynth. Res.*, **61**(3), 281–290 (1999).)

TRIS/Tricine SDS/PAGE system in gels of 15 cm length (data not shown). On the other hand, the small microsequence performed does not reveal any amino acid differences and the molecular mass measured by mass spectrometry (MS) indicates that the difference is of the order of 60–80 kDa. More experiments and a complete amino acid sequence are necessary to give an explanation for these different subpopulations of Type I. Nevertheless, the existence of more type I LHCII is in accordance with molecular genetic data reported in the literature showing that higher plants have several Lhcb1 genes encoding different Lhcb1 proteins for each species.[31] However, it is important to note that the resolution by SDS/PAGE of two Lhcb1 proteins usually requires special experimental conditions such as the use of polyclonal and monospecific antibodies,[43] dedicated electrolyte solutions, and a gel of extended length,[44] whereas using HPLC the two different subpopulations may be easily separated.

The RPHPLC method reported here, in addition to being rapid, simple, and precise, has proven to be effective at detecting differences in the protein components of LHCII isolated from different plants that might be not evidenced by denaturating SDS/PAGE, as in the case of Type I of spinach. This knowledge is expected to shed light on the composition and supramolecular organization of LHCII and may increase the understanding of the molecular mechanisms underlying the physiological adaptations of higher plants to environmental conditions.

4.2.4 Separation of Proteins Present in Photosystem I

PSI prepared from the thermophilic cyanobacterium *Synechococcus elongatus* has been crystallized and the structure determined at a resolution of 6 Å.[45] The determination of the structure of higher plant PSI has been less successful, most likely because it is a much larger complex than its cyanobacterial counterpart. The

cyanobacterial PSI complex consists of 11 subunits.[46] In the case of higher plants, although a lot of PSI genes have been identified and sequenced, there is little information on the protein, and the relationship between the gene and relative protein expressed is lacking. This is related to the fact that the molecular masses of most of these proteins range between 4 and 25 kDa, and consequently they are not well separable by SDS/PAGE.

4.2.4.1 Isolation of Photosystem I by Sucrose-gradient Ultracentrifugation Supernatants of the first purification were concentrated by ultrafiltration and then solubilized by β-DM at a final concentration of 1%. After stirring for 20 min at 4 °C, the sample was centrifuged for 10 min at 40 000 *g*, and 6-mL aliquots of the supernatant were loaded on 0.1–1 M sucrose gradients (35 mL) layered over 2 mL of 2 M sucrose, containing 5 mM Tricine (pH 7.8) and 0.03% β-DM. After centrifugation for 42 h at 28 000 rpm in an SW28 rotor (Beckman) at 4 °C, four green bands were distinguishable. The bands were collected and analyzed by SDS/PAGE.[47] The lowermost band, containing PSI-200, was diluted in 5 mM Tricine (pH 7.8) and centrifuged for 3 h at 70 000 rpm in an 80 Ti rotor (Beckman). The pellet was resuspended in 5 mM Tricine (pH 7.8) and 50 mM sorbitol, frozen in liquid nitrogen, and stored at −80 °C.

4.2.4.2 Purification of Photosystem I Core and Light-harvesting Complex of Photosystem I The pellet from PSI-200 preparation was resuspended at 0.3 mg mL^{-1} in distilled water and solubilized by 1% β-DM and 0.5% Zwittergent-16. After stirring for 20 min at 4 °C, the sample was rapidly frozen in liquid nitrogen and slowly thawed to improve the detachment between the PSI core and the light-harvesting complex of photosystem I (LHCI). Samples (1.5-mL aliquots) were loaded on a 12-mL 0.1–1 M sucrose gradient, also containing 5 mM Tricine (pH 7.8) and 0.03% β-DM. After centrifugation for 24 h at 4 °C at 41 000 rpm in an SW41 rotor (Beckman), five green bands were obtained. The bands were analyzed by SDS/PAGE.[47] The second band from the top contained all of the LHCI polypeptides and the third the PSI core particles. The fractions were frozen in liquid nitrogen and stored at −80 °C.

4.2.4.3 Separation of Photosystem I Proteins by Reversed-phase High-performance Liquid Chromatography The complete resolution of the protein components of the PSI was performed by RPHPLC as for PSII. The experiments were carried out subjecting to chromatography all proteins of PSI-200 and then separately the antenna and core proteins previously isolated by sucrose-gradient ultracentrifugation.

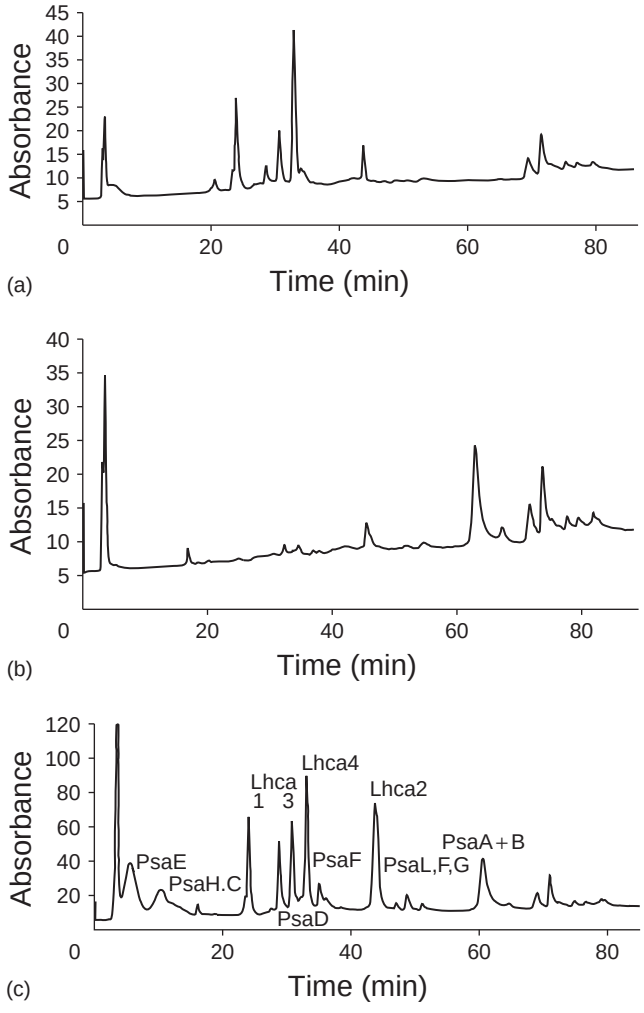

Figure 7 RPHPLC separation of (a) spinach LHCI proteins contained in band 2 and (b) core proteins in band 3 of PSI-200 fractionated by sucrose-gradient ultracentrifugation and (c) in whole PSI-200. Column, Vydac Protein C-4 (250 × 10 mm ID), eluted with an increasing acetonitrile concentration gradient as described in the text. Flow rate, 4.7 mL min^{-1}; detection at 214 nm.

Figure 7(a–c) compares the chromatograms obtained upon injection of (a) isolated antenna proteins, (b) the core proteins, and (c) the total PSI-200.[48]

Different gradients were used to find the optimum experimental conditions to resolve the protein components of the antenna and core complex. In the case of PSI-200, the optimum separation of all proteins was obtained with a multisegment gradient. Good complementarity between the three chromatograms is observed, indicating that the presence of all proteins in the PSI-200 system does not affect the elution time of each protein. The material harvested from the second band of the sucrose-gradient ultracentrifugation, which was previously checked by SDS/PAGE and contained a mixture

of the protein components of the PSI antenna system, is resolved into five main peaks (Figure 7a). Most of the peaks are symmetrical, but the last one at 43.4 min is broad, indicating the presence of more than one protein. In contrast, the material harvested from the third band of the sucrose-gradient ultracentrifugation, containing essentially the core protein components of PSI, shows a main peak at a long elution time of 61 min, and numerous smaller peaks. The main peak is broad and elutes at a long time, suggesting the presence of more than protein showing high hydrophobicity. Finally, in the chromatogram for the total PSI-200, five of the same peaks with retention times corresponding to those observed in Figure 7(a) are present, which are the most abundant separated proteins and allow one to hypothesize that they represent the antenna complex, although only four proteins are expected. Moreover, the main peak observed in the core complex (Figure 7b) is also observed in Figure 7(c), which may be hypothesized to contain psaA and psaB, and other smaller core peaks are underestimated when all proteins are present.

Once the optimized conditions to obtain a better separation of each component had been found, we attempted the identification of each protein contained in each peak by coupling on-line HPLC with MS (see section 5). The mass values measured by ESIMS corresponded to the molecular mass expected on the basis of the DNA sequence, allowing the assignment of each peak as reported in Figure 7(c).[49] These results corroborate the utility of on-line coupled HPLC/MS with an electrospray ion source.

Finally, the separation of both PSI and PSII may be achieved by direct injection of thylakoid membrane into the column and elution with a multisegmented gradient. All the protein components of the PSII can be well resolved without interference from the other protein components of PSI, indicating that the use of the crude thylakoid membrane preparation does not affect the resolution and retention times of each protein.

4.3 Separation of Lipoproteins

4.3.1 Preparation of Lipoprotein Fractions

Plasma or serum can be used for lipoprotein analysis. However, plasma is preferred since it can be kept cold in order to slow most enzymatic processes that degrade lipoproteins. The following procedure is commonly used to prepare lipoproteins.[15] Human blood (ca. 20–30 mL) was collected from normolipidemic males by venepuncture after 12–16 h of fasting. The blood was allowed to stand for 2–3 h at room temperature until agglutination was complete. The plasma was withdrawn after centrifugation at 1000 *g* at 15 °C for 15 min. The plasma density was adjusted to 1.225 g mL^{-1} by adding solid potassium bromide (0.3517 g KBr per milliliter of plasma). Plasma (1.225 g mL^{-1}, ca. 3–5 mL) was then placed in ultracentrifuge tubes, which were centrifuged in a swinging-bucket rotor at 200 000 *g* at 10 °C for 40 h. The lipoprotein fraction prepared by this procedure did not contain serum proteins, except for a small amount of albumin. The lipoprotein fraction in the KBr solution was dialyzed against 0.154 M sodium chloride solution.

4.3.2 Preparation of Individual High-density Lipoprotein, Low-density Lipoprotein and Very Low-density Lipoprotein Fractions

Many separations subfractionate the lipoprotein fraction of plasma. Hence each main class of lipoproteins may be collected by ultracentrifugation using a multiple discontinuous density gradient. For this purpose, human blood was collected from fasting normolipidemic healthy males in tubes containing 0.15% EDTA. The plasma was separated by centrifugation at 700 rpm at 7 °C for 20 min. A discontinuous NaCl–KBr density gradient (total volume 18.5 mL) was formed by adjusting the density of plasma to 1.30 g mL^{-1} with solid KBr and sequentially layering on the adjusted plasma salt (NaCl–KBr) solutions with densities of 1.240, 1.063, 1.019 and 1.006 g mL^{-1} and 0.5 mL of distilled water. Tubes loaded with the discontinuous density gradient were placed in a vertical rotor and centrifuged at 313 500 *g* at 7 °C for 80 min.

4.3.3 Separation of Particles and Subfractions

In an alternative to ultracentrifuge separation, the GPC mode of HPLC is the most commonly used method to separate particles such as HDL, VHDL, and LDL. Two commercial columns are commonly used: Superose 6 (from Pharmacia) and TSK-type.[50] Rigid, fast-flow agarose gels, such as Superose 6, are able to accept large sample volumes (up to 5 mL) and it is preferred for the preparation of larger quantities of plasma lipoproteins. Furthermore, multiple automated separations of lipoproteins from whole plasma may be performed with a preparative-grade Superose 6 column. Thus 2 mL of plasma are loaded and eluted with 0.9% sodium chloride and 2 mM sodium phosphate (pH 7.4). The eluent also contains 0.02% or 1% sodium azide for serum or plasma separation, respectively. VLDL, LDL, and HDL can be separated in a single run. A single separation can be completed in about 160 min. Hence the system can be used to analyze up to six different plasma samples (2 mL per sample) overnight.

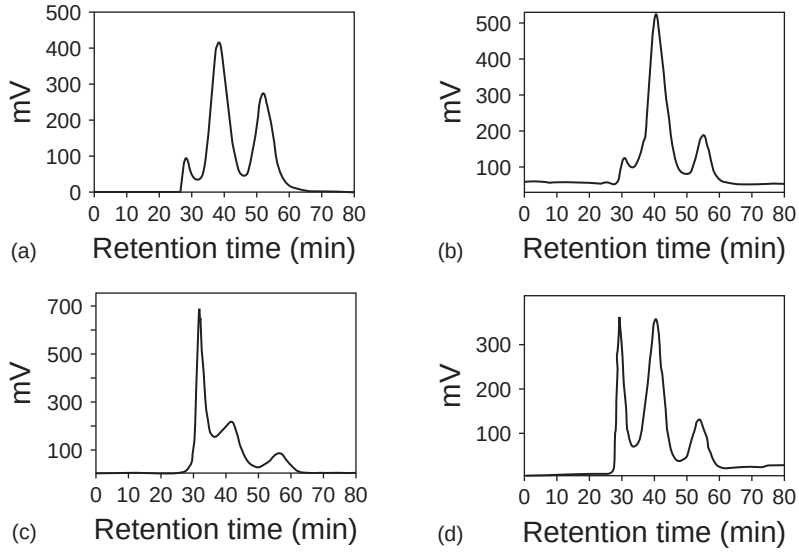

Figure 8 Chromatographic profile of normal and HLP sera. Conditions: 300-mm Superose 6 column eluted with 100 mM Na_2HPO_4, 200 mM NaCl (pH 7.4) at 300 mL min^{-1}. (a) Normolipidemic sample; (b) Type IIa HLP; (c) Type III HLP; (d) Type IV HLP. Peaks from left to right: VLDL, LDL, HDL. (Reproduced by permission of *Clinical Chemistry* from W. Marz, R. Skeimer, H. Scharnagi, U.B. Seiffert, W. Gross, *Clin. Chem.*, **39**(11), 2276–2281 (1993).)

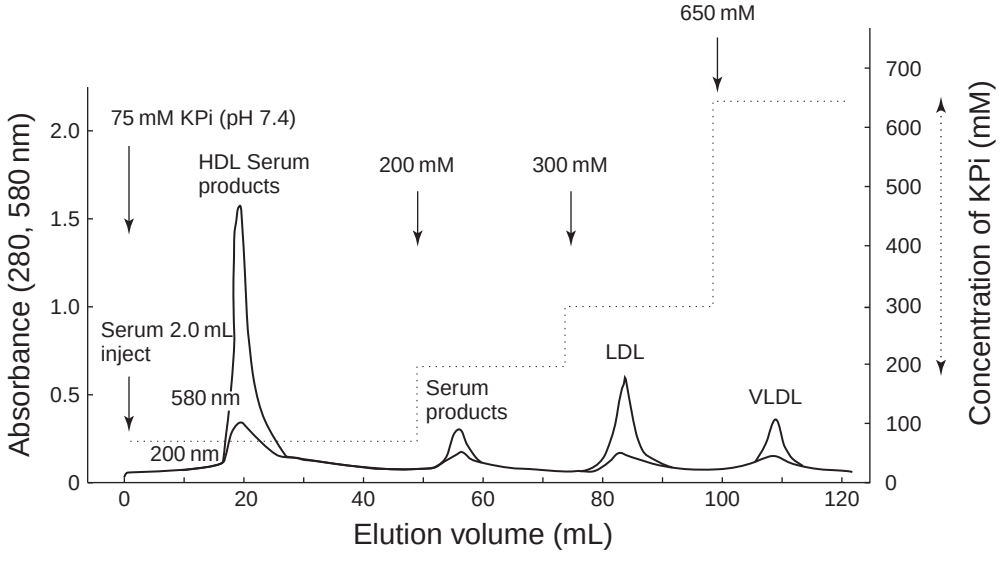

Figure 9 Stepwise elution profile of human serum lipoproteins. Column, Bio-Gel HTP DNA grade 25 × 1.0 cm ID; eluents, 25, 200, 300 and 650 mM potassium phosphate buffer (KPi) (pH 7.4), flowrate, 12.0 mL h^{-1}, sample, 2.0 mL of human serum. (Reproduced by permission of Elsevier Science from Y. Shibusawa, *J. Chromatogr. B*, **699**, 419–437 (1997).)

Fast-flow separation of plasma is also used, where lipoproteins are detected on-line at 500 nm after post-column derivatization with an enzymatic cholesterol reagent (CHOD-PAP, Boehringer Mannheim). The eluting lipoproteins and the cholesterol reagent are mixed in a chamber attached to the column outlet. The mixture is then passed through a reaction capillary and detected. Complete separation of lipoprotein fractions is usually achieved in less than 80 min. The chromatographic profiles of normal and several types of hyperlipoproteinemic (HLP) serum samples are illustrated in Figure 8(a–d). The three peaks correspond to, in order, VLDL, LDL, and HDL.[17]

Separation of lipoproteins by HPLC on TSK-type gel permeation columns is faster than on a Superose 6 column but smaller sample volumes (<300 μL) may be loaded. Furthermore, serum instead of plasma is preferred for TSK-gel HPLC of lipoproteins. With plasma there is a

danger of fibrin forming during the analysis. A further disadvantage of the TSK method is that this type of column is generally more prone to clogging than a Superose 6 column. Adsorption of lipoproteins on the support has also been noted by some workers. It was found that a combination of G4000 SW and G3000 SW eluted with 0.15 M NaCl (pH 7) was best at separating serum lipoproteins into VLDL, LDL, HDL_2 and HDL_3.[51] Separation by a TSK column allows the determination of lipids using a combination of two TSK columns and on-line enzymatic reaction. The system used two detectors. The first was placed immediately after the column and monitored protein absorbance at 280 nm. The second detector was placed after the enzymatic reactor and detected lipid absorbance at 500–600 nm.[51]

Separation of large amounts of LDLs and VLDLs from human serum may be performed by using a hydroxyapatite Bio-Gel HTP DNA grade column (25 × 1.0 cm ID) by one[52] or four stepwise elutions.[15] In the latter case, a 2-mL volume of human serum is loaded on the column and eluted with a discontinous gradient of 75, 200, 300 and 650 mM potassium phosphate buffer at pH 7.4. Figure 9 shows the elution profile of human serum obtained with a hydroxyapatite column.

Four peaks are detected, the first containing HDLs and serum proteins, the second the serum proteins, eluted at 200 mM, the third mainly LDLs, eluted at 300 mM, and the fourth VLDLs, eluted at 650 mM.

The whole lipoprotein fraction ($d < 1.21$) can be further subfractionated into VLDL, LDL and HDL by GPC.[18] Lipoprotein fractions prepared by ultracentrifugation are often used for this further chromatographic separation. In this way, the procedure is both analytical and preparative, since the isolated fractions contain no contamination from serum proteins.

Subfractionation of HDL particles accomplished by GPC gives rise to subfractions having different sizes (3.81–528 nm) and different amounts of proteins: HDL_{2a}, HDL_{2b}, HDL_{3b}, and HDL_{3c}.[18]

HDL particles may also be subfractionated by heparin affinity chromatography according to the apoE content.[20] HDLs rich in apoE were separated into five subfractions using an elution buffer with varying Mn^{2+} concentration.[19]

In the case of LDL, charge heterogeneity was demonstrated by ion-exchange chromatography. These types of separations are important since oxidatively modified LDLs have been associated with atherosclerosis.[53] LDL particles modified by Cu^{2+} were analyzed by anion-exchange chromatography using a Mono Q HR 515 column (Pharmacia).[21] Five different oxidatively modified LDLs were identified. The isolated fractions exhibited differences in density and lipid content.

4.3.4 Separation of Apolipoproteins and Isomeric Forms

The growing evidence that apolipoproteins are better markers of CHD than serum cholesterol levels[14] has increased the interest in analyzing lipoproteins according to the apolipoprotein compositions.

Affinity chromatography, which employs specific and reversible interactions between lipoproteins and ligands, is capable of separating lipoproteins on the basis of their apolipoprotein content by using nonspecific binding and IA. Heparin was successfully immobilized on to glyceryl controlled-pore glass (CPG) and used to bind specifically β-lipoproteins (apoB).[54] Serum samples applied to heparin–CPG were separated into two fractions. The unretained fraction was washed though the column with 0.1 M NaCl and the retained fraction was eluted with 1 M NaCl. Radial immunodiffusion studies confirmed the complete resolution of apoA- and apoB-containing lipoproteins. However, attempts at incorporating the heparin–CPG column into a high-performance affinity chromatography system were unsuccessful. Two types of lipoproteins have been clearly established as atherogenic, the apoB100-containing LDL and apoB48-containing chylomicron remnants.[55] Human genetic disorders resulting in increased circulating levels of either of these lipoproteins cause premature atherosclerosis. Experiments were reported to demonstrate that the NH_2-terminal region of apoB binds to heparin affinity gels with an affinity equal to or greater than that of apoB100-containing LDL.[56] In addition, apoB48-containing lipoproteins were observed to bind to heparin as well as LDL. Hence, based on these findings, it was proposed that NH_2-terminal apoB contributes to the atherogenicity of LDL and remnant lipoproteins.

Dextran sulfate is a synthetic analog of heparin, and also binds preferentially to apoB-containing lipoproteins. Commercial sulfated dextran beads from Sigma consist of sulfated cross-linked dextran and are useful for rapid (15 min) separation of apoA- and apoB-containing lipoproteins on elution of proteins by NaCl at 80 mM.

IA chromatography provides the highest specificity for separation and allows the separation of relatively large volumes (1–10 mL) of whole plasma or lipoproteins. Preparations of both polyclonal[57] and monoclonal[58] antibodies to apolipoproteins, produced using purified delipidized apolipoproteins, are commonly used as antigens. Antibodies to apolipoproteins coupled to cross-linked dextran (Sephadex) or agarose derivatives (Sepharose) serve as selective ligands for the separation of lipoproteins by IA chromatography.

Separation of apoproteins into their isomers is of interest especially for apoA and apoB, which may be performed by sequential chromatographic stages using more

IA or affinity followed by gel permeation, ion-exchange or reversed-phase techniques. In the first procedure, whole plasma is first fractionated into apoB and apoA lipoproteins by affinity chromatography on concanavalin A (con A). Further subfractionation proceeds by subsequent IA chromatography using a specific antibody: anti-apoA-II immunosorbent, which retains a fraction containing lipoprotein (A-1 + A-II) and lipoprotein A-II. Further, separation of this fraction on an anti-apoA-I immunosorber leads to the isolation of three types of apoA lipoproteins: lipoprotein A-1, lipoprotein A-II and lipoprotein (A-I + A-TI).[59]

Conventional GPC or ion-exchange chromatography on soft gel supports can be used for the preparation of large quantities (10–50 mg) of pure apolipoproteins,[60] whereas by HPLC only small amounts of sample (<5 mg) can be applied to the column but the analysis is completed within 60 min. Separation by high-performance GPC is strongly dependent on the completeness of the delipidation step, which may be performed satisfactorily with an organic solvent. However, high-performance GPC separation of HDL apolipoproteins has been achieved without prior organic solvent delipidation of the HDL fraction.[61]

GPC is often used as a preparative technique for isolating apolipoproteins derived from HDL or VLDL, while HDL apolipoproteins, apoA-I, apoA-II, and apoC, can be separated on Sephadex G-200 columns using TRIS–HCI buffer (pH 8.6) containing 8 M urea.[62] ApoC and apoE from VLDL can be isolated by Sepharyl S-200 (Pharmacia) chromatography.[62] Separation of apolipoprotein isoforms can be achieved by either ion-exchange or RPHPLC.

ApoC isoforms (C-III$_0$, C-III$_1$, and C-III$_2$) can be separated on Mono Q-HR 5/5 (Pharmacia) anion-exchange columns[63] in less than 30 min using a linear gradient of 0.15 M NaCl in TRIS–HCl containing 6 M urea (pH 8.2). This mode of apolipoprotein separation is based on differences in isoelectric points. Chromatography of VLDL and HDL apolipoproteins is performed on anion-exchange columns such as SynChropak AX300 (SynChrom IUC) and Mono Q-HR (Pharmacia).[64]

A rapid and highly efficient technique for monitoring heterogencity in apolipoproteins is the separation on the basis of differences in hydrophobicity by RPHPLC. Unfortunately, the small amounts of apolipoprotein (50–100 µg) which may be loaded, render this method qualitative but highly resolutive. Successful separations of apoA and apoC were performed on C-18 columns with a gradient of acetonitrile–water in the presence of 0.1% of TFA.[65] Figure 10 shows the separation of human apolipoproteins A-IV, A-I and E on a TSK Phenyl-5PW column.[66]

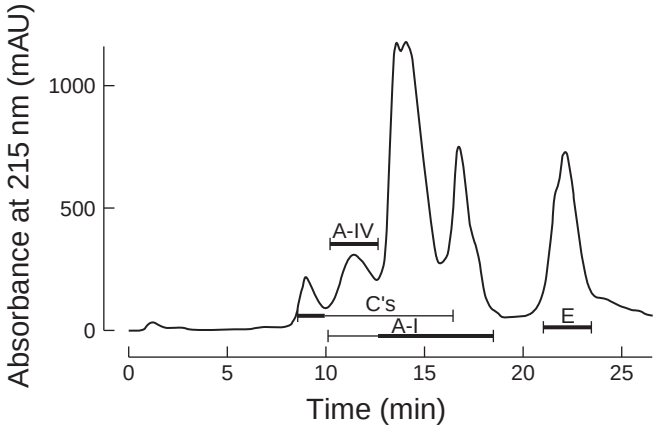

Figure 10 RPHPLC separation of delipidated rat HDL (200 mg protein) on a TSK Phenyl-5PW column. Gradient between 20 mM H_3PO_4 in water (pH 2.3) and 20 mM H_3PO_4 in 60% acetonitrile. Flow rate, 1 mL min^{-1} at 45 °C. Fractions were analyzed by SDS/PAGE and their lipoprotein contents as assessed by Comassie Brilliant Blue staining (bold bars) and by silver staining (normal) are as indicated. (Reproduced by permission of Elsevier Science from B. Meyer, E. Kecorius, P. Barter, N. Fidge, T. Tetaz, *J. Chromatogr. A*, **540**, 386–391 (1991).)

5 CURRENT TRENDS

5.1 Membrane Proteins

5.1.1 Perfusion Chromatography

The advent of perfusion chromatography, which consists of matrix containing flow-through particles, has further increased the interest in using HPLC systems, especially for membrane proteins.[67] In contrast to conventional particles, the new matrix has two classes of pores. Large ''throughpores'' allow convective flow through the particles, quickly carrying sample molecules to short ''diffusive'' pores inside. By reducing the distance and therefore the time over which molecules diffuse to access the particle binding surface area, the flow rate can be increased 10–100-fold with little or no loss in resolution or capacity. Thus, once a particular mode of chromatography has been selected, it is possible to perform it by using perfusion particles. In general, the perfusion chromatography allows:

- routine achievement of high-resolution laboratory-scale separations in 30 s–3 min;
- rapid and systematic determination of the best separation method for optimum purity;
- a significant reduction in the time to process large-volume samples;
- improved yields and recoveries of biological activity through faster processing.

5.1.2 High-performance Liquid Chromatography Coupled On-line with Mass Spectrometry

Further developments in HPLC system have been achieved by coupling the outlet on-line with other instrumentation. The introduction of electrospray ion sources that are coupled with quadrupole mass filters has produced a mass spectrometer that is easily compatible with HPLC and CE, and therefore analysis and measurement of biological samples at a reasonable cost may be performed.[42,68,69] Since most MS ion sources require lower flow rates than common chromatographic levels, a split of the HPLC eluent stream must be accomplished. Such systems are highly advantageous in that they facilitate the simultaneous separation of complex mixtures along with ultraviolet (UV) adsorption and mass detection.

With regard to the separation of thylakoid proteins of the PSII, RPHPLC interfaced with MS with an electrospray ion source allows the separation and accurate molecular mass determination of the individual membrane proteins contained in each peak on the HPLC trace, in particular for the identification of the major light-harvesting complex LHCII and minor (CP24, CP26 and CP29) antenna system, whose molecular masses range between 22 and 29 kDa.[42]

Table 2 reports the comparison of the molecular mass values calculated from the protein sequence derived from the isolated genes with those determined by reversed-phase high-performance liquid chromatography/electrospray mass spectrometry (RPHPLC/ESIMS) together with the apparent molecular masses given by SDS/PAGE. The molecular mass values determined are in good agreement with the computed molecular masses of these protein based on their DNA sequences.

Thus, the assignment of each peak resolved by RPHPLC performed as above by electrophoresis, immunoblotting, and amino acid sequencing is corroborated by the values of molecular masses determined by the combined use of HPLC coupled on-line with a mass spectrometer equipped with an electrospray ion source (ESIMS). Furthermore, the resolution of two variants of type I proteins is in agreement with previous findings reporting the resolution of more than one type I protein by both high-resolution polyacrylamide gel electrophoresis[37] and HPLC[38] and is consistent with the high copy numbers of *Lhcb1* genes isolated in higher plants,[31] giving confidence in the real potential of these coupled techniques. On the other hand, the isolation of multiple copy numbers of the same gene from several higher plant species accounts for the resolution of different variants of the same proteins reported in the literature.[31]

In the case of glycoproteins, ESIMS can detect whether an oligosaccharide is O or N-linked.[70] It can also differentiate between complex, high-mannose, or hybrid forms. Moreover, this technique may be used to gain limited linkage order information using collision induced dissociation (CID) with both a single- and a triple-quadrupole mass spectrometer. In fact, in a complex map the region of the map that contains the glycopeptides can be deduced by looking for characteristic patterns in the two-dimensional (2D) plot or by the observation of oxonium ions produced by CID. The plot of m/z against retention time (contour maps) as a facile approach to the rapid 2D mapping of complex samples has some similarity to the popular 2D techniques currently used in biochemistry, such as a combination of IEF and SDS/PAGE, but offers a different combination of orthogonal separation methods. Such maps are readily available from the data generated by an HPLC/MS analysis and can give valuable information about glycosylation patterns and product consistency.

This new analytical method can be used to deduce possible carbohydrate structures by determining both the mass and elution position of individual glycopeptides.

RPHPLC/ESIMS has greatly expanded the power of peptide mapping to identify carbohydrate structures that are attached to asparagine, serine, or threonine

Table 2 Comparison between experimental and computed values of the molecular masses (Da) of the protein constituents of the PSII major and minor antenna system

Antenna proteins	SDS/PAGE: apparent molecular mass (Da)	Molecular mass (measured)	Molecular mass (expected)	
			Spinach	Other species[a]
Type 1 I	27 000–28 000	24 936	25 030	25 026 24 969 (M)
Type 1 II		24 942		25 041 24 906 (P)
Type 2	25 000–27 000	24 761		24 834 (T)
Type 3	24 000–25 000	24 323		24 285 (B) 24 308 (T)
CP 29	29 000–31 000	28 076		27 804 (B) 27 642 (T)?
CP 26	26 000–29 000	27 068		26 607 (M) 27 642 (T)?
CP 24	20 000–22 000	22 820	22 813	

[a] Plant species: P, petunia; M, maize; T, tomato; B, barley.

residues. One can use in-source CID to scan the map for regions with a high concentration of glycopeptides. Such information should prove invaluable in determining the role in biology of the carbohydrate moiety in glycoproteins and in reducing the approval barriers for the pharmaceutical use of glycosylated proteins produced by mammalian fermentation systems.

5.1.3 High-performance Liquid Chromatography Coupled On-line with Light Scattering, Ultraviolet Absorbance and Refractive Index Detectors

Since protein molecules self-associate to oligomers for specific purposes and we often have no idea whether the protein exists in solution as a monomer, dimer, or other oligomer, the possibility of determining the molecular weights of proteins or their complexes is an important step in understanding proteins and their functions. In addition to HPLC coupled on-line with a mass spectrometer, techniques of using SEC with on-line light-scattering, UV absorbance, and refractive index detectors have recently been reported.[71] Conventional SEC is a simple and fast method for estimating the molecular mass of a protein in its native form based on its elution position. However, there are several problems with this simple SEC approach. One is that the elution position depends not only on the molecular mass of the protein but also on its shape. Another problem is that the elution position will change if the protein has any tendency to interact with the column matrix. In addition, when a protein or a protein complex contains carbohydrates, the carbohydrates usually have a disproportionately large effect on its elution position, so SEC may not be able to determine its polypeptide molecular mass. However, as reported by Wen et al.,[71] by using two or more detectors proteins and glycoproteins may be easily determined.

5.1.4 Affinity Chromatography and Biosensors

Affinity chromatography, which in the past yielded success as a preparative tool, has stimulated the development of analytical applications for biomolecular recognition and most recently of molecular recognition biosensors. Affinity-based methods on other solid-phase surfaces such as blots and microliter plates have further expanded affinity-based technology. The principles of analytical affinity chromatography recently have been adapted for a nonchromatographic affinity technology, namely real-time optical biosensors for interaction analysis. As a real-time method, biosensors offer the opportunity to measure not only the equilibrium affinity constant but also the on and off rate constants for interactions of biological macromolecules.[72]

5.2 Lipoproteins

CCC is a generic name for various liquid–liquid partition chromatographic methods which are used without solid support matrices.[73] The stationary phase is retained in the column with the aid of gravity or centrifugal force. In this way the system eliminates all complications arising from the solid support. Previous studies have shown that a mixture of HDL and LDL fractions prepared by ultracentrifugation could be separated by *x*-axis ultracentrifugation.[74] Recently the complementary use of CCC was attempted for the separation of three main classes of lipoproteins. Consequentely, the fractionation of HDLs, VLDLs and LDLs may be performed by the combined use of polymer-phase CCC and hydroxyapatite chromatography without prior ultracentrifugation.[75]

In the last decade, CE has emerged as a powerful separation technique for separation of plasma apolipoproteins.[76] By adding the surfactant SDS to the separation buffer, the main apolipoproteins of HDL (apoA-I and apoA-II) and LDL (apo-B) may be separated in a single run in less than 12 min. The CE separation of VLDL apolipoproteins has also been studied. High-performance capillary isotachophoresis (ITP) of lipoproteins has been reported, which provided for rapid screening of lipoprotein abnormalities.[77]

6 COMPARISON WITH OTHER METHODS

To summarize, it is not surprising that HPLC techniques, given the wide versatility, relative ease of use and high resolution, may be considered the most valuable tool for the characterization of virtually any hydrophobic protein. New supports offer a good compromise between mechanical stability and nonspecific binding, and an excellent recovery of the protein loaded. Moreover, these matrices also have excellent chemical stability within the working range of biological separations. In this context, the use of a reversed phase C-4 column, where the interaction of proteins with the matrix is very soft, has allowed a good separation and the complete recovery of proteins when applied to the separation of photosynthetic apparatus.[33,38] In this particular case, where the protein molecular masses are similar and the amino acid sequence is sufficiently conserved, electrophoresis and immunoblotting appeared expensive and technically demanding. Moreover, for quantitative analysis, traditional approaches by SDS/PAGE are not only cumbersome but also rather ineffective for evaluating differences in the relative quantity of each component unless time-consuming antibody titration is used.[78]

Table 3 Comparison of the advantages and disadvantages of separations performed by HPLC and other methods

Parameter	RPHPLC	CZE	IEF	SDS/PAGE
Analysis time	30 min	10 min	Hours	18–24 h
Reproducibility	Excellent	Reasonable	Operator dependent	Operator dependent
Sample amount	1–50 μL	1–5 μL	1–100 μL	1–100 μL
Resolution	Good	Good	Excellent	Excellent
Sensitivity	Nanograms	Picograms	Micrograms	Micrograms
Quantitative analysis	Excellent	Good	Reasonable	Reasonable
Automation	Yes	Yes	No	No

CZE, capillary zone electrophoresis.

In Table 3 are summarized the advantages and disadvantages of separations performed by HPLC with others alternative separation systems such as SDS/PAGE, IEF, and CE. The number of reports dealing with the separation of membrane proteins by high-performance CE is still rather small. First results indicate that virtually all methods used for the separation of water-soluble proteins can also be applied to the separation of hydrophobic membrane proteins. Apart from the guidelines mentioned above, the know-how acquired in high-performance CE and in the preparative free-flow (ITP) of serum lipoproteins can be transferred to the separation of membrane proteins.[79] Recently, we have demonstrated that CZE can be successfully applied for the complete resolution of the LHCII thylakoid membrane proteins by using the neutral detergent OD at a concentration lower than its critical micelle concentration in the electrolyte solution.[80] This method was revealed to be rapid and sensitive for identifying and determining quantitatively the several components of PSII, but the RPHPLC method,[33] in addition to being rapid, simple, and precise, has proven to be effective at detecting differences in the protein components of LHCII isolated from different plants that might not be evidenced by denaturing SDS/PAGE, as in the case of many species (data not shown). In addition, the possibility of separating all protein components of the PSII major and minor antenna system in samples not subjected to sucrose-gradient ultracentrifugation, as in the case of injection of BBY, is expected to be advantageous for evaluating the relative content of the different protein components of PSII and their variation related to physiological adaptation to environmental conditions. The injection of BBY directly allows one to obtain an exact evaluation of the quantitative relationships between chlorophyll *a*/*b* binding present in PSII, avoiding fractionated separation, Coomassie Brilliant Blue staining, quantification by densitometry, and correction of the results according to the specific binding of Coomassie Brilliant Blue to isolated proteins, as required by SDS/PAGE. The use of this method in screening photosynthetic mutants and plants adapted to different environmental conditions will be useful in the elucidation of the composition and supramolecular organization of LHCII and will possibly increase the understanding of the molecular mechanisms underlying the physiological adaptations.

As shown in the application examples in section 4, all HPLC methods developed for membrane protein separations offer the further advantage of analyzing the content of each HPLC peak by a multidimensional approach. A single peak observed in the RPHPLC separation of a protein is not an indication that the protein is highly pure, because it is common for related substances to co-elute with the parent protein. The main peak can be isolated and reinjected into an alternative separation system such as an SDS/PAGE, IEF, CE, or dissimilar HPLC system, taking appropriate measures to avoid decomposition of the protein after collection from the HPLC system. If the HPLC solvent system is not compatible with the second separation system, or if suitable stability of the isolated main peak is not attainable, then the nonfractionated protein product can be subjected to one of the alternative separation procedures. On-line spectroscopic monitoring using diode-array detection can be of value in establishing peak purity in favorable cases, but usually protein spectra are too similar to provide much discrimination.

The RPHPLC/ESIMS method holds several advantages over SDS/PAGE, the conventional technique for studying membrane proteins, including better protein separation, mass accuracy, speed and efficiency. Our study has shown that RPHPLC/ESIMS is an effective method for separating and characterizing the integral membrane proteins comprising the PSII major and minor antenna system, both as isolated complexes by sucrose-gradient ultracentrifugation and as the BBY grana membrane preparation directly. In accordance with molecular genetic data reported in the literature, showing

that higher plants have several *Lhcb1* genes encoding different type I proteins for each species, two type I proteins of similar molecular mass have been resolved in spinach leaves. The experimental data are in good agreement with the molecular masses of the individual antenna proteins calculated on the basis of their nucleotide-derived amino acid sequences. In addition, the RPHPLC/ESIMS method allows the separation of protein constituents of the major and minor antenna system which are not resolved by conventional SDS/PAGE methods. Other advantages of RPHPLC/ESIMS over SDS/PAGE include the accuracy in determining the molecular mass and the higher speed and efficiency.

The use of chromatographic methods offers the possibility of combining more techniques, and therefore both SEC and RPHPLC methods will be employed to provide high-resolution, quantitative data, in addition to various spectroscopic, electrophoretic, and immunochemical procedures for more specialized purposes. Conventional electrophoretic techniques often do not directly add to the information obtained, because HPLC frequently offers better resolution of related substances. However, a key advantage of conventional electrophoretic techniques is the ability to detect virtually any protein, thereby ensuring that any unanticipated protein impurities would be detected. In developing the HPLC methods to be used in the battery of tests, one must also consider the information obtained from the other techniques and seek to understand the relationship between these sets of information. The objective of such an assessment is to obtain the maximum amount of relevant information consistent with an efficient use of analytical resources.

Concerning lipoproteins separation, it was reported in section 5 that by using CE, high separation efficiencies and short analysis times for lipoproteins can be achieved. CE exhibits advantages over HPLC for lipoprotein analysis: CE separations are faster and do not require the use of expensive columns, and CE does not suffer from slow mass transfer rates which lead to band broadening in HPLC separations of apolipoproteins. The main advantage of CE over conventional electrophoresis is speed and instrumental format, which eliminates the need for labor-intensive steps such as gel preparation and staining.

ACKNOWLEDGMENTS

I am grateful to former and present members of the research group who have contributed to much of the work presented in this review, and to Dr Anna Maria Timperio and Dr Sara Rinalducci for valuable assistance in the preparation of the manuscript. This work was supported by CE Project CIPA CT93 0202 and COST Contract ERB IC15CT 980126.

ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immune Deficiency Syndrome
BLM	Perfusion Planar Lipid Membrane
CCC	Countercurrent Chromatography
CE	Capillary Electrophoresis
CHAPS	3-[(3-Cholamidopropyl)dimethyl-ammonio]-l-propanesulfonate
CHD	Coronary Heart Disease
CID	Collision Induced Dissociation
CPG	Glyceryl Controlled-pore Glass
CZE	Capillary Zone Electrophoresis
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Ethylene Glycol Bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Assay
ESIMS	Electrospray Ionization Mass Spectrometry
FPLC	Fast Protein Liquid Chromatography
GPC	Gel Permeation Chromatography
HDL	High-density Lipoprotein
HEPES	*N*-(2-Hydroxyethyl)piperazine-*N*-(2-ethanesulfonic acid)
HIV	Human Immunodeficiency Virus
HLP	Hyperlipoproteinemic
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
IA	Immunoaffinity
IEF	Isoelectric Focusing
ITP	Isotachophoresis
LDL	Low-density Lipoprotein
LHCI	Light-harvesting Complex of Photosystem I
LHCII	Light-harvesting Complex of Photosystem II
MAb	Monoclonal Antibody
MES	Morpholinoethanesulfonic Acid
MS	Mass Spectrometry
NADP	Nicotinamide Adenine Dinucleotide Phosphate
OD	n-Octyl β-D-Glucopyranoside
PSI	Photosystem I
PSII	Photosystem II

RPHPLC	Reversed-phase High-performance Liquid Chromatography
RPHPLC/ESIMS	Reversed-phase High-performance Liquid Chromatography/ Electrospray Mass Spectrometry
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size-exclusion Chromatography
SEHPLC	Size-exclusion High-performance Liquid Chromatography
TFA	Trifluoroacetic Acid
TRIS	Tris(hydroxymethyl)-aminomethane
UV	Ultraviolet
VLDL	Very Low-density Lipoprotein
2D	Two-dimensional
β-DM	*n*-Dodecyl β-D-Maltoside

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Carbohydrate Analysis **(Volume 1)**
Glycoprotein Analysis: General Methods

Clinical Chemistry **(Volume 2)**
Biochemical Markers of Acute Coronary Syndromes • Capillary Electrophoresis in Clinical Chemistry • Lipid Analysis for Important Clinical Conditions • Phospholipids of Plasma Lipoproteins, Red Blood Cells and Atheroma, Analysis of • Serum Proteins

Particle Size Analysis **(Volume 6)**
Centrifugation in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Gel Electrophoresis in Protein and Peptide Analysis • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Capillary Electrophoresis • Gradient Elution Chromatography • Ion Chromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

REFERENCES

1. P. Larsson, 'Chromatographic Matrices in Affinity Chromatography', in *Handbook of Affinity Chromatography, Chromatographic Science Series*, Vol. 63, ed. T. Kline, Marcel Dekker, New York, 61–75, 1993.
2. M. Wilchek, T. Miron, J. Kohn, 'Affinity Chromatography', *Methods Enzymol.*, **104**, 3–55 (1984).
3. C. Jones, A. Patel, S. Griffln, J. Martin, P. Young, K. O'Donnell, C. Silverman, T. Porter, I. Chaiken, 'Current Trends in Molecular Recognition and Bioseparation', *J. Chromatogr. A*, **707**, 3–22 (1995).
4. P. Lundahl, Q. Yang, 'Liposome Chromatography: Liposomes Immobilized in Gel Beads as a Stationary Phase for Aqueous Column Chromatography', *J. Chromatogr.*, **544**(2), 283–304 (1991).
5. A. Lundqvist, P. Lundahl, 'Chromatography on Cells and Biomolecular Assembles', *J. Chromatogr. B*, **699**, 209–220 (1997).
6. I.M. Chaiken, 'Separation of Receptors by Affinity Chromatography', in *Handbook of Affinity Chromatography, Chromatographic Science Series*, Vol. 63, ed. T. Kline, Marcel Dekker, New York, 90–105, 1993.
7. D. Josic, J. Reusch, K. Loster, O. Baum, W. Reutter, 'High-performance Membrane Chromatography of Serum and Plasma Membrane Proteins', *J. Chromatogr.*, **590**(1), 59–76 (1992).
8. J.T. Skare, E.S. Shang, D.M. Foley, D.R. Blanco, C. Champion, T. Mirzabekov, Y. Sokolov, B.L. Kagan, J.N. Miller, M.A. Lovett, 'Virulent Strain Associated Outer Membrane Proteins of *Borrelia burgdorferi*', *J. Clin. Invest.*, **96**(5), 2380–2392 (1995).
9. T.A. Mirzabrkov, A.Y. Silberstein, B.L. Kagan, 'Use of Planar Lipid Bilayer Membranes for Rapid Screening of Membrane Active Compounds', *Methods Enzymol.*, **271**, 661–674 (1996).
10. J.T. Skare, T.A. Mirzabekov, E.S. Shang, D.R. Blanco, H. Erdjument-Bromage, J. Bunikis, S. Bergstrom, P. Tempst, B.L. Kagan, J.N. Miller, M.A. Lovett, 'The Oms66 (p66) Protein is a *Borrelia burgdorferi* Porin', *Infect. Immun.*, **65**(9), 3654–3661 (1997).
11. T. Mirzabekov, M.C. Lin, W.L. Yuan, P.J. Marshall, M. Carman, K. Tomaselli, I. Lieberburg, B.L. Kagan, 'Channel Formation in Planar Lipid Bilayers by a Neurotoxic Fragment of the Beta-amyloid Peptide', *Biochem. Biophys. Res. Commun.*, **202**(2), 1142–1148 (1994).

12. C.G. Helleqvist, B.J. Sweetman, 'Biomedical Application of Mass Spectroscopy', *Methods Biochem. Anal.*, **34**, 97–143 (1990).
13. L.J. Sernan, J.R. McNannara, E.J. Schaefer, 'Lipoprotein(a), Homcystein, and Remnantlike Particles: Emerging Risk Factors', *Curr. Opin. Cardiol.*, **14**, 186–191 (1999).
14. N. Rifai, J.F. Chapman, L.M. Silverman, J.T. Gwynnes, 'Review of Serum Lipids and Apolipoproteins in Risk-assessment of Coronary Heart Disease', *Ann. Clin. Lab. Sci.*, **18**(6), 429–439 (1988).
15. Y. Shibusawa, 'Lipoproteins: Comparison of Different Separation Strategies', *J. Chromatogr. B*, **699**, 419–437 (1997).
16. T. Sata, D.L. Estrich, P.D.S. Wood, L.W. Kinsell, 'Evaluation of Gel Chromatography for Plasma Lipoprotein Fractionation', *J. Lipid Res.*, **11**(4), 331–340 (1970).
17. W. Marz, R. Skeimer, H. Scharnagi, U.B. Seiffert, W. Gross, 'Fast Lipoprotein Chromatography: New Method of Analysis for Plasma Lipoproteins', *Clin. Chem.*, **39**(11), 2276–2281 (1993).
18. P.M. Clifton, A.M. MacKinnon, P.J. Barter, 'Gel Permeation Chromatography of Lipoproteins', *J. Chromatogr.*, **414**, 25–34 (1987).
19. H.M. Wilson, B.A. Griffin, C. Watt, E.R. Skinner, 'The Isolation and Characterization of High-density-lipoprotein Subfractions Containing Apolipoproteins E from Human Plasma', *Biochem. J.*, **284**(2), 477–482 (1992).
20. T. O'Brìen, J. Buithieu, T.T. Nguyen, L. Klein, N. Bren, M. Wentworth, B.J. Hallaway, B.A. Kottke, 'Separation of High-density Lipoproteins into Apolipoprotein E-poor and Apolipoprotein E-rich Subfractions by Fast Protein Liquid Chromatography Using a Heparin Affinity Column', *J. Chromatogr.*, **613**(2), 239–246 (1993).
21. B. Vedie, I. Myara, M.A. Pech, I.C. Maziere, C. Maziere, A. Capranì, N. Moatti, 'Fractionation of Charge-modified Low Density Lipoproteins by Fast Protein Liquid Chromatography', *J. Lipid Res.*, **32**(8), 1359–1369 (1991).
22. J.A.K. Harmony, E.I. Cordes, 'Interaction of Human Plasma Low Density Lipoprotein with Concanavalin A and with Ricin.', *J. Biol. Chem.*, **250**(22), 8614–8617 (1975).
23. J.W. McLean, N.A. Elshourbagy, D.J. Chang, R.W. Mahley, J.M. Taylor, 'Human Apolipoprotein E mRNA. cDNA Cloning and Nucleotide Sequencing of a New Variant', *J. Biol. Chem.*, **259**(10), 6498–6504 (1984).
24. P.M. Young, T.M. Boehrn, 'Rapid and Efficent Separation of Apolipoproteins', *J. Chromatogr. A*, **31**, 79–85 (1984).
25. T.A. Hughes, M.A. Moore, P. Neame, M.F. Medley, B.H. Cheung, 'Rapid Quantitative Apolipoprotein Analysis by Gradient Ultracentrifugation and Reversed-phase High Performance Liquid Chromatography', *J. Lipid. Res.*, **29**(3), 363–376 (1988).
26. I. Sekigawa, J.E. Groopmen, J.D. Allan, K. Ikeuchi, G. Biberfield, K. Takatsuki, R.A. Byrn, 'Characterization of Autoantibodies to the CD4 Molecule in Human Immunodeficiency Virus Infection', *Clin. Immunol. Immunopathol.*, **58**(1), 145–153 (1991).
27. P.A. Well, B. Beiderman, R.L. Garlick, S.B. Lyle, J.P. Martin, Jr, J.T. Rerberg, H.F. Meyer, S.L. Henderson, F.M. Eckenrode, 'Large-scale Immunoaffinity Purification of Recombinant Soluble Human Antigen CD4 from *Escherichia coli* Cells', *Biotechnol. Appl. Biochem.*, **18**, 341–357 (1993).
28. B.S. Welinder, H.F. Sorensen, B. Hansen, 'High-performance Liquid Chromatographic Separation of Membrane Proteins Isolated from Erythrocyte Ghosts', *J. Chromatogr.*, **462**, 255–268 (1989).
29. B. Hankamer, J. Barber, E.J. Boekema, 'Structure and Membrane Organization of Photosystem II in Green Plants', *Annu. Rev. Plant Physiol. Mol. Biol.*, **48**, 641–671 (1997).
30. B.R. Green, D. Shen, R. Aebersold, E. Pichersky, 'Identification of the Polypeptides of the Major Light-harvesting Complex of Photosystem II (LHCII) with Their Genes in Tomato', *FEBS Lett.*, **305**(1), 18–22 (1992).
31. S. Jansson, 'The Light-harvesting Chlorophyll *a*/*b* Binding Proteins', *Biochim. Biophys. Acta*, **1184**, 1–19 (1994).
32. R. Bassi, F. Rigoni, R. Barbato, G.M. Giacometti, 'Light-harvesting Chlorophyll *a*/*b* Protein (LHCI) Populations in Phosphorylated Membranes', *Biochim. Biophys. Acta*, **936**, 29–38 (1988).
33. L. Zolla, M. Bianchetti, A.M. Timperio, D. Corradini, 'Rapid Resolution by Reversed-phase High-performance Liquid Chromatography of the Thylakoid Membrane Proteins of the Photosystem II Light-harvesting Complex', *J. Chromatogr. A*, **779**, 131–138 (1997).
34. D.A. Berthold, G.T. Babcock, C.A. Yocum, 'A Highly Resolved Oxygen-evolving Photosystem II Preparation from Spinach Thylakoid Membranes', *FEBS Lett.*, **134**, 231–234 (1981).
35. R.J. Porra, W.A. Thompson, P.E. Kriedemann, 'Determination of Accurate Extinction Coefficients and Simultaneous Equations for Assaying Chlorophyll *a* and *b* Extracted with Four Different Solvents: Verification of the Concentration of Chlorophyll Standards by Atomic Absorption Spectroscopy', *Biochim. Biophys. Acta*, **975**, 384–394 (1989).
36. S. Jansson, B. Andersen, H.V. Scheller, 'Nearest-neighbor Analysis of Higher-plant Photosystem I Holocomplex', *Plant Physiol.*, **112**(1), 409–420 (1996).
37. R. Bassi, P. Dainese, 'A Supramolecular Light-harvesting Complex from Chloroplast Photosystem-II Membranes', *Eur. J. Biochem.*, **204**, 317–326 (1992).
38. L. Zolla, A.M. Timperio, M.G. Testi, M. Bianchetti, R. Bassi, D. Corradini, 'Isolation, Characterization and Subunits Stoichiometry of Chloroplast Photosystem II Antenna of Spinach by Reversed-phase Liquid Chromatography', *Photosynth. Res.*, **63**, 281–290 (1999).

39. H. Schagger, G. von Jagow, 'Tricine–Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis for the Separation of Proteins in the Range 1 to 100 kDa', *Anal. Biochem.*, **166**, 368–379 (1987).
40. M.L. Di Paolo, A. Peruffo Dal Belin, R. Bassi, 'Immunological Studies on Chlorophyll *a*/*b* Proteins and Their Location in Chloroplast Membrane Domains', *Planta*, **191**, 275–286 (1990).
41. R. Bassi, F. Rigoni, G.M. Giacometti, 'Chlorophyll Binding Proteins with Antenna Function in Higher Plants and Green Algae', *Photochem. Photobiol.*, **52**, 1187–1206 (1990).
42. D. Corradini, C.G. Huber, A.M. Timperio, L. Zolla, 'Use of Reversed-phase Liquid Chromatography with Electrospray Mass Spectrometric Detection for the Resolution and Identification of the Protein Components of the Photosystem II Antenna System of Higher Plants', *J. Chromatogr.*, submitted.
43. M. Sigrist, L.A. Staehelin, 'Identification of Type 1 and Type 2 Light-harvesting Chlorophyll *a*/*b*-binding Proteins Using Monospecific Antibodies', *Biochim. Biophys. Acta*, **1098**, 191–200 (1992).
44. P. Dainese, R. Bassi, 'Subunit Stoichiometry of the Chloroplast Photosystem II Antenna System and Aggregation State of the Component Chlorophyll *a*/*b* Binding Proteins', *J. Biol. Chem.*, **266**(13), 8136–8142 (1991).
45. N. Krauss, W. Hinrichs, I. Witt, P. Fromme, M. Pritzkow, Z. Dauter, C. Betzel, K.S. Wilson, H.T. Witt, W. Saenger, 'Three-dimensional Structure of System I of Photosynthesis at 6 Å Resolution', *Nature (London)*, **361**, 326–331 (1993).
46. P.R. Chitnis, Q. Xu, V.P. Chitnis, R. Nechushtai, 'Function and Organization of Photosystem Polypeptides', *Photosynth. Res.*, **44**(1), 23–40 (1995).
47. R. Bassi, G. Hoyer-Hansen, R. Barbato, G.M. Giacometti, D.J. Simpson, 'Chlorophyll Proteins of the Photosystem II Antenna System', *J. Biol. Chem.*, **262**(27), 13 333–13 341 (1987).
48. A.M. Timperio, R. Croce, L. Zolla, 'Separation of Proteins from Photosysthem I by High Performance Liquid Chromatography in Higher Plants', in preparation.
49. L. Zolla, A.M. Timperio, 'Electrospray–Mass Spectrometer On-line with Reversed-phase Liquid Chromatography for the Simultaneous Resolution and Identification of Intrinsic Thylakoid Membrane Proteins', *Proteins*, submitted.
50. I. van Gent, A. van Tol, 'Automated Gel Permeation Chromatography of Plasma Lipoproteins by Preparative Fast Protein Liquid Chromatography', *J. Chromatogr.*, **525**(2), 433–441 (1990).
51. I. Hara, M. Okazaki, 'High-performance Liquid Chromatography of Serum Lipoproteins', *Methods Enzymol.*, **129**, 57–78 (1986).
52. U. Matsumoto, H. Nakayama, Y. Shibusawa, T. Niimura, 'Separation of Human Serum Lipoproteins into Three Major Classes by Hydroxyapatite Chromatography', *J. Chromatogr.*, **566**(1), 67–76 (1991).
53. W. Palinski, M.E. Rosenfeld, S. Yla-Herttuala, G.C. Gurtner, S.S. Socher, S.W. Butler, S. Parthasarathy, T.E. Carew, D. Steinberg, J.L. Witztum, 'Low Density Lipoprotein Undergoes Oxidative Modification in Vivo', *Proc. Natl. Acad. Sci. USA*, **86**(4), 1372–1376 (1989).
54. A. Carpenter, W.C. Purdy, 'Preparation of Heparinglyceryl Controlled-pore Glass Affinity Media for the Separation of Alpha- and Beta-lipoproteins', *J. Chromatogr.*, **573**(1), 132–135 (1992).
55. G. Steiner, 'Triglyceride-rich Lipoproteins and Atherosclerosis, from Fast to Feast', *Ann. Med.*, **25**(5), 431–451 (1993).
56. I.J. Goldberg, W.D. Wagner, L. Pang, L. Paka, L.K. Curtiss, J.K. DeLozier, G.S. Shelness, C.S.H. Young, S. Pillarisetti, 'The NH_2-terminal Region of Apolipoprotein B is Sufficient for Lipoprotein Association with Glycosaminoglycans', *J. Biol. Chem.*, **273**, 35 355–35 361 (1998).
57. R.W. James, A. Proudfoot, D. Pometta, 'Immunoaffinity Fractionation of High-density Lipoprotein Subclasses 2 and 3 Using Anti-apolipoprotein A-I and A-II Immunosorbent Gels', *Biochim. Biophys. Acta*, **1002**(3), 292–301 (1989).
58. E. Koren, P. Alaupovic, D.M. Lee, N. Dashti, H.U. Kloer, G. Wen, 'Selective Isolation of Human Plasma Low-density Lipoprotein Particles Containing Apolipoproteins B and E by Use of a Monoclonal Antibody to Apolipoprotein B', *Biochemistry*, **26**(10), 2734–2740 (1987).
59. M.C. Cheung, 'Methods for Separation of Apolipoproteins', *Methods Enzymol.*, **129**, 130–151 (1986).
60. C. Edeistein, in *Analysis of Oils, Fats and Lipoproteins*, ed. E.G. Perkins, American Oil Chemists Society, 589–603, 1991.
61. M. Okazaki, M. Kinoshita, I. Hara, 'Improved High-performance Liquid Chromatographic Method for the Determination of Apolipoproteins in Serum High-density Lipoproteins', *J. Chromatogr.*, **430**(1), 135–142 (1988).
62. G.L. Mill, P.A. Lane, P.K. Weech, *A Guide Book to Lipoprotein Techniques: Techniques in Biochemistry and Molecular Biology*, Vol. 14, Elsevier, Amsterdam, 1984.
63. H. Mezdour, V. Clavey, I. Kora, M. Koffigan, A. Barkia, J.C. Fruchart, 'Anion-exchange Fast Protein Liquid Chromatographic Characterization and Purification of Apolipoproteins A-I, A-II, C-I, C-II, C-III0, C-III1, C-III2 and E from Human Plasma', *J. Chromatogr.*, **414**(1), 35–45 (1987).
64. G. Hocke, H. Kaffarnik, G. Minscher, A. Steinemetz, 'Purification of Human Serum Amyloid A by Anion-exchange Fast Protein Liquid Chromatography', *J. Chromatogr.*, **526**(1), 203–209 (1990).
65. R. Weinberg, C. Patton, B. DaGue, 'Analytic and Preparative Separation of Apolipoproteins A-I, A-II, and A-IV by Reverse Phase High Pressure Liquid Chromatography', *J. Lipid Res.*, **29**(6), 819–824 (1988).

66. B. Meyer, E. Kecorius, P. Barter, N. Fidge, T. Tetaz, 'Application of Reversed-phase High-performance Liquid Chromatography to the Separation of Apolipoproteins A-IV, A-I and E from Rat High-density Lipoprotein', *J. Chromatogr. A.*, **540**, 386–391 (1991).
67. M.L. Zeidel, S. Nielsen, B.L. Smith, S.V. Ambudkar, A.B. Maunsbach, P. Agre, 'Ultrastructure, Pharmacological Inhibition and Transport Selectivity of Aquaporin Channel-forming Integral Membrane Protein in Proteoliposomes', *Biochemistry*, **33**, 1606–1611 (1994).
68. J.P. Whitelegge, B.C. Gundersen, K.F. Faull, 'Electrospray-ionization Mass Spectrometry of Intact Intrinsic Membrane Proteins', *Protein Sci.*, **7**, 1423–1430 (1998).
69. L. Zolla, C.H. Huber, A.M. Timperio, M. Bianchetti, D. Corradini, 'Use of RPHPLC with Electrospray Mass Spectrometric Detection for the Resolution and Identification of the Protein Components of PSII of Wheat', in *Photosynthesis: Mechanisms and Effects*, ed. L. Garab, Kluwer, Dordrecht, **V**, 4377–4380 (1998).
70. W. Hancock, A. Apffel, J. Chakel, C. Souders, T.M. Timkulu, E. Pungor, Jr, A.W. Guzzetta, 'Reversed-phase Peptide Mapping of Glycoproteins Using Liquid Chromatography/Electrospray lonization Mass Spectrometry', *Methods Enzymol.*, **271**, 403–427 (1996).
71. J. Wen, T. Arakawa, J.S. Philo, 'Size-exclusion Chromatography with On-line Light-scattering, Absorbance, and Refractive lndex Detectors for Studying Proteins and Their lnteractions', *Anal. Biochem.*, **2401**, 155–166 (1996).
72. U. Jonsson, L. Fagerstam, B. Ivarsson, B. Johnsson, R. Karlsson, K. Lundh, S. Lofas, B. Persson, H. Roos, I. Ronnberg, 'Real-time Biospecific Interaction Analysis Using Surface Plasmon Resonance and Sensor Chip Technology', *Biotechniques*, **11**(5), 620–627 (1991).
73. Y. Ito, in *Counter-current Chromatography: Theory and Practice*, eds. N.B. Mandava, Y. Ito, Marcel Dekker, New York, 1988.
74. Y. Shibusawa, Y. Ito, K. Ikewaki, D.J. Rader, B. Brewer, 'Counter-current Chromatography of Lipoproteins with a Polymer Phase System Using the Cross-axis Synchronous Coil Planet Centrifuge', *J. Chromatogr.*, **596**(1), 118–122 (1992).
75. Y. Shibusawa, M. Magiyama, U. Matsumoto, Y. Ito, 'Complementary Use of Counter-current Chromatography and Hydroxyapatite Chromatography for the Separation of Three Main Classes of Lipoproteins from Human Serum', *J. Chromatogr. B*, **664**(2), 295–301 (1995).
76. T. Tadey, W.C. Purdy, 'Effect of Detergents on the Electrophoretic Behaviour of Plasma Apolipoproteins in Capillary Electrophoresis', *J. Chromatogr. A*, **652**(1), 131–138 (1993).
77. G. Schmitz, U. Borgmann, G. Assman, 'Analytical Capillary Isotachophoresis: A Routine Technique for the Analysis of Lipoproteins and Lipoprotein Subfractions in Whole Serum', *J. Chromatogr.*, **320**(1), 253–262 (1985).
78. R. Bassi, J. Marquardt, J. Lavernie, 'Biochemical and Functional Properties of Photosystem II in a Agranal Membranes from Maize Mesophyll and Bundle Sheath Chloroplasts', *Eur. J. Biochem.*, **233**(3), 709–719 (1995).
79. G. Schmitz, C. Mollers, 'Analysis of Lipoproteins with Analytical Capillary Isotachophoresis', *Electrophoresis*, **15**(1), 31–39 (1994).
80. L. Zolla, M. Bianchetti, A.M. Timperio, G. Scarascia Mugnozza, D. Corradin, 'Capillary Electrophoresis of Closely Related Intrinsic Membrane Proteins of the Photosystem II Light-harvesting Complex (LHC II)', *Electrophoresis*, **17**, 1597–1601 (1996).

Electron Spin Resonance Spectroscopy Labeling in Peptide and Protein Analysis

Peter G. Fajer
Florida State University, Tallahassee, USA

1 Introduction **2**
2 Historical Perspective **2**
3 Sample Preparation **3**
3.1 Nitroxide Spin Labels 3
3.2 Labeled Sites 3
3.3 Attachment Rigidity 4
3.4 Impairment of Function/Structure of Labeled Proteins 4
4 Techniques and Instrumentation **4**
4.1 Continuous Wave Electron Spin Resonance 4
4.2 Saturation Transfer Electron Spin Resonance 8
4.3 Time Domain Methods 8
5 Applications of Spin Labeling **10**
5.1 Protein Orientation 10
5.2 Protein Dynamics 13
5.3 Kinetic Experiments 20
5.4 Protein Folding 21
5.5 Ligand Binding 22
5.6 Distance Measurements 22
5.7 Structural Biology 26
6 Conclusion **30**
Acknowledgments **30**
List of Symbols **30**
Abbreviations and Acronyms **31**
Related Articles **32**
References **32**

Electron spin resonance (ESR) is a powerful analytical tool used in protein and peptide biochemistry. It is used in the determination of secondary, tertiary and quaternary protein structure and associated conformational changes. Protein dynamics and the relative orientation of protein components in ordered systems can also be measured. The majority of proteins do not contain unpaired electrons whose spin transitions give rise to an ESR signal, hence necessitating the use of extrinsic probes called spin labels. Spin labels are nitroxide derivatives with a stable unpaired electron and a functional group for specific attachment to the protein (covalent or as a ligand). The most popular covalent sites are cysteine residues, which, if necessary, can be introduced into the protein structure using molecular biology techniques.

The physical basis for nearly all ESR applications is the anisotropy of the nitroxide signal and the sensitivity of the ESR spectra to various relaxation pathways. The interaction between an electron of a spin label and an external magnetic field depends on their relative orientations. The splitting and the center of ESR spectra of an oriented sample are used to determine the orientation of labeled domains. For samples with little disorder the orientational sensitivity is better than 1°. The width of the signal is proportional to the orientational disorder, which is used to measure conformational heterogeneity of proteins.

If the spin label reorientates itself on the ESR timescale (nanoseconds) then the spectral anisotropy is averaged. The extent of averaging defines the ESR line shape which is used to determine the rotational rate and anisotropy of motion. The dynamic range of ESR is very broad, rotational correlation times range from 10^{-12} to 10^{-7} s for conventional ESR and the sensitivity can be extended to slower motions (10^{-3} s) with nonlinear saturation transfer electron spin resonance (STESR). Protein (spin label) mobility is used to follow conformational changes, steric restrictions on the spin label and the formation of large complexes.

Spin labels are also sensitive to the presence of other paramagnetic species. Collisions with water and lipid-soluble relaxing agents provide additional relaxation pathways measured by changes in relaxation times. The probability of these collisions reflects the accessibility of a spin label to the relaxant. The periodic patterns along the polypeptide chain of this accessibility are used to determine the secondary and tertiary structure of proteins. In the presence of another bound spin label or a paramagnetic metal complexed by histidine residues, spectra become broadened by dipolar or exchange interactions. Both mechanisms depend on the distance between the paramagnetic centers. Thus ESR can be used to determine intra- and intermolecular distances. The range of sensitivity is 5–25 Å and there are intensive efforts to increase the upper range to >50 Å. ESR as a spectroscopic ruler is used in protein structure determination and the investigation of macromolecular assembly processes and protein folding.

The foremost limitation of spin labeling ESR is the necessity to modify a protein with a spin probe. In some cases, the spin labels may perturb protein function and therefore cannot be used for spectroscopy. However, even an unsuccessful modification that results in functional

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

loss identifies functional regions of proteins and as such represents successful ''mutational analysis'' experiments.

1 INTRODUCTION

A spinning electron orbiting around a nucleus is a magnetic dipole. When placed in an external magnetic field, the dipole aligns parallel or antiparallel with the external field. These two orientations of the magnet represent two energy levels, with the difference in energy levels of the electron spin proportional to the strength of the magnetic field. The electron can be excited from one level (i.e. parallel dipole orientation) to another (antiparallel orientation) by an oscillating magnetic field. The energy of the oscillating field has to match the energy difference between the two levels. For a free electron in a magnetic field with a strength of a few hundred gauss, the frequency range of the exciting field is in the microwave region of the electromagnetic wave spectrum. The resonance between the orbiting electron and the microwave field forms the basis of ESR, also known as electron paramagnetic resonance or electron magnetic resonance.

ESR is commonly used to investigate protein and peptide structure, particularly studies of molecular orientation, protein dynamics and ligand binding. Observation of a resonance requires samples containing an unpaired electron, e.g. transition metals or organic radicals. Proteins and peptides are generally not paramagnetic and therefore require the use of extrinsic probes called spin labels. Spin labels are derivatives of nitroxides, small stable organic radicals, which are covalently attached to protein side chains or to metabolic substrates. In the last decade, the development of site-directed spin labeling (SDSL), which utilizes molecular biology to introduce new labeling sites, has established ESR as a protein structural determination technique. Patterns of side-chain mobility, accessibility to quenchers and the measurement of distances between spin labels have allowed the determination of the secondary, tertiary and quaternary structure of proteins.

This article is focused exclusively on spin labeling applications in protein and peptide biochemistry. The vast literature on metalloproteins, photosynthesis and reactive radicals in biology is not discussed here, and interested readers are directed to the many excellent reviews on these topics.[1–6]

2 HISTORICAL PERSPECTIVE

The first ESR experiments were performed by Zavoisky at the University of Kazan (Russia) during the Second World War.[7] Inspired by the experiments of Gorter[8] and Rabbi et al.[9] on paramagnetic relaxation and atomic beams, Zavoisky demonstrated resonance between microwaves and the precession of Cu^{2+} ions in a magnetic field. Resonance was observed as an absorption of microwaves whenever the frequency of the oscillating microwave field was equal to the ion precession frequency.

In the decade following the Second World War, ESR was the domain of physical chemists and physicists, with the first biological applications appearing in the mid-1950s. This early work included structural studies of metalloproteins,[10] measurement of free radicals in biological tissues,[11] carbonized carbohydrates,[12] and X-ray irradiated silk and hair.[13] Assenheim provides an excellent review of this early work with intrinsic ESR signals.[14]

In 1965, McConnell introduced extrinsic spin labels designed to label proteins. Using nitroxide derivatives first synthesized in Russia,[15,16] McConnell et al. demonstrated a helix–coil transition of a polylysine peptide.[17] Since then, ESR spin labeling has been used to study conformational changes in a number of proteins modified by maleimide nitroxides, which specifically target cysteine residues. However, reliance on the naturally occurring cysteine residue was a severe limitation. The SDSL strategy developed by Hubbell in 1989 employs molecular biology to introduce new cysteines for spin label attachment. The use of SDSL to scan the protein sequence with cysteines has stimulated the resurgence of ESR as a structural biology method.

The methodology of ESR was also undergoing an evolution. In 1957, Feher invented electron–nuclear double resonance (ENDOR) spectroscopy, a combination of both ESR and nuclear magnetic resonance (NMR),[18] in which nuclear spin transitions are observed indirectly by monitoring electron spin transitions. A few years later, electron–electron double resonance (ELDOR) spectroscopy was developed by Hyde et al.[19] and Benderskii et al.[20] which allowed the measurement of spectral diffusion between distinct spin populations. The development of spin-echo instruments by Mims et al.[21] introduced time-domain ESR in the 1960s. This was followed by Fourier transform electron spin resonance (FTESR), developed independently in the 1980s by Eliav and Freed,[22] Dinse et al.[23] and Bowman.[24] The first spin label applications appeared in 1986 when Gorester and Freed performed two-dimensional (2-D) FTESR experiments to measure spin dynamics.[25]

ESR moved towards high field (high frequency) with Lebedev et al.'s construction of a 150-GHz spectrometer,[26] followed by Freed et al.'s 250-GHz spectrometer, which was based on quasi-optics. The latter instrument was used extensively to investigate spin labels

in biological systems.[27] Ultra-high-field spectrometers operating at 550 GHz now exist and their application to nitroxide labels is only a question of time.[28]

Readers interested in the history of ESR are directed to a series of historic recollections of the ESR pioneers assembled by Eaton et al.[29]

3 SAMPLE PREPARATION

3.1 Nitroxide Spin Labels

Proteins are ESR silent, with the exception of metalloproteins, and must therefore be ''labeled'' with paramagnetic probes. These probes, or spin labels, are nitroxide derivatives containing an unpaired electron in the pπ orbital of the N–O bond (Figure 1a–c). The nitroxide radical is stable owing to the presence of methyl groups on neighboring carbon atoms. To limit flexibility, the NO group is enclosed in either a six-membered piperidine or a five-membered pyrrole ring. Pyrrole rings with an unsaturated bond are the least flexible.

The unpaired electron in the pπ orbital also interacts with the spin of the nitrogen nucleus, splitting the ESR signal into resonances corresponding to different nitrogen nuclear manifolds. Thus, the number of resonant peaks depends on the nitrogen isotope, three for ^{14}N and two for ^{15}N. ^{15}N labels have the advantage of less spectral dispersion which increases the signal amplitude 1.5-fold in conventional ESR and allows for full spectral coverage in FTESR. Reduction of the nuclear manifolds also simplifies the interpretation of nuclear relaxation and accelerates computer simulations of ESR line shapes. ^{15}N labels, however, are considerably more expensive than ^{14}N and only a handful of them are available commercially.

A weaker interaction occurs between the electron spin and the hydrogen nuclei of the ring and methyl groups. Each resonance peak is split by the nuclear spin, but the splittings are unresolved, resulting in a broad peak. The broadening can be removed by the substitution of hydrogen with deuterium which increases the peak height 1.5-fold for Gaussian and 5-fold for Lorentzian lines.

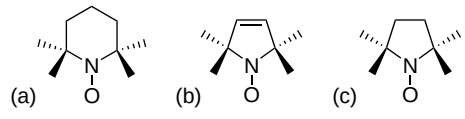

Figure 1 Commonly used nitroxides: (a) six-membered piperidine ring; (b) saturated five-membered pyrroline ring; (c) unsaturated pyrrolidine ring.

3.2 Labeled Sites

Nitroxide spin labels are used either covalently as modifiers of selected amino acids or noncovalently as analogs of substrates or enzymatic cofactors. The specificity of the label is conferred by the functional group attached to the nitroxide. For example, maleimide, iodoacetamide, indanedione and α-ketone groups attached to the nitroxide moiety target cysteine residues, while lysines are modified by activated esters in Figure 2(a–d). Attachment of the nitroxides by disulfide bonds allows for reversible modification. Reduction of the disulfide bonds with a mild reducing agent yields the unmodified protein. Bifunctional spin labels with two linker groups facilitate attachment to two sites on a protein, reducing probe mobility with respect to the protein. The ability to engineer neighboring attachment sites in a protein using molecular biology is likely to increase the use of bifunctional labels.

The molecular biology revolution has had a profound impact on spin label ESR. The limitations of using naturally occurring binding sites are circumvented by the site-directed spin-labeling method pioneered by Hubbell et al.[30] In SDSL, native cysteines are mutated out and new cysteines are introduced at desired vantage points. The power of this method is best illustrated by cysteine scanning where each residue along the polypeptide chain is changed to a cysteine and labeled with nitroxide.

Noncovalent labels are used in the investigation of active sites, e.g. substrate or cofactor analogs, adenosine triphosphate (ATP) or nicotinamide adenine dinucleotide (NAD) nitroxide adducts (Figure 3a–c). The binding and function of these substrates are often not compromised by the presence of the nitroxide. Both approaches

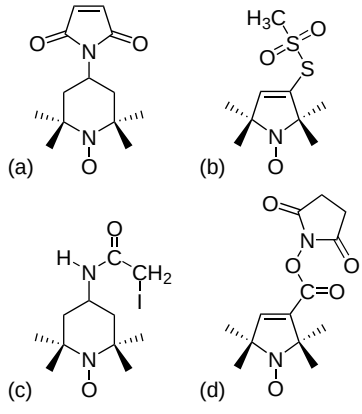

Figure 2 Various spin labels used in covalent modification of proteins: (a) maleimide spin label; (b) methyl thiosulfonate spin label; (c) iodoacetamide spin label; (d) hydroxysuccinamide (lysines).

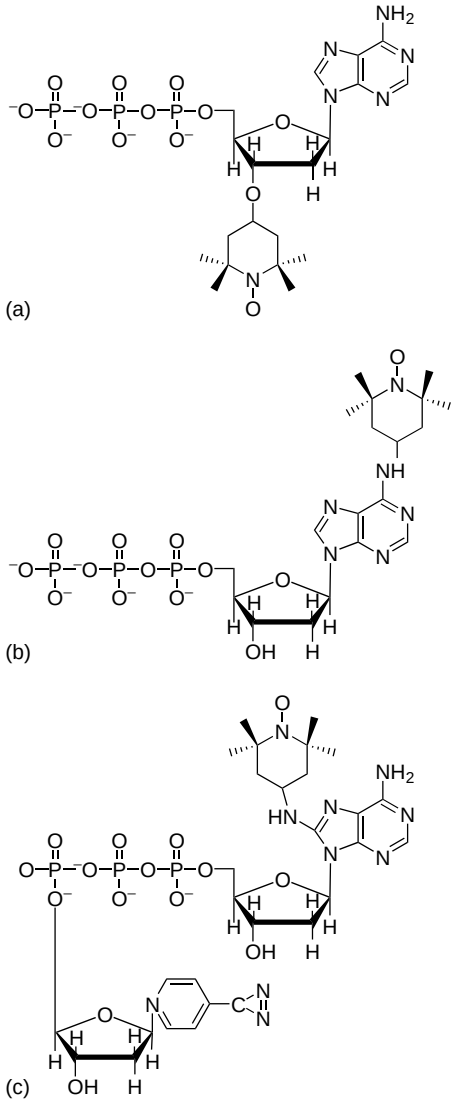

Figure 3 (a, b) ATP spin labels and (c) NAD spin label.

are combined in photoactivated labels with an additional azido or nitrene groups. The label is guided by a substrate analog moiety to an active site and photoactivation attaches the label covalently to the protein.

3.3 Attachment Rigidity

Spin labels are attached to a protein via one or more single bonds about which the nitroxides can rotate on a subnanosecond timescale. In studies of protein orientation and dynamics, such an independent (librational) motion is a major hindrance since it averages the orientational dependence of magnetic tensors – anisotropy. Anisotropy of magnetic tensors is the basis for ESR orientational and motional sensitivity as discussed in sections 5.1 and 5.2. The extent of probe motion is estimated by immobilizing the protein on either glass or ion-exchange beads and comparing spectral parameters such as effective splitting (in conventional ESR) or line-height ratios (in STESR) to their rigid limit values. Alternatively, the protein mobility can be reduced by increasing the medium viscosity, η. The observed spectral parameters can then be plotted against η (Perrin plots) and extrapolated to infinite viscosity. If the extrapolated values are lower than the rigid limit of the nitroxide, or if discontinuities exist in the Perrin plots, then it can be concluded that the probe moves independently of the protein.

3.4 Impairment of Function/Structure of Labeled Proteins

Covalent modification of proteins with extrinsic probes carries the danger of damaging the function of the molecule. Certain labels are innocuous at certain sites while others are not. No generalizations can be made. For example, out of 32 spin-labeled cysteine mutants of T4 lysozyme, 11 displayed intact activity and 11 had activity $\geq$50%. Modification of buried residues and residues in tertiary contacts decreased appreciably the enzymatic activity.[31] In K^+ channels, ion pumping was affected by <50% in 46 out of 66 spin-labeled mutants and very few were completely inactivated.[32] Therefore, a careful functional characterization is necessary if the objective of the study is the correlation of structure and function.

The overall global structure of a protein is generally less affected than its function by labeling. Nitroxide labels are relatively small and a labeled cysteine is not much larger than a tryptophan residue. Surface labeling has been shown to have little effect on protein folding or stability. However, when buried residues are labeled, side chains and backbone can shift to accommodate the labels and restore packing of the protein core.[33]

4 TECHNIQUES AND INSTRUMENTATION

4.1 Continuous Wave Electron Spin Resonance

4.1.1 Theory

The magnetic moment, μ, of an electron interacts with an external magnetic field just like a compass needle interacts with the earth's magnetic field. The interaction of electron spin with the field is often referred to as the *Zeeman interaction*. The energy (E) of a magnetic dipole (μ) in a static magnetic field is given by Equation (1):

$$E = -\mu H \tag{1}$$

where H is the magnetic field strength. The magnetic moment of an electron is generated by its spin (S) (Equation 2):

$$\mu = -g\beta S \tag{2}$$

with β denoting the Bohr magneton (intrinsic unit of electron magnetic moment) and g denoting the spectroscopic splitting factor (relates contribution of spin and orbital motion of the electron to its total angular momentum).

Unlike a compass needle, electron spin is quantized. For a single electron, the projection of S on the magnetic field axis S_z can only take values of $\pm 1/2$. Thus the energy levels from Equations (1) and (2) are $E = \pm(1/2)g\beta H$, resulting in an energy gap which increases linearly with the magnetic field, $\Delta E = g\beta H$ (Figure 4). An oscillating magnetic field can flip the electrons from one energy level to the other if its own energy, defined by the oscillating frequency ν, equals the energy gap. Hence, for resonance between the oscillating field (microwave) and the electron spin, the condition in Equation (3) has to be satisfied:

$$h\nu = g\beta H \tag{3}$$

The resonance condition can also be obtained by considering a spinning electron moving in an orbit around a nucleus placed in a magnetic field. From classical mechanics, the rate of change of the magnetic moment is proportional to the torque produced by the interaction of the moment and the magnetic field and given by their vector product (Equation 4):

$$\frac{\delta\mu}{\delta t} = \mu \otimes \gamma H \tag{4}$$

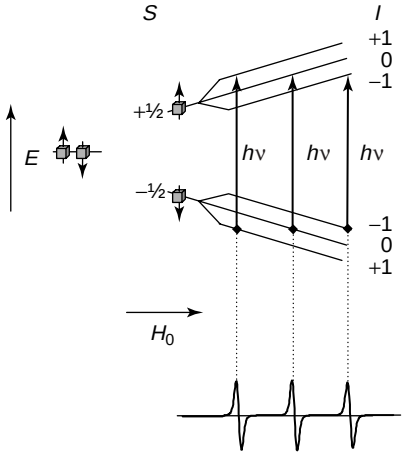

Figure 4 Energy level diagram of Zeeman and hyperfine interactions for a [^{14}N]nitroxide ($S = 1/2$, $I = 1$). The vertical arrows denote ESR transitions with the resulting first-derivative spectrum below.

where γ is a magnetogyric ratio (ratio of magnetic and inertia moments) characteristic of a given electron (Equation 5)

$$\gamma = \frac{g\beta}{\hbar} \tag{5}$$

where $\hbar$ is Planck's constant. The torque will force the magnetic dipole (μ) to precess around the static field at a defined frequency, the *Larmor frequency*, w, given by Equation (6):

$$w = \gamma H \tag{6}$$

Substitution of γ from Equation (5) into Equation (6) yields the resonant condition $h\nu = g\beta H$ stated in Equation (3). Applying an oscillating microwave field of the same frequency as the Larmor frequency cancels the orienting effect of the static magnetic field. The spin then rotates about an axis perpendicular to the static field direction, periodically aligning itself with or against the static field. This is equivalent to dipole (μ) flipping between the two energy levels.

The extent of microwave absorption, which defines the intensity of the ESR signal, is proportional to the difference in spin populations, N, between the upper and lower energy states. The ratio of the two populations is determined by the Boltzmann distribution (Equation 7):

$$\frac{N_{+1/2}}{N_{-1/2}} = \exp\left(-\frac{\Delta E}{kT}\right. \tag{7}$$

The difference in spin populations is increased by either increasing the magnetic field H or reducing the temperature T. For example, at 0.35 T and room temperature the population difference is 0.1% but it can be increased to 13% by reducing the temperature to 3 K. The difference between the levels decreases with absorption and an efficient relaxation pathway has to exist to restore the Boltzmann equilibrium. Relaxation pathways include the *dipolar spin–spin* relaxation – sharing of energy between electrons or nuclei – and *spin–lattice* relaxation – sharing vibrational modes with the lattice. They are characterized by relaxation times T_2 and T_1, respectively. Relaxation times are defined as the time interval between initial perturbation and when the deviation from equilibrium decays to 1/e of its initial value. The relaxation rates are additive and their sum defines the width (at half-height) of the resonance, Γ (Equation 8):

$$\Gamma = \frac{1}{\gamma}\left(\frac{1}{T_2} + \frac{1}{T_1}\right. \tag{8}$$

Faster relaxation (shorter T_1 or T_2) results in broader line widths. For paramagnetic ions, the strong coupling of spins to lattice (short T_1) produces broad lines. Lowering the temperature weakens lattice coupling (increases

T_1) and is commonly used to observe resonance of transition metal ions. The coupling of free radicals (including spin labels) to the lattice is weak, therefore spin–spin relaxation is more efficient and the line width is determined by T_2.

Electrons orbiting around a nucleus experience a small local field produced by the nuclear magnetic moment. This field enhances or counteracts the external field depending on the orientation of the nuclear dipole. This interaction between the nucleus and the electron is known as *hyperfine interaction*. Similar to electron spin, nuclear spin (I) is also quantized to $2I + 1$ levels. Since the selection rule for spin transitions dictates that the total spin quantum number can only change by 1, the hyperfine interactions lead to $2I + 1$ transitions (Figure 4). In the case of nitroxide labels, it is the nitrogen nucleus which interacts with the unpaired electron. For ^{15}N the nuclear spin number m_I is 1/2 so that two electron transitions are observed; for ^{14}N, $I = 1$ and therefore there are three electron transitions.

The resonance condition of Equation (3) is thus modified to include hyperfine interactions, A (Equation 9):

$$h\nu = g\beta H + m_I A \tag{9}$$

The hyperfine interaction between an electron and the nucleus has both an isotropic and dipolar component. The magnitude of the *isotropic splitting*, a_0, is proportional to the electron spin density on the nucleus. Since the unpaired electron is located between the oxygen and nitrogen, increasing the polarity of the medium decreases the oxygen's attraction and increases the electron density on the nitrogen. Thus, a_0 is a sensitive measure of the spin environment.

The pπ orbital of an unpaired electron is asymmetric, making the dipolar interactions of the electron and nucleus orientation dependent. For example, hyperfine interactions are stronger when the z-axis of the orbital is aligned with the magnetic field and weaker when the field is aligned perpendicular. The hyperfine interactions are best described by a second rank tensor, $\boldsymbol{A}$ (Equation 10):

$$\boldsymbol{A} = \begin{matrix} A_{xx} & 0 & 0 \\ 0 & A_{yy} & 0 \\ 0 & 0 & A_{zz} \end{matrix} \tag{10}$$

Typical values for nitroxide spin labels are $A_{xx} \approx A_{yy} \approx 7\,\text{G}$ and $A_{zz} \approx 35\,\text{G}$. The difference between the x- and y-components is small and often the hyperfine tensor is assumed to be axially symmetric.

The Zeeman interaction of the electron spin with the static magnetic field (Equation 1), is also anisotropic. Asymmetry of orbital motion in the pπ orbital results in different contributions of spin and orbital momenta and thus the g-value can also be described by a tensor (Equation 11):

$$\boldsymbol{g} = \begin{matrix} g_{xx} & 0 & 0 \\ 0 & g_{yy} & 0 \\ 0 & 0 & g_{zz} \end{matrix} \tag{11}$$

In contrast to the hyperfine tensor, the g tensor is rhombic with typical values $g_{xx} \approx 2.0085$, $g_{yy} \approx 2.0065$ and $g_{zz} \approx 2.0027$. The asymmetry of the Zeeman and hyperfine interactions defines ESR sensitivity to orientation and to rotational motion.

4.1.2 Electron Spin Resonance Spectrometer

A modern ESR instrument consists of three basic units: (a) a microwave bridge and resonator, (b) a variable field magnet and (c) signal amplification circuitry (Figure 5).

Microwaves of the desired frequency are generated by either a klystron or Gunn diode. Their intensity is adjusted by an attenuator and transmitted via a waveguide to the sample chamber/resonator. During resonance, a small amount of microwaves is reflected from the resonator and detected by a Shottky diode. To separate the reflected and incident microwaves, a circulator is placed between the attenuator and resonator. The circulator channels the microwaves in a forward direction: incident microwaves to the resonator and reflected microwaves to the detector. The bridge often contains an additional pathway – a reference arm which taps off a small fraction of the microwaves from the source – which bypasses the resonator and falls on to the detector to ensure its bias for the optimal detection of small intensity changes during resonance.

A static magnetic field is provided by an electromagnet stabilized by a Hall probe. The field is slowly swept by varying the amount of current passing through the electromagnet. In order to decrease microwave noise,

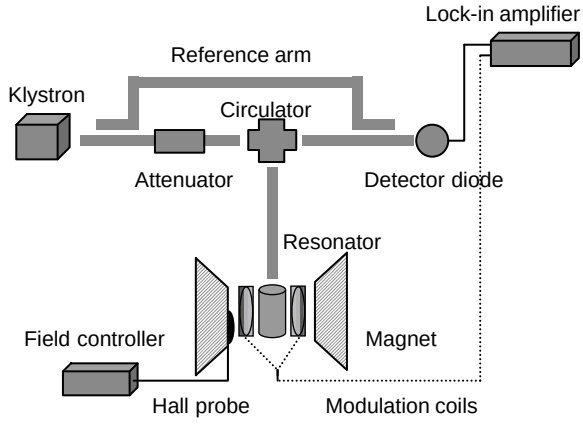

Figure 5 Block diagram of a typical ESR spectrometer.

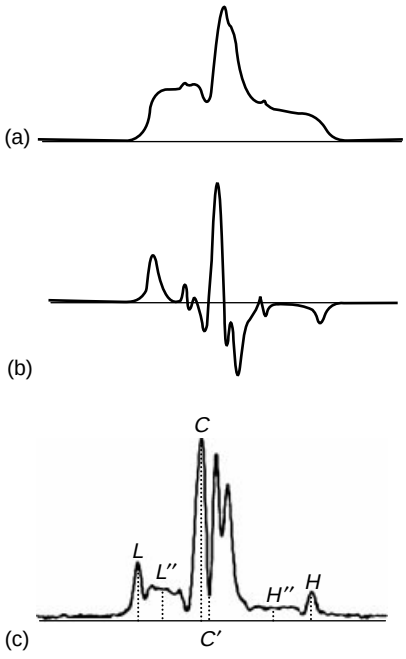

Figure 6 Conventional ESR signals: (a) absorption, V_0; (b) first derivative V_1; (c) STESR spectrum, second derivative, 90° out-of-phase display, V_2'.

the resonance signal is encoded by modulating the static field with a small magnetic field generated by modulation coils. The modulation field sweeps periodically through the nitroxide resonance field. Therefore, the changes in microwave absorption due to resonance occur with the modulation frequency. A lock-in amplifier selects and amplifies only the signal which is in phase and in frequency with the modulation field and rejects all other microwave fluctuations as noise. The signal detected using field modulation is proportional to the changes in the microwave intensity during one cycle of modulation, i.e. the signal is the first derivative of the absorption (Figure 6a–b).

The microwave field produced by the klystron is too weak to induce any detectable absorption by the sample. Resonant cavities, loop gap resonators (LGRs) or, more recently, dielectric resonators (DRs) are used to increase the microwave magnetic field at the sample. The cavities rely on the generation of a standing wave pattern of microwaves whose intensity builds up during cavity resonance. The main drawback of cavities is the presence of an electric component of the microwave. The electric component is absorbed by ''lossy'', aqueous samples (common in biology) causing sample heating and loss of cavity resonance. To avoid this problem, sample volume is restricted to the nodal planes of the electric field, limiting the usable volume of the cavity and thus resulting in a low filling factor, η. Cavities are high-Q structures ($Q = E_{\text{stored}}/E_{\text{dissipated}}$), storing thousands of times more energy than is dissipated on the walls. High Q can only be achieved within a narrow frequency bandwidth of stored microwaves as $Q = \nu/\Delta\nu$. Small changes in the sample, cavity geometry or temperature can all cause frequency shifts and mismatching of the incident microwave with the cavity. Automatic frequency control (AFC) circuitry is employed to track the frequency of the klystron to that of the cavity. However, the AFC feedback response time limits the deadtime of signal changes in transient experiments such as stop-flow. In pulse experiments it is necessary to wait until the energy of the perturbing pulse is fully dissipated. This ring-downtime is proportional to Q; thus in the high-Q structures a longer time has to elapse before a relatively weak spin echo or free induction decay (FID) signal can be collected.

Most of these problems with cavities have been overcome by low-Q resonators such as an LGR or DR. These resonators condense the magnetic component of the microwave, separating it from the electric component. Lossy samples are no longer heated by the electric component. Small sample volumes and large filling factors offer an additional advantage especially when dealing with genetically engineered proteins which are often purified in picomolar quantities. Furthermore, fast dissipation of energy and the large bandwidths of LGRs and DRs make them suitable for pulsed and transient experiments.

4.1.3 Instrumental Variables Affecting the Electron Spin Resonance Spectrum

Two instrumental parameters influence the line shape of experimental spectra: *modulation amplitude* and *microwave power*. The amplitude of the ESR signal initially increases with the modulation amplitude (H_m) as it approaches the intrinsic line width (ΔH_{pp}). Maximum amplitude is attained at $H_m = 3.5\Delta H_{\text{pp}}$ for Lorentzian and at $H_m = 1.8\Delta H_{\text{pp}}$ for Gaussian line shapes. Any further increases in H_m result in a decrease of the signal. The broadening of ΔH_{pp} is observed well before the maximum amplitude of the signal. When $H_m = \Delta H_{\text{pp}}$, the observed Lorentzian and Gaussian widths are 25% and 15% larger, respectively. As a rule of thumb, modulation should be kept at one-fifth of the intrinsic line width when resolution or line width is of importance. The errors in line width are then <1%.

The ESR signal is proportional to the static microwave field, H_1, and hence to the square root of the microwave power ($P^{1/2}$), in the absence of saturation. When the rate of relaxation lags behind the rate of excitation and when the spin populations in the ground and excited states are equalized, the signal saturates and decreases to nil. The amplitude of the ESR signal (y) for a Lorentzian line shape as a function of the microwave field is given by

Equation (12):

$$y = \frac{y_0 H_1}{\sqrt{(1 + H_1^2 \gamma^2 T_1 T_2)^3}} \tag{12}$$

where y_0 is a field-independent parameter. Hence it is important to keep power levels well below the maximum amplitude whenever spectra are used to quantify the number of spins. Line-width distortions are less pronounced than those due to the modulation field, but at powers giving maximum amplitude, the observed line width increases 1.2 times over the intrinsic line width.

4.2 Saturation Transfer Electron Spin Resonance

4.2.1 Qualitative Theory

STESR was developed to study slower molecular dynamics with rotational correlation times (τ_r) of >200 ns.[34]

The timescale of conventional ESR is determined by the spin–spin relaxation time T_2 (nanoseconds). The ESR timescale can be extended to a longer spin–lattice relaxation time T_1 (microseconds) if a signal sensitive to spin saturation is observed. This can be done by saturating the signal with intense microwaves, creating a ''hole'' in the absorption spectrum, and subsequently observing signal recovery. When the saturating microwave is switched off (or decreased to nonsaturating levels) the signal recovers with the rate determined by the spin–lattice relaxation time, T_1. The onset of motion provides an additional relaxation mechanism: *spectral diffusion*. The saturation is relieved as the resonating-saturated spins rotate away from the resonance field and the unsaturated spins come into resonance. The ''hole'' broadens out across the spectrum and the intensity of the signal increases. The second harmonic, out-of-phase ESR signal (V_2'), collected at moderate saturation, is particularly sensitive to spectral diffusion. The line shape of V_2', in the presence of saturation, bears a strong resemblance to the absorption spectrum with the intensity lowered in the spectral regions most sensitive to the spectral diffusion (see Figure 6c).

4.2.2 Instrumental Parameters

The STESR signal is influenced by nitroxide relaxation times, spectral diffusion, spin saturation level and the modulation frequency with which the ''hole'' is observed. Hence the instrumental parameters which affect any of these must be precisely controlled. The saturating microwave field averaged over the sample volume is set to 0.25 G. The microwave power is adjusted to this level using the microwave field conversion factor (c), corrected for a filling factor (η) and dielectric losses, which lower the Q factor (Equation 13):

$$\langle H_1^2 \rangle_s = c\eta PQ \tag{13}$$

where P is incident microwave power. The power-to-field conversion factor is determined experimentally by the saturation of Fremy salt [peroxylamine disulfonate (PADS)] for which the half-saturation field is 0.1067 G.

The modulation frequency and amplitude, which determine the frequency of the stepping on- and off-resonance, i.e. the interval between burning and observing the ''hole'', must also be calibrated. Modulation broadening of a narrow line-width sample, e.g. Fremy salt, is used for this purpose. The observed line width (ΔH_{pp}) is dominated by modulation broadening when the modulation amplitude is ~10 times the intrinsic line width [$\Delta H_{pp}(0)$], i.e. $\Delta H_{pp} = H_m - \Delta H_{pp}(0)$. Commonly used values for the modulation field are 5 G and 50 kHz.

Finally, since the V_2' STESR signal is 90° out of phase (phase quadrature) with modulation, the precise phase nil must be found. An error of 1° in setting the phase quadrature can result in significant line-shape changes due to leakage of a more intense in-phase signal. The most popular phase nulling method is by interpolation of the unsaturated in-phase signal: two or three readings are taken within 15° on each side of the putative nil and the phase at which the signal is zero is found by linear interpolation. A general description of the experimental procedures and calibration can be found in Fajer and Marsh[35] and Squier and Thomas.[36] Digital post-acquisition methods have also been proposed but are not widely used.

As a footnote, protein mobilities measured by conventional ESR and STESR were independently verified by optical methods – fluorescence and phosphorescence anisotropy. Bovine serum albumin labeled with a dual probe bearing a spin label moiety and the optical probe eosin was measured using optical methods and ESR/STESR.[37] The agreement between the fluorescence/phosphorescence and ESR was excellent.

4.3 Time Domain Methods

Time domain ESR relies on the perturbation of the equilibrium magnetization by an intense microwave pulse which is then followed by one of the following: (a) conventional ESR to observe the return of magnetization to equilibrium – saturation recovery ESR; (b) refocusing of the magnetization in the xy-plane – spin-echo ESR; or (c) free induction decay (FID) of the magnetization in the xy-plane which is then Fourier transformed (FTESR).

The development of time domain ESR posed a formidable technical challenge. The microwave pulse must be short (a few nanoseconds) and strong enough to

cover a 70-G wide spectrum of nitroxides. The resonators must dissipate the pulse energy within tens of nanoseconds before the loss of magnetization coherence and the signal must be digitized with a subnanosecond dwell time owing to the short nitroxide relaxation times. Fortunately, technological advances in microwave sources, resonator design and data acquisition electronics in the last decade have facilitated development of commercial Fourier transform spectrometers and the time domain method has become increasingly popular. The various time domain ESR techniques are illustrated in Figure 7(a–c).

Saturation recovery electrospin resonance (SRESR) is a hybrid of continuous wave and pulse methods in which the pulse saturates a spin population at a desired field thereby creating a ''hole'' in the absorption spectrum (Figure 7a). The kinetics of recovery are determined by various relaxation pathways: spin–lattice relaxation, nuclear relaxation, Heisenberg spin exchange (HSE) or spectral diffusion. These competing pathways can be resolved by varying the pulse duration. SRESR has been used successfully in the determination of spin–lattice correlation times and spin exchange.

Spin-echo electron spin resonance (SEESR) uses a sequence of pulses; in Hahn echo a 90° pulse is followed by a 180° pulse t_1 seconds later (Figure 7b). The first pulse tips the magnetization into the *xy*-plane where individual spins rotate with their respective Larmor frequency, w. The difference in Larmor frequencies, which arises from different resonant fields, leads to dephasing of the magnetization in the *xy*-plane which is then refocused by the 180° pulse. Spins lagging by $\Delta w t_1$ before the refocusing pulse are now $\Delta w t_1$ ahead. At time $2t_1$, spins are brought into coherence and an echo is formed. In this way static differences in Larmor frequency due to different resonant fields or different local fields (inhomogeneous broadening) are annihilated. The dependence of the echo amplitude on time t_1 reveals Larmor frequency fluctuations that cannot be refocused by the 180° pulse. These fluctuations contain information about molecular dynamics, spin exchange and dipolar interactions. The decrease of the spin echo as a function of t_1 is a measure of the T_2 relaxation time. The decay of the echo amplitude is often recorded as a function of spectral position by stepping the magnetic field, resulting in a 2-D spectrum: an inhomogeneously broadened spectrum along the field axis and a homogeneous line shape on the t_1 axis. Since the inhomogeneous broadening often obscures a multitude of phenomena affecting ESR line width, then the ability to obtain a pure, homogeneously broadened spectrum is of considerable value.

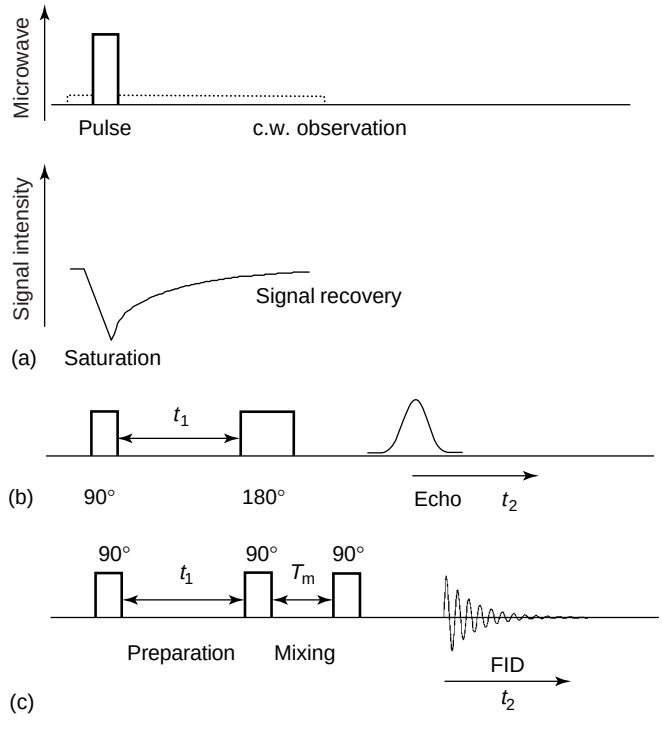

Figure 7 Time domain ESR methods: (a) saturation recovery; (b) spin echo; (c) 2-D FTESR (ELDOR).

2-D FTESR is the most versatile technique of time domain ESR (Figure 7c). All the spins are excited simultaneously with a strong, short microwave pulse which tips the magnetization into the *xy*-plane. The length and strength of the signal determine the spectral range covered, e.g. for nitroxides with a 200-MHz spectral range, 2-kW pulses 5 ns in duration are needed.

Coherently excited spins precess about a magnetic field at their Larmor frequency. This precession can be detected as an oscillating signal in the *xy*-plane which decays in time as the spins lose coherence. This FID signal is Fourier transformed into the frequency domain to yield an absorption spectrum. Application of two or more pulses spaced by varying intervals allows sampling of spin coherences in multiple dimensions. In the simplest of these experiments, two-pulse spin-echo correlation spectroscopy (SECSY), the first pulse tips the magnetization into the *xy*-plane where the spins become frequency labeled during the evolution time t_1. A second pulse reverses the magnetization during the collection time t_2, canceling inhomogeneous broadening. Fourier transformation with respect to t_1 and t_2 yields an absorption spectrum along the f_2 axis and a homogeneous line width along the f_1 axis. Thus SECSY is an FTESR equivalent of field-stepped SEESR. The isolation of the homogeneous line shapes out of an inhomogeneously broadened spectrum is used to study molecular dynamics.

Three pulse sequences are used in 2-D ELDOR experiments (e.g. Figure 7c). The first pulse creates a transverse magnetization in the *xy*-plane which evolves for time t_1. The second pulse stores this frequency encoded magnetization along the *z*-axis allowing for

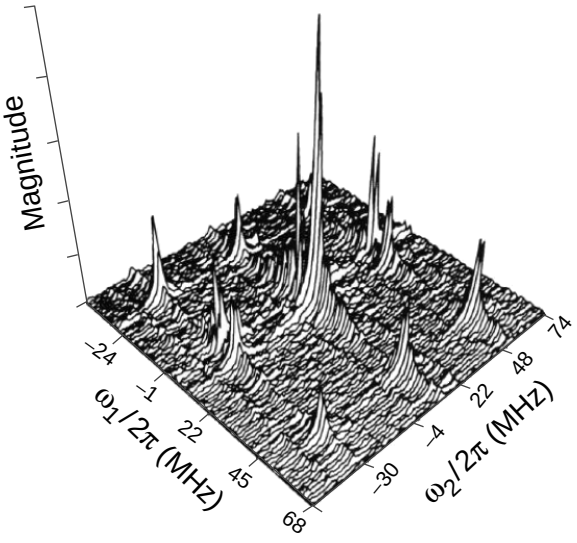

Figure 8 2-D ELDOR of PD-TEMPONE. In addition to autopeaks occurring at $f_1 = f_2$, there are off-diagonal cross-peaks due to HSE.[25] (Reproduced with permission from J. Gorcester, J.H. Freed, *J. Phys. Chem.*, **88**, 4678–4693 (1986).)

magnetization transfer to take place during mixing time T_M. A third pulse transforms the magnetization back into the xy-plane where it is observed during time t_2. Magnetization transfer changes the resonant frequency of a spin from f_a to f_b, creating an off-diagonal cross peak at (f_a, f_b) (Figure 8). The time evolution of the amplitude of the cross peak, measured by varying T_M, is used to determine the time course of magnetization transfer. The biophysically relevant phenomena causing magnetization transfer include HSE, modulation of dipolar interactions, nuclear flips and, most importantly, spectral diffusion due to rotational motion.

5 APPLICATIONS OF SPIN LABELING

The orientational difference of magnetic interactions (referred to anisotropy) forms the basis of spin labeling techniques in biological research. In the absence of motion, each field position corresponds to a defined orientation of the label with respect to the field. The intensity of the signal at a particular field position is directly proportional to the population of molecules with that given orientation. Hence an ESR spectrum can be used to determine the range of orientations present in a sample. Partially or fully averaging the g- and hyperfine tensor anisotropy results in spectral line shapes determined by the frequency and amplitude of molecular motion. ESR can also be used to measure intra- and intermolecular distances. The presence of paramagnetic centers in the vicinity of spin labels modulates spin relaxation pathways in a distance-dependent manner. In this section we shall discuss how ESR is used in the investigation of molecular orientation, molecular dynamics, ligand binding, intra- and intermolecular distance measurements and the determination of various levels of proteins structure.

For each of these applications a qualitative description of the physical principles allowing for these measurements will be given, followed by examples. More extensive reviews of these topics can be found in a monograph by Lichtenstein,[38] a series edited by Berliner and Reuben[39,40] and separate reviews by Hubbell et al.,[41–43] Marsh and Horvath,[44] Millhauser et al.[45,46] and, most recently, Hustedt and Beth.[47]

5.1 Protein Orientation

5.1.1 Orientation of a Single Molecule

The anisotropy of the Zeeman and hyperfine interactions confers orientational sensitivity to ESR spectra. Nitroxide spin labels with a z-axis parallel to the magnetic field generate a spectrum with a splitting of 70 G. Spins oriented perpendicular to the field display a splitting of 14 G (Figure 9).

The center position of the spectrum, determined by the g-tensor, is sensitive not only to the position of the z-axis but also to the orientation of the x- and y-axes. At 9 GHz the center is shifted 5 G downfield (left) for a spin with its y-axis aligned with the magnetic field and another 4 G for spins with x-axis parallel to H_0 (Figure 9).

The effective g- and hyperfine splitting tensors for a spin placed at an arbitrary polar angle (θ, ϕ) with respect to the field are given by Equations (14) and (15), respectively:

$$g(\theta, \phi) = g_{xx} \sin^2\theta \cos^2\phi + g_{yy} \sin^2\theta \sin^2\phi + g_{zz} \cos^2\theta \tag{14}$$

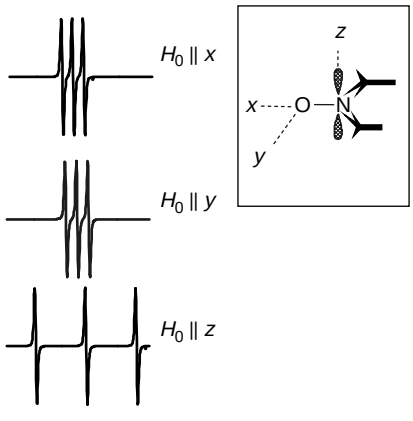

Figure 9 Orientational sensitivity of ESR spectra. Splitting of the spectrum changes when the z-axis of nitroxide rotates with respect to the magnetic field. The center of the spectrum changes when nitroxide rotates about any axis.

$$A^2(\theta,\phi) = A_{xx}^2 \sin^2\theta\cos^2\phi + A_{yy}^2 \sin^2\theta\sin^2\phi + A_{zz}^2\cos^2\theta \quad (15)$$

It follows, then, that the resonance field for a given orientation (θ, ϕ) is given by Equation (16):

$$H_{\text{res}}(\theta,\phi,m_{\text{I}}) = \frac{h\nu}{\beta g(\theta,\phi)} + m_{\text{I}}A(\theta,\phi) \quad (16)$$

Equation (16) describes the orientational resolution of ESR line shapes: a spin with a specific orientation can be found at a defined position along the field axis. The ESR intensity at any field position is directly proportional to the number of spins at the orientation defining H_{res}. An ESR spectrum can also be considered as an orientational distribution function, $N(\theta)$. $N(\theta)$ is approximated by an orthonormal set of spherical harmonics and has been developed and applied to samples with cylindrical and planar symmetry.[48,49] Alternatively, the orientation can be modeled in terms of a Gaussian distribution with a width $\Delta\theta$ and center θ_0 (Equation 17):

$$\rho(\theta) = \exp\ -\ln 2\frac{(\theta-\theta_0)^2}{\Delta\theta^2} \quad (17)$$

The ESR spectrum, $Y(H)$, is created by calculating a resonance field H_{res} for every θ within the Gaussian distribution of orientations and placing a Lorentzian first derivative line width at H_{res} with the intensity weighted by $\rho(\theta)$ (Equation 18):

$$Y(H) = \rho(\theta)\frac{(H-H_{\text{res}})\Delta H_{\text{pp}}}{[(H-H_{\text{res}})^2 + \Delta H_{\text{pp}}^2]^2} \quad (18)$$

where the peak-to-peak width of the Lorentzian (ΔH_{pp}) is defined by the spin–spin relaxation time, T_2 (Equation 19):

$$\Delta H_{\text{pp}} = \frac{2}{\sqrt{3}\gamma T_2} \quad (19)$$

ESR is one of the very few biophysical techniques directly sensitive to orientational disorder. The spectra in Figure 10(a–c) illustrate this sensitivity. As the width of the Gaussian distribution increases, the ESR resonances broaden to a powder pattern limit which is characteristic of isotropically disordered spins.

In summary, the parameters describing orientational distribution (axial, azimuthal angles and their disorder) can be obtained from spectral parameters: spectral splitting, center of the spectrum and the line width respectively. This can be achieved either by graphical methods or from the automated fitting of full spectral line-shape parameters. Graphical methods compare the effective splitting and width of the resonance to graphs obtained from computer simulations. The automated method allows modeling of more complex bimodal distributions, in both θ and ϕ, and can be linked to standard fitting routines such as Levenberg–Marquardt[50] or Simplex.[51]

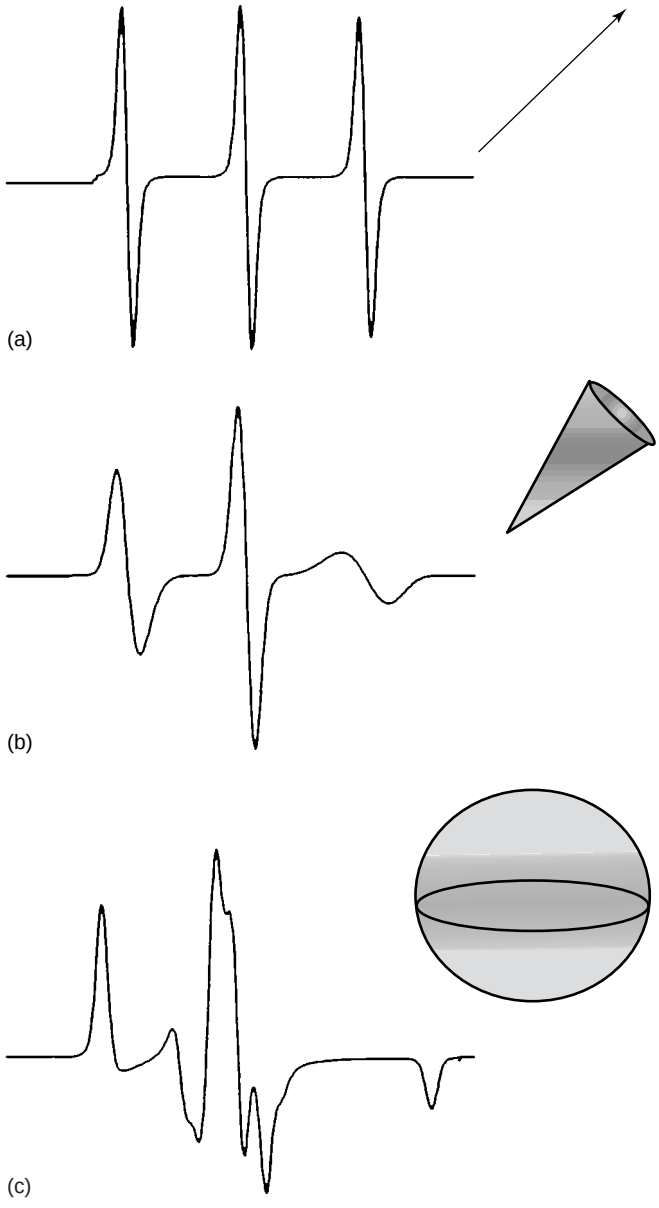

Figure 10 Broadening of the spectra with increasing disorder: Gaussian disorder of (a) $\pm1^\circ$ and (b) $\pm5^\circ$ and (c) completely disordered (isotropic) distribution.

5.1.2 Macromolecular Assemblies

Of considerable interest is the orientation of molecules within macromolecular assemblies, e.g. proteins within lipid membranes or contractile proteins in the muscle

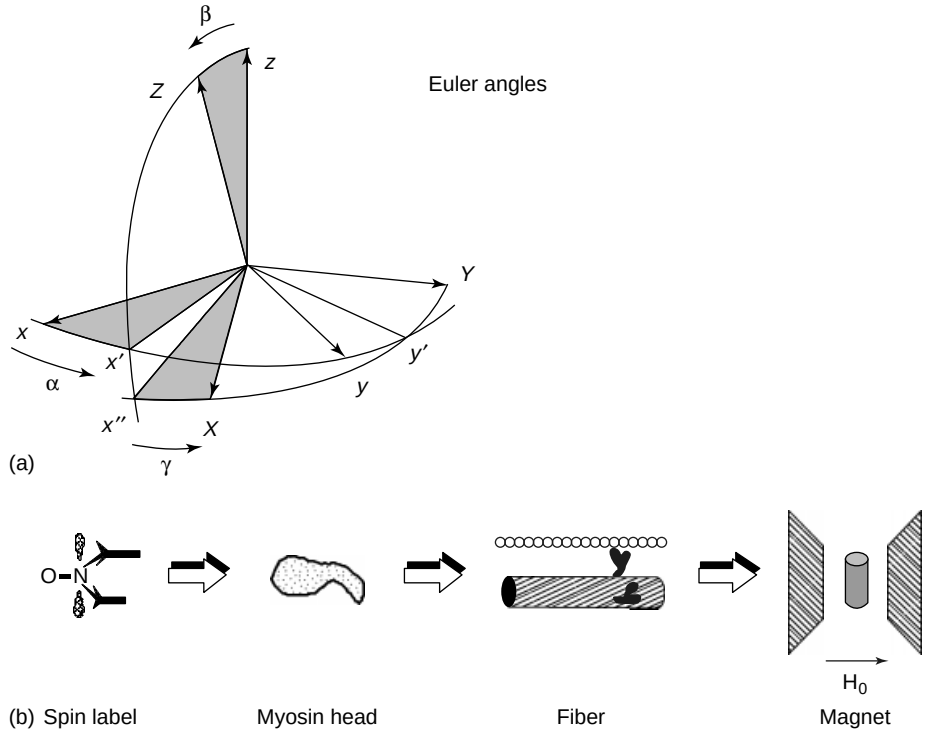

Figure 11 Definition of (a) Eulerian angles α, β and γ; (b) Eulerian transformations of the spin label orientation through the protein (myosin head); sample (fiber) frame of reference into the laboratory frame (magnet).

fibers. If the assembly of proteins is ordered and oriented at a specific angle with respect to the magnetic field, the orientational distribution of the labeled components of the assembly can be easily determined. The spectra of such samples oriented with the symmetry axis parallel to the field are the same as for a single spin (Equation 18). When the orientation of the spin label with respect to the protein is known, the ESR spectra are interpreted in terms of the orientation of the labeled domain with respect to the assembly which is of biological interest. This is achieved using *Eulerian transformations* between three frames:[52,53] *molecular* frame (defining orientation of the label within the protein), *sample* frame (orientation of proteins within the assembly) and *laboratory* frame (orientation of the sample in the magnetic field) (Figure 11a and b).

The ESR spectra are simulated taking into account the orientational distribution in each of the frames. The magnetic tensors are rotated from molecular to laboratory axes using directional cosine matrices (L) according to Equation (20):

$$A_{\text{lab}} = L^{\tau}_{\text{mol}} L^{\tau}_{\text{sam}} L^{\tau}_{\text{lab}} A_{\text{NO}} L_{\text{lab}} L_{\text{sam}} L_{\text{mol}} \qquad (20)$$

where L_{mol}, L_{sam} and L_{lab} are cosine matrices defined in Equation (21) for each of the Eulerian transformations and L^{τ} is their transpose:

$$L = \begin{matrix} \cos\beta\cos\alpha\cos\gamma - \sin\alpha\sin\gamma & \sin\beta\sin\alpha\cos\gamma + \cos\alpha\sin\gamma & -\sin\beta\cos\gamma \\ -\cos\beta\cos\alpha\sin\gamma - \sin\alpha\cos\gamma & -\cos\beta\sin\alpha\sin\gamma + \cos\alpha\cos\gamma & \sin\beta\sin\gamma \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{matrix} \qquad (21)$$

The resonant field for each spin packet is calculated as shown in Equation (22):

$$H_{\text{res}} = \frac{h\nu}{\beta g_{zz}} + m_{\text{I}}\sqrt{A^2_{xz} + A^2_{yz} + A^2_{zz}} \qquad (22)$$

Note that the subscripts of the g- and hyperfine tensors in Equation (22) denote elements in the laboratory frame.

5.1.2.1 Protein Orientation in Membranes The Eulerian transformation approach was introduced by Griffith et al. to determine lipid orientation in membrane bilayers. The orientation of the spin label nitroxide with respect to the lipid molecule is well defined. A stack of lipid membranes is tilted with respect to the magnetic field at a known angle, and the spectra can be defined solely by the orientational distribution in the sample frame of reference (Ω_{sample}).[52,54]

Membrane-bound proteins are investigated in a similar manner. The orientation of a spin-labeled ligand of the

erythrocyte anion transporter, Band 3, was determined by flowing red blood cells into thin, flat, sample cells. The flow shear oriented the red blood cells parallel to the flow and the sample cell was rotated both parallel and perpendicular to the field in order to vary Ω_{lab}.[55] A global analysis of the tilt series resulted in a full description of the label orientation with respect to the normal membrane axis. The derived spin label orientation was found to be consistent with orientation results determined independently by analyzing the anisotropy of motion.

5.1.2.2 Muscle Proteins Similar approaches have been used to describe the orientation of various muscle proteins. Force generation in muscle is believed to be brought about by the reorientation of the myosin cross bridges. Since ESR is sensitive to the orientation of the proteins, muscle field proved to be a fertile ground for ESR applications. Muscle fibers form a naturally ordered assembly with the fiber axis defining the cylindrical symmetry required for ESR. Proteins can be labeled directly in muscle cells or labeled as isolated components and exchanged for corresponding unlabeled proteins in the muscle sample. These include most of the thin filament proteins – actin, troponin C (TnC), troponin I (TnI) and tropomyosin – and also the thick filament components – myosin heavy chain, regulatory light chain and essential light chain. The orientation of many of these components has been extensively studied as a function of the intermediate states of the acto-myosin cycle and also during muscle activation. Thomas and Cooke have established that in the absence of ATP, myosin heads attach themselves strongly and stereospecifically to actin. Muscle relaxation, in the presence of ATP, produced disorder consistent with head detachment and Brownian motion.[56] Subsequent studies using various nucleotide analogs to trap the intermediate adenosine triphosphatase (ATPase) states have revealed a sequence of orientational changes of the catalytic domain: a nonstereospecific attachment of transient, weakly bound heads followed by an equally large orientational disorder of the strongly attached heads in the prepower stroke state.[75,76] Force generation was associated with the disorder-to-order transition. The postpower stroke state, with the hydrolysis product adenosine diphosphate (ADP) in the active site, showed a local domain heterogeneity, but overall the catalytic domain was well oriented, $\Delta\theta \pm 8^\circ$ (Figure 12). Release of ADP (rigor state) resulted in a slight change in the twist and the tilt angle of the heads.[77] During isometric contraction, when most of the myosin heads should be in the prepower stroke state (immediately prior to the rate-limiting step of the cycle), no species were observed at a different angle to that of the postpower stroke heads.[102] These findings excluded a simple model in which the catalytic domain (accounting for most of the myosin head mass) generates a force while rotating by 45° from one well-defined angle to another.

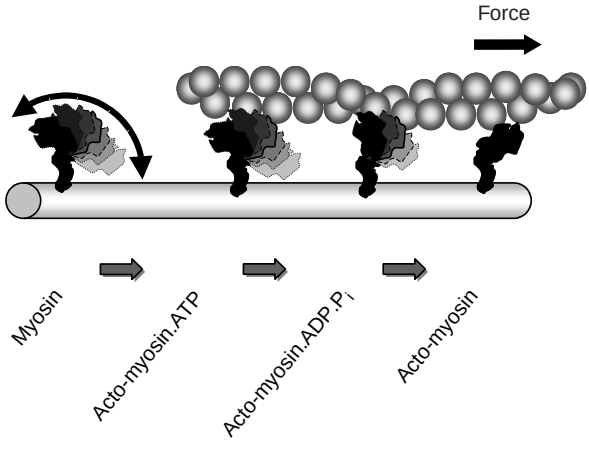

Figure 12 Orientation of catalytic domain of myosin head in the intermediate stages of the contractile cycle. As the head progresses through the cycle, the dynamic disorder is continually diminished, culminating in the disorder-to-order transition associated with force generation.

A different story emerged when the labels were placed in the regulatory domain of the myosin head. Two distinct populations, centered 36° apart, were observed in contraction (predominantly prepower stroke heads), whereas in rigor, only one population was observed.[57] Clearly, the rotation of the head is limited to the regulatory domain with the catalytic domain shifting from a disordered to ordered structure.

The disorder-to-order transition was also found for other muscle proteins. For example, spin-labeled TnC is well ordered prior to Ca^{2+} activation but becomes disordered in the presence of Ca^{2+} or activating myosin heads.[58] The loss of stereospecific, protein–protein interactions is reflected by changes to the conformational homogeneity and is the basis of many molecular mechanisms. ESR, with its capacity to see directly both the ordered and the disordered populations, complements more popular methods such as X-ray crystallography or electron microscopy image reconstructions which ignore disorder.

5.2 Protein Dynamics

5.2.1 Sensitivity of Conventional Electron Spin Resonance

Conventional ESR is used to study molecular dynamics on the nanosecond timescale. This timescale corresponds to motions of peptides and small proteins, or the mobility of labels on the surface of large proteins. Sensitivity of the ESR signal to motion arises from rotational modulation of the magnetic tensor anisotropy. The anisotropy

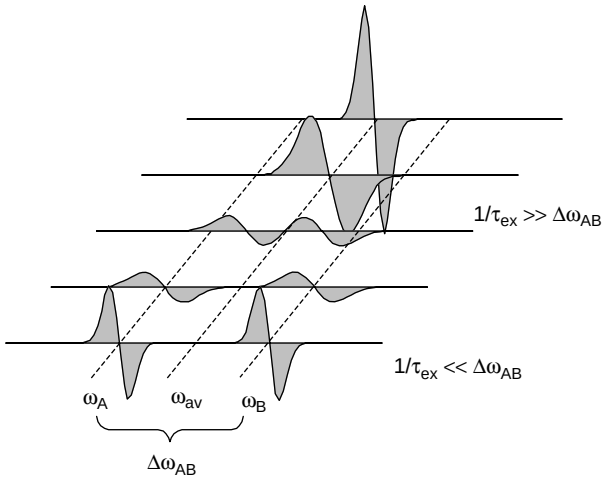

Figure 13 Two-site exchange between environments giving rise to two resonances separated by Δw. Increased rate of exchange initially broadens the resonances then averages their resonant frequencies.

of Zeeman or hyperfine interactions results in different resonant fields for different spin orientations. Therefore, rotational motions which change the spin label orientation will modulate the ESR line shape. This can be explained by using a two-site exchange example: two spins, A and B, resonate with frequencies w_A and w_B and exchange their positions at a rate $1/\tau_{ex}$. When the exchange rate is significantly slower than the difference in resonant frequencies ($1/\tau_{ex} \ll \Delta w^0_{AB} = w_A - w_B$, slow exchange), the spectrum consists of resonances A and B centered at w_A and w_B, respectively, and the line widths are determined by T_2 (Figure 13). The effective relaxation rate $1/T_2^{eff}$ is the sum of the intrinsic $1/T_2$ rate and the exchange rate $1/\tau_{ex}$. When the exchange frequency increases, the line widths start to broaden. Further increase of the exchange rate causes partial averaging of the resonance positions. The observed difference in resonant frequencies is reduced according to Equation (23):

$$\Delta w_{AB} = \Delta w^0_{AB} \sqrt{1 - \frac{8}{\tau_{ex}^2 \Delta w^0_{AB}}} \tag{23}$$

For exchange rates faster than the difference in resonant frequencies ($1/\tau_{ex} \gg \Delta w_{AB}$, fast exchange), the two resonant peaks coalesce into one peak at the average frequency (assuming equal populations of A and B) (Figure 13). The line width is determined by the effective relaxation time, T_2^{eff}, with contributions from T_2 of species A and B and the exchange broadening (Equation 24):

$$\frac{1}{T_2^{eff}} = \frac{1}{T_{2A}} + \frac{1}{T_{2B}} + \Delta w^2_{AB} \frac{\tau_{ex}}{8} \tag{24}$$

In other words, at the fast exchange limit ($1/\tau_{ex} \gg \Delta w_{AB}$), the contribution of the exchange rate to line width disappears and the spectrum consists of a single peak at $1/2(w_A + w_B)$ with an average line width of $1/T_{2A} + 1/T_{2B}$ (Figure 13).

The above considerations generally hold true for any spins exchanging between different environments such as different local magnetic fields, dipolar interactions or association with different macromolecular assemblies. Spin label reorientation with respect to the magnetic field is also a form of exchange where the rotational correlation time τ_r is the exchange rate and the anisotropy of the g- and hyperfine interactions defines the frequency difference. At X-band, $\Delta w = (A_{zz} - A_{xx})/\hbar = 500\,\text{MHz}$ or $(g_{xx} - g_{zz})\beta H/\hbar = 185\,\text{MHz}$.

As for the two-site exchange discussed above, various motional regimes of ESR can be defined: *fast* ($\tau_r \approx 10^{-11}$–10^{-9} s), *slow* ($\tau_r \approx 10^{-9}$–2×10^{-7} s) and *very slow* ($\tau_r > 2 \times 10^{-7}$ s). Fast and slow motion are of the order of T_2 (15–30 ns) for electron spin and have visible effects on conventional ESR spectra which measure the transverse component of magnetization. The very slow motions do not affect transverse magnetization, but they do affect longitudinal magnetization which decays with T_1 (1–15 μs). These slow motions can be detected (indirectly) using saturation transfer, pulsed ELDOR or saturation recovery ESR.

5.2.1.1 Fast Motion ($\tau_r \approx 10^{-11}$–10^{-9} s) In the fast motional regime, the motion completely averages the anisotropy of the g- and hyperfine tensors. The rotational rate is obtained from the line width broadening using Redfield's perturbation theory.[59] The broadening itself is a function of the nuclear quantum spin number as different nuclear manifolds have varying anisotropy values (Equation 25):

$$\Delta H(m_I) = A + Bm_I + Cm_I^2 \tag{25}$$

The coefficient A is equal to homogeneous broadening and coefficients B and C assure differential broadening of lines belonging to different nuclear manifolds. These coefficients are obtained from the line widths of the Lorentzian lines according to Equations (26) and (27):

$$B = \frac{\sqrt{3}}{4}\Delta H(0) \left\{ \overline{\frac{V(0)}{V(+1)}} - \overline{\frac{V(0)}{V(-1)}} \right\} \tag{26}$$

$$C = \frac{\sqrt{3}}{4}\Delta H(0) \left\{ \overline{\frac{V(0)}{V(+1)}} + \overline{\frac{V(0)}{V(-1)}} - 2 \right\} \tag{27}$$

where $V(m_I)$ is the peak-to-peak height of a given nuclear manifold resonance and $\Delta H(0)$ is the peak-to-peak line width of the central line.

Factors B and C are equal for isotropic motion and τ_r^{iso} is calculated directly from Equations (28) and (29):

$$\tau_B^{iso} = -1.22 \times 10^{-9} B \tag{28}$$

$$\tau_C^{iso} = 1.19 \times 10^{-9} C \tag{29}$$

where B and C are expressed in gauss and τ in seconds.[61]

In the case of anisotropic motion, $B \neq C$, and the rates of rotation about the nitroxide z-, x- and y-axes are different. The ratio of C and B can be used to define the anisotropy as the coefficients are independent of the rate of motion. Various models of anisotropic motion are considered in excellent reviews by Marsh[61] and Beth and Robinson.[67] If the molecule is diffusing in an isotropic medium, then the rotational correlation times about the nitroxide z-axis ($\tau_\parallel$) and about an axis perpendicular to z($\tau_\perp$) are given by Equations (30) and (31):

$$\tau_\parallel = \frac{2\tau_{20}\tau_{22}}{3\tau_{20} - \tau_{22}} \tag{30}$$

$$\tau_\perp = \tau_{20} \tag{31}$$

where τ_{20} and τ_{22} describe spin relaxation and are related to the anisotropy of the magnetic interactions (Equations 32 and 33):

$$\tau_{20} = \frac{1.11 \times 10^{-7}}{H\Delta A} \frac{5(\delta A)B - 8(\delta g)HC}{\Delta g \delta A - \delta g \Delta A} \tag{32}$$

$$\tau_{22} = \frac{3.69 \times 10^{-8}}{H\delta A} \frac{8\Delta g HC - 5\Delta A B}{\Delta g \delta A - \delta g \Delta A} \tag{33}$$

where ΔA and δA are given by hyperfine anisotropy (Equations 34 and 35):

$$\Delta A = A_{zz} - \tfrac{1}{2}(A_{xx} + A_{yy}) \tag{34}$$

$$\delta A = \tfrac{1}{2}(A_{xx} - A_{yy}) \tag{35}$$

with equivalent equations for g-anisotropy. The indices in Equations (34) and (35) are permutated to calculate the values of $\tau_\parallel$ and $\tau_\perp$ for the rotation about the x- and y-axes of the nitroxide.

The above equations hold for anisotropic motion about a specific nitroxide axis in an isotropic medium. An additional complication arises when diffusion takes place within a strongly orientating potential such as in a lipid membrane, or within the steric confines of a protein. The field position of resonances now depends on the amplitude of motion, which defines the time average of available angular space; e.g. if the nitroxide can move only within an angular cone, then only the resonances corresponding to the orientations within the cone are averaged. Motionally averaged spectra are described in terms of order parameters (S) – time averages of the direction cosines of the diffusion axis with respect to the local director axis. For an isotropic diffusion within the cone angle θ_c, components of an ordering tensor are given by Equations (36) and (37):

$$S_{zz} = \tfrac{1}{2}(\cos^2\theta_c + \cos\theta_c) \tag{36}$$

$$S_{xx} = S_{yy} = -\tfrac{1}{2} S_{zz} \tag{37}$$

S_{zz} and S_{xx} are used to define the motionally averaged magnetic g-tensors $g_\parallel$ and $g_\perp$ in terms of its average value g_0 and the anisotropy Δg and δg (Equations 38 and 39):

$$g_\parallel = g_0 + \frac{2}{3}\Delta g S_{zz} + \frac{2}{3}\delta g(S_{xx} - S_{yy}) \tag{38}$$

$$g_\perp = g_0 - \frac{1}{3}\Delta g S_{zz} - \frac{1}{3}\delta g(S_{xx} - S_{yy}) \tag{39}$$

with equivalent expressions for the effective hyperfine splitting $A_\parallel$ and $A_\perp$. The latter two are resolved in the experimental line shapes; see Figure 14. Thus S_{zz} of the nitroxide z-axis can be easily obtained from Equations (38) and (39) and the corresponding cone angle from Equation (36).

5.2.1.2 Slow Motion ($\tau_r \approx 10^{-9}$–2×10^{-7} s) For $\tau_r >$ 2 ns, Redfield's theory does not hold. The equation of motion for an electron spin is solved using a stochastic Liouville equation (SLE), developed by Schneider and Freed.[60] Although the description of the SLE approach is beyond the scope of this article, one can rationalize the effect of slow motion on ESR spectra in terms of the two-site exchange. The low- and high-field extremes of the powder spectra correspond to nitroxides lying with the z-axis parallel to the magnetic field. Rotation (i.e. exchange with any other orientation) results first in an exchange broadening of the line width and then partial averaging of the anisotropy. Line width and effective splitting are

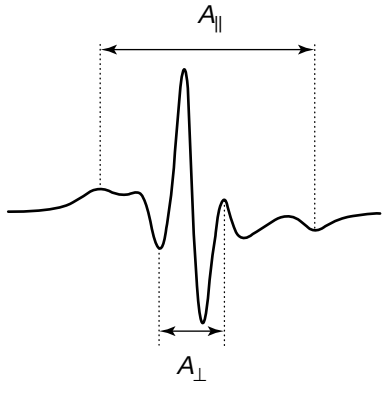

Figure 14 Definition of the parallel and perpendicular hyperfine splitting for calculation of order parameters.

used to determine τ_r (Equations 40 and 41):

$$\tau_r = a'_m \left(\frac{\Delta H_m}{\Delta H_m^R - 1} \right)^{b'_m} \quad (40)$$

$$\tau_r = a \left(1 - \frac{A'_{zz}}{A_{zz}^R} \right)^{b} \quad (41)$$

where ΔH_m and A'_{zz} are the line widths at half-height and hyperfine splitting, respectively, and the superscripts denote their rigid limit values. Coefficients a'_m, b'_m, a and b are calculated from SLE simulations. Their precise values depend on the motional model used for the simulations. For a Lorentzian line width $\delta = 3.0\,G$ and isotropic Brownian diffusion, $a'_{m=1} = 11.5\,ns$, $b'_{m=1} = -0.943$, $a'_{m=-1} = 21.2\,ns$, $b'_{m=1} = -0.778$, $a = 0.54\,ns$ and $b = -1.36$. Values for different line widths or motional models can be found in Marsh.[61] It should be noted that the calculated τ_r values depend strongly on the chosen rigid limit values. A user-friendly simulation and optimization program based on the SLE was developed by Budil et al.[50] Sensitivities of the conventional ESR and STESR spectra are illustrated in Figure 15(a) and (b).

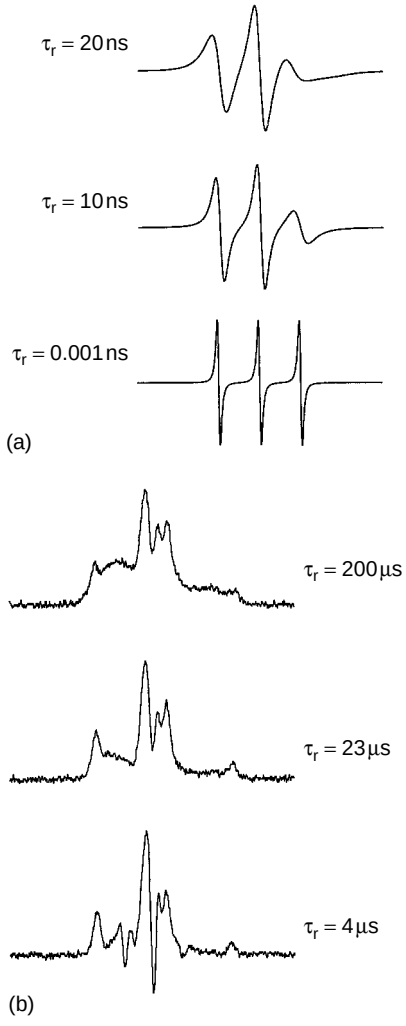

Figure 15 Sensitivities of (a) the conventional ESR spectra and (b) STESR spectra.

Examples. Side-chain and polypeptide backbone dynamics are determined using the above formalism. Spin labels attached to the surface of small α-helical peptides exhibit subnanosecond motions observed by ESR which compare well with motions predicted by molecular dynamics simulation programs.[45,62] Scanning of the label position along a peptide length reveals a V-shaped gradient of the label mobility. The cone angle for random motion in the middle of the peptide was half the value found at either terminus. Interestingly, the C-terminus was found to be more flexible than the N-terminus, which explains the decreased stability of the C-terminus as compared with the N-terminus in α-helices.[45,46] Backbone dynamics observed in isolated peptides are further modulated by tertiary interactions. A survey of 30 cysteine mutants of T4 lysozyme with spin labels at various structural sites (on the surface of helices, within the helix termini, interhelical loops, buried sites and sites involved in tertiary contacts) revealed a characteristic pattern of spin label mobility in relation to the secondary structure of the protein.[31] When the second moment of the spectrum (defined as the reciprocal

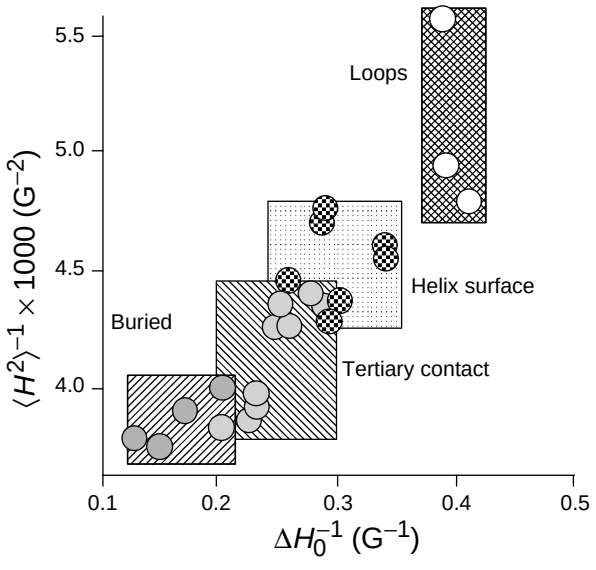

Figure 16 Reciprocal of the square of the splitting versus reciprocal of the central resonance line width. The spectral parameters cluster according to the labeled protein structural elements.[31] [Reprinted with permission from H.S. McHaourab, M.A. Lietzow, K. Hideg, W.L. Hubbell, *Biochemistry*, **35**, 7692–7704 (1996). Copyright 1996 American Chemical Society.]

of maximum splitting squared) is plotted against the reciprocal of the central field line width ($\Delta H_{m=0}^{-1}$), sites in similar environments are clustered together (Figure 16). The clustering reflects the degree of motional restrictions, with the second moment related to the averaging of the hyperfine anisotropy and the central line width to the averaging of the *g*-tensor. When motional restrictions increase, the averaging decreases and both the second moment and the line width of the resonances increase. Hence the second moment and line width can be used as semiempirical diagnostic tools to evaluate the secondary and tertiary structure of a labeled site.

5.2.1.3 Very Slow Motion ($\tau_r > 10^{-9}$ s) When $\tau_r >$ 100 ns, conventional ESR line shapes are no longer sensitive to motion. The rate of angular exchange is too small to affect the hyperfine or *g*-anisotropy and the line shapes become insensitive to very slow motions. To study these biologically important motions, a related ESR technique was developed, STESR.

Isotropic Motion. In the presence of power saturation, the second harmonic out-of-phase (V_2') line shape resembles an absorption spectrum (Figure 6c), with the intensity reflecting the effective relaxation at that point. Since effective relaxation is related to spectral diffusion and spectral diffusion is a function of the rotational correlation time, the V_2' line shape reflects rotational mobility. The rate of spectral diffusion (τ_{sd}) is a function of the resonant field H_{res}. Some field positions are more sensitive to angular rotation than others and $\partial H_{res}/\partial\theta$ varies across the spectral line shape. For instance, the rate of spectral diffusion is zero at the turning point $H^* = H_{res}(\theta = 0^\circ)$, but increases in the intermediate fields (Equation 42):

$$\tau_{sd}(H_{res})^{-1} = \left(\frac{8}{3\pi^2}\right)\left(\frac{\partial H_{res}}{\partial\theta}\right)^2 T_2^2\tau_r^{-1} \qquad (42)$$

To the first approximation,[63] the change of the signal intensity (I) at any field position is proportional to the change of the spin–lattice relaxation time due to spectral diffusion (Equation 43):

$$I(H_{res}) = I_0(H_{res})\frac{T_1^{eff}(H_{res})}{T_1} \qquad (43)$$

where I_0 is the rigid limit intensity in absence of motion and T_1^{eff} is the intrinsic T_1 modified by spectral diffusion according to Equation (44):

$$T_1^{eff}(H_{res}) = T_1^0\frac{1 + (I_0(H_{res})/T_2)T_1^0\tau_{sd}(H_{res})^{-1}}{1 + T_1^0\tau_{sd}(H_{res})^{-1}} \qquad (44)$$

Since T_1^{eff} is a function of the field position (spin angle with respect to field), it is customary to define the line height at precise positions in the spectrum: L'', C' or H'' at $\theta = 35^\circ$ (two-thirds of the way between resonant field corresponding to $\theta = 90^\circ$ and $\theta = 0^\circ$) and normalize it to the intensity at H^* (L, C and H positions) for which spectral diffusion is zero.

By substituting Equation (44) for the effective relaxation rate in Equation (43) a semiempirical expression for the P'/P ratio dependence on τ_r is obtained (Equation 45):

$$\frac{P'}{P(H_{res})} = \frac{I_0(H_{res})}{I_0(H^*)}\frac{1 + a/\tau_r}{1 + b/\tau_r} \qquad (45)$$

The parameters a, b and $I_0(H_{res})/I(H^*)$ can be estimated from Equations (42) and (44) by numerically evaluating sensitivity $\partial H_{res}/\partial\theta$ at each spectral position. In practice, these values are obtained from fits to the experimental curves of line-height ratios from spectra of molecules undergoing Brownian diffusion with a known τ_r. Spin-labeled hemoglobin or bovine serum albumin tumbling in media of a known viscosity (water–glycerol mixtures) is used for this purpose.[63] The rotational rate of hemoglobin (the abscissa in Figure 17) is calculated from the Stokes–Einstein equation for a sphere of radius r, tumbling in a medium with viscosity η (Equation 46):

$$\tau_r = \frac{4\pi\eta r^3}{3kT} \qquad (46)$$

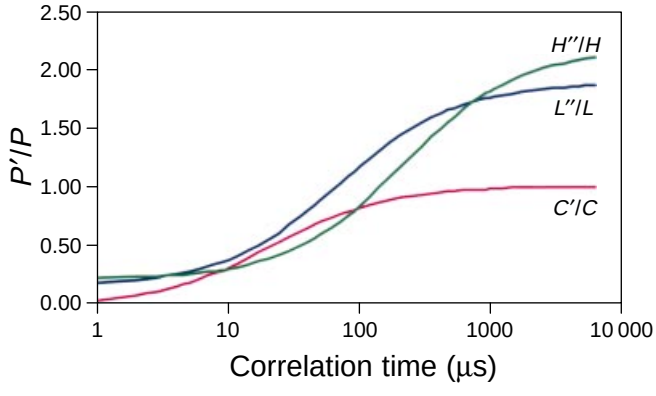

Figure 17 Dependence of V_2' diagnostic ratios on the rotational correlation time. The curves are simulated with Equation (63) using the parameter values from Table 1.

Table 1 STESR parameters from maleimide spin label–hemoglobin calibration curves[63]

Parameter	$I_0(H_{res})/I(H^*)$	a (μs)	b (μs)
L''/L	1.88	6.18	67.9
C'/C	1.01	0	21.1
H''/H	2.17	21.7	210

Table 2 Useful constants

Constant	Symbol	Value	Units
Planck's constant	h	6.63×10^{-34}	J s
	$\hbar$	1.06×10^{-34}	J s
Bohr magneton	β	9.27×10^{-24}	$J T^{-1}$
Free electron g factor	g	2.00232	
Electron magnetic moment	μ	-9.29×10^{-21}	$J T^{-1}$
Magnetogyric ratio	γ	$-1.76 \times 10^{+11}$	$s^{-1} T^{-1}$
Boltzmann constant	k	1.38×10^{-23}	$J K^{-1}$

The parameter values listed in Table 1 were obtained for maleimide spin-labeled hemoglobin tumbling in different water–glycerol mixtures. In principle, these values should be transferable from one laboratory to another. In practice, each cavity is sufficiently different in its microwave distribution and modulation fields that separate calibrations are often constructed. New calibrations are also necessary for different spin labels. Changes to the magnetic tensors and relaxation times alter STESR line shapes. In rare cases, full numerical simulation of the V_2' line shape is used to determine the correlation time, but the computational time required is still prohibitive.[68]

Anisotropic Motion. The effective rotational correlation times (τ^{eff}) obtained from such calibration curves reflect rates for isotropic rotation. However, isotropic motion is not very common in biological systems. For example, the nonspherical shape of the diffusing molecules or the restoring potential of the media results in anisotropic motion. Intuitively, rotation about the long axis of a cylinder is faster than the tumbling motion around its short axis. Assigning an isotropic τ_r to an anisotropic motion is obviously in error. For elongated molecules correlation times for rotation about the major and minor axes are given by Equations (47) and (48):

$$\tau_{\parallel} = \frac{f_{\parallel}}{4kT(1 + f_{\parallel}/2f_{\perp})} \tag{47}$$

$$\tau_{\perp} = \frac{f_{\perp}}{6kT} \tag{48}$$

where T is absolute temperature and the frictional coefficients $f_{\parallel}$ and $f_{\perp}$ are a function of the shape of the molecule.[64,65] For a cylinder of length $2a$ and radius b the frictional coefficients are given by Equations (49) and (50):[66]

$$f_{\parallel} = 8\pi\eta ab^2[0.96(1 + \delta_{\parallel})] \tag{49}$$

$$f_{\perp} = \frac{8\pi\eta a^3}{3[\ln(a/b) + \delta_{\perp}]} \tag{50}$$

where $\delta_{\perp}$ and $\delta_{\parallel}$ are as given by Beth and Robinson (Equations 51 and 52):[67]

$$\delta_{\parallel} = 0.688\left(\frac{b}{a}\right) - 0.202\left(\frac{b}{a}\right)^2 \tag{51}$$

$$\delta_{\perp} = -0.661 + 0.891\left(\frac{b}{a}\right) \tag{52}$$

The anisotropic diffusion tensor (D) creates an additional complication. The effect of the molecular rotation on the spectral line shape is a function of the label orientation with respect to the diffusion axis. If the principal axis of diffusion is parallel to the z-axis of the spin label, the motion interconverts the x- and y-components only. If it is parallel to the x-axis, then the y- and z-components will be mixed. To describe fully anisotropic diffusion of the anisotropic tensor, six parameters are needed: three diffusion coefficients about the x-, y- and z- axes and three Eulerian angles describing the orientation of the diffusion tensor with respect to the magnetic tensor.

The problem is simplified if either the diffusion tensor (D) and/or the magnetic tensors (g or A) are axially symmetric: the elements of the diffusion tensor are related to the correlation times by $\tau_{\perp} = 1/(6D_{\perp})$ with a corresponding expression for $\tau_{\parallel}$. It has been shown that the effective correlation time obtained from the L''/L and H''/H line-height ratios of STESR spectra ($m_I = \pm 1$) can be described in terms of $D_{\perp}$, $D_{\parallel}$ and the angle θ between the diffusion and magnetic tensor axis (Equation 53):[68]

$$\tau_R^{\text{eff}}(\pm 1) = \frac{1}{3[D_{\parallel}\sin^2\theta + D_{\perp}(1 + \cos^2\theta)]} \tag{53}$$

When $\theta = 0^{\circ}$ the outer manifolds reflect $D_{\perp}$ which defines the z- and x-(y-) element conversion ($D_{\parallel}$ leaves the nitroxide z-direction unchanged). If $\theta = 90^{\circ}$ the intensity of L'' and H'' is determined by $D_{\parallel}$, which now interconverts the z- and x-, y-axes.

In some cases, anisotropic rotation is about a single axis and the motion can be described by a uniaxial model. The mobility of transmembrane peptides or proteins in lipid membranes is a good example. A uniaxial model, with a single diffusion tensor element $D_{\parallel}$ and an angle θ defining the relative orientation of the magnetic and diffusion axes, is sufficient to simulate STESR spectra of membrane-bound proteins.[69] If θ is not known, then $\tau^{\text{eff}}(\pm 1)$ gives an upper estimate of $0.5\tau_{\parallel}$ ($\theta = 90^{\circ}$). It is important to realize that the changes of the θ angle, brought about by conformational changes, might result in STESR line-shape changes which can be mistakenly interpreted as changes in protein dynamics. A quick diagnostic for the presence of anisotropic motion is the comparison of the effective correlation times estimated from the C'/C ratio and from L''/L (H''/H). If they agree,

then the motion is likely to be isotropic. If they are different, then either the overlap of the nuclear manifolds is different in hemogiobin calibration spectra (unlikely) or the motion is anisotropic.[70] If the change in the STESR spectra is to be interpreted in terms of changed motional rate and not changed anisotropy of motion, then at least the ratio of the correlation times $\tau^{eff}(\pm 1)/\tau^{eff}(0)$ should stay constant.

Another motional model commonly encountered in biology is restricted diffusion. In such a model, the motion is isotropic but constrained in amplitude. The smaller the amplitude of motion, the slower is the apparent mobility derived from isotropic calibration. When the amplitude is $<30°$, the effective correlation time can be a factor of 10 larger than the actual τ_r and for amplitudes $>60°$, τ^{eff} approaches τ_r.[71] For small amplitudes, there is no isotropic line shape which will match the STESR spectrum of restricted motion. The amplitude effect is not just reflected in one or two places in the spectrum, but rather it is distributed across the whole line width.

An extensive review by Beth and Robinson[67] deals with the effects of anisotropic motion on STESR spectra and the theoretical simulations of line shapes. Numerical simulations are based on the transition rate matrix which couples neighboring angular zones with the rate of angular reorientation. The SLE approach and spin density matrix method have also been used. Both approaches have been applied successfully to isotropic and anisotropic motional models. The continuous increase in computational speed bodes well for the routine application of STESR simulations to analyze experimental data.

5.2.1.4 Examples in Muscle Proteins Microsecond motions are common for large macromolecular complexes (1 MDa) such as are present in muscle. The timescales of force generation, the actomyosin ATPase cycle and muscle activation coincide with the micro-to millisecond timescale of STESR, thereby making it the method of choice. The first application of the method established the dynamics of myosin, its subfragments and actin.[72] Thomas et al. showed that the myosin head is capable of moving independently of the large myosin filament. Such motion was a prerequisite for force production. When bound to actin, in the rigor state (no ATP) the myosin heads were immobilized but when ATP was added the heads detached and were free to move.[73,74]

Subsequent studies in muscle fibers at various intermediate states of the acto-myosin ATPase cycle have established a progressive decrease of catalytic domain mobility during the contractile cycle: the 10-μs motion of relaxed and weakly attached heads[75] became 80 μs just before force was generated[76] and was completely "frozen out" in the postpower stroke states of ADP and rigor. In the ADP state the head, although globally rigid, retained "breathing motions", which were suppressed on the release of nucleotide.[77] It is believed that this gradient in protein mobility reflects tighter and more stereospecific binding as myosin progresses through the contractile cycle (Figure 12).

The dynamics of the myosin head are complicated by the fact that this elongated protein does not behave like a rigid body. A comparison of the dynamics of the catalytic and regulatory domains revealed a three-fold difference in the rate of motion for the two domains.[78] Moreover, the two domains were found to have dramatically different orientational distributions.[57] These results highlight the complexity of the conformational changes in the actomyosin system: force generation is not synonymous with force transmission and both events involve changes of dynamics and orientation.

This complex behavior of myosin is in contrast to that of actin. Neither the orientation nor the dynamics of actin monomers, as probed by labels attached near the myosin binding site, were affected by head attachment.[79–82] The absence of any orientational changes in contracting muscle fibers was also observed using spin-labeled toxin phalloidin bound rigidly to the interface between the actin monomers.[83] This agrees with the current model of actin's passive role in force production in providing "tracks" for myosin motor protein to "walk" on.

Force activation involves a complex pathway with subtle changes in protein–protein interactions. It is mediated by the conformations and dynamics of the participating molecules. Smooth muscle is activated via phosphorylation of myosin light chain 2, whereas skeletal muscle is regulated by a thin filament based system involving Ca^{2+} binding to TnC. STESR spectra of phosphorylated myosin with a probe bound to myosin light chain 1 have implied increased motional freedom of the head. This finding supports a model[86] in which unphosphorylated heads are tied to the surface of myosin filaments and inhibited from binding to actin.[87] Phosphorylation abolishes the electrostatic attraction to the filament surface allowing the heads to interact with actin.

In skeletal muscle, binding of Ca^{2+} to TnC initiates a signaling pathway from the thin to thick filament which ultimately activates muscle contraction. Biochemical changes in the affinity of myosin for actin and of TnC for Ca^{2+} have a structural basis that is readily observed by both STESR and conventional ESR. The mobility and orientation of TnC (and TnI) has been found to be similarly affected by the binding of myosin heads to actin or by Ca^{2+} binding to TnC.[84] Interestingly, TnC was capable of sensing not only the binding of the myosin heads to actin but also the intermediate ATPase states.[85]

5.2.1.5 Examples in Membranes Rotational diffusion of membrane-bound proteins is often the best way of

determining the oligomerization state in their native environment without the possible dissociative effect of detergents. STESR of various integral proteins has revealed monomers such as rhodopsin,[88] dimers such as cytochrome oxidase[89] and ADP–ATP carrier[90] or even higher oligomers such as Na, K-ATPase.[91]

The biological significance of dynamic structural changes is best illustrated by the Ca-ATPase,[92] whose molecular dynamics correlate with transport activity.[93,94] As shown by ST-EPR, allosteric interactions between Ca-ATPase polypeptide chains and catalytically important domain interactions involved in the transport cycle are regulated by both alterations in membrane lipid composition, anesthetics, and the regulatory protein phospholamban.[95–97] Thus, physiological regulators of calcium transport modulate catalytically important motions and provide a structural basis for β-adrenergic stimulation in the heart.

Protein dynamics measured by STESR and conventional ESR have differentiated between two models of steroid biosynthesis in mitochondria: the shuttle mechanism and the ternary complex of adrenodoxin, P450 and adrenodoxin reductase. Adrenoxin was found to form binary complexes (but not ternary complexes) with either P450 or adrenodoxin reductase, supporting a shuttle mechanism.[98]

An excellent example of the potential of STESR in describing complex anisotropic motions is in the study of the transmembrane anion transporter Band 3 by Hustedt and Beth.[69] The STESR spectra were simulated using a uniaxial model for protein rotation. The diffusion rate and the angle between the magnetic and diffusion tensor were freely floated in the least-squares fits to experimental spectra. The uniqueness of the solution was corroborated by the orientational study of Band 3 in oriented erythrocytes.[55]

5.2.2 Mobility and Time Domain Methods

The measurement of the molecular dynamics by time-resolved ESR methods is still in its infancy. Specialized hardware is necessary to perform such experiments. Spectral diffusion due to the reorientation of spins can be observed either by recovery from saturation at the resonant field (saturation recovery ESR) or by arrival of saturation originally induced at some other nonresonant field (pulsed ELDOR). The initial promise of these methods was not fulfilled when it was shown that the nuclear relaxation, which couples different nuclear manifolds, contributed significantly to spectral diffusion. Combining pulsed ELDOR and saturation recovery differentiates between nuclear relaxation and rotational spectral diffusion and can be used to measure the true rotational correlation time.[99]

2-D FTESR methods appear to be more promising. Nuclear relaxation is seen as cross peaks between the manifolds and can easily be distinguished from homogeneous broadening and spectral diffusion broadening.[100] In the limits of fast motion, τ_r is obtained directly from the homogeneous line width and is defined by the pure T_2 (similar information is obtained from the spin-echo experiments). For slower motions, mixing time between the pulses is varied (2-D ELDOR) and the dependence of spectral broadening on mixing time is used to determine τ_r. Correlation times in the range $1-30\,\mu s$ have been measured for small peptides tumbling in viscous media.[101]

5.3 Kinetic Experiments

Elucidation of molecular mechanisms involves primarily two approaches: (a) entrapment of reaction intermediates with a subsequent reconstruction of the sequence of events and (b) transient kinetics in which the reactions are synchronized with the observed spectral changes. Each of these approaches have potential problems. In the ''trapping'' approach, the states have to be related to the kinetic intermediates. There are cases in which states trapped with substrate or product analogs are not lying on the kinetic pathway. On the other hand, transient experiments are easier to interpret, but technically more challenging owing to lower signal levels, fast acquisition times and difficulties in spectral assignment. The two approaches should be considered complementary. In an ideal world, ''trapping'' approaches should be used to identify and assign signals collected during transient experiments.

Historically, optical spectroscopy was used for transient kinetics owing to inherently higher sensitivity, but ESR is making substantial inroads.[102] Recent advances in resonator design allow for millisecond resolution on microliter samples in the submillimolar concentration range.[103] The DR developed for this purpose is capable of measuring millisecond kinetics in a single shot on $100\,\mu L$ of a $40\,\mu M$ sample with an 8-ms deadtime.[104] The further development of this DR/stop-flow configuration allows the recording of a full spectrum within 100 ms.[105]

Transient ESR was used to resolve the stages of channel formation in lipid membranes. Phospholipid vesicles and membrane channel collicin were mixed rapidly and the time course of the protein absorption to the membrane surface was clearly resolved from the insertion of the channel into the membrane.[106] For collicin the process was fairly slow, with a timescale of seconds, but the formation of another channel annexin was followed on the millisecond timescale.[107] The millisecond time resolution makes ESR a viable

alternative to optical methods in investigations of kinetic processes.

The photolysis of caged compounds, cATP and cCa^{2+}, to study conformational transients has been used primarily in the muscle field and in the study of Ca^{2+}-ATPase. A single ultraviolet (UV) pulse (10 ns in duration with an energy flux of 150 mJ cm^{-2} at 351 nm) from an excimer laser is capable of liberating 0.5 mM ATP (Figure 18a and b). The magnetic field is locked into a position where the initial and final states display a large spectral difference and the intensity at that position is followed in time. In myosin, the pre- and post-ATP hydrolysis states have different mobilities at a labeled residue near the catalytic site, with correlation times of $\tau_r > 100$ ns and $\tau_r \approx 80$ ns. This mobility difference and the associated line-shape difference was utilized in measuring the rate of transition between the two states ($43\,s^{-1}$) and was found to correspond to the hydrolysis rate of ATP[108] (Figure 18a and b). Similar experiments in muscle fibers, measuring both the orientation and the mobility, established that the rapid disorder of myosin heads follows nucleotide binding but precedes hydrolysis. These experiments also determined that the rate of hydrolysis is the same in fibers as in solution.[102]

Transient ESR of the Ca^{2+}-ATPase following cATP photolysis revealed local domain changes around the labeled site which correlate well with the formation of the phosphoenzyme intermediate.[109] A larger and more motionally restricted label was also used to observe global changes, e.g. shape or oligomerization state. No such changes were observed by transient STESR during the ATPase cycle.[110]

Biological photocycles encountered in rhodopsin and bacteriorhodopsin are special cases of cycles that are easily synchronized. The photoisomerization of retinal in bacteriorhodopsin initiates a series of proton transfer reactions via short-lived intermediates culminating in the loss of H^+ at the extracellular surface. Some 50 μs after photoactivation, an intermediate M decays to N when Asp96 transfers a proton to the Schiff base. During the decay of the N state, Asp96 regains a proton from the cytoplasmic site and bacteriorhodopsin reverts to the ground state, thereby completing the cycle. Labels have been attached to a number of cytoplasmic, interhelical and extracellular loops in the vicinity of Asp96 and their mobility was followed after irradiation with light.[111–113] Cytoplasmic sites and those near Asp96 all showed significant changes which coincided with the decay of the M state and recovered with the decay of the N state.[112] The efforts of Hubbell's and Steinhoff's groups to describe the molecular mechanism of rhodopsin and bacteriorhodopsin are an excellent example of the power of ESR methods to evaluate both static and transient molecular structures. In summary, the combination of high sensitivity, short mixing deadtimes, and temporal resolution makes ESR an increasingly popular method to study transient kinetics.

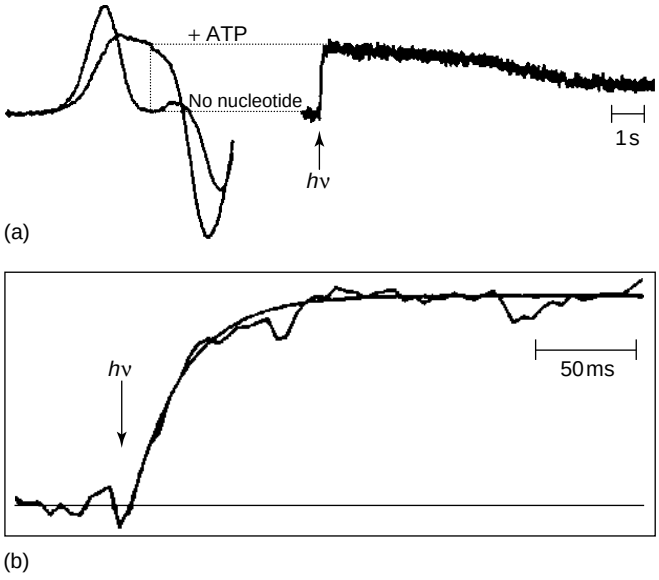

Figure 18 (a) Transient ESR of myosin head labeled with iodoacetamide spin label after the photolysis of cATP and (b) expanded transient.[108] (Reprinted with permission from E.M. Ostap, H.D. White, D.D. Thomas, *Biochemistry*, **32**, 6712–6720 (1993). Copyright 1993 American Chemical Society.)

5.4 Protein Folding

In the last few years, site-specific spin labeling has been applied to protein folding problems.[114,115] The advantage of the ESR approach to protein folding lies in site specificity as the denaturation of local domains can be followed independently of global denaturation. This approach relies on differences in the mobility of spin labels in folded and denatured proteins. The folded protein provides steric restrictions due to secondary structure and tertiary contacts whereas the denatured one does not. The ESR spectra for the denatured fraction are a composite of sharp, motionally averaged line shapes in contrast to broader, immobilized spectra observed for the folded protein. Fractions of protein in each form are easily calculated by spectral subtractions and by line shape integrations.[115] Cooperativity and stability of the given region are determined from spectral titration with a denaturing agent, e.g. GdnCl, urea or heat. Differences in the melting of hydrophobic and aqueous surfaces of the β-strand pore of FepA receptor were observed by cysteine scanning of the polypeptide chain lining the channel. The hydrophilic surface was more stable and cooperative in the transmembrane portion of the strand than the

extramembraneous strand ends. The residues exposed to the lipid exhibited noncooperative melting and did not denature completely even at the highest concentrations of denaturants.[116]

ESR is also capable of sensing multiphasic folding intermediates. Carbonic anhydrase was found to denature via an intermediate characterized by a compact and stable molecular core with a more dynamic periphery.[117] In the presence of the chaperonin GroEL, the intermediate core was destabilized and partially melted, explaining how GroEL allows for the refolding of misfolded proteins.[118] The development of new resonators, as mentioned previously, has facilitated detailed analysis of folding kinetics. The initial phases (<20 ms), ascribed to helix formation, were recently resolved by stop-flow ESR.[104]

5.5 Ligand Binding

Two less well known applications of spin labeling are the determination of the binding of small ligands and the aggregation of large (the latter reviewed in section 5.7.4). The binding of small ligands is followed by changes in their mobility. Spin label analogs of ligands have sharp, motionally narrowed spectra when free in solution. Binding to a larger target slows the motion of the label and the spectra become broadened. Spectral resolution between the broad/bound species and the narrow/free ligands allows quantification of bound species and hence calculation of binding isotherms. The binding of mellitin to α-crystallin[119] and the binding of nucleotide analogs to ATPases were determined by this method. Binding studies are not limited to ligands carrying a spin label. Competition of unlabeled and labeled analogs is used to determine the K_d of unlabeled ligands. The binding of ATP, ADP, adenosine thiotriphosphate (ATPγS), adenosine imidotriphosphate (AMPPNP) and adenosine methylenetriphosphate (AMPPCP) to myosin was determined by the displacement of spin-labeled ATP.[120]

A good example of ESR applications in ligand binding is the association of lipid spin labels with intrinsic membrane proteins. The differences in mobility are small and the spectra of bound and free labels are not resolved. Small broadening due to exchange between free and restricted environment was simulated to extract the equilibrium constant and the number of binding sites.[61]

Finally, the binding of metals to proteins can be established by ESR. Mn(II) has a characteristic six-line spectrum when free in solution, but no signal when bound to protein. The decrease of free Mn(II) signal upon addition of a protein identifies the fraction of bound metal.[121] As shown in the example with ATP analogs, displacement of bound Mn(II) by other metals can be used to determine their binding affinity.

5.6 Distance Measurements

ESR is capable of measuring short distances (2–25 Å) between selected sites. The method relies on the distance-dependent interactions between two spin labels (spin label–spin label method), or an interaction between a spin label and a paramagnetic metal (spin label–spin probe method). The physical basis for the coupling between the two spins is the exchange interaction (J) arising from the overlap of the orbitals of unpaired electrons and the dipolar interaction between magnetic moments of the two spin labels. The spin label–spin probe method relies on the enhancement of the relaxation of the spin labels by paramagnetic metals which have considerably faster relaxation rates and provide an efficient relaxation pathway for nitroxides.

5.6.1 Spin Label–Spin Label Method

5.6.1.1 Exchange Interaction, Distances <8 Å At distances shorter than 8 Å, the σ- or π-orbitals of the neighboring unpaired electrons can overlap, creating a single spin state: singlet (spins are antiparallel) or triplet (spins are parallel) state. Unlike nuclear spin, electron spin coupling propagates both through space and through covalent bonds. Through-bond coupling drops off very quickly and is insignificant at a distance greater than a few bonds. The through-space coupling strength (J) is a function of the interspin distance r_{dd} (Equation 54). J diminishes exponentially, with a 1-Å interaction distance from the initial value of $J_0 = 300$ G. In the strong exchange regime, when J is much larger than hyperfine interactions, the combined nuclear spin is 2 and the spectrum displays a five-line pattern. For weak coupling ($J < A_0$), the interaction asymmetrically broadens the low- and high-field resonances. The shift of the downfield edge of the low-field resonance in the presence of the second label (ΔH_{dd}) is proportional to J and hence to the interspin distance.

$$\Delta H_{dd} \approx J = J_0 \exp(-\beta r_{dd}) \tag{54}$$

For example, a 0.1-Gauss broadening implies an interspin distance of ~8 Å, defining the upper limit of the sensitivity.

Spin exchange in peptides labeled with two probes has been used to determine helical folds.[122] Some proteins contain 3_{10}-helices between β-strands or at the end of α-helices and their detection in solution is difficult. Fiori and Millhauser utilized a difference of distance between the i and $i+4$ residues along the α- and 3_{10}-helices to distinguish between the two helical forms. The i to $i+3$ distances are similar in both helices (6–8 Å), but the i to $i+4$ distance for the 3_{10}-helix is outside the range of the exchange interaction (10–16 Å) and within the

range for an α-helix (7–11 Å). Placing pairs of labels in the i, $i+3$ and $i+4$ positions along the synthetic helices identified the predominantly α-helical regions in the N-terminus (equal broadening of i, $i+3$ and i, $i+4$ pairs) with the 3_{10}-helix at the C-end (no broadening of the i, $i+4$ pair). A comparison of 16- and 21-residue peptides revealed a length-dependent equilibrium between the α- and 3_{10}-helices. Shorter peptides favored the 3_{10}-helix whereas the longer peptides favored the α-helix.

Transition between the two helical forms might provide a mechanical pathway for allosteric mechanisms. The α-helix has more residues per turn and is significantly shorter, hence the 3_{10}-helix to α-helix transition will mechanically pull on the neighboring domains.

5.6.1.2 Dipolar Coupling Distances: 8–25 Å In addition to the exchange interaction, neighboring spins experience local fields induced by their respective magnetic dipoles. The local fields can add to or subtract from the external field splitting each of the resonances. This dipolar interaction has an r_{dd}^{-3} dependence which results in discernible line-shape changes for distances up to 25 Å. The contribution of the exchange interaction for distances over 10 Å is negligible.

The strength of the dipolar interaction is a function of eight parameters: the interspin distance r_{dd}, the angle between the static magnetic field and the interspin vector and the six angles describing the orientation of each of the spin labels with respect to the interspin vector. Protein dynamics also affects these parameters on the timescale of the experiment: the global motion of the molecule modulates the angle with respect to the field, intradomain motions modulate the interspin distance and the interspin angles. The full solution of dipolar ESR line shapes in the presence of motion poses a formidable challenge.

The geometry of the NAD coenzyme bound to glyceraldehyde dehydrogenase (GAPDH) is the best example of the approaches taken to tackle this problem. Since GAPDH is a tetramer it can bind four molecules of coenzyme NAD. The relative orientation of the coenzyme was determined from the dipolar splitting of a spin label analog of NAD. Hustedt et al. used different microwave frequencies (9, 35 and 94 GHz) coupled with sophisticated fitting procedures to solve independently for most of the parameters listed above.[123] The ESR structure was consistent with the geometry derived from molecular modeling using a crystal structure of the apo-enzyme.

Another example is the open form of α-helices in protic solvents.[124] Alanine-rich peptides, incorporating the unnatural spin-labeled amino acid TOAC (2,2,6,6-Tetramethylpiperidine-1-oxyl-4-amino-4-carboxylic acid), showed a strong dipolar interaction for the i, $i+4$ labeled peptides with less interaction between the i, $i+3$ sites. Molecular modeling revealed a shorter interspin distance and larger backbone torsional angles, consistent with 3.8–3.9 residues per turn compared with 3.6 residues for a standard α-helix. The open configuration allows for the formation of additional hydrogen bonds with amide carbonyls while preserving helical hydrogen bonds. ESR proves to be an important tool in identifying subtle changes in helical folds induced by the local environment.

Short of solving the general case as above, two simpler cases are often encountered: (1) the static dipolar case in which all motions are frozen out, the labels are disordered on the surface of the protein and the protein itself is isotropically disordered in the magnetic field, and (2) the motionally averaged case in which the anisotropy of the dipolar interactions is averaged ($\tau_r < 6$ ns) and the spectrum is homogeneously broadened by spin–spin interactions.

In the first case, the nitroxides and the sample are rigid and each of the resonance peaks are broadened by the Pake function given in Equation (55):

$$\Delta H_{dd} = \pm \frac{3g\beta(3\cos^2\theta - 1)}{4r^3} \tag{55}$$

where θ is the angle of the interspin vector in the magnetic field.

Convolution of the resonances with the Pake pattern yields a characteristic line shape with dipolar wings in the low- and high-field regions of the spectra. The Pake function is obtained from the spectrum by a Fourier transform,[125] line shape fitting[126] or by empirical calibrations.[38] Empirical calibrations relate the line heights of the low- and high-field resonances to the line height of the central peak (Equation 56):

$$r_{dd} = 9.3 + \frac{0.77}{(d_1/d) + 0.36A_{zz} - 1.76} \tag{56}$$

where r_{dd} is in ångstroms and A_{zz} is in gauss, d is the line height of central resonance, and d_1 is the difference between low-field peak and high-field trough.

Alternatively, the Van Vleck relation between the second moment of the central resonance M_2 and r_{dd} is used in obtaining the distance (Equation 57):

$$M_2 = \frac{3}{20}\frac{g^2\beta^2}{r_{dd}^6} \tag{57}$$

where r_{dd} is in ångstroms. The second moment is given by the Gaussian part of the peak-to-peak line width (Equation 58):

$$M_2 = \left(\frac{\Delta H_{pp}^G}{2}\right)^2 \tag{58}$$

The Pake broadening approach was verified for interspin distances of doubly labeled rigid synthetic polypeptides in which distances in the range 8–24 Å were predicted from the structure.[125] The method was also verified for larger molecules by comparison with an X-ray determined structure of spin-labeled insulin. The interspin distances in a crystal were in excellent agreement with those found in solution by ESR. Distribution of distances due to protein flexibility in solution was found to be twice as large in solution as in the crystal.[126] It is important to note that ESR is one of the very few techniques which can estimate conformational heterogeneity of a protein in solution.

Static dipolar broadening is rapidly gaining popularity. It was used in the analysis of helix packing of lactose permease in the membrane.[127] Spin labels placed on surfaces which faced each other displayed dipolar broadening. Analysis of the extent of broadening as the label is moved around the helix identified the relative rotation of neighboring helices. Dipolar broadening was also used in the elucidation of the opening/closure of K-channels[128] as described in detail in section 5.7.4.

The static dipolar approach (case 1) fails when the proteins are not rigid on the nanosecond timescale of an ESR experiment. If the spin mobility is high enough to modulate the anisotropy of dipolar interactions (Equation 59)

$$\tau_r \leq \left(\frac{g\beta\Delta H_{dd}}{h}\right)^{-1} = \left(\frac{3\pi g^2\beta^2}{r_{dd}^3 h}\right)^{-1} \qquad (59)$$

the rotational modulation of the interspin vector results in the broadening related to both the distance and the correlation time (Equation 60):

$$\Delta H_{dd} = \frac{3}{10}\frac{\gamma^4}{g\beta}\hbar^3\frac{\tau_r}{r_{dd}^6}\left(3 + \frac{5}{1+\nu^2\tau_r^2} + \frac{2}{1+4\nu^2\tau_r^2}\right. \qquad (60)$$

Line-shape comparisons of singly and doubly labeled samples reveal the extent of dipolar broadening from which the r_{dd} distance is calculated assuming (or measuring) an appropriate correlation time for the molecule. McHaourab et al.[129] tested this approach on a series of labeled sites in T4 lysozyme (8 Å < r_{dd} < 23 Å) and found them to be in excellent agreement with the structure of the protein (Figure 19a and b). Note that the motional modulation regime is a function of both the interspin distance and the rotational correlation time, e.g. dipolar interactions of labels 15 Å apart are averaged for motions with correlation times of ≤6 ns.

The above examples illustrate the static and motionally averaged cases of spin–spin interactions. However, often the precise mechanism of spin–spin interactions is not known as there are contributions from static dipolar interactions, from the rotational modulation of the interspin vector and from the modulation of the interspin distance. Although there is currently no theory fully describing the changes in the ESR line shape, the presence of line-width broadening is always indicative of two labels being in the range 10–20 Å from each other. Even the simplest, qualitative statement of spin–spin distance can yield important structural information. The presence of spin–spin interactions has helped to elucidate changes in the topology of the cytoplasmic portion of rhodopsin following light activation. It has explained how rhodopsin initiates the phosphorylation cascade by rhodopsin kinase.[130]

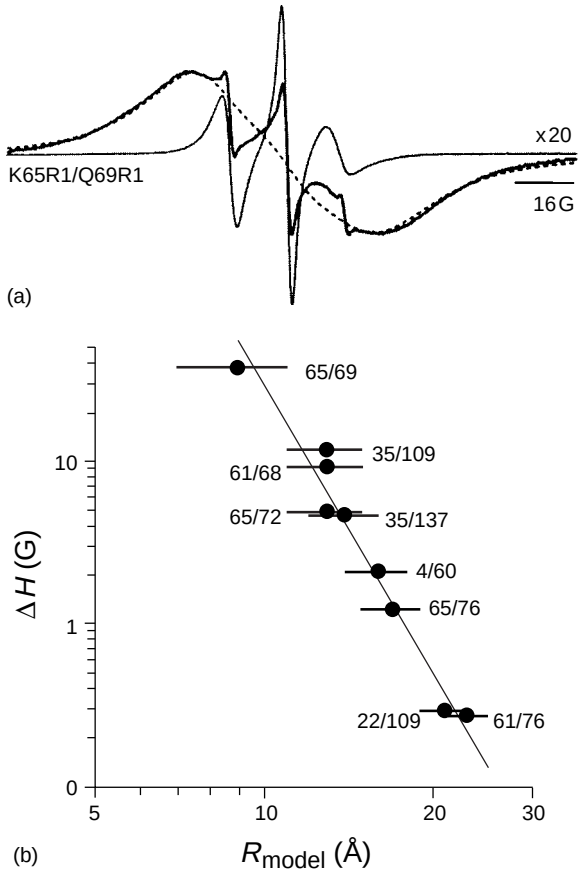

Figure 19 (a) An ESR spectrum broadened by the presence of another nitroxide and (b) the line width as a function of the interspin distance.[129] (Reproduced with permission from H.S. McHaourab, K.J. Oh, C.J. Fang, W.L. Hubbell, *Biochemistry*, **36**, 307–316 (1997).)

Interested readers are directed to an excellent review by Hustedt and Beth.[47]

5.6.1.3 Very Long Distances: >25 Å The upper limit of 25 Å for the detection of dipolar interactions is defined by an observable broadening: ~3 G for inhomogenously

broadened samples of rigid samples and 0.2 G for the sharp, motionally narrowed line shapes. This limit can be extended by time-resolved techniques, which unlike the continuous-wave methods do not rely on line broadening. A variety of pulsed ESR methods have been developed for this purpose. Double electron–electron resonance (DEER) is one method in which the spin echo is observed as a function of time of an intervening pulse applied at a second frequency,[131,132] or at the same frequency as in "2 + 1" resonance.[133] In both cases, an amplitude of the spin echo is modulated by Pake dipolar interactions which are extracted by Fourier transformation of the echo modulation. In model compounds, the distances recovered by DEER ranged between 20 and 33 Å,[132] whereas the "2 + 1" scheme elucidated an interspin distance of 35 Å between nitroxides attached to β-93 cysteine in tetrameric hemoglobin.[133] Double quantum 2-D FTESR is the most recent technique that looks directly at spin–spin interactions i.e. the filtering out of a "normal" ESR spectrum originating from isolated spins. This method has been tested on solid polyproline peptides with an interspin distance of 18 Å.[101]

5.6.2 Spin Label–Paramagnetic Metal Method

An entirely different way of determining distances is by coupling an unpaired spin of a nitroxide radical to a fast-relaxing spin of a paramagnetic metal, e.g. Cu(II) or Fe(III). The spin–lattice relaxation times of many metals are three orders of magnitude faster than those of nitroxide spin labels and thus, when coupled to nitroxides, they provide an efficient relaxation path. The relaxation enhancement can be detected by either increased Lorentzian broadening or directly by the shortening of T_2 or T_1 of the spin label in the presence of the metal (Equation 61):

$$\Delta\left(\frac{1}{T_{1s}}\right) = \frac{\mu^2\gamma^2}{6r^6T_{1f}}\left[\frac{4}{5(w_1-w_2)^2}+\frac{24}{5(w_1+w_2)^2}+\frac{12}{5w_1^2}\right] \quad (61)$$

All three methods have been tested in spin labeled hemoglobin in which a nitroxide at residue 93 of the β-chain was coupled to Fe(III) bound to the heme. The distance of 15 Å between the label and iron was in excellent agreement with molecular modeling.[134]

The metal relaxation is not limited to naturally occurring metal binding sites. Site-specific spin labeling has recently been extended to the engineering of metal binding sites.

Voss et al. engineered Cu(II) binding sites to T4 lysozyme and lactose permease by introducing histidine residues in consecutive turns of the α-helix.[135,136] Spin labels were placed between 8 and 18 Å away from the metal site by cysteine mutagenesis. The ESR line shapes were broadened by paramagnetic Cu(II) chelated by the histidine residues. Binding of diamagnetic metals did not affect the line shape. The spectra were analyzed in terms of the dipolar model of Leigh,[137] which relates line-shape broadening to the interspin distances r_{dd} and metal relaxation time T_{1m} (Equation 62):

$$\Delta H_{dd} = \frac{g\beta\mu^2 T_{1m}}{\hbar r^6}(1-3\cos^2\theta)^2 \quad (62)$$

where μ is the magnetic moment of the metal.

A convenient experimental parameter is the amplitude of the central line. The amplitude decreases with decreasing spin–spin separation. Calibration curves derived from computer simulations were used to estimate the interspin distances in two model systems, lactose permease and T4 lysozyme. Excellent agreement was found, not only for the rigid samples for which Leigh's model was originally developed, but also for motionally narrowed spectra due to the mobility of the nitroxides with respect to the protein. For Leigh's model to hold, the interspin vector must not move on the same timescale as dipolar interactions, which is T_1 of the metal. This is satisfied for molecules larger than 15 kDa ($\tau_r > 6$ ns) and for metals such as Cu(II) where T_1 is 1–3 ns. The distances obtained for T4 lysozyme in solution at room temperature were approximately 1 Å shorter than those obtained from frozen proteins.[136] This small underestimation of the distance is compensated by the biological advantage of performing experiments at room temperature and by the increased fidelity in measurements of small-amplitude changes in sharper, motionally narrowed spectra. Another elegant application of this method involved the determination of helix packing in lactose permease. Interspin distances between three helices labeled with nitroxide labels and a metal site containing Cu(II) determined the relative orientation of the helices and their relative tilt.[138]

Currently this method is limited to distances between 10 and 20 Å for Cu(II) with the X-band. Inspection of Leigh's Equation (61) suggests that the distance range might be extended to ~50 Å by using lower ESR frequencies (S-band), metals with a larger magnetic moment (Gd^{3+}) or shorter T_1 (Ni^{2+}) and also by direct measurement of relaxation times.[139]

5.6.3 Collision Exchange

A variation of spin–spin interactions is the relaxation of spin labels by collisions with soluble paramagnetic agents such as metals or O_2. Collisions lead to the HSE, enhancing spin–lattice relaxation according to Equation (63):

$$\Delta T_1^{-1} = kW_x \quad (63)$$

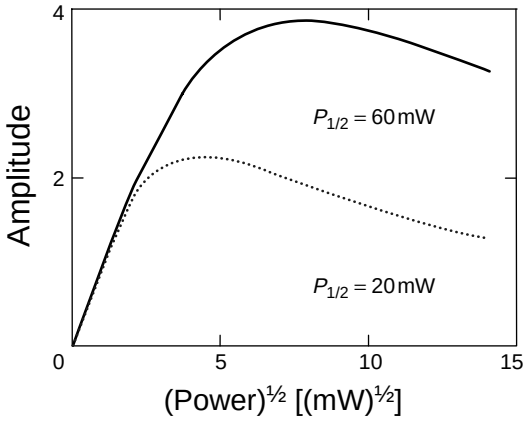

Figure 20 Power saturation of two samples with different $P_{1/2}$ values.

where k is a factor accounting for the efficiency of collisions and statistics of diffusion in two or three dimensions and W_x is the bimolecular collision frequency. The collision frequency is of interest because it reflects the accessibility of the labeled site to the relaxant. Comparison of W_x for various sites reveals which residues are exposed or hidden and their secondary structure content and identifies tertiary interactions.

Relaxation enhancement is measured directly by pulse methods (saturation recovery or spin echo ESR) or by continuous-wave power saturation. The amplitude of the ESR signal increases linearly with the microwave magnetic field ($H_1 \propto P^{1/2}$) until the Boltzmann equilibrium population difference is perturbed and the signal between excited and ground states decreases. Samples with a long T_1 saturate easily and addition of relaxing agents relieves this saturation (Figure 20).

The peak-to-peak amplitude (A) of the first-derivative spectrum is given by Equation (64):[140]

$$A = \frac{A_0\sqrt{P}}{1 + \left(\sqrt[\varepsilon]{2} - 1\right)(P/P_{1/2})^{\varepsilon}} \tag{64}$$

where ε depends on the resonance line shape and varies between 0.5 for purely Lorentzian and 1.5 for Gaussian line shapes; A_0 is an instrument scaling factor and $P_{1/2}$ is the half-saturation power (the power at which the signal is half of what it would be in the absence of saturation). $P_{1/2}$ is determined either graphically or by the fitting of experimental curves to Equation (64). The $P_{1/2}$ value is then used in calculating T_1 according to Equation (65):

$$T_1 = \frac{2^{2/3} - 1}{\gamma^2 \Lambda^2 P_{1/2} T_2} \tag{65}$$

where Λ is an instrumental factor which depends on the power-to-magnetic field conversion of the resonator. Since T_2 for nitroxides is 2–3 orders of magnitude smaller than T_1 and because T_2 is proportional to the peak-to-peak line width of the central line ($T_2 \propto 1/\Delta H_0$), the collision frequency is determined from Equations (63) and (65) (Equation 66):

$$W_x \propto \Delta P_{1/2} T_2 = \frac{\Delta P_{1/2}}{\Delta H_0} \tag{66}$$

In order to account for differences in resonators and spectrometers between various laboratories, a dimensionless accessibility parameter π was defined; π normalizes W_x to the half-saturation power and line width of a diphenylpicrylhydrazyl (DPPH) standard.[141] (Equation 67):

$$\pi = \frac{\Delta P_{1/2}}{\Delta H_0} \frac{\Delta H_0^{\mathrm{DPPH}}}{P_{1/2}^{\mathrm{DPPH}}} \tag{67}$$

Trends in accessibility to various relaxing agents are used to determine the local environment of spin labels. The relaxants can be nonpolar such as O_2 partitioning into lipid bilayers, or polar with preference for the aqueous phase. The latter includes neutral relaxants such as NiAA [nickel(II) acetylacetonate] and NiEDDA [nickel(II) ethylenediaminediacetate] and charged relaxants such as CROX [potassium tris(oxalatochromate)].[142]

An important application of collisional relaxation is the determination of the secondary structure of peptides and proteins. Patterns of collisional accessibility along the polypeptide chain can reveal α-helical folds, β-sheet strands, immersion in membranes and chain tilt within the membranes. These applications are described at length in section 5.7.2.

5.7 Structural Biology

The advent of site-specific spin labeling established ESR as a structural technique.[41–43,46,142] In the first study,[30] comparison of the relaxations enhancement of four labeled cysteine mutants of bacteriorhodopsin identified membrane embedded and surface exposed residues. Since then, these approaches have been refined and extended to establish (a) the topology of membrane bound proteins, (b) the secondary structure of proteins by cysteine scanning and following trends in accessibility and mobility of residues and (c) the tertiary folding of proteins by distance measurements between engineered sites. Most of the examples discussed below are from work of Hubbell et al.

5.7.1 *Side-chain Environment – Immersion in Membranes*

Differential effects of nonpolar (O_2) and polar (NiAA, CROX) relaxants are used to measure the immersion depth of membrane proteins. This technique relies on opposite concentration gradients for polar and nonpolar relaxants within lipid membranes. The concentration of nonpolar reagents increases with the immersion depth and the concentration of polar reagents decreases. The further the nitroxide is from the aqueous interface, the stronger is the relaxation enhancement by nonpolar reagents and the weaker is the effect of polar relaxants. The difference (Φ) between the polar and nonpolar reagents as defined in Equation (68) is thus a function of the immersion depth:[140]

$$\Phi = \ln \frac{\Delta P_{1/2}^{\text{nonpolar}}}{\Delta P_{1/2}^{\text{polar}}} \tag{68}$$

Calibration curves of Φ are constructed using lipid spin labels with nitroxides at defined positions along the acyl chains and these curves are used to determine the immersion depth of labels attached to membrane-bound proteins (Figure 21a and b).

This relatively simple method, when used in conjunction with cysteine scanning, differentiates between membrane-bound and solvent-accessible surfaces of membrane-associated proteins or peptides. As the nitroxide is moved along the length of a polypeptide chain, $\Delta P_{1/2}$ for the polar agent shows minima and maxima for the residues interacting with the lipid bilayer and the surface-exposed residues, respectively. The nonpolar agent has a similar pattern of minima and maxima, but it is offset by 180° with respect to the polar relaxant, i.e. maximum relaxation will be observed for the residues interacting with the membrane and minimum relaxation for water-exposed residues. A similar phase shift is observed for the residues of helices lining the aqueous pores of channels. Residues facing the lumen of the pore show maximum relaxation enhancement for polar agents (minima for the nonpolar oxygen), while the residues facing the membrane environment have maxima for oxygen and minima for CROX, NiAA and NiEDDA.

Such is the case for the ferric enterobactin receptor FepA, the transmembrane β-strand of which was found to line an aqueous channel. The maxima of accessibility to NiEDDA was alternating with maxima to O_2, identifying the β-strand face as lining the channel and the side of the β-strand facing the lipid bilayer.[143] Similar results have identified residues lining the aqueous channels in collicin,[144] diphtheria toxin[145] and annexin.[107]

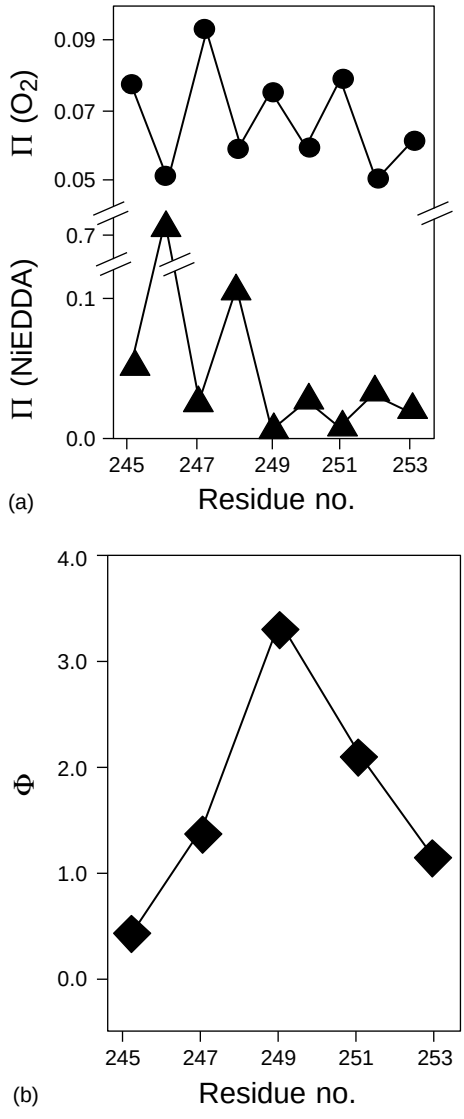

Figure 21 Immersion of a polypeptide chain in the lipid bilayer. Differential accessibility to polar ( ) and nonpolar (●) relaxants, Φ, identifies the distance from the aqueous surface. Calibration curves of Φ are constructed using lipid spin labels with nitroxides at defined positions along the acyl chains.[143] (Reproduced with permission from C.S. Klug, W. Su, J.B. Feix, *Biochemistry*, **36**, 13027–13033 (1997).)

5.7.2 *Secondary Structure Determination*

Cysteine scanning also allows for secondary structure determination. One method, based on the exchange interactions between nitroxides attached to the i, $i+3$ and i, $i+4$ residues, identifies α- and 3_{10}-helices and was discussed in section 5.5.1.1. Other methods rely on changes in nitroxide mobility and accessibility to relaxing agents. The periodicity of steric interactions varies along the polypeptide chain, which in turn determines the nitroxide mobility and/or periodicity of relaxation effects.[146] For example, an α-helix in an unevenly

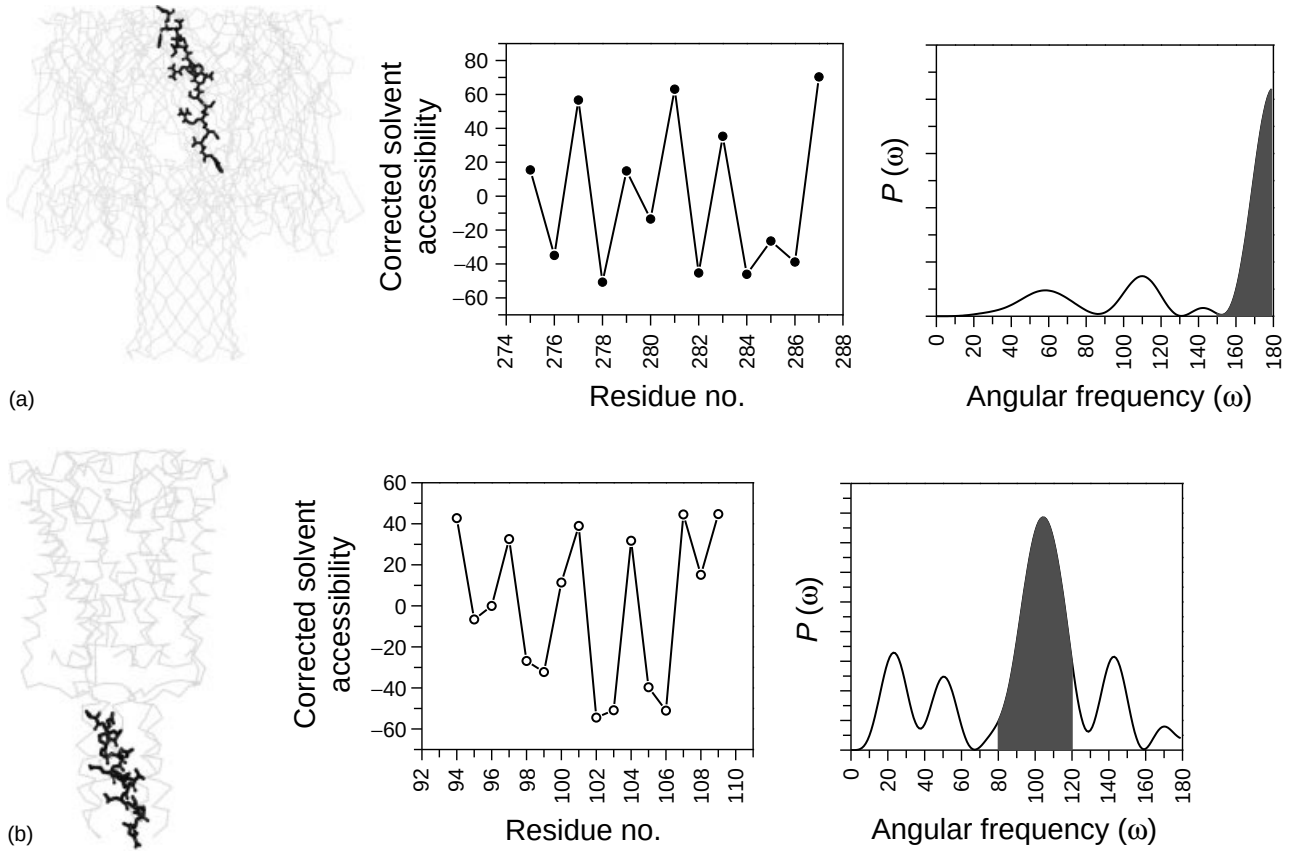

Figure 22 Patterns of solvent accessibility for (a) β-strands of α-hemolysin (two-residue periodicity) and (b) an α-helix of *Streptomyces* K^+ channel (3.6-residue periodicity). The Fourier transform identifying angular periodicity characteristic of α-helix and β-strands is illustrated. (Courtesy of E. Perozo, unpublished.)

solvated environment (owing to the interaction with a membrane surface or another polypeptide chain) shows a pattern of flexibility and solvent accessibility with a 3.6-residue periodicity. β-Strands, on the other hand, will display a two-residue periodicity (Figure 22a and b).

This characteristic periodicity of 3.6 residues was observed for a number of helices of transmembrane proteins: rhodopsin,[146–149] collicin,[143] K-channel[32] and the soluble protein T4 lysozyme.[31] Periodicity of β-strands was observed for the transmembrane protein FepA receptor[143] and water-soluble α-crystallin.[150,151] In some cases ESR has extended the structural information obtained by other methods, for example interhelical loops in rhodopsin.[147] However, in other cases, the secondary structure determined by ESR was the only available source for example FepA receptor[143] and α-crystallin.[150]

An interesting use of SDSL ESR is to extend monomeric (subunit) structures determined by NMR and X-ray crystallography to the structures of functioning macromolecular complexes. The monomeric structure of the soluble (nonfunctional) form of the membrane pore annexin has been solved by X-ray crystallography. A mobility and accessibility profile of 26 single cysteine mutants in the helix–loop–helix motif has revealed a dramatic structural transition when annexin is inserted into the membrane to form a continuous, transmembrane α-helix. As was expected for a membrane pore made of the annexin trimer, one side of the helix was found to be highly solvated.[152]

5.7.3 Tertiary Structure: Conformational Changes

The greatest potential for the above methodologies is the determination of tertiary structure. The current rate of structure determination by X-ray crystallography or NMR (~1000 per year) is too slow to solve for all 120 000 gene products. Fortunately, most proteins are built from well defined, common structural motifs, but are packed in different ways to give proteins their unique three-dimensional structure. It seems that instead of solving ab initio the atomic structure of each protein it will be simpler to determine the relative arrangement of common structural motifs. For instance,

a few chosen mutations can quickly establish whether given helices or β-strands are in a parallel or antiparallel arrangement.[143,153]

Qualitative information about tertiary structure is obtained from mobility and solvent accessibility values, both of which are limited to sites of tertiary contacts, e.g. π (O_2) values are 0–0.05 for buried sites and 0.3 for solvent-accessible surface sites.[119] Additionally, in-phase tracking of accessibility to polar and nonpolar reagents and tracking of mobility patterns identifies surface residues and residues buried within a protein core. The surfaces involved in helix packing in rhodopsin[147,148] and collicin[144] and β-strand packing in α-crystallin[151] have been identified by this in-phase behavior.

The tilt angle of polypeptide chains within lipid membranes is easily determined from the immersion depth of selected residues (see section 5.7.1). The immersion depth is calculated from the Φ parameter (ratio of nonpolar to polar accessibility), which is calibrated in terms of the distance from the membrane surface of lipid spin labels at defined positions. The average depth (d) of consecutive residues is compared with the distance along the chain (d_0). The tilt of the chain with respect to the bilayer normal is given by Equation (69):

$$\alpha = \cos^{-1}\left(\frac{d}{d_0}\right) \tag{69}$$

For β-barrels, the tilt information, combined with number of β-strands, can be used to estimate the diameter of the barrel.[144]

Most of the examples identifying conformational changes are from membrane-bound proteins. In rhodopsin, helices flanking the ionone ring of retinal have been labeled with nitroxides and the interspin distance tracked upon photoisomerization of the retinal.[130,154] The observed rigid body rotation, with an associated change of the tilt angle in one of the flanking helices, resulted in increased accessibility of the cytoplasmic loop. Increased exposure of the loop facilitates binding of transducin to rhodopsin, which is the first step in the phosphorylation cascade of signal transduction pathway.

Conformational changes accompanying insertion into a membrane and pore formation were observed by ESR for the small cytosolic protein annexin. A water-soluble monomer with a helix–loop–helix motif was rebuilt to form a continuous transmembrane helix in the presence of Ca^{2+}. The formation of long helix induced membrane insertion of annexin.[107] In another example, smaller conformational changes were observed by varying the lipid environment of transmembrane proteins. Reconstitution of lactose permease into proteoliposomes induced a small 2-Å movement of neighboring helices.[127]

A particularly rewarding example is that of conformational changes in T4 lysozyme: two structures solved by X-ray crystallography implied a hinge movement which opened the active site by 8 Å. Using strategically placed cysteines near the active site, this predicted opening of the active site was verified in solution.[129] In addition to corroborating the presence of the two conformers, ESR was able to measure an equilibrium of closed and open structures, yielding a unique estimate of activation energy associated with catalysis.

The most spectacular application of ESR to the tertiary/quaternary structure of proteins was that of the bacterial potassium channel by Perozo.[32,128] Nearly a third of the entire protein including two transmembrane helices and the interhelical region flanking a selectivity filter have been scanned with spin labels. This is a total of 62 mutants for the 160 amino acid polypeptide chain. The channel is formed by the tetrameric assembly of the two helices, with one helix (TM2) forming an aqueous pore and the other helix (TM1) located on the periphery. The structure has been solved independently by X-ray crystallography[156] and by ESR from accessibility and mobility profiles.[32] The ESR protein structure determination was further extended to the structural description of the channel opening. The channel is activated by lowering the pH. Sequence profiles of mobility and interspin distances were compared for the open and closed forms, revealing a physical opening of the central pore. The open form was brought upon by a rigid body rotation and tilting of the TM2 helices with an accompanying movement of the peripheral TM1 helices.[128]

5.7.4 Assembly of Polypeptide Chains: Quaternary Structure

Differences in mobility between monomers and oligomers can be used to identify the oligomerization state of proteins. For small proteins, ESR line shapes are motionally narrowed, whereas the spectra of aggregates are considerably broader. Spectral resolution of monomeric and oligomeric forms in a composite spectrum allows for the determination of their respective concentrations in solution and hence thermodynamic parameters of oligomer formation. Oligomerization in a variety of solvents of cecropin AD, a small ion channel, was studied in this way.[157]

The kinetics of the formation of amyloid plaques were monitored by the disappearance of the sharp central line of the spin-labeled amyloid protein monomer.[158] Monomers were found to aggregate initially into an amorphous plaque precursor in which the protein was in equilibrium between soluble monomers and the aggregated protein. The precursor was an initiation site for fibril formation of which the amyloid

plaques are subsequently formed. Characterization of various assemblies and the equilibria between monomers and aggregates are of direct interest in understanding the molecular basis for diseases such as Creutzfeldt–Jacob (''mad cow disease'') and Alzheimer's disease.

An alternative way of following the formation of aggregates is to utilize spin–spin interactions. Interacting monomers lead to a broadening of the ESR spectra, provided that the labels are within the range of spin exchange or dipolar interactions. Spin-labeled insulin B chain was found to aggregate on reduction of the interchain disulfide bonds, but the presence of α-crystallin was found to prevent aggregation. In the absence of crystallin, the spectra of B insulin displayed a broad Lorentzian pattern, characteristic of closely placed spins. This turned into a normal powder pattern upon binding to crystallin.[119]

Titration of spin–spin interactions with unlabeled proteins can be used to estimate the number of monomers forming an oligomer. As the concentration of unlabeled monomers increases, the probability of spin–spin interactions decreases and dipolar broadening is relieved leading to an increase in signal amplitude. For small oligomers, dilution with small amounts of unlabeled protein results in a greater increase of signal amplitude than for larger oligomers. The titration follows a binomial expansion and has been used to establish that the membrane-bound form of annexin is a trimer.[107]

This qualitative approach has been used for annexin pores. While the crystal structure of the soluble form of annexin suggested a hexameric assembly, nothing was known about annexin in membranes. Among the possibilities were a trimeric ring and a hexamer consisting of stacked trimers. To distinguish between these alternatives, cysteines were introduced at the interface between the monomers forming a trimer and on the interface between the trimers forming a putative hexamer. In the soluble form, no dipolar interactions between any of the sites were observed, consistent with the monomeric form in solution. Addition of Ca^{2+}, which triggers membrane binding, resulted in the broadening of the spectra within the trimer but not between the trimers, proving that a trimer and not a hexamer was forming the pore.[107]

6 CONCLUSION

ESR of protein is currently enjoying a renaissance of sorts. In addition to its contributions in studies of protein dynamics and orientation, ESR is being increasingly used as a structural technique. The advances of molecular biology facilitate targeting of chosen domains or scanning of the whole structure with spin labels. Comparison of dynamics, accessibility and distances at consecutive positions along the polypeptide chain is used in the determination of the secondary, tertiary and quaternary structure of proteins, which is of enormous importance in the post-genomic era. It is likely that the ease of determination of relative orientation of known domain motifs will make ESR a method of choice in high-throughput structural biology.

Technical advances in ESR, which include new probes, FTESR, higher magnetic fields, increase in absolute sensitivity, spectral dispersion and diversity of applications bode well for the continued development of ESR spectroscopy. Lastly, the development of powerful computational simulations makes ESR user-friendly and increases the number of ESR practitioners outside the die-hard community of spectroscopists.

ACKNOWLEDGMENTS

This work was sponsored by the National Science Foundation (NSF-IBN-9808708), NHMFL (in-house grant) and the American Heart Association (GIA-995024N).

LIST OF SYMBOLS

A	hyperfine interactions
A	peak-to-peak amplitude
$\boldsymbol{A}$	hyperfine interaction tensor
$A_{\max}, A'_{zz}$	maximum hyperfine splitting
a_0, A_0	isotropic hyperfine splitting
c	microwave field conversion factor
C'/C	central manifold line-height ratio of the V'_2 spectrum
D	diffusion tensor
$D_\perp, D_\parallel$	elements of diffusion tensor for motion parallel and perpendicular to the z-axis of a nitroxide
E	energy
f	resonant frequency of a spin
$f_\parallel, f_\perp$	frictional coefficients
$\boldsymbol{g}$	Zeeman interaction tensor
g_{eff}	effective g-value
H, H_0	magnetic field strength
$h, \hbar$	Planck's constant
H_1	microwave field
H_c	center field of a spectrum
H_m	modulation amplitude
H_{res}	resonant field
$I_0(H_{\text{res}})$	amplitude of the resonance line
J	coupling strength
L	directional cosine matrices
L, C, H	turning points of the spectra, $\theta = 0^\circ$

L''/L; H''/H	line-height ratios of STESR line shape
M_2	second moment of the central resonance
m_I	nuclear spin quantum number, nuclear manifold
P	microwave power
P'/P	line-height ratios of V_2' STESR spectra
$P_{1/2}$	half-saturation power
Q	resonator quality factor
r_{dd}	interspin distance
S	electron spin
S	order parameter
T_1	spin–lattice relaxation time
T_1^{eff}	effective spin–lattice relaxation time
T_2	spin–spin relaxation time
T_2^{eff}	effective spin–spin relaxation time
$V(m_I)$	peak-to-peak height of a given nuclear manifold resonance
V_0	absorption spectrum
V_1	first-derivative spectrum
V_2'	second-derivative, 90° out-of-phase display
W_x	rate of bimolecular collisions
$Y(H)$	electron spin resonance spectrum
β	Bohr magneton
ΔH_m	line width at half-height of a given nuclear manifold resonance line
ΔH_{pp}	peak-to-peak resonance line width
$\Delta\theta$	width of Gaussian angular distribution
Δw_{AB}^0	difference in resonant frequencies between spins A and B
η	resonator filling factor
η	viscosity
γ	magnetogyric ratio
Γ	width (at half-height) of the resonance
μ	magnetic moment of an electron
ν	Microwave frequency
w	Larmor frequency
Ω	orientational distribution
π	dimensionless accessibility parameter to relaxants
π	normalized solvent accessibility to various quenchers
Φ	differential accessibility to polar and nonpolar relaxants
$\rho(\theta)$	probability of the spins being orientated at angle θ with respect to the magnetic field
τ^{eff}	effective rotational correlation time
$\tau^{eff}(m_I)$	effective correlation time obtained from the STESR calibration curves of P'/P ratios
τ_{ex}	exchange time
$\tau_\perp$	correlation time for rotation about axis perpendicular to nitroxide z-axis
τ_r	rotational correlation time
τ_r^{iso}	isotropic rotational correlation time
$\tau_\parallel$	correlation time for rotation about the nitroxide z-axis
θ_c	cone angle
θ_0	center of Gaussian angular distribution
θ, ϕ	axial and azimuthal polar angles

ABBREVIATIONS AND ACRONYMS

ADP	Adenosine Diphosphate
AFC	Automatic Frequency Control
AMPPCP	Adenosine Methylenetriphosphate
AMPPNP	Adenosine Imidotriphosphate
ATP	Adenosine Triphosphate
ATPase	Adenosine Triphosphatase
ATPγS	Adenosine Thiotriphosphate
CROX	Potassium Tris(oxalatochromate)
DEER	Double Electron–Electron Resonance
DPPH	Diphenylpicrylhydrazyl
DR	Dielectric Resonator
ELDOR	Electron–Electron Double Resonance
ENDOR	Electron–Nuclear Double Resonance
ESR	Electron Spin Resonance
FID	Free Induction Decay
FTESR	Fourier Transform Electron Spin Resonance
GAPDH	Glyceraldehyde Dehydrogenase
HSE	Heisenberg Spin Exchange
LGR	Loop Gap Resonator
NAD	Nicotinamide Adenine Dinucleotide
NiAA	Nickel(II) Acetylacetonate
NiEDDA	Nickel(II) Ethylenediaminediacetate
NMR	Nuclear Magnetic Resonance
PADS	Peroxylamine Disulfonate
SDSL	Site-directed Spin Labeling
SECSY	Spin-echo Correlation Spectroscopy
SEESR	Spin-echo Electron Spin Resonance
SLE	Stochastic Liouville Equation
STESR	Saturation Transfer Electron Spin Resonance
TnC	Troponin C

TnI	Troponin I
TOAC	2,2,6,6-Tetramethylpiperidine-1-oxyl-4-amino-4-carboxylic Acid
UV	Ultraviolet
2-D	Two-dimensional

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Electron Spin Resonance Spectroscopy

REFERENCES

1. H. Thomann, M. Bernardo, 'Pulsed Electron Nuclear Multiple Resonance Spectroscopic Methods for Metalloproteins and Metalloenzymes', *Methods Enzymol.*, **227**, 118–189 (1993).
2. R. Cammack, C.E. Cooper, 'Electron Spin Resonance Spectroscopy of Iron Complexes and Iron-containing Proteins', *Methods Enzymol.*, **227**, 353–384 (1993).
3. C.J. Bender, P. Aisen, 'Continuous Wave Electron Nuclear Double Resonance Spectroscopy', *Methods Enzymol.*, **227**, 190–231 (1993).
4. G.W. Brudvig, 'Electron Spin Resonance Spectroscopy', *Methods Enzymol.*, **246**, 536–554 (1995).
5. V.J. DeRose, B.M. Hoffman, 'Protein Structure and Mechanism Studied by Electron Nuclear Double Resonance Spectroscopy', *Methods Enzymol.*, **246**, 554–589 (1995).
6. M.P. Klein, 'Perspectives on Magnetic Resonance and X-ray Absorption Spectroscopy in Biochemistry', *Methods Enzymol.*, **246**, 529–536 (1995).
7. E. Zavoisky, 'Spin-magnetic Resonance in Paramagnetics', *J. Phys. USSR*, **9**, 211 (1945).
8. C. Gorter, 'Paramagnetic Resonance', *Physica*, **3**, 503–513 (1936).
9. I. Rabi, S. Millman, P. Kush, J. Zacharias, 'The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments', *Phys. Rev.*, **55**, 526–535 (1939).
10. D. Ingram, J. Bennett, 'Paramagnetic Resonance in Phthalocyanine, Haemoglobin, and Other Organic Derivatives', *Discuss Faraday Soc.*, **19**, 140–146 (1955).
11. B. Commoner, J. Townsend, G. Pake, 'Free Radicals in Biological Materials', *Nature (London)*, **174**, 689–691 (1954).
12. D. Ingram, J. Tapley, R. Jackson, R. Bond, A. Murnaghan, 'Paramagnetic Resonance in Carbonaceous Solids', *Nature (London)*, **174**, 797–798 (1954).
13. W. Gordy, W.B. Ard, H. Shields, 'Microwave Spectroscopy of Biological Substances – I: Paramagnetic Resonance in X-irradiated Amino Acids and Proteins', *Proc. Natl. Acad. Sci. USA*, **41**, 983–1004 (1955).
14. H.M. Assenheim, *Introduction to Electron Spin Resonance*, Plenum Press, New York, 1967.
15. E.G. Rozantsev, in *Free Nitroxyl Radicals*, ed. H. Ulrich, Plenum Press, New York, 1970.
16. O. Lebedev, S. Kazarnovskii, *Tr. Khim. Khim. Tekhnol. (Gorkii)*, **8**, 649 (1959).
17. T.J. Stone, T. Buckman, P.L. Nordio, H.M. McConnell, 'Spin-labeled Biomolecules', *Proc. Natl. Acad. Sci. USA*, **54**, 1010–1017 (1965).
18. G. Feher, 'Method of Polarizing Nuclei in Paramagnetic Substances', *Phys. Rev.*, **103**, 500–501 (1956).
19. J.S. Hyde, J.C.W. Chien, J.H. Freed, 'Electron–Electron Double Resonance of Free Radicals in Solution', *J. Chem. Phys.*, **48**, 4211 (1968).
20. V. Benderskii, L. Blumenfeld, P. Stunzas, E. Sokolov, 'Double Electron–Electron Resonance of Triplet Excitons in Ion–Radicals Salts', *Nature (London)*, **220**, 365–367 (1968).
21. W.B. Mims, K. Nassau, J.D. McGee, 'Spectral Diffusion in Electron Resonance Lines', *Phys. Rev.*, **123**, 2059–2069 (1961).
22. U. Eliav, J.H. Freed, 'The Oscillatory Nature of Polarization Evolution in CIDEP', *J. Phys. Chem.*, **88**, 1277–1280 (1984).
23. O. Dobbert, T. Prisner, K.P. Dinse, 'Single-channel Quadrature FT ESR', *J. Magn. Reson.*, **70**, 173–175 (1986).
24. M.K. Bowman, 'Fourier-transform Electron Spin Resonance', *Bull. Am. Phys. Soc.*, **31**, 524 (1986).
25. J. Gorcester, J.H. Freed, 'Two-dimensional Fourier Transform ESR Spectroscopy', *J. Chem. Phys.*, **85**, 5375–5377 (1986).
26. O.Y. Grinberg, A.A. Dubinski, V.F. Shuvalov, L.G. Oranskii, V.I. Kurochkin, Y.S. Lebedev, 'Submilimeter EPR Spectroscopy of Free Radicals', *Dokl. Phys. Chem. (Engl. Transl.)*, **230**, 884–886 (1976).
27. W.B. Lynch, K.A. Earle, J.H. Freed, '1-mm Wave ESR Spectrometer', *Rev. Sci. Instrum.*, **59**, 1345–1351 (1985).
28. F. Muller, M.A. Hopkins, N. Coron, M. Grynberg, L.C. Brunel, G. Martinez, 'A High Magnetic Field EPR Spectrometer', *Rev. Sci. Instrum.*, **60**, 3681–3684 (1989).
29. G. Eaton, S. Eaton, K. Salikhov, *Foundations of Modern ESR*, World Scientific, Singapore, 1998.
30. C. Altenbach, S.L. Flitsch, H.G. Khorana, W.L. Hubbell, 'Structural Studies on Transmembrane Proteins. 2. Spin Labeling of Bacteriorhodopsin Mutants at Unique Cysteines', *Biochemistry*, **28**, 7806–7812 (1989).

31. H.S. McHaourab, M.A. Lietzow, K. Hideg, W.L. Hubbell, 'Motion of Spin-labeled Side Chains in T4 Lysozyme. Correlation with Protein Structure and Dynamics', *Biochemistry*, **35**, 7692–7704 (1996).
32. E. Perozo, D.M. Cortes, L.G. Cuello, 'Three-dimensional Architecture and Gating Mechanism of a K^+ Channel Studied by EPR Spectroscopy', *Nature Struct. Biol.*, **5**, 459–469 (1998).
33. B.W. Matthews, 'Studies on Protein Stability with T4 Lysozyme', *Adv. Protein Chem.*, **46**, 249–278 (1995).
34. D.D. Thomas, L.R. Dalton, J.S. Hyde, 'Rotational Diffusion Studied by Passage Saturation Transfer Electron Spin Resonance', *J. Chem. Phys.*, **65**, 3006–3024 (1976).
35. P.G. Fajer, D. Marsh, 'Microwave and Modulation Field Inhomogeneities and the Effect of Cavity *Q* in Saturation Transfer ESR Spectra. Dependence on Sample Size', *J. Magn. Reson.*, **49**, 212–224 (1982).
36. T.C. Squier, D.D. Thomas, 'Methodology for Increased Precision in Saturation Transfer Electron Spin Resonance Studies of Rotational Dynamics', *Biophys. J.*, **49**, 921–935 (1986).
37. E.J. Hustedt, C.E. Cobb, A.H. Beth, J.M. Beechem, 'Measurement of Rotational Dynamics by the Simultaneous Nonlinear Analysis of Optical and EPR Data', *Biophys. J.*, **64**, 614–621 (1993).
38. G.I. Likhtenshtein, *Biophysical Labeling Methods in Molecular Biology*, Cambridge University Press, Cambridge, 1993.
39. L. Berliner (ed.), 'Spin Labeling Theory and Applications', in *Biological Magnetic Resonance*, Plenum Press, New York, 1998.
40. L. Berliner, J. Reuben (eds.), *Biological Magnetic Resonance, Spin Labeling and the Next Millennium*, Plenum Press, New York, 1989.
41. W.L. Hubbell, C. Altenbach, 'Investigation of Structure and Dynamics in Membrane Proteins Using Site-directed Spin Labelling', *Curr. Opin. Struct. Biol.*, **4**, 566–573 (1994).
42. W.L. Hubbell, H.S. McHaourab, C. Altenbach, M.A. Lietzow, 'Watching Proteins Move Using Site-directed Spin Labeling', *Structure*, **4**, 779–783 (1996).
43. W.L. Hubbell, A. Gross, R. Langen, M.A. Lietzow, 'Recent Advances in Site-directed Spin Labeling of Proteins', *Curr. Opin. Struct. Biol.*, **8**, 649–656 (1998).
44. D. Marsh, L.I. Horvath, in *Advanced ESR: Applications in Biology and Biochemistry*, ed. A.J. Hoff, Elsevier, Amsterdam, 707–752, 1989.
45. G.L. Millhauser, W.R. Fiori, S.M. Miick, 'Electron Spin Labels', *Methods Enzymol.*, **246**, 589–610 (1995).
46. G.L. Millhauser, 'Selective Placement of Electron Spin Resonance Spin Labels: New Structural Methods for Peptides and Proteins', *Trends Biochem. Sci.*, **17**, 448–452 (1992).
47. E.J. Hustedt, A.H. Beth, 'Nitroxide Spin–Spin Interactions: Applications to Protein Structure and Dynamics', *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 129–153 (1999).
48. T.P. Burghardt, N.L. Thompson, 'Model-independent Electron Spin Resonance for Measuring Order of Immobile Components in a Biological Assembly', *Biophys. J.*, **48**, 401–409 (1985).
49. T.P. Burghardt, A.R. French, 'Reconstruction of the Probe Angular Distribution from a Series of Electron Spin Resonance Spectra of Tilted Oriented Samples', *Biophys. J.*, **56**, 525–534 (1989).
50. D.E. Budil, S. Lee, S. Saxena, J.H. Freed, 'Nonlinear-least-squares Analysis of Slow-motion EPR Spectra in One and Two Dimensions Using a Modified Levenberg–Marquardt Algorithm', *J. Magn. Reson.*, **120**, 155–189 (1996).
51. P.G. Fajer, B. Bennett, C. Polnaszek, E.A. Fajer, D. Thomas, 'General Method for Multiparameter Fitting of High-resolution EPR Spectra Using a Simplex Algorithm', *J. Magn. Reson.*, **88**, 111–125 (1990).
52. P. Jost, O.K. Griffith, in *Spin Labeling Theory and Applications*, ed. L.J. Berliner, Academic Press, New York, 453–523, 1976.
53. P.G. Fajer, 'Determination of Spin-label Orientation Within the Myosin Head', *Proc. Natl. Acad. Sci. USA*, **91**, 937–941 (1994).
54. J. Seelig, in *Spin Labeling Theory and Applications*, ed. L.J. Berliner, Academic Press, New York, 373–409, 1976.
55. E.J. Hustedt, A.H. Beth, 'Determination of the Orientation of a Band 3 Affinity Spin-label Relative to the Membrane Normal Axis of the Human Erythrocyte', *Biochemistry*, **35**, 6944–6954 (1996).
56. D.D. Thomas, R. Cooke, 'Orientation of Spin-labeled Myosin Heads in Glycerinated Muscle Fibers', *Biophys. J.*, **32**, 891–906 (1980).
57. J.E. Baker, I. Brust-Mascher, S. Ramachandran, L.E. LaConte, D.D. Thomas, 'A Large and Distinct Rotation of the Myosin Light Chain Domain Occurs Upon Muscle Contraction', *Proc. Natl. Acad. Sci. USA*, **95**, 2944–2949 (1998).
58. H.C. Li, P.G. Fajer, 'Orientation Changes of Troponin C Associated with Thin Filament Activation', *Biochemistry*, **33**, 14 324–14 332 (1994).
59. A. Redfield, 'The Theory of Relaxation Processes', *Adv. Magn. Reson.*, **1**, 1 (1965).
60. D.J. Schneider, J.H. Freed, in *Spin Labeling Theory and Applications*, eds. L.J. Berliner, J. Reuben, Plenum Press, New York, 1–76, 1989.
61. D. Marsh, in *Spin Labeling Theory and Applications*, eds. L.J. Berliner, J. Reuben, Plenum Press, New York, 255–303, 1989.
62. H.J. Steinhoff, W.L. Hubbell, 'Calculation of Electron Spin Resonance Spectra from Brownian Dynamics Trajectories: Application to Nitroxide Side Chains in Proteins', *Biophys. J.*, **71**, 2201–2212 (1996).

63. D. Marsh, L. Horvath, 'A Simple Theoretical Treatment of the Sensitivity of Saturation-transfer EPR Spectra to Slow Rotational Diffusion', *J. Magn. Reson.*, **99**, 323–331 (1992).
64. S. Koening, 'Brownian Motion of an Ellipsoid. A Correction to Perrin's Result', *Biopolymers*, **14**, 2421–2423 (1975).
65. S. Broersma, 'Rotational Diffusion Constant of a Cylindrical Particle', *J. Chem. Phys.*, **32**, 1626–1631 (1960).
66. M.M. Tirado, J.G.d.I. Torre, 'Translational Friction Coefficients of Rigid Symmetric Top Macromolecules. Application to Circular Cylinders', *J. Chem. Phys.*, **71**, 2581–2587 (1979).
67. A. Beth, B.H. Robinson, in *Spin Labeling Theory and Applications*, eds. L.J. Berliner, J. Reuben, Plenum Press, New York, 179–253, 1989.
68. B.H. Robinson, L.R. Dalton, 'Anisotropic Rotational Diffusion Studied by Passage Saturation Transfer Electron Spin Resonance', *J. Chem. Phys.*, **72**, 1312–1324 (1980).
69. E.J. Hustedt, A.H. Beth, 'Analysis of Saturation Transfer Electron Spin Resonance Spectra of a Spin-labeled Integral Membrane Protein, Band 3, in Terms of the Uniaxial Rotational Diffusion Model', *Biophys. J.*, **69**, 1409–1423 (1995).
70. P.G. Fajer, D. Marsh, 'Sensitivity of Saturation Transfer ESR Spectra to Anisotropic Rotation. Application to Membrane Systems', *J. Magn. Reson.*, **51**, 446–459 (1983).
71. E.C. Howard, K.M. Lindahl, C.F. Polnaszek, D.D. Thomas, 'Simulation of Saturation Transfer Electron Spin Resonance Spectra for Rotational Motion with Restricted Angular Amplitude', *Biophys. J.*, **64**, 581–593 (1993).
72. D.D. Thomas, J.C. Seidel, J.S. Hyde, J. Gergely, 'Motion of Subfragment-1 in Myosin and Its Supramolecular Complexes: Saturation Transfer Electron Spin Resonance', *Proc. Natl. Acad. Sci. USA*, **72**, 1729–1733 (1975).
73. D.D. Thomas, S. Ishiwata, J.C. Seidel, J. Gergely, 'Submillisecond Rotational Dynamics of Spin-labeled Myosin Heads in Myofibrils', *Biophys. J.*, **32**, 873–889 (1980).
74. V.A. Barnett, D.D. Thomas, 'Saturation Transfer Electron Spin Resonance of Spin-labeled Muscle Fibers. Dependence of Myosin Head Rotational Motion on Sarcomere Length', *J. Mol. Biol.*, **179**, 83–102 (1984).
75. P.G. Fajer, E.A. Fajer, M. Schoenberg, D.D. Thomas, 'Orientational Disorder and Motion of Weakly Attached Cross-Bridges', *Biophys. J.*, **60**, 642–649 (1991).
76. D. Raucher, P.G. Fajer, 'Orientation and Dynamics of Myosin Heads in Aluminum Fluoride Induced Prepower Stroke States: an EPR Study', *Biochemistry*, **33**, 11 993–11 999 (1994).
77. D. Raucher, C.P. Sar, K. Hideg, P.G. Fajer, 'Myosin Catalytic Domain Flexibility in MgADP', *Biochemistry*, **33**, 14 317–14 323 (1994).
78. B. Adhikari, K. Hideg, P.G. Fajer, 'Independent Mobility of Catalytic and Regulatory Domains of Myosin Heads', *Proc. Natl. Acad. Sci. USA*, **94**, 9643–9647 (1997).
79. E.M. Ostap, D.D. Thomas, 'Rotational Dynamics of Spin-labeled F-actin During Activation of Myosin S1 ATPase Using Caged ATP', *Biophys. J.*, **59**, 1235–1241 (1991).
80. E.M. Ostap, T. Yanagida, D.D. Thomas, 'Orientation Distribution of Spin-labeled Actin Oriented by Flow', *Biophys. J.*, **63**, 966–975 (1992).
81. N. Naber, M. Lorenz, R. Cooke, 'The Orientation of Spin-probes Attached to Cys374 on Actin in Oriented Gels', *J. Mol. Biol.*, **236**, 703–709 (1994).
82. M. Mossakowska, J. Belagyi, H. Strzelecka-Golaszewska, 'An EPR Study of the Rotational Dynamics of Actins from Striated and Smooth Muscle and Their Complexes with Heavy Meromyosin', *Eur. J. Biochem.*, **175**, 557–564 (1988).
83. N. Naber, E.M. Ostap, D.D. Thomas, R. Cooke, 'Orientation and Rotational Dynamics of Spin-labeled Phalloidin Bound to Actin in Muscle Fibers', *Proteins*, **17**, 347–354 (1993).
84. H.C. Li, K. Hideg, P.G. Fajer, 'The Mobility of Troponin C and Troponin I in Muscle', *J. Mol. Recognit.*, **10**, 194–201 (1997).
85. H.C. Li, P.G. Fajer, 'Structural Coupling of Troponin C and Actomyosin in Muscle Fibers', *Biochemistry*, **37**, 6628–6635 (1998).
86. H.L. Sweeney, B.F. Bowman, J.T. Stull, 'Myosin Light Chain Phosphorylation in Vertebrate Striated Muscle: Regulation and Function', *Am. J. Physiol.*, **264**, C1085–C1095 (1993).
87. B.B. Adhikari, J. Somerset, J.T. Stull, P.G. Fajer, 'Dynamic Modulation of the Regulatory Domain of Myosin Heads by pH, Ionic Strength, and RLC Phosphorylation in Synthetic Myosin Filaments', *Biochemistry*, **38**, 3127–3132 (1999).
88. N.J. Ryba, D. Marsh, 'Protein Rotational Diffusion and Lipid/Protein Interactions in Recombinants of Bovine Rhodopsin with Saturated Diacylphosphatidylcholines of Different Chain Lengths Studied by Conventional and Saturation-transfer Electron Spin Resonance', *Biochemistry*, **31**, 7511–7518 (1992).
89. P. Fajer, P.F. Knowles, D. Marsh, 'Rotational Motion of Yeast Cytochrome Oxidase in Phosphatidylcholine Complexes Studied by Saturation-transfer Electron Spin Resonance', *Biochemistry*, **28**, 5634–5643 (1989).
90. L.I. Horvath, A. Munding, K. Beyer, M. Klingenberg, D. Marsh, 'Rotational Diffusion of Mitochondrial ADP/ATP Carrier Studied by Saturation-transfer Electron Spin Resonance', *Biochemistry*, **28**, 407–414 (1989).
91. M. Esmann, H.O. Hankovszky, K. Hideg, D. Marsh, 'A Novel Spin-label for Study of Membrane Protein Rotational Diffusion Using Saturation Transfer Electron

Spin Resonance. Application to Selectively Labeled Class I and Class II-SH Groups of the Shark Rectal Gland Na^+/K^+-ATPase', *Biochim. Biophys. Acta*, **978**, 209–215 (1989).
92. T.C. Squier, S.E. Hughes, D.D. Thomas, 'Rotational Dynamics and Protein–Protein Interactions in the Ca-ATPase Mechanism', *J. Biol. Chem.*, **263**, 9162–9170 (1988).
93. J.E. Mahaney, D.D. Thomas, 'Effects of Melittin on Molecular Dynamics and Ca-ATPase Activity in Sarcoplasmic Reticulum Membranes: Electron Spin Resonance', *Biochemistry*, **30**, 7171–7180 (1991).
94. D.A. Ferrington, T.E. Jones, Z. Qin, M. Miller-Schlyer, T.C. Squier, D.J. Bigelow, 'Decreased Conformational Stability of the Sarcoplasmic Reticulum Ca-ATPase in Aged Skeletal Muscle', *Biochim. Biophys. Acta*, **1330**, 233–247 (1997).
95. R.L. Cornea, L.R. Jones, J.M. Autry, D.D. Thomas, 'Mutation and Phosphorylation Change the Oligomeric Structure of Phospholamban in Lipid Bilayers', *Biochemistry*, **36**, 2960–2967 (1997).
96. S. Negash, L.T. Chen, D.J. Bigelow, T.C. Squier, 'Phosphorylation of Phospholamban by cAMP-dependent Protein Kinase Enhances Interactions Between Ca-ATPase Polypeptide Chains in Cardiac Sarcoplasmic Reticulum Membranes', *Biochemistry*, **35**, 11 247–11 259 (1996).
97. D. Thomas, L. Reddy, C. Karim, M. Li, R. Cornea, J. Autry, L. Jones, J. Stamm, 'Direct Spectroscopic Detection of Molecular Dynamics and Interactions of the Calcium Pump and Phospholamban', *Ann. NY Acad. Sci.*, **853**, 186–194 (1998).
98. D. Schwarz, A. Chernogolov, P. Kisselev, 'Complex Formation in Vesicle-reconstituted Mitochondrial Cytochrome P450 Systems (CYP11A1 and CYP11B1) as Evidenced by Rotational Diffusion Experiments Using ESR and STESR', *Biochemistry*, **38**, 9456–9464 (1999).
99. B.H. Robinson, D.A. Haas, C. Mailer, 'Molecular Dynamics in Liquids: Spin–Lattice Relaxation of Nitroxide Spin Labels', *Science*, **263**, 490–493 (1994).
100. S. Saxena, J.H. Freed, 'Absorption Line shapes in Two-dimensional Electron Spin Resonance and the Effects of Slow Motions in Complex Fluids', *J. Magn. Reson.*, **124**, 439–454 (1997).
101. S. Saxena, J.H. Freed, 'Theory of Double Quantum Two-dimensional Electron Spin Resonance with Application to Distance Measurements', *J. Chem. Phys.*, **107**, 1317–1340 (1997).
102. D.T. Thomas, E.M. Ostap, C.L. Berger, P.G. Fajer, J.E. Mahaney, in *Biological Magnetic Resonance – EMR of Paramagnetic Molecules*, eds. L. Berliner, J. Reuben, Plenum Press, New York, 323–351, 1993.
103. A. Sienkiewicz, K. Qu, C. Scholes, 'Dielectric Resonator-based Stopped-flow ESR', *Rev. Sci. Instrum.*, **65**, 68–74 (1994).
104. K. Qu, J.L. Vaughn, A. Sienkiewicz, C.P. Scholes, J.S. Fetrow, 'Kinetics and Motional Dynamics of Spin-labeled Yeast Iso-1-cytochrome *c*: 1. Stopped-flow Electron Spin Resonance as a Probe for Protein Folding/Unfolding of the C-terminal Helix Spin-labeled at Cysteine 102', *Biochemistry*, **36**, 2884–2897 (1997).
105. A. Sienkiewicz, A.M. da Costa Ferreira, B. Danner, C.P. Scholes, 'Dielectric Resonator-based Flow and Stopped-flow ESR with Rapid Field Scanning: a Methodology for Increasing Kinetic Information', *J. Magn. Reson.*, **136**, 137–142 (1999).
106. Y.K. Shin, C. Levinthal, F. Levinthal, W.L. Hubbell, 'Colicin E1 Binding to Membranes: Time-resolved Studies of Spin-labeled Mutants', *Science*, **259**, 960–963 (1993).
107. R. Langen, J.M. Isas, H. Luecke, H.T. Haigler, W.L. Hubbell, 'Membrane-mediated Assembly of Annexins Studied by Site-directed Spin Labeling', *J. Biol. Chem.*, **273**, 22 453–22 457 (1998).
108. E.M. Ostap, H.D. White, D.D. Thomas, 'Transient Detection of Spin-labeled Myosin Subfragment 1 Conformational States During ATP Hydrolysis', *Biochemistry*, **32**, 6712–6720 (1993).
109. S.M. Lewis, D.D. Thomas, 'Resolved Conformational States of Spin-labeled Ca-ATPase During the Enzymatic Cycle', *Biochemistry*, **31**, 7381–7389 (1992).
110. S.M. Lewis, D.D. Thomas, 'Microsecond Rotational Dynamics of Spin-labeled Ca-ATPase During Enzymatic Cycling Initiated by Photolysis of Caged ATP', *Biochemistry*, **30**, 8331–8339 (1991).
111. Z.T. Farahbakhsh, K. Hideg, W.L. Hubbell, 'Photoactivated Conformational Changes in Rhodopsin: a Time-resolved Spin Label Study', *Science*, **262**, 1416–1419 (1993).
112. H.J. Steinhoff, R. Mollaaghababa, C. Altenbach, K. Hideg, M. Krebs, H.G. Khorana, W.L. Hubbell, 'Time-resolved Detection of Structural Changes During the Photocycle of Spin-labeled Bacteriorhodopsin', *Science*, **266**, 105–107 (1994).
113. T. Rink, J. Riesle, D. Oesterhelt, K. Gerwert, H.J. Steinhoff, 'Spin-labeling Studies of the Conformational Changes in the Vicinity of D36, D38, T46, and E161 of Bacteriorhodopsin During the Photocycle', *Biophys. J.*, **73**, 983–993 (1997).
114. M. Lindgren, M. Svensson, P.O. Freskgard, U. Carlsson, P. Jonasson, L.G. Martensson, B.H. Jonsson, 'Characterization of a Folding Intermediate of Human Carbonic Anhydrase II: Probing Local Mobility by Electron Spin Resonance', *Biophys. J.*, **69**, 202–213 (1995).
115. C.S. Klug, W. Su, J. Liu, P.E. Klebba, J.B. Feix, 'Denaturant Unfolding of the Ferric Enterobactin Receptor and Ligand-induced Stabilization Studied by Site-directed Spin Labeling', *Biochemistry*, **34**, 14 230–14 236 (1995).
116. C.S. Klug, J.B. Feix, 'Guanidine Hydrochloride Unfolding of a Transmembrane Beta-strand in FepA Using

Site-directed Spin Labeling', *Protein Sci.*, **7**, 1469–1476 (1998).
117. M. Svensson, P. Jonasson, P.O. Freskgard, B.H. Jonsson, M. Lindgren, L.G. Martensson, M. Gentile, K. Boren, U. Carlsson, 'Mapping the Folding Intermediate of Human Carbonic Anhydrase II. Probing Substructure by Chemical Reactivity and Spin and Fluorescence Labeling of Engineered Cysteine Residues', *Biochemistry*, **34**, 8606–8620 (1995).
118. M. Persson, P. Hammarstrom, M. Lindgren, B.H. Jonsson, M. Svensson, U. Carlsson, 'EPR Mapping of Interactions Between Spin-labeled Variants of Human Carbonic Anhydrase II and GroEL: Evidence for Increased Flexibility of the Hydrophobic Core by the Interaction', *Biochemistry*, **38**, 432–441 (1999).
119. Z.T. Farahbakhsh, Q.L. Huang, L.L. Ding, C. Altenbach, H.J. Steinhoff, J. Horwitz, W.L. Hubbell, 'Interaction of Alpha-crystallin with Spin-labeled Peptides', *Biochemistry*, **34**, 509–516 (1995).
120. P.G. Fajer, E.A. Fajer, N.J. Brunsvold, D.D. Thomas, 'Effects of AMPPNP on the Orientation and Rotational Dynamics of Spin-labeled Muscle Cross-bridges', *Biophys. J.*, **53**, 513–524 (1988).
121. M. Zhao, K.C. Zen, W.L. Hubbell, H.R. Kaback, 'Proximity Between Glu126 and Arg144 in the Lactose Permease of *Escherichia coli*', *Biochemistry*, **38**, 7407–7412 (1999).
122. W.R. Fiori, G.L. Millhauser, 'Exploring the Peptide 3_{10}-helix Reversible Alpha-helix Equilibrium with Double Label Electron Spin Resonance', *Biopolymers*, **37**, 243–250 (1995).
123. E.J. Hustedt, A.I. Smirnov, C.F. Laub, C.E. Cobb, A.H. Beth, 'Molecular Distances from Dipolar Coupled Spin-labels: the Global Analysis of Multifrequency Continuous Wave Electron Spin Resonance Data', *Biophys. J.*, **72**, 1861–1877 (1997).
124. P. Hanson, D. Anderson, G. Martinez, G. Millhauser, F. Formaggio, M. Crisma, C. Toniolo, C. Vita, 'Electron Spin Resonance and Structural Analysis of Water Soluble Alanine-rich Peptides Incorporating TOAC', *Mol. Phys.*, **95**, 957–966 (1998).
125. M.D. Rabenstein, Y.K. Shin, 'Determination of the Distance Between Two Spin Labels Attached to a Macromolecule', *Proc. Natl. Acad. Sci. USA*, **92**, 8239–8243 (1995).
126. H.J. Steinhoff, N. Radzwil, W. Thevis, V. Lenz, D. Brandenburg, A. Antson, G. Dodson, A. Wollmer, 'Determination of Interspin Distances Between Spin Labels Attached to Insulin: Comparison of Electron Spin Resonance Data with the X-ray Structure', *Biophys. J.*, **73**, 3287–3298 (1997).
127. J. Wu, J. Voss, W.L. Hubbell, H.R. Kaback, 'Site-directed Spin Labeling and Chemical Crosslinking Demonstrate that Helix V is Close to Helices VII and VIII in the Lactose Permease of *Escherichia coli*', *Proc. Natl. Acad. Sci. USA*, **93**, 10 123–10 127 (1996).
128. E. Perozo, D.M. Cortes, L.G. Cuello, 'Structural Rearrangements Underlying K^+-channel Activation Gating', *Science*, **285**, 73–78 (1999).
129. H.S. McHaourab, K.J. Oh, C.J. Fang, W.L. Hubbell, 'Conformational of T4 Lysozyme in Solution. Hinge-bending Motion and the Substrate-induced Conformational Transition Studied by Site-directed Spin Labeling', *Biochemistry*, **36**, 307–316 (1997).
130. K. Cai, R. Langen, W.L. Hubbell, H.G. Khorana, 'Structure and Function in Rhodopsin: Topology of the C-terminal Polypeptide Chain in Relation to the Cytoplasmic Loops', *Proc. Natl. Acad. Sci. USA*, **94**, 14 267–14 272 (1997).
131. D.J. Singel, R.G. Larsen, 'Double Electron–Electron Resonance Spin-echo Modulation: Spectroscopic Measurement of Electron Spin Pair Separations in Orientationally Disordered Solids', *J. Chem. Phys.*, **98**, 5134–5146 (1993).
132. V. Pfannebecker, H. Klos, M. Hubrich, T. Volkmer, A. Heuer, U. Wiesner, H.W. Spiess, 'Determination of End-to-end Distance in Oligomers by Pulsed EPR', *J. Phys. Chem.*, **100**, 13 428–13 432 (1996).
133. A. Raitsimring, J. Peisach, H.C. Lee, 'Measurement of Distance Distribution Between Spin Labels in Spin-labeled Hemoglobin Using an Electron Spin Echo Method', *J. Phys. Chem.*, **96**, 3526–3531 (1992).
134. V. Budker, J.L. Du, M. Seiter, G.R. Eaton, S.S. Eaton, 'Electron–Electron Spin–Spin Interaction in Spin-labeled Low-spin Methemoglobin', *Biophys. J.*, **68**, 2531–2542 (1995).
135. J. Voss, W.L. Hubbell, H.R. Kaback, 'Distance Determination in Proteins Using Designed Metal Ion Binding Sites and Site-directed Spin Labeling: Application to the Lactose Permease of *Escherichia coli*', *Proc. Natl. Acad. Sci. USA*, **92**, 12 300–12 303 (1995).
136. J. Voss, L. Salwinski, H.R. Kaback, W.L. Hubbell, 'A Method for Distance Determination in Proteins Using a Designed Metal Ion Binding Site and Site-directed Spin Labeling: Evaluation with T4 Lysozyme', *Proc. Natl. Acad. Sci. USA*, **92**, 12 295–12 299 (1995).
137. J.S. Leigh, 'ESR Rigid-lattice Line Shape in a System of Two Interacting Spins', *J. Chem. Phys.*, **52**, 2608–2612 (1970).
138. J. Voss, W.L. Hubbell, H.R. Kaback, 'Helix Packing in the Lactose Permease Determined by Metal–Nitroxide Interaction', *Biochemistry*, **37**, 211–216 (1998).
139. H.R. Kaback, J. Voss, J. Wu, 'Helix Packing in Polytopic Membrane Proteins: the Lactose Permease of *Escherichia coli*', *Curr. Opin. Struct. Biol.*, **7**, 537–542 (1997).
140. C. Altenbach, D.A. Greenhalgh, H.G. Khorana, W.L. Hubbell, 'A Collision Gradient Method to Determine the Immersion Depth of Nitroxides in Lipid Bilayers: Application to Spin-labeled Mutants of bacteriorhodopsin', *Proc. Natl. Acad. Sci. USA*, **91**, 1667–1671 (1994).

141. Z.T. Farahbakhsh, C. Altenbach, W.L. Hubbell, 'Spin Labeled Cysteines as Sensors for Protein–Lipid Interaction and Conformation in Rhodopsin', *Photochem. Photobiol.*, **56**, 1019–1033 (1992).
142. J.B. Feix, C.S. Klug, in *Biological Magnetic Resonance: Spin Labeling the Next Millenium*, ed. L.J. Berliner, Plenum Press, New York, 251–281, 1998.
143. C.S. Klug, W. Su, J.B. Feix, 'Mapping of the Residues Involved in a Proposed Beta-strand Located in the Ferric Enterobactin Receptor FepA Using Site-directed Spin-labeling', *Biochemistry*, **36**, 13 027–13 033 (1997).
144. L. Salwinski, W.L. Hubbell, 'Structure in the Channel Forming Domain of Colicin E1 Bound to Membranes: the 402–424 Sequence', *Protein Sci.*, **8**, 562–572 (1999).
145. K.J. Oh, H. Zhan, C. Cui, K. Hideg, R.J. Collier, W.L. Hubbell, 'Organized of Diphtheria Toxin T Domain in Bilayers: a Site-directed Spin Labeling Study', *Science*, **273**, 810–812 (1996).
146. C. Altenbach, T. Marti, H.G. Khorana, W.L. Hubbell, 'Transmembrane Protein Structure: Spin Labeling of Bacteriorhodopsin Mutants', *Science*, **248**, 1088–1092 (1990).
147. C. Altenbach, K. Yang, D.L. Farrens, Z.T. Farahbakhsh, H.G. Khorana, W.L. Hubbell, 'Structural Features and Light-dependent Changes in the Cytoplasmic Interhelical E–F Loop Region of Rhodopsin: a Site-directed Spin-labeling Study', *Biochemistry*, **35**, 12 470–12 478 (1996).
148. C. Altenbach, K. Cai, H.G. Khorana, W.L. Hubbell, 'Structural Features and Light-dependent Changes in the Sequence 306–322 Extending from Helix VII to the Palmitoylation Sites in Rhodopsin: a Site-directed Spin-labeling Study', *Biochemistry*, **38**, 7931–7937 (1999).
149. C. Altenbach, J. Klein-Seetharaman, J. Hwa, H.G. Khorana, W.L. Hubbell, 'Structural Features and Light-dependent Changes in the Sequence 59–75 Connecting Helices I and II in Rhodopsin: a Site-directed Spin-labeling Study', *Biochemistry*, **38**, 7945–7949 (1999).
150. H.S. McHaourab, A.R. Berengian, H.A. Koteiche, 'Site-directed Spin-labeling Study of the Structure and Subunit Interactions Along a Conserved Sequence in the Alpha-crystallin Domain of Heat-shock Protein 27. Evidence of a conserved Subunit Interface', *Biochemistry*, **36**, 14 627–14 634 (1997).
151. A.R. Berengian, M.P. Bova, H.S. McHaourab, 'Structure and Function of the Conserved Domain in AlphaA-crystallin. Site-directed Spin Labeling Identifies a Beta-strand Located Near a Subunit Interface', *Biochemistry*, **36**, 9951–9957 (1997).
152. R. Langen, J.M. Isas, W.L. Hubbell, H.T. Haigler, 'A Transmembrane Form of Annexin XII Detected by Site-directed Spin Labeling', *Proc. Natl. Acad. Sci. USA*, **95**, 14 060–14 065 (1998).
153. S.J. Anthony-Cahill, P.A. Benfield, R. Fairman, Z.R. Wasserman, S.L. Brenner, W.F.D. Stafford, C. Altenbach, W.L. Hubbell, W.F. DeGrado, 'Molecular Characterization of Helix–Loop–Helix Peptides', *Science*, **255**, 979–983 (1992).
154. D.L. Farrens, C. Altenbach, K. Yang, W.L. Hubbell, H.G. Khorana, 'Requirement of Rigid-body Motion of Transmembrane Helices for Light Activation of Rhodopsin', *Science*, **274**, 768–770 (1996).
155. K. Yang, D.L. Farrens, C. Altenbach, Z.T. Farahbakhsh, W.L. Hubbell, H.G. Khorana, 'Structure and Function in Rhodopsin. Cysteines 65 and 316 are in Proximity in a Rhodopsin Mutant as Indicated by Disulfide Formation and Interactions Between Attached Spin Labels', *Biochemistry*, **35**, 14 040–14 046 (1996).
156. D.A. Doyle, J. Morais Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, 'The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity', *Science*, **280**, 69–77 (1998).
157. H.S. McHaourab, J.S. Hyde, J.B. Feix, 'Aggregation State of Spin-labeled Cecropin AD in Solution', *Biochemistry*, **32**, 11 895–11 902 (1993).
158. K.M. Lundberg, C.J. Stenland, F.E. Cohen, S.B. Prusiner, G.L. Millhauser, 'Kinetics and Mechanism of Amyloid Formation by the Prion Protein H1 Peptide as Determined by Time-dependent ESR', *Chem. Biol.*, **4**, 345–355 (1997).
159. J. Sun, J. Voss, W.L. Hubbell, H.R. Kaback, 'Proximity Between Periplasmic Loops in the Lactose Permease of *Escherichia coli* as Determined by Site-directed Spin labeling', *Biochemistry*, **38**, 3100–3105 (1999).

Fluorescence Spectroscopy in Peptide and Protein Analysis

Alexey S. Ladokhin*
University of California, Irvine, USA

1 Introduction **1**
1.1 What is Measured in a Protein Fluorescence Experiment? 2
2 Protein Fluorescence **3**
2.1 Spectral Properties of Protein Fluorophores 3
2.2 Spectral Classes of Tryptophan Residues 5
2.3 Protein Dynamics and Heterogeneity of Fluorescence Decay 7
3 Fluorescence Studies of Protein Conformation and Interactions **8**
3.1 Protein Folding 8
3.2 Protein–Ligand Interactions 10
3.3 Membrane Proteins and Peptides 11
Acknowledgments **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **15**

Fluorescence spectroscopy and its multiple applications to the life sciences have undergone rapid development. This is due to numerous technical advances in both instrumentation and methods of data analysis as well as to a vast proliferation of basic techniques. Applications of fluorescence spectroscopy to protein and peptide analysis are governed by three principal factors: the dynamic nature of the signal, its localized nature, and its redundancy. Although these features can complicate interpretation of the experimental result, they also can be exploited to obtain unique structural and dynamic information. The availability and simplicity of basic data acquisition and analysis are important practical features behind the popularity of fluorescence as compared to other spectroscopic techniques. Yet this simplicity does not appear to compromise its advantages. This article is intended first to provide an overview of the fluorescence phenomenon in proteins, and second to illustrate applications of fluorescence spectroscopy in advanced (but not necessarily high-tech) studies. Three areas of protein studies, namely protein–ligand interactions, protein folding and studies of membrane proteins, have been chosen to demonstrate key advantages of fluorescence spectroscopy: it is sensitive, versatile, and it lends itself readily to fast data acquisition.

1 INTRODUCTION

Fluorescence spectroscopy and its multiple applications to protein analysis, and to the life sciences in general, have undergone rapid development during the past decade. This progress appears to be driven on two levels. First, numerous technical advances in time resolution, methods of data analysis, and improved instrumentation have enabled researchers to probe the structural and dynamic features of proteins, membranes and nucleic acids, to acquire multidimensional (space–time) microscopic images of the distributions of various molecules in cell cultures, to follow conformational changes of single molecules, etc. These developments, pioneered by a handful of research groups, have spilled over into multiple areas as diverse as basic analytical chemistry and practical clinical applications.

The second, and sometimes under-appreciated, level of development of fluorescence spectroscopy in studies of biological macromolecules involves the proliferation of the basic technique. This progress is driven mainly by researchers whose primary scientific interests lie well outside the field of spectroscopy, and who, while admiring the elegance of the custom-designed multiphoton excitation experimental scheme, might not necessarily want to implement one in their own laboratory. Instead, they will use a commercially available fluorimeter, almost standard equipment these days in many laboratories involved in biochemical and biophysical studies. Nevertheless, their studies have produced not only numerous answers to a variety of important biological problems but have broadened the range of fluorescence techniques as well.

The availability and simplicity of basic data acquisition and analysis are important practical features behind the popularity of fluorescence as compared to other spectroscopic techniques. Yet this simplicity does not appear to compromise its main advantages, one of which, exceptional sensitivity, allows routine detection of fluorescent substances on a subnanomolar scale on the one hand, and provides the ability to collect data in a kinetic regime during fast submillisecond reactions on the other. Applications of fluorescence spectroscopy to protein and peptide analysis are governed by three principal factors: (1) the dynamic nature of the signal, (2) its localized nature, and (3) its redundancy. Fluorescence is a dynamic phenomenon and the

* On leave from: Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kiev 252143, Ukraine.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

lifetime of the excited state is sufficient for a variety of chemical and physical reactions to take place prior to emission. The usual nanosecond time-window of fluorescence is normally shorter than that of other dynamic techniques (e.g. nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR)). Fluorescence spectroscopy is essentially a probe technique sensing changes in the local environment of the fluorophore, which distinguishes it from generalized techniques, such as calorimetry, far-ultraviolet circular dichroism (CD), and infrared (IR) spectroscopy. Also, various possibilities of structural rearrangements in the environment of the fluorophore may lead to a similar fluorescence signal. None of these features taken alone is either beneficial or detrimental. Although they can complicate interpretation of the experimental result, they also can be exploited to obtain unique structural and dynamic information.

Fluorescence spectroscopy and its various applications have been reviewed over the years in several excellent publications, e.g. basic[1] and advanced textbooks,[2,3] monographs[4,5] and specialized editions.[6,7] This article is not intended as a condensed version of the material in those contributions. Instead it focuses on two tasks: first to provide an overview of the fluorescence phenomenon in proteins, and second to illustrate applications of fluorescence spectroscopy in state-of-the-art (but not necessarily high-tech) studies. The goal is to concentrate on fundamental principles and advanced applications.

1.1 What is Measured in a Protein Fluorescence Experiment?

The four principal aspects of the fluorescence phenomenon are: energetics, probability, kinetics and vectoriality. Each can be studied experimentally by measuring the intensity of the emitted light as a function of wavelength, orientation of polarizers, time elapsed since excitation, etc. Therefore the question posed in the title of this section has a simple answer – fluorescence intensity, I. Let us now consider how this basic measurement can be converted into multiple measures characterizing various aspects of fluorescence.

Intensity measured as a function of wavelength (or wavenumber) comprises the fluorescence spectrum. The position of the emission spectrum reflects changes in energetics of the excited and ground state between excitation and emission of a photon. In practice the position of the emission band is expressed as a wavelength of the maximum λ_{max}.

Total intensity of the emission at all wavelengths is related to fluorescence quantum yield, q. Quantum yield is defined as a ratio of the number of emitted quanta to the number of absorbed quanta. It reflects the probabilities of various photochemical and photophysical processes leading to radiational or radiationless deactivation of the excited state.

The time-dependence of fluorescence intensity following excitation in the ensemble of fluorophores is related to the excited state lifetime, τ. This kinetic parameter characterizes the average time a molecule spends in the excited state before emitting the photon:

$$I(t) \sim e^{-t/\tau}$$

More often than not, fluorescence intensity follows a more complex law of decay than simple exponential (see section 2.3). Two experimental schemes allowing determination of the fluorescence lifetime are known as pulse domain and frequency-modulation domain spectroscopy (for more on lifetime measurements see Lakowicz[1,3]).

Anisotropy, r (along with the interchangeably related polarization, P), is a vectorial characteristic reflecting changes in the direction of the emission transition moment with regard to the excitation transition moment. Anisotropy is affected by the relative orientation of the dipoles in the ground and excited states and by changes in the orientation of the excited state with respect to external (laboratory) coordinates due to the molecule's rotation or radiationless energy transfer to another molecule prior to emission. Anisotropy and polarization are defined in Equations (1) and (2) by intensities of vertically ($I_{\parallel}$) and horizontally ($I_{\perp}$) polarized light measured when excitation light is vertically polarized:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{1}$$

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{2}$$

Anisotropy and polarization can be measured in a steady-state and kinetic regime (for more on anisotropy measurements and their applications see Steiner[8]).

The ratio of intensities in the absence, I_0, and in the presence, I, of an external quencher at concentration $[Q]$ is used to determine a Stern–Volmer constant, K_{SV}, as shown in Equation (3):

$$\frac{I_0}{I} = 1 + K_{SV}[Q] \tag{3}$$

If the changes in lifetime follow the changes in intensity, the quenching is called dynamic and the biomolecular quenching rate, k_q, can be determined using Equation (4):

$$\frac{\tau_0}{\tau} = 1 + K_{SV}[Q] = 1 + k_q\tau_0[Q] \tag{4}$$

The biomolecular rate gives information on chromophore accessibility to the aqueous phase and on the distribution

of charges surrounding the fluorophore if a charged quencher is used. In the case of static quenching, only the intensity decreases with addition of quencher while the lifetime remains the same (for more on quenching measurements and their applications see Eftink[9]).

Another useful parameter is the efficiency of the long-range energy transfer, E, between an initially excited molecule (donor) and another chromophore (acceptor). Efficiency of this dynamic process can be calculated from either intensity or lifetime of donor alone (I_D, τ_D) and in the presence of acceptor (I_{DA}, τ_{DA}) as shown in Equation (5):

$$E = 1 - \frac{I_{DA}}{I_D} = 1 - \frac{\tau_{DA}}{\tau_D} \tag{5}$$

Efficiency of transfer is used to estimate the distance between donor and acceptor, R, according to Förster theory, as shown in Equation (6):

$$E = \frac{R_0^6}{R_0^6 + R^6} \tag{6}$$

where R_0 is a Förster distance for half-transfer and is a characteristic of a donor–acceptor pair and its environment (for more on energy transfer measurements and their applications see Cheung,[10] Wolber and Hudson,[11] and Wu and Brand[12]).

It is important to keep in mind that neither distance, nor quenching rate, nor strictly speaking lifetime are *measured* in a fluorescence experiment. All these important parameters are derived from the basic measurements of intensity, using more or less restrictive assumptions (e.g. absence of nonspecific perturbations after introduction of the acceptor fluorophore, or fluorescence decay being adequately described by a sum of exponential components) and therefore are to some extent model-dependent.

2 PROTEIN FLUORESCENCE

2.1 Spectral Properties of Protein Fluorophores

2.1.1 Tryptophan and its Derivatives

Absorption of tryptophan is due to $\pi \rightarrow \pi^*$ transitions in the indole ring **(1)**. The short wavelength band (220 nm) is due to the ^1B_b transition while the long wavelength band (260–290 nm) consists of the two overlapping transitions ^1L_a and ^1L_b with vectors almost perpendicular to each other. The ^1L_a transition is sensitive to the polarity giving rise to a slight shift in absorbance spectrum. The ^1L_a transition is believed to be the main contributor to the emission.

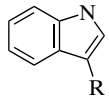

(1)

In aqueous solution tryptophan displays a wide, structureless fluorescence band with a maximum of about 350 nm and width of about 60 nm. It is worth noting that tryptophan fluorescence spectra measured under the same conditions but using different instruments may differ slightly in position of the maximum (348–353 nm), width and shape because of the differences in spectral calibration. Unfortunately, at present there are no generally accepted standards for the spectral calibration of instruments in the near-ultraviolet region. Because excitation leads to a substantial increase in the dipolar moment (about 4 D), shifts in the emission are much stronger. This shift is caused by the orientational relaxation processes involving the dipole of the chromophore and solvent dipoles. This sensitivity of tryptophan emission to polarity and mobility of environment makes tryptophan fluorescence an important tool in studies of protein structure and dynamics.

The absolute quantum yield of tryptophan in an aqueous environment was reported to be 0.13.[13] In most cases, however, a knowledge of the absolute value is not necessary and relative quantum yield is used instead. Tryptophan fluorescence is susceptible to the quenching by water which occurs with high activation energy (11–12 kcal mol^{-1}).[14,15] This leads to the frequently observed decreases in quantum yield upon denaturation when tryptophan becomes exposed to an aqueous environment. However, there are numerous exceptions to this rule because tryptophan fluorescence can also be quenched in a protein's native state.

Almost all polar protein groups can quench tryptophan fluorescence, to some extent.[14,16–21] Possible nonradiative processes include photoionization, intersystem crossing, exciplex formation, and excited state proton and excited state electron transfer. Aspartic and glutamic acid residues are effective dynamic quenchers in their neutral, but not charged, forms. Lysine and arginine are also dynamic quenchers, but are more effective when charged. Histidine at low pH appears to quench by formation of a stacking complex with the indole ring. The nonprotonated histidine is also able to quench tryptophan fluorescence, although with much-reduced efficiency. Disulfide is one of the strongest quenchers of tryptophan fluorescence, although single cysteine can be an effective quencher too. In addition, amide and peptide groups were shown to act as dynamic quenchers. In heme-containing proteins, long-range energy transfer quenching is important due to the spectral overlap

of tryptophan emission and heme absorbance. While such effects complicate the analysis of protein fluorescence, they can also be used to extract valuable structural information. For example, in membrane-bound cytochrome b_5, heme quenching of Trp-109 was used to estimate the distribution of distances between two protein domains.[22]

Tryptophan appears to be uniquely sensitive to quenching by a variety of substances, such as oxygen, iodide, bromide, acrylamide, succinimide, hydrogen peroxide, dichloroacetamide, pyridinium hydrochloride, NO_3^-, Cs^{2+}, Cu^{2+}, Pb^{2+}, Cd^{2+} and Mn^{2+}.[14,16,23] This sensitivity to quenchers allows determination of the accessibility of the tryptophan residues in proteins by quenching measurements.

The fluorescence decay of tryptophan does not follow a single exponential. Szabo and Rayner[24] had reported two decay times of 0.5 and 3.1 ns for tryptophan zwitterion in an aqueous environment at room temperature. The short-lived component has a blue-shifted spectrum. The heterogeneity of decay of tryptophan derivatives in fast-relaxing environments is attributed to the side chain rotamers. Decay becomes even more complex when solvent relaxation occurs on a nanosecond timescale (see section 2.3).

The existence of two overlapping transitions and energy transfer occurring predominantly from ^1L_b to ^1L_a complicates the appearance of the excitation polarization spectrum of tryptophan. In their classic work, Valeur and Weber[25] have resolved two transitions in the excitation spectrum of indole and tryptophan in frozen propylene glycol using polarization data. The limiting anisotropy of tryptophan depends on the excitation wavelength and is about 0.3 at 300 nm, and 0.2 at 270 nm. The spectral region in-between has several sharp bands, which are expected to shift depending on the solvent conditions, making the region from 280 to 300 nm difficult to use for polarization measurements.

An important emerging field in protein fluorescence is related to the use of spectrally enhanced protein mutants.[26] This approach takes advantage of tryptophan analogs with different photophysical properties incorporated as intrinsic fluorophores either by chemical synthesis or, biosynthetically, using tryptophan auxotroph *Escherichia coli* strains.[27,28] Three analogs are suggested to be most useful for studies of protein–protein and protein–nucleic acid interactions: 5-hydroxytryptophan **(2)**, 4-fluorotryptophan **(3)** and 7-azatryptophan **(4)**. 5-Hydroxytryptophan absorbance has a long wavelength shoulder going as far as 320 nm, allowing this fluorophore to be selectively excited in the presence of multiple normal tryptophan residues. In addition, when excited in this region 5-hydroxytryptophan has a high limiting

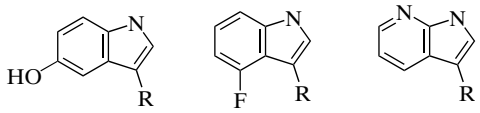

(2) **(3)** **(4)**

anisotropy, making it useful for polarization studies. 4-Fluorotryptophan is nonfluorescent, making a ''silent'' analog. It also has been used in nonfluorescence experiments because of its altered ground state dipole. 7-Azatryptophan has a dramatic change in quantum yield upon exposure to water, making it potentially useful in protein folding experiments.

2.1.2 Other Natural Fluorophores and Fluorescence Labels

A variety of biological molecules contain naturally occurring or intrinsic fluorophores. Tryptophan is the most highly fluorescent amino acid in proteins (see sections 2.1.1 and 2.2). It is so widely used that the term ''natural protein fluorescence'' is almost always associated with tryptophan fluorescence. The second most fluorescent amino acid is tyrosine, but its application is mostly limited to tryptophan-free proteins. Fluorescence of phenylalanine is weak and almost never used in protein studies. Another class of natural fluorophores consists of cofactors, NADH (reduced β-nicotinamide adenine dinucleotide) being the most prominent.

The fluorescence properties of NADH have been the subject of multiple studies.[29–31] In aqueous solution the quantum yield of fluorescence is low and the average lifetime is in the subnanosecond range due to stacked conformation of the molecule. Binding to liver alcohol dehydrogenase leads to a blue-shift in absorbance and fluorescence and an increase in quantum yield (see section 3.2). This effect is enhanced by formation of a ternary complex with the substrate analog isobutyramide. The lifetime distribution for NADH shifts into a nanosecond range and undergoes a complex change in binary and ternary complex, indicating the existence of an excited state reaction.[31]

The fluorescence properties of natural chromophores are frequently inadequate for certain studies. In these cases fluorophores foreign to the system under study but displaying improved spectral properties are chosen. These extrinsic fluorophores when covalently linked to a protein (usually at a cysteine or lysine side chain) are called fluorescence labels. Modern labeling procedures in combination with mutagenesis allow labels to be introduced selectively at a specific location in a protein. The *Handbook of Fluorescent Probes and Research Chemicals*[32] provides an excellent source of information on various dyes and labeling procedures.

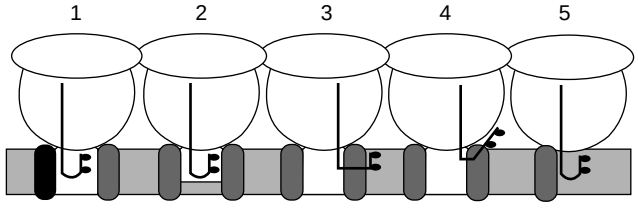

Figure 1 Example of the experimental system requiring the use of fluorescence labels in protein studies. Neither intrinsic fluorescence nor conventional labeling can specifically probe the environment of the nascent chain (solid line) in the presence of ribosome–translocon complex (large ovals). To overcome this limitation Crowley et al.[34] have developed a way to biosynthetically incorporate NBD-labeled (small ovals) lysine residues in the nascent chain. Fluorescence of NBD dye is sensitive to the polarity of its environment and to addition of water-soluble quenchers, such as iodide. Combination of steady-state and phase-modulation measurements allowed differentiation between various possibilities of organization of the translocon complex. (Reproduced by permission of Cell Press.)

In an ingenious example of the use of a fluorescence label, the fluorescence properties of 6-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)aminohexonic acid (NBD) were utilized to determine the environment of the nascent chain in the translocon complex[33–35] (Figure 1). Emission of NBD is very sensitive to the properties of its environment: moving from an aqueous to apolar environment increases the lifetime and the quantum yield and causes blue-shift of emission. Nascent chains, however, cannot be labeled selectively in the presence of ribosomes and endoplasmic reticulum-membrane proteins. Site-specific positioning of the probes in the nascent polypeptides can only be achieved by using aminoacyl-tRNA analogs that incorporate amino acids labeled with suitable probes into the chain as it is being synthesized. Johnson et al.[36] devised a method for preparing analogs in which probes are covalently attached to the side chain of lysine in Lys-tRNA. The validity of the approach was demonstrated by showing that the modified lysines were incorporated successfully into proteins in vitro. Since different models of translocon organization differ primarily in the exposure of the nascent chain to the cytoplasm, they could be tested directly by determining whether fluorescent dyes in the nascent chains of membrane-bound ribosomes were accessible to water-soluble quenchers in the cytoplasm. NBD dyes located at different positions along the nascent chain pathway in free ribosomes were each found to have very short fluorescence lifetimes (less than a nanosecond) and to be quenched when iodide ions were added to the solvent.[34] Thus, the nascent chain is in an aqueous environment and exposed to the cytoplasm when bound to a free ribosome. But when iodide ions were added to membrane-bound samples of the same ribosome–nascent-chain complexes, the Stern–Volmer constant for fluorescence quenching was reduced by an order of magnitude. This observation led to the conclusion that, in early stages of translocation, the nascent chain is tightly surrounded by the ribosome and membrane components of the endoplasmic reticulum.

2.2 Spectral Classes of Tryptophan Residues

The position of the maximum of the fluorescence spectrum of tryptophan residues in proteins varies from 307 to 353 nm. According to the model of discrete states of Burstein et al.[37,38] there are five most probable spectral forms of tryptophan residues (Figure 2). Formation of various exciplexes (complexes in the excited state) and subsequent dipole relaxation is believed to cause the red-shift of emission.

Spectral form A corresponds to the emission of the unperturbed indole chromophore in the extremely nonpolar environment inside the protein globule. Such emission in its pure form was found only in azurin, however the class A spectrum contributes to the total spectra of a few other proteins. Azurin, a small globular metal-binding protein, contains a single tryptophan residue that displays a rather unusual fluorescence spectrum. It is extremely blue-shifted, with a main maximum of about 307 nm, and it possesses a distinct vibrational structure. The emission from the ^1L_b state is assumed to contribute significantly in the spectrum, giving rise to mirror symmetry of emission and absorbance.

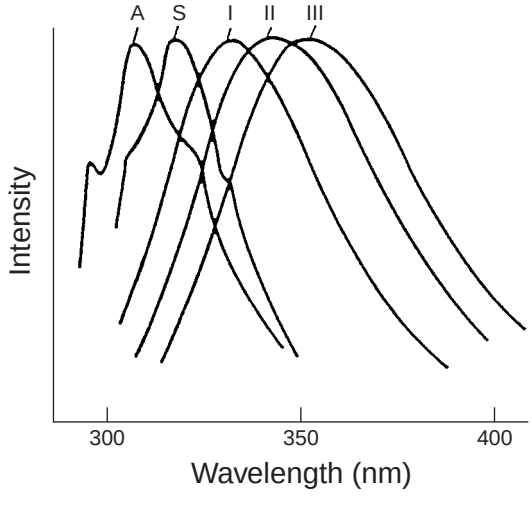

Figure 2 Normalized fluorescence spectra of tryptophan residues belonging to five spectral classes A, S, I, II and III according to Burstein.[38] Extreme variability of the emission of the indole fluorophore, depending on polarity and mobility of its environment, makes tryptophan fluorescence a sensitive tool in protein analysis. (Reproduced by permission of ONTI NCBI.)

Freezing an azurin solution does not change the position of its fluorescence spectrum.

Spectral form S corresponds to the emission of the indole chromophore located in the relatively nonpolar environment inside the protein globule and forming a 1:1 exciplex with some polar protein group. The S spectrum has a maximum at 316–317 nm and shoulders at 305–307 nm and 320–330 nm. It is of interest that a pure S spectrum has never been reported, but is always accompanied in proteins by a contribution from the class I spectrum (see below), which corresponds to a 2:1 exciplex. For example, types S and I contribute almost equally to emission of L-asparaginase.[39] It is assumed that during the excitation lifetime the 2:1 complex is formed to a different extent in different proteins depending on the mobility and abundance of polar groups in the tryptophan environment. Freezing of the solution does not shift class S spectra, indicating that a 1:1 complex can be formed at temperatures as low as −196 °C.

Spectral form I corresponds to the emission of the indole chromophore in the somewhat polar but perhaps rigid environment inside the protein globule, forming a 2:1 exciplex with two neighboring polar protein groups. The fluorescence spectrum of class I has a maximum at about 330 nm, width of about 50 nm and no vibrational structure. A class I spectrum is observed for actin, chymotrypsin and tetrameric melittin under high-salt conditions.

Spectral form II corresponds to the emission of the indole chromophore at the protein surface. It is assumed to be in contact with bound water and other polar groups. The fluorescence spectrum of class II is also structureless, but red-shifted as compared to class I. The position of the maximum is at about 340 nm and the width is about 55 nm. Many proteins contain tryptophan residues of this class, e.g. human serum albumin and myosin.

Spectral form III corresponds to the emission of the indole chromophore at the protein surface in contact with free water molecules. The spectrum of class III, with its maximum at about 350 nm and width of about 60 nm, nearly coincides with the spectrum of free tryptophan. Tryptophan residues of spectral class III very seldom occur in the native proteins, but are typical for unfolded states.

Due to their location on the protein interface, tryptophans of spectral classes II and III are easily accessible to external quenchers in aqueous phase. At the same time, tryptophans of classes A, S and I are located inside the protein structure and are poorly accessible to water-soluble quenchers. Freezing of the water solution of a protein shifts the fluorescence spectra of classes I, II and III towards shorter wavelengths due to the immobilization of their polar environment. It should be noted that the model of discrete states of tryptophan residues in proteins is a statistical one, i.e. tryptophan residues located in environments that result in one of the five likeliest spectral forms is more probable than finding them in intermediate environments.

Most proteins, however, exhibit spectra that contain contributions from different classes. To resolve those contributions a fitting procedure was designed based on a parameterized description of fluorescence spectra. At the heart of this analysis lies approximation of emission spectra with the log-normal distribution[40] usually expressed on the scale of wavenumbers, ν, as shown in Equations (7) and (8):

$$I(\nu) = I_{\max} \exp\left(-\frac{\ln 2}{\ln^2 \rho} \ln^2 \frac{a-\nu}{a-\nu_{\max}}\right) \qquad \text{for } \nu < a \qquad (7)$$

$$I(\nu) = 0 \qquad \text{for } \nu \geq a \qquad (8)$$

where $I_{\max}$ is intensity at position of maximum, $\nu_{\max}$, and $\rho = (\nu_{\max} - \nu^-)/(\nu^+ - \nu_{\max})$ is spectral asymmetry. Parameter $a = \nu_{\max} + (\nu^+ - \nu^-)\rho/(\rho^2 - 1)$, where ν^- and ν^+ are the positions on the wings of the spectrum where intensity equals half of $I_{\max}$. The total of four fitting parameters ($I_{\max}$, $\nu_{\max}$, ν^-, ν^+) per spectrum is adequate to get a unique solution, but to resolve two or more spectral components additional constraining factors are required. They are obtained from the empirical observation that width and position of the maximum of tryptophan derivatives in isotropic media are related as shown in Equations (9) and (10):

$$\nu^+ = 0.8308\nu_{\max} + 7071\ \text{cm}^{-1} \qquad (9)$$

$$\nu^- = 1.1768\nu_{\max} + 7681\ \text{cm}^{-1} \qquad (10)$$

An advanced version also utilizes spectra collected at different concentrations of quenchers to increase reliability and robustness of fit.[41]

In a simplified version of the analysis, the spectrum is fitted to a single log-normal distribution (often on a wavelength scale) to recover full width at half-height, γ, and position of maximum, $\lambda_{\max}$.[37,42] If the point on a spectral width vs position plot appears above the standard line for tryptophan in model isotropic media, several spectral classes are present. For example, a conformational change in bee venom peptide melittin can be followed by position-width analysis of steady-state spectra of Trp-19 (Figure 3). At low ionic strength melittin exists as a mainly unstructured monomer with highly exposed tryptophan (spectral class III). Additions of salt induce a monomer–tetramer transition, resulting in the removal of the tryptophan side chain from the aqueous phase (spectral class I). Intermediate states have significantly broadened spectra and follow theoretically

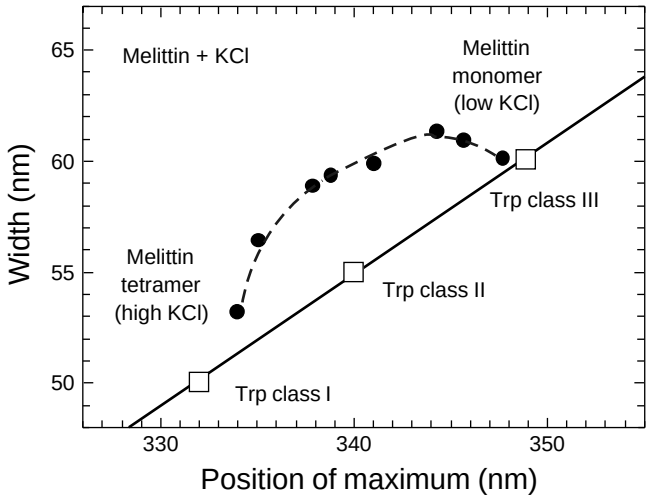

Figure 3 Example of protein conformational transition followed by spectral changes in tryptophan emission spectrum. Steady-state fluorescence spectra were fitted to a log-normal distribution to recover position of maximum and width. Open squares represent discrete states of tryptophan residues in proteins belonging to spectral classes I, II and III. Fluorescence is heterogeneous when the position–width point is above the straight line. Closed circles represent titration of melittin solution with KCl, which in this peptide causes conformational change from monomer to tetramer. This transition results in changes of exposure of the tryptophan residue from completely exposed (class III) directly to buried (class I), without going through the intermediate state (class II).

predicted semi-arc patterns.[37] Several advantages of position–width analysis over simple analysis of λ_{max} are immediately obvious. First, it is clear that spectral class II is not involved in the transition, because the spectra in the mid-transition are too broad, although they have the same maximum as II. Second, λ_{max} is not changed appreciably in the last stages of transition, and completion of conformational change is best judged by the narrowing of the spectral width. It should be emphasized, however, that despite its sensitivity the analysis of spectral position cannot be used to quantitate the conformational change or ligand binding, because λ_{max} is not a linear response function (see section 3.2). To calculate the thermodynamic parameters one should use one of the linear response spectroscopic functions as the observable, e.g. the intensity at any constant wavelength.

2.3 Protein Dynamics and Heterogeneity of Fluorescence Decay

It has long been noted that even for single tryptophan-containing proteins, fluorescence decay does not follow simple exponential kinetics.[43] Due to the nature of the lifetime analysis certain assumptions have to be made, which usually means that complex decay is assumed to be described by a number of exponential components. As instruments become more sophisticated, more components have been reported and even the distributions of lifetimes have been utilized to fit the data.[44–47] Regardless of the mode of analysis (discrete exponentials or distributions), the main empirical result is that it takes several parameters to describe fluorescence decay. Exponentials, however, do not form an orthogonal system of functions and multiple solutions can exist which fit the same data equally well. Therefore, one should use the term multi-exponential decay with caution, keeping in mind that the exact nature of nonexponentiality might not be known, and that recovered components may or may not have a physical meaning.

Interpretations of complex decay of tryptophan fluorescence fall into two main categories: ground state heterogeneity (e.g. rotamer model) and excited state reaction (e.g. relaxation model). Based mainly on model studies done with tryptophan and tyrosine derivatives, several groups have suggested that different rotameric forms are responsible for nonexponential decay.[24,48,49] In recent years there has been a breakthrough in understanding the mechanisms governing the decay of rotamers of restricted indole derivatives.[20,50,51] However, those model studies were carried out in isotropic fast-relaxing media. Proteins, in contrast, are heterogeneous systems with a hierarchy of internal motions that cover a wide range of correlation times, including the nanosecond time window of fluorescence. On the other hand, the excited state reaction approach relies on protein dynamics to convert electronically excited indole into other spectroscopic species. More specifically, the relaxation model assumes that such reaction involves reorientation of polar groups surrounding tryptophan in a protein.[52–54] Neither of the models has been unequivocally proven for any specific protein.

Recently it has been suggested that excitational dependence can be used as a criterion to distinguish between alternative mechanisms causing deviations from exponential decay in proteins.[55] This approach takes advantage of considerations developed to explain inhomogeneous broadening of the electronic spectra of dye molecules caused by the distribution in configurational energy in solvate.[56–58] Such systems are characterized by the following features: (1) decay is nonexponential, (2) decay depends on the excitation wavelength, λ_{exc}, and becomes faster at longer λ_{exc} (red-edge effect), and (3) in extreme cases a rising component in intensity (negative pre-exponential) is observed at the longer emission wavelength.

An abbreviated summary of these features for fluorescence decay in simple model systems and in proteins

Table 1 Summary of heterogeneity of fluorescence decay of indole fluorophore in model media and in proteins

System	Is fluorescence decay monoexponential?	Does decay depend on excitation wavelength?	Is the rising intensity component observed?	Nature of heterogeneity of decay
Indole in water or methanol	Yes	No	No	None (no rotamers, fast relaxation)
Indole in 40% glycerol	No (2 components)	Yes	No	Mild relaxation effects
Indole in 90% glycerol	No (3 components)	Yes	Yes	Relaxation effects
Trp in water	No (2 components)	No	No	Rotamers
Trp in 40% glycerol	No (3 components)	Yes	No	Rotamers + mild relaxation effects
Proteins (single tryptophan residue)	No (2–4 components, usually 3)	Both possibilities can be observed	No (with rare exceptions)	? (possible combination of several mechanisms)

is presented in Table 1 (Ladokhin, unpublished).[55] Slowing of dipolar relaxation by additions of glycerol leads to progressive heterogeneity and excitational dependence in fluorescence decay of the indole molecule, which obviously lacks rotameric forms. Note that deviation from exponentiality starts at low glycerol concentrations, before the negative pre-exponential could be observed under commonly used instrumental conditions (the width of the excitation pulse was 60 ps). As described above (see section 2.1.1), tryptophan zwitterion in aqueous solution shows a double exponential decay due to rotameric forms.[24] In this case, however, the decay is independent of the excitation wavelength. When a tryptophan molecule is placed under conditions of mild viscosity, decay becomes more complex, and in this case excitational dependence can be detected. In proteins the excitation dependence can be small or significant and the negative exponential normally is not observed unless four exponential components are used in the analysis, and then only for a few proteins. Melittin in methanol, when it is helical and monomeric, constitutes an exception when a negative pre-exponential is observed with one of the three components.[55] This overall behavior of proteins is consistent with contributions from both ground state heterogeneity and excited state reactions.

The model studies presented above indicate the importance of protein dynamics for interpretation of fluorescence decay in proteins. This is also true for other fluorescence parameters, such as dynamic depolarization, biomolecular quenching rate, temperature-induced quenching in the native state, red-edge excitation shift, distribution of distances calculated in energy transfer experiments, etc. In a sense all protein fluorescence spectroscopy is influenced by the dynamic nature of protein. Further understanding of this link will allow one to solve the reverse task and to determine dynamic parameters of protein structure from its fluorescence.

3 FLUORESCENCE STUDIES OF PROTEIN CONFORMATION AND INTERACTIONS

3.1 Protein Folding

Fluorescence spectroscopy plays a major role in protein folding studies.[59,60] For example, the fluorescence increase of 8-anilino-1-naphthalensulfonic acid was used to identify the molten globule state during refolding of various proteins,[61] and intrinsic fluorescence, along with other methods, was utilized to provide evidence for a two-state transition in chymotrypsin inhibitor.[62] A variety of fluorescence observables can be used to follow denaturation or renaturation of the protein native structure induced by changes in temperature, pH or additions of solutes. One such example has already been discussed in previous sections – changes in the emission bandshape following tetramerization of melittin was presented in Figure 3 (see section 2.2). Another example is the change in emission maximum and polarization of tryptophan fluorescence of apohorseradish peroxidase upon addition of a denaturing agent, guanidinium chloride.[63] Unfolding of this protein results in increased exposure of a single tryptophan residue to aqueous environment revealed by the fluorescence red-shift (Figure 4). Increased mobility of the tryptophan due to the loss of native packing contacts upon denaturation results in the parallel loss of emission polarization. Note that an excitation wavelength of 300 nm was used for polarization measurements to avoid strong depolarization (insert, Figure 4), due to overlapping ^1L_a and ^1L_b transitions at the shorter wavelengths (see section 2.1.1). In addition to emission maximum and polarization, a number of other parameters (e.g. intensity, lifetime distribution, Stern–Volmer constant, efficiency of the radiationless energy transfer) can be utilized in steady-state protein folding studies. But the advantages of fluorescence studies of folding are demonstrated most vividly, perhaps, with kinetic measurements.

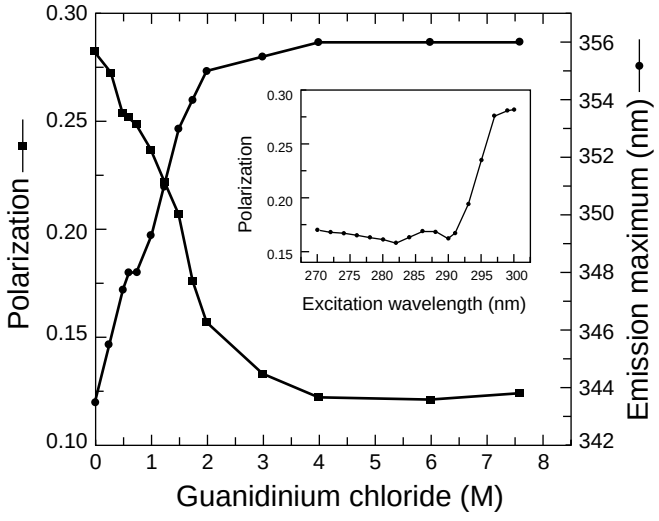

Figure 4 Example of chemical-induced protein unfolding studied by fluorescence spectroscopy.[63] Additions of guanidinium result in red-shift of fluorescence and in decrease of emission polarization of a single tryptophan in the apohorseradish peroxidase. Insert shows the excitation polarization spectrum. (Reproduced by permission of the Biophysical Society.)

Among the spectroscopic methods used in combination with stopped-flow mixing, fluorescence is probably the simplest, yet most sensitive and widely employed. Recent years have been marked by considerable advances in both mixing techniques and kinetic acquisition of spectroscopic data. Application of advanced continuous-flow mixing to follow the energy transfer from tryptophan and tyrosines to heme in cytochrome *c* has allowed probing of the barrier-free submillisecond transition between the initially collapsed state and the native state. Using a combination of both continuous-flow microsecond mixing and conventional stopped-flow mixing, Shastry et al.[64] demonstrated how continuous kinetic data from 50 μs to >10 s after initiation of the folding event can be obtained (Figure 5). Ramsay et al.[65,66] have developed a fluorimeter/CD photometer hybrid allowing global changes in secondary structure and local changes in the tryptophan environment to be determined simultaneously for the same sample.

Beechem et al.[67–69] have pioneered the development of another important technical advance in spectroscopic analysis of fast kinetic effects by combining the advantages of time-resolved spectroscopy with kinetic stopped-flow measurements. This double-kinetics technique achieves simultaneous measurement of changes in fluorescence on both the picosecond/nanosecond and millisecond timescale. After initiation of folding the decay of tryptophan anisotropy is measured in millisecond intervals of time, thus allowing the determination

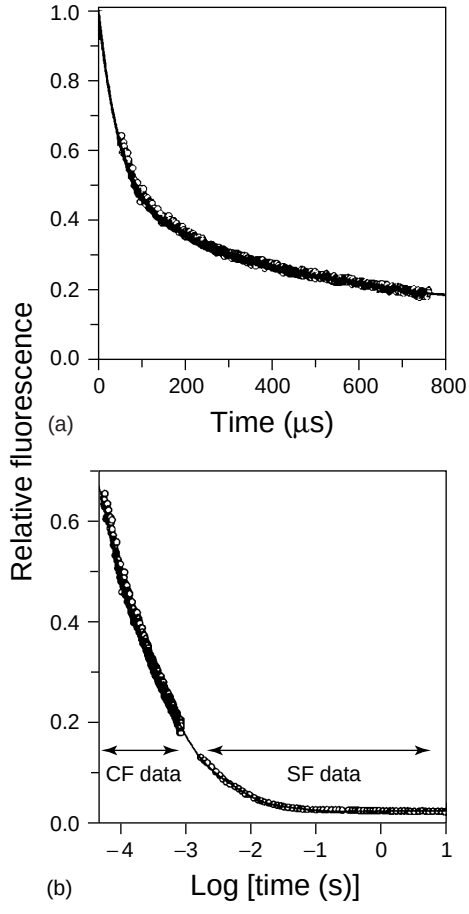

Figure 5 Example of protein folding kinetic measurement utilizing tryptophan fluorescence.[64] (a) Refolding of acid-unfolded horse cytochrome *c* was followed on the submillisecond timescale in a continuous flow capillary mixing apparatus. (b) Combination of continuous flow (CF) and conventional stopped-flow (SF) experiments allows collection of data over a wide range of time (symbols) and determination of the five kinetic phases with correlation times ranging from tens of microseconds to tens of milliseconds (solid line). (Reproduced by permission of the Biophysical Society.)

of tryptophan rotation mobility "in real time" during protein refolding. This allows ascertaining the timescale associated with tryptophan "packing" into intermediate structures and eventually into the final native state. Such experiments can also follow the changes in lifetime due to different fluorescence quenching in the folded and unfolded state. If two fluorophores capable of making a donor–acceptor pair are introduced, such studies can be used to determine the changes in distance between labeled sites on the millisecond timescale.

In addition to the water-soluble systems discussed above, fluorescence spectroscopy is now being employed to study folding of membrane proteins[70–74] (see section 3.3).

3.2 Protein–Ligand Interactions

There are two principal possibilities that can be exploited to study protein–ligand interactions. The fluorescence of either the protein or the ligand can change upon binding. These two situations require somewhat different data analysis procedures.[75] When changes in the fluorescence of a protein are observed, the fluorophore may be one or more of the intrinsic tryptophan residues, or it may be the covalently attached or tightly bound fluorophore.

Tryptophan is a relatively sensitive fluorophore. Its spectrum can vary more than 40 nm (see section 2.2) and its quantum yield in proteins and peptides ranges from virtually zero to almost 0.5. The latter change is usually accompanied by a corresponding variation in the excited state lifetime. The binding of a ligand to a protein may directly affect the fluorescence of a tryptophan residue by acting as a quencher (i.e. by a collisional or energy transfer mechanism) or by physically interacting with the fluorophore and thereby changing the polarity of its environment and/or its accessibility to solvent. Alternatively, a ligand may bind at a site remote from the tryptophan residue and act via a generalized mechanism by inducing protein conformational change, which alters the microenvironment of the tryptophan. Both direct and generalized effects may result either in enhancement or quenching of fluorescence and/or in shifts in the spectrum to the red or blue. If there are multiple tryptophan residues in a protein, this will make it difficult to assign any fluorescence changes to a specific residue. To overcome this limitation, structural studies are now being conducted on single tryptophan-containing mutants. However, for many other purposes, e.g. the thermodynamics of ligand binding, the presence of multiple fluorophores is a minor consideration. More important is the fact that the observable be a linear-response function.

3.2.1 Linear-response Spectroscopic Techniques

Consider the simple case of a two-state equilibrium (the general case is described by Toptygin and Brand[76]) that can be followed by a change in a certain spectroscopic parameter S that changes from a value $S_{\min}$ in the absence of binding to $S_{\max}$ when all of the molecules are bound. The fractional change in the spectroscopic parameter f_{signal}, defined as $f_{\text{signal}} = (S - S_{\min})/(S_{\max} - S_{\min})$, is often used to characterize the extent of completion of the binding reaction. The fraction of bound molecules $f_{\text{bound}} = P_{\text{bound}}/P_{\text{total}}$ will coincide with f_{signal} only if S is a linear-response function. For that case, the signal observed can be represented as a linear combination of molar fractions of bound and free molecules weighted by their corresponding molar spectroscopic characteristics, s_{bound} and s_{free}, as shown in Equation (11):

$$\begin{aligned}\frac{S}{P_{\text{total}}} &= f_{\text{bound}}s_{\text{bound}} + f_{\text{free}}s_{\text{free}} \\ &= f_{\text{bound}}s_{\text{bound}} + (1 - f_{\text{bound}})s_{\text{free}}\end{aligned} \tag{11}$$

Not all spectroscopic data satisfy this equation. For example, the following spectroscopic parameters are not linear-response functions and therefore cannot be utilized for measurements of binding: position of maximum of the fluorescence spectrum and related parameters such as center of weight, center of the chord at any intensity level, or ratio of intensities on the wings; fluorescence polarization (or anisotropy); intensity at the maximum of the spectrum; the phase ϕ and modulation m from frequency domain fluorimetry; fluorescence decay curve collected to a constant peak; and transmittance from spectrophotometric measurements. Parameters that are linear-response functions and that can be utilized for measurements of binding include the following: fluorescence quantum yield; steady-state intensity at a constant wavelength I_{ss}; $I_{ss}m\cos\phi$ and $I_{ss}m\sin\phi$;[76] fluorescence decay curve collected to a constant time; vertical and horizontal components of fluorescence intensity (and of any other angle too); and absorbance from spectrophotometric measurements.

In addition to general rules that apply to all binding experiments, there are some peculiarities relevant to specific aspects of experimental schemes. For example, in membrane partitioning experiments there is an explicit assumption that s_{bound} does not depend on the completion of binding or on the number of bound molecules per lipid vesicle. However, in the case of fluorescence, the intensity can be a nonlinear function of bound molecules due to nonradiative energy homotransfer or other self-quenching mechanisms. This possibility is often overlooked during the analysis of complex membrane partitioning. Various spectroscopic and nonspectroscopic methods to quantitate binding to membranes are discussed elsewhere.[77]

Toptygin and Brand[76] have developed special software to fit multiple families of titration data obtained with any of the linear-response spectroscopic techniques. The SPECTRABIND program uses model equilibrium equations as constraints to determine the basic components associated with the actual chemical species and their concentrations. A statistical test is used to discriminate between adequate and inadequate models, the decision being independent of any knowledge and/or assumptions about the spectroscopic characteristics of chemical species. This approach is aimed at systems involving multiple interactions, where the changes in spectroscopic signal cannot be attributed to a single interaction. An example

of the application of SPECTRABIND to the interaction of horse liver alcohol dehydrogenase with NADH and isobutyramide is given in Figure 6.[76] Fluorescence spectra of NADH in various binary and ternary complexes recovered from the same titration data sets using two different binding models are presented. Both models assume that the two binding sites (one on each of the identical subunits of the protein dimer) are identical and independent in terms of association constants. Figure 6(a) contains results obtained under the additional generic assumption that binding to the two sites is also spectroscopically independent. The three essential spectroscopic species correspond to a free NADH (N), its binary complex with the protein subunit (SN), and its ternary complex with additional isobutyramide (SNI). Protein (S) and isobutyramide (I) by themselves produce negligible signal, while the background signal (B) contains a sharp peak of Raman scattering band of water. The model from Figure 6(b) does not assume spectroscopic independence of binding sites and contains multiple species of various combinations of the entire dimeric protein (P) with one or two NADH and isobutyramide molecules. The association constants for each binding step are also determined.[76] This example illustrates the ability of the SPECTRABIND program to analyze complex binding equilibria.

The basic limitation of spectroscopic techniques for studying protein–ligand interactions is the necessity for a sufficiently strong change in some spectroscopic signal. The lack of an adequate signal can hinder rigorous quantitative determination of binding. To overcome this limitation in studies of nucleic acid-binding proteins, Jezewska et al.[78,79] applied an inventive approach that takes advantage of the strong fluorescence of the reference ligand. An example of application of the Macromolecule Competition Titration method for binding of nonfluorescent single-stranded nucleic acid poly(dA) to DnaB helicase is presented in Figure 7. The observed fluorescence signal was due to the poly(dεA), an etheno derivative of poly(dA), used as a reference. This method allows the determination of absolute average binding density and construction of a model-independent true binding isotherm.

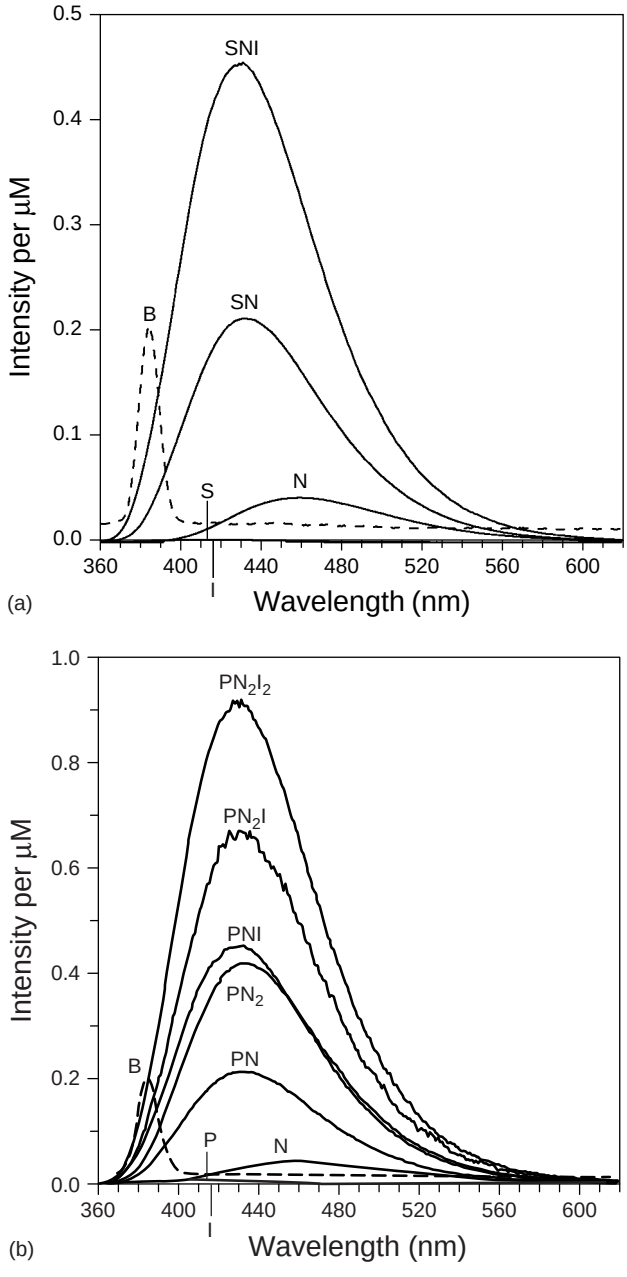

Figure 6 Example of the protein–ligand binding study followed by fluorescence. Individual spectra of NADH when free in solution (N), bound in a binary complex with horse liver alcohol dehydrogenase ((a) SN; (b) PN and PN_2), and in a ternary complex in the presence of isobutyramide ((a) SNI; (b) PNI, PN_2I and PN_2I_2), along with corresponding equilibrium binding constants, were obtained with the help of the SPECTRABIND program of Toptygin and Brand.[76] Fluorescence spectra presented in (a) and (b) are obtained from the same titration data sets using two different binding models. Both models assume that the two binding sites (one on each of the identical subunits of the protein dimer) are identical and independent in terms of association constants. Model (a) also assumes that binding to the two sites is also spectroscopically independent. See text and Toptygin and Brand[76] for details. (Reproduced by permission of Academic Press, Inc.)

3.3 Membrane Proteins and Peptides

Determination of membrane organization and dynamics is one of the most challenging problems of structural biology because many methods (even of low resolution) developed for water-soluble systems are not directly applicable to membranes. For example, while the position of the fluorescence spectrum of tryptophan in a globular protein indicates a degree of exposure to the aqueous phase, it will not be sensitive to exposure to the

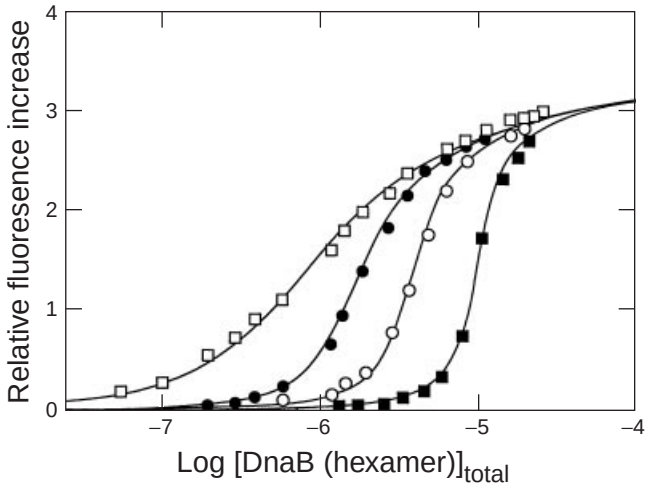

Figure 7 Example of protein–ligand interactions followed by means of fluorescence spectroscopy. The exceptional sensitivity of fluorescence is utilized in a variety of binding studies using one of the linear-response spectroscopic functions (e.g. intensity at a constant wavelength). Normally the fluorescence of the protein or ligand is altered as a result of the interaction. Jezewska and Bujalowski[78] have introduced a Macromolecule Competition Titration Method to study those interactions that do not directly result in such changes of fluorescence. Instead, the fluorescence of the reference ligand (an etheno derivative of a polynucleotide in this case) is measured in the presence of various concentrations of the ligand under study (nonlabeled polynucleotide). Fluorescence increase, plotted as a function of the protein concentration (DnaB helicase, symbols), is fitted using a binding model (solid lines) to calculate various parameters. Binding constants, parameters of cooperativity of binding and stoichiometry can be studied by fluorescence titration methods. (Reproduced by permission of the American Chemical Society.)

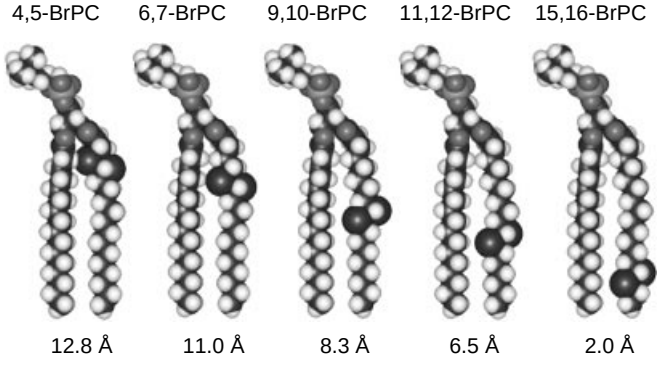

Figure 8 Set of bromolipids utilized to examine membrane penetration of proteins and peptides. Selective labeling of different positions along one of the acyl chains with bromine atoms, which are capable of quenching fluorescence via dynamic collisional mechanism,[42] allows the transverse position of the fluorophore to be determined. Two bromine atoms (large spheres) per lipid are attached in neighboring positions. This ''molecular ruler'' has been calibrated in an independent diffraction experiment yielding the distances (given in angstroms) from bromines to the bilayer center.[86] Depth-dependent fluorescence quenching profiles measured utilizing bromolipids can be evaluated with the help of distribution analysis[70,87,88] (see Figure 9 for examples) to obtain a variety of structural information.

lipid phase in a membrane-bound protein. Therefore, new approaches should be developed for structural studies in membranes. The depth-dependent fluorescence quenching technique is a useful tool to explore the structure of membrane proteins and peptides along the depth coordinate.[74,80–85] To achieve this objective it utilizes lipids with bromine atoms (Figure 8) or spin labels selectively attached to certain positions along acyl chains. This is a dynamic quenching resulting in decrease of both intensity and lifetime.[42,71]

In a depth-dependent fluorescence quenching experiment one normally determines the fluorescence intensity, F, of a probe as a function of the known depth of the quencher, h. Data are usually normalized to the intensity in the absence of quenching, F_0. Distribution analysis is one of the methods that can be used to quantitate the quenching in order to extract information on membrane penetration. It assumes that the quenching profile can be approximated by a symmetrical double-Gaussian function, which accounts for *cis*- and *trans*-leaflet quenching,[70,87] as shown in Equations (12) and (13):

$$\ln \frac{F_0}{F(h)} = G(h, \sigma, S) + G(-h, \sigma, S) \qquad (12)$$

$$G(h, \sigma, S) = \frac{S}{\sigma\sqrt{2\pi}} \exp\ -\frac{(h - h_m)^2}{2\sigma^2} \qquad (13)$$

where h_m, σ and S are the three fitting parameters: mean position, dispersion and area, respectively.

The position of the maximum represents the most probable transverse location of the fluorophore, while dispersion arises from the several broadening terms, such as finite size of the probe and the quencher and fluctuational distribution of their depth due to thermal motion. Multiple conformations will result in additional broadening of the profile. Equation (14) shows that the area under the quenching profile is a product of the inherent quenching constant, γ, determined by the nature of the quenching mechanism, excited state lifetime in the absence of quenching, τ, the degree to which the probe is exposed to a lipid phase, w, and the concentration of the quencher, C:

$$S = \gamma w \tau C \qquad (14)$$

Variation of exposure arising from the shielding of tryptophan side chains by the protein moiety provides important structural information on protein conformation

in the lipid bilayer. Equation (15) shows that the relative exposure, Ω, can be estimated as the ratio of absolute exposures of tryptophan residue in a protein, w_P, to that in a model compound (e.g. tryptophan octyl ester[42,88]), w_M:

$$\Omega = \frac{w_P}{w_M} = \frac{S_P \tau_M}{S_M \tau_P} \approx \frac{S_P Q_M}{S_M Q_P} \qquad (15)$$

If the lifetime measurements are not available, the ratio of τ values can be approximated by the ratio of quantum yields, Q, of a protein and a model compound in a nonquenching lipid membrane.

Application of the distribution analysis technique to a mutant cytochrome b_5 and to a model membrane-spanning peptide are presented in Figure 9. A single fluorophore of the membrane-binding peptide of cytochrome b_5 mutant, Trp-108, was found to be located close to the membrane interphase and its distribution along the depth coordinate is relatively narrow, suggesting the lack of conformational freedom.[70,87] The relative exposure of Trp-108 to lipid phase equals 0.66, as compared to tryptophan octyl ester. A single tryptophan in a membrane-spanning peptide is located deep in the hydrophobic core of the bilayer. For this peptide, the use of a double-Gaussian fitting function was especially important due to strong trans-leaflet quenching. Thus for this peptide the unimodal distribution of transverse position (Figure 9) differs from that of a quenching profile.

It has been demonstrated that all three parameters – average depth of the tryptophan, the width of its transverse distribution and the degree of its exposure to the lipid phase – are different in various proteins and peptides. Studies of melittin and various mutants of cytochrome b_5 and of outer membrane protein A indicate that these parameters can be altered by temperature-induced conformational change or during the kinetic insertion and folding.[70–73,88]

Another important aspect of the structural characterization of membrane proteins and peptides is determination of their topology. The critical issue is: Does a particular peptide equilibrate freely across the bilayer, form a stable transmembrane structure, or remain only on one surface? Recently Wimley and White[91] have developed a method that uses fluorescence quenching arising from resonance energy transfer for determining the topology of the tryptophan residues of peptides partitioned into phospholipid bilayer vesicles. This is accomplished through the use of a novel lyso-phospholipid quencher, lyso-methylcoumarin (LysoMC). Methylcoumarin quenches the fluorescence of membrane-bound tryptophan by the long-range resonance energy transfer mechanism with an apparent Förster distance that is comparable to the thickness of the hydrocarbon core of a lipid bilayer (~25 Å). Consequently, the methylcoumarin acceptor predominantly quenches tryptophans that reside in the same monolayer as the probe. The topology of a peptide's tryptophan in membranes can be determined by comparing the quenching in symmetric and asymmetric LysoMC-labeled vesicles (Figure 10).

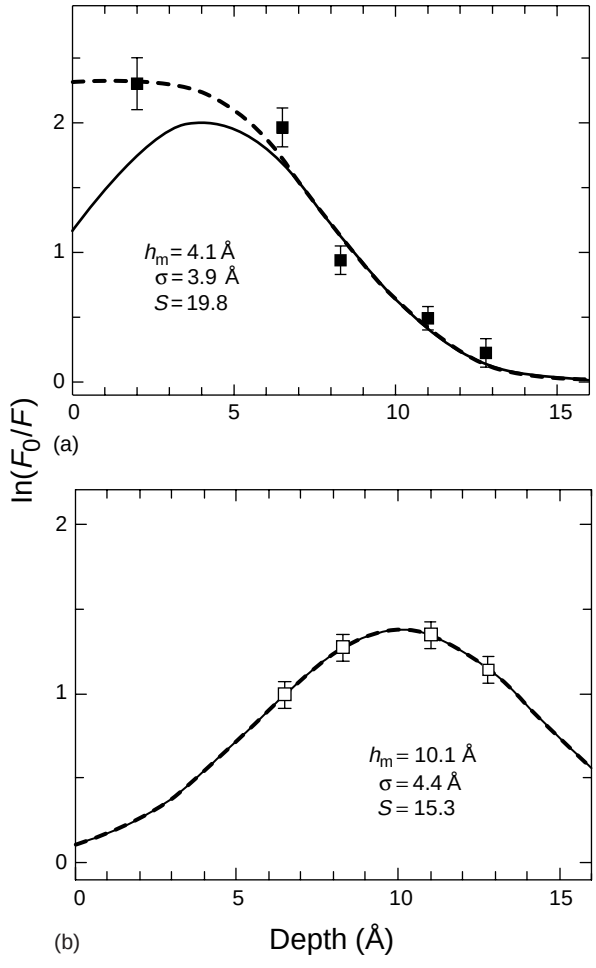

Figure 9 Examples of structural studies of membrane proteins and peptides using depth-dependent fluorescence quenching. Symbols represent quenching data for (a) membrane-spanning peptide[89] and (b) cytochrome b_5 mutant[90] obtained with a set of bromolipids (Figure 8). Dashed lines are the fitting curves of the distribution analysis[70,87,88] utilizing Equations (12) and (13). Solid lines are the individual Gaussian components with three parameters h_m, σ and S, representing average position of the fluorophore in the lipid bilayer, heterogeneity of its transverse penetration, and quenching efficiency, respectively.

Both structural and thermodynamic studies on membrane proteins are clearly lagging behind those of their soluble counterparts. Only recently have the first basic principles of integral protein structure and stability begun to emerge.[77,92] The general reason for this is that the lipid bilayer membrane appears to be too large and

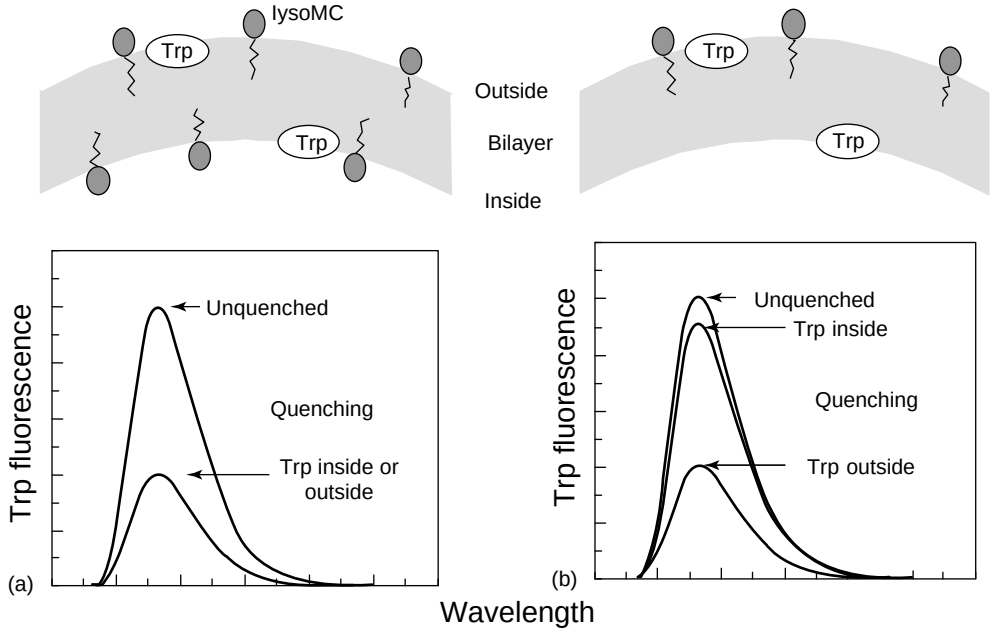

Figure 10 Schematic representation of the resonance energy transfer method of Wimley and White[91] for the determination of the topology of tryptophan in a membrane protein. LysoMC, which acts as the acceptor of electronic energy of tryptophan via long-range Förster mechanism, can be easily incorporated into large unilamellar vesicles either symmetrically (a) or asymmetrically (b). Comparison of the efficiency of energy transfer reveals the topology of the tryptophan.

slow-tumbling for solution NMR studies and too ''two-dimensional'' for X-ray crystallography. These limitations hindering applicability of high-resolution structural techniques do not compromise in principle the application of fluorescence and other spectroscopic methods (e.g. EPR[93]). Therefore, in addition to its traditional role as a technique sensing dynamic fluctuational aspects of protein structure and its kinetic metamorphoses, fluorescence spectroscopy is assuming a new role as a direct structural tool in studies of membrane proteins.

ACKNOWLEDGMENTS

I am very grateful to Drs L. Brand, W.M. Bujalowski, E.A. Burstein, A.E. Johnson, D. Toptygin, S.H. White and W.C. Wimley for providing me with the original figures from their publications, to Dr P.W. Holloway for creating an image of brominated lipids, and to Mr M.A. Myers for proofreading and editing the manuscript. During writing I was supported by grant GM-46823 (Prof S.H. White, PI) from the NIH. I would like to dedicate this contribution to Profs G. Weber, S.V. Konev, M. Kasha, L. Brand and E.A. Burstein who shaped my understanding of fluorescence phenomena by memorable conversations and by their groundbreaking research.

ABBREVIATIONS AND ACRONYMS

CD	Circular Dichroism
EPR	Electron Paramagnetic Resonance
IR	Infrared
LysoMC	Lyso-methylcoumarin
NADH	Reduced β-Nicotinamide Adenine Dinucleotide
NBD	6-(7-Nitrobenz-2-oxa-1,3-diazol-4-yl)aminohexonic Acid
NMR	Nuclear Magnetic Resonance

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Peptides and Proteins **(Volume 7)**
Electron Spin Resonance Spectroscopy Labeling in Peptide and Protein Analysis • Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Molecular Modeling in Peptide and Protein Analysis • Protein–Drug Interactions • Protein–Oligonucleotide Interactions • Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis • X-ray Crystallography of Biological Macromolecules

Electronic Absorption and Luminescence **(Volume 12)**
Detectors, Absorption and Luminescence • Fluorescence Imaging Microscopy • Fluorescence Lifetime Measurements, Applications of • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

REFERENCES

1. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publishers, New York, 1–698, 1999.
2. J.R. Lakowicz, *Topics in Fluorescence Spectroscopy: Volume 1 – Techniques*, Plenum Press, New York, 1991.
3. J.R. Lakowicz, *Topics in Fluorescence Spectroscopy: Volume 2 – Principles*, Plenum Press, New York, 1–432, 1991.
4. A.P. Demchenko, *Ultraviolet Spectroscopy of Proteins*, Springer-Verlag, Berlin, 1986.
5. E.A. Permyakov, *Luminescent Spectroscopy of Proteins*, CRC Press, Boca Raton, 1–164, 1993.
6. J.R. Lakowicz, *Topics in Fluorescence Spectroscopy: Volume 3 – Biochemical Applications*, Plenum Press, New York, 1992.
7. L. Brand, M.L. Johnson, *Fluorescence Spectroscopy*, Academic Press, San Diego, 1–628, 1997.
8. R.F. Steiner, 'Fluorescence Anisotropy: Theory and Applications', in *Topics in Fluorescence Spectroscopy: Volume 2 – Principles*, ed. J.R. Lakowicz, Plenum Press, New York, 1–52, 1991.
9. M.R. Eftink, 'Fluorescence Quenching: Theory and Applications', in *Topics in Fluorescence Spectroscopy: Volume 2 – Principles*, ed. J.R. Lakowicz, Plenum Press, New York, 53–126, 1991.
10. H.C. Cheung, 'Resonance Energy Transfer', in *Topics in Fluorescence Spectroscopy: Volume 2 – Principles*, ed. J.R. Lakowicz, Plenum Press, New York, 128–176, 1991.
11. P.K. Wolber, B.S. Hudson, 'An Analytic Solution to the Förster Energy Transfer Problem in Two Dimensions', *Biophys. J.*, **28**, 197–210 (1979).
12. P.G. Wu, L. Brand, 'Resonance Energy Transfer: Methods and Applications', *Anal. Biochem.*, **218**, 1–13 (1994).
13. I. Tatischeff, R. Klein, 'Influence of the Environment on the Excitation Wavelength Dependence of the Fluorescence Quantum Yield of Indole', *Photochem. Photobiol.*, **22**, 221–229 (1975).
14. E.A. Burstein, *Intrinsic Protein Luminescence (Origins and Applications)*, VINITI, Moscow, 1977.
15. T.L. Bushueva, E.P. Busel, E.A. Burstein, 'Relationship of Thermal Quenching of Protein Fluorescence to Intramolecular Structural Mobility', *Biochim. Biophys. Acta*, **534**, 141–152 (1978).
16. E.A. Burstein, *Luminescence of Protein Chromophores (Model Studies)*, VINITI, Moscow, 1976.
17. T.L. Bushueva, E.P. Busel, V.N. Bushuev, E.A. Burstein, 'The Interaction of Protein Functional Groups with Indole Chromophore. 1. Imidazole Group', *Stud. Biophys.*, **44**, 129–139 (1974).
18. T.L. Bushueva, E.P. Busel, E.A. Burstein, 'Interaction of Protein Functional Groups with Indole Chromophore. Temperature Dependence of the Rate of Fast Fluorescence Quenching Processes', *Stud. Biophys.*, **51**, 173–182 (1975).
19. T.L. Bushueva, E.P. Busel, E.A. Burstein, 'Interaction of Protein Functional Groups with Indole Chromophore. Amine, Amide and Thiol Groups', *Stud. Biophys.*, **52**, 41–52 (1975).
20. W.J. Colucci, L. Tilstra, M.C. Sattler, F.R. Fronczek, M.D. Barkley, 'Conformational Studies of a Constained Tryptophan Derivative: Implications for the Fluorescence Quenching Mechanism', *J. Am. Chem. Soc.*, **112**, 9182–9190 (1990).
21. Y. Chen, M.D. Barkley, 'Toward Understanding Tryptophan Fluorescence in Proteins', *Biochemistry*, **37**, 9976–9982 (1998).
22. A.S. Ladokhin, H. Malak, M.L. Johnson, J.R. Lakowicz, L. Wang, A.W. Steggles, P.W. Holloway, 'Frequency-domain Fluorescence of Mutant Cytochrome b_5', in *Time-Resolved Laser Spectroscopy in Biochemistry III*, ed. J.R. Lakowicz, SPIE, Bellingham, WA, 1–807, 1992.
23. T.L. Bushueva, E.P. Busel, E.A. Burstein, 'Some Regularities of Dynamic Accessibility of Buried Fluorescence Residues to External Quenchers in Proteins', *Arch. Biochem. Biophys.*, **204**, 161–166 (1980).
24. A.G. Szabo, D.M. Rayner, 'Fluorescence Decay of Tryptophan Conformers in Aqueous Solution', *J. Am. Chem. Soc.*, **102**, 554–563 (1980).
25. B. Valeur, G. Weber, 'Resolution of the Fluorescence Excitation Spectrum of Indole into the ^1L_a and ^1L_b Excitation Bands', *Photochem. Photobiol.*, **25**, 441–444 (1977).
26. J.B.A. Ross, A.G. Szabo, C.W.V. Hogue, 'Enhancement of Protein Spectra with Tryptophan Analogs: Fluorescence Spectroscopy of Protein–Protein and Protein–Nucleic Acid Interactions', *Methods Enzymol.*, **278**, 151 (1997).
27. C.W.V. Hogue, I. Rasquinha, A.G. Szabo, J.P. MacManus, 'A New Intrinsic Fluorescent Probe for Proteins: Biosynthetic Incorporation of 5-Hydroxytryptophan into Oncomodulin', *FEBS Lett.*, **310**, 269–272 (1992).

28. J.B. Ross, D.F. Senear, E. Waxman, B.B. Kombo, E. Rusinova, Y.T. Huang, W.R. Laws, C.A. Hasselbacher, 'Spectral Enhancement of Proteins: Biological Incorporation and Fluorescence Characterization of 5-Hydroxytryptophan in Bacteriophage λ cI Repressor', *Proc. Natl. Acad. Sci. USA*, **89**, 12 023–12 027 (1992).
29. T.G. Scott, R.D. Spencer, N.J. Leonard, G. Weber, *J. Am. Chem. Soc.*, **92**, 687–695 (1970).
30. A.J.W.G. Visser, A. van Hoek, *Photochem. Photobiol.*, **33**, 35–40 (1981).
31. A.S. Ladokhin, L. Brand, 'Evidence for an Excited-state Reaction Contributing to NADH Fluorescence', *J. Fluoresc.*, **5**, 99–106 (1995).
32. R.P. Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, Molecular Probes, Inc., Eugene, OR, 1996.
33. A.E. Johnson, 'Protein Translocation Across the ER Membrane: A Fluorescent Light at the End of the Tunnel', *Trends Biochem. Sci.*, **18**, 456–458 (1993).
34. K.S. Crowley, G.D. Reinhart, A.E. Johnson, 'The Signal Sequence Moves Through a Ribosomal Tunnel into a Noncytoplasmic Aqueous Environment at the ER Membrane Early in Translocation', *Cell*, **73**, 1101–1115 (1993).
35. B.D. Hamman, J.-C. Chen, E.E. Johnson, A.E. Johnson, 'The Aqueous Pore Through the Translocon has a Diameter of 40–60 Å During Cotranslational Protein Translocation at the ER Membrane', *Cell*, **89**, 535–544 (1997).
36. A.E. Johnson, W.R. Woodward, E. Herbert, J.R. Menninger, 'N^{ε}-Acetyllysine Transfer Ribonucleic Acid: a Biologically Active Analog of Aminoacyl Transfer Ribonucleic Acids', *Biochemistry*, **15**, 569–575 (1976).
37. E.A. Burstein, N.S. Vedenkina, M.N. Ivkova, 'Fluorescence and the Location of Tryptophan Residues in Protein Molecules', *Photochem. Photobiol.*, **18**, 263–279 (1973).
38. E.A. Burstein, 'Studies of Fast Mobility of Protein Structure Using Methods of Intrinsic Fluorescence', in *Equilibrium Dynamics of Native Structure of Protein*, ed. E.A. Burstein, ONTI NCBI, Pushchino, 60–83, 1977.
39. V.M. Grischenko, V.I. Emelyanenko, M.N. Ivkova, S.I. Bezborodova, E.A. Burstein, *Bioorg. Chem.*, **2**, 207 (1976).
40. E.A. Burstein, V.I. Emelyanenko, 'Log–Normal Description of Fluorescence Spectra of Organic Fluorophores', *Photochem. Photobiol.*, **64**, 316–320 (1996).
41. S.M. Abornev, E.A. Burstein, 'Resolution of Protein Tryptophan Fluorescence Spectra into Elementary Components', *Mol. Biol.*, **26**, 890–897 (1992).
42. A.S. Ladokhin, P.W. Holloway, 'Fluorescence of Membrane-bound Tryptophan Octyl Ester: a Model for Studying Intrinsic Fluorescence of Protein–Membrane Interactions', *Biophys. J.*, **69**, 506–517 (1995).
43. J.M. Beechem, L. Brand, 'Time-resolved Fluorescence of Proteins', *Annu. Rev. Biochem.*, **54**, 43–71 (1985).
44. J.R. Lakowicz, H. Cherek, 'Phase-sensitive Fluorescence Spectroscopy: a New Method to Resolve Fluorescence Lifetimes or Emission Spectra of Components in a Mixture of Fluorophores', *J. Biochem. Biophys. Methods*, **5**, 19–35 (1981).
45. J.R. Alcala, E. Gratton, F.G. Prendergast, 'Resolvability of Fluorescence Lifetime Distributions Using Phase Fluorimetry', *Biophys. J.*, **51**, 587–596 (1987).
46. J.R. Alcala, E. Gratton, F.G. Prendergast, 'Interpretation of Fluorescence Decays in Proteins Using Continuous Lifetime Distributions', *Biophys. J.*, **51**, 925–936 (1987).
47. A.K. Livesey, J.C. Brochon, 'Analyzing the Distribution of Decay Constants in Pulse-fluorimetry Using the Maximum Entropy Method', *Biophys. J.*, **52**, 693–706 (1987).
48. W.R. Laws, J.B. Ross, H.R. Wyssbrod, J.M. Beechem, L. Brand, J.C. Sutherland, 'Time-resolved Fluorescence and ^{1}H NMR Studies of Tyrosine and Tyrosine Analogs: Correlation of NMR-determined Rotamer Populations and Fluorescence Kinetics', *Biochemistry*, **25**, 599–607 (1986).
49. M.L. McLaughlin, M.D. Barkley, 'Time-resolved Fluorescence of Constrained Tryptophan Derivatives: Implications for Protein Fluorescence', *Methods Enzymol.*, **278**, 190 (1997).
50. L. Tilstra, M.C. Sattler, W.R. Cherry, M.D. Barkley, 'Fluorescence of a Rotationally Constrained Tryptophan Derivative, 3-Carboxyl-1,2,3,4-tetrahydro-2-carboline', *J. Am. Chem. Soc.*, **112**, 9176–9182 (1990).
51. L.P. McMahon, H.-T. Yu, M.A. Vela, G.A. Morales, L. Shui, F.R. Fronczek, M.L. McLaughlin, M.D. Barkley, 'Conformer Interconversion in the Excited State of Constrained Tryptophan Derivatives', *J. Phys. Chem. B*, **101**, 3269–3280 (1997).
52. J.R. Lakowicz, H. Cherek, 'Resolution of Heterogeneous Fluorescence from Proteins and Aromatic Amino Acids by Phase-sensitive Detection of Fluorescence', *J. Biol. Chem.*, **256**, 6348–6353 (1981).
53. J.R. Lakowicz, A. Balter, 'Direct Recording of the Initially Excited and the Solvent Relaxed Fluorescence Emission Spectra of Tryptophan by Phase Sensitive Detection of Fluorescence', *Photochem. Photobiol.*, **36**, 125–132 (1982).
54. J.R. Lakowicz, A. Balter, 'Resolution of Initially Excited and Relaxed States of Tryptophan Fluorescence by Differential-wavelength Deconvolution of Time-resolved Fluorescence Decays', *Biophys. Chem.*, **15**, 353–360 (1982).
55. A.S. Ladokhin, 'Red-edge Excitation Study of Nonexponential Fluorescence Decay of Indole in Solution and in a Protein', *J. Fluoresc.*, **9**, 1–9 (1999).
56. N.A. Nemkovich, A.N. Rubinov, V.I. Tomin, 'Inhomogeneous Broadening of Electronic Spectra of Dye Molecules in Solutions', in *Topics in Fluorescence Spectroscopy:*

Volume 2 – Principles, ed. J.R. Lakowicz, Plenum Press, New York, 367–428, 1991.
57. A.P. Demchenko, A.S. Ladokhin, 'Red-edge-excitation Fluorescence Spectroscopy of Indole and Tryptophan', *Eur. Biophys. J.*, **15**, 369–379 (1988).
58. A.P. Demchenko, A.S. Ladokhin, 'Temperature-dependent Shift of Fluorescence Spectra Without Conformational Changes in Protein: Studies of Dipole Relaxation in the Melittin Molecule', *Biochim. Biophys. Acta*, **955**, 352–360 (1988).
59. M.R. Eftink, 'The use of Fluorescence Methods to Monitor Unfolding Transitions in Proteins', *Biophys. J.*, **66**, 482–501 (1994).
60. M.R. Eftink, M.C.R. Shastry, 'Fluorescence Methods for Studying Kinetics of Protein-folding Reactions', *Methods Enzymol.*, **278**, 25 (1997).
61. O.B. Ptitsyn, R.H. Pain, G.V. Semisotnov, E. Zerovnik, O.I. Razgulyaev, 'Evidence for a Molten Globule State as a General Intermediate in Protein Folding', *FEBS Lett.*, **262**, 20–24 (1990).
62. S.E. Jackson, A.R. Fersht, 'Folding of Chymotrypsin 2. 1. Evidence for a Two-state Transition', *Biochemistry*, **30**, 10 428–10 435 (1991).
63. M. Lasagna, E. Gratton, D.M. Jameson, J.E. Brunet, 'Apohorseradish Peroxidase Unfolding and Refolding: Intrinsic Tryptophan Fluorescence Studies', *Biophys. J.*, **76**, 443–450 (1999).
64. M.C.R. Shastry, S.D. Luck, H. Roder, 'A Continuous-flow Capillary Mixing Method to Monitor Reactions on the Microsecond Timescale', *Biophys. J.*, **74**, 2714–2721 (1998).
65. G. Ramsay, M.R. Eftink, 'A Multidimensional Spectrophotometer for Monitoring Thermal Unfolding Transitions of Macromolecules', *Biophys. J.*, **66**, 516–523 (1994).
66. G. Ramsay, R. Ionescu, M.R. Eftink, 'Modified Spectrophotometer for Multidimensional Circular Dichroism/Fluorescence Data Acquisition in Titration Experiments: Application to the pH and Guanidine-HCl Induced Unfolding of Apomyoglobin', *Biophys. J.*, **69**, 701–707 (1995).
67. B.E. Jones, J.M. Beechem, C.R. Matthews, 'Local and Global Dynamics During the Folding of *Escherichia coli* Dihydrofolate Reductase by Time-resolved Fluorescence Spectroscopy', *Biochemistry*, **34**, 1867–1877 (1995).
68. O. Bilsel, L. Yang, J.A. Zitzewitz, J.M. Beechem, C.R. Matthews, 'Time-resolved Fluorescence Anisotropy Study of the Refolding Reaction of the α-Subunit of Tryptophan Synthase Reveals Nonmonotonic Behavior of the Rotational Correlation Time', *Biochemistry*, **38**, 4177–4187 (1999).
69. J.M. Beechem, 'Picosecond Fluorescence Decay Curves Collected on Millisecond Timescale: Direct Measurement of Hydrodynamic Radii, Local/Global Mobility, and Intramolecular Distances During Protein-folding Reactions', *Methods Enzymol.*, **278**, 24 (1997).
70. A.S. Ladokhin, 'Analysis of Protein and Peptide Penetration into Membranes by Depth-dependent Fluorescence Quenching: Theoretical Considerations', *Biophys. J.*, **76**, 946–955 (1999).
71. A.S. Ladokhin, P.W. Holloway, 'Fluorescence Quenching Study of Melittin–Membrane Interactions', *Ukrainian Biochem. J.*, **67**, 34–40 (1995).
72. J.H. Kleinschmidt, L.K. Tamm, 'Time-resolved Distance Determination by Tryptophan Fluorescence Quenching: Probing Intermediates in Membrane Protein Folding', *Biochemistry*, **38**, 4996–5005 (1999).
73. J.H. Kleinschmidt, T. den Blaauwen, A.J. Driessen, L.K. Tamm, 'Outer Membrane Protein A of *Escherichia coli* Inserts and Folds into Lipid Bilayers by a Concerted Mechanism', *Biochemistry*, **38**, 5006–5016 (1999).
74. S.D. Zakharov, M. Lindeberg, W.A. Cramer, 'Kinetic Description of Structural Changes Linked to Membrane Import of the Colicin E1 Channel Protein', *Biochemistry*, **38**, 11 325–11 332 (1999).
75. M.R. Eftink, 'Fluorescence Methods for Studying Equilibrium Macromolecule–Ligand Interactions', *Methods Enzymol.*, **278**, 221 (1997).
76. D. Toptygin, L. Brand, 'Analysis of Equilibrium Binding Data Obtained by Linear-response Spectroscopic Techniques', *Anal. Biochem.*, **224**, 330–338 (1995).
77. S.H. White, W.C. Wimley, A.S. Ladokhin, K. Hristova, 'Protein Folding in Membranes: Determining the Energetics of Peptide–Bilayer Interactions', *Methods Enzymol.*, **295**, 62–87 (1998).
78. M.J. Jezewska, W. Bujalowski, 'A General Method of Analysis of Ligand Binding to Competing Macromolecules using the Spectroscopic Signal Originating from a Reference Macromolecule: Application to *Escherichia coli* Replicative Helicase DnaB Protein–Nucleic Acid Interactions', *Biochemistry*, **35**, 2117–2128 (1996).
79. M.J. Jezewska, U.-S. Kim, W. Bujalowski, 'Binding of *Escherichia coli* Primary Replicative Helicase DnaB Protein to Single-stranded DNA: Long-range Allosteric Conformational Changes Within the Protein Hexamer', *Biochemistry*, **35**, 2129–2145 (1996).
80. T. Markello, A. Zlotnick, J. Everett, J. Tennyson, P.W. Holloway, 'Determination of the Topography of Cytochrome b_5 in Lipid Vesicles by Fluorescence Quenching', *Biochemistry*, **24**, 2895–2901 (1985).
81. L.A. Chung, J.D. Lear, W.F. DeGrado, 'Fluorescence Studies of the Secondary Structure and Orientation of a Model Ion Channel Peptide in Phospholipid Vesicles', *Biochemistry*, **31**, 6608–6616 (1992).
82. J.M. González-Mañas, J.H. Lakey, F. Pattus, 'Brominated Phospholipids as a Tool for Monitoring the Membrane Insertion of Colicin A', *Biochemistry*, **31**, 7294–7300 (1992).
83. A.S. Ladokhin, L. Wang, A.W. Steggles, H. Malak, P.W. Holloway, 'Fluorescence Study of a Temperature-induced Conversion from the ''Loose'' to the ''Tight'' Binding

Form of Membrane-bound Cytochrome b_5', *Biochemistry*, **32**, 6951–6956 (1993).
84. A. Chattopadhyay, E. London, 'Parallax Method for Direct Measurement of Membrane Penetration Depth Utilizing Fluorescence Quenching by Spin-labeled Phospholipids', *Biochemistry*, **26**, 39–45 (1987).
85. F.S. Abrams, E. London, 'Calibration of the Parallax Fluorescence Quenching Method for Determination of Membrane Penetration Depth: Refinement and Comparison of Quenching by Spin-labeled and Brominated Lipids', *Biochemistry*, **31**, 5312–5327 (1992).
86. T.J. McIntosh, P.W. Holloway, 'Determination of the Depth of Bromine Atoms in Bilayers Formed from Bromolipid Probes', *Biochemistry*, **26**, 1783–1788 (1987).
87. A.S. Ladokhin, 'Distribution Analysis of Depth-dependent Fluorescence Quenching in Membranes: a Practical Guide', *Methods Enzymol.*, **278**, 462–473 (1997).
88. A.S. Ladokhin, 'Evaluation of Lipid Exposure of Tryptophan Residues in Membrane Peptides and Proteins', *Anal. Biochem.*, **276**, 65–71 (1999).
89. E.J. Bolen, P.W. Holloway, 'Quenching of Tryptophan Fluorescence by Brominated Phospholipids', *Biochemistry*, **29**, 9638–9643 (1990).
90. A.S. Ladokhin, L. Wang, A.W. Steggles, P.W. Holloway, 'Fluorescence Study of a Mutant Cytochrome b_5 with a Single Tryptophan in the Membrane-binding Domain', *Biochemistry*, **30**, 10 200–10 206 (1991).
91. W.C. Wimley, S.H. White, 'Determining the Membrane Topology of Peptides by Fluorescence Quenching', *Biochemistry*, **39**, 161–170 (2000).
92. S.H. White, W.C. Wimley, 'Membrane Protein Folding and Stability: Physical Principles', *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 319–365 (1999).
93. Y.-K. Shin, C. Levinthal, F. Levinthal, W.L. Hubbell, 'Colicin E1 Binding to Membranes: Time-resolved Studies of Spin-labeled Mutant', *Science*, **259**, 960–963 (1993).

Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Heinz Fabian
Max Delbrück Center for Molecular Medicine, Berlin, Germany

Christian P. Schultz*
Institute for Biodiagnostics, Winnipeg, Canada

1 Introduction 1
2 Historical Overview 2
3 Instrumental Techniques and Sample Procedures 3
3.1 General Considerations 3
3.2 The Role of Water in Buffer Subtraction from Protein Spectra 5
3.3 Effects and Consequences of the Replacement of Water with Deuterium Oxide in Buffers 6
3.4 Attenuated Total Reflection as an Alternative to Transmission 7
3.5 Water Vapor and its Effects on Protein Analysis 8
4 Determination of Protein Secondary Structure 9
4.1 Approaches Based on Curve Fitting 9
4.2 Approaches Based on Pattern Recognition 11
4.3 Advantage of Comparing Spectra of Known Structures with Those Containing Alterations 11
5 Unfolding and Folding Events Under Equilibrium Conditions 12
5.1 Thermal Unfolding of Proteins – Irreversible Aggregation 12
5.2 The Effect of Point Mutations on the Thermal Stability of Proteins 13
5.3 Impact of Point Mutations on the Structure of Proteins Examined by Infrared Difference Spectroscopy 15
5.4 Chemically Denatured States of Proteins 15
6 Analysis of Peptides 16
6.1 Alzheimer β-Amyloid Peptides 16
6.2 Membrane-associated Peptides/Proteins 17
7 Analysis of Peptide–Protein and Protein–Protein Complexes 19
8 Sulfhydryl Group Absorptions as Local Probes of Protein Structure 20
9 Near-infrared Spectroscopy as a Tool for Monitoring Structural Changes of Proteins in Aqueous Solution 20
10 Time-resolved Fourier Transform Infrared Spectroscopy, Folding Events on the Millisecond to Minute Timescale 21
Abbreviations and Acronyms 23
Related Articles 23
References 23

Infrared (IR) spectroscopy is one of the two forms of vibrational spectroscopy, the other being Raman spectroscopy. IR spectroscopy measures absorptions of vibrating molecules and yields information about molecular structures and structural interactions. The development of computerized Fourier transform infrared (FTIR) techniques has opened up new dimensions in biological IR spectroscopy owing to the increase in achievable signal-to-noise ratios, wavenumber accuracy, and data aquisition rates, and the ability to perform measurements with strongly absorbing samples. High-quality FTIR spectra can be obtained with relative ease and rapidly with very small amounts of sample in a variety of environments. Measurements of proteins in aqueous solution are almost routine now, and can be performed under equilibrium and nonequilibrium conditions. There are many IR absorption bands characteristic of peptide groups and amino acid side-chain groups from which information on protein structures can be obtained. The information provided by FTIR spectroscopy may be a global one or highly specific for a single vibrating chemical group. In some cases, the usefulness of the method is limited by difficulties in extracting the structural information contained in the IR absorption bands.

1 INTRODUCTION

Over the last few decades, IR spectroscopy has incorporated technological advances in physics, biochemistry, engineering, computer science, and molecular biology to become an invaluable tool in structural biology. FTIR spectroscopy now provides molecular information in systems ranging from small peptides and isolated enzymes

* Bruker Optics Inc., Billerica, USA

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

to peptide–protein complexes and membrane proteins. The information can either be highly specific for a single vibrating group in an enzyme or it may report global changes, e.g. due to some experimental procedure. Furthermore, the process of obtaining structural information is not restricted to a static picture but can also be achieved in real time by applying time-resolved IR techniques. An important feature to that, the size of a protein or the nature of the environment does not limit the application of FTIR spectroscopy. Measurements can be performed in aqueous solution or organic solvents, in oriented films or dispersions, and in detergents or membrane-like environments. High-quality spectra can rapidly be aquired (seconds to minutes) and require only relatively small amounts of protein sample (as low as 10–100 μg). Advances in instrumentation and numerical data analysis today permit the quantification of protein secondary structures from IR spectra. One of the strengths of FTIR spectroscopy in this respect is the sensitivity to assess, in relative terms, the extent and nature of changes in secondary and tertiary structure of peptides and proteins. The effects of environmental factors, point mutations, or ligand binding on the structure of soluble proteins or enzymes can be examined with high sensitivity by using peptide backbone and side-chain IR bands as conformation-sensitive monitors. Changes in external factors (such as temperature) and data aquisition may be performed under computer control, thereby permitting the design of complex experiments using automatic scanning under highly reproducible conditions. In addition to the normal spectroscopic analysis, these types of experiments can also provide typical structural and standard thermodynamic parameters of proteins and enzymes.

FTIR spectroscopy has proven to be particularly valuable for studying membrane-associated peptides and proteins, since they are often difficult or impossible to investigate by other spectroscopic methods (e.g. owing to light scattering artifacts, solubility, and structural integrity). Beyond the ability to measure even large membrane proteins, IR spectroscopy also allows the study of lipids and membrane proteins simultaneously, without the use of marker molecules or extra probes. Morever, polarized IR spectroscopy on oriented samples can provide structural information on the orientation of polypeptide chains within the membrane environment.

2 HISTORICAL OVERVIEW

IR spectroscopy is one of the earliest experimental methods recognized as potentially useful in providing information on structural features of peptides and proteins. As early as 1950 it was demonstrated that a strong correlation exists between the position of certain bands in IR spectra of homopolypeptides and their secondary structure.[1] Later, these experimental observations were refined by making detailed vibrational analyses of these structure-sensitive amide bands in order to establish a correlation between the frequencies of these bands and the type of secondary structure, such as purely α-helical or purely β-sheet structures.[2,3] Nine such IR bands exist, which are assigned as amide A, amide B and amides I–VII, in order of decreasing frequency[3,4] (Table 1).

Table 1 Characteristic IR bands of the peptide linkage

Designation (amide)	Approximate frequency (cm^{-1})	Description
A	~3300	NH stretching in resonance with first amide II overtone
B	~3100	
I	1610–1695	CO stretching
II	1480–1575	NH bending and CN stretching
III	1220–1320	CN stretching and NH bending
IV	625–765	OCN bending, mixed with other modes
V	640–800	Out-of-plane NH bending
VI	535–605	Out-of-plane CO bending
VII	~200	Skeletal torsion

Of all the amide bands, the most intense and most useful for the analysis of the secondary structure of proteins was found to be the amide I band, which occurs in the region 1600–1700 cm^{-1}. In contrast to the other amide vibrations, this band reflects an almost pure vibrational character, since it consists mainly of the carbonyl stretching vibration mode of the amide group. Figure 1(a) and (b) illustrates the different IR spectroscopic characteristics of the amide I band for two proteins fundamentally different in their secondary structure. The α-helically structured myoglobin shows a relatively sharp band centered at 1654 cm^{-1} (a), whereas the antiparallel β-sheet structured concanavalin A displays an asymmetrically shaped amide I band with a maximum at 1635 cm^{-1} and a weaker feature at 1694 cm^{-1} (b).

Despite the well-recognized potential of the amide I band for conformational studies, interest in the IR spectroscopy of proteins remained limited for nearly three decades. Experimental conditions made it difficult to obtain reproducible spectra and to extract the structural information encoded in the IR bands. Major drawbacks that prevented widespread use of IR spectroscopy in the past were (i) the poor sensitivity and reproducibility of the classical dispersive IR spectrometers, which recorded spectra with low signal-to-noise ratio and low

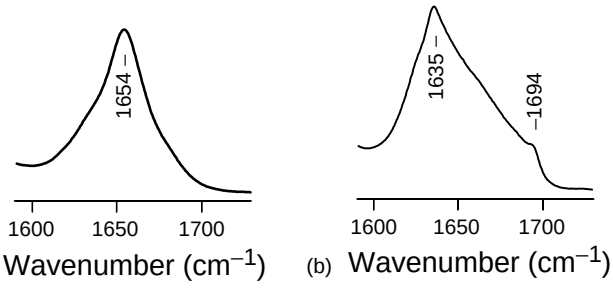

Figure 1 The amide I band of the α-helical protein myoglobin (a) and the β-sheet protein concanavalin A (b), as measured by IR spectroscopy in H_2O.

wavenumber accuracy, (ii) strongly interfering absorptions from the surrounding media or solvents such as water, and the difficulty of completely subtracting their contributions from the protein spectrum to reveal the weak IR bands of the protein, and (iii) the difficulty of extracting structural information contained in the IR amide bands. The first two constraints have largely been overcome by the development of FTIR spectrometers, which are far superior to dispersive IR instruments in their high energy throughput, achievable signal-to-noise ratio ($>10\,000:1$), wavenumber accuracy ($0.01\,cm^{-1}$), and data processing capabilities. The third difficulty, however, is of a more fundamental nature. Since the polypeptide backbone of proteins is normally folded in more than one conformation, amide bands of these proteins result from the superposition of bands corresponding to the different types of secondary structure such as helical, sheet, and unordered/loop structures. Moreover, the spectral features of the amide bands also vary slightly for each type of secondary structure, depending on environment, structural distortion, and domain interactions. All this illustrates the extraordinarily high conformational sensitivity of these amide modes as being specific for each protein, but it also creates at the same time the difficulty that these individual band components cannot easily be classified and quantified in more general terms of secondary structure.

In the last decade, different methods for determining the secondary structural composition of a protein by mathematical manipulation of the experimental data have been developed, each of them with certain weaknesses[5] as described later in this article. In contrast to these indirect methods, a new, different, and very promising direct approach has been applied, which is based on the introduction of isotopic labeling of specific sites of the polypeptide chain.[6] This allows FTIR spectroscopy to locate a particular secondary structure within the polypeptide chain and helps in analyzing conformational changes that originate exclusively from the labeled site. Since most of the IR labels have no structural consequences [similar to ^{13}C-labeling experiments in nuclear magnetic resonance (NMR) spectroscopy] and isotopic replacement is relatively easily achieved by chemical synthesis of peptides and is biosynthetically feasible for proteins, the behavior of single amino acids or even atoms can be monitored within intact peptides or proteins. This information can be obtained in real time by applying different time-resolved FTIR techniques, which closes the gap between this methodology and visible and fluorescence spectroscopy.

3 INSTRUMENTAL TECHNIQUES AND SAMPLE PROCEDURES

3.1 General Considerations

Water (as one of the strongest IR-absorbing molecules) restricts the measurement of IR bands of molecules in biological water-containing systems if they occur in the range between 1600–1700 and 3100–3700 cm^{-1} where water shows the strongest IR bands. Most of the early experiments were therefore limited to sample preparations in which the water was partially reduced or completely removed. Samples were prepared as dehydrated protein films deposited on IR-transparent substrates or as solid protein powders ground and pressed into potassium bromide pellets. These experiments should be regarded as good preliminary tests, but were of only limited relevance due to the nonphysiological nature of the measurements. Today's sample preparations aim for the measurement of molecules in buffer solutions, which is permitted by the described inherent advantages of FTIR spectroscopy.

To record a spectrum of a sample in H_2O buffer solution, very short pathlength cells of 3–8 µm are required, to prevent total IR absorption in the spectral regions of the water. According to the Lambert–Beer law, such short pathlengths limit the intensities of the IR bands and the signal-to-noise ratio at a given sample concentration. Consequently, relatively high sample concentrations ($>10\,mg\,mL^{-1}$) are required for the measurement. To obtain the spectrum of a sample, digital subtraction of solvent/buffer absorptions from the spectrum of the sample is required. For appropriate subtraction, the spectrum of the solvent/buffer should be recorded under conditions (such as temperature, ionic strength, pH, number of scans, resolution, etc.) identical with those of the sample spectrum. In practice, identical physico-chemical parameters are very important for this procedure since variations will certainly cause changes to spectral features of the H_2O bands, preventing an ideal subtraction of the spectral buffer contributions.

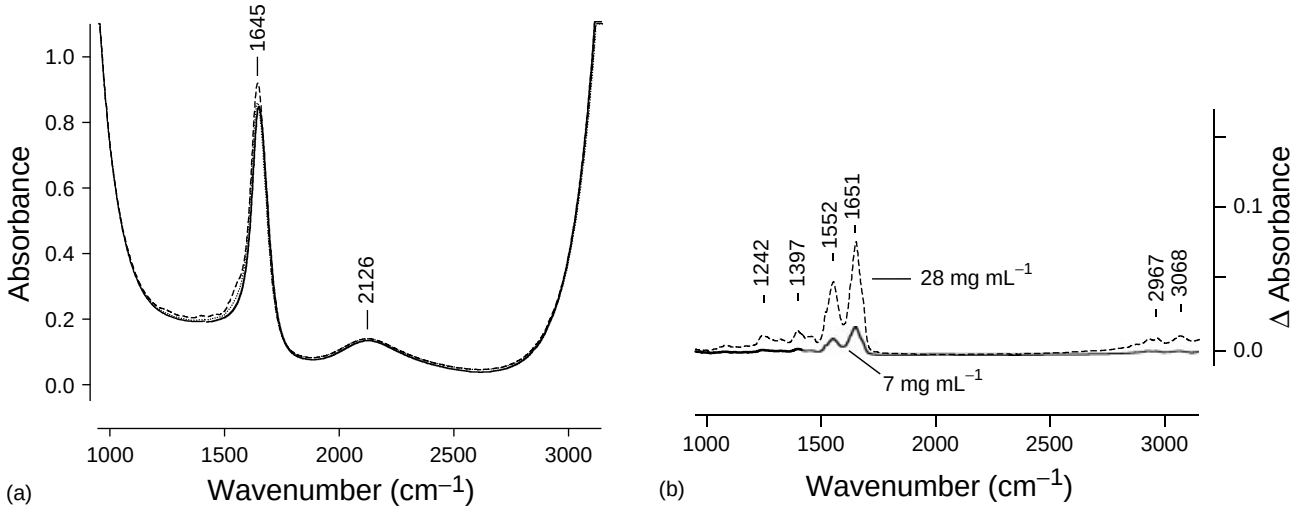

Figure 2 (a) IR spectra of ribonuclease (RNase) T1 in H_2O buffer at a protein concentration of 28 mg mL^{-1} (dashed line) and 7 mg mL^{-1} (dotted line), together with the buffer spectrum (solid line). (b) IR spectra of RNase T1 after subtraction of the buffer spectrum.

For instance, the temperatures of the sample in aqueous solution and the solvent should coincide within 0.1 °C in order to avoid artifacts caused by temperature differences.

Figure 2(a) shows IR spectra of the enzyme RNase T1 in H_2O buffer at two different concentrations. For these transmission measurements, a micro drop of 2 µL of the enzyme solution (7 or 28 mg mL^{-1}) was placed between a pair of calcium fluoride (CaF_2) windows separated by a pathlength of 8 µm. The solvent spectrum (Figure 2a, solid line) was measured in an optimally matched second cell of slighly reduced pathlength, which takes into account the slightly lower water concentration in the protein sample measured. Among the IR window materials available for experiments in solution, CaF_2 is the most common because (i) it has a low refractive index which is similar to that of water (high light throughput, low tendency for interference fringe pattern in the spectra); (ii) it is relatively rugged; and (iii) it is transparent to IR light (>1000 cm^{-1}) as far as the ultraviolet (UV) region of the spectrum. Barium fluoride (BaF_2), a similar window material, has a lower spectral cutoff than CaF_2 (at ~800 cm^{-1}), enabling additional IR spectral bands to be observed, but it is less rugged than CaF_2 and significantly more soluble in aqueous solution. Others [such as ZnSe, thallium iodide (KRS-5), AgCl, or Irtran] are characterized by unfortunately high refractive indices, which results in major reflection losses and persistent fringing in the spectra.

Although CaF_2 is the most suitable window material for protein measurements, it is not an ideal window material for long-term measurements of calcium-binding proteins. The solubility of CaF_2 in water is very low, but possible contamination of the sample by dissolution of Ca^{2+} from the CaF_2 window cannot completely be excluded. In direct consequence, the Ca^{2+}-free form of a sample may not persist during the collection of spectra over longer timescales. Nevertheless, flow-through demountable cells with Luer-lock fittings, CaF_2 windows and spacers covering pathlengths from 6 to 200 µm are often used for protein measurements. These cells are available from virtually any IR accessories supplier, but reveal some disadvantages in practice such as being difficult to clean, accidental injection of gas bubbles and the application of high pressure in the case of short pathlength cells.

To circumvent these problems, our design of a custom-made IR cell consists of a flat cover disk (w_1, typically made of CaF_2) and a second disk of the same material (w_2), with the center hollowed out to form a recessed parallel surface surrounded by a groove (see Figure 3). The groove (gr) prevents direct contact of the sample with the outer part of the disk. Pressing the cover disk (w_1) on to the sample disk (w_2) is sufficient to prevent the evaporation of water for many hours at room temperature and seals the cell. Two spacer discs (g_1 and g_2) e.g. made of lead or graphite are used as gaskets to fit the windows into the base mounting (h_2), which is then sealed with the lid mounting (h_1). For measurements at higher temperatures and/or for long-term experiments, the sealing surface of the disks is lubricated with mineral oil dissolved in a volatile organic solvent prior to filling and assembling the cell. This prevents both evaporation of the solvent (e.g. water) at high temperatures (e.g. 95 °C), for up to several weeks, or changes of the isotope content when working with deuterium oxide

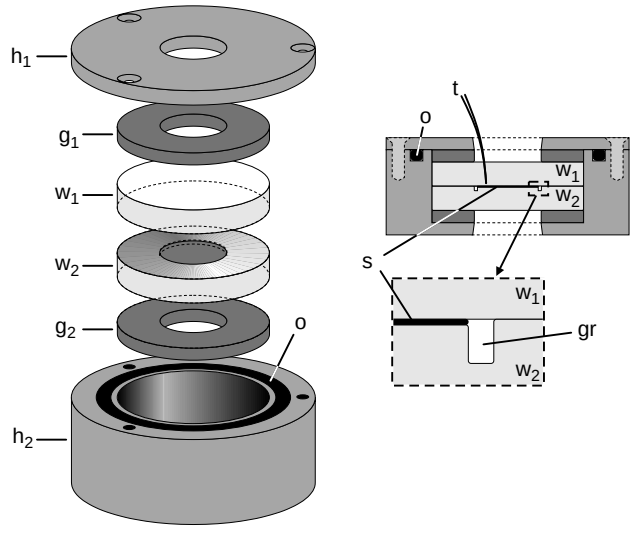

Figure 3 Schematic representation of our custom-made IR cell used to record spectra from samples in aqueous solutions or as dispersions.

(D_2O) solutions. For temperature control, the gas-tight IR cell is placed into a jacket through which heating or cooling liquid from an external bath circulates. The temperature is monitored at the jacket/cell interface with a thermocouple (t).

Depending upon the diameter and the depth of the recessed surface of the window (i.e. the pathlength of the cell), only a few microliters are required to fill the cell. Moreover, this type of cell can very easily be filled with solution, assembled and disassembled, cleaned between measurements, and provides a constant pathlength, which is very difficult to achieve with conventional tin or polyethylene spacers. The pathlength of the cell can be estimated very precisely by measuring the phase difference in interference fringes caused by internal light reflection of the window surfaces in the empty cell. The relationship between the pathlength of the cell (in centimeters) and the peak-to-peak fringes (ripples) in the spectrum is given by Equation (1):

$$\text{pathlength} = \frac{n}{2}(\nu_1 - \nu_2) \qquad (1)$$

where n represents the number of complete peak-to-peak fringes between two maxima (or minima) at the wavenumbers ν_1 and ν_2 (in cm^{-1}). From a practical point of view, it is interesting to note that all IR transmission cells, if made from suitable high-quality window material, can also be used for measurements in the visible to far-UV spectral region. This allows a variety of measurements such as IR, circular dichroism (CD), and absorption spectra, of the same sample prepared in the same sampling cell.

3.2 The Role of Water in Buffer Subtraction from Protein Spectra

The dominance of water in the spectra of proteins in water-buffered solutions is demonstrated in Figure 2(a), comparing the IR spectra of RNase T1 with a spectrum of the H_2O-based buffer solution applied in the experiment. Whereas some additional spectral features can be seen in the spectrum of the concentrated protein sample ($28\,mg\,mL^{-1}$), the original spectrum of the protein at a lower concentration of $7\,mg\,mL^{-1}$ (dotted line) is practically indistinguishable from the spectrum of the buffer (solid line). The actual protein bands can only be seen after subtracting the H_2O buffer spectrum from the protein spectrum when measured under identical conditions (Figure 2b). A fourfold increase in protein concentration consistently results in a difference spectrum of higher spectral quality, in which all of the major absorptions expected in the IR spectra of proteins can be identified, but not all proteins tolerate this relatively high concentration without self-association or aggregation.

The precision of the subtraction of water from a protein spectrum requires a reference water band that does not overlap with those of the sample. Often, the combination band of water at around $2126\,cm^{-1}$ is used to ensure proper subtraction of water, which then results ideally in a straight baseline between 1900 and $2400\,cm^{-1}$. If positive or negative features are visible in the $2100\,cm^{-1}$ region of the water-subtracted spectrum, incomplete or overcompensation of the water features is indicated. If derivative-type artifacts are observed in this region (positive and negative bands), this implies that the spectroscopic characteristics of the water in the sample water are not identical with those of the reference water. The latter artefact may often be observed at very high protein concentrations ($>50\,mg\,mL^{-1}$), since the interaction of a protein with water can cause the creation of hydration shells around the protein and therefore modify the water vibrations and hence the shape of the water bands in the IR spectrum. Consequently, the reference spectrum of the H_2O buffer cannot completely match the spectral conditions of the water in the protein solution (containing bulk water plus protein-bound water). While the weak water band at $2126\,cm^{-1}$ (which is more than six times less intense than the water band of interest at around $1645\,cm^{-1}$) may serve as a good approximation to subtract the water features interactively, it is necessary to perform the final water subtraction based on a different spectral region with stronger water absorptions. The side lobes of the strong water band of the stretching vibrations (range $3000–3700\,cm^{-1}$) are ideal for this subtraction.

Peptides or proteins have only very weak IR bands in the 3000–3100 cm^{-1} region (see Figure 2b). Residual absorbance in the difference spectrum at 3080–3100 or ~3650 cm^{-1} may serve as a very sensitive monitor for incomplete water compensation.

3.3 Effects and Consequences of the Replacement of Water with Deuterium Oxide in Buffers

One way of overcoming the problems related to the strong water absorbance in the amide I region is to replace H_2O in the buffer with D_2O. The IR bands of D_2O occur at lower wavenumbers than those of H_2O, owing to the downshifted vibrations of the almost twice as heavy deuterium atoms. This isotopic effect creates a region of relatively low absorbance between 1400 and 1800 cm^{-1}, an ideal window for observing the weak IR bands of the dissolved peptide or protein. Much longer pathlengths of 40–80 μm may then be used, hence much lower sample concentrations are required to obtain high-quality spectra, which is of particular practical interest for proteins of low solubility. Figure 4(a) shows IR spectra of RNase T1 in D_2O buffer at protein concentrations of 14 (dashed line) and 1.4 mg mL^{-1} (dotted line), together with the spectrum of the D_2O buffer (solid line). The buffer-subtracted spectrum (Figure 4b) illustrates that a spectrum with a good signal-to-noise ratio may be obtained from 10 μL of a 1.4 mg mL^{-1} protein solution prepared in an IR transmission cell with a 45 μm pathlength. Measurements of lower protein concentrations are feasible, but below values of approximately 0.5–1 mg mL^{-1} the noise in the spectra becomes too great to enable some of the approaches described later in this article to be applied. A specific feature of protein studies in D_2O is the exchange of amide protons into deuterons, which in consequence causes changes in the spectrum after a peptide or protein has been introduced into D_2O. The most dramatic change is a substantial decrease in the intensity of the amide II band (predominantly originating from peptide N–H bending modes, coupled with C–N stretching modes) which is centered near 1545 cm^{-1} in H_2O. This band shifts down in frequency by nearly 100 cm^{-1} to 1450 cm^{-1} (amide II′) in D_2O, thereby revealing side-chain vibration bands of the protein such as those of arginine, tyrosine, aspartic acid and glutamic acid in the range 1490–1610 cm^{-1}. The intensity changes of the amide II band of a protein in D_2O can be efficiently used to monitor hydrogen–deuterium exchange, providing valuable information concerning the structure and flexibility of a protein.

Compared with the large effects on the amide II band, the shift of the nearby amide I band upon deuteration of the backbone hydrogens (labeled amide I′ by convention) is relatively small (5–10 cm^{-1}). This can be explained by the fact that this band reflects mainly the carbonyl stretching vibration of the amide group and is therefore less sensitive to the H–D exchange. Nevertheless, individual spectral components of the amide I band reveal different exchange kinetics. Amide protons exposed to the solvent and involved in irregular or turn conformations exchange rapidly, whereas NH groups in stable secondary structures are typically resistant to exchange even after prolonged exposure to D_2O. This greatly assists the assignment of absorption bands arising from different secondary structural classes. Despite the positive aspects, it can also complicate the interpretation of the amide I′ region if a protein cannot be completely exchanged. In order to ensure that the spectral changes observed arise solely from changes in secondary structure, it is necessary to exchange completely all peptide protons

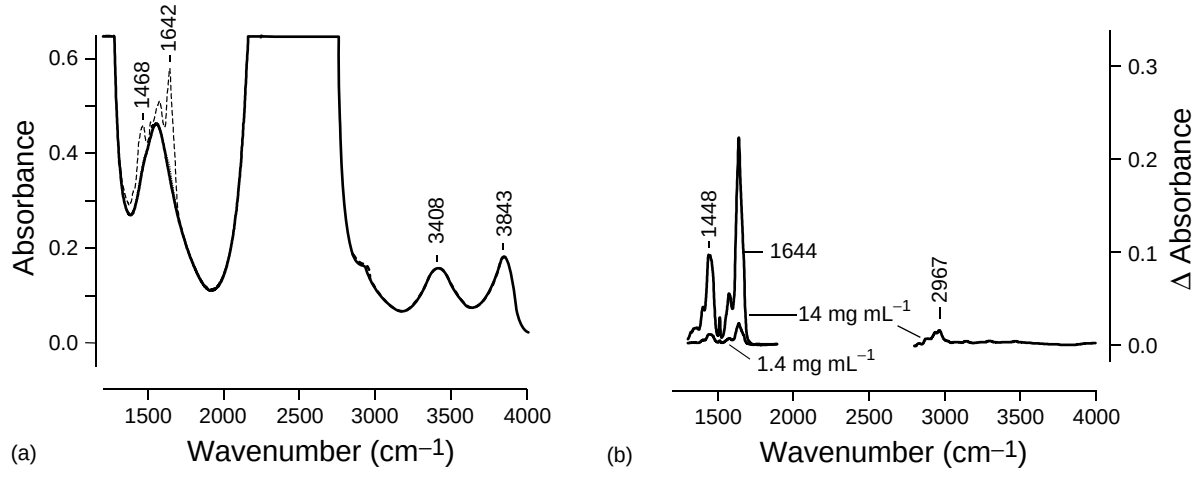

Figure 4 (a) IR spectra of RNase T1 in D_2O buffer at a protein concentration of 14 mg mL^{-1} (dashed line) and 1.4 mg mL^{-1} (dotted line), together with the buffer spectrum (solid line). (b) IR spectra of RNase T1 after subtraction of the buffer spectrum.

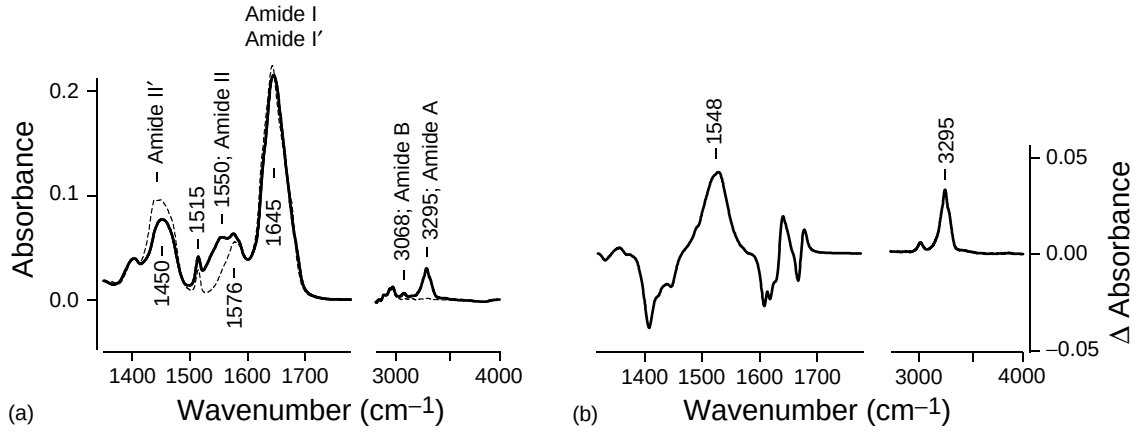

Figure 5 (a) IR spectra (after subtraction of the buffer spectrum) of RNase T1 3 h after dissolution in D_2O (solid line) and after complete H−D exchange (dashed line). (b) Difference spectrum obtained by subtraction of the spectrum after complete H−D exchange from the spectrum of the partially exchanged protein.

of the protein backbone before any conformational changes can be investigated in detail. While complete H−D exchange is often attainable in peptides, it typically requires considerable effort in the case of proteins, and often it cannot be achieved at all without irreversible alterations to the structure (e.g. causing aggregation). Figure 5(a) and (b) summarizes the spectral effects of partial and complete deuterium exchange on the amide bands of a protein spectrum. The solid line in Figure 5(a) represents the IR spectrum of the small globular protein RNase T1 that has been allowed to exchange in D_2O buffer for 3 h at room temperature. The residual intensity in the amide II region at about 1550 cm^{-1}, together with the presence of the amide A band (N−H stretching vibrations of the peptide groups) centered at 3295 cm^{-1}, indicate that a significant number of the amide protons are not exchanged after 3 h of exposure to D_2O. This situation is common in many proteins.

The amide A band is the best indicator for residual nonexchanged N−H groups owing to the lack of other protein absorptions in the range 3200–3300 cm^{-1}. The same assessment cannot easily be made based on the residual intensity in the amide II region, since IR bands arising from amino acid side-chain groups overlap with the remaining amide II band, e.g. in RNase T1 the two bands are at 1515 and 1576 cm^{-1}. Complete deuteration of RNase T1 was achieved by keeping the protein solution close to the denaturation temperature for 10 min before cooling back down to room temperature (Figure 5a, broken line). The thermal unfolding of RNase T1 is known to be fully reversible, whereas other proteins may start to aggregate irreversibly upon thermal denaturation. In most cases, keeping the temperature 10 °C below the denaturation temperature accelerates the exchange without changing the protein structure. It also seems to be possible to destabilize the protein structure slightly by changing the pH and/or adding chemical denaturants to achieve the accelerated exchange. In any case however, it is necessary to demonstrate that the pretreatment does not irreversibly alter the native protein structure.

3.4 Attenuated Total Reflection as an Alternative to Transmission

Another sampling technique widely used for obtaining IR spectra of biological systems is attenuated total reflection (ATR).[7,8] For ATR measurements the sample is prepared on the surface of an IR-transparent medium of high refractive index, usually a zinc selenide, diamond, silicon, or germanium crystal. The IR beam is guided through the crystal in such a way that one, three or several total reflections take place at the surface. Since the IR beam penetrates slightly into the surrounding medium, the deposition of an IR absorber on the crystal surface causes the IR light to be partially absorbed. In this way, a spectrum of the sample may be recorded. The penetration depth of the IR radiation in this arrangement is strictly dependent on the wavelength and the IR spectrum measured therefore contains only information on a very thin layer of the sample that is in close proximity to the surface of the crystal. This allows a spectrum of a protein in H_2O solution to be obtained relatively easily, without much interference from IR absorption of the bulk water.

Specially designed temperature-jacketed ATR cell configurations are available, which also permit experiments under temperature control. Surface adsorption, however, may significantly change the secondary structure of the proteins molecules, which are in direct contact with the crystal. Although the contribution of those moleclues to the total absorbance measured may be small, one should

proceed with caution in structural studies of water-soluble proteins by ATR spectroscopy.

Since the amount of sample required to obtain spectra with adequate signal-to-noise ratios by ATR techniques is similar to that required for the transmission experiments described previously, there seems to be no real advantage to using the ATR IR approach for the analysis of water-soluble proteins. ATR techniques, however, are particularly well suited for the study of membrane-associated peptides or proteins. Here, strong adhesion of a membrane film on the ATR crystal may be of advantage. Studies of immobilized samples oriented on the surface of an ATR crystal, such as in membrane-oriented proteins, may be performed with polarized IR light (IR dichroism), providing information on the spatial orientations of functional groups. Compared with classical IR dichroism measurements in transmission mode, the number of layers deposited on the crystal may be much lower, usually resulting in a more uniform orientation. Furthermore, sophisticated ATR techniques allow periodic stimulation of the membrane system by modulation of an external parameter, such as temperature, concentration, or electric field. This results in periodic modulation of only those absorption bands which are related to molecules, or parts of them, that are affected by the external perturbation,[9] and even minor spectral changes within the large background of the total absorbance can be detected.

3.5 Water Vapor and its Effects on Protein Analysis

Another general sampling consideration is that of water vapor and carbon dioxide in the instrument atmosphere. Water vapor and carbon dioxide absorb strongly, and it is necessary to purge the instrument with dry air or nitrogen. This will drastically reduce their level, but it is almost impossible to remove all water vapor and carbon

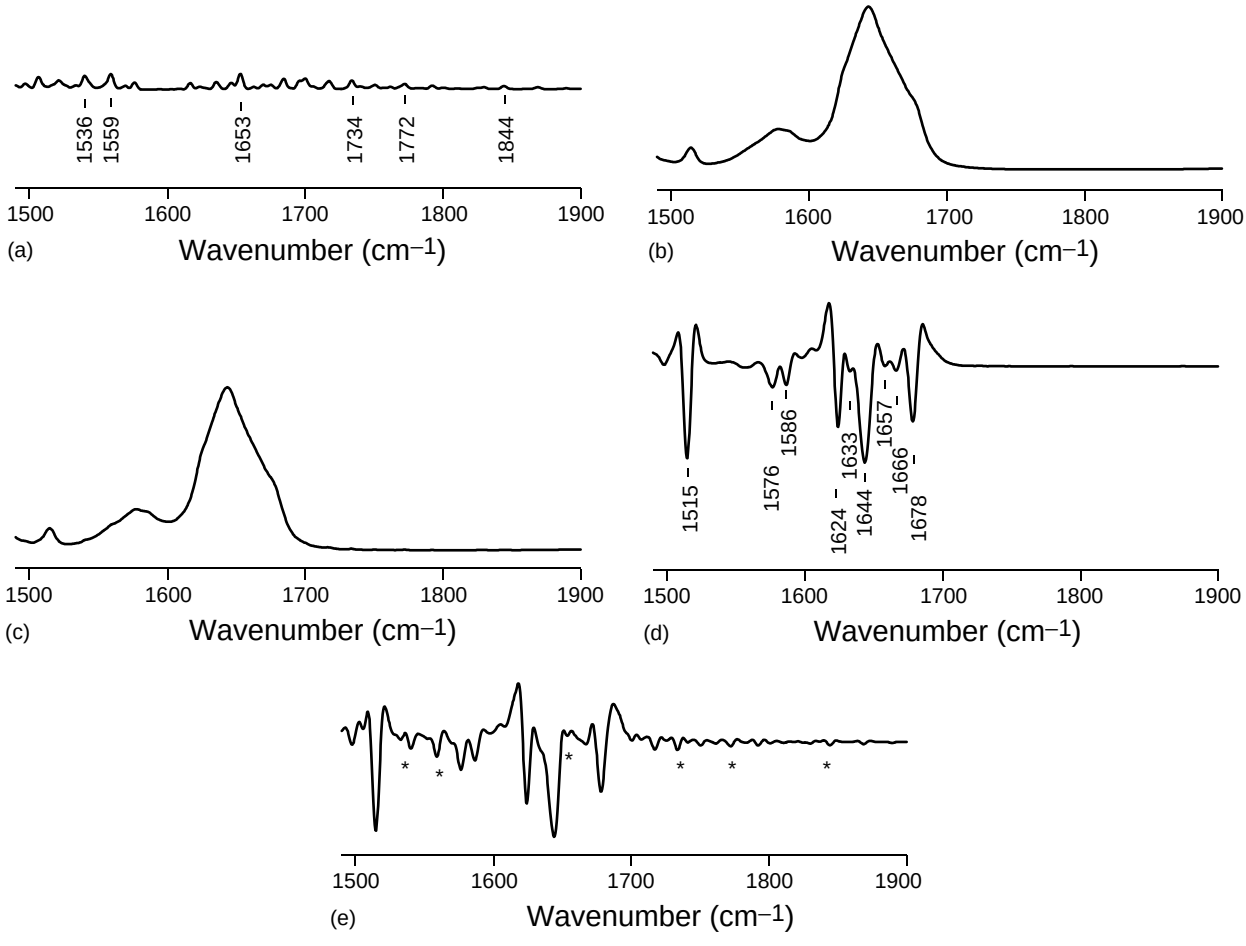

Figure 6 Effect of improper water vapor subtraction on the IR spectra of RNase T1 in D_2O buffer. (a) Water vapor obtained by subtracting two buffer spectra collected at different instrument purge levels; (b) spectrum of the protein at correct level of water vapor subtraction; (c) spectrum of the protein with water vapor slightly undersubtracted; (d) second derivative of the protein spectrum shown in (b); (e) second derivative of the slightly undercompensated spectrum shown in (c). Some of the peaks that result from undersubtraction of water vapor are marked with asterisks.

dioxide no matter how long the spectrometer is purged. In addition, the level always changes when the sample chamber is opened up and the effectiveness of the purging of the instrument varies during the measurements. It is therefore convenient to record spectra of water vapor at low but well-matched levels for the sample and the reference, rather than waiting for the final purge level of the system between different experiments. A sample shuttle that allows the background to be signal averaged almost concurrently with the sample is instrumentally very helpful. Residual water vapor can then simply be subtracted from the sample spectrum by using prerecorded water vapor spectra. This is mandatory for peptide or protein measurements, since the narrow water vapor bands overlap with the conformation-sensitive amide I/II bands (see Figure 6a–e). For aqueous solutions, subtracting two buffer spectra from the same cell collected at different instrument purge levels (Figure 6a) should generate the correct reference water vapor spectrum. Any over- or under-subtraction of water vapor can best be visualized by calculating the second-derivatives of the spectra, which enhances narrow bands in particular (Figure 6d and e). The subtraction factor must be varied until the second-derivative spectrum is featureless in the range 1850–1750 cm^{-1}, which is normally free of any bands due to the lack of functional groups in the sample molecule studied (see Figure 6d).

4 DETERMINATION OF PROTEIN SECONDARY STRUCTURE

Specific information on the secondary structure of peptides or proteins is obtained from the analysis of the various conformation-sensitive amide bands. Some of these bands are more useful than others for conformational studies. By far the best characterized in this respect is the amide I/I′ band, which represents primarily the C=O stretching vibrations of the amide groups (coupled to in-plane bending of the N–H and stretching of the C–N bonds) and gives rise to IR band(s) between approximately 1600 and 1700 cm^{-1}. The amide I/I′ bands (I stands for the band measured in water and I′ the band measured in D_2O) are established indicators of the protein secondary structure because of their sensitivity to hydrogen-bonding pattern, dipole–dipole interaction and the geometry of the peptide backbone. Typically, the amide I/I′ band of proteins or peptides consists of a series of overlapped component bands which occur as a result of the secondary structures present in such molecules. As a consequence, the individual component bands that represent different structural elements, such as α-helices, β-sheets, turns, and irregular structures, are often not resolved and difficult to identify in the broad amide I/I′ band contour of the experimentally measured spectra. Two approaches are currently used for the quantitative estimation of protein secondary structure from FTIR spectra: those which involve curve fitting of the amide I/I′ band profiles and those which are based on pattern recognition algorithms.

4.1 Approaches Based on Curve Fitting

Curve-fitting analysis of amide I/I′ band profiles starts with the choice of input parameters, such as the number of component bands and their positions. Here, techniques for band narrowing such as Fourier deconvolution or derivative spectroscopy are very useful and necessary tools.[10] Fourier deconvolution decreases the widths of IR bands and is therefore often referred to as a resolution enhancement technique, although it does not increase the instrumental resolution. It significantly improves the degree to which the individual component bands can be resolved, whereby the relative integrated intensities are maintained. Visualization of overlapping bands can also be achieved by calculating the *n*th derivative of the spectrum. Often the second derivative is calculated, which gives a negative peak for every band or shoulder in the spectrum. Because sharp bands are enhanced at the expense of broad ones, derivatization does not preserve the integrated areas of individual components.

Band-narrowing techniques greatly amplify features in the spectra originating from random noise and/or uncompensated water vapor. Therefore, Fourier deconvolution and derivation should only be performed on spectra with a very high signal-to-noise ratio (preferably better than 5000 : 1) and after elimination of water vapor bands. The estimated number of component bands plus their approximated width, height, and shape are then used as input parameters in an iterative least-squares procedure that attempts to reproduce the measured amide I/I′ band profile by varying these parameters. For practical reasons, self-deconvoluted spectra should be subjected to curve fitting, because least-squares algorithms work significantly more reliably on spectra with an enhanced profile.

Good-quality fitting is accomplished when different starting conditions generate the same fitting and a difference spectrum between the fitted curve and original spectrum appears to be featureless (within 1% of the total intensity at any position in the spectrum). When a reasonable fit is obtained, the fractional areas of the fitted components are taken as directly proportional to the relative quantities of structural elements they represent. The percentages of different secondary structure elements are then estimated by adding the areas of all component bands assigned to each of these structures and

expressing the sum as a fraction of the total amide I/I′ band area.

This curve-fitting approach provided estimates of the secondary structure of about 20 selected proteins, which correlated well with the corresponding values based on X-ray data for these proteins.[11]

4.1.1 *Critical Aspects in Defining Typical Bands for Secondary Structure*

Further studies revealed that this approach (like all curve-fitting applications) shows some significant inherent problems.[5,12] First, it is always assumed that the number of component bands estimated by self-deconvolution or derivatization reflects the real number of components. In cases where bands significantly overlap, even the applied band-narrowing procedures will certainly fail to separate the components present. This is especially critical in cases in which these components describe different types of secondary structure. It is therefore not surprising that it is sometimes necessary to include extra component band(s) not identified in the spectra after band narrowing, in order to achieve agreement between IR estimates of secondary structure and structure data provided by the X-ray crystallography. In this context, it is important to be aware that the structural information obtained from crystal structures is based on the hydrogen-bonding pattern observed, whereas the IR spectroscopic information is based on the vibrational behavior of the bonds involved. The middle and the end of an α-helix, for example, show slightly different spectral components, while the definition of a helix in a crystal structure does not consider these effects.

NMR spectroscopy on peptides and proteins in solution confirms the findings of inhomogeneous secondary structure elements as described for α-helices. Nevertheless, the number of band components used in the curve-fitting analysis is therefore always somewhat arbitrary. Another critical step is the assignment of the component bands. The assignment is normally guided by theoretical calculations and by emperical spectral–structural correlations experimentally established for model polypeptides and proteins of known three-dimensional structure. The data in Table 2 can be taken as a general guideline for the approximation of protein secondary structure[5,8,11–13] based on their characteristic amide I′ frequencies. They illustrate that amide bands arising from α-helical and irregular structures appear in the same spectral range, while parallel and antiparallel β-sheet structures have relatively distinct features. It is obvious that in some cases overlapping bands of α-helical and irregular structures significantly complicate the assignment of these structures to their band components. If a protein, for example, contains both larger quantitites of irregular structures such as extended loops and many different longer and shorter α-helices, additional information should be obtained from CD data, since this methodology can distinguish and quantify these two structural elements. The situation is different in the case of β-sheet structures. Theoretical calculations[3] suggest that it might be possible to distinguish antiparallel β-sheet structures from their parallel counterparts, since the latter lack the high-frequency band component characteristic of antiparallel allignment of β-strands. Few attempts to verify this suggestion experimentally have been made owing to the lack of model structures for parallel β-sheet conformations. Turns are associated with various bands between 1660 and 1690 cm^{-1}, β-turns in particular with a band around 1641 cm^{-1}. The complexity in the amide I region of a protein creates difficulties in unique assignment of an IR band component to a specific type of turn structure, although well-defined bands in IR spectra of defined turns in model structures exist.

The band assignment of 3_{10}-helices is still uncertain but two bands at around 1640 and 1662 cm^{-1} are under consideration. Some proteins contain secondary structures that absorb outside the frequency ranges shown in Table 2. In membrane proteins, α-helices appear to absorb in the range 1658–1662 cm^{-1}, and weak component bands below 1640 cm^{-1} have been observed in α-helical proteins known to contain no β-sheet structure. Such variations may result from an unusual degree of solvent–protein interactions[14] (e.g. in bifurcated hydrogen bonds), from distortion of structural elements, or from other factors.

Table 2 Amide I′ frequencies for protein secondary structures

Structure	Amide I′ frequency (cm^{-1})
Aggregated strands	~1615
β-Sheet	1620–1635
Irregular	1640–1650
3_{10}-Helix	~1640
α-Helix	1650–1658
3_{10}-Helix	~1660
Turns and loops	1655–1685
β-Sheet (antiparallel)	1675–1695
Aggregated strands	~1685

4.1.2 *Comparability of the Bands Representing Different Secondary Structures*

Another assumption in the curve-fitting approach is that the molar absorptivities (integrated IR intensities) of the bands associated with different secondary structural elements are identical. This assumption is at its best a rough approximation. Studies on polypeptides and proteins have revealed that the integrated area of the amide I/I′ absorption may vary as the structure of the

polypeptide backbone is altered, indicating a change in molar absorptivities. Moreover, some amino acid side-chains also display absorption bands in the amide I/I′ spectral region, which in some proteins may account for as much as 15–20% of the total integrated intensity in this region. Since their spectral contributions are not balanced in the amide I/I′ region, a subtraction of the side-chain contributions from the experimental protein spectrum would be desirable. It has been suggested that the IR spectra of pure amino acids or simple peptides may be used for this purpose.[15,16] The spectral parameters of the side-chain absorption bands in the model compounds, however, provide only qualitative information since the spectral features of the side-chain groups in a protein are influenced by the specific microenvironment of the corresponding group.

4.2 Approaches Based on Pattern Recognition

Quite different approaches to estimate the secondary structure of a protein are pattern recognition methods. These methods use IR spectra of proteins with known three-dimensional structures as a calibration matrix. Spectra are either used directly as a spectroscopic pattern to model the spectrum of the protein to be determined, or indirectly after reduction of the protein spectra into a number of eigenspectra which represent the most important spectral information. Mathematical tools such as partial least-squares analysis[17,18] and factor analysis[19–21] are normally applied for the evaluation of the protein spectra. In order to validate the method, a leave-one-out rotation procedure is applied. This involves sequential removal of the spectrum of each protein from the calibration set and estimation of the secondary structure using the remaining proteins as reference. Good correlation has been observed between the secondary structure contents estimated and those derived from corresponding X-ray structures, although the number of proteins included in these studies was limited (only 13–23 proteins were selected by different workers). The best results were obtained for proteins recorded in H_2O solution, whereas the analysis of the spectra of proteins in D_2O was less successful. This is not surprising, since variations in the extent of H−D exchange for the different proteins used to create the calibration set results in several band shifts in the amide I′ region caused by partially exchanged structures. Nevertheless, the advantage of pattern recognition approaches is that these methods eliminate the inherent subjectivity of the curve-fitting techniques discussed above and they do not require the assignment of individual component bands to different types of secondary structure. However, these methods cannot be freely applied without considering certain limitations. For simplicity, most of the approaches involve normalization procedures that assume equal molar absorptivities for various secondary structures, similar to the curve-fitting analysis described previously. More importantly, the methods encounter difficulties in cases where the spectral features of the protein under study do not reflect the characteristics of the spectra within the calibration set. In such situations incorrect estimation of the secondary structure is very likely, even though the mathematical treatment of the spectral data is formally correct.

An increasing number of observations confirm that some proteins have spectral characteristics that cannot be recognized in the spectra of a reference database, when the number of spectra is still relatively small and does not cover the full range of possible types of conformations. Furthermore, the usefulness of these strategies in secondary structure estimation of partially or completely unfolded states of proteins is not always satisfactory. In this context, the IR methods have many features in common with the corresponding strategies used to analyze CD spectra. The potential sources of error in the IR and the CD methods, however, are different. Thus, the combination of the two approaches is highly recommended in order to obtain a complementary overall picture of protein secondary structure. In addition to these pattern recognition techniques, methods with learning capabilities such as neural nets are now emerging and these will widen the range of spectral analyses possible.[22]

4.3 Advantage of Comparing Spectra of Known Structures with Those Containing Alterations

Notwithstanding the difficulties and some limitations in the quantitative assessment of protein secondary structure – a situation common to all ''low-resolution'' methods – IR spectroscopy provides a very sensitive and reliable tool for monitoring, in relative terms, even marginal changes in the conformation of the polypeptide backbone. The IR approach can also provide detailed information about local changes in the microenvironment of side-chain groups difficult to assess with other methods. For example, if the three-dimensional structure of a protein is known and one is interested in demonstrating changes in protein structure caused by external factors such as temperature, solvents, ions or, in more complex systems, the whole environment (i.e. a membrane lipid composition), IR spectroscopy can estimate the induced structural changes. Difference spectra generated between spectra of the original state and the altered state of a protein allow a critical assessment of the new structures induced, based on the knowledge of the three-dimensional structure. Since the strength of IR spectroscopy is the ability to identity spectra by

simple comparison, very small spectral changes can easily be observed and analyzed. In the following sections, selected examples of the types of conformational changes that can be probed by FTIR methods are introduced, together with a description of how this information can be extracted from the experimental data. An extensive review of the tremendous number of papers in this field and a description of all IR-based experimental strategies is beyond the scope of this article, however. In particular, we will not present or discuss in detail the wealth of IR studies on photobiological systems (e.g. hemoproteins, bacteriorhodopsin, or photosynthetic reaction centers). These experiments normally require step-scanning techniques and (using specialized tools) may provide unique information on the structural and functional aspects of the system under study on the nanosecond or picosecond timescale. These methods can also be very specific down to the level of single functional groups in a large molecule. The interested reader is referred to recent reviews.[23–29]

5 UNFOLDING AND FOLDING EVENTS UNDER EQUILIBRIUM CONDITIONS

Studies of unfolding/folding of proteins under equilibrium conditions can be monitored by a number of techniques, including FTIR spectroscopy. The structural alterations to be studied in a protein can be induced as for other methods, ranging from simple changes in temperature and pressure to extremes in pH and the addition of denaturants. The purpose of this section is to discuss the general effects that can be observed by IR spectroscopy when simple parameters (e.g. temperature) or the entire environment (e.g. denaturants) are changed. For a typical pressure experiment, the protein is placed in a specially designed diamond anvil cell, and the pressure is calibrated using α-quartz as an internal standard.[30]

5.1 Thermal Unfolding of Proteins – Irreversible Aggregation

Illustrative applications are FTIR studies on the conformation of the enterotoxic protein cholera toxin, and the effect of receptor binding on the structure and thermal stability of the B subunit.[31] The toxin molecule is composed of two structurally and functionally distinct subunits A and B. While the A subunit is directly responsible for the toxic activity of the protein, the major role of the B subunit is to initiate the toxin–target cell interaction by binding to the monosialoganglioside G_{M1} receptor on the membrane surface. As shown in Figure 7(a), the amide I′ band contour of the free B subunit remains unchanged upon an increase in temperature from 24

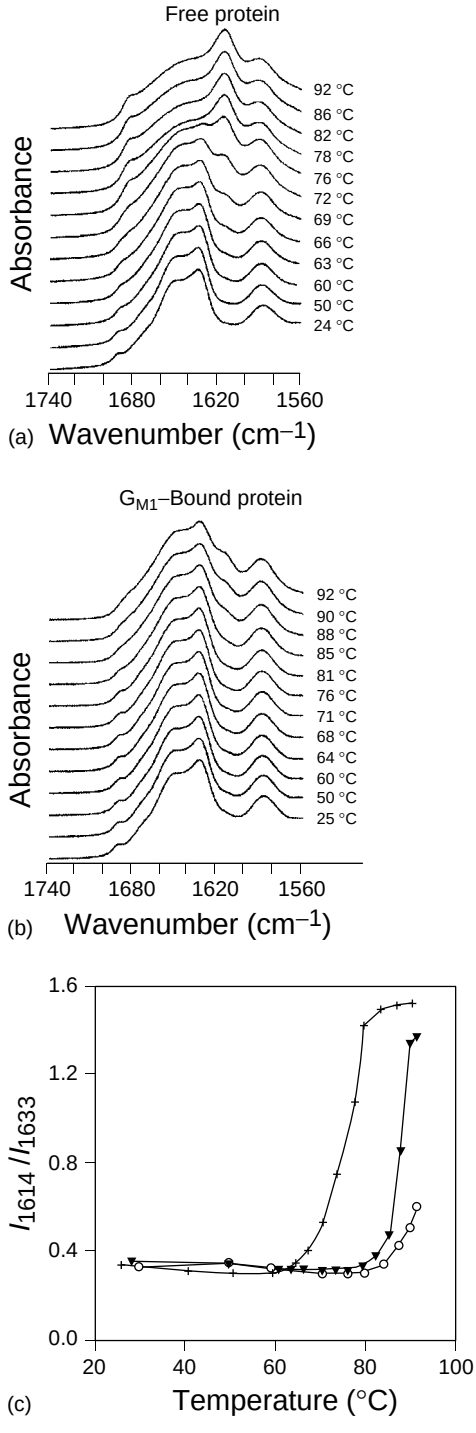

Figure 7 Temperature dependence of IR spectra of the cholera toxin B subunit free in D_2O solution (a) and in the presence of ganglioside G_{M1} at the glycolipid to the B subunit monomer molar ratio of 2 : 1 (b). Ratio of cholera toxin B subunit amide I′ band intensity at 1614 cm^{-1} to that at 1633 cm^{-1} as a function of temperature (c). Free B subunit in aqueous solution (+) and B subunit in the presence of oligo G_{M1} ( ) or intact ganglioside G_{M1} (○). (Reprinted with permission from W.K. Surewicz, J.J. Leddy, H.H. Mantsch, *Biochemistry*, **29**, 8106–8111 (1990). Copyright 1990 American Chemical Society.)

up to about 66 °C, indicating that the protein secondary structure did not change in this temperature interval. A further increase in temperature results in major spectral changes, indicating the denaturation of the B subunit. Upon denaturation, the features representing β-sheet structures (around 1633 cm^{-1}) and α-helices (around 1652 cm^{-1}) in the spectrum of the native protein are transformed into new features displaying a strong band centered at 1614 cm^{-1} and weaker bands around 1683 and 1644 cm^{-1}. For the analysis of these changes, the ratio of the band intensities at 1614 and 1633 cm^{-1} was found to be a useful parameter to follow thermal denaturation (Figure 7c). Although the IR spectrum of the receptor also shows a broad IR band overlapping with the amide I′ band of the toxin, this particular ganglioside absorption was found to be very weak in comparison with the protein amide I′ band. In this model system it was therefore permitted to analyze the spectra of the complex in the same way as for the free protein, ignoring the minor spectral contributions of the receptor (Figure 7b). The evaluation shows that binding the toxin either to the ganglioside G_{M1} receptor or to only the isolated oligosaccharide moiety of the glycolipid can substantially increase the thermal stability of the B subunit.

A noticeable feature present in the IR spectra of the thermally denatured cholera toxin subunits A and B is a strong and sharp band at 1614 cm^{-1}, normally paired with a second weaker band at higher wavenumbers (here at around 1685 cm^{-1}). These two sharp IR bands (the low-wavenumber band in particular) are indicators (marker bands) for protein aggregation. Their origin is a special kind of extended structure, which is often formed upon aggregation of thermally unfolded proteins.

5.2 The Effect of Point Mutations on the Thermal Stability of Proteins

RNase T1 belongs to the few proteins that show no signs of aggregation under conditions typical for standard FTIR protein measurements, such as protein concentrations of 5–10 mg mL^{-1} (for measurements in D_2O) and an extended time frame necessary for collection of the spectra. The two major elements of secondary structure of RNase T1 are a single elongated α-helix and an extended antiparallel β-sheet composed of three long and two short β-strands. The IR spectrum of native RNase T1 in D_2O buffer after complete H–D exchange and subtraction of the buffer is dominated by a broad amide I′ band, which exhibits a maximum at 1644 cm^{-1} and two well-defined shoulders around 1626 and 1678 cm^{-1} (solid line in Figure 8a). Further details are obscured by the overlapping of the amide I′ band components that represent different elements of secondary structure present in the protein.

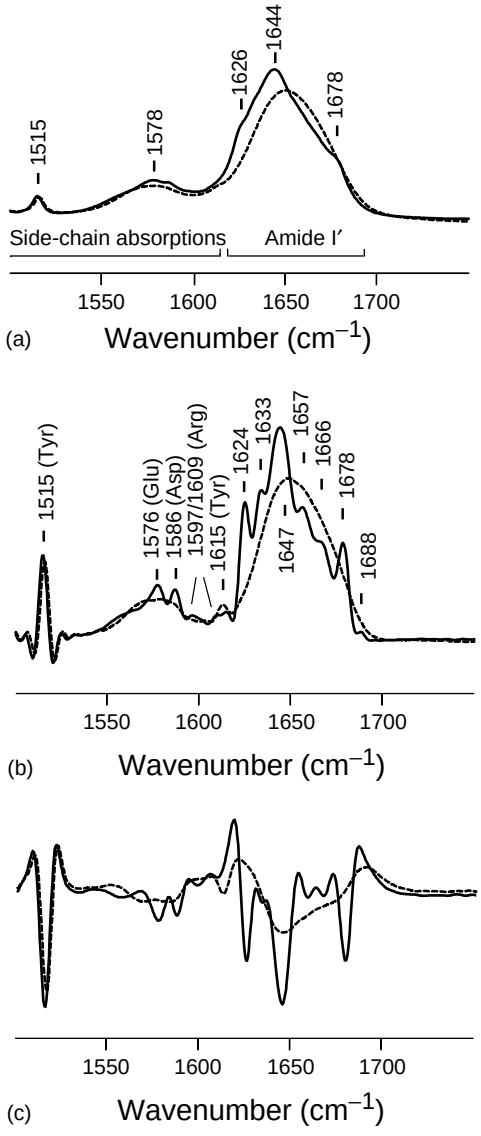

Figure 8 (a) Original IR spectra of RNase T1 recorded in D_2O buffer solution at 20 °C (solid line) and at 70 °C (dashed line). (b) Comparison of the IR spectrum of native RNase T1 at 20 °C (solid line) with that of thermally denatured RNase T1 at 70 °C (dashed line) after band narrowing by Fourier self-deconvolution performed with identical deconvolution parameters. (c) Second derivatives of the IR spectrum of RNase T1 at 20 °C (solid line) and at 70 °C (dashed line).

As for many other proteins, most of these band components can be resolved by a mathematical procedure of band narrowing such as Fourier self-deconvolution or derivatization. The deconvoluted IR spectrum of native RNase T1 (solid line in Figure 8b) and also the second-derivative spectrum (solid line in Figure 8c) reveal the presence of seven individual bands in the amide I′ region. They can be assigned to characteristic secondary structure elements, e.g. the IR bands at 1624 and 1633 cm^{-1}

are assigned to different β-sheet structures.[32] Based on comparative analysis of the spectral features of RNase T1 in H_2O, in partially deuterated states, and after complete exchange of all amide protons, it was suggested that the band component at 1657 cm^{-1} can be associated with the α-helix present in RNase T1. The component at 1644 cm^{-1} could then be assigned to irregular structures with minor contributions from 3_{10} and/or β-turn structures all present in RNase T1. The band component at 1666 cm^{-1} was solely attributed to turn structures, whereas both antiparallel β-sheet structures and turns were associated with the band components at 1678 and 1688 cm^{-1}. After complete thermal denaturation of RNase T1 at 70 °C, the IR spectrum of the protein (dashed lines in Figure 8a–c) exhibits only a broad, nearly featureless amide I′ band contour centered at 1647 cm^{-1}, completely different from the spectrum recorded at 20 °C. The broad, featureless amide I′ band contour suggests that the thermally unfolded state of RNase T1 is predominantly irregular, comparable to a randomly oriented polypeptide chain.

It is interesting to note that the thermal unfolding of the protein also resulted in clear spectral changes in the range 1500–1615 cm^{-1}. In fully exchanged proteins, this region is entirely dominated by amino acid side-chain absorptions of tyrosine (two bands at 1515 and 1615 cm^{-1}), glutamate (around 1576 cm^{-1}), aspartate (around 1586 cm^{-1}), and arginine (around 1597/1609 cm^{-1}).

From IR spectra collected as a function of temperature, intensity/temperature and frequency/temperature profiles for selected IR ''marker bands'' can be constructed. The measurement of these spectral changes provides a means of determining standard thermodynamic parameters of a protein, such as transition temperatures (T_m) and enthalpy changes (ΔH).[32,33] Figure 9(a) shows such intensity/temperature plots for the amide I band at 1625 cm^{-1} in the spectra of wild-type RNase T1 and three of its variants with altered amino acid composition. This band provides a good monitor of the unfolding of the secondary structure of the slightly different RNase T1 samples. The tyrosine band around 1515 cm^{-1}, on the other hand, provides a particularly useful local monitor to probe the thermal denaturation of the proteins, because the shift in frequency indicates changes in the microenvironment around the tyrosines. Figure 9(b) shows frequency–temperature plots for the tyrosine band of the four RNase T1 samples.

The data in Figure 9(a) and (b) indicate that all three mutant proteins display a decrease in the thermal transition temperature (thermal stability). The mutation Tyr-45 → Trp affects T_m less than the replacement of Trp-59 by Tyr in both the single (W59Y) and the double (Y45W/W59Y) variant.

The temperature dependence of the different IR ''marker bands'', which probe different structural features of the protein, permits us also to address the question of whether or not unfolding can be fitted by a two-state unfolding model. If the temperature profiles of two (or more) bands coincide then this is a strong indication in favor of such a model, although it cannot be excluded that corresponding signals may be insensitive to one of the transitions in some cases. If corresponding profiles do not coincide, it clearly indicates that intermediates are present at equilibrium and hence a simple two-state transition cannot be used for analysis of the data. It is a specific advantage of the FTIR spectroscopic approach that such information can be derived from a single experiment with a single sample.

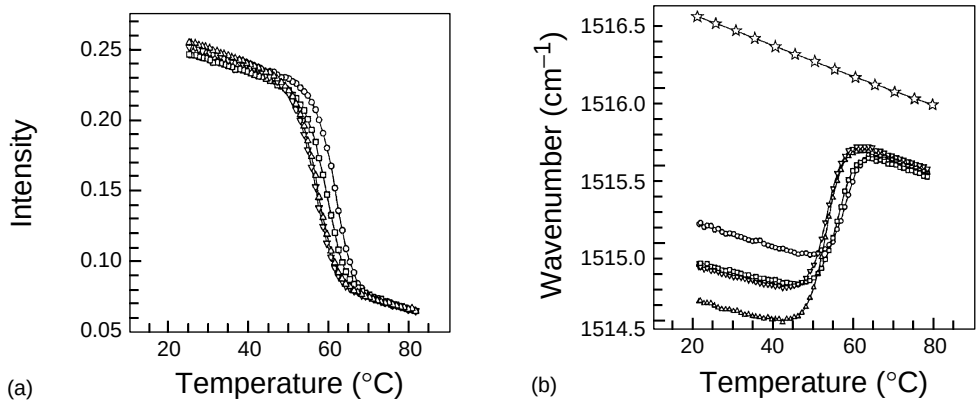

Figure 9 Temperature dependence of the peak height intensity of the amide I′ band at 1625 cm^{-1} (a) and temperature dependence of the position of the tyrosine band (b) for wild-type RNase T1 (○) and for the variants Y45W/W59Y (△), W59Y (▽), and Y45W ( ). For comparison, the temperature dependence of the position of the tyrosine band (☆) in a short peptide known to lack any secondary structure at all temperatures is also shown, which illustrates the general effect of temperature on the position of the corresponding tyrosine band. (Reprinted with permission from H. Fabian, C. Schultz, J. Backmann, U. Hahn, W. Saenger, H.H. Mantsch, D. Naumann, *Biochemistry*, **33**, 10 725–10 730 (1994). Copyright 1994 American Chemical Society.)

5.3 Impact of Point Mutations on the Structure of Proteins Examined by Infrared Difference Spectroscopy

The study of marginal changes in protein structure by IR difference spectroscopy can provide information on changes of single peptide bonds or side-chain groups, but has been limited for many years to studies of light-triggered chromophoric proteins, as those spectra can be measured with minimal sample manipulation and without removal from the FTIR instrument. The quantitative comparison of spectra of different samples, often recorded at different protein concentrations, requires an internal standard for their normalization. The well-separated tyrosine band at around 1516 cm^{-1} is a useful internal reference to normalize the spectra of a wild-type protein and its mutants. In the case of proteins which do not aggregate upon heat denaturation, the IR spectra of thermally unfolded proteins provide an ideal internal intensity standard for the quantitative comparison of spectra of related native proteins.[33] This is illustrated in Figure 10(a) and (b) for the protein RNase T1. Figure 10(a) shows the IR spectra of wild-type RNase T1 (dashed line) and the mutant Y45W/W59Y (solid line) after normalizing both data sets based on the integrated intensity of their spectra in the thermally unfolded state. Figure 10(b) shows IR difference spectra between the wild-type protein and some variants. Positive and negative features in these difference spectra reflect fine structural differences, in whose absence only a flat line is obtained.

An essential prerequisite for this kind of data evaluation is that the spectra have been recorded with a very high signal-to-noise ratio, and that the stability of the instrument is also very high, otherwise minor spectral differences will be ''buried'' within the noise. Clear spectral differences (at 20 °C) can be observed between wild-type RNase T1 on the one hand and the two variants Y45W/W59Y and W59Y on the other, whereas practically identical difference spectra in the amide I′ region were obtained for the double variant Y45W/W59Y and the single variant W59Y (compare traces A and B in Figure 10b). This demonstrates that it is the change from Trp to Tyr in position 59 which affects the overall protein conformation. In particular, the positive bands at 1628 and 1635 cm^{-1}, along with the negative band at 1622 cm^{-1} in these difference spectra, reflect fine differences in the hydrogen-bonding pattern of β-structures in the two variants, compared to the wild-type protein. The mutation Tyr to Trp in position 45, on the other hand, has practically no impact on the polypeptide backbone conformation, as indicated by the almost flat difference spectrum (trace C in Figure 10b). This was found to be in good agreement with X-ray data, and demonstrates the sensitivity of the IR spectroscopic approach.

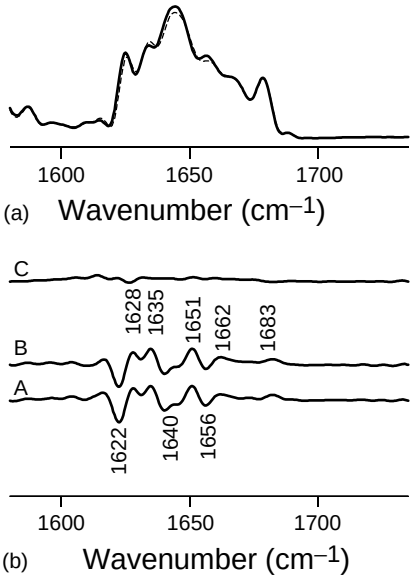

Figure 10 (a) Deconvoluted IR spectra of wild-type RNase T1 (dashed line) and of the double mutant Tyr-45 → Trp/Trp-59 → Tyr (solid line) at 20 °C after normalization using the spectra of the corresponding thermally unfolded proteins. (b) IR difference spectra: (A) wild-type at 20 °C minus Y45W/W59Y variant at 20 °C; (B) wild-type at 20 °C minus W59Y variant at 20 °C; (C) wild-type at 20 °C minus Y45W variant at 20 °C. Note that the absorbance scale for the difference spectra in (b) was expanded by a factor of two compared with the scale in (a).

5.4 Chemically Denatured States of Proteins

Obtaining IR spectra of proteins in the presence of the most commonly used denaturating agents, urea and guanidinium chloride (GdmCl) is not simple, and has been unsuccessful for a long time for several reasons. First, the denaturant that must be used at high concentrations to achieve a significant structural alteration has very strong IR bands of its own. Therefore, IR cells with very short pathlengths (5–7 μm) are required to prevent saturation of the IR detector by absorption of the chemical denaturant. Second, the major IR band of urea around 1613 cm^{-1} (Figure 11a, dashed line) or GdmCl around 1600 cm^{-1} (Figure 11b) masks the much weaker protein backbone modes of interest and prevents their analysis. Isotopic labeling of the denaturant (e.g. using $^{13}C{=}O$-labeled urea) is a suitable method for helping to circumvent this problem by shifting the urea band from 1613 to 1560 cm^{-1} (compare the dashed and solid lines in Figure 11a). In this way, isotope labeling creates a clear window in the IR spectrum for observing the protein amide I′ bands above 1600 cm^{-1}.[34] This strategy allows the measurement of chemically induced unfolding transition of proteins as described by other methods and enables us to compare conformational features of thermally and chemically denatured states

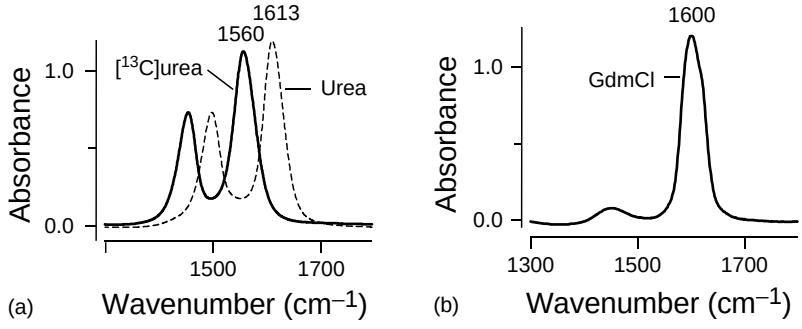

Figure 11 (a) IR spectra in D_2O of deuterated urea (dashed line) and of deuterated $^{13}C{=}O$-labeled urea (solid line). (b) IR spectrum of deuterated GdmCl. The samples were placed between a pair of CaF_2 windows separated by path lengths of 6–6.5 μm; the concentration of the denaturants was in the range 6–7 M.

of the same proteins[34–36] or of slightly different proteins.

6 ANALYSIS OF PEPTIDES

FTIR spectroscopy is also extensively used in conformational studies of shorter and longer peptides.[37] Typical applications include the analysis of the formation of secondary structures and their changes when peptides were subjected to environmental perturbations (e.g. temperature, pressure, pH, addition of organic or membrane-mimicking solvents). In some cases, the impact of point mutations or amino acid modifications (e.g. D-amino acid substitution, amidation, phosphorylation) on the structure and stability of peptides was studied. More difficult, but very beneficial to the understanding of membrane effects on proteins structures, are studies aiming to monitor changes in the conformation of a peptide resulting from an interaction with a membrane.

Nevertheless, some difficulties exist in the quantitative assessment of the secondary structure of peptides from IR spectra. Primarily, this is because the position of an IR band associated with a certain secondary structure in a peptide can be significantly different from that of the corresponding structure in a globular protein. For instance, α-helical structures in globular proteins typically show absorption in the range 1650–1658 cm^{-1}, while the position of a band of an α-helical structure in a solvent-exposed peptide may occur below 1640 cm^{-1}.

Since the pattern recognition methods discussed previously rely on calibration spectra obtained almost exclusively from globular proteins, these approaches may not (yet) be applied for secondary structure estimation of peptides. Hence the curve-fitting approach is often applied for this purpose, despite the weaknesses associated with band assignment. In proteins, differences between the H–D exchange kinetics of amide protons in irregular structures (very fast) and in α-helical structures (often very slow) can aid band assignment. In peptides, however, this strategy usually does not help very much, since isotopic exchange in α-helical segments of peptides is typically also fast. Here, FTIR and CD spectroscopy should both be applied, since the latter method is known to be more suitable for quantifying α-helical structures. Moreover, IR spectra of peptides often reveal relatively structureless amide I band contours, even after the application of resolution-enhancement techniques. This implies a large degree of subjectivity in the estimation of the number of band components and their relative contribution to the total amide I band contour. An exception is the analysis of antiparallel β-type structures, which can readily be identified by characteristic amide I′ band components near 1620–1635 cm^{-1} (very strong) and near 1675–1690 cm^{-1} (weak).

The potential of FTIR spectroscopy to indicate very sensitively the presence of, and changes in, β-pleated sheet structures has been employed to characterize conformational properties of synthetic analogs of peptides known to be associated with pathological conditions such as Alzheimer's disease and the spongiform encephalopathies. These neurological disorders have in common an aberrant conformational transition in an underlying peptide/protein, characteristically leading to aggregated β-sheet assemblies and tissue deposition. The IR approach does not provides structural details of the arrangement of such assemblies of a synthetic analog, but is useful in studying the impact of environmental factors, naturally occurring point mutations or chemical modifications on conformation and stability.

6.1 Alzheimer β-Amyloid Peptides

As an example for peptide analysis, Figure 12(a–c) shows IR spectra of analogs of a 42-amino acid peptide, called βA4, known to be involved in the formation of neuritic plaques in patients with Alzheimer's disease.

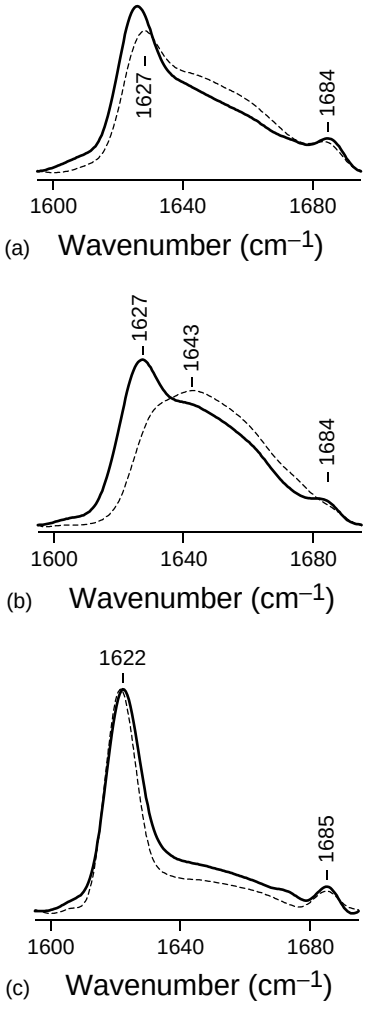

Figure 12 IR spectra of synthetic β-amyloid peptides corresponding to the normal human sequence [H(1–42)] (dashed line) and the variant containing isoaspartic residues in positions 1 and 7 [I(1–42)] (solid line) recorded in different environments. In D_2O at pH 8 (a), in D_2O at pH 11 (b), and in 10% octyl-β-D-glucoside at pH 8 (c). All spectra are shown after band narrowing by Fourier self-deconvolution performed with identical deconvolution parameters.

The comparative study of two peptides, a synthetic peptide corresponding to the normal human sequence [H(1–42)] with an analog of the human sequence containing isoaspartic residues in positions 1 and 7 [I(1–42)], revealed that the naturally occurring post-translational modification of the aspartyl residues result in more stable β-aggregates of the corresponding synthetic peptides.[38] At pH 8, the characteristic β-sheet marker bands at 1626/27 and 1684 cm^{-1} are more pronounced in the spectrum of the I-type peptide (Figure 12a). Differences in the spectra of the two peptides are even more striking at pH 11 (Figure 12b). The major band in the spectrum of the normal H-peptide is centered at 1643 cm^{-1}, characteristic of a peptide with only little regular secondary structure. In contrast, the spectrum of the I-peptide still exhibits strong bands at 1627 and 1684 cm^{-1}, indicating a significant amount of β-type structure in the modified peptide at pH 11. For peptide analysis, organic solvents or others are often employed to probe the propensity of a peptide to form an α-helix (in trifluoroethanol) or to mimic a micellar environment (e.g. sodium dodecyl sulfate, octyl-β-D-glucoside, or aqueous acetonitrile solutions). In principle, IR spectra of peptides in these solvents can be recorded as readily as in pure aqueous solution. However, care must be taken to subtract properly the solvent/buffer spectrum (the solvents have IR bands of their own overlapping with bands of the peptide) and to avoid evaporation of the very volatile solvents after filling the IR cell. The spectra of the β-amyloid peptides in octyl-β-D-glucoside (Figure 12c) can be taken as textbook examples of antiparallel β-sheet structures.

6.2 Membrane-associated Peptides/Proteins

FTIR spectroscopy has proven to be of particular value in probing conformational transitions in peptides/proteins caused by the interaction and association with lipid bilayers or natural membranes. Such changes are thought to be important for understanding a number of physiological processes. Studies of membrane-associated peptides or proteins are often difficult to perform by other spectroscopic techniques, e.g. owing to spectral distortions caused by light scattering. An advantage of FTIR spectroscopy is that it provides structural and dynamic information on both the peptide/protein and the lipid components simultaneously, and that extra probes and marker molecules are not required.[39,40] In addition, polarized IR spectroscopy can be used to measure the average orientation of peptides/proteins with respect to the plane of the membrane. Moreover, isotope labeling, combined with site-directed mutagenesis, now makes it possible to scrutinize conformational properties at the level of individual chemical groups. The latter strategy has been elegantly demonstrated on phospholamban (PLB), a small membrane protein which can form cation-selective ion channels in lipid bilayers. The protein is anchored in the membrane by a 28-amino acid fragment corresponding to its C-terminal region (hPLB). Polarized ATR FTIR spectroscopy along with site-directed isotope labeling was used to probe the local structure of hPLB.[41]

Figure 13 shows the spectra of dehydrated films of hPLB and hPLB labeled at both Leu-39 and Leu-42 amide carbonyl groups. The amide I and amide II bands of hPLB are located at 1658 and 1544 cm^{-1}, indicating a predominantly α-helical structure for the peptide reconstituted into dimyristoylphosphatidylcholine (DMPC)

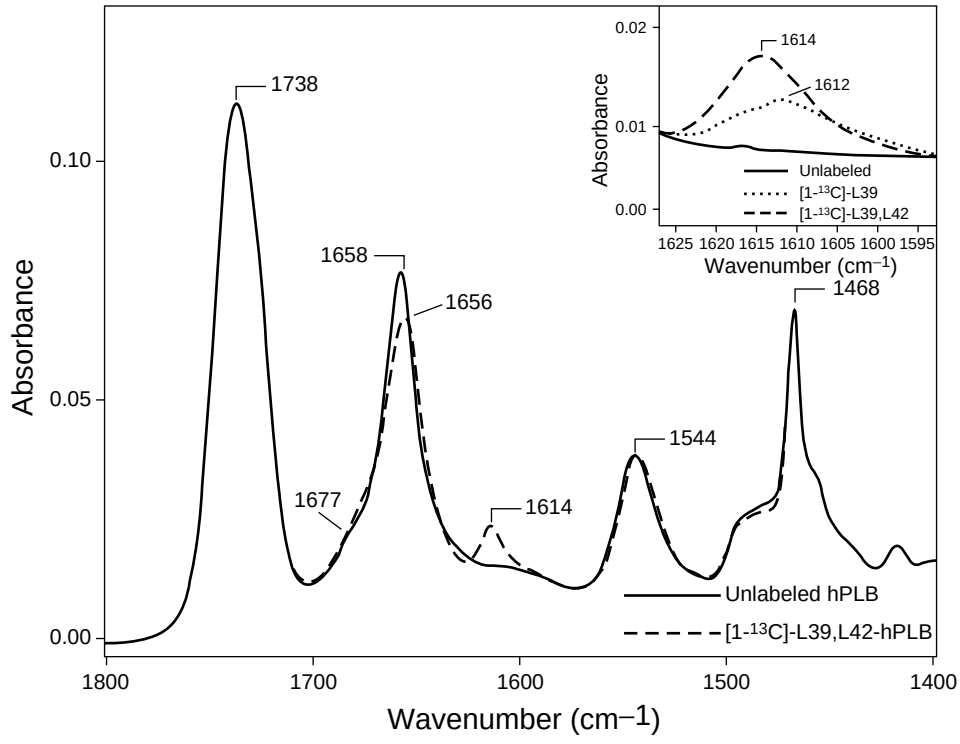

Figure 13 Comparison of ATR FTIR spectra of the unlabeled C-terminal sequence of hPLB and [1-^{13}C]-L39,L42-hPLB recorded using parallel polarization. Measurements were performed on dehydrated films. Inset: comparison of the expanded region near 1614 cm^{-1} of unlabeled hPLB, [1-^{13}C]-L39-hPLB, and [1-^{13}C]-L39,L42-hPLB. (Reproduced by permission of the Biophysical Society from Lundlam et al.[41])

bilayer membranes. The other two intense bands in the spectrum, which are well separated from the peptide backbone modes, are associated with the DMPC ester carbonyl stretching mode (at 1738 cm^{-1}) and the CH_2 scissoring mode (at 1486 cm^{-1}). The replacement of a ^{12}C=O group with a ^{13}C=O group decreases the amide I vibration by 40–45 cm^{-1} owing to the increased mass. Labeled amino acids are commercially available and can be incorporated during peptide synthesis without additional effort. Utilizing this approach, comparison of the spectra of unlabeled and labeled peptides allows the identification and analysis of the amide I bands that originate exclusively from the labeled site. This is clearly shown in the spectra of hPLB (Figure 13). A drop in intensity at 1658 cm^{-1}, together with the appearance of a new band at 1614 cm^{-1}, is observed, which can be assigned to the isotope-induced downshifted amide I modes of Leu-39 and Leu-42. This assignment is supported by the fact that the intensity of the band near 1614 cm^{-1} has only approximately half the intensity in the spectrum of hPLB labeled at only Leu-39 (Figure 13, inset). The position of the amide I modes of Leu-39 and Leu-42 indicates that both residues are part of an α-helical structure and therefore residue $i + 4$ (e.g. $46 = 42 + 4$) must also be in a helical geometry. The same holds for the nearby residues, which contribute to the hydrogen bonding scheme of the α-helix. That means that a single label can report on the local secondary structure of four residues. Hence it was concluded that six isotopic labels would be the minimum needed to define a transmembrane helix (~22 amino acid residues).[41] Moreover, the IR linear dichroism (data not shown) of the amide I bands assigned to the labeled sites suggested an α-helix with an axial orientation of 30° relative to the membrane normal. H–D exchange measurements revealed that 70% of the amide protons, including the peptide groups of Leu-39 and Leu-42, are inaccessible to the solvent, indicating that most of the hPLB fragment is embedded within the lipid bilayer.

Site-directed isotope labeling is so valuable to FTIR spectroscopy that its application allowed the study of local changes in the polypeptide backbone occurring during the photocycle in an integral membrane protein, bacteriorhodopsin.[6,42] In this case, L-tyrosine containing a ^{13}C isotope at the carbonyl carbon was selectively introduced in a biosynthetic process at specific positions along the polypeptide chain of bacteriorhodopsin. Then structural changes during the transition from the M to the N state were monitored by IR difference spectroscopy. Similar results to those of the hPLB peptide discussed

previously were observed, i.e. a minor drop in intensity of the amide I band assigned to the α-helical structure in bacteriorhodopsin, along with the appearance of a weak downshifted amide I band. The characteristic spectral changes were only recorded after labeling at Tyr-185. This demonstrated that of the 11 tyrosines, only the peptide carbonyl group of Tyr-185 undergoes a significant conformational change during the early photocycle. This is an excellent example which demonstrates the sensitivity of FTIR spectroscopy in monitoring one selective bond in a rather large membrane protein.

7 ANALYSIS OF PEPTIDE–PROTEIN AND PROTEIN–PROTEIN COMPLEXES

Extensive overlap of the diagnostic IR bands limits the application of FTIR spectroscopy when a peptide and a protein or two different proteins form a complex. These problems can be overcome by mixing a completely ^{13}C-labeled protein with an unlabeled peptide or protein. This is illustrated in Figure 14(a–c), which compares the IR spectra of (a) unlabeled calmodulin, (b) ^{15}N uniformly labeled calmodulin, and (c) $^{13}C/^{15}N$ uniformly labeled calmodulin. As expected, ^{15}N labeling produces almost no changes in the amide I′ band contour, owing to the relatively minor C–N stretching contribution to the amide I vibration. On the other hand, ^{15}N labeling causes a significant shift of the amide II′ band of the protein (from 1457 to 1430 cm^{-1}). Complete (biosynthetic) labeling solely with ^{13}C results in a large low-frequency shift of the amide I′ band of the protein away from the original spectral region, and allows the conformational changes associated with binding of unlabeled synthetic target peptides to calmodulin[43] to be studied individually.

Figure 15(a) shows the IR spectra of $^{13}C/^{15}N$ uniformly labeled calmodulin (dashed line) and its complex with the calmodulin-binding domain of the peptide MLCK (solid line). It is important to ensure that the generally larger protein, in this case calmodulin (148 amino acids), be labeled by over 99% to prevent the appearence of any residual intensity from the unlabeled protein. This allows the detection of the generally weaker amide I′ bands of the smaller target peptides, in this case MLCK (22 amino acids). A comparison of the spectrum of the MLCK peptide alone and in complex with calmodulin reveals clear differences (compare the solid and dashed lines in Figure 15b). In this case, the carryover counterion TFA used in peptide synthesis, which has a strong IR band at 1674 cm^{-1}, was used as an internal reference to normalize the spectra. This permits comparison not only of frequencies but also of intensities between the spectra

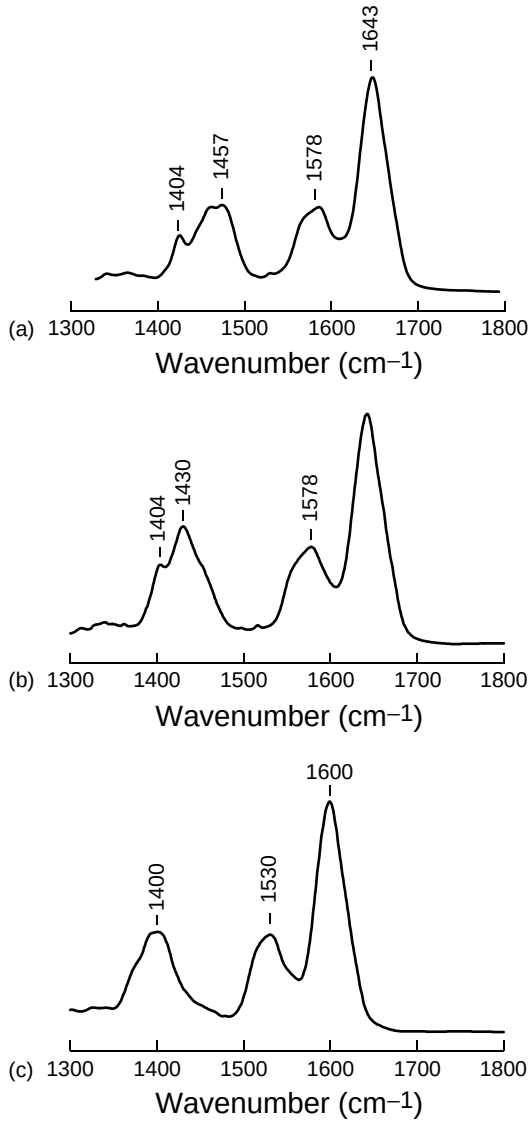

Figure 14 Amide I′ and amide II′ regions of the IR spectra of (a) unlabeled, (b) ^{15}N-uniformly labeled, and (c) $^{13}C/^{15}N$-uniformly labeled calmodulin in D_2O buffer saturated with Ca^{2+}.

of the peptide alone and in the peptide–protein complex. The broad and featureless amide I′ band contour of the MLCK peptide centered at 1642 cm^{-1} is typical of an unstructured peptide. The amide I′ band of the peptide in the complex is considerably more intense, narrower, and centered at 1647 cm^{-1}, indicating that the peptide has changed from an irregular structure to an α-helical conformation upon binding to calmodulin. Spectral differences between 1550 and 1610 cm^{-1} suggest that the binding of the peptide also leads to a slight but perceptible perturbation of the conformation of calmodulin. Recently, the isotope-edited strategy was employed to determine the protein secondary structures of both ligand and receptor in a signal transduction complex.[44]

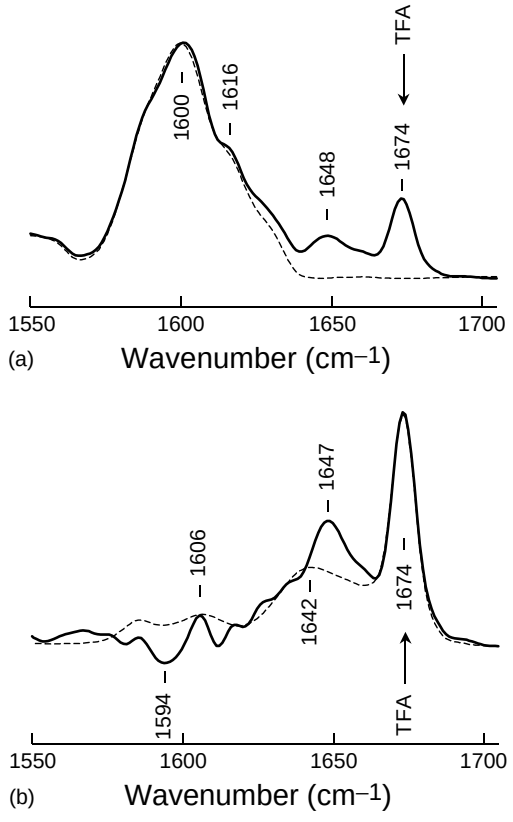

Figure 15 (a) Deconvoluted IR spectra of $^{13}C/^{15}N$-labeled calmodulin (dashed line) and its complex with the myosin light chain kinase (MLCK) peptide at a 1 : 1 ratio (solid line). (b) Comparison of the deconvoluted IR spectrum of the MLCK peptide in D_2O buffer (dashed line) and the spectrum of the peptide in complex with $^{13}C/^{15}N$-labeled calmodulin (solid line). The IR band resulting from residual internal trifluoroacetic acid (TFA) was used to normalize the intensity of the corresponding peptide spectra.

8 SULFHYDRYL GROUP ABSORPTIONS AS LOCAL PROBES OF PROTEIN STRUCTURE

IR bands arising from the S−H stretching mode of cysteine residues occur between 2520 and 2600 cm^{-1}, a region devoid of interference from any other protein absorptions. Many proteins contain only a few S−H groups, which permits the analysis of them individually and thus enables information about a defined local site within a protein to be obtained. One difficulty is that S−H groups absorb rather weakly in the IR region. Hence the use of highly concentrated samples (>50 mg mL^{-1}) and cells with path lengths of 200–250 μm is required, which is feasible because the water absorption between 2500 and 2600 cm^{-1} is minimal (see Figure 2a and b). Variations in band parameters (frequency, intensity, half bandwidth) of the S−H absorptions reflect differences in the local environment and the hydrogen bonding experienced by the S−H groups.[23,24] Strong hydrogen bonding between an S−H group and a peptide carbonyl leads to a low-frequency shift of the corresponding IR band. S−H groups located in a solvent with nonhydrogen-bonding properties have an extinction coefficient much smaller than those that are involved in hydrogen bonding. Studying the S−H to S−D exchange kinetics and behavior of different sulfhydryl groups can provide additional information, since the S−H vibration shifts down to lower frequencies by approximately 700 cm^{-1} on deuteration (to 1820–1900 cm^{-1}). As an example, sulfhydryl exchange in combination with site-directed substitution of cysteine has been employed for mapping the lipid-exposed surface in a synthetic transmembrane segment of PLB.[45] The basic concept behind this approach is that S−H groups within the lipid interface can undergo rapid H−D exchange, while it is assumed that S−H groups in helical interfaces are more shielded from this exchange. The comparative analysis of the exchange of peptides corresponding to the wild-type sequence and three peptides each containing a substitution of an alanine for a wild-type cysteine revealed that only one cysteine S−H exchanges with D_2O. Based on these results, it was possible to conclude that this cysteine is oriented towards the lipid interface, while the other two residues (shielded from exchange) are packed in the interior of the system studied.

9 NEAR-INFRARED SPECTROSCOPY AS A TOOL FOR MONITORING STRUCTURAL CHANGES OF PROTEINS IN AQUEOUS SOLUTION

Thus far, IR spectroscopy in the spectral range above 4000 cm^{-1}, the near-infrared (NIR) region, has seldom been employed for structural studies of peptides or proteins. The absorptions observed in the NIR region are overtones (multiples) or combinations of the fundamental stretching bands that occur in the mid-IR region. The bands involved are usually due to C−N, N−H, and O−H stretching vibrations. Generally, all these bands are much weaker than in the mid-IR region. In consequence, most of the NIR studies were limited to solid samples. For measurements in aqueous solutions, relatively high sample concentrations (>30 mg mL^{-1}) and cells with pathlengths of 0.2–1 mm are required. The latter is feasible, since the water absorption in the NIR region is relatively weak. An illustrative example of the potential of NIR spectroscopy for the characterization of structural changes in proteins is a study of the thermal unfolding of the model protein RNase A in H_2O buffer.[46] Spectra

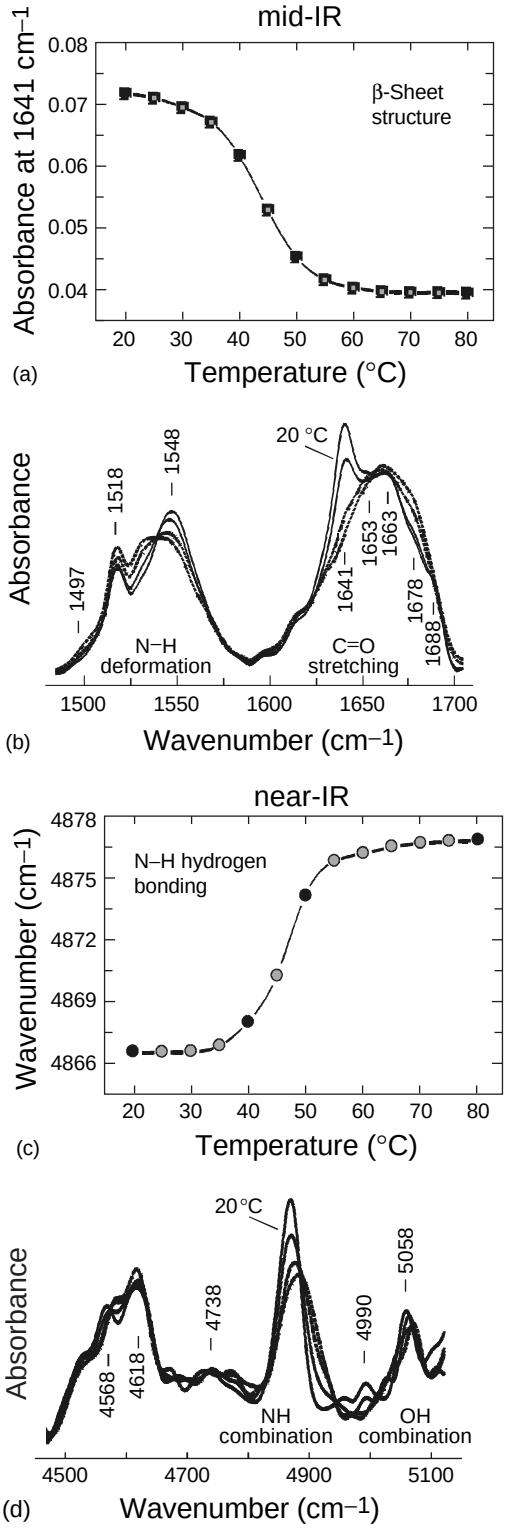

Figure 16 (a), (c) Temperature dependence of the peak-height intensity of the amide I band at 1641 cm^{-1} (a) and temperature dependence of the position of the NH combination band (c). (b), (d) Spectra of RNase A in H_2O buffer at four discrete temperatures (20, 40, 50, and 80 °C) in the mid-IR (b) and NIR region (d).

in the mid-IR and NIR regions were measured under identical experimental conditions using the same protein solution with the exception of the cell path length which was required to be about 30 times longer for the NIR studies. The heat-denaturation of RNase A is accompanied by large spectral changes, as illustrated in Figure 16(a) and (b) for the mid-IR region and Figure 16(c) and (d) for the NIR region.

As discussed previously, characteristic changes in the amide I region (e.g. intensity changes of the β-structure "marker band" at 1641 cm^{-1}) can be used to monitor thermal unfolding of the protein. In the NIR region, the most pronounced spectral changes were observed in the range between 4820 and 4940 cm^{-1}, where NH combination modes of amide A and amide II can be expected. In the spectrum of RNase A at 20 °C the N–H combination band is located at 4867 cm^{-1}. With increasing temperature this band loses intensity, becomes broader, and shifts to higher wavenumbers. The frequency–temperature plot for this band reveals that the temperature dependence is similar to the intensity–temperature plot of the backbone C=O amide I band component in the mid-IR region (compare Figure 16a and c). This demonstrates that the N–H combination band can be used as a monitor for the thermally induced unfolding of proteins in H_2O solutions. The fact that in the mid-IR N–H stretching vibrations are overlaid with strong water bands and N–H bending vibrations are always overlapped with C–N stretching vibrations makes it difficult to evaluate amide N–H vibrations of proteins in H_2O. The N–H combination band offers a new way of achieving this.

10 TIME-RESOLVED FOURIER TRANSFORM INFRARED SPECTROSCOPY, FOLDING EVENTS ON THE MILLISECOND TO MINUTE TIMESCALE

For many years, the use of time-resolved FTIR spectroscopy was restricted to events in photobiological systems where reactions could be triggered with light and were often reversible.[25–28] Only very recently has progress been made in the adaptation of conventional stopped-flow or temperature-jump technologies to the specific requirements of IR spectroscopy. A major hurdle was the construction of an apparatus which ensured a rapid flow of a concentrated viscous protein solution through an IR cell with a pathlength of only 50 µm. To achieve this, high pressure is needed and the apparatus must withstand this pressure. Furthermore, the dead volume must be kept as small as possible to minimize the amount of protein necessary for these experiments.

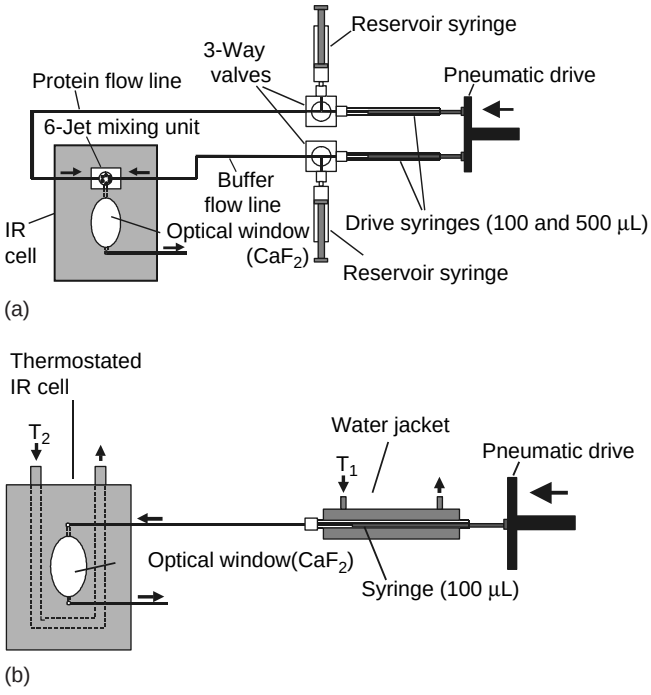

Figure 17 Schematic diagrams of the (a) stopped-flow and (b) temperature-jump system developed to monitor unfolding/folding events by time-resolved FTIR spectroscopy.

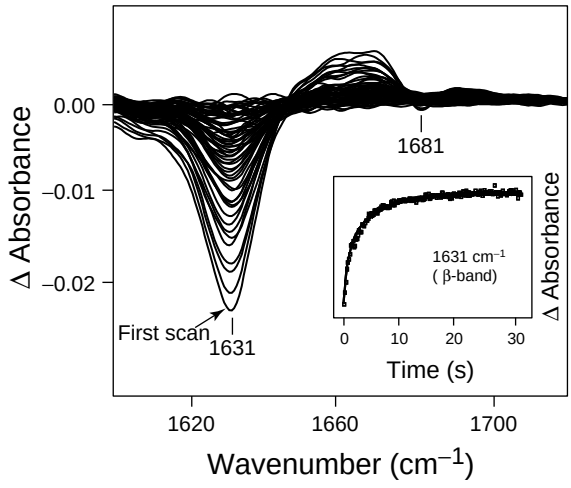

Figure 18 Typical IR difference spectra obtained after injection of 100 μL of thermally unfolded RNase A into the IR cell kept at 20 °C. The difference spectra were calculated according to $\log(S_n/S_1)$, where S_n is the single channel spectrum of the nth scan and S_1 is the single-channel spectrum of the last scan of the injection experiment. Inset: refolding followed by measuring the amide I difference peak intensity at 1631 cm^{-1} as a function of time.

A prototype of a stopped-flow apparatus for IR spectroscopy of aqueous protein solutions has recently been developed.[47] Figure 17 shows a schematic diagram of an adapted experimental set-up,[35,48] which permits the initiation of the refolding process either by applying a temperature jump on the heat-denatured protein or by rapid dilution of a concentrated ^{13}C urea solution containing the chemically unfolded protein. A pneumatic drive is used to induce either the mixing process of the chemically denatured protein with buffer or the temperature jump of the thermally unfolded protein. The drive is triggered by an electric signal from the spectrometer. The dead time of the injecting or mixing devices and the time resolution of the FTIR spectrometer allow refolding kinetics on the millisecond to minute timescale to be monitored.

Figure 18 shows a series of difference spectra obtained from a typical temperature-jump experiment with the protein RNase A. Decreasing negative features in the difference spectra indicate the formation of secondary structure after a temperature-jump from 80 to 20 °C, whereas reduced positive features correspond to a decrease in structure characteristic of the unfolded protein. The negative band at 1631 cm^{-1}, which dominates the IR difference spectrum, directly indicates the formation of antiparallel β-sheet structures in the protein. The inset in Figure 18 represents the intensity–time plot obtained employing the prominent difference band at 1631 cm^{-1}. As is already known from studies under equilibrium conditions, IR spectroscopy also permits the monitoring of the impact of unfolding/folding on the microenvironment of certain side-chain groups, which primarily indicates changes in tertiary structure arrangements. Tyrosine-detected folding of RNase A employing the band at around 1515 cm^{-1} exhibited slower kinetics than folding monitored by the β-sheet band at 1631 cm^{-1}. This clearly indicates that the formation of secondary structure precedes the formation of stable tertiary contacts in RNase A.[35] Moreover, a comparison of the intensity changes observed during kinetic experiments with those observed in equilibrium studies revealed that a significant part of the secondary structure in RNase A is already formed within the dead time (~50 ms) of the present experimental set-up.

The FTIR approach discussed above certainly has great potential to complement established techniques in describing structural events that occur during unfolding/folding of proteins on the millisecond to minute timescale. It is one of the strongest advantages of FTIR spectroscopy that a complete spectrum is available for each time point of measurement. In this way, several spectral windows are accessible simultaneously for the observation of the formation of different secondary structure elements and also folding events that can be attributed to the formation of tertiary contacts. Thus, perfect time correlation is ensured between folding events that can otherwise only be probed by different techniques, which often require different experimental conditions.

On the other hand, the kinetic IR data for RNase A and experimental data for many other proteins also clearly show that there are much faster events in protein folding that must be explored. This, however, requires new and different approaches. One very promising new strategy, which offers both the temporal resolution and the structural specificity, is to combine laser temperature-jump techniques to initiate impulsively the process with single-wavelength IR spectroscopy (using an IR diode laser) to probe the structural evolution of the event.[49]

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflection
CD	Circular Dichroism
DMPC	Dimyristoylphosphatidylcholine
FTIR	Fourier Transform Infrared
GdmCl	Guanidinium Chloride
IR	Infrared
MLCK	Myosin Light Chain Kinase
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
PLB	Phospholamban
RNase	Ribonuclease
TFA	Trifluoroacetic Acid
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Circular Dichroism in Analysis of Biomolecules • Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction

REFERENCES

1. A. Elliot, E.J. Ambrose, 'Structure of Synthetic Polypeptides', *Nature (London)*, **165**, 921–922 (1950).
2. T. Miyazawa, 'Infrared Spectra and Helical Conformations', in *Poly-α-Amino Acids*, ed. G.D. Fasman, Marcel Dekker, New York, 69–103, 1967.
3. S. Krimm, J. Bandekar, 'Vibrational Spectroscopy and Conformation of Peptides, Polypeptides and Proteins', *Adv. Protein Chem.*, **38**, 181–364 (1986).
4. H. Susi, 'Infrared Spectroscopy – Conformation', *Methods Enzymol.*, **22**, 455–472 (1972).
5. W.K. Surewicz, H.H. Mantsch, D. Chapman, 'Determination of Protein Secondary Structure by Fourier Transform Infrared Spectroscopy: A Critical Assessment', *Biochemistry*, **32**, 389–394 (1993).
6. S. Sonar, C.-P. Lee, M. Coleman, N. Patel, X. Liu, T. Marti, H.G. Khorana, U.L. RajBhandary, K.J. Rothschild, 'Site-directed Isotope Labelling and FTIR Spectroscopy of Bacteriorhodopsin', *Nature Struct. Biol.*, **1**, 512–517 (1994).
7. U.P. Fringeli, 'In Situ Infrared Attenuated Total Reflection Membrane Spectroscopy', in *Internal Reflection Spectroscopy, Theory and Applications*, ed. F.M. Mirabella, Marcel Dekker, New York, 255–324, 1992.
8. E. Goormaghtigh, V. Cabiaux, J.-M. Ruysschaert, 'Determination of Soluble and Membrane Protein Structure by Fourier Transform Infrared Spectroscopy', in *Subcellular Biochemistry, Vol. 23: Physicochemical Methods in the Study of Biomembranes*, eds. H.J. Hilderson, B.G. Ralston, Plenum Press, New York, 329–450, 1994.
9. U.P. Fringeli, J. Goette, G. Reuter, M. Siam, D.Baurecht, 'Structural Investigations of Oriented Membrane Assemblies by FTIR/ATR Spectroscopy', in *Fourier Transform Spectroscopy: 11th International Conference*, ed. J.A. de Haseth, American Institute of Physics, Woodbury, New York, 729–747, 1998.
10. D.J. Moffatt, H.H. Mantsch, 'Fourier Resolution Enhancement of Infrared Spectral Data', *Methods Enzymol.*, **210**, 192–200 (1992).
11. D.M. Byler, H. Susi, 'Examination of the Secondary Structure of Proteins by Deconvoluted FT-IR Spectra', *Biopolymers*, **25**, 469–487 (1986).
12. M. Jackson, H.H. Mantsch, 'The Use and Misuse of FTIR Spectroscopy in the Determination of Protein Structure', *Crit. Rev. Biochem. Mol. Biol.*, **30**, 95–120 (1995).
13. J.L.R. Arrondo, A. Muga, J. Castresana, F.M. Goni, 'Quantitative Studies of the Structure of Proteins in Solution by Fourier-transform Infrared Spectroscopy', *Prog. Biophys. Mol. Biol.*, **59**, 23–56 (1993).
14. W.C. Reisdorf, S. Krimm, 'Infrared Amide I′ Band of the Coiled Coil', *Biochemistry*, **35**, 1383–1386 (1996).
15. Y.N. Chirgadze, O.V. Fedorov, N.P. Trushina, 'Estimation of Amino Acid Residue Side-chain Absorption in the Infrared Spectra of Protein Solutions in Heavy Water', *Biopolymers*, **14**, 679–694 (1975).
16. S.Yu. Venyaminov, N.N. Kalnin, 'Quantitative IR Spectrometry of Peptide Compounds in Water (H_2O) Solutions. I. Spectral Parameters of Amino Acid Residue Absorption Bands', *Biopolymers*, **30**, 1243–1257 (1990).
17. F. Dousseau, M. Pezolet, 'Determination of the Secondary Structure Content of Proteins in Aqueous Solutions from Their Amide I and Amide II Infrared Bands.

Comparison Between Classical and Partial Least-squares Methods', *Biochemistry*, **29**, 8771–8779 (1990).
18. N.N. Kalnin, S.Yu. Venyaminov, 'Quantitative IR Spectrometry of Peptide Compounds in Water (H_2O) Solutions. III. Estimation of the Protein Secondary Structure', *Biopolymers*, **30**, 1273–1280 (1990).
19. D.C. Lee, P.I. Haris, D. Chapman, R.C. Mitchell, 'Determination of Protein Secondary Structure Using Factor Analysis of Infrared Spectra', *Biochemistry*, **29**, 9185–9193 (1990).
20. R.W. Sarver, W.C. Krueger, 'Protein Secondary Structure from Fourier Transform Infrared Spectroscopy: A Data Base Analysis', *Anal. Biochem.*, **194**, 89–100 (1991).
21. V. Baumruk, P. Pancorska, T.A. Keiderling, ' Predictions of Secondary Structure Using Statistical Analyses of Electronic and Vibrational Circular Dichroism and Fourier Transform Infrared Spectra of Proteins in H_2O', *J. Mol. Biol.*, **259**, 774–791 (1996).
22. P. Pancorska, H. Fabian, G. Yoder, V. Baumruk, T.A. Keiderling, 'Protein Structural Segments and Their Interconnections Derived from Optical Spectra. Thermal Unfolding of Ribonuclease T1 as an Example', *Biochemistry*, **35**, 13 094–13 106 (1996).
23. A. Dong, W.S. Caughey, 'Infrared Methods for Study of Hemoglobin Reactions and Structures', *Methods Enzymol.*, **232**, 139–175 (1994).
24. J.O. Alben, 'Fourier Transform Infrared Spectroscopy of Enzyme Systems', in *Infrared Spectroscopy of Biomolecules*, eds. H.H. Mantsch, D. Chapman, John Wiley & Sons, New York, 19–37, 1996.
25. K.J. Rothschild, 'FTIR Difference Spectroscopy of Bacteriorhodopsin: Toward a Molecular Model', *J. Bioenerg. Biomembr.*, **24**, 147–167 (1992).
26. W. Mäntele, 'Infrared Vibrational Spectroscopy of the Photosynthetic Reaction Center', in *The Photosynthetic Reaction Center*, eds. H. Deisenhofer, J.R. Norris, Academic Press, San Diego, CA, 239–283, Vol. II, 1993.
27. E. Nabedryk, 'Light-induced Fourier Transform Infrared Difference Spectroscopy of the Primary Electron Donor in Photosynthetic Reaction Centers', in *Infrared Spectroscopy of Biomolecules*, eds. H.H. Mantsch, D. Chapman, John Wiley & Sons, New York, 39–81, 1996.
28. F. Siebert, 'Equipment: Slow and Fast Infrared Kinetic Studies', in *Infrared Spectroscopy of Biomolecules*, eds. H.H. Mantsch, D. Chapman, John Wiley & Sons, New York, 83–106, 1996.
29. V. Cepus, C. Ulbrich, C. Allin, A. Troullier, K. Gerwert, 'Fourier Transform Infrared Photolysis Studies of Caged Compounds', *Methods Enzymol.*, **291**, 223–245 (1998).
30. P.T.T. Wong, K. Heremans, 'Pressure Effects on Protein Secondary Structure and Hydrogen Deuterium Exchange in Chymotrypsinogen: a Fourier Transform Infrared Spectroscopic Study', *Biochim. Biophys. Acta*, **956**, 1–9 (1988).
31. W.K. Surewicz, J.J. Leddy, H.H. Mantsch, 'Structure, Stability, and Receptor Interaction of Cholera Toxin as Studied by Fourier-transform Infrared Spectroscopy', *Biochemistry*, **29**, 8106–8111 (1990).
32. H. Fabian, C. Schultz, D. Naumann, O. Landt, U. Hahn, W. Saenger, 'Secondary Structure and Temperature-induced Unfolding and Refolding of Ribonuclease T1 in Aqueous Solution', *J. Mol. Biol.*, **232**, 967–981 (1993).
33. H. Fabian, C. Schultz, J. Backmann, U. Hahn, W. Saenger, H.H. Mantsch, D. Naumann, 'Impact of Point Mutations on the Structure and Thermal Stability of Ribonuclease T1 in Aqueous Solution Probed by Fourier Transform Infrared Spectroscopy', *Biochemistry*, **33**, 10 725–10 730 (1994).
34. H. Fabian, H.H. Mantsch, 'Ribonuclease A Revisited: Infrared Spectroscopic Evidence for Lack of Native-like Secondary Structure in the Thermally Denatured State', *Biochemistry*, **34**, 13 651–13 655 (1995).
35. D. Reinstädler, H. Fabian, J. Backmann, D. Naumann, 'Refolding of Thermally and Urea Denatured Ribonuclease A Monitored by Time-resolved FTIR Spectroscopy', *Biochemistry*, **35**, 15 822–15 830 (1996).
36. N.B. From, B.E. Bowler, 'Urea Denaturation of Staphylococcal Nuclease Monitored by Fourier Transform Infrared Spectroscopy', *Biochemistry*, **37**, 1623–1631 (1998).
37. P.I. Haris, D. Chapman, 'The Conformational Analysis of Peptides Using Fourier Transform IR Spectroscopy', *Biopolymers*, **37**, 251–263 (1995).
38. H. Fabian, G.I. Szendrei, H.H. Mantsch, B.D.Greenberg, L. Ötvös, Jr, 'Synthetic Post-translationally Modified Human Aβ Peptide Exhibits a Markedly Increased Tendency to Form β-Pleated Sheets in Vitro', *Eur. J. Biochem.*, **221**, 959–964 (1994).
39. M. Jackson, H.H. Mantsch, 'Biomembrane Structure from FT-IR Spectroscopy', *Spectrochim. Acta Rev.*, **15**, 53–69 (1993).
40. L.K. Tamm, S.A. Tatulian, 'Infrared Spectroscopy of Proteins and Peptides in Lipid Bilayers', *Q. Rev. Biophys.*, **30**, 365–429 (1997).
41. C.F.C. Lundlam, I.T. Arkin, X.-M. Liu, M.S. Rothman, P. Rath, S. Aimoto, S.O. Smith, D.M. Engelman, K.J. Rothschild, 'Fourier Transform Infrared Spectroscopy and Site-directed Isotope Labeling as a Probe of Local Secondary Structure in the Transmembrane Domain of Phospholamban', *Biophys. J.*, **70**, 1728–1736 (1996).
42. C.F.C. Lundlam, S. Sonar, C.-P. Lee, M. Coleman, J. Herzfeld, U.L. RajBhandary, K.J. Rothschild, 'Site-directed Isotope Labeling and ATR/FTIR Difference Spectroscopy of Bacteriorhodopsin: the Peptide Carbonyl Group of Tyr-185 is Structurally Active During the bR → N Transition', *Biochemistry*, **34**, 2–6 (1995).
43. M. Zhang, H. Fabian, H.H. Mantsch, H.J. Vogel, 'Isotope-edited Fourier Transform Infrared Spectroscopy Studies of Calmodulins's Interaction with its Target Peptides', *Biochemistry*, **33**, 10 883–10 888 (1994).
44. T. Li, T. Horan, T. Osslund, G. Stearns, T. Arakawa, 'Conformational Changes in G-CSF/Receptor Complex

as Investigated by Isotope-edited FTIR Spectroscopy', *Biochemistry*, **36**, 8849–8857 (1997).

45. I.T. Arkin, K.R. MacKenzie, L. Fisher, S. Aimoto, D.M. Engelman, S.O. Smith, 'Mapping the Lipid-exposed Surfaces of Membrane Proteins', *Nature Struct. Biol.*, **3**, 240–243 (1996).
46. C.P. Schultz, H. Fabian, H.H.Mantsch, 'Two-dimensional Mid-IR and Near-IR Correlation Spectra of Ribonuclease A: Using Overtone and Combination Modes to Monitor Changes in Secondary Structure', *Biospectroscopy*, **4**, S19–S29 (1998).
47. A.J. White, K. Drabble, C. Wharton, 'A Stopped-flow Apparatus for Infrared Spectroscopy of Aqueous Solutions', *Biochem. J.*, **306**, 843–849 (1995).
48. J. Backmann, H. Fabian, D. Naumann, 'Temperature-jump Induced Refolding of Ribonuclease A: a Time Resolved FTIR Spectroscopic Study', *FEBS Lett.*, **364**, 175–178 (1995).
49. R.B. Dyer, F. Gai, W.H. Woodruff, R. Gilmanshin, R.H. Callender, 'Infrared Studies of Fast Events in Protein Folding', *Acc. Chem. Res.*, **31**, 709–716 (1998).

Gel Electrophoresis in Protein and Peptide Analysis

Batia Kaplan
Sheba Medical Center, Tel Hashomer, Israel

1 Introduction 1
2 History 1
3 Methods of Gel Electrophoresis: Principles, Basic Techniques, and Equipment 3
3.1 One-dimensional Polyacrylamide Gel Electrophoresis 3
3.2 Isoelectrofocusing 4
3.3 Two-dimensional Electrophoresis 5
3.4 Peptide Mapping 6
3.5 Immunoelectrophoresis 6
4 Major Applications 6
4.1 Determination of the Molecular Mass of Proteins and Polypeptides by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis 6
4.2 Immunochemical Characterization of Proteins Using Western Blotting 8
4.3 Electrophoretic Separation of Complex Mixtures of Proteins and Peptides for their Chemical Analysis – Microsequencing and Mass Spectrometry 10
4.4 Detection of Enzyme Activity Following Gel Electrophoresis 12
4.5 Gel Electrophoresis in Protein Conformation and Quaternary Structure Studies 13
4.6 Gel Electrophoresis in Protein Interaction Studies 14
4.7 Gel Electrophoresis in Preparative Small-scale Purification of Proteins 15
Abbreviations and Acronyms 16
Related Articles 17
References 17

Gel electrophoresis is a widely used method in biochemical research; a number of different forms of gel electrophoresis have been developed and applied to the analysis of proteins and peptides. The most popular form is polyacrylamide gel electrophoresis (PAGE), in which proteins are separated within a gel matrix on the basis of differences in their charge density and/or relative molecular mass (M_r). Polyacrylamide gel is also widely used as a media for the generation of pH gradients in an isoelectrofocusing (IEF) technique, where separation of proteins is due to their differences in isoelectric point (pI). Two-dimensional (2D) electrophoresis is an extremely powerful separation technique, allowing differentiation of proteins on the basis of their pI in IEF, and their M_r in sodium dodecyl sulfate (SDS)/PAGE. Agarose gels are rarely used today in the electrophoretic analysis of proteins; their utility is mainly restricted to the immunoelectrophoretic techniques, which allow the characterization of proteins by their migration in gel and immunological properties.

This article provides a brief history of gel electrophoresis and describes the principles of major electrophoretic techniques. The utility of gel electrophoresis in different fields of biochemical research is also demonstrated. Topics included are analytical and small-scale preparative separations of complex proteins samples; determination of molecular mass of the electrophoretically separated proteins and polypeptides; electrophoretic micropreparation of a protein sample for its further immunochemical, enzymatic, and chemical analysis; study of protein interactions with different ligands; examination of protein unfolding and its quaternary structure. The advantages and limitations of the described techniques are discussed.

1 INTRODUCTION

Fifty years after the introduction of gel electrophoresis for protein separation, this technique continues to play one of the major roles in protein and peptide research. Different electrophoretic methods, developed using agarose or polyacrylamide as supporting media, have allowed separation of proteins on the basis of their differences in size, charge, and pI. During the 1990s, the importance of electrophoretic techniques, especially polyacrylamide electrophoresis, have increased significantly due to advances in the immunochemical and chemical methods of protein and peptide analysis. The combined use of high-resolution electrophoresis with sensitive immunodetection methods, microsequencing techniques, and mass spectrometry (MS), has become a powerful analytical approach in many biochemical studies. This article presents a brief history of gel electrophoresis, and describes the principles and applications of the major electrophoretic techniques presently used in the analysis of proteins and peptides.

2 HISTORY

Historically, the first form of electrophoresis was moving-boundary electrophoresis in free solution, in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

which an electric field applied to a solution of protein mixture resulted in migration of proteins to electrodes at a rate depending on the charge density of the molecules. This technique was modified to zone electrophoresis in free solution, where the proteins to be separated were placed as a narrow zone at a suitable distance from electrodes. This method was subsequently replaced by zone electrophoresis in solutions stabilized within supporting media, such as paper, cellulose acetate, thin-layer material (silica gel, alumina, cellulose), and gels of agar, agarose, and polyacrylamide. These supporting media minimized the deleterious effects of convection and diffusion associated with electrophoresis in free solution. They also allowed fixation of separated proteins at their final position immediately after electrophoresis. The use of gels as supporting media was of particular importance – in contrast to paper, cellulose acetate, or thin-layer materials, gels can be considered as porous media in which the separation of proteins is dependent on both charge density and size.

In the early 1950s, zone electrophoresis in agar gels was a popular technique for separation of serum proteins, especially when combined with the immunotechnique of Grabar and Williams.[1] However, owing to the presence of sulfate and carboxyl groups, agar gels suffered from marked electro-endosmosis and from protein losses due to absorption. Therefore, after the introduction of agarose – a neutral component of agar – in electrophoresis of proteins,[2] agar itself is now very seldom used in electrophoretic studies. The method of starch gel electrophoresis pioneered by Smithes in 1965[3] dramatically improved the resolution of serum proteins and was widely adopted for analytical purposes. Starch gels have pores of the same order as protein molecules, thus contributing an efficient molecular sieving effect. In this respect starch is advantageous over agarose, where pore size is sufficiently larger and thus the molecular sieving of most proteins is minimal. However, as starch is a natural product, its composition can vary affecting the gelling ability, resolution and reproducibility of the results. Therefore, after the introduction of polyacrylamide gel – a synthetic polymer of acrylamide monomer – for protein electrophoresis,[4] starch gel electrophoresis became almost entirely superseded by this technique. Similarly, electrophoretic separation of proteins in agarose gels was eclipsed by the use of polyacrylamide. However, agarose gel remains widely used in immunoelectrophoretic procedures and in the separation of very large molecules, such as nucleic acids and lipoproteins.

During the period from about 1960 to 1970 many different forms of PAGE were developed, including continuous and dicontinuous, dissociating and nondissociating buffer systems.[5–9] The technique of SDS/PAGE, first described by Shapiro et al. in 1967[7] and further modified by Laemmli in 1970,[9] has become one of the most widely used electrophoretic methods. Other remarkable events of this decade were the publication of the theory of IEF by Svensson in 1961[10] and invention of the procedure for the synthesis of carrier ampholytes (CAs) by Vesterberg and Svensson in 1964.[11] CAs, the synthetic amphoteric buffers, were used to generate and stabilize the pH gradient in IEF, thus allowing the proteins to be focused at their isoelectric points. The utilization of polyacrylamide gel matrix[12–14] made the IEF technique extremely popular. Remarkable progress in immunoelectrophoretic techniques has been made with the introduction of rocket[15] and crossed[16] immunoelectrophoresis.

The 1970s marked further rapid development of polyacrylamide slab gel electrophoretic techniques. High-resolution 2D electrophoresis combining IEF and SDS/PAGE, was reported in 1975.[17–19] A method of peptide mapping was developed which involved partial proteolysis of the electrophoretically separated proteins and further analysis of the resulting peptides by SDS/PAGE.[20] Silver staining, introduced in 1979 as a novel procedure for protein staining in gels, was 100-fold more sensitive than the commonly used Coomassie Blue R-25.[21] The method of electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets was described in 1979;[22] it gained widespread application in the immunochemical identification of proteins.

In the early 1980s, the first report on IEF in immobilized pH gradients (IPGs) appeared.[23] In the IPG technique, the buffers forming the pH gradient are covalently attached to the polyacrylamide matrix, thus enabling the generation of more reproducible and stable pH gradients in comparison to conventional IEF.

Since the mid-1980s, with the advent of gas-phase automated protein sequencing and the introduction of inert transfer membranes compatible with microsequencing procedure, gel electrophoresis has emerged as an extremely useful sample preparation technique for amino acid sequence analysis. Separation of complex protein mixtures by one-dimensional (1D) or 2D electrophoresis followed by electroblotting of proteins on to poly(vinyl difluoride) (PVDF™) membranes[24] became a common approach for direct N-terminal sequencing. The techniques have been described for the generation of peptide fragments in order to obtain information on internal sequences of gel-separated proteins.[25–28]

In the early 1990s, MS demonstrated its ability to analyze gel separated proteins, especially using 2D separations.[29,30] Both matrix-assisted laser desorption/ionization (MALDI)/MS and electrospray ionization (ESI)/MS provided routes to determine accurate masses

of intact proteins and peptide fragments. Computer programs have been developed to identify proteins by using the experimentally obtained peptide masses for peptide-mass database search.

At present, polyacrylamide slab gel electrophoresis, in conjunction with highly sensitive detection techniques, represents one of the most powerful analytical tools in protein and peptide research. Polyacrylamide gel matrix is also used in the rapidly developing technique of capillary electrophoresis (this issue is discussed in article **Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for**). In light of the interface of electrophoretic techniques with amino acid sequence and mass spectral analyses, gel electrophoresis has gained wide application as an analytical method and as an important microtechnique for the isolation and purification of proteins and peptides.

3 METHODS OF GEL ELECTROPHORESIS: PRINCIPLES, BASIC TECHNIQUES, AND EQUIPMENT

3.1 One-dimensional Polyacrylamide Gel Electrophoresis

3.1.1 Polyacrylamide Gels

The most widespread technique for electrophoretic analysis of proteins and peptides is zone electrophoresis in polyacrylamide gels, where the macromolecules are separated on the basis of their charge and/or size. Polyacrylamide gel is obtained by the polymerization of acrylamide monomers into long chains and their cross-linking using *N*,*N*′-methylene bisacrylamide. Polymerization is initiated by the addition of riboflavin or ammonium persulfate. *N*,*N*,*N*′,*N*′-tetramethylenediamine (TEMED) is added to accelerate this process by catalyzing the formation of free radicals from persulfate, which in turn initiate the polymerization. The effective pore size of the gels decreases as the acrylamide concentration increases. For any given total concentration of monomers (%T, i.e. gram acrylamide plus bisacrylamide per 100 mL), the effective pore size and physical properties of polyacrylamide gel vary depending on the proportion of the cross-linker (%C; concentration of crosslinker relative to %T, weight for weight).

3.1.2 Buffer Systems

Electrophoresis of proteins can be carried out under nondenaturing conditions in order to study native proteins, or under denaturing conditions by employing the dissociating agents, i.e. detergents (usually the anionic detergent SDS), urea, and disulfide cleavage agents (β-mercaptoethanol or dithiothreitol (DTT)). The SDS binds to different polypeptides in an approximately constant ratio of 1.4 g SDS per gram of polypeptide. The resulting SDS–polypeptide complexes have essentially identical densities of negative charge provided by the detergent, and thus migrate in PAGE according to their molecular mass (section 4.1). Urea acts by disrupting the hydrogen bonds. In contrast to SDS, urea does not affect the intrinsic charge of proteins, so the electrophoretic separation of the constituent polypeptides in presence of urea occurs on the basis of both charge and size. The systems employing both urea and SDS have been described and found to be effective in obtaining high resolution of proteins and small peptides (section 4.1).

The electrophoretic techniques also vary in respect of whether the buffers used are homogeneous (continuous) or multiphasic (discontinuous). In homogeneous buffer systems the same buffer ions are present at constant pH throughout the sample, gel and electrode vessel reservoirs. In multiphasic systems the composition and pH of the buffers are different in the electrode reservoir and in different parts of the gel. In most multiphasic systems a sample is loaded on to a large-pore stacking (upper) gel, in which the proteins are effectively concentrated prior to their separation in the small-pore resolving (lower) gel. In the Ornstein–Davies discontinuous system[5,6] the sample and the stacking gel contain Tris (tris(hydroxymethyl)aminomethane)–HCl (pH 6.7) buffer, whereas the reservoir buffer consists of Tris–glycine (pH 8.3). In continuous SDS/PAGE as described by Weber and Osborn[8] and by Swank and Munkres,[31] sodium phosphate buffer was applied. The popular Laemmli method[9] is based on the discontinuous Ornstein–Davies system with SDS present. The alternative SDS dissociating system described by Neville[32] employed Tris–borate buffer. The discontinuous Tris–Tricine (*N*-[tris(hydroxymethyl)methyl]-glycine)[33] and Tris–Bicine (*N*,*N*-bis(2-hydroxyethyl)-glycine)[34] SDS/PAGE methods have been developed, allowing effective separation of many proteins over a broad range of their molecular mass. The commonly used SDS discontinuous buffer systems are presented in Table 1.

The SDS/PAGE techniques described above are not always suitable in cases where the native protein conformation and biological activity must be preserved. Unfortunately, there is no universal buffer system ideal for the separation of native proteins. Contrary to SDS/PAGE, where the protein–SDS complexes are negatively charged over a wide range of pH, separation of proteins in nondissociating systems is highly dependent on pH. Here, changes in pH can alter the net charge of protein and affect negatively its separation, as well as its biological activity. Therefore, a large number of

systems have been designed for use at any pH; systems resolving proteins at pH 3.8,[6] pH 8.0,[35] and pH 9.5[36] are commonly used.[37]

3.1.3 Gel Concentration

The commonly used concentrations of polyacrylamide gels vary from 5 to 20% depending on size or charge of protein under study. For the analysis of proteins with an unknown molecular mass the optimal gel concentration can be determined experimentally by starting with 7.5% acrylamide gel and then checking the range from 5 to 15%. Protein mixtures can also be separated by using a concentration gradient in which the concentration of acrylamide increases in the direction of protein migration. Gradient gels are commonly used in SDS/PAGE and are especially useful when complex protein mixtures that cover a wide range of molecular mass must be displayed in a single gel.

3.1.4 Equipment

The equipment for both rod and slab gel electrophoresis is commercially available from many suppliers. However, in many cases the flat slab gel configuration is preferred. First, by using slab gels many samples can be run simultaneously under identical conditions in a single gel. Second, proteins run on the slab gel can be easily transferred to thin blotting membranes for their further analysis. The vertical slab gel apparatus available from Bio-Rad Laboratories and Hoeffer-Scientific Instruments are designed to run small-format slab gels (10×8 cm or 12×10 cm) and represent the most widely used models. NuPAGE electrophoretic systems including running cell, precast gels (of single and gradient concentration), ready-to-use buffers of neutral pH, and staining kits are now available from Novex (USA). This system provides rapid, reproducible and high-resolution separations under both dissociating and nondissociating conditions.

3.1.5 Detection

Various methods have been reported for the detection of proteins separated on polyacrylamide gels. In early studies the commonly used organic dye was Amido Black 108. Later it was replaced by the more sensitive Coomassie Blue R-250 (or G-250) allowing detection of 0.2–0.5 μg of any protein in a sharp band. With the introduction of silver staining[21] the detection sensitivity was increased up to 100 times. However, silver staining is a more laborious and expensive procedure as compared to Coomassie Blue. Silver staining can result in high backgrounds; in addition, it stains not only proteins, but also DNA, lipopolysaccharides, and polysaccharides. Different procedures have been described for detection of radioactive proteins separated on gels, which include autoradiography, fluorography and liquid scintillation counting. These methods are far more sensitive than staining methods for unlabeled proteins. It is remarkable that incorporation of the isotope does not alter the biological or enzymatic properties of proteins, nor change their net molecular charge.[38,39] The immunological methods for direct detection of proteins in a gel have now been superseded by the immunoblotting technique.[38] Immunoblotting, as well as the techniques for detection of enzymes following gel electrophoresis, are described in sections 4.2 and 4.4, respectively.

3.2 Isoelectrofocusing

This powerful technique is based on the electrophoretic separation of proteins according to their isoelectric points along a continuous pH gradient. Two major forms of IEF are presently used: conventional IEF utilizing CAs, namely CA/IEF, and IEF in an IPG.[39–43] CAs (known under different trade names such as Ampholine®, Pharmalyte®, Biolyte®, and Servalyte®) are synthetic amphoteric buffers – the oligoamino and oligocarboxylic acids – used to generate and stabilize the pH gradient over the pH range 3–10. The CA/IEF method is usually performed on a polyacrylamide matrix; however, for separation of very large proteins, agarose is preferable. The protein mixture is placed on a gel containing the CA buffers. When the electric field is applied, the small ampholytes migrate to their pI points, forming a continuous pH gradient in the gel, whereas the larger sample proteins begin a slower migration directed by the initial pH environment. Finally, each protein molecule reaches it's isoelectric point within the established pH gradient.[39] Special procedures have been adopted to improve the performance of this technique.[39,41,43] These include addition of specific amphoteric compounds (good and bad spacers) to a given pH interval for local flattening of the pH gradient. Good spacers focus into a narrow zone and require low concentrations to induce a pronounced flattening of the pH curve around their own pI. Poor spacers are used in high concentrations and form a broad plateau in their pI region. Another way to improve the separation consists of subfractioning of the carrier buffers into narrow pH cuts – commercially available ranges span two or more pH units, and the technologies for preparation of the narrower pH cuts have been developed.

The most serious limitations of conventional IEF include cathodic drift of the pH gradient, uneven conductivity, and protein precipitation at the pI value. Some of these problems have been solved with the development of IPG.[23] In contrast to CA/IEF, in the IPG technique

the IPG are generated prior to the electrophoretic step by casting a gradient gel with the aid of a two-vessel gradient mixer. The buffers, termed Immobilins™ (Pharmacia), represent eight weak acrylamide derivates containing either carboxyl or tertiary amino groups, supplemented by one strongly acidic and one strongly basic derivate. During polymerization these buffering species are covalently attached and immobilized via vinyl bonds to the backbone polyacrylamide, thus providing reproducible and indefinitely stable pH gradients. The protein mixture is placed on a gel with a preformed IPG; when the electric field is applied, the sample molecules migrate to their pI regions. Receipts have been calculated for the generation of narrow, ultranarrow and extended pH gradients.[39,41,42]

The protocols for casting polyacrylamide gels and the electrophoretic procedures used in IEF are described elsewhere.[39] Usually, the polyacrylamide composition varies from 7%T and 5%C to 4%T and 2.5%C. Commercial precast Immobiline™ gels are now available. The major problem common for both CA/IEF and IPG is the precipitation and smearing of proteins near their pI. To cure this problem, IEF is performed in the presence of urea (6–8 M) and nonionic or zwitterionic detergents. Encouraging results have been obtained when using as additives the nondetergent sulfobetains and a mixture of sugars and taurine.[43]

The equipment for CA/IEF and IPG is similar to that of PAGE.[39] Gel rod and vertical or horizontal slab gel systems are available (from LKB, Bio-Rad, and Pharmacia). The methods used for detection of proteins separated by IEF are also essentially the same as those for PAGE.[38,39] Ampholytes should be removed prior to Amido Black staining [by soaking the gel in 10–15% trifluoroacetic acid (TFA)].[38] The IPG-separated gels give strong backgrounds; therefore, in this case, medium-sensitivity stains are used.[39]

3.3 Two-dimensional Electrophoresis

2D electrophoresis methods are designed to separate complex polypeptide mixtures on the basis of a different molecular property in each dimension.[17] Typically, the most common 2D techniques utilize separation of proteins on the basis of charge by using IEF in the first dimension, and on the basis of molecular mass by using SDS/PAGE in the second dimension. The first dimension is performed by employing the CAs or IPG (these techniques are discussed above). The CA/IEF is usually carried out using cylindrical gels in capillary tubes as described elsewhere.[38,44] One of the major problems of this technique is protein–ampholyte interactions which cause the artifactual spots on 2D maps. The other problem is cathodic drift, which leads to pH instability and loss of basic proteins from 2D maps. Therefore, for the analysis of basic proteins, nonequilibrium pH gradient electrophoresis (NEPHGE) has been developed.[45] In this technique the polarity of the IEF apparatus is reversed – the sample is applied at anode, and the basic proteins migrate towards the basic end of the gel. Short time runs are used to avoid the gel running to equilibrium. Under these conditions the separation of proteins occurs on the basis of protein mobility in a rapidly-forming pH gradient. As the more acidic proteins are absent from the resulting 2D patterns, a combination of 2D patterns using equilibrium IEF and NEPHGE is helpful.

The introduction of IPG has provided a solution for pH instability occurring in conventional IEF. However, when using IPG for the first dimension in 2D separations, difficulties have been found in elution and transfer of proteins from IPG gels to the second-dimension SDS/PAGE gels (due to the presence of fixed charges on the Immobiline™ matrix leading to increased electro-endosmosis). This problem has been overcome by the development of a special standardized protocol: IPG gels on GelBond PAG™ support films are dried, cut into strips, and rehydrated in urea and nonionic or zwitterionic detergents.[46] Studies have also been carried out to optimize the pH gradients, because the narrow and ultranarrow pH gradients were not generally suitable for 2D separations.[47] Receipts are now available for producing wide pH gradients (pH 2.5–11), which can be effectively applied in 2D electrophoresis.[48] However, as the analysis of 2D maps obtained using wide-range IPG is difficult, the pH gradients spanning 3–4 pH units are commonly employed.

The second dimension is usually carried out by the Laemmli technique[9] using polyacrylamide of single or gradient concentration. The cylindrical IEF rod gel or IPG gel strip (bonded to plastic supports) is usually cemented on the top of the vertical SDS/PAGE slab gel with agarose. Using horizontal SDS/PAGE systems, the IPG strip is transferred on to the surface of the stacking gel.[44]

Although 2D separations are usually carried out by combining IEF with SDS/PAGE, other 2D systems are used for the separation of some types of proteins, such as histones or ribosomal proteins. These include discontinuous PAGE in the first dimension followed by continuous PAGE in the second, both dimensions containing urea; acid–urea PAGE in the first dimension with Triton–acid–urea PAGE in the second; acid–urea PAGE in the first and SDS/PAGE at high pH in the second dimension.[44]

Computerized systems designed for quantitative 2D gel analysis are now available. World Wide Web sites for 2D gel databases and imaging software developers are listed elsewhere.[49]

3.4 Peptide Mapping

Peptide mapping is obtained by breaking down proteins into a number of peptide fragments in a specific and controlled manner, separating the peptide mixture, and then comparing the separation pattern with that of standard proteins treated in the same way.[50] This can be done in many ways, by utilizing different protein fragmentation and separation techniques. Protein fragmentation can be accomplished enzymatically by proteases (such as trypsin, pepsin, or chymotrypsin) or chemically by employing chemical cleavage reagents (such as cyanogen bromide, hydroxylamine or *N*-chlorosuccinimide). For the separation of cleaved peptides the techniques of 1D or 2D electrophoresis, capillary zone electrophoresis, and either reversed-phase (RP) high-performance liquid chromatography (HPLC) or ion-exchange HPLC are generally used.

The utility of PAGE for peptide mapping was demonstrated in 1977 by Cleveland et al.:[20] SDS/PAGE was applied both for purification of proteins from complex protein mixtures, as well as for subsequent analysis of the enzymatically cleaved peptides. The pattern of peptide fragments obtained was characteristic of the protein substrate and the proteolytic enzyme. Although at the present time RP/HPLC is the most popular method for separation of small peptides, peptide mapping by gel electrophoresis is still widely used.

Three different strategies are currently applied for the fragmentation of proteins separated by gel electrophoresis: elution of proteins from the gel with subsequent fragmentation in solution; direct digestion of the protein in the gel; electroblotting of proteins on to supporting membranes and fragmentation on a membrane.[51] (Procedures for elution and electroblotting are described in sections 4.2 and 4.7.) Optimization of the cleavage conditions is especially important for peptide mapping studies. The commercially available proteinases vary in activity depending on their specificity and purity, upon the nature of substrate proteins, as well as on the temperature and pH. Thus, in many cases, preliminary experiments are required in order to establish the optimal hydrolysis conditions. Generally, complete fragmentation produces small fragments whose separation is more difficult, even when using high-resolution RP/HPLC. Partial proteolysis with a considerable proportion of larger peptides is often desirable, especially when using SDS/PAGE slab gel as a mapping gel. First, small peptides may be poorly resolved and run close to the buffer front. Second, they are poorly fixed and stained in the gel, and may be washed out. Therefore, in many instances it is also advisable to use chemical cleavage agents that target the low-frequency residues (Met or Trp). Use of the Tris–Tricine–SDS system[33] is preferable (instead of the Laemmli technique[9]) by allowing separation of the fragments of a lower molecular mass.

Separated proteins can be localized using the standard detection methods used for gel electrophoresis, namely by autoradiography or fluorography when the radioactively labeled proteins are run, or by using staining techniques for the unlabeled proteins (e.g. Coomassie R-250 or G-250 or silver staining).[37,38,50]

During the 1990s the importance of peptide mapping has increased due to the combination of this technique with amino acid sequencing and mass spectroscopy of peptide fragments (section 4.3).

3.5 Immunoelectrophoresis

Immunoelectrophoresis is the procedure in which proteins and other antigenic substances are characterized by both their electrophoretic migration in a gel (usually agarose) and their immunological properties. In classic immunoelectrophoresis[1] proteins are first separated by zone electrophoresis; antiserum is then placed in the central slot, and proteins are allowed to diffuse to the gel. Proteins are precipitated by their antibodies during diffusion; the precipitation lines are formed at positions where antigen to antibody ratio is optimal. In cross-immunoelectrophoresis[16] the protein sample is first subjected to gel electrophoresis in agarose, and then the separated proteins are driven into the gel containing antibodies by an electric field applied at right-angles to the initial separation. When a sufficient antibody is present the precipitates are formed, and the area enclosed by the precipitate is proportional to the amount of antigen. A major advantage of this technique is its potential for identification and resolution of a complex mixture of proteins. In rocket immunoelectrophoresis[15] antigen is electrophoresed in agarose gel containing a specific antibody. This results in the formation of a long rocket-like immunoprecipitate; the height of the rocket correlates with the amount of antigen, thus allowing quantification of a specific protein. Various modifications of this technique have been developed, such as fused rocket immunoelectrophoresis, line immunoelectrophoresis, and intermediate gel technique. The techniques are simple and inexpensive; some have gained wide application in clinical laboratories. The equipment and the procedures for immunoelectrophoresis techniques are described elsewhere.[52]

4 MAJOR APPLICATIONS

4.1 Determination of the Molecular Mass of Proteins and Polypeptides by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

PAGE can be used for determination of the molecular size of native, as well as denaturated proteins. In the

case of native proteins, where separation occurs on the basis of differences in size and charge, the M_r estimation technique relies on a mathematical canceling of charge effects following measurements of protein mobility in gels of different concentration. This is done by constructing the Ferguson plot,[53] a plot of $\log_{10}$ relative mobility (R_f) versus gel concentration – when %C is kept constant, a linear relationship can be obtained. The slope of the line is the retardation coefficient, K_r, which is a measure only of molecular size. A linear relationship between K_r and molecular mass is established, thus allowing construction of standard curves for determination of the M_r of native proteins. However, this method is only valid if the standard proteins used to generate the calibration curves have the same shape as the unknowns.

In an alternative approach for the M_r estimation, the differences in molecular charge are cancelled chemically. Proteins are treated with the anionic detergent SDS and then separated by SDS/PAGE by using the common Laemmli method.[9] The SDS binds to proteins at an approximately constant ratio (1.4 g SDS per gram of protein), that masks the intrinsic charge of polypeptide chain and forms SDS–protein complexes having essentially identical densities of negative charge. As SDS eliminates charge density differences amongst proteins, migration of the polypeptides in polyacrylamide gels depends only on their size and not on their charge. Under these conditions, a plot of $\log_{10}$ polypeptide M_r versus R_f reveals a linear relationship over a certain range of M_r. Thus the appropriately selected proteins can be used as markers for the construction of a standard curve for the M_r estimation of the denatured proteins. The $\log_{10}$ M_r of standard proteins is plotted against R_f, and the M_r of unknowns can be calculated on the basis of their R_f values. In the discontinuous buffer system of Laemmli[9] the linear relationship holds true for M_r values from 12 000 to 45 000 on 15% polyacrylamide gels, and from 16 000 to 60 000 on 10 and 5% gels. The commonly used M_r markers are commercially available (Sigma, Bio-Rad, Amersham).

As a linear relationship between M_r and mobility is essential for accurate M_r estimation, precautions should be taken to ensure maximal SDS binding in order to overwhelm the intrinsic charge of the polypeptide. It is therefore important to use an excess of SDS to polypeptide of at least 3:1. An excess of thiol reagent, usually β-mercapthoetanol or DTT, is needed to break disulfide bridges, which otherwise would prevent saturation of polypeptide with SDS.

The behavior of SDS-treated proteins can be analyzed by using Ferguson plots. Ideally, the intercept at 0%T, which is the apparent relative free mobility Y_0, could be identical for all SDS–polypeptide complexes having identical charge densities. In fact, many proteins behave "normally", i.e. they approximate this situation. However, some proteins (glycoproteins, very basic or very acidic proteins, small peptides) behave anomalously by demonstrating markedly different Y_0 values even in an excess of SDS and thiol reagent.

Significant improvement in the M_r determination was achieved by using gradients of an increased acrylamide concentration (instead of the gels of uniform concentration). During electrophoresis in the gradient gels, proteins migrate until the decreasing pore size impedes further process. The SDS/PAGE of proteins in linear gradient gels revealed a linear relationship between $\log_{10}$ M_r and $\log_{10}$ polyacrylamide concentration.[54] Thus, with a suitable set of standard proteins, calibration curves can be constructed from which the M_r of sample protein may be estimated (Figure 1). The M_r range available for M_r estimation depends on the conditions of gradient used: M_r 14 000–33 000 in a linear gradient from 7 to 25%T (1%C); M_r 14 000–210 000 in 5–20%T (2.6%C); M_r 13 000–950 000 in 3–30%T (8.4%C).[37] When compared with the uniform concentration gel, the gradient gels give sharper bands and allow the analysis of proteins over a wider M_r range. Furthermore, glycoproteins and

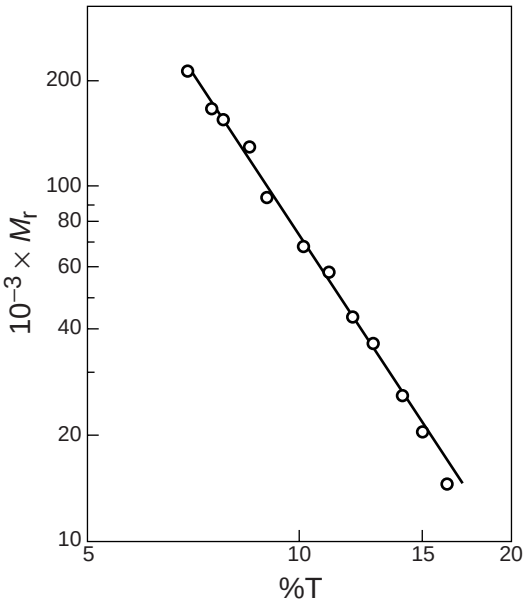

Figure 1 Calibration curves of $\log_{10}$ polypeptide molecular mass versus $\log_{10}$%T for a 5–20% linear gradient gel. Molecular mass markers: myosin, M_r 212 000; RNA polymerase β', M_r 165 000, and β, M_r 155 000, subunits; β-galactosidase, M_r 130 000; phosphorylase, M_r 92 500; bovine serum albumin, M_r 68 000; catalase, M_r 57 500; ovalbumin, M_r 43 000; glyceraldehyde-3-phosphate dehydrogenase, M_r 36 000; chymotrypsinogen A, M_r 25 750; soybean trypsin inhibitor, M_r 20 100; and lysosyme, M_r 14 300. (Reproduced from B.D. Hames[37] in *Gel Electrophoresis of Proteins*, eds. B.D. Hames, D. Rickwood, 1990, by permission of Oxford University Press.)

some proteins (such as papain, lysosyme, and ribonuclease) which behave anomalously in uniform concentration gels, demonstrate normal behavior in the gradient gels.[54] In general, gradient gels are especially useful in the study of complex mixtures of proteins varying significantly in their molecular sizes. However, when the proteins of interest are within a narrow M_r interval, uniform concentration gels may be superior.

Special SDS/PAGE techniques have been developed for the determination of M_r of small peptides, because the standard SDS/PAGE method[9] failed to resolve or provide useful size estimates for peptides below M_r 12 000. Swank and Munkres[31] demonstrated that inclusion of 8 M urea in the continuous buffer system and the increase in the proportion of cross-linker was effective in obtaining a linear relationship between mobility and M_r over the range M_r from 2400 to 17 000. The improved separation was obtained by the inclusion of urea in the discontinuous buffer system; this allowed the M_r estimation within the M_r range 2000–92 000.[55] Hashimoto et al.[56] utilized 10–18% gradient SDS gels in presence of urea (Table 1) which enabled good resolution of peptides down to M_r 1500. In general, these urea-containing systems were effective for separation of proteins and peptides. However, the presence of urea could cause band smearing of polypeptide bands of M_r above 40 000;[57] in addition, other disadvantages of urea arise from its instability, as well as its crystalization at low temperatures. Thus, the nonurea SDS/PAGE techniques employing high molarity Tris,[57] Tris–Tricine,[33] and Tris–Bicine[34] buffer systems have been introduced (Table 1), allowing separation of proteins and peptides ranging from M_r 1000 to 100 000, and estimation of their M_r over this broad M_r range (Figure 2).

Although the rapidly developing methods of mass spectroscopy permit more precise determination of the M_r of proteins and peptides as compared with the electrophoretic techniques, the latter continue to play an important role in different fields of biochemical research as they are simple, rapid and inexpensive.

4.2 Immunochemical Characterization of Proteins Using Western Blotting

4.2.1 General

The immunochemical characterization of proteins after their electrophoresis and subsequent electrotransfer to

Table 1 Commonly used SDS/PAGE discontinuous buffer systems

Reference	System
9	Stacking gel (3%T, 2.6%C) buffer: 0.125 M Tris–HCl (pH 6.8), 0.1% SDS Resolving gel (5–20%T, 2.6%C) buffer: 0.375 M Tris–HCl (pH 8.8), 0.1% SDS Reservoir buffer: 0.025 M Tris, 0.192 M glycine, 0.1% SDS Sample buffer: 0.625 M Tris–HCl (pH 6.8), 3% SDS, 1% β-mercaptoethanol, 0.2 M sucrose, 0.003% BPB[a]
56	Stacking gel (5%T, 2.4%C) buffer: 0.067 Tris–HCl (pH 6.8), 0.1% SDS Resolving gel (gradient 10.5–19%T, 4.7%C) buffer: 0.45 M Tris–HCl (pH 8.8), 0.1% SDS, 7 M urea, sucrose gradient 0–10% Reservoir buffer: 0.05 M Tris, 0.38 M glycine (pH 8.5), 0.1% SDS Sample buffer: 0.0625 M Tris–HCl (pH 6.8), 2% SDS, 10 mM DTT, 0.0025% BPB[a]
33	Stacking gel (4%T, 3%C) buffer: 0.75 M Tris–HCl (pH 8.45), 0.75% SDS Spacer gel (10%T, 3%C) buffer: 0.3 M Tris–HCl (pH 8.45), 0.1% SDS Resolving gel (10%T, 3%C, 16.5%T, 3%C, 16.5%T, 6%C, 16.5%T, 6%C, 6 M urea) buffer: 0.3 M Tris–HCl (pH 8.45), 0.1% SDS, glycerol 0.1 mL mL^{-1} gel solution Anode buffer: 0.2 M Tris–HCl (pH 8.8) Cathode buffer: 0.1 M Tris, 0.1 M Tricine (pH 8.25), 0.1% SDS Sample buffer: 50 mM Tris–HCl (pH 6.8), 4% SDS, 12% glycerol, 2% β-mercaptoethanol, 0.01% Serva Blue G[b]
34	Comb gel (7.5%T, 5%C) buffer: 0.359 M Bistris,[c] 0.159 M Bicine (pH 7.7), 0.1% SDS Stacking gel (6%T, 5%C) buffer: 0.4 M Bistris, 0.1M H_2SO_4 (pH 6.7), 0.1% SDS Resolving gel (12 or 18%T, 5%C) buffer: 0.4 M Tris, 0.1 M H_2SO_4 (pH 8.1), 0.1% SDS, glycerol 0.2 mL mL^{-1} gel solution Anode buffer: 0.2 M Tris, 0.05 M H_2SO_4 (pH 8.1) Cathode buffer: 0.2 M Bicine, 0.1 M NaOH (pH 8.2), 0.1% SDS Sample buffer: 0.359 M Bistris, 0.159 M Bicine (pH 7.7), 1% SDS, 2.5% β-mercaptoethanol, 15% sucrose, 0.004% BPB[a]

[a] BPB, bromophenol blue, tracking dye.
[b] Tracking dye.
[c] Bistris, [bis(2-hydroxyethyl)-amino]tris(hydroxymethyl)methane.

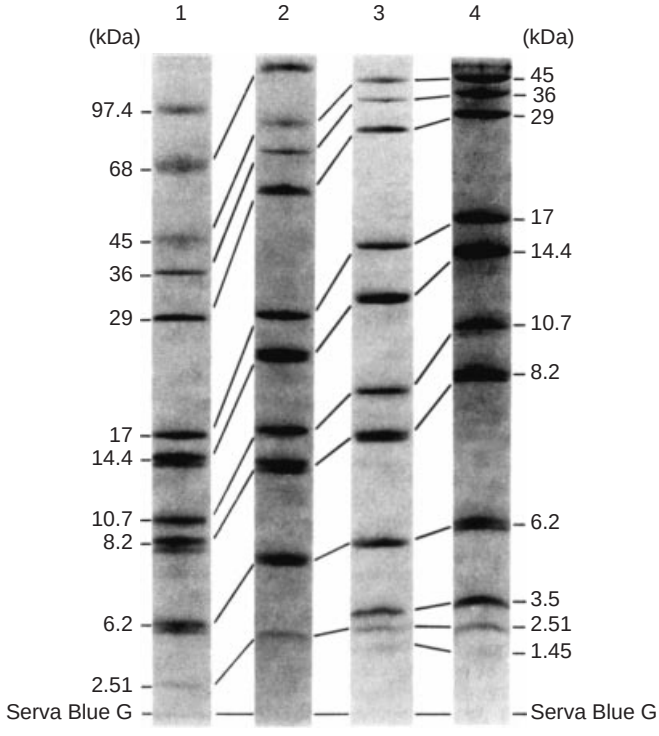

Figure 2 Protein SDS/PAGE using the discontinuous Tris–Tricine buffer system of Schagger and Von Jagow:[33] 1 = resolving gel of 10%T, 3%C; 2 = 16.5%T, 3%C; 3 = 16.5%T, 6%C; 4 = 16.5%T, 6%C, 6 M urea (see Table 1). (Reproduced from Schagger and Jagow[33] by permission of Academic Press.)

the support membranes has become one of the most popular techniques in protein analysis. The electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets was first described by Towbin et al.[22] This technique is now known as blotting (electroblotting) or Western blotting; when the electrotransfer is followed by antibody probing, the whole procedure is called immunoblotting. In general, introduction of the electroblotting technique essentially simplified the analysis of the electrophoresed proteins. The procedures for staining and destaining the blotted proteins are significantly shorter than in gels. The support membranes are easier to handle than the polyacrylamide gels; Western blots can be dried and stored for long periods prior to further analysis. Use of an appropriate blotting matrix allows multiple probing, i.e. examination of a single blot with several different detection systems. As a result, the immunological detection of proteins in gels is now almost completely replaced by immunoblotting. Finally, electroblotting has found especially wide application in amino acid sequencing and mass spectral analysis of proteins and peptides (section 4.3).

Various electrophoretic techniques of 1D and 2D electrophoresis are employed in Western blotting. Although electrophoresis is often carried out under denaturing conditions involving heat, reducing agents, SDS, and urea, which can reduce subsequent immunological activity, the experience usually shows no difficulties in the immunological detection of proteins. It is possible that the residual activity is sufficient to be detected by the sensitive methods used. It is also possible that in some cases the immunological activity is unaffected or even enhanced under the dissociating conditions.

4.2.2 Support Membranes

Among the variety of support matrices introduced in Western blotting the nitrocellulose membranes first applied by Towbin et al.[22] are still most widely used and are available from different suppliers (Schlecher & Schuell, Millipore, Amersham International, Bio-Rad). Nylon membranes (e.g. Zetaprobe from Bio-Rad, Hybond N from Amersham) have the advantage of being stronger and of higher binding capacity than nitrocellulose, but have the limitation of a high non-specific background with general stains (Coomassie Blue or Amido Black). Another type of supporting membrane includes diazo-modified papers; as the proteins are covalently bound to these matrices, a diazotized paper may be useful when a number of detection methods have to be applied sequentially to the same membrane. However, the diazotized paper matrices have lower binding capacity, they must be activated prior to their use, and are incompatible with the commonly used transfer buffers containing glycine (due to its interaction with diazo groups). At present, chemically stable PVDF™ membranes compatible with general protein stains and immunodetection methods are widely used (Millipore, Bio-Rad). Due to their chemical stability, the PVDF™ membranes have gained especially widespread application in amino acid sequence and mass spectral analyses (section 4.3).

4.2.3 Equipment and Transfer Buffers

Equipment for the electrotransfer of proteins is available commercially and includes tank-type apparatus (Bio-Rad, Hoeffer Scientific Instruments, Novex) and semidry electroblotting systems (Bio-Rad, Hoeffer Scientific Instruments, Millipore). The electric field is applied transversely to the gel, allowing migration of proteins from the gel to the membrane. In tank-type apparatus the gel–membrane sandwich is immersed in a buffer tank. The sandwich is prepared for sequentially placed layers: (1) porous pad (foam sponge or ScotchBrite scouring pad); (2) Whatman paper, 3MM; (3) gel; (4) membrane; (5) Whatman paper, 3MM; (6) porous pad. When the SDS gels are electroblotted at neutral or alkaline pH,

the membrane should face the anodic side. In contrast, under acidic conditions when the proteins behave as cations, the membrane should be at the cathodic side. The typical buffer systems used are 20 mM Tris–150 mM glycine (pH 8.3), 7.5 mM Tris–1.2 mM boric acid and 25 mM sodium phosphate (pH 6.5); methanol (20%) is often added because it minimizes the swelling of the gel during blotting and increases the binding capacity of nitrocellulose to protein.[37] As methanol reduces protein elution from the gels, 0.1% SDS is often incorporated in the transfer buffer to enhance the efficiency of elution. Another commonly employed transfer buffer is 10 mM 3-(cyclohexylamino-1-propanesulfonic) acid (CAPS)–10% methanol (pH 11) which is used for the electrotransfer of proteins to PVDF™ membranes.[51] The buffer system 25 mM Bicine–25 mM Bistris, pH 7.2, containing 1.02 mM ethylenediaminetetraacetic acid (EDTA) and 20% methanol, is recommended for electrotransfer of proteins to nitrocellulose, as well as to PVDF™ in Novex Western Transfer Apparatus, Model X Cell II.

In semidry transfer apparatus the voltage is applied through filter papers soaked in transfer buffer. Due to the high field strength, the transfer is rapid. It requires less transfer buffer than tank-type apparatus. Another advantage of semidry apparatus is the possibility of using multiple buffers by soaking filter papers in different solutions. As alkaline pH and SDS favor protein elution from gel, whereas methanol and acid pH favor their adsorption on membrane, the appropriate asymmetrical disposition of methanol, SDS, and pH on each side of the membrane can be extremely useful.[58–61] Highly efficient semidry transfer has been achieved by creating a stable pH boundary between the two faces of membrane – basic on the gel side (in the presence of SDS), and acidic on the membrane side (in the presence of methanol).[61]

4.2.4 Detection Systems

Once the proteins are electroblotted, they can be immunodetected by specific antibodies. As antibodies can be bound nonspecifically by the supporting membrane, the nonspecific binding sites should be blocked. Blocking is usually performed by incubating the membrane with bovine serum albumin or skim milk solution for several hours. The immunodetection procedure includes the following steps: incubation with the specific first antibodies, washing away the excess of unreacted antibodies, and incubation with second labeled antibodies directed against the first antibodies. Finally, the excess of unreacted antibodies is washed out, and the bound second-labeled antibodies are detected. In early studies radioiodinated and fluorescently labeled second antibodies were used and detected with autoradiography and ultraviolet illumination, respectively. Most studies are now carried out by enzyme labeling, usually horseradish peroxidase (HRP) and alkaline phosphatase (AP), which are visualized by converting suitable soluble substrates into colored precipitates in the presence of special color development reagents (such as 3,3′-diaminobenzidine or 4-chloro-1-naphthol). The advantage of the AP method is that the colored reaction product does not fade, in contrast to the HRP method. A variation of these methods is utilization of biotinylated antibodies that are recognized by avidin or by streptavidin labeled with an enzyme or fluorescent tag. Recent progress has been made with the introduction of the enhanced chemiluminiscence (ECL) Western blotting technique which is a light-emitting nonradioactive method for detection of the HRP-labeled antibodies. The commercially available substrate reagents include luminol/enhancer solution and a stable peroxidase buffer solution (Amersham, Pierce). When the blot containing the antigen bound to the HRP-labeled antibody is incubated with the mixture of these solutions, luminol is oxidated, resulting in emission of light which can be detected by exposure to autoradiography film. The recently developed ECL substrate systems (Pierce) allow detection of proteins at mid-femtogram and even at low femtogram levels.

4.3 Electrophoretic Separation of Complex Mixtures of Proteins and Peptides for their Chemical Analysis – Microsequencing and Mass Spectrometry

4.3.1 General

With the development of highly sensitive automated protein sequencing procedures in the 1980s[24–27] and after the introduction of MS for examination of proteins in 1985,[62,63] gel electrophoresis emerged as an extremely useful protein sample micropreparation technique for the subsequent determination of its amino acid sequence and mass. At the present time electrophoretic separation and purification of proteins remains one of the most widely used approaches, despite the recent progress made in the protein separation field, especially in development of new HPLC systems and capillary electrophoresis. There are three reasons for the popularity of gel electrophoresis:

- electrophoretic separations are reproducible and provide high resolution of proteins and peptides by appropriate selection of either 1D or 2D techniques;
- electrophoretic techniques are simple, rapid and inexpensive;
- proteins separated on gels can be blotted on the supporting membranes compatible with sequencing

and mass spectral procedures that in many instances facilitate protein analysis.

It is notable that the amount of purified proteins obtained by employing analytical scale gels is sufficient for their further analysis with highly sensitive microsequencing and mass spectral techniques, allowing detection of proteins and peptides at a picomole and subpicomole level (the latter techniques are described in the article **Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis**).

The electrophoretic purification of proteins used in amino acid sequencing and MS studies is commonly conducted by employing Tris–glycine[9] or Tris–Tricine[33] SDS/PAGE (sections 3.1 and 4.1). The 2D separations, which include IEF with IPG gels in the first dimension and SDS/PAGE in the second, are also widely used (section 3.2). Precautions should be taken to avoid the gel electrophoresis-induced modifications of proteins which could affect mass determination by MS. Some of these modifications are due to incomplete polymerization of acrylamide. Usually more than 30 mM of free acrylamide remain, which may form covalent adducts with cystein during gel electrophoresis. In addition, oxidation of methionine residues can occur, probably due to the presence of residual persulfate in the gel. Other modifications include formation of oxidized cystein–acrylamide and cystein–2-mercaptoethanol adducts. In order to prevent these undesirable modifications, several reagents are in use, such as mercaptoacetic acid, 3-mercaptopropanoic acid (in Tris–Tricine SDS/PAGE), glutathione or sodium thioglycolate (in SDS/PAGE), or free cystein (for IPG gels).[58,64]

4.3.2 Applications in Amino Acid Sequence Analysis

After electrophoresis, subsequent processing of the sample depends on whether N-terminal or internal amino acid sequence analysis is required. For N-terminal sequencing, the gel-separated proteins can be electroblotted on the supporting membranes (section 4.2) or electroeluted (section 4.7). Electroblotting is commonly preferred over electroelution because fewer manipulations are required.[24,25,58,65] After the introduction of PVDF™ as a sequencing substrate in 1987,[24] different types of PVDF™-based membranes were developed and applied for amino acid sequencing. High-retention PVDF™ membranes such as Immobilon-PSQ™ (Millipore), ProBlot™ (Applied Systems, Inc.), and Trans-Blot PVDF™ (Bio-Rad) are preferable over low-retention membranes, when the N-terminal sequence is required.[58,65]

The major problem in protein sequencing is the N-terminal blockage. In fact, about 80% of cellular proteins in eukaryotes are blocked at their N-termini; in most cases these are acetyl, formyl, and pyroglutamyl groups. Techniques for deblocking have been developed;[65] however, they are not always effective and therefore internal sequence analysis may be needed. For internal sequencing, the gel-separated proteins are cleaved enzymatically or chemically, and the cleavage products obtained are separated, usually by HPLC or gel electrophoresis, for their further analysis. The strategies applied for generation and isolation of peptides from the gel-separated proteins are the same as those described for peptide mapping (section 3.4). First, proteins can be electroeluted from the gel and then digested. Second, proteins can be in-gel digested and then the resulting peptides are eluted. Third, proteins can be blotted on to the membrane and then either eluted from membrane and digested, or digested on the membrane. The major advantage of using blotting membranes is easy removal of salts and detergents from proteins prior to their sequencing. However, for proteins of low blotting efficiencies, in-gel digestion or elution plus digestion can be recommended. In any case, SDS removal is of special importance because it can reduce the extent of enzymatic digestion, as well as interfere with the subsequent RP/HPLC separation of the resulting digest. When the proteins are eluted from the SDS gels, SDS is removed by trichloroacetic acid (TCA) or acetone precipitation; however, these procedures may be complicated when the protein concentration is too low or the SDS concentration is too high.[66,67] However, SDS removal from the PVDF™-blotted proteins is simple and can be achieved by washing the membrane by methanol or acetone or during staining/destaining procedures.[64] In many instances the low-retention membranes (such as Immobilon-P (Millipore) or Westran™ (Schleicher and Schuell, Inc.)),[58,65] as well as nitrocellulose (Schleicher and Schuell, Inc.)[26,68,69] have been found suitable for internal sequencing. Good recoveries have been obtained with membranes having attached charged groups, such as carboxymethylnitrocellulose[68] and Immobilon-CD.[58,70]

4.3.3 Applications in Mass Spectral Analysis

Utilization of MS for accurate mass determination of the gel-separated proteins and peptides has gained increasing popularity. Although SDS/PAGE is still widely used for M_r estimation, this method has an accuracy of about $\pm 10\%$[64] (section 4.1). In contrast to electrophoretic techniques, the high accuracy of MS techniques allows detection of mass differences between protein isoforms; MS analysis is also able to provide information on post-translational modifications of

protein, and makes possible determination of the number of proteins comigrating in a band or spot in the gel.

Both MALDI/MS and ESI/MS have been found useful for the analysis of intact proteins electroeluted from gel, and for proteins electroblotted on PVDF™ membranes. Removal of SDS is essential in these studies because it inhibits ionization in both MALDI/MS and ESI/MS.[29,30]

These MS techniques are applied to the analysis of enzymatic digests generated from gel-separated proteins. In general, the common methods of electroelution, or in-gel or on-membrane digestion are applicable to MS. However, most protocols used for in-gel digestions utilize neutral detergents (Tween-20™, Triton X-100™, NP-40™) to enhance the recovery of peptides. The presence of neutral detergents poses a significant problem in MS analysis in that they suppress ionization in MALDI and can interfere with ESI/MS analysis. Thus, several modifications have been introduced to eliminate the use of neutral detergents in the in-gel digestions.[64] The protocols used for on-membrane digestion usually employ additives prior to the enzymatic digestion in order to prevent absorption of enzyme to membrane (PVDF™-based membranes and nitrocellulose). Addition of polyvinylpyrrolidene, average M_r 40 000 (PVP-40) has enabled generation of samples compatible with MALDI/MS and with liquid chromatography (LC)/MS (ESI/MS coupled with HPLC). Triton X-100™ was effectively utilized to enhance the recovery of peptides from nitrocellulose and high-retention PVDF™; however, the above-mentioned problems in MS analysis were observed when using this neutral detergent. Some other useful modifications to increase the peptide recovery have been described, such as reduction and alkylation prior to digestion or addition of polyvinylpyrrolidene, average M_r 360 000 (PVP-360).[64] The MS-compatible protocols employing the charge-modified membranes and allowing good recovery of peptides have also been reported.[64,70,71]

Accurate mass determination of gel-separated peptides by MALDI/MS, ESI/MS, and LC/MS has been found useful in conjunction with peptide sequencing in the peptide-mapping studies. In fact, peptide mass data obtained by MS analysis can provide a unique fingerprint for that protein. Thus, when using specific cleavage rules, the peptide mass data obtained can be interpreted by consulting the existing protein sequence databases. These peptide-mass databases can be searched by using the experimentally obtained peptide masses to identify and characterize the protein; different computer programs have been developed and are now available commercially.[72]

4.4 Detection of Enzyme Activity Following Gel Electrophoresis

4.4.1 General

The detection of enzyme activity following gel electrophoresis is a popular technique in the analysis of crude enzyme preparations. There are two different approaches to the electrophoretic separation of enzyme proteins. Separation can be accomplished under the nondenaturing (native) conditions[6,36] that retain their enzymatic activity. Alternatively, electrophoresis can be carried out under denaturing conditions, such as in the presence of SDS and disulfide cleavage agents.[9] In this case, a renaturation step that includes SDS removal is usually needed after the electrophoresis; however, some enzymes retain activity even in the presence of SDS.[73,74] Although in some instances SDS/PAGE was found to be unsuitable for the detection of protein activity due to its irreversible denaturation,[75] SDS/PAGE has been successfully applied to the detection of a large number of different proteases (such as trypsin, chymotrypsin, collagenase, pronase, endopeptidase I, carboxypeptidases A and B), protein phosphatases, kinases, peroxidases and others. The popularity of SDS/PAGE in these studies is due to the high resolution power and reproducibility of this technique.

The different factors that influence the renaturation of enzymes separated by SDS/PAGE should be taken into consideration. These include the preparation of a sample prior to electrophoresis, the electrophoretic conditions, and the postelectrophoretic treatment. Some enzymes are successfully renaturated following the classical sample preparation procedure of heating the SDS–protein sample at 100 °C under reducing conditions. In other cases, however, much milder conditions are needed to retain the enzymatic activity (for example by preparing the SDS–protein complexes at lower temperature, by avoiding the use of thiol reducing agents, or by including the carrier proteins). Some studies include exogenous proteins in the gel polymerization mixture; it has been supposed that the included proteins act either as molecular chaperones, or as absorbents of heavy metals or other toxic materials.[74] Different strategies have been applied for SDS removal from the gel-separated proteins. These include washing the gel in aqueous buffers in absence or in presence of anion exchange resins,[76] adding the organic solvents,[77,78] or by using competing nonionic detergents[79,80] or chaotropic agents (e.g. urea or guanidine hydrochloride).[81–83]

4.4.2 Enzymatic Activity Assays

Different approaches are used for assaying the enzymatic activity. The assay can be carried out in situ on the gels,

or following the extraction of proteins (by diffusion or electroelution), or after electroblotting on to membranes (nitrocellulose or PVDF™). Removal of proteins from gel matrices allows easy application of different protocols for regeneration of enzymes, as well as for detection of the enzymatic activity. However, the extraction procedures are laborious and time-consuming; they can also lead to loss of resolution because slight differences in enzyme migration can be missed. The resolution of electrophoretically separated proteins can be preserved by electroblotting. However, the limitations of using this technique become apparent in enzyme detection assays in which the reaction products are soluble and are of low M_r, and hence cannot be confined to the region in which they were produced. Electroblotting techniques have been utilized for the detection of protein kinases which autophosphorylate and therefore act by themselves as the immobilized substrate/product; when [^{32}P] ATP (adenosine triphosphate) is used as the cosubstrate, the radiolabeled autophosphorylated protein kinases can be detected by autoradiography.[83–85] However, this autophosphorylation assay is now commonly carried out on gels in situ.[78] In general, the in situ on-gel assays have advantages over the assays performed following extraction of proteins or after their electroblotting for the following reasons. First, in situ assays require less manipulation and time, thus assisting in maintaining the original electrophoretic resolution. Second, they permit diffusion of small molecules into the polyacrylamide matrix, which is helpful in confining the enzymatic reaction products to the region they are produced.

Two major types of in situ assay are employed. The assay can be carried out by using diffusible substrates where the substrate molecules diffuse into the gel after the completion of electrophoresis, or by using substrate molecules immobilized within the gel prior to the electrophoretic run. Many assays, with diffusible substrates relying on the production of fluorogenic or chromogenic products, are used for the detection of hydrolases. Other popular assays are based on the reduction of tetrazolium salts to colored formazan, which is insoluble in water and precipitates close to the site of its formation. These assays were widely applied for detection of dehydrogenases. Other assays involving release and subsequent oxidation of indole derivates to form blue indigo dyes have been described.[37,74] Techniques employing immobilized substrates have been found effective for detection of enzymes acting on large biopolymers (e.g. proteases, nucleases, glycosylases). The macromolecules entrapped within the gel are cleaved by an enzyme, and the resulting fragments migrate out of the region; if the polymer is labeled (with a chromogenic, fluorogenic, or radioactive tag), the clear zones can become visible. The detailed description of these assays, their application to different enzymes classes, and numerous references are given in review articles.[73,74,85]

4.5 Gel Electrophoresis in Protein Conformation and Quaternary Structure Studies

Protein folding is one of the basic issues in the biochemistry of macromolecules. Under different denaturing conditions, such as rise in temperature, variations in pH, or addition of denaturants, proteins are unfolded and their native secondary and tertiary structure is lost. Unfolding transitions can be studied by any of the physicochemical techniques that are sensitive to changes in protein conformation.[86,87] The denaturant-induced unfolding leads to a significant increase in dimensions of polypeptide. In 1979, Creighton first applied the technique of transverse urea-gradient electrophoresis to monitor the denaturant-induced unfolding.[88] In this technique, protein is electrophoresed in a direction perpendicular to urea gradient (usually from 0 to 8 M), where the macromolecules are exposed to various urea concentrations at different positions of the gel. At the position where the concentration of urea is high enough to induce unfolding, the mobility of protein decreases due to the greater hydrodynamic volume of the unfolded form. After visualization of protein, the resulting electrophoretic pattern is interpreted as the unfolding curve. The analysis of the relationship between the transition rates from native to unfolded states and the electrophoretic patterns is described elsewhere.[87–90] As compared with other physicochemical methods used in protein folding studies,[86,87] transverse urea-gradient electrophoresis is a simple and inexpensive technique requiring only small amounts of protein. It was also found useful to follow the changes occurring in quaternary structure, i.e. dissociation in urea of the oligomeric proteins into monomers, where transition to faster-migrating species occurs at a particular urea concentration.[87,89,91–93] The dissociation may occur prior to or after unfolding; in some cases these two processes cannot be differentiated.

Different modifications to the basic technique have been described. In some studies both urea and Triton gradients were included;[94] other authors used the urea gradient in presence of SDS.[95] As the presence of different concentrations of urea may cause differences in the viscosity across the gel, the inverse gradient of the polyacrylamide concentration was used in some studies, e.g. 10–8%T, or 10–7.5%T for a 0–8 M urea gradient.[91,92] In most cases, homogenous buffer systems (of alkaline or acid pH) were used; however, applications of discontinuous systems have also been described.[87] For visualization of the electrophoretic patterns, general protein-staining techniques (e.g. Coomassie or silver staining) are useful when the analyzed sample contains a single purified

protein. When the sample contains several different proteins or when the functional protein properties are to be examined, specific detection techniques such as immunoblotting or enzyme activity assay are employed. Detailed information on the preparation of gels, running conditions, and detection systems is available.[96,97]

Protein unfolding can influence not only changes in volume, but also the surface charge. In some studies,[87] unfolding was followed by using IEF across the transfer urea gradient. Another useful electrophoretic technique was developed to monitor the conformational changes induced by heat denaturation. In this technique, first described by Thatcher and Hodson in 1981,[98] a temperature gradient was applied in the direction perpendicular to the electrophoretic run by using an aluminum heating jacket clamped around a vertical polyacrylamide slab gel.

Numerous biochemical applications have demonstrated the ability of these techniques to distinguish subtle differences in the behavior of various isomers and mutant proteins, in the analysis of folding/unfolding pathways, and to examine the dissociation of multimeric proteins. These studies have been reviewed.[87,90]

4.6 Gel Electrophoresis in Protein Interaction Studies

Proteins are able to bind a large number of biochemical, organic and inorganic compounds; different electrophoretic techniques have been employed to study these interactions.

Interaction may occur between proteins and ions of the electrophoresis buffer media. This has been observed when using borate buffer systems – borate ions interact reversibly with protein, resulting in the appearance of two zones for a single protein. Generation of multiple-protein zones due to interaction of proteins with buffers containing carboxylic and amino acids has also been observed. In general, these effects are undesirable, because the resulting multiplicity of bands may be misinterpreted as inherent heterogeneity of protein. In some cases, however, such interactions may be beneficial; thus, borate ion binding to glycoproteins allowed extension of the semilog relationship between migration distance and M_r in SDS/PAGE.[99,100]

The technique of counterion electrophoresis was originally developed to study calcium binding,[101] but it can also be applied to study protein interaction with other cationic or anionic ligands.[99] In this technique, radioactive Ca ions (added to the anode reservoir) migrate up to the gel in the direction of cathode, whereas the protein (placed in the cathodic end) migrates down the gel where it binds calcium. The system reaches a steady state in which calcium is uniformly distributed, and the amount of bound calcium can be determined by subtracting the apparent baseline concentration from the total calcium concentration in the protein zone.

Transverse urea concentration gradient electrophoresis (section 4.5) has been employed to follow the dissociation of protein bound to small charged ligand complexes, where removal of the small ligand resulted in changes in the electrophoretic pattern.[102]

IEF (section 3.2) has been effective in studying the binding of proteins with neutral molecules, such as oxygen or vitamin D. In these cases, the noncharged compounds affect the surface charge of binding protein and its pI.[100]

The binding of proteins with detergents is widely exploited in electrophoretic studies. The best example is SDS electrophoresis, where the stoichiometric binding of SDS to proteins allows estimation of M_r for many proteins and peptides (sections 3.1 and 4.1). In contrast to SDS, neutral detergents bind to proteins in different amounts depending on even slight differences in their hydrophobicities. Small additions of neutral detergents to urea-containing gel media of IEF may result in pI shifts, thus permitting discrimination of minute differences in protein structure.[103] Charge-shift electrophoresis[104] is another example in which detergents are used in electrophoretic separations. Electrophoresis is carried out in the presence of nonionic detergents, and the resulting patterns are used as references; test runs are performed by adding cationic or anionic detergents. Depending on the charge of added detergent, amphiphatic proteins display either anodal or cathodal shifts in their electrophoretic behavior. The immunoelectrophoretic techniques applied in these studies allowed identification of proteins in complex mixtures and provided information on the extent and nature of their interaction with detergents.

Affinity electrophoresis is applied to study protein interactions with different molecular mass ligands, such as lectins, enzymes, various binding proteins, and polysaccharides. The protein sample is electrophoresed under nondenaturing conditions in polyacrylamide gels (or occasionally in agarose) in the presence of ligands (free or immobilized) that interact with the protein. Under these conditions the mobility of protein is retarded as compared with a control run, i.e. in the absence of any ligand. The theory and measurements of the dissociation constants of the affinity interaction are described elsewhere.[38,99,100]

Electrophoretic methods are especially useful in studying interactions within the quarternary structure. Application of transverse urea-concentration gradient electrophoresis in analysis of protein subunit dissociation/association is described in section 4.6. Another popular and simple approach in subunit analysis is a comparison of the electrophoretic profiles obtained by SDS/PAGE separations in the presence and absence of thiol cleavage agents (sections 3.1 and 4.1). In some studies a similar approach was taken by utilizing IEF; in this case, the pI of the subunits resulting from the

reduction of protein usually differs from that of intact protein.[100]

The immunochemical techniques used to study antigen–antibody interactions are described in the previous sections. These include classic immunoelectrophoresis, rocket immunoelectrophoresis, crossed immunoelectrophoresis and other variations of these techniques (section 3.5), as well as immunoblotting (section 4.2).

The ligand-blotting technique is conceptually similar to immunoblotting. The electrophoresed proteins transferred on to membranes can be recognized by their binding with different ligands: low-M_r components, such as ^{45}Ca, ^{59}Fe, ^{14}C-phospholipids, iodinated proteins (e.g. insulin growth factor, tumor necrosis factor), and lectins conjugated with different reporter molecules have been reviewed.[100] As in immunoblotting, 1D or 2D electrophoretic separations are applied. The binding capacity is retained when using SDS/PAGE, but may be lost in the presence of thiol cleavage agents.

4.7 Gel Electrophoresis in Preparative Small-scale Purification of Proteins

Despite recent progress in the technology of chromatographic separations, especially HPLC, simple and inexpensive PAGE remains one of the most popular techniques for small-scale purification of proteins and peptides. In many cases preparative electrophoresis of a crude protein sample provides a level of purity that would otherwise require multiple chromatographic procedures. Although various designs of electrophoretic equipment are commercially available for large-scale purification purposes, at present they are rarely used for the preparative work. First, the resolution is lower than for analytical gel electrophoresis. Second, the amounts of protein purified using analytical-scale gels are sufficient for their analysis by modern analytical techniques.

Proteins are recovered from gels in different ways. They may be extracted (eluted) from the gels or electroeluted. Alternatively, they can be electrotransferred on to membranes and then extracted. Elution from gels is possible following protein staining, but usually staining of gels reduces the efficiency of extraction. Different procedures have been described for localization of proteins in unstained gels.[37,38] A simple protein localization in unstained gel is also possible using prestained molecular weight markers (Bio-Rad, Amersham); in this case, the electrophoretic mobility of the protein under study is determined by staining the gel in a preliminary run.

Buffers used in the extraction range procedures in most cases contain SDS (0.1%)[37,105] or urea (4–8 M).[38] The localized gel area is excised, and the gel slice is chopped and incubated in about three volumes of buffer overnight, stirring constantly. A repeated extraction is usually carried out the next day.

Electrophoretic elution can be performed using different devices, such as Electro-Elutor Model 422 (Bio-Rad) or Biodyalizer (Sialomed, USA). However, in many instances a more simple procedure is applied. The gel slice is transferred into a dialysis bag, which is then placed into a horizontal electrophoresis tank filled with elution buffer (e.g. 0.1 M sodium phosphate, 0.1% SDS (pH 7.4),[37] or 25 mM Tris, 192 mM glycine, 0.1% SDS (pH 8.3)[106–108]). In an electric field (20–200 V), proteins migrate out of the gel into the dialysis bag. Upon complete elution (2–8 h), the contents of the bag are centrifuged, the gel particles are removed, and the supernatant containing the eluted protein is collected.[37]

In an alternative approach to recover protein from the gel matrix, proteins are electrotransferred to PVDF™ or nitrocellulose (section 4.2) and then extracted from the membranes with detergents or organic solvents.[109,110] Use of volatile organic solvents, such as isopropanol, acetonitrile, formic acid, or TFA, is especially recommended, because they are readily removed by Speed-Vac without the need for dialysis.

The purification of proteins employing the elution-from-gel technique requires a subsequent SDS removal step which is not simple to perform (SDS cannot be completely removed from proteins by dialysis). In contrast, SDS removal is not a problem when the blotted proteins are recovered from membranes. First, SDS is partially dissociated from the protein during electrotransfer due to inclusion of methanol in a transfer buffer; second, SDS can be removed from membranes during the staining/destaining procedure (sections 4.2 and 4.3). However, the limitation of extraction-from-membrane techniques is that the efficiency of extraction may depend on the M_r and nature of the blotted protein, thus the extraction conditions should be carefully optimized for the proteins under study.

Several techniques have been developed and applied to SDS removal following the elution of proteins from gels. In early studies, the use of ion exchange resins in the presence of urea was found effective for SDS removal.[111] However, this procedure had to be followed by TCA precipitation of proteins or extensive dialysis to remove urea; this is not always suitable for the small amounts of proteins recovered from analytical-scale gels. The application of an ion-retardation resin (Bio-Rad, AGIIA8) with aqueous buffers containing no chaotropic agents has been reported;[112] however, this technique may not be appropriate for purification of insoluble hydrophobic proteins. A simple technique relying upon ion-pair extraction of SDS into a mixture of organic solvents (acetone, thiethylamine, acetic acid, and water, 85 : 5 : 5 : 5) and acetone precipitation of protein has been described.[113] At

present, techniques based on acetone or TCA precipitation are widely used.[66] However, they are efficient only at a protein concentration above 100 μg per milliliter and at SDS concentrations not higher than 0.05%. Thus, it might be necessary to reduce the SDS level by dialysis (especially after protein concentration by Speed-Vac) which is less appropriate if working with small amounts of protein. An SDS-removal technique based on dissociation of SDS–protein complexes in 50% acetonitrile–0.1% TFA solution has been reported, where proteins were purified by employing gel permeation chromatography on a Fractogel™ TSK HW-40 (F) column (Merck) in this aqueous organic solution.[114] Although this technique is more complex than acetone or TCA precipitation methods, it allows the proteins to be recovered in a soluble state thus avoiding the problems associated with acetone or TCA precipitation.

The most common form of electrophoresis used for preparative purposes is SDS/PAGE. However, this technique may not be efficient enough when different proteins of similar M_r have to be separated. 2D electrophoresis is an extremely powerful separation technique allowing differentiation of proteins on the basis of pI (in the IEF dimension) and M_r (in the SDS/PAGE dimension; section 3.3). Combination of 2D electrophoresis with a blotting procedure is widely used as a sample preparation technique for subsequent chemical and immunochemical analysis (sections 4.2 and 4.3). Unfortunately, 2D electrophoresis is less suitable for preparative work when using elution from gel techniques. The loading capacity (sample volume or protein amount) is low, and the protein bands in the second dimension are not sharp (as compared with those obtained by 1D separations); this makes it difficult to obtain efficient protein elution from gels. A new protocol for micropreparative purification of proteins employing 2D separation has been reported,[115] in which the protein mixture was separated by SDS/PAGE, blotted, and extracted from membranes. The material obtained from 10 blots was then subjected to 2D electrophoresis, thus significantly increasing the loading capacity of the sample.

Consecutive use of SDS/PAGE and chromatographic techniques, especially RP/HPLC, can be another useful approach in small-scale preparative separation of proteins. This approach combines the resolution power of analytical gels with the high separation speed of HPLC, and permits separation of proteins – first, on the basis of their differences in M_r, and, second, due to their differences in hydrophobicity. This technique has been developed[106] and applied to the small-scale purification of amyloid A proteins[107] and, for micropreparative separation of major acute-phase reactants of mice, the isomers of serum amyloid A (SAA1 and SAA2).[108] In these studies, proteins separated by SDS/PAGE were electroeluted from the gels, purified from SDS,[114] and applied to Vydac™ 214 TP54 RP/HPLC column (Alltech, USA).

Further improvements in the technology of 1D and 2D electrophoretic separations, and the development of new techniques combining gel electrophoresis and modern chromatography, could be promising for efficient purification and analysis of proteins and peptides.

ABBREVIATIONS AND ACRONYMS

AP	Alkaline Phosphatase
ATP	Adenosine Triphosphate
Bicine	*N*,*N*-Bis(2-hydroxyethyl)glycine
Bistris	[Bis(2-hydroxyethyl)-amino]tris-(hydroxymethyl)methane
BPB	Bromophenol Blue
CA	Carrier Ampholyte
CAPS	3-(Cyclohexylamino-1-propanesulfonic) Acid
DTT	Dithiothreitol
ECL	Enhanced Chemiluminiscence
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray Ionization
HPLC	High-performance Liquid Chromatography
HRP	Horseradish Peroxidase
IEF	Isoelectrofocusing
IPG	Immobilized pH Gradient
LC	Liquid Chromatography
MALDI	Matrix-assisted Laser Desorption/ Ionization
MS	Mass Spectrometry
NEPHGE	Nonequilibrium pH Gradient Electrophoresis
PAGE	Polyacrylamide Gel Electrophoresis
pI	Isoelectric Point
PVDF	Poly(Vinyl Difluoride)
PVP-40	Polyvinylpyrrolidene, Average M_r 40 000
PVP-360	Polyvinylpyrrolidene, Average M_r 360 000
RP	Reversed-phase
SDS	Sodium Dodecyl Sulfate
TCA	Trichloroacetic Acid
TEMED	*N*,*N*,*N*′,*N*′-tetramethylenediamine
TFA	Trifluoroacetic Acid
Tricine	*N*-[Tris(hydroxymethyl)methyl]glycine
Tris	Tris(hydroxymethyl)aminomethane
1D	One-dimensional
2D	Two-dimensional
%C	Proportion of the Cross-linker
%T	Total Concentration of Monomers

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Proteolytic Mapping • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

REFERENCES

1. P. Grabar, C.A. Williams, 'Methode Permettant L'Etude Conjuguee des Proprietes Electrophoretiques et Immunochimiques d'un Melange de Proteines. Application au Serum Saguin', *Biochim. Biophys. Acta*, **10**, 193–194 (1953).
2. S. Hjerten, 'Agarose as an Anticonvection Agent in Zone Electrophoresis', *Biochim. Biophys. Acta*, **53**, 514–517 (1961).
3. O. Smithies, 'Zone Electrophoresis in Starch Gels: Group Variations in the Serum Proteins of Normal Human Adults', *Biochem. J.*, **61**, 624–641 (1955).
4. S. Raymond, L.S. Weintroub, 'Acrylamide Gel as a Supporting Medium for Zone Electrophoresis', *Science*, **130**, 711 (1959).
5. L. Ornstein, 'Disc Electrophoresis – I. Background and Theory', *Ann. N.Y. Acad. Sci.*, **121**, 321–349 (1964).
6. B.J. Davies, 'Disc Electrophoresis – II. Method and Application to Human Serum Proteins', *Ann. N.Y. Acad. Sci.*, **1**, 404–427 (1964).
7. A.L. Shapiro, E. Vinuela, J.V. Maizel, 'Molecular Weight Estimation of Polypeptide Chains by Electrophoresis in SDS-Polyacrylamide Gels', *Biochem. Biophys. Res. Commun.*, **28**(5), 815–820 (1967).
8. K. Weber, M. Osborn, 'The Reliability of Molecular Weight Determination by Dodecyl Sulfate Polyacrylamide Gel Electrophoresis', *J. Biol. Chem.*, **244**, 4406–4412 (1969).
9. U.K. Laemmli, 'Cleavage of Structural Proteins During Assembly of the Head of the Bacteriophage', *Nature*, **227**, 680–685 (1970).
10. H. Svensson, 'Isoelectric Fractionation, Analysis, and Characterization of Ampholytes in Natural pH Gradients. I. The Differerential Equation of Solute Concentrations at a Steady State and its Solution for Simple Cases', *Acta Chem. Scand.*, **15**, 325–341 (1961).
11. O. Vesterberg, H. Svensson, 'Isoelectric Fractionation, Analysis, and Characterization of Ampholytes in Natural pH Gradients. IV. Further Studies on Resolving Power in Connection with Separation of Myoglobins', *Acta Chem. Scand.*, **20**(3), 820–834 (1966).
12. G. Dale, A.L. Latner, 'Isoelectric Focusing in Polyacrylamide Gels', *Lancet*, **1**(7547), 847–848 (1968).
13. Z.L. Awdeh, A.R. Williamson, B.A. Askonas, 'Isoelectric Focusing in Polyacrylamide Gels and its Application to Immunoglobulins', *Nature*, **219**, 66–67 (1968).
14. P.G. Righetti, J.W. Drysdale, 'Isoelectric Focusing in Polyacrylamide Gels', *Biochim. Biophys. Acta*, **236**, 17–28 (1971).
15. C.B. Laurell, 'Quantitative Estimation of Proteins by Electrophoresis in Agarose Gel Containing Antibodies', *Anal. Biochem.*, **15**, 45–52 (1966).
16. H.G.M. Clarke, T. Freeman, 'Quantitative Immunoelectrophoresis of Human Serum Proteins', *Clin. Sci.*, **35**, 403–413 (1968).
17. P.H. O'Farrell, 'High Resolution Two-dimensional Electrophoresis of Proteins', *J. Biol. Chem.*, **250**, 4007–4021 (1975).
18. J. Klose, 'Protein Mapping by Combined Isoelectric Focusing and Electrophoresis of Mouse Tissues. A Novel Approach to Testing for Induced Mutations in Mammals', *Humangenetik*, **26**, 231–243 (1975).
19. G.A. Scheele, 'Two-dimensional Gel Analysis of Soluble Proteins. Characterization of Guinea Pig Exocrine Pancreatic Proteins', *J. Biol. Chem.*, **250**(14), 5375–5385 (1975).
20. D.W. Cleveland, S.G. Fischer, M.W. Kirschner, U.K. Laemmli, 'Peptide Mapping by Limited Proteolysis in Sodium Dodecyl Sulfate and Analysis by Gel Electrophoresis', *J. Biol. Chem.*, **252**, 1102–1106 (1977).
21. R.C. Switzer, C.R. Merril, S. Shifrin, 'A Highly Sensitive Silver Stain for Detecting Proteins and Peptides in Polyacrylamide Gels', *Anal. Biochem.*, **98**, 231–237 (1979).
22. H. Towbin, T. Staehelin, J. Gordon, 'Electrophoretic Transfer of Proteins from Polyacrylamide Gels to Nitrocellulose Sheets: Procedure and Some Applications', *Proc. Natl. Acad. Sci. USA*, **76**(9), 4350–4354 (1979).
23. B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Gorg, R.J. Westermeier, W. Postel, 'Isoelectric Focusing in Immobilized pH Gradients: Principles, Methodology and Some Applications', *J. Biochem. Biophys. Methods*, **6**, 317–379 (1982).
24. P. Matsudaira, 'Sequence from Picomole Quantities of Proteins Electroblotted onto Polyvinylidene Difluoride Membranes', *J. Biol. Chem.*, **262**(21), 10035–10038 (1987).
25. G. Bauw, M. De Loose, D. Inze, M. Van Montagu, J. Vandekerckhove, 'Alterations in the Phenotype of Plant Cells Studied by NH_2 Terminal Amino Acid-Sequence Analysis of Proteins Electroblotted from Two-Dimensional Gel-Separated Total Extracts', *Proc. Natl. Acad. Sci. USA*, **84**, 4806–4810 (1987).
26. R.H. Aebersold, J. Leavitt, R.A. Savedra, L.E. Hood, S.B.H. Kent, 'Internal Amino Acid Sequence Analysis of Proteins Separated by One- or Two-dimensional Gel Electrophoresis After In Situ Protease Digestion

on Nitrocellulose', *Proc. Natl. Acad. Sci. USA*, **84**, 6970–6974 (1987).
27. G. Bouw, J.V. Damme, M. Puype, J. Vandekerckhove, G. Gesser, G.P. Ratz, J.B. Lauridsen, J.E. Celis, 'Protein Electroblotting and Microsequencing Strategies in Generating Protein Data Bases from Two-dimensional Gels', *Proc. Natl. Acad. Sci. USA*, **86**, 7701–7705 (1989).
28. J.R. Vanfleteren, J.G. Raymackers, S.M. Van Bun, L.A. Meheus, 'Peptide Mapping and Microsequencing of Proteins Separated by SDS/PAGE After Limited In Situ Acid Hydrolysis', *Biotechniques*, **12**(4), 550–557 (1992).
29. M. Le Maire, S. Deschamps, J.V. Moller, J.P. Le Caer, J. Rossier, 'Electrospray Ionization Mass Spectrometry on Hydrophobic Peptides Electroeluted from Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis. Application to the Topology of the Sarcoplasmic Reticulum Ca^{2+} ATPase', *Anal. Biochem.*, **214**, 50–57 (1993).
30. S. Haebel, C. Jensen, S.O. Andersen, P. Roepstorff, 'Isoforms of a Cuticular Protein from Larvae of the Meal Beetle, Tenebrio Molitor; Studies by Mass Spectrometry in Combination with Edman Degradation and Two-dimensional Polyacrylamide Gel Electrophoresis', *Protein Sci.*, **4**, 394–401 (1995).
31. R.T. Swank, K.D. Munkres, 'Molecular Weight Analysis of Oligopeptides by Electrophoresis in Polyacrylamide Gel with Sodium Dodecyl Sulfate', *Anal. Biochem.*, **39**, 462–477 (1971).
32. D.M. Neville, 'Molecular Weight Determination of Protein – Dodecyl Sulfate Complexes by Gel Electrophoresis in Discontinuous Buffer Systems', *J. Biol. Chem.*, **246**(20), 6328–6334 (1971).
33. H. Schagger, G. von Jagow, 'Tricine–Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis for the Separation of Proteins in the Range from 1 to 100 kDa', *Anal. Biochem.*, **166**, 368–379 (1987).
34. J. Wiltfang, N. Arold, V. Neuhoff, 'A New Multiphasic Buffer System for Sodium Dodecyl Sufate/Polyacrylamide Gel Electrophoresis of Proteins and Peptides with Molecular Masses 100 000–1000, and their Detection with Picomolar Sensitivity', *Electrophoresis*, **12**, 352–366 (1991).
35. D.E. Williams, R.A. Reisfeld, 'Disc Electrophoresis in Polyacrylamide Gels: Extension to New Conditions of pH and Buffers', *Ann. NY Acad. Sci.*, **121**, 373–381 (1964).
36. R.A. Reisfeld, V.J. Lewis, D.E. Williams, 'Disc Electrophoresis of Basic Protein and Peptides on Polyacrylamide Gels', *Nature*, **195**, 281–283 (1962).
37. B.D. Hames, 'One-dimentional Polyacrylamide Gel Electrophoresis', in *Gel Electrophoresis of Proteins. A Practical Approach*, 2nd edition, eds. B.D. Hames, D. Rickwood, IRL Press, Oxford, 1–148, 1990.
38. A.T. Andrews, *Electrophoresis: Theory, Techniques, and Biochemical and Clinical Applications*, 2nd edition, eds. A.R. Peacocke, W.F. Harrington, Clarendon Press, Oxford, 1–452, 1988.
39. P.G. Righetti, E. Gianazza, C. Gelfi, M. Chiari, 'Isoelectric Focusing', in *Gel Electrophoresis of Proteins. A Practical Approach*, 2nd edition, eds. B.D. Hames, D. Rickwood, IRL Press, Oxford, 149–216, 1990.
40. P.G. Righetti, C. Gelfi, M. Chiari, 'Isoelectric Focusing in Immobilized pH Gradients', *Methods Enzymol.*, **270**, 235–255 (1996).
41. P.G. Righetti, A. Bossi, C. Gelfi, 'Conventional Isoelectric Focusing in Gel Slabs in Capillaries, and Immobilized pH Gradients', in *Gel Electrophoresis of Proteins, a Practical Approach*, 3rd edition ed. B.D. Hames, Oxford University Press, New York, 127–187, 1998.
42. P.G. Righetti, *Immobilized pH Gradients. Theory and Methodology*, Elsevier, Amsterdam, The Netherlands, 1–400, 1990.
43. P.G. Righetti, A. Bossi, 'Isoelectric Focusing of Proteins and Peptides in Gel Slab and in Capillaries', *Anal. Chim. Acta*, **372**, 1–20 (1998).
44. O. Rickwood, J.A.A. Chambers, S.P. Spragg, 'Two-dimensional Gel Electrophoresis', in *Gel Electrophoresis of Proteins, a Practical Approach*, 2nd edition, eds. B.D. Hames, D. Rickwood, IRL Press, Oxford, 217–272, 1990.
45. P.Z. O'Farrel, H.M. Goodman, P.H. O'Farrel, 'High Resolution Two-dimensional Electrophoresis of Basic, as Well As Acidic Proteins', *Cell*, **12**, 1133–1142 (1977).
46. A. Gorg, W. Postel, S. Gunter, 'The Current State of Two-dimensional Electrophoresis with Immobilized pH Gradients', *Electrophoresis*, **9**, 531–546 (1988).
47. M.J. Dunn, Y.M. Corbett, 'Two-dimensional Polyacrylamide Gel Electrophoresis', *Methods Enzymol.*, **271**, 177–203 (1996).
48. P. Sinha, E. Kottgen, R. Westermeier, P.G. Righetti, 'Immobilized pH 2.5–11 Gradients for Two-dimensional Electrophoresis', *Electrophoresis*, **13**, 210–214 (1992).
49. S.M. Hanash, 'Two-dimensional Gel Electrophoresis', in *Gel Electrophoresis of Proteins, a Practical Approach*, 3rd edition, ed. B.D. Hames, Oxford University Press, New York, 189–211, 1998.
50. A.T. Andrews, 'Peptide Mapping', in *Gel Electrophoresis of Proteins, a Practical Approach*, 2nd edition, eds. B.D. Hames, D. Rickwood, IRL Press, Oxford, 301–320, 1990.
51. H. Charbonneau, 'Strategies for Obtaining Partial Amino Acid Sequence Data from Small Quantities (<500 pmol) of Partially Purified Protein', in *A Practical Guide to Protein and Peptide Purification For Microsequencing*, 2nd edition, ed. P. Matsudara, Academic Press, San Diego, 17–44, 1993.
52. T.C. Bog-Hansen, 'Immunoelectrophoresis', in *Gel Electrophoresis of Proteins, a Practical Approach*, 2nd edition, eds. B.D. Hames, D. Rickwood, IRL Press, Oxford, 273–300, 1990.
53. K.A. Ferguson, 'Starch-Gel Electrophoresis – Application to the Classification of Pituitary Proteins and Polypeptides', *Metabolism*, **13**, 985–1002 (1964).

54. P. Lambin, 'Reliability of Molecular Weight Determination of Proteins by Polyacrylamide Gel Gradient Electrophoresis in the Presence of Sodium Dodecyl Sulfate', *Anal. Biochem.*, **85**, 114–125 (1978).
55. B.L. Anderson, R.W. Berry, A. Telser, 'A Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis System that Separates Peptides and Proteins in the Molecular Weight Range of 2500 to 90 000', *Anal. Biochem.*, **132**, 365–375 (1983).
56. F. Hashimoto, T. Horigana, M. Kanbayashi, K. Yoshida, H. Sugano, 'An Improved Method for Separation of Low-molecular-weight Polypeptides in Electrophoresis in Sodium Dodecyl Sulfate/Polyacrylamide Gels', *Anal. Biochem.*, **129**, 192–199 (1983).
57. S.P. Fling, D.S. Gregerson, 'Peptide and Protein Molecular Weight Determination by Electrophoresis Using High-molarity Tris Buffer System Without Urea', *Anal. Biochem.*, **155**, 83–88 (1986).
58. S.D. Patterson, 'From Electrophoretically Separated Protein to Identification: Strategies for Sequence and Mass Analysis', *Anal. Biochem.*, **221**, 1–15 (1994).
59. R.T.M. Vaessen, J. Kreike, G.S.P. Groot, 'Protein Transfer to Nitrocellulose Filters. A Simple Method for Quantitation of Single Proteins in Complex Mixtures', *Febs Lett.*, **124**, 193–196 (1981).
60. M. Svoboda, S. Meuris, C. Robyn, J. Christopher, 'Rapid Electrotransfer of Proteins from Polyacrylamide Gel to Nitrocellulose Membrane Using Surface/Conductive Glass as Anode', *Anal. Biochem.*, **151**, 16–23 (1985).
61. M. Laurerie, 'A Semidry Electroblotting System Efficiently Transfers Both High-and Low-molecular-weight Proteins Separated By SDS/PAGE', *Anal. Biochem.*, **212**, 206–211 (1993).
62. R.T. Covey, R.F. Bonner, B.I. Shushan, J. Henion, 'The determination of Protein, Oligonucleotide and Peptide Molecular Weights by Ion Spray Mass Spectrometry', *Rapid Comm. Mass Spectrom.*, **2**, 249–259 (1988).
63. G.B. Fenn, M. Mann, C.K. Heng, S.F. Wong, G.M. Whitehouse, 'Electrospray Ionization of Mass Spectrometry of Large Molecules', *Science*, **246**, 64–71 (1989).
64. S.D. Patterson, R. Aebersold, 'Mass Spectrometric Approaches for the Identification of Gel-separated Proteins', *Electrophoresis*, **16**, 1791–1814 (1995).
65. N. Le Gendre, M. Mansfield, A. Weiss, P. Matsudaira, 'Purification of Proteins by SDS/PAGE', in *A Practical Guide to Protein and Peptide Purification for Microsequencing*, 2nd edition, ed. P. Matsudara, Academic Press, 74–104, 1993.
66. K. Stone, K.R. Williams, 'Enzymatic Digestion of Proteins and HPLC Peptide Isolation', in *A Practical Guide to Protein and Peptide Purification For Microsequencing*, 2nd edition, ed. P. Matsudara, Academic Press, San Diego, 45–73, 1993.
67. B. Kaplan, 'Application of Polyacrylamide Slab Gel Electrophoresis to the Analysis and Small-scale Purification of Amyloid Proteins', *Anal. Chim. Acta*, **372**, 161–172 (1998).
68. R. Aebersold, 'Internal Amino Acid Sequence Analysis of Proteins After In Situ Protease digestion on Nitrocellulose', in *A Practical Guide to Protein and Peptide Purification For Microsequencing*, 2nd edition, ed. P. Matsudara, Academic Press, San Diego, 105–126, 1993.
69. E. Alimi, A. Martinage, P. Sautiere, P. Shevaillier, 'Electroblotting Proteins onto Carboxymethylcellulose Membranes for Sequencing', *BioTechniques*, **15**, 912–917 (1993).
70. S.D. Patterson, D. Hess, T. Yungwirth, R. Aeberson, 'High-yield Recovery of Electroblotted Proteins and Cleavage Fragments from a Cationic Polyvinidene Fluoride-based Membrane', *Anal. Biochem.*, **202**, 193–203 (1992).
71. D. Hess, T.C. Covey, R. Winz, R.W. Browney, R. Aeberson, 'Analytical and Micropreparative Peptide Mapping by High Performance Liquid Chromatography/Electrospray Mass Spectrometry of Proteins Purified by Gel Electrophoresis', *Protein Sci.*, **2**, 1342–1351 (1993).
72. R.Y. Simpson, G.E. Reid, 'Sequence Analysis of Gel-resolved Proteins', in *Gel Electrophoresis of Proteins, a Practical Approach*, 3rd edition, ed. B.D. Hames, Oxford University Press, 237–267, 1998.
73. O. Gabriel, D.M. Gerstein, 'Staining for Enzymatic Activity after Gel Electrophoresis', *Anal. Biochem.*, **203**, 1–21 (1992).
74. K.M. Bischop, L. Shi, P.J. Kennelly, 'The Detection of Enzyme Activity Following Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis', *Anal. Biochem.*, **260**, 1–17 (1998).
75. D. Ruggiero-Lopez, P. Louisot, A. Martin, 'A Nondenaturing Preparative Gel Electrophoresis System for the Recovery of Functional Proteins. Application to the Identification of an Endogenous Protein Inhibitor of Fucosyl-transferase Activities', *Anal. Biochem.*, **212**, 247–252 (1993).
76. K. Weber, D.J. Kuter, 'Reversible Denaturation of Enzymes by Sodium Dodecyl Sulfate', *J. Biol. Chem.*, **246**, 4505–4509 (1971).
77. A. Blank, R.H. Sugiyama, C.A. Dekker, 'Activity Staining of Nucleolytic Enzymes after Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis: Use of Aqueous Isopropanol to Remove Detergents from Gels', *Anal. Biochem.*, **120**, 267–275 (1982).
78. J.E. Hutchcroft, M. Anostario, Jr, M.L. Harrison, R.L. Geahlen, 'Renaturation and Assay of Protein Kinases After Electrophoresis in Sodium Dodecyl Sulfate Polyacrylamide Gels', *Methods Enzymol.*, **200**, 417–423 (1991).

79. S. Clarke, 'The Size and Detergent Binding of Membrane Proteins', *J. Biol. Chem.*, **250**, 5459–5469 (1973).
80. S. Hjerten, M. Sparrman, J. Liao, 'Purification of Membrane Proteins in SDS and Subsequent Renaturation', *Biochim. Biophys. Acta*, **939**, 476–484 (1988).
81. R.E. Manrow, R.P. Dottin, 'Renaturation and Localization of Enzymes in Polyacrylamide Gels: Studies with UDP Glucose Pyrophosphorylase of Dictyostelium', *Proc. Natl. Acad. Sci. USA*, **77**, 730–734 (1980).
82. G.P. Tuszynski, C.A. Buck, L. Warren, 'A Two-dimensional Polyacrylamide Gel Electrophoresis (PAGE) System Using Sodium Dodecyl Sulfate/PAGE in the First Dimension', *Anal. Biochem.*, **93**, 329–338 (1979).
83. J.L. Celenza, M. Carlson, 'Renaturation of Protein Kinase Activity in Protein Blots', *Methods Enzymol.*, **200**, 423–430 (1991).
84. J.E. Ferrel, Jr, G.S. Martin, 'Assessing Activities of Blotted Protein Kinases', *Methods Enzymol.*, **200**, 430–435 (1991).
85. D. Michaud, 'Gel Electrophoresis of Proteolytic Enzymes', *Anal. Chim. Acta*, **372**, 173–185 (1998).
86. T.E. Creighton, *Proteins*, 2nd edition, W.E. Freeman and Company, New York, 1–507, 1996.
87. E. Gianazza, I. Eberini, O. Santi, M. Vignati, 'Denaturant-gradient Gel Electrophoresis: Technical Aspects and Practical Applications', *Anal. Chim. Acta*, **372**, 99–120 (1998).
88. T.E. Creighton, 'Electrophoretic Analysis of the Unfolding of Proteins by Urea', *J. Mol. Biol.*, **129**, 235–264 (1979).
89. T.E. Creighton, 'Kinetic Study of Protein Unfolding and Refolding Using Urea Gradient Electrophoresis', *J. Mol. Biol.*, **137**, 61–80 (1980).
90. D.P. Goldenberg, T.E. Creighton, 'Gel Eelectrophoresis in Studies of Protein Conformation and Folding', *Anal. Biochem.*, **138**, 1–118 (1984).
91. D.R. Thatcher, R. Sheikh, 'The Relative Conformational Stability of the Alcohol Dehydrogenase Alleloenzymes of the Fruitfly Drosophila Melanogaster', *Biochem. J.*, **197**, 111–1117 (1981).
92. P. Carter, H. Bedouelle, G. Winter, 'Construction of Heterodimer Tyrosyl-tRNA Synthetase Shows $tRNA^{TYR}$ Interacts with both Subunits', *Proc. Natl. Acad. Sci. USA*, **83**, 1189–1992 (1986).
93. F. Gentile, B.M. Veneziani, C. Sellitto, 'Polyacrylamide Gel Electrophoresis in Discontinuous Transverse Urea Gradient Gels', *Anal. Biochem.*, **244**, 228–232 (1997).
94. P.B. Fernandes, R.V. Nardi, S.G. Franklin, 'The Resolution of Membrane Proteins Based upon Size, Charge, and Hydrophobicity', *Anal. Biochem.*, **91**, 101–114 (1978).
95. D.L. Feinstein, E.N. Moudrianakis, 'Hydrophobic and Ionic Effects upon Electrophoretic Mobilities of the Subunits of Coupling Factor 1 from Mitochondria', *Anal. Biochem.*, **136**, 362–371 (1984).
96. D.P. Goldenberg, 'Transverse Urea Gradient Gel Electrophoresis', in *Current Protocols in Protein Science*, eds. J.E. Coligan, B. Dunn, H.L. Ploegh, D.W. Speicher, P.T. Wingfield, John Wiley and Sons, New York, 7.4.1.–7.4.13, 1996.
97. D.P. Goldenberg, 'Analysis of Protein Conformation by Gel Electrophoresis', in *Protein Structure: A Practical Approach*, 2nd edition, ed. T.E. Creighton, IRL Press, Oxford, 187–218, 1997.
98. D.R. Thatcher, B. Hodson, 'Denaturation of Proteins and Nucleic Acids by Thermal Gradient Electrophoresis', *Biochem. J.*, **197**, 105–109 (1981).
99. J.R. Cann, 'Theory and Practice of Gel Electrophoresis of Interacting Molecules', *Anal. Biochem.*, **237**, 1–16 (1996).
100. E. Gianazza, P. Arnaud, 'Application of Gel Electrophoresis in the Determination of Protein – Low M_r Substances and Protein – Protein Interactions', *Anal. Chim. Acta*, **372**, 67–69 (1998).
101. T.H. Ueng, F. Bronner, 'Cellular and Luminal Forms of Rat Intestinal Calcium Binding as Studied by Counter Ion Electrophoresis', *Arch. Biochem. Biophys.*, **197**, 205–217 (1979).
102. R.W. Evans, J. Williams, 'The Electrophoresis of Transferrins in Urea/Polyacrylamide Gels', *Biochem. J.*, **189**, 541–546 (1980).
103. G. Saglio, G. Ricco, U. Mazza, et al., 'Human $^{T}\gamma$ Globin Chain is a Variant of $^{A}\gamma$ Chain ($^{A}\gamma^{Sardinia}$)', *Proc. Natl. Acad. Sci. USA*, **76**, 3420–3424 (1979).
104. A. Helenius, K. Simons, 'Charge Shift Electrophoresis: Simple Method for Distinguishing Between Amphiphilic and Hydrophylic Proteins in Detergent', *Proc. Natl. Acad. Sci. USA*, **74**, 529–532 (1977).
105. D.A. Hager, R.R. Burges, 'Elution of Protein from Sodium Dodecyl Sulfate/Polyacrylamide Gels, Removal of Sodium Dodecyl Sulfate, and Renaturation of Enzymatic Activity: Results with Sigma Subunit of *Escherichia coli* RNA Polymerase, Wheat Germ DNA Topoisomerase, and Other Enzymes', *Anal. Biochem.*, **109**, 76–86 (1980).
106. B. Kaplan, M. Pras, 'Separation of Proteins by Consecutive Sodium Dodecyl Sulfate (SDS)/Polyacrylamide Gel Electrophoresis and Reversed Phase High Performance Liquid Chromatography', *Biomed. Chromatogr.*, **5**, 86–89 (1992).
107. B. Kaplan, M. Pras, M. Ravid, 'Isolation and Purification of Amyloid Protein A by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis and Reversed Phase High Performance Chromatography', *J. Chromatogr.*, **573**, 17–22 (1992).
108. B. Kaplan, S. Yakar, Y. Balta, M. Pras, B. Martin, 'Isolation and Purification of Two Major Amyloid A Isotopes SAA1 and SAA2 from Acute Phase Plasma of Mice', *J. Chromatogr.*, **704**, 69–76 (1997).
109. B.S. Parekh, H.B. Mehta, M.D. West, R.C. Motelaro, 'Preparative Elution of Proteins from Nitrocellulose Membranes After Separation by Sodium Dodecyl

Sulfate/Polyacrylamide Gel Electrophoresis', *Anal. Biochem.*, **148**, 87–92 (1985).
110. B. Szewszyk, D.F. Summers, 'Preparative Elution of Proteins Blotted in Immobilon Membranes', *Anal. Biochem.*, **168**, 48–53 (1988).
111. K. Weber, J. Kutter, 'Reversible Denaturation of Enzymes by Sodium Dodecyl Sulfate', *J. Biol. Chem.*, **246**, 4504–4509 (1971).
112. O.M. Kapp, S.N. Vinogradov, 'Removal of Sodium Dodecyl Sulfate from Proteins', *Anal. Biochem.*, **91**, 230–235 (1978).
113. L.E. Henderson, S. Oroszlan, W. Konigsberg, 'A Micromethod for Complete Removal of Dodecyl Sulfate from Proteins by Ion-pair Extraction', *Anal. Biochem.*, **93**, 153–157 (1979).
114. B. Kaplan, M. Pras, 'Removal of Sodium Dodecyl Sulfate from Proteins by Gel Permeation Chromatography', *J. Chromatogr.*, **423**, 376–379 (1987).
115. F.T. Liang, D.E. Granstrom, J.F. Timoney, Y.F. Shi, 'Micropreparative High Resolution Purification of Proteins by a Combination of Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis, Isoelectric Focusing, and Membrane Blotting', *Anal. Biochem.*, **250**, 61–65 (1997).

High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of

E.C. Nice
Ludwig Institute for Cancer Research, Parkville, Victoria, Australia, and The Walter and Eliza Hall Institute, Victoria, Australia

B. Catimel
Ludwig Institute for Cancer Research, Parkville, Victoria, Australia

1 Introduction 1

2 Principles of Micropreparative High-performance Liquid Chromatography 2
- 2.1 Chromatographic Principles 2
- 2.2 Resolution 2
- 2.3 Large Sample Volume Trace Enrichment 3
- 2.4 Column Loading Capacity 3
- 2.5 Column Length 4
- 2.6 Micropreparative Size-exclusion Chromatography 5
- 2.7 Sample Recovery 5
- 2.8 Multidimensional Purification 6

3 Instrumentation 6
- 3.1 Narrow-bore and Microbore High-performance Liquid Chromatography 6
- 3.2 Capillary High-performance Liquid Chromatography 7
- 3.3 Columns 7

4 Micro Liquid Chromatography Versus Capillary Electrophoresis 8

5 Capillary Electrochromatography 8

6 Applications of Micropreparative High-performance Liquid Chromatography 8
- 6.1 Microsequence Analysis 9
- 6.2 The Use of Micropreparative Ion-exchange Columns to Generate Samples for Microsequence Analysis 9
- 6.3 Synergy of Micro Liquid Chromatography with Mass Spectrometry 10
- 6.4 Chromatographic Recovery of Proteins from Polyacrylamide Gels 11
- 6.5 Database Searching Strategies 11
- 6.6 Synergies Between Micropreparative High-performance Liquid Chromatography and Instrumental Biosensors 12
- 6.7 Liquid Chromatography/Nuclear Magnetic Resonance Spectroscopy 13

7 Future Directions 14
- 7.1 Microfabrication Technology for Miniaturized Separation Systems 14

Abbreviations and Acronyms 14

Related Articles 15

References 15

The use of short narrow-bore (2.1-mm i.d.), microbore (1-mm i.d.) or capillary (<1-mm i.d.) columns allows the recovery of purified proteins and peptides in reduced volume at concomitantly increased concentration compared with larger bore columns. Additionally, sample detection is correspondingly improved for concentration-dependent detectors. Sample volumes are not limiting: large sample volumes can be loaded on to interactive supports (e.g. reversed-phase (RP), ion-exchange (IEX), hydrophobic interaction (HI), affinity) under appropriate mobile phase conditions (trace enrichment). Suitable instrumentation requires careful consideration with regard to minimized pre- and postcolumn dead volumes, accurate gradient formation at low flow rates, detector flow cell geometry and sample recovery. These techniques are ideally suited for the purification of trace protein or peptide components in bulk biological samples prior to application to highly specific and sensitive downstream analytical techniques (e.g. microsequence analysis, mass spectrometry (MS), biosensor analysis, biological assay). They also facilitate the recovery of low levels of material following chemical or enzymatic modification (e.g. radioiodination, enzymatic semi-synthesis, chemical derivatization). The techniques can also be applied to the development of large-scale processes, whereby multidimensional purification protocols can be rapidly developed using minimum amounts of sample, solvent and column packing materials and then scaled up appropriately.

1 INTRODUCTION

High-performance liquid chromatography (HPLC) methodologies for the purification of proteins and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

peptides were developed largely during the 1970s, aided in particular by the availability of optimized column packing materials for the separation of higher molecular weight proteins based on the use of wide-pore silica supports.[1–5] These methods offered unrivaled advantages in terms of speed, resolution, sensitivity and recovery when used for the purification of small (low-micrograms level) quantities of biologically important compounds such as growth factors or their receptors from complex biological sources. The generation of column packing materials with alternative selectivities (e.g. RP, HI, anion-exchange, cation-exchange, size-exclusion, chromatofocusing, affinity) facilitated the development of high-resolution multidimensional methods, often resulting in high cumulative purification factors.[6] However, conventional HPLC columns (e.g. 4.6-mm i.d. operated at flow rates of ca. 1 mL min^{-1}) typically resulted in peak volumes of ca. 1 mL, or even larger. The resultant sample concentrations (μg mL^{-1}) were not ideally suited to subsequent manipulations, and at such low concentrations losses due to non-specific adsorption on either the chromatographic support or associated equipment (syringes, sample vials, concentration membranes, etc.) were commonplace. Even when using volatile chromatographic solvent systems (e.g. trifluoroacetic acid (TFA)–acetonitrile (CH_3CN), heptafluorobutyric anhydride (HFBA)–CH_3CN, ammonium hydrogencarbonate) to allow subsequent sample concentration, poor recoveries were frequently observed after lyophilization.[7–10] In addition, solvent evaporation was shown, in some cases, to result in chemical modification even when performed in an inert atmosphere.[11] Attempts to reduce peak volumes by operating at lower flow rates were shown to be associated with poor recoveries of, in particular, more hydrophobic proteins.[12,13]

Chromatographic theory suggested that the use of columns with smaller internal diameter (narrowbore, microbore or capillary columns) would enable samples to be recovered in reduced volumes at increased concentration. However, while microbore HPLC systems had been described previously for low-molecular-weight compounds, and had been shown to be associated with very high operating efficiencies,[14,15] they had not been generally accepted because of perceived sample volume limitations (injection volumes of <1 μL), limiting their application when working with large-volume samples at low concentration. Fortuitously, proteins and peptides exhibit very large capacity factors (k') on interactive supports below critical secondary solvent compositions,[16] allowing large-volume samples to be loaded, and hence overcoming the perceived limitations.

2 PRINCIPLES OF MICROPREPARATIVE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

2.1 Chromatographic Principles

It can be shown that eluent peak volumes from chromatographic columns are directly proportional to the column length and inversely proportional to the square of the column diameter.[17] Thus, if we compare a 4.6-mm i.d. column operated at a flow rate of 1 mL min^{-1} (a practical optimum for many chromatographic columns of these dimensions, as evidenced by a minimum in plots of linear flow velocity (u) against height equivalent to a theoretical plate (HETP) in the Van Deemter equation)[18] with a 1-mm i.d. microbore column, the flow rate for the 1-mm i.d. column would need to be reduced by $(4.6/1)^2$, i.e. 21-fold, to approximately 50 μL min^{-1} to operate at the same linear flow velocity (Figure 1). Under such operating conditions, if the same mass of protein is loaded on to each column, and the chromatographic efficiency of each column is identical (high packing efficiencies can be achieved with microbore and capillary columns), then samples will be recovered from the microbore column in concomitantly reduced peak volumes (50 μL compared with 1 mL for a peak 1 min wide). For concentration-dependent detectors, this increase in sample concentration will give a directly proportional increase in the detection signal if the flow cell geometry is conserved. It has been shown that this predicted performance can be achieved in practice.[16,19] Nanogram quantities of peptides and proteins can be readily detected using small-bore columns (Figure 2), with peak elution volumes of 100 μL or less.

2.2 Resolution

Resolution is a function of both the packing selectivity and the efficiency with which the column has been packed. Therefore, columns of the same efficiency but of different dimensions operated under equivalent elution conditions (solvent composition, gradient profile, temperature and linear flow velocity) should give identical separations of complex mixtures. This has been shown in practice,[16,19,20] and facilitates direct scale-up or scale-down of separation protocols using matched columns. There will be small differences in resolution/retention time not related to the columns themselves, but to instrumental variations mainly due to extra-column dead volumes (e.g. pumping configurations, plumbing design, mixer volumes, loading loop volume, tubing size).

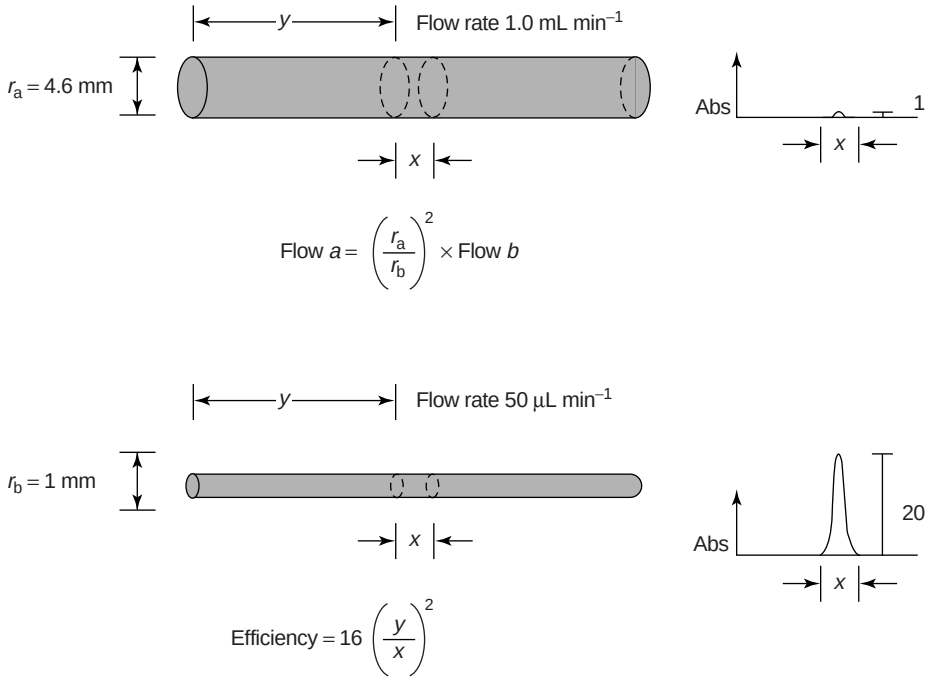

Figure 1 Theory of micropreparative HPLC. Abs = absorbance.

2.3 Large Sample Volume Trace Enrichment

When purifying trace components from complex mixtures, it is imperative that the total sample be loaded if the optimum benefits of microcolumns are to be achieved. A typical application would be the repurification of a sample recovered from a conventional HPLC column, typically in volumes of several milliliters, on to a narrow-bore or microbore column at a suitable stage of a multidimensional purification protocol. Proteins and peptides bind very strongly (i.e. exhibit large capacity factors) to interactive supports (e.g. RP, IEX, HI or affinity matrices) below critical secondary solvent compositions (Figure 3). Under these conditions, trace enrichment of large sample volumes directly on to micropreparative columns is possible before recovering retained proteins or peptides in small volumes at high concentration by gradient or step elution.[21] An important practical point is to ensure that the sample is in a buffer that will allow trace enrichment. Samples may require dilution to reduce the salt or organic solvent concentration used for elution at the previous chromatographic step. Samples are typically loaded at high flow rate (see the discussion on column back-pressures below) by means of a large-volume injection loop (up to 2 mL), often using multiple injections, a Superloop (up to 10 mL) (Amersham Pharmacia Biotech) or by direct loading via the primary or auxiliary pump. Large-volume injection loops should be switched out of the main flow path following sample loading to minimize the system volume between the pumps and the column, which would otherwise result in unacceptably long delays in gradient delivery. We have frequently loaded samples in volumes of up to 100 mL on to 1-mm i.d. RP columns and recovered the retained proteins in peak volumes of <100 µL (>1000-fold concentration).

2.4 Column Loading Capacity

Short narrow-bore, microbore or capillary columns must have adequate sample capacity to be a viable proposition for the micropreparative HPLC of proteins and peptides. We have shown that even short 30 × 2.1-mm i.d. RP columns [7-µm particle size, 300-Å pore size ($1\text{Å} = 10^{-10}$ m)] have a capacity of over 4 mg of protein.[21] However, at such loadings the column was obviously grossly overloaded, as evidenced by the recovery volume after stepwise elution (1.4 mL). An investigation of protein load versus peak volume showed that optimum efficiency could be maintained for individual protein loads up to 10 µg when using a 100 × 1-mm i.d. column. Comparison of varying column lengths (30 and 100 mm) showed that the optimum load was independent of length until the critical loading was exceeded, when the shorter column overloaded more rapidly with increasing sample (Figure 4).[22] However, even when operating under overload conditions there are still significant advantages to be gained in terms of the sample recovery volume when using columns of reduced inner diameter. With complex mixtures, we

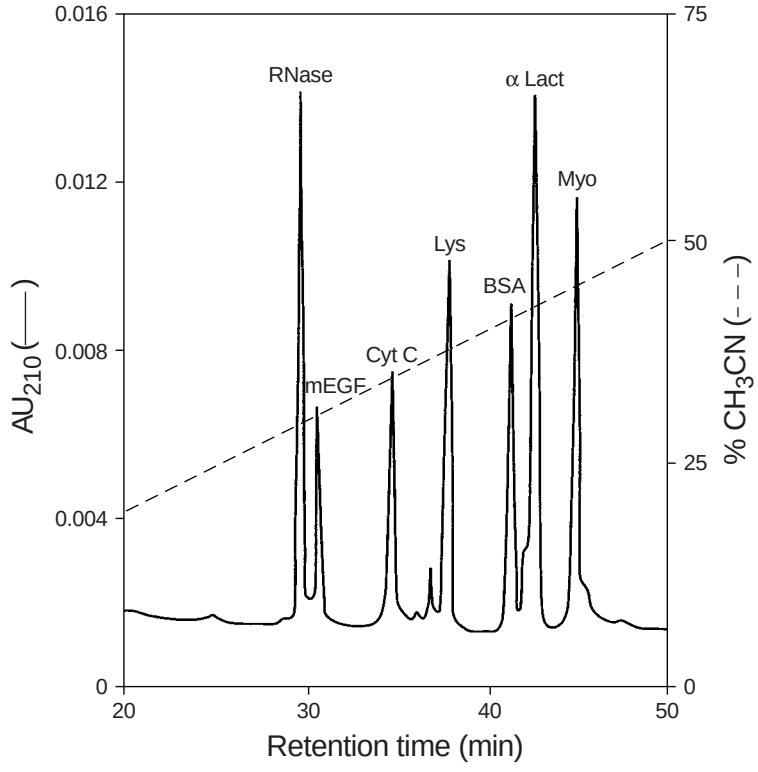

Figure 2 Separation of low-level protein standards by micropreparative reversed-phase high-performance liquid chromatography (RPHPLC). Protein standards were separated on a Brownlee RP 300 narrow-bore ''guard column'' (30 × 2.1-mm i.d.) using a linear 1% min^{-1} gradient between 0.9% (w/v) NaCl (pH 2.1) and CH_3CN as indicated (dashed line). The flow rate was 100 µL min^{-1} and the column temperature was 45 °C. Detection was by absorbance measurement at 210 nm. The standards were ribonuclease A (RNase, 92 ng), murine epidermal growth factor (mEGF, 25 ng), cytochrome *c* (cyt *c*, 134 ng), lysozyme (lys, 80 ng), bovine serum albumin (BSA, 140 ng), bovine α-lactalbumin (α lact, 94 ng) and myoglobin (myo, 79 ng). (Reprinted from *Biochem. Int.*, **11**, E.C. Nice, B. Grego, R.J. Simpson, 'Application of Short Microbore HPLC Columns for the Preparation of Samples for Protein Microsequencing', 187–195, Copyright (1985), with permission from Harcourt Brace and Company.)

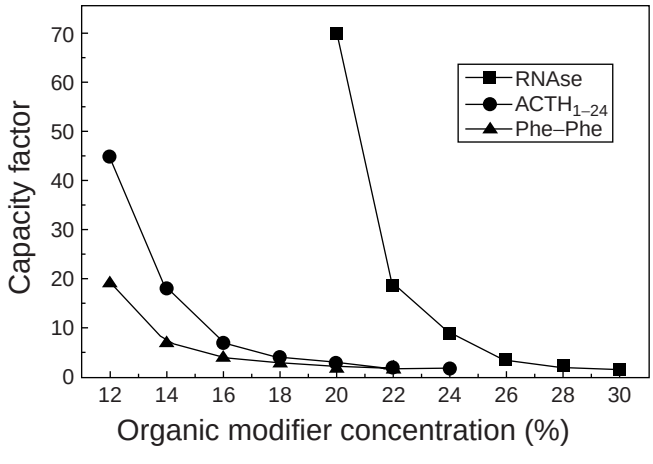

Figure 3 Relationship between capacity factor (k') and organic solvent concentration for the separation of proteins and peptides by RPHPLC. Proteins and peptides were chromatographed on a Brownlee RP 300 column (30 × 2.1 mm i.d.) at 25 °C and a flow rate of 100 µL min^{-1}. The mobile phase was 0.15% (v/v) TFA mixed with CH_3CN at the concentrations shown.

have frequently applied total protein loads of >1 mg on to 30 × 2.1-mm i.d. columns and recovered purified minor components of interest in the small peak volumes associated with such columns.

For capillary columns, it has been shown that similar criteria apply.[23] Detection of proteins and peptides at the low-picogram level has been reported with resolution equivalent to that observed on larger bore columns. The optimum length appeared to be 20 cm.[23]

2.5 Column Length

An important practical consideration with the use of columns of reduced inner diameter is the increased operational back-pressures associated with these columns, particularly if large sample volumes are to be loaded at high flow rates to minimize the total run time, since most chromatographic systems have operating pressure limits of 6000–10 000 psi. Back-pressures are further exacerbated when using either small particle size or nonporous supports, which have been shown to

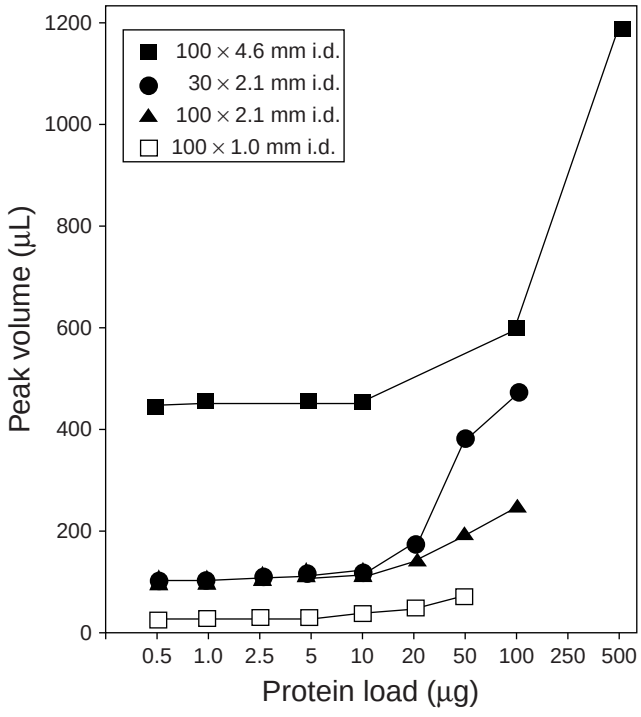

Figure 4 Relationship between protein load and recovery volume. α-Lactalbumin (0.5–500 µg) was chromatographed on Brownlee RP 300 columns of the dimensions indicated. Proteins were eluted using a linear 60 min gradient between 0.15% (v/v) TFA and 60% aqueous CH_3CN–0.125% (v/v) TFA. The flow rates were 1000, 200 and 50 µL min^{-1} for the 4.6, 2.1 and 1-mm i.d. columns, respectively. Column temperature, 45 °C. (Reproduced with permission from E.C. Nice, R.J. Simpson, 'Micropreparative HPLC of Proteins and Peptides', *J. Pharm. Biomed. Anal.*, **7**, 1039–1053 (1989), Copyright 1989, with permission from Elsevier Science.)

offer potential advantages in terms of both resolution and chromatographic efficiency for protein and peptide separations.[24–26]

Since the operating back-pressure is directly proportional to column length,[27] an obvious solution is to use short columns. By extrapolation of experimentally determined pressure–flow relationships for 4-mm i.d. columns packed with small (1.5-µm diameter) nonporous silica particles, we have calculated that for short narrow-bore columns (30 × 2-mm i.d.) packed with the same nonporous support, trace enrichment could be performed at flow rates of up to 0.5 mL min^{-1}, which would allow large sample volumes to be loaded in a reasonable time frame.[22]

It has been shown that column length plays a negligible role in the resolution of proteins or peptides on strongly interactive supports when using gradient elution conditions,[28,29] since the steep dependence of secondary solvent composition on k' under isocratic conditions (Figure 3) results in the peak volumes being largely a function of the gradient profile, with only a narrow time window during which the true chromatographic process is operative. Thus, effective micropreparative protein and peptide separations can be achieved on columns (30 × 2.1-mm i.d.) that were originally designed as guard columns for use before 4.6-mm i.d. analytical columns and which only have a relatively low plate count.[16,19,30] Such guard columns offer significant advantages in terms of cost, and can be virtually regarded as disposable when working with potent biologically active molecules where cross-contamination between samples can be disastrous.

2.6 Micropreparative Size-exclusion Chromatography

Different criteria apply when considering the dimensions of micropreparative columns for use with noninteractive size-exclusion supports. Efficiency (plate count) is of the utmost importance with such columns, and column length must therefore be maintained. In addition, samples cannot be trace enriched on to these columns and injection volumes become limiting [ideally the sample volume should be less than 2% of the column bed volume[31]]. Because of these sample volume limitations, micropreparative size-exclusion columns are best used in the middle of multidimensional protocols following the use of interactive supports from which recovery volumes of 50–100 µL can be routinely achieved, allowing direct injection of concentrated, small-volume samples on to the size-exclusion support with the potential for further trace enrichment of recovered fractions in the subsequent chromatographic stages. We have also found micropreparative size-exclusion columns ideal for analytical screening of biologically active compounds in bulk biological samples when separation can be readily achieved in buffers compatible with direct bioassay (e.g. 0.15 M sodium chloride), or for buffer exchange of samples prior to bioassay or chemical or enzymatic manipulation.

2.7 Sample Recovery

Recovery is of primary importance in any purification technique. The use of short microcolumns has been shown to result in increased sample recovery compared with wider-bore columns.[32,33] This is presumably due to the greatly reduced surface area of the packing exposed to the sample, with a corresponding reduction in the number of available sites for irreversible nonspecific adsorption.

Losses during chromatographic procedures can be due to both on-column losses and losses occurring during associated nonchromatographic sample manipulations (e.g. sample dilution, pH adjustment, storage, chemical manipulation). We have found that the use of an injection

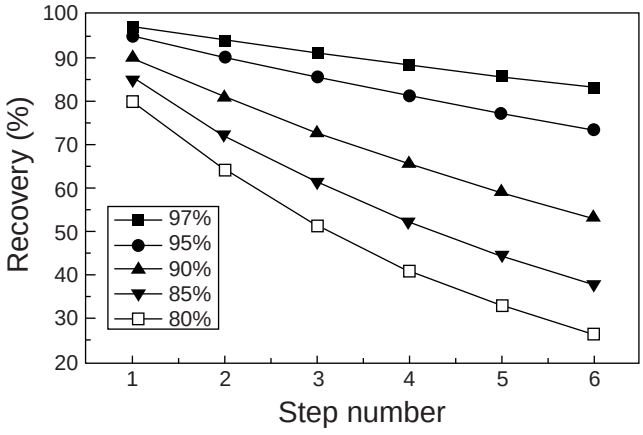

Figure 5 Relationship between overall recovery and repetitive yield in multidimensional chromatography.

syringe as a vehicle for performing sample dilution prior to reinjection and the use of low levels of nonionic detergent in the diluent (e.g. 0.02% Tween 20) help to reduce losses during sample manipulation.[6,16,19] If the overall total system recovery can be kept high throughout the procedure, it is possible to take a sample through a number of successive chromatographic steps (multidimensional purification) and still have sufficient purified material available for structural and/or biological analysis. If the total recovery for each stage is >90%, a purification strategy of more than six sequential steps is possible with an overall recovery of >50% (Figure 5). Moritz et al. were able to demonstrate a repetitive total system recovery of 97% for five sequential micromanipulations of lysozyme (50 ng) on a 0.32-mm i.d. capillary column.[34]

2.8 Multidimensional Purification

The sequential use of microcolumns of varying selectivity, assuming they have good recovery characteristics, allows very high purification factors to be achieved. Indeed, it can be calculated (Table 1) that purification factors far in excess of those achieved by two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE), which is generally accepted as being a high-resolution technique for the separation of complex mixtures of proteins and peptides, can be achieved using multidimensional purification protocols, with the added advantage that the purified sample is in a form directly compatible with downstream analysis. By careful consideration of the order in which columns are used in a multidimensional purification scheme, the need to manipulate the sample between stages can be minimized, thereby further improving overall recoveries. For example, samples recovered in small volumes from micropreparative IEX or HI columns can be taken directly on to size exclusion and then, being in aqueous solvents, can be loaded directly on to RP columns. We have used such an approach for the purification to homogeneity of murine EGF from salivary glands,[35] a GTPase-activating protein from bovine brain[36] and the membrane-associated A33 antigen from a colonic carcinoma cell line.[37] Opiteck et al.[38,39] have described automated 2D narrow-bore and capillary HPLC systems, using the orthogonal chromatographic modes of size exclusion and RP or cation exchange and RP, to separate complex protein mixtures.

Table 1 Potential peak resolution in multidimensional HPLC

Chromatographic mode	Peak capacity[a]	Cumulative resolution
Hydrophobic interaction	50	50
Size-exclusion	10	500
Ion-exchange	50	25 000
Reversed-phase	100	2 500 000

[a] Peak capacity is defined as the number of peaks that could theoretically be resolved in a typical chromatographic run.

3 INSTRUMENTATION

3.1 Narrow-bore and Microbore High-performance Liquid Chromatography

Suitable instrumentation for micropreparative HPLC requires (a) precise, pulseless pumping at low flow rates, (b) the ability to generate accurately and reproducibly gradients at low flow rates, (c) minimum dead volumes throughout the instrument and, for true micropreparative applications, (d) reliable and accurate recovery of small-volume fractions. These key requirements have been reviewed.[40–43] A prototype liquid-delivery system in which thermal expansion is used to reproducibly deliver pulse-free flow at flow rates down to 10 nL min^{-1} has also been described.[44]

Most conventional HPLC systems using high-pressure mixing can be readily modified for use with narrow-bore or microbore columns.[21,43] In such cases, recovery will usually be effected by manual trapping, allowance being made for the dead volume between the detector flow cell and the collection port, which is significant when operating at low flow rates (<200 μL min^{-1}). Manual trapping usually involves holding the tip of the collection device either against the side of the collection tube (we usually use 0.5- or 1.5-mL Eppendorf tubes which have particularly good recovery and low-level background contamination properties) or under the liquid meniscus to prevent droplet formation. The drop size (typically around 10 μL) becomes extremely significant in terms of both resolution and overall recovery when dealing

with peak volumes of 20–100 μL. Systems using prepump low-pressure mixing devices are generally not suitable for gradient elution at very low flow rates because of excessive instrumental volumes between the point of generation of the gradient and the head of the column, often of the order of several milliliters.

Relatively few HPLC systems have been designed specifically for micropreparative operation in terms of the requirements for both the chromatographic separation and microfraction collection. As an alternative to direct automatic collection of fractions into tubes or microtiter plates, the use of moving blotting membranes has also been described.[45,46] Commercially available micropreparative instruments include the Amersham Pharmacia SMART system (http://www.apbiotech.com), which has a built in microfraction collector,[35,47] and the LC Packings instrument with the PROBOT fraction collector (http://www.lcpackings.nl),[48] which can collect fractions on to a number of different substrates [matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) targets, poly(vinylidene fluoride) (PVDF) membranes, 96-well microtiter plates].

One instrumental component which deserves further comment is the detector flow cell. The requirement for specially designed low-volume flow cells for narrow-bore and microbore chromatography is often overstated. Kirkland et al. have shown that combined postcolumn and detector flow-cell volumes up to 10% of the peak volume are allowable.[49] In practice, we have found that detector flow-cell volumes of 4.6 μL are compatible with peak volumes of >50 μL. Flow-cell design, per se, is important, however, since poorly designed flow cells can cause turbulence or peak hold-up, and hence poor sample clearance and peak mixing. In addition, small-volume flow cells with reduced optical pathlengths have proportionally reduced sensitivity.

3.2 Capillary High-performance Liquid Chromatography

The recovery of protein and peptide samples in small (microliter) volumes that can be readily manipulated off-line for use in ''downstream'' high-sensitivity analysis is readily achieved with the use of 1-mm i.d. columns, where peak volumes of 20–50 μL are routine (volumes much less than 10 μL are not really desirable for manual off-line manipulations since post-column sample handling becomes difficult and even small volume losses are significant). However, the use of on-line MS for the high-sensitivity analysis of low (femtomole) levels of high-molecular-weight proteins,[50] in which chromatographic techniques are the optimum method of introducing samples into the instrument,[50–52] has stimulated the development of capillary HPLC techniques.[23,34,53] As indicated above, the increases in performance observed with microbore columns are equally applicable to capillary systems, with concomitant reductions in peak volume and increases in sample concentration and hence sensitivity of detection.[23,34,53] Gradients need to be developed and supplied at flow rates of 10 μL min^{-1} or less with capillary columns (Figure 6). This is usually achieved by the use of stream-splitting techniques, in which gradients are generated at higher flow rates (e.g. 50 μL min^{-1}) but only 1–5% of the resulting gradient is passed through the column, the remainder of the solvent being diverted to waste.

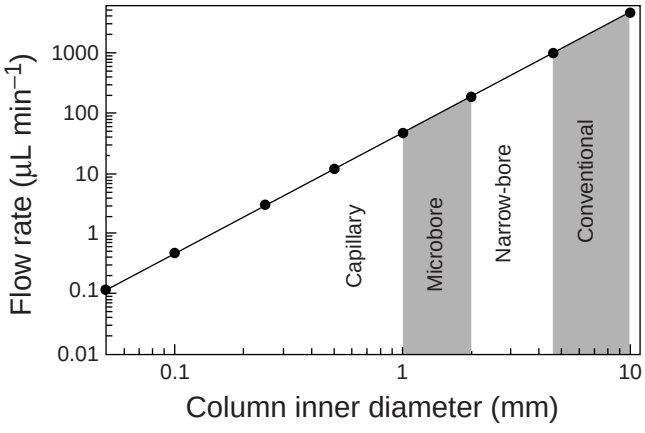

Figure 6 Relationship between operating flow rate and column diameter. The linear flow velocities for the columns of varying i.d. were calculated using the equation shown in Figure 1.

The small peak volumes (low-microliter volumes or less) achievable with capillary HPLC of proteins and peptides means that on-column detection[54] or specially designed low-volume flow cells [e.g. axial-beam longitudinal alignment (Z-cell), which yields longer pathlengths[23,34,53,55]] are required. Commercially available instruments include the Ultimate Integrated Capillary HPLC system from LC Packings (http://www.lcpackings.nl), the Micro Pro modular system from Eldex (http://www.eldex.com), the Micro-Tech MicroLC (http://www.mtscientific.com) and the PE-ABI 173A Microblotter cLC system (http://www.perkin-elmer.com).[56] The Ultimate system is capable of running at flow rates from 50 nL min^{-1} to 200 μL min^{-1}.

3.3 Columns

Short (<10-cm) narrow-bore and microbore columns are now commercially available from a wide range of manufacturers, although they may be readily packed ''in-house'' using conventional slurry packing techniques.[21] A range of packed capillary columns are currently available from LC Packings. However, the packing of

capillary columns in-house is also not difficult. Detailed instructions are given in the publications by Tong et al.,[57] Huber et al.[58] and Kennedy and Jorgenson.[59] It has been noted[60] that capillary columns can be used for extended periods (up to 18 months) since they can be readily regenerated by trimming the top of the column to remove debris.

4 MICRO LIQUID CHROMATOGRAPHY VERSUS CAPILLARY ELECTROPHORESIS

The relative merits of micro liquid chromatography (LC) and capillary electrophoresis (CE) have been the topic of much discussion. CE is undoubtedly a rapid ultrasensitive analytical technique, exhibiting high separation efficiencies under relatively mild operating conditions (in terms of biological stability), with detection levels at the single molecule level for suitably labeled proteins using laser-induced fluorescence detection.[61–63] However, with on-line ultraviolet (UV) absorbance because of the small detection cell volume and short optical pathlength, detection limits are in the micromolar range.[64] A further significant limitation is the very small sample volumes (low-nanoliter level) that can be applied directly to CE. Whilst the development of on-line concentration devices has helped alleviate this problem,[65–68] large sample loadings will probably not be possible owing to significant heat generation under the operating conditions used, which leads to a deterioration in performance. However, Figeys et al.[68] have successfully used a 5-μm, 300-Å C_{18} packing as a solid-phase extraction device coupled to CE and electrospray ionization tandem mass spectrometry (ESIMS/MS) to identify low-nanogram quantities of yeast proteins previously separated by 2D electrophoresis. The volume loaded on to the extraction device was 15 μL.

By contrast, as we have described above, micro LC is compatible with the loading of large-volume dilute samples by trace enrichment, a necessity when purifying trace components. Thus, for many micropreparative applications the use of micro LC would appear to be the method of choice.

Jorgenson et al.[69–71] have coupled the concentration potential of micro LC with the resolving power of CE in a multidimensional approach. Micropreparative size-exclusion chromatography (SEC)/CE and RPHPLC/CE were presented. Such systems may be even more applicable if capillary chromatography is used in the first dimension, where it would be possible to recover proteins and peptides in nanoliter volumes if using, for example, 50-μm columns.

5 CAPILLARY ELECTROCHROMATOGRAPHY

Capillary electrochromatography (CEC) is a hybrid between CE and HPLC, coupling the very high efficiency of CE with the high selectivity of HPLC. In the simplest (isocratic) form an apparatus identical with that in CE is used, solvent flow being generated by electroendosmotic flow at high voltage (typically 10–30 kV). Under these conditions, and using low buffer concentrations in the mobile phase (2–5 mM) to avoid Joule heating effects, a virtually linear relationship exists between solvent velocity and applied voltage. CEC can be performed in both open-tubular columns and capillaries packed with conventional supports.[72,73] Neutral species are separated by a partition-based mechanism whereas the separation of charged species is based on both electrophoresis and partition. The velocity profile in electrochromatography approaches perfect plug flow (contrasting with parabolic flow observed in pressure-driven systems), resulting in very high chromatographic efficiencies limited only by axial diffusion.[74] Because flow is generated electrically rather than by pressure, small particle diameter supports can be used, further enhancing the separation efficiency.[75]

Initially, CEC tended to be performed under isocratic conditions. However, modifications of the instrumentation to allow gradient elution have recently been described,[58,76–78] enabling large samples to be trace enriched on to interactive supports and the full potential of such systems for the separation of proteins and peptides to be achieved. CEC has recently been coupled to ion trap/reflectron time-of-flight MS for the analysis of protein digests.[79,80]

6 APPLICATIONS OF MICROPREPARATIVE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

Micropreparative HPLC is suited to a wide range of scientific disciplines where the recovery of trace components present in bulk biological samples is required prior to sensitive and specific downstream analysis (Figure 7). Although this article relates specifically to applications related to protein and peptide purification, the use of micropreparative HPLC is equally applicable to other classes of chemical compounds (e.g. steroids, pesticides, oligonucleotides), provided that sensitive and specific analytical methods are available for subsequent analysis.

For protein and peptide applications, the concentration of samples recovered from micropreparative HPLC

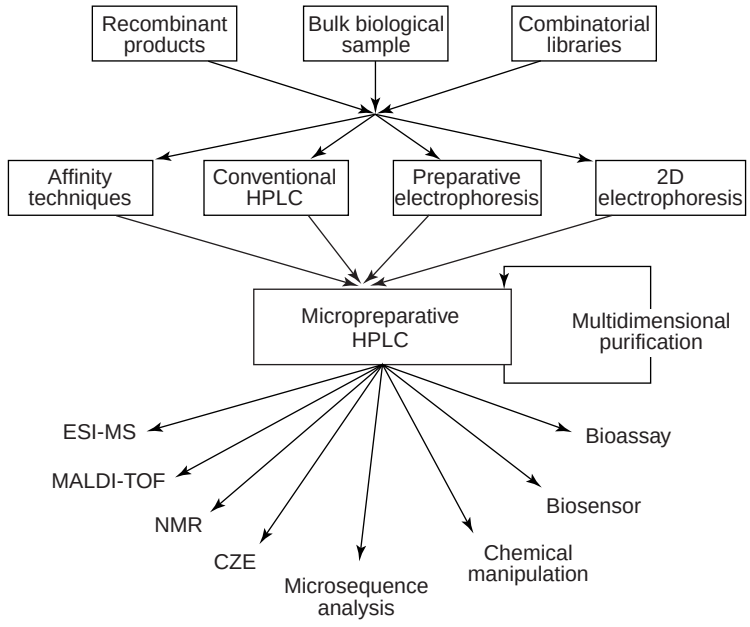

Figure 7 Applications of micropreparative HPLC in protein and peptide research.

columns is such that a number of complementary downstream analytical techniques can be used (Figure 7). Indeed, even if only nanogram quantities of purified material are available, the recovered sample can often still be subjected to a number of different assays (e.g. bioassay, biosensor analysis, MS, immunochemical assays, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE), and N-terminal sequence analysis).

6.1 Microsequence Analysis

Historically, the most significant use of micropreparative HPLC has been in the preparation of low-level samples for amino acid sequence analysis. Indeed, it was this application that initially led us[16,19,21,22] and others[20] to develop the use of microbore HPLC methodologies for the purification of proteins and peptides. These techniques have been used extensively in the isolation and characterization of a number of important biological molecules, including the hematopoietic growth factors and other members of the cytokine family.[81–84] Typically, such purifications have involved multiple chromatographic stages utilizing several column packing materials with alternative different selectivity to achieve high purification factors as described above.

6.2 The Use of Micropreparative Ion-exchange Columns to Generate Samples for Microsequence Analysis

The use of narrow-bore and microbore RPHPLC columns to purify proteins and peptides has been generally accepted by most laboratories involved in microsequence analysis, and has resulted in numerous scientific publications which make use of this technique.[18,19,30,85–88] However, RPHPLC is not a panacea, poor recoveries in particular being frequently observed for high-molecular-weight hydrophobic proteins. The use of micropreparative columns with different selectivities immediately prior to microsequence analysis would therefore greatly extend the general applicability of these microtechniques, provided that such columns could (a) resolve complex mixtures of peptides and proteins, (b) recover purified proteins in small peak volumes with high yield and (c) use mobile phases compatible with subsequent Edman chemistry, enabling samples to be loaded directly on to the sequencer following the final chromatographic purification.

We have shown that micropreparative IEX columns can be used in multidimensional chromatographic strategies for the purification of trace components in bulk biological samples.[35,36] By careful selection of appropriate solvent systems the small eluent volumes obtained can be loaded directly on to the gas-phase sequencer and the N-terminal sequence obtained. In particular, the use of an aqueous sodium chloride gradient (20 mM–1 M, pH 6.5, i.e. a buffer-free system) was found to be compatible with subsequent microsequence analysis[35,36] since sodium chloride does not interfere with the Edman chemistry. Indeed, sodium chloride appears to change the partitioning of the anilinothiazolinone (ATZ) amino acids between the polybrene film used for immobilization on to the glass-fiber sample

disk of the sequencer and the extraction solvent (1-chlorobutane), leading to increased recovery of charged amino acids.[89]

Recent developments in sequencer technology have also facilitated the analysis of buffer-containing samples. For example, the Hewlett-Packard G1005A protein sequencer[90] uses a novel biphasic two-stage hydrophobic and hydrophilic reaction cartridge which permits processing, concentration and cleanup of samples directly on the sequencer cartridge. Large volumes (up to 5 mL) can be processed in a single application. Using the hydrophobic support, inorganic salts and buffers present in the purified fractions can be washed out using aqueous washes that can be either mildly basic or acidic. PE-ABI also offers a sample preparation cartridge (ProSorb) which will accommodate up to 400-μL samples and which is compatible with a wide variety of buffers and reagents (http://perkin-elmer.com).

Although microsequence analysis continues to be an important tool in the modern protein structure–function laboratory, recent advances in MS (see below) have demonstrated the use of gel electrophoresis, narrow-bore or capillary RPHPLC and MS to be a very powerful combination for the rapid characterization of proteins and peptides.[91–94] Indeed, these technologies, coupled with sophisticated database analysis,[95] form the basis of modern proteomics,[96] the determination of the total protein complement of an individual genome.

6.3 Synergy of Micro Liquid Chromatography with Mass Spectrometry

MS has become an important technique for the analysis of biomolecules. In particular, MALDI/TOF[97,98] and electrospray ionization mass spectrometry (ESIMS),[99] which are both capable of accurately determining the molecular weight of proteins of over 100 000 Da with low-picomole to femtomole sensitivity,[50] have had a major impact on protein and peptide analysis. In practice, these two techniques are very different. MALDI/TOF generates predominately singly protonated ions that can be analyzed using a time-of-flight mass detector. Electrospray ionization (ESI), by contrast, generates a series of multiply protonated protein ions, many of which fall in the mass/charge range 500–4000 Da, which is compatible with modern quadrupole instruments. Additionally, structural analysis by tandem mass spectrometry (MS/MS) of ESI-generated peptide ions can be achieved.[85] Fundamental to both of these techniques is the ability to generate the sample in a suitable form for direct analysis. Microchromatography systems appear ideal for this purpose.[23,34,50–52]

6.3.1 Matrix-assisted Laser Desorption/Ionization Time-of-Flight

MALDI/TOF is an ''off-line'' technique, in which the sample is purified and recovered prior to analysis. The sample is then mixed with a UV-absorbing matrix [e.g. sinapinic acid,[100] α-cyano-4-hydroxycinnamic acid[51]] and applied to the inert target probe of the mass spectrometer. The matrix absorbs energy from a pulsed UV laser, causing ionization of the sample, whose mass is then determined by time-of-flight MS. MALDI/TOF is suitable for the analysis of mixtures of proteins or peptides, and it is relatively insensitive to salts [a key feature of the technique first demonstrated by Beavis and Chait[100]], although improved sample application methods have been described[101] that enable salts and impurities to be washed off the prepared samples, improving both resolution and sensitivity. MALDI/TOF has small sample volume requirements (typically 0.5–1 μL of sample is mixed with an equal volume of matrix solution) and high (femtomole) sensitivity, making the technique ideally suited to the direct analysis of samples purified by micropreparative HPLC.[102] The proven compatibility of these two techniques has resulted in the development of low-volume fraction collectors designed specifically for direct application to MALDI/TOF targets (see section 3). A method has been described in which proteins and peptides are further trace enriched during vacuum concentration on to RP packing material if even higher sensitivity is required.[103]

MALDI/TOF has also been used for direct protein sequence analysis.[104] First, a concatenated series of peptide fragments, differing from each other by a single residue, is generated in a controlled fashion by rapid stepwise degradation in the presence of a chain-terminating agent (ladder-generating chemistry), or by using the enzyme carboxypeptidase Y (specifically for generating a C-terminal sequence). The complete ladder generated in this manner can be analyzed by MALDI/TOF. The mass differences correspond to the amino acid residue removed at each step, and hence define the sequence of amino acids in the original chain. Subpicomole amounts of protein can be rapidly analyzed with a mass accuracy of up to 1 part in 10 000. However, as with other methods of sequence analysis, a homogeneous sample facilitates analysis.

6.3.2 Electrospray Ionization Mass Spectrometry

ESIMS is frequently used as an ''on-line'' technique in which an aqueous solution is introduced into the instrument through a needle held at high voltage. Ionization is caused by desolvation of droplets created under atmospheric pressure in an electrospray process between the needle and a grounded nozzle. Separation techniques

such as microbore or capillary HPLC,[34] CE[65,105] or more recently CEC[79,80] are therefore the methods of choice for sample introduction. Unfortunately, CE is sample limited (only low-nanoliter volumes are compatible), although on-capillary sample concentration methods have partially alleviated this problem and both mass analysis and sequence determination at the low- or even sub-femtomole level have been reported.[105] Additionally, ESIMS compatibility with the electrolytes used for capillary zone electrophoresis (CZE) is essential. The use of a suitable sheath liquid can minimize buffer-related effects, but the dilution effect of the sheath liquid can reduce the sensitivity of the analysis.[105]

By comparison, as we have illustrated above, trace enrichment on to microscale RPHPLC columns permits the whole sample to be applied, irrespective of volume, provided that concentrations of organic solvent which would prevent interaction with the support are not exceeded (see Figure 3). Narrow-bore, microbore and capillary systems are now used widely for the introduction of proteins and peptides into the mass spectrometer. The column eluate can also be split prior to the mass spectrometer, allowing part of the separated sample to be recovered for additional analyses (e.g. microsequence analysis).[91]

The success of this technique has facilitated the low-level analysis of peptide maps from novel proteins[91,92] and is now the method of choice for the assignment of disulfide bonds in native and recombinant proteins,[106–108] identification of post-translational and other modifications of proteins and peptides,[109,110] identification of peptides bound to major histocompatibility complex (MHC) molecules[111–112] and the identification of naturally occurring protein variants.[113] Narrow-bore and capillary chromatography have also been used for the analysis of UV-cross-linked protein–nucleic acid complexes,[114] where the identity of a tryptic peptide was confirmed by on-line capillary RPHPLC/ESIMS/MS, using 25 pmol of the peptide digest.

6.4 Chromatographic Recovery of Proteins from Polyacrylamide Gels

PAGE is now a widely used technique for the separation of complex mixtures of proteins and peptides. Up to 10 000 individual protein species have been separated in a single run using high-resolution 2D techniques.[115] The general applicability of PAGE is further enhanced when coupled to ancillary protein recovery techniques (e.g. passive elution, electroelution, electroblotting). Because gel electrophoresis is compatible with the separation of low levels of protein, micropreparative HPLC techniques have often proved beneficial in subsequent sample manipulation. A number of alternative protocols have been proposed, including ''inverse gradients'' for the efficient recovery of pure proteins from gel electroeluates, free of high concentrations of sodium dodecyl sulfate- and polyacrylamide gel-related artefacts[116] and a micropreparative technique in which protein spots from 2D gels were excised, minced and placed directly in a precolumn connected in series with a short RPHPLC cartridge (20×2 mm i.d.).[117] In an extension of this method, Tetaz et al.[118] performed ''in-gel'' digestion in situ in the precolumn and then recovered the generated peptides.

Methods utilizing micropreparative HPLC for the purification of peptide fragments generated using a gel-based approach have also been described.[119–123] Aebersold et al. described a generally applicable method in which proteins were electroblotted from 2D gels on to nitrocellulose.[120] Spots were then excised and the samples eluted with 100% TFA. The volatile TFA was then evaporated and the eluted protein digested either chemically or enzymatically followed by separation of the resulting peptides by narrow-bore RPHPLC. However, poor recoveries (<10%) have been observed following electroblotting and electroelution,[119] and therefore methods have been developed for in-gel digestion followed by passive elution and chromatographic separation of the generated peptides.[93,119,121–124] The method of Hellman and Renlund[124] has been shown to be compatible with low-level samples separated on precast minigels (Phastgels, Amersham Pharmacia Biotech).

Although 2D electrophoresis is now widely used for the separation of complex protein mixtures from whole cell lysates for proteome studies, one of the major limitations is that trace components are frequently not detected, often being overshadowed by more abundant proteins.[125,126] This can be avoided by the use of preliminary fractionation, either chromatographically using micropreparative columns to facilitate gel loading, using preparative electrophoretic techniques such as free-flow isoelectric focusing,[127] by subcellular fractionation[126] or by a combination of these techniques.

6.5 Database Searching Strategies

The combination of 2D electrophoretic separations, micro LC and MS is complemented by another emerging technology, peptide mass fingerprinting,[95] in which observed masses are compared with computer-generated peptide molecular weights derived from site-specific proteolytic cleavage of proteins whose sequences are already in the continually expanding DNA and protein databases. For this technology, MALDI/TOF/MS or capillary HPLC/ESIMS have been used to generate the required peptide mass information from chemical or

enzymatic digests of the unknown protein of interest. It has been suggested that as few as 5–7 peptides are required for unambiguous matching with a particular protein present in the databases.[94] This technique is particularly effective when studying organisms where the complete nucleotide sequence of the genome is available.

A second method involves the use of data generated by MS/MS.[95] This generates highly specific information in the fragmentation pattern as well as sequence information and is a general solution when solving data from incompletely resolved protein mixtures. A database (SEQUEST) has been developed[128] to use the fragmentation pattern to identify the amino acid sequence represented by a tandem mass spectrum.

6.6 Synergies Between Micropreparative High-performance Liquid Chromatography and Instrumental Biosensors

The application of molecular genetics and recombinant DNA technology in biomedical research and the development of the Human Genome project have led to the isolation of many genes encoding proteins whose biological function and possible binding partners remain unknown. Furthermore, an exponential number of new compounds with potential therapeutic effects have been produced by combinatorial techniques or obtained from natural product libraries. New strategies and complementary technologies including novel affinity methods are therefore required for the screening and functional characterization of these proteins and small molecules. In response to this need, interdisciplinary studies in chemistry, electronics and biology have led to the research and development of a number of novel interfaces for biomolecular electronic devices and sensors. Reflecting this impetus in research and development, a number of instrumental biosensors, suitable for a wide range of research applications, are emerging and are rapidly becoming a major tool in the field of biomedical research, in particularly in the area of biomolecular interaction.[129] The optical biosensors [BIAcore™ range from BIAcore, Uppsala, Sweden (http://www.biacore.com) and IAsys systems from Affinity Sensors, Cambridge, UK (http://www.affinity-sensors.com)] are successful illustrations of this new generation of instrumental biosensors.

These evanescent wave sensors do not require the use of labeled reagents for detection since signals are generated through a change in mass loading when an interaction occurs. The BIAcore™ uses the technique of surface plasmon resonance[130] (see also **Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis**) while IAsys uses a waveguide technique called a prism coupler or resonant mirror.[131] These instruments continuously monitor the resonance angle and thus can detect changes in the refractive index caused by changes in sensor surface mass. The BIAcore™ sensor surface consists of a glass slide coated with a thin (50 nm) gold film to which is attached, by an inert (alkanethiol) linker layer, a chemical matrix on to which one of the binding partners can be immobilized using well-defined chemistries. The IAsys sensor surface is the bottom of a microcuvette possessing its own integrated optics; a choice of derivatized surfaces is available.

Specific ligands (proteins, peptides, carbohydrates, lipids, DNA) can be immobilized on the sensor surface at low concentration using various strategies: covalent attachment through amine, thiol or aldehyde chemistry,[132,133] capture via biotin–avidin interaction or specific tags[134] or immobilization using HI chemistries.[135] Real-time binding studies can be performed by sequentially flowing reagents of interest over the biospecific sensor surface. Concentration and kinetic data can be obtained by analyzing the binding curves generated.[136–140] Recovery of specifically bound material for downstream analysis (e.g. MS, N-terminal sequence analysis, PAGE) is possible with these instruments. In particular, the microcuvette of the IAsys sensor is very suitable for this purpose owing to the availability of a high-capacity sensor surface (CMD select cuvette, 10–20-μg range), the potential for long contact times and the confined environment, which facilitates recovery postdesorption in small eluate volumes.

6.6.1 Sample Compatibility Between Micropreparative High-performance Liquid Chromatography and Biosensor Analysis

Optical biosensors have been shown to be capable of detecting biomolecular interactions at low nanograms per milliliter concentrations, corresponding to levels of ligand often observed in biological samples.[141] However, absolute sensitivity for a particular interaction depends on several factors, including the relative masses of the interactants, immobilization level, association and dissociation rate constants for the interaction and injection volume. The autoinjection systems of these biosensors are capable of accurately injecting samples down to 1 μL. As we have described above, short micropreparative HPLC columns (1–2-mm i.d.) can detect low levels (nanograms to micrograms) of proteins or peptides with sample recovery in small volumes (20–100 μL) at high concentration. Therefore, the sensitivity and sample recovery volumes characteristic of micropreparative HPLC are complementary with the biosensor requirements, allowing the two instruments to be used in a ''bidirectional

synergy'': micropreparative HPLC can generate highly purified reagents in a form suitable for the biosensor, and the biosensor can provide specific information to aid the chromatographic process.[142] Indeed, the biosensor can be considered as a micropreparative affinity surface with on-line detection comparable to a chromatographic immunoaffinity system. Samples recovered from micropreparative columns can often be analyzed without further manipulation by taking samples from the micropreparative HPLC fraction collector directly to the autosampler racks of the biosensor, except following RPHPLC where volatile organic and acidic solvents are either buffer exchanged using micropreparative SEC or removed by lyophilization prior to sensor analysis. In a further example of the compatibility between chromatography and biosensor analysis, the column effluent of a Superose 75 size-exclusion column has been coupled directly to an optical biosensor:[143] the elution of a human scFV against 2-phenyloxazol-5-one (OX) present in concentrated conditioned media was detected on-line using immobilized OX–BSA hapten on the sensor surface.

6.6.2 The Use of Micropreparative High-performance Liquid Chromatography to Generate Reagents for the Biosensor

Biosensor specificity depends critically on the homogeneity of immobilized material. Samples should be purified to homogeneity and characterized immediately prior to immobilization, since heterogeneity of immobilized material has been shown to cause complex kinetics.[144–146] It is also important to remove high-molecular-weight aggregates, which might display avidity effects complicating kinetic studies.[143] Micropreparative SEC or desalting columns are ideally suited for the buffer exchange of samples into immobilization or running buffers prior to analysis. This purification step minimizes buffer-related refractive index changes (evanescent wave optical detectors are exquisitely refractive index sensitive) which may hinder interpretation of binding data for weak or low-level interactions, and removes buffer components which might interfere with subsequent binding (e.g. inappropriate pH, high salt concentrations, amine-containing buffers which would poison the *N*-hydroxysuccinimide/*N*-ethyl-*N*′-dimethylaminopropyl carbodiimide immobilization).

6.6.3 The Use of Biosensors to Monitor Chromatographic Purification

Having immobilized a specific reagent on the sensor surface, the biosensor can be used to screen a wide range of biological extracts (e.g. tissue and cell culture conditioned media, cell extracts, biological fluids, combinatorial libraries) for possible sources of interacting proteins. These samples can be screened automatically using the robotic autosamplers of automated instruments. This screening approach has been shown to be particularly useful in the search for unknown ligands of, for example, monoclonal antibodies raised against complex mixture of proteins[37] or orphan receptors identified by the polymerase chain reaction or database-searching strategies.[129,141] In such cases, the sensor can then be used as a specific detector for monitoring fractions during chromatographic purification of the ligand of interest,[142] for detailed structure–function studies or may even be used directly as a micropreparative affinity surface.[147–150] Since the instrument can measure concentration, recovery and specific activity data can be readily calculated by reference to an appropriate calibration graph. In the case of novel proteins, this can be calculated retrospectively.[37]

Obviously, the complexity of a specific purification protocol will depend largely on the relative abundance of the ligand and the nature of the contaminant proteins present. In the identification and purification of the ligand for the orphan human HPH-like receptor tyrosine kinase (HEK) from human placental conditioned medium,[151] where the specific biosensor response observed during the initial screening was only 10 RU (close to the detection limit of the system), seven chromatographic steps were necessary to purify the ligand to homogeneity for amino acid sequence analysis, including micropreparative anion exchange (MonoQ PC1.6/5) followed by micropreparative RPHPLC (Brownlee RP 300, 50 × 1-mm i.d.). Using this protocol, 1.5 μg of protein was recovered with a purification factor of 1.8×10^6 and an overall yield of 9%.

6.7 Liquid Chromatography/Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy used in conjunction with HPLC and MS has the potential to generate both chemical and structural information for the characterization of small molecules separated from complex mixtures. LC/NMR analysis has been applied in the fields of polymer, synthetic combinatorial and clinical chemistries and also in pharmaceutical, natural product and environmental analysis.

The current sensitivity obtained in conventional HPLC/NMR in a continuous-flow mode is in the low-micrograms range.[152] However, stop-flow analysis, where the flow is stopped after the peak has entered the detector cell prior to NMR analysis, increases the sensitivity to the high-nanograms level. The volumes of NMR flow cells used in conventional HPLC/NMR are in the range 60–180 μL, which can lead to a loss

of chromatographic resolution. However, the introduction of a second dimension, the proton chemical shift, results in partial compensation of the chromatographic dispersion and allows the differentiation of co-eluting peaks.[152] The current direction in research and development is towards the miniaturization of the coupled chromatographic separation NMR detection system [e.g. development of microcoil probe heads,[153] interfacing with micro[154] or capillary HPLC[152]]. This miniaturization leads to a reduction in the use of solvents (making the use of fully deuterated solvents feasible) and therefore removes the need for solvent suppression, which leads to peak distortion.

The recovery of samples at high concentration from microbore and capillary HPLC is compatible with the use of microcoil detectors in NMR. These microcoils are suited for mass-limited applications (e.g. biological analysis, combinatorial libraries). It has been calculated that a 400-fold change in cell volume will lead to only a marginal reduction in signal-to-noise ratio.[154] In practice, the use of a 31-nL microcoil actually gave a 10-fold enhancement in mass sensitivity compared with a 40-μL nanoprobe, allowing the acquisition of a detectable signal in 9 min compared with 15 h with the nanoprobe. The technical aspect of combining microcoil and capillary HPLC and the future direction of microcoil NMR have been reviewed by Olson et al.[155]

7 FUTURE DIRECTIONS

The potential of microsystems in the field of the separation sciences is now well established, in particular for the recovery of trace components in bulk biological systems. In the field of micro LC systems, further advances in instrumental design, particularly in the direct generation of gradients at submicroliter flow rates for capillary systems, is likely to facilitate operation.

The major benefits of such systems are realized by coupling them, either on- or off-line, with downstream sensitive and specific methods of detection. Current methods are compatible with femtomole levels of material for many applications, sufficient even for the analysis of single cell constituents in certain biological systems.[156] However, the search for even more sensitive methods, aiming ultimately at single molecule detection, will continue to be a priority. To this end, it is worth noting that atomic force microscopy has recently been shown to be compatible with the characterization of ligand-binding forces in molecular repertoires at the single molecule level.[157]

Another area that will almost certainly see further interesting developments in the near future is that of column packing materials and column technology. Affinity-based systems will almost certainly continue to gain prominence, often exploiting the exquisite selectivity obtainable with direct on-line analysis.[158–161] Molecular imprinting, which creates selective recognition sites by polymerization of functional monomers in the presence of the imprint molecule, should play a role in this area.[162] Particle size will continue to drop, or be replaced by monolithic stationary phases (where the packing is a continuous porous bed) or open-tubular columns in which the stationary phase is bonded to the column wall. These approaches are particularly amenable to CEC, where flow is not pressure generated.

However, perhaps the main limitation of existing micro LC systems is in the potential for high-throughput screening, since even with column-switching devices the number of concurrent analyses which can be performed is limited. An automated system incorporating an affinity extraction column and two RP columns coupled to a mass spectrometer has been described for the characterization of combinatorial chemical libraries of small drug compounds.[163] However, this was only illustrated using a small 20-component library, a far cry from the current state of the art. The use of microfabricated systems may increase throughput.

7.1 Microfabrication Technology for Miniaturized Separation Systems

Microfabrication technology is suitable for both automation and integration, and is ideally suited for massively parallel high-throughput tasks. Additionally, the use of such systems could further reduce both sample losses and the introduction of contaminating proteins from the environment (e.g. keratins), a major problem for the micromanipulation of samples at the sub-nanogram level.[66] These devices are usually fabricated from silicon, glass or plastic, often using techniques emanating from the microelectronics industry (e.g. photolithography, chemical etching, thermal bonding). An ultimate goal of this research is the development of fully integrated miniaturized analytical systems: a ''lab-on-a-chip''.[164] Chips with heaters, valves, pumps, microfluidics and sensitive detection systems have been described, in addition to integrated systems incorporating several techniques.[164]

Reflecting these developments, a number of instruments have already been described for LC,[165,166] CE[167,168] and CEC,[169] particularly for direct coupling to MS.[170–173]

ABBREVIATIONS AND ACRONYMS

ATZ	Anilinothiazolinone
BSA	Bovine Serum Albumin

CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CZE	Capillary Zone Electrophoresis
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization Mass Spectrometry
ESIMS/MS	Electrospray Ionization Tandem Mass Spectrometry
HEK	HPH-like Receptor Tyrosine Kinase
HETP	Height Equivalent to a Theoretical Plate
HFBA	Heptafluorobutyric Anhydride
HI	Hydrophobic Interaction
HPLC	High-performance Liquid Chromatography
IEX	Ion-Exchange
LC	Liquid Chromatography
MALDI/TOF	Matrix-assisted Laser Desorption/Ionization Time-of-Flight
MHC	Major Histocompatibility Complex
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
OX	2-phenyloxazol-5-one
PAGE	Polyacrylamide Gel Electrophoresis
PVDF	Poly(vinylidene Fluoride)
RP	Reversed-phase
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size-exclusion Chromatography
TFA	Trifluoroacetic Acid
UV	Ultraviolet
2D	Two-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Chromatography of Membrane Proteins and Lipoproteins • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Proteolytic Mapping • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis • Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Biopolymer Chromatography • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. J.D. Pearson, W.C. Mahoney, M.A. Hermodson, F.E. Regnier, 'Reversed-phase Supports for the Resolution of Large Denatured Protein Fragments', *J. Chromatogr.*, **207**, 325–332 (1981).
2. E.C. Nice, M. Capp, N.H.C. Cooke, M.J. O'Hare, 'A Comparison of Short and Ultrashort-chain Alkylsilane-bonded Silicas for the High-performance Liquid Chromatography of Proteins by Hydrophobic Interaction Methods', *J. Chromatogr.*, **218**, 569–580 (1981).
3. K.K. Unger, R. Janzen, G. Jilge, 'Packings and Stationary Phases for Biopolymer Separations by HPLC', *Chromatographia*, **24**, 144–154 (1987).
4. F.E. Regnier, 'HPLC of Biological Macromolecules: The First Decade', *Chromatographia*, **24**, 241–259 (1987).
5. K. Stulik, V. Pacakova, J. Suchankova, H.A. Claessens, 'Stationary Phases for Peptide Analysis by High-performance Liquid Chromatography', *Anal. Chim. Acta*, **352**, 1–19 (1997).
6. E.C. Nice, R.J. Simpson, N.A. Nicola, 'A Micropreparative Chromatographic Strategy for the Purification and Sequence Analysis of Murine Granulocyte-colony Stimulating Factor (mG-CSF)', *Chromatographia*, **24**, 449–468 (1987).

7. F.S. Esch, 'Polypeptide Microsequence Analysis with the Commercially Available Gas-phase Sequencer', *Anal. Biochem.*, **136**, 39–47 (1984).
8. A.D. Strosberg, L. Camoin, J.L. Guillaume, 'Microsequencing of Peptides and Proteins', *Trends Anal. Chem.*, **8**, 292–298 (1989).
9. B. Grego, I.R. Van Driel, P.A. Stearne, J.W. Goding, E.C. Nice, R.J. Simpson. 'A Microbore High-performance Liquid Chromatography Strategy for the Purification of Polypeptides for Gas-phase Sequence Analysis. Structural Studies on the Murine Transferrin Receptor', *Eur. J. Biochem.*, **148**, 485–491 (1985).
10. M.W. Hunkapiller, R.M. Hewick, W.J. Dreyer, L.E. Hood, 'High-sensitivity Sequencing with a Gas-phase Sequenator', *Methods Enzymol.*, **91**, 399–413 (1983).
11. J.M. Zanelli, J.C. Kent, B. Rafferty, R.A. Nissenson, E.C. Nice, M.W. Capp, M.J. O'Hare, 'High-performance Liquid Chromatographic Methods for the Analysis of Human Parathyroid Hormone in Reference Standards, Parathyroid Tissue and Biological Fluids', *J. Chromatogr.*, **276**, 55–68 (1983).
12. M.J. O'Hare, M.W. Capp, E.C. Nice, N.H. Cooke, B.G. Archer, 'Factors Influencing Chromatography of Proteins on Short Alkylsilane-bonded Large Pore-size Silicas', *Anal. Biochem.*, **126**, 17–28 (1982).
13. N.H.C. Cooke, B.G. Archer, M.J. O'Hare, E.C. Nice, M. Capp, 'Effects of Chain Length and Carbon Load on the Performance of Alkyl-bonded Silicas for Protein Separations', *J. Chromatogr.*, **255**, 115–123 (1982).
14. D. Ishii, T. Tsuda, K. Ibi, T. Takeuchi, T. Nakanishi, 'Study of Open-tubular Micro-capillary Liquid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**, 21–25 (1979).
15. R.P.W. Scott, P. Kucera, 'Use of Microbore Columns for the Separation of Substances of Biological Origin', *J. Chromatogr.*, **185**, 27–41 (1979).
16. R.J. Simpson, E.C. Nice, 'Strategies for the Purification of Subnanomole Amounts of Protein and Polypeptides for Microsequence Analysis', in *The Use of HPLC in Receptor Biochemistry*, ed. A.R. Kerlavage, Alan R. Liss, New York, 201–244, Vol. 8, 1989.
17. P. Kucera, 'Design and Use of Short Microbore Columns in Liquid Chromatography', *J. Chromatogr.*, **198**, 93–109 (1980).
18. J.J. Van Deemter, F.J. Zuiderweg, A. Klinkenberg, 'Longitudinal Diffusion and Resistance to Mass Transfer as Causes for Non-ideality in Chromatography', *Chem. Eng. Sci.*, **5**, 271–289 (1956).
19. R.J. Simpson, R.L. Moritz, G.S. Begg, M.R. Rubira, E.C. Nice, 'Micropurification Procedures for High Sensitivity Sequencing of Peptides and Proteins', *Anal. Biochem.*, **177**, 1–16 (1989).
20. K.J. Wilson, 'An Introduction to Protein/Peptide Isolation Strategies for Primary Structure Determinations', *Applied Biosystems User Bull.*, No. 1 (1985).
21. E.C. Nice, C.J. Lloyd, A.W. Burgess, 'The Role of Short Microbore HPLC Columns for Protein Separation and Trace Enrichment', *J. Chromatogr.*, **296**, 153–170 (1984).
22. E.C. Nice, R.J. Simpson, 'Micropreparative High-performance Liquid Chromatography of Proteins and Peptides', *J. Pharm. Biomed. Anal.*, **7**, 1039–1053 (1989).
23. R.L. Moritz, R.J. Simpson, 'Purification of Proteins and Peptides for Sequence Analysis Using Microcolumn Chromatography', *J. Microcol. Sep.*, **4**, 485–489 (1992).
24. K.K. Unger, G. Jilge, J.N. Kinkel, M.T. Hearn, 'Evaluation of Advance Silica Packing for the Separation of Biopolymers by High-performance Liquid Chromatography. Performance of Non-porous Monodisperse 1.5 μm Silica Bead in the Separation of Proteins by Reversed-phase Gradient Elution HPLC', *J. Chromatogr.*, **359**, 61–72 (1986).
25. D.J. Burke, J.K. Duncan, C. Siebert, G.S. Ott, 'Rapid Cation-exchange Chromatography of Hemoglobins and Other Protein', *J. Chromatogr.*, **359**, 533–540 (1986).
26. K. Kalghatgi, C. Horvath, 'Rapid Analysis of Proteins and Peptides by Reversed-phase Chromatography', *J. Chromatogr.*, **398**, 335–339 (1987).
27. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, 2nd edition, John Wiley & Sons, New York, 1979.
28. F.E. Regnier, 'High-performance Liquid Chromatography of Proteins', *Methods Enzymol.*, **91**, 137–190 (1983).
29. J.D. Pearson, N.T. Lin, F.E. Regnier, 'High-performance Liquid Chromatography of Proteins', *Anal. Biochem.*, **124**, 217–230 (1982).
30. E.C. Nice, B. Grego, R.J. Simpson, 'Application of Short Microbore HPLC Columns for the Preparation of Samples for Protein Microsequencing', *Biochem. Int.*, **11**, 187–195 (1985).
31. K. Unger, 'High-performance Size-exclusion Chromatography', *Methods Enzymol.*, **104**, 154–169 (1984).
32. J.D. Pearson, 'High-performance Liquid Chromatography Column Length Designed for Submicrogram Scale Protein Isolation', *Anal. Biochem.*, **152**, 189–198 (1986).
33. T.D. Schlabach, K.J. Wilson, 'Microbore Flow-rates and Protein Chromatography', *J. Chromatogr.*, **385**, 65–74 (1987).
34. R.L. Moritz, G.E. Reid, L.D. Ward, R.J. Simpson, 'Capillary HPLC: A Method for Protein Isolation and Peptide Mapping', *Methods: Compan. Methods Enzymol.*, **6**, 213–226 (1994).
35. E.C. Nice, L. Fabri, A. Hammacher, K. Andersson, U. Hellman, 'Micropreparative Ion Exchange High-performance Liquid Chromatography: Applications to Microsequence Analysis', *Biomed. Chromatogr.*, **7**, 104–111 (1993).
36. E.C. Nice, L. Fabri, A. Hammacher, J. Holden, H. Maruta, R.J. Simpson, A.W. Burgess, 'The Purification of a RAP 1 Dependent GTP′ase Activating Protein from

Bovine Brain Cytosol', *J. Biol. Chem.*, **267**, 1546–1553 (1992).
37. B. Catimel, G. Ritter, S. Welt, L.J. Old, L. Cohen, M.A. Nerrie, S.J. White, J.K. Heath, B. Demediuk, T. Domagala, F.T. Lee, A.M. Scott, G.F. Tu, H. Ji, R.L. Moritz, R.J. Simpson, A.W. Burgess, E.C. Nice, 'Purification and Characterization of a Novel Restricted Antigen Expressed by Normal and Transformed Human Colonic Epithelium', *J. Biol. Chem.*, **271**, 25 664–25 670 (1996).
38. G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, 'Comprehensive On-line LC/LC/MS of Proteins', *Anal. Chem.*, **69**, 1518–1524 (1997).
39. G.J. Opiteck, S.M. Ramirez, J.W. Jorgenson, M.A. Moseley III, 'Comprehensive Two-dimensional High-performance Liquid Chromatography for the Isolation of Overexpressed Proteins and Proteome Mapping', *Anal. Biochem.*, **258**, 349–361 (1998).
40. M.V. Novotny, 'Microcolumn Liquid Chromatography in Biochemical Analysis', *Methods Enzymol.*, **270**, 101–133 (1996).
41. J.P.C. Vissers, H. Cleassens, C.A. Cramers, 'Microcolumn Liquid Chromatography: Instrumentation, Detection and Application', *J. Chromatogr. A.*, **779**, 1–28 (1997).
42. T. Takeuchi, D. Ishii, 'Microcolumn Liquid Chromatography', *Methods Chromatogr.*, **1**, 168–198 (1996).
43. E.C. Nice, 'Micropreparative HPLC of Proteins and Peptides: Principles and Applications', *Biopolymers (Pept. Sci.)*, **40**, 319–341 (1996).
44. C. Ericson, S. Hjerten, 'Pump Based on Thermal Expansion of a Liquid for Delivery of a Pulse-free Flow Particularly for Capillary Chromatography and Other Microvolume Application', *Anal. Chem.*, **70**, 366–372 (1998).
45. K.O. Eriksson, A. Palm, S. Hjerten, 'Preparative Capillary Electrophoresis Based on Adsorption of the Solutes (Proteins) onto a Moving Blotting Membranes as They Migrate Out of the Capillary', *Anal. Biochem.*, **201**, 211–215 (1992).
46. J.L. Liao, C.M. Zeng, A. Palm, S. Hjerten, 'Continuous Beds for Microchromatography: Detection of Proteins by a Blotting Membrane Technique', *Anal. Biochem.*, **241**, 195–198 (1996).
47. E.C. Nice, 'Micropreparative LC of Proteins', *Nature*, **348**, 462–463 (1990).
48. R. Grimm, M. Serve, J.P. Chervet, 'Capillary LC Using Conventional HPLC Instrumentation', *LC/GC*, **15**, 960–967 (1997).
49. J.J. Kirkland, W.W. Yau, H.J. Stoklosa, C.H. Dilks, 'Sampling and Extra-column Effects in High-performance Liquid Chromatography; Influence of Peak Skew on Plate Count Calculations', *J. Chromatogr. Sci.*, **15**, 303–316 (1977).
50. L. Jardine, 'Electrospray Ionization Mass Spectrometry of Biomolecules', *Nature*, **345**, 747–748 (1990).
51. P. Roepstorff, 'Mass Spectrometry of Proteins', *Trends Anal. Chem.*, **12**, 413–421 (1993).
52. D. Ishii, T. Takeuchi, 'New Interfacing Techniques for Capillary High-performance Liquid Chromatography–Mass Spectrometry', *Trends Anal. Chem.*, **8**, 25–29 (1989).
53. M.T. Davis, T.D. Lee, 'Analysis of Peptide Mixtures by Capillary High-performance Liquid Chromatography: A Practical Guide to Small-scale Separations', *Protein Sci.*, **1**, 935–944 (1992).
54. J.W. Jorgenson, E.J. Guthrie, 'Liquid Chromatography in Open-tubular Columns. Theory of Column Optimization with Limited Pressure and Analysis Time, and Fabrication of Chemically Bonded Reversed-phase Columns on Etched Borosilicate Glass Capillaries', *J. Chromatogr.*, **255**, 335–338 (1983).
55. J.C. Chervet, R.E.J. van Soest, M. Ursem, 'Z-shaped Flow Cell for UV Detection in Capillary Electrophoresis', *J. Chromatogr.*, **543**, 439–449 (1991).
56. K.L. Hsi, S.A. O'Neill, D.R. Dupont, P.M. Yuan, 'Visualization of Proteins by Modification of Lysines, Cysteines, and Phosphorylated Serines Facilitates Sample Preparation for Microsequencing', *Anal. Biochem.*, **258**, 38–47 (1998).
57. D. Tong, R.L. Moritz, J.S. Eddes, G.E. Reid, R.K. Rasmussen, D.S. Dorow, R.J. Simpson, 'Fabrication of Stable Packed Capillary Reversed-phase Columns for Protein Structural Analysis', *J. Protein Chem.*, **16**, 425–431 (1997).
58. C.G. Huber, G. Choudhary, C. Horvath, 'Capillary Electrochromatography with Gradient Elution', *Anal. Chem.*, **69**, 4429–4436 (1997).
59. R.T. Kennedy, J.W. Jorgenson, 'Preparation and Evaluation of Packed Capillary Liquid Chromatography Columns with Inner Diameter from 20 to 50 μm', *Anal. Chem.*, **61**, 1128–1135 (1989).
60. G.E. Reid, R.K. Rasmussen, D.S. Dorow, R.J. Simpson, 'Capillary Column Chromatography Improves Sample Preparation for Mass Spectrometric Analysis: Complete Characterization of Human Alpha-enolase from Two-dimensional Gels Following In Situ Proteolytic Digestion', *Electrophoresis*, **19**, 946–955 (1998).
61. D.Y. Chen, N.J. Dovichi, 'Single-molecule Detection in Capillary Electrophoresis: Molecular Shot Noise as a Fundamental Limit to Chemical Analysis', *Anal. Chem.*, **68**, 690–696 (1996).
62. K. Peck, L. Stryer, A.N. Glazer, R.A. Mathies, 'Single-molecule Fluorescence Detection: Autocorrelation Criterion and Experimental Realization with Phycoerythrin', *Proc. Natl. Acad. Sci. USA*, **86**, 4087–4091 (1989).
63. M.A. Cousino, T.B. Jarbawi, H.B. Halsall, W.R. Heineman, 'Pushing Down the Limits of Detection: Molecular Needles in a Haystack', *Anal. Chem.*, **69**, 544A–549A (1997).

64. E. Svoko, 'Protein and Peptide Analysis by Capillary Zone Electrophoresis and Micellar Electrokinetic Chromatography', *Electrophoresis*, **18**, 74–81 (1997).
65. D. Figeys, A. Ducret, I.V. Osstveen, R. Aebersold, 'Protein Identification by Capillary Zone Electrophoresis/Microelectrospray Ionization–Tandem Mass Spectrometry at the Subfemtomole Level', *Anal. Chem.*, **68**, 1822–1828 (1996).
66. G.A. Valaskovic, N.L. Kelleher, F.W. McLafferty, 'Attomole Protein Characterization by Capillary Electrophoresis–Mass Spectrometry', *Science*, **273**, 1199–1202 (1996).
67. R. Zhang, S. Hjerten, 'A Micromethod for Concentration and Desalting Utilizing Hollow Fiber, with Special Reference to Capillary Electrophoresis', *Anal. Chem.*, **69**, 1585–1592 (1997).
68. D. Figeys, A. Ducret, J.R. Yates III, R. Aebersold, 'Protein Identification by Solid Phase Microextraction–Capillary Zone Electrophoresis–Microelectrospray–Tandem Mass Spectrometry', *Nature Biotechnol.*, **14**, 1579–1583 (1996).
69. M.M. Bushey, J.W. Jorgenson, 'Automated Instrumentation for Comprehensive Two-dimensional High-performance Liquid Chromatography/Capillary Zone Electrophoresis', *Anal. Chem.*, **62**, 978–984 (1990).
70. A.V. Lemmo, J.W. Jorgenson, 'Transverse Flow Gating Interface for the Coupling of Microcolumn Liquid Chromatography in a Comprehensive Two-dimensional System', *Anal. Chem.*, **65**, 1576–1581 (1993).
71. J.P. Larmann Jr, A.V. Lemmo, A.W. Moore Jr, J.W. Jorgenson, 'Two-dimensional Separations of Peptides and Proteins by Comprehensive Liquid Chromatography–Capillary Electrophoresis', *Electrophoresis*, **14**, 439–447 (1993).
72. L.A. Colon, K.J. Reynolds, R.A. Maldonado, A.M. Fermier, 'Advances in Capillary Electrochromatography', *Electrophoresis*, **18**, 2162–2174 (1997).
73. J.J. Peseck, M.T. Matyska, 'Column Technology in Capillary Electrophoresis and Capillary Electrochromatography', *Electrophoresis*, **18**, 2228–2238 (1997).
74. J.H. Knox, I.H. Grant, 'Miniaturization in Pressure and Electroendosmotically Driven Liquid Chromatography: Some Theoretical Considerations', *Chromatographia*, **24**, 135–145 (1987).
75. J.E. MacNair, G.J. Opiteck, J.W. Jorgenson, M.A. Moseley III, 'Rapid Separation and Characterization of Protein and Peptide Mixtures Using 1.5 μm Diameter Non-porous Silica in Packed Capillary Liquid Chromatography/Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 1279–1285 (1997).
76. M.R. Taylor, S.A. Westwood, D. Perrett, 'Direct Monitoring of Enzyme Reactions Using Micellar Electrokinetic Capillary Chromatography. Optimization of Drug Glucuronide and Sulfate Conjugate Hydrolysis', *J. Chromatogr. A*, **768**, 67–71 (1997).
77. C. Yan, R. Dadoo, R.N. Zare, D.J. Rakestraw, D.S. Anex, 'Gradient Elution in Capillary Electrochromatography', *Anal. Chem.*, **68**, 2726–2730 (1996).
78. M.R. Taylor, P. Teale, S.A. Westwood, D. Perrett, 'Analysis of Corticosteroids in Biofluids by Capillary Electrochromatography with Gradient Elution', *Anal. Chem.*, **69**, 2555–2558 (1997).
79. P. Huang, J.T. Wu, D.M. Lubman, 'Separation of Tryptic Digests Using a Modified Buffer in Pressurized Capillary Electro-chromatography with an Ion Trap Storage/Reflectron Time-of-flight Mass Spectrometer', *Anal. Chem.*, **70**, 3003–3006 (1998).
80. J.T. Wu, P. Huang, M.X. Li, D.M. Lubman, 'Protein Digest Analysis by Pressurized Capillary Electrochromatography Using an Ion Trap Storage/Reflectron Time-of-flight Mass Spectrometer', *Anal. Chem.*, **69**, 2908–2913 (1997).
81. R.J. Simpson, D.J. Hilton, E.C. Nice, M.R. Rubira, D. Metcalf, D.P. Gearing, N.M. Gough, N.A. Nicola, 'Structural Characterization of Murine Myeloid Leukaemia Inhibitory Factor', *Eur. J. Biochem.*, **175**, 541–547 (1988).
82. A.W. Burgess, D. Metcalf, I.J. Kozka, R.J. Simpson, G. Vairo, J.A. Hamilton, E.C. Nice, 'Purification of Two Forms of Colony Stimulating Factor from Mouse L-Cell Conditioned Media', *J. Biol. Chem.*, **260**, 16 004–16 011 (1985).
83. A.W. Burgess, D. Metcalf, L.G. Sparrow, R.J. Simpson, E.C. Nice, 'Granulocyte-macrophage Colony Stimulating Factor from Mouse Lung Condition Medium: Purification of Multiple Forms and Radio-iodination', *Biochem. J.*, **235**, 805–814 (1986).
84. R.J. Simpson, R.L. Moritz, M.R. Rubira, J. Van Snick, 'Murine Hybridoma/Plasmacytoma Growth Factor. Complete Amino-acid Sequence and Relation to Human Interleukin-6', *Eur. J. Biochem.*, **176**, 187–197 (1988).
85. R.H. Aebersold, J. Leavitt, R.A. Saavedra, L.E. Hood, S.B.H. Kent, 'Internal Amino Acid Sequence Analysis of Proteins Separated by One- or Two-dimensional Gel Electrophoresis After In Situ Protease Digestion on Nitrocellulose', *Proc. Natl. Acad. Sci. USA*, **84**, 6970–6974 (1987).
86. S.A. Hefta, R.J. Paxton, J.E. Shively, 'Sequence and Glycosylation Site Identity of Two Distinct Glycoforms of Nonspecific Cross-reacting Antigen as Demonstrated by Sequence Analysis and Fast Atom Bombardment Mass Spectrometry', *J. Biol. Chem.*, **265**, 8618–8626 (1990).
87. T.S. Jardetzky, W.S. Lane, R.A. Robinson, D.R. Madden, D.C. Wiley, 'Identification of Self Peptides Bound to Purified HLA-B27', *Nature*, **353**, 326–329 (1991).
88. M.W. Hunkapiller, J.E. Strickler, K.J. Wilson, 'Contemporary Methodology for Protein Structure Determination', *Science*, **226**, 304–311 (1984).
89. P.M. Yuan, S. Yuen, J. Bergot, M.H. Hunkapiller, K.J. Wilson, 'Progress Toward Polybrene Purification and

Utility in Micro-peptide/Protein Sequencing', in *Proteins, Structure and Function*, ed. J. L'Italien, Plenum Press, New York, 307–313, 1987.
90. W. Burkhart, 'In Situ Proteolytic Digestions Performed on Protein Bound to the Hewlett-Packard Hydrophobic Sequencing Column', in *Techniques in Protein Chemistry IV*, ed. R.H. Angeletti, Academic Press, San Diego, CA, 399–406, 1993.
91. D. Hess, T.C. Covey, R. Winz, R.W. Brownsey, R. Aebersold, 'Analytical and Micropreparative Peptide Mapping by High-performance Liquid Chromatography/Electrospray Mass Spectrometry of Proteins Purified by Gel Electrophoresis', *Protein Sci.*, **2**, 1342–1351 (1993).
92. A.M. Mahrenholz, C.H. Yeh, J.E. Shively, S.A. Hefta, 'Microsequence and Mass Spectral Analysis of Nonspecific Cross-reacting Antigen 160, a CD15-positive Neutrophil Membrane Glycoprotein. Demonstration of Identity with Biliary Glycoprotein', *J. Biol. Chem.*, **268**, 13 015–13 018 (1993).
93. K. Klarskov, D. Roecklin, B. Bouchon, J. Sabatie, A.V. Dorsselaer, R. Bischoff, 'Analysis of Recombinant *Schistosoma mansoni* Antigen rSmp28 by On-line Liquid Chromatography–Mass Spectrometry Combined with Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis', *Anal. Biochem.*, **216**, 127–134 (1994).
94. H. Ji, R.H. Whitehead, G.E. Reid, R.L. Moritz, L.D. Ward, R.J. Simpson, 'Two-dimensional Electrophoretic Analysis of Proteins Expressed by Normal and Cancerous Human Crypts: Application of Mass Spectrometry to Peptide Mass Fingerprinting', *Electrophoresis*, **15**, 391–405 (1994).
95. J.R. Yates III, 'Database Searching Using Mass Spectrometry Data', *Electrophoresis*, **10**, 893–900 (1998).
96. I. Humphery-Smith, S.J. Cordwell, W.P. Blackstock, 'Proteome Research: Complementarity and Limitations in Respect to the RNA and DNA Worlds', *Electrophoresis*, **18**, 1217–1242 (1998).
97. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**, 2299–2301 (1988).
98. F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**, 1193A–1203A (1991).
99. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, 'Electrospray Ionization for the Mass Spectrometry of Large Biomolecules', *Science*, **246**, 64–71 (1989).
100. R.C. Beavis, B.T. Chait, 'High-accuracy Molecular Mass Determination of Proteins Using Matrix-assisted Laser Desorption Mass Spectrometry', *Anal. Chem.*, **62**, 1836–1840 (1990).
101. O. Vorm, P. Roepstorff, M. Mann, 'Improved Resolution and Very High Sensitivity on MALDI TOF of Matrix Surfaces Made by Fast Evaporation', *Anal. Chem.*, **66**, 3281–3287 (1994).
102. M.W.F. Nielen, 'Polymer Analysis by Micro-scale Size-exclusion Chromatography/MALDI Time-of-flight Mass Spectrometry with a Robotic Interface', *Anal. Chem.*, **70**, 1563–1568 (1998).
103. N.J. Clarke, F. Li, A.J. Thomlison, S. Naylor, 'One Step Microelution Concentration Method for Efficient Coupling of Sodium Dodecylsulfate Gel Electrophoresis and Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry for Protein Analysis', *J. Am. Soc. Mass. Spectrom.*, **9**, 88–91 (1998).
104. B.T. Chait, R. Wang, R.C. Beavis, S.B.H. Kent, 'Protein Ladder Sequencing', *Science*, **262**, 89–91 (1993).
105. D. Figeys, R. Aebersold, 'High Sensitivity Analysis of Proteins and Peptides by Capillary Electrophoresis–Tandem Mass Spectrometry: Recent Developments in Technology and Applications', *Electrophoresis*, **19**, 885–892.
106. K. Mock, M. Hail, I. Mylchreest, J. Zhou, K. Johnson, L. Jardine, 'Rapid High-sensitivity Peptide Mapping by Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **646**, 169–174 (1993).
107. D.L. Smith, Z. Zhou, 'Strategies for Locating Disulfide Bonds in Proteins', *Methods Enzymol.*, **193**, 374–389 (1990).
108. Y. Abe, M. Odaka, F. Inagaki, I. Lax, J. Schlessinger, D. Kohda, 'Disulfide Bond Structure of Human Epidermal Growth Factor Receptor', *J. Biol. Chem.*, **273**, 11 150–11 157 (1998).
109. J. Ding, W. Burkhart, D.B. Kassel, 'Identification of Phosphorylated Peptides from Complex Mixture Using Negative-ion Orifice-potential Stepping and Capillary Liquid Chromatography Electrospray Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 94–98 (1994).
110. Z.I. Randhawa, H.E. Witkowska, J. Cone, J.A. Wilkins, P. Hughes, K. Yamanishi, S. Yasuda, Y. Masui, P. Arthur, C. Kletke, F. Bitsch, C.H.L. Shackleton, 'Incorporation of Norleucine at Methionine Positions in Recombinant Human Macrophage Colony Stimulating Factor (M-CSF, 4-153) Expressed in *Escherichia coli*: Structural Analysis', *Biochemistry*, **33**, 4352–4362 (1994).
111. V.H. Engelhard, 'Structure of Peptides Associated with MHC class I Molecules', *Curr. Opin. Immunol.*, **6**, 13–23 (1994).
112. D.F. Hunt, R.A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A.L. Cox, E. Appella, V.H. Engelhard, 'Peptides Presented to the Immune System by the Murine Class II Major Histocompatibility Complex Molecule I-Ad', *Science*, **255**, 1261–1263 (1992).
113. H.E. Witkowska, F. Bitsch, C.H. Shackleton, 'Expediting Rare Variant Hemoglobin Characterization by Combined HPLC/Electrospray Mass Spectrometry', *Hemoglobin*, **17**, 227–242 (1993).
114. O.N. Jensen, D.F. Barofsky, M.C. Young, P.H. von Hippel, S. Swenson, S.E. Seifried, 'Mass Spectrometric

Protocol for the Analysis of UV-crosslinked Protein–Nucleic Acid Complexes', in *Techniques in Protein Chemistry V*, Academic Press, San Diego, 27–37, 1994.

115. K. Klose, U. Kobalz, 'Two-dimensional Electrophoresis of Proteins: An Updated Protocol and Implication for a Functional Analysis of the Genome', *Electrophoresis*, **16**, 1034–1039 (1995).
116. R.J. Simpson, R.L. Moritz, E.C. Nice, B. Grego, 'A High-performance Liquid Chromatography Procedure for Recovering Subnanomole Amounts of Protein from SDS-gel Eluates for Gas-phase Sequence', *Eur. J. Biochem.*, **165**, 21–29 (1987).
117. J.D. Pearson, D.B. DeWald, H.A. Zurcher-Neely, R.L. Heinrikson, R.A. Poorman, 'Sequencing Proteins from Two-dimensional Gels Using a Liquid Pressure Extraction Technique', in *Methods in Protein Sequence Analysis*, ed. K.A. Walsh, Humana Press, Clifton, NJ, 295–302, 1986.
118. T. Tetaz, E. Bozas, J. Kanellos, I. Walker, I. Smith, 'In Situ Tryptic Digestion of Proteins Separated by SDS-PAGE: Improved Procedures for Extraction of Peptides Prior to Microsequencing', in *Techniques in Protein Chemistry IV*, ed. R.H. Angeletti, Academic Press, New York, 389–397, 1993.
119. L.D. Ward, G.E. Reid, R.L. Moritz, R.J. Simpson, 'Strategies for Internal Amino Acid Sequence Analysis of Proteins Separated by Polyacrylamide Gel Electrophoresis', *J. Chromatogr.*, **519**, 199–216 (1990).
120. R. Aebersold, J. Leavitt, L. Hood, S. Kent, in *Methods in Protein Sequence Analysis*, ed. K.A. Walsh, Humana Press, Clifton, NJ, 277–290, 1986.
121. H. Kawasaki, Y. Emori, K. Suzuki, 'Production and Separation of Peptides from Proteins Stained with Coomassie Brilliant Blue R-250 After Separation by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis', *Anal. Biochem.*, **191**, 332–336 (1990).
122. C. Eckerskorn, F. Lottspeich, 'Internal Amino Acid Sequence Analysis of Proteins Separated by Gel Electrophoresis After Tryptic Digestion in Polyacrylamide Matrix', *Chromatographia*, **28**, 92–94 (1990).
123. J. Rosenfeld, J. Capdevielle, J.C. Guillemot, P. Ferrara, 'In-gel Digestion of Proteins for Internal Sequence Analysis After One- or Two-dimensional Gel Electrophoresis', *Anal. Biochem.*, **203**, 173–179 (1992).
124. U. Hellman, S. Renlund, 'Micropurification of Peptides for Amino Acid Sequence Analysis after SDS-PAGE Separation', presented at the Seventeenth International Symposium on Column Liquid Chromatography, Hamburg, Germany, May 9–14, 1993.
125. H. Ji, R.L. Moritz, G.E. Reid, G. Ritter, B. Catimel, E.C. Nice, J.K. Heath, S.J. White, S. Welt, L.J. Old, A.W. Burgess, R.J. Simpson, 'Electrophoretic Analysis of the Novel Antigen for the Gastrointestinal-specific Monoclonal Antibody, A33', *Electrophoresis*, **18**, 614–621 (1997).
126. T. Rabilloud, S. Kieffer, V. Procaccio, M. Louwagie, P.L. Courchesne, S.D. Patterson, P. Martinez, J. Garin, J. Lunardi, 'Two-dimensional Electrophoresis of Human Placental Mitochondria and Protein Identification by Mass Spectrometry: Toward a Human Mitochondrial Proteome', *Electrophoresis*, **19**, 1006–1014 (1998).
127. G. Weber, P. Bocek, 'Recent Developments in Preparative Free Flow Isoelectric Focusing', *Electrophoresis*, **19**, 1649–1653 (1998).
128. J.R. Yates III, J.K. Eng, A.L. McCormack, 'Mining Genomes: Correlating Tandem Mass Spectra of Modified and Unmodified Peptides to Sequences in Nucleotide Databases', *Anal. Chem.*, **67**, 1426–1436 (1995).
129. E.C. Nice, B. Catimel, 'Instrumental Biosensors: New Perspectives for the Analysis of Biomolecular Interactions', *Bioessays*, **21**, 339–352 (1999).
130. M. Malmqvist, 'Biospecific Interaction Analysis Using Biosensor Technology', *Nature*, **361**, 186–187 (1993).
131. R.D. Davies, P.R. Edwards, H.J. Watts, C.R. Lowe, P.E. Buckle, D. Yeung, T.M. Kinning, D.V. Pollard-Knight, 'The Resonant Mirror: A Tool for the Study of Biomolecular Interactions', in *Techniques in Protein Chemistry V*, ed. J.W. Crabb, Academic Press, San Diego, 285–292, 1994.
132. D.J. O'Shannessy, M. Brigham-Burke, K. Peck, 'Immobilization Chemistries Suitable for Use in the BIAcore Surface Plasmon Resonance Detector', *Anal. Biochem.*, **205**, 132–136 (1992).
133. B. Johnsson, S. Lofas, G. Lindquist, A. Edstrom, R.M. Muller Hillgren, A. Hansson, 'Comparison of Methods for Immobilization to Carboxymethyl Dextran Sensor Surfaces by Analysis of the Specific Activity of Monoclonal Antibodies', *J. Mol. Recognit.*, **8**, 125–131 (1995).
134. J. Nilsson, M. Larsson, S. Stahl, P.A. Nygren, M. Uhlen, 'Multiple Affinity Domains for the Detection, Purification and Immobilization of Recombinant Proteins', *J. Mol. Recognit.*, **9**, 585–594 (1996).
135. B. Catimel, A.M. Scott, F.T. Lee, N. Hanai, G. Ritter, S. Welt, L.J. Old, A.W. Burgess, E.C. Nice, 'Direct Immobilization of Gangliosides onto Gold Carboxymethyldextran Sensor Surfaces by Hydrophobic Interaction: Applications to Antibody Characterization', *Glycobiology*, **8**, 927–938 (1998).
136. R. Karlsson, L. Fagerstam, H. Nilshans, B. Persson, 'Analysis of Active Antibody Concentration. Separation of Affinity and Concentration Parameters', *J. Immunol. Methods*, **166**, 75–84 (1993).
137. R. Karlsson, 'Kinetic Analysis of Monoclonal Antibody–Antigen Interactions with a New Biosensor Based Analytical System', *J. Immunol. Methods*, **145**, 229–240 (1991).
138. R. Karlsson, D. Altschuh, M.V.H. Van Regenmortel, 'Measurement of Antibody Affinity', in *Structure of Antigens*, ed. M.H.V. Van Regenmortel, CRC Press, Boca Raton, FL, 127–148, Vol. 1, 1992.

139. D.J. O'Shannessy, M. Brigham-Burke, K.K. Soneson, P. Hensley, I. Brooks, 'Determination of Rate and Equilibrium Binding Constants for Macromolecular Interactions Using Surface Plasmon Resonance: Use of Nonlinear Least Squares Analysis', *Anal. Biochem.*, **212**, 457–468 (1993).
140. L.D. Roden, D.G. Myszka, 'Global Analysis of a Macromolecular Interaction Measured on BIAcore', *Biochem. Biophys. Res. Commun.*, **225**, 1073–1077 (1996).
141. E.C. Nice, B. Catimel, M.S. Stacker, A. Runting, A. Wilks, N. Nicola, A.W. Burgess, 'Strategies for the Identification and Purification of Orphan Biomolecules', *Lett. Pept. Sci.*, **4**, 107–120 (1997).
142. E.C. Nice, M. Lackmann, F. Smyth, L. Fabri, A.W. Burgess, 'Synergies Between Micropreparative High-performance Liquid Chromatography and an Instrumental Optical Biosensor', *J. Chromatogr.*, **660**, 169–185 (1994).
143. M. Malmqvist, 'Real-time Biospecific Interaction Analysis for the Direct Measurement of Antibody–Antigen Interactions', *Immunotechnology Colloqium at 22nd FEBS Letters Meeting, Dublin*, 61–75, 1993.
144. E.C. Nice, T.L. McInerney, D.J. Jackson, 'Analysis of the Interaction Between a Synthetic Peptide of Influenza Virus Hemagglutinin and Monoclonal Antibodies Using an Optical Biosensor', *Mol. Immunol.*, **33**, 659 (1996).
145. D.G. Myszka, 'Kinetic Analysis of Macromolecular Interactions Using Surface Plasmon Resonance Biosensors', *Curr. Opin. Biotechnol.*, **8**, 50–57 (1997).
146. R. Karlsson, A. Falt, 'Experimental Design for Kinetic Analysis of Protein–Protein Interactions with Surface Plasmon Resonance Sensors', *J. Immunol. Methods*, **200**, 121–133 (1997).
147. B. Catimel, M. Nerrie, F.T. Lee, A.M. Scott, G. Ritter, S. Welt, L.J. Old, A.W. Burgess, E.C. Nice, Kinetic Analysis of the Interaction Between the Monoclonal Antibody A33 and Its Colonic Epithelial Antigen by the Use of an Optical Biosensor. A Comparison of Immobilization Strategies', *J. Chromatogr. A*, **776**, 15–30 (1997).
148. G. Ritter, L.S. Cohen, E.C. Nice, B. Catimel, A.W. Burgess, R.L. Moritz, H. Ji, J.K. Heath, S. White, S. Welt, L.J. Old, R.J. Simpson, 'Characterization of Post-translational Modifications of Human A33 Antigen, a Novel Palmitoylated Surface Glycoprotein of Human Gastrointestinal Epithelium', *Biochem. Biophys. Res. Commun.*, **236**, 682–686 (1997).
149. E.C. Nice, F. Smyth, T. Domagala, L. Fabri, B. Catimel, A.W. Burgess, 'Synergies Between Micropreparative HPLC and an Optical: Application to Structure–Function Studies', *Sci. Tools*, **1**, 3–5 (1996).
150. B. Catimel, T. Domagala, H. Nerrie, J. Weinstock, H. Abud, J. Heath, E. Nice, 'Recent Applications of Instrumental Biosensors for Protein and Peptide Structure–Function Studies', *Protein Pept. Lett.*, (in press).
151. M. Lackmann, T. Bucci, R.J. Mann, L.A. Kravets, E. Viney, F. Smith, R.L. Moritz, W. Carter, R.J. Simpson, N.A. Nicola, K. Mackwell, E.C. Nice, A.F. Wilks, A.W. Boyd, 'Purification of a Ligand for the EPH-like Receptor HEK Using a Biosensor-based Affinity Detection Approach', *Proc. Natl. Acad. Sci. USA*, **93**, 2523–2527 (1996).
152. K. Albert, U. Braumann, L.H. Tseng, G. Schlotterbeck, 'On-line ^{1}H NMR Spectroscopy', *Biomed. Chromatogr.*, **12**, 158–159 (1997).
153. D.L. Olson, T.K. Peck, A.G. Webb, R.L. Margin, J.V. Sweedler, 'High Resolution Microcoil ^{1}H-NMR for Mass-limited Nanoliter Volume Samples', *Science*, **270**, 1967–1970 (1997).
154. N. Wu, A. Webb, T.L. Peck, J.V. Sweedler, 'On-line Detection of Amino Acids and Peptides in Microbore LC', *Anal. Biochem.*, **67**, 3101–3107 (1995).
155. D.L. Olson, M.E. Lacey, J.V. Sweedler, 'The Nanoliter Niche. NMR Detection for Trace Analysis and Capillary Separations', *Anal. Chem.*, **70**, 257A–264A (1998).
156. S. Hsieh, K. Dreisewerd, R.C. van der Schors, C.R. Jimenez, J. Stahl-Zeng, F. Hillenkamp, J.W. Jorgenson, W.P. Geraerts, K.W. Li, 'Separation and Identification of Peptides in Single Neurons by Microcolumn Liquid Chromatography–Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry and Postsource Decay Analysis', *Anal. Chem.*, **70**, 1847–1852 (1998).
157. R. Ros, F. Schwesinger, D. Anselmetti, M. Kubon, R. Schafer, A. Pluckthun, L. Tiefenauer, 'Antigen Binding Forces of Individually Addressed Single-chain Fv Antibody Molecules', *Proc. Natl. Acad. Sci. USA*, **95**, 7402–7405 (1998).
158. R.W. Nelson, J.R. Krone, O. Jansson, 'Surface Plasmon Resonance Biomolecular Interaction Analysis Mass Spectrometry', *Anal. Chem.*, **69**, 4363–4368 (1997).
159. R.W. Nelson, 'The Use of Bioreactive Probes in Protein Characterization', *Mass Spectrom. Rev.*, **16**, 353–376 (1997).
160. H. Kuwata, T.T. Yip, C.L. Yip, M. Tomita, T.W. Hutchens, 'Bactericidal Domain of Lactoferrin: Detection, Quantitation, and Characterization of Lactoferricin in Serum by SELDI Affinity Mass Spectrometry', *Biochem. Biophys. Res. Commun.*, **245**, 764–773 (1998).
161. T.T. Yip, J. Van de Water, M.E. Gershwin, R.L. Coppel, T.W. Hutchens, 'Cryptic Antigenic Determinants on the Extracellular Pyruvate Dehydrogenase Complex/Mimeotope Found in Primary Biliary Cirrhosis. A Probe by Affinity Mass Spectrometry', *J. Biol. Chem.*, **271**, 32 825–32 833 (1996).
162. K. Mosbach, 'Molecular Imprinting', *Trends Biochem. Sci.*, **19**, 9–14 (1994).
163. M.L. Nedved, S. Habibi-Goudarzi, B. Ganem, J.D. Henion, 'Characterization of Benzodiazepine ''Combinatorial'' Chemical Libraries by On-line Immunoaffinity Extraction, Coupled Column HPLC–Ion Spray

Mass Spectrometry – Tandem Mass Spectrometry', *Anal. Chem.*, **68**, 4228 – 4236 (1996).
164. L.J. Kricka, 'Revolution on a Square Centimetre', *Nature Biotechnol.*, **16**, 513 – 514 (1998).
165. G. Ocvirk, E. Verpoorte, A. Manz, M. Grasserbauer, H.M. Widmer, 'High-performance Liquid Chromatography Partially Integrated onto a Silicon Chip', *Anal. Methods Instrum.*, **2**, 74 – 82 (1996).
166. J. Dzubian, A. Gorecka-Drzazga, L. Nieradko, J. Mroz, 'Microstructures and Microfabricated Systems', *Proc. Electrochem. Soc.*, **97**, 956 – 963 (1997).
167. D.J. Harrison, K. Fluri, K. Seiler, Z.H. Fan, C.S. Effenhauser, A. Manz, 'Micromachining a Miniaturized Capillary Electrophoresis-based Chemical Analysis System on Chip', *Science*, **261**, 895 – 897 (1993).
168. P.C. Simpson, D. Roach, A.T. Woolley, T. Thorsen, R. Johnston, G.F. Sensabaugh, R.A. Mathies, 'High-throughput Genetic Analysis Using Microfabricated 96-sample Capillary Array Electrophoresis Microplates', *Proc. Natl. Acad. Sci. USA*, **95**, 2256 – 2261 (1998).
169. S.C. Jacobson, R. Hergenroder, L.B. Koutry, J.M. Ramsey, 'Open Channel Chromatography on a Microchip', *Anal. Chem.*, **66**, 2369 – 2373 (1994).
170. D. Figeys, Y. Ning, R. Aebersold, 'A Microfabricated Device for Rapid Protein Identification by Microelectrospray Ion Trap Mass Spectrometry', *Anal. Chem.*, **69**, 3153 – 3160 (1997).
171. Q. Xue, F. Foret, Y.M. Dunayevskiy, P.M. Zavracky, N.E. McGruer, B.L. Karger, 'Multichannel Microchip Electrospray Mass Spectrometry', *Anal. Chem.*, **69**, 426 – 430 (1997).
172. Q. Xue, Y.M. Dunayevskiy, F. Foret, B.L. Karger, 'Integrated Multichannel Microchip Electrospray Ionization Mass Spectrometry: Analysis of Peptides from On-chip Tryptic Digestion of Melittin', *Rapid Commun. Mass Spectrom.*, **11**, 1253 – 1256 (1997).
173. R.S. Ramsey, J.M. Ramsey, 'Generating Electrospray from Microchip Devices Using Electroosmotic Pumping', *Anal. Chem.*, **69**, 1174 – 1178 (1997).

High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis

Tun-Li Shen
Brown University, Providence, RI, USA

1 Introduction **2**
2 Instrumentation of High-Performance Liquid Chromatography/Mass Spectrometry for Peptide and Protein Analysis **2**
2.1 Early Development of High-performance Liquid Chromatography/Mass Spectrometry 2
2.2 Electrospray Ionization 3
2.3 Instrumentation for High-performance Liquid Chromatography/Mass Spectrometry 3
2.4 Capillary High-performance Liquid Chromatography/Mass Spectrometry 5
2.5 Experimental Parameters 7
3 High-Performance Liquid Chromatography/Mass Spectrometry of Peptides **8**
3.1 Analysis of Peptides by Mass Spectrometry 8
3.2 Study of Natural and Synthetic Peptides by High-performance Liquid Chromatography/Mass Spectrometry 8
4 High-Performance Liquid Chromatography/Mass Spectrometry of Proteins **9**
4.1 General Strategy of Characterization of Proteins by Mass Spectrometry 9
4.2 Posttranslational Modification of Proteins 10
4.3 High-performance Liquid Chromatography/Mass Spectrometry Characterization of Phosphoproteins 11
4.4 High-performance Liquid Chromatography/Mass Spectrometry Characterization of Glycoproteins 13
4.5 Characterization of Recombinant Proteins by High-performance Liquid Chromatography/Mass Spectrometry 14
5 High-Performance Liquid Chromatography/Mass Spectrometry Analysis in the "Proteome" ERA **16**
5.1 Protein Analysis and Proteome 16
5.2 Analyzing Gel-separated Proteins by Mass Spectrometry 16
5.3 Protein Analysis by High-performance Liquid Chromatography/Tandem Mass Spectrometry 17
5.4 Analysis of Large-scale Proteomic Datasets Generated by High-throughput LC-MS/MS 19
Abbreviations and Acronyms **20**
Related Articles **21**
References **21**

Since the 1970s, high-performance liquid chromatography/mass spectrometry (HPLC/MS) has gone through many stages of development. The main technical challenge has been the coupling of liquid chromatography (LC) systems with mass spectrometers. Soft ionization techniques that have contributed to HPLC/MS for peptide and protein study include fast atom bombardment (FAB), matrix-assisted laser desorption/ionization (MALDI), and electrospray ionization (ESI). Among these techniques, FAB was developed first and used extensively in the 1980s; it lacked the sensitivity of the other two more recent techniques. MALDI and ESI are comparable in terms of sensitivity and the types of biomolecules that can be analyzed. However, MALDI has the slight advantage of being more tolerant to salts and able to analyze complex mixtures. On the other hand, the ability of ESI to interface mass spectrometry (MS) directly with LC makes it extremely useful for protein and peptide analysis. Another advantage that ESI offers is its ability to produce multiply charged protein ions, which allows the observation of large proteins with mass analyzers of limited mass range. In this article, we discuss the experimental parameters that are important in HPLC/MS and the considerable progress made in ESI source design and capillary HPLC/MS.

In the study of natural and synthetic peptides, HPLC/MS has been used to separate, characterize, and sequence the peptides being investigated. In protein analysis, HPLC/MS is most useful in characterizing protein posttranslation modification such as glycosylation and phosphorylation. We discuss specific HPLC/MS methods for detecting glycopeptides and phosphopeptides during chromatographic separation by monitoring marker ions: Hex^+ (m/z 163), $HexNAc^+$ (m/z 204), $NeuAc^+$ (m/z 292), and $HexHexNAc^+$ (m/z 366) for glycopeptides,

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

and PO_2^- (m/z 63) and PO_3^- (m/z 79) for phosphopeptides. Another important application of HPLC/MS is in proteomic analysis. Here, the most significant development in the first decade of proteomics analysis has been the implementation of MS as the primary analytical technique to identify proteins separated by two-dimensional polyacrylamide gel electrophoresis (2-D/PAGE). Additionally, the need to identify gel-separated proteins has promoted various MS-based protein database search strategies. These include peptide mass mapping using proteolytic peptide masses for database search and using peptide fragmentation data obtained from tandem mass spectrometry (MS/MS) to search protein sequence databases. Finally, we examine the more recent advances in proteomics in integrating multidimensional LC–MS/MS and the analysis of large-scale proteomics datasets.

1 INTRODUCTION

HPLC/MS is one of the most powerful tools in peptide and protein analysis today. Complementary to HPLC/MS, capillary electrophoresis/mass spectrometry (CE/MS) is another highly efficient separation technique that can be coupled directly with MS using ESI, and the technique has proven to be very sensitive.[1] The popularity of HPLC/MS has resulted from the developments in both chromatography and MS.[2–5] The advances in soft ionization, in particular MALDI[6] and ESI,[7–9] have not only broadened the range of biomolecules that can be analyzed but also provided unprecedented sensitivity. Parallel to the development of MS, the utilization of capillary high-performance liquid chromatography (HPLC) column technology in conjunction with electrospray ionization mass spectrometry (ESIMS) has further extended the detection limits. HPLC/MS has been applied in many areas of peptide and protein study. For peptide analysis, this method has proved to be useful from the routine screening of synthetic peptide products to the study of naturally occurring peptides at the femtomole level. One of the major contributions HPLC/MS had made in protein analysis is in the characterization of protein posttranslational modifications. Notable examples are the analysis of phosphoproteins and glycoproteins, primarily by using the technique of monitoring diagnostic marker ions. As a result of genome projects in the 1990s, there is an increasing need to develop analytical tools for proteomics research. The combination of database search, multidimensional HPLC, and MS/MS has become an ideal format for large-scale high-throughput automatic identification of proteomics analysis.

2 INSTRUMENTATION OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY FOR PEPTIDE AND PROTEIN ANALYSIS

2.1 Early Development of High-performance Liquid Chromatography/Mass Spectrometry

The coupling of LC and MS combines the advantages of two powerful techniques of separation and detection. The resulting liquid chromatography/mass spectrometry (LC/MS) method provides one of the most popular analytical tools for peptide and protein study today. Since the 1970s, the development of the LC/MS has focused on resolving the fundamental problems of the coupling of the two techniques. The major technical issues that need to be addressed are the compatibility of the flow rate from the LC system to the vacuum system of the mass spectrometer, the compatibility of the mobile phase of the LC separation and the operation of MS, and the ionization of the analyte molecules.

In the 1980s, LC/MS systems that reached maturity and were made commercially available included moving-belt interface,[10] thermospray,[11,12] particle beam,[13] and continuous-flow FAB.[14] The moving belt was one of the first commercial LC/MS interfaces available in the 1970s. In a moving-belt interface, the flow from an LC column is deposited on a moving belt and the solvent is evaporated by heating. The analyte is desorbed from the belt and introduced into the mass spectrometer for mass analysis. The introduction of thermospray interface in 1983 by Vestal[15] was a major advancement in the development of LC/MS. Adding heat to the ionization source, thermospray is produced when column effluent passes through the capillary vaporizer, the liquid becomes partially vaporized by the heat, resulting in a supersonic jet of vapor with particles or droplets entrained. This technique was used for LC/MS studies on peptides and proteins.[15] In a particle beam LC/MS interface, the column effluent is nebulized, either pneumatically or by thermospray, into a desolvation chamber, which is connected to a momentum separator to transfer the analyte into an electron impact/chemical ionization (EI/CI) ion source for mass analysis. The continuous-flow FAB interface adapted for LC/MS uses FAB ionization[16] and an interface connects a small liquid flow of 5–15 $\mu L\,min^{-1}$ consisting of FAB matrix such as glycerol to the FAB target.[14] Ions are formed by bombarding the liquid film formed on the target with fast atoms or ions. Although these interfaces provide ways to perform LC/MS, the capability for protein and peptide applications is still limited owing to the difficulty of producing molecular ions directly from the liquid phase.

2.2 Electrospray Ionization

Mass spectrometric measurement of the m/z values of proteins and peptides requires ionization of these polar and charged molecules in the gas phase. A major breakthrough in biological MS has been the development of ESI for biomolecules.[7–9] At present, the majority of LC/MS systems for peptide and protein analysis are based on this ionization technique. The ESI source had been investigated by Dole in the 1960s while attempting to produce charged macromolecules by electrospraying a solution of polystyrene. Experimentally, the electrospray is produced by applying an electrical potential of several kilovolts to a sample liquid flowing from a metal capillary.[17] As effluent passes through the needle tip, electrical charge is deposited on the surface of the liquid emerging from the capillary, and Coulomb repulsion forces are produced, which break down the liquid into small charged droplets. The charged droplets shrink by solvent evaporation, and the charge migrates to the surface of the droplets. When the repulsive forces exceed the surface tension, the droplet disintegrates, producing analyte ions in the gas phase for mass analysis. Since electrospray (optionally) operates without heat, nonvolatile thermal labile and polar molecules can be ionized without degradation. Furthermore, ESI is uniquely suited for LC/MS because it can couple directly to HPLC effluent. The practical importance is that the ESI spectra of the peptide and protein ions display a characteristic multiple charged ion series $[M + nH]^{n+}$, which give rise to a series of consecutive peaks of m/z $(M + n)/n$; molecular mass can be determined from any two consecutive peaks differing by one proton in the series. In general, the maximum number of protons attached to proteins or peptides correlates with the total number of basic residues (e.g. Arg, Lys, His), plus the N-terminal amino group. In practice, a distribution of many ions is seen, each differing by one charge from its neighbor and corresponding to a variable number of protons added to the protein. A protein of 10 000 Da with 10 added protons will yield an m/z 1000 protein ion in ESI, m/z 1112 with 9 added protons, m/z 1251 with 8, and so on. As a result, measurement of m/z values of biomolecules with high mass can be made in the mass range m/z 500–2000, which allows common mass analyzers with limited mass range, such as a simple quadrupole instrument, to study large proteins.

2.3 Instrumentation for High-performance Liquid Chromatography/Mass Spectrometry

Built on the early research by Dole, Fenn et al. presented a practical implementation of the ESI for LC/MS in 1985.[18] In their design (Figure 1), a sample solution passes through a 100-μm-i.d. stainless steel hypodermic needle at a flow rate of 5–20 μL min^{-1}.

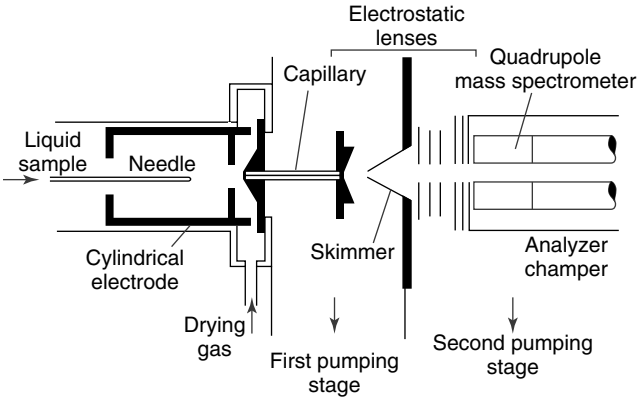

Figure 1 An electrospray interface design for LC/MS by Whitehouse et al.[18] (Reproduced by permission of the American Chemical Society from C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, *Anal. Chem.*, **57**, 675–679 (1985). Copyright 1985 American Chemical Society.)

The electrospray needle is held at ground potential and the cylindrical electrode is held at −3.5 kV (for positive-ion detection). The liquid droplets emerging from the tip of the electrospraying needle are dispersed by a countercurrent of heated nitrogen gas. The solvent that evaporates from the droplets is swept away by the bath gas, and the analyte ions enter the mass analyzer through a glass capillary. In a later design described by Chait et al.,[19] the transport and desolvation of analyte ions are carried out with a heated stainless steel transfer capillary. Pneumatic nebulization, an ion-spray LC/MS interface that combines the ion evaporation principle suggested by Iribarne and Thomson, with ESI was demonstrated by Bruins et al. in 1987.[20–23] In this design, the effluent from the liquid chromatograph is fed through a pneumatic nebulizer floating at several kilovolts relative to the sampling orifice (OR) of the ion source, liquid is dispersed into charged droplets in dry nitrogen gas at atmospheric pressure, and the nebulizing action is assisted by the coaxial high-velocity gas flow. The main advantage of the pneumatic-assisted ion-spray interface is its ability to handle higher flow rates of up to 200 μL min^{-1}. In a coaxial electrospray interface designed for CE/MS, Smith et al. have developed a liquid sheath electrospray source.[24] The electrospray needle of this interface consisted of three coaxial capillaries, with the two outer capillaries made of stainless steel and the inner capillary made of fused silica. The organic sheath liquid flow, usually consisting of pure methanol or acetonitrile, sometimes with a small amount of acetic acid, is introduced from the annular space of the outer capillaries.

Various commercial LC/MS sources have been reported.[25,26] Much of the progress made in the source

design focuses on the central issue of optimizing the nebulization process in ESI. For example, a curtain gas that opposes the path of the droplets to sweep away larger undesolvated droplets has been incorporated into Perkin-Elmer SCIEX's quadrupole instruments. In a different approach, increased vaporization was achieved by delivering heat into the nebulized plume. A heated electrospray source has been designed by Vestec, without the countercurrent or nitrogen curtain gas. In a design different from the conventional in-axial (relative to the sampling OR) electrospray source, the Hewlett-Packard/Bruker orthogonal electrospray instrument ESQUIRE-LC™ ion trap uses an electrospray ion source that is orthogonal to the sampling OR (Figure 2).[27]

The desolvation, transport, and declustering of analyte ions are achieved by the combined use of heated nitrogen gas and a dielectric glass capillary sampling device. The orthogonal geometry offers the advantage of reducing the contamination of the sampling OR and enhances the signal-to-noise ratio by minimizing the sampling charged droplets. An orthogonal sampling adapter used in the Finnigan's LCQ Deca™ ion trap electrospray instrument was reported to be very robust and capable of handling nonvolatile buffers. In a different approach, the electrospray source called a *pepperpot* was made by Micromass for their quadrupole instrument. This device and the modified cross-flow version allow sampling of the ions from the electrospray plume with desolvation and trapping of the nonvolatile materials. In the latest version of this device called *Z-spray*, analyte ions are extracted orthogonally into the mass analyzer sampling OR.[5]

Almost all the major types of mass analyzers have been coupled to HPLC as detectors. In general, mass spectrometers are used to simply measure the intact molecular weight of peptides or proteins; in these cases, single-stage mass spectrometers (MS) are used.

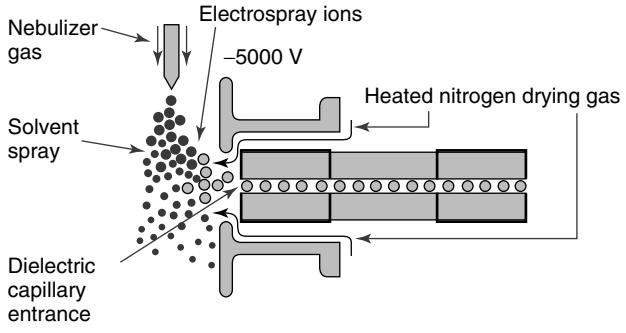

Figure 2 An orthogonal ESI source interface. Desolvation, transport, and declustering of analyte ions are optimized by using heated nitrogen drying gas and a glass capillary.[27] (Reproduced by permission of Hewlett-Packard from Hewlett-Packard Technical Note MS/MS TN97-1.)

To determine additional structural features of peptides and proteins, including amino acid sequence information, multistage mass analyzers capable of performing MS/MS are commonly used. These mass analyzers may differ in their designs and operational principles; the more widely used mass analyzers for LC/MS applications belong to the four basic types: ion trap, time of flight (TOF), quadrupole, and Fourier transform mass spectrometry (FT-MS). These mass analyzers can also be put together in tandem to combine the strength of individual analyzers. Notable examples are triple quadrupole (Q–Q–Q), hybrid quadrupole time of flight (Q-TOF), tandem time of flight (TOF–TOF), and linear ion-trap-FT/ICR (LIT-FT/ICR). Very recently, a new type of mass analyzer called *orbitrap* has emerged as one of the more powerful tools that has the characteristics of an FT/MS spectrometer but does not need a superconducting magnet.[9] Compared to the previous generation of mass analyzers, these mass spectrometers offer improved mass accuracy, resolving power, sensitivity, and MS/MS capabilities to meet the evolving analytical demands.

MS/MS is an important tool in LC/MS experiments. The four most common MS/MS experiments are (i) product ion scanning, (ii) precursor ion scanning, (iii) neutral-loss scanning, and (iv) multiple ion monitoring. The selection of a specific type of MS/MS operation depends on the information desired. The product ion scan selects a precursor with the first stage of a mass analyzer, fragments it by collision-induced dissociation (CID), and scans the second-stage mass analyzer to obtain a product mass spectrum. Product ion scanning mode is frequently used for determining the peptide sequence. In precursor ion scan, a specific product ion is selected with the second mass analyzer, and the first mass analyzer scans over a specific mass range, selecting precursor ions of different m/z that fragment to yield the specific product ion. This mode is used to screen a class of compounds that fragment to a common substructure. In a neutral-loss scan, both mass analyzers scan with a constant difference in mass. The resulting neutral-loss spectrum produces product ions that arise from the loss of a neutral fragment from the precursor ion. Precursor ion scanning and neutral-loss scanning are often used in detecting posttranslational modifications such as phosphorylation and glycosylation, using specific marker ions produced in the fragmentation process. The multiple ion monitoring is commonly used for quantitative analysis, where a specific precursor–product ion pair is monitored for a particular molecule of interest. In addition to the conventional MS/MS techniques for proteins and peptides, mass spectrometric-based methods such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) offer alternatives for peptide fragmentation; unlike CID, however, these techniques

tend to fragment in a more evenly distributed manner along the peptide backbone, and they are useful in identifying posttranslational modifications.

2.4 Capillary High-performance Liquid Chromatography/Mass Spectrometry

The detection limits for peptides and proteins by MS have dropped significantly from approximately 1 nmol in the 1980s to the low picomole to the femtomole range in the 1990s. Parallel to this development, the trend of using smaller diameter columns become clear. The applications have moved from analytical columns (4.6 mm i.d.) to microbore (1 mm i.d.), and more recently to capillary columns (<0.5 mm i.d.). The basic chromatography theory predicts that reducing the column diameter will increase sensitivity in a relationship, which is inversely proportional to the square of the column diameter.[28] This relationship is summarized in Table 1.

One of the main technical issues in HPLC and MS interfacing is what HPLC flow-rate ESI can accommodate. For effective nebulization, the optimum operation of ESI requires a stable spray of small charged droplets produced by low flow-rate conditions, typically of the order of 1–10 μL min^{-1}. Although initially the interest in developing capillary columns was to improve separation efficiency, the decrease in flow rate with reduced column inner diameter is ideal for HPLC/MS. As shown in Table 1, the flow rate of a packed capillary is decreased to about 1–15 μL min^{-1}. As the column inner diameter decreases, the injection volume and loading capacity also decrease, leading to improved detection limits (Table 1). In peptide and protein analysis, the characteristics of capillary columns match the ever-increasing demand for techniques to analyze the small amount of materials in a minute sample size (e.g. proteins isolated by 2-D/PAGE). Generally, for low-sensitivity analysis (>1 nmol), 4.6 or 2.1-mm-i.d. columns are sufficient. For analysis at the 1–100-pmol level, 1-mm-i.d. microbore columns should be appropriate. Capillary HPLC columns of <500 μm i.d. would be preferred for analysis of proteins or peptides at the <1 pMol level. Figure 3 shows a typical on-line microbore HPLC/MS system for the separation and detection of proteolytic digests of proteins.[28] A small fraction of the sample eluted from the column is directed into the electrospray ion source.

Although there are still challenges in capillary HPLC/MS, such as reproducibility between capillary columns, clogging associated with capillary columns, and the requirement for HPLC pumps capable of submicroliter per minute flow rates, recent progress in this field has shown great promise for peptide and protein analysis.[29,33–39] Recently, HPLC systems capable of delivering gradients in the 20–1000-nL min^{-1} flow rate without splitting are commercially available (e.g. Eksigent, Dionex/LC Packing, and Agilent). These systems are especially critical in proteomic experiments that perform large-scale separation of peptides for analysis by a mass spectrometer.

There have been attempts to improve the ESI detection limit. The first is to use a low flow-rate infusion technique to increase the length of mass spectrometric sampling time.[40–43] By reducing the diameter of the electrospray needle, flow rates in the nanoliters per minute range (''nanospray'') have been used to achieve low-femtomole levels of detection of peptide and protein samples.[43] In a conventional electrospray source, high voltage is used to generate aerosols of charged liquid droplets utilizing flow rates from 10 μL min^{-1} to 1 mL min^{-1}; the formation of the aerosol is assisted by pneumatic nebulization to produce a stable spray. When the flow rate of the source is reduced to nanoliters per minute, commonly referred to as *nanospray*, droplets form more readily without the assistance of sheath gas. The ionization efficiency of the nanospray is improved as the flow rate is lowered and thus less volume of the mobile phase passes through the source OR, producing smaller aerosol droplets.

A microfabricated device coupled to ESI is another promising technique that has subfemtomole per microliter detection limits.[44,45] Another approach employs capillary HPLC separation methods in conjunction with microelectrospray ionization.[30–32,46–56] This approach is capable of on-line peptide separation, which has the advantages of sample purification from contaminants and sample concentration. Research in the design of a smaller microelectrospray needle interface (tip dimension in the range of 1–25 μm), used in combination with capillary HPLC, has shown a significant improvement in sensitivity.[32,55,56] Figure 3(a) shows a

Table 1 Comparison of HPLC columns and performance[29]

Type	Column inner diameter	Flow rate	Loading	Sensitivity enhanced (-fold)
Analytical	4.6 mm	0.5–2 mL min^{-1}	2–10 mg	1
Narrow-bore	2.1 mm	200–400 μL min^{-1}	0.2–2 mg	5
Microbore	0.8–1.0 mm	25–60 μL min^{-1}	50–500 μg	20–25
Capillary	100–500 μm	1–15 μL min^{-1}	1–50 μg	80–2000
Nano	<100 μm	<1 μL min^{-1}	<1 μg	2000–10 000

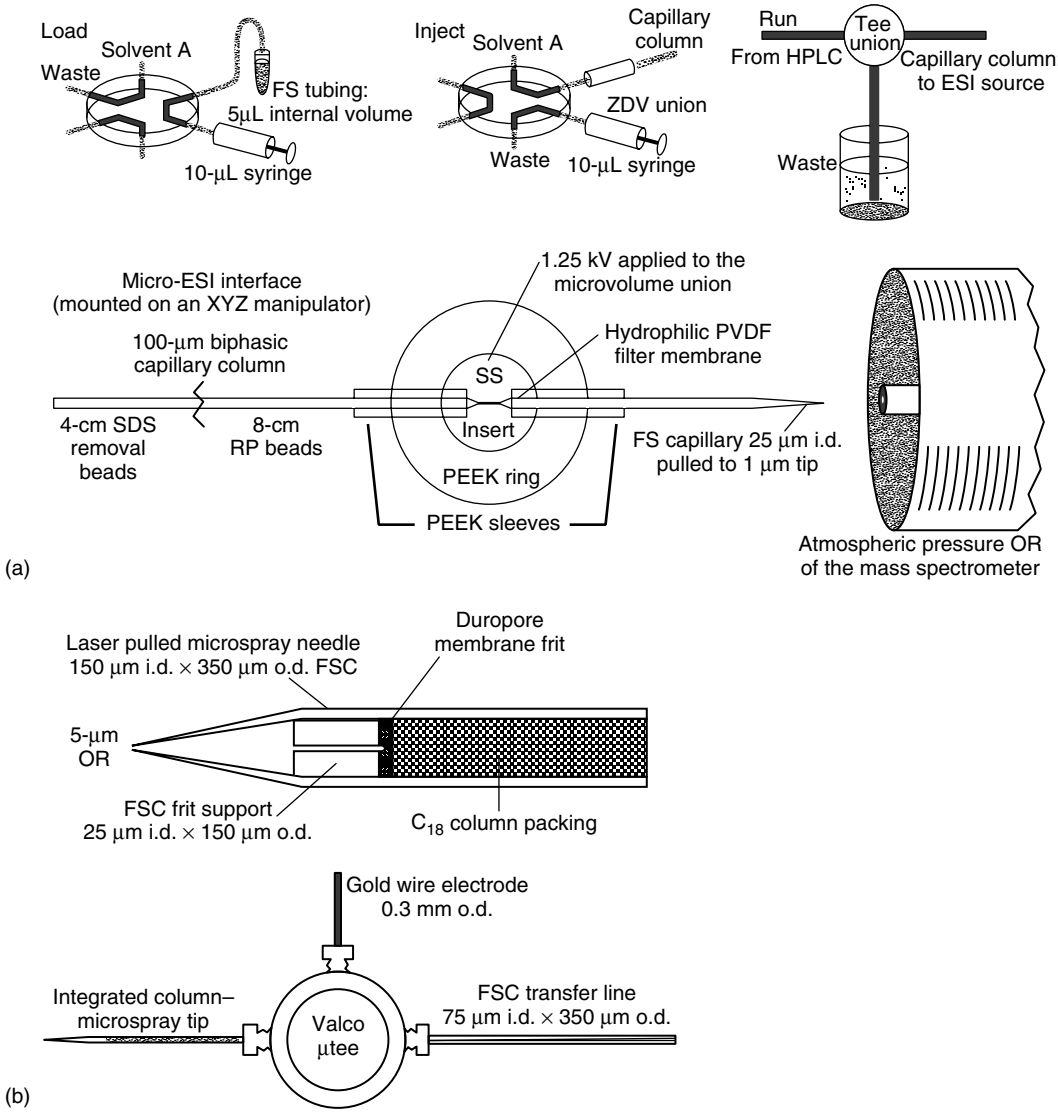

Figure 3 Capillary HPLC/MS. (a) Sample is loaded by a modified injection valve onto a biphasic capillary HPLC column (100 μm × 12 cm) with sodium dodecyl sulfate (SDS) removal and reversed-phase packing; the capillary column is coupled to a microelectrospray ionization source.[30,31] zero- dead-volume (ZVD); polyvinylidene fluoride (PVDF); polyetheretherketone (PEEK). (Reproduced by permission of Elsevier Science from M.T. Davis, T.D. Lee, *J. Am. Soc. Mass Spectrom.*, **9**, 194–201 (1998). Copyright 1998 by American Society of Mass Spectrometry.)

capillary HPLC/MS interface for proteome analysis.[55,56] The capillary HPLC column was packed with sodium dodecyl sulfate (SDS) removal and POROS (Applied-Biosystems perfusion chromatography) reversed-phase medium. Samples were loaded from an injection valve as shown. A flow rate of 500 nL min^{-1} was maintained by splitting the 150 μL min^{-1} from the HPLC pump, and the effluent was delivered into a microelectrospray needle. Figure 3(b) shows a detailed view of the microelectrospray needle designed by Davis and Lee.[32] The spray needle is pulled from a fused-silica capillary (FSC). The solvent gradient is delivered to the tee using an FSC transfer line, while a 0.3-mm-gold wire is connected to the electrospray potential. This technique can be used for routine analysis of peptide and protein sample amounts ranging from 50 to 500 fmol. In recent years, multidimensional chromatography based on capillary LC column has been used routinely to analyze highly complex peptide samples, such as samples from large-scale proteomic analysis.[30,31] Figure 4 shows the multidimensional protein identification technology (MudPIT). This technique has been developed to separate peptide by microcapillary biphasic LC column that is packed with strong cation exchange (SCX) and C_{18} reversed-phase (RP) resin, which is coupled to MS/MS detection (see Figure 4a). In a typical MudPIT

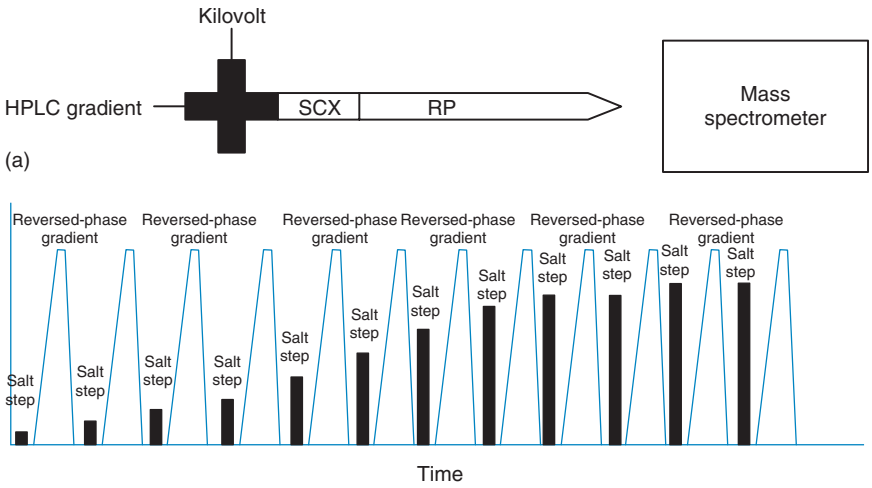

(b)

Salt step	1	2	3	4	5	6	7	8	9	10	11	12
Percentage	10	20	30	40	50	60	70	80	90	90	100	100

Figure 4 (a) An integrated MudPIT two-dimensional (SCX/RP) liquid chromatography column. A typical microcapillary biphasic column is a 75–100-μm-i.d. fused-silica capillary column with 4-cm SCX resin and 10 cm of C_{18} resin. (b) Gradient setting for the two-dimensional chromatography and an example of SCX concentration of 12-step MudPIT (concentration is in percent of salt).

experiment, the complex protein mixture is first digested and loaded onto the 75–100-μm-i.d. fused-silica microcapillary columns with tip pulled to a 5 μm diameter. By using a series of salt steps with increasing salt concentration (e.g. ammonium acetate), peptides can be eluted from the SCX resin onto the RP resin and then eluted off the RP phase typically using an increasing gradient of organic mobile phase (Figure 4b); the number of salt steps used will depend on the complexity of the sample. The strength of the MudPIT separation is the orthogonality of the two chromatographic phases – the SCX phase separates peptides according to their charge and the RP phase according to their hydrophobicity.

2.5 Experimental Parameters

n-Alkylsilica-based reversed-phased high-performance liquid chromatography (RPHPLC) is the most common medium for peptide and protein analysis.[57] Typically, peptide and protein samples are eluted with solvent gradients with an increasing concentration of organic component such as acetonitrile. Optimum separations are usually achieved by changing the slope of the gradient, temperature, ionic modifier, or solvent composition. The packing materials used in RPHPLC are porous silica based and are chemically modified by alkyl chains of variable length. The nature of these hydrophobic ligands is responsible for the chromatographic characteristics of the column. The most common materials used for proteins and peptides are C_{18} (*n*-octadecyl) and C_4 (*n*-butyl). The retention of the proteins and peptides on the reversed-phase columns depends on their hydrophobicity. The shorter chain packing works better for more hydrophobic samples (e.g. proteins), while the longer chain packing C_{18} columns are often used for peptide separation (e.g. peptide mapping). The geometry of the packing material is an important factor for separation. The pore size of the RPHPLC packing material for peptide analysis is usually 100–300 Å and for protein analysis 300–4000 Å. The most common particle diameter in RPHPLC is 3–5 μm. The conditions for a linear solvent gradient are usually an increase from 5% up to 50–100% organic over 20–120 min, using aqueous acetonitrile solution with 0.1% trifluoroacetic acid (TFA) as the mobile phase. Acetonitrile is widely used because of its high volatility and ultraviolet (UV) transparency. However, an alcohol such as 2-propanol is employed for more hydrophobic peptides or larger proteins. At low pH associated with RPHPLC, basic residues and the amino terminus exist as cations, so anionic ion-pairing reagents such as TFA can interact with proteins and peptides through a complex with the positively charged groups to enhance separation. However, volatile strong acids such as TFA can cause signal suppression and reduce spray stability in ESI; this is attributed to the high conductivity and increased surface tension.[58,59] TFA can change the acid–base equilibria in the solution, which makes attracting H^+ more difficult in the ESI process. The effect of TFA on the HPLC/MS

analysis of peptides has been studied.[58–60] To improve the quality of electrospray, the surface tension of the effluent can be reduced by mixing the sample flow with a sheath flow of organic solution such as methanol or 2-propanol.[39] A different mobile phase comprising formic acid in a 5 : 2 mixture of EtOH and PrOH for capillary HPLC/MS has been suggested.[61] To optimize the HPLC/MS performance in the presence of TFA, 10–50-fold signal enhancements have been reported for peptide analysis by adding 75% (v/v) propionic acid–25% (v/v) 2-propanol to the effluent post column, before the ESI process.[62] In addition, non-TFA-requiring columns are available commercially.

3 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY OF PEPTIDES

3.1 Analysis of Peptides by Mass Spectrometry

The sequence information of the peptide is derived from the fragment ions produced in the gas phase. Experimentally, this is achieved in two stages of mass analysis, that is, MS/MS.[63–66] The first stage of mass analysis selects the singly charged protonated peptide ion $[M + H]^+$ of interest. This is followed by CID and the fragment ions produced are analyzed in the second stage of mass analysis. Under favorable conditions, peptide sequence can be interpreted from the tandem mass spectra. At present, the most common mass analyzers for MS/MS include four-sector instruments (mainly for fast atom bombardment mass spectrometry (FABMS)), TOF instruments capable of postsource decay (PSD), Q–Q–Q instruments, and ion trap instruments. The main difference between the four-sector and the other configurations is that four-sector geometry is capable of high-energy collisions (5–10 keV), which can generate extensive side-chain fragmentation, whereas the other modes can only perform low-energy collisions (<100 eV). Figure 5 shows the nomenclature for the fragmentation of a protonated peptide ion and the mechanism for fragment ions formation.[67–70] The a_n, b_n, and c_n ions are fragment ions with charge retained on the N-terminal portion and the x_n, y_n, and z_n ions are fragment ions with charge retained on the C-terminal portion. For example, if the most basic site in the peptide ion is the amide group, in an MS/MS experiment the protonated amide nitrogen will undergo CID and produce fragment ions b_2 and y_2. The b_2 ion represents the cleavage of the –CO–NH bond between R_2 and R_3, with the charge retained on the N-terminal fragment.

The formation of the y_2 ion involves the transfer of a hydrogen atom from the N-terminal side, and the charge is retained on the C-terminal fragment. The peptide

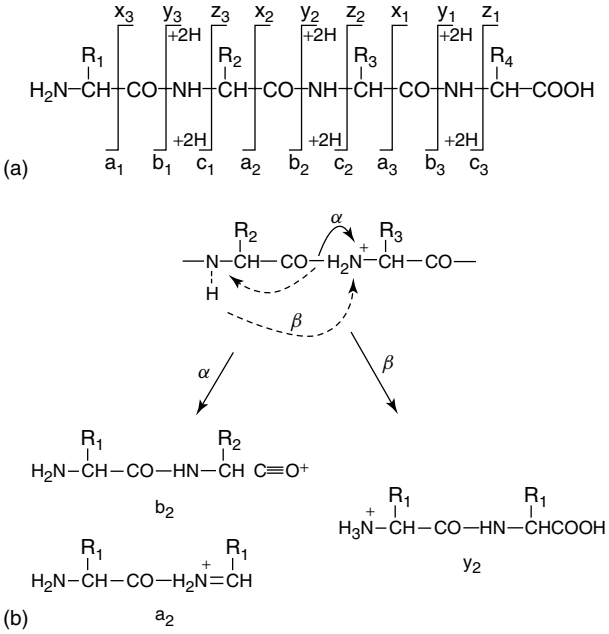

Figure 5 (a) Nomenclature of MS/MS peptide ion fragmentation by Roepstorff and Biemann.[67] (b) The formation of b_2 and y_2 ions.[64,68–70]

sequence can be derived from the same type of fragment ions in the tandem mass spectra. Generally, fragment ions a_n, b_n, and y_n form continuous series; the ions belong to the same series with the adjacent n values differing in the mass of the amino acid. The interpretation of peptide sequence involves identifying consecutive pairs of fragment ions with mass peaks that differ by the mass of the amino acid residue, or establishing combined incomplete but overlapping series that are present in the spectra.

3.2 Study of Natural and Synthetic Peptides by High-performance Liquid Chromatography/Mass Spectrometry

A wide variety of naturally occurring peptides, including toxins, neuropeptides, and antigens, have been studied by MS.[71] An elegant example of using HPLC/MS for peptides analysis is demonstrated by Hunt et al.'s studies of peptides bound to class I major histocompatibility complex (MHC).[72] One of the key events in immune response is the recognition of peptide fragments originating from degradation of foreign proteins by T lymphocytes; peptides bound to class I MHC are derived from the degradation of intracellular proteins. Determination of the sequence information of the antigenic peptide presented on the surface of the cell in association with class I MHC is an important step in designing vaccines and therapeutic treatments for diseases. Information about these peptides can be obtained from studies

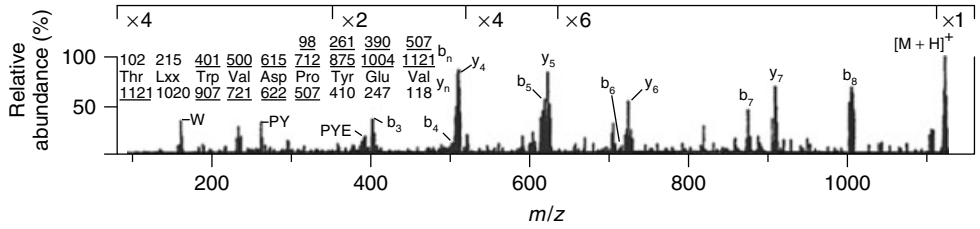

Figure 6 Tandem mass spectrum of the peptide ion $[M + H]^+$ at m/z 1121. The peptide sequence T(I/L)WVDPYEV is deduced from the b- and y-type fragment ions.[72] (Reprinted with permission from D.F. Hunt, R.A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A.J. Cox, E. Appelia, V.H. Engelhard, *Science*, **255**, 1261–1263 (1992). Copyright 1992, American Association for the Advancement of Science.)

using synthetic peptides and from Edman degradation of mixtures of peptides extracted from class I MHC. However, because HPLC is unable to resolve the complex peptide mixtures, this approach is limited to fractions containing one or two major peptides. To resolve this difficulty, Hunt et al.[72] employed capillary high-performance liquid chromatography/tandem mass spectrometry (HPLC/MS/MS) to study peptides bound to class I MHC HLA-A2.1. In these studies, a nanoscale HPLC column (75 μm × 10 cm) was used to concentrate and deliver into the ESI Q–Q–Q instrument. The detection limit of this method was 30 fmol. Figure 6 shows the tandem mass spectrum of a peptide with m/z 1121 observed in the separation. Predicted masses for type b and y fragment ions are shown; the ions observed are underlined. Any two consecutive fragment ions differing by the mass of a single amino acid yield the identity of the residue at that position. Residues 5–9 were deduced from b-type ion series, and residues 3–5 from y-type ion series.

In a study of the toxin conopeptides, which comprised highly constrained and multiple disulfide-bonded peptides, HPLC/MS coupled with differential alkylation was used to determine the disulfide connectivity.[73,74]

Another application for HPLC/MS is to evaluate synthetic peptides.[75,76] Synthetic peptides have been used as probes of biological structures, intermediates for enzyme inhibitors, and therapeutic agents, and the automation of solid-phase synthesis has made these peptides more available to users. The peptide synthesis involves the combined processes of synthesis, cleavage, purification, and the final characterization of the product. Analysis of synthetic peptides by a variety of analytical methods has been regularly reported by the Association of Biomolecular Resources Facility (ABRF, http://www.abrf.org).[77–81] The purpose of these studies is to identify optimum methods for characterizing synthetic peptides and to provide data regarding common side-chain reactions and products in peptide synthesis.

Although characterization of postsynthetic peptides can be carried out with analytical tools such as an HPLC profile or an amino acid analysis, MS provides the most direct and most efficient verification of the peptide structure. There are potential problems associated with peptide synthesis, occurring during cleavage of the peptide from the resin and removal of protecting groups from the side chains of the amino acids. Therefore, a single major peak in an HPLC profile of a crude peptide only provides a good indication of the content of the desired product and cannot prove that the structure is correct. Mass spectrometric analysis, on the other hand, allows the identification of the peptide and also the contaminant side products. For characterizing complex mixtures of peptide products that have several major HPLC peaks, a single HPLC/MS run would be very useful to corroborate the assignments of HPLC peaks as the desired peptide or side product. For structural verification of the location of a modification or the identity of a by-product, fragmentation of the peptide is needed. For example, the location of a by-product from modification of the Cys residue of a synthetic peptide has been identified by an MS/MS experiment.[76]

4 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY OF PROTEINS

4.1 General Strategy of Characterization of Proteins by Mass Spectrometry

There are three basic types of MS experiments one can perform in a protein analysis:

- mass measurement on intact protein;
- mass measurement on the proteolytic digest of the proteins (peptide mass mapping); and
- amino acid sequencing by MS/MS.

In the first category of experiments, the accurate mass of an intact protein is measured. This information alone could provide sufficient evidence to verify the correctness of a protein sequence and homogeneity of the protein sample and also to detect any chemical or posttranslation

Table 2 Common cleavage reagents for protein analysis

Cleavage reagent	Specificity
Endoproteinase Lys–C	Lys–X
Endoproteinase Arg–C	Arg–X
Endoproteinase Asp–N	X–Asp; X–cysteic acid
Asparaginylendopeptidase	Asn–X
V8 proteinase	Glu–X; Asp–X
Trypsin	Arg–X, Lys–X; X ≠ Pro
Chymotrypsin	Trp–X; Tyr–X; Phe–X; Leu–X; Met–X; X ≠ Pro
Cyanogen bromide	Met–X

modifications.[82–85] However, the problems that can be solved using intact protein molecular weights alone depend on the mass accuracy required. For example, to differentiate Asn 114.1 Da/Asp 115.1 Da or Gln 128.1 Da/Glu 129.1 Da substitutions, a mass accuracy of ±0.5 Da is needed. On the other hand, to characterize known glycosylation (e.g. an addition of 162 Da for hexose) or known phosphorylation (an addition of 80 Da), a mass accuracy of ±50 Da would be sufficient. The advancements in the FT/MS capable of extremely high mass accuracy allow applying MS to measure intact proteins and many of the posttranslational modifications. For example, a heterogeneous 42-kDa protein was found to have ragged N-terminus and alkylated Cys residue.[83,84]

The second category of experiments can achieve the same goal by determining the mass values of the peptide mixture after a proteolytic digest of the protein. Table 2 lists the common cleavage reagents used in protein analysis.

Applications include verification or correction of DNA sequence for proteins, determination of disulfide bridges, and modifications or substitutions of amino acids present in peptides by comparing the predicted molecular weights of the proteolytic peptides with those measured by MS.[63,86] This method is particularly effective when confirming peptides that are derived from proteins of known sequence, such as recombinant proteins. At present, on-line HPLC/MS analysis of protein digests has become one of the routine methods of analysis in biological MS. Finally, structural information of a protein can be obtained by MS/MS.[63,87] Conventional methods to determine protein primary structure based on MS involve enzymatic or chemical degradation. The proteolytic peptide mixture is fractionated by RPHPLC and the peptide(s) in the HPLC fractions are analyzed by MS. In CID/MS/MS, gas-phase protonated peptide molecules $[M + H]^+$ are generated and selected. These either dissociate spontaneously or dissociate following collisional activation to yield a variety of N- and C-terminal fragment ions. The peptide sequence is obtained from interpretation of the resulting tandem mass spectra. This strategy has been established as a protein sequencing tool by the elegant work of Hunt, Biemann, and others.[88–92] Although MS and Edman sequencing approach the same level of sensitivity, the former offers advantages in requiring a less demanding purification (separation into individual peptides is preferred in Edman sequencing), efficiency (each Edman sequencing cycle requires at least 20 min[93]), and the ability to identify amino acid modifications. Electrospray ionization tandem mass spectrometry (ESIMS/MS) of tryptic peptides is especially effective. Since there are two charges located on the tryptic peptides, the basic N-terminus and the C-terminus lysine or arginine residue, they generally produce a doubly charged (+2) peptide ion as the base peak.[94,95] MS/MS of the doubly charged tryptic peptide ions yields primarily singly charged fragment ions, dominated by b- and y-type series ions. The combination of enzymatic digestion and HPLC/MS/MS could provide a very powerful method to characterize protein primary structure.

4.2 Posttranslational Modification of Proteins

Proteins undergo a variety of structural changes after synthesis. As an integral part of the protein structural information, the posttranslational modifications occur immediately after translation of the polypeptide chain. It should be noted that posttranslational modifications cannot be deduced from amino acid sequence.[96] The traditional methods to detect posttranslational modification include Edman degradation, amino acid analysis, isotopic labeling, and immunochemistry. In recent years, MS and MS/MS experiments have become some of the most powerful methods for the direct identification of these modifications.[97–100] Applications include the characterization of N-terminal (e.g. formation of formyl and acetyl groups and pyroglutamic acid) and C-terminal (e.g. amidation) blocking groups, new amino acids (e.g. γ-carboxyglutamic acid, and β-hydroxyaspartic acid), disulfide bond formation, glycosylation, and phosphorylation. Since the presence of covalent modifications in proteins produce a change in molecular weight of the modified amino acid, the molecular weight of the protein or peptide that contains the modified amino acid residue, comparison of the experimentally observed intact molecular mass and the calculated mass from amino acid sequence, can reveal mass increase or decrease; the mass difference defines the specific modifications. These intact protein mass measurements can reveal heterogeneous modifications and processing of the protein, from multiple species each corresponding to an individual modification state. Furthermore, the peptide sequencing by MS/MS has the potential to reveal the presence of modification at an individual amino acid residue in proteins. Table 3

Table 3 Mass changes due to posttranslational modifications of proteins

Modification	Mass change (Da)
Pyroglutamic acid (N-terminal Gln)	−17.03
Disulfide bond formation	−2.02
C-terminal amide	−0.98
Deamidation of Asn and Gln	+0.98
Oxidation of Met or Trp	+16.00
N-terminal formylation	+28.01
N-terminal acetylation	+42.04
Carboxylation of Glu	+44.01
Phosphorylation	+79.98
Hexoses	+162.14

summarizes the mass changes due to some of the common protein modifications.

Although in some cases the ability to measure intact protein molecular weights accurately can determine protein modification directly without protein digestion, it is far more common to combine proteolytic digestion, HPLC separation, and MS to achieve this goal. Mass spectrometric strategies and techniques for the identification of modified amino acids in proteins have been reported.[101–105]

4.3 High-performance Liquid Chromatography/Mass Spectrometry Characterization of Phosphoproteins

Phosphorylation is one of the most common posttranslational modifications of protein. Protein phosphorylation takes place at specific sites and is frequently reversible. The addition of a phosphoryl group is carried out by specific protein kinases that use ATP as the phosphoryl donor, and it can be removed by phosphatases. Usually, the phosphorylation sites are hydroxyl groups of specific Ser, Thr, and Tyr residues. Short consensus linear sequences of amino acids have been identified as protein phosphorylation sites. However, because phosphorylation happens to folded proteins after protein synthesis is completed, secondary and tertiary structures are also important for determining phosphorylation sites. Therefore, any phosphorylation site predicted from consensus sequence information has to be verified experimentally.

An experimental approach based on Edman sequence analysis has been used for the determination of phosphorylation sites. However, such a method generally requires incorporation of ^{32}P and detection of ^{32}P-labeled peptides.[102] Several mass spectrometric approaches have been established for the determination of phosphorylation sites.[103–105] The phosphorylation sites have been determined by off-line HPLC and ESIMS.[105] In this approach, phosphoproteins are enzymatically cleaved, and the phosphopeptides are then separated and identified by HPLC and ESIMS. When the protein sequence is known, the phosphopeptide is identified by comparing the observed peptide molecular weight with the predicted value based on the sequence. A peptide that has a mass of 80 Da ($+HPO_3$) higher than the prediction indicates a potential phosphorylation site, and a mass higher by a multiple of 80 Da could indicate several phosphorylation sites.[106] An on-line mass spectrometric approach utilizing immobilized metal-ion chromatography coupled to capillary HPLC/MS has been reported.[107] Additionally, the potential and limitations of mapping phosphorylation sites using nanospray have also been discussed.[108]

A novel LC/MS procedure that utilizes neutral-loss scanning to identify phosphopeptides in a mixture has been described by Covey et al.[109] This technique is based on the tendency of phosphopeptides to lose the neutral phosphoric acid moiety. Such a loss indicates phosphoserine and phosphothreonine; these can eliminate β-eliminate after treating the proteolytic digest of the phosphorylated protein with a weak alkali. The phosphopeptide is identified by searching for a decrease of 98 Da (H_3PO_4), a loss of 49 Da ($m/z = 98/2 = 49$) from the doubly charged peptide, and 32.7 Da (m/z 98/3) from the triply charged peptide ion. Separate experiments are needed to scan for decreases corresponding to each charge state. Carr et al. have developed a simple strategy for HPLC/MS identification of peptides containing phosphate linked to Ser, Thr, and Tyr.[110–112] In an HPLC/MS experiment, chromatographic effluent containing a phosphopeptide experiences ''in-source'' CID, and the phosphopeptide will produce diagnostic fragment ions at m/z 63 (PO_2^-) and 79 (PO_3^-) in the negative-ion mode, under a high-collision excitation potential (high OR potential) (see Table 4).

Figure 7(a) shows the UV absorbance chromatogram and single-ion monitoring at m/z 63 + 79 in the negative-ion detection mode of the selective detection of phosphopeptides in a mixture of six peptides, including Ser-, Thr-, and Tyr-phosphorylated peptides.[110]

The effect of OR voltage on the CID of phosphopeptides performed in a quadrupole instrument is shown in Figure 7(b). Notably, no characteristic low-mass fragment

Table 4 Marker ions for protein posttranslational modifications

Carbohydrate moiety	Phosphate
Hex^+ (m/z 163)[a]	PO_2^- (m/z 63)
$HexNAc^+$ (m/z 204)	PO_3^- (m/z 79)
$NeuAc^+$ (m/z 292)	–
$HexHexNAc^+$ (m/z 366)	–

[a] The mass of sugar unit Hex is 162 Da ($C_6H_{10}O_5$) and that of carbohydrate fragment ion Hex^+ ($C_6H_{11}O_5$) is 163 Da, which is equivalent to the oxonium ion of hexose.

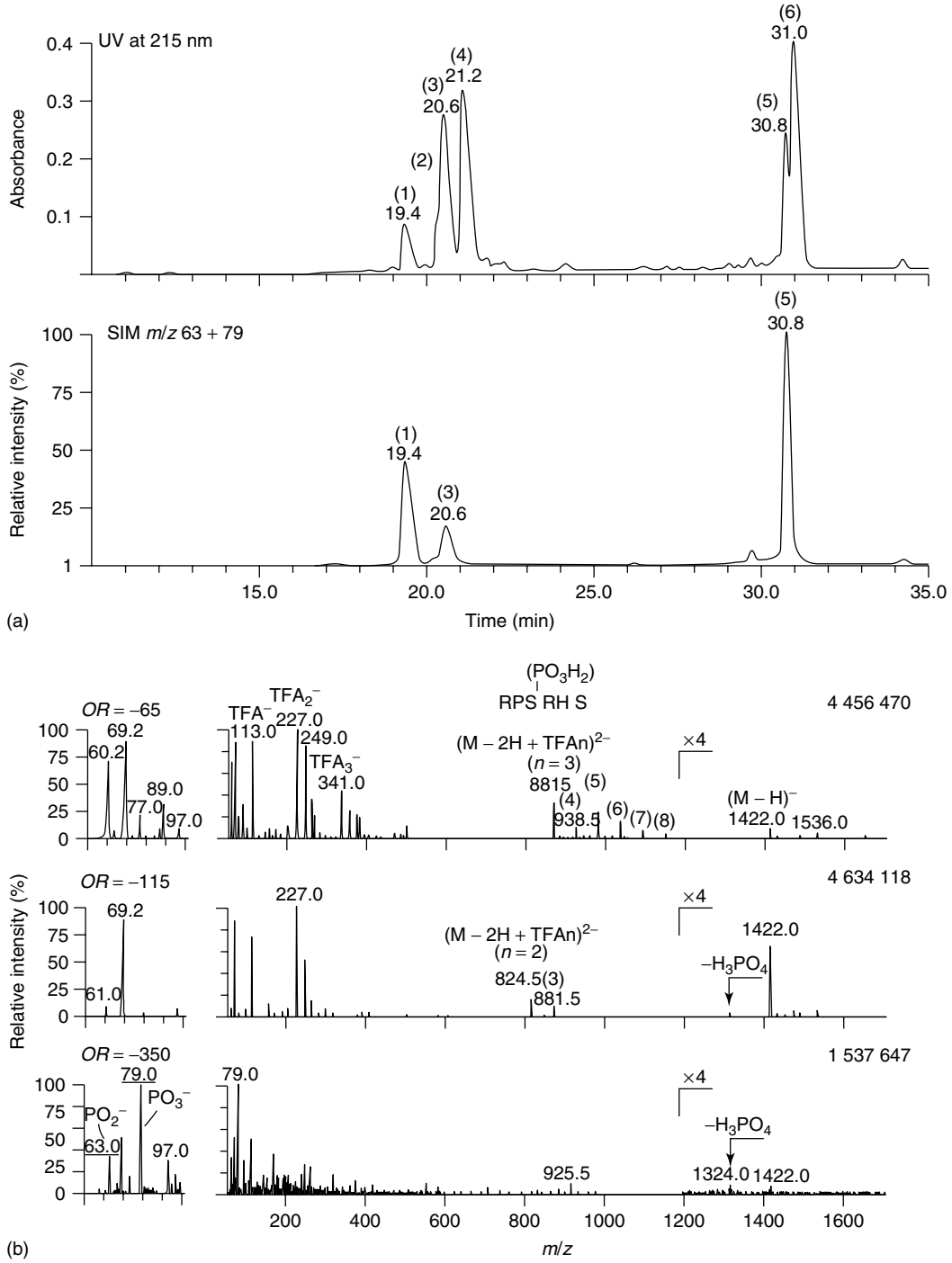

Figure 7 HPLC/MS of a peptide mixture of three phosphopeptides and nonphosphorylated analogs: (1) KRT(PO_3H_2)IRR; (2) KRTIRR; (3) KRPS(PO_3H_2)GRHGSKY; (4) KRPSGRHGSKY; (5) Ac-RRLIEDAEY(PO_3H_2)AARG-NH2; (6) Ac-RRLIEDAEY (PO_3H_2)AARG-NH_2. (a) UV absorbance chromatogram and negative-ion single-ion monitoring at m/z 63 + 79. (b) Negative-ion ESI spectra of phosphopeptide (3) recorded at three different OR voltages.[110] (Reproduced by permission of Elsevier Science from M.J. Huddleston, R.S. Annan, M.F. Bean, S.A. Carr, *J. Am. Soc. Mass Spectrom.*, **4**, 710–717 (1993). Copyright 1993 by American Society of Mass Spectrometry.)

ions occur at a low absolute OR voltage of −65 V (low collision energy). However, at a higher OR voltage of −350 V, the phosphopeptide produces the characteristic fragment ions at m/z 63 and 79. Nevertheless, final determination of phosphorylation sites requires direct sequencing. To take full advantage of the observation of

the marker ions at high OR voltage and the ability to detect the molecular phosphopeptide ion at a lower OR voltage, a full-scan LC/MS scheme can be implemented with a quadrupole instrument, scanning m/z 59–99 with OR voltage −350 V and m/z 444–2000 with OR voltage −115 V. In a multidimensional ESIMS approach, LC/MS is used to identify fractions containing phosphopeptide by single-ion monitoring at m/z 79. In the second dimension, the phosphopeptide-containing fractions are analyzed by nanoelectrospray to confirm the presence of phosphopeptide by precursor ion scan and sequence the phosphopeptide by-product ion scan.[110,113,114] It should be noted that matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS) can be used for analyzing phosphopeptides by detecting the loss of the diagnostic fragment ion. In the positive-ion mode, the presence of $[MH - HPO_3]^+$ (−80 Da) and $[MH - H_3PO_4]^+$ (−98 Da) reveal the loss of HPO_3, and consecutive loss of HPO_3 and H_2O can be used to locate phosphorylation sites. MALDI/MS can distinguish tyrosine phosphorylation from serine and threonine phosphorylation; phosphorylation of serine and threonine yield abundant $[MH - H_3PO_4]^+$ ion and weaker $[MH - HPO_3]^+$ ion; phosphotyrosine, however, shows an $[MH - HPO_3]^+$ ion and a weaker $[MH - H_3PO_4]^+$ ion.

4.4 High-performance Liquid Chromatography/Mass Spectrometry Characterization of Glycoproteins

A carbohydrate unit is attached to glycoproteins by either an N- or an O-linkage, depending on the attached amino acid residues (Asn, Thr, or Ser).[115] In the N-linked structure, the carbohydrate moiety is linked through the amide N atom of the side chain of an Asn residue. The attachment site of the Asn-linked glycopeptide is generally observed in the consensus sequence Asn–X–Ser/Thr (where X can be any amino acid except proline); however, only selective consensus sites are glycosylated. There are three types of common N-linked carbohydrate structures: complex, high mannose, and hybrid.[115] The O-linked structures are defined by the linkage through the O atom of the side-chain hydroxyl of Thr and Ser residues. No consensus sequence for their attachment sites has been reported. O-linked structures are generally less complex than the N-linked structures. At each attachment site, there is usually a heterogeneous population of different glycans, a phenomenon referred to as *site heterogeneity*. This heterogeneity of the glycoproteins creates distinct subsets (different structural classes – complex, hybrid, or high mannose – or different molecular structures within the same structural class – differences in number, type, or linkage of the sugar residues) or glycoforms that have different physical and biochemical properties.

Chromatographically, RPHPLC mapping of enzymatic peptides of glycoproteins is difficult because glycopeptides can be present in the peptide map as broad peaks due to the carbohydrate heterogeneity. The advent of HPLC/MS combined with enzymatic digestion has dramatically changed the characterization of glycoproteins.[116–120] The procedure to characterize glycoproteins by HPLC/MS is as follows: select the appropriate enzyme for cleavage, isolate the glycosylation sites into separate peptides by HPLC, and perform an MS analysis of the general nature of the glycosylation. If the primary structure of a protein is known, possible N-linked glycosylation sites can be predicted on the basis of a simple enzymatic digestion (e.g. trypsin), together with the knowledge of the N-linked consensus sequence. The main objective here is to select a degradation scheme to isolate peptides containing glycosylation sites. The separation of the resulting peptides and subsequent mass analysis can be straightforward, provided the glycopeptides do not coelute during chromatographic separation. In the analysis of large, complex glycoproteins, cleavage with trypsin may result in a very complex peptide map, in which glycopeptides can be obscured by the large number of nonglycosylated peptides. In these cases, digestion with a different proteolytic enzyme such as Lys–C or Arg–C produces larger peptide fragments, and can simplify the detection of glycopeptides. A single HPLC/MS run of the glycoprotein digest may consist of over 1000 mass scans, so to interpret and identify the glycosylation sites using HPLC/MS data requires further methodology development. An approach to locate glycopeptides in complex HPLC/MS runs has been suggested by Hancock et al.[116,117] In this method, glycoproteins are digested and then separated and detected by HPLC/MS. The glycopeptides are observed as a cluster of negatively sloping ions in a contour plot of m/z versus retention time from the HPLC/MS run.

A very specific method was developed for selectively detecting glycopeptides during the HPLC separation of glycoprotein digests by Carr et al.[120] In this method, HPLC/MS is used to locate and differentiate N- and O-linked glycopeptides, as outlined in Figure 8. The glycoprotein is cleaved by a protease, producing a mixture of peptides and glycopeptides. The digestion mixture is analyzed by three separate experiments: (i) an LC/MS experiment to locate glycopeptides, (ii) an LC/MS/MS experiment to identify specifically both N- and O-linked glycopeptides, and (iii) an experiment to differentiate N- and O-linked glycopeptides by treating the sample with PNGase F and performing LC/MS/MS experiments to identify O-linked glycopeptides. In essence, this technique is based on the observation of carbohydrate-specific diagnostic oxonium fragment ions formed by CID to locate glycopeptides in HPLC/MS (see Table 4). The

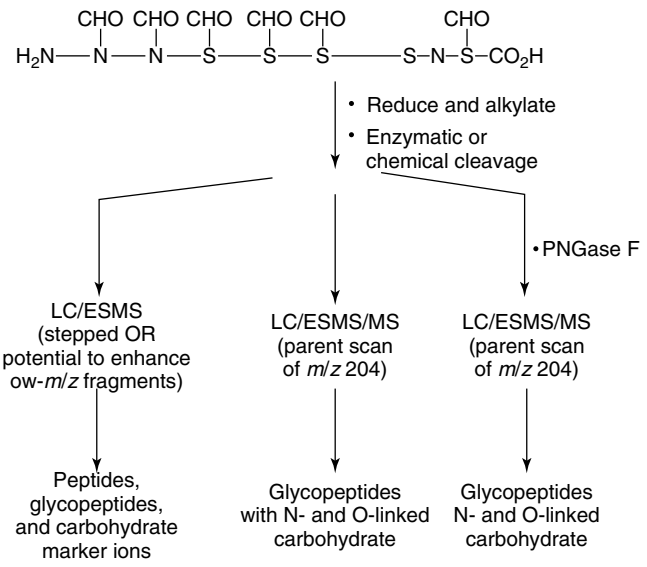

Figure 8 A strategy to locate and differentiate N- and O-linked glycopeptides by enzymatic digestion and LC/MS.[120] (Reproduced by permission of Cambridge University Press from S.A. Carr, M.J. Huddleston, M.F. Bean, *Protein Sci.*, **2**, 183–196 (1993).)

common low-mass carbohydrate oxonium ions observed are m/z 147 ($dHex^+$, fucose, or xylose), m/z 163 (Hex^+, hexose), m/z 204 ($HexNAc^+$, *N*-acetylhexosamine), m/z 292, ($NeuAc^+$, *N*-acetylneuraminic acid), and m/z 366 ($HexHexNAc^+$). The production of oxonium fragment ions takes place in the high-pressure source before the quadrupole mass analyzer (pre-Q1); this experiment requires only a single-quadrupole instrument. Alternatively, in a Q–Q–Q instrument, it can be achieved by selectively detecting only the parent ions that fragment in the Q2 to yield a $HexNAc^+$ ion at m/z 204; both N- and O-linked carbohydrates produce m/z 204. Analogous to the detection of phosphopeptides, the first experiment involves increasing the OR voltage to induce fragmentation during the m/z 150–500 mass scan to observe the carbohydrate oxonium ion fragments, and lowering the OR voltage to a setting that does not cause fragmentation of the glycopeptide ion in the m/z 500–2000 mass scan. This allows for the observation of both the marker oxonium ion fragments and the parent glycopeptides. Figure 9(a–c) shows an HPLC/MS analysis of a trypsin/Asp–N digest of bovine fetuin, which is a 42-kDa glycoprotein containing three N-linked (Asn^{81}, Asn^{138}, and Asn^{158}) and at least three O-linked (Ser^{253}, Thr^{262}, Ser^{264}) glycosylation sites. Although the location of glycopeptides in the HPLC/MS experiment can be determined by examining the total ion current (TIC) trace (Figure 9a) and the reconstructed ion current trace for m/z 204 (Figure 9b), it does not indicate which parent ions produced the marker ions.

For example, there are at least seven components in the peak eluted at about 30 min, labeled Asn^{81} in Figure 9(c). The second type of experiment can selectively detect glycopeptides by using parent ion scanning (scanning Q1) and monitoring m/z 204, produced by CID in Q2. In the parent ion scan, a specific daughter ion is selected with the second mass analyzer (Q3), and the first mass analyzer (Q1) is scanned over a specific mass range, selecting parent ions that will fragment to yield the fragment ion with the specific m/z value. A parent ion scan is usually employed to screen compounds containing a common substructure. The parent ion scan tandem mass spectrum of the chromatographic peak at 30 min reveals an Asn^{81} glycopeptide [Asp^{75}–Asn^{81}]. In the third experiment, O-linked glycopeptides are selectively detected by LC/MS/MS analysis of a trypsin/Asp-N digest after it has been treated with PNGase F. There are two enzymes commonly used to determine the sites of carbohydrate attachment by MS. By treating glycoproteins with the enzyme Endo H, the glycosidic bond between the carbohydrate and the *N*-acetylglucosamine is cleaved, and the peptide containing a glycosylation site will be 203 Da higher in mass than the corresponding unglycosylated peptide. The enzyme PNGase F, used in this study, hydrolyzes N-linked oligosaccharides. After the PNGase treatment, LC/MS/MS (parent ion scan m/z 204), the only significant remaining signals are Ser^{253}- and Thr^{262}/Ser^{264}-containing peaks (see Figure 9c), and three O-linked glycopeptides were identified.

4.5 Characterization of Recombinant Proteins by High-performance Liquid Chromatography/Mass Spectrometry

The therapeutic value of a recombinant protein is highly dependent on its structure. The protein's structural properties may vary widely according to the cell types used as the host system and the protein expression conditions.[121] In vitro or in vivo biological activities may depend on the primary structure of the recombinant protein or the extent of posttranslational modifications that are influenced by the specific host expression system. Other factors that may also have an impact on the structure of recombinant protein are cell culture conditions, method of cell harvesting, and the protein purification process. Improvement of analytical methodologies is very important in characterizing the structural details of the recombinant proteins. MS is an analytical method that can be used for the determination of the natural posttranslational modifications (e.g. glycosylation, acylation, phosphorylation, and disulfide bond formation) that are influenced by the host expression systems. It can also determine other modifications such as the carbamylation of lysine (in high concentrations

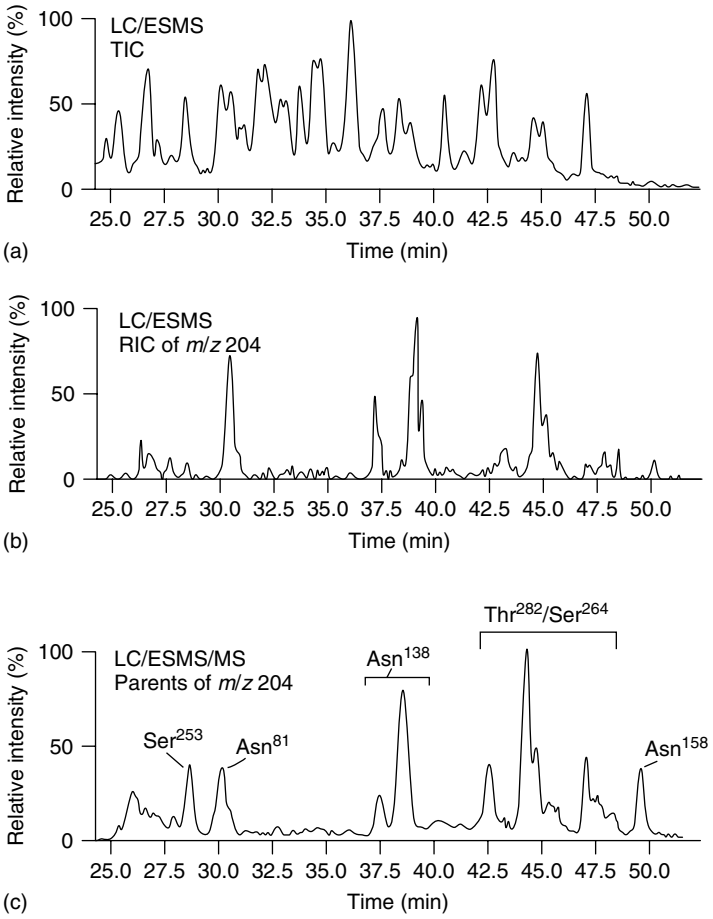

Figure 9 (a) TIC of an LC/MS/MS run of a trypsin/Asp-N digest of bovine fetuin. OR voltage ramp: 120–65 V (m/z 150–500); 65 V (m/z 500–2000). (b) Reconstructed ion chromatogram (RIC) of m/z 204 ($HexNAc^+$) from the ramped OR voltage LC/MSMS data. (c) LC/ESMS, parent ion scan of m/z 204 ($HexNAc^+$), TIC trace.[120] (Reproduced by permission of Cambridge University Press from S.A. Carr, M.J. Huddleston, M.F. Bean, *Protein Sci.*, **2**, 183–196 (1993).)

of urea), the cyclization of the N-terminal glutamine to form pyroglutamic acid, and N- and C-terminal heterogeneity. The modifications, such as deamidation of asparagine or glutamine to their acids and the oxidation of methionine or tryptophan occurring during expression or purification, can have a profound influence on the chemical and physical stability of the recombinant proteins.

Many recombinant protein pharmaceuticals are glycoproteins, including antibodies, hormones, growth factors, clotting factors, and cell-surface receptors. The oligosaccharide structure of these recombinant proteins has a direct impact on their half-lives, stability, biological responses, and resistance to proteolysis. The characterization of recombinant human erythropoietin (rEPO) is a well-documented example.[121–123] EPO is a hormone that stimulates the formation of mature erythrocytes. Off-line RPHPLC coupled with MS has shown that the rEPO from Chinese hamster ovary undergoes both N- and C-terminal processing.[121] Variation of the carbohydrate structure of rEPO has contributed to the different activities; desialylated rEPO can lose its activity completely. The glycoprotein has three N-linked (N-24, N-38, and N-83) and one O-linked (O-126) glycosylation sites. By combining peptide mapping, deglycosylation, periodate oxidation, and glycan permethylation, Linsley et al.[122] were able to study the glycan profiles of four glycopeptides from rEPO by ESIMS. Using HPLC/MS, Rush et al.[123] identified 52 different N-linked oligosaccharides. HPLC/MS has been used to study another recombinant protein, factor VIII, a glycoprotein involved in the blood coagulation cascade. In these experiments, the oligosaccharide site occupancy and site-specific microheterogeneity were determined by HPLC/MS.[124,125]

Disulfide bonds in proteins play an important role in protein folding and refolding processes and in maintaining the three-dimensional structure.[126] Although the amino acid sequence can be deduced from the DNA sequence databases, sequence information alone cannot predict the disulfide bond linkages. Characterization of the

correct formation of the disulfide bridges in recombinant proteins is critical, because an incorrect disulfide linkage may cause loss of biological activity. HPLC/MS tryptic mapping has been used to investigate the disulfide linkages in recombinant factor VIII[127] and the heparin-binding domain of the endothelial growth factor.[128] Characterization of the possible disulfide linkage was carried out by comparing the tryptic digest of the intact with the reduced and alkylated proteins. The masses of the unassigned components in the tryptic digest of the nonreduced protein were compared with the calculated molecular weights of all the possible combinations of Cys-containing tryptic peptides.

5 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY ANALYSIS IN THE "PROTEOME" ERA

5.1 Protein Analysis and Proteome

The term *proteome* was introduced in 1995 to describe a growing field that monitors cellular events by identifying the thousands of proteins that cells produce.[129] More specifically, proteome can be defined as the protein complement expressed by the genome of an organism. With the completion of genome sequencing projects, DNA sequences of at least 18 organisms are now available, and over 30 more are expected in the near future.[130,131] Large-scale proteomic projects have been initiated on a number of different organisms. Proteome projects on microbes including *Saccharomyces cervisiae*, *Haemophilus influenzae*, and *Escherichia coli* are currently in progress, and other projects for more complex organisms are also under way. The central dogma of biology dictates the path of biological information as DNA $\rightarrow$ RNA $\rightarrow$ protein. Given the recent advances in techniques such as polymerase chain reaction, cDNA microarray, and the DNA chip, it is now possible to obtain quantitative global information on mRNA expression of cells and tissues. Proteomics can provide information at the level of protein expression (which cannot be predicted from the level of expression of mRNA), and also protein posttranslational modifications. Generally, proteomic analysis represents the analysis of total protein mixture generated from a particular type cell, tissue, or multicellular organisms. The development of several technologies is key to MS-based proteome analysis.[132–134] First, the separation of proteins extracted from cell or tissue by 2-D/PAGE and identification of gel-separated proteins by MS; and secondly, the analysis of complex peptide mixtures generated by proteolytic digestion of the whole proteome sample ("shotgun proteomics") by LC and MS/MS. The latter approach has become routine in the proteomics research in the last decade.

5.2 Analyzing Gel-separated Proteins by Mass Spectrometry

The high separation power of gel electrophoresis makes it one of the most useful analytical tools for protein analysis. The size-based sodium dodecyl sulfate polyacylamide gel electrophoresis (SDS/PAGE) method developed by Laemmli is a standard tool used to separate and characterize proteins with molecular weights of up to 400 kDa. Molecular weights of proteins separated by SDS/PAGE are estimated by comparing the electrophoretic mobility of the protein being analyzed with the mobility of standard proteins with an accuracy of $\pm 10\%$. For complex proteomic protein samples, 2-D/PAGE is required. The first dimension in 2-D/PAGE is isoelectric focusing, where proteins are separated on the basis of charge. The second dimension is SDS/PAGE, which separates proteins on the basis of their molecular weights. In 2-D/PAGE, thousands of proteins are resolved in a single experiment, and the challenge is to retrieve the information. To study the protein expression pattern, sometimes gel images have to be compared through computer image analysis and database search. Databases of gel images from a number of tissues are available on the Internet (e.g. http://ca.expasy.org/ch2d/2d-index.html). Ultimately, to take full advantage of the resolution of the 2-D/PAGE for protein analysis, the identity of the protein spots on a two-dimensional gel need to be investigated. Traditionally, protein spots on the gel are visualized by Coomassie brilliant blue or silver staining. The identification of proteins is achieved by performing N-terminal or internal Edman sequencing on proteins extracted from the gels, either by electroblotting or electroeluting proteins. However, in addition to other constraints, the Edman sequencing protocol is not rapid enough to process efficiently thousands of proteins separated in a single 2-D/PAGE experiment. Another important factor is sensitivity. The amount of protein material available from a single gel spot is usually in the 1–2-μg range or lower, which sets the minimum sensitivity requirement for any analytical techniques to be feasible. Recently, the development of rapid and highly sensitive MS techniques has changed the protein identification from 2-D/PAGE completely.[135–139] Figure 10 shows a strategy for identifying proteins separated by 2-D/PAGE by integrating separation, protein digestion, MS, and database search.[138]

There are three common approaches to performing digestion on gel-separated proteins for subsequent

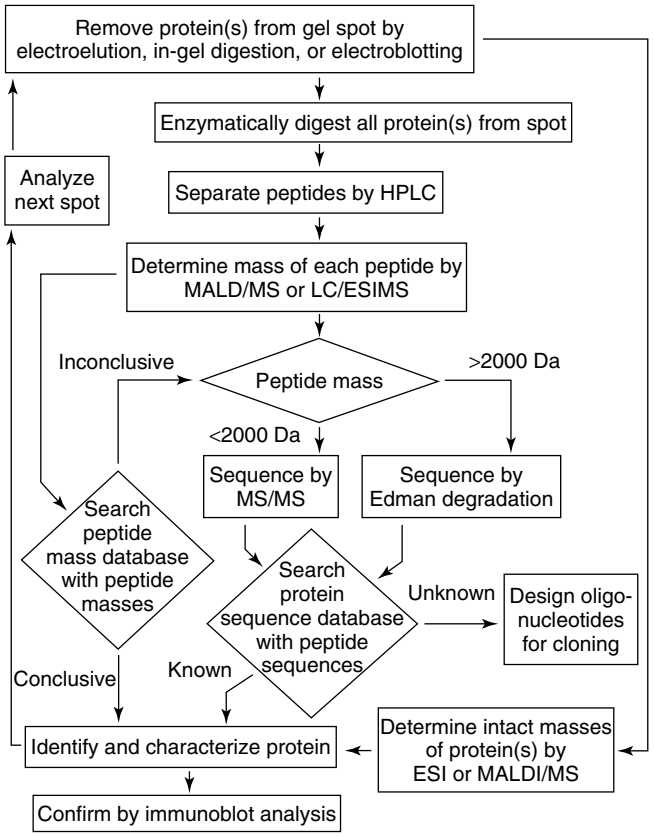

Figure 10 A mass spectrometric-based strategy for identification of 2-D/PAGE-separated proteins.[136] (Republished with permission of the Proceedings of the National Academy of Sciences USA, 2101 Constitution Ave., NW, Washington, DC 20418, from K.R. Clauser, S.C. Hall, D.M. Smith, J.W. Webb, L.E. Andrews, H.M. Tran, L.B. Epstein, A.L. Burlingame, *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 5072–5076 (1995). Reproduced by permission of the publisher via Copyright Clearance Center, Inc.)

identification by MS: (i) electroelution of protein from gels and then carrying out enzymatic digestion in solution; (ii) electrotransfer of proteins from gel to membrane and then carrying out an enzymatic digest; and (iii) direct in-gel digestion.[139,140] The uses of in-gel and on-membrane digestions in conjunction with HPLC/MS are well established.[141,142] The main consideration for HPLC/MS analysis of in-gel and on-membrane digestions is the compatibility of the chemical reagents used in the protocols. For in-gel digestion, Rosenfeld et al.[143] reported the method of washing the gels with aqueous acetonitrile–0.02% Tween 20 solution for removal of SDS in Coomassie brilliant blue-stained gels. The removal of the detergent is important for HPLC/MS analysis, since its presence is known to interfere with RPHPLC and because most detergents are ionized by ESI and undergo fragmentation, which can obscure the mass signal from peptides. Analysis of proteins transferred from gels to membranes (e.g. poly(vinylidene difluoride)) usually involves enzyme digestion of membrane-bound protein in the presence of nonionic detergents such as OGP (n-octyl-b-glucopyranoside), Triton X-100, or Tween-20, to assist peptide elution from the membrane and to block enzyme adsorption or peptide readsorption on the membrane.[144] The efficacies of organic solvents, organic bases, and detergents in dissociating electroblotted proteins from nitrocellulose have been reported by Lui et al.[145] The use of RTX-100 or PVP-360 can increase peptide recovery, but they can also cause signal suppression problems in HPLC/MS analysis.

Although a very useful technique, 2-D/PAGE has some drawbacks. It is time consuming, difficult to reproduce, and very difficult to automate. Furthermore, low abundant and membrane proteins, which represent a high percentage of cellular proteins and frequently those with important biological functions, are usually not well resolved.

5.3 Protein Analysis by High-performance Liquid Chromatography/Tandem Mass Spectrometry

Figure 11 illustrated an early example of large-scale on-line protein identification by database search that involves HPLC/MS using SEQUEST™ to perform database searches with MS/MS fragmentation data.[146,147] The SEQUEST™ program is designed to perform database searches utilizing the fragmentation pattern in a reversed manner. Specifically, the peptide fragmentation pattern acquired by the MS/MS is compared with the calculated fragmentation pattern from every peptide sequence in the database (see Figure 12) (the peptide sequences are first generated from the protein sequences in the database according to user-specified cleavage enzyme(s)). The predicted and the experimental spectra are compared on the basis of matching of peptide fragment ions, their intensities, and the presence of immonium ions. The similarity of the spectra is evaluated by a correlation function that automatically generates a list of possible proteins found, ranked according to the correlation values used by the program. The most important value to be noted is the ΔC_n score in the SEQUEST™ output. The parameter ΔC_n is an indicator of the quality of the match between the experimental and a theoretical tandem mass spectrum; the match is significant when $\Delta C_n > 0.1$. The advantage of this program is that it does not require manual interpretation of the tandem mass spectra. Mascot is a recently introduced search algorithm, evolved from the molecular weight search (MOWSE) algorithms which performs SEQUEST™-type searches but many times faster.

The usefulness of the HPLC/MS/MS method in proteome research was demonstrated by Ducret et al.[146]

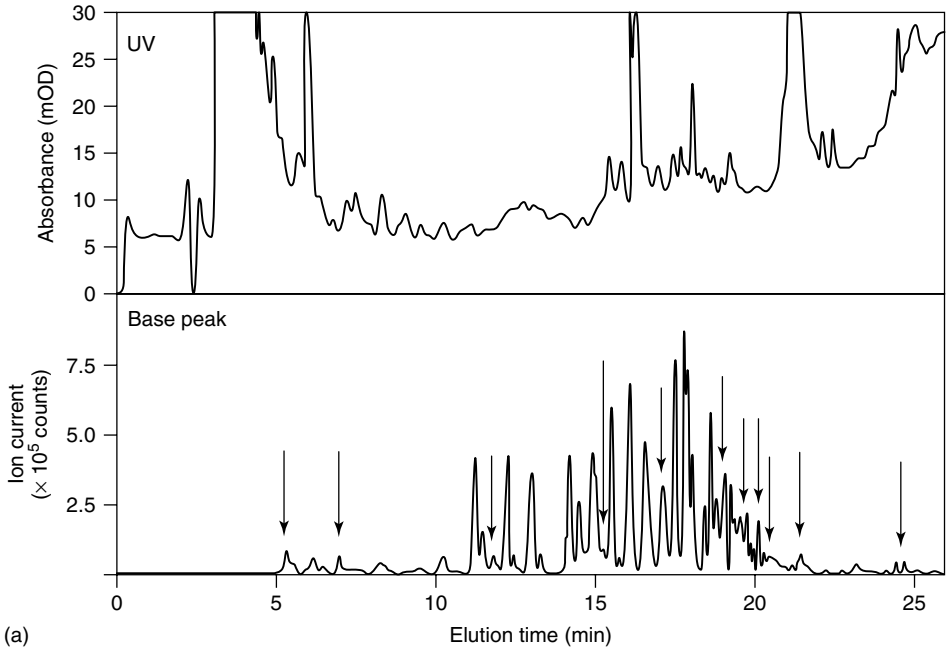

(a)

Partial SEQUEST_SUMMARY output of spot 59:

Scan #	Mass	Xcorr	ΔCoff	Reference	Sequence
199-203	1087.4	3.0055	0.300	**ADK_YEAST**	(K) ENDAILVDAK
261-265	1357.3	3.6829	0.254	K2C1_HUMAN	(K) LNDLEDALQQAK
441-445	1699.7	3.0725	0.495	**ADK_YEAST**	(K) M#AIFDELLQM#PETK
577-581	1128.4	3.4239	0.365	**ADK_YEAST**	(K) LVAGGAAQNTAR
654-658	1479.6	3.6053	0.408	**ADK_YEAST**	(K) GTSTYPVKPLDSSK
737-741	1699.2	3.4459	0.510	**ADK_YEAST**	(K) M#AIFDELLQM#PETK
765-769	2013.1	5.8912	0.563	**ADK_YEAST**	(K) IVDTNGAGDAFAGGFMAGLTK
779-783	2063.3	4.7486	0.487	**ADK_YEAST**	(K) SLVTDLGAANFFTPDHLDK
793-797	1302.9	2.5986	0.087	RAD4_SCHPO	(R) SILWEELNNGK
826-830	1666.5	4.3031	0.347	**ADK_YEAST**	(K) MAIFDELLQMPETK
946-950	2659.9	2.1392	0.063	CW41_YEAST	(K) FMTNNLEANPGLLTEYAKKIYPK

1.	**ADK_YEAST**	**80**	**(8,0,0,0,0)**
2.	RAD4_SCHPO	10	(1,0,0,0,0)
3.	CW41_YEAST	10	(1,0,0,0,0)
4.	K2C1_HUMAN	10	(1,0,0,0,0)

(b)

Figure 11 Automated HPLC/MS/MS characterization of protein of a protein spot isolated by 2-D/PAGE of yeast cell lysate. (a) UV adsorbance at 214 nm and TIC. The tandem mass spectra correlated with the sequence of the ADK_YEAST are indicated by arrows. (b) The output from SEQUEST™ program.[148] (Reproduced by permission of Cambridge University Press from A. Ducret, I.V. Oostveen, J.K. Eng, J.R. Yates III, R. Aubersold, *Protein Sci.*, **7**, 706–719 (1998).)

in a large-scale automated characterization of yeast proteins separated by 2-D/PAGE. Digests extracted from an in-gel digestion were loaded by an autosampler and subsequently injected onto a capillary HPLC system that was coupled to a mass spectrometer. Batches of 40 samples were analyzed in unattended runs during a 24-h period. The rapid scanning rate of the ion trap instrument allows the acquisition of a full-scan mass spectrum every second; it automatically determines the masses of peptide ions in an HPLC run. Target peptide ions were selected and subjected to MS/MS experiments. The resulting spectra were automatically analyzed by SEQUEST™. Figure 11(a) and (b) shows the identification of a protein spot (''spot 59'') separated from yeast cell lysate by 2-D/PAGE and followed by HPLC/MS/MS analysis.

The UV and the ion current traces are shown in Figure 11(a); the tandem mass spectra that were found to correlate with the protein identified in the database are labeled with arrows. The output from the

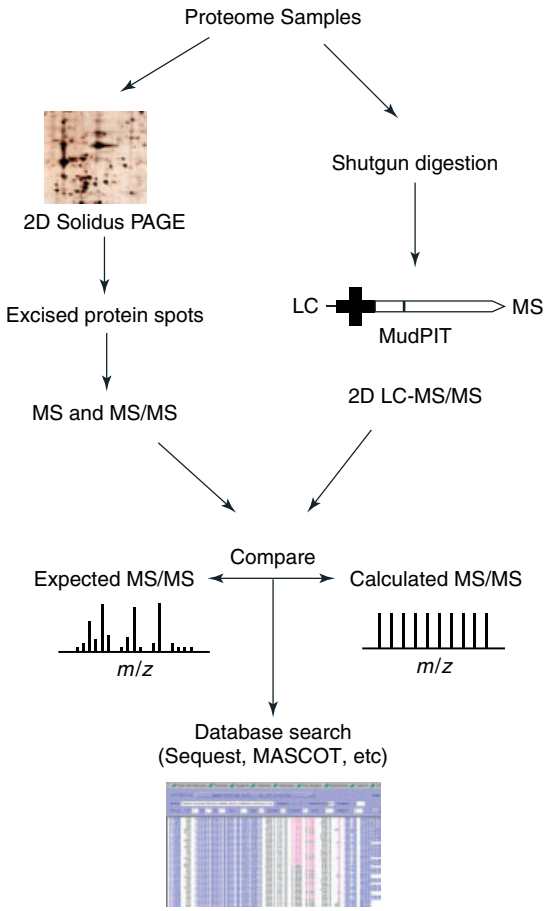

Figure 12 Overview of mass spectrometry-based proteomics.

SEQUEST™ program is shown in Figure 11(b). The spot was identified as ADK_YEAST, a putative adenosine kinase. The protein identity is established by the matches of eight proteolytic peptides with the ADK_YEAST sequence. One peptide is identified as human keratin (K2C_HUMAN), a very common contaminant. For large-scale 2-D/PAGE-separated protein identification, a database search based on tandem mass spectra has the advantage of eliminating ambiguities due to comigrating proteins, which can interfere with peptide mass mapping, and lessens the need for other search constraints such as protein molecular weight, PI (the experimental isoelectric point), or sequence tags. However, it should be noted that this strategy is limited to the identification of proteins existing in the sequence databases.

5.4 Analysis of Large-scale Proteomic Datasets Generated by High-throughput LC-MS/MS

Early on, peptide mass mapping was the most commonly used protein sequence database search method for the identification of 2-D/PAGE-separated proteins.[149–153] The method involves in-gel or on-membrane protein digestion by a specific protease (e.g. trypsin). The peptide mixture is extracted from the gel or membrane after digestion for subsequent MS analysis. The experimentally measured peptide masses are compared with all the possible peptide masses calculated from known protein sequences in the databases according to the enzyme specificity. Protein identification usually requires correct matching of observed masses from several proteolytic peptides. The computer algorithms for peptide mass mapping were developed by several laboratories in 1993.[149–152]

Usually the user defines the input parameters for the search such as taxonomic category, protein mass range, peptide mass tolerance, possible amino acid modifications (e.g. Met, Cys), and number of missed cleavages. There are several factors that will affect the results of a search using mass mapping. Obviously, the key to the success of mass mapping is the mass accuracy of the molecular weight of the peptides. A mass accuracy of ± 3 ppm or better is now achievable. Other factors that will affect the protein identification include the modifications of proteins either artificially or by posttranslational modifications. Some optional modifications such as oxidation of methionine, protection of cysteine residues, and phosphorylation have been incorporated into the programs. Finally, gel electrophoresis may induce protein modifications including cysteine–acrylamide adduct (propionamidocysteine, +71 Da), cysteine–β-mercaptoethanol adduct (+76 Da), and methionine sulfoxide (+16 Da).[154]

Rather than using only the information of peptide molecular masses, another strategy for protein identification is to use peptide fragmentation data obtained by MS/MS.[148,155–158] The information available from the tandem mass spectra includes the molecular weights of the peptides, the composition of the amino acids, and the sequence of the peptides. Mann and Wilm[155] have developed a method using peptide sequence tags (partial manual interpretation of tandem mass spectra to obtain the sequence of two or three amino acids) to search the protein sequence database. Yates et al. have developed a method that allows the search of the protein sequence database by using the uninterpreted MS/MS fragmentation spectra to search the sequence databases.[148,158] Since the latter method does not require mass spectral interpretation, it is particularly well suited for automated HPLC/MS.

Presently, the use of LC-MS/MS for analysis of complex peptide mixtures has become one of the main approaches in MS-based proteomics.[132–134] Significant progress has been made in the last decade in chromatographic separation, mass spectrometry instrumentation, and protein database search. For peptide mixtures as complex as those generated from proteomic samples (protein from a complex of 10^2–10^4 with a dynamic range of at least 10^5),

Table 5 Selective publicly available protein database search tools based on mass spectrometric data[160–162]

Program	Website
Database search tools	
SEQUEST	http://www.thermo.com
MASCOT	http://matrixscience.com
ProteinProspector	http://prospector.ucsf.edu
TANDEM	http://www.thegpm.org
SpectrumMill	http://chem.agilent.com
OMSSA	http://pubchem.ncbi.nlm.nih.gov/omssa
Spectral matching tools	
SpectraST	http://www.peptideatlas.org/spectrast
X!P3	http://p3.thegpm.org/tandem/ppp.html
De novo sequencing tools	
Lutefisk	http://www.hairyfatguy.com/lutefisk/
PEAKS	http://www.bioinformaticssolutions.com
Hybrid approaches	
Guten Tag	http://fields.scripps.edu/GutenTag
Popitam	http://expasy.org/tools/popitam
Database for storing and mining	
PeptideAtlas	http://www.peptideatlas.org
SBEAMS	http://sbeams.org

a single-dimension chromatography does not provide sufficient peak capacity for separation. To increase peak capacity, two-dimensional (strong cation/reversed phase) chromatographic separation of peptide generated by tryptic digestion of proteomic samples have been explored. With subsequent MS/MS and database search, hundreds of proteins can be identified in a single experiment. In automated 2-D LC separation, tryptic peptide mixtures are loaded onto the column and eluted stepwise with a series of salt plugs of increase ionic strength in the first dimension (SCX). The eluted peptides are then separated in the second dimension, the RP phase of the column, and then analyzed by MS/MS. In a typical experiment, thousands of MS and MS/MS peptide spectra are generated. The assignment of such a spectrum to a specific peptide sequence is a critical step in retrieving biological information from such experiments. Indeed, significant progress has been made in this area.

A number of computational tools that have been developed for automatic assignment of peptide fragment ion spectra are summarized in Table 5.[159–162] These search tools for MS/MS spectra can be broadly classified into three categories: database searching, de novo sequencing, and hybrid approach.

Sequence database searching of uninterpreted MS/MS spectra is widely used in high-throughput proteomic studies (see Figure 12); SEQUEST and MASCOT are two popular algorithms. In this approach, experimentally acquired MS/MS spectra are compared with predicted spectra from protein sequences database. The majority of the known protein sequences in these databases were obtained by translation of nucleotide sequences data into amino acid sequences. The main difference between different database search tools is the scoring scheme used to quantify the degree of similarity between the compared spectra. Alternatively, peptide sequences can be identified from MS/MS spectra by using de novo sequencing algorithms; however, it requires higher-quality MS/MS spectra and the process is computationally intensive and thus not suitable for high-throughput studies. The hybrid approaches combine the use of short partial peptide sequence tags generated from MS/MS spectra and de novo algorithms, and error-tolerant database searching.

Peptide sequence assignments for large amount of data generated in these high-throughput experiments still present significant challenges. To address some of the technical issues, there have been efforts to improve the understanding of gas-phase peptide ion fragmentation process, processing MS/MS spectra before submitting for database search, and refining scoring schemes and search algorithms to diminish false positives and false negatives. One such advancement is the development of spectral library–searching method for peptide identification from MS/MS spectra.[163] The concept of library searching – to infer the chemical structure of unknowns by comparing the observed EI spectra with a compiled spectral library from the characteristic fragmentation patterns – has been used successfully by analytical chemists for a very long time. The same approach has been used to develop the software program SpectraST, a tool that performs spectral library search on peptide MS/MS spectra against compiled spectral libraries for peptides.

ABBREVIATIONS AND ACRONYMS

ABRF	Association of Biomolecular Resources Facility
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CID	Collision-induced Dissociation
2-D/PAGE	Two-dimensional Polyacrylamide Gel Electrophoresis
ECD	Electron Capture Dissociation
EI/CI	Electron Impact/Chemical Ionization
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization Mass Spectrometry
ESIMS/MS	Electrospray Ionization Tandem Mass Spectrometry

ETD	Electron Transfer Dissociation
FAB	Fast Atom Bombardment
FABMS	Fast Atom Bombardment Mass Spectrometry
FSC	Fused-silica Capillary
FT-MS	Fourier transform mass spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
HPLC/MS/MS	High-performance Liquid Chromatography/Tandem Mass Spectrometry
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LIT-FT/ICR	Linear Ion-Trap-FT/ICR
MALDI	Matrix-assisted Laser Desorption/Ionization
MALDI/MS	Matrix-assisted Laser Desorption/Ionization Mass Spectrometry
MHC	Major Histocompatibility Complex
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MudPIT	Multidimensional Protein Identification Technology
OGP	n-Octyl-b-Glucopyranoside
OR	Orifice
PSD	Postsource Decay
Q–Q–Q	Triple Quadrupole
Q-TOF	Quadrupole Time of flight
rEPO	Recombinant Human Erythropoietin
RIC	Reconstructed Ion Chromatogram
RP	Reversed-Phase
RPHPLC	Reversed-phased High-performance Liquid Chromatography
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacylamide Gel Electrophoresis
TFA	Trifluoroacetic Acid
TIC	Total Ion Current
TOF	Time of flight
TOF–TOF	Tandem Time of flight
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Posttranslational Oxidative Modifications of Proteins • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Liquid Chromatography **(Volume 13)**
Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Fourier Transform Ion Cyclotron Resonance Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Mass Spectrometry: Overview and History • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. G.A. Valaskovic, N.L. Kelleher, F.W. McLafferty, 'Attomole Protein Characterization by Capillary Electrophoresis–Mass Spectrometry', *Science*, **273**, 1199–1202 (1996).
2. W.M.A. Niessen, A.P. Tinke, 'Liquid Chromatography Mass Spectrometry – General Principles and Instrumentation', *J. Chromatogr. A*, **703**, 37–57 (1995).
3. E. Gelpi, 'Biomedical and Biochemical Applications of Liquid Chromatography Mass Spectrometry', *J. Chromatogr. A*, **703**, 59–80 (1995).
4. W.M.A. Niessen, 'Advances in Instrumentation in Liquid Chromatography–Mass Spectrometry and Related Liquid-introduction Techniques', *J. Chromatogr. A*, **794**, 407–435 (1998).
5. W.M.A. Niessen, *Liquid Chromatography–Mass Spectrometry, Chromatographic Science Series*, Marcel Dekker, New York, Vol. 79, 1999.
6. F. Hillenkamp, M. Karas, R.C. Beavis, B. T. Chait, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**, 1193A–1203A (1991).
7. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, 'Electrospray Ionization for Mass Spectrometry of Large Biomolecules', *Science*, **246**, 64–71 (1989).

8. G. L. Glish, R. W. Vachet, 'The Basics of Mass Spectrometry in the Twenty-First Century', *Nat. Rev. Drug Discov.*, **2**, 140–150 (2003).
9. B. Domon, R. Aebersold, 'Mass Spectrometry and Protein Analysis', *Science*, **312**, 212–217 (2006).
10. P.J. Arpino, 'Combined Liquid Chromatography Mass Spectrometry. Part I. Coupling by Means of a Moving Belt Interface', *Mass Spectrom. Rev.*, **8**, 35 (1989).
11. P.J. Arpino, 'Combined Liquid Chromatography Mass Spectrometry. Part II. Techniques and Mechanisms of Thermospray', *Mass Spectrom. Rev.*, **9**, 631 (1990).
12. P.J. Arpino, 'Combined Liquid Chromatography Mass Spectrometry. Part III. Applications of Thermospray', *Mass Spectrom. Rev.*, **11**, 3 (1992).
13. A. Cappiello, 'Is Particle Beam an Up-to-date LC–MS Interface? State of the Art and Perspectives', *Mass Spectrom. Rev.*, **15**, 283–296 (1996).
14. R.M. Caprioli, M.J.-F. Suter, 'Continuous-flow Fast-atom Bombardment – Recent Advances and Applications', *Int. J. Mass Spectrom. Ion Processes*, **118**, 449–476 (1992).
15. M.L. Vestal, 'High-performance Liquid Chromatography–Mass Spectrometry', *Science*, **226**, 275–281 (1984).
16. M. Barber, R.S. Bordoli, R.D. Sedgwick, A.N. Tyler, 'Fast Atom Bombardment of solids as an Ion-source in Mass spectrometry', *Nature*, **293**, 270–275 (1981).
17. P. Kebarle, L. Tang, 'From Ions in Solution to Ions in the Gas Phase: the Mechanism of Electrospray Mass Spectrometry', *Anal. Chem.*, **65**, 972A–986A (1993).
18. C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, 'Electrospray Interface for Liquid Chromatography and Mass Spectrometers', *Anal. Chem.*, **57**, 675–679 (1985).
19. S.K. Chowdhury, V. Katta, B.T. Chait, 'An Electrospray-ionization Mass Spectrometer with New Features', *Rapid Commun. Mass Spectrom.*, **4**, 81–87 (1990).
20. A.P. Bruins, T.R. Covey, J.D. Henion, 'Ion Spray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 2642–2646 (1987).
21. E.C. Huang, T. Wachs, J.J. Conboy, J.D. Henion, 'Atmospheric Pressure Ionization Mass Spectrometry: Detection for the Separation Sciences', *Anal. Chem.*, **62**, 713A–725A (1990).
22. A.P. Bruins, 'Liquid Chromatography–Mass Spectrometry with Ion Spray and Electrospray Interfaces in Pharmaceutical and Biomedical Research', *J. Chromatogr.*, **554**, 39–46 (1991).
23. A.P. Bruins, 'Mass Spectrometry with Ion Sources Operating at Atmospheric Pressure', *Mass Spectrom. Rev.*, **10**, 53–77 (1991).
24. R.D. Smith, J.A. Loo, R.R.O. Loo, M. Busman, H.R. Udseth, 'Principles and Practice of Electrospray Ionization–Mass Spectrometry for Large Polypeptides and Proteins', *Mass Spectrom. Rev.*, **10**, 359–451 (1991).
25. L. Voress, 'Electrospray MS Charging Full Steam Ahead', *Anal. Chem.*, **66**, 481A–486A (1994).
26. C. Henry, 'Electrospray Continues to Evolve', *Anal. Chem.*, **69**, 427A–432A (1997).
27. Ion Trap MS/MS, 'Detection for High Performance Liquid Chromatography Applications', Hewlett-Packard Technical Note MS/MS TN97-1, Hewlett-Packard, Avondale, PA, 1997.
28. D. Hess, T.C. Covey, R. Winz, R.W. Brownsey, R. Aubersold, 'Analytical and Micropreparative Peptide Mapping by High Performance Liquid Chromatography/Electrospray Mass Spectrometry of Proteins Purified by Gel Electrophoresis', *Protein Sci.*, **2**, 1342–1351 (1993).
29. F.J. Yang, W. Lai, D.C. Park, C.T. Yang, 'Recent Advances and Future Prospects in Micro HPLC Columns', *LC–GC*, S34–S43 (1997).
30. M.P. Washburn, D. Wolters, J.R. Yates III, 'Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology', *Nat. Biotechnol.*, **19**, 242–247 (2001).
31. C.M. Delahunty, J.R. Yates III, 'MudPit: Multidimensional Protein Identification Technology', *BioTechniques*, **43**, 563–569 (2007).
32. R.T. Kelly, J.S. Page, R. Zhao, W.-J. Qian, H.M. Mottaz, K. Tang, R.D. Smith, 'Capillary-Based Multi Nanoelectrospray Emitters: Improvements in Ion Transmission Efficiency and Implementation with Capillary Reversed-Phase LC-ESI-MS', *Anal. Chem.*, **80**, 143–149 (2008).
33. M.V. Novotny, 'Recent Advances in Microcolumn Liquid Chromatography', *Anal. Chem.*, **60**, 500A–510A (1988).
34. R.T. Kennedy, J.W. Jorgenson, 'Preparation and Evaluation of Packed Capillary Liquid-chromatography Columns with Inner Diameters from 20-μm to 50-μm', *Anal. Chem.*, **61**, 1128–1135 (1989).
35. M.A. Moseley, L.J. Deterding, K.B. Tomer, J.W. Jorgenson, 'Nanoscale Packed-Capillary Liquid-Chromatography Coupled with Mass-spectrometry Using a Coaxial Continuous-flow Fast-atom-bombardment Interface', *Anal. Chem.*, **63**, 1467–1473 (1991).
36. E.C. Huang, J.D. Henion, 'Packed-capillary Liquid Chromatography/Ion-spray Tandem Mass Spectrometry Determination of Biomolecules', *Anal. Chem.*, **63**, 732–739 (1991).
37. R.L. Moritz, R.J. Simpson, 'Application of Capillary Reversed-phase High Performance Liquid Chromatography to High-sensitivity Protein Sequence Analysis', *J. Chromatogr.*, **599**, 119–130 (1992).
38. K.B. Tomer, M.A. Moseley, L.J. Deterding, C.E. Parker, 'Capillary Liquid Chromatography/Mass Spectrometry', *Mass Spectrom. Rev.*, **13**, 431–457 (1994).

39. J. R. Yates III, A.L. Ashley, L. McCormack, A.J. Link, D. Schieltz, J. Eng, L. Hays, 'Future Prospects for the Analysis of Complex Biological Systems Using Micro-column Liquid Chromatography–Electrospray Tandem Mass Spectrometry', *Analyst*, **121**, 65R–76R (1996).
40. D.C. Gale, R.D. Smith, 'Small-volume and Low Flow-rate Electrospray-ionization Mass-spectrometry of Aqueous Samples', *Rapid Commun. Mass Spectrom.*, **7**, 1017–1021 (1993).
41. G.A. Valaskovic, J.P. Murphy III, M.S. Lee, 'Automated Orthogonal Control System for Electrospray Ionization', *J. Am. Soc. Mass Spectrom.*, **15**, 1201–1215 (2004).
42. M.R. Emmett, R.M. Caprioli, 'Microelectrospray Mass Spectrometry: Ultra-high-sensitivity Analysis of Peptides and Proteins', *J. Am. Soc. Mass Spectrom.*, **5**, 605–613 (1994).
43. M.S. Wilm, M. Mann, 'Femtomole Sequencing of Proteins from Polyacrylamide Gels by Nano-electrospray Mass Spectrometry', *Nature*, **379**, 466–469 (1996).
44. D. Figeys, Y. Ning, R. Aebersold, 'A Microfabricated Device for Rapid Protein Identification by Microelectrospray Ion Trap Mass Spectrometry', *Anal. Chem.*, **69**, 3153–3160 (1997).
45. D. Figeys, C. Lock, L. Taylor, R. Aebersold, 'Microfabricated Device Coupled with an Electrospray Ionization Quadrupole Time-of-flight Mass Spectrometer: Protein Identifications Based on Enhanced-resolution Mass Spectrometry and Tandem Mass Spectrometry Data', *Rapid Commun. Mass Spectrom.*, **12**, 1435–1444 (1998).
46. M.T. Davis, T.D. Lee, 'Analysis of Peptide Mixtures by Capillary High Performance Liquid Chromatography: a Practical Guide to Small-scale Separations', *Protein Sci.*, **1**, 935–944 (1992).
47. M.T. Davis, D.C. Stahl, T.D. Lee, 'Low Flow High-performance Liquid Chromatography Solvent Delivery System Designed for Tandem Capillary Liquid Chromatography–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **6**, 571–577 (1995).
48. D.C. Stahl, K.M. Swiderek, M.T. Davis, T.D. Lee, 'Data-controlled Automation of Liquid Chromatography/Tandem Mass Spectrometry Analysis of Peptide Mixtures', *J. Am. Soc. Mass Spectrom.*, **7**, 532–540 (1996).
49. M.T. Davis, T.D. Lee, 'Variable Flow Liquid Chromatography–Tandem Mass Spectrometry and the Comprehensive Analysis of Complex Protein Digest Mixtures', *J. Am. Soc. Mass Spectrom.*, **8**, 1059–1069 (1997).
50. M.T. Davis, T.D. Lee, 'Rapid Protein Identification Using a Microscale Electrospray LC/MS System on an Ion Trap Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **9**, 194–201 (1998).
51. M.T. Davis, D.G. Stahl, S.A. Hefta, T.D. Lee, 'A Microscale Electrospray Interface for On-line, Capillary Liquid Chromatography/Tandem Mass Spectrometry of Complex Peptide Mixtures', *Anal. Chem.*, **67**, 4549–4556 (1995).
52. R.D. Smith, Y. Shen, K. Tang, 'Ultrasensitive and Quantitative Analyses from Combined Separations-Mass Spectrometry for the Characterization of Proteomes', *Acc. Chem. Res.*, **37**, 269–278 (2004).
53. Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, 'High-Efficiency Nanoscale Liquid Chromatography Coupled On-line with Mass Spectrometry using Nanoelectrospray Ionization for Proteomics', *Anal. Chem.*, **74**, 4235–4249 (2002).
54. Y. Shen, R.J. Moore, R. Zhao, J. Blonder, D.L. Auberry, C. Masselon, L. Pasa-Tolic, K.K. Hixson, K.J. Auberry, R.D. Smith, 'High-Efficiency On-Line Solid-Phase Extraction Coupling to 15–150-μm-i.d. Column Liquid Chromatography for Proteomic Analysis', *Anal. Chem.*, **75**, 3596–3605 (2003).
55. C.L. Gatlin, G.R. Kleemann, L.G. Hays, A.J. Link, J.R. Yates III, 'Protein Identification at the Low Femtomole Level from Silver-stained Gels Using a New Fritless Electrospray Interface for Liquid Chromatography–Microspray and Nanospray Mass Spectrometry', *Anal. Biochem.*, **263**, 93–101 (1998).
56. D. Arnott, K.L. O'Connell, K.L. King, J.T. Stults, 'An Integrated Approach to Proteome Analysis: Identification of Proteins Associated with Cardiac Hypertrophy', *Anal. Biochem.*, **258**, 1–18 (1998).
57. M.-I. Aguilar, M.T. Hearn, 'High-resolution Reversed-phase High-performance Liquid Chromatography of Peptides and Proteins', *Methods Enzymol.*, **270**, 3–26 (1996).
58. J. Eshraghi, S.K. Chowdhury, 'Factors Affecting Electrospray Ionization of Effluents Containing Trifluoroacetic Acid for High-performance Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **65**, 3528–3533 (1993).
59. J.F. Banks Jr, C.M. Whitehouse, 'Electrospray Mass Spectrometry', *Methods Enzymol.*, **270**, 486–518 (1996).
60. U.A. Mirza, B.T. Chait, 'Effects of Anions on the Positive Ion Electrospray Ionization Mass Spectra of Peptides and Proteins', *Anal. Chem.*, **66**, 2898–2904 (1994).
61. K.F. Medzihradszky, D.A. Maltby, S.C. Hall, C.A. Settineri, A.L. Burlingame, 'Characterization of Protein N-Glycosylation by Reversed-phase Microbore Liquid Chromatography/Electrospray Mass Spectrometry, Complementary Mobile Phases, and Sequential Exoglycosidase Digestion', *J. Am. Soc. Mass Spectrom.*, **5**, 350–358 (1994).
62. F.E. Kuhlmann, A. Apffel, S.M. Fischer, G. Glodberg, P.C. Goodley, 'Signal Enhancement for Gradient

Reversed-phased High-performance Liquid Chromatography–Electrospray Ionization Mass Spectrometry Analysis with Trifluoroacetic and Other Strong Acid Modifiers by Postcolumn Addition of Propionic Acid and Isopropanol', *J. Am. Soc. Mass Spectrom.*, **6**, 1221–1225 (1995).

63. K. Biuemann, 'Contribution of Mass Spectrometry to Peptide and Protein Structure', *Biomed. Environ. Mass Spectrom.*, **16**, 99–111 (1988).
64. R.S. Johnson, S.A. Martin, K. Biemann, 'Collision-induced Fragmentation of [M + H]+ Ions of Peptides. Side Chain Specific Sequence Ions', *Int. J. Mass Spectrom. Ion Processes*, **86**, 137–154 (1988).
65. S.A. Carr, M.E. Hemling, M.F. Bean, G.D. Roberts, 'Integration of Mass Spectrometry in Analytical Biotechnology', *Anal. Chem.*, **63**, 2802–2824 (1991).
66. I.A. Papayannopoulos, 'The Interpretation of Collision-induced Dissociation Tandem Mass Spectra of Peptides', *Mass Spectrom. Rev.*, **14**, 49–73 (1995).
67. K. Biemann, 'Nomenclature for Peptide Fragment Ions (Positive Ions)', *Methods Enzymol.*, **193**, 886–887 (1990).
68. K. Biemann, 'Sequencing Peptides by Tandem Mass Spectrometry and High-energy Collision-induced Dissociation', *Methods Enzymol.*, **193**, 455–479 (1990).
69. H. Steen, M. Mann, 'The ABC's (and XYZ's) of Peptide Sequencing', *Nat. Rev. Mol. Cell Biol.*, **5**, 699–711 (2004).
70. K.F. Medzihradszky, 'Peptide Sequence Analysis', *Methods Enzymol.*, **402**, 209–244 (2005).
71. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', *Anal. Chem.*, **70**, 647R–716R (1998).
72. D.F. Hunt, R.A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A.J. Cox, E. Appelia, V.H. Engelhard, 'Characterization of Peptides Bound to Class I MHC Molecule HLA-A2.1 by Mass Spectrometry', *Science*, **255**, 1261–1263 (1992).
73. A. Jones, J.-P. Bingham, J. Gehrmann, T. Bond, M. Loughnan, A. Atkins, R.J. Lewis, P.F. Alewood, 'Isolation and Characterization of Conopeptides by High-performance Liquid Chromatography Combined with Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 138–143 (1996).
74. T. Nakamura, Z. Yu, M. Fainzilber, A.L. Burlingame, 'Mass Spectrometric-based Revision of the Structure of a Cysteine-rich Peptide Toxin with -Carboxyglutamic Acid, TxVIIA, from Sea Snail, Conus Textile', *Protein Sci.*, **5**, 524–530 (1996).
75. W.T. Moore, 'Integration of Mass Spectrometry into Strategies for Peptide Synthesis', *Biol. Mass Spectrom.*, **22**, 149–162 (1993).
76. D.J. Burdick, J.T. Stults, 'Analysis of Peptide Synthesis Products by Electrospray Ionization Mass Spectrometry', *Methods Enzymol.*, **289**, 499–519 (1996).
77. A.J. Smith, J.D. Young, S.A. Carr, D.R. Marshak, L.C. Williams, K.R. Williams, 'State-of-the-art Peptide Synthesis: Comparative Characterization of a 16-mer Synthesized in 31 Different Laboratories', in *Techniques in Protein Chemistry III*, ed. R.H. Angeletti, Academic Press, New York, 219, 1992.
78. G.B. Fields, S.A. Carr, D.R. Marshak, A.J. Smith, J.T. Stults, L.C. Williams, K.R. Williams, J.D. Young, 'Evaluation of Peptide Synthesis as Practiced in 53 Laboratories', in *Techniques in Protein Chemistry IV*, ed. R.H. Angeletti, Academic Press, New York, 229, 1993.
79. G.B. Fields, R.H. Angeletti, S.A. Carr, A.J. Smith, J.T. Stults, L.C. Williams, J.D. Yong, 'Variable Success of Peptide–resin Cleavage and Deprotection Following Solid-phase Synthesis', in *Techniques in Protein Chemistry V*, ed. J.W. Crabb, Academic Press, New York, 501, 1993.
80. R.H. Angeletti, L. Bibbs, L.F. Bonewald, G.B. Fields, J.S. McMurray, W.T. Moore, J.T. Stults, 'Formation of a Disulfide Bond in an Octreotide-like Peptide: a Multicenter Study', in *Techniques in Protein Chemistry VII*, ed. D.R. Marshak, Academic Press, New York, 261, 1996.
81. R.H. Angeletti, L. Bibbs, L.F. Bonewald, G.B. Fields, J.W. Kelly, J.S. McMurry, W.T. Moore, S.T. Weintraub, 'Analysis of Racemization During 'Standard' Solid Phase Peptide Synthesis: a Multicenter Study', in *Techniques in Protein Chemistry VIII*, ed. D.R. Marshak, Academic Press, New York, 875, 1997.
82. B.T. Chait, B.H. Kent, 'Weighting Naked Proteins: practical, High-accuracy Mass Measurement of Peptides and Proteins', *Science*, **257**, 1885–1894 (1992).
83. G.E. Reid, S.A. McLuckey, 'Top down' Protein Characterization via Tandem Mass Spectrometry'. *J. Mass Spectrom.*, **37**, 663–675 (2002).
84. N.L. Kelleher, 'Top-Down Proteomics'. *Anal. Chem.*, **76**, 197A–203A (2004).
85. B. Macek, L.F. Waanders, J.V. Olsen, M. Mann, 'Top-down Protein Sequencing and MS3 on a Hybrid Linear Quadrupole ion Trap-orbitrap Mass Spectrometer', *Mol. Cell Proteomics*, **5**, 949–958 (2006).
86. J.A. Loo, 'Bioanalytical Mass Spectrometry: Many Favors to Choose', *Bioconjugate Chem.*, **6**, 644–665 (1995).
87. G. Siudzak, 'The Emergence of Mass Spectrometry in Biochemical Research', *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 11290–11297 (1994).
88. D.F. Hunt, J.R. Yates III, J. Shabanowitz, S. Winston, C.R. Hauer, 'Protein Sequencing by Tandem Mass Spectrometry', *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 6233–6237 (1986).
89. R.S. Johnson, K. Biemann, 'The Primary Structure of Thioredoxin from Chromatium Vinosum Determined

by High-performance Tandem Mass Spectrometry', *Biochemistry*, **26**, 1209–1214 (1987).

90. R.S. Johnson, W.R. Mathews, K. Biemann, S. Hopper, 'Amino Acid Sequence of Thioredoxin Isolated from Rabbit Bone Marrow Determined by Tandem Mass Spectrometry', *J. Biol. Chem.*, **263**, 9589–9597 (1988).
91. K.F. Medzihradszky, B.W. Gibson, S. Kaur, Z. Yu, D. Medzihradszky, A.L. Burlingame, N.M. Bass, 'The Primary Structure of Fatty-acid-binding Protein from Nurse Shark Liver', *Eur. J. Biochem.*, **203**, 327–339 (1992).
92. J. Lingner, T.R. Hughes, A. Shevchenko, M. Mann, V. Lundblad, T.R. Cech, 'Reversed Transcriptase Motifs in the Catalytic Subunit of Telomerase', *Science*, **276**, 561–567 (1997).
93. W.J. Henzel, J. Tropea, D. Dupont, 'Protein Identification Using 20-minute Edman Cycles and Sequence Mixture Analysis', *Anal. Biochem.*, **267**, 148–160 (1999).
94. E.C. Huang, J.D. Henion, 'LC/MS and LC/MS/MS Determination of Protein Tryptic Digests', *J. Am. Soc. Mass Spectrom.*, **1**, 158–165 (1990).
95. T.R. Covey, E.C. Huang, J.D. Henion, 'Structural Characterization of Protein Tryptic Peptides via Liquid Chromatography/Mass Spectrometry and Collision-induced Dissociation of Their Doubly Charged Molecular Ions', *Anal. Chem.*, **63**, 1193–1200 (1991).
96. F. Wold, 'In Vivo Chemical Modification of Proteins', *Annu. Rev. Biochem.*, **50**, 783–814 (1981).
97. K. Biemann, H.A. Scoble, 'Characterization by Tandem Mass Spectrometry of Structural Modification in Proteins', *Science*, **237**, 992–998 (1987).
98. M. Mann, O.N. Jensen, 'Proteomic Analysis of Post-Translational Modifications', *Nat. Biotechnol.*, **21**, 255–261 (2003).
99. O.N. Jensen, 'Interpreting the Protein Language using Proteomics', *Nat. Rev. Mol. Cell Biol.*, **7**, 392–403 (2006).
100. M.R. Larsen, M.B. Trelle, T.E. Thingholm, O.N. Jensen, 'Analysis of Posttranslational Modifications of Proteins by Tandem Mass Spectrometry', *BioTechniques*, **40**, 790–798 (2006).
101. S.A. Carr, K. Biemann, 'Identification of Posttranslational Modified Amino Acid in Proteins by Mass Spectrometry', *Methods Enzymol.*, **106**, 29–58 (1984).
102. W.J. Boyle, P. van der Geer, T. Hunter, 'Phosphopeptide Mapping and Phosphoamino Acid Analysis by Two-dimensional Separation on Thin-layer Cellulose Plates', *Methods Enzymol.*, **201**, 110–149 (1991).
103. P. Cohen, B.W. Gibson, C.F. Holmes, 'Analysis of the in Vivo Phosphorylation States of Proteins by Fast Atom Bombardment Mass Spectrometry and Other Techniques', *Methods Enzymol.*, **201**, 153–168 (1991).
104. R.S. Annan, S.A. Carr, 'The Essential Role of Mass Spectrometry in Characterizing Protein Structure: Mapping Posttranslational Modifications', *J. Protein Chem.*, **16**, 391–402 (1997).
105. K.A. Resing, N.G. Ahn, 'Protein Phosphorylation Analysis by Electrospray Ionization–Mass Spectrometry', *Methods Enzymol.*, **283**, 29–44 (1997).
106. K.A. Resing, R.S. Johnson, K.A. Walsh, 'Mass Spectrometry Analysis of 21 Phosphorylation Sites in the Internal Repeat of Rat Profilaffrin, Precursor of an Intermediate Filament Associated Protein', *Biochemistry*, **34**, 9477–9487 (1995).
107. J.D. Watts, M. Affolter, D.L. Krebs, R.L. Wange, L.E. Samelson, R. Aebersold, 'Identification by Electrospray Ionization Mass Spectrometry of the Sites of Tyrosine Phosphorylation Induced in Activated Jurkat T Cells on the Protein Tyrosine Kinase ZAP-70', *J. Biol. Chem.*, **269**, 29520–29529 (1994).
108. G. Neubauer, M. Mann, 'Mapping of Phosphorylation Sites of Gel-isolated Proteins by Nanoelectrospray Tandem Mass Spectrometry: Potentials and Limitations', *Anal. Chem.*, **71**, 235–242 (1999).
109. T.R. Covey, B. Shushan, B. Bonner, W. Schroder, F. Hucho, 'LC/MS and LC/MS/MS Screening for Sites of Post-translational Modification in Proteins', in *Methods in Protein Sequence Analysis*, eds. H. Jorvall, J.-O. Hoog, A.M. Gustavsson, Birkhauser Verlag, Basel, 249–256, 1991.
110. M.J. Huddleston, R.S. Annan, M.F. Bean, S.A. Carr, 'Selective Detection of Phosphopeptides in Complex Mixtures by Electrospray Liquid Chromatography/Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 710–717 (1993).
111. J.J. Conboy, J.D. Henion, 'The Determination of Glycopeptides by Liquid Chromatography/Mass Spectrometry with Collision-induced Dissociation', *J. Am. Soc. Mass Spectrom.*, **3**, 804 (1992).
112. J. Ding, W. Burkhary, D.B. Kassel, 'Identification of Phosphorylated Peptides from Complex Mixtures Using Negative-ion Orifice-potential Stepping and Capillary Liquid Chromatography/Electrospray Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 94–98 (1994).
113. R. Verma, R.S. Annan, M.J. Huddleston, S.A. Carr, G. Reynard, R.J. Deshaies, 'Phosphorylation of Sic1p by G1 Cdk Required for its Degradation and Entry into S Phase', *Science*, **278**, 455–460 (1997).
114. R.S. Annan, M.J. Huddleston, R. Verma, R.J. Deshaies, S.A. Carr, 'A Multi-dimensional Electrospray MS-Based Approach to Phosphopeptide Mapping', *Anal. Chem.*, **73**, 393–404 (2001).
115. H. Lis, N. Sharon, 'Protein Glycosylation: Structural and Functional Aspects', *Eur. J. Biochem.*, **218**, 1–27 (1993).

116. V. Ling, A.W. Guzzetta, E. Canova-Davis, J.T. Stults, W.S. Hancock, T.R. Covey, B.I. Shusan, 'Characterization of the Tryptic Map of Recombinant DNA Derived Tissue Plasminogen Activator by High-performance Liquid Chromatography–Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **63**, 2909–2915 (1991).

117. W.S. Hancock, A. Apffel, J. Chankel, C. Souders, T.M. Timkulu, E. Pungor, A.W. Guzzetta, 'Reversed-phase Peptide Mapping of Glycoproteins Using Liquid Chromatography/Electrospray Ionization Mass Spectrometry', *Methods Enzymol.*, **271**, 403–427 (1996).

118. M.J. Huddleston, M.F. Bean, S.A. Carr, 'Collisional Fragmentation of Glycopeptides by Electrospray Ionization LC/MS and LC/MS/MS: Methods for Selective Detection of Glycopeptides in Protein Digests', *Anal. Chem.*, **65**, 877–884 (1993).

119. P.T. Jedrzejewski, W.D. Lehmann, 'Detection of Modified Peptides in Enzymatic Digests by Capillary Liquid Chromatography/Electrospray Mass Spectrometry and a Programmable Skimmer CID Acquisition Routine', *Anal. Chem.*, **69**, 294–301 (1997).

120. S.A. Carr, M.J. Huddleston, M.F. Bean, 'Selective Identification and Differentiation of N- and O-linked Oligosaccharides in Glycoproteins by Liquid Chromatography–Mass Spectrometry', *Protein Sci.*, **2**, 183–196 (1993).

121. H.A. Scoble, S.A. Martin, 'Characterization of Recombinant Proteins', *Methods Enzymol.*, **193**, 519–536 (1990).

122. K.B. Linsley, S.-Y. Chan, S. Chan, B.B. Reinhold, P.J. Lisi, V.N. Reinhold, 'Applications of Electrospray Mass Spectrometry to Erythropoietin N- and O-linked Glycans', *Anal. Biochem.*, **219**, 207–217 (1994).

123. R.S. Rush, P.L. Derby, D.M. Smith, C. Merry, G. Rogers, M.F. Rohde, V. Katta, 'Microheterogeneity of Erythropoietin Carbohydrate Structure', *Anal. Chem.*, **67**, 1442–1452 (1995).

124. I. Mazsaroff, W. Yu, B.D. Kelley, J.E. Vath, 'Quantitative Comparison of Global Carbohydrate Structures of Glycoproteins Using LC-MS and In-source Fragmentation', *Anal. Chem.*, **69**, 2517–2524 (1997).

125. K.F. Medzihradszky, M.J. Besman, A.L. Burlingame, 'Structural Characterization of Site-specific N-glycosylation of Recombinant Human Factor VIII by Reversed-phase High-performance Liquid Chromatography–Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **69**, 3986–3994 (1997).

126. D.L. Smith, Z. Zhou, 'Strategies for Locating Disulfide Bonds in Proteins', *Methods Enzymol.*, **193**, 374–389 (1990).

127. K.F. Medzihradszky, M.J. Besman, A.L. Burlingame, 'Reversed-phase Capillary High Performance Liquid Chromatography/High Performance Electrospray Ionization Mass Spectrometry: An Essential Tool for the Characterization of Complex Glycoprotein Digests', *Rapid Commun. Mass Spectrom.*, **12**, 472–478 (1998).

128. R.G. Keck, L. Berleau, R. Harris, B.A. Keyt, 'Disulfide Structure of the Heparin Binding Domain in Vascular Endothelial Growth Factor: Characterization of Posttranslational Modifications in VEGF', *Arch. Biochem. Biophys.*, **344**, 103–113 (1997).

129. P. Kahn, 'From Genome to Proteome: Looking at a Cell's Proteins', *Science*, **270**, 369–370 (1995).

130. W.P. Blackstock, M.P. Weir, 'Proteomics: Quantitative and Physical Mapping of Cellular Proteins', *Trends Biotechnol.*, **17**, 121–127 (1999).

131. P.A. Haynes, S.P. Gygi, D. Figeys, R. Aebersold, 'Proteome Analysis: Biological Assay or Data Archive', *Electrophoresis*, **19**, 1862–1871 (1998).

132. M. Tylers, M. Mann, 'From Genomics to Proteomics', *Nature*, **422**, 193–197 (2003).

133. J. Malmstrom, H. Lee, R. Abeersold, 'Advances in Proteomics Workflows for System Biology', *Curr. Opin. Biotechnol.*, **18**, 378–384 (2007).

134. B.F. Cravatt, G.M. Simon, J.R. Yates III, 'The Biological Impact of Mass-Spectrometry Based Proteomics', *Nature*, **450**, 991–1000 (2007).

135. W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, 'Identifying Proteins from Two-dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Database', *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5011–5015 (1993).

136. K.R. Clauser, S.C. Hall, D.M. Smith, J.W. Webb, L.E. Andrews, H.M. Tran, L.B. Epstein, A.L. Burlingame, 'Rapid Mass Spectrometric Peptide Sequencing and Mass Matching for Characterization of Human Melanoma Proteins Isolated by Two-dimensional PAGE', *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 5072–5076 (1995).

137. A. Shevchenko, O.N. Jensen, A.V. Potelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, 'Linking Genome and Proteome by Mass Spectrometry: Large-scale Identification of Yeast from Two Dimensional Gels', *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 14440–14445 (1996).

138. J.R. Yates III, 'Mass Spectrometry and the Age of the Proteome', *J. Mass Spectrom.*, **33**, 1–19 (1998).

139. S.D. Patterson, 'Electrophoretically Separated Protein to Identification: Strategies for Sequence and Mass Analysis', *Anal. Biochem.*, **221**, 1–15 (1994).

140. S.D. Patterson, R. Aebersold, 'Mass Spectrometric Approaches for Identification of Gel-separated Proteins', *Electrophoresis*, **16**, 1791–1814 (1995).

141. J. Fernandez, F. Gharahdaghi, S.M. Mische, 'Routine Identification of Proteins from Sodium Dodecyl

Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) Gels or Polyvinyl Difluoride Membranes Using Matrix Assisted Laser Desorption/Ionization–Time-of-flight–Mass Spectrometry (MALDI–TOF–MS), *Electrophoresis*, **19**, 1036–1045 (1998).

142. K.L. Stone, R. DeAngelis, M. LoPresti, J. Jones, V.V. Papov, K.R. Williams, 'Use of Liquid Chromatography–Electrospray Ionization Tandem Mass Spectrometry (LC–ESIMS/MS) for Routine Identification of Enzymatically Digested Proteins Separated by Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis', *Electrophoresis*, **19**, 1046–1052 (1998).

143. J. Rosenfeld, J. Capevielle, J.C. Guillemot, P. Ferrara, 'In-gel Digestion of Proteins for Internal Sequence Analysis After One- or Two-dimensional Gel Electrophoresis', *Anal. Biochem.*, **203**, 173–179 (1992).

144. J. Fernandez, L. Andrews, S.M. Mische, 'An Improved Procedure for Enzymatic Digestion of Polyvinylidene Difluoride-bound Proteins for Internal Sequence Analysis', *Anal. Biochem.*, **218**, 112–117 (1994).

145. M. Lui, P. Tempst, H. Erdjument-Bromage, 'Methodical Analysis of Protein–Nitrocellulose Interactions to Design a Refined Digestion Protocol', *Anal. Biochem.*, **241**, 156–166 (1996).

146. A. Ducret, I.V. Oostveen, J.K. Eng, J.R. Yates III, R. Aubersold, 'High Throughput Protein Characterization by Automated Reverse-phase Chromatography/Electrospray Tandem Mass Spectrometry', *Protein Sci.*, **7**, 706–719 (1998).

147. A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates III, 'Direct Analysis of Protein Complexes Using Mass Spectrometry', *Nat. Biotechnol.*, **17**, 676–682 (1999).

148. J.R. Yates, J.K. Eng, K.R. Clauser, A.L. Burlingame, 'Search of Sequence Database with Uninterpreted High-energy Collision-induced Dissociation Spectra of Peptides', *J. Am. Soc. Mass Spectrom.*, **7**, 1089–1098 (1996).

149. D.J.C. Pappin, P. Hojrup, A.J. Bleasby, 'Rapid Identification of Proteins by Peptide Mass Fingerprinting', *Curr. Biol.*, **3**, 327–332 (1993).

150. M. Mann, P. Hojrup, P. Roepstorff, 'Use of Mass Spectrometric Molecular Weight Information to Identify Proteins in Sequence Databases', *Biol. Mass Spectrom.*, **22**, 338–345 (1993).

151. P. James, M. Quadroni, E. Carafoli, G. Gonnet, 'Protein Identification by Mass Profile Fingerprinting', *Biochem. Biophys. Res. Commun.*, **195**, 58–64 (1993).

152. J.R. Yates III, S. Speicher, P.R. Griffin, T. Hunkapiller, 'Peptide Mass Maps: a Highly Informative Approach to Protein Identification', *Anal. Biochem.*, **214**, 397–408 (1993).

153. P. James, M. Quadroni, E. Carafoni, G. Gonnet, 'Protein Identification in DNA Database by Peptide Mass Fingerprinting', *Protein Sci.*, **3**, 1347–1350 (1994).

154. K.M. Swiderek, M.T. Davis, T.D. Lee, 'The Identification of Peptide Modifications Derived from Gel-separated Proteins Using Electrospray Triple Quadrupole and Ion Trap Analyses', *Electrophoresis*, **19**, 989–997 (1998).

155. M. Mann, M. Wilm, 'Error-tolerant Identification of Peptides in Sequence Database by Peptide Sequence Tags', *Anal. Chem.*, **66**, 4390–4399 (1994).

156. E. Mortz, P.B. O'Connor, P. Roepstorff, N.L. Keleher, T.D. Wood, F.W. McLafferty, M. Mann, 'Sequencing Tag Identification of Intact Proteins by Matching Tandem Mass Spectral Data Against Sequence Data Bases', *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8264–8267 (1996).

157. D. Fenyo, J. Qin, B.T. Chait, 'Protein Identification Using Mass Spectrometric Information', *Electrophoresis*, **19**, 998–1005 (1998).

158. J.R. Yates, 'Database Searching Using Mass Spectrometry Data', *Electrophoresis*, **19**, 893–900 (1998).

159. R.G. Sadygov, D. Cociorva, J.R. Yates III, 'Large-Scale Database Searching Using Tandem Mass Spectra: Looking up the Answer in the Back of the Book', *Nat. Methods*, **1**, 195–202 (2004).

160. A.I. Nesvizhskii, R. Aebersold, 'Analysis, Statistical Validation and Dissemination of Large-Scale Proteomics Datasets Generated by Tandem MS', *Drug Discovery Today*, **9**, 173–181 (2004).

161. B. Kuster, M. Shirle, P. Mallick, R. Abersold, 'Scoring Proteomes with Proteotypic Peptide Probes', *Nat. Rev. Mol. Cell Biol.*, **6**, 577–583 (2005).

162. A.I. Nesvizhskii, O. Vitek, R. Aebersold, 'Analysis and Validation of Proteomics Data Generated by Tandem Mass Spectrometry', *Nat. Methods*, **4**, 787–797 (2007).

163. H. Lam, E. W. Deutsch, J. S. Eddes, J. K. Eng, N. King, S.E. Stein, R. Aebersold, 'Development and Validation of a Spectral Library Searching Method for Peptide Identification from MS/MS', *Proteomics*, **7**, 655–667 (2007).

Hydrophilic-interaction Chromatography in Peptide and Protein Analysis

Paul Jenö
Biozentrum of the University of Basel, Basel, Switzerland

1 Introduction 1
2 The "Hydrophilic" Effect 2
3 Applications 3
3.1 Hydrophilic Interaction/Cation-exchange Chromatography for Separation of Amphipathic α-Helical Peptides 3
3.2 Analysis of Protein Glycosylation by Hydrophilic-interaction Liquid Chromatography 5
3.3 Analysis of Protein Phosphorylation and Acetylation by Hydrophilic-interaction Liquid Chromatography 6
3.4 Removal of Detergent from Membrane Proteins 10
4 Conclusions 12
Abbreviations and Acronyms 12
Related Articles 12
References 12

Hydrophilic-interaction liquid chromatography (HILIC) has been introduced as a highly efficient chromatographic technique for the separation of a wide range of solutes. A hydrophilic stationary phase is eluted with a hydrophobic mobile phase whereby retention of solutes increases with the hydrophilicity of solutes. Bound solutes are eluted by decreasing the hydrophobicity of the mobile phase. Mixed mode effects based on hydrophilic and ionic interactions can also be exploited by proper selection of the stationary and the mobile phase. Typically, the order of elution is the opposite of that obtained by reversed-phase chromatography (RPC). HILIC is particularly interesting for solutes which are difficult to chromatograph on conventional reversed-phase materials. This review focuses on the application of HILIC to the separation of amphipathic α-helical peptides, the analysis of protein glycosylation, phosphorylation, and acetylation, and to the removal of detergents from membrane proteins.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

It is axiomatic in biological sciences that understanding the function of proteins or peptides requires determination of their structures. Advances in chemical technology, the development of cDNA chip technology and the accumulation of DNA data banks make it possible to obtain a global and quantitative picture of mRNA and protein expression in model organisms such as the yeast *Saccharomyces cerevisiae*.[1] Changes in relative abundance of mRNAs and proteins in response to environmental changes[2–4] or in response to alteration of the expression of transcriptional factors[2] have been investigated at a genome-wide level.

Instrumental to these breakthroughs was the development of interfacing high-resolution protein separation such as two-dimensional (2-D) gel electrophoresis[5] with various chromatographic techniques and mass spectrometry (MS) for high sensitivity detection and analysis of peptides and proteins. Reversed-phase high-performance liquid chromatography (RPHPLC) is one of the most widely used chromatographic separations. The combination of RPHPLC and electrospray ionization (ESI) represents a particularly attractive configuration for the on-line analysis of complex peptide mixtures.[6] Automation of chromatography and mass spectral analysis in combination with powerful data base searching algorithms allow the identification of hundreds of proteins separated by 2-D gel electrophoresis.[7–9] As the number of protein sequences compiled in databases increased, it has become feasible to match a gel spot to a known protein without extensive sequence information. In favorable cases, as few as two peptides are sufficient to identify proteins in gels at the low picomole level, even when several unresolved proteins are present.[10,11]

As powerful as these methods are for protein identification, the structure and the function of proteins is in many cases critically regulated by modifications. Co- and posttranslational modifications such as phosphorylation, sulfation, glycosylation, and *N*-terminal modifications are means to regulate the cellular distribution and function of proteins and are of particular importance in cell signalling.[12] To date, approximately 250 physiological covalent modifications are known.[13,14] For the characterization of such modifications, various analytical techniques such as radiolabeling followed by chemical cleavage (e.g. Edman degradation) are often applied to unravel the nature of such modifications. Common to all of these experiments is the need to chemically, or enzymatically, cleave the protein into smaller pieces, isolate the fragment of interest and determine its precise location and nature of modification. The majority of such peptide separations is carried out by RPHPLC. However,

the nature of the modification, or the tactical considerations leading to the isolation of peptides carrying the modification may necessitate the use of alternative chromatographic techniques.

Since its introduction in 1990 by Alpert[15] HILIC has been used in a number of applications to solve chromatographic problems not amenable to conventional types of chromatographies, such as ion-exchange, or RPHPLC. This article is intended to give a selection of recent applications of HILIC in the field of bioanalytical chromatography in the hope that they can be taken as a basis to solve chromatographic problems other than the ones outlined below.

2 THE "HYDROPHILIC" EFFECT

In the search for new stationary phases which can complement RPC, the hydrophilic strong-cation-exchange (CEX) material PolySulfoethyl A was introduced in 1988.[16] Retention of solutes in mobile phases containing low concentrations of acetonitrile on such PolySulfoethyl A supports is mainly governed by charge effects due to the electrostatic interactions between solutes and the stationary phase. Thus, elution of peptides, for example, is governed by the number of basic residues when chromatography is performed at pH 3. Typically, columns of this material are eluted with a salt gradient.

The selectivity of the stationary phase changes markedly when the proportion of organic solvent is increased above 70%: retention of solutes is drastically increased. These data suggest that an increase in organic solvent leads to a higher sensitivity of the stationary phase to interact with hydrophilic residues of peptides. If this were the case, elution of solutes should occur according to increasing hydrophilicity. When the elution behavior of a series of peptides with increasing hydrophobicity was compared between RPC and HILIC, it was shown that the order of elution between the two modes of chromatography was reversed, strongly suggesting that the interaction of solutes and the stationary phase is indeed governed by hydrophilic interactions.[17] Therefore, in contrast to "normal-phase" chromatography, where solutes are adsorbed to a polar stationary phase and elution is promoted by increasing the polarity of the mobile phase, the combination of a hydrophilic stationary phase and hydrophobic, mostly organic phases was designated as HILIC.[15]

As an example, the adsorption behavior of amino acids on two different, polar stationary phases, PolySulfoethyl A and PolyHydroxyethyl A is shown (Figure 1a). At pH 2.8, no amino acids are retained to a significant extent on PolyHydroxyethyl A in the absence of acetonitrile, whereas the basic amino acids Arg, His, and Lys are slightly retained on PolySulfoethyl A, probably due to ionic interactions between the negatively charged sulfonic group and the positively charged side chains of the amino acids. A strong increase in retention of all amino acids occurs when the acetonitrile concentration is raised above 60%. Interestingly, the most polar amino acids Arg, His, and Lys show the most pronounced increase in retention, arguing that the interactions between the stationary phase and the solutes are indeed of hydrophilic nature (Figure 1a). Elution

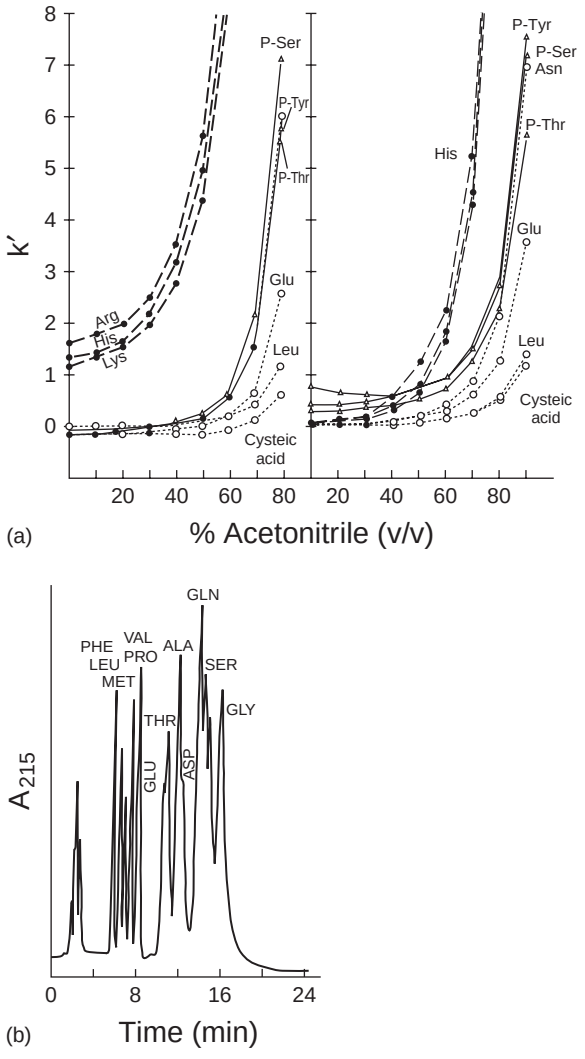

Figure 1 (a) Capacity factors of free amino acids chromatographed on 4.6 × 200 mm columns packed with PolySulfoethyl A (left) or PolyHydroxyethyl A (right). The mobile system was 25 mM triethylamine phosphate adjusted to pH 5.0 containing acetonitrile with the concentrations indicated on the *x*-axis. (b) HILIC of free amino acids on a column (4.6 × 200 mm) packed with PolySulfoethyl A. Solvent system: 5 mM triethylamine phosphate, pH 2.8, containing 80% acetonitrile. Flow rate 2 ml min^{-1}. (Reproduced from Alpert[15] with permission of Elsevier Science © 1990.)

of bound amino acids from PolySulfoethyl A occurs in the order of their hydrophilicities: the most unpolar amino acids Phe, Leu, Met, Pro, and Val elute first, followed by the more polar amino acids (Figure 1b). Also, when compared to RPC, the order of elution is reversed: the amino acids Met, Leu, Ile, Trp, and Phe eluting last from reversed-phase columns are the first to elute from PolySulfoethyl A or PolyHydroxyethyl A,[15] again arguing that HILIC exploits interactions based on hydrophilicity in the chromatographic process.

The mechanism by which solutes interact with a hydrophilic stationary phase at high organic solvent concentration is not fully understood. In simple terms, HILIC can be regarded as a variant of normal-phase chromatography where retention is proportional to the polarity of the solute and inversely proportional to the polarity of the mobile phase. However, in contrast to normal-phase chromatography, there is evidence that the mechanism of HILIC involves partitioning between the hydrophobic mobile phase and a layer of mobile phase which is enriched with water and is partially immobilized on the stationary phase.[15]

From a practical point of view, the chromatographer now has the choice of two well defined stationary phases for HILIC: PolySulfoethyl A and PolyHydroxyethyl A. The former matrix can be used primarily as a CEX material, whose polar matrix is well suited for solving specific separation problems such as the isolation of C-terminal peptides,[18] disulfide-linked peptides,[19] and pyroglutamyl-terminated peptides[20] from protein digests. Interesting applications arise, when the two properties, ion-exchange- and hydrophilic interactions of the matrix at high concentrations of organic solvent, are exploited simultaneously. The latter matrix (PolyHydroxyethyl A) is more hydrophilic than PolySulfoethyl A and should therefore be used for substances which fail to bind to PolySulfoethyl A. Although not evident from its structure, in addition to the hydrophilic interactions the PolyHydroxyethyl A support also exhibits weak ionic interactions.[17] Therefore, the 'ideal' sorbent exhibiting solely hydrophilic interactions has yet to be developed. However, these mixed-mode HILIC/ion-exchange (HILIC/IEX) interactions allow a great deal of chromatographic flexibility. The CEX properties of the support separates solutes based on net positive charge and this separation mode is overlaid by the presence of acetonitrile overcoming undesirable hydrophobic interactions with the sorbent while promoting favorable hydrophilic interactions. Thus, mixed-mode separations are typically carried out with linearly increasing salt gradients at pH 3, with mobile phases containing 15–80% acetonitrile. On the other hand, separations exploiting primarily hydrophilic interactions require gradients of decreasing organic solvent (also termed 'inverse' gradients) or increasing salt. Which of the two types of separations is to be chosen depends on the goal of the individual separation process and has to found on a trial basis.

3 APPLICATIONS

3.1 Hydrophilic Interaction/Cation-exchange Chromatography for Separation of Amphipathic α-Helical Peptides

A particularly illustrative example of HILIC combined with ion-exchange chromatography was recently published by Mant and coworkers.[21] The goal was to assess the relative contributions of α-helical structural elements on the influence of chromatographic behavior during RPC and HILIC. A series of amphipathic α-helical peptides was synthesized whose amino acids in the centre of their hydrophilic or hydrophobic faces were varied to assess their relative contributions for their retention behavior during HILIC or RPC.

In all of these studies, HILIC was carried out in the mixed-mode where hydrophilic and ionic interactions are superimposed. Consequently, the strong CEX material PolySulfoethyl A was operated at pH 3.0 to ensure a net positive charge of the peptides by protonation of acidic and basic side chains (the pK_a value of glutamate and aspartate is approximately 4.5, and all basic amino acids are fully protonated). Also, pH 3 leaves a net negative charge on the sulphonate CEX functional group ($pK_a \sim 1-1.5$). Hence, the use of 20 mM aqueous triethylammonium phosphate (TEAP) in the presence of $NaClO_4$ was chosen. The choice of salts for elution of ionically bound substances is of particular importance if hydrophilic interactions are to be exploited at relatively high organic solvent concentrations. $NaClO_4$ is suitable for this purpose due to its excellent solubility characteristics in the presence of high concentrations of organic solvent (up to 400 mM $NaClO_4$ in 80% acetonitrile).

To study the effect of amphipathicity on the retention behavior, two series of peptides were chromatographed: the two nonamphipathic analogs Ac-Glu-Glu-Ala-Lys-Ala-Lys-Ala-Glu-Ala-Glu-Ala-Lys-Ala-Lys-Ala-Glu-Glu-Lys-NH_2 (naA) and Ac-Glu-Glu-Leu-Lys-Leu-Lys-Leu-Glu-Leu-Glu-Leu-Lys-Leu-Lys-Leu-Glu-Glu-Lys-NH_2 (naL) and the two amphipathic analogs Ac-Glu-Ala-Glu-Lys-Ala-Ala-Lys-Glu-Ala-Glu-Lys-Ala-Ala-Lys-Glu-Ala-Glu-Lys-NH_2 (AA9) and Ac-Glu-Leu-Glu-Lys-Leu-Leu-Lys-Glu-Leu-Glu-Lys-Leu-Leu-Lys-Glu-Leu-Glu-Lys-NH_2 (LL9). The periodic distribution of the non-polar residues alanine in AA9 and leucine in LL9 ensures a wide hydrophobic face on the amphipathic helix,

whereas the hydrophilic face is made up of lysine and glutamic acids. Furthermore, the peptide pairs naA and AA9, and naL and LL9, respectively, have the same amino acid composition, but different amino acid sequences. RPC of the mixture of the α-helical peptides AA9 and LL9 and their nonamphipathic counterparts naA and naL not only separated all four peptides, separation also occurred between the nonamphipathic and their amphipathic analogs (Figure 2a). For both pairs of peptides (AA9/naA and LL9/naL) the amphipathic analogs AA9 and LL9 eluted later in the gradient than their nonamphipathic analogs AA9 and LL9. This is in accordance with findings previously reported that conformational effects can have a profound effect on the retention behavior of peptides on RPC.[22–24] Upon binding to a hydrophobic stationary phase, α-helical structures are stabilized[22] and due to the preferred binding domain created by the non-polar face of an α-helix, they are more retentive than peptides having the same composition (and therefore the same overall hydrophobicity) but lacking secondary structural elements creating such preferred binding domains.

In the absence of any conformational effects, peptide retention between HILIC and RPC is reversed, which means that the most hydrophilic peptides are eluted last during HILIC and first during RPC. If adsorption of peptides to the stationary phase occurs predominantly through preferred binding domains formed by secondary structural elements such as α-helices, one would expect in HILIC that amphipathic peptides bind to the hydrophilic stationary phase via their polar face, and the unpolar face will generally be oriented away and will have a lesser effect on retention if these residues were distributed evenly throughout the helix. This is clearly demonstrated due to the fact that the two amphipathic peptides LL9 and AA9 eluted later during HILIC/IEX than their nonamphipathic counterparts naL and naA (Figure 2b). Also, within the corresponding amphipathic and nonamphipathic analogs, the less polar Leu analog elutes earlier than the Ala analog. These results clearly demonstrate that binding of peptides in the HILIC/IEX mode occurs through interactions between hydrophilic portions of peptides and the hydrophilic surface of the stationary phase and that these interactions can become stronger when secondary structural elements such as α-helices form preferred binding domains.

A distinct advantage of HILIC lies in the fact that due to the high content of organic solvents used to promote hydrophilic interactions, extremely hydrophobic peptides can be chromatographed under more favorable conditions than by RPC.[21] Some examples are the extremely hydrophobic ampipathic α-helical peptides LL7 (Ac-Glu-Leu-Glu-Lys-Leu-Leu-Leu-Glu-Leu-Glu-Lys-Leu-Leu-Lys-Glu-Leu-Glu-Lys-amide), LV7 (Ac-Glu-Leu-Glu-Lys-Leu-Leu-Val-Glu-Leu-Glu-Lys-Leu-Leu-Lys-Glu-Leu-Glu-Lys-amide), LT7 (Ac-Glu-Leu-Glu-Lys-Leu-Leu-Thr-Glu-Leu-Glu-Lys-Leu-Leu-Lys-Glu-Leu-Glu-Lys-amide), and LS7 (Ac-Glu-Leu-Glu-Lys-Leu-Leu-Ser-Glu-Leu-Glu-Lys-Leu-Leu-Lys-Glu-Leu-Glu-Lys-amide). These peptides have a long hydrophilic face formed by the acid and basic amino acids at positions 1, 3, 4, 8, 10, 11, 14, 15, 17, and 18. In addition, the center of the polar face formed by amino acid seven is varied in terms of its polarity from an unpolar leucine to a polar serine residue. The hydrophobic face is formed by the extremely unpolar amino acid leucine at positions 2, 5, 6, 9, 12, 13, and 16. Thus, the preferred binding sites for the interaction between reversed-phase supports is formed by identical

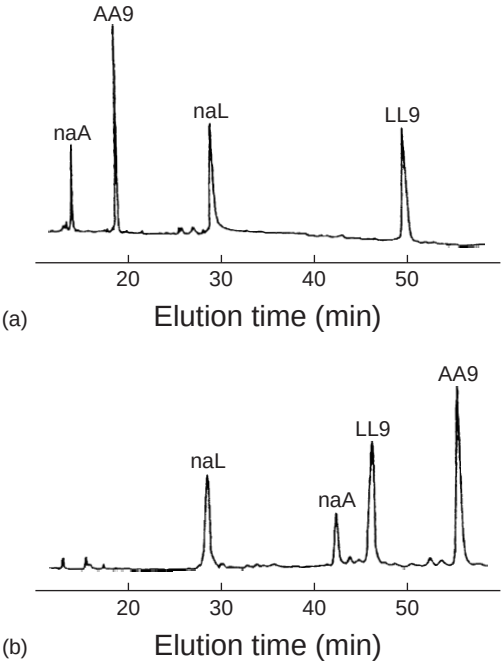

Figure 2 Reversed-phase liquid chromatography (RPLC) (a) and HILIC/CEX (b) of amphipathic peptides AA9 and LL9 and nonamphipathic peptides naA and naL (for the peptide sequences, see text). Chromatographic conditions for RPLC were: column: 4.6 × 150 mm packed with Zorbax SB300-C_8, 5 µm particle size, 300 Å pore size), solvent A: 20 mM triethylamine phosphate, pH 3.0, 100 mM $NaClO_4$, solvent B: solvent A containing 60% acetonitrile. The gradient was 0.5% acetonitrile change per min at a flow rate of 1 ml min^{-1}. Chromatographic conditions for HILIC/CEX were: column: 4.6 × 200 mm packed with PolySulfoethyl A (5 µm particle size, 300 Å pore size), solvent A: 20 mM triethylamine phosphate buffered to pH 3, containing 80% acetonitrile, solvent B: solvent A containing 400 mM $NaClO_4$. Elution was with a linear gradient starting from 0 to 400 mM $NaClO_4$ at 5 mM $NaClO_4$ change per minute. Peptide detection was at 210 nm. (Reproduced from Mant et al.[21] with permission of Elsevier Science © 1998.)

sequences in all four peptides and should therefore result in coelution of the peptides during RPLC. Indeed, coelution of all peptides during RPC occurred at pH 2 and 3 (Figure 3). Obviously, substitutions in the hydrophilic face has little effect on the retention behavior of the peptides. The increased retention time at pH 3 is either a consequence of the binding of the negatively charged ClO_4^- ion to the positively charged lysine groups of the peptides, or might be caused by enhanced hydrophobic effects between the peptides and the stationary phase induced by the relatively high concentration of the salt present in the mobile phase.

In contrast to RPC, all four peptides were well separated by HILIC/IEX (Figure 3). Since the four peptides have all identical net positive charges and therefore, act equally strongly with the ion exchange matrix, the separation is a consequence of the different substitutions introduced into the hydrophilic face of these peptides. As previously observed, the elution order follows the decreasing hydrophobicity with the leucine analog eluting first, followed by the Val, Thr, and Ser analogs, underscoring the fact that the interaction with the hydrophilic stationary phase occurs predominantly through the hydrophilic face present in these amphipathic peptides. Also note the dramatic improvement of peak shape in HILIC/IEX when compared to RPC, indicating that HILIC might be a better alternative for the chromatography of extremely hydrophobic peptides.

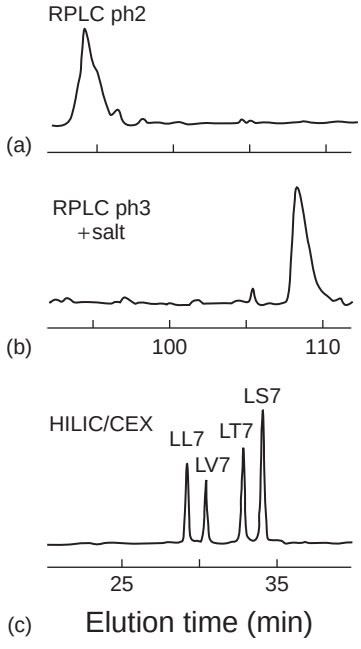

Figure 3 RPLC (a) and (b) and HILIC/CEX (c) of amphipathic α-helical peptides (for the sequences of the LL7, LV7, LT7, and LS7 peptides see text). Chromatographic conditions for RPLC at pH 2 (a): solvent A: 0.05% TFA, solvent B: 0.05% TFA in acetonitrile, linear gradient at 0.5% acetonitrile change per minute; chromatographic conditions at pH 3 (b): same as for HILIC/CEX in Figure 2, except that the concentration of $NaClO_4$ was 100 mM in both solvents A and B. Chromatographic conditions for HILIC/CEX: see Figure 2. (Reproduced from Mant et al.[21] with permission of Elsevier Science © 1998.)

3.2 Analysis of Protein Glycosylation by Hydrophilic-interaction Liquid Chromatography

Expression of proteins for therapeutic use in mammalian cell culture systems has created demands for extensive analytical characterization of the recombinant product. For many glycoproteins, the carbohydrate moieties endow their polypeptide portion with important physical properties such as conformational stability, protease resistance, charge, and water binding capacity. Interest in the biological roles of the carbohydrate chains of glycoproteins has been intensified because of interesting examples of their importance as recognition markers for protein targeting and cell–cell and cell–matrix interactions. Characterization of the glycosylation pattern of a protein is challenging in as much as glycosylation sites can be occupied with carbohydrates of different structures. Traditionally, characterization of glycosylation sites is carried out by generating glycopeptides from enzymatic digests of glycoproteins which are usually separated by RPHPLC. However, because the hydrophobic stationary phase exploits interactions with hydrophobic residues of the peptide chain, separation of peptides having the same glycosylation site, but carrying different oligosaccharide structures, is not always successful. In contrast, it can be assumed that the high initial concentration of organic solvent used in HILIC promotes interactions primarily between the hydrophilic stationary phase and the hydrophilic oligosaccharide moiety of a glycopeptide, therefore exploiting interactions more favorable for the separation of differentially glycosylated peptides. This was convincingly demonstrated by analyzing the glycosylation pattern of interferon-γ expressed in Chinese hamster ovary cell culture.[25] The two tryptic glycopeptides containing *N*-linked carbohydrates at Asn_{25} and Asn_{97} were first isolated by RPHPLC and further analyzed by hydrophilic interaction chromatography on a PolyHydroxyethyl A column at pH 6.0 in 85% acetonitrile. Bound glycopeptides were eluted with a gradient of decreasing acetonitrile and increasing $NaClO_4$ concentration, thus separation by HILIC was carried out in mixed-mode (ionic-overlaid with hydrophilic interactions). Both seemingly pure glycopeptides isolated by RPC were further resolved by HILIC into a number of peaks assumed to contain different oligosaccharide structures (Figure 4). By measuring the mass of each of the components by matrix-assisted laser

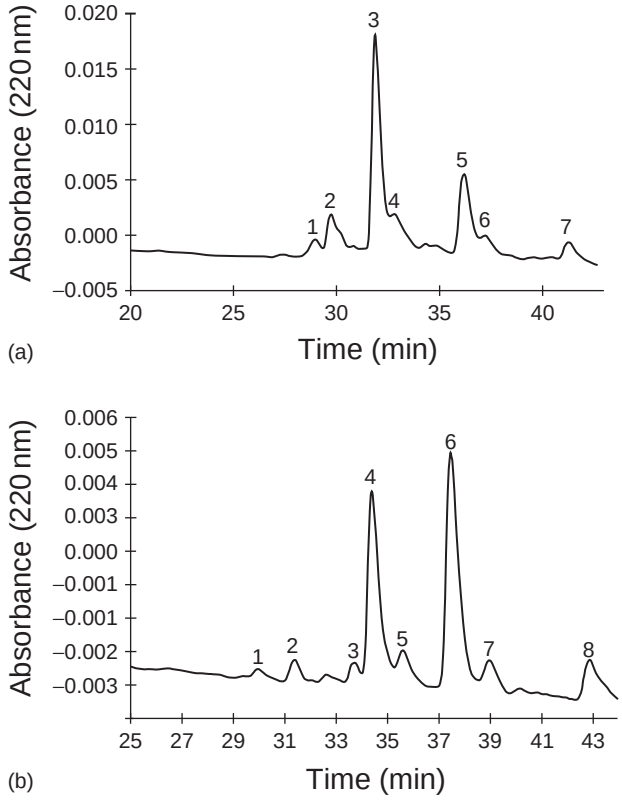

Figure 4 HILIC of tryptic glycopeptides of Asn_{25} (a) and Asn_{97} (b) from recombinant human γ-interferon isolated from chinese hamster ovary cell cultures. Chromatographic conditions: column (1 × 150 mm) packed with PolyHydroxyethyl A, solvent A: 10 mM triethylamine, pH 6.0, 85% acetonitrile, solvent B: 10 mM triethylamine, pH 6.0, 25 mM $NaClO_4$, 20% acetonitrile, 60 min gradient from 100% solvent A to 100% solvent B at a flow rate of 50 µl min^{-1}. Peptide monitoring was at 220 nm. (Reproduced from Zhang and Wang[25] with permission of Elsevier Science © 1998.)

desorption/ionization/time-of-flight (MALDI/TOF) MS the glycan structures for the two glycopeptides were identified by comparing the mass difference between the observed mass of the glycopeptides and the known mass of its peptide fragments. The most abundant species on the Asn_{25} peptide were fully sialylated bi- and triantennary glycans, containing one fucose unit, whereas the Asn_{97} glycopeptide contained mostly sialylated biantennary glycans lacking fucose (Figure 5). From the structures shown in Figure 5(a), it is apparent that the sialic acids have the most profound effect on retention of glycopeptides during HILIC: the glycopeptides are eluted according to their degree of sialylation (monosialo elutes first, tetrasialo last). Whereas in RPC separation the peptide portion of the glycopeptides is exploited for separation, HILIC offers an alternative separation method for the resolution of differentially glycosylated peptides by exploiting contact regions other than the peptide backbone.

Another observation interesting to note is that HILIC is able to separate different glycans with the same number of sialic acids. For example, the bisialo biantennary glycopeptide (peak 6, Figure 4b) elutes earlier than the bisialo triantennary glycopeptide (peak 7, Figure 4b). Also, two glycopeptides with the same mass elute in two separate peaks (peaks 3 and 4, Figure 4b) which suggests that two monosialo biantennary structures but with different branching (MAN(α1–3) or MAN(α1–6)) can be separated by HILIC. Therefore, HILIC not only holds great promise for further research on separation of oligosaccharide chromatography, it can also provide a means for quantitative monitoring of site-specific glycosylation in process monitoring of proteins produced for pharmaceutical production.

3.3 Analysis of Protein Phosphorylation and Acetylation by Hydrophilic-interaction Liquid Chromatography

Histones belong to a family of proteins which are thought to play an important role in organizing and modulating the structure of eukaryotic chromatin. They become phosphorylated, acetylated and ubiquitinated and it is believed that reversible posttranslational modification of histones are associated with dynamic changes in chromosome packaging during cell cycle. To be able to study the relationship between histone phosphorylation and chromatin structure/function, a system is needed in which precisely defined phosphorylated H1 histones and DNA sequences can be reconstituted. A number of possible strategies exist to resolve and isolate differentially phosphorylated proteins, such as 2-D polyacrylamide gel electrophoresis (PAGE), ion-exchange chromatography, or capillary electrophoresis. Because phosphorylation increases the polarity of a protein due to the presence of the phosphate group, HILIC offers an alternative to the above mentioned separations.

In their studies to separate modified histone H1 subtypes into single components, Lindner et al.[26] developed a two-step high-performance liquid chromatography (HPLC) procedure by combining RPHPLC and HILIC. In a first step, separation of whole linker histones were isolated from exponentially growing Raji cells (a cell line originally derived from patients with Burkitt's lymphoma) and separated into two components, H1.1 and H1.2 (Figure 6a). Capillary zone electrophoresis (CZE) of the H1.1 fraction separated the seemingly pure protein into four distinct components (Figure 6b), presumably consisting of nonphosphorylated (p0), singly- (p1), doubly- (p2) and a small amount of triply- (p3) phosphorylated histone H1. Treatment of the H1.1 pool with alkaline

Peak #	Glycopeptide structure	$(M + H)^+$ Expected mass	$(M + H)^+$ Observed mass	Peak area %
1	T4-GlcNAc-GlcNAc-Man (Fuc) / Man-GlcNAc-Gal-NeuAc, Man-GlcNAc-Gal	4314.2	4315.3	3.2
2	T4-GlcNAc-GlcNAc-Man (Fuc) / Man-GlcNAc-Gal-NeuAc, Man-GlcNAc-Gal	4314.2	4315.5	9.8
3	T4-GlcNAc-GlcNAc-Man (Fuc) / Man-GlcNAc-Gal-NeuAc, Man-GlcNAc-Gal-NeuAc	4605.5	4605.4	45.2
4	T4-GlcNAc-GlcNAc-Man (Fuc) / Man-GlcNAc-Gal-NeuAc, Man (GlcNAc-Gal-NeuAc, GlcNAc-Gal)	4970.8	4968.9	9.9
5	T4-GlcNAc-GlcNAc-Man (Fuc) / Man-GlcNAc-Gal-NeuAc, Man (GlcNAc-Gal-NeuAc, GlcNAc-Gal-NeuAc)	5262.1	5266.1	20.5
6	T4-GlcNAc-GlcNAc-Man (Fuc) / Man (GlcNAc-Gal-NeuNAc, GlcNAc-Gal-NeuNAc), Man (GlcNAc-Gal-NeuNAc, GlcNAc-Gal)	5627.4	5628.3	6.1
7	T4-GlcNAc-GlcNAc-Man (Fuc) / Man (GlcNAc-Gal-NeuNAc, GlcNAc-Gal-NeuNAc), Man (GlcNAc-Gal-NeuNAc, GlcNAc-Gal-NeuNAc)	5918.7	5915.8	5.3

(a)

Peak #	Glycopeptide structure	$(M + H)^+$ Expected mass	$(M + H)^+$ Observed mass	Peak area %
1	T97-GlcNAc-GlcNAc-Man / Man, Man (Man, Man)	2740.6	2741.9	2.3
2	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal, Man-GlcNAc-Gal	3147.0	3148.2	3.5
3	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal-NeuAc, Man-GlcNAc-Gal	3438.3	3440.2	3.1
4	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal-NeuAc, Man-GlcNAc-Gal	3438.3	3439.3	30.7
5	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal-NeuAc, Man (GlcNAc-Gal, GlcNAc-Gal)	3803.6	3805.3	7.2
6	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal-NeuAc, Man-GlcNAc-Gal-NeuAc	3729.6	3726.2	39.8
7	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal-NeuAc, Man (GlcNAc-Gal-NeuAc, GlcNAc-Gal)	4094.9	4093.5	5.6
8	T97-GlcNAc-GlcNAc-Man / Man-GlcNAc-Gal-NeuAc, Man (GlcNAc-Gal-NeuAc, GlcNAc-Gal-NeuAc)	4386.2	4388.0	7.9

(b)

Figure 5 Oligosaccharide structures for Asn_{25} (a) and Asn_{97} (b) from recombinant human γ-interferon which were separated by HILIC into the peaks shown in Figure 4. The structures were identified by measuring the masses of the peptides present in the individual peaks and comparing them with the expected masses for a given oligosaccharide structure. (Reproduced from Zhang and Wang[25] with permission of Elsevier Science © 1998.)

phosphatase followed by CZE showed that the p1, p2, and p3 components were converted to p0 and therefore indeed represented singly-, doubly-, and triply phosphorylated H1 histones. CEX chromatography of the H1.1 pool on a PolyCAT A column (a weak CEX material with a bonded coating of polyaspartic acid) at pH 3.0 with a linear gradient of sodium perchlorate was unable to resolve the differentially phosphorylated histones into single components. Since phosphorylation of histone H1 takes place exclusively in hydrophilic *N*- and *C*-terminal regions

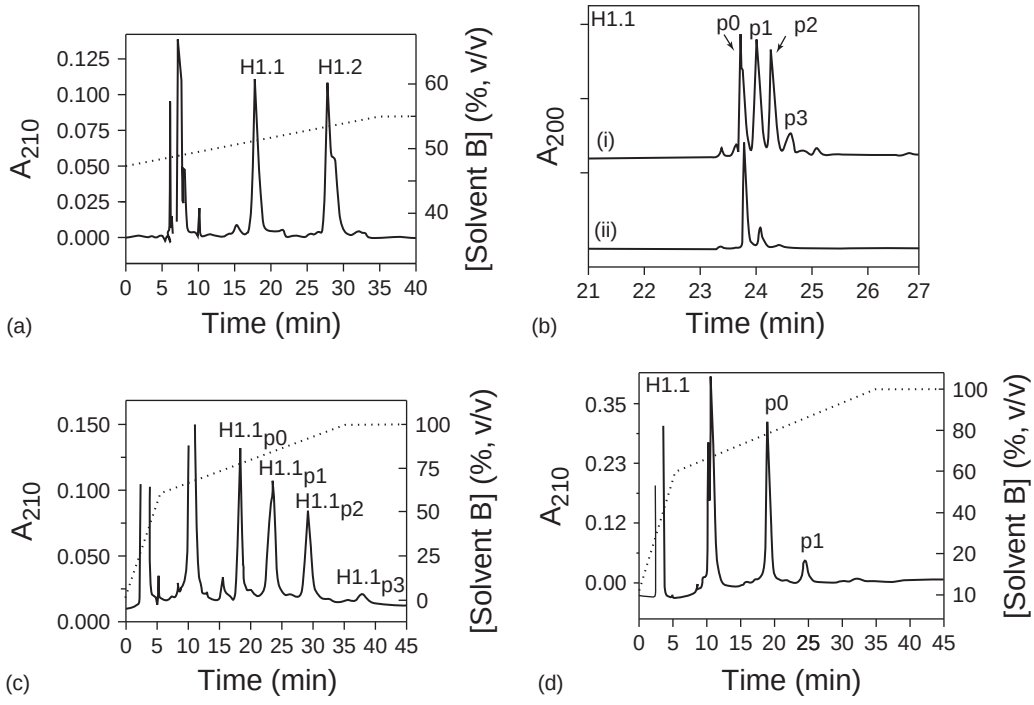

Figure 6 (a) Separation of histones from Raji cells by RPLC. Column: C_4 reversed-phase Nucleosil 300-5 (8 × 250 mm, 5 µm particle size, 300 Å pore size). Solvent A: 0.1% TFA, solvent B: 70% acetonitrile, 0.1% TFA. Flow: 1.5 ml min^{-1}. Gradient: linear from 53 : 47 solvent A : solvent B to 45 : 55 solvent A : solvent B in 35 min. Approximately 200 µg total protein was applied. (b) CZE of histone H1.1 before (i) and after (ii) treatment with alkaline phosphatase. p0, p1, p2, and p3 denote unphosphorylated, singly-, doubly-, and triply phosphorylated H1.1 histones. (c) HILIC of multiply phosphorylated histones H1.1 isolated by RPLC (s. panel A). The column used was a PolyCAT A (4.6 × 200 mm) which was developed at 1 ml min^{-1} with a two step gradient from solvent A (10 mM triethylamine-methanephosphonic acid, pH 3.0, containing 70% acetonitrile) to solvent B (10 mM triethylamine-methanephosphonic acid, pH 3.0, 70% acetonitrile, 1 M $NaClO_4$). Solvent B was increased from 0 to 60% B in 5 min and from 60% to 100% B in 30 min. Protein load was between 20–150 µg. p0, p1, p2, and p3 denote unphosphorylated, singly-, doubly-, and triply phosphorylated H1.1 histones. (d) HILIC of H1.1 histones dephosphorylated with alkaline phosphatase. Chromatography was performed as described in (c). p0 and p1: unphosphorylated and monophosphorylated histone H1.1. (Reproduced from Lindner et al.[26] with permission of Elsevier Science © 1997.)

flanking a hydrophobic central domain, the authors speculated that HILIC should be better suited to exploit interactions between the hydrophilic domains of the protein and the stationary phase. Indeed, increasing the acetonitrile concentration from 40% to 70% to promote hydrophilic interactions with the PolyCAT A support resulted in the separation of the H1.1 pool into four distinct peaks, when elution was carried out with a salt gradient at pH 3.0 (Figure 6c). Again, the use of $NaClO_4$ and triethylamine-methanephosphonic acid to adjust the pH of the solvent system is dictated by the use of such high concentrations of acetonitrile in the solvent system. Treatment of the H1.1 pool by alkaline phosphatase followed by HILIC led to the disappearance of the doubly- and triply phosphorylated H1.1 with some residual mono phosphorylated H1.1 due to incomplete removal of phosphate by the enzyme (Figure 6d). The order of elution of these differentially phosphorylated species during HILIC follows the degree of phosphorylation with the nonphosphorylated protein eluting first, followed by the singly-, doubly-, and triply-phosphorylated protein, therefore following the order of hydrophilicity.

Similar observations with respect to the separation of acetylated histone H4 were made from the same team.[27] Acetylation of the *N*-terminal domains of the so-called core histones H2A, H2B, H3, and H4 have been described, but their biological implications are not well understood. To be able to study nucleosome and chromatin structure–function relationships of histone acetylation, chromatographic procedures have to be developed which yield defined acetylated histones. RPH-PLC has proved to be successful in separating histones, but failed to resolve posttranslationally modified histones. Therefore, initial separation of whole histones is first carried out to separate them into linker histone subtypes (H1o, H1b, H1a, H1e + H1d, H1c) and core histones (H2B, H2A, H4, and H3) (Figure 7a). Among them, histone H4 has been shown to occur in multiply acetylated forms. Although histone H4 appeared to be pure based on the peak shape, the protein could be further separated

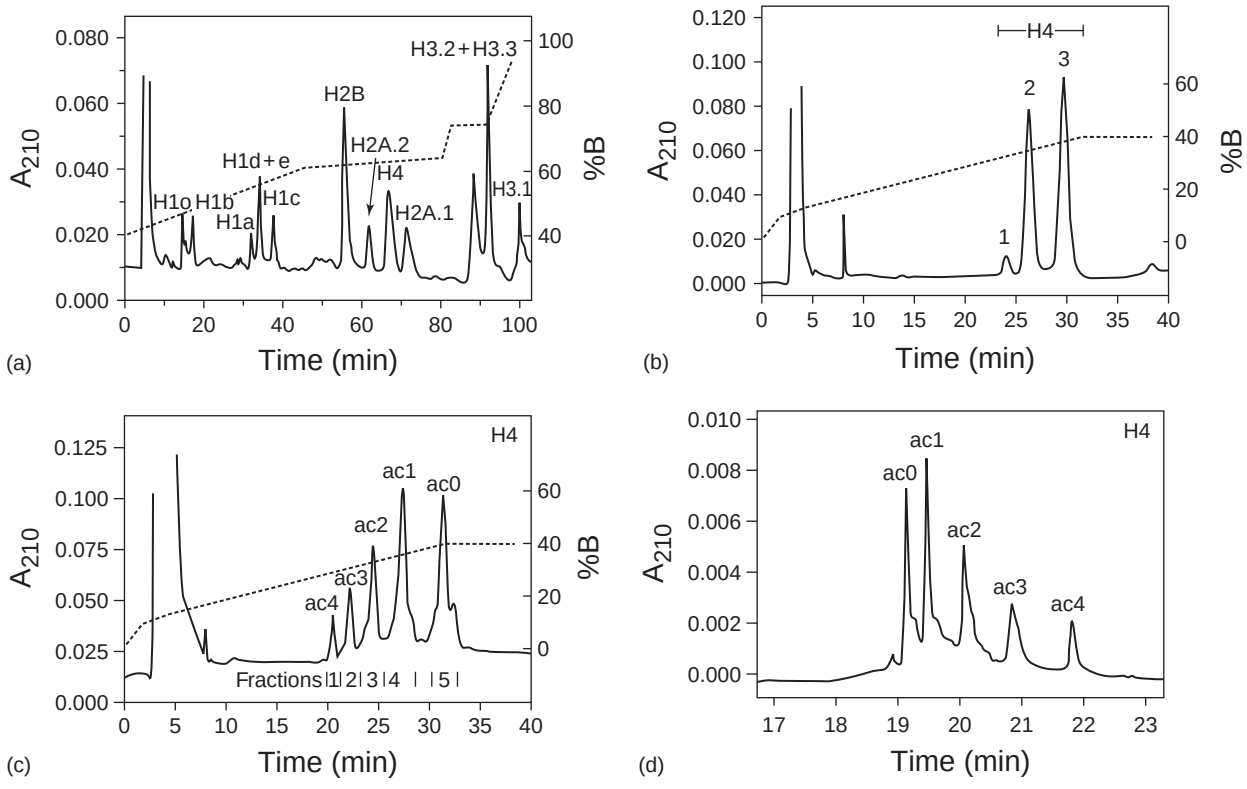

Figure 7 (a) RPLC of whole histones from Friend erythroleukaemic cells. Column: C_4 Nucleosil 300-5 (8 × 125 mm, 5 µm particle size, 300 Å pore size). Solvents: A: 0.1% TFA, 10% ethylene glycol monomethyl ether, B: 70% acetonitrile in A. The multistep gradient was developed at 1 ml min^{-1} according to the following program: from 40 to 61% (45 min), from 61 to 64% (35 min), from 64 to 74% (2 min), isocratic at 74% for 9 min and from 74 to 100% in 10 min. 250 µg total protein was injected. (b) HILIC of the histone H4 pool isolated by RPLC (panel A). The column used was a SynChropak CM300 (4.6 × 250 mm, 6.5 µm particle size, 300 Å pore size) developed at 1 ml min^{-1} with the following solvents: A: 15 mM triethylamine phosphate, pH 3.0, containing 70% acetonitrile, solvent B: solvent A containing 0.68 M $NaClO_4$. The concentration of solvent B was increased from 0 to 10% B during 2 min and from 10 to 40% B during 30 min. (c) HILIC of H4 histones from butyrate treated Friend erythroleukaemic cells. Column and chromatographic conditions as in panel B. ac0–ac4 denote nonacetylated, singly-, doubly-, triply-, and quadruply-acetylated histone H4. (d) CZE of hyperacetylated histone H4 from butyrate treated Friend erythroleukaemic cells. ac0–ac4 are nonacetylated, singly-, doubly-, triply-, and quadruply-acetylated histones H4. (Reproduced from Lindner et al.[27] with permission of Elsevier Science © 1996.)

into distinct components by capillary electrophoresis. Ion-exchange chromatography, although potentially well suited to resolve proteins whose net positive charge decreases upon acetylation, was unable to separate the various acetylated forms of histone. Since upon acetylation, not only the net positive charge decreases, but also the hydrophilicity, HILIC was investigated as a possible separation method for the differentially acetylated histone H4. Elution of H4 on a SynChropak CM300 CEX column at pH 3.0 with a salt gradient in the presence of acetonitrile successfully resolved the proteins into three different species consisting of the di-acetylated, mono-acetylated and non-acetylated histone. The decrease of hydrophilicity brought about by acetylation is in good agreement with the observed order of elution, with the most hydrophilic species (non-acetylated histone) eluting last, preceded by the mono-, and di-acetylated histone, respectively (Figure 7b). Interestingly, hyperacetylation brought about by butyrate treatment of Friend erythroleukaemic cells could easily be followed by HILIC. Butyrate treatment increased modification of histone H4 by the addition of two more acetate groups which could all be separated by HILIC (Figure 7c). Furthermore, the extent of modification occurring in vivo can be monitored with the HILIC method, since a close correlation of the observed peak heights of the various acetylated proteins exists between HILIC and CZE (compare Figure 7c and d). It should also be noted that the order of elution of the hyperacetylated H4 histones is reversed between HILIC and CZE. Since CZE separation is mainly depending on charge effects, elution of the non-acetylated histone occurs first, followed by the increasingly acetylated forms which contain decreasing net positive charges due to increasing acetylation.

3.4 Removal of Detergent from Membrane Proteins

Structural characterization of membrane proteins is often hampered by their hydrophobic nature and tendency to aggregate. A particularly useful detergent in this respect is sodium dodecylsulfate (SDS). Although of strongly denaturing character, SDS is often the last resort to solubilize extremely hydrophobic proteins. The use of SDS however, poses particular problems when primary structure determinations are to be carried out: it inhibits enzymatic digestion of SDS-coated proteins, and residual detergent severly affects subsequent fragment separation by RPHPLC.[28] In this context, it is interesting to note that a procedure for detergent removal has been suggested by Simpson et al.[29] which can be regarded as an early type of hydrophilic interaction chromatography. By investigating a variety of reversed-phase supports, it was noted proteins bind to certain reversed-phase packings at high concentrations of organic solvent. Bound proteins can then be eluted with a decreasing gradient of organic solvent. The authors attributed this finding due to the presence of residual silanol groups which create a hydrophilic environment for proteins to bind. Such behavior is typical for proteins whereas small molecular weight compounds are not retained at such concentrations of organic solvent. This offers a convenient method for desalting proteins into volatile buffer systems for subsequent structural characterization such as automated Edman degradations and MS.

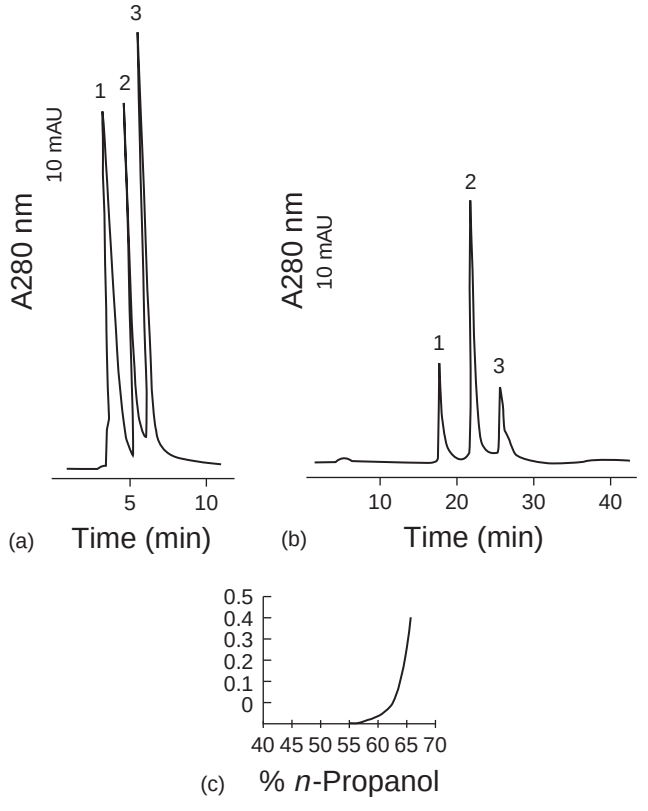

Figure 8 Chromatography of polypeptides on PolyHydroxyethyl A. (a) Gel filtration of ovalbumin (1, 25 µg), synthetic decamer GKEGLRFDMKGA (2, 25 µg), and tryptophan (3, 5 µg). The column (4.6 × 200 mm) was packed with PolyHydroxyethyl A (5 µm particle size, 200 Å pore size). The column was developed at 0.5 ml min^{-1} in 50 mM formic acid containing 20% *n*-propanol. (b) Gradient elution of a mixture of ovalbumin (1, 10 µg), cytochrome c (2, 10 µg), and bovine serum albumine (3, 10 µg) at 0.5 ml min^{-1} with a step gradient from 75% *n*-propanol, 50 mM formic acid (solvent A) to 50 mM formic acid (solvent B) according to the following program: 5 min 0% B, 5–10 min 20% B, 10–15 min 40% B, 15–20 min 60% B, 20–25 min 80% B, and 25–30 min 100% B. (c) Dependence of capacity factor k′ for cytochrome c on the composition of the mobile phase. The effluent was monitored at 280 nm.

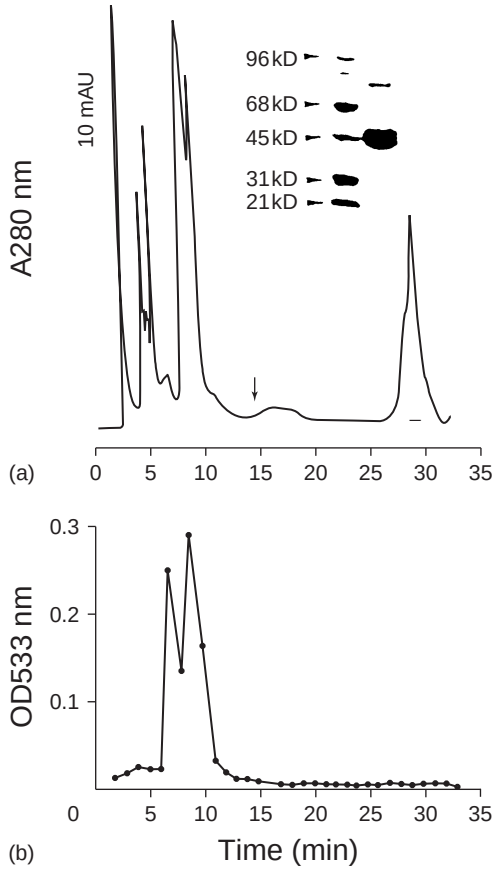

Figure 9 HILIC of electroeluted mitochondrial 45 kDa outer membrane protein. (a) 20 µg of electroeluted 45 kDa outer membrane protein was injected onto a PolyHydroxyethyl A column which had been equilibrated in 70% *n*-propanol, 50 mM formic acid. Elution of the protein was carried out as described in Figure 8(b). The inset shows an SDS polyacrylamide gel with molecular weight standards (left) and the 45 kDa outer membrane protein eluted from the column (right). (b) Elution of SDS from the PolyHydroxyethyl A column tested with Fuchsin red.

A similar method for detergent removal was developed based on the hydrophilic stationary phase PolyHydroxyethyl A.[30] The selectivity of this material changes markedly with the concentration of the organic modifier used. At organic modifier concentrations below 50%, no interaction occurs between solutes and the stationary phase. Separation in such cases is governed by molecular sieving effects (Figure 8a). Increasing the concentration of the organic modifier above 55% causes retention of solutes to increase dramatically and bound solutes can subsequently be eluted with a decreasing gradient of *n*-propanol (Figure 8b). In an application aimed at identifying import components of the yeast mitochondrial membrane,[31] the PolyHydroxyethyl A support was used to desalt proteins isolated by electroelution. Proteins obtained in such a way contain large amounts of SDS which interfere with subsequent analysis. To remove the detergent, electroeluted 45 kDa outer membrane protein was injected onto PolyHydroxyethyl A column equilibrated in 70% *n*-propanol. Testing the eluate for the presence of SDS with Fuchsin red showed that all of the detergent eluted with the injection front. Elution of the protein was achieved with an 'inverse' gradient. The 45 kDa outer membrane protein was eluted at the end of the gradient and was completely devoid of SDS (Figure 9). Other interfering low molecular contaminants from the electroelution process such as residual stain are also effectively removed. Spectral data obtained on a 30 kDa mitochondrial membrane protein after HILIC desalting showed no ultraviolet absorbing components above 280 nm which would indicate the presence of residual stain (Figure 10b). RPC of the electroluate containing

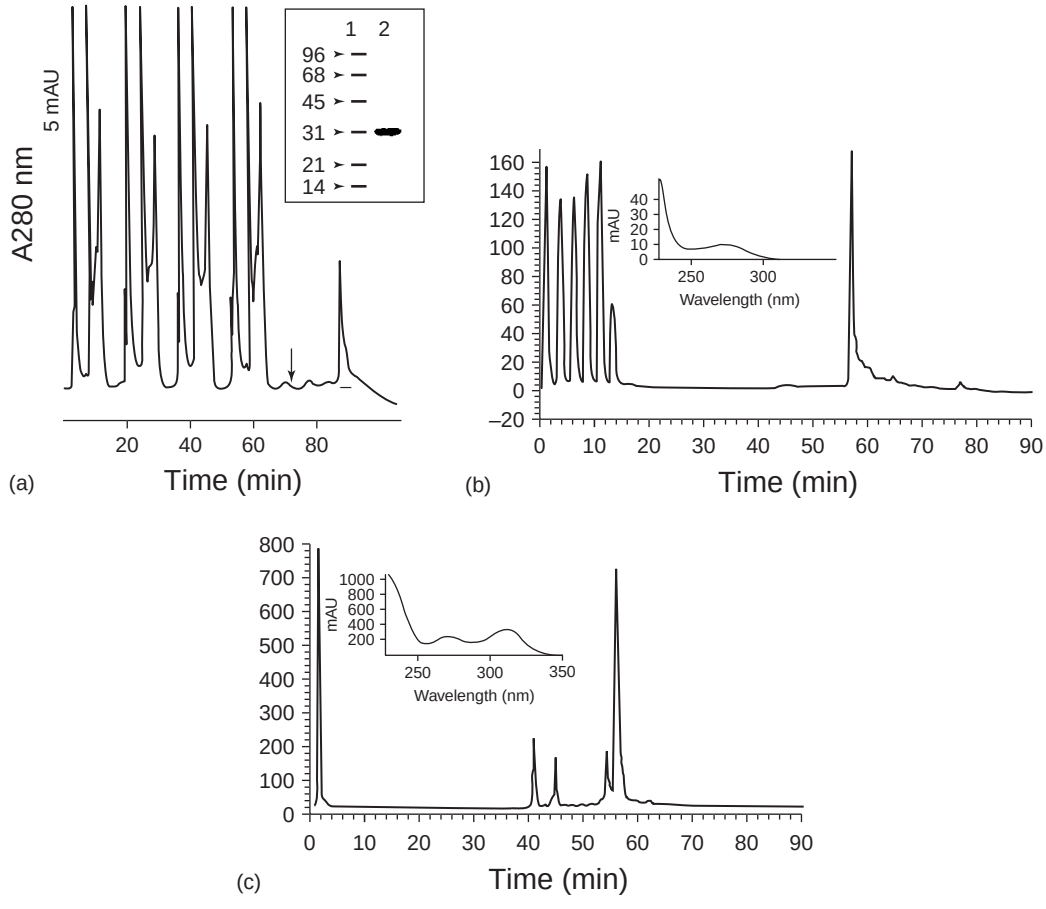

Figure 10 HILIC of an electroeluted 30 kDa membrane protein. (a) An electroeluate containing approximately 5 µg of a 30 kDa mitochondrial membrane protein was applied with multiple 50 µl injections onto a PolyHydroxyethyl A column equilibrated in 70% *n*-propanol, 50 mM formic acid. After the baseline was stabilized, elution of the protein was started with a linear gradient from 70% *n*-propanol, 50 mM formic acid to 50 mM formic acid. The effluent was recorded at 280 nm. The inset shows the SDS polyacrylamide gel of the protein collected from the column. (b) RPLC of the desalted 30 kDa membrane protein. The protein pool from the PolyHydroxyethyl A column was diluted fivefold with 0.1% TFA and repeatedly injected onto a C_4 Aquapore column (2.1 × 30 mm) and eluted with a linearly increasing gradient from 0.1% TFA to 75% acetonitrile, 0.1% TFA in 70 min at a flow rate of 150 µl min^{-1}. Approximately 2 µg protein was injected. (c) The electroeluate containing an equal amount of the 30 kDa protein was injected onto the C_4 Aquapore column without prior desalting by HILIC and eluted from the column under identical conditions. The insets show the absorption spectra of the peaks eluting at 57 min. The arrow marks the elution position of the 30 kDa protein.

the same 30 kDa membrane without HILIC desalting showed that the protein peak was totally obscured by the eluting Coomassie Blue stain (Figure 10c). However, it should be pointed out that, although HILIC and the inverse-gradient mode described by Simpson et al.[29] are very effective in removing SDS from proteins, care should be taken when extremely hydrophobic proteins are chromatographed. The removal of SDS from proteins can lead to aggregation, especially when they are denatured. Whether organic phases such as acetonitrile or *n*-propanol are able to retain the protein in solution must be checked carefully in each application, otherwise severe protein losses can occur.

4 CONCLUSIONS

Hydrophilic interaction chromatography as introduced by Alpert[15] is a versatile chromatographic tool which exploits hydrophilic interactions between solutes and a hydrophilic stationary phase. While RPHPLC is the most commonly used separation method for a variety of biological molecules, HILIC offers distinct advantages for the separation of specific mixtures of compounds. This is particularly true when HILIC is operated in the mixed-mode where separation is based on hydrophilicity/hydrophobicity differences combined with ion-exchange chromatography for the separation of solutes of varying net charge. For instance, separation of synthetic peptides from deletion impurities differing only slightly from the desired target sequence was achieved by HILIC/CEX, but not by RPHPLC.[32] Furthermore, for resolving very complex mixtures, HILIC/CEX and RPHPLC can be ideally interfaced. Due to the high amount of organic modifier required for hydrophilic interactions, compounds resolved by RPHPLC in a first dimension are ideally suited for subsequent HILIC. One disadvantage of HILIC/CEX though, is that solutes are present at relatively high salt concentrations. The need to identify components is great, e.g. by coupling HPLC to MS. In such cases, gelfiltration on PolyHydroxyethyl A columns at low organic modifier efficiently separates salts from components down to a few hundred daltons. Alternatively, the use of volatile mobile phase additives for HILIC/CEX is currently under active investigation.

ABBREVIATIONS AND ACRONYMS

CEX	Cation-exchange
CZE	Capillary Zone Electrophoresis
ESI	Electrospray Ionization
HILIC	Hydrophilic-interaction Liquid Chromatography
HILIC/IEX	HILIC/Ion-exchange
HPLC	High-performance Liquid Chromatography
MALDI/TOF	Matrix-assisted Laser Desorption/ Ionization/Time-of-flight
MS	Mass Spectrometry
PAGE	Polyacrylamide Gel Electrophoresis
RPC	Reversed-phase Chromatography
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RPLC	Reversed-phase Liquid Chromatography
SDS	Sodium Dodecylsulfate
TEAP	Triethylammonium Phosphate
2-D	Two-dimensional

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Chromatography of Membrane Proteins and Lipoproteins • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

REFERENCES

1. A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, *Science*, **274**, 563–567 (1996).
2. J.L. DeRisi, V.L. Iyer, P.O. Brown, *Science*, **278**, 680–686 (1997).
3. L. Wodicka, H. Dong, M. Mittmann, M.H. Ho, D.J. Lockhart, *Nature Biotechnology*, **15**, 1359–1367 (1997).
4. N.C. Hauser, M. Vingron, M. Scheideler, B. Krems, K. Hellmuth, K.D. Entian, J.D. Hoheisel, *Yeast*, **14**, 1209–1221 (1998).
5. P.H. O'Farrell, *J. Biol. Chem.*, **250**, 680–685 (1975).
6. R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, *Anal. Chem.*, **62**, 882–899 (1990).
7. D.C. Stahl, K. Swiderek, M.T. Davis, T.D. Lee, *J. Am. Soc. Mass Spectrom.*, **7**, 532–540 (1996).
8. A. Ducret, I. van Oosten, J.K. Eng, J.R. Yates, R. Aebersold, *Protein Sci.*, **7**, 706–719 (1998).
9. J. Eng, A.L. McCormack, J.R. Yates, *J. Am. Soc. Mass Spectrom.*, **5**, 976–989 (1994).

10. O.N. Jensen, A. Podtelejnikow, M. Mann, *Rapid Commun. Mass Spectrom.*, **10**, 1371–1378 (1996).
11. K. Geveart, J.L. Verschelde, M. Puype, J. Van Damme, M. Goethals, S. De Boeck, J. Vandekerckhove, *Electrophoresis*, **17**, 918–924 (1996).
12. F. Wold, *Annu. Rev. Biochem.*, **50**, 783–814 (1981).
13. R.G. Krishna, F. Wold, in *Methods in Protein Sequence Analysis*, eds. K. Imahori, F. Sakiyama, Plenum Press, New York, 167–172, 1993.
14. Ken Mitchelhill, John Holt Protein Structure Laboratory, St. Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy 3065, Victoria, Australia (*http://www.medstv.unimelb.edu.au/wwwdocs/svimrdocs/MassSpec/deltamassV2.html*).
15. A.J. Alpert, *J. Chromatogr.*, **499**, 177–196 (1990).
16. A.J. Alpert, P.C. Andrews, *J. Chromatogr.*, **443**, 85–96 (1988).
17. B.-Y. Zhu, C. Mant, R.S. Hodges, *J. Chromatogr.*, **548**, 13–24 (1991).
18. D.H.W. Cheng, P.-J. Lee, W.K. Watt, presented at the *7th International Symposium on HPLC of Proteins, Peptides and Polynucleotides, Washington, DC*, abstract 311, 1987.
19. P.C. Andrews, *Pept. Res.*, **1**, 93–99 (1988).
20. D.L. Crimmins, D.W. McCourt, B.D. Schwartz, *Biochem. Biophys. Res. Commun.*, **156**, 910–916 (1988).
21. C.T. Mant, J.R. Litowski, R.S. Hodges, *J. Chromatogr. A*, **816**, 65–78 (1998).
22. N.E. Zhou, C.T. Mant, R.S. Hodges, *Peptide Res.*, **3**, 8–20 (1990).
23. M.I. Aguilar, S. Mougas, J. Boublik, J. Rivier, M.T.W. Hearn, *J. Chromatogr.*, **646**, 53–65 (1993).
24. T.J. Sereda, C.T. Mant, R.S. Hodges, *J. Chromatogr. A*, **695**, 205–221 (1995).
25. J. Zhang, D.I.C. Wang, *J. Chromatogr. B*, **712**, 73–82 (1998).
26. H. Lindner, B. Sarg, W. Helliger, *J. Chromatogr. A*, **782**, 55–62 (1997).
27. H. Lindner, B. Sarg, C. Meraner, W. Helliger, *J. Chromatogr. A*, **743**, 137–144 (1996).
28. A. Bosserhoff, J. Wallach, R. Frank, *J. Chromatogr.*, **437**, 71–77 (1989).
29. R.J. Simpson, R.L. Moritz, E.E. Nice, B. Grego, *Eur. J. Biochem.*, **165**, 21–29 (1987).
30. P. Jenö, P.E. Scherer, U. Manning-Krieg, M. Horst, 'Desalting Electroeluted Proteins with Hydrophilic Interaction Chromatography', *Anal. Biochem.*, **215**, 292–298 (1993).
31. M. Horst, P. Jenö, N. Kronidou, L. Bolliger, W. Oppliger, P. Scherer, U. Manning-Krieg, G. Schatz, *EMBO J.*, **12**, 3035–3041 (1993).
32. C.T. Mant, P.J. Kondejewski, O.D. Cachia, R.S. Monera, R.S. Hodges, *Meth. Enzymol.*, **289**, 426–469 (1997).

Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis

J. Kathleen Lewis, Jing Wei, and Gary Siuzdak
The Scripps Research Institute, La Jolla, USA

1 Introduction to Matrix-assisted Laser Desorption/Ionization Mass Spectrometry 1
1.1 Mechanism of Matrix-assisted Laser Desorption/Ionization 1
1.2 Matrix-assisted Laser Desorption/Ionization Mass Analyzers 2
2 Analysis of Peptides and Proteins by Matrix-assisted Laser Desorption/Ionization 3
2.1 Sample Preparation 4
2.2 Protein Primary Sequence Analysis 5
2.3 Characterizing Protein Modifications (Co- and Posttranslational) 7
2.4 Protein Structure Elucidation 7
3 Applications 9
3.1 Diagnostic 9
3.2 Quantitative Aspects of Matrix-assisted Laser Desorption/Ionization 10
3.3 Characterizing Peptides and Reactions Directly from the Solid Phase 12
4 Conclusion 12
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 13
References 13

Matrix-assisted laser desorption/ionization mass spectrometry (MALDIMS), first introduced in 1988 by Hillenkamp and Karas, has become a widespread analytical tool for peptides, proteins and most other biomolecules. In MALDI (matrix-assisted laser desorption/ionization) analysis, the analyte is first co-crystallized with a large molar excess of a matrix compound, usually an ultraviolet (UV)-absorbing weak organic acid, after which laser radiation of this analyte–matrix mixture results in the vaporization of the matrix which carries the analyte with it. The matrix therefore plays a key role by strongly absorbing the laser light energy and causing, indirectly, the analyte to vaporize. The matrix also serves as a proton donor and receptor, acting to ionize the analyte in both positive and negative ionization modes, respectively. The efficient and directed energy transfer during a matrix-assisted laser-induced desorption event provides high ion yields of the intact analyte and allows for the measurement of compounds with high accuracy and subpicomole sensitivity. The ability to generate such accurate information can be extremely useful for protein identification and characterization. For example, a protein can often be unambiguously identified by the accurate mass analysis of its constituent peptides (produced by either chemical or enzymatic treatment of the sample). This article discusses basic MALDI concepts and instrumentation and focuses on applications in the field of peptides and proteins, specifically on the utility of MALDI in protein identification, protein structural studies, and as a clinical assay.

1 INTRODUCTION TO MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY

MALDIMS, first introduced in 1988 by Hillenkamp and Karas,[1] has become a widespread analytical tool for peptides, proteins, and most other biomolecules (oligonucleotides, carbohydrates, natural products, and lipids). The efficient and directed energy transfer during a matrix-assisted laser-induced desorption event provides high ion yields of the intact analyte, and allows for the measurement of compounds with high accuracy and subpicomole sensitivity.[2–4] This article discusses basic MALDI concepts and instrumentation and focuses on applications in the field of peptides and proteins, specifically on the utility of MALDI in protein identification, protein structural studies, and as a clinical assay.

1.1 Mechanism of Matrix-assisted Laser Desorption/Ionization

MALDI provides for the nondestructive vaporization and ionization of both large and small biomolecules (Figure 1). In MALDI analysis, the analyte is first co-crystallized with a large molar excess of a matrix compound, usually a UV-absorbing weak organic acid, after which laser radiation of this analyte–matrix mixture results in the vaporization of the matrix which carries the analyte with it. The matrix therefore plays a key role by strongly absorbing the laser light energy and causing, indirectly, the analyte to vaporize. The matrix also serves as a proton donor and receptor, acting to ionize the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

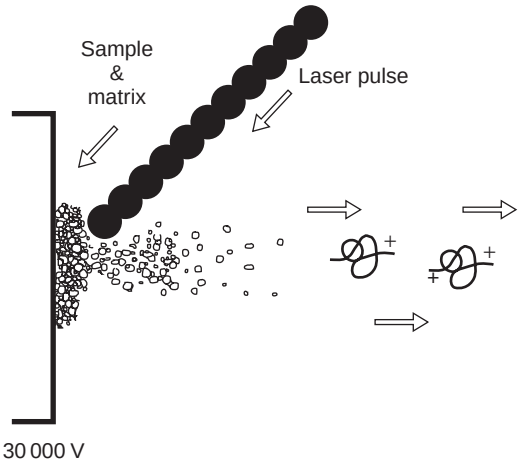

Figure 1 MALDI.

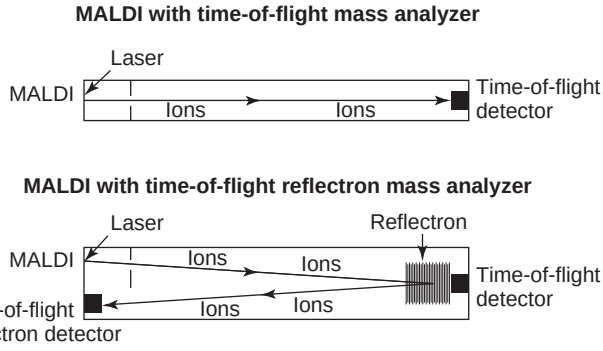

Figure 2 MALDI with TOF and TOF reflectron.

analyte in both positive and negative ionization modes, respectively.[5]

Several theories have been developed to explain desorption of large molecules by MALDI. The thermal-spike model[6] proposes that the matrix molecules sublime from the surface as a result of local heating at low laser fluence, but above a certain laser intensity, a rapid rise in desorption efficiency is observed. The ejection of intact molecules is attributed to poor vibrational coupling between the matrix and analyte which leads to a bottleneck in the energy transfer from the matrix to the internal vibrational modes of the analyte molecule. The pressure pulse theory[7] proposes that a pressure gradient is created normal to the surface and desorption of large molecules might be enhanced by momentum transfer from collisions with fast-moving matrix molecules.

It is generally thought that ionization occurs through proton transfer or cationization. The ionization depends critically on the matrix–analyte combination, but is not critically dependent on the number of acidic or basic groups of the analyte.[8] This suggests that a more complex interaction of analyte and matrix, rather than simple acid–base chemistry, is responsible for ionization.

1.2 Matrix-assisted Laser Desorption/Ionization Mass Analyzers

There are three types of mass analyzers typically used with the MALDI ionization source: a linear time-of-flight (TOF), a TOF reflectron, and a Fourier transform mass analyzer (Figure 2). The linear TOF mass analyzer is the simplest of the three devices and has enjoyed a renaissance with the invention of MALDI. TOF analysis is based on accelerating a set of ions to a detector where all of the ions are given the same amount of energy. Because the ions have the same energy, yet a different mass, the ions reach the detector at different times. The smaller ions reach the detector first because of their greater velocity while the larger ions take longer owing to their larger mass. Hence, the analyzer is called TOF because the mass is determined from the ions' time of flight. The arrival time at the detector is dependent upon the mass, charge, and kinetic energy (KE) of the ion. Since KE is equal to $1/2\ mv^2$ or velocity $v = (2\mathrm{KE}/m)^{1/2}$, ions will travel a given distance, d, within a time, t, where t is dependent upon their mass-to-charge ratio (m/z).

The TOF reflectron combines TOF technology with an electrostatic analyzer, the reflectron. The reflectron serves to increase the amount of time, t, ions need to reach the detector while reducing their KE distribution, thereby reducing the temporal distribution Δt. Since resolution is defined by the mass of a peak divided by the width of a peak or $m/\Delta m$ (or $t/\Delta t$ since m is related to t), increasing t and decreasing Δt results in higher resolution. This increased resolution, however, often comes at the expense of sensitivity and a relatively low mass range, typically $<10\,000\ m/z$. One innovation that has had a dramatic effect on increasing the resolving power of MALDI/TOF instruments has been delayed extraction (DE). DE is a relatively simple means of cooling the ions (and possibly focusing them) immediately after the MALDI ionization event. In traditional MALDI instruments the ions were accelerated out of the ionization source as soon as they were formed; however, with DE the ions are allowed to "cool" for ~150 ns before being accelerated to the analyzer (Figure 3). This cooling period generates a set of ions with a much lower KE distribution, ultimately reducing the temporal spread of ions once they enter the TOF analyzer. Overall, this results in increased resolution and accuracy. Incidentally, the benefits of DE significantly diminish with higher-molecular-weight proteins (>30 000 Da).

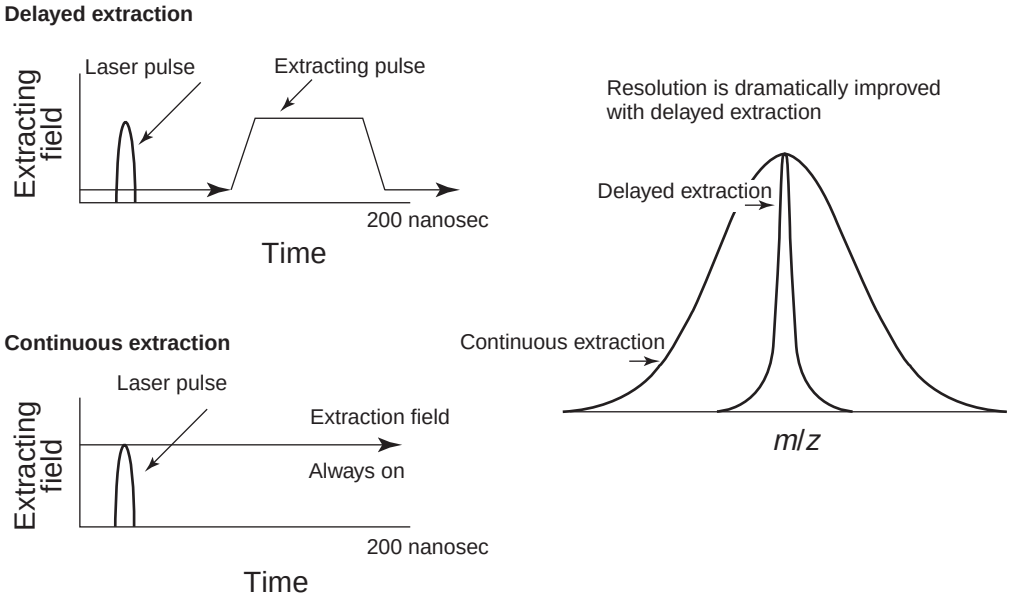

Figure 3 DE versus continuous extraction.

MALDIMS is most commonly combined with the TOF mass analyzers. However, their somewhat limited resolution (10^2–10^4) results in accuracy on the order of 0.2% to a high of 0.005% (with internal calibrant). Alternatively, MALDI instruments are also being coupled to ultrahigh-resolution ($>10^5$) mass analyzers such as the Fourier transform ion cyclotron resonance (ICR) mass analyzer. First introduced in 1974 by Comisarow and Marshall,[9] Fourier transform mass spectrometry (FTMS) is based on the principle of a charged particle orbiting in the presence of a magnetic field. While the ions are orbiting, a radio frequency (RF) signal is used to excite them and as a result of this RF excitation, the ions produce a detectable image current. The time-dependent image current can then be Fourier-transformed to obtain the component frequencies of the different ions which correspond to their m/z. FTMS has become an important research tool offering high accuracy with errors as low as ±0.001% (ppm accuracy). Different types of MALDI mass analyzers are compared in Table 1.

2 ANALYSIS OF PEPTIDES AND PROTEINS BY MATRIX-ASSISTED LASER DESORPTION/IONIZATION

The utility of MALDI for protein and peptide analyses lies in its ability to provide highly accurate molecular-weight information on intact molecules. The ability to generate such accurate information can be extremely useful for protein identification and characterization. For

Table 1 General comparison of MALDI mass analyzers

	TOF	TOF reflectron	FTMS
Accuracy	0.05–0.2% (50–200 ppm)	0.03% (30 ppm)	0.005% (5 ppm)
Resolution	2000	10 000	100 000
m/z range	>300 000	10 000	10 000
Tandem MS	MS	MS^2	MS^4
Scan speed	ms	ms	s
Comments	Highest mass range. Very fast scan speed. Simple design. Low cost.	Lower sensitivity than TOF. Good resolving power. Limited m/z range. Very fast scan speed. Simple design.	Capable of high resolution and exact mass measurements. Well-suited for tandem MS. Instrumentation is expensive, large. Requires high vacuum ($<10^{-7}$ torr). Requires superconducting magnet.

MS, mass spectrometry.

example, a protein can often be unambiguously identified by the accurate mass analysis of its constituent peptides (produced by either chemical or enzymatic treatment of the sample). Furthermore, protein identification can also be facilitated by analysis of the protein's proteolytic peptide fragments in the gas phase; fragment ions generated inside MALDI mass spectrometers via collision-induced dissociation (CID) yield information about the primary structure and modifications. Tandem mass spectrometry (*n*th series) (MS^n) experiments, previously allowed with quadrupole and ion trap mass spectrometers are now attainable with MALDI sources (known as postsource decay or PSD).

While the MALDI mass spectrometer is a powerful tool for the accurate mass determination of peptide mixtures, obtaining accurate mass measurements is highly dependent upon the sample and the sample preparation. Acquiring optimum MALDI data depends on the choice of suitable matrices and solvents, the functional and structural properties of the analyte, sample purity, and how the sample is prepared on the MALDI sample plate. Because the solvent or matrix is the medium by which the analyte will be transported to the gas phase and provides the conditions that make ionization possible, contamination of the sample with excessive salt (>10 mM) will affect these conditions and lead to reduced sensitivity. This is also true of chaotropic agents, including urea and guanidinium salts, and solvents like dimethyl sulfoxide and glycerol. Dialysis and reversed-phase liquid chromatography (RPLC), or exchange chromatography are useful methods for purifying samples of such contaminants prior to mass spectral analysis.

2.1 Sample Preparation

For peptides and proteins, the standard matrices[1,10–13] are α-cyano-4-hydroxycinnamic acid **(1)** (α-cyano or CCA), 3,5-dimethoxy-4-hydroxycinnamic acid **(2)** (sinapinic acid or SA), and 2,5-dihydroxybenzoic acid **(3)** (DHB). CCA **(1)** is mainly used for peptides, glycopeptides and small proteins. SA **(2)** is commonly used for both peptide and protein analysis, and DHB **(3)** is used for glycopeptides, glycoproteins, small proteins, and oligonucleotides (<10 bases). In the analysis of protein digests (peptide mapping), the choice of matrix plays an important role in the observation of the peptides produced. Some matrices can complement each other with respect to obtaining complete protein sequence coverage, for instance peptide mapping with CCA can yield more low-mass peptide ions (<2500 Da), while SA may provide better coverage for the higher-mass peptides (>2500 Da).

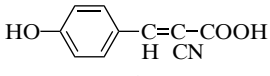

(1)

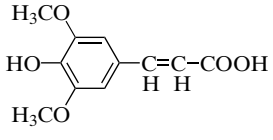

(2)

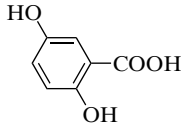

(3)

Sample–matrix preparation procedures greatly influence the quality of MALDI mass spectra of peptides/proteins. Among the variety of reported preparation methods, the dried-droplet method is the most frequently used. In this case, a saturated matrix solution is mixed with the analyte solution, giving a matrix-to-sample ratio of about 5000 : 1. An aliquot (0.5–2.0 μL) of this mixture is then applied to the sample target where it is allowed to dry. Below is one specific example of the dried-droplet method:

- Pipet 0.5 μL of sample to the sample plate (preferable concentration ~ 1.0 pmol μL^{-1}).
- Pipet 0.5 μL of matrix to the sample plate.
- Mix the sample and matrix by drawing in and out of the pipette.
- Allow to air dry.
- For peptides, small proteins, and most compounds a saturated solution of CCA **(1)** in 50 : 50 H_2O : acetonitrile (ACN) with 0.1% TFA (trifluoroacetic acid) is used.
- For glycopeptides/proteins and small compounds a saturated solution of DHB **(3)** in 50 : 50 H_2O : ACN with 0.1% TFA is used.
- For proteins and other large molecules a saturated solution of SA **(2)** in 50 : 50 H_2O : ACN with 0.1% TFA is used.

Alternatively, samples can be prepared in a stepwise manner. In the thin-layer method,[14,15] a matrix homogeneous "film" is formed on the target first, and the sample is then applied and absorbed by the matrix. This method yields good sensitivity, resolving power, and mass accuracy. Similarly, in the thick-layer method,[16] nitrocellulose (NC) is used as matrix additive and once a uniform NC-matrix layer is obtained on the target, the sample is applied. This preparation method suppresses alkali adduct formation and significantly increases the detection sensitivity, especially for peptides and proteins extracted from gels. The sandwich method[16] is another variant in this category. A thin layer of matrix crystals is prepared as in the thin-layer method, followed by the subsequent addition of droplets of aqueous TFA, sample and matrix.

2.2 Protein Primary Sequence Analysis

Amino acid sequence information can be obtained by one of several methods using MALDIMS. The first is called protein mass mapping which consists of the site-specific enzymatic or chemical degradation of a protein followed by mass spectrometric analysis of the released peptides. Owing to the complexity of mixtures generated during proteolysis, MALDI/TOF MS is most ideally suited for such analyses. The traditional analytical methods used to characterize the released peptides consist of high-performance liquid chromatography (HPLC) or gel electrophoresis followed by N-terminal (Edman) sequencing and/or amino acid analysis. However, such methods are considerably more time-consuming and in some cases are not capable of separating individual peptides because of their low resolution. Consequently, high-resolution MALDI/TOF analysis can provide for a more rapid, accurate, and highly sensitive analysis of these complex peptide mixtures.

The purpose of protein mass mapping is not to sequence the entire protein but instead to generate information that can be used in protein database searches either to identify the protein or determine whether the protein is novel. Protein identification via database searching is facilitated by the accurate m/z values of the digest fragments, the specificity of the enzyme used, and the accurate m/z of the intact protein. Entire scientific conferences have been devoted to this approach which has been coined ''proteomics.'' Table 2 previews some of the protein databases available on the internet.

Typically, the type of protease used in the mass mapping approach is an endoprotease, a protease which cleaves internal residues. An extended version of this method follows the endoproteolytic digestion with an exoprotease, which cleaves successively from either the n-terminal or c-terminal side of a peptide or protein. The result is a mass spectrum with peptides differing by one amino acid residue. The mass difference between ion signals determines the identity of the released amino acid and typically provides a sequence tag on the order of 1–12 amino acids. This method further increases the possibility of a database ''hit'' by providing one more piece of valuable information, a partial sequence tag. Figure 4 illustrates this approach.

Another MALDI-based approach that provides amino acid sequence information on peptides relies on the fragmentation of the ions in the field-free drift region (between the ion source and the detector) of the TOF mass analyzer. This fragmentation, referred to as PSD, is due to collisions between analyte ions and neutral matrix molecules or residual gas molecules during the desorption and acceleration stage. Initially these fragments have the same velocities as their precursor ion. Consequently, in a linear TOF mass analyzer these fragments arrive at the detector at the same time as their precursor ion and are therefore not represented in the mass spectrum. In a reflectron TOF mass analyzer, however, these fragments penetrate the reflectron at different depths and are therefore spatially differentiated, meaning that individual fragments arrive at the reflectron detector at different flight times. As a result, these fragments are represented in the mass spectrum and provide valuable structural information (Figure 4). Amino acid sequencing by PSD has become a popular method for sequencing peptides and is now available on most commercial MALDI reflectron instruments. Figure 5 shows the PSD spectrum of 10 pmol of angiotensin (m/z 1297 Da).

A third approach to obtaining peptide sequence information from MALDI is MS^n within an FTMS analyzer cell. One of the advantages of FTMS, in addition to its high resolution capabilities, is its ability to perform MS^n experiments. Dissociation of the precursor ion can be accomplished by the introduction of a collision gas or

Table 2 Protein databases available on the internet

NCBInr	A nonredundant database compiled by the NCBI by combining most of the public domain databases (ESTs not included).
Swiss Prot	A curated protein sequence database which strives to provide a high level of annotation, such as the description of the function of a protein, its domain's structure, post-translational modifications, variants, etc. This database offers a minimal level of redundancy and high level of integration with other databases.
OWL	A nonredundant composite of four publicly available primary sources: SWISSPROT, PIR, (1-3), GenBank (translation) and NRL-3D. SWISSPROT is the highest priority source, all others being compared against it to eliminate identical and trivially different sequences.
Genpept	Protein translation of Genbank (ESTs not included).
Unknome	A theoretical database used in de novo MS/MS spectral interpretation that is created on-the-fly and contains all amino acid sequence permutations consistent with the parent mass and amino acid composition information contained in an MS/MS spectrum.

NCBI, National Center for Biotechnology Information; EST, expressed sequence tag; MS/MS, tandem mass spectrometry (2nd series).

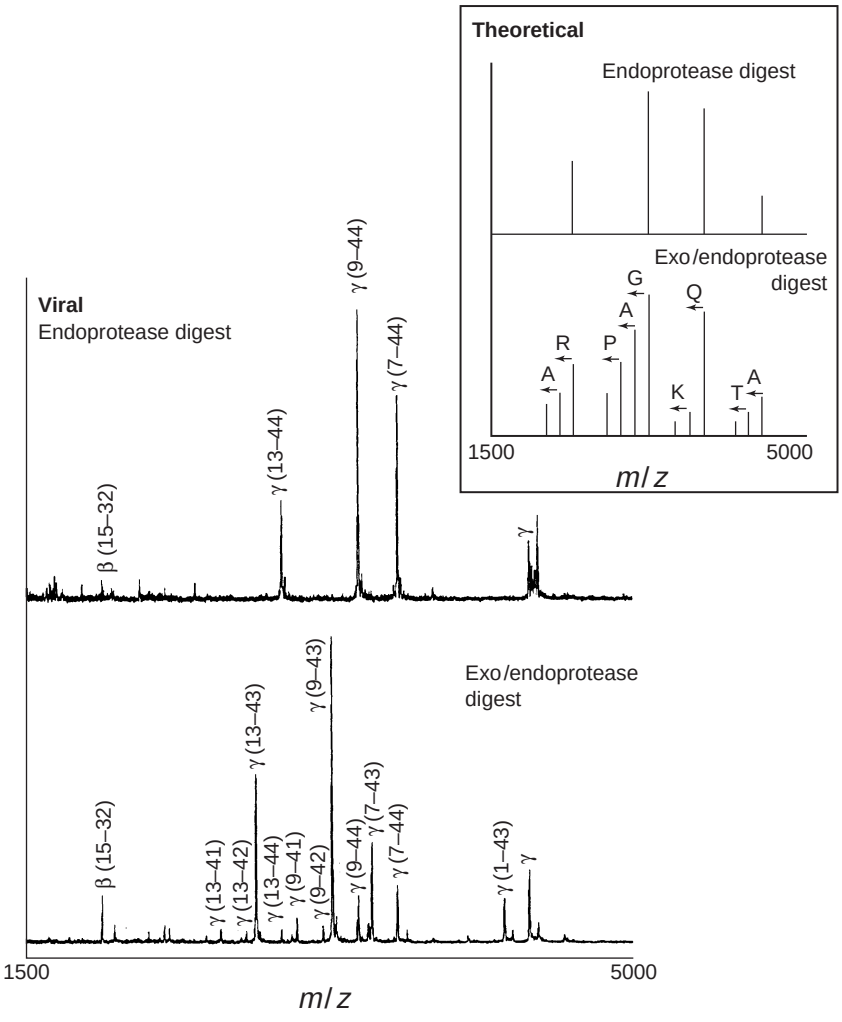

Figure 4 C-terminal ladder sequencing.

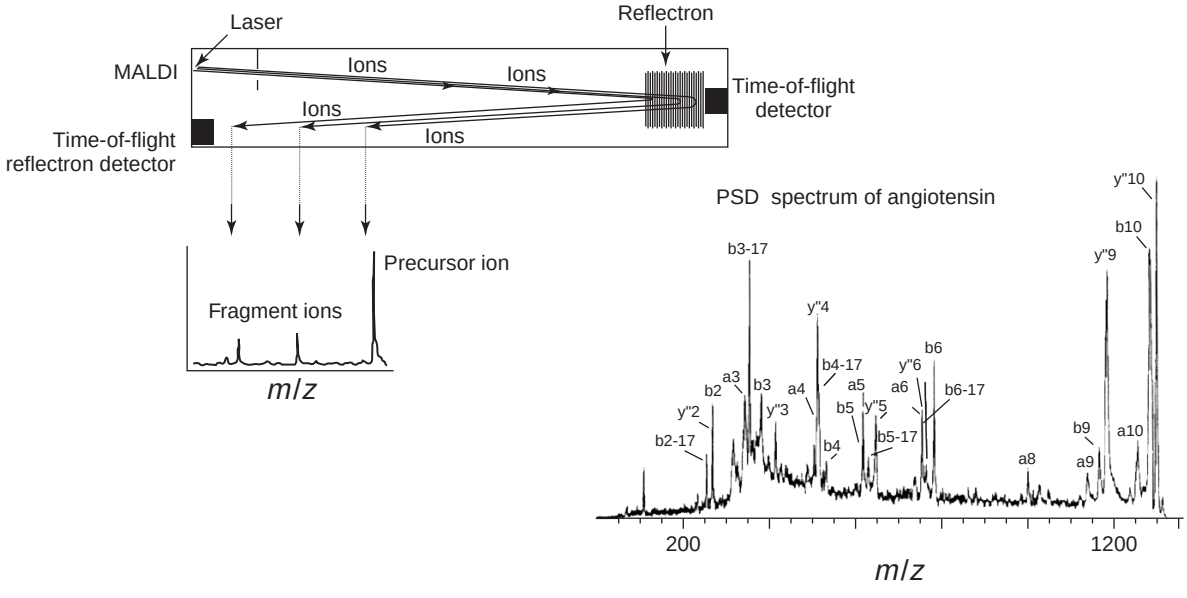

Figure 5 Postsource delay.

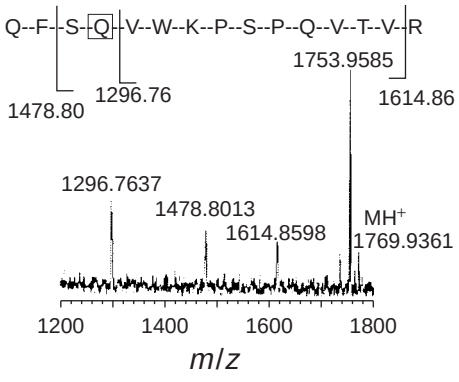

Figure 6 MS/MS with MALDI/FTMS.

by use of pulsed lasers to photodissociate selected ions. Other advantages associated with using FTMS include the simultaneous high mass resolving power of the fragments, the high sensitivity over a wide mass range and the high mass accuracy. The primary disadvantage is the lack of multiply charged species generated in MALDI/FTMS experiments, the lack of which reduces the amount of fragmentation. Figure 6 shows the MALDI/FTMS MS/MS spectrum of a viral peptide obtained from a mutant common cold (rhino) virus. Such analysis provided for the rapid (<20 min) identification of the mutated amino acid with <5 ppm mass accuracy.

2.3 Characterizing Protein Modifications (Co- and Posttranslational)

Protein modifications, including both co- and post-translational modifications (e.g. phosphorylation, sulfation, glycosylation, and N-terminal modifications), are recognized as important means of regulating the cellular distribution and modulating protein functions. For example, protein phosphorylation is known to play a critical role in cell signal transduction pathways.[17,18] Carbohydrates provide various functional, immunological, and structural aspects of glycoprotein. Characterization of protein modifications is an important aspect for analyzing protein structure and function. The utility of MALDI for this purpose relies on its ability to perform accurate mass determination of peptide mixtures, its very high sensitivity and resolving power, and its buffer tolerance.

Typical MALDI protocols for protein modification characterization begin with the enzymatic digestion of the protein sample, resulting in a complex mixture of peptides. When a protein's primary structure is known, the molecular weight of a peptide can be used to corroborate the presence of any modification by its specific molecular-weight difference between the observed mass of peptide and that calculated on the basis of the sequence. This approach is straightforward but has very little specificity and selectivity. The results can be ambiguous if a complicated digestion mixture is presented. This difficulty can be overcome by using MALDI PSD MS/MS to generate and record metastable fragment ions. The aid of a separation technique, usually HPLC, is sometimes needed before the MALDI analysis. This protein modification characterization approach has been employed by several groups to identify and sequence either glycoproteins[19–22] or phosphoprotein.[23,24]

2.4 Protein Structure Elucidation

Proteins in their native state are typically folded into well-defined, three-dimensional structures by relatively weak intramolecular forces (i.e. hydrogen bonds, electrostatic, or hydrophobic interactions). Because the function of the protein relies heavily on this structure, information on higher-order structure (i.e. tertiary protein conformation and quaternary protein–protein interactions) is of fundamental biological importance. There are a variety of spectroscopic methods available to study tertiary and quaternary protein structure, including X-ray crystallography and nuclear magnetic resonance (NMR), along with a variety of other spectroscopic techniques. It should be noted, however, that while these techniques constitute the state of the art as far as structural determination, they can be time-consuming and require deconvolution of complex data and a significant amount of material.

An alternative approach to investigate higher-order protein structure using MS is to probe the surface availability of regions of the molecule using the protein mass mapping method. Proteolytic digestion is dependent on the availability of cleavage sites which is a reflection of the protein structure. Therefore, proteolysis performed in modified solution conditions can sample different protein conformations, and relative differences in these conformations can then be discerned through direct comparison of the resulting mass spectrometric data. For example, digestion of a protein in its denatured or unfolded state will most likely result in the release of additional digest fragments when compared to the digest of the protein in its native or folded state. In the unfolded state, more cleavage sites will be accessible to the protease owing to lack of tertiary structure.

Using this method, Lewis was able to recognize conformational differences (i.e. changes in tertiary structure) in regulatory proteins brought on by the interaction with an effector molecule.[25] The allosteric protein calmodulin (CaM) is known to take on different conformational states depending on the presence or absence of its effector, calcium. Different proteolytic mass maps were observed for CaM, indicating structural differences in the protein depending on the presence of calcium.

The overall tertiary structure of CaM reveals a dumbbell shaped molecule (148 amino acids) with two

globular domains separated by a single long central α-helix (amino acids 65–92). It has been proposed that calcium activates CaM by exposing hydrophobic residues (near the two ends of the central helix) which interact with CaM target proteins. Small-angle X-ray scattering analysis has shown that the central helix acts as a flexible ''tether'' which, upon binding calcium, allows the two globular domains to come together and in effect ''hug'' the target protein, via the hydrophobic regions. Mass maps resulting from digests by trypsin, chymotrypsin, and pepsin all indicated that the protein had undergone a tertiary structural change in the presence of calcium and in doing so hindered the activity of the enzyme in the central helical region. Presumably, the coming together of the two globular domains places certain residues of the central helix, normally cleaved in a calcium-free environment, in a conformation which does not allow the enzyme accessibility.

Thus, in observing qualitative differences between spectra obtained in the absence and in the presence of an effector, the mass mapping approach allows different solution conformations (differences in tertiary structure) of a protein to be identified and which regions of a protein are effected by such conformational changes to be determined. Figure 7 shows the trypsin digests of CaM in the presence and absence of calcium. Comparison of the two mass spectra reveals obvious qualitative differences corresponding to cleavages in the central helical region of the protein. Digestion fragments resulting from cleavages in the central region (residues 65–92) of the molecule are not present in the calcium-free environment.

Investigating protein/protein interactions (quaternary protein structure) has also been pursued with limited

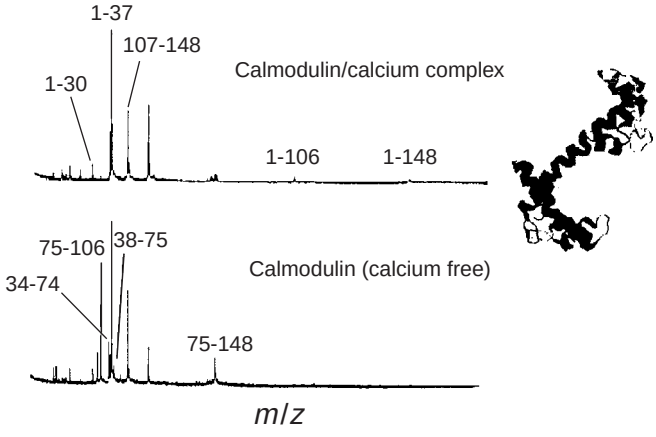

Figure 7 Protein mass mapping.

proteolysis combined with MS, the goal being to determine the topographical regions involved in their interaction.[26] Again, since enzymatic digestion of a protein is dependent on the availability of cleavage sites, it can be assumed that digestion of a protein complex should hinder access of the protease to cleavage sites normally accessible in the digestion of the individual proteins. Toward this end, comparative mass mapping should reveal regions of the protein(s) involved in the formation of the complex. Changes in the susceptibility of cleavage sites toward proteases can arise from several physical effects as a result of the complex formation. One is a direct consequence of the formation of the complex, the steric hindrance, and another is an indirect result of the formation of the complex, a conformational change, that may alter the accessibility of not only

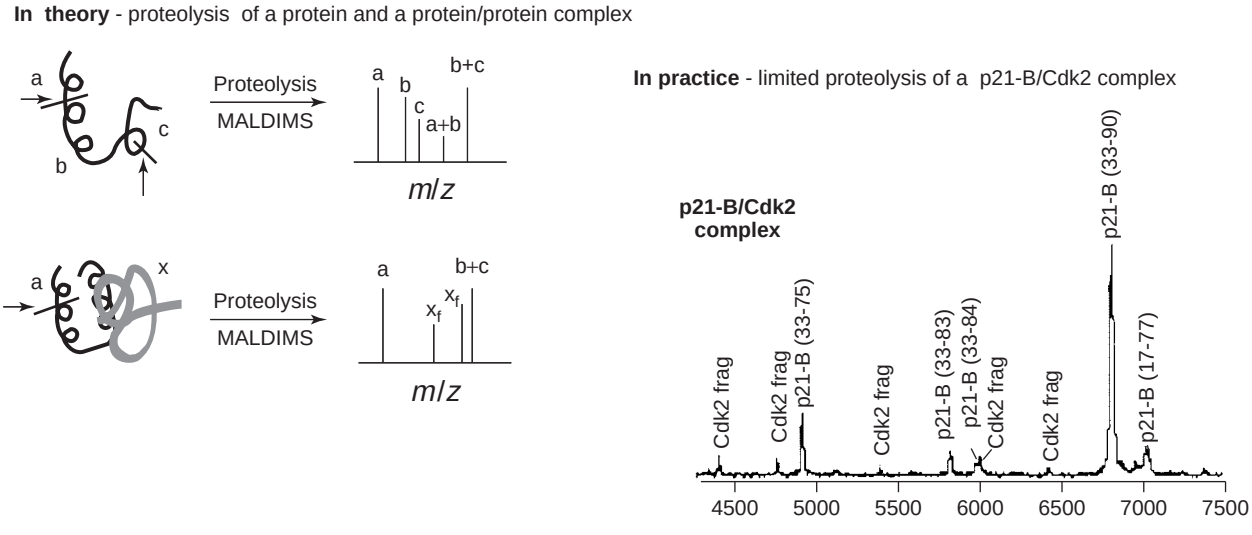

Figure 8 Protein–protein interactions with mass mapping. The symbols represent a library of potential ligands for the receptor IgG.

the sites involved in the interaction, but remote sites as well.

Siuzdak et al. performed MALDI analysis on the digests of the kinase inhibitory domain of the cell cycle regulatory protein, p21-B in both the free solution state and in the 1 : 1 complex of p21-B and cyclin-dependent kinase 2 (Cdk2) as shown in Figure 8.[26] Mass mapping of the trypsin digests of the free p21-B revealed peaks corresponding to all of the 12 potential trypsin cleavage sites (amino acids 9, 16, 19, 20, 32, 46, 48, 67, 69, 75, 83, and 84). However, MALDI mass spectra for the trypsin digests of p21-B in the presence of Cdk2 showed that several p21-B fragments that were produced in the absence of Cdk2 disappear in the presence of Cdk2. Specifically, Siuzdak et al. found that amino acids 46, 48, 67 were protected from trypsin cleavage in the presence of Cdk2. This data was later crystallographically supported by the crystal structure data on a homologous system using a p21 homolog p27.[27]

Other methods of probing higher-order structure include investigating the chemical reactivity of individual amino acids in a protein and chemical cross-linking studies. Both approaches stem from the fact that MALDI has proved to be a powerful method with which to characterize covalent post-translational modifications.[28] Using simple modification chemistries such as acylation or succinylation, Glocker et al. have shown that there is a clear correlation between the relative reactivity of specific amino acids and their surface (accessibility) topology in a protein.[29] Chemical cross-linking studies consist of treating proteins with cross-linkers prior to digestion. Cross-linkers covalently attach adjacent protein regions or protein subunits, therefore the resulting proteolytic fragments are good indications of the overall tertiary and/or quaternary protein structure.

3 APPLICATIONS

3.1 Diagnostic

As mentioned previously, the power of MALDI lies in its ability to analyze complex mixtures, making MALDI a promising method for the diagnostic screening of biological fluids (serum, cerebral spinal fluid, urine). However, low analyte concentrations and signal suppression due to large amounts of host protein (e.g. hemoglobin in blood) often hinder the screening of such complex

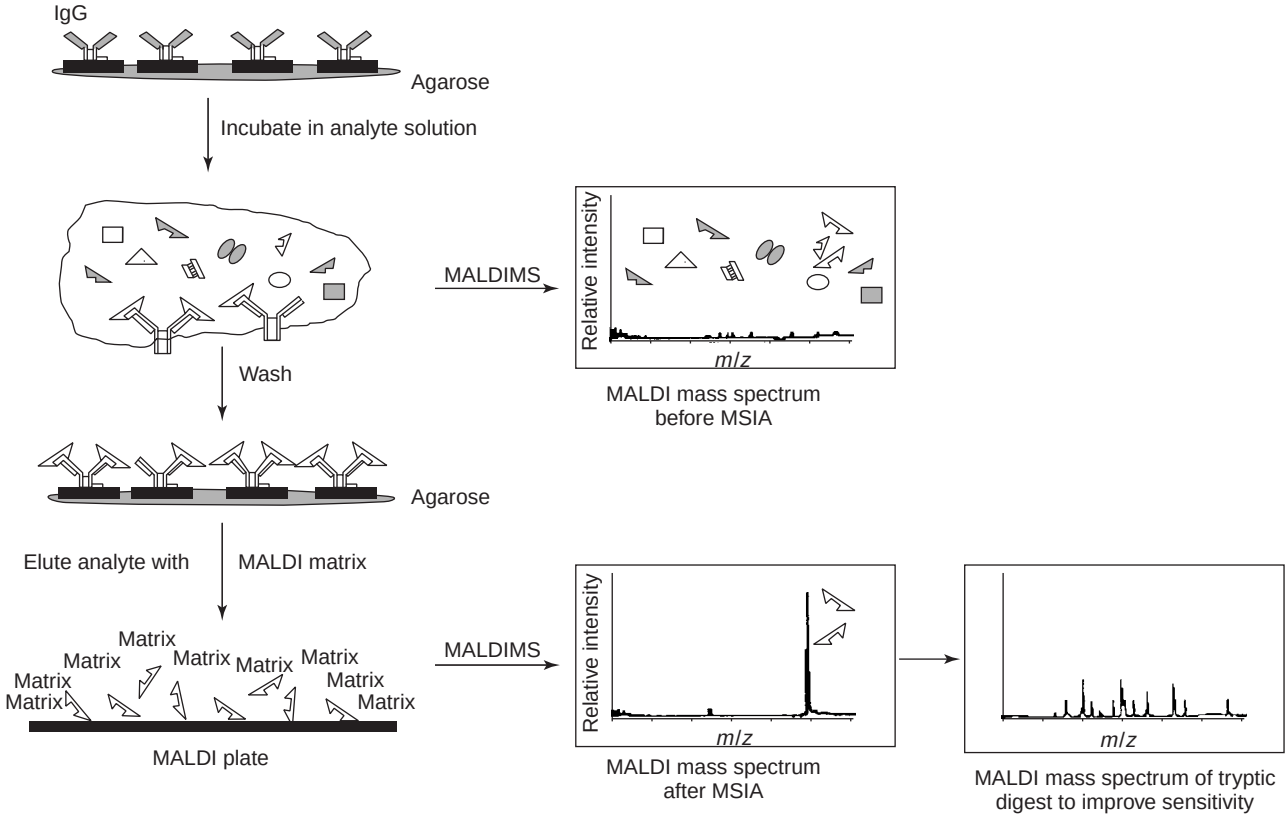

Figure 9 MSIA.

natural fluids. It therefore becomes necessary to isolate the analyte from the fluids prior to MALDI analysis. Mass spectrometric immunoassay (MSIA) (Figure 9), developed by Nelson et al., relies on the affinity capture (antibody–antigen recognition) of the analyte from the biological matrix prior to mass spectrometric analysis.[30] This method not only preconcentrates the analyte, but also provides for the most selective form of analyte isolation, making this an extremely powerful technique for screening biological fluids. Nelson et al., for example, were able to detect unambiguously femtomole amounts (nanomolar concentrations) of a snake toxin doped into human whole blood.

Similar to the enzyme-linked immunosorbent assay (ELISA), MSIA consists of the nanoscale immunoaffinity purification of the analyte of interest. Isolation is performed by incubating the biological matrix with antibodies immobilized on a solid support, such as resin beads. During incubation, the analyte is captured on the immobilized antibody and, after a series of washes, the antibody–antigen interaction is disrupted by the addition of the acidic MALDI matrix. Subsequent mass spectrometric analysis of the eluant confirms the presence of the antigen (analyte) at a specific m/z value. Unlike the ELISA, however, any nonspecifically bound molecules can be unambiguously differentiated from the analyte of interest owing to unique m/z values measured. In addition, incubation times are on the order of 15–20 min, significantly shorter than with the conventional ELISA.

In preliminary experiments aimed at identifying cardiac troponin T (cTnT) present in the urine of chronic renal failure patients, Fitzgerald et al. used biotinylated-anti-troponin mAb immobilized on streptavidin beads to affinity capture 0.15 $\mu g\,mL^{-1}$ (~44 pmol) cTnT from Tris buffer.[31] In the MALDI mass spectrum of 9 $\mu g\,mL^{-1}$ cTnT in 100-mM Tris buffer (Figure 10b), no ion signals of cTnT are observed. Figure 10(a) shows the mass spectrum obtained after the MSIA isolation and elution, where cTnT is observed at $m/z \sim 34$ kDa. Thus, the combination of MS with immunoaffinity capture serves as a rapid, sensitive, and selective means of detecting biologically relevant compounds.

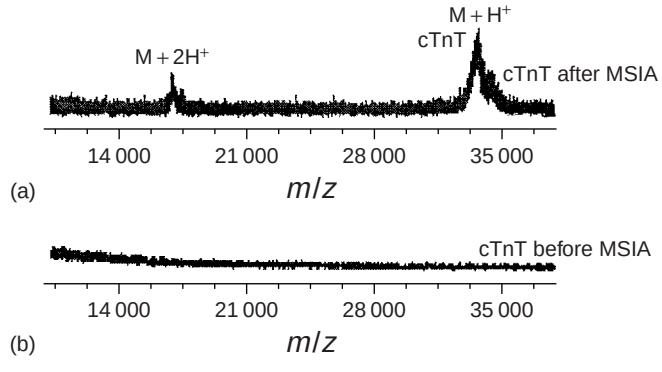

Figure 10 MSIA example.

3.2 Quantitative Aspects of Matrix-assisted Laser Desorption/Ionization

The ability of MALDI to analyze complex heterogeneous mixtures has made it a common and valuable technique in the analysis of biological fluids, analysis commonly performed by immunoassays and HPLC. Immunoassays generally have low reproducibility and reliability and provide little or no selectivity between a drug and its metabolites. This lack of specificity is a significant limitation since metabolites, although structurally similar to the parent compound, often have different biological activity. Mass spectrometric analysis, on the other hand, allows for co-extracted metabolites to be identified and quantitatively monitored (unless they have the same molecular weight). Although HPLC is relatively selective and accurate, the sensitivity is compound dependent and method development can be time-consuming. Thus, while both techniques are useful, both suffer when compared to the speed, sensitivity, and accuracy offered by MS.

In the following example, an automated MALDI MS procedure was used to improve the clinical analysis of a cyclic peptide (the immunosuppressant drug cyclosporin A (CsA)) **(4)** from whole blood extracts.[32] A MALDI mass spectrometer equipped with automated multisampling capabilities was used to facilitate data collection and analysis of CsA. Extraction optimization was performed by generating an array of solvent systems to identify successful extractions. The first generation of experiments revealed four effective binary solvent systems (hexane/EtOH, ACN/H_2O, ACN/MeOH, hexane/$CHCl_3$). A new array based on these solvent systems was generated and in a second iteration of these experiments, hexane/$CHCl_3$ (70 : 30) was found to provide the most effective single-step extraction for cyclosporin and its metabolites (Figure 11). In order to determine the efficiency of the new extraction

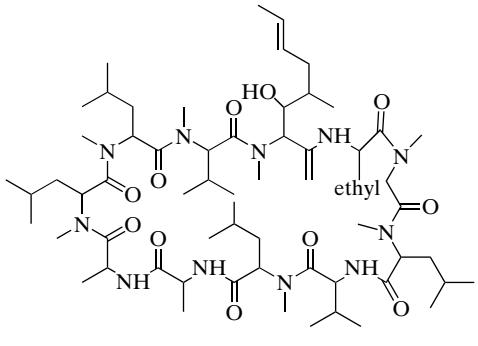

(4)

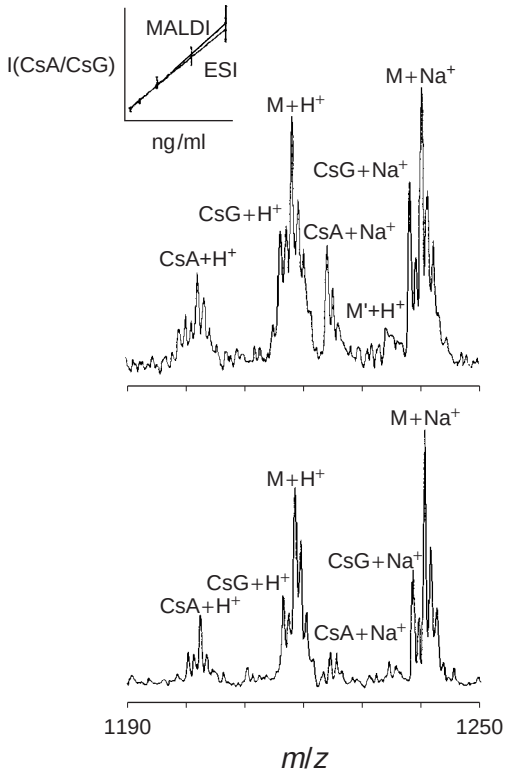

Figure 11 MALDI pharmacokinetics.

procedure it was compared to the previously developed ''ether'' extraction.

Initially, calibration curves were obtained using the ''ether'' extraction (a plot of the ratio of int.$[CsAH^+ + CsANa^+]/[CsGH^+ + CsGNa^+]$ vs CsA concentration) with both the ESI (electrospray ionization) and MALDI mass spectrometers (cyclosporin G, CsG, was used as the internal standard). The data exhibited a linear relationship in the concentration range 0–1500 ng mL^{-1} (Table 3) with an excellent correlation coefficient ($R = 0.999$) for both electrospray and MALDI. The new extraction generated standard curves for CsA that were very similar to those from the ''ether'' extraction method.

Table 3 Results obtained from the ESI/MS and MALDIMS analysis of extracted SA from standard blood samples using the 70/30 hexane/$CHCl_3$ extraction

Standard ng mL^{-1}	ESI ng mL^{-1} (error)	MALDI ng mL^{-1} (error)
100	98 (2%)	94 (6%)
250	231 (8%)	238 (5%)
500	523 (5%)	543 (9%)
1000	1009 (1%)	961 (4%)
1500	1404 (6%)	1514 (1%)

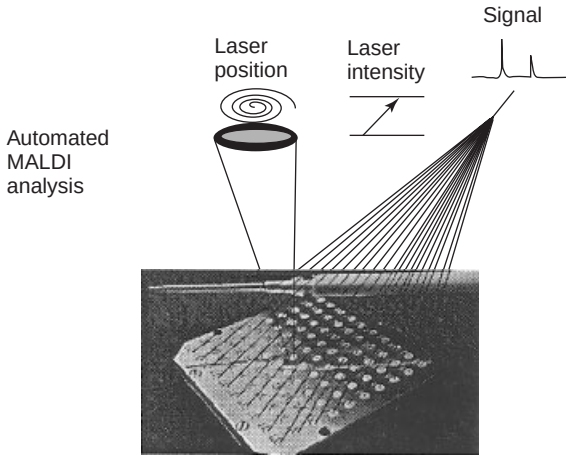

Figure 12 MALDI automation.

The automation of MALDI analyses, such as that used in the cyclosporin study, is becoming increasingly important in proteomics and combinatorial chemistry. These analyses are driven by a computer-controlled procedure to monitor the ion signal as a function of laser position and laser intensity (Figure 12). To accomplish this, the computer workstation automatically adjusts laser intensity and searches the sample well until a signal (within the specified mass range and intensity threshold) is obtained. Based on a careful preselection of autosampler options, each parameter (laser intensity, search pattern and step size in well, signal within a specified m/z range, and m/z range) is adjusted to minimize time of analysis and maximize signal quality.

What follows is a brief description of a MALDI automation procedure performed for the analysis of peptides. Here, the laser intensity was initially set at a minimum energy setting and then allowed to increase up to a maximum (~2–50 μJ/pulse, as controlled by a variable neutral density filter). Step sizes were made in five increments, resulting in an increase in laser intensity of approximately 10 μJ/pulse per step. The laser intensity was increased until an acceptable data signal was acquired, whereby if no signal was observed the laser beam was repositioned on the well and MALDI analysis resumed at the lower laser power. To adjust the laser position on the MALDI sample plate, a preprogrammed spiral search pattern was used which began in the center of each circular well and spiraled outward in 0.2 mm increments. For each sample well analysis, only signals that reached a specified intensity were saved and once this signal was observed the analyses would automatically move to the next well. On average the total time spent for each sample well analysis was 140 s. This included acquisition time for averaging 64 scans and a delay for adjusting laser intensity and repositioning the sample plate.

3.3 Characterizing Peptides and Reactions Directly from the Solid Phase

Several reports have shown that MS can be very useful in characterizing compounds covalently bound to solid polymeric supports subsequent to their chemical cleavage from the resin. More specifically, the characterization of peptides and carbohydrates covalently linked to a polymeric support through a photolabile linker is feasible using MALDIMS. The scheme outlined in Figure 13(a) permits the characterization of resin-bound analytes in a single step and requires no pretreatment of the sample to induce cleavage from the support. In addition, this approach also shows that the technique is suitable for monitoring chemical reactions on the solid phase.[33,34]

Figure 13(b) shows the mass spectrum of the unprotected resin-bound peptide. MALDI analysis yielded a characteristic $[M+H]^+$ signal at m/z 1247 which is consistent with the expected mass of the free peptide (MW = 1245.5 Da). The spectrum in Figure 13(b) was acquired from a single bead. The peptide (in a chemical cleavage step) was recovered from resin samples and then subjected to MALDI analysis, indicating that the sample was not completely photolyzed in the experiment. This is not surprising, as only a small amount of material (<femtomoles) is typically consumed during such analysis. Figure 13(c) shows the MALDI analysis of resin-bound carbohydrate, illustrating the application of this technique to another class of compounds.

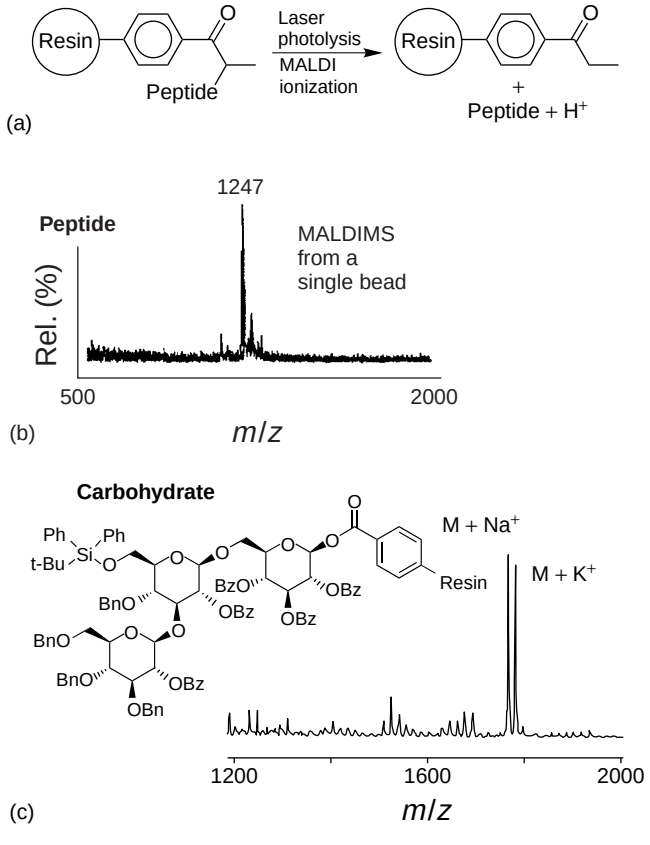

Figure 13 MALDI analysis from the solid phase.

The one-step MALDI procedure for the direct analysis of resin-bound molecules described above is also well suited to studying chemical reactions on the solid phase. As a demonstration of this utility, the coupling reaction of a Boc-Arg(Tos) residue (preactivated as the hydroxybenzotriazole ester) to the protected 11 amino acid peptide (described above) was monitored as a function of time. Interestingly, the Fmoc protected peptide was not amenable to MALDI analysis; the absence of appropriate protonation sites in this sample probably prevents ionization. Therefore, prior to analysis resin samples were treated with 20% (v/v) piperidine in dimethylformamide (DMF) in order to remove the Fmoc protecting groups, after which the reaction proceeded quickly and coupling appeared nearly complete in just 6 min.

The direct analysis of resin-bound molecules by MALDI offers several important advantages. The first is the lack of an additional cleavage step prior to mass analysis, which is often required by other methods of characterization. Performing MALDI directly on the solid phase requires less sample handling and more efficient management since the resin-bound compound can be easily recovered for subsequent manipulations (again, <femtomoles of material is consumed in typical MALDI analyses). The most significant advantage is that it can be used to monitor chemical reactions on the solid phase in real time, in much the same way that thin-layer chromatography is used to monitor reactions in solution. Furthermore, all analytes amenable to MALDI ionization should prove suitable for routine analysis by this procedure.

4 CONCLUSION

Since the late 1970s, tremendous progress has been made in the analysis of peptides and proteins by MS. Much of this enormous growth can be attributed to advances in desorption/ionization techniques such as MALDI, namely in resolution and accuracy. In addition, it is now routine to analyze peptides and proteins at femtomole levels which makes it well suited for protein mass mapping and structural determination. We are all familiar with the prominence of gel electrophoresis methods in molecular biology studies; we posit that MS methods will capture this position as a primary research tool in coming years owing to its superior sensitivity, precision, accuracy, and throughput. The applications discussed here are just a few examples of current approaches that will see expanded use in the future.

ACKNOWLEDGMENTS

The authors would like to thank Jennifer Boydston for helpful editorial comments. This work was supported by grants from the National Institutes of Health, 1 R01 GM55775-01A1 and 1 S10 RR07273-01 (G.S.).

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
CaM	Calmodulin
CCA	α-Cyano-4-hydroxycinnamic acid
Cdk2	Cyclin-dependent Kinase 2
CID	Collision-induced Dissociation
CsA	Cyclosporin A
CsG	Cyclosporin G
cTnT	Cardiac troponin T
DE	Delayed Extraction
DHB	2,5-Dihydroxybenzoic acid
DMF	Dimethylformamide
ELISA	Enzyme-linked Immunosorbent Assay
ESI	Electrospray Ionization
EST	Expressed Sequence Tag
FTMS	Fourier Transform Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ICR	Ion Cyclotron Resonance
KE	Kinetic Energy
MALDI	Matrix-assisted Laser Desorption/ Ionization
MALDIMS	Matrix-assisted Laser Desorption/ Ionization Mass Spectrometry
MS	Mass Spectrometry
MS^n	Tandem Mass Spectrometry (nth series)
MSIA	Mass Spectrometric Immunoassay
MS/MS	Tandem Mass Spectrometry
NC	Nitrocellulose
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
PSD	Postsource Decay
RF	Radio Frequency
RPLC	Reversed-phase Liquid Chromatography
SA	3,5-Dimethoxy-4-hydroxycinnamic acid
TFA	Trifluoroacetic Acid
TOF	Time-of-flight
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Carbohydrate Analysis **(Volume 1)**
Glycolipid Analysis

Industrial Hygiene **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Mass Spectrometry **(Volume 13)**
Artificial Intelligence and Expert Systems in Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**, 259–280 (1988).
2. R.J. Cotter, 'Time-of-flight Mass Spectrometry for the Structural Analysis of Biological Molecules', *Anal. Chem.*, **64**, A1027–A1039 (1992).
3. B.T. Chait, S.B.H. Kent, 'Weighing Naked Proteins – Practical, High-accuracy Mass Measurement of Peptides and Proteins', *Science*, **257**, 1885–1894 (1992).
4. F.Hillenkamp, M.Karas, R.C.Beavis, B.T.Chait, 'Matrix-assisted Laser Desorption Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**, A1193–A1202 (1991).
5. F. Hillenkamp, M. Karas, D. Holtkamp, P. Klusener, 'Energy Deposition in Ultraviolet Laser Desorption Mass Spectrometry of Biomolecules', *Int. J. Mass Spectrom. Ion Phys.*, **69**, 265–276 (1986).
6. A. Vertes, R.D. Levine, 'Sublimation Versus Fragmentation in Matrix-assisted Laser Desorption', *Chem. Phys. Lett.*, **171**, 284–290 (1990).
7. R.E. Johnson, B.U.R. Sundquist, 'Laser-pulse Ejection of Organic Molecules from A Matrix – Lessons from Fast-ion-induced Ejection', *Rapid Commun. Mass Spectrom.*, **5**, 574–578 (1991).
8. R.C. Beavis, B.T. Chait, K.G. Standing, *Physics*, **66**, 269 (1990).
9. M.B. Comisarow, A.G. Marshall, 'Frequency-sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **26**, 489–490 (1974).
10. R.C. Beavis, B.T. Chait, 'Cinnamic Acid Derivatives as Matrices for Ultraviolet Laser Desorption Mass Spectrometry of Proteins', *Rapid Commun. Mass Spectrom.*, **3**, 432–435 (1989).
11. R.C. Beavis, B.T. Chait, 'Factors Affecting the Ultraviolet Laser Desorption of Proteins', *Rapid Commun. Mass Spectrom.*, **3**, 233–237 (1989).
12. K. Strupat, M. Karas, F. Hillenkamp, '2,5-Dihydroxybenzoic Acid–a New Matrix for Laser Desorption

Ionization Mass Spectrometry', *Int. J. Mass Spectrom. Ion Process.*, **111**, 89–102 (1991).
13. S.L. Cohen, B.T. Chait, 'Influence of Matrix Solution Conditions on the MALDI-MS Analysis of Peptides and Proteins', *Anal. Chem.*, **68**, 31–37 (1996).
14. O. Vorm, P. Roepstorff, M. Mann, 'Improved Resolution and Very High Sensitivity in MALDI TOF of Matrix Surfaces Made by Fast Evaporation', *Anal. Chem.*, **66**, 3281–3287 (1994).
15. O. Vorm, M. Mann, 'Improved Mass Accuracy in Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry of Peptides', *J. Am. Soc. Mass Spectrom.*, **5**, 955–958 (1994).
16. M. Kussmann, et al., 'Matrix-assisted Laser Desorption/ Ionization Mass Spectrometry Sample Preparation Techniques Designed for Various Peptide and Protein Analytes', *J. Mass Spectrom.*, **32**, 593–601 (1997).
17. F. Wold, K. Moldave, *Methods Enzymol.*, **106** (1984).
18. F. Wold, 'In Vivo Chemical Modification of Proteins (Post-translational Modification)', *Annu. Rev. Biochem.*, **50**, 783 (1981).
19. M. Wilm, et al., in *Mass Spectrometry in the Biological Sciences*, eds. A.L. Burlingame, S.A. Carr, Humana Press, Totowa, NJ, 245–263, 1995.
20. Y. Wu, J.E. Vath, M.C. Huberty, S.A. Martin, 'Identification of the Facile Gas-phase Cleavage of the ASP PRO and ASP XXX Peptide Bonds in Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Anal. Chem.*, **65**, 3015–3023 (1993).
21. J. Machold, Y. Utkin, D. Kirsch, R. Kaufmann, 'Photolabeling Reveals the Proximity of the Alpha-neurotoxin Binding Site to the M2 Helix of the Ion Channel in the Nicotinic Acetylcholine Receptor', *Proc. Natl. Acad. Sci. USA*, **92**, 7282–7286 (1995).
22. M.C. Huberty, J.E. Vath, Y. Wu, S.A. Martin, 'Site-specific Carbohydrate Identification in Recombinant Proteins using MALD-TOF MS', *Anal. Chem.*, **65**, 2791–2800 (1993).
23. D.C.A. Neville, C.R. Rozanas, E.M. Price, D.B. Gruis, 'Evidence for Phosphorylation of Serine 753 in CFTR Using a Novel Metal-ion Affinity Resin and Matrix-assisted Laser Desorption Mass Spectrometry', *Prot. Sci.*, **6**, 2436–2445 (1997).
24. R.S. Annan, S.A. Carr, 'Phosphopeptide Analysis by Matrix-Assisted Laser Desorption Time-of-flight Mass Spectrometry', *Anal. Chem.*, **68**, 3413–3421 (1996).
25. J.K. Lewis, Arizona State University, 'Protein Structural Characterization Using Bioreactive Mass Spectrometer Probes', *Ph.D. Thesis Dissertation* (1997).
26. R.W. Kriwacki, J. Wu, G. Siuzdak, P.E. Wright, 'Probing Protein/Protein Interactions with Mass Spectrometry and Isotopic Labeling–Analysis of the P21/CDK2 complex', *J. Am. Chem. Soc.*, **118**, 5320–5321 (1996).
27. A.A. Russo, P.D. Jeffrey, A.K. Patten, J. Massague, N.R. Pavletich, 'Crystal Structure of the P27(KIP1) Cyclin-dependent-Kinase Inhibitor Bound to the Cyclin A CDK2 Complex', *Nature*, **382**, 325–331 (1996).
28. R.W. Kriwacki, J. Wu, G. Siuzdak, P.E. Wright, 'Probing Protein Structure Using Biochemical and Biophysical Methods–Proteolysis, Matrix-assisted Laser Desorption/Ionization Mass Spectrometry, High-performance Liquid Chromatography and Size-exclusion Chromatography of p21 (Waf1/Cip1/Sdi1)', *J. Chromatog.*, **777**, 23–30 (1997).
29. M.O. Glocker, S. Nock, M. Sprinzl, M. Przybylski, 'Characterization of Surface Topology and Binding Area in Complexes of the Elongation Factor Proteins EF-Ts and EF-Tu Center Dot GDP from *Thermus Thermophilus*: A Study by Protein Chemical Modification and Mass Spectrometry', *Chem. Eur. J.*, **4**, 707–715 (1998).
30. R.W. Nelson, J.R. Krone, A.L. Bieber, P. Williams, 'Mass Spectrometric Immunoassay', *Anal. Chem.*, **67**, 1153–1158 (1995).
31. R.L. Fitzgerald, K.D. Harris, W.L. Frankel, D.A. Herold, 'Cardiac Troponin-T Analysis by Immunoaffinity Purification and MALDI Mass Spectrometry', *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, 948 (1996).
32. J. Wu, K. Chatman, K. Harris, G. Siuzdak, 'An Automated MALDI Mass Spectrometry Approach for Optimizing Cyclosporin Extraction and Quantitation', *Anal. Chem.*, **69**, 3767–3771 (1997).
33. G. Siuzdak, 'Probing Molecular Diversity with Mass Spectrometry', *J. Assoc. Lab. Autom.*, **3**, 34–37 (1998).
34. M.C. Fitzgerald, K. Harris, C.G. Shevlin, G. Siuzdak, 'Direct Characterization of Solid Phase Resin-bound Molecules by Mass Spectrometry', *Bioorg. Med. Chem. Lett.*, **6**, 979–982 (1996).

Molecular Modeling in Peptide and Protein Analysis

Krzysztof Kuczera
University of Kansas, Lawrence, USA

1 Introduction **1**
2 History **2**
3 Peptide and Protein Structure **3**
4 Overview of Modeling Methods **3**
4.1 Potential Energy Functions 3
4.2 Energy Minimization 6
4.3 The Multiple Minimum Problem 7
4.4 Normal Mode Analysis 7
4.5 Molecular Dynamics Simulations 9
4.6 Monte Carlo Simulations 13
4.7 Conformational Search 14
4.8 Free Energy Simulations 15
4.9 Solvation and Environmental Effects 18
4.10 Treatment of Electrostatic Interactions 21
4.11 Path Exploration Methods 22
5 Relation to Experimental Approaches **23**
5.1 X-ray Diffraction 23
5.2 Nuclear Magnetic Resonance Spectroscopy 24
5.3 Hydrogen Isotope Exchange 24
5.4 Fluorescence Depolarization 24
5.5 Vibrational Spectroscopy 25
5.6 Neutron Scattering 25
6 Examples of Applications **25**
7 Future Directions **28**
Acknowledgments **29**
Abbreviations and Acronyms **29**
Related Articles **29**
References **30**

This article describes the theoretical background and practical applications of classical molecular mechanics and molecular dynamics (MD). Following a brief historical outline of molecular modeling, the concept of the molecular potential energy function is introduced and the basic algorithms of molecular mechanics: energy minimization, normal mode analysis, MD and Monte Carlo (MC) simulations are defined. The advantages of molecular modeling methods – speed of calculations and detailed microscopic insight – are described. The shortcomings – approximate nature and difficulties with conformational sampling – are also analyzed. A number of modern approaches aimed at improving the reliability of computer simulations are briefly presented, including advanced energy optimization methods, treatment of long-range forces and accelerated conformational exploration. Examples of applications of molecular modeling methods to study of various aspects of structure, dynamics and function of peptides and proteins are presented. A brief exposition of the relationship between molecular modeling and some fundamental experimental techniques, such as X-ray crystallography, neutron scattering, nuclear magnetic resonance (NMR), fluorescence and vibrational spectroscopies is provided. Finally, an outline of future directions of modeling studies is suggested.

1 INTRODUCTION

The term ''molecular modeling'' covers a wide range of approaches used to calculate, analyze and visualize molecular properties. The spectrum of methods ranges from physical models, made of plastic or wire, to structural rendering by computer graphics and mathematical models of different levels of sophistication. This article focuses on the theoretical background and practical applications of classical molecular mechanics and MD, a mathematical model widely used in computational studies of physical, chemical and biological properties of molecular systems. The wider field of molecular modeling, including protein homology modeling, docking studies, extended atom and lattice models, electrostatic PB (Poisson–Boltzmann) methods and quantum chemical calculations will be only briefly mentioned.

In molecular mechanics, the fundamental quantity is the potential energy function, which enables assignment of energy values to the different configurations of the atoms comprising the system. The availability of the energy as a function of atomic positions allows the relative stability of different structures (conformational search) to be determined. First derivatives of the potential with respect to atomic coordinates yield the forces acting on the atoms. The forces may be used to numerically solve equations of motion of the molecular system (MD) or to locate energy minima (energy minimization or structure optimization). Second derivatives of the potential energy yield the harmonic force constants, which are the basis for normal mode analysis and predictions of vibrational spectra and collective modes of motion.

Using basic principles of physics and chemistry, computer simulations employing the model molecular mechanics potential may be used to predict a wide range

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of properties for molecular systems. These properties include the stable structure (or group of structures), intermolecular interactions, types of motion present and their timescales. These general quantities can be related to specific observables – structures, dipole moments, diffusion coefficients, vibrational spectra, NMR spectral parameters, etc. Using specialized simulation techniques, free energy differences between various molecular states may be evaluated, allowing calculations of such quantities as population ratios of peptide conformers or binding constants for protein-inhibitor complexes.

Molecular mechanics models have two main advantages – speed and interpretive power. Standard molecular mechanics calculations use simple potential energy functions, which can be rapidly evaluated for large molecular systems – such as solvated peptides or proteins, including 10 000 or more atoms. At this time such large systems cannot be extensively studied by the (in principle) more accurate quantum mechanical approaches. By considering motions of individual atoms over timescales starting in the sub-picosecond range, simulations provide an enormous amount of microscopic detail about molecular structure and dynamics. Once the reliability of the simulation protocol has been verified by comparison with available experimental data, the detailed atomic-level predictions may be used to provide valuable insight into the microscopic mechanisms of observed phenomena. Such insight is highly valuable in developing basic understanding of complex molecular systems, guiding the development of new experiments and manipulation of molecular properties for practical purposes.

The main limitations of classical molecular mechanics lie in the approximate nature of the potential energy functions and in the difficulty of sampling the conformational space of large, flexible molecules such as peptides or proteins. In order to make the calculations fast and applicable to a wide range of systems, a number of simplifying assumptions about the form of the potential are usually made, such as the harmonic and diagonal form of the internal deformation force field, use of fixed point charges to describe electrostatic interactions, pair-wise additivity of atom–atom nonbonded interactions and truncation of long-range forces. These approximations limit the accuracy of the calculated energies. Further, the classical potentials refer to a fixed electronic state, so that processes involving chemical reactions cannot be modeled using the standard molecular mechanics approach. The problem of potential accuracy and modeling of chemical reactions may in principle be overcome by using the more expensive quantum mechanical methods. The conformational sampling problem is even more difficult, as there is currently no viable approach for solving it. This problem is due to the fact that the number of possible states of a molecular system grows exponentially with the number of atoms. As a consequence, it is currently impossible to exhaustively explore the conformational space of large flexible molecules. Thus, molecular mechanics cannot predict the three-dimensional structure of a small protein with 100 amino acid residues from first principles. Using explicit molecular solvent models, even structure prediction of a decapeptide in solution is beyond the range of current methods. For large molecular systems we have to employ experimental information on structure and use special simulation techniques to obtain reliable results.

In spite of its limitations, and thanks to its speed and insights, the molecular mechanics approach is being extensively used in studying peptides, proteins and other complex systems. Due to continuing improvements in potentials, simulation algorithms and computer power, more reliable results are being obtained for ever larger systems. As the number of successful applications grows, the computer simulations based on molecular mechanics methods are becoming widely accepted as a useful tool in the arsenal of modern chemists and biologists, a tool complementary to both experimental and theoretical approaches.

2 HISTORY

The development of quantum mechanics in the 1920s and 1930s led to the concept of the molecular potential energy surface, expressing the energy as a function of positions of nuclei. Attempts to model molecular properties using simple potentials soon followed. Interpretation of molecular vibrational spectra in terms of normal modes and the underlying harmonic force constants of small symmetric molecules developed during the 1930s and 1940s.[1] An early study involving molecular motion is the 1936 classical trajectory calculation for the $H + H_2 \longrightarrow H_2 + H$ reaction.[2] Quantitative description of molecular strain in simple organic molecules through harmonic bond stretching and angle deformations, and Lennard–Jones interactions or exponential repulsion between nonbonded groups was proposed in 1946 by Hill[3] and Westheimer and Mayer.[4] An internal rotation barrier of the form $V_0[1 - \cos(3\phi)]$ for methyl groups was employed to explain the entropy of ethane in 1936.[5] Potentials including electrostatic interactions between atomic partial charges were used by Kitaigorodskij[6] in the 1950s to investigate the properties of organic crystals. A simple hard-sphere model was employed to predict the allowed conformations of amino acids in 1963.[7] An interesting account of the early history of molecular modeling is given by Rappe and Casewit.[8]

Potentials for modeling wider classes of polyatomic molecules, including peptides and proteins, became available in the 1960s, 1970s and 1980s.[9–13] This period coincided with the introduction of computers and computer-based numerical calculations into chemistry, physics and biology. The basic tools of molecular modeling – energy minimization, normal mode analysis, MD and MC simulations were also developed in those years.[14] Simulations started with simple systems – hard-sphere fluids,[15] liquid argon with a Lennard–Jones potential[16] and pure water.[17] Studies of other liquids and solutions followed.[18] The first MD simulation of a small protein, a 9 ps trajectory of bovine pancreatic trypsin inhibitor (BPTI) in vacuum, was reported in 1977.[19] Further MD and normal mode studies of proteins in vacuum followed in the 1980s: myoglobin,[20] lysozyme,[21] BPTI, crambin, ribonuclease.[22] Molecular free energy simulations were introduced in the late 1970s and early 1980s, first for simple model systems[23–25] and then for point mutations in proteins.[26] Attempts to include electric polarization were introduced.[27–29] Methods to efficiently and accurately model chemical reactions in condensed phases based on the valence bond,[30] Car-Parinello[31] and quantum mechanics/molecular mechanics (QM/MM) approach[32,33] were implemented. Finally, further increases in computer power and algorithm development in the 1990s brought studies of solvated peptides[34] and proteins,[35] inclusion of long-range interactions,[36] explicit treatment of mobile counterions,[37] simulations of protein folding/unfolding events,[38–40] efficient path exploration,[41] multiple time stepping (MTS),[42] and advanced conformational search strategies.[43,44]

3 PEPTIDE AND PROTEIN STRUCTURE

Peptides and proteins are polymeric chains built from 20 commonly occurring amino acids.[45] The primary structural coordinates are the flexible backbone dihedrals ϕ and ψ, and sidechain dihedrals $\chi_1, \chi_2, \ldots$ (Figure 1[45,46]). Proteins and many peptides exhibit a tendency to exist in a single, well defined three-dimensional structure, denoted as the folded form. The folded structures are characterized by the presence of secondary structure elements – mainly α-helices, β-sheets and several kinds of regular turns.[47] In proteins, the secondary structure elements are further organized into structural motifs. Several examples are shown in Figure 1. Further complexity is added by the presence of nonpeptide cofactors such as porphyrins, nucleotides or metal ions, the possibility of forming covalent disulfide bonds between Cys sidechains, and aggregation of individual subunits to form quaternary structure. In principle, the properties of a peptide/protein system are determined by the sequence. However, in practice the complex problem of predicting protein structure, dynamics and function from the sequence alone is at present intractable. A summary of the difficulties encountered and solutions proposed to this challenging problem is presented below.

4 OVERVIEW OF MODELING METHODS

The molecular mechanics approach to molecular modeling starts with the definition of the potential energy function. The potential is crucial for the basic simulation algorithms – energy minimization, normal mode analysis, MD and MC methods. These algorithms may be applied to explore various properties of the modeled systems – from favored conformations to structural fluctuations, detailed microscopic dynamics and free energies of conformational transitions or mutation processes. In the course of the modeling, special attention needs to be paid to the complexity of the energy landscape, treatment of solvation, long-range interactions and large-scale structural change.

4.1 Potential Energy Functions

At the fundamental level a molecular potential energy function U may be identified with the total energy of a system of electrons and nuclei in the Born–Oppenheimer approximation of quantum mechanics.[48] From the quantum mechanical viewpoint U consists of the kinetic energy of the electrons and three Coulombic terms, describing electron–electron, electron–nuclear and nuclear–nuclear interactions, all in a given electronic state. Thus, in the classical description the potential depends only on positions of nuclei (or atoms). To facilitate calculations and relate the energy to chemical structure, the form of U is usually parameterized in terms of simple structural descriptors – bond lengths, bond angles, dihedral angles and nonbonded atom–atom distances. A typical form of the potential, shared by many publicly available force fields, such as CHARMM,[49] AMBER,[50] OPLS,[51] or GROMOS[52] is shown in Equation (1):

$$U = \frac{1}{2}\sum_{\text{bonds}} k_b(b-b_0)^2 + \cdots + \frac{1}{2}\sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \frac{1}{2}\sum_{\text{impropers}} k_w(w-w_0)^2 + \frac{1}{2}\sum_{\text{dihedrals}} k_\phi \times [1+\cos(n\phi+\delta)] + \sum_{\text{atom pair } ij}\left(\frac{q_iq_j}{R_{ij}} - \frac{A_{ij}}{R_{ij}^6} + \frac{B_{ij}}{R_{ij}^{12}}\right. \quad (1)$$

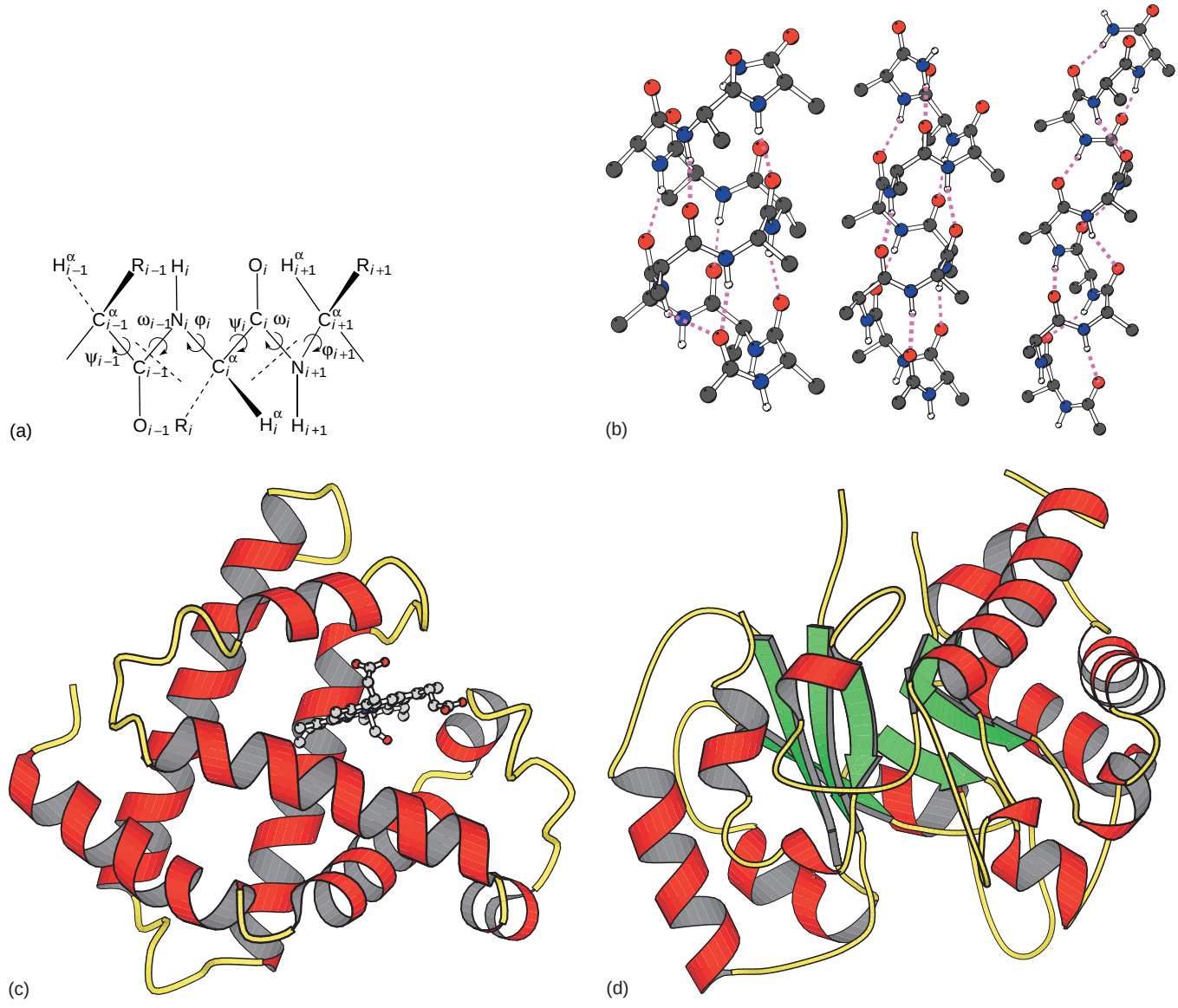

Figure 1 Peptide and protein structure. (a) Dihedrals determining peptide conformation. (b) Helical structures of $(Ala)_{10}$. From left to right: π-helix, with $i - i + 5$ hydrogen bonds, α-helix, with $i - i + 4$ hydrogen bonds and the 3_{10}-helix, with $i - i + 3$ hydrogen bonds. (c) Myoglobin – an α-helical protein, protein data bank (PDB) structure 1 MBC.[45] (d) Triosephosphate isomerase – example of a β-barrel, from PDB structure 3 TIM.[46]

In this equation b, θ, w, ϕ and R_{ij} are the actual bond lengths, valence bond angles, improper and proper dihedral angles and atom–atom distances for a given configuration of the system, respectively; $k_b, b_0, k_\theta, \theta_0, k_w, w_0, k_\phi, n, \delta, q_i, A_{ij}$ and B_{ij} are parameters. One way of looking at potential energy functions of this type is to treat them simply as mathematical models leading to certain predictions. However, both the method of derivation and the form of the function gives the parameters physical meaning.

The first three terms in U describe energies of deformation from equilibrium of chemical bond length (b), bond angles (θ) and improper dihedrals (w). These distortions are usually small under normal conditions, allowing the energy to be described in the form of a sum of harmonic (Hooke's law) terms. The presence of these terms assures that throughout the simulation the molecule preserves its basic chemical structure.

The fourth term in U above describes the energy of torsion around chemical bonds – it is periodic and is modeled as a cosine term (or sum of several cosines with different periodicities). The presence of the $\cos(n\phi)$ term assures that the molecule will sample n distinct potential energy minima as the complete rotation around

the chemical bond is performed. Taken together, the first four terms might be considered as the strain energy of deforming a molecular structure from equilibrium. For further discussion about the approximations involved in the first four terms see section 4.4.

The last term of the formula for U describes the nonbonded interactions between pairs of atoms, and has two contributions: electrostatic interactions between point charges (usually located at atom centers) and a Lennard–Jones 6–12 potential which models the universal features of interatomic interactions: weak attraction at moderate distances and strong repulsion on close contact. For a review of theoretical and experimental aspects of intermolecular interactions see Margenau and Kestner[53] and Israelachvili.[54] The internal strain is the dominant factor in determination of the structure of small, relatively rigid molecules such as water or benzene. In larger, flexible molecules such as peptides and proteins, the torsional deformations and nonbonded interactions play a major role.

There are several approaches to determine the potential parameters.[8] Most force fields, e.g. AMBER, CHARMM, GROMOS, OPLS, MM2, MM3,[55] and Tripos, primarily aim to reproduce experimental data, such as molecular geometries, vibrational frequencies, torsional barriers, conformational energy differences, heats of vaporization or sublimation, densities, etc. Due to paucity of accurate experimental measurements, increase in computer power and proliferation of quantum chemistry codes, the use of electronic structure calculations is gaining popularity as a source of data. The mostly empirical force fields listed above usually use some quantum chemical information, e.g. to calculate atomic point charges or vibrational frequencies. The Discover force field is an example of a potential parameterized mainly on quantum calculations, with some empirical corrections.[56] In both the quantum and empirical approaches, some parameters may be determined directly (e.g. bond lengths), while others are optimized by indirect fitting to data (e.g. force constants derived from vibrational spectra, van der Waals radii from heats of vaporization). Generally, parameter values are determined in studies of small molecule models such as amino acids or their fragments in condensed phases – crystals or liquids. The assumption that the same parameters may be used to describe macromolecules is verified by comparison of calculated and observed properties. The above force fields provide relatively accurate models for classes of molecules that have been used in the parameterization. The price of this accuracy is that extension of the models to new types of molecules typically requires some rather time-consuming parameter development. An alternative to this approach is provided by force fields such as DREIDING[57] or UFF (universal force field[58]) which assign most parameters based on simple general rules. In the rule-based methods, the loss of accuracy is offset by the more general applicability of the models.[8]

Even the relatively simple formula in Equation (1) is not evaluated directly as is, and further approximations are usually made in nonbonded energy evaluation. For a system of N atoms there are $N(N-1)/2$ distinct pair interactions; e.g. with $N = 10\,000$ we get almost 50 million pairs. Calculating all these terms would take too much computer time, especially in MD or MC simulations, which typically require millions of energy evaluations. To increase the speed of computation, the nonbonded terms for atoms separated by more than a specified cutoff distance are ignored. The cutoff distances used in different force fields tend to increase with time, as computer power grows. Typical cutoffs used in the 1990s were 9–12 Å.[49] In order to remove any discontinuities in the energy or its derivatives due to the cutoff, the calculated energies had to be further modified by applying smoothing procedures.[12,59] The effects of truncation are most pronounced in the electrostatic energy, which falls off slowly with increasing distance. Schemes aimed at improved electrostatic energy calculations, including avoiding truncation effects are described in section 4.10. A decision about inclusion of bonded atom pairs in the nonbonded interaction list has to be made. Usually, atoms separated by one or two chemical bonds (so-called 1–2 and 1–3 interactions) are excluded from the nonbonded calculations, as it is assumed that the internal strain terms should describe their energetics sufficiently well. Atoms separated by three chemical bonds (1–4 interactions) are usually included, sometimes with modified values of the van der Waals parameters.[49] Finally, in order to reduce the number of nonbonded parameters, two parameters, typically a van der Waals radius and an interaction well depth are assigned to each atom. The parameters for pair interactions are then obtained from simple combination rules.[8]

For a peptide or protein system, the potential energy would involve the terms describing deformations of all the bonds, angles and dihedrals from their optimum values, as well as nonbonded interactions between atoms which are not chemically bonded, but are brought close together due to the three-dimensional folding of the molecule (tertiary structure). To simulate a solvated protein, we add additional terms in the potential, describing bond and angle deformations of the solvent molecules as well as solvent–solvent and solvent–protein interactions. It is truly amazing that such a simple functional form of the potential can describe the wide range of interactions occurring in macromolecules – formation of salt links, hydrogen bonds, solvation by polar and nonpolar solvents, complementary fitting of hydrophobic surfaces. The simple form is also misleading. For a protein of about

100 amino acids there might be 100 000 individual terms in the potential U, and 500 000 terms if solvation by explicit solvent is considered. This is a complicated multidimensional function, the basic properties of which are not yet fully understood.

4.2 Energy Minimization

Energy minimization is a versatile tool of molecular modeling, used to identify stable structures, ensure consistency of model structures with a given potential, and remove thermal excitation from structures obtained from MD or MC simulations. An excellent introduction to the theoretical aspects may be found elsewhere,[8,60,61] while examples of algorithms are presented in Flannery et al.[62]

4.2.1 Potential Energy Minima

The potential energy $U(q)$ of a large molecular system depends on positions q of all atoms ($3N$ coordinates for N atoms) and on a large number of parameters. A useful way of exploring properties of U and characterizing available molecular conformations is the identification of energy minima. A minimum of $U(q)$ may be generally defined as some value of the coordinates $q = q_{\min}$ such that $U(q) > U(q_{\min})$ for all possible values in some (possibly small) neighborhood of $q_{\min}$. By exploring different regions of variable space q, we may identify a series of minima. These are called *local minima*; each corresponds to the lowest possible value of U in some local neighborhood of coordinate space. Of special interest is the *global minimum*, i.e. the minimum which corresponds to the lowest possible of all values of $U(q)$. For functions of a large number of variables, the global minimum is rather difficult to find, see Wales and Scheraga[43] and below.

The minima represent possible stable structures of the molecular system. Rather than comparing some arbitrary structures in two different regions of conformation space, it is much more meaningful to compare energies of minima located in the two regions of interest. Some arbitrariness will always remain for large molecular systems, such as proteins, which have large numbers of local minima corresponding to a single macroscopically observable state, e.g. the ''native form''. In that case we have to look at a sample of minima from each region of conformational space (e.g. the ''folded structure'', ''substrate bound structure'', ''unfolded structure'').

Mathematically, energy minimization is a special case of a more general optimization problem: finding the lowest possible value for a function of many variables. The necessary condition for a minimum is that the function gradient is zero, as shown in Equation (2):

$$\nabla U = \mathbf{g(q)} = 0 \quad \text{or} \quad \frac{\partial U}{\partial x_i} = 0, \quad i = 1, 2, 3, \ldots, 3N \tag{2}$$

where x_i denote atomic Cartesian coordinates and N is the number of atoms. The gradient **g(q)** is a vector with $3N$ components. Points at which the gradient is zero are called *stationary points* of a function. Besides minima, stationary points include maxima and saddle points.

The sufficient condition for a minimum is that the second derivative matrix is positive definite. Equation (3) shows that for any $3N$-dimensional vector **u**

$$\sum_{i,j=1}^{3N} F_{ij} u_i u_j > 0 \tag{3}$$

A simpler operational definition of this property is that all eigenvalues of F are positive at a minimum. The second-derivative matrix **F** is defined by Equation (4):

$$F_{ij} = \frac{\partial^2 U}{\partial x_i \partial x_j} \tag{4}$$

F is discussed further in section 4.4. One measure of the distance from a stationary point is the root mean square gradient shown in Equation (5):

$$GRMS = \sqrt{\frac{1}{3N} \sum_{i=1}^{3N} \left(\frac{\partial U}{\partial x_i} \right)^2} \tag{5}$$

4.2.2 Minimization Algorithms

Because $U(q)$ is a complicated function varying quickly with atomic coordinates q, molecular energy minimization is an iterative procedure, in which the coordinates at step $n+1$ are determined from coordinates at previous step n, as shown in Equation (6):

$$q_{n+1} = q_n + \delta_n \tag{6}$$

where δ_n is called a minimization step. The step may be determined in a number of ways in different minimization algorithms. The conceptually simplest approach is the *steepest descent* method. The step is performed in the direction of fastest local decrease of U, i.e. opposite to the gradient g, as shown in Equation (7)

$$\delta_n = -\alpha_n g(q_n) \tag{7}$$

where α is a factor determining the length of the step. For multi-dimensional systems characterized by complex energy landscapes, this algorithm is rather inefficient; a relatively large number of iterations is needed to locate a local minimum. A whole family of *conjugate gradient* methods have been developed, which modify the steepest descent step in order to increase efficiency. The idea is

to make sure that the current step vector is not similar to the previous one, two or more step vectors. In the Fletcher–Powell algorithm the gradients from successive steps are used to form a numerical estimate of the second derivative matrix **F** in order to speed up convergence. Since energy and gradient evaluations are a standard part of molecular mechanics codes, all these gradient-based methods carry a low computational cost per optimization step, but require a relatively large number of steps to complete.

A basic method which uses both the first and second derivatives of U is the *Newton–Raphson* algorithm, which tries to predict the position q_{n+1} at which the gradient will be zero, using the first-order Taylor expansion shown in Equation (8)

$$g(q_{n+1}) = g(q_n) + \mathbf{F} \cdot (q_{n+1} - q_n) = 0 \tag{8}$$

which leads to Equation (9)

$$q_{n+1} = q_n - \mathbf{F}^{-1} \cdot g(q_n) \tag{9}$$

where $\mathbf{F}^{-1}$ is the inverse of the second-derivative matrix **F**. The Newton–Raphson approach leads to fast convergence of the minimization. Its drawback is the necessity for evaluation and storage of the $3N \times 3N$ matrix **F**, which makes the method difficult to apply for large molecules. More efficient alternatives involve neglecting most of the small elements of **F** (truncated Newton method) or limiting the matrix equations to a subspace of smaller dimensionality (e.g. adopted basis Newton–Raphson).[12,60]

4.3 The Multiple Minimum Problem

The basic problem of molecular modeling is that the number of minima of a molecular potential energy function (i.e. the number of stable conformations) increases exponentially with the number of atoms N. This is known as the protein folding problem or global minimum problem.[63,64] Several hundred distinct minima were found explicitly for a tetrapeptide in vacuum,[65] a cluster of 55 Lennard–Jones atoms has an estimated 10^{10} minima,[43] while the protein myoglobin explores a new potential energy minimum every 0.2 ps in a MD simulation at 300 K.[66]

A simple example may be used to illustrate the multiple minimum problem. Consider a small protein of 100 amino acid residues. Let each amino acid be able to assume only two distinct conformers. The total number of protein conformations would then be $2^{100} \approx 10^{30}$. A processor on a good modeling workstation available in the year 1999 might perform about 10 energy evaluations per second for our system. Even if we had available a large number of much faster processors, and were able to perform as many as 10^{18} energy evaluations per second, testing all the conformers to find the lowest energy one would take 10^{12} s, or about 30 000 years.

The presence of such a large number of minima leads to severe difficulties in modeling properties of large, flexible molecular systems such as peptides and proteins. For example, in *first principles* protein structure prediction the global minimum of the molecular potential energy is sought as a model of the native, folded state.[43] As the example above shows, finding the global minimum is impractical by simple systematic or random search techniques. Recent advances in specialized methods of global optimization for molecular systems are reviewed in Wales and Scheraga.[43] Further discussion of problems related to multiple minima is presented in sections 4.5 and 4.6.

There are three additional complications in protein structure prediction. First, it is not certain that the native state corresponds to a global minimum or a local minimum in which the system has been kinetically trapped. Second, considering energies only ignores entropic effects; at finite temperatures we might expect the native form to be the free energy, rather than the energy minimum. Finally, there is no certainty that currently available potentials have sufficient accuracy to discriminate between folded and misfolded protein conformations.[43] Simplified model calculations point to an important property expected of proteins that take on well defined three-dimensional structures: the global minimum should be separated from the rest of the minima by a significant energy gap.[67] Such strong discrimination between the global minimum and the closest excited states has not been demonstrated so far in simulations at the atomic level of detail, which may be attributable to the crude models of solvation used so far. Although progress is being made, accurate, systematic structure prediction for complex systems such as large clusters, crystals or proteins is still not within reach of modeling techniques. Thus, experimental structural information is still needed as input in most molecular modeling studies.

A promising family of global optimization methods is based on the idea of deforming (smoothing, coarse graining) the potential energy surface. The smoothed energy surface is simpler than the original, allowing faster conformational exploration and identification of the global minimum. The different algorithms, such as the diffusion equation method, Gaussian phase packet, Gaussian density annealing, distance scaling method, conformational space annealing are reviewed in Wales and Scheraga[43] and Straub.[68]

4.4 Normal Mode Analysis

Normal mode analysis deals with systems executing small amplitude vibrations around an equilibrium position.

Under these conditions the potential energy may be taken to be a quadratic function of the coordinates and the physical system of N interacting particles (atoms) becomes equivalent to a set of $3N$ independent harmonic oscillators. As a result, the dynamics and thermodynamics of the system may be determined in closed form. An excellent book on this topic is by Wilson et al.[1]

4.4.1 The Harmonic Approximation

For a molecule of N atoms, described by $3N$ Cartesian coordinates $x_i, i = 1, 2, \ldots, 3N$, the kinetic energy is given by Equation (10)

$$T = \frac{1}{2}\sum_{i=1}^{3N} m_i \dot{x}_i^2 \tag{10}$$

where $\dot{x}_i$ denotes the derivative of x_i with respect to time. In the vicinity of the origin of the coordinates the potential energy U may be expanded as a Taylor series in the powers of x_i. Requiring that the origin $x = 0$ be a minimum of U and choosing the zero of energy $U(x = 0) = 0$ allows the representation of U for small displacements around the origin in the form of Equation (11):

$$U(x_1, x_2, \ldots, x_{3N}) = \frac{1}{2}\sum_{i=1}^{3N}\sum_{j=1}^{3N}\left(\frac{\partial^2 U}{\partial x_i \partial x_j}\right)_{x=0} x_i x_j \tag{11}$$

This is called the harmonic approximation. Here the linear term in the Taylor expansion is zero due to the condition that U has a minimum at $x = 0$, and the higher order terms are neglected because they depend on higher powers of the small displacements x_i. The second derivatives of the potential energy with respect to atomic positions are the well-known harmonic force constants F_{ij}, already encountered in section 4.2.

4.4.2 Normal Coordinates

In principle, all the particles (atoms) in the system may interact with each other. Thus, all elements of the matrix $\mathbf{F}$ may be non-zero, with the only condition being that $F_{ij} = F_{ji}$. The harmonic potential in Equation (11) couples together all coordinates x. Writing down Newton's equation of motion in the x coordinates leads to a set of $3N$ coupled differential equations. There exists a different set of coordinates y_j, $j = 1, 2, \ldots, 3N$ in which both the kinetic energy T and potential energy V have *diagonal form*, as shown in Equations (12) and (13):

$$V(y_1, y_2, \ldots, y_{3N}) = \frac{1}{2}\sum_{i=1}^{3N} \lambda_i y_i^2 \tag{12}$$

$$T = \frac{1}{2}\sum_{i=1}^{3N} \dot{y}_i^2 \tag{13}$$

In these coordinates the equations of motion are uncoupled, i.e. they represent a system of $3N$ noninteracting quasi-particles, called normal modes.

The normal coordinates y and the original Cartesian displacements from equilibrium x are related by a linear transformation defined by a $3N \times 3N$ matrix $\mathbf{L}$, as shown in Equations (14) and (15):

$$x_i = \sum_{j=1}^{3N} L_{ij} y_j \quad \text{or} \quad x = Ly \tag{14}$$

$$y_i = \sum_{j=1}^{3N} L_{ij}^{-1} x_j \quad \text{or} \quad y = L^{-1}x \tag{15}$$

$\mathbf{L}$ is a matrix of constant coefficients, which depend, in principle, on *all* N atomic masses m_i and *all* $3N(3N + 1)/2$ harmonic force constants F_{ij}.

It is easy to show that the matrix $\mathbf{L}$ is given by[1] Equation (16):

$$\mathbf{L} = \mathbf{M}^{-1/2}\mathbf{L}' \tag{16}$$

where $\mathbf{M}^{-1/2}$ is a diagonal matrix with inverses of the square roots of atomic masses on the diagonal as shown in Equation (17)

$$(\mathbf{M}^{-1/2})_{i,j} = m_i^{-1/2}\delta_{ij} \tag{17}$$

with δ_{ij} the Kronecker symbol, while $\mathbf{L}'$ is the matrix having as columns the eigenvectors of the mass-weighted Cartesian force constant matrix $\mathbf{F}'$ ($F'_{ij} = (m_i m_j)^{-1/2} F_{ij}$). Equation (18) shows:

$$\mathbf{F}'\mathbf{L}' = \mathbf{L}'\Lambda \quad \Lambda_{ij} = \lambda_i \delta_{ij} \tag{18}$$

The eigenvalues λ_k, $k = 1, \ldots, 3N$ of the matrix $\mathbf{F}'$ are exactly the coefficients appearing in Equation (12).

4.4.3 Normal Modes

In the normal coordinate representation, the motion of the system (molecular nuclei) is a superposition of independent harmonic vibrations in each of the coordinates $y_i, i = 1, 2, \ldots, 3N$. These $3N$ types of motion are called the *normal modes* of the system. In the classical picture, the motion in the kth normal mode is a harmonic oscillation with cyclic frequency $w_k = \sqrt{\lambda_k}$, as shown in Equation (19)

$$y_k(t) = A_k \cos(w_k t + \delta_k) \tag{19}$$

where the amplitude A_k and phase δ_k are determined by the initial conditions. In the quantum picture, the energies $\hbar w_k$ represent possible excitations of the system from the ground state, observable in vibrational spectra.[1] The matrix $\mathbf{L}$ contains a description of the form of motion:

L_{ik} gives the amplitude of displacement of the Cartesian coordinate x_i upon excitation of normal mode k.

Describing vibrations in Cartesian coordinates is counterintuitive. The ''natural'' way to look at molecular vibrations is in the same way as we look at structure: in internal coordinates, i.e. deformations from equilibrium of chemical bonds, angles and dihedrals. Traditionally, vibrations of small molecules have been described in internal coordinates. Because of the difficulty of making the transformation to internal coordinates, vibrations of large molecules are traditionally described in Cartesian coordinates.

4.4.4 Anharmonic Effects

While the potential energy of a molecular system may be adequately represented by a quadratic function for small fluctuations, a full harmonic force field is needed to properly describe frequencies and forms of molecular normal modes.[1] For larger amplitude motions anharmonic terms might be needed to correctly describe structural deformations, such as the Morse function for bond stretching.[8] In comparison, the model potential in Equation (1) contains only the diagonal harmonic terms. Some force fields use more complicated forms for the bond stretching and angle bending terms (see MM3, CHARMM). However, the full harmonic potential is not employed in molecular mechanics, due to the number of parameters needed and their limited transferability. This limits the accuracy of calculated vibrational spectra, especially in the higher frequency range, corresponding to localized bond stretching and angle deformation modes. Usually, more effort is put into parameterizing low frequency vibrations, describing delocalized motions involving dihedral angle deformations. It is these motions that are of most interest, since they are related to conformational transitions in flexible molecules.[8]

4.4.5 Applications

Protein motions cannot be described by classical harmonic vibrations at low temperatures (below 100 K), where quantum corrections need to be applied, or at room temperature over long timescales (over 100 ps), where hopping between different energy minima dominates.[61] However, normal modes remain a useful tool in studies of peptide/protein structure and dynamics. Once the vibrational frequencies and normal modes are determined, the vibrational free energy and the mean atomic fluctuations may be calculated, and forms of collective motions of the systems proposed. Early studies looked at the structural fluctuations in α-helices and β-sheets (see Brooks et al.[61] for a review), as well modeling the mobilities of atoms in different parts of proteins, to compare with X-ray temperature factors.[21,22] Differences in entropy between molecular structures were estimated based on normal mode analysis, e.g. for A and Z DNA,[69] and the 3_{10} and α helices of an octapeptide.[70] Low-frequency vibrational motions have been used as approximate reaction coordinates in complex systems, e.g. the inter-domain or ''hinge-bending'' motion in lysozyme[21,71] and the α-helix $\rightarrow$ 3_{10}-helix transition.[70] Normal modes have also been employed to provide step directions in MC simulations[72] and in algorithms for long time step MD.[73] Detailed normal mode analysis of short model peptides aimed at understanding their vibrational spectra and improving model force field description of vibrations have also been performed.[74,75]

4.5 Molecular Dynamics Simulations

MD simulations involve solving the equations of motion for a system of particles moving in a known potential U. The discussion below deals with the case of Newton's equations of classical mechanics. Other equations may also be used, which incorporate some average interactions with the environment (e.g. Langevin, Brownian dynamics[61]). Newton's equations provide a systematic way to generate a series of configurations of the simulated system corresponding the classical time evolution. These configurations not only provide information on the types of structures that may be sampled by the system (averages and fluctuations), but also completely describe the detailed dynamics of structural change and energy flow within the classical model, allowing calculation of time correlation functions and conformational transition rates.

4.5.1 Numerical Solution of Newton's Equations

Newton's equations of motion for a system of N atoms can be expressed by Equation (20):

$$m_i\vec{a}_i = m_i\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2} = \vec{F}_i \quad i = 1, \ldots, N \tag{20}$$

where m_i is the mass, $\vec{a}_i$ the acceleration, $\vec{r}_i = (x_i, y_i, z_i)$ the position, and $\vec{F}_i$ the force acting on atom i, and t is the time. Given a potential $U = U(q)$ describing the potential energy as a function of coordinates q, the forces can be expressed as partial derivatives of U with respect to the atomic positions (i.e. components of the gradient):

$$\vec{F}_i = -\vec{\nabla}_i U = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right. \qquad i = 1, \ldots, N \tag{21}$$

A widely used and popular method of numerically solving Newton's equations of motion (Equation 20) with the forces (Equation 21) is the Verlet algorithm,[76] Equation (22), in which the position at time $t + \Delta t$ is calculated from the positions at $t - \Delta t$ and t, and the forces at time t:

$$\vec{r}_i(t+\Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t-\Delta t) + \frac{\vec{F}_i}{m_i}(\Delta t)^2 + O[(\Delta t)^4] \tag{22}$$

while the velocity at time t is obtained from positions at $t - \Delta t$ and $t + \Delta t$, as shown in Equation (23):

$$\vec{v}_i(t+\Delta t) = \frac{\vec{r}_i(t+\Delta t) - \vec{r}_i(t-\Delta t)}{2\Delta t} + O[(\Delta t)^2] \tag{23}$$

The above equations provide a numerical solution for Newton's equations of motion by generating positions $\vec{r}_i$, and velocities $\vec{v}_i$ for all atoms $i = 1, \ldots, N$ at times $t + \Delta t$, $t + 2\Delta t, \ldots$ Such solutions are usually called trajectories of the simulated system. Trajectories generated using the Verlet algorithm have very interesting properties similar to those of the exact solutions.[77] Other algorithms in use include velocity Verlet and leapfrog.[78]

4.5.2 The Timescales

There are two extreme timescales involved in MD simulations. The shortest timescale is the integration time step Δt, while the longest one is the total length of the simulation τ. The length of the time step is determined by the numerical algorithm used and the highest frequency motions present in the simulated system. For typical biological or organic molecules, with or without solvent, the highest frequency motions correspond to stretching and bending of bonds involving hydrogen atoms and stretching double bonds. X−H (C−H, O−H, N−H, ...) stretching frequencies fall in the 3000–3600 cm^{-1} range, X−H bending and double bond stretching (C=O, C=C, C=N, ...) in 1400–1800 cm^{-1}.[79] The period of a 3300 cm^{-1} vibration is equal to 10 fs (1 fs = 10^{-15} s). In standard approaches, to obtain an accurate numerical solution to the equations of motion the integration time step must be much smaller that this value. Thus, typical integration time steps Δt of 1 fs or less are usually used.[18,61] Elimination of the highest frequency X−H stretching motions through application of constraint algorithms such as SHAKE or RATTLE allows the use of the 2 fs time step.[78,80] Since the forces have to be evaluated at every time step, longer simulations could be performed at the same cost if Δt were increased. This goal is being addressed by several novel approaches described at the end of this section.

Ideally, the total length of the simulation τ should be at least comparable to the timescale of the phenomenon under study. For peptide and protein systems this has rarely been the case so far, and the current standard of 1–10 ns length simulations is determined mainly by the computational cost. Such trajectories are useful in predicting structures and conformational dynamics of short peptides (2–6 residues), but are of insufficient length for many phenomena in larger peptides or proteins. For example, helix formation in peptides has been observed on the 100 ns timescale, the structural changes in hemoglobin accompanying oxygen binding take place in the microsecond to millisecond range (10^{-6} to 10^{-3} s), and different stages of folding of small globular protein have been characterized on millisecond to second timescales. Special simulation techniques which enable simulation of large scale structural transitions occurring over long timescales are briefly described in sections 4.8 and 7. Difficulties of conformational exploration are compounded by the fact that not only are there multiple minima, but the energy landscape is rugged, with different regions separated by barriers with a wide range of heights. This leads to problems of kinetic trapping and the necessity to sample rare barrier crossing events.[42]

4.5.3 System Preparation

To start a simulation, we need initial coordinates and velocities for all atoms. The starting structure is usually obtained from X-ray crystallography or NMR experiments.[47,61] First, energy minimization is performed, to assure consistency of the structure with the potential energy function employed, and to bring the coordinates close to a local minimum, where forces acting on atoms are small. The next stage is heating, consisting of a series of consecutive random velocity assignments from Maxwell–Boltzmann distributions corresponding to increasing temperatures, interspersed with periods of unperturbed dynamical evolution. The purpose of this phase is to bring the system close to the desired final temperature and evenly distribute kinetic energy throughout the macromolecule. Heating is followed by an equilibration phase, when the system is allowed to evolve, with temperature (and other variables, if needed) monitored and adjusted as necessary to the desired target value. The equilibration is usually assumed to be complete when the monitored properties of the system do not change in a systematic way. In MD simulations, the temperature T is defined through the kinetic energy K, as shown in Equation (24):

$$K = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{1}{2} f k_B T \tag{24}$$

where k_B is the Boltzmann constant and f is the number of degrees of freedom. This formula is obtained using the classical equipartition theorem.[81] For a free molecule, usually $f = 3N - 6$ usually, since we eliminate the overall translations and rotations; for a system under periodic boundary conditions (PBC) $f = 3N - 3$, since rotations are no longer free. The final stage of a simulation is the generation of a trajectory which will be used to analyze

the properties of interest. The trajectory is simply a set of coordinates (and possibly velocities) of the system, providing "snapshots" of its time evolution.

4.5.4 Information Content of a Trajectory

A significant amount of information can be derived from MD simulations. Since we are solving Newton's equations of motion, the resulting dynamics represents the true classical time evolution of our system, yielding forms, amplitudes, timescales and energetics for different microscopic events, such as atomic fluctuations, side chain rotations, helix–helix reorientations, etc. which occur within the time frame of the simulation.[61] The time course of the dynamics is usually described in terms of *time correlation functions*, which are related to observable dynamical properties. Trajectory averages of dynamical quantities – structural parameters, energies and their fluctuations – correspond to the observed values of these quantities.

4.5.5 Time Correlation Functions

For a time dependent quantity $A(t)$ the autocorrelation function is defined by Equation (25):

$$C_{AA}(\tau) = \langle A(0)A(\tau)\rangle \tag{25}$$

where $\langle\cdots\rangle$ denotes averaging over the initial times $t = 0$.[81] Time correlation functions are related to a variety of observable dynamical quantities, e.g. the translational diffusion coefficient is shown by Equation (26):

$$D_{\mathrm{t}} = \frac{1}{3}\int_0^\infty \langle\vec{v}(0)\vec{v}(\tau)\rangle\,\mathrm{d}\tau \tag{26}$$

where $\vec{v}$ is the velocity of the molecular center of mass,[81] and the rotational diffusion coefficient is given by Equation (27):

$$D_{\mathrm{r}} = \frac{1}{3}\int_0^\infty \langle\vec{w}(0)\vec{w}(\tau)\rangle\,\mathrm{d}\tau \tag{27}$$

where $\vec{w}$ is the angular velocity of the center of mass.[82] Formulae for reaction rates,[83] dielectric relaxation,[84] vibrational frequencies and other spectroscopic parameters have been obtained. Examples of correlation functions are shown in Figure 2.

4.5.6 Trajectory Averages

Most experimental measurements are performed not on individual molecules, but on macroscopic samples. As a result, observed macroscopic quantities are averages, both over the molecules in the sample and over the timescale of the measurement. For systems at equilibrium, statistical mechanics provides a connection between such averages and properties of microscopic states, which are directly sampled in simulations. In the canonical ensemble, describing a system at constant number of particles N, volume V and temperature T (NVT), the probability distribution for finding the system in a microscopic state corresponding to a set of particle momenta p and positions q is given by the well known Boltzmann formula[85] shown in Equation (28)

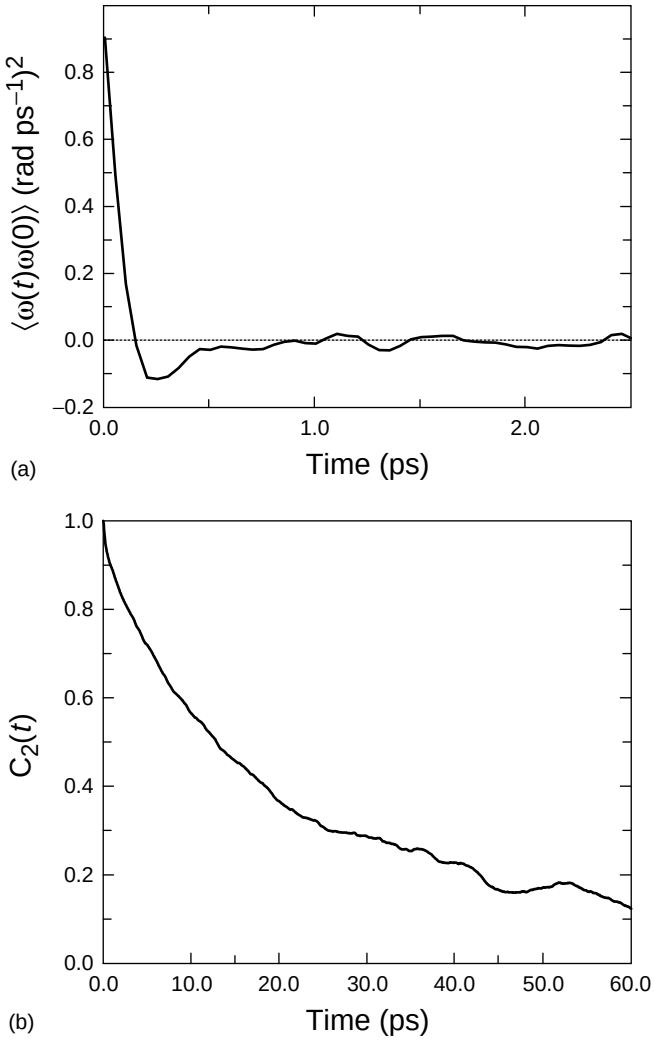

Figure 2 Description of tryptophan reorientation in aqueous solution. (a) Autocorrelation function of the center of mass angular velocity. (b) C_2 correlation function for the long in-plane axis. (Adapted from G. Jas, K. Kuczera, unpublished results.)

$$\rho(p, q) = \frac{\mathrm{e}^{-H(p,q)/k_{\mathrm{B}}T}}{\mathrm{d}p\,\mathrm{d}q\,\mathrm{e}^{-H(p,q)/k_{\mathrm{B}}T}} \tag{28}$$

and the ensemble average of any quantity $X(p, q)$ is given by Equation (29)

$$\langle X\rangle = \int \mathrm{d}p\,\mathrm{d}q X(p,q)\rho(p,q) \tag{29}$$

The probability distribution of finding the system at a given position q in the configuration space independent of the value of the momenta (velocities) is given by Equation (30):

$$\rho(q) = \frac{\mathrm{e}^{-U(q)/k_\mathrm{B}T}}{\int \mathrm{d}q\,\mathrm{e}^{-U(q)/k_\mathrm{B}T}} \tag{30}$$

which gives for an average of a physical quantity $Y(q)$ which depends on positions only, as shown in Equation (31):

$$\langle Y\rangle = \int \mathrm{d}q Y(q)\rho(q) \tag{31}$$

In the equations above $H(p,q) = K(p) + U(q)$ is the Hamiltonian, which gives the energy of the system as a function of p and q, k_B is the Boltzmann constant, K is the kinetic energy, and U the potential energy. In calculating ensemble averages we are considering an infinite number of copies of our system prepared with different initial conditions at a fixed point in time.

Ensemble averages may be directly evaluated from MC simulations of many-particle systems (see section 4.6). In MD simulations we can directly evaluate time (or trajectory) averages, defined as Equation (32):

$$\overline{X}_t = \frac{1}{t}\int_0^t X(\tau)\mathrm{d}\tau \approx \frac{1}{N_t}\sum_{k=1}^{N_t} X(t_k) \tag{32}$$

where $X(\tau) = X(p(\tau), q(\tau))$ is the value of quantity X at time τ and the final equation shows how in practice the time average is estimated as the arithmetic average over the N_t trajectory points $t_k, k = 1, \ldots, N_t$ ($t = N_t\Delta t$). Thus, in MD the time evolution of a single copy of the system is determined and X is averaged over the microstates sampled in the trajectory.

The *ergodic hypothesis* states that these two averages should be equivalent in the long time limit,[86] as shown in Equation (33):

$$\langle X\rangle = \overline{X} = \lim_{t\to\infty}\overline{X}_t \tag{33}$$

Real computer simulations can only be propagated for finite time intervals, sampling a part of available structures, and providing approximate estimates of true averages. If the finite set of configurations and momenta sampled during a simulation is able to correctly reproduce the properties of the system, it is called a representative sample. To check whether a given sample is representative, the usual way is to compare calculated and observed properties of the simulated system, monitor averages for systematic change, and compare results from separate trajectories started from different initial conditions. That configuration samples obtained by MD simulations are usually representative follows from the method of preparing the trajectory and from Equation (21). Since the forces acting on the atoms are driving them towards regions of lower potential energy U, we expect that after neglecting the initial heating and equilibration phase, the system will sample regions of relatively low U. These are the physically most meaningful configurations, which have the largest weights in the ensemble average (Equations 29 and 31). MD simulations may thus be envisioned as a systematic method of sampling system configurations. It must be underlined that due to the enormous number of states available to a large molecular system, we can never be sure that all important structures have been sampled in a simulation of finite length.

4.5.7 Equilibrium versus Nonequilibrium Dynamics

MD simulations may be used to study two general classes of motion – equilibrium fluctuations and nonequilibrium, or relaxation type dynamics. The first type of motion may be generated by the procedure described above, where the system was carefully prepared in a conformation close to the experimentally observed (and presumably biologically active) one, with a random thermal distribution of velocities. The simulated dynamics is then expected to model thermal fluctuations at equilibrium of the experimental system.

To understand the difference between equilibrium and nonequilibrium dynamics, let us consider the standard example of irreversible gas expansion.[86,87] The gas initially occupies one part of the volume of a container divided by a partition. After the partition is removed, the gas irreversibly expands to fill the whole available volume. Analogously, in simulations of nonequilibrium dynamics, we prepare a system in a nonequilibrium state and follow its time evolution as it relaxes towards a new equilibrium. This nonequilibrium initial state may have unequal distributions of kinetic[88] or potential energy.[89–92] Observations of relaxation in a macroscopic sample will provide results averaged over the constituent molecules. The individual molecules will generally exhibit somewhat different timescales and pathways of relaxation, due to differences in their initial conditions. In order to adequately simulate observed relaxation phenomena, it is not sufficient to calculate a single long trajectory, since the system will tend to reach its new equilibrium after some characteristic time. Rather, it is necessary to generate a number of independent trajectories, starting with different initial conditions, and obtain from them the average time evolution. Clearly, variations in relaxation timescales and pathways between the different

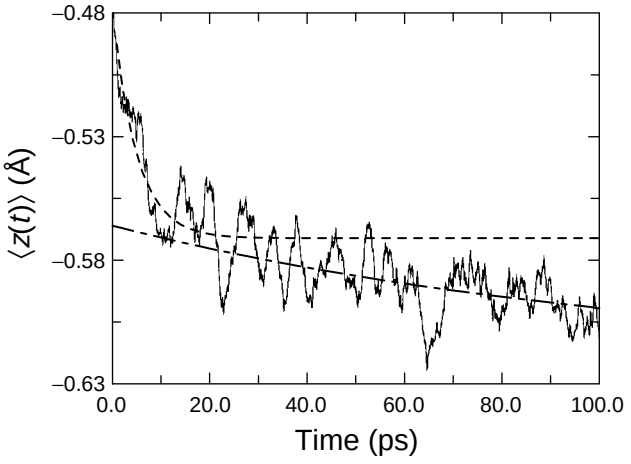

Figure 3 Time course of Fe atom displacement out of heme plane subsequent to CO dissociation from CO–myoglobin.[93] Noisy solid line – average evolution from four 100 ps MD trajectories smoothed with a 2 ps running window. Smooth solid line – fit with two-exponential decay. Dashed lines – the two-component exponentials. (Adapted from K. Kuczera et al.[93])

trajectories also constitute valuable information about properties of the studied system. Sample results of a nonequilibrium relaxation are shown in Figure 3.[93]

4.5.8 Constant Temperature and Pressure Molecular Dynamics

The simplest MD simulation protocol involves solution of Newton's equations of motion, as described above. The simulation conditions correspond roughly to the microcanonical ensemble of statistical mechanics, i.e. constant energy E, number of particles N and volume V.[18,78] For some processes, more realistic conditions would involve constant temperature T or constant temperature T and pressure P. Several different approaches have been developed that allow constant temperature and/or constant pressure simulations. These approaches involve some sort of adjustment to the kinetic energy (Equation 24) or the virial.[18,78]

4.5.9 New Molecular Dynamics Algorithms

Several approaches are being pursued in development of novel numerical algorithms for MD. Progress has been achieved in understanding the fundamental properties of the dynamics in nonlinear systems. One aspect that has emerged is that it is desirable to develop numerical algorithms which have mathematical properties of the original dynamical equations, such as being *symplectic* (conserving volume in phase space[94]) or *time reversible*.[42] Symplectic numerical algorithms display exceptional long-term stability. Thus, while high-order nonsymplectic algorithms may be formulated that give more precise solutions with smaller local energy fluctuations, the symplectic methods have smaller long-term energy drift.[73] Time reversible algorithms have become the basis of the whole family of force splitting or multiple time stepping (MTS) methods such as the reversible reference system propagator algorithms (r-RESPA).[42] In the MTS approach the forces in the system are split into several categories – fast and slow, long and short range. Each category is associated with a timescale. Thus, the fast forces (due to bond stretching and angle bending) are evaluated every time step Δt, and the slow forces (dihedral deformation and nonbonded interactions) every $m\Delta t$ steps. By choosing $\Delta t = 0.5$ fs and $m = 5$–6 integration of the equations of motion becomes more accurate, since the relatively few highest frequency terms are evaluated more often, and cheaper, since the most costly nonbonded term is evaluated less often than with a standard simulation with a single time step of 1 fs. Analogous effects are obtained by updating more often the strong but relatively few in number nonbonded interactions of an atom with its nearest neighbors and less often the weaker but more numerous interactions with more distant atoms. Combining these two subdivisions leads to significant acceleration of computations, by factors of 5–40, dependent on details of the studied system.[42] The additional advantage is that the MTS methods allow the abandonment of SHAKE constraints, together with their modifying influence on the system dynamics. Using MTS methods, a 4–5 fs limit may be reached for the time step of updating the slowest (and most costly) force component. Further extension of the time step in Newtonian dynamics appears to be difficult, as instabilities arise due to resonances with the highest frequency motions.[73] One approach of further extending the time step is based on time averaging of the forces.[95] Other approaches use normal modes to describe the high frequency motions: analytical normal vibrations combined with MTS[96] or efficient numerical normal vibrations with force extrapolation and splitting, combined with using stochastic dynamics to eliminate effects of resonance.[73]

Difficulties arise when attempts to assess accuracy of MD trajectories are made. We are simulating model systems, with potentials that only approximate real molecules. Thus, it is not clear that highly accurate simulations are really necessary. The motion of nonlinear systems is chaotic, i.e. trajectories starting from very similar initial coordinates will eventually diverge to some maximum distance.[18,73] Thus, it appears that statistical measures, such as spectral densities of dynamical variables are the most appropriate quantities for comparison.

4.6 Monte Carlo Simulations

MC methods may be defined as ways of calculating integrals using special types of random sampling.[97]

The integrals in question are of the type given by Equations (29) and (31). In molecular simulations MC methods are used to generate representative samples of possible system structures with the correct distribution corresponding to a given temperature. The averages and fluctuations of structural or energetic parameters calculated numerically over such a sample should then correspond to statistical mechanics ensemble averages on the theoretical side and to experimental observables on the empirical side. For systems with a large number of atoms N, e.g. $N = 100–1000$, numerical evaluation of the averages of the type shown in Equations (29) and (31) becomes impractical by simple grid or random approaches. In the grid method some number m of values is sampled for each coordinate involved. The total number of points (structures) sampled is then m^{3N}, which becomes prohibitively large even with $m = 2$. When random samples of structures are used, it is found that most points have a high potential energy U and do not contribute significantly to the average (Equation 31) due to a low Boltzmann factor.[97]

4.6.1 Markov Chains

The MC solution to these difficulties is to create samples of structures that are partly correlated – so-called *Markov chains*, which may be defined in the following way. Label the possible states of the system (structures): $q_1, q_2, \ldots, q_i, \ldots$ Consider a random walk over the states, which reaches state q_i at step k. If the probability P_{ij} of reaching a given state q_j in the next step $(k+1)$ of the walk depends only on the current state q_i (and not on the states reached at previous steps $k-1, k-2, \ldots$), then the random walk is called a Markov process.[98] The set of states explored in a Markov process is called a Markov chain. The advantage of Markov chains is that the transition probability P_{ij} may be adjusted so that a large fraction of sampled points correspond to large weights in the target average.

4.6.2 Metropolis Monte Carlo

In the widely used Metropolis MC method for the algorithm of generating the Markov chain, we start with the system in state q_i at step k. The current system configuration q_i is then perturbed in a random way to generate a new *trial configuration* q_i'. The potential energies $U(q_i)$ and $U(q_i')$ are compared. If the trial configuration has a lower energy, it is accepted, i.e. the state of the system at step $k+1$ is set to q_i'. If the trial configuration has a higher energy than the current one, then the trial configuration is accepted with a probability P and rejected with probability $1-P$. Rejection of the trial step means that the state of the system at step $k+1$ is set to q_i again. The probability P is given by Equation (34):

$$P = \frac{\mathrm{e}^{-U(q_i')/k_\mathrm{B}T}}{\mathrm{e}^{-U(q_i)/k_\mathrm{B}T}} = \mathrm{e}^{-[U(q_i')-U(q_i)]/k_\mathrm{B}T} \tag{34}$$

As long as the perturbations used to generate trial configurations obey some simple conditions, it can be shown that in a sufficiently long simulation the ratio of probabilities of finding any two structures q_i and q_j in a sample generated by the Metropolis MC method is given by Equation (35):

$$\frac{P(q_i)}{P(q_j)} = \mathrm{e}^{-[U(q_i)-U(q_j)]/k_\mathrm{B}T} \tag{35}$$

i.e. it is equal to ratio of Boltzmann factors, the same as would be expected in the canonical ensemble. Due to this property we can calculate statistic mechanical averages directly from the sample of structures generated in the MC simulation. Equation (36) shows that for any quantity $Y(q)$ which is a function of particle positions only

$$\langle Y \rangle = \frac{1}{M}\sum_{i=1}^{M} Y(q_i) \tag{36}$$

The MC simulation yields a sample of structures from which we can obtain statistics. The method of generating the Markov chains assures us that a large fraction of sampled conformations correspond to high Boltzmann factors, where the integrand is appreciably different from zero.

4.6.3 Convergence

Theoretically, infinitely long MD or MC simulations will provide the complete classical characterization of the system with a given model potential $U(q)$. In a finite simulation we never explore all possible conformations or all possible motions/fluctuations, since we do not know what structures or motions have been missed, and can never be completely certain that we have a representative sample of microscopic states which can be used to characterize the observable macroscopic properties. In finite simulations the choice of the starting point and the length of the simulation become important. Especially difficult is the problem of sampling the rare barrier crossing events, corresponding to high relative values of U and low Boltzmann factors.

4.7 Conformational Search

Conformational search aims to determine the lowest energy structures of flexible molecular systems. In studies of biological and organic molecules the chemical bond lengths and angles do not vary significantly between

conformers; the search may be limited to exploring dihedral angles.[8,99] In the case when aggregation of two or more molecules is being studied (docking, crystal structure prediction), the relative distances and orientations between the species has to be taken into account.[43,100]

For systems with $n = 1–3$ flexible dihedrals it is possible to provide a full description of conformational space using adiabatic mapping. In this approach energy minimization is performed from starting points on a grid covering the full range of the angles. In each minimization the mapped variables are constrained to remain close to the starting point. The optimized structure corresponds to the lowest possible energy of the system for the given values of the mapped variables. An example of this approach is the alanine dipeptide, which is used as a test case for parameterizing peptide/protein potentials,[49] for which the $\phi–\psi$ map is shown in Figure 4. For larger systems, with $n = 5–10$ flexible dihedrals systematic, or grid search may easily be performed. If each of the explored coordinates is allowed to take on m values, m^n minimizations have to be performed. This number may be significantly reduced by rejecting whole families of starting structures due to unfavorable strain or nonbonded clashes (pruning).[99,101] For medium-sized molecules the random search method may also be effectively applied, in which geometry optimizations start from successive randomly generated conformations.[8] Random searches in Cartesian and dihedral space have been implemented, and applied to both linear and cyclic molecules.[101]

The above approaches quickly become impractical for large molecular systems, as the number of possible conformations to be explored becomes too large and the fraction of low energy structures among all possible structures decreases. For large systems, methods based on MD and MC simulations are often used.[8] These methods are based on the fact that both MD and MC simulations explore structures of low energy (relative to neighboring conformations). For small systems the *direct* approach may be used, i.e. the configuration samples from MD or MC simulations are used as representative structures of the system directly (or after energy minimization). Owing to the nature of the potential energy surface, exploration of new conformations is slow in simulations at room temperature and for larger molecules the simulated annealing procedure is used. Simulated annealing starts with an MD or MC simulation run at an elevated temperature (as high as 500–1500 K) in order to enhance the rate of conformational exploration. Selected structures from the high temperature trajectory are then cooled to room temperature, and energy minimized. Simulated annealing works well for small peptides. The rate of conformational space exploration is still insufficient to effectively predict stable structures of larger peptides or proteins. The difficulties of including solvation effects in conformational search are described in section 4.9.

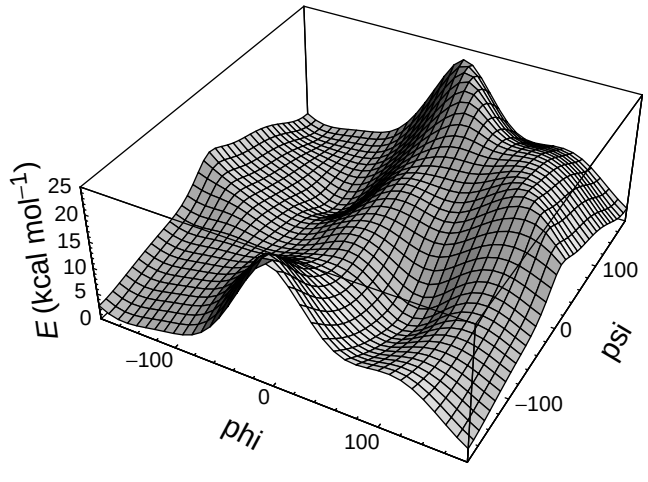

Figure 4 Adiabatic $\phi - \psi$ map of the alanine dipeptide. (Produced using CHARMM program and parameters.[12,49])

Improved methods being developed to enhance the rate of conformational sampling include genetic algorithms, locally enhanced sampling (LES), hybrid Monte Carlo (HMC), J-walking, S-walking, using Tsallis statistics and multicanonical sampling simulations. The genetic algorithm approach is modeled on the evolution of genetic material.[102] An initial pool of conformations is generated, with the structural information encoded in bit strings, called chromosomes. The population evolves by reproduction, crossover and mutation, with survivability measured by the energy. Ideally, the final result is a population consisting of a single chromosome representing the global minimum. In the LES method n noninteracting copies of a molecule are present simultaneously, with their potential energies weighted by $1/n$.[103,104] As a result, multiple conformations are sampled simultaneously and transition barriers are lowered. The HMC approach uses MD methods to generate large MC moves with high acceptance ratios in arbitrary systems.[44,105] J-walking and S-walking algorithms expand the repertoire of MC to include occasional jumps to conformation states corresponding to a higher temperature, thus enabling faster barrier crossing. The Tsallis statistics[106] and multicanonical sampling[107,108] methods increase the rate of conformational transitions in MC or MD simulations by modifying the underlying distribution of states from the relatively narrow Boltzmann distribution to some wider form.

4.8 Free Energy Simulations

4.8.1 Introduction

The free energy is a fundamental quantity in chemistry and physics. Free energy differences between different

molecular states are related to equilibrium constants (through thermodynamics) and kinetic rates (through transition state theory). In recent years computational methods have been proposed which make possible calculation of free energy changes from computer simulations of molecular systems. For reviews of methods and applications refer to the literature.[61,109–112] Typical biomolecular applications for which free energy simulations have met with success involve two kinds of processes – structural transitions and chemical change. In structural transitions the chemical nature of the system does not vary, and the process is followed by evaluating the free energy profile PMF (potential of mean force) along a selected reaction coordinate, e.g. a dihedral angle for a conformational transition, a distance between two species for aggregation. The PMF provides direct information on the free energy differences between various stable conformers as well as about the free energy barriers that separate them. These data can then be converted into population ratios and transition rates. In simulations of chemical change, or mutations, the free energy change of substituting one chemical group with another in a condensed phase system (crystal, solution, protein) is evaluated. This allows predictions of such effects as influence of point mutations on thermal stability and ligand binding of proteins, or role of different substituents in determining the affinity of an inhibitor for a target enzyme.

Free energy simulations have two main advantages relative to standard simulation approaches. First, free energy values are obtained, rather than just energies. The free energies are directly related to experimental observables – binding affinities, population ratios, rate constants. Second, the free energy differences obtained may be much more precise than energy values. This follows from some special properties of the evaluation algorithms, as described below. Additionally, the structural transition simulations provide a way to partially overcome the conformational sampling problem. Namely, using free energy simulations of feasible length, the system may be forced to overcome quite high barriers along a reaction coordinate, effectively reaching states that are practically inaccessible in a direct simulation. As with all modeling approaches free energy simulations are able to provide important insights and improved understanding of complex macromolecular processes, such as ligand binding, folding, aggregation, solvation and mutation by decomposing the total free energy change into contributions from solvent, protein and even individual residues, chemical groups or atoms. Free energy simulations also share some shortcomings with the other modeling techniques. Obviously, any simulation, free energy or otherwise, is as reliable as the underlying potential function. To obtain reliable estimates of free energies, representative samples of system structures still have to be generated, which is difficult due to the problem of multiple minima. Although sampling along the reaction coordinate may be accelerated, exploration of conformational space in the orthogonal direction proceeds at the same pace as in standard MD or MC simulations.

Fundamentally, the free energy is related to the statistic mechanical partition function, i.e. the sum of weights of all possible states.[85] To circumvent the difficult problem of calculating absolute value partition functions, free energy simulation methods concentrate on calculating their ratios, which yield the differences between free energies of thermodynamic states for molecular systems. Since free energies are functions of state, their changes can be evaluated along any path connecting the two states under consideration; the path can be chosen so as to make calculations easiest. The Gibbs free energy G is the appropriate function under constant temperature and pressure, while the Helmholtz free energy A is used at constant temperature and volume. In condensed phase systems, such as proteins or peptides in solution or crystal, we expect that $\Delta G \approx \Delta A$.

4.8.2 The Coupling Parameter

The three main approaches to free energy calculations are thermodynamic perturbation (TP), thermodynamic integration (TI) and umbrella sampling. In TI and TP a coupling parameter λ is introduced to construct a hybrid potential energy function $U(q, \lambda)$. For conformational processes the parameter λ is an actual structural coordinate, e.g. a distance between two atoms or a dihedral angle; $U(q, \lambda)$ then represents the potential energy corresponding to a fixed value of the chosen coordinate, with all other coordinates q allowed to vary. For mutation processes, in which the chemical structure of the system is modified, $U(q, \lambda)$ mixes the properties of the two molecular states in such a way that varying λ from 0 to 1 changes the potential energy function from that of the initial state U_0 to the final state U_1.[113] In this case the intermediate states are artificial constructs having a potential i.e. a mixture of the initial and final state potentials. The simplest way to do this is the linear coupling approach shown in Equation (37):

$$U(\lambda) = (1 - \lambda)U_0 + \lambda U_1 = U_0 + \lambda(U_1 - U_0) \quad (37)$$

Simulations are performed for a series of values of the coupling parameter λ; each such simulation is called a step or window.

4.8.3 Thermodynamic Perturbation

In the TP approach the free energy difference between states corresponding to $U(q, \lambda')$ and $U(q, \lambda'')$ we use

Equation (38):[109,110,113]

$$\Delta G(\lambda' \rightarrow \lambda'') = -k_B T \ln \left\langle e^{-\beta[U(\lambda'')-U(\lambda')]} \right\rangle_{\lambda'} \quad (38)$$

where the angled brackets signify a statistical ensemble average for a system with potential energy $U(q, \lambda')$. As usual, this is estimated numerically by a trajectory average of an MD simulation (or sample average of an MC simulation) over the window corresponding to $U(q, \lambda')$. The total free energy change is calculated using Equation (39) by accumulating the differences between windows so as to span the whole interval between 0 and 1:

$$\Delta A = \sum_k \Delta A_k \quad (39)$$

The formula in Equation (38) involves averaging over all possible configurations, or, equivalently over infinitely long trajectories (Equation 33). In simulations we always deal with limited samples and finite trajectories. With sufficiently careful preparation, involving choice of starting structure, energy minimization, heating and equilibration, a representative sample of configurations of the initial state can usually be generated by MC or MD simulations. In order to obtain an accurate estimate of the average from the TP method, Equation (38), these same configurations also have to be representative for the final state. To assure this condition, in practical applications the perturbation from the initial to the final state is not performed in a single step, but is broken up into several stages, with introduction of one or more intermediate states. This is called multistep or multiwindow TP. It should be noted that we only need to have simulations for the starting values (λ') of the coupling parameter in Equation (38), while the end-points (λ'') are arbitrary. Thanks to this property, simulations for the initial ($\lambda = 0$) and final ($\lambda = 1$) states are not necessary to calculate ΔG. Another consequence is the possibility of using the double-wide sampling technique, in which one evaluates the free energy changes half a step forward and half a step backward in λ.[23] This procedure reduces the size of the perturbation and leads to more accurate free energy differences. If the initial and final surfaces are approximately parallel, we can expect to obtain good estimates of the free energy change even for large values of ΔG and relatively small samples of configurations.[114] Novel approaches which use special ''soft potentials'' enable evaluations of many perturbations from a single simulation.[115]

4.8.4 Thermodynamic Integration

In the TI approach the derivative of the free energy with respect to the coupling parameter λ is evaluated. Equation (40) shows that for mutational processes, when λ is an additional external coordinate:[109,110,113]

$$\frac{\partial G(\lambda)}{\partial \lambda} = \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (40)$$

i.e. the derivative of the free energy with respect to the coupling parameter λ is equal to the average of the corresponding derivative of the hybrid potential over the ensemble for the system with potential $U(\lambda)$. For conformational processes, for which λ is an internal coordinate which is also involved in defining the system configuration we obtain Equation (41):[116,117]

$$\frac{\partial G(\lambda)}{\partial \lambda} = \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} - k_B T \left\langle \frac{\partial \ln J(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (41)$$

where the averages are evaluated over configurations with a fixed, definite value of λ. According to Equation (41), the derivative of the free energy with respect to the conformational coordinate involves the averages of two terms: the derivative of the potential with respect to the reaction coordinate λ and a term involving the Jacobian of the coordinate transformation between λ and atomic Cartesian coordinates. The Jacobian term is easily evaluated analytically for distance coordinates[116] and has been found to be negligible for dihedral angle coordinates.[117] Proceeding analogously as in TP, simulations are performed for several intermediate states or windows with different values of λ_i. These simulations yield samples of configurations over which we can estimate the averages involved. The desired ΔG between the final and initial states is obtained using Equation (42) by numerical integration of the derivative

$$\Delta G = \int_i^f \frac{\partial G(\lambda)}{\partial \lambda} d\lambda \approx \sum_k \left(\frac{\partial G(\lambda)}{\partial \lambda} \right)_k d\lambda_k \quad (42)$$

where the summation is over the windows k, and the integral bounds are $i = 0$ and $f = 1$ for mutation processes, and $i = \lambda_0$ and $f = \lambda_1$ (the initial and final values of the reaction coordinate) for conformational processes. While for mutation processes the quantity $G(\lambda)$ is just a computational device for the calculation of ΔG, for conformational processes $G(\lambda)$ has physical meaning – it is the conformational free energy, or PMF along the reaction coordinate λ.[116,117]

An advantageous property of the TI method is that Equation (42) is linear in potential energy. This allows free energy changes to be decomposed into contributions from the different terms in the potential energy function (bond stretching, angle and dihedral deformation, van der Waals and electrostatic) as well as different parts of the system (solvent, protein, donor and acceptor chains, individual residues).[26] Decompositions allow insight into the microscopic details of the studied

phenomena. Such decompositions are not rigorous. Only the overall free energy is a thermodynamic function of state and corresponds to an experimental observable. The values of the different components may depend on the force field parameterization and on the path connecting the initial and final thermodynamic state. Nevertheless decompositions are widely used to provide qualitative insight into the microscopic mechanism of the studied phenomena. An additional property of the TI method is that it is completely local, i.e. we only have to consider whether the simulations provide representative samples for the current value of λ (whereas in TP the sample also has to be representative for the neighboring windows). In practice, comparable results are obtained when TP and TI approaches are used to extract free energy values from the same data set.[26,118]

4.8.5 Umbrella Sampling

The umbrella sampling method differs from TI in that the reaction coordinate is not fixed, but restrained to remain in a limited range of values by a constraint potential $U_c(\lambda)$.[97,119] Thus, the simulations are performed in an effective potential U^*, given by Equation (43)

$$U^* = U(q) + U_c(\lambda) \tag{43}$$

where $U(q)$ is the molecular potential function and the constraint potential U_c is some (usually simple) function of the reaction coordinate λ, given by Equation (44):

$$U_c = k(\lambda - \lambda_{ref})^2 \tag{44}$$

where k is a force constant and λ_{ref} is a reference value around which we wish to sample values of λ. Simulations with U^* are employed to calculate the corresponding biased distribution $P^*(\lambda)$ of the reaction coordinate in the vicinity of λ_{ref}. This distribution is then used to calculate the PMF or conformational free energy $G(\lambda)$ in the unbiased potential U by Equation (45):[97,119]

$$G(\lambda) = -k_B T \ln P^*(\lambda) - U_c + \text{const} \tag{45}$$

The arbitrary constant in the equation above is different for every simulation window. Values of the constants are determined from the condition that probability densities from neighboring windows overlap.[119] By calculating a series of windows with different values λ_{ref} which have overlapping distributions $P^*(\lambda)$, the free energy profile (or PMF) $G(\lambda)$ for the process corresponding to the motion of the system along the reaction coordinate may be generated.

One additional difficulty appears for the mutational processes. The potential energy parameters refer to a given electronic state of a fixed covalent chemical structure, and we cannot use classical potentials to calculate electronic energy differences between such different structures. This difficulty can be overcome through the use of thermodynamic cycles.[120] Calculating the difference $\Delta\Delta G$ of the effects of the same mutation process in two different environments, such as protein interior and solution, the unphysical terms are cancelled out, leaving only the physical contributions from the changed interactions with the environment.

An interesting feature of free energy simulations is that the free energy values obtained are more precise than the corresponding energy values. This appears to be due to the fact that only those energy terms that directly vary with the coupling parameter/reaction coordinate λ contribute to Equations (38), (40), (41) and (45). Also, it is found in practice that the accuracy of components is usually lower than that of the total free energy change. This is true both for the unphysical components, such as electrostatic or solvation free energies, but also for thermodynamic functions such as entropy S and enthalpy H when extracted from $\Delta G = \Delta H - T\Delta S$.

4.9 Solvation and Environmental Effects

Peptides, proteins and other biomolecules realize their functions in a condensed phase molecular environment, such as aqueous solution, lipid membrane or macromolecular aggregate. An example of the influence of solvent on biomolecular properties is the recognition of the hydrophobic effect (i.e. clustering together of nonpolar groups to exclude water) as the major driving force in protein folding.[64] Current molecular simulations employ a host of methods to include the influence of the external molecular environment (the ''solvent'') on the modeled system of interest (the ''solute'').[99,121]

4.9.1 Explicit Solvent Methods

Conceptually simplest are explicit solvent methods. Here the solvent is treated in full atomic detail, on par with the solute. On the level of the model potential U (Equation 1), additional terms appear, describing internal deformations of the solvent molecules as well as solute–solvent and solvent–solvent nonbonded interactions. Within the explicit solvent approach the problem of dealing with the system boundary may be dealt with in several ways.[18,61] An elegant solution is to use PBC, in which the simulated system is placed in a space-filling cell (the primary cell) surrounded by translational copies of itself (images). Image atoms within the nonbond cutoff distance of a primary atom are included in calculating the nonbonded interaction energy. The image atoms thus provide a molecular environment for primary atoms close to the cell surface. Since image atoms are not independent entities but translational copies of primary atoms, this procedure effectively introduces correlations

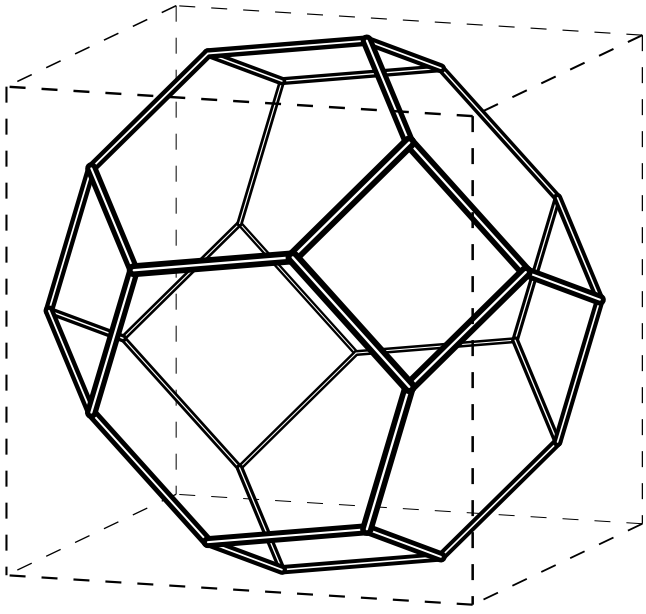

Figure 5 Illustration of PBC: the truncated octahedron space-filling cell.

between primary atoms at opposing cell surfaces. To minimize this correlation, in cutoff-based methods the minimum image convention is usually applied, in which each primary atom does not interact with itself and interacts with, at most, one image (the closest) of any other atom in the primary cell.[18] The most widely used are the cubic and truncated octahedral cell, although other shapes have been proposed.[99,122] An example of a simulation cell is presented in Figure 5. In the cubic cell, the minimum image convention is satisfied when the nonbond cutoff is less than half the cell size. The advantage of the PBC approach is that a realistic molecular environment is generated for the primary cell. By making the primary cell large enough, the effect of spatial correlations introduced by the PBC on the structure and dynamics of the central solute may be minimized. Additionally, shaping the system into a space filling cell with PBC is a starting point for some more advanced simulation algorithms – constant pressure simulations, Ewald summation and the fast multipole method (FMM). The main disadvantage of explicit solvent simulations with PBC is the computational cost. The atoms included in nonbonded force calculations are those in the primary cell and a shell of image atoms on the surface, of the thickness equal to the cutoff distance. In typical solvated peptide/protein simulations, the solute, which is the focal point of our interest, makes up at most 10% of the system atoms. Since the nonbonded term calculations are the largest component of energy and force evaluation, this means that a majority of the computational time is spent computing solvent–solvent interactions. A number of approaches to overcome this difficulty have been proposed, as described below.

The simplest solution is to use a droplet or cluster solvent model, in which the solute is placed in a sphere of solvent. Adding a boundary potential enables control of system pressure and shape.[61] Using this scheme a 3 ns simulation of the protein calmodulin was recently performed. The system included the 148 amino acid residues of calmodulin with four bound Ca^{2+} ions, 6 Cl^{-} and 22 Na^{+} ions in a water sphere of 44 Å radius.[123] Significant reorganization of the protein structure compared to the starting X-ray structure was seen in the simulation, including deformation of the central helix and reorientation of the N- and C-terminal domains.

The stochastic boundary molecular dynamics (SBMD) method was originally proposed as a tool for performing cost-effective simulations of enzyme active sites.[61,124] The starting point is a full protein, partitioned into a central region of interest (e.g. the active site), called the reaction zone, with the rest of the protein assigned to the reservoir region. The atoms in the reservoir region are either neglected or kept at fixed positions to generate a static background field. The reaction zone is solvated by introducing solvent molecules in to any available space, and is further divided into a central part – the reaction region and the surface layer – the buffer region. All atoms within the reaction region undergo standard Newtonian dynamics. Atoms in the buffer region experience forces due to the molecular potential energy function as well as some additional terms that compensate for the lack of external environment. To avoid major structural changes in the reaction zone, protein atoms in the buffer region are restrained to their initial positions by harmonic constraint potentials while solvent molecules are restrained to remain within a given distance of the center by a mean field spherical potential. To mimic the effect of collisions with molecules from the surroundings, all buffer atoms undergo Langevin dynamics, which includes a friction force and a random force.[61] The surface constrained all-atom model is an approach similar in spirit to SBMD.[30] The SBMD approach has been used to study local dynamics in several proteins, including ribonuclease A[125] and hemoglobin.[26] An illustration of a system set up for SBMD simulation is shown in Figure 6.[126,127]

4.9.2 Langevin Dipoles–Protein Dipoles

Intermediate between explicit and continuum solvent models is the Langevin dipoles–protein dipoles model of Warshel.[30] The solute is placed in a cavity inside a cubic grid of rotatable dipoles, with the grid spacing and average polarizability of the dipoles obtained from atomistic

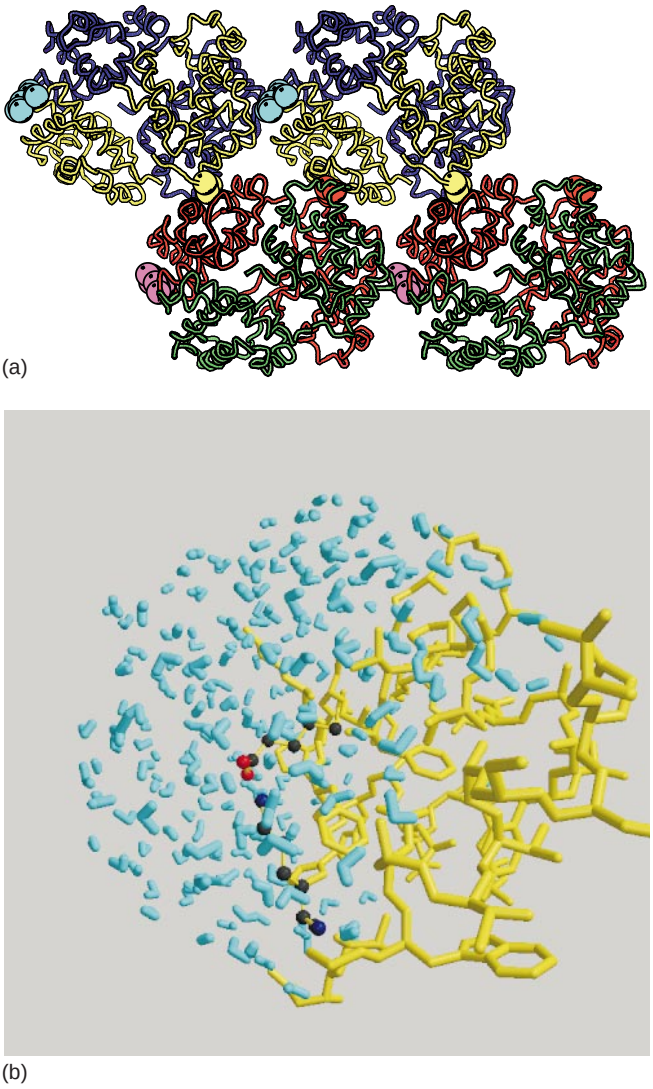

(a)

(b)

Figure 6 (a) Intermolecular contacts in sickle-cell hemoglobin (HbS). Four HbS tetramers shown, each with two α and two β chains, designated by different colors. Yellow and red spheres show sites of the sickle-cell mutation β6Glu $\rightarrow$ Val, located at the interface between two Hb strands. Blue and purple spheres show sites of β121Glu $\rightarrow$ Gln mutation, located at interface of two hemoglobins within the same strand and responsible for aggravated sickling. (b) The simulation system for performing the mutation of the β121Glu $\rightarrow$ Gln at the surface of an isolated hemoglobin. Structures from PDB file 1 HbS.[126] (Adapted from K. Kuczera.[127])

simulations of the solvent being modeled (usually water). An iterative procedure is used to calculate self-consistent induced dipoles on the grid and on the solute (''protein'') atoms, due to the electric field of the fixed solute charges and the other dipoles. The resulting solvation free energy captures the key features of the electrostatic term of solute–solvent interactions, while significantly reducing the computational cost.

4.9.3 Continuum Models

In continuum models the solvent is treated as a bulk dielectric medium. The calculated quantities are related to the solvation free energy ΔG_{sol}, i.e. the free energy of transferring a molecule from the gas phase to dilute solution. ΔG_{sol} is usually taken to be the sum of two terms as shown in Equation (46)[128]

$$\Delta G_{\mathrm{sol}} = \Delta G_{\mathrm{el}} + \Delta G_{\mathrm{np}} \tag{46}$$

ΔG_{el} is the electrostatic contribution, which is the most important factor for polar or charged solutes in polar solvents, such as proteins and peptides in water. ΔG_{np} describes van der Waals solute–solvent interactions, and reorganizing the solvent molecules to form the cavity which the solute occupies. The starting point of the electrostatic solvation free energy calculations are the Born formula, Equation (47), for a point charge q:[129]

$$\Delta G_{\mathrm{el}} = -\frac{q^2}{2a}\left(1 - \frac{1}{\varepsilon}\right. \tag{47}$$

and the Onsager formula, Equation (48), for the point dipole μ:[130]

$$\Delta G_{\mathrm{el}} = -\frac{\varepsilon - 1}{(2\varepsilon + 1)a^3}\mu^2 \tag{48}$$

at the center of a spherical cavity in a medium of dielectric permittivity ε. For more complicated charge distributions inside a spherical cavity the reaction field method of Kirkwood may be used.[131] To treat molecules of arbitrary shape, two main methods are currently in use – the generalized Born (GB) and PB equations. In the GB method the solute atoms are assigned individual charges and radii. ΔG_{el} is evaluated as the sum of Born-type and interaction terms.[99,132]

In approaches using the PB equation, the system is represented as a low dielectric cavity (the solute) in a high dielectric medium (the solvent), including the presence of mobile ions.[128,133] The PB equation shown in Equation (49)

$$\nabla \cdot [\varepsilon(\vec{r})\nabla\phi(\vec{r})] - \varepsilon(\vec{r})\kappa^2(\vec{r})\sinh[\phi(\vec{r})] = -4\pi\rho(\vec{r}) \tag{49}$$

determines the electrostatic potential $\phi(\vec{r})$ in the presence of a given spatial distribution of dielectrics $\varepsilon(\vec{r})$ and fixed charge density $\rho(\vec{r})$; $\varepsilon\kappa^2 = 8\pi e^2 N_{\mathrm{A}} I/k_{\mathrm{B}}T$ is determined by the ionic strength of the bulk solution I, with e the electronic charge, N_{A} Avogadro's number, k_{B} the Boltzmann constant and T the temperature. In PB calculations the solute atoms are assigned van der Waals radii and partial charges, both of which may be taken from model force fields or separately parameterized for this purpose.[128] The atomic radii are used to define a contact surface between the solute and a probe solvent molecule.

Regions inside the surface are assigned a low dielectric permittivity (typically $\varepsilon \approx 2$–4), while outside ε is set to the permittivity of the bulk solvent (e.g. ≈80 for water). Numerical solutions of the PB equation give the potential $\phi(\vec{r})$ and the electric field $\vec{E} = -\nabla\phi$ on a grid in space. Visualizing the electric field with molecular graphics tools has yielded valuable information that is not evident from the protein structure alone.[128] The calculation of electrostatic energies for macromolecules in different environments has allowed the prediction of changes of pK_a values of ionizable groups in proteins, influence of salt concentration on nucleic acid conformations and ligand binding, and the electrostatic components of solvation free energies ΔG_{el}.

Equation (50) shows that the nonpolar/cavity contribution to the solvation free energy is usually approximated by a term proportional to the total accessible area of the solute A[128]

$$\Delta G_{np} = \gamma A + b \tag{50}$$

where b is a constant (usually small) and γ is a microscopic surface tension term. Fitting the above expression to experimental data for small organic molecules gives $\gamma = 6\,\text{cal mol}^{-1}\,\text{Å}^{-2}$ and $b = 0$ for vacuum to water transfer and $\gamma = 19\,\text{cal mol}^{-1}\,\text{Å}^{-2}$ and $b = 0.6\,\text{kcal mol}^{-1}$ for cyclohexane to water transfer.[134] Incorporation of continuum models into standard molecular modeling programs is currently under way.[135–137]

An especially simple solvent model has been developed by Eisenberg and McLachlan.[138] Equation (51) shows:

$$\Delta G_{sol} = \sum_{\text{atoms},k} a_k S_k \tag{51}$$

where S_k are the solvent accessible surface areas of solute atoms and a_k are parameters obtained by fitting to experimental data.

4.10 Treatment of Electrostatic Interactions

The electrostatic term in the potential energy describes interactions between the charge distributions on neighboring molecules. In the Born–Oppenheimer approximation, this molecular charge distribution is represented by point charges at the nuclei and a continuous function $\rho(\vec{r})$ giving the probability of finding an electron at point $\vec{r}$. Most current force fields use Equation (1) which expresses electrostatic energy in terms of interactions between point charges located at atoms. This approach involves several approximations. First, only charges are used in describing the molecular charge distribution, and higher multipole moments (dipoles, quadrupoles, ...) are ignored.[99] Second, the charges are assigned to atoms and are fixed, independent of conformation of the molecule or the nature of the surrounding environment. Third, the interaction is truncated at some cutoff distance to enable faster nonbonded calculations. The truncation at reasonable distances of 10 Å or higher does not greatly influence the van der Waals energy, which has terms varying as $1/R^6$ and $1/R^{12}$ and thus quickly decays to zero. However, the $1/R$ electrostatic interaction energy decays quite slowly; for two proton charges separated by 10 Å the energy is $33\,\text{kcal mol}^{-1}$, and even at 100 Å it is still significant at $3\,\text{kcal mol}^{-1}$.

Because of the presence of both nuclear charges and high electronic charge densities at the nuclei, it is natural and convenient to consider nuclei as centers for placing point charges. The disadvantage of such an approach is that it does not account for the asymmetric distribution of charge around atoms in molecules. Generalized schemes for placing charges off atomic nuclei have been proposed to study specific effects such as the structure of liquid water,[139] aromatic–aromatic interactions[140] or accurate modeling of hydrogen bonds.[111] Distributed multipole analysis, in which charges, dipoles, quadrupoles, etc., are assigned to an arbitrary set of centers is a systematic approach to improved description of the details of molecular charge distributions.[141,142] Such models suffer from technical difficulties in parameterization and energy/force calculation, and they have not been widely accepted. Attempts to include the effect of charge on molecular conformations have also been undertaken.[143–145]

An important effect in condensed phases is polarization, which is described by the induction of an additional dipole moment μ_{ind} in a selected molecule by an external electric field E generated by all the other molecules in the system,[84] as shown in Equation (52):

$$\mu_{ind} = \alpha E \tag{52}$$

where the coefficient α is the polarizability. In the most general case α is a tensor quantity. Due to polarization, the effective charge on a given atom will vary with environment. Introduction of polarization into molecular potentials leads to several difficulties – parameters describing polarizabilities have to be introduced and the simple pair-additivity of the potential is lost, because E depends on positions of all atoms. Simple force fields usually include an average correction for polarizability. For example, the TIP3 water model[139] has a dipole moment equal to the average molecular dipole moment in liquid water, which is about 15% larger than in vacuum. As a result, the average water–water interactions in liquid are correctly reproduced, but the energy of the water dimer in vacuum is exaggerated. Several potentials with explicit polarization have been introduced. The results indicate that for polar molecules and monovalent

ions the polarization term is small, contributing 10–20% of the overall interaction energy and that the average correction approach should work acceptably. However for divalent ions, which generate higher electric fields in their vicinity, the polarization correction is expected to be significantly larger.

Even when the nonspherical distribution of charge around atoms, dependence of atomic charges on conformation, and polarization are taken into account, classical potentials still do not include the charge transfer effect, i.e. the fact that when molecules are brought together, electron density may flow from one molecule to another. In the final analysis, very high level quantum mechanical calculations have to be employed to correctly describe all aspects of nonbonded interactions.

Several approaches have been introduced to eliminate the need for the problematic truncation of electrostatic interactions. The whole class of continuum solvent models is discussed in section 4.9. Two powerful methods are discussed below – Ewald summation and fast multipoles. The Ewald summation method is a natural extension of PBC simulations.[146,147] An infinite crystal lattice is constructed from the primary simulation cell. A ''neutralizing charge distribution'' is introduced, which surrounds each atomic point charge by a diffuse charge distribution (usually a spherical Gaussian function) with the same total magnitude but opposite sign. The electrostatic interaction energy of the primary cell is split into two contributions – from the neutralized charges and from a second charge distribution, which cancels the first neutralizing charge distribution. The first term converges quickly with distance, and is evaluated by a direct sum over neighboring atoms within a cutoff distance. The second term converges quickly in reciprocal space and is evaluated by a finite sum involving the Fourier transform of the charge density.[99,146,147] The disadvantage of the Ewald method is that it breaks the minimum image convention and tends to reinforce artifacts due to the PBC (e.g. opposing charges separated by half the cell size will have a zero interaction energy). Current implementations use the particle mesh Ewald (PME) algorithm, which redistributes the atomic charges on a grid and uses the fast Fourier transform to accelerate evaluation of the reciprocal space sum.[147]

An alternative approach to summing up long-range interactions is the FMM.[148] In FMM space is divided into cubic cells, and multipole moments are calculated for each cell. While normal atomic pairwise summation is used at closer distances, the interaction energies of particles which are more than one cell away are evaluated using local Taylor expansions of the multipole potentials. Grouping together neighboring cells is employed to further speed up calculations.

The PME and FMM method become quite efficient for large systems The cost of direct summation of pairwise interactions scales like $O(N^2)$ with the number of atoms, since there are $N(N-1)/2$ distinct pairs. In cutoff-based methods each particle interacts with some fixed average number of other particles within the cutoff radius and the scaling is $O(N)$. With the PME algorithm, the cost is $O(N \ln N)$, while with the periodic version of FMM it is only $O(N)$.[147,149] It is expected that starting at $N \approx 20\,000$ the periodic FMM method will be the most cost effective and accurate way to evaluate long-range interactions.[42]

4.11 Path Exploration Methods

Modeling of structural transitions in large molecular systems encounters two types of difficulties. First is the problem of timescale, as large scale structural changes occur over periods of time longer than those accessible to current direct simulations. Second, while the end-points of the transition (initial and final state) may be known (e.g. crystal structures of oxy- and deoxyhemoglobin), the details of the reaction path are often unclear. In many cases the reaction coordinate involves hundreds or thousands of atoms, leading to the need for specific new methods of path exploration.[41] Approaches to efficiently describe such transitions include the self-penalty walk (SPW), locally updated planes (LUP) and conjugate peak refinement (CPR). All these methods consider a straight line connecting the initial and final state as a starting approximation to the reaction coordinate, and proceed to refine a numerical description of this coordinate by optimizing structures of intermediate states. In the SPW method M copies of the system are considered, initially in an equally spaced arrangement along the straight line connecting the initial and final state in Cartesian space. The potential energy of this ''supersystem'' is taken as the sum of the (standard) molecular potential energies of the individual copies and two penalty terms depending on the distance between copies. The penalty terms are a harmonic contribution due to deviation of the nearest-neighbor distance from the average and a Gaussian repulsive contribution between copies. Energy minimization of this ''supersystem'' leads to an optimized continuous path connecting the initial and final state, with the intermediate structures approximately equally spaced. In the LUP again a ''necklace'' or ''polymer'' of equally spaced intermediate structures is the starting point.[41] These structures are refined by constrained optimization in the direction orthogonal to the local reaction coordinate. This method is quite simple to implement and efficient due to the use of linear Cartesian constraints. In the CPR method alternate searches along the reaction coordinate and in the orthogonal direction

allow the identification of one or more transition states and/or intermediate minima between a given initial and final point.[150]

4.11.1 Applications

The LUP approach has been employed to determine the path for the transition between the α-helix and 3_{10}-helix in short model peptides[151] (see Figure 1). The sequential transition, in which residues change conformation one at a time was found to have a barrier of 2 kcal mol^{-1}.[151] A combination of LES and SPW methods was employed to study ligand escape in the protein leghemoglobin.[152] Leghemoglobin is an oxygen binding protein from plants, analogous to myoglobin in vertebrates. The reactive oxygen ligand is tightly bound to an iron atom of a heme, which is sequestered inside the protein, with no evident path of escape in the crystal structure. Using LES, a swarm of noninteracting ligands was first used to obtain a statistical picture of the pathways of ligand escape. Based on the generated ligand distribution, three somewhat different exit paths were proposed, and potentials of mean force along them were evaluated using the SPW. Two general properties of the ligand exit paths were: an initial activated step with a barrier of 2.5 kcal mol^{-1}, involving local motions of the ligand and sidechains of Phe 29 and Phe 44 lining the binding pocket; and a barrierless, diffusive step involving displacement of the G helix, with hundreds of atoms displaced along the reaction coordinate.

5 RELATION TO EXPERIMENTAL APPROACHES

A number of peptide and protein properties are accessible to both modeling and experimental studies. Verification that the calculated property values agree with measured data is an important test of the quality of modeling techniques. On the other hand, microscopic explanation of the observations is the goal of all modeling studies. Thus, availability of both sorts of information is highly desirable for complex molecular systems. Relationships between some fundamental calculated and observed quantities are outlined below. An excellent, in-depth discussion of this topic is given in Brooks et al.[61]

5.1 X-ray Diffraction

X-ray diffraction experiments are most often the source of structural data for starting macromolecular simulations. In many cases, the diffraction experiments also yield the Debye–Waller factors (isotropic temperature factors). For a given atom i the relationship between the temperature factor B_i and the mean-square position fluctuation is (Equation 53):[61]

$$\langle \Delta r_i^2 \rangle = \frac{3}{8\pi^2} B_i \tag{53}$$

Thus, accuracy of MD motional amplitudes may be tested by comparing calculated atomic fluctuations with the temperature factors. Some care has to be taken in the comparison due to the presence of a static disorder contribution in the X-ray data.[153]

Combining the information content of a molecular force field and the X-ray diffraction pattern has led to the development of an improved X-ray structure refinement protocol.[154] In this approach, the refinement residual, i.e. the difference between the observed and calculated structure factors $F_{h,k,l}$ is added to the molecular potential $U(q)$ as a penalty function, or constraint. In "low energy" regions of the combined potential the geometry and interatomic contacts are reasonable while the structures correspond to low values of the crystallographic R-factor. In combination with simulated annealing or elevated temperature MD, use of such potentials is a powerful tool in the final stages of X-ray structure refinement, used e.g. in the popular program XPLOR. A comparison of observed and measured atomic fluctuation amplitudes for myoglobin is given in Figure 7.[90,155–158]

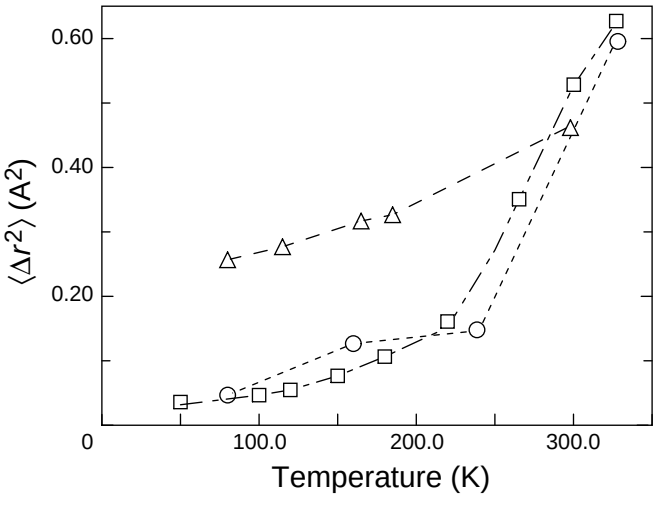

Figure 7 Myoglobin atom fluctuations: MD vs experiment. Root mean square position fluctuations averaged over all non-exchangeable hydrogens: from 50 ps vacuum MD trajectories of myoglobin bound with CO (circles) and from incoherent neutron scattering (squares).[90,155,156] Triangles: average over backbone atom root mean square position fluctuations calculated from X-ray Debye–Waller factors.[157] (Adapted from K. Kuczera.[158])

5.2 Nuclear Magnetic Resonance Spectroscopy

The measured NMR vicinal coupling constants $^3J_{HH}$ are used to determine the dihedral angle θ formed by the atoms involved through the Karplus equation shown in Equation (54):

$$^3J_{HH} = A\cos^2\theta - B\cos\theta + C \tag{54}$$

where A, B and C are parameters determined from model studies. The trajectory values of the dihedral $\theta(t)$ may be used to predict an average value and fluctuations of the dihedral and of the $^3J_{HH}$ function given above. Such calculations can test the similarity of families of conformations sampled in simulations and in experiments.

Nuclear Overhauser effect (NOE) experiments are sensitive to the distance between atoms. NOE results may be compared to simulation averages of $\langle r_{ij}^{-6}\rangle$, where r_{ij} is the distance between atoms i and j.[159] In the process of structure determination from NOE data, a penalty function U_{NOE} is defined by Equation (55):

$$U_{NOE} = \sum_k A_k (R_k - R_{NOE,k})^2 \tag{55}$$

where k numbers atom pairs, and R_k and $R_{NOE,k}$ are the current and NOE determined distance between atoms in the pair, respectively. The coefficients A_k are used to weight the NOE penalty function U_{NOE} relative to the molecular potential $U(q)$. Simulations for crambin, a small protein of 46 residues, showed that inclusion of NOE interproton distance information enables simulating a transition from a fully extended polypeptide chain to the correctly folded structure on timescales of tens of picoseconds.[160]

Order parameters S^2 have been used to extract the information on internal motions of biomolecules from NMR measurements.[161] An order parameter measures the degree of spatial restriction of the motion of a vector and is defined model-independently as Equation (56):[161]

$$S^2 = \lim_{t\to\infty} C_I(t) = \frac{4\pi}{5}\sum_{m=-2}^{2} |\langle Y_{2m}(\theta,\phi)\rangle|^2 \tag{56}$$

where $C_I(t)$ is the correlation function describing internal motion and the Y_{2m} are the second-order spherical harmonics as functions of the angular spherical coordinates θ and ϕ of a vector between two atoms in the molecule-fixed frame. Calculating order parameters from an MD trajectory provides a way to obtain information about internal motions i.e. directly comparable to experimental data.[161]

5.3 Hydrogen Isotope Exchange

Experiments involving exchange of amide NH protons with deuterium from solvent provide direct evidence for protein internal motions and can be used to study pathways of protein folding.[162] Exchange rates for NH protons in native proteins span a wide range of values, and are up to 10^8 slower than in small molecules where the protons are exposed to solvent. As longer simulations of solvated proteins are performed, it is expected that structural fluctuations exposing buried protons to solvent will be observed. This will help resolve the controversy about the microscopic mechanism of the exchange.[163]

5.4 Fluorescence Depolarization

Fluorescence depolarization measurements can probe internal motions of peptides and proteins by following the reorientation of fluorescent labels, such as aromatic residues.[164] For the case of parallel absorption and emission dipoles the fluorescence anisotropy decay is given by Equation (57):[165]

$$r(t) = \frac{I_{||}(t) - I_{\perp}(t)}{I_{||}(t) + 2I_{\perp}(t)} = \frac{2}{5}C_2(t) \tag{57}$$

where $I_{||}$ and $I_{\perp}$ are the intensities of emitted light with polarization parallel and perpendicular to that of the exciting pulse and $C_2(t)$ is the correlation function of the emission dipole, as shown in Equation (58):

$$C_2(t) = \frac{\langle 3\cos^2(\theta) - 1\rangle}{2} \tag{58}$$

where $\theta(t)$ is the angle of rotation of the emission/absorption dipole during time t. Similar relations may be obtained for linear dichroism and transient absorption measurements.[166]

For a system undergoing simple rotational diffusion, it is expected that the principal axes of the rotational diffusion tensor will reorient according to Equation (59):[165,166]

$$C_{2,i}(t) = a_i\,\mathrm{e}^{-(6D_r+2\Delta)t} + b_i\,\mathrm{e}^{-(6D_r-2\Delta)t} \quad i = x, y, z \tag{59}$$

where a_i, b_i are weights that depend on the diffusion coefficients around the individual axes, D_x, D_y, D_z; $D_r = (D_x + D_y + D_z)/3$ is the average rotational diffusion coefficient; and Δ is a measure of anisotropy. Fitting the $C_2(t)$ functions to exponential decays can provide information on timescales of molecular reorientation. Results of simulations and experimental measurements may thus be compared. The comparison becomes more difficult in the case of asymmetric chromophores, where up to five different exponents are expected in C_2,[166] and in the case when the motion may have several components, e.g. for a tryptophan in a protein, where overall protein tumbling and local sidechain reorientations may be differentiated.[61] Recently, some interesting and unexpected results have been obtained in simulations of simple

aromatic fluorophores in solution. Significant effects of microscopic solvent packing and interactions on rotational diffusion around individual molecular axes were found.[167] Extending such studies to peptide and protein systems could greatly enhance the amount of information that can be extracted from fluorescence depolarization studies.

5.5 Vibrational Spectroscopy

Vibrational spectroscopy of peptides and proteins provides information on the frequencies of molecular vibrations. This information is used in various ways. Infrared and Raman spectroscopy are most often used to study the relatively high frequency, localized motions of the amide groups. These frequencies are highly sensitive to the environment, and can be used to characterize protein secondary structure.[168,169] Specific isotope labeling and isotopic frequency shifts are employed to dissect enzymatic reaction mechanisms.[170] Resonance Raman spectroscopy is employed to analyze structure and interactions of chromophoric prosthetic groups.[171] As described in section 4.4, the model potentials used for peptide and protein simulations give a rather approximate picture of the high frequency localized group motions. However, normal mode analysis is highly useful in providing models for the low frequency delocalized motions which cannot otherwise be assigned.[61]

5.6 Neutron Scattering

In incoherent inelastic neutron scattering, the intensity of neutrons scattered from a sample is measured as a function of energy and momentum transfer. This yields information about the amplitudes and timescales of hydrogen atom motions.[172] Average hydrogen positional fluctuations can be determined from neutron scattering experiments and directly compared with simulation results (see Figure 7 for an example). Atomic motions from MD trajectories may be used to generate simulated incoherent neutron scattering profiles which are in excellent agreement with experimental observations. Additionally, based on the simulations, types of molecular motions that are responsible for different features of the observed neutron spectra may be determined.[172]

6 EXAMPLES OF APPLICATIONS

MD simulations of a wide range of peptide and protein properties have been reported. Protein simulations started with vacuum trajectories, focusing on calculation of atomic fluctuation amplitudes of small globular proteins: BPTI,[19] myoglobin[20] and lysozyme.[173,174] These studies showed qualitative agreement with X-ray temperature factors, but also found that atomic fluctuations in proteins tend to be highly anisotropic and anharmonic. For lysozyme, a change in the pattern in atomic mobility between the free and liganded forms of the enzyme was found, consistent with the observed greater rigidity of the liganded form.[173] Analysis of timescales of sidechain motions in MD trajectories indicated that folded globular proteins exhibit an unexpected mobility – even over picosecond timescales rotations of Tyr groups by up to 30° were found.[175] Calculations of dielectric relaxation, translational diffusion and rotational diffusion of BPTI and lysozyme showed good agreement with experimental data,[176] with dielectric relaxation and rotational diffusion occurring on the nanosecond timescale. A recent 1 μs simulation of a 38-residue peptide from the vilin headpiece subdomain in 3000 waters progressed from the totally unfolded structure to a collapsed compact form with formed α-helices and favorable solvation free energy. However, the final conformation of the peptide still exhibited some significant deviations from the folded structure.[40] Thus, protein folding still falls outside the timescales accessible to atomistic simulations. Several studies have approached the reverse process of protein unfolding by changing the temperature, solvent composition, perturbing structure or making a point mutation.[38,177] The general conclusions from these simulations are that unfolding is due to a change in the balance between protein–protein and protein–solvent interactions. The details of the unfolding and the final state reached vary when different methods of initiating unfolding are used.

A number of studies have gone beyond answering basic physical questions to modeling aspects of protein function. Early work focused on the protein myoglobin, which may be considered a model for understanding relations between biological function and protein motions. Two types of motion occur in myoglobin – equilibrium fluctuations and nonequilibrium relaxation.[178] The structural fluctuations are necessary for exit of small ligands (O_2, CO, NO, etc.) from the heme binding site to external environment. The relaxation is the directional structural transition upon ligand dissociation from the heme, involving motion of the iron atom out of the heme plane and subsequent tertiary structure change (see Figure 8[45,158,179]). Studies of ligand escape pathways included suggestion of an intuitive simple path obtained by minimal distortion of the protein structure through rotation of two surface sidechains.[61] The energetics of this pathway were explored using adiabatic mapping[180] and umbrella sampling,[119] suggesting that it should be accessible due to thermal fluctuations. Further simulations employing the LES approach suggested existence of several alternative exit pathways for ligands

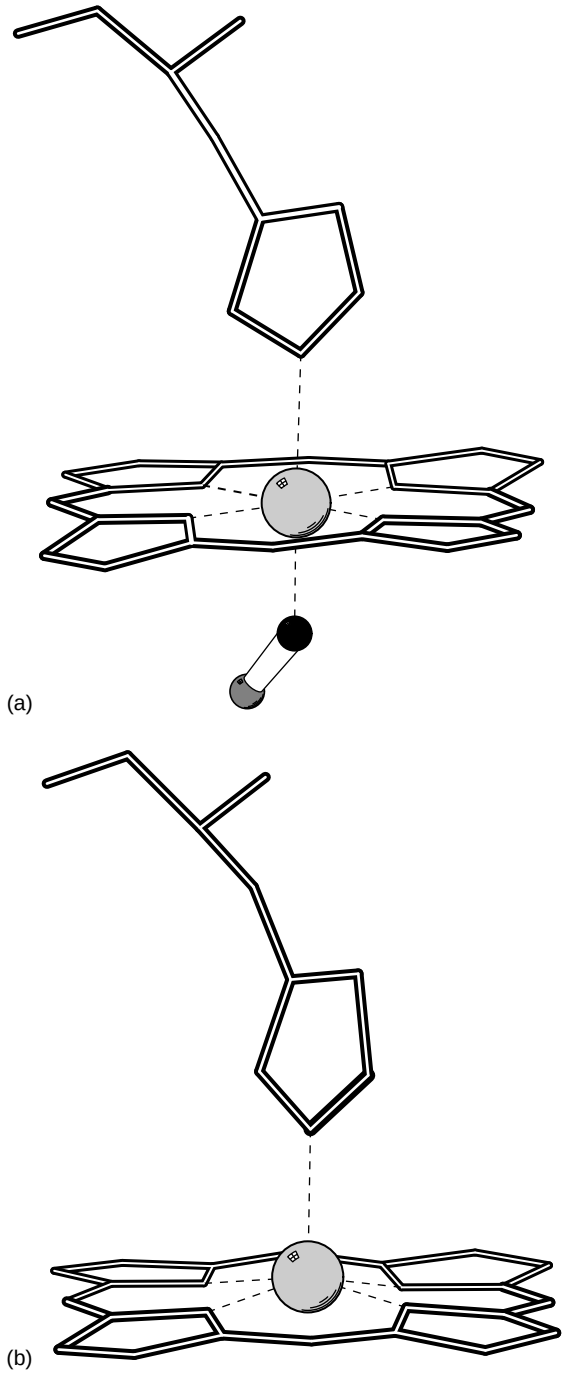

Figure 8 Fe atom displacement of the heme in myoglobin upon CO dissociation. Also shown is proximal histidine 93. (a) Six-liganded planar heme.[45] (b) Five-liganded domed heme.[179] (Adapted from K. Kuczera.[158])

from the myoglobin interior, some involving extended ligand diffusion through the protein matrix.[181,182] Protein structural relaxation after ligand dissociation were simulated by initiating either a sudden potential parameter change[89,183] or injection of kinetic energy into the heme.[88] These studies found a complex, nonexponential behavior. The simulated time course of the iron atom out-of-plane displacement agreed with time resolved infrared measurements and was used to explain why the observed rebinding for NO is nonexponential, while for CO or O_2 it is exponential at room temperature.[93,183] Simulations of the actual rebinding process did not find significant long-term structural relaxations, but suggested that nonexponential recombination may be explained by ligand diffusion through the protein.[91,92] An extended discussion of this still controversial question may be found in Kuczera.[158]

The structure and function of RNAse A were investigated in Haydock et al.[125] For this protein the X-ray structure of the enzyme–inhibitor complex was inconsistent with biochemical and mechanistic data on the reaction. Using the SBMD approach to model the active site and its surroundings, the actual substrate was built in place of the inhibitor in a crystal structure. After energy minimization and a short period of MD, the sidechains in the active site rearranged in a manner consistent with experimental aspects of the reaction mechanisms. This application clearly demonstrated the flexibility and utility of modeling methods.[125] Several studies focused on motions of ''flaps'' or ''lids'' which open to allow entry of substrate into the active site and then close to sequester the substrate in a nonaqueous environment. For triosephosphate isomerase, static analysis of the ''open'' and ''closed'' conformations from X-ray crystallography was supplemented by high-temperature MD simulations to show a possible transition pathway.[184] Analogous simulations for lactate dehydrogenase led to identification of the loop hinges from cross-correlations of atomic displacements.[185] In the case of human immunodeficiency virus (HIV) protease, simulations in water showed significant rearrangement of the enzyme flaps (which sequester the active site from solvent) compared to the crystal structure, where the flaps are involved in crystal contacts (see Figure 9[186]). The proposed solution structure of the free enzyme corresponded to a more closed flap state than expected.[187]

Molecular modeling of peptides has been progressing along two main paths – detailed simulations of small model peptides aimed at developing new methods and building basic understanding, and approximate studies of larger systems targeted at practical structure prediction of biologically active molecules. A number of examples of practical applications involving peptide-based drug design have been described.[188] The procedure starts conformational analysis of a biologically active peptide such as bradykinin (9 residues), angiotensin (8 residues) or substance P (11 residues) with highly simplified interactions with the environment (vacuum or constant

Figure 9 HIV protease structure, showing bound inhibitor sequestered under flaps. From PDB structure 1AAQ.[186]

dielectric treatment). A rough grid search, build-up procedure or simulated annealing are used to obtain a set of low energy conformers, representing several possible structures. These structures are then used to design constrained analogs of the parent compounds. The activity of the analogs is assayed, providing information about the putative structure of the biologically active parents, about the receptors as well as development of compounds with improved potency, selectivity and stability.[8,188] More advanced methods are typically used for fundamental simulations of simpler systems. Force fields for peptides and proteins are usually calibrated on predictions of relative conformational energies of the ''alanine dipeptide'' (α-(formylamino)propanamide)[34] (Figure 4). Nanosecond length simulations in explicit solvent have been performed for a series of short peptides.[34,104,189,190] In several simple cases such as $(Ala)_3$, $(Val)_3$ or Tyr-Pro-Gly-Asp-Val, complete conformational sampling was achieved in several nanoseconds, as the same set of structures was visited several times, allowing the calculation of equilibrium constants directly from trajectories.[34]

The role of salt in the simulations of solvated biomolecules has been studied, including the distribution of counterions, influence on conformational equilibria and dynamics.[37,189] Comparative simulations of cyclic (disulfide bridged) and linear versions of the RGD peptide Ac-Pen-Arg-Gly-Asp-Cys-NH_2 (Pen – penicillamine) and opioid peptide DPDPE were used to quantify the effects of disulfide constraints.[190,191] Using several different measures, rates of conformational exploration of the linear forms were about two to three times faster than those of the cyclic ones. The free energy cost of restricting the linear peptide to the conformations available to the cyclic form was estimated at 4–6 kcal mol^{-1}.[192] A number of studies have approached the problem of unfolding and initiation of peptide α-helices in water.[193–196] A complex overall picture was found, with partial unfolding proceeding from the C-terminus toward the N-terminus on a nanosecond timescale, and transient population of alternative helix forms. Recent simulations indicate that helix initiation is associated with a barrier of ca. 3 kcal mol^{-1}, corresponding to a rate of 0.1–0.01 ps^{-1}, faster that previously expected, and that the complete helix formation from a disordered structure should occur within 20–70 ns at room temperature.[197,198] These timescales are much longer than used in many previous simulations, but are quickly coming within reach of current computational studies.

Recently, β-peptides, a class of peptide analogs with an additional carbon atom in the backbone, have attracted significant attention.[199] β-peptides have the same functional groups as regular peptides, exhibit a tendency to form regular secondary structures, and are resistant to degradation by proteases. These molecules are thus a promising source of potential biologically active agents. They are also good targets for simulations, as relatively small systems are capable of forming regular structures. MD simulations of a seven residue β-peptide in aqueous solution at several temperatures on 50 ns timescales generated an equilibrium distribution of conformers.[200] The β-peptide repeatedly visited a relatively small number of conformations ($\simeq 10^2$), with the experimentally observed left-handed 3_1 helix being predominantly populated (Figure 10). Free energy differences between conformers, the melting temperature and folding free energy were determined from the trajectory. Multiple pathways to the folded state were found.[200]

Early free energy simulations of peptide/protein systems involved successful calculations of solvation free energy differences between sidechain models.[25] Further studies used the idea of thermodynamic cycles[120] to study influence of protein point mutations on thermal stability of T4 lysozyme,[201] cooperative oxygen binding to hemoglobin[26] and the hemoglobin sickling aggregation.[118] In all these studies the calculated overall free energy change was in qualitative agreement with experimental results. Additionally, interesting and unexpected effects were found in analysis of the detailed simulation results, supplying valuable insight into the observed effects. The hemoglobin studies found effects of removal of favorable interactions as well as the creation of favorable ones, differential solvation of wild type and mutant structures, and nonobvious contributions from individual residues to the overall process.[118,127] An

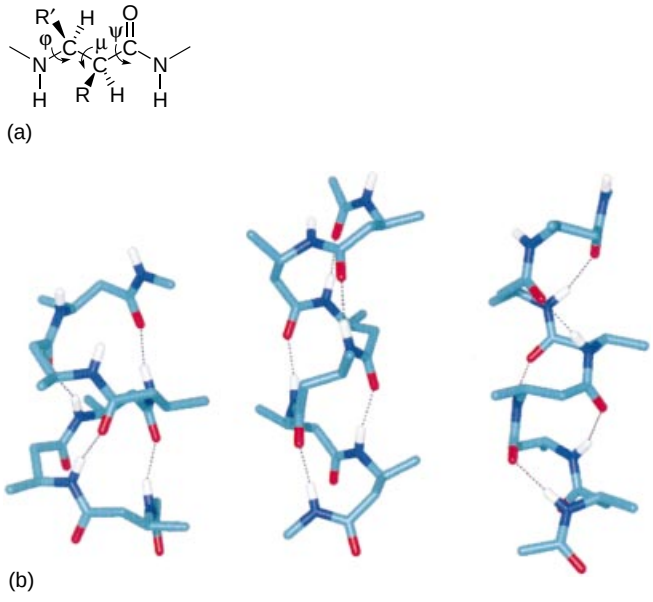

Figure 10 (a) The β-peptide backbone structure. (b) Examples of possible β-peptide helices. From left to right: H_{14}-helix, H_{12}-helix and H_{10}-helix.

example of a challenging and interesting application are the free energy simulations of binding of biotin and the protein streptavidin.[111] This is probably the strongest known protein–organic ligand interaction, with a binding free energy ΔG_{assoc} of $-18\,\text{kcal mol}^{-1}$. The simulations were able to correctly predict the relative binding free energies of biotin, thiobiotin and iminobiotin, and also to determine the absolute free energy of biotin binding.[111] An interesting conclusion from the calculations is that the unusual stability of the biotin–streptavidin complex is due to van der Waals/hydrophobic type energy terms, rather than electrostatics. This is because the charged and polar groups involved in binding tend to interact equally well with each other in the bound state as with water in the uncomplexed state.

The helix and turn propensity of different amino acid residues was investigated in a series of cleverly designed free energy simulations.[202] By mutating residues in different peptide conformations and using thermodynamic cycles, the authors were able to dissect the relative roles of solvation and intramolecular interactions in structure formation. Calculations of the potential mean force for helix terminal unwinding[197] and for folding of a three-helix bundle with the radius of gyration as reaction coordinate were also undertaken.[39] Recently, an extension of the TI approach has been proposed, allowing the calculation of the free energy gradient with respect to several conformational coordinates in a single simulation.[114] This method has been applied to the study of the relative and intrinsic stability of different helical states of model peptides (Figure 11[203]). The standard α-helix was found to be a local minimum on the free energy surface of the $(\text{Ala})_{10}$ peptide both in vacuum and in aqueous solution. A stable π-helix (with $i - i + 5$ hydrogen bonds) was also found in the study. Unexpectedly, the π-helical state was quite stable in solution, due to stronger interactions with solvent than the standard, widely studied α-helix.

7 FUTURE DIRECTIONS

Emerging future trends in computer simulations of biomolecules involve more accurate description of larger and more complex systems over longer time intervals, as well as development and application of a wide range of simplified models.

- System size and complexity. Current simulations involve systems of 10^3–10^4 atoms; within the next 5–10 years sizes of 10^4–10^5 atoms will become standard. The systems themselves, from the current solvated peptides, proteins or nucleic acids, will become more complex, including buffer ions and interfacial environments. This will enable studies of many new interesting biochemical systems, including large, multisubunit proteins, proteins in membranes and protein–nucleic acid interactions.
- Simulation length. The standard of simulation length is changing from 100 ps several years ago to about 1–5 ns currently. This trend will certainly continue and longer simulations can be expected in the future. With longer simulations, structural and energetic dynamics will be probed on longer timescales. This will also provide larger samples of configurations and yield more accurate averages. Parallel computing will be used to extend both system size and simulation length. Improved algorithms involving symplectic, reversible, splitting and stochastic approaches will be more widely implemented.
- Potentials. More accurate potential energy functions are being developed. Improvements include both description of internal structural deformations and nonbonded interactions. In the internal deformation force field off-diagonal, anharmonic and conformation-dependent terms are being added. More accurate representations of the static charge distribution, inclusion of polarization and avoidance of truncation effects will become widespread in electrostatic energy calculations. At the level of van der Waals interactions, nonadditive corrections are being introduced to improve agreement with experimental data.[111] Mixed QM/MM approaches will be more widely used, which allow the quantum mechanical

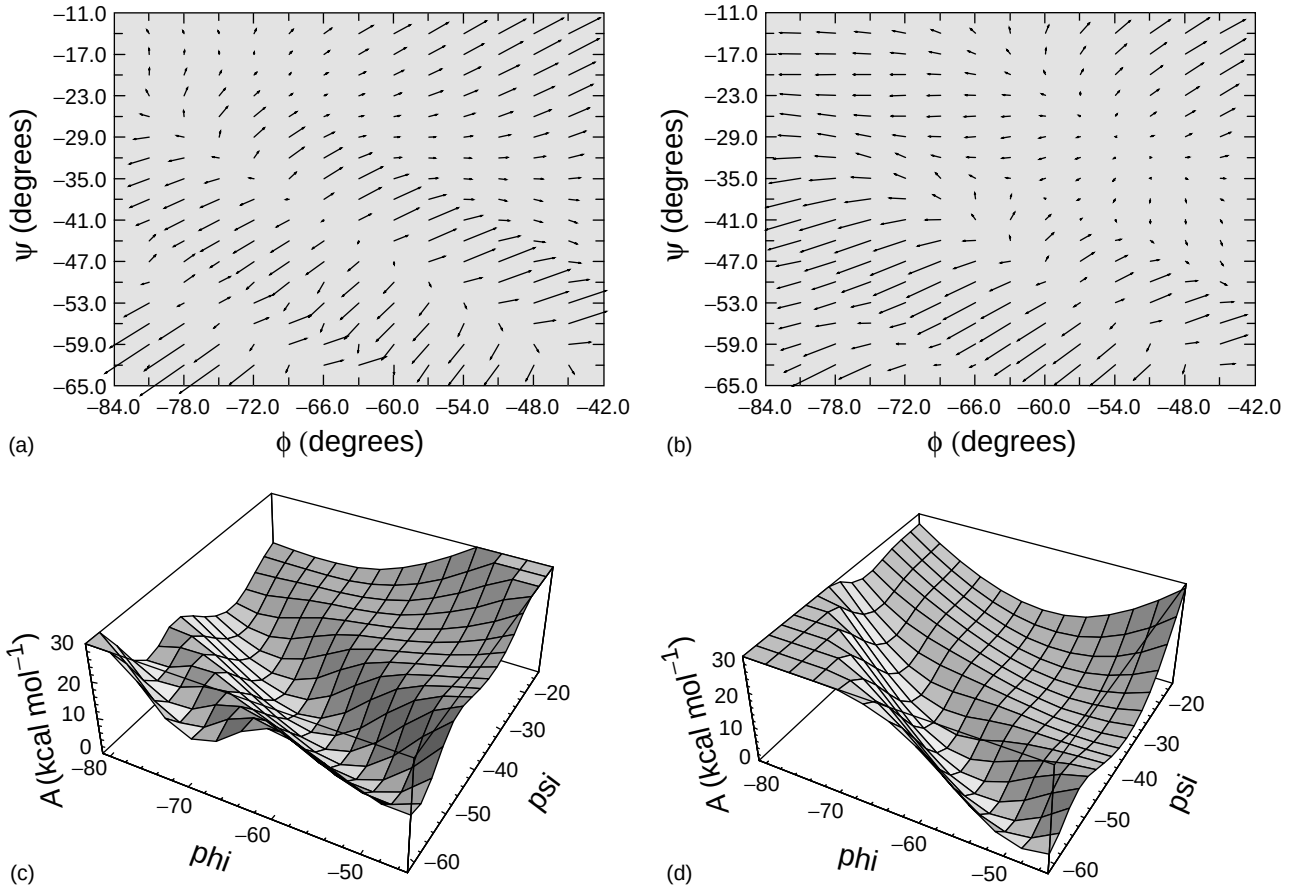

Figure 11 Two-dimensional conformational free energy maps of $(Ala)_{10}$ and $(Aib)_{10}$ in vacuum. Aib is α-methylalanine. (a) Map of free energy gradient for $(Ala)_{10}$ showing two minima, corresponding to the α- and π-helices. (b) Map of free energy gradient for $(Aib)_{10}$ showing two minima, corresponding to the α- and 3_{10}-helices. (c) Free energy surface corresponding to (a). (d) Free energy surface corresponding to (b). (Adapted from Y. Wang, K. Kuczera.[203])

treatment of selected parts of the system.[204,205] Fully quantum calculations of macromolecular systems are also being developed.

- Sampling. Enhanced sampling algorithms are becoming part of standard tools of molecular modeling. These include approaches such as LES, path exploration algorithms, and other approaches such as essential dynamics (based on singular value decomposition)[206] or self-guided dynamics (based on potential energy smoothing).[207] In the field of free energy simulations soft-sphere potentials are being considered as tools for accelerating convergence of the statistical mechanical averages.[115]
- Simplified models. The revolution in molecular biology is generating ever increasing numbers of protein sequences and information about interactions between macromolecules with each other and with small effectors. This is driving the growing use of simplified approaches which are able to yield quick though approximate results. The treatment of solvation through the PB and GB methods described in section 4.9 belong in this class. Another class of methods decreases the number of interaction centers through the use of coarse-grained models.[208] Comparative modeling aims to predict protein structures based on known three-dimensional structures of proteins with similar sequences.[209] Docking methods that predict binding of small ligands to macromolecules using simplified interaction potentials are of growing importance for drug design.[210,211] Another class of interesting approaches are lattice methods, which are generating highly useful insights into protein folding.[67,212,213]

ACKNOWLEDGMENTS

I would like to thank Dr Gouri Jas, Mr Robert Guenther and Mrs Terrie Saunders for help with preparing the

figures for this article. The molecular graphics were prepared using MolScript[214] and Raster3D.[215,216]

ABBREVIATIONS AND ACRONYMS

BPTI	Bovine Pancreatic Trypsin Inhibitor
CPR	Conjugate Peak Refinement
FMM	Fast Multipole Method
GB	Generalized Born
HbS	Sickle-cell Hemoglobin
HIV	Human Immunodeficiency Virus
HMC	Hybrid Monte Carlo
LES	Locally Enhanced Sampling
LUP	Locally Updated Planes
MC	Monte Carlo
MD	Molecular Dynamics
MTS	Multiple Time Stepping
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PB	Poisson–Boltzmann
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
QM/MM	Quantum Mechanics/Molecular Mechanics
r-RESPA	Reversible Reference System Propagator Algorithms
SBMD	Stochastic Boundary Molecular Dynamics
SPW	Self-penalty Walk
TI	Thermodynamic Integration
TP	Thermodynamic Perturbation
UFF	Universal Force Field

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Infrared Spectroscopy of Biological Applications • Nuclear Magnetic Resonance of Biomolecules • Raman Spectroscopy in Analysis of Biomolecules

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis • Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis • X-ray Crystallography of Biological Macromolecules

Pharmaceuticals and Drugs **(Volume 8)**
Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery • Vibrational Spectroscopy in Drug Discovery, Development and Production

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

REFERENCES

1. E.B. Wilson, Jr, J.C. Decius, P.C. Cross, *Molecular Vibrations*, McGraw-Hill, New York, 1955.
2. J.A. Hirschfelder, H. Eyring, B. Topley, 'Reactions Involving Hydrogen Molecules and Atoms', *J. Chem. Phys.*, **4**, 170 (1936).
3. T.J. Hill, 'On Steric Effects', *J. Chem. Phys.*, **14**, 465 (1946).
4. F.H. Westheimer, J.E. Mayer, 'The Theory of Racemization of Optically Active Derivatives of Diphenyl', *J. Chem. Phys.*, **14**, 733–738 (1946).
5. J.D. Kemp, K.S. Pitzer, 'Hindered Rotation of the Methyl Groups in Ethane', *J. Chem. Phys.*, **4**, 749 (1936).
6. A.I. Kitaigordskij, *Chemical Organic Crystallography*, Consultants Bureau, New York, 1959.
7. G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, 'Stereochemistry of Polypeptide Chain Configurations', *J. Mol. Biol.*, **7**, 95 (1963).
8. A.K. Rappe, C.J. Casewit, *Molecular Mechanics Across Chemistry*, University Science Books, Sausalito, CA, 1997.
9. K.D. Gibson, H.A. Scheraga, 'Minimization of Polypeptide Energy I. Preliminary Structures of Bovine Pancreatic Ribonuclease S-peptide', *Proc. Natl. Acad. Sci. USA*, **58**, 420 (1967).
10. A. Hagler, E. Huber, S. Lifson, 'Energy Functions for Peptides and Proteins I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals', *J. Am. Chem. Soc.*, **96**, 5319 (1974).
11. M. Levitt, 'Molecular Dynamics of Native Protein I. Computer Simulation of Trajectories', *J. Mol. Biol.*, **168**, 595 (1983).
12. B.R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, 'CHARMM: A Program for Macromolecular Energy, Minimization and Dynamics Calculations', *J. Comp. Chem.*, **4**, 187–217 (1983).

13. S.J. Weiner, P. Kollman, D. Case, U. Singh, C. Ghio, A. Alagona, S. Profeta, P. Weiner, 'A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins', *J. Am. Chem. Soc.*, **106**, 765–784 (1984).
14. U. Burkert, N.L. Allinger, *Molecular Mechanics*, American Chemical Society, Washington, DC, 1982.
15. B.J. Alder, T.E. Wainwright, 'Studies in Molecular Dynamics I. General Method', *J. Chem. Phys.*, **31**, 459 (1959).
16. A. Rahman, 'Correlations in the Motion of Atoms in Liquid Argon', *Phys. Rev. A*, **136**, 405 (1964).
17. F.H. Stillinger, A. Rahman, 'Improved Simulation of Liquid Water by Molecular Dynamics', *J. Chem. Phys.*, **60**, 1545 (1974).
18. M.P. Allen, D.J. Tildesley, *Computer Simulations of Liquids*, Oxford Science Publications, London, 1987.
19. J.A. McCammon, B.R. Gelin, M. Karplus, 'Dynamics of Folded Proteins', *Nature (London)*, **267**, 585–590 (1977).
20. R.L. Levy, R.P. Sheridan, J.W. Keepers, G.S. Dubey, S. Swaminathan, M. Karplus, 'Molecular Dynamics of Myoglobin at 298 K. Results from a 300-ps Computer Simulation', *Biophys. J.*, **48**, 509–518 (1985).
21. B.R. Brooks, M. Karplus, 'Normal Modes for Specific Motions of Macromolecules: Application to the Hinge-bending Mode of Lysozyme', *Proc. Natl. Acad. Sci. USA*, **82**, 4995–4999 (1985).
22. M. Levitt, C. Sander, P.S. Stern, 'Protein Normal Mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme', *J. Mol. Biol.*, **181**, 423 (1985).
23. W.L. Jorgensen, C. Ravimohan, 'Monte Carlo Simulation of Differences in Free Energies of Hydration', *J. Am. Chem. Soc.*, **83**, 3050–3054 (1985).
24. T.P. Lybrand, I. Ghosh, J.A. McCammon, 'Hydration of Chloride and Bromide Anions: Determination of Relative Free Energy by Computer Simulation', *J. Am. Chem. Soc.*, **107**, 7793–7794 (1985).
25. P.A. Bash, U.C. Singh, R. Langridge, P.A. Kollman, 'Free Energy Calculations by Computer Simulation', *Science*, **236**, 564–568 (1987).
26. J. Gao, K. Kuczera, B. Tidor, M. Karplus, 'Hidden Thermodynamics of Mutant Proteins: A Molecular Dynamics Analysis', *Science*, **244**, 1069–1072 (1989).
27. T.P. Lybrand, P.A. Kollman, 'Water–Water and Water–Ion Potential Functions Including Terms for Many-body Effects', *J. Chem. Phys.*, **83**, 2923–2933 (1985).
28. S.T. Russel, A. Warshel, 'Calculation of Electrostatic Energies in Proteins. The Energetics of Ionized Groups in BPTI', *J. Mol. Biol.*, **185**, 389–404 (1985).
29. D. Van Belle, I. Couplet, M. Prevost, S.J. Wodak, 'Calculations of Electrostatic Properties in Proteins. Analysis of Contributions from Induced Protein Dipoles', *J. Mol. Biol.*, **198**, 721–735 (1987).
30. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley, New York, 1991.
31. R. Car, M. Parinello, 'Unified Approach for Molecular-dynamics and Density-functional Theory', *Phy. Rev. Lett.*, **55**, 2471–2476 (1985).
32. U.C. Singh, P.A. Kollman, 'A Combined *ab initio* Quantum Mechanical and Molecular Mechanical Method for Carrying Out Simulations on Complex Molecular Systems: Application to the CH_3Cl+Cl^- Exchange Reaction and Gas Phase Protonation of Polyethers', *J. Comp. Chem.*, **7**, 718–729 (1986).
33. P.A. Bash, M.J. Field, M. Karplus, 'Free Energy Perturbation Method for Chemical Reactions in the Condensed Phase: A Dynamical Approach Based on a Combined Quantum and Molecular Mechanics Potential', *J. Am. Chem. Soc.*, **109**, 8092–8094 (1987).
34. C.L. Brooks III, D.A. Case, 'Simulations of Peptide Conformational Dynamics and Thermodynamics', *Chem. Rev.*, **93**, 2487–2502 (1993).
35. L.A. Findsen, S. Subramanian, V. Lounnas, B.M. Pettitt, 'Timescales and Fluctuations of Protein Dynamics: Metmyoglobin in Aqueous Solution', in *Principles of Molecular Recognition*, eds. A.D. Buckingham, A.C. Stanley, S.M. Legon, Blackie, Glasgow, 168–193, 1993.
36. H.-Q. Ding, N. Karasawa, W.A. Goddard, III, 'Atomic Level Simulations on a Million Particles: The Cell Multipole Method for Coulomb and London Nonbonding Interactions', *J. Chem. Phys.*, **97**, 4309–4315 (1992).
37. P.E. Smith, B.M. Pettitt, 'Amino Acid Side-chain Populations in Aqueous and Saline Solution: Bis(penicillamine) Enkephalin', *Biopolymers*, **32**, 1623–1629 (1992).
38. P.H. Huenenberger, A.E. Mark, W.F. van Gunsteren, 'Computational Approaches to Study of Protein Unfolding: Hen Egg White Lysozyme as a Case Study', *Proteins*, **21**, 196–213 (1995).
39. E.M. Boczko, C.L. Brooks III, 'First Principles Calculation of the Folding Free Energy of a Three Helix Bundle Protein', *Science*, **269**, 393–396 (1995).
40. Y. Duan, P.A. Kollman, 'Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution', *Science*, **282**, 740–744 (1998).
41. R. Elber, 'Reaction Path Studies of Biological Molecules', in *Recent Developments in Theoretical Studies of Proteins*, ed. R. Elber, World Scientific, Singapore, 65–136, 1996.
42. B.J. Berne, 'Molecular Dynamics in Systems with Multiple Timescales: Reference System Propagator Algorithms', in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, eds. P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, R.D. Skeel, Springer, Berlin, 297–317, 1999.
43. D.J. Wales, H.A. Scheraga, 'Global Optimization of Clusters, Crystals and Biomolecules', *Science*, **285**, 1368–1372 (1999).
44. B.J. Berne, J.E. Straub, *Curr. Topics Struct. Biol.*, **7**, 181–189 (1997).

45. J. Kuriyan, M. Wilz, S. Karplus, G.A. Petsko, 'X-ray Structure and Refinement of Carbon-monoxy Fe(II)-myoglobin at 1.5 Å Resolution', *J. Mol. Biol.*, **192**, 133–154 (1986).
46. R.K. Wierenga, M.E.M. Noble, J.P.M. Postma, H. Groendijk, K.H. Kalk, W.G.J. Hol, F.R. Opperdoes, 'The Crystal Structure of the Open and the Closed Conformation of the Flexible Loop of Trypanosomal Triosephosphate Isomerase', *Proteins*, **10**, 867 (1991).
47. T.E. Creighton, *Proteins. Structures and Molecular Properties*, W. H. Freeman, New York, 1993.
48. I.N. Levine, *Quantum Chemistry*, 5th edition, Prentice-Hall, Upper Saddle River, NJ, 2000.
49. A.D. MacKerell, Jr, D. Bashford, M. Bellott, R.L. Dunbrack, Jr, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, M. Karplus, 'All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins', *J. Phys. Chem. B*, **102**, 3386–3616 (1998).
50. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179–5197 (1995).
51. G. Kaminski, E.M. Duffy, T. Matsui, W.L. Jorgensen, 'Free Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the OPLS All-atom Model', *J. Phys. Chem.*, **98**, 13 077–13 082 (1994).
52. W.F. van Gunsteren, H.J.C. Berendsen, 'Computer Simulations of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry', *Angew. Chem. Int. Ed. Engl.*, **29**, 992–1023 (1990).
53. H. Margenau, N.R. Kestner, *Theory of Intermolecular Forces*, Pergamon, Oxford, 1971.
54. J. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, London, 1992.
55. N.L. Allinger, Z.S. Zhu, K. Chen, 'Molecular Mechanics (mm3) Studies of Carboxylic Acids and Esters', *J. Am. Chem. Soc.*, **114**, 6120–6133 (1992).
56. J.R. Maple, U. Dinur, A.T. Hagler, 'Derivation of Force Fields for Molecular Mechanics and Molecular Dynamics from *ab initio* Energy Surfaces', *Proc. Natl. Acad. Sci. USA*, **85**, 5350 (1988).
57. S.L. Mayo, B.D. Olafson, W.A. Goddard III, 'DREIDING: A Generic Force Field for Molecular Simulations', *J. Phys. Chem.*, **94**, 8897 (1990).
58. A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard III, W.M. Skiff, 'UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations', *J. Am. Chem. Soc.*, **114**, 10 024 (1992).
59. R.J. Loncharich, B.R. Brooks, 'The Effect of Truncating Long-range Forces on Protein Dynamics', *Proteins*, **6**, 32–45 (1989).
60. T. Schlick, 'Optimization Methods in Computational Chemistry', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH, New York, 1–71, Vol. 3, 1992.
61. C.L. Brooks III, M. Karplus, B.M. Petitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Wiley & Sons, New York, 1988.
62. W.H.B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.
63. H.A. Scheraga, 'Predicting Three-dimensional Structures of Oligopeptides', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH, New York, 73–141, Vol. 3, 1992.
64. T.E. Creighton (ed.), *Protein Folding*, W.H. Freeman, New York, 1992.
65. R. Czerminski, R. Elber, 'Reaction Path Study of Conformational Transitions in Flexible Systems: Applications to Peptides', *J. Phys. Chem.*, **92**, 5580–5601 (1990).
66. R. Elber, M. Karplus, 'Multiple Conformational States of Proteins: A Molecular Dynamics Analysis of Myoglobin', *Science*, **235**, 318–321 (1987).
67. A. Sali, E. Shakhnovich, M. Karplus, 'How does a Protein Fold?', *Nature*, **369**, 248–251 (1994).
68. J.E. Straub, 'Optimization Techniques with Applications to Proteins', in *Recent Developments in Theoretical Studies of Proteins*, ed. R. Elber, World Scientific, Singapore, 137–196, 1996.
69. K. Irikura, B. Tidor, B.R. Brooks, M. Karplus, 'Transition from A to Z DNA: Contribution of Internal Fluctuations to the Configurational Entropy Difference', *Science*, **229**, 571–572 (1985).
70. G. Basu, F. Hirata, N. Go, 'A Collective Motion Description of the 3_{10}-/α-helix Transition: Implications for a Natural Reaction Coordinate', *J. Am. Chem. Soc.*, **116**, 6307–6315 (1994).
71. R. Bruccoleri, M. Karplus, J.A. McCammon, 'The Hinge-bending Mode of a Lysozyme–Inhibitor Complex', *Biopolymers*, **25**, 1767–1802 (1986).
72. T. Noguti, N. Go, 'Structural Basis of Hierarchical Multiple Substates of a Protein. i: Introduction', *Proteins: Structure, Function and Genetics*, **5**, 97–103 (1989).
73. T. Schlick, 'Some Failures and Successes of Long-timestep Approaches to Biomolecular Simulations', in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, eds. P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, R.D. Skeel, Springer, Berlin, 227–262, 1999.
74. S. Krimm, 'Spectra and Structure of Polypeptides', in *Vibrational Spectra and Structure*, ed. J.R. Durig, Elsevier, Amsterdam, 1–72, Vol. 16, 1987.
75. T.C. Cheam, S. Krimm, 'Ab Initio Force Fields of Alanine Dipeptide in C_5 and C_7 Conformations', *J. Mol. Struct.*, **188**, 15–43 (1989).

76. L. Verlet, 'Computer ''Experiments'' on Classical Fluids I. Thermodynamical Properties of Lennard–Jones Molecules', *Phys. Rev.*, **159**, 98 (1967).
77. J.M. Sanz-Serna, 'Symplectic Integrators for Hamiltonian Problems: An Overview', *Acta Numerica*, **1**, 243–286 (1991).
78. D. Frenkel, B. Smit, *Understanding Molecular Simulation*, Academic Press, San Diego, 1996.
79. D.C. Harris, M.D. Bertolucci, *Symmetry and Spectroscopy*, Dover, New York, 1989.
80. E. Barth, K. Kuczera, B. Leimkuhler, R.D. Skeel, 'Algorithms for Constrained Molecular Dynamics', *J. Comp. Chem.*, **16**, 1192–1209 (1995).
81. D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, Oxford, 1987.
82. J. McConnell, *Rotational Brownian Motion and Dielectric Theory*, Academic Press, London, 1980.
83. D. Brown, J.H. Clarke, 'A Direct Method of Studying Reaction Rates by Equilibrium Molecular Dynamics: Application to the Kinetics of Isomerization in Liquid *n*-Butane', *J. Chem. Phys.*, **92**, 3062–3073 (1990).
84. H. Frohlich, *Theory of Dielectrics*, Clarendon, Oxford, 1958.
85. D.A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York, 1976.
86. M. Toda, R. Kubo, N. Saito, *Statistical Physics I. Equilibrium Statistical Mechanics*, Springer-Verlag, Berlin, 1992.
87. R. Chang, *Physical Chemistry with Applications to Biological Systems*, Macmillan, New York, 1981.
88. E.R. Henry, W.A. Eaton, R.M. Hochstrasser, 'Molecular Dynamics Simulations of Cooling in Laser-excited Heme Proteins', *Proc. Natl. Acad. Sci. USA*, **83**, 8982–8986 (1986).
89. E.R. Henry, M. Levitt, W.A. Eaton, 'Molecular Dynamics Simulation of Photodissociation of Carbon Monoxide from Hemoglobin', *Proc. Natl. Acad. Sci. USA*, **82**, 2034–2038 (1985).
90. K. Kuczera, J. Kuriyan, M. Karplus, 'Molecular Dynamics of CO-myoglobin at 325 and 80 K', *J. Mol. Biol.*, **213**, 351–373 (1990).
91. O. Schaad, H.-X. Zhou, A. Szabo, W.A. Eaton, E.R. Henry, 'Simulation of the Kinetics of Ligand Binding to a Protein by Molecular Dynamics: Geminate Rebinding of Nitric Oxide to Myoglobin', *Proc. Natl. Acad. Sci. USA*, **90**, 9547–9551 (1993).
92. H. Li, R. Elber, J.E. Straub, 'Molecular Dynamics Simulation of NO Recombination to Myoglobin Mutants', *J. Biol. Chem.*, **268**, 17 908–17 916 (1993).
93. K. Kuczera, J.-C. Lambry, J.-L.M. Martin, M. Karplus, 'Nonexponential Relaxation After Ligand Dissociation from Myoglobin: A Molecular Dynamics Simulation', *Proc. Natl. Acad. Sci.*, **90**, 5805–5807 (1993).
94. J.M. Sanz-Serna, M.P. Calvo, *Numerical Hamiltonian Problems*, Chapman & Hall, London, 1994.
95. R.D. Skeel, J.A. Izaguirre, 'The Five Femtosecond Time Step Barrier', in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, eds. P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, R.D. Skeel, Springer, Berlin, 318–321, 1999.
96. D. Janezic, F. Merzel, 'Split Integration Symplectic Method for Molecular Dynamics Algorithms', *J. Chem. Inform. Comp. Sci.*, **37**, 1048–1054 (1997).
97. J.P. Valleau, G.M. Torrie, 'A Guide to Monte Carlo for Statistical Mechanics 2. Byways', in *Statistical Mechanics. Part A. Equilibrium Techniques*, ed. B.J. Berne, Plenum, New York, 169–194, 1977.
98. N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland Publishing Company, Amsterdam, 1981.
99. A.R. Leach, *Molecular Modeling. Principles and Applications*, Longman, Harlow, 1996.
100. D.S. Goodsell, A.J. Olson, 'Automated Docking of Substrates to Proteins by Simulated Annealing', *Proteins*, **8**, 195–202 (1990).
101. M. Saunders, K.N. Houk, Y.-D. Wu, W.C. Still, M. Lipton, W.C. Guida, 'Conformations of Cycloheptadecane. A Comparison of Methods for Conformational Searching', *J. Am. Chem. Soc.*, **112**, 1419–1427 (1990).
102. R. Judson, 'Teaching Polymers How to Fold', *J. Phys. Chem.*, **96**, 10 102 (1992).
103. A. Roitberg, R. Elber, 'Modeling Sidechains in Peptides and Proteins: Annealing Methods to Find Minimum Energy Conformations', *J. Chem. Phys.*, **95**, 9277 (1992).
104. C. Simmerling, R. Elber, 'Computer Determination of Peptide Conformations in Water: Different Roads to Structure', *Proc. Natl. Acad. Sci. USA*, **92**, 3190–3193 (1995).
105. S. Duane, A.D. Kennedy, B.J. Pendleton, D. Roweth, 'Hybrid Monte Carlo', *Phys. Rev. Lett. B*, **195**, 216–222 (1987).
106. I. Andricioaei, J.E. Straub, 'Generalized Simulated Annealing Algorithm using Tsallis Statistics: Application to Conformational Optimization of a Tetrapeptide', *Phys. Rev. E*, **53**, 3055 (1996).
107. B.A. Berg, T. Neuhaus, 'Multicanonical Algorithm for First Order Phase Transitions', *Phys. Lett. B*, **267**, 249 (1991).
108. N. Nakajima, H. Nakamura, A. Kidera, 'Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides', *J. Phys. Chem. B*, **101**, 817–824 (1997).
109. D.L. Beveridge, F.M. DiCapua, 'Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems', *Annu. Rev. Biophys. Biophys. Chem.*, **18**, 431–492 (1989).
110. T.P. Straatsma, J.A. McCammon, 'Computational Alchemy', *Annu. Rev. Phys. Chem.*, **43**, 407–435 (1992).
111. P.A. Kollman, 'Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the

Properties of Organic and Biological Molecules', *Acc. Chem. Res.*, **29**, 461–469 (1996).
112. M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, 'A Statistical-thermodynamic Basis for Computation of Binding Affinities: A Critical Review', *Biophys. J.*, **72**, 1047–1069 (1997).
113. M. Mezei, D.L. Beveridge, 'Free Energy Simulations', *Ann. N.Y. Acad. Sci.*, **482**, 1–23 (1986).
114. K. Kuczera, 'Exploration of Peptide Free Energy Surfaces', in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, eds. P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, R.D. Skeel, Springer, Berlin, 163–175, 1999.
115. H. Liu, A.E. Mark, W.F. van Gunsteren, 'Estimating the Relative Free Energy of Different Molecular States with Respect to a Single Reference State', *J. Phys. Chem.*, **100**, 9485–9494 (1996).
116. E.A. Carter, G. Ciccotti, J.T. Hynes, R. Kapral, 'Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events', *Chem. Phys. Lett.*, **156**, 472–477 (1989).
117. K. Kuczera, 'One- and Multidimensional Conformational Free Energy Simulations', *J. Comp. Chem.*, **17**, 1726–1749 (1996).
118. K. Kuczera, J. Gao, B. Tidor, M. Karplus, 'Free Energy of Sickling: A Simulation Analysis', *Proc. Natl. Acad. Sci. USA*, **87**, 8481–8485 (1990).
119. J. Kottalam, D.A. Case, 'Dynamics of Ligand Escape from the Heme Pocket of Myoglobin', *J. Am. Chem. Soc.*, **110**, 7690–7697 (1988).
120. B.L. Tembe, J.A. McCammon, 'Ligand–Receptor Interactions', *Comput. Chem.*, **8**, 281–283 (1984).
121. R.M. Levy, E. Gallicchio, 'Computer Simulations with Explicit Solvent: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic Effects', *Annu. Rev. Phys. Chem.*, **49**, 531–567 (1998).
122. D.J. Adams, 'Alternatives to the Periodic Cube in Computer Simulation', *CCP5 Quarterly*, **10**, 30–36 (1983).
123. W. Wriggers, E. Mehler, F. Pitici, H. Weinstein, K.S. Schulten, 'Structure and Dynamics of Calmodulin in Solution', *Biophys. J.*, **74**, 1622–1639 (1998).
124. C.L. Brooks III, M. Karplus, 'Solvent Effects on Protein Motion and Protein Effects on Solvent Motion. Dynamics of the Active Site Region of Lysozyme', *J. Mol. Biol.*, **208**, 159–181 (1989).
125. K. Haydock, C. Lim, A. Brunger, M. Karplus, 'Simulation Analysis of Structures on the Reaction Pathway of RNAse A', *J. Am. Chem. Soc.*, **112**, 3826–3831 (1990).
126. E.A. Padlan, W.E. Love, 'Refined Crystal Structure of Deoxyhemoglobin S. I. Restrained Least-squares Refinement at 3.0 Å Resolution and II. Molecular Interactions in the Crystal', *J. Biol. Chem.*, **260**, 8272–8279 and 8280–8291 (1985).
127. K. Kuczera, 'Free Energy Simulations of Axial Contacts in Sickle-cell Hemoglobin', *Biopolymers*, **39**, 221–242 (1996).
128. B. Honig, A. Nicholls, 'Classical Electrostatics in Biology and Chemistry', *Science*, **268**, 1144–1149 (1995).
129. M. Born, 'Volumen und Hydrationswarme Der Ionen', *Z. Phys.*, **1**, 45–48 (1920).
130. L. Onsager, 'Electric Moments of Molecules in Liquids', *J. Am. Chem. Soc.*, **58**, 1486–1493 (1936).
131. J.G. Kirkwood, 'Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions', *J. Chem. Phys.*, **2**, 351–361 (1934).
132. R. Constanciel, R. Contreras, 'Self-consistent Field Theory of Solvent Effect Representation by Continuum Models – Introduction of Desolvation Contribution', *Theor. Chim. Acta*, **65**, 1–11 (1984).
133. K.A. Sharp, B. Honig, 'Electrostatic Interactions in Macromolecules: Theory and Application', *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 301 (1990).
134. T. Simonson, A. Brunger, 'Solvation Free Energies Estimated from Macroscopic Continuum Theory: An Accuracy Assessment', *J. Phys. Chem.*, **98**, 4683–4694 (1994).
135. W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, 'Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics', *J. Am. Chem. Soc.*, **112**, 6127–6129 (1990).
136. J.D. Madura, M.E. Davis, M.K. Gilson, R.C. Wade, B.A. Luty, J.A. McCammon, 'Biological Applications of Electrostatic Calculations and Brownian Dynamics Simulations', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH Publishers, New York, Vol. V, 1994.
137. J.L. Smart, T.J. Marrone, J.A. McCammon, 'Conformational Sampling with Poisson–Boltzmann Forces and a Stochastic Dynamics/Monte Carlo Method: Application to Alanine Dipeptide', *J. Comput. Chem.*, **18**, 1750–1759 (1997).
138. D. Eisenberg, A.D. McLachlan, 'Solvation Energy in Protein Folding and Binding', *Nature*, **319**, 199–203 (1986).
139. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, 'Comparison of Simple Potential Functions for Simulating Liquid Water', *J. Chem. Phys.*, **79**, 926–935 (1983).
140. C.A. Hunter, M. Saunders, 'The Nature of $\pi-\pi$ Interactions', *J. Am. Chem. Soc.*, **112**, 5525–5534 (1990).
141. A.J. Stone, M. Alderton, 'Distributed Multipole Analysis Methods and Applications', *Mol. Phys.*, **56**, 1047–1064 (1985).
142. S.L. Price, R.J. Harrison, M.F. Guest, 'An *ab initio* Distributed Multipole Study of the Electrostatic Potential Around an Undecapeptide Cyclosporin Derivative and a Comparison with Point Charge Electrostatic Models', *J. Comp. Chem.*, **10**, 552–567 (1989).

143. D.E. Williams, 'Alanyl Dipeptide Potential-derived Net Atomic Charges and Bond Dipoles, and their Variation with Molecular Conformation', *Biopolymers*, **29**, 1367–1386 (1990).
144. A.K. Rappe, W.A. Goddard, II, 'Charge Equilibration for Molecular Dynamics Simulations', *J. Phys. Chem.*, **95**, 3358–3363 (1991).
145. U. Dinur, A.T. Hagler, 'Geometry-dependent Atomic Charges: Methodology and Application to Alkanes, Aldehydes, Ketones and Amides', *J. Comp. Chem.*, **11**, 159–169 (1995).
146. P. Ewald, 'Die Berechnung Optischer und Elektrostatische Gitterpotentiale', *Ann. Phys.*, **64**, 253–287 (1921).
147. T.A. Darden, A. Toukmaji, L.G. Pedersen, 'Long-range Electrostatic Effects in Biomolecular Simulations', *J. Chim. Phys.-Chim. Biol.*, **94**, 1346–1364 (1997).
148. L. Greengard, V.I. Roklin, 'A Fast Algorithm for Particle Simulations', *J. Comput. Phys.*, **73**, 325–348 (1987).
149. H.-Q. Ding, N. Karasawa, W.A. Goddard III, 'The Reduce Cell Multipole Method for Coulomb Interactions in Periodic Systems with Million-atom Unit Cells', *Chem. Phys. Lett.*, **196**, 6–10 (1992).
150. S. Fischer, M. Karplus, 'Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom', *Chem. Phys. Lett.*, **194**, 252–261 (1992).
151. S.E. Huston, G.E. Marshall, '$\alpha/3_{10}$-helix Transitions in α-methylalanine Homopeptides: Conformational Transition Pathway and Potential of Mean Force', *Biopolymers*, **34**, 75–90 (1994).
152. W. Nowak, R. Czerminski, R. Elber, 'Reaction Path Study of Ligand Diffusion in Proteins: Application of the Self Penalty Walk (SPW) Method to Calculate Reaction Coordinates for the Motion of CO Through Leghemoglobin', *J. Am. Chem. Soc.*, **113**, 5627–5637 (1991).
153. D. Ringe, G.A. Petsko, 'Fluctuations in Protein Structure from X-ray Diffraction', *Ann. Rev. Biophys. Bioeng.*, **13**, 331–371 (1984).
154. A.T. Brunger, J. Kuriyan, M. Karplus, 'Crystallographic R Factor Refinement by Molecular Dynamics', *Science*, **235**, 458–460 (1987).
155. J. Smith, B. Tidor, W. Doster, S. Cusack, M. Karplus, 'Internal Dynamics of Globular Proteins: Comparison of Neutron Scattering Measurements and Theoretical Models', *Physica B*, **156–157**, 437–443 (1989).
156. J. Smith, K. Kuczera, M. Karplus, 'Molecular Dynamics of Myoglobin: Comparison of Simulation Results with Neutron Scattering Spectra', *Proc. Natl. Acad. Sci. USA*, **87**, 1601–1605 (1990).
157. F. Parak, H. Hartmann, K.D. Aumann, H. Reuscher, G. Rennekamp, H. Bartunik, W. Steigemann, 'Low Temperature X-ray Investigation of Structural Distributions in Myoglobin', *Eur. Biophys. J.*, **15**, 237–249 (1987).
158. K. Kuczera, 'Dynamics and Thermodynamics of the Globins', in *Recent Developments in Theoretical Studies of Proteins*, ed. R. Elber, World Scientific, Singapore, 1–63, 1996.
159. A.E. Torda, W.F van Gunsteren, 'Molecular Modeling Using Nuclear Magnetic Resonance Data', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH Publishers, Inc., New York, Vol. 3, 1992.
160. A.T. Brunger, G.M. Clore, A.M. Gronenborn, M. Karplus, 'Three-dimensional Structure of Proteins Determined by Molecular Dynamics with Interproton Distance Restraints: Application to Crambin', *Proc. Natl. Acad. Sci. USA*, **83**, 3801–3805 (1986).
161 G. Lipari, A. Szabo, *J. Am. Chem. Soc.*, **104**, 4546–4559 (1982).
162. Y. Bai, T.R. Sosnick, L. Mayne, S.W. Englander, 'Protein Folding Intermediates: Native-state Hydrogen Exchange', *Science*, **269**, 192–197 (1995).
163. S.W. Englander, N.R. Kallenbach, 'Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids', *Quart. Rev. Biophys.*, **16**, 521 (1983).
164. J.R. Lakowicz, B.P. Maliwal, H. Cherek, A. Balter, 'Rotational Freedom of Tryptophan Residues in Proteins and Peptides', *Biochemistry*, **22**, 1741–1752 (1983).
165. M. Ehrenberg, R. Rigler, 'Polarized Fluorescence and Rotational Brownian Motion', *Chem. Phys. Lett.*, **14**, 539–544 (1972).
166. T.J. Chuang, K.B. Eisenthal, 'Theory of Fluorescence Depolarization by Anisotropic Rotational Diffusion', *J. Chem. Phys.*, **57**, 5094–5097 (1972).
167. G.S. Jas, Y. Wang, S.W. Pauls, C.K. Johnson, K. Kuczera, 'Influence of Temperature and Viscosity on Anthracene Rotational Diffusion in Organic Solvents: Molecular Dynamics Simulations and Fluorescence Anisotropy Study', *J. Chem. Phys.*, **107**, 8800–8812 (1997).
168. A. Dong, P. Huang, W.S. Caughey, 'Protein Secondary Structures in Water from Second-derivative Amide I Infrared Spectra', *Biochemistry*, **29**, 3303–3308 (1990).
169. F. Dousseau, M. Pezolet, 'Determination of the Secondary Structure Content of Proteins in Aqueous Solutions from their Amide I and Amide II Infrared Bands. Comparison Between Classical and Partial Least-squares Methods', *Biochemistry*, **29**, 8771–8779 (1990).
170. L.A. Komives, L.C. Chang, E. Lolis, R.F. Tilton, G.A. Petsko, J.R. Knowles, 'Electrophilic Catalysis in Triosephosphate Isomerase: The Role of Histidine-95', *Biochemistry*, **30**, 3011–3019 (1991).
171. D.L. Rousseau, M.R. Ondrias, 'Resonance Raman Spectra of Dissociated Hemoglobins', *Biophys. J.*, **47**, 537–545 (1985).
172. J.C. Smith, 'Dynamics of Biomolecules: Simulation Versus X-ray, Neutron and Infrared Experiment', in *Computer Simulations of Biomolecular Systems. Theoretical*

and Experimental Applications, eds. W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson, Kluwer Academic, Dordrecht, 305–360, 1997.

173. C.B. Post, B.R. Brooks, M. Karplus, C.M. Dobson, P.J. Artymiuk, J.C. Cheetham, D.C. Phillips, 'Molecular Dynamics Simulations of Native and Substrate-bound Lysozyme', *J. Mol. Biol.*, **190**, 455–479 (1986).
174. T. Ichiye, M. Karplus, 'Anisotropy and Anharmonicity of Atomic Fluctuations in Proteins: Analysis of a Molecular Dynamics Simulation', *Proteins: Struct., Funct., Genet.*, **2**, 236–259 (1987).
175. J.A. McCammon, P.G. Wolynes, M. Karplus, 'Picosecond Dynamics of Tyrosine Sidechains in Proteins', *Biochemistry*, **18**, 927–942 (1979).
176. W.F. van Gunsteren, P.H. Huhnenberger, A.E. Mark, P.E. Smith, I.G. Tironi, 'Computer Simulation of Protein Motion', *Comput. Phys. Commun.*, **91**, 305–319 (1995).
177. W.F. van Gunsteren, P.H. Huhnenberger, H. Kovacs, A.E. Mark, C.A. Schiffer, 'Investigation of Protein Unfolding and Stability by Computer Simulation', *Phil. Trans. R. Soc. Lond. B*, **348**, 49–59 (1995).
178. P.J. Steinbach, A. Ansari, J. Berendzen, D. Braunstein, K. Chu, B.R. Cowen, D. Ehrenstein, H. Frauenfelder, J.B. Johnson, D.C. Lamb, S. Luck, J.R. Mourant, G.U. Nienhaus, P. Ormos, R. Philipp, A. Xie, R.D. Young, 'Ligand Binding to Heme Proteins: Connection Between Dynamics and Function', *Biochemistry*, **30**, 3988–4001 (1991).
179. S.E.V. Phillips, 'Structure and Refinement of Oxomyoglobin at 1.6 Å Resolution', *J. Mol. Biol.*, **142**, 531–554 (1980).
180. D.A. Case, M. Karplus, 'Dynamics of Ligand Binding to Heme Proteins', *J. Mol. Biol.*, **132**, 343–368 (1979).
181. R. Elber, M. Karplus, 'Enhanced Sampling in Molecular Dynamics: Use of the Time-dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion Through Myoglobin', *J. Am. Chem. Soc.*, **112**, 9161–9175 (1990).
182. A. Ulitsky, R. Elber, 'Application of the Locally Enhanced Sampling (LES) and a Mean Field with Binary Collision Correction (cLES) to the Simulation of Ar Diffusion and NO Recombination', *J. Phys. Chem.*, **98**, 1034–1043 (1994).
183. J.W. Petrich, J.-C. Lambry, K. Kuczera, M. Karplus, C. Poyart, J.-L. Martin, 'Ligand Binding and Protein Relaxation in Heme Proteins: A Room Temperature Analysis of NO Geminate Recombination', *Biochemistry*, **30**, 3975–3987 (1991).
184. D. Joseph, G.A. Petsko, M. Karplus, 'Anatomy of a Conformational Change: Hinged ''Lid'' Motion of the Triosephosphate Isomerase Loop', *Science*, **249**, 1425–1428 (1990).
185. M. Philippopoulos, Y. Xiang, C. Lim, 'Identifying the Mechanism of Protein Loop Closure: A Molecular Dynamics Simulation of the Bacillus Stearothermophilus LDH Loop Closure', *Protein Eng.*, **8**, 565–573 (1995).
186. G.B. Dryer, D.M. Lambert, T.D. Meek, T.J. Carr, T.A. Tomaszek, A.V. Fernandez, H. Bartus, E. Cacciavilani, A.M. Hassell, M. Minnich, S.R. Petteway Jr, B. Metcalf, 'Hydroxyethylene Isostere Inhibitors of HIV-1 Protease Structure-activity Analysis Using Enzyme Kinetics, X-ray Crystallography, and Infected T-cell Assays', *Biochemistry*, **31**, 6646 (1992).
187. S. Swaminathan, W.E. Harte Jr, D.L. Beveridge, 'Investigation of Domain Structure in Proteins Via Molecular Dynamics Simulation: Application to HIV-1 Protease Dimer', *J. Am. Chem. Soc.*, **113**, 2717 (1991).
188. G.V. Nikiforovich, 'Computational Molecular Modeling in Peptide Drug Design', *Int. J. Peptide Protein Res.*, **44**, 513–531 (1995).
189. G.E. Marlow, J.S. Perkyns, B.M. Pettitt, 'Salt Effects in Peptide Solutions: Theory and Simulation', *Chem. Rev.*, **93**, 2503–2521 (1993).
190. Y. Wang, S.-Y. Goh, K. Kuczera, 'Molecular Dynamics Study of Disulfide Bond Influence on Properties of an RGD Peptide', *J. Peptide Res.*, **53**, 188–200 (1999).
191. Y. Wang, K. Kuczera, 'Molecular Dynamics Simulations of Cyclic and Linear DPDPE: Influence of the Disulfide Bond on Peptide Flexibility', *J. Phys. Chem.*, **100**, 2555–2563 (1996).
192. Y. Wang, K. Kuczera, 'Conformational Free Energy Surface of the Linear DPDPE Peptide: Cost of Preorganization for Disulfide Bond Formation', *Theor. Chem. Acc.*, **101**, 274–281 (1999).
193. J. Tirado-Rives, W.L. Jorgensen, 'Molecular Dynamics Simulations of the Unfolding of an α-Helical Analog of Ribonuclease A S-peptide in Water', *Biochemistry*, **30**, 3864–3871 (1991).
194. K.V. Soman, A. Karimi, D.A. Case, 'Unfolding of an α-Helix in Water', *Biopolymers*, **31**, 1351–1361 (1991).
195. V. Daggett, M. Levitt, 'Molecular Dynamics Simulations of Helix Denaturation', *J. Mol. Biol.*, **223**, 1121–1138 (1992).
196. C.L. Brooks III, 'Molecular Simulations of Peptide and Protein Unfolding: In Quest of a Molten Globule', *Curr. Opin. Struct. Biol.*, **3**, 92–98 (1993).
197. W.S. Young, C.L. Brooks III, 'A Microscopic View of Helix Propagation: N and C-terminal Helix Growth in Alanine Helices', *J. Mol. Biol.*, **259**, 560–572 (1996).
198. C.L. Brooks III, 'Helix-coil Kinetics: Folding Timescales for Helical Peptides from a Sequential Kinetic Model', *J. Phys. Chem.*, **100**, 2546–2549 (1996).
199. D.H. Appella, J.J. Banchi Jr, S.R. Durell, S.H. Gellman, 'Formation of Short, Stable Helices in Aqueous Solution by β-Amino Acid Hexamers', *J. Am. Chem. Soc.*, **121**, 2309–2310 (1999).
200. X. Daura, W.F. Van Gunsteren, A.E. Mark, 'Folding–Unfolding Thermodynamics of a β-heptapeptide from Equilibrium Simulations', *Proteins: Struct., Funct., Genet.*, **34**, 269–280 (1999).

201. B. Tidor, M. Karplus, 'Simulation Analysis of the Stability Mutant R96H of T4 Lysozyme', *Biochemistry*, **30**, 3217–3228 (1991).
202. J. Hermans, 'Molecular Dynamics Simulations of Helix and Turn Propensity in Model Peptides', *Curr. Opin. Struct. Biol.*, **3**, 270–276 (1993).
203. Y. Wang, K. Kuczera, 'Exploration of Conformational Free Energy Surface of Helical Ala_n and Aib_n Peptides. *J. Phys. Chem. B*, **101**, 5205–5213 (1997).
204. M.J. Field, P.A. Bash, M. Karplus, 'A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations', *J. Comp. Chem.*, **11**, 700–733 (1990).
205. J. Gao, X. Xia, 'A Priori Evaluation of Aqueous Polarization Effects Through Monte Carlo QM/MM', *Science*, **258**, 631–635 (1992).
206. H.J.C. Berendsen, 'Molecular Dynamics Simulations: The Limits and Beyond', in *Computational Molecular Dynamics: Challenges, Methods, Ideas*, eds. P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, R.D. Skeel, Springer, Berlin, 3–36, 1999.
207. X. Wu, S. Wang, 'Self-guided Molecular Dynamics Simulations for Efficient Conformational Search', *J. Phys. Chem. B*, **102**, 7238–7250 (1998).
208. R.L. Jernigan, I. Bahar, 'Structure-derived Potentials and Protein Simulations', *Curr. Opinion Struct. Biol.*, **6**, 195–209 (1996).
209. A. Šali, J.P. Overington, M.S. Johnson, T.S. Blundell, 'From Comparisons of Protein Sequences and Structures to Protein Modelling and Design', *Trends Biochem. Sci.*, **15**, 235–240 (1990).
210. B.K. Shoichet, R.M. Stroud, D.V. Santi, I.D. Kuntz, K.M. Perry, 'Structure-based Discovery of Inhibitors for Thymidylate Synthase', *Science*, **259**, 1445–1450 (1993).
211. G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, 'Automated Docking Using Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function', *J. Comp. Chem.*, **19**, 1639–1662 (1998).
212. H.S. Chan, K.A. Dill, 'The Protein Folding Problem', *Physics Today*, Feb., 24–32 (1993).
213. A. Godzik, A. Kolinski, J. Skolnick, 'De novo and Inverse Folding Predictions of Protein Structure and Dynamics', *J. Computer-aided Mol. Des.*, **7**, 397–438 (1993).
214. P.J. Kraulis, 'MOLSCRIPT: a Program to Produce Both Detailed and Schematic Plots of Protein Structures', *J. Appl. Cryst.*, **24**, 946–950 (1991).
215. D.J. Bacon, W.F. Anderson, 'A Fast Algorithm for Rendering Space-filling Molecular Pictures', *J. Mol. Graphics*, **6**, 219–220 (1988).
216. E.A. Merritt, E.A.P. Murphy, 'Raster3D Version 2.0 – A Program for Photorealistic Molecular Graphics', *Acta Cryst. D*, **50**, 869–873 (1994).

Peptide Diastereomers, Separation of

Gerhard K.E. Scriba
University of Jena, Jena, Germany

1 Introduction **1**
2 Chromatographic Separations **2**
2.1 Paper Chromatography 2
2.2 Thin-layer Chromatography 3
2.3 Gas Chromatography 4
2.4 High-performance Liquid Chromatography 4
2.5 Method Development 8
3 Electrophoretic Separations **8**
3.1 Capillary Zone Electrophoresis 10
3.2 Micellar Electrokinetic Chromatography 10
3.3 Method Development 11
Abbreviations and Acronyms **11**
Related Articles **12**
References **12**

Diastereomeric peptides, isomers in which one or more of the chiral centers have been converted to the opposite configuration (R or S), often possess different biological activities and/or conformational properties. Therefore, the separation and determination of diastereomeric peptides is particularly significant to the pharmaceutical industry for quality control of peptide synthesis and stability as well as for regulatory requirements. Diastereomers differ in their physicochemical properties which can be utilized for their determination by numerous techniques. In contrast to many other methods such as optical rotation, circular dichroism, differential scanning calorimetry, infrared (IR) spectroscopy, or nuclear magnetic resonance (NMR) spectroscopy only chromatographic techniques and capillary electrophoresis (CE) allow the sensitive and simultaneous identification and quantification of peptide diastereomers.

The current chapter will cover the analytical scale separation of peptide diastereomers by chromatographic methods, including paper chromatography (PC), thin-layer chromatography (TLC), gas chromatography (GC), and high-performance liquid chromatography (HPLC) as well as CE. HPLC is currently the predominant method for peptide diastereomer separations. However, due to the high resolving power particularly for polar compounds CE has been applied to the separation of diastereomeric peptides in recent years. This article will cover methods employing achiral stationary phases or buffers without chiral additives for the separations. Moreover, only examples of diastereomeric peptides containing natural amino acids will be discussed.

1 INTRODUCTION

The separation and identification of optical isomers of synthetic or naturally occurring peptides is of great importance. Diastereomeric peptides, isomers in which one or more of the chiral centers have been converted to the opposite configuration, (*R*) or (*S*), often possess different biological activities and/or conformational properties. Peptide diastereomers may result from the synthesis methodology, which may lead to enantiomerization of one or more amino acid residues, the use of optically impure amino acid derivatives during synthesis, or enantiomerization during molecular ageing. Therefore, the separation and determination of diastereomeric peptides is particularly significant to the pharmaceutical industry for quality control of peptide synthesis and stability as well as for regulatory requirements.

Diastereomers are stereoisomers that are not related as mirror images, whereas epimers are defined as diastereomers differing in the configuration of only one chiral center.[1] Diastereomers differ in their physicochemical properties and often possess different conformations. They can be determined by many techniques such as crystallization, enzymatic reactions, optical rotation, circular dichroism, differential scanning calorimetry, IR spectroscopy, NMR spectroscopy, mass spectrometry (MS), as well as chromatographic and capillary electrophoretic separation methods.[1,2] Although it is generally possible to distinguish between diastereomers with all techniques mentioned, only chromatography and CE allow the simultaneous separation, identification and determination of even complex mixtures of diastereomers. Epimeric impurities well below 1% can be detected and quantified.

Applications of the analysis of peptide diastereomers include the indirect determination of the optical purity of amino acids, the detection of epimerization of amino acid residues during peptide synthesis and degradation, and the direct determination of the optical purity of peptides. Although systematic studies have been performed primarily with di- and tripeptides, specific procedures have been developed for certain compounds or closely related analogs.

This article covers analytical-scale separations of peptide diastereomers by chromatographic and electrophoretic techniques. Preparative procedures are not discussed. The article also focuses on peptides containing naturally occurring amino acids, although some analogs

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

are listed in a few cases. Diastereomeric diketopiperazines derived from dipeptides are included. A review on the chromatographic separation of peptides containing β-alkyl amino acids has been published recently by Peter and Toth.[3]

2 CHROMATOGRAPHIC SEPARATIONS

Chromatographic separations of diastereomeric peptides have been reported by PC, TLC, GC and HPLC. No peptide diastereomer resolution by supercritical fluid chromatography (SFC) has been published. This may be due to the low solubility of native peptides in supercritical carbon dioxide. Today, HPLC is the most common technique for the resolution of peptide diastereomers.

Separation mechanisms in chromatography include adsorption and/or distribution between the stationary phase and the mobile phase as well as ion exchange phenomena. Differences in the adsorption/distribution characteristics between diastereomers result in their separation.

With a few exceptions in which resolutions of diastereomeric peptides were performed using chiral stationary phases or chiral additives, most published separations were achieved using achiral stationary phases. Although it is generally possible to separate diastereomers on chiral phases, this topic is not covered in the article. Chiral stationary phases and separation methods are treated in monographs on chiral chromatographic separations[4–6] as well as in the relevant articles of this encyclopedia – **Gas Chromatography: Introduction**; **Column Technology in Gas Chromatography**; **Liquid Chromatography: Introduction**; **Column Theory and Resolution in Liquid Chromatography**; **Normal-phase Liquid Chromatography**; **Reversed Phase Liquid Chromatography** and **Thin-layer Chromatography**.

2.1 Paper Chromatography

The first separation of the diastereomers of a dipeptide was reported in 1950 by PC.[7] Since then numerous di- and tripeptides have been separated primarily on Whatman® No. 1 paper by ascending or descending techniques (Table 1). The peptides were detected employing the ninhydrin reaction. Although most early reports used acidic solvent systems,[7–11] many basic solvent mixtures have been applied even more successfully in subsequent work.[9–14] Generally, dipeptides are resolved more efficiently than tripeptides. Higher R_f values were found for the LL/DD diastereomeric pairs of dipeptides compared to the DL/LD epimers. A diastereomeric hexapeptide has also been resolved, the LDLLLL epimer displaying a higher R_f value than the LLLLLL isomer.[11] The diastereomers

Table 1 Separation of diastereomeric peptides by PC

Peptide	Stationary phase	Mobile phase composition (v/v)[a]	Detection	R_f sequence[b]	Ref.
Val-Val	Whatman® No. 1	*n*-Butanol–acetic acid–water (2:1:1)	–	LL > LD	7
Ala-Ala	Whatman® 3MM	*n*-Butanol–acetic acid–water (4:1:1)	Ninhydrin	LL > LD	10
		Pyridine–water (4:1)	Ninhydrin	LL > LD	
Dipeptides	Whatman® No. 1	Ethyl acetate–pyridine–acetic acid–water (5:5:1:3)	Ninhydrin	LL > LD	12
		Pyridine–water (4:1)	Ninhydrin	LL > LD	
Dipeptides	Whatman® No. 1	Ethyl acetate–pyridine–acetic acid–water (5:5:1:3) *n*-Butanol–acetic acid–water (4:1:5) *n*-Butanol–acetic acid–water (4:1:1) Pyridine–water (4:1) Isoamylalcohol–pyridine–water (7:7:6)	Ninhydrin	LL > LD	9
Lys-Glu, Lys-Asp, Gly-Glu-Lys, Gly-Asp-Lys	Toyo Roshi® 52	*n*-Butanol–acetic acid–pyridine–water (4:1:1:2)	–	LL > LD	13
Leu dipeptides	Toyo Roshi® 52	Pyridine–water (4:1) *n*-Butanol–acetic acid–pyridine–water (4:1:1:2)	–	LL > LD	14
Leu dipeptides	Amberlite® SA-II, NH_4^+ form	0.2 M NH_4OAc, pH 4.5 or pH 7.0	–	LL > LD	14
Glu tripeptides	Whatman® No. 1	*n*-Butanol–acetic acid–water (5:1:4)	–	–	8
Val-D/L-Tyr-Val-His-Pro-Phe	Whatman® No. 1	2-Butanol–2-propanol–monochloroacetic acid–water (7:1:0.3:4)	–	LDLLLL > LLLLLL	11

[a] Optimized conditions.
[b] Higher R_f value listed first.

of Leu dipeptides have been separated by ion-exchange PC using Amberlite® SA-II paper in its ammonium form.[14]

2.2 Thin-layer Chromatography

Separations of diastereomeric di- and tripeptides as free[15,16] or protected derivatives[17–19] by TLC have been achieved using cellulose,[15] silica gel,[15,19,20] reversed-phase (RP) materials,[16] and RP phases impregnated with cationic or anionic detergents[16] or ammonium tungstophosphate (AWP)-coated plates[16] as stationary phases (Table 2). With a few exceptions acidic solvent systems have been employed.

Higher R_f values were observed for the LL/DD diastereomeric pairs of free dipeptides compared to the corresponding LD/DL epimers on cellulose,[15] silica gel[15,20] and RP materials.[16] In contrast, AWP as adsorbent retained the LL/DD epimers more strongly than the LD/LD diastereomers, resulting in a reversal of the R_f sequence.[16] The retention order on

Table 2 Separation of diastereomeric peptides by TLC

Peptide	Stationary phase[a]	Mobile phase composition (v/v)[b]	Detection	R_f sequence[c]	Ref.
Phe-Phe, Phe-Val	Silica gel 60	*n*-Butanol–acetic acid–water (4:1:1)	Ninhydrin/cupric nitrate	LL > LD	20
Z-Phe-LeuOMe, Z-Phe-Leu	Silica gel 60	*n*-Butanol–acetic acid–water (4:1:1) *n*-Butanol–acetic acid–pyridine–water (15:10:3:2)	Ninhydrin or iodine	Depending on derivative and mobile phase	19
Phe-Leu		Phenol–water (3:1) Isopropanol–water (5:1)			
Met-Ala, Ala-Phe, Ala-Tyr, Val-Tyr	Cellulose MN 300	Pyridine–water (1:1)	Ninhydrin	LL > LD	15
	Silica gel 60	Methanol–water (100:1)	Ninhydrin	LL > LD	
Dipeptides, Ala-Ala-Ala	OPTI-UP® C_{12}	1 M Acetic acid 3% KCl in 1 M Acetic acid 0.5 M NaOAc in water 1 M Acetic acid in water–methanol (30%)	Ninhydrin	LL > LD	16
Ala-Leu, Leu-Leu, Leu-Tyr	Sil C_{18}-50	1 M Acetic acid in water–methanol (30%)	Ninhydrin	LL > LD	16
Ala-Leu, Leu-Leu, Leu-Tyr, Tyr-Arg, Ala-Ala-Ala	4% HDBS coated Sil C_{18}-50	1 M Acetic acid + 0.1 M HCl in water–methanol (20%) 1 M Acetic acid + 1 M HCl in water–methanol (20%)	Ninhydrin	LL > LD	16
Leu-Tyr, Tyr-Arg	4% *N*-DPC coated Sil C_{18}-50	0.5 M NaOAc in water–methanol (20%)	Ninhydrin	LD > LL	16
Ala-Ala, Lal-Leu, Leu-Tyr, Tyr-Arg, Ala-Ala-Ala	AWP	0.5–3 M NH_4NO_3 in water	Ninhydrin	LD > LL	16
NPS-Met-Ala-ONP[d]	Silica gel 60	Acetic acid–diethyl ether (1.5:20)	Iodine	LL > LD	18
NPS-Met-Met-Ala-ONP[d]	Silica gel 60	Acetic acid–diethyl ether (0.2:20)	Iodine	LLD > LLL	18
N-Protected dipeptide esters	Silica gel 60	Diisopropyl ether–isopropanol (10:1 or 4:1) Benzene–acetone (2:1 or 3:1) Diisopropyl ether; ethyl acetate	40% NH_4SO_4 and heating to 180 °C	Depending on derivative and mobile phase	17
Leu-Leu, Ala-Ala, Ala-Phe, Met-Met	Chiralplate®	Methanol–water–acetonitrile (5:5:3)	Ninhydrin	LD > LL	21
Leu-Leu	Chir® HPTLC	Methanol–*n*-propanol–water (5:1:4)	Ninhydrin	LD > LL	22
Somatostatin	Silica gel 60	2-Propanol–1 M acetic acid (1:1)	Ninhydrin	–	23
Neurotensin analogs	Silica gel	1-Butanol–acetic acid–water (4:1:5, upper phase) 2-Butanol–water (1:1, upper phase)	Ninhydrin	–	24
Bombesin analogs	Silica gel 60	2-Butanol–0.1 M acetic acid (1:1, upper phase)	Ninhydrin	–	25

[a] *N*-DPC, *N*-Dodecylpyridinium chloride; HDBS, dodecylbenzene sulfuric acid; HPTLC, high-performance thin-layer chromatography.
[b] Optimized conditions.
[c] Higher R_f value listed first.
[d] ONP, *o*-Nitrophenyl ester.

detergent-impregnated RP materials depended on the type of detergent used. Thus, the LL/DD pair displayed higher R_f values than the LD/DL pair when the anionic detergent HDBS was used in combination with acetic solvent mixtures, whereas the retention order was reversed for the cationic detergent *N*-DPC and for basic mobile phases.[16] The retention order of protected dipeptides strongly depended on the type of derivatization of the N- or C-terminus.[17] The all-L stereoisomer of the tripeptide Ala-Ala-Ala was less retained than the mixed epimers with L-L-D or L-D-L configuration on RP/TLC plates and RP materials coated with an anionic detergent, whereas the L-L-L isomer was retained more strongly than the L-D-L epimer on AWP-coated TLC plates.[16]

The resolution of free dipeptides by ligand exchange on chiral stationary phases has been reported.[21,22] The DL isomers were less retained than the LL isomers of all dipeptides investigated.

Diketopiperazines derived from diastereomeric dipeptides could also be resolved on silica gel plates using acetic solvent mixtures[26–28] (Table 3). The DL epimers with *trans*-substituted piperazine-2,5-dione moiety displayed higher R_f values than the LL isomers with *cis* conformation of the substituents.

2.3 Gas Chromatography

The separation of peptides by GC generally requires the derivatization of the amino and the carboxyl group. *N*-Acyl-peptide esters have been primarily investigated. Only one report described trimethylsilylation for derivatization. The majority of analyses have been performed on packed columns. Various stationary phases have been investigated. Weygand et al.[29] first succeeded in separating the diastereomers of Ala-Phe as the *N*-trifluoroacetyl (*N*-TFA) methyl ester. Later the same group reported the resolution of many *N*-TFA dipeptide methyl esters.[30] Except for Ala-Ala, the DL epimers were retained stronger than the LL isomers. The trimethylsilyl (TMS) derivatives of dipeptides could be resolved on a 12-m SE-54 fused-silica capillary column.[31] The LL/DD isomers had shorter retention times than the LD/DL isomers. The separation of *N*-TFA and *N*-tetrafluoropropionyl (*N*-TFP) tetrapeptides obtained by reaction of *N*-TFA-propyl and *N*-TFP-propyl derivatizing reagents and tripeptide methyl esters using packed columns were described, the LDLL epimers eluting first (Table 4).[32,33]

Diastereomeric diketopiperazines can also be resolved by GC using several stationary phases[28,37] (Table 5).

2.4 High-performance Liquid Chromatography

This is definitely the most widely applied technique for the separation of peptide diastereomers. Many papers on systematic studies and specific applications have been published and numerous additional applications exist in synthetic papers and patents. Two short reviews on the resolution of peptide diastereomers have appeared [38,39] and several monographs are dedicated to the HPLC of peptides and proteins.[40,41]

Systematic studies have been performed with di-, tri-, and tetrapeptides, whereas the resolution of larger peptides addressed specific problems (Table 6). The great majority of peptides were analyzed as free compounds. Only a few studies were performed with N-derivatized esters of di-, tri-, and tetrapeptides.[42–52]

Three basic principles have been applied in HPLC: RP chromatography (which dominates), ion-exchange chromatography and normal-phase chromatography. Only one publication gives a single example for the use of a peptide-based column.[53]

The reversed-phase high-performance liquid chromatography (RPHPLC) method has been primarily performed using octadecyl silica (ODS) supports, although C-8, phenyl, cyano, or macroporous polystyrene–divinylbenzene copolymers were also employed. To ensure sufficient interaction between the highly polar peptides and the hydrophobic stationary phase, analyses are usually carried out in the acidic pH range although separations up to pH 9 have been reported.[54] The mobile phases consisted of aqueous solutions of acetic acid, TFA), phosphoric acid, phosphate buffers, ammonium acetate, TEAP or triethylammonium acetate (TEAA) mixed with water-soluble organic solvents such as acetonitrile, methanol, or

Table 3 Separation of diastereomeric diketopiperazines by TLC

Diketopiperazine	Stationary phase	Mobile phase composition (v/v)[a]	Detection	R_f sequence[b]	Ref.
c(Leu-Tyr)	Silica gel 60	2-Butanol–methanol–acetic acid (70:10:5) Isopropyl ether–chloroform–acetic acid (6:3:1)	Hypochlorite–iodine	LD > LL	26
Diketopiperazines	Silica gel 60	Isopropyl ether–chloroform–acetic acid (6:3:1) Chloroform–methanol–acetic acid (14:2:1)	Hypochlorite–iodine	LD > LL	27
Diketopiperazines	Silica gel 60	Isopropyl ether–chloroform–acetic acid (6:3:1) Chloroform–methanol–acetic acid (14:2:1)	Hypochlorite–iodine	LD > LL	28

[a] Optimized conditions.
[b] Higher R_f value listed first.

Table 4 Separation of diastereomeric peptides by GC

Peptide derivative	Stationary phase	Mobile phase	Detection[a]	Sequence[b]	Refs.
N-TFA-Ala-Phe-OMe	Silicone on Celite	Helium	FID	LL < LD	29
N-TFA-Val-Ala-OMe	Carbowax® 20 M capillary column	Nitrogen	FID	LD < LL	27
N-TFA-dipeptide-OMe	QF-1 on Chromosorb® W	Nitrogen	FID	Primarily LL < LD	27
N-TFA-dipeptide-OMe	Polypropylene glycol-coated steel capillary Column polyphenyl ether-coated steel capillary column	Nitrogen	FID	LL < LD	30
N-Chloroacetyl-dipeptide-OMe	5% SE-30 on Chromosorb® W 5% FFAP[c] on Chromosorb® W 20% DEGS[d] on Chromosorb® W	Nitrogen	FID	LD < LL	34, 35
N-TFA-propyl-dipeptide-O-TMS	0.5% Polyethylene glycol on Chromosorb® W	Nitrogen	FID	LD < LL	36
N-TFA-Pro-D/L-Val-Pro-Val-OMe	5% OV-17 on Chromosorb® W	Nitrogen	FID	LDLL < LLLL	32
N-TFA-Pro-Val-OMe				LD < LL	
N-TFP-Pro-D/L-Val-Pro-Val-OMe				LDLL < LLLL	
N-TFP-Pro-Val-OMe				LD < LL	
N-TFA-tetrapeptide-OMe	5% OV-17 on Chromosorb® W	Nitrogen	FID	–	33
TMS-dipeptides	SE-54 Fused-silica capillary	Helium	FID	LL < LD	31

[a] FID, Flame ionization detection.
[b] Shorter retention time listed first.
[c] FFAP, Free fatty acid phase.
[d] DEGS, Diethyleneglycol succinate.

Table 5 Separation of diastereomeric diketopiperazines by GC

Derivative	Stationary phase	Mobile phase	Detection	Sequence[a]	Ref.
Underivatized	5% QF-1 on Aeropak eO	Nitrogen	FID	Primarily LD < LL	28
Underivatized	3% EGS on Gas Chrom Q	Nitrogen	FID	Primarily LD < LL	28
Underivatized	3% EGSP-Z[b] on Gas Chrom Q	Argon	FID	Primarily LL < LD	37
TMS derivatives	3% SE-30 on Gas Chrom Q	Argon	FID	LD < LL	37

[a] Shorter retention time listed first.
[b] EGSP-Z, Ethyleneglycol succinate phenylsilicone copolymer.

ethanol. In some cases alkylsulfonates as ion-pair reagents or metal ions were added. Ion-exchange chromatography was sometimes applied successfully. Dipeptides,[55] neurotensin,[56] or angiotensins[57] were resolved on weak anion exchange columns while cation exchange packings were used for gramicidin analogs[58] and hGRF.[59] Few reports on normal-phase HPLC for the resolution of diastereomeric N-derivatized di-, tri-, and tetrapeptide esters using mixtures of organic solvents as eluents have been published.[44,48]

With a few exceptions the LL epimers of dipeptides were eluted faster than the LD diastereomers on RP[45,47,60–66] or anion exchange columns,[55] regardless of whether the peptides were analyzed as free peptides or as N-derivatives. This behavior is attributed to the spatial arrangement of the side chains of the amino acid residues leading to different hydrophobicities of the diastereomers. In the LD or DL forms the side chains are on the same side of the peptide bond and in close proximity to one another. This results in a higher overall hydrophobic surface area and in increased retention compared to the LL and DD forms where the side chains are on opposite sides. The all-L epimer of tripeptides eluted faster in most cases than the mixed epimers.[42,43,54] As the size of the peptides increases, further factors contribute to the overall hydrophobicity so that the retention behavior of the peptide diastereomers becomes less predictable. The retention order of the diastereomers of N-derivatized dipeptide esters depended on the type of the derivative.[44] Table 6 summarizes separations of diastereomeric peptides ranging from dipeptides to peptides containing 44 amino acids.

The HPLC technique can also be applied to the separation of diketopiperazines derived from diastereomeric dipeptides.[62,75,92,93] Examples are summarized in Table 7.

Table 6 Separation of diastereomeric peptides by HPLC

Peptide	Stationary phase	Mobile phase composition (v/v)[a]	Detection	Sequence[b]	Refs.
Dipeptides	Nucleosil® C-18	CH_3CN–0.1 M KH_2PO_4, pH 1.6–4.5 (0 : 100–20 : 80)	210 nm	LL < LD	61
Asp dipeptides	Econosphere® C-18	CH_3CN–50 mM KH_2PO_4, pH 3.15, gradient	215 nm	LL < LD	62
Dipeptides	Hamilton® PRP-1	CH_3CN–0.1 M NaH_2PO_4, pH 5.25, gradient	208 nm	LL < LD	63
Z-Glu-D/L-Pro	Hypercarb®	10 mM Citric acid in MeOH	273 nm	LD < LL	46
Dipeptides	Spherisorb® ODS ODS Hypersil®	MeOH–10 mM NH_4OAc (10 : 90–30 : 70)	–	LL < LD	65
Dipeptides, Ala-Ala-Ala	Hamilton® PRP-1	0.05 mM $Ru(bipy)_3Cl_2$–0.075–0.1 mM LeuPhe in CH_3CN–water (1 : 99)	Indirect detection	LL < LD	66
Dipeptides, tripeptides	LiChrosorb® RP-8	CH_3CN–20 mM NaH_2PO_4, pH 6.4	208 nm	LL < LD	60
	µBondapak® C-18	CH_3CN–20 mM NaH_2PO_4, pH 2.6		LLD < LDD < LLL < LDL	
Dipeptides, tripeptides	LiChrosorb® C-8	EtOH–20 mM NaH_2PO_4, pH 3.4 (10 : 90)	208 nm	LL < LD	64
Dipeptides, tripeptides	µBondapak® C-18	0.01 M NH_4OAc, pH 6.6 0.01 M NH_4OAc, pH 4.4 0.01 M NaH_2PO_4, pH 3.2	215 nm	LL < LD	67
Arg-Lys-Asp	LiChrosorb® RP-18	15 mM NaH_2PO_4, pH 9.0	215 nm	LLL < LLD < DLL < LDL	54
Pyroglutamyl-His-dimethylprolinamide	Spherisorb® CN	CH_3CN–0.01 mM $Cu(OAc)_2$ (30 : 70)	210 nm	–	68
Pentapeptides	Spherisorb® ODS	EtOH–0.1 M NaH_2PO_4, pH 3, (1 : 9) cont. 0.015 M PSA[c]	210 nm	–	69
Val-Tyr-Pro-L/D-Asp-Val-Ala	Hypersil® ODS C-18	CH_3CN–0.1% TFA (12 : 88)	214 nm	–	70
Trp-D/L-Asn-Pro-Ala-Asp-Tyr-Gly-Gly-Ile-Lys	Spherisorb® ODS Supelcosil® LC-ABZ	CH_3CN–0.1% TFA, gradient	210 nm	D-Asn < L-Asn	71
Neuropeptide Y (1–17) analog	Nucleosil® C-18	CH_3CN–0.1% TFA, gradient	–	–	72
Neuropeptide Y	Vydac® C-18	CH_3CN–0.1% TFA (60 : 40)		–	73
	Vydac® diphenyl	CH_3CN–0.1% TEAP, pH 2.25 (60 : 40)			
Endorphin analogs	µBondapak® Phenyl	CH_3CN–0.01 M NH_4OAc, pH 4.5 (40 : 60)	254 nm	–	74
Met- and Leu-enkephalin	Spherisorb® ODS	0.1% NH_4OAc in CH_3CN–water (10 : 90) 0.1% NH_4OAc in CH_3CN–0.025% acetic acid (70 : 30)	215 nm	–	75
	Spherisorb® CN	0.01 M TEAP in CH_3CN–water (80 : 20) 0.035 M $Cu(OAc)_2$ in CH_3CN–water (70 : 30)	210 nm	–	
Cyclic enkephalin analogs	Vydac® C-18	CH_3CN–0.1% TFA in water (25 : 75)	280 nm	–	39
Thymopoietin fragments	Ultrasphere® C-18	MeOH–20 mM phosphate, pH 3.0 (30 : 70), 30 mM HSA[d]	–	–	76
Arg pentapeptide, vasopressin	Spherisorb® C-18	EtOH–0.1 M phosphate, pH 3.2 (78 : 22), 0.015 M HSA[d]	210 nm	–	77
Arginine vasopressin	Spherisorb® ODS	MeOH–0.025 M TEAA, pH 4.0 (25 : 75)	215 nm	–	78
	µBondapak® C-18	CH_3CN–0.025 M TEAA, pH 4.0 (12 : 88)			
	LiChrosorb® RP-18	THF–0.025 M TEAA, pH 4.0 (7 : 93)			
Vasopressin analogs	Nucleosil® C-18	MeOH–0.05 M NH_4OAc, pH 6.5 (39 : 61)	220 nm	–	79
Oxytocin	µBondapak® C-18	CH_3CN–0.1 M NH_4OAc, pH 4.0 (18 : 82)	254 nm	–	80, 81
Oxytocin	Separon® Si-C-18	MeOH–NH_4OAc, pH 7 (70 : 30) MeOH–0.05% TFA (60 : 40–40 : 60)	–	–	82

Table 6 (*continued*)

Peptide	Stationary phase	Mobile phase composition (v/v)[a]	Detection	Sequence[b]	Refs.
α-MSH analogs[e]	Vydac® C-18	CH_3CN–0.1% TFA, pH 2.0 (20 : 80) CH_3CN–0.25 M TEAP, pH 2.2 (14 : 86) CH_3CN–0.1% HFBA,[g] pH 2.0 (30 : 70)	214 nm	–	83
Cyclic α-MSH analogs[e]	Vydac® C-18	CH_3CN–0.1% TFA (21 : 79) CH_3CN–0.25 M TEAP, pH 2.2 (15 : 85) MeOH–0.25 M TEAP, pH 2.2 (37 : 63)	220 nm	–	84
Angiotensin I and II	ODS column	CH_3CN–0.1 M TEAP, pH 3.5 (19 : 81)	254 nm	–	85
Bombesin analogs	µBondapak® C-18	CH_3CN–0.01 M NH_4OAc, pH 4.5 (30 : 70)	–	–	25
ACTH (1–18)	Nucleosil® C-18	CH_3CN–0.005 M BSA[h] and 0.05 M Na_2SO_4 in 0.005 M sodium tartrate, pH 3.0 (18 : 82–22.5 : 77.5)	–	–	86
Neurotensin	µBondapak® C-18	CH_3CN–0.01 M NH_4OAc, pH 4.5 (25 : 75)	210 nm	–	24
Corticotropin releasing factor	Vydac® C-18	0.1% TFA in 5% CH_3CN–0.1% TFA in 80% CH_3CN, gradient	210 nm	–	87
Somatostatin	µBondapak® C-18	CH_3CN–0.01 M NH_4OAc, pH 4.1, gradient	220 nm	–	88
Somatostatin, LH-RH[f]	µBondapak® C-18	EtOH–0.01 M NH_4OAc, pH 4.0 (22 : 78–35 : 65)	210 nm	–	89
Phosphonoheptapeptides	C18 RP column	0.1% in CH_3CN–0.1% in water, gradient	–	–	90
Z-Gly-dipeptide-OMe	Cosmosil® C-18	MeOH–water (45 : 55–65 : 35)	254 nm	LL < LD	45
Z-tripeptide-OMe	LiChrospher® C-18	MeOH–water (60 : 40)	220 nm	LLL < LDL	42, 43
N-Derivatized di-, tri-, and tetrapeptides	µBondapak® C-18	CH_3CN–0.01 M NH_4OAc, pH 4.4 (3 : 97–6 : 94) CH_3CN–0.1% acetic acid (15 : 85–30 : 70) MeOH–0.1% acetic acid (35 : 65–50 : 50)	208 nm	LL < LD	47
Z-Gly-tripeptide-OMe	Cosmosil® C-18	MeOH–water (65 : 35)	271 nm	LL < LD	49
Z-tripeptide-OMe	Cosmosil® C-18	MeOH–water (65 : 35)	254 nm	LL < LD	52
Z-tripeptide-OMe, Z-tetrapeptide-OMe	Cosmosil® C-18	MeOH–water (65 : 35)	254 nm	LL < LD	50, 51
Dipeptides	MicroPak® AX-10	CH_3CN–10 mM TEAA, pH 4.3 (68 : 32)	200 nm	LL < LD	55
Neurotensin	MicroPak® AX-10	CH_3CN–10 mM TEAA, pH 6.0, gradient	220 nm	–	56
Angiotensin I	MicroPak® AX-10	CH_3CN–10 mM TEAA, pH 6.0, gradient	225 nm	–	57
Atosiban	Spherisorb® NH_2	CH_3CN–2.5 mM NH_4OAc, 0.25 M $NaClO_4$ (92.3 : 7.7)	210 nm	–	91
	YMC-Pack® Polyamine II	CH_3CN–5 mM NH_4OAc (91.8 : 88.2)			
	Polyhydroxyethyl aspartamide	10 mM TEAP in CH_3CN–2-propanol–MeOH (88.6 : 0.8 : 10.6)			
	Inertsil® ODS-2	aq. TFA, pH 3.3–MeOH–CH_3CN (75 : 10 : 15)			
Gramicidin analogs	Zorbax® 300XDB C8	0.05% TFA in CH_3CN–0.05% TFA in water, gradient	210 nm	–	58
	PolySulfoethyl A	20 mM TEAP, pH 3, in 90% CH_3CN–20 mM TEAP and 0.4 M $NaClO_4$, pH 3, in 80% CH_3CN	210 nm	–	

(*continued overleaf*)

Table 6 (*continued*)

Peptide	Stationary phase	Mobile phase composition (v/v)[a]	Detection	Sequence[b]	Refs.
hGRF	TSK® gel CM-2SW	A: CH_3CN–20 mM Na_2HPO_4, pH 6.0 (90:10) B: CH_3CN–0.5 M NaCl in 20 mM Na_2HPO_4, pH 7.0 (90:10), gradient A/B	280 nm	–	59
N-Derivatized dipeptide esters	silica gel	2-Propanol–*n*-hexane or 2-propanol–chloroform	254 nm	Depends on derivative	44
BOC- Met_3-OMe, BOC-Met_4-OMe	μPorasil®	Cyclohexane–2-propanol–MeOH (93:6:1)	220 nm	LLL < DLL LDLL < LLL	48
D/L-Leu-Tyr	Val-Ala-Pro bonded to silica gel	Water	254 nm	DL < LL	53

TFA, trifluoroacetic acid; TEAP, triethylammonium phosphate; hGRF, human growth hormone releasing factor.
[a] Optimized conditions.
[b] Shorter retention time listed first.
[c] PSA, 1-Pentanesulfonic acid sodium salt.
[d] HSA, 1-Hexanesulfonic acid sodium salt.
[e] MSH, Melanocyte-stimulating hormone.
[f] LH-RH, Luteinizing hormone releasing hormone.
[g] HFBA, Heptafluorobutyric acid.
[h] BSA, 1-Butanesulfonic acid sodium salt.

Table 7 Separation of diastereomeric diketopiperazines by HPLC

Derivatives	Stationary phase	Mobile phase (v/v)[a]	Detection	Sequence[b]	Ref.
c-Asp-Phe	Econosphere® C-18	CH_3CN–0.05 M KH_2PO_4, pH 3.15, gradient	215 nm	LD < LL	62
c-His-Pro	Partisil ODS	0.1% NH_4OAc in CH_3CN–water (10:90)	210 nm	LL < LD	75
Diketopiperazines	Zorbax® ODS	CH_3CN–water (10:90–50:50) MeOH–water (30:70–50:50)	210 nm	Depending on sequence and mobile phase	92
c-Thiodipeptides	Hypersil-silica	CH_2Cl_2–EtOAc (98:2)	270 nm	LL < LD	93

[a] Optimized conditions.
[b] Shorter retention time listed first.

2.5 Method Development

Method development will be only addressed briefly for HPLC as this is the most important chromatographic technique for peptide diastereomer separations. In the RP mode ODS (C-18) columns have been most frequently used. However, the packing material from various manufacturers may differ with respect to carbon loading, endcapping of residual silanol groups, etc. The separation of peptide diastereomers is particularly sensitive to differences in the composition of the stationary phase. Moreover, the individual should be aware of the mechanism that dictates the selectivity for an HPLC separation.

In most RP analyses low-pH aqueous solvents should be employed – TFA is probably most versatile but sodium phosphate, TEAP, ammonium acetate, or TEAA is also very suitable. The ionic strength of the buffer can also be crucial for a successful separation. Following optimization of the pH and the type and the concentration of the organic modifier, additives such as ion-pair reagents or metal ions can be tested.

In cation-exchange chromatography TFA/acetonitrile mixtures are most frequently used ensuring protonation of the carboxylic acid groups of the peptides. In anion-exchange HPLC solvent systems with higher pH values have been employed.

3 ELECTROPHORETIC SEPARATIONS

Electrophoretic methods comprise several techniques including paper electrophoresis, gel electrophoresis, and CE. Despite the fact that the resolution of the diastereomers of the dipeptides Lys-Glu and Lys-Asp by paper electrophoresis has been reported,[13] virtually all separations of diastereomeric peptides have been carried out by CE in recent years. Due to its extremely high resolution ability, CE has become a powerful tool for the analysis of polar compounds such as peptides.

The separation mechanism of electrophoretic techniques differs fundamentally from chromatographic

methods. In electrophoresis ionic analytes migrate with a constant velocity v that can be expressed as the product of the applied electric potential E and the electrophoretic mobility μ (Equation 1):

$$v = \mu E \quad (1)$$

with the mobility being dependent on the charge density (charge/mass ratio) of an ionic species. For spherical molecules μ can be expressed (Equation 2) as:

$$\mu = \frac{q}{6\pi r\eta} \quad (2)$$

where q is the charge of the analyte, r is the Stokes radius of the hydrated species and η is the viscosity of the separation medium. For peptides and proteins, which may not always be spherical, interpretation of the electrophoretic mobility in terms of molecular size and charge is more complex. However, as can be deduced from the theory and empirical data available, the electrophoretic mobility of a peptide or protein is proportional to the actual charge and inversely proportional to the size. The charge of weak electrolytes depends on the pH of the separation buffer according to the acid–base equilibrium of the electrolyte. The degree of ionization of weak acids and bases and, thus, the effective electrophoretic mobility, is therefore a function of the buffer pH and the dissociation constant (pK_a) of the compounds. Moreover, in CE the velocity of an analyte will also depend upon the magnitude of the unspecific electroosmotic flow (EOF) generated by the pH-dependent ionization of the silanol groups of the fused-silica capillary wall, resulting in a flow of the bulk solution towards the cathode.

Among the different CE techniques free solution capillary zone electrophoresis (CZE) and micellar electrokinetic chromatography (MEKC) have been applied to the separation of peptide diastereomers. In MEKC, surfactants above their critical micelle concentration are applied as pseudostationary phases. A separation in MEKC is the result of differences in the partitioning of solutes between the micelles and the aqueous buffer; MEKC is typically applied to the separation of hydrophobic compounds. Whereas in CZE a separation is based on

Table 8 Separation of diastereomeric peptides by CZE

Peptide	Run buffer composition[a]	Detection	Sequence[b]	Refs.
α/β-Asp-PheOMe, α/β-Asp-PheNH$_2$	50 mM NaH_2PO_4, pH 2.7	215 nm	α: LD < LL	102, 103
	50 mM Na_2HPO_4, pH 8.0	215 nm	β: LL < LD	97
Phe-α/β-Asp-GlyNH$_2$, Gly-α/β-Asp-PheNH$_2$	50 mM NaH_2PO_4, pH 3.0	215 nm	α: LD < LL	100
	17.5 mg mL^{-1} β-Cyclodextrin in 50 mM NaH_2PO_4, pH 3.0		β: LL < LD	
α/γ-GluPheOMe, α/γ-GluPheNH$_2$, Gly-α/γ-Glu-PheOMe	50 mM NaH_2PO_4, pH 2.5	215 nm	α: LD < LL	98
	50 mM Na_2HPO_4, pH 8.5	215 nm	γ: LD < LL	
Ala-Phe, Leu-Phe	70 mM NaH_2PO_4, 2 M urea, pH 3.0	215 nm	LL < LD	104
Ala-Phe-Gly	10 mM Tris citrate, pH 2.5	210 nm	DL < LL	105
N-(8-Guanidinooctanoyl)-α/β-Asp-Phe	40 mM NaH_2PO_4, pH 3.2	190 nm	α: LL < LD	106
	40 mM NaOAc, pH 3.5		β: LL < LD	
[L/D-Neopentylglycine]^5-dalargin, [NCH$_3$-L/D-Phe]^4-dalargin	150 mM Phosphoric acid, pH 2.0		–	101
	10 mM β-Cyclodextrin in 50 mM NaH_2PO_4, pH 2.5			
[L/D-Arg 6]-dynorphin (1–13), [D/L-Ala 2]-enkephalin, Tyr-Ala-Gly-Phe-D/L-Met	20 mM sodium citrate, pH 2.5	200 nm	–	107
Atosiban	5 mM γ-Cyclodextrin in 100 mM NaH_2PO_4, pH 2.5	195 nm	–	91
Neuropeptide Y (1–17) analogs	50 mM Formic acid and 10 mM NaCl, pH 2.5	–	–	72
Neuropeptide Y	100 mM Phosphoric acid, pH 2.5	214 nm	–	73
Human growth hormone fragment	10 mM β-Cyclodextrin and 50 mM NaCl in 25 mM aspartic acid, pH 3.8	214 nm	–	108
Cyclic analog of hGRF	125 mM NaH_2PO_4, pH 2.2	–	–	109
Yeast tridecapeptide pheromone analogs	100 mM NaH_2PO_4, pH 2.5	220 nm	–	110
Ala-Phe	25 mM NH_4Cl in MeOH–CH_3CN (1:1)	214 nm	LL < LD	99
	50 mM $AcONH_4$ + 1 M AcOH in MeOH–CH_3CN (1:1)			

[a] Optimized conditions.
[b] Shorter migration time listed first.

the mobility difference between analytes, MEKC is governed by a chromatographic separation principle and an electrophoretic migration principle. For a comprehensive treatment of the theory and the various CE separation techniques the reader is referred to the numerous monographs on this topic as well as to the relevant articles of this encyclopedia – **Capillary Electrophoresis** and **Micellar Electrokinetic Chromatography**.

With a few exceptions where resolutions of diastereomeric peptides were performed using chiral additives, the published separations were primarily achieved in achiral environments. It is usually possible to separate diastereomers using chiral additives as well, but this is outside the scope of the current article. The use of chiral selectors also allows the simultaneous separation of enantiomers and diastereomers of small peptides, which has been mentioned in several recent reviews.[94,95] A comprehensive treatment of chiral CE has been published by Chankvetadze.[96]

3.1 Capillary Zone Electrophoresis

Diastereomeric peptides have been preferentially separated by CZE at acidic buffer pH (Table 8), i.e. at pH values in the region of the pK_a values of the carboxyl groups of the C-terminus or in side chains of Asp or Glu. Small differences in the pK_a values of the carboxylic acid groups of the diastereomers result in different effective mobilities of the peptides. However, it is also possible to resolve diastereomers at basic pH values in the region of the pK_a of the amino groups, as demonstrated for Asp and Glu di- and tripeptides.[97,98] One reason for the preference of acidic buffers is the fact that the unspecific EOF is negligible at low pH. The large EOF at higher pH values (>4) might interfere with the separation. The EOF can be suppressed by dynamic or permanent coating of the inner capillary wall, but this approach has not been exploited for the resolution of peptide diastereomers at present. Peptides containing up to 36 amino acids have been separated by CZE. Moreover, the first resolution of peptide diastereomers by nonaqueous CZE has been reported.[99]

Chiral selectors such as cyclodextrins have been added in a few cases.[100,101] The observed separations are due to the formation of transient diastereomeric complexes between the selectors and the peptide analytes. This results in different effective charge densities caused by different spatial orientations of the diastereomers, or by specific interactions between the selectors and the host molecules.

In contrast to the chromatographic separation techniques no rule concerning the migration sequence of the diastereomers of di- and tripeptides can be deducted from the published data (Table 8). It is interesting to note that replacement of Asp by Glu, i.e. insertion of a methylene moiety, led to opposite migration behavior. Although the LL diastereomer of the α isomers of Asp-PheOMe and Asp-PheNH_2 migrated faster than the respective LD epimers,[102,103] the LD diastereomers of α-GluPheOMe and α-GluPheNH_2 migrated faster than the LL epimers.[98]

3.2 Micellar Electrokinetic Chromatography

As peptides are usually polar compounds, most analyses including stereoisomer separations have been performed in the CZE mode. However, the resolution of the diastereomers of several hydrophobic peptides and N-derivatized peptides have been reported using neutral as well as ionic detergents as pseudostationary phase.

Anionic surfactants are typically used in MEKC, with the analyte distribution between the phases governed mainly by hydrophobic interactions with the sodium dodecylsulfate (SDS) alkyl side chains. Small peptides interact only to a limited extent with the micelles, often resulting in a separation. Larger peptides and proteins can interact much more strongly so that a separation is not achieved.[111] Water-soluble organic solvents such as acetonitrile and methanol have been successfully employed to decrease these interactions by reducing the polar nature of the separation buffer. An alternative approach is the use of polar nonionic surfactants such as Brij® 35 or Tween® 20 as micelle-forming agents having weaker interactions with peptides.[112] Although ionic detergents such as SDS or sodium cholate possess an electrophoretic self-mobility, neutral detergents such as Tween® 20, Tween® 40 and Brij® 35 migrate only with the EOF.

Typically, high-pH buffers are used in MEKC separations of peptides (Table 9).[106,113–115] At these pH values the peptides are negatively charged and possess a self-mobility towards the anode as do the negatively charged surfactants such as SDS. Thus, high pH values are required to generate a sufficiently high EOF so that the peptides can still be detected at the cathodic end of the capillary. When neutral surfactants are employed peptides can also be separated at low pH when they are positively charged and migrate towards the cathode.[106,116] Compared to the system using SDS and high-pH buffers, this results in a reversal of the diastereomer migration order.[106] It is interesting to note that a reversal of the migration order was also observed for the peptide drug lisinopril by switching the surfactant from sodium cholate to SDS.[114] This may be attributed to the chiral recognition ability of the chiral surfactant cholated towards the lisinopril diastereomers as compared to the achiral SDS. The use of a cationic detergent

Table 9 Separation of diastereomeric peptides by MEKC

Peptide	Run buffer composition[a]	Detection	Sequence[b]	Ref.
CBI-dipeptides[c]	50 mM SDS in 100 mM sodium borate, pH 9.0	LIF[d]	LD < LL	113
(−)-FLEC-dipeptides[e]	15 mM SDS in 40 mM sodium borate, pH 9.2	254 nm	–	115
N-(8-Guanidinooctanoyl)-α/β-Asp-Phe (SC-49992)	30 mM SDS in 40 mM NaH_2PO_4, pH 6.0	190 nm	LD < LL	106
	1 mM Tween® 20 in 40 mM NaH_2PO_4, pH 3.0		LL < LD	
	1 mM Tween® 40 in 40 mM NaH_2PO_4, pH 3.0		LL < LD	
Lisinopril (*SSS*/*RSS*)	50 mM Sodium cholate in 25 mM Tris phosphate, pH 9.55–methanol (45 : 55)	210 nm	(*RSS*) < (*SSS*)	114
	40 mM SDS in 25 mM Na_2HPO_4, pH 9.0		(*SSS*) < (*RSS*)	
SB-238592-DB (bis-decapeptide)	15 mM Brij® 35 in 200 mM lithium phosphate, pH 2.5–acetonitrile (95 : 5)	200 nm	–	116
Neuropeptide Y (1–17) analogs	50 mM CTAC[f] in 50 mM Na_2HPO_4, pH 7.0	–	–	72

[a] Optimized conditions.
[b] Shorter migration time listed first.
[c] CBI, Cyano[*f*]isoindole.
[d] LIF, Laser-induced fluorescence.
[e] (−)-FLEC, (−)-1-(9-Fluorenyl)-ethylchloroformiate.
[f] CTAC, Hexadecyltrimethylammonium chloride.

hexadexyltrimethylammonium chloride in pH 7.0 buffer has also been described.[72]

3.3 Method Development

The electrophoretic mobility of peptides is proportional to the charge and inversely proportional to the size. As peptides are weak electrolytes their degree of ionization and, thus, the effective electrophoretic mobility are therefore a function of the buffer pH and the pK_a of ionizable groups. The factor that most effects electrophoretic mobility is charge, hence pH is the most significant variable for peptide CE separations.

Most diastereomer separations in the CZE mode have been carried out between pH 2 and pH 4, i.e. in the range of the pK_a values of carboxyl groups located either at the C terminus or in the side chain of Asp or Glu. Method development for CE peptide diastereomer separations starts by selecting the appropriate pH, the most important parameter for optimizing the selectivity. Phosphate buffers have been primarily used, but acetate or other buffers can be applied to match ion mobility. The second step is the optimization of the type of buffer ion, the ionic strength of the buffer, the separation voltage, and the length of the capillary. Further optimization may include the use of buffer additives such as organic solvents, metal ions, ion-pair reagents, surfactants, chiral selectors, etc. in order to maximize differences between the diastereomers or to mask interactions between the peptides and the capillary wall. As CE is a relatively young technique not all factors influencing a separation are fully understood. Thus, trial-and-error still governs method development in CE, including the separation of peptide diastereomers.

ABBREVIATIONS AND ACRONYMS

AWP	Ammonium Tungstophosphate
CE	Capillary Electrophoresis
CZE	Capillary Zone Electrophoresis
DEGS	Diethyleneglycol Succinate
EGSP-Z	Ethyleneglycol Succinate Phenylsilicone Copolymer
EOF	Electroosmotic Flow
FFAP	Free Fatty Acid Phase
FID	Flame Ionization Detection
GC	Gas Chromatography
HDBS	Dodecylbenzene Sulfuric Acid
HFBA	Heptafluorobutyric Acid
hGRF	Human Growth Hormone Releasing Factor
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
HSA	1-Hexanesulfonic Acid Sodium Salt
IR	Infrared
LH-RH	Luteinizing Hormone Releasing Hormone
MEKC	Micellar Electrokinetic Chromatography
MS	Mass Spectrometry
MSH	Melanocyte-stimulating Hormone
N-DPC	*N*-Dodecylpyridinium Chloride
NMR	Nuclear Magnetic Resonance
N-TFA	*N*-Trifluoroacetyl
N-TFP	*N*-Tetrafluoropropionyl
ODS	Octadecyl Silica
ONP	*o*-Nitrophenyl Ester
PC	Paper Chromatography
RP	Reversed-phase
RPHPLC	Reversed-phase High-performance Liquid Chromatography

SDS	Sodium Dodecylsulfate
SFC	Supercritical Fluid Chromatography
TEAA	Triethylammonium Acetate
TEAP	Triethylammonium Phosphate
TFA	Trifluoroacetic Acid
TLC	Thin-layer Chromatography
TMS	Trimethylsilyl

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Electron Spin Resonance Spectroscopy Labeling in Peptide and Protein Analysis • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis • Proteins and Peptides Purification in Pharmaceuticals Analysis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography • Micellar Electrokinetic Chromatography

REFERENCES

1. E.L. Eliel, S.H. Wilen, *Stereochemistry of Organic Compounds*, John Wiley & Sons, New York, 1994.
2. J.S. Splitter, F. Turecek (eds.), *Applications of Mass Spectrometry to Organic Stereochemistry*, VCH, New York, 1994.
3. A. Peter, G. Toth, 'Chromatographic Methods for the Separation of Enantiomers and Epimers of β-Alkyl Amino Acids and Peptides Containing them', *Anal. Chim. Acta*, **352**, 335–356 (1997).
4. S. Allenmark, *Chromatographic Enantioseparations: Methods and Applications*, John Wiley & Sons, New York, 1988.
5. W.A. König, *Gas Chromatographic Enantiomer Separations with Modified Cyclodextrins*, John Wiley & Sons, Chichester, 1989.
6. G. Subramanian (ed.), *A Practical Approach to Chiral Separations by Liquid Chromatography*, VCH, New York, 1994.
7. J.W. Hinman, E.L. Caron, H.N. Christensen, 'The Isomeric Dipeptides of Valine Including a Correction', *J. Am. Chem. Soc.*, **72**, 1620–1626 (1950).
8. H. Sachs, E. Brand, 'Optical Rotation of Peptides VIII. Glutamic Acid Tripeptides', *J. Am. Chem. Soc.*, **76**, 1811–1814 (1954).
9. E. Taschner, T. Sokolowska, J.F. Biernat, A. Chimiak, C. Wasielewski, B. Rzeszoarska, 'Nachweis der Racemisierung bei Peptidsynthesen durch Papierchromatographie', *Liebigs Ann. Chem.*, **663**, 197–202 (1963).
10. H. Kihara, M. Ikawa, E.E. Snell, 'Peptides and Bacterial Growth. X. Relation of Uptake and Hydrolysis to Utilization of D-Alanine Peptides for Growth of *Streptomyces faecalis*', *J. Biol. Chem.*, **236**, 172–176 (1961).
11. B. Riniker, R. Schwyzer, 'Die sterische Einheitlichkeit des synthetischen Val^5-Hypertensin II-Asp^1-β-amids', *Helv. Chim. Acta*, **44**, 658–667 (1961).
12. T. Sokolowska, J.F. Biernat, 'The Chromatographic Separation of Diastereomeric Dipeptides', *J. Chromatogr.*, **13**, 269–270 (1964).
13. M. Muraoka, N. Yoshida, K. Noda, N. Izumiya, 'Studies on Separation of Amino Acids and Related Compounds. IV. Separation of Diastereomeric Tripeptides Containing Basic and Acidic Amino Acid Residue', *Bull. Chem. Soc. Jpn.*, **41**, 2134–2140 (1968).
14. K. Noda, H. Okai, T. Kato, N. Izumiya, 'Studies on Separation of Amino Acids and Related Compounds. III. Separation of Diastereomers of Leucyl Dipeptides by Ion-exchange Chromatography', *Bull. Chem. Soc. Jpn.*, **41**, 401–407 (1968).
15. T. Wieland, H. Bende, 'Chromatographische Trennung einiger diastereomerer Dipeptide und Betrachtungen zur Konformation', *Chem. Ber.*, **98**, 504–515 (1965).
16. L. Lepri, P.G. Desideri, D. Heimler, S. Giannessi, 'High-performance Thin-layer Chromatography of Diastereomeric Di- and Tripeptides on Ready for Use Plates of Silanized Silica Gel and on Ammonium Tungstophosphate Layers', *J. Chromatogr.*, **265**, 328–334 (1983).
17. A. Arendt, A. Kolodziejczyk, T. Sokotowska, 'Separation of Diastereomers of Protected Dipeptides by Thin-layer Chromatography', *Chromatographia*, **9**, 123–126 (1976).
18. P. Hubert, E. Dellacherie, 'Thin-layer Chromatographic Separation of Dipeptide and Tripeptide Diastereoisomers', *J. Chromatogr.*, **80**, 144–145 (1973).
19. Z. Pravda, K. Poduska, K. Blaha, 'Amino Acids and Peptides. XLIII. Preparation of Derivatives of D-Phenylalanyl-L-Leucine and L-Phenylalanyl-L-Leucine from Diastereomeric Mixtures', *Coll. Czech. Chem. Commun.*, **29**, 2626–2631 (1964).

20. E. Taschner, B. Rzeszotarska, L. Lubiewska, 'Darstellung aktiver Pyridyl-(3)-ester N-geschützer Aminosäuren und Peptide sowie deren Anwendung zur Peptid-Synthese', *Liebigs Ann. Chem.*, **690**, 177–181 (1965).
21. K. Günther, J. Martens, M. Schickedanz, 'Dünnschichtchromatographische Trennung stereoisomerer Dipeptide', *Angew. Chem., Int. Ed. Engl.*, **25**, 278–279 (1986).
22. M. Mack, H.E. Hauck, 'Separation of Enantiomers in Thin-layer Chromatography', *Chromatographia*, **26**, 197–205 (1988).
23. J. Rivier, M. Brown, W. Vale, 'D-Trp^8-Somatostanin: an Analog of Somatostatin more Potent than the Native Molecule', *Biochem. Biophys. Res. Commun.*, **65**, 746–751 (1975).
24. J.E. Rivier, L.H. Lazarus, M.H. Perrin, M.R. Brown, 'Neurotensin Analogs. Structure–Activity Relationships', *J. Med. Chem.*, **20**, 1409–1412 (1977).
25. J.E. Rivier, M.R. Brown, 'Bombesin, Bombesin Analogs, and Related Peptides: Effect on Thermoregulation', *Biochemistry*, **17**, 1766–1771 (1978).
26. K.D. Kopple, H.G. Ghazarian, 'A Convenient Synthesis of 2,5-Piperazinediones', *J. Org. Chem.*, **35**, 682–683 (1968).
27. D.E. Nitecki, B. Halpern, J.W. Westley, 'A Simple Route to Sterically Pure Diketopiperazines', *J. Org. Chem.*, **33**, 864–866 (1968).
28. J.W. Westley, V.A. Close, D.N. Nitecki, B. Halpern, 'Determination of Steric Purity and Configuration of Diketopiperazines by Gas Chromatography, Thin-layer Chromatography, and Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **40**, 1888–1890 (1968).
29. F. Weygand, B. Kolb, A. Prox, M.A. Tilak, I. Tomida, 'Gaschromatographische Trennung von *N*-TFA-Dipeptid-Methylestern', *Hoppe-Seylers Z. Physiol. Chem.*, **322**, 38–51 (1960).
30. F. Weygand, A. Proux, L. Schmidhammer, W. König, 'Gaschromatographische Untersuchung der Racemisierung bei Peptidsynthesen', *Angew. Chem.*, **75**, 282–287 (1963).
31. M. Dizdaroglu, M.G. Simic, 'Separation of Dipeptide Diastereomers by High-resolution Gas Chromatography', *J. Chromatogr.*, **244**, 293–298 (1982).
32. I. Tomida, M. Matsuzaki, 'Optical Purity of Tfa and Tfp-L-Prolylating Reagents. The Anhydrides as the Recommended Acylating Reagents for the Separation of Diastereomers of Tripeptide by GLC', *Agric. Biol. Chem.*, **43**, 925–928 (1979).
33. I. Tomida, N. Nishimura, 'A Convenient Test for Racemization in Peptide Synthesis by Separation of Diastereomers of Tetrapeptide by GLC on Packed Column', *Agric. Biol. Chem.*, **44**, 1241–1244 (1980).
34. B. Halpern, J.W. Westley, 'High Sensitivity Optical Resolution of D,L Amino Acids by Gas Chromatography', *Biochem. Biophys. Res. Commun.*, **19**, 361–363 (1965).
35. B. Halpern, J.W. Westley, 'High Sensitivity Optical Resolution of D,L-Amino Acids by Gas Chromatography', *Chem. Commun.*, 246–247 (1965).
36. B. Halpern, J.W. Westley, 'High Sensitivity Optical Resolution of Polyfunctional Amino Acids by Gas Liquid Chromatography', *Tetrahedron Lett.*, **21**, 2282–2286 (1966).
37. A.B. Mauger, 'Gas Liquid Chromatography of Diketopiperazines', *J. Chromatogr.*, **37**, 315–317 (1968).
38. D.D. Blevins, M.F. Burke, V.J. Hruby, 'Diastereomer Separations', in *CRC Handbook of HPLC Separations of Amino Acids, Peptides and Proteins*, ed. W.S. Hancock, CRC Press, Boca Raton, 137–143, Vol. 2, 1984.
39. V.J. Hruby, A. Kawasaki, G. Toth, 'Reversed-phase High-performance Liquid Chromatography for the Separation of Peptide Diastereomers', in *High-performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis, and Conformation*, eds. C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 379–387, 1991.
40. W.S. Hancock (ed.), *CRC Handbook of HPLC Separations of Amino Acids, Peptides and Proteins*, CRC Press, Boca Raton, Vols. 1 and 2, 1984.
41. C.T. Mant, R.S. Hodges (eds.), *High-performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis, and Conformation*, CRC Press, Boca Raton, 1991.
42. C. Griehl, J. Weigt, H. Jeschkeit, 'Use of Tryptophan Peptides for an HPLC Racemization Test with Fluorescence Detection', *J. High Res. Chromatogr.*, **17**, 700–704 (1994).
43. C. Griehl, S. Merkel, 'Synthesis and Separation of Protected Tripeptide Epimers by RP-HPLC', *Int. J. Pept. Protein Res.*, **45**, 217–224 (1995).
44. L.-C. Lo, S.-T. Chen, S.-H. Wu, K.-T. Wang, 'Separation of Diastereomers of Protected Dipeptides by Normal-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **472**, 336–339 (1989).
45. T. Miyazawa, T. Otomatsu, T. Yamada, T. Kuwata, 'Racemization Studies in Peptide Synthesis Through the Separation of Protected Epimeric Peptides by Reversed-phase High-performance Liquid Chromatography', *Int. J. Peptide Protein Res.*, **39**, 229–236 (1992).
46. A. Karlsson, O. Karlsson, 'Enantiomeric Separation of Amino Alcohols Using Z-L-Glu-L-Pro or Z-L-Glu-D-Pro as Chiral Counter Ions and Hypercarb® as the Solid Phase', *Chirality*, **9**, 650–655 (1997).
47. N.L. Benoiton, Y. Lee, B. Liberek, R. Steinauer, F.M.F. Chen, 'High-performance Liquid Chromatography of Epimeric N-Protected Peptide Acids and Esters for Assessing Racemization', *Int. J. Pept. Protein Res.*, **31**, 581–586 (1988).
48. F. Naider, A.S. Steinfeld, J.M. Becker, 'Separation of Oligopeptides by Normal Phase Chromatography', in *CRC Handbook of HPLC Separations of Amino Acids, Peptides and Proteins*, ed. W.S. Hancock, CRC Press, Boca Raton, 89–97, Vol. 2, 1984.

49. T. Yamada, M. Shimamura, T. Miyazawa, S. Kuwata, 'Gly-L-Phe(pNO_2)-OMe, a Highly Sensitive Derivatization Reagent for Chiral Separation of N-(Benzyloxycarbonyl)-amino Acids by Reversed-phase HPLC', *Chem. Express*, **8**, 793–796 (1993).
50. T. Yamada, S. Nonomura, H. Fujiwara, T. Miyazawa, S. Kuwata, 'Separation of Peptide Diastereomers by Reversed-phase High-performance Liquid Chromatography and its Applications. IV. New Derivatization Reagent for the Enantiomeric Analysis of α- and β-Amino Acids', *J. Chromatogr.*, **515**, 471–482 (1990).
51. T. Yamada, M. Nakao, K. Tsuda, S. Nonomura, T. Miyazawa, S. Kuwata, M. Sugiura, 'HPLC Separation and Conformation of Peptide Diastereomers', *Pept. Chem.*, 97–100 (1988).
52. T. Yamada, M. Nakao, S. Kuwata, 'Separation of Diastereomers of Protected Achiral Amino Acid Middled Tripeptides. VII. On Peptides Containing 1-Aminocyclopropanecarboxylic Acid', *Pept. Chem.*, 333–338 (1985).
53. E. Grushka, E.J. Kikta, Jr, 'Chemically Bonded Stationary Phases in Chromatography', *Anal. Chem.*, **49**, 1004A–1014A (1977).
54. B. Herenyi, S. Görög, O. Nyeki, I. Schön, L. Kisfalu, 'The Effect of pH on the HPLC Separation of some Diastereomeric Tripeptides', *Chromatographia*, **29**, 395–396 (1990).
55. M. Dizdaroglu, M.G. Simic, 'Separation of Underivatized Dipeptides by High-Performance Liquid Chromatography on a Weak Anion-exchange Bonded Phase', *J. Chromatogr.*, **195**, 119–126 (1980).
56. M. Dizdaroglu, M.C. Simic, F. Rious, S.St.- Pierre, 'Separation of Diastereomers and Analogs of Neurotensin by Anion-exchange High-performance Liquid Chromatography', *J. Chromatogr.*, **245**, 158–162 (1982).
57. S.A. Margolis, M. Dizdaroglu, 'Separation and Purification of Diastereomers of Angiotensin I by Weak Anion-exchange High-performance Liquid Chromatography', *J. Chromatogr.*, **322**, 117–128 (1985).
58. C.T. Mant, L.H. Kondejewski, R.S. Hodges, 'Hydrophilic Interaction/Cation-exchange Chromatography for Separation of Cyclic Peptides', *J. Chromatogr. A*, **816**, 79–88 (1998).
59. K. Kumagaye, M. Takai, N. Chino, T. Kimura, S. Sakakibara, 'Comparison of Reversed Phase and Cation Exchange High-performance Liquid Chromatography for Separation of Closely Related Peptides: Separation of Asp^{76}-Human Parathyroid Hormone (1-84) from Asn^{76}-Human Parathyroid Hormone (1-84)', *J. Chromatogr.*, **327**, 327–332 (1985).
60. W.R. Cahill, Jr, E.P. Kroeff, D.J. Pietrzyk, 'Application of tert-Butyloxycarbonyl-L-Amino Acid N-Hydroxysuccinimide Esters in the Chromatographic Separation and Determination of D,L-Amino Acids and Diastereomeric Dipeptides', *J. Liq. Chromatogr.*, **3**, 1319–1334 (1980).
61. T. Takaya, Y. Kishida, S. Sakakibara, 'Determination of the Optical Purity of Amino Acids by High-performance Liquid Chromatography. Modification of the Manning Moore Procedure', *J. Chromatogr.*, **215**, 279–287 (1981).
62. S.M. Gaines, J.L. Bada, 'Reversed-phase High-performance Liquid Chromatographic Separation of Aspartame Diastereomeric Decomposition Products', *J. Chromatogr.*, **389**, 219–225 (1987).
63. Z. Iskandarani, D.J. Pietrzyk, 'Liquid Chromatographic Separation of Amino Acids, Peptides, and Derivatives of a Porous Polystyrene–divinylbenzene Copolymer', *Anal. Chem.*, **53**, 489–495 (1981).
64. E.P. Kroeff, D.J. Pietrzyk, 'High-performance Liquid Chromatographic Study of the Retention and Separation of Short Chain Peptide Diastereomers on a C8 Bonded Phase', *Anal. Chem.*, **50**, 1353–1358 (1978).
65. E. Lundanes, T. Greibokk, 'Reversed-phase Chromatography of Peptides', *J. Chromatogr.*, **149**, 241–254 (1978).
66. D. Yuan, D.J. Pietrzyk, 'Separation and Indirect Detection of Small-chain Peptides Using Chromatographic Mobile Phase Additives', *J. Chromatogr.*, **509**, 357–368 (1990).
67. R. Steinauer, F.M.F. Chen, N.L. Benoiton, 'Determination of Epimeric Peptides for Assessing Enantiomeric Purity of Starting Materials and Studying Racemization in Peptide Synthesis Using High-performance Liquid Chromatography', *J. Chromatogr.*, **325**, 111–126 (1985).
68. K. Sudgen, C. Hunter, J.G. Lloyd-Jones, 'Separation of the Diastereomers of Pyroglutamyl-Histidyl-3,3-Dimethylprolinamide by Ligand-exchange Chromatography', *J. Chromatogr.*, **204**, 195–200 (1981).
69. B. Nyasse, B. Fransson, U. Ragnarsson, 'Separation of a Complete Set of Basic, Diastereomeric Pentapeptides by Reversed-phase Ion-pair Chromatography', *Int. J. Pept. Protein Res.*, **36**, 316–319 (1990).
70. C. Oliyai, R.T. Borchardt, 'Chemical Pathways in Peptide Degradation. VI. Effect of the Primary Sequence on the Pathways of Degradation of Aspartyl Residues in Model Hexapeptides', *Pharm. Res.*, **11**, 751–758 (1994).
71. D. Korakas, K. Valko, C. Sakarellos, M. Sakarellos-Daisiotis, W.A. Gibbons, 'Analytical and Semi-preparative Separation of Diastereomeric Lipidic Amino Acid Conjugates', *J. Pharm. Biomed. Anal.*, **12**, 1141–1145 (1994).
72. H.-J. Gaus, A.G. Beck-Sickinger, E. Bayer, 'Optimization of Capillary Electrophoresis of Mixtures of Basic Peptides and Comparison with HPLC', *Anal. Chem.*, **65**, 1399–1405 (1993).
73. D.A. Kirby, C.L. Miller, J.E. Rivier, 'Separation of Neuropeptide Y Diastereomers by High-performance Liquid Chromatography and Capillary Zone Electrophoresis', *J. Chromatogr.*, **648**, 257–265 (1993).
74. B.L. Currie, J.K. Chang, R. Cooley, 'High Performance Liquid Chromatography of Enkephalin and Endorphin Peptide Analogs', *J. Liq. Chromatogr.*, **3**, 513–527 (1980).

75. C. Hunter, K. Sugden, J.G. Lloyd-Jones, 'HPLC of Peptides and Peptide Diastereomers on ODS- and Cyanopropyl–Silica Gel Column Packing Materials', *J. Liq. Chromatogr.*, **3**, 1352–1355 (1980).
76. M. Gazdag, K. Mihalyfi, S. Görög, 'Ion-pair RP-HPLC Separation of Thymopoietin Fragments and Related Peptides', *Anal. Chim. Acta*, **352**, 231–237 (1997).
77. U. Ragnarsson, B. Fransson, Ö. Zetterqvist, 'Separation of Basic Hydrophilic Peptides by Reversed-phase HPLC', in *CRC Handbook of HPLC Separations of Amino Acids, Peptides and Proteins*, ed. W.S. Hancock, CRC Press, Boca Raton, 75–88, Vol. 2, 1984.
78. D.D. Blevins, M.F. Burke, V.J. Hruby, B.R. Larsen, 'Factors Affecting the Separation of Arginine Vasopressin Peptide Diastereomers by HPLC', *J. Liq. Chromatogr.*, **3**, 1299–1318 (1980).
79. G. Lindeberg, 'Separation of Vasopressin Analogs by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **193**, 427–431 (1980).
80. B. Larsen, V. Viswanatha, S.Y. Chang, V.J. Hruby, 'Reversed Phase High Pressure Liquid Chromatography for the Separation of Peptide Hormone Diastereomers', *J. Chromatogr. Sci.*, **16**, 207–210 (1978).
81. B. Larsen, B.L. Fox, M.F. Burke, V.J. Hruby, 'The Separation of Peptide Hormone Diastereomers by Reversed Phase High Pressure Liquid Chromatography. Factors Affecting Separation of Oxytocin and its Diastereomers – Structural Implications', *Int. J. Pept. Protein Res.*, **13**, 12–18 (1979).
82. M. Lebl, 'Amino Acids and peptides CLXXXI. Separation of Diastereomers of Oxytocin Analogs', *J. Chromatogr.*, **264**, 459–462 (1983).
83. W.L. Cody, B.C. Wilkes, V.J. Hruby, 'Reversed-phase High-performance Liquid Chromatography Studies of α-MSH Fragments', *J. Chromatogr.*, **314**, 313–321 (1984).
84. M. Lebl, W.L. Cody, V.J. Hruby, 'Cyclic Melanotropins. Part VI. Reversed-phase HPLC Studies, *J. Liq. Chromatogr.*, **7**, 1195–1210 (1984).
85. S.A. Margolis, P.L. Konash, 'The High-performance Liquid Chromatographic Analysis of Diastereomers and Structural Analogs of Angiotensins I and II', *Anal. Biochem.*, **134**, 163–169 (1983).
86. S. Terabe, R. Konaka, K. Inouye, 'Separation of some Polypeptide Hormones by High-performance Liquid Chromatography', *J. Chromatogr.*, **172**, 163–177 (1979).
87. J. Rivier, C. Rivier, R. Galyean, A. Miranda, C. Miller, A.G. Craig, G. Yamamoto, M. Brown, W. Vale, 'Single Point D-Substituted Corticotropin-releasing Factor Analogs: Effect on Potency and Physicochemical Characteristics', *J. Med. Chem.*, **36**, 2851–2859 (1993).
88. C.A. Meyers, D.H. Coy, W.Y. Huang, A.V. Schally, T.W. Redding, 'Highly Active Position Right Analogs of Somatostatin and Separation of Peptide Diastereomers by Partition Chromatography', *Biochemistry*, **17**, 2326–2330 (1978).
89. R. Burgus, J. Rivier, 'Use of High Pressure Liquid Chromatography in the Purification of Peptides', in *Peptides*, ed. C.A. Loffet, Editions de L'Universite de Bruxelles, Brussels, 85, 1976.
90. C. Yang, R. Qamar, S.J. Norton, P.F. Cook, D.E. Minter, 'Synthesis and Separation of a Diastereomeric Pair of Phosphonopeptide Inhibitors of a Cyclic AMP-dependent Protein Kinase Catalytic Subunit', *Tetrahedron*, **50**, 1919–1926 (1994).
91. A.R. Oyler, B.L. Armstrong, J.Y. Cha, M.X. Zhou, Q. Yang, R.I. Robinson, R. Dunphy, D.J. Burinsky, 'Hydropholic Interaction Chromatography of Aminosilica Phases Complements Reversed-phase High-performance Liquid Chromatography and Capillary Electrophoresis for Peptide Analysis', *J. Chromatogr. A*, **724**, 378–383 (1996).
92. N. Funasaki, S. Hada, S. Neya, 'Conformational Effects in Reversed-phase Liquid Chromatographic Separation of Diastereomeric Cyclic Dipeptides', *Anal. Chem.*, **65**, 1861–1867 (1993).
93. G. Szokan, Z. Majer, E. Kollat, M. Kajtar, M. Hollosi, M. Peredy-Kajtar, 'Chromatographic Studies on the Racemization of Thiopeptides', *J. Liq. Chromatogr.*, **18**, 941–964 (1995).
94. K. Verleysen, P. Sandra, 'Separation of Chiral Compounds by Capillary Electrophoresis', *Electrophoresis*, **19**, 2798–2833 (1998).
95. G. Gübitz, M.G. Schmid, 'Chiral Separation Principles in Capillary Electrophoresis', *J. Chromatogr. A*, **792**, 179–225 (1997).
96. B. Chankvetadze, *Capillary Electrophoresis in Chiral Analysis*, John Wiley & Sons, Chichester, 1997.
97. S. Sabah, *Stereoisomer Separation of Aspartyl and iso-Aspartyl Peptides by Capillary Electrophoresis*, Ph.D. Thesis, University of Münster, 1998.
98. S. Schücker, *Separation of Glutamyl Di- and Tripeptides by Capillary Electrophoresis*, Ph.D. Thesis, University of Münster, 1999.
99. S.H. Hansen, I. Bjornsdottir, J. Tjornelund, 'Separation of Cationic *cis-trans* (Z-E) Isomers and Diastereomers Using Non-aqueous Capillary Electrophoresis', *J. Chromatogr. A*, **792**, 49–55 (1997).
100. S. Sabah, G.K.E. Scriba, 'Electrophoretic Stereoisomer Separation of Aspartyl Dipeptides and Tripeptides in Untreated Fused-silica and Polyacrylamide-coated Capillaries Using Charged Cyclodextrins', *J. Chromatogr. A*, **822**, 137–145 (1998).
101. V. Pacakova, J. Suchankova, K. Stulik, 'Separation of Biologically Active Peptides by Capillary Electrophoresis and High-performance Liquid Chromatography', *J. Chromatogr. B*, **681**, 69–76 (1996).
102. S. Sabah, G.K.E. Scriba, 'Determination of Aspartame and its Degradation and Epimerization Products by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **16**, 1089–1096 (1998).

103. S. Sabah, G.K.E. Scriba, 'Resolution of Aspartyl Dipeptide and Tripeptide Stereoisomers by Capillary Electrophoresis', *J. Microcolumn Sep.*, **10**, 255–258 (1998).
104. J. Li, K.C. Waldron, 'Estimation of the pH-Independent Binding Constants of Alanylphenylalanine and Leucylphenylalanine Stereoisomers with β-Cyclodextrin in the Presence of Urea', *Electrophoresis*, **20**, 171–179 (1999).
105. R. Kuhn, D. Riester, B. Fleckenstein, K.-H. Wiesmüller, 'Evaluation of an Optically Active Crown Ether for the Chiral Separation of Di- and Tripeptides', *J. Chromatogr. A*, **716**, 371–379 (1995).
106. D.J. Skanchy, R. Wilson, T. Poh, G.-H. Xie, C.W. Demarest, J.F. Stobaugh, 'Resolution of Acylated Dipeptide Stereoisomers by Capillary Electrophoresis Using Sulfobutylether Derivatized Cyclodextrin', *Electrophoresis*, **18**, 985–995 (1997).
107. J.C. Colburn, P.D. Grossman, S.E. Moring, H.H. Lauer, 'Analysis of Proteins and Peptides by Free Solution Capillary Electrophoresis', in *High-performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, eds. C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, 895–901, 1991.
108. A. Vinther, A. Holm, T. Hoeg-Jensen, A.M. Jespersen, N.K. Klausen, T. Christensen, H.H. Sorensen, 'Synthesis of Stereoisomers and Isoforms of a Tryptic Heptapeptide Fragment of Human Growth Hormone and Analysis by Reversed-phase HPLC and Capillary Electrophoresis', *Eur. J. Biochem.*, **235**, 304–309 (1996).
109. J. Bongers, T. Lambros, A.M. Felix, E.P. Heimer, 'Capillary Zone Electrophoresis of Degradative and Cyclic Lactam Derivatives of the Growth Hormone-releasing Factor Peptide', *J. Liq. Chromatogr.*, **15**, 1115–1128 (1992).
110. Y.L. Zhang, J.M. Becker, F.R. Naider, 'Electrophoretic Behavior of L- and D-Alanine-Scanning Analogs of a Yeast Tridecapeptide Pheromone in a Fused-silica Capillary', *Anal. Biochem.*, **241**, 220–227 (1996).
111. T. Yashima, A. Tsuchiya, O. Morita, S. Terabe, 'Separation of Closely Related Large Peptides by Micellar Electrokinetic Chromatography with Organic Modifiers', *Anal. Chem.*, **64**, 2981–2984 (1992).
112. N. Matsubara, S. Terabe, 'Separation of 24 Dansylated Amino Acids by Capillary Electrophoresis with a Nonionic Surfactant', *J. Chromatogr. A*, **680**, 311–315 (1994).
113. K. DeSilva, Q. Jiang, T. Kuwana, 'Chiral Separation of Naphthalene-2,3-Dialdehyde labelled Peptides by Cyclodextrin-modified Electrokinetic Chromatography', *Biomed. Chromatogr.*, **9**, 295–301 (1995).
114. X.-Z. Qin, D.-S.T. Nguyen, D.P. Ip, 'Separation of Lisinopril and its RSS Diastereomer by Micellar Electrokinetic Chromatography', *J. Liq. Chromatogr.*, **16**, 3713–3734 (1993).
115. H. Wan, L.G. Blomberg, 'Enantiomeric and Diastereomeric Separation of Di- and Tripeptides by Capillary Electrophoresis', *J. Chromatogr. A*, **758**, 303–311 (1997).
116. M. Gilges, M. Hadley, 'Resolution of the Diastereomers of a Large Synthetic Peptide by Capillary Electrophoresis Using Nonionic Surfactants', *Electrophoresis*, **18**, 2944–2949 (1997).

Posttranslational Oxidative Modifications of Proteins

Christian Schöneich and Lorena B. Barrón
The University of Kansas, Lawrence, USA

1 Introduction 1
2 General Strategies 1
3 Specific Strategies for Selected Modifications 2
3.1 Protein Carbonyls 2
3.2 Cysteine 2
3.3 3-Nitrotyrosine 3
3.4 Other Modifications of Tyrosine 4
3.5 Oxidative Modifications of Tryptophan 4
3.6 2-Oxo-histidine 4
3.7 Hydroxylation of Valine, Leucine, and Lysine 5
3.8 Methionine Sulfoxide 5
3.9 Advanced Glycation Endproducts 6
4 Conclusions 6
Acknowledgments 6
Abbreviations and Acronyms 6
Related Articles 7
References 7

Specific oxidative modifications of amino acids in proteins are used to monitor the exposure of biological tissue to oxidative stress. This article introduces the most prevalent oxidative amino acid modifications which have been characterized in vivo and summarizes the analytical methodology for their identification and quantification.

1 INTRODUCTION

The oxidative posttranslational modification of proteins is an important feature of biological oxidative stress, aging,[1,2] and the manufacturing and storage of protein pharmaceuticals.[3] Often, but not always, protein oxidation is accompanied by structural and functional changes, loss of thermal stability,[4] and accelerated turnover.[4,5] However, to what degree these parameters are affected by oxidation depends on the nature and the extent of modification. Much effort has been spent on the analysis of certain amino acid and protein oxidation products as general 'biomarkers' of oxidative stress, e.g. protein carbonyls,[6] and indicators of the action of certain reactive oxygen and nitrogen species, e.g. 3-nitrotyrosine (3-NY)[7] **(1)** or *o*- and *m*-hydroxyphenylalanine.[8] Such strategies are valuable to demonstrate the exposure of tissue to oxidative processes. However, attempts to deduce the severity of oxidative stress from a quantitative measurement of protein oxidation products need to be discussed carefully with regard to the following issues. Based on sequence, structure and location in tissue, different proteins may show (i) different sensitivities towards oxidation in general, and (ii) different product patterns for a single oxidant. For example, by measurement of protein carbonyl formation, Shacter et al.[9] identified fibrinogen as particularly sensitive to oxidative stress in vivo. In contrast, the age-dependent in vivo oxidation of calmodulin showed no significant accumulation of protein carbonyls but high levels of methionine (Met) sulfoxide instead.[10,11] It is also important to evaluate the experimental method relative to the question to be answered. For example, Leeuwenburgh et al. quantified 3-NY in skeletal muscle homogenates of rats and concluded that biological aging is not accompanied by an increase in 3-NY formation.[12] However, when specific skeletal muscle proteins were analyzed, a significant accumulation of 3-NY (up to 3.5 mol/mol) on the SERCA2a isoform of the sarcoplasmic reticulum Ca-ATPase was observed, associated with a loss of function.[13,14] In essence, functionally important oxidative modifications of proteins can be missed by the analysis of whole tissue homogenate when the proteins are expressed in low yields. In the following, we shall briefly review some general analytical strategies, followed by methods developed for the most prevalent oxidative modifications, especially in view of low sample volumes derived from tissue or cell cultures.

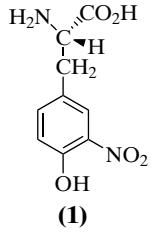

(1)

2 GENERAL STRATEGIES

If sample volume permits, several standard techniques are available for the characterization of oxidative protein

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

modifications. Proteins may be isolated or separated by gel electrophoresis. Both the isolated protein or specific bands on a gel may be subjected to microsequencing techniques, proteolytic digestion, high-performance liquid chromatography (HPLC) or capillary electrophoresis, and mass spectrometric analysis. These techniques have been described in detail elsewhere in this volume.

3 SPECIFIC STRATEGIES FOR SELECTED MODIFICATIONS

3.1 Protein Carbonyls

There are various routes to protein carbonyls involving either the direct oxidation of an amino acid side chain (predominantly lysine (Lys), arginine (Arg), proline (Pro), threonine (Thr))[15] or the covalent attachment of lipid peroxidation products (e.g. 4-hydroxy-α-β-unsaturated aldehydes via Michael addition) to a protein (protein-associated carbonyls).[16,17] Hence, the measurement of protein carbonyls does not necessarily indicate protein oxidation only but, more generally, the occurrence of oxidative processes in the environment of a protein. Because of its sensitivity, it has evolved into a routine tool to assay oxidative stress. Usually, proteins are derivatized with 2,4-dinitrophenylhydrazine (2,4-DNPH) to yield the corresponding hydrazones.[6,9] The latter can be directly quantified by ultraviolet (UV) spectroscopy ($\lambda = 360-370$ nm). However, specific care must be taken with regard to loss of protein during washing steps and the potential contamination by excess 2,4-DNPH or nucleotides.[18] Such problems may be circumvented using size-exclusion chromatography and dual wavelength detection at 280 and 370 nm.[6] Initially, proteins can be concentrated by precipitation with trichloroacetic acid or ammonium sulfate, and resolubilized for derivatization with either 6.0 M guanidine hydrochloride or sodium dodecyl sulfate (SDS). However, in their work on soluble proteins from rat brain, Floor and Wetzel noted that about half of the proteins were not redissolved unless 6.0 M guanidine was supplemented with 50% formic acid.[19] Initially, a Zorbax GF450 column was recommended for the chromatographic step, mostly based on its performance tolerating higher flow-rates and back pressure associated with the use of a guanidine/phosphate buffer as the eluent. However, these problems may be avoided, and other columns used, when the size exclusion chromatography is performed in SDS.[6] Once derivatized with 2,4-DNPH, proteins can also be separated by gel electrophoresis and detected by Western blotting[6] or enzyme-linked immunosorbent assay (ELISA)[20] using an antibody against the 2,4-dinitrophenyl moiety. Alternative labeling reagents for protein carbonyls include digoxigenin-hydrazide (followed by staining with an anti-digoxigenin antibody)[21] and biotin-4-amidobenzoic hydrazide, visualized by streptavidin conjugated to a fluorescent dye, dichlorotriazinylaminofluorescein.[22] Protein carbonyls can also be reduced to the corresponding alcohols with NaB_3H_4 and proteins or amino acid hydrolysates monitored by scintillation counting.[23]

A specific pathway to the formation of protein-associated carbonyls involves the addition of the lipid peroxidation product 4-hydroxynonenal (4-HNE) to proteins.[16,17,24] Antibodies against protein-bound 4-HNE have been used for the screening of 4-HNE-protein adducts formed during conditions of oxidative stress, e.g. for oxidized low density lipoprotein (LDL) where 4-HNE-adducts were formed with Lys and histidine (His).[25] Subsequently, 4-HNE-addition to various His residues of apoB was confirmed by mass spectrometry (MS).[26]

3.2 Cysteine

The sulfhydryl group of cysteine is easily modified by various reactive oxygen and nitrogen species. Products include sulfenic, sulfinic or sulfonic acid, disulfide, nitrosocysteine and nitrocysteine. As the oxidation state of cysteine often determines the structure, function and properties of a protein[27] it is important to assess the cysteine content of proteins under investigation. The most common method for determining protein cysteinyl residues is derivatization with Ellman's reagent (5,5′-dithiobis (2-nitrobenzoate) DTNB).[28] This assay, although rapid and reproducible, has the disadvantage of lack of sensitivity at cysteine concentrations below 3 μM.[29] More sensitive assays have been developed based on fluorescent reagents such as rhodamine, dansyl chloride, fluoresceine iodoacetamide derivatives, monobromobimane, and maleimide derivatives. After modification, the protein can be digested and the labeled fragments can be quantified by fluorescence spectroscopy and/or reversed phase high-performance liquid chromatography (RPHPLC) coupled to a fluorescence detector. For the detection of protein disulfides, cysteine residues are alkylated (i.e. with iodoacetate) before the protein is reduced and the resulting additional cysteine residues labeled with the fluorescent tag.[30] When a fluorescence detector is not available, cysteine residues can be labeled with derivatizing agents containing UV chromophores. The use of hydrophobic chromophores for the selective isolation of cysteine-containing peptides was first reported by Chang et al.[31] Recently, the identification of the functionally important cysteine residues in Ca-ATPase through

derivatization with 4-dimethylaminophenyl-azophenyl-4′-maleimide (DABMI) was reported by Viner et al.[32] After modification, the protein was digested and the resulting peptides were analyzed by HPLC coupled with MS. The excess label needs to be removed by washing with organic solvents (dimethylformamide (DMF)).[33]

Papain amplification was reported as a sensitive assay for the determination of cysteine in proteins.[29] The basis of this assay is to generate a stoichiometric amount of active enzyme rather than a level of a protein-bound chromophore or fluorophore. Once the active enzyme is generated, it produces a chromophoric product proportional to the amount of active enzyme. Since this is an amplification assay the sensitivity and detection limits are proportional to the incubation time of the reactivated papain with substrate. Wright et al. reported a 10-fold increase in sensitivity over an improved Ellman's assay.[29]

Other chemical modifications have been used in conjunction with chromatographic methods. Ming et al. successfully quantified the eight cysteines of brazzein, a natural high potency protein from a wild West African plant. The protein's cysteinyl residues were modified by 4-vinylpyridine and then desalted by RPHPLC. The eluted samples were analyzed by electrospray ionization/mass spectrometry (ESI/MS).[34] Another chemical modification to quantitate cysteines employees a selective cleavage followed by analysis with matrix-assisted laser desorption ionization/time-of-flight/mass spectrometry (MALDI/TOF/MS).[35] The modification involves the specific reaction between free sulfhydryls with 2-nitro-5-thiocyanobenzoic acid (NCTB). The *N*-terminal peptide bond is cleaved under alkaline conditions to form an amino-terminal peptide and a series of 2-iminothiazolidine-4-carboxylyl peptides can be mapped to a sequence by MALDI/TOF/MS. This assay can be performed with native and reduced protein in order to differentiate between free and disulfide-linked cysteine residues.

Weiner[36] described a method to quantitate protein cysteine residues by spin labeling, employing a symmetrical disulfide-linked biradical, bis(2,2,5,5-tetramethyl-3-imidazoline-1-oxyl-4-yl)-disulfide. The reaction of protein cysteine residues with the biradical cleaves the disulfide bridge and generates monoradicals which can be monitored by electron paramagnetic resonance (EPR). The reaction rate is proportional to the reactivity, i.e. accessibility of the cysteine residue. This sensitive method allows the quantitation of $<10^{-12}$ M, and provides a useful tool to study protein folding.

With the discovery of nitric oxide (NO) as an important biological messenger, the detection of cysteine *S*-nitrosylation has become an important area of research. The location of *S*-nitrosocysteine residues in proteins can be achieved by MS; however, during the preparation of the protein reducing agents need to be avoided. There is also always the problem that *trans*-nitrosation reactions may occur in proteins which contain several free cysteine residues. Fluorometric and colorimetric techniques for the detection of *S*-nitroso compounds utilize the metal-catalyzed liberation of NO which combines with an appropriate precursor(s), i.e. 2,3-diaminonaphthalene to yield a fluorescent 2,3-naphthotriazole (NAT), or sulfanilamide and *N*-(1-naphtyl)ethylenediamine dihydrochloride to yield a chromophore.[37,38]

Thomas and colleagues developed an isoelectrofocusing method for the separation of *S*-nitrosylated protein from other isoforms such as native, oxidized and *S*-glutathiolated protein.[39] This method has been applied to detect protein heterogeneity in cell lysates. However, in order to utilize this method for the identification of specific protein modification products, the method shall be standardized carefully in vitro.

3.3 3-Nitrotyrosine

Since it has been recognized that 3-NY is a product of the reaction of peroxynitrite ($ONOO^-$) with tyrosine (Tyr) residues,[7] and that 3-NY is present in human atherosclerotic tissue,[40] there has been much interest to analyze the presence of 3-NY in tissue as a 'biomarker' for the formation and reaction of peroxynitrite. It is now clear, however, that, besides peroxynitrite, several reactive nitrogen species and enzymatic systems, utilizing nitrite, can generate 3-NY so that the presence of 3-NY in tissue is indicative collectively for the formation of reactive nitrogen species.[41]

Both free and protein-bound 3-NY can be detected by various analytical techniques and it is recommended to use at least two independent methods to monitor 3-NY. Potential artefacts based on a single technique have been reported recently. When Kaur et al. analyzed human brain tissue from patients with various neurodegenerative diseases (Parkinson's disease, Huntington chorea, Alzheimer's disease, etc.) for 3-NY they detected by HPLC analysis a peak with a retention time matching that of authentic 3-NY.[42] However, further analysis by diode array UV spectroscopy and MS revealed that the monitored peak was not 3-NY. The presence of this artefact was confirmed by Althaus et al.[43]

Various techniques have been developed for the analysis of both free and protein-bound 3-NY in/from tissue. Tissue homogenate can be screened by Western-blotting using a monoclonal anti-3-NY antibody.[44,45] Two important control experiments involve the incubation of the antibody with 3-NY (antigen-competed antibody) and the chemical reduction of 3-NY to 3-aminotyrosine (3-AY) with dithionite (removal of antigen) prior to Western blotting. With the same antibody, 3-NY-containing

proteins can be immunoprecipitated for further characterization by sequencing, peptide mapping, amino acid analysis and/or MS.[46] Immunoprecipitation is, of course, also possible with antibodies specific to the protein of interest. However, it was noted that nitrated proteins are sometimes less well recognized by protein-specific antibodies as are the native non-modified proteins.[47] When protein-bound 3-NY is to be analyzed as free 3-NY after hydrolysis of the protein, special attention has to be paid to the hydrolysis procedure. It has been shown that 3-NY can be formed during acid hydrolysis when contaminating nitrite is present in the reaction mixture.[48] Contaminating nitrite can be removed by exhaustive dialysis. Alternatively, the protein may be proteolyzed enzymatically, for example by pronase E. However, enzymatic digestion always raises the possibility of incomplete cleavage or contamination of the protein sample through autodigestion of the proteolytic enzyme.

For the most part, the analysis of free 3-NY involves HPLC coupled to UV or electrochemical detection (after reduction to 3-AY, fluorescence detection may be used as well; however, 3-AY is unstable and slowly reoxidizes). 3-NY has been analyzed in its free or *N*-derivatized form (*N*-acetylation, derivatization with phenyl isothiocyanate; PITC) and, electrochemically, after reduction to 3-AY.[14,45,48–50] Electrochemical detection has been combined with post-column UV photolysis in a TiO_2-coated reaction coil.[51,52] An alternative strategy for a more specific detection of 3-NY (and other Tyr oxidation products) involves isotope dilution MS. After reaction with *n*-propanol (*C*-terminus) and heptafluorobutyric acid anhydride (*N*-terminus and phenoxy group) the resulting 3-NY derivative is analyzed by gas chromatography (GC)/negative-ion electron capture MS.[53]

3.4 Other Modifications of Tyrosine

Several other characteristic products of Tyr modification **(2–8)** have been identified in vitro and in vivo for several pathologic conditions associated with oxidative stress (atherosclerosis, nuclear cataractogenisis, aging etc.).[12,53–55]

Some of these products are characteristic for specific modifying agents and intermediates such as 3,4-dihydroxyphenylalanine (DOPA) **(2)** (hydroxyl radicals),[54,55] 3-chlorotyrosine (HOCl, Cl_2) **(3)**,[49,53,56] 3-bromotyrosine (HOBr) **(4)**,[57] *o*,*o*′-dityrosine **(5)**, isodityrosine **(6)**, trityrosine **(7)**, and pulcherosine **(8)** (tyrosyl radicals).[58] GC/MS analysis has been useful especially for the analysis of 3-chloro- and 3-bromotyrosine as well as *o*,*o*′-dityrosine. Alternatively, these products can be detected by HPLC coupled to UV (direct detection or derivatization with orthophthalaldehyde; OPA), fluorescence (dityrosine, trityrosine, pulcherosine; 3-chlorotyrosine after derivatization with 1-nitroso-2-naphtol) or electrochemical detection (3-chlorotyrosine). Especially, *o*- and *m*-hydroxyphenylalanine have been used as indices for the formation and reaction of hydroxyl radicals.[8,12] These products can be monitored by several amino acid analysis techniques as well as GC/MS.

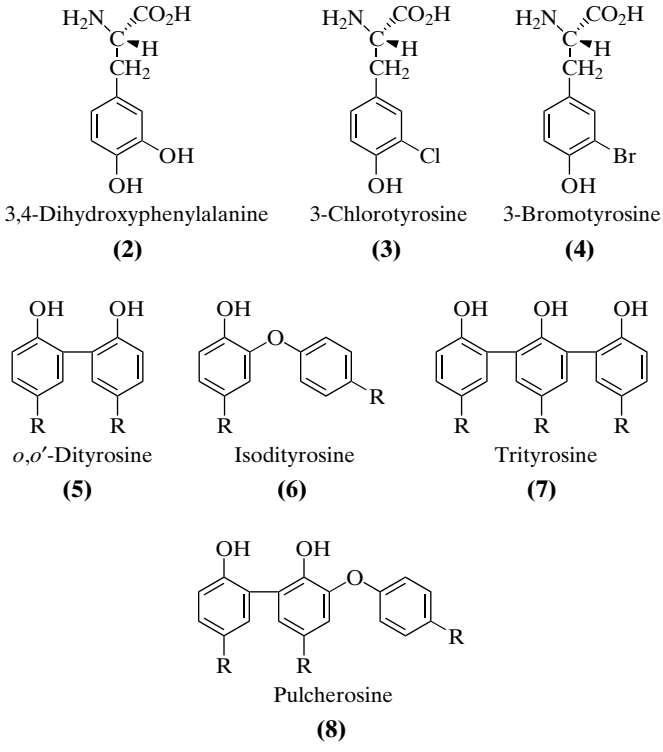

3.5 Oxidative Modifications of Tryptophan

The oxidation of tryptophan (Trp) leads to various aromatic hydroxylation products as well as fluorescent products such as *N*-formylkynurenine.[59] The latter exhibits an emission maximum around $\lambda_{em} = 430\,nm$ and peptides containing *N*-formylkynurenine are conveniently detected by HPLC coupled to fluorescence detection. We note that a common Tyr oxidation product, *o*,*o*′-dityrosine shows an emission maximum around $\lambda_{em} = 420\,nm$. Hence, when fluorescent oxidation products are present in proteins their identity shall be confirmed by amino acid analysis and/or MS.

3.6 2-Oxo-histidine

Based on their metal-binding properties, protein His residues are preferred sites for site-specific metal-catalyzed oxidation. General principles underlying the site-specific metal-catalyzed oxidation of proteins have been established with the protein glutamine synthetase[60,61] which was also shown to be a target

for oxidation during oxidative stress.[62] The specific sensitivity of His to metal-catalyzed oxidation is probably best documented in glutamine synthetase where His[269], located in the sequence Met/His[269]-Cys-His/Met, is oxidized whereas the generally more oxidation-sensitive residues Cys and Met are left intact.[61] Subsequently, the metal-catalyzed modification of His residues was demonstrated for several other peptides and proteins (for a recent review, see Schöneich[63]).

One characteristic oxidation product of His is 2-oxo-His **(9)** though, depending on the extent of His oxidation, higher oxidation products of His may be observed.

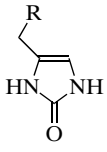

(9)

2-Oxo-His can be conveniently detected by classical amino acid analysis methods including ion-exchange chromatography and postcolumn derivatization or reversed-phase chromatography coupled to precolumn derivatization with OPA or 6-aminoquinoyl-*N*-hydroxysuccinimidyl carbamate (AQC).[64,65] Importantly, 2-oxo-His is relatively unstable during acid hydrolysis (6 M HCl) unless dithiothreitol (>20 mM) is included in the reaction mixture. Especially for the derivatization with OPA, the pH of the reaction mixture needs to be lowered towards pH 8.0 as 2-oxo-His is chemically too unstable at higher pH values. Alternatively to the fluorescence detection of derivatized 2-oxo-His, 2-oxo-His may be selectively monitored directly by HPLC coupled to an electrochemical detector (ECD) set at 0.85 V.[66]

3.7 Hydroxylation of Valine, Leucine, and Lysine

Especially when exposed to the highly oxidizing hydroxyl radical, proteins can accumulate oxidative modifications also on otherwise less reactive aliphatic amino acid side chains such as those of leucine and valine.[67,68] Initially, the reaction of hydroxyl radicals with these amino acids generates hydroperoxides of which several isomers and diastereomers have been characterized in detail for both valine and leucine. The latter can be reduced to the corresponding hydroxy derivatives by $NaBH_4$ or other, biologically relevant reductants, and analyzed by reversed-phase chromatography using pre- or postcolumn derivatization with OPA. Some structures of hydroxyleucine[68] are shown **(10–12)**:

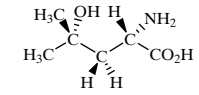

(2*S*)-γ-Hydroxyleucine

(10)

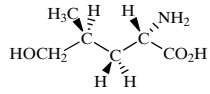

(2*S*,4*S*)-δ-Hydroxyleucine

(11)

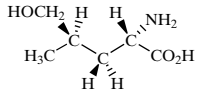

(2*S*,4*R*)-δ-Hydroxyleucine

(12)

It was proposed that the hydroxylation of valine and leucine serves as a biomarker specifically for the exposure of proteins to hydroxyl radical-generating systems. In vivo, hydroxyvaline and hydroxyleucine have been characterized for various diseased tissue such as advanced human atherosclerotic plaques[55] and nuclear cataractogenisis.[54] The oxidation of valine and leucine was also observed during the incubation of collagen with glucose, i.e. as a result of glycation-dependent oxidation processes.[69]

Lys is quite sensitive to metal-catalyzed oxidation which usually results in *N*-dealkylation and the formation of protein carbonyls (see above). However, when exposed to free hydroxyl radicals other characteristic oxidation products have been detected such as 3-, 4- and 5-hydroxylysine (in vitro the hydroxylation is preceded by hydroperoxide formation and the hydroperoxides reduced with $NaBH_4$; in vivo, alternative biological reductants may operate).[70] Whereas 5-hydroxylysine is also generated enzymatically by lysyl oxidase, 3- and 4-hydroxylysine have been proposed as indices for (hydroxyl) radical-mediated protein oxidation. These products have been monitored by HPLC analysis following derivatization to their 9-fluorenylmethylchloroformate (FMOC)-derivatives.

3.8 Methionine Sulfoxide

L-Methionine sulfoxide (L-MetSO) is a common oxidation product of L-Met in vivo and in vitro which can be conveniently detected by classical amino acid analysis methods. Proteins can be hydrolyzed with 1 M NaOH or 4 M methanesulfonic acid (CH_3SO_3H) whereas hydrolysis with 6 M HCl is not recommended as under such conditions MetSO can convert back to Met.

An important detail of L-Met oxidation is the potential formation of two MetSO diastereomers, L-Met-*d*-SO **(13)** and L-Met-*l*-SO **(14)**. The occurrence of diastereoselective oxidation can indicate the diastereoselective chemistry of an oxidant or a specific conformational arrangement of the target L-Met residue in a protein. In vivo, other parameters may contribute to the accumulation of specific L-MetSO diastereomers such as the diastereoselective repair of L-MetSO by the

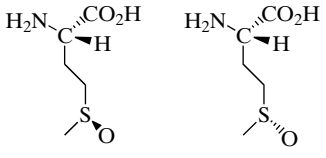

enzyme Met sulfoxide reductase or a selective turnover of oxidized protein. L-Met-*d*-SO and L-Met-*l*-SO can be separated by reversed phase chromatography after precolumn derivatization with 4-dimethylamino-azobenzene-4′-sulfonylchloride (DABSYL)[71] or OPA.[72] In order to avoid reduction or racemization of the L-MetSO diastereomers during recovery from an oxidized protein, the protein was subjected to proteolytic cleavage, and isolated proteolytic fragments to further digestion by carboxypeptidase Y.[71] Both L-MetSO diastereomers were separated well from other amino acids in these digests by reversed-phase chromatography after precolumn derivatization with OPA.

The location of Met residues in proteins has been routinely assessed through cyanogen bromide (CNBr) cleavage followed by *N*-terminal sequencing of the resulting fragments. By comparison of the fragments from a native and oxidized protein, the location of MetSO residues can be characterized. Recently, Levine et al. have optimized a simultaneous sequencing procedure on a Hewlett Packard G1005A automated sequencer in order to detect the locations of MetSO in oxidized glutamine synthetase.[73] Glutamine synthetase was cleaved with CNBr, the resulting fragments loaded onto the sequencer column, and the pattern of phenylthiohydantoin (PTH)-amino acids recorded. The subtle but evident changes in these patterns caused by oxidation of the protein were sufficient to identify the locations of the MetSO residues.

3.9 Advanced Glycation Endproducts

Protein glycation results in a large manifold of structures of which only a fraction has been chemically characterized in detail. One prominent glycation endproduct is carboxymethyllysine (CML) which can be conveniently monitored by GC/MS as trifluoroacetyl methyl ester derivative.[74] Furthermore antibodies against CML have been developed and used, for example, to detect CML in atherosclerotic lesions of human aorta.[75] The mechanisms of CML formation have been studied in quite detail and it is interesting to note that CML formation occurred faster as a result of lipid peroxidation as compared to glucose oxidation.[74] Hence, CML formation is not specifically indicative of glucose oxidation.

4 CONCLUSIONS

This brief article illustrates the variety of possible products resulting from the posttranslational oxidation of proteins. It can be anticipated that an increasing amount of products will be characterized in the future as our understanding of free radical reactions with proteins develops. It is now mandatory not only to study which of these products are characteristic for various pathological conditions but also to identify the affected proteins and potential functional consequences. The latter may directly correlate with physiological dysfunctions.

ACKNOWLEDGMENTS

Much of our research on the oxidative modifications of peptides and proteins was supported by the NIH (PO1AG12993 and GM08359).

ABBREVIATIONS AND ACRONYMS

AQC	6-Aminoquinoyl-*N*-hydroxy-succinimidyl Carbamate
Arg	Arginine
CML	Carboxymethyllysine
DABMI	4-Dimethylaminophenyl-azophenyl-4′-maleimide
DABSYL	4-Dimethylamino-azobenzene-4′-sulfonylchloride
DMF	Dimethylformamide
DOPA	3,4-Dihydroxyphenylalanine
DTNB	5,5′-Dithiobis (2-nitrobenzoate)
ECD	Electrochemical Detector
ELISA	Enzyme-linked Immunosorbent Assay
EPR	Electron Paramagnetic Resonance
ESI/MS	Electrospray Ionization/ Mass Spectrometry
FMOC	9-Fluorenylmethylchloroformate
GC	Gas Chromatography
His	Histidine
HPLC	High-performance Liquid Chromatography
LDL	Low Density Lipoprotein
L-MetSO	L-Methionine Sulfoxide
Lys	Lysine

MALDI/TOF/MS	Matrix-assisted Laser Desorption Ionization/Time-of-flight/ Mass Spectrometry
Met	Methionine
MS	Mass Spectrometry
NAT	2,3-Naphthotriazole
NCTB	2-Nitro-5-thiocyanobenzoic Acid
OPA	Orthophthalaldehyde
PITC	Phenyl Isothiocyanate
Pro	Proline
PTH	Phenylthiohydantoin
RPHPLC	Reversed Phase High-performance Liquid Chromatography
SDS	Sodium Dodecyl Sulfate
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultraviolet
2,4-DNPH	2,4-Dinitrophenylhydrazine
3-AY	3-Aminotyrosine
3-NY	3-Nitrotyrosine
4-HNE	4-Hydroxynonenal

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Gel Electrophoresis in Protein and Peptide Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Proteolytic Mapping • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

REFERENCES

1. E.R. Stadtman, 'Protein Oxidation and Aging', *Science*, **257**, 1220–1224 (1992).
2. B.S. Berlett, E.R. Stadtman, 'Protein Oxidation in Aging, Disease, and Oxidative Stress', *J. Biol. Chem.*, **272**, 20 313–20 316 (1997).
3. Ch. Schöneich, M. Hageman, R.T. Borchardt, 'Stability of Peptides and Proteins', in *Controlled Drug Delivery*, ed. K. Park, American Chemical Society, Washington, DC, 205–228, 1997.
4. A.J. Rivett, R.L. Levine, 'Metal-catalyzed Oxidation of *Eschericia coli* Glutamine Synthetase: Correlation of Structural and Functional Changes', *Arch. Biochem. Biophys.*, **278**, 26–34 (1990).
5. R.E. Pacifici, Y. Kono, K.J.A. Davies, 'Hydrophobicity as the Signal for Selective Degradation of Hydroxyl Radical-modified Hemoglobin by the Multicatalytic Proteinase Complex, Proteasome', *J. Biol. Chem.*, **268**, 15 405–15 411 (1993).
6. R.L. Levine, J.A. Williams, E.R. Stadtman, E. Shacter, 'Carbonyl Assays for Determination of Oxidatively Modified Proteins', *Methods Enzymol.*, **233**, 346–363 (1994).
7. J.S. Beckman, 'Oxidative Damage and Tyrosine Nitration from Peroxynitrite', *Chem. Res. Toxicol.*, **9**, 836–844 (1996).
8. I.N. Acworth, M.B. Bogdanov, D.R. McCabe, M.F. Beal, 'Estimation of Hydroxyl Free Radical Levels in vivo based on Liquid Chromatography with Electrochemical Detection', *Methods Enzymol.*, **300**, 297–313 (1999).
9. E. Shacter, J.A. Williams, M. Lim, R.L. Levine, 'Differential Susceptibility of Plasma Proteins to Oxidative Modification: Examination by Western Blot Immunoassay', *Free Radic. Biol. Med.*, **17**, 429–437 (1994).
10. M.L. Michaelis, D.J. Bigelow, Ch. Schöneich, T.D. Williams, L. Ramonda, D. Yin, A.F.R. Hühmer, Y. Yao, J. Gao, T.C. Squier, 'Decreased Plasma Membrane Transport Activity in Aging Brain', *Life Sci.*, **59**, 405–412 (1996).
11. J. Gao, D. Yin, T.D. Williams, T.C. Squier, 'Progressive Decline in the Ability of Calmodulin Isolated from Aged Brain to Activate the Plasma Membrane Ca-ATPase', *Biochemistry*, **37**, 9536–9548 (1998).
12. C. Leeuwenburgh, P. Hansen, A. Shaish, J.O. Holloszy, J.W. Heinecke, 'Markers of Protein Oxidation by Hydroxyl Radical and Reactive Nitrogen Species in Tissues of Aging Rats', *Am. J. Physiol.*, **274**, R453–R461 (1998).
13. R.I. Viner, D.A. Ferrington, A.F.R. Hühmer, D.J. Bigelow, Ch. Schöneich, 'Accumulation of Nitrotyrosine on the SERCA2a Isoform of SR Ca-ATPase of Rat Skeletal Muscle During Aging: A Peroxynitrite-mediated Process?', *FEBS Lett.*, **379**, 286–290 (1996).
14. R.I. Viner, D.A. Ferrington, T.D. Williams, D.J. Bigelow, Ch. Schöneich, 'Protein Modification During Biological Aging: Selective Tyrosine Nitration of the SERCA2a Isoform of the Sarcoplasmic Reticulum Ca-ATPase in Skeletal Muscle', *Biochem. J.*, **340**, 657–669 (1999).
15. E.R. Stadtman, 'Oxidation of Free Amino Acids and Amino Acid Residues in Proteins by Radiolysis and by Metal-catalyzed Reactions', *Annu. Rev. Biochem.*, **62**, 797–821 (1993).
16. L.I. Szweda, K. Uchida, L. Tsai, E.R. Stadtman, 'Inactivation of Glucose-6-phosphate Dehydrogenase by 4-Hydroxy-2-nonenal', *J. Biol. Chem.*, **268**, 3342–3347 (1993).
17. B. Friguet, E.R. Stadtman, L.I. Szweda, 'Modification of Glucose-6-phosphate Dehydrogenase by 4-Hydroxy-2-nonenal', *J. Biol. Chem.*, **269**, 21 639–21 643 (1994).
18. G. Cao, R.G. Cutler, 'Protein Oxidation and Aging. I. Difficulties in Measuring Reactive Protein Carbonyls in Tissues Using 2,4-Dinitrophenylhydrazine', *Arch. Biochem. Biophys.*, **320**, 106–114 (1995).

19. E. Floor, M.G. Wetzel, 'Increased Protein Oxidation in Human Substantia Nigra Pars Compacta in Comparison with Basal Ganglia and Prefrontal Cortex Measured with an Improved Dinitrophenylhydrazine Assay', *J. Neurochem.*, **70**, 268–275 (1998).
20. C.C. Winterbourne, I.H. Buss, 'Protein Carbonyl Measurement by Enzyme-linked Immunosorbent Assay', *Methods Enzymol.*, **300**, 106–111 (1999).
21. J. Bautista, M.D. Mateos-Nevado, 'Immunological Detection and Quantification of Oxidized Proteins by Labelling with Digoxigenin', *Biosci., Biotechnol., Biochem.*, **62**, 419–423 (1998).
22. R.A. Leedle, S.K. Wynia, 'Oxidation Products in Endothelial Cells: A Fluorescence Microscopic Method for Visualizing Carbonyls In Situ', in *The Oxygen Paradox*, eds. K.J.A. Davies, F. Ursini, CLEUP University Press, Padova, Italy, 123–135, 1995.
23. A. Amici, R.L. Levine, L. Tsai, E.R. Stadtman, 'Conversion of Amino Acid Residues in Proteins and Amino Acid Homopolymers to Carbonyl Derivatives by Metal-catalyzed Oxidation Reactions', *J. Biol. Chem.*, **264**, 3341–3346 (1989).
24. E.R. Stadtman, 'Covalent Modification Reactions are Marking Steps in Protein Turnover', *Biochemistry*, **29**, 6323–6331.
25. K. Uchida, S. Toyokuni, K. Nishikawa, S. Kawakishi, H. Oda, H. Hiai, E.R. Stadtman, 'Michael Addition-type 4-Hydroxy-2-nonenal Adducts in Modified Low-density Lipoproteins: Markers for Atherosclerosis', *Biochemistry*, **33**, 12 487–12 494 (1994).
26. M.S. Bolgar, C.-H. Yang, S.J. Gaskell, 'First Direct Evidence for Lipid/Protein Conjugation in Oxidized Human Low Density Lipoprotein', *J. Biol. Chem.*, **271**, 27 999–28 001 (1996).
27. T.-Y. Lie, *The Proteins*, 3rd edition, Academic Press, London, 329–402, 1977.
28. G.L. Ellman, 'Tissue Sulfhydryl Groups', *Arch. Biochem. Biophys.*, **82**, 70–77 (1959).
29. S.K. Wright, R.E. Viola, 'Evaluation of Methods for the Quantitation of Cysteines in Proteins', *Anal. Biochem.*, **265**, 8–14 (1998).
30. M. Cardamone, K. Ashman, M. Brandon, N.K. Puri, 'Identification of Cysteine-containing Peptides During the Peptide Mapping of Recombinant Proteins', *Peptide Res.*, **6**, 242–248 (1993).
31. J.-Y. Chang, R. Knecht, D.G. Braun, 'A New Method for the Selective Isolation of Cysteine-containing Peptides', *Biochem. J.*, **211**, 163–171 (1983).
32. R.I. Viner, T.D. Williams, Ch. Schöneich, 'Peroxynitrite Modification of Protein Thiols: Oxidation, Nitrosylation and *S*-Glutathiolation of Functionally Important Cysteine Residue(s) in the Sarcoplasmic Reticulum Ca-ATPase', *Biochemistry*, **38**, 12 408–12 415 (1999).
33. A. Wawrzynów, J.H. Collins, 'Chemical Modification of the Ca^{2+}-ATPase of Rabbit Skeletal Muscle Sarcoplasmic Reticulum: Identification of Sites Labeled with Aryl Isothiocyanates and Thiol-directed Conformational Probes', *Biochim. Biophys. Acta*, **1203**, 60–70 (1993).
34. D. Ming, J.L. Markley, G. Hellekant, 'Quantification of Cysteinyl Sulfhydryl Residues in Peptides and Proteins by ESI-MS and MALDI-MS', *BioTechniques*, **18**, 808–810 (1995).
35. J. Wu, D.A. Gage, J.T. Watson, 'Strategy to Locate Cysteine Residues in Proteins by Specific Chemical Cleavage Followed by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Anal. Biochem.*, **235**, 161–174 (1996).
36. L.M. Weiner, 'Quantitative Determination of Thiol Groups in Low and High Molecular Weight Compounds by Electron Paramagnetic Resonance', *Methods Enzymol.*, **251**, 87–105 (1995).
37. D.A. Wink, S. Kim, D. Coffin, J.C. Cook, Y. Vodovotz, D. Chistodoulou, D. Jourd'heuil, M.B. Grisham, 'Detection of *S*-Nitrosothiols by Fluorometric and Colorimetric Methods', *Methods. Enzymol.*, **301**, 201–211 (1999).
38. P. Kostka, J.K.J. Park, 'Fluorometric Detection of *S*-Nitrosothiols', *Methods Enzymol.*, **301**, 227–235 (1999).
39. Y. Ji, T.P.M. Akerboom, H. Sies, J.A. Thomas, 'Gel Electrofocusing Method for Studying Protein *S*-Nitrosylation', *Methods Enzymol.*, **301**, 145–151 (1999).
40. J.S. Beckman, Y.Z. Ye, P.G. Anderson, J. Chen, M.A. Accavitti, M.M. Tarpey, C.R. White, 'Extensive Nitration of Protein Tyrosines in Human Atherosclerosis Detected by Immunohistochemistry', *Biol. Chem. Hoppe-Seyler*, **375**, 81–88 (1994).
41. H. Ischiropoulos, 'Biological Tyrosine Nitration: A Pathophysiological Function of Nitric Oxide and Reactive Oxygen Species', *Arch. Biochem. Biophys.*, **356**, 1–11 (1998).
42. H. Kaur, L. Lyras, P. Jenner, B. Halliwell, 'Artefacts in HPLC Detection of 3-Nitrotyrosine in Human Brain Tissue', *J. Neurochem.*, **70**, 2220–2223 (1998).
43. J.S. Althaus, R.T. Carroll, R.L. Root, M.T. Tseng, E.D. Hall, 'Eight-channel Electrochemical Detection of Nitrotyrosine by HPLC: A Survey of Animal Models of Brain Injury', *Nitric Oxide*, **3**, 22 (abstract).
44. Y.Z. Ze, M. Strong, Z.-Q. Huang, J.S. Beckman, 'Antibodies that Recognize Nitrotyrosine', *Methods Enzymol.*, **269**, 201–209 (1996).
45. J.P. Crow, H. Ischiropoulos, 'Detection and Quantitation of Nitrotyrosine Residues in Proteins: In Vivo Marker of Peroxynitrite', *Methods Enzymol.*, **269**, 185–195 (1996).
46. L.A. MacMillan-Crow, J.A. Thompson, 'Immunoprecipitation of Nitrotyrosine-containing Proteins', *Methods Enzymol.*, **301**, 135–145 (1999).
47. C.B. Hellberg, S.E. Boggs, E.G. Lapetina, 'Phosphatidylinositol-3-kinase is a Target for Protein Tyrosine Nitration', *Biochem. Biophys. Res. Commun.*, **252**, 313–317 (1998).
48. M. Shigenaga, 'Quantitation of Protein-bound 3-Nitrotyrosine by High-performance Liquid Chromatography

with Electrochemical Detection', *Methods Enzymol.*, **301**, 27–40 (1999).
49. J.P. Crow, 'Measurement and Significance of Free and Protein-bound 3-Nitrotyrosine, 3-Chlorotyrosine, and Free 3-Nitro-4-hydroxyphenylacetic Acid in Biologic Samples: A High-performance Liquid Chromatography Method Using Electrochemical Detection', *Methods Enzymol.*, **301**, 151–160 (1999).
50. K. Hensley, M.L. Maidt, Z. Yu, H. Sang, W.R. Markesberry, R.A. Floyd, 'Electrochemical Analysis of Protein Nitrotyrosine and Dityrosine in the Alzheimer Brain Indicates Region-specific Accumulation', *J. Neurosci.*, **18**, 8126–8132 (1998).
51. H. Lie, C.T. Duda, T. Huang, W.O. Aruda, P.T. Kissinger, 'Optimization of Post-column Photolysis and Electrochemical Detection for the Liquid Chromatographic Determination of 3-Nitro-L-tyrosine', *J. Chromatogr. A*, **818**, 69–75 (1998).
52. H. Liu, T. Huang, C.B. Kissinger, P.T. Kissinger, 'Comparison of Detection Modes for Liquid Chromatographic Determination of 3-Nitro-L-tyrosine', *J. Chromatogr. B Biomed. Sci. Appl.*, **713**, 289–295 (1998).
53. J.W. Heinecke, F.F. Hsu, J.R. Crowley, S.L. Hazen, C. Leeuwenburgh, D.M. Mueller, J.E. Rasmussen, J. Turk, 'Detecting Oxidative Modification of Biomolecules with Isotope Dilution Mass Spectrometry: Sensitive and Quantitative Assays for Oxidized Amino Acids in Proteins and Tissues', *Methods Enzymol.*, **300**, 124–144 (1999).
54. S. Fu, R. Dean, M. Southan, R. Truscott, 'The Hydroxyl Radical in Lens Nuclear Cataractogenisis', *J. Biol. Chem.*, **273**, 28 603–28 609 (1998).
55. S. Fu, M.J. Davies, R. Stocker, R.T. Dean, 'Evidence for Roles of Radicals in Protein Oxidation in Advanced Human Atherosclerotic Plaque', *Biochem. J.*, **333**, 519–525 (1998).
56. A.J. Kettle, 'Detection of 3-Chlorotyrosine in Proteins Exposed to Neutrophil Oxidants', *Methods Enzymol.*, **300**, 111–120 (1999).
57. W. Wu, Y. Chen, A. d'Avignon, S.L. Hazen, '3-Bromotyrosine and 3,5-Dibromotyrosine are Major Products of Protein Oxidation by Eosinophil Peroxidase: Potential Markers for Eosinophil-dependent Tissue Injury In Vivo', *Biochemistry*, **38**, 3538–3548 (1999).
58. J.S. Jacob, D.P. Cistola, F.F. Hsu, S. Muzaffar, D.M. Mueller, S.L. Hazen, J.W. Heinecke, 'Human Phagocytes Employ the Myeloperoxidase-hydrogen Peroxide System to Synthesize Dityrosine, Trityrosine, Pulcherosine, and Isodityrosine by a Tyrosyl Radical-dependent Pathway', *J. Biol. Chem.*, **271**, 19 950–19 956 (1996).
59. K. Itakura, K. Uchida, S. Kawakishi, 'Selective Formation of Oxindole- and Formylkynurenine-type Products from Tryptophan and its Peptides Treated with a Superoxide-generating System in the Presence of Iron(III)-EDTA: A Possible Involvement with Iron-oxygen Complex', *Chem. Res. Toxicol.*, **7**, 185–190 (1994).
60. R.L. Levine, 'Oxidative Modification of Glutamine Synthetase', *J. Biol. Chem.*, **258**, 11 823–11 827.
61. J.M. Farber, R.L. Levine, 'Sequence of a Peptide Susceptible to Mixed-function Oxidation. Probable Cation Binding Site in Glutamine Synthetase', *J. Biol. Chem.*, **261**, 4574–4578 (1986).
62. C.N. Oliver, P.E. Starke-Reed, E.R. Stadtman, G.J. Lie, J.M. Carney, R.A. Floyd, 'Oxidative Damage to Brain Proteins, Loss of Glutamine Synthetase Activity, and Production of Free Radicals During Ischemia/Reperfusion Injury to Gerbil Brain', *Proc. Natl. Acad. Sci. USA*, **87**, 5144–5147 (1990).
63. Ch. Schöneich, 'Mechanism of Metal-catalyzed Oxidation of Histidine to 2-Oxohistidine in Peptides and Proteins', *J. Pharm. Biomed. Anal.*, in press (1999).
64. S.A. Lewisch, R.L. Levine, 'Determination of 2-Oxohistidine by Amino Acid Analysis', *Anal. Biochem.*, **231**, 440–446 (1995).
65. S.A. Lewisch, R.L. Levine, 'Determination of 2-Oxohistidine by Amino Acid Analysis', *Methods Enzymol.*, **300**, 120–124 (1999).
66. K. Uchida, S. Kawakishi, '2-Oxo-histidine as a Novel Biological Marker for Oxidatively Modified Proteins', *FEBS Lett.*, **332**, 208–210 (1993).
67. S. Fu, L.A. Hick, M.M. Sheil, R.T. Dean, 'Structural Identification of Valine Hydroperoxides and Hydroxides on Radical-damaged Amino Acid, Peptide, and Protein Molecules', *Free Radic. Biol. Med.*, **19**, 281–292 (1995).
68. S.-L. Fu, R.T. Dean, 'Structural Characterization of the Products of Hydroxyl-radical Damage to Leucine and their Detection on Proteins', *Biochem. J.*, **324**, 41–48 (1997).
69. S. Fu, M.-X. Fu, J.W. Baynes, S.R. Thorpe, R.T. Dean, 'Presence of DOPA and Amino Acid Hydroperoxides in Proteins Modified with Advanced Glycation End Products (AGEs): Amino Acid Oxidation Products as Possible Source of Oxidative Stress Induced by AGE Proteins', *Biochem. J.*, **330**, 233–239 (1998).
70. B. Morin, W.A. Bubb, M.J. Davies, R.T. Dean, S. Fu, '3-Hydroxylysine, a Potential Marker for Studying Radical-induced Protein Oxidation', *Chem. Res. Toxicol.*, **11**, 1265–1273 (1998).
71. G. Minetti, C. Balduini, A. Brovelli, 'Reduction of DABS-L-methionine-*d*,*l*-sulfoxide by Protein Methionine Sulfoxide Reductase from Polymorphonuclear Leukocytes: Stereospecificity Towards the *l*-Sulfoxide', *Ital. J. Biochem.*, **43**, 273–283 (1994).
72. V.S. Sharov, D.A. Ferrington, T.C. Squier, Ch. Schöneich, 'Diastereoselective Reduction of Protein-bound Methionine Sulfoxide by Methionine Sulfoxide Reductase', *FEBS Lett.*, **455**, 247–250 (1999).
73. R.L. Levine, L. Mosoni, B.S. Berlett, E.R. Stadtman, 'Methionine Residues as Endogenous Antioxidants in Proteins', *Proc. Natl. Acad. Sci. USA*, **93**, 15 036–15 040 (1996).

74. M.X. Fu, J.R. Requena, A.J. Jenkins, T.J. Lyons, J.W. Baynes, S.R. Thorpe, 'The Advanced Glycation End Product, *N*-Epsilon-(carboxymethyl)lysine, is a Product of Both Lipid Peroxidation and Glycoxidation Reactions', *J. Biol. Chem.*, **271**, 9982–9986 (1996).
75. S. Kume, M. Takeya, T. Mori, N. Araki, H. Susuki, S. Horiuchi, T. Kodama, Y. Miyauchi, K. Takahashi, 'Immunohistochemical and Ultrastructural Detection of Advanced Glycation End Products in Atherosclerotic Lesions of Human Aorta with a Novel Specific Monoclonal Antibody', *Am. J. Pathol.*, **147**, 654–667 (1995).

Protein Purification: Theoretical and Methodological Considerations

Béla János Takács
F. Hoffmann-La Roche Ltd, Basel, Switzerland

1 Introduction **1**
2 Purification of Recombinant Proteins, Expressed in *Escherichia coli* **2**
2.1 Purification of Soluble Proteins 2
2.2 Purification of Proteins from Inclusion Bodies 8
2.3 Refolding of Inclusion Body Proteins 9
3 Purification of Native Intrinsic Membrane Proteins from Mammalian Cells **11**
3.1 Preparation of Sub-cellular Fragments 11
3.2 Solubilization of Cell Membranes 12
3.3 Immunoaffinity Purification of Membrane Proteins 12
4 Summary **13**
Acknowledgments **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **14**

With the rapid advances in recombinant DNA technologies, it became possible to express large amounts of virtually any protein in different host cells, including bacteria, yeast, insect, and mammalian cells. Efficient purification methods are becoming more and more important as the need for protein drugs continues to expand. In this article we cover some basic procedures for the purification of soluble recombinant proteins expressed in Escherichia coli *and methods to obtain soluble and active protein from insoluble and inactive inclusion bodies. Finally, methods for the purification of native intrinsic membrane proteins from human lymphoblastoid cell lines, using immobilized monoclonal antibodies (mAbs), are presented.*

As a first step, the protein of interest has to be released into an aqueous extract. Following extraction, residual debris and intact cells are removed by centrifugation. Nucleic acids are either hydrolyzed or precipitated to reduce the viscosity of the extract. A crude enrichment for the protein of interest is made based on its differential solubility in ammonium sulfate. This is then followed by high-resolution chromatographic methods. The type of chromatographic matrices and the order of their use are evaluated in pilot experiments for column selection. By combining the specificity and discriminatory powers of various chromatographic techniques, it is possible to isolate rapidly a single protein species from a heterogeneous mixture.

For the purification of native intrinsic membrane proteins, the use of immunoaffinity methods is emphasized, since specific mAbs may be obtained even against low-copy-number membrane proteins by reducing a multispecific antibody response to a series of monospecific responses by cloning. The use of immobilized, highly specific mAbs is demonstrated in the purification of human leukocyte antigen D-related (HLA-DR) antigens.

Rapid advances in complete genome sequences provide us with a continuously expanding number of new potential therapeutic targets. These proteins will have to be purified and characterized. The existence of a functional protein blueprint will be invaluable not only for the diagnosis but possibly also for the treatment of many diseases.

1 INTRODUCTION

The term ''protein'' comes from the Greek *proteios*, meaning ''of first importance'', and its use is attributed to Berzelius around 1838.[1] Proteins were isolated from various plant and animal tissues already in the early part of the 19th century. Fractionation methods for proteins using alcohols, acetone and various salt solutions were developed and formed the basis of protein purifications for the next 100 years. Until relatively recently, most protein purification took place in university laboratories in order to obtain samples for enzymatic and structural studies. With the advances in recombinant DNA technology, it became possible to express large amounts of virtually any protein in different host cells, including bacteria, yeast, insect, and mammalian cells.[2] Cloning genes, encoding human proteins, such as interferon, growth hormone and insulin, into bacteria has allowed their production in amounts required for their widespread therapeutic use. Efficient purification methods become more and more important as the need for these protein drugs continues to expand. A high level of expression, of course, facilitates protein purification, where we have to remove contaminants from a highly enriched product rather than try to purify a small component from a complex mixture. Nevertheless, separating the desired recombinant protein from a background of endogenous bacterial contaminants remains an essential and often difficult task.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

With rapid advances in complete genome sequences and the establishment of two-dimensional gel databases for entire organisms or tissues, it is becoming possible to study gene expression in disease states and under various environmental conditions. The rapid identification of proteins using peptide mass fingerprinting coupled with gene knock-out experiments provides us with a continuously expanding number of new potential therapeutic targets. These proteins will have to be purified and characterized. Purified proteins are needed at several steps along the drug development process: (1) for development of assays; (2) for crystallization of apo- and holo-forms of the protein; (3) for high-throughput screening; and (4) as drugs themselves. Purification protocols are also used for the fractionation and enrichment of low-copy-number proteins to facilitate proteome analysis.

Each protein has unique properties, many of which cannot be predicted from the amino acid sequence. For this reason, there are no standard purification protocols available. Nevertheless, there are certain general steps that we can follow to analyze quickly the behavior of a protein and to be able to predict purification methods. Some of these steps are desorbed in this article, first with respect to the purification of soluble recombinant proteins, then the solubilization and refolding of proteins from insoluble inclusion bodies and finally the purification of native integral membrane proteins from human cells.

2 PURIFICATION OF RECOMBINANT PROTEINS, EXPRESSED IN *ESCHERICHIA COLI*

2.1 Purification of Soluble Proteins

2.1.1 Extracting the Protein

As a first step in recombinant protein purification, we have to disrupt the bacterial host in order to release the protein of interest into an aqueous ''extract''. Bacterial cells are typically lysed by using any of the mechanical or chemical methods listed in Table 1.

Repeated cycles of freezing and thawing apparently liberate recombinant proteins from the bacterial cytoplasm, presumably through transient pores, without releasing the bulk of endogenous host proteins.[3] Bacterial pellets, expressing a recombinant protein, are frozen in a dry-ice–ethanol bath and thawed in an ice–water bath. The cycle is repeated three times before the addition of a standard buffer for extraction.

Grinding in alumina is a good way to open cells for the extraction of labile proteins. One part by weight of cell paste is mixed with 2.5 parts by weight of alumina and ground in a precooled mortar until a smooth consistency is obtained. Buffer (2 parts by weight) and deoxyribonuclease (DNase) (10 μg mL^{-1}) are added before the alumina and cell debris are removed by centrifugation at 5000 *g* for 15 min.

Table 1 Opening cells

Freezing and thawing
Grinding in alumina (in a mortar)
Grinding with glass beads (in a blender)
Agitation in the presence of glass beads in a Mickle® apparatus
Sonication
French® pressure cell treatment
Manton–Gaulin® device
Lysozyme–EDTA treatment
Urea
Guanidine–HCl
Sodiumlauryl sarcosinate
Sodiumlauryl sulfate

Larger amounts (100–1000 g wet weight of cell paste) can be opened by grinding with glass beads in a water-jacketed Waring® blender. Frozen bacterial paste is mixed with 1 part by weight of buffer and 2.5 parts by weight of glass beads (Superbrite® 100 beads from Minnesota Mining and Manufacturing Co.) in a cooled Waring® blender and treated for 10–15 min at high speed. To reduce viscosity, DNase is added to 10 μg mL^{-1} of cell suspension.

Small amounts of cells, up to 1 g (about 10^{12} *E. coli*), can be easily opened by sonication, although spherical bacteria such as *Staphylococcus* sp. are relatively resistant to this method. Cell suspension, up to 2×10^{11} cells mL^{-1}, is treated on ice for 3–4 min with bursts of 30 s of sonication at about 50 W.

A combination of lysozyme (0.2 mg mL^{-1})–ethylenediaminetetraacetic acid (EDTA) (2 mM) treatment followed by osmotic lysis is often successful on a small to medium scale. Gram-positive species are most susceptible to lysozyme, although the inclusion of a nonionic detergent (0.1%) in combination with lysozyme–EDTA works well for Gram-negative species also.[4] Enzymatic procedures may be expensive on a large scale; moreover, the enzyme added is one more contaminant that must be subsequently removed later.

Some common chemical release methods include strong denaturing agents such as urea, guanidine hydrochloride[5] and detergents such as sodium lauryl sarcosinate[6] and sodium lauryl sulfate. These methods, although effective in opening recombinant *E. coli* cells, should be reserved for cases where the recombinant protein is present in insoluble inclusion bodies.

Perhaps the most popular methods to open bacteria on a medium to large scale include French® pressure cell treatment, Manton–Gaulin® homogenization and Dyno-Mill® cell disruption. For French® pressure cell

treatment, 25 g (wet weight) of *E. coli* paste, expressing a recombinant protein, is suspended in 50 mL of breaking buffer (BB): 100 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid) (HEPES)–NaOH, pH 8.0, containing 10% glycerol, 1 mM $MgSO_4$ and 150 mM NaCl. Benzonase® (Merck) is usually added to 250 units mL^{-1} to hydrolyze DNA and RNA present. The protease inhibitors Trasylol® (Bayer) and ε-aminocaproic acid (Serva) are added to 100 units mL^{-1} and 5 mM, respectively. A nonionic detergent, either Triton X-100®, Nonidet P-40 (NP-40) or Brij-58, is added to a 0.1% final concentration and the cells are broken in a precooled French® pressure cell (SLM Instruments) at 20 000 lb in^{-2}. One or two passages are usually sufficient to obtain complete cell breakage. Some proteins, such as peptide deformylase, are extremely sensitive to oxidation. This protein is an amide hydrolase and utilizes a ferrous ion as the catalytic metal.[7] Atmospheric oxygen can rapidly oxidize the catalytic Fe^{2+} ion of deformylase into a catalytically inactive Fe^{3+} form. The inclusion of scavengers of O_2, such as the glucose oxidase and catalase system, was found to be effective in preserving deformylase activity. We found that for some recombinant proteins the inclusion of catalase and superoxide dismutase (SOD) at 10 and 0.1 μg mL^{-1}, respectively, in most of the buffers during purification has helped to preserve enzymatic activity.

The pH of the BB is dictated by the isoelectric point (p*I*) of the protein to be purified, and should be at least 1 pH unit higher, to prevent isoelectric precipitation. Proteins are least soluble at their p*I* and once precipitated they might need denaturing conditions for resolubilization. We use a BB with high buffering capacity (100 mM) since we found that biomasses, especially fermenter grown, have a pH around 5. Glycerol is included because it minimizes hydrophobic interactions with various surfaces, stabilizes proteins[8] and at higher concentration (30–50%) can also act as a cryoprotectant.

For the purification of DNA binding recombinant proteins (DNA gyrases, helicases, topoisomerases, etc.), the use of nucleases should be avoided since residual nuclease in the purified protein preparation would hydrolyze the DNA used in the DNA shift experiments. In these cases the BB should contain higher salt concentration (300 mM NaCl) to reduce ionic interactions between the negatively charged DNA backbone and positively charged domains on protein molecules. To reduce viscosity, after cell breakage DNA can be precipitated by polyethyleneimine (PEI) (Polymin P®), protamine sulfate or streptomycin sulfate.[9]

Proteolytic digestion, especially during the early stages of purification, might be a problem. Many investigators therefore include a whole cocktail of protease inhibitors in the extraction buffer.[10] We routinely use Trasylol® and ε-aminocaproic acid. Trasylol®, also known as aprotinin or kallikrein inhibitor, is a broad-spectrum protease inhibitor that inhibits trypsin, chymotrypsin, trypsinogen, urokinase, kallikrein, elastase, plasmin, and tissue and lysosomal proteinases. ε-Aminocaproic acid inhibits chymotrypsin, fibrinolysin and plasmin. The use of protease inhibitors that covalently react with proteins, such as phenylmethylsulfonyl fluoride (PMSF), diisopropyl fluorophosphate (DFP), *N*-α-*p*-tosyl-L-lysine chloromethyl ketone (TLCK) and *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK) is avoided. These inhibitors can modify amino acid side chains on proteins[11] and, if the recombinant protein was prepared for use in humans, such a modification could lead to reduced efficacy of the antigenic material itself and might elicit cross-reactive autoimmune responses. These inhibitors might also interfere with protein crystallization, especially if the modifications are at contact points in the crystal lattice. Furthermore, modified proteins could display multiple spots on two-dimensional gel-electrophoretic analysis, that is, spots with the same molecular size but varying p*I*.

Soon after cell breakage, EDTA is added to the lysed cell suspension to 5 mM to chelate Mg^{2+}, which could cause protein aggregation,[12] to activate metalloproteases and to accelerate protein oxidation.

2.1.2 *Clarification of the Extract*

Following cell disruption by chemical or mechanical procedures, intact cells and residual cell debris are removed by centrifugation at 3000 *g* for 30 min. The supernatant fraction is then subjected to ultracentrifugation at 150 000 *g* for 2 h to sediment cell envelopes. Pellets are usually resuspended in BB with magnetic stirring and the distribution of the protein of interest is assessed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE) analysis. If the recombinant protein is soluble, most of it should be present in the 150 000 *g* supernatant. If it is found in the 3000 *g* pellet, it is an indication that the recombinant protein is present in inclusion bodies. Solubilization and refolding of recombinant proteins from inclusion bodies is covered in sections 2.2 and 2.3.

2.1.3 *Choosing Assays for Protein or Biological Activity*

It is important to be able to assay for the protein of interest early in the purification. The presence of a prominent band on a sodium dodecyl sulfate (SDS) gel might not be enough. I know of cases where investigators purified a protein by following a fat band on SDS gels. After the protein had been crystallized and its structure had been determined, it was realized that they had purified an *E. coli* protein whose structure

was already known. If antibody is available the protein can be followed by immunoblotting;[13] if it is an enzyme, by its enzymatic activity; if it is a receptor, by ligand binding. Amino acid analysis, N-terminal amino acid sequencing or most conveniently in-gel proteolytic digestion followed by matrix-assisted laser desorption ionization/mass spectrometry (MALDI/MS) analysis[14] can also provide us with a positive identification. The biological activity divided by the total amount of protein present yields the specific activity of the assayed protein. These measurements should be routinely performed during the purification train in order to assess the efficiency of each purification step.

2.1.4 Differential Solubility Techniques

Fractionation of complex protein mixtures using organic solvents, mainly ethanol and acetone, was widely used in the early days of protein purification. More recently, these reagents are being displaced by poly(ethylene glycol) (PEG) and ammonium sulfate. Organic solvents destabilize proteins, whereas PEG and ammonium sulfate stabilize them, thereby minimizing losses of biological activity. If it is necessary to interrupt purification, then storing samples as ammonium sulfate suspensions perhaps could provide us with optimum conditions. Montoya and Castell[15] studied the long-term storage of peroxidase-labeled immunoglobulins for use in enzyme immunoassays. Conjugates were stored at various temperatures (−40, +4, +25, +37 °C) in liquid form in the presence of 50% glycerol, lyophilized or as ammonium sulfate precipitates. They evaluated the immunological and enzymatic activities of the conjugates at regular intervals. The best results were obtained with the conjugates stored in ammonium sulfate at 4 °C. Under these conditions they estimated that the half-life of the conjugates was 9 years.

We routinely subject the 150 000 *g* supernatant to stepwise ammonium sulfate fractionation. Solid ammonium sulfate is added to 30% saturation, according to the table provided by England and Seifter.[16] The suspension is stirred until the ammonium sulfate is dissolved and incubated on ice for 30 min. The precipitate formed is removed by centrifugation at 18 000 *g* for 20 min. The volume of the supernatant is measured and more ammonium sulfate is added to 50% saturation. The precipitate formed is pelleted as above and the supernatant fraction is brought to 80% saturation with respect to ammonium sulfate. The suspension is stirred and left on ice for 30 min. Precipitated proteins are pelleted and the distribution of proteins in the three ammonium sulfate pellets is determined by SDS/PAGE analysis. We usually pellet 200–500 µL of each ammonium sulfate suspension separately for SDS/PAGE analysis and the bulk of the ammonium sulfate pellets are kept at 4 °C until the analysis is complete. Ammonium sulfate precipitation has a high capacity but low resolving power; nevertheless, it not only serves as a crude enrichment step for the protein of interest but can also separate fibrous host proteins that have a tendency to polymerize with time and to clog membrane filters and chromatography columns. A word of caution is warranted here. Although most proteins are stable for long periods even in saturated ammonium sulfate, we have encountered some recombinant proteins (e.g. recombinant human retinoic acid receptor-γ, expressed in *E. coli*) that could not be resolubilized, after exposure to 80–100% saturated ammonium sulfate, in an aqueous buffer in which it was previously freely soluble. The same protein, however, could be resolubilized when it was exposed to 50–60% saturated ammonium sulfate. The pH of a saturated ammonium sulfate solution is close to 5 and it should be neutralized with NaOH or aqueous ammonia before use.

2.1.5 Medium- and High-resolution Chromatographic Techniques

From Table 2, it is clear that of the common isolation and purification methods, chromatography provides us with the highest purification capability. The ammonium sulfate

Table 2 Separation methods in a typical isolation and purification train

Method	Separation basis	Isolation, purification, capability
Centrifugation	Density/size	+
Filtration	Size	+
Liquid–liquid extraction	Solubility	+
Precipitation	Solubility	+
Chromatography:		
Anion-exchange	Charge	+++
Cation-exchange	Charge	+++
Hydrophobic interaction	Hydrophobicity	++
Affinity:	Structure	+++++
Immunoaffinity		
Heparin affinity		
Mimetic™ dye–ligand affinity		
Metal–chelate affinity		
Immobilized affinity "tails"		
Reversed-phase	Hydrophobicity	++++
Chromatofocusing	Charge	+++
Preparative isoelectric focusing	Charge	++++
Immobilized artificial membrane	Hydrophobicity	+++
Adsorption	Charge/shape	++
Gelfiltration	Size/Shape	++

The + symbol refers to the purification capability of each method, ranging from low (+) to high (+++++) level.

fraction that contains most of the desired recombinant protein is therefore subjected to chromatographic fractionation. At first a pilot chromatographic experiment is performed manually for column selection, using small (2-mL) disposable columns packed with various gels. We routinely test six different chromatography matrices (Table 3); many others, of course, could be included, such as various immobilized dye–ligand adsorbents.[17] Each column, and the sample to be applied to that column, is equilibrated in an appropriate buffer (Table 3). The sample is applied and the column is washed with loading buffer. The first 5 mL of flow-through material is saved to represent the unbound fraction. The column is washed with an additional 10 column volumes of loading buffer and the bound molecules are eluted by washing the column with elution buffer. Ten 0.5-mL fractions are collected. Volumes of 10 μL of each fraction are spotted on to nitrocellulose membranes and the membranes are stained for 5 min in a solution of 0.1% aniline blue, prepared in methanol–acetic acid–water (5 : 1 : 5). The membranes are then destained in the same solution, without the dye. This allows us to determine quickly the fractions that should be analyzed by SDS/PAGE. The distribution of the target protein in the bound and unbound fractions permits us to make decisions concerning a purification train. The advantage of this approach over a ''blind'' column selection is that it allows us to predict not only the type but also the order of column use (Figure 1), thereby reducing the number of manipulations that

Table 3 Separation methods for column selection

Chromatographic gel	Chromatographic technique	Separation parameter	Loading buffer	Elution buffer
Source™ 30Q (Pharmacia) Heparin-Actigel™ (Sterogene) Phospho-Ultrogel® (BioSepra)	Anion exchange Affinity/ion exchange Affinity/ion exchange	Charge Structure/charge Structure/charge	20 mM TRIS–HCl, pH 8, 10% glycerol, 1 mM EDTA	20 mM TRIS–HCl, pH 8, 10% glycerol, 1 mM EDTA, 1 M NaCl
Source™ 30S (Pharmacia)	Cation exchange	Charge	20 mM NaOAc, pH 6, 10% glycerol, 1 mM EDTA	20 mM NaOAc, pH 6, 10% glycerol, 1 mM EDTA, 1 M NaCl
Ceramic Hydroxyapatite (Bio-Rad)	Adsorption	Anion exchange Cation exchange Hydrogen bonding	2 mM $NaHP0_4^-$, pH 7, 10% glycerol	500 mM $NaHPO_4^-$, pH 7, 10% glycerol
TSK® Phenyl 5PW (TosoHaas)	Hydrophobic interaction	Hydrophobicity	50 mM $NaHPO_4^-$, pH 7, 1 M $(NH_4)_2SO_4$, 1 M glycine	50 mM $NaHPO_4^-$, pH 7

TRIS, tris(hydroxymethyl)aminomethane.

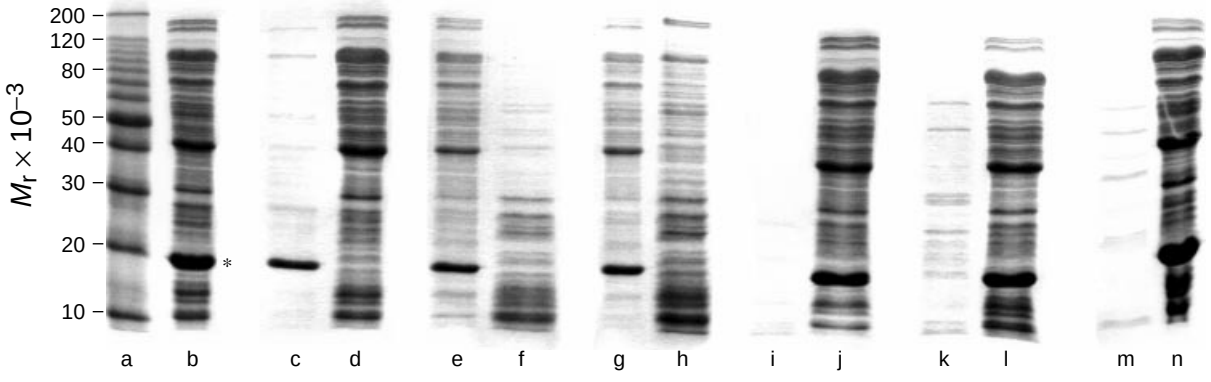

Figure 1 Pilot chromatographic experiment for column selection. An ammonium sulfate fraction of a total *E. coli* cell extract, containing a recombinant protein (∗) (lane b) was applied to six different columns [Ceramic Hydroxyapatite (c, d), Phospho-Ultrogel® (e, f), Heparin-Actigel™ (g, h), Source™ 30Q (i, j), Source™ 30S (k, l) and TSK® Phenyl 5PW (m, n)]. Flow-through (lanes c, e, g, i, k and m) and eluted fractions (lanes d, f, h, j, l and n) were subjected to Coomassie® Blue stained SDS/PAGE analysis. Lane a contains molecular weight standards. Ceramic Hydroxyapatite, which retains most of the contaminating host proteins (lane d), was chosen as the first column. The recombinant protein was not retained on this column (lane c). Source™ 30Q, which binds the recombinant protein (lane j), was chosen as the second column.

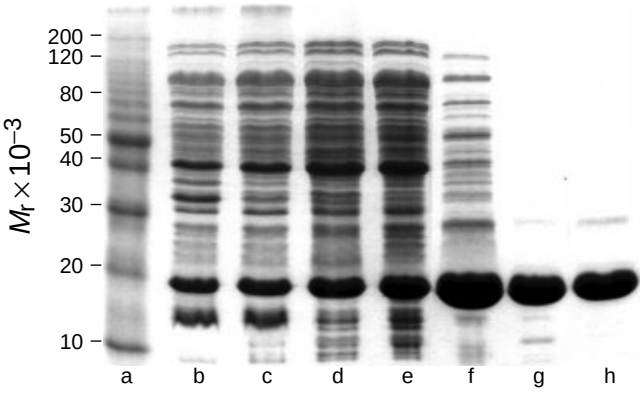

Figure 2 Purification of the recombinant protein described in Figure 1. The total cell extract (lane b) was subjected to two centrifugation steps. The 3000 *g* and 150 000 *g* supernatants are shown in lanes c and d, respectively. Lane e represents proteins precipitated between 30% and 60% ammonium sulfate saturation from the 150 000 *g* supernatant. Lanes f–h show purification after hydroxyapatite, Source™ 30Q and gel filtration chromatography, respectively. Lane a contains molecular weight standards.

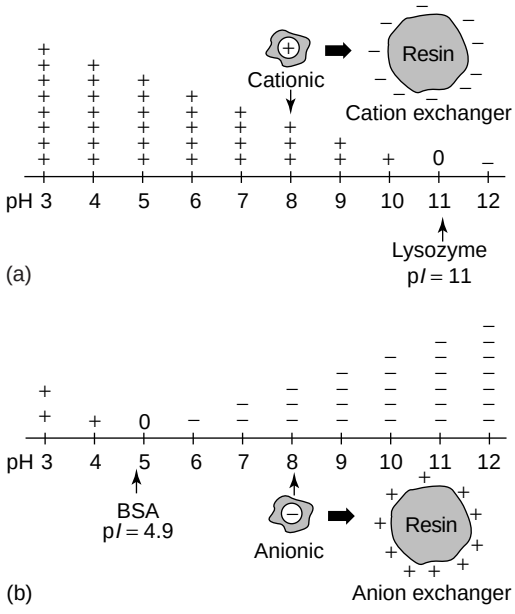

Figure 3 Diagrammatic representation of the charge and binding properties of (a) a basic protein, lysozyme (p*I* = 11), and (b) an acidic protein, BSA (p*I* = 4.9). At pH values above their p*I* proteins are negatively charged and below their p*I* they are positively charged. At pH 8 lysozyme would be positively charged (cationic) and would bind to a cation exchanger which is negatively charged. At pH 8 BSA would be negatively charged (anionic) and would bind to an anion exchanger which is positively charged.

could cause protein denaturation. By combining the specificity and discriminatory powers of the various chromatographic techniques, it is possible to isolate a single protein species from a heterogeneous mixture (Figure 2).

Before starting protein purification, it is important to review the protein's physical and chemical properties. Knowledge of the size, isoelectric point and number of cysteine residues can greatly influence the choice of extraction and fractionation methods. To demonstrate, let us take a hypothetical case depicted in Figure 3(a) and (b). We would like to separate a mixture of two proteins, lysozyme and bovine serum albumin (BSA). Lysozyme, a basic protein, has a p*I* of 11, and BSA, an acidic protein, has a p*I* of 4.9. Proteins at their p*I* have a net charge of zero. At pH values above their p*I* they are negatively charged and below their p*I* they are positively charged. The further away we are from the p*I*, the greater is the charge on the protein. At pH 8, for example, lysozyme would be positively and BSA would be negatively charged. At this pH lysozyme would bind to a cation-exchange resin which has a negative charge and BSA would bind to an anion-exchange resin which has a positive charge. Conversely, at pH 8 lysozyme would be in the flow-through, unbound fraction of an anion exchanger and BSA would be in the flow-through fraction of a cation-exchange column. Figure 4 demonstrates that this reasoning also works in a real-life situation. We were purifying a recombinant procaryotic glycosyltransferase from *E. coli*. This protein is associated with the cell envelope and has a p*I* of 9.38. We were able to purify it from a detergent-solubilized membrane fraction using a single cation-exchange chromatographic step at a pH of 8. At this pH glycosyltransferase was positively charged and bound to a cation exchanger whereas most of the *E. coli* proteins were negatively charged and were found in the unbound fractions (Figure 4, lane c).

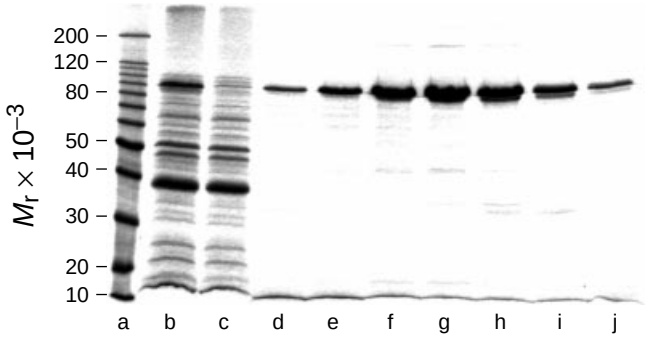

Figure 4 Purification of a recombinant glycosyltransferase from *E. coli* cell membranes. Nonionic detergent extract (lane b) of *E. coli* cell envelope fraction was applied to a Source™ 30S column at pH 8. Unbound proteins (lane c) were washed through with loading buffer. Bound proteins were eluted with a linear NaCl gradient (lanes d–j). Lane a contains molecular weight standards.

The order of the various techniques is important and their consideration must be included in the overall strategy to reduce time spent on sample treatment. For example if most of the target protein is found in the 30–50% saturated ammonium sulfate pellet, then it could be resolubilized and applied directly on to a hydrophobic interaction chromatography (HIC) column in 25% saturated ammonium sulfate. Elution from the HIC column is performed by decreasing the ionic strength, which can make the eluted fractions suitable for application to an ion-exchange or an affinity column.

Of the various chromatographic techniques that are available to us (Table 2), affinity chromatography can provide us with the best isolation and purification capability. For this reason, many investigators, using genetic engineering methods, have added ''tags'' or ''affinity tails'' to the recombinant polypeptide chain to permit their purification in a simplified way. Recombinant proteins, containing protein A or G sequences, can be purified on immobilized immunoglobulin columns.[18,19] The addition of DYKDDDDK amino acid sequences to recombinant polypeptides allows their purification by immunoaffinity on antibody columns specific for this sequence.[20] Furthermore, after purification the affinity tag can be removed with enterokinase. Thiol-containing tails allow the covalent attachment of the tagged protein to thiolated matrices.[21] Genetically engineered C-terminal polyarginine was used[22] to make the fusion protein very basic and thereby to bind to a cation-exchange column under conditions where the majority of the host proteins were found in the flow-through fraction. Removal of the polyarginine tag with carboxypeptidase B and rechromatography resulted in a large change in the elution profile of the recombinant protein but not of the contaminants. Biotin tail has been used to purify biotin-tagged proteins on monomeric avidin resin.[23] Recombinant proteins fused to glutathione *S*-transferase (GST) can be purified on immobilized glutathione-Sepharose® 4B.[24] Polyhistidines have been used extensively to tag recombinant proteins and to allow their purification by metal–chelate affinity chromatography.[5,25,26] The list of various affinity tails is extensive and there is not enough space here to cover them all, so the purification of only two affinity tagged proteins, GST- and hexa-His-tagged fusion proteins, both expressed in *E. coli*, is described. The GST fusion protein contained a thrombin cleavage site to facilitate the removal of the tag by thrombin treatment.

2.1.5.1 Purification of Glutathione S-Transferase Fusion Proteins For purification of GST fusion proteins containing thrombin cleavage site, the 150 000 *g* supernatant is mixed with 50 mL (settled volume) of glutathione-Sepharose® 4B matrix (Pharmacia) and agitated by end-to-end rotation for 16 h at 4 °C. The binding capacity is about 5 mg of GST per milliliter of packed gel. The gel is then pelleted by centrifugation at 650 *g* for 5 min and resuspended in about 100 mL of wash buffer [50 mM HEPES, pH 8, containing 10% glycerol, 150 mM NaCl, 1 mM EDTA-Na_2, 1 mM D, L-dithiothreitol (DTT) and 1% Triton X-100®]. The gel is pelleted as above, resuspended in 100 mL of wash buffer and divided into six 50-mL centrifuge tubes. Each fraction is then underlayered with 25 mL of 10% sucrose, prepared in wash buffer. The tubes are centrifuged at 980 *g* for 20 min and the gel is washed four times with wash buffer, once with wash buffer containing 0.5 M NaCl, and with wash buffer again. A small aliquot of the gel suspension is treated with SDS sample buffer and analyzed by SDS/PAGE to estimate the amount of fusion protein adsorbed. The pelleted gel is resuspended in 50 mL of wash buffer containing thrombin (Sigma® T-3010) and incubated by end-to-end rotation for 2 h at room temperature. The amount of thrombin required can vary between 1 and 15 units per milligram of fusion protein, and should be determined for each recombinant protein (Figure 5a and b). After

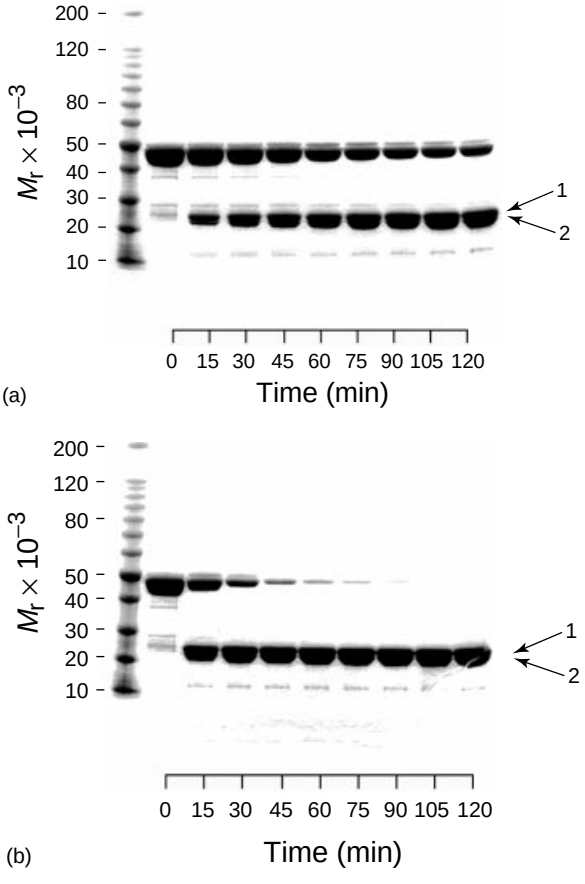

Figure 5 Thrombin cleavage of a GST fusion protein incubated at room temperature with (a) 1 or (b) 3 units of thrombin for the indicated times. The cleaved recombinant protein (1) and the GST affinity tag (2) are indicated. Molecular weight standards are shown on the left.

cleavage, *n*-dodecyl-β-D-maltoside is added to the gel suspension to 0.15% to improve the recovery of the cleaved protein. The suspension is transferred into a column, the effluent is saved and the column is washed with about 50 mL of wash buffer supplemented with 0.15% of dodecyl maltoside. The effluent and the wash fractions are combined and dialyzed for 16 h at 4 °C against 2 L of 25 mM TRIS–HCl, pH 8, containing 5% glycerol, 1 mM DTT and 1 mM EDTA. The dialysate is filtered through a 0.22-μm pore-size membrane (Millex®-GV, Millipore®) and applied to a 25-mL diethylaminoethyl (DEAE)-Sephacel® column (Sigma®). The column is washed with three column volumes of loading buffer to remove the detergents. Bound protein is then eluted by washing the column in the reverse direction with 500 mM sodium phosphate buffer, pH 8, containing 10% glycerol. This chromatographic step removes residual detergents and also concentrates the cleaved recombinant protein and allows its direct application to a gel filtration column for final polishing (see section 2.1.8).

2.1.5.2 Purification of Poly-histidine-tagged Proteins

Immobilized metal chelate affinity chromatography was first introduced in 1975[25] and it has become a widely used technique. Originally the chelating ligand iminodiacetic acid (IDA), charged with Ni^{2+} or Zn^{2+}, was used to purify various peptides and proteins. Since IDA has only three chelating sites, it does not bind metal ions containing six coordination sites with high affinity. More recently, a new chelating adsorbent, nitrilotriacetic acid (NTA), was introduced to overcome this problem.[26] The affinity of hexa-His-tagged proteins to NTA is 1000 times higher than to IDA, which allows its use in strong denaturants such as 6 M guanidine–HCl, 8 M urea or high concentrations (1 M) of chaotropic salt solutions. This permits the use of highly stringent washing conditions to remove nonspecifically bound host proteins. The purification of recombinant hexa-His-tagged proteins under denaturing conditions was described in an earlier publication,[5] and here a nondenaturing purification method for a hexa-His-tagged recombinant bacterial protein is presented.

NTA–agarose resin (Qiagen) is provided derivatized with Ni^{2+}. Nevertheless, we usually go through the following preparation steps: Ni^{2+}–NTA–agarose gel (50 mL) is washed in a column with 0.1 M $NiSO_4$, until $NiSO_4$ appears in the breakthrough. Excess Ni^{2+} is washed out with three column volumes of 0.1 M $NaHPO_4^-$, pH 6, containing 10% glycerol, 300 mM NaCl and 1% Triton X-100®. The column is then equilibrated in loading buffer (0.1 M $NaHPO_4^-$, pH 8, 10% glycerol, 300 mM NaCl, 1% Triton X-100®). The 150 000 *g* supernatant (see section 2.1.2) is made 1% with respect to Triton X-100® and filtered through a 0.22-μm pore-size membrane. The filtrate is mixed with the Ni^{2+}-derivatized resin in a bottle and rotated for 1 h at room temperature. The suspension is transferred into a column and the effluent, representing the unbound fraction, is saved. The gel is washed with five column volumes of loading buffer, followed by five column volumes of loading buffer containing 5 mM β-mercaptoethanol. This washing step usually reduces background due to disulfide cross-linked proteins. Washing is continued with three column volumes of 0.1 M $NaHPO_4^-$, pH 6, containing 10% glycerol, 300 mM NaCl and 1% Triton X-100®, and three column volumes of the same buffer without Triton®. The recombinant protein is then eluted with five column volumes of a linear gradient of 500 mM imidazole–HCl, pH 6, containing 10% glycerol and 300 mM NaCl. The column is regenerated by washing with the elution buffer containing 100 mM EDTA-Na_2, until all the Ni^{2+} is eluted from the column. Washing is continued with 6 M guanidine–HCl, in 0.1 M sodium acetate buffer, pH 4, and the column is stored in 0.1 M NaCl–20% ethanol at 4 °C. The gel can be derivatized with Ni^{2+} as described above and reused repeatedly.

2.1.6 The Final Steps

First we use less specific techniques of centrifugation, nucleic acid removal and protein precipitation. These are followed by high-resolution chromatographic methods such as ion-exchange, affinity and HIC. Gel filtration complements precipitation and the high-resolution methods, and it enables us to determine molecular size in a simple way. It is a very gentle purification method since there is no or little interaction of the protein with the gel matrix, and the whole separation is conducted in a constant buffer system. Gel filtration chromatography is usually used in the later stages as a final polishing step to remove the last traces of impurities, to separate multimer and aggregated forms of the protein from the monomers and to exchange into buffer systems suitable for assays, for crystallization or for formulation.

2.2 Purification of Proteins from Inclusion Bodies

Very often heterologous proteins, expressed in *E. coli*, are precipitated in the cells in the form of highly refractile inclusion bodies. Purifying recombinant proteins from inclusion bodies is similar to the isolation of soluble egg albumin from a hard-boiled egg. The presence of the recombinant protein of interest in the 3000 *g* pellet indicates that it is in inclusion bodies. Inclusion bodies fall into two categories: (1) paracrystalline type and (2) amorphous aggregates. Both contain most of the recombinant protein in a precipitated, denatured form. In addition to the recombinant protein, inclusion bodies also contain coprecipitated host proteins, especially proteins

of the host translation and transcription machinery, and also nucleic acids. Originally it was thought that inclusion body formation was harmful and hindered purification of the protein of interest. In recent years many investigators have found the formation of inclusion bodies to be desirable since they can simplify protein purification. Inclusion bodies are usually washed with weak detergents and chaotropic agents to remove some of the host proteins. This can result in the enrichment of the recombinant protein to 65% or better.

2.2.1 Recovery of Inclusion Bodies

To purify recombinant proteins from inclusion bodies, the 3000 *g* pellet, after opening the cells, is resuspended in 50 mM TRIS–HCl, pH 8, containing 10% glycerol, 2 mM EDTA, 0.1 mM DTT, 200 μg mL^{-1} lysozyme and 2% NP-40 or 1% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), by homogenization. The inclusion bodies are pelleted by centrifugation at 12 000 *g* for 15 min and the extraction is repeated three more times. The pellet is resuspended in 20 mL of extraction buffer as above and layered on top of 50% glycerol (2 × 30 mL) and centrifuged at 12 000 *g* for 30 min. The inclusion body pellet is then washed free of detergent by two cycles of resuspension in 50 mM TRIS–HCl buffer, pH 8, and centrifugation at 12 000 *g* for 15 min.

2.2.2 Solubilization of Inclusion Body Protein

The washed inclusion bodies are then solubilized in 20 mL of 7 M urea in 25 mM TRIS–HCl buffer, pH 8, containing 100 units mL^{-1} Trasylol®, 5 mM ε-aminocaproic acid, 10% glycerol, 1 mM EDTA and 10 mM lysine or 10 mM ethylenediamine. Lysine or ethylenediamine is included whenever urea is used as a scavenger of amine-reactive cyanates originating from the breakdown of urea. Cyanate ions can react with amino groups of proteins to form carbamylated derivatives.[27] The solution is stirred at room temperature for 30 min and centrifuged at 150 000 *g* for 1 h. The supernatant is filtered through a 0.22-μm pore-size membrane and applied to a 50-mL Source™ 30Q column, equilibrated in solubilization buffer. Unbound proteins are washed out with solubilization buffer and bound components are eluted with 10 column volumes of a linear gradient of 0 to 1.0 M NaCl in solubilization buffer. Fractions are collected and location and purity of the recombinant protein are assessed by SDS/PAGE analysis. Depending on the purity of the recombinant protein in question, an additional purification step, for example gel filtration, can be performed or the denatured protein may be refolded by one of the methods described below.

2.3 Refolding of Inclusion Body Proteins

There are many published refolding methods and there is no single method that would be applicable to all proteins. The denaturant used to solubilize inclusion bodies must be removed to allow refolding of the protein. The recovery of activity depends on the formation of correct intramolecular interactions including disulfide bonds. The principal change that is being reversed is not chemical but physico-chemical, involving noncovalent and conformational changes.[28] The refolding methods usually consist of either chaotrope-mediated folding, cosolvent-assisted folding, detergent-mediated folding, chaperone-assisted folding or a combination of these methods. Strict guidelines to refold proteins cannot be given since the requirements are protein specific, but the general consensus is as follows: (1) during refolding the protein concentration must be low, <0.5 mg mL^{-1} or preferably <0.1 mg mL^{-1}, in order to favor intramolecular interactions in preference to intermolecular interactions; (2) expose recombinant proteins to alkaline pH (>9.0) in the presence of thiol reagents, 50 mM DTT or 0.1 M β-mercaptoethanol, and a denaturant, 8 M urea or 6 M guanidine–HCl, to solubilize aggregates; the use of DTT over mercaptoethanol is preferred since it does not form mixed disulfides with proteins, as mercaptoethanol does;[29] (3) slow removal of the denaturant by dilution or preferentially by dialysis at around 10 °C; (4) the inclusion of several additives in the dialysis buffer,[30–32] 0.01% Triton X-100®, 0.4–0.5 M L-arginine, 5 mM EDTA, 25% glycerol, 150–200 mM KCl; and (5) the presence of redox reagents such as reduced (4 mM) and oxidized (0.4 mM) glutathione to promote disulfide shuffling. The number of ways in which cysteine residues can combine to form disulfide bonds increases very rapidly with increase of the number of cysteines (Table 4) and soon reaches astronomical figures. The recombinant protein human tissue plasminogen activator,[33] for example, has 17 disulfide bonds. The 35 cysteine residues present in this protein theoretically can form disulfide bonds 2.2×10^{20} ways. If disulfide bond formation were to occur at random, then statistically one would need 200 000 tons

Table 4 The number of possibilities in which cysteine residues can combine to form disulfide bonds

Number of disulfide bonds	Number of combinations
1	1
2	3
3	15
5	945
10	654 729 075
20	319 830 986 772 877 770 815 625
25	58 435 841 445 947 272 053 455 474 390 625

of inclusion body protein to correctly refold 1 ng of this protein. However, luckily for us, in a real-life situation refolding is biased towards the formation of the correct three-dimensional structure. This is an important piece of information because it means that the three-dimensional functional state of a folding protein is more stable than the unfolded state. This thermodynamic statement underlies all our efforts to refold proteins. Over millions of years nature has selected for those proteins that are most active in their most stable form.

Very rarely a denatured and reduced protein will refold to a soluble and active form by simple dialysis against a physiological buffer to remove the denaturant. A relatively simple protocol was sufficient to refold the recombinant protein, expressed in *E. coli*, and shown in Figure 6(a). After washing of inclusion bodies, solubilization and purification of the recombinant protein on a Source™ 30Q anion-exchange column, the denaturant (7 M urea) was removed by dialysis against 10 mM sodium phosphate buffer, pH 6.8, containing 2 mM dodecyl maltoside. Dialysis was continued against 25 mM HEPES–NaOH, pH 8, containing 50 mM NaCl. The refolded protein was concentrated by dialysis against 4 M ammonium sulfate, centrifuged, the pellet dissolved in 50 mM HEPES–NaOH, pH 7.8, containing 150 mM NaCl. The solution was filtered through a 0.22-μm pore-size membrane and subjected to gel filtration chromatography on a Superdex™ 200 (2.6 × 60 cm) column in the same buffer.

Figure 6(b) shows SDS/PAGE analysis of fractions eluted from this column. The refolded protein was active and chromatographed as a monomer on the gel filtration column. More often, however, we have to ''guide'' an unfolded protein through a series of conformational states to allow refolding. This is usually done by subjecting the denatured and reduced protein to high concentration of labilizing agents such as 0.5 M L-arginine or lysine. At the concentration used these basic amino acids are perturbents and destabilizers of structure. However, they are not strong enough to destabilize correctly folded polypeptides. They allow the reshuffling of the less stable misfolded molecules being trapped in side reactions. In our laboratory we had good experience with the following refolding protocol: the purified inclusion body protein is diluted to 50–100 μg mL^{-1} with the 7 M urea buffer described above for the solubilization of inclusion bodies. The diluted protein is dialyzed against 3 × 2 L of refolding buffer (25 mM glycylglycine–HCl, pH 9.5, containing 0.4 M L-arginine, 0.5 M NaCl, 1 mM EDTA, 25% glycerol, 4 mM reduced glutathione and 0.4 mM oxidized glutathione) for 3 days at 10 °C. Dialysis is continued against 2 × 2 L of 20 mM glycylglycine–NaOH, pH 9, containing 25% glycerol, 150 mM NaCl, 1 mM EDTA, 0.01% Triton X-100® and 50 mM $(NH_4)_2SO_4$

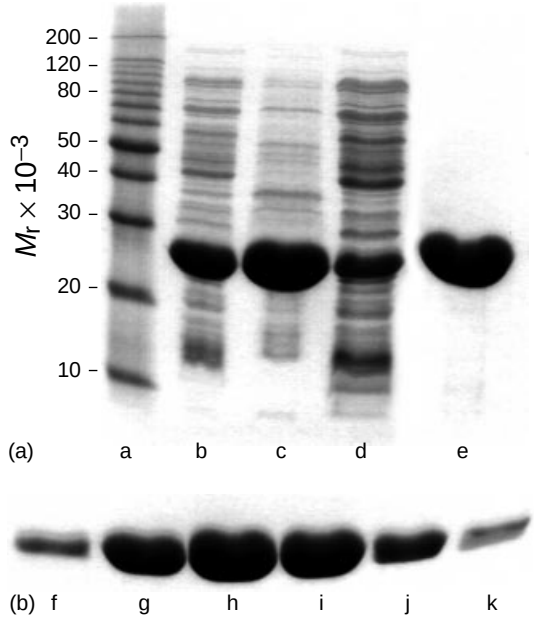

Figure 6 (a) Purification and refolding of a recombinant protein from inclusion bodies. *E. coli* cell lysate (lane b) was subjected to centrifugation at 3000 *g*. The supernatant (lane d) was removed and the pellet (lane c) was solubilized and purified by anion-exchange chromatography (lane e) in the presence of urea. (b) The protein was refolded as described in the text and subjected to gel filtration chromatography. Lanes f–j show SDS/PAGE analysis of fractions eluted from the gel filtration column. Lane a contains molecular weight standards.

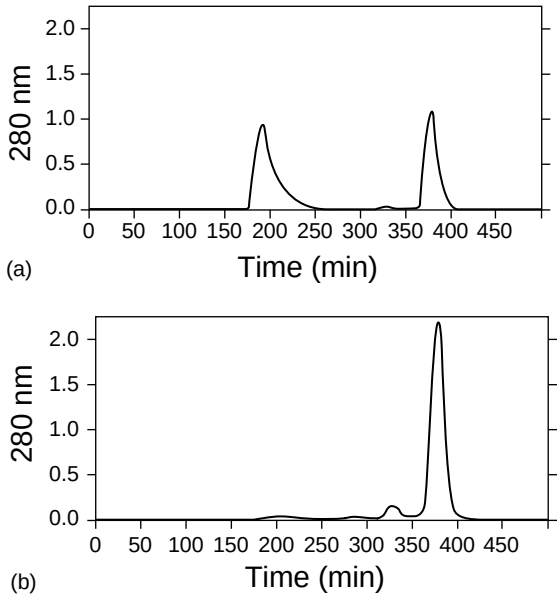

Figure 7 Gel filtration chromatography of a recombinant protein (a) following dialysis of the urea solubilized protein against 20 mM TRIS–HCl, pH 8.3, 50 mM NaCl buffer and (b) following dialysis against the refolding buffers described in section 2.3.

buffer for 2 days at 10 °C. The dialyzate is centrifuged at 48 000 *g* for 30 min and the supernatant is filtered through a 0.22-μm pore-size membrane and concentrated to 1–5 mg mL^{-1} using either a stirred cell (Amicon) or a tangential-flow device (Millipore®). The buffer is usually exchanged for 20 mM glycylglycine–NaOH, pH 9, containing 5% glycerol, 1 mM EDTA and 150 mM NaCl, using a PD-10 column (Pharmacia). Refolding is monitored by activity measurements and by gel filtration chromatography to assess the amount of monomers and multimers in the preparation.

In Figure 7(a) and (b) the refolding of recombinant *Staphylococcus aureus* DNA gyrase B fragment, expressed in *E. coli*, is shown. Dialysis of urea solubilized protein against 20 mM TRIS–HCl, pH 8.3, 50 mM NaCl buffer resulted in the formation of about 50% of aggregated forms of the protein as revealed by gel filtration on a Superdex™ 200 column (Figure 7a). However, subjecting the protein to the refolding protocol described above resulted mainly in monomers (Figure 7b).

3 PURIFICATION OF NATIVE INTRINSIC MEMBRANE PROTEINS FROM MAMMALIAN CELLS

The cell membrane not only functions as a permeability barrier but is also the site of a number of specific enzymes and other proteins that play a role in cell recognition and interaction and in the multitude of specific transport processes. Biological membranes are composed of lipids and proteins to which carbohydrate may be covalently bound. According to the model of Singer and Nicolson,[34] the membrane is fluid or dynamic and the proteins are swimming in or on the viscous lipid bilayer. Lipids, which constitute almost half of the plasma membrane mass, not only serve as an anchoring medium for membrane proteins but also may be responsible for the activity of membrane-associated enzymes. Lipids are amphipathic, having hydrophilic and hydrophobic ends. The majority of the membrane lipids are phospholipids, the rest being made up of glycolipids and neutral lipids, mainly cholesterol. The glycolipids are restricted to the outer half of the lipid bilayer, indicating that an asymmetry exists between the two leaflets of lipid layers. This asymmetry is also evident when we examine the distribution of membrane proteins. Most, if not all, of the intrinsic membrane proteins are glycosylated and have their carbohydrate moieties on the non-cytoplasmic side of the membrane. Proteins that are directly or indirectly associated with the lipid bilayer are frequently differentiated into three groups depending on whether they are (1) artificially adsorbed or are a (2) peripheral (extrinsic) or (3) integral (intrinsic) part of the membrane. Before the purification of intrinsic membrane proteins can be undertaken, they must be extracted from the membrane matrix. For this purpose one might employ urea, EDTA, distilled water, organic solvents, controlled proteolysis, guanidine and other chaotropic agents or manipulation of ionic strength and pH. These methods, however, do not lead to the complete solubilization of proteins that are strongly bound to the lipid matrix of the membrane. For intrinsic membrane proteins, detergents (synthetic detergents, bile salts and saponins) appear to provide the best extraction method. There are many detergents available for the solubilization of membrane proteins.[35] The detergent of choice should be able to disintegrate the membrane and keep the proteins in solution without destroying the protein's antigenic structure, thereby interfering with antigen–antibody interactions. Neutral or nonionic detergents fulfill these requirements. Of the many neutral detergents that have been employed for membrane solubilization, we prefer NP-40 [poly(ethylene glycol) *p*-isooctylphenyl ether] for the solubilization of lymphocyte plasma membranes. NP-40 is a relatively mild, yet highly effective solubilizing agent for membrane proteins.[36] At the concentration used, 0.5%, it reacts mainly with the hydrophobic, lipophilic portions of proteins without interfering with antibody interactions directed at the hydrophilic portions of cell surface antigens. Furthermore, at this concentration NP-40 does not solubilize nuclear membranes, an important consideration when detergent extraction of intact cells is undertaken.

Optimum conditions for solubilization with neutral detergents include the use of low ionic strength buffer (<0.2 M), alkaline pH (8–8.5), a detergent to protein ratio of at least 1.5 and a protein concentration below 3 mg mL^{-1}. Since the degree of solubilization is only slightly influenced by temperature, solubilization should be done at 0 °C to reduce proteolytic degradation of membrane proteins. Enzymatic degradation of the solubilized proteins can be a serious problem and many investigators include a cocktail of protease inhibitors in the solubilization mixture.[37]

For the purification of membrane proteins one may solubilize whole cells with detergents. However, since the plasma membrane proteins represent only about 1% of the total cellular proteins, it is an advantage to prepare membranes first. Purified membranes represent approximately a 50-fold purification of cell surface antigens.

3.1 Preparation of Sub-cellular Fragments

Large-scale purification of plasma membranes is best achieved by nitrogen decompression of cell suspensions, followed by differential centrifugation of the resulting

homogenate.[12] If possible, the use of established cell lines is recommended for the purification of cell surface antigens. Cultured cell lines provide a uniform and continuous source of antigens.

Cells (5×10^8–1×10^9) are washed twice with BB (20 mM HEPES, pH 7.5, containing 0.13 M NaCl, 0.5 mM $MgCl_2$ and 1 mM PMSF) by centrifugation at 800 *g* for 10 min. Cell pellets are resuspended to 10^7 cells mL^{-1} in BB and an equal volume of 0.5 M sucrose is added with constant gentle agitation. The cell suspension is equilibrated with nitrogen gas in a Parr® nitrogen cavitator (Parr® Instruments) for 15 min at 50 atm (5.07 MPa) at 0–4 °C. Cell breakage is achieved by dropwise release of the cell suspension. EDTA is added to a 1 mM final concentration and the homogenate is fractionated by differential centrifugation. Nuclei, unbroken cells and debris are removed by centrifugation at 800 *g* for 15 min. The supernatant is centrifuged at 20 000 *g* for 20 min to pellet mitochondria and lysosomes. Membrane vesicles in the supernatant are then pelleted by centrifugation at 120 000 *g* for 60 min. Pelleted membranes are washed consecutively with 10 and 1 mM HEPES, pH 7.5. For most experiments these membranes may be used without further purification. However, they may be further separated into plasma membrane and endoplasmic reticulum by isopycnic centrifugation through discontinuous Ficoll gradients.[38] The gradient is prepared from 35, 25, 17.5, and 9% Ficoll in 1 mM HEPES, pH 7.5, representing specific gravities of 1.12, 1.09, 1.06 and 1.03 g cm^{-3}, respectively, on top of a 70% sucrose cushion. After centrifugation at 30 000 *g* for 20 h at 4 °C the plasma membrane banding at a density of 1.06 is removed and washed with 20 mM HEPES, pH 7.4, by centrifugation at 140 000 *g* for 60 min.

3.2 Solubilization of Cell Membranes

Cell surface proteins may be directly solubilized at a cell density of 2×10^7 mL^{-1} on ice for 30 min in 1% NP-40 in phosphate-buffered saline (PBS) (0.02 M sodium phosphate, pH 7.8, 0.15 M NaCl, 1 mM $MgCl_2$, 0.5 mM $CaCl_2$) containing 1 mM iodoacetamide, 1 mM PMSF and 10 μg mL^{-1} of each of soybean trypsin inhibitor, antipain, pepstatin, leupeptin and chymostatin (a 100× stock solution of the protease inhibitor cocktail is prepared from 0.1 M PMSF and 1 mg mL^{-1} of each of the other components in 10% methanol in dimethyl sulfoxide (DMSO) and kept frozen at −70 °C). The lysate is cleared of nuclei and debris by centrifugation at 37 000 *g* for 30 min. Some cell surface proteins form aggregates in the presence of neutral detergents.[39] This is prevented by the addition of 1–2% sodium deoxycholate (DOC) to the extract after the nuclei have been removed by centrifugation.

Purified cell membranes (from 10^9 cells) are solubilized in 2 mL of 1% DOC in PBS, containing 1 mM iodoacetamide, 1 mM PMSF and 10 μg mL^{-1} of each of the protease inhibitors listed above, at 0 °C for 30 min with magnetic stirring. The extract is centrifuged at 150 000 *g* for 60 min. If not immediately used, detergent extracts should be stored at −70 °C.

3.3 Immunoaffinity Purification of Membrane Proteins

For the preparation of immunoaffinity columns, antibodies may be immobilized directly to CNBr-activated Sepharose® (Pharmacia) or to activated gels that contain spacer arms, e.g. Affi-Gel® 10 (Bio-Rad), according to the manufacturer's instructions. A spacer arm is used to keep the antibody away from the gel matrix so that the combining site of the antibody molecule is free to bind the antigen. Perhaps the best immunoadsorbent to purify membrane proteins is prepared by oriented coupling of antibodies through their Fc portions to immobilized protein A or protein G resins with the bifunctional cross-linking reagent dimethyl pimelimidate.[40] Protein A–Sepharose® CL-4B (Pharmacia) or protein G–agarose (Oncogene™, No. IP08) is mixed with the antibody solution in 0.1 M borate buffer, pH 8.2, at about 10 mg antibody per milliliter of packed beads. The slurry is rotated at room temperature for 60 min, transferred into a sintered glass filter and washed extensively with 0.1 M borate buffer, pH 8.2, followed by washing with 0.2 M triethanolamine–HCl, pH 8.2. The gel is transferred to a tube with about 25 volumes of 0.2 M triethanolamine, 50 mM dimethyl pimelimidate dihydrochloride (Pierce), pH 8.2. The tube is rotated for 45 min and the gel is centrifuged and treated with a fresh aliquot of cross-linking solution as above for another 45 min. After this incubation time, the gel is centrifuged and resuspended in 25 volumes of blocking buffer (50 mM ethanolamine–HCl, pH 8.2, 0.15 M NaCl), and incubated at room temperature for 30 min. Finally, the gel is washed twice with 0.1 M borate buffer, pH 8.2, and stored at 4 °C in the presence of 0.02% sodium azide.

For the immunoaffinity purification of intrinsic membrane proteins, detergent extracts obtained either from solubilized whole cells or from purified membranes are applied using a recycling system (peristaltic pump) to two successive columns attached in series, containing (1) immobilized nonspecific mouse mAb, and (2) a specific mouse mAb column recognizing the desired antigen. The precolumn is used to remove components from the lysates that bind nonspecifically either to the gel matrix or to the immobilized antibodies.

The specific antibody column is usually prewashed with elution buffer (50 mM diethylamine–HCl, pH 11.5, 1% *n*-octyl-β-D-glucopyranoside (OG), 150 mM NaCl,

1 mM EDTA, 1 mM iodoacetamide) and both columns are equilibrated in wash buffer 1 (50 mM TRIS–HCl, pH 8, 150 mM NaCl, 0.5% NP-40, 0.5% DOC, 10% glycerol), before application of the detergent extracts. After recirculation of the extract through both columns (2–3 times), the precolumn is removed and the specific immunoaffinity column is washed with about 20 column volumes of wash buffer 1. This is followed by a wash with five column volumes of wash buffer 2 (50 mM TRIS–HCl, pH 9, 500 mM NaCl, 0.5% NP-40, 0.5% DOC, 10% glycerol). This high-pH and high-salt wash will elute a lot of nonspecifically adsorbed components. The column is then washed with five column volumes of wash buffer 3 (2 mM TRIS–HCl, pH 8, 1% OG, 10% glycerol). This washing step exchanges the detergents for the dialyzable OG and reduces the ionic strength of the buffer in preparation for elution. Bound antigen is then eluted by washing the column with elution buffer (described above). From a 5-mL immunoaffinity column we usually collect 2-mL fractions into tubes that contain 250 µL of 1 M TRIS–HCl, pH 6.8, 0.15 M NaCl and 1% OG, in order to neutralize the pH. Figure 8 shows the immunoaffinity purification of HLA-DR1 molecules from a human lymphoblastoid cell line.

4 SUMMARY

The biotechnological revolution of the past 20 years has given a strong impetus to interest in methods for the separation and purification of proteins. Purified proteins are needed in several steps during the drug development process: (1) for development of assays; (2) for crystallization of apo- and holo-forms of the protein; (3) for high-throughput screening; and (4) as drugs themselves. Purification protocols are also used for the fractionation and enrichment of low-copy-number proteins in order to facilitate proteome analysis. Recent developments in chromatographic separation methods, in two-dimensional polyacrylamide gel technology and large-scale sample throughput mass spectrometry now allow us to prepare master purification steps.[14,41,42] Complex protein mixtures of entire organisms are being separated by chromatographic methods, the protein content of the resolved fractions are analyzed by two-dimensional polyacrylamide gel electrophoresis and the proteins present are identified by mass spectrometry. Such master schemes will become valuable tools in the isolation of many known or hypothetical proteins.

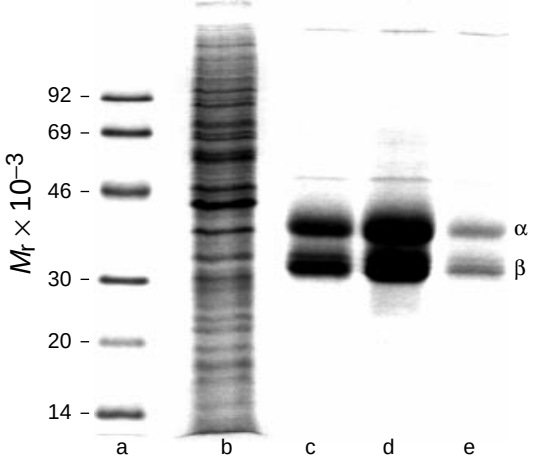

Figure 8 SDS/PAGE analysis of immunoaffinity-purified HLA-DR1 molecules. Detergent extract (lane b), obtained from a DR1-homozygous lymphoblastoid cell line, was applied to a column containing mouse monoclonal anti-HLA-DR antibodies. After washing the column, DR1 molecules (two noncovalently associated α and β chains) were eluted. Three consecutively eluted fractions are shown in lanes c–e. Lane a contains molecular weight standards.

With advances in DNA technology it became possible to express large amounts of virtually any protein in various host cells. In fact, the technology for cloning genes for a magnitude of proteins has developed at a faster rate than the technology for purifying the expressed gene products.

In this article we have covered some basic procedures for the purification of soluble recombinant proteins expressed in *E. coli* and methods to obtain soluble and active protein from insoluble and inactive inclusion bodies. As a first step the protein of interest has to be released into an aqueous extract. Following extraction, residual debris and intact cells are removed by centrifugation. Nucleic acids are either hydrolyzed or precipitated to reduce the viscosity of the extract. A crude enrichment for the protein of interest is made based on its differential solubility in ammonium sulfate. This is then followed by high-resolution chromatographic methods. The type of chromatographic matrices and the order of their use are evaluated in pilot experiments for column selection. Gel filtration chromatography is used as a final polishing step to remove last traces of impurities and to separate multimer and aggregated forms of the protein from the monomers.

In the last section, the purification of a native intrinsic membrane protein, using immobilized mAbs, was also covered. The use of immunoaffinity purification of integral membrane proteins is emphasized since specific mAbs may be obtained even against low-copy-number membrane proteins by reducing a multispecific antibody response to a series of monospecific responses by cloning. The use of immobilized, highly specific mAbs should allow the purification of many minor and weakly antigenic membrane proteins that might be missed by other methods.

The rapid identification of proteins using peptide mass fingerprinting provides us with a continuously expanding number of new potential therapeutic targets. For many of these proteins we do not even have a biological function. In order to understand the molecular mechanisms in which they are involved, these proteins will have to be purified and characterized. The cataloging of the total protein content of various organisms is progressing rapidly. The existence of a functional protein blueprint would be invaluable not only for the diagnosis but possibly also for the treatment of many diseases.

ACKNOWLEDGMENTS

The author acknowledges the invaluable technical assistance and efficient secretarial help of Marie-Françoise Takács and thanks Dr Michael Fountoulakis for reading the manuscript. The author is also indebted to Dr Paul Hadváry, without whose understanding these studies would not have been possible.

ABBREVIATIONS AND ACRONYMS

BB	Breaking Buffer
BSA	Bovine Serum Albumin
CHAPS	3-[(3-cholamidopropyl)dimethyl ammonio]-1-propanesulfonate
DEAE	Diethylaminoethyl
DFP	Diisopropyl Fluorophosphate
DMSO	Dimethyl Sulfoxide
DNase	Deoxyribonuclease
DOC	Sodium Deoxycholate
DTT	D, L-dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
GST	Glutathione *S*-transferase
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid)
HIC	Hydrophobic Interaction Chromatography
HLA-DR	Human Leukocyte Antigen D-related
IDA	Iminodiacetic Acid
mAb	Monoclonal Antibody
MALDI/MS	Matrix-assisted Laser Desorption Ionization/Mass Spectrometry
NP-40	Nonidet P-40
NTA	Nitrilotriacetic Acid
OG	*n*-octyl-β-D-glucopyranoside
PBS	Phosphate-buffered Saline
PEG	Poly(Ethylene Glycol)
PEI	Polyethyleneimine
PMSF	Phenylmethylsulfonyl Fluoride
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate/ Polyacrylamide Gel Electrophoresis
SOD	Superoxide Dismutase
TLCK	*N*-α-*p*-tosyl-L-lysine Chloromethyl Ketone
TPCK	*N*-tosyl-L-phenylalanine Chloromethyl Ketone
TRIS	Tris(hydroxymethyl)aminomethane

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Peptides and Proteins **(Volume 7)**
Chromatography of Membrane Proteins and Lipoproteins • Gel Electrophoresis in Protein and Peptide Analysis • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography

REFERENCES

1. H.B. Vickery, 'The Origin of the Word Protein', *Yale J. Biol. Med.*, **22**, 387–393 (1950).
2. M.H. Marino, 'Expression Systems for Heterologous Protein Production', *BioPharm*, **2**, 18–33 (1989).
3. B.H. Johnson, M.H. Hecht, 'Recombinant Proteins can be Isolated from *E. coli* Cells by Repeated Cycles of Freezing and Thawing', *Bio/Technology*, **12**, 1357–1360 (1994).
4. E.A. Schwinghamer, 'A Method for Improved Lysis of some Gram-negative Bacteria', *FEMS Microbiol. Lett.*, **7**, 157–162 (1980).

5. B.J. Takács, M.-F. Girard, 'Preparation of Clinical Grade Proteins Produced by Recombinant DNA Technologies', *J. Immunol. Methods*, **143**, 231–240 (1991).
6. S. Frankel, R. Sohn, L. Leinwand, 'The Use of Sarkosyl in Generating Soluble Protein after Bacterial Expression', *Proc. Natl. Acad. Sci. USA*, **88**, 1192–1196 (1991).
7. P.T. Rajagopalan, D. Pei, 'Oxygen-mediated Inactivation of Peptide Deformylase', *J. Biol. Chem.*, **273**(35), 22 305–22 310 (1998).
8. T.C. Thomas, M.G. McNamee, 'Purification of Membrane Proteins', *Methods Enzymol.*, **182**, 499–520 (1990).
9. J. Jendrisak, 'The Use of Polyethyleneimine in Protein Purification', in *Protein Purification: Micro to Macro, UCLA Symposia on Molecular and Cellular Biology*, ed. R. Burgess, Alan R. Liss, New York, 75–97, Vol. 68, 1987.
10. M. Hassel, G. Klenk, M. Frohme, 'Prevention of Unwanted Proteolysis During Extraction of Proteins from Protease-rich Tissue', *Anal. Biochem.*, **242**, 274–275 (1996).
11. D.T. Liu, N. Goldman, F. Gates, III, 'Consideration of the Proteins and Peptides Produced by New Technology for Use as Therapeutics', in *Delivery Systems for Peptide Drugs, NATO ASI Series, Series A: Life Sciences*, eds. S.S. Davis, L. Illum, E. Tomlinson, Plenum Press, New York, 341–349, Vol. 125, 1986.
12. E. Ferber, K. Resch, D.F.H. Wallach, W. Imm, 'Isolation and Characterization of Lymphocyte Plasma Membranes', *Biochim. Biophys. Acta*, **266**, 494–504 (1972).
13. B.J. Takács, C. Staehli, 'Activated Macrophages and Antibodies Against the Plant Lectin, GSI-B_4, Recognize the Same Tumor-associated Structures (TAS)', *J. Immunol.*, **138**(6), 1999–2007 (1987).
14. M. Fountoulakis, M.-F. Takács, B.J. Takács, 'Enrichment of Low-copy-number Gene Products by Hydrophobic Interaction Chromatography', *J. Chromatogr. A*, **833**, 157–168 (1999).
15. A. Montoya, J.V. Castell, 'Long-term Storage by Peroxidase-labelled Immunoglobulins for Use in Enzyme Immunoassay', *J. Immunol. Methods*, **99**, 13–20 (1987).
16. S. England, S. Seifter, 'Precipitation Techniques', *Methods Enzymol.*, **182**, 285–300 (1990).
17. E. Stellwagen, 'Chromatography on Immobilized Reactive Dyes', *Methods Enzymol.*, **182**, 343–357 (1990).
18. B. Nilsson, L. Abrahmsén, M. Uhlén, 'Immobilization and Purification of Enzymes with Staphylococcal Protein A Gene Fusion Vectors', *EMBO J.*, **4**(4), 1075–1080 (1985).
19. M. Uhlén, T. Moks, 'Gene Fusions for Purpose of Expression: an Introduction', *Methods Enzymol.*, **185**, 129–143 (1990).
20. T.P. Hopp, K.S. Prickett, V.L. Price, R.T. Libby, C.J. March, D.P. Cerretti, D.L. Urdal, P.J. Conlon, 'A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification', *Bio/Technology*, **6**, 1204–1210 (1988).
21. R. Axén, H. Drevin, J. Carlsson, 'Preparation of Modified Agarose Gels Containing Thiol Groups', *Acta Chem. Scand., Ser. B*, **29**, 471–474 (1975).
22. H.M. Sassenfeld, S.J. Brewer, 'A Polypeptide Fusion Designed for the Purification of Recombinant Proteins', *Bio/Technology*, **2**, 76–81 (1984).
23. J.E. Cronan, Jr, 'Biotination of Proteins *In Vivo*', *J. Biol. Chem.*, **265**(18), 10 327–10 333 (1990).
24. D.P. Gearing, N.A. Nicola, D. Metcalf, S. Foote, T.A. Willson, N.M. Gough, R.L. Williams, 'Production of Leukemia Inhibitory Factor in *Escherichia coli* by a Novel Procedure and its Use in Maintaining Embryonic Stem Cells in Culture', *Bio/Technology*, **7**, 1157–1161 (1989).
25. J. Porath, J. Carlsson, I. Olsson, G. Belfrage, 'Metal Chelate Affinity Chromatography, a New Approach to Protein Fractionation', *Nature*, **258**, 598–599 (1975).
26. E. Hochuli, H. Döbeli, A. Schacher, 'New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighbouring Histidine Residues', *J. Chromatogr.*, **411**, 177–184 (1987).
27. F.A.O. Marston, D.L. Hartley, 'Solubilization of Protein Aggregates', *Methods Enzymol.*, **182**, 264–276 (1990).
28. R. Jaenicke, R. Rudolph, 'Folding Proteins', in *Protein Structure, a Practical Approach*, ed. T.E. Creighton, IRL Press, Oxford, 191–223, 1989.
29. W.W. Cleland, 'Dithiothreitol, a New Protective Reagent for SH Groups', *Biochemistry*, **3**(4), 480–482 (1964).
30. D.S. Thorpe, S. Niu, E. Morkin, 'Refolding Parameters for the Allosteric Homodimeric Guanylyl Cyclase Catalytic Core from the Atrial Natriuretic Peptide Receptor', *Biochem. Biophys. Res. Commun.*, **224**, 765–771 (1996).
31. J. Buchner, R. Rudolph, 'Renaturation, Purification and Characterization of Recombinant Fab Fragments Produced in *Escherichia coli*', *Bio/Technology*, **9**, 157–162 (1991).
32. J. Buchner, I. Pastan, U. Brinkmann, 'A Method for Increasing the Yield of Properly Folded Recombinant Fusion Proteins: Single-chain Immunotoxins from Renaturation of Bacterial Inclusion Bodies', *Anal. Biochem.*, **205**, 263–270 (1992).
33. D. Pennica, W.E. Holmes, W.J. Kohr, R.N. Harkins, G.A. Vehar, C.A. Ward, W.F. Bennett, E. Yelverton, P.H. Seeburg, H.L. Heyneker, D.V. Goeddel, D. Collen, 'Cloning and Expression of Human Tissue-type Plasminogen Activator cDNA in *E. coli*', *Nature*, **301**, 214–221 (1983).
34. S.J. Singer, G.L. Nicolson, 'The Fluid Mosaic Model of the Structure of Cell Membranes', *Science*, **175**, 720–731 (1972).
35. A. Helenius, K. Simons, 'Solubilization of Membranes by Detergents', *Biochim. Biophys. Acta*, **415**, 29–79 (1975).
36. B.D. Schwartz, S.G. Nathenson, 'Isolation of H-2 Alloantigens Solubilized by the Detergent NP-40', *J. Immunol.*, **107**(5), 1363–1367 (1971).

37. B. Axelsson, A. Kimura, S. Hammarström, H. Wigzell, K. Nilsson, H. Mellstedt, '*Helix pomatia* A Hemagglutinin: Selectivity of Binding to Lymphocyte Surface Glycoproteins on T Cells and Certain B Cells', *Eur. J. Immunol.*, **8**, 757–764 (1978).
38. B.J. Takács, T. Staehelin, 'Biochemical Characterization of Cell Surface Antigens Using Monoclonal Antibodies', in *Immunological Methods*, eds. I. Lefkovits, B. Pernis, Academic Press, New York, 27–56, Vol. 2, 1981.
39. R. Standring, W.R. McMaster, C.A. Sunderland, A.F. Williams, 'The Predominant Heavily Glycosylated Glycoproteins at the Surface of Rat Lymphoid Cells are Differentiation Antigens', *Eur. J. Immunol.*, **8**, 832–839 (1978).
40. C. Schneider, R.A. Newman, D.R. Sutherland, U. Asser, M.F. Greaves, 'A One-step Purification of Membrane Proteins Using a High Efficiency Immunomatrix', *J. Biol. Chem.*, **257**(18), 10 766–10 769 (1982).
41. M. Fountoulakis, B. Takács, 'Design of Protein Purification Pathways: Application to the Proteome of *Haemophilus influenzae* Using Heparin Chromatography', *Protein Expression Purif.*, **14**, 113–119 (1998).
42. M. Fountoulakis, H. Langen, C. Gray, B. Takács, 'Enrichment and Purification of Proteins of *Haemophilus influenzae* by Chromatofocusing', *J. Chromatogr. A*, **806**(2), 279–291 (1998).

Protein–Drug Interactions

Jana Oravcová
GlaxoWellcome Export Ltd., Bratislava, Slovak Republic

Wolfgang Lindner
University of Vienna, Vienna, Austria

1 Introduction **1**
2 Most Relevant Characteristics of Drug–Protein Binding Assays **3**
2.1 Binding Stoichiometry 3
2.2 Binding Specificity 3
2.3 Stereochemical Aspects 3
2.4 Drug–Drug Binding Interactions 4
3 Methodology of Binding Experiments **5**
3.1 Conventional Methods 5
3.2 Chromatographic Methods 6
3.3 Capillary Electrophoresis 8
3.4 Affinity Information on Biospecific Interaction with Respect to Immobilization of Interacting Species 13
3.5 Spectroscopic Methods 14
3.6 Other Methods and Future Trends in Analytical Methodology 15
4 In Vivo Monitoring of Free Fraction of Drugs: Microdialysis **15**
5 Evaluation of Binding Data **15**
6 Conclusion **17**
Abbreviations and Acronyms **18**
Related Articles **18**
References **18**

Noncovalent (reversible) drug–protein binding interactions are an integral part of intermolecular recognition processes taking place in a cell or organ environment. Characterization of basic constraints of target protein binding site(s) (affinity, stoichiometry, specificity, saturability and stereoselectivity) for particular ligand(s) is an essential output of many different drug–protein binding studies. Nonlinear and stereoselective binding phenomena, including drug–drug (or enantiomer–enantiomer) interactions and their e.g. genetically or pathophysiologically determined variability, can be predicted in this way. In general, the methods used for drug–protein binding studies can be divided into separation (nonspectroscopic) and spectroscopic methods. Separation methods are particularly suitable for the quantitative determination of drug binding, whereas spectroscopic techniques are able to characterize the qualitative features of drug–protein complexes (e.g. conformational status/changes). The primary objective was to summarize the most relevant trends in separation techniques [affinity high-performance liquid chromatography (HPLC) and affinity capillary electrophoresis (ACE)] employed in the framework of drug–protein/biopolymer binding assays and to compare them with so-called conventional approaches (equilibrium dialysis (ED), ultrafiltration). Focused attention has been paid particularly to HPLC and capillary electrophoresis (CE) methods designed for drug–protein studies when so-called ''dynamic'' and ''secondary'' equilibria are involved and created. A common feature of this strategy is to use binding protein(s) or interacting drug(s) without their covalent immobilization as mobile phase or background electrolyte (BGE) (buffer) additives. This offered a new, elegant tool for the separation of ''bound'' and ''free'' forms of interacting species in a ''miniaturized free solution system'' operating under nearly physiological conditions. Undoubtedly, it permits simplification of the experimental protocols where it is crucial to perform repetitive analyses under ''the same starting conditions'', as e.g. the occupancy of binding sites is concerned.

1 INTRODUCTION

Drug–protein interactions are an integral part of intermolecular recognition and/or mediating processes taking place in cells or tissues of a living organism, including transport phenomena through membranes. In the recent decades, implementation of advanced methodological tools has enabled a competent and comprehensive insight to be obtained into various relevant aspects and operating mechanisms of drug binding to different functional, transport or depot proteins. Undoubtedly, there are many similarities encountered in the mechanisms of drug binding to proteins compared with protein–protein, protein–carbohydrate, protein–DNA, drug–DNA, enzyme–protein (peptide), etc., binding interactions. However, since most of them are primary subjects of other articles here, we shall just highlight the most relevant trends, particularly using examples on drug attachment to receptor and nonreceptor proteins. The availability of sophisticated analytical instrumentation and techniques facilitated the recognition of the complexity of interlocked biomolecular signaling networks after a drug/xenobiotic has entered the biological system and the resulting pharmacodynamic (toxicological) or pharmacokinetic impact in vivo.

Molecular topology of the most common ''protein drug acceptors'' (Table 1) suggests that interacting proteins

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Examples of pharmacologically relevant drug–protein interactions and molecular localization of drug acceptors

Acceptor protein structure	Localization	Interacting drugs/ligands	Selected references
DNA[a]	Intracellular	Antitumor antibiotics (vancomycin group)	1
P-glycoprotein[a]	Membrane	Verapamil, dihydropyridines, cyclosporin A, etc.	2–4
Calmodulin[a]	Membrane	Ca^{2+}, Ca-antagonists, neuroleptics, benzodiazepines, etc.	5–7
Carbonic anhydrase[a]	Erythrocytes	Carbonic anhydrase inhibitors, sulfonamides	8–10
Hemoglobin	Erythrocytes	Melatonin	11
Caeruloplasmin	Plasma	CNS-active drugs	
Albumin[a]	Plasma	Mainly acidic drugs and endogenous substances such as FFA, steroids, thyroxine, etc.	12–15
AGP[a]	Plasma	Mainly basic drugs, progesterone, etc.	16–18
α_1-Fetoprotein	Plasma of fetus or newborn	Estradiol, bilirubin, retinoids	19
Lipoproteins	Plasma	Lipophilic drugs, e.g. dihydropyridines, cyclosporin A, etc.	20–23
Immunophilins	Intracellular	cyclosporin A, rapamycin	24

[a] Stereoselective binding of drugs has been considered and/or reported.
FFA, free fatty acids; AGP, α_1-acid glycoprotein.

are usually integrated in, for example, intra- or extracellular surfaces of membranes, building functional units of ''true'' receptors or ''receptor-like'' structures, ion channels or pumps. Only to a lesser degree are they available in matrix-soluble form (such as carbonic anhydrase in erythrocytes[8–10] or well-known plasma drug transporters, albumin[25,26] and AGP[16]). There is clear evidence that at least some of the commonly applied purification or solubilization steps may cause subtle structural and/or steric changes of the protein binding site(s), resulting in profound, including stereoselective, differences in binding profiles of interacting drugs. Therefore, detailed knowledge of the conformational status and adaptability of ''drug acceptors'', as well as preserving of integrity of their molecular environment (e.g. membrane, lipid or coupling proteins), is a critical point for the determination of ''true'', molecularly plausible, binding affinities.

The essential output of drug–protein binding studies is to characterize at least some of the principal functional criteria of protein binding sites, i.e.:[27]

- affinity;
- stoichiometry (i.e. monovalent vs multivalent binding);

Table 2 The most common analytical methodologies adopted for evaluation of drug–protein interactions

Method	Calculation of binding parameters (n, K_a)		Screening purposes	Direct assay of stereoselective binding	Qualitative information	Microscale set-up	Complex biological samples
	Monovalent binding	Multivalent binding					
ED	✓	(✓)	✓	–	–	–	✓
Ultrafiltration	✓	(✓)	✓	–	–	(✓)	✓
Ultracentrifugation	✓	–	–	–	–	–	✓
HPLC:							
PSPs	(✓)	–	✓	✓	–	–	–
HD	✓	✓	–	–	–	–	–
FA	✓	✓	–	✓	–	(✓)	✓
CE:							
ACE	✓	–	✓	✓	–	✓	(✓)
VACE	✓	✓	✓	–	–	✓	(✓)
HD	✓	✓	–	–	–	✓	–
FA	✓	✓	–	✓	–	✓	(✓)
CEC	✓	–	✓	✓	–	✓	–
Spectroscopic methods	(✓)	–	–	–	✓	–	–
Microdialysis	(✓)	–	(✓)	✓	–	✓	✓

PSP, protein stationary phase; HD, Hummel–Dreyer; FA, frontal analysis; CEC, capillary electrochromatography; VACE, vacancy affinity capillary electrophoresis. ✓ Parameter could be determined by some methods given in this table; (✓), parameter could be determined with some exceptions or small restricions; – , parameter could not be determined at all.

- (species or tissue) specificity;
- saturability [i.e. presence of eventual nonlinear binding phenomena;[28,29]]
- stereoselectivity.

More specifically, relevant information on drug–drug (or enantiomer–enantiomer) interactions[30–34] and their, for example, genetically[17,35] or pathophysiologically[16,36–38] determined variability has to be elucidated frequently. End-points of many binding studies include understanding of the complex biological impact of drug–protein or protein–protein interactions at the cellular or subcellular level (i.e. interlinking, ''on–off'' or ''tuning'' effect).

All of the above-mentioned aspects have to be reflected by using an appropriate analytical/experimental methodology providing an objective and reliable picture of the situation in vivo or in vitro. Table 2 summarizes the fulfillment of the most relevant criteria when different approaches are adapted for evaluation of the various aspects of binding interactions. The following sections of this article outline on a comparative basis the balance between potential advantages and limitations of different methods and their combinatory potential and/or complementary features.

2 MOST RELEVANT CHARACTERISTICS OF DRUG–PROTEIN BINDING ASSAYS

2.1 Binding Stoichiometry

Information on the stoichiometry of biomolecular interactions not only implicates their specificity, but also enables one to gain an insight into the network of interlinked molecular mechanisms inevitable for a living organism [regulation of programmed cell death, polyclonal humoral immune response or bacterial resistance.[39,40]] This holds true particularly for multivalent recognition, stemming from association of multiple complementary ligands with multiple binding sites of receptors, acceptors or their dimers/oligomers. The situation, which has to be analyzed in case of relatively transparent bivalent binding, is depicted schematically in Figure 1. The most common problems concerning the derivation of binding data are further outlined in section 5. Taking into account possible interfering factors (e.g. generation of intermediates in the course of tight vs weak binding attachments), the methodological approach used has to fulfill these specific demands or should be refined appropriately.[39–42]

2.2 Binding Specificity

The specificity of drug–protein binding interactions is very variable, ranging from ''multidrug transporters''

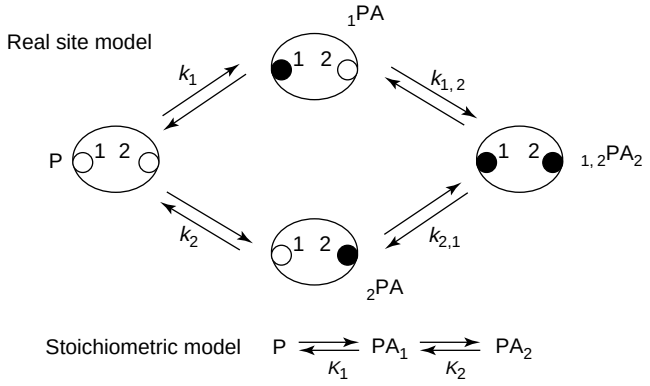

Figure 1 Schematic diagram of a ''two-site receptor'' or ''acceptor'' (P) system interacting with analyte (A) leading to protein–analyte (PA) associates. Visualization of ''stoichiometric'' and ''site binding'' constants. (Reproduced with permission from Klotz.)

such as P-glycoprotein[2–4,43] or albumin[12,25,44] to highly specialized drug carriers such as cyclophilin engaged in cyclosporin A transport[24] or β-tubulin for *Vinca* alkaloid binding[45] (further examples are listed in Table 1). Hence the characteristic spectrum of drug affinities for target functional proteins is capable of mediating, more or less selectively addressed, many recognition and transduction processes on a molecular level simultaneously. This implicates direct modulation of drug efficacy and toxicity in target tissues.

Drug binding properties may differ considerably among the species[46–49] [interspecies differences have been described in detail particularly for drug binding sites of albumins (Figure 2)[50]], and any extrapolation of preclinical drug binding data should be done carefully. Similarly, pathophysiologically, genetically or artificially (in vitro) induced posttranslational modifications of proteins [e.g. tyrosylation, acetylation, glutamylation, phosphorylation of tubulin,[45] glysosylation of albumin[51,52] or carbohydrate content of AGP[53]] may alter the surface charges or hydrophobicity of the protein subunits, leading to profound changes in drug binding affinity and complicating the prediction of drug–drug interactions beyond those seen under normal conditions (i.e. in a healthy population).

2.3 Stereochemical Aspects

Chiral recognition properties of protein targets, distinguishing the three-dimensional conformation of an attaching pharmacophore, are responsible for quantitative or qualitative differences in affinity and/or efficacy of individual drug enantiomers.[54–56] Thereby, stereospecificity is restricted to

- ''true receptor sites'', including ion channels and other ''receptor-like'' structures;

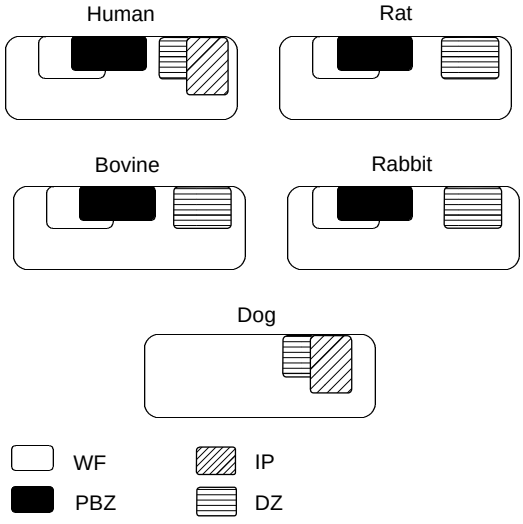

Figure 2 Interspecies binding differences: binding model proposed for drug binding sites on different albumins. Model binding drugs as depicted are warfarin (WF), phenylbutazone (PBZ), ibuprofen (IP) and diazepam (DZ), where WF and PBZ are markers of site I and IP and DZ are markers of site II. Human albumin site I appears to exist for rabbit and rat albumins, but not bovine albumin, and especially dog albumin. With respect to site II, both the binding constants and characteristics of IP and DZ are to dog albumin and are similar to human albumin, whereas the binding in other species is different, particularly the binding characteristics of DZ. (Reproduced with permission from Kosa et al.)

- ''silent receptors'' or ''acceptors'', differentiating enantiomers without observing any obvious pharmacological response (Table 1).

Only in a limited number of cases are the enantiomeric differences for a given drug similar for ''receptor'' and ''acceptor'' sites, suggesting that both binding targets may exhibit some degree of structural homology. Typically, the stereoselectivity is much more pronounced in the case of ''receptor'' than ''acceptor'' sites. As a rule, it increases with the affinity displayed towards the more tightly bound and pharmacologically more active enantiomer (eutomer) of a stereoisomeric pair.[57] Apart from the ''conventional receptor concept'', of particular pharmacodynamic interest is the interaction of chiral drugs with ion channels.[58–66] The properties and access to the drug-binding site at channels termed ''moving (dualistic) targets'' may vary according to their different functional states (resting, open or inactivated) represented molecularly by different, dynamically changing conformations.[59–61,67] Consequently, binding sites must be structurally rearranged as the channel changes its conformation. It therefore seems very likely that open and closed ion channels may have different (state-dependent) binding affinity as well as different stereoselectivity toward interacting drugs.[63]

From a pharmacokinetic point of view, intensive attention has been paid to stereoselective affinity differences of drug towards transport proteins acting as drug acceptors, transporters and storage targets in

- plasma (albumin and AGP);[68,69]
- cell membranes with overexpression of specific proteins (P-gp);[2]
- other compartments (e.g. isoenzymes of carbonic anhydrase in erythrocytes[8–10]).

Of particular importance also are the chiral discriminative properties of cytochrome P450 isoenzymes,[70–72] mediating stereoselective metabolic pathways. The enantiomeric ratios observed for transporters such as albumin or AGP rarely exceed a factor of 2 under physiological conditions. However, if saturation of binding sites can be approached or significant disease-induced changes in protein concentration are occurring, this situation could reach clinical significance.[73] Of course, there are a number of further interfering factors that have to be taken into account for a final interpretation of the binding results and their overall pharmacological impact in vivo. Stereoselective interspecies binding differences, enantiomer–enantiomer (enantiomer–metabolite) interactions at binding sites, stereoselective covalent binding of metabolites, metabolic chiral inversion or racemization of the ligand or stereoselective cooperative binding of drugs represent the most important examples.[9,30,49,74–77]

2.4 Drug–Drug Binding Interactions

It is commonly known that competitive, displacement interactions at the same binding site[30,31,36] as well as allosteric (non-primary site)[33,78–80] drug–drug or enantiomer–enantiomer interactions may be revealed by testing for mutual interactions in binding assays. Allosteric influences may change the conformation of a receptor even if it is at some distance from the gating part of the channel. However, the conventional concept (i.e. Scatchard binding model) is no longer valid when allosteric binding phenomena are involved in the binding process. In some cases, competitive and allosteric binding interactions may undergo transformation into each other depending on the competitor's concentration.[33] Besides the more or less traditional views on drug–drug (enantiomer–enantiomer) displacement models, also cooperativity of binding sites[81] or stereoselective ''site-to-site'' displacement of drug enantiomers[34] have been considered. These findings underline further the importance of considering

the (enantioselective) affinity differences when an objective picture of multidrug–protein interactions has to be created.

3 METHODOLOGY OF BINDING EXPERIMENTS

3.1 Conventional Methods

Membrane separation methods, represented particularly by ED and ultrafiltration, have become very popular for the evaluation of drug binding to proteins owing to their simplicity and general applicability to many different systems in vitro and ex vivo. ED is based on the establishment of an equilibrium state between a ''protein'' compartment and ''buffer'' compartment which are separated by a membrane permeable only to low-molecular-weight components (i.e. drugs). On the other hand, ultrafiltration with semipermeable membranes produces a separation of the free/unbound drug by employing a chemical potential and pressure gradient, respectively, which forces the small drug molecules through the membrane. The possible interfering factors and potential shortcomings of ED and ultrafiltration as compared with ultracentrifugation, another conventional separation method, are summarized in Table 3. Although there is no declared ''standard method'' for the validation of binding data, ED is often regarded as the ''reference method'' for the determination of a drug–protein binding profile. In spite of this, ED suffers from a variety of problems, including a relatively long time to reach equilibrium, volume shifts, Donnan effects, hindering the passage of free ligand, nonspecific adsorption to dialysis apparatus (particularly important for lipophilic drugs and/or proteins), presence of radiochemical impurities and radiochemical instability when a radiolabeled drug is used, protein leakage into the dialyzate and difficulty in the control of some experimental variables (e.g. pH of the dialyzate).[82–89] The major controversy with ultrafiltration involves the stability of the binding equilibrium during the separation process. It is therefore advisable to verify, especially in the case of low-affinity interactions, the basic assumption that the binding ratio of protein bound to free drug remains constant. Ultracentrifugation is an alternative to both ED and ultrafiltration, since it eliminates the problems associated with membrane effects and permits the separation of the free and protein-bound fractions in a ''natural environment'', i.e. without addition of buffer systems and, therefore, dilution problems. However, comparative binding studies with different drugs have revealed that there are quantitative discrepancies between results obtained by ED (or ultrafiltration) and by ultracentrifugation.[82,90–93] As was pointed out, the error in the estimation of the free drug concentration can be influenced by physical phenomena such as sedimentation, back-diffusion, viscosity and binding to plasma lipoproteins in the supernatant fluid. The possible floating of lipoprotein fractions interfered mainly with

Table 3 Comparison of conventional separation methods for drug–protein binding studies

Interfering factor	ED	Ultrafiltration	Ultracentrifugation
Equilibrium time	Can be long (up to 20 h)	Short (10–15 min)	Usually long (12–15 h)
Amount of sample	Usually 500–1000 μL	Small (<1 mL), depending on the K_a of the interaction	Usually >1 mL
Temperature control	Yes	Usually not possible	Usually not possible
pH control	Necessary	Not necessary	Usually necessary
Donnan effects	Yes	Yes	No
Adsorption on membranes and apparatus	Yes	Yes	No
Dilution of the sample: volume shifts	Yes	No	No
Membrane permselectivity	Presence of ''sieve'' effects	Usually satisfactory	–
Other shortcomings of the method	Not suitable for drugs undergoing hydrolysis in serum. Changes in stability of ligand and/or protein may occur	Stability of the free fraction (i.e. the binding ratio of protein-bound drug unbound protein) may change during the separation process according to the K_a of the interaction. Optional leakage of drug or protein	Binding equilibrium may be altered during separation process (sedimentation, back-diffusion). Expensive equipment

the determination of the binding characteristics of basic drugs.[92]

3.2 Chromatographic Methods

There are two most important approaches in chromatographic methods which have been introduced as a powerful tool in drug–protein binding studies:

- affinity (bio)chromatography based on immobilized proteins (or enzymes) in form of a PSP;[94–97]
- HPLC methods operating without chemical (covalent) immobilization of proteins, creating so-called ''dynamic equilibria''.

The selection of a particular method is dependent on the binding strength of the drug–protein complex and on the amount and the purity of protein fractions which are injected into the system. Since only relatively few proteins are available in highly purified form and in microgram amounts, this becomes a limiting factor for the wider expansion of so-called conventional HPLC techniques in drug–protein binding screening studies. Major attention has been drawn to miniaturization of the experimental set-up leading to microscale techniques (and this trend has been successfully explored in CE). In addition, in order to establish physiologically relevant correlations (prediction of drug uptake through epithelial cell membrane), immobilized-(proteo)liposome chromatography[95,98,99] and immobilized artificial membrane (IAM) HPLC[96] have been introduced. Focused attention should be drawn to conformational changes of immobilized proteins (also if the pH is adjusted or an organic modifier is introduced), since binding behavior and thus derivation of the binding data/binding model could be affected significantly in this way.

3.2.1 Affinity Biochromatography: Protein Stationary Phase

Proteins (receptor proteins) as immobilized (chiral) selectors in HPLC have been used for bioaffinity screening purposes in order to select active ligands/drugs. This is also the subject of screening for combinatorial libraries (for e.g. phytochemicals). The advantages of this approach are represented particularly by the following points:[94,100,101]

- relative stability and constant binding behavior of immobilized proteins (PSPs);
- precision, reproducibility and time efficiency of the chromatographic system permitting large sets of comparative and screening binding experiments;
- effectiveness of the PSPs achieved by changing the protein conformation through modifications of the mobile phase (pH, presence of organic modifier or ''additives'' of particular interest, e.g. metal ions such as Ca^{2+}, Mg^{2+} or Zn^{2+}) and using small amounts of drugs (also in the form of their enantiomers) as well as different competitors in order to study drug–drug or enantiomer–enantiomer interactions;
- ability to perform (enantioselective) studies even if there is a tendency towards instability of drugs (e.g. if the enantiomers undergo racemization in aqueous media).

Apart from some inconsistencies regarding results obtained with human serum albumin (HSA) versus bovine serum albumin (BSA), studies with, for example, 1,4-benzodiazepine derivatives,[101–104] 2-arylpropionic nonsteroidal anti-inflammatory drugs (NSAIDs),[105] coumarin derivatives,[79,100,106] *Vinca* alkaloids[107] or triazole derivatives[100] have confirmed that HSA/PSP is able to reflect accurately the binding behavior of ''non-immobilized'' HSA, including its inherent enantioselectivity. Thus, HSA/PSP has been successfully used for the extraction of quantitative binding parameters and a qualitative probe for the evaluation of drug–drug interactions (noncooperative, cooperative and independent binding). In contrast to HSA, immobilized AGP seems not to be a suitable screening tool for molecular mimicking of the binding properties of AGP in solution. This is most probably caused by the immobilization procedure leading to ''reorientation'' or ''cross-linking'' of specific protein functional groups (particularly sialic acid residues), which are co-responsible for high-affinity attachment of (cationic) drugs.

Owing to their complexity, many specific points concerning the recognition abilities and operating mechanisms of PSPs remain to be clarified. However, assuming that the immobilization procedure does not influence the binding properties of the protein, this experimental approach may provide useful information on (a) binding affinity (and eventually also on number of binding sites involved) and (b) binding area(s) where the specific interaction takes place. This has been transparently done, e.g. when interspecies differences for drug interaction with albumin have been addressed.[94,97]

3.2.2 High-performance Liquid Chromatography Methods Using ''Secondary and Dynamic Equilibria''

Several HPLC modifications of size-exclusion chromatographic techniques for binding interactions have been described and compared based on their efficiency to extract reasonable binding characteristics:[108]

- HD method;[109–112]
- vacancy peak (VP) method;[113]
- FA;[114–120]
- zonal elution;[121]
- retention analysis.[122–124]

The basic principle and methodological differences between HD, VP and FA methods, as these methods are the most commonly used, are illustrated in Figure 3. HPLC methods operating on the size-exclusion principle were further optimized by introducing the internal surface reversed-phase (ISRP)-type ''restricted access'' materials,[125,126] combining the principles of size-exclusion and bonded-phase partitioning separations. Applying this approach under equilibrium conditions, (free) drug molecules are able to enter the pores of the ISRP column and partition on the internal surface of pores contrarily to large protein (and protein-associated drug) molecules, eluting in the column interstitial void volume.

3.2.2.1 Hummel and Dreyer Method In the HD method, the column is equilibrated with a mobile phase containing a given concentration of drug in phosphate buffer. When the drug–protein complex is injected, equilibrium between the ligand and protein is rapidly re-established in accordance with the free ligand concentration in the mobile phase. The protein and drug–protein complex elute in the column interstitial void volume and the ligand migrates according to its pore volume penetration and distribution in the stationary phase. For the quantification of bound drug, internal or external calibration procedures have been established and compared.[109,110,112] If a column meets the requirements of nonadsorption but size exclusion of proteins with concurrent chromatographic partitioning of the ligand (e.g. diol or ISRP columns), the bound concentration can be measured directly, while the free ligand concentration is controlled in the mobile phase as the true independent variable. Obviously, this method provides an efficient means of determining accurate binding parameters for protein–ligand equilibria:

- occurring at more than one type of binding site (concomitant presence of high- and low-affinity binding sites);
- if multidrug–protein interactions are to be taken into account;[111]
- in the case of highly protein-bound hydrophobic drugs;
- when binding differences of individual drug enantiomers have to be compared precisely.[127–129]

The most relevant disadvantage of this method is related to the restriction to use only purified protein (or biopolymer) samples. It remains to be clarified whether the downsizing of the analytical set-up [e.g. using CEC] could be adopted in case of this method.

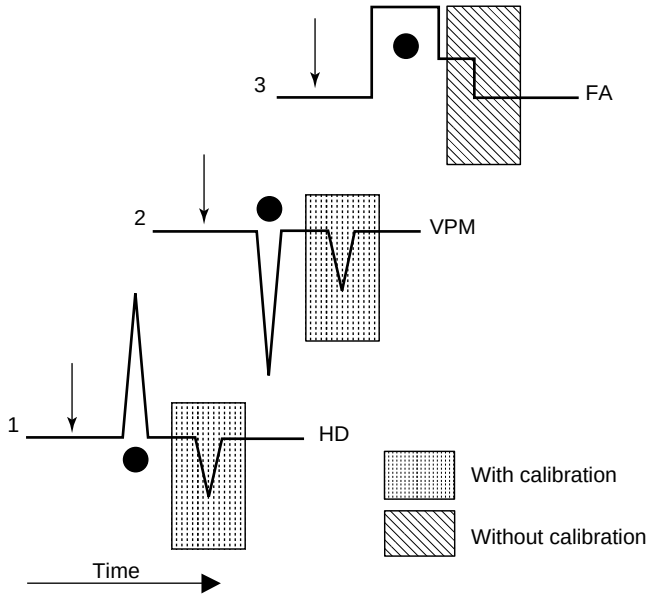

Figure 3 Schematic comparison between HD method, VP method and FA in HPLC. Protein and protein-associated ligand fraction (●) elutes in the void volume of the column. The highlighted area indicates in all cases the signal corresponding to ligand and its quantitative measurement (in the case of HD and VP methods an internal or external calibration procedure is applied). 1, 2 and 3 denote equilibrium conditions prior to injection of the sample established with ligand in phosphate buffer, ligand + protein in phosphate buffer and phosphate buffer alone, respectively.

3.2.2.2 Vacancy Peak Method (Equilibrium Saturation Method) When both interacting species (ligand and protein) are implemented as eluent constituents, after achieving equilibrium, injection of a small buffer volume results typically in two negative peaks. The first corresponds to the protein–ligand complex and from the second the free drug concentration can be determined.[110,113] Since the solubility of a substance increases in the presence of a protein, this approach is of particular interest for poorly soluble drugs[108] and for studying competitive binding aspects [e.g. influence of FFA or sodium dodecyl sulfate (SDS) on warfarin–HSA binding[110]].

Overlooking some other limitations (e.g. excessive amounts of material required), the presence of protein in the mobile phase results in high background absorbance and an internal calibration is usually needed.

3.2.2.3 Frontal Analysis In the FA method [high-performance frontal analysis (HPFA), FA], a large volume of drug–protein mixed solution is applied continuously to a size-exclusion column to achieve a steady-state

concentration. The elution profile obtained consists of α-, β- and γ-plateau zones corresponding to the free protein, mixture of protein-bound ligand and free ligand, respectively. Earlier applications of this method indicated that large sample volumes are necessary to observe a clear γ-plateau for hydrophilic drugs (e.g. >10 mL for warfarin).[109,117] Updated analytical set-ups [using, e.g., hydrophilic poly(vinyl alcohol) gels] enabled the sample volume to be decreased dramatically (by a factor of at least 10) and the free drug concentration to be determined also for relatively hydrophilic drugs (e.g. salicylate or acetazolamide) exhibiting weak protein binding.[130] The HPFA method is advantageous particularly for the estimation of the concentration of unbound hydrophobic drugs, since

- relatively small sample volumes are injected (100 µL–2 mL, depending on the free drug fraction);
- untreated plasma samples may be injected;
- strong adsorption effects interfering by using conventional separation approaches are eliminated;
- column-switching systems facilitating simple, rapid and (stereo)specific separation could be used.[120]

On-line coupling of HPFA and chiral HPLC offers the prospect of direct and enantioselective quantification of unbound drug under protein binding equilibrium conditions, as has been shown for enantiomers of warfarin,[115] ketoprofen,[116] fenoprofen,[114] nilvadipine,[120] xanthine oxidase inhibitor, BOF-4272[113] and levo-/semotiadil.[132] In addition, this methodological approach is suitable for revealing species-dependent or concentration-dependent aspects of stereoselective binding.[131]

3.2.2.4 Other Chromatographic Methods

Retention Analysis: Protein as Mobile Phase Additive. Since the retention volume of a ligand eluted by a protein solution depends on the association of the ligand with the protein, the binding affinity of interaction can be determined using the protein as a mobile phase additive (i.e. protein pseudostationary phase). Differences in retention volumes in the absence and presence of the binding protein could be used to calculate the corresponding binding characteristics (i.e. total affinity, denoted $\sum_{i=1}^{m} n_i k_i$), according to Equation (1):[122,123]

$$\frac{V'_R - V_0}{V_R - V_0} = \frac{1}{1 + \sum_{i=1}^{m} n_i k_i P_t} \quad (1)$$

where V_R and V'_R are the retention volume in the absence and presence of the protein, respectively, V_0 is the void volume of the chromatographic column, n is the number of binding sites, k is the binding constant of the interaction and P_t is the total concentration of the biomacromolecule (receptor). Since proteins interact strongly with silica-based solid phases, questions have been addressed about the adsorption of, for example, albumin to different materials comparing hydrophobic (RP-8, phenyl) and hydrophilic (diol) groups; optimal conditions (with only negligible protein adsorption) were achieved on a diol-type support. The introduction of diluted albumin solution ($<80\,\mu mol\,L^{-1}$) functioning as a chiral pseudostationary phase resulted in the successful separation of, for example, enantiomers of lipophilic omeprazole[124] and hydrophilic tryptophan. The main requirement of this methodology is related primarily to the stability of the bound/free drug ratio during analysis and this is solved experimentally using small amounts of drug injected.[108]

Zonal (Elution) Chromatography. If the binding equilibrium is not disturbed or modified by the interaction of a drug with a buffer-equilibrated stationary phase, samples containing both ligand and protein could be analyzed by direct injection. In principle, the protein–drug complex elutes first and is followed by a second peak corresponding to the free ligand: the validity of binding parameters is dependent on the interaction kinetics, assuming stability (nondissociation) as one of the components is eluted from the zone. This method is appropriate particularly for high-affinity complexes (with $K_a > 10^6$–$10^7\,L\,mol^{-1}$, i.e. values corresponding to affinities of hormones to their specific protein transporters). Its application for ''middle-range'' drug binding affinities ($K_a = 10^3$–$10^6\,L\,mol^{-1}$) is not possible, since their dissociation kinetics are faster than the chromatographic elution process. Along this line, some authors have demonstrated the use of zonal chromatography for clinical therapeutic drug monitoring of free drugs in serum [e.g. free phenytoin with a limit of detection (LOD) of $0.1\,\mu g\,mL^{-1}$ by injecting sample volumes of 250–450 µL[121]].

3.3 Capillary Electrophoresis

In the past decade, various forms and modifications of CE have been rationalized and applied to the characterization of noncovalent ligand–protein binding interactions. Proteins may be used as:

- dynamic protein coating of the capillary (as renewable PSPs);
- soluble additives of running buffers in ACE, VACE;
- cross-linked/immobilized in capillary affinity gel electrophoresis (CAGE);
- immobilized in affinity CEC;
- similarly as in HPLC, the FA, HD or VP method can be adopted.

3.3.1 Affinity Capillary Electrophoresis

The main advantage of ACE over affinity HPLC with dynamic equilibria for the evaluation of biospecific binding interactions resides in the following:[133–137]

- miniaturization of the experimental set-up and, therefore, significantly smaller amounts of (non-purified) protein or ligand have to be introduced when "equilibrium conditions" should be established;
- creating of a "pseudophysiological" environment for biospecific drug–protein or protein–protein interactions, eliminating the difficulties related to the use of different (more or less) biocompatible HPLC stationary phases;
- gathering quantitative (or qualitative) information on binding interactions by using a ligand or protein as a buffer (BGE) additive or implementation of both interacting species in the running buffer (VACE), i.e. no protein immobilization is required;
- direct enantioseparation if stereoselective drug binding occurs.

If the protein is used as a buffer or BGE additive and the solution completely fills the separation capillary including both buffer reservoirs (so-called complete-filling technique), by relating the electrophoretic mobility to binding it is possible to determine ligand (L)–protein (P) binding parameters (Equation 2):[138]

$$\mu = \mu_f \frac{[L]}{[L] + [LP]} + \mu_p \frac{[LP]}{[L] + [P]} \quad (2)$$

where μ is the overall mobility of the analyte, μ_f is the mobility of the free analyte, μ_p is the mobility of the analyte–protein complex and [L] and [LP] are the concentrations of the free and complexed ligand, respectively. When monovalent binding is considered, by substituting $K_a = [LP]/[L][P]$, Equation (2) transforms to Equation (3):

$$\mu = \frac{\mu_f + \mu_p K_a [P]}{1 + K_a [P]} \quad (3)$$

where K_a is the equilibrium association constant and [P] is the free protein concentration. However, determination of binding affinity by using this approach may become complicated, as can be seen in Table 4, demonstrating that

Table 4 ACE: protein used as a buffer additive

Protein	Drug	Aspect(s) considered[a]			Refs.
		M/S	BC	SS	
BSA, derivatized BSA, albumins	(±)-Ofloxacin (D,L)-Tryptophan	•		•	141
BSA	(6*R*)-, (6*S*)-Leucovorin	•		•	143
BSA, HSA	(±)-Ofloxacin (*RS*)-Warfarin	•		•	142
BSA	Porphyrins Porphyrin isomers			•[b,c]	144
HSA	(*RS*)-Benzoin (±)-Promethazine (±)-Thioridazine (±)-Propiomazine	•	•[d]	•	138
BSA, HSA	(D,L)-Tryptophan	•		•	145
BSA, AGP (ovomucoid, cellulase)	(D,L)-Tryptophan (±)-Benzoin (*RS*)-Pindolol Promethazine			•[e]	146
HSA	(D,L)-Tryptophan (*RS*)-Warfarin	•		•	147
Vancomycin	Peptides	•	•		148, 149
Heparin	Chloroquine, etc.	•		•	150, 151

[a] •, Particular aspects considered: M/S, relative information on binding mechanism and/or binding stoichiometry; BC, binding constants (characterizing binding capacity and affinity); SS, stereoselective aspects.
[b] Geometrical isomers.
[c] Addition of SDS.
[d] Valid only for (*S*)-benzoin.
[e] Addition of organic modifier (2-propanol).

the measurement of quantitative binding characteristics has been performed rather exceptionally. The main difficulties reside in the accurate determination of μ_p, in a large detector response itself, in the presence of zone-broadening processes or in exact temperature control.[138] Special attention should be paid further to protein–wall adsorption phenomena and an appropriate (reproducible) capillary coating is often necessary. In addition, as has been pointed out, enantioseparation of some drugs achieved using proteins as a buffer additive in CE cannot always easily be reproduced by affinity HPLC (with PSPs) and vice versa.[139,140] This could be explained most probably by different protein concentrations (and thus available protein binding sites) in each system. In CE, the HSA used as a buffer additive is usually at tens of micromolar concentration, whereas HSA immobilized as a chiral selector in affinity HPLC may approach millimolar levels.[139] As already suggested, the main advantage of this approach is that the injected ligand may be a racemate, and if stereoselective binding occurs, a chiral separation could be achieved. By these means, stereoselective interactions of (±)-ofloxacin with different albumins (either chemically modified or from different biological species) were investigated,[141,142] and stereoselective species-dependent binding aspects and relevant displacement interactions were described. Further examples and aspects evaluated are listed in Table 4 and the principal working schemes are depicted in Figures 4 and 5.

An alternative approach is to use the so-called partial-filling technique, i.e. conditions in which the protein solution fills only part of the capillary (starting from the injection end of the capillary). Since the protein is not present at the detector window, this approach can achieve better sensitivity than complete-filling ACE methods.[133,134] This could be further optimized by coupling of ACE with mass spectrometry (MS), as demonstrated by Terabe et al.[152]

When the ligand is used as buffer additive, derivation of the respective binding parameters is straightforward also considering numerous interfering factors as described above, and this alternative has been proved successfully in several binding studies (Table 5). A ''historic'' form of ACE for drug–protein studies is represented by the conventional affinity electrophoresis in agarose gels [agarose gel electrophoresis (AGE)].[165,166] The experimental set-up of AGE was updated recently with heterogeneous (multi)sectioned slab gels, improving the run efficiency.[167] Dissociation constants in the nanomolar range could be determined in this way,[166] but if only minor migration shifts occur the estimation of binding characteristics is not possible for practical reasons.[148,157,158] In principle, if the protein binds a (charged) ligand of relatively small mass, the change in

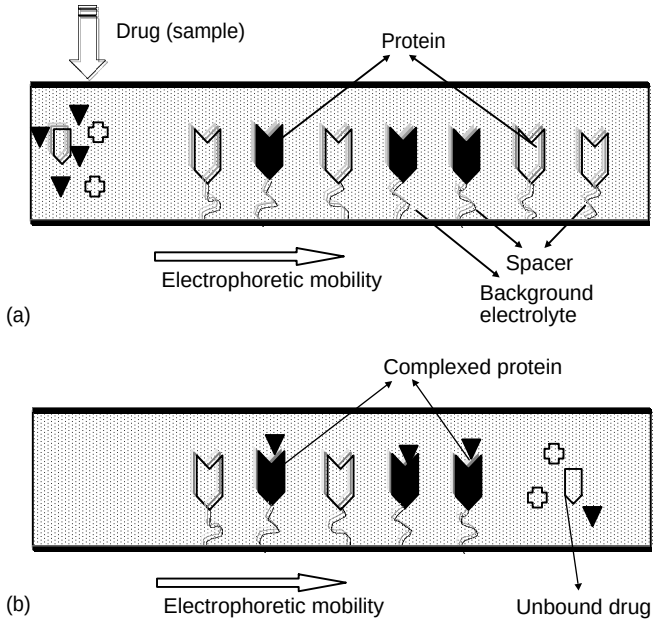

Figure 4 Schematic illustration of (a) sample injection and (b) electrophoretic separation of protein-bound and free drug using immobilized (or adsorbed) protein in CE. In this format, the applied electric field creates a flow of solvent and solute (drug) through the system. Direct enantioseparation of chiral drugs suggesting enantioselective binding differences could also be detected in this way.

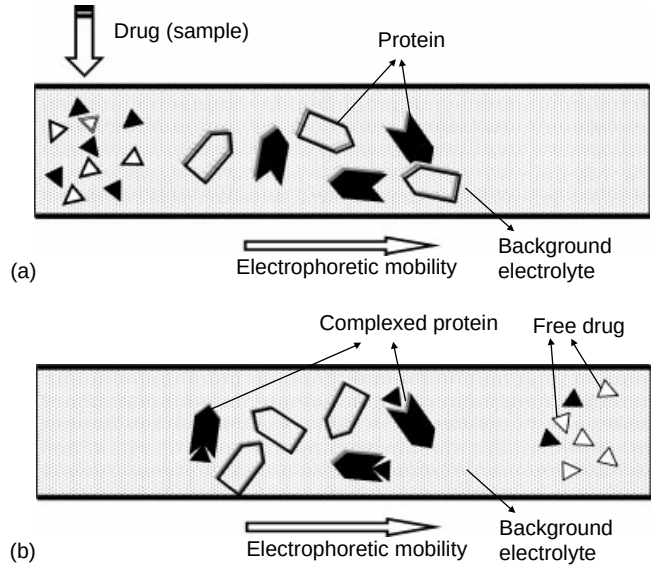

Figure 5 Schematic illustration of (a) sample injection and (b) electrophoretic separation of protein-bound and free drug using nonimmobilized protein (e.g. protein as buffer additive) in CE. The CE system contains a fixed concentration of protein in the running buffer and if solutes (drugs) are recognized via more or less specific binding interaction(s), a change in their mobility could be detected. Enantioselective binding could be resolved based on the corresponding shifts in drug mobilities.

Table 5 ACE: measurement of migration shifts by using ligand as a buffer additive

Protein (peptide/enzyme)	Ligand	Further specification of the method	Affinity constant(s)[a]	Refs.
HSA BSA	(*RS*)-Ibuprofen (*RS*)-Ketoprofen (*RS*)-Flurbiprofen		•	153
Concavalin A	Neutral mono-saccharides	Fluorescence affinity probe CE	••	154
Pea lectin	Neutral sugars	Affinophoresis	•••	155
Phosphotyrosine (Ag)	MAbs	Immuno CE	••	156
Bovine carbonic anhydrase B	Charged vs neutral arylsulfonamides		•	157, 158
Vancomycin Carbonic anhydrase B	Dipeptide *N*-acetyl-D-Ala-D-Ala Arylsulfonamides	Multiple plug binding assay	• •	159
Human C-reactive protein	Ca^{2+}, phosphorylcholine		••	160
Heat shock protein Hsc70	Deoxyspergualin		•	161, 162
DNA	MAbs	Immuno CE[b]	•••	163
Vancomycin	Peptidoglycan precursors		•	164
Rat myeloma monoclonal IgG_{2b}	Ligands containing *N*-2,4-dinitrophenyl group		•	41

[a] •, Association/dissociation constant obtained from Scatchard plot; ••, Association/dissociation constant obtained by linearization of the data using double reciprocal plot; •••, Association/dissociation constant obtained by linearization of the data using Woolf–Hofstee plot.
[b] MAbs, monoclonal antibodies; Compared with biosensor (surface plasmon resonance) technology.

its electrophoretic mobility (μ) due to the change in mass [from $M^{2/3}$ to $(M+m)^{2/3}$] is small relative to the change in μ due to the change in charge (from Z to $Z+z$). Therefore, the protein–ligand complex migrates at a different rate to the uncomplexed protein: by measuring migration times as a function of the concentration of a (charged) ligand present in the buffer, it is possible to use Scatchard analysis to estimate the affinity constant, K_b (Equation 4):

$$\frac{\delta\Delta t}{\delta\Delta t_{max}}\frac{1}{[L]} = K_b - K_b\frac{\delta\Delta t}{\delta\Delta t_{max}} \qquad (4)$$

where Δt is the difference between the migration time of the protein of interest and the reference protein ($\Delta t = t_{ref} - t_{interest\ sample}$) at concentration [L] of the charged ligand, $\delta\Delta t = \Delta t - \Delta t_0$ (Δt_0 is the value of Δt at $[L] = 0$), and Δt_{max} is the value of $\delta\Delta t$ at saturating concentrations of L. (The reference protein is usually another protein having a similar value of migration time but which does not bind the ligand studied.) If the migration time is affected by changes in electroosmotic flow (EOF), Equation (4) is not accurate for estimating binding affinity and has to be corrected accordingly. Direct analysis of the interaction of protein with electrically neutral, low-molecular-weight ligands using this approach does not seem possible, since the neutral ligand does not introduce a sufficient change in charge (or charge distribution) of the protein–ligand complex. Hence the mobilities of uncomplexed and complexed protein cannot be distinguished from each other. An alternative estimation of the binding constant is possible when, for example, a neutral ligand competes with a charged ligand of known binding constant.[158] The validity of Equation (4) is restricted to monovalent binding. Nevertheless, an optimized experimental protocol has been suggested also for the determination of bi- and multivalent binding stoichiometries.[40]

A typical design of ACE experiments for deducing quantitative binding characteristics requires multiple experimental runs at different concentrations of the ligand; the use of a multiple-plug binding assay has been attempted recently in order to avoid this practical limitation.[159]

For specific purposes, labeling of one molecular species [affinity probe CE] using a fluorophore [followed by laser-induced fluorescence (LIF) detection][154] could expand the applicability of ACE for the evaluation of drug (ligand)-binding equilibria. Fluorophore-labeled molecules can be detected at very low concentration ($\sim 10^{-9}$ mol L^{-1}), which is particularly important in the case of high-affinity complexes (dissociation constant in the nanomolar range). When the ligand of interest is not charged or the number of the charges is not sufficient to induce the required mobility change of the binding protein, it could be modified with an affinophore. Affinity constants of divalent lectins for neutral sugars have been derived in this way on the microscale.[155]

The major problem of CE methodology remains the tendency of proteins to adsorb on the wall of uncoated capillaries. This particular problem is most commonly minimized either by coating the capillary wall or by adjusting the pH of the electrophoresis buffer.[134]

3.3.2 *Vacancy Affinity Capillary Electrophoresis*

In the VACE method, both interacting species are added to the buffer, whereby the concentration of one is fixed and the concentration of the other is varied (most commonly, the protein concentration is kept constant).

Similarly as in the VP method in HPLC, the injection of a small volume of buffer leads to the occurrence of two negative peaks. The area of the first peak corresponds to the concentration of free protein and the second peak reflects the concentration of the complex and free drug in the buffer. Unlike the calculation of peak areas, information on the association constant can be obtained from their migration shifts. There are two principal possibilities for the extraction of binding parameters when using VACE in binding assays, the shift in the mobility of the drug ($\mu_{D,P}$) and the shift in the mobility of protein ($\mu_{P,D}$), and both can be plotted against the ''total'' drug concentration in the buffer. As stressed by Busch,[168] the main advantage of the VACE method is that the plot of $\mu_{D,P}$ vs total drug contains information on the absolute number of different binding sites present on the protein. In the other case (and similarly as in ACE), i.e. plotting $\mu_{P,D}$ versus drug concentration in the buffer, it is possible to obtain the affinity constant value, but not the absolute number of binding sites. As with the VP method in HPLC, the VACE method is well suited for hydrophobic ligands, even in cases of multiple equilibria.

3.3.3 *Hummel and Dreyer and Frontal Analysis Method Adapted to Capillary Electrophoresis Conditions*

Similarly as in HPLC, FA and HD methods have been adapted to CE conditions.[128,168–170] Using FA [high-performance capillary electrophoresis/frontal analysis (HPCE/FA)], both nonstereoselective[119] and stereoselective[171] aspects of drug–protein interactions have been described. Compared with HPLC/FA, the sample volume in HPCE/FA is small (~80–200 nL) and since the separation is based on the measurement of differences in electrophoretic mobility of interacting components, interactions between molecules of similar size (but different electrophoretic mobility) should also be considered. HPCE/FA with electrokinetic injection could be adopted only when the drug and the protein migrate in directions opposite to each other, hence it is readily applicable particularly for positively charged drugs.[171] In turn, HPCE/FA following hydrodynamic injection is much more robust and is applicable for basic, acidic and neutral drugs (as long as electrophoretic mobility differences exist between drug and protein). The use of HPCE/FA has been proved useful also for the determination of binding parameters of a hapten (drug)–antibody complex (with dissociation constants in the nanomolar range): the hapten concentration can be measured directly, enabling one to check also the stoichiometry of the binding model.[170] The major disadvantage of HPCE/FA is related at present to the relatively high detection limit and, thus, insufficient sensitivity for the analysis of clinical samples. This holds true also for the HD method which is also more laborious, since a simplified experimental design[111] does not provide satisfactory smooth binding isotherms in CE.[128,169] In contrast to FA, the HD method could not be coupled with chiral separation, since detection is indirect. Its use could be recommended, e.g. for monitoring the binding parameters of highly lipophilic drugs [e.g. for carvedilol enantiomers the analysis time could be effectively reduced, i.e. from ca. 40 min in HPLC to ca. 15 min in CE[128]].

Summarizing, if both interacting species are detectable and their mobilities are different, both ACE and VACE can be considered as complementary techniques. Specifically, with the VACE method it is possible not only to derive the association constant(s), but also to characterize the number of binding sites (this is particularly important when multivalent binding models are considered). Thereby, a very significant contribution to minimizing systematic errors involved in VACE or ACE method could be obtained by decreasing the concentration of one of the interacting species to as low a level as possible.[168]

Assuming that the mobility of the protein is equal to the mobility of the drug–protein complex, as an alternative, both FA and HD methods can be used for deriving quantitative binding parameters. Since the HD method always requires a certain calibration procedure (internally or externally), FA could be seen as methodologically advantageous in terms of its experimental versatility (including material and handling requirements) and ability to derive complete sets of reliable binding characteristics.[119,168,171]

3.3.4 *Capillary Electrochromatography*

CEC is a hybrid technique of HPLC and CE attempting to eliminate their individual disadvantages (e.g. large-scale set-up of HPLC consuming large amounts of samples, mobile phases, mobile phase additives and packing materials) and exploring certain advantages (e.g. the use of neutral species as compared with CE, leading to unique selectivity).[98] The main operating difficulties are often related to bubble formation within the column (leading to the breakdown of current and EOF) and

local overheating phenomena resulting in drying out of the capillary. Although not experimentally proved and confirmed yet, CEC could be of interest if miniaturization and/or enhanced separation selectivity for separation of interacting species is addressed. Nonspecific adsorption phenomena of proteins to silica-based surfaces may alter the zeta-potential and thus the EOF resulting in retention time shifts and low reproducibility.

3.4 Affinity Information on Biospecific Interaction with Respect to Immobilization of Interacting Species

Immunoassays such as an enzyme-linked immunosorbent assay (ELISA) or radioimmunoassay (RIA) employing covalent immobilization of one of the interacting species (either an antibody or an antigen) on a solid support or matrix (microtiter plate, glass-fiber surface, etc.) are frequently used to characterize binding models, including the extraction of binding parameters. Beyond some technical difficulties (development of a rational and reproducible immobilization procedure, lot-to-lot variation in solid-phase materials, nonspecific binding of the reactant to a solid matrix, many rinsing steps to wash out unreacted receptor molecules), the reaction kinetics using solid-phase assays are often very slow, requiring several hours (overnight) of incubation.[172] Although the ''solid-phase'' techniques are experimentally relatively convenient, measurement of reliable binding parameters by this approach seems not to be a straightforward process. Several factors (e.g. antibody valency, steric hindrance, antigen density and mobility on the surface) may severely affect the equilibrium in this heterogeneous phase system. This implies further difficulties in the quantification of ''real'' affinity constants. As a consequence, only an ''apparent'' equilibrium constant (differing from the thermodynamically defined affinity in solution) could be obtained using, for example, ELISA methodology. Ultimately, this has to be taken into account by comparing data sets generated with heterogenous methodologies.

An inventive approach is represented by the development of receptor-based LC stationary phases, e.g. if solubilized nicotinic acetylcholine receptor (nAchR) subtypes are immobilized in the phospholipid monolayer of an IAM.[173] It has been shown that (a) immobilized nAchR preserved its biorecognizing properties after solubilization and immobilization procedure and (b) the affinities of model agonists [epibatidine, (−)-nicotine, carbachol or atropine] were well correlated with the reference method. Hence the method could be employed for the rapid on-line screening of combinatorial pools of potential drug candidates (as in this case of nicotinic-like activity).

Biomolecular interaction analysis (BIA) has been introduced to monitor the kinetics of molecular binding complexation in real time and without the use of labels.[174] In this method, one interactant is immobilized on the sensory chip and the other is injected across its surface using a sophisticated microfluidics system providing a continuous, precisely controlled flow of analyte to the surface of the sensor chip. BIA is based on the principle that two or more molecules interacting in solution induce a change in the refractive index detectable by surface plasmon resonance (at the surface of a sensor chip). Compared with conventional solid-phase methodologies (ELISA, RIA, etc.), one of the most attractive experimental advantages of BIA technique resides in the use of biosamples without prior purification or solubilization. In some cases, changed binding characteristics have been reported owing to its insufficient regeneration capabilities of the chip and another limitation is that a measurable signal requires analytes to be at least of $M_r = 2000–10\,000$.[175] Comparison of immuno-CE with BIA technology for the characterization of a monoclonal antibody against DNA confirmed good agreement of the binding characteristics derived. However, gradual loss of DNA from the chip during regeneration procedures was noted with BIA as compared with CE, where neither the ligand nor the analyte is reused.[175]

It has been demonstrated that ''solution-phase assays'' such as ACE can be used for checking of stoichiometric model and estimation of affinity constants regardless of cooperativity.[41] Binding assays taking place in solution eliminate some of the problems associated with solid-phase matrices, since interacting molecules are free from attachment (i.e. without steric hindrance effects), improving the kinetics of binding and the binding equilibria significantly. Recently, it has been shown that for the quantification of antibody–antigen interactions, statistically reliable binding parameters can be obtained with FA in CE conditions.[170] However, except for a larger sample volume required for a particular binding isotherm (500 μL in CE vs 10 μL with ELISA), the major limitation of CE measurement is related to the poor ultraviolet (UV) sensitivity, so that is not possible to describe quantitatively a high-affinity complex (e.g. antigen–antibodies). If a sophisticated design of the optical path length (capillaries with a bubble-cell or Z- or U-shaped cell) does not lead to a sufficient improvement, another detection method [laser-induced fluorescence immunodetection (LIFID)] could be employed (if available) as an alternative. However, either native fluorescence of the ligand or a specific labeling technology is required to use LIF. High-intensity laser systems are several orders of magnitude more sensitive than UV absorbance (LOD ca. $2 \times 10^{-10}–2 \times 10^{-9}$ mol L^{-1}) providing also greater specificity.[172] The lack of an opportunity to run multiple samples simultaneously could

be solved in CE by introducing so-called multicapillary arrays.

3.5 Spectroscopic Methods

Compared with separation methods, spectroscopic methods facilitate predominantly an insight into three-dimensional protein structure, elucidating complementary structural and/or conformational variations of a protein molecule resulting from drug attachment. Of the different spectroscopic methods [UV, visible, fluorescence, nuclear magnetic resonance (NMR)] and circular dichroism (CD), fluorescence spectroscopy has been used most widely.[43,81,176–178] Derivation of quantitative binding parameters is based on measurements of the changes in the intrinsic molar fluorescence (ψ_B) of the protein-bound drug (fluorescence quenching). It is a common practice to measure fluorescence at a constant drug concentration and gradually increasing protein concentration, whereby extrapolation of these two parameters in a double-reciprocal plot enables one to calculate the value of the intrinsic molar fluorescence of bound drug (F_b). The most relevant shortcomings are related to nonlinearity of the plots derived and to incorrect derivation of binding characteristics, since binding affinity is dependent on protein concentration. Similarly, ignoring nonspecific quenching phenomena may lead to the misinterpretation of binding parameters.

Of other spectroscopic methods which have been used to characterize drug–protein interactions, proton (^{1}H) NMR spectroscopy should be mentioned. However, in order to obtain reproducible and reliable spectra of subtle alterations in protein structure after ligand complexation, high-field NMR spectrometers with advanced resolution and data acquisition and processing technology are required. The interpretation of results is based on the fact that drugs interacting with protein in a nonspecific (i.e. low-affinity) manner show different spectra, often with only somewhat broadened but not resolved signals of drug molecules. In contrast, the more specifically bound drugs usually cause a definite structural change of the ''surrounding'' protein molecule (and/or the structural changes facilitate the high-affinity binding) and, consequently, in NMR difference spectra the drug signals are either very broadened, shifted or cannot be determined at the positions where the free drug signals were observed.[179,180] For special purposes, (^{19}F)-NMR spectroscopy has been successfully used to monitor the interactions of fluorinated drugs (5-fluorotryptophan, 5-fluorosalicylic acid, flurbiprofen and sulindac) with HSA.[181] Drug-dependent changes in the NMR heteronuclear (^{1}H–^{13}C) spectra have been used for monitoring of drug (bepridil, trifluoperazine) binding sites on cardiac troponin C.[182] It is of importance that the signals derived by NMR spectroscopy not only indicate the location of the respective binding site, but also provide information on the residues of protein involved in the drug binding. Current developments in high-field (500–600 MHz) NMR technology offer, owing to the very high resolution of NMR signals, many promising opportunities for detailed mapping of binding interactions including the stoichiometry, kinetics and conformational properties of drug–protein complexes.[183–185] One- and two-dimensional homonuclear (e.g. ^{1}H, ^{13}N) or heteronuclear high-field NMR spectroscopy are suitable for the investigation of equilibrium conformational states in order to determine the most favorable structures of the intercalated complexes of different stoichiometries.[185] Predominantly, however, the mapping of primary (high-affinity) binding sites is usually achieved by using NMR spectroscopy. Monitoring of the additional presence of secondary low-affinity binding sites does not seem possible[179] and/or has not been described yet. A further limitation of this method is the use of drug concentrations that are not pharmacologically relevant (typically in the millimolar range).

Chiroptical methods, optical rotatory dispersion (ORD) or CD, have been proposed as a useful tool to describe dynamic movements of protein conformation (and the role of relevant protein functionalities) in order to obtain more complex information on the binding mechanisms involved.[186–188] The conformational status of a protein and modification of its secondary and tertiary structure due to complexation with a particular drug are of crucial importance for the binding properties of the relevant binding sites. This has been demonstrated for protein-linked conformation changes of albumin in the physiological pH range 7.0–9.0, with the N-form occurring below neutral pH and the B-form at higher pH (the N–B transition).[189–192] Consequently, in nonstereoselective assays it has been shown that ligands are bound differently to the N- and B-forms of HSA: warfarin and diazepam exhibited higher affinity for the B-form,[193,194] pirprofen and benoxaprofen were bound preferentially to the N-form,[195] whereas suprofen affinity did not seem to be affected by the N–B transition.[196] Evaluation of possible conformational alterations of acceptor molecules are of major significance for the description of the binding mechanism of stereoselective interactions and should be considered in the experimental study design.

In comparisons of binding studies evaluating both qualitative and quantitative aspects of drug–protein interactions by employing a variety of complementary techniques (ED, UV, fluorescence, CD and NMR), a clear discrepancy between results acquired from CD vs ED has been observed as related to the binding mechanism. Monitoring of heterogeneous binding processes (concomitant presence of low-affinity binding

components) is usually not possible by CD, hence quantitative information derived from such studies should be interpreted carefully.

3.6 Other Methods and Future Trends in Analytical Methodology

Some other methods have occasionally been used in the framework of drug–protein binding studies, with respect to unique features of the ligand or to reveal specific quantitative or qualitative aspects of binding interaction. Examples include the use of

- polarography;[113,197]
- (micro)calorimetry;[198,199]
- stopped-flow experiments [this approach was able to elucidate, for example, different albumin binding kinetics in diabetics[200]];
- fluorescence polarization immunoassay (FPIA), introduced mainly for routine therapeutic monitoring of free levels of drugs with a narrow therapeutic index.

Apart from ''main-stream'' analytical (separation) techniques, interesting potential resides also in other methods, such as field flow fractionation (FFF), permitting also analyses of raw and complex biological samples (protein mixtures, aggregates, cells, etc.).[201,202]

4 IN VIVO MONITORING OF FREE FRACTION OF DRUGS: MICRODIALYSIS

The sampling and determination of the ''true'' free drug fraction in vivo, i.e. in a dynamically stirred biological system with many interlocked processes, is of primary pharmacological importance. Obviously, a complex methodological approach should reflect the existing organ-specific differences as well as ''free intermediate action'', i.e. the fact that part of the initially bound drug is eventually released, depending on the capillary transit time, the rate of dissociation of the drug–protein complex and the permeability and surface area of the capillaries.[203–205] Although measurements of saliva or cerebrospinal fluid have been proved to estimate the free drug fraction, they are of only limited general utility for free drug therapeutic monitoring.[206,207] The use of the microdialysis perfusion technique was proposed as an in vivo alternative for the study of the steady-state free fraction of drugs in different tissues and body compartments.[208] In microdialysis sampling, a hydrophilic capillary (small-diameter dialysis tubing) is implanted in compartment(s)/tissue(s) of particular interest. By perfusion of the capillary with a perfusate solution at a low flow rate (usually less than $1–2\,\mu L\,min^{-1}$), low-molecular-weight compounds diffuse into the perfusate, in contrast to proteins (and other higher-molecular-weight compounds), which do not enter the perfusion medium. In summary, microdialysis offers advantages in terms of maintaining equilibria and experimental versatility in vivo (time-dependent sampling, rapid continuous sampling avoiding enzymatic degradation of the sample, or sampling in awake, freely moving animals). The alternative use of this method for in vitro plasma protein-binding studies has been verified in modified experimental set-ups for determination of the free concentration of drugs (paracetamol, procainamid, caffeine, theophylline, lidocaine, carbamazepine, phenobarbital and phenytoin) in vitro.[209] The results obtained confirmed that microdialysis is capable of operating effectively by preserving binding equilibria over a wide range of drug concentrations and protein binding. Since it is possible in animal experiments to implant the microdialysis probe(s) intravenously, the extent of drug binding to proteins could be studied temporarily in vivo. Owing to the relatively small volumes of samples acquired (a few microliters depending on the collection rate, which is usually $0.1–5\,\mu L\,L^{-1}$, and typical sampling intervals are 1–5 min), the major limitation of microdialysis could be related to the poor sensitivity of most of the currently available HPLC (or CE) systems. Capillary ultrafiltration, employing an active-pressure gradient instead of a passive diffusion concentration gradient for the sampling process, is another new technique with the ability to monitor unbound drug in biological systems in vivo.[210]

5 EVALUATION OF BINDING DATA

In principle, there are two possibilities for the interpretation of reversible ligand (L)–protein (P) interactions, i.e. ''real binding site'' (Equation 5) and ''stoichiometric'' (Equation 6) concepts:[211]

$$B = \frac{k_1 L}{1 + k_1 L} + \frac{k_2 L}{1 + k_2 L} + \cdots + \frac{k_n L}{1 + k_n L} \tag{5}$$

$$B = \frac{K_1(L) + 2K_1K_2(L)^2 + \cdots + n(K_1K_2\cdots K_n)(L)^n}{1 + K_1(L) + K_1K_2(L)^2 + \cdots + (K_1K_2\cdots K_n)(L)^n} \tag{6}$$

The binding constants (i.e. k_1, k_2, k_n vs K_1, K_2, K_n) derived by the two approaches are not the same either in magnitude or in the binding step to which they have to be assigned.[42,212] Stoichiometric formulation using Adair's binding equation (Equation 6) neglects the rather conventional concept of rigid and independent

binding sites attached by a particular ligand. It can be applied for any multiple binding mechanism, irrespective of whether the ligand molecules are bound cooperatively, independently or anticooperatively. Moreover, in this model, the sequential binding of ligands is reflected by the conformational adaptability of protein binding sites and it could explain more transparently its ability to bind two or more ligand molecules to one binding locus. From a theoretical point of view, the stoichiometric model provides the most versatile format and it could be reduced by additional conditions to the ''site'' (Scatchard) model. In practice, Adair's equation is not as popular as Scatchard's; however, it offers a more complete view of so-called ''nonreceptor'' (e.g. albumin) drug-binding interactions.[213,214]

The usual approach in ligand binding studies is to fit the experimental data to Equation (5) and to plot them in the form of a so-called Scatchard plot[215] (Figure 6a–c). The most important limitations concerning this approach include either the oversimplification of ligand attachment to the binding site(s) by fitting of curvilinear plots with straight lines or, contrarily, the detection of visually, biochemically or pharmacologically not interpretable acceptor heterogeneity.[216–220] At least in part, a number of experimental artifacts depending on the methodology applied may explain the curvilinear nature of Scatchard plots. The reasons for downward curvature are represented predominantly by incomplete recovery of the bound fraction, irreversible ligand binding or its internalization, ligand or acceptor degradation and nonequilibrium binding conditions. Upward curvature is usually caused by an affinity difference between labeled and unlabeled ligands, imprecise estimation of nonspecific binding or by contamination of the bound fraction with unbound ligand.[220] Some factors (impurities in labeled ligand preparation or influence of membrane microenvironment) may result in either upward or downward curvature of Scatchard plots. On the other hand, by choosing an inappropriate analytical method for measuring binding data sets, some low-affinity binding components may be overlooked (Figure 6a–c). In addition, another obvious reason for variability of binding parameters is that a ligand's association constant and/or the number of binding sites increases when the acceptor (receptor or protein) preparation is diluted.[216] This point is related partly to the very common situation when a competitive inhibitor (or contaminant) of binding is present in the acceptor preparation. Obviously, all the above-mentioned aspects are plausible also for the evaluation of stereospecific drug–protein interactions (including enantiomeric contamination).

Generally, although Scatchard plots will probably be used further for illustration and comparison of different binding data sets, for quantitative evaluation purposes some alternative graphical representations must be considered. The major advantage of these[217,218]

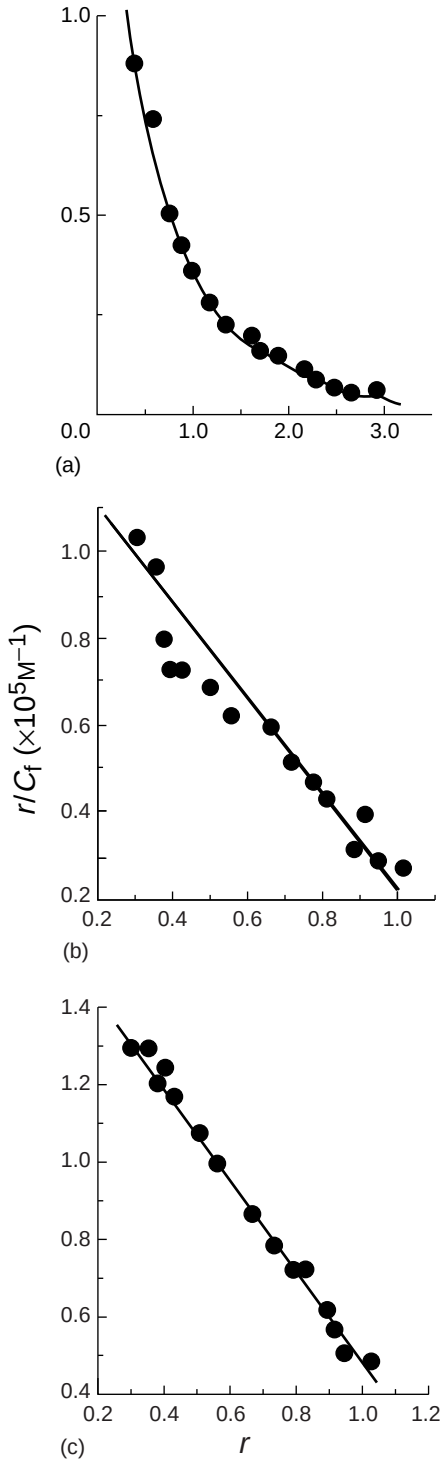

Figure 6 Comparison of binding data obtained for (±)-suprofen interaction with HSA determined by (a) ED (b) CD and (c) fluorescence quenching at pH 7.4 and 25 °C.

is the direct representation of experimental data. As pointed out,[220] the application of linear regression to transformed (Scatchard) data in order to calculate the binding parameters could no longer be accepted as an appropriate quantitative approach.

6 CONCLUSION

Description of interactions of proteins (and also peptides, enzymes, DNA, antigens, fragments of membranes or cells, etc.) with ligands (drugs and other xenobiotics, endogens, ions, metals, etc.) is an issue of vital importance in bioscience. Although some exceptions for specific purposes may exist, an ''optimal methodology'' for characterizing the principal functional criteria of binding interactions (i.e. bioaffinity, selectivity, specificity, stoichiometry, saturability and stereoselectivity) should exhibit

- methodological (bioanalytical) power (e.g. sensitivity, speed, precision, validity, versatility, cost effectiveness);
- broad applicability (microscale set-up for, for example, rarely available interacting species, use of crude or complex samples of biological origin);
- ability to reflect (or mimic) the structural or functional complexity of biological systems [creating pseudophysiological experimental equilibria for measuring and derivation of binding characteristics, including implementation of (proteo)liposomes, artificial membranes, etc., and also optional extrapolation of binding results from in vitro conditions to an in vivo situation];
- screening abilities for direct and rapid screening of combinatorial libraries of drug candidates.

Unlike conventional (e.g. ED, ultrafiltration) or alternative analytical methods suffering from intrinsic methodological restrictions (e.g. spectroscopy, NMR), particularly (affinity) HPLC and CE are of general applicability for monitoring of biospecific intermolecular interactions.[221] The practical utility of conventional affinity HPLC is somewhat limited, since relatively large (milligram) amounts of purified protein/biopolymer are needed [applicability almost exclusively to binding studies with albumins.[94,103,222–225]]. Of other HPLC methods for screening of binding affinity, HPLC/FA is undoubtedly the most robust alternative, especially with regard to the complexity of the biological matrix applied and the possibility of coupling FA with a stereoselective assay via different column-switching devices.

On the other hand, it seems likely that different variants of (affinity) CE may play a pivotal role in future methodological developments, since

- they exhibit great experimental versatility for large sample sets;
- several orders of magnitude smaller amounts (volumes) of ligands and purified but also nonpurified biosamples (membranes, liposomes, biomolecular ''aggregates'', vesicles, cells) are required compared with HPLC;
- binding studies can be performed in ''free solution'', nondenaturing ''pseudophysiological'' buffered aqueous media, hence a broad variety of drug–protein, but also ligand (drug)–DNA, protein–protein, (drug)–vesicles–protein and (drug)–cell–cell interactions may be studied by this methodology.

As drugs or proteins may be implemented as buffer additives and immobilization of interacting species in CE is only optional (if needed for specific purposes), derivation of binding characteristics and the re-establishment of ''starting'' conditions for repetitive analyses are relatively uncomplicated. The dynamically growing number of applications of ACE underlines further that the CE approach offers effective and convenient toolkits, even if only minute amounts of biopolymer are available or if multiple equilibria, nonspecific (weak) or stereoselective binding interactions are considered. Further technical improvements (e.g. laser detection systems, multicapillary arrays) and their practical availability will probably broaden further the use of CE for both ''nonroutine'' and ''routine'' applications.

The most challenging future prospects are represented by the on-line coupling of microscale bioanalytical methods with desired combinatory potential (e.g. microdialysis and tandem CE coupled with MS) bridging, for example, the ''in vivo approach'' with highly advanced analytical tools. Most recently, CE has been carried out in microchannels (made by photolithographic etching) in a planar chip of glass or fused silica.[137] The advantages of such CE-based competitive immunoassays (using fluorescence-labeled haptenic antigens and LIF) include, in addition to the integrally anchored experimental flexibility, the small amount of sample required (less than 1 amol), short separation times (from 30 s to 1 min), ability to determine multiple analytes simultaneously (drug-screening devices for combinatorial ligand libraries), high sensitivity (LOD of 0.1 nmol L^{-1}) and accuracy of the instrumentation. Broadening the field of applicability of these or similar techniques in the future may be forecast and this will certainly impact on further achievements in elucidating the biorecognition properties of different biomolecular species, systems and processes.

ABBREVIATIONS AND ACRONYMS

ACE	Affinity Capillary Electrophoresis
AGE	Agarose Gel Electrophoresis
AGP	α_1-Acid Glycoprotein
BGE	Background Electrolyte
BIA	Biomolecular Interaction Analysis
BSA	Bovine Serum Albumin
CAGE	Capillary Affinity Gel Electrophoresis
CD	Circular Dichroism
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
ED	Equilibrium Dialysis
ELISA	Enzyme-linked Immunosorbent Assay
EOF	Electroosmotic Flow
FA	Frontal Analysis
FFA	Free Fatty Acids
FFF	Field Flow Fractionation
FPIA	Fluorescence Polarization Immunoassay
HD	Hummel–Dreyer
HPCE/FA	High-performance Capillary Electrophoresis/Frontal Analysis
HPFA	High-performance Frontal Analysis
HPLC	High-performance Liquid Chromatography
HSA	Human Serum Albumin
IAM	Immobilized Artificial Membrane
ISRP	Internal Surface Reversed-phase
LIF	Laser-induced Fluorescence
LIFID	Laser-induced Fluorescence Immunodetection
LOD	Limit of Detection
MAbs	Monoclonal Antibodies
MS	Mass Spectrometry
nAchR	Nicotinic Acetylcholine Receptor
NMR	Nuclear Magnetic Resonance
NSAID	Nonsteroidal Anti-inflammatory Drug
ORD	Optical Rotatory Dispersion
PSP	Protein Stationary Phase
RIA	Radioimmunoassay
SDS	Sodium Dodecyl Sulfate
UV	Ultraviolet
VACE	Vacancy Affinity Capillary Electrophoresis
VP	Vacancy Peak

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of

Forensic Science **(Volume 5)**
Chiroptical Spectroscopy in Drug Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Chemical Reagents and Derivatization Procedures in Drug Analysis • Chiral Purity in Drug Analysis

REFERENCES

1. D.H. Williams, M.J. Stone, R.J. Mortishire-Smith, P.R. Hauck, 'Molecular Recognition by Secondary Metabolites', *Biochem. Pharmacol.*, **40**, 27–34 (1990).
2. J.M. Ford, W.N. Hait, 'Pharmacology of Drugs that Alter Multidrug Resistance in Cancer', *Pharmacol. Rev.*, **42**, 155–199 (1990).
3. A.R. Safa, C.J. Glover, J.L. Sewell, M.B. Meyers, J.L. Biedler, R.L. Felsted, 'Identification of the Multidrug Resistance-related Membrane Glycoprotein as an Acceptor for Calcium Channel Blockers', *J. Biol. Chem.*, **262**, 7884–7888 (1987).
4. P.R. Twentyman, 'Cyclosporins as Drug Resistance Modifiers', *Biochem. Pharmacol.*, **43**, 109–117 (1992).
5. P. Schaeffer, C. Lugnier, J.-C. Stoclet, 'Interactions of Calmodulin Antagonists with Calcium Binding Sites', *Eur. J. Pharmacol., Mol. Pharmacol. Sect.*, **206**, 325–332 (1991).
6. I. Fitos, J. Visy, J. Kardos, M. Simonyi, 'Stereoselective Binding of Benzodiazepines and Related Compounds on Calmodulin-sepharose', *Chirality*, **7**, 69–73 (1995).
7. D.G. Reid, L.K. McLachlan, S.P. Robinson, P. Camilleri, C.A. Dyke, C.J. Thorpe, 'Calmodulin Discriminates Between the Two Enantiomers of the Receptor-operated Calcium Channel Blocker SK&F 96365: a Study Using ^{1}H-NMR and Chiral HPLC', *Chirality*, **2**, 229–232 (1990).
8. J.H. Lin, E.H. Ulm, L.E. Los, 'Dose-dependent Stereopharmacokinetics of 5,6-Dihydro-4H-4-(isobutylamino) thieno(2,3-B)thiopyran-2-sulfonamide-7,7-dioxide, a Potent Carbonic Anhydrase Inhibitor, in Rats', *Drug Metab. Dispos.*, **19**, 233–238 (1991).
9. J.H. Lin, I.-W. Chen, T.-H. Lin, 'Species-dependent Stereopharmacokinetics of MK-927, a Potent Carbonic Anhydrase Inhibitor', *Drug Metab. Dispos.*, **19**, 816–822 (1991).
10. J.H. Lin, T.-H. Lin, H. Cheng, 'Uptake and Stereoselective Binding of the Enantiomers of MK-927, a Potent Carbonic Anhydrase Inhibitor, by Human Erythrocytes in Vitro', *Pharm. Res.*, **9**, 339–344 (1992).
11. E. Gilad, N. Zisapel, 'High-affinity Binding of Melatonin to Hemoglobin', *Biochem. Mol. Med.*, **56**, 115–120 (1995).
12. B. Honore, 'Conformational Changes in Human Serum Albumin Induced by Ligand Binding', *Pharmacol. Toxicol.*, **66** (Suppl. II), 1–26 (1990).
13. O.J.M. Bos, M.J.E. Fischer, J. Wilting, L.H.M. Janssen, 'Drug-binding and other Physicochemical Properties of a Large Tryptic and a Large Peptic Fragment of Human

Serum Albumin', *Biochim. Biophys. Acta*, **953**, 37–47 (1988).
14. K. Kamikubo, S. Sagata, S. Nakamura, T. Komaki, K. Miura, 'Thyroxine Binding to Human Serum Albumin Immobilized on Sepharose and Effects of Nonprotein Albumin-binding Plasma Constituents', *J. Protein Chem.*, **9**, 461–465 (1990).
15. M.J.E. Fischer, O.J.M. Bos, R.F.V.d. Linden, J. Wilting, L.H.M. Janssen, 'Steroid Binding to Human Serum Albumin and Fragments Thereof. Role of Protein Conformation and Fatty Acid Content', *Biochem. Pharmacol.*, **45**, 2411–2416 (1993).
16. J.M.H. Kremer, J. Wilting, L.H.M. Janssen, 'Drug Binding to Human Alpha-1-acid Glycoprotein in Health and Disease', *Pharmacol. Rev.*, **40**, 1–47 (1988).
17. F. Herve, J.-C. Duche, P. D'Athis, C. Marche, J. Barre, J.-P- Tillement, 'Binding of Disopyramide, Methadone, Dipyridamole, Chlorpromazine, Lignocaine and Progesterone to the Two Main Genetic Variants of Human Alpha1-acid Glycoprotein: Evidence for Drug-binding Sites on Alpha1-acid Glycoprotein', *Pharmacogenetics.*, **6**, 403–415 (1996).
18. W.E. Muller, 'A Common Single Binding Site for Many Psychotropic Drugs on Human Alpha-1-acid Glycoprotein. Therapeutically Relevant Observation?', *Pharmacopsychiatry*, **18**, 257–258 (1985).
19. C. Aussel, 'Presence of Three Binding Sites for Retinoids Bilirubin and Estrogen or Arachidonic Acid on Rat Alpha-fetoprotein', *Tumour Biol.*, **6**, 179–193 (1985).
20. T.A. Hughes, A.O. Gaber, C.E. Montgomery, 'Plasma Distribution of Cyclosporine Within Lipoproteins and ''in Vitro'' Transfer Between Very Low-density Lipoproteins, Low-density Lipoproteins, and High-density Lipoproteins', *Ther. Drug Monit.*, **13**, 289–295 (1991).
21. S. Urien, R. Zini, M. Lemaire, J.-P. Tillement, 'Assessment of Cyclosporine A Interactions with Human Plasma Lipoproteins in Vitro and in Vivo in the Rat', *J. Pharmacol. Exp. Ther.*, **253**, 305–309 (1990).
22. S. Urien., 'Interaction of Drugs with Human Plasma Lipoproteins', in *Protein Binding and Drug Transport*, eds. P. Baumann, C.B. Eap, W.E. Meuller, J.P. Tillement, Schattauer Verlag, New York, 63–75, 1986.
23. J. Oravcova, D. Sojkova, 'The Binding of Two Dihydropyridines to Isolated and Native Plasma Lipoproteins', *J. Pharm. Pharmacol.*, **47**, 75–78 (1995).
24. S.L. Schreiber, G.R. Crabtree, 'The Mechanism of Action of Cyclosporin A and FK506', *Immunol. Today*, **13**, 136–142 (1992).
25. R. Brodersen, B. Honore, F.G. Larsen, 'Serum Albumin – a Non-saturable Carrier', *Acta Pharmacol. Toxicol.*, **54**, 129–133 (1984).
26. U. Kragh-Hansen, 'Molecular Aspects of Ligand Binding to Serum Albumin', *Pharmacol. Rev.*, **33**, 17–53 (1981).
27. P.M. Laduron, 'Stereospecificity in Binding Studies. A Useful Criterion Though Insufficient to Prove the Presence of Receptors', *Biochem. Pharmacol.*, **37**, 37–40 (1988).
28. K.M. Giacomini, T.F. Blaschke, 'Effect of Concentration-dependent Binding to Plasma Proteins on the Pharmacokinetics and Pharmacodynamics of Disopyramide', *Clin. Pharmacokinet.*, **9** (Suppl. 1), 42–48 (1984).
29. M.N. Dudley, W.-C. Shyu, C.H. Nightingale, R. Quintiliani, 'Effect of Saturable Serum Protein Binding on the Pharmacokinetics of Unbound Cefonicid in Humans', *Antimicrob. Agents Chemother.*, **30**, 565–569 (1986).
30. K.M. Giacomini, W.L. Nelson, R.A. Pershe, L. Valdivieso, K. Turner-Tamiyasu, T.F. Blaschke, 'In Vivo Interaction of the Enantiomers of Disopyramide in Human Subjects', *J. Pharmacokinet. Biopharm.*, **14**, 335–356 (1986).
31. F. Jamali, B.W. Berry, M.R. Wright, 'Dose-dependency of Flurbiprofen Enantiomer Pharmacokinetics in the Rat', *J. Pharm. Sci.*, **83**, 1077–1080 (1994).
32. Y.Z. Yang, B. Little, S.R. Meshnick, 'Alkylation of Proteins by Artemisinin Effects of Heme, pH, and Drug Structure', *Biomed. Pharmacol.*, **48**, 569–573 (1994).
33. T.A.G. Noctor, I.W. Wainer, D.S. Hage, 'Allosteric and Competitive Displacement of Drugs from Human Serum Albumin by Octanoic Acid, as Revealed by High-performance Liquid Affinity Chromatography, on a Human Serum Albumin-based Stationary Phase', *J. Chromatogr. B*, **577**, 305–315 (1992).
34. M.H. Rahman, T. Maruyama, T. Okada, T. Imai, M. Otagiri, 'Study of Interaction of Carprofen and its Enantiomers with Human Serum Albumin – II. Stereoselective Site-to-site Displacement of Carprofen by Ibuprofen', *Biochem. Pharmacol.*, **46**, 1733–1740 (1993).
35. F. Herve, E. Gomas, J.C. Duche, J.P. Tillement, 'Evidence for Differences in the Binding of Drugs to the Two Main Genetic Variants of Human Alpha-1-acid Glycoprotein', *Br. J. Clin. Pharmacol.*, **36**, 241–249 (1993).
36. R. Zini, P. Riant, J. Barre, J.-P. Tillement, 'Disease-induced Variations in Plasma Protein Levels. Implications for Drug Dosage Regimens (Part I)', *Clin. Pharmacokinet.*, **19**, 147–159 (1990).
37. R. Zini, P. Riant, J. Barre, J.-P. Tillement, 'Disease-induced Variations in Plasma Protein Levels. Implications for Drug Dosage Regimens (Part II)', *Clin. Pharmacokinet.*, **19**, 218–229 (1990).
38. D.J. Morgan, A.J. McLean, 'Clinical Pharmacokinetic and Pharmacodynamic Considerations in Patients With Liver Disease. An Update', *Clin. Pharmacokinet.*, **29**, 370–391 (1995).
39. Z.-X. Wang, D.K. Srivastava, 'A Graphical Method for Determining the Number of Essential Sites in Enzymes With Multiple Binding Sites for a Ligand', *Anal. Biochem.*, **216**, 15–26 (1994).
40. Y.-H. Chu, W.L. Lees, A. Stassinopoulos, C.T. Walsh, 'Using Affinity Capillary Electrophoresis to Determine

Binding Stoichiometries of Protein–Ligand Interactions', *Biochemistry*, **33**, 10616–10621 (1994).
41. M. Mammen, F.A. Gomez, G.M. Whitesides, 'Determination of the Binding of Ligands Containing the *N*-2,4-Dinitrophenyl Group to Bivalent Monoclonal Rat Anti-DNP Antibody Using Affinity Capillary Electrophoresis', *Anal. Chem.*, **67**, 3526–3535 (1995).
42. I.M. Klotz, D.L. Hunston, 'Protein Interactions with Small Molecules. Relationships Between Stoichiometric Binding Constants, Site Binding Constants, and Empirical Binding Parameters', *J. Biol. Chem.*, **250**, 3001–3009 (1975).
43. F.J. Sharom, 'The P-glycoprotein Efflux Pump: How Does it Transport Drugs?', *J. Membr. Biol.*, **160**, 161–175 (1997).
44. G. Menke, W. Worner, W. Kratzer, N. Rietbrock, 'Kinetics of Drug Binding to Human Serum Albumin: Allosteric and Competitive Inhibition at the Benzodiazepine Binding Site by Free Fatty Acids of Various Chain Lengths', *Naunyn-Schmiedeberg's Arch. Pharmacol.*, **339**, 42–47 (1989).
45. S. Lobert, A. Frankfurter, J.J. Correia, 'Energestics of *Vinca* Alkaloid Interactions with Tubulin Isotypes: Implications for Drug Efficacy and Toxicity', *Cell Motil. Cytoskeleton*, **39**, 107–121 (1998).
46. H. Takahashi, H. Ogata, S. Kanno, H. Takeuchi, 'Plasma Protein Binding of Propanol Enantiomers as a Major Determinant of Their Stereoselective Tissue Distribution in Rats', *J. Pharmacol. Exp. Ther.*, **252**, 272–278 (1990).
47. S.A. Bai, S.M. Lankford, L.M. Johnson, 'Pharmacokinetics of the Enantiomers of Verapamil in the Dog', *Chirality*, **5**, 436–442 (1993).
48. O. Okazaki, T. Kurata, H. Hakusui, H. Tachizawa, 'Species-related Stereoselective Disposition of Ofloxacin in the Rat, Dog and Monkey', *Xenobiotica*, **22**, 439–450 (1992).
49. J.H. Lin, F.A. DeLuna, E.H. Ulm, D.J. Tocco, 'Species-dependent Enantioselective Plasma Protein Binding of MK-571, a Potent Leukotriene D4 Antagonist', *Drug Metab. Dispos.*, **18**, 484–487 (1990).
50. T. Kosa, T. Maruyama, M. Otagiri, 'Species Differences of Serum Albumins: I. Drug Binding Sites', *Pharm. Res.*, **14**, 1607–1612 (1997).
51. M. Nishikawa, Y. Ohtsubo, J. Ohno, T. Fujita, Y. Koyama, F. Yamashita, M. Hashida, H. Sezaki, 'Pharmacokinetics of Receptor-mediated Hepatic Uptake of Glycosylated Albumin in Mice', *Int. J. Pharm.*, **85**, 75–85 (1992).
52. H. Koyama, N. Sugioka, A. Uno, S. Mori, K. Nakajima, 'Effects of Glycosylation of Hypoglycaemic Drug Binding to Serum Albumin', *Biopharm. Drug Dispos.*, **18**, 791–801 (1997).
53. S. Kishino, A. Nomura, M. Saitoh, M. Sugawara, K. Iseki, A. Kitabatake, K. Miyazaki, 'Single-step Isolation Method for Six Glycoforms of Human Alpha-1-acid Glycoprotein by Hydroxyapatite Chromatography and Study of Their Binding Capacities for Disopyramide', *J. Chromatogr. B*, **703**, 1–6 (1997).
54. R.L. Nation, 'Chirality in New Drug Development. Clinical Pharmacokinetics Considerations', *Clin. Pharmacokinet.*, **27**, 249–255 (1994).
55. A.J. Hutt, S.C. Tan, 'Drug Chirality and Its Clinical Significance', *Drugs*, **52**, 1–12 (1996).
56. E.J.D. Lee, K.M. Williams, 'Chirality. Clinical Pharmacokinetic and Pharmacodynamic Considerations', *Clin. Pharmacokinet.*, **18**, 339–345 (1990).
57. F.P.A. Lehmann, 'A Quantitative Stereo-structure Activity Relationship Analysis of the Binding of Promiscuous Chiral Ligands to Different Receptors', *Quant. Struct.–Act. Relat.*, **6**, 57–65 (1987).
58. Y.-W. Kwon, D.J. Triggle, 'Chiral Aspects of Drug Action at Ion Channels: a Commentary on the Stereoselectivity of Drug Actions at Voltage-gated Ion Channels with Particular Reference to Verapamil Actions at the Ca^{2+} Channel', *Chirality*, **3**, 393–404 (1991).
59. D.J. Triggle, 'On the Other Hand: the Stereoselectivity of Drug Action at Ion Channels', *Chirality*, **6**, 58–62 (1994).
60. R.P. Hof, A. Hof, U.T. Ruegg, N.S. Cook, A. Vogel, 'Stereoselectivity at the Calcium Channel: Different Profiles of Hemodynamic Activity of the Enantiomers of the Dihydropyridine Derivative PN 200-110', *J. Cardiovasc. Pharmacol.*, **8**, 221–226 (1986).
61. R.P. Hof, U.T. Ruegg, A. Hof, A. Vogel, 'Stereoselectivity at the Calcium Channel: Opposite Action of the Enantiomers of a 1,4-Dihydropyridine', *J. Cardiovasc. Pharmacol.*, **7**, 689–693 (1985).
62. W. Schreibmayer, W. Lindner, 'Stereoselective Interactions of (*R*)- and (*S*)-Propafenone with the Cardiac Sodium Channel', *J. Cardiovasc. Pharmacol.*, **20**, 324–331 (1992).
63. G.K. Wang, 'Binding Affinity and Stereoselectivity of Local Anesthetics in Single Batrachotoxin-activated Na^+ Channels', *J. Gen. Physiol.*, **96**, 1105–1127 (1990).
64. N.P. Franks, W.R. Lieb, 'Stereospecific Effects of Inhalational General Anesthetic Optical Isomers on Nerve Ion Channels', *Science*, **254**, 427–430 (1991).
65. D. Tricarico, B. Fakler, W. Spittelmeister, J.P. Ruppensberg, R. Stutzel, C. Franchini, V. Tortorella, D. Conte-Camerino, R. Rudel, 'Stereoselective Interaction of Tocainide and Its Chiral Analogs With the Sodium Channels in Human Myoballs', *Pfluegers Arch.*, **418**, 234–237 (1991).
66. R.J. Hill, H.J. Duff, R.S. Sheldon, 'Determinants of Stereospecific Binding of Type I Antiarrhythmic Drugs to Cardiac Sodium Channels', *Mol. Pharmacol.*, **34**, 659–663 (1988).
67. P.M. Colangelo, R.A. Blouin, J.E. Steinmetz, P.J. McNamara, A.N. DeMaria, P.J. Wedlund, 'Age and Beta-adrenergic Receptor Sensitivity to *S*(−)- and

R,*S*(±)-Propanol in Humans', *Clin. Pharmacol. Ther.*, **51**, 549–554 (1992).
68. W.E. Muller, 'Stereoselective Plasma Protein Binding of Drugs, in *Drug Stereochemistry. Analytical Methods and Pharmacology.*, eds. I.W. Wainer, D.E. Drayer, Marcel Dekker, New York, 227–244, 1988.
69. F. Brunner, W.E. Müller, 'The Stereoselectivity of the ''Single Drug Binding Site'' of Human Alpha1-acid Glycoprotein (Orosomucoid)', *J. Pharm. Pharmacol.*, **39**, 986–990 (1987).
70. M. Eichelbaum, 'Pharmacokinetic and Pharmacodynamic Consequences of Stereoselective Drug Metabolism in Man', *Biochem. Pharmacol.*, **37**, 93–96 (1988).
71. H.K. Kroemer, C. Fischer, C.O. Meese, M. Eichelbaum, 'Enantiomer/Enantiomer Interaction of (*S*)- and (*R*)-Propafenone for Cytochrome P450IID6-catalyzed 5-Hydroxylation: in Vitro Evaluation of the Mechanism', *Mol. Pharmacol.*, **40**, 135–142 (1991).
72. M. Eichelbaum, L. Bertilsson, A. Kupfer, E. Steiner, C.O. Meese, 'Enantioselectivity of 4-Hydroxylation in Extensive and Poor Metabolizer of Debrisoquine', *Br. J. Clin. Pharmacol.*, **25**, 505–508 (1988).
73. P.J. Hayball, R.L. Nation, F. Bochner, L.N. Sansom, M.J. Ahem, M.D. Smith, 'The Influence of Renal Function on the Enantioselective Pharmacokinetics and Pharmacodynamics of Ketoprofen in Patients with Rheumatoid Arthritis', *Br. J. Clin. Pharmacol.*, **36**, 185–193 (1993).
74. J.J. Lima, 'Species-dependent Binding of Disopyramide Enantiomers', *Drug Metab. Dispos.*, **16**, 563–567 (1988).
75. P.J. Hayball, R.L. Nation, F. Bochner, 'Stereoselective Interactions of Ketoprofen Glucuronides With Human Plasma Protein and Serum Albumin', *Biochem. Pharmacol.*, **44**, 291–299 (1992).
76. C. Volland, H. Sun, J. Dammeyer, L.Z. Benet, 'Stereoselective Degradation of the Fenoprofen Acyl Glucuronide Enantiomers and Irreversible Binding to Plasma Protein', *Drug Metab. Dispos.*, **19**, 1080–1086 (1991).
77. J.X. Mazoit, L.S. Cao, K. Samii, 'Binding of Bupivacaine to Human Serum Proteins, Isolated Albumin and Isolated Alpha-1-acid Glycoprotein. Differences Between the Two Enantiomers are Partly due to Cooperativity', *J. Pharmacol. Exp. Ther.*, **276**, 109–115 (1996).
78. I. Fitos, Z. Tegyey, M. Simonyi, M. Kajtar, 'Stereoselective Allosteric Interaction in the Binding of Lorazepam Methylether and Warfarin to Human Serum Albumin', Presented at the 4th FECHEM Conference on Heterocycles in Bio-organic Chemistry, Houthalen, Belgium, 1986.
79. I. Fitos, M. Simonyi, 'Stereoselective Effect of Phenprocoumon Enantiomers on the Binding of Benzodiazepines to Human Serum Albumin', *Chirality*, **4**, 21–23 (1992).
80. C. Martin, G. Berridge, C.F. Higgins, R. Callaghan, 'The Multi-drug Resistance Reversal Agent SR33557 and Modulation of *Vinca* alkaloid Binding to P-glycoprotein by an Allosteric Interaction', *Br. J. Pharmacol.*, **122**, 765–771 (1997).
81. A.B. Shapiro, 'Positively Cooperative Sites for Drug Transport by P-glycoprotein with Distinct Drug Specificities', *Eur. J. Biochem.*, **250**, 130–137 (1997).
82. H. Kurz, H. Trunk, B. Weitz, 'Evaluation of Methods to Determine Protein-binding of Drugs: Equilibrium Dialysis, Ultrafiltration, Ultracentrifugation, Gel Filtration', *Drug Res.*, **27**, 1373–1380 (1977).
83. W.F. Bowers, S. Fulton, J. Thompson, 'Ultrafiltration vs Equilibrium Dialysis for Determination of Free Fraction', *Clin. Pharmacokinet.*, **9** (Suppl. 1), 49–60 (1984).
84. J.-D. Huang, 'Errors in Estimating the Unbound Fraction of Drugs Due to the Volume Shift in Equilibrium Dialysis', *J. Pharm. Sci.*, **72**, 1368–1369 (1983).
85. W.W. Mapleson, 'Computation of the Effect of Donnan Equilibrium on pH in Equilibrium Dialysis', *J. Pharmacol. Methods.*, **17**, 231–242 (1987).
86. R.A. Fois, J.A. Ashley, 'Drug Binding to Apparatus: a Factor Controlling Time to Equilibrium in Equilibrium Dialysis Studies', *J. Pharm. Sci.*, **80**, 300–302 (1991).
87. J.A. Henry, S.N. Mitchell, 'Effect of pH on Human Plasma Protein Binding of a Series of Beta-adrenoceptor Antagonists', *Br. J. Clin. Pharmacol.*, **11**, 119P–120P (1981).
88. O. Brors, O.G. Nilsen, G. Sager, D. Sandnes, S. Jacobsen, 'Influence of pH and Buffer Type on Drug Binding in Human Serum', *Clin. Pharmacokinet.*, **9**, 85–86 (1984).
89. O. Brors, S. Jacobsen, 'pH Lability in Serum During Equilibrium Dialysis', *Br. J. Clin. Pharmacol.*, **20**, 85–88 (1985).
90. M. Oellerich, H. Muller-Vahl, 'The EMIT FreeLevel Ultrafiltration Technique Compared with Equilibrium Dialysis and Ultracentrifugation to Determine Protein Binding of Phenytoin', *Clin. Pharmacokinet.*, **9** (Suppl. 1), 61–70 (1984).
91. R.K. Verbeeck, J.-A. Cardinal, 'Plasma Protein Binding of Salicylic Acid, Phenytoin, Chlorpromazine, Propanol and Pethidine Using Equilibrium Dialysis and Ultracentrifugation', *Arzneim.-Forsch.*, **35**, 903–906 (1985).
92. J. Barre, J.M. Chamouard, G. Houin, J.-P. Tillement, 'Equilibrium Dialysis, Ultrafiltration and Ultracentrifugation Compared for Determining the Plasma-protein-binding Characteristics of Valproic Acid', *Clin. Chem.*, **31**, 60–64 (1985).
93. J. Clausen, M.H. Bickel, 'Prediction of Drug Distribution in Distribution Dialysis and in Vivo from Binding to Tissues and Blood', *J. Pharm. Sci.*, **82**, 345–349 (1993).
94. I.W. Wainer, R. Kaliszan, T.A.G. Noctor, 'Biochromatography Using Immobilized Biopolymers: a New Approach to the Determination of Pharmacological Properties', *J. Pharm. Pharmacol.*, **45**, 367–373 (1993).
95. Q. Yang, P. Lundahl, 'Immobilized Proteoliposome Affinity Chromatography for Quantitative Analysis of Specific Interactions Between Solutes and Membrane

Proteins. Interaction of Cytochalasin B and D-glucose with the Glucose Transporter Glut1', *Biochemistry*, **34**, 7289–7294 (1995).
96. W.K. Chui, I.W. Wainer, 'Enzyme-based High-performance Liquid Chromatography Supports as Probes of Enzyme Activity and Inhibition: the Immobilization of Trypsin and Alpha-chymotrypsin on an Immobilized Artificial Membrane High-performance Liquid Chromatography Support', *Anal. Biochem.*, **201**, 237–245 (1992).
97. G. Massolini, A.F. Aubry, A. McGann, I.W. Wainer, 'Determination of the Magnitude and Enantioselectivity of Ligand Binding to Rat and Rabbit Serum Albumins Using Immobilized High-performance Liquid Chromatography Stationary Phases', *Biochem. Pharmacol.*, **46**, 1285–1293 (1993).
98. F. Beigi, Q. Yang, P. Lundahl, 'Immobilized-liposome Chromatographic Analysis of Drug Partitioning into Lipid Bilayers', *J. Chromatogr. A*, **704**, 315–321 (1995).
99. Y. Zhang, S. Aimoto, L. Lu, Q. Yang, P. Lundahl, 'Immobilized Liposome Chromatography for Analysis of Interactions Between Lipid Bilayers and Peptides', *Anal. Biochem.*, **229**, 291–298 (1995).
100. T.A.G. Noctor, M.J. Diaz-Perez, I.W. Wainer, 'Use of a Human Serum Albumin-based Stationary Phase for High-performance Liquid Chromatography as a Tool for the Rapid Determination of Drug–Plasma Protein Binding', *J. Pharm. Sci.*, **82**, 675–676 (1993).
101. T.A.G. Noctor, C.D. Pham, R. Kaliszan, I.W. Wainer, 'Stereochemical Aspects of Benzodiazepine Binding to Human Serum Albumin. I. Enantioselective High Performance Liquid Affinity Chromatographic Examination of Chiral and Achiral Binding Interactions Between 1,4-benzodiazepines and Human Serum Albumin', *Mol. Pharmacol.*, **42**, 506–511 (1992).
102. I. Fitos, Z. Tengyey, M. Simonyi, I. Sjoholm, T. Larsson, C. Lagercrantz, 'Stereoselective Binding of 3-Acetoxy-, and 3-Hydroxy-1,4-benzodiazepine-2-ones to Human Serum Albumin', *Biochem. Pharmacol.*, **35**, 263–269 (1986).
103. E. Domenici, C. Bertucci, P. Salvadori, I.W. Wainer, 'Use of a Human Serum Albumin-based High-performance Liquid Chromatography Chiral Stationary Phase for the Investigation of Protein Binding: Detection of the Allosteric Interaction Between Warfarin and Benzodiazepine Binding Sites', *J. Pharm. Sci.*, **80**, 164–166 (1991).
104. R. Kaliszan, T.A.G. Noctor, I.W. Wainer, 'Stereochemical Aspects of Benzodiazepine Binding to Human Serum Albumin. II. Quantitative Relationships Between Structure and Enantioselective Retention in High-performance Liquid Affinity Chromatography', *Mol. Pharmacol.*, **42**, 512–517 (1992).
105. T.A.G. Noctor, G. Felix, I.W. Wainer, 'Stereochemical Resolution of Enantiomeric 2-Arylpropionic Acid Non-steroidal Anti-inflammatory Drugs on a Human Serum Albumin-based High-performance Liquid Chromatographic Chiral Stationary Phase', *Chromatographia*, **31**, 55–59 (1991).
106. I. Fitos, J. Visy, A. Magyar, J. Kajtar, M. Simonyi, 'Inverse Stereoselectivity in the Binding of Acenocoumarol to Human Serum Albumin and to Alpha1-acid Glycoprotein', *Biochem. Pharmacol.*, **38**, 2259–2262 (1989).
107. I. Fitos, J. Visy, M. Simonyi, 'Binding of *Vinca* Alkaloid Analogs to Human Serum Albumin and to Alpha1-acid Glycoprotein', *Biochem. Pharmacol.*, **41**, 377–383 (1991).
108. B. Sebille, 'Methods of Drug Protein Binding Determinations', *Fundam. Clin. Pharmacol.*, **4**, 151s–161s (1990).
109. B. Sebille, N. Thuaud, J.-P. Tillement, 'Study of Binding of Low-molecular-weight Ligand to Biological Macromolecules by High-performance Liquid Chromatography. Evaluation of Binding Parameters for Two Drugs Bound to Human Serum Albumin', *J. Chromatogr.*, **167**, 159–170 (1978).
110. B. Sebille, N. Thuaud, J.-P. Tillement, 'Equilibrium Saturation Chromatographic Method for Studying the Binding of Ligands to Human Serum Albumin by High-performance Liquid Chromatography. Influence of Fatty Acids and Sodium Dodecyl Sulfate on Warfarin–Human Serum Albumin Binding', *J. Chromatogr.*, **180**, 103–110 (1979).
111. T.C. Pinkerton, K.A. Koeplinger, 'Determination of Warfarin–Human Serum Albumin Protein Binding Parameters by an Improved Hummel–Dreyer High-performance Liquid Chromatographic Method Using Internal Surface Reversed-phase Columns', *Anal. Chem.*, **62**, 2114–2122 (1990).
112. S.F. Sun, C.L. Hsiao, 'Hummel–Dreyer Method in High-performance Liquid Chromatography for the Determination of Drug–Protein Binding Parameters', *J. Chromatogr.*, **648**, 325–331 (1993).
113. N. Thuaud, B. Sebille, M.H. Livertoux, J. Bessiere, 'Determination of Diazepam–Human Serum Albumin Binding by Polarography and High-performance Liquid Chromatography at Different Protein Concentrations', *J. Chromatogr.*, **282**, 509–518 (1983).
114. A. Shibukawa, M. Nagao, A. Terakita, J. He, T. Nakagawa, 'High-performance Frontal Analysis/High-performance Liquid Chromatographic system for the Enantioselective Determination of Unbound Fenoprofen Concentration in Protein Binding Equilibrium', *J. Liq. Chromatogr.*, **16**, 903–914 (1993).
115. A. Shibukawa, M. Nagao, Y. Kuroda, T. Nakagawa, 'Stereoselective Determination of Free Warfarin Concentration in Protein Binding Equilibrium Using Direct Sample Injection and an On-line Liquid Chromatographic System', *Anal. Chem.*, **62**, 712–716 (1990).
116. A. Shibukawa, A. Terakita, J.-Y. He, T. Nakagawa, 'High-performance Frontal Analysis–High-performance Liquid Chromatographic System for

Stereoselective Determination of Unbound Ketoprofen Enantiomers in Plasma After Direct Sample Injection', *J. Pharm. Sci.*, **81**, 710–715 (1992).
117. A. Shibukawa, T. Nakagawa, N. Nishimura, M. Miyake, H. Tanaka, 'Determination of Free Drug in Protein Binding Equilibrium by High-performance Frontal Analysis Using Internal-surface Reversed-phase Silica Support', *Chem. Pharm. Bull.*, **37**, 702–706 (1989).
118. A. Shibukawa, N. Nishimura, K. Nomura, Y. Kuroda, T. Nakagawa, 'Simultaneous Determination of the Free and Total Carbamazepine Concentrations in Human Plasma by High-performance Frontal Analysis Using an Internal-surface Reversed-phase Silica Support', *Chem. Pharm. Bull.*, **38**, 443–447 (1990).
119. A. Shibukawa, Y. Yoshimoto, T. Ohara, T. Nakagawa, 'High-performance Capillary Electrophoresis/Frontal Analysis for the Study of Protein Binding of a Basic Drug', *J. Pharm. Sci.*, **83**, 616–619 (1994).
120. A. Shibukawa, C. Nakao, T. Sawada, A. Terakita, N. Morokoshi, T. Nakagawa, 'Determination of the Unbound Concentration of Hydrophobic Drugs in Albumin Solutions by High-performance Frontal Analysis Using a Diol–silica Column', *J. Pharm. Sci.*, **83**, 868–873 (1994).
121. T. Ohshima, I. Johno, T. Hasegawa, S. Kitazawa, 'Direct Determination of Protein Binding of Phenytoin in Serum by High-performance Liquid Chromatography', *J. Pharm. Sci.*, **79**, 77–81 (1990).
122. B. Sebille, N. Thuaud, 'Determination of Tryptophan–Human Serum Albumin Binding from Retention Data and Separation of Tryptophan Enantiomer by High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **3**, 299–308 (1980).
123. B. Sebille, N. Thuaud, J.-P. Tillement, 'Retention Data Methods for the Determination of Drug–Protein Binding Parameters by High-performance Liquid Chromatography', *J. Chromatogr.*, **204**, 285–291 (1981).
124. I. Marle, C. Petterson, T. Arvidsson, 'Determination of Binding Affinity of Enantiomers to Albumin by Liquid Chromatography', *J. Chromatogr.*, **456**, 323–336 (1988).
125. S.E. Cook, T.C. Pinkerton, 'Characterization of Internal Surface Reversed-phase Silica Supports for Liquid Chromatography', *J. Chromatogr.*, **368**, 233–248 (1986).
126. I.H. Hagestam, T.C. Pinkerton, 'Internal Surface Reversed-phase Silica Supports for Liquid Chromatography', *Anal. Chem.*, **57**, 1757–1763 (1985).
127. J. Oravcova, D. Sojkova, E. Bencsikova, P. Bohov, T. Trnovec, 'Stereoselective Plasma Protein Binding of Isradipine', *Chirality*, **7**, 167–172 (1995).
128. J. Oravcova, D. Sojkova, W. Lindner, 'Comparison of the Hummel–Dreyer Method in High-performance Liquid Chromatography and Capillary Electrophoresis Conditions for Study of the Interaction of (*RS*)-, (*R*)- and (*S*)-Carvedilol with Isolated Plasma Proteins', *J. Chromatogr. B*, **682**, 349–357 (1996).
129. J. Oravcova, S. Bystricky, T. Trnovec, 'Different Binding of Propanol Enantiomers to Human Alpha-1-acid Glycoprotein', *Biochem. Pharmacol.*, **38**, 2575–2579 (1989).
130. N. Nishimura, A. Shibukawa, T. Nakagawa, 'High-performance Frontal Analysis for the Determination of the Free Drug Concentration in Protein Binding Equilibrium Using Porous Polymer Stationary Phase', *Anal. Sci.*, **6**, 355–359 (1990).
131. A. Shibukawa, M. Kadohara, J. He, M. Nishimura, S. Naito, T. Nakagawa, 'Erratum to Study of Enantioselective Binding Between BOF-4272 and Serum Albumins by Means of High-performance Frontal Analysis', *J. Chromatogr. A*, **708**, 364 (1995).
132. M.E.R. Rosas, A. Shibukawa, K. Ueda, T. Nakagawa, 'Enantioselective Protein Binding of Semotiadil and Levosemotiadil Determined by High-performance Frontal Analysis', *J. Pharm. Biomed. Anal.*, **15**, 1595–1601 (1997).
133. D.S. Hage, 'Chiral Separations in Capillary Electrophoresis Using Proteins as Stereoselective Binding Agents', *Electrophoresis*, **18**, 2311–2321 (1997).
134. D.K. Lloyd, A.-F. Aubry, E.D. Lorenzi, 'Selectivity in Capillary Electrophoresis: the Use of Proteins', *J. Chromatogr. A*, **792**, 349–369 (1997).
135. G. Rippel, H. Corstjens, H.A.H. Billiet, J. Frank, 'Affinity Capillary Electrophoresis', *Electrophoresis*, **18**, 2175–2183 (1997).
136. S. Bose, J. Yang, D.S. Hage, 'Guidelines in Selecting Ligand Concentrations for the Determination of Binding Constants by Affinity Capillary Electrophoresis', *J. Chromatogr. B*, **697**, 77–88 (1997).
137. K. Shimura, K. Kasai, 'Affinity Capillary Electrophoresis: a Sensitive Tool for the Study of Molecular Interactions and Its Use in Microscale Analyses', *Anal. Biochem.*, **251**, 1–16 (1997).
138. D.K. Lloyd, S. Li, P. Ryan, 'Investigation of Enantioselective Ligand–Protein Binding and Displacement Interactions Using Capillary Electrophoresis', *Chirality*, **6**, 230–238 (1994).
139. A. Ahmed, H. Ibrahim, F. Pastore, D.K. Lloyd, 'Relationship Between Retention and Effective Selector Concentration in Affinity Capillary Electrophoresis and High-performance Liquid Chromatography', *Anal. Chem.*, **68**, 3270–3273 (1996).
140. D.K. Lloyd, A. Ahmed, F. Pastore, 'A Quantitative Relationship Between Capacity Factor and Selector Concentration in Capillary Electrophoresis and High-performance Liquid Chromatography: Evidence from the Enantioselective Resolution of Benzoin Using Human Serum Albumin as a Chiral Selector', *Electrophoresis*, **18**, 958–964 (1997).
141. T. Arai, N. Nimura, T. Kinoshita, 'Investigation of Enantioselective Ofloxacin–Albumin Binding and Displacement Interactions Using Capillary Affinity Zone Electrophoresis', *Biomed. Chromatogr.*, **9**, 68–74 (1995).
142. T. Arai, M. Ichinose, H. Kuroda, N. Nimura, T. Kinoshita, 'Chiral Separation by Capillary Affinity Zone

Electrophoresis Using an Albumin-containing Support Electrolyte', *Anal. Biochem.*, **217**, 7–11 (1994).
143. G.E. Barker, P. Russo, R.A. Hartwick, 'Chiral Separation of Leucovorin with Bovine Serum Albumin Using Affinity Capillary Electrophoresis', *Anal. Chem.*, **64**, 3024–3028 (1992).
144. N. Wu, G.E. Barker, C.H. Huie, 'Separation of Porphyrins and Porphyrin Isomers in Capillary Electrophoresis Using Mixed Ionic Surfactant–Bovine Serum Albumin Buffer System', *J. Chromatogr. A*, **659**, 435–442 (1994).
145. Z. Simek, R. Vespalec, 'Interpretation of Enantioselective Activity of Albumin Used as the Chiral Selector in Liquid Chromatography and Electrophoresis', *J. Chromatogr. A*, **685**, 7–14 (1994).
146. S. Busch, J.C. Kraak, H. Poppe, 'Chiral Separations by Complexation with Proteins in Capillary Zone Electrophoresis', *J. Chromatogr. A*, **635**, 119–126 (1993).
147. J. Yang, D.S. Hage, 'Chiral Separations in Capillary Electrophoresis Using Human Serum Albumin as a Buffer Additive', *Anal. Chem.*, **66**, 2719–2725 (1994).
148. Y.H. Chu, G.M. Whitesides, 'Affinity Capillary Electrophoresis can Simultaneously Measure Binding Constants of Multiple Peptides to Vancomycin', *J. Org. Chem.*, **57**, 3524–3525 (1992).
149. J.L. Carpenter, P. Camilleri, D. Dhanak, D. Goodall, 'A Study of the Binding of Vancomycin to Dipeptides Using Capillary Electrophoresis', *J. Chem. Soc. Chem. Commun.*, 804–806 (1992).
150. A.M. Stalcup, N.M. Agyel, 'Heparin: a Chiral Mobile-phase Additive for Capillary Zone Electrophoresis', *Anal. Chem.*, **66**, 3054–3059 (1994).
151. A.M. Stalcup, K.H. Gahm, M. Baldueza, 'Chiral Separation of Chloroquine Using Heparin as a Chiral Selector in High-performance Liquid Chromatography', *Anal. Chem.*, **68**, 2248–2250 (1996).
152. S. Terabe, K. Otsuka, H. Nishi, 'Separation of Enantiomers by Capillary Electrophoretic Techniques', *J. Chromatogr. A*, **666**, 295–319 (1994).
153. P. Sun, A. Hoops, R.A. Hartwick, 'Enhanced Albumin Protein Separations and Protein–Drug Binding Constant Measurements Using Anti-inflammatory Drugs as Run Buffer Additives in Capillary Electrophoresis', *J. Chromatogr. A*, **661**, 335–340 (1994).
154. K. Shimura, K. Kasai, 'Determination of the Affinity Constants of Concanavalin A for Monosaccharides by Fluorescence Affinity Probe Capillary Electrophoresis', *Anal. Biochem.*, **227**, 186–194 (1995).
155. K. Shimura, K. Kansai, 'Determination of the Affinity Constants of Pea Lectin for Neutral Sugars by Capillary Affinophoresis With a Monoligand Affinophore', *J. Biochem.*, **120**, 1146–1152 (1996).
156. N.H.H. Heegaard, 'Determination of Antigen–Antibody Affinity by Immuno-capillary Electrophoresis', *J. Chromatogr. A*, **680**, 405–412 (1994).
157. Y.H. Chu, L.Z. Avila, H.A. Biebuyck, G.M. Whitesides, 'Use of Affinity Capillary Electrophoresis to Measure Binding Constants of Ligands to Proteins', *J. Med. Chem.*, **35**, 2915–2917 (1992).
158. L.Z. Avila, Y.H. Chu, E.C. Blossey, G.M. Whitesides, 'Use of Capillary Electrophoresis to Determine Kinetic and Equilibrium Constants for Binding of Arylsulfonamides to Bovine Carbonic Anhydrase', *J. Med. Chem.*, **36**, 126–133 (1993).
159. F.A. Gomez, J.N. Mirkovich, V.M. Dominguez, K.W. Liu, D.M. Macias, 'Multiple-plug Binding Assays Using Affinity Capillary Electrophoresis', *J. Chromatogr. A*, **727**, 291–299 (1996).
160. N.H.H. Heegaard, F.A. Robey, 'A Capillary Electrophoresis-based Assay for the Binding of Ca^{2+} and Phosphorylcholine to Human C-reactive Protein', *J. Immunol. Methods*, **166**, 103–110 (1993).
161. J. Liu, K.J. Volk, M.S. Lee, E.H. Kerns, I.E. Rosenberg, 'Affinity Capillary Electrophoresis Applied to the Studies of Interactions of a Member of Heat Shock Protein Family with an Immunosupressant', *J. Chromatogr. A*, **680**, 395–403 (1994).
162. K. Nadeau, S.G. Nadler, M. Saulnier, M.A. Tepper, C.T. Walsh, 'Quantitation of the Interaction of the Immunosupressant Deoxyspergualin and Analogs with Hsc70 and Hsp90', *Biochemistry*, **33**, 2561–2567 (1994).
163. N.H.H. Heegaard, P.M.H. Heegaard, P. Roepstorff, F.A. Robey, 'Ligand-binding Sites in Human Serum Amyloid P Component', *Eur. J. Biochem.*, **239**, 850–856 (1996).
164. J. Liu, K.J. Volk, M.S. Lee, M. Pucci, S. Handwerger, 'Binding Studies of Vancomysin to the Cytoplasmic Peptidoglycan Precursors by Affinity Capillary Electrophoresis', *Anal. Chem.*, **66**, 2412–2416 (1994).
165. V. Horejsi, M. Ticha, 'Qualitative and Quantitative Applications of Affinity Electrophoresis for the Study of Protein–Ligand Interactions: a Review', *J. Chromatogr.*, **376**, 49–67 (1986).
166. N.H.H. Heegaard, O.J. Bjerrum, 'Affinity Electrophoresis Used for Determination of Binding Constants for Antibody–Antigen Reactions', *Anal. Biochem.*, **195**, 319–326 (1991).
167. Y.H. Chu, J.K. Chen, G.M. Whitesides, 'Affinity Electrophoresis in Multisectional Polyacrylamide Slab Gels is a Useful and Convenient Technique for Measuring Binding Constants of Aryl Sulfonamides to Bovine Carbonic Anhydrase B', *Anal. Chem.*, **65**, 1314–1322 (1993).
168. M.H.A. Busch., 'Study and Use of Protein Binding in Capillary Zone Electrophoresis', PhD Thesis, University of Amsterdam, 1997.
169. J.C. Kraak, S. Busch, H. Poppe, 'Study of Protein–Drug Binding Using Capillary Zone Electrophoresis', *J. Chromatogr.*, **608**, 257–264 (1992).
170. M.H.A. Busch, H.F.M. Boelens, J.C. Kraak, H. Poppe, A.A.P. Meekel, M. Resmini, 'Critical Evaluation of the Applicability of Capillary Zone Electrophoresis for the

Study of Hapten–Antibody Complex Formation', *J. Chromatogr. A*, **744**, 195–203 (1996).
171. T. Ohara, A. Shibukawa, T. Nakagawa, 'Capillary Electrophoresis/Frontal Analysis for Microanalysis of Enantioselective Protein Binding of a Basic Drug', *Anal. Chem.*, **67**, 3520–3525 (1995).
172. T. Prichett, R.A. Evangelista, F.-T.A. Chen, 'Capillary Electrophoresis-based Immunoassays. A Speedy Alternative to Solid-phase Immunoassays', *Bio/Technology*, **13**, 1449–1450 (1995).
173. Y. Zhang, Y. Xiao, K.J. Kellar, I.W. Wainer, 'Immobilized Nicotinic Receptor Stationary Phase for On-line Chromatographic Determination of Drug–Receptor Affinities', *Anal. Biochem.*, **263**(1), 22–25 (1998).
174. J.-C. Mani, V. Marchi, C. Cucurou, 'Effect of HIV-1 Peptide presentation on the Affinity Constants of Two Monoclonal Antibodies by BIAcore Technology', *Mol. Immunol.*, **31**, 439–444 (1994).
175. N.H.H. Heegaard, D.T. Olsen, K.-L.P. Larsen, 'Immunocapillary Electrophoresis for the Characterization of a Monoclonal Antibody Against DNA', *J. Chromatogr. A*, **744**, 285–294 (1996).
176. Y. Sugiyama, Y. Suzuki, Y. Sawada, S. Kawasaki, T. Beppu, T. Iga, M. Hanano, 'Auramine O as a Fluorescent Probe for the Binding of Basic Drugs to Human Alpha-1-acid Glycoprotein (Alpha-1-AG)', *Biochem. Pharmacol.*, **34**, 821–829 (1985).
177. R. Zini, C. Copigneaux, J.P. Tillement, '1-anilino-naphthalene-8-sulfonic acid (ANS) as a Probe for the Binding of Antidepressant Drugs to Human Alpha-1-acid Glycoprotein (AAG)', in *Alpha-1-acid Glycotein: Genetics, Biochemistry, Physiological Functions, and Pharmacology*, eds. P. Baumann, C.B. Eap, W.E. Muller, J.P. Tillement, Alan R. Liss, New York, 417–421, 1989.
178. M.H. Rahman, T. Maruyama, T. Okada, K. Yamasaki, M. Otagiri, 'Study of Interaction of Carprofen and Its Enantiomers with Human Serum Albumin–I. Mechanism of Binding Studied by Dialysis and Spectroscopic Methods', *Biochem. Pharmacol.*, **46**, 1721–1731 (1993).
179. T. Oida, 'Proton Nuclear Magnetic Resonance Study on the Multimode Interactions of Human Serum Albumin With Drug Molecules', *J. Biochem.*, **100**, 99–113 (1986).
180. T. Oida, '^{1}H-NMR Study on the Interactions of Human Serum Albumin with Free Fatty Acid', *J. Biochem.*, **100**, 1533–1542 (1986).
181. B.G. Jenkins, 'Detection of Site-specific Binding and Cobinding of Ligands to Macromolecules Using ^{19}F NMR', *Life Sci.*, **48**, 1227–1240 (1991).
182. Q. Kleerekoper, W. Liu, D. Choi, J.A. Putkey, 'Identification of Binding Sites for Bepridil and Trifluoperazine on Cardiac Troponin C', *J. Biol. Chem.*, **273**, 8153–8160 (1998).
183. D.A. Keire, S.V.S. Mariappan, J. Peng, D.L. Rabenstein, 'Nuclear Magnetic Resonance Studies of the Binding of Captopriland Penicillamine by Serum Albumin', *Biomed. Pharmacol.*, **46**, 1059–1069 (1993).
184. S.W. Bligh, H.A. Boyle, A.B. McEwen, P.J. Sadler, R.H. Woodham, '^{1}H NMR Studies of Reactions of Copper Complexes with Human Blood Plasma and Urine', *Biochem. Pharmacol.*, **43**, 137–145 (1992).
185. D.B. Davies, V.I. Pahomov, A.N. Veselkov, 'NMR Determination of the Conformational and Drug Binding Properties of the DNA Heptamer 5'-d(GpCpApApGpC) in Aqueous Solution', *Nucleic Acids. Res.*, **25**, 4523–4531 (1997).
186. C.F. Chignell, 'Optical Studies of Drug–Protein Complexes. III. Interaction of Flufenamic Acid and Other *N*-Arylanthranilates with Serum Albumin', *Mol. Pharmacol.*, **5**, 455–462 (1969).
187. C.F. Chignell, 'Optical Studies of Drug–Protein Complexes. II. Interaction of Phenylbutazone and Its Analogues with Human Serum Albumin', *Mol. Pharmacol.*, **5**, 244–252 (1969).
188. C.F. Chignell, D.K. Starkweather, 'Optical Studies of Drug–Protein Complexes. V. The Interaction of Phenylbutazone, Flufenamic Acid, and Dicoumarol with Acetylsalicylic Acid-treated Human Serum Albumin', *Mol. Pharmacol.*, **7**, 229–237 (1971).
189. J. Wilting, M.M. Weideman, A.C.J. Roomer, J.H. Perrin, 'Conformational Changes in Human Serum Albumin Around the Neutral pH from Circular Dichroic Measurements', *Biochim. Biophys. Acta*, **579**, 469–473 (1979).
190. J. Wilting, B.J.T. Hart, J.J.D. Gier, 'The Role of Albumin Conformation in the Binding of Diazepam to Human Serum Albumin', *Biochim. Biophys. Acta*, **626**, 291–298 (1980).
191. O.J.M. Bos, J.P.M. Remijn, M.J.E. Fischer, J. Wilting, L.H.M. Janssen, 'Location and Characterization of the Warfarin Binding Site of Human Serum Albumin. A Comparative Study of Two Large Fragments', *Biochem. Pharmacol.*, **37**, 3905–3909 (1988).
192. S. Watanabe, T. Saito, 'Conformational Effects in the Interaction of Phenylbutazone With Albumin Studied by Circular Dichroism', *Biochem. Pharmacol.*, **43**, 931–935 (1992).
193. J. Wilting, W.F. Van der Giesen, L.H.H. Janssen, M.M. Weideman, M. Otagiri, J.H. Perrin, 'Effect of Albumin Conformation on the Binding of Warfarin to Human Serum Albumin: Dependence of the Binding of Warfarin to Human Serum Albumin on the Hydrogen, Calcium and Chloride Ion Concentrations as Studied by Circular Dichroism, Fluorescence and Equilibrium Dialysis', *J. Biol. Chem.*, **255**, 3032–3037 (1980).
194. S. Wanwimolruk, D.J. Birkett, 'The Effect of N–B Transition of Human Serum Albumin on the Specific Drug–Protein Binding Sites', *Biochem. Biophys. Acta*, **709**, 247–255 (1982).
195. J. Fleitman, J.H. Perrin, 'The Effects of pH, Calcium and Chloride Ions on the Binding of Benoxaprofen to Human Serum Albumin: Circular Dichroic and Dialysis Measurements', *Int. J. Pharm.*, **11**, 227–236 (1982).

196. T. Maruyama, C.C. Lin, K. Yamasaki, T. Miyoshi, T. Imai, M. Yamasaki, M. Otagiri, 'Binding of Suprofen to Human Serum Albumin. Role of the Suprofen Carboxyl Group', *Biochem. Pharmacol.*, **45**, 1017–1026 (1993).
197. M.H. Livertoux, J. Bessiere, 'Polarographic Analysis of Human Serum Albumin–Benzodiazepine Interactions', *Biochem. Bioenerg.*, **17**, 535–548 (1987).
198. H. Aki, M. Yamamoto, 'Thermodynamics of the Binding of Phenothiazines to Human Plasma, Human Serum Albumin and Alpha-1-acid Glycoprotein: a Calorimetric Study', *J. Pharm. Pharmacol.*, **41**, 674–679 (1989).
199. T. Miyoshi, K. Sukimoto, M. Otagiri, 'Investigation of the Interaction Mode of Phenothiazine Neuroleptics with Alpha-1-acid Glycoprotein', *J. Pharm. Pharmacol.*, **44**, 28–33 (1992).
200. N. Shaklai, R.L. Garlick, H.F. Bunn, 'Nonenzymatic Glycosylation of Human Serum Albumin Alters Its Conformation and Function', *J. Biol. Chem.*, **259**, 3812–3817 (1984).
201. J.C. Giddings, 'Field-flow Fractionation: Analysis of Macromolecular, Colloidal and Particulate Materials', *Science*, **260**, 1456–1465 (1993).
202. M.H. Moon, J.C. Giddings, 'Size Distribution of Liposomes by Flow Field-flow Fractionation', *J. Pharm. Biomed. Anal.*, **11**, 911–920 (1993).
203. F. Herve, S. Urien, E. Albengres, J.-C. Duche, J.-P. Tillement, 'Drug Binding in Plasma. A Summary of Recent Trends in the Study of Drug and Hormone Binding', *Clin. Pharmacokinet.*, **26**, 44–58 (1994).
204. W.M. Pardridge, R. Sakiyama, G. Fierer, 'Transport of Propanol and Lidocaine through the Rat Blood Brain Barrier: Primary Role of Globulin-bound Drug', *J. Clin. Invest.*, **71**, 900–908 (1983).
205. W.M. Pardridge, E.M. Landaw, 'Tracer Kinetic Model of Blood–Brain Barrier Transport of Plasma Protein-bound Ligands', *J. Clin. Invest.*, **74**, 745–752 (1984).
206. C.S. Svensson, M.N. Woodruff, J.G. Baxter, D. Lalka, 'Free Drug Concentration Monitoring in Clinical Practice. Rationale and Current Status', *Clin. Pharmacokinet.*, **11**, 450–469 (1986).
207. R.K. Drobitch, C.K. Svensson, 'Therapeutic Drug Monitoring in Saliva. An Update', *Clin. Pharmacokinet.*, **23**, 365–379 (1992).
208. L.A. Dawson, 'Capillary Electrophoresis and Microdialysis: Current Technology and Applications', *J. Chromatogr. B*, **697**, 89–99 (1997).
209. S. Sarre, K.v. Belle, I. Smolders, G. Krieken, Y. Michotte, 'The Use of Microdialysis for the Determination of Plasma Protein Binding of Drugs', *J. Pharm. Biomed. Anal.*, **10**, 735–739 (1992).
210. M.C. Linhares, P.T. Kissinger, 'Capillary Ultrafiltration: in Vivo Sampling Probes for Small Molecules', *Anal. Chem.*, **64**, 2831–2835 (1992).
211. I.M. Klotz, 'Ligand–Receptor Interaction', in *Introduction to Biomolecular Energetics Including Ligand–Receptor Interactions*, ed. I.M. Klotz, Academic Press, Orlando, FL, 103–133, 1986.
212. R. Brodersen, B. Honore, A.O. Pedersen, I.M. Klotz, 'Binding Constants for Ligand–Carrier Complexes', *Trends Pharmacol. Sci.*, **9**, 252–257 (1988).
213. H. Vorum, K. Fisker, R. Brodersen, 'High-affinity Binding of Two Molecules of Warfarin and Phenprocoumon to Human Serum Albumin', *Biochim. Biophys. Acta*, **1205**, 178–182 (1994).
214. R. Brodersen, F. Nielsen, J.C. Christiansen, K. Andersen, 'Characterization of Binding Equilibrium Data by a Variety of Fitted Isotherms', *Eur. J. Biochem.*, **169**, 487–495 (1987).
215. G. Scatchard, 'The Attractions of Proteins for Small Molecules and Ions', *Ann. N.Y. Acad. Sci.*, **51**, 660–672 (1949).
216. J.B. Pedersen, W.E. Lindup, 'Interpretation and Analysis of Receptor Binding Experiments Which Yield Nonlinear Scatchard Plots and Binding Constants Dependent Upon Receptor Concentration', *Biochem. Pharmacol.*, **47**, 179–185 (1994).
217. I.M. Klotz, 'Numbers of Receptor Sites from Scatchard graphs: Facts and Fantasies', *Science*, **217**, 1247–1249 (1982).
218. I.M. Klotz, 'Number of Receptor Sites From Scatchard and Klotz Graphs: A Constructive Critique', *Science*, **220**, 981 (1983).
219. K. Zierler, 'Misuse of Nonlinear Scatchard Plots', *Trends Biochem. Sci.*, **14**, 314–317 (1989).
220. J.C. Kermode, 'The Curvilinear Scatchard Plot. Experimental Artifact or Receptor Heterogeneity?', *Biochem. Pharmacol.*, **38**, 2953–2060 (1989).
221. J. Oravcova, B. Bohs, W. Lindner, 'Drug-binding Studies. New Trends in Analytical and Experimental Methodology', *J. Chromatogr. B*, **677**, 1–28 (1996).
222. E. Domenici, C. Bertucci, P. Salvadori, S. Motellier, I.W. Wainer, 'Immobilized Serum Albumin: Rapid HPLC Probe of Stereoselective Interactions', *Chirality*, **2**, 263–268 (1990).
223. I.W. Wainer, 'Enantioselective High-performance Liquid Affinity Chromatography as a Probe of Ligand–Biopolymer Interactions: an Overview of a Different Use for High-performance Liquid Chromatographic Chiral Stationary Phases', *J. Chromatogr. A*, **666**, 221–234 (1994).
224. D.S. Hage, T.A.G. Noctor, I.W. Wainer, 'Characterization of the Protein Binding of Chiral Drugs by High-performance Affinity Chromatography. Interactions of *R*- and *S*-Ibuprofen with Human Serum Albumin', *J. Chromatogr. A*, **693**, 23–32 (1995).
225. J. Yang, D.S. Hage, 'Role of Binding Capacity Versus Binding Strength in the Separation of Chiral Compounds on Protein-based High-performance Liquid Chromatography Columns. Interactions of D- and L-Tryptophan with Human Serum Albumin', *J. Chromatogr. A*, **725**, 273–285 (1996).

Protein–Oligonucleotide Interactions

Odd Stokke Gabrielsen, Vilborg Matre, and Stine Bergholtz
Department of Biochemistry, University of Oslo, Norway

1 Principles of Protein–Oligonucleotide Interactions **1**
1.1 Protein–DNA Interactions 1
1.2 Protein–RNA Interactions 3
2 Methods Used to Monitor Protein–Oligonucleotide Interactions **3**
2.1 Electrophoretic Mobility Shift Assay 3
2.2 DNA Footprinting Methods 12
Abbreviations and Acronyms **17**
Related Articles **18**
References **18**

The specific binding of a protein to a nucleic acid is a first step in several central processes in a living cell. Sequence-specific protein–DNA interactions are crucial for the functional read-out of genetic information. Sequence recognition is the result of a concerted action of many weak chemical interactions of different types between the protein and its DNA target, including nonspecific electrostatic interactions, hydrogen bonding and van der Waals contacts. The precise complementarity of shape between the two macromolecules facilitates specific chemical recognition to be established. The electrophoretic mobility shift assay (EMSA) and several variants of footprinting are simple electrophoretic methods developed to study protein–DNA interactions. Because the specificity is determined by the nucleic acid sequence, the same methods can be exploited for a wide range of proteins simply by changing the sequence of the nucleic acid. EMSA detects sequence-specific DNA-binding activity in a protein sample as a separate migrating band in a nondenaturating gel. A footprinting method provides more detailed information on the precise location of a bound protein along the DNA fragment through the removal of specific bands in a pattern of cleaved fragments separated by electrophoresis. Both methods are highly sensitive due to the use of radioactively labeled oligonucleotides and can be performed with protein samples of low purity. When combined these methods are capable of providing a picture of the protein–DNA complex with a great deal of molecular detail, surpassed only by the more demanding methods of crystallography and nuclear magnetic resonance (NMR).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 PRINCIPLES OF PROTEIN–OLIGONUCLEOTIDE INTERACTIONS

Nucleic acids function in living cell as molecules that store and transport information. The genetic information encoded in the nucleotide sequences can only be converted into functional processes after interaction with specific proteins that are able to recognize and ''read'' the sequence information. Hence, the specific binding of a protein to a nucleic acid is a first step in several processes where sequence information is controlling cellular functions, like transcription, splicing and DNA repair. Protein–DNA and protein–RNA interactions are therefore instrumental for the functional read-out of genetic information. Although the subjects of research in this area are quite complex and diverse, the basic methods used to study such interactions are rather simple and uniform. This is largely due to the development of uncomplicated and sensitive methods, like EMSA and several variants of footprinting methods. In these procedures, the specificity is determined by the nucleic acid sequences. Hence, the same methods can be exploited for a wide range of proteins simply by changing the sequence of the nucleic acid. The easy access to synthetic oligonucleotides has further facilitated the study of protein–nucleic acid interactions. Finally, several highly developed systems for expression of recombinant proteins and protein domains encoded by cloned genes have given easy access to the many low-abundance proteins involved in reading genetic information.

1.1 Protein–DNA Interactions

The ability of a protein to recognize and bind preferentially to a specific DNA sequence element in the genome is the result of a concerted action of many weak interactions of different types. The principles governing protein–DNA recognition can be described at two levels, as *shape recognition* describing the structures in protein and DNA that form complementary surfaces that are able to interact, and as *chemical recognition* describing the nature and arrangement of chemical bonds involved in nucleotide–amino acid contacts. The precise complementarity of shape between the two macromolecules facilitates specific chemical recognition to be established. The chemical rules can be generalized for any DNA-binding motif, while the stereochemical rules are specific to a particular DNA-binding motif.[1] A brief summary

of some major points is given below. For a more detailed discussion of this topic, several reviews are available.[1–12]

1.1.1 Nonspecific Electrostatic Interactions

Even if a DNA-binding protein can distinguish a few specific recognition sites in the genome from the vast amount of nonspecific sites, it will also bind DNA nonspecifically. It is important also, from a methodological point of view, to realize that all proteins with a sequence-specific binding property also have a significant affinity for DNA in general independent of sequence. This is because electrostatic interactions with the negatively charged phosphates in DNA make a major contribution to the binding affinity for most of these proteins. It is evident that phosphate interaction will be essentially sequence-independent, and therefore cannot account for specificity. In vivo, the high concentration of nonspecific sites cause these proteins to be DNA-bound all the time, irrespective of sequence, at physiological conditions of pH and ionic strength.[13] The ratio of specific versus nonspecific binding constants will often be in the range 10^4–10^6. The electrostatic contribution to the protein–DNA interaction is also the reason why complexes are highly sensitive to the concentrations of mono- and divalent cations. In the absence of proteins the negatively charged phosphodiester chains in DNA will be partially neutralized by counterions with positive charge. In a normal NaCl solution, it has been estimated that about 70% of the negative charges in DNA will be neutralized at any time by Na^+ ions.[10] When a protein binds, the counterions in the binding site will be displaced and liberated to the solution, a process that is associated with an increase in entropy. The liberated counterions will have to be incorporated in the true equilibrium constant for the binding process, with the consequence that the apparent binding constant becomes an exponential function of the cation concentration. The affinity of a protein for its DNA target will therefore drop rapidly with increasing salt concentration. For any specific protein it is not only the number of positive charges that determines its nonspecific affinity for DNA. Affinity will also be determined by the precise geometric location of the charged groups in the protein relative to DNA and to which degree a complementary surface of charge optimized for interaction is generated. In addition, the electrostatic interaction also has a role in orienting the protein for a second docking step where more precise hydrogen bond interactions can be established.[14]

1.1.2 Hydrogen Bonding as Basis for Sequence Recognition

A specific DNA-binding protein has to distinguish its particular recognition sequence in DNA without strand separation. This is possible because each specific base pair exposes a unique pattern of hydrogen bond donors and acceptors in the major groove. Hydrogen bonds can only be formed between closely positioned chemical groups, and therefore the geometry of the interacting protein and DNA structures becomes critical. In B-form DNA the major groove is just wide enough to fit an α-helix. Accordingly, many DNA-binding proteins use an α-helix protruding from the surface of the protein as ''recognition helix'',[4–6,12] although β-sheet DNA-binding motifs also occur.[15] The geometry of an α-helix docked in the major groove allows hydrogen bond formations between side chains protruding from the α-helix and the edges of the base pairs. When these two structures fit, a complementary pattern of hydrogen bonds and van der Waals interactions lead to specific binding. When the protein is in contact with a nonsite, complementarity is not established. Since the side chains in the recognition helix of a protein will form hydrogen bonds with water in the unbound state, the number of hydrogen bonds are not necessarily altered upon specific complex formation. However, this number is probably reduced when the protein docks on to a nonsite, making nonspecific binding unfavorable. A closer inspection of the structures involved led early on to the conclusion that two hydrogen bonds are required for each base to be uniquely recognized by a protein (bidentate contact[16]). Sufficient hydrogen bonding possibilities exist in the major groove for each base pair to expose a unique pattern (ternary code), while the minor groove only allows distinction between AT/TA versus GC/CG base pairs (binary code). The determinations of three-dimensional (3D) structures for many DNA-bound proteins have confirmed the importance of hydrogen bonds in sequence recognition. However, they have underscored the importance of an intricate complementary network of interactions where also water can mediate protein–DNA contacts and thus contribute to the recognition process.[4]

The minor groove in B-form DNA is deep and narrow and thus less accessible for protein contact. However, several examples are described where the long side chain of arginine extends into the minor groove. The groove width will vary with the sequence being narrower in AT-rich DNA than in GC-rich DNA. In A-form DNA (or in duplex RNA segments) the groove geometries are nearly the opposite of B-form DNA, the minor groove being shallow and wide and the major groove deep and narrow. Accordingly, proteins that bind A-form DNA (like many RNA-binding proteins) often interact with the minor groove. It should be noted that the detailed structure of any stretch of DNA will deviate to some extent (groove width, helical periodicity, orientation of bases etc.) from the canonical A- and B-forms and this form is determined by the sequence of base pairs. A specific shape induced

in DNA by a particular sequence can be an important aspect recognized by a DNA-binding protein.[4]

The lesson drawn from the study of many 3D structures of protein–DNA complexes is that it seems impossible to construct a recognition code predicting which amino acid residue will make close contact with which particular base pair.[1] Within specific families of transcription factors some success along these lines has been obtained.[2,6] In general however some overall preferences seem to exist. In many structures the guanidine group of an arginine makes bidentate hydrogen bonds with the edge of a guanine base, and glutamine is a frequent contact point in the recognition of adenine bases, also through double hydrogen bonds. In addition, many examples are described of van der Waals contacts between hydrophobic side chains and the 5-methyl group in thymine.[4]

1.1.3 Structural Changes Induced by Complex Formation

When a protein binds specifically to a DNA sequence element, it is not uncommon to observe induced changes in the local structure of DNA. These changes can be a bending of the DNA duplex, alterations in twist and roll angles, and changes in groove width.[17] How easily a given DNA segment will undergo the induced changes will depend on its sequence. AT-rich tracks will for example prefer to undergo bending that compresses the minor groove, while GC-rich tracks will prefer kinks that make the major groove more narrow. These differences in deformabilities have as consequence an indirect sequence dependence of protein–DNA complexes that comes in addition to hydrogen bond formation.[18] The asymmetrical neutralization of charge by a protein bound along one side of the DNA helix will by itself cause some bending.[19] In addition, architectural factors interact with DNA through minor groove contacts and alter DNA conformation in a more dramatic way.[20] Protein-induced bending in the binding site can very easily be analyzed by versions of the EMSA method described below.

It has been realized that many proteins also undergo conformational changes upon DNA binding. Part of the DNA-binding surface will in these cases be generated during the interaction. The thermodynamics of the 'induced fit' model for DNA recognition, where local folding is coupled to binding, have been treated in detail by Spolar and Record.[21] Recently it was proposed that specific binding sites bound by a transcription factor might in fact act as a ligand for the protein capable of inducing distinct conformations in the factor, some activating and others repressing.[22] In this view, the DNA-induced conformational changes in the transcription factors could be central to their functional output, and these factors can be regarded as true information-processing proteins.

1.2 Protein–RNA Interactions

The specific recognition of RNA by a protein has many distinct features.[7,23–26] At the level of shape recognition, RNA recognition differs considerably from DNA recognition. RNAs are generally single-stranded molecules that are intramolecularly folded into defined 3D structures of great variety. Even in stem regions RNAs have distinct structures since the presence of a 2′-hydroxyl group in RNA causes the duplex segments to be close to A-form DNA. Hence, protein interactions in duplex regions often occur through the wide accessible minor groove. Also induced fit and rearrangements of the RNA structure by the bound protein are common themes in protein–RNA interactions. At the level of chemical recognition, many of the principles of protein–DNA interaction also apply to protein–RNA interactions.

2 METHODS USED TO MONITOR PROTEIN–OLIGONUCLEOTIDE INTERACTIONS

Two electrophoretic methods dominate the analysis of protein–DNA interactions. EMSA is technically the simplest and is used in its basic form to reveal sequence-specific DNA-binding activity in a protein sample or cell extract. Here complex formation generates a separate migrating band in a nondenaturating gel, and the protein–DNA binding thus leads to generation of a positive signal. The same method is also called band shift assay, gel retardation assay, gel shift assay or gel mobility shift assay (GMSA). A footprinting method provides more detailed information on the precise location of a bound protein along the DNA fragment. Here binding lead to removal of specific bands (negative signal) in a pattern of cleaved fragments separated by electrophoresis and reflecting the sequence of the DNA. Both methods are highly sensitive due to the use of radioactively labeled oligonucleotides that enable detection and analysis of proteins in the femtomole range. Both methods can be performed with protein samples of low purity. Since their specificity resides in the DNA sequence, they both require some prior knowledge of the sequence around the binding site. When combined these methods are capable of providing a picture of the protein–DNA complex with a great deal of molecular detail, surpassed only by the more demanding methods of crystallography and NMR.

2.1 Electrophoretic Mobility Shift Assay

In EMSA the complex that is formed between a sequence-specific DNA-binding protein and its recognition sequence (encoded by a duplex oligonucleotide or

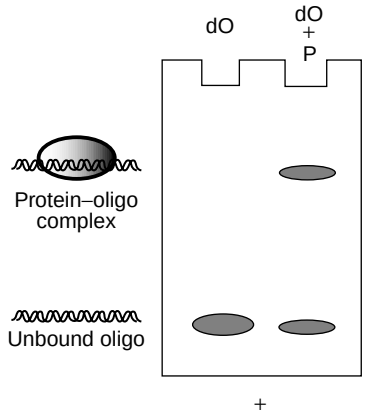

Figure 1 Schematic representation of the basic principle of the EMSA.

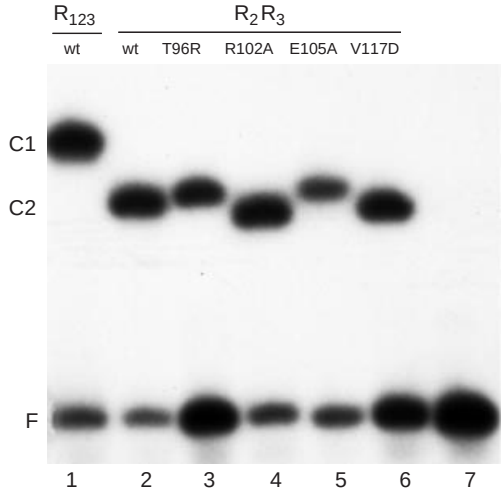

Figure 2 DNA-binding of recombinant c-Myb proteins analyzed by EMSA. DNA-binding was monitored by the EMSA using purified chicken c-Myb-derived proteins expressed in *E. coli* (Protocol 4) and a labelled MRE(mim) oligonucleotide as DNA-probe (using Protocols 1B, 2, and 3A).[41] Each binding reaction contained 20 fmoles of recombinant protein and 10 fmoles labelled oligonucleotide and the analysis was performed as described in Protocols 6 and 7. On the autoradiogram shown C1 indicates the complex formed with a protein spanning the three domains R_1, R_2 and R_3 in c-Myb (residues 39–192, protein of 18.8 kDa) (lane 1), while C2 indicates the complex formed with the smaller protein containing only the $R_2 + R_3$ domains (residues 89–192, wild type protein of 12.7 kDa) (lane 2–6). In addition to the wild type version of R_2R_3 (lane 3), four mutants of the same protein were used: T96R, R102A, E105A and V117D (lanes 3–6). Lane 7 shows the free DNA-probe only.

a DNA fragment) is subjected to electrophoretic separation in a nondenaturating polyacrylamide (PA) gel. Since the complex is fairly stable, the macromolecules remain associated during electrophoresis. In its standard version the duplex oligonucleotide is radioactively labeled and its migration is detected after electrophoresis by autoradiography. Unbound DNA is seen as the fastest migrating band, while complex formation leads to bands of reduced (''shifted'') mobility, hence its name. This principle is illustrated in Figure 1. This simple method was first described by Fried and Crothers[27] and by Garner and Revzin,[28] and has now become one of the standard methods in molecular biology. It has been the subject of several reviews.[29–37]

2.1.1 Principle of Electrophoretic Mobility Shift Assay

During EMSA both free and bound DNA migrate towards the positive pole driven by the dominating negative charge of the oligonucleotide. The shifted mobility of the complex is caused mainly by two effects. Neutralization of several phosphate groups in the oligonucleotide by basic residues in the protein leads to a reduced net negative charge of the complex compared to free DNA. In addition, the increased mass of the complex caused by the added protein will affect migration. Both these contributions work in the same direction to slow the migration of the complex. Which of these effects dominate depends on the specific proteins under study and experimental conditions used. Even small DNA-binding domains, with a size that barely contributes to retardation in a 6% PA gel, may give well shifted complex bands, suggesting that the neutralization of charge can make a major contribution to the shift in mobility (see Figure 2). The effect of molecular mass can be seen in the analysis of truncations of the same DNA-binding protein. When the mass of the protein increases, the migration is normally retarded as illustrated in Figure 2 (compare lanes 1 and 2). On the other hand, when the mass of the oligonucleotide increases, the shift in migration may be reduced since the relative mobility of bound versus free DNA tends to increase with increasing oligo length.[38] Hence, it has been concluded that it is the ratio of protein and nucleic acid masses rather than their absolute mass which is important for resolution in EMSA.[30] The contribution of the associated protein will of course be one of both mass and charge. The latter can be illustrated by an EMSA analysis of mutants of the same c-Myb-derived protein (Figure 2). On a 6% PA gel, where the effects of differences in masses are expected to be negligible, a clear effect of charge differences can be seen. When a mutation eliminates a positive charge (R102A, lane 4) or adds a negative charge (V117D, lane 6), the complex migrates slightly faster, while the introduction of an extra positive charge (T96R, lane 3) or elimination of a negative charge (E105A, lane 5) have the opposite effect and cause slower migration. The importance of the charge effect is best illustrated by the E105A protein (lane 5), which migrates slower due to its reduced charge even if its mass is reduced by 58 Da. An interesting use of this phenomenon

is in studies of phosphorylation of DNA-binding proteins where the altered charge may be followed by mobility changes in EMSA (see below). A third factor that also will affect the migration of the complex is its conformation, a phenomenon that is directly exploited to study protein-induced conformational changes in DNA (see below).

It has been suggested that protein–DNA complexes may in fact be stabilized by the PA matrix itself. This phenomenon termed ''caging'' was interpreted as an effect of the gel matrix that impeded the diffusion and separation of the dissociated components.[27,30] Concentrations of protein, DNA and complex would thus remain high, close to the migrating complex favoring reformation of complexes after dissociation.[31,39,40]

For a more extensive discussion of theory and principles of EMSA, see the recent review by Cann.[31]

2.1.2 Usage of Electrophoretic Mobility Shift Assay

The attractiveness of EMSA lies both in its simplicity and its wide range of applications. Its potency is illustrated as follows.

EMSA is used to assay for the presence of a particular sequence-specific DNA-binding protein in cell extracts. EMSA is widely used in combination with nuclear extracts (NE) obtained from cultured cells. The potency of the method lies in its ability to assay several different protein factors in the same extract simply by changing the duplex oligonucleotide. In a crude cell extract it is usually important to confirm the specific nature of a shifted band either by competition experiments or by performing a supershift experiment as discussed below.

EMSA is used to follow the purification of a particular sequence-specific DNA-binding protein. For several proteins like transcription factors, this is often the only simple in vitro assay available that is specific for the factor and applicable to rapid screening of fractions.

EMSA is used to analyze the effect of mutants in the DNA-binding domain of a specific protein. Mutations that have a significant effect on DNA-binding can easily be identified in EMSA studies.[38]

EMSA can also be used to analyze the effect of protein modifications (phosphorylation, oxidation etc.) on DNA-binding of a specific protein.[41,42] Again, alterations that have a significant effect on DNA-binding can easily be identified. Even modifications that do not reduce or increase DNA-binding can be observed if they lead to an alteration of charge, as in cases of phosphorylation. Since a phosphate group will contribute both with a small increase in size and a negative charge (working in opposite directions with respect to migration), it may be necessary to optimize the running conditions to get a maximum effect of either the charge difference or the increased mass in order to get a separate band for the modified form (low/high gel concentration and extended running times).

By using a series of systematically mutated duplex oligonucleotides, EMSA can be used to analyze the importance of each base pair in the binding site for a particular DNA-binding protein. Similarly, related but different response elements found in different genes can be systematically compared with EMSA. See Stormo and Fields[13] for a recent discussion of this topic.

EMSA with purified proteins can be used to determine thermodynamic binding constants for complex formation.[32] Even if EMSA separates the components in the binding reaction after equilibrium is obtained, it is generally assumed that what is observed after electrophoresis reflects equilibrium concentrations in the binding reaction before loading on the gel. The caging effect discussed above is probably one reason why this seems to be a good approximation in many cases making it possible to obtain reasonable estimates of equilibrium constants from EMSA. Quantitation of the unbound DNA and the protein–DNA complex (by a phosphoimager, by scanning of the autoradiogram or by direct scintillation counting of excised bands) provide concentration information on two of the three components in the simple equilibrium reaction (Equation 1):

$$P + dO = C \tag{1}$$

where P is the equilibrium concentration of unbound DNA-binding protein, dO is the equilibrium concentration of unbound duplex oligonucleotide, and C is the equilibrium concentration of the protein–DNA complex. The information provided by EMSA-quantitation is C and dO. Then the equilibrium constant is given by Equation (2)

$$K_{app} = \frac{[C]}{[P][dO]} \tag{2}$$

A simple estimate of binding constant can be made from a series of binding reactions with increasing protein input (or decreasing probe input). Typically, the ratio of bound to unbound ([C]/[dO]) DNA is determined at several concentrations of protein and then curve-fitting is employed to give an estimate of K_{ass} (or K_d). Scatchard plot is one convenient way of plotting the data for curve fitting (plotting [C]/[dO] versus [C]). A very simple alternative is the following. At the input where the retarded band and the unbound oligonucleotide becomes equal in intensity on the autoradiogram, the following simplification can be made (Equation 3):

$$K_{app} = \frac{1}{[P]} \tag{3}$$

[P] can then be calculated from Equation (4):

$$[P] = P_t - [C] \tag{4}$$

where P_t indicates the total input of purified protein. More sophisticated approaches have been reviewed.[29,31,32,43,44]

EMSA can also be used to estimate kinetic parameters by loading samples from a binding reaction on a running EMSA gel.[32] Again, the underlying assumption is that the electrophoresis ''freezes'' the ratios of the components at the time of entrance into the gel matrix.

EMSA can also be used to analyze the assembly of multiprotein complexes, to demonstrate synergistic assembly of complexes where two or more protein factors bind closely positioned sites or where macromolecular assemblies are formed with only one DNA-binding component. If the original DNA probe is large enough to cover several binding sites, and a complex pattern of shifted complexes are observed, these may be analyzed by competition with distinct duplex oligonucleotides representing individual binding sites for specific transcription factors. Cooperative assemblies of protein–DNA complexes have been recently reviewed.[45]

Even if the most common use of EMSA is analytical with sensitive labeled probes, the method can also be used preparatively to isolate a particular sequence-specific DNA-binding protein from a partially purified sample.[46]

EMSA is often performed with a specific antibody added to demonstrate the presence of a previously identified DNA-binding protein in the complex. This is of particular importance when analyzing crude cell extract preparations where a shifted band may be caused by factors other than that expected. Most EMSA protocols can easily be adapted to this type of analysis simply by allowing the antisera and the protein sample to be preincubated for about 30 min on ice before proceeding with the DNA-binding reaction. Addition of antibodies may either block complex formation, if the epitope recognized is one critical for DNA-binding, or lead to a complex with reduced mobility, if the epitope recruits the antibody to the complex without affecting protein–DNA interaction. The latter is generally called ''supershift'' and is the result that is the simplest to interpret. In these kinds of experiments, two types of controls are desired: analysis of the antisera alone in EMSA and a ''supershift'' control with an unrelated antibody or pre-immune serum, both to exclude artificial nonspecific bands.

Since the DNA-binding function of a protein often resides in a small domain representing only a fraction of the protein's molecular mass, it is possible to subject a protein to limited proteolysis and thus generate a fingerprint pattern of digestion products and analyze these in EMSA. All fragments with the DNA-binding domain untouched should normally give shifted bands, but these will migrate differently due to their different masses. The digestion pattern will provide a unique fingerprint pattern characteristic of the protein in question and can be used for comparative purposes. Protocols for this ''proteolytic clipping band shift assay'' are found in Dent and Latchman.[35]

Several transcription factors induce bending and other conformational changes in DNA upon binding (as referred to above). When a protein induces a bend in the DNA helix, this may result in changes of mobility of the complex during electrophoresis in a way that depends on the bend position within the migrating DNA fragment.[17,30,32,37,47] By using a series of fragments of equal size, but with the protein binding site differently located along the fragments (circular permutation), it is possible, using EMSA, to demonstrate induced bending from the position-dependent migration of the complexes. It is even possible to estimate the bending angle. In addition, by combining with a static sequence-induced bend it is possible to determine the direction of the protein-induced bend (phasing analysis). These methods are normally performed with restriction fragments of DNA (generated from plasmids designed for the purpose[48,49]), rather than with oligonucleotides. Information on these techniques can be found in several reviews.[18,47,50–53]

2.1.3 *Practical Aspects of Electrophoretic Mobility Shift Assay*

To set up a standard EMSA experiment, the following steps have to be performed:

- preparation of oligo probe
- preparation of protein sample
- making of the PA gel
- performing the binding reaction
- electrophoresis
- autoradiography.

In the following paragraphs, these steps will be described and commented on.

2.1.4 *Preparation of Labeled Oligo Probe*

There are several possibilities to label an oligonucleotide for use in EMSA. Usually, the oligonucleotide is labeled by ^{32}P on one or both ends of the duplex. In contrast to footprinting probes, the location of label is not critical for binding probes used in EMSA. Since commercially synthesized oligonucleotides generally are delivered with a 5′-OH, they are ideal substrates for T4 polynucleotide kinase. This kinase will transfer the labeled γ-phosphate from [γ-^{32}P] adenosine triphosphate (ATP) to the 5′-end of the oligonucleotide leading to a probe with high specific activity (>1000 cpm/fmol). In our lab we routinely use the Ready-To-Go™ T4 Polynucleotide Kinase kit (Pharmacia) to label up to 10 pmole oligo,

but a standard reaction mix will work as well for this purpose (see Protocols 1A and 1B). Since the enzyme is more efficient with single-stranded termini than with blunt ends, it is recommended to first kinase the oligonucleotide and then to anneal it to the second complementary oligonucleotide to generate the duplex probe. A fast method to prepare EMSA probes is to kinase the first oligonucleotide, anneal it with a slight molar excess of the complementary oligonucleotide (to force quantitative annealing of the labeled strand) and purify the duplex by a spin column to eliminate unincorporated isotope and buffer components. This can be performed in a few hours. To obtain good, clean EMSAs and for more critical experiments, in particular more quantitative experiments, it is recommended to purify the duplex on a PA gel. For quantitation of ^{32}P-incorporation it is advocated to take a small aliquot of the kinase reaction, mix it with carrier tRNA, spot half on a 3MM paper filter disk that is dried (total cpm), the other half on a filter disk that is precipitated and washed in cold 10% trichloroacetic acid (TCA) (incorporated cpm). After scintillation counting, it is then possible to calculate the efficiency of labeling and the specific activity of the probe (cpm or dpm per fmole oligonucleotide). After purification, an aliquot of the final probe is also counted. From these values (specific activity and activity in final probe) it is easy to calculate the recovery and final molar concentration of the probe.

If several EMSA probes are needed with the same specific activity, it is possible to design the different oligonucleotides such that they all contain a common sequence in the 3′ end (approximately 10 nucleotides). A small oligonucleotide complementary to this common region is then labeled first, and thereafter annealed to the different oligonucleotides followed by a fill-in reaction using DNA polymerase I (DNaseI) (Klenow fragment) and the four deoxyribonucleotide triphosphates. In this manner, all probes get identical specific activities and therefore directly comparable intensities in EMSA.[54,55]

As an alternative to use of T4 polynucleotide kinase, the two oligonucleotides needed to generate a duplex can be designed to generate a 5′-protruding single-stranded terminus after annealing. Such a duplex will be a template for a fill-in labeling using DNaseI (Klenow fragment) and the relevant [α-^{32}P] deoxyribonucleotide triphosphate. Advantages of this approach are the possibility to incorporate more than one ^{32}P-labeled nucleotide and to incorporate unlabeled nucleotides after the labeled one and thus obtain a probe where the ^{32}P is better protected toward enzymatic degradation. Degradation of the label in crude cell extracts may be a problem reducing the sensitivity of the assay.[56] In general, a duplex oligonucleotide of 20–30 base pairs will be sufficient for EMSA. If larger fragments are needed, these may be isolated as restriction fragments that can be labeled by fill-in as described above, or produced as polymerase chain reaction (PCR) fragments. In the latter case, labeling is achieved either by kinasing of the PCR-primer(s) or by including an [α-^{32}P] deoxyribonucleotide triphosphate in the PCR reaction.

2.1.4.1 Protocol 1A: Labeling Oligonucleotides by Polynucleotide Kinase

- Decide the amount of oligonucleotide to be labeled, and calculate the equivalent amount of [γ-^{32}P]-ATP isotope. A standard isotope with specific activity of 3000 Ci mmole^{-1} and a concentration of 10 μCi μL^{-1}, will have an ATP concentration of 3.3 pmol μL^{-1}. Using a slight excess of isotope, 10 pmol oligo requires 3.5–4 μL [γ-^{32}P]-ATP isotope.
- Mix in a PCR tube on ice the following (add enzyme last):
  - 1.0 μL 10 μM oligonucleotide (single-stranded)
  - 3.5 μL [γ-^{32}P]-ATP 3000 Ci mmole^{-1}
  - 1.0 μL 10× polynucleotide kinase buffer
  - 1.0 μL 10 U μL^{-1} polynucleotide kinase
  - 3.5 μL H_2O
- Incubate at 37 °C for 45 min, followed by enzyme inactivation at 65 °C for 10 min.
- Remove an aliquot of 1 μL for scintillation counting and estimation of incorporation, as follows:
  - Mix the 1-μL aliquot with 25 μL 2 mg mL^{-1} tRNA.
  - Spot 12 μL of the mixture on one 3MM paper circle disk and leave to dry.
  - Spot 12 μL of the mixture on another 3MM paper circle disk and put it into a vessel of ice-cold 5% TCA.
  - Wash with gentle shaking for 5 min, pour off and repeat washing three times with 5% TCA.
  - Dry the filter disks.
  - Put the two filters into scintillation vials, add 5 mL of water and count incorporation using the ^{3}H channel of the scintillation counter. Take 1-μL aliquots of the final purified probe, spot, dry and count similarly. Use the cpm values and concentrations to calculate specific activity of the probe.

2.1.4.2 Protocol 1B: Labeling Oligonucleotides by Ready-To-Go™ PNK-kit

- Add 25 μL H_2O to one tube from the Ready-To-Go™ PNK-kit (Pharmacia) containing freeze-dried

material. Leave it for 2–5 min at room temperature to allow solubilization, and then mix by gentle pipetting.

- Transfer the mixture to a PCR tube to allow the incubation and subsequent annealing to be performed in a PCR machine.
- Add the following to the tube:
  - 1.0 μL 10 μM oligonucleotide (single-stranded)
  - 3.5 μL [γ-^{32}P]-ATP 3000 Ci mmole^{-1}
  - 20.5 μL H_2O
- Incubate the mixture for 30 min at 37 °C.
- If no fill-in reaction is necessary, stop the reaction by addition of 2 μL 0.5 M ethylene diamine tetraacetic acid (EDTA). If fill-in is required, no EDTA must be added at this stage.
- Remove an aliquot of 3 μL for scintillation counting and estimation of incorporation (see Protocol 1A).

2.1.4.3 Protocol 2: Annealing to Complementary Strand and Rapid Purification

- The PCR tube with the kinasing mixture (Protocol 1B) can be processed directly. If protocol 1A is used, transfer the kinasing mixture to a PCR tube and adjust the protocol to the amounts and volumes used.
- Add a slight excess of oligonucleotide that is complementary to the kinased oligo. 50% excess usually works well. The ionic strength is increased by addition of NaCl to facilitate annealing. Following the protocol 1B this means the following addition:
  - 1.5 μL 10 μM complementary oligonucleotide (single-stranded)
  - 5 μL 5 M NaCl
- Incubate in a PCR machine at 70 °C for 5 min, followed by a programmed slow decrease in temperature (one degree per minute) until the temperature has dropped to below room temperature.
- If the oligos that have been annealed are equal in length, the duplex probe is ready for purification (proceed with protocol 3A or 3B). If a shorter primer oligo has been labeled and annealed to a longer complementary oligonucleotide, a fill-in step is required at this stage to produce a full-length duplex.
- Optional fill-in reaction. Add the following to the tube on ice:
  - 2 μL 5 mM dNTP (mixture of dGTP, dATP, dTTP and dCTP; 5 mM each)
  - 0.5 μL 5 U μL^{-1} DNaseI Klenow-fragment.
- Incubate on ice for 2 min, followed by incubation for 45 min at 15 °C.
- Stop the reaction by addition of 2 μL 0.5 M EDTA.

2.1.4.4 Protocol 3A: Rapid Purification by Spin Columns

- Duplex oligo probes can be rapidly purified from unincorporated nucleotides by a spin column version of gel filtration. Suitable for this purpose are the MicroSpin G-25 Columns (Pharmacia). The disadvantage with this rapid method relative to purification on PA gels, is that the purified probe will still contain some labeled oligonucleotides that have not been annealed to duplex form, as well as the remaining complementary oligonucleotide.
- Follow the instructions of the kit for preparing the column, and load carefully 25–50 μL of the kinasing reaction on the column. Spin according to the instructions given by the manufacturer.

2.1.4.5 Protocol 3B: Purification on Polyacrylamide Gels

- Assemble the glass plates for the electrophoresis unit. The recipe below is for a 20 × 20 × 0.1 cm PA gel.
- To make a 15% PA gel, mix the gel solutions in a vacuum flask:
  - 11.25 mL 40% acrylamide [C = 2.7% (relative concentration of bis/bis + acrylamide)]
  - 0.6 mL 1 M Tris–HCl, pH 8.0
  - 6 μL 0.5 M EDTA
  - 18 mL H_2O.
- Degas and start polymerization by adding:
  - 100 μL 10% APS (ammonium persulfate)
  - 20 μL TEMED (*N*,*N*,*N*′,*N*′-tetramethylenediamine).
- Pour the gel solution immediately between the plates and mount the well-forming comb.
- Use as electrophoresis buffer 20 mM Tris–HCl, 0.1 mM EDTA:
  - 20 mL 1 M Tris–HCl, pH 8.0
  - 0.2 mL 0.5 M EDTA
  - H_2O to 1000 mL.
- Run a pre-electrophoresis at 250 V for 15–30 min.
- Add a blue marker to the labeled oligos, and load each of them in 1–3 lanes.
- Perform the electrophoresis at 15 V cm^{-1} until the marker has migrated two-thirds of the length of the gel.
- While the electrophoresis is running, prepare three small adhesive pieces of paper (3 × 3 mm) and spot 0.5–1.0 μL of any radioactive solution of suitable

activity. Allow to dry. These will be used for orientation purposes. Alternatively, special fluorescent paper strips may be used.

- Stop the electrophoresis and remove one of the glass plates. Attach the three adhesive papers to the other glass plate in an asymmetric pattern around the gel. Wrap in plastic film and place an X-ray film over the gel in the dark room. Expose for 3–5 min, or longer for weaker probes. Develop the film.
- By alignment of the paper spots to the corresponding spots on the film, it is possible to precisely orient the gel over the film. Cut out the labeled oligos and place them in a microcentrifuge tube. Add 1 mL 10% TE-buffer [1 mM Tris–HCl (pH 8.0), 0.1 mM EDTA] and place the tubes in a rotating device (like a hybridization oven) at 37 °C overnight.
- Next day, transfer the eluate to a new tube. Spin briefly the tube with the gel to collect all the liquid. Spin the tube with the pooled eluate in a benchtop centrifuge for 1 min and transfer the eluate carefully to a new tube leaving behind the last 10–20 μL to avoid traces of gel pieces.
- Concentrate the eluate with 2-butanol as follows. Add 500 μL 2-butanol, vortex and spin for 1 min to separate the phases. Remove the upper phase of 2-butanol. Add a second portion of 500 μL 2-butanol, vortex and spin. Remove the upper phase of 2-butanol. Continue this extraction procedure until the volume of the lower water phase is reduced to 100–150 μL. Carefully remove traces of 2-butanol.
- Spot three aliquots of 1 μL on a 3MM filter paper disk, dry and count to determine the activity and concentration of the final probe.

2.1.5 Recombinant Proteins as Protein Source

When the DNA-binding properties of a specific DNA-binding protein are to be characterized, it is advantageous to express the protein of interest in bacteria or another system for efficient production of recombinant protein. The DNA-binding function in proteins like transcription factors often resides in a small independent domain of the protein. In cases where it is difficult to express a soluble functional full-length protein, it is possible to express only a subdomain carrying the full DNA-binding function of the protein. Smaller DNA-binding domains will often be soluble in *E. coli* and expressed more efficiently than larger proteins. Many commercial systems for expression of recombinant proteins in *E. coli* are available, generally using vectors that allow the inducible expression of tagged proteins that are easy to purify in a single step on affinity resins. Tags (like glutathione S-transferase (GST), His-tags, etc.) are in most cases very useful, but it should be verified that a particular tag does not alter the DNA-binding affinity or properties of a protein. In our work with the c-Myb DNA-binding domain, we have observed 10–100-fold reductions in binding constants due to fusion to tags. If the tag is only used for purification purposes, it should also be noted that unfused DNA-binding domains overexpressed in *E. coli* may be very easy to purify by more classical approaches. Columns of choice are either a negatively charged heparin matrix (like Heparin-Sepharose or HiTrap, Pharmacia) or classical cation exchangers (S-Sepharose). The efficiency of these matrices is probably due to the presence of a basic domain in DNA-binding proteins able to bind efficiently to the negative phosphodiester chain in DNA and hence also with high affinity to cationic matrices. In addition, *E. coli* appears to produce few basic proteins at comparable high levels that will contaminate the actual fractions. In our experience, it is possible in a one-step procedure to obtain in the order of 90% pure recombinant proteins in this way, and if the two types of columns are combined, a very high purity can be obtained without the need of any tags.

DNA-binding studies with purified recombinant proteins is recommended when well-defined and quantitative biochemical experiments are to be performed, since a precise estimation of protein concentrations is possible. On the other hand, the specificity of EMSA allows detection of specific DNA-binding proteins expressed in *E. coli* directly in a cell-free extract obtained after induction, lysis and high-speed centrifugation of the cell lysate. In crude systems it is important to include a nonspecific DNA like poly dI–dC competitor to eliminate interference from nonspecific DNA-binding proteins. If only analytical quantities are needed to demonstrate a particular protein–DNA interaction, an alternative approach is to use coupled in vitro transcription/translation systems.

An obvious disadvantage of expression of mammalian protein in bacteria is of course the lack of post-translational modifications. Properties deduced from in vitro experiments with recombinant proteins should therefore be validated by studies of the same protein in NE from relevant cells where modifications may be expected to occur.

2.1.5.1 Protocol 4: Purification of Untagged Recombinant DNA-binding Proteins Expressed in E. coli

- The DNA-binding protein or subdomain is expressed in *E. coli* BL21(DE3)LysS using the T7 system (or other commercial systems). Recombinant *E. coli* cells are grown in SOB medium (20 g L^{-1} bactotryptone, 5 g L^{-1} bacto-yeast extract, 0.5 g L^{-1} NaCl, 2.5 mM KCl, pH 7.0; before use add 20 mM $MgSO_4$) at 37 °C to an optical density of 0.6–0.9 and then induced by addition of isopropyl-β-D-thiogalactopyranoside ((IPTG), final concentration

0.4 mM). The cells are harvested after two hours of induction, washed once in cold TEN buffer (10 mM Tris–HCl, 1 mM EDTA and 100 mM NaCl), and resuspended on ice in buffer A (20 mM Tris–HCl, pH 8.0, 1 mM EDTA, 10% glycerol, and 1 mM dithiothreitol (DTT)) containing 500 mM NaCl. To the suspension on ice is added phenylmethylsulfonyl fluoride (PMSF, 1 mM), and Triton X-100 (0.1%). After lysis (5–10 min), the viscous suspension was centrifuged at 40 000 rpm for 1 hr in a TFT 65.13 rotor at 4 °C. The nonviscous part of the supernatant is collected. At this stage, the extract may be stored at −80 °C.

- The proteins can then be purified on a heparin affinity column using a buffer containing 20 mM Tris–HCl, pH 8.0, 1 mM EDTA, 10% glycerol, 1 mM DTT, 0.01% Triton X-100 and 0.1 mM PMSF. The samples are loaded on the column using 250 mM NaCl (this may be adjusted for each protein), and the proteins are purified in an NaCl gradient from 250 to 1500 mM.

2.1.6 Nuclear Extracts or Whole Cell Extracts as Protein Source

By far the most common use of EMSA is to demonstrate the presence of a particular DNA-binding activity in an extract from a particular cell line or from a tissue. Identification of the pattern of endogenous transcription factors expressed in a particular cell type is an important step in the characterization of a promoter active in that cell line. Alternatively, a cell line may be transfected with expression plasmids for a particular DNA-binding protein resulting in higher levels of the protein of interest. The preparation of extracts for EMSA may be performed from entire cells giving whole cell extracts (WCE) or from isolated nuclei giving NE. The advantage of WCE is a simpler procedure with fewer steps where loss of activity can occur. In cases where a DNA-binding protein shuttles between cytoplasm and the nucleus, a WCE may also reflect the whole DNA-binding capacity of the cell in question. The disadvantage is mainly one of lower relative abundance since nuclear proteins become diluted with a large amount of cytoplasmic proteins. It is also the only option when the cellular material has been frozen, since this destroys the nuclear membrane and precludes the isolation of nuclei. If sufficient cell material is available, it may be advantageous to prepare a nuclear extract. Several protocols are available for this,[35,57,58] most of them derivatives of the original protocol of Dignam et al.[59]

2.1.6.1 Protocol 5: Preparation of Whole-cell Extract from COS-cells Transfected with Expression Plasmids for a DNA-binding Protein A convenient EMSA source for the protein of interest can be obtained by transfection of COS-1 cells with an appropriate expression plasmid. COS-1 cells are African green monkey kidney-derived cells transformed by simian virus 40 large T antigen. The cells can be transfected with the expression plasmid for the protein of interest using one of several standard methods. A WCE can easily be prepared using the protocol of Oelgeschläger et al.[60] as follows:

- Cells are seeded in 10-cm dishes at a density of 5×10^5 per plate and transfected with 30 μg of DNA per plate. Cells are harvested 36 h after transfection in 500 μL of F buffer [10 mM Tris–HCl (pH 7.05) 50 mM NaCl, 30 mM sodium pyrophosphate, 50 mM NaF, 5 μM $ZnCl_2$, 100 μM Na_3VO_4, 1% Triton X-100, 1 mM PMSF, and a cocktail of protease inhibitors].
- The lysate is vortexed for 30 s before centrifugation at 14 000 rpm for 30 min at 4 °C. The supernatants are stored as aliquots at −80 °C after being rapidly frozen in liquid nitrogen.
- For EMSA, use 1–3 μL as protein source.

2.1.7 Protocol 6: Preparation of Polyacrylamide Gel

- Assemble the glass plates for the electrophoresis unit.
- Mix the gel solutions in a vacuum flask:
  - 4.5 mL 40% acrylamid [C = 2.7 (relative concentration of bis/bis+acrylamide)]
  - 3.0 mL 5× TBE [0.45 M Tris-base, 0.45M boric acid, 10 mM EDTA]
  - 3.0 mL 50% glycerol
  - 19.5 mL H_2O.
- Degas and start polymerization by adding:
  - 150 μL 10% APS
  - 20 μL TEMED.
- Pour the gel solution immediately between the plates and mount the well-forming comb.
- Use 0.5× TBE as electrophoresis buffer.
- When the gel has polymerized, remove the comb and the bottom spacer and mount in the electrophoresis unit. After the electrophoresis buffer has been filled up, wash the wells with buffer.
- Since the samples to be loaded are without color, it might be practical to mark the wells on the outside by a permanent ink pen.
- To avoid dissociation due to heating during electrophoresis, place the gel apparatus in the cold room.

2.1.8 Conditions for Complex Formation

The optimal conditions for complex formation will to some extent depend on the protein under study and

should be optimized in each case. Some general guidelines may be followed. Many of the typical additions to enzymatic reactions (like buffer, salt, glycerol) will also help to keep a DNA-binding protein active. A comment should be made on the ionic strength in the reaction mixture (adjusted by the addition of salt). The latter is important for the stability and half-life of the complex. Since low ionic strength will stabilize protein–DNA interactions in general, too low salt concentration may strengthen binding to nonspecific sites as well (like on competitor DNA) and slow the exchange between specific and nonspecific sites. If the competitor is added to the protein before the probe under such conditions, the protein may remain associated with the competitor hampering specific complex formation. Too high ionic strength will on the other hand destabilize protein–DNA binding and preclude specific complex formation. It should also be noted that the ionic strength normally drops when the complex migrates into the gel matrix, a factor that contributes to the high stability and slow dissociation of the complex during electrophoresis. This is also the reason why EMSA is not ideal to measure the salt-dependence of protein–DNA complexes.

Some DNA-binding proteins may need specific cofactors to be active. Typical examples are the family of zinc finger proteins that require zinc to form an active DNA-binding structure. Addition of EDTA here can inhibit the activity of the factor.

Many transcription factors contain easily oxidized cystein residues in their DNA-binding domains. Addition of reducing agents like DTT or β-mercaptoethanol will be necessary to obtain full activity.

Due to the sensitivity of EMSA, operating below the picomole range of protein input, it may be necessary to take precautions to avoid adsorption of proteins to surfaces that may cause loss of active material. This is particularly important when working with purified proteins, and additions of non-ionic detergents (like NP-40 or Triton X-100) and/or inert proteins (like bovine serum albumine) are recommended.

Since the protein sample is the most fragile ingredient, we prefer to add this as the last part of the reaction mixture. After mixing the reaction components, they are usually incubated for a specific time at a specific temperature. Many complexes form very rapidly and extended time of incubation is not necessary. Incubation on ice may be sufficient. In some cases the binding may be a two-step reaction where a rapid bimolecular association is followed by a slower docking reaction involving conformational changes and leading to more stable complexes.[61] In such cases incubation for 10–30 min at room temperature may be preferably to obtain stable complexes.

A typical reaction mixture contains sufficient glycerol to be loaded directly on the gel after incubation. The addition of tracking dye before loading is not generally recommended since the dye may inhibit the binding activity of specific proteins. Loading dye may be added to a reference lane containing the probe only in order to follow the electrophoresis.

2.1.8.1 Protocol 7: Mixing Samples and Electrophoretic Separation of Complexes

- Estimate the amount of labeled probe and protein to be used in the binding reaction. Using the labeling protocol described above, a specific activity in the order of 1000 cpm fmole^{-1} should be attained and 10 fmol of duplex probe will be sufficient per lane to give strong signals on the autoradiogram after overnight exposure. The amount of protein input required depends on its binding strength and purity. With purified recombinant proteins 10–50 fmol will usually be sufficient. With WCE 1–2 μL is recommended.
- Dilute the protein/protein extract in an appropriate buffer on ice [like $TGED^{T}_{500}$ (20 mM Tris–HCl, 10% glycerol, 1 mM EDTA, 500 mM NaCl, 1 mM DTT and 0.05% Triton X-100) with DTT and Triton freshly added]. The proteins should be freshly diluted and not stored in diluted form.
- Keep all samples on ice.
- Prepare the required number of tubes with the components added in the order listed below, giving a total reaction volume of 20 μL. Make one tube without added protein (free probe).
  - 4 μL 5× binding buffer (100 mM Tris–HCl, 50% glycerol, 5 mM EDTA)
  - 1 μL 0.5 mg mL^{-1} poly dI-dC (input adjusted depending on the protein purity, not necessary for pure recombinant proteins)
  - 12 μL H_2O
  - 1 μL of duplex probe diluted to 10 fmol μL^{-1}
  - 2 μL diluted protein/protein extract (giving a final [NaCl] = 50 mM).
- Incubate for 15 min at 25 °C (or on ice, see text).
- Assemble the electrophoresis apparatus, and place it in the cold room. Start the pre-electrophoresis (10 min at 200–250 V).
- Add a blue marker only to the lane with free probe (Coomassie Brilliant Blue, no detergent) to follow migration and to orient the gel later on.
- Load the samples in the respective lanes.
- Perform the electrophoresis at 10–20 V cm^{-1} for the required time (we use 200 V for 2 h; this will depend on the size and design of apparatus).

- Stop the electrophoresis, disassemble one of the plates leaving the gel on the other, cut a mark in one edge of the gel (for orientation purposes), and transfer to a dry piece of 3MM paper sheet. Cover the gel with a thin plastic film. Put in a film cassette with intensifying screen and film, and expose at $-70\,^{\circ}C$ for 4–48 h (depending on the signals) before development of the film.

2.2 DNA Footprinting Methods

The information provided by EMSA is basically whether a DNA-binding protein recognizing part of the oligo sequence is present in a protein sample or not. It does not tell exactly what part of the sequence is recognized by the protein, even if such information can be obtained indirectly by EMSA by using a series of mutant oligos. Information on the location of a bound protein along a DNA sequence can be obtained in great detail using one or several types of footprinting methods. Footprinting is a designation used on a set of methods that combine modification of a specific DNA (chemically or enzymatically) with binding to a sequence specific protein. The modified DNA is finally cleaved into a series of fragments that are separated with base pair resolution on a sequencing gel. Since the original DNA fragment or oligonucleotide is end-labeled on one strand, the analysis can be restricted to the fragments retaining this label. This generates a pattern of cleaved fragments where all have one labeled end in common and where the other end marks the site of cleavage, which is also a site of modification. In this ''ladder'' of bands, each cleavage product corresponds to a specific nucleotide in the DNA sequence. To precisely align the ladder of cleavage products to the DNA sequence, a Maxam–Gilbert type of chemical sequencing reaction is performed on the same labelled DNA fragment, and run in parallel alongside the cleavage products. By comparing the generated pattern in the presence and absence of the bound protein, it is possible to locate the protein relative to the DNA sequence. In some versions protein binding creates a clear zone free of bands in the ladder, and this region is usually called a ''footprint''.

The size of a binding site for a typical sequence-specific DNA-binding protein varies, but will most often be in the range of 5–15 base pairs. Although DNA fragments of a few hundred base pairs have often been used in these kinds of analyses, the methods are easily adapted for use with duplex oligonucleotides of some length. In the design of oligonucleotides for footprinting, it should be taken into consideration that the binding site becomes situated at some distance from the end that is labeled (>25 base pairs). This is because most methods rely on the precipitation of DNA cleavage products, which becomes challenging when fragment lengths are too short. Labeling strategies for these methods are the same as for EMSA probes, but now only one end can be labeled. To fully analyze a region of interest by DNA footprinting, two probes should be prepared, each with one of the strands labeled.

Depending on the types of modifications used, the protein–DNA interaction can be mapped with different degrees of accuracy. Footprinting methods can be divided into two main groups, *protection* and *interference* types of footprinting, depending on the order of operations involved. Methods where protein binding precedes DNA modification are called protection footprinting methods, since these methods show at what positions the bound protein protects the DNA from subsequent modification reactions. Methods where DNA modification precedes protein binding are called interference footprinting methods, since these methods show at what positions a specific DNA modification interferes with subsequent protein binding. Although these methods were originally developed as in vitro methods, some of them have also been adapted to analysis of protein binding in vivo. Footprinting methods have been the subject of several reviews.[29,36,37,62–71]

2.2.1 Protection Footprinting Methods

Protection footprinting involves treatment of a preformed protein–oligonucleotide complex with agents that modify the DNA. The specific methods vary with respect to the type of modification agent used. Their common feature is that they provide information on how sequence-specific protein binding affects subsequent DNA modification reactions in the binding site. The general principle of protection footprinting is illustrated in Figure 3.

An end-labeled duplex oligonucleotide or DNA fragment is allowed to bind to an excess of a specific DNA-binding protein so that the DNA is quantitatively complexed with the protein. A control without the protein is processed in parallel. Then the chemical or enzymatic agent is allowed to modify the complex and the control DNA. The agent will attack randomly, but will be excluded from certain positions along the DNA due to the presence of the protein. When the DNA finally is deproteinized and cleaved at the modified positions, followed by separation on a sequencing gel, the pattern of labeled bands will reveal at which positions the presence of the protein has protected against modification. From this information, the location of the bound protein can be deduced.

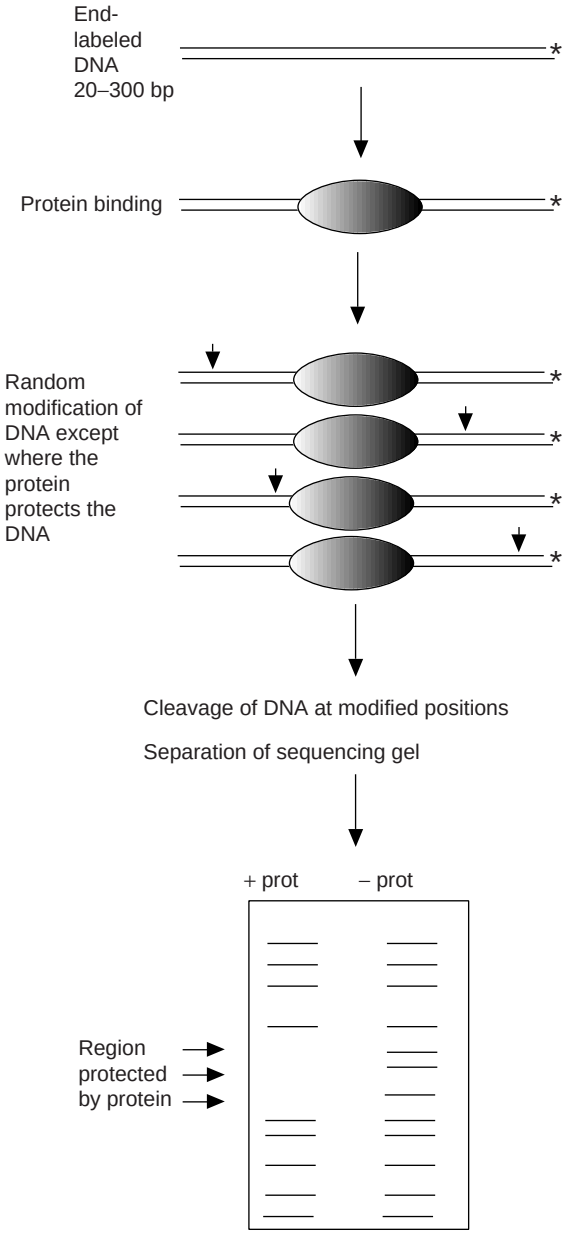

Figure 3 Schematic representation of the principle of protection footprinting methods.

2.2.2 Footprinting using the Enzyme DNase I

DNaseI footprinting was the first DNA footprinting method developed[72] and today is one of the most commonly used variants of protection footprinting. As the name implies, the modification agent here is the enzyme DNaseI. The complex is exposed to a fixed (titrated) amount of DNaseI for a brief period of time (15–60 s), which will introduce nicks in DNA in a fairly random fashion, but will be excluded from the region where the bound protein is located. This creates an open zone free of cleavage products indicating the position of the protein (a ''footprint''). Because the enzyme introduces cleavages directly, no separate cleavage step is required, making this variant one of the simpler versions of protection footprinting methods. To obtain an optimal ladder of bands, it is important to find conditions that give a ''single hit'' pattern where the complexes on the average are only cleaved at one position. It is therefore recommended to carefully titrate the input of DNaseI under the reaction conditions to be used, to find the amounts of enzyme that give the best ladder of cleavage products. It is noteworthy in this context that DNaseI is highly sensitive to the ionic strength in the reaction mixture.

While simplicity is the major advantage, a few disadvantages of DNaseI footprinting should also be mentioned. The first is one of limited resolution. Since the cleavage agent is a protein of 31 kDa, it will easily be excluded by steric hindrance from the binding site by the bound protein. Although this generally contributes to a good visual footprint, the protected region will usually be slightly larger than the DNA region actually in direct contact with the bound protein. A related problem is that DNaseI does not cleave in a totally random fashion, but has some secondary structure preference. The consequence is an uneven pattern of bands and some regions may have few cleaved fragments, reducing the resolution of the footprint analysis.

Another problem common to most of the protection methods is that a bound protein produces negative signals, that is the disappearance of specific bands on a gel. To obtain a clear pattern it is therefore crucial to start with conditions where the labeled DNA is quantitatively complexed to the protein. If not, the footprinted zone will not be free of signals since the remaining uncomplexed DNA will be attacked and produce cleavage products in the zone. This is not necessarily a major problem if it is possible to scan the footprint and treat it quantitatively. In cases where it is difficult to obtain optimal conditions of binding, an improved footprint can be obtained by isolating the complex before deproteinization and denaturation. The DNaseI reaction can be stopped by EDTA and the nicked complex separated from free DNA using EMSA. The retarded complex band is cut out from the EMSA gel, eluted and then processed according to the standard DNaseI footprint protocol. In this way, only complexed DNA is analyzed. This latter version has been called ''indirect'' footprinting to distinguish it from the ''direct'' footprinting that lacks the EMSA step.

A phenomenon that is often observed, in particular in the border regions of a footprint, is bands that instead of being protected become enhanced in the presence of a bound protein. The likely explanation is that the protein induces some kind of structural alteration in the DNA, in a way that makes the DNA a better substrate for the DNaseI just outside the protected region.

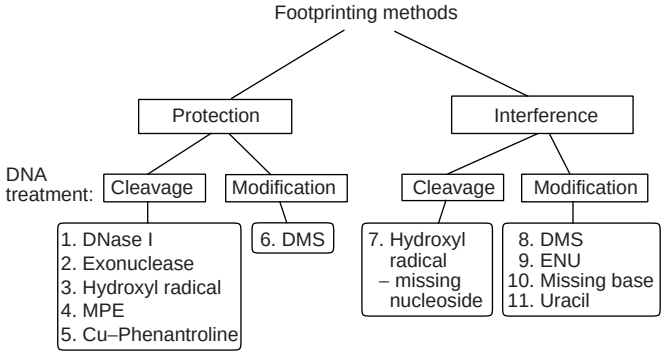

Figure 4 Schematic overview of enzymatic and chemical agents used to modify DNA in either protection or interference footprinting methods.

DNaseI footprinting has also been used in more quantitative approaches to analyze protein–DNA interactions.[65,67,73–76] Protocols for DNaseI footprinting can be found in specific reports,[77] in several reviews,[29,66,71] or in practically-oriented book chapters.[36,37,70]

Several agents have been used to generate supplementary methods of protection footprinting. Some of the common agents used are listed in Figure 4. Most of these are cleavage agents able to make incisions in DNA under mild conditions compatible with protein–DNA complex formation. In contrast, many of the agents used to modify DNA in interference methods cannot be used under such conditions.

2.2.3 Footprinting Using Dimethyl Sulfate

One very useful chemical agent is dimethyl sulfate (DMS) ($(CH_3)_2SO_4$). The use of DMS for DNA footprinting came after the development of methods for chemical DNA sequencing by Maxam and Gilbert.[78,79] DMS is a strong methylating agent that will alkylate DNA in unprotected guanine bases in the N-7 position of the purine ring under mild conditions. The N-7 position is exposed in the major groove and is often involved in direct protein–DNA hydrogen bonding contacts in specific complexes. DMS also attacks, to a lower extent, adenine residues at N-3 positions exposed in the minor groove, but this has been less exploited for footprinting. Protein bound in the major groove will therefore protect guanine residues from methylation in the N-7 position. DNA fragments with this type of modification can be selectively cleaved by treatment with hot piperidine. Since DMS is a rather small chemical agent, it will be able to penetrate many regions where DNaseI is excluded, and thus provide a footprint of higher resolution. DMS is exceptional in that it may be used both in protection and interference types of footprinting. It will therefore be treated further in the section below.

DMS also has the ability to penetrate cells and modify DNA in chromatin. This has been the basis for development of methods for in vivo footprinting allowing the analysis of protein binding sites in a living cell. The first version of in vivo footprinting using DMS was presented by Church and Gilbert.[80] They used DMS treatment of whole cells, DNA isolation followed by restriction digest, piperidine cleavage followed by electrophoretic separation and blotting of the sequencing gel and finally indirect end-labeling by hybridization to visualize the footprint. The introduction of ligation-mediated PCR was an important improvement to increase signal strength.[81,82] The basic principle in this version is that DNA is isolated from DMS-treated cells, cleaved at the modified positions by piperidine, denatured and annealed to a target-specific oligonucleotide. After a fill-in reaction, the generated blunt end is then ligated to a common linker oligonucleotide. This generates fragments of different sizes with know sequences at both ends, and thus amenable to PCR amplification. The PCR amplification is performed using one target-specific primer and one primer complementary to the linker. The generated series of fragments will have one common end (the target-specific primer) and one end corresponding to the modified site that is cleaved, similar to what is the case in classical footprinting methods. To transform it into a series of radioactively labeled fragments, the technique ends with annealing a labeled primer that is extended to the end of the fragments. In parallel, a deproteinized preparation of genomic DNA is processed similarly. By comparing the two patterns obtained with chromatin versus naked template, the protective presence of proteins in the analyzed region can be demonstrated.

Several further developments of these principles have been reported. In particular, the analysis of genomic DNA in permealized cells or isolated nuclei by similar strategies expands largely the types of modifying agents that can be used.[63] Genomic and in vivo footprinting have been treated in several reviews.[62,83,84]

2.2.4 Exonuclease Footprinting

Probably the simplest version of protection footprinting methods is exonuclease footprinting. A complex is formed between an end-labeled DNA fragment and a specific protein, and the complex is treated with an exonuclease that digests the DNA in a directional fashion until it is hindered by the bound protein. The type of labeling must match the type of exonuclease used. A 5′-labeled DNA is suitable for 3′–5′ types of exonucleases (like exonuclease III), while a 3′-labeled DNA can be treated with 5′–3′ types of exonucleases (like the λ-exonuclease). The information provided by this type of footprinting is limited since only the borders of a protein–DNA complex

are mapped. On the other hand, it has the advantage of generating a positive signal in that the complex border is evidenced by formation of a novel specific band. In its simplest version, only one band is formed and a gel with lower resolution may be used to determine the size of the signal and thus to map the border of the complex.

The principle of exonuclease footprinting is illustrated in Figure 5.

Protocols for exonuclease footprinting can be found in specific reports,[54,77] and in several reviews.[63,85]

2.2.5 Hydroxyl Radical Footprinting

The smallest reactive agent used in protection footprinting is the hydroxyl radical, OH•. This highly reactive molecule is comparable in size to water and will penetrate many parts of the complex and attack DNA at most regions that are not in tight contact with the protein. No sequence preference is operating. This provides a footprint of high resolution with a lot of detailed information.[86] The hydroxyl radical is generated by the Fenton reaction (Equation 5):

$$Fe^{2+} + H_2O_2 = Fe^{3+} + OH^- + OH\bullet \qquad (5)$$

Fe^{2+} is complexed with EDTA under these conditions and will not associate with DNA due to the negative charge of the complex. The hydroxyl radicals formed will attack the deoxyribose moieties and lead to chain cleavage. No separate cleavage step is therefore required. This method often does not produce totally clear zones in the footprint and it is an advantage to be able to scan the autoradiograms to quantitate the footprint. It is a technically challenging footprinting method in particular because of the highly reactive agent involved. In some cases problems may arise because the hydroxyl radical attacks and inactives the protein. The short life-time of the reactive radical also demands rapid ways of mixing the hydroxyl radical-producing reagents with the protein–DNA complex. One should take care not to use protein solutions containing glycerol, since this is a radical scavenger and inhibits cleavage when present at concentrations higher than 0.5%.[29] Protocols for hydroxyl radical footprinting can be found in several reviews,[87–89] or in practically-oriented book chapters.[36,37]

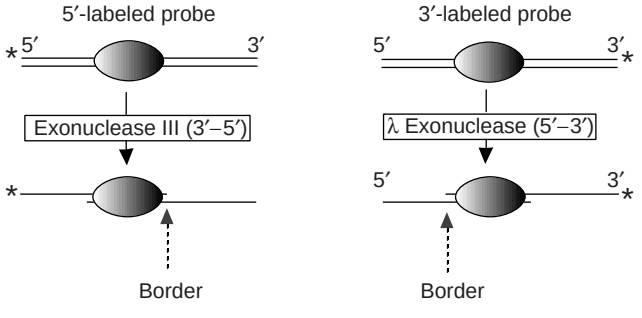

Figure 5 Schematic representation of the principle of exonuclease footprinting methods.

2.2.6 Footprinting by other Chemical Cleavage Agents

Other chemical cleavage reagents used in footprinting methods include methidium–propyl–EDTA–Fe(II) (MPE-Fe(II)) which becomes associated with DNA by intercalation through its methidium group and thus acts more locally than hydroxyl radicals generated by the Fenton reaction. MPE-Fe(II) will, in the presence of ferrous ions, DTT, dioxygen and other reducing agents, efficiently cleave DNA. It cleaves the DNA backbone by an oxidative degradation of the deoxyribose ring, most likely by hydroxyl radicals.[90] One advantage is the low sequence specificity producing a near-random cleavage pattern. This footprinting method was originally developed to map location of the binding sites for small molecules on DNA,[91,92] but it has later been used in protein–DNA footprinting analysis, including analysis of specific DNA–protein complexes from a crude nuclear extract.[93]

A related chemical DNA cleavage agent used for footprinting is 1,10-phenanthroline–copper. In its cuprous form, the coordination complex with hydrogen peroxide as a coreactant cleaves DNA by oxidatively attacking the deoxyribose moiety. An interesting practical application of this version is the direct footprinting of protein–DNA complexes embedded within the acrylamide matrix of an EMSA gel.[37,94] This chemical agent associates with DNA in the minor groove where it leads to oxidative attack on the deoxyribose followed by strand cleavage. Protocols for these versions of footprinting can be found in specific reports,[93,95–97] in reviews,[87] or in practically-oriented book chapters.[37]

2.2.7 Footprinting of Structural Alterations in DNA

Binding of a protein to DNA may induce structural alterations including strand separations as occur in transcription and replication complexes. Such alterations can be probed by different footprinting agents[23,63,98–100] that modify DNA in a structure-dependent fashion. Examples are $KMnO_4$ and OsO_4 which have a strong preference for single-stranded pyrimidines. Similarly, diethylpyrocarbonate (DEPC) carbethoxylates purines at the N-7 position with a strong preference for single-stranded loops and also for purine residues in Z-DNA which appear structurally more accessible than in B-DNA.

2.2.8 Interference Footprinting Methods

Interference footprinting involves pretreatment of a duplex oligonucleotide or DNA fragment with agents that modify the DNA. The modified DNA is then used as a binding probe for a specific DNA-binding protein. The different methods vary with respect to the type of modification agent used. Their common feature is that they provide information on how DNA modification reactions in the binding site affect sequence-specific protein binding. The general principle of interference footprinting is illustrated in Figure 6.

An end-labeled duplex oligonucleotide or DNA fragment is here first treated by a modification agent under conditions that introduces ''single hit'' modifications randomly distributed along the sequence of DNA. This pretreated DNA is then allowed to bind to a specific DNA-binding protein under conditions that in an EMSA analysis would have given a reasonable amount of both complex and free probe (that is nonsaturating conditions). Then the complex is separated from noncomplexed DNA using an EMSA gel or using nitrocellulose filter binding. Both fractions of DNA (bound and nonbound) are isolated (eluted from the EMSA gel or from the filter) and subjected to cleavage by a reagent that cleaves specifically at the modified residues. A control of pretreated DNA without the protein is processed in parallel. When the generated fragments of DNA are then separated on a sequencing gel, the pattern of labeled bands will reveal at which positions the presence of a modified residue has interfered with protein binding. Fragments corresponding to these critical positions will be under-represented in the bound fraction, over-represented in the nonbound fraction, and have a normal intensity in the control fraction (note that the simplified version shown in Figure 6 only shows the bound and the control fraction). From this information, the location of the bound protein can be deduced.

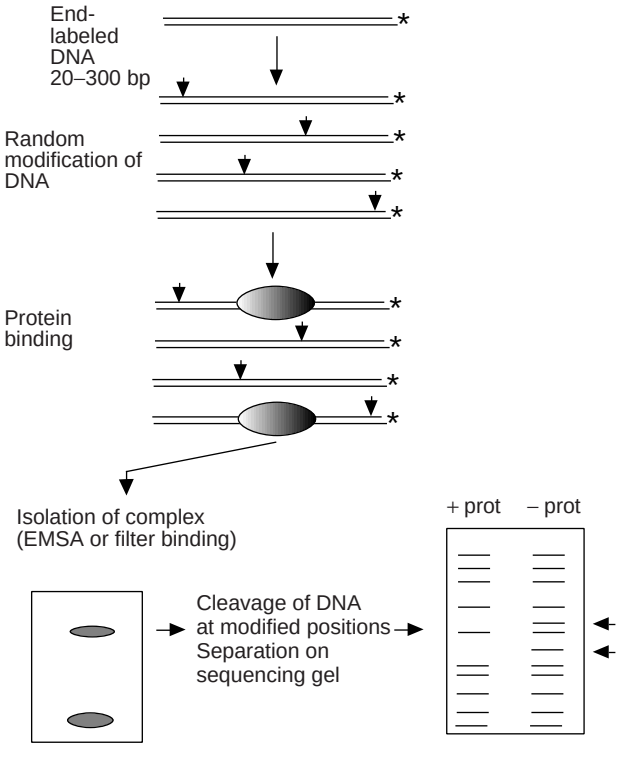

Figure 6 Schematic representation of the principle of interference footprinting methods.

2.2.8.1 Interference Footprinting by Dimethyl Sulfate Probably the most common version of interference footprinting is analysis of DMS interference.[36,37,70,101–103] As mentioned above, the alkylating agent DMS efficiently methylates N-7 of guanine bases, and thereby destroys critical contact points used for hydrogen bonding in many protein–DNA complexes. This type of interference footprinting therefore reveals guanine bases where interaction with N7 is of critical importance for sequence-specific DNA binding, as illustrated in Figure 7.

2.2.8.2 Missing Base Interference Footprinting While DMS interference is limited to analysis of guanine bases (in the major groove), missing base interference footprinting is an informative version of interference footprinting that may provide data on the importance of all four bases in a binding site.[104] Different chemical treatments generate either single-hit depurinated DNA (by treatment with formiate pH 2) or depyrimidated

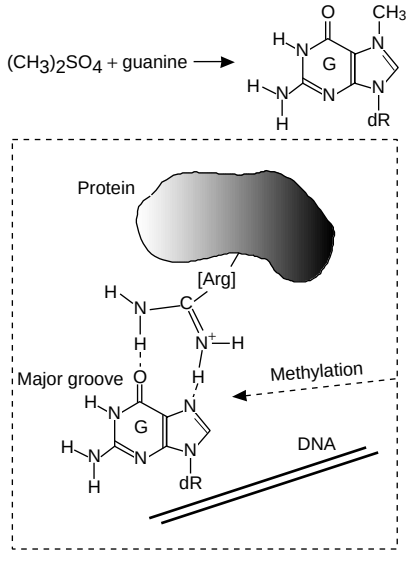

Figure 7 Schematic representation of the principle of DMS interference footprinting. *Protection*: maps DMS accessibility. *Interference*: maps guanines where N7 contributes to binding. Methylation interferes with specific H-bond interactions between protein and DNA in the major groove.

DNA (by treatment with hydrazine). In these cases the entire bases are removed, and the argument is that if their presence is important for specific interactions with the protein, their absence will abolish or reduce binding. The cleavage reaction specific for the apurinic/apyrimidinic site is treatment with a solution of hot piperidine.

Protocols for missing base interference analysis can be found in Ording et al.[54] and Hørring et al.[105]

2.2.8.3 Missing Nucleoside Interference Footprinting
This footprinting method is related to hydroxyl radical protection footprinting, but is performed as an interference version. This means that the end-labeled DNA is pretreated with hydroxyl radicals as described above. This modification leads to destruction of the base and sugar attacked (nucleoside) creating randomly distributed one-residue gaps in the strands of the duplex. This modified DNA can then be used as a probe according to a standard interference strategy. One advantage is that all nucleotides in a binding site can be probed. However, the gaps introduced may possibly alter the flexibility or conformation of the DNA, which may affect protein–DNA interactions.[106–108]

2.2.8.4 Ethylation Interference Footprinting While most modification reagents attack either the bases or the sugar moieties in DNA, one reagent also attacks the phosphate groups and thus allows the analysis of critical phosphate interactions in the protein–DNA complex. This reagent is *N*-ethyl, *N*-nitroso-urea (ENU). ENU is an alkylating agent that will ethylate phosphate groups in DNA in a random fashion. This modification will interfere with protein binding at critical positions both due to the removal of a negative charge on the alkylated phosphate and because it introduces a bulky group that may interfere with binding for steric reasons. The modification specific cleavage reaction is performed under alkaline conditions.[103,109–111]

2.2.8.5 Uracil Interference Footprinting Interactions between hydrophobic side chains and the 5-methyl group in thymine have been observed in many protein–DNA complexes.[4] A method has been devised to specifically analyze the importance of these interactions.[112] The thymine bases are randomly replaced with uracil (lacking the 5-methyl group) using a PCR reaction in the presence of a mixture of TTP and dUTP that create ''single hit'' frequencies of substitutions. This modified PCR product is then used as a probe for binding according to the general interference strategy. The uracil-specific cleavage that is required after separation of bound and nonbound DNA is performed in two steps: first enzymatically using uracil-*N*-glycosylase that creates an apyrimidinic site, then chemically with piperidine. In this way it is possible to identify at which positions the removal of a 5′-methyl group in thymines will affect the binding of sequence-specific protein.[112–114]

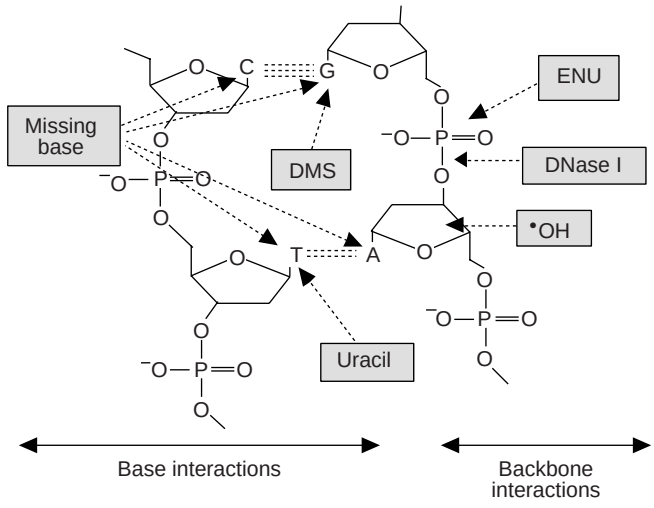

Figure 8 Schematic representation of the different parts of the DNA duplex targeted by the various footprinting methods.

2.2.9 Summary

The different versions of protection and interference footprinting methods described in this review exploit the modification of different parts of the DNA duplex, either specific bases or backbone components. For a summary, these various points of attack are illustrated in Figure 8.

ABBREVIATIONS AND ACRONYMS

APS	Ammonium Persulfate
ATP	Adenosine Triphosphate
DEPC	Diethylpyrocarbonate
DMS	Dimethyl Sulfate
DNaseI	DNA Polymerase I
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetraacetic Acid
EMSA	Electrophoretic Mobility Shift Assay
ENU	*N*-Ethyl, *N*-Nitroso-urea
GMSA	Gel Mobility Shift Assay
GST	Glutathione S-Transferase
IPTG	Isopropyl-β-D-Thiogalactopyranoside
MPE-Fe(II)	Methidium–Propyl–EDTA–Fe(II)
NE	Nuclear Extracts
NMR	Nuclear Magnetic Resonance
PA	Polyacrylamide
PCR	Polymerase Chain Reaction
PMSF	Phenylmethylsulfonyl Fluoride
TCA	Trichloroacetic Acid
TEMED	*N*,*N*,*N*′,*N*′-Tetramethylenediamine

WCE	Whole Cell Extracts
3D	Three-dimensional

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Polymerase Chain Reaction and Other Amplification Systems • RNA Tertiary Structure • Sequencing Strategies and Tactics in DNA and RNA Analysis • X-ray Structures of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Protein Purification: Theoretical and Methodological Considerations • X-ray Crystallography of Biological Macromolecules

REFERENCES

1. M. Suzuki, 'A Framework for the DNA–Protein Recognition Code of the Probe Helix in Transcription Factors: the Chemical and Stereochemical Rules', *Structure*, **2**(4), 317–326 (1994).
2. Y. Choo, A. Klug, 'Physical Basis of a Protein–DNA Recognition Code', *Curr. Opin. Struct. Biol.*, **7**(1), 117–125 (1997).
3. G. Patikoglou, S.K. Burley, 'Eukaryotic Transcription Factor–DNA Complexes', *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 289–325 (1997).
4. D. Rhodes, J.W. Schwabe, L. Chapman, L. Fairall, 'Towards an Understanding of Protein–DNA Recognition', *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **351**(1339), 501–509 (1996).
5. M. Billeter, 'Homeodomain-type DNA Recognition', *Prog. Biophys. Mol. Biol.*, **66**(3), 211–225 (1996).
6. M. Suzuki, S.E. Brenner, M. Gerstein, N. Yagi, 'DNA Recognition Code of Transcription Factors', *Protein Eng.*, **8**(4), 319–328 (1995).
7. D.E. Draper, 'Protein–RNA Recognition', *Annu. Rev. Biochem.*, **64**, 593–620 (1995).
8. W.J. Gehring, Y.Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A.F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, K. Wuthrich, 'Homeodomain-DNA Recognition', *Cell*, **78**(2), 211–223 (1994).
9. S.C. Harrison, A.K. Aggarwal, 'DNA Recognition by Proteins with the Helix-turn-helix Motif', *Annu. Rev. Biochem.*, **59**, 933–969 (1990).
10. M.T. Record Jr, C.F. Anderson, T.M. Lohman, 'Thermodynamic Analysis of Ion Effects on the Binding and Conformational Equilibria of Proteins and Nucleic Acids: the Roles of Ion Association or Release, Screening, and Ion Effects on Water Activity', *Q. Rev. Biophys.*, **11**(2), 103–178 (1978).
11. T.A. Steitz, 'Structural Studies of Protein–Nucleic Acid Interaction: the Sources of Sequence-specific Binding,' *Q. Rev. Biophys.*, **23**(3), 205–280 (1990).
12. P.S. Freemont, A.N. Lane, M.R. Sanderson, 'Structural Aspects of Protein–DNA Recognition', *Biochem. J.*, **278**(Pt 1), 1–23 (1991).
13. G.D. Stormo, D.S. Fields, 'Specificity, Free Energy and Information Content in Protein–DNA Interactions', *Trends Biochem Sci.*, **23**(3), 109–113 (1998).
14. D.H. Ohlendorf, J.B. Matthew, 'Electrostatics and Flexibility in Protein–DNA Interactions', *Adv. Biophys.*, **20**, 137–151 (1985).
15. M. Tateno, K. Yamasaki, N. Amano, J. Kakinuma, H. Koike, M.D. Allen, M. Suzuki, 'DNA Recognition by Beta-sheets', *Biopolymers*, **44**(4), 335–359 (1997).
16. N.C. Seeman, J.M. Rosenberg, A. Rich, 'Sequence-specific Recognition of Double Helical Nucleic Acids by Proteins', *Proc. Natl. Acad. Sci. USA*, **73**(3), 804–808 (1976).
17. D.M. Crothers, M.R. Gartenberg, T.E. Shrader, 'DNA Bending in Protein–DNA Complexes', *Methods Enzymol.*, **208**, 118–146 (1991).
18. R.K. Allemann, M. Egli, 'DNA Recognition and Bending', *Chem. Biol.*, **4**(9), 643–650 (1997).
19. J.K. Strauss, L.J. Maher III, 'DNA Bending by Asymmetric Phosphate Neutralization', *Science*, **266**(5192), 1829–1834 (1994).
20. C.A. Bewley, A.M. Gronenborn, G.M. Clore, 'Minor Groove-binding Architectural Proteins: Structure, Function, and DNA Recognition', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 105–131 (1998).
21. R.S. Spolar, M.T. Record Jr, 'Coupling of Local Folding to Site-specific Binding of Proteins to DNA', *Science*, **263**(5148), 777–784 (1994).
22. J.A. Lefstin, K.R. Yamamoto, 'Allosteric Effects of DNA on Transcriptional Regulators', *Nature*, **392**(6679), 885–888 (1998).
23. K.M. Weeks, D.M. Crothers, 'Major Groove Accessibility of RNA', *Science*, **261**(5128), 1574–1577 (1993).
24. D.E. Draper, L.P. Reynaldo, 'RNA Binding Strategies of Ribosomal Proteins', *Nucleic Acids Res.*, **27**(2), 381–388 (1999).
25. G. Varani, K. Nagai, 'RNA Recognition by RNP Proteins During RNA Processing', *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 407–445 (1998).
26. K. Nagai, 'RNA–Protein Complexes', *Curr. Opin. Struct. Biol.*, **6**(1), 53–61 (1996).
27. M. Fried, D.M. Crothers, 'Equilibria and Kinetics of Lac Repressor-operator Interactions by Polyacrylamide Gel Electrophoresis', *Nucleic Acids Res.*, **9**(23), 6505–6525 (1981).
28. M.M. Garner, A. Revzin, 'A Gel Electrophoresis Method for Quantifying the Binding of Proteins to Specific

DNA Regions: Application to Components of the Escherichia coli Lactose Operon Regulatory System', *Nucleic Acids Res.*, **9**(13), 3047–3060 (1981).
29. M.J. Guille, G.G. Kneale, 'Methods for the Analysis of DNA–Protein Interactions', *Mol. Biotechnol.*, **8**(1), 35–52 (1997).
30. D. Lane, P. Prentki, M. Chandler, 'Use of Gel Retardation to Analyze Protein–Nucleic Acid Interactions', *Microbiol. Rev.*, **56**(4), 509–528 (1992).
31. J.R. Cann, 'Theoretical Studies on the Mobility-shift Assay of Protein–DNA Complexes', *Electrophoresis*, **19**(2), 127–141 (1998).
32. M.G. Fried, 'Measurement of Protein–DNA Interaction Parameters by Electrophoresis Mobility Shift Assay', *Electrophoresis*, **10**(5–6), 366–376 (1989).
33. J.A. Ceglarek, A. Revzin, 'Studies of DNA–Protein Interactions by Gel Electrophoresis', *Electrophoresis*, **10**(5–6), 360–365 (1989).
34. V. Scott, A.R. Clark, K. Docherty, 'The Gel Retardation Assay', *Methods Mol. Biol.*, **31**, 339–347 (1994).
35. C.L. Dent, D.S. Latchman, 'The DNA Mobility Shift Assay', in *Transcription Factors A Practical Approach*, ed. D.S. Latchman, IRL Press, Oxford, 1–26, 1993.
36. D. Rhodes, 'Analysis of Sequence-specific DNA-binding Proteins', in *Protein Function. A Practical Approach*, ed. T.E. Creighton, IRL Press, Oxford, 177–198, 1989.
37. M.J. Garabedian, J. LaBaer, W.-H. Liu, J.R. Thomas, 'Analysis of Protein–DNA Interactions', in *Gene Transcription. A Practical Approach*, IRL Press, eds. B.D. Hames, S.J. Higgins, IRL Press, Oxford, 243–293, 1993.
38. O.S. Gabrielsen, A. Sentenac, P. Fromageot, 'Specific DNA Binding by c-Myb: Evidence for a Double Helix-turn-helix Related Motif', *Science*, **253**(5024), 1140–1143 (1991).
39. J.R. Cann, 'Phenomenological Theory of Gel Electrophoresis of Protein–Nucleic Acid Complexes', *J. Biol. Chem.*, **264**(29), 17032–17040 (1989).
40. J.R. Cann, 'Comment on Mobility-shift Computations Featuring Cage Effects', *Electrophoresis*, **17**(10), 1535–1536 (1996).
41. E.M. Brendeford, A.H. Myrset, A.B. Hegvold, M. Lundin, O.S. Gabrielsen, 'Oncogenic Point Mutations Induce Altered Conformation, Redox Sensitivity, and DNA Binding in the Minimal DNA Binding Domain of Avian Myeloblastosis Virus v-Myb', *J. Biol. Chem.*, **272**(7), 4436–4443 (1997).
42. A.H. Myrset, A. Bostad, N. Jamin, P.N. Lirsac, F. Toma, O.S. Gabrielsen, 'DNA and Redox State Induced Conformational Changes in the DNA-binding Domain of the Myb Oncoprotein', *EMBO J.*, **12**(12), 4625–4633 (1993).
43. J.R. Cann, O. Pfenninger, D.E. Pettijohn, 'Theory of the Mobility-shift Assay of Nonspecific Protein–DNA Complexes Governed by Conditional Probabilities: the HU : DNA Complex', *Electrophoresis*, **16**(6), 881–887 (1995).
44. J.R. Cann, 'Theoretical Studies on the Mobility-shift Behavior of Binary Protein–DNA Complexes', *Electrophoresis*, **14**(8), 669–679 (1993).
45. D.F. Senear, J.B. Ross, T.M. Laue, 'Analysis of Protein and DNA-mediated Contributions to Cooperative Assembly of Protein–DNA Complexes', *Methods*, **16**(1), 3–20 (1998).
46. O.S. Gabrielsen, N. Marzouki, A. Ruet, A. Sentenac, P. Fromageot, 'Two Polypeptide Chains in Yeast Transcription Factor Tau Interact with DNA', *J. Biol. Chem.*, **264**(13), 7505–7511 (1989).
47. J.D. Kahn, E. Yun, D.M. Crothers, 'Detection of Localized DNA Flexibility', *Nature*, **368**(6467), 163–166 (1994).
48. S.J. Sperbeck, G.J. Wistow, 'pBendBlue: Modification of the pBend System for Color Selectability', *Biotechniques*, **24**(1), 66–68 (1998).
49. J. Kim, C. Zwieb, C. Wu, S. Adhya, 'Bending of DNA by Gene-regulatory Proteins: Construction and Use of a DNA Bending Vector', *Gene*, **85**(1), 15–23 (1989).
50. G. McGill, D.E. Fisher, 'DNA Bending and the Curious Case of Fos/Jun', *Chem Biol.*, **5**(2), R29–R38 (1998).
51. J.D. Kahn, D.M. Crothers, 'Protein-induced Bending and DNA Cyclization', *Proc. Natl. Acad. Sci. USA*, **89**(14), 6343–6347 (1992).
52. S.S. Zinkel, D.M. Crothers, 'DNA Bend Direction by Phase Sensitive Detection', *Nature*, **328**(6126), 178–181 (1987).
53. P.C. van der Vliet, C.P. Verrijzer, 'Bending of DNA by Transcription Factors', *Bioessays*, **15**(1), 25–32 (1993).
54. E. Ording, W. Kvavik, A. Bostad, O.S. Gabrielsen, 'Two Functionally Distinct Half Sites in the DNA-recognition Sequence of the Myb Oncoprotein', *Eur. J. Biochem.*, **222**(1), 113–120 (1994).
55. E. Ording, S. Bergholtz, E.M. Brendeford, N. Jamin, O.S. Gabrielsen, 'Flexibility in the Second Half-site Sequence Recognised by the c-Myb R2 domain – In Vitro and In Vivo Analysis', *Oncogene*, **13**(5), 1043–1051 (1996).
56. M.A. Laniel, S.L. Guerin, 'Improving Sensitivity of the Electrophoretic Mobility Shift Assay by Restricting Tissue Phosphatase Activities', *Biotechniques*, **24**(6), 964–970 (1998).
57. J.D. Dignam, 'Preparation of Extracts from Higher Eukaryotes', *Methods Enzymol.*, **182**, 194–203 (1990).
58. J.D. Dignam, P.L. Martin, B.S. Shastry, R.G. Roeder, 'Eukaryotic Gene Transcription with Purified Components', *Methods Enzymol.*, **101**, 582–598 (1983).
59. J.D. Dignam, R.M. Lebovitz, R.G. Roeder, 'Accurate Transcription Initiation by RNA Polymerase II in a Soluble Extract from Isolated Mammalian Nuclei', *Nucleic Acids Res.*, **11**(5), 1475–1489 (1983).
60. M. Oelgeschlager, J. Krieg, J.M. Luscher-Firzlaff, B. Luscher, 'Casein Kinase II Phosphorylation Site Mutations in c-Myb Affect DNA Binding and Transcriptional Cooperativity with NF-M', *Mol. Cell. Biol.*, **15**(11), 5966–5974 (1995).

61. L. Zargarian, V. Le Tilly, N. Jamin, A. Chaffotte, O.S. Gabrielsen, F. Toma, B. Alpert, 'Myb-DNA Recognition: Role of Tryptophan Residues and Structural Changes of the Minimal DNA Binding Domain of c-Myb', *Biochemistry*, **38**(6), 1921–1929 (1999).
62. A.L. Paul, R.J. Ferl, 'In Vivo Footprinting of Protein–DNA Interactions', *Methods Cell. Biol.*, **49**, 391–400 (1995).
63. H.P. Saluz, J.P. Jost, 'Approaches to Characterize Protein–DNA Interactions In Vivo', *Crit. Rev. Eukaryot. Gene Expr.*, **3**(1), 1–29 (1993).
64. H.P. Saluz, K. Wiebauer, A. Wallace, 'Studying DNA Modifications and DNA–Protein Interactions In Vivo. A Window onto the Native Genome', *Trends Genet.*, **7**(7), 207–211 (1991).
65. J.C. Dabrowiak, J. Goodisman, B. Ward, 'Quantitative DNA Footprinting', *Methods Mol. Biol.*, **90**, 23–42 (1997).
66. K.R. Fox, 'DNaseI Footprinting', *Methods Mol. Biol.*, **90**, 1–22 (1997).
67. M. Shubsda, H. Kishikawa, J. Goodisman, J. Dabrowiak, 'Quantitative Footprinting Analysis', *J. Mol. Recognit.*, **7**(2), 133–139 (1994).
68. P.E. Nielsen, 'In Vivo Footprinting: Studies of Protein–DNA Interactions in Gene Regulation', *Bioessays*, **11**(5), 152–155 (1989).
69. T.D. Tullius, 'Physical Studies of Protein–DNA Complexes by Footprinting', *Annu. Rev. Biophys. Biophys. Chem.*, **18**, 213–237 (1989).
70. N.D. Lakin, 'Determination of DNA Sequences that Bind Transcription Factors by DNA Footprinting', in *Transcription Factors. A Practical Approach.*, ed. D.S. Latchman, IRL Press, Oxford, 27–47, 1993.
71. B. Leblanc, T. Moss, 'DNaseI Footprinting', *Methods Mol. Biol.*, **30**, 1–10 (1994).
72. D.J. Galas, A. Schmitz, 'DNAse Footprinting: a Simple Method for the Detection of Protein–DNA Binding Specificity', *Nucleic Acids Res.*, **5**(9), 3157–3170 (1978).
73. V. Petri, M. Brenowitz, 'Quantitative Nucleic Acids Footprinting: Thermodynamic and Kinetic Approaches', *Curr. Opin. Biotechnol.*, **8**(1), 36–44 (1997).
74. M. Hsieh, M. Brenowitz, 'Quantitative Kinetics Footprinting of Protein–DNA Association Reactions', *Methods Enzymol.*, **274**, 478–492 (1996).
75. M. Brenowitz, D.F. Senear, M.A. Shea, G.K. Ackers, 'Quantitative DNase Footprint Titration: a Method for Studying Protein–DNA Interactions', *Methods Enzymol.*, **130**, 132–181 (1986).
76. M. Brenowitz, D.F. Senear, M.A. Shea, G.K. Ackers, ''Footprint'' Titrations Yield Valid Thermodynamic Isotherms', *Proc. Natl. Acad. Sci. USA*, **83**(22), 8462–8466 (1986).
77. S. Camier, O. Gabrielsen, R. Baker, A. Sentenac, 'A Split Binding Site for Transcription Factor tau on the tRNA3Glu Gene', *EMBO J.*, **4**(2), 491–500 (1985).
78. A.M. Maxam, W. Gilbert, 'Sequencing End-labeled DNA with Base-specific Chemical Cleavages', *Methods Enzymol.*, **65**(1), 499–560 (1980).
79. A.M. Maxam, W. Gilbert, 'A New Method for Sequencing DNA', *Proc. Natl. Acad. Sci. USA*, **74**(2), 560–564 (1977).
80. G.M. Church, W. Gilbert, 'Genomic Sequencing', *Proc. Natl. Acad. Sci. USA*, **81**(7), 1991–1995 (1984).
81. G.P. Pfeifer, S.D. Steigerwald, P.R. Mueller, B. Wold, A.D. Riggs, 'Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR', *Science*, **246**(4931), 810–813 (1989).
82. P.R. Mueller, B. Wold, 'In Vivo Footprinting of a Muscle Specific Enhancer by Ligation Mediated PCR', *Science*, **246**(4931), 780–786 (1989).
83. K. Zaret, 'Identifying Specific Protein–DNA Interactions within Living Cells, or in 'In Vivo Footprinting'', *Methods*, **11**(2), 149–150 (1997).
84. I.K. Hornstra, T.P. Yang, 'In Vivo Footprinting and Genomic Sequencing by Ligation-mediated PCR', *Anal. Biochem.*, **213**(2), 179–193 (1993).
85. W. Metzger, H. Heumann, 'Footprinting with Exonuclease III', *Methods Mol. Biol.*, **30**, 11–20 (1994).
86. T.D. Tullius, 'DNA Footprinting with Hydroxyl Radical', *Nature*, **332**(6165), 663–664 (1988).
87. D.S. Sigman, C.H. Chen, 'Chemical Nucleases: New Reagents in Molecular Biology', *Annu. Rev. Biochem.*, **59**, 207–236 (1990).
88. W.J. Dixon, J.J. Hayes, J.R. Levin, M.F. Weidner, B.A. Dombroski, T.D. Tullius, 'Hydroxyl Radical Footprinting', *Methods Enzymol.*, **208**, 380–413 (1991).
89. T.D. Tullius, B.A. Dombroski, M.E. Churchill, L. Kam, 'Hydroxyl Radical Footprinting: a High-resolution Method for Mapping Protein–DNA Contacts', *Methods Enzymol.*, **155**, 537–558 (1987).
90. R.P. Hertzberg, P.B. Dervan, 'Cleavage of DNA with Methidiumpropyl–EDTA–Iron(II): Reaction Conditions and Product Analyzes', *Biochemistry*, **23**(17), 3934–3945 (1984).
91. M.W. Van Dyke, R.P. Hertzberg, P.B. Dervan, 'Map of Distamycin, Netropsin, and Actinomycin Binding Sites on Heterogeneous DNA: DNA Cleavage-inhibition Patterns with Methidiumpropyl–EDTA–Fe(II)', *Proc. Natl. Acad. Sci. USA*, **79**(18), 5470–5474 (1982).
92. I.L. Cartwright, R.P. Hertzberg, P.B. Dervan, S.C. Elgin, 'Cleavage of Chromatin with Methidiumpropyl–EDTA–Iron(II)', *Proc. Natl. Acad. Sci. USA*, **80**(11), 3213–3217 (1983).
93. N.F. Landolfi, X.M. Yin, J.D. Capra, P.W. Tucker, 'Protection Analysis (or 'Footprinting') of Specific Protein–DNA Complexes in Crude Nuclear Extracts Using Methidiumpropyl–EDTA–Iron (II)', *Biotechniques*, **7**(5), 500–504 (1989).
94. M.D. Kuwabara, D.S. Sigman, 'Footprinting DNA–Protein Complexes In Situ Following Gel Retardation Assays Using 1,10-Phenanthroline–Copper Ion:

Escherichia Coli RNA Polymerase-lac Promoter Complexes', *Biochemistry*, **26**(23), 7234–7238 (1987).
95. Q. Guo, M. Lu, M. Shahrestanifar, R.D. Sheardy, N.R. Kallenbach, 'Drug Binding to a DNA BZ Molecule: Analysis by Chemical Footprinting', *Biochemistry*, **30**(51), 11735–11741 (1991).
96. C.P. Vary, J.N. Vournakis, 'RNA Structure Analysis Using Methidiumpropyl–EDTA–Fe(II): a Base-pair-specific RNA Structure Probe', *Proc. Natl. Acad. Sci. USA*, **81**(22), 6978–6982 (1984).
97. A.S. Kimball, M.L. Kimball, M. Jayaram, T.D. Tullius, 'Chemical Probe and Missing Nucleoside Analysis of Flp Recombinase Bound to the Recombination Target Sequence', *Nucleic Acids Res.*, **23**(15), 3009–3017 (1995).
98. P.E. Nielsen, 'Uranyl Photofootprinting of DNA–Protein Complexes', *Methods Mol. Biol.*, **30**, 141–149 (1994).
99. J.A. McClellan, 'Osmium Tetroxide Modification and the Study of DNA–Protein Interactions', *Methods Mol. Biol.*, **30**, 97–112 (1994).
100. M.J. McLean, 'Diethyl Pyrocarbonate as a Probe of Protein–DNA Interactions', *Methods Mol. Biol.*, **30**, 89–95 (1994).
101. P.E. Shaw, A.F. Stewart, 'Identification of Protein–DNA Contacts with Dimethyl Sulfate. Methylation Protection and Methylation Interference', *Methods Mol. Biol.*, **30**, 79–87 (1994).
102. M. Affolter, A. Percival-Smith, M. Muller, W. Leupin, W.J. Gehring, 'DNA Binding Properties of the Purified Antennapedia Homeodomain', *Proc. Natl. Acad. Sci. USA*, **87**(11), 4093–4097 (1990).
103. J.A. Nye, B.J. Graves, 'Alkylation Interference Identifies Essential DNA Contacts for Sequence-specific Binding of the Eukaryotic Transcription Factor C/EBP', *Proc. Natl. Acad. Sci. USA*, **87**(10), 3992–3996 (1990).
104. A. Brunelle, R.F. Schleif, 'Missing Contact Probing of DNA–Protein Interactions', *Proc. Natl. Acad. Sci. USA*, **84**(19), 6673–6676 (1987).
105. P.I. Hovring, A. Bostad, E. Ording, A.H. Myrset, O.S. Gabrielsen, 'DNA-binding Domain and Recognition Sequence of the Yeast BAS1 Protein, a Divergent Member of the Myb Family of Transcription Factors', *J. Biol. Chem.*, **269**(26), 17663–17669 (1994).
106. J.J. Hayes, T.D. Tullius, 'The Missing Nucleoside Experiment: a New Technique to Study Recognition of DNA by Protein', *Biochemistry*, **28**(24), 9521–9527 (1989).
107. P. Schickor, H. Heumann, 'Hydroxyl Radical Interference', *Methods Mol. Biol.*, **30**, 33–41 (1994).
108. G. Chalepakis, M. Beato, 'Hydroxyl Radical Interference: a New Method for the Study of Protein–DNA Interactions', *Nucleic Acids Res.*, **17**(4), 1783 (1989).
109. I.H. Kim, G.P. Rodgers, 'In Vivo Footprinting Using *N*-ethyl, *N*-nitrosourea: Improved Resolution of the DNA–Protein Interactions in the Human Gamma-globin Gene Promoter Region', *Anal. Biochem.*, **254**(1), 1–8 (1997).
110. I. Manfield, P.G. Stockley, 'Ethylation Interference', *Methods Mol. Biol.*, **30**, 125–139 (1994).
111. F.D. Bushman, J.E. Anderson, S.C. Harrison, M. Ptashne, 'Ethylation Interference and X-ray Crystallography Identify Similar Interactions Between 434 Repressor and Operator', *Nature*, **316**(6029), 651–653 (1985).
112. W.T. Pu, K. Struhl, 'Uracil Interference, a Rapid and General Method for Defining Protein–DNA Interactions Involving the 5-methyl Group of Thymines: the GCN4–DNA Complex', *Nucleic Acids Res.*, **20**(4), 771–775 (1992).
113. P.R. Devchand, J.D. McGhee, J.H. Van de Sande, 'Uracil–DNA Glycosylase as a Probe for Protein–DNA Interactions', *Nucleic Acids Res.*, **21**(15), 3437–3443 (1993).
114. P.R. Devchand, J.D. McGhee, J.H. Van de Sande, 'Analysis of the Tet Repressor-operator Interactions Using the Uracil–DNA Glycosylase Footprinting System', *Ann. N.Y. Acad. Sci.*, **726**, 309–311 (1994).

Proteolytic Mapping

Tim Nadler
PE Biosystems, Framingham, USA

1 Introduction **1**
1.1 Description of Proteolytic Mapping 1
1.2 Purpose and Uses of Proteolytic Mapping 2
1.3 History of Proteolytic Mapping 2
1.4 Limitations 2
2 Sample Preparation **3**
2.1 Solubility Requirements 3
2.2 Reduction 3
2.3 Alkylation 4
3 Proteolytic Digestion **5**
3.1 Chemistry of Digestion 5
3.2 Selection of Enzyme or Reagent 5
3.3 Typical Protocol for Trypsin 7
3.4 Troubleshooting 7
4 Map Development **8**
4.1 Column Selection 8
4.2 Mobile Phase Selection 9
4.3 Gradient Optimization 9
4.4 Reproducibility 9
4.5 Typical Protocol 9
5 Identification of Peptide Peaks **10**
5.1 Retention Time Prediction 10
5.2 Ultraviolet Spectral Analysis 10
5.3 Amino Acid Analysis 10
5.4 Peptide Sequence Analysis 10
5.5 Mass Spectrometry 11
6 Automation of Proteolytic Mapping **12**
6.1 Automating Reduction and Alkylation 12
6.2 On-line Digestion with Immobilized Enzymes 13
6.3 Liquid Chromatography/Mass Spectrometry to Automate Peak Identification 13
6.4 Database Searching to Identify Proteins 14
7 Selected Applications **14**
7.1 Mapping in the Quality Control Environment 14
7.2 Miniaturization for Increased Sensitivity in Research and Development 15
7.3 Application to Proteomics 15
Abbreviations and Acronyms **15**
Related Articles **16**
References **16**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Proteolytic mapping is a technique used to verify the primary structure of a protein. It involves the cleavage of the protein at specific amino acids to generate a unique set of peptides which is then separated and analyzed. The chromatogram generated during the separation is known as the proteolytic map of the protein, which serves as a unique ''fingerprint'' for that protein. A proteolytic map can be used to verify the identity of a protein and to detect mutations in a protein. This technique is faster than two-dimensional (2D) nuclear magnetic resonance (NMR), X-ray crystallography, or complete amino acid sequencing for verifying the primary structure of a protein. However, it does not allow determination of the sequence of a protein de novo *as can be accomplished with sequencing, nor does it yield the higher order structural information that may be obtained with 2D NMR or X-ray crystallography.*

The process of proteolytic mapping involves sample preparation, enzyme digestion, and map development. Each of these steps should be optimized for each new protein to obtain a good map; alternatively, generic protocols can be used to produce satisfactory results for some proteins. A typical protocol requires about a day to complete owing to the time required to digest a protein with an enzyme. However, more recent automated techniques are capable of completing the process in a few hours because they employ a highly concentrated immobilized enzyme which speeds digestion. Proteolytic maps are used in both quality control (QC) and research environments, with the most recent application in the field of proteomics.

1 INTRODUCTION

1.1 Description of Proteolytic Mapping

Proteins are polymers of amino acids that have a defined amino acid sequence and a specific three-dimensional structure when properly folded. Protein structure is classified into four levels. The sequence of the amino acids is the primary structure of a protein. Short lengths of amino acids can form helices, sheets or turns which are defined as the secondary structure of the protein. These elements fold together to define the overall three-dimensional structure of the protein which is known as the tertiary structure. Finally, some proteins combine with other proteins or additional copies of the same protein to form a complex (e.g. hemoglobin is composed of two α- and two β-chains) which defines a quaternary structure.

However, not all proteins have a quaternary structure. It is the primary structure (amino acid sequence) which defines the higher order structure (how the protein folds), and it is the three-dimensional structure of the protein which defines the function and activity of the protein.

During the synthesis of a protein, or afterwards, amino acids may be modified; for example, by phosphorylation, addition of carbohydrate, fatty acid acylation, reduction, oxidation, and proteolytic cleavage. These modifications are known as post-translational modifications (PTMs) and can be very important to the activity and proper function of a protein. Many techniques are available to probe the structure of a protein including X-ray crystallography, 2D NMR, SDS/PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis), isoelectric focusing, mass spectrometry (MS), circular dichroism, activity assays, Western blots, immunoaffinity techniques, and proteolytic maps. The focus of this chapter will be on proteolytic mapping and how it is used to verify the primary structure of a protein. The first step involves specifically cutting the protein into small peptides with either endoproteases (e.g. trypsin) or chemical cleavage (cyanogen bromide, CNBr). Cleavage is only performed at specified amino acids so the fragments generated are not random. The peptides are then separated, usually by chromatography, to generate a map that is specific to that protein. Changes in the amino acid sequence or in the PTMs will result in changes in the map. There are a number of excellent reviews discussing analysis by the proteolytic map technique.[1–3]

1.2 Purpose and Uses of Proteolytic Mapping

Proteolytic mapping is typically used to verify the primary structure of a protein, although it can also be a useful tool in determining disulfide linkages and ascertaining tertiary structure. It is primarily the PTMs and amino acid substitutions which can be found by proteolytic mapping. Current peptide mapping techniques are sensitive enough to detect these modifications even when they are only present at a level of about 5% of the total protein product.[4–6]

The peptide map alone is only sufficient to determine that there is a change in structure in relation to a known standard. Proteolytic mapping can also be used to identify an unknown protein (provided the sequence is in a database) by analyzing the peptide fragments by MS. With the list of masses of the peptide fragments and known conditions for digestion (e.g. tryptic digestion at only arginine and lysine residues) one may search the databases of known proteins to determine if the masses in the list match a theoretical digest of a protein in the database. This technique is used in proteomics to identify proteins from 2D gels.

1.3 History of Proteolytic Mapping

The basic protocols for proteolytic mapping are based on work from the 1960s by Crestfield et al.[7] and Cleland[8] that has since been developed to the point of its current use in QC.[9–11] Initially, proteolytic digests were analyzed by thin-layer chromatography or SDS/PAGE. These map development techniques have since been replaced by high-performance liquid chromatography (HPLC) (see Figure 1 for an example peptide map); however, the digestion protocols have changed little. Although immobilized enzymes have been used for a long time,[12,13] they have only gained acceptance for proteolytic mapping since more robust immobilized-enzyme cartridges have become available.[14–19] These cartridges are based on a hydrophilic-coated polystyrene with wide pores (2000–4000 Å in diameter). The polystyrene base provides a pH-stable rigid support while the hydrophilic coating minimizes interactions with the proteins and peptides. Using these immobilized-enzyme cartridges, it has been possible to automate the entire process from reduction and alkylation, through digestion, peptide separation, and peak identification by MS.[14]

1.4 Limitations

Again, the peptide map alone is only sufficient either to confirm the identity of a protein by comparison with a previous map of the same protein or to determine if there is a change relative to a previous map. The map in itself does not identify the nature of the change (amino acid mutation or different PTM). MS or peptide sequence analysis may be used to identify the peaks in the map. The combination of liquid

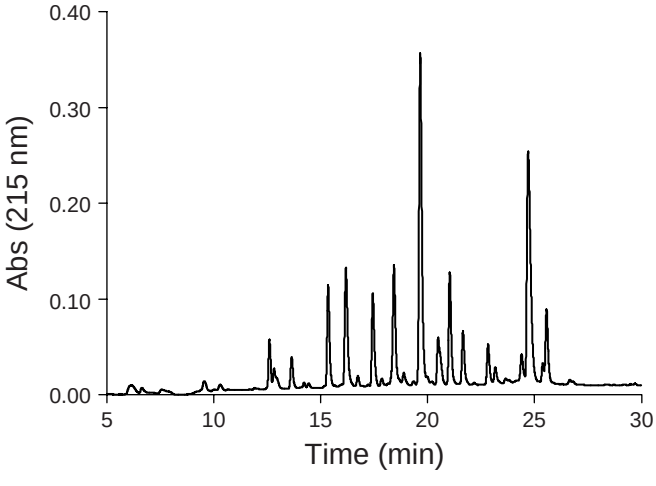

Figure 1 Example proteolytic map of equine heart cytochrome c digested with trypsin and developed on a reversed-phase HPLC system monitoring absorbance at 215 nm.

chromatography with MS in peptide mapping is a very powerful tool for protein structure elucidation and protein identification.

2 SAMPLE PREPARATION

2.1 Solubility Requirements

2.1.1 Reagents to Aid Solubility and Denaturation

Prior to proteolytic digestion, it is generally desirable to reduce the disulfide linkages between cysteine residues within the protein to assure complete digestion. To prevent disulfide reformation during later stages of the process, the sulfhydryl groups are blocked by alkylation. During the reduction process, proteins will often unfold and become insoluble in standard aqueous buffers. Also, the reduction will proceed more quickly if the protein is purposely denatured. Furthermore, the general chemistry of proteolytic digestion is most easily accomplished if the protein is soluble. For these reasons, chaotropic denaturants (e.g. 8-M urea [CAS 57-13-6] or 6-M guanidine hydrochloride [CAS 50-01-1]) are generally added prior to reduction in order to aid in the unfolding and solubilization of the protein. When using urea, it is important to prepare the solution fresh from dry stock in order to avoid carbamoylation from the cyanates which form upon storage of a urea solution. However, the cyanates may be removed from a stock solution by passing the solution through a bed of charged anion-exchange beads (e.g. Dowex™ beads) or placing the beads in the stock solution. Although detergents also aid in solubilization and denaturation of proteins, interference with either the digestion of the protein or in the analysis of the peptides in later steps of the proteolytic mapping can occur. If detergents are used, they may be removed by ion-exchange chromatography,[20] but currently most protocols favor the use of chaotropes.

2.1.2 Digestion of Precipitated Proteins

When digesting a protein in a solution-phase format, endoproteinases will still digest a precipitated protein, and the liberated peptides will often be soluble. Generally, the digest will require longer incubation times than if the protein remained in solution. The chemical cleavage reagents also tend to work on precipitated protein but may require longer reaction times. However, if an immobilized proteolytic enzyme is to be used, steps must be taken to resuspend the protein for injection on to a column containing the immobilized protease.

2.2 Reduction

2.2.1 Purpose

Tightly folded proteins are difficult to digest. Denaturants are used to unfold a protein partially, but they have no effect on the covalent disulfide bonds that cross-link many proteins. Since these disulfide linkages between cysteine residues are quite effective in protecting a protein against denaturation, it is important to reduce these bonds to get an effective and complete digestion. Reduction is not required for proteins lacking internal disulfide linkages. Also, in the case of disulfide linkage assignment, it is desirable to digest the protein as much as possible without reducing these linkages.

2.2.2 Chemistry of Reduction

Typically, disulfide bonds are broken by adding excess reducing agents such as dithiothreitol (DTT) or mercaptoethanol [CAS 60-24-2] under neutral to alkaline conditions. However, phosphine reagents such as tris(2-carboxyethyl)phosphine (TCEP; CAS 51805-45-9) can also be employed under acidic conditions.[21–24] During the reduction process one proton is added to each sulfur, hence, two reducing equivalents are required to reduce one disulfide bond. Although two molecules of mercaptoethanol will be required to reduce a disulfide bond, note that a single molecule of DTT (CAS 27656-41-9) is capable of completely reducing a disulfide bond since it carries two reducing equivalents.

2.2.3 Reagents

A very common reducing agent used for peptide mapping is DTT. It performs best under slightly alkaline conditions (pH 8) and more quickly if the protein is denatured and an elevated temperature is employed. Mercaptoethanol is more commonly used as a reducing agent for proteins prior to SDS/PAGE, but it can also be employed in peptide mapping. TCEP is being used with increasing frequency since phosphine reagents are effective even under acidic conditions.[21–24]

2.2.4 Typical Protocol

1. Dissolve 2 mg of protein in 1 mL of freshly prepared 50-mM TRIS (tris(hydroxymethyl)aminomethane; CAS 77-86-1), pH 8.0 with 8-M urea.
2. Add 100 µL of 50-mM DTT to bring the concentration to about 4.5 mM.
3. Heat the solution at 50 °C for 20 min.
4. Cool to room temperature prior to alkylation.

2.2.5 Troubleshooting

A number of problems can occur during reduction. If the protein is highly disulfide-linked, more DTT may be required to reduce all the bonds. The conditions above assume that the protein contains one disulfide bond for each 3000 Da of mass (i.e. cysteine content of about 6%). If the number of disulfides is unknown and reduction seems to be incomplete, double the concentration of DTT in the reaction. If the cysteine content is at or below 6% and reduction is incomplete, increase the temperature for reduction (up to 90 °C). Upon reduction, many proteins will precipitate, illustrating the importance of the chaotrope not only to aid in denaturation but also aid in solubility. In some cases 6-M guanadine or SDS (sodium dodecyl sulfate) may be a good alternative to urea.

2.3 Alkylation

2.3.1 Purpose

Upon exposure to normal atmosphere, a solution of reduced protein will tend to reoxidize slowly. Oxidation can lead to reformation of disulfide linkages (often biologically incorrect linkages) or conversion to cysteic (SO_3) or cysteinesulfinic (SO_2) acid. Alkylation is often performed to quench the reduced cysteines to prevent these oxidation events (see Figure 2 for diagram of reduction and alkylation process). Common alkylation reagents include iodoacetic acid (IAA; CAS 64-69-7), (IAM; CAS 144-48-9), and 4-vinylpyridine (4VP; CAS 100-43-6).[25]

2.3.2 Chemistry of Alkylation

Alkylation is most often done immediately after reduction. An excess of alkylating reagent is required in order to ensure a complete reaction. Usually a molar excess of 10% over the reducing agent is used. Note that 1 mol of alkylating reagent is only equivalent to 0.5 mol of DTT since DTT will reduce a disulfide bond that generates two free cysteine residues. The alkylation reaction is very rapid and usually occurs in minutes; however, note that with the iodo alkylating reagents the reaction should be carried out in the dark or in amber vials in order to reduce the possibility of activating the iodine liberated in the reaction which will then modify tyrosine residues (light can generate an iodine free radical).

Also note that alkylation can be performed prior to reduction of a protein to derivatize any free cysteine residues that may be present in the protein. For example, if a protein contains an odd number of cysteine residues, there will be at least one that is not disulfide linked. The protein may be alkylated first with 4VP, then the remaining cysteine residues may be reduced and alkylated with IAM. In this way, the free cysteine may be identified since it will be the only cysteine alkylated with 4VP.

2.3.3 Reagents

Alkylation with IAA will add a negative charge to each cysteine. This charge may be desirable to avoid aggregation; however, the peptides digested from the protein will be more polar and, therefore, less retained on a reversed-phase column. If the peptide is too polar, it will not be retained. Also, when using IAA for alkylation, it is important to neutralize the reagent prior to use to avoid altering the pH of the reaction. Some vendors sell the sodium salt hence no neutralization is necessary. Alkylation with IAM results in a neutral side chain, less polar than that from IAA, and does not require neutralization prior to derivatization. Both IAM and IAA preferentially react with free cysteine residues, but will react more slowly with primary amines. It is important not to use an overwhelming excess of the reagent or allow the reaction to proceed too long since over-alkylation may result. The alkylation reaction may be quenched by addition of cysteine or thiomalic acid.

4VP has become popular because it will impart a positive charge for each cysteine. The extra charge helps

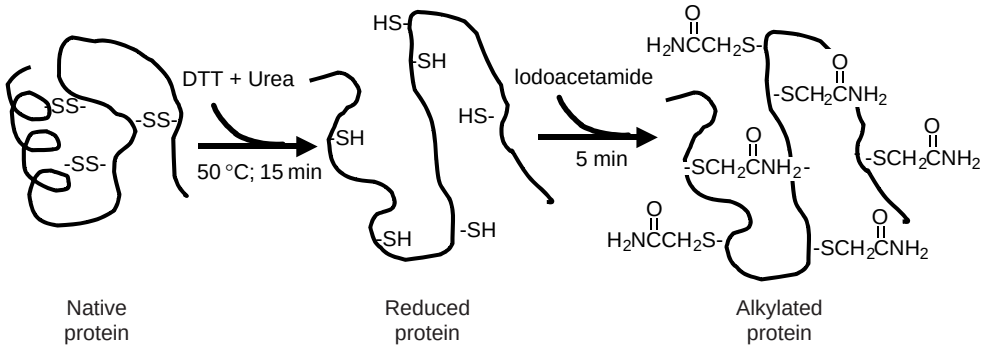

Figure 2 Diagram of the reduction and alkylation of a protein with DTT and iodoacetamide (IAM), respectively.

improve ionization efficiency and therefore the detection of cysteine containing peptides by MS becomes more feasible. Vinylpyridine also has an absorption maxima at 254 nm and can be used for detection of cysteine-containing peptides spectophotometrically. The ratio of the absorbance at 254 nm versus 280 nm is quite useful for this determination.[26]

Acrylamide (CAS 79-06-1) can also be used to alkylate cysteine residues but is generally not used purposely in peptide mapping. Acrylamide adduction of cysteine residues sometimes occurs during polyacrylamide gel electrophoresis and is specifically mentioned since many proteins are purified for analysis on gels.

2.3.4 Typical Protocol

1. Using the reduced protein from the previous section, add 110 µL of 100-mM IAM.
2. Allow to react in the dark for 5 min at room temperature.
3. Dialyze overnight in the dark against 50-mM TRIS, pH 8, containing 10-mM $CaCl_2$ (trypsin digest buffer).

2.3.5 Troubleshooting

The degree of alkylation can be checked by analyzing the protein before and after alkylation by MS. The mass should increase stoichiometrically with the number of expected cysteine residues. If the protein is under-alkylated either allow more time for alkylation or increase the amount of alkylating reagent. If the protein is over-alkylated (lysine residues can be alkylated), reduce the alkylating reagent concentration or time for the reaction. Proteins will often precipitate during alkylation, especially when using IAM that results in a neutral side chain. Precipitation is not always a serious problem since solution-phase enzymatic digestions will still digest precipitates. However, if an immobilized enzyme will be used, it is best to keep the protein soluble by adding urea to a concentration of about 6–8 molar.

3 PROTEOLYTIC DIGESTION

3.1 Chemistry of Digestion

There are two general methods for fragmenting proteins for peptide mapping. Enzymatic digestion is most often employed using endoproteases which specifically recognize certain amino acids. The proteases catalyze the hydrolysis of the peptide bond on either the amino or carboxy side of a specific amino acid. The alternative is a specific chemical cleavage like CNBr (CAS 506-68-3) which will only cleave at methionine residues under the proper conditions.

3.2 Selection of Enzyme or Reagent

Although trypsin is the most common enzyme used in proteolytic mapping, it is not always the best choice. Generally, the sequence of the protein of interest is known and by scanning the sequence for known proteolytic sites, an enzyme can be chosen which will generate appropriately sized fragments (see Table 1 for list of enzymes used for proteolytic mapping). The appropriate size is determined by the technique used to analyze the digest (usually reversed-phase HPLC). The purpose of digesting the protein is to generate smaller fragments so that small subtle changes to the overall protein constitute a larger and more significant change in a peptide fragment (e.g. deamidation leads to a 1-Da increase in mass which would be difficult to detect in an intact protein but much easier to detect in a peptide of just a few amino acids).

In some cases, more than one enzyme may be used either to generate two separate maps or in series to generate smaller fragments.[27] It is also possible to use CNBr in combination with one of the proteases.[28–30]

3.2.1 Trypsin

Since trypsin (EC 3.4.21.4) is the most widely used protease for proteolytic mapping, it will be discussed in the greatest detail. Trypsin is a serine protease which is very closely related to chymotrypsin (EC 3.4.21.1) and elastase (EC 3.4.21.36). Trypsin catalyzes the hydrolysis of peptide bonds on the carboxyl side of lysine and arginine residues, while chymotrypsin cleaves peptides on the carboxyl side of the bulky hydrophobic residues: tryptophan, tyrosine, and phenylalanine. Elastase, which is rarely used for proteolytic mapping, cleaves on the carboxyl side of small amino acids such as glycine and

Table 1 Common proteases used in peptide mapping

Protease	pH optimum	Specificity
Trypsin	pH 8	Carboxyl side of lysine (Lys, K), and arginine (Arg, R)
Chymotrypsin	pH 8	Carboxyl side of phenylalanine (Phe, F), tryptophan (Trp, W) and tyrosine (Tyr, Y)
Lys-C	pH 8	Carboxyl side of lysine (Lys, K)
Asp-N	pH 8	Amino side of aspartate (Asp, D)
Glu-C (V8 protease)	pH 4 or pH 7	Carboxyl side of glutamate (Glu, E) in ammonium acetate at pH 4 and asparate (Asp, D) in phosphate at pH 7.8
Arg-C	pH 8	Carboxyl side of arginine (Arg, R)

alanine. The serine proteases are known as such because they contain a critical serine residue in their catalytic cleft that participates in stabilizing the transition state in hydrolysis. The specificity of the enzymes is determined by the size, shape, and charge in the catalytic cleft. Trypsin has a relatively deep, negatively charged cleft, accounting for its preference for positively charged amino acids, while chymotrypsin has a deep hydrophobic cleft and elastase has a shallow cleft accounting for their specificities. Trypsin has difficulty cleaving peptide bonds at lysine and arginine residues followed by an acidic amino acid such as glutamate or aspartate and does not cleave the bond at all if the next amino acid is a proline.

Cleavage of a protein with trypsin typically forms peptides on the order of 3–10 residues in length. With the exception of the C-terminal peptide, all of the tryptic fragments are guaranteed to contain at least one basic residue making analysis by positive ion MS much easier. However, trypsin is not the best enzyme for every protein. If there are too many cleavage sites (too many lysine and arginine residues) the resulting fragments may be too small and too numerous to generate a good peptide map. If there are too few cleavage sites then only a few large peptide fragments will be generated making detection of modifications more difficult.

Trypsin requires a calcium ion in order to maintain its most active structure; however, calcium is not absolutely required in the digestion buffer. On the other hand, chelators such as ethylene diamine tetraacetic acid (EDTA; CAS 60-00-4) will inhibit digestion because it essentially strips calcium from the enzyme. Trypsin, as well as the other serine proteases are also irreversibly inhibited by phenylmethyl sulfonyl fluoride (PMSF; CAS 329-8-98-6) which covalently attaches to the active site serine. PMSF and EDTA are common reagents used during cell lysis for the purpose of inhibiting proteolytic digestion and it is important that they be removed from the preparation prior to proteolytic mapping. Trypsin has an optimum activity at about pH 8 and its activity increases with temperature up to about 60 °C. However, most digestion protocols call for 37 °C since other side reactions can occur at elevated temperature. Trypsin also remains active in the presence of low concentrations of denaturants and organic solvents. The stability of trypsin can be improved by immobilizing it and by alkylating its lysine residues.

Trypsin is also autocatalytic, digesting itself as well as its substrate proteins. However, in the case of standard solution-phase digests, only small amounts of trypsin are added relative to the substrate protein (typically a ratio of 1 : 50 enzyme : substrate) to prevent the autohydrolytic fragments from trypsin from significantly interfering with the analysis of the protein of interest. Reductive methylation with formaldehyde and sodium borohydride decreases the autohydrolyis.[31] Trypsin preparations often contain chymotryptic activity as well, hence they are treated with *N*-tosyl-L-phenylalanylchloromethyl ketone (TPCK; CAS 402-71-1) which specifically and irreversibly inactivates chymotrypsin. Almost all proteolytic mapping protocols call for TPCK-treated trypsin which is available from many vendors.

Since trypsin is relatively stable and has a neutral to basic pH optimum, it may be stored in 6-mM hydrochloric acid (HCl; CAS 7647-01-0) with minimal loss of activity over several days with refrigeration. Aliquots may also be frozen. HCl is chosen as the storage buffer since it has little buffering capacity. Thus, the HCl will have negligible effect on the pH of the digest solution upon its addition to a protein which is prepared in a buffered solution. Usually, the trypsin will be added at a ratio of 1 : 50 compared to the protein to be digested; hence, its overall volume contribution will also be negligible.

3.2.2 Chymotrypsin

Like trypsin, chymotrypsin (EC 3.4.21.1) is a serine protease and as mentioned earlier cleaves on the carboxyl side of the bulky hydrophobic residues: tryptophan, tyrosine, and phenylalanine. The optimum conditions for activity are pH 8 with activity increasing with temperature to about 60 °C, although most digestions are performed at 37 °C. Preparations of chymotrypsin are often contaminated with trypsin; hence, they are treated with *N*-tosyl-L-lysylchloromethyl ketone (TLCK; CAS 4238-41-9) to irreversibly inactivate the trypsin activity. TLCK-treated chymotrypsin is commercially available for peptide mapping applications. Chymotrypsin is chosen for proteolytic mapping only on rare occasions, but is nonetheless useful in special circumstances.[32] Enzymes such as trypsin, Lys-C (EC 3.4.21.50), Arg-C (EC 3.4.22.8), Asp-N (EC 3.4.24.33), and Glu-C (EC 3.4.21.19) all insure that most of the peptides generated contain a charged amino acid; hence analysis by MS is straightforward. Since chymotrypsin cleaves at bulky hydrophobic residues, no such assurance is given. However, sometimes proteins contain many charged residues, thus cleavage with the other proteases leads to too many small peptides in the proteolytic map, leaving chymotrypsin as a better choice for sequence coverage.

3.2.3 Lys-C

Lys-C (EC 3.4.21.50), from *Lysobacter enzymogenes*, cleaves on the carboxyl side of lysine residues (thus its name). The optimum activity is at pH 8 and digestions are generally done at 37 °C. This protease is generally chosen when the digest fragments that are generated by trypsin are too numerous or too small because it cleaves

at only one amino acid, lysine, instead of two amino acids, lysine and arginine. This enzyme is favored by mass spectrometrists because it guarantees the presence of at least one basic amino acid on every fragment except the C-terminal and generates relatively large fragments (typically twice as large as tryptic fragments).

3.2.4 Asp-N

Asp-N (EC 3.4.24.33), from *Pseudomonas fragi*, cleaves on the amino side of aspartic acid residues thus assuring that each proteolytic fragment, except the *N*-terminal, contains at least one acidic residue. It is not often used for proteolytic mapping but has specificity complementary to the more commonly used proteases. The optimum activity is at pH 8 and digestions are performed at 37 °C.

3.2.5 Glu-C

Glu-C (EC 3.4.21.19), from *Staphylococcus aureus*, also known as V8 protease, cleaves on the carboxyl side of glutamic acid (thus its name Glu-C) in pH 4 ammonium acetate buffer. Glu-C will also cleave on the carboxyl side of aspartic acid residues in pH 7.8 phosphate buffer. Note that under either set of conditions at least one acidic residue is guaranteed on each fragment except the C-terminal fragment.

3.2.6 Arg-C

Clostripain or Arg-C (EC 3.4.22.8), which cleaves on the carboxyl side of arginine residues (hence its name), may be obtained from *Clostridium histolyticum*. Arg-C can also be obtained from the mouse submaxillary gland (EC 3.4.21.40).[33] Similar to Lys-C, the optimum activity is at pH 8 and digestions are generally done at 37 °C. Again, like Lys-C, this protease is generally chosen when the digest fragments generated by trypsin are too numerous or too small since it cleaves at only one amino acid, arginine, instead of two amino acids, arginine and lysine. This enzyme is also favored by mass spectrometrists because it guarantees the presence of at least one basic amino acid on every fragment except the C-terminal and generates relatively large fragments (typically twice as large as trypsin fragments).

3.2.7 Other Proteases Including Pepsin and Papain

Other proteases may be used for proteolytic mapping including pepsin, which has optimal activity under acidic conditions. This feature of pepsin is sometimes used when performing deuterium-exchange experiments.[34] Once the deuterium exchange is complete, the protein is kept cold and at low pH to inhibit re-exchange. Pepsin and papain are also both used in the preparation of antibody fragments. Many of the new therapeutic proteins are either antibodies or based on antibodies. In performing a peptide mapping experiment on an antibody, it is sometimes useful to digest it first into the Fc and either two Fab fragments (pepsin) or one (Fab′)2 fragment (papain) which are separated from one another prior to an additional digestion with another enzyme.

3.2.8 Cyanogen Bromide

CNBr cleaves proteins at methionine residues. Since methionine is a relatively rare amino acid, CNBr tends to generate very large fragments. Caution must be used when working with this reagent since cyanide gas can be liberated. Generally fume hoods are utilized and reactions are performed in sealed vessels. Upon cleavage, the product peptide will be terminated with either a homoserine or homoserine lactone moiety. Especially when analyzing digests by MS, it is important to identify both potential products.

3.3 Typical Protocol for Trypsin

1. Dissolve 1 mg of trypsin in 1 mL of 6-mM HCl (this may be frozen in aliquots for use up to 6 months later).
2. Remove the solution in the previous section which was dialyzed against 50-mM TRIS, pH 8.0 containing 10-mM $CaCl_2$ from the dialysis apparatus and transfer it to a polypropylene test tube. If the protein precipitates, transfer the precipitate also. Trypsin will often digest precipitated protein.
3. Add 20 µL of the trypsin solution.
4. Incubate at 37 °C for 2 h.
5. Add another 20 µL of trypsin and incubate at 37 °C overnight.
6. Quench the reaction the next morning by adding 50 µL of trifluoroacetic acid (TFA; CAS 76-05-1).

3.4 Troubleshooting

The best way to check for completeness of a tryptic digest is by MS. If the digest is not complete, consider adding more enzyme (1 : 25) and/or allowing more time for digestion (up to 24 h). If MS reveals the presence of peptides with inappropriate cleavage sites (i.e. cleavage at chymotryptic sites) consider using less enzyme, decreasing digestion time, or finding a new source for the enzyme. Verify that the trypsin was treated with TPCK. Occasionally, peptides will precipitate upon cleavage from the intact protein. Often these peptides can be resolubilized with the addition of up to 5% TFA. Avoid the addition of organic solvents if possible, since

the next step is often reversed-phase chromatography in which organic solvent is used to elute the peptides from the column.

4 MAP DEVELOPMENT

Although there are a number of ways to separate and identify the peptides generated in a digest, the most common method for developing the peptide map after a digestion is reversed-phase chromatography with UV (ultraviolet) detection at around 210–230 nm. Older techniques included 2D thin-layer chromatography followed by detection with ninhydrin (CAS 485-47-2). Of course, the digest may also be analyzed directly by MS without any chromatographic separation (e.g. this is often the case with proteins digested from SDS/PAGE gel slices). A diagram for a typical chromatographic system for peptide mapping is shown in Figure 3.

4.1 Column Selection

The most common reversed-phase column for peptide mapping is the silica-based C_{18} column. The notation "C_{18}" refers to the length of the carbon chain covalently attached to the surface of the silica. Other reversed-phase columns of this classification include C_4 and C_8 columns which are less hydrophobic than the C_{18}, and therefore allow elution of hydrophobic peptides with much less organic solvent. There are also polystyrene-based reversed-phase columns that have a hydrophobicity similar to a C_8 column, but they are difficult to compare directly with the silica-based columns because the mechanism of adsorption is slightly different, and therefore the elution order may also be different. Although resolution is sometimes poorer on a polystyrene column compared with the C_{18} columns, the polystyrene columns may be operated under basic conditions (the silica-based support will start to dissolve in alkaline media), thus allowing more flexibility in method development.

The appropriate column for developing the map depends upon the size, charge, hydrophobicity, and number of peptides generated. Typically, the C_{18} column is the best first choice. However, if it is known that the peptides generated from the digest are large and hydrophobic (e.g. CNBr digest), it may be better to use a less hydrophobic column such as the C_4 or C_8 columns. If for some reason alkaline reversed-phase conditions yield a better separation of the peptides than the typical acidic conditions, then a polystyrene column may be a better choice because it will withstand alkaline conditions much better than the silica-based columns. Unfortunately, there is no good way to select the best column a priori, so a number of columns and mobile phase conditions must be evaluated to determine which are appropriate for a particular digest.

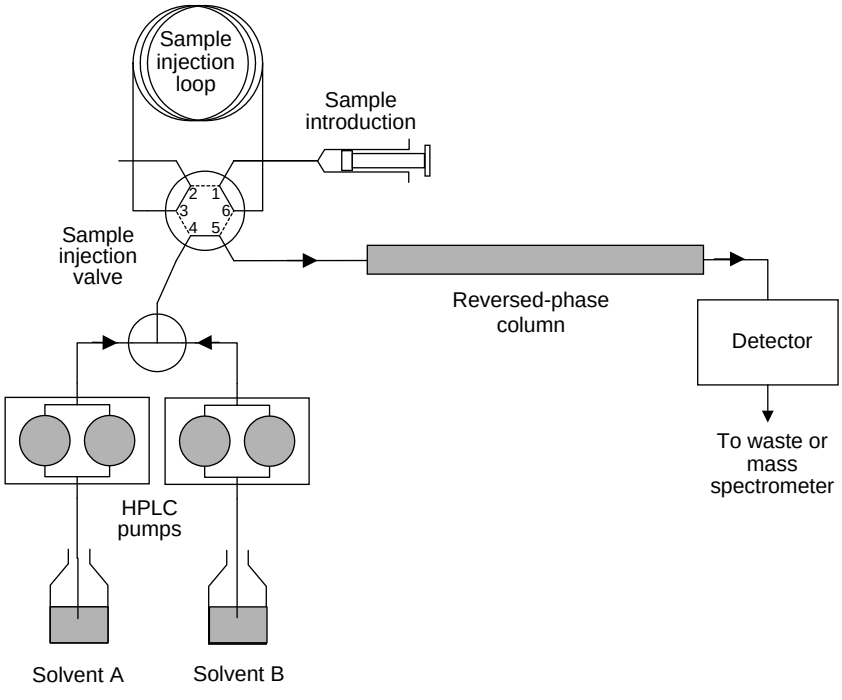

Figure 3 Diagram of a chromatographic system for developing peptide maps.

4.2 Mobile Phase Selection

4.2.1 Solvents

Reversed-phase HPLC peptide maps are developed with a gradient from aqueous to organic solvent. Typical organic solvents for peptide mapping include acetonitrile (CAS 75-05-8), methanol (CAS 67-56-1), propanol (CAS 71-23-8), isopropanol (CAS 67-63-0), and ethanol (CAS 64-17-5), with the most popular reversed-phase solvent for peptide mapping being acetonitrile. Again, there is no good way of predicting the best solvent, but method developers usually begin with acetonitrile and only move to the other solvents if acetonitrile does not provide acceptable results. In addition to resolution, other factors to consider in the selection of a solvent include viscosity (higher viscosity will increase back pressure), UV cutoff (detector compatibility in general), disposal, and safety.

4.2.2 Mobile Phase Modifiers

The mobile phase modifier (e.g. TFA, HFBA [heptafluorobutyric acid; CAS 375-22-4], HCl, ammonium bicarbonate [CAS 1066-33-7]) usually serves more than one purpose. Not all of the surface silanol groups are covalently modified with a carbon chain in silica-based columns. Under neutral to basic conditions these silanol groups will be deprotonated leaving a negative charge on the surface which will interact with positive charges on peptides in the sample. Since these interactions will diminish resolution, they are suppressed by adding acids to the mobile phase to insure protonation and thus neutralization of the silanol groups.

Some of the modifiers are also known as ion-pairing agents. For example, TFA will interact with the positively charged residues on a peptide. Because TFA is hydrophobic, peptides that ion pair with it will be more retained on a reversed-phase column. Likewise, HFBA is even more hydrophobic than TFA and will cause these same peptides to be even more highly retained. A cationic ion-pairing agent, like *tert*-butylamine (CAS 75-64-9), will pair with negatively charged residues on the peptide thus causing the acidic peptides to be more highly retained. However, since this agent is used under basic conditions, it may be wise to consider the use of a polymeric column.

4.3 Gradient Optimization

Finding the right gradient for the best resolution in a minimum analysis time is one of the most challenging tasks in developing a proteolytic mapping protocol. In general, if there are large gaps in the peptide map, one tries to increase the slope of the gradient in these areas to save analysis time. In regions of overlapping peaks, the slope of the gradient is decreased to provide better separation even to the extent that an isocratic (no gradient) step is used. If the analysis will only be run occasionally, it may not be worth the investment in effort to reduce the analysis time. In fact, there are many cases in which a one- or two-hour gradient is used as a matter of routine with no further optimization.

4.4 Reproducibility

Reproducibility is measured in both retention time and peak area. Hoff and Pickering propose that the reproducibility of retention times of peaks in a peptide map have a standard deviation of less than 0.05 min in order to generate useful data.[35] Many, but not all, HPLC systems are capable of delivering this performance. The performance capability of the HPLC system is also dependent on the column used to develop the map. Smaller columns (2.1 and 1.0 mm diameter) are more challenging since they operate at lower flow rates. Map reproducibility can also be greatly influenced by experimental technique. It is important to prepare the solvents precisely, and that they are degassed and filtered. Column temperature is another critical factor. Resolution can be improved with elevated temperature, but to attain reproducibility, it is important to keep a constant temperature from analysis to analysis. Typically, a column oven is used to maintain the temperature at 37 °C. Of course, sample injection must also be precise and accurate, and the injection valve should have less than 0.1% carryover between injections. Sample handling is another important consideration and is somewhat protein dependent. It is generally a good idea to refrigerate a stored digest taking care to avoid oxidation, evaporation (change concentration), or contamination.

4.5 Typical Protocol

1. Select a 4.6-mm-internal diameter × 250-mm-length C_{18} reversed-phase column.
2. Prepare 0.1% TFA in water for solvent A and 0.08% TFA in acetonitrile as solvent B filtering through a 0.2-μm solvent compatible filter (like nylon).
3. Set the detector wavelength to 214 nm.
4. Set up the column and solvents on an HPLC and program a method that will equilibrate the column for 15 min at 1 mL min^{-1} with 100% solvent A, then inject 50 μL of a 1 mg mL^{-1} digest, and run a linear gradient to 30% solvent A + 70% solvent B in 60 min. At the end of the gradient, wash the column for 10 min in 100% solvent B. The entire method should be run at 1 mL min^{-1}.

5 IDENTIFICATION OF PEPTIDE PEAKS

Although the peaks in the map can be used as a fingerprint to identify a protein compared to a control, it is often desirable or necessary to determine the identity of the peptides in each of the peaks. This is especially true when a new unexpected peak is found. There are a number of ways to identify the peaks, but currently the most popular and most definitive techniques are peptide sequence analysis and MS.

5.1 Retention Time Prediction

Retention times of peptides can be estimated based on the sequence. Although these models are very good, they are not perfect and reality often deviates from the theoretical. If two peptides are very similar (e.g. same sequence except for one residue) prediction is very successful, but in the case of proteolytic mapping it is best to use these models only as a guide.

Another route to retention time prediction is to generate all of the peptides expected in a map synthetically and then analyze them individually to determine their retention time. Although this may seem like the most difficult way to accomplish the task, it can have some advantages. It can be especially useful to find the retention time of expected contaminants or breakdown products. For example, if one is concerned whether an active site cysteine is oxidized, a fragment may be synthesized with the cysteine oxidized and analyze it on the column. When a complete digest is analyzed and that peak does not appear one can say that cysteine oxidation at that site was not observed. Essentially, this technique may be used to guide the researcher in his interpretation of the chromatogram for potential problems.

5.2 Ultraviolet Spectral Analysis

Although most peptides have a similar UV spectra, they are not identical (they especially depend upon the aromatic amino acid content). Further, derivatization (e.g. alkylation with vinylpyridine) may create unique spectral signatures which can be identified. For example, if vinylpyridine is used for alkylation, cysteine-containing peptides may be identified by determining the ratio of the absorbance at 254 nm and 280 nm. However, not all peptides can be uniquely identified this way. The purity of peaks in the peptide map may also be checked by a UV spectral analysis. If a peak contains only one pure peptide, the spectrum at the beginning, middle and end of the peak should remain constant, except for intensity. Any change in the spectrum throughout the peak will indicate that a second peptide or impurity is present.

5.3 Amino Acid Analysis

Amino acid analysis involves the hydrolysis of the peptide fragment into its component amino acids. This gives a composition of the peptide, not the sequence. Unfortunately, asparagine and glutamine are converted to asparate and glutamate, respectively, and tryptophan is degraded. However, it is still often possible to deduce the sequence from the composition when the sequence of the intact protein is known. Also, amino acid analysis is a quantitative determination giving the amount of peptide present. It is also possible to find modified amino acids with this technique. For example, the amino terminal residue can be located by specifically labeling it with fluorodinitrobenzene.

5.4 Peptide Sequence Analysis

To determine the sequence of a peptide or protein, *N*-terminal sequence analysis by Edman degradation has been the dominant technique for many years. However, there are now viable competing techniques which include C-terminal sequence analysis and MS.

5.4.1 N-*terminal*

Edman degradation involves the systematic removal of one amino acid at a time from the amino terminus of the peptide or protein. First, phenyl isothiocyanate (CAS 103-72-0) is used to derivatize the neutral amino terminal of the peptide under alkaline conditions. Then, by changing the buffer conditions to a slightly acidic environment, a cyclic derivative is formed with the terminal amino acid as it cleaves from the peptide which is then ready for another cleavage cycle. Since the peptide or protein is immobilized on a resin or membrane, reagent addition, buffer change, and analysis of the liberated amino acid by HPLC are easily automated. Unfortunately, this method of sequence analysis is not useful if the amino terminal end of the protein or peptide is blocked.

5.4.2 C-terminal

The C-terminal sequence analysis process proceeds much the same way as *N*-terminal sequence analysis except by different chemistry so that a single amino acid is removed from the carboxyl terminus. The peptide or protein is first attached to a Zitex™ membrane, then coupled with diphenyl phosphoroisothiocyanatidate (DPPITC). The coupled, membrane-bound peptide is washed with acetonitrile to remove excess DPPITC and then treated with pyridine to effect ring closure, anhydrous TFA, and finally trimethylsilanolate to cleave a single thiohydantoin-amino acid which is identified by retention on a reversed-phase column. The shortened

peptide is then cycled through the process again to determine the next amino acid.[36]

5.5 Mass Spectrometry

MS is rapidly becoming the dominant technique for protein and peptide characterization. This is partly because of a decrease in price for performance (i.e. lower-priced MS systems are improving in sensitivity and resolution). There are a variety of mass spectrometers available, but they are generally classified by the type of ionization employed (e.g. electrospray ionization (ESI) and MALDI (matrix-assisted laser desorption ionization) are most popular for proteins and peptides) and the type of mass analyzer used (e.g. time-of-flight [TOF], quadrupole [Q], or ion trap). Because most ionization sources can be fitted to almost any mass analyzer, many combinations are possible. Further, mass analyzers may be coupled in tandem mass spectrometry (known as MS/MS) first to select an ion based upon its intact mass, fragment the ion, and then analyze its fragments on a second mass analyzer. Although many MS/MS systems use the same type of mass analyzer for both steps, it is not a requirement as is the case for ESI/Q/TOF. Some combinations are more logical than others, hence, not every conceivable product is available. As technology improves, the combinations which appear to be logical are subject to change.

There are also a number of options available for coupling a reversed-phase column used to develop the map of a mass spectrometer. The most direct way is though an electrospray source which will ionize the peptides in the flow stream as they elute from the column and desolvate them prior to introduction into the high vacuum of the mass analyzer. However, fractions may also be collected from the reversed-phase column and then injected into an electrospray source or applied to a MALDI target with the matrix. Application of fractions to a MALDI target have been automated using either robots or a specialized fraction collector which will apply the fractions directly to the plate as they elute. In this case, matrix is applied in a second pass.

5.5.1 Peptide Intact Mass

The simple molecular mass of a peptide from a protein of a known sequence digested with a known enzyme is often more than sufficient to assign the identity of a peptide unambiguously. Of course, the probability that this is true improves as the accuracy of the measurement improves. Also, given several peptide masses from a protein digested with a specific enzyme, even an unknown protein may be uniquely identified from a database given sufficient mass accuracy.[37]

5.5.2 Sequence Analysis with Enzymes (Carboxyl and Amino Peptidases)

Peptides can be sequenced by MS utilizing either amino or carboxypeptidases to digest from one end of the peptide, one amino acid at a time. The objective is to generate a series of peptides which differ by only one amino acid in length, in effect generating a ''ladder'' from which the sequence may be read by determining the difference in mass between each peak in the ladder. The difficulty with this technique is that the enzymes used for this purpose often do not cleave all amino acids at equal rates, hence, there may be gaps or stops in the ladder sequence. Unfortunately, this technique cannot distinguish between the isomers leucine and isoleucine (lysine and glutamine can also be difficult to distinguish because they differ by only 0.04 Da).

5.5.3 Sequence Analysis by Mass Spectrometry

When peptides are ionized in a gas phase, they can be fragmented in a number of ways. In fact, if the ionization method is harsh they may spontaneously fragment. Peptides can fragment at three points along the peptide bond: between the carbon and nitrogen that make the peptide bond, between the nitrogen and α-carbon, or between the α-carbon and carbonyl. Also upon fragmentation of a singly charged peptide, only one fragment or the other receives the charge. The neutral species will not be detected. With three potential fragmentation points multiplied by the two possible directions in which the charge may transfer, there are six possible fragments that can be generated known as a, b, c, x, y, and z (see Figure 4 for fragmentation examples). Unfortunately, this is only the simplest case. It is also possible to fragment within the amino acid side chains, and it is possible to generate internal fragments (i.e. two bonds broken and the charge carried by the middle fragment). Often, a great deal of expertise is required to interpret an MS/MS spectrum.

5.5.3.1 Tandem Mass Spectrometry The most common way to cause fragmentation once the peptide is in the mass spectrometer is by introduction of a collision gas (collision-induced dissociation [CID]). In many MS/MS experiments, an initial mass analyzer is tuned to the mass of interest allowing only that ion to continue to the area where fragmentation occurs. Then, the fragments which retain the charge are separated on the second mass analyzer. A good example of this technique is demonstrated with a triple Q instrument in which one Q is used for the initial mass separation, a second for fragmentation, and a third to analyze the fragments. The major advantage of MS/MS is that relatively complex mixtures may be infused, and a single species may be selected for further

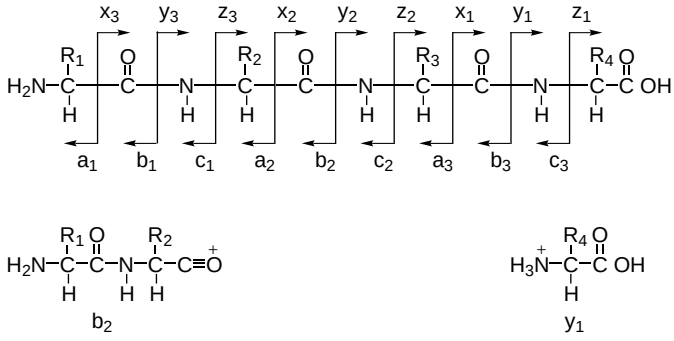

Figure 4 Diagram of peptide fragmentation which leads to a, b, c, x, y, and z ions. The lines indicate the bond along which the peptide backbone is cleaved, and the arrows show the direction of the charge transfer. The groups R_1-R_4 represent the various amino acid side chains. An example of a b_2 and a y_1 ion are also shown for this theoretical peptide.

fragmentation and MS analysis. This is not the case with in-source fragmentation and post-source decay where relatively pure peptides are required in order to interpret the fragmentation patterns.

5.5.3.2 In-source Fragmentation Fragmentation of peptides can also be achieved in the source of a mass spectrometer. In this experiment, only one mass analyzer is used to analyze the fragments, but the major limitation is that only one component can be infused into the instrument at a time or else data interpretation becomes extremely complicated with fragments from multiple peptides.

5.5.3.3 Post-source Decay Post-source decay applies peptides that fragment after initial acceleration in a MALDI TOF MS. Fragmentation can also occur prior to acceleration but the ionized fragments from such a decomposition would be found in the original spectrum. Post-source decay allows for detection of fragments generated post acceleration due to fission of bonds along the peptide backbone. To separate and detect these fragments in a TOF instrument, an ion mirror (reflector) that will cause the ionized fragments to reverse course must be used and the neutral fragments must be allowed to pass through undetected. The reflected peptide fragments are then detected using stepped-mirror voltages to reflect lower-mass fragments.

6 AUTOMATION OF PROTEOLYTIC MAPPING

The entire proteolytic mapping process including reduction and alkylation has been automated.[14] A number of authors have shown various ways of accomplishing this depending on the goals of the mapping process.[15,16,18] Central to these methods is an immobilized proteolytic enzyme that rapidly digests the substrate protein. For example, when trypsin is immobilized, its autolytic capability is greatly reduced due to its immobility. Therefore, high concentrations of trypsin can be used to speed the digestion considerably (normally the slowest step of proteolytic mapping, taking several hours to overnight to complete), so that it only takes a few minutes. Immobilization also improves the stability and lifetime of the enzyme so it can be used repeatedly.

Using column switching valves, a column packed with trypsin can be used first to digest an intact protein and pass the fragments on to an analytical column. Then, the enzyme column can be switched off-line while the analytical column is resolved (see Figure 5 for an example plumbing diagram). Further, the reduction and alkylation of the protein can be performed in some autosamplers allowing these modifications to be completed while the reversed-phase column equilibrates with the starting solvents. Given this type of automated system, it takes no longer to analyze an intact protein than one which had been prepared manually using an overnight digestion. Automation can make method development easier as well as lead to improved reproducibility, especially between operators.

6.1 Automating Reduction and Alkylation

Many autosamplers and robotic workstations are capable of automating the process of reduction and alkylation of a protein.[14] To perform this process, the machine must be capable of adding reagents, heating the sample, and mixing the sample. Automation of these steps has a number of advantages including the fact that the protein is freshly prepared and less likely to be oxidized, carbamylated, or otherwised modified by storage for long periods of time in the alkylation reagents. Automation generally leads to better reproducibility, especially between operators, and finally automation allows around-the-clock operation.

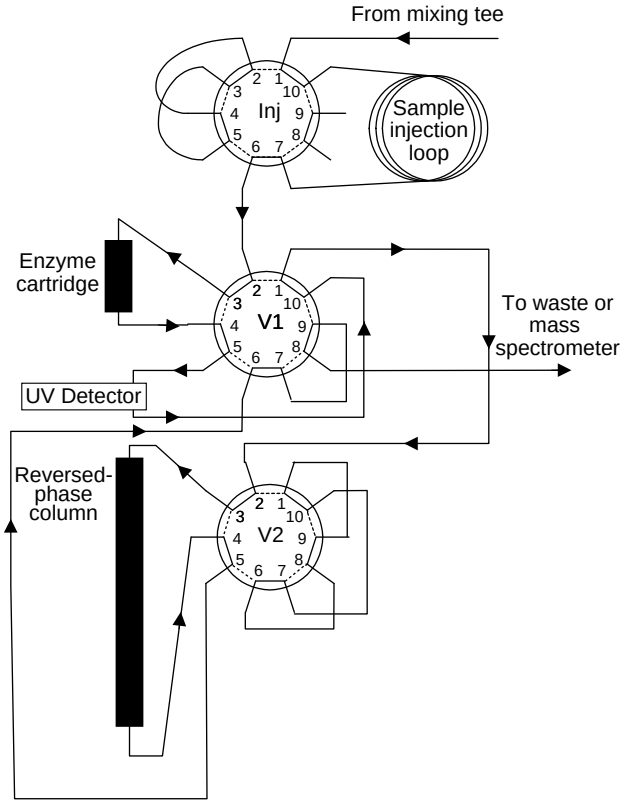

Figure 5 Example plumbing diagram for automated peptide mapping with an immobilized enzyme column. One valve is used for injection of sample (denoted Inj), another valve is used to engage and disengage the enzyme cartridge (denoted V1) and the third valve is used to engage and disengage the reversed-phase analytical column (denoted V2).

6.2 On-line Digestion with Immobilized Enzymes

The use of immobilized enzymes for on-column digestion has long been reported. However, many problems with early attempts were due to the instability of the silica support at the basic pH used for tryptic digestions. Trypsin autohydrolytic fragments were the major problem encountered until a hydrophilic-coated polystyrene-based resin was used for enzyme immobilization. The base polystyrene is mechanically rigid under typical HPLC pressures and chemically inert over a wide pH range. The hydrophilic coating minimizes hydrophobic interactions with either the protein or its digest fragments. Immobilized enzymes are generally more stable than those in free solution. For example, immobilized trypsin can be used in either organic solvents (up to 50% acetonitrile) or at high temperature (up to 60 °C).

Immobilized enzyme columns are plumbed in tandem with an analytical reversed-phase column.[14–18,37,38] Since the enzymes do not tolerate the harsh organic solvents used in reversed-phase, switching valves are used to take each of the columns in and out of line at the appropriate times. Also, the HPLC pumps must be capable of delivering at least three different solvents: digestion buffer, aqueous reversed-phase solvent, and organic reversed-phase solvent. A typical method includes flushing the system with reversed-phase solvents with both columns off-line followed by engaging the reversed-phase column and equilibrating it with the aqueous starting solvent. Then, the columns are switched out of the flow path again to allow the system to be flushed with digestion buffer after which the enzyme column is cleaned and equilibrated with digestion buffer. Then, the flow rate is slowed (about 10–50 $\mu L\,min^{-1}$), both columns are placed in-line, and a protein sample is injected. First, the sample slowly passes through the enzyme column, it is digested, and then it passes through the reversed-phase column which captures the peptide fragments. Once digestion and capture are complete, both columns are switched off-line to be flushed with reversed-phase solvents. Then, the reversed-phase column is engaged and eluted with a gradient of organic solvent as in standard proteolytic mapping (see Figure 6 for an example of reproducibility of an automated system).

6.3 Liquid Chromatography/Mass Spectrometry to Automate Peak Identification

As with a standard solution-phase digestion peptide mapping protocol, the automated protocol is easily coupled to MS. Coupling to electrospray is straightforward,[18] but automated methods are also available to couple to MALDI.

6.3.1 Electrospray Ionization

Coupling to electrospray is quite simple. The outlet of the reversed-phase column is coupled directly to the mass spectrometer. However, there are some tips that will improve performance and sensitivity. First, the flow rate from the column may not be optimal for sensitivity on the mass spectrometer and a flow splitter may be needed. Also, the initial flow-through peak from the injection of the peptides often contains salts, buffers, and urea. It is best not to pass these substances through the spray tip of the mass spectrometer, even if the shutter is closed. It is better to use a diverter valve to send this fraction to waste. Finally, the ion-pairing agent used for optimum resolution on the reversed-phase column may not be optimal for sensitivity in the ionization source. For example, TFA sometimes interferes with sensitivity while ion-pairing agents like formic or acetic acid are more well behaved. Either a switch may take place to a more favorable ion-pairing agent or a make-up solvent may be added to the flow stream (also known as a TFA

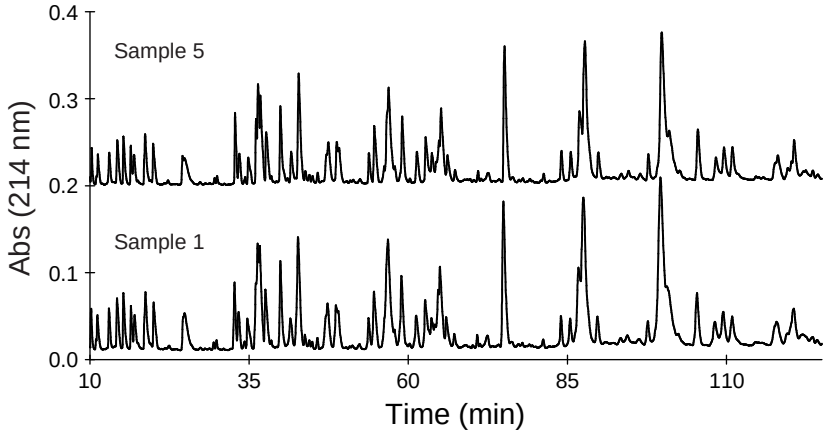

Figure 6 Example of the reproducibility of automated peptide mapping in each case starting with a fresh sample of native human serum albumin which is automatically reduced and alkylated, digested, and then analyzed by reversed-phase HPLC.

"fix") to minimize the interference from the ion-pairing agent.

Direct coupling of the column to the mass spectrometer is perhaps most desirable since the mass data is collected in real time; but, this is not required. Fractions may be collected and then infused at a later time. This may be desirable if concentration of the fraction is necessary or if additional chemistry is to be performed prior to mass spectral analysis.

6.3.2 Matrix-assisted Laser Desorption Ionization

Although MALDI can be coupled directly to reversed-phase chromatography,[39] it is often desirable simply to deposit fractions on to the MALDI target. This may be done by collecting fractions and manually or robotically redispensing fractions from the original tubes on to the MALDI target. Alternatively, the fractions may be applied to the target as they elute from the reversed-phase column using a modified fraction collector and flow splitter. A matrix must also be added to each sample to aid in desorption and ionization.

6.4 Database Searching to Identify Proteins

A reversed-phase peptide map provides a fingerprint of a protein that can be matched against the maps of other proteins in an attempt to identify it. The map is highly technique-, column-, and instrument-dependent so although the identity of a known protein can be confirmed, it is difficult to identify an unknown protein. However, if the masses of these proteolytic fragments are known, the observed masses can be compared with the expected theoretical masses generated from a digest of all proteins in a sequence database. It is not necessary to digest and analyze the database proteins, but simply to have a correct sequence from which a theoretical digest is generated. In addition to searching databases with a set of peptide masses known to originate from a single protein, databases may also be searched for specific peptide sequences. For example, if a peptide fragment from a protein is sequenced by Edman degradation, the database can be searched with that sequence to find the protein from which it came. Also, since MS/MS fragmentation contains sequence information, MS/MS data can be submitted for a database search to find peptides that would theoretically produce a similar fragmentation pattern.

7 SELECTED APPLICATIONS

There are a variety of applications for proteolytic mapping and the use of the data generated depends upon the application. For example, mapping is performed in the QC environment to confirm the identity of a protein by simply using its "fingerprint" from the HPLC chromatogram. All peaks are important, and any loss of peaks or any new peaks indicate problems. In contrast, in typical proteomics applications, one only wishes to identify a protein based on the masses of the observed fragments. Not all of the masses are needed to identify the protein. In addition, any extraneous masses may come from a variety of sources, may not indicate a problem, and can often be ignored.

7.1 Mapping in the Quality Control Environment

The most important feature of a proteolytic mapping procedure in the QC environment is robustness. The method must give consistent reproducible results over time and between operators. This has been one of the driving forces for automating the proteolytic mapping

process. Typically, 4.6-mm diameter columns are used for reversed-phase analysis since most HPLC pumping systems are able to deliver very reproducible gradients using a column of this size (typically operated at $1\,mL\,min^{-1}$). Owing to their large volume, these columns are not the most sensitive. However, because the sample is usually not limiting in the QC environment, the sample is sacrificed for robustness.

Considerable effort is placed on developing a method specific to the protein to be analyzed. To diminish variability, complete digestion is usually required with no partially digested fragments. Great care is taken to avoid additional modifications such as oxidation during the process which may obscure modifications incurred prior to proteolytic mapping. After all of the peaks in the HPLC chromatogram have been identified once (usually by sequence analysis and/or MS) they are not re-identified in the final routine assay. Chromatographic retention time is considered to be sufficient identification henceforward.

7.2 Miniaturization for Increased Sensitivity in Research and Development

The research and development (R&D) environment is quite different from the QC environment in that large amounts of sample are often difficult to obtain. For example, one may wish to identify contaminants in a product that are present at the 1% level. Since the sample identity is unknown, the mapping procedure will not be optimized for the sample. Consequently, a generic method employed for most proteins will be used. Further, complete information about the sample may not be required, and a simple identification may be sufficient. If only identification is needed, the chromatographic separation may be circumvented and the digest may be analyzed directly by MS since only a few fragment masses will be required to identify uniquely the protein.

However, if further characterization is required (i.e. the contaminant may be a product that is oxidized and verification is needed), chromatographic separation may be necessary prior to MS in order to allow each of the peptides to ionize more efficiently. When analyzing a mixture by MS, quenching of some analytes by others is often observed (analytes which ionize easily obscure those that ionize less readily). A chromatographic separation allows the peptides to be presented individually to the mass spectrometer to avoid this competition. Again, to attain the highest level of sensitivity, digestions are usually performed in solution in small volumes using microbore or capillary HPLC columns to develop the map. The robustness of a larger column is sacrificed for the sensitivity of a capillary column.

7.3 Application to Proteomics

Generally, in a proteomic application, the goal of proteolytic mapping is to identify the protein based upon its mass fingerprint, hence a reversed-phase separation of the peptide fragments is rarely used. Instead, the digest is analyzed directly by MS. Also, unlike mapping in the QC and R&D environment, relatively large amounts of trypsin (or other proteolytic enzyme) may be used. The enzyme autohydrolysis peptides which interfere in the other analyses may be used as internal calibration standards in proteomic applications. The peptide masses are then submitted to a database searching program that looks for proteins that would give the same masses if digested with the same enzyme as used in the experiment. Results of the search are improved by using more peptides and by having good mass accuracy.

ABBREVIATIONS AND ACRONYMS

CID	Collision-induced Dissociation
CNBr	Cyanogen Bromide
DPPITC	Diphenyl Phosphoroisothiocyanatidate
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetraacetic Acid
ESI	Electrospray Ionization
HCl	Hydrochloric Acid
HFBA	Heptafluorobutyric Acid
HPLC	High-performance Liquid Chromatography
IAA	Iodoacetic Acid
IAM	Iodoacetamide
MALDI	Matrix-assisted Laser Desorption Ionization
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PMSF	Phenylmethyl Sulfonyl Fluoride
PTM	Post-translational Modification
Q	Quadrupole
QC	Quality Control
R&D	Research and Development
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic Acid
TLCK	*N*-Tosyl-L-lysylchloromethyl Ketone
TOF	Time-of-flight
TPCK	*N*-Tosyl-L-phenylalanylchloromethyl Ketone
TRIS	Tris(hydroxymethyl)aminomethane

UV	Ultraviolet
2D	Two-dimensional
4VP	4-Vinylpyridine

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Posttranslational Oxidative Modifications of Proteins • Protein Purification: Theoretical and Methodological Considerations • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Gradient Elution Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. E.R. Hoff, R.C. Chloupek, 'Analytical Peptide Mapping of Recombinant DNA-derived Proteins by Reversed-phase High-performance Liquid Chromatography', *Methods Enzymol.*, **271**, 51–68 (1996).
2. A. Apffel, J. Chakel, S. Udiavar, S. Swedberg, W.S. Hancock, C. Souders, E. Pungor, Jr, 'Application of New Analytical Technology to the Production of a 'Well-characterized Biological'', *Dev. Biol. Stand.*, **96**, 11–25 (1998).
3. J.R. Yates III, 'Mass Spectrometry and the Age of the Proteome', *J. Mass Spectrom.*, **33**(1), 1–19 (1998).
4. R.J. Harris, A.A. Murnane, S.L. Utter, K.L. Wagner, E.T. Cox, G.D. Polastri, J.C. Helder, M.B. Sliwkowski, 'Assessing Genetic Heterogeneity in Production Cell Lines: Detection by Peptide Mapping of a Low Level Tyr to Gln Sequence Variant in a Recombinant Antibody', *Biotechnology*, **11**, 1293–1297 (1993).
5. R.L. Garnick, 'Peptide Mapping for Detecting Variants in Protein Products', *Dev. Biol. Stand.*, **76**, 117–130 (1992).
6. C.L. Clogston, Y.R. Hsu, T.C. Boone, H.S. Lu, 'Detection and Quantitation of Recombinant Granulocyte Colony-stimulating Factor Charge Isoforms: Comparative Analysis by Cationic-exchange Chromatography, Isoelectric Focusing Gel Electrophoresis, and Peptide Mapping', *Anal. Biochem.*, **202**, 375–383 (1992).
7. A.M. Crestfield, S. Moore, W.H. Stien, 'The Preparation and Enzymatic Hydrolysis of Reduced and *S*-carboxymethylated Proteins', *J. Biol. Chem.*, **238**, 622–627 (1963).
8. W.W. Cleland, 'Dithiothreitol, a New Protective Reagent for SH Groups', *Biochemistry*, **3**, 480–482 (1964).
9. S. Borman, 'Analytical Biotechnology of Recombinant Products', *Anal. Chem.*, **59**, 969A–973A (1987).
10. R.L. Garnick, N.J. Solli, P.A. Papa, 'The Role of Quality Control in Biotechnology: an Analytical Perspective', *Anal. Chem.*, **60**, 2546–2557 (1988).
11. N. Lundell, T. Schreitmuller, 'Sample Preparation for Peptide Mapping – A Pharmaceutical Quality-control Perspective', *Anal. Biochem.*, **266**(1), 31–47 (1999).
12. G.E. Swaisgood, G.L. Catignani, 'Use of Immobilized Proteinases and Peptidases to Study Structural Changes in Proteins', *Methods Enzymol.*, **135**, 596–604 (1987).
13. A.W. Burgess, L.I. Weinstein, D. Gabel, H.A. Scheraga, 'Immobilized Carboxypeptidase A as a Probe for Studying the Thermally Induced Unfolding of Bovine Pancreatic Ribonuclease,' *Biochemistry*, **14**(2), 197–200 (1975).
14. T. Nadler, C. Blackburn, J. Mark, N. Gordon, F.E. Regnier, G. Vella, 'Automated Proteolytic Mapping of Proteins, *J. Chromatogr., A*, **743**(1), 91–98 (1996).
15. J. Lippincott, E. Hess, I. Apostol, 'Mapping of Recombinant Hemoglobin Using Immobilized Trypsin Cartridges,' *Anal. Biochem.*, **252**(2), 314–325 (1997).
16. Y.L. Hsieh, H. Wang, C. Elicone, J. Mark, S.A. Martin, F. Regnier, 'Automated Analytical System for the Examination of Protein Primary Structure,' *Anal. Chem.*, **68**(3), 455–462 (1996).
17. B.J. Harmon, X. Gu, D.I. Wang, 'Rapid Monitoring of Site-specific Glycosylation Microheterogeneity of Recombinant Human Interferon-gamma,' *Anal. Chem*, **68**(9), 1465–1473 (1996).
18. D.B. Kassel, T.G. Consler, M. Shalaby, P. Sekhri, N. Gordon, T. Nadler, 'Direct Coupling of an Automated

2-dimensional Microcolumn Affinity Chromatography–Capillary HPLC System with Mass Spectrometry for Biomolecule Analysis', in *Techniques in Protein Chemistry VI*, ed. J.W. Crabb, Academic Press, San Diego, CA, 39–46, 1995.

19. J. Gobom, E. Nordhoff, R. Ekman, P. Roepstorff, 'Rapid Microscale Proteolysis of Proteins for MALDI-MS Peptide Mapping Using Immobilized Trypsin', *Int. J. Mass Spectrom.*, Ion Processes, **169/170**, 153–163 (1997).
20. K. Stoney, K. Nugent, 'Online Preparation of Complex Biological Samples Prior to Analysis by HPLC, LC/MS and/or Protein Sequencing', in *Techniques in Protein Chemistry VI*, ed. J.W. Crabb, Academic Press, San Diego, CA, 277–284, 1995.
21. J.A. Burns, J.C. Butler, J. Moran, G.M. Whitesides, 'Selective Reduction of Disulfides by Tris(2-carboxyethyl) phosphine', *J. Org. Chem.*, **56**, 2648–2650 (1991).
22. W.R. Gray, 'Disulfide Structures of Highly Bridged Peptides: a New Strategy for Analysis', *Protein Sci.*, **2**(10), 1732–1748 (1993).
23. C.E. White, M.J. Hunter, D.P. Meininger, S. Garrod, E.A. Komives, 'The Fifth Epidermal Growth Factor-like Domain of Thrombomodulin does not have an Epidermal Growth Factor-like Disulfide Bonding Pattern', *Proc. Natl. Acad. Sci. USA*, **93**(19), 10 177–10 182 (1996).
24. J. Wu, J.T. Watson, 'A Novel Methodology for Assignment of Disulfide Bond Pairings in Proteins,' *Protein Sci.*, **6**(2), 391–398 (1997).
25. S. Renlund, I.M. Klintrot, M. Nunn, J.L. Schrimsher, C. Wernstedt, U. Hellman, 'Peptide Mapping on HIV Polypeptides Expressed in *Escherichia coli*. Quality Control of Different Batches and Identification of Tryptic Fragments Containing Residues of Aromatic Amino Acids or Cysteine,' *J. Chromatogr.*, **512**, 325–335 (1990).
26. J.J. Holthuis, R.J. Driebergen, 'Chromatographic Techniques for the Characterization of Proteins,' *Pharm. Biotechnol.*, **7**, 243–299 (1995).
27. G.M. Neumann, L.A. Bach, 'The *N*-terminal Disulfide Linkages of Human Insulin-like Growth Factor-binding Protein-6 (hIGFBP-6) and hIGFBP-1 are Different as Determined by Mass Spectrometry,' *J. Biol. Chem.*, **274**(21), 14 587–14 594 (1999).
28. B.S. Antharavally, B.R. DasGupta, 'Covalent Structure of Botulinum Neurotoxin Type B; Location of Sulfhydryl Groups and Disulfide Bridge and Identification of C-termini of Light and Heavy Chains,' *J. Protein Chem.*, **17**(5), 417–428 (1998).
29. M.J. Feild, D.C. Nguyen, F.B. Armstrong, 'Amino Acid Sequence of *Salmonella typhimurium* Branched-chain Amino Acid Aminotransferase,' *Biochemistry*, **28**(12), 5306–5310 (1989).
30. J. Hofsteenge, W.J. Weijer, P.A. Jekel, J.J. Beintema, '*p*-Hydroxybenzoate Hydroxylase from *Pseudomonas fluorescens*. 1. Completion of the Elucidation of the Primary Structure,' *Eur. J. Biochem.*, **133**(1), 91–108 (1983).
31. R.H. Rice, G.E. Means, W. Duane Brown, 'Stabilization of Bovine Trypsin by Reductive Methylation', *Biochim. Biophys. Acta*, **492**, 316–321 (1977).
32. J.O. Hui, J. Le, V. Katta, R. Rosenfeld, M.F. Rohde, M. Haniu, 'Human Neurotrophin-3: a One-step Peptide Mapping Method and Complete Disulfide Characterization of the Recombinant Protein', *J. Protein Chem.*, **15**(4), 351–358 (1996).
33. S.A. Ahmed, T. Fairwell, S. Dunn, K. Kirschner, E.W. Miles, 'Identification of Three Sites of Proteolytic Cleavage in the Hinge Region Between the Two Domains of the beta 2 Subunit of Tryptophan Synthase of *Escherichia coli* or *Salmonella typhimurium*', *Biochemistry*, **25**(11), 3118–3124 (1986).
34. P. Guy, M. Jaquinod, H. Remigy, J.P. Andrieu, J. Gagnon, B. Bersch, A. Dolla, L. Blanchard, F. Guerlesquin, E. Forest, 'New Conformational Properties Induced by the Replacement of Tyr-64 in *Desulfovibrio vulgaris* Hildenborough ferricytochrome c553 using Isotopic Exchanges Monitored by Mass Spectrometry', *FEBS Lett.*, **395**(1), 53–57 (1996).
35. E.R. Hoff, M.V. Pickering, 'A Biotechnology Approach to Assessing HPLC Gradient Performance', *LC-GC*, **7**(4), 320–326 (1989).
36. C.G. Miller, D.H. Hawke, J. Tso, S. Early, 'Automated C-Terminal Protein Sequence Analysis using the Hewlett-Packard G1009A C-Terminal Protein Sequencing System', in *Techniques in Protein Chemistry VI*, ed. J.W. Crabb, Academic Press, San Diego, 219–227, 1995.
37. K.R. Clauser, S.C. Hall, D.M. Smith, J.W. Webb, L.E. Andrews, H.M. Tran, L.B. Epstein, A.L. Burlingame, 'Rapid Mass Spectrometric Peptide Sequencing and Mass Matching for Characterization of Human Melanoma Proteins Isolated by Two-dimensional PAGE', *Proc. Natl. Acad. Sci. USA*, **92**(11), 5072–5076 (1995).
38. J. Lei, D.A. Chen, F.E. Regnier, 'Rapid Verification of Disulfide Linkages in Recombinant Human Growth Hormone by Tandem Column Tryptic Mapping', *J. Chromatogr., A*, **808**(1–2), 121–131 (1998).
39. J. Preisler, F. Foret, B.L. Karger, 'On-line MALDI-TOF MS using a Continuous Vacuum Deposition Interface', *Anal. Chem.*, **70**(24), 5278–5287 (1999).

Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Marie-Isabel Aguilar
Monash University, Clayton, Australia

1 Introduction 1
2 Nature and Origin of Samples 2
3 The Role of Peptide and Protein Structure in Retention 4
4 Optimization of Resolution 6
4.1 Stationary Phase 6
4.2 Mobile Phase 7
4.3 Column Geometry 7
4.4 Operating Parameters 8
5 Mass Spectrometry Detection 10
6 Conclusions and Future Directions 11
Abbreviations and Acronyms 12
Related Articles 12
References 13

Reversed-phase high-performance liquid chromatography (RPHPLC) involves the separation of molecules on the basis of hydrophobicity. The separation depends on the hydrophobic binding of the solute molecule from the mobile phase to the immobilized hydrophobic ligands attached to the stationary phase, i.e. the sorbent. The solute mixture is initially applied to the sorbent in the presence of aqueous buffers, and the solutes are eluted by the addition of organic solvent to the mobile phase. Elution can proceed either under isocratic conditions, where the concentration of organic solvent is constant, or by gradient elution, whereby the amount of organic solvent is increased over a period of time. The solutes are therefore eluted in order of increasing molecular hydrophobicity. RPHPLC is a very powerful technique for the analysis of peptides and proteins because of a number of factors which include (1) the excellent resolution that can be achieved under a wide range of chromatographic conditions for very closely related molecules in addition to structurally distinct molecules, (2) the experimental ease with which chromatographic selectivity can be manipulated through changes in mobile phase characteristics, (3) the generally high recoveries and hence high productivity and (4) the excellent reproducibility of repetitive separations carried out over a long period of time, which is due partly to the stability of the sorbent materials under a wide range of mobile phase conditions. However, RPHPLC can cause the irreversible denaturation of protein samples, thereby reducing the potential recovery of material in a biologically active form.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

RPHPLC is now the central chromatographic technique in the analysis, structure elucidation and isolation of peptides and proteins and has therefore played a critical role in the enormous advances made in the biological sciences over the last 20 years. The extraordinary success of RPHPLC can be attributed to a number of factors. These include (1) the excellent resolution that can be achieved under a wide range of chromatographic conditions for very closely related molecules in addition to structurally distinct molecules, (2) the experimental ease with which chromatographic selectivity can be manipulated through changes in mobile phase characteristics, (3) the generally high recoveries and hence high productivity and (4) the excellent reproducibility of repetitive separations carried out over a long period of time, which is due partly to the stability of the sorbent materials under a wide range of mobile phase conditions.

The RPHPLC experimental system for the analysis of peptides and proteins usually consists of an *n*-alkylsilica-based sorbent from which the solutes are eluted with gradients of increasing concentrations of organic solvent such as acetonitrile containing an ionic modifier such as trifluoroacetic acid (TFA).[1,2] Complex mixtures of peptides and proteins can be routinely separated and low picomolar–femtomolar amounts of material can be collected for further characterization. Separations can be easily manipulated by changing the gradient slope, the operating temperature, the ionic modifier or the organic solvent composition.

The extensive use of RPHPLC for the purification of peptides, small polypeptides with molecular masses up to 10 000 Da and related compounds of pharmaceutical interest has not been replicated to the same extent for larger polypeptides (molecular mass > 10 000 Da) and globular proteins. The combination of the traditionally used acidic buffering systems and the hydrophobicity of the *n*-alkylsilica supports which can result in low mass yields or the loss of biological activity of larger polypeptides and proteins have often discouraged practitioners from using RPHPLC methods for large-scale protein separations. The loss of enzymatic activity, the formation of multiple peaks for compositionally pure samples and poor

yields of protein can all be attributed to the denaturation of protein solutes during the separation process using RPHPLC.[3–6] Whilst these features detract from the use of RPHPLC as a technique of choice in preparative purification protocols with proteins, these same characteristics can provide a unique opportunity to study protein folding and stability. Thus, the widespread practical application of RPHPLC has been recently accompanied by a significant improvement in our understanding of the molecular basis of the retention process and its impact on conformational stability of both peptides and proteins. As a result, RPHPLC can now also be used as a physicochemical tool for the analysis of the dynamic behavior of peptides and proteins at hydrophobic surfaces.[7–15]

The challenge facing the scientist who wishes to analyze and/or purify a peptide or protein sample is the selection of the initial separation conditions and subsequent optimization of the appropriate experimental parameters. This article provides an overview of the different experimental options available to achieve a high-resolution separation of a peptide or protein mixture using RPHPLC. The interested reader is also referred to a number of recent publications which provide a comprehensive theoretical and practical overview of RPHPLC of peptides and proteins together with other modes of high-performance chromatography.[16–20]

2 NATURE AND ORIGIN OF SAMPLES

RPHPLC is extremely versatile for the isolation of peptides and proteins from a wide variety of synthetic or biological sources and is used for both analytical and preparative applications.[1,2] Analytical applications

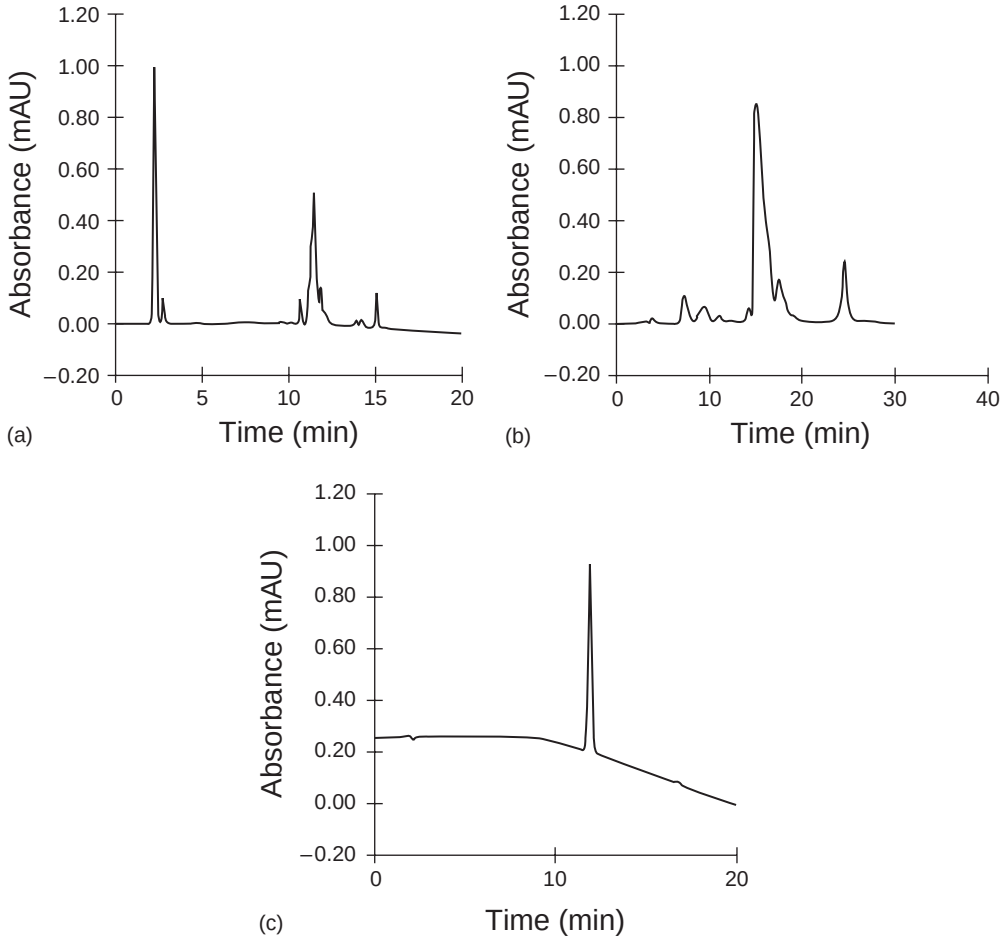

Figure 1 RPHPLC elution profiles illustrating the purification of a synthetic peptide. (a) Analytical profile (1 mg) of a crude peptide mixture from solid-phase peptide synthesis. Column: Zorbax 300 RP-C_{18}, 25 cm × 4.6 mm internal diameter (ID), 5 µm particle size, 30 nm pore size. Conditions: linear gradient from 0 to 60% acetonitrile with 0.1% TFA over 30 min, flow rate 1 mL min^{-1}, 25 °C. (b) Semi-preparative profile (10 mg). Column: Brownlee RP-C_{18}, 15 cm × 10 mm ID, 5 µm particle size, 30 nm pore size. Conditions: linear gradient from 0 to 60% acetonitrile with 0.1% TFA over 60 min, flow rate 1 mL min^{-1}, 25 °C. (c) Analytical profile (1 mg) of purified peptide. Chromatographic conditions as in (a).

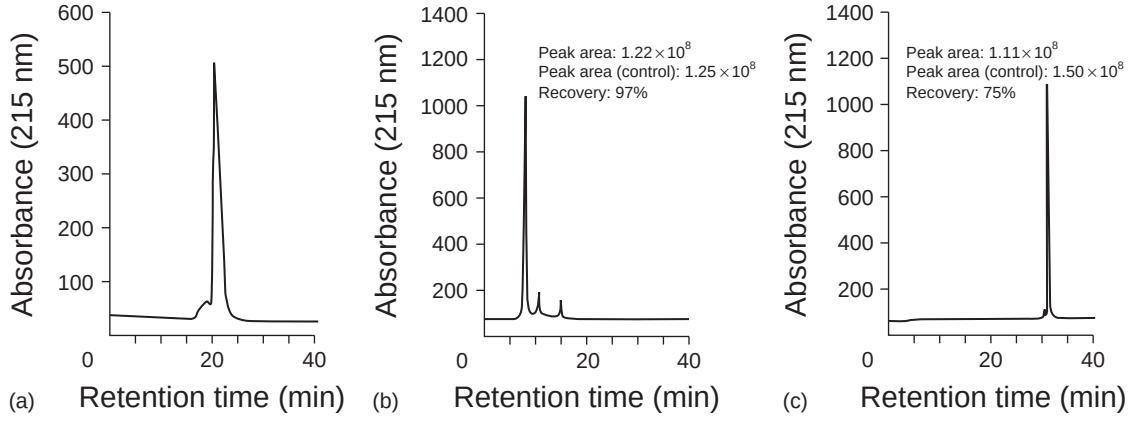

Figure 2 Multidimensional micropreparative HPLC of murine epidermal growth factor (mEGF), using size-exclusion, anion-exchange and reversed-phase chromatography. (a) mEGF (4.5 µg, 750 pmol) was chromatographed on a Superose 12 PC 3.2/30 size-exclusion chromatography column (Pharmacia, Uppsala, Sweden), 30 cm × 3.2 mm ID, using a mobile phase of 1% ammonium hydrogen carbonate–0.02% Tween 20 at a flow rate of 0.1 mL min^{-1}. (b) The major peak in (a) was collected and applied to a Mono-Q PC 1.6/5 anion-exchange column (Pharmacia), 30 cm × 3.2 mm ID, using a linear gradient from 0 to 1 M sodium chloride in 20 mM Tris, pH 7.5, over 50 min at a flow rate of 0.1 mL min^{-1}. (c) The major peak from (b) was collected and further purified by reversed-phase chromatography using a Brownlee RP-300 column, 30 cm × 2.1 mm ID, and a linear gradient from 0 to 60% acetonitrile in 0.15% TFA over 60 min at a flow rate of 0.1 mL min^{-1}. The integrated peak areas and the calculated recovery between stages are indicated in (b) and (c). (Reproduced by permission from Nice et al.[25])

range from the assessment of the purity of peptides following solid-phase peptide synthesis[21] to the analysis of tryptic maps of proteins.[22] Preparative RPHPLC is also used for applications from the micropurification of protein fragments for sequencing[23] to large-scale purification of synthetic peptides[21] and recombinant proteins.[24]

The complexity of the mixture to be chromatographed will depend on the nature of the source and the degree of preliminary cleanup that can be performed. In the case of synthetic peptides, RPHPLC is generally employed for both the initial analysis and the final large-scale purification. The purification of synthetic peptides usually involves an initial separation on an analytical scale to assess the complexity of the mixture followed by large-scale purification and collection of the target product. A sample of the purified material can then be subjected to RPHPLC analysis under the same or different elution conditions to check for purity. An example of this procedure is illustrated in Figure 1(a–c) for the purification of a synthetic pentapeptide.

The isolation of proteins from a biological cocktail derived from a tissue extract or biological fluid, for example, often requires a combination of techniques to produce a homogeneous sample. High-performance liquid chromatography (HPLC) techniques are then introduced at the later stages following initial precipitation, clarification and preliminary separations using soft gels. Purification protocols therefore need to be tailored to the specific target molecule and often involve the strategic combination of a range of high-performance techniques including

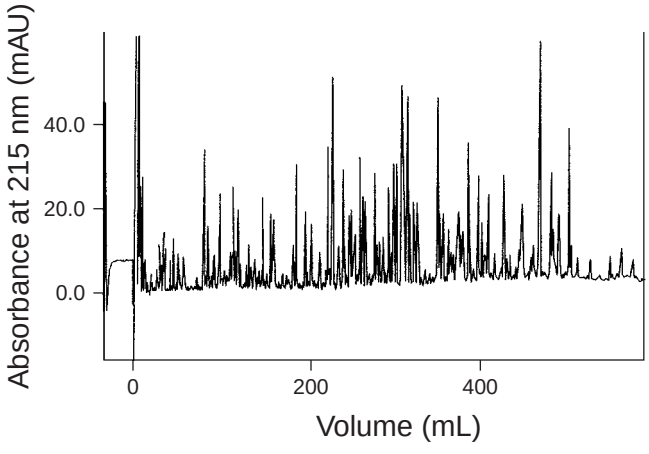

Figure 3 High-resolution reversed-phase chromatographic separation of a tryptic digest of a 165 kDa protein on a µRPC C_2/C_{18} ST 4.6/100 column (10 cm × 4.6 mm ID, 3 µm particle size, 12 nm pore size). Eluent A, 0.065% TFA in water; eluent B, 0.050% TFA in 84% acetonitrile. Gradient elution was carried out with 100% A for 2 column volumes (CV), 0–50% B for 392 CV (650 min), 50–100% B for 55 CV (91 min) and 100% B for 10 CV (17 min), at a flow rate of 1 mL min^{-1} and with detection at 215 nm. (Reproduced from Application Note No. 18-1119-53[26] by permission of Amersham Pharmacia Biotech.)

ion-exchange, size-exclusion, hydrophobic interaction, reversed-phase and affinity chromatography methods. One example of the use of multidimensional separations is shown in Figure 2(a–c), where the judicious use of size-exclusion, ion-exchange and RPHPLC methods

resulted in the efficient purification of mEGF[25] in which manipulation of the sample between stages has been minimized and selectivity has been maximized.

Peptide mapping by RPHPLC is now an indispensable tool in the characterization of new proteins and the quality control of proteins derived from recombinant DNA.[22] An example of the high-resolution analysis of a tryptic digest of a protein is shown in Figure 3.[26] This figure, in which 150 protein fragments were resolved, demonstrates the highly selective separation that can be achieved with enzymatic digests of proteins using RPHPLC as part of the quality control or structure determination of a recombinant or natural protein. The chromatographic separation was obtained with an C_2/C_{18} stationary phase packed in a column of dimensions 10 cm × 4.6 mm internal diameter (ID). Separated components can then be directly subjected to further analysis such as automated Edman sequencing or electrospray mass spectrometry (ESMS).

3 THE ROLE OF PEPTIDE AND PROTEIN STRUCTURE IN RETENTION

The mechanism by which peptides and proteins are retained in RPHPLC involves the hydrophobic expulsion of the solute from the polar mobile phase followed by adsorption on a nonpolar sorbent.[27] Peptides and proteins are thus retained to different extents depending on their surface hydrophobicity, the eluotropicity of the mobile phase and the nature of the hydrophobic ligands. The physicochemical basis of RPHPLC lies in the hydrophobic interaction between a peptide or protein and the hydrophobic sorbent, and can be described in terms of the solvophobic theory where the isocratic retention factor can be expressed as Equation (1):[27]

$$\ln k' = \log k_0 - \left[N\Delta A_h + 4.836 N^{\frac{1}{3}} (\kappa^e - 1) V^{\frac{2}{3}} \right] \frac{\gamma}{RT} \quad (1)$$

where N is Avogadro's number, ΔA_h is the hydrophobic contact area of the interacting solute, V is the mean molar volume of the solvent, R is the gas constant, T is the temperature and γ is the surface tension of the mobile phase. The parameter κ^e is the ratio of the energy required to create a cavity for a solvent molecule and the energy required to extend the planar surface of the liquid by the surface area of the solute molecule. Thus, simply stated, elution in RPHPLC is achieved through a decrease in the microscopic surface tension associated with the solute–sorbent interface. Experimentally this is achieved through changes in the water content by variation in the mole fraction of organic solvent in RPHPLC. However, high-resolution isocratic elution of peptides and proteins can rarely be achieved as the experimental window of organic solvent concentration required for solute elution is typically very narrow. Hence peptides and proteins are generally separated using gradient elution conditions. Under these conditions, the experimentally observed retention data can be analyzed according to the linear solvent strength (LSS) model [Equation 2]:[28–31]

$$\log \overline{k} = \log k_0 - S\overline{\varphi} \quad (2)$$

where $\overline{k}$ is the solute median capacity factor and $\overline{\varphi}$ is the corresponding organic mole fraction. The $\log k_0$ value is the affinity of the solute for the sorbent in the absence of organic solvent. By analogy with the solvophobic Equation (1) above, the S value is related to the hydrophobic contact region established between the solute and the sorbent surface. Both S and $\log k_0$ can be derived from plots of $\log \overline{k}$ versus $\overline{\varphi}$ and the LSS model provides the computational basis for the rational optimization of peptide and protein separations. An example of these plots is shown in Figure 4 for a series of peptide analogs related to neuropeptide Y-[18–36] which differ in sequence only by the substitution of a single D-amino acid residue.[7] The data presented in these plots provide a clear example of how to optimize

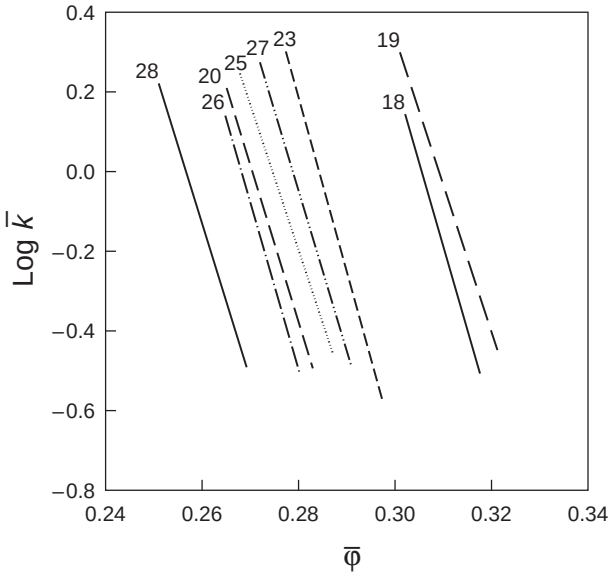

Figure 4 Plots of $\log \overline{k}$ versus $\overline{\varphi}$ for a series of D-analogs of neuropeptide Y-[18–36] separated on C_{18} silica with acetonitrile as the organic modifier at 4 °C. Column: Bakerbond (J.T. Baker, Phillipsburg, NJ, USA) RP-C_{18}, 25 cm × 4.6 mm ID, 5 µm particle size, 30 nm pore size. The retention plot for each analog is designated by the residue position of the D-amino acid substitution. The plots were derived from best fit analysis to the data points. (Reproduced from Lazoura et al.[7] by permission of The Biophysical Society.)

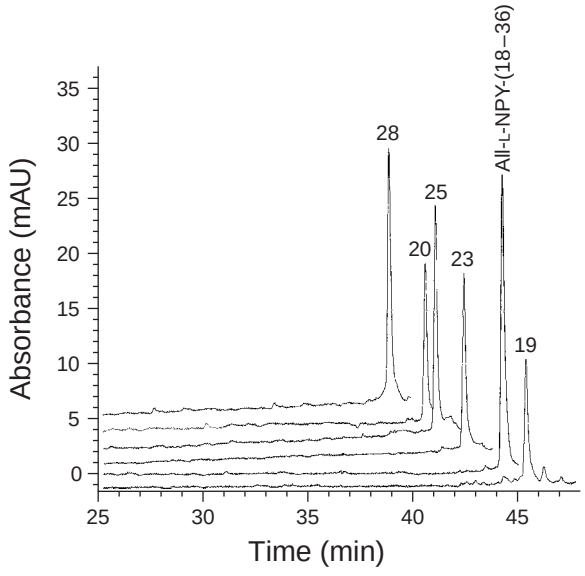

Figure 5 Chromatograms showing the separation of the all-L-NPY-[18–36] peptide and NPY-[18–36] analogs with D-substitutions at positions 19, 20, 23, 25 and 28 separated on a C_{18} silica [Bakerbond (J.T. Baker, Phillipsburg, NJ, USA) RP-C_{18}, 25 cm × 4.6 mm ID, 5 μm particle size, 30 nm pore size] with a 60-min gradient from 0 to 50% acetonitrile at 1 mL min^{-1}. The separation was used to derive the retention plots shown in Figure 4. (Reproduced from Lazoura et al.[7] by permission of The Biophysical Society.)

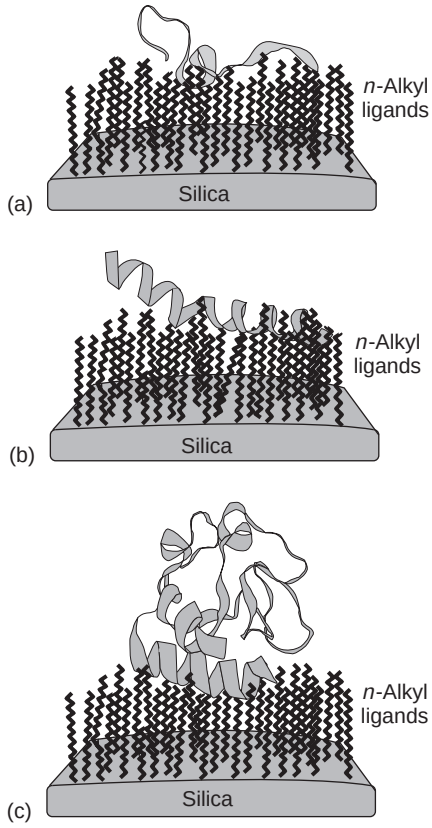

Figure 6 Schematic representation of the binding of (a) a small peptide (2–10 amino acid residues), (b) a large peptide (11–30 amino acid residues) and (c) a polypeptide or protein (>30 amino acid residues) to an RPHPLC silica-based sorbent. The peptide or protein interacts with the immobilized hydrophobic ligands through the hydrophobic chromatographic contact region.

separations, whereby maximum optimization will be achieved by selecting elution conditions where there is the greatest separation between the retention plots.[28–31] The elution profiles of five of the NPY-[18–36] analogs are shown in Figure 5 and demonstrate that the order of elution of the peaks corresponds to the relative position of the retention plots shown in Figure 4 and also illustrates the corresponding degree of resolution that is obtained.[7]

The plots shown in Figure 4 also illustrate the ability of RPHPLC to resolve very small differences in peptide structure as each NPY-[18–36] analog exhibited significant differences in α-helical structure in different solution environments. These[7] and other studies[8–15] thus serve to illustrate how RPHPLC can act as a molecular probe of peptide surface topography. In particular, analysis of the S and $\log k_0$ values as a function of temperature can be used to gain significant insight into the conformational integrity of peptides and proteins during interaction with hydrophobic surfaces.[7–9,12]

Peptides and proteins are retained in RPHPLC through hydrophobic interactions. As depicted in Figure 6(a), the hydrophobic contact region for small peptides involves the contribution from all or a large proportion of the peptide structure. As a consequence, the retention time for small peptides can be predicted on the basis of the amino acid composition through summation of the hydrophobicity coefficients for the constituent amino acid residues.[32,33] In contrast, it has been well established that proteins interact with the chromatographic surface in an orientation-specific manner, in which their retention time is determined by the molecular composition of specific contact regions.[34–36] For larger polypeptides and proteins which adopt a significant degree of secondary and tertiary structure, there is no correlation between retention times and the summated hydrophobicity coefficients of all constituent amino acid residues. In these cases, the chromatographic contact region comprises a small proportion of the total molecular surface as shown in Figure 6(b) and (c). It is also generally observed that the retention time increases with increase in solute molecular mass as a result of increases in hydrophobic surface area. However, it is the molecular composition of the contact region that determines the retention and bandwidth properties of the peptide or protein solute. Thus, in the presence of any degree of preferred secondary structure,

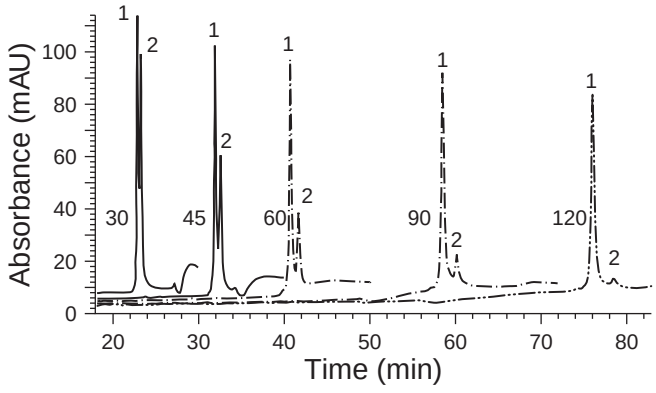

Figure 7 Chromatograms showing split peak elution profiles obtained for equine apocytochrome *c* chromatographed on a C_{18} silica [Bakerbond (J.T. Baker, Phillipsburg, NJ, USA) RP-C_{18}, 25 cm × 4.6 mm ID, 5 µm particle size, 30 nm pore size] over gradient times of 30, 45, 60, 90 and 120 min from 0 to 50% acetonitrile at 1 mL min^{-1} at 5 °C. The first- and second-eluted conformers are denoted 1 and 2, respectively. (Reproduced from K.L. Richards, M.I. Aguilar, M.T.W. Hearn, 'A Comparative Study of the Retention Behaviour and Stability of Cytochrome *c* in RPHPLC', *J. Chromatogr.*, **676**, 17–31 (1994), Copyright 1994, by permission from Elsevier Science.)

no simple relationship will exist between the retention time and the summated retention coefficients unless the identity of the contact region can be established.

An example of the potential impact that protein conformational changes can have on the elution profiles in RPHPLC is shown in Figure 7. This figure shows the elution profiles of horse apocytochrome *c* separated under gradient conditions with aqueous TFA–acetonitrile mixtures on an octadecylsilica sorbent at 5 °C.[35] It is evident that a single peak was observed at longer gradient times. However, at shorter gradient times between 30 and 60 min, two peaks were observed which corresponds to the existence of two conformational species. In summary, under conditions where peptides or proteins undergo very rapid or very slow conformational interconversions, a single Gaussian shape peak will be observed. However, if the peptide or protein solute experiences structural changes characterized by a rate constant which is comparable to the chromatographic separation time, then more complex elution profiles, including asymmetric and/or split peaks, may be observed.[3–6,8,9,12,35]

4 OPTIMIZATION OF RESOLUTION

An appreciation of the experimental factors that control the resolution of peptides and proteins in interactive modes of chromatography is essential in the development and manipulation of separation protocols to obtain the desired separation. The optimization of high-resolution separations of peptides and proteins involves the separation of sample components through manipulation of both retention times and solute peak shape. In the following sections, the different RPHPLC techniques which are commonly employed for the analysis and purification of peptides and proteins are described. In particular, the experimental approaches that can be used to manipulate resolution in order to achieve a particular level of separation are discussed below in terms of the parameters that control resolution and are illustrated with relevant examples.

4.1 Stationary Phase

The most commonly employed experimental procedure for the RPHPLC analysis of peptides and proteins generally involves the use of an octadecylsilica-based sorbent and a mobile phase. The chromatographic packing materials that are generally used are based on microparticulate porous silica which allows the use of high linear flow velocities, resulting in favorable mass transfer properties and rapid analysis times.[37–39] The silica is chemically modified by a derivatized silane bearing an *n*-alkyl hydrophobic ligand. The most commonly used ligand is *n*-octadecyl (C_{18}), while *n*-butyl (C_4) and *n*-octyl (C_8) also find important application and phenyl and cyanopropyl ligands can provide different selectivity.[40] The process of chemical immobilization of the silica surface results in approximately half of the surface silanol group being modified. The sorbents are therefore generally subjected to further silanization with a small reactive silane to produce an end-capped packing material.

The type of *n*-alkyl ligand significantly influences the retention of peptides and proteins and can therefore be used to manipulate the selectivity of peptide and protein separations. Although the detailed molecular basis of the effect of ligand structure is not fully understood, a number of factors including the relative hydrophobicity and ligand chain length, flexibility and the degree of exposure of surface silanols all play a role in the retention process.[41–45] An example of the effect of chain length on peptide separations is shown in Figure 8(a) and (b).[1] It can be seen that the peaks labelled T_3 and T_{13} are fully resolved on the C_4 packing but cannot be separated on the C_{18} material. In contrast, the peptides T_5 and T_{18} are unresolved on the C_4 column but fully resolved on the C_{18} material. In addition to effects on peptide selectivity, the choice of ligand type can also influence protein recovery and conformational integrity of protein samples. Generally higher protein recoveries are obtained with the shorter and less hydrophobic *n*-butyl ligands. However, proteins have also been obtained in high yield with *n*-octadecylsilica.[46–48]

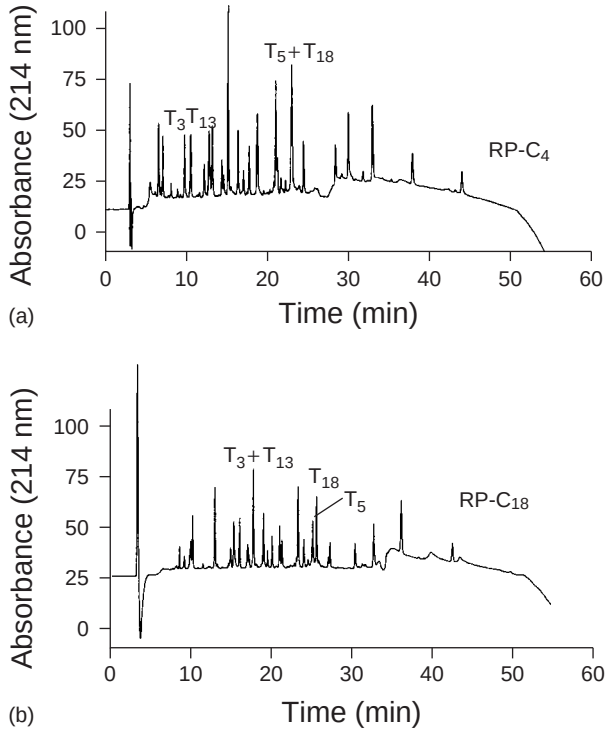

Figure 8 Influence of *n*-alkyl chain length on the separation of tryptic peptides derived from porcine growth hormone. (a) Bakerbond (J.T. Baker, Phillipsburg, NJ, USA) RP-C_4, 25 cm × 4.6 mm ID, 5 µm particle size, 30 nm pore size; (b) Bakerbond RP-C_{18}, 25 cm × 4.6 mm ID, 5 µm particle size, 30 nm pore size. Conditions: linear gradient from 0 to 90% acetonitrile with 0.1% TFA over 60 min, flow rate 1 mL min^{-1}, 25 °C. (Reproduced from Aguilar and Hearn[1] by permission of Academic Press.)

The denaturation of proteins by RPHPLC sorbents can be controlled by using silica which has been coated with polymethacrylate-based polymers which provide a series of sorbents with varying surface hydrophobicity, and in which the surface silanol groups have been masked.[49,50] Silica-based packings are also susceptible to cleavage at pH > 7. This limitation can severely restrict the utility of these materials for separations which require basic pH conditions to effect resolution. In these cases, alternative stationary phases have been developed, including cross-linked poly(styrene–divinylbenzene)[51,52] and porous zirconia,[53,54] which are all stable to hydrolysis at alkaline pH.

The geometry of the particle in terms of the particle diameter and pore size is also an important feature of the packing material. Improved resolution can be achieved by decreasing the particle diameter and the most commonly used range of particle diameters for analytical-scale RPHPLC is 3–5 µm. There are also examples of the use of nonporous particles of smaller diameter.[55] For preparative-scale separations, 10–20-µm particles are utilized. The pore size of RPHPLC sorbents is also an important factor that must be considered. For peptides, the pore size generally ranges between 100 and 300 Å depending on the size of the peptides. Porous materials of ≥300 Å pore size are necessary for the separation of proteins, as the solute molecular diameter must be at least one tenth the size of the pore diameter to avoid restricted diffusion of the solute and to allow the total surface area of the sorbent material to be accessible. The recent development of particles with 6000–8000 Å pores with a network of smaller pores of 500–1000 Å has also allowed very rapid protein separations to be achieved.[56,57]

4.2 Mobile Phase

One of the most powerful characteristics of RPHPLC is the ability to manipulate solute retention and resolution through changes in the composition of the mobile phase. In RPHPLC, peptide and protein retention is due to multisite interactions with the ligands. The practical consequence of this is that high-resolution isocratic elution of peptides and proteins can rarely be achieved as the experimental window of solvent concentration required for their elution is very narrow. Mixtures of peptides and proteins are therefore routinely eluted by the application of a gradient of increasing organic solvent concentration. RPHPLC is generally carried out with an acidic mobile phase, with TFA being the most commonly used additive owing to its volatility. Phosphoric acid, perchloric acid, formic acid, hydrochloric acid, acetic acid and heptafluorobutyric acid have also been used.[58–61] Alternative additives such as nonionic detergents can be used for the isolation of more hydrophobic proteins such as membrane proteins.[62]

The three most commonly employed organic solvents in RPHPLC are acetonitrile, methanol and 2-propanol, which all exhibit high optical transparency at the detection wavelengths used for peptide and protein analysis. Acetonitrile provides the lowest viscosity solvent mixtures and 2-propanol is the strongest eluent. An example of the influence of organic solvent is shown in Figure 9, where changes in selectivity can be observed for a number of peptide peaks in the tryptic map. In addition to the eluotropic effects, the nature of the organic solvent can also influence the conformation of both peptides and proteins and will therefore have an additional effect on selectivity through changes in the structure of the hydrophobic contact region.[5,63] In the case of proteins, this may also impact on the level of recovery of biologically active material.

4.3 Column Geometry

The desired levels of efficiency and sample loading size determine the dimensions of the column to be used. For

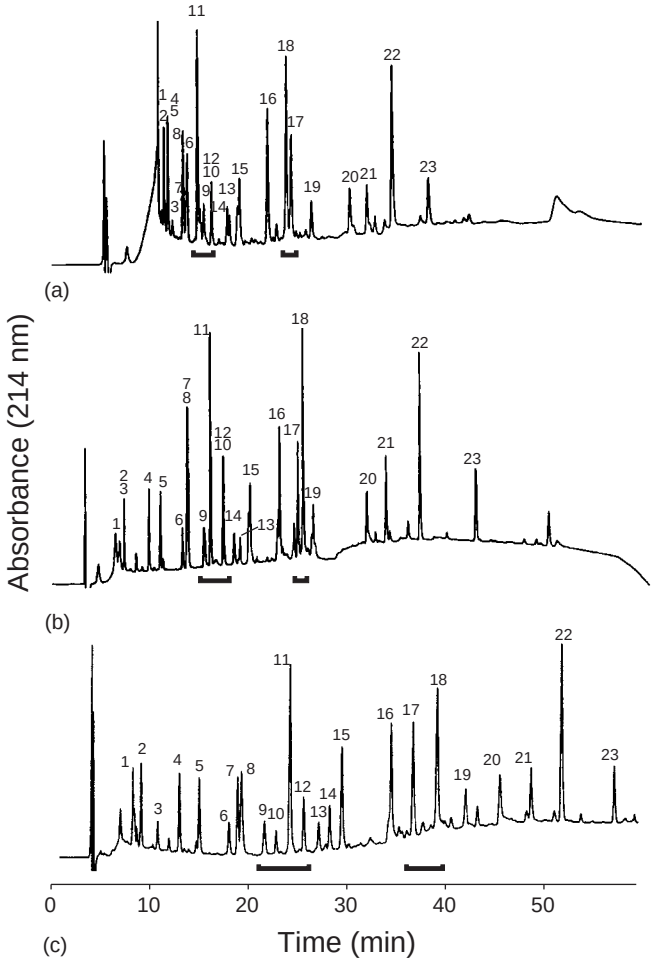

Figure 9 Influence of organic solvent on the reversed-phase chromatography of tryptic peptides derived from porcine growth hormone. Column: Bakerbond (J.T. Baker, Phillipsburg, NJ, USA) RP-C_4, 25 cm × 4.6 mm ID, 5 µm particle size, 30 nm pore size. Conditions: linear gradient from 0 to 90% (a) 2-propanol, (b) acetonitrile or (c) methanol with 0.1% TFA over 60 min, flow rate 1 mL min^{-1}, 25 °C. (Reproduced from Aguilar and Hearn[1] by permission of Academic Press.)

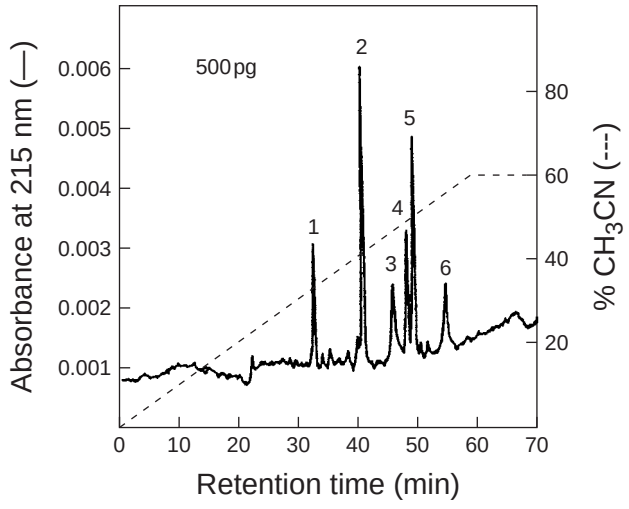

Figure 10 Separation of proteins by reversed-phase capillary chromatography. Column: Brownlee RP-300, 15 cm × 0.2 mm ID, 5 µm particle size, 30 nm pore size. Conditions: linear gradient from 0 to 60% acetonitrile over 60 min, flow rate 1.4 µL min^{-1}. Peaks: 1 = ribonuclease B; 2 = lysozyme; 3 = bovine serum albumin; 4 = carbonic anhydrase; 5 = myoglobin; 6 = ovalbumin. (Reproduced from Moritz and Simpson[65] by permission of Plenum Publishing Corp.)

small peptides and proteins, increased resolution will be obtained with increases in column length. Thus, for applications such as tryptic mapping, column lengths between 15 and 25 cm and ID of 4.6 mm are generally employed. However, for larger proteins, low mass recovery and loss of biological activity may result with these columns due to irreversible binding and/or denaturation. In these cases, shorter columns of between 2 and 20 cm in length can be used. For preparative applications on the 1–500 mg scale, such as the purification of synthetic peptides, so-called semipreparative columns of dimensions 30 × 1 cm ID and preparative columns of 30 × 2 cm ID can be used.

The selection of the ID of the column is based on the sample capacity and detection sensitivity. While most analytical applications are carried out with columns of 4.6 mm ID (as shown in Figures 3, 5 and 8), for samples derived from previously unknown proteins where there is a limited supply of material, the task is to maximize the detection sensitivity. In these cases, the use of narrow-bore columns of 1–2 mm ID can be used, which allow the elution and recovery of samples in much smaller volumes of solvent.[23] Capillary chromatography is also finding increasing application where capillary columns of ID between 0.2 and 0.4 mm and a length of 15 cm result in the analysis of femtomole amounts of sample, as shown in Figure 10.[64] The effect of decreasing column ID on detection sensitivity is shown in Figure 11 for the analysis of lysozyme on a C_{18} material packed into columns of 4.6, 2.1 and 0.3 mm ID.[65]

4.4 Operating Parameters

There are several operating parameters that can be changed in order to manipulate the resolution of peptide and protein mixtures in RPHPLC. These parameters include the gradient time, the gradient shape, the mobile phase flow rate and the operating temperature. A typical experiment with an analytical-scale column would utilize a linear gradient from 5% to between 50 and 100% organic solvent over a time range of 20–120 min while flow rates are between 0.5 and 2.0 mL min^{-1}. With microbore columns (1–2 mm ID) flow rates of 50–250 µL min^{-1} are used, while for capillary columns of 0.2–0.4 mm ID, flow rates of 1–4 µL min^{-1} are applied.

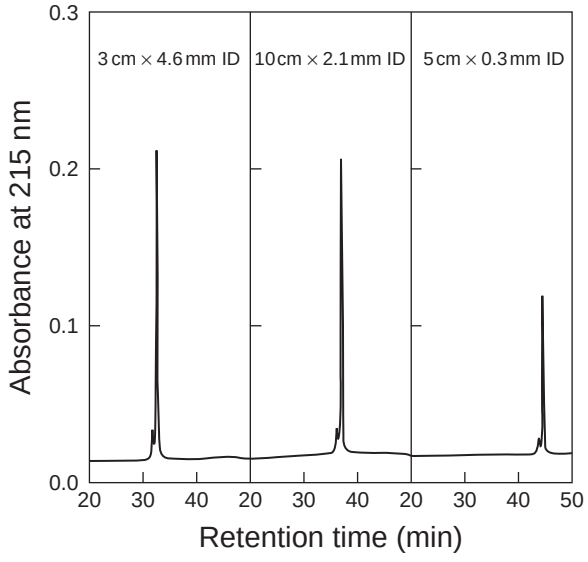

Figure 11 Effect of column ID on detector sensitivity. Column: Brownlee RP-300 C_8 (7 μm particle size, 30 nm pore size), 3 cm × 4.6 mm ID and 10 cm × 2.1 mm ID (Applied Biosystems) and 5 cm × 0.32 mm ID. Conditions: linear gradient from 0 to 60% acetonitrile with 0.1% TFA over 60 min, 45 °C. Flow rates: 1 mL min^{-1}, 200 μL min^{-1} and 4 μL min^{-1} for the 4.6, 2.1 and 0.32 mm ID columns, respectively. Sample loadings: lysozyme, 10, 4 and 0.04 μg for the 4.6, 2.1 and 0.32 mm ID columns, respectively. (Reproduced from R.L. Moritz, R.J. Simpson, 'Application of Capillary Reversed-phase High-performance Liquid Chromatography to High Sensitivity Protein Sequence Analysis', *J. Chromatogr.*, **599**, 119–130 (1992), Copyright 1992, by permission from Elsevier Science.)

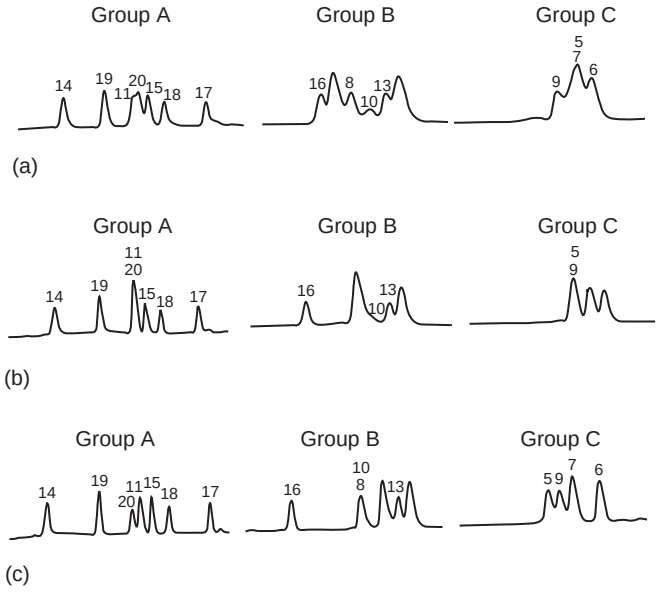

Figure 12 Effect of gradient time on the resolution of 30 S ribosomal proteins by reversed-phase chromatography. Column: DuPont Bioseries Protein PLUS. Conditions: linear gradient from 25 to 46% acetonitrile with 0.1% triethylamine and 0.042% TFA over (a) 60, (b) 120 or (c) 240 min at a flow rate of 0.7 mL min^{-1}. (Reproduced by permission from Ghrist et al.[66])

At the preparative end of the scale with columns of 10–20 mm ID, flow rates between 5 and 20 mL min^{-1} are required.

The choice of gradient conditions will depend on the nature of the molecules of interest. The influence of gradient time on the separation of a series of ribosomal proteins is shown in Figure 12(a–c).[66] Generally, the use of longer gradient times provides improved separation. However, these conditions also increase the residence time of the peptide or protein solute at the sorbent surface, which may then result in an increase in the degree of denaturation.

The operating temperature can also be used to manipulate resolution. While the separation of peptides and proteins is normally carried out at ambient temperature, solute retention in RPHPLC is influenced by temperature through changes in solvent viscosity. In addition, peptide and protein conformation can also be manipulated by temperature. Changes in temperature can therefore also be used to manipulate the structure and retention of peptide mixtures. For peptides, it has been shown that secondary structure can actually be enhanced through binding to the hydrophobic sorbent.[7–15] In the case of proteins which are to be subjected to further chemical analysis and thus where recovery of a biologically active protein is not essential, increasing temperature can be used to modulate retention via denaturation of the protein structure.[3–6] However, if the efficient recovery of both mass and biological activity is of paramount importance, the use of elevated temperatures is not an option.

The chromatographic recovery of proteins from polyacrylamide gels is another important application of RPHPLC. Inverse gradient elution chromatography has been successfully utilized for the micropreparative isolation of proteins from sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) electroeluates.[67] This approach is based on observations that certain RPHPLC packings display strong interactions with proteins at high organic solvent concentrations. This allows the loading of the electroeluate under conditions where the protein is retained while the sodium dodecyl sulfate and other gel-related contaminants are washed through the column. The protein is then recovered in high yield by a gradient of decreasing organic solvent.

Detection of peptides and proteins in RPHPLC, generally involves detection at between 210–220 nm which is specific for the peptide bond, or at 280 nm which corresponds to the aromatic amino acids tryptophan and tyrosine. The use of photodiode array detectors can enhance the detection capabilities by the on-line accumulation of complete solute spectra. The spectra can

then be used to identify peaks specifically on the basis of spectral characteristics and for the assessment of peak purity.[68–70] In addition, second derivative spectroscopy can provide information on the conformational integrity of proteins following elution.[71,72]

5 MASS SPECTROMETRY DETECTION

One of the most significant recent advances in bioanalytical technology has been the advent of mass spectrometry for the analysis and measurement of peptide and protein molecular mass.[73] In particular, the development of on-line ESMS following RPHPLC [liquid chromatography/electrospray mass spectrometry (LC/ESMS)] has provided a powerful detection system for the rapid analysis of peptides and proteins.[74] For example, LC/ESMS significantly facilitates the identification of peptide fragments in peptide mapping studies,[22] as illustrated in Figure 13(a) and (b), which shows the analysis of an Arg-C digest of plasminogen activator separated on a C_{18} column with (a) the total ion current and (b) the elution profile at 214 nm. Figure 14 shows the RPHPLC separation of a Lys-C specific endopeptidase (Lys-C) digest of two forms of apolipoprotein A1 in which there is a significant difference in the retention time of the last major peak K15.[75] Subsequent on-line LC/ESMS revealed the presence of an arginine–cysteine substitution and illustrates the power of LC/ESMS in the analysis of microheterogeneous forms of naturally derived proteins or protein mutants with as little as one change in the amino acid sequence. Other important applications involve the identification of post-translational modifications of peptides and proteins[76,77] including sites of N- and O-linked glycosylation,[78–80] assignment of disulfide bonds[81,82] and the identification

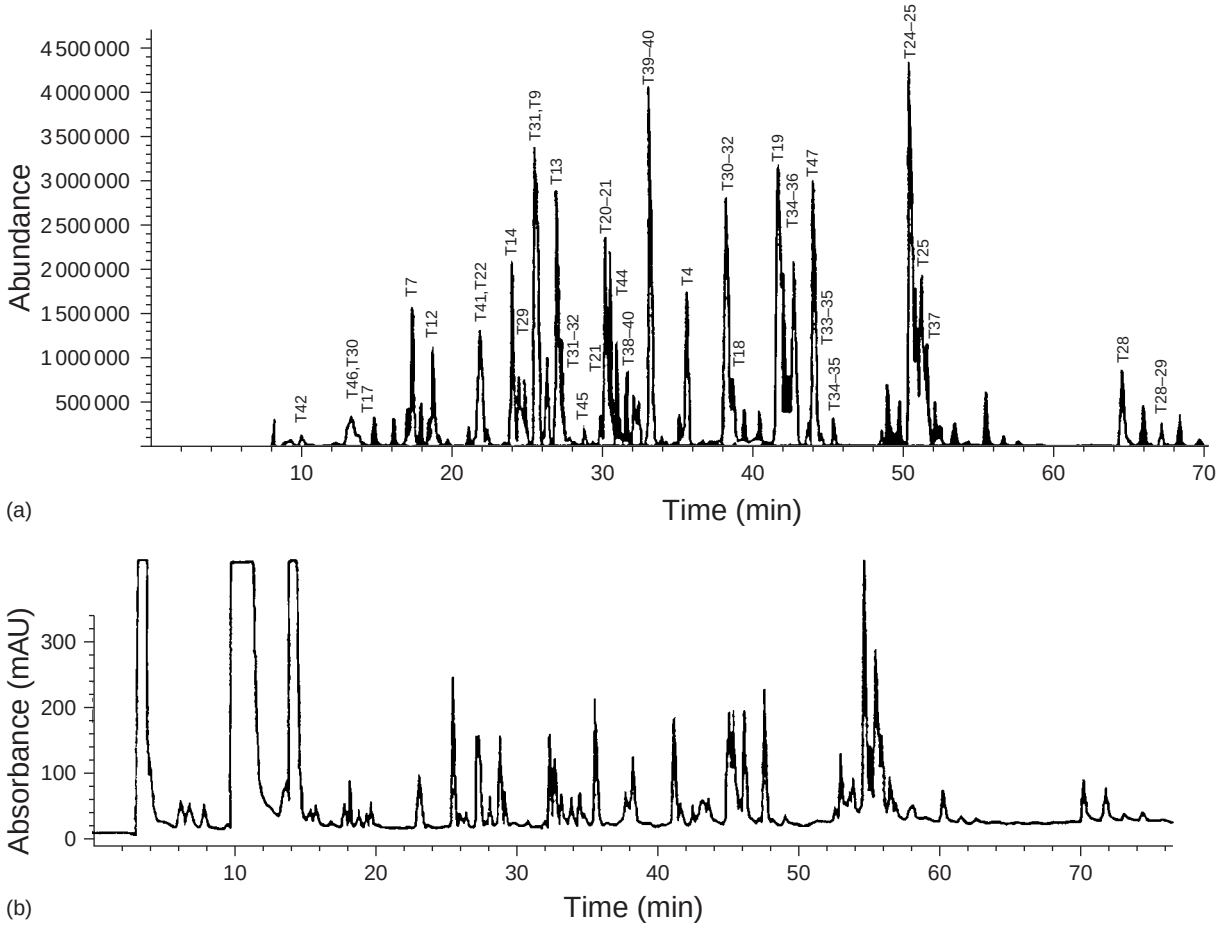

Figure 13 LC/ESMS of a tryptic digest of single-chain plasminogen activator. Column: Vydac C_{18}, 5 µm particle size, 30 nm pore size. Conditions: linear gradient from 0 to 60% acetonitrile with 0.1% TFA over 90 min, 45 °C, flow rate 0.2 mL min^{-1}. (a) ESMS total ion current; (b) detection at 214 nm. (Reproduced from A. Apffel, J. Chakel, S. Udivar, W.S. Hancock, C. Souders, E. Pungor, Jr, 'Application of Capillary Electrophoresis, High-performance Liquid Chromatography, On-line Electrospray Mass Spectrometry and Matrix-assisted Laser Desorption Ionization Time of Flight Mass Spectrometry to the Characterization of Single-chain Plasminogen Activator', *J. Chromatogr.*, **A717**, 41–60 (1995), Copyright 1995, by permission from Elsevier Science.)

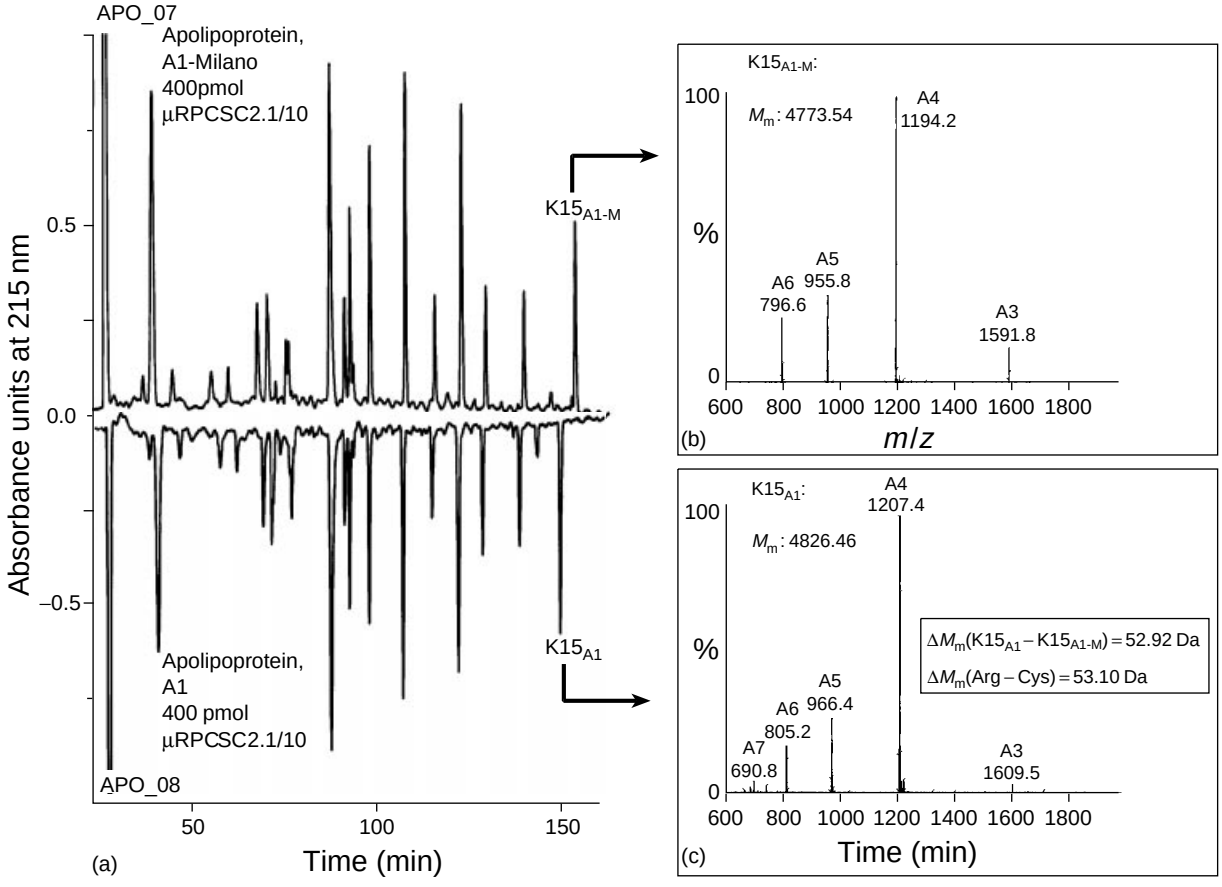

Figure 14 LC/ESMS analysis of two forms of apolipoprotein A1. (a) Reversed projections of the separation of 400 pmol each of Apo A1-M (upper trace) and Apo A1 (lower trace). Column: μRPC C_2/C_{18} ST 4.6/100 (10 cm × 4.6 mm ID, 3 μm particle size, 12 nm pore size). Eluent A, 0.25% pentafluoropropionic acid (PFPA) in water; eluent B, 0.25% PFPA in 100% acetonitrile. Conditions: linear gradient from 10 to 60% B over 200 min, flow rate 25 μL min^{-1}, 25 °C. (b) LC/ESMS trace obtained from peptide material eluted in peak K15A1-M in (a). (c) LC/ESMS trace obtained from peptide material eluted in peak K15A1 in (a). (Reproduced from Renlund et al.[75] by permission of the Protein Research Foundation, Osaka, Japan.)

of peptides bound to major histocompatibility complex molecules.[83] Although LC/ESMS has clearly enhanced the analytical power of RPHPLC, the analysis of large complex proteins can be hindered by the wide range of peptides generated by enzymatic digestion, particularly in the case of heterogeneous glycoproteins. In these circumstances, multidimensional analytical techniques such as capillary electrophoresis and matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) mass spectrometry can be used to complement the information obtained by LC/ESMS and further expand the analytical capability of RPHPLC techniques.[74]

6 CONCLUSIONS AND FUTURE DIRECTIONS

RPHPLC is now an indispensable tool for the high-performance separation of complex mixtures of peptides and proteins. The number of applications of RPHPLC in peptide and protein purification continues to expand at an extremely rapid rate. Solid-phase peptide synthesis and recombinant DNA techniques have allowed the production of large quantities of peptides and proteins which need to be highly purified. The design of multidimensional purification schemes to achieve high levels of product purity highlight the power of RPHPLC techniques in the production of peptide- and protein-based therapeutics.

Consequently, RPHPLC is now firmly entrenched as the central tool for the analysis of peptides and proteins and, in particular, plays a pivotal role in the pharmaceutical industry, providing the core analytical technique at all stages of the development of peptide- and protein-based therapeutics. Following purification, mass spectrometry can be used to confirm the structural identity of synthetic peptides, while recombinant proteins require further structural analysis by high-resolution analytical

fingerprinting to confirm the amino acid sequence. RPHPLC will therefore continue to be the central method for the characterization and quality control analysis of synthetic peptides and recombinant proteins. Moreover, the coupling of mass spectrometry to allow on-line identification of samples will become routine and continue to expand the analytical power of HPLC.

Other areas of separation technology in which significant advances are expected to emerge are in the areas of miniaturization and high-speed analysis to allow the efficient purification of femtomolar to attomolar levels of material. These techniques will have an important impact on the discovery of new bioactive peptides and novel proteins as potential candidates for new therapeutics.

The development of fully mechanistic models which describe the underlying thermodynamic and kinetic processes involved in the interaction of peptides and proteins with hydrophobic stationary phase surfaces has not kept pace with the growth of RPHPLC applications. This is predominantly due to the complex nature of peptide and protein structure and difficulties in determining the precise structure of the solvated sorbent material. However, both spectroscopic and molecular modeling techniques are starting to provide detailed molecular information on peptide and protein surface interactions, which will not only allow further advances in the rational design of new sorbent materials but will further establish RPHPLC as a physicochemical tool for the analysis of peptide and protein surface interactions.

Finally, there has been a strong synergy between the biotechnology industry and the field of biomacromolecular RPHPLC as several significant recent advances in separation technology have been driven by the stringent and continually evolving regulatory requirements of the biotechnology industry. RPHPLC will therefore continue to be at the heart of the analytical techniques with which scientists in all areas of biomolecular research must arm themselves to be able to characterize fully the identity, purity and potency of peptides and proteins.

ABBREVIATIONS AND ACRONYMS

ESMS	Electrospray Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ID	Internal Diameter
LC/ESMS	Liquid Chromatography/Electrospray Mass Spectrometry
LSS	Linear Solvent Strength
Lys-C	Lys-C Specific Endopeptidase
MALDI/TOF	Matrix-assisted Laser Desorption/ Ionization Time-of-flight
mEGF	Murine Epidermal Growth Factor
PFPA	Pentafluoropropionic Acid
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TFA	Trifluoroacetic Acid

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Food **(Volume 5)**
Liquid Chromatography in Food Analysis • Proteins, Peptides, and Amino Acids Analysis in Food

Peptides and Proteins **(Volume 7)**
Chromatography of Membrane Proteins and Lipoproteins • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Molecular Modeling in Peptide and Protein Analysis • Peptide Diastereomers, Separation of • Posttranslational Oxidative Modifications of Proteins • Protein–Drug Interactions • Protein–Oligonucleotide Interactions • Proteolytic Mapping

Pharmaceuticals and Drugs **(Volume 8)**
Combinatorial Chemistry Libraries, Analysis of • Mass Spectrometry in Pharmaceutical Analysis • Proteins and Peptides Purification in Pharmaceuticals Analysis

Process Instrumental Methods **(Volume 9)**
Chromatography in Process Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Biopolymer Chromatography • Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Reversed Phase Liquid Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. M.I. Aguilar, M.T.W. Hearn, 'High Resolution Reversed-phase High-performance Liquid Chromatography of Peptides And Proteins', *Methods Enzymol.*, **270**, 3–26 (1996).
2. C.T. Mant, R.S. Hodges, 'Analysis of Peptides by High-performance Liquid Chromatography', *Methods Enzymol.*, **271**, 3–50 (1996).
3. A.W. Purcell, M.I. Aguilar, M.T.W. Hearn, 'Conformational Effects in the RPHPLC of Polypeptides. II: The Role of Insulin A and B Chains in the Chromatographic Behavior of Insulin', *J. Chromatogr.*, **711**, 71–79 (1995).
4. K.L. Richards, M.I. Aguilar, M.T.W. Hearn, 'The Effect of Protein Conformation on Experimental Bandwidths in RPHPLC', *J. Chromatogr.*, **676**, 33–41 (1994).
5. P. Oroszlan, S. Wicar, G. Teshima, S.-L. Wu, W.S. Hancock, B.L. Karger, 'Conformational Effects in the Reversed-phase Chromatographic Behavior of Recombinant Human Growth Hormone (rhGH) and *N*-Methionyl Recombinant Human Growth Hormone (met-hGH)', *Anal. Chem.*, **64**, 1623–1631 (1992).
6. S. Lin, B.L. Karger, 'Reversed-phase Chromatographic Behavior of Proteins in Different Unfolded States', *J. Chromatogr.*, **499**, 89–102 (1990).
7. E. Lazoura, J. Maidonis, E. Bayer, M.T.W. Hearn, M.I. Aguilar, 'The Conformational Analysis of NPY-[18–36] Analogs at Hydrophobic Surfaces', *Biophys. J.*, **72**, 238–246 (1997).
8. T.-Z. Lee, P.T. Thompson, M.T.W. Hearn, M.I. Aguilar, 'Conformational Stability of a Type II′ β-Turn Motif in Human Growth Hormone [6–13] Peptide Analogs at Hydrophobic Surfaces', *J. Pept. Res.*, **49**, 394–403 (1997).
9. A.W. Purcell, M.I. Aguilar, R.E.H. Wettenhall, M.T.W. Hearn, 'The Induction of Amphipathic Helical Structures in Peptides During RPHPLC', *Pept. Res.*, **8**, 160–170 (1995).
10. S.E. Blondelle, R.A. Houghten, 'Design of Model Peptides Having Potent Antimicrobial Activities', *Biochemistry*, **31**, 12 688–12 694 (1992).
11. S.E. Blondelle, K. Buttner, R.A. Houghten, 'Evaluation of Peptide–Peptide Interactions Using Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **625**, 199–206 (1992).
12. D.L. Steer, P.E. Thompson, S.E. Blondell, R.A. Houghten, M.I. Aguilar, 'Comparison of the Binding of α-Helical and β-Sheet Peptides to a Hydrophobic Surface', *J. Pept. Res.*, **51**, 401–412 (1998).
13. N.E. Zhou, C.T. Mant, R.S. Hodges, 'Effect of Preferred Binding Domains on Retention Behavior in Reversed-phase Chromatography', *Pept. Res.*, **3**, 8–20 (1990).
14. R.S. Hodges, B.-Y. Zhu, N.E. Zhou, C.T. Mant, 'Reversed-phase Liquid Chromatography as a Useful Probe of Hydrophobic interactions involved in Protein Folding and Protein Stability', *J. Chromatogr.*, **676**, 3–15 (1994).
15. E. Krause, M. Beyermann, M. Dathe, S. Rothemund, M. Bienert, 'Location of an Amphipathic α-Helix in Peptides Using Reversed-phase HPLC Retention Behavior of D-Amino Acid Analogs', *Anal. Chem.*, **67**, 252–258 (1995).
16. *Methods Enzymol.*, **270**, (1996), entire volume.
17. *Methods Enzymol.*, **271**, (1996), entire volume.
18. C.T. Mant, R.S. Hodges (eds.), *High-performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, CRC Press, Boca Raton, FL, 1991.
19. M.T.W. Hearn (ed.), *HPLC of Proteins, Peptides and Polynucleotides: Contemporary Topics and Applications*, VCH, Deerfield Beach, FL, 1991.
20. K.M. Gooding, F.E. Regnier (eds.), *HPLC of Biological Macromolecules: Methods and Applications*, Marcel Dekker, New York, 1990.
21. C. Miller, J. Rivier, 'Peptide Chemistry: Development of High-performance Liquid Chromatography and Capillary Zone Electrophoresis', *Biopolymers*, **40**, 265–318 (1996).
22. E.R. Hoff, R.C. Chloupek, 'Analytical Peptide Mapping of Recombinant DNA-derived Proteins by Reversed-phase High-performance Liquid Chromatography', *Methods Enzymol.*, **271**, 51–67 (1996).
23. E. Nice, 'Micropreparative High-performance Liquid Chromatography of Proteins and Peptides: Principles and Applications', *Biopolymers*, **40**, 319–341 (1996).
24. C.V. Olson, D.H. Reifsnyder, E. Canova-Davis, V.T. Ling, S.E. Builder, 'Preparative Isolation of Recombinant Human Insulin-like Growth Factor 1 by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. A*, **675**, 101–112 (1994).
25. E.C. Nice, L. Fabri, A. Hammacher, K. Andersson, U. Hellman, 'Micropreparative Ion-exchange HPLC. Applications to Microsequence Analysis', *Biomed. Chromatogr.*, **7**, 104–111 (1993).
26. 'Efficient Isolation of Protein Fragments for Structure Analysis', ÄKTA™ Purifier, Application Note No. 18-1119-53, Pharmacia Biotech, Uppsala, 1996.

27. C. Horvath, W. Melander, I. Molnar, 'Solvophobic Interactions in Liquid Chromatography with Non-polar Stationary Phases', *J. Chromatogr.*, **125**, 129–156 (1976).
28. L. Snyder, 'Gradient Elution', in *HPLC – Advances and Perspectives*, ed. C.S. Horvath, Academic Press, New York, 207–316, Vol. 1, 1980.
29. M.A. Stadalius, H.S. Gold, L.R. Snyder, 'Optimization Model for the Gradient Elution Separation of Peptide Mixtures by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **296**, 31–59 (1984).
30. L.R. Snyder, 'Automated Method Development in High-performance Liquid Chromatography', *Methods Enzymol.*, **270**, 151–175 (1996).
31. M.I. Aguilar, A.N. Hodder, M.T.W. Hearn, 'Studies on the Optimization of the Reversed-phase Gradient Elution of Polypeptides: Evaluation of Retention Relationships with β-Endorphin-related Polypeptides', *J. Chromatogr.*, **327**, 115–138 (1985).
32. D. Guo, C.T. Mant, A.K. Taneja, R.S. Hodges, 'Prediction of Peptide Retention Times in Reversed-phase High-performance Liquid Chromatography. II. Correlation of Observed and Predicted Peptide Retention Times and Factors Influencing the Retention Times of Peptides', *J. Chromatogr.*, **359**, 519–532 (1986).
33. M.C.J. Wilce, M.I. Aguilar, M.T.W. Hearn, 'The Physicochemical Basis of Hydrophobicity Scales: Evaluation of Four New Sets of Amino Acid Coefficients Derived from Reversed-phase High-performance Liquid Chromatography', *Anal. Chem.*, **67**, 1210–1219 (1995).
34. F.E. Regnier, 'The Role of Protein Structure in Chromatographic Behavior', *Science*, **238**, 319–323 (1987).
35. K.L. Richards, M.I. Aguilar, M.T.W. Hearn, 'A Comparative Study of the Retention Behavior and Stability of Cytochrome *c* in RPHPLC', *J. Chromatogr.*, **676**, 17–31 (1994).
36. A.W. Purcell, M.I. Aguilar, M.T.W. Hearn, 'Conformational Effects in the RPHPLC of Polypeptides. I: The Resolution of Insulin Variants', *J. Chromatogr.*, **711**, 61–70 (1995).
37. K.K. Unger, B. Anspach, R. Janzen, G. Jilge, K.D. Lork, 'Bonded Silica Phases for the Separation of Biopolymers by Means of Column Liquid Chromatography', in *HPLC – Advances and Perspectives*, ed. C.S. Horvath, Academic Press, New York, 2–93, Vol. 5, 1988.
38. K.K. Unger, 'Silica as a Support', in *HPLC of Biological Macromolecules: Methods and Applications*, eds. K.M. Gooding, F.E. Regnier, Marcel Dekker, New York, 3–24, 1990.
39. M. Henry, 'Design Requirements of Silica-based Matrices for Biopolymer Chromatography', *J. Chromatogr.*, **544**, 413–443 (1991).
40. N.E. Zhou, C.T. Mant, J.J. Kirkland, R.S. Hodges, 'Comparison of Silica-based Cyanopropyl and Octyl Reversed-phase Packings for the Separation of Peptides and Proteins', *J. Chromatogr.*, **548**, 179–193 (1991).
41. I. Yarovsky, M.I. Aguilar, M.T.W. Hearn, 'Influence of the Chain Length and Surface Density on the Conformation and Mobility of *n*-Alkyl Ligands Chemically Immobilised to a Silica Surface', *Anal. Chem.*, **67**, 2145–2153 (1995).
42. I. Yarovsky, M.T.W. Hearn, M.I. Aguilar, 'Molecular Simulation of Peptide Interactions with an RPHPLC Sorbent', *J. Phys. Chem.*, **101**, 10 962–10 970 (1997).
43. K. Albert, E. Bayer, 'Characterization of Bonded Phases by Solid-state NMR Spectroscopy', *J. Chromatogr.*, **544**, 345–370 (1991).
44. L.A. Cole, J.G. Dorsey, 'Temperature Dependence of Retention in Reversed-phase Liquid Chromatography 1. Stationary Phase Considerations', *Anal. Chem.*, **64**, 1317–1323 (1992).
45. D.M. Bleisner, K.B. Sentell, 'Deuterium Nuclear Magnetic Resonance Spectroscopy as a Probe for Reversed-phase Liquid Chromatographic Bonded Phase Solvation; Methanol and Acetonitrile Mobile Phase Components', *J. Chromatogr.*, **631**, 23–35 (1993).
46. F.J. Moy, Y.-C. Li, P. Rauenbeuhler, M.E. Winkler, H.A. Scheraga, G.T. Montelione, 'Solution Structure of Human Type-α Transforming Growth Factor Determined by Heteronuclear NMR Spectroscopy and Refined by Energy Minimization with Restraints', *Biochemistry*, **32**, 7334–7353 (1993).
47. M.A. Chlenov, E.I. Kandyba, L.V. Nagornaya, I.L. Orlova, Y.V. Volgin, 'High-performance Liquid Chromatography of Human Glycoprotein Hormones', *J. Chromatogr.*, **631**, 261–267 (1993).
48. B.S. Welinder, H.H. Sorenson, B. Hansen, 'Reversed-phase High-performance Liquid Chromatography of Insulin. Resolution and Recovery in Relation to Column Geometry and Buffer Components', *J. Chromatogr.*, **361**, 357–363 (1986).
49. M. Hanson, K.K. Unger, C.T. Mant, R.S. Hodges, 'Polymer-coated Reversed-phase Packings with Controlled Hydrophobic Properties. I. Effect on the Selectivity of Protein Separations', *J. Chromatogr.*, **599**, 65–76 (1992).
50. M. Hanson, K.K. Unger, C.T. Mant, R.S. Hodges, Polymer-coated Reversed-phase Packings with Controlled Hydrophobic Properties. II. Effect on the Selectivity of Peptide Separations', *J. Chromatogr.*, **599**, 77–86 (1992).
51. N. Tanaka, K. Kimata, Y. Mikawa, K. Hosoya, T. Araki, Y. Ohtsu, Y. Shiojima, R. Tsuboi, H. Tsuchiya, 'Performance of Wide-pore Silica- and Polymer-based Packing Materials in Polypeptide Separation: Effect of Pore Size and Alkyl Chain Length', *J. Chromatogr.*, **535**, 13–32 (1990).
52. B.S. Welinder, 'Use of Polymeric Reversed-phase Columns for the Characterization of Polypeptides Extracted from Human Pancreata. II. Effect of the Stationary Phase', *J. Chromatogr.*, **542**, 83–99 (1991).
53. H.-J. Wirth, K.-O. Eriksson, P. Holt, M.I. Aguilar, M.T.W. Hearn, 'Ceramic Based Particles as Chemically

Stable Chromatographic Supports', *J. Chromatogr.*, **646**, 129–141 (1993).
54. L. Sun, P.W. Carr, 'Chromatography of Proteins Using Polybutadiene-coated Zirconia', *Anal. Chem.*, **67**, 3717–3721 (1995).
55. G. Jilge, R. Janzen, K.K. Unger, K.N. Kinkel, M.T.W. Hearn, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by High-performance Liquid Chromatography. III. Retention and Selectivity of Proteins and Peptides in Gradient Elution on Non-porous Monodisperse 1.5 μm Reversed-phase Silicas', *J. Chromatogr.*, **397**, 71–80 (1987).
56. N.B. Afeyan, N.F. Gordon, I. Mazsaroff, L. Varady, S.P. Fulton, Y.B. Yang, F.E. Regnier, 'Flow-through Particles for the High-performance Liquid Chromatographic Separations of Biomolecules: Perfusion Chromatography', *J. Chromatogr.*, **519**, 1–29 (1990).
57. S.K. Paliwal, M. deFrutos, F.E. Regnier, 'Rapid Separations of Proteins by Liquid Chromatography', *Methods Enzymol.*, **270**, 133–151 (1996).
58. P.M. Young, T.E. Wheat, 'Optimization of High-performance Liquid Chromatographic Peptide Separations with Alternative Mobile and Stationary Phases', *J. Chromatogr.*, **512**, 273–28 (1990).
59. D.J. Poll, D.R.K. Harding, 'Formic acid as a Milder Alternative to Trifluoroacetic Acid and Phosphoric Acid in Two-dimensional Peptide Mapping', *J. Chromatogr.*, **469**, 231–239 (1989).
60. G. Thevenon, F.E. Regnier, 'Reversed-phase Liquid Chromatography of Proteins with Strong Acids', *J. Chromatogr.*, **476**, 499–511 (1989).
61. C.T. Mant, R.S. Hodges, 'The Effects of Anionic Ion-pairing Reagents on Peptide Retention in Reversed-phase Chromatography', in *High-performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*, eds. C.T. Mant, R.S. Hodges, CRC Press, Boca Raton, FL, 327–341, 1991.
62. G.W. Welling, R. Van der Zee, S. Welling-Wester, 'Column Liquid Chromatography of Integral Membrane Proteins', *J. Chromatogr.*, **418**, 223–243 (1987).
63. M.I. Aguilar, S. Mougos, J. Boublik, J. Rivier, M.T.W. Hearn, 'The Effect of D-Amino Acid Substitutions on the RPHPLC Behavior of Neuropeptide-Y [18–36] Analogs', *J. Chromatogr.*, **646**, 53–65 (1993).
64. R.L. Moritz, R.J. Simpson, 'Application of Capillary Reversed-phase High-performance Liquid Chromatography to High Sensitivity Protein Sequence Analysis', *J. Chromatogr.*, **599**, 119–130 (1992).
65. R.L. Moritz, R.J. Simpson, 'Capillary Liquid Chromatography: A Tool for Protein Structural Analysis', in *Methods in Protein Sequence Analysis*, eds. K. Imahori, F. Sakiyama, Plenum Press, New York, 3–10, 1993.
66. B.F.D. Ghrist, L.R. Snyder, B.S. Cooperman, 'HPLC of Ribosomal Proteins', in *HPLC of Biological Macromolecules: Methods and Applications*, eds. K.M. Gooding, F.E. Regnier, Marcel Dekker, New York, 403–427, 1990.
67. R.J. Simpson, R.L. Moritz, E.C. Nice, B. Grego, 'A High-performance Liquid Chromatography Procedure for Recovering Subnanomole Amounts of Protein from SDS–Gel Electroeluates for Gas-phase Sequence Analysis', *Eur. J. Biochem.*, **165**, 21–29 (1987).
68. J. Frank, A. Braat, J.A. Duine, 'Assessment of Protein Purity by Chromatography and Multiwavelength Detection', *Anal. Biochem.*, **162**, 65–73 (1987).
69. F. Nyberg, C. Pernow, U. Moberg, R.B. Eriksson, 'High-performance Liquid Chromatography and Diode Array Detection for the Identification of Peptides Containing Aromatic Amino Acids in Studies of Endorphin-degrading Activity in Human Cerebrospinal Fluid', *J. Chromatogr.*, **359**, 541–551 (1986).
70. G.P. Rozing, 'Diode Array Detection', *Methods Enzymol.*, **270**, 201–234 (1996).
71. X.M. Lu, K. Benedek, B.L. Karger, 'Conformational Effects in the High-performance Liquid Chromatography of Proteins. Further Studies of the Reversed-phase Chromatographic Behavior of Ribonuclease A', *J. Chromatogr.*, **359**, 19–29 (1986).
72. M. Fridman, M.I. Aguilar, M.T.W. Hearn, 'A Comparative Study of the Equilibrium Refolding of Bovine, Porcine and Human Growth Hormone Using Size Exclusion Chromatography', *J. Chromatogr.*, **512**, 57–75 (1990).
73. S.A. Carr, M.E. Hemling, M.F. Bean, G.D. Roberts, 'Integration of Mass Spectrometry in Analytical Biotechnology', *Anal. Chem.*, **63**, 2802–2824 (1991).
74. A. Apffel, J. Chakel, S. Udiavar, W.S. Hancock, C. Souders, E. Pungor, Jr, 'Application of Capillary Electrophoresis, High-performance Liquid Chromatography, On-line Electrospray Mass Spectrometry and Matrix-assisted Laser Desorption Ionization–Time-of-flight Mass Spectrometry to the Characterization of Single-chain Plasminogen Activator', *J. Chromatogr. A*, **717**, 41–60 (1995).
75. S. Renlund, H. Wadensten, P. Persson, 'Studies of Different Forms of Apolipoprotein A1 and Insulin-like Growth Factor 1 by Electrospray LC/MS', in *Peptide Chemistry 1995*, ed. N. Nishi, Protein Research Foundation, Osaka, 69–72, 1996.
76. J. Ding, W. Burkhart, D.B. Kassel, 'Identification of Phosphorylated Peptides from Complex Mixtures Using Negative-ion Orifice-potential Stepping and Capillary Liquid Chromatography/Electrospray Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 94–98 (1994).
77. Z. Randhawa, H.E. Witkowska, J. Cone, J.A. Wilkins, P. Hughes, K. Yamanishi, S. Yashuda, Y. Masui, P. Arthur, C. Kletke, F. Bitsch, C.H.L. Shackleton, 'Incorporation of Norleucine at Methionine Positions in Recombinant Human Macrophage Colony Stimulating Factor (m-CSF,4-153)', *Biochemistry*, **33**, 4352–4362 (1994).
78. M.J. Huddleston, M.F. Bean, S.A. Carr, 'Collisional Fragmentation of Glycopeptides by Electrospray Ionization

LC/MS and LC/MS/MS: Methods for Selective Detection of Glycopeptides in Protein Digests', *Anal. Chem.*, **65**, 877–884 (1993).

79. S.A. Carr, M.J. Huddleston, M.F. Bean, 'Selective Identification and Differentiation of N- and O-linked Oligosaccharides in Glycoproteins by Liquid Chromatography–Mass Spectrometry', *Protein Sci.*, **2**, 183–196 (1993).
80. K.L. Duffin, J.K. Welply, J.D. Henion, 'Characterization of N-linked Oligosaccharides by Electrospray and Tandem Mass Spectrometry', *Anal. Chem.*, **64**, 1440–1448 (1992).
81. K. Mock, M. Hail, I. Mylchreest, J. Zhou, K. Johnson, I. Jardine, 'Rapid High Sensitivity Peptide Mapping by Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **646**, 169–174 (1993).
82. D.L. Smith, Z.R. Zhou, 'Strategies for Locating Disulfide Bonds in Proteins', *Methods Enzymol.*, **193**, 374–389 (1990).
83. D.F. Hunt, R.A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A.L. Cox, E. Appella, V.H. Engelhard, 'Characterization of Peptides Bound to the Class I MHC Molecule HLA-A2.1 by Mass Spectrometry', *Science*, **255**, 1261–1263 (1992).

Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

Zdzislaw Salamon and Gordon Tollin
University of Arizona, Tucson, USA

1 Introduction **1**
1.1 General Overview of the Surface Plasmon Resonance Spectroscopy Technique 1
1.2 Use of Coupled Plasmon-waveguide Resonators to Measure Anisotropic Properties of Thin Films 3
1.3 Advantages and Limitations of Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis 6
2 Sensitivity of the Surface Plasmon Resonance Method **6**
3 Immobilization of Proteins on Plasmon-generating Solid Surfaces **7**
3.1 Chemical Modifications of the Surface 7
3.2 Self-assembled Solid-supported Lipid Membranes as Biocompatible Matrices 8
4 Biological Applications of Surface Plasmon Resonance Spectroscopy **8**
4.1 Structural Parameters of Biomolecular Films 9
4.2 Evaluation of Binding Affinities 9
4.3 Determination of Rate Constants for Binding Reactions 9
Acknowledgment **10**
Abbreviations and Acronyms **10**
Related Articles **10**
References **10**

This chapter presents a detailed overview of the theoretical basis of surface plasmon resonance (SPR) spectroscopy, emphasizing both principles and practical considerations such as thin-film designs for SPR devices which can be employed in characterization of biological materials, including peptides and proteins. It is demonstrated that the analysis of the resonance spectrum yields structural parameters of anisotropic films of interacting biomolecules, including distribution of mass, long-range molecular order, and molecular conformational changes. Binding parameters, including binding affinities and rate constants, can also be determined. A summary is presented of different techniques for biomolecule immobilization on SPR-generating surfaces applicable in peptide and protein analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 General Overview of the Surface Plasmon Resonance Spectroscopy Technique

The concept of surface plasmons (SPs) derives from the plasma formulation of Maxwell's theory of electromagnetism, where the free electrons of a metal (or a conductive electron gas) are treated as a high-density liquid (plasma). Plasma oscillations in metals are collective longitudinal excitations of the conductive electron gas, and plasmons are the quanta representing these charge-density oscillations. When such oscillations occur on an interface between a metallic and a dielectric surface they are called SPs, or surface polaritons. These propagating electron-density fluctuations generate evanescent SP electromagnetic waves (or guided surface electromagnetic waves), which propagate along the plane metal/dielectric media interface, with the electric field normal to this interface.[1] It has been shown that Maxwell's equations have solutions resulting in the generation of SP electromagnetic waves only when the frequency, w, and wave vector, κ, meet special requirements which produce the following constraints:[1,2]

1. one of the adjacent media (i.e. the *surface-active medium* which generates SP waves) has a negative value for the real part of its complex dielectric constant, ε;
2. a component of the wave vector along the interface between these two media satisfies an equation which involves the dielectric constants of both media (see below).

The complex dielectric constant, ε, is directly related to the complex index of refraction $N = c/v = n - ik$, by Equation (1):

$$\varepsilon = \varepsilon' + i\varepsilon'' = N^2 \tag{1}$$

where the real part of the complex dielectric constant is $\varepsilon' = n^2 - k^2$, and the imaginary part is $\varepsilon'' = 2nk$, n is the refractive index, k is the extinction coefficient, and v is the velocity of light in the dielectric medium. In the optical region of the electromagnetic spectrum, the

complex index of refraction is numerically equivalent to the optical admittance, Y, which is defined by the ratio of the amplitudes of the electric and magnetic fields of the electromagnetic wave.[3]

The condition that the surface-active medium has a negative dielectric constant results in several experimental requirements which have to be fulfilled in order to be able to generate SP waves. First of all, not all materials can be used as surface-active media; gold (Au), and silver (Ag) are the best examples of materials that can support SPs. In addition, the electromagnetic wave in the surface-active medium is an evanescent wave under all circumstances. Therefore, SPs can only be optically generated by an evanescent wave excited under total internal reflection conditions (obtained with either prisms or gratings), by a transverse magnetic (TM) component, p, of incident light of wavelength, λ (or frequency, w), whose component wave vector parallel to the metal interface, κ_{ph}, which is described by Equation (2):

$$\kappa_{ph} = \frac{w}{c} \varepsilon_0^{1/2} \sin\alpha \tag{2}$$

matches that of the evanescent SP electromagnetic wave κ_{SP} (Equation 3):

$$\kappa_{ph} = \kappa_{SP} \tag{3}$$

The latter is given by Equation (4):

$$\kappa_{SP} = \frac{w}{c\,[\varepsilon_1\varepsilon_2/(\varepsilon_1+\varepsilon_2)]^{1/2}} \tag{4}$$

where w is the frequency of the SP wave (and excitation light wavelength), c is the velocity of light in vacuo, and ε_1 and ε_2 are the complex dielectric constants for the surface active and dielectric (emerging) media, respectively.[2] At the resonance condition (Equation 3), the incident light is coupled into a plasmon wave travelling along and bound to the outer active (metal) surface, and the phenomenon is known as SPR.

As Equation (3), taken together with Equations (2) and (4), indicates, the resonance condition can be achieved by changing either the incident angle, α, with constant value of plasmon excitation light frequency, w (or wavelength, λ), or by changing w (or λ) at a constant value of the incident light angle, α. During the resonance interaction, energy is transferred from photons to plasmons, so that the effect of plasmon excitation can be observed as a sharp minimum of the reflectance when either angle of incidence is varied at the same light wavelength, or light wavelength is varied at the same incident angle, thus defining an SPR spectrum. In both instances this spectrum reflects the resonance in absorption of incident photons.[4]

SPR can also be detected with a fluorescence technique known as total internal reflectance fluorescence (TIRF), employed with waveguided systems. The application of TIRF to monitor SPR is based on the following properties of the surface modes. First, there is a possibility that the nonradiative SP modes can, under specific conditions, decay into light, thereby allowing emission techniques to be used to detect the resonance.[5] Nonradiative SPs can decay into photons of the same frequency if coupling by the surface roughness takes place. The intensity of emitted light images the SPR occurring at silver and gold surfaces, producing an emission resonance curve similar to that of the reflectance resonance curve obtained under attenuated total reflection (ATR) conditions. Second, the surface bound electric field generated by SP modes can be used as an efficient source to excite fluorescence emission of molecules adsorbed at the SPR active surface. This property of the plasmon electromagnetic field allows the monitoring of resonance by using fluorescent labeling of molecules adsorbed (or immobilized) on the active surface of the SPR-producing medium.

In practical terms there are two configurations, both based on the ATR technique, available to optically excite and to monitor SPR at the metal/dielectric (or emerging medium) interface. In the first, the Kretschmann configuration,[6] the prism is in direct contact with the surface-active (metal) medium. In the second, the Otto configuration,[7] the prism is separated by a thin layer of a dielectric (inactive) medium at a distance of approximately one wavelength of excitation light from the metal film.

There are three principal designs of resonators resulting in four different types of resonances:

(i) Conventional Surface Plasmon Resonance. This can be generated by a prism whose hypotenuse is coated with a single high-performance metal (Ag or Au) layer.[4,8–16] Such a device produces a resonance that is characterized by a very large increase of electromagnetic field amplitude of the evanescent wave bound to the outer surface of the metal,[4] compared with that of the incident wavelength. It has the characteristics of a sharp resonance and can only be obtained with p-polarized excitation light. Figure 1 compares the calculated resonances observed with reflected light intensity for both metals and both variables i.e. either as a function of α with $\lambda = \text{const}$ (panel a), or vs λ at $\alpha = \text{const}$ (panel b).

(ii) Long-range Surface Plasmon Resonance. This is generated in the same way as conventional SPR, but with a thinner metal layer, which is surrounded by dielectric media that are beyond the critical angle so that they support evanescent waves.[3] Figure 2 demonstrates an example of calculated LRSPR reflectance vs incident angle (panel a) and LRSPR reflectance as a function of light wavelength (panel b) spectra. As can be seen these

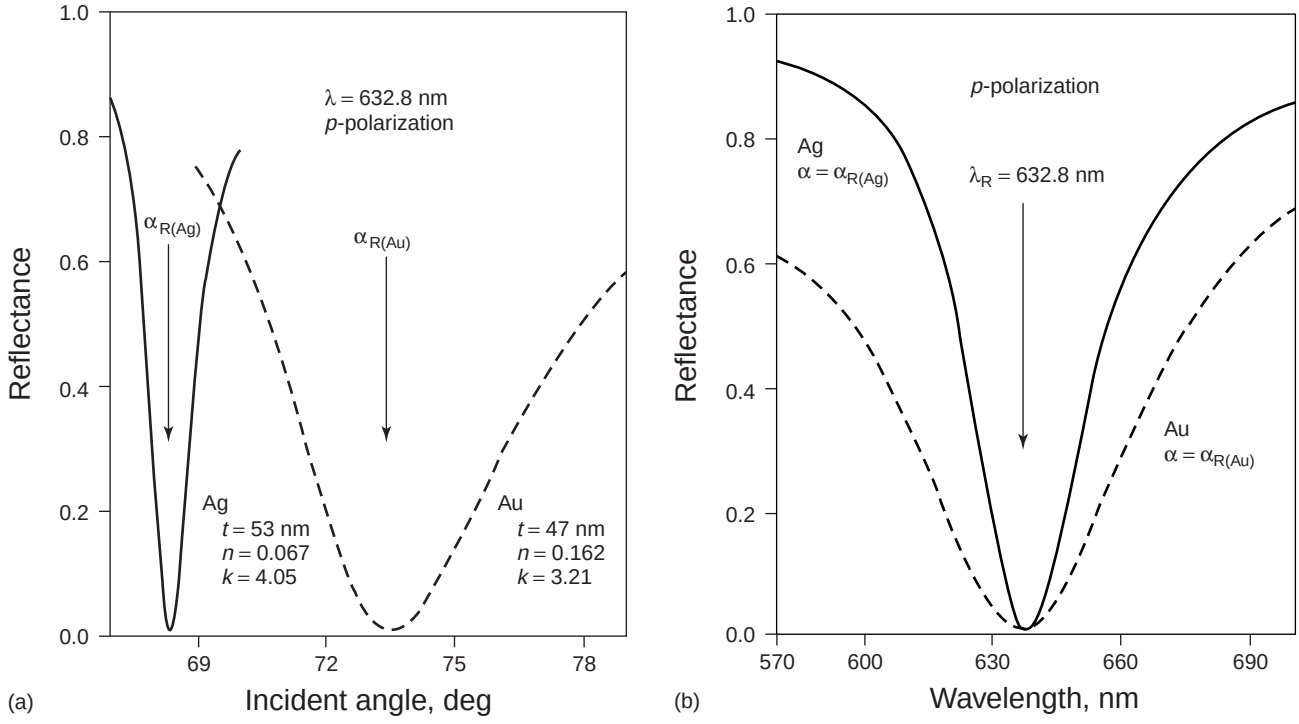

Figure 1 Conventional SPR spectra of the two most frequently used metal films: silver (solid line; Ag), and gold (dashed line; Au), with the indicated optical parameters, presented as reflectance vs either the incident angle, α, (panel a), obtained with *p*-polarized (TM) light of constant wavelength, ($\lambda = 632.8$ nm), or light wavelength (panel b) obtained at a constant value of $\alpha = \alpha_R$. $\alpha_{R(Ag)}$ and $\alpha_{R(Au)}$ are the incident angles at which resonance excited with $\lambda = 632.8$ nm achieves its maximum (i.e. minimum intensity of reflected light), for Ag and Au, respectively, as indicated in panel a. The calculation has been made assuming a glass prism as an incident medium and water as an emerging medium.

are much narrower resonances than conventional SPR (Figure 1), which indicates a much higher evanescent electromagnetic field intensity generated by LRSPR. Furthermore, LRSPR involves two surface-bound waves on both the inner and outer surfaces of the metal layer compared with one wave in the conventional resonance.

(iii) Coupled Plasmon-waveguide Resonance. This involves even more complex assemblies in which SPRs in a thin metal film are coupled with guided waves in a dielectric overcoating, resulting in excitation of both plasmon and waveguide resonances, a combination known as coupled plasmon-waveguide resonance (CPWR).[17] A coupled plasmon-waveguide resonator contains a metallic layer (the same as in a conventional SPR assembly), which is deposited on either a prism or a grating and is overcoated with either a single dielectric layer or a system of dielectric layers, characterized by appropriate optical parameters so that the assembly is able to generate surface resonances upon excitation by both *p*- and *s*-polarized light components. As can be seen from Figure 3, the addition of such a dielectric layer (or layers) to a conventional SPR assembly plays several important roles. First, it functions as an optical amplifier which significantly increases electromagnetic field intensities at the dielectric surface in comparison to conventional SPR. This results in an increased sensitivity and spectral resolution (the latter due to decreased resonance linewidths). Second, it enhances spectroscopic capabilities (due to excitation of resonances with both *p*- and *s*-polarized light components), which results in the ability to directly measure anisotropies in refractive index and optical absorption coefficient in a thin film adsorbed onto the surface of the overcoating. Third, the dielectric overcoating also serves as a mechanical and chemical shield for the thin metal layer in practical applications.

(iv) Long-range Coupled Plasmon-waveguide Resonance. This is generated by a resonator which combines both the long-range SP and coupled plasmon-waveguide resonators into one device. The resulting resonance spectra are similar in shape and intensity to those obtained with CPWR devices.

1.2 Use of Coupled Plasmon-waveguide Resonators to Measure Anisotropic Properties of Thin Films

The standard mathematical tool used to describe the SPR phenomenon, as well as the optical properties of

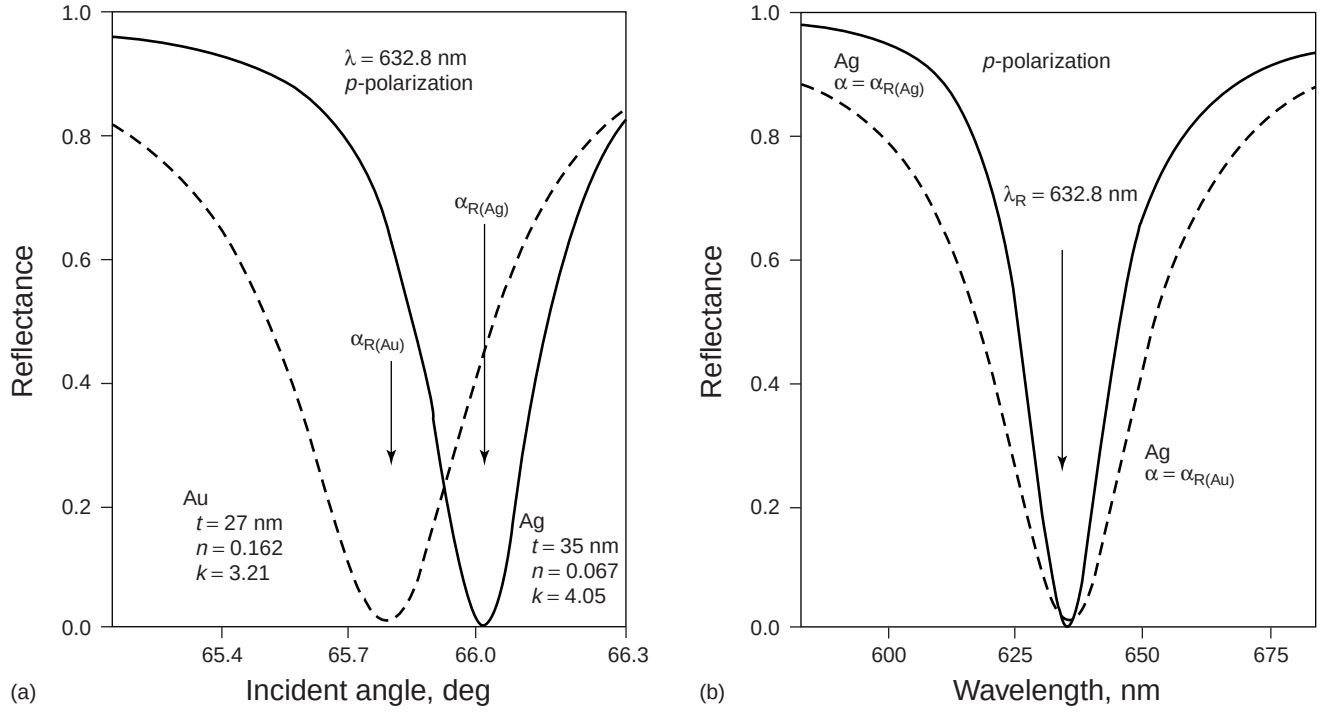

Figure 2 Calculated LRSPR spectra of silver (solid line), and gold (dashed line), with indicated optical parameters, presented as reflectance vs either incident angle (panel a), obtained with *p*-polarized excitation light and a glass prism coated with a MgF_2 layer (thickness = 950 nm, and refractive index = 1.379 at light wavelength = 632.8 nm), on top of which either 35-nm-thick silver, or 27-nm-thick gold has been deposited, or light wavelength (panel b), obtained at $\alpha = \alpha_R$. $\alpha_{R(Ag)}$ and $\alpha_{R(Au)}$ are the incident angles at which the resonance excited with $\lambda = 632.8$ nm reaches its maximum for silver and gold layers, respectively, as indicated in panel a.

multilayered thin-film devices such as the SPR resonators, can be seen as a straightforward result of the application of thin-film electromagnetic theory.[3] In the case of SPR which is generated optically with the ATR technique, the reflectance, R, of a multilayer system, defined as the ratio of the energy reflected at the surface of such a structure to the energy which is incident, depends not only on the optical parameters of the sensing layer, but also on the optical parameters of both incident and emerging media as well as the metal film (or metal/dielectric layers in CPWR), generating the resonance. It can be calculated from Equations (5a, 5b, 5c):

$$R = rr^* \tag{5a}$$

where:

$$r = \frac{Y_0 - Y}{Y_0 + Y} \tag{5b}$$

is the amplitude reflection coefficient (reflectivity or Fresnel reflection coefficient), and:

$$r^* = \left\{ \frac{Y_0 - Y}{Y_0 + Y} \right\}^* \tag{5c}$$

is the complex conjugate of r. Y is the optical admittance of a multilayer-thin-film assembly and can be calculated with the help of the characteristic matrix of such an assembly.[3] Y_0 is the admittance of the incident medium (which in the case of the present application to SPR is a non-light-absorbing glass prism, i.e. with $k = 0$, and therefore the Y_0 value becomes real and equal to the refractive index of the incident medium, n_0). Equations (5a), (5b), and (5c) comprise a set of mathematical tools to examine optically excited SPR by analyzing the SPR spectrum represented by the reflectance, R, as a function of either the incident angle, α, or the wavelength, λ. Such an analysis can be applied to all three types of resonance: SPR,[4,18–21] LRSPR, and CPWR,[17] and results in three optical parameters (refractive index, n, extinction coefficient, k, and thickness, t) which characterize the properties of a dielectric layer. These three optical parameters characterizing a deposited dielectric film (including protein, lipid, and lipid–protein layers) can be evaluated in the case of CPWR for both exciting light polarizations, at different angles of light incidence, and using different light wavelengths, thereby producing enough experimental data to characterize all of the structural parameters of thin films, i.e. thickness, orientation of molecules (by measuring the anisotropy in n), and the orientation of chromophores attached to the molecules within the sensing layer (by measuring the anisotropy

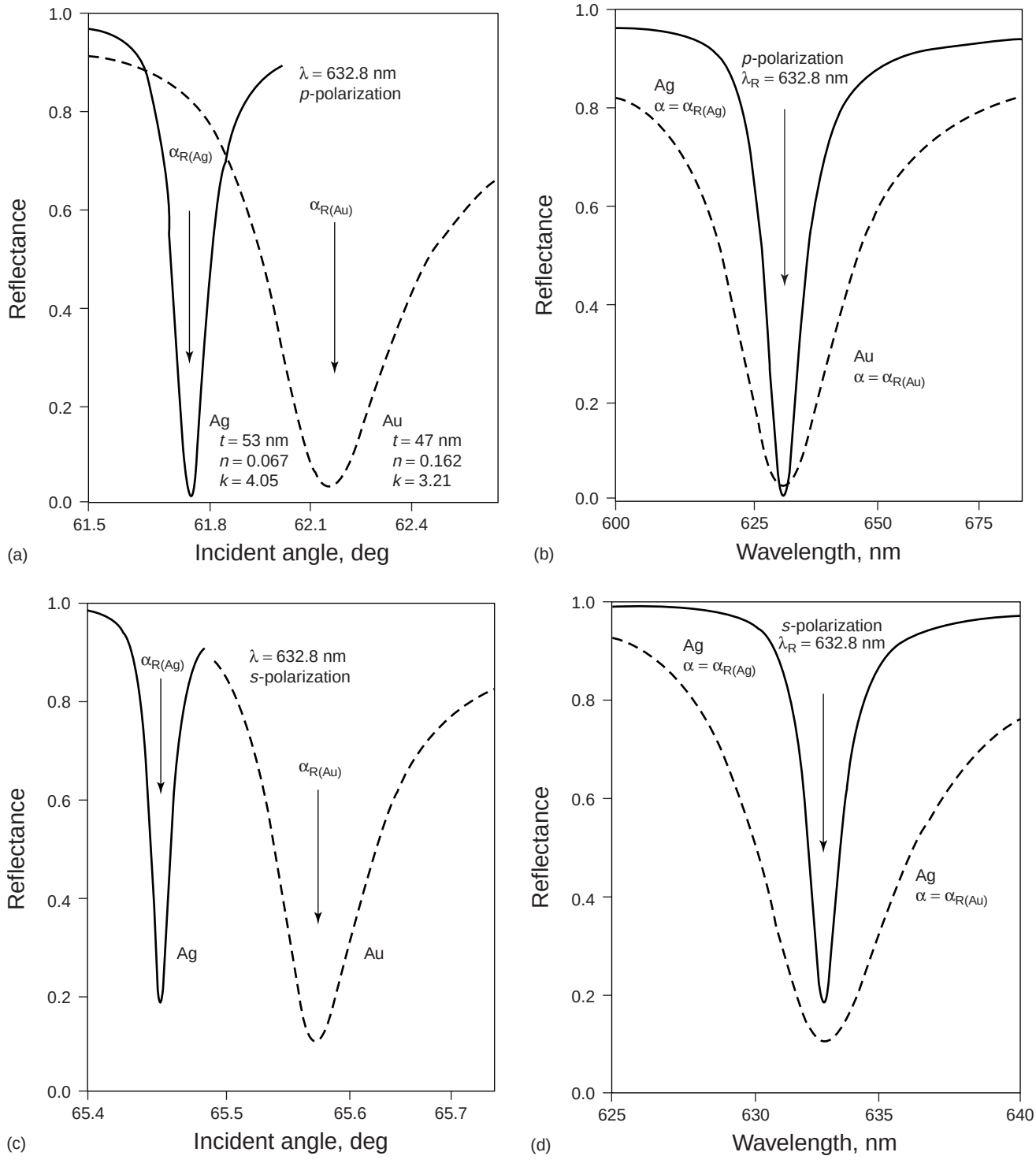

Figure 3 CPWR spectra obtained with either *p*-polarized (TM; panels a and b), or *s*-polarized (transverse electric (TE); panels c and d) excitation light ($\lambda = 632.8$ nm), and a glass prism coated with either silver or gold layer with the optical parameters as indicated in Figure 1, overcoated with a 450-nm SiO_2 film. The emerging medium is water. The resonance spectra of silver (solid line), and gold (dashed line), are presented as reflectance versus either the incident angle (panels a and c), with constant value of light wavelength ($\lambda = 632.8$ nm), or the light wavelength (panels b and d), at constant value of incident angle ($\alpha = \alpha_R$, as indicated in panels a and c, are the incident angles at which resonance excited with light wavelength = 632.8 nm reaches its maximum, for silver and gold).

of k). In addition, the mass of the deposited layer can also be calculated, which allows binding isotherms to be measured.[17] All of these characterizations can be obtained using a single device covered with a sensing layer, and using a measurement method that involves only a determination of reflected light intensity under total internal reflection conditions. Furthermore, because the electromagnetic field decays exponentially within the emerging medium, the measurement is sensitive only to the interface region between the dielectric overcoat and the emerging medium, and is not interfered with by the bulk properties of the medium. In addition, there is no limitation on the dielectric material that can be used in such coatings (as long as the optical characteristics are favorable), and therefore the dielectric film can be formed from any number of layers designed and optimized for different uses.[17]

1.3 Advantages and Limitations of Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

As described above, SPs create a surface-bound evanescent electromagnetic wave which propagates along the surface of an active medium (usually a thin metallic film), with the electric field intensity maximized at this surface and diminishing exponentially on both sides of the interface. As a consequence of this property, the phenomenon has been utilized extensively in studies of surfaces and of thin dielectric films, including lipid, protein, and lipid–protein layers, deposited on the active medium. The SPR technique is able to probe both kinetic and thermodynamic processes, as well as to provide microstructural information, which makes it a very important component of the experimental methodology available to probe molecular interactions occurring at surfaces.[4,14,15,17,22] Further, it allows some of the limitations of other techniques to be overcome. For example, other methods often require one of the partners to be labeled in some way in order to allow it to be detected. Fluorescent probes, radioactive labels, and attachment of independently detectable molecules (e.g. enzymes) have all been used for this purpose. These suffer from the drawback that they may interfere with the binding of the labeled partner to the unlabeled one, or cause unwanted structural perturbations. SPR observations are based solely on the dielectric properties of molecules, or their intrinsic light absorption characteristics, and thus require no specific labeling. The more versatile CPWR technique offers a relatively simple optical method to obtain information about both mass (i.e. binding affinity) and structural parameters, including anisotropy in lipid and proteolipid films, overcoming the problems and challenges related to detecting small amounts of material located at the interface between two immiscible phases, which may be labile and structurally heterogeneous.[4,15,17]

Although numerous other optical techniques have also been applied to such systems (e.g. ellipsometry, interferometry, spectrophotometry, and microscopy), the SPR methodology has some important advantages over all other optical techniques, as follows. It utilizes a relatively simple optical system, it has a superior sensitivity, and the complete system of measurement is located on the side of apparatus which is remote from the sample, and thus there is no optical interference from the bulk medium. Further, the surfaces of the sample need no extra treatment to increase reflectivity, because this is achieved by operating at the critical angle for total reflectance. Additional benefits include the fact that there are three parameters of the resonance that can readily be measured, thereby yielding much more information about the sample and changes within it than the simple interferometric step height used in other sensitive optical techniques. Finally, the method allows direct measurements of mass in real time. CPWR expands spectroscopic sensitivities and capabilities even further, allowing the measurement of anisotropies in both the refractive index and the extinction coefficient.

2 SENSITIVITY OF THE SURFACE PLASMON RESONANCE METHOD

The sensitivity, S, of an SPR measurement can be defined as the change in reflectance, measured either at a specified incident angle, α_1, or a specified excitation wavelength, λ_1, within the range of the resonance curve, divided by the change in one of the optical parameters, (Equations 6a and 6b):

$$S_{\alpha 1} = \left.\frac{\mathrm{d}R_{(\alpha)}}{\mathrm{d}n\,\mathrm{d}k\,\mathrm{d}t}\right|_{\alpha 1} \tag{6a}$$

$$\text{or} \quad S_{\lambda 1} = \left.\frac{\mathrm{d}R_{(\lambda)}}{\mathrm{d}n\,\mathrm{d}k\,\mathrm{d}t}\right|_{\lambda 1} \tag{6b}$$

respectively. As can be seen from the explicit expression for R given by the characteristic matrix of a thin film assembly,[3] the reflectance is not only a function of the optical parameters of the sensing layer, but also depends on the optical parameters of the incident and emerging media, as well as those of the metal film which generates the SPs. In addition, the refractive indices and extinction coefficients of these media are related to one another by the complex form of Snell's Law, which complicates the function R even further. In general, however, changes in the experimental value of R are generated by two factors: the shift of the position and the change of the shape of

the resonance spectrum. The latter parameter is usually described either by the sharpness of the SPR spectrum, i.e. its half-width, or by the slope of the reflectance function, and characterizes the resolution in either the resonance angle or the resonance wavelength.

In general, the overall sensitivity of the SPR device depends on both the metallic-layer material (e.g. gold or silver), and the type of SPR being measured. The increase in the evanescent field is much smaller (about 2-fold) with gold than silver (see Figures 1–3), which translates into about a 4-fold smaller overall sensitivity for an SPR device based on gold.[4] The different types of SPRs (Figures 1–3) show different distributions of the evanescent electromagnetic field within the resonator device, resulting in widely varying sensitivities. LRSPR has about 2.5-fold higher overall sensitivity than conventional SPR, whereas CPWR shows an even higher increase in sensitivity, about 3.5-fold for *p*-polarization and about 8-fold for *s*-polarization, as compared to conventional SPR.[17] Therefore, the final design of an SPR device is usually a compromise between the different factors influencing overall sensitivity, durability, and any other requirements of the device for a specific practical application.

3 IMMOBILIZATION OF PROTEINS ON PLASMON-GENERATING SOLID SURFACES

The SPR (or CPWR) technique can only be applied to investigate surface or interfacial phenomena when combined with an appropriate thin metal (or dielectric) film surface which is modified to allow the desired types of molecular interactions to occur at the surface or interface. Therefore, one important function of such solid-surface modifications is to immobilize the molecules of interest at the interface. A second important feature of the surface modification is to prevent molecules on the solid surface from undergoing structural changes due to adsorption, which can result in a loss of biological activity or a modification of the thermodynamic properties of the system. This is especially important in the case of immobilization of proteins at a solid–liquid interface.

There are two principal techniques that have been employed in order to immobilize proteins at solid–liquid interfaces, while avoiding unwanted structural perturbations. The first of these involves transfer onto a solid surface of an immobilizing medium, for example, dextran gels,[23] lipid monolayers generated at a liquid–gas interface by the Langmuir–Blodgett method,[24] or lipid vesicles suspended in an aqueous solution,[25] followed by binding of the protein to the immobilizing medium by electrostatic interactions, hydrophobic interactions, hydrogen bonding, etc. The second method uses self-assembly techniques that directly form an immobilization layer on a solid surface. The first of these involves chemical modification of the solid surface, whereas the most commonly used second method applies the principles governing the spontaneous formation of freely suspended lipid bilayer membranes within an orifice in a hydrophobic barrier separating two aqueous regions.[26,27]

3.1 Chemical Modifications of the Surface

Chemical modifications usually involve formation of film coatings by covalent attachment to the solid surface. This type of surface modification includes alkylsilane surfactants on hydroxylated surfaces such as silica and aluminum oxide, alkanethiolates on gold, silver, and copper, alcohols and amines on platinum, and carboxylic acids on aluminum oxide and silver oxide.[10,12,27,28–32] The technique allows formation of oriented monolayer films. If combined with polymeric material, such procedures can be used to create much thicker structures, as has been done in a commercially available SPR device, where an approximately 100–200-nm-thick layer of covalently bound dextran hydrogel has been produced on a thin gold film.[23] The synthesis and characterization of coatings of organic surfactants on metal or metal oxide surfaces has become a separate field of physical chemistry.

Although both of these solid-surface modifications, i.e. thin monolayer coatings and thick layers of dextran hydrogel, have been successfully used in protein immobilization, they have some inherent limitations. In general, there are two types of restrictions:

1. those associated with the amount of biological material which can be immobilized at the modified surface; and
2. those related to the types of biological material that can be immobilized.

The first of these influences the dynamic concentration range (defined as the range below the saturation level) that can be measured by SPR. Thus, the quantity of immobilized molecules will be proportional to the amount of the chemical coating, and by increasing the thickness of the chemical coating to produce a three-dimensional multilayer surface, one can thereby increase the extent of molecule immobilization. However, this introduces some very serious disadvantages, which result both from complexities in the kinetic pattern of protein–protein interactions within the three-dimensional structure of the modifier, as well as from the exponentially decaying sensitivity of the SPR response across the thickness of the chemical modifier.[33]

The second limitation of chemical modification procedures results from the fact that they generate either a densely packed monolayer of modifier or a three-dimensional hydrophilic gel. Although such systems have been used with water-soluble proteins they are not very well suited to the immobilization of integral membrane proteins, which require a lipid bilayer or a surfactant micelle to retain their native properties. Therefore, there have been some attempts to adjust the chemical surface modifications to allow membrane proteins to be immobilized at a solid–liquid interface, by combining them with Langmuir–Blodgett monolayers or by using lipid vesicles.[12,13,27,34] However, the use of self-assembled lipid bilayer membranes as described in the next section appears to be superior, and provides a more flexible and biocompatible environment.

3.2 Self-assembled Solid-supported Lipid Membranes as Biocompatible Matrices

The application of SPR to membrane proteins is more difficult than to water-soluble proteins, because the former materials require a hydrophobic environment, and they are especially difficult to maintain in a functional state. These two factors become even more important when the proteins have to be associated with a solid surface.[27,34] Several approaches to generating a surface coating that would be compatible with membrane proteins have been developed. As noted above, one such approach involves the transfer onto a solid surface of either monolayers generated at a liquid–gas interface by the Langmuir–Blodgett method, or lipid bilayer vesicles suspended in an aqueous solution. A second approach is based on self-assembly techniques that directly form a bilayer on a solid surface. Although both of these methodologies can produce a lipid bilayer film on a solid surface, the properties of the resulting membranes are quite different.

Tamm and McConnell[35] have described the formation of a supported phospholipid bilayer generated by sequential transfer of two Langmuir–Blodgett monolayers from the air–water interface to the solid substrate, and Kalb et al.[36] have developed a procedure for forming supported phospholipid bilayers by fusion of vesicles to supported phospholipid monolayers. These two techniques have been further elaborated by combining them with chemical modification of the solid surface (mainly by silanization and alkanethiolization). Although such planar supported lipid bilayer membranes can in principle be used in SPR measurements of membrane-associated proteins, a major limitation involves the difficulty of reconstituting transmembrane proteins. This is a result of a membrane-generating procedure in which two monolayers are transferred separately to form a bilayer system.

Generation of self-assembled lipid bilayers employs the principles that govern the spontaneous formation of a freely suspended lipid bilayer membrane within an orifice in a hydrophobic barrier separating two aqueous solutions.[26] This procedure, when applied to a solid surface in contact with an aqueous medium, generates a lipid bilayer membrane which rests on an ultrathin water layer, which makes it freely supported and accessible for transmembrane proteins which usually protrude outside the membrane.[15,19,20,27] Furthermore, because the membrane is formed from a lipid solution, it contains small amounts of hydrocarbon solvent molecules, which are incorporated within the lipid monolayers. This results in two significant alterations in the bilayer structure. It decreases the tilt angle of the aliphatic tails of the lipid molecules, that is it increases the thickness of the lipid bilayer, and, what is even more important in the context of the present applications, also separates the lipid molecules, introducing both more flexibility and extra space between the head groups. The method of preparation involves spreading a small amount of lipid bilayer-forming solution across a small hole in a Teflon® sheet that separates a metal (or semiconductor) support from an aqueous phase.[15,18,37] The hydrophilic surface of the solid support, which is covered with an ultrathin water layer, attracts the polar headgroups of the lipid molecules, thus forming an adsorbed lipid monolayer with the hydrocarbon chains oriented toward the bulk lipid phase. Subsequent to this first step of lipid membrane formation, the main body of the sample cell is filled with an appropriate aqueous buffer solution. This initiates the second step, which involves a thinning process, that is formation of the second monolayer and a plateau–Gibbs border that anchors the bilayer film to the Teflon® spacer. The plateau–Gibbs border, which results from the membrane-forming process, and is composed of the lipid solution used to generate the membrane, plays the role of a reservoir of lipid and solvent molecules, which can be moved either into or out of the storage region during the process of bilayer development or membrane protein incorporation.

4 BIOLOGICAL APPLICATIONS OF SURFACE PLASMON RESONANCE SPECTROSCOPY

As a consequence of the abovementioned characteristics, SPR is ideally suited to studying both structural and mass changes of thin films, including molecular interactions such as protein–protein or lipid–protein, occurring at

surfaces and interfaces. These can be examined using either steady-state SPR spectroscopy, when Equation (7) applies:

$$R = f(\alpha)_{\text{time=const, } \lambda=\text{const}};$$

$$\text{or} \quad R = f(\lambda)_{\text{time=const, } \alpha=\text{const}} \tag{7}$$

or time-resolved SPR spectroscopy, when Equation (8) is valid:

$$R = f(\alpha, \text{time})_{\lambda=\text{const}};$$

$$\text{or} \quad R = f(\lambda, \text{time})_{\alpha=\text{const}} \tag{8}$$

and can be applied to a wide range of materials. The time-resolved mode expands the capability of SPR techniques to allow probing of the dynamics of structural and mass alterations of thin films.

4.1 Structural Parameters of Biomolecular Films

The influence of optical parameters on the angular (or wavelength) position, the angular (or wavelength) width, and the depth of the SPR spectrum is completely contained in the characteristic matrix of the thin-film assembly which allows a determination of their values. The experimental SPR (or CPWR) spectra interpreted in the context of the characteristic matrix of the assembly allow a unique evaluation of n, k, and t.[4] The evaluation procedure is based on fitting a theoretical resonance curve to the experimental one. In addition, the CPWR method provides a means for determining the optical parameters using both TM (p) and TE (s) polarizations of the excitation light, resulting in two values of the refractive index (n_p and n_s), and two values of the extinction coefficient (k_p and k_s). These parameters can then be used to calculate the anisotropy of n and k, thereby describing the degree of both molecular order (by the anisotropy in n) and orientation of chromophoric groups attached to the molecules comprising the thin film (by the anisotropy in k).[4,17] Such information, taken together with the film thickness (t), provides insights into the microscopic structure of the film. Furthermore, the optical parameters can also be employed to calculate the mass of a deposited thin layer (see next section for details).

4.2 Evaluation of Binding Affinities

As noted above, the values of the n, k, and t parameters of a thin-film layer deposited on the surface of a metal film contain information about the amount of material in the layer. The adsorbed mass can be calculated from both extinction coefficient and refractive index. The former is based on the relationship of mass (or concentration) of the adsorbed substance and the amount of light which is absorbed, expressed by the well-known Lambert–Beer law. In CPWR experiments the mass (or concentration), can be calculated from the value of the extinction coefficient, k, which is a measure of the light absorption in the medium and is related to the more familiar absorption coefficient, β, which describes the absorption of light of wavelength λ per unit of absorbing substance thickness and concentration C, by Equation (9):

$$k = \beta \left(\frac{C\lambda}{4\pi} \right. \tag{9}$$

There are two different ways of calculating the adsorbed mass from the refractive index value. The first of these is based on the assumption that the refractive index increment, $\mathrm{d}n/\mathrm{d}C$, is independent of the concentration, C, of the adsorbed substance. If this is so, the surface mass density, m_s, i.e. the amount of material per unit surface area, can be evaluated by Equation (10):

$$m_s = t \quad \frac{n - n_2}{\mathrm{d}n/\mathrm{d}C} \tag{10}$$

where n and n_2 are refractive indices of the adsorbed thin film and the emerging medium, respectively, and $\mathrm{d}n/\mathrm{d}C$ is the refractive index increment of the adsorbed substance. Equation (10) can, of course, only be used if the refractive index increment is known, and depends on the assumption that the value of the increment is constant over the concentration range of the adsorbed material.

The second method of mass calculation is based on the Lorentz–Lorenz relation, which can be presented in the most general case, i.e. when the deposited layer contains a mixture of substances, by Equation (11):

$$\frac{n^2 - 1}{n^2 + 2} = A_1 M_1 + A_2 M_2 + \cdots \tag{11}$$

where A_i and M_i are the molar refractivity and the number of moles of substance per unit volume, respectively.[38]

The most straightforward application of this aspect of SPR (or CPWR) technology is the measurement of binding isotherms and binding affinities.[15] Steady-state SPR measurements, by determining mass density changes occurring in molecular assemblies accompanying binding interactions, allow an evaluation of the binding affinities.[15]

4.3 Determination of Rate Constants for Binding Reactions

The measurements can be greatly simplified when the SPR (CPWR) technique is used in a kinetic mode, i.e. when reflectance is only a function of time at $\alpha = \text{const}$ and $\lambda = \text{const}$. Under the stated conditions, the reflectance is taken at one specific point of the resonance curve as a function of time and is described by

Equation (12):

$$R = f(\text{time})_{\alpha=\text{const},\ \lambda=\text{const}} \quad (12)$$

Such a simplified experimental approach can provide a means for real-time analysis of molecular interactions and allows one to directly probe the molecular interaction, including protein–protein and lipid–protein interactions, by monitoring association and dissociation processes in real time,[33,39] under the following conditions. First, the measurement must be performed with an optically isotropic system characterized by a refractive index value that is independent of the light polarization. Second, the binding of one molecule to another that is immobilized on the surface of an SPR device must produce a change in the refractive index, n, of the interface that results in an alteration of the SPR spectrum. This latter condition is only valid when the SPR-exciting light wavelength is outside of an absorption spectral region of both interacting molecules, and when the interaction does not induce any structural changes in the interface. Third, the changes in n are assumed to be proportional to the surface concentration of bound molecules, C, i.e. $dn/dC = \text{constant}$ over the whole concentration range. This assumption can give rise to some error, especially at high concentrations where the proportionality might not be fulfilled. Fourth, the changes in n must only cause a shift of the resonance curve without any alterations in its shape, which is a simplification of the theoretical influence of n on the resonance curve as determined by the characteristic matrix of the thin-film assembly. Under these conditions, changes in the n-value can be measured directly by the shift of the SPR resonance curve, which is the basic requirement of using kinetic measurements to monitor association and dissociation processes in real time.

ACKNOWLEDGMENT

The preparation of this article was supported by a grant from the National Science Foundation (MCB-9404702).

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflection
CPWR	Coupled Plasmon-waveguide Resonance
LRSPR	Long-range Surface Plasmon Resonance
SP	Surface Plasmon
SPR	Surface Plasmon Resonance
TE	Transverse Electric
TIRF	Total Internal Reflectance Fluorescence
TM	Transverse Magnetic

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Fluorescence Spectroscopy in Peptide and Protein Analysis

Surfaces **(Volume 10)**
Differential Reflectance Spectroscopy in Analysis of Surfaces • Ellipsometry in Analysis of Surfaces and Thin Films • Photoluminescence in Analysis of Surfaces and Interfaces

Electronic Absorption and Luminescence **(Volume 12)**
Surface Measurements using Absorption/Luminescence

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics

REFERENCES

1. H. Raether, 'Surface Plasma Oscillations and their Applications' in *Physics of Thin Films*, Academic Press, New York, 145–261, Vol. 9, 1977.
2. F. Abeles, 'Surface Electromagnetic Waves Ellipsometry', *Surf. Sci.*, **56**, 237–251 (1976).
3. H.A. Macleod, *Thin-film Optical Filters*, McGraw Hill, New York, 1988.
4. Z. Salamon, H.A. Macleod, G. Tollin, 'Surface Plasmon Resonance Spectroscopy as a Tool for Investigating the Biochemical and Biophysical Properties of Membrane Protein Systems. I: Theoretical Principles', *Biochim. Biophys. Acta*, **1331**, 117–129 (1997).
5. K. Holst, H. Raether, 'The Influence of Thin Surface Films on the Plasma Resonance Emission', *Optics Comm.*, **2**(7), 312–316 (1970).
6. E. Kretschmann, 'Die Bestimmung Optischer Konstanten von Metallen durch Anregung von Oberfflachenplasmasmaschwingungen', *Z. Phys.*, **241**, 313–324 (1971).
7. A. Otto, 'Excitation of Nonradiative Surface Plasma Waves in Silver by the Method of Frustrated Total Reflection', *Z. Physik*, **216**, 398–402 (1968).
8. I. Pockrand, J.D. Swalen, J.G. Gordon II, M.R. Philpott, 'Surface Plasmon Spectroscopy of Organic Monolayer Assemblies', *Surf. Sci.*, **74**, 237–244 (1977).
9. T. Lopez-Rios, G. Vuye, 'Use of Surface Plasmon Excitation for Determination of the Thickness and Optical Constants of Very Thin Surface Layers', *Surf. Sci.*, **81**, 529–538 (1979).
10. L. Häussling, H. Ringsdorf, F.J. Schmitt, W. Knoll, 'Biotin-functionalized Self-assembled Monolayers on Gold:

Surface Plasmon Optical Studies of Specific Recognition Reactions', *Langmuir*, **7**(9), 1837–1840 (1991).
11. W.M.K.P. Wijekoon, B. Asgharian, M. Casstevens, M. Samoc, G.B. Talapatra, P.N. Prasad, T. Geisler, S. Rosenkilde, 'Electrooptic Effect in Langmuir–Blodgett Films of 2-(Docosylamino)-5-nitropyridine Probed by Surface Plasmon Waves', *Langmuir*, **8**, 135–139 (1992).
12. A.L. Plant, 'Self-assembled Phospholipid/Alkanethiol Biomimetic Bilayers on Gold', *Langmuir*, **9**, 2764–2767 (1993).
13. S. Terrettaz, T. Stora, C. Duschl, H. Vogel, 'Protein Binding to Supported Lipid Membranes: Investigation of the Cholera Toxin–Ganglioside Interaction by Simultaneous Impedance Spectroscopy and Surface Plasmon Resonance', *Langmuir*, **9**, 1361–1369 (1993).
14. P.B. Garland, 'Optical Evanescent Wave Methods for the Study of Biomolecular Interactions', *Q. Rev. Biophys.*, **29**, 91–117 (1996).
15. Z. Salamon, H.A. Macleod, G. Tollin, 'Surface Plasmon Resonance Spectroscopy as a Tool for Investigating the Biochemical and Biophysical Properties of Membrane Protein Systems. II: Applications to Biological Systems', *Biochim. Biophys. Acta*, **1331**, 131–152 (1997).
16. Z. Salamon, G. Tollin, 'Surface Plasmon Resonance Spectroscopy: A new Biophysical Tool for Probing Membrane Structure and Function', in *Biomembrane Structures*, eds. D. Chapman, P. Haris, IOS Press, Amsterdam, 186–204, 1998.
17. Z. Salamon, H.A. Macleod, G. Tollin, 'Coupled Plasmon-waveguide Resonators: A New Spectroscopic Tool for Probing Proteolipid Film Structure and Properties', *Biophys. J.*, **73**, 2791–2797 (1997).
18. Z. Salamon, Y. Wang, G. Tollin, H.A. Macleod, 'Assembly and Molecular Organization of Self-assembled Lipid Bilayers on Solid Substrates Monitored by Surface Plasmon Resonance Spectroscopy', *Biochim. Biophys. Acta*, **1195**, 267–275 (1994).
19. Z. Salamon, Y. Wang, M.F. Brown, H.A. Macleod, G. Tollin, 'Conformational Changes in Rhodopsin Probed by Surface Plasmon Resonance Spectroscopy', *Biochemistry*, **33**, 13706–13711 (1994).
20. Z. Salamon, Y. Wang, J.L. Soulages, M.F. Brown, G. Tollin, 'Surface Plasmon Resonance Spectroscopy Studies of Membrane Proteins: Transducin Binding and Activation by Rhodopsin Monitored in Thin Membrane Films', *Biophys. J.*, **71**, 283–294 (1996).
21. Z. Salamon, G. Tollin, 'Surface Plasmon Resonance Studies of Complex Formation between Cytochrome c and Bovine Cytochrome c Oxidase Incorporated into a Supported Planar Lipid Bilayer. I: Binding of Cytochrome c to Cardiolipin/Phosphatidylcholine Membranes in the Absence of Oxidase', *Biophys. J.*, **71**, 848–857 (1996).
22. P. Schuck, 'Use of Surface Plasmon Resonance to Probe the Equilibrium and Dynamic Aspects of Interactions between Biological Macromolecules', *Ann. Rev. Biophys. Biomol. Struct.*, **26**, 541–566 (1997).
23. S. Löfås, B. Johnsson, 'A Novel Hydrogel Matrix on Gold Surfaces in Surface Plasmon Resonance Sensors for Fast and Efficient Covalent Immobilization of Ligands', *J. Chem. Soc. Chem. Commun.*, 1526–1528 (1990).
24. I.R. Peterson, 'Langmuir–Blodgett Films', *J. Phys. Appl. Phys.*, **23**, 379–395 (1990).
25. A.A. Brian, H.M. McConnell, 'Simulation of Cytotoxic T Cells by Supported Planar Membranes', *Proc. Natl. Acad. Sci.*, USA, **81**, 6159–6166 (1984).
26. P. Mueller, D.O. Rudin, H.T. Tien, W.C. Wescott, 'Reconstitution of Cell Membrane Structure in vitro and its Transformation into an Excitable System', *Nature*, **194**, 979–980 (1962).
27. E. Sackmann, 'Supported Membranes: Scientific and Practical Applications', *Science*, **271**, 43–48 (1996).
28. R.G. Nuzzo, D.L. Allara, 'Adsorption of Bifunctional Organic Disulfides on Gold Surfaces', *J. Am. Chem. Soc.*, **105**, 4481–4483 (1983).
29. P.B. Daniels, J.K. Deacon, M.J. Eddowes, D.G. Pedley, 'Surface Plasmon Resonance Applied to Immunosensing', *Sensors and Actuators*, **25**, 11–18 (1988).
30. H. Morgan, D.M. Taylor, C. D'Silva, 'Surface Plasmon Resonance Studies of Chemisorbed Biotin-streptavidin Multilayers', *Thin Solid Films*, **209**, 122–126 (1992).
31. M. Wilchek, E.A. Bayer, 'The Avidin-biotin Complex in Bioanalytical Applications', *Anal. Chem.*, **171**, 1–32 (1988).
32. M. Egger, S.P. Heyn, H.E. Gaub, 'Synthetic Lipid-anchored Receptors Based on the Binding Site of Monoclonal Antibody', *Biochim. Biophys. Acta*, **1104**, 45–54 (1992).
33. P. Schuck, 'Kinetics of Ligand Binding to Receptor Immobilized in a Polymer Matrix, as Detected with an Evanescent Wave Biosensor. I. A Computer Simulation of the Influence of Mass Transport', *Biophys. J.*, **70**, 1230–1249 (1996).
34. D. Chapman, 'Biomembranes and New Hemocompatible Materials', *Langmuir*, **9**, 39–45 (1993).
35. L.K. Tamm, H.M. McConnell, 'Supported Phospholipid Bilayers', *Biophys. J.*, **47**, 105–113 (1985).
36. E. Kalb, S. Frey, L.K. Tamm, 'Formation of Supported Planar Bilayers by Fusion of Vesicles to Supported Phospholipid Monolayers', *Biochim. Biophys. Acta*, **1103**, 307–316 (1992).
37. Z. Salamon, R.A. Schmidt, G. Tollin, H.A. Macleod, 'Reusable Biocompatible Interface for Immobilization of Materials on Solid Support', U.S. Patent 5 521 702, 1996.
38. P.A. Cuypers, J.W. Corsel, M.P. Janssen, J.M.M. Kop, W.Th. Hermens, H.C. Hemker, 'The Adsorption of Prothrombin to Phosphatidylserine Multilayers Quantitated by Ellipsometry', *J. Biol. Chem.*, **258**, 2426–2431 (1983).
39. R. Karlsson, 'Real-time Competitive Kinetic Analysis of Interactions between Low-molecular-weight Ligands in Solution and Surface-immobilized Receptors', *Anal. Chem.*, **221**, 142–151 (1994).

X-ray Crystallography of Biological Macromolecules

Albrecht Messerschmidt and Robert Huber
Max-Planck-Institut für Biochemie, Martinsried, Germany

1 Introduction **1**
1.1 Crystals and Symmetry 1
1.2 Protein Solubility 3
1.3 Experimental Techniques 4
1.4 Crystallization Screenings 6
2 Experimental Techniques **6**
2.1 X-ray Sources 6
2.2 Detectors 8
2.3 Crystal Mounting and Cooling 11
2.4 Data Collection Techniques 12
3 Principles of X-ray Diffraction by a Crystal **13**
3.1 Scattering of X-rays by an Atom 13
3.2 Scattering of X-rays by a Unit Cell 14
3.3 Scattering of X-rays by a Crystal 15
3.4 The Reciprocal Lattice and Ewald Construction 16
3.5 The Temperature Factor 17
3.6 Symmetry in Diffraction Patterns 17
3.7 Electron Density Equation and Phase Problem 17
3.8 The Patterson Function 18
3.9 Integrated Intensity Diffracted by a Crystal 18
3.10 Intensities on an Absolute Scale 18
3.11 Resolution of the Structure Determination 19
3.12 Diffraction Data Evaluation 19
3.13 Solvent Content of Protein Crystals 19
4 Methods for Solving the Phase Problem **20**
4.1 Isomorphous Replacement 20
4.2 Anomalous Scattering 22
4.3 Patterson Search Methods (Molecular Replacement) 26
4.4 Phase Calculation 27
4.5 Phase Improvement 29
4.6 Difference Fourier Technique 31
5 Model Building and Refinement **32**
5.1 Model Building 32
5.2 Crystallographic Refinement 32
5.3 Accuracy and Verification of Structure Determination 33
6 Applications **34**
6.1 Enzyme Structure and Enzyme–Inhibitor Complex 34
6.2 Metalloproteins 38
6.3 Large Molecular Assembly 42
Abbreviations and Acronyms **43**
Related Articles **44**
References **44**

X-ray crystallography of biological macromolecules comprises the determination of the three-dimensional structures of proteins, nucleic acids and other biological macromolecules at atomic resolution by diffraction of X-rays on crystals of such macromolecules. The atomic structure is obtained from the electron density distribution of the macromolecular crystal, which is the Fourier transform of the waves diffracted by the crystal. The amplitudes of the diffracted waves are determined directly from the diffraction experiment. The diffraction phases are revealed (i) from additional diffraction data of isomorphous heavy atom derivatives [multiple isomorphous replacement (MIR) technique], (ii) from multiple anomalous diffraction (MAD), if suitable anomalous scatterers are in the crystal and using tunable synchrotron radiation, and (iii) by Patterson search techniques (molecular replacement), if structural information on the biological macromolecule under investigation is available. X-ray crystallography of biological macromolecules is the unique method for the elucidation of the spatial structures at atomic resolution of complex biomolecules with molecular masses greater than 30 kDa. The structures represent a time average and spatial average of molecules packed in a crystal. The preparation of suitable crystals may be a limiting factor.

1 INTRODUCTION

1.1 Crystals and Symmetry

In a crystal, atoms or molecules are arranged in a three-dimensionally periodic manner by translational symmetry. The crystal is formed by a three-dimensional stack of unit cells which is called the crystal lattice (Figure 1a and b). The unit cell is built up by three non-collinear vectors **a**, **b** and **c**. In the general case these vectors have unequal magnitudes and their mutual angles deviate from 90°. The arrangement of the molecule(s) in the unit cell may be asymmetrical but very often

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

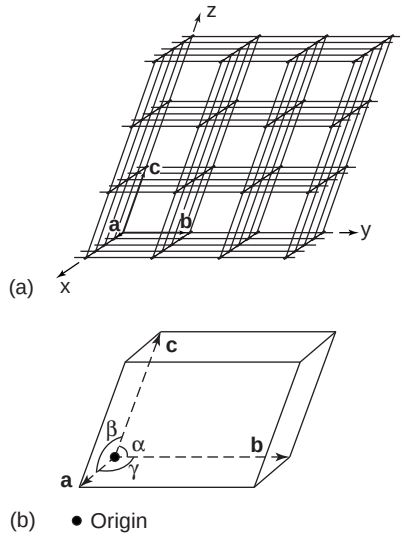

Figure 1 (a) Crystal lattice and (b) unit cell.

it is symmetrical. This is illustrated in Figure 2(a–e) in two-dimensional lattices for rotational symmetries. In crystals, only 1-, 2-, 3-, 4- and 6-fold rotations are allowed. This follows from the combination of the lattice properties with rotational operations. Other possible symmetry elements are mirror plane *m*, inversion center and combination of rotation axis with inversion center (inversion axis). These are the point group symmetries. They can only occur among each other in a few certain combinations of angles. Other angle orientations would violate the lattice properties. The number of all possible combinations reveals the 32 point groups. The crystal morphology obeys the point group symmetries.

Adding an inversion center to the point group symmetry leads to the 11 Laue groups. These are of importance for the symmetry of X-ray diffraction patterns. Their symbols are: $\bar{1}$, 2/*m*, 2/*mmm*, $\bar{3}$, $\bar{3}m$, 4/*m*, 4/*mmm*, 6/*m*, 6/*mmm*, *m*3, *m*3*m*. Proteins and nucleic acids are chiral molecules. Therefore, they can crystallize only in the 11 enantiomorphic point groups: 1, 2, 3, 4, 6, 23, 222, 32, 422, 622 and 432.

The combination of point group symmetries with lattices leads to seven crystal systems,

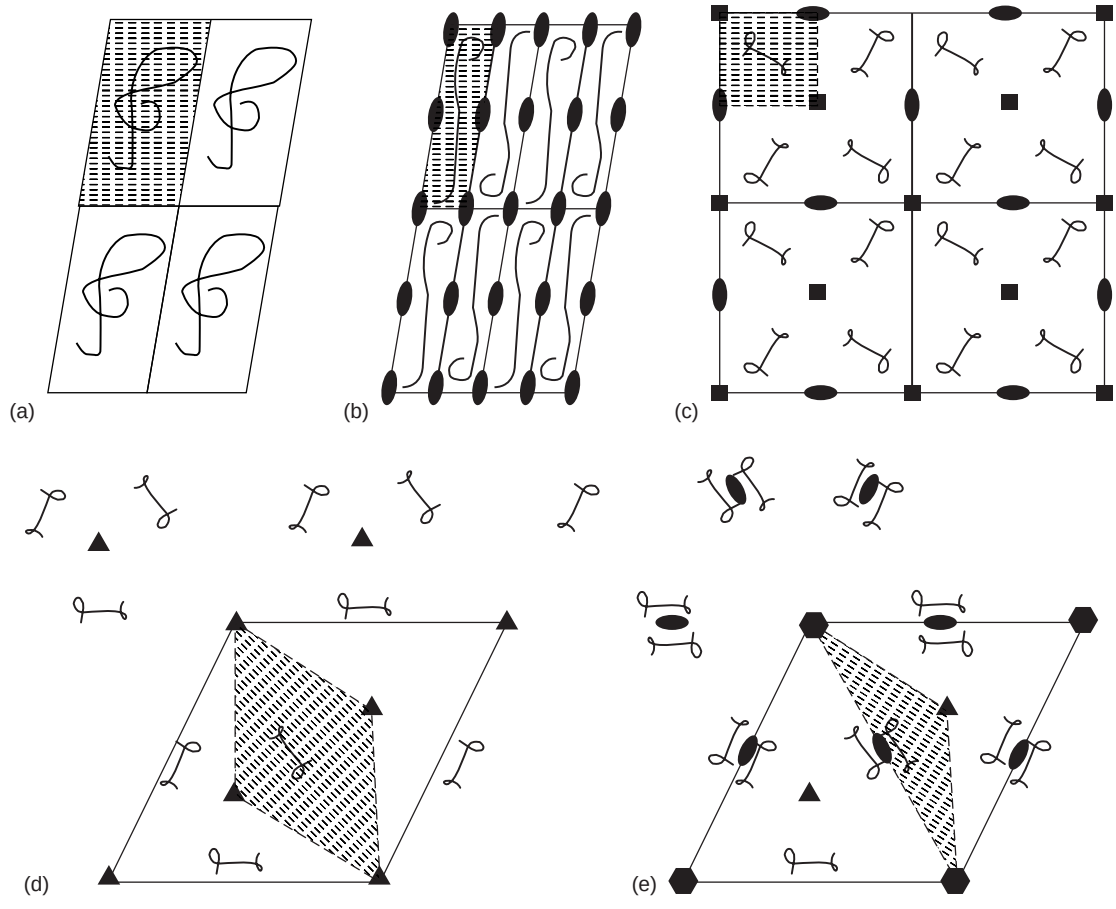

Figure 2 Rotational symmetry elements in two-dimensional lattices: (a) 1, (b) 2, (c) 4, (d) 3 and (e) 62. The asymmetric unit is hatched.

triclinic, monoclinic, orthorhombic, trigonal, tetragonal, hexagonal and cubic, with 14 different Bravais-lattice types which can be primitive, face-centered, all-face-centered and body-centered. Furthermore, additional symmetry elements are generated having translational components such as screw axes or glide mirror planes. There exist 230 space groups of which 65 are enantiomorphic (for chiral molecules such as proteins). Figure 3 shows the graphical representation for the space group $P2_12_12_1$ as listed in the *International Tables for Crystallography*.[1] The asymmetric unit is one fourth of the unit cell and can contain one or several molecules. Multimeric molecules may have their own symmetries which are called noncrystallographic symmetries. Here axes which are 5-fold, 7-fold, etc., are also allowed.

1.2 Protein Solubility

Figure 4 shows a typical phase diagram illustrating the solubility properties of a macromolecule. In the labile

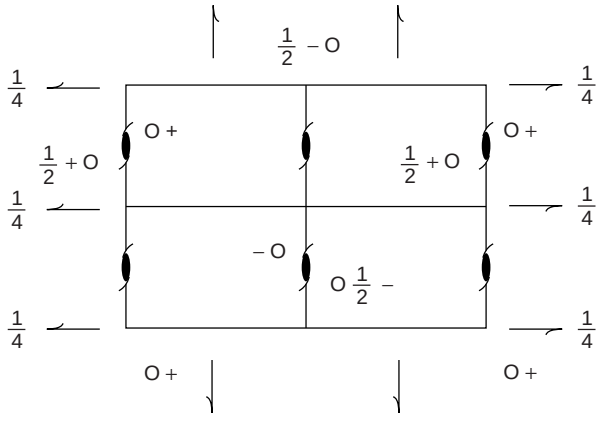

Figure 3 Graphical representation of space group $P2_12_12_1$.

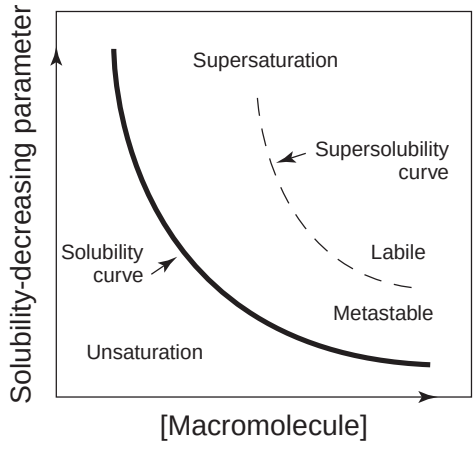

Figure 4 Phase diagram illustrating the solubility properties of macromolecules. (Reproduced by permission of Academic Press, Inc., from P.C. Weber, 'Overview of Protein Crystallization Methods', *Methods Enzymol.*, **276**, 13–23 (1997).)

phase crystal nucleation and growth compete whereas in the metastable region only crystal growth appears. In the unsaturated region crystals dissolve. The solubility of proteins is influenced by several factors, as follows.

1.2.1 Ionic Strength

A protein can be considered as a polyvalent ion and therefore its solubility can be discussed on the basis of the Debye–Hückel theory.

In aqueous solution, each ion is surrounded by an "atmosphere" of counter ions. This ionic atmosphere influences the interactions of the ion with water molecules and hence the solubility.

1.2.1.1 "Salting-in" At low ionic concentration, the "ionic atmosphere" increases the solubility as it increases the possibilities for favorable interactions with water molecules. We obtain (Equations 1 and 2)

$$\log S - \log S_0 = \frac{AZ_+Z_-\sqrt{\mu}}{1 + aB\sqrt{\mu}} \tag{1}$$

$$\mu = \frac{1}{2}\sum c_j Z_j^2 \tag{2}$$

where μ = ionic strength, S = solubility of the salt at a given ionic strength μ, S_0 = solubility of the salt in absence of the electrolyte, Z_+, Z_- = ionic charge of salt ions, A, B = constants depending on the temperature and dielectric constant, a = average diameter of ions and c_j = concentration of the jth chemical component.

Ions with higher charge are more effective for changes in solubility. Most salts and proteins are more soluble in low ionic strength than in pure water. This is termed as "salting-in" (Figure 5).

1.2.1.2 "Salting-out" At higher ionic strength the ions compete for the surrounding water. Therefore, water molecules are taken away from the dissolved agent and the solubility decreases according to Equation (3):

$$\log S - \log S_0 = \frac{AZ_+Z_-\sqrt{\mu}}{1 + aB\sqrt{\mu}} - K_s\mu \tag{3}$$

The term $K_s\mu$ predominates at high ionic strengths, which means that "salting-out" is then proportional to the ionic strength (Figure 5). In a medium with low ionic strength the solubility of a protein can be decreased by increasing or decreasing the salt concentration. Salts with small, highly charged ions are more effective than those with large, lowly charged ions. Ammonium sulfate is often used because of its high solubility.

1.2.2 pH and Counterions

A protein is more soluble the larger is its net charge. The minimum solubility is found at the isoelectric

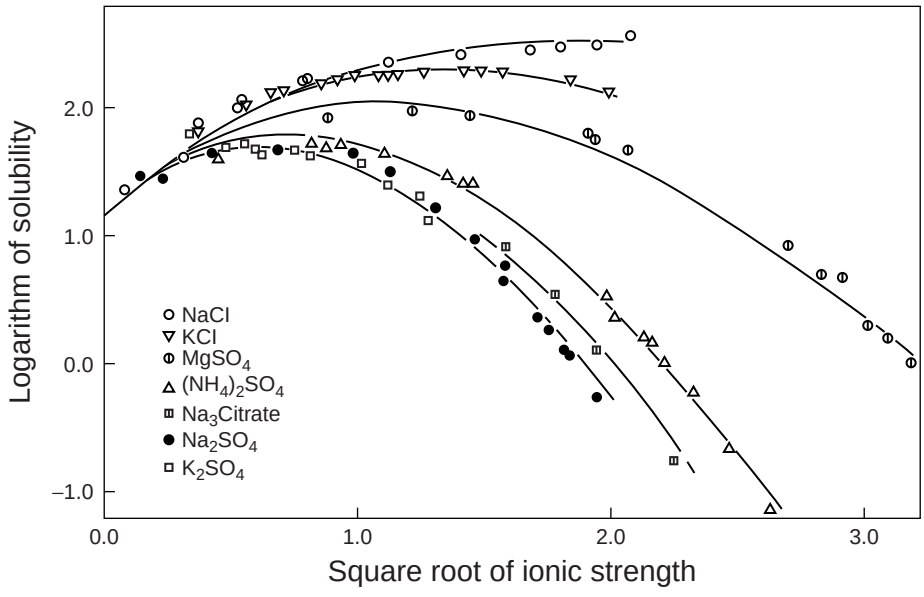

Figure 5 Solubility of carboxyhemoglobin in various electrolytes at 25 °C. (Reproduced by permission of the American Society for Biochemistry and Molecular Biology, from Green.[2])

point. The net charge is zero and hence the packing in the solid state (in the crystal) is possible owing to electrostatic interactions without the accumulation of a net charge of high energy. All ''salting-out'' curves are parallel, K_s remains constant and S_0 varies with pH (Figure 6a and b). In some cases the isoelectric point is different at low and high ionic strength owing to the interactions of the protein with counterions which can cause a net charge at the pH of the isoelectric point.

1.2.3 Temperature

Many factors governing protein solubility are temperature dependent. The dielectric constant decreases with increasing temperature. In the solution energy, $\Delta G = \Delta H - T\Delta S$, the entropy term has an increasing influence with increasing temperature. The temperature coefficient of the solubility depends on other conditions (ionic strength, presence of organic solvents, etc.). At high ionic strength most proteins are less soluble at 25 °C than at 4 °C, e.g. the temperature coefficient is negative. The opposite is valid for low ionic strength.

1.2.4 Organic Solvents

The presence of organic solvents leads to a decrease in the dielectric constant. This causes an augmentation of the electric attraction between opposite charges on the surface of the protein molecule and hence to a reduction in solubility. In general, the solubility of a protein is reduced in the presence of an organic solvent if the temperature decreases. Often organic solvents denature proteins. Therefore, one should work at low temperatures.

1.3 Experimental Techniques

The whole field of macromolecular crystallography has been excellently reviewed in Volumes 114[4] and 115[5] and Volumes 276[6] and 277[7] of *Methods in Enzymology*. A collection of review articles concerning the theory and practice of crystallization of biomacromolecules is given in Part A of Carter and Sweet.[6]

A protein preparation to be used in crystallization should be ''pure'' or ''homogeneous'' at a level that established chromatographic methods are providing (protein content $\geq$ 95%). Furthermore, it should meet the requirements of ''structural homogeneity''. These requirements can be enumerated as follows. It is first necessary to prepare the protein in an isotypically pure state free from other cellular proteins. It may then be necessary to maintain the homogeneity of the protein preparation against covalent modification during crystallization by adding inhibitors of sulfhydryl group oxidation, proteolysis and the action of reactive metals. It may be necessary to suppress the slow denaturation/aggregation of the protein and to restrict its conformational flexibility to reduce the entropic barrier to crystallization presented by extensive conformational flexibility.

For the crystallization of biomacromolecules a broad spectrum of crystallization techniques exists. The most

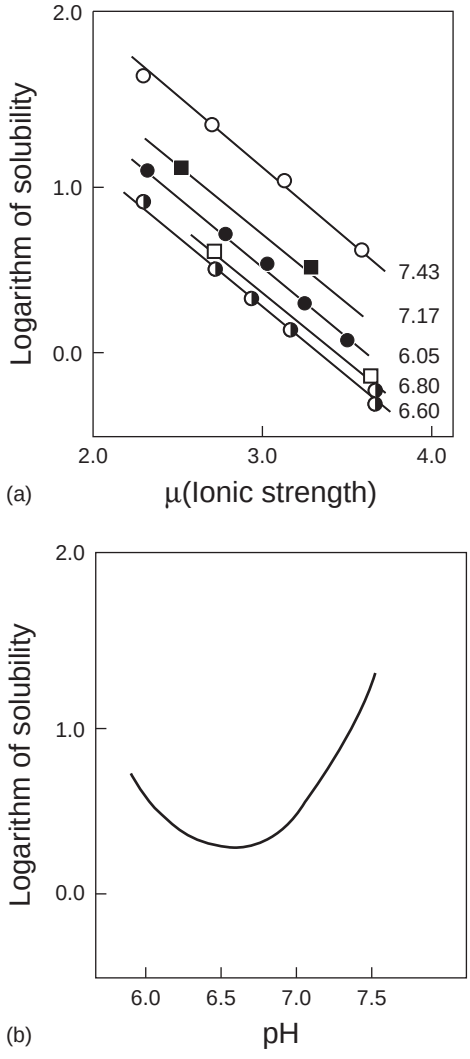

Figure 6 (a) Solubility of hemoglobin at different pH values in concentrated phosphate buffers; (b) extracted from (a). (Reproduced by permission of the American Society for Biochemistry and Molecular Biology, from Green.[3])

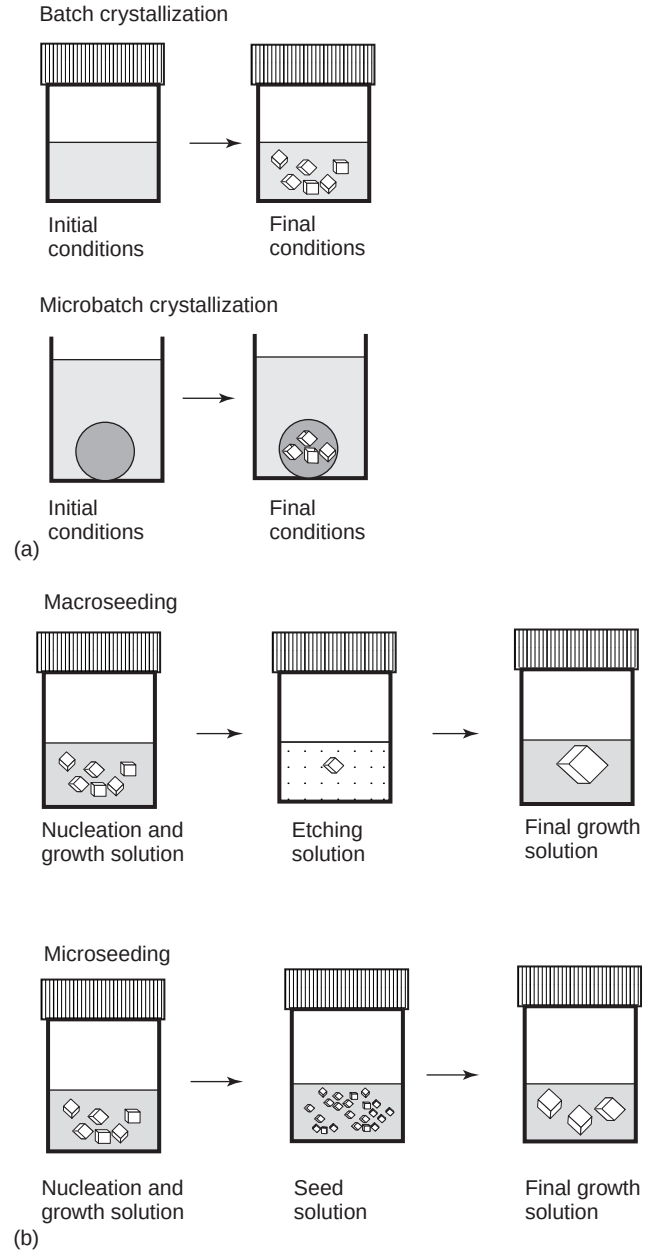

Figure 7 Schematic presentation of (a) batch crystallization and (b) seeding techniques. (Reproduced by permission of Academic Press, Inc., from P.C. Weber, 'Overview of Protein Crystallization Methods', *Methods Enzymol.*, **276**, 13–23 (1997).)

common techniques are described here. The oldest and simplest method is batch crystallization (Figure 7a). In batch experiments, vials containing supersaturated protein solutions are sealed and left undisturbed. In microbatch methods, a small (2–10 μL) droplet containing both protein and precipitant is immersed in an inert oil which prevents droplet evaporation. In the case that ideal conditions for nucleation and growth are different, it is useful to undertake the separate optimization of these processes. This can be done by seeding, a technique where crystals are transferred from nucleation conditions to those that will support only growth (Figure 7b). For macroseeding, a single crystal is transferred to an etching solution, then to a solution of optimal growth. In microseeding experiments, a solution containing many small seed crystals, occasionally obtained by grinding a larger crystal, is transferred to a crystal growth solution.

The method of crystallization by vapor diffusion is depicted in Figure 8(a). In this method, unsaturated precipitant-containing protein solutions are suspended over a reservoir. Vapor equilibration of the droplet

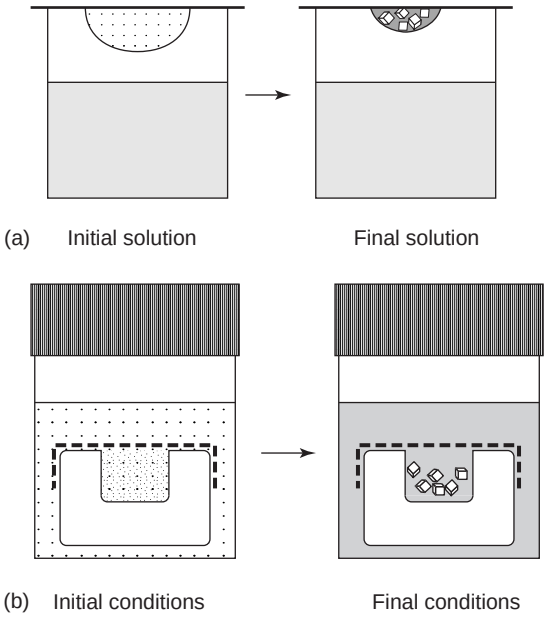

Figure 8 Schematic representation of (a) vapor diffusion and (b) dialysis. (Reproduced by permission of Academic Press, Inc., from P.C. Weber, 'Overview of Protein Crystallization Methods', *Methods Enzymol.*, **276**, 13–23 (1997).)

and reservoir causes the protein solution to reach a supersaturation level where nucleation and initial crystal growth occur. Changes in soluble protein concentration in the droplet are likely to decrease supersaturation over the time course of the experiment. The vapor diffusion technique can be carried out as hanging drop or sitting drop method.

In crystallization by dialysis, the macromolecular concentration remains constant as in batch methods (Figure 8b) because the molecules are forced to stay in a fixed volume. The solution composition is changed by diffusion of low-molecular-weight components through a semipermeable membrane. The advantage of dialysis is that the precipitating solution can be easily changed. Dialysis is also uniquely suited to crystallizations at low ionic strength and in the presence of volatile reagents such as alcohols.

1.4 Crystallization Screenings

Screening schemes have been developed which change the most common parameters of this multiparameter problem such as protein concentration, the nature and concentration of the precipitant, pH and temperature. Each screening can be extended by adding specific additives in low concentrations that affect the crystallization. Sparse matrix crystallization screens are widely applied. The sparse matrix formulation allows one to screen efficiently a broad range of the most popular and effective salts (e.g. ammonium sulfate, sodium and potassium phosphate, sodium citrate, sodium acetate, lithium sulfate), polymers [e.g. poly(ethyleneglycol) (PEG) of different molecular masses (from 400 to 8000)] and organic solvents [e.g. 2,4-methylpentanediol (MPD), 2-propanol, ethanol] versus a wide range of pH.[8] Another approach is the systematic screening of the statistically most successful precipitants. A single precipitant is screened at four unique concentrations versus seven precise levels of pH between 4 and 10. Such grid screens can be done with ammonium sulfate, PEG 6000, MPD, PEG 6000 in the presence of 1.0 M lithium chloride or sodium chloride. For the crystallization of membrane proteins [see Michel[9]] for each detergent which is necessary to make the membrane protein soluble a whole grid screen or sparse matrix screen must be constructed. In principle, all three techniques can be applied for the different screening schemes but mostly the vapor diffusion technique is applied because it is easy to use and the protein consumption is low. For a typical broad screening, about 2 mg of protein are sufficient. Chryschem plates (sitting drop) or Linbro plates (hanging drops) may be used for the vapor diffusion crystallization screening experiments. Once crystals have been obtained, their size and quality can be optimized by additional fine screens around the observed crystallization conditions. General rules do not exist that indicate which method for crystallization one has to use for which type of protein. Suggestions for crystallization conditions to be tested can be obtained from the Biological Macromolecule Crystallization Database.[10]

2 EXPERIMENTAL TECHNIQUES

2.1 X-ray Sources

2.1.1 Conventional X-ray Generators

X-rays are produced when a beam of high-energy electrons, which have been accelerated through a voltage V in a vacuum, hit a target. An X-ray tube run at voltage V will emit a continuous X-ray spectrum with a minimum wavelength given by Equation (4):

$$\lambda_{\min} = \frac{hc}{eV} = \frac{12\,398}{V} \tag{4}$$

with λ in ångströms ($1\,\text{Å} = 10^{-10}\,\text{m}$) and V in volts. The critical voltage, V_0, which is required to excite the characteristic line of a particular element can be calculated from the corresponding wavelength for the appropriate absorption edge. For the copper absorption edge $\lambda_{ae} = 1.380\,\text{Å}$. Hence (Equation 5):

$$V_0 = \frac{12\,398}{\lambda_{ae}} = 8.98\,\text{kV} \quad (5)$$

Provided that $V > V_0$, the characteristic line spectra will be produced (Figure 9). The oldest and cheapest X-ray sources are sealed X-ray tubes. The cathode and anode are situated under vacuum in a sealed glass tube and the heat generated at the anode is removed by a water cooling system. For the generation of higher intensities, as needed in protein crystallography, one has to use a rotating anode (Figure 10). Here the anode is rotated, which allows a higher power loading at the focal spot. In protein crystallography copper targets are usually taken. The used take-off angle is near 4°, which results in apparent focal spot sizes of about 0.3×0.3 mm.

2.1.2 *Synchrotron Radiation*

As electrically charged particles such as electrons or positrons of high energy are kept under the influence of magnetic fields and travel in a pseudocircular trajectory, synchrotron radiation is emitted and can be used in many different types of experiments [for a comprehensive discussion of synchrotron radiation in macromolecular crystallography, see Helliwell[11]]. The particles are injected into the storage ring directly from a linear accelerator or through a booster ring. They circulate in a high vacuum for several hours at a relative constant energy. To keep the bunched particles traveling in a nearly circular path, a lattice of bending magnets is set up around the storage ring. As the particle beam traverses each magnet, the path of the beam is altered, and synchrotron radiation is emitted. The loss of energy of the particle beam is compensated by a radiofrequency

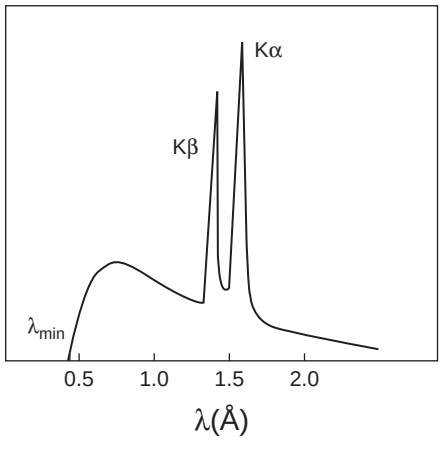

Figure 9 X-ray spectrum emitted from a copper anode. It shows the continuous ''Bremsspektrum'' starting at λ_{min} and the two characteristic copper lines λ K$\alpha = 1.5418$ Å (superposition of λ K$\alpha_1 = 1.5405$ Å and λ K$\alpha_2 = 1.5443$ Å) and λ K$\beta = 1.3922$ Å.

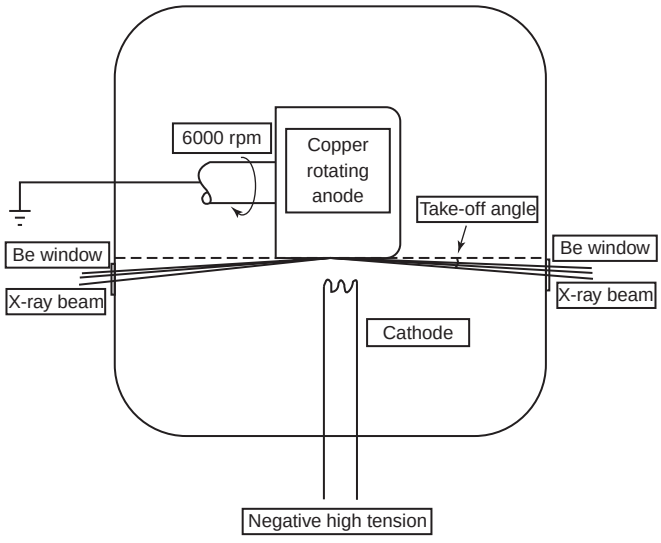

Figure 10 Schematic drawing of a rotating anode tube. Take-off angle is near 4°. For copper the tube is normally operated at 50 kV high tension and 100 mA cathode current.

input at each cycle. The synchrotron radiation can be channeled through different beamlines for use in research.

Other types of magnets – insertion devices called ''wigglers'' and ''undulators'' – can be assembled in the storage ring. Unlike the bending magnets, the primary purpose of which is to maintain the circular trajectory, wigglers and undulators are used to increase the intensity of the emitted radiation. Bending magnets and wigglers cause a continuous spectrum of radiation. In contrast, the radiation produced by an undulator has a discontinuous spectrum and can be tuned to various wavelengths. The importance of synchrotron radiation for macromolecular crystallography lies in the high brilliance (photons s^{-1} $mrad^{-2}$ mm^{-2} per $\Delta\lambda/\lambda$; that is, how small is the source and how well collimated are the X-rays?) of the beam, the high intensity and the tunability of the wavelength in the relevant range from 0.5 to 3.0 Å. The time structure of the beam is of interest for time-resolved crystallography.[12] The particles circulate in bunches with widths of 50–150 ps and repeat every few microseconds.

About 15 synchrotron radiation facilities equipped with beamlines for macromolecular crystallography are available throughout the world operated at energies from about 1.5 to 6–8 GeV for third-generation machines. An aerial view of the European Synchrotron Radiation Facility (ESRF) in Grenoble, a third-generation machine, is shown in Figure 11. The ESRF storage ring is operated at 6 GeV and has a circumference of 844.39 m. Its critical wavelength, λ_c, is 0.6 Å.

Figure 11 Aerial view of ESRF in Grenoble. Courtesy of ESRF.

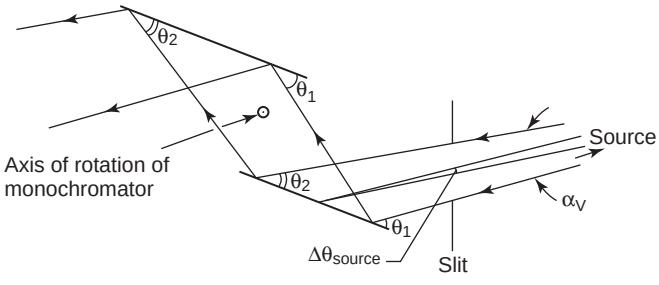

Figure 12 Schematic drawing of a double monochromator system. (Reproduced by permission of Cambridge University Press, from Helliwell.[11])

2.1.3 Monochromators

In the majority of applied diffraction techniques, monochromatic X-rays are used. Therefore, the emitted white radiation of X-rays must be further monochromatized. With copper Kα radiation generated by a sealed or rotating anode tube, the Kβ radiation can be removed with a nickel filter. Much better results can be achieved with a monochromator. The simplest monochromator is a piece of a graphite crystal which reflects the copper Kα radiation at a Bragg angle of 13.1° and a glancing angle of 26.2°. Improved beam focusing is obtained by a double mirror system. The mirror assembly is composed of two perpendicular bent nickel-coated glass optical flats, each with translation, rotation and slit components housed in a helium gas flashed chamber, which is commercially available (Molecular Structure Corporation, The Woodlands, TX, USA). The prototype and basic theory in the use of this system were discussed in detail by Phillips and Rayment.[13]

For synchrotron radiation with its much higher intensity, germanium or silicon single crystals can be applied as monochromators which filter out a bandwidth of $\delta\lambda/\lambda$ from 10^{-4} to 10^{-5}, two orders of magnitude smaller than with graphite. Single or double monochromators can be used which are either flat or bent. The bent monochromators have the advantage that they simultaneously focus the beam. The double monochromator (Figure 12) has the advantage that the emergent monochromatic beam is parallel to and only slightly displaced from the incident synchrotron radiation beam. This makes necessary only small adjustments of the X-ray optics and detector arrangement when it is tuned to another wavelength compared with a single monochromator where the whole X-ray diffraction assembly must be moved.

2.2 Detectors

2.2.1 General Components of an X-ray Diffraction Experiment

A principal arrangement for a macromolecular X-ray diffraction experiment is depicted in Figure 13. The primary beam leaves the X-ray source and passes the X-ray optics, which may be a simple collimator or the various types of monochromators or mirror systems described above and terminated with a collimator. The crystal is mounted on a goniometer head either in a quartz capillary or in a cryo-loop shock frozen at low temperature. The goniometer head is attached to a device which can perform spatial movements of the crystal around the center of the crystal. The simplest kind of such a movement is the rotation of the crystal about a spindle axis as indicated in Figure 13. This device can be a multiple axis goniostat (2–4 axes) which allows the crystal to be brought into any spatial orientation around its center. The

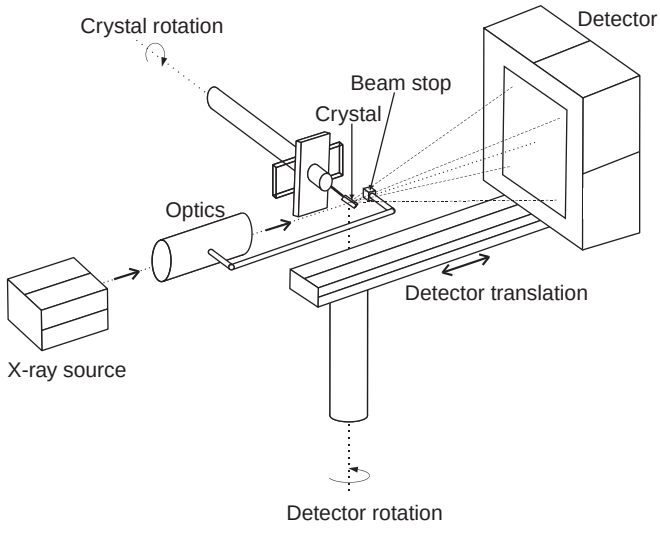

Figure 13 Principal arrangement for a macromolecular X-ray diffraction experiment.

X-ray detector which registers the diffracted intensities is mounted on a device which permits the translation and rotation of the detector. If the active area of the detector is large enough to collect all generated diffracted beams at a given wavelength, detector rotation is not necessary and the detector is arranged normal to the primary beam. A small piece of lead is placed in the path of the primary beam just behind the crystal to prevent damage to the detector and superfluous gas scattering.

The classical detectors in macromolecular crystallography have been photographic films and single-photon counters. The photographic films were used on specially designed X-ray cameras and the single-photon counters on four-circle diffractometers. The main disadvantage of these detectors was their low sensitivity and with films the limited dynamic range (1 : 200). Over the last 15 years, powerful detectors have been developed which will be discussed briefly. These new detectors have almost completely replaced photographic films and single-photon counters.

2.2.2 *Image Plates*

An image plate (IP) consists of a support (either a flexible plastic plate or a metal base) coated with a photostimulable phosphor (150 μm) and a protective layer (10 μm). The photostimulable phosphor is a mixture of very thin crystals of e.g. $BaF(Br,I):Eu^{2+}$ and an organic binder. This phosphor can store a fraction of the absorbed X-ray energy by electrons trapped in color centers. It emits photostimulated luminescence whose intensity is proportional to the absorbed X-ray intensity, when later stimulated by visible light. The wavelength of the photostimulated luminescence ($\lambda \approx 390$ nm) is reasonably separated from that of the stimulating light ($\lambda \approx 633$ nm, in practice a red laser), allowing it to be collected by a conventional high quantum efficiency photomultiplier tube. The output of the photomultiplier is amplified and converted to a digital image, which can be processed by a computer. The residual image on the IP can be erased completely by irradiation with visible light, to allow repeated use.

IPs have several excellent performance characteristics as integrating X-ray area detectors that make them well suited in X-ray diffraction. The sensitivity is at least 10 times higher than for X-ray films and the dynamic range is much broader ($1:10^4-10^5$). Important for synchrotron radiation is their high sensitivity at shorter wavelengths (e.g. 0.65 Å). A disadvantage is the relatively long readout times for each exposure (from 45 s to several minutes). IP diffractomer systems are commercially available from several companies. All systems work reliably and deliver good-quality data. A photograph of the newest IP system produced by Mar Research (Hamburg, Germany) is shown in Figure 14.

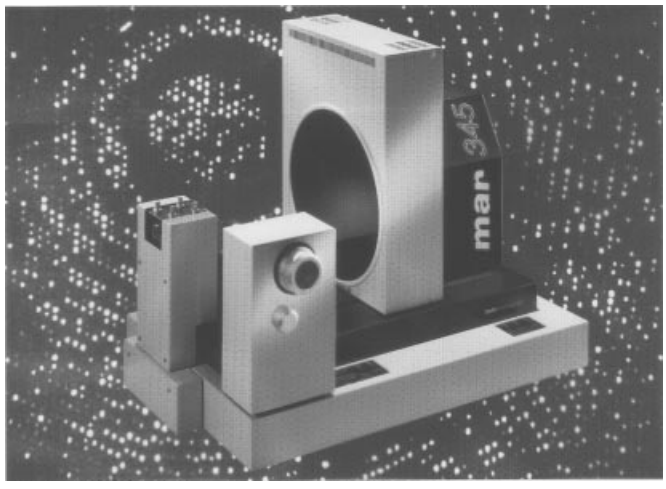

Figure 14 Mar 345 IP diffractometer system from Mar Research, Hamburg, Germany. The circular plate rotates for scanning and the laser is moved along a radial line. Courtesy of Mar Research.

2.2.3 *Gas Proportional Detectors*

As X-ray counters, gas proportional detectors provide unrivaled dynamic range and sensitivity for photons in the range important for macromolecular crystallography (for a review, see Kahn and Fourme[14]). The classical gas proportional detector is a multiwire proportional chamber (MWPC), widely used as an in-house detector with conventional X-ray sources. Two MWPC diffractometer systems are commercially available. Gas proportional detectors use as a first step the absorption of an X-ray photon in a gas mixture high in xenon or argon. This photoabsorption produces one electron–ion pair whose total energy is just the energy of the initial X-ray photon. The ion returns to its neutral state either by emission of Auger electrons or by fluorescence. Since the kinetic energy of these first electrons is far greater than the energy of the first ionization level of the xenon or argon atoms, fast collisions with atoms (or molecules) in the gas very quickly produce a cascade of new electron–ion pairs in a small region extending over a few hundred micrometers around the conversion point. The total number of primary electrons that are produced during this process is proportional to the energy of the absorbed X-ray photon and is thus a few hundred for ~10 keV photons. These primary electrons then drift to the nearest anode wire where an ionization avalanche of 10 000–1 000 000 as many ion pairs results. The motion of the charged particles in this avalanche (chiefly the motion of the heavy positive ions away from the anode wire) causes a negative-going pulse on the anode wire and positive-going pulses on a few of the nearest wires in the back (cathode) wire plane (see Figure 15).

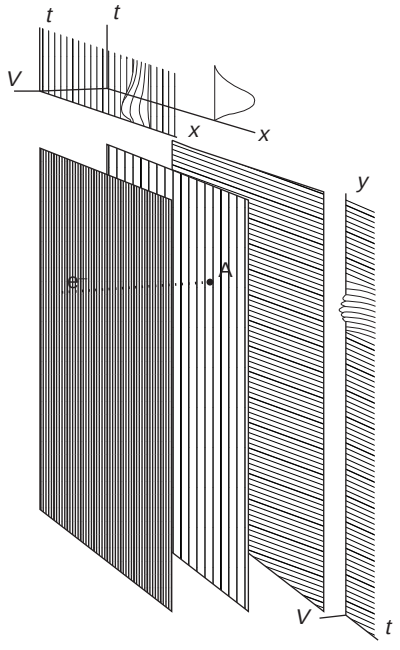

Figure 15 Expanded view of a MWPC showing the anode plane sandwiched between the two cathode planes. A is the position of the avalanche. The centers of the induced charge distributions are used to determine the coordinates, *x* and *y*, of the avalanche. (Reproduced by permission of Academic Press, Inc., from Kahn and Fourme.[14])

Disadvantages of the MWPC detector are the limited counting rate due to the build-up of charges in the chamber and limitations in the readout electronics and the lower sensitivity at shorter wavelengths. This makes the application of MWPCs with synchrotron radiation poorly effective.

2.2.4 Charge-coupled Device-based Detectors

A remarkable development for the use with synchrotron radiation is the design and construction of charge-coupled device (CCD) detectors (for a review, see Westbrook and Naday[15]). CCDs were developed originally as memory devices, but the observation of localized light-induced charge accumulation in CCDs quickly led to their development as imaging sensors. These CCD detectors are integrating detectors like the conventional X-ray sensitive film, IPs and analog electronic detectors using either silicon intensified target (SIT) or CCD sensors. Integrating detectors have virtually no upper rate limits because they measure the total energy deposited during the integration period (although individual pixels may become saturated if the signal exceeds its storage capacity). The first commercially available analog electronic detector was the fast area television detector (FAST) detector produced by Enraf-Nonius (Delft, The Netherlands). This detector contained a SIT vidicon camera as an electronically readable sensor. The SIT vidicon exhibits higher noise than CCDs, which have therefore replaced SIT sensors during the past few years. Because of their high intrinsic noise, detectors with SIT vidicon sensors need an analog image-amplification stage and this limits the overall performance of such detectors. Several CCD detector systems have also been developed that incorporate image intensification. The most important development in detector design for macromolecular crystallography has been the incorporation of scientific-grade CCD sensors into instruments with no image intensifier. These detector designs are based on direct contact between the CCD and a fiber-optic taper. There are several commercial systems available based on this construction (Mar Research, Hamburg Germany; Hamlin Detector).

A schematic representation of such a detector is shown in Figure 16. An X-ray phosphor (commonly $Gd_2O_2S:Tb$) is attached to a fiber-optic faceplate which is tightly connected to a fiber-optic taper. The X-ray sensitive phosphor surfaces at the front convert the incident X-rays into a burst of visible-light photons. Although it is possible to permit the X-rays to strike the CCD directly, this method has several drawbacks, such as radiation damage to the CCD, signal saturation and poor efficiency. The use of a larger phosphor as active detector area and the demagnifying fiber-optic taper is also necessary because the size of the scientific-grade CCD sensors is not as large as needed for the demands of the X-ray diffraction experiment. The fiber-optic taper is then bonded to the CCD which is connected to the electronic readout system. The CCD must be cooled to temperatures ranging from -40° C to -90° C, depending on the various systems. The great advantage of CCD detectors is their short readout time, which lies in the range from 1 to a few seconds.

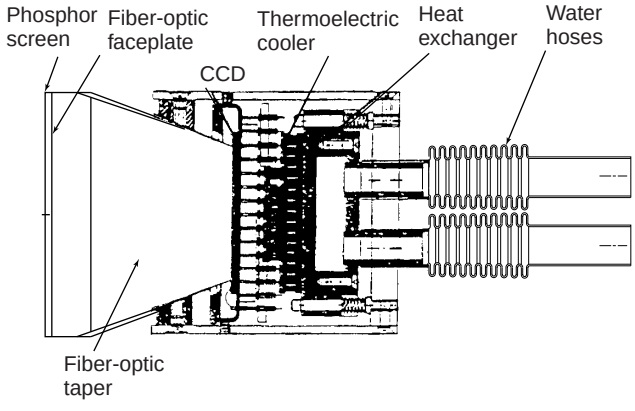

Figure 16 Schematic representation of a CCD/taper detector. (Reproduced by permission of Academic Press, Inc., from Westbrook and Naday.[15])

2.3 Crystal Mounting and Cooling

2.3.1 Conventional Crystal Mounting

The purpose of crystal mounting is to isolate a single crystal from its growth medium so that it can be used in the X-ray diffraction experiment to study its diffraction properties. It is important that the manipulation of the crystal introduce as little damage as possible to its three-dimensional structure. The most important aspect of crystal mounting is to preserve the crystal in its state of hydration. This is accomplished by sealing the crystal in a thin-walled (0.001 mm thick) glass or quartz capillary tube. The important steps in conventional crystal mounting are illustrated in Figure 17(a–c). The crystal must be dislodged from the surface on which it grew, then it may be drawn into the capillary using suction from a small-volume (0.25 mL) syringe, micropipet or mouth aspirator which are connected to the funnel of the capillary by a flexible plastic hose of appropriate diameter. Next, the capillary should be inverted to allow the crystal to fall to the inner meniscus. Then the surrounding solution may be removed using thin strips of filter paper or by a small glass pipet. The extent to which the crystal should be dried must be determined by experience. The final step is to place a small-volume of mother liquor in the capillary and seal both ends. The capillary is then glued to a metal base which can be attached to a goniometer head.

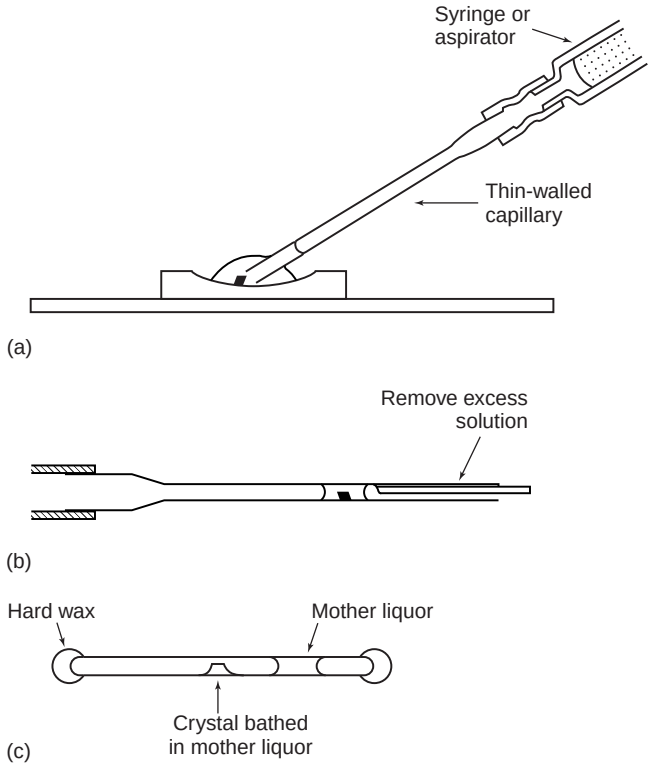

Figure 17 Mounting of a crystal in a glass capillary. (Reproduced by permission of Academic Press, Inc., from I. Rayment, 'Treatment and Manipulation of Crystals', *Methods Enzymol.*, **114**, 136–140 (1985).)

2.3.2 Cryocrystallography

Many macromolecular crystals suffer from radiation damage when exposed to X-rays with energies and intensities as used in macromolecular X-ray diffraction experiments with both conventional sources and synchrotron radiation. A possibility for reducing radiation damage of the crystal during the measurement is to cool the crystal to low temperatures, usually to 100 K (for a review on cryocrystallography, see Rodgers[16]). For this purpose the crystal is flash-frozen to prevent ice formation or damage to the crystal. One method of crystal treatment is the removal of external solution by transferring the crystal in a small drop to a hydrocarbon oil and either teasing the liquid away or drawing it off with filter paper or a small pipet. The oil-coated crystal is then mounted on a glass fiber or small glass "spatula". Oil protects the crystal from drying and acts as an adhesive that hardens on cooling to hold the sample rigidly. Much more frequently used is a technique in which the crystal is suspended in a film of mother liquor in a small loop.[17] This method avoids problems with damage by the oil or mechanical damage when removing the external liquid and it has proven successful for most samples. It does, however, require the use of a cryoprotectant to prevent ice formation. The most commonly used cryoprotectants are glycerol, PEG of different molecular weights, glucose and MPD.

The loop is produced from fine fibers which permit unobstructed data collection in nearly all sample orientations. The crystal is held within the loop, suspended in a thin film of cryoprotectant-containing harvest buffer. The loop is supported by a fine wire or pin, which itself is attached to a steel base used for placing the assembly on a goniometer head and in storing mounted crystals. Once in the loop, the crystal is cooled to a temperature at which the increasing viscosity of the liquid prevents molecular rearrangement. The rate of cooling must be rapid enough to reach this point before ice-crystal nucleation occurs. Two methods are used: cooling directly in the gas stream of a cryostat or plunging the crystal into a cryogenic liquid. The first method is explained in Figure 18(a) and (b). The loop assembly (with crystal) is attached to the goniometer head with the cold stream deflected. Then the cold stream is unblocked to flash-freeze the crystal. In this gas-stream position the goniometer head must be heated to prevent ice formation on the goniometer head. Cryostats and cryocrystallographic tools are commercially available.

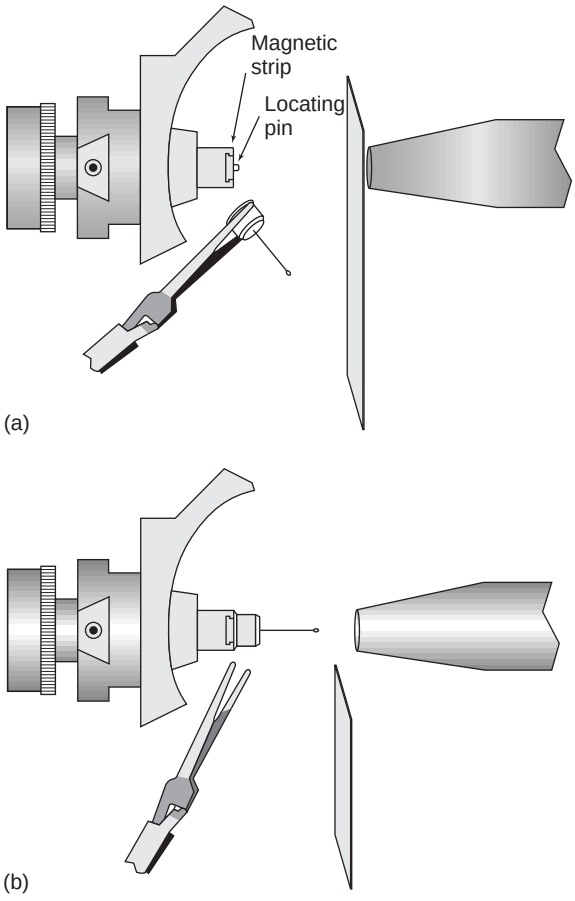

Figure 18 Flash cooling in the direct cold gas stream of a cryostat. (Reproduced by permission of Academic Press, Inc., from D.W. Rodgers.[16])

Cryocrystallography has had a great impact on macromolecular crystallography by dramatically increasing the lifetime of a crystal during the X-ray experiment allowing, for example, the collection of several data sets from one crystal at different wavelengths using synchrotron radiation.

2.4 Data Collection Techniques

2.4.1 Rotation Method

Most macromolecular X-ray diffraction systems use the rotation method for data collection [a detailed discussion of the rotation method is given by Arndt and Wonnacott[18]]. For each crystal a reciprocal lattice can be constructed which is very useful in the interpretation of crystallographic crystal diffraction experiments. Diffraction theory (discussed later) tells us that an X-ray reflection is generated when a point of this reciprocal lattice lies on a sphere of radius $1/\lambda$ whose origin is $1/\lambda$ away from the origin of the reciprocal lattice in the direction of the primary beam (Figure 19). The direction of

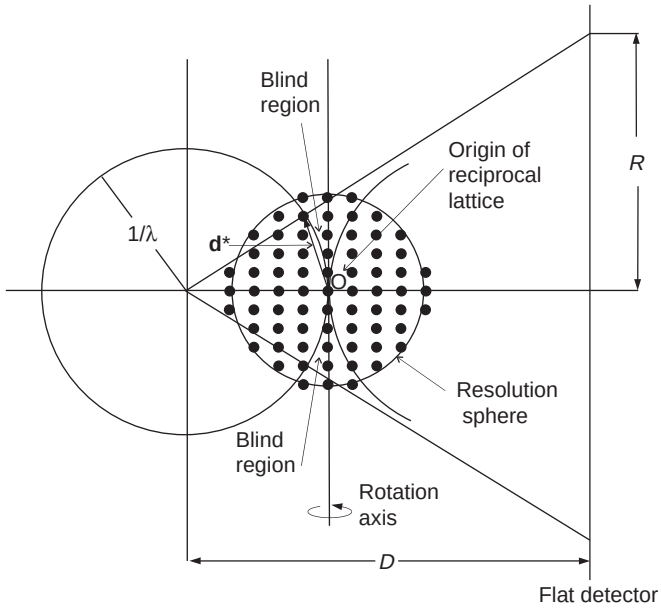

Figure 19 Diffraction geometry in the rotation method usually applied in macromolecular X-ray diffraction systems.

such a diffracted beam is along the connection of the center of the so-called Ewald sphere (radius $1/\lambda$) and the intersection of the reciprocal lattice point on the Ewald sphere. Owing to certain factors which will be discussed later, the apparent reciprocal lattice extends to a given radius only which defines the resolution sphere. To bring all reciprocal lattice points within the resolution sphere into the reflection position, the crystal must be rotated around its center. Nearly all macromolecular X-ray diffraction systems apply the rotation technique in the normal beam case where the rotation axis is normal to the incident X-ray beam. Rotating the crystal around 360° brings all reciprocal lattice points within the resolution sphere in the reflection position except for the region between the rotation axis and the Ewald sphere, which is therefore called the blind region. This region can be collected when the crystal has been brought into another orientation. The diffracted beams are usually registered with a flat detector at distance D from the crystal which is also normal to the primary beam. To avoid overlapping of reflection spots on the detector, the crystal is rotated by rotation angle increments which can vary from tenths of degrees to 1 to a few degrees depending on the size of the crystal unit cell, crystal mosaicity, beam collimation and other factors. Each individual exposure is processed and electronically stored in a computer. These raw data images are evaluated subsequently with relevant computer programs (discussed in some detail later) to give the intensities and geometric reference values (indices) for each collected intensity.

2.4.2 Precession Method

The rotation method delivers a distorted image of the reciprocal lattice for each geometry of the detector (flat or curved) and orientation with respect to the rotation axis. An undistorted image of the reciprocal lattice can be obtained by using the precession method.[19] The principle of this technique is shown in Figure 20. The detector is a flat film. During the motion of a given reciprocal lattice plane (in Figure 20 a so-called zeroth plane going through the origin O of the reciprocal lattice) the flat detector must always be parallel to this reciprocal lattice plane to obtain an undistorted image of this plane. The normal of the reciprocal lattice plane and consequently also the detector are inclined with respect to the primary X-ray beam by an angle μ. When the normals of the reciprocal lattice plane and the detector carry out a concerted precession motion of angle μ around the primary X-ray beam, a circular region of the reciprocal lattice plane is registered on the detector (these regions are shown as dashed circles in Figure 20). In a precession camera construction, the crystal and the film cassette are both held in a universal joint which are linked so that the film and crystal move together in phase with precession angle μ. In Figure 20 the joints are symbolized as forks and their linkage by a line. Parallel to the zeroth reciprocal lattice plane is a set of lattice planes that also carry out this precession movement. The parts of them that are swung through the Ewald sphere also give rise to an image on the flat detector (first and second reciprocal lattice layers are also indicated in Figure 20). The images of these layers would superimpose on the detector and use of the technique in this way is denoted the screenless precession method. The insertion of a screen with a suitable annular aperture between the crystal and the detector at an appropriate distance can be used to screen out the desired reciprocal layer. This screen is also inclined by the precession angle μ and is coupled to the concerted precession movement of crystal and film cassette. The strength of the precession technique is that in addition to the undistorted imaging of the reciprocal lattice planes, the indexing of the diffraction spots is straightforward and the symmetry of the diffraction pattern is readily obtained by inspection. For this reason, the precession method has been broadly applied in macromolecular crystallography for a long time. The replacement of film by the new generation of detectors has almost completely stopped the use of the precession method. Nevertheless, the use of a precession camera has great teaching benefits in becoming familiar with the reciprocal lattice concept.

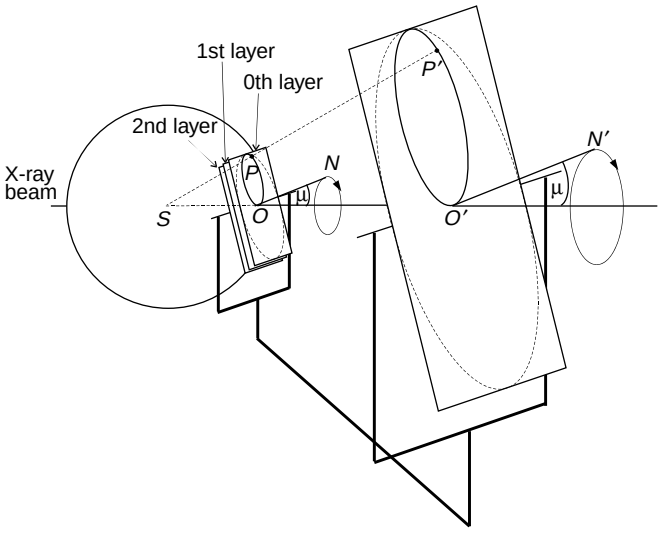

Figure 20 Principle of the precession method [modified from Buerger[19]].

3 PRINCIPLES OF X-RAY DIFFRACTION BY A CRYSTAL

3.1 Scattering of X-rays by an Atom

A component of the electrical field of the incident wave has the following form (Equation 6) in free space referred to an origin at $\mathbf{x} = 0$:

$$E(\mathbf{x}) = E_0 \exp[2\pi i(\nu t - \mathbf{x}\mathbf{s}_0)] \tag{6}$$

where $\mathbf{s}_0$ = wave vector of incident wave, $\nu = c/\lambda$ = frequency, λ = wavelength and $s_0 = 1/\lambda$ = absolute value of incident wave vector. This wave interacts with a scattering center at position **r** (Figure 21). The electric field component of the incident wave causes the electron at position **r** to oscillate. Together with the positively charged nucleus of the atom, which does not oscillate, this can be considered as a classical dipole-oscillator. This dipole-oscillator emits a spherical wave which is denoted as a scattered wave and can be given in the following form

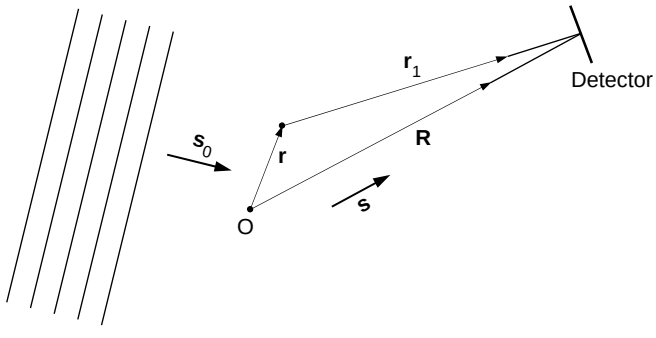

Figure 21 Scattering of a planar X-ray wave by an electron at position **r** with respect to origin O.

(Equations 7 and 8):

$$E_{sc} = CE(\mathbf{r})\frac{e^{-2\pi i s_0 r_1}}{r_1} \tag{7}$$

$$E_{sc} = CE_0 e^{-2\pi i \mathbf{s}_0 \mathbf{r}}\frac{e^{2\pi i(\nu t - s_0 r_1)}}{r_1} \tag{8}$$

The phase-angle-dependent factor has been omitted. The amplitude of the scattered wave is proportional to the amplitude $E(\mathbf{x})$ of the wave incident at $\mathbf{r}$. This gives the factor $E(\mathbf{r})$. C is a proportionality factor taking into account peculiarities of the scattering center. The factor $1/r_1$ considers the conservation of the scattered energy flux. We add all wavelets scattered from different volume elements of an atom to get the total amplitude of the scattered wave at point $\mathbf{R}$ relative to the origin O in the atom.

Equations (9–12) follow from Figure 21:

$$\mathbf{r} + \mathbf{r_1} = \mathbf{R} \tag{9}$$

$$r_1{}^2 = (\mathbf{R} - \mathbf{r})^2 = R^2 + r^2 - 2rR\cos(\mathbf{r}, \mathbf{R}) \tag{10}$$

$$r_1 = R\left[1 - \frac{2r}{R}\cos(\mathbf{r}, \mathbf{R}) + \left(\frac{r}{R}\right)^2\right]^{1/2} \tag{11}$$

$$r_1 \approx R\left[1 - \frac{r}{R}\cos(\mathbf{r}, \mathbf{R}) + \cdots\right] \tag{12}$$

where $(r/R)^2$ and higher terms were neglected in the expansion of the square root. It follows that (Equation 13):

$$r_1 \approx R - r\cos(\mathbf{r}, \mathbf{R}) \tag{13}$$

Now we can combine the spatial phase factors in the equation for E_{sc}. With the approximation for r_1 we obtain Equation (14):

$$e^{-2\pi i(\mathbf{s}_0\mathbf{r} + s_0 r_1)} = e^{-2\pi i s_0 R} e^{-2\pi i[\mathbf{s}_0\mathbf{r} - s_0 r\cos(\mathbf{r},\mathbf{R})]} \tag{14}$$

As $s_0 = s$ and $\mathbf{s}$ parallel to $\mathbf{R}$, we obtain Equation (15):

$$s_0 r\cos(\mathbf{r}, \mathbf{R}) = sr\cos(\mathbf{r}, \mathbf{s}) = \mathbf{sr} \tag{15}$$

It follows for the phase factor (Equation 16):

$$\begin{aligned} e^{-2\pi i(\mathbf{s}_0\mathbf{r} + s_0 r_1)} &= e^{-2\pi i s_0 R} e^{-2\pi i(\mathbf{s}_0\mathbf{r} - \mathbf{sr})} \\ &= e^{-2\pi i s_0 R} e^{2\pi i[\mathbf{r}(\mathbf{s} - \mathbf{s}_0)]} \end{aligned} \tag{16}$$

Now we can write for the wave scattered by a center at $\mathbf{r}$ (Equation 17):

$$E_{sc}(r_1) = \left(\frac{CE_0 e^{-2\pi i s_0 R} e^{2\pi i \nu t}}{R}\right) e^{2\pi i \mathbf{r}(\mathbf{s} - \mathbf{s}_0)} \tag{17}$$

In a useful approximation, r_1 has been replaced by R in the denominator.

In general, the scattering from an atom comes from the distribution of electrons in the atom. If the scattering of a volume element dv of the atom is proportional to the local electron density $\rho(\mathbf{r})$, then the scattering amplitude will be proportional to the integral (Equation 18):

$$f(\mathbf{S}) = \int_{\text{vol. of atom}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{rS})\, dv \tag{18}$$

with $\mathbf{S} = \mathbf{s} - \mathbf{s_0}$.

3.2 Scattering of X-rays by a Unit Cell

A unit cell may contain N atoms at positions of their internal origins at $\mathbf{r}_j$ $(j = 1, 2, 3, \ldots, N)$ with respect to the origin of the unit cell (Figure 22). For atom 1 we obtain Equation (19):

$$\mathbf{f}_1 = \int_{\text{vol. of atom}} \rho(\mathbf{r}) \exp[2\pi i(\mathbf{r}_1 + \mathbf{r})\mathbf{S}]\, dv = f_1 \exp(2\pi i \mathbf{rS}) \tag{19}$$

with (Equation 20)

$$f_1 = \int_{\text{vol. of atom}} \rho(\mathbf{r}) \exp(2\pi i \mathbf{r}_1 \mathbf{S})\, dv \tag{20}$$

where f_1 is the atomic form factor for atom 1. It reflects the characteristics of the scattering of the individual atoms and is real if the wavelength of the incident X-ray is not close to an absorption edge of the atom. The atomic form factor f is equal to Z, the ordinary number of the scattering atom, at a diffraction angle of 0° and decreases with increasing diffraction angle.

For N atoms this adds up to the total scattered wave of a unit cell $\mathbf{G}(\mathbf{S})$ (Figure 23) according to Equation (21):

$$\mathbf{G}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{r}_j \mathbf{S}) \tag{21}$$

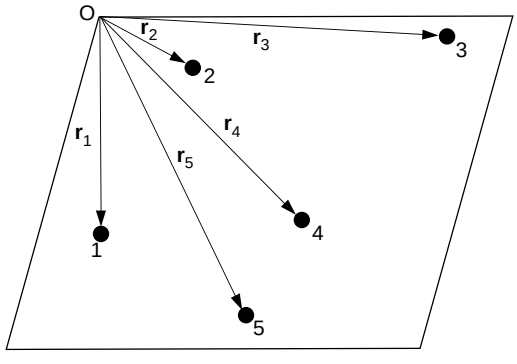

Figure 22 Atomic positions in a unit cell.

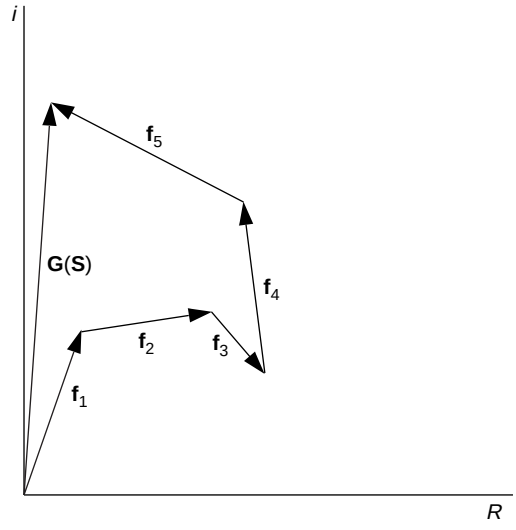

Figure 23 Vector diagram for the total scattered wave in direction **S** added up for five atoms.

3.3 Scattering of X-rays by a Crystal

3.3.1 One-dimensional Crystal

In a one-dimensional crystal the unit cells are separated by the unit cell vector **a**. The contribution of the scattered wave from the unit cell at the origin of the crystal is **G**(**S**). All scatterers in the second unit cell are displaced by the vector **a** relative to the origin, which introduces a corresponding phase factor and reveals for the second unit cell relative to the origin $\mathbf{G}(\mathbf{S})\exp 2\pi i\mathbf{aS}$. For the nth unit cell relative to the origin we obtain $\mathbf{G}(\mathbf{S})\exp 2\pi i(n-1)\mathbf{aS}$.

This sums up for the total wave to Equation (22):

$$\mathbf{F}(\mathbf{S}) = \sum_{n=1}^{T} \mathbf{G}(\mathbf{S}) \exp 2\pi i(n-1)\mathbf{aS} \qquad (22)$$

Generally, **F**(**S**) is of the same order of magnitude as **G**(**S**) and no strong scattering effect is observed (Figure 24a). However, when $2\pi\mathbf{aS} = 2\pi h$ or an integral multiple of 2π or $\mathbf{aS} = h$ (h is an integer), the waves add up constructively to a scattered wave proportional to $T|\mathbf{G}(\mathbf{S})|$ (Figure 24b). $T = 10^5$ for a 1 mm long crystal with a 100 Å lattice constant. The intensity distribution of the scattered waves is concentrated around the values where **aS** is equal to an integer and depends on the number of contributing unit cells. The more unit cells are contributing, the sharper is the concentration of the intensity around these values.

3.3.2 Three-dimensional Crystal

In this case, the unit cell is spanned by the unit cell vectors **a**, **b** and **c** and is repeated periodically by

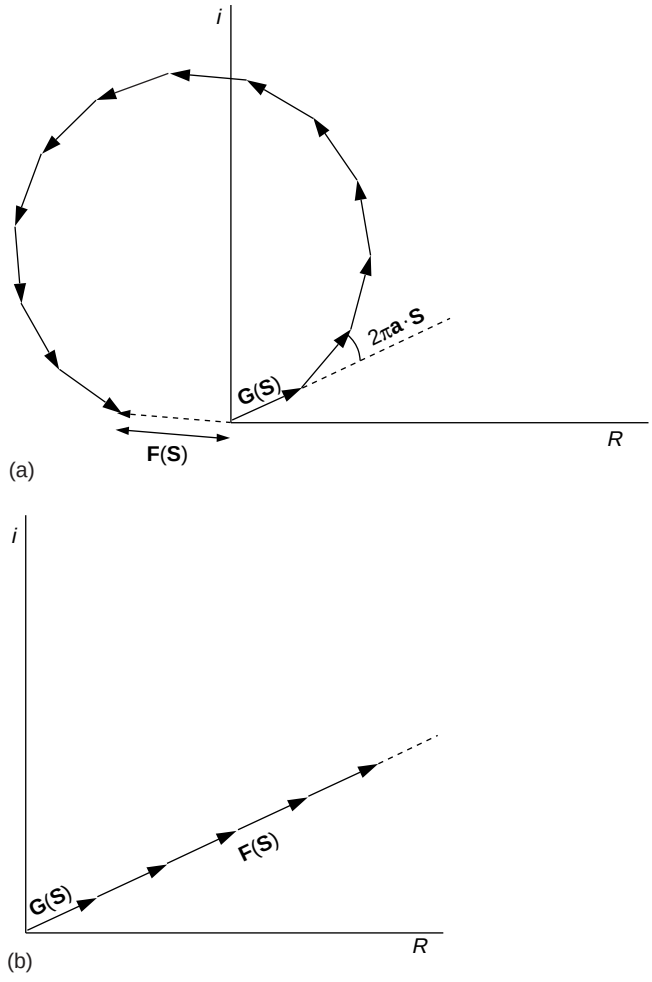

Figure 24 Vector diagrams displaying the total wave scattered by a molecule in a crystal. (a) The phase differences between waves scattered by adjacent unit cells is $2\pi\mathbf{aS}$, (b) the phase difference is an integral multiple of 2π (adapted from Blundell and Johnson[20]).

the corresponding vector shifts in the respective spatial directions. This means that we will obtain scattered waves of measurable intensities when the three subsequent conditions are fulfilled (Equation 23):

$$\mathbf{aS} = h; \quad \mathbf{bS} = k; \quad \mathbf{cS} = l \qquad (23)$$

These conditions are known as Laue equations.

If we neglect the proportionality constant T we obtain Equation (24) for the total scattered wave for a three-dimensional crystal with a unit cell containing N atoms:

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp 2\pi i\mathbf{r}_j\mathbf{S} \qquad (24)$$

with (Equation 25)

$$\mathbf{r}_j = \mathbf{a}x_j + \mathbf{b}y_j + \mathbf{c}z_j \qquad (25)$$

Hence we have (Equations 26 and 27)

$$\mathbf{r}_j\mathbf{S} = x_j\mathbf{aS} + y_j\mathbf{bS} + z_j\mathbf{cS} = hx_j + hy_j + lz_j \quad (26)$$

(from Laue's equation) and

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp 2\pi i(hx_j + ky_j + lz_j) = |\mathbf{F}(hkl)|e^{i\alpha(hkl)} \quad (27)$$

with $|\mathbf{F}(hkl)|$ – amplitude and α – phase angle. We obtain the intensity of the scattered wave as the structure factor $\mathbf{F}(hkl)$ multiplied by its complex conjugate value according to (Equation 28):

$$I(hkl) = \mathbf{F}(hkl)\mathbf{F}^*(hkl) = |\mathbf{F}(hkl)|^2 \quad (28)$$

3.4 The Reciprocal Lattice and Ewald Construction

In section 2.4 we mentioned the usefulness of the concept of the reciprocal lattice in understanding the diffraction of X-rays from a crystal. Now we have the necessary relations for the derivation of the reciprocal lattice. One can write the scattering vector **S** as Equation (29):

$$\mathbf{S} = h_x\mathbf{a}^* + k_y\mathbf{b}^* + l_z\mathbf{c}^* \quad (29)$$

where **S** is a vector in reciprocal space with the metric $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$. The relation to the direct space with metric **a**, **b**, **c** is still unknown. The vector **S** must obey the Laue equations (Equation 30):

$$\mathbf{aS} = \mathbf{a}(h_x\mathbf{a}^* + k_y\mathbf{b}^* + l_z\mathbf{c}^*) = h = h_x\mathbf{aa}^* + k_y\mathbf{ab}^* + l_z\mathbf{ac}^* = h \quad (30)$$

This is fulfilled only when $\mathbf{aa}^* = 1$, $h_x = h$ and $\mathbf{ab}^*$ and $\mathbf{ac}^* = 0$. Similar equations can be derived for the other two Laue conditions. Thus, vector **S** is a vector of a lattice in reciprocal space. The relation between direct and reciprocal lattice is given by the following set of nine equations (Equations 31–39):

$$\mathbf{aa}^* = 1 \quad (31)$$

$$\mathbf{ba}^* = 0 \quad (32)$$

$$\mathbf{ca}^* = 0 \quad (33)$$

$$\mathbf{ab}^* = 0 \quad (34)$$

$$\mathbf{bb}^* = 1 \quad (35)$$

$$\mathbf{cb}^* = 0 \quad (36)$$

$$\mathbf{ac}^* = 0 \quad (37)$$

$$\mathbf{bc}^* = 0 \quad (38)$$

$$\mathbf{cc}^* = 1 \quad (39)$$

It follows from these that $\mathbf{a}^* \perp \mathbf{b}, \mathbf{c}$; $\mathbf{b}^* \perp \mathbf{a}, \mathbf{c}$; $\mathbf{c}^* \perp \mathbf{a}, \mathbf{b}$; and vice versa. The metric relations can also be derived from these relations. They are given explicitly e.g. in Drenth.[21] It means that the inverse lattice vectors are perpendicular to the plane which is spanned by the two other non-inverse lattice vectors.

Bragg's law can be derived now by inspection of Figure 25. The wave vectors for the incident wave $\mathbf{s}_0$ and the scattered wave **s** have the same absolute value of $1/\lambda$. Vector **S** must be a vector of the reciprocal lattice and its absolute value is equal to d^*. From Figure 25 we obtain Equations (40–42):

$$\sin\theta = \frac{d^*}{2}\lambda \quad (40)$$

$$\lambda = \frac{2\sin\theta}{d^*} \quad (41)$$

$$\lambda = 2d\sin\theta \quad \text{for } n = 1 \quad (42)$$

The general equation for Bragg's law is Equation (43):

$$2d\sin\theta = n\lambda \quad (43)$$

where n is the order of reflection and d the interplanar distance in the direct lattice.

The Ewald construction is contained in Figure 19. A sphere of radius $1/\lambda$ is drawn. The origin of the reciprocal lattice is located where the wave vector $\mathbf{s}_0$ ends on the Ewald sphere. A diffracted beam is generated if a reciprocal lattice vector $\mathbf{d}^*_{hkl}$ with an absolute value of $1/d_{hkl}$ cuts the Ewald sphere. The beam is diffracted in the direction of the connection of the origin of the

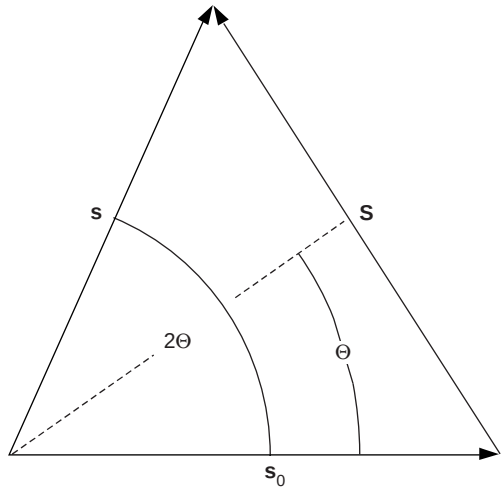

Figure 25 Geometric representation of diffraction geometry, 2Θ, glance angle; Θ, Bragg angle.

Ewald sphere and the intersection point of the reciprocal lattice point on the Ewald sphere. The diffraction pattern of a lattice is itself a lattice with reciprocal lattice dimensions.

3.5 The Temperature Factor

The thermal motion of the atoms causes a decrease of the scattering power by a factor of $\exp[-B(\sin^2\theta/\lambda^2)]$ with (Equation 44)

$$B = 8\pi^2\bar{u}^2 \tag{44}$$

where $\bar{u}$ is the mean displacement of the atoms due to the thermal motion. The atomic scattering factor f must be multiplied with this factor. In this model the thermal motion has been assumed to be isotropic. Therefore, B is denoted the isotropic temperature factor. In crystals this is usually not the case and the thermal motion is described by a tensor ellipsoid. Here we obtain a set of six independent anisotropic temperature factors. In protein crystallography, isotropic B values for each atom of the molecules are used normally. The thermal motion of the atoms is one main reasons for the fall-off of the diffraction intensity especially at higher diffraction angles. This limits the possible recordable number of diffraction spots and, as will be seen later, the resolution of the diffraction experiment.

3.6 Symmetry in Diffraction Patterns

An X-ray diffraction data set from a crystal represents its reciprocal lattice with the corresponding diffraction intensities at the reciprocal lattice points (hkl). As the reciprocal lattice is closely related to its direct partner it reveals symmetries, lattice properties and other peculiarities (e.g. systematic extinctions) that are connected to the direct crystal symmetry such as unit cell dimensions and space group. A detailed discussion of this problem is given in Buerger.[22]

In the case of real atomic scattering factors f the diffraction intensities are centrosymmetric according to Friedel's law (Equation 45):

$$I(hkl) = I(\overline{hkl}) \tag{45}$$

This is illustrated in Figure 26(a) and (b). The square of a complex number is the product of this number by its complex conjugate. This is shown for $\mathbf{F}(hkl)$ in Figure 26(a) and for $\mathbf{F}(\overline{hkl})$ in Figure 26(b). The resulting intensities are equal in both cases.

3.7 Electron Density Equation and Phase Problem

Inspection of the equation for the structure factor (Equation 46):

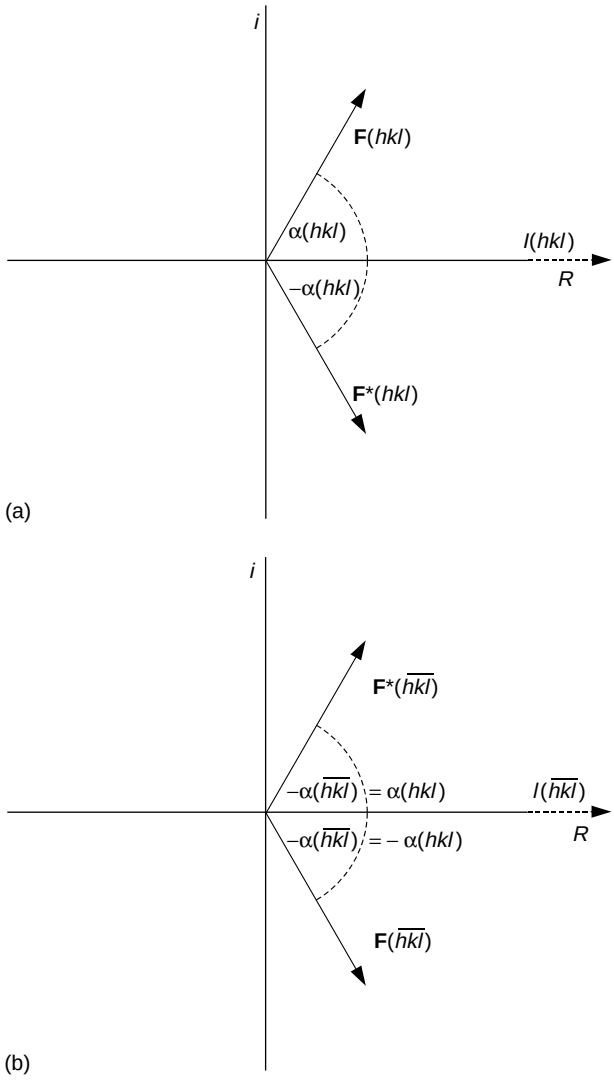

Figure 26 Diagram illustrating the basis of Friedel's law.

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{N} f_j \exp 2\pi i \mathbf{r}_j \mathbf{S}$$

$$= \int_{\text{vol. of unit cell}} \rho(\mathbf{r}) \exp 2\pi i \mathbf{r}\mathbf{S}\, \mathrm{d}v \tag{46}$$

shows that it is the Fourier transform of the electron density $\rho(\mathbf{r})$.

The electron density $\rho(\mathbf{r})$ is then the inverse Fourier transform of the structure factor $\mathbf{F}(\mathbf{S})$ according to Equation (47):

$$\rho(\mathbf{r}) = \int_{\text{vol. of diffraction space}} \mathbf{F}(\mathbf{S}) \exp -2\pi i \mathbf{r}\mathbf{S}\, \mathrm{d}v_s \tag{47}$$

The integration is replaced by summation since $\mathbf{F}(\mathbf{S})$ is not continuous and is nonzero only at the reciprocal lattice

points. Hence we have (Equation 48)

$$\rho(xyz) = \frac{1}{V}\sum_{h=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\mathbf{F}(hkl) \times \exp[-2\pi i(hx+ky+lz)] \quad (48)$$

Knowing the structure factors (Equation 49):

$$\mathbf{F}(hkl) = |\mathbf{F}(hkl)|e^{i\alpha(hkl)} \quad (49)$$

one can calculate the electron density distribution in the unit cell and thus determine the atomic positions of the scattering molecule(s). Unfortunately, the measured quantities are only the absolute values $|\mathbf{F}(hkl)|$ of the structure factor. Information on the phase angles $\alpha(hkl)$ is lost during the diffraction experiment. The determination of these phases is the basic problem in any crystal structure determination and we discuss the methods for solving the phase problem later.

3.8 The Patterson Function

The measured X-ray intensities are proportional to the square of the absolute value of the structure factor according to Equation (28). Would it be possible to use the intensities directly to calculate from these a function which contains structural information? The answer is ''yes''. If one calculates a convolution of the electron density with itself Patterson[23] showed that this is just the Fourier transform of the intensities (Equation 50):

$$P(uvw) = \int_{\text{Vol. of unit cell}} \rho(xyz)\rho(x+u, y+v, z+w)\,dv \quad (50)$$

From this it follows that (Equation 51):

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} F_{\mathbf{h}}^2 \exp -2\pi i\mathbf{hu} \quad (51)$$

The function $P(\mathbf{u})$ will have maxima if the positions $\mathbf{x}$ and $\mathbf{x}+\mathbf{u}$ correspond to atoms. Thus, we obtain a function that contains the interatomic vectors as maxima. We expect N^2 peaks for N atoms. The maxima are proportional to Z_iZ_j. The Patterson function is a very useful tool to locate atoms when the number of atoms in the asymmetric unit of the unit cell is not too high (e.g. <20) or it contains a subset of heavy atoms among not too many (e.g. <100) light atoms such as C, N, O or S. Here, the heavy atom–heavy atom vectors are clearly prominent. If a protein with 1000 or a multiple of that light atoms holds one or several heavy atoms per molecule, the signal resulting from the heavy atoms can no longer be resolved. However, when using the method of isomorphous replacement (discussed later), a Patterson function of the heavy atom structure can be calculated, from which it is possible to locate the heavy atoms.

3.9 Integrated Intensity Diffracted by a Crystal

Real crystals are not perfect. They can be regarded as consisting of small blocks of perfect crystals (sizes in the range of 0.1 μm) which have an average tilt angle among each other of 0.1–0.5° for protein crystals and which diffract independently of each other. Such a real crystal is denoted a mosaic crystal. The total energy, $E_{\mathbf{h}}$, in a diffracted beam for a mosaic crystal rotating with uniform angular velocity w through the reflecting position in a beam of X-rays of incident intensity I_0 is given by Darwin's equation (Equation 52):

$$E(\mathbf{h}) = \frac{I_0}{w}\lambda^3\frac{e^4}{m^2c^4}p\frac{LAV_x}{V^2}|\mathbf{F}(\mathbf{h})|^2 \quad (52)$$

where λ = wavelength of X-rays, e = electronic charge, m = mass of electron, c = velocity of light, p = polarization factor, L = Lorentz factor (geometrical factor taking into account the relative time each reflection spends in the reflection position), A = absorption factor, V = volume of the unit cell and V_x = irradiated crystal volume.

Owing to the mosaicity (0.1–0.5°), each reflection has a corresponding reflection width. The integrated intensity equation is valid under the assumption that apart from ordinary absorption the incident intensity, I_0, is constant within the crystal (kinematic theory of X-ray diffraction) and the mosaic blocks are so small that no multiple scattering occurs within an individual mosaic block. The integrated intensity depends on λ to the third power. Increasing the wavelength causes appreciably stronger diffraction intensities but is accompanied by larger absorption. Cu Kα radiation with a wavelength of 1.5418 Å is an optimal choice for protein crystallography when using X-ray generator sources. Also important is the dependence of the integrated intensity on the unit cell volume V by its negative second power. Doubling of the unit cell volume with twice as many molecules, taking into account the increase in $|\mathbf{F}(\mathbf{h})|^2$ by having now $2n$ molecules per unit cell, reduces the average intensity for the reflected beams by a factor of two.

In Equation (52) $(\lambda^3/wV^2) \times (e^4/m^2c^4) \times V_x \times I_0$ is a constant for a given experiment. The corrected intensity on a relative scale $I(\mathbf{h})$ is obtained from Equation (53):

$$I(\mathbf{h}) = \frac{E(\mathbf{h})}{p \times L \times A} \quad (53)$$

3.10 Intensities on an Absolute Scale

The corrected intensity on a relative scale $I(\mathbf{h})$ can be converted to an intensity given by Equation (54):

$$I(\text{abs}, \mathbf{h}) = \mathbf{F}(\mathbf{h})\mathbf{F}(\mathbf{h})^* = |\mathbf{F}(\mathbf{h})|^2 \quad (54)$$

on an absolute scale by applying a so-called Wilson plot. The basis for this plot is an equation which connects the average intensity on an absolute scale with the average intensity on a relative scale by a scale factor C and considers the isotropic thermal motion of the scattering atoms by the temperature factor given in Equation (44). This is written in the form of Equation (55):

$$\ln \frac{\overline{I(\mathbf{h})}}{\sum_j (f_j)^2} = \ln C - 2B\frac{\sin^2\theta}{\lambda^2} \tag{55}$$

This is the equation of a straight line. B, the overall temperature factor, and C, the scale factor, can be obtained by plotting $\ln \overline{I(\mathbf{h})}/\ _j(f_j)^2$ against $(\sin^2\theta)/\lambda^2$.

3.11 Resolution of the Structure Determination

The concept of resolution in X-ray diffraction has the same meaning as the concept in image formation in the optical microscope.

After the Abbe theory, we obtain Equation (56):

$$d_m = \frac{\lambda}{2NA} \tag{56}$$

where NA is the numerical aperture of the objective lens. In protein crystallography, the nominal resolution of an electron density map is expressed in d_m, the minimum interplanar spacing for which Fs are included in the Fourier series. The maximum attainable resolution at a given wavelength is $\lambda/2$. For copper Kα radiation it is 0.7709 Å and would suffice to determine protein structures at atomic resolution (the distance of a carbon–carbon single bond is about 1.5 Å). However, usually the thermal vibrations of the atoms in a protein crystal are so high that the diffraction data cannot be observed to the full theoretical resolution limit. The polypeptide chain fold can be determined at a resolution of better than 3.5 Å. A medium-resolution structure is in the resolution range of 3.0–2.2 Å and makes the amino acid side-chains clearly visible. A high-resolution structure has a nominal resolution better than 2.2 Å and can be as good as 1.2 Å. In such structures the main-chain carbonyl oxygens become visible as prominent bumps and at a resolution better than 2.0 Å aromatic side-chains acquire a hole in the middle of their ring systems. For some very well diffracting crystals from small proteins, diffraction data extending to resolutions below 1.2 Å could be collected with synchrotron radiation.[24] Such structures reveal real atomic resolution where each atom is visible as an isolated maximum in the electron density map.

3.12 Diffraction Data Evaluation

The analysis and reduction of diffraction data from a single crystal consists of seven main steps: (1) visualization and preliminary analysis of the raw, unprocessed data; (2) indexing of the diffraction patterns; (3) refinement of the crystal and detector parameters; (4) integration of the diffraction spots; (5) finding the relative scale factors between measurements; (6) precise refinement of crystal parameters using the whole data set; and (7) merging and statistical analysis of the measurements related by space-group symmetry. When using electronic area detectors with short read-out times such as CCD or MWPC detectors it is possible to collect diffraction images with small rotational increments (0.05–0.2°). In this case, the reflection profile over the crystal rotation angle can be registered giving a three-dimensional picture of the spot. The evaluation of such diffraction data can be done with computer programs MADNES,[25] XDS,[26,27] the San Diego programs[28] and related programs XENGEN[29] and X-GEN. IP systems with their longer read-out times are operated in a film-like mode with rotational increments of 0.5–2.0°. Here, mainly the program systems MOSFLM[30] and Denzo[31] are applied. The most important developments in the data evaluation of macromolecular diffraction measurements are autoindexing,[26,32] profile fitting,[33,34] transformation of data to a reciprocal-space coordinate system[27,32] and demonstration[35] that a single rotation image contains all of the information necessary to derive the diffraction intensities from that image.

Scaling, merging and statistical analysis of the intensity data are either done with corresponding programs of the CCP4 program suite[36] or with Scalepack.[37] The principles of these operations are given in the manuals for these programs. With modern data-collection methods, the completeness should approach 100% (including the low-resolution data which are very important for molecular replacement), the ratio $I/\sigma(I)$ should be significant even for the highest resolution shell and undue emphasis should not be given to the reliability factor for merging the data (R-merge) unless factors such as multiplicity are taken into account. Nowadays it is customary using synchrotron radiation techniques and fast CCD detectors to collect as much data as possible (before radiation damage becomes significant) in order to produce good statistics.

3.13 Solvent Content of Protein Crystals

The Matthews parameter V_M, which is defined according to Equation (57):

$$V_M = \frac{V_{\text{unit cell}}}{M_{\text{Prot}}} \tag{57}$$

where $V_{\text{unit cell}}$ is the volume of the unit cell and M_{Prot} is the molecular mass of the protein in the unit cell, has values that are in the range 1.6–3.5 $\text{Å}^3\,\text{Da}^{-1}$ for proteins. This allows a rough estimation of the number of molecules in the unit cell.

Furthermore, V_{M} can be used for the assessment of the solvent content of a protein crystal. Calling V_{Prot} the crystal volume occupied by the protein, V'_{P} its fraction with respect to the total crystal volume V and M_{Prot} the mass of protein in the cell, we obtain Equation (58):

$$V'_P = \frac{V_{\text{Prot}}}{V} = \frac{V_{\text{Prot}}/M_{\text{Prot}}}{M_{\text{Prot}}/V} \tag{58}$$

The first term is the specific volume of the protein, the second the reciprocal of V_{M} and, remembering that the molecular weight is expressed in daltons, we have Equation (59):

$$V'_P = \frac{1.6604}{d_{\text{Prot}} V_{\text{M}}} \tag{59}$$

Taking 1.35 g cm^{-3} as the protein density, we obtain as a first approximation Equations (60) and (61):

$$V'_{\text{P}} \approx \frac{1.23}{V_{\text{M}}} \tag{60}$$

$$V'_{\text{Solv}} = 1 - V'_{\text{P}} \tag{61}$$

The solvent content in a protein crystal may vary from 75 to 40%.

4 METHODS FOR SOLVING THE PHASE PROBLEM

4.1 Isomorphous Replacement

4.1.1 Preparation of Heavy Metal Derivatives

If one can attach one or several heavy metal atoms at defined binding site(s) to the protein molecules without disturbing the crystalline order, one can use such isomorphous heavy atom derivatives for the phase determination. The lack of isomorphism can be monitored by a change in the unit cell parameters compared with the native crystal and a deterioration of the quality of the diffraction pattern. The preparation of heavy atom derivatives is undertaken by soaking the crystals in mother liquor containing the dissolved heavy metal compound. Soaking times may be in the range from several minutes to months. Concentrations of the heavy metal compound may vary from tenths of millimolar to 50 mM.

Favorite heavy atoms are Hg, Pt, U, Pb, Au, rare earth metals, etc. Potential ligands can be classified as hard and soft ligands according to Pearson.[38] Hard ligands are electronegative and undergo electrostatic interactions. In proteins, such ligands are glutamate, aspartate, terminal carboxylates, hydroxyls of serines and threonines and in the buffer acetate, citrate and phosphate. Soft ligands are polarizable and form covalent bonds such as cysteine, cystine, methionine and histidine in proteins and Cl^-, Br^-, I^-, S-ligands, CN^- and imidazole in the buffer solution.

Metals are classified according to their preference for hard or soft ligands. Class (a) metals bind preferentially to hard ligands. They comprise the cations of A-metals such as alkali and alkaline earth metals, the lanthanides, some actinides and groups IIIA, IVA and VA of the transition metals. Class (b) metals are rather soft and polarizable and can form covalent bonds to soft ligands. They include heavy metals at the end of the transition metal groups such as Hg, Pt and Au. Thus, in the protein, the class (b) metals Hg, Pt and Au and complex compounds of them bind to soft ligands such as cysteine, histidine or methionine and the class (a) metals U and Pb to hard ligands such as the carboxylate groups of glutamate or aspartate.

4.1.2 Single Isomorphous Replacement

The structure factor $\mathbf{F}_{\text{PH}}$ for the heavy atom derivative structure (Figure 27) becomes (Equation 62):

$$\mathbf{F}_{\text{PH}} = \mathbf{F}_{\text{P}} + \mathbf{F}_{\text{H}} \tag{62}$$

where $\mathbf{F}_{\text{P}}$ = structure factor of the native protein and $\mathbf{F}_{\text{H}}$ = contribution of the heavy atoms to the structure factor of the derivative. The isomorphous differences, $F_{\text{PH}} - F_{\text{P}}$, which can be calculated from experimental

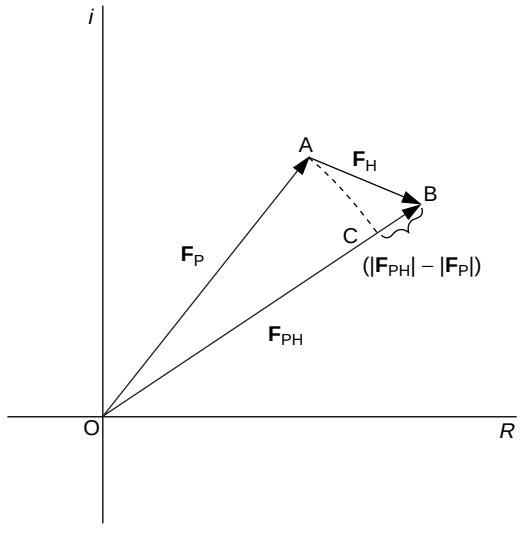

Figure 27 Vector diagram for the vector addition of the structure factor of the native protein $\mathbf{F}_{\text{P}}$ and the heavy atom contribution $\mathbf{F}_{\text{H}}$ to the heavy atom derivative structure factor $\mathbf{F}_{\text{PH}}$.

intensity data sets of the native and derivative protein, correspond to the distance CB in Figure 27 and are given by Equation (63):

$$F_{PH} - F_P = F_H \cos(\alpha_{PH} - \alpha_H) - 2F_P \sin^2\left(\frac{\alpha_P - \alpha_{PH}}{2}\right. \tag{63}$$

If F_H is small compared with F_P and F_{PH}, the sine term will be very small and we have (Equation 64)

$$F_{PH} - F_P \approx F_H \cos(\alpha_{PH} - \alpha_H) \tag{64}$$

When vectors $\mathbf{F}_P$ and $\mathbf{F}_H$ are collinear, then (Equation 65)

$$|\mathbf{F}_{PH} - \mathbf{F}_P| = F_H \tag{65}$$

The square of the isomorphous differences, $F_{PH} - F_P$, can be used as coefficients in a Patterson synthesis. We get

$$\begin{aligned}(F_{PH} - F_P)^2 = 4F_P^2 \sin^4\left(\frac{\alpha_P - \alpha_{PH}}{2}\right. & \quad \text{(i)} \\ + F_H^2 \cos^2(\alpha_{PH} - \alpha_H) & \quad \text{(ii)} \\ - 4F_P F_H \sin^2\left(\frac{\alpha_P - \alpha_{PH}}{2}\right. & \\ \times \cos(\alpha_{PH} - \alpha_H) & \quad \text{(iii)}\end{aligned} \tag{66}$$

It is a theorem of Fourier theory that the Fourier transform of the sum of Fourier coefficients is equal to the sum of the Fourier transforms of the individual Fourier coefficients. Here there are three different terms. $(\alpha_P - \alpha_{PH})$ is small if F_H is small and term (i) which gives the protein–protein interaction will be of low weight. The transform of term (iii) is zero if sufficient terms are included. However, if $F_H \ll F_P$, $(\alpha_{PH} - \alpha_H)$ is effectively random and term (ii) will give heavy atom vectors with half the expected peak heights (Equation 67):

$$F_H^2 \cos^2(\alpha_{PH} - \alpha_H) = \tfrac{1}{2}F_H^2 + \tfrac{1}{2}F_H^2 \cos 2(\alpha_{PH} - \alpha_H) \tag{67}$$

with the second term on the right contributing only noise to the Patterson map because the angles α_{PH} and α_H are not correlated. Such an isomorphous heavy atom difference Patterson map allows the determination of the positions of the heavy metals on the condition of isomorphism and a not too large heavy atom partial structure. The interpretation of these difference Patterson maps is undertaken by vector verification routines which are part of the CCP4 program suite.[36] In these routines the asymmetric unit of the unit cell is systematically scanned by calculating on each scan point the corresponding heavy atom–heavy atom vectors, determining their peak height in the Patterson map and evaluating a meaningful correlation value (e.g. the sum of the correlated maxima). Prominent heavy atom sites should show up with high correlation values.

It is important to know what intensity changes are generated by the attachment of heavy atoms to the macromolecule. According to Crick and Magdoff,[39] the relative root mean square intensity change is given by Equation (68) for centric reflections:

$$\frac{\sqrt{\overline{(\Delta I)^2}}}{\overline{I_P}} = 2 \times \frac{\overline{I_H}}{\overline{I_P}} \tag{68}$$

and by Equation (69) for acentric reflections:

$$\frac{\sqrt{\overline{(\Delta I)^2}}}{\overline{I_P}} = \sqrt{2} \times \frac{\overline{I_H}}{\overline{I_P}} \tag{69}$$

where $\overline{I_H}$ is the average intensity of the reflections if the unit cell were to contain the heavy atoms only and $\overline{I_P}$ is the average intensity of the reflections of the native protein. Attaching one mercury atom ($Z = 80$) to a macromolecule with varying molecular mass and assuming 100% occupancy gives the following average relative changes in intensity: 0.51 for 14 000 Da, 0.25 for 56 000 Da, 0.18 for 112 000 Da, 0.13 for 224 000 Da and 0.09 for 448 000 Da. From this estimation it is evident that with increasing molecular mass more heavy atoms or for large molecular masses heavy metal clusters such as $Ta_6Br_{12}^{2+}$[40] must be introduced to generate intensity changes which can be statistically measured (precision for intensity measurements between 5 and 10%) and which are sufficient for the phasing.

The phase calculation for single isomorphous replacement can be seen from the so-called Harker construction for this case (Figure 28). $\mathbf{F}_H$, which can be calculated from the known heavy atom positions, is drawn in its negative direction from the origin O ending at point A. Circles are drawn with radii F_P and F_{PH} from points O and A, respectively. The connections of the intersection points of both circles B and C with origin O determine two possible phases for $\mathbf{F}_P$. This means that the single isomorphous replacement leaves an ambiguity in the phase determination for the acentric reflections.

4.1.3 Multiple Isomorphous Replacement

The phase ambiguity can be overcome if two or more isomorphous heavy atom derivatives are used which exhibit different heavy atom partial structures. In Figure 29 the Harker construction for two different heavy atom derivatives is shown. In addition to Figure 28, $-\mathbf{F}_{H_2}$ is drawn from the origin O and a third circle with radius F_{PH_2} is inserted around its end-point B. The intersection point, H, of all three circles determines the protein phase, α_P. In the case of n isomorphous derivatives there are $n + 1$ circles which have one common intersection point

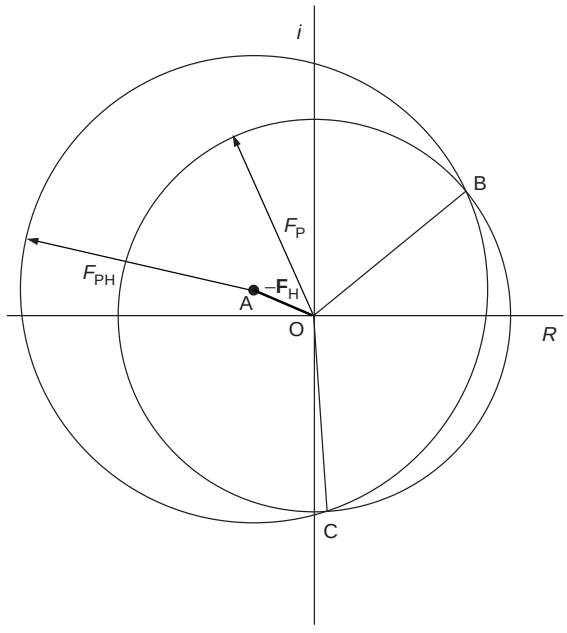

Figure 28 Harker construction for the phase calculation by the method of single isomorphous replacement.

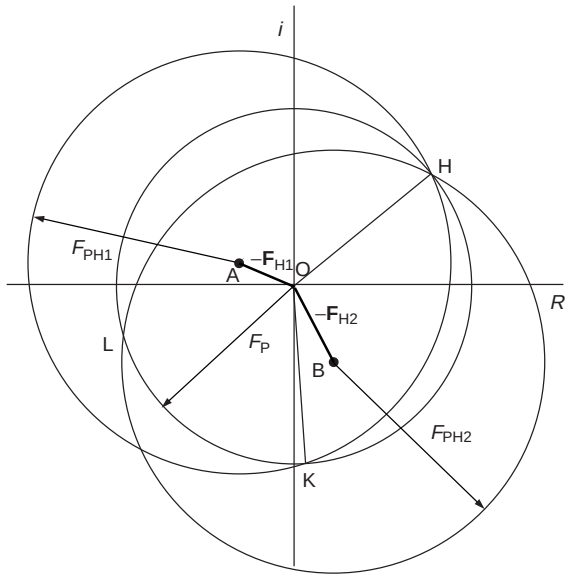

Figure 29 Harker construction for the phase calculation by the method of MIR for two different heavy atom derivatives PH_1 and PH_2.

whose connection to origin O determines the protein phase, α_P.

4.2 Anomalous Scattering

4.2.1 Theoretical Background

So far, in the normal Thomson scattering of X-rays the electrons in the atom have been treated as free electrons that vibrate as a dipole-oscillator in response to the incident electromagnetic radiation and generate elastic scattering of the X-rays. However, the electrons are bound to atomic orbitals in atoms and this treatment is valid only if the frequency w of the incident radiation is large compared with any natural absorption frequency w_{kn} of the scattering atom. For the light atoms in biological macromolecules (H, C, N, O, S, P) with frequency w of the used radiation (in the λ range of 0.4 to 3.5 Å), this condition is fulfilled and these atoms really scatter normally. For heavier elements the assumption $w \gg w_{kn}$ is no longer valid and the frequency w may be higher for some and lower for other absorption frequencies. If w is equal to an absorption frequency w_{kn}, absorption of radiation will occur which is manifested by the ejection of a photoelectron with an energy corresponding to the ionization energy for this electron. This transition goes to a state in the continuous region because the discrete energy states are all occupied in the atom. The absorption frequencies for the K, L, or M shells are connected with the corresponding absorption edges which are characterized by a sharp drop in the absorption curve (absorption vs λ) at the edge position. It is evident that the scattering from the electrons with their resonance frequencies close or equal to the frequency of the incident radiation will deliver a special contribution which is called anomalous scattering.

The classical treatment [see James[41]] is briefly outlined. It is assumed that the atoms scatter as if they contain electric dipole-oscillators having certain definite natural frequencies. The classical differential equation of the motion of a particle of mass m and charge e in an alternating electric field $\mathbf{E} = \mathbf{E}_0 e^{iwt}$ is Equation (70):

$$\ddot{\mathbf{x}} + k\dot{\mathbf{x}} + w_s^2\mathbf{x} = \frac{e\mathbf{E}_0}{m}\mathrm{e}^{iwt} \tag{70}$$

where the damping factor, k, is proportional to the velocity of the displayed charge and w_s is the natural circular frequency of the dipole if the charge is displaced. The steady-state solution for this equation for the moment of the dipole which executes forced oscillations of frequency w under the action of the incident wave is Equation (71):

$$\mathbf{M} = e\mathbf{x} = \frac{\mathrm{e}^2}{m}\frac{\mathbf{E}_0\mathrm{e}^{iwt}}{w_s^2 - w^2 + ikw} \tag{71}$$

The amplitude A of the scattered wave at unit distance in the equatorial plane is given by Equation (72):

$$A = \frac{\mathrm{e}^2}{mc^2}\frac{w^2E_0}{w_s^2 - w^2 + ikw} \tag{72}$$

The scattering factor of the dipole, f, is now defined as the ratio of the amplitude scattered by the oscillator to

that scattered by a free classical electron under the same conditions. This amplitude at unit distance and in the equatorial plane is given by Equation (73):

$$A' = -\frac{e^2}{mc^2}E_0 \tag{73}$$

Hence we obtain Equation (74) for f:

$$f = \frac{w^2}{w^2 - w_s^2 - ikw} \tag{74}$$

If f is positive the scattered wave has a phase difference of π with respect to the primary beam (introduced by the negative sign in the equation for A'). If $w \gg w_s$, f is unity. In the case of $w \ll w_s$, f is negative and the dipole then scatters a wave in phase with the primary beam. Equation (74) can be split into real and imaginary parts so that we obtain Equation (75):

$$f = f' + if'' \tag{75}$$

with (Equations 76 and 77)

$$f' = \frac{w^2(w^2 - w_s^2)}{(w^2 - w_s^2)^2 + k^2w^2} \tag{76}$$

$$f'' = \frac{kw^3}{(w^2 - w_s^2)^2 + k^2w^2} \tag{77}$$

We now extend this for an atom consisting of s electrons each acting as a dipole-oscillator with oscillator strength $g(s)$ and resonance frequency w_s. We have to multiply the contribution for each electron by $g(s)$ and form the sum over all electrons. For the total real part of the atomic scattering factor we obtain Equation (78):

$$f' = \sum_s \frac{g(s)w^2}{w^2 - w_s^2} \tag{78}$$

which assumes w is not very nearly equal to w_s, and a small damping. f' can be written as Equation (79):

$$f' = f_0 + \Delta f' = \sum_s g(s) + \sum_s \frac{g(s)w_s^2}{w^2 - w_s^2} \tag{79}$$

For free electrons we have $w_s = 0$ and $f' = f_0 = {}_s g(s)$. The real part of the increment of the scattering factor is due to the binding of electrons. $\Delta f'$ is the dispersion component of the anomalous scattering.

If w is comparable to w_s but slightly greater, ikw must not be neglected. f becomes complex (Equation 80):

$$f = f' + if'' = f_0 + \Delta f' + i\Delta f'' \tag{80}$$

The imaginary part lags $\pi/2$ behind the primary wave, i.e. is always $\pi/2$ in front of the scattered wave. $\Delta f''$ is known as the absorption component of the anomalous scattering.

In the quantum mechanical treatment of the problem the oscillator strengths are calculated from the atomic wave functions. Hönl,[42] in theoretical work, used hydrogen-like atomic wave functions. In the frame of this approach, to each natural dipole frequency w_s in the classical expression there corresponds in the quantum expression a frequency w_{kn}, which is the Bohr frequency associated with the transition of the atom from the energy state k to the state n in which it is supposed to remain during the scattering. Modern quantum mechanical calculations of anomalous scattering factors on isolated atoms, based on relativistic Dirac–Slater wave functions, have been carried out by Cromer and Liberman.[43] It follows from the theory of the anomalous scattering of X-rays that f_0 is real, independent of the wavelength of the incident X-rays but dependent on the scattering angle. $\Delta f'$ and $\Delta f''$ depend on the wavelength, λ, of the incident radiation but are virtually independent of the scattering angle.

4.2.2 Experimental Determination

$\Delta f''$ is related to the atomic absorption coefficient μ_0 by Equation (81):

$$\Delta f''(w) = \frac{mcw}{4\pi}\mu(w_0) \tag{81}$$

$\Delta f'$ can now be calculated by the Kramers–Kronig transformation (Equation 82):

$$\Delta f'(w) = \frac{2}{\pi} \int_0^\infty \frac{w'\Delta f''(w')}{w^2 - w'^2}\, dw' \tag{82}$$

As fluorescence is closely related to absorption, fluorescence measurements varying the X-ray radiation frequency are used to determine the frequency dependence of the dispersive components of the different chemical elements. Instead of the radiation frequency w, the radiation is often characterized by its wavelength, λ, or photon energy, E. The dispersion correction terms $\Delta f'$ and $\Delta f''$ are often simply denoted f' and f''. Figure 30 shows the anomalous scattering factors near the absorption K edge of selenium from a crystal of *E. coli* selenomethionyl thioredoxin. The spectrum was measured with tunable synchrotron radiation. Apart from the "white line" feature at the absorption edge, f'' drops by about 4 electrons approaching the edge from the short wavelength side; f' exhibits a symmetrical drop of -8 electrons around the edge. Similar values can be observed at the K edges for Fe, Cu, Zn and Br, whose wavelengths all lie in the range 0.9–1.8 Å, which is well suited for biological macromolecular X-ray diffraction experiments. For other interesting heavy atoms such as Sm, Ho, Yb, W, Os, Pt and Hg the L_{II} (Sm) or L_{III} edges are in this range.

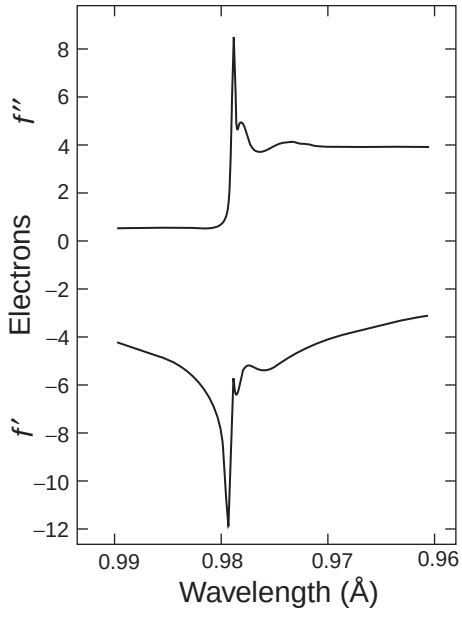

Figure 30 Anomalous scattering factors near the absorption K-edge of selenium from a crystal by *E. coli* selenomethionyl thioredoxin. (Reproduced by permission of Academic Press, Inc., from Hendrickson and Ogata.[44])

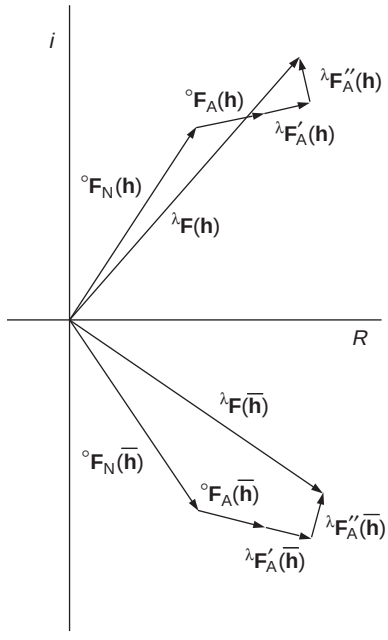

Figure 31 Vector diagram explaining the breakdown of Friedel's law.

Here, the effects are even greater. Considerably larger changes are found for several lanthanides, such as Yb, where the minimum f' is -33 electrons and the maximum f'' is 35 electrons.

4.2.3 Breakdown of Friedel's Law

Under the assumption that the crystal contains a group of anomalous scatterers, one can separate the contributions from the distinctive components of the scattering factor according to Hendrickson and Ogata[44] to obtain Equation (83):

$$^{\lambda}\mathbf{F}(\mathbf{h}) = {}^{\circ}\mathbf{F}_{\mathrm{N}}(\mathbf{h}) + {}^{\circ}\mathbf{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}\mathbf{F}'_{\mathrm{A}}(\mathbf{h}) + i\,{}^{\lambda}\mathbf{F}''_{\mathrm{A}}(\mathbf{h}) \quad (83)$$

where ${}^{\circ}\mathbf{F}_{\mathrm{N}}$ is the contribution of the normal scatterers and ${}^{\circ}\mathbf{F}_{\mathrm{A}}$, ${}^{\circ}\mathbf{F}'_{\mathrm{A}}$ and ${}^{\circ}\mathbf{F}''_{\mathrm{A}}$ are the contributions for the corresponding components of the complex atomic form factor. For the centrosymmetric reflection we obtain Equation (84):

$$^{\lambda}\mathbf{F}(-\mathbf{h}) = {}^{\circ}\mathbf{F}_{\mathrm{N}}(-\mathbf{h}) + {}^{\circ}\mathbf{F}_{\mathrm{A}}(-\mathbf{h}) + {}^{\lambda}\mathbf{F}'_{\mathrm{A}}(-\mathbf{h}) + i\,{}^{\lambda}\mathbf{F}''_{\mathrm{A}}(-\mathbf{h}) \quad (84)$$

The geometric presentation for both structure factors is given in Figure 31. The inversion of the sign of **h** causes a negative phase angle for all contributions where the components of the scattering factor are real. For the f''-dependent part this is also valid but owing to the imaginary factor i this vector has to be constructed with a phase angle $+\pi/2$ with respect to ${}^{\circ}\mathbf{F}_{\mathrm{A}}(-\mathbf{h})$ and ${}^{\lambda}\mathbf{F}'_{\mathrm{A}}(-\mathbf{h})$.

The resultant absolute values for ${}^{\lambda}\mathbf{F}(\mathbf{h})$ and ${}^{\lambda}\mathbf{F}(-\mathbf{h})$ are no longer equal, which means that their intensities (square of the amplitude) are different (breakdown of Friedel's law).

4.2.4 Anomalous Difference Patterson Map

One can show that (Equation 85)

$$^{\lambda}F(\mathbf{h}) - {}^{\lambda}F(-\mathbf{h}) \approx \frac{2}{k}[{}^{\circ}F_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})]\sin(\alpha_{\mathbf{h}} - \alpha_{\mathrm{A}}) \quad (85)$$

where $\alpha_{\mathbf{h}}$ is the phase angle of ${}^{\lambda}F(\mathbf{h})$, α_{A} the phase angle of the anomalous scatterers and (Equation 86)

$$k = \frac{{}^{\circ}F_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})}{{}^{\lambda}F''_{\mathrm{A}}(\mathbf{h})} \quad (86)$$

As coefficients for an anomalous difference Patterson we obtain Equation (87):

$$\Delta F^2_{\mathrm{ano}} = [{}^{\lambda}F(\mathbf{h}) - {}^{\lambda}F(-\mathbf{h})]^2 \sim \frac{4}{k^2}[{}^{\circ}\mathbf{F}_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}\mathbf{F}'_{\mathrm{A}}(\mathbf{h})]^2 \sin^2(\alpha_{\mathbf{h}} - \alpha_{\mathrm{A}}) \quad (87)$$

The ΔF_{ano}s will be maxima if the phase angle α_{A} is perpendicular to the phase angle $\alpha_{\mathbf{h}}$ and zero if both vectors are collinear, which is opposite to the MIR case. The anomalous Patterson map contains peaks of the anomalous scatterers with heights proportional to half of $(4/k^2)$ $[{}^{\circ}F_{\mathrm{A}}(\mathbf{h}) + {}^{\lambda}F'_{\mathrm{A}}(\mathbf{h})]^2$ owing to the $\sin^2$ term and is therefore suited to determine the structure of the anomalous scatterers.

4.2.5 Phasing Including Anomalous Scattering Information

The combination of anomalous scattering information with isomorphous replacement permits the unequivocal determination of the protein phases as shown in Figure 32. Using the anomalous scattering information alone gives two possible solutions for the protein phase characterized by the intersection points H and L in Figure 32. Combining it with the corresponding intensities from the native protein without the anomalous scatterers leaves only one solution for the protein phase (vector O–H in Figure 32). The case in Figure 32 is called single isomorphous replacement anomalous scattering (SIRAS). Having n isomorphous heavy atom derivatives, each with anomalous scattering contributions, the Harker construction can be extended for this situation and the phasing method is then designated multiple isomorphous replacement anomalous scattering (MIRAS).

4.2.6 Multiwavelength Anomalous Diffraction Technique

During the last few years, the MAD technique has matured to be a routine method and led to a revolution in biological macromolecular crystallography. If there are one or a few anomalous scatterers in the biological macromolecule it is possible to determine the whole spatial structure from one crystal (exact isomorphism) by the MAD technique. The anomalous scatterers may be intrinsic as in metalloproteins (e.g. Fe, Zn, Cu, Mo, Mn) or exogenous (e.g. Hg in a heavy atom derivative or Se in selenomethionyl proteins[45,46]). A prerequisite for the MAD technique is well diffracting crystals (resolution better than 2.8 Å) because the anomalous components of the atomic form factor are virtually independent of the diffraction angle and acquire increasing weight with increasing scattering angle. This advantageous property together with exact isomorphism serves for the determination of good phases down to the full resolution and leads to the production of excellent experimental MAD-phased electron density maps. A typical MAD experiment is carried out at three different wavelengths (tunable synchrotron radiation), at minimum f' and maximum f'' at the absorption edge of the anomalous scatterer(s) and at a remote wavelength where anomalous scattering effects are small.

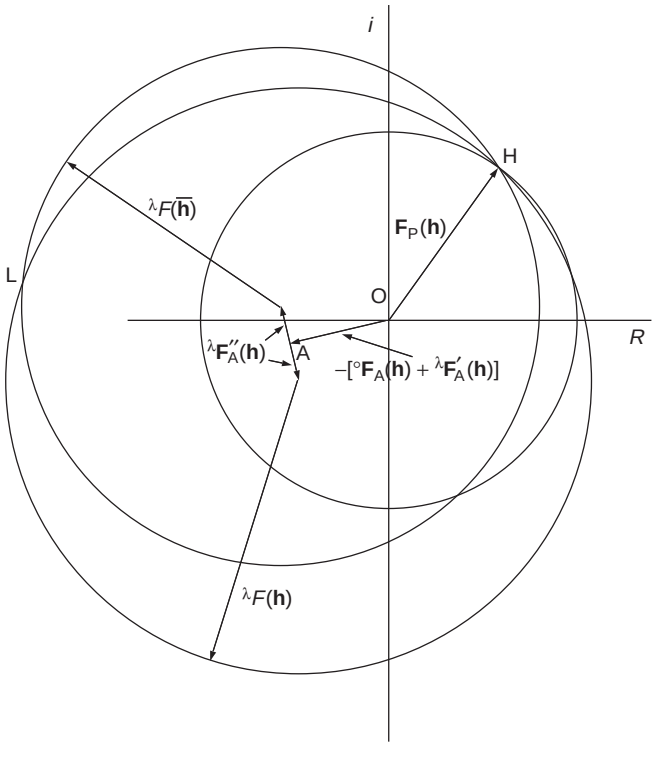

Figure 32 Harker construction illustrating the phase determination combining information from anomalous scattering and isomorphous replacement.

The basic equations for the MAD technique as formulated by Hendrickson and Ogata[44] are as follows. Equation (83) can be written as Equation (88):

$$^{\lambda}\mathbf{F}(\mathbf{h}) = {}^{\circ}\mathbf{F}_T(\mathbf{h}) + {}^{\lambda}\mathbf{F}'_A(\mathbf{h}) + i^{\lambda}\mathbf{F}''_A(\mathbf{h}) \tag{88}$$

where (Equation 89):

$$^{\circ}\mathbf{F}_T = {}^{\circ}\mathbf{F}_N + {}^{\circ}\mathbf{F}_A \tag{89}$$

with subscript T for the totality of atoms in the structure. Furthermore, we have Equations (90–93):

$$^{\circ}\mathbf{F}_T(f^{\circ}) = {}^{\circ}F_T \exp(i^{\circ}\phi_T) \tag{90}$$

$$^{\circ}\mathbf{F}_A(f^{\circ}) = {}^{\circ}F_A \exp(i^{\circ}\phi_A) \tag{91}$$

$$^{\lambda}\mathbf{F}'_A = f(f') \tag{92}$$

$$^{\lambda}\mathbf{F}''_A = f(f'') \tag{93}$$

In the common case of a single kind of anomalous scatterer Equations (94) and (95):

$$^{\lambda}\mathbf{F}'_A = \frac{f'(\lambda)}{f^{\circ}}{}^{\circ}\mathbf{F}_A \tag{94}$$

$$^{\lambda}\mathbf{F}''_A = \frac{f''(\lambda)}{f^{\circ}}{}^{\circ}\mathbf{F}_A \tag{95}$$

Separating the experimentally observable squared amplitude into wavelength-dependent and wavelength-independent terms gives Equation (96):

$$\begin{aligned} {}^{\lambda}F(\pm\mathbf{h})^2 = {}& {}^{\circ}F_T^2 + a(\lambda){}^{\circ}F_A^2 \\ & + b(\lambda){}^{\circ}F_T{}^{\circ}F_A \cos({}^{\circ}\phi_T - {}^{\circ}\phi_A) \\ & \pm c(\lambda){}^{\circ}F_T{}^{\circ}F_A \sin({}^{\circ}\phi_T - {}^{\circ}\phi_A) \end{aligned} \tag{96}$$

with (Equations 97–99)

$$a(\lambda) = \frac{f'^2 + f''^2}{f^{\circ 2}} \tag{97}$$

$$b(\lambda) = 2\frac{f'}{f^{\circ}} \tag{98}$$

$$c(\lambda) = 2\frac{f''}{f^{\circ}} \tag{99}$$

Maximum anomalous scattering effects can be expected in intensity differences of reflections that would be equal for exclusively normal scattering. This is the case for Friedel pairs, **h** and −**h**, or their rotational symmetry partners and the relation for such differences is given in Equation (85). Of further interest are dispersive differences between structure amplitudes at different wavelengths Equation (100):

$$\Delta F_{\Delta\lambda} \equiv {}^{\lambda i}F(\mathbf{h}) - {}^{\lambda j}F(\mathbf{h}) \tag{100}$$

The anomalous or dispersive intensity differences can be used to determine the structure of the anomalous scatterers. The methods are the same as for isomorphous replacement. They include vector verification procedures of difference Patterson maps or direct methods.

4.2.7 Determination of the Absolute Configuration

As anomalous scattering destroys the centrosymmetry of the diffraction data, this effect can be used to determine the absolute configuration of chiral biological macromolecules. The most common method is to calculate protein phases based on both hands of the heavy atom or anomalous scatterer structures and check the quality of the relevant electron density map which should be better for the correct hand. Furthermore, secondary structural elements in proteins (consisting of L-amino acids) such as α-helices should be right-handed.

4.3 Patterson Search Methods (Molecular Replacement)

If the structures of molecules are similar (virtually identical) or contain a major similar part, this can be used to determine the crystal structure of the related molecule if the structure of the other molecule is known. This is done by systematically exploring the Patterson function of the crystal structure to be determined with the Patterson function of the search model. Let us first consider some important features of the Patterson function. The relation between two identical molecules in the search crystal structure (Figure 33a) can generally be formulated as Equation (101):

$$\mathbf{X}_2 = [C]\mathbf{X}_1 + \mathbf{d} \tag{101}$$

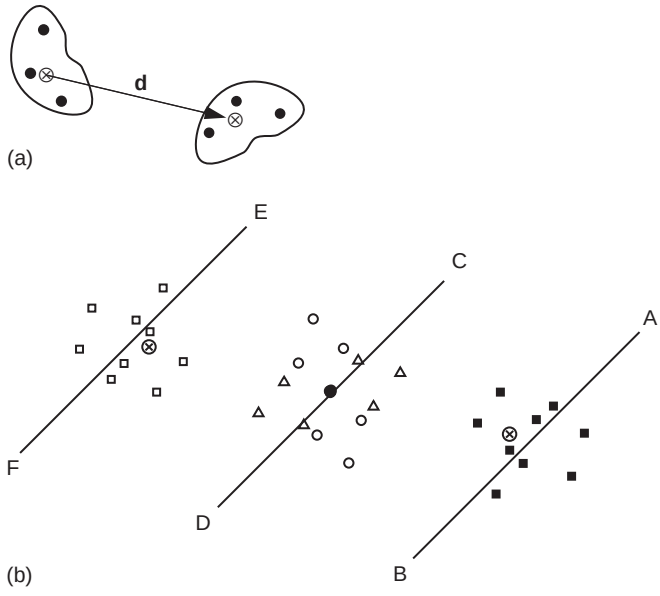

Figure 33 Patterson function of two identical molecules separated by the spatial movement given in Equation (101). (a) Positions of the two molecules; (b) interatomic vectors of structure in (a). △, ○, Intramolecular vectors of the left and right molecule; , ■, intermolecular vectors.

Equivalent positions $\mathbf{X}_1$ in molecule 1 are at positions $\mathbf{X}_2$ of molecule 2. $[C]$ is the rotation matrix and **d** the translation vector of the movement of the molecule. Figure 33(b) shows the Patterson function belonging to the molecular arrangement in Figure 33(a). It is evident that around the origin vectors are assembled that are intramolecular whereas the vectors around lines AB and EF are intermolecular. The intramolecular vectors depend on the molecule orientation only and therefore can be used for its determination. Once the orientation of the molecule(s) has been elucidated this can be used to reveal the translation of the molecule(s) by analyzing the intermolecular vector part of the Patterson function. The distinction between intra- and intermolecular vector sets and exploiting them for orientation and translation determination was given by Hoppe.[47] The extension to protein crystallography and the first mathematical formulation of the rotation and translation functions were given by Rossman and Blow.[48]

4.3.1 Rotation Function

The intramolecular vectors are arranged in a volume around the origin of the Patterson function with a radius equal to the dimension of the molecule. The rotational search is then carried out in this volume u. The search Patterson (deduced either from the search model or from the crystal Patterson itself) is rotated to any possible rotational orientation $\mathbf{X}_2$ characterized by three

rotational angles α, β, γ which may be defined in different ways (polar angles, Eulerian angles, etc.). At each angular position the actual functional values are correlated with those of the crystal Patterson all through the volume u and integrated over this volume. The correlation function may be the sum or the product of each corresponding pair of values. Rossman and Blow proposed a product function and the rotation function for this case is given by Equation (102):

$$R(\alpha, \beta, \gamma) = \int_u P_2(\mathbf{X}_2) P_1(\mathbf{X}_1)\, \mathrm{d}\mathbf{X}_1 \qquad (102)$$

The function has maxima if the intramolecular vector sets are coincident. The calculation can be carried out in both direct and reciprocal space.

The self-rotation function is a special form of the rotation function. If an asymmetric unit contains more than one copy of a molecule, the rotation matrix between the molecules can be determined by a self-rotation function. Here, the crystal Patterson is rotated against itself and the integration is taken over the volume u around the origin in the same manner. The identical molecules may have an arbitrary orientation to each other or they may be related by local or so-called noncrystallographic symmetries. Searching for local rotation axes is done best in a polar angle system. The search Patterson is brought into each polar orientation and then rotated around the angle value for the local axis being sought, e.g. 120° for a threefold local axis.

4.3.2 Translation Function

Once the orientation of the molecule(s) has been determined, the translation of the molecule(s) can be obtained from a translation function. The model Patterson $P_2(\mathbf{u})$ revealed from the model in the correct orientation is calculated for different translations $\mathbf{t}$ and correlated with the crystal Patterson $P_1(\mathbf{u})$. The translation function proposed by Crowther and Blow[49] has the form shown in Equation (103):

$$T(\mathbf{t}) = \int P_1(\mathbf{u}) P_2(\mathbf{u}, \mathbf{t})\, \mathrm{d}\mathbf{u} \qquad (103)$$

$T(\mathbf{t})$ reveals a maximum peak at the correct translation $\mathbf{t}$ if the center of gravity of the search model was at the origin for $\mathbf{t} = 0$.

4.3.3 Computer Programs for Molecular Replacement

An early program for molecular replacement, working in direct space, was written by Huber.[50] Nowadays, several program packages are available, either being exclusively dedicated to the molecular replacement technique or having integrated relevant modules. Pure molecular replacement programs are e.g. AMORE[51] and GLRF.[52] The rotational and translational search starting from the search model is fully automated in AMORE and includes a final rigid body refinement of each proposed solution. GLRF offers different types of rotation and translation functions, all operating in reciprocal space, and a Patterson correlation refinement.[53] A peculiarity of the GLRF program is the locked rotation function. This function takes into account possible noncrystallographic symmetries and is an average of n independent rotation functions with an improved peak-to-noise ratio. Frequently used program packages including molecular-replacement modules are the CCP4 program suite,[36] X-PLOR[54] and PROTEIN.[55]

4.4 Phase Calculation

4.4.1 Refinement of Heavy Atom Parameters

Before the protein phases can be calculated, it is necessary to refine the heavy atom parameters. These are the coordinates x, y, z, the temperature factor (either isotropic or anisotropic) and the occupancy. The refinement modifies the parameters in such a way that $|\mathbf{F}_{\mathrm{PH}}(\mathrm{obs})|$ becomes as close as possible to $|\mathbf{F}_{\mathrm{PH}}(\mathrm{calc})|$. Using the method of least squares, the refinement according to Rossmann[56] minimizes Equation(104):

$$\varepsilon = \sum_{\mathbf{h}} w(\mathbf{h}) [(F_{\mathrm{PH}} - F_{\mathrm{P}})^2 - k F^2_{\mathrm{H\,calc}}]^2 \qquad (104)$$

where k is a scaling factor to correct $\mathbf{F}^2_{\mathrm{H\,calc}}$ to a theoretically more acceptable value because according to Equation (64) $\mathbf{F}_{\mathrm{PH}} - \mathbf{F}_{\mathrm{P}}$ and $\mathbf{F}_{\mathrm{H}}$ have approximately the same length when $\mathbf{F}_{\mathrm{PH}}$, $\mathbf{F}_{\mathrm{P}}$ and $\mathbf{F}_{\mathrm{H}}$ point in the same direction. The probability for this case will be high if the difference between F_{PH} and F_{P} is large. An improvement can be obtained if the contribution from the anomalous scattering is included.[57]

For the parameter refinement of anomalous scattering sites, the differences between the observed and calculated structure factor amplitudes for $^{\circ}F_{\mathrm{A}}$ are subjected to minimization. Another approach treats the anomalous or dispersive contributions as in MIR phasing.

From the refined heavy atom parameters, preliminary protein phase angles α_{P} can be obtained as shown in the corresponding Harker construction. A further refinement of the heavy atom parameters can be achieved by the "lack of closure" method[58] incorporating this knowledge. The definition of this "lack of closure" ε is illustrated in Figure 34(a) and (b). In the case of perfect isomorphism, the vector triangle $\mathbf{F}_{\mathrm{P}} + \mathbf{F}_{\mathrm{H}} = \mathbf{F}_{\mathrm{PH}}$ closes exactly (Figure 34a). In practice, this condition will not be fulfilled and a difference ε between the observed F_{PH}

and the calculated F_{PH} will remain (Figure 34b). F_{PH}(calc) can be obtained from the triangle OAB (Figure 34b) with the cosine rule (Equation 105):

$$F_{PH} = [F_P^2 + F_H^2 + 2F_P \times F_H \cos(\alpha_H - \alpha_P)]^{1/2} \quad (105)$$

The function which is minimized by the least-squares method is Equation (106):

$$E_j = \sum_{\mathbf{h}} m_{\mathbf{h}} \varepsilon_j(\mathbf{h})^2 \quad (106)$$

where (Equation 107)

$$\varepsilon_j = k_j F_{PHj}(\text{obs}) - F_{PHj}(\text{calc}) \quad (107)$$

is the ''lack of closure'' for the heavy atom derivative j, k_j is a scaling factor and $m_{\mathbf{h}}$ is a weighting factor.

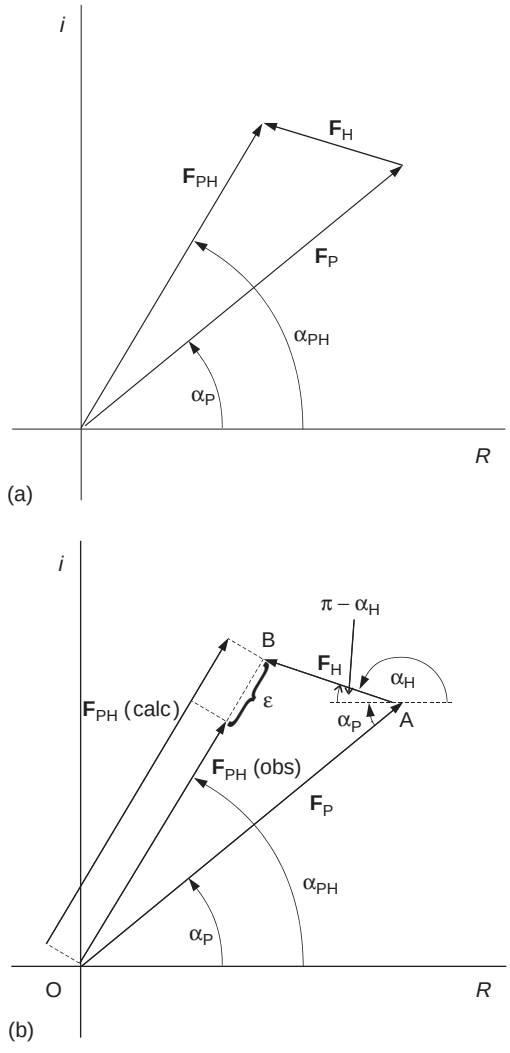

Figure 34 Definition of ''lack of closure''. (a) Perfect isomorphism; (b) usually, the observed and calculated values for $\mathbf{F}_{PH}$ differ by the ''lack of closure'' ε.

4.4.2 Protein Phases

As the structure factor amplitudes F_P, F_{PH}, F_H and α_H are known, the protein phase angle α_P can be calculated. For the single isomorphous replacement situation (Figure 27) ε is zero only for the two protein phase angles α_P where the two circles for F_P and F_{PH} intersect. In practice, all these observed quantities exhibit errors. For the treatment of these errors it is assumed that all errors are in $\mathbf{F}_{PH}$ and that both $\mathbf{F}_H$ and $\mathbf{F}_P$ are error-free. For each protein phase angle α, $\varepsilon(\alpha)$ is calculated. The smaller $\varepsilon(\alpha)$ is, the higher is the probability of a correct phase angle α. For each reflection of a derivative j a Gaussian probability distribution is assumed for ε according to Equation (108):

$$P(\alpha) = P(\varepsilon) = N \exp\ -\frac{\varepsilon^2(\alpha)}{2E^2} \quad (108)$$

where N is a normalization factor and E^2 the mean square value of ε. Small values of E are related to probability curves with sharp peaks and well-determined phase angles and the opposite is true for large E values.

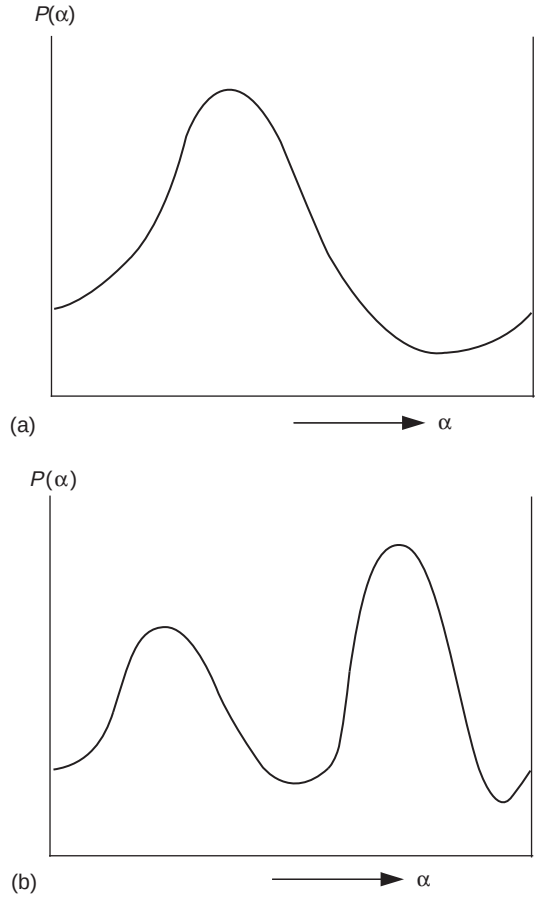

Figure 35 Total probability curves $P(\alpha)$ for two different reflections: (a) for one derivative; (b) for more than one derivative.

Such phase-angle probability curves can be calculated for each individual reflection and derivative. For single isomorphous replacement this curve is symmetric with two high peaks corresponding to the two possible solutions for α_P. We obtain the total probability for each reflection with contributions from n heavy atom derivatives by multiplying the individual probabilities (Equation 109):

$$P(\alpha) = \prod_{j=1}^{n} P_j(\alpha) = N' \exp\left(-\sum_j \frac{\varepsilon_j^2(\alpha)}{2E_j^2}\right) \quad (109)$$

These curves will be nonsymmetric with one or several maxima (see Figure 35a and b).

The question arises of which phases should be taken in the electron density equation to calculate the best electron density function. An immediate guess would be to use the phases where $P(\alpha)$ has the highest value. This approach would be appropriate for unimodal distributions but not for bimodal distributions. Blow and Crick[59] derived the phase value that must be applied under the assumption that the mean square error in electron density over the unit cell is minimal. For one reflection this is given by Equation (110):

$$\langle \Delta\rho^2 \rangle = \frac{1}{V^2}(\mathbf{F}_s - \mathbf{F}_t)^2 \quad (110)$$

where $\mathbf{F}_t$ is the true and $\mathbf{F}_s$ the structure factor applied in the Fourier synthesis. The mean square error is then obtained as Equation (111):

$$\langle \Delta\rho^2 \rangle = \frac{1}{V^2} \frac{\int_{\alpha=0}^{2\pi} (\mathbf{F}_s - F \exp i\alpha)^2 P(\alpha)\, d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha)\, d\alpha} \quad (111)$$

$\mathbf{F}_t$ has a phase probability of $P(\alpha)$ and has been given as $\mathbf{F}_t = F \exp i\alpha$. It can be shown that the numerator integral in Equation (111) is minimal if (Equation 112):

$$\mathbf{F}_{s(best)} = F \frac{\int_{\alpha=0}^{2\pi} \exp(i\alpha) P(\alpha)\, d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha)\, d\alpha} = mF \exp(i\alpha_{best}) \quad (112)$$

Equation (112) corresponds to the center of gravity of the probability distribution with polar coordinates (mF, α_{best}), where m is defined as magnitude of $\mathbf{m}$ given by Equation (113):

$$\mathbf{m} = \frac{\int_{\alpha=0}^{2\pi} P(\alpha) \exp(i\alpha)\, d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha)\, d\alpha} \quad (113)$$

This magnitude of $\mathbf{m}$ is equivalent to a weighting function and is designated the ''figure of merit''. The electron density map calculated with mF and α_{best} is known as ''best Fourier'' and should represent a Fourier map with minimum least-squares error from the true Fourier map.

For the total error of the ''best Fourier'', Equation (114) has been derived:

$$\langle \Delta\rho^2 \rangle = \frac{1}{V^2} \sum_{\mathbf{h}} F^2(\mathbf{h})(1 - m^2) \quad (114)$$

The order of magnitude of this error may be illustrated by the example of the structure determination of lysozyme. The root mean square error in the Fourier synthesis was 0.35 e Å^{-3} with values of 2.0 Å resolution for the diffraction data and a mean ''figure of merit'' of 0.6.

The program systems CCP4[36] and PROTEIN[55] contain all routines necessary to calculate protein phases according to the MIRAS technique and a number of different kinds of Fourier maps. Alternative probabilistic approaches for the phase calculation are used in programs MLPHARE[60] and SHARP.[61] Both programs can also carry out MAD phasing. The MADSYS program[62] is based on the algebraic approach outlined in the MAD section of this contribution and executes all tasks of a MAD analysis from scaling to phase-angle calculation.

4.5 Phase Improvement

With the methods so far described an experimental electron density map can be calculated and if its quality is high enough the atomic model can be constructed. However, there are methods for further phase improvement available which may be applied in general or depending on given prerequisites. Such phase improvement routines have been used routinely over the last 10 years and have had a large impact on the advancement of biological macromolecular crystallography.

4.5.1 Solvent Flattening

Protein crystals have a solvent content of 75–40%. In a highly refined protein crystal structure the solvent space between the molecules is rather flat owing to the dynamic nature of this region. Usually the initial experimental starting phases are of lower quality than the final ones and as a result the solvent region (if the molecular boundaries can be identified) contains noise peaks. It is now obvious to set the noisy solvent space to a low constant value and calculate new improved phases by Fourier back-transforming this corrected electron density map. However, it is evident that the definition of the molecular boundaries will be tedious and depend on the quality of the electron density map. Wang[63]

has proposed an automatic procedure which smooths the electron density to define the protein region. This smoothed electron density map is traced against a threshold value which separates this map into molecule and solvent space according to their ratio of volumes in the unit cell. The space inside the molecular envelope is polished to avoid internal voids. Now, a new electron density map is calculated using the observed structure factor amplitudes and the phases revealed from the solvent corrected map. The solvent corrected map is obtained by setting all electron density values inside the molecular envelope to those of the initial map and all values outside the envelope to a low constant value. These phases from the solvent flattening procedure can be combined with the MIR or MAD phases. This procedure can be repeated in several iterative cycles because after each cycle of solvent flattening the quality of the electron density map is improved. There are no prerequisites for the application of the method of solvent flattening. It is evident that solvent flattening is most effective for crystals with a high solvent content.

4.5.2 Histogram Matching

Histogram matching is a technique emanating from image processing. In the application to electron density maps it is assumed that a high-quality protein crystal structure has a characteristic frequency distribution of electron density which serves as a standard reference distribution for other electron density maps. Such maps of lower quality exhibit a frequency distribution of electron density which deviates from the standard distribution. The electron density map of low quality is then scaled in such a way that its frequency distribution of electron density now corresponds to the standard distribution. Histogram matching is normally used together with solvent flattening and is incorporated into the density modification programs SQUASH[64] and DM from the CCP4 program package.[36]

4.5.3 Molecular Averaging

If there is more than one identical subunit in the asymmetric unit of the crystal, molecular averaging can be used to improve the protein phases. The spatial relations between the single identical subunits in the asymmetric unit may be determined by Patterson search methods or from the arrangement of the heavy atoms or anomalous scatterers. The spatial relation between the identical subunits can be improper (the relevant spatial movement consists of a rotation about an unsymmetrical angle value and a translation component) or proper (the spatial movements form a symmetry group which is composed of rotational symmetry elements only). Such additional symmetries are called noncrystallographic or local and there are no limitations concerning the Zähligkeit of the symmetry axes (e.g. five-, seven- and higher-fold axes are allowed). It is evident that averaging about the different related subunits, whose electron density should be equal in each subunit, must result in an improved electron density map and therefore in improved protein phases. Molecular averaging is best done in direct space and several programs (e.g. RAVE[65] or MAIN[66]) are available.

The procedure of molecular averaging is composed of several steps. First, the molecular envelope must be determined from the initial electron density map or from a molecular model which, for example, has been obtained from molecular replacement. Next, the particular electron density averaging between the related subunits is performed. This is followed by the reconstitution of the complete crystal unit cell with the averaged electron density. The space outside the molecular envelope is flattened. This map is then Fourier back-transformed. The obtained phase angles can either be taken directly or combined with known phase information to calculate a new and improved electron density map. This cycle can be repeated several times until convergence of the electron density map improvement has been reached. It is very useful to refine the local symmetry operations after every macrocycle of molecular averaging. Furthermore, molecular averaging can be applied if proteins crystallize in more than one crystal form.

Molecular averaging is especially efficient if a high noncrystallographic symmetry is present as in virus structures, but the averaging over two related subunits alone (the lowest case of local symmetry) can give a considerable improvement. In special cases where high noncrystallographic symmetry exists and the phase information extends to low resolution only, cyclic molecular averaging can be used to extend the phase angles to the full resolution of the native protein. This was first shown in the structure analysis of hemocyanin from *Panulirus interruptus*.[67,68] It is extensively used in the analysis of icosahedral structures (see, e.g. Rossmann et al.[69] and Ladenstein et al.[70]) and for large molecular assemblies (e.g. Löwe et al.).[71]

4.5.4 Phase Combination

In the course of a crystal structure analysis of a biological macromolecule, phase information from different sources may be available, such as information from isomorphous replacement, anomalous scattering, partial structures, solvent flattening and molecular averaging. An overall phase improvement can be expected when these factors are combined and a useful method to do this was proposed by Hendrickson and Lattman.[72] The probability curve

for each reflection is written in an exponential form as Equation (115):

$$P_s(\alpha) = N_s \exp(K_s + A_s \cos\alpha + B_s \sin\alpha + C_s \cos 2\alpha + D_s \sin 2\alpha) \quad (115)$$

Subscript s stands for the source from which the phase information has been derived. K_s and the coefficients A_s, B_s, C_s and D_s depend on the structure factor amplitudes and other magnitudes, e.g. the estimated standard deviation of the errors in the derivative intensity, but are independent of the protein phase angles α. The overall probability function $P(\alpha)$ is obtained by a multiplication of the individual phase probabilities and this turns out to be a simple addition of all K_s and of the related coefficients in the exponential term. We obtain Equation (116):

$$\begin{aligned} P(\alpha) &= \prod_s P_s(\alpha) \\ &= N' \exp\left(\sum_s K_s + \left(\sum_s A_s\right)\cos\alpha + \left(\sum_s B_s\right)\sin\alpha \right. \\ &\quad \left. + \left(\sum_s C_s\right)\cos 2\alpha + \left(\sum_s D_s\right)\sin 2\alpha\right) \end{aligned} \quad (116)$$

K_s and the coefficients $A_s - D_s$ have special expressions for each source of phase information and they are explicitly given e.g. in Drenth.[21]

4.6 Difference Fourier Technique

Supposing that one has solved the crystal structure of a biological macromolecule and has isomorphous crystals of this macromolecule which contain small structural changes caused by a substrate-analog or inhibitor binding, a metal removal or replacement or a local mutation of one or several amino acids. Then these structural changes can be determined by the difference Fourier technique. The difference Fourier map is calculated with the differences between the observed structure factor amplitudes of the slightly altered molecule $F_{\mathrm{DERI}}(\mathrm{obs})$ and the native molecule $F_{\mathrm{NATI}}(\mathrm{obs})$ as Fourier coefficients and the phase angles of the native molecule α_{NATI} as phases according to Equation (117):

$$\rho_{\mathrm{DERI}} - \rho_{\mathrm{NATI}} \simeq \frac{1}{V}\sum_{\mathbf{h}} m[F_{\mathrm{DERI}}(\mathrm{obs}) - F_{\mathrm{NATI}}(\mathrm{obs})] \times \exp(i\alpha_{\mathrm{NATI}}) \exp(-2\pi i \mathbf{hx}) \quad (117)$$

where m may be the figure of merit or another weighting scheme. The difference Fourier map can alternatively be calculated with coefficients $F_{\mathrm{DERI}}(\mathrm{obs}) - F_{\mathrm{DERI}}(\mathrm{calc})$ and phases $\alpha_{\mathrm{DERI}}(\mathrm{calc})$. $F_{\mathrm{DERI}}(\mathrm{calc})$ and $\alpha_{\mathrm{DERI}}(\mathrm{calc})$ do not include the unknown contribution of the structural change.

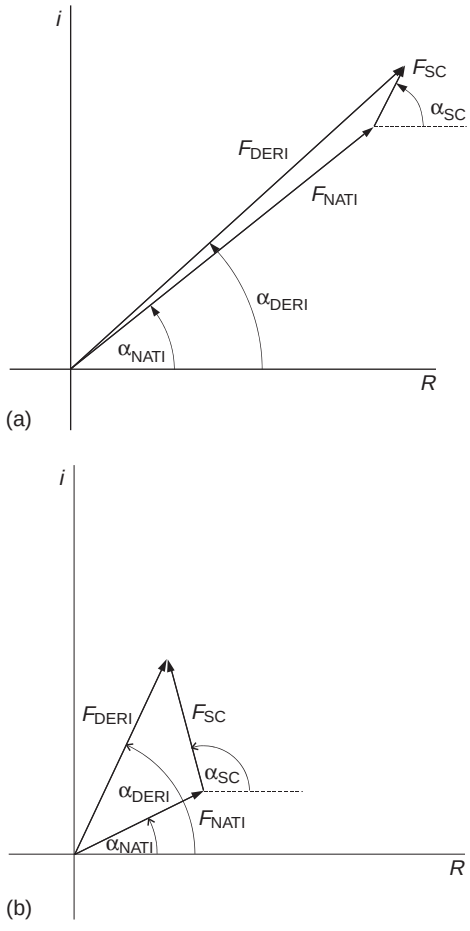

Figure 36 Vector diagrams illustrating different situations (a) and (b) in the difference Fourier technique for the involved structure factors.

Figure 36(a) and (b) illustrate the relation for the structure factors involved in the difference Fourier technique. We assume that the structural change is small. If F_{NATI} is large, the structure factor amplitude of the structural change F_{SC} will be small compared with F_{NATI}, and α_{DERI} will be close to α_{NATI}. This is no longer valid if F_{NATI} is small. Now F_{SC} is comparable to F_{NATI}, and α_{DERI} may deviate considerably from α_{NATI}. This implies the necessity to introduce a weighting scheme that scales down the contributions where the probability is high that α_{NATI} differs appreciable from the correct phase angle. Various weighting schemes have been elaborated such as those of Sim[73] and Read.[74] The weighting scheme of Sim has the following form (Equation 118):

$$w = \frac{I_1(X)}{I_0(X)} \quad (118)$$

for acentric reflections and (Equation 119):

$$w = \tanh \frac{X}{2} \tag{119}$$

for centric reflection with (Equation 120)

$$X = \frac{2F_{\mathrm{DERI}} \times F_{\mathrm{NATI}}}{\sum_{1}^{n} f_j^2} \tag{120}$$

$I_0(X)$ and $I_1(X)$ are modified Bessel functions of zeroth and first order, respectively. These equations and weighting schemes can also be used for the calculation of OMIT maps (where parts of the model have been omitted from the structure factor evaluation) or when a complete structure must be developed from a known partial model. F_{DERI} must be replaced by the observed structure factor F, F_{NATI} by the structure factor of the known or included part of the model F_{K} and α_{NATI} by the phase angle α_{K} of the known or included part of the model.

5 MODEL BUILDING AND REFINEMENT

5.1 Model Building

Once the quality of the MIRAS or MAD maps is good enough, model building can be started. This is done on a computer graphics system and the main modeling programs are O[75] and TURBO-FRODO.[76,77] An interesting alternative is the program MAIN,[66] which additionally contains routines for molecular averaging, molecular docking and other features. The visualization of the relevant electron density map on the computer graphics system is done as cage-like structures. For this purpose, the standard deviation from the mean value of the map is calculated and the cage-like structure is built up for a given contour level (normally 1.0 σ). The first task in a de novo protein crystal structure analysis is to localize the trace of the polypeptide chain. This can be assisted by routines for automatic chain tracing such as the *bones* routine, an auxiliary program of O. Such automatic chain-tracing programs generate a skeleton of the electron density map. This representation was introduced by Greer.[78] When the trace of the polypeptide chain has been identified, the atomic model can be built into the electron density. The atomic model is represented as sticks connecting the atomic centers of bonded atoms. The individual building blocks (amino acids) of the protein molecule can be generated, interactively manipulated (e.g. linked with each other, moved, rotated, etc.) and fitted into the corresponding part of the electron density map. The geometry of the atomic model is regularized according to protein standard geometries.

The success of model building depends on the quality and resolution of the experimental electron density map. Usually, the quality of the electron density map is not so good that the complete model can be constructed in one cycle. In this case, the partial atomic model is refined crystallographically against the observed structure factor amplitudes. This phase information can be used directly to calculate a new electron density map commonly with $2F_{\mathrm{obs}} - F_{\mathrm{calc}}$ Fourier coefficient amplitudes. This kind of map is the sum of a normal F_{obs} Fourier and a difference Fourier synthesis. It displays the atomic model with normal weight and indicates errors in the model by its contribution of the difference Fourier map. The parallel determination and inspection of a difference Fourier map is also very helpful. As already mentioned, the model phases can be combined with phases present from other sources or incorporated in procedures of phase improvement. A further model-building cycle can be started with such new and improved electron density maps. After several cycles of model building and crystallographic refinement the atomic model will be so well defined that the solvent structure of internally bound solvent molecules can be developed. The atomic model is complete now and the biochemical interpretation can be started.

5.2 Crystallographic Refinement

The structural model has to be subjected to a refinement procedure. In practical macromolecular crystallography, one does not always have atomic resolution. Therefore, the single atoms cannot be treated as moving independently. They must be refined using energy or stereochemistry restraints, taking care to maintain a reasonable stereochemistry of the macromolecule.

There exist different approaches to structure refinement of macromolecules. The minimization of a potential energy function E together with a diffraction term D according to Equation (121):

$$S = E + D \tag{121}$$

where (Equations 122 and 123):

$$E = \sum k_{\mathrm{b}}[b_{j(\mathrm{calc})} - b_j^0]^2 + \sum k_{\tau}[\tau_{j(\mathrm{calc})} - \tau_j^0]^2 + \sum k_{\Theta}[1 + \cos(m\Theta_k + \delta)] + \sum (Ar^{-12} + Br^{-6}) \tag{122}$$

$$D = \sum_i w_i[F_{i(\mathrm{obs})} - kF_{i(\mathrm{calc})}]^2 \tag{123}$$

is applied in the programs EREF[79] and XPLOR,[54] which is now used frequently. The four terms of the right-hand side of E describe bond, valence angle, dihedral torsion angle and non-bonded interactions, k_{b}

is the bond stretching constant, k_τ is the bond angle bending force constant, k_Θ is the torsional barrier, m and δ are the periodicity and the phase of the barrier, A and B are the repulsive and long-range nonbonded parameters, D is the crystallographic contribution with w_i a weighting factor, F_{obs} the observed structure factor, F_{calc} the calculated structure factor and k a scaling factor. Programs TNT[80] and PROLSQ[81] use stereochemically restrained least-squares refinement. For both refinement schemes parameters are employed which were derived from small molecule crystal structures of amino acids, small peptides, nucleic acids, sugars, fatty acids, cofactors etc.[82] If noncrystallographic symmetry is present a corresponding term may be introduced in the energy or stereochemistry part of the expression to be minimized. It is possible to divide the structural model into several individual parts and refine these parts as rigid bodies. This is especially useful with solutions from molecular replacement.

A measure of the quality of the crystallographic model is calculated from the crystallographic R-factor (Equation 124):

$$R = \frac{\sum_i |F_{obs}| - k|F_{calc}|}{\sum_i |F_{obs}|} \tag{124}$$

Typical R-factors are below 0.2 for a well-refined macromolecular structure.

Beside the atomic coordinates x, y, z, the atomic temperature factor B may be refined at a resolution better than 3.5 Å. This is done in most programs in a separate step where e.g. in program XPLOR[54] the target function (Equation 125):

$$T = E_{XRAY} + E_R \tag{125}$$

is minimized, where (Equation 126)

$$E_R = W_B \sum_{(i,j)-\text{bonds}} \frac{(B_i - B_j)^2}{\sigma^2_{\text{bonds}}} + W_B \sum_{(i,j,k)-\text{angles}} \frac{(B_i - B_k)^2}{\sigma^2_{\text{angles}}} + W_B \times \sum_{k\text{-group}} \sum_{j\text{-equivalences}} \sum_{i\text{-unique atoms}} \frac{(B_{ijk} - \overline{B}_{ik})^2}{\sigma^2_{\text{ncs}}} \tag{126}$$

The last term is used only if noncrystallographic symmetry restraints should be imposed on the molecules. Normally isotropic B-factors are applied and refined in macromolecular crystallography only. Even for a high-resolution structure (1.7 Å), the ratio of observations (observed structure factors) to parameters to be refined (x, y, z, B for each atom) is only about 3. Therefore, as many as possible additional "observations" (energy or stereochemistry restraints) are incorporated. In some cases it is useful to refine the individual occupancy of certain atoms such as bound metal ions or solvent atoms. This must be performed in a separate step.

All of the mentioned refinement procedures are based on the least-squares method. The radius of convergence for this method is not very high because it follows a downhill path to its minimum. If the model is too far away from the correct solution, the minimization may end in a local minimum corresponding to an incorrect structure. Brünger[83,84] introduced the method of molecular dynamics (MD) which is able to overcome barriers in the S-function and find the correct global minimum. MD calculations simulate the dynamic behavior of a system of particles. The basic idea of the MD refinement technique is to increase the temperature sufficiently high for the atoms to overcome energy barriers and then to cool slowly to approach the energy minimum. This MD protocol is designated as simulated annealing (SA). The crystallographic application of the MD or SA technique includes a crystallographic term D as given in Equation (121) treated as a pseudoenergy term. A crystallographic MD or SA refinement is capable e.g. of overcoming a high-energy barrier occurring in the flipping of a peptide plane. It can be useful in removing model bias from the system.

A new approach is the refinement of macromolecular structures by the maximum likelihood method. Programs working on the basis of this method are REFMAC[85] and CNS.[86] The results derived using the maximum likelihood residual are consistently better than those from least-squares refinement.

If the resolution of a biological macromolecular crystal structure is equal to or better than 1.2 Å it is in the range of real atomic resolution and the ratio of observations to parameters is high enough to carry out, in principle, an unrestrained crystallographic refinement. Advances in cryogenic techniques, area detectors and the use of synchrotron radiation enabled macromolecular data to be collected to atomic resolution for an increasing number of proteins.[87] SHELXL[88] is a program with all tools for the crystallographic refinement of biological macromolecules at real atomic resolution.

5.3 Accuracy and Verification of Structure Determination

A measure of the quality of a structure determination is the crystallographic R-factor given in Equation (124). For a high-resolution structure, e.g. 1.6 Å, it should not be much larger than 0.16. As this R-factor is an overall number, it does not indicate major local errors. This can be obtained by the evaluation of a real space R-factor[75] which is calculated on a grid for nonzero

elements according to Equation (127):

$$R_{\text{real space}} = \frac{\sum |\rho_{\text{obs}} - \rho_{\text{calc}}|}{\sum |\rho_{\text{obs}} + \rho_{\text{calc}}|} \qquad (127)$$

where ρ_{obs} is the observed and ρ_{calc} the calculated electron density.

It has been shown that the conventional *R*-factor may reach rather low values in a crystallographic refinement with structural models that turned out to be wrong later. To overcome this unsatisfactory situation, Brünger[89,90] proposed the additional calculation of a so-called free *R*-factor. For this purpose the reflections are divided into a working set (e.g. 90%) and a test set (e.g. 10%). The reflections in the working set are used in the crystallographic refinement. The free *R*-factor is calculated with reflections from the test set which were not used for the crystallographic refinement and is thus unbiased by the refinement process. There exists a high correlation between the free *R*-factor and the accuracy of the atomic model phases.

The accuracy of the final model expressed by the mean coordinate error can be determined alternatively by a Luzzati[91] or σ_A plot.[74,92] The mean coordinate error for a macromolecular structure determined with a resolution of 2.0 Å and a crystallographic *R*-factor of 0.2 is in the region of ±0.2 Å.

The stereochemistry of the final model must also be checked. The root-mean-square deviation of bond lengths and bond angles from ideal geometry should not be greater than 0.015 Å and 3.0°, respectively. The conformation of the main chain folding is verified by a Ramachandran plot.[93] The dihedral angles Φ and Ψ are plotted against each other for each residue. The data points should lie in the allowed regions of the plot which correspond to energetically favorable secondary structures such as α-helices, β-sheets and defined turn structures. Exceptions are glycine residues which may occur at any position in the Ramachandran plot. Further stereochemical parameters to be checked are bond lengths and angles, dihedral angles (e.g. determinating side-chain conformations), noncovalent interactions, geometry of H-bonds and interactions in the solvent structure. This can be done with the programs PROCHECK[94–96] or WHAT_CHECK.[97]

Nearly all spatial structures of biological macromolecules determined either by X-ray crystallography or nuclear magnetic resonance (NMR) techniques have been and will be deposited with the Protein Data Bank[98] at Brookhaven National Laboratory or its successor. The information of the structural model is in a file that contains for each individual atom of the model a record with atom number, atom name, residue type, residue name, coordinates *x*, *y*, *z*, *B*-value(s) and occupancy. The header records hold useful information such as crystal parameters, amino acid sequence, secondary structure assignments and references.

6 APPLICATIONS

6.1 Enzyme Structure and Enzyme–Inhibitor Complex

6.1.1 X-ray Structure of Cystathionine β-Lyase

Cystathionine β-lyase (CBL) is a member of the γ-family of pyridoxal-5′-phosphate (PLP)-dependent enzymes that cleaves Cβ–S bonds of a broad variety of substrates. The crystal structure of CBL from *E. coli* has been solved using MIR phases in combination with density modification.[99] The enzyme has been crystallized by the hanging drop vapor diffusion method using either ammonium sulfate or PEG 400 as precipitating agent. The crystals belong to the orthorhombic space group $C222_1$ with unit cell parameters $a = 60.9$ Å, $b = 154.7$ Å and $c = 152.7$ Å. There is one dimer per asymmetric unit. A native data set has been collected using synchrotron radiation (wavelength 1.1 Å) at the wiggler beamline BW6 at the storage ring DORIS at the Deutsches Elektronensynchrotron (DESY) in Hamburg, Germany. Data sets for three heavy atom derivatives [thiomersalate, 2-mercuri-4-diazobenzoic acid and platinum(II)-2,2′-6,6″-terpyridinium chloride] were registered on an imaging plate scanner (MAR Research, Hamburg, Germany) using graphite monochromatized Cu Kα radiation from an RU200 rotating anode generator (Rigaku, Tokyo, Japan) operating at 5.4 kW. The reflection data were processed with the MOSFLM package[30] and scaled with programs from the CCP4 program suite.[36] All data sets reveal satisfactory symmetry consistency factors ($R_{\text{merge}} \leq 0.08$) and completenesses (>90%). The heavy atom positions were determined from isomorphous difference Patterson maps using the vector verification routines of program PROTEIN.[55] All derivatives have one common heavy atom binding site at Cys72 and one of the mercury derivatives shows a second binding site at Cys229. The dimer in the asymmetric unit is related by a local twofold axis which lies parallel to the *x*-axis. The translation of this local axis was determined from the distribution of the heavy atom sites in the unit cell. An initial MIR map was calculated, followed by solvent flattening, twofold averaging about the local symmetry and density modification and phase extension to 2.5 Å resolution. Phase calculations, solvent flattening, density modification and phase extension were done with programs of the CCP4 package. Program AVE[65] was used for the averaging. The quality of the resulting electron density map was sufficiently high to build an

almost complete atomic model. Model building was performed on an ESV-30 Graphic system workstation (Evans and Sutherland, Salt Lake City, UT, USA) using program O.[75] The atomic model was refined by energy restrained crystallographic refinement with XPLOR.[54]

The final model of CBL is made up of two monomers with 391 amino acids each, one cofactor and one hydrogencarbonate molecule per monomer and 581 solvent water molecules. The final crystallographic *R*-factor is 0.152 for data from 8.0 to 1.83 Å resolution and the free *R*-factor is 0.221. The mean positional error of the atoms as estimated from a Luzzati plot[91] is ±0.19 Å. A homotetramer with 222 symmetry is built up by crystallographic and noncrystallograpic symmetry (Figure 37). Each monomer of CBL (Figure 38) can be described in terms of three spatially and functionally different domains. The N-terminal domain (residues 1–60) consists of three α-helices and one β-strand. It contributes to tetramer formation and is part of the active site of the adjacent subunit. The second domain (residues 61–256) harbors PLP and has an α/β structure with a seven-stranded β-sheet as the central part. The remaining C-terminal domain (residues 257–395), connected by a long α-helix to the PLP-binding domain, consists of four helices packed on the solvent-accessible side of an antiparallel four-stranded β-sheet. The fold of the C-terminal and the PLP-binding domain and the location of the active site are similar to aminotransferases. Most of the residues in the active site are strongly conserved among the enzymes of the transulfuration pathway. Figure 39 shows the final $2F_{obs} - F_{calc}$ map superimposed with the refined atomic model around the active site of CBL. The cage-like structures for the representation of the electron density correspond to a contour level of 1.2σ. The protein part (main- and side-chain atoms), the PLP cofactor and the hydrogencarbonate molecule are well defined in the electron density map.

The knowledge of the spatial structure of a given enzyme structure forms the basis for understanding its functional properties. It is now possible to design rational site-directed mutants or determine the

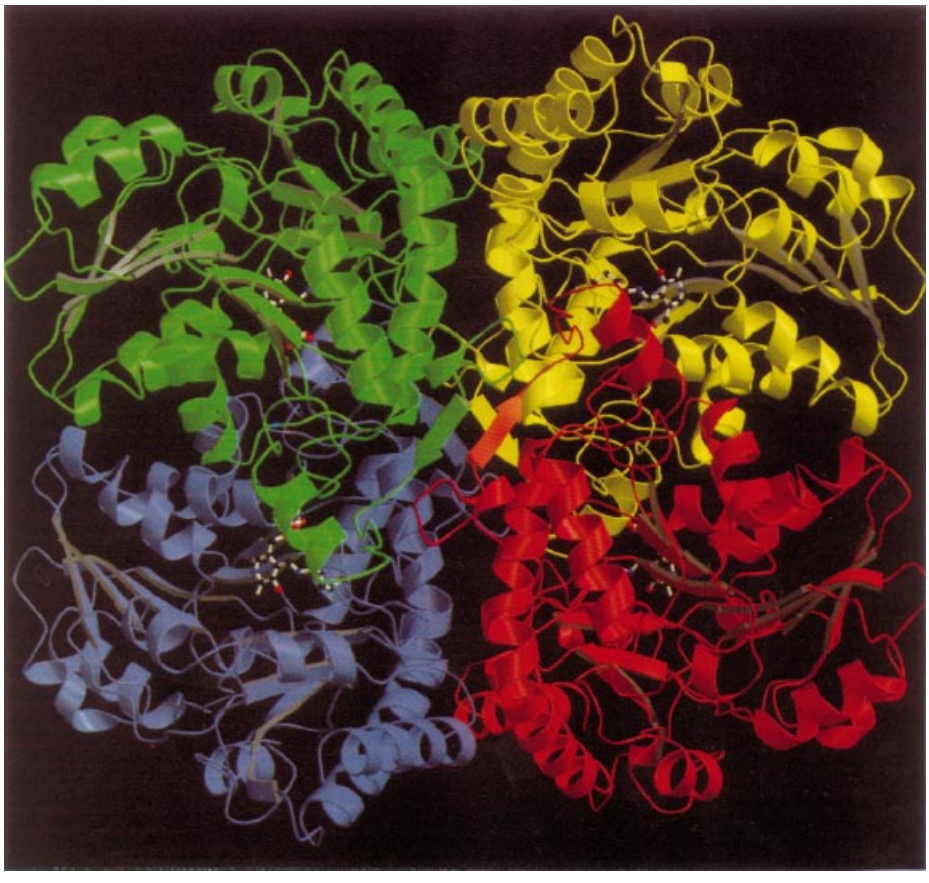

Figure 37 Ribbon plot of the CBL tetramer viewed along the *x*-axis. The monomers are colored differently. The blue- and green-colored monomers, which are related by a crystallographic axis (horizontal, in the plane of the paper), build up one catalytic active dimer, and the yellow and red ones the other. The location of the PLP-binding site is shown in a ball-and-stick presentation; MOLSCRIPT[100] and RASTER3D.[101] (Reproduced by permission of Academic Press, Ltd., from Clausen et al.[99])

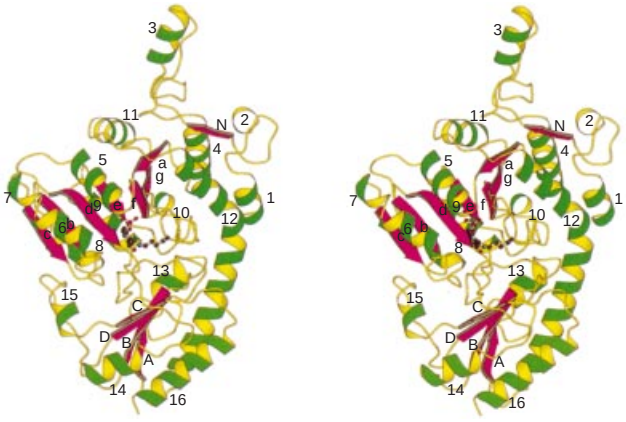

Figure 38 Stereo ribbon presentation of the CBL monomer, emphasizing secondary structure elements. α-Helices are drawn as green spirals, β-strands as magenta arrows. PLP and PLP-binding Lys210 are shown in a ball-and-stick representation; MOLSCRIPT[100] and RASTER3D.[101] (Reproduced by permission of Academic Press, Ltd., from Clausen et al.[99])

structures of enzyme–substrate, enzyme–substrate analog or enzyme–inhibitor complexes which will deliver invaluable information in understanding the enzyme's functional properties.

6.1.2 *Enzyme–Inhibitor Complex Structures of Cystathione β-Lyase*

The enzyme–inhibitor X-ray structures of β,β,β-trifluoroalanine (TFA) and L-aminoethoxyvinylglycine (AVG) with CBL could be determined.[99,102] In both cases, crystals of the complexes were obtained by incubating the enzyme solution with inhibitor in the millimolar range and subsequent cocrystallization. The resultant crystals were isomorphous with the native enzyme, making it possible to apply the difference Fourier technique in the structure solution. The technical details for the structure analyses are given in the relevant references [CBL/TFA[99] and CBL/AVG[102]]. The CBL/TFA complex structure was determined to substantiate that the ε-amino group of the active-site Lys210 can react with the nucleophile at the active site via Michael addition, which leads to covalent labeling and inactivation of the enzyme. The final $F_{obs} - F_{calc}$ and $2F_{obs} - F_{calc}$ electron density maps for the CBL/TFA complex around the active site are displayed in Figure 40. Clear, continuous electron density between the cofactor and Lys210 can be seen in this map, indicating a covalent lysine–inactivator–PLP product. In Figure 40, the blue $F_{obs} - F_{calc}$ map reveals well-defined electron density for the bound inhibitor. This binding mode of TFA to CBL corresponds to an intermediate in the reaction of TFA with CBL.[99] The structure of the inactivation product proves that Lys210 is the active-site nucleophile reacting via Michael addition with the inactivator. It must also be the residue which transfers a proton from Cα to Sγ in the reaction with the substrates.

The CBL/AVG structure[102] has been determined at 2.2 Å resolution and a crystallographic *R*-factor of 0.164.

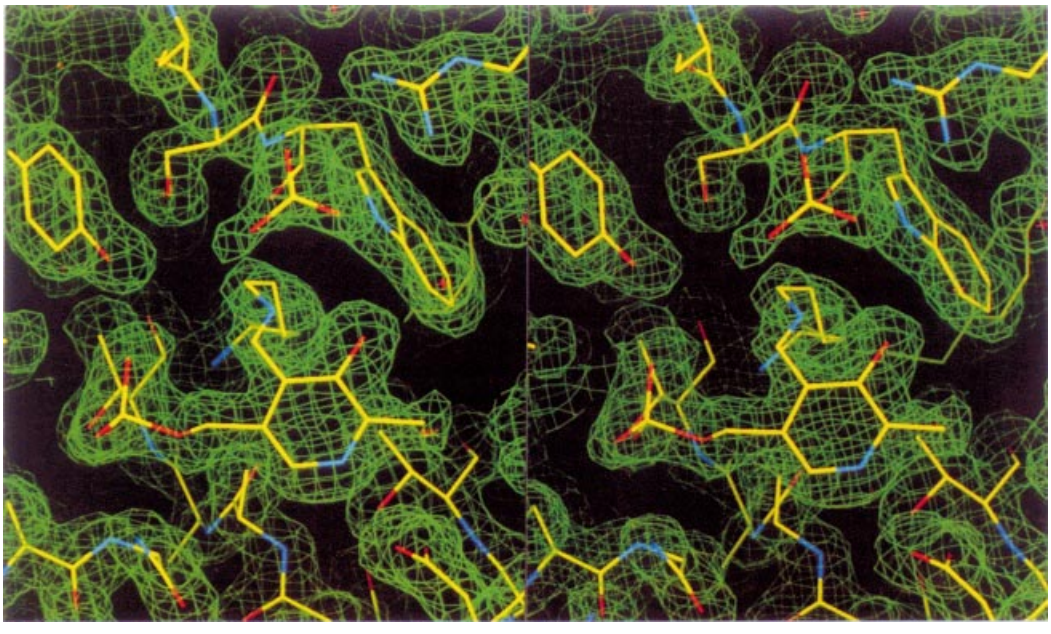

Figure 39 Stereo plot of the electron density in the active site of CBL, superimposed with the refined model of the region around the cofactor. The $2F_{obs} - F_{calc}$ map is contoured at 1.2σ and calculated at 1.83 Å resolution. (Reproduced by permission of Academic Press, Ltd., from Clausen et al.[99])

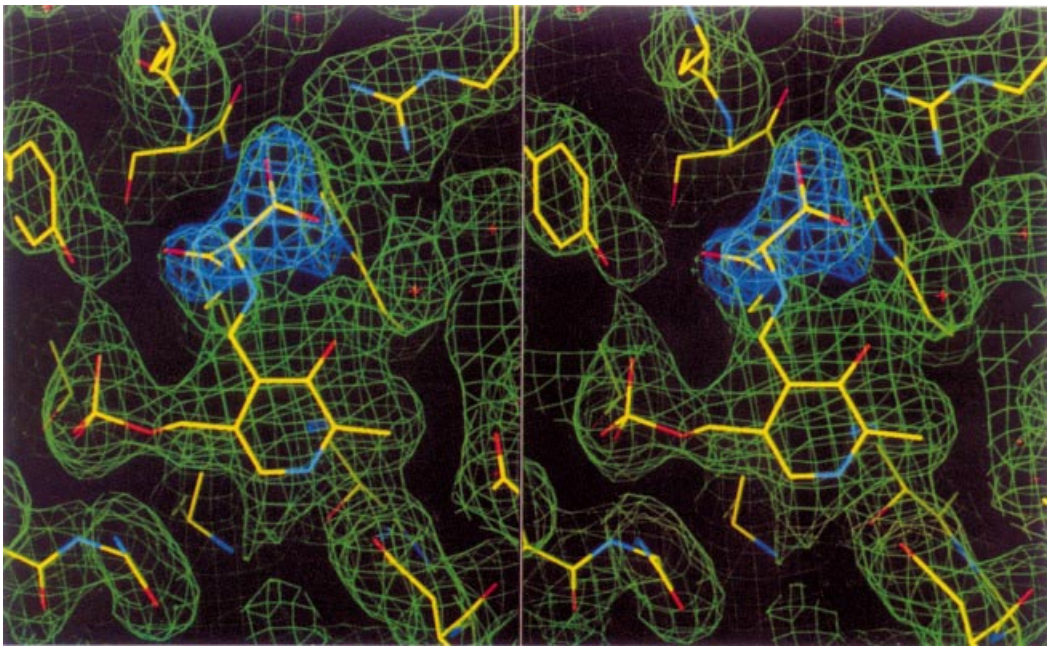

Figure 40 $F_{obs} - F_{calc}$ (blue) and $2F_{obs} - F_{calc}$ (green) electron density map of the CBL/TFA complex around the active site contoured at 3.5σ and 1.0σ, respectively, at 2.3 Å resolution. (Reproduced by permission of Academic Press, Ltd., from Clausen et al.[99])

The X-ray structure shows that AVG binds to the PLP cofactor forming the external aldimine. Lys210 is no longer bound to the PLP cofactor. Figure 41 is an overlay of the atomic models of native CBL (magenta), CBL/TFA (yellow) and CBL/AVG (green). The main difference in inhibitor binding is the location of Cβ and its substituents; in the TFA complex, the inactivator is directed toward the protein interior (the A face of the cofactor), whereas in CBL/AVG, Cβ is located at the B side of the cofactor. The α-carboxylate group in CBL/AVG is located in the same position as the hydrogencarbonate in the native and the α-carboxylate group of TFA in the CBL/TFA complex. The terminal amino group of AVG is held in place mainly by interactions with the hydroxyl group of Tyr111.

The experimental determination of the external aldimine structure in the CBL/AVG complex is of high relevance because it can serve as a rational basis for modeling of substrate and inhibitor binding, leading to more effective herbicides.

6.1.3 Crystal Structure of the Thrombin–Rhodniin Complex

This complex structure is an example related to pharmaceutical research. The goal of this special application is the development of more efficient blood anticoagulants. The target enzyme is α-thrombin. This enzyme is a serine proteinase of trypsin-like specificity. α-Thrombin, the key enzyme in hemostasis and thrombosis, exhibits both enzymatic and hormone-like properties and can be both pro- and anticoagulatory [see e.g. Stubbs and Bode[104]]. Rhodniin is a highly specific inhibitor of thrombin isolated from the assassin bug *Rhodnius prolixus*. Such blood-sucking animals have developed various anticlotting mechanisms to prevent local clotting of the victim's blood. These natural thrombin inhibitors are polypeptides of 60–120 amino acid residues.

The crystal structure of the noncovalent complex between recombinant rhodniin and bovine α-thrombin has been determined at 2.6 Å resolution.[105] Crystals were obtained by cocrystallization of thrombin with rhodniin in an approximately 1 : 1 molar ratio. The structure could be solved by molecular replacement because the spatial structure of the major constituent, bovine α-thrombin, was known. Only a diffraction data set of the complex crystal had to be collected. Rotational and translational searches for the orientation and position of the thrombin molecules in the unit cell were performed with the program AMORE.[51] The rotational search showed two solutions with correlation values of 0.22 and 0.20 over 0.09 for the next highest peak. Translational search and rigid body fitting for these two solutions resulted in a correlation value of 0.54 with the two independent complex molecules in the asymmetric unit. The quality of the electron density map calculated from the thrombin phases was good enough in principle to build the model of the rhodniin molecule (noncrystallographic averaging

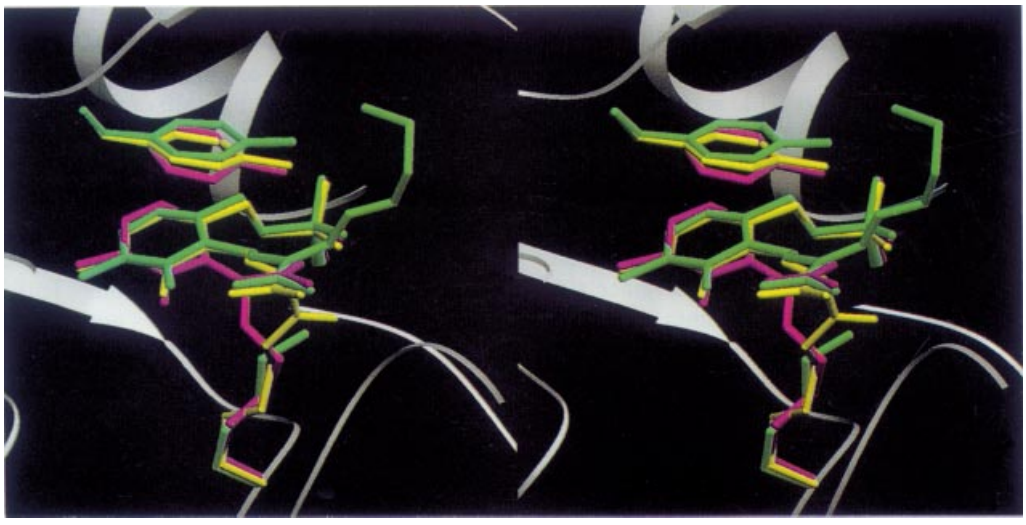

Figure 41 Stereo view of a superposition of Tyr111 and the PLP derivative of the unliganded enzyme (magenta), the CBL/AVG adduct (green) and the TFA-inactivated enzyme (yellow); SETOR.[103] (Reproduced by permission of the American Chemical Society, from Clausen et al.[102])

was also applied). The structure was refined with XPLOR to an *R*-factor of 0.189 and a free *R*-factor of 0.262. Further technical aspects of the crystal structure analysis are given in van de Locht et al.[105]

Figure 42 shows the structure of the complex between thrombin and rhodniin as a ribbon plot with α-helices represented as ribbon spirals and β-strands as arrows. The N-terminal domain binds in a substrate-like manner to the narrow active-site cleft of thrombin. The C-terminal domain, whose distorted reactive-site loop cannot adopt the canonical conformation, docks to the fibrinogen recognition exosite via extensive electrostatic interactions. The peculiarity of this complex structure is that the two KAZAL-type domains of rhodniin bind to two different sites of thrombin.

6.2 Metalloproteins

Metals bound as cofactors in proteins have a great variety of functions. They may be involved in the activation of small inorganic or organic molecules, in oxygen storage and transport, in electron transport or in

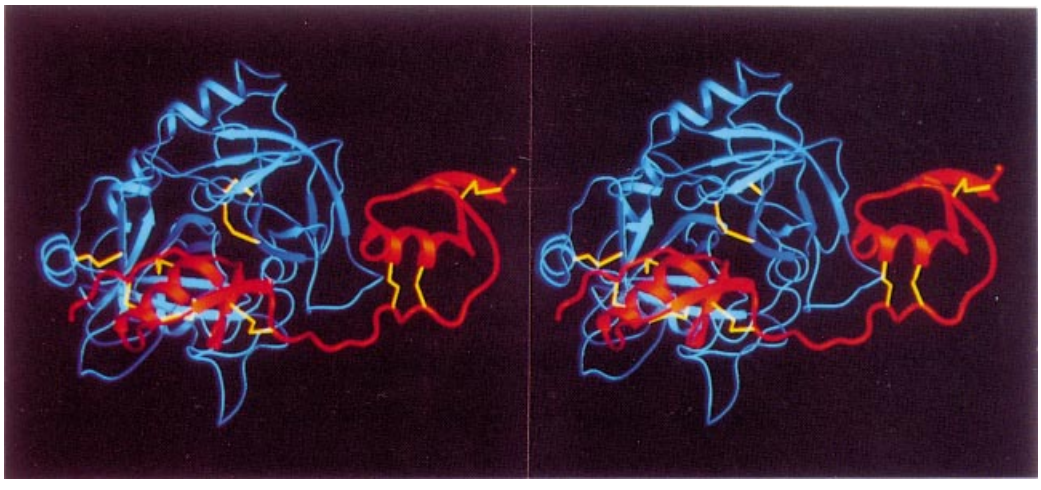

Figure 42 Stereo view of the complex formed between thrombin (blue) and rhodniin (red) in the thrombin standard orientation, i.e. with the active-site cleft facing the viewer and a bound inhibitor chain from left to right. Yellow connections indicate disulfide bridges. Rhodniin interacts through its N-terminal domain in a canonical manner with the active site and through its C-terminal domain with the fibrinogen recognition exosite of thrombin; SETOR.[103] (Reproduced by permission of Oxford University Press, from van de Locht et al.[105])

stabilizing a transition state during enzymatic catalysis. Their role may also be solely structural. The chemistry of metals in proteins has attracted the interest of inorganic chemists and has led to the formation of the field of biological inorganic chemistry. As an example of a complex metalloprotein, the multicopper oxidase ascorbate oxidase (AO) is presented.

6.2.1 Crystal Structure of the Multicopper Enzyme Ascorbate Oxidase

The blue protein AO belongs to the group of "blue" oxidases with laccase and ceruloplasmin. These are multicopper enzymes catalyzing the four-electron reduction of molecular oxygen with concomitant one-electron oxidation of the substrate. The crystal structure of AO has been solved by the MIR technique and refined to 1.9 Å resolution.[106,107] The peculiarity of this structure determination is briefly described. It consists in the utilization of the information of two different crystal forms. In both crystal forms the molecules arrange themselves as homotetramers with 222 symmetry, but in crystal form 1, one of these twofold axes is realized by a crystallographic twofold axis resulting in two subunits per asymmetric unit. In crystal form 2, one homotetramer is found per asymmetric unit. In crystal form 1, six isomorphous heavy atom derivatives could be found and interpreted. An initial MIR map was calculated, solvent flattened and averaged about the local twofold axis. For crystal form 2, no phase information was available. From the averaged uninterpreted MIR map a whole tetramer was selected and used for rotational and translational searches in crystal form 2. This was successful and provided the necessary phase information for crystal form 2 and additionally the local symmetry. Now averaging could be performed both separately in the two crystal forms and subsequently between both crystal forms. The averaged electron density was transported into both unit cells and a new macrocycle of averaging could be started. These macrocycles were used to extend the phases from 3.5 Å to the full attainable resolution. This structure analysis was the first example where a molecular replacement was carried out with an uninterpreted MIR electron density-based model.

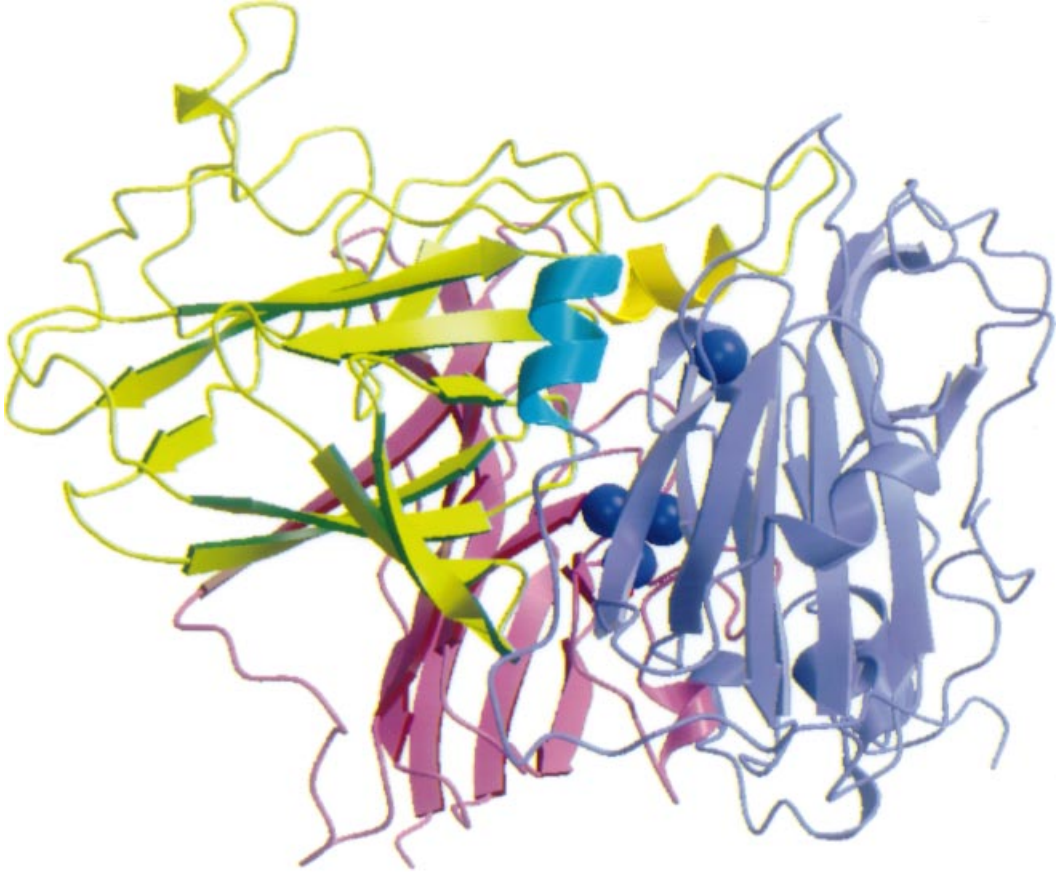

Figure 43 Schematic representation of the monomer structure of AO. (Reproduced by permission of World Scientific Publishing Co., from A. Messerschmidt, 'Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins', in *Multi Copper Oxidases*, ed. A. Messerschmidt, World Scientific Publishing, Singapore, 1997.)

AO is a homodimeric enzyme with a molecular mass of 70 kDa and 552 amino acid residues per subunit (zucchini). The three-domain structure and the location of the mononuclear and trinuclear copper centers in the AO monomer as derived from the crystal structure are shown in Figure 43. The folding of all three domains is of a similar β-barrel type. The mononuclear copper site is located in domain 3 and the trinuclear copper species is bound between domains 1 and 3. The coordination of the mononuclear copper site is depicted in Figure 44. It has the four canonical type-1 copper ligands (His, Cys, His, Met) also found in plastocyanin and azurin. The copper is coordinated to the ND1 atoms of His445 and His512, the SG atom of the Cys507 and the SD atom of Met517 in a distorted trigonal geometry. This unusual coordination geometry confers this copper site with its blue color. The trinuclear copper site (see Figure 45) has eight histidine ligands symmetrically supplied by domains 1 and 3 and two oxygen ligands. The trinuclear copper site may be divided into a pair of copper (CU2, CU3) with six histidine ligands in a trigonal prismatic arrangement. The pair is bridged by an OH^-, which leads to a strong antiferromagnetic coupling and makes this copper pair electron paramagnetic resonance silent. The remaining copper has two histidine ligands and an OH^- or H_2O ligand. A binding pocket for the reducing substrate which is complementary to an ascorbate molecule is located near the mononuclear copper site and accessible from solvent. A broad channel providing access from the solvent to the trinuclear copper species which is the binding and reaction site for the dioxygen is present in AO. During catalysis an intramolecular electron transfer between the mononuclear copper site and the trinuclear copper cluster occurs. The distances between the mononuclear copper and the three coppers of the trinuclear center are 12.20, 12.69 and 14.87 Å, respectively. Furthermore, the crystal structures of functional derivatives of AO such as the reduced, azide and peroxide form have been determined.[108] They show considerable changes at the trinuclear copper site in the reduced form and the peroxide or azide binding to the trinuclear copper species in the relevant structures. The X-ray studies on AO delivered essential information in understanding the catalytic mechanism.

6.2.2 *Crystal Structure of Cytochrome-c_{552} from* Thermus thermophilus *Determined by Multiwavelength Anomalous Diffraction Phasing*

The spatial structure of cytochrome-c_{552} from *Thermus thermophilus* has been solved by the MAD technique using synchrotron radiation and refined at a resolution of 1.28 Å to a *R*-factor of 0.19.[109] The data collection for the MAD phasing was performed at the wiggler beamline BW6 at the storage ring DORIS at DESY in Hamburg, Germany. At the beginning, an X-ray fluorescence spectrum around the Fe K-absorption edge

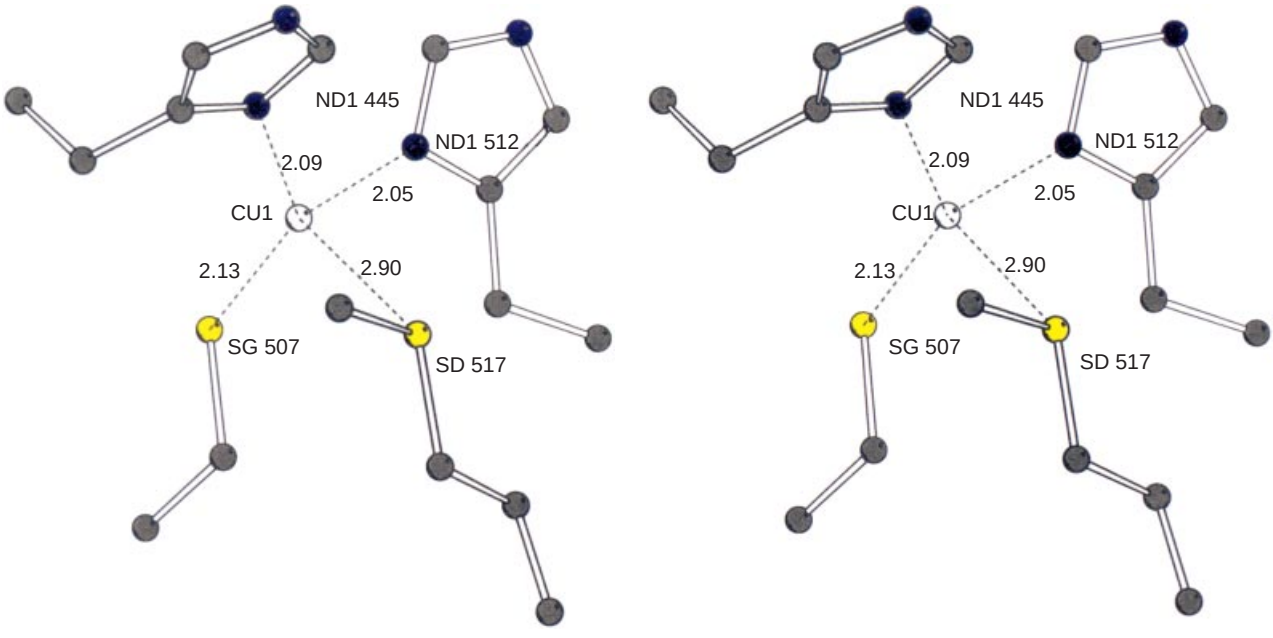

Figure 44 Stereo drawing of the mononuclear copper site in domain 3 of AO. The displayed bond distances are for subunit A; MOLSCRIPT.[100] (Reproduced by permission of World Scientific Publishing Co., from A. Messerschmidt, 'Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins', in *Multi Copper Oxidases*, ed. A. Messerschmidt, World Scientific Publishing, Singapore, 1997.)

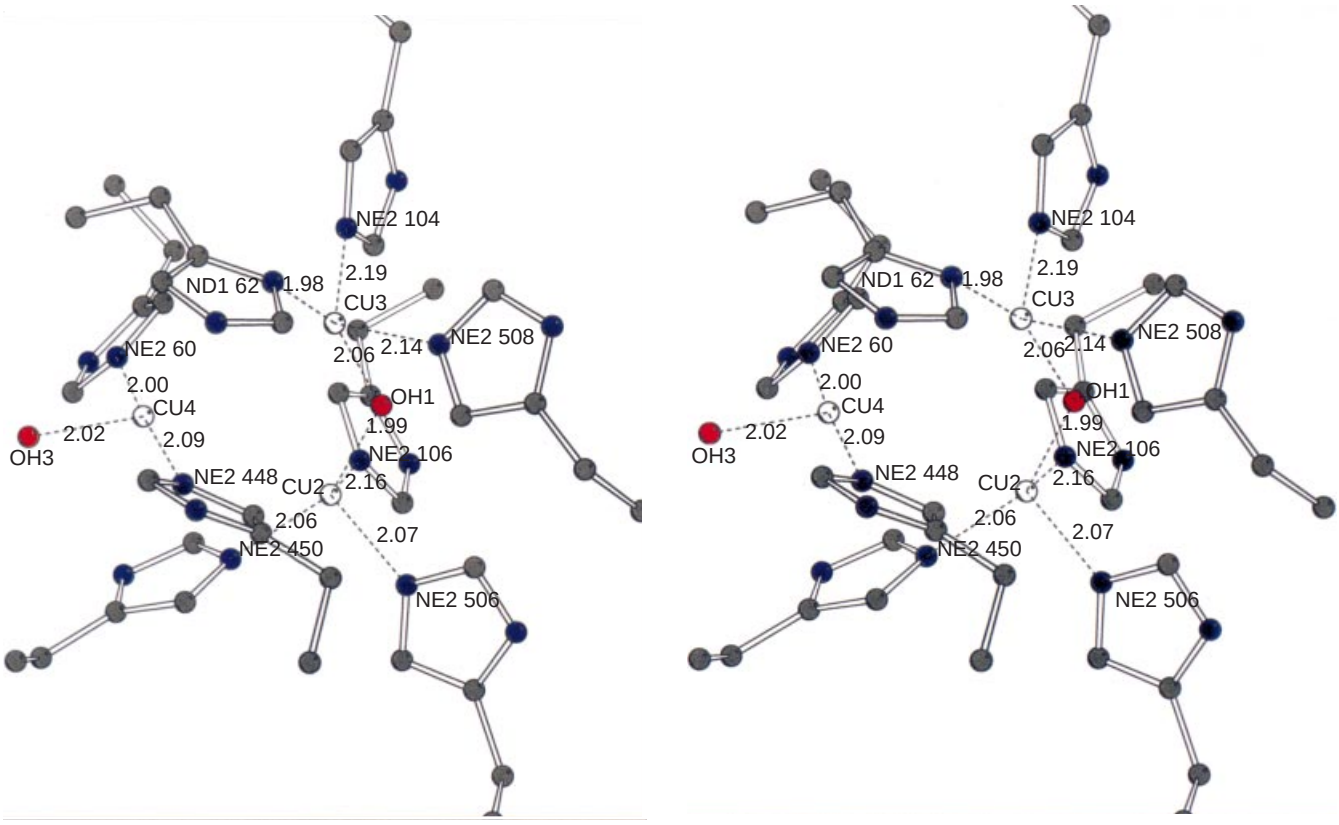

Figure 45 Stereo drawing of the trinuclear copper site of AO. The displayed bond distances are for subunit A; MOLSCRIPT.[100] (Reproduced by permission of World Scientific Publishing Co., from A. Messerschmidt, 'Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins', in *Multi Copper Oxidases*, ed. A. Messerschmidt, World Scientific Publishing, Singapore, 1997.)

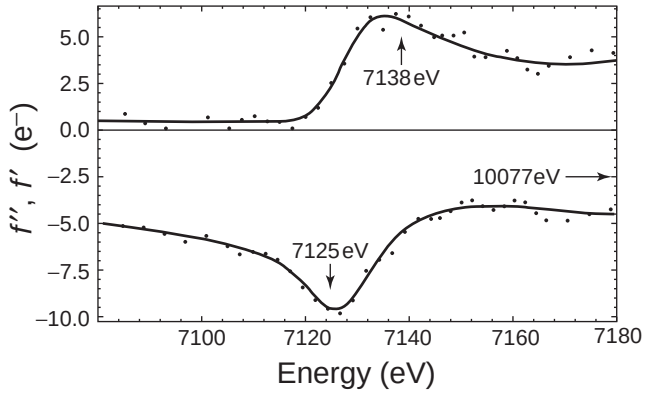

Figure 46 Dispersion terms as a function of the photon energy as obtained from an X-ray fluorescence scan from the cytochrome-c_{552} sample crystal and evaluation with DISCO.[110] (Reproduced by permission of Academic Press, Ltd., from Than et al.[109])

was registered using an NaI(Tl) scintillation counter. Evaluation of the spectrum with program DISCO[110] gave the anomalous dispersion contributions, f' and f'', as a function of photon energy (Figure 46). Subsequently, diffraction data were collected at three different photon energies. Two of them, 7138 and 7125 eV, correspond to maximum f'' and minimum f', respectively. The choice of the third, 10 077 eV, was for scaling purposes. A fourth high-resolution data set (1.28 Å) was collected at a photon energy of 11 810 eV. All data sets were obtained from a single crystal which had been shock-frozen at 90 K. Anomalous and dispersive Patterson syntheses were calculated from each of the MAD data sets. All showed strong Harker peaks from which the coordinates of the heme Fe atom (one Fe atom per asymmetric unit) could be determined easily. MAD phases were calculated with programs MADSYS[62] and MADPRB.[111] The resulting initial electron density map at a resolution of 2.1 Å was of high quality and allowed the construction of the complete atomic model into that map.

Cytochrome-*c* transports electrons between complexes III (cytochrome-*c* reductase) and IV (cytochrome-*c* oxidase) within the respiration chain of prokaryotic and eukaryotic organisms. Cytochrome-c_{552} from *Thermus thermophilus* is a soluble monoheme cytochrome with an unusual high molecular mass (14.2 kDa) and its

primary structure is only distantly related to other monoheme cytochrome amino acid sequences. Remarkable for this cytochrome is its high thermostability. The molecular structure of cytochrome-c_{552} from *Thermus thermophilus* is shown in Figure 47. The polypeptide chain is folded around the heme group. The N-terminal two-thirds of cytochrome-c_{552} cover spatial regions around the heme group that are similar to those observed for other cytochromes. The N- and C-termini and also the C-terminal clamp (thermohelices) with a stabilizing function are labeled in Figure 47. The core of the protein is highly hydrophobic and there is no internal water molecule present which is typical for other cytochromes. The presence of the ''thermohelices'' with their high concentration of ion pairs and hydrophobic contacts to the residual part of the molecule and the complete absence of internal water molecules could be responsible for the high thermal stability.

6.3 Large Molecular Assembly

6.3.1 Crystal Structure of 20S Proteasome from Yeast

The controlled degradation of proteins in the interior of cells is of central significance for many processes that range from cell cycle control and differentiation to cellular immune response. The target protein is labeled for destruction by the covalent linkage of a small protein called ubiquitin and is degraded after adenosine triphosphate (ATP)-driven unfolding in the closed internal chamber of a large protein complex which is known as 26S proteasome (molecular mass 2 000 000 Da). The core and the proteolytic chamber of the 26S proteasome are formed by the catalytic 20S particle (molecular mass 700 000 Da). This particle is flanked at each end by a so-called 19S cap which seems to be responsible for the recognition and unfolding of the ubiquitin-labeled target proteins. The eukaryotic catalytic 20S machinery consists of 14 different but related protein subunits that assemble to the overall structure of 28 protein chains, a monster of the molecular world. The structure of the 20S proteasome from *Saccharomyces cerevisiae* has been elucidated at atomic resolution.[112] This is one of the most complex structures that has been determined up to now, excluding symmetrical systems such as viral capsids. It shows that complexity itself is no limit to understanding structure on the atomic level.

The structure analysis is briefly outlined. Crystals could be obtained by the hanging drop vapor diffusion method with a final precipitant concentration of 12% MPD in the drop (pH 6.5). The crystals are monoclinic, space group $P2_1$, cell parameters $a = 135.7$ Å, $b = 301.8$ Å, $c = 144.7$ Å and $\beta = 112.6°$, and have one 20S particle per asymmetric unit. Data sets with two different inhibitors were collected at the wiggler beamline BW6 at DESY with radiation of $\lambda = 1.1$ Å and at 90 K temperature. The mean attainable resolution was 2.4 Å. A self-rotation function calculated at 5 Å resolution revealed the orientation of the local twofold axis. The atomic model of the 20S proteasome from *Thermoplasma acidophilum*[71] could be used for molecular replacement calculations [program AMORE[51]] at 3.5 Å resolution and a prominent solution could be found. The resultant electron density map was cyclically averaged about the local twofold axis. The individual subunits were identified according to their characteristic amino acid

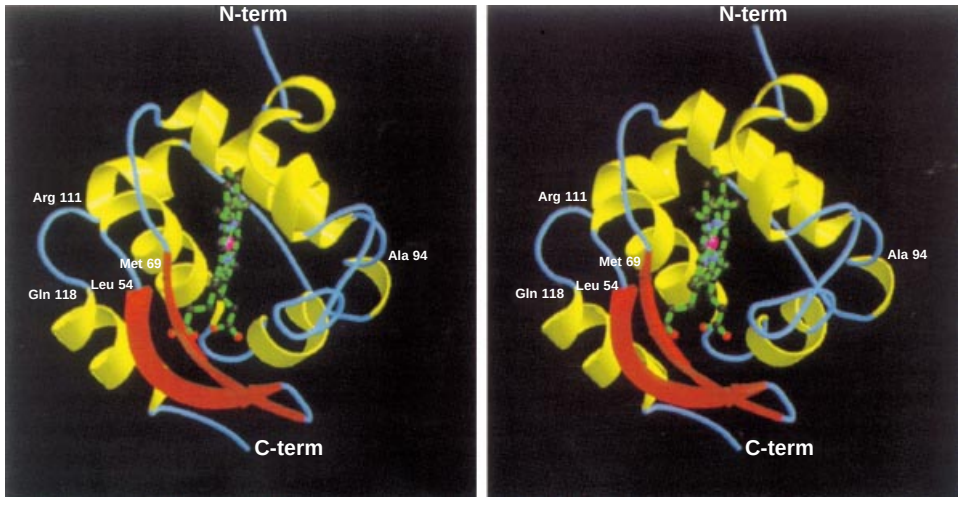

Figure 47 Stereo view of the cytochrome-c_{552} molecule from *Thermus thermophilus*. The prostetic heme group is presented as a ball-and-stick model. The N- and C-termini and the start (Ala94) of the C-terminal ''thermohelices'' are labeled. (Reproduced by permission of Academic Press, Ltd., from Than et al.[109])

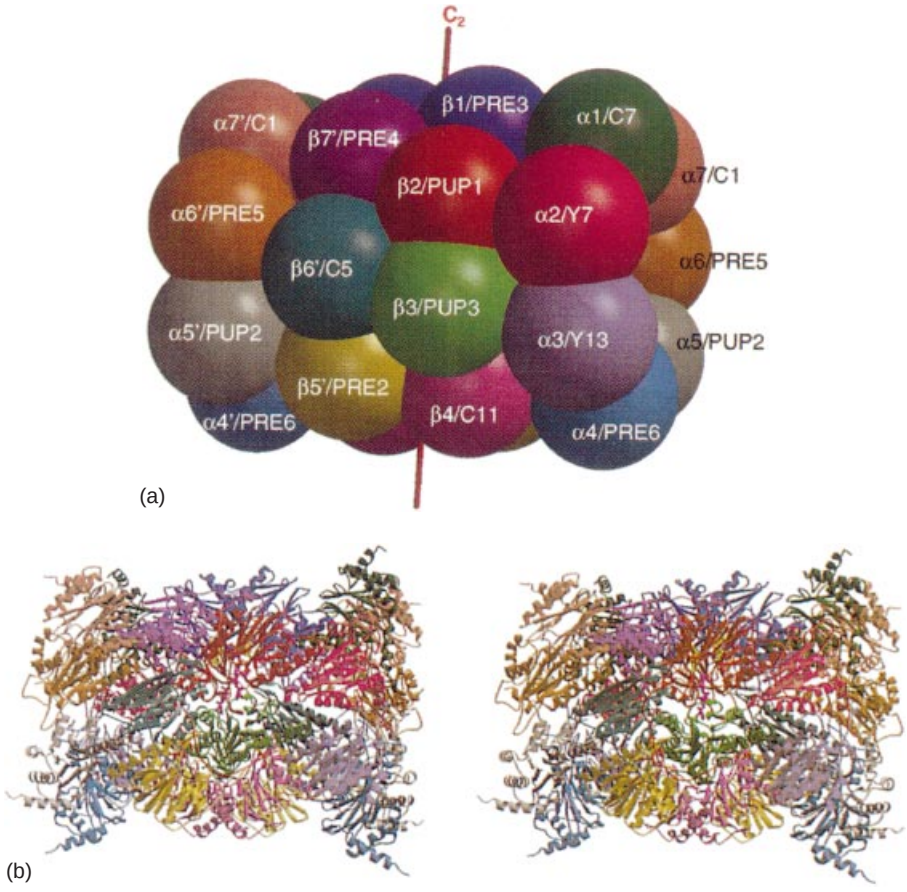

Figure 48 Topology of the 28 subunits of the yeast 20S proteasome drawn as (a) spheres and (b) ribbon representation; MOLSCRIPT,[100] RASTER3D.[101] (Reproduced by permission of MacMillan Magazines Ltd., from Groll et al.[112])

sequences and were built into the map. The final model with the lactacystin inhibitor was refined at a resolution of 2.4 Å to an *R*-factor of 0.26 consisting of 48 888 protein, 30 inhibitor, 18 magnesium atoms and 1800 solvent molecules.

The structure of the eukaryotic proteasome is important because it is much more complex than its archaebacterial relative which has two types of subunits, α and β, only. The α-subunits do not seem to be catalytic but they may self-assemble to form sevenfold rings. In contrast, the β-subunits show catalytic activity but are not able to assemble themselves. The two components assemble themselves to a stack of four rings, two outer α-rings which enclose an inner pair of β-rings. The eukaryotic proteasome retains the important property, however, that the single subunits of the sevenfold rings are different, probably reflecting the increased biological functions. The exact sevenfold symmetry of the particle is lost and a twofold symmetry remains solely to give seven different α- and seven different β-subunits (Figure 48a and b). In the proteasome, the unfolded protein is cut into peptide products in the inner chamber of the 20S particle and these have a length distribution with a center around octa- or nonapeptides. These sizes are appropriate to bind to MHC-class-I molecules.

ABBREVIATIONS AND ACRONYMS

AO	Ascorbate Oxidase
ATP	Adenosine Triphosphate
AVG	L-Aminoethoxyvinylglycine
CBL	Cystathionine β-Lyase
CCD	Charge-coupled Device
DESY	Deutsches Elektronensynchrotron
ESRF	European Synchrotron Radiation Facility
FAST	Fast Area Television Detector
IP	Image Plate
MAD	Multiple Anomalous Diffraction
MD	Molecular Dynamics
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement Anomalous Scattering
MPD	2,4-Methylpentanediol
MWPC	Multiwire Proportional Chamber

NMR	Nuclear Magnetic Resonance
PEG	Poly(ethyleneglycol)
PLP	Pyridoxal-5′-Phosphate
SA	Simulated Annealing
SIRAS	Single Isomorphous Replacement Anomalous Scattering
SIT	Silicon Intensified Target
TFA	β,β,β-Trifluoroalanine

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules

Nucleic Acids Structure and Mapping **(Volume 6)**
X-ray Structures of Nucleic Acids

Particle Size Analysis **(Volume 6)**
Diffraction in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
Protein–Drug Interactions • Protein–Oligonucleotide Interactions

X-ray Spectrometry **(Volume 15)**
Structure Determination, X-ray Diffraction for

REFERENCES

1. T. Hahn (ed.), *International Tables for Crystallography*, Reidel, Dordrecht, Vol. A, 1993.
2. A.A. Green, 'Studies in the Physical Chemistry of the Proteins. X. The Solubility of Hemoglobin in Solutions of Chlorides and Sulfates of Varying Concentration', *J. Biol. Chem.*, **95**, 47–66 (1932).
3. A.A. Green, 'Studies in the Physical Chemistry of the Proteins. VIII. The Solubility of Hemoglobin in Concentrated Salt Solutions. A Study of the Salting out of Proteins', *J. Biol. Chem.*, **93**, 495–516 (1931).
4. H.W. Wyckoff, C.H.W. Hirs, S.N. Timasheff (eds.), 'Diffraction Methods for Biological Macromolecules, Part A', *Methods Enzymol.*, **114**, 1–588 (1985).
5. H.W. Wyckoff, C.H.W. Hirs, S.N. Timasheff (eds.), 'Diffraction Methods for Biological Macromolecules, Part B', *Methods Enzymol.*, **115**, 1–485 (1985).
6. C.W. Carter Jr, R.M. Sweet (eds.), 'Macromolecular Crystallography, Part A', *Methods Enzymol.*, **276**, 1–700 (1997).
7. C.W. Carter Jr, R.M. Sweet (eds.), 'Macromolecular Crystallography, Part B', *Methods Enzymol.*, **277**, 1–664 (1997).
8. J. Jancarik, S.H. Kim, 'Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins', *J. Appl. Crystallogr.*, **24**, 409–411 (1991).
9. H. Michel, *Crystallization of Membrane Proteins*, CRC Press, Boca Raton, FL, 1991.
10. G.L. Gilliland, M. Tung, D.M. Blakeslee, J.E. Ladner, 'Biological Macromolecule Crystallization Database, Version 3.0: New Features, Data and the NASA Archive for Protein Crystal Growth Data', *Acta Crystallogr.*, **D50**, 408–413 (1994).
11. J.R. Helliwell, *Macromolecular Crystallography with Synchrotron Radiation*, Cambridge University Press, Cambridge, 1992.
12. K. Moffat, 'Ultrafast Time-resolved Crystallography', *Nature Struct. Biol., Synchrotron Suppl.*, 641–643 (1998).
13. W.C. Phillips, I. Rayment, 'A Systematic Method for Aligning Double-focusing Mirrors', *Methods Enzymol.*, **114**, 316–329 (1985).
14. R. Kahn, R. Fourme, 'Gas Proportional Detectors', *Methods Enzymol.*, **276**, 268–288 (1997).
15. E.M. Westbrook, I. Naday, 'Charge-coupled Device-based Area Detectors', *Methods Enzymol.*, **276**, 244–268 (1997).
16. D.W. Rodgers, 'Practical Cryocrystallography', *Methods Enzymol.*, **276**, 183–203 (1997).
17. T.-Y. Teng, 'Mounting of Crystals for Macromolecular Crystallography in a Free-standing Thin Film', *J. Appl. Crystallogr.*, **23**, 387–391 (1990).
18. U.W. Arndt, A.J. Wonnacott (eds.), *The Rotation Method in Crystallography*, North Holland, Amsterdam, 1977.
19. M.J. Buerger, *The Precession Method in X-ray Crystallography*, John Wiley & Sons, New York, 1964.
20. T.L. Blundell, L.N. Johnson, *Protein Crystallography*, Academic Press, New York, 1976.
21. J. Drenth, *Principles of Protein X-ray Crystallography*, Springer, New York, 90, 1994.
22. M.J. Buerger, *Crystal-structure Analysis*, John Wiley & Sons, New York, 1961.
23. A.L. Patterson, 'A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals', *Phys. Rev.*, **46**, 372–376 (1934).
24. K.S. Wilson, 'Illuminating Crystallography', *Nature Struct. Biol., Synchrotron Suppl.*, 627–630 (1998).
25. A. Messerschmidt, J.W. Pflugrath, 'Crystal Orientation and X-ray Pattern Prediction Routines for Area-detector Diffractometer Systems in Macromolecular Crystallography', *J. Appl. Crystallogr.*, **20**, 306–315 (1987).
26. W. Kabsch, 'Automatic Indexing of Rotation Diffraction Patterns', *J. Appl. Crystallogr.*, **21**, 67–81 (1988).
27. W. Kabsch, 'Evaluation of Single-crystal X-ray Diffraction Data from Position-sensitive Detector', *J. Appl. Crystallogr.*, **21**, 916–924 (1988).
28. A.J. Howard, C. Nielsen, Ng.H. Xuong, 'Software for a Diffractometer with Multiwire Area Detector', *Methods Enzymol.*, **114**, 452–472 (1985).

29. A.J. Howard, G.L. Gilliland, B.C. Finzel, T.L. Poulos, D.H. Ohlendorf, F.R. Salemne, 'The Use of an Imaging Proportional Counter in Macromolecular Crystallography', *J. Appl. Crystallogr.*, **20**, 383–387 (1987).
30. A. Leslie, 'Mosflm User Guide, Mosflm Version 5.41', in *Data Collection and Processing*, Proceedings of the CCP4 Study Weekend, 29–30 January 1993, eds. L. Sawyer, N. Isaac, S. Bailey, HRC Laboratory of Molecular Biology, Cambridge, UK, 44–51, 1994.
31. Z. Otwinowski, 'The Denzo Program Package', in *Data Collection and Processing*, Proceedings of the CCP4 Study Weekend, 29–30 January 1993, eds. N. Isaac, S. Bailey, SERC Daresbury Laboratory, Warrington, UK, 56–62, 1993.
32. W. Kabsch, 'Data Collection and Evaluation with Program XDS', in *Data Collection and Processing*, Proceedings of the CCP4 Study Weekend, 29–30 January 1993, eds. L. Sawyer, N. Isaac, S. Bailey, SERC Daresbury Laboratory, Warrington, UK, 63–70, 1993.
33. R. Diamond, 'Profile Analysis in Single Crystal Diffractometry', *Acta Crystallogr.*, **A25**, 43–55 (1969).
34. G. Ford, 'Intensity Determination by Profile Fitting Applied to Precession Photographs', *J. Appl. Crystallogr.*, **7**, 555–564 (1974).
35. M.G. Rossmann, 'Processing Oscillation Diffraction Data for Very Large Unit Cells with an Automatic Convolution Technique and Profile Fitting', *J. Appl. Crystallogr.*, **12**, 225–238 (1979).
36. CCP4, 'Collaborative Computational Project No. 4', *Acta Crystallogr.*, **D50**, 760–763 (1994).
37. D. Gerwirth, *The HKL Manual – A Description of Programs Denzo, XDisplayf and Scalepack*, 4th edition, Yale University, New Haven, CT, 1994.
38. R.G. Pearson, 'Hard and Soft Acids and Bases', *Survey*, **5**, 1–52 (1969).
39. F.H.C. Crick, B.S. Magdoff, 'The Theory of the Method of Isomorphous Replacement', *Acta Crystallogr.*, **9**, 901–908 (1956).
40. J. Knäblein, T. Neuefeind, F. Schneider, A. Bergner, A. Messerschmidt, J. Löwe, B. Steipe, R. Huber, '$Ta_6Br_{12}^{2+}$, a Tool for Phase Determination of Large Biological Assemblies by X-ray Crystallography', *J. Mol. Biol.*, **270**, 1–7 (1997).
41. R.W. James, in *The Optical Principles of the Diffraction of X-rays: Vol. II, The Crystalline State*, Bell, London, 135–167, 1960.
42. H. Hönl, 'Zur Dispersionstheorie der Röntgenstrahlen', *Z. Phys.*, **84**, 1–16 (1933).
43. D.T. Cromer, D. Liberman, 'Relativistic Calculation of Anomalous Scattering Factors for X-rays', *J. Chem. Phys.*, **53**, 1891–1898 (1970).
44. W.A. Hendrickson, C.M. Ogata, 'Phase Determination from Multiwavelength Anomalous Diffraction Measurements', *Methods Enzymol.*, **276**, 494–523 (1997).
45. W.A. Hendrickson, J.R. Horton, D.M. LeMaster, 'Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle for Direct Determination of Three-dimensional Structure', *EMBO J.*, **9**, 1665–1672 (1990).
46. W. Yang, W.A. Hendrickson, E.T. Kalman, R.J. Crouch, 'Expression, Purification, and Crystallization of Natural and Selenomethionyl Recombinant Ribonuclease H from *Escherichia coli*', *J. Biol. Chem.*, **265**, 13553–13559 (1990).
47. W. Hoppe, 'Die Faltmolekülmethode: Eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannten Molekülstrukturen', *Z. Elektrochem.*, **61**, 1076–1083 (1957).
48. M.G. Rossman, D.M. Blow, 'The Detection of Subunits Within the Crystallographic Asymmetric Unit', *Acta Crystallogr.*, **15**, 24–31 (1962).
49. R.A. Crowther, D.M. Blow, 'A Method of Positioning a Known Molecule in an Unknown Crystal Structure', *Acta Crystallogr.*, **23**, 544–548 (1967).
50. R. Huber, 'Die Automatisierte Faltmolekülmethode', *Acta Crystallogr.*, **19**, 353–356 (1965).
51. J. Navaza, 'AMoRe: An Automated Package for Molecular Replacement', *Acta Crystallogr.*, **A50**, 157–163 (1994).
52. L. Tong, M.G. Rossmann, 'The Locked Rotation Function', *Acta Crystallogr.*, **A46**, 783–792 (1990).
53. A.T. Brünger, 'Extension of Molecular Replacement: New Search Strategy Based on Patterson Correlation Refinement', *Acta Crystallogr.*, **A46**, 46–57 (1990).
54. A.T. Brünger, *XPLOR, Version 3.1, a System for Crystallography and NMR*, Yale University Press, New Haven, CT, 1992.
55. W. Steigemann, 'Die Entwicklung und Anwendung von Rechenverfahren und Rechenprogrammen zur Strukturanalyse von Proteinen am Beispiel des Trypsin-Trysininhibitorkomplexes, des freien Inhibitors und der L-Asparaginase,' PhD Thesis, Technische Universität München, 1974.
56. M.G. Rossmann, 'The Accurate Determination of the Position and Shape of Heavy-atom Replacement Groups in Proteins', *Acta Crystallogr.*, **13**, 221–226 (1960).
57. E. Dodson, M. Vijayan, 'The Determination and Refinement of Heavy-atom Parameters in Protein Heavy-atom Derivatives: Some Model Calculations Using Acentric Reflections', *Acta Crystallogr.*, **B27**, 2402–2411 (1971).
58. R.E. Dickerson, J.E. Weinzierl, R.A. Palmer, 'A Least-squares Refinement Method for Isomorphous Replacement', *Acta Crystallogr.*, **B24**, 997–1003 (1968).
59. D.M. Blow, F.H.C. Crick, 'The Treatment of Errors in the Isomorphous Replacement Method', *Acta Crystallogr.*, **12**, 794–802 (1959).
60. Z. Otwinowski, 'Maximum Likelihood Refinement of Heavy Atom Parameters', in *Isomorphous Replacement and Anomalous Scattering*, eds. W. Wolf, P.R. Evans, A.G.W. Leslie, SERC Daresbury Laboratory, Warrington, 80–86, 1991.

61. E. de La Fortelle, G. Bricogne, 'Maximum-likelihood Heavy-atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods', *Methods Enzymol.*, **276**, 472–494 (1997).
62. W.A. Hendrickson, 'Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation', *Science*, **254**, 51–58 (1991).
63. B.-C. Wang, 'Resolution of Phase Ambiguity in Macromolecular Crystallography', *Methods Enzymol.*, **115**, 90–112 (1985).
64. K.Y.J. Zhang, P. Main, 'The Use of Sayre's Equation with Solvent Flattening and Histogram Matching for Phase Extension and Refinement of Protein Structures', *Acta Crystallogr.*, **A46**, 377–381 (1990).
65. G.J. Kleywegt, T.A. Jones, 'Halloween ... Masks and Bones', in *From First Map to Final Model*, eds. S. Bailey, R. Hubbard, D. Waller, Proceedings of the Study Weekend, SERC Daresbury Laboratory, Warrington, UK, 59–66, 1994.
66. D. Turk, 'Weiterentwicklung eines Programmes für Molekülgrafik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Struktur-Aufklärungen', PhD Thesis, Technische Universität München, 1992.
67. W.P.J. Gaykema, W.G.J. Hol, J.M. Vereijken, N.M. Soeter, H.J. Bak, J.J. Beintema, '3.2 Å Structure of the Copper-containing, Oxygen-carrying Protein *Panulirus interruptus* Haemocyanin', *Nature (London)*, **309**, 23–29 (1984).
68. W.P.J. Gaykema, A. Volbeda, W.G. Hol, 'Structure Determination of *Panulirus interruptus* Haemocyanin at 3.2 Å Resolution: Successful Phase Extension by Sixfold Density Averaging', *J. Mol. Biol.*, **187**, 255–275 (1985).
69. M.G. Rossmann, E. Arnold, J.W. Erikson, E.A. Frankenberger, J.P. Griffith, H.-J. Hecht, J.E. Johnson, G. Kamer, M. Luo, A.G. Mosser, R.R. Rueckert, B. Sherry, G. Vriend, 'Structure of a Human Common Cold Virus and Functional Relationship to Other Picornaviruses', *Nature (London)*, **317**, 145–153 (1985).
70. R. Ladenstein, M. Schneider, R. Huber, H. Bartunik, K. Schott, A. Bacher, 'Heavy Riboflavin Synthase from *Bacillus subtilis*: Crystal Structure Analysis of the Icosahedral Beta-60 Capsid at 3.3 Å Resolution', *J. Mol. Biol.*, **203**, 1045–1070 (1988).
71. J. Löwe, D. Stock, B. Jap, P. Zwickl, W. Baumeister, R. Huber, 'Crystal Structure of the 20S Proteasome from the Archaeon *Thermoplasma acidophilum* at 3.4 Å Resolution', *Science*, **268**, 533–539 (1995).
72. W.A. Hendrickson, E.E. Lattman, 'Representation of Phase Probability Distributions for Simplified Combination of Independent Phase Information', *Acta Crystallogr.*, **B26**, 136–143 (1970).
73. G.A. Sim, 'The Distribution of Phase Angles for Structures Containing Heavy Atoms. II. A Modification of the Normal Heavy-atom Method for Non-centrosymmetrical Structures', *Acta Crystallogr.*, **12**, 813–815 (1959).
74. R.J. Read, 'Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors', *Acta Crystallogr.*, **A42**, 140–149 (1986).
75. T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjelgaard, 'Improved Methods for Building Protein Models in Electron Density Maps and Location of Errors in These Models', *Acta Crystallogr.*, **A47**, 110–119 (1991).
76. T.A. Jones, 'A Graphics Model Building and Refinement System for Macromolecules', *J. Appl. Crystallogr.*, **15**, 24–31 (1978).
77. A. Roussel, C. Cambilleau, *Turbo-Frodo in Silicon Graphics Geometry, Partners Directory*, Silicon Graphics, Mountain View, CA, 1989.
78. J. Greer, 'Three-dimensional Pattern Recognition: An Approach to Automated Interpretation of Electron Density Maps of Proteins', *J. Mol. Biol.*, **82**, 279–301 (1974).
79. A. Jack, M. Levitt, 'Refinement of Large Structures by Simultaneous Minimization of Energy and *R*-Factor', *Acta Crystallogr.*, **A34**, 931–935 (1978).
80. D.E. Tronrud, L.F. Ten Eyk, B.W. Matthews, 'An Efficient General-purpose Least-squares Refinement Program for Macromolecular Structures', *Acta Crystallogr.*, **A43**, 489–501 (1987).
81. W.A. Hendrickson, 'Stereochemically Restrained Refinement of Macromolecular Structures', *Methods Enzymol.*, **115**, 252–270 (1985).
82. R.A. Engh, R. Huber, 'Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement', *Acta Crystallogr.*, **A47**, 392–400 (1991).
83. A.T. Brünger, J. Kuriyan, M. Karplus, 'Crystallographic *R*-Factor Refinement by Molecular Dynamics', *Science*, **235**, 458–460 (1987).
84. A.T. Brünger, M. Nilges, 'Computational Challenges for Macromolecular Structure Determination by X-ray Crystallography and Solution NMR Spectroscopy', *Q. Rev. Biophys.*, **26**, 49–125 (1993).
85. G.N. Murshudov, A.A. Vagin, E.J. Dodson, 'Refinement of Macromolecular Structures by the Maximum-likelihood Method', *Acta Crystallogr.*, **D53**, 240–255 (1997).
86. A.T. Brünger, P.D. Adams, G.M. Clore, W.L. Delano, P. Gros, R.W. Grossekunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, 'Crystallography and NMR System – a New Software Suite for Macromolecular Structure Determination', *Acta Crystallogr.*, **D54**, 905–921 (1998).
87. Z. Dauter, V.S. Lamzin, K.S. Wilson, 'The Benefits of Atomic Resolution', *Curr. Opin. Struct. Biol.*, **7**, 681–688 (1997).
88. G.M. Sheldrick, 'SHELX: High-resolution Refinement', *Methods Enzymol.*, **277**, 319–343 (1997).

89. A.T. Brünger, 'Free *R* Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures', *Nature (London)*, **355**, 472–475 (1992).
90. A.T. Brünger, 'Assessment of Phase Accuracy by Cross Validation: The Free *R* Value. Methods and Applications', *Acta Crystallogr.*, **D49**, 24–36 (1993).
91. V. Luzzati, 'Traitement Statistique des Erreurs dans la Determination des Structures Crystallines', *Acta Crystallogr.*, **A5**, 802–810 (1952).
92. R.J. Read, 'Structure-factor Probabilities for Selected Structures', *Acta Crystallogr.*, **A46**, 900–912 (1990).
93. G.N. Ramachandran, C. Ramakrishnan, V.J. Sasisekharan, 'Stereochemistry of Polypeptide Chain Configurations', *J. Mol. Biol.*, **7**, 95–99 (1963).
94. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, 'PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures', *J. Appl. Crystallogr.*, **26**, 283–291 (1993).
95. R.A. Laskowski, D.S. Moss, J.M. Thornton, 'Main-chain Bond Lengths and Bond Angles in Protein Structures', *J. Mol. Biol.*, **231**, 1049–1067 (1993).
96. M.W. MacArthur, R.A. Laskowski, J.M. Thornton, 'Knowledge-based Validation of Protein-structure Coordinates Derived by X-ray Crystallography and NMR Spectroscopy', *Curr. Opin. Struct. Biol.*, **4**, 731–737 (1994).
97. R. Hooft, G. Vriend, '*WHAT_IF Program Manual*, EMBL, Heidelberg, Chapter 23, 1996.
98. E.E. Abola, J.L. Sussman, J. Prilusky, N.O. Manning, 'Protein Data Bank Archives of Three-dimensional Macromolecular Structures', *Methods Enzymol.*, **277**, 556–590 (1997).
99. T. Clausen, R. Huber, B. Laber, H.-D. Pohlenz, A. Messerschmidt, 'Crystal Structure of the Pyridoxal-5′-phosphate Dependent Cystathionine β-Lyase from *Escherichia coli* at 1.83 Å', *J. Mol. Biol.*, **262**, 202–224 (1996).
100. P.J. Kraulis, 'MOLSCRIPT: A Program to Produce both Detailed and Schematic Plots of Protein Structures', *J. Appl. Crystallogr.*, **24**, 946–950 (1991).
101. E.A. Merrit, N.E.P. Murphy, 'RASTER3D Version 2.0. A Program for Photorealistic Molecular Graphics', *Acta Crystallogr.*, **D50**, 869–873 (1994).
102. T. Clausen, R. Huber, A. Messerschmidt, H.-D. Pohlenz, B. Laber, 'Slow-binding Inhibition of *Escherichia coli* Cystathionine β-Lyase by L-Aminoethoxyvinylglycine: A Kinetic and X-ray Study', *Biochemistry*, **36**, 12633–12643 (1997).
103. S.V. Evans, 'SETOR: Hardware Lighted Three-dimensional Solid Model Representation of Macromolecules', *J. Mol. Graphics*, **11**, 134–138 (1993).
104. M.T. Stubbs, W. Bode, 'A Player of Many Parts: The Spotlight Falls on Thrombin's Structure', *Thromb. Res.*, **69**, 1–58 (1993).
105. A. van der Locht, D. Lamba, M. Bauer, R. Huber, T. Friedrich, B. Kröger, W. Höffken, W. Bode, 'Two Heads are Better than One: Crystal Structure of the Insect Derived Double Domain Kazal Inhibitor Rhodniin in Complex with Thrombin', *EMBO J.*, **14**, 5149–5157 (1995).
106. A. Messerschmidt, A. Rossi, R. Ladenstein, R. Huber, M. Bolognesi, G. Gatti, A. Marchesini, R. Petruzzelli, A. Finazzi-Agro, 'X-ray Crystal Structure of the Blue Oxidase Ascorbate Oxidase from Zucchini: Analysis of the Polypeptide Fold and a Model of the Copper Sites and Ligands', *J. Mol. Biol.*, **206**, 513–529 (1989).
107. A. Messerschmidt, R. Ladenstein, R. Huber, M. Bolognesi, L. Avigliano, R. Petruzzelli, A. Rossi, A. Finazzi-Agro, 'Refined Crystal Structure of Ascorbate Oxidase at 1.9 Å Resolution', *J. Mol. Biol.*, **224**, 179–205 (1992).
108. A. Messerschmidt, H. Luecke, R. Huber, 'X-ray Structures and Mechanistic Implications of Three Functional Derivatives of Ascorbate Oxidase from Zucchini', *J. Mol. Biol.*, **230**, 997–1014 (1993).
109. M.E. Than, P. Hof, R. Huber, G.P. Bourenkov, H.D. Bartunik, G. Buse, T. Soulimane, '*Thermus thermophilus* Cytochrome-c_{552}: A New Highly Thermostable Cytochrome-*c* Structure Obtained by MAD Phasing', *J. Mol. Biol.*, **271**, 629–644 (1997).
110. K.D. Eichhorn, *DISCO: Calculation of the Anomalous Dispersion Corrections, f′ and f″, to the Atomic X-ray Form Factor from Both EXAFS and Theory*, Deutsches Elektronensynchrotron DESY, Hamburg, 1985.
111. A.M. Friedman, 'Data Reduction and Scaling for MAD (with Some Comments on Phase Determination)', *Acta Crystallogr.*, **A52**, C56–C56 (1996).
112. M. Groll, L. Ditzel, J. Löwe, D. Stock, M. Bochtler, H.D. Bartunik, R. Huber, 'Structure of 20S Proteasome from Yeast at 2.4 Å Resolution', *Nature (London)*, **386**, 463–471 (1997).

Pesticide Analysis: Introduction

Antonio Di-Corcia
Università di Roma ''La Sapienza'', Rome, Italy

The term *pesticide* is used to indicate any substance, preparation or organism used for destroying pests. This broad definition covers substances used for many purposes. They are as follows: insecticides, herbicides, fungicides, nematocides, acaricides, lumbricides, growth regulators, insect repellants. The latter two groups are usually included as pesticides, even though they do not kill the target organisms. An enormous spectrum of chemical types of pesticides covers the definition of pesticides. According to their chemical nature, a first rough classification distinguishes between organic and inorganic pesticides. Actually, organic chemical pesticides receive virtually all of the regulatory attention and public concern and these will be the subject of this series of articles. The most important classes of pesticides are the following: organochlorines, organophosphorus, carbamates, triazines, phenoxyacids, phenyl and sulfonyl ureas, acetoanilides, benzimidazoles, pyrethroids, diphenylethers, and the recently introduced imidazolinones (see **Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of**).

The meaning of the term *trace* has changed through time. During the past 35 years, the meaning of trace analysis of pesticides in real-life samples has changed from parts per million (ppm: 1 part in 10^6), to parts per billion (ppb: 1 part in 10^9), and currently to parts per trillion (ppt: 1 part in 10^{12}). Presently, some methods can also detect pesticides at parts per quadrillion levels. In this case, the term *ultratrace* is sometimes used.

Before the introduction of organochlorine and organophosphate insecticides in the mid-1940s, the history of pesticide trace analysis coincides with the history of the determination of arsenic-based pesticide residues in foodstuffs. Following a number of deaths occurring between 1855 and 1900, the English Royal Commission, chaired by Lord Kelvin, set a residue limit for arsenic of 1.43 ppm. Even after setting this standard, the controversy about arsenic and some other inorganics (lead, mercury, copper and thallium) used as insecticides and fungicides continued and ceased only after World War II, when less toxic synthetic organic insecticides replaced inorganic ones. At that time, apart from some biochemical assays involving inhibition of the cholinesterase enzyme for monitoring organophosphate insecticides, spectroscopic methods of analysis in connection with color-producing reactions were the only methods used that had the sensitivity needed for analysis of a pesticide at the ppm level. Following the introduction of dichlorodiphenyltrichloroethane (DDT), the first practical method for determining DDT residues in foods was proposed in 1945 and persisted until it was later replaced by gas chromatographic methods. While laboratory scientists were slowly developing new spectrophotometry-based methods for determining ppm residues in food of new insecticides of the same class of DDT, that is organochlorine insecticides (see **Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of**), and some recently introduced organophosphate (see **Organophosphorus Pesticides in Water and Food, Analysis of**) insecticides, the latter scenario was changing both rapidly and greatly. Starting from the early 1950s, the number of organic pesticides adopted for crop protection dramatically increased. Several other synthetic compounds increased the classes of organochlorines and organophosphates which, together with the new class of carbamates (see **Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of**), came into wide use against insects. Phenoxyacids (see **Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of**), triazines (see ***s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of**) and phenylureas (see **Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of**) replaced chlorates and borates for inhibiting the growth of weeds. Dithiocarbamates replaced copper salts and sulfur as fungicides. In addition, a number of organics were increasingly used as nematocides, rodenticides and fumigants. As compared to inorganic pesticides, the relatively low acute toxicity of these new pesticides encouraged their widespread use and abuse for agricultural and nonagricultural purposes. As a consequence of their widespread use, any kind of environmental compartment was increasingly contaminated by pesticides.

In the early 1960s, the vulnerability of surface water to pesticides was recognized and bioaccumulation in fish, zooplankton and aquatic plants investigated. In the same years, evidence was obtained for leaching of pesticides through the soil and contamination of groundwaters. At the same time, the science of modern toxicology brought to light that chronic exposure to pesticides was the most serious problem. Several long-term pathologies, including cancer, mutagenesis and teratogenesis, could be provoked by daily intake of very small amounts of pesticides and related compounds. It appeared, then, that colorimetric methods were inadequate for monitoring a very large number of small amounts of pesticides having a broad range of polarity and different functional groups in a wide variety of complex sample matrices.

In the early 1970s, the development of gas chromatography (GC) with capillary columns coupled to sensitive and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

selective detectors (see **Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis**) revolutionized pesticide trace analysis. Application of this technique allowed separation at nanogram levels of many pesticides from each other and often from potential interfering compounds into a single run. This initiated the development of the multiresidue analysis of pesticides (see **Multiclass, Multiresidue Analysis of Pesticides, Strategies for**).

Since the first research papers reported in the mid-1960s describing coupling of a mass spectrometer with a gas chromatograph (see **Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis**) and later with a liquid chromatograph (see **High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis**), it appeared that this highly specific technique could become the ultimate tool to solve problems of false positives which may frequently occur in the analysis of pesticide traces in complex matrices. Data obtained by using a mass spectrometer as a detector are so reliable that today no analytical result is definitively accepted if it is not supported by mass spectra. The evolution of liquid chromatography into the analytical technique known as high-performance liquid chromatography (see **High-performance Liquid Chromatography Methods in Pesticide Residue Analysis**), occurring in the late 1970s, afforded the tool for analyzing those pesticides and pesticide degradation products that, being nonvolatile, highly polar, and thermally labile, were not amenable to direct gas chromatographic analysis.

In the last decade, the use of enzyme-linked immunoadsorbant assays for specifically and rapidly determining traces of a single pesticide or a single class of pesticides has been rapidly growing (see **Immunochemical Assays in Pesticide Analysis**). Immunoassays require decreased sample preparation and offer high sample throughput, so they can greatly help analysts when chromatographic methods are difficult and time-consuming.

Recently, capillary electrophoresis has emerged as a valid alternative technique to liquid chromatography for analyzing ionic and ionizable pesticides, such as phenoxyacids, some cationic herbicides (see **Pesticides (New Generation) and Related Compounds, Analysis of**) and sulfonylureas (see **Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of**).

Analytical methods for determining pesticides have different purposes. The particular objective to be achieved can influence the choice of the analytical methodology.

- Compliance with standards. Any material destined for human consumption has to be tested to enforce national residue limits. Analytical methods for determining as many pesticides as possible are needed, even though single pesticide methods may also be required to attain the desired scope. The method sensitivity is not always the parameter of choice, as different countries can have very different maximum acceptable concentration levels.
- Emergency response. In this case, the speed of analysis is the principal requirement. Pollution incidents can pose a sudden need for analytical methods for an unpredictable range of pesticides. Method sensitivity and specificity are not key factors in these situations.
- Behavior and fate of pesticides. Methods of analysis are needed for monitoring studies aimed at understanding how pesticides behave in the environment so that predictive models can be developed and tested. The sensitivity and accuracy of such methods should be adequate for likely predicting compliance with national standards.
- Laboratory and field tests on new pesticides. Manufacturers of pesticides need sensitive and specific methods of analysis to study the persistence, the mobility and the degradation pathway of the pesticides in the laboratory and field.

A particular type of analysis is that aimed at determining concentration levels in a certain matrix of chiral pesticides (see **Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of**). Chiral compounds show enantiomeric selectivity in reactions with biological systems. Biological activity in soil or water environments may result in the preferential reactivity of one enantiomer of a pesticide in terms of microbial degradation, biological uptake, metabolism and/or toxicity.

Today, there are more than 1800 basic chemicals that are used as active ingredients of pesticides dispersed in approximately 33 600 formulations. Over the last 20 years, in the United States about 15×10^6 kg of pesticides were spread in the environment. This situation has urged local governments to introduce more restrictive regulations banning some dangerous pesticides and lowering the maximum admissible concentrations of pesticides in drinking water and foodstuffs. An enacted European Community Directive states that a single pesticide cannot be present in water destined for human consumption in concentrations higher than $0.1\ \mu g\ L^{-1}$.

The complex matter of pesticide trace analysis in environmental samples can be afforded by three basic approaches. The first approach is that of developing/using multi-residue methods (MRMs) able to determine as many as possible pesticides and pesticide metabolites having a broad range of polarity and structural properties in a wide array of sample matrices at the required sensitivity limit. These methods are employed usually by analysts in regulatory and compliance laboratories and by those conducting monitoring and surveillance

programs. Still today, GC remains the most popular technique for accomplishing MRMs, as demonstrated by the fact that the five principal MRMs invariably involve this technique as an effective identification/confirmation tool. After some pioneering works in 1963, several other MRMs followed as analysts have attempted to deal with the increasing number of pesticides. However, many pesticides and pesticide metabolites cannot be determined by these MRMs, because they are not amenable to direct GC analysis. In 1988, the Congress's Office of Technology Assessment of the USA reported that official MRMs can determine only 200 of the approximately 750 pesticides and related compounds that must be tested. This gap has stimulated researchers to develop MRMs based upon liquid chromatography with ultraviolet and mass spectrometric detection for determining those pesticide classes, such as phenylureas and carbamates, and pesticide metabolites which cannot be easily determined by GC. Very recently, a Canadian group of researchers has elaborated a method for determining 199 pesticides in vegetable matrices which involves both high-resolution GC/mass spectrometry and liquid chromatography with fluorimetric detection.

A second approach is that of developing/using selective MRMs for monitoring compounds of the same class or having similar structural properties. The US Environmental Protection Agency (USEPA) has recently developed a method (Method 531.1) for selectively determining carbamate insecticide residues in drinking water involving liquid chromatography and fluorescence detection with a post-column derivatization reactor.

The third approach is that of developing/using single residue methods (SRMs) that are effective for the analysis of one particular pesticide and its degradation products in a particular substrate. This approach is generally used by pesticide manufacturers who are responsible for supplying data from field and metabolism studies in support of registration of pesticides. SRMs are also followed by geochemists who are interested to know the fate of a pesticide in the environment and by analysts when a compound of particular interest cannot be included for a number of reasons in any of the usually adopted MRMs. Recently, the USEPA has developed a suitably tailored method based on high-resolution GC with selective detection for determining ethylenethiourea, which is a toxic degradation product of dithiocarbamate fungicides, in water.

These methods usually share at least one of the analytical preparative steps before final analyte identification and quantification. They may consist of isolating the analytes from the matrix and enriching them by solvent or solid-phase extraction, supercritical fluid extraction, of cleanup by liquid–liquid partitioning, adsorption onto suitable adsorbents and/or gel permeation chromatography, bioaffinity chromatography, sweep codistillation.

The aim of this series of articles is that of giving an up-to-date overview of the analytical methodologies and instruments used and emerging strategies for monitoring pesticide traces in environmental samples, such as water (see **Pesticides in Water: Sampling, Sample Preparation, Preservation**), vegetables, soil and sediments (see **Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation**) and biota (see **Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation** and **Herbicide Residues in Biota, Analysis of**).

ABBREVIATIONS AND ACRONYMS

DDT	Dichlorodiphenyltrichloroethane
GC	Gas Chromatography
MRMs	Multi-residue Methods
SRMs	Single Residue Methods
USEPA	US Environmental Protection Agency

Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation

Seiji Kawasaki and Jutaro Tadano
Saga Medical School Hospital, Nabeshima Saga, Japan

1 Introduction 1
1.1 Need for Pesticide Analysis 1
1.2 Characterization of the Measurement Method 1
1.3 Selection of Targets for Measurement 2
2 Sampling 2
2.1 Which Sample Should be Measured? 2
2.2 What Should be Measured? 3
3 Sample Preparation 4
3.1 Organophosphates 4
3.2 Carbamates 5
3.3 Organochlorinates 8
3.4 Herbicides 8
3.5 Other Pesticides 9
4 Sample Preservation 9
4.1 Blood (Plasma and Serum) 9
4.2 Urine 9
4.3 Other Biomedical Fluids 9
4.4 Tissues 9
5 Conclusion 10
Abbreviations and Acronyms 10
Related Articles 10
References 11

Environmental pollution is a world-wide problem. Consequently, the demand for pesticide analysis in biological samples is increasing. In most cases, the availability of the test result depends on the suitability of sampling, sample preservation, and sample preparation which must satisfy the need. However, there is no reference method. This paper summarizes sampling and sample preparation and preservation methods for the case when the compound to be detected is not known.

1 INTRODUCTION

1.1 Need for Pesticide Analysis

Environmental pollution has been known as a world-wide problem for many years, and pesticides are one of its primary causes. Pesticide use has been increasing in households, in agriculture, and on golf courses. It is recognized that pesticides are harmful to human beings, but it should also be remembered pesticides are chemical substances that are essential to maintaining a high quality of life for human beings. Therefore, we must learn to live with pesticides.

Residual pesticides in foods, drinks, and the environment (water, soil, etc.) have been analyzed in the fields of environmental chemistry, forensic chemistry, and preventive medicine, so the need for pesticide analysis is well established.

In the field of clinical medicine as well, it is extremely important, for therapeutic reasons, to identify and analyze pesticides taken into living bodies intentionally or accidentally. Nevertheless, neither standardized nor recommended analytical methods have yet been established for pesticides in biomedical samples.

1.2 Characterization of the Measurement Method

In many cases, the object of clinical medicine is a living body. Accordingly, pesticide analytical methods in clinical medicine must have three characteristics: specificity, rapidity, and sensitivity.

The pesticide analytical methods currently in use in clinical medicine are serum pseudo-cholinesterase (Ps-ChE, EC: 3.1.1.8) activity measurement as an indirect indicator of intoxication, and spectrophotometry,[1–2] thin-layer chromatography (TLC),[3–4] gas chromatography (GC),[5–15] gas chromatography/mass spectrometry (GC/MS),[16–20] high-performance liquid chromatography (HPLC),[21–26] and high-performance liquid chromatography/mass spectrometry (HPLC/MS)[27–29] as direct analytical methods.

Among them, serum Ps-ChE activity measurement can be employed in most laboratories. Serum Ps-ChE, however, shows low activity only for intoxication caused by organophosphorus pesticides (OP) and carbamate pesticides (CPs). The inhibition rate of Ps-ChE activity varies by type of pesticide, and does not reflect blood concentration. Accordingly, the measurement of Ps-ChE activity in serum is only an indirect indicator of pesticide intoxication. Even in cases in which Ps-ChE activity is utilized, in consideration of this limitation on examination, Ps-ChE activity also decreases in hepatic malfunction and hereditary hypocholinestrasemia cases, so other hepatic function

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

tests and/or acetyl-cholinesterase (Ac-ChE, EC: 3.1.1.7) activity in erythrocytes must be conducted at the same time. In addition, when treating pesticide-intoxicated patients with low Ps-ChE activity in serum, it is necessary to identify the causal substance – OP or CPs. This is because in the case of OP intoxication, the administration of paralidoxime-indole (PAM) is effective in enabling the early recovery of Ps-ChE activity. However, in the case of CP intoxication, PAM administration tends to delay and restrict the natural recovery of Ps-ChE activity, so its use is prohibited.

GC and HPLC, which feature easy operability, high measurement sensitivity, and specificity, are useful analytical methods employed in many laboratories in recent years. In both methods, however, peak identification is based on only the elution from the used column. Accordingly, when complex matrices such as biomedical samples are measured, a complex clean-up operation for samples is necessary in some cases.

However, although MS, particularly GC/MS, is regarded as an analytical method with extremely high sensitivity and specificity, it has the disadvantage of requiring an expensive instrument and a specialized operator.

1.3 Selection of Targets for Measurement

''Pesticide'' is a generic term for several hundred kinds of chemical substances. In life-saving treatment, however, there is no great difference among intoxication cases caused by different pesticides (Table 1). The only matter worthy of note is deciding whether PAM should be administered to victims of poisoning. As previously mentioned, the administration of PAM, which is an antagonist to relatively high-toxicity OP, promotes the recovery of the Ac-ChE activity in OP intoxication cases, but delays the natural recovery of Ac-ChE activity in cases of CP intoxication. Therefore, in clinical medicine, particularly emergency medicine, two groups – OP and CPs – are the most important targets for measurement.

In the fields of legal and preventive medicine, a wide range of compounds, including organochlorinates, herbicides, and dioxin-like compounds, should be selected as measurement targets.

Table 1 Basic treatment for pesticide intoxication cases

I. Remove the nonabsorbed pesticides
1. Promote vomiting
2. Gastric lavage
3. Dose with cathartic and adsorbent
4. Intestinal lavage
II. Remove the absorbed pesticides
1. Dose with diuretic
2. Blood purification (plasma exchange, blood perfusion)
3. Dose with antidote and antagonist

2 SAMPLING

2.1 Which Sample Should be Measured?

When pesticide intoxication is suspected, blood, urine, digestive tract contents or abluents, and tissues are mainly selected as the measurement sample. Among these, blood is the most important sample, as the detection of a residual pesticide in a biomedical sample in clinical medicine is conducted to save life. The subsequent course of pesticide-intoxicated patients normally depends on the type of pesticide (toxicity) and its concentration in the blood, so the blood concentration should be measured first. It is recommended that not less than 50 ml (if possible) of the circulating blood be sampled prior to the administration of other therapeutic drugs.

At the time of intoxication caused by trans-respiratory tract ingestion, however, the causal substance in the blood can rarely be identified. In addition, at oral ingestion, before the causal substance enters the bloodstream from the digestive tract, it is obviously pointless to sample the blood for measurement target. In such a case, it is effective to use the digestive tract contents or abluents as samples. Fortunately, these samples generally contain unchanged pesticides that have not yet been metabolized at high concentration, so pesticides can easily be detected. However, it is impossible to determine the prognosis of the patient, even though the causal substance can be identified and the ingested amount can be estimated.

If pesticide intoxication is suspected but cannot be proven based on testing of the blood, it is often effective to use urine as a measurement sample. In particular, most OP and CPs are easily metabolized in living bodies, as they decompose in urine and are excreted relatively early. The concentration of pesticides in the blood shows an elimination pattern similar to that of pharmaceuticals. That is, the pesticide concentration immediately after ingestion quickly decreases as the pesticides spread to the tissues (α curve), then, gradually drops due to metabolism and excretion (β curve). Regrettably, the half-life of individual pesticides in human blood is unclear, so we must refer to data obtained from experimental animals. In our experience, a major portion is excreted within 2–5 days.

Low-polarity pesticides (halogen-element OP, organochlorinates, thiophosphate esters, etc.) move from the bloodstream to tissues quickly, and tend to remain in organs. In such a case, it is useful to use fat tissues and organs as measurement samples, but it is often difficult to collect them from a living body.

2.2 What Should be Measured?

Pesticides taken into living bodies are metabolized through various routes, either enzymatically or non-enzymatically. Metabolites are excreted into urine at a high concentration for an extended period. It is therefore important to understand the metabolism of pesticides, and to detect not only pesticides but also metabolites at the same time. Pesticide metabolism in living bodies is briefly described below.

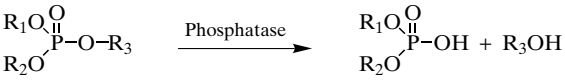

Scheme 1 Phosphate ester hydrolysis in phosphates.

2.2.1 Organophosphorus

OP are easily metabolized in living bodies. The primary metabolic route can be classified as (1) hydrolysis, (2) oxidation, and (3) transmethylation.

The hydrolysis of phosphate ester by phosphatase occurs specifically at a site of the elimination group in three ester bonds of OP (Scheme 1). Phosphatase, however, does not act on compounds (phosphorothionates and phosphorodithionates) that have the P=S bond. In malathion and dimethoate, side-chain ester hydrolysis takes place (Scheme 2).

Drug-metabolizing enzymes present in hepatic microsome cause various oxidation reactions, such as oxidation of the P=S bond to the P=O bond, oxidation of the thioether bond, the hydrolysis-like reaction of the P=S compound, and the O-deethylation or O-demethylation reaction (Scheme 3). Methyl-transferase causes O-demethylation, and then transmethylation to glutathione.

2.2.2 Carbamates

Here, we will explain metabolism in living bodies, using carbaryl (NAC), a typical CP, as an example.

NAC does not accumulate in living bodies, but is excreted through two metabolic routes.

In the first route, NAC is hydrolyzed into 1-naphthol and *N*-methylcarbamic acid. The former, 1-naphthol, further becomes glucuronic acid or sulfuric acid conjugates to be excreted outside the body. The latter, *N*-methylcarbamic acid, is decomposed into carbonic acid and methylamine.

In the second route, NAC metabolites are produced by methylation of the *N*-methyl group and hydroxylation of the naphthyl group (4-OH, 5-OH, and 5,6-dihydroxyl). These metabolites then turn to glucuronic acid or sulfuric acid conjugates to be excreted outside the body.

2.2.3 Organochlorinates

Generally, organochlorinates undergo virtually no enzymatic or non-enzymatic metabolism, and easily

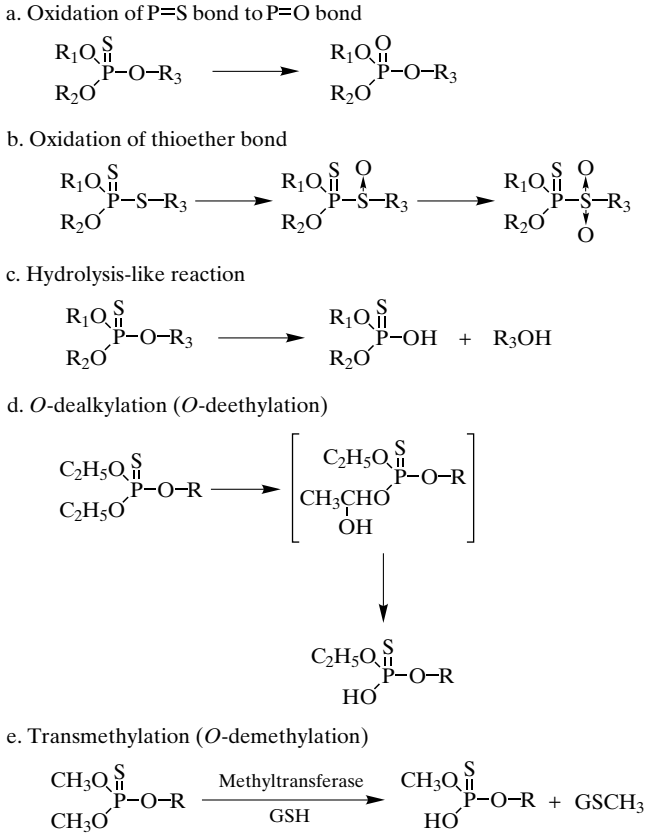

Scheme 3 Metabolism by drug-metabolizing enzyme in phosphorothionates, phosphorothioates and phosphorodithioates.

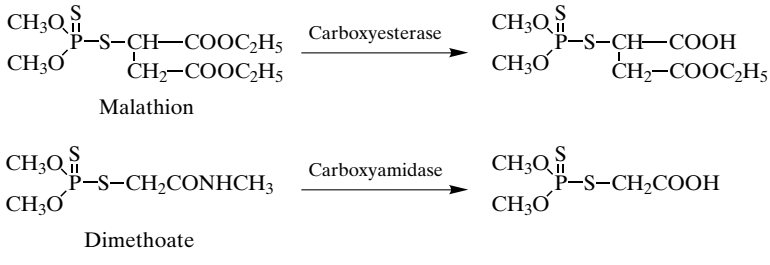

Scheme 2 Side-chain hydrolysis in malathion and dimethoate.

accumulate in living bodies. This, however, does not mean that they are never metabolized (Schemes 4–10).

Organochlorinates accumulated in fat circulate through fat tissues, the blood stream, and the liver. They are metabolized in the liver by various enzymes at a considerably slower rate than that of fat metabolism. That is, organochlorinates are converted to highly polar substances through various oxidation reactions, such as aromatic-ring and side-chain hydroxylation, O- or N-dealkylation, S- or N-oxidation, and epoxidization, and combinations of dehydrogenation, dehydrochlorination, reduction, hydrolysis, and exchange reaction. These substances are excreted as they are or as conjugated substances (glucuronic acid, sulfuric acid, amino acid). In such a case, unchanged organochlorinates and their metabolites are excreted not only into urine but also into bile in large quantities, and via the duodenum in the form of human waste.

3 SAMPLE PREPARATION

3.1 Organophosphates

Table 2 summarizes recent reports on sample preparation in the analysis of OP in biomedical samples.

Since OP are low-polarity compounds, organic solvents with a polarity lower than that of the sample matrix should be used for extraction, purification, and concentration.

3.1.1 Liquid–Liquid Extraction

In sample preparation for the analysis of OP in tissues or body fluids, extraction is firstly performed using MeCN.

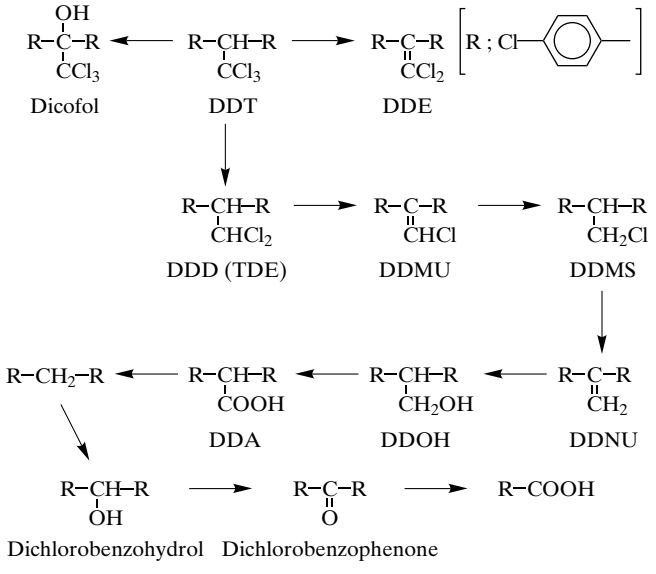

Scheme 4 Metabolism of *pp′*-DDT.

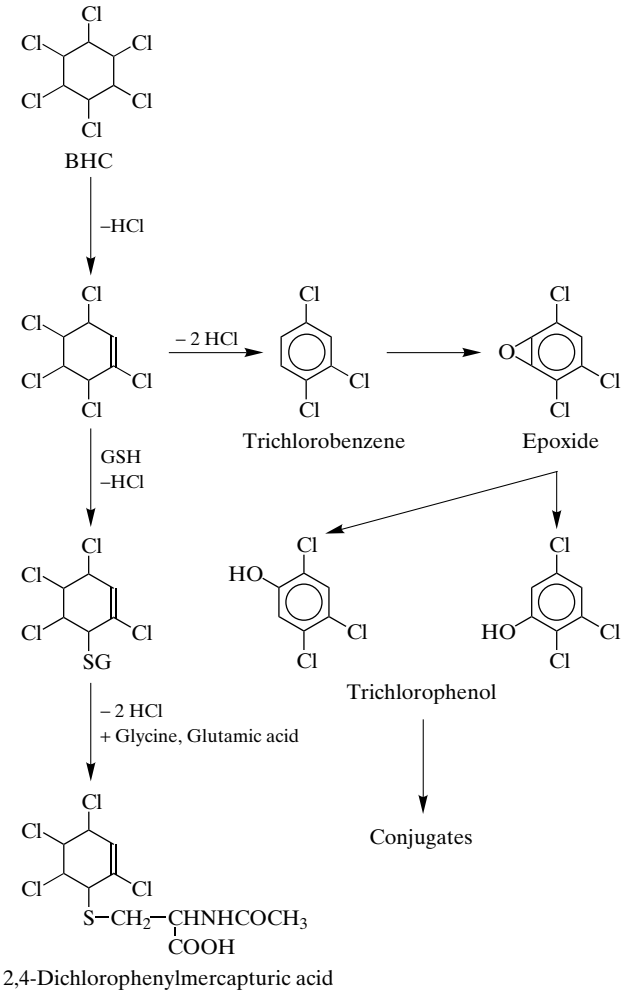

Scheme 5 Metabolism of γ-BHC.

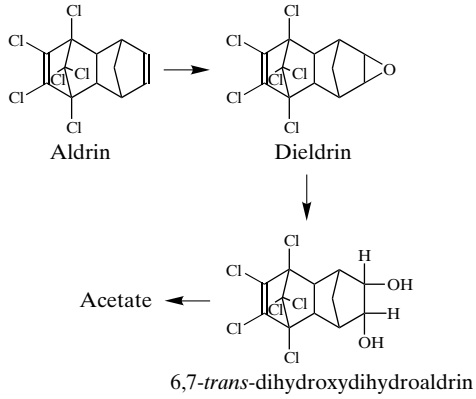

Scheme 6 Metabolism of aldrin and dieldrin.

Generally, OP in the extract are transferred to the *n*-hexane layer by distribution with NaCl and *n*-hexane, and become suitable for use as a measurement sample through the processes of drying using sodium sulfate anhydride and concentration at 50 °C or below. If the matrix contains

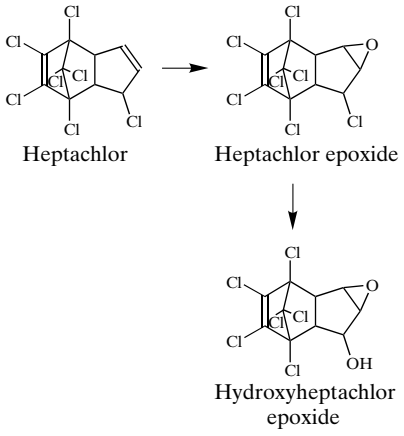

Scheme 7 Metabolism of heptachlor.

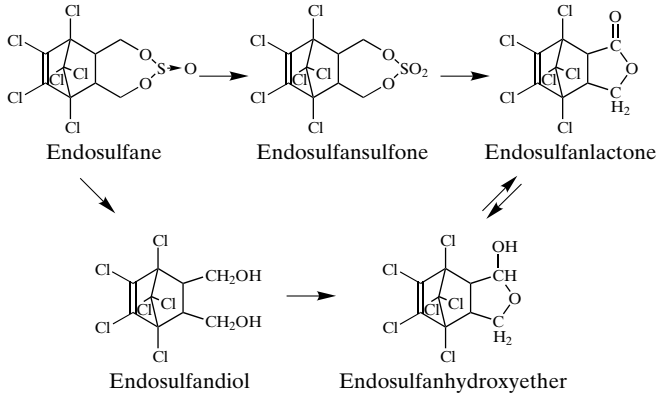

Scheme 8 Metabolism of endosulfane.

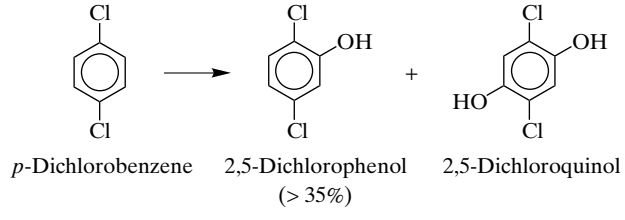

Scheme 9 Metabolism of *p*-dichlorobenzene.

a large quantity of fats, fats in the concentrated *n*-hexane layer must be removed. In the analysis of OP in the bovine liver, D.M. Holstage et al. removed fat components by a clean-up method combining GPC using an S-X3 biobeads column with solid-phase extraction (SPE) using a silica gel cartridge.[9] Organic solvents other than *n*-hexane used for distribution include chloroform, dichloromethane, and *n*-hexane/acetone liquid mixture.

Incidence of intoxication caused by glyphosate, a highly water-soluble OP, has increased in recent years. M. Tomita et al. analyzed glyphosate in human serum by deproteinization using perchloric acid (PCA) washing the supernatant with ethyl ether, and performing CE,[21] or

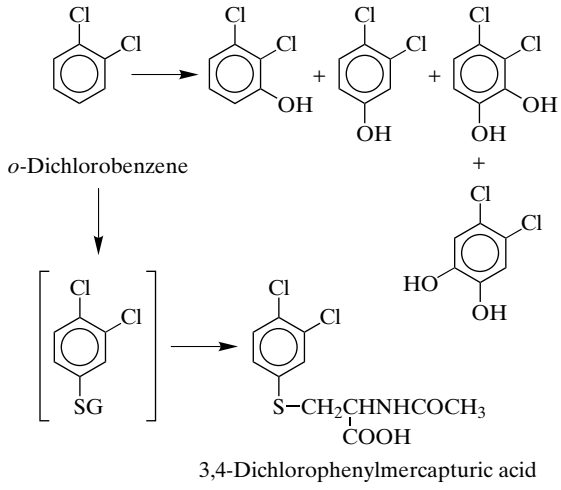

Scheme 10 Metabolism of *o*-dichlorobenzene.

by cleaning the deproteinized supernatant with PCA by anion exchange and using HPLC.[23]

3.1.2 Solid-phase Extraction

SPE is a technique frequently used to analyze pesticides in environmental water, but only a few cases have been reported in which only SPE has been used for analysis in biomedical samples. Kawasaki et al. used HPLC/MS to analyze 21 types of OP in human serum and urine. Using adsorption to a diatomaceous earth cartridge, elution with an *n*-hexane/diethyl ether mixture,[27] and an octadecasil (ODS) column, they employed the one-step sample clean-up method using elution with MeCN.[29]

3.2 Carbamates

3.2.1 Liquid–Liquid Extraction

Most CPs are insoluble or hardly soluble in water. CPs are highly soluble in organic solvents such as MeOH, EtOH, acetone, chloroform, and dichloromethane, but hardly soluble in petroleum-based hydrocarbons and xylene. Accordingly, organic solvents to be used for extracting CPs from samples must be selected with consideration of the solubility of the CPs to be detected, and the types of coexistent substances in sample matrices. For analysis of CPs in biomedical samples, an extraction method using acetone is employed in many cases. For analysis of matrices containing a large quantity of fat, dichloromethane is used for extraction. In such a case, fat removal may be necessary, depending on the CP detection methods. After the extract is dried under reduced pressure, CPs are recovered into the MeCN layer by distribution using *n*-hexane and MeCN, and the MeCN layer dried again under reduced pressure for use as measurement samples.

Table 2 Clean-up methods for organophosphates from biomedical samples

Analyte	Sample	Species	Clean-up procedure	Determination	Ref.
Diazinon, Dimethoate, Malathion, Parathion, EPN, Phosalone	Milk	Cattle	LLE (MeCN, Dichloromethane, *n*-hexane/acetone)	GC/FPD	5
cis-Mevinfos, Methyl-parathion, Paraoxon	Milk	Cow	LLE (MeCN, Dichloromethane, *n*-hexane/acetone)	GC/NPD	6
Acephate, Azinphos, Carbophenothion, Chlorfenvinphos, Chlorpyriphos, Coumaphos, Crufomate, Crotoxyphos, DDVP, Demeton, Diazinon, Dicrotophos, Dimethoate, Dioxathion, Disulfoton, EPN, Ethion, Ethoprop, Fenamiphos, Fensulfothion, Fenthion, Fonofos, Isofenphos, Malathion, Merphos, Methamidophos, Methidation, Methyl parathion, Mevinphos E, Mevinphos Z, Monocrotophos, Naled, Parathion, Phorate, Phosalone, Phosmet, Phosphamidone, Profenphos, Propetamphos, Ronnel, Terbufos, Tetrachlorvinphos, Triazophos	Liver	Bovine	Homogenize and LLE (5%-EtOH in ethyl acetate) Clean-up by GPC (S-X3 Biobeads) and SPE (Silica gel)	GC/ECD	9
Parathion, Paraoxone, Malathion	Plasma Brain Heart Muscle Liver Kidney	Fish	Deproteinization and LLE (Isooctane, MeOH, NaCl)	GC/ECD	12
Malathion, Fenitrothion	Blood	Human	Head-space–solid-phase microextraction	GC/MS	18
Glyphosate	Serum	Human	Deproteinization (10%-TCA) Wash (Diethyl ether) Clean-up with anion-exchange column (AG-1X8)	HPLC/UV	21
Chlorpyriphos, Chlorpyriphos-methyl, DDVP, Diazinon, Dimethoate, Dimethylvinphos, Disulfoton, Edifenphos, EPN, Ethion, Fenitrothion, Fenthion, IBP, Isoxathion, Malathion, Methidathion, Parathion, Parathion-methyl, Phenthoate, Propaphos, Pyridaphenthion	Serum	Human	SPE (Extrelut No.3)	HPLC/MS	27
Chlorpyriphos, Chlorpyriphos-methyl, DDVP, Diazinon, Dimethoate, Dimethylvinphos, Disulfoton, Edifenphos, EPN, Ethion, Fenitrothion, Fenthion, IBP, Isoxathion, Malathion, Methidathion, Parathion, Parathion-methyl, Phenthoate, Propaphos, Pyridaphenthion	Serum	Human	SPE (Bond Elut C18)	HPLC/MS	29
Glyphosate, Metabolite	Serum	Human	Deproteinization (10%-TCA) Wash (Ethyl acetate, diethyl ether)	CE/UV	30

LLE, liquid–liquid extraction; FPD, flame photometric detection; GC/FPD, gas chromatography/flame photometric detection; NPD, nitrogen phosphate detection; GC/NPD, gas chromatography/nitrogen phosphate detection; UV, ultraviolet; HPLC/UV, high-performance liquid chromatography/ultraviolet; CE, capillary electrophoresis; CE/UV, capillary electrophoresis/ultraviolet; GPC, gel permeation chromatography; EPN, *O*-ethyl *O*-*p*-nitrophenyl phenylphosphonothioate; DDVP, dichlonos, 2,2-dichlorovinyl dimethyl phosphate; IBP, 1 profenos, *O*-*o*-diisopropyl *S*-benzyl phosphorothioate.

Table 3 Clean-up methods for carbamates from biomedical samples

Analyte	Sample	Species	Clean-up procedure	Determination	Ref.
Aldicarb, Aldicarb sulfone, Bendiocarb, NAC, Carbofuran, 3-Hydroxycarbofuran, Methiocarb, Methomyl, Mexacarbate, Oxamyl, Propoxur	Liver	Bovine	Homogenize LLE (5%-EtOH in ethyl acetate) Clean-up by GPC (S-X3 Biobeads) and SPE (Silica gel)	GC/ECD	9
1-Naphthol (metabolite of NAC)	Liver		Deproteinization and SPE	GC/MS	16
Bendiocarb, Bufencarb, NAC, Carbofuran, 3-Hydroxycarbofuran, Dioxacarb, Isoprocarb, Methicarb, Promecarb, Propoxur	Serum	Human	Homogenize SPE(Sep-Pak C_{18})	GC/MS	19
NAC, Ethiofencarb, Fenobucarb, Isoprocarb, Metolcarb, Propoxur, XMC, Xylylcarb	Serum	Human	SPE (Extrelut No.3)	HPLC/MS	28
NAC, Ethiofencarb, Fenobucarb, Isoprocarb, Metolcarb, Propoxur, XMC, Xylylcarb	Serum Plasma Urine	Human	SPE (Bond Elut C_{18})	HPLC/MS	29

XMC, 3,5-xylyl methyl carbamate.

Table 4 Clean-up method for organochlorinates from biomedical samples

Analyte	Sample	Species	Clean-up procedure	Determination	Ref.
DDE, PCB	Blood	Human	SPE (Florisil, charcoal)	GC/ECD	7
α-BHC, β-BHC, Lindane, Aldrin, Heptachlor, Heptachlor-epoxide, *p*,*p*′-DDE, *p*,*p*′-DDD, *p*,*p*′-DDT	Plasma	Human	LLE (Hexane, dichloromethane) SPE (Silica gel)	GC/ECD	8
Aldrin, β-BHC, γ-Chlordane, *p*,*p*′-DDD, *p*,*p*′-DDE, *p*,*p*′-DDT, Dicofol, Dieldrin, Endosulfan I, Endosulfan II, Endrin, HCB, Heptachlor, Heptachlor epoxide, Lindane, *p*,*p*′-Methoxychlor, Mirex	Liver	Bovine	Homogenize LLE (5%-EtOH in ethyl acetate) Clean-up by GPC (S-X3 Biobeads) and SPE (Silica gel)	GC/ECD	9
PCDDs, PCDFs, PCB	Serum	Human	SPE (Silica gel, activated carbon "AX-21")	GC/HRMS GC/ECD	13
DDT, BHC	Serum	Human	SPE (Silica gel)	GC/ECD	14
β-BHC, Oxychlordane, *trans*-Nonachlor, HCB, *p*,*p*′-DDE, *p*,*p*′-DDT, PCB	Breast adipose tissue	Human	LLE (*n*-Hexane) Clean-up with SPE (Florisil)	GC/ECD	15
β-BHC, Oxychlordane, *trans*-Nonachlor, HCB, *p*,*p*′-DDE, *p*,*p*′-DDT, PCB	Serum	Human	LLE (Ethyl ether/*n*-hexane = 1/1) Clean-up with SPE (Florisil)	GC/ECD	15
Toxaphene	Serum		LLE (Hexane/dichlormethane - 9/1) Fractionation with SPE (Silica gel, Florisil, Carbopack)	GC/MS	17
PCB, PCDDs, PCDFs	Plasma	Human	Deproteinization (Ammonium sulfate, EtOH) LLE (*n*-Hexane) Clean-up (Florisil)	GC/MS	20

HRMS, high-resolution mass spectrometry; GC/HRMS, gas chromatography/high-resolution mass spectrometry; DDE, Dichlorodiphenyl-dichloroethylene, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene; PCB, polychlorinated biphenyls; BHC, Benzene hexachloride, 1,2,3,4,5,6-hexachlorocyclohexane; DDD, Dichlorodiphenyldichloroethane, 1,1-dichloro-2,2-bis(4-chlorophenyl)ethane; DDT, 2,2-bis(*p*-chlorophenyl)-1,1,1-trichloroethane; HCB, hexachlorobenzene; PCDF, polychlorinated dibenzofuran; PCDD, polychlorinated dibenzo-*p*-dioxin.

3.2.2 Solid-phase Extraction

Table 3 summarizes recent reports on sample preparation in the analysis of CPs in biomedical samples.

Using an ODS cartridge, Bakowski et al. eluted 10 types of CPs in liver with 70% MeOH (in water), and analyzed them by GC/MS.[16] Kawasaki et al. eluted eight types of CPs in serum with dichloromethane, using a diatomaceous earth cartridge, and analyzed them by HPLC/MS.[28]

Holstage et al.[9] and Kawasaki et al.[29] have reported on the simultaneous extraction and analysis of OP and CPs.

3.3 Organochlorinates

Organochlorinates are used extensively as insecticides, sterilizers, and herbicides. In particular, those used as insecticides are extremely toxic to living bodies, so the highest priority should be given to the clinical analysis of these organochlorine insecticides. The use of most organochlorinates in insecticides is currently prohibited or limited. However, they remain in animal foods such as meat and fish, and accumulate in living bodies when such foods are consumed.

Low-polarity organic solvents such as *n*-hexane and chloroform are used to extract organochlorinates from biomedical samples. In such a case, pigments and fats in samples are extracted simultaneously, so a clean-up method using a florisil column may also be employed (Table 4).

Though PCB are not pesticides, they are extracted by the above method. They can be measured at the same time as organochlorinates, by gas chromatography/electron capture detection (GC/ECD) or GC/MS.

3.4 Herbicides

Alkyldipyrimidium salts, organochlorinates, triazines, phenolate, phenoxylate, and anilinate compounds are used as herbicides. Among them, compounds other than alkyldipyrimidium salts and triazines have a relatively weak toxicity to the human body.

Alkyldipyrimidium salts (paraquat, diquat, morfamquat), particularly, are highly toxic compounds. In particular, the number of acute intoxication patients caused by paraquat and/or diquat has not decreased. In these intoxications, the blood concentration must be quickly lowered in order to save the patient's life. Accordingly, it is clinically important to measure the concentration of the two kinds of pesticides in the blood. HPLC is frequently used to measure paraquat and diquat in the blood or urine. Supernatant obtained following deproteinization with PCA (approximately 5%) can be used for test samples. In some cases, such protein-free supernatant is cleaned by ion exchange or through the use of a reversed-phase column ODS.

Table 5 Clean-up method for herbicides from biomedical samples

Analyte	Sample	Species	Clean-up procedure	Determination	Ref.
Paraquat	Blood	Human	Remove various metal ions (5%-EDTA)	Spectrophotometry	2
	Urine		Deproteinization and extraction (1%-TCA)		
Benfluralin, Ethalfliralin, Isopropalin, Nitralin, Pendimethalin, Profluralin, Trifluralin	Urine	Human	SPE (Sep-Pak C18)	GC/ECD	10
Terbacil, Bromacil, Norflurrazon, PAC	Whole blood Plasma Urine	Human	SPE (Bond Elut C18)	GC/FID	11
Paraquat, Diquat	Tissues	Human	Homogenize and deproteinization (3%-PCA)	HPLC/UV	22
	Whole blood				
Diquat	Blood Bile	Rat	Deproteinization (70%-MeOH containing 21 mM PCA)	HPLC/UV	24
	Urine				
Diquat	Liver	Rat	Homogenize and deproteinization (70%-MeOH containing 21 mM PCA)	HPLC/UV	24
	Kidney				
Simazine, Atrazine, Propazine	Tissue	Catfish	Extract with ethyl acetate Solvent partitioning (MeCN, Petroleum ether) Clean-up by SPE (C18 cartridge)	HPLC/UV	25
Triallate	Lipid-rich tissue		Not necessary	MS/MS	32

FID, flame ionization detection; GC/FID, gas chromatography/flame ionization detection; EDTA, ethylenediaminetetraacetic acid; PAC, chloridazon, 5-amino-4-chloro-2 phenylpyridazin-3(2H)-one; TCA, trichloroacetic acid.

Table 6 Clean-up method for ivermectin from biomedical samples

Analyte	Sample	Species	Clean-up procedure	Determination	Ref.
Ivermectin	Serum	Cattle	LLE (Methyl *tert*-butyl ether)	TLC	4
Ivermectin	Plasma	Human	LLE (Acetone, isooctane)	HPLC/FD	23
Ivermectin	Serum	Sheep	SPE (Immunoaffinity column)	HPLC/UV	25

FD, fluorimetric detection; HPLC/FD, high-performance liquid chromatography/fluorimetric detection.

Triazines also show comparatively high toxicity for mammals. Dichlormethane is commonly used as the extraction agent from the biomedical samples. However, solvent partitioning and/or additional clean-up with a C_8 or C_{18} column are sometimes needed.[25,31]

Table 5 summarizes recent reports of herbicides analysis in biomedical samples.

3.5 Other Pesticides

Ivermectin, which has a broad spectrum, has been used as an antibiotic in recent years, and its effect on the human body has been extensively investigated.

LLE and SPE extraction methods are utilized to determine ivermectin concentration in human, bovine, and sheep serum[4,23,26] (Table 6).

4 SAMPLE PRESERVATION

Pesticides in biomedical samples should be measured as quickly as possible after sampling. Since most pesticides are unstable in highly polar solvents and acidic or alkaline solutions, their decomposition during sample preservation may result in erroneous measurements and incorrect interpretations. However, if samples must be preserved, they should be preserved in completely sealed glass containers at as low a temperature as possible. Antiseptics may interfere with measurement, so their use should be avoided.

If the compounds to be detected are known, make sure you fully understand their characteristics and select the optimum preservation method for them.

4.1 Blood (Plasma and Serum)

Blood is the biomedical sample most frequently used, and its value for clinical utilization is very high. As previously mentioned, many types of pesticide decompose naturally in solution, so the blood must be very carefully preserved.

Figure 1 illustrates a typical example of natural decomposition of the pesticide in serum. In this experiment, NAC was added to drug-free serum, and stored at 4 °C and −20 °C. The serum concentrations of NAC and its metabolite (1-naphthol) were measured in a time series. Even with preservation at −20 °C, NAC was decomposed.

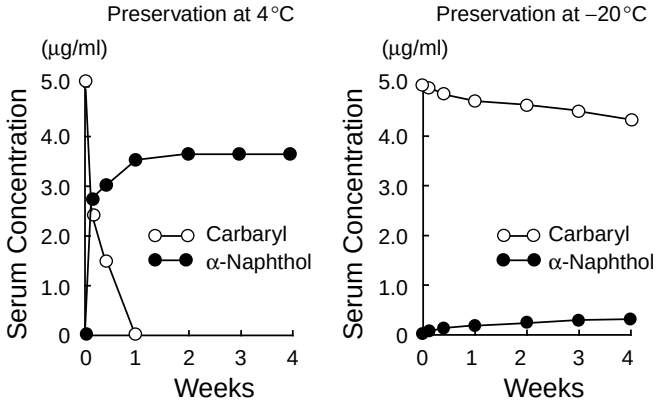

Figure 1 Natural decomposition of pesticides during sample preservation. NAC was added to drug-free serum to produce a final concentration of 10 μg mL^{-1} and the sample was preserved at 4 °C and −20 °C. 400 μL of acetonitrile was added to 200 μL of serum for deproteinization, and NAC and 1-naphthol concentrations of supernatant was measured by HPLC/UV (220 nm).

4.2 Urine

Principally, the preservation of urine samples for pesticide analysis should be the same as that of blood. After sampling the entire amount of urine confirm that its pH is neutral and store it at a temperature of −20 °C or below. As antiseptics may make measurement difficult, their use should be avoided.

4.3 Other Biomedical Fluids

The contents or abluents of the digestive tract can be used as biomedical samples for pesticide analysis, in addition to blood and urine. Whole capsules or tablets are occasionally found in the digestive tract contents. Such residues should be measured individually. Accordingly, after centrifuging the sample, the residue and supernatant should be preserved in separate containers.

4.4 Tissues

The liver is the main detoxifying organ in living bodies, and many drug toxins accumulate there. For this reason,

the liver can be the optimum organ for sampling. Since it is difficult to obtain a liver sample from a living body, the liver is often used for pesticide analysis in dead bodies. After examining marked lesions such as necrosis and color change in the liver, collect at least 100 g and freeze-preserve it at −20 °C or below.

5 CONCLUSION

"Pesticide" is a generic term for several hundred types of compounds. In pesticide analysis, there is currently no reference or recommended method for sampling, sample preservation, or pesticide detection. If it is clear which pesticides are to be detected, methods appropriate for individual compounds should be selected. In the analysis of pesticides in biomedical samples, however, it is often unclear which compounds are to be detected. Accordingly, primarily for such cases, this article describes sampling, preparation of test solution, and preservation of samples.

If humankind is to coexist successfully with pesticides, demand for pesticide analysis in biomedical samples is expected to increase in the future. Standardization in this field and the establishment of information networks are urgently needed.

ABBREVIATIONS AND ACRONYMS

Ac-ChE	Acetyl-cholinesterase
BHC	Benzene Hexachloride, 1,2,3,4,5,6-hexachlorocyclohexane
CE	Capillary Electrophoresis
CE/UV	Capillary Electrophoresis/Ultraviolet
CP	Carbamate Pesticide
DDD	Dichlorodiphenyldichloroethane, 1,1-dichloro-2,2-bis(4-chlorophenyl)ethane
DDE	Dichlorodiphenyldichloroethylene, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene
DDT	2,2-bis(*p*-chlorophenyl)-1,1,1-trichloroethane
DDVP	Dichlonos, 2,2-dichlorovinyl dimethyl phosphate
EDTA	Ethylenediaminetetraacetic Acid
EPN	*O*-ethyl *O*-*p*-nitrophenyl phenylphosphonothioate
FD	Fluorimetric Detection
FID	Flame Ionization Detection
FPD	Flame Photometric Detection
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/FID	Gas Chromatography/Flame Ionization Detection
GC/FPD	Gas Chromatography/Flame Photometric Detection
GC/HRMS	Gas Chromatography/High-resolution Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
GC/NPD	Gas Chromatography/Nitrogen Phosphate Detection
GPC	Gel Permeation Chromatography
HCB	Hexachlorobenzene
HPLC	High-performance Liquid Chromatography
HPLC/FD	High-performance Liquid Chromatography/Fluorimetric Detection
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
HPLC/UV	High-performance Liquid Chromatography/Ultraviolet
HRMS	High-resolution Mass Spectrometry
IBP	1 profenos, *O*-*O*-diisopropyl *S*-benzyl phosphorothioate
LLE	Liquid–Liquid Extraction
NAC	Carbaryl
NPD	Nitrogen Phosphate Detection
ODS	Octadecasil
OP	Organophosphorus Pesticides
PAC	Chloridazon, 5-amino-4-chloro-2 phenylpyridazin-3(2H)-one
PAM	Paralidoxime-indole
PCA	Perchloric Acid
PCB	Polychlorinated biphenyls
PCDD	Polychlorinated dibenzo-*p*-dioxin
PCDF	Polychlorinated dibenzofuran
Ps-ChE	Pseudo-cholinesterase
SPE	Solid-phase Extraction
TCA	Trichloroacetic acid
TLC	Thin-layer Chromatography
UV	Ultraviolet
XMC	3,5-xylyl methyl carbamate

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Sampling Considerations for Biomonitoring

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

REFERENCES

1. S.K. Handa, 'Spectrophotometric Method for the Microdetermination of Bendiocarb Standard Residues in Water', *J. Assoc. Off. Anal. Chem.*, **71**(1), 51–52 (1988).
2. R. Kesari, M. Rai, V.K. Gupta, 'Spectrophotometric Method for Determination of Paraquat in Food and Biological Samples', *J. AOAC Int.*, **80**(2), 388–391 (1997).
3. P. Corti, E. Dreassi, N. Polite, C. Aprea, 'Comparison of an HPTLC and an HPLC procedure for the Determination of Chlorpropham, Propham and Thiabendazole Residues in Potatoes', *Food Addit. Contam.*, **8**(5), 607–616 (1991).
4. W.G. Taylor, T.J. Danielson R.L. Orcutt, 'Thin-layer Chromatographic Detection of Ivermectin in Cattle Serum', *J. Chromatogr. B*, **661**(2), 327–333 (1994).
5. M. Toyoda, K. Adachi, T. Ida, K. Noda, N. Minagawa, 'Simple Analytical Method for Organophosphorus Pesticide Residues in Milk', *J. Assoc. Off. Anal. Chem.*, **73**(5), 770–772 (1990).
6. C.M. Lino, M.I. Noronha da Silveira, 'Organophosphorus Pesticide Residues in Cow's Milk: Levels of *cis*-Mevinfos, Methyl-parathion, and Paraoxon', *Bull. Environ. Contam. Toxicol.*, **49**, 211–216 (1992).
7. R. Frank, H.E. Braun, B. Thorpe, 'Comparison of DDE and PCB Residues in the General Diet and in Human Blood – Ontario 1986–87', *Bull. Environ. Contam. Toxicol.*, **51**, 146–152 (1993).
8. R. Rugama, S. Calero, I. Fomsgaard, M.L. Lacayo, V. Martinez, J. Pitty, 'Levels of Organochlorine Residues in Blood Plasma from Three Populations in Nicaragua', *Bull. Environ. Contam. Toxicol.*, **51**, 153–159 (1993).
9. D.M. Holstage, D.L. Scharberg, E.R. Tor, L.C. Hart, F.D. Galey, 'A Rapid Multiscreen for Organophosphorus, Organochlorine, and *N*-Methyl Carbamate Insecticides in Plant and Animal Tissues', *J. AOAC Int.*, **77**(5), 1263–1274 (1994).
10. T. Kumazawa, K. Sato, H. Seno, A. Ishii, O. Suzuki, 'Capillary Gas Chromatography with Four Different Detectors for Dinitroaniline Herbicides in Human Body Fluids', *J. Anal. Toxicol.*, **19**, 95–98 (1995).
11. X.-P. Lee, T. Kumazawa, K. Sato, 'Rapid Extraction and Capillary Gas Chromatography for Diazine Herbicides in Human Body Fluids', *Forensic Sci. Int.*, **72**, 199–207 (1995).
12. R. Abbas, W.L. Hayton, 'Gas Chromatographic Determination of Parathion and Paraoxon in Fish Plasma and Tissues', *J. Anal. Toxicol.*, **20**, 151–154 (1996).
13. H.R. Johansen, J. Alexander, O.J. Rossland, S. Planting, M. Lovik, P.I. Gaarder, W. Gdynia, K.S. Bjerve, G. Becher, 'PCDDs, PCDFs, and PCBs in Human Blood in Relation to Consumption of Crabs from a Contaminated Fjord Area in Norway', *Environ. Health Perspect.*, **104**(7), 756–764 (1996).
14. E.V. Minelli, M.L. Ribeiro, 'DDT and HCH Residues in the Blood Serum of Malaria Control Sprayers', *Bull. Environ. Contam. Toxicol.*, **57**, 691–696 (1996).
15. S.L. Archibeque-Engel, J.D. Tessari, D.T. Winn, T.J. Keefe, T.M. Nett, T. Zheng, 'Comparison of Organochlorine Pesticide and Polychlorinated Biphenyl Residues in Human Breast Adipose Tissue and Serum', *J. Toxicol. Environ. Health*, **52**, 285–293 (1997).
16. R.S. Bakowski, M.S. Ali, J.D. White, E.T. Phillippo, R.L. Ellis, 'Gas Chromatographic/Mass Spectrometric Confirmation of Ten *N*-Methylcarbamate Insecticides in Liver', *J. AOAC Int.*, **77**(6), 1568–1574 (1994).
17. U.S. Gill, H.M. Schwartz, B. Wheatly, H. Parlar, 'Congener Specific Analysis of Toxaphene in Serum using ECNI-MS', *Chemosphere*, **33**(6), 1021–1025 (1996).
18. A. Namera, M. Yashiki, N. Nagasawa, Y. Iwasaki, T. Kojima, 'Rapid Analysis of Malathion in Blood using Head Space-Solid Phase Microextraction and Selected Ion Monitoring', *Forensic Sci. Int.*, **88**, 125–131 (1997).
19. D.B. Shealy, J.R. Barr, D.L. Ashley, D.G. Patterson, Jr, D.E. Camann, A.E. Bond, 'Correlation of Environmental Carbaryl Measurements with Serum and Urinary 1-Naphthol Measurements in Farmer Applicator and His Family', *Environ. Health Perspect.*, **105**(5), 510–513 (1997).
20. P. Ayotte, E. Dewailly, J.J. Ryan, S. Bruneau, G. Lebel. 'PCBs and Dioxin-like Compounds in Plasma of Adult Inuit Living in Nunavik (Arctic Quebec)', *Chemosphere*, **34**(5–7), 1459–1468 (1997).
21, M. Tomita, T. Okuyama, S. Watanabe, B. Uno, S. Kawai, 'High-performance Liquid Chromatographic Determination of Glyphosate and (Aminomethyl)Phosphonic Acid in Human Serum after Conversion into *p*-Toluenesulfonyl Derivatives', *J. Chromatogr.*, **566**(1), 239–243 (1991).
22. S. Ito, T. Nagata, K. Kudo, K. Kimura, T. Imamura, 'Simultaneous Determination of Paraquat and Diquat in Human Tissues by High-performance Liquid Chromatography', *J. Chromatogr.*, **617**(1), 119–123 (1993).
23. D.R. Krishna, U. Klotz, 'Determination of Ivermectin in Human Plasma by High-performance Liquid Chromatography', *Arzneim.-Forsch./Drug Res.*, **43**(5), 609–611 (1993).

24. C. Madhu, Z. Gregus, C.D. Klassen, 'Simple Method for Analysis of Diquat in Biological Fluids and Tissues by High-performance Liquid Chromatography', *J. Chromatogr. B*, **674**(1), 193–196 (1995).
25. D.C. Holland, R.K. Munns, J.E. Poybal, J.A. Hurlbut, 'Liquid Chromatographic Determination of Simazine, Atrazine and Propazine in Catfish', *J. AOAC Int.*, **78**(4), 1067–1071 (1995).
26. J. Li, S. Zhang, 'Immunoaffinity Column Cleanup and Liquid Chromatographic Method for Determining Ivermectin in Sheep Serum', *J. AOAC Int.*, **79**(6), 1300–1302 (1996).
27. S. Kawasaki, H. Ueda, H. Itoh, J. Tadano, 'Screening of Organophosphorus Pesticides using Liquid Chromatography – Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr.*, **595**(1), 193–202 (1992).
28. S. Kawasaki, F. Nagumo, H. Ueda, Y. Tajima, M. Sano, J. Tadano, 'Simple, Rapid and Simultaneous Measurement of Eight Different Types of Carbamate Pesticides in Serum using Liquid Chromatography – Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. B*, **620**(1), 61–71 (1993).
29. H. Itoh, S. Kawasaki, J. Tadano, 'Application of Liquid Chromatography – Atmospheric Pressure Chemical Ionization Mass Spectrometry to Pesticide Analysis', *J. Chromatogr. A*, **754**(1 + 2), 61–76 (1996).
30. M. Tomita, T. Okuyama, Y. Nigo, B. Uno, S. Kawai, 'Determination of Glyphosate and Its Metabolite, (Aminomethyl)Phosphonic Acid, in Serum using Capillary Electrophoresis', *J. Chromatogr.*, **571**(1 + 2), 324–330 (1991).
31. V. Pacakova, K. Stulik, J. Jiskra, 'High-performance separations in the determination of triazine herbicides and their residues', *J. Chromatogr. A*, **754**(1 + 2), 17–31 (1996).
32. J.V. Headley, K.M. Peru, M.T. Arts, 'Tandem Mass Spectrometry of Herbicide Residues in Lipid-rich Tissue', *Anal. Chem.*, **67**(23), 4349–4353 (1995).

Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of

Maarten Honing
NV Organon, Oss, The Netherlands

1 Introduction **1**
2 Sample Preparation **2**
2.1 Aqueous Samples 3
2.2 Biological and Soil Samples 4
3 Column Chromatography **5**
3.1 Gas Chromatography 5
3.2 Column Liquid Chromatography 7
3.3 Capillary Electrophoresis 7
4 Liquid Chromatography Detection Methods **8**
4.1 Ultraviolet Detection 9
4.2 Electrochemical Detection 9
4.3 Fluorescence Detection 9
4.4 Mass Spectrometric Detection 10
5 Conclusions **14**
Abbreviations and Acronyms **14**
Related Article **15**
References **15**

Large quantities of modern pesticides such as the N-methyl- and carbamoyloxime carbamates (over 1000 t of the N-methylcarbamate carbaryl) are used for agricultural as well as for nonagricultural purposes. The European Union (EU)-permitted concentration levels of these insecticides and their related transformation products in drinking water is 0.1 μg L^{-1}, and the maximum for the total pesticide content has been set to 0.5 μg L^{-1}. Many analytical procedures, focusing on sample pretreatment, column chromatographic separation, and sensitive detection, have been reported for the trace-level determination of these compounds. This article surveys the application of these techniques, and their potential with regard to recovery, resolution, sensitivity, and robustness is discussed.

For example the reproducibility of results with a variety of carbamates from water samples or pretreated biological/soil samples is an essential criteria during the development of extraction procedures. The percentage of recovery influences the sensitivity of the total analytical method. Good results with solid-phase extraction (SPE) are obtained with a C_{18}/OH phase which is used on a routine basis by various analytical laboratories for the analysis of surface waters and vegetables. This technique is superior to liquid–liquid extraction (LLE) for reasons explained in this article. Chromatographic resolution is most effective, when gas chromatography (GC) is used. Unfortunately, thermal degradation of some carbamates in the injector takes place, making liquid chromatography (LC) the separation technique of choice despite the fact that some carbamates cannot now be separated. Capillary electrophoresis (CE) offers better separation efficiency but is still not a robust technique. Another way to solve the coelution problem is the application of mass spectrometry (MS). Utilizing the selected ion monitoring mode, coeluting compounds can be quantified. MS adds an extra separation step because ionized compounds are separated by the mass-to-charge ratio. Application of the ionspray (ISP) and the heated nebulizer (HN) interfacing techniques makes possible the quantification of carbamates at trace levels as low as 0.1 ng L^{-1} in drinking water, without any sample preconcentration. Furthermore, MS also offers the ability to confirm the identity of the compounds. On-line SPE in combination with liquid chromatography/mass spectrometry (LC/MS) or, better, liquid chromatography/tandem mass spectrometry (LC/MS/MS) is the best method for the quantification and structural confirmation of carbamate pesticides at concentration levels of 0.1 ng L^{-1} or lower.

1 INTRODUCTION

The use of nonpersistent carbamate and carbamoyloxime pesticides (which have replaced organochlorine and organophosphorus compounds) is common practice, mainly because of their broad spectrum of activity, high pesticide effectiveness, and generally low mammalian toxicity. A large variety of carbamates is currently in use. Their application as an insecticide, fungicide, or herbicide is related to the molecular structure. The insecticides have the *N*-substituted carbamate moiety and, generally, an aromatic ester or an oxime function. The fungicides have either a benzimidazolyl ester, a pyrimidyl ester or a (bis)dithiocarbamate group. The herbicides are *N*-alkyl or *N*-arylthiocarbamates.[1–3] Figure 1 shows a general representation of these subclasses. The degradation products are generally oxidized derivatives (such as aldicarb sulfone from aldicarb), or alcohols formed by saponification of the carbamic acid group (such as 1α-naphthol from carbaryl). The general degradation pathways of carbamate pesticides have been discussed by Schlagbauer and Schlagbauer,[4] and several reports dealt with the fate of specific carbamates: aryl *N*-methylcarbamates,[5,6] benomyl,[7] and oxime *N*-methylcarbamates.[8–10] The (bio)degradation of some carbamates in various matrices

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

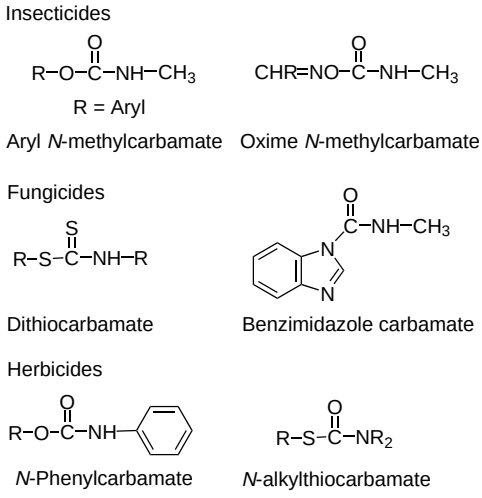

Figure 1 General structures of some carbamate classes.

has also been studied.[11,12] The effects of photolysis have also been investigated.[13]

Many chromatographic separation techniques have been developed for the determination of carbamate pesticides and in some cases their degradation products. Other methods in use are bioassays,[14,15] thin-layer chromatography,[16] and calorimetry.[17] More recently, biosensor-based[18,19] techniques have been developed and applied to the determination of various types of pesticides. In most cases, these techniques are used for the quantitative analyses of a limited amount of compounds, typically one target compound in a foodstuff, for example. CE, a chromatographic technique using a potential difference between the two ends of the capillary, has been applied to the analysis of polar pesticides.[20,21] However, the applicable sample volume is restricted and the robustness of the technique has still to evolve. This article focuses mainly on GC- and LC-based analytical methods for the determination and identification of carbamates and their transformation products, in various environmental matrices.

The large amounts of carbamate pesticides used since the late 1970s requires extensive monitoring in all types of environmental samples. In order to do this, validated analytical procedures need to be developed that are able to determine concentration levels down to $0.1\,\mu g\,L^{-1}$ for drinking water.[22] Ideally, such procedures should he able to assess the carbamates and their transformation products, so as to reduce analysis time and expense.

Although not exhaustive the references given here provide a good representation of the items of interest in carbamate determination. The majority of the target and multiresidue analytical procedures reported are based on GC and LC techniques.[22–24] This article is divided into three sections: (1) sample preparation from water, plant materials, and soil; (2) GC, LC and CE separations; and (3) ultraviolet (UV), fluorescence, electrochemical detection (ECD) and MS detection in LC. The discussions lead to the conclusion that LC/MS with proper clean-up and preconcentration procedures is the method of choice for the multiresidue analysis of carbamate and carbamoyloxime carbamate insecticides.[25]

2 SAMPLE PREPARATION

It is of utmost importance to pay care to the sample, because the compounds of interest are susceptible to degradation by hydrolysis at basic or neutral pH,[23] prolonged exposure to light, and metabolism. For example, adjustment of the pH of an aqueous sample, from pH 9.5 to 6.5, keeps the solution stable for 1 week. The best procedure is to acidify and extract the aqueous sample directly after collection. Light-induced degradation may be minimized by storage in the dark, and metabolic degradation may be prevented by performing the extraction directly after sampling. As the increased polarity of the (bio)degradation products will generally add to the complexity of the problem, extraction and clean-up also require careful attention.

Various methods have been applied to the extraction of carbamates and some of their transformation products, from environmental matrices such as surface or groundwater,[26–38] plants,[39–50] and soil.[50–53] The extraction procedures for water samples, biological materials and soil have been reviewed.[53] The Environmental Protection Agency (EPA) procedure for the determination of *N*-methylcarbamates in drinking water[38,54,55] requires neither extraction nor clean-up (500 µL samples are injected directly into the analytical system). Many extraction procedures are available, such as LLE, solid–liquid extraction (SLE), SPE, or supercritical fluid extraction (SFE).[56–60] With LLE the analytes are partitioned between two immiscible liquid phases, generally water and an organic solvent. In SPE the analytes are first adsorbed from a liquid sample onto a solid sorbent and then selectively desorbed, often after washing of the precolumn. In SLE the compounds are partitioned between the solid sample and a liquid phase, generally an organic solvent and/or water. In SFE the analytes are directly extracted into a supercritical fluid. LLE and SPE are commonly used for water samples, whereas SLE (using for instance Soxhlet extractors or sonication) and SFE are mainly applied to foodstuff, soil, and sediment samples. Clean-up is generally required for foodstuff, sediment, or soil extracts, using liquid–liquid partitioning or adsorption chromatography. Extracts of aqueous samples hardly ever require further clean-up.

The efficiency of sample preparation, both for extraction and clean-up, is usually tested by recovery

experiments, in which samples are spiked with known quantities of the analyte. The recovery itself as well as its reproducibility – as a relative standard deviation (RSD) – are the key quality parameters for sample treatment. If the efficiency is high (over 80%) and reproducible (RSD values of a few percent), a correction for analyte losses can safely be carried out. The selection of a sample preparation method is strongly dependent on the recovery and its reproducibility and on the sample type (aqueous, solid, biological), but aspects such as treatment time and consumption of chemicals should also be considered. In the following discussion, sample preparation for aqueous samples, biological materials, and soil samples (mostly solids) are treated separately.

2.1 Aqueous Samples

A variety of parameters influence the extraction efficiency of LLE and SPE, which are commonly used for the extraction of water samples. Although LLE is still applied, it has been replaced by SPE mainly because it can be coupled on-line to subsequence analytical procedures and smaller quantities of organic solvents are used. Both the procedures and the parameters influencing the extraction efficiency are discussed below.

Various solvents may be used in LLE to obtain a satisfactory partitioning of the analytes between water and an organic phase. Dichloromethane is mainly applied for carbamate extraction, although chloroform, toluene, and benzene are also used.[53] The extraction efficiency is influenced by the acidity[24] and ionic strength[28] of the solution, the nature of the sample (such as the presence of particulate material), the volume ratio of the phases, and the number of subsequent extractions. In a typical extraction procedure, one liter of water is acidified and extracted with three portions of the solvent (totaling 100 mL). The extract is subsequently evaporated to dryness and, prior to analysis, redissolved in a small volume of a suitable solvent, there by gaining sensitivity.

The sorbents used for extraction and clean-up with SPE include Amberlite™, Carbopack B™, and bonded silica phases, many of which are commercially available, often prepacked in disposable cartridges. The choice of an appropriate sorbent, washing and elution solvent is of crucial importance for the extraction and, hence, several studies on suitable phases for carbamate extraction have appeared.[29–32]

The differences in the sorption behavior of carbamates on a polar, alkyl-bonded silica phases may give rise to a wide range of breakthrough volumes. Low breakthrough volumes often cause a low recovery for the more polar carbamates, such as aldicarb, methomyl, and oxamyl[28] (see Table 1). As a consequence, the simultaneous determination of the oxime and aryl *N*-methylcarbamates is rather difficult when using a polar phases. Although SPE with a polar phases generally provides a better reproducibility than LLE, the recoveries for the polar carbamates are lower than with LLE.

Table 1 Extraction recoveries (expressed as RSDs) of 2 L water samples, employing LLE (dichloromethane) and SPE (C_{18} packed pre-columns) with samples containing seven oxime and aryl *N*-methylcarbamates and two degradation products (spiking levels 0.1 and 1 ng mL^{-1})[28] and of 10 mL of sample, using SPE C_{18} extraction disks (spiking level 0.1 ng mL^{-1})[33]

Compound	LLE (%)	SPE (%)	SPE disks (%)
Aldicarb[a]	66 (15)	86 (14)	94 (3)
Aldicarb sulfoxide	19 (18)	10 (38)	87 (6)
Aldicarb sulfone	95 (17)	16 (33)	90 (4)
Methomyl[a]	68 (20)	25 (44)	88 (4)
Oxamyl[a]	69 (11)	26 (46)	85 (4)
Carbaryl[b]	88 (12)	90 (12)	80 (5)
Carbofuran[b]	83 (9)	97 (4)	95 (3)
Methiocarb[b]	93 (3)	98 (8)	96 (4)
Propoxur[b]	69 (16)	93 (17)	93 (3)

[a] Oxime *N*-methylcarbamate.
[b] Aryl *N*-methylcarbamate.

A comparative study[31] on SPE with C_{18}-, C_8-, and C_2-bonded silica, using aldicarb, its sulfoxide and sulfone (and 10 mL water samples), has shown that retention on the C_2 phase is higher than that on the C_8 and C_{18} phases; however, the sorption of other organic material on C_2 is irreversible. As a consequence, the C_2 phase can only be used for a few analyses, whereas the C_8 phase did not deteriorate even after 100 analyses. From the three bonded phases, C_8 generally showed the best performance, although the recovery was never over 80%.

A further comparative study on LiChrosorb™ RP-18 and Carbopack B™, using a variety of *N*-methylcarbamates (and 2 L samples), showed that Carbopack B™ gives a recovery of about 90% for some of the oxime and aryl *N*-methylcarbamates (oxamyl 89%, methomyl 93%, carbofuran 98%, and carbaryl 96%; RSDs not reported). The extraction time could be seriously reduced by applying a flow rate of 150–160 mL min^{-1} (65 × 14 mm inner diameter (i.d.) column). However, the use of Carbopack B™ resulted in contaminated blanks, and extensive conditioning of the sorbent was therefore necessary. Besides, Carbopack B™ does not lend itself for the analysis of large numbers of samples.[29]

In another study, cartridges containing a low-carbon LiChrosorb™ RP-18 sorbent (C_{18}/OH) were tested, using a variety of carbamates (and 50 mL samples).[27] All compounds tested gave recoveries of 80–100% (RSDs 1–7% at the 0.1 μg L^{-1} level), except for butocarboxim sulfoxide (76%), ethiofencarb sulfoxide (55%) and thiofanox sulfoxide (42%). The low recovery of the latter three compounds remains unexplained and contrasts with

Table 2 Recoveries of some carbamates, using 100 mL samples spiked at 5 ug L^{-1} (carbaryl 0.5 ug L^{-1})[202]

Compound	CPP-50[a]	PLRP-S[b]	C_{18}[c]
Aldicarb sulfoxide	100		
Oxamyl	100	40	
Methomyl	100	15	
Propoxur	100	130	35
Carbaryl	100	115	50
Carbofuran	100	100	100

[a] CPP-50 is a graphitized carbon.
[b] PLRP-S is a styrene–divinylbenzene copolymer.
[c] Bondesil C_{18}/OH.

the satisfactory recovery when using LiChrosorb™ RP-8 and RP-18.[27] The C_{18}/OH phase has a retention capacity comparable to that of Carbopack B™ but, fortunately, contamination is much less of a problem. By way of illustration, the recoveries for six selected carbamates with three commercially available phases are listed in Table 2.

In subsequent studies, various parameters which influence the preconcentration step, such as the rate of sample loading, the particle size of the sorbent, and the amount of sorbent needed, were investigated. No significant influence of the sampling rate was found; small particles (10 μm) and minimal amounts of sorbent are to be preferred.[30,31]

On-line coupling of SPE with LC, employing precolumns with bonded-silica phases or so-called membrane extraction disks and desorption by the LC eluent, has been reported.[33] If a sufficiently sensitive mode of detection is used, sample volumes can be reduced dramatically,[27] and breakthrough is largely prevented – on-line extraction of 10 mL samples using these disks, and LC combined with on-line hydrolysis, derivatization, and fluorescence detection,[27] gave good results (recoveries of 80–100%, RSDs of 3–6%) for the oxime *N*-methylcarbamates.[32,33]

The selection of either LLE or SPE should primarily be based on performance, i.e. on recovery and reproducibility. Although most official methods are based on LLE procedures, such as from the National Pesticide Survey,[54,55] these procedures generally have several drawbacks related to more practical aspects.

The treatment of (typically) 1 L of sample precludes on-line coupling of LLE to LC and makes the procedure laborious and time-consuming. Furthermore, the formation of emulsions and the necessary solvent evaporation may cause practical problems with regard to evaporation and concentration. Moreover, the considerable volumes of inflammable and sometimes toxic solvents required for LLE cause waste problems. The main advantage of SPE over LLE is the possibility of on-line coupling to chromatographic systems. In addition, less sample is required for SPE and no solvent disposal problems occur. Both LLE and SPE require appropriate choices of materials (solvents and sorbents). Current developments in water analysis will probably lead to the large-scale introduction of on-line SPE as a basis for validated methods.

2.2 Biological and Soil Samples

Extraction and clean-up of plant and soil samples is complicated by the variations in matrix composition (often characterized by parameters such as percentage of organic matter, sand, and clay). In addition, the composition of the solid samples may negatively influence the extraction efficiency – analytes that have been in prolonged contact with a solid matrix are often desorbed less readily than freshly applied analytes. As a consequence, the recovery of freshly spiked analytes can frequently provide an erroneous impression. The above-mentioned preventive measure of extraction directly after sampling is especially applicable to soil and biological matrices.

Extraction procedures for carbamates (and their transformation products) from plant or soil matrices are mainly hampered by the presence of coextractives. Therefore, a conventional SLE procedure, for instance Soxhlet extraction, blending, or sonication, is usually followed by one or more clean-up steps. These mostly involve column chromatography over a material such as Florisil™; gel permeation chromatography (GPC) is used for the clean-up of fruit and vegetable extracts. Alternatively, SFE may be applied; it often provides sufficient selectivity to make further clean-up superfluous. The extraction and clean-up of plant and soil samples is discussed separately.

For the determination of *N*-methylcarbamates, plant material may be treated by SLE using blending, sonication and, less appropriately, Soxhlet extraction. Generally, acetone or mixtures of this solvent with dichloromethane, methanol and/or light petroleum are employed.[53] Although it has been stated that methanol is the most suitable solvent for the extraction of carbamates from grain, because it extracts up to 15% more of the carbamate residues than acetone or acetonitrile, the presence of coextractives such as the chlorophylls and carotenes makes methanol less attractive.

Separation of the analytes from coextractives by column chromatography over Florisil™, silica, or alumina produces good results for most carbamates. Multiple desorption is generally applied using different solvents. When employing the oxide-type phases, clean-up recoveries are generally bad (below 50%) for the more polar compounds, such as aldicarb sulfoxide and sulfone.[47] In contrast, good recoveries (70–95%, RSDs 5–9%) were obtained for the simultaneous determination of aldicarb, its oxide, and sulfone, if these compounds were targeted. The use of an aminopropyl-bonded silica column,[45,46] with dichloromethane as the extraction

solvent showed good recoveries (85–100%, RSDs 2–4%) for all carbamates tested, although the recoveries of the total pretreatment procedure did not exceed 80% for the polar compounds (aldicarb sulfoxide, butocarboxim sulfone, 1α-naphthol, and oxamyl). The advantage of aminopropyl-bonded silica over the oxide-type phases is that small extraction volumes can be used and that the number of desorption steps can be kept low. GPC of methanol extracts from blended plant material was reported to give good results with low-pigment material (apples, potatoes), whereas high-pigment materials (cabbage) required a second separation step. Such a second step was realized by connecting a polypropylene minicolumn, packed with 0.5 g Nuchar-Celite™, on-line with the GPC column. Recoveries for both procedures (GPC only or GPC and Nuchar-Celite™) were comparable. Recoveries of 85–100% were obtained for some 10 *N*-methylcarbamates, whereas the recovery for aldicarb sulfoxide was 50%.[47] Unfortunately, GPC was only used off-line from the analytical column.

High selectivity can be obtained using immunoaffinity precolumns. The application of such a column[48] resulted in the liquid chromatography/ultraviolet spectroscopy (LC/UV) detection (220 nm) of carbofuran from a crude potato extract injected directly on the precolumn. With a C_{18} bonded-silica phase, overloading of the UV signal was observed, and carbofuran could not be detected. The presence of the coextractives retained on the C_{18} also caused the suppression of the ion signal in atmospheric pressure chemical ionization (APCI)/MS. This was eliminated when the immunoaffinity phase was used. Although the development of new immunoaffinity precolumns opens a new area within the field of sample pretreatment, up till now the application range of this type of columns is limited to one target compound or a to group of compounds with the same antigen.

Soil can often be conveniently subjected to Soxhlet extraction, using various solvent mixtures and subsequent clean-up over Florisil™.[50,51] A comparative study showed that acetone–dichloromethane (50 : 50 v/v) gave the best extraction results for oxamyl, carbaryl, carbofuran, propoxur, 1α-naphthol, chloropropham, and pirimicarb, whereas aldicarb and its oxidation products could successfully be extracted (85–90% recovery) in two successive extractions with acetone–water (40 : 60 v/v) and methanol–water (50 : 50 v/v). In contrast to findings for plant material,[47] methanol was shown to extract all carbamates insufficiently from soil samples. So far, no single solvent mixture has been shown to extract carbamates and their transformation products with nearly similar recoveries and reproducibility.

Clean-up of soil extracts is commonly performed by column chromatography. The use of Florisil™ and aminopropyl-bonded silica gives good recoveries (75–100%, no RSDs reported) for all carbamates, including the oxidation products of aldicarb.[50,51] It should be emphasized that the elution of all compounds requires more than one solvent mixture and that, so far, no procedure for the simultaneous clean-up of all carbamates (and their transformation products) has been reported.

A major advantage of SFE over SLE is that clean-up can be omitted if the extraction parameters are chosen to provide maximum selectivity. SFE with carbon dioxide (typical conditions 50 °C, 250 atm, 15 min) has been successfully applied to the extraction of aldicarb and carbaryl from liver samples (recoveries 75–100%, no RSDs reported[56]) and methomyl, methiocarb from fruit and soil samples (recoveries 100%, no RSDs reported[57]). Despite these promising results, no reports have appeared on SFE treatment of soil samples yet. However, the obvious potential of the method justifies expectations that SFE will be used to a wider extent for carbamate extraction from soil and plant materials in the near future.

3 COLUMN CHROMATOGRAPHY

Analytical systems for the determination of carbamates generally involve a column chromatographic step. Initially GC was used, but column LC is now generally preferred. Supercritical fluid chromatography (SFC) was occasionally used as separation technique, but has lost favor during the 1990s. CE has also been used, but the robustness of the technique has still to evolve. In general the separation potential of a particular chromatographic technique is the key issue.

3.1 Gas Chromatography

Many column types, detection techniques, and derivatization reagents have been studied for the determination of carbamates and their degradation products by GC.[61–93] The reasons for this were the main problems with GC analysis of carbamates,[61–88] namely thermal decomposition of the analytes on the column[61] and/or in the injector,[66] and insufficient sensitivity of detection.

Except for the halogen-containing compounds, such as barban, carbamates in general show a low response to electron-capture detection, flame ionization detection (FID) and thermal sulfur detection (TSD). Detectability was initially improved by converting the carbamates into, for instance, the *N*-perfluoroacyl,[69] *N*-perchloroacyl,[72] or pentafluorobenzyl[73] derivatives for electron-capture detection. Later, with the advent of nitrogen/phosphorus detection (NPD) and MS, such derivatization methods became obsolete.

Thermally assisted hydrolysis is a key problem in the GC analysis of carbamates. For example, *N*-methyl- and *N*-phenylcarbamates are labile at higher temperatures (>100 °C), whereas *N*,*N*-dimethylcarbamates, such as pirimicarb, do not suffer thermal degradation under GC conditions. Thermally assisted hydrolysis may be prevented by the use of special GC systems or by appropriate derivatization. Alternatively, derivatization may be required to yield more volatile derivatives for degradation products such as amino hydroxypyrimidine compounds (from pirimicarb), because these compounds are not directly amenable to GC. Derivatization and instrument modifications, as a means of prevention of thermal degradation, are discussed next.

In the 1960s and early 1970s, insufficiently deactivated column material caused thermal degradation with a variety of packed-column stationary phases (Carbowax™ 20M, polar; Apiezon N™, nonpolar; SE 30 and QF-1, mixed polarity[61]). Extensive deactivation of the stationary phase and the column support material and the exchange of a glass injection liner by quartz were shown to suppress degradation; however, decomposition tended to increase upon column aging. With the introduction of high-quality deactivated capillary columns (SE-54[61]), thermal degradation is effectively limited to the injector. Recently, this problem was solved by the application of programmed temperature vaporization (PTV) and cold on-column injection for the GC analysis of some thermolabile aryl *N*-methylcarbamates.[64–66] (Note that the injector inlets and packing still required deactivation.) Figure 2 shows some chromatograms to illustrate the response dependence of selected carbamates upon the injector type. Despite the fact that contamination of the injection port due to poor clean-up of extracts is known to promote the degradation of carbamates,[66] thermal lability remains an intrinsic property of some carbamates.

Prevention of thermally assisted hydrolysis by derivatization is mainly used for aryl *N*-methylcarbamates, because these compounds are easily hydrolyzed (to give phenol-type products) and subsequently converted into thermally stable esters.[61] This kind of derivatization may lead to serious overestimation of pesticide levels, because the natural phenol-type degradation products are also derivatized. An alternative derivatization (used for methomyl, an oxime *N*-methylcarbamate) is hydrolysis to the oxime, subsequent derivatization of which provides a useful conversion into a thermally stable compound. Although derivatization of carbamates is not generally applicable, it may be of use for target compound analysis.

Despite these thermolability problems, some GC/MS methods, using electron ionization (EI) or chemical ionization mass spectrometry (CI/MS), with isobutane as the reagent gas, have been reported for selected compounds.[86–93] Six GC stationary phases were tested

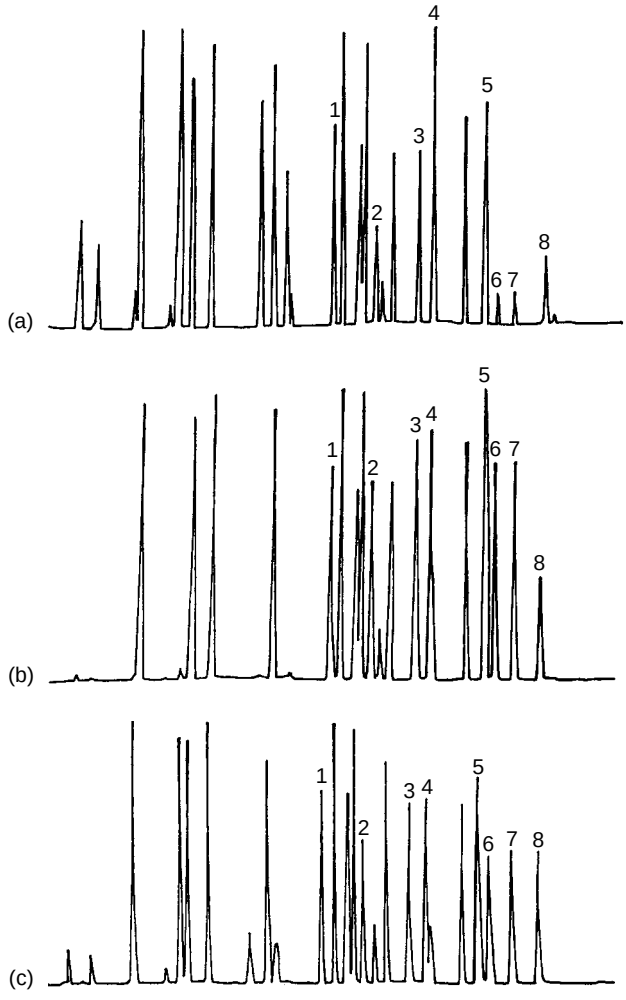

Figure 2 GC/NPD chromatograms of eight aryl *N*-methylcarbamates and one *N*,*N*-dimethylcarbamate (pirimicarb) using (a) hot splitless, (b) cold on-column and (c) PTV injection.[65] Peak assignment: (1) propoxur, (2) bendiocarb, (3) carbofuran, (4) aminocarb, (5) pirimicarb + ethiofencarb, (6) dioxacarb, (7) carbaryl, (8) methiocarb.

in packed columns, in combination with CI/MS detection, for 32 carbamates.[93] Moderate column temperatures (<185 °C) and short open tubular allowed detection of intact carbamates, except for carbaryl (ca. 50% degradation), from spiked soil sample extracts (recoveries 80–112%, RSD 2–9%, 19 carbamates).[92] Packed-column gas chromatography/chemical ionization/mass spectrometry (GC/CI/MS) was used for aldicarb and butocarboxime and their metabolites.[89] Short capillary columns and CI/MS detection were also used to overcome thermolability problems.[92] In both cases, GC with chemical ionization (CI)-triple quadrupole tandem mass spectrometry (MS/MS) detection was used, either to facilitate identification from the CI spectra or to gain selectivity in compensation for loss of chromatographic resolution. No quantitative data are reported in the latter

study (on aldicarb, its sulfoxide and sulfone), but the former study specifies absolute limits of detection (LODs) of 0.15 ng (aldicarb nitrile), 0.3 ng (aldicarb) and 1.2 ng (aldicarb oxime) with a linear response from the LOD up to 150 ng and in full-scan mode.

Although some compounds, particularly aldicarb and its sulfoxide and sulfone, and carbaryl, are definitely not amenable to GC; GC with CI/MS detection (and MS/MS) may lead to unequivocal identification and low LODs for some carbamates. In other words, although GC is not the method of choice for the simultaneous determination of a wide variety of thermal labile carbamates and their degradation products, GC with CI/MS detection, possibly combined with derivatization, is a powerful method for target analysis.

3.2 Column Liquid Chromatography

Since the late 1970s LC has become a reliable technique for the separation of all types of carbamates. A 1975 review on the determination of carbamates[94] identified the potential of LC as opposed to GC–LC in that it generally requires less extensive sample pretreatment and enables on-line trace enrichment and sample clean-up. Most importantly no heat is used, which prevents thermal degradation. A comparative study on the separation of carbamates,[95] using various normal-phase and reversed-phase modes of operation, and a report on pesticide analysis,[96] then gave LC experimental conditions for many carbamates.

Several reports on the target LC analysis of carbamates and their degradation products have been published since.[97–105] The quality as well as the nature of the stationary phase are key parameters in most LC studies and both subjects are discussed below.

The presence of free silanol groups may lead to bad peak shapes, especially for polar carbamates and their transformation products. This has been demonstrated for the simultaneous separation of aldicarb, its oxidation products, and their oximes,[98–104] and for pirimicarb and three of its degradation products.[105] In all cases the use of a buffer in the LC eluent or of a specially end-capped phase[102] was necessary to maintain separation and constant retention times with alkyl-bonded silica phases. Unfortunately, end-capped phases tend to be lost during use[104] and the separation capacity has to be maintained by using buffered eluents; however, buffered eluents can result in irreproducible retention times for polar carbamates.[98] Regular monitoring of the separation quality and the retention times is therefore recommended.

One study[95] on the use of a wide variety of phases (silica, cyano-bonded, amino-bonded, and C_{18}-bonded silica and ether phase-ETH) showed that most compounds could be separated on silica and cyano-bonded silica, whereas amino-bonded silica gave peak broadening, especially for the more polar compounds. In the reversed-phase mode, C_{18}-bonded silica gave the best results;[95] no separation of the polar compounds was achieved on ETH. Twenty-two carbamates could be separated on C_{18}, using gradient elution (20–60% acetonitrile in water). Furthermore, Zorbax™ C_8 and Zorbax™ cyano both give insufficient separation – the former case did not produce baseline separation of propoxur, carbofuran, bendiocarb, carbaryl, and 1α-naphthol. Although the cyano phase did not separate aldicarb sulfone and oxamyl (Figure 3b), it was found to be a good phase for the separation of aldicarb and its metabolites.[103] Incomplete separation for pirimicarb and its seven degradation products on polar Hibar RP-8 and amino columns in the reversed-phase mode has also been reported.[97] The LC available phases have a limited capacity for the separation of mixtures containing all carbamates[105] and their degradation products, with problems being encountered especially for polar compounds.

3.3 Capillary Electrophoresis

Despite the many advantages of LC over GC, the inferior separation potential of LC is still a major drawback. Micellar electrokinetic chromatography (MEKC) is a promising technique for the complete separation of all carbamates. Being a unique mode of CE, the separation of these neutral carbamates is realized on the basis of their different affinities to the micellar pseudostationary phase and consequently shows similar behavior as in reversed-phase LC. The difference is the much higher column efficiency in the MEKC system.

However, some important parameters, such as the surfactant concentration, ionic strength, and buffer modifiers still strongly influence the separation power of this technique. The pH, which is important in conventional CE, is of less importance because the carbamates are neutral at pH 7. Moreover, the tendency of carbamates to hydrolyze at higher pH prevents the use of an alkaline buffer. These parameters were investigated extensively by Wu et al.[20,21] and the resultant method was applied to real-life samples. The optimum sodium dodecyl sulfate (SDS) concentration and the effect of the phosphate buffer (ionic strength) were varied, and the influence of urea and cyclodextrine as additives to the CE buffer were evaluated.

A major advantage in MEKC is the ability to change the capacity factor (k') by changing the SDS concentration. Indeed an increase of the SDS concentration from 20 to 100 mM leads to an increase of k' for all carbamates. The k' increase for aminocarb was so large that, at SDS concentrations of 50 mM, aminocarb started to elute after

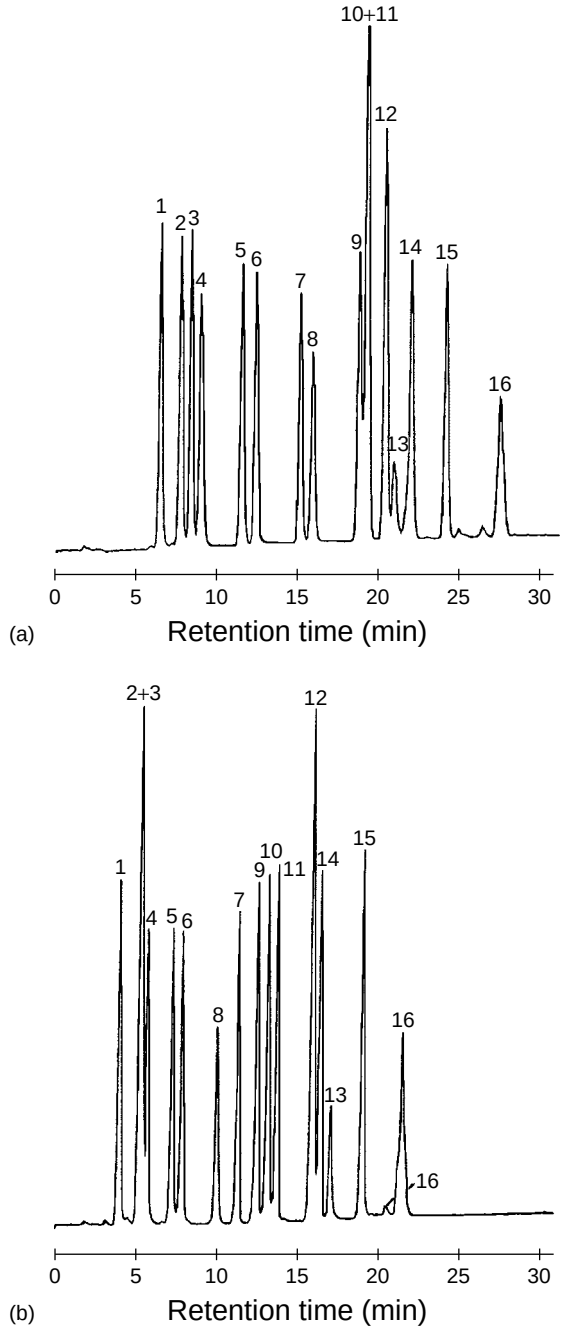

Figure 3 LC chromatogram with fluorescence detection of 10 carbamates and six degradation products, obtained with (a) C_8 and (b) cyano phases, using: 25 cm × 4.6 mm i.d. Zorbax™ column with 6 μm spherical particles; a 30 min linear gradient from 12% to 70% acetonitrile in water; flow rate 1.5 mL min. Peak assignment: (1) aldicarb sulfoxide, (2) aldicarb sulfone, (3) oxamyl, (4) methomyl, (5) 3-hydroxycarbofuran, (6) methiocarb sulfoxide, (7) methiocarb sulfoxide, (8) aldicarb, (9) propoxur, (10) carbofuran, (11) bendiocarb, (12) carbaryl, (13) 1α-naphthol, (14) landrin, (15) methiocarb, (16) bufencarb.

isoprocarb. For this reason an SDS concentration of 60 mM was considered optimal, even though carbofuran and bendiocarb were coeluting, whereas incomplete separation was still obtained for isoprocarb/aminocarb and carbaryl/trimethacarb. Baseline separation for these four was finally obtained after adjusting the phosphate buffer concentration, and consequently the ionic strength, to 40 mM. Unfortunately the electrophoretic current increased from 25 to 55 μA making the control of the capillary temperature necessary. The separation of carbofuran and bendiocarb was finally achieved after addition of 4 M urea. Another consequence was the reversal of the migration orders of the peakpairs, such as pirimicarb–trimethacarb, isoprocarb–carbaryl, promecarb–methiocarb, isoprocarb–pirimicarb, and carbofuran–propoxur, without the loss of the chromatographic resolution. Also, the addition of hydroxypropyl-β-cyclodextrine resulted in the complete separation of the carbamates and moreover led to a significant reduction of the migration time, making the total analysis time approximately 14 min.

After optimization of all experimental parameters, the reproducibilities of the migration times and peak areas for all carbamates were better than 3 and 8%, respectively. The migration-time value was within 1% when the relative migration time with respect to the first-eluting carbamates was used. These data clearly prove the stability of the system. Nevertheless, the sensitivity using UV detection was still low, typically 200 ppb (only for carbaryl a detection limit of 50 ppb). This drawback was finally overcome by the application the enrichment of the sample with SPE and the increase of the injection volume by stacking. Using ODS-6 cartridges and reducing the sample volume from 50 mL to 100 μL, a 500-fold enrichment of the compounds was achieved. In addition, an approximately 5-fold in gain sensitivity was achieved with the positive-polarity stacking mode, bringing the detectability well below 0.1 ppb – this being the requirement for drinking water levels in the EU. Although, these data make this analytical tool a serious candidate for the routine analysis of carbamates, the method was only tested for a test mixture of 15 different carbamates, which is half of the amount used by de Kok et al.[27] Also, no data were reported using environmental samples, such as soil, sediments, groundwater, and other surface waters.

4 LIQUID CHROMATOGRAPHY DETECTION METHODS

Because thermal degradation often occurs in GC, LC is preferred for the separation of thermolabile carbamates in combination with UV,[106–119] ECD,[119–128] fluorescence,[129–147] or MS[148–196] detection. These techniques differ in terms of analyte detectability and selectivity and

are therefore the most important characteristics of the detection methods discussed below.

4.1 Ultraviolet Detection

Monitoring UV absorbance at properly selected wavelengths and/or recording UV absorption spectra, if diode array detection (DAD) is applied, provides a convenient means of detecting most organic compounds. Many carbamates have a UV absorption maximum in the range of 190–225 nm, and additional maxima at wavelengths between 245 and 280 nm.[24,89] It is obvious that multiple-wavelength detection is required to obtain near-maximum sensitivity if a large number of analytes has to be determined. UV detection is best combined with acetonitrile–water mixtures as an eluent, because this allows detection at 195–200 nm (compared with methanol–water at 205–210 nm and tetrahydrofuran–water at >220 nm). Interferences may, of course, be caused by compounds which strongly absorb at 190–220 nm, such as humic acids – in many LC separations a broad hump occurs at low retention times. It has been stated that such interferences may be efficiently trapped by using a cyanopropyl-bonded silica precolumn.[104]

Generally speaking, the detection limits obtainable with UV for carbamates are rather average, typically in the lower nanogram range.[111] In systems involving on-line trace enrichment this may still be sufficient, such as for surface water analysis with DAD, and will provide some structural information. One should also be aware of problems encountered with complex samples (i.e. insufficient clean-up) which may explain reported differences in carbamate LODs by an order of magnitude when comparing water and soil samples.[110] In conclusion, for most sample types of LC, UV detection is suitable for the trace-level determination of carbamates only if it is combined with concentration and/or elaborate clean-up procedures.

4.2 Electrochemical Detection

Despite the inherent sensitivity and selectivity of ECD it has not been as extensively used for LC as might be expected. Problems encountered with the robustness of LC/ECD (liquid chromatography/electrochemical detection) systems (electrode materials, electrode fouling, required experience) are probably the main reasons. Analyte detectability in LC/ECD is strongly influenced by electrode geometry and composition,[120,122,124] the LC eluent composition (especially the pH for carbaryl[120]), and the presence of electroactive coextractives.

LC/ECD of most aryl *N*-methylcarbamates and *N*-phenylcarbamates is easily performed, because these compounds can be oxidized within the available potential range. For example, aminocarb can be detected at a concentration as low as 5 $\mu g\,L^{-1}$.[120] The use of a detection potential of about 1.1 V is commonly considered best, although the oxime *N*-methylcarbamates and some aryl *N*-methylcarbamates (carbaryl and carbofuran) are electroinactive at this potential. Despite the practical limitations due to the oxidation of water, ECD at 1.9 V has been reported for some *N*-methylcarbamates (aldicarb and its sulfoxide and sulfone, aminocarb, bendiocarb, carbaryl, methiocarb, and methomyl); unfortunately, the LOD and oxidation potential were specifically mentioned for aminocarb only.[123] Alternatively, it has been that aryl *N*-methylcarbamates, such as carbaryl and carbofuran, can be hydrolyzed on-line to the corresponding phenols, which can be subsequently oxidized at potentials of 0.5–0.8 V.[125,126] The general tendency that lower potentials lead to a lower background current and, thus, to better LODs is illustrated by the LODs of 10–35 $\mu g\,L^{-1}$ obtained for some aryl *N*-methylcarbamates.[126] For the rest, it has been observed that LODs for ECD are one order of magnitude better than those obtained with UV detection; however, the reproducibility of ECD is lower than with UV detection, probably due to the slowly increasing contamination of the electrode surfaces referred to above.[120–126] From the above it is evident that the main use for LC/ECD is the target analysis of aryl *N*-methyl- and *N*-phenylcarbamates.

4.3 Fluorescence Detection

The number of compounds that display native fluorescence is rather restricted. As a consequence, fluorescence monitoring is a selective – and often highly sensitive – method of detection in LC. The selectivity of fluorescence over UV absorption detection is further improved by the fact that two wavelengths (excitation and emission) rather than one (absorption maximum) have to be selected. Compounds which do not exhibit natural fluorescence, such as most carbamates, can only be detected if they are derivatized or otherwise chemically converted into highly fluorescent products.

Fluorimetric detection of carbamates was used in the 1960s, either for direct analysis (e.g. carbaryl) or for the analysis of derivatives (e.g. benomyl).[129] The determination of carbamates and their degradation products by a combination of fluorescent labeling and LC was first demonstrated in 1974 (e.g. see Figure 3).[130] Precolumn derivatization with dansyl chloride and subsequent normal-phase separation gave detection limits of 0.5–2.5 $mg\,L^{-1}$.[130] The method was successfully applied to the determination of carbaryl in potato and corn samples.[171] On-line postcolumn derivatization, using *o*-phthalaldehyde (OPA) as the reagent, was first reported for carbamate analysis in 1978;[131] a scheme of the on-line analytical system is given in Figure 4.

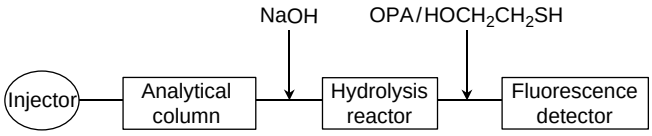

Figure 4 Scheme of an LC–hydrolysis–fluorescence detection system for *N*-methylcarbamates.

The OPA derivatization is preceded by alkaline hydrolysis, to convert the *N*-methylcarbamates into methylamine which is subsequently reacted with OPA in the presence of a reducing agent such as 2-mercaptoethanol; the resulting fluorophore is monitored. Several important parameters were studied and the reactions conditions were optimized.[133–136] The OPA derivatization finally formed the basis for the reference method (for *N*-methylcarbamates) of the US EPA[38] and was commercialized by Perkin-Elmer and the Pickering Laboratories. One main advantage is that OPA itself is completely nonfluorescent, so there is no need to remove the (large) excess of reagent after postcolumn reaction, prior to the actual detection.

The initial set-up of the OPA postcolumn reaction detection system had some disadvantages. Peak broadening occurred, due to the extra dead volume generated by the reaction coils and tee-pieces (the reagent delivery system for the two reaction steps), the hydrolytic conversion was not complete and reproducibility deteriorated over time due to the limited stability of 2-mercaptoethanol. As an alternative, catalytic conversion and heating[136] using a reactor packed with a basic anion-exchange resin[137–139] or magnesium oxide[140] proved successful. Although the rate of hydrolysis of the *N*-methylcarbamates differs widely, with temperature optima from 90–100 °C to over 140 °C,[137] LODs as low as 0.1 ng (aldicarb) and 0.85 ng (methiocarb) were reported for the modified system.[137] Miniaturization of the reactor, to achieve compatibility with narrow-bore LC, led to even lower LODs.[141] Photolytically initiated hydrolysis has been introduced, which makes *N*,*N*-dialkylcarbamates accessible to reaction with OPA. Photolytic hydrolysis sometimes requires the addition of sensitizers (e.g. acetone) to the eluent to effect the conversion of for example, aldicarb sulfone. Typical LODs of 2.5 $\mu g\,L^{-1}$ were reported.[142] Further optimization of the reactor geometry and of the solvent conditions allowed a one-step conversion of the carbamates into the OPA derivatives,[133] with similar LODs. Finally, the use of 3-mercaptopropionic acid and *N*,*N*-dimethyl-2-mercaptoethylamine hydrochloride (thiofluor) reagents instead of 2-mercaptoethanol greatly improved reproducibility over a longer time period.[144] The combination of this analytical method with on-line precolumn extraction and SPE-based trace enrichment resulted in detection limits of 0.02–0.03 $\mu g\,L^{-1}$ for all *N*-methylcarbamates, and some of their degradation products, in water samples.[27] The application of this method to fruit samples was optimized and validated.[46] In conclusion, on-line postcolumn OPA derivatization is an excellent tool for the determination of a large number of, though not all (such as *N*-phenylcarbamates), carbamates. Being essentially a derivatization-type procedure, it cannot be used for degradation products that do not yield a primary amine upon hydrolysis, nor does it provide structural information. Even so, it is possibly one of the best procedure available today for the trace-level analysis of carbamates.

4.4 Mass Spectrometric Detection

MS may be considered as a nearly universal method of detection, as gas-phase ions can be generated from volatile polyaromatic compounds and thermally labile proteins. The high-vacuum system of MS (typically operated at pressures below 1 Pa) was successfully coupled to GC in the late 1960s, and to LC in the early 1980s. Modern GC/MS instruments generally use capillary column GC and direct introduction of the column effluent into the MS source (operating either under EI or CI conditions). Commercially available GC/MS is routinely used and still the method of choice when it comes to identification (typically at microgram per liter levels) and/or of compound quantitation (typically at low microgram per liter levels) of volatile compounds. Since the late 1980, LC/MS has become an proven analytical tool for both qualitative and quantitative biopharmaceutical analysis. The most popular interfaces in the 1980s, TSP (thermospray) and PB (particle beam), have been replaced by the more robust atmospheric pressure ionization (API) techniques, ESP (electrospray) and APCI.

This section discusses the use of LC/MS in combination with TSP, PB, ESP, and APCI, for the determination of carbamates and their degradation products. As regards ionization methods, EI mass spectra of carbamates can only be generated by the PB interface and generally produces the most information to allow unequivocal identification, if necessary with the help of spectrum libraries. CI mass spectra of carbamates give molecular mass and some structural information depending on the applied reagent gas. Spectra from desorption ionization, such as fast atom bombardment (FAB), and from LC/MS-related methods of ionization (TSP, ESP) produce mass spectra with the protonated molecule or adduct ions and do not give any structural information.

The EI/MS of *N*-methylcarbamates has been reviewed in a book on environmental analysis.[148] Most *N*-monosubstituted carbamates show a molecular ion, $[M]^{+\bullet}$, whereas at least one characteristic fragment ion is usually

observed – the loss of neutral alkyl- or arylisocyanate produces the alcohol radical cation $[M - RNCO]^{+\bullet}$. The EI spectra of the *N*,*N*-dimethylcarbamates, as well as those of most degradation products, cannot be described in general terms.

The positive ion chemical ionization (PCI) mass spectra of carbamates generally show signals of a protonated molecule, $[M + H]^+$, and of the protonated alcohol, $[M + H - RNCO]^+$.[149–152] The loss of CH_3NCO from protonated *N*-methylcarbamates has been investigated in some detail by deuterium labeling and MS/MS experiments.[151,152] The intensity ratio of the $[M + H]^+$ and $[M + H - RNCO]^+$ signals varies with the reaction gas used. Moreover, some carbamates show adduct ion signals, $[M + NH_4]^+$, if ammonia is used as the reaction gas. In general, carbamates have a proton affinity close to that of ammonia and protonation should therefore result in good detection sensitivity.

The negative ion chemical ionization (NCI) mass spectra of carbamates have been less well studied. Studies[153–155] show that molecular anions, $M^{-\bullet}$, are only observed with some halogen-containing carbamates and that deprotonated molecules have a low abundance or are completely absent. Typical negative fragment ions are the alcoholate ions, $[RO]^-$, and the carbamic acid anion, (for example $[CH_3-NH-C(=O)-O]^-$ from *N*-methylcarbamates).[154,155] Deprotonation and negative fragment ion formation is inefficient for most carbamates, giving rise to poor detection sensitivity. Deprotonation is potentially more efficient for the degradation products, such as phenol-type compounds, and better sensitivity is to be expected. The general characteristics make NCI less suitable for carbamate analysis.

FAB ionization,[156] as well as LC/MS-related ionization (see below) most often generate mass spectra comparable to those obtained with chemical ionization and typically exhibit the loss of RNCO from $[M + H]^+$ in the positive ion mode. FAB has been applied to the specific detection of benomyl, a compound that is known to decompose to carbendazim under all other MS conditions.[156] FAB ionization, and any of the other desorption-type ionization techniques, is of very limited use for the determination of carbamates, because quantitation is extremely difficult, if not impossible, and because CI produces essentially the same information with fewer practical limitations.

As regards the potential of LC/MS, LC provides adequate separations and MS has the capability of detection and identification of trace amounts of compounds (typically at the nanogram level). This was recognized early in the development of LC/MS interfacing and, consequently, reports have appeared on carbamate analysis using moving belt (MB),[157] direct liquid introduction (DLI),[158,159] PB,[160–166] TSP,[166–181] ESP or ISP,[166,182–188] and APCI.[169,189–196] Recently, PB, TSP, ISP, and APCI were compared in a single study, using eight selected carbamates.[166] The impact of these latter interfacing methods on carbamate analysis is briefly evaluated.

Only the PB interface is capable of (nearly) separating the analyte from the eluent solvents and transfering it to the MS high vacuum. In principle, this allows operation of the mass spectrometer with common MS ionization modes, EI, and CI (eluent independent). However, PB interfacing generally has poor sensitivity (typically at the milligram per liter level) compared with the TSP, ESP, and APCI interfaces. As observed for TSP, the response sensitivity was shown to decrease if more organic modifier is used in the mobile phase.[161] The LC-to-MS transfer efficiency of analytes with PB is concentration dependent[161] and calibration curves can therefore be nonlinear over a large range. It has been claimed, but not convincingly demonstrated, that linear calibration curves at low concentration can be obtained by adding a carrier compound to the LC eluent.[161] The relatively low sensitivity of detection may be improved by sample preconcentration, using on-line SPE, thus allowing LC/PB/MS identification from EI mass spectra at low microgram per liter levels.[163,164] Another factor complicating the analysis of polar pesticides in general, and carbamates in particular, is the occurrence of thermal degradation during the desolvation process. Complete desolvation of the analyte is achieved by colliding the analyte–solvent cluster against the heated ion-source wall, resulting in an increase of the relative abundance of the fragment ion of *N*-methylcarbamates, namely $[M + H - CH_3NCO]^+$ in CI and $[M + H - CH_3NCO]^{+\bullet}$ in the EI mode.[165] Under conditions of ammonia CI the situation becomes even more complicated, because the intensities of the $[M + H + NH_3]^+$ and $[M + H]^+$ ions are strongly dependent upon the ammonia pressure in the ion source. It is clear that many parameters need to be optimized[163,164] before good analytical conditions in terms of response linearity, reproducibility, and detection limits can be obtained.[160] The lack of sensitivity, especially when compared with those of the other three interfaces,[166] and the danger of thermal degradation hampering the reproducibility does not make the PB interface the preferred interface for the quantitative analysis of the carbamate pesticides. Nevertheless, it is still a powerful tool for the identification of unknowns.

TSP and APCI typically yield CI-like mass spectra with the protonated solvent and additive ions acting as the reagent gas. In addition, ion–molecule complex formation of the analyte and eluent components is frequently observed. However, the reagent gas pressure and ion source temperature strongly influence the equilibrium of the gas-phase reaction and the mass spectra obtained with

Table 3 Ions with m/z 239 $[M + NH_4]^+$, m/z 222 $[M + H]^+$, and m/z 165 $[M + H - CH_3NCO]^+$ reported for carbofuran, when using various MS ionization methods

Ionization (specification)	Ion m/z (relative abundance)		Ref.
CI (NH_3 reagent)	222 (100)	165 (11)	51
CI (CH_4 reagent)	165 (100)	222 (24)	51
PB/CI (NH_3 reagent)	165 (100)	165 (100)	166
TSP (filament off)	222 (100)	239 (38)	8
TSP (filament off)	239 (100)	222 (67)	59
TSP (filament on)	239 (100)	222 (40)	17
TSP (filament on)	239 (100)	222 (25)	166
APCI	222 (100)	165 (64)	166

these interfacing methods can differ widely, as illustrated for carbofuran in Table 3. The LC eluent composition, source temperature, source pressure, flow-rate, and ion source tuning should be optimized to produce maximum sensitivity. Optimization of the experimental conditions often requires compromises, because most parameters are interdependent and different parameter optima for different compounds are most often obtained.

TSP has been the most popular interfacing technique, especially for the analysis of carbamate pesticides. The TSP interface places some restrictions on the chromatographic system (non-volatile salts or reagents should not be present), but is effective in the presence of an ionogenic compound (typically ammonium acetate).[159] This additive is required to induce proton transfer reactions during and after evaporation, thus causing the formation of detectable ions. In the PCI mode, protonation and ammoniation of the analyte are typically observed (giving rise to $[M + H]^+$ and $[M + NH_4]^+$ ions, respectively). In the NCI mode, electron capture, deprotonation, and acetate adduct formation are commonly observed (giving rise to $M^{-\bullet}$, $[M - H]^-$ and $[M + CH_3COO]^-$ ions, respectively[177]).

Most carbamates are amenable to TSP/MS detection, despite the thermal character of the desolvation process.[177] Detection of carbamates is generally most efficient in the PCI mode, whereas the NCI mode is best for detection of halogenated carbamates and hydroxyl-containing degradation products. The performance of TSP/MS is improved, at least for carbamates, if the ionization is assisted either by using a discharge or by a filament.[166] Under all TSP conditions the carbamates tend to fragment by losing of CH_3NCO from protonated molecular ions or even from adduct ions. More specific fragmentation may be obtained from MS/MS experiments, such as with triple quadrupole MS, but MS/MS techniques generally cause the analyte detectability to deteriorate. Furthermore, many carbamates tend to form adduct ions, with ammonia or with solvent molecules, or (apparently) with strange ions, such as $[M + 32]^+$ and $[M + 59]^+$.[181] Although fragmentation may provide limited means of structure attribution, the occurrence of odd and adduct ions may hamper identification. Moreover, fragmentation, odd ion, and adduct ion formation represent side reactions which may be undesirable from the point of view of quantitative analysis.

The parameters that characterize TSP conditions have been studied extensively for carbamates.[180,181] Instrument parameters, specifically the geometry of the TSP interface and the MS source, are (generally) invariant parameters, which may have a major influence. With filament or discharge-assisted ionization and PCI detection being the most effective, ion source temperature, vaporizer temperature, and eluent composition are important optimizable parameters, which may exhibit different optima for different instruments.

High source temperatures (up to 300 °C) lead to more intense fragment ion signals ($[M + H - CH_3NCO]^+$ and $[M + NH_4 - CH_3NCO]^+$) with asulam, carbaryl, carbofuran, oxamyl, chloropropham, desmedipham, and phenmedipham.[175] This increase in intensity is probably due to thermal dissociation by loss of CH_3NCO, with subsequent protonation (ammoniation) of the alcohol-type reaction product, because near-thermal equilibration conditions occur inside the source under the high pressure applied (estimated at <13.3 Pa). Hence, the source temperature for detection of carbamates, as such, has a maximum.

The vaporizer temperature may also induce thermal dissociation, as shown for methiocarb.[180] This temperature is the more critical because adequate evaporation and stable ion currents are only obtained above a certain minimum temperature. Thermal dissociation is easily established by monitoring fragment and protonated molecule currents over a range of temperatures. The vaporizer temperature may well influence the degree of desolvation of the analytes, but no effects on adduct ion intensities have been reported.

The LC eluent composition affects the efficiency of detection, because it primarily determines the possibility for adduct ion formation, for protonation, and for fragmentation.[180] Moreover, the eluent components influence initial vapor formation and different eluents therefore generally require different vaporizer temperatures.

The most widely used LC/MS coupling techniques are the API interfaces, ESP, ISP and the HN. The latter is better known as the APCI interface. They provide superior performance in terms of detectability, reproducibility, and response linearity over TSP and PB. The major drawback of the original ESP interface was the limited flow rate, typically 2–10 $\mu L\,min^{-1}$, making postcolumn splitting of the eluents necessary. In order to increase the applicable flow rate, ultrasonically,[197] thermally,[198]

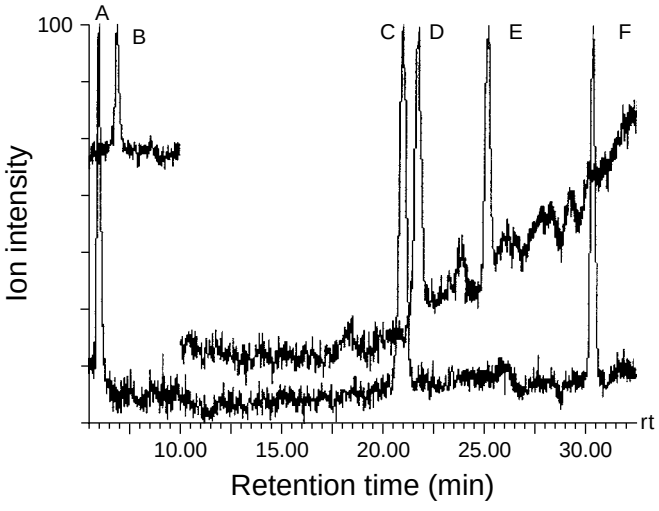

Figure 5 Time-scheduled selected ion mode LC/ISP/MS chromatogram of tap water, spiked with (A) oxamyl, (B) methomyl, (C) propoxur, (D) carbofuran, (E) pirimicarb, and (F) promecarb at $0.1\,\mu g\,L^{-1}$, utilizing a $500\,\mu L$ injection onto a $12.5\,cm \times 3\,mm$ i.d. LC column.

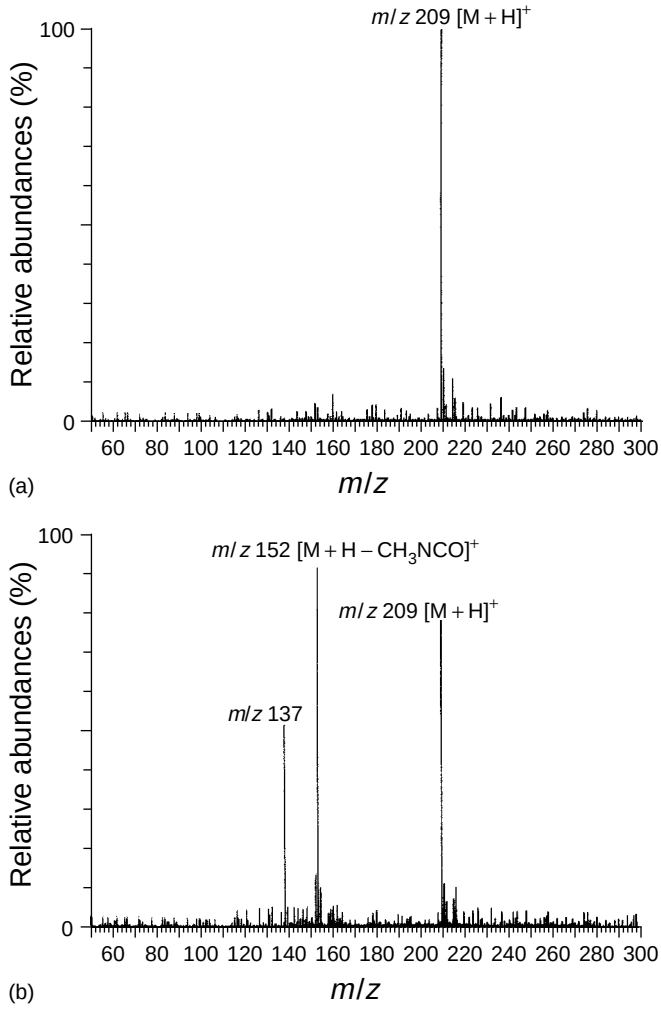

Figure 6 Full-scan (m/z 50–300) ISP mass spectra of aminocarb at extraction voltages of (a) 25 V and (b) 45 V. The analyte was $0.5\,\mu g\,mL^{-1}$ dissolved in methanol–water (50 : 50), with a flow rate of $0.3\,mL\,min^{-1}$.

and pneumatically[199] assisted ESP interfaces have been developed. The latter type can be used with flow rates of up to $1\,mL\,min^{-1}$ and has been applied to the determination of carbamates.[184] Nevertheless, optimal sensitivity for the carbamates is obtained at flow rates of approximately $0.3\,mL\,min^{-1}$,[185,186] and postcolumn splitting still remains an option, because the interface design is responsible for the concentration dependence of the MS signal. LODs are typically in the low-picogram range,[186] making possible the determination of carbamates at $0.1\,\mu g\,L^{-1}$ in tap water without any sample preconcentration (see Figure 5). Moreover, the reproducibility of ISP is far better than TSP. The reproducibility of ISP is strongly dependent on the stability of the ion evaporation (IEV) and the desolvation processes in the ion source. Complete or stable desolvation in the API source is achieved by optimizing the counter-current nitrogen gas flow and the voltage on the sampling cone of the MS. At high cone voltages, collisionally induced dissociation can take place generating fragment ions, which facilitates structural identification (see Figure 6). However, high cone voltages can cause dispersion of the ions (the intensity of the protonated molecule is also reduced due to the fragmentation process) lowering the signal intensity and increasing the chemical noise. The interface was successfully applied to identify the carbamates carbofuran and methiocarb together with their degradation products in groundwater.[187]

Both parameters are also of importance with the HN. With this interface the eluent is evaporated by heat-assisted pneumatic nebulization, allowing the use of conventional LC systems with eluent flow rates of $1–2\,mL\,min^{-1}$. Despite heating the probe to temperatures of typically 500 °C, achieving complete desolvation, no thermal degradation of the carbamates was observed.[166] The counter-current gas flow and the cone voltage show a comparable behavior to that observed with the ISP interface, and ions are formed by gas-phase reaction at atmospheric pressure of the desolved analytes with a reagent gas, a process better known as APCI. By changing the eluent composition, and consequently the reagent gas, fragment ions can be generated, making structural confirmation possible. Selectivity can also be influenced by the choice of the reagent gas. However, before ionization, the compound is evaporated to the gas phase, which makes the HN interface unsuitable for truly nonvolatile compounds, such as azo dyes. Sensitivity is good with detection limits of typically $2–750\,ng\,L^{-1}$ using

analytical columns of a few centimeters and the selected ion mode.[195]

Regarding its sensitivity, reproducibility, and dynamic range, APCI can be considered as the method of choice[166] for the qualitative and quantitative determination of carbamates and other less polar compounds, whereas ESP and ISP are best suited for polar species or molecules with acidic (phenoxyacids, sulfonated surfactants) or basic groups. Application of proper off- or on-line preconcentration techniques, in combination with TSP and PB, will make the determination of carbamates at concentration levels of $0.1\,\mu g\,L^{-1}$ possible. Nevertheless, the application of these interfaces is mainly hampered by thermal degradation, which is responsible for poor reproducibility.

5 CONCLUSIONS

Carbamate pesticides and their transformation products can be determined in various matrices at low and even sub microgram per liter levels. Table 4 gives typical LODs for oxime and aryl *N*-methylcarbamates in aqueous samples. It should be noted that the methods of sample preparation and of separation in the quoted papers differ widely, and that the compound classes selected are not representative for all carbamates. Moreover, the listed detection limits can be decreased up to three order of magnitude when off- or on-line preconcentration techniques are applied.

Isolation of carbamates from water samples is preferably carried out by SPE, whereas SLE still is the method of choice for biological materials and soil. With SFE good results are obtained for the extraction of analytes from biological materials and soil, but more research is required. LC is the most applied separation technique for the determination of the carbamate pesticides, preferably performed on C_8- or cyanopropyl-bonded silica with methanol–water or acetonitrile–water mixtures as eluent. GC is still a good alternative with PTV. With this technique thermal degradation of the carbamates in the injector is prevented.

Table 4 LODs of oxime and aryl *N*-methylcarbamates, as reported for LC with various detectors

Detection mode	LOD ($\mu g\,L^{-1}$)	Ref.
UV/DAD	20–200	107
ECD	1–7	123
Fluorescence	1–10	32
PB/MS	2000–88 000	160
TSP/MS	40–150	166
ESP/MS	1–10	182
ISP/MS	10–100	184
APCI/MS	2–750	159

UV is an universal detection technique and therefore the conventional choice when all carbamates need to be determined. The lack of sensitivity can be compensated by the application of on-line preconcentration techniques, resulting in the determination of the carbamates at low microgram per liter levels. Even more sensitivity and selectivity is achieved by subjecting them on-line, to first, hydrolysis and then reaction with OPA, to form a highly fluorescent product, making fluorescence detection the method of choice for the quantitative analysis of *N*-methyl and carbamoyl carbamate pesticides. Comparable detection limits are obtained with MS detection, which is a sensitive and universal detector. Combination with on-line preconcentration, which may easily improve analyte detectability by up to 1000-fold, provides the means to confirm, and sometimes unambiguously identify, all carbamates and their degradation products.

LC/PB/MS has a good identification potential that requires substantial preconcentration if the analytes have to be detected at low levels.[201,202] However, many parameters influence sensitivity and reproducibility, making it and difficult tool to handle. LC/TSP/MS in combination with preconcentration has been successfully applied for many years. Nevertheless, problems comparable with those of the PB type of interface have been encountered, hampering quantitative analysis.

The reports on API-type interfaces show that LC/APCI/MS and LC/ISP/MS are the main competitors of the former interfacing techniques with respect to sensitivity and identification potential.[166] No evidence of thermal degradation is observed with either API interface. Consequently, the requirements for multiresidue analysis of carbamate and carbamoyloxime insecticides at trace levels ($<0.1\,\mu g\,L^{-1}$) are easily met by the combination of LC/API/MS and preconcentration techniques (with sample volumes of less than 50 mL), be it off- or on-line.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
CE	Capillary Electrophoresis
CI	Chemical Ionization
CI/MS	Chemical Ionization Mass Spectrometry
DAD	Diode Array Detection
DLI	Direct Liquid Introduction
ECD	Electrochemical Detection
EI	Electron Ionization
EPA	Environmental Protection Agency
ESP	Electrospray
EU	European Union

FAB	Fast Atom Bombardment
FID	Flame Ionization Detection
GC	Gas Chromatography
GC/CI/MS	Gas Chromatography/Chemical Ionization/Mass Spectrometry
GPC	Gel Permeation Chromatography
HN	Heated Nebulizer
i.d.	inner diameter
IEV	Ion Evaporation
ISP	Ionspray
LC	Liquid Chromatography
LC/ECD	Liquid Chromatography/ Electrochemical Detection
LC/MS	Liquid Chromatography/ Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LC/UV	Liquid Chromatography/Ultraviolet Spectroscopy
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
MB	Moving Belt
MEKC	Micellar Electrokinetic Chromatography
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NCI	Negative Ion Chemical Ionization
NPD	Nitrogen/Phosphorus Detection
OPA	*o*-Phthalaldehyde
PB	Particle Beam
PCI	Positive Ion Chemical Ionization
PTV	Programmed Temperature Vaporization
RSD	Relative Standard Deviation
SDS	Sodium Dodecyl Sulfate
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SLE	Solid–Liquid Extraction
SPE	Solid-phase Extraction
TSD	Thermal Sulfur Detection
TSP	Thermospray
UV	Ultraviolet

RELATED ARTICLE

Pesticides **(Volume 7)**
Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of

REFERENCES

1. R.J. Kuhr, H.W. Dorough (eds.), *Carbamate Insecticides: Chemistry Biochemistry and Toxicology*, CRC Press, Cleveland, 1976.
2. K.A. Hassall, *The Biochemistry and Uses of Pesticides*, VCH Verlag, Weinheim, 125–154, 1990.
3. R.T. Krause, 'Multiresidue Methods for Carbamate Pesticides', in *Emerging Strategies for Pesticide Analysis*, eds. T. Cairns, J. Sherma, CRC Press, Boca Raton, 99, 1992.
4. B.G.L. Schlagbauer, A.W. Schlagbauer, 'The Metabolism of Carbamates, Literature Analysis I and II', in *The Metabolism of Carbamates*, eds. F.A. Gunther, J. Davies Gunther, Residue Reviews, Springer Verlag, New York, 1, Vol. 42, 1972.
5. N.B.K. Murthy, K. Raghu, 'Fate of 14C-Carbaryl in Soils as a Function of pH', *Bull. Environ. Contam. Toxicol.*, **46**, 374–379 (1991).
6. K. Ramanand, M. Sharmila, N. Singh, N. Sethunathan, 'Metabolism of Carbamate Insecticides by Resting Cells and Cell-free Preparations of a Soil Bacterium Arthrobacter Sp.', *Bull. Environ. Contam. Toxicol.*, **46**, 380–386 (1991).
7. M. Chiba, D.F. Veres, 'Fate of Benomyl and Its Degradation Compound Methyl-2-benzimidazolecarbamate on Apple Foliage', *J. Agric. Food Chem.*, **29**(3), 588–590 (1985).
8. J. Harvey, Jr, J.C-Y. Han, 'Metabolism of Oxamyl and Selected Metabolites in the Rat', *J. Agric. Food Chem.*, **26**(4), 902–910 (1978).
9. S.W. Fisher, 'Changes in the Toxicity of Three Pesticides as a Function of Environmental pH and Temperature', *Bull. Environ. Contam*, *Toxicol*, **46**(2), 197–202 (1991).
10. T. Cairns, E.G. Siegmund, T.S. Savage, 'Persistence and Metabolism of Aldicarb in Fresh Potatoes', *Bull. Environ. Contam. Toxicol.*, **32**(3), 274–281 (1984).
11. E.K. Dzantor, A.S. Felsot, 'Soil Differences in the Biodegradation of Carbofuran and Trimethacarb Following Pretreatment With These Insecticides', *Bull. Environ. Contam. Toxicol.*, **45**(4), 531–537 (1990).
12. C.J. Miles, M.L. Trehy, R.A. Yost, 'Degradation of *N*-Methylcarbamate and Carbamoyl Oxime Pesticides in Chlorinated Water', *Bull. Environ. Contam. Toxicol.*, **41**(6), 838–843 (1988).
13. D.A. Volmer, 'Investigation of Photochemical Behavior of Pesticides in a Photolysis Reactor Coupled On-line With a Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry System. Application to Trace and Confirmatory Analysis in Food Samples', *J. Chromatogr. A*, **794**(1–2), 129–146 (1998).
14. G.S. Nunes, I.A. Toscano, D. Barceló, 'Analysis of Pesticides in Food and Environmental Samples by Enzyme-linked Immunosorbent Assays', *Trends Anal. Chem.*, **17**(2), 79–87 (1998).
15. R.J. Bushway, B.E.S. Young, L.R. Paradis, L.B. Perkins, S.K. Martin, M.P. Brown, 'Determination of Methyl 2-Benzimidazolecarbamate in Bulk Fruit Juice Concentrates by Competitive-inhibition Enzyme Immunoassays', *J. Assoc. Off. Anal. Chem.*, **77**(5), 1237–1243 (1994).

16. K.M. Sundaram, S.Y. Szeto, R. Hindle, 'Detection of Aminocarb and Its Metabolites by Thin-layer Chromatography', *J. Chromatogr*, **194**(1), 100–103 (1980).
17. B.C. Verma, B.S. Atwal, D.P. Thakur, H.K. Thakur, 'Spectrometric Methods for the Determination of Carbaryl, Carbofuran and Baygon in Commercial Formulations', *J. Indian Chem. Soc.*, **75**(1), 61–62 (1998).
18. G.S. Nunes, P. Skladal, H. Yamanaka, D. Barceló, 'Determination of Carbamate Residues in Crop Samples by Cholinesterase-based Biosensors and Chromatographic Techniques', *Anal. Chim. Acta*, **362**(1), 59–68 (1998).
19. G.S. Nunes, D. Barceló, 'Electrochemical Biosensors for Pesticide Determination in Food Samples', *Analysis*, **26**(6), M156–M159 (1998).
20. S. Wu, H.K. Lee, S.F.Y. Li, 'Separation and Determination of Pesticides by Capillary Electrophoresis. I. Rapid Separation of Fifteen *N*-Methylcarbamate Insecticides via Micellar Electrokinetic Chromatography', *J. Microcolumn Sep.*, **10**(6), 239–247 (1998).
21. Y.S. Wu, H.K. Lee, S.F.Y. Li, 'Separation and Determination of Pesticides by Capillary Electrophoresis. II. Determination of *N*-Methylcarbamates in Drinking Water by Micellar Electrokinetic Chromatography With SPE and On-column Enrichment', *J. Microcolumn Sep.*, **10**(6), 529–535 (1998).
22. D. Barceló, 'A Review of Liquid Chromatography in Environmental Pesticide Analysis', *Chromatographia*, **25**(10), 928–936 (1988).
23. D. Barceló, 'Occurrence, Handling and Chromatographic Determination of Pesticides in the Aquatic Environment. A Review', *Analyst*, **116**(7), 681–689 (1991).
24. D. Barceló, 'Environmental Protection Agency and Other Methods for the Determination of Priority Pesticides and Their Transformation Products', (Review), *J. Chromatogr.*, **643**(1), 117–143 (1993).
25. D. Barceló, 'Mass Spectrometry in Environmental Organic Analysis. A Review', *Anal. Chim. Acta*, **263**(1), 1–19 (1992).
26. D. Chaput, 'On-line Trace Enrichment for Determination of Aldicarb Species in Water, Using Liquid Chromatography With Post-column Derivatization', *J. Assoc. Off. Anal. Chem.*, **69**(6), 985–989 (1986).
27. A. de Kok, M. Hiemstra, U.A.Th. Brinkman, 'Low ng L^{-1}-level Determination of Twenty *N*-Methylcarbamate Pesticides and Twelve of Their Polar Metabolites in Surface Water via off-line Solid-phase Extraction and High Performance Liquid Chromatography With Post-column Reaction and Fluorescence Detection', *J. Chromatogr. A*, **623**(1), 265–276 (1992).
28. T.A. Bellar, W.L. Budde, 'Determination of Nonvolatile Organic Compounds in Aqueous Environmental Samples Using Liquid Chromatography Mass Spectrometry', *Anal. Chem.*, **60**(19), 2076–2083 (1988).
29. A. Di Corcia, M. Marchetti, 'Multi Residue Method for Pesticides in Drinking Water Using Graphitized Carbon Black Cartridge Extraction and Liquid Chromatographic Analysis', *Anal. Chem.*, **63**(6), 580–585 (1991).
30. C.H. Marvin, I.D. Brindle, C.D. Hall, M. Chiba, 'Development of an Automated High-performance Liquid Chromatographic Method for the On-line Preconcentration and Determination of Trace Concentrations of Pesticides in Drinking Water', *J. Chromatogr.*, **503**(1), 167–176 (1990).
31. C.H. Marvin, I.D. Brindle, C.D. Hall, M. Chiba, 'Rapid On-line Precolumn High-performance Liquid Chromatographic Method for the Determination of Benomyl, Carbendazim and Aldicarb Species in Drinking Water', *J. Chromatogr.*, **555**(1–2), 147–154 (1991).
32. S. Chiron, A. Fernandez-Alba, D. Barceló, 'Comparison of On-line Solid-phase Disk Extraction and Liquid–Liquid Extraction for Monitoring Selected Pesticides in Environmental Waters', *Environ. Sci. Technol.*, **27**(9), 2352–2359 (1993).
33. S. Chiron, D. Barceló, 'Determination of Pesticides in Drinking Water by On-line Solid-phase Disk Extraction Followed by Various Liquid Chromatographic Systems', *J. Chromatogr.*, **645**(1), 125–134 (1993).
34. G. Durand, S. Chiron, V. Bouvot, D. Barceló, 'Use of Extraction Disks for Trace Enrichment of Various Pesticides From River and Seawater Samples', *Int. J. Environ. Anal. Chem.*, **49**(1), 31–42 (1992).
35. M.R. Diss, M.-C. Hennion, M.L. Bouguerra, 'Determination of Carbaryl and Some Organophosphorus Pesticides in Drinking Water Using On-line Liquid Chromatographic Preconcentration Techniques', *J. Chromatogr.*, **639**(2), 352–358 (1993).
36. S. Chiron, S. Dupas, P. Scribe, D. Barceló, 'Application of On-line Solid-phase Extraction Followed by Liquid Chromatography Thermospray Mass Spectrometry to the Determination of Pesticides in Environmental Waters', *J. Chromatogr.*, **665**(2), 295–305 (1994).
37. U.A.Th. Brinkman, 'On-line Sample Treatment for or via Column Liquid Chromatography', *J. Chromatogr.*, **665**(2), 217–231 (1994).
38. D.L. Foerst, H.A. Moye, EPA/600/D-85/051, US Environmental Protection Agency, Off. Res. Dev., Cincinnatti, 1985; R.L. Graves, Revision 3.0, 1989.
39. J. Rouchaud, F. Gustin, F. van de Steene, C. Pelerents, L. VanParys, M. de Proft, E. Seutin, J. Gillet, F. Benoit, N. Ceustermens, 'Effects of Organic Fertilizers on Aldicarb Soil Biodegradation in Sugar Beet Crops', *Arch. Environ. Contam. Toxicol.*, **24**(1), 67–74 (1993).
40. I.H. Williams, M.J. Brown, 'Determination of Carbofuran and 3-Hydroxycarbofuran Residues in Small Fruits', *J. Agric. Food Chem.*, **21**(3), 399–401 (1973).
41. G.J. Sharp, J.G. Brayan, S. Dilli, P.R. Haddad, J.M. Desmarchelier, 'Extraction, Clean-up and Chromatographic Determination of Organophosphate, Pyrethroid and Carbamate Insecticides in Grain and Grain Products', *Analyst*, **113**(10), 1493–1507 (1988).

42. C.E. Goewie, E.A. Hoogendoorn, 'Precolumn Clean-up and Liquid Chromatographic Determination of Residues of *N*-Methylcarbamates in Extracts of Total Diet', *J. Chromatogr.*, **404**(2), 352–358 (1987).
43. B.D. McGarvey, M. Chiba, A.B. Broadbent, 'Simplified Cleanup and Liquid Chromatographic Determination of Oxamyl Residues in Potato Tubers', *J. Assoc. Off. Anal. Chem.*, **68**(4), 753–756 (1985).
44. R.T. Krause, 'Liquid Chromatographic Determination of *N*-Methylcarbamate Insecticides and Metabolites in Crops. II. Analytical Characteristics and Residue Findings', *J. Assoc. Off. Anal. Chem.*, **68**(4), 734–741 (1985).
45. A. de Kok, M. Hiemstra, C.P. Vreeker, 'Improved Cleanup Method for the Multi Residue Analysis of *N*-Methylcarbamates in Grains, Fruits and Vegetables by Means of HPLC With Postcolumn Reaction and Fluorescence Detection', *Chromatographia*, **24**(3), 469–476 (1987).
46. A. de Kok, M. Hiemstra, 'Optimization, Automation, and Validation of the Solid-phase Extraction Clean-up and On-line Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides in Fruits and Vegetables', *J. Assoc. Off. Anal. Chem.*, **75**(6), 1063–1073 (1992).
47. D. Chaput, 'Simplified Multi Residue Method for Liquid Chromatographic Determination of *N*-Methylcarbamate Insecticides in Fruits and Vegetables', *J. Assoc. Off. Anal. Chem.*, **71**(3), 542–546 (1988).
48. G.S. Rule, A.V. Mordehai, J. Henion, 'Determination of Carbofuran by On-line Immuno Affinity Chromatography With Coupled-column Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **66**(2), 230–235 (1994).
49. D.M. Holstege, D.L. Scharberg, E.R. Tor, L.C. Hart, F.D. Galey, 'A Rapid Multiresidue Screen for Organophosphorus, Organochlorine, and *N*-Methylcarbamate Insecticides in Plant and Animal Tissues', *J. Assoc. Off. Anal. Chem.*, **77**(5), 1263–1274 (1994).
50. D.W. Woodham, R.R. Edwards, R.G. Reeves, R.L. Schutzman, 'Total Toxic Aldicarb Residues in Soil, Cotton Lint Following a Soil Treatment With the Insecticide on the Texas High Plains', *J. Agric. Food Chem.*, **21**(2), 303–307 (1973).
51. M. Galoux, J.C. van Damme, A. Bernes, J. Potvin, 'Gas–Liquid Chromatographic Determination of Aldicarb, Aldicarb Sulfoxide and Aldicarb Sulfone in Soils and Water Using a Hall Electrolytic Conductivity Detector', *J. Chromatogr.*, **177**(1), 245–253 (1979).
52. K.A. Johnson, F.D. Harper, C.P. Weisskopf, 'Solid-phase Extraction of Aldicarb and Its Metabolites From Water and Soil', *J. Environ. Qual.*, **26**(5), 1435–1438 (1997).
53. B.D. McGarvey, 'High-performance Liquid Chromatographic Methods for the Determination of *N*-Methylcarbamate Pesticides in Water, Soil, Plants and Air', *J. Chromatogr.*, **642**(1), 89–105 (1993).
54. 'National Pesticide Survey, Phase I Report', US Environmental Protection Agency, PB-91-125765, National Technical Information Service, Springfield, 1990.
55. 'National Survey of Pesticides in Drinking Water Wells, Phase II Report', US Environmental Protection Agency, EPA 570/9-91-20, National Technical Information Service, Springfield, 1990.
56. A. Izquierdo, M.T. Tena, M.D. Luque de Castro, M. Valcarcel, 'Supercritical Fluid Extraction of Carbamate Pesticides From Soils and Cereals', *Chromatographia*, **42**(3–4), 206–212 (1996).
57. A.L. Howard, C. Braue, L.T. Taylor, 'Feasibility of Thiocarbamate Pesticide Analysis in Apples by Supercritical Fluid Extraction and High-performance Liquid Chromatography', *J. Chromatogr. Sci.*, **31**(8), 323–329 (1993).
58. J.W. King, M.L. Hopper, R.G. Luchtefeld, S.L. Taylor, W.L. Orton, 'Optimization of Experimental Conditions for the Supercritical Carbon Dioxide Extraction of Pesticide Residues from Grains', *J. Assoc. Off. Anal. Chem.*, **76**, 857 (1993).
59. K.-S. Nam, J.W. King, 'Supercritical Fluid Extraction and Enzyme Immunoassay for Pesticide Detection in Meat Products', *J. Agric. Food Chem.*, **42**, 1469–1472 (1994).
60. J.J. Jimenez, J. Atienza, J.L. Bernal, L. Toribio, 'Determination of Carbendazim in Lettuce Samples by SFE/HPLC', *Chromatographia*, **38**(5/6), 395–399 (1994).
61. I.H. Williams, 'Carbamate Insecticide Residues in Plant Material: Determination by Gas Chromatography', *Residue Rev.*, **30**, 1 (1971).
62. M.J. Climent, M.A. Miranda, 'Gas Chromatographic/Mass Spectrometric Study of Photodegradation of Carbamate Pesticides', *J. Chromatogr. A*, **738**(2), 225–231 (1996).
63. M.L. Bao, F. Pantani, K. Barbieri, D. Burrini, O. Griffini, 'Multi-residue Pesticide Analysis in Soil by Solid-phase Disk Extraction and Gas Chromatography/Ion-trap Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **64**(4), 23–245 (1996).
64. H.M. Müller, H.J. Stan, 'Pesticide Residue Analysis in Food With CGC – Study of Long-term Stability by the Use of Different Injection Techniques', *J. High Resol. Chromatogr.*, **13**(10), 697–701 (1990).
65. H.M. Müller, H.J. Stan, 'Thermal Degradation Observed With Different Injection Techniques: Quantitative Estimation by the Use of Thermolabile Carbamate Pesticides', *J. High Resol. Chromatogr.*, **13**(11), 759–763 (1990).
66. H.M. Müller, PhD Thesis, Technical University Berlin, Germany, 1989.
67. J.E. Bagness, W. Sharples, 'Determination of the Substituted Pyrimidine Pesticides Ethirimol, Dimethirimol, Piriphos-methyl, Pirimiphos-ethyl and Pirimicarb in Technical and Formulated Materials by Gas Chromatography', *Analyst*, **99**(177), 225–232 (1974).

68. D.L. Lewis, D.F. Paris, 'Direct Determination of Carbaryl by Gas–Liquid Chromatography Using Electron Capture Detection', *J. Agric. Food Chem.*, **22**(1), 148–149 (1974).
69. J.N. Sieber, '*N*-Perfluoroacyl Derivatives for Methylcarbamate Analysis by Gas Chromatography', *J. Agric. Food Chem.*, **20**(2), 443–446 (1972).
70. R.L. Tilden, C.H. van Middelem, 'Determination of Carbaryl as an Amide Derivative by Electron-capture Gas Chromatography', *J. Agric. Food Chem.*, **18**(1), 154–158 (1970).
71. S.C. Lau, R.L. Marxmiller, 'Residue Determination of Landrin Insecticide by Trifluoroacetylation and Electron-capture Gas Chromatography', *J. Agric. Food Chem.*, **18**(3), 413–415 (1970).
72. L.I. Butler, L.M. McDonough, 'Determination of Residues of Carbofuran and Its Toxic Metabolites by Electron-capture Gas Chromatography After Derivative Formation', *J. Assoc. Off. Anal. Chem.*, **54**(6), 1357–1360 (1971).
73. J.J. Ryan, J.F. Lawrence, 'Comparison of Electron-capture and Electrolytic Conductivity Detection for the Gas–Liquid Chromatographic Analysis of Some Perfluoro Derivatives of Four Agricultural Chemicals', *J. Chromatogr.*, **135**(1), 117–122 (1977).
74. J.A. Coburn, B.D. Ripley, A.S.Y. Chau, 'Analysis of Pesticide Residues by Chemical Derivatization. II. *N*-Methylcarbamates in Natural Water and Soils', *J. Assoc. Off. Anal. Chem.*, **59**(1), 188–196 (1976).
75. L. Wong, F.M. Fisher, 'Determination of Carbofuran and its Toxic Metabolites in Animal Tissues by Gas Chromatography of Their *N*-Trifluoroacetyl Derivatives', *J. Agric. Food Chem.*, **23**(2), 315–318 (1975).
76. B.C. Leppert, J.C. Markle, R.C. Helt, H.H. Fujie, 'Determination of Carbosulfan and Carbofuran Residues in Plants, Soil and Water by Gas Chromatography', *J. Agric. Food Chem.*, **31**(2), 220–223 (1983).
77. R.J. Nakamoto, M. Page, 'Simplified Determination of Carbaryl in Rainbow Trout Liver Tissues', *Bull. Environ. Contam. Toxicol.*, **37**(3), 415–420 (1986).
78. M.A. Luke, J.E. Froberg, G.M. Doose, H.T. Masumoto, 'Improved Multiresidue Gas Chromatographic Determination of Organophosphorus, Organonitrogen, and Organohalogen Pesticides in Procedure, Using Flame Photometric and Electrolytic Conductivity Detectors', *J. Assoc. Off. Anal. Chem.*, **64**(5), 1187–1195 (1981).
79. H. Pyysalo, 'A Rapid Gas–Liquid Chromatography Method for Determining Benomyl Residues in Foods', *J. Agric. Food Chem.*, **25**, 995 (1977).
80. H.J. Stan, P. Klaffenbach, 'Determination of Carbamate Pesticides and Some Urea Pesticides After Derivatization With Acetic Anhydride by Means of GC/MSD', *Fresenius' J. Anal. Chem.*, **339**(2), 151–157 (1991).
81. A. Vincze, J. Yinon, 'Analysis of Thermally Labile Pesticides by Gas Chromatography/Mass Spectrometry and Gas Chromatography/Tandem Mass Spectrometry With a Temperature-programmed Injector', *Rapid Commun. Mass Spectrom*, **10**(13), 1638–1644 (1996).
82. H.A. Moye, 'Esters of Sulfonic Acids as Derivatives for the Gas Chromatographic Analysis of Carbamate Pesticides', *J. Agric. Food Chem.*, **23**(3), 415–418 (1975).
83. J.P. Rouchaud, J.R. Decallonne, 'A Gas Chromatographic Method for the Analysis of MBC in Plants and Soil', *J. Agric. Food Chem.*, **22**(2), 259–260 (1974).
84. L.J. Sullivan, J.M. Eldridge, J.B. Knaak, M.J. Tallant, '5,6-Dihydro-5,6-dihydroxycarbaryl Glucuronide as a Significant Metabolite of Carbaryl in the Rat', *J. Agric. Food Chem.*, **120**(5), 980–985 (1972).
85. Y. Pico, C. Albelda, J.C. Molto, G. Font, J. Manes, 'Aldicarb Residues in Citrus Soil, Leaves and Fruits', *Food Add. Contam.*, **7**(Suppl. 1), S29–34 (1990).
86. L.G.M.Th. Tuinstra, F.R. Povel, A.H. Roos, 'Multi-matrix-multi-pesticide Method for Agricultural Products', *J. Chromatogr.*, **552**(1), 259–264 (1991).
87. L.G.M.Th. Tuinstra, A.H. Roos, A.M. Master, W.A. Traag, J.A. van Rhijn, 'Development of a Multi-residue/Multi-matrix Method for Pesticide Analysis in Agricultural Products', *Fresenius' J. Anal. Chem.*, **339**, 384–386 (1991).
88. R.C. Hall, P.E. Harris, 'Direct Gas Chromatographic Determination of Carbamate Pesticides Using Carbowax™ 20M-modified Supports and the Electrolytic Conductivity Detector', *J. Chromatogr.*, **169**, 245–259 (1979).
89. L. Muszkat, N. Aharonson, 'GC/CI/MS Analysis of Aldicarb, Butocarboxime, and Their Metabolites', *J. Chromatogr. Sci.*, **21**(4), 411–414 (1983).
90. S. Dagan, A. Amirev, 'Fast, Very Fast, and Ultra Fast Gas Chromatography/Mass Spectrometry of Thermally Labile Steroids, Carbamates, and Drugs in Supersonic Molecular Beams', *J. Am. Soc. Mass Spectrom*, **7**(8), 737–752 (1996).
91. M.L. Trehy, R.A. Yost, J.J. McCreary, 'Determination of Aldicarb, Aldicarboxime, and Aldicarbnitrile in Water by Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **56**(8), 1281–1285 (1984).
92. M.L. Trehy, R.A. Yost, J.G. Dorsey, 'Short Open Tubular Columns in Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **58**(1), 14–19 (1986).
93. R.S. Bakowski, M. Sher Ali, J.D. White, E.T. Phillippo, R.L. Ellis, 'Gas Chromatographic/Mass Spectrometric Confirmation of Ten *N*-Methylcarbamate Insecticides in Liver', *J. Assoc. Off. Anal. Chem.*, **77**(6), 1568–1574 (1994).
94. H.W. Dorough, J.H. Torstenson, 'Analysis for Carbamate Insecticides and Metabolites. (Review)', *J. Chromatogr. Sci.*, **13**(5), 212–224 (1975).
95. C.M. Sparacino, J.W. Hines, 'High Performance Liquid Chromatography of Carbamate Pesticides', *J. Chromatogr. Sci.*, **14**(11), 549–556 (1976).
96. J.F. Lawrence, D. Turton, 'High Performance Liquid Chromatographic Data for 166 Pesticides', *J. Chromatogr.*, **159**(1), 207–226 (1978).

97. P. Cabras, L. Spanedda, C. Tuberoso, M. Gennari, 'Separation of Pirimicarb and Its Metabolites by High Performance Liquid Chromatography', *J. Chromatogr.*, **478**(1), 250–254 (1989).
98. W. Cochrane, M. Lanouette, 'High Pressure Chromatography Determination of Aldicarb, Aldicarb Sulfoxide, and Aldicarb Sulfone in Potatoes', *J. Assoc. Off. Anal. Chem.*, **64**(3), 724–728 (1981).
99. W. Cochrane, M. Lanouette, S. Trudeau, 'Determination of Aldicarb, Aldicarb Sulfoxide, Aldicarb-sulfone and Carbofuran Residues in Water Using High Performance Liquid Chromatography', *J. Chromatogr.*, **243**(1), 307–314 (1982).
100. J. Spalik, G.E. Janauer, M. Lau, A.T. Lemley, 'Rapid Twostep Preconcentration Procedure for Aldicarb Determination in Water by High-performance Liquid Chromatography Using a 254-nm Detector', *J. Chromatogr.*, **253**(1), 289–294 (1982).
101. L.-Y. Lin, W.T. Cooper, 'Isocratic High Performance Liquid Chromatographic Separation and Multi-wavelength Ultraviolet Detection of Aldicarb and Its Soil Degradation Products. Optimization of Stationary Phase Selectivity', *J. Chromatogr.*, **390**(1–2), 285–295 (1987).
102. C.J. Miles, J.J. Delfino, 'Determination of Aldicarb and Its Derivatives in Groundwater by High Performance Liquid Chromatography With UV Detection', *J. Chromatogr.*, **299**(1), 275–280 (1984).
103. K.C. Ting, P.K. Kho, 'High Performance Liquid Chromatographic Method for the Determination of Aldicarb Sulfoxide in Watermelon', *Bull. Environ. Contam. Toxicol.*, **37**(2), 192–198 (1986).
104. S.S. Yang, I. Smetena, 'Determination of Aldicarb, Aldicarb Sulfoxide and Aldicarb Sulfone in Tobacco Using High-performance Liquid Chromatography with Dual Post-column Reaction and Fluorescence Detection', *J. Chromatogr. A*, **664**(2), 289–294 (1994).
105. M. Honing, D. Barceló, R.T. Ghijsen, U.A.Th. Brinkman, 'Optimization of the Liquid Chromatographic Separation of Pirimicarb and Its Metabolites V–VII: Application to a Soil Sample Used as a Candidate Reference Material', *Anal. Chim. Acta*, **286**(4), 457–468 (1994).
106. J.F. Lawrence, 'Direct Analysis of Some Carbamate Pesticides in Foods by High Pressure Liquid Chromatography', *J. Agric. Food Chem.*, **25**(1), 211–212 (1977).
107. J.F. Lawrence, R. Leduc, 'Direct Analysis of Carbofuran and Two Nonconjugated Metabolites in Crops by High Performance Liquid Chromatography With UV Absorption Detection', *J. Agric. Food Chem.*, **25**(6), 1362–1365 (1977).
108. J.E. Farrow, R.A. Hoodless, M. Sargent, J.A. Sidwell, 'Fungicide Residues. Part VI. Determination of Residues of Post-harvest Fungicides on Citrus Fruit by High Performance Liquid Chromatography', *Analyst*, **102**, 752–758 (1977).
109. R.A. Hoodless, J.A. Sidwell, J.C. Skinner, R.D. Treble, 'Application of High Performance Liquid Chromatography to the Determination of Pesticides Included in the European Economic Community Directive on Fruit and Vegetables', *J. Chromatogr.*, **166**(1), 279–286 (1978).
110. T.R. Nelsen, R.F. Cook, 'Determination of Carbofuran and 3-Hydroxycarbofuran Residues in Plant Tissue by Nitrogen Selective Gas Chromatography', *J. Agric. Food Chem.*, **28**(1), 98–102 (1980).
111. Y.W. Lee, N.D. Westcott, 'Direct Analysis of Carbofuran and 3-Hydroxycarbofuran in Rape Plants by Reversed-phase High-pressure Liquid Chromatography', *J. Agric. Food Chem.*, **28**(4), 719–722 (1980).
112. M. Chiba, D.F. Veres, 'High Performance Liquid Chromatographic Method for Simultaneous Determination of Residual Benomyl and Methyl 2-Benzimidazole Carbamate on Apple Foliage Without Clean-up', *J. Assoc. Off. Anal. Chem.*, **63**(6), 1291–1295 (1980).
113. G. Zweig, R. Gao, 'Determination of Benomyl by Reversed-phase Liquid Chromatography', *Anal. Chem.*, **55**(8), 1448–1451 (1983).
114. K.C. Ting, P.K. Kho, A.S. Musselman, G.A. Root, G.R. Tichelaar, 'High Performance Liquid Chromatographic Method for Determination of Six *N*-Methylcarbamates in Vegetables and Fruits', *Bull. Environ. Contam. Toxicol.*, **33**(5), 538–547 (1984).
115. P.C. Bardalaye, W.B. Wheeler, 'Simplified Method for the Clean-up and Reversed-phase High-Performance Liquid Chromatographic Determination of Benomyl in Mangoes', *J. Chromatogr.*, **330**(2), 403–407 (1985).
116. Th.J. Spierenburg, M.B.H. Kemmeren-van Dijk, P.E.F. Zoun, 'High-performance Liquid Chromatographic Analysis of Aldicarb in the Stomach of Birds of Prey', *J. Chromatogr.*, **393**(1), 137–139 (1987).
117. S.A. Ward, G. May, R.A. Branch, 'Quantitative Determination of Carbaryl and Carbaryl Metabolites by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **388**(2), 462–466 (1987).
118. R.J. Bushway, 'High Performance Liquid Chromatographic Determination of Carbaryl in Fruit Juices', *J. Chromatogr.*, **457**, 437–441 (1988).
119. J.G. Brayan, P.R. Haddad, G.J. Sharp, S. Dilli, 'Determination of Organophosphate Pesticides and Carbaryl on Paddy Rice by Reversed Phase High Performance Liquid Chromatography', *J. Chromatogr.*, **447**(1), 249–255 (1988).
120. M. Lanouette, R.K. Pike, 'Analysis of Aminocarb Formulations by High Performance Liquid Chromatography', *J. Chromatogr.*, **190**(1), 208–211 (1980).
121. J.L. Anderson, D.J. Chesney, 'Liquid Chromatographic Determination of Selected Carbamate Pesticides in Water With Electrochemical Detection', *Anal. Chem.*, **52**(13), 2156–2161 (1980).
122. G. Chiavari, V. Concialini, P. Vitali, 'High Performance Liquid Chromatography of Some Hydroxy-substituted Aromatic Compounds', *J. Chromatogr.*, **249**(2), 385–392 (1982).

123. J.L. Anderson, K.K. Whiten, J.D. Brewster, T-Y. Ou, W.K. Nonidez, 'Microarray Electrochemical Flow Detectors at High Applied Potentials and Liquid Chromatography With Electrochemical Detection of Carbamate Pesticides in River Water', *Anal. Chem.*, **57**(7), 1366–1373 (1985).
124. M.B. Thomas, P.E. Sturrock, 'Determination of Carbamates by High Performance Liquid Chromatography With Electrochemical Detection Using Pulsed-potential Cleaning', *J. Chromatogr.*, **357**(2), 318–324 (1986).
125. Q.G. van Nehring, J.W. Hightower, J.L. Anderson, 'Liquid Chromatography With Electrochemical Detection and Coloumetric Investigations of Carbamate and Urea Pesticides', *Anal. Chem.*, **58**(13), 2777–2781 (1986).
126. S. Kawai, K. Goto, K. Kano, T. Kubota, 'Determination of Carbaryl by High Performance Liquid Chromatography With Electrochemical Detection', *J. Chromatogr.*, **442**, 451–454 (1988).
127. R.T. Krause, 'High-performance Liquid Chromatographic Determination of Aryl *N*-Methylcarbamate Residues Using Post-column Hydrolysis Electrochemical Detection', *J. Chromatogr.*, **442**(2), 333–343 (1988).
128. L.H. Leenheers, R. Engel, W.E.T. Spruit, W.J.A. Meuling, M.J.M. Jongen, 'Determination of 5-Hydroxy-2-benzimidazole Carbamate in Urine by High Performance Liquid Chromatography With Electrochemical Detection', *J. Chromatogr.*, **613**(1), 89–94 (1993).
129. R.J. Argauer, J.D. Warthen, Jr, 'Separation of 1- and 2-Naphthols and Determination of Trace Amounts of 2-Naphthyl Methylcarbamate in Carbaryl Formulations by High Pressure Liquid Chromatography With Confirmation by Spectrofluorometry', *Anal. Chem.*, **47**(14), 2472–2474 (1975).
130. R.W. Frei, J.F. Lawrence, J. Hope, R.M. Cassidy, 'Analysis of Carbamate Insecticides by Fluorigenic Labelling and High-speed Liquid Chromatography', *J. Chromatogr. Sci.*, **12**(1), 40–44 (1974).
131. J.F. Lawrence, R. Leduc, 'High Pressure Liquid Chromatography With Ultraviolet Absorbance or Fluorescence Detection of Carbaryl in Potato and Corn', *J. Assoc. Off. Anal. Chem.*, **61**(4), 872–876 (1978).
132. H.A. Moye, S.J. Scherer, P.A. St John, 'Dynamic Fluorogenic Labeling of Pesticides for High Performance Liquid Chromatography: Detection of *N*-Methylcarbamates with *O*-phthalaldehyde', *Anal. Lett.*, **10**, 1049 (1977).
133. R.T. Krause, 'Multiresidue Method for Determining *N*-Methylcarbamate Insecticides in Crops, Using High-performance Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **63**(5), 1114–1124 (1980).
134. R.T. Krause, 'Determination of Fluorescent Pesticides and Metabolites by Reversed Phase High Performance Liquid Chromatography', *J. Chromatogr.*, **255**(2), 497–510 (1983).
135. R.T. Krause, 'Liquid Chromatographic Determination of *N*-Methylcarbamate Insecticides and Metabolites in Crops. I. Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **68**(4), 726–733 (1985).
136. H. Engelhardt, N. Lillig, 'Optimization of a Reaction Detector for Analysis at the Femtomol Level of Carbamate Insecticides by HPLC', *Chromatographia*, **21**(3), 136–142 (1986).
137. L. Nondek, R.W. Frei, U.A.Th. Brinkman, 'Heterogeneous Catalytic Post-column Reaction Detectors for High Performance Liquid Chromatography. Application to *N*-Methylcarbamates', *J. Chromatogr.*, **282**(1), 141–150 (1983).
138. L. Nondek, U.A.Th. Brinkman, R.W. Frei, 'Band Broadening in Solid-phase Reactors Packed With Catalyst for Reactions in Concentration-flow Systems', *Anal. Chem.*, **55**(9), 1466–1470 (1983).
139. L.K. She, U.A.Th. Brinkman, R.W. Frei, 'Liquid Chromatographic Residue Analysis of Carbaryl Based on a Post-column Catalytic Reactor Principle and Fluorogenic Labeling', *Anal. Lett.*, **17**, 915 (1984).
140. A. de Kok, M. Hiemstra, C.P. Vreeker, 'Optimization of the Post-column Hydrolysis Reaction on Solid-phases for the Routine High Performance Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides in Food Products', *J. Chromatogr.*, **507**, 459–472 (1990).
141. H. Jansen, U.A.Th. Brinkman, R.W. Frei, 'Miniaturization of Solid-phase Reactors for On-line Post-column Derivatization in Narrow-bore Liquid Chromatography', *Chromatographia*, **20**(3), 453–462 (1985).
142. C.J. Miles, H.A. Moye, 'Postcolumn Photolysis of Pesticides for Fluorometric Determination by High Performance Liquid Chromatography', *Anal. Chem.*, **60**(3), 220–226 (1988).
143. B.D. McGarvey, 'Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides Using a Single-stage Post-column Derivatization Reaction and Fluorescence Detection', *J. Chromatogr.*, **481**(2), 445–451 (1989).
144. V.A. Simon, K.S. Pearson, A. Taylor, 'Determination of *N*-Methylcarbamates and *N*-Methylcarbamoyloximes in Water by High Performance Liquid Chromatography With the Use of Fluorescence Detection and a Single *O*-Phthalaldehyde Post-column Reaction', *J. Chromatogr.*, **643**(2), 317–320 (1993).
145. H.A. Moye, C.J. Miles, S.J. Scherer, 'A Simplified High Performance Liquid Chromatographic Residue Procedure for the Determination of Glyphosate Herbicide and (aminomethyl)Phosphonic Acid in Fruits and Vegetables Employing Postcolumn Fluorogenic Labeling', *J. Agric. Food Chem.*, **31**(1), 69–72 (1983).
146. M. deBernardinis, Jr, W.A. Wargin, 'High-performance Liquid Chromatographic Determination of Carbaryl and 1-Naphthol in Biological Fluids', *J. Chromatogr.*, **246**(1), 89–94 (1982).
147. K.M. Hill, R.H. Hollowell, L.A. dal Cortivo, 'Determination of *N*-Methylcarbamate Pesticides in Well Water

by Liquid Chromatography With Post Column Fluorescence Derivatization', *Anal. Chem.*, **56**(13), 2465–2468 (1984).

148. F.I. Onuska, *Mass Spectrometry in Environmental Sciences*, eds. F.W. Karasek, O. Hutzinger, S. Safe, Plenum Press, New York, 367–404, 1985.
149. T. Cairns, E.G. Siegmund, J.J. Stamp, 'Chemical Ionization Mass Spectrometry of Carbamates: 1-*N*-Phenylcarbamates-Propham and Chloropropham', *Biomed. Environ. Mass Spectrom.*, **11**(6), 301–308 (1984).
150. T. Cairns, E.G. Siegmund, J.J. Stamp, 'A Basic Mechanistic Fragmentation Concept in Chemical Ionization', *Org. Mass Spectrom.*, **21**(3), 161–162 (1986).
151. J.J. Stamp, E.G. Siegmund, T. Cairns, K.K. Chan, 'Chemical Ionization Mass Spectrometry of Carbamate Pesticides: A Major Dissociation Pathway', *Anal. Chem.*, **58**(4), 873–881 (1986).
152. D.R. Hunt, J. Shabanowitz, T.M. Harvey, M. Coates, 'Scheme for the Direct Analysis of Organics in the Environment by Tandem Mass Spectrometry', *Anal. Chem.*, **57**(6), 525 (1985).
153. M.J.I. Mattina, L.Q. Huang, 'Adduct Ion Formation in the Methane Enhanced Negative Ion Mass Spectrometry of Aryloxy-*N*-Methoxycarbamates', *Org. Mass Spectrom.*, **24**, 360 (1989).
154. B.C. Lynn, Jr, G.D. Marbury, J.R. Tuschall, Jr, 'Fragment-molecule Adduct Formation from Oxime Carbamates During Negative Ion Chemical Ionization Mass Spectrometry', *Org. Mass Spectrom.*, **23**, 736 (1988).
155. B.C. Lynn, Jr, G.D. Marbury, J.R. Tuschall, Jr, Proc. 36th. ASMS Conference, San Fransisco, June 5–10, 908, 1988.
156. R.P. Singh, I.D. Brindle, T.R.B. Jones, J.M. Miller, M. Chiba, 'Fast-atom Bombardment Mass Spectrometric Determination of Methyl[1-(butylcarbamoyl)-1*H*-Benzimidazol-2-yl]carbamate (Benomyl) in Wettable Powder Formulations', *Rapid Commun. Mass Spectrom.*, **7**(9), 167–171 (1993).
157. T. Cairns, E.G. Siegmund, G.M. Doose, 'Liquid Chromatography Mass Spectrometry of Thermally Labile Pesticides', *Biomed. Environ. Mass Spectrom.*, **10**(1), 24–29 (1983).
158. R.D. Voyksner, J.T. Bursey, 'Modification and Optimization of Source and Interface Parameters for Direct Liquid Introduction Liquid Chromatography/Mass Spectrometry of Pesticides', *Anal. Chem.*, **56**(9), 1582–1587 (1984).
159. R.D. Voyksner, J.T. Bursey, E.D. Pellizzari, 'Analysis of Selected Pesticides by High Performance Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **312**(1), 221–235 (1984).
160. T.D. Behymer, T.A. Bellar, W.L. Budde, 'Liquid Chromatography/Particle Beam/Mass Spectrometry of Polar Compounds of Environmental Interest', *Anal. Chem.*, **62**, 1686 (1990).
161. T.A. Bellar, T.D. Behymer, W.L. Budde, 'Investigation of Enhanced Ion Abundances From a Carrier Process in High Performance Liquid Chromatography Particle Beam Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **1**(1), 92–98 (1989).
162. A. Cappiello, G. Famiglini, F. Bruner, 'Determination of Acidic and Base-neutral Pesticides in Water With a New Micro Flow Rate LC/MS Particle Beam Interface', *Anal. Chem.*, **66**, 1416 (1994).
163. J. Slobodnik, S.J.F. Hoekstra-Oussoren, M.E. Jager, M. Honing, B.L.M. van Baar, U.A.Th. Brinkman, 'Online Solid-phase Extraction Liquid Chromatography Particle Beam Mass Spectrometry and Gas Chromatography Mass Spectrometry of Carbamate Pesticides', *Analyst*, **121**(9), 1327–1334 (1996).
164. J. Slobodnik, M.E. Jager, S.J.F. Hoekstra-Oussoren, M. Honing, B.L.M. van Baar, U.A.Th. Brinkman, 'Identification of Carbamates by Particle Beam/Mass Spectrometry', *J. Mass Spectrom.*, **32**(1), 43–54 (1997).
165. M. Honing, D. Barceló, M.E. Jager, J. Slobodnik, B.L.M. van Baar, U.A.Th. Brinkman, 'Effect of Ion Source Pressure on Ion Formation of Carbamates in Particle Beam Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **712**(1), 21–30 (1995).
166. S. Pleasance, J.F. Anacleto, M.R. Bailey, D.H. North, 'An Evaluation of Atmospheric Pressure Ionization Techniques for the Analysis of *N*-Methylcarbamate Pesticides by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **3**(3), 378–391 (1992).
167. R.D. Voyksner, J.T. Bursey, E.D. Pellizzari, 'Post-column Addition of Buffer for Thermospray Liquid Chromatography/Mass Spectrometry. Identification of Pesticides', *Anal. Chem.*, **56**(8), 1507–1514 (1984).
168. T. Cairns, E.G. Siegmund, J.J. Stamp, 'Thermospray Mass Spectrometry of Aldicarb and its Metabolites', *Rapid Commun. Mass Spectrom.*, **1**, 89 (1987).
169. J. Saar, A. Salomon, 'Thermospray Mass Spectrometry of *N*-Methylcarbamates', *Org. Mass Spectrom.*, **25**, 209 (1990).
170. R.D. Voyksner, 'Characterization of Dyes in Environmental Samples by Thermospray High Performance Liquid Chromatography Mass Spectrometry', *Anal. Chem.*, **57**(13), 2600–2605 (1985).
171. D. Barceló, G. Durand, R.J. Vreeken, G.J. de Jong, H. Lingeman, U.A.Th. Brinkman, 'Evaluation of Eluents in Thermospray Liquid Chromatography–Mass Spectrometry for Identification and Determination of Pesticides in Environmental Samples', *J. Chromatogr.*, **553**(1–2), 311–328 (1991).
172. G. Durand, N. de Bertrand, D. Barceló, 'Mobile-phase Variations in Thermospray Liquid Chromatography–Mass Spectrometry of Pesticides', *J. Chromatogr.*, **562**(1–2), 507–523 (1991).
173. G. Durand, N. de Bertrand, D. Barceló, 'Applications of Thermospray Liquid Chromatography Mass

Spectrometry in Photochemical Studies of Pesticides in Water', *J. Chromatogr.*, **554**(1–2), 233–250 (1991).
174. D. Volmer, A. Preis, K. Levsen, G. Wünsch, 'Thermospray Mass Spectral Studies of Pesticides. Temperature and Salt Concentration Effects on the Ion Abundances in Thermospray Mass Spectra', *J. Chromatogr.*, **647**(1–2), 235–259 (1993).
175. D. Volmer, K. Levsen, G. Wünsch, 'Thermospray Liquid Chromatographic–Mass Spectrometric Multiresidue Determination of 128 Polar Pesticides in Aqueous Environmental Samples', *J. Chromatogr.*, **660**(1–2), 231–248 (1993).
176. R.J. Vreeken, W.D. van Dongen, R.T. Ghijsen, U.A.Th. Brinkman, 'The Use of Mobile Phase Additives in the Determination of 55 (polar) Pesticides by Column Liquid Chromatography–Thermospray–Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **54**(2), 119–128 (1994).
177. M. Honing, D. Barceló, B.L.M. van Baar, R.T. Ghijsen, U.A.Th. Brinkman, 'Ion Formation of *N*-Methylcarbamates in Thermospray Mass Spectrometry: The Effects of Additives to the Liquid Chromatographic Eluent and of the Vaporizer Temperature', *J. Am. Soc. Mass Spectrom.*, **5**(6), 913–927 (1994).
178. D. Volmer, K. Levsen, 'Mass Spectrometric Analysis of Nitrogen- and Phosphorus Containing Pesticides by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**(7), 655–675 (1994).
179. J. Slobodnik, B.L.M. van Baar, U.A.Th. Brinkman, 'Column Liquid Chromatography/Mass Spectrometry: Selected Techniques in Environmental Applications for Polar Pesticides and Related Compounds', *J. Chromatogr.*, **703**(1), 81–90 (1995).
180. M. Honing, D. Barceló, B.L.M. van Baar, R.T. Ghijsen, U.A.Th. Brinkman, 'Ion Formation of *N*-Methylcarbamate Pesticides in Thermospray Mass Spectrometry: The Effects of Additives to the Liquid Chromatographic Eluent and of the Vaporizer Temperature', *J. Am. Soc. Mass Spectrom.*, **5**(8), 913–927 (1994).
181. D. Volmer, K. Levsen, M. Honing, D. Barceló, J. Abian, E. Gelpi, B.L.M. van Baar, U.A.Th. Brinkman, 'Comparative Study of Different Thermospray Interfaces With Carbamate Pesticides: Influence of the Ion Source Geometry', *J. Am. Soc. Mass Spectrom.*, **6**(7), 656–667 (1995).
182. H-Y. Lin, R.D. Voyksner, 'Determination of Environmental Contaminants Using a Electrospray Interface Combined With an Iontrap Mass Spectrometer', *Anal. Chem.*, **65**(4), 451–456 (1993).
183. R.D. Voyksner, T. Pack, 'Investigation of Collisional-activation Decomposition Process and Spectra in the Transport Region of an Electrospray Single-quadrupole Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **5**, 263 (1991).
184. G. Hopfgartner, T. Wachs, K. Bean, J. Henion, 'High-flow Ion Spray Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **65**(4), 439–446 (1993).
185. G. Hopfgartner, K. Bean, J. Henion, R. Henry, 'Ion Spray Mass Spectrometric Detection for Liquid Chromatography: A Concentration- or a Mass-flow-sensitive Device?', *J. Chromatogr.*, **647**(1), 51–61 (1993).
186. M. Honing, J. Riu, D. Barceló, B.L.M. van Baar, U.A.Th. Brinkman, 'Determination of Ten Carbamate Pesticides in Aquatic and Sediment Samples by Liquid Chromatography Ionspray and Thermospray Mass Spectrometry', *J. Chromatogr.*, **733**(1), 283–294 (1996).
187. S. Chiron, J.A. Torres, A. Fernandez-Alba, D. Barceló, 'Identification of Carbofuran and Methiocarb and Their Transformation Products in Estaurine Waters by On-line Solid Phase Extraction Liquid Chromatography Mass Spectrometry', *Int, J. Environ. Anal. Chem.*, **65**(1–4), 37–52 (1996).
188. D. Giraud, A. Ventura, V. Camel, A. Bermond, P. Arpino, 'Determination of Traces of Pesticides in Water by Solid-phase Extraction and Liquid Chromatography Ionspray Mass Spectrometry', *J. Chromatogr. A*, **777**(1), 115–125 (1997).
189. D.R. Doerge, S. Bajic, 'Analysis of Pesticides Using Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**(9), 663–666 (1992).
190. S. Kawasaki, F. Naguno, H. Ueda, Y. Fajima, M. Sano, J. Tadano, 'Simple, Rapid and Simultaneous Measurement of Eight Different Types of Carbamate Pesticides in Serum Using Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr.*, **620**(1), 61–71 (1993).
191. J.A. Schoemaker, T.D. Behymer, J.W. Eichelberger, W.L. Budde, 'Evaluation of C_{18}/OH Liquid–Solid Extraction and Atmospheric Pressure Ionization Liquid Chromatography/Mass Spectrometry for the Determination of Nonvolatile Organic Analytes in Aqueous Matrices', Abstracts 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, 257, 1994.
192. T.D. Behymer, D.D. Kryak, W.L. Budde, 'Quantitative Aspects of Atmospheric Pressure Chemical Ionization Liquid Chromatography/Mass Spectrometry for the Measurement of Compounds of Environmental Interest', Abstracts 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, IL, 325, 1994.
193. R.D. Voyksner, 'Atmospheric Pressure Ionization LC/MS. New Solutions for Environmental Analysis', *Environ. Sci. Technol.*, **28**(3), 118A–127A (1994).
194. C. Aguilar, I. Ferrer, F. Borrull, R.M. Marce, D. Barceló, 'Comparison of Automated On-line Solid-phase Extraction Followed by Liquid Chromatography Mass Spectrometry With Atmospheric Pressure Chemical Ionization and Particle Beam Mass Spectrometry for the Determination of a Priority Group of Pesticides in Environmental Waters', *J. Chromatogr. A*, **794**(1–2), 147–163 (1998).

195. A.C. Hogenboom, J. Slobodnik, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Single Short Column Liquid Chromatography With Atmospheric Pressure Chemical Ionization/(tandem) Mass Spectrometric Detection for Trace Environmental Analysis', *Chromatographia*, **42**(9/10), 506–514 (1996).

196. K.A. Barnes, R.J. Fussell, J.R. Startin, M.K. Pegg, S.A. Thorpe, S.L. Reynolds, 'High Performance Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry With Ionization Polarity Switching for the Determination of Selected Pesticides', *Rapid Commun. Mass Spectrom.*, **11**(1), 117–123 (1997).

197. J.F. Banks, S. Shen, G.M. Whitehouse, J.B. Fenn, 'Ultrasonically Assisted Electrospray Ionization for LC/MS Determination of Nucleotides From a Transfer RNA Digest', *Anal. Chem.*, **66**(3), 406–414 (1994).

198. M.G. Ikonomou, P. Kebarle, 'A Heated Electrospray Source for Mass Spectrometry of Analytes From Aqueous Solutions', *J. Am. Soc. Mass Spectrom.*, **5**(9), 791–800 (1994).

199. A.P. Bruins, T.R. Covey, J. Henion, 'Ion Spray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 2642 (1987).

200. J. Slobodnik, 'Automation of Hyphenated Liquid and Gas Chromatographic Systems in Environmental Analysis', PhD Thesis, Amsterdam, 1996.

201. C.S. Creaser, J.W. Stygall, 'Particle Beam Liquid Chromatography/Mass Spectrometry: Instrumentation and Applications', *Analyst*, **118**(10), 1467–1480 (1993).

202. A. Cappiello, 'Is Particle Beam an up-to-date LC/MS Interface? State of the Art and Perspectives', *Mass Spectrom. Rev.*, **15**, 283–296 (1996).

Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of

Arthur W. Garrison
US Environmental Protection Agency, Athens, USA

1 Introduction 1
2 Analysis of Chiral Pesticides and Polychlorinated Biphenyls 3
2.1 Chiral Separations by Gas Chromatography 3
2.2 Chiral Separations by High-performance Liquid Chromatography 5
2.3 Chiral Separations by Capillary Electrophoresis 7
3 Applications of Enantiomeric Separations to the Environmental Chemistry of Pesticides and Polychlorinated Biphenyls 9
Acknowledgments 10
Abbreviations and Acronyms 10
Related Articles 11
References 11

Over 25% of pesticides and other toxic organic pollutants are chiral, as are 19 of the 209 polychlorinated biphenyl (PCB) congeners; that is, they exist as two mirror image species called enantiomers (PCB enantiomers are called atropisomers). The enantiomers of a chiral compound have identical physical and abiotic chemical properties, but usually differ in biological properties such as microbial degradation, uptake and transport across membranes, metabolism rate, and toxicity. Chirality in the environment has become recognized as an important phenomenon since the mid-1980s, and related research has centered on the need to understand the environmental fate and effects of pesticides and PCBs. This article provides a brief background on chiral chemistry in general and the importance of chirality in the drug industry, followed by a summary of examples of the enantioselectivity of pesticide and PCB occurrences and degradation in the environment as gleaned from the scientific literature. The main emphasis is on the analysis of these compounds, with the primary focus on chiral separations. Chiral separations are generally performed by gas chromatography (GC) or high-performance liquid chromatography (HPLC) using chiral columns, or, to a lesser extent, by capillary electrophoresis (CE) using cyclodextrin (CD) or other chiral selectors in the run buffer. Summaries of these three separation methods are provided, with figures giving examples of pesticide and PCB enantiomeric separations. Finally, to show the usefulness of these separation methods, applications to several environmental chemistry problems involving chiral pesticides and PCBs are described and illustrated.

1 INTRODUCTION

In 1860, Louis Pasteur observed that tartaric acid crystals, when observed under a microscope, existed in two forms; the crystal face of each form was the mirror image of the other. This was the first observation of the natural phenomenon known as *chirality*, which shortly gained recognition as a critical factor in biological chemistry. *Chiral* compounds are those that contain at least one center of asymmetry, which results in the occurrence of mirror image isomers called *enantiomers*. There are two main types of asymmetry – tetrahedral asymmetry about sp^3-bonded carbon or phosphorus atoms, in which case each of the four attached groups must be different, and axial asymmetry with sterically restricted rotation about, for example, a carbon–carbon or nitrogen–carbon bond (Scheme 1). Mirror image isomers resulting from axial asymmetry are called *atropisomers*. A more general name for mirror image isomers is *optical isomers,* because such isomers differ only in their direction of rotation of plane-polarized light. There are no other chemical or physical differences except for one critical exception – each enantiomer of a chiral compound reacts at a different rate with each enantiomer of another chiral molecule. As complex biochemical molecules are generally chiral, and usually exist as only one of the possible enantiomers, the enantiomers of active substrates such as pharmaceuticals and pesticides react with such biochemicals at different rates and, hence, the great recent interest in chiral drugs, and now in chiral pesticides.

A *racemate*, or racemic mixture, is an equal mixture of the enantiomers of a chiral compound; thus, a racemate is itself not optically active. Enantiomers were formerly designated as (d) and (l), for dextrorotatory and levorotatory, corresponding to the rotation of the plane of light to the right or to the left, respectively; this has been superseded by (+) and (−). The direction of rotation of light is incidental to the *absolute configuration* of the molecule, which is the actual arrangement in space of the groups around the asymmetric center and is the real determinant of the molecule's chiral properties (see the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

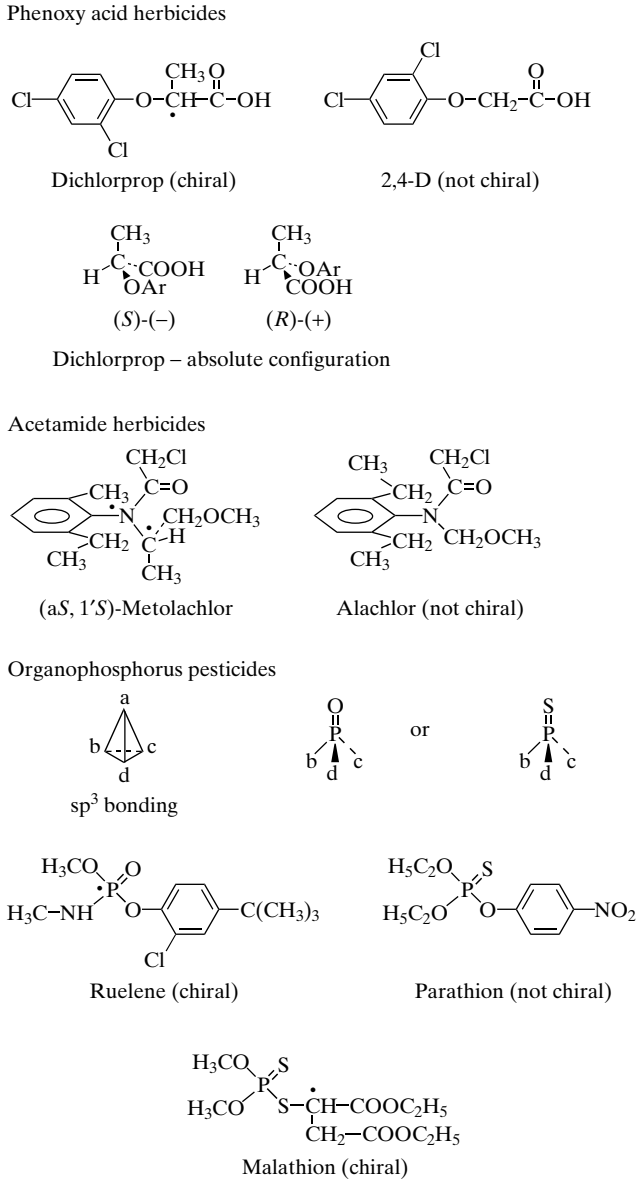

Scheme 1 Structures of representative chiral and achiral pesticides. Notice: (1) the absolute configuration of the dichlorprop enantiomers; (2) that organophosphorus compounds can be chiral because of asymmetry about a carbon atom or the phosphorus atom; and (3) the two centers of asymmetry in metolachlor, one because of asymmetry about a carbon atom and the other because of sterically restricted rotation about a nitrogen to carbon bond. 2,4-D, 2,4-dichlorophenoxy acetic acid.

example in Scheme 1). The absolute configuration about a chiral center is designated (*R*) or (*S*) (sometimes D and L) according to established rules of spatial arrangement. If a molecule has more than one center of asymmetry, there will exist more than two enantiomers, and it is necessary to designate the absolute configuration of each for a complete description of the molecule's chirality. A complete description of a chiral molecule, or of an enantiomer, entails notation of both the absolute configuration (if known) and the direction of light rotation, such as (*R*)-(+) and (*S*)-(−). It is fairly common, however, to simply refer to a particular enantiomer as the (−) or (+) species to distinguish one from the other. More information about chirality is given in the comprehensive treatise on the subject by Eliel et al.[1]

The same factors that make chirality so important in the interactions of drugs with complex biological materials in the human body are also effective in pesticide interactions with receptors. Until the mid to late 1980s, most chiral drugs were marketed as the racemate, such as is produced in typical synthetic processes. However, there are many examples showing that single-enantiomer drugs are more effective or less harmful than the racemic mixture.[2,3] The (+) enantiomer of ibuprofen, the racemate of which is used as an analgesic and muscle relaxant, is as effective as the racemate, without the side effects. Only the (−) enantiomers of barbiturates have the desired narcotic effects; the (+) enantiomers sometimes have convulsive properties. Methadone is used in drug-dependence treatment as a substitute for more harmful drugs such as cocaine; (−) methadone effects last for 72 h, whereas racemic methadone effects last for only 24 h. Racemic retalin, used in treatment of hyperactivity in children, has side effects of insomnia and appetite suppression; the (2*R*,3*R*) enantiomer avoids these side effects. With US Food and Drug Agency approval and encouragement to use the effective enantiomer, the production and marketing of single-isomer drugs has become an extremely important business. Of the 100 top-selling drugs worldwide, 50 are now single enantiomers with sales of $42.8 billion of the $85.2 billion total for the 100.[3]

Chirality in the environment has become recognized as an important phenomenon since the late 1980s,[4–6] and related research has centered on the need to understand the environmental fate and effects of pesticides and PCBs. As with drugs, chiral pesticides are usually *enantioselective* in their interactions with naturally occurring biological molecules; that is, one enantiomer will react faster than the other. The processes of microbial and other enzymatically catalyzed transformations, uptake into plants and animals, translocation across protein membranes, and metabolism in various organs are all likely to be enantioselective. This can lead to enantioselective occurrences in, for example, soils where enantioselective microbial degradation has depleted one enantiomer faster than the other.

It is important to remember, however, that abiotic processes are not inherently enantioselective. Each enantiomer of a chiral compound behaves identically in reactions involving abiotic hydrolysis, photolysis, reduction, oxidation, etc. because it is not reacting with other chiral molecules or systems.

Several examples of enantioselectivity in the environment have been published in the scientific literature, mostly in the mid to late 1990s. For example, it was shown in 1993[7] that there exists considerable diversity in the enantioselectivity of various enzymatic degradation pathways for α-hexachlorocyclohexane (α-HCH). Research in 1994 on the occurrence of toxaphene congeners in tissues of aquatic vertebrates showed enantioselectivity for several of the chiral congeners.[8] In 1995, researchers reported the rapid enantioselective degradation of α-HCH in Arctic lakes;[9] they later showed reversal in selectivity of the two enantiomers from one end of an Arctic sea to the other, indicating a difference in microbial populations.[10] In the same year, the enantioselective degradation of several acetamide herbicides was observed in sewage sludge and soil.[11] Enantioselectivity can also occur during human metabolism of toxic substances – it has been shown that the ratios of the atropisomers of PCB 132 ranged from 0.45 to 0.85 in human milk.[12] Later, experiments showed that 1,1-dichloro-2-(*o*-chlorophenyl)-2-(*p*-chlorophenyl) ethane (*o*,*p*′-DDD) residues in about 20 whole-body samples from three fish species were enriched in the (−)-enantiomer.[13] Most recently, the enantio-selective degradation of 1,1,1-trichloro-2-(*o*-chlorophenyl)-2-(*p*-chlorophenyl) ethane (*o*,*p*′-DDT), *cis*- and *trans*-chlordane, and heptochlor epoxide was demonstrated to occur in soils in the US cornbelt.[14] Finally, the enantioselective occurrence of the methylsulfonyl metabolites of several chiral PCB congeners was measured in polar bears and seals.[15]

In addition to these fate processes, the mode of action and effects of chiral pesticides and other anthropogenic compounds are generally enantioselective. For example, the (*R*) enantiomers of dichlorprop and mecoprop, important phenoxypropionic acid herbicides, are herbicidally active, whereas the (*S*) enantiomers are ineffective; however, the (*S*) enantiomers of metolachlor are the most active of its four enantiomers.[16] Finally, the (−) enantiomer of *o*,*p*′-DDT is much more active in endocrine disrupter tests than the (+) enantiomer, whereas both enantiomers of *o*,*p*′-DDD have low activity. More information on the synthesis and mode of action of enantiomers of chiral pesticides is given by Tombo and Belluš.[6]

Many pesticides and other environmental pollutants are chiral – 25% of the pesticides manufactured in 1995 were chiral compounds.[17] However, chiral pesticides are almost always manufactured and formulated as their racemic mixtures. Some individual chiral pesticides, as well as some pesticide classes containing several chiral members, are:

- phenoxypropionic acid herbicides
- *o*,*p*′-DDT and *o*,*p*′-DDD
- α-HCH
- *cis*- and *trans*-chlordane
- many toxaphene congeners
- 19 of the 209 PCB congeners (atropisomers)
- imidazolinone herbicides
- acetamide herbicides
- pyrethroid pesticides
- organophosphorus pesticides.

Scheme 1 gives the structures of some chiral pesticides, with some achiral compounds for comparison.

2 ANALYSIS OF CHIRAL PESTICIDES AND POLYCHLORINATED BIPHENYLS

Samples for analysis of chiral pesticides and PCBs may include air, surface water, groundwater, drinking water, bottom sediments, soils, and a variety of biological species. In the analysis of such samples, whether environmental or biological, the main objective is usually to determine the enantiomeric ratio (ER) of the chiral analyte(s). The first goal is the separation of the enantiomers; the second is the separation of both enantiomers from other sample components. These samples differ from conventional complex environmental or biological samples only in that special techniques and/or reagents are employed to achieve final separation of the enantiomers. An overview of chiral separations of pesticides is available.[16] Preparation of samples for instrumental analysis involves the same basic steps as required for any sample: extraction of the analyte(s) from the bulk matrix and clean-up of the extract to remove interfering materials. None of these operations can affect the ER of the chiral analyte because all such processes are abiotic and nonchiral, unless specifically designed for enantioselectivity (such as with chiral adsorbents in column chromatography).

Conventional analytical instruments are generally used for chiral analysis of prepared environmental and biological samples. Capillary GC and HPLC are by far the most commonly used instruments for chiral analysis of pesticides and PCBs, although CE is becoming more important because of its demonstrated effectiveness in separating drug enantiomers. Each of these instrumental techniques is addressed in the following sections, with illustrations of enantiomer separations.

2.1 Chiral Separations by Gas Chromatography

As in conventional capillary GC, volatile and semivolatile chiral analytes, including the PCB atropisomers and many pesticides, are separated as their vapors are passed across a stationary liquid phase. Separation of these analytes is based on their preferential partitioning between the gas

and liquid phases; however, separation of the enantiomers requires complementary, specific, steric interactions. This is usually accomplished by preferential association between each enantiomer and a chiral component of the liquid phase, to form two transient diastereomers that have different properties, and so are separable. Most of these chiral components, and the great majority of those used for chiral pesticide analysis, are CDs, which are either dissolved in a conventional liquid phase, usually a polysiloxane, or chemically linked to the polysiloxane backbone for added stability. Even more stable columns are now available; these have the chemically linked phases bonded to the fused silica column surface.

The CDs are cyclic molecules composed of six, seven, or eight linked carbohydrate moieties called α-, β-, or γ-CDs, respectively; these cyclic structures are in the shape of a truncated cone (Scheme 2). Because each of the enantiomers of a chiral compound has a different spatial configuration, there is often a preferred fit into the cone so that the two complexes formed between the two enantiomers and the CD are of different strengths. In addition, each carbohydrate moiety contains one primary and two secondary hydroxyl groups, any of which may form hydrogen bonds with the enantiomers; this is a secondary complexation mechanism that often enhances separation. Various derivatives of these CDs have been developed since the mid 1990s. The most common of these result from methylation of two or three of the hydroxyl groups on each carbohydrate moiety. Other derivatives include combinations of *O*-pentyl, *O*-acetyl, *O*-(*S*)-2-hydroxypropyl, and *O*-*tert*-butyldimethylsilyl substituents. Thus, there is currently available such a variety of CD-based capillary columns that the enantiomers of most gas chromatographable chiral compounds can be separated.

Vetter and Schurig[18] published a comprehensive article about GC analysis of chiral organochlorine

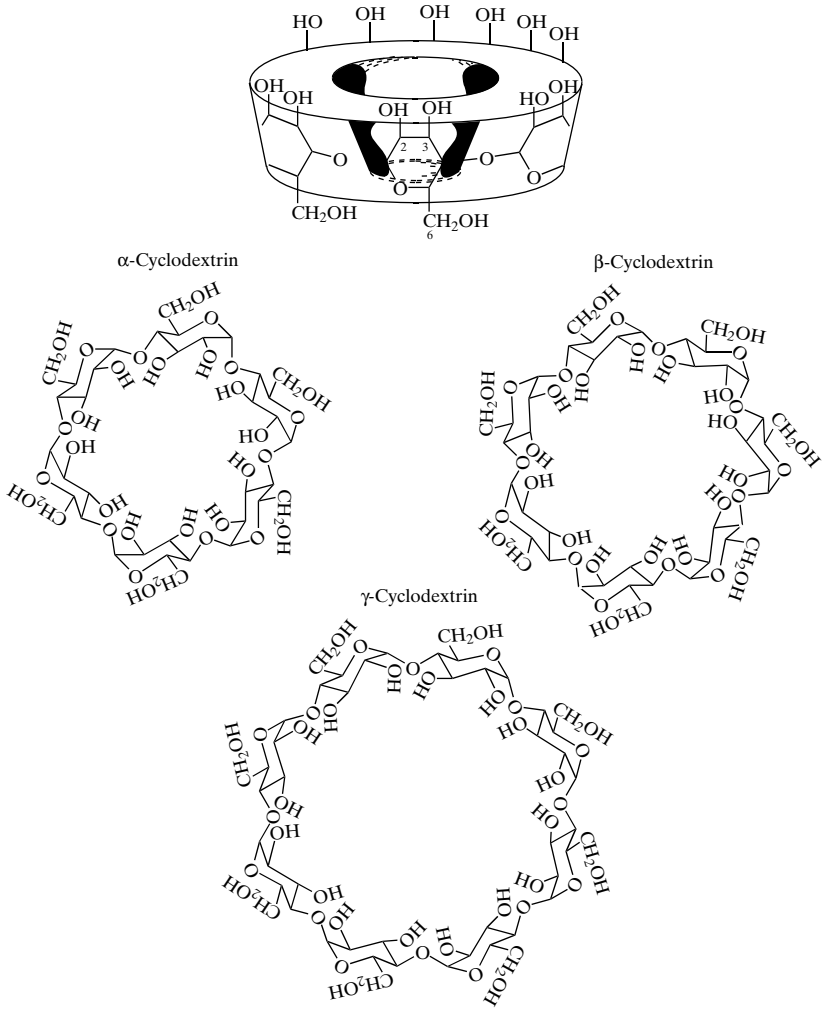

Scheme 2 Molecular model of α-CD (above) and the ring structures of α-, β-, and γ-CD.

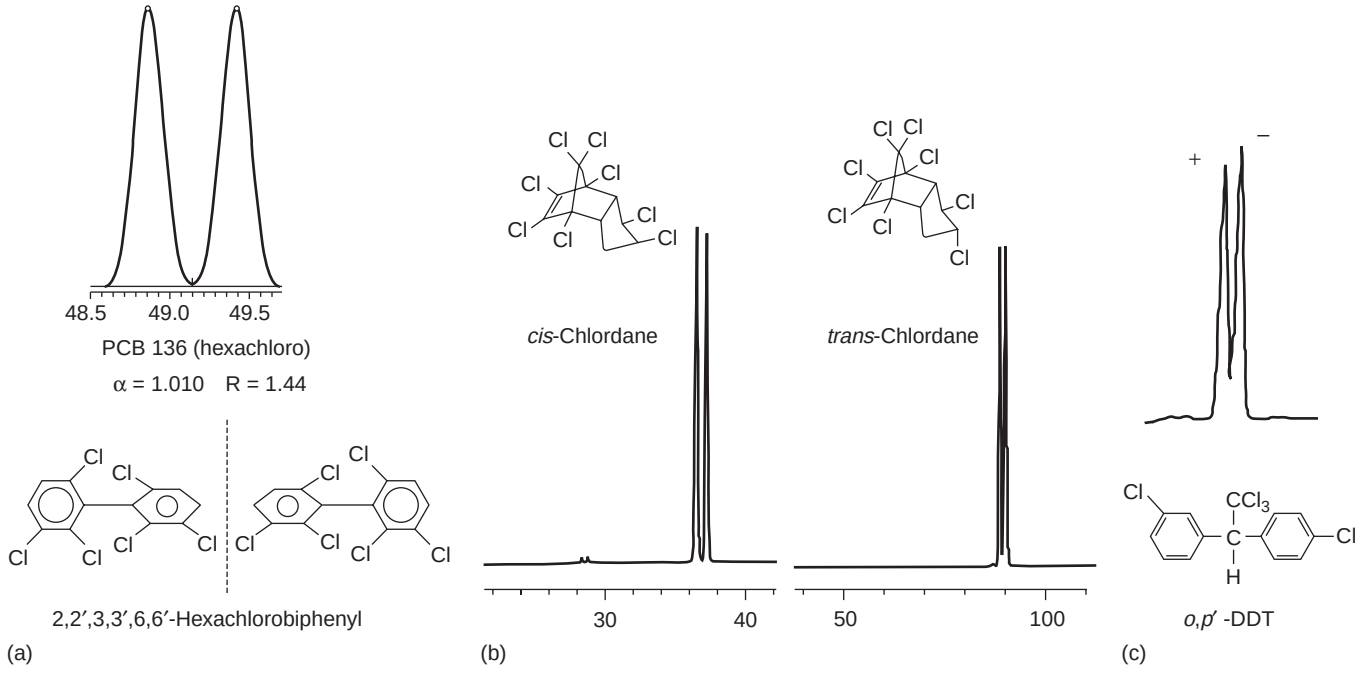

Figure 1 GC separations: (a) structures of PCB 136 atropisomers and their separation using a Chiraldex™ 20 m dimethyl-β-CD column; (b) *cis*- and *trans*-chlordane enantiomers using the same column; (c) *o*,*p*′-DDT enantiomers using a Chiraldex™ G-PT 12 m column with a permethyl trifluoroacetoxypropyl-γ-CD chiral phase in tandem with an achiral DB-5, 95% dimethyl : 5% diphenylpolysiloxane, type column. (Chiraldex™ is a trademark of Advanced Separation Technologies, Inc., Whippany, NJ, USA. This figure is supplied courtesy of the US Environmental Protection Agency.)

compounds, emphasizing separation by CD-based chiral phases and including examples of chiral separations. Other examples of chiral GC separations, including the enantiomers of *cis*- and *trans*-chlordane, *o*,*p*′-DDT, and a pair of PCB atropisomers, are given in Figure 1.

Detection of enantiomers after GC separation can be by any of the conventional GC detectors, depending on the properties of the chiral analyte. These detectors include mass spectrometers and infrared spectrophotometers for compound detection and identification.

2.2 Chiral Separations by High-performance Liquid Chromatography

The method of choice for chiral compounds that are not gas chromatographable is HPLC. Such compounds are ionic or polar and/or heat labile and include many pesticides and their metabolites. The HPLC separations are based on liquid–solid partitioning which, for chiral separations, may involve several mechanisms of complexation between solute (analyte) and a chiral stationary phase in an otherwise conventional column; separations may be by the reversed phase or normal phase mode. According to a common classification system, these chiral stationary phases are divided into five basic types, depending on the complexation mechanism. These mechanisms may include hydrogen bonding, dipole–dipole interactions, $\pi-\pi$ interactions, steric hindrance, ligand exchange with metal complexes, and formation of inclusion complexes within chiral cavities. In most cases, the chiral phase is covalently bound to the silica support to achieve greater column stability. Type I chiral stationary phases include the popular Pirkle phases, which are based on $\pi-\pi$ and other attractive interactions. Type III phases include the important phases based on CDs (section 2.1), which are usually covalently bonded to the silica through their primary hydroxyl groups. More recently, a variety of derivatives of these, made by reacting some of the secondary hydroxyl groups, have become commercially available. A major work on chiral separations by HPLC gives more information on chiral columns.[19]

Detection of enantiomers is usually with conventional HPLC detectors; most often these are ultraviolet/visible (UV/VIS) filter photometric or UV/VIS diode array detectors. Fluorescence detectors provide greatly increased sensitivity for compounds that fluoresce or can be derivatized with a fluorescent reagent. Contrary to GC, there are commercially available HPLC detectors that can measure the optical activity of an eluting compound. These involve the principle of polarimetry, a technique that has been used for almost 150 years for the

measurement of optical activity. Modern HPLC detectors of this type use a plane-polarized laser beam that rotates left or right in the presence of an optically active compound, be it an enantiomer or a nonracemic mixture of unseparated enantiomers. Laser-based polarimetric HPLC detectors can measure as little as 12 ng of sample. The simultaneous measurement of absorbance and optical rotation during the resolution of chiral substances on a chiral HPLC column makes possible the determination of the enantiomer composition in spite of extensive peak overlap.[20]

Figure 2 gives examples of enantiomeric separations by HPLC. Included are methamidophos, which requires a low temperature for separation on the particular column used, and *o*,*p*′-methoxychlor, which also separates better at lower temperatures. Although resolution of enantiomers usually increases with decreasing temperature, lower temperatures diminish column efficiency.[19] The enantiomers of malathion, which is not asymmetric about the phosphorus atom but does contain an asymmetric carbon, are separated by 5.5 min at 25 °C.

Preparative HPLC is the primary technique used for separation and collection of adequate amounts of the enantiomers of a pesticide for the determination of optical rotation, absolute configuration, biological activity and/or toxicity of the respective enantiomers, as well as for studies of their environmental fate. For example, a permethylated γ-CD column was used to separate and collect about 50 µg of each enantiomer of *o*,*p*′-DDT and *o*,*p*′-DDD for determination

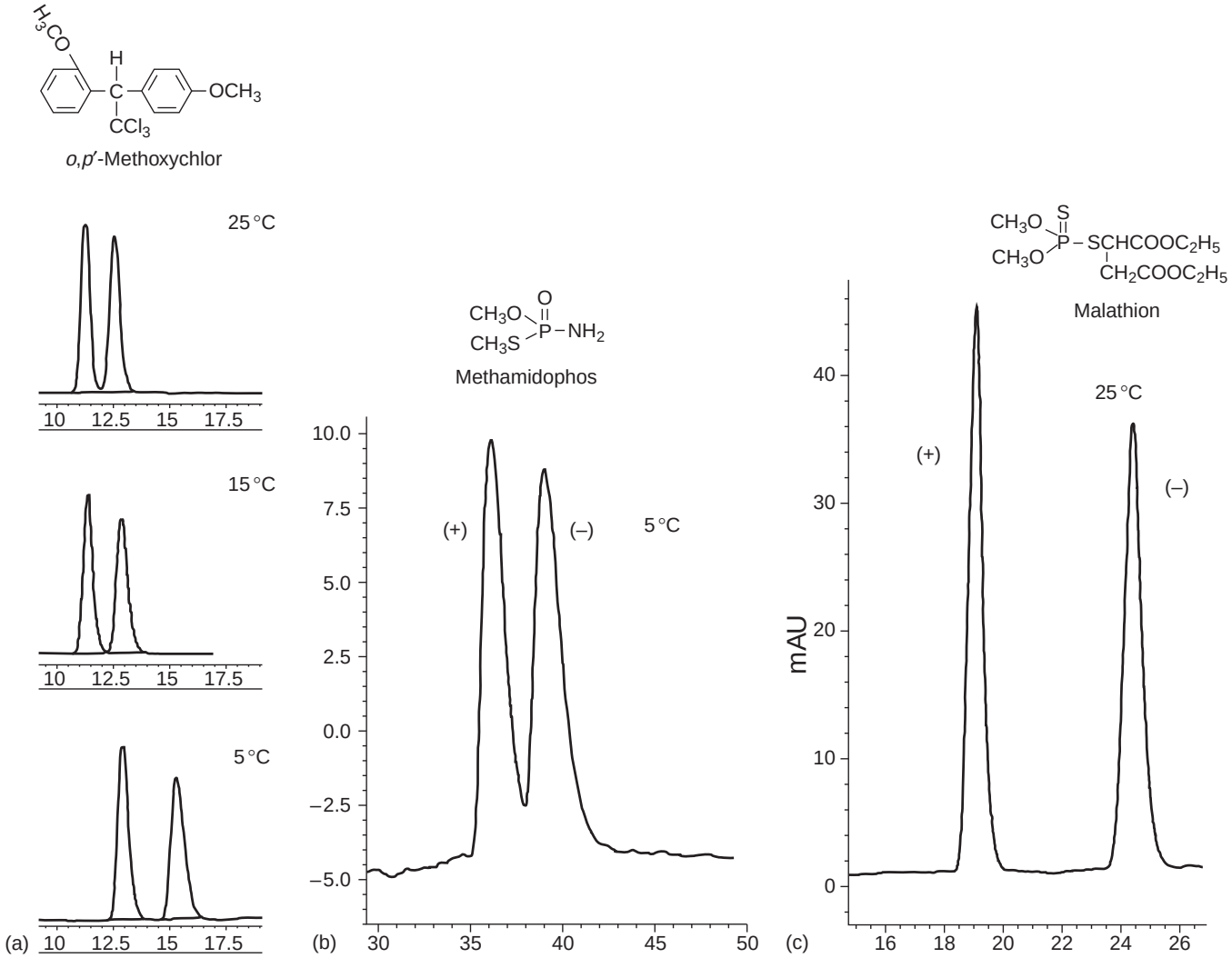

Figure 2 Enantiomeric separations by HPLC using a Chiracel® OJ column (Chiral Technologies, Inc., Exton, PA, USA): (a) *o*,*p*′-methoxychlor separates more efficiently at lower temperatures; (b) methamidophos requires a low temperature for separation; (c) malathion separates very efficiently at 25 °C. (This figure is supplied courtesy of the US Environmental Protection Agency and Chiral Technologies, Inc.)

of optical rotation (using a polarimetric detector during separation) and absolute configuration (by subsequent chemical means).[21] More recently, Type II chiral HPLC columns were used to separate 200–300 mg of each enantiomer of these two compounds, as well as the enantiomers of dichlorprop acid, dichlorprop methyl ester ruelene, and *o*,*p*′-methoxychlor, for subsequent testing of their respective endocrine-disrupting activities.

2.3 Chiral Separations by Capillary Electrophoresis

Commercial CE instruments have been available only since the mid-1980s, but have made a great impact in the drug industry in analysis of polar, especially ionizable, pharmaceutical products, including chiral products. The fair to high water solubility of such compounds lend them to the particular advantages of capillary zone electrophoresis (CZE), the most common mode of CE. In CZE, an electric current flowing through a fused silica capillary column creates an electro-osmotic flow (EOF) of the aqueous buffer, or carrier electrolyte, which generally moves toward the cathode; this flow also carries along the analytes. The pH of the buffer is adjusted in accordance with the pK_a of the analyte to create either a positive or negative charge on the analyte, so that it is attracted to either the cathode or the anode. However, the electro-osmotic flow is generally strong enough to eventually move both cations and then anions toward the

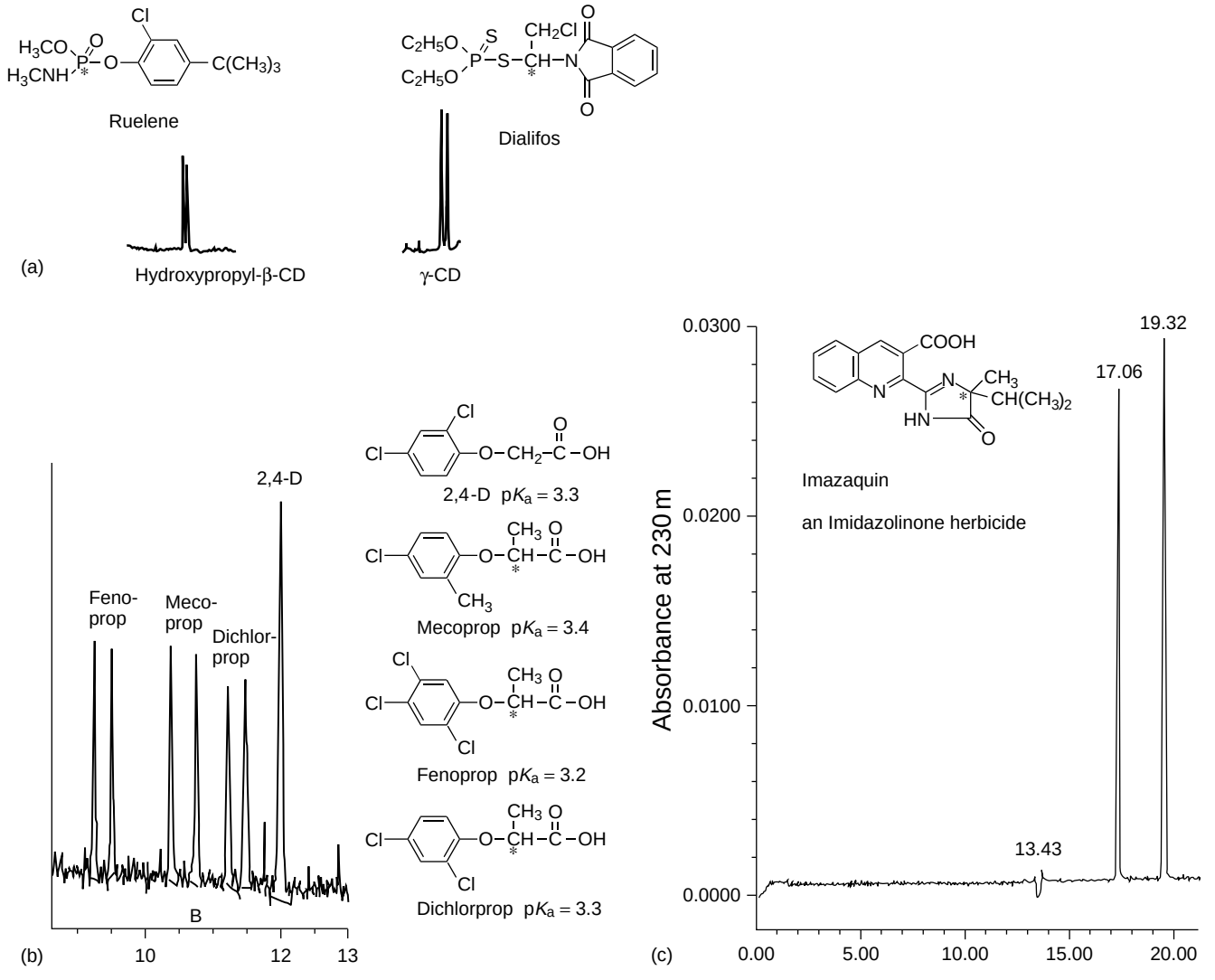

Figure 3 Enantiomeric separations by CE: (a) ruelene and dialifos (dialifor) by the MEKC mode of CE using hydroxypropyl-β-CD and γ-CD chiral selectors, respectively; (b) three phenoxy acid herbicides by the CZE mode using trimethyl-β-CD (notice that 2,4-D is not chiral); (c) imazaquin by CZE using dimethyl-β-CD. (This figure is supplied courtesy of the US Environmental Protection Agency.)

cathode. This movement carries the analytes through the detector window, which is simply a small section of the column, close to the anode, with the protective polyimide coating removed.[22]

Detection in CE analysis is by UV/VIS filter photometers or diode array detectors. Because the detector cell is in the column, the path length is very short, only 25–100 μm, so that sensitivity is proportionately less than with a UV/VIS HPLC detector. However, excellent sensitivity for fluorescent compounds or fluorescent derivatives can be achieved using a laser-induced fluorescence detector. In addition, columns widened to about 1 mm just at the detector window and columns arranged in a Z configuration to provide an extended cell pathlength at the window are available and these provide moderate sensitivity improvements. Finally, various types of electrochemical detectors that provide increased sensitivity and selectivity have been devised, but none of these are commercially available.

As is the case with pharmaceutical products, CZE is a good technique for analysis of ionizable pesticides. There are several classes of these for which CZE methods have been developed, including sulfonylureas, imidazolinones, triazines, and phenoxyacids. Many pesticides are not ionizable, of course, so cannot be analyzed by CZE – they simply move with the EOF without separating from one another. Micellar electrokinetic chromatography (MEKC) is a mode of CE designed for the analysis of nonionic compounds. In this mode, a surfactant, such as sodium dodecyl sulfate, is added to the CE buffer to create charged micelles. These micelles interact with the nonionic analytes (and often the ionic ones), resulting in a charged micelle–analyte complex that moves with the electric current, just as in CZE. The micelles act as a pseudostationary phase to provide the chromatographic part of the separation.

Separation of the enantiomers of chiral compounds by CE is simpler, faster, and often more efficient than with HPLC or GC. With CE, the appropriate chiral reagent (chiral selector) is simply added to the carrier electrolyte. In CZE, complexation of the enantiomers with the selector occurs in the aqueous buffer solution and, if the reaction is selective (complexation constants differ), enantiomeric separation occurs. For example, the enantiomers of three chiral phenoxy acid herbicides were separated by CZE using trimethyl-β-CD as the chiral selector (Figure 3).[23] In MEKC, complexation also occurs, but the neutral complexes of both enantiomers are then incorporated into the micelle for electrophoresis to occur. Schmitt et al. accomplished the enantiomeric separation of a variety of neutral chiral pesticides using MEKC with CD chiral selectors.[24]

Chiral selectors for CE are usually similar to the CDs used for chiral separations in HPLC or GC, except they are used in their neat chemical form rather than being part of a solid or liquid phase in a column. Most CDs are water soluble at the concentrations required for optimum enantiomer separation. Organic modifiers, usually methanol or acetonitrile, are often added to

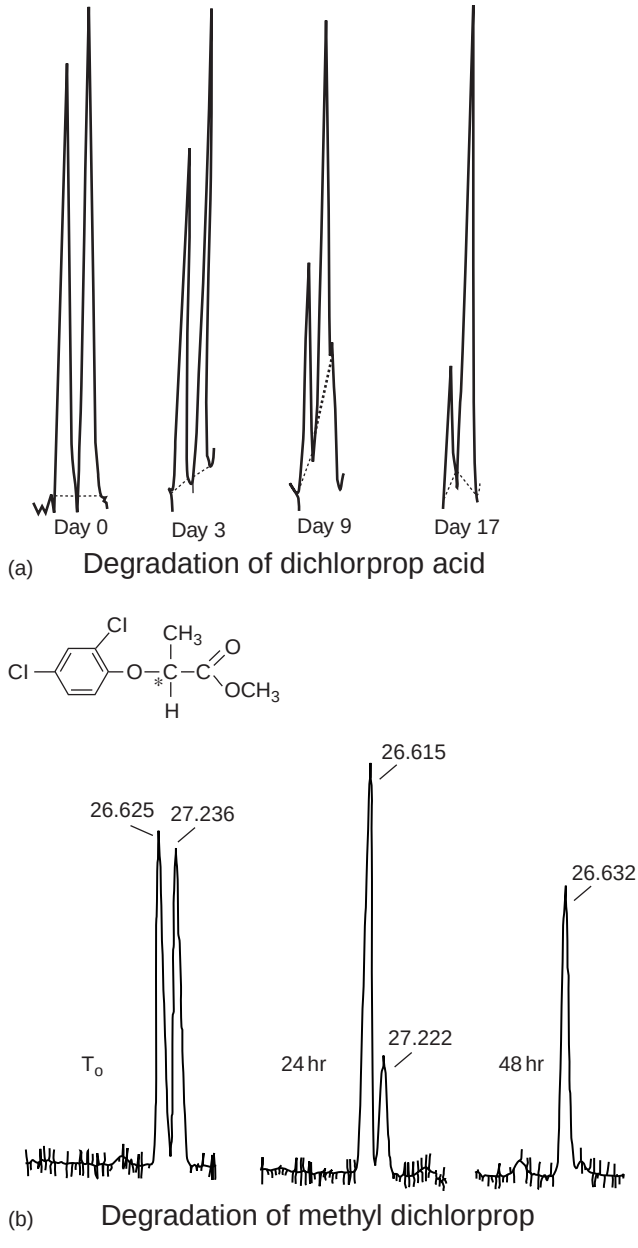

Figure 4 (a) Enantioselective microbial degradation of dichloroprop (acid) followed by CE using trimethyl-β-CD as the chiral selector. (b) Enantioselective microbial degradation of methyl dichlorprop followed by GC using a dimethyl-β-CD chiral column; numbers above peaks are retention times. (This figure is supplied courtesy of the US Environmental Protection Agency.)

the CZE or MEKC electrolyte to enhance separation; these modifiers also enhance the solubility of the CD as well as hydrophobic analytes. Sulfated ionic CDs have been developed that are not only more water soluble, but also eliminate the need for micelles; they selectively complex with enantiomers of both neutral and ionic analytes to produce ionic species for direct electrophoresis.[25]

Figure 3 shows MEKC enantiomeric separation of two organophosphorus pesticides: ruelene, which requires the chiral selector hydroxypropyl-β-CD for separation, is chiral because of asymmetry about the phosphorus atom; dialifos, requiring γ-CD, contains an asymmetric carbon. Figure 3 also illustrates CZE enantiomeric separations of an imidazolinone herbicide and three phenoxypropionic acid herbicides.

3 APPLICATIONS OF ENANTIOMERIC SEPARATIONS TO THE ENVIRONMENTAL CHEMISTRY OF PESTICIDES AND POLYCHLORINATED BIPHENYLS

The illustrations in the preceding section show successful enantiomeric separations of pesticide and PCB standards. The following figures illustrate the usefulness of these separation techniques in the study of environmental processes and the analysis of real samples that relate to problems in environmental chemistry.

Figure 4 shows the enantioselective microbiological degradation of the herbicide dichlorprop (2,4-dichlorophenoxy-2-propionic acid) in soil with time.[26]

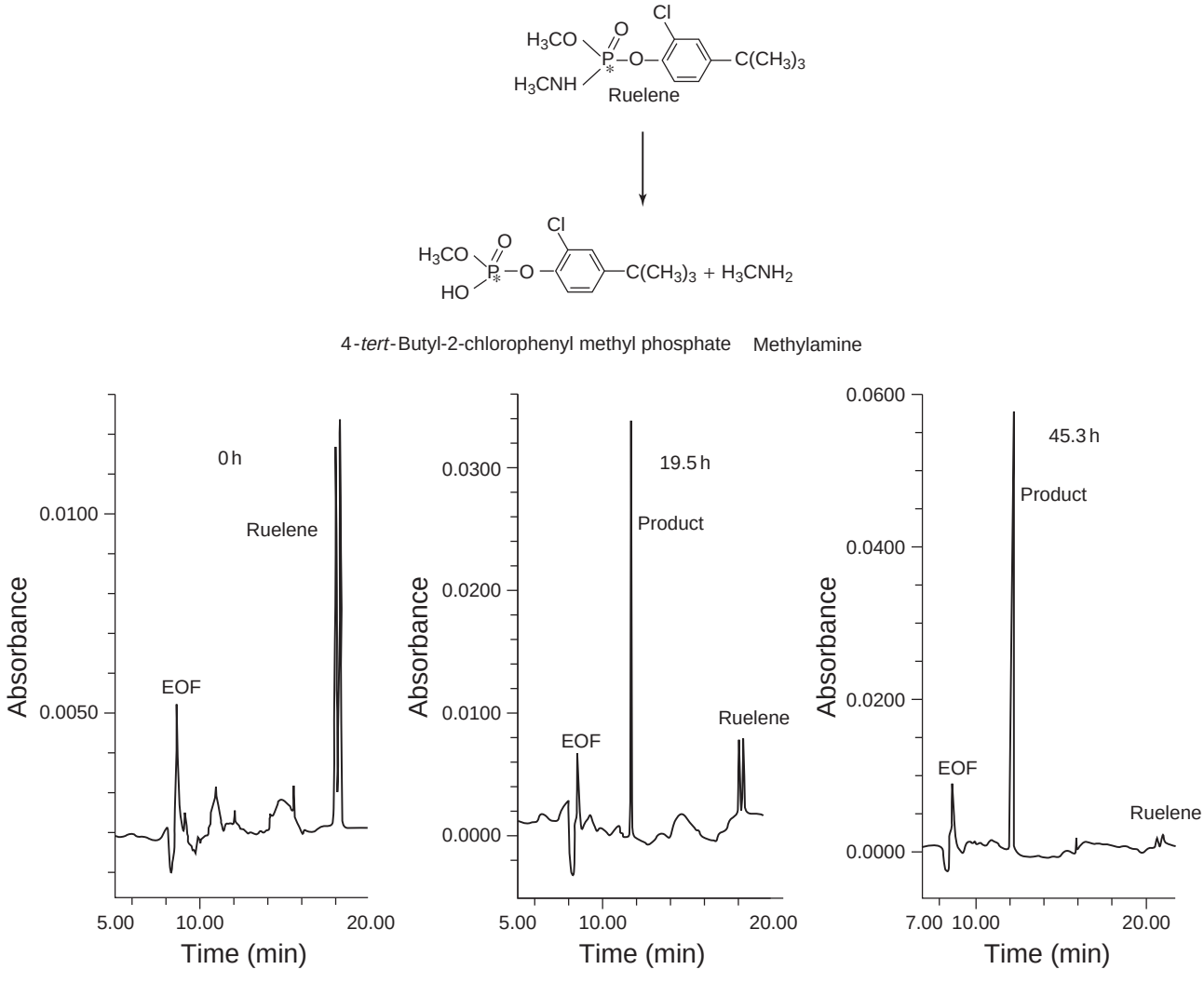

Figure 5 Acid hydrolysis of ruelene at pH 2.5 followed by CE in the MEKC mode using hydroxypropyl-β-CD as the chiral selector. Notice the formation of the product and the lack of enantioselectivity (equal enantiomer concentrations) in the loss of ruelene because this is an abiotic reaction. (This figure is supplied courtesy of the US Environmental Protection Agency.) EOF, electro-osmotic flow.

This herbicide is a component of Foxtril, a commercial formulation of herbicides. The *S*-(−)-enantiomer, with a half life of 4.4 days in this soil environment, degrades faster than the *R*-(+)-enantiomer, which has a half life of 8.7 days. Separation was by CZE, using heptakis (2,3,6-tri-*O*-methyl)-β-CD as the chiral selector.

The methyl ester of dichlorprop was used as a probe to assess the enantioselective activity of microbial populations in a variety of soil samples that had received different treatments. For example, exposure of this herbicide to one organic-rich soil resulted in complete degradation of the (−)-enantiomer in less that 48 h, as shown in Figure 4. This degradation process was followed by GC, using a dimethyl-β-CD column.

Figure 5 includes MEKC electropherograms of the organophosphorus pesticide ruelene (or crufomate, 4-*t*-butyl-2-chlorophenylmethyl *N*-methyl phosphoramidate) before and during acid hydrolysis at pH 2.5. Enantiomeric separation of ruelene is achieved by the addition of hydroxypropyl-β-CD to the MEKC buffer. After 19.5 h, most of the ruelene has hydrolyzed and a product peak, thought to be the phosphate ester resulting from hydrolysis of the P−N bond, is evident. (This product is chiral, but the particular chiral selector used was not appropriate for its enantiomeric separation.) A third electropherogram at 45 h shows the ruelene to be completely degraded; the half-life of this reaction was measured in this way as 16.4 h. It is obvious that the degradation was not enantioselective, which is as expected because this is an abiotic reaction.

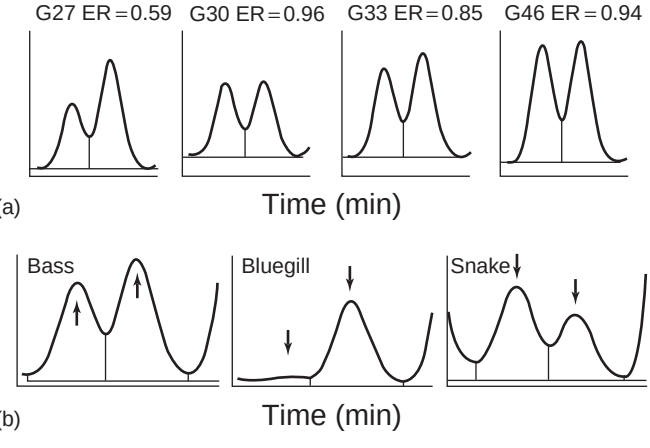

Figure 6 GC traces of the atropisomers of PCB 91 showing enantioselectivity in (a) sediment and (b) biota samples from Lake Hartwell. Increase in sample number (G27 to G46) corresponds to increase in distance from PCB spill. Separation is by chiral GC using a permethylated-β-CD bonded to a DB-5 stationary phase. (This figure is supplied courtesy of the US Environmental Protection Agency.)

PCBs are extremely persistent in the environment and are known to bioaccumulate in various animal species. Figure 6 shows GC traces of the enantiomers of PCB 91, one of the chiral congeners, in a series of sediment samples from Lake Hartwell, South Carolina, USA. A mixture of Aroclors, which are PCB commercial formulations, was accidentally spilled into the lake from a manufacturing facility some 25 years ago and has accumulated in the sediments. The chromatograms were obtained using a column containing permethylated-β-CD bonded to a DB-5 stationary phase. The data indicate that PCB 91 has degraded enantioselectively; the degree of selectivity decreasing with distance away from the spill as the Aroclor concentration decreases below a threshold necessary for microbial degradation. Figure 6 also depicts the high degree of enantioselectivity of PCB 91 that had accumulated in bass and bluegill fish and in a water snake from Lake Hartwell. Apparently, either bioaccumulation or metabolism in these aquatic species is enantioselective.

ACKNOWLEDGMENTS

Appreciation is expressed to Charles Wong, Jimmy Avants and Darrell Rennels for the examples of GC separations shown in Figure 1; to Jackson Ellington and John Evans for the HPLC examples in Figure 2; to Jimmy Avants and Tracy Cash for the CE separation of Imazaquin shown in Figure 3; and to Philippe Schmitt for assistance with the separations and kinetic data illustrated in Figures 3 and 4. Alton Wittemore and Glenn Chapman were responsible for the methyl dichlorprop data and GC separation in Figure 4; Lorrie Howell performed most of the kinetic work depicted in Figure 5; and Charles Wong produced the GC data of Figure 6. These persons are all associated with the US Environmental Protection Agency Laboratory in Athens, Georgia, USA, except for Philippe Schmitt, who is with the Institute of Ecological Chemistry in Munich, Germany. Tom Wiese of the US Environmental Protection Agency, Research Triangle Park, NC, performed the endocrine disrupter screening tests on separated enantiomers. Bill Champion and coworkers of Chiral Technologies, Inc., Exton, PA, USA performed the preparative HPLC separation of the enantiomers of *o*,*p*′-DDT, *o*,*p*′-DDD, dichlorprop acid, dichlorprop methyl ester, and *o*,*p*′-methoxychlor.

ABBREVIATIONS AND ACRONYMS

CD	Cyclodextrin
CE	Capillary Electrophoresis

CZE	Capillary Zone Electrophoresis
DB-5	95% Dimethyl : 5% Diphenylpolysiloxane
EOF	Electro-osmotic Flow
ER	Enantiomeric Ratio
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
MEKC	Micellar Electrokinetic Chromatography
o,*p*′-DDD	1,1-Dichloro-2-(*o*-chlorophenyl)-2-(*p*-chlorophenyl) ethane
o,*p*′-DDT	1,1,1-Trichloro-2-(*o*-chlorophenyl)-2-(*p*-chlorophenyl) ethane
PCB	Polychlorinated Biphenyl
UV/VIS	Ultraviolet/Visible
2,4-D	2,4-Dichlorophenoxy Acetic Acid
α-HCH	α-Hexachlorocyclohexane

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Matrix Isolation Spectroscopy in Atmospheric Chemistry

Environment: Water and Waste **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis • Polychlorinated Biphenyls Analysis in Environmental Samples

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography • Micellar Electrokinetic Chromatography

REFERENCES

1. E.L. Eliel, S.H. Wilen, L.N. Mander, *Stereochemistry of Organic Compounds*, John Wiley & Sons, New York, 1994.
2. E.L. Eliel, S.H. Wilen, L.N. Mander, *Stereochemistry of Organic Compounds*, John Wiley & Sons, New York, 201–208, 1994.
3. S.C. Stinson, 'Counting on Chiral Drugs', *Chem. Eng. News*, September 21, (1998).
4. H.-P.E. Kohler, A. Werner, W. Giger, C. Kanz, S. Muller, M.J.-F. Suter, 'Environmental Fate of Chiral Pollutants – the Necessity of Considering Stereochemistry', *Chimia*, **51**, 947–951 (1997).
5. D.W. Armstrong, G.L. Reid III, M.L. Hilton, C.-D. Chang, 'Relevance of Enantiomeric Separations in Environmental Science', *Environ. Pollut.*, **79**, 51–58 (1993).
6. G.M.R. Tombo, D. Belluš, 'Chirality and Crop Protection', *Angew. Chem., Int. Ed. Engl.*, **30**(10), 1193–1215 (1991).
7. K. Möller, H. Hühnerfuss, G. Rimkus, 'On the Diversity of Enzymatic Degradation Pathways of α-Hexachlorocyclohexane as Determined by Chiral Gas Chromatography', *J. High Resolut. Chromatogr.*, **16**, 672–673 (1993).
8. H.-R. Buser, M.D. Müller, 'Isomer- and Enantiomer-selective Analysis of Toxaphene Components Using Chiral High-resolution Gas Chromatography and Detection by Mass Spectrometry/Mass Spectrometry', *Environ. Sci. Technol.*, **28**(1), 119–128 (1994).
9. R.L. Falconer, T.F. Bidleman, D.J. Gregor, R. Semkin, C. Teixeira, 'Enantioselective Breakdown of α-Hexachlorocyclohexane in a Small Arctic Lake and Its Watershed', *Environ. Sci. Technol.*, **29**(5), 1297–1302 (1995).
10. L.M. Jantunen, T. Bidleman, 'Air–Water Gas Exchange of Hexachlorocyclohexanes (HCHs) and the Enantiomers of α-HCH in Arctic Regions', *J. Geophys. Res.*, **101**(D22), 28 837–28 846 (1996).
11. M.D. Müller, H.-R. Buser, 'Environmental Behavior of Acetamide Pesticide Stereoisomers. 2. Stereo- and Enantioselective Degradation in Sewage Sludge and Soil', *Environ. Sci. Technol.*, **29**(8), 2031–2037 (1995).
12. A. Glausch, J. Hahn, V. Schurig, 'Enantioselective Determination of Chiral 2,2′,3,3′,4,6′-Hexachlorobiphenyl (PCB 132) in Human Milk Samples by Multidimensional Gas Chromatography/Electron Capture Detection and by Mass Spectrometry', *Chemosphere*, **30**(11), 2079–2085 (1995).
13. A.W. Garrison, V.A. Nzengung, J.K. Avants, J. Ellington, N.L. Wolfe, 'Determining the Environmental Enantioselectivity of *o*,*p*′-DDT and *o*,*p*′-DDD', *Organohalogen Compd.*, **31**, 256–261 (1997).
14. E.J. Aigner, A.D. Leone, R.L. Falconer, 'Concentrations and Enantiomeric Ratios of Organochlorine Pesticides in Soils from the U.S. Corn Belt', *Environ. Sci. Technol.*, **32**(9), 1162–1168 (1998).
15. K. Wiberg, R. Letcher, C. Sandau, J. Duffe, R. Norstrom, P. Haglund, T. Bidleman, 'Enantioselective Gas Chromatography/Mass Spectrometry of Methylsulfonyl PCBs with Application to Arctic Marine Mammals', *Anal. Chem.*, **70**(18), 3845–3852 (1998).

16. H.-P. Buser, E. Francotte, 'Stereoselective Analysis in Crop Protection', in *Chiral Separations*, ed. S. Ahuja, American Chemical Society, Washington, DC, USA, 1997.
17. A. Williams, 'Opportunities for Chiral Chemicals', *Pestic. Sci.*, **46**(1), 3–9 (1996).
18. W. Vetter, V. Schurig, 'Enantioselective Determination of Chiral Organochlorine Compounds in Biota by Gas Chromatography on Modified Cyclodextrins', *J. Chromatogr. A*, **774**, 143–175 (1997).
19. G. Subramanian, *A Practical Approach to Chiral Separations by Liquid Chromatography*, VCH, Weinheim, Germany, 1994.
20. E.L. Eliel, S.H. Wilen, L.N. Mander, *Stereochemistry of Organic Compounds*, John Wiley & Sons, New York, 1072, 1994.
21. H.-R. Buser, M.D. Muller, 'Isomer-selective and Enantiomerselective Determination of DDT and Related Compounds Using Chiral High-resolution Gas Chromatography/Mass Spectrometry and Chiral High-performance Liquid Chromatography', *Anal. Chem.*, **67**(15), 2691–2698 (1995).
22. D.R. Baker, *Capillary Electrophoresis*, John Wiley & Sons, New York, 1995.
23. A.W. Garrison, P. Schmitt, A. Kettrup, 'Separation of Phenoxy Acid Herbicides and Their Enantiomers by High-performance Capillary Electrophoresis', *J. Chromatogr. A*, **688**, 317–327 (1994).
24. Ph. Schmitt, A.W. Garrison, D. Freitag, A. Kettrup, 'Application of Cyclodextrin-modified Micellar Electrokinetic Chromatography to the Separations of Selected Neutral Pesticides and Their Enantiomers', *J. Chromatogr. A*, **792**, 419–429 (1997).
25. J.B. Vincent, D.M. Kirby, T.V. Nguyen, G. Vigh, 'A Family of Single-isomer Chiral Resolving Agents for Capillary Electrophoresis. 2. Hepta-6-sulfato-β-cyclodextrin', *Anal. Chem.*, **69**(21), 4419–4428 (1997).
26. A.W. Garrison, P. Schmitt, D. Martens, A. Kettrup, 'Enantiomeric Selectivity in the Environmental Degradation of Dichlorprop as Determined by High-performance Capillary Electrophoresis', *Environ. Sci. Technol.*, **30**(8), 2449–2455 (1996).

Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis

Pilar Bou Carrasco and Josep M. Bayona
IIQAB-CSIC, Barcelona, Spain

1 Introduction **1**

2 Gas Chromatography of Pesticides **1**
2.1 Injection Techniques 2
2.2 Analytical Columns 4
2.3 Stationary Phases 4
2.4 Detection Systems 6
2.5 On-line Techniques 9
2.6 Fast-speed Gas Chromatography 9

3 Supercritical Fluid Chromatography of Pesticides **10**
3.1 Packed Column Supercritical Fluid Chromatography 10
3.2 Capillary Column Supercritical Fluid Chromatography 13

Abbreviations and Acronyms **15**

Related Articles **15**

References **16**

Pesticide determination by gas chromatography (GC) and supercritical fluid chromatography (SFC) techniques has been presented in this article. Instrumentation used in the different pesticide class analyses, such as injection (vaporizing, cold injection and large volume) and detection techniques (nitrogen–phosphorus detection, NPD; flame photometric detection, FPD; electron capture detection, ECD), and hyphenated systems (atomic emission detection, AED; mass spectrometry, MS; Fourier transform infrared, FTIR) used in GC and SFC determination of pesticides are presented. The figures of merit of each detection technique are outlined. Instrumental set-ups which combine extraction and separation techniques are also presented. Stationary phases used in GC and the type of SFC columns used in the determination of pesticides are also reported. Finally, the application of recent GC developments in the determination of pesticides is mentioned.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The generic term ''pesticide'' includes a variety of chemical classes with completely different physical–chemical properties and applications (i.e. biocide, herbicide, nematocide, fungicide, etc.).[1] In addition to process control, which can be fulfilled by many analytical techniques such as spectroscopic and electrochemical methods, there is also an analytical need for multiresidue analysis in a variety of matrices including environmental and food products. Analytical schemes include sample pretreatment (preconcentration, clean-up) followed by determination oriented to the chromatographic technique. GC using capillary columns and selective detection systems is the preferred analytical technique because of its high resolution, speed of analysis, low cost and easy coupling to highly sensitive and selective detectors. However, due to the large variety of compounds with different physical–chemical properties, this analytical technique cannot be applied to all the pesticide classes. Other chromatographic and related techniques such as liquid chromatography (LC), SFC and capillary electrophoresis (CE) are needed for some specific classes of pesticides. In this article, the application of GC (section 2) and SFC (section 3) to pesticide determination will be covered but no other important analytical aspects dealing with sample preparation for pesticide analysis. Since most of the analytical demands of pesticide analysis include degradation products, the relevant ones are also considered. Furthermore, the high toxicity at very low concentration levels demands analytical techniques and, particularly, injection and detection systems compatible with trace analysis. Some integrated approaches combining in-line sample preparation to chromatographic determination, developed to improve the sensitivity of pesticide determinations, are also included. Finally, due to the complexity of matrices containing pesticides, only selective detection systems will be considered.

2 GAS CHROMATOGRAPHY OF PESTICIDES

As mentioned earlier, capillary GC combined with selective detection systems is widely used in pesticide determination. However, pesticide thermal lability is an aspect to be considered carefully. In fact, some chemical classes of pesticides, such as *N*-methylcarbamates (NMCs) and most ureas, are susceptible to degradation at the usual injection temperatures (250–310 °C) used in conventional vaporizing injectors. To avoid this problem, the thermal labile analytes can be either derivatized (alkylation, esterification, etc.) or injected cold

on-column or in a programmable temperature vaporizer (PTV) injector which allows injection at low temperatures or minimization of the residence time at high injector temperatures. However, for extremely heat-sensitive analytes, despite the fact that the injection can be carried out at low temperatures, they can be thermally degraded in the column during the chromatographic run. Therefore, the derivatization step is unavoidable for those analytes if GC determination is envisaged. Generation of volatile derivatives allows the elution of pesticides at lower temperatures, avoiding their thermal degradation during the chromatographic run and also minimizing the adsorption on the active sites of the capillary column. Despite its clear analytical benefits, derivatization introduces an additional step in the analytical scheme, increasing the analysis time and the chances of analytical errors (i.e. partial derivatization yields, formation of by-products, etc.). Therefore, alternative analytical techniques such LC, SFC or CE are preferred for these specific analytes.

Nevertheless, derivatization reactions of pesticides are used in several laboratories and are summarized here for the most important pesticide classes.

Phenylureas. Derivatizations commonly used for some phenylurea determinations are alkylation with alkyl iodide and acylation with heptafluorobutyric acid (HFBA). Another option is their hydrolysis to obtain the corresponding anilines, followed by derivatization to obtain the fluoroacyl derivatives since they are more stable and show higher response than their precursors.[2]

Sulfonylureas. Sulfonylureas cannot be determined in their free forms by GC because of their thermolability. This can be overcome either by methylation with diazomethane in order to obtain the monomethyl or the dimethyl derivatives, or by derivatization with pentafluorobenzylbromide (PFBBr).

Ureas and Carbamates. Acetylation with acetic anhydride for determination of carmabate pesticides and some ureas is also reported.[2] Another approach for the determination of NMCs is the hydrolysis using trimethylsulfonium hydroxide followed by derivatization with pentafluoropropionic anhydride.[3]

Phenoxyacids. They need to be derivatized to obtain volatile compounds: esters or ethers. Alkyl, chloroalkyl, silyl or pentafluorobenzyl derivatives are obtained. Silyl derivatives do not provide enough sensitivity for trace level determination, whereas the methylesters are commonly used for residue analysis. However, these last derivatives of common phenoxyacids provide low retention times and present interferences. Typical derivatization reagents are diazomethane, boron trifluoride or the complex boron trichloride–methanol and PFBBr. The last derivatization increases the response in ECD but a sample clean-up prior to injection is required to avoid interferences.[2]

2.1 Injection Techniques

2.1.1 Vaporizing Injectors

Pesticide analysis requirements based on existing regulations in environmental and food products demand determinations at trace level. Therefore, the split injection is not suitable for the former mentioned instances, because the carrier-gas mixed together with the evaporated sample is vented and only a small fraction of the sample reaches the analytical column. Furthermore, it is not suitable for thermal labile compounds. The splitless injection technique is preferred because most of the evaporated sample injected is transferred to the column, increasing the sensitivity, but its volume is limited to a few microliters by the injector volume. Focusing methods of analytes must be used to obtain narrow peak bands. They are usually achieved by injecting the sample at lower column temperatures than the solvent boiling point which temporarily acts as a stationary phase.[4] Splitless injection is suitable for samples containing nonvolatile by-products which are retained in the injector port.[5] Hajslová et al.[6] evaluated the matrix effects when determining pesticide residues in water. Figure 1 shows a chromatogram illustrating the effect of crude extract clean-up. In Figure 1(a) a narrower ''pesticide fraction'' is collected (from 16 mL instead of 15 mL when the first analyte λ-cyhalothrin appears in the eluate). Reduced recoveries are obtained for some analytes when collecting the narrower fraction. Nevertheless the analysis of real matrices justifies this election because many interfering coextractants are detected in the broader ''pesticide fraction'' as shown in Figure 1(b).

Splitless injection is also used in solventless injection techniques such as solid-phase microextraction (SPME).[7] As in the conventional splitless injection with an organic solvent, the purge valve is closed before the injection and, in this case, activated when the thermal desorption period is completed (few minutes). In order to assure a suitable peak shape, column temperature should be kept low enough to obtain the band focusing effect.

One disadvantage of vaporizing injection when a wide range of pesticides is simultaneously analyzed is the so-called discrimination of the high boiling vs the volatile analytes. This effect can be minimized by optimizing the injection variables (i.e. hot needle, injection speed, boiling point of solvent, etc.) but not completely eliminated.[4] Therefore, calibration should be carried out with the same analyte or, if not available, with a closely eluting

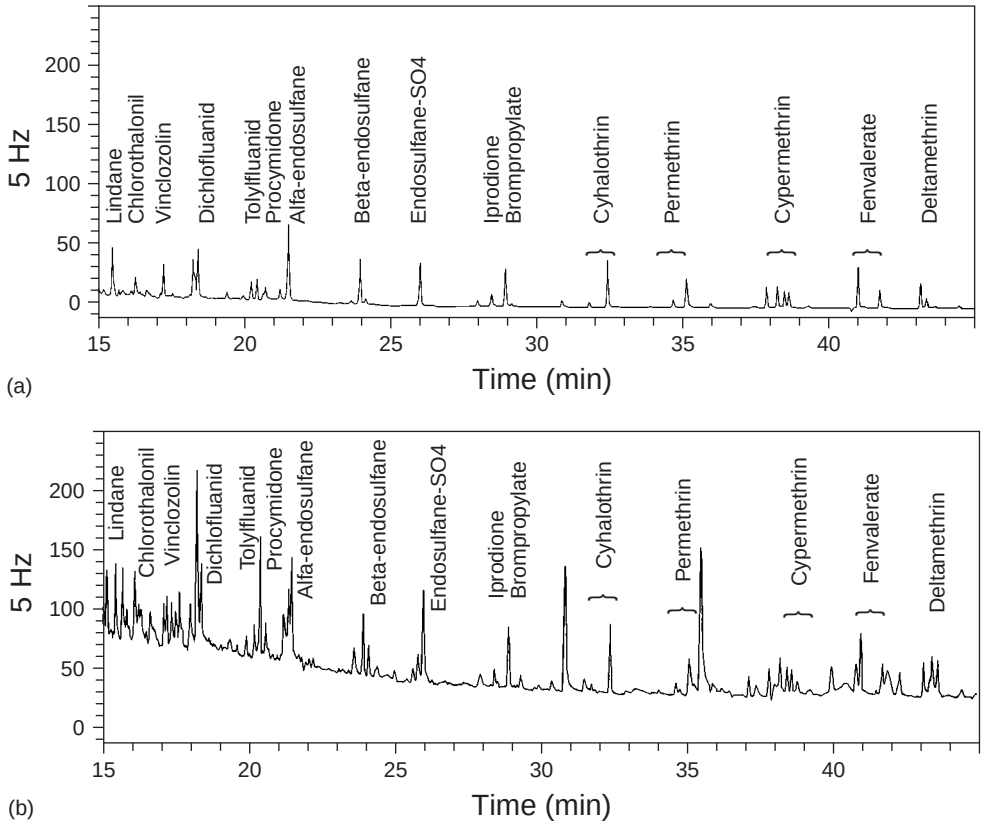

Figure 1 GC/ECD chromatograms illustrating the effect of crude extract clean-up (orange matrix, spiking level 10.8–101 ppb). Column: DB-5MS (60 m × 0.25 mm, 0.25 µm). Splitless injection: purge off, 120 s; injector temperature, 250 °C. (Reproduced from Hajslová et al.,[6] 288, Copyright (1998), by permission of Elsevier Science.)

compound. In the following discussion, the major classes of pesticides that are GC amenable with conventional vaporizing injectors are reported.

The use of electronic pressure programming injectors, which increase the pressure of the column inlet immediately after the injection, reduces thermal degradation of analytes.[3]

Organochlorine. Most organochlorine pesticides (OCPs) can be analyzed in conventional vaporizing injectors. However, at high injector temperatures (280 °C), thermal degradation of chlorinated diphenylethanes leads to the formation of dehydrochlorination products. These alkene-type products, namely dichlorodiphenylethylenes (DDEs), can occur in environmental samples since they are metabolites leading to biased results. Thus, injector temperature must be optimized to avoid this thermal degradation reaction during the GC analysis.

Organophosphates. These constitute a broad class of GC amenable pesticides. Only very few organophosphorus pesticides (i.e. CL 26691, pyraclofos and azamethiphos) among 68 examined have been identified as not being GC amenable.[1]

Pyrethroids and Triazines. These can be determined in conventional vaporizing injectors. Only cypermethrin among the 32 pesticides belonging to the pyrethroid class has been identified as not being GC amenable in conventional GC vaporizing injectors.[1] No single 1,3,5-triazine among 14 compounds examined has been identified as not being GC amenable.

2.1.2 Cold Injection Techniques

Programmable Temperature Vaporizer. Liquid samples are injected into a cool vaporizing chamber that is rapidly heated following the syringe needle being withdrawn.[4] Discrimination of the high-boiling analytes is minimized when the sample is injected and its precision is much better than the conventional hot injection techniques.[4] A PTV injector does not completely solve the thermal degradation of pesticides but is preferred to splitless injection for this particular analysis in terms of minimum discrimination and the larger injection volumes allowed without detector configuration modification.

On-column Injection Techniques. The sample is directly injected into the oven-thermostated column inlet

or into an uncoated precolumn which is connected to the analytical column.[4] The injector body is continuously cooled (primary cooling), as is the inlet of the precolumn during the injection and the post-injection time (5–20 s) (secondary cooling). In this way, not only the sample evaporation is avoided but also peak broadening. Furthermore, the evaporation process is highly reproducible. Whereas this technique is optimal to inject labile solutes, it is not convenient for highly polar analytes prone to adsorb on the active sites of the capillary wall surface. Many determinations of NMCs by using GC in the cold on-column injection have been carried out in order to reduce their thermal degradation.[3]

Several extraction techniques have been coupled to GC by means of an on-column interface in order to improve the procedural detection limits. Jongenotter et al.[8] described a gel permeation coupled to GC with FPD for organophosphate pesticides (OPPs) in olive oil with a flow-regulated on-column transfer of the gel permeation chromatography (GPC) fraction, controlled in a time basis. The temperature range is wider than in other methods, and allows the detection of a broad volatility range of OPPs (i.e. mevinphos, acephate and monocrotophos). Pocurull et al.[9] describe a solid-phase extraction (SPE) coupled to gas chromatography/mass spectrometry (GC/MS) using an on-column interface to determine organophosphates, organochlorines, triazines and thiocarbamates in tap water.

2.1.3 Large Volume Injection

Injection devices have been developed to accommodate large sample volumes (up to several hundreds of microliters) into the capillary column in order to minimize tedious extract concentration steps prior to the analytical determination. Large volume injections (LVIs) can be carried out by using on-column injection, where the sample is slowly injected and the solvent is concurrently evaporated from the analytes in a retention gap (e.g. bare deactivated fused silica tubing). Solute focusing is achieved in the column inlet and the chromatographic separation begins when column temperature is ramped. A modified PTV injector also allows LVIs.[10] Low volatility analytes are separated from the solvent in the injector port. PTV is preferred to an on-column injector because the latter is suitable for dirty samples.[5] Another possibility is using splitless injection for LVI. Analyte focusing is achieved by means of a precolumn with increasing film thickness with injection port distance coupled to a solvent venting valve where the analytical column is fitted.

Applications. Steen et al.[11] obtained limits of detection (LODs) in the range from 5 ng L^{-1} for deethylatrazine to 0.2 ng L^{-1} for alachlor and metolachlor in a PTV LVI tandem mass spectrometric detector. The initial sample volume was only 200 mL of marine and estuarine waters and the injection volume was 40 μL after the sample pretreatment.

López et al.[12] combined extraction procedures using low sample volumes (5–50 mL) with LVI/GC. A total of 16 organophosphates in clean water samples at the low concentration levels (nanogram per liter) were determined. The simplification in the extraction step by means of the LVI/GC technique allows less solvent consumption, produces less waste and reduces the analysis time. Suzuki et al.[13] reported injections of 100 μL of extract from water samples, allowing pesticide determinations at concentrations of 0.1 μg L^{-1} with an enrichment factor of 20 in the splitless LVI. On the other hand, in the conventional splitless injection only a few microliters of sample can be injected. Thus, to determine pesticides at the former concentration, an enrichment factor of 500–1000 by liquid–liquid extraction (LLE) with dichloromethane is required.

2.2 Analytical Columns

Open tubular capillary columns of different length and diameter are used in the determination of pesticides (Table 1). Typically, 25–30 m length and 0.25 mm internal diameter (ID) are used in multiresidue analytical procedures where high efficiency is needed for the determination of a large number of pesticides. For some specific applications where a class of pesticides such as organochlorines (i.e. toxaphene) co-elute with complex mixtures of compounds (polychlorinated biphenyls, PCBs), longer columns are used (i.e. 50–60 m). The main disadvantage in the use of longer columns is their long analysis time. Recently, narrow bore (100–150 mm ID), shorter columns (5–10 m length) have been gaining acceptance because analysis time is significantly reduced without sacrificing resolution. The main limitation of narrow bore columns is their low sample capacity. The determination of mixtures of pesticides containing very few components can be accomplished in very short columns (1–2 m length × 0.25 mm ID) (Table 1) in a few minutes. Wide bore columns (0.53 mm ID) are also used in the determination of simple mixtures of pesticides because they offer a large sample capacity but due to their larger flow rate are not useful in most GC/MS applications.

2.3 Stationary Phases

Current pesticide determinations are accomplished by gas–liquid chromatography. Stationary phase selection is based on the polarity of pesticides. Furthermore,

Table 1 Chromatography systems used in GC analysis of pesticides

Pesticide class	Matrix	Sample preparation[a]	Injection	Column dimension	Stationary phase	Detector	Ref.
Carbamates	Water	LLE, SPE	On column, LVI	1–2 m × 0.25 mm ID, 0.1 μm	5% Diphenyl methylPS[b]	MS	3
	Soil, plants	(1) Soxhlet, SFE, sonication (2) Clean-up: LLP, silica (3) Derivatization (PFPA, TMSH)				AED NPD ECD	
Chloroacetanilides	Water	SPE (SDB)	LVI, PTV	25 m × 0.22 mm ID, 0.25 μm	5% Diphenyl methylPS	MS/MS	11
Organochlorines	Water	(1) LLE, SPE (C_{18}, CN, SDB) (2) Florisil, silica GPC, H_2SO_4	All	25–50 m × 0.25 mm ID, 0.25 μm 10–30 m × 0.25 mm ID	0–5% Diphenyl methylPS Cyclodextrins	ECD/MS	14
	Soil	(1) Soxhlet, SFE (2) Clean-up					
	Cereal	(1) Soxhlet (2) Clean-up					
Phenylureas	Soils	(1) Ethylation (2) Hydrolization (3) Extraction	Splitless	12 m × 0.22/0.53 mm ID	0–5% Diphenyl methylPS	NPD MS/ITD	21
Phenoxyacids	Soils	(1) Soxhlet, SFE (2) Derivatization (BCl_3-MeOH)	Splitless	20 m × 0.25 mm ID	CyanophenylPS Cyclodextrin	MS/ITD	15
Pyrethroids	Vegetables	(1) MSPD (Florisil)	On column	30 m × 0.25 mm ID, 0.25 μm	0–5% Diphenyl methylPS	ECD	39
Triazines	Water	SPE (various)	LVI, PTV	25 m × 0.22 m ID, 0.25 μm	25% Diphenyl methylPS	MS/MS	11
	Soil	(1) Soxhlet, SFE, ASE (2) Florisil				NPD	2

[a] LLP, liquid–liquid partitioning; SFE, supercritical fluid extraction; PFPA, pentafluoropropionic anhydride; TMSH, trimethyl sulfonium hydroxide; SDB, styrene–divinylbenzene; MSPD, matrix solid-phase dispersion.
[b] PS, polysiloxane.

enantioselective stationary phases are needed to separate quiral pesticides.

Another practical consideration in the stationary phase selection is the detector selectivity used in the pesticide determination. Ideally, the stationary phase composition should not contain any heteroatom detectable to minimize the column bleeding. For instance, halogenated stationary phases cannot be used in the ECD and cannot be used for cyanopropyl in the NPD because they give a strong baseline drift in temperature-programmed separation, thereby decreasing the detection limits. Finally, specialty columns are recommended in GC/MS applications since the outlet of the column is fitted to high vacuum in the ion source leading to an increased background at high programming temperatures which can contaminate the ion source.

2.3.1 Nonpolar

Typically, nonpolar stationary phases chemically bonded to a column surface are the most widely used in capillary GC because of their high efficiency and thermal stability (low bleeding and elevated maximum operational temperature) and long lifetime. These stationary phases contain 0–5% diphenyl and 0–95% dimethylpolysiloxanes (OV1, SE-54, DB-5, HP-5, etc.). More recently, silylarene-type stationary phases were introduced for GC/MS applications because they exhibited a lower bleeding rate compared to conventional polysiloxanes.

These stationary phases are used in the determination of chlorinated and other lipophilic pesticides such as pyrethroids. Other applications of the nonpolar stationary phases are in the determination of derivatized polar pesticides such as NMCs or phenoxyacids (Table 1).

2.3.2 Medium Polarity

These stationary phases contain a higher degree (35–50%) of either phenyl and/or low substitution percentage (14%) of cyanopropyl polysiloxanes (OV-17, DB-35, DB-1701 or equivalents). They are used as secondary columns in the determination of OCPs. In addition, they are the stationary phase of choice for the determination of organophosphorus and triazines (Table 1).

2.3.3 High Polarity

Polar stationary phases contain a higher percentage of trifluoropropyl or cyanopropyl substituted polysiloxanes (DB-210, DB-200, DB-225 or equivalents). These stationary phases are useful for the determination of highly polar underivatized pesticides and herbicides (Table 1).

2.3.4 Chiral

Stationary phases used for chiral separations of pesticides are either derivatized or underivatized β- or γ-cyclodextrins which are usually chemically bonded to a typical polysiloxane (i.e. methyl, phenyl, or cyano substituted). Very few stationary phases contain the cyclodextrin directly bonded to the column surface. They have been used for the determination of chiral OCPs such as α-hexachlorocyclohexane, *cis*- and *trans*-chlordane, heptachlor, heptachlorepoxide, oxychlordane, 2,4′-dichlorodiphenyltrichloroethane (2,4′-DDT) and chlorinated camphenes (toxaphene congeners). Figure 2 shows the resolution of a racemic mixture of 2,4′-DDT enantiomers. In the case of polar pesticides such as the aryloxyalkanoic acids (i.e. dichlorprop and mecoprop), derivatization is necessary before the GC determination.[15]

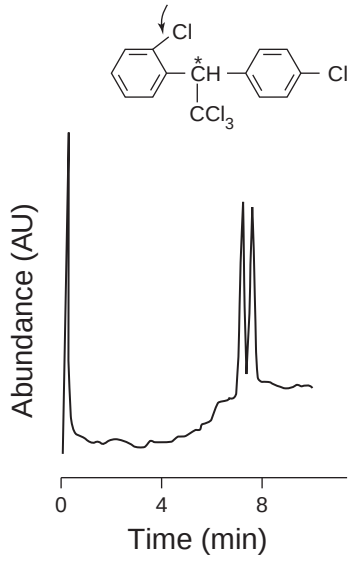

Figure 2 Enantiomeric separation of 2,4′-DDT on a 10-m capillary column (0.25 mm ID) coated with heptakis (2,6-di-*O*-methyl-3-*O*-trifluoro-acetyl)-*b*-cyclodextrin bonded to a dimethylpolysiloxane. (Reproduced from Vetter and Schurig,[14] 163, Copyright (1997), by permission of Elsevier.)

2.4 Detection Systems

Several properties, such as sensitivity, linear dynamic range, kind of response and its speed, noise, stability, etc., should be taken into consideration to select the convenient detector. Due to the fact that most of the pesticides contain heteroatoms, element-selective detectors are preferentially used in pesticide determination in real matrices. In Table 2, the properties of the main detectors used in the determination of pesticides are listed.

2.4.1 Nitrogen–Phosphorus Detector

This detector is widely employed in the determination of pesticides containing nitrogen or phosphorus because of its high sensitivity and selectivity towards these elements. The higher the molecular N : C ratio, the higher the sensitivity.

Thus, the detector has been applied to the determination of triazines, substituted ureas, thiocarbamates, dinitroanilines and chloroacetamides and concentrations in the parts per million to parts per billion range for residues of pesticides in soils, plants and waters.[2]

Draper[16] found a severe tailing in the organophosphorus pesticide analysis associated with the NPD mechanism being enhanced when the alkaline source ages. Tailing can be reduced or eliminated by recoating the source with rubidium and alumina powder. Response factors were increased 2- to 20-fold for phosphorus-containing pesticides by cleaning the bonded phase with an organic solvent, whereas the same factors for nitrogen-containing pesticides remain the same.

Organonitrogen. Sánchez-Brunete et al.[17] reported LODs in the range of $0.001–0.02\,\mu g\,g^{-1}$ for some chloroacetamides, thiocarbamates and triazines in soils. García-Valcárcel et al.[18] analyzed several dinitroanilines in air, plants and soils. LODs detected were at least $1\,ng\,L^{-1}$, near $0.05\,\mu g\,g^{-1}$ and $0.01\,\mu g\,g^{-1}$, respectively. A total of 16 organonitrogen-containing carbamate pesticides, such as pirimicarb, bitertanol and aldicarb, were monitored among a group of 48 pesticide residues in 10 different agricultural commodities (e.g. potato, cabbage, and lettuce). Detection limits were 0.01 ppm. No peak interference was detected (except for the cabbage sample) and the only pesticide detected in the

unfortified samples (banana) was bitertanol.[19] Triazines show the highest sensitivity in NPD. Low detection limits (0.4–4 ng L^{-1}) were reached for triazines in water samples. Triazines were detected in the 3–52 ng L^{-1} range.[20] Residues of phenylureas such as chlorotoluron, isoproturon and metoxuron from soil samples were detected directly or after alkylation. Similar results were obtained for both methods and LODs were 0.01 μg g^{-1} dry wt for each herbicide.[21] Several pyrethroids in plants were analyzed by NPD. No interference peaks were observed except for the apple and carrot samples. LODs were 0.01 ppm.[19]

Organophosphates. LODs in the range of 0.003 μg g^{-1} (diazinon) and 0.01 μg g^{-1} (quinalphos) in vegetables are reported.[22] For some matrices, such as alfalfa, broccoli and cauliflower, NPD is not a suitable detector method for organophosphates because it suffers from serious interference when crude extracts are analyzed, making it necessary to use a higher selective detector or to introduce a clean-up step in the analytical scheme.[23]

2.4.2 *Electron Capture Detector*

This detector is highly sensitive towards compounds with electronegative atoms or functional groups, e.g. halogens, nitro, etc. Thus, it is very convenient to analyze halogenated pesticides or organohalogen derivatized pesticides in environmental samples (Table 2). The best response is obtained with polyhalogenated pesticides (i.e. hexachlorobenzene, lindane, DDTs) and some dinitroanilines. ECD is also used to detect phenoxyacids, benzonitriles, pyrethroids and substituted ureas.[2] For trace level determination of pesticides, halogen atoms are

Table 2 Detection systems used in GC of pesticides

Type	LOD ($g\,s^{-1}$)	Linear range	Characteristics	Pesticide class (reference)	Matrix	LOD (ppb)
Nitrogen–phosphorus	10^{-8}–10^{-14}	10^5–10^7	• Tailing peaks (OPs)[16]	*Containing N–P*:		
			• Interferences in crude extracts for some vegetables matrices (OPs)[23]	Chloroacetamides[17]	Soils	1–20
				Dinitroanilines[18]	Air, vegetables and soils	1×10^{-3}; 50; 10
				Organonitrogens[19]	Fruits and vegetables	10
				Organophosphates[22]	Vegetables	3–10
				Pyrethroids[19]	Vegetables	10
				Phenylureas[21]	Soils	10
				Triazines[20]	Water	0.4×10^{-3}–4×10^{-3}
Electron capture	10^{-14}	10^2–10^3	• Not very selective	*Halogenated or derivatized*:		
			• Narrow linear range	Benzonitriles[2]	Several	
				Dinitroanilines[2]	Air	1 ng
			• Limited linearity towards OCs[25]	Organochlorines[38]	Water	
				Phenoxiacids[2]	Several	
			• Preferrable for clean samples	Pyrethroids[39]	Vegetables	5.1–91.5
				Thiocarbamates[24]	Several	
Atomic emission	N: 7×10^{-12} S: 1.7×10^{-12} P: 1.5×10^{-12} Cl: 39×10^{-12}	10^3–10^4	• Appropriate for "dirty" matrices	Organochlorine[24]	Several	
				Organofluorine[24]	Several	15–300
				Organophosphates[24]	Several	
MS						
EI	0.5–2×10^{-12}	10^3	• Not very selective	All classes[11,28]	Several	0.1–1.0
NCI	$>0.1 \times 10^{-12}$	10^4	• Appropriate for dirty samples			0.2–5.0 ppt

EI, electron impact (70 eV); NCI, negative chemical ionization; MS/MS, tandem mass spectrometry.

introduced in the pesticide molecule via derivatization (see section 1). In some particular ''dirty'' matrices, such as alfalfa, halogenated pesticides cannot be detected with ECD.[23] This problem is usually overcome either with a clean-up of crude extract or with two-dimensional GC. Screening analysis of pesticide residues can be carried out reliably with the majority of foodstuffs with GC in parallel to ECD/NPD. Comparing cost–profit calculations of GC/ECD/NPD with GC/AED, the former is a cost effective choice to analyze pesticides in foodstuffs.[24]

Some disadvantages of ECD are its narrow linearity range, limited selectivity and the detector maintenance due to the radioactive source. Linearity range and detector sensitivity have been improved by detector redesign[25] such as Micro-ECD (HP 6890 series). In addition, nonradioactive ionization sources have been developed to replace conventional ECD, as in pulse discharge detection (PDD) which uses low energy ionization on helium doped with methane. Detection limits are in the femtogram or picogram range, depending on the analyte, and high detector temperatures can be reached with this detector (i.e. 400 °C).

2.4.3 Flame Photometric Detector

The elemental selectivity allows the determination of organophosphorus- and organosulfur-containing pesticides using appropriate bandpass filters. Quenching effects in the detection of organosulfur pesticides in matrices containing large concentrations of hydrocarbons and the nonlinear response for sulfur are the major drawbacks of this detector. Dual-FPD was introduced to solve this problem but the time-consuming optimization process has focused analytical interest towards pulsed flame photometric detection (PFPD) which improves the sensitivity in the determination of pesticides in complex matrices. Absolute detection limits of this detector are in the low picogram range for organophosphorus pesticides in the sulfur mode.[26] Procedural detection limits in the range of 100 ppt for methylparathion from eggplants in the phosphorus mode have been reported.[27]

2.4.4 Atomic Emission Detector

The multielemental detection capability of this detector makes it very powerful in the characterization of unknown compounds. In addition, detection limits of most common elements found in herbicides are in the picogram per second range (Table 2). Quantitative analysis of each element makes calculation of the approximate empirical formula of analytes feasible.[2]

For some difficult matrices, such as leek, garlic, onion and cabbage, AED appears to be the optimal detection method for screening. The reliability of quantitative results at low concentration levels is remarkable.[24] Lee and Wylie[23] have found that AED has much better selectivity than NPD or ECD, and could be used to determine organochlorines, organofluorine and OPPs in agricultural matrices. Moreover, for some matrices it is possible to obviate several clean-up steps. Nevertheless, this detector is not very suitable if nitrogen is the only heteroatom present in the pesticide because the AED sensitivity and selectivity for nitrogen are lower than for other elements of interest.

2.4.5 Mass Spectrometry

The low cost of the electronic impact mass spectrometric detectors as well as the structural information provided in the full scan mode or the high sensitivity and selectivity in the selected ion monitoring mode, means that they replace other conventional detectors used in pesticide multiresidue analysis. Furthermore, mass spectral libraries contain over 120 000 compounds, which together with the powerful algorithms for analyte identification by library matching and co-eluted peak deconvolution make GC/MS the analytical technique of choice for the identification of unknown pesticides and their metabolites in real samples. All pesticide classes, provided that they are thermally stable (otherwise derivatives are needed), can be analyzed by GC/MS (see section 1). The combination of GC/MS with SPE allows the determination of a large number of pesticides at sub-microgram per liter levels in groundwaters by using the selected ion monitoring technique.[9,28] In addition to a quantitation ion, at least one additional ion is used for confirmation purposes in the selected ion monitoring mode to assess the compound identity.

However, GC/MS in electron impact mode does not possess enough selectivity for the determination of pesticides from more difficult matrices. Two different approaches can be used to circumvent this limitation. One of them is negative chemical ionization using methane or ammonia as reagent gas. GC/MS in negative chemical ionization mode is very useful for the determination of pesticides containing electron-withdrawing groups such as organohalogens, organophosphorus and the nitro-substituted aromatics because only these analytes are ionized. Usually, the molecular ion is the base peak of the mass spectra and the isotopic distribution gives additional information on the heteroatom content in the molecule.[29] The molecular anion abundance strongly depends on the ability to stabilize it by charge delocalization of the radical formed in the electron capture reaction. In the case of organophosphates, the molecular anion is obtained only when the aromatic moiety is substituted with electron-withdrawing groups.[30] Pyrethrin pesticides can be determined by a combination of negative chemical

ionization and positive chemical ionization.[31] In fact, the preferred fragmentation modes involved cleavage around the ester oxygen single bond, with one fragment appearing in the positive and the other in the negative ionization mass spectrum. The main limitation of negative chemical ionization is the difficulty in reproducing mass spectral patterns because it depends on the type of instrumentation used and many other instrumental variables, such as ion source temperature and pressure, which make it difficult to build mass spectral libraries.[32] Furthermore, the linearity range is narrower than for electron impact and the response factor strongly varies with the compound or even the isomer. Nevertheless, negative chemical ionization is useful in the identification of unknown compounds detected by ECD.

Another useful approach for the determination of pesticides in difficult matrices is the application of the MS/MS analyzer such as ion trap detector or triple stage quadrupole. The low cost of the ion trap detector and the possibility to perform *n* times MS/MS increased the interest of this instrumentation in the determination of pesticides because of the selectivity (Figure 3). In the literature, methods for the detection, quantification, and confirmation of more than 100 pesticides by GC/ion trap detection are available.[33] Using MS/MS, very low detection limits (low parts per billion) and reliable confirmation (one precursor ion and two or more product ions) are simultaneously achieved.

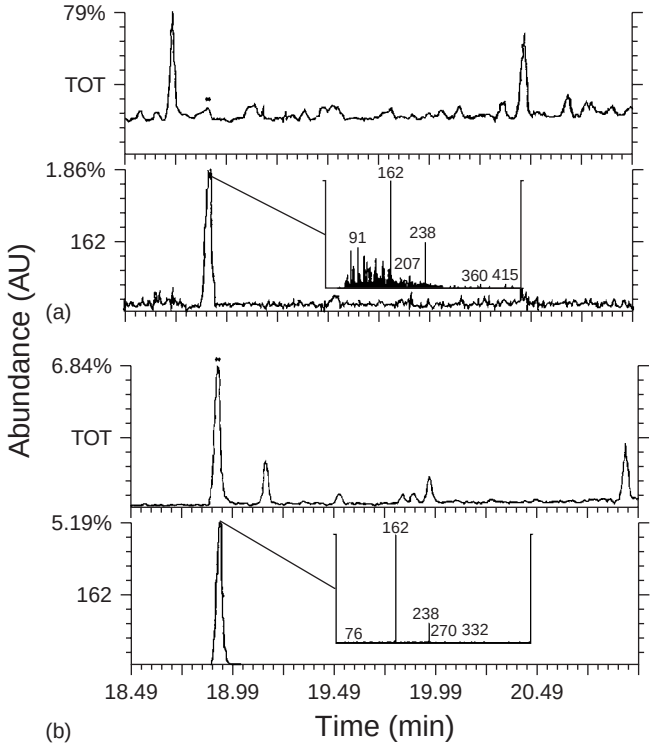

Figure 3 Comparison of GC/ion trap detection: (a) electron impact/MS; (b) electron impact/MS/MS. The asterisks indicate the peak position in the total ion chromatograms. (Reproduced from Steen et al.,[11] Copyright (1997), by permission of Elsevier.)

2.4.6 Fourier Transform Infrared Spectrometry

This detection system gives additional structural information to GC/MS and GC/AED and is particularly useful for the identification of unknown degradation products of pesticides. Light-pipe was the first commercially available interface for GC/FTIR coupling. The main drawback of this interface was its limited sensitivity. This limitation has been overcome by the introduction of cryotrapping open split interfaces which allow detection limits in the picogram and subpicogram range by signal averaging.[34]

2.5 On-line Techniques

Sample preparation techniques such as SPE or LLE techniques have been combined with GC, aiming to automate sample preparation steps. These techniques use LVI systems to introduce the organic extract obtained by elution of the SPE of the dried cartridge into the GC port. The main limitation of this approach is the large number of coextractants that are introduced into the analytical column. Heartcuts of LC in the normal phase, size-exclusion chromatography (as a clean-up) and reversed-phase[35] mode have been coupled to GC using a multiport switching valve. The latter technique involves the direct injection of aqueous samples into the GC column, which can be troublesome. SPME is another approach used in sample preparation to avoid the use of solvents or the direct injection of water into the column. The extraction can be performed in the aqueous phase or in the headspace in equilibrium with the aqueous phase continuously stirred during the extraction.[36] A variety of fibers have been used according to the polarity of the pesticide class.

2.6 Fast-speed Gas Chromatography

Very recently the increased demand for the determination of a large number of samples in food and environmental analysis has increased attention on the application of fast-speed determination of pesticides. Fast-speed GC can be accomplished with the use of short, narrow-bore columns (2–5 m length $\times$ 0.50–0.100 μm ID columns) leading to fast separations in minutes.[37] The main limitation of this approach is the limited sample capacity of the column and its high back-pressure, which demands in most applications a modification of the GC port. A GC oven with very fast heating rates (up to 400 °C) using inductive heating (so-called flash GC) has been recently introduced

on the market. This technique allows the determination of simple mixtures of pesticides in only a few seconds.

3 SUPERCRITICAL FLUID CHROMATOGRAPHY OF PESTICIDES

Supercritical CO_2 is the most widely used fluid because of its relatively low critical constants and inertness. However, it can react in the presence of pesticides containing primary amine substituents that should be derivatized before SFC determination. Another limitation of CO_2 as a mobile phase is its low polarity but in the presence of polar analytes it is polarizable, which allows their solvatation. An advantage of SFC in the determination of pesticides compared to GC is the use of lower temperatures, allowing the determination of thermal labile pesticides that are not GC amenable. In comparison to LC and capillary column supercritical fluid chromatography (cSFC), analysis time in packed column supercritical fluid chromatography (pSFC) is shorter. Four different mobile phase operational modes are currently used in SFC: (i) density programming; (ii) pressure programming; (iii) isobaric; and (iv) gradient programming using modified CO_2. Usually, density and pressure programming are applied in cSFC while isobaric or gradient programming is used in pSFC.

Until now, SFC has been applied to the determination of a large variety of pesticides and functional groups using different detection systems. Furthermore, it has been coupled to SPE for sample preparation. Although most of the SFC applications in the pesticide residue analysis have been focused on environmental matrices, several analytical schemes including SFC determination have been developed for pesticide determination in food matrices. Despite the potential advantages of SFC over more widely used chromatographic techniques, the high cost of SFC instrumentation precludes its widespread acceptance.

3.1 Packed Column Supercritical Fluid Chromatography

Most recent applications of SFC to pesticide determination use packed columns because of easier sample handling and because they allow LVIs without sophisticated hardware. Also, analysis time is much shorter than for cSFC or LC if a single column is used.[38,39] The large flow rate used in pSFC allows its electronic regulation independent of pressure, which permits operation at optimum chromatographic conditions. Conventional LC columns are currently used in pSFC but polar modifiers must be added to supercritical CO_2 to elute polar pesticides. However, due to the small pressure drop along the column, pSFC offers the unique possibility to couple up to 10 analytical packed columns (5 μm particle size, 4.6 ID × 40 mm length) in series to increase the global efficiency, yielding high resolution of complex mixtures of pesticides.[40] Several applications have been developed involving SFE/SFC coupling either to clean the sample or to increase the sensitivity of the detection system (see section 3.1.3).

3.1.1 Stationary Phases

The efficiency of different pSFCs for pesticide analysis has been evaluated for three different column packings (Si-60, 3 μm; C-18, 5 μm; and CN, 5 μm).[41] The highest efficiency was obtained with the Si-60 column (45 790 N m^{-1}, 100 °C, 160 atm). It could be attributable to the smaller particle size of this packing. Stationary phases used in pSFC depend on the class of pesticides (Table 3). Consequently, further discussion will be organized according to pesticide chemical classes.

Carbamates. Different stationary phases have been evaluated for the determination of carbamates (i.e. cyanopropyl, hypersil, silica, diol). The most selective stationary phase for the 11 carbamates included in USEPA (United States Environmental Protection Agency) method 531.1 was a diol stationary phase using CO_2–methanol.[42] Pressure, mobile phase composition and temperature are key variables affecting the analyte retention. A careful optimization of variables affecting the peak retention (i.e. pressure, temperature, modifier volume) is necessary since they control the retention time and reversal elution (change in the elution order) can be obtained if the pressure, temperature or mobile phase composition is slightly modified (Figure 4). The

Table 3 Stationary phases and detection systems used in pSFC of pesticides

Analyte	Matrix	Mobile phase	Stationary phase	Detection systems	LOD	Ref.
Carbamates	Standard	10% MeOH–CO_2 40 °C, 200 bar	Diol	UV, NPD in line	15 pg	43
Phenylureas	Standard	6% MeOH–CO_2	Diol	ECD, NPD		44
Sulfonylureas	Standard	10% MeOH–CO_2	Diol	UV, ECD, NPD		45
Organochlorine	Carrots	CO_2	ODS	ECD		47
Organophosphorus	Vegetable	MeOH–CO_2	ODS	NPD		48

UV, ultraviolet.

SFC determination of carbamates is at least six times faster than the standard USEPA method 531.1 using LC involving post-column derivatization since the determination can be carried out without gradient programming and the chromatographic run is shorter, thus increasing the sample throughput.

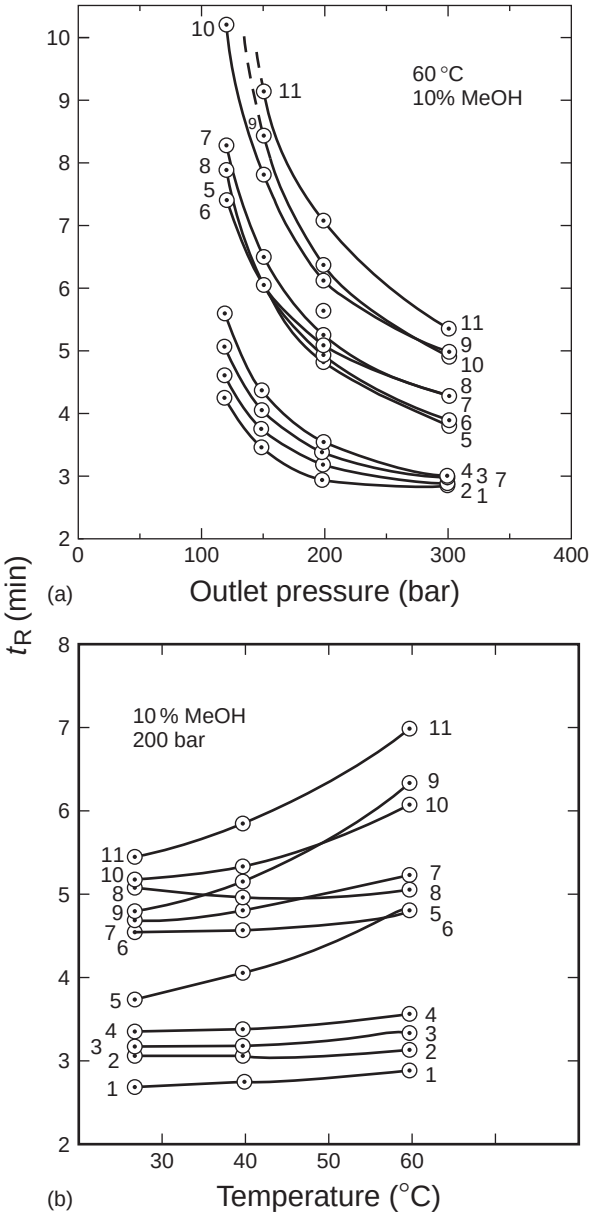

Figure 4 Effect of (a) pressure and (b) temperature on the retention of carbamates and selected metabolites. Column: 4.6 × 250 mm, 5-µm Lichrosphere diol, 2.0 mL min^{-1} CO_2–methanol. Compound identification: 1, aldicarb; 2, baygon; 3, carbofuran; 4, methiocarb; 5, aldicarb sulfone; 6, carbaryl; 7, methomyl; 8, 1-naphthol; 9, 3-hydroxycarbofuran; 10, aldicarb sulfoxide; 11, oxamyl. (Reproduced from the *Journal of Chromatographic Science* by permission of Preston Publications, a division of Preston Industries, Inc.)

Phenyl Urea and Sulfonylurea Herbicides. Phenylureas can be analyzed by pSFC. The best resolution was obtained with a diol column at low temperatures. At 60 °C, thermal degradation was noticed.[43] Sulfonylureas can also be analyzed with a diol column using methanol–CO_2 in a short analysis time (6 min) at 40 °C and 200 bar (Figure 5). An increase of temperature led to a longer retention time due to a drop in the density of the mobile phase but no reversal in the elution of analytes was found.[44] Sulfonylurea degradation products also can be analyzed by pSFC.[40]

Triazines. This was one of the first herbicide classes that was analyzed by SFC.[45] Hypersil silica (5 µm, 2.0 m × 4.6 mm) is particularly effective for atrazines, allowing a baseline resolution of eight analytes in less than 30 min.[43]

The retention behavior of OCPs (i.e. methoxychlor, DDT, DDE, aldrin and lindane) according to temperature on the reversed-phase C_{18} showed no decrease of capacity factors (k') with temperature (60–130 °C). Hexachlorocyclohexanes and 4,4′-dichlorodiphenyldichloroethane were analyzed with an octadecylsilane silica column and eluted with neat CO_2 using an ECD determination. This analytical procedure was applied to SFE extracts recovered from carrots.[46] Since most OCPs are thermally stable, GC/ECD is the preferred analytical technique because higher resolution can be obtained than in SFC/ECD (see section 2.4.2).

The separation of OPPs has been evaluated in different stationary phases (i.e. silica, NH_2, diol, CN, C-8, C-18 and PRP-1).[47] An organic modifier of CO_2, such as methanol, was needed to improve the peak shape and

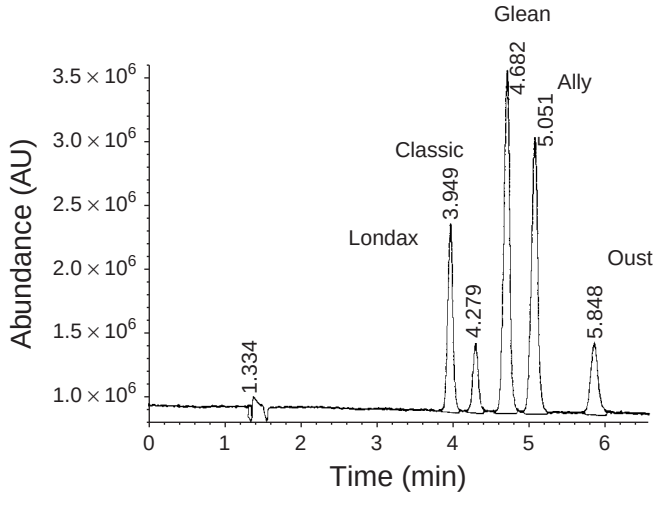

Figure 5 Separation of sulfonylurea herbicides on a diol column. Column: 4.6 × 250 mm, 5-µm Lichrosphere diol, 2 mL min^{-1}, 10% methanol in CO_2, 40 °C, 200 bar. (Reproduced by permission of The Royal Society of Chemistry.)

to reduce the analysis time. The retention behavior of selected OPPs depended on the interactions between functionalities of analytes and stationary phase and decreased with analyte solvation (density increase). Cyanopropyl stationary phase was the optimum in terms of resolution and analysis time (15 min). This stationary phase was used for the determination of spiked OPPs from vegetable extract (i.e. cucumber, lettuce, and grapes).

The low pressure drop along the chromatographic column allows the possibility to use long packed columns providing high resolution in complex mixtures of pesticides. In this regard, nearly 100 thermal labile pesticides were analyzed with a 1.6-m-long column.[48]

Some of the enantioselective stationary phases available for LC can be applied in pSFC for chiral separations of agrochemicals.[49] Although the resolution obtained in LC and pSFC is comparable, the latter can provide shorter analysis time as in other columns.

3.1.2 Detection Systems

pSFC uses LC-like detectors modified to the operational conditions of SFC because decompression is carried out following the detection. In this regard, pressure-resistant optical cells are mandatory in pSFC when spectroscopic detectors are used. Furthermore, GC-like detectors can be used but the large flow rates used in pSFC make splitting to other detectors necessary, allowing simultaneous multiple detection signals. Since modifiers are currently used in pSFC, detectors should be compatible with them. NPD, ECD, photoionization detection (PID) and sulfur chemiluminescence detection (SCD) have been used in the determination of pesticides. Applications of the different detectors according to pesticide classes are described hereafter.

Carbamates. The determination of carbamates has been carried out using NPD and ultraviolet (UV) detectors.[42] However, the effluent of the analytical column must be split (1/1000) to allow the introduction of a few microliters of mobile phase for NPD or thermionic detection (TID). The signal ratio for UV/NPD detectors provides structural information that is useful for identity confirmation purposes.

Sulfonylureas. Sulfonylureas and their degradation products can be detected by ECD, NPD or UV.[37] Other pesticides with strong UV-absorbing chromophores (i.e. diflubenzuron, teflubenzuron, triflumuron) were readily detected at 254 nm in the UV region.[50]

Phenylureas. Response surfaces of the ECD detector working at various detector conditions have been used to optimize the response using CO_2 modified with methanol.[51] The ECD response of phenylureas is dependent on the amount of modifier but the LODs of phenylureas were in the low picogram level when 5% methanol in CO_2 was used as the mobile phase.

Organophosphate and Organosulfur Pesticides. A new generation of SCD systems was interfaced and tested for SFC with 100% CO_2 and CO_2 modified with methanol.[52] The detection chemistry of SCD is based on ozone-induced chemiluminescence following a two-step combustion process of consecutive oxidation and reduction of sulfur-containing compounds. The LOD was 3 pg of sulfur ($0.2\,pg\,s^{-1}$) and the molar response is independent of the analyte bonding environment.

A photoionization detector was coupled to an SFC apparatus to evaluate its performance.[53] Organophosphorus and organosulfur pesticides (i.e. disulfoton, ethion, and sulfotep) could be detected at low nanogram levels using methanol-modified CO_2. In addition, TID has been used in the determination of OPPs in spiked vegetable samples allowing LODs in the low picogram level (50–165 pg), depending on the analyte.[47] This detector exhibited at least two orders of magnitude of dynamic range.

Chloroatrazines. A comparison between the thermospray LC/MS and electrospray SFC/MS interphase has been carried out in the determination of atrazine metabolites.[54] It was found that SFC/MS was the more sensitive for the less polar chloroatrazine compounds whereas the thermospray presented better sensitivity for the dealkylated products of atrazines.

3.1.3 In-line Supercritical Fluid Extraction/Solid-phase Extraction/Supercritical Fluid Chromatography

Two different approaches have been developed in order to improve the detection limits and sample throughput in pSFC. SPE and SFE for aqueous and solid matrices, respectively, were coupled to pSFC.[55,56] In-line SFE/SFC interfaces have been developed[48] and utilized for trace analyses of OCPs and OPPs from complex matrices (i.e. chicken fat, ground beef, and lard).[57] A clean-up step, incorporating pSFC, allowed the fractionation of relatively small-sized, non-polar pesticides from the coextracted fatty materials. Clean extracts were analyzed by GC coupled to ECD and NPD.

Large volumes of aqueous samples (i.e. 100 mL) have been preconcentrated in-line in a precolumn (2.0×1.0 mm ID) or SPE disks and dried under a nitrogen flow at room temperature.[48] This approach allows very low detection limits (10 ppt) to be reached for pesticides and herbicides in aqueous samples. After drying, desorption with a supercritical fluid to the

analytical column was performed by switching restrictors mounted on valves. During desorption, analytes were focused on the top of the analytical column and the final determination was carried out by switching valves to transfer the extract to the analytical column.

3.2 Capillary Column Supercritical Fluid Chromatography

Most of the cSFC applications on the determination of pesticides published until the end of the last decade have used open-tubular small-diameter columns (typically 50–100 μm ID). More recently the interest was focused on pSFC because of its easier operation, higher sample capacity and less instrumental constraints. In fact, the low flow rates used in cSFC demand the use of high-accuracy mobile phase delivery and flow control by fixed restrictors that do not allow its regulation independently of pressure or density programming. Therefore, the high theoretical efficiencies of the chromatographic column are partially lost during the density or pressure programming combined with the higher than optimum flow rates used in order to reduce the analysis time. Since multiresidue pesticide analysis demands high resolution, cSFC is not well suited for the determination of complex mixtures of pesticides analyzed in screening analysis. pSFC using multiple columns or packed microbore columns is more suitable for the determination of pesticides in real samples. Nevertheless, the easier interface of cSFC to GC-like detectors because of its low flow rates makes this chromatographic technique still attractive in the determination of selected classes of pesticides. In particular, packed capillary columns appear to be an area for development because analysis time, resolution and sample capacity are optimized.

3.2.1 Injection Systems

Theoretical considerations show that there are stringent requirements for the allowable injection volume in open-tubular cSFC. In this regard, injections lower than 96 nL should be performed in order to prevent no more than 1% loss of resolution in a typical column (10 m × 50 μm ID).[56] Typical injector hardware used in cSFC is a high-pressure valve equipped with a low-volume internal loop (hundreds of nanoliters). Nevertheless, this volume is still too high for most analytical columns and different injection systems have been developed to allow the injection of small sample volumes (dynamic split, delayed split and timed split). However, low sensitivities and relatively poor reproducibility are usually achieved in these split injectors and solvent venting systems have been developed in order to allow the injection of larger sample volumes such as solvent venting in a precolumn, or with gas purging and solvent backflush.[58–64]

3.2.2 Chromatographic Conditions

Analytical columns used in cSFC are narrow-bore open-tubular columns usually 25–100 μm ID coated with bonded or highly cross-linked stationary phases to prevent the washout during the SFC programming conditions (Table 4). The most used stationary phase in the determination of pesticides is the substituted highly cross-linked polysiloxanes containing methyl or phenyl substituents because resolution is based on the high efficiency of the narrow-bore capillary columns. More polar stationary phases like cyanopropyl exhibited poorer coating efficiency because the low diameter of the

Table 4 Chromatography systems used in cSFC

Pesticide class	Matrix	Column dimensions	Stationary phase[a]	Mobile phase/ temperature	Analytical conditions	Ref.
Carbamate and phenoxyacids	Standard	1.5 m × 25 μm ID, 0.15 μm	5% Ph	CO_2 100 °C	135–350 atm, 50 atm min^{-1}	61
Carbamate	Bell pepper	10 m × 50 mm ID,	100% Me	CO_2	–	62
Linuron	Wheat grain	10 m × 100 μm ID, 0.5 μm	100% Me	CO_2 110 °C	0.2–0.6 g mL^{-1} min^{-1}	65
Carbamate	Parsley	10 m × 50 μm ID, 0.25 μm	100% Me	CO_2 100 °C	0.2–0.75 g mL^{-1} min^{-1}	64
Multiple classes	Standard	6 m × 50 μm ID, 0.25 μm	5% Ph	CO_2 100 °C	0.3–0.4 g mL^{-1} min^{-1}	65
Phenoxy acids	Soil extract	20 m × 50 μm ID, 0.25 μm	100% Me	CO_2 100 °C	120–280 g mL^{-1} min^{-1}	66
Pyrethrins and pyrethroids	Standard			CO_2 90 °C/ 130–80 °C	110–220 g mL^{-1} min^{-1}	67

[a] Unless specified, stationary phases are polysiloxanes. Ph: phenyl, Me: methyl.

column creates instability in the film distribution along the column. Despite this, polar stationary phases are useful because they provide a higher selectivity in comparison to nonpolar stationary phases.

Another important aspect is the column surface inertness. It must be minimized by deactivation to avoid the reversible or irreversible adsorption of polar pesticides which is particularly apparent at the low amounts of compound injected in cSFC. Table 4 reports some of the applications developed in pesticide determination using cSFC and the analytical conditions used. Density or pressure programming is used in most applications in order to increase the mobile phase solvating power, allowing the elution of a variety of pesticides. Alternatively, negative temperature gradient programming keeping isobaric conditions has been used for the elution of pyrethrins and pyrethroids.[67]

Chlorinated Pesticides. These show a strong selectivity according to the chromatographic temperature since these analytes possess a remarkable vapor pressure. At high temperatures, chlorinated pesticide retention is mainly controlled by temperature, whereas at low temperatures fluid solubility is the more prevalent property affecting the retention of these analytes. A comparison between the effect of temperature in pSFC and cSFC shows that it is more relevant in the former columns.[60]

Chromatographic reproducibility in terms of retention time, peak area and peak height has been studied in detail with respect to routine analysis of agricultural analytes. Typical reproducibilities using cSFC for herbicide standards at different concentrations ($10–100\,\mu g\,mL^{-1}$) were 2–3% RSD (relative standard deviation) based on peak area and peak heights ($n = 5$).[60] Direct injection (without splitting) using a loop size of 1 mL gave a retention time reproducibility of 0.1% RSD.

3.2.3 Detection Systems

cSFC has great detector flexibility because of the low flow rates and in most applications neat CO_2 is used as the mobile phase. Interfaces for cSFC coupling in ECD, TID, NPD, FPD, SCD, UV, MS, and FTIR have been developed.[61]

UV absorbance detection has been applied to the detection of polar pesticides using modifiers to elute polar analytes that are difficult to elute from the analytical column by density programming. The photodiode array detector has been interfaced to cSFC for the sensitive determination of pesticides and herbicides.[66] The sensitivity is similar to single channel UV but allows the deconvolution of overlapped chromatographic peaks and peak confirmation. The LODs of SFC/UV were compared to those of LC/UV for analyzing sulfonylurea herbicides.[67] The sensitivity of both chromatographic systems was comparable due to the broad UV transparent window of CO_2 in comparison to organic solvents.

NPD was applied to the determination of carbamates and other nitrogen-containing pesticides by using neat CO_2 as the mobile phase. The method has been applied to standard mixtures and real samples reaching LODs below 1.6 ppb in complex matrices.[68]

The flame photometric detector is element selective to sulfur, phosphorus or tin depending on the bandpass filter mounted between the flame and the photomultiplier. However, when CO_2 is used as the mobile phase, signal quenching occurs because the luminescence of organosulfur compounds is at the same wavelength that CO_2 absorbs. Therefore, the application of FPD has been restricted to organophosphorus[69] and organotin[70] compounds. An increased sensitivity in tin-selective FPD was obtained when large hydrogen flow rates were used, reaching detection limits in the low picogram range (Figure 6). Different detector

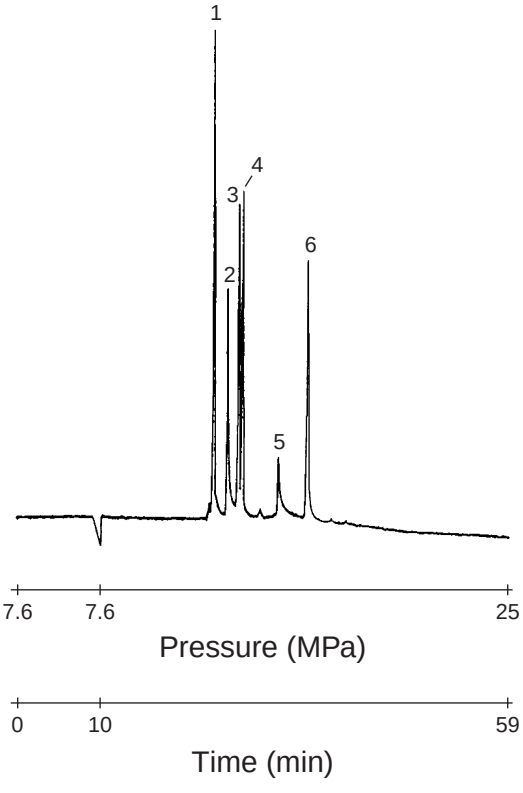

Figure 6 cSFC chromatograms of organotin chlorides obtained with tin-selective FPD at 50 °C. Injection was performed at 76 atm and then programmed to 250 atm at 3.5 atm min^{-1} after a 10-min isobaric period. Compound identification: 1, TPrTCl; 2, DBTCl; 3, TBTCl; 4, TeBTCl; 5, DPhTCl; 6, TPhTCl. (Reproduced from Dachs and Bayona[70] by permission of Elsevier.)

configurations have been used for detection in single, dual and pulsed FPD.[71] Selective and sensitive determination of labile sulfur-containing pesticides has been achieved by using SCD (see section 2.1.2).[71] Its detection limit was 12 pg, it exhibited three orders of magnitude of linearity and the sulfur to carbon selectivity was 10^7. The limited signal quenching allows the use of moderate amounts of polar modifiers, enabling the elution of polar sulfur-containing pesticides.

The selective determination of OCPs has been carried out using two different detectors (ECD[72] and TID[73]) in a nitrogen atmosphere. Detection limits of both detectors were strongly compound dependent and ranged from high femtogram to low picogram. The dynamic range of ECD is four orders of magnitude. The Fourier transform ion mobility detector response operated in the negative ion collection mode and a commercially available ECD detector were compared for organochlorine compounds.[74] Pressure programming led to a baseline drift in the electron capture detector in contrast with the Fourier transform ion mobility detector that always exhibited a stable baseline.

Several classes of thermal labile pesticides such as carbamates or nonvolatile phenoxyacids have been successfully analyzed by cSFC/MS in the chemical ionization mode. Accordingly, different sensitivities and fragmentation patterns in mass spectra can be obtained depending on the reagent gas used. In this regard, when ammonia is used, less fragmentation was obtained, showing a base peak corresponding to $(M + 18)^+$ in comparison with methane which led to a base peak corresponding to the substituent cleavage.[75]

cSFC/MS has been used in the determination of OCPs using a double focusing instrument and different ionization modes. Detection limits were in the low nanogram level independently of the ionization procedure (positive or negative ion) used.[76] Positive-ion methane chemical ionization resulted in a relatively abundant $[M + H]^+$ ion, whereas positive ion isobutane and ammonia appeared not to be amenable to the detection of chlorinated pesticides. The electron impact charge exchange mass spectra of the pesticides did not match the library spectra. CO_2 was an efficient moderating gas giving relatively large amounts of $[M]^-$, in addition to some fragment ions.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
CE	Capillary Electrophoresis
cSFC	Capillary Column Supercritical Fluid Chromatography
DDE	Dichlorodiphenylethylene
ECD	Electron Capture Detection
FPD	Flame Photometric Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
HFBA	Heptafluorobutyric Acid
ID	Internal Diameter
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
LVI	Large Volume Injection
MS	Mass Spectrometry
NMC	*N*-Methylcarbamate
NPD	Nitrogen–Phosphorus Detection
OCP	Organochlorine Pesticide
OPP	Organophosphate Pesticide
PCB	Polychlorinated Biphenyl
PDD	Pulse Discharge Detection
PFBBr	Pentafluorobenzylbromide
PFPD	Pulsed Flame Photometric Detection
PID	Photoionization Detection
pSFC	Packed Column Supercritical Fluid Chromatography
PTV	Programmable Temperature Vaporizer
RSD	Relative Standard Deviation
SCD	Sulfur Chemiluminescence Detection
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TID	Thermionic Detection
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
2,4′-DDT	2,4′-Dichlorodiphenyltrichloroethane

RELATED ARTICLES

Pesticides **(Volume 7)**
Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of

Pesticides cont'd **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

REFERENCES

1. C. Tomlin (ed.), *The Pesticide Manual*, 11th edition, The Royal Society of Chemistry, Cambridge, 1997.
2. J.L. Tadeo, C. Sánchez-Brunete, A.I. García-Valcárcel, L. Martínez, R.A. Pérez, 'Determination of Cereal Herbicide Residues in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **754**, 347–365 (1996).
3. S.S. Yang, A.I. Goldsmith, I. Smetena, 'Recent Advances in the Residue Analysis of *N*-methylcarbamate Pesticides', *J. High Res. Chromatogr. A*, **754**, 3–16 (1996).
4. K. Grob, 'Injection Techniques in Capillary GC', *Anal. Chem.*, **66**(20), 1009A–1019A (1994).
5. K. Grob, M. Bossard, 'Effect of Dirt on Quantitative Analyses by Capillary Gas Chromatography with Splitless Injection', *J. Chromatogr.*, **294**, 65–75 (1984).
6. J. Hajslová, K. Holadová, V. Kocourek, J. Poustka, M. Godula, P. Cuhra, M. Kempny, 'Matrix-induced Effects: a Critical Point in the Gas Chromatographic Analysis of Pesticides Residues', *J. Chromatogr. A*, **800**, 283–295 (1998).
7. J. Pawliszyn, *Applications of Solid Phase Microextraction*, Royal Society of Chemistry, Cambridge, 1999.
8. G.A. Jongenotter, M.A.T. Kerkhoff, H.C.M. van der Knaap, B.G.M. Vandeginste, 'Automated On-line GPC/GC/FPD Involving Co-solvent Trapping and the On-column Interface for the Determination of Organophosphorus Pesticides in Olive Oils', *J. High Resolut. Chromatogr.*, **22**(1), 17–23 (1999).
9. E. Pocurull, C. Aguilar, F. Borrull, R.M. Marcé, 'On-line Coupling of Solid-phase Extraction to Gas Chromatography with Mass Spectrometric Detection to Determine Pesticides in Water', *J. Chromatogr. A*, **818**, 85–93 (1998).
10. J.V. Hinshaw, 'Large-volume Injection in Capillary GC: Problems and Solutions', *LC–GC*, **2**(2), 24–30 (1989).
11. R.J.C.A. Steen, I.L. Freriks, W.P. Cofino, U.A.Th. Brinkman, 'Large-volume Injection in Gas Chromatography – Ion Trap Tandem Mass Spectrometry for the Determination of Pesticides in the Marine Environment at the Low ng/L Level', *Anal. Chim. Acta*, **353**, 153–163 (1997).
12. F.J. López, J. Beltrán, M. Forcada, F. Hernández, 'Comparison of Simplified Methods for Pesticide Residue Analysis. Use of Large-volume Injection in Capillary Gas Chromatography', *J. Chromatogr. A*, **823**, 25–33 (1998).
13. T. Suzuki, K. Yaguchi, K. Ohnishi, T. Yamagishi, 'Determination of Pesticides in Water by Capillary Gas Chromatography with Splitless Injection of Large Sample Volumes', *J. Chromatogr. A*, **662**, 139–146 (1994).
14. W. Vetter, V. Schurig, 'Enantioselective Determination of Chiral Organochlorine Compounds in Biota by Gas Chromatography on Modified Cyclodextrins', *J. Chromatogr. A*, **774**, 143–175 (1997).
15. F. Sánchez-Rasero, M.B. Matallo, G. Dios, E. Romero, A. Peña, 'Simultaneous Determination and Enantiomeric Resolution of Mecoprop and Dichloroprop in Soil Samples by High-performance Liquid Chromatography and Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **799**, 355–360 (1998).
16. W.M. Draper, 'Optimizing Nitrogen–Phosphorus Detector Gas Chromatography for Pesticide Analysis', *J. Agric. Food Chem.*, **43**, 2077–2082 (1995).
17. C. Sánchez-Brunete, R.A. Pérez, E. Miguel, J.L. Tadeo, 'Multiresidue Herbicide Analysis in Soil Samples by Means of Extraction in Small Columns and Gas Chromatography with Nitrogen–Phosphorus and Mass Spectrometric Detection', *J. Chromatogr. A*, **823**, 17–24 (1998).
18. A.I. García-Valcárcel, C. Sánchez-Brunete, L. Martínez, J.L. Tadeo, 'Determination of Dinitroaniline Herbicides in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **719**, 113–119 (1996).
19. Y. Nakamura, Y. Tonogai, Y. Sekiguchi, Y. Tsumura, N. Nishida, K. Takakura, M. Isechi, K. Yuasa, M. Nakamura, N. Kifune, K. Yamamoto, S. Terasawa, T. Oshima, M. Miyata, K. Kamakura, Y. Ito, 'Multiresidue Analysis of 48 Pesticides in Agricultural Products by Multiresidue Gas Chromatography', *J. Agric. Food Chem.*, **42**, 2508–2518 (1994).
20. H. Sabik, R. Jeannot, 'Determination of Organonitrogen Pesticides in Large Volumes of Water by Liquid–Liquid and Solid Phase Extraction Using Gas Chromatography with Nitrogen Phosphorus Detection and Liquid Chromatography with Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **818**, 197–207 (1998).
21. S. Pérez, J.M. García-Baudín, J.L. Tadeo, 'Determination of Chlorotoluron, Isoproturon and Metoxuron in Soil by GLC-NPD and Confirmation Using GLC-MS', *Fresenius J. Anal. Chem.*, **339**, 413–416 (1991).
22. L. Kadenczki, Z. Arpad, I. Gardi, A. Ambrus, L. Gyorfi, G. Reese, W. Ebing, 'Column Extraction of Residues of Several Pesticides from Fruits and Vegetables: a Simple Multiresidue Analysis Method', *J. AOAC Int.*, **75**(1), 53–61 (1992).

23. S.M. Lee, P.L. Wylie, 'Comparison of the Atomic Emission Detector to Other Element-selective Detectors for the Gas Chromatographic Analysis of Pesticides Residues', *J. Agric. Food Chem.*, **39**, 2192–2199 (1991).
24. J. Tekel, S. Hatrík, 'Pesticide Residue Analyses in Plant Material by Chromatographic Methods: Clean-up Procedures and Selective Detectors', *J. Chromatogr. A*, **754**, 397–410 (1996).
25. M.S. Klee, M.D. Williams, I. Chang, J. Murphy, 'Superior ECD Performance Through Design and Application', *J. High Resolut. Chromatogr.*, **22**(1), 24–28 (1999).
26. A. Amirav, H. Jing, 'Pulsed Flame Photometer Detector for Gas Chromatography', *Anal. Chem.*, **67**, 3305–3318 (1995).
27. A. Amirav, H. Jing, 'Pulsed Flame Photometric Detector for Gas Chromatography', Technical Report, Tel Aviv University, 1999.
28. A.R. Fernández-Alba, A. Agüera, M. Contreras, G. Peñuela, I. Ferrer, D. Barceló, 'Comparison of Various Sample Handling and Analytical Procedures for the Monitoring of Pesticides in Groundwaters', *J. Chromatogr. A*, **823**, 35–47 (1998).
29. E.A. Stemmler, R.A. Hites, 'The Fragmentation of Negative Ions Generated by Electron Capture Negative Ion Mass Spectrometry: a Review with New Data', *Biomed. Environ. Mass Spectrom.*, **17**, 311–328 (1988).
30. H.-J. Stan, G. Kellner, 'Negative Chemical Ionization Mass Spectrometry of Organophosphorus Pesticides', *Biomed. Mass Spectrom.*, **9**, 483–492 (1982).
31. R.O. Lidgard, A.M. Duffield, R.J. Wells, 'Positive and Negative Ion Chemical Ionization Mass Spectrometry of Pyrethrin Pesticides', *Biomed. Environ. Mass Spectrom.*, **13**, 677–680 (1986).
32. E.A. Stemmler, R.A. Hites, B. Arbogast, W.L. Budde, M.L. Deinzer, R.C. Dougherty, J.W. Eichelberger, R.L. Foltz, C. Grimm, E.P. Grimsrud, C. Sakashita, L.J. Sears, 'Interlaboratory Comparison of Methane Electron Capture Negative Ion Mass Spectra', *Anal. Chem.*, **60**, 781–787 (1988).
33. R.S. Sheridan, J.R. Meola, 'Analysis of Pesticide-residues in Fruits, Vegetables and Milk by Gas Chromatography/Tandem Mass-spectrometry', *J AOAC Int.*, **82**, 982–990 (1999).
34. T. Hankemeier, E. Hooijschuur, R.J.J. Vreuls, U.A.T. Brinkman, T. Visser, 'On-line Solid-phase Extraction Gas-chormatography Cryotrapping Infrared Spectrometry for the Trace-level Determination of Microcontaminants in Aqueous Samples', *J. High Resolut. Chromatogr.*, **21**, 341–346 (1998).
35. A.J.H. Louter, J.J. Vreuls, U.A.T. Brinkman, 'On-line Combination of Aqueous-sample Preparation and Capillary Gas-chromatography', *J. Chromatogr.*, **842**, 391–426 (1999).
36. S.J. Crook, 'The Application of SPME to Pesticide Residue Analysis', in *Applications of Solid-phase Microextraction*, ed. J. Pauliszyn, Royal Society of Chemistry, Cambridge, 188–200, 1999.
37. J.R. Barr, V.E. Green, Ch.R. Lapeza, Jr, V.L. Maggio, W.E. Turner, A.R. Woolfit, J. Grainger, L.L. Needman, D.G. Patterson, Jr, 'New Mass Spectrometric Approaches for the Analysis of PCBs and Persistent Pesticides in Human Serum', *Organohal. Comp.*, **31**, 276–281 (1997).
38. I. Mukherjee, M. Gopal, 'Chromatographic Techniques in the Analysis of Organochlorine Pesticide Residues', *J. Chromatogr. A*, **754**, 33–42 (1996).
39. Y. Ling, I.-P. Yuang, 'Multi-residue Matrix Solid-phase Dispersion Method for the Determination of Six Synthetic Pyrethroids in Vegetables Followed by Gas Chromatography with Electron Capture Detection', *J. Chromatogr. A*, **695**, 75–82 (1995).
40. T.A. Berger, *Packed Column SFC*, ed. R.M. Smith, RSC Chromatography Monographs, Royal Society of Chemistry, Cambridge, 1995.
41. B. Wenclawiak, J. Schipke, 'Supercritical Fluid Chromatography of Organochlorine Pesticides on Different Columns', *Fresenius J Anal. Chem.*, **344**, 521–523 (1992).
42. T.A. Berger, W.H. Wilson, J.F. Deye, 'Analysis of Carbamate Pesticides by Packed Column Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **32**, 179–184 (1994).
43. T.A. Berger, 'Separation of Phenylurea Herbicides by Packed Column Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **32**, 25–29 (1994).
44. M.E. McNally, J.R. Wheeler, 'Comparison of Packed Column Supercritical Fluid Chromatography and High-performance Liquid Chromatography Using Representative Herbicides and Pesticides as Typical Moderate Polarity and Molecular Weight Range Molecules', *J. Chromatogr.*, **410**, 343–353 (1988).
45. S. Shah, L.T. Taylor, 'Analysis of Triazine and Triazole Herbicides by Gradient-elution Supercritical Fluid Chromatography', *J. Chromatogr.*, **505**, 293–298 (1990).
46. T. Yarita, A. Nomura, Y. Horimoto, J. Yamada, 'Determination of Chlorinated Pesticides by Packed-column Supercritical-fluid Chromatography Electron Capture Detector', *J. Microcol. Sep.*, **49**, 145–149 (1994).
47. B.N. Zegers, A.C. Hogenboom, S.E.G. Dekkers, H. Lingeman, U.A.T.H. Brikman, 'Packed Capillary Supercritical Fluid Chromatography of Organophosphorus Pesticides: Selective Detection Applications', *J. Microcol. Sep.*, **6**, 55–62 (1994).
48. T.A. Berger, 'Feasibility of Screening Large Aqueous Samples for Thermally-unstable Pesticides Using High Efficiency Packed Column Supercritical Fluid Chromatography with Multiple Detectors', *Chromatographia*, **41**, 471–484 (1995).
49. P.R. Massey, M.J. Tandy, 'Challenges and Frustrations in the Separation and Analysis of Chiral Agrochemicals', *Chirality*, **6**, 63–71 (1994).

50. C. Bicchi, C. Balbo, A. Damato, O. Panero, 'SFC/UV Determination of Dibenzuron and Triflumuron in Apple and Pear Pulps for Baby Food', *Chromatographia*, **43**, 439–443 (1996).
51. J.T.B. Strode, L.T. Taylor, 'Optimization of Electron-capture Detector When Using Packed Column Supercritical Fluid Chromatography with Modified Carbon Dioxide', *J. Chromatogr.*, **723**, 361–369 (1996).
52. H. Shi, L.T. Taylor, E.M. Fujinari, X. Yan, 'Sulfur-selective Chemiluminescence Detection with Packed-column Supercritical Fluid Chromatography', *J. Chromatogr.*, **779**, 307–313 (1997).
53. B.N. Zegers, R. Hessels, J. Jagesar, J. Rozenbrand, H. Lingeman, U.A.T. Brinkman, 'Photoionization Detection in Packed Capillary and Supercritical Fluid Chromatography', *J. Liq. Chromatogr.*, **18**, 413–440 (1995)
54. S. Nelieu, M. Stobiecki, F. Sadoun, H. Virelizier, L. Kerhoas, J. Einhorn, 'Solid Phase Extraction and LC/MS or SFC/MS for the Analysis of Atrazine Metabolites in Water', *Analysis*, **22**, 70–75 (1994).
55. C.H. Kirschner, L.T. Taylor, 'Recent Advances in Sample Introduction for Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr.*, **16**, 73–84 (1993).
56. M.L. Lee, K.E. Markides, *Analytical Supercritical Fluid Chromatography and Extraction*, Chromatography Conf. Inc., Provo, 1990.
57. K.S. Nam, J.W. King, 'Coupled SFE/SFC/GC for the Trace Analysis of Pesticide-residues in Fatty Food Samples', *J. High Resolut. Chromatogr.*, **17**(8), 577–582 (1994).
58. B.N. Zegers, H.J. Degeus, S.H.J. Wildenburg, H. Lingeman, U.A.T. Brinkman, 'Large-volume Injection in Packed-capillary Supercritical Fluid Chromatography', *J. Chromatogr.*, **677**, 141–150 (1994).
59. B.E. Berg, T. Greibrokk, 'The Solvent Venting Injection Technique in Capillary Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr.*, **12**, 322–326 (1989).
60. D.E. Knowles, B.E. Richter, M.R. Andersen, D.W. Later, 'Supercritical Fluid Chromatography: a New Approach to Pesticide Analysis', *Chimicaoggi*, **Jan/Feb**, 11–15, (1989).
61. D.W. Later, D.L. Bornhop, E.D. Lee, J.D. Henion, R.C. Wieboldt, 'Detection Techniques for Capillary Supercritical Chromatography', *LC–GC Int.*, **1**, 36–42 (1987).
62. Y. Hirata, M. Tanaka, K. Inomata, 'Purged Splitless Injection in Microcolumn Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **27**, 395–398 (1989).
63. Y. Hirata, K. Inomata, 'Splitless Injection Method with Mixing Chamber in Capillary Supercritical Fluid Chromatography', *J. Microcol. Sep.*, **1**, 242–248 (1989).
64. S.B. Hawthorne, D.J. Miller, 'An Almost On-column Injector for Capillary Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **27**, 197–202 (1989).
65. B. Wenclawiak, A. Otterbach, M. Krappe, 'Capillary Supercritical Fluid Chromatography of Pyrethrins and Pyrethroids with Positive Pressure and Negative Temperature Gradients', *J. Chromatogr.*, **799**(1–2), 265–273 (1998).
66. J.E. France, K.J. Vorhees, 'Capillary Supercritical Fluid Chromatography with Ultraviolet Multichannel Detection of Some Pesticides and Herbicides', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**, 692–696 (1988).
67. T. Greibokk, B.E. Berg, A.L. Blilie, J. Doehl, A. Farbot, E. Lundanes, 'Techniques and Applications in Supercritical Fluid Chromatography', *J. Chromatogr.*, **394**, 429–441 (1987).
68. B.E. Richter, M.R. Andersen, D.E. Knowles, E.R. Campbell, N.L. Porter, L. Nixon, D.W. Later, *ACS Symp. Ser.*, **366**, 179 (1988).
69. H. Koizumi, Y. Suzuki, 'Capillary Supercritical Fluid Chromatography with Flame Photometric Detection Using a Large Bore Column SFC Pumping System', *J. High Resolut. Chromatogr.*, **14**, 257–261 (1991).
70. J. Dachs, J.M. Bayona, 'Optimization of a Flame Photometric Detector for Supercritical Fluid Chromatography of Organotin Compounds', *J. Chromatogr.*, **636**, 277–283 (1993).
71. H.C.K. Chang, L.T. Taylor, 'Use of Sulfur Chemiluminescence Detection after Supercritical Fluid Chromatography', *J. Chromatogr.*, **517**, 491–501 (1990).
72. H.-C.K. Chang, L.T. Taylor, 'The Performance of Electron-capture Detection after Capillary Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **28**, 29–33 (1990).
73. R.K. Jew, B.E. Richter, 'Thermionic Ionization Detection of Halogenated Compounds Using Supercritical Fluid Chromatography', *J. Microcol. Sep.*, **5**, 41–45 (1993).
74. E.E. Tarver, H.H. Hill, 'Comparison of a Pulsed Electron-capture Detector and a Fourier Transform Ion Mobility Detector after Capillary Fluid Chromatography', *Fresenius J. Anal. Chem.*, **344**, 453–459 (1992).
75. B.W. Wright, H.T. Kalinoski, H.R. Udseth, R.D. Smith, 'Capillary Supercritical Fluid Chromatography–Mass Spectrometry', *J. High Resolut. Chromatogr.*, **9**, 145–152 (1986).
76. A. Jablonska, M. Hansen, D. Ekeberg, E. Lundanes, 'Determination of Chlorinated Pesticides by Capillary Supercritical Fluid Chromatography Mass Spectrometry with Positive-ion and Negative-ion Detection', *J. Chromatogr.*, **647**, 341–350 (1993).

Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Emilio Benfenati, Marco Natangelo, and Simona Tavazzi
Istituto di Ricerche Farmacologiche ''Mario Negri'', Milan, Italy

1 Introduction 1

2 Application of the Main Techniques for Pesticide Analysis in Gas Chromatography/Mass Spectrometry 2
2.1 Electron Ionization 2
2.2 Chemical Ionization 3
2.3 Tandem Mass Spectrometry 4
2.4 Derivatization 4
2.5 Ion Analyzers 4

3 Identification of Pesticides and Related Products by Gas Chromatography/Mass Spectrometry 6
3.1 Pesticides 6
3.2 Transformation Products of Pesticides 6
3.3 Impurities in Pesticides 7

4 Quantification of Pesticides by Gas Chromatography/Mass Spectrometry 7
4.1 General Characteristics of the Method and Comparison with Other Methods 7
4.2 Detection Limits and Ranges 9
4.3 Isotope Dilution Mass Spectrometry 9

5 Application to Different Matrices 9
5.1 Environmental Matrices 9
5.2 Food Samples 13
5.3 Other Matrices 14

6 Applications for Different Pesticide Classes 14
6.1 Introduction 14
6.2 Multiclass Methods 15
6.3 Organochlorine Pesticides 16
6.4 Organophosphorus Pesticides 16
6.5 Herbicides 17
6.6 Pyrethroids 17
6.7 Carbamates and Related Pesticides 18
6.8 Fungicides 18
6.9 Fumigants 19
6.10 Organotin Compounds 19

7 Conclusions 19

CAS Numbers 19

Acknowledgments 20

Abbreviations and Acronyms 20

Related Articles 21

References 21

Gas chromatography/mass spectrometry (GC/MS) is irreplaceable in pesticide analysis. Nowadays international bodies consider mass spectrometry (MS) the method of choice for pesticide analysis and quantitative methods using GC/MS are more and more numerous.[1,2] The key characteristics of GC/MS in pesticide analysis are its selectivity and sensitivity, but other important features will be considered in detail in the specific sections on identification and quantitative analysis.

A great advantage of GC/MS in identification is that it gives a mass spectrum that can be easily and quickly compared on-line with a library of more than 200 000 mass spectra. Specific pesticide libraries are available with hundreds of compounds. Mass spectra can be obtained with just a few picograms of compound injected and gas chromatography (GC) can separate analytes. This is an unbeatable feature of GC/MS.

In quantitative analysis, sensitivity is often excellent and selectivity is better than with other GC detectors. Multiresidue analyses can easily be done. The use of internal standards labeled with stable isotopes allows an accuracy which is not achieved with common methods. These characteristics facilitate sample processing because cleanup stages are not necessary in many cases.

Major limitations in the use of GC/MS depend on GC limits. Thus, very polar and thermally labile pesticides, not suitable for GC analysis, cannot be analyzed by GC/MS. Other limitations are the cost of the apparatus and the relative complexity of the method, compared with the simpler GC detectors. However, these limitations are gradually becoming less important as GC/MS instruments become simpler and cheaper.

1 INTRODUCTION

Pesticides are ubiquitous pollutants and have effects on the environment and human beings even in minute amounts. Fortunately, this is not the case for all compounds, but this category has nevertheless been viewed with suspicion for years. Since the adverse

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

activity of pesticides may appear at low doses, analysts have been pushed to develop sensitive methods for analyzing them. However, the fact that there are hundreds of pesticides belonging to many different chemical classes complicates analysis. Furthermore, many more compounds originating from the parent pesticide may be present and cause adverse effects.

Selective detectors offer good sensitivity and enough specificity towards certain classes of pesticides. For instance, electron capture detection (ECD) is widely used for chlorinated pesticides, while nitrogen–phosphorus detection (NPD) is responsive to nitrated and phosphorated compounds. The selectivity of ECD and NPD limits their applicability to specific pesticide chemical classes, but this is not the case with GC/MS. Furthermore, even selective detectors may be not selective enough since many interfering compounds respond to ECD or NPD. GC/MS overcomes the selectivity of these detection methods, offering a solution.

The so-called interfering peaks may actually be related to the pesticide of interest, owing to metabolism or transformation, for instance. Selective detectors rely on a specific response for the analyte, indicating that an atom of a given class is present in the molecule. GC/MS can do the same but better. Thus GC/MS has been used to analyze chlorinated and brominated pesticides, looking at chlorine or bromine.[3] GC/MS can detect these atoms without any possibility of error, because it measures both isotopes of these atoms, while ECD can detect a halogen ''aspecifically'', but we cannot establish whether it is chlorine, bromine, fluorine or something else that interferes.

However, the real power of GC/MS goes much further. It can consider not only a specific atom, but also a molecule (as molecular ion) or a molecular (charged) fragment, characteristic of the pesticide. This further increases the specificity. Then, in addition, we can look at many of these characteristic ions for a single molecule, improving the specificity and obtaining a sort of fingerprint. As regards specificity, GC/MS is also superior to immunoassays, which tend to be affected more by interfering compounds (cross-reactivity). In fact, MS is considered the validating technique.[1,4]

GC/MS is not just a sophisticated procedure to validate other simpler methods. It is itself a technique which may became a popular routine method for pesticide analysis. There may be problems (for instance, due to low volatility and thermal instability of the analytes, the cost of the instrument and the complexity of the method. However, the first point is common to all GC methods, while the gap between GC/MS and the other techniques is closing as regards cost and complexity.)

There are several kinds of GC/MS instruments, several ionization modes and techniques, and different analytical approaches. Thus different results can be obtained for the same pesticide, useful in different situations. There are magnetic instruments, quadrupoles and ion traps. The scope of electron ionization (EI) and chemical ionization (CI) will be discussed. The different techniques will be illustrated for identification purposes in section 3 and for quantitative analyses in section 4. GC/MS can be used to identify pesticides, considering the whole mass spectrum, in order to have full details of the fragmentations, or only selected ions for more sensitive and accurate quantitative methods. Applications will be illustrated in section 5 as far as regards the different matrices and in section 6 considering the different pesticide classes.

2 APPLICATION OF THE MAIN TECHNIQUES FOR PESTICIDE ANALYSIS IN GAS CHROMATOGRAPHY/MASS SPECTROMETRY

2.1 Electron Ionization

In EI (previously called electron impact), the analyte is introduced into the source, which is under vacuum, and bombarded with electrons, which break the molecules and give many charged species. For MS the positive ions are the most useful. A more detailed description of MS is given in the article **Electron Ionization Mass Spectrometry**. For the introduction of the analyte GC is employed in GC/MS, which separates the mixture of compounds present in a typical pesticide analysis from an environmental sample, a biota or a food matrix. Further details on GC/MS are given in the article **Gas Chromatography/Mass Spectrometry**. GC/MS offers a unique possibility to analyze halogenated pesticides. Chlorine and bromine in the molecule are easily detected by the appearance of ions as doublets, separated by two mass units. Chlorine presents two isotopes in nature, chlorine-35 and chlorine-37, in a definite ratio (about 3 : 1), while bromine has two isotopes, bromine-79 and bromine-81, almost equally abundant. As a result, ions containing one of these atoms show a typical doublet: see Figure 1 for an example of a pesticide with one chlorine. In this figure, the ion at m/z 215 is due to the molecule of atrazine bearing the isotope of chlorine-35, while at m/z 217 the sister ion, due to the molecule with chlorine-37, appears. Similarly, the ion at m/z 200 corresponds to the loss of one methyl group from the molecular ion at m/z 215; the ion at m/z 202 corresponds to the same loss from the ion at m/z 217. All ions with chlorine (molecular or fragment ions) present this behavior (they are split into two in a typical relative abundance), while ions not containing chlorine are singlets (plus the ion at one higher mass due to carbon-13, mainly). This helps in interpreting mass spectra.

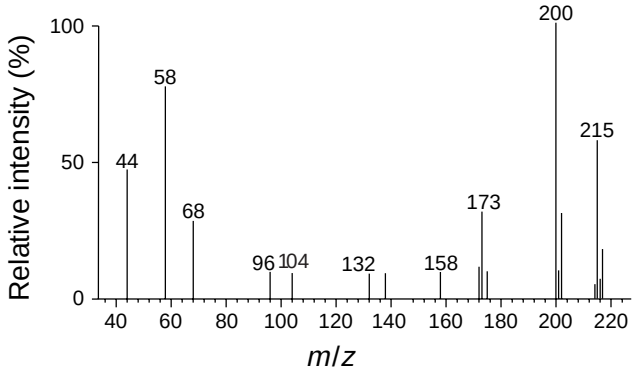

Figure 1 Mass spectrum of atrazine.

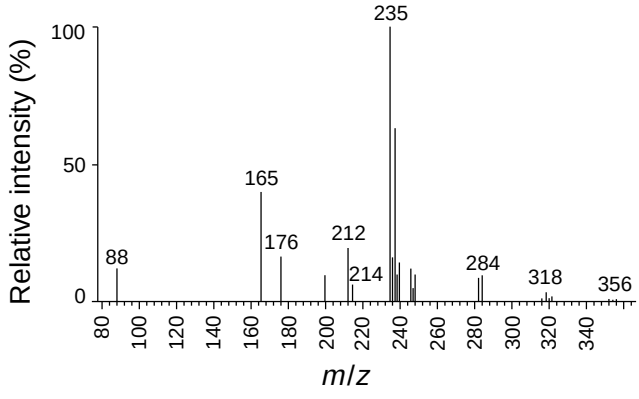

Figure 2 Mass spectrum of DDT.

If more than one of these atoms is present the situation is more complicated, but still manageable, e.g. see the mass spectrum of 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) (Figure 2). The pattern of these ions is typical of molecules with five chlorine atoms and can be easily calculated with probabilistic algorithms. From inspection of the mass spectrum the analyst can understand whether and how many chlorines and bromines are present in the pesticide (and in the individual fragment ions).

2.2 Chemical Ionization

If a softer method is used to ionize the molecules, there is less fragmentation so the mass spectrum shows a more abundant molecular ion. In CI a gas is introduced into the source, such as methane, isobutane or ammonia. Both positive and negative ions can be used. More details are given in the article **Chemical Ionization Mass Spectrometry: Theory and Applications**.

Positive ion chemical ionization (PICI) is used less than negative ion chemical ionization (NICI). NICI offers high sensitivity mainly because there is less noise, because abundant negative ions occur only in special cases. The opposite happens with PICI: generally there is more noise because of the high abundance of ions produced in the source. In PICI, addition of protons and in some cases of protonated gas may happen. The molecules give ions with multiplets in many cases, and pseudomolecular ions (at higher mass then expected). Nitrated compounds present MH^+ ions. An example is given in Figure 3(a–c).

NICI is useful if the pesticide contains chlorine or bromine or aromatic nitro compounds. An example of the same compound under different ionization conditions is given in Figure 3(a–c). CI mass spectra are less reproducible than EI spectra. NICI may be valuable for identifying these compounds, for instance in the case of transformation products (TPs) when we know that the parent compound contains these atoms or residues (see section 3.2). Halogens can be intentionally introduced into the molecule using derivatization with fluorinated residues.

CI is used for identification for specific purposes and not as a regular tool, for several reasons. It is more likely to produce different mass spectra depending on the experimental conditions and instruments. For instance, the different gases used may produce different mass spectra and the pressure of the gas itself is important. CI is less common than EI and mass spectral libraries

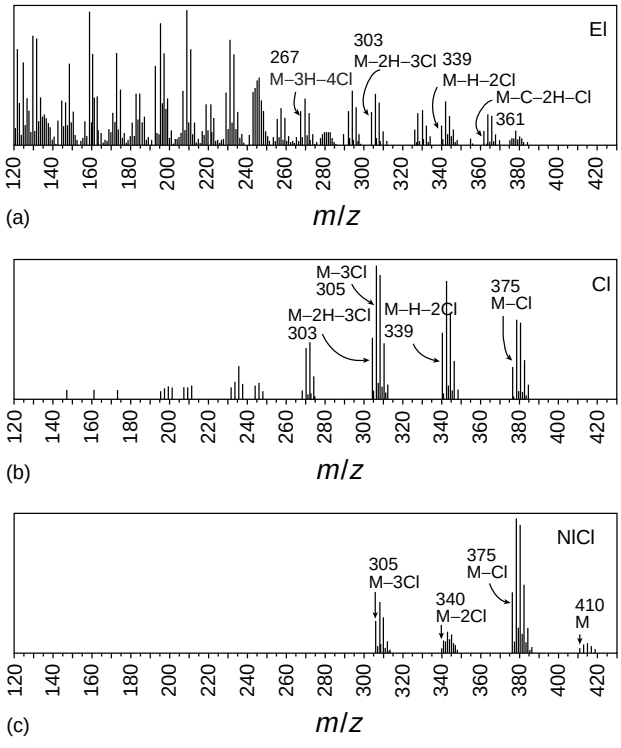

Figure 3 The mass spectrum of octachlorobornane (component of toxaphene) in three different ionization modes [(a) EI, (b) CI and (c) NICI]. (Reprinted with permission from D.L. Swackhamer, M.J. Charles, R.A. Hites, *Anal. Chem.*, **59**, 914 (1987). Copyright 1987, American Chemical Society.)

relate to EI. However, if we need confirmation of the molecular weight, PICI can be used or NICI will give useful mass spectra of halogenated pesticides, or very high sensitivities in quantitative analysis.

2.3 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) uses an ion to generate other ions as a result of fragmentation. To do this we need to separate the ion of interest from the others. Generally this is done in the MS analyzer; only the ion of interest is transmitted. Then at the exit of the first ion analyzer, the selected ion is dissociated with the help of a gas. In the case of the triple quadrupole, this is done within the second quadrupole. The selected ion produces a spectrum of ions (daughter ions), which are the ions it generates (see Figure 4a and b). This spectrum is typical of a given parent ion and informative on the structure of this ion. If a parent ion P gives a daughter ion D, the second spectrum may be highly specific and further increases the specificity of MS.

A different mass spectrometer, the ion trap, can separate the ions on a timescale, within the same physical space. Thus, by selecting appropriate potential values it is possible to allow only a specific ion to "survive" in the first phase and then to record its daughter ions in a second stage. This process can be repeated, obtaining MS^n capabilities.

MS/MS can be used for identification, to gain information on the structure of the ions. It can be used in quantitative analyses where it generally offers high sensitivity and selectivity. The selectivity comes from the typical dissociation involved in the transition from the parent ion P to the daughter ion D. The sensitivity is due to the low noise, which is typical in this case.

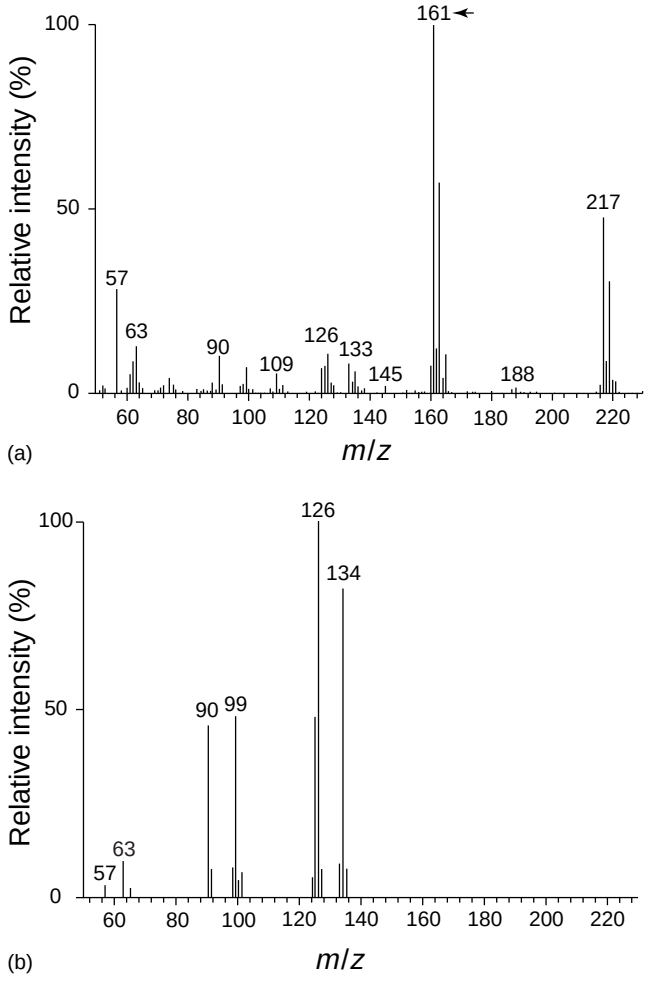

Figure 4 (a) Mass spectrum of propanil. The arrow indicates the parent ion which gave the daughter ion spectrum (b).

2.4 Derivatization

Derivatization is used in GC to produce volatile derivatives of compounds with little or no volatility. For instance, silylation is a typical procedure for derivatizing hydroxylated compounds. These may give mass spectra where the molecular ion is not present, because they undergo loss of water after EI. Hydroxylated molecules also present some problems in their chromatographic behavior, often resulting in tailing peaks with the commonly used GC column phases. Silylation, with *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) for example, can be used to characterize hydroxylated compounds better and to obtain more clearly resolved GC peaks. After derivatization, the chromatographic peak of the hydroxylated compound will disappear and a new chromatographic peak will appear, corresponding to the derivatized compound. The mass spectrum of this new compound can be interpreted considering some common losses presented by silylated compounds, such as the loss of the methyl (15 u) or trimethylsilanol (90 u). Another example is trichlorfon. This compound is unstable in GC conditions, decomposing at the injector temperature. Special conditions, such as glass wool in the injector, enhance this degradation.[5] However, if trichlorfon is derivatized with trifluoroacetic anhydride it remains stable in GC analysis and the limit of detection (LOD) is improved.

The most commonly used derivatizations for pesticides are summarized in Table 1. In the following sections we shall consider specific cases for the different pesticides.

2.5 Ion Analyzers

There is no single mass spectrum for a given compound. We shall not deal here with mass spectra obtained using interfaces which are not GC, such as liquid chromatography/mass spectrometry (LC/MS),

Table 1 Derivatization procedures for pesticide analysis in GC/MS

Chemical classes	Derivatizing reagent	Derivative product	Refs.
Phenols	Pentafluorobenzyl bromide	Pentafluorobenzyl aryl ethers	6
	Diazomethane	Anisoles	7, 8
Carboxylic pesticides	Pentafluorobenzyl bromide	Ester derivatives	9
Phenoxy acid herbicides	Pentafluorobenzylbromide	Ester derivatives	6, 10
	Benzyl bromide	Ester derivatives	11
Bentazone	Pentafluorobenzyl bromide	*N*-Fluoroaryl derivative	9
Organotin compounds	RMgX	Alkylated derivatives	12–16
Phenylurea herbicides	EtI, NaH, Me_2SO	Ethylated derivatives	17
Glyphosate, glufosinate, bialaphos and their metabolites	*N*-Methyl-*N*-(*tert*-butyldimethylsilyl)-trifluoroacetamide (MTBSTFA)	*tert*-Butyldimethylsilyl esters	18
	Fluorinated alcohols, perfluorinated anhydrides	Ester fluoroamide derivative	19
Urea herbicides	Trimethylanilinium hydroxide or trimethylsulfonium hydroxide		20
	Acetic anhydride	Acetyl derivate of the corresponding amine	21
Sulfonylurea herbicides	Diazomethane	*N*,*N*′-Dimethyl derivatives	22, 23
Trichlorphon	Acetic anhydride	Acetyl derivative	5
Carbamate pesticides	Acetic anhydride	Acetyl derivatives of the original molecule, or of a by-product	21

fast atom bombardment (FAB), capillary electrophoresis/mass spectrometry (CE/MS) and others, and shall confine ourselves to spectra obtained under GC/MS conditions. The mass spectra of pesticides obtained under CI conditions (positive and negative), described above, generally look different from the EI mass spectra (see Figure 3a–c). EI is by far the most widely used mode, but in this case too different mass spectra can be obtained. The electron voltage can be modified, in order to obtain more information on the fragments or on the molecular ion. Using different mass spectrometers the mass spectra present some differences, mainly in the relative abundances of the ions. Magnetic instruments were widely used in the past and mass spectra libraries are based on spectra recorded in the EI mode. Some mass magnetic instruments operate at high resolution. This means that they can distinguish masses which are very close, less than 10^{-4} u. For pesticide analysis, however, high resolution is not necessary in most cases, although it certainly improves the selectivity and gives cleaner ion chromatograms, with fewer interfering peaks. Consequently, better resolution may reduce the need to purify the sample. We referred to high resolution in relation to MS. High resolution is a term sometimes used also for GC, with capillary columns, but for GC high resolution is actually common so we shall not use this term in this context.

Quadrupole instruments contributed to the spread of MS because they are less expensive. They have some advantages, such as the absence of hysteresis, which allows faster scans and the possibility of recording ions in the same group using selected ion monitoring (SIM) regardless of the mass ratio. On the other hand, quadrupoles have lower resolution (high resolution is not possible), a more limited mass range (depending on the quadrupole) and the sensitivity may be lower. As we have seen, generally there is no need for high resolution for pesticide analyses. The mass range is usually enough and this is not a limiting factor, although some exceptions may appear with very particular quadrupole instruments operating at a very limited mass range. As regards sensitivity in quantitative analysis, in most cases quadrupole instruments reach enough sensitivity for common pesticide analysis.

Recently another mass spectrometer has appeared on the market, offering a good performance/price ratio: the ion trap. This instrument has been produced for several years now and has the advantage of offering MS/MS at low cost, as described above. The ion trap can record spectra of pesticides, injecting only a few picograms of analyte in several cases. Its high scanning sensitivity reduces the need to select specific ions for selective monitoring to reach enough sensitivity in quantitative analysis. A problem arising mainly in the first models was that the spectra showed behavior between those of EI and CI. Indeed, depending on the concentration of the analyte, the ion trap had a tendency to show reduced fragmentation. This affected quantitative analysis, on account of the different behavior of the analyte at different concentrations. Modern ion trap instruments give more reproducible results and quantitative analysis of pesticides is comparable to that with other mass spectrometers.

Time-of-flight instruments can be expected to catch on more in the future, but at the moment there are only limited applications for pesticide analysis and we shall not go into details.

3 IDENTIFICATION OF PESTICIDES AND RELATED PRODUCTS BY GAS CHROMATOGRAPHY/MASS SPECTROMETRY

3.1 Pesticides

GC/MS is unique in its ability to provide informative spectra of pesticides. The mass spectrum contains a kind of fingerprint specific to individual pesticides and can be used for automatic comparison with mass spectral libraries containing up to 200 000 spectra. Specific pesticide libraries are available. The computer compares the acquired mass spectrum with those in the library in seconds. The software suggests one or more spectra similar to the unknown spectrum. Generally, ions at higher masses are considered more significant, because they are more characteristic; at low masses many compounds have fragments. The software considers this feature of mass spectra.

The comparison is based on the presence of reference ions in the mass spectra of the unknown compound, but not vice versa. This makes the procedure more flexible towards impurities. In theory we might, for instance, have one extra ion in the mass spectrum of the unknown compound, suggesting impurities. If, however, we have an extra ion in the mass spectrum of the standard which is missing in the unknown mass spectrum, this should be considered with suspicion as it is indicative of a wrong identification.

The mass spectral library is a collection of spectra obtained in different laboratories and reported in the literature. The library may have more than one spectrum for the same compound. The library reports spectra obtained in the EI mode, as we said before, but not necessarily in GC/MS. EI can be obtained by introducing the analyte directly into the source [the direct inlet system (DIS)]. This gives a mass spectrum of a compound which is not suitable for GC analysis. For instance, if we analyze a sediment it may be common to find molecular sulfur. This gives a typical mass in which several ions are separated by 32 u. The library may suggest the spectrum of maneb. This is because maneb, once introduced into the ion source with the DIS, gives a decomposition mass spectrum dominated by sulfur ions, similar to that of molecular sulfur (see Figure 5a and b). The analyst must be aware of possible mistakes in identifying compounds on the basis only of the mass spectrum. The real identification requires confirmation with an authentic standard. This principle applies to all identification techniques, not only MS.

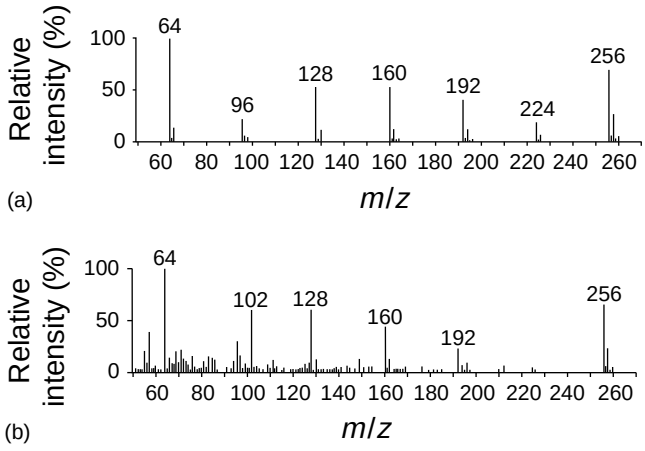

Figure 5 (a) Mass spectrum of sulfur, as obtained during an analysis of a sediment sample. (b) Mass spectrum of maneb, suggested by the National Bureau of Standards library for identification of the spectrum in (a).

3.2 Transformation Products of Pesticides

It is not necessary to interpret the mass spectra when identifying pesticides using the library, which compares the occurrence and the abundance of the main ions (see below). However, interpretation of the mass spectrum is essential for identifying unknown pesticides and related compounds, such as degradation products, metabolites, transformation compounds and chemical impurities related to the synthesis of the pesticide. Mass spectra have been reported for many of these compounds and included in the libraries, but often they are not considered in as much detail as their precursors.

The literature provides many examples where GC/MS proved essential for the identification of pesticides and related compounds. Several studies refer to the characterization of a TP, this expression including metabolites, degradation products and adducts. In this case the analyst can take advantage of the specific knowledge available, stemming from the problem considered. For instance, in the case of the identification of a TP we know the parent compound and it is therefore important to study its mass spectrum. A TP is a related compound, and its structure can be expected to be similar or at least related to that of the parent compound, even though there may be hundreds of potential metabolites. A fragment ion in common in the mass spectrum of the TP and of the parent compound is a strong indication that that moiety has been maintained in the TP, so transformation affected another part of the molecule. Similarly, the loss of the same residue indicates that the group which is lost in

the two molecules, the parent compound and the TP, is unchanged.

If the molecule contains two or more atoms of chlorine or bromine, they can be easily detected in a TP. Of course, other information may be useful too, such as GC retention time of a TP compared to the parent compound. As an example, see Mangiapan et al.[24]

3.3 Impurities in Pesticides

Pesticides are produced, for their final use, as a technical grade product. This means they are often contaminated by impurities. For instance, when atrazine was found in many parts of northern Italy, it was often together with simazine and propazine. MS is a major tool for confirming the real identity of the impurities.

Even when analytical standards are used for laboratory experiments, pesticides come with impurities. This was the case for alachlor, which complicated the identification of TPs, since some of the compounds possible after degradation were already present in the original compound.[24] Glyphosate was also contaminated by some impurities, determined by GC/MS.[25]

Particular impurities are nitrosamines, which occur in pesticides with amines, and are subject to special control on account of their possible carcinogenic effects. The common method is to use GC with a thermal energy analyzer (TEA), which is selective for NO and NO_2 residues; as an alternative, we used GC/MS looking at the ion at m/z 30, corresponding to NO, thus improving the selectivity of the method.[26]

A special situation may arise when derivatives are used. The derivatization compounds have been found to be contaminated with impurities, which may complicate identification of the derivatives.[25]

4 QUANTIFICATION OF PESTICIDES BY GAS CHROMATOGRAPHY/MASS SPECTROMETRY

4.1 General Characteristics of the Method and Comparison with Other Methods

For quantification, GC/MS employs the ions specific to the pesticide of interest, as they appear in the mass spectrum. This is SIM or selected ion recording (SIR). The use of these ions gives MS a unique ability. The analyst should select the ions on the basis of their abundance, but this is not the only criterion. The ions at higher mass are more specific, while those at lower mass are shared with many compounds, and this increases the likelihood of interfering compounds. Furthermore some ionic traces are ''dirtier'', because there are often contributions from common compounds or from column bleeding. For instance, the ion at m/z 149 commonly contains peaks due to phthalates, which are ubiquitous contaminants, while the ions at m/z 207 and 281 and several others show high noise due to bleeding from silicone columns. On the other hand, GC/MS is insensitive to the other contaminants, which may complicate analysis for some specific GC detectors, such as ECD and NPD. For these selective detectors, the problems are more serious than contamination in GC/MS.

For quantitative analysis the selectivity is due to the retention time (as with the other selective detectors) and to the ion chromatograms, which is a more specific feature than the response to a flame ionization detection (FID), ECD or NPD. However, in many cases the selectivity is further improved, because a second ion is added for the quantitative analysis. When two ions are used, the likelihood of a compound interfering with both is very low. Furthermore, we have to consider the relative intensity of the two ions: if in the reference mass spectrum the two ions are in a given ratio, this ratio has to be confirmed in the sample, with 10–15% imprecision, for instance. This ratio is a fourth element of specificity, with the retention time and the specific responses to the two ion traces. Of course, we can add yet another ion trace to improve the selectivity.

What happens if we find that the ratio of the two ions is not correct? It means that there is an interfering peak, and we can easily establish on which of the two ion traces it appears. We can then consider the cleaner trace, and calculate the amount of the analyte; if this is below a given fixed amount, which should be checked, we can conclude that the concentration of the pesticide is in any case below the fixed limit. If our aim is to determine its concentration or if the limit is exceeded considering the less contaminated trace, we can probably choose a third trace, to be used for the analysis, which may not be contaminated by interfering peaks. In this sense, GC/MS is a flexible technique, and the method can be modified according to the sample to be analyzed.

MS/MS generally gives higher selectivity than GC/MS since it refers to a very specific ionic transition. If we consider two ion traces, as is common in the case of quantitative analysis with GC/MS, one of the two traces may have impurities, but with MS/MS to see the signal both the ions must be involved, the parent and the daughter ion, and they must be linked by a particular transition, which means less likelihood of interfering peaks and a cleaner background, and thus a higher signal-to-noise (S/N) ratio.

In several cases the complete mass spectrum of the pesticide has been recorded with the ion trap instrument and used for quantitative analysis; this procedure couples quantitative and qualitative analysis, which in all the other cases are split in GC/MS, and is thus particularly

Table 2 LOD reported for GC/MS methods in pesticide analysis

Compound	Matrix	Method	Method LOD	Ref.
Organochlorines, triazines, thiocarbamates	Water	SPME/GC/MS (scan)	$\mu g\,L^{-1}$	27
Organochlorines, triazines, thiocarbamates	Water	SPME/GC/MS (SIM)	$ng\,L^{-1}$	27
Carbamates, chloroacetanilides, acetamides, azole fungicides	Foods and environmental waters	GC/MS (SIM)	$ng\,mL^{-1}$	28
Organochlorines	Spruce needles	GC/MS (SIM)	pg	3
Organophosphorus	Fruits	GC/MS (scan)	$10\,\mu g\,kg^{-1}$	29
Organophosphorus	Waters	GC/MS (SIM)	100 pg	5
Herbicides	Soils	GC/MS (scan)	$0.01\,mg\,kg^{-1}$	30
Herbicides	Soils	GC/MS (SIM)	$0.002\,mg\,kg^{-1}$	30
Herbicides	Waters	GC/MS (scan)	$0.1\,\mu g\,L^{-1}$	30
Herbicides	Soils	GC/MS (SIM)	$0.01\,mg\,kg^{-1}$	31
Herbicides	Drinking waters	GC/MS (SIM)	$0.1–10\,ng\,L^{-1}$	10
Phenoxy acid herbicides	Aqueous samples	SPME/GC/MS	$1\,ng\,L^{-1}$	11
Herbicides	Environmental waters	GC/MS (scan)	$6\,ng\,mL^{-1}$	32
Alachlor and its derivatives	Water samples	GC/MS (scan)	$5–50\,mg\,L^{-1}$	33
Alachlor and its derivatives	Water samples	GC/MS (SIM)	$0.05–1\,mg\,L^{-1}$	33
Alachlor, metolachlor, atrazine and simazine	Soils	GC/MS (SIM)	$0.5\,\mu g\,kg^{-1}$	34
Alachlor, metolachlor, atrazine and simazine	Water samples	GC/MS (SIM)	$0.05\,\mu g\,L^{-1}$	34
Vinclozolin	Spiked honey	GC/MS (scan)	$2\,mg\,L^{-1}$	35
Vinclozolin	Spiked honey	GC/MS (SIM)	$0.1\,mg\,L^{-1}$	35
Vinclozolin	Spiked larvae	GC/MS (scan)	$3\,mg\,L^{-1}$	35
Vinclozolin	Spiked larvae	GC/MS (SIM)	$0.1\,mg\,L^{-1}$	35
Cyhexatin	Spiked diet samples	GC/MS (SIM)	$20\,ng\,g^{-1}$	16
Synthetic pyrethroids	Soils	GC/NICI/MS	$0.1–40\,\mu g\,kg^{-1}$	36
Triazines and acetanilide herbicides	Waters	USEPA Method 525	$0.08–0.14\,mg\,L^{-1}$	37
Atrazine	Underground waters	GC/MS (SIM)	$1\,ng\,L^{-1}$	38
Atrazine	Ground and tap waters	GC/MS (SIM)	$200–500\,fg\,L^{-1}$	39
Organochlorine, herbicides, organophosphorus	Tap and river waters	EI positive ions	$2–20\,ng\,L^{-1}$	40
Organochlorine, herbicides, organophosphorus, thiocarbamates, carbamates, fungicides	River waters	GC/MS (SIM)	$0.5–50\,ng\,L^{-1}$	41
Organochlorine, herbicides	Drinking waters	EI positive ions	$2–9\,ng\,L^{-1}$	7
Amines, anilides, organophosphorus, triazines	Waters and sewage plant effluents	SPME/GC/MS	$5–90\,ng\,L^{-1}$	42
Organochlorine, triazines, thiocarbamates	Tap water	SPME/GC/MS (scan)	$10–200\,ng\,L^{-1}$	27
Organochlorine, triazines, thiocarbamates	Tap water	SPME/GC/MS (SIM)	$1–20\,ng\,L^{-1}$	27
Triazines, nitroanilines, uracils, thiocarbamates, chloroacetanilides	Water	SPME/GC/MS	$0.02–15\,ng\,L^{-1}$	43
Organophosphorus, triazines, chloroacetanilides, uracils	Water	SPME/GC/MS	$0.01–8.13\,\mu g\,L^{-1}$	44
Organophosphorus, organochlorine, organonitrogen	Water	SPME/GC/MS	$0.1–60\,ng\,L^{-1}$	56
Organotin compounds	Soil and sediments	GC/MS (SIM)	$0.4–2\,ng\,g^{-1}$	45
Toxaphene	Soil and sediments	GC/NICI/MS	$1\,\mu g\,kg^{-1}$	46

interesting. Further studies are still needed to explore the potential of this approach.

4.2 Detection Limits and Ranges

The LOD differs for different compounds. This is because each analyte gives fragmentations in different ways and yields. This may complicate quantitative analysis, because it requires standard and calibration curves obtained with the authentic compound. This should in any case be the procedure followed with all detectors. In the case of unknown compounds, semiquantitative analysis is less accurate in GC/MS than with other detectors. In quantitative analysis of pesticides, the LOD with GC/MS is adequate in most cases, once the best procedure is adopted. For instance, for pesticides in water the LOD was good enough for 50 compounds even applying the strict European Union (EU) limit of 0.1 ppb. ECD may offer better sensitivity in some cases, but generally this is not required, and very good sensitivities are achieved with GC/MS using NICI for the compounds which give good sensitivity with ECD.

Table 2 shows some indicative (not exhaustive) examples of LOD. With GC/MS some of the LODs are achieved without any purification, which is commonly required for other detectors. In GC/MS the ranges are often excellent, allowing a wide selection of theoretical analysis. There may be some limitations, which apply not only to GC/MS, however, but also to GC in general or to more specific cases. For instance, memory effects are possible with solid-phase microextraction (SPME), and this must always be checked, not only with SPME.

Ion trap instruments, especially the old ones, may show saturation of the source, which induces a sort of chemical autoionization. In the worst cases this produces a different mass spectrum for the analyte, depending on the concentration, with a consequent loss of linearity in the curve at higher concentrations. Modern instruments take this into consideration.

4.3 Isotope Dilution Mass Spectrometry

GC/MS offers a unique opportunity to improve the accuracy of measurement. The analyst may add a standard with isotopic labeling to the sample before any processing step. This compound is almost identical from the physicochemical point of view with the analyte of interest. The only difference is that it contains some atoms labeled with the corresponding non-natural isotope. Carbon-13 or deuterium is commonly used. MS distinguishes the labeled compound from the natural compound on the basis of the mass of the ion. We can easily keep account of the actual recovery of the analyte of interest or of loss due to evaporation or degradation. This procedure is particularly useful in the case of labile pesticides.

The number of labeled atoms to be introduced into the molecule varies. The rule is that one should check the mass spectrum of the analyte, looking carefully for the abundance of the satellite ions that commonly occur at higher masses. This is because some carbon-13 and deuterium are also present in nature. In the case of chlorinated or brominated compounds the situation is worse, because there is a high abundance of chlorine-37 or bromine-81. As a consequence, a shift of at least three or four mass units is required.

Several labeled pesticides are commercially available. Their price may be high and this limits their use. An alternative is to prepare the necessary standard. For deuterium labeling this may be simple because deuterium exchanges easily in several cases. However, care is needed when using these deuterated standards, since the labeling may be lost under particular conditions, such as altered pH.

5 APPLICATION TO DIFFERENT MATRICES

The specificity of pesticide action has led to the development of specific analytical methods for the various matrices. These methods are important from a toxicological point of view, to monitor food for sale and protect human health from dangerous residues of pesticides in the diet. National legislations establish different admissible residue levels for each pesticide in each commodity, and the analyst must take account of these specifications.

The GC/MS approach with its high specificity is particularly useful for analyzing pesticides in several matrices. This technique is now widely used not only for confirmation of residues, but also as a primary quantitation method. The wide applicability of GC/MS instruments for solving problems in the analyses of complex matrices is illustrated in this section.

5.1 Environmental Matrices

5.1.1 Water Samples

Most GC/MS applications for pesticide determination relate to the analysis of environmental and drinking waters. There are different reasons. First, the regulations on water quality are severe in several countries as regards maximum allowable concentrations of pesticides: the EU sets a limit of $0.1\,\mu g\,L^{-1}$ for individual pesticides and related products and $0.5\,\mu g\,L^{-1}$ for total pesticides.

Standard analytical protocols for water monitoring have been widely developed in the USA.[2] The United States Environmental Protection Agency (USEPA) is responsible, and this agency has introduced GC/MS-based protocols during the past few years. The standard

Method 525 is an example:[37] the sample is prepared with a solid phase extraction (SPE) cartridge or an Empore disk containing octadecylsilica as solid-phase; after elution with dichloromethane, the extract can be analyzed by GC/MS with LOD in the range 0.08–0.14 $\mu g\,L^{-1}$ in water for some triazines and acetanilide herbicides.

The SPE approach for water sample preparation for these analyses is particularly favorable: the liquid sample is passed through a column containing a solid phase, analytes are adsorbed and concentrated on this phase and subsequently eluted from the column with an appropriate organic solvent. Typical sample volumes in SPE are from a few milliliters to 10 L, depending on the column size and the sensitivity required. SPE is rapid and reduces the use of toxic and expensive solvent; furthermore, the formation of possible emulsions in the liquid–liquid extraction (LLE) procedure is avoided.

Much research work has been done on SPE, as described later, and the development of specific SPE procedures by the USEPA is a clear indication of the success of this approach. Most of the other USEPA methods for pesticide residues in water are based on LLE procedures. This sample preparation is still frequently used in laboratories, also in GC/MS based methods, in view of its good reproducibility and simplicity. For example, LLE is employed in important official methods from the UK Standing Committee of Analysis (SCA):[4] 1 L of water is extracted with 100 mL of hexane (in the case of pyrethroid insecticides) or diethyl ether after acidification of the water (for chlorinated phenoxy acids). These procedures comprise some successive cleanup steps, such as purification on a Florisil column (for pyrethroids) or hydrolysis of the extract (for phenoxy acids). GC/MS analysis is used in these UK methods only for confirmation of analyte identity using the NICI mode, after the usual determination by gas chromatography/electron capture detection (GC/ECD) and derivatization, in the case of chlorinated phenoxy acids. These procedures offer better sensitivity than the USEPA methods because of the use of a highly specific technique such as gas chromatography/negative ion chemical ionization mass spectrometry (GC/NICIMS).

Procedures based on LLE for water samples have been developed to analyze organochlorine, triazine and organotin pesticides by GC/MS.[15,37,47,48] LLE methods are now being progressively replaced by SPE, as this modern sample preparation technique is gaining in popularity. The use of specific solid sorbents for extraction has been extensively reviewed[49,50] and different SPE applications together with GC/MS have been developed (Table 3).

Most SPE applications are off-line procedures: the solid-phase column is eluted with a solvent, and a fraction

Table 3 SPE methods for pesticide analysis by GC/MS

Analyte	Water matrix	SPE phase	GC/MS technique	Method LOD	Ref.
Atrazine	Underground water	Octadecylsilica	EI positive ions	1 $ng\,L^{-1}$	38
Atrazine	Ground and tap water	Octadecylsilica	EI positive ions	200–500 $fg\,L^{-1}$	39
Organochlorine and OPPs, herbicides, thiocarbamates	Groundwater	Octadecylsilica	EI positive ions		51
PCP	Groundwater	Octylsilica Octadecylsilica Cyclohexylsilica Phenylsilica	EI positive ions		52
Organochlorine and OPPs, herbicides, thiocarbamates carbamates, fungicides	River water	Octadecylsilica Phenylsilica	EI positive ions	0.5–50 $ng\,L^{-1}$	41
Organochlorine pesticides, herbicides,	Drinking water	Octadecylsilica	EI positive ions	2–9 $ng\,L^{-1}$	7
Acetochlor	Environmental water	Octadecylsilica	EI positive ions	1.5 $ng\,L^{-1}$	53
Organochlorine and OPPs, herbicides	Tap and river water	Styrene–divinyl -benzene copolymer	EI positive ions	2–20 $ng\,L^{-1}$	40

OPPs, organophosphorus pesticides; PCP, pentachlorophenol.

of this extract is injected with a microsyringe, manually or automatically. Solid phase extraction/gas chromatography/mass spectrometry (SPE/GC/MS) methods have also been developed with the extraction column coupled on-line to the GC system; extraction and analysis of the water samples for pesticide analysis are done on the same apparatus.[40] Generally, the elution solvent is injected on-column, interfacing a retention gap between the injector port and the analytical GC column. This means that most of the organic solvent injected (typically 100 μL) is purged from the system through a solvent vapor exit, inserted between this gap and the column.[40] The instruments can automatically dry the SPE column before the solvent elution step, using a purge gas line.

On-line SPE/GC/MS methods are sometimes called liquid chromatography/gas chromatography/mass spectrometry (LC/GC/MS) procedures.[54] These methods have been used to detect triazines at low parts per trillion (nanograms per liter) levels; a short polymer-packed liquid chromatography (LC) precolumn is used as the SPE column. Typical extracted water volumes in LC/GC/MS are a few milliliters.

Another attractive recent sample preparation technique for water samples is SPME: a 1-cm silica capillary coated externally with a thin layer of a sorbent phase [a viscous polydimethylsiloxane or a crystalline polymer such as polyacrylate (PA)], is immersed in the liquid sample under stirring; analytes partition in this capillary and in the sample to reach an equilibrium. After the desired extraction time, the fiber is withdrawn from the sample and the extract is directly injected into the gas chromatograph by thermal desorption, exposing the fiber inside a GC injector port.

The large number of methods with SPME coupled with GC that have appeared in recent years shows that this technique can be a valid alternative to SPE and LLE. Ground, tap and river waters, effluents from sewage plants and leachates from landfills have been extracted with SPME, and GC/MS was performed without any purification step.[42,55]

Most SPME methods for pesticide analysis used a polar polymeric fiber phase (PA), on account of the mid-polarity of most analytes. LOD are in the low nanograms per liter region in the EI and SIM modes, and linearity of quantitation ranges over three orders of magnitude (up to 100 $\mu g\,L^{-1}$).[27,42–44,56]

Supercritical fluid extraction (SFE) is another sample preparation technique applied in pesticide analysis. The extraction solvent is a fluid maintained at supercritical conditions (pressure and temperature above the critical point), CO_2 being the most widely used solvent because of its low critical values and low toxicity. Furthermore, it can be easily separated from the extracted material, since it is gaseous in normal conditions. Some SFE methods for water have been developed for pesticide analysis with GC/MS.[5,57,58] The liquid sample (1 mL) is immobilized in the extraction cartridge using a hydrophilic solid support (sodium sulfate, diatomaceous earth or ChemElut); supercritical CO_2 is used and the extracted compounds are recovered by bubbling the depressurized CO_2 in a vial containing few milliliters of an organic solvent; this organic phase is analyzed by GC/MS without any purification. Quantitative recoveries have been obtained with this approach for fenamiphos and its metabolites, fenamiphos sulfoxide and fenamiphos sulfone.[5]

5.1.2 Soils and Sediments

The soil matrix contains more interfering and co-extractive compounds than water. Consequently, the high specificity of GC/MS instruments can help the analyst overcome problems associated with the application of element-specific GC detectors for the analysis of dirty soil extracts. GC/MS in soil analysis also serves as a confirmatory technique for positive results for pesticides, as discussed later.

Bruns et al.[31] pointed out that ECD was insufficient for confirmation of herbicides in soil because of interference in the chromatogram. A multiresidue GC/MS method for herbicide analysis was specifically developed to solve these problems: analytes were identified with a high degree of confidence to the 0.003 $mg\,kg^{-1}$ level, but with relatively low precision (15–23% standard deviation in the quantitation),[31] using SIM.

Corn herbicides in soil were determined using GC/MS,[30] but only as confirmation besides gas chromatography/nitrogen–phosphorus detection (GC/NPD). Samples were extracted with ethyl acetate and analyzed without any purification step; GC/MS was sufficiently specific and the LOD in the SIM mode was more than 10 times lower than that with full-scan operation.

GC/MS was also used as confirmation for the analysis of some thermally labile phenylurea herbicides;[17] an ion trap GC/MS system was applied to elucidate the structures of the thermal decomposition products formed in the GC injector and selected for quantitation in the direct GC analysis of these herbicides. Each phenylurea gave a characteristic phenylisocyanate from thermal decomposition, and ion fragments from each of these products were selected for confirmation by GC/MS down to the 0.01 $mg\,kg^{-1}$ level, in the SIM mode. The SIM mode was also used to confirm some sulfonylurea herbicides in soil using membrane disk SPE.[23] In membrane SPE the phase is settled on the surface of a Teflon™ filtering disk; compared with the SPE cartridges, the extraction process here is faster since higher water fluxes can be achieved in the disk than with an extraction column.

Klaffenbach and Holland[23] extracted soil samples with 0.1 M sodium hydrogencarbonate buffer; analytes were separated from this solution with membrane disk octadecylsilica SPE. After solvent elution of the disk and derivatization with diazomethane, GC/MS operating at resolution 1000 was used to analyze the *N*,*N*′-dimethyl derivatives of these herbicides in the eluate.

GC/NICIMS has been considered recently for the determination of toxaphene in soils.[46,59] This organochlorine pesticide is a mixture of about 700 polychlorinated bornane congeners; it is difficult to analyze because of this complex composition, and highly selective methods are needed. In view of the presence of chlorine atoms in the analytes the NICI mode of MS was used, and selectivity was increased using a magnetic sector instrument operating at resolution 10 000; these operating conditions were effective in the analysis of soil samples containing interfering compounds such as polychlorinated biphenyl (PCB) compounds.[59]

GC/NICIMS in the SIM mode was also used to analyze some pyrethroid insecticides in soil and sediments;[36] samples were extracted with a mixture of hexane and dichloromethane in an ultrasonic bath and extracts were cleaned up with Florisil chromatography. The NICI mode of MS was chosen because all the selected insecticides contain halogen atoms (chlorine or bromine). The authors reported a limit of quantification for this method in the range 0.012–4.4 $\mu g\,kg^{-1}$.

The reported NICI mass spectrum of deltamethrin presents some peculiarity: the most intense peaks are due to the bromine anions at *m*/*z* 79 and 81. A similar NICI mass spectrum was observed for another halogenated insecticide, DDT, but of course in this case chlorine was detected.[60]

Organotin compound are widely used as pesticides and antifouling agents. A GC/MS method has been recently developed to monitor their accumulation in sediments with accelerated solvent extraction (ASE).[45] ASE is a new sample preparation technique: solid samples are extracted with a pressurized fluid held at 50–100 °C. Volatile solvents can be used at high temperatures because of the pressurization, so less solvent is required and extraction times are shorter than with traditional methods. Organotin compounds have been extracted with a methanolic mixture of sodium acetate and acetic acid. GC/MS detection allowed the use of perdeuterated analogs to achieve high precision (relative standard deviations <5%) and low LOD (0.4–2 $ng\,g^{-1}$, starting with 2.5 g of sediment).[45]

Durand et al. compared the performances of GC/MS at two resolving powers (1000 and 10 000) and that of gas chromatography/tandem mass spectrometry (GC/MS/MS) in the determination of chlorotriazine pesticides in soil;[61] the MS/MS technique (on a hybrid magnetic quadrupole instrument) achieved higher specificity in the analysis of real soil samples than the other two single MS procedures.

Some preliminary experiments indicate that SPME methods can be applied for the analysis of pesticide residues in soils using GC/MS.[43]

5.1.3 Air

GC/MS has been applied by different research groups in the field of air monitoring. The common procedures for sample collection are for adsorption: air is forced through a tube containing a specific solid medium, using a sampling pump. In some cases, filter sheets made of glass[62] or quartz[63] fibers are placed before the adsorption column to collect analytes present in the air particulate. Different solid media are used: Amberlite™ XAD-2,[63,64] silica gel,[62] Tenax™ TA[65,66] and a polyurethane foam (PUF) plug.[67] High volumes of air (100 L to 1000 m^3) are sampled with these materials, in view of the low concentrations of the pesticides (1–40 $pg\,m^{-3}$ for DDT in continental air).[62]

The trapped pesticides are desorbed by solvent extraction[62–65,67] or thermal desorption;[66] in this latter case, a thermostable material (Tenax™ TA) is used as adsorption medium so the sampling tube can be heated directly connected to the GC/MS system.

5.1.4 Biota

GC/MS has been successfully applied in the analysis of very complex matrices such as animal tissues. The NICI mode has been employed to analyze different organochlorine pesticides in mouse and rat brain and mouse plasma;[68,69] these methods were highly selective, permitting direct analysis of the sample extracts without any purification. Identification of some lindane metabolites in the rat brain was possible using NICI in the full-scan MS mode.[68] The fragmentation was limited in the NICI spectra of endosulfan reported by these authors and the spectrum was dominated by the molecular ion signal; in another study,[69] however, the target endosulfan ions chosen for SIM in NICI were the chlorine ion signals (*m*/*z* 35 and 37). This is an example of how NICI conditions influence the spectra; ionization depends on several experimental factors (such as ion source temperature and geometry, reagent gas and pressure).

Different pesticides were detected in pine and spruce needles with GC/MS techniques: PCP was analyzed at low parts per billion levels after derivatization with diazomethane to give the corresponding methyl ether.[8] Other organochlorines such as hexachlorocyclohexane (HCH), hexachlorobenzene (HCB), DDT and metabolites has been found in conifer needles using GC/MS in the

EI[70] and NICI[3] modes. The EI procedure employed a Florisil chromatography purification step and isotopically labeled internal standards in the quantitation. As an alternative, the NICI method avoided the chromatographic cleanup step using a concentrated sulfuric acid treatment on the vegetable extract, and employed a brominated analog of DDT as an internal standard.

An example of the use of the technique to monitor organochlorine bioaccumulation in aquatic organisms is provided by Galceran and Santos:[71] homogenized and dried samples of different mussel species were Soxhlet extracted with hexane–dichloromethane (1 : 1), and the extract was cleaned as described elsewhere.[70] GC/MS in the SIM mode was used for confirmation of results obtained first with ECD.

Development of analytical procedures for toxaphene in biological samples is complicated by its complex nature (as mentioned in the soil section) and by the numerous interferences from other chlorinated compounds. A procedure was developed for toxaphene analysis in fish samples:[72] after extraction with dichloromethane, three successive chromatographic cleanup steps are used to separate first the analytes from the lipids [with gel permeation chromatography (GPC)] and then from the other chlorinated contaminants (with Florisil and silica gel chromatography). GC/MS was used in the SIM mode with EI positive ion detection, with an LOD of $1\,\mu g\,kg^{-1}$. Compared with this method, the NICI procedure developed for these same matrices was less time-consuming, since only GPC and silica gel purifications were needed,[73] demonstrating the higher selectivity of the NICI technique. The NICI spectrum of toxaphene is characterized by limited ion fragmentation with respect to PICI and EI (see Figure 3); the presence of these high-intensity NICI ion signals explains the higher sensitivity of this technique compared with PICI and EI methods. In particular, the same authors[73] found that the NICI mode was approximately 100 times more sensitive than EI or PICI for the analysis of toxaphene.The minimum detectable amount of toxaphene with this method was 75 pg.

5.2 Food Samples

GC applications to pesticide residue analysis in foods are as numerous as those to water, and the number of GC/MS methods has grown continuously since the 1980s. GC/MS has been considered as a secondary confirmation technique in multiresidue methods for samples exceeding the maximum residue level (MRL),[74] but its use as a primary screening tool is becoming frequent.[75]

Advantages over the classical element-sensitive detectors convinced the Food and Drug Administration (FDA) to use GC with ion trap MS in their multiresidue analytical protocol based on the Luke extraction method.[76] This instrument was compared with the other detectors for the determination of over 100 residues in real contaminated food samples. The FDA concluded that GC/MS was able to provide a determination and confirmation simultaneously in samples using the full-scan mode down to the sub-parts per million level, but probably an additional cleanup step was necessary to obtain acceptable quantification for all the target compounds and better sensitivity.

Sample preparation in most of the GC/MS procedures is based on the traditional solvent extraction (see Table 4 for a tentative list and section 6). In addition to the official extraction methods (Luke and Mills methods),[89,90] specific procedures have been considered.

Solventless procedures such as SFE and SPME are other extraction methods considered for foods analysis using GC/MS. SFE methods have gained ground in the last decade.[75] Sample preparation can be highly selective for pesticide extraction, and the extracts can be injected directly into the instrument without any purification step. The possibility of dramatically reducing the use of toxic solvents is another advantage that encouraged food analysts to consider this technique. SFE and GC/MS have been used together in the development of a multiresidue method for 46 pesticides in fruits and vegetables:[91] CO_2 was used as a supercritical fluid, and extracts collected using an octadecylsilica solid trap directly connected to the extraction chamber. Analytes were recovered from this trap with acetonitrile elution. No quantitative recoveries were achieved for all the pesticides with this method (values ranged from 0% to 114%, depending on the analyte), meaning this supercritical fluid extraction/gas chromatography/mass spectrometry (SFE/GC/MS) approach needs some improvement before its application in regulatory laboratories.

Unlike the previous method, SFE using CO_2 modified with methanol was effective in the analysis of fluazifop-P-butyl in onions, with GC coupled with quadrupole MS;[92] it is common practice to add a polar solvent to the supercritical fluid in SFE methods in order to increase the solvent strength and the diffusion of the fluid into the sample matrix.

Applications of SPME in food analysis have been reported for wines;[43,93] with these methods, the SPME fiber is directly immersed in the liquid sample, without any pretreatment. Quantitations were based on a calibration curve obtained with pesticide-spiked wine samples[93] or with a standard addition method,[43] and LOD in wines for 14 most frequently used pesticides were in the $0.1–5.5\,\mu g\,L^{-1}$ range.[93] Orange juice was also analyzed with this SPME/GC/MS approach.[56]

Validation of immunoanalytical methods for pesticide analysis is an important task in the development of such

Table 4 GC/MS methods for pesticide analysis in foods

Analyte	Matrix	Extraction	Cleanup procedure	GC/MS detection	Refs.
Chlorothalonil	Cucumber, pepper, cherry tomato	Mills official procedure			77
Multiclasses	Apple, pear	Acetone–dichloromethane–hexane mixture	None	EI positive ions, SIM mode	78
Dacthal, diazinon, atrazine	Leafy vegetables, root, fruits, citrus, spices	Acetonitrile	(a) C_{18} LC (b) SPE (Florisil)	EI positive ions, scan mode	79
o-Phenylphenol, diphenylamine, propargite	Fruits, vegetables	Acetonitrile	None	EI positive ions, SIM mode	80
Paraquat, diquat	Potato, rapeseed	Diethyl ether extraction after reaction with sodium borohydride–nickel(II) chloride		EI positive ions, SIM mode	81
Multiclasses	Apple, cucumber, wheat, salad, cauliflower, tomato, endive, potato, pepper, chicory	Acetone	(a) water–hexane partition (b) GPC	EI positive ions, scan mode	82, 83
Profenofos	Tomato	Hexane–acetone mixture	Charcoal treatment	EI positive ions, scan mode	84
Procymidone	Wine	C_{18} SPE	None	EI positive ions, SIM mode	85
Multiclasses	Cocoa bean	(a) Hexane after conc. H_2SO_4 treatment (b) UNITREX distillation system		EI positive ions, SIM mode	86
Multiclasses	Fruits, vegetables	Dichloromethane	None	EI positive ions, ion trap MS/MS	87
Linuron	Potato	Modified Luke official procedure		CI positive ions, SIM mode	88

procedures. GC/MS is generally accepted as the leading technique for comparison in most types of enzyme-linked immunosorbent assay (ELISA).[94–96] Recently, two commercial ELISA kits for procymidone residue analysis in wines were compared with an SPME/GC/MS method.[97]

5.3 Other Matrices

Some specific matrices such as solid wastes and animal feed preparations have been analyzed for pesticides using GC/MS methodologies. Official GC/MS methods to evaluate the content of different organochlorine pesticides (DDT, toxaphene) in solid wastes have been developed by the USEPA, namely Methods 8250 and 8270.[98,99] A GC/MS procedure has been reported for the determination of the organotin acaricide cyhexatin (tricyclohexyltin hydroxide) in animal feed samples;[16] a derivatization step with a Grignard reagent (ethylmagnesium bromide) is used to give the alkyl derivative of this pesticide, allowing GC analysis. Tricyclopentyltin hydroxide is used as an internal standard.

6 APPLICATIONS FOR DIFFERENT PESTICIDE CLASSES

6.1 Introduction

We shall consider here the major pesticide classes individually, indicating the peculiarities of the major pesticides. This section reports applications and examples related to the previous ones. Since some examples have already been presented, mainly in section 5, we shall not repeat them here.

The most frequently used methods for analyzing pesticide contamination in the environment rely on GC separation and detection with various selective detectors. The choice of detection technique depends on the

chemical characteristics of the target analytes and on the required specificity and sensitivity.

6.2 Multiclass Methods

Maximizing the number and types of analytes to be determined with given resources is considered the most practical and economical approach in a monitoring program. One of the main aims of recent analytical procedures is to monitor the overall content of pesticide contamination, often involving different chemical classes and different applications in environmental matrices.

GC/MS can monitor a wide series of different pesticides, regardless of their chemical class, in a single analysis (see Figure 6 for an example). In order to achieve the low legal limits set for pesticide contamination, extraction and enrichment steps are necessary before GC/MS analysis.

The preconcentration step can be carried out by different extraction techniques, such as solvent extraction, including solid–liquid extraction (SLE) and LLE, used in several USEPA methods, SPE, SPME and the more recent SFE.

Solvent extraction has been the main extraction procedure for a long time. Depending on the solvent's characteristics (volatility, dissolving ability and miscibility with other solvents), different pesticides chemical classes can be extracted from food and environmental matrices. For multiresidue analysis of pesticides in fruits and vegetables, acetone partition/extraction and ethyl acetate extraction have mainly been used.[75,89] Organochlorine, organophosphate and organonitrogen pesticides are extracted well from food, feed and fat samples using an acetone–water mixture, or from fruit and vegetables using acetonitrile. SPE can be applied for multiresidue extraction procedures on a large number of pesticides belonging to different chemical classes.[41,100] Since various analytes with different chemical properties have to be extracted and analyzed in a multiresidue procedure, solid-phase columns or disks are chosen on the basis of their affinity toward the targets. Different extraction materials may be used simultaneously, in order to achieve exhaustive recoveries.

C_{18}, phenyl and NH_2 silica-bonded phase performances have been evaluated in the multiresidue/multiclass extraction of 50 different chemical pesticides from water samples.[41] The C_{18} phase was suitable for pesticides of medium polarity, whereas the most polar ones were extracted better with the NH_2 phase. To shorten

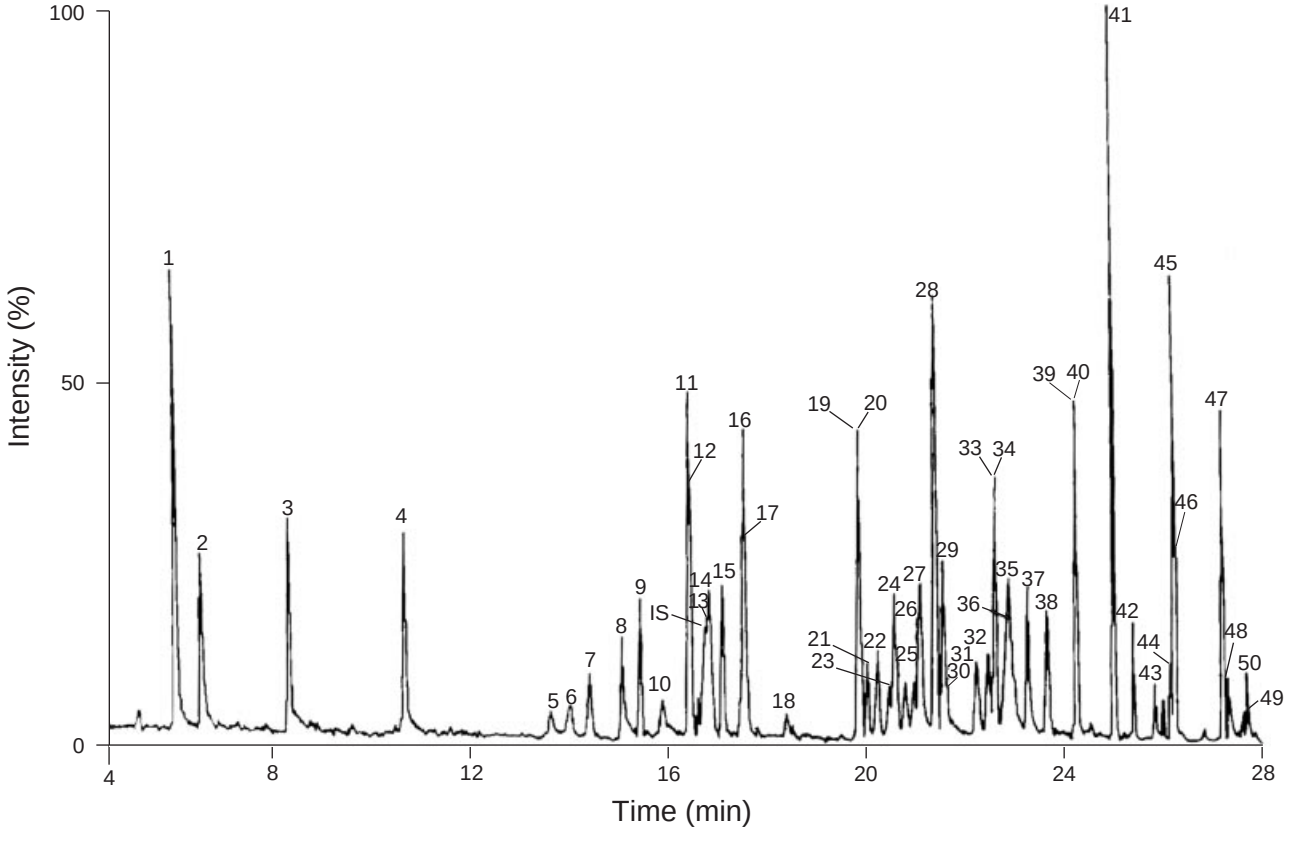

Figure 6 Total ion chromatogram of a mixture of 50 pesticides. (Reprinted from E. Benfenati, P. Tremolada, L. Chiappetta, R. Frassanito, G. Bassi, N. Di Toro, R. Fanelli, G. Stella, *Chemosphere*, **21**, 1417 (1990), with permission from Elsevier Science.)

the sample preparation time and to achieve consistent recoveries, these different phases can be used simultaneously on the same column, so analytes can be extracted in a single preparation step.

Other materials have been tested for multiresidue pesticide extraction from water, including XAD-2 and XAD-7 resins.[100] The XAD method can extract a wide variety of pesticides, including triazine, phenylurea, acetamide and aniline herbicides, carbamates and organophosphates and the simultaneous use of both resins achieves good recoveries.

SFE is another useful technique for the extraction of pesticide residues from food and environmental matrices. It has been successfully employed with the GC/MS procedure in the analyses of pesticides in vegetables.[91]

GC/MS procedures for multiresidue/multiclass screening of fresh fruit and vegetables have been developed for 143 selected compounds, several pesticides, some of their metabolites and other chemicals of concern in the safety of domestic or imported food. Including organophosphorus, organohalogen, organonitrogen, carbamate, triazine, thiourea and phenolic compounds and aromatic hydrocarbons and not considering chemicals not suitable for GC analysis, it has been demonstrated that the GC/MS procedure is powerful and one of the major analytical tools in multiresidue screening because it permits rapid routine and confirmatory monitoring of hundred of chemicals of food safety concern.[101]

SPME/GC/MS procedures have been used in multiclass analyses; organochlorine, atrazine and thiocarbamate pesticides in environmental aqueous samples can be identified and quantitated using a polar fiber (PA, 85 μm) and employing GC/MS in both the scan and SIM modes.[27] Good linearity and satisfactory LOD can be achieved, demonstrating the suitability of the newly developed extraction technique and the effectiveness of GC/MS for multiresidue pesticides analyses.

This technique can also be applied for pesticide extraction from food and environmental samples.[28] Employing the appropriate fiber for the target analytes and limiting the interferences that complex matrices can generate, multiclass pesticide analysis can be performed after optimization of the experimental parameters. Both MS and MS/MS can be used for pesticides. Using the SPME procedure in a preparation step, samples for high-sensitivity studies can be obtained in a short time, achieving good specificity and low LOD.

Further improvements can be achieved in the evaluation of GC/MS data using macro programs which automatically run all the necessary steps for manual data evaluation.[102] These programs take less time than manual data processing and the quality of evaluation can be maintained constant throughout the analysis. The use of a macro program in combination with a selected library containing a limited number of target compounds gives reliable results which the analyst can verify quickly.

6.3 Organochlorine Pesticides

Different environmental matrices have been extracted in order to quantitate organochlorine pesticides and several analytical methods have been developed. Here, we describe some of them.

The official USEPA method for the analysis of organochlorinated pesticides in water is Method 508, in which these compounds are monitored by GC/ECD, after LLE. ECD is mainly used in this kind of analyses, on account of its high sensitivity, but the need for confirmatory studies makes a GC/MS procedure more suitable for organochlorine pesticides.

Several GC/MS methods have been developed to analyze these chemicals not only in water samples but also in other environmental and food matrices.[103] A Soxhlet apparatus is usually employed for extraction. SFE with CO_2 has been compared with the classical method for sediment analyses;[104] furthermore, low LOD have been obtained with coupled SFE/GC/MS, making this a useful method for organochlorine pesticides in sediments.

Ultrasonic extraction has also been studied for analyzing DDT, 1,1-dichloro-2,2-bis(4-chlorophenyl) ethylene (DDE) and 1,1-dichloro-2,2-bis(4-chlorophenyl) ethane (DDD) in soil, with a view to saving time and minimizing the use of hazardous solvents in comparison with the classical Soxhlet extraction procedure. GC/MS indicated that ultrasonic extraction is suitable for the preparation of soil samples in organochlorine analysis. In addition, the lower temperature required may reduce the thermal degradation of DDT to DDE that can occur during Soxhlet extraction.[105]

In organochlorine pesticide analyses it is better to use an internal standard for the standardization of the extraction procedure and subsequent instrumental determination. ^{13}C-labeled HCB has been employed in quantification of unlabeled organochlorine pesticides in vegetable samples using MS in the NICI mode.[106] A brominated compound, 1,1-bis(4-chlorophenyl)-2-bromoethane has been used as an internal standard in organochlorine pesticide determination in vegetable samples, with ECD and NICI mode MS detection.[3]

6.4 Organophosphorus Pesticides

Organophosphorus pesticides (OPPs) can be analyzed by different methods although the commonest involves GC or LC. In many studies, GC with phosphorus flame photometric detection (FPD-P) has been the technique of choice because of its high selectivity, sensitivity and reproducibility. However, interference may occur when FPD-P is used in the phosphorus mode and sulfur is

present, or when thermally labile OPPs are present or when the liquid phase bleeds. To avoid mistakes, GC/MS can be used as a confirmatory technique. After some pretreatment steps, depending on the nature of the matrix to be analyzed (e.g. maceration and addition of drying agent for food samples and just solvent partition for water samples), LLE or SPE with various sorbents can be used to isolate them.[5,29,107] LiChrolut EN and C_{18} cartridges are suitable for the extraction of OPPs and the recently developed SFE can be considered a useful alternative in some cases.[5]

These simple extraction procedures are suitable for the determination of OPPs at the nanogram level and there are few interferences. To achieve picogram-level determination, high-performance GC has been employed, using two columns with different polarity coatings, with a column switching technique.[108] To avoid sample loss during the cleanup procedure and to correct the deviation of injection volume and the variability in detector response, internal calibration is often useful, particularly when labile or volatile OPPs have to be analyzed. Generally, different OPPs not involved in the analysis[29] or deuterated analogs of the targets[5] can be chosen for this purpose.

Different MS application modes have been tested. EI in the scan mode gives the most information about the compounds but it suffers from limited detection sensitivity; in the SIM mode the sensitivity is higher, but it is susceptible to interference. CI of OPPs generates base peaks with high mass in the positive mode (PICI), frequently the quasi-molecular ion, and in the negative mode (NICI) it gives mostly group-specific fragments of high intensity.[108] In general OPPs, determination was more sensitive with NICI than with PICI or EI. Reliability, however, was better in PICI thanks to the abundant quasi-molecular ion. Pulsed positive–negative ion chemical ionization (PPNICI) was indicated as a good compromise with regard to sensitivity and reliability.[108,109]

6.5 Herbicides

LLE is widely used for herbicide extraction from environmental solid samples, such as soils and particulates suspended in environmental waters. Depending on the chemical characteristics of the target analytes and on the type of matrix, different solvents have been used.[30] To enhance recovery, ultrasonic extraction can be performed and, if herbicides with different acidity are present in the same sample, a double-space extraction procedure can be performed, using extraction solvents with different acidity.[31,110]

SPE is also suitable for herbicide extraction from water samples. Different solid phases have been tested. For instance, styrene–divinylbenzene copolymer is suitable for a wide range of compounds[110] and RP-C_{18} cartridges are used for phenoxyalkanoic acid and other acidic herbicides.[31] SPME procedures too are suitable for herbicide extraction and, since they are simple and can be automated easily, are ideal for in-field water measurements.[11]

When target analytes are not suitable for GC/MS determination because of their chemical characteristics, a derivatization procedure has to be employed. As illustrated above, it is generally applied for non-volatile, unstable compounds or when there is the possibility of undesired reactions between the analytes and the stationary phase of the GC column. Many derivatization procedures have been developed to convert herbicides not suitable for GC/MS analyses into easier to handle chemicals. Acid herbicides can be converted to benzyl esters by reaction with pentafluorobenzyl bromide or benzyl bromide;[10,31] phosphorus-containing herbicides can be derivatized by reaction with *N*-methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide;[18] urea herbicides can be converted by pyrolytic methylation with trimethylanilinium hydroxide or trimethylsulfonium hydroxide[20] and sulfonylurea herbicides can be converted into thermostable derivatives by reaction with diazomethane in ethyl acetate.[22]

When selective detectors are chosen as a primary screening tool for herbicide determination, MS detection is a good confirmatory technique.[30,111] However, considering the low contamination levels in environmental matrices, GC/MS is one of the main techniques for identification and quantification.

Often, an ion trap detector is chosen for trace determination because of its sensitivity. It identifies the contaminants on the basis of their total spectrum[18,32] and fragmentation patterns of the compounds of interest can be studied at ultratrace levels. When co-extracted substances interfere, ion trap MS is used in the SIM mode, reaching high levels of confidence for identification in environmental matrices.[10,31,33]

GC/MS analyses of water samples can be speeded up by introducing the aqueous sample directly into the instruments, without any extraction or preparation steps. GC/MS systems equipped with a silanized-glass reverse-cup liner and the splitless injection mode can be employed for direct injection analyses of water samples.[33] Isotope dilution GC/MS is now widely used in herbicide analyses because it gives better accuracy and precision, and high-isotopic-purity, stable labeled standards are available.[34]

6.6 Pyrethroids

Pyrethrin and pyrethroid pesticides are non-polar and non-systemic in plants, and their extraction is simple compared with the organophosphate or carbamate pesticides.

The solvent and the method used depend on the nature of the sample. Basically, the samples are homogenized with a single nonpolar solvent, such as benzene or hexane, or with a binary solvent mixture, such as hexane–acetone, hexane–2-propanol or light petroleum–diethyl ether. The pyrethroids are co-extracted together with a wide variety of lipophilic compounds during this process. If too many compounds interfere with the final determination, the sample can be cleaned up through liquid–liquid partition or column chromatography. The cleanup procedure is also influenced by the final determination. If the detection is selective enough, the cleanup procedure need not be so rigorous, otherwise more efficient cleanup is needed.

Samples with low water content, such as tea, tobacco and straw, are usually homogenized with a binary solvent mixture such as acetone–hexane (1:1) or hexane–2-propanol (3:1) or a single polar solvent, such as methanol, acetone or acetonitrile. Vegetables and fruits are usually homogenized with a binary mixture such as acetone–hexane (1:1 or 1:4) or hexane–2-propanol (3:1) in the presence of granular anhydrous sodium sulfate. Soil samples, after mixing and removing stones and vegetation, are extracted with acetone–hexane, methanol, acetone and acetonitrile by shaking or end-over-end tumbling for many hours.[112]

The clean-up procedure consists of liquid–liquid partition and column chromatography. The solvent partition systems used for pyrethroid residue analyses include acetone–hexane, acetone–dichloromethane, acetone–light petroleum, acetonitrile–hexane, methanol–toluene and acetonitrile–light petroleum.[112]

SPE can be used in place of liquid–liquid partition methods for the extraction of pyrethroids from matrices. Octyl (C_8), ethyl (C_2) and cyclohexyl (CH) sorbents are superior to the others in strength and selective elution.

Adsorption chromatography is often used as an additional cleanup step to eliminate interference from co-extractives. Commonly used sorbents include Fluorisil, silica gel and alumina. These adsorbents are polar, thus retaining the lipid fraction on elution with organic solvents of low polarity. They are suitable for cleaning up apolar analytes such as pyrethrin, pyrethroids and organochlorine pesticides.[112]

GPC can be applied to cleanup procedures for environmental matrices in pyrethroid analyses.[112,113]

Generally, GC/ECD is the technique of choice for pyrethroid residues since most of these compounds contain halogen. For nonhalogen-containing pyrethroids, it is recommended to form halogenated derivatives before GC/ECD determination.[112]

However, MS, operating in the scan mode, serves as a structure-selective detection technique and permits the characterization and identification of TPs formed under special environmental conditions.[114] High detection sensitivities have been achieved in the SIM mode at the parts per billion level.[36] Ion trap GC/MS/MS is another means of determining the concentration and structures of pyrethroid insecticides. The characteristic ester cleavage of these compounds combined with selected ejection chemical ionization (SECI) and collisionally activated dissociation (CAD) procedure speeds up the analysis of protonated pyrethroids.[115] In the SECI mode, specific ions from the CI gas are isolated before they are allowed to react with the neutral analyte molecules. This avoids EI by-products created during the ionization of the CI gas and the eluting analytes from being present in the final spectrum. The CAD procedure provides the ability to isolate specific ions after an EI event or CI reaction time, to excite them resonantly or nonresonantly and to detect the resulting product ions, providing more specific fragmentation patterns. Furthermore, the low cost of ion trap mass spectrometers makes them affordable for the structural characterization of complex mixtures.

6.7 Carbamates and Related Pesticides

Generally, LC techniques are preferred for the analyses of carbamate insecticides, but these methods require additional postcolumn derivatization steps to achieve the necessary detection limits and, except in LC/MS experiments, confirmatory MS analyses are necessary. Most of the carbamates now in use are not stable under standard GC conditions using a capillary column and hot splitless injection. They decompose in the hot injector forming phenols and isocyanates. The carbamates must therefore be derivatized. Various procedures have been optimized for single residues and for the whole group of *N*-methylcarbamates, to convert them to derivatives more suitable for GC/MS analysis. A rapid derivatization procedure involves conversion to phenol acetates by reaction with acetic anhydride.[21] Since phenol acetates containing no heteroatoms are not suitable for ECD or NPD, these derivatives need MS determination. Working in the SIM mode, the GC/MS analysis of phenol acetates shows that high sensitivity can be achieved, making the procedure suitable for trace analyses of carbamates. The GC/MS technique has been successfully employed for the determination of metal–dialkyldithiocarbamate complexes in water. The EI technique has been tested for qualitative and quantitative analyses. On the basis of the mass spectra, the fragmentation patterns of the complexes have been studied.[116]

6.8 Fungicides

Fungicide residues have been studied in many matrices such as soil, plants, wine and water, by extraction with an organic solvent of low to medium polarity or by SPE. Different sorbents have been tested for fungicide

residue extraction. Good recoveries have been obtained using octadecylsilanes[35] and the newer porous carbon sorbents CARB GR, which offer thermal stability, chemical resistance and stability over a wide range of pH.[117] GC/MS experiments have been carried out for confirmation of previous GC with selective detection analyses and as a primary screening tool, giving low detection limits.[35]

Fungicide residues have also been monitored using GC/MS/MS, providing confirmation of pesticides with a high degree of certainty and little sample preparation.[87]

6.9 Fumigants

Methyl bromide is one of the most widely used chemicals in fumigation. Its residues have been determined indirectly as total inorganic bromide. Various methods are available for its direct determination in different environmental matrices in field conditions.

The detection of methyl bromide in air is important at three levels: control readings for warning fumigation workers, workplace measurements and the measurement of methyl bromide in the atmosphere. Pellizari et al. developed a GC/MS method for air samples in which, after two adsorption steps on charcoal, the analyte is vaporized onto a GC glass capillary column, followed by MS to achieve detection limits at the nanograms per cubic meter level.[118]

For methyl bromide in water samples, headspace and purge-and-trap sampling methods have been generally used prior to its GC/MS determination.[119–121] Using an isotope-labeled internal standard, methyl bromide has been detected at the milligrams per liter level in water samples.[122] With methanol or polyethylene glycol extraction before GC/MS analyses of soil samples, as suggested in two different USEPA methods, detection limits of 1 mg g^{-1} have been achieved.[120,123]

GC/MS determination of methyl bromide has been also performed with fish and post-mortem tissues. For fish samples, homogenization and purge-and-trap sampling methods have been coupled with GC/MS, reaching detection limits at the milligrams per kilogram level;[124] for post-mortem tissues, generally GC/ECD procedures have been used and MS detection has been employed as a confirmatory technique.[125]

6.10 Organotin Compounds

Generally, GC/MS analysis of tributyltin (TBT) is preceded by derivatization procedures. TBT can be converted to butyltin hydrides[126,127] or tetraalkyl compounds by Grignard reactions.[12–15] The Grignard derivatives are less volatile than organotin hydrides and are thermally more stable, thus facilitating MS analysis. In estuarine waters, after solvent extraction and derivatization with hexylmagnesium bromide, the tetraalkyl derivatives have been quantified with gas chromatography/flame photometric detection (GC/FPD), performing GC/MS experiments as confirmatory analyses. CI was chosen rather than EI because of the better sensitivity and because the formation of high-mass, high-intensity ions is ideal for SIM analyses of these compounds. Internal calibration was used to monitor nanograms per liter levels in environmental and laboratory samples.[15]

Tricyclohexyltin hydroxide (cyhexatin) is an acaricide effective against mites, used in a wide range of vegetables. It has been analyzed in animal feed samples using GC/MS (SIM) in the EI mode. Converting the analyte to the alkylated derivative and using a tricyclopentyl analog as an internal standard, LOD and quantitation at the nanograms per gram level have been achieved, showing more specificity than the other conventional detection techniques.[16]

7 CONCLUSIONS

GC/MS is now an established method for pesticide analysis and is versatile, sensitive, selective and accurate. Many different methods can be developed. GC/MS is used to confirm quantitative analyses with other detectors. However, with minimal effort laboratories could fully apply the powerful capabilities of GC/MS making a single GC/MS quantitative analysis, which is more robust, avoiding the need for confirmation after a first analysis with a less selective detector. GC/MS is an excellent technique for the identification of pesticides and related compounds and for their quantification. The approaches for these two different analyses are different and generally GC/MS is not as fast as other GC methods. Furthermore, its cost is higher. However, these factors are becoming less important as new, easier and cheaper instruments come on to the market, allowing the wider use of the technique.

CAS NUMBERS

Acetic acid	64-19-7
Acetic anhydride	108-24-7
Acetochlor	34256-82-1
Acetone	67-64-1
Acetonitrile	75-05-8
Alachlor	15972-60-8
Alumina	1344-28-1
Atrazine	1912-24-9
Benzene	71-43-2
Benzyl bromide	100-39-0
BSTFA	25561-30-2
Chlorothalonil	1897-45-6
Cyhexatin	13121-70-5

Dacthal	1861-32-1
DDD	72-54-8
DDE	72-55-9
DDT	50-29-3
Deltamethrin	52918-63-5
Diazinon	333-41-5
Diazomethane	334-88-3
Diphenylamine	122-39-4
Diquat	85-00-7
Endosulfan	115-29-7
Ethyl acetate	141-78-6
Ethylmagnesium bromide	925-90-6
Fenamiphos	22224-92-6
Fluazifop-P-butyl	79241-46-6
Glyphosate	1071-83-6
HCB	118-74-1
HCH	608-73-1
Hexane	110-54-3
Linuron	330-55-2
Maneb	12427-38-2
Methanol	67-56-1
Methyl bromide	74-83-9
Metolachlor	51218-45-2
MTBSTFA	77377-52-7
Paraquat	4685-14-7
Pentafluorobenzyl bromide	1765-40-8
o-Phenylphenol	90-43-7
Polyethylene glycol	25322-68-3
Procymidone	32809-16-8
Profenofos	41198-08-7
2-Propanol	67-63-0
Propargite	2312-35-8
Propazine	139-40-2
Simazine	122-34-9
Sodium acetate	127-09-3
Sodium sulfate	7757-82-6
Sulfur	7704-34-9
Toxaphene	8001-35-2
Trichlorfon	52-68-6
Trimethylanilinium hydroxide	1899-02-1
Vinclozolin	50471-44-8

ACKNOWLEDGMENTS

We acknowledge the financial support of the European Commission (IC15 CT96-0802).

ABBREVIATIONS AND ACRONYMS

ASE	Accelerated Solvent Extraction
BSTFA	*N*,*O*-bis(trimethylsilyl)trifluoroacetamide
CAD	Collisionally Activated Dissociation
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CI	Chemical Ionization
DDD	1,1-Dichloro-2,2-bis (4-chlorophenyl)ethane
DDE	1,1-Dichloro-2,2-bis (4-chlorophenyl)ethylene
DDT	1,1,1-Trichloro-2,2-bis (4-chlorophenyl)ethane
DIS	Direct Inlet System
ECD	Electron Capture Detection
EI	Electron Ionization
ELISA	Enzyme-linked Immunosorbent Assay
EU	European Union
FAB	Fast Atom Bombardment
FDA	Food and Drug Administration
FID	Flame Ionization Detection
FPD-P	Phosphorus Flame Photometric Detection
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/FPD	Gas Chromatography/Flame Photometric Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GC/MS/MS	Gas Chromatography/Tandem Mass Spectrometry
GC/NICIMS	Gas Chromatography/Negative Ion Chemical Ionization Mass Spectrometry
GC/NPD	Gas Chromatography/Nitrogen–Phosphorus Detection
GPC	Gel Permeation Chromatography
HCB	Hexachlorobenzene
HCH	Hexachlorocyclohexane
LC	Liquid Chromatography
LC/GC/MS	Liquid Chromatography/Gas Chromatography/Mass Spectrometry
LC/MS	Liquid Chromatography/Mass Spectrometry
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NICI	Negative Ion Chemical Ionization
NPD	Nitrogen–Phosphorus Detection
OPP	Organophosphorus Pesticide
PA	Polyacrylate
PCB	Polychlorinated Biphenyl
PCP	Pentachlorophenol
PICI	Positive Ion Chemical Ionization

PPNICI	Pulsed Positive–Negative Ion Chemical Ionization
PUF	Polyurethane Foam
SCA	Standing Committee of Analysis
SECI	Selected Ejection Chemical Ionization
SFE	Supercritical Fluid Extraction
SFE/GC/MS	Supercritical Fluid Extraction/Gas Chromatography/Mass Spectrometry
SIM	Selected Ion Monitoring
SIR	Selected Ion Recording
SLE	Solid–Liquid Extraction
S/N	Signal-to-noise
SPE	Solid Phase Extraction
SPE/GC/MS	Solid phase Extraction/Gas Chromatography/Mass Spectrometry
SPME	Solid-phase Microextraction
TBT	Tributyltin
TEA	Thermal Energy Analyzer
TP	Transformation Product
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Mass Spectrometry **(Volume 13)**
Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry

REFERENCES

1. M. Fielding, D. Barceló, A. Helweg, S. Galassi, L. Torstensson, P. Van Zoonen, G. Angeletti, 'Pesticides in Ground and Drinking Water', Commission of the European Communities: Water Pollution Research Report 27, 1992.
2. http://www.epa.gov/OGWDW/Pubs/new/dwsp/html #anal1, 1998.
3. M. Natangelo, S. Romagnano, L. Guzzella, E. Benfenati, 'GC Analysis of Some Organochlorine Pesticides using a Brominated Internal Standard', *Int. J. Environ. Anal. Chem.*, **58**, 55–66, (1995).
4. D. Barceló, 'EPA and Other Methods for the Determination of Priority Pesticides and Their Transformation Products in Water: A Review', *J. Chromatogr. A*, **643**, 117–143 (1993).
5. P. Grasso, E. Benfenati, M. Terreni, M. Pregnolato, M. Natangelo, G. Pagani, 'Deuterated Internal Standards for Gas-chromatographic–Mass Spectrometric Analysis of Polar Organophosphoros Pesticides in Water Samples', *J. Chromatogr. A*, **822**, 91–99 (1998).
6. R.E. Cline, G.D. Todd, D.L. Ashley, J. Grainger, J.M. McCraw, C.C. Alley, R.H. Hill, Jr, 'Gas Chromatographic and Spectral Properties of Pentafluorobenzyl Derivatives of 2,4-Dichlorophenoxyacetic Acid and Phenolic Pesticides and Metabolites', *J. Chromatogr. Sci.*, **28**, 167–172 (1990).
7. T.S. Thompson, B.D. Miller, 'Use of Solid-phase Extraction Disks for the GC/MS Determination of Acidic and Neutral Herbicides in Drinking Water', *Chemosphere*, **36**, 2867–2878 (1998).
8. T.S. Thompson, R.G. Treble, 'Use of Pine Needles as an Indicator of Atmospheric Contamination of Pentachlorophenol', *Chemosphere*, **31**, 4387–4392 (1995).
9. H.D. Meiring, G. den Engelsman, A.P.J.M. de Jong, 'Determination of Polar Pesticides by Phase-transfer Catalyzed Derivatization and Negative-ion Chemical Ionization Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **644**, 357–365 (1993).
10. S. Butz, Th. Heberer, H.-J. Stan, 'Determination of Phenoxyalkanoic Acids and Other Acidic Herbicides at the Low ppt Level in Water Applying Solid-phase Extraction with RP-C_{18} Material', *J. Chromatogr. A*, **667**, 63–74 (1994).
11. T. Nilsson, D. Baglio, I. Gualdo-Miguez, 'Solid-phase Microextraction for the Analysis of Phenoxy Acid Herbicides', presented at the 20th International Symposium on Capillary Chromatography, May 26–29, 1998, Riva del Garda, Italy, 1998.
12. R.J. Maguire, Y.G.A. Bengert, E.J. Hale, P.T.S. Wong, O. Kramer, 'Occurrence of Organotin Compounds in Ontario Lakes and Rivers', *Environ. Sci. Technol.*, **16**, 698–702 (1982).
13. H.A. Meinema, T. Burger-Wiersma, G. Verslius-de Haan, E.M. Gevers, 'Determination of Trace Amounts of Butyltin Compounds in Aqueous Systems by Gas Chromatography/Mass Spectrometry', *Environ. Sci. Technol.*, **12**(3), 288–293 (1978).
14. M.D. Müller, 'Tributyltin Detection at Trace Levels in Waters and Sediments by GC with Flame Photometric Detection and GC/MS', *Fresenius' Z. Anal. Chem.*, **317**, 32–36 (1984).
15. M.A. Unger, W.G. MacIntyre, J. Greaves, R.J. Huggett, 'GC Determination of Butyltins in Natural Waters by Flame Photometric Detection of Hexyl Derivatives with Mass Spectrometric Confirmation', *Chemosphere*, **15**, 461–470 (1986).
16. E. Benfenati, M. Natangelo, E. Pallucca, R. Tridico, E. Borghetti, G. Lualdi, 'A Specific Analytical Method to Quantify Cyhexatin in Animal Feed Using Gas Chromatography/Mass Spectrometry', *J. Chromatogr. B*, **605**, 129–133 (1992).
17. J. Perez, M. Garcia-Baudin, J.L. Tadeo, 'Determination of Chlorotoluron, Isoproturon and Metoxuron in Soil by GLC/NPD and Confirmation using GLC/MS', *Fresenius' Z. Anal. Chem.*, **339**, 413–416 (1991).

18. N. Tsunoda, 'Simultaneous Determination of the Herbicides Glyphosate, Glufosinate, and Bialaphos and Their Metabolites by Capillary Gas Chromatography/Ion Trap Mass Spectrometry', *J. Chromatogr. A*, **637**, 167–173 (1993).
19. C.L. Deyrup, S.-M. Chang, R.A. Weintraub, H. Anson Moye, 'Simultaneous Esterification and Acylation of Pesticides for Analysis by Gas Chromatography. 1. Derivatization of Glyphosate and (Aminomethyl)phosphonic Acid with Fluorinated Alcohols–Perfluorinated Anhydrides', *J. Agric. Food Chem.*, **33**, 944–947 (1985).
20. H. Faerber, H.F. Schoeler, 'Gas Chromatographic Determination of Urea Herbicides in Water after Methylation with Trimethylanilinium Hydroxide or Trimethylsulfonium Hydroxide', *Vom Wasser*, **77**, 249–262 (1991).
21. H.-J. Stan, P. Klaffenbach, 'Determination of Carbamate Pesticides and Some Urea Pesticides after Derivatization with Acetic Anhydride by means of GC/MSD', *Fresenius' J. Anal. Chem.*, **339**, 151–157 (1991).
22. P. Klaffenbach, P.T. Holland, 'Analysis of Sulfonylurea Herbicides by Gas/Liquid Chromatography. III – Mass Spectrometry and Multiresidue Determination', *Biol. Mass Spectrom.*, **22**, 565–578 (1993).
23. P. Klaffenbach, P.T. Holland, 'Analysis of Sulfonylurea Herbicides by Gas–Liquid Chromatography. Determination of Chlorsulfuron and Metsulfuron-methyl in Soil and Water Samples', *J. Agric. Food Chem.*, **41**, 396–401 (1993).
24. S. Mangiapan, E. Benfenati, P. Grasso, M. Terreni, M. Pregnolato, G. Pagani, D. Barcelò, 'Metabolites of Alachlor in Water: Identification by Mass Spectrometry and Chemical Synthesis', *Environ. Sci. Technol.*, **31**, 3637–3646 (1997).
25. M. Natangelo, E. Benfenati, G. De Gregorio, R. Fanelli, G. Ciotti, 'GC Analysis of Glyphosate Through Derivatization with a Perfluoroanhydride and Trifluoroethanol: Identification of Systematic Sources of Error', *Toxicol. Environ. Chem.*, **38**, 225–232 (1993).
26. R. Frassanito, E. Benfenati, G. Ciotti, R. Fanelli, 'Determination of *N*-Nitrosamine Traces in Pesticide Samples by GC/MS', *Toxicol. Environ. Chem.*, **45**, 199–204 (1994).
27. C. Aguilar, S. Penalver, E. Pocurull, F. Borrull, R.M. Marcé, 'Solid-phase Microextraction and Gas Chromatography with Mass Spectrometry Detection for the Determination of Pesticides in Aqueous Samples', *J. Chromatogr. A*, **795**, 105–115 (1998).
28. S. Tavazzi, M. Natangelo, E. Benfenati, 'SPME Methodology for the Analysis of Some Pesticides in Environmental Waters and Agricultural Products using GC/MS and GC/MS/MS', 2nd Euroconference on Environmental and Analytical Chemistry, 31 October–6 November, 1998, Cordoba, Spain, Book of Abstracts, CL 24, 1998.
29. A. Aguera, M. Contreras, 'Gas Chromatographic Analysis of Organophosphorus Pesticides of Horticultural Concern', *J. Chromatogr. A*, **655**, 293–300 (1993).
30. C. Sanchez-Brunete, L. Martinez, J.L. Tadeo, 'Determination of Corn Herbicides by GC/MS and GC/NPD in Environmental Samples', *J. Agric. Food Chem.*, **42**, 2210–2214 (1994).
31. W.Bruns, S. Nelson, D.G. Erickson, 'Determination of MCPA, Bromoxynil, Trifluralin, Triallate, Picloram and Diclofop-methyl in Soil by GC/MS Using Selected Ion Monitoring', *J. Assoc. Off. Anal. Chem.*, **74**, 550 (1991).
32. W. Pereira, C.L. Rostad, T.J. Leiker, 'Determination of Trace Level Herbicides and Their Degradation Products in Surface and Groundwaters by Gas Chromatography/Ion Trap Mass Spectrometry', *Anal. Chim. Acta*, **228**, 69–75 (1990).
33. S. Hong, A.T. Lemley, 'Gas Chromatographic–Mass Spectrometric Determination of Alachlor and its Degradation Products by Direct Aqueous Injection', *J. Chromatogr. A*, **822**(2), 253–261 (1998).
34. L.Q. Huang, 'Simultaneous Determination of Alachlor, Metolachlor, Atrazine, and Simazine in Water and Soil by Isotope Dilution Gas Chromatography/Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **72**, 349–354 (1989).
35. J.L. Bernal, M.J. de Nozal, J.M. Rivera, J.J. Jimenez, J. Atienza, 'Determination of Fungicide Vinclozolin in Honey and Bee Larvae by Solid-phase and Solvent Extraction with Gas Chromatography and Electron Capture and Mass Spectrometric Detection', *J. Chromatogr. A*, **754**, 507–513 (1996).
36. M. Yasin, P.J. Baugh, G.A. Bonwick, D.H. Davies, P. Hancock, M. Leinoudi, 'Analytical Method Development for the Determination of Synthetic Pyrethroid Insecticides in Soil by Gas Chromatography/Mass Spectrometry Operating in Negative-ion Chemical Ionization Mode', *J. Chromatogr. A*, **754**, 235–243 (1996).
37. 'Determination of Organic Compounds in Drinking Water by Liquid–Solid Extraction and Capillary Column Gas Chromatography/Mass Spectrometry', EPA Method 525, Environmental Protection Agency, Washington, DC, 1988.
38. E. Davoli, E. Benfenati, R. Bagnati, R. Fanelli, 'Analysis of Atrazine in Groundwaters at Part per Trillion Levels as an Early Warning Method for Contamination and Soil Distribution Studies', *Chemosphere*, **16**, 1425–1430 (1987).
39. Z. Cai, V.M. Sadagopa Ramanujam, D.E. Giblin, M.L. Gross, 'Determination of Atrazine in Water at Low- and Sub-part-per-trillion Levels by Using Solid-phase Extraction and Gas Chromatography/High-resolution Mass Spectrometry', *Anal. Chem.*, **65**, 21–26 (1993).
40. E. Pocurull, C. Aguilar, F. Borrul, R.M. Marcé, 'On-line Coupling of Solid-phase Extraction to Gas Chromatography with Mass Spectrometric Detection to Determine Pesticides in Water', *J. Chromatogr. A*, **818**, 85–93 (1998).

41. E. Benfenati, P. Tremolada, L. Chiappetta, R. Frassanito, G. Bassi, N. Di Toro, R. Fanelli, G. Stella, 'Simultaneous Analyses of 50 Pesticides in Water Samples by Solid-phase Extraction and GC/MS', *Chemosphere*, **21**(12), 1411–1421 (1990).
42. R. Eisert, K. Levsen, 'Determination of Pesticides in Aqueous Samples by Solid-phase Microextraction Inline Coupled to Gas Chromatography/Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **6**, 1119–1130 (1995).
43. A. Boyd-Boland, J.B. Pawliszyn, 'Solid-phase Microextraction of Nitrogen Containing Herbicides', *J. Chromatogr. A*, **704**, 163–172 (1995).
44. T.K. Choudhury, K.O. Gerrhardt, T.P. Mawhinney, 'Solid-phase Microextraction of Nitrogen and Phosphorus-containing Pesticides from Water and Gas Chromatographic Analysis', *Environ. Sci. Technol.*, **30**, 3259–3265 (1996).
45. C.G. Arnold, M. Berg, S.R. Müller, U. Dommann, R.P. Schwarzenbach, 'Determination of Organotin Compounds in Water, Sediments, Sewage Sludge Using Perdeuterated Internal Standards, Accelerated Solvent Extraction, Large Volume Injection GC/MS', *Anal. Chem.*, **70**, 3094–3101 (1998).
46. F.I. Onuska, K.A. Terry, A. Seech, M. Atonic, 'Determination of Toxaphene in Soil by Electron Capture Negative-ion Mass Spectrometry and Capillary Column Gas Chromatography', *J. Chromatogr. A*, **665**, 125–132 (1994).
47. H.-J. Brauch, S. Schullerer, 'Strategy for Pesticide Control in Groundwater and Drinking Water', *Fresenius' J. Anal. Chem.*, **339**, 420–422 (1991).
48. H.-J. Stan, 'Automated Recognition of Target Compounds at Low Levels in Environmental Samples by Means of Capillary Gas Chromatography/Mass Spectrometry with Dedicated Mass Spectral Libraries and the Macro Program AUTARG', *J. Chromatogr. A*, **653**, 55–62 (1993).
49. D. Barceló, 'Occurrence, Handling and Chromatographic Determination of Pesticides in Aquatic Environment', *Analyst*, **116**, 681–689 (1991).
50. I. Liska, J. Krupcikk, P.A. Leclercq, 'The Use of Solid Sorbent for Direct Accumulation of Organic Compounds from Water Matrices – A Review of Solid-phase Extraction Technique', *J. High Resolut. Chromatogr.*, **12**, 577–590 (1989).
51. R. Bagnati, E. Benfenati, E. Davoli, R. Fanelli, 'Screening of 21 Pesticides in Waters by Single Extraction with C_{18} Silica Bonded Phase Column and HRGC/MS', *Chemosphere*, **17**, 59–65 (1988).
52. C.E. Rostad, W.E. Pereira, S.M. Ratcliff, 'Bonded-phase Extraction Column Isolation of Organic Compounds in Groundwater at a Hazardous Waste Site', *Anal Chem.*, **56**, 2856–2860 (1984).
53. C.E. Lindley, J.T. Stewart, M.W. Sandstorm, 'Determination of Low Concentration of Acetochlor in Water by Automated Solid-phase Extraction and Gas Chromatography with Mass-selective Detection', *J. AOAC Int.*, **79**(4), 962–966 (1996).
54. J. Vreuls, A.-J. Bulterman, R.T. Ghijinson, U.A.Th. Brinkman, 'On-line Preconcentration of Aqueous Samples for Gas Chromatographic–Mass Spectrometric Analysis', *Analyst*, **117**, 1701–1705 (1992).
55. E. Benfenati, G. Facchini, P. Pierucci, R. Fanelli, 'Identification of Organic Contaminants in Leachates in Industrial Waste Landfields', *Trends Anal. Chem.*, **15**, 305–310 (1996).
56. A. Boyd-Boland, S. Magdic, J.B. Pawliszyn, 'Simultaneous Determination of 60 Pesticides in Water Using Solid-phase Microextraction and Gas Chromatography/Mass Spectrometry', *Analyst*, **121**, 929–938 (1996).
57. M. Natangelo, D. Giavini, E. Benfenati, R. Fanelli, 'Estrazione in Fase Supercritica (SFE): Teoria ed Esperienze Pratiche', *Laboratorio 2000*, **10**, 46–52 (1996).
58. M. Terreni, E. Benfenati, M. Pregnolato, A. Bellini, D. Giavini, S.F. Bavetta, C. Molina, D. Barcelò, 'Synthesis and Use of Deuterated Fenamiphos and Metabolites as Internal Standards for Mass Spectrometry Analysis in Water', *J. Chromatogr. A*, **754**, 207–219 (1996).
59. W.C. Brumley, C.M. Brownrigg, 'Determination of Toxaphene in Soil by Electron Capture Negative-ion Mass Spectrometry after Fractionation by High Performance Gel Permeation Chromatography', *J. Chromatogr. A*, **633**, 177–183 (1993).
60. R.C. Dougherty, J. Dalton F.J. Biros, 'Negative Chemical Ionization Mass Spectra of Polycyclic Chlorinated Pesticides', *Org. Mass Spectrom.*, **6**, 1171–1181 (1972).
61. G. Durand, P. Gille, D. Fraisse, D. Barceló, 'Comparison of Gas Chromatographic–Mass Spectrometric Methods for Screening of Chlorotriazine Pesticides in Soil', *J. Chromatogr. A*, **603**, 175–184 (1992).
62. R. Wittlinger, K. Ballschmiter, 'Global Baseline Pollution Study XI: Congener Specific Determination of Polychlorinated Biphenyls (PCB) and Occurrence of Alpha- and Gamma-Hexachlorocyclohexane (HCH), 4,4′-DDE and 4,4′-DDT in Continental Air', *Chemosphere*, **16**(10–12), 2497–2513 (1987).
63. K. Haraguchi, E. Kitamura, T. Yamashito, K. Azuma, 'Simultaneous Determination of Trace Pesticides in Urban Air', *Atmos. Environ.*, **28**(7), 1319–1325 (1994).
64. M.A. Brown, M.X. Petreas, H.S. Hokamoto, T.M. Mischke, R.D. Stephens, 'Monitoring of Malathion and Its Impurities and Environmental Transformation Products on Surface and in Air Following Aerial Application', *Environ. Sci. Technol.*, **27**, 388–397 (1993).
65. K.S. Roinestad, J.B. Louis, J.D. Rosen, 'Determination of Pesticides in Indoor Air and Dust', *J. Assoc. Off. Anal. Chem.*, **76**(5), 1121–1126 (1993).
66. T. Yamashita, K. Haraguchi, A. Kido, 'Determination of Chlordane in Air by Gas Chromatography Mass Spectrometry with Selected Ion Monitoring', *J. Chromatogr. A*, **657**, 405–411 (1993).

67. J.P. Hsu, H.G. Wheeler, Jr, D.E. Camann, H.J. Schattenberg 3rd, R.G. Lewis, A.E. Bond, 'Analytical Methods for Detection of Nonoccupational Exposure of Pesticides', *J. Chromatogr. Sci.*, **26**(4), 181–189 (1988).
68. F. Artigas, E. Martinez, E. Gelpi, 'Organochlorine Pesticides by Negative Ion Chemical Ionization. Brain Metabolites of Lindane', *Biomed. Environ. Mass. Spectrom.*, **16**, 279–284 (1988).
69. G. Mariani, E. Benfenati, R. Fanelli, 'A NICI/GC/MS Method to Analyze Endosulfan in Biological Samples', *Int. J. Environ. Anal. Chem.*, **58**, 67–72 (1995).
70. A. Reischl, M. Reissinger, O. Hutzinger, 'Occurrence and Distribution of Atmospheric Organic Micropollutants in Conifer Needles', *Chemosphere*, **16**, 2647–2652 (1987).
71. M.T. Galceran, F.J. Santos, 'PCBs and Chlorinated Pesticides in Shellfish of a Deltraic Environment', *Chemosphere*, **27**(7), 1183–1200 (1993).
72. F.I. Onuska, K.A. Terry, 'Quantitative High Resolution Gas Chromatography and Mass Spectrometry of Toxaphene Residues in Fish Samples', *J. Chromatogr. A*, **471**, 161–171 (1989).
73. D.L. Swackhamer, M.J. Charles, R.A. Hites, 'Quantitation of Toxaphene in Environmental Samples using Negative Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 913–917 (1987).
74. A. Anderson, T. Bergh, 'Pesticides Residues in Fresh Fruits and Vegetables on the Swedish Market', *Fresenius' J.Anal. Chem.*, **339**, 387–389 (1991).
75. N. Motohashi, H. Nagashima, C. Parkanyi, B. Subrahmanyam, G. Zhang, 'Official Multiresidue Method for Pesticides Analysis in Vegetables, Fruits and Soil', *J. Chromatogr. A*, **754**, 333–346 (1996).
76. T. Cairns, K.S. Chiu, D. Navarro, E. Siegmund, 'Multiresidue Pesticide Analysis by Ion Trap Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 971–988 (1993).
77. A. Valverde-Garcia, E. Gonzales-Pradas, A. Aguilera-Del Real, M.D. Urena-Amate, 'Determination and Degradation Study of Chlorotalonil Residues in Cucumbers, Peppers and Cherry Tomatoes', *Anal. Chim. Acta*, **276**, 15–23 (1993).
78. E. Lacassie, M.-F. Dreyfuss, J.L. Daguet, M. Vignaud, P. Marquet, G. Lachâtre, 'Multiresidue Determination of Pesticides in Apples and Pears by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **805**, 319–326 (1998).
79. K.-C. Ting, G.S. Tamashiro, 'Off-line High Performance Liquid Chromatography and Solid-phase Extraction Clean-up of Pesticides Residues in Fresh Produce by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **754**, 455–462 (1996).
80. L. Yu, R. Schoen, A. Dunkin, M. Firman, H. Cushman, A. Fontalnilla, 'Determination of *o*-Phenylphenol, Diphenylamine, and Propargite Pesticides Residues in Selected Fruits and Vegetables by Gas Chromatography–Mass Spectrometry, *J. AOAC Int.*, **80**, 651–656 (1997).
81. J. Hajslová, P. Cuhra, T. Davidek, 'Gas Chromatographic Determination of Diquat and Paraquat in Crops', *J. Chromatogr. A*, **479**, 243–250 (1989).
82. L.G.M. Tuinstra, A.H. Roos, A.M. Matser, W.A. Traah, J.A. van Rhijn, 'Development of a Multiresidue/Multimatrix Method for Pesticides Analysis in Agricultural Products', *Fresenius' J. Anal. Chem.*, **339**, 384–386 (1991).
83. L.G.M. Tuinstra, F.R. Povel, A.H. Roos, 'Multi-matrix–Multi-pesticide Method for Agricultural Products', *J. Chromatogr. A*, **552**, 259–264 (1991).
84. S.M.M. Ismail, H.M. Ali, R.A. Habiba, 'GC/ECD and GC/MS Analyses of Profenofos Residues and its Biochemical Effects in Tomatoes and Tomato Products', *J. Agric. Food Chem.*, **41**, 610–615 (1993).
85. J. Simal Gàndara, P. Paseiro Losada, V. Gonzalez Rodriguez, A. Romero Rodriguez, 'Comparison of GC/MS and HPLC/UV for Determination of Procymidone in Wine', *J. Agric. Food Chem.*, **41**, 674–677 (1993).
86. P. Rodríguez, J. Permanyer, J.M. Grases, C. Gonzalez, 'Confirmation Method for the Identification and Determination of Some Organophosphorus and Organochlorine Pesticides in Cocoa Beans by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **562**, 547–553 (1991).
87. S. Schachterle, C. Feigel, 'Pesticide Residue Analysis in Fresh Produce by Gas Chromatography–Tandem Mass Spectrometry', *J. Chromatogr. A*, **754**, 411–422 (1996).
88. G.C. Mattern, G.M. Singer, J. Louis, M. Robson, J.D. Roseen, 'Determination of Linuron in Potatoes using Capillary Column Gas Chromatography/Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **72**, 970–974 (1989).
89. M.A. Luke, J.E. Froberg, H.T. Masumoto, 'Extraction and Clean-up of Organochlorine, Organophosphate, Organonitrogen and Hydrocarbon Pesticides in Produce for Determination by Gas–Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **58**, 1020–1026 (1975).
90. P.A. Mills, 'Variation of Fluorisil Activity; Simple Method for Measuring Adsorbent Capacity and its Use in Standardizing Fluorisil Column', *J. Assoc. Off. Anal. Chem.*, **51**(1), 29–32 (1968).
91. S.J. Lehotay,K.I. Heller, 'Development of a Method of Analysis for 46 Pesticides in Fruits and Vegetables by Supercritical Fluid Extraction and Gas Chromatography/Ion Trap Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **78**, 821–830 (1995).
92. Y.Y. Wigfield, M. Lanouette, 'Supercritical Fluid Extraction of the Fortified Residues of Fluazifop-P-butyl (Fusilade II) and Its Major Metabolites, Fluazifop-P, in Onions', *J. Agric. Food Chem.*, **41**, 84–88 (1993).
93. M. Vitali, M. Guidotti, R. Giovinazzo, O. Cedrone, 'Determination of Pesticides Residues in Wine by SPME and

GC/MS for Consumer Risk Assessment', *Food Addit. Contam.*, **15**, 280–287 (1998).
94. J.D. Gaynor, D.A. Cancilla, G.R. Barrie Webster, L.P. Sarna, K.N. Graham, H.Y.F. Ng, Chin S. Tan, C.F. Drury, T. Welacky, 'Comparative Solid-phase Extraction, Solid-phase Microextraction and Immunoassay Analyses of Metolachlor in Surface Runoff and Tile Drainage', *J.Agric. Food Chem.*, **44**, 2736–2741 (1996).
95. E.M. Thurman, M. Meyer, M. Pomes, C. A. Perry, A.P. Schwab, 'Enzyme-linked Immunosorbent Assay Compared with Gas Chromatography/Mass Spectrometry for the Determination of Triazine Herbicides in Water', *Anal. Chem.*, **62**, 2043–2048 (1990).
96. B. Gruessner, N.C. Shambaugh, M.C. Watzin, 'Comparison of Enzyme Immunoassay and Gas Chromatography/Mass Spectrometry for the Detection of Atrazine in Surface Water', *Environ. Sci. Technol.*, **29**, 251–254 (1995).
97. L. Urruty, M. Montury, M. Braci, J. Fournier, J.-M. Dournel, 'Comparison of Two Recent Solventless Methods for the Determination of Procymidone Residues in Wines: SPME/GC/MS and ELISA Test', *J.Agric. Food Chem.*, **45**, 1519–1522 (1997).
98. 'GC/MS Method For Semivolatiles Organics: Packed Column Technique', Test Methods For Evaluating Solid Wastes (SW-846), EPA Method 8250, Environmental Protection Agency, Washington, DC, 1988.
99. 'GC/MS Method For Semivolatiles Organics: Capillary Column Technique', Test Methods For Evaluating Solid Wastes (SW-846), EPA Method 8270, Environmental Protection Agency, Washington, DC, 1987.
100. G.C. Mattern, J.B. Louis, J.D. Rosen, 'Multipesticide Determination in Surface Water by Gas Chromatography/Chemical Ionization/Mass Spectrometry/Ion Trap Detection', *J. Assoc. Off. Anal. Chem.*, **74**(6), 982–986 (1991).
101. W. Liao, T. Joe, W.J. Cusick, 'Multiresidue Screening Method for Fresh Fruit and Vegetables with Gas Chromatographic/Mass Spectrometric Detection', *J.Assoc. Off. Anal. Chem.*, **74**(3), 554–565 (1991).
102. P. Klaffenbach, H.J. Stan, 'Automated Screening of Food and Water Samples for Pesticide Residues by means of Capillary GC/MSD using a Library of Pesticide Spectra and Macro Programming', *J.High Resolut. Chromatogr.*, **14**, 754–756 (1991).
103. W. Specht, M. Tillkes, 'Gas-chromatographische Bestimmung von Rückständen an Pflanzenbehandlungsmitteln nach Clean-up über Gel-Chromatographie und Mini-Kieselgel-Säulen-Chromatographie', *Fresenius' Z. Anal. Chem.*, **301**, 300–307 (1980).
104. Y. Naude, W.H.J. deBeer, S. Jooste, L. vanderMerwe, S.J. vanRensburg, 'Comparison of Supercritical Fluid Extraction and Soxhlet Extraction for the Determination of DDT, DDD and DDE in Sediment', *Water SA*, **24**(3), 205–214 (1998).
105. R.H. Kaake, M. Orr, M.W. Watwood, 'A Comparison of Ultrasonical Extraction and Soxhlet Methods for DDT Extraction from Soil', *J. Soil Contam.*, **7**(5), 589–597 (1998).
106. K. Warmann, 'Report on the Distribution of the Chlorinated Aromatics Around a Remote-located Industrial Site', *Chemosphere*, **20**, 1511–1516 (1990).
107. S. Stout, A.R. daCunha, J.E. Boyd, J.M. Devine, 'Confirmation of Phorate, Terbufos and their Sulfoxides and Sulfones in Water by Capillary Gas Chromatography/Chemical Ionization Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **72**(6), 971–992 (1989).
108. H.-J. Stan, G. Kellner, 'Confirmation of Organophosphorus Pesticide Residue in Food Applying Gas Chromatography/Mass Spectrometry with Chemical Ionization and Pulsed Positive Negative Detection', *Biomed. Mass Spectrom.*, **18**, 645–651 (1989).
109. D.F. Hunt, G.C. Stafford, Jr, F.W. Crow, 'Pulsed Positive Negative Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **48**(14), 2098 (1976).
110. A. Tanabe, H. Mitobe, K. Kuwata, M. Sakai, 'Monitoring of Herbicides in River Water by Gas Chromatography/Mass Spectrometry and Solid-phase Extraction', *J. Chromatogr. A*, **754**(1–2), 159–168 (1996).
111. G. Durand, V. Bouvot, D. Barceló, 'Determination at Trace Level of Herbicides in Estuarine Waters by Gas and Liquid Chromatographic Techniques', *J. Chromatogr. A*, **607**, 319–327 (1992).
112. Z.-M. Chen, Y.-H. Wang, 'Chromatographic Methods for the Determination of Pyrethrin and Pyrethroid Pesticide Residues in Crops, Foods and Environmental Samples', *J. Chromatogr. A*, **754**, 367–395 (1996).
113. *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th edition, AOAC, Arlington, VA, 1990.
114. T.J. Class, 'Determination of Pyrethroids and Their Degradation Products in Indoor Air and on Surface by HRGC/ECD and HRGC/MS(NCI)', *J. High Resolut. Chromatogr.*, **14**, 446–450 (1991).
115. G.F. Bauerle, K.L. Ray, J.S. Brodbelt, 'Determination of Pyrethroid Insecticides by Ion Trap GC/MS/MS', *Anal. Chim. Acta*, **317**, 137–148 (1995).
116. J. Krupcik, P.A. Leclercq, J. Garaj, J. Masaryk, 'Analysis of Divalent Metal Dialkyldithiocarbamates by Gas Chromatography and Mass Spectrometry', *J. Chromatogr.*, **171**, 285–304 (1979).
117. E. Matisovà, L. Kakalicova, J. Lesko, J. de Zeeuw, 'Application of Porous Carbon for Solid-phase Extraction of Dicarboxyimide Fungicide Residue from Wine in Combination with High-resolution Capillary Gas Chromatography and Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **754**, 445–454 (1996).
118. E.D. Pellizari, R.A. Zweidinger, M.D. Erickson, 'Environmental Monitoring Near Industrial Sites: Brominated Chemicals Part II: Appendix', EPA-560/6-78-002A; US

NTIS PB-286483, US Environmental Protection Agency, Washington, DC, 1978.
119. D.E. Gryder-Boutey, J.M. Kennish, 'Using Headspace Sampling with Capillary Column GC/MS to Analyze Trace Volatile Organics in Water and Wastewater', *Am. Water Works Assoc. J.*, **80**, 52–55 (1988).
120. 'GC/MS Method for Volatile Organics', Test Method for Evaluating Solid Wastes (SW 846), Method 8240, Environmental Protection Agency, Washington, DC, 1987.
121. 'Purgeables', Method 624, EPA/EMSL, Cincinnati, OH, 1982.
122. 'Volatile Organic Compounds by Isotope Dilution by GC/MS', Method 1624, Industrial Technology Division (ITD), Environmental Protection Agency, Washington, DC, 1989.
123. 'Halogenated Volatile organics', Method 8010, Test Method for Evaluating Solid Wastes (SW 846), Environmental Protection Agency, Washington, DC, 1987.
124. D.M. Easley, R.D. Kloepfer, A.M. Carasca, 'Gas Chromatographic–Mass Spectrometric Determination of Volatile Organic Compounds in Fish', *J. Assoc. Off. Anal. Chem.*, **64**, 653–656 (1981).
125. T. Honma, M. Miyagawa, M. Sato, H. Haegawa, 'Neurotoxicity of Methyl Bromide in Rats', *Toxicol. Appl. Pharmacol.*, **81**, 183–191 (1985).
126. V.F. Hodge, S.L. Seidel, E.D. Goldberg, 'Determination of Tin(IV) and Organotin Compounds in Natural Waters Coastal Sediment and Macro Algae by Atomic Absorption Spectrometry', *Anal. Chem.*, **51**(8), 1256–1259 (1979).
127. C.L. Mathias, G.J. Olson, F.E. Brinckman, J.M. Bellama, 'A Comprehensive Method for the Determination of Aquatic Butyltin and Butylmethyl Species at Ultra-trace Levels Using Simultaneous Hybridization/Extraction with GC/FPD', presented at the National Meeting of American Chemical Society, Miami Beach, FL, April 1985.

Herbicide Residues in Biota, Analysis of

J.V. Headley and L.C. Dickson
National Hydrology Research Centre, Saskatoon, Canada

A.J. Cessna
Agriculture and Agri-Food Research Centre, Lethbridge, Canada

1 Introduction 1
2 General Strategies 2
2.1 Quality Assurance/Quality Control 2
2.2 Type of Matrix 3
2.3 Type of Herbicide 3
2.4 Multiresidue Methods 3
2.5 Unknown Analytes 4
2.6 Miniaturization of Methodology 4
2.7 Performance Criteria for Analytical Methods 5
2.8 Sampling Biota 5
2.9 Extraction, Derivatization and Clean-up 6
2.10 Instrumental Determination 7
3 Sampling 8
3.1 Sampling Plants 8
3.2 Sampling Animals 9
4 Extraction, Derivatization and Clean-up 9
4.1 Extraction of Plant and Animal Tissues 10
4.2 Derivatization of Plant and Animal Tissue Extracts 13
4.3 Clean-up of Plant and Animal Tissue Extracts 13
5 Instrumental Determination 14
5.1 Gas Chromatographic Separations 14
5.2 High-performance Liquid Chromatographic Separations 16
5.3 Capillary Electrophoretic Separations 17
5.4 Immunochemical Methods 17
6 Official Multiresidue Methods for Pesticide Residue Analysis 18
Abbreviations and Acronyms 19
Related Articles 19
References 19

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Current extraction, derivatization and clean-up techniques, and instrumental methods are reviewed for the analysis of herbicide residues in biota. Sampling procedures are shown to be an integral part of the methodology. Herbicide analysis is seldom based on analyte-specific methods but is usually integrated into multiresidue methods (MRMs). Current methods generally utilize relatively small sample sizes and miniaturized apparatus to take advantage of advances in instrumental performance and detection of analytes. These methods reduce the amount of solvent used for sample preparation and help to minimize waste generation. As well as using less solvent, the resulting miniaturized methods tend to be generally cheaper, faster, and less labor-intensive than conventional methods and, furthermore, they reduce analyst exposure to hazardous materials.

Mass spectrometry (MS), interfaced with high-resolution gas chromatography (GC), high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE), has become the detection method of choice for herbicide analysis. MS is well suited to the confirmation of target analytes and the tentative identification of unknown analytes. HPLC–mass spectrometers are becoming more widely available and less expensive. CE is a relatively new separation technique providing many advantages over traditional gas and liquid chromatography, including shorter analysis times and smaller injection volumes. There have been advances in the development of immunoassays primarily for the rapid screening for herbicide residues. These methods, once optimized, can facilitate high sample throughput at relatively low cost compared to conventional approaches. To date however, development has been limited to aqueous systems and little immunoassay work has been done for the direct determination of herbicides in biota.

1 INTRODUCTION

Herbicides are generally synthetic organic compounds used to control unwanted vegetation or weeds. Application may either be directly to plants or soil to control unwanted terrestrial vegetation, or to water bodies to control aquatic weeds. Herbicide use plays an integral role in most crop production systems and agricultural production accounts for their greatest use. However, significant amounts of herbicides are also used in the forestry industry and to maintain pipeline, power-line and highway rights-of-way. Significant use also occurs within the urban setting to control weeds in lawns and gardens, and to maintain parks and golf courses.

Following application, a herbicide can undergo microbial or chemical degradation in soil, metabolism/conjugation in plants, and photodegradation by sunlight. It is

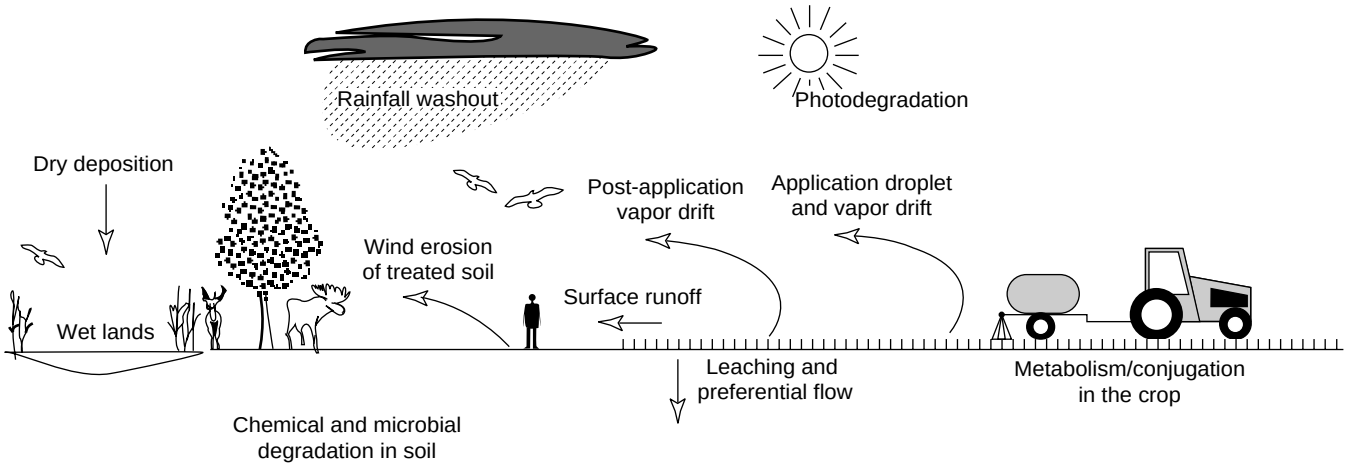

Figure 1 Herbicide degradation processes which occur following application, and transport mechanisms which remove a portion of the applied herbicide from the treated area.

also well established that not all of the herbicide or its metabolites/degradation products remain within the area to which the herbicide was applied (Figure 1). Atmospheric transport mechanisms include droplet and vapor drift during application, post-application volatility losses, and wind erosion of treated soil. Transport in water occurs with snow melt, rainfall or irrigation runoff from treated land and via leaching and preferential flow of herbicides under both dry land and irrigated agriculture. Finally, herbicides and their metabolite/degradation products may also be removed from treated areas as residues in edible portions of crops.

Such transport can contaminate the air, surface and groundwaters and wildlife habitat leading to human and wildlife exposure and decreased environmental quality through wildlife habitat degradation and decreased biodiversity. Public interest in human safety and environmental quality with respect to pesticide use, as well as the requirement for the agrochemical industry to provide toxicology, soil persistence and crop residue data to support the registration of pesticide uses, have contributed to the driving force for development of sensitive MRMs. Such methods may be used to quantify residues of more than 100 pesticides and their metabolites/degradation products in a wide variety of matrices including soils, air, surface and groundwaters, plant tissues, human and animal tissues and body fluids.

Although several classes of pesticides can be analyzed in many matrices with these methods, this article will address only the determination of herbicides, and exclusively in biota; that is, terrestrial and aquatic animals and plants. While it is acknowledged that body fluids (urine, saliva, blood, milk) and environmental matrices (soil, air and water) are important matrices in toxicology/human health studies and environmental fate studies, respectively, analysis of these matrices will not be discussed.

The intent of this article is to review current extraction, derivatization, clean-up and detection methods for analysis of herbicide residues in plant and animal tissues, as well as highlighting emerging technologies. Since herbicides are targeted for uptake by plants, in contrast to animals which are only indirectly exposed, the majority of information regarding the determination of herbicide residues in biota derives from plants and discussion within the article will be weighted accordingly. Discussion will include considerations for the type of matrix, sampling, known or unknown analytes, acidic or base-neutral herbicides, metabolites/conjugates, and the level of sensitivity and degree of confirmation required.

Throughout the article, herbicides will be referred to by their common names as adopted by the British Crop Protection Council.[1] Acidic herbicides are occasionally formulated as the free acid, but more commonly either as an alkali metal salt, an amine salt or as an ester. Unless reference is made to a specific salt or ester, acidic herbicides will be referred to by the common names for the parent acids.

2 GENERAL STRATEGIES

2.1 Quality Assurance/Quality Control

Sampling and analysis plans are defined by the hypotheses to be tested and the level of confidence required in drawing conclusions about the hypotheses. The statements about the level of confidence are called the data quality objectives. The quality assurance/quality control (QA/QC) program ensures that the study objectives are

met in an efficient and cost-effective manner. Quality control (QC) refers to the technical activities, such as the use of blanks, controls, and fortified blanks, and evaluation of the condition and integrity of analytical standards and stored samples which are used to control and assess the quality of the analytical data being produced. Quality assurance (QA) is the management strategy that ensures that the QC system is working properly.[2] The strategy involves defining the QC activities necessary to meet the objectives, assessing the QC results to ensure that the system is in control, and reporting the QC data along with the analytical results so that the overall quality of the results can be assessed.[2] The overall purpose of the QA/QC program is to provide analytical data of a known level of reliability that is acceptable to the end user as defined by the study objectives. The importance of QA activities in chemical analyses, in general,[3] and pesticide analyses, in particular,[4] has been reviewed.

An important QC activity in herbicide residue analysis is the control of inadvertent contamination of samples. To prevent cross-contamination between samples, it is essential that equipment and glassware, used for field sampling, subsampling and laboratory analyses, have undetectable levels of herbicide and metabolite/degradation product residues. Glassware is generally washed with hot softened water containing detergent, rinsed several times with hot water followed by several rinses with distilled/deionized water. In situations where very low residues are expected, a further rinse with an organic solvent such as acetone or hexane is recommended. Detergents should be chosen that do not introduce possible interfering compounds to the tissue extracts. Glassware used for pesticide residue analysis should be kept segregated from other glassware during the washing process. To minimize herbicide residue losses due to sorption during analysis, all surfaces wetted by the tissue extracts should be constructed of glass, stainless steel or Teflon®.

2.2 Type of Matrix

The most basic division of sample types is plant versus animal. Analytical methods can be very different for plants and animals because of differences in tissue, expected levels of contaminants, and the nature of metabolites and degradation products.

It is the edible portions of plants and animals that are most frequently sampled for herbicide residue analysis. Plant parts, such as fruiting bodies or kernels, leaves, stems, and roots are often subsampled and analyzed separately. With animals, study objectives may require that muscle tissue be sampled or fatty tissues and the liver in which persistent residues are most likely to accumulate. The choice of extraction and clean-up will be affected by the moisture level of the sample, the level of oils or fat, and the presence of co-extracted compounds and conjugates specific to plant or animal tissues.[5] The concerns and approaches to the extraction and clean-up of plant and animal tissues are discussed in subsequent sections.

2.3 Type of Herbicide

Although there are exceptions, such as paraquat and diquat, herbicides fall into two basic chemical classes: base-neutral and acidic herbicides. Acidic herbicides, which comprise a relatively large group and account for approximately one third of currently used herbicides, include benzoic, oxyphenoxyalkanoic, phenoxyalkanoic, pyridinecarboxylic, quinolinecarboxylic, and substituted alkanoic acids, phenols and some miscellaneous acidic herbicides. Salts or esters of acidic herbicides usually undergo hydrolysis in the environment, or following uptake by plants or ingestion by animals. The parent acid herbicides may form conjugates in plants and are more readily excreted by animals than the base-neutral herbicides. Base-neutral herbicides include amides, carbamates and dinitroanilines. These herbicides have relatively low water solubilites compared to acidic herbicides and are more likely to accumulate in fatty tissues. The choice of analytical methods will depend on whether determination of the parent compound is sufficient to meet the study objectives, or whether the suite of target compounds should include hydrolysis products, metabolites, and conjugates.

2.4 Multiresidue Methods

There is generally a need to analyze biota samples for residues of not only the parent herbicide but also of metabolites and/or degradation products. Thus, when a method is developed, it is generally multiresidue in nature. In general, this approach tends to be cost-effective, that is, the cost per analyte per sample is low. Other attractive features are that the MRMs are usually rugged and well suited to handle large volumes of samples with relatively short turnaround times, with acceptable levels of performance as defined by the end-use of the data. There are, however, some drawbacks to the use of MRMs that must be evaluated by the analyst and users of the analytical data. When a MRM is used, not all analytes will be extracted from a given matrix with the same efficiency, nor will they respond with the same sensitivity to a specific detector. Thus it is not unusual that a range of detection/quantification limits will result for a given method. In those situations in which analysis of a single analyte is required, rather than develop a new method, the analyst will generally choose or adapt an appropriate existing MRM.

Early MRMs were developed primarily by agrochemical companies for the purpose of providing information to support the registration of pesticide uses. These early methods were heavily focused toward the more stable and low polarity compounds, such as the organochlorine and organophosphorus pesticides, and were designed to quantify the parent compound and a small suite of toxicologically significant byproducts, metabolites, and degradation products. Similar MRMs involving analysis of a few herbicides and their metabolites or degradation products are currently used for research purposes in crop residue studies, toxicological studies, environmental fate studies, etc.

MRMs have been used by regulatory agencies to monitor a broad spectrum of several pesticides, including insecticides, herbicides, fungicides, etc.[6] However, for best results, the MRMs should be optimized for a selected class of compounds. In reality, MRMs for pesticides are usually based on physico-chemical properties of the pesticides and/or analytical approaches, and do not distinguish between herbicides and other pesticides. Thus, methods commonly used for insecticides and other pesticides can also be used for herbicides. Some of the current MRMs, however, do not include acidic herbicides in the suite of pesticides monitored. Acidic herbicides require pH adjustment of the aqueous homogenate for efficient extraction, as well as derivatization prior to determination by GC.

2.5 Unknown Analytes

The strategies adopted for the analysis of herbicide residues in biota fall into two main categories, namely cases where only analysis for one or more specific herbicides is required and authentic standards are available. The second case is where the identity of a peak in a chromatogram is unknown and authentic standards may not be available. For the latter, the results may be based on tentative identifications and follow-up analysis may be conducted for possible transformation products.

When analysis of specific target analytes is required, as in the case of residue determinations on sprayed crops, the options for sample extraction and clean-up available to the analyst can be optimized to give the highest recoveries for the chosen compounds. The instrumental analysis, usually employing a chromatographic separation, can be optimized to give maximum resolution between closely eluting compounds. Specific detectors are generally used in these situations. In recent years, MS detectors have been more commonly used, particularly as these detectors have become less expensive and easier to maintain. For a small subset of possible target analytes, the level of effort necessary to confirm that a given analyte, rather than an interfering substance, has actually been detected will be lower than in the case where it is not known which herbicides and metabolites/degradation products may be in the sample.

When it is not known which analytes are in the sample, as in the case of many environmental monitoring studies, compromises must be made to ensure that a wide range of herbicides and their metabolites and degradation products can be extracted and isolated from the matrix components. The recoveries of some compounds will not be as high as in the optimized methodology for known target analytes but will be sufficiently high to be quantifiable with a given degree of confidence. MS or atomic emission detection (AED) is desirable in this case because these detectors not only provide quantitative data but will give qualitative information which can be used for confirming the presence of an analyte in the sample.

2.6 Miniaturization of Methodology

There is growing concern about the amount of reagents, solvents, and contaminated water released to the environment as a result of analytical operations. Steinwandter[7] has estimated that such operations release a factor of 10^8 more materials to the environment than the amounts of residues being determined. The benefits of herbicide residue analytical programs – reduction of the detrimental effects of herbicide use on the environment and human health through improved understanding of the levels, distribution, fate and effects of these substances – should not be negated by the detrimental effects of laboratory chemicals released into the environment as a result of these programs. Wan and Wong[8] have argued that the issues of solvent consumption and increased environmental impact of proposed analytical methodologies should be as important as those of sensitivity and accuracy. Thus strategies which limit or reduce the consumption of reagents and solvents and the generation of contaminated wastewater are being adopted. This is due in part to safety considerations in reducing the exposure of laboratory personnel to hazardous substances and the relatively high cost associated with disposal of waste solvents. Presently, disposal costs for solvents and contaminated solutions (water plus solvents) can equal or exceed the original purchase price of the solvents.

There is a growing trend to extract smaller samples, and utilize miniaturized apparatus for sample extraction procedures. Micro methods can reduce solvent and reagent use by a factor of 10 to 100 over conventional methods. Proportionally less solvent is required for methods that use 1 to 5 g of sample instead of the 20 to 100 g used in conventional methods. As well as using less solvent, the resulting micro methods generally

tend to be cheaper, faster, less labor-intensive and more amenable to automation using on-line techniques than conventional methods. However, this approach demands higher instrumental performance and lower detection limits. Other approaches include: the use of a single extraction rather than exhaustive multiple extractions; clean-up and analysis of an aliquot of an extract rather than using the whole extract; use of combined extraction and clean-up techniques; and the elimination of the clean-up step by use of a highly selective instrumental method.[8]

2.7 Performance Criteria for Analytical Methods

Government regulatory agencies will often dictate minimum performance parameters for analytical methods to be used to determine herbicide residues in commodities and food products. The levels of residues in food are monitored for two purposes: to enforce regulations on acceptable levels, and to estimate dietary intakes. These two programs make different demands on analytical method performance, especially method detection limits.

Regulatory monitoring programs require analytical methods that can detect pesticide residues at or below a stated maximum residue limit (MRL). This limit, sometimes called a tolerance, is the legal maximum concentration of a particular pesticide, metabolite or degradation product allowed in a particular commodity or food. Although national regulatory agencies set their own MRLs, many agencies attempt to set their MRLs to the same levels as the Food and Agriculture Organization/World Health Organization (FAO/WHO)-Codex Alimentarius Commission[9] in order to facilitate international trade.[10] MRLs typically range from 0.001 to 110 mg kg^{-1} in Canada,[11] 0.1 to 50 mg kg^{-1} in the United States,[12] 0.05 to 70 mg kg^{-1} in European countries,[13] and 0.02 to 50 mg kg^{-1} in Japan.[14]

For regulatory monitoring, it is desirable that the method detection limit for a particular herbicide be at or below the MRL. Detection limits are variable and depend upon the particular herbicide, commodity, choice of sampling and subsampling methods, and choice of analytical method. Detection limits for methods used for regulatory monitoring in the United States range from 0.005 to 1 mg kg^{-1}, although 0.01 mg kg^{-1} is a typical value. Detection limits are usually well below the MRL for any particular combination of herbicide and commodity.[12] Using Belgium as being typical of European countries, the method detection limits range from 0.02% to 100% of the MRLs.[13]

Total diet studies require detection limits to be as low as practically possible to give accurate information on dietary residue intakes. Detection limits for analyses used in these studies are 5 to 10 times lower than those used for regulatory monitoring, and in general are about 0.001 mg kg^{-1}.[12] Detection limits can be lowered by a combination of larger sample size and smaller final extract volume,[13] a larger injection volume for chromatographic methods, use of higher efficiency columns for chromatographic separations or a more sensitive detector.

2.8 Sampling Biota

In brief, there are two main approaches commonly used for sampling biota, namely judgmental or intuitive sampling and statistical or probabilistic sampling. Judgmental sampling is nonstatistical and is based only on the judgment of the person(s) choosing the samples or devising some process to do so.[15,16] It can be most effective when there is prior knowledge of the sampling site based on historical data or experience of the sampling site in question. Statistical sampling is based on random sampling of a population and, in general, provides a statistical basis for making predictions and inferences about a large target population based on data from a relatively small number of samples.[15,17] Random sampling requires the largest number of samples, whereas judgmental sampling incorporates the greatest amount of bias. Often a combination of judgmental and random sampling becomes the most feasible approach in order to control costs but still obtain as much useful data as possible.

When designing protocols for sampling plants and animals in the field, many factors have to be considered including large variations within study populations, inhomogeneities in matrices and in the distribution of the analyte within a matrix, and biological processes such as metabolism and conjugation, bioaccumulation and the efficiency of uptake of individual organisms.[2] The overall sampling protocol may also be influenced by the choice of analytical method, the sampling device, the type of sample container, and possible deterioration (for example, biologically mediated reactions, thermal instability, photodegradation, etc.) of analytes of interest.

2.8.1 Sample Handling in the Field

Appropriate care must be taken in the field to avoid compromising sample integrity. QA/QC protocols must be established for sampling procedures to account for sample preservation, sample transportation and storage, collection of field blanks, field background or control samples, fortified field control samples, and for subsampling. The field QC samples should be handled in exactly the same manner as the environmental samples. QA/QC measures taken to safeguard the confidentiality and chain of custody of samples should be documented and this is mandatory where possible litigation may be an issue of concern. This requirement for good

documentation also extends to situations in which judgmental sampling is performed. For a full discussion of these factors the reader should consult reviews available in the literature.[2,18]

2.8.2 Sampling Plants

Herbicide residues in plants may result through direct application to the plants, or be present as inadvertent residues as a consequence of herbicide transport. Both situations require different considerations in planning appropriate field sampling. In the case of direct application of herbicides, the type (pre- or post-emergence) and method of application will be known and thus some idea of the uniformity of the application. In addition, the rate of application, the growth stage of the crop at application, the interval between application and sampling, and other pertinent information such as rainfall or irrigation after application will be known. Because of such information, crop sampling will most likely be a combination of both judgmental and statistical approaches.

In cases of inadvertent residues, a lower level of contamination and a less uniform distribution of the herbicide can usually be assumed compared to those resulting from agricultural applications. Although the analyte may be unknown, symptoms of phytotoxicity may identify the mode of action of the herbicide and, thus, possibly the herbicide class to which it belongs. In these situations where information is much more limited, or in situations involving litigation and sampling for legal purposes, the approach to sampling would most likely be more statistical in nature so that personal bias is minimized.

2.8.3 Sampling Animals

A fundamental difference between plants and animals for sampling purposes is that animals are mobile. Sampling protocols for animals must therefore account for factors such as movement into and out of areas of exposure, migration patterns, and feeding, spawning, or other periodic activities that could affect the level of exposure and the concentration and location of residues in the organism.[2] Because of such factors, approaches in animal sampling will most likely be more judgmental than statistical in nature.

Proper attention must be given at the study design stage to ensure that all procedures adopted are humane, particularly where the animal must be killed. Likewise due care must be given to the medical ethics pertaining to the sampling procedures to ensure that the overall program is based on sound and defensible science. Furthermore, where the collection of body fluids is necessary, the sampler must exercise due care and attention to safety procedures to avoid infectious diseases.

2.9 Extraction, Derivatization and Clean-up

2.9.1 Considerations in the Analysis of Plant Tissues

Herbicide residues, and residues arising from metabolites or degradation products, can be present as foliar deposits on plant surfaces or, following uptake by the plant, as residues within plants. Depending upon the data required, either type of residue or both could be determined. In situations where both surface deposits and residues taken up by plants are required, two samples are generally analyzed. One sample would consist of intact plants or plant parts for surface deposit determination, and the other would consist of milled/chopped tissue to determine both surface deposits plus residues taken up by plants. The difference between the total residues and surface deposits would be equivalent to the residues taken up by the plants.

Plant surface deposits can include the parent herbicide, photodegradation products, and metabolites formed by enzymatically catalyzed reactions. Residues that occur within plants as a direct result of plant uptake can include the parent herbicide and possibly metabolites or photodegradation products formed on the plant surface. Subsequent to uptake, residues can also include metabolites produced by enzymatically catalyzed reactions such as oxidation, reduction, hydrolysis and conjugation.[19] Unlike the nonsynthetic reactions of oxidation, reduction and hydrolysis, conjugation involves the reaction of the herbicide or metabolite/degradation product with endogenous plant substrates. Conjugation with plant substituents is quite common because these reactions are used by plants to deactivate or detoxify herbicides. Common conjugation reactions include the reaction of acidic herbicides with sugars to form an ester linkage,[19,20] formation of amido linkages with amino acids, and glutathione conjugation. Unlike the glutathione conjugates, the ester and amide conjugates can be cleaved by hydrolysis to release the parent herbicide.

In choosing an appropriate solvent system for the extraction of residues, one of the first considerations of the analyst in developing a method, or choosing/adapting an existing method, is whether the parent herbicide and its metabolite/degradation product(s) include acidic compounds, base-neutral compounds, or both. When acidic herbicides/metabolites/degradation products are present, an hydrolysis step can be incorporated to cleave possible conjugates. In contrast, base-neutral herbicides can be extracted from plant tissue with a number of organic solvents including acetone, hexane, diethyl ether, dichloromethane (DCM), ethyl acetate (EtOAc), acetonitrile (ACN), methanol (MeOH) and ethanol (EtOH). As the moisture content of the plant tissue increases, water miscible solvents are preferred.

Table 1 Groupings of food samples for extraction

Group	Food type	Water (%)	Fat (%)	Extraction
I	Vegetables, fresh fruits	>45	<2	Acetone
II	Whole milk, green cheese	>45	>2	Acetone
IIa	Eggs, meat	>45	>2	Acetone
III	Cheese, oil, dried legumes	<45	>2	Acetone–Water
IV	Wheat meal, pasta, rice, bread	<45	<2	Acetone–Water

Reproduced from Tekel and Hatrik, *J. Chromatogr. A*, **754**, 397–410.[21] © 1996, with permission from Elsevier Science.

Since benzene and chloroform are suspected carcinogens, these solvents are seldom used in current analytical methods. When both acidic and base neutral herbicides/metabolites/degradation products are present, a basic aqueous extraction solution is generally employed. Base-neutral analytes can be removed from the basic extract by liquid–liquid partitioning (LLP) with an appropriate organic solvent. After acidification of the aqueous extract, the acidic analytes can be isolated also by partitioning with an appropriate organic solvent.

The majority of crops used for human or livestock consumption are generally low in oil or lipid content. Thus, analytical methods to determine pesticide residues in such crops generally do not require clean-up procedures for lipid or oil removal. However, some MRMs include a clean-up step for lipid and pigment removal regardless of the matrix being analyzed. In those crops grown for their oil (canola, flax, sunflower, soybean, etc.), clean-up for lipid content is usually only required for analysis of the seed, the plant component in which the oil is most concentrated. Table 1 presents types of foods and commodities grouped by water and fat content. Although plant and animal tissues share many matrix components, plants contain significant levels of matrix co-extractives not found in animal tissues which can interfere with herbicide residue analysis unless appropriate clean-up procedures are employed. Such co-extractives include pigments, such as chlorophyll, and alkaloids.

2.9.2 Considerations in the Analysis of Animal Tissues

Animals are only indirectly exposed to herbicides, usually through contact with herbicide drift or by ingesting contaminated food and water. Herbicides generally have low to moderate toxicity to mammals and other animals because the herbicides and/or their metabolites are readily metabolized or excreted. Thus, because herbicides generally do not bioaccumulate in animals, sampling animal tissues for herbicide residue determination has not been a priority from a regulatory point of view. Herbicide residues in animal tissue used for human consumption are thus rarely looked for unless an animal (livestock) has been directly sprayed or has eaten contaminated feed. However, in ecotoxicology and environmental fate studies, herbicide residues have been determined in the tissues of several organisms (for example, birds, fish, mussels, benthic and terrestrial invertebrates), especially those which may serve as bioindicators of herbicide exposure.[22]

Extraction of animal tissues can be complicated by interferences from lipids and fatty acids (present as di- and triglycerides) as well as minor components such as sterols. Thus appropriate clean-up procedures must also be employed for these matrices. Animal tissues are generally classified according to the level of moisture and the amount of fat; such considerations affect the choice of the appropriate analytical method. Low moisture samples have less than 75% water. Low fat samples have less than 2 g of fat per 20 to 100 g of sample. Using these criteria, muscle tissue is low fat–low moisture, whereas lard, fat, blubber and adipose tissue are high fat–low moisture.

2.10 Instrumental Determination

There are several factors to consider in the selection of an instrumental method for the determination of herbicide residues in biota. The choice is usually governed by the availability of instrumentation in a laboratory, the human resources for conducting the analyses, including the level of training and expertise required for operation of an instrument, the need to meet specified detection limits and turn around times, and the overall cost of the analyses.

Earlier MRMs have relied largely on GC with selective detectors, such as the nitrogen–phosphorus detector (NPD), the Hall electrolytic conductivity detector (HECD), the flame photometric detector (FPD) and the electron capture detector (ECD). More recently, MRMs have incorporated the use of MS detectors which provide much greater specificity and essentially equivalent sensitivity. Because of the increased specificity provided by the MS detectors, larger numbers of compounds can be analysed in a given sample extract. It is not unusual for regulatory agencies to use a MRM based on automated MS detection to screen greater than 100 analytes.

3 SAMPLING

The quality of the determination of the levels of herbicides in biota is only as good as the ''weakest link in the chain''. This weakest link, in general, is often the wide variance associated with field sampling compared to the variance introduced by laboratory procedures. In the past, this fact was grossly overlooked in the scientific literature. More recently, however, there has been increasing attention given to sampling, as evidenced by several excellent reviews on sampling procedures for a wide range of environmental matrices, including biota.[2,15,17,18,23,24]

3.1 Sampling Plants

3.1.1 Sampling Treated Crops

In general, only the portions of the crop plant utilized for human or livestock consumption are sampled for residue analysis. This may include roots or tubers, bulbs, leaves and stalks, seed and fruit from trees, bushes and vines. The manner in which field sampling is carried out will be determined largely by the type of crop grown, the part of the plant to be analyzed, the objectives of doing the analyses and the size of the plot or number of plants available for sampling. For example, the number of treated trees or bushes available for sampling may be limited and thus dictate how many individual fruits are sampled from each tree or bush. Use of small field plots may require that the entire plot or complete center rows (to avoid edge effects of spraying) be used to get sufficient sample size. Use of larger plots will require the random collection of samples, each of which could be a defined length of row, a defined area, or individual plants at a defined sampling frequency. In some instances, these samples could be combined to form composite samples. Larger plots are also required when multiple samplings are necessary to monitor herbicide dissipation in the crop over time (for example, the complete growing season). Unless the study has a temporal aspect, samples are usually collected when the crop would normally be harvested for human or livestock consumption. Foliar portions of plants may be cut at the soil surface or at a given height to simulate normal harvest procedures.

3.1.2 Sampling Environmentally Contaminated Plants

The most common source of inadvertent residues in crops or other vegetation is from atmospheric deposition of herbicide drift. Contamination occurs downwind of the herbicide application and damage to nontolerant plants is often characterized by the browning/yellowing, spotting or deformation of leaves.

Since a major objective is to confirm the presence of the herbicide in the matrix of interest, sampling is generally carried out in a manner to maximize the probability of detecting the herbicide. Thus, because of possible dissipation losses of the relatively small amounts of herbicide involved, sampling should be conducted as soon as possible after deposition occurs or as soon as plant damage is observed. The extent of crop or plant damage and the fact that environmental residues decrease with distance from the application can also be determinants in how sampling is carried out. Generally, leaves, which make up the greatest surface area of plants, are randomly sampled for extraction and analysis. In the case of coniferous plants, the needles, which have been shown to be efficient collectors of atmospheric pesticide residues, are sampled for analysis.

3.1.3 Sample Handling in the Field

Once representative field samples have been collected, precautions must be taken to maintain their integrity. Control and treated samples are usually kept separate to prevent possible cross-contamination of the controls. Generally, samples are protected from sunlight to prevent photodegradation of foliar deposits and to slow further metabolism by the plant. Immediately after collection, samples are generally placed in a cool environment, usually near 0 °C, to slow possible metabolism and microbiological processes. Although sample extraction immediately after collection would be preferred, samples can seldom be accommodated in such a manner and therefore must be maintained at −20 °C or lower to minimize metabolic and microbial processes. The interval between sample collection and deep cooling is usually kept to a minimum to maintain the integrity of the sample until extraction.

Generally, matrices of high moisture content are collected into polyethylene bags or aluminum foil-lined paper bags which help to maintain the moisture content of the sample. Relatively dry matrices can be collected into paper or cotton bags, cardboard boxes, or wooden boxes/crates. These matrices are generally further dried prior to milling and such containers facilitate air-drying.

3.1.4 Representative Subsamples for Residue Analysis

In situations in which the field sample is quite large (for example, several kilograms), it is necessary to take a subsample representative of the field sample. Generally, the subsample is large enough to permit duplicate analysis, if required. In most cases, the matrix is prepared for subsampling as it normally would be used for human or livestock consumption. For example, carrots and beets would be washed to remove soil particles, oranges and cooking onions peeled, the bracts removed from strawberries, and lettuce and celery trimmed. For

naturally segmented matrices, one or more segments can be removed from each fruit or bulb to achieve a representative subsample. For non-segmented larger matrices, segments can be removed by cutting. For small matrices, individual fruits or seeds can be randomly selected until a sufficient sample size is obtained.

3.1.4.1 High Moisture Content Matrices High moisture content subsamples are often chopped with powdered dry ice in a food chopper which eliminates the separation of the sample into pulp and juice components. The sample is readily transferred into a polyethylene freezer bag for maintenance at $-20\,^{\circ}C$ until extraction. The advantage of using dry ice during chopping is that, as the dry ice sublimes from the sample, the sample becomes a porous block which can be easily crushed in order to weigh a sample for extraction. The rotating bowl and knives of the chopper are cleaned by rinsing with water after each sample has been chopped to prevent cross-contamination between subsamples.

3.1.4.2 Low Moisture Content Matrices Subsamples with relatively low moisture content are generally allowed to air dry or may be dried in a forced-air oven to facilitate milling using either a Wiley mill or a hammer mill. In the latter case, relatively low temperatures (30 to $40\,^{\circ}C$) are preferred to prevent volatilization or thermal decomposition of herbicide residues. Dry ice is often employed during the milling process to cool the mill and prevent possible thermal decomposition of herbicide residues within the matrix. The mill is generally cleaned between samples, using a blast of forced air, to prevent cross-contamination between subsamples.

3.2 Sampling Animals

The key to representative sampling of fish, birds, insects and other animals for herbicide residues is effective planning prior to conducting the field program. Detailed knowledge of variables such as the migration routes of birds, the life cycles of insects, and the behavior patterns of the animals or fish species being studied may be required. The number and frequency of sampling events to account for seasonal variations in time and space may have to be determined. Decisions regarding what constitutes a representative sample and the appropriate number of samples to meet study objectives will have to be made. Careful attention must be given to the type of sample containers, the sampling technique employed to avoid cross-contamination between samples, appropriate sizes and number of coolers, and contingency plans for timely transportation to the testing laboratory to minimize the uncertainty associated with compromising sample integrity.

For investigations requiring sampling of live specimens of either animals or human beings, sampling procedures must be humane and be as nonintrusive as possible. Exposure to herbicides is often assessed indirectly by sampling body fluids such urine, milk, blood or feces for residue analysis. Such sampling of live specimens generally requires the presence of a professional from the veterinary or medical disciplines. However, other less intrusive samples, for example, dermal patches and inhalation samplers, used for assessment of exposure of farmers during field application of herbicides, can also be collected.

3.2.1 Sample Handling in the Field

Precautions to be taken to avoid compromising sample integrity during sampling are similar to those described for sampling plants. Since judgmental sampling is frequently performed (for example, the collection of dead fish following a fish-kill accident, or sampling selected birds and/or eggs following a spill in the impacted area), it is particularly important that control samples are also collected from nonimpacted areas to establish background levels of the herbicides.

3.2.2 Representative Subsampling for Residue Analysis

Following field sample collection, representative subsamples may be taken in the laboratory for extraction and instrumental analysis. The type of subsamples taken will depend on the objectives of the study. For example, if there is a need to address human health and safety issues pertaining to consumption of meat, then muscle tissue and other edible portions of the animal would be subsampled. In contrast, for assessment of the environmental exposure of seals to herbicides, for example, the blubber which contains lipid-rich tissue and is therefore expected to be a sink for organic contaminants would be subsampled. For similar reasons, sampling of the lipid-rich tissue in marine or fresh water shrimp, mussels or clams would also be favored over analysis of the whole animal. The latter two animals are known to be fairly localized and not move over large distances within a given area and thus can serve as good integrators of contaminants over a period of time. The change in levels and distribution of contaminants within the organism with age can be a useful indicator of the change in contaminant loading with time.

4 EXTRACTION, DERIVATIZATION AND CLEAN-UP

This section describes the most common techniques employed to extract and derivatize herbicide residues, and clean-up extracts for subsequent instrumental determination. While these techniques are discussed

Table 2 Extraction and clean-up of cereal herbicide residues from plants

Herbicide	Extraction	Clean-up	Refs.
Acidic herbicides, phenoxyacids and phenols			
Acidic herbicides	NaOH 0.1 M-homogenizer	LLP-Florisil® column	29–32
Acidic herbicides	NaOH 0.1 M-homogenizer	LLP	33, 34
Acidic herbicides	EtOH–H_2O (4 : 1)-homogenizer	LLP-Florisil® column	30, 32, 35
Acidic herbicides	MeOH-homogenizer	LLP-Florisil® column	36, 37
2,4-D	Acetone:$CHCl_3$-reflux, pH = 1.5	LLP-Florisil® column	38
Phenoxyacids	Aqueous basic buffer-homogenizer	–	39
Bromoxynil [CAS 1689-84-5]	NaOH 0.1 M-homogenizer	LLP-Florisil® column	31
Urea herbicides			
Chlorotoluron [CAS 15545-48-9]	MeOH–H_2O (4 : 1)-homogenizer	Silica column	40, 41
Linuron	Acetone-homogenizer	Florisil® column	42
Metoxuron [CAS 19937-59-8]	Acetone (basic pH)	LLP-silica column	43
Phenylureas	MeOH-homogenizer	LLP-Florisil® column	44
Phenylureas	EtOH-homogenizer	LLP	45
Phenylureas	ACN-homogenizer	Florisil® : MgO : cellulose column	46
Sulfonylureas	EtOAc-homogenizer	C_{18} column	47
Sulfonylureas	SFE	–	48
Triazines			
Atrazine	MeOH-homogenizer	Acidic alumina column	49
Cyanazine [CAS 21725-46-2]	MeOH–H_2O (4 : 1)-homogenizer	LLP-alumina column	50
Metribuzin [CAS 21087-64-9]	Acetone–H_2O (3 : 1)	LLP-silica column	51
Metribuzin	ACN–H_2O (4 : 1)-reflux	LLP-Florisil® column	52
Simazine	ACN-homogenizer	LLP-alumina column	53
Simazine	H_2O, $CHCl_3$-homogenizer, shaker	Alumina column	54
Triazines	DCM-maceration	Silica column	55
Triazines	MeOH-blender	LLP-alumina column	56, 57
Dinitroanilines			
Dinitroanilines	MeOH-homogenizer	LLP-Florisil® column	58, 59
Dinitroanilines	MeOH-homogenizer	LLP	60
Trifluralin [CAS 1582-09-8]	MeOH-homogenizer	Florisil® column	61
Trifluralin	EtOH 95%-homogenizer	–	62
Chloroacetamides			
Alachlor [CAS 15972-60-8]	ACN–H_2O (4 : 1)-shaker	Florisil® column	63
Chloroacetamides	MeOH-shaker	Acidic alumina + Florisil® column	51
Chloroacetamides	ACN-homogenizer	LLP	64
Thiocarbamates			
EPTC [CAS 759-94-4]	Steam distillation	Silica column	65
Triallate [CAS 2303-17-5]	ACN : homogenizer, shaker	Alumina column	66–69
Triallate	Hexane : ACN-blender	Alumina column	70
Triallate	Steam distillation	Florisil® column	71

Adapted from Tadeo et al., *J. Chromatogr. A*, **754**, 347–365.[28] © 1996 with permission from Elsevier Science. Abbreviation: C_{18}, octadecyl; SFE, supercritical-fluid extraction.

in separate subsections as though they are distinct operations, the distinctions between extraction and clean-up have become blurred by the use of techniques in which these operations are combined and performed simultaneously.[25]

Table 2 summarizes selected methods available for the extraction and clean-up of residues of some herbicide residues from plants. For more details about the techniques discussed below, the reader should consult one of the several excellent reviews which have appeared in the literature.[21,26–28]

4.1 Extraction of Plant and Animal Tissues

The term ''extraction'' in this context refers to the transfer of herbicide residues from the solid sample matrix to the extracting liquid phase, usually an organic solvent, while leaving most of the bulk matrix behind. Some of the

operations involved include homogenization, filtration and LLP. The latter operation involves partitioning solutes between two immiscible liquid phases, usually by shaking in a separatory funnel. Solutes partition between the two phases depending on their relative solubilities in the solvents.

In many of the methods currently in use, tissue samples are first blended with a water-miscible solvent such as ACN or acetone. Such solvents have become popular because a large number of different herbicide residues can be extracted without also extracting large amounts of lipids which can cause problems later in the determination. The raw extract is then filtered to remove suspended tissue, and the aqueous filtrate partitioned against a water-immiscible organic solvent to transfer the herbicide residues to the organic phase. This approach is termed ''shake-filter extraction'' or ''liquid–liquid extraction''. A second technique, which replaces LLP with liquid–solid partitioning, is termed ''solid-phase extraction'' (SPE). A third approach is to mix and grind the homogeneous solid sample together with a granular material, place this mix into a column, and elute the herbicide residues directly with a suitable organic solvent. This approach, called ''matrix solid-phase dispersion'' (MSPD), combines homogenization and partition into one step and eliminates the need to filter a liquid homogenate.

In the following sections, these conventional extraction techniques will be discussed in more detail. SFE, a relatively new technique that shows promise as an alternative to the conventional methods for the extraction of herbicide residues from biota, is also discussed.

4.1.1 Shake-filter Methods

The simplest and most common methods of extracting herbicide residues from tissues are the shake-filter methods. They are applicable both to residues on plant surfaces and to incurred residues in plant and animal tissues.

Herbicide residue deposits on plant surfaces are removed either by rinsing the plant surface with a stream of solvent or by submersing the whole plant or plant part in a suitable solvent. The choice of solvent is based on the solubility of the herbicide or its metabolite/degradation product(s), and whether or not it is the intention to remove herbicide residues present in plant cuticular waxes. In the latter case, *n*-hexane or DCM are frequently used. If not, MeOH, EtOH or aqueous solutions of these alcohols have been used. The contact time of the plant tissue with the solvent is generally kept short (5 min or less) to minimize extraction of herbicide residues (apart from residues on the surface) taken up by the plants. Subsequent to extraction, the resulting solvent wash is subjected to any necessary clean-up or derivatization steps, and concentrated to a suitable volume prior to analysis and quantification of the residues.

The extraction of herbicide residues from plant or animal tissues, either chopped or milled, requires mechanical disruption of cell membranes using processes such as sonication, blending or homogenization in an organic solvent or aqueous-organic solvent mixture, or a dilute acidic or basic aqueous solution. Extraction recoveries, in some cases, can be further enhanced by shaking (in a sealed container), heating or refluxing. The extract is then separated from the solid matrix by centrifugation, decanting, filtration, or a combination of these. The most common approach is to homogenize the tissue sample with the extracting solvent mixture. The homogenate is then either filtered prior to LLP to isolate the residues, or centrifuged prior to the removal of the supernatant containing the isolated residues.

MRMs commonly use so-called ''universal extraction solvents'' to extract and isolate herbicide residues.[6,7,25] The sample is usually blended with one of the following solvent mixtures: water–acetone–light petroleum, water–acetone–EtOAc, or water–ACN–DCM. The amount of water added to the solvent mixture depends upon the moisture content of the sample and is adjusted to give a total volume of 100 mL water; for example, very high moisture commodities such as tomatoes need no additional water, whereas wheat kernels may have to be soaked overnight in 100 mL of water prior to extraction. Sodium chloride is usually added to the mixture to aid in phase separation. The advantages of this approach include: extraction, partitioning and removal of water in one step; and the possibility of method miniaturization. Further details can be found in the reports by Steinwandter[7,25] which review modern pesticide extraction methodologies.

Fatty tissues, such as those in fish, are often extracted by blending the tissue with a mixture of light petroleum and sodium sulfate. The extract is dried over anhydrous sodium sulfate, the solvent evaporated, and then subjected to a clean-up step, typically LLP or gel permeation chromatography (GPC; see below) to isolate the herbicide residues.

4.1.2 Solid-phase Extraction

SPE involves the partitioning of solutes between a stationary solid phase and a mobile liquid phase. Two basic forms of SPE exist: normal-phase, in which the stationary phase material, usually alumina (aluminum oxide), Florisil® (a manufactured coprecipitate of magnesium and silicon dioxides containing 16% MgO) or silica gel, interacts strongly with polar components, and eluting solvents are generally of low or medium polarity (e.g.

light petroleum, hexane or DCM); and reversed-phase, in which the stationary phase material is generally a bonded-phase silica gel material, and eluting solvents are of higher polarity (MeOH or ACN). In both forms, fractions are collected with step-wise addition of solvents of increasing solvent strength. Since most samples are homogenized with water-miscible solvents or aqueous solvent mixtures, reversed-phase is the most commonly used form of SPE.

SPE columns can be prepared in the laboratory by placing the granular solid into a column or a syringe barrel made from glass, polyethylene or polypropylene. Alternatively, manufactured SPE cartridges, containing from 100 mg to several grams of sorbent material prepacked into the cartridges, are also available. These disposable cartridges are convenient to use, have a controlled lot-to-lot variability, and can be used in automated sample preparation procedures. The most common granular materials used in SPE are bonded-phase silica gels. The silica gel is coated with a layer of organic ligands that are chemically bonded to its surface. Depending on the nature of the functional group, the selectivity of the extraction can be tailored for different applications. The most common organic functional groups used are C_{18}, octyl (C_8) and cyanopropyl (CN).

SPE can be accomplished in two ways as the filtered homogenate is passed through the sorbent bed. In the first case, the stationary phase has a greater affinity for the herbicide residues in the filtrate than the solvent in which the residues are dissolved. The residues accumulate on the surface of the granular solid while the filtrate with the majority of co-extracted matrix components pass through unadsorbed. The herbicide residues can then be eluted with a second, stronger solvent, producing a concentrated solution with a minimum of interfering matrix components.[72] In the second case, the stationary phase retains the majority of the co-extracted matrix components while the herbicide residues elute with the solvent. For example, SPE columns containing C_{18} bonded-phase silica gel are often used to retain fats and lipids while facilitating the extraction of pesticides by the eluting solvent.

SPE has been used to replace the LLP step of shake-filter extraction procedures. Casanova has reported the use of C_{18}-SPE in the Luke MRM extraction procedure for the determination of organochlorines and organophosphates in high moisture–low fat foods.[73] Hopper[74] used a different SPE approach to replace the LLP step of the Luke procedure.[75] A column containing a bed of Hydromatrix® (a specially prepared pelletized form of diatomaceous earth) is conditioned by eluting in sequence: 0.1 M KH_2PO_4 buffer, acetone and DCM. The aqueous acetone sample extract is then added to the column, where water separates from the DCM–acetone mixture and is absorbed by the solid support. The column is then eluted with DCM. These approaches save time, eliminate the emulsions typical of LLP, offer batch processing, and combine extraction and clean-up into the same operation.[72]

4.1.3 Matrix Solid-phase Dispersion

MSPD, sometimes referred to as ''column extraction'' has been used to extract pesticide residues from fruits and vegetables[76] and from animal tissues.[77,78] The sample is ground together with the solid support, the mixture packed into a column, and residues eluted with a suitable solvent. The solid supporting material can be either a relatively inert material such as anhydrous sodium sulfate or diatomaceous earth,[79] or a material chosen to give some selectivity to the partition and extraction of the herbicide residues. Solid sorbents such as alumina or Florisil®, or bonded-phase silica gels, are commonly used. In some applications, the tissue–sorbent mixture is placed into a column prepacked with another adsorbent such as Florisil® to absorb additional co-extracted material and combine extraction and clean-up in the same step. This approach combines homogenization and partition into one step, eliminating the need to filter a liquid homogenate. It also uses less solvent, is less labor intensive and faster than conventional liquid–liquid extraction methods.

4.1.4 Supercritical-fluid Extraction

SFE is one of the most promising alternatives to the conventional methods of residue extraction. It takes advantage of the unique solvating properties of supercritical fluids to reduce greatly the use of toxic organic solvents.[80]

A supercritical fluid is formed when a substance is heated and pressurized to its critical state in which the density of the liquid and gas phases are the same. This combination of critical temperature and critical pressure is referred to as the critical point. At temperatures and pressures at or above the critical point, a substance is neither a gas nor a liquid but a dense fluid with some of the properties of both. The critical point of CO_2, which is commonly used for SFE, is 31.3 °C and 7.38 MPa.

Typically, the sample is ground up with or dispersed on a medium such as diatomaceous earth or anhydrous sodium sulfate and transferred to a stainless steel extraction cell. The extraction cell is then placed in an oven and a pump delivers supercritical fluid to the cell at pressures generally ranging from 10 to 70 MPa. The fluid with the extracted analytes passes through a restrictor and depressurizes into a collection vial or other analyte trapping region. When the supercritical fluid is depressurized, the resulting gas has no solvating power,

and the extracted solutes dissolved in the supercritical fluid precipitate out and can be easily collected.

SFE has many advantages over traditional extraction techniques. Highly selective extractions can be made with control of fluid density and the addition of modifiers such as MeOH, ACN or water.[81] Supercritical fluids have lower viscosities and higher diffusivities than liquid organic solvents. The lower viscosities lead to better penetration of matrix pores and interstitial spaces and higher diffusivities lead to faster molecular diffusion. Both properties lead to faster and more efficient extractions. The extraction can be selective, automated and can be successfully used to extract thermally labile compounds. In addition, SFE can be placed on-line with GC or HPLC. When supercritical CO_2 is used, the extraction fluid is relatively non-toxic. Disadvantages include extraction efficiencies that can be matrix dependent and different recoveries obtained for fortified analytes versus aged or incurred analytes. The high pressures utilized in SFE require special safety precautions.

A few reports have reviewed the methodology and approaches to method development.[80–82] The factors controlling SFE of environmental samples, which also apply to the extraction of biological samples, have been reviewed.[83]

4.2 Derivatization of Plant and Animal Tissue Extracts

Derivatization can be incorporated into a method at different points in the procedure. The herbicide residues can be derivatized as part of the sample preparation step, or on-line after separation on an HPLC column. This latter technique is often referred to as ''post-column derivatization''. Derivatization may be performed to render analytes more suitable for GC analyses by increasing volatility and thermal stability. Derivatization may also facilitate confirmation and increase sensitivity to specific detectors.[6,84] Most current methods do not incorporate derivatization reactions in the sample preparation step prior to instrumental determination because these reactions can be time-consuming and can produce unwanted interferences due to side-reactions. In some situations, improvements in chromatographic resolution, such as the adoption of capillary columns in GC, and the use of HPLC, MS detectors and immunoassays have made derivatization reactions unnecessary.[6]

A notable exception to the trend away from pre-analysis derivatization is the use of esterification reactions to convert acidic herbicides such as phenoxyalkanoic, aryloxyalkanolic, benzoic and picolinic acids into their more volatile esters for GC analysis. Methyl esters, generated using a boron trifluoride/MeOH reagent or diazomethane, are frequently utilized. If co-extractives interfere with the methyl esters, higher molecular weight esters, such as the propyl or butyl esters, can be utilized. To enhance sensitivity to specific detectors, reagents containing 2-chloroethanol, trifluoroethanol or pentafluorobenzyl bromide are also available to form the analogous alkylated derivatives.

Post-column derivatization has proven useful in HPLC analyses of certain analytes such as *N*-methylcarbamates which cannot be detected with sufficient sensitivity by commonly used ultraviolet/visible (UV/VIS) absorbance detectors. Kok and Heimstra[85] developed a method that uses a solid-phase catalyst in a heated metal column to hydrolyze the analytes, followed by reaction with a *o*-phthalaldehyde–2-mercaptoethanol reagent to produce derivatives that can be detected by a fluorescence detector with great sensitivity.

Derivatization techniques and their applicability to pesticide residue analyses have been reviewed.[28,84,86]

4.3 Clean-up of Plant and Animal Tissue Extracts

In addition to herbicide residues, sample extracts also contain co-extracted matrix components, some of which may interfere in the analysis and quantification of the analytes. Matrix components can accumulate in GC injectors or be retained on GC or HPLC columns and decrease the resolution of chromatographic separations. For these reasons, it is generally necessary to remove co-extractives from tissue extracts using clean-up techniques. Unfortunately, there are also disadvantages to clean-up procedures. Not only do they add time and expense to the analyses, analyte losses may be incurred which can be unpredictable in some matrices. With the addition of clean-up procedures to MRMs, compromises are often made which can lead to poor recoveries for some analytes.

4.3.1 Liquid–Liquid Partitioning

LLP is commonly used as a group separation technique to fractionate on the basis of herbicide classes or to clean-up samples prior to analysis. The main disadvantages of LLP are the use of relatively large volumes of solvent, and the formation of emulsions which can be time-consuming to break. In many methods, LLP has been replaced by SPE techniques.

Light petroleum extracts of fatty tissue can be partitioned against ACN that is saturated with light petroleum.[5] The more polar pesticides partition into the ACN layer, leaving the nonpolar lipids in the light petroleum layer. The ACN is mixed with water and sodium chloride, back-extracted with light petroleum, and dried over anhydrous sodium sulfate before further clean-up.

LLP has proven to be very useful for the clean-up of basic extracts of biota tissue for determinations of

acidic herbicides. The basic extract is first extracted with an organic solvent to remove base-neutral compounds. The aqueous phase is then made acidic (pH < 2) by the addition of sulfuric or hydrochloric acid. The neutralized acids are then back-extracted into DCM or diethyl ether. The solvent layer may be washed with distilled water prior to drying over anhydrous sodium sulfate before concentration and derivatization.

4.3.2 Gel Permeation Chromatography

GPC, also known as size exclusion chromatography, separates components on the basis of molecular size.[87] The GPC resin, which contains a narrow distribution of pore sizes, is slurry-packed into a column using the eluting solvent. Prior to packing, the GPC resin is allowed to swell in the eluting solvent. The sample extract is dissolved in the elution solvent and is added to the top of the column. Large molecules such as lipids, which have molecular weights in the range 600 to 1500 Daltons (atomic mass units), are not able to enter the pores of the resin so they move through the column with the elution solvent and are eluted first. The smaller analyte molecules,which are typically 200 to 400 Daltons, can enter the pores and thus elute more slowly than the larger molecules. The flow of the eluate from the column is diverted to waste while the lipids and other co-extractives are eluting from the column, after which the smaller analytes are collected. A single column can be used for many samples before its performance deteriorates. The performance and calibration of the column is checked using corn oil fortified with pesticides. GPC clean-up of samples is usually performed in batches of 20 or more using an automated system, such as the Autoprep (OI Analytical).

The original GPC method developed by Stalling et al.[87] used 2.5-cm internal diameter columns. Reports by Patterson[88] and van Rijn and Tuinstra[89] have described the use of 1-cm and 2-mm internal diameter columns, respectively. The main benefit of these modifications was to reduce solvent consumption. Other advantages include a reduction in the amount of resin required, and smaller elution volumes. One disadvantage of the modifications is a decrease in the amount of fat that can be retained by the column. However, with the trend to smaller sample sizes, this should not be an impediment to the use of smaller columns.

4.3.3 Solid-phase Extraction

The principles of SPE for the clean-up of extracts are essentially the same as described above under extraction techniques. One of the most frequent uses of SPE is the removal of lipids from extracts of biota using a column of Florisil®.[90] Silica gel columns are often used to further clean-up fractions collected from Florisil® columns.[91]

5 INSTRUMENTAL DETERMINATION

Following the extraction, derivatization and clean-up steps, common instrumental analyses are used for the determination of herbicides and their degradation products in extracts of both animal and plant tissues. The most common separation techniques used are GC and HPLC.[86,92] In recent years, there has been advances in CE,[93–95] a relatively new technique, providing many advantages over traditional gas and liquid chromatographic separations. Selected examples of instrumental techniques will not be repeated here but are summarized in Table 3. The current most widely used instrumental technique for the determination of herbicide residues is MS coupled to GC or HPLC. This technique provides sensitive multichannel detection of components, with reliable identification and quantification of herbicide residues. These features are particularly attractive for investigations of herbicide residues in complex mixtures containing co-extracted matrix compounds. The MS methods, once optimized, can facilitate high sample throughput at competitively low cost compared to conventional approaches.

5.1 Gas Chromatographic Separations

GC is a technique for separating volatile and semi-volatile components of a mixture based on differences in distribution or partitioning of substances between

Table 3 Summary of commonly used instrumental techniques for determination of herbicides

Physical/chemical properties	Herbicide class	Instrumentation
Volatile and thermally stable	Halogens Chlorophenoxy-alkanoic acids Chlorophenols Chlorobenzoic acids	GC with ECD, HECD or MS
	Sulfur Thiocarbamates	GC with FPD in S mode, HECD, or MS
	Nitrogen Dinitroanilines Triazines	GC with NPD, HECD or MS
Polar and thermally labile	Sulfonylureas Ureas Bipyridylium Phosphonic acids Carbamates	HPLC with UV, fluorescence or MS detector

a stationary liquid phase and an mobile gas phase. The separations take place within a column, usually a fused-silica open tubular (FSOT) column, 0.1 to 0.5 mm internal diameter and 10 to 60 m long. As the separated components elute from the column, they are detected using a variety of detectors to produce a chromatogram. Current gas chromatographs are microprocessor-controlled and are equipped with desktop microcomputer data handling systems. The advantages of GC include the ability to resolve several hundred components in a mixture, excellent quantitative performance, high sensitivity, and ease of use. However, it is limited to compounds that are thermally stable, non-reactive and volatile at typical operating temperatures of 50 to 350 °C. Fortunately, most herbicides or their derivatives are amenable to GC separations. There are several detectors which can be utilized for the analysis of herbicides residues in biota. Most give reproducible linear responses over wide dynamic ranges, high sensitivity, are stable and rugged, and have low noise characteristics.

5.1.1 Mass Spectrometric Detectors

MS is generally the detector of choice for selectivity and sensitivity of detection of analytes in extracts of plant and animal tissues. This is largely due to the fact that these detectors can be operated such that the response is either universal (positive response for all compound classes) or selective (very sensitive and specific responses to particular elements or chemical classes). The MS can thus provide information for identification of unknown components. The combination of the separating power of the GC with the identification power of the MS detector and the data handling capabilities of modern desktop computers has produced a powerful instrumental technique for residue analysis.[96] Key characteristics of the MS detector are discussed below.

The differences between MS instruments are mainly in size and functionality, and in the type of mass filter used to separate ions according to mass. Each type has advantages and disadvantages in performance, utility, and cost. MS instruments can be classified as benchtop instruments, or larger, more traditional instruments. Benchtop MS instrumentation has been reviewed by Wach.[97] Generally, benchtop instruments are best suited for routine well-defined applications while larger instruments are better suited for research applications, although the distinction is becoming blurred. Benchtop instruments are usually limited to ion masses less than 1000 and have smaller vacuum pumps which limit the choice of sample inlets and interfaces that can be used. The larger instruments have higher mass range capabilities, more inlet and interface options, and generally higher mass resolution. Benchtop instruments are generally less expensive than their larger counterparts.

Benchtop MS instruments are relatively inexpensive, rugged, easy to use, and employ technology based on either the conventional quadrupole mass filter or the relatively new quadrupole ion-trap. A good example of the former is the quadrupole mass selective detector (MSD) manufactured by Hewlett-Packard. This compact instrument is equipped with an electron impact (EI) source and is optimized for use with FSOT-column gas chromatographs. It has proven very useful for analyses of a variety of organic compounds, including herbicides. The quadropole ion-trap employs a relatively new technique in which the ionization, mass selection, and detection occur within the same space, but are separated in time. The principles and practice of ion-traps have been reviewed by McLuckey et al.[98] Examples of ion-trap instruments are the Varian Saturn and the Finnigan GCQ. Other instruments use either the conventional quadrupole mass filter, magnetic sectors or time-of-flight mass selection.

The mass spectrometer can be operated in two main modes of detection: full scan mode or selected-ion monitoring mode. In the full scan mode, the MS repeatedly scans from one end of the mass range to the other as components elute from the chromatographic column, creating a collection of mass spectra. The mass spectrum from a single component peak can be used to identify or confirm the component by comparison of the mass spectrum of the unknown with that of an authentic standard. In this mode, the linear range of the MS is roughly comparable to that of the flame ionization detector (FID), 100 pg to 100 µg. For quantitative determinations, the mass spectrometer is usually operated in the selected-ion monitoring mode. In this mode the mass spectrometer is tuned to detect only a few ions characteristic of the target herbicides, metabolites, or degradation products. This can result in an increase in sensitivity by a factor of 10 to 1000 compared to the full scan mode. In favorable cases, the relative abundances of the ions can be matched with the expected values observed for the authentic standards for both sensitive detection and confirmation in a single instrumental analysis. In the selected-ion monitoring mode, the MS detector can be as sensitive as specific detectors for halogenated compounds, as discussed below.

Tandem mass spectrometry (MS/MS),[98] is a powerful technique for characterizing unknown components in a mixture and providing sensitive, selective detection of trace levels of target compounds.[99] In conventional quadrupole and sector instruments, the first mass filter selects an ion, passes it to a collision cell where further fragmentation takes place, and a second mass filter scans or selects the fragment ions.[100] In the ion-trap, the processes take place in the same volume but are separated in time. The technique provides an added

dimension of information which can improve selectivity, reduce chemical noise, and increase sensitivity. For example, MS/MS has been used to determine residues of diclofop-methyl and triallate in the lipid-rich tissues of amphipods without prior clean-up or GC separation.[101] The high cost and complexity of traditional MS/MS instruments employing quadrupole mass filters and/or magnetic sectors have hampered the widespread adoption of this powerful technique. However, the relatively low cost, ease of use, and sensitivity of ion-trap instruments is expected to result in a wider acceptance of MS/MS techniques for the analysis of organic residues.

5.1.2 Other Specific Detectors

In addition to the MS, the NPD and the ECD are the most frequently used specific detectors for herbicide determinations. The operation of these detectors has been described elsewhere[102] and will not be repeated here. Instead, emphasis will be given to examples of herbicide classes for which the detectors are best suited, a summary of which is given in Table 3.

Except for the MS detector, specific detectors have one common disadvantage: they provide insufficient qualitative information for identification of components. Component identification using such specific detectors is based only on a characteristic elution time using a specified column under defined conditions. While acceptable in well-defined situations, such as routine determinations of known compounds in crops or food commodities, the lack of qualitative information is a disadvantage in screening of samples for unknown analytes.

The NPD, sometimes referred to as the thermionic emission detector, is a variant of the FID. It offers improved selectivity for nitrogen- and phosphorus-containing compounds over hydrocarbons. Compared to the FID, the NPD in general has 1 to 2 orders of magnitude greater response for nitrogen-containing compounds and 2 to 3 orders of magnitude greater response for phosphorus-containing compounds. In general, the linear range of the NPD is about 0.1 pg to 1 ng of analyte.

The primary advantages of the ECD is its very high sensitivity (0.1 pg) and high selectivity for herbicides containing halogen atoms (F, Cl, Br, I). Many compounds which do not contain halogens can also be detected by the ECD after derivatization with reagents that add halogen-containing functional groups. The ECD responds with lower sensitivity to oxygen-, sulfur-, and nitrogen-containing compounds as compared to halogen-containing compounds.

Other less commonly used specific detectors include the HECD and the FPD. The HECD can be operated in halogen-, nitrogen- and sulfur-specific modes of operation. However, it has a somewhat limited linear range of about 1 ng to 100 ng. The FPD can also be used for the selective detection of organic compounds containing sulfur atoms as well as phosphorus. In the phosphorus mode, it is linear from about 10 pg to 100 ng. In the sulfur mode, the response is proportional to the square of the quantity of sulfur with a quantification range of about 200 pg to 100 ng.

5.1.3 Atomic Emission Detector

The AED manufactured by Hewlett-Packard is a relatively new instrument with capabilities that complement the MS for determination of herbicides and other organic compounds in a variety of matrices. The GC column effluent is mixed with helium and a small amount of oxygen and introduced into a microwave-induced plasma source operating at 1700 °C. At this temperature, the organic compounds are broken down into their individual elements which are thermally excited and emit light at characteristic wavelengths. The light is dispersed by a monochromator and detected by a photodiode array. Multiple elements can be quantified in one injection, giving information that can be used to determine the empirical formula of each compound. The instrument is also able to distinguish isotopes of some elements, such as deuterium, carbon-13 and nitrogen-15. Combined with MS data, the data from the AED can be used to identify unknown components with great reliability. The detector can provide picogram level detection of elements with a linear response that is largely independent of the molecular structure of the analyte. It is generally more sensitive than the MSD in full-scan mode, five-times more sensitive than FID for carbon and ten times more sensitive than the FPD for sulfur. The utility of the AED for the analysis of pesticide residues has been described.[103,104]

5.2 High-performance Liquid Chromatographic Separations

HPLC is a separation technique complementary to GC in that a large portion of the known organic compounds (about 90%), that for reasons of low volatility, high polarity and water solubility, or thermal instability are not amenable to GC analysis, can be separated by some form of HPLC technique. The separation is based on partition between a mobile liquid phase and an immiscible stationary phase. HPLC columns are typically stainless steel cylinders 1 to 4 mm internal diameter, 10 to 30 cm long, and packed with 1 to 10 μm diameter beads of silica or other material that functions as the stationary phase. The most common mode of operation in HPLC is reversed-phase.

The most commonly used detectors for residue analyses are the UV/VIS absorption detector and the fluorescence

Table 4 Principles, operation and selected applications of commonly used HPLC detectors

Specific detectors	Principle of operation	Selected applications
UV/VIS	Specific absorption of the compound at a specific wavelength & low or zero absorption by mobile phase. Operation using fixed single-wavelength, variable wavelength, scanning wavelength, or diode-array	Sulfonylurea herbicides in complex matrices,[48] fluridone residues in meat, milk, eggs and crops[105] and bromoxynil, ioxynil and chlorophenoxyalkanolic acid herbicides in tissue homogenate[106]
Fluorescence	Excitation of the compound at a specific operating wavelength and low or zero absorbance by mobile phase. Operation using interference filters or grating monochromators	Following post-column derivatization, *N*-methylcarbamate pesticides are commonly determined by HPLC with fluorescence detection[85]
MS with thermospray	Ions expelled from heated droplets in ionization chamber under partial vacuum; flow rates 1 to 4 mL min^{-1}	Sulfonylurea herbicides in wheat seed[107]
MS with electrospray	High electric field causes column eluent to disperse into a fine spray; free ions directed into MS; flow rates 5 to 30 μL min^{-1}	Imazathapyr [CAS 8135577-5] in crops[108] diquat and paraquat[109]

detector although, in recent years, there is a growing trend of the use of MS detectors. A summary of the principle of operation and examples of selected applications of these popular detectors is given in Table 4.

The UV/VIS detector is based on the absorption of ultraviolet and/or visible light by components of the sample as they pass through the flow-cell. Sensitive detection of components depends on high specific absorption of the compound at the operating wavelength and very low or zero absorption by the mobile phase. In contrast, the fluorescence detector gives excellent sensitivity and selectivity for target compound analyses of substances or derivatives which strongly fluoresce. Two other selective detectors which have been used for herbicide analysis include the radiochemical detector (for radiolabeled analytes) and the photoconductivity detector.[110]

In recent years, technology has developed such that interfaces and ionization techniques are now starting to meet the quite disparate requirements of the HPLC and the mass spectrometer. For example, the thermospray interface (TSP) is a popular, versatile, reliable and rugged interface suitable for routine use. Another popular MS source is the electrospray ionization (ESI) source. This interface is well suited to capillary columns and can produce very reproducible spectra of large or thermally labile compounds. The key concepts and techniques for MS detectors have been reviewed by Garcia and Barcelo.[111]

5.3 Capillary Electrophoretic Separations

CE is a relatively new technique that provides high-speed separations of herbicide residues in extracts of environmental and biological matrices. Separations are carried out in fused silica capillaries, 25 to 75 μm internal diameter, about 50 cm long and filled with a buffer solution. A potential difference of 20 to 40 kV is applied between the ends of the capillary. The high electric field creates an electroosmotic flow of the buffer from the anode (positive) to the cathode (negative). Ionized organic molecules also migrate under the influence of the electric field and are separated according to differences in their electrophoretic mobility. Very specific separations of analytes can be accomplished by adjusting the pH, careful control of the nature of the buffer, the addition of organic solvents such as ACN, and the nature and concentration of surfactants. A variety of detectors are available for CE analyses, including UV/VIS, fluorescence, amperometric, conductivity, and MS detectors.

CE provides many advantages over traditional HPLC and GC separations. Peak widths are very narrow and resolution is high, resulting in very short analysis times. Very small injection volumes of less than 100 nL can be utilized. The operation of the instrument can be easily automated. CE separations have been demonstrated for the triazine herbicides prometryn [CAS 7287-19-6], terbutryn [CAS 886-50-0], desmetryn [CAS 1014-69-3], simazine [CAS 122-34-9] and atrazine [CAS 1912-24-9],[112] the cationic herbicides paraquat [CAS 4685-14-7] and diquat [CAS 2764-72-9],[93,95] and glyphosate [CAS 1071-83-6].[95]

5.4 Immunochemical Methods

Immunoassay methods are based on the ability of antibodies to ''recognize'' and attach themselves to molecules with very great specificity. These methods have several advantages over conventional chromatographic methods. Analyses can be performed using crude extracts of tissues, reducing clean-up costs and preparation time. Once optimized, they can facilitate high sample throughput at relatively low cost compared to conventional approaches for screening of herbicide residues. Furthermore, the immunoassay kits are easily field portable.

The more popular methods are: the indirect enzyme-linked immunosorbent assay (indirect ELISA), direct ELISA, and radioimmunoassay (RIA). The details of these methods can be found in various reports.[113–115] Immunoassay systems have been developed for such herbicides as atrazine, terbutryn, chlorsulfuron [CAS 64902-72-3], 2,4-D [CAS 94-75-7], 2,4,5-T [CAS 93-76-5], diuron [CAS 330-54-1], linuron [CAS 330-55-2], monolinuron [CAS 1746-81-2], diclofop-methyl [CAS 51338-27-3], paraquat, and picloram [CAS 1918-02-1] in water samples.[113,114] Detection limits are typically 0.08 to 5 $\mu g\,L^{-1}$, and the methods are best used as a screening tool. While there has been much work done on detection of herbicide residues in soils and water,[114] little work has been done for the direct determination of herbicides in biota. Methods have been described for the analysis of paraquat residues in milk, beef and potatoes[116] and imazamethabenz [CAS 100728-84-5] in cereal grain.[117]

6 OFFICIAL MULTIRESIDUE METHODS FOR PESTICIDE RESIDUE ANALYSIS

Over the last 40 years a number of MRMs have been adopted by various government regulatory agencies to screen, quantify and confirm pesticide residues, including herbicides in foods and commodities.[26] Since little is known, prior to analysis, about the nature and levels of possible contamination, MRMs have to be capable of determining a broad spectrum of analytes. The MRMs in common use are rugged, well-suited to handle large numbers of samples with relatively short turnaround times, and cost-effective, that is, the cost per analyte per sample is low. The most widely used official MRMs are based on the following methods: Mills,[118] Mills, Onley, Gaither,[119] Storherr,[120] Luke,[75] and Krause.[121] United States agencies, such as the Food and Drug Administration (FDA) and the Department of Agriculture, use methods validated by the Association of Official Analytical Chemists (AOAC).[91] European agencies commonly use the methods of Specht.[122,123] Many of the MRMs in use have been reviewed by Motohashi et al.[27]

The methods used by the US FDA Total Diet Program typify the MRMs in use for screening foods and commodities for pesticide residues.[124] One procedure, AOAC method 970.52, is used to screen for organochlorine and organophosphorus pesticide residues.[91] Samples of non-fatty foods are extracted with ACN or with aqueous ACN depending on water and sugar content. Fatty tissues are first extracted with diethyl ether then partitioned between light petroleum and ACN. The ACN phase containing the extracted residues is diluted with water and partitioned into light petroleum. The partitioned extracts are cleaned-up on a Florisil® column. Additional clean-up, when necessary, is performed using a column of MgO (magnesia) and Celite™. The residues are then determined by GC with various specific detectors. This ACN extraction-based method is based in part on methods developed by Mills and co-workers.[118,119] The California Department of Food and Agriculture improved this method by using SPE columns to eliminate the use of DCM and reduce the number of labor-intensive LLP steps.[125]

An alternative to the ACN-based method is AOAC method 885.22,[91] which is based on the acetone extraction method of Luke.[75] Nonfatty samples are blended with acetone and filtered. The filtrate is partitioned with light petroleum and DCM. The DCM is evaporated with additions of light petroleum followed by acetone to remove DCM and leave any pesticide residues dissolved in acetone. Pesticide residues in this acetone solution can be determined by GC without further clean-up. The absence of a chromatographic clean-up step permits the determination of many of the polar residues which are not recovered from Florisil®, or charcoal–MgO–Celite™ clean-up columns.

The methods of Specht[122,123] are the basis of the MRM DFG S19 adopted by Germany.[27] Samples are extracted with aqueous acetone, the amount of water dependent upon the amount of moisture in the samples. After the addition of NaCl and partitioning with DCM, the extract is exchanged into cyclohexane–EtOAc and cleaned up using GPC. The fraction containing the residues is exchanged into EtOAc. Phosphorus-, nitrogen- and sulfur-containing residues are determined by GC with specific detectors. Organochlorine pesticide residues are determined by GC after further clean-up on a silica gel column.

The MRMs described above do not recover acidic herbicides such as chlorophenoxy acids and their chlorophenol transformation products. Such herbicide residues can be determined by a method described by Hopper[126] and incorporated into the US FDA Pesticide Analytical Manual I as method 402.[127] Acidic and phenolic residues are extracted with acidic DCM, MeOH or aqueous MeOH, depending on the commodity. Extracts are concentrated and residues exchanged into hexane prior to clean-up by GPC. Acids and phenols in the cleaned up extract are methylated using methyl iodide in the presence of tetrabutylammonium hydroxide, and further cleaned up using a miniature Florisil® column. Methyl esters and ether derivatives of the acidic and phenolic residues are determined by GC with ECD or HECD operated in halogen or nitrogen modes.

A useful source of information on MRMs for determination of herbicide residues in foods and commodities is the US FDA Pesticide Analytical Manual, Part I, which is

available in its entirety in electronic format via the Internet from the US FDA Center for Food Safety and Applied Nutrition at the following site: http://vm.cfsan.fda.gov/.

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
AED	Atomic Emission Detection
AOAC	Association of Official Analytical Chemists
C_8	Octyl
C_{18}	Octadecyl
CE	Capillary Electrophoresis
CN	Cyanopropyl
DCM	Dichloromethane
ECD	Electron Capture Detector
EI	Electron Impact
ELISA	Enzyme-linked Immunosorbent Assay
ESI	Electrospray Ionization
EtOAc	Ethyl Acetate
EtOH	Ethanol
FAO/WHO	Food and Agriculture Organization/ World Health Organization
FDA	Food and Drug Administration
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
FSOT	Fused-silica Open Tubular
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HECD	Hall Electrolytic Conductivity Detector
HPLC	High-performance Liquid Chromatography
LLP	Liquid–Liquid Partitioning
MeOH	Methanol
MRL	Maximum Residue Limit
MRM	Multiresidue Method
MS	Mass Spectrometry
MSD	Mass Selective Detector
MS/MS	Tandem Mass Spectrometry
MSPD	Matrix Solid-phase Dispersion
NPD	Nitrogen–Phosphorus Detector
QA	Quality Assurance
QA/QC	Quality Assurance/Quality Control
QC	Quality Control
RIA	Radioimmunoassay
SFE	Supercritical-fluid Extraction
SPE	Solid-phase Extraction
TSP	Thermospray Interface
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Pesticides **(Volume 7)**
Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

REFERENCES

1. Anonymous, *The Pesticide Manual – A World Compendium*, 11th edition, ed. C.D.S. Tomlin, The British Crop Protection Council, Farnham, UK, 1997.
2. L.H. Keith, *Environmental Sampling and Analysis: A Practical Guide*, Lewis Publishers, Boca Raton, FL, 1991.
3. R.J. Mesley, W.D. Pocklington, R.F. Walker, 'Analytical Quality Assurance: A Review', *Analyst*, **116**, 97–99 (1991).
4. H.B.S. Conacher, 'Importance of Analytical Assurance in Canadian Pesticide Analysis', *J. Assoc. Off. Anal. Chem.*, **70**, 941–943 (1987).
5. S.M. Walters, 'Clean-up of Samples', in *Analytical Methods for Pesticides and Plant Growth Regulators*, Vol. XV, *Principles, Statistics and Applications*, eds. G. Zweig, J. Sherma, Academic Press, Orlando, FL, 67–110, 1986.
6. J.N. Seiber, 'New Trends in Analytical Methods for Pesticide Residues in Foods', in *Pesticide Residues and Food Safety: A Harvest of Viewpoints*, eds. B.G. Tweedy, H.T. Dishburger, L.G. Ballantine, J. McCarthy, J. Murphy, ACS Symp. Ser. 446, American Chemical Society, Washington, DC, 1991.
7. H. Steinwandter, 'Development of Microextraction Methods in Residue Analysis', in *Emerging Strategies for Pesticide Analysis*, eds. T. Cairns, J. Sherma, CRC Press, Boca Raton, FL, 3–38, 1992.
8. H.B. Wan, M.K. Wong, 'Minimization of Solvent Consumption in Pesticide Residue Analysis', *J. Chromatogr. A*, **754**, 43–47 (1996).
9. Joint FAO/WHO Food Standards Program, 'Status of Codex Maximum Residue Limits for Pesticides in Food and Animal Feed', Codex Committee on Pesticide Residues, 25th Session, Havana, Cuba, 1993.
10. C.L. Trichilo, R.D. Schmitt, 'Tolerance Setting Process in the US Environmental Protection Agency', *J. Assoc. Off. Anal. Chem.*, **72**, 536–538 (1989).
11. Canada Food and Drug Regulations, 'Maximum Residue Limits for Agricultural Chemicals', Division 15, Table II, 1996.

12. Anonymous, 'US Food and Drug Administration Pesticide Program Residue Monitoring 1993', *J. AOAC Int.*, **77**, 161A–185A (1994).
13. W. Dejonckheere, W. Steurbaut, S. Drieghe, R. Verstraeten, H. Braeckman, 'Monitoring of Pesticide Residues in Fresh Vegetables, Fruits and Other Selected Food Items in Belgium, 1991–1993', *J. AOAC Int.*, **79**, 97–110 (1996).
14. 'Japan Pesticide Information No. 54', Third Update, Ministry of Agriculture, Forestry and Fisheries, Tokyo, Japan, 1996.
15. J.A. Graham, 'Monitoring Groundwater and Well Water for Crop Protection Chemicals', *Anal. Chem.*, **63**, 613A–614A, 616A, 618A–620A (1991).
16. J.V. Headley, 'Sampling in the Field', Proceedings of CIC Symposium, 'Introduction to GC/MS Environmental Analysis – Uses, Limitations and Interpretation', Calgary, AB, May 10–11, 1990.
17. S.Z. Cohen, C. Eiden, M.N. Lorber, 'Monitoring Ground Water for Pesticides', in *Evaluation of Pesticides in Ground Water*, ACS Symp. Ser. 315, American Chemical Society, Washington, DC, 1986.
18. L.H. Keith (ed.), *Principles of Environmental Sampling*, American Chemical Society, Washington, DC, 1988.
19. K.K. Hatzios, 'Biotransformations of Herbicides in Higher Plants', in *Environmental Chemistry of Herbicides*, eds. R. Grover, A.J. Cessna, CRC Press, Boca Raton, FL, 141–185, Vol. 2, 1991.
20. A.J. Cessna, N.D. Westcott, 'Fate of Pesticides Applied to Cereals Under Field Conditions', in *Pesticide Interactions in Crop Production: Beneficial and Deleterious Effects*, ed., J. Altman, CRC Press, Boca Raton, FL, 59–85, 1993.
21. J. Tekel, Š. Hatrík, 'Pesticides Residue Analyses in Plant Material by Chromatographic Methods: Clean-up Procedures and Selective Detectors', *J. Chromatogr. A*, **754**, 397–410 (1996).
22. M.T. Arts, J.V. Headley, K.M. Peru, 'Persistence of Herbicide Residues in *Grammarus lacustris* (Crustacea: Amphipoda) in Prairie Wetlands', *Environ. Toxicol. Chem.*, **15**, 481–488 (1996).
23. R. Grover, 'Nature, Transport, and Fate of Airborne Residues', in *Environmental Chemistry of Herbicides*, eds. R. Grover, A.J. Cessna, CRC Press, Boca Raton, FL, Vol. 2, 89–117, 1988.
24. W.J. Fielding, F.B. Lauckner, 'Biometry Training in Developing Countries: Caribbean Experiences', *The Statistician*, **41**, 105–111 (1992).
25. H. Steinwandter, 'Universal Extraction and Clean-up Methods', in *Analytical Methods for Pesticides and Plant Growth Regulators*, Vol. XVII, *Advanced Analytical Techniques*, ed. J. Sherma, Academic Press, New York, 35–73, 1989.
26. C.M. Torres, Y. Picó, J. Mañes, 'Determination of Pesticide Residues in Fruit and Vegetables', *J. Chromatogr. A*, **754**, 301–331 (1996).
27. N. Motohashi, H. Nagashima, C. Párkányi, B. Subrahmanyam, G. Zhang, 'Official Multiresidue Methods of Pesticide Analysis in Vegetables, Fruits and Soils', *J. Chromatogr. A*, **754**, 333–346 (1996).
28. J.L. Tadeo, C. Sanchez-Brunete, A.I. Garciá-Valcarcel, L. Martínez, R.A. Pérez, 'Determination of Cereal Herbicide Residues in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **754**, 347–365 (1996).
29. A.E. Smith, 'Gas Chromatographic Method for Analysis of 2,4-D in Wheat: Interlaboratory Study', *J. Assoc. Off. Anal. Chem.*, **67**, 794–798 (1984).
30. A.J. Cessna, 'The Determination of Residues of 2,4-D in Postemergence-treated Triticale', *Pestic. Sci.*, **30**, 141–147 (1990).
31. A.J. Cessna, 'Simultaneous Extraction and Detection of Residues of (2,4-dichlorophenoxy)acetic Acid and Bromoxynil from Wheat', *J. Agric. Food Chem.*, **28**, 1229–1239 (1980).
32. A.J. Cessna, 'Comparison of Extraction/Hydrolysis Procedures for the Determination of Acidic Herbicides in Plants: Residues of Mecoprop in Barley Following Post-emergence Application', *J. Agric. Food Chem.*, **40**, 1154–1157 (1992).
33. C. Sanchez-Brunete, A.I. Garcia-Valcarcel, J.L. Tadeo, 'Determination of Residues of Phenoxy Acid Herbicides in Soil and Cereals by Gas-chromatography–Ion Trap Detection', *J. Chromatogr. A*, **675**, 213–218 (1994).
34. C. Sanchez-Brunete, A.I. Garcia-Valcarcel, J.L. Tadeo, 'Determination of Bromoxynil and Ioxynil Residues in Cereals and Soils by GC/ITD', *Chromatographia*, **38**, 624–628 (1994).
35. A.J. Cessna, 'Gas Chromatographic Determination of Bromoxynil in Onions', *J. Agric. Food Chem.*, **38**, 1844–1847 (1990).
36. T. Tsukioka, T. Murakami, 'Capillary Gas Chromatographic–Mass Spectrometric Determination of Acid Herbicides in Soils and Sediments', *J. Chromatogr.*, **469**, 351–359 (1989).
37. J.F. Lawrence, L.G. Panopio, H.A. McLeod, 'Comparison of Liquid and Gas Chromatography for the Determination of Bromoxynil Octanoate and Benzoylprop Ethyl in Wheat Products', *J. Agric. Food Chem.*, **28**, 1019–1022 (1980).
38. G.R. Leather, L.E. Forrence, 'Residues of 4-(2,4-dichlorophenoxy)butyric Acid in Winter Wheat', *J. Environ. Qual.*, **11**, 345–347 (1982).
39. H. Ruckendorfer, W. Lindners, 'Trace Analysis of 2,4,5-TP and Other Acidic Herbicides in Wheat Using Multicolumn-HPLC', *Int. J. Environ. Anal. Chem.*, **18**, 87–99 (1984).
40. J.L. Tadeo, J.M. Garcia-Baudin, T. Matienzo, S. Perez, H. Sixto, 'Chlortoluron Levels and Selectivity in Wheat Cultivars', *Chemosphere*, **18**, 1673–1679 (1989).

41. S. Perez, M.T. Matienzo, J.L. Tadeo, 'GC/ITD Analysis of Chlortoluron and its Metabolites in Cereals', *Chromatographia*, **36**, 195–200 (1993).
42. J.F. Lawrence, 'A Comparison of Electron Capture GLC, Electrolytic-conductivity GLC and UV-absorption HPLC for the Analysis of Some Herbicides in Foods', *J. Chromatogr. Sci.*, **14**, 557–559 (1976).
43. J. Lantos, U.A.Th. Brinkman, R.W. Frei, 'Residue Analysis of Methoxuron and Its Breakdown Product (3-chloro-4-methoxyaniline) by Thin-layer and High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **292**, 117–127 (1984).
44. R.G. Luchtefeld, 'Multiresidue Method for Determining Substituted Urea Herbicides in Foods by Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **70**, 740–745 (1987).
45. J.F. Lawrence, G.W. Laver, 'Analysis of Some Carbamate and Urea Herbicides in Foods by Gas–Liquid Chromatography after Alkylation', *J. Agric. Food Chem.*, **23**, 325–329 (1975).
46. J.H. Onley, G. Yip, 'Analysis of a Single Crop Extract for Substituted Urea Herbicides and Metabolites, Chlorinated Insecticides and Amitrole', *J. Assoc. Off. Anal. Chem.*, **52**, 526–532 (1969).
47. E.G. Cotterill, 'Determination of the Sulfonylurea Herbicides Chlorsulfuron and Methsulfuron Methyl in Soil, Water and Plant Materials by Gas Chromatography of their Pentafluorobenzyl Derivatives', *Pestic. Sci.*, **34**, 291–296 (1992).
48. M.E.P. McNalley, J.R. Wheeler, 'Supercritical Fluid Extraction Coupled with Supercritical Fluid Chromatography for the Separation of Sulfonylurea Herbicides and their Metabolites from Complex Matrices', *J. Chromatogr.*, **435**, 63–71 (1988).
49. C. Marucchini, L. Scarponi, P. Perucci, 'Effect of Alachlor, Metolachlor, Atrazine and Simazine Residues on Some Enzyme-activities of Maize Tissues', *Agrochimica*, **32**, 536–540 (1988).
50. K.J. Beynon, 'The Analysis of Crops and Soils for the Triazine Herbicides Cyanazine and Some of its Degradation Products 1. Development of Method', *Pestic. Sci.*, **3**, 389–400 (1972).
51. J.M. Garcia-Baudin, M. Villarroya, M.C. Chueca, J.L. Tadeo, 'Different Tolerance of Two Cultivars to Triticum Turgidum L. to Metribuzin', *Chemosphere*, **21**, 223–230 (1990).
52. J.S. Thornton, C.W. Stanley, 'Gas Chromatographic Determination of Sencor and Metabolites in Crops and Soils', *J. Agric. Food Chem.*, **25**, 380–386 (1977).
53. C.B. Ely, R.E. Frans, T.L. Lavy, R.E. Talbert, J.D. Mattice, 'Determining Diuron, Simazine, and Methiocarb Residues in Highbush Blueberries (*Vaccinium corymbosum*), *Hort. Sci.*, **28**, 33–35 (1993).
54. J.C. Pringle, L.W. Anderson, R.W. Raines, 'Residues in Crops Irrigated with Water Containing Simazine', *J. Agric. Food Chem.*, **26**, 1143–1147 (1978).
55. H. Roseboom, H.A. Herbold, 'Determination of Triazine Herbicides in Various Crops by Capillary Gas Chromatography with Thermionic Detection', *J. Chromatogr.*, **202**, 431–438 (1980).
56. K. Ramsteiner, W.D. Hormann, D.O. Eberle, 'Multiresidue Method for the Determination of Triazine Herbicides in Field-grown Agricultural Crops, Water and Soil', *J. Assoc. Off. Anal. Chem.*, **57**, 192–201 (1974).
57. J.F. Lawrence, 'Direct Analysis of Some Carbamate Pesticides in Foods by High Pressure Liquid Chromatography', *J. Agric. Food Chem.*, **25**, 211–212 (1977).
58. A.I. Garcia-Valcarcel, C. Sanchez-Brunete, L. Martinez, J.L. Tadeo, 'Determination of Dinitroaniline Herbicides in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **719**, 113 (1996).
59. E.W. Day, in *Analytical Methods for Pesticides and Plant Growth Regulators*, eds. G. Zweig, J. Sherma, Academic Press, New York, 341–352, Vol. X, 1978.
60. J.W. Vandeventer, W.F. Meggitt, D. Penner, 'Absorption, Translocation and Metabolism of Ethalfluralin and Trifluralin in *Solanum ssp.*', *Pestic. Sci.*, **17**, 380–384 (1986).
61. J.B. Tepe, R.E. Scroggs, in *Analytical Methods for Pesticides and Plant Growth Regulators*, eds. G. Zweig, J. Sherma, Academic Press, New York, 627–535, Vol. V, 1967.
62. K. Hawxby, E. Basler, P.W. Santelman, 'Temperature Effects on Absorption and Translocation of Trifluralin and Methazole in Peanuts', *Weed Sci.*, **20**, 285–289 (1972).
63. R.A. Conkin, in *Analytical Methods for Pesticides and Plant Growth Regulator*, eds. G. Zweig, J. Sherma, Academic Press, New York, 255–265, Vol. X, 1978.
64. A. Balinova, 'Gas Chromatographic Determination of Chloracetamide Herbicides in Plants and Soils', *J. Chromatogr.*, **455**, 391–395 (1988).
65. W.J. Patchett, W.J. Smith, in *Analytical Methods for Pesticides and Plant Growth Regulators*, eds. G. Zweig, J. Sherma, Academic Press, New York, 644–651, Vol. VI, 1972.
66. C. Sanchez-Brunete, T. Salto, J.M. Garcia-Baudin, J.L. Tadeo, 'Analysis of Triallate Residues in Cereals and Soil by Gas Chromatography with Ion-trap Detection', *J. Chromatogr.*, **562**, 525–530 (1991).
67. J.M. Garcia-Baudin, C. Sanchez-Brunete, T. Salto, M.C. Chueca, J.L. Tadeo, 'Enhancement of Wheat Tolerance to Triallate with the Safener MON 5501', *Weed Res.*, **32**, 275–278 (1992).
68. A.J. Cessna, N.W. Hoet, B.N. Drew, 'Tolerance and Residue Studies of Triallate in Lentils', *Can. J. Plant Sci.*, **60**, 1283–1288 (1980).
69. A.J. Cessna, 'Determination of Triallate Residues in Alfalfa and Companion Crops of Wheat and Barley', *Can. J. Plant Sci.*, **61**, 765–768 (1981).

70. G.H. Thiele, R.L. Zimdahl, 'Triallate Accumulation in Wild Oats and Barley', *Weed Sci.*, **24**, 183–185 (1976).
71. L.W. Cook, F.W. Zach, J.R. Fleeker, 'Steam Distillation and Gas–Liquid Chromatographic Determination of Triallate and Diallate in Milk and Plant Tissue', *J. Assoc. Off. Anal. Chem.*, **65**, 215–217 (1982).
72. M.L. Hopper, 'Solid-phase Partition Column Technology', in *Emerging Strategies for Pesticide Analysis*, eds. T. Cairns, J. Sherma, CRC Press, Boca Raton, FL, 39–50, 1992.
73. J.A. Casanova, 'Use of Solid-phase Extraction Disks for Analysis of Moderately Polar and Non-polar Pesticides in High Moisture Foods', *J. AOAC Int.*, **79**, 936–940 (1996).
74. M.L. Hopper, 'Improved Method for Partition of Organophosphate Pesticide Residues on a Solid Phase Partition Column', *J. Assoc. Off. Anal. Chem.*, **71**, 731–734 (1988).
75. M.A. Luke, J.E. Froberg, H.T. Masumoto, 'Extraction and Clean-up of Organochlorine, Organophosphate, Organonitrogen and Hydrocarbon Pesticides in Produce for Determination by Gas–Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **58**, 1020–1026 (1975).
76. L. Kandenczki, Z. Arpad, I. Garda, A. Ambrus, L. Gyorfi, G. Reese, W. Ebing, 'Column Extraction of Residues of Several Pesticides from Fruits and Vegetables: A Simple Multiresidue Analysis Method', *J. AOAC Int.*, **75**, 53 (1992).
77. S.A. Barker, A.R. Long, M.E. Hines II, 'Disruption and Fractionation of Biological Materials by Matrix Solid-phase Dispersion', *J. Chromatogr.*, **629**, 23–34 (1993).
78. C.C. Walker, H.M. Lott, S.A. Barker, 'Matrix Solid-phase Dispersion Extraction and the Analysis of Drugs and Environmental Pollutants in Aquatic Species', *J. Chromatogr.*, **642**, 225–242 (1993).
79. A. Di Muccio, A. Ausili, R. Donnarca, D.A. Barbini, A. Santilio, F. Vergori, G. De Meruilis, L. Sernicola, 'Solid-matrix Partition for Separation of Organochlorine Pesticide Residues from Fatty Materials', *J. Chromatogr.*, **552**, 241–247 (1991).
80. B.E. Richter, 'Supercritical Fluid Extraction Methods', in *Emerging Strategies for Pesticide Analysis*, eds. T. Cairns, J. Sherma, CRC Press, Boca Raton, FL, 1992.
81. K.G. Furton, J. Rein, 'Trends in Techniques for the Extraction of Drugs and Pesticides from Biological Specimens Prior to Chromagraphic Separation and Detection ', *Anal. Chim. Acta*, **23**, 99–114 (1990).
82. M.E.P. McNally, 'Method Development in Supercritical Fluid Extraction', *J. AOAC Int.*, **79**, 380–387 (1997).
83. S.B. Hawthorne, D.J. Miller, M.D. Burford, J.J. Langenfeld, S. Eckert-Tilotta, P.K. Louie, 'Factors Controlling Quantitative Supercritical Fluid Extraction of Environmental Samples', *J. Chromatogr.*, **642**, 301–317 (1993).
84. P. Brown, A. Charlton, M. Cuthbert, L. Barnett, L. Ross, M. Green, L. Gilles, K. Shaw, M. Fletcher, 'Identification of Pesticide Poisoning in Wildlife', *J. Chromatogr., A*, **754**, 463–478 (1996).
85. A. de Kok, M. Hiemstra, 'Optimization, Automation and Validation of the Solid-phase Extraction, Clean-up and On-line Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides in Fruits and Vegetables', *J. AOAC Int.*, **75**, 1063–1072 (1992).
86. J. Tekel, J. Kovačičá, 'Chromatographic Methods in the Determination of Herbicide Residues in Crops, Food and Environmental Samples', *J. Chromatogr.*, **643**, 291–303 (1993).
87. D.L. Stalling, R.C. Tindle, J.L. Johnson, 'Clean-up of Pesticide and Polychlorinated Biphenyl Residues in Fish Extracts by Gel Permeation Chromatography', *J. AOAC*, **55**, 32–38 (1972).
88. J.R. Patterson, 'Reducing Solvent Consumption in Automated Gel Permeation Chromatographic Clean-up for Pesticide Residue Analysis: A Modified GPC Autoprep 1002', *J. Assoc. Off. Anal. Chem.*, **74**, 1016–1018 (1991).
89. J.A. van Rijn, L.G.M.Th. Tuinstra, 'Miniaturization of Size-exclusion Chromatography as a Powerful Clean-up Tool in Residue Analysis', *J. Chromatogr.*, **552**, 517–526 (1991).
90. J.E. Quintanilla-Lopez, R. Lebron-Aquilar, L.M. Polo-Diez, 'Comparative Study of Clean-up and Fractionation Methods for the Determination of Organochlorine Pesticides in Lipids by Gas Chromatography', *J Chromatogr.*, **91**, 303–311 (1992).
91. 'AOAC Official Methods of Analysis', 16th edition, AOAC International, Baltimore, MD, 1996.
92. K.K. Hatzios, 'Biotransformations of Herbicides in Higher Plants', in *Environmental Chemistry of Herbicides*, eds. R. Grover, A.J. Cessna, CRC Press, Boca Raton, FL, 141–185, Vol. 2, 1991.
93. J. Cai, Z. el Rassi, 'Capillary Zone Electrophoresis of Two Cationic Herbicides, Paraquat and Diquat', *J. Liq. Chromatogr.*, **15**, 1193–1200 (1992).
94. M. Tomita, T. Okuyama, Y. Nigo, 'Simultaneous Determination of Paraquat and Diquat in Serum Using Capillary Electrophoresis', *Biomed. Chromatogr.*, **6**, 91–94 (1992).
95. M. Tomita, T. Okuyama, Y. Nigo, B. Uno, S. Kawai, 'Determination of Glyphosate and its Metabolite, (Aminomethyl)phosphonic Acid, in Serum Using Capillary Electrophoresis', *J. Chromatogr.*, **571**, 324–330 (1991).
96. W. Liao, T. Joe, W.G. Cusick, 'Multiresidue Screening Method for Fresh Fruits and Vegetables with Gas Chromatographic/Mass Spectrometric Detection', *J. Assoc. Off. Anal. Chem.*, **74**, 554–565 (1991).

97. F. Wach, 'Benchtop GC/MS Instruments', *Anal. Chem.*, **66**, 927A–930A (1994).
98. S.A. McLuckey, G.J. Van Berkel, D.E. Goeringer, G.L. Glish, 'Ion Trap Mass Spectrometry of External Generated Ions', *Anal. Chem.*, **66**, 689A–696A (1994).
99. K.L. Busch, G.L. Glish, S.A. McLuckey, *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*, VCH Publishers, New York, 1988.
100. R.A. Yost, C.G. Enke, 'Triple Quadrupole Mass Spectrometry for Direct Mixture Analysis and Structure Elucidation', *Anal. Chem.*, **51**, 1251A, 1252A, 1256A, 1258A, 1260A, 1262A (1979).
101. J.V. Headley, K.M. Peru, M.T. Arts, 'Detection of Herbicide Residues in Lipid-rich Tissue Using Tandem Mass Spectrometry', *Anal. Chem.*, **67**, 4349–4353 (1995).
102. J.V. Headley, L.C. Dickson, A.J. Cessna, 'Herbicide Residues in Biota, Determination', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, John Wiley and Sons, New York, 2139–2165, 1998.
103. R. Buffington, *GC–Atomic Emission Spectroscopy Using Microwave Plasmas*, Hewlett-Packard, Avondale, CA, 1988.
104. S.M. Lee, P.L. Wylie, 'Comparison of the Atomic Emission Detector to Other Element Selective Detectors for the Gas Chromatographic Analysis of Pesticide Residues', *J. Agric. Food Chem.*, **39**, 2192–2199 (1991).
105. S.D. West, E.W. Day Jr, 'Determination of Fluridone Residues in Meat, Milk, Eggs and Crops by High Performance Liquid Chromatography or Gas Chromatography', *J. Agric. Food Chem.*, **36**, 53–56 (1988).
106. R.J. Flanagan, M. Ruprah, 'HPLC Measurement of Chlorophenoxy Herbicides Bromoxynil and Ioxynil in Biological Specimens to Aid Diagnosis of Acute Poisoning', *Clin. Chem.*, **35**, 1342–1347 (1988).
107. L.M. Shalaby, S.W. George, 'Liquid Chromatography/Mass Spectrometry Multiresidue Methods for Thermally Labile Pesticides', in *Emerging Strategies for Pesticide Analysis*, eds. T. Cairns, J. Sherma, CRC Press, Boca Raton, FL, 295–307, 1992.
108. S.J. Stout, A.R. da Cunha, G.L. Picard, M.M. Sarpour, 'Simplification of Analytical Methods in Pesticide Residue Analysis by Liquid Chromatography/Electrospray Ionization Mass Spectrometry and Tandem Mass Spectrometry', *J. AOAC Int.*, **81**, 685–690 (1998).
109. V.Y. Taguchi, S.W.D. Jenkins, P.W. Crozier, D.T. Wang, 'Determination of Diquat and Paraquat in Water by Liquid Chromatography–(Electrospray Ionization) Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **9**, 830–839 (1998).
110. E.W. Zahnow, 'Analysis of the Herbicide Sulfometuron Methyl in Fish and Green Plants by Liquid Chromatography', *J. Agric. Food. Chem.*, **33**, 1206 (1985).
111. J.F. Garcia, D. Barcelo, 'An Overview of LC/MS Interfacing Systems with Selected Applications', *J. High Resolut. Chromatogr.*, **16**, 633–641 (1993).
112. F. Foret, V. Sutacek, P. Pocek, 'Separation of Some Triazine Herbicides and their Solvolytic Products by Capillary Zone Electrophoresis', *Electrophoresis*, **11**, 95–97 (1990).
113. J.C. Hall, R.J.A. Deschamps, M.R. McDermott, 'Immunoassays to Detect and Quantitate Herbicides in the Environment', *Weed Technol.*, **4**, 226–234 (1990).
114. J.P. Sherry, 'Immunodetection of Ecosystem Contaminants. Research, Application, and Acceptance in Canada', in *Immunoanalysis of Agrochemicals Emerging Technologies*, eds. J.O. Nelson, A.E. Karu, R.B. Wong, ACS Symp. Ser. No. 586, American Chemical Society, Washington, DC, 1995.
115. B.D. Hammock, S.J. Gee, 'Impact of Emerging Technologies on Immunochemical Methods for Environmental Analysis', in *Immunoanalysis of Agrochemicals. Emerging Technologies*, eds. J.O. Nelson, A.E. Karu, R.B. Wong, ACS Symp. Ser. No. 586, American Chemical Society, Washington, DC, 1995.
116. J. Van Emon, J. Seiber, B. Hammock, 'Application of an Enzyme-linked Immunosorbent Assay (ELISA) to Determine Paraquat and Diquat Residues in Milk, Beef and Potatoes', *Bull. Environ. Contam. Toxicol.*, **39**, 490–497 (1987).
117. W.H. Newsome, P.G. Collins, 'Determination of Imazamethabenz in Cereal Grain by Enzyme-linked Immunosorbant Assay', *Bull. Environ. Contam. Toxicol.*, **47**, 211–216 (1991).
118. P.A. Mills, 'Detection and Semi-quantitative Estimation of Chlorinated Organic Pesticide Residues in Foods by Paper Chromatography', *J. Assoc. Off. Anal. Chem.*, **42**, 734 (1959).
119. P.A. Mills, J.H. Onley, R.A. Gaither, 'Rapid Method for Chlorinated Pesticide Residues in Non-fatty Foods', *J. Assoc. Off. Anal. Chem.*, **46**, 186 (1963)
120. R.W. Storherr, M.E. Getz, R.R. Watts, 'Identification and Analyses of Five Organophosphate Pesticides: Recoveries from Crops Fortified at Different Levels', *J. Assoc. Off. Anal. Chem.*, **47**, 1087–1093 (1964).
121. R.T. Krause, 'Multiresidue Method for Determining *N*-Methylcarbamate Insecticides in Crops, Using High Performance Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **63**, 1114 (1980).
122. W. Specht, M. Tilkes, 'Gas Chromatographic Determination of Pesticide Residues After Clean-up by Gas-permeation Chromatography and Mini Silica Gel Column Chromatography. 3. Communication – Clean-up of Foods and Feeds of Vegetable and Plant Origin for Multiresidue Analysis of Fat-soluble and Water-soluble Pesticides', *Fresenius' J. Anal. Chem.*, **310**, 300 (1971).
123. W. Specht, M. Tilkes, 'Gas Chromatographic Determination of Pesticide Residues After Clean-up by Gas-permeation Chromatography and Mini Silica Gel

Column Chromatography. Part 5. Clean-up of Foods and Feeds of Vegetable and Plant Origin for Multiresidue Analysis of Fat-soluble and Water-soluble Pesticides', *Fresenius' J. Anal. Chem.*, **322**, 443–445 (1985).

124. D.L. Heikes, 'Environmental Contamination in Table-ready Foods from the Total Diet Program of the Food and Drug Administration', in *Advances in Environmental Science and Technology*, eds. J.O. Nriagu, M.S. Simmons, John Wiley and Sons, New York, 31–57, 1990.
125. S.M. Lee, M.L. Papathakis, H.C. Feng, G.F. Hunter, J.E. Carr, 'Multipesticide Residue Method for Fruits and Vegetables – California Department-of-Food-and-Agriculture', *Fresenius' J. Anal. Chem.*, **339**, 376–383 (1994).
126. M.L. Hopper, B.M. McMahon, K.R. Griffitt, K. Kline, M.E. Fleming-Jones, D.C. Kendall, 'Analysis of Fatty and Nonfatty Foods for Chlorophenoxy Alkyl Acids and Pentachlorophenol', *J. AOAC Int.*, **75**, 707–713 (1992).
127. Anonymous, *Pesticide Analytical Manual*, eds. B.M. McMahon, R.F. Wagner, US Food and Drug Administration, Washington, DC, Vol. 1, Section 402, 1994.

Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of

Aldo Laganà
Università La Sapienza, Rome, Italy

1 Introduction **1**
2 Regulations **2**
3 Analytical Methods for Determination of Post-emergence Herbicides in Environmental and Food Samples: General **3**
3.1 Extraction 3
3.2 Cleanup Procedures 3
3.3 Determination 4
4 Imidazolinone Herbicides **4**
4.1 Chemical and Physical Properties 4
4.2 Analytical Methods for Imidazolinone Analysis 8
4.3 Determination 8
5 Aryloxyphenoxypropionic Acids/esters **14**
5.1 Chemical and Physical Properties 14
5.2 Analytical Methods for Aryloxyphenoxypropionate Analysis 14
5.3 Determination 21
6 Diphenyl Ether Herbicides **26**
6.1 Chemical and Physical Properties 26
6.2 Analytical Method for Diphenyl Ether Analysis 28
6.3 Determination 31
Abbreviations and Acronyms **35**
Related Articles **36**
References **37**

This article critically examines the existing methodologies including residue isolation and preconcentration, cleanup, chromatographic analysis, such as gas chromatography (GC), high-performance liquid chromatography (HPLC), hyphenated techniques, such as gas chromatography/mass spectrometry (GC/MS) and high-performance liquid chromatography/mass spectrometry (HPLC/MS), and some nonchromatographic techniques, such as immunoassay (IA), of herbicide residues in crops, food, and environmental samples. The main structural groups of herbicides discussed are imidazolinones (IMIs), aryloxyphenoxy propionic acids/esters, and diphenyl ethers (DPhEs).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The development of chemicals in selective control of weed growth is a fascinating success story that has unfolded largely since the early 1980s. At present, herbicides represent the largest single agrochemical sector, worth approximately US$15 billion.

In order to limit environmental pollution and damage to humans, the main goals should be to produce herbicides that are more specific towards target plants, that can be applied in low doses, and that are rapidly degraded and inactivated by the environment. Until recently herbicide treatment was made following the pre-emergence (PrE) procedure, namely the herbicide was spread over the soil after sowing, but before crop germination.

Continuous research progress has produced herbicides that are so selective and efficacious that they may be applied at very low concentrations (down to $1\,g\,ha^{-1}$ rather than the $500–2000\,g\,ha^{-1}$ of traditional pesticides), directly on foliage, and only when the danger threshold is approached. Thus PrE treatments have been superseded by post-emergence (PoE) treatments in which herbicide application takes place after germination.

A PoE herbicide can be defined as a foliage-applied product used to control weeds that have emerged in competition with the developing crop. PoE herbicides' high selectivity make it necessary to treat specific crops with specific compounds, (i.e. compounds that are well tolerated by the crop but not by the weeds), thus avoiding the more persistent and environmentally less safe PrE herbicides. However, many PoE herbicides retain residual activity in the soil and can thus control late-germinating weeds. The more used classes of PoE herbicides, which are predicted to become increasingly popular in the near future, are listed in Table 1.

The opportunity offered to farmers, to use small amounts of chemicals with a very low toxicity to mammals, has encouraged a fast worldwide diffusion of PoE herbicides as shown in Figure 1.

Even though the risk to humans of ingesting toxic doses of herbicide residues in food seems low, it is important to monitor herbicide levels in the environment and food commodities because of their extensive use and documented occurrence both in the environment and in foods. By ''residues'' is meant low levels of pesticides and their metabolites that remain in food after application. To protect the health of their consumers, most countries

Table 1 Principal PoE herbicide classes and their primary target site of action

Chemical class	Primary target site of action
Sulfonylureas	acetolactate synthetase
IMIs	acetolactate synthetase
Aryloxyphenoxypropionic acids/esters	acetyl-CoA carboxylase
DPhEs	protoporphyrinogen oxidase
Triazolopyrimidine sulfonanilides	acetolactate synthetase
Phenylaminosulfonylureas	acetolactate synthetase
Cyclohexanediones	acetyl-CoA carboxylase

CoA, coenzyme A.

have introduced laws not only governing the use of pesticides but also setting limits for the levels of pesticide residues that can be tolerated in drinking water and foods. The maximum residue limit (MRL) is the maximum acceptable level of a residue that regulatory authorities will tolerate on the basis of toxicological food intake data.

2 REGULATIONS

The ''Maximum Admissible Concentration'' of pesticides in drinking water is defined by the European Community Water Directive as $100\,ng\,L^{-1}$ for individual pesticides and as $500\,ng\,L^{-1}$ for the sum of pesticides.[1] Pesticide residues in food are regulated at the international and national levels according to the toxicity of the pesticide and the human intake of a particular crop. The levels of pesticide residues are controlled by the MRLs, established by each country. These sometimes cause conflict because residue levels acceptable in one country can be unacceptable in another. This problem of MRL harmonization has been mainly dealt with by two international organizations: the European Union (EU) at European level, and the Codex Alimentarius Commission,[2,3] an international body established in 1962, to implement the Joint FAO/WHO Food Standards Program. Nations that are members of the Food and Agriculture Organization (FAO) and World Health Organization (WHO) may become members of the Codex Alimentarius Commission.[4]

The EU directives promulgated between 1976 and 1989 achieved only partial harmonization because they fixed the MRLs of only 64 out of the more than 600 active ingredients on the market.[5] In 1990, the EU issued a new directive that fixes MRLs for all EU countries, thus eliminating the higher MRLs acceptable in some countries.[2,3]

There are EU directives that establish sampling and analytical methods for the official control of pesticide

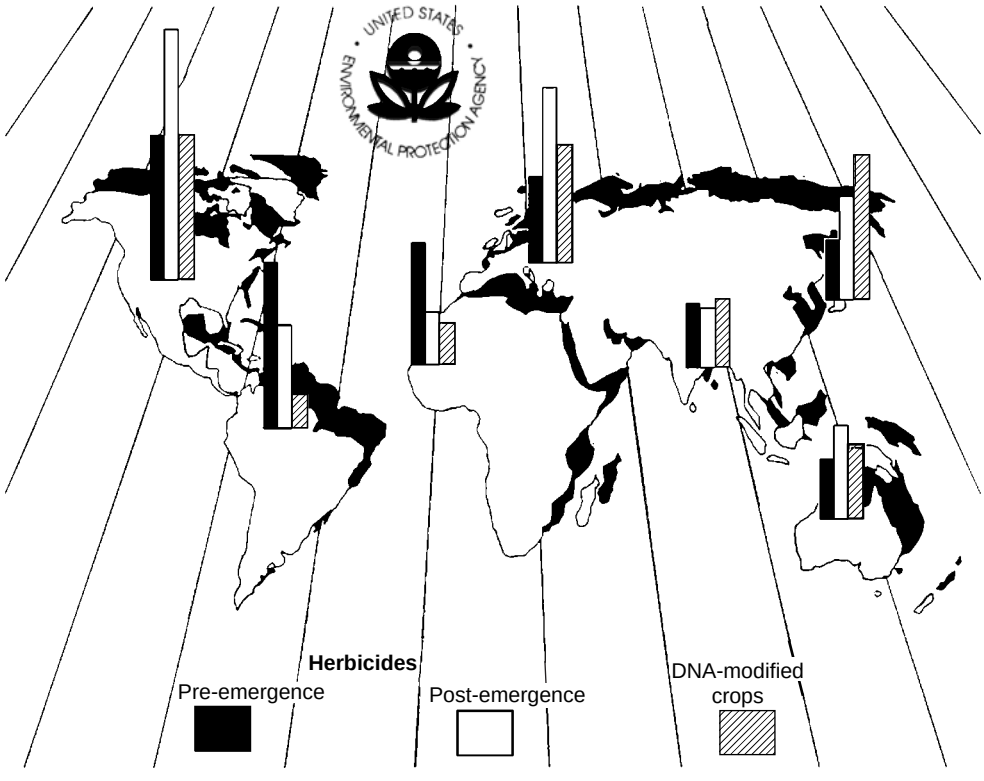

Figure 1 Herbicide diffusion in the world. (Source: United States Environmental Protection Agency (USEPA).)

residues in fruit, vegetables, and products for human consumption. Also, an EU directive published in 1991 covers the marketing of pesticide products. This demands a large number of studies on residues before an active ingredient can be authorized at European level.[3]

In the USA, food safety is an important responsibility not only of the Food and Drug Administration (FDA), but also of the United States Department of Agriculture (USDA), the USEPA, and other agencies at the federal, state, and local levels. The USEPA registers or approves the use of pesticides and establishes a tolerance level if the use of a pesticide may lead to residues in food.[6]

The FDA is responsible for enforcing the USEPA-established pesticide tolerances for foods shipped in interstate commerce, with the exception of meat and poultry, which are the responsibility of the USDA.[7]

The detection and determination limits required for routine analytical methods for herbicide residues should not be higher than 10–50% of the corresponding MRL as recommended by Frehse.[8] This puts the highest requirements of methods for the determination of residues in drinking water, where the maximum permissible levels are sometimes as low as $100\,ng\,L^{-1}$, and in food.

3 ANALYTICAL METHODS FOR DETERMINATION OF POST-EMERGENCE HERBICIDES IN ENVIRONMENTAL AND FOOD SAMPLES: GENERAL

In order to analyze PoE herbicides, knowledge of the physical and chemical properties of the classes of pesticides is essential to develop the best residue analytical methods with the most appropriate instruments.

The sequence of steps in the analysis are the sample preparation and determination. The sample preparation consists in the extraction and cleanup procedures.

3.1 Extraction

The extraction of PoE residues from the matrix is a requisite for almost all the methods of analysis, either for the whole body of water sample or for solid or semisolid substances. A wide variety of extraction procedures have been used to remove PoE residues from environmental water, soil, and food.

For liquid samples, extraction will be accomplished by liquid–liquid extraction (LLE) partition using an organic solvent nonmiscible with water, by solid-phase extraction (SPE), or solid-phase microextraction (SPME) techniques, and supercritical fluid extraction (SFE) after enrichment on a precolumn packed with a polymeric phase.

Extraction of a solid or semisolid sample depends on the polarity of the PoE residues as well as the polar matrix. If the PoE residues are ionic or polar compounds, samples are usually extracted by mixing with pH-adjusted water or organic solvent diluted with pH-adjusted water using a shaker, blender, Polytron™, sonication, or Soxhlet. For less polar PoE residues, organic solvents (acetonitrile, methanol, acetone, ethyl acetate) and solvent mixtures are ground with the samples. The extraction solvents and residual solids are then separated by filtration or centrifugation.

Newer techniques for the extraction of PoE residues are:

- SFE using a supercritical fluid, usually carbon dioxide, with or without organic modifier (methanol, acetonitrile);
- microwave-assisted extraction (MAE), which has the advantages of lower solvent consumption and greater speed than conventional extraction;
- soil column extraction (SCE), using organic solvent mixed with pH-adjusted water. This technique reduces the complexity of the analytical method and the sample preparation time.

3.2 Cleanup Procedures

An extraction technique is selected to obtain the highest recovery for a wide range of contaminants. Therefore, the extract is likely to contain a high proportion of coextracted material. The goal of cleanup is to remove as much interfering coextractive material and as little of the analyte as possible. A cleanup procedure for sample extracts is often required, depending on the type of compound and kind of sample to be analyzed and for the selectivity of the analytical equipment used in determination.

Liquid–liquid partition (LLP) is the most common first step in cleaning up the extract to remove water-miscible solvents or reduce the amount of fat present in the extracts. The extraction efficiency can be enhanced by adding sodium chloride or sodium sulfate to salt out the analyte. If the PoE analyte is an acid, it can be brought into the aqueous or organic phase by modulating the pH value.

The cleanup techniques most commonly employed for extracts containing PoE residues are column chromatographic SPE systems and gel permeation chromatography (GPC).

Open-column chromatography uses columns packed with florisil, silica or alumina in the normal phase, or with reversed-phase (RP) columns of C_{18}, cyanopropyl, phenyl

or graphitized carbon black (GCB). These columns can be prepared in the laboratory, or use can be made of SPE cartridges.

To some extent, interference can be overcome by using size-exclusion chromatography (SEC) with gel, for example, GPC. The most useful gel for PoE cleanup is BioBead SX-3. A disadvantage of the GPC step is that it is slow and requires the use of comparatively large volumes of solvent with the attendant risk of contamination.

3.3 Determination

Nowadays the classic techniques for determination and quantification of PoE residues in environmental and food samples are GC and HPLC. GC and HPLC, in conjunction with mass spectrometry (MS), can be powerful tools for confirmatory analysis. IAs are now being seen as useful supplements to classical chromatographic analytical systems. IA provides a specific reaction with a compound or a group of compounds, which could reduce matrix effect and the need for sample cleanup and concentration. Classical direct current (DC) polarography and fast scan differential pulse voltammetry (FSDVP) have also been applied.

3.3.1 Gas Chromatography

GC on a packed column or capillary column is still the most common separatory technique in PoE residue analysis, and the most common detection systems are nitrogen–phosphorus detection (NPD), electron capture detection (ECD), atomic emission detection (AED), and flame ionization detection (FID). The direct analysis of PoE acids by GC is obstructed because of their acidic character and low volatility. Therefore, derivatization is usually necessary before GC analysis. In environmental applications, the most-used derivatization procedures employ diazomethane, 2,2,2-trifluoroethanol (TFE), pentafluorobenzyl bromide, methyl, ethyl and butyl chloroformate, and trifluoroethyl bromine.

3.3.2 High-performance Liquid Chromatography

Nowadays, HPLC is being extensively used in PoE residue analysis and related areas where the chemicals of interest are frequently of low volatility for GC separation. Briefly, HPLC methods for determination of the PoE residue in environmental samples could employ RP chromatography with C_{18}, C_8, cyanopropyl, and phenyl columns followed by ultraviolet (UV) absorption, UV diode array detector (DAD), fluorimetric (FI) detection, amperometric electrochemical detection (AD) and photoconductivity detection (PCD). Isocratic or gradient elution with a large change in mobile-phase composition is also used. Buffer solutions with the pH adjusted below the pK_a values of the acidic compounds are commonly employed. For ionic PoE residues an ion-pairing mechanism has been reported.[9]

3.3.3 Hyphenated Techniques

The confirmation of all results is a fundamental requirement of PoE residue analysis, as in all other trace analyses. As chromatography is a poor identification technique, hyphenated techniques have been developed and have gained popularity for increasing the selectivity and the certainty of identification.

MS, coupled with GC or HPLC, is being used more often.

3.3.4 Immunoassays

Although IAs have been widely used in clinical analysis since the 1960s, their application in pesticide residue determination is recent. At the present time, PoE residual determination has only been developed for IMIs and for diclofop-methyl. Application of IAs in PoE analysis will probably increase in future. This technique has great potential because it offers many advantages including sensitivity, specificity, and analysis speed.

However, the disadvantage is that each kit is specific to a single analyte.

3.3.5 Other Techniques

Nitropesticides have been studied with the aim of establishing their electrochemical reduction and adsorption characteristics at the mercury-drop electrode with regard to the medium acidity and the compound structure.

4 IMIDAZOLINONE HERBICIDES

4.1 Chemical and Physical Properties

The chemical structures, names, and chemical abstracts service registry numbers of IMI herbicides are reported in Table 2. The basic physical, chemical, and other properties of the IMI herbicides are listed in Tables 3–5. The data in these tables are taken from *The Pesticide Manual*.[10]

The IMIs are a relatively new class of herbicides, used to control a wide spectrum of broad-leaved weeds and grasses in a variety of agricultural commodities.[11] These herbicides are very potent weedkillers and are used in doses that are substantially lower than conventional herbicides.

The water solubility and vapor pressure of the IMI carboxylic acids are low, whereas the octanol–water

Table 2 Chemical structures, IUPAC-recommended names[10] and Chemical Abstracts Service registry numbers of IMI herbicides

Chemical structure	Name	CAS
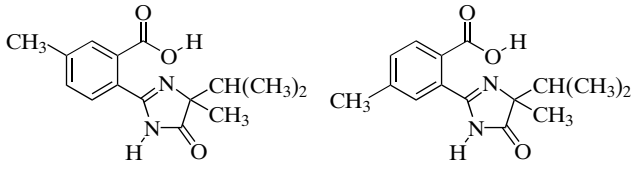	***m*-, *p*-imazamethabenz** (±)-6-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)-*m*-toluic acid (±)-2-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)-*p*-toluic acid	[100728-84-5]
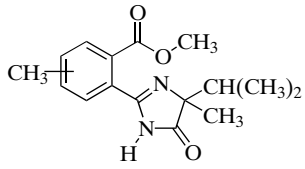	**imazamethabenz-methyl** methyl-(±)-6-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)-*m*-toluate methyl-(±)-2-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)-*p*-toluate	[81405-85-8]
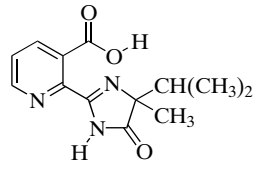	**imazapyr** 2-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)nicotinic acid	[81334-34-1]
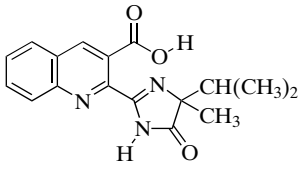	**imazaquin** (*R,S*)-2-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)quinoline-3-carboxylic acid	[81335-37-7]
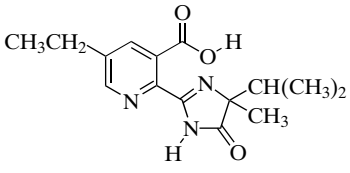	**imazethapyr** (*R,S*)-5-ethyl-2-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)nicotinic acid	[81335-77-5]
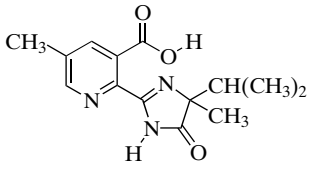	**imazmethapyr** (*R,S*)-5-methyl-2-(4-isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)nicotinic acid	–

partition coefficients are pH dependent. At pH 4, the IMI carboxylic acids are neutral and therefore lipophilic, whereas at pH 7 the carboxylic acids are ionized. The pK_a values of several IMI herbicides lie between 1.9 and 11.4. Up to three inflection points have been observed in the titration of IMIs. The IMI ring is amphoteric and behaves as a weak base or a weak acid. The first inflection point corresponds to protonation of the IMI ring at pK_1 of about 1.9–3.3, whereas the second reflects ionization of the carboxylic acid at pK_2 values around 3.6–3.9. The third inflection point, determined only for imazapyr, corresponds to IMI ring deprotonation at a pK_3 of about 11.4.

The hydrolytic stability of the IMI ring is pH dependent. At environmentally relevant pH and temperature, hydrolysis is very slow. At pH less than 7, the IMI ring is stable.

As shown in Table 2, the members of this class of herbicides have similar structural features centered around the IMI ring and an attached aromatic system bearing a carboxylic acid moiety. IMIs have an excellent activity against annual and perennial grasses and broad-leaved

Table 3 Chemical and physical properties of IMI herbicides

Property	Imazamethabenz	Imazamethabenz-methyl	Imazapyr	Imazaquin	Imazethapyr	Imazmethapyr
Empirical formula	$C_{15}H_{18}N_2O_3$	$C_{16}H_{20}N_2O_3$	$C_{13}H_{15}N_3O_3$	$C_{17}H_{17}N_3O_3$	$C_{15}H_{19}N_3O_3$	$C_{14}H_{17}N_3O_3$
Molecular weight	274.3	288.3	261.3	311.3	289.3	275.3
Physical form	–	off-white fine to lumpy powder with slight musty odor	white to tan powder with a slight odor of acetic acid	tan solid with slightly pungent odor	colorless crystals	–
Melting point (°C)	–	113–153	169–173	219–224	169–173	–
Vapor pressure (mPa)	–	1.5×10^{-3} (25 °C)	<0.013 (60 °C)	<0.013 (60 °C)	<0.013 (60 °C)	–
Solubility at 25 °C						
water ($g\,L^{-1}$)	–	1.370 *p* isomer 0.857 *m* isomer	11.3	0.06–0.12	1.4	–
acetone ($g\,L^{-1}$)	–	230 ($g\,kg^{-1}$)	33.9	–	48.2	–
toluene	–	45	1.80	0.4	5	–
dimethyl sulfoxide	–	216	471	159	422	–
n-heptane	–	0.6	–	–	0.9	–
n-hexane	–	0.4	0.0095	–	–	–
methanol	–	309	105	–	105	–
isopropanol	–	183	–	–	17	–
methylene chloride	–	–	87.2	14	185	–
pK_a	–	2.9 (23.5 °C)	1.9 (pK_{a1}) 3.6 (pK_{a2}) 11.4 (pK_{a3})	3.8	2.1 (pK_{a1}) 3.9 (pK_{a2})	–

Table 4 Stability of the IMI herbicides

Parameter	Imazamethabenz	Imazamethabenz-methyl	Imazapyr	Imazaquin	Imazethapyr	Imazmethapyr
Stability	–	24 months at 25 °C	2 years at 25 °C	3 months at 45 °C	$DT_{50} \cong 3$ days	–
	–	12 months at 37 °C	1 year at 37 °C	2 years at room temperature in the dark		
	–	3 months at 45 °C	3 months at 45 °C		–	–
Other observations	–	hydrolysis is rapid at pH 9, slow at pH 5–7	avoid storage >45 °C, in solution is decomposed in simulated sunlight	rapidly degraded on exposure to UV light	rapidly degraded in sunlight	–
DT_{50} in water	–	–	6 days at pH 5–9	–	–	–
DT_{50} in soil	–	25–45 days	residual activity:	60 days	1–3 months	–
	–	slowly degrades to corresponding free acids in sandy loam and clay loam soils under both aerobic and anaerobic conditions	6 months to 2 years in temperate climates; 3–6 months in tropical climates	steadily degraded by microbial activity and photolysis, but may remain active for several weeks to several months depending on environmental conditions		

DT_{50} = time to 50% loss.

weeds when applied either PrE or PoE. They function by inhibiting acetohydroxy acid synthase (AHAS), the feedback enzyme in the biosynthesis of branched-chain essential acids.[12–14] This enzyme is not present in animals.

In order for a herbicide to kill plants, the chemical must be adsorbed and translocated to the site of action. In the case of IMIs, the site of action is the AHAS enzyme, which is concentrated in plant meristematic tissues. IMIs enter plants through the root or shoot tissue and are then translocated throughout the plant via the xylem, the water transport system, or the phloem, the sugar transport system. The mechanism by which IMIs are adsorbed and retained within plant cells is called

Table 5 Applications and environmental intake of the IMI herbicides

Parameter	Imazametha-benz	Imazamethabenz-methyl	Imazapyr	Imazaquin	Imazethapyr	Imazmetha-pyr
Acute oral LD_{50} rat (mg kg^{-1} body weight)	–	>5000	>5000	>5000	>5000	–
Acute dermal LD_{50} rabbit (mg kg^{-1} body weight)	–	>2000	>2000	>2000	>2000	–
WHO toxicity class[a]	–	III	III	III	III	–
Fate in:						
animals	–	rapidly excreted	excreted within 24 h	excreted within 2 days	excreted within 24 h	–
plants	–	methyl group of benzene moiety is oxidized to hydroxymethyl	parent compound	rapidly metabolized by soyabean to inactive compound	rapidly metabolized in nonsusceptible plants	–
Uses[b]	–	U1	U3	U4	U5	–

[a] WHO toxicity class: Ia, extremely hazardous; Ib, highly hazardous; II, moderately hazardous; III, slightly hazardous.
[b] U1 = PoE control of *Avena* spp., *Alopecurus myosuroides, Apera spica-venti* and some dycotyledonous weeds in wheat, barley, rye, and sunflowers; U2 = PoE weed control in soybeans, and other legumes grown in rotation with sugar beet and other crops where more persistent IMIs are not suited; U3 = PoE control of annual and perennial grasses, sedges and broad-leaved weeds, as well as many brush and deciduous tree species; U4 = PrE and PoE control of broad-leaved weeds in soybeans; U5 = PrE and PoE control of annual and perennial grasses and broad-leaved weeds in soybeans and other leguminous crops.

ion-trapping.[15] The pH gradient that occurs within plants is typically pH 4.5–5 outside the cell and 7–7.5 inside the cell.[16] Although IMIs are water-soluble compounds, under acidic conditions the molecules are relatively more lipophilic and, as such, are able to diffuse across cell membranes. Once inside, the molecule becomes charged in the alkaline environment with a concomitant decrease in lipophilicity. Because the charged form of the IMI is slower to diffuse across the plasma membrane, the compound accumulates, or is trapped within the cell.

Imazapyr is a broad-spectrum herbicide that controls most annual and perennial grasses and broad-leaved weeds, woody brush, and deciduous trees in noncrop areas such as railroads, utility areas, pipeline and highway rights-of-way, utility plant sites, petroleum tank farms, pumping installations, fence rows, storage areas, and nonirrigation ditchbanks. It is used to establish and maintain wildlife openings, to prepare sites for conifer planting, and to release conifers from competing vegetation. Imazapyr has also been developed for use in sugar cane and plantation crops, such as rubber and oil palm.

Imazapyr PoE application is preferred in most situations, particularly for the control of perennial weeds. For maximum herbicidal activity, imazapyr should be applied when the weeds are growing vigorously.

Imazaquin was the first IMI to be widely accepted as a herbicide for broad-spectrum weed control. Initial results have demonstrated that imazaquin can be applied preplant incorporated, PrE, or PoE for the control of many difficult-to-control soybean weed species.

In early testing, it was discovered that imazaquin applied as a PoE treatment uniquely provided both burndown of susceptible weeds present and residual control of later-germinating weeds.

Imazethapyr is a selective IMI used to control a wide spectrum of broad-leaved weeds and grasses in soybeans and several leguminous crops. Imazethapyr is a uniquely flexible herbicide that can be applied early preplant, at planting, or PoE. The herbicide is very effective in no-till as well as conventional tillage situations. Residual season-long control was recognized as an important feature of imazethapyr, and in general it was determined that optimum application rates for soil and PoE applications are the same, and efficiency is not affected by soil type or characteristics.

Imazamethabenz-methyl is a selective, PoE herbicide used to control weeds in cereal crops. It controls many of the economically important weeds that infest wheat, barley, and rye. Imazamethabenz-methyl is applied as ester, but it is the acidic form that is herbicidally active.

The first sign of herbicidal activity is chlorosis of the youngest leaves; later in the season the whole plant becomes necrotic or remains as a small, noncompetitive weed at the base of the crop.

Imazamethabenz-methyl also has soil residual activity; thus, susceptible weeds that germinate after application can be controlled.

The IMI herbicides have been subjected to exhaustive toxicological studies to determine their toxicity to mammals and other nontarget organisms and to evaluate the potential hazard of handling and applying these herbicides. As demonstrated by the results of those studies, the IMIs have a low toxicological potential, partially because they act by inhibiting a biosynthetic process at a site present only in plants. In addition, these herbicides are excreted rapidly by rats before they can accumulate in the tissues or blood.

4.2 Analytical Methods for Imidazolinone Analysis

4.2.1 Extraction

The extractive solvent and the method used depend on the nature of sample. For liquid samples, particularly for groundwater and surface water, extraction has been obtained by using a 6-mL SPE column packed with 1 g of Baker Light-load Octadecil or Waters Sep-Pak Plus Environmental,[17] and 1 g of Carbograph-1.[18] Extraction from solid samples depends on the polarity of the pesticides as well as the sample matrix. For food samples, extractions are usually performed by mixing them with methanol[19,20] or a methanol–water (25 : 75) solution acidified with HCl.[21] For solid samples such as environmental soils, extraction can be carried out in two ways: (1) mechanically stirring the sample with different organic solvents, such as methanol, acetone, ethyl acetate, diethyl ether, 2-propanol, ethanol, and dichloromethane,[22] ammonium hydrogen carbonate,[23] 0.1 mol L^{-1} sodium carbonate,[24] methanol–water (50 : 50, v/v),[25] 0.5 mol L^{-1} NaOH, and methanol;[26] (2) extracting imazamethabenz-methyl[22] and imazaquin[27] by SFE, usually using CO_2 with methanol as organic modifier.[22] Two new techniques that have been developed to extract IMIs from solid samples are MAE[28–30] and SCE.[18] Stout et al.[28,29] found that MAE with NH_4OAc/NH_4OH (0.1 mol L^{-1}, pH 10) as the extractant achieved the best results with different soils. They applied the same technique, using several types of extractants: (namely H_2O, CH_3OH, 10% H_2O/CH_3OH, and 0.1 mol L^{-1} NH_4OAc/H_2O) to the extraction of imazethapyr, its hydroxy metabolite, and the glucose conjugate from 11 different preparations.[30] *m*-imazamethabenz, *p*-imazamethabenz, *m*-,*p*-imazamethabenz-methyl, imazapyr, imazaquin and imazethapyr have been extracted by SCE, using as the extraction phase a methanol–ammonium carbonate mixture (0.1 mol L^{-1}, pH 8.2, 50 : 50, v/v).[18] The SCE technique makes the sample preparation method efficient and economical, and the preparation time is reduced.

Extraction and cleanup procedures used in the determination of IMI herbicides are shown in Table 6.

4.2.2 Cleanup

The extract obtained during an extraction step usually has to undergo a further cleanup procedure. After solid samples have been ground in the presence of a mixture of polar solvent and water-immiscible solvent, the resulting mixture is cleaned up by LLP and/or column chromatography.

The solvent partition system used in IMI residue analysis includes dichloromethane,[24,26] and hexane.[20] For the solvent partition system of oil obtained after the filtrate evaporation under reduced pressure, Mortimer and Weber[19] use 1 N NaOH and 25 mL of hexane–EtOAc (1 : 1). The aqueous layer was removed, and the organic layer was extracted with an additional 10 mL of 1 N NaOH. The combined aqueous layers were first back-extracted with 15 mL of hexane–EtOAc (1 : 1), acidified to pH 3 with 10 N H_2SO_4, and extracted with two 20-mL portions of dichloromethane. Currently SPE is predominant in the cleanup of IMI residues. One or more SPE columns packed with various types of sorbents have been developed and applied. The nonpolar material octadecyl (C_{18}) is the most commonly used sorbent in IMI analysis.[20,21,23,28,30] For a better cleanup the C_{18} sorbent is used in tandem with an SCX cartridge,[30] and an SAX cartridge.[23] Curran et al.[26] use solid-phase octyl (C_8) and aromatic sulfonic acid cartridges for imazaquin cleanup from soil. A sorbent of diol-bounded silica in a single column has also been used[19] for the cleanup of imazethapyr in soybeans. A GCB, Carbograph-1, has been evaluated[18] as a cleanup system in the soil analysis of six main IMIs.

4.3 Determination

All of the analytical methods for IMI residues are based on chromatography, mainly GC and HPLC, hyphenated and immunochemical techniques (see Tables 7–10).

4.3.1 Gas Chromatography

Two gas chromatographic methods have been developed for the determination of imazethapyr in soybeans[19] and soil.[26] Both make use of a capillary column with a bonded phase of low polarity (DB-5) and NPD to determine the residue. For GC analysis, the method[19] exploits the ready cyclization of the nonvolatile acid imazethapyr to the volatile imidazoisoindol-3,5-dione using an excess of dimethylaminopropylethylcarbodiimide hydrochloride.

Table 7 shows examples of IMI determination by GC, and Figure 2 presents the chromatograms of imazethapyr in soybeans analyzed by GC/NPD.

Table 6 Extraction and cleanup procedures used in the determination of IMI herbicides

Compound	Matrix	Extraction	Partition step	Cleanup	Ref.
m-Imazamethabenz *p*-Imazamethabenz *m*-, *p*-Imazamethabenz-methyl Imazapyr Imazaquin Imazethapyr	groundwater, lake water, river water	Carbograph-1	–	–	18
m-Imazamethabenz *p*-Imazamethabenz *m*-, *p*-Imazamethabenz-methyl Imazapyr Imazaquin Imazethapyr	soil	SCE	–	Carbograph-1	18
Imazapyr	surface runoff water	SPE C_{18}	–	–	17
Imazethapyr Imazmethapyr	soil	MAE	–	SPE C_{18}	28
Imazethapyr	soil	MAE	–	–	29
Imazethapyr Hydroxy metabolite Glucoside of the hydroxy metabolite	plant tissue	MAE	–	SCX and SPE C_{18}	30
Imazapyr	corn	acetone–water (25 : 75) acidified with HCl	–	Celite 545 AW and SPE C_{18}	21
Imazamethabenz-methyl	soil	different solvents, SPE, or SFE	–	–	22
Imazapyr	soil	ammonium hydrogen carbonate	–	SAX and SPE C_{18}	23
Imazaquin	soil	SFE	–	–	27
Imazapyr Imazethapyr	soil	sodium carbonate	dichloromethane	–	24
Imazethapyr	soybeans	methanol	NaOH 1 N : hexane-EtOAc (1 : 1) (50 : 50, v/v) and dichloromethane acidified at pH 3	SPE–diol	19
Imazapyr	soil	methanol–water (50 : 50, v/v)	–	–	25
Imazamethabenz	cereal, grain	methanol	hexane	SPE C_{18}	20
Imazaquin	soil	0.5 M NaOH, methanol	dichloromethane	tandem SPE C_8–SCX	26

SCX, strong cation exchange; SAX, strong anion exchange.

Table 7 Gas chromatographic determination of IMI herbicides

Compound	Matrix	Column, dimensions, temperature	Detector	DL	Ref.
Imazethapyr	soybeans	DB-5 30 m × 0.25 mm ID 100–275 °C	NPD	0.01 ppm	19
Imazethapyr	soil	DB-5 30 m × 0.25 mm ID 60–230 °C	NPD	5 ng g^{-1}	26

DL, detection limit; ID, internal diameter.

Table 8 HPLC determination of IMI herbicides

Compound	Matrix	Column	Mobile phase	Detector	λ (nm)	DL	Ref.
m-Imazamethabenz *p*-Imazamethabenz *m*-, *p*-Imazamethabenz-methyl Imazapyr Imazaquin Imazethapyr	groundwater, lake water, river water	Alltima LC-18	solvent A: water acidified with 30 mmol L^{-1} HCOOH solvent B: CH_3CN acidified with 5 mmol L^{-1} HCOOH	UV	240	30–39 ng L^{-1} 43–51 ng L^{-1} 55–67 ng L^{-1}	18
Imazapyr	soil	Nova-Pak	methanol–17 mM phosphoric acid (30 : 70, v/v)	UV	235	5 ng g^{-1}	23
Imazapyr Imazethapyr	soil	Supelcosil LC-ABZ and Hypersil APS	solvent A: phosphate buffer pH 3.1 solvent B: CH_3CN	UV	236 260	3 ng g^{-1}	24
Imazapyr	soil, water	Bondapack C_{18}	acetonitrile–ortho-phosphoric acid (pH 3) (30 : 70, v/v)	UV	250	50 ng g^{-1} 0.01 mg L^{-1}	25
Imazaquin	soil	C_8 reversed phase	water–methanol–acetic acid (60 : 40 : 4, v/v)	UV	240	5 ng g^{-1}	26

Table 9 Determination of IMI herbicides by hyphenated techniques

Compound	Matrix	Technique	MS conditions	DL	Ref.
m-Imazamethabenz *p*-Imazamethabenz *m*-, *p*-Imazamethabenz-methyl Imazapyr Imazaquin Imazethapyr	groundwater, river water, soil	HPLC/MS	ESIMS, TIC and SIM	4–7 ng L^{-1} 9–13 ng L^{-1} 0.1–0.05 ng g^{-1}	18
Imazethapyr Imazmethapyr	soil	HPLC/MS	ESIMS, ESIMS/MS	0.1 ng g^{-1}	28
Imazethapyr	soil	GC/MS	ECNCI	1 ng g^{-1}	29
Imazethapyr Hydroxy metabolite Glucoside of the hydroxy metabolite	plant tissue	HPLC/MS	ESIMS	<5 ng g^{-1}	30
Imazapyr	corn	GC/MS	ECNCI, EI	<5 ng g^{-1}	21
Imazamethabenz Imazamethabenz-methyl Imazapyr Imazaquin Imazethapyr Imazmethapyr	tap water, lake water, well water	HPLC/MS	ESIMS, ESIMS/MS	1 μ L^{-1}	31
Imazamethabenz-methyl	soil	GC/MS, HPLC/MS	EI, PCI, NCI PBMS	500 ng g^{-11}	22
Imazethapyr	soybeans	GC/MS	EI, SIM	100 ng g^{-1}	19

ECNCI, electron capture negative chemical ionization; PBMS, particle beam mass spectrometry; TIC, total ion current.

Table 10 Immunochemical techniques for the determination of IMI herbicides

Compound	Matrix	Technique	DL	Possible cross-reactions	Ref.
Imazapyr	surface runoff water	ELISA	$0.5\,\mu g\,L^{-1}$	–	17
Imazethapyr	tap water, groundwater	capillary IA	$5\,\mu g\,L^{-1}$	–	34
Imazaquin	soil	bioassay	$<0.5\,ng\,g^{-1}$	–	32
Imazaquin	soil	bioassay	$15\,ng\,g^{-1}$	–	26
Imazethapyr			$5\,ng\,g^{-1}$		
Imazaquin	agrochemicals	ELISA	$0.4\,ng\,L^{-1}$	The specificity of the assay system is generic for the IMI class of compounds, but no cross-reactivity was obtained with all nonimidazolinone-containing agrochemicals	33
Imazamethabenz	cereal grain	ELISA	$0.5\,ng\,L^{-1}$	Imazamethabenz-methyl	20

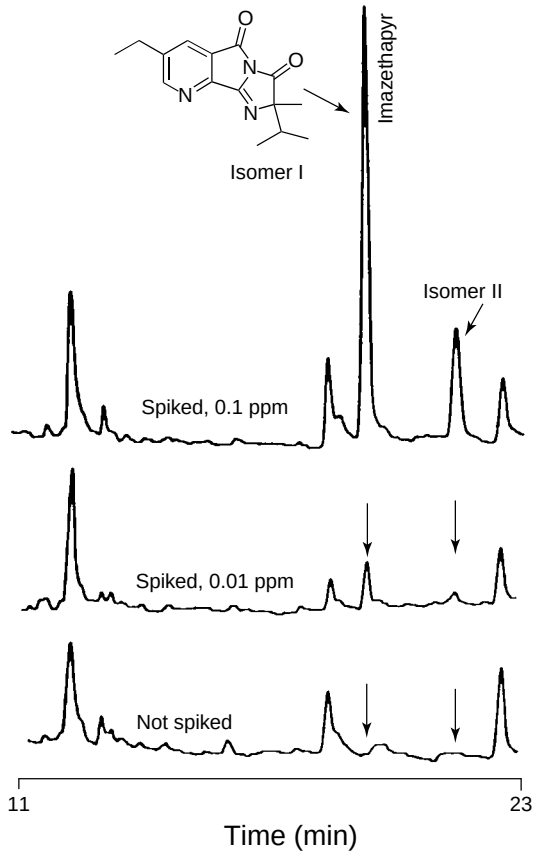

Figure 2 GC/NPD chromatograms of derivatized extracts of soybean meal from ground dry beans (40 mesh, Wiley mill) spiked at various levels with imazethapyr after cleanup on the diol-bonded columns. (Reprinted from *The Journal of AOAC International*, **76**, 377–381 (1993). Copyright, 1993, by AOAC International.)

4.3.2 High-performance Liquid Chromatography

HPLC is being extensively used in IMI analysis. The advantage of HPLC over GC is that the quantification of acidic IMI is carried out without need of derivatization.

The diverse methods used to determine IMI in environmental samples are shown in Table 8. Summarizing them, HPLC methods for the determination of residues employ RP chromatography with C_{18}[18,23–25] or C_8[26] columns. For increasing the selectivity in the analysis of imazapyr and imazethapyr from soil, Nováková[24] employs a column-switching method using two separation columns. Whether in isocratic[23,25,26] or gradient mode,[18,24] all mobile phases are formed from aqueous solutions acidified with different acids: formic,[18] phosphoric,[23–25] and acetic.[26] UV detection, usually operated in the range 235–260 nm, is the only detection method utilized in the analysis of IMI residues. Table 8 summarizes representative papers using this technique. HPLC/UV chromatograms obtained by analyzing groundwater, lake, and river water spiked with six IMI herbicides are shown in Figure 3.

4.3.3 Hyphenated Techniques

The successful combination of MS with GC and subsequently with liquid chromatography (LC) has enabled not only the determination of IMI pesticides but also the identification of their residues at trace levels. The three modes of GC/MS that have been used are electron impact (EI), positive chemical ionization (PCI), and negative chemical ionization (NCI). GC/MS in EI mode using selected ion monitoring (SIM) is used in the determination imazethapyr in soybeans,[19] imazamethabenz-methyl in soil,[22] and imazapyr in corn.[21] PCI and NCI are good alternative ionization methods which offer better selectivity and/or sensitivity than EI.[22] In the work proposed by Stout et al.,[29] the authors demonstrate that analyte detecting with the sensitivity and specificity of GC/ECNCI/MS reduces the required cleanup. Stout

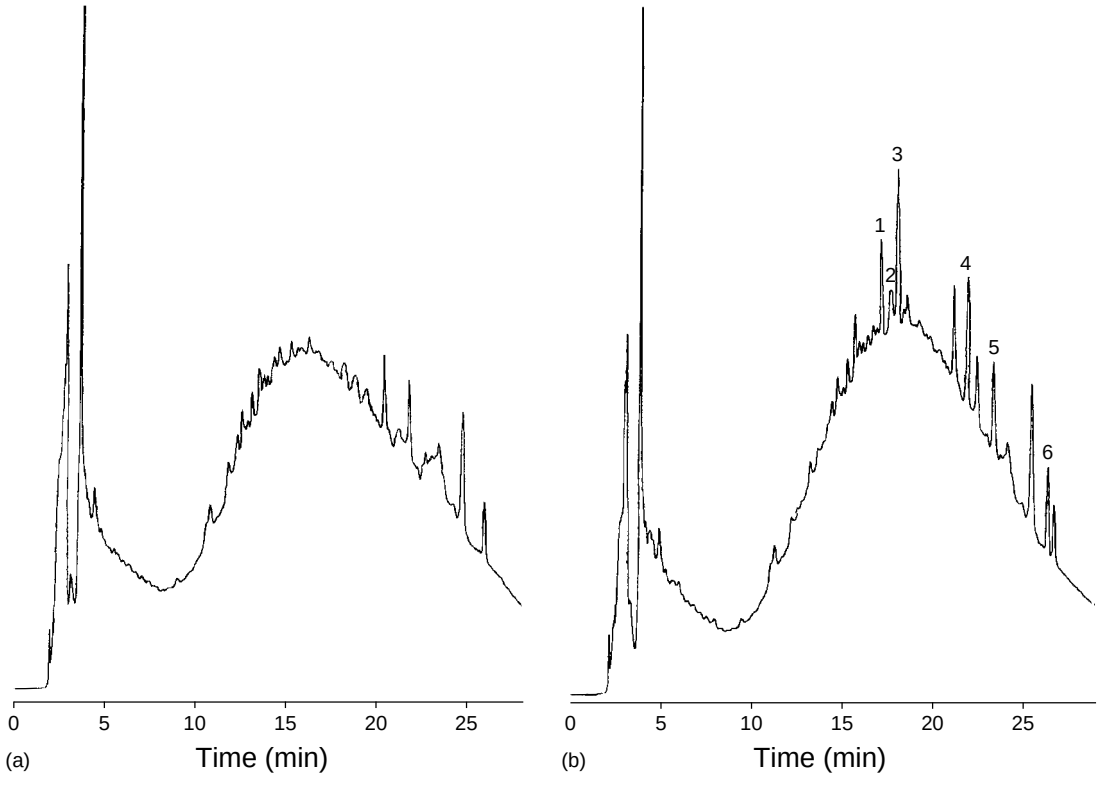

Figure 3 HPLC/UV chromatograms obtained by analyzing (a) 0.5 L of river water and (b) 0.5 L of river water, spiked at 1000 ng L^{-1} (*m*-imazamethabenz at 200 ng L^{-1}). Peaks: 1 = imazapyr, 2 = *m*-imazamethabenz, 3 = *p*-imazamethabenz, 4 = *m*-, *p*-imazamethabenz-methyl, 5 = imazethapyr, 6 = imazaquin. (Reproduced by permission of American Chemical Society from A. Laganà, G. Fago, A. Marino, *Anal. Chem.*, **70**, 121–130 (1998).)

et al.[21] wanted to evaluate the potential utility of GC with EI/MS on a less expensive GC/ECNCI/MS for the determination of imazapyr in corn.

HPLC coupled with MS provides the analyst with a powerful tool for IMI residue determination. Interfacing these devices is most commonly done by employing a particle beam (PB) or electrospray ionization (ESI) technique. HPLC/PB/MS has been applied for confirming imazamethabenz-methyl[22] in soil. By combining LC with electrospray ionization/mass spectrometry (ESIMS) or electrospray ionization tandem mass spectrometry (ESIMS/MS)[31] IMIs present in water can be monitored simultaneously at the 1 ppb level with only a simple filtration prior to analysis. The same authors developed a method for the rapid and direct determination of imazethapyr and imazmethapyr in soil by using LC/ESIMS and LC/ESIMS/MS;[28] an analytical procedure based on LC/ESIMS[30] is employed for targeted imazethapyr and its metabolites in a wide range of crop commodities. LC/ESIMS is a selective approach particularly suitable for confirming the presence of imazapyr, *m*-imazamethabenz, *p*-imazamethabenz, *m*-, *p*-imazamethabenz-methyl, imazethapyr, and imazaquin in two types of samples:[18] (1) for groundwater, lake water, and river water samples the analysis is made by HPLC/ESIMS in full-scan mode; (2) for soil samples LC/ESIMS under three ion SIM conditions is employed.

Table 9 shows examples of IMI herbicides determined by these hyphenated techniques and Figure 4 depicts chromatograms from LC/ESIMS of an unfortified, and a fortified soil with imazethapyr at the 1 ppb level.

4.3.4 Immunoassay

IAs are now being seen as useful supplements to classical chromatographic analytical systems. The near future will also see an integration of immunochemical and classical procedures.

IA provides rapid, sensitive, and cost-effective analysis for a variety of IMI residues. The main disadvantage is that only one compound at a time can be determined. The usefulness of these techniques is experienced during screening analysis when a large number of samples have to be analyzed in parallel for a simple analyte within a short time. Among the different IA procedures for determining the IMI residue, the most explored is enzyme-linked immuno-adsorbent assay (ELISA).

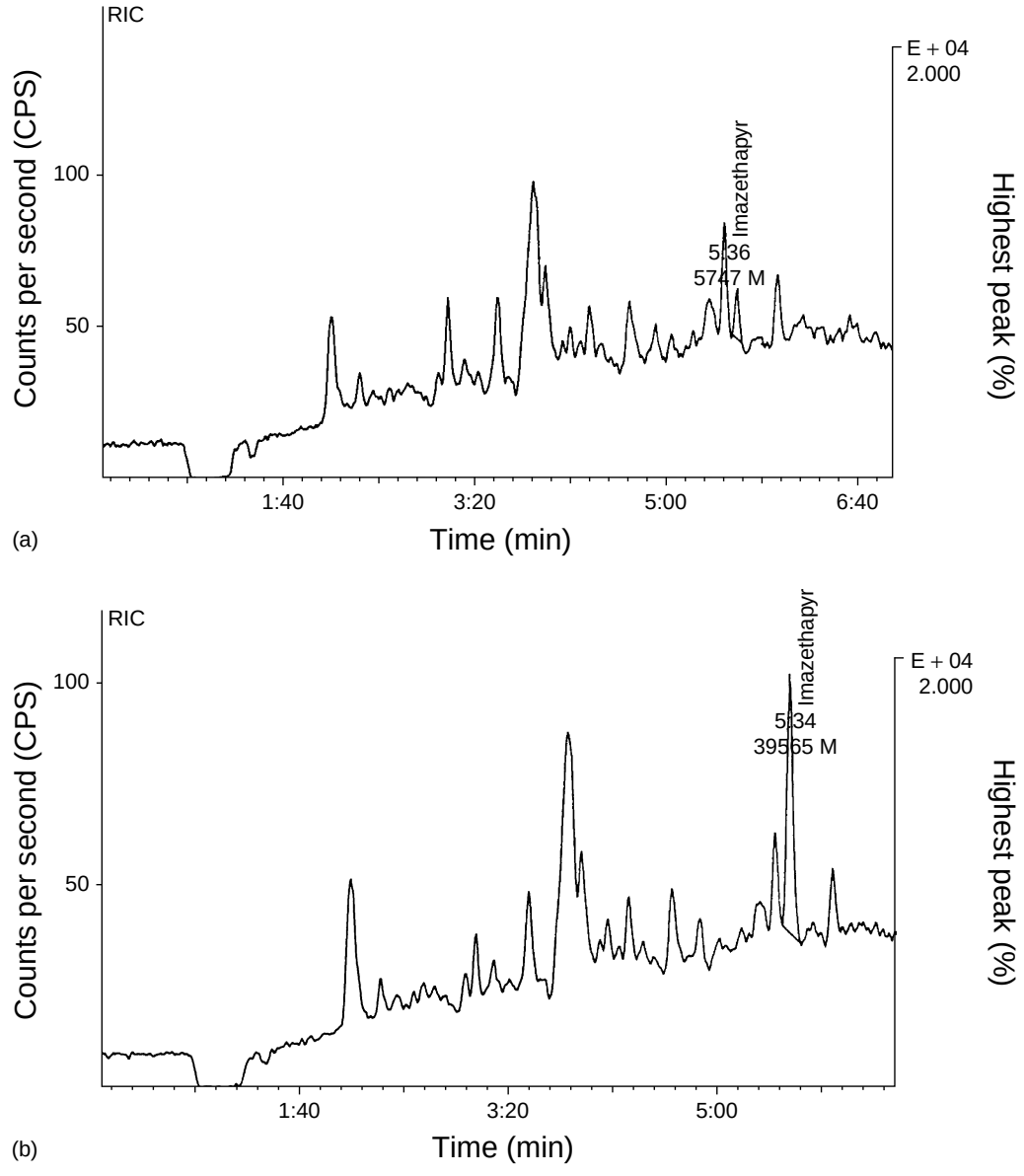

Figure 4 Chromatograms from LC/ESIMS of (a) blind Indiana soil with apparent imazethapyr at 0.17 ppb and (b) Indiana soil fortified with imazethapyr at 1 ppb. (Reprinted from *The Journal of AOAC International*, **80**, 426–432 (1997). Copyright, 1997, by AOAC International.)

ELISA is applied to the analysis of surface runoff water[17] and to the analysis of more complex matrices such as agrochemicals[33] and cereal grain.[20] In the analysis of imazaquin[33] and imazamethabenz,[20] erroneous results could be produced by matrix effects or the inability to differentiate between structurally similar compounds (cross-reactivity).

For tap and groundwater, Lee and Durst[34] have developed an assay based on sequential competitive binding of imazethapyr and liposomes for a limited number of antibody binding sites. A capillary tube with immobilized antibody was used as the immunoreactor column. Liposomes that entrap fluorescent molecules as the detectable label provide instantaneous, rather than time-dependent, enhancement, common with enzyme IAs. Bioassay techniques are applied for the determination of imazaquin and imazethapyr in soil.[26,32] In the method proposed by O'Bryan et al.[32] three corn-root bioassays were evaluated for detecting imazaquin in soil. Two techniques, one which utilized a cone-shaped tube as a growth container and another a petri dish, were compared to a method that utilized a thin layer of soil between two 20×20 cm glass plates. Representative examples are summarized in Table 10.

5 ARYLOXYPHENOXYPROPIONIC ACIDS/ESTERS

5.1 Chemical and Physical Properties

Since the late 1980s, many heterocyclic oxyphenoxypropionic acid derivatives have been prepared as a new class of selective herbicides for the control of graminaceous weeds in broad-leaved crops and tolerant cereals. Esters of aryloxyphenoxypropionic acids are a new series of highly selective PoE herbicides, often termed ''phenoxyphenoxys''. In the treated plants they decompose fairly rapidly, yielding the corresponding free acids as the main metabolites.

The chemical structures, names, and chemical abstracts registry numbers of aryloxyphenoxypropionate (ArPP) herbicides are shown in Table 11. Tables 12–14 enumerate the chemical, physical, and other properties of ArPP herbicides.

Considerable progress has been made towards understanding the mode of action of the ArPP herbicides. Earlier studies suggested that these compounds disabled de novo fatty acid biosynthesis, but no specific site of action was identified.[35] Recently, it has been demonstrated that they specifically inhibit acetyl-CoA carboxylase in susceptible plants. This enzyme catalyzes the adenosine triphosphate (ATP)-dependent formation of malonyl-CoA from acetyl-CoA and bicarbonate, that is the first committed step for fatty acid synthesis de novo.

In spite of their different structures they show the same symptomatology and selectivity of herbicidal action. In fact, they are PoE herbicides that are used for the control of annual and perennial grasses in a large variety of broad-leaved crop plants.[36]

ArPP esters form colorless crystals or odorless solids, except for fluazifop-butyl which is a pale straw-colored liquid. This class of compounds is relatively stable in acidic media (pH 3), and in some cases in neutral conditions, undergoing slow hydrolysis to the free acid between 30 and more than 1000 days. This process is much faster in alkaline media (pH 9), ranging from 2.2 to 60 h. The pK_a values of the free acids are between 2.9 and 3.4.

ArPPs are selective herbicides with a systemic action. They are absorbed principally by the leaves, and in some cases by the roots, to undergo rapid transformation into the acidic form which is then translocated throughout the plant via the xylem or phloem system, accumulating in the meristematic tissues. All these compounds can be regarded as heterocyclic homologs of the wild oat herbicide diclofop, in which the terminal dichlorophenyl ring is replaced by a heterocycle, such as pyridine, quinoline, quinoxaline, benzothiazole, or benzoxazole. ArPPs posses a chiral center and exist as a mixture of two optical isomers. The (*R*) enantiomer of fluazifop-butyl was the second case of a herbicide in which the introduction of the resolved form largely replaced the racemate. Although biological activity resides exclusively with the (*R*) enantiomer in PoE applications, the (*R*) and (*S*) enantiomers show equivalent activity when applied in PrE. This has been shown to be due to microbial conversion of the (*S*) to the (*R*) enantiomer in the soil.

The main uses are on soybeans, sugar beet, and cotton. Particular uses include control of volunteer cereals and other grass weeds in wheat, barley, rye, legumes, oil seed rape, sunflowers, fodder beet, flax, alfalfa, peanuts, beetroot, lettuce, spinach, potatoes, cucumber, peas, tomatoes, fennel, and strawberries.

With so many products already on the market or becoming available for the same use on the same crops in a relatively short period of time, severe competition is inevitable. Some efforts have been made to search for new applications.

Thus fenoxaprop-ethyl is only marginally selective in wheat, but addition of the safener fenchlorazole-ethyl has allowed the use of the herbicide on all wheat varieties except durum wheat. Racemic fenoxaprop-ethyl is formulated with the safener in a 4:1 ratio, whereas the active (*R*) isomer utilizes a 2:1 ratio. The safener does not influence either foliar uptake of the herbicide by wheat plants, or the degree of inhibition of acetyl-CoA carboxylase. However, levels of fenoxaprop-ethyl and of its herbicidally active free acid in shoot tissues of wheat decrease more rapidly when applied with fenoxaprop-ethyl; at the same time the formation of inactive metabolites of the herbicide increases.

The ArPPs have undergone many tests to determine their toxicity to mammals in order to prevent serious hazard to farmers. Test conducted on rats have shown that in most cases the ArPP is metabolized into the free acid and then excreted within 3–7 days in urine and feces.[10] The toxicity class according to WHO is III for all compounds except for haloxyfop, which is II.[10]

5.2 Analytical Methods for Aryloxyphenoxypropionate Analysis

5.2.1 Extraction

Table 15 summarizes extraction and cleanup procedures used in the determination of ArPP herbicides.

Currently SPE is predominant in analytical methods used for ArPP herbicidal residues in water matrices. Conventional LLE with dichloromethane has been used in the monitoring of diclofop and haloxyfop,[37] and diclofop-methyl[38] in groundwater.

Table 11 Chemical structures, IUPAC-recommended names[10] and Chemical Abstracts Service registry numbers of ArPP herbicides

Chemical structure	Name	CAS
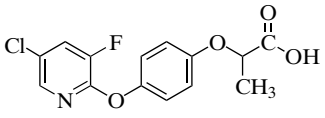	**clodinafop** (*R*)-2-[4-(5-chloro-3-fluoro-2-pyridyloxy)phenoxy]propionic acid	[114420-56-3]
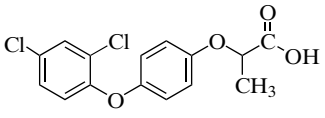	**diclofop** (*R,S*)-2-[4-(2,4-dichlorophenoxy) phenoxy]propionic acid	[40843-25-2]
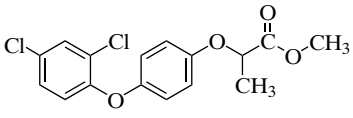	**diclofop-methyl** methyl-(*R,S*)-2-[4-(2,4-dichlorophenoxy) phenoxy]propionic acid	[51338-27-3] (*R*) [71283-65-3] (*S*) [75021-72-6]
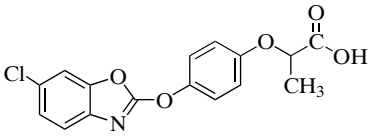	**fenoxaprop** (*R*)-2-[4-(6-chloro-1,3-benzoxazol-2-yloxy)phenoxy]propionic acid	[113158-40-0]
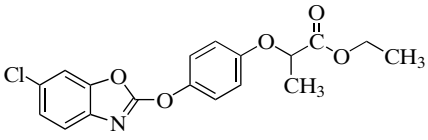	**fenoxaprop-ethyl** ethyl-(*R*)-2-[4-(6-chloro-1,3-benzoxazol-2-yloxy)phenoxy]propionate	[71283-80-2]
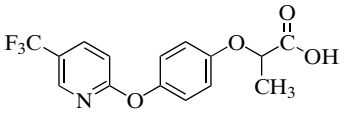	**fluazifop** (*R,S*)-2-[4-(5-trifluoromethyl-2-pyridyloxy)phenoxy]propionic acid	[69335-91-7]
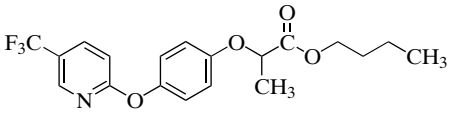	**fluazifop-butyl** butyl-(*R,S*)-2-[4-(5-trifluoromethyl-2-pyridyloxy)phenoxy]propionate	[69806-50-4]
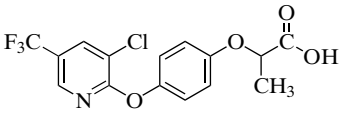	**haloxyfop** (*R,S*)-2-[4-(3-chloro-5-trifluoromethyl-2-pyridyloxy)phenoxy]propionic acid	[69806-34-4]
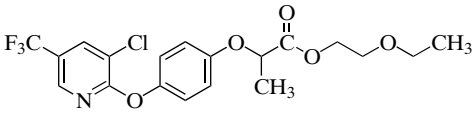	**haloxyfop-ethoxyethyl** (*R,S*)-2-ethoxyethyl-2-[4-(3-chloro-5-trifluoromethyl-2-pyridyloxy)phenoxy] propionic acid	[87237-48-7]
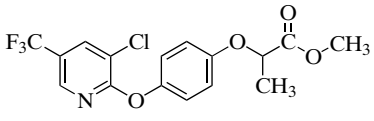	**haloxyfop-methyl** methyl-(*R,S*)-2-[4-(3-chloro-5-trifluoromethyl-2-pyridyloxy)phenoxy]propionate	[69806-40-2]
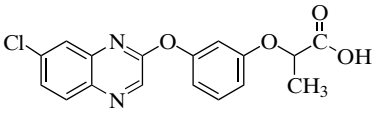	**quizalofop** (*R,S*)-2-[4-(6-chloroquinoxalin-2-yloxy)phenoxy]propionic acid	[76578-12-6]
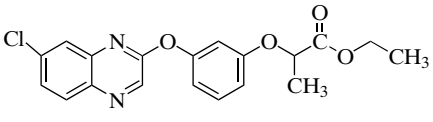	**quizalofop-ethyl** ethyl (*R,S*)-2-[4-(6-chloroquinoxalin-2-yloxy)phenoxy]propionate	[100760-10-9]

Table 12 Chemical and physical properties of ArPP herbicides

Property	Clodinafop	Diclofop	Fenoxaprop	Fluazifop	Haloxyfop	Quizalofop
Empirical formula	$C_{14}H_{11}ClFNO_4$	$C_{15}H_{12}Cl_2O_4$	$C_{16}H_{12}ClNO_5$	$C_{15}H_{12}F_3NO_4$	$C_{15}H_{11}ClF_3NO_4$	$C_{17}H_{13}ClN_2O_4$
Molecular weight	311.7	327.2	333.7	327.3	361.7	344.8
Physical form	–	yellowish-white solid	light beige, weakly pungent, fine powder	–	colorless crystals	–
Melting point (°C)	–	118–122	155–161	–	107–108	–
Vapor pressure (mPa)	–	3.1×10^{-6} (20 °C) 9.7×10^{-6} (25 °C)	0.18 (20 °C)	–	$<1.33 \times 10^{-3}$ (25 °C)	–
Solubility in water at 20 °C ($g\,L^{-1}$)	–	0.453 at pH 5 122.7 at pH 7 127.4 at pH 9	0.27 at pH 5.1 61 at pH 7	–	43.4 at pH 2.6 (25 °C) 1.59 at pH 5 6.98 at pH 9	–
pK_a	–	3.43	–	3.2	2.9	–

Property	Clodinafop-propargyl	Diclofop-methyl	Fenoxaprop-ethyl	Fluazifop-butyl	Haloxyfop-ethoxyethyl	Haloxyfop-methyl	Quizalofop-ethyl
Empirical formula	$C_{17}H_{13}ClFNO_4$	$C_{16}H_{14}Cl_2O_4$	$C_{18}H_{16}ClNO_5$	$C_{19}H_{20}F_3NO_4$	$C_{19}H_{19}ClF_3NO_5$	$C_{16}H_{13}ClF_3NO_4$	$C_{19}H_{17}ClN_2O_4$
Molecular weight	349.8	341.2	361.8	383.4	433.8	375.7	372.8
Physical form	colorless crystals	colorless crystals	white odorless solid	pale straw-colored liquid	colorless crystals	–	colorless crystals
Melting point (°C)	59.5	39–41	89–91	13	58–61	55–57	91.7–92.1
Vapor pressure (mPa)	3.19×10^{-3} (25 °C)	0.25 (20 °C)	5.3×10^{-4} (20 °C)	55×10^{-3} (20 °C)	1.64×10^{-5} (20 °C)	0.80 (25 °C)	8.65×10^{-4} (20 °C)
Solubility:	(25 °C)	(20 °C)	(20 °C)	(20 °C)	(20 °C)	(25 °C)	(20 °C)
water ($mg\,L^{-1}$)	4	0.8	0.7 at pH 7.8	1 at pH 6.5	0.58	9.3	0.3
acetone ($g\,L^{-1}$)	880	>500	200	miscible	>1000	$3.5\,kg\,kg^{-1}$	111
toluene	690	>500	200	–	>1000	–	–
n-hexane	8.6×10^{-3}	50	–	miscible	44	–	2.6
methanol	–	120	–	miscible	233	–	–
ethanol	97	–	24	–	–	–	9

Different SPE methods have been developed using C_{18} bonded-silica cartridges.[39–45] On-line trace enrichment for diclofop-methyl and fluazifop-butyl screening in river water is obtained with a cartridge packed with a polymeric RP material.[46] For the extraction of fenoxaprop in groundwater, Dupas et al.[47] suggest an on-line device that utilizes a precolumn packed with the apolar PLRP-S. The same material is used by Wuchner and Grob.[48] In such work, SFE is employed for the elution of diclofop-methyl, after enrichment on a precolumn packed with a polymeric phase. For the simultaneous extraction of clodinafop, diclofop, fenoxaprop, fluazifop, haloxyfop and quizalofop from drinking, spring, and groundwater, Laganà et al.[49] employ three different SPE systems:

- a disposable SPE cartridge packed with Carbograph-1;
- a LiChrolut-EN cartridge, an ethylvinylbenzene–divinylbenzene (DVB) resin;
- an SPE disk, made of polystyrene–divinylbenzene (PSDVB).

Chromatograms obtained by analyzing 1 L of a groundwater sample spiked with six ArPPs, using two different SPE systems, are reported in Figure 5.

For solid samples, extraction with a water-miscible solvent is necessary. When the analyzed sample is food, the subsample can be homogenized with acetone,[50–53], acetone–ethanol–2 mol L^{-1} HCl (2 : 1 : 1, by vol),[54] 0.1 mol L^{-1} NaOH,[55] 0.2 mol L^{-1} NaOH in methanol,[56]

Table 13 Stability of ArPP herbicides

Parameter	Clodinafop-propargyl	Diclofop-methyl	Fenoxaprop-ethyl	Fluazifop-butyl	Haloxyfop-ethoxyethyl	Haloxyfop-methyl	Quizalofop-ethyl
Stability	–	stable to light	90 days at 50 °C	3 years at 25 °C 6 months at 37 °C	–	–	90 days at 40 °C
Other observations	relatively stable in acidic media at 50 °C, hydrolyzes in alkaline media	–	not sensitive to light, decomposed by acids and alkalis	–	hydrolyzes to haloxyfop under acidic and alkaline conditions	–	DT_{50} 10–30 days to light
DT_{50} in water:	(25 °C)	(25 °C)	(20 °C)	–	(25 °C)		–
ester	64 h at pH 7 2.2 h at pH 9	363 days at pH 5 31.7 days at pH 7 0.52 days at pH 9	>1000 days at pH 5 100 days at pH 7 2.4 days at pH 9	reasonably stable in acidic and neutral conditions, rapidly hydrolyzed at pH 9	33 days at pH 5 5 days at pH 7 few hours at pH 9	–	stable at pH 3–7
acid	–	–	–	–	–	78 days at pH 5 73 days at pH 7 51 days at pH 9	
DT_{50} in soil:							
ester	<2 h to free acid	to free acid	–	<1 week to free acid	>1 day at 20 °C	<1 day to free acid	<1 day to free acid
acid	5–20 days	1–57 days 30–281 (DT_{90})	–	<3 weeks	55–100 days at pH 7, 20–25 °C microbially degraded		–

Table 14 Applications and environmental uptake of ArPP herbicides

Parameter	Clodinafop	Diclofop	Fenoxaprop	Fluazifop	Haloxyfop	Quizalofop
Acute oral LD_{50} rat (mg/kg body weight)	1829	481–693	3150–4000	3600	337	1480
Acute dermal LD_{50} rabbit (mg/kg body weight)	nonirritating	–	–	>2420	>5000	–
WHO toxicity class[a]	III	III	–	III	II	III
EC risk[b]	R22, R43	Xn, R22, R43	–	–	Xn, R22	–
Fate in						
animals	acid	excreted	–	acid	acid	excreted
plants	acid	acid	–	acid	acid	unchanged
Uses[c]	U1	U2	U3	U4	U5	U6

[a] WHO toxicity class: Ia, extremely hazardous; Ib, highly hazardous; II, moderately hazardous; III, slightly hazardous.
[b] EC Risk: Xn, sensitizing; R22, harmful if swallowed; R43, may cause sensitization by skin contact.
[c] U1: PoE control of annual grasses including *Avena, Lolium, Setaria, Phalaris* and *Alopecurus*, in cereals. U2: PoE control of wild oats, wild millets and annual grass weeds in wheat, barley, rye, red fescue and broad-leaved crops such as soybeans, sugar, beet, fodder beet, flax, legumes, oilseed rape, sunflowers, clover, alfalfa, peanuts, brassicas, carrots, celery, beetroot, parsnips, lettuce, spinach, potatoes, cucumbers, peas, beans, tomatoes, fennel, alliums, herbs, etc. U3: PoE control of annual and perennial grass weeds in potatoes, beans, soybeans, beets, vegetables, peanuts, flax, oilseed rape and cotton. U4: PoE control of annual and perennial grass weeds in broad-leaved crops. U5: PrE and PoE control of annual and perennial grasses in sugar beet, fodder beet, oilseed rape, potatoes, leaf vegetables, onions, flax, sunflowers, soybeans, vines, and strawberries. U6: PoE control of annual and perennial grass weeds in potatoes, soybeans, sugar beet, peanuts, oilseed rape, sunflowers, vegetables, cotton and flax.

acetone–light petroleum (50:50, v/v) and Celite, or 4 mol L^{-1} HCl–(acetone–light petroleum, 2:1) (10:60, v/v).[57] An SFE procedure is described to isolate fluazifop-butyl and its major metabolite, fluazifop, directly from onions without any further cleanup procedure.[58] To reduce the requirement for solvent and time, a column extraction method has been developed by Kadenczki et al.[59] and Laganà et al.[60] In the work of Kadenczki et al.[59] homogeneous pulp, prepared from fruit and vegetables of different content with or without additional

Table 15 Extraction and cleanup procedures used in the determination of ArPP herbicides

Compound	Matrix	Extraction	Partition step	Cleanup	Ref.
Clodinafop Diclofop Fenoxaprop Fluazifop Haloxyfop Quizalofop	drinking water, spring water, groundwater	–	–	Carbograph-1 cartridge, LiChrolut-EN cartridge, PV-DVB Empore disk	49
Diclofop Haloxyfop Fluazifop	soil	SCE	–	Carbograph-1 cartridge	60
Diclofop-methyl Fenoxaprop Haloxyfop-ethoxyethyl Quizalofop-ethyl	fruit and vegetables	sonication with acetone	dichloromethane	GPC (Envirosep S-X3)	50
Diclofop-methyl Fluazifop-butyl	river water	PRP-1 cartridge	–	–	46
Diclofop-methyl	soil	sonication with water and methanol	–	SPE C_{18} disk	62
Fenoxaprop	groundwater	PLRP-S	–	–	47
Fenoxaprop Fenoxaprop-ethyl	drinking water	SPE C_{18}	–	–	44
Diclofop Haloxyfop	groundwater	dichloromethane	–	–	37
Diclofop-methyl	river water, wastewater	PLRP-S column + SFE	–	–	48
Fluazifop Fluazifop-butyl Haloxyfop	drinking water, surface water, sewage water	SPE C_{18}	–	–	43
Fluazifop Fluazifop-butyl Haloxyfop	drinking water, groundwater	SPE C_{18}	–	–	39
Fluazifop	drinking water	SPE C_{18}	–	–	40
Diclofop Fenoxaprop Fluazifop Haloxyfop Quizalofop	tap water, groundwater	SPE C_{18}	–	–	41
Diclofop-methyl Fenoxaprop-ethyl	fruit, vegetable matrix	acetone	petroleum ether and dichloromethane	–	51
Fluazifop Fluazifop-butyl	groundwater	SPE C_{18}	–	–	42
Diclofop-methyl	onions, red radishes	acetone	dichloromethane	GPC (Bio-Beads S-X3), silica gel column	52
Fluazifop Fluazifop-butyl	onions	SFE	–	–	58
Fenoxaprop Fenoxaprop-ethyl	soil	0.1 M HCl–methanol (50 : 50, v/v), ethyl acetate	dichloromethane	Florisil or Al_2O_3	65
Fluazifop-butyl Haloxyfop	fruit, vegetables	Florisil	–	–	59
Fluazifop Fluazifop-butyl	soil	soxhlet with acetone	–	SPE-phenyl SPE-cyanopropyl	9
Fluazifop Fluazifop-butyl	soil	soxhlet with acetone	dichloromethane	GPC (Bio-Beads S-X3)	61
Diclofop-methyl	groundwater	dichloromethane	–	–	38
Fluazifop Fluazifop-butyl	soil, leaves, cotton, seed, peanuts	sonication with water and acetone or 6 mol L^{-1} HCl and acetone	chloroform, light petroleum	Al_2O_3 and Florisil	63

Table 15 (*continued*)

Compound	Matrix	Extraction	Partition step	Cleanup	Ref.
Diclofop-methyl	soil	acetone–water–acetic acid (80:19:1, v/v), reextract by sonication with 0.1 mol L^{-1} KOH	diazomethane	–	64
Diclofop Diclofop-methyl	soil, crops	sonication with acetone–light petroleum (50:50, v/v) and Celite, or 4 M HCl: (acetone:light petroleum, 2:1) (10 mL + 60 mL)	chloroform, light petroleum	Al_2O_3, Ag–Al_2O_3, Florisil	57
Quizalofop-ethyl	fruit, vegetables	acetonitrile	–	–	67
Fluazifop	agricultural products	acetone	–	GPC (Bio-Beads S-X3)	53
Fluazifop Fluazifop-butyl	groundwater	SPE C_{18}	–	–	45
Fluazifop Fluazifop-butyl Haloxyfop Haloxyfop-ethoxyethyl Quizalofop Quizalofop-ethyl	fruit, vegetables	acetone–ethanol–2 M HCl (2:1:1, v/v)	hexane–diethyl ether (50:50, v/v), and diethyl ether	silica gel	54
Fluazifop	soybeans, soybean oil	0.2 N NaOH in methanol	dichloromethane	–	56
Fluazifop Fluazifop-butyl	crops, soil	0.1 M NaOH	hexane, dichloromethane pH < 2	Florisil	55

[a] PRP-1, polymeric RP material; PLRP-S, copolymer styrene divinylbenzene.

water, is adsorbed on the surface of activated florisil to obtain a free-flowing powder, which is extracted in a glass column with ethyl acetate or dichloromethane–acetone (90:10, v/v). The method proposed by Laganà et al.,[60] for sample preparation involves a simple treatment of the soil sample via isolation and trace enrichment of fluazifop, haloxyfop, and diclofop by SCE.

Several methods have been proposed for the extraction of ArPPs from soil samples. These methods are as follows: Soxhlet with acetone;[9,61] sonication with water and methanol,[62] acetone–light petroleum (50:50, v/v),[57] 4 mol L^{-1} HCl and acetone–light petroleum (2:1) (10:60, v/v),[57] water and acetone or 6 mol L^{-1} HCl and acetone;[63] and batch extraction with acetone–water–acetic acid (80:19:1, v/v) followed by re-extraction by sonication with 0.1 mol L^{-1} KOH;[64] mechanical shaker with 0.1 mol L^{-1} NaOH,[55] 0.1 mol L^{-1} HCl–methanol (50:50, v/v) and ethyl acetate,[65] dichloromethane and then 0.1 mol L^{-1} NaOH–methanol (50:50, v/v).[66]

5.2.2 Cleanup

In the majority of instances, the initial extract contains a too high level of coextracted interfering compounds to allow direct analysis, and some form of cleanup is required.

In the LLP step, the extraction efficiency can be enhanced by adding sodium chloride to salt out the analytes[50,52–55,61] or sodium sulfate.[57] Dichloromethane is most commonly used in LLP.[52,55,56,65] Solvent mixtures such as petroleum ether and dichloromethane,[51] hexane–diethyl ether (50:50, v/v),[54] and chloroform and light petroleum[57,63] are also used. For ArPP analysis, the most commonly used cleanup techniques are SPE packed with Florisil, alumina, silica gel, GCB, phenyl and cyanopropyl bonded phases, SPE-disk and size exclusion (gel permeation).

An alternative sorbent that is especially effective for cleanup of acidic ArPP herbicides is Carbograph-1,[49,60] composed of nonporous spherical graphite particles. The advantage of using the Carbograph-1 cartridge was that extraction, concentration, and fractionation into classes were possible simultaneously (the acid analytes were separated from the neutral to basic ones by differential elution). The presence of some active centers bearing a positive charge enables Carbograph-1 to behave as both a nonspecific and an anion-exchange sorbent. It follows that anionic ArPP herbicides were specifically adsorbed

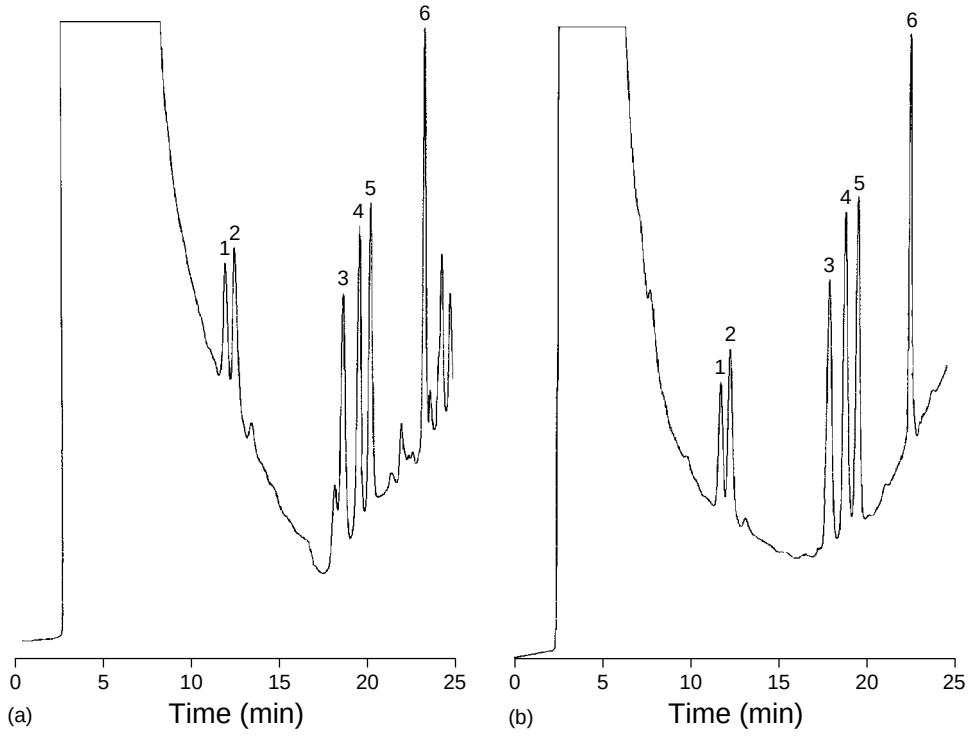

Figure 5 HPLC/DAD chromatograms obtained by analyzing 1 L of a groundwater sample spiked with ArPP at the individual level of 100–200 ng L^{-1} by different procedures involving the use of (a) the LiChrolut-EN extraction cartridge and (b) the Carbograph-1 cartridge. Peaks: 1 = fluazifop; 2 = clodinafop; 3 = quizalofop; 4 = fenoxaprop; 5 = haloxyfop; 6 = diclofop. (Reprinted from *J. Chromatogr.*, **796**, A. Laganà, G. Fago, A. Marino, 'Determination of Aryloxyphenoxypropionic Acid Herbicides in Water Using Different SPE Procedures and LC/DAD', 309–318, Copyright 1998, with permission from Elsevier Science.)

on the Carbograph-1 surface via electrostatic forces and that they could be desorbed simply by adding a displacing agent (such as formic acid) to the organic solution. Carbograph-1 behaves like conventional exchangers if the stepwise elution of acidic compounds can be achieved on the basis of the acid strands, by appropriate selection of the displacing agents.

Open-column chromatography using a column packed with Florisil is used to remove the chromatographic interference in the determination of fluazifop-butyl and fluazifop in potatoes, soybeans, and soil samples.[55] The six principal ArPP herbicides in fruit and vegetable samples are subjected to activated silica gel column cleanup.[54] Silica gel is also used for cleanup of fluazifop-butyl and fluazifop in soil.[66] Fenoxaprop and fenoxaprop-ethyl in different soils are purified on a Florisil or alumina cartridge.[65] Diclofop-methyl and its metabolite diclofop were extracted and derivatized, and then the products were purified on a chromatographic column containing alumina, silver–alumina, and Florisil.[57] The bromo derivative of fluazifop-butyl and the pentafluorobenzyl derivative of fluazifop are purified on a chromatographic column containing alumina and Florisil.[63]

A C_{18} extraction disk, used in off-line mode, has been evaluated by Bao et al.[62] for the cleanup of diclofop-methyl. Higher flow rates can be applied with an SPE disk and the extraction time is much shorter.

The large difference in polarities between fluazifop-butyl and fluazifop causes difficulties in the application of the sorbent extraction technology for the simultaneous determination of both compounds from soil extracts. Therefore, in the handling of soil extracts containing fluazifop-butyl and fluazifop, Zanco et al.[9] have developed a combination of a phenyl phase and cyanopropyl-phase cartridges for the cleanup operation.

A method used for the cleanup of ArPP in complex matrices is GPC. It has been used for cleanup in the analysis of diclofop-methyl, fenoxaprop, haloxyfop-ethoxyethyl, quizalofop-ethyl,[50] diclofop-methyl,[52] fluazifop, and fluazifop-butyl.[61] Separation is performed using Bio Beads SX-3, a polystyrene type gel, using solvents such as a mixture of cyclohexane and ethyl acetate (50 : 50, v/v),[52] or cyclohexane and acetone (50 : 50, v/v).[53] A liquid chromatographic step on silica gel was therefore inserted between the GPC and the GC step to filter out polar products.[52] Figure 6 shows a typical GC/ECD chromatogram of a fruit

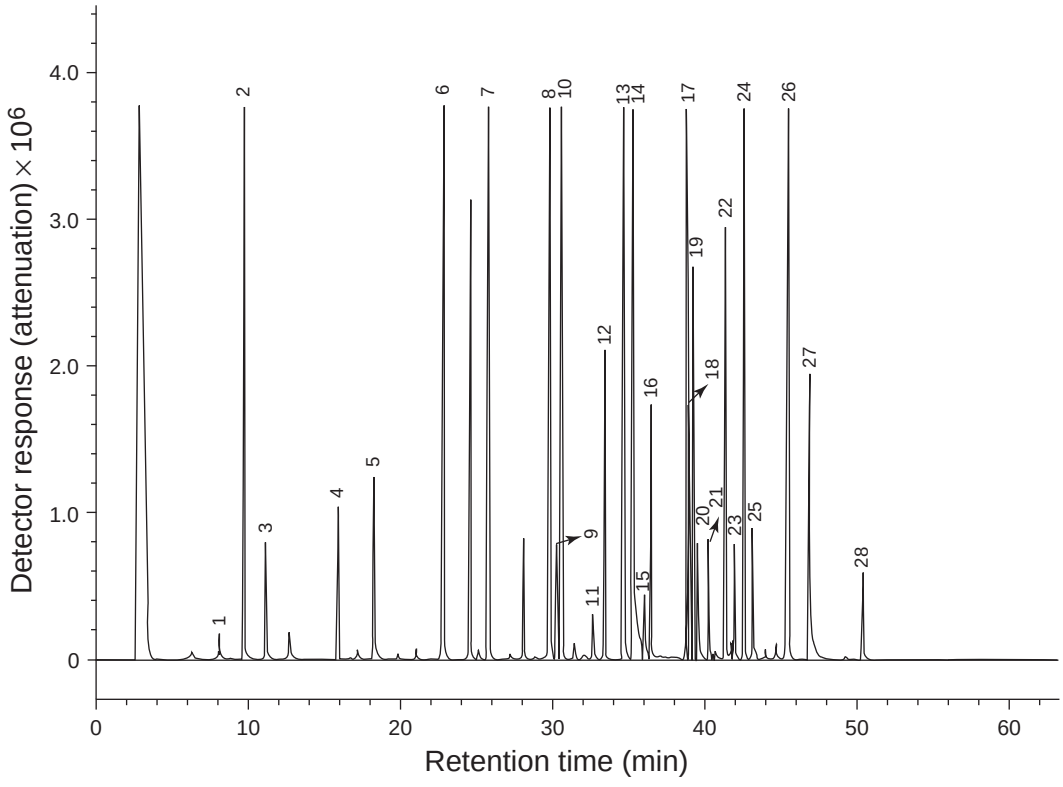

Figure 6 GC/ECD chromatogram of melon extract fortified with haloxyfop-ethoxyethyl (0.33 mg kg^{-1}), fenoxaprop (0.28 mg kg^{-1}), and quizalofop-ethyl (0.29 mg kg^{-1}). (Reproduced from *J. Chromatogr. A*, **782**, A. Gelsomino, B. Petrovicová, S. Tiburtini, E. Magnani, M. Felici, 'Multiresidue Analysis of Pesticides in Fruit and Vegetables by GPC Followed by GC with Electron Capture and Mass Spectrometric Detection', 105–122, Copyright 1997, with permission of Elsevier Science.)

extract containing haloxyfop-ethoxyethyl, fenoxaprop, and quizalofop-ethyl, cleaned up by gel permeation.

5.3 Determination

Chromatographic methods are most widely used for analytical separation, identification, and quantification of ArPP residues in different matrices.

5.3.1 Gas Chromatography

Table 16 summarizes the column, detector, and DL currently employed in routine GC ArPP residue analysis, and Figure 7 shows chromatograms of diclofop-methyl residues in soil and crops. Capillary GC in combination with selective detectors, mainly ECD[37,41,54,57,59,67] and NPD,[54,55] are the most common techniques for ArPP determination. The use of other detectors, such as FID[48] and AED[52] have been reported in only one case. The GC/AED is one type of spectrochemical method, and the incorporation of a microwave-induced plasma generator and a photodiode array has demonstrated its accuracy, reproducibility and simplicity.

The data on the elemental composition of pesticides available from AED chromatograms makes the information content much higher than that obtained from conventional GC detectors. Several stationary bonded phases have been used in the determination of ArPP in different samples, including different polarities: SE-54,[41,55] HP-1,[52] DB-5,[59] HP-5,[54] and SE-52[45] are nonpolar, and 65% dimethyl–35% diphenylpolysiloxane,[37] DB-1701,[48] OV-17,[57,63] and SPB-608[50] are polar.

The acidic ArPPs in the extract were derivatized with TFE,[37] pentafluorobenzyl bromide,[39,42,43,47] methyl, ethyl, and butyl chloroformate,[41] diazomethane,[55,58,66] and bromine.[63]

Figure 8 is a chromatogram of a groundwater sample spiked with fluazifop and haloxyfop, derivatized to their methyl esters.

5.3.2 High-performance Liquid Chromatography

HPLC is used extensively in ArPP residue determination. HPLC methods employ RP chromatography with C_{18}[40,44–47,49,58,65,66] or C_8[9] columns. For the determination of fluazifop the authors utilize a phenyl-bonded phase,[9] and cyano-bonded phase[40] column.

Table 16 Gas chromatographic determination of ArPP herbicides

Compound	Matrix	Column, dimension, temperature	Detector	DL	Ref.
Diclofop-methyl Fenoxaprop Haloxyfop-ethoxyethyl Quizalofop-ethyl	fruit and vegetables	SBP-608, 30 m × 0.25 mm ID, 50–280 °C	ECD	<10 ng g^{-1}	50
Diclofop Haloxyfop	groundwater	cross-bonded 65% dimethyl–35% diphenyl polysiloxane, 30 m × 0.53 mm ID, 130–260 °C	ECD	500 ng L^{-1}	37
Diclofop-methyl	river water, wastewater	DB-1701, 15 m × 0.25 mm ID, 60–285 °C	FID	≤10 ng L^{-1}	48
Diclofop Fenoxaprop Fluazifop Haloxyfop Quizalofop	tap water, groundwater	SE-54, 25 m × 0.20 mm ID, 100–260 °C	ECD	<100 ng L^{-1}	41
Diclofop-methyl	onions, red radishes	HP-1, 25 m × 0.32 mm ID, 90–240 °C	AED	100 ng g^{-1}	52
Fluazifop-butyl Haloxyfop	fruit, vegetables	DB-5, 60 m × 0.245 mm ID, 85–300 °C	ECD	10 ng g^{-1} 40 ng g^{-1}	59
Fluazifop Fluazifop-butyl	soil, leaves, cotton, seed, peanuts	2% OV-17, 1.6 m × 3.2 mm ID, 220 °C	ECD	4–5 ng g^{-1}	63
Diclofop Diclofop-methyl	soil, crops	2% OV-17, 1.6 m × 3.2 mm ID, 215 °C	ECD	10–50 ng g^{-1}	57
Fluazifop Fluazifop-butyl	groundwater	SE-52/4, 25 m × 0.22 mm ID, 50–260 °C	NPD	<100 ng L^{-1}	45
Fluazifop Fluazifop-butyl Haloxyfop Haloxyfop-ethoxyethyl Quizalofop Quizalofop-ethyl	fruit, vegetables	HP-5, fused silica capillary, 10 m × 0.53 mm ID, 60–260 °C	ECD, NPD	10 ng g^{-1} 50 ng g^{-1}	54
Fluazifop Fluazifop-butyl	crops, soils	SE-54, 30 m × 0.25 mm ID, 90–250 °C	NPD	10 ng g^{-1}	55

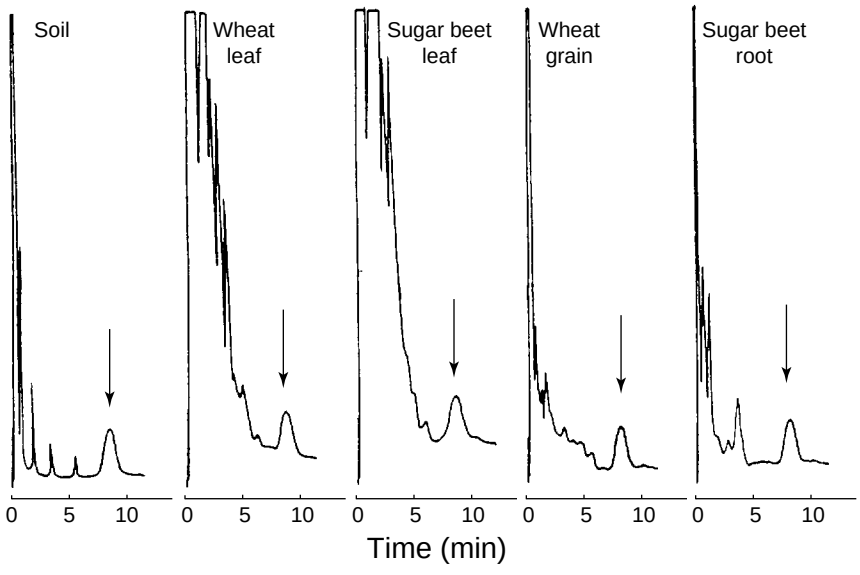

Figure 7 GC/ECD chromatograms of diclofop-methyl residues in soil and crops. Residual level in samples, 0.10 mg kg^{-1}; final volume, 20 mL; injection volume, 1 µL; detection current, 2.0 nA; attenuation, 4. (Reprinted from *J. Chromatogr.*, **547**, W. Liu, Z. Chen, H. Xu, Y. Shi, Y. Chen, 'Determination of Diclofop-Methyl and Diclofop Residues in Soil and Crops by GC', 509–515, Copyright 1991, with permission from Elsevier Science.)

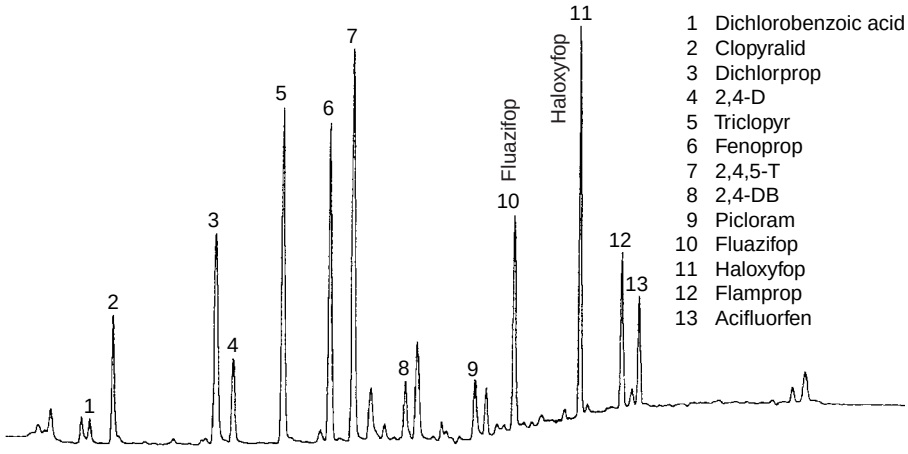

Figure 8 Gas chromatogram of a groundwater sample spiked with 100 ng L^{-1} of a mixture of 13 acidic herbicides derivatized to their methyl esters; detection by ECD. 2,4-D, (2,4-dichlorophenoxy) acetic acid; 2,4,5-T, (2,4,5-trichlorophenoxy) acetic acid; 2,4-DB, (2,4-dichlorophenoxy) butyric acid. (Reprinted from *J. Chromatogr.*, **643**, S. Butz, H.-J. Stan, 'Determination of Chlorophenoxy and Other Acidic Herbicide Residues in Groundwater by Capillary Gas Chromatography of Their Alkyl Esters Formed by Rapid Derivatization Using Various Chloroformates', 227–238, Copyright 1993, with permission from Elsevier Science.)

Gradient elution with a large change in mobile-phase composition is used. Buffer solutions with the pH-adjusted below the pK_a values of the acid ArPP are commonly used to increase the retention capacity of RP sorbents for HPLC separation of acidic herbicides.[40,49,56,58,65]

A possible method with which to increase the separation efficiency and retention capacity for acid herbicides is RP/HPLC under mild conditions, at nearly neutral pH and utilizing an ion-pairing mechanism with the ion-pair reagent tetrabutylammonium hydrogensulfate ($TBAHSO_4$)[9] and triethylamine.[44]

Variable-wavelength UV has been the routinely used detection technique, enabling sufficient sensitivity for most ArPPs to be detected.[9,40,44,58,65] It is usually operated in the range 220–280 nm. The problem of

Table 17 HPLC determination of ArPP herbicides

Compound	Matrix	Column	Mobile phase	Detector	Wavelength (nm)	DL	Ref.
Clodinafop Diclofop Fenoxaprop Fluazifop Haloxyfop Quizalofop	drinking water, spring water, groundwater	Supelcosil LC-18	solvent A: H_2O 0.1% TFA solvent B: CH_3OH-CH_3CN (70:30, v/v) 0.025% TFA gradient	DAD	240	2–6 ng L^{-1} 5–11 ng L^{-1} 5–11 ng L^{-1}	49
Diclofop-methyl Fluazifop-butyl	river water	LiChrosphere	solvent A: H_2O pH 5.8 solvent B: CH_3CN gradient	DAD	210	50 ng L^{-1} 100 ng L^{-1}	46
Fenoxaprop	groundwater	TSK ODS80-TM	solvent A: 0.005 M phosphate buffer pH 7.2 solvent B: CH_3CN gradient	DAD	220	100 ng L^{-1}	47
Fenoxaprop-ethyl	drinking water	LiChrosorb RP-18	solvent A: CH_3OH–triethylamine buffer (9:1, v/v) solvent B: CH_3OH–triethylamine buffer (4:6, v/v)	UV	240	100 ng L^{-1} 50 ng L^{-1}	44
Fluazifop	drinking water	LiChrosorb-CN, RP-18 Spheri	CH_3OH–0.1 M $CH_3COOH-CH_3COONa$ (pH 3.8) (50:50, v/v)	UV	230	50 ng L^{-1}	40

(*continued overleaf*)

Table 17 (*continued*)

Compound	Matrix	Column	Mobile phase	Detector	Wavelength (nm)	DL	Ref.
Fluazifop Fluazifop-butyl	onions	Zorbax ODS	solvent A: KH_2PO_4 (0.005 M)–CH_3OH (96 : 4, v/v) pH 2.3 solvent B: CH_3OH	UV	270	200 ng g^{-1}	58
Fenoxaprop Fenoxaprop-ethyl	soil	Supelcosil LC-18	solvent A: H_2O pH 3 with *o*-phosphoric acid solvent B: CH_3CN gradient	UV	280	20 ng g^{-1}	65
Fluazifop Fluazifop-butyl	soil	Hypersil-phenyl	CH_3CN–H_2O (54 : 46, v/v) 0.5 mmol L^{-1} $TBAHSO_4$	UV	222	40 ng g^{-1} 100 ng g^{-1}	9
		Hypersil-MOS (C_8)	solvent A: H_2O solvent B: CH_3CN				
Fluazifop Fluazifop-butyl	groundwater	Ultrasphere ODS	acetonitrile–water (50 : 50, v/v) + 0.4% acetic acid	UV	270	<100 ng L^{-1}	45
Fluazifop	soybeans, soybean oil	PRP-1	CH_3CN–0.15 M *o*-phosphoric acid (30 : 70, v/v)	AD	+125 V[a]	100 ng g^{-1}	56
Fluazifop Fluazifop-butyl	soil	C_{18} RP Carlo Erba Silica B/5 RP	methanol–H_2O–acetic acid (55 : 44 : 1, v/v) chloroform–hexane (10 : 90, v/v)	UV	245	50 ng g^{-1}	66

TFA, trifluoroacetic acid.
[a] Potential applied in the oxidation mode.

Table 18 Determination of ArPP herbicides by hyphenated techniques

Compound	Matrix	Technique	MS conditions	DL	Ref.
Diclofop Fluazifop Haloxyfop	soil	HPLC/MS	APCI/NI, SIM	0.1–0.3 ng g^{-1}	60
Diclofop-methyl Fenoxaprop Haloxyfop-ethoxyethyl Quizalofop-ethyl	fruit and vegetables	GC/MS	EI, SIM	<10 ng g^{-1}	50
Diclofop-methyl	soil	GC/MS	IT	5 ng g^{-1}	62
Diclofop Haloxyfop	groundwater	GC/MS	EI, TIC	500 ng L^{-1} 200 ng L^{-1}	37
Fluazifop Fluazifop-butyl Haloxyfop	drinking water, surface water, sewage water	GC/MS	EI, SIM	10 ng L^{-1}	43
Fluazifop Fluazifop-butyl Haloxyfop	drinking water, groundwater	GC/MS	EI, SIM	≤1 ng L^{-1}	39
Diclofop Fenoxaprop Fluazifop Haloxyfop Quizalofop	tap water, groundwater	GC/MS	EI, TIC	<100 ng L^{-1}	41
Diclofop-methyl Fenoxaprop-ethyl	fruit, vegetable matrix	GC/MS	IT	sub-μg g^{-1}	51
Fluazifop Fluazifop-butyl	groundwater	GC/MS	EI, TIC	<100 ng L^{-1}	42
Fluazifop Fluazifop-butyl	onions	GC/MS	EI, SIM	20 ng g^{-1}	58

Table 18 (*continued*)

Compound	Matrix	Technique	MS conditions	DL	Ref.
Diclofop-methyl	groundwater	GC/MS	EI, SIM	$<100\,ng\,L^{-1}$	38
Diclofop-methyl	soil	GC/MS	SIM	$3\,ng\,g^{-1}$	64
Quizalofop-ethyl	fruit and vegetables	GC/MS	EI, TIC	$50\,ng\,g^{-1}$	67
Fluazifop	agricultural products	GC/MS	TIC	below $250\,ng\,g^{-1}$	53
Fluazifop Fluazifop-butyl	groundwater	GC/MS	EI, TIC, SIM	$<100\,\mu g\,L^{-1}$	45
Fluazifop Fluazifop-butyl Haloxyfop Haloxyfop-ethoxyethyl Quizalofop Quizalofop-ethyl	fruit, vegetables	GC/MS	EI, SIM	$<10\,ng\,g^{-1}$	54
Fluazifop Fluazifop-butyl	crops, soil	GC/MS	EI, SIM	$10\,ng\,g^{-1}$	55
Fluazifop Fluazifop-butyl	soil	GC/MS	EI	$10\,ng\,g^{-1}$	66

APCI/NI, atmospheric pressure chemical ionization with negative ion mode of operation; IT, ion-trap.

identification, a characteristic of chromatographic methods, cannot be solved with the application of a UV detector. An increase of confirmation ability has been achieved with the DAD. This technique has been used in the simultaneous HPLC analysis of clodinafop, quizalofop,[49] diclofop-methyl, fluazifop-butyl,[46] and fenoxaprop[47] in aqueous samples. In the determination of fluazifop in soybean oil by HPLC, AD in oxidation mode[56] is used to achieve selectivity and sensitivity.

Table 17 summarizes the column, mobile phase, detector, and DLs currently employed in routine ArPP residue analyses.

5.3.3 Hyphenated Techniques

The identification and confirmation of analytical results is been obtained by combining MS with GC (GC/MS) or subsequently with LC (HPLC/MS). Table 18 summarizes representative papers using hyphenated techniques.

A newly developed scan function and methane gas control system has enabled classical chemical ionization (CI) spectra to be obtained in the IT, thus allowing direct comparison with reference spectra developed on magnetic and quadrupole instruments. Gas chromatography ion-trap mass spectrometry (GC/IT/MS) is used for the determination of diclofop-methyl and fenoxaprop-ethyl in fruit and vegetable matrices,[51] and diclofop-methyl in soil.[62]

Mass selective detection (MSD), using either full-scan mode or SIM, can be employed to achieve selective detection of target pesticides in the presence of a complex matrix. MSD is a highly sensitive and specific technique suitable for environmental organic analysis. The most widely used technique for ArPP residue analyses is MSD with EI ionization. Quantification is usually achieved with the SIM technique. With this technique selectivity is also improved.[38,39,43,45,50,54,55,58,64] Lee et al.[37] have determined diclofop and haloxyfop in groundwater; after esterification with 2,2,2,-trifluoromethanol, the esters are identified by GC with MSD in full-scan mode. In the method proposed by Butz and Stan,[41] for the simultaneous determination of diclofop, fenoxaprop, fluazifop, haloxyfop, and quizalofop in tap water and groundwater, all MS measurements are performed with MSD combined with GC in full-scan mode. MSD in full-scan mode is also utilized for the determination of fluazifop and fluazifop-butyl in groundwater,[41] and of quizalofop-ethyl in fruit and vegetables.[67]

To date only one report[60] discusses HPLC coupled with MS. The aim of this work was to attain lower levels of detection while maintaining sufficiently high specificity in order to unequivocally identify and quantify selected target ArPPs in soil samples; all data were obtained using HPLC/ESIMS in NI mode. Figure 9 shows a SIM chromatogram obtained in NI discharge on electrospray interface coupling HPLC to MS of a soil sample spiked with fluazifop, haloxyfop, and diclofop.

5.3.4 Other Techniques

IAs are now being seen as useful supplements to classical chromatographic analytical systems. Schwalbe et al.[68] describe the development of enzyme IA and fluoro-IA for the herbicide diclofop-methyl in soil, urine,

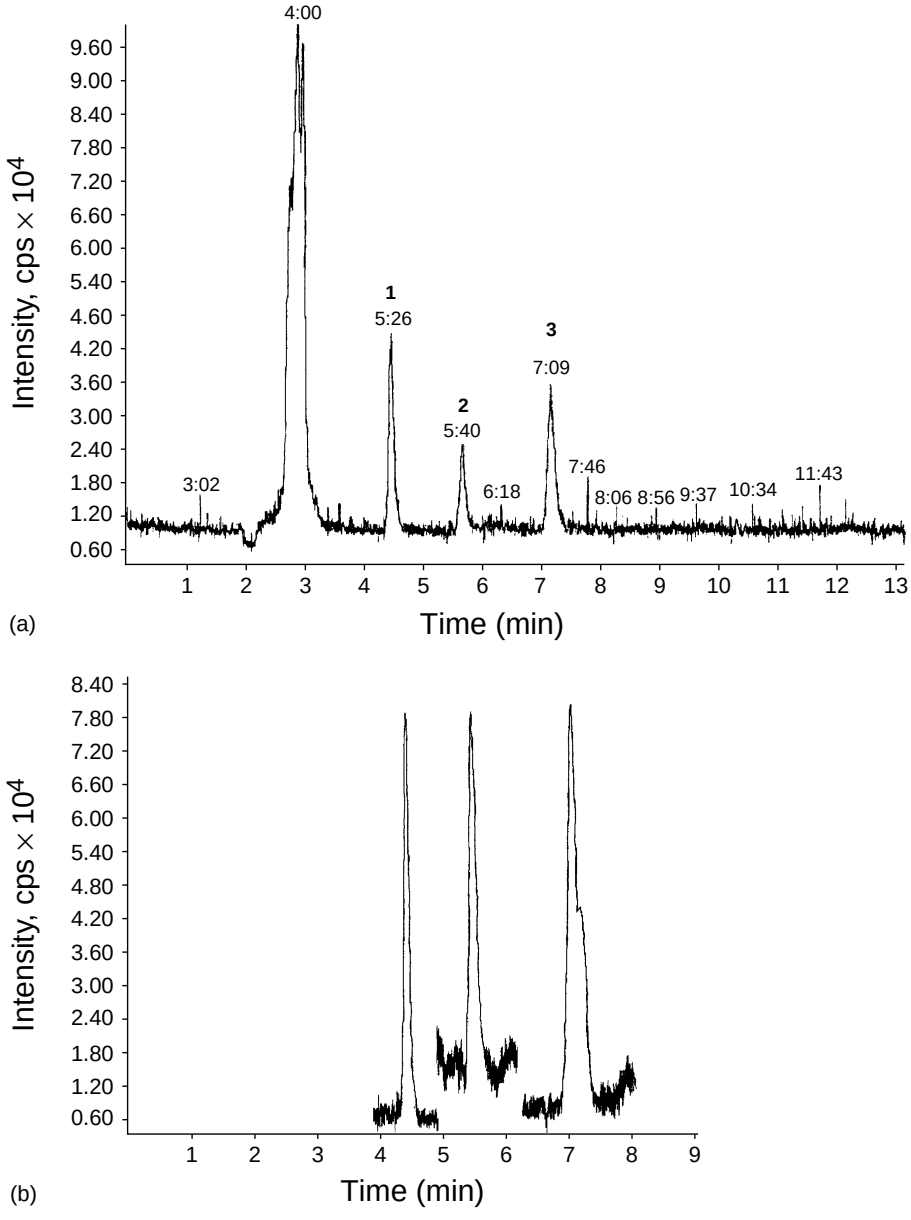

Figure 9 HPLC/ESIMS chromatogram obtained by Pomaia soil sample spiked with ArPP herbicides at the 1 ng g^{-1} level using (a) full-scan and (b) time-scheduled SIM conditions under negative ion (NI) mode: (1) fluazifop, (2) haloxyfop, and (3) diclofop. (Reprinted from *Anal. Chim. Acta*, **375**, A. Laganà, G. Fago, A. Marino, M. Mosso, 'SCE Followed by LC/ESIMS for the Efficient Determination of Aryloxyphenoxypropionic Herbicides in Soil Samples at ppb Levels', 107–116, Copyright 1998, with permission from Elsevier Science.)

serum, wheat, soybeans, and sugar beet. To produce antibodies against the herbicide, rabbits were immunized with a conjugate of diclofop acid and bovine serum albumin, which was prepared by use of a water-soluble carbodiimide. The resulting antisera showed a high specificity towards the (dichlorophenoxy)phenoxy moiety of the herbicide, which was demonstrated by the determination of the cross-reactions of structurally related compounds.

6 DIPHENYL ETHER HERBICIDES

6.1 Chemical and Physical Properties

The *p*-nitro DPhE herbicides are commonly known as the DPhE herbicides. This class of herbicides is mainly composed of esters, but a few compounds are acids or have an acidic behavior, with pK_a values between 2.7 and 3.8. They are yellow to brown crystals,

Table 19 Chemical structures, IUPAC-recommended names[10] and Chemical Abstracts Service registry numbers of DPhE herbicides

Chemical structure	Name	CAS
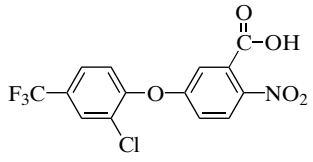	**acifluorfen** 5-(2-chloro-α,α,α-trifluoro-*p*-tolyloxy)-2-nitrobenzoic acid	[50594-66-6]
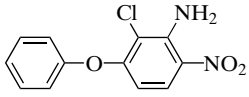	**aclonifen** 2-chloro-6-nitro-3-phenoxybenzenamine	[74070-46-5]
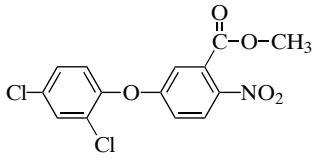	**bifenox** methyl 5-(2,4-dichlorophenoxy)-2-nitrobenzoate	[42576-02-3]
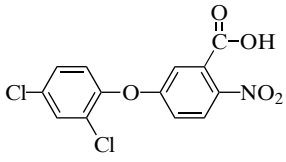	**bifenox acid** 5-(2,4-dichlorophenoxy)-2-nitrobenzoic acid	N/A
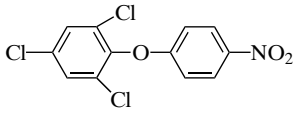	**chlornitrofen** 4-nitrophenyl 2, 4, 6-trichlorophenyl ether	[1836-77-7]
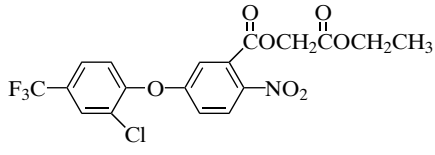	**fluoroglycofen-ethyl** ethyl *O*-[5-(2-chloro-α,α,α-trifluoro-*p*-tolyloxy)-2-nitrobenzoyl]glycolate	[77501-90-7]
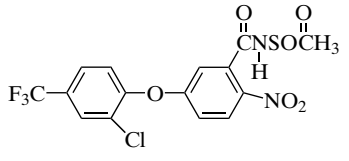	**fomesafen** 5-(2-chloro-α,α,α-trifluoro-*p*-tolyloxy)-*N*-methylsulfonyl-2-nitrobenzamide	[72178-02-0]
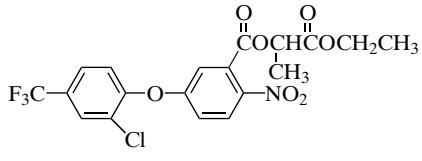	**lactofen** ethyl *O*-[5-(2-chloro-α,α,α,-trifluoro-*p*-tolyloxy)-2-nitrobenzoyl]-DL-lactate	[77501-63-4]
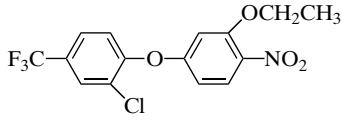	**oxyfluorfen** 2-chloro-α,α,α-trifluoro-*p*-tolyl 3-ethoxy-4-nitrophenyl ether	[42874-03-3]

presenting in some cases a slightly aromatic odor. All compounds of this class are stable in slightly acid or slightly alkaline media (pH 5–9), but are hydrolyzed quite rapidly above pH 9. The exception is fomesafen which is stable even at pH 11. DPhE herbicides have been subjected to many studies in order to understand their mode of action. Initial studies into the mechanism of action of these compounds suggested they were interfering with some aspect of photosynthetic electron transport.[69–73] Then Bugg et al.[74] obtained evidence

which indicated the site of inhibition by the DPhE was in the plastoquinone-Cyt *f* region between PS-I and PS-II. However, there is evidence to indicate the inhibition of electron transport is a secondary effect.[75–81] The most recent studies have determined that the site of action is protoporphyrinogen oxidase, and that the R group at the 3′ position of the DPhE molecule is an absolute requirement for the short term (8 h) expression of herbicidal activity.[82] However, the nature of the R group is also important.[82]

Structure/activity studies have shown that the highest biological activity of DPhE is generated by the substituents 2-Cl, 4-CF_3, 4-NO_2 and various structures in the 3′ position (see Table 19). DPhE derivatives having oxime substituent groups at the 3′ position were found to exhibit high herbicidal activity and soybean selectivity.[83]

DPhEs are systemic contact herbicides, absorbed more readily by the foliage, and especially by the shoots, then by the roots with very little translocation. In some cases herbicide activity is enhanced by sunlight. DPhE symptomatology is necrosis of shoot and root meristems with rapid desiccation of stems and leaves following PoE treatment of sensitive species.

This class of compounds is mainly used for PrE or PoE control of a wide spectrum of annual broad-leaved weeds and grasses in soybeans, peanuts, rice, and cereals. Particular crops include also potatoes, carrots, peas, maize, peppermint, onions, and a variety of tropical and subtropical crops.

The WHO toxicity class is III for all compounds except lactofen, which is WHO IV. Tests conducted on rabbits have demonstrated that DPhEs can be slight to moderate eye and skin irritants, except acifluorfen, which is a severe eye irritant. From studies of oral application in rats, it appears that fast and almost complete absorption and excretion occurs. Multiple applications did not indicate a cumulative effect, which suggests these herbicides do not present a substantial hazard to aquatic or terrestrial wildlife.

The basic physical chemical and other properties of the DPhE herbicides are listed in Tables 20–22.

6.2 Analytical Method for Diphenyl Ether Analysis

6.2.1 Extraction

The extractive methods for DPhEs from environmental and food samples are given in Table 23.

LLE is the method chosen by several authors[37,84,85] for DPhE extraction from aqueous samples. In some work[37,84] the sample is shaken with an immiscible organic solvent such as dichloromethane; Edgell et al.[85] extracted acifluorfen at pH < 2 with ethyl ether.

Sorption on to a solid sorbent provides a useful method for the isolation of DPhEs from aqueous samples. The most common sorbent used in SPE is C_{18} bonded silica cartridges.[39,41,43,45,86–88]

Junker-Buchheit and Witzenbacher[89] have developed an alternative SPE sorbent material to extract bifenox from tap water. This adsorbent is based on a highly cross-linked, porous ethylvinylbenzene–DVB copolymer and exhibits a unique adsorption efficiency due to its high specific surface area. The ability of this adsorbent (LiChrolut EN) is comparable to a conventional RP-18 solid phase. Increased attention has been devoted to carbon sorbents,[87,90] in addition to those with an SiO_2 matrix. GCB is a nonspecific sorbent and is generally hydrophobic. Contrary to the sorbents based on SiO_2, this may be used without taking into account the pH of the treated solutions.

Comparisons of different sorbents for SPE are proposed. SPE materials and techniques (C_{18} Empore disks, polystyrenedivinylbenzene, C_{18} Bond Elut cartridges, and ENVI-Carb cartridges) have been compared for the extraction of bifenox from water.[87] Hodgeson et al.[91] evaluated C_{18} and polystyrene DVB resin disks for the extraction of DPhEs in aqueous samples. Fibers coated with PDMS or liquid crystals of PAC have been developed for the extraction of oxyfluorfen from water samples[92] by SPME. SPME is an alternative technique that involves direct extraction of the analyte with the use of a small diameter optical fiber that has been coated with a polymeric stationary phase and housed in a syringe assembly for protection. SPME eliminates the separate concentration step from the SPE and LLE methods because the analytes diffuse directly into the coating of the SPME device and are concentrated there. XAD-2 has been utilized for the extraction of bifenox from river water.[93]

DPhE residues in soil and food samples are traditionally extracted from the solvent with some physical assistance (sonication, shaking, homogenizing). The solvents used include acetone,[51,53,94,95] acetonitrile,[67,96] mixed methanol–water (3 : 1, v/v),[62] dichloromethane–ethanol–acetone (6 : 2 : 1, v/v),[97] and methanol–0.1 mol L^{-1} NaOH (80 : 20, v/v).[98]

Homogeneous sample pulp containing oxyfluorfen, prepared from fruit and vegetables of different water content with or without additional water, is adsorbed on the surface of activated Florisil to obtain a free-falling powder, which is extracted in a glass column with ethyl acetate or dichloromethane–acetone (9 : 1, v/v).[59]

6.2.2 Cleanup

The choice of cleanup technique closely depends on how the extraction was done; in other words, whatever else enters the extract determines the purification procedure.

Table 20 Chemical and physical properties of DPhE herbicides

Property	Acifluorfen	Aclonifen	Bifenox	Chlornitrofen	Fluoroglycofen-ethyl	Fomesafen	Lactofen	Oxyfluorfen	Bifenox acid
Empirical formula	$C_{14}H_7ClF_3NO_5$	$C_{12}H_9ClN_2O_3$	$C_{14}H_9Cl_2NO_5$	$C_{12}H_6Cl_3NO_3$	$C_{18}H_{13}ClF_3NO_7$	$C_{15}H_{10}ClF_3N_2O_6S$	$C_{19}H_{15}ClF_3NO_7$	$C_{15}H_{11}ClF_3NO_4$	$C_{13}H_7Cl_2NO_5$
Molecular weight	361.7	264.7	342.1	318.5	447.8	438.8	461.8	361.7	328.1
Physical form	light brown solid	yellow crystals	yellow crystals with a slightly aromatic odor	pale yellow crystals	dark amber solid	colorless crystalline solid	dark brown to tan	orange crystalline solid	yellowish crystals
Melting point (°C)	–	81–82	84–86	107	65	220–221	–	85–90	169–171
Vapor pressure (mPa)	0.01 (20 °C)	0.016 (20 °C)	0.32 (30 °C)	3.2 (20 °C)	$<1.33 \times 10^5$ (25 °C)	<0.1 (50 °C)	–	0.0267 (25 °C)	–
Solubility	(25 °C)	(20 °C)	(25 °C)	(25 °C)	(25 °C)	(20 °C)	(20 °C)	(25 °C)	–
water (mg L^{-1})	120	1.4	0.35	0.25	0.6	50 and <1 at pH 1	<1	0.116	–
acetone (g kg^{-1})	600	–	400	–	readily soluble in most organic solvents except hexane	300 (g L^{-1})	–	725	–
toluene	–	390	–	–		–	–	–	–
xylene	<10	–	300	360		1.9	–	–	–
n-hexane	–	4.5	–	–		0.5	–	–	–
cyclohexanone	–	–	–	–		150	–	615	–
methanol	–	50	–	–		–	–	–	–
ethanol	500	–	<50	–		–	–	–	–
dichloromethane	50	–	–	–		10	–	–	–

Table 21 Stability of the DPhE herbicides

Parameter	Acifluorfen	Aclonifen	Bifenox	Chlornitrofen	Fluoroglycofen-ethyl	Fomesafen	Lactofen	Oxyfluorfen
Stability	pH 3–9 (40 °C)	pH 3–9	pH 5–7.3 (22 °C)	–	–	pH 3–11 (40 °C)	–	–
Other observations	decomposes at 235 °C, DT_{50} 110 h under UV light	slowly decomposes when exposed to light	hydrolyzed at pH 9; stable up to 175 °C, totally decomposed above 290 °C	stable under neutral conditions up to 250 °C	aqueous suspensions are rapidly decomposed by UV light	stable >6 months at 50 °C, unstable in light; stable to aqueous photolysis for 32 days at pH 7, 25 °C	–	stable up to 50 °C; DT_{50} 3 days under UV light
DT_{50} in water	2 h in light	1 month in presence of micro-organisms	24 min, in saturated solution, at $\lambda = 250–400$ nm	–	0.25 mg L^{-1}, 22 °C: 231 days at pH 5 15 days at pH 7 0.15 days at pH 9	–	–	rapid photodegradation
DT_{50} in soil	108 days in silt loam 200 days in clay loam	36–80 days (22 °C)	ca. 5 h for a thin film on soil	–	11 h, microbial degradation	>6 months under aerobic conditions, <1–2 months under anaerobic conditions, 100–104 days by photo-degradation	–	slow photodegradation; in the dark 292 days under aerobic conditions 580 days under anaerobic conditions
DT_{50} in field	–	–	–	–	–	15 weeks	–	5–55 days

Table 22 Applications and environmental intake of the DPhE herbicides

Parameter	Acifluorfen	Aclonifen	Bifenox	Chlornitrofen	Fluoroglycofen-ethyl	Fomesafen	Lactofen	Oxyfluorfen
Acute oral LD_{50} rat (mg/kg body weight)	1540	>5000	>5000	–	1500	1250–2000	>5000	>5000
Acute dermal LD_{50} rabbit (mg/kg body weight)	>2000	–	>2000	–	>5000	>1000	–	>10 000
WHO toxicity class[a]	III	III	III	–	III	III	–	III
Fate in:								
animals	excreted	excreted	excreted	–	metabolized	–	–	–
plants	metabolized	hydroxylated	–	–	metabolized	metabolized in soy	–	metabolized
DT_{50} in plants	1 week	14 days	–	–	–	–	–	–
Uses[b]	U1	U2	U3		U4	U5	U6	U7

[a] WHO toxicity class: Ia, extremely hazardous; Ib, highly hazardous; II, moderately hazardous; III, slightly hazardous.

[b] U1 = PoE control of annual broad-leaved weeds (*Abutilon, Amaranthus, Datura, Euphorbia, Poligonum, Ipomea, Xanthium*) with some effects on grasses in soybeans, peanuts and rice; U2 = PrE control of grass, and broad-leaved weeds in winter wheat, potatoes, sunflowers, peas, carrots, maize and other crops; U3 = PrE and PoE control of annual broad-leaved weeds and some grasses in cereals, sorghum, soybeans, rice, and some other crops; U4 = PrE and PoE control of broad-leaved weeds and grasses, particularly *Gallium, Viola* and *Veronica*, in wheat, balley, oats, peanuts, rice, and soybeans; U5 = early PoE control of broad-leaves weeds in soybeans; U6 = PoE control of annual broad-leaved weeds in cereals, potatoes, soybeans, rice and peanuts; U7 = PrE and PoE control of annual broad-leaved weeds and grasses in a variety of tropical and subtropical crops, particularly in tree fruit, including citrus, vines, nuts, cereals, maize, soybeans, peanuts, rice, cotton, bananas, peppermints, onions, garlic, ornamental trees, and shrubs, and conifer seedbeds.

Table 23 Extraction and cleanup procedures used in the determination of DPhE herbicides

Compound	Matrix	Extraction	Partition step	Cleanup	Ref.
Oxyfluorfen	soil	sonication with water and methanol (1 : 3, v/v)	–	SPE C_{18} disk	62
Acifluorfen	water	Carbograph-1	–	–	90
Bifenox	tap water	LiChrolut EN or LiChrolut RP-18 cartridge	–	–	89
Oxyfluorfen	water	SPE C_{18}	–	–	86
Oxyfluorfen	water	SPME/PAC or SPME/PDMS	–	–	92
Acifluorfen	groundwater	dichloromethane	–	–	37
Bifenox	water	SPE C_{18} or Empore C_{18} disks or SPE-Envi-Carb	–	–	87
Oxyfluorfen	water	dichloromethane	hexane	–	84
Acifluorfen	drinking water, groundwater	SPE C_{18}	–	–	39
Acifluorfen	tap water, groundwater	SPE C_{18}	–	–	41
Fluoroglycofen Bifenox	fruit and vegetable matrices	acetone	petroleum ether and dichloromethane	–	51
Acifluorfen	drinking water, surface water, sewage water	SPE C_{18}	–	–	43
Acifluorfen	drinking water	pH < 2 with ethyl ether	–	–	85
Acifluorfen	reagent water, tap water, surface water, groundwater	Empore C_{18} and PS/DVB disks	–	–	91
Bifenox	vegetables	acetone	dichloromethane	GPC (Bio-Beads-X3)	94
Chlornitrofen	river water	SPE C_{18}	–	–	88
Oxyfluorfen	fruit and vegetables	Florisil	–	–	59
Bifenox	soil	dichloromethane–ethanol–acetone (6 : 2 : 1, v/v/v)	–	GPC (Bio-Beads S-X3)	97
Bifenox Oxyfluorfen	fruit and vegetables	acetonitrile	–	–	67
Bifenox	agricultural products	acetone	–	GPC (Bio-Beads S-X3)	53
Fomesafen	groundwater	SPE C_{18}	–	–	45
Bifenox	agricultural products	acetone	hexane	GPC (Bio-Beads S-X3)	95
Bifenox Oxyfluorfen	crops	acetonitrile	–	Attagel, Florisil, SPE C_{18}	96
Acifluorfen	soil, water	methanol–0.1 M NaOH (80 : 20, v/v)	dichloromethane	Sep-Pac silica cartridge	98
Bifenox	river water	XAD-2	–	–	93

PAC, polyacrylate; PDMS, polydimethylsiloxane; XAD-2, styrene–divinylbenzene copolymer.

LLP followed by column chromatography or SPE has been the most common cleanup procedure in DPhE analysis. Dichloromethane,[94,98] hexane,[84,95] and petroleum ether and dichloromethane[51] are used in LLP. Hsu et al.[96] have investigated the potential of a SPE cartridge, Florisil, and SPE-C_{18}, to replace Attagel in the bifenox cleanup procedure for raw agricultural crops. SPE with a silica cartridge was used by Gennari et al.[98] for the cleanup of the organic extracts of soil and water in acifluorfen determination. An alternative mode of SPE is the use of the ''SPE disks''. Bao et al.[62] reported the use of C_{18} Empore extraction disks for isolation and cleanup of oxyfluorfen in soil.

The most universally applicable cleanup of organic extracts of DPhE from soil,[97] vegetables,[94] and agricultural products[53,95] is GPC. Separation is generally performed by using DVB-linked polystyrene gels, mostly Bio Beads SX-3. For the elution of DPhE herbicides, the mixtures cyclohexane–EtOAc (50 : 50, v/v) and cyclohexane–acetone (50 : 50, v/v) are used.

6.3 Determination

Once the sample has been extracted from its matrix, usually into a suitable solvent, it is then ready to be analyzed by the appropriate technique. Chromatographic methods, and in particular GC and HPLC, are the methods of choice. GC is often the first choice in the analysis of DPhEs because of their good response with ECD; however, DPhEs are also well analyzed by HPLC, and this

is often used in conjunction with UV detection. GC or HPLC, coupled with MS, provides the analyst with a powerful tool for both determination and confirmatory analysis. Because of the good voltammetric behavior of fluoroglycofen, the mercury-drop electrode is also applied.

6.3.1 Gas Chromatography

Fused capillary columns with phases of different polarity have been used: (nonpolar examples are SBP-5,[90] 5% phenylmethylsilicone,[86] DB-5,[59,84,91] 5% phenyl–95% methylpolysiloxane;[85] low-polar examples are 65% dimethyl–35% diphenylpolysiloxane,[37] SBP-608,[41] 50% phenyl–50% methylsilicone,[96] OV-17.[93]); typical dimensions are IDs of 0.22–0.53 mm, film thicknesses of 0.25–1 μm, and lengths of 12.5–60 m.

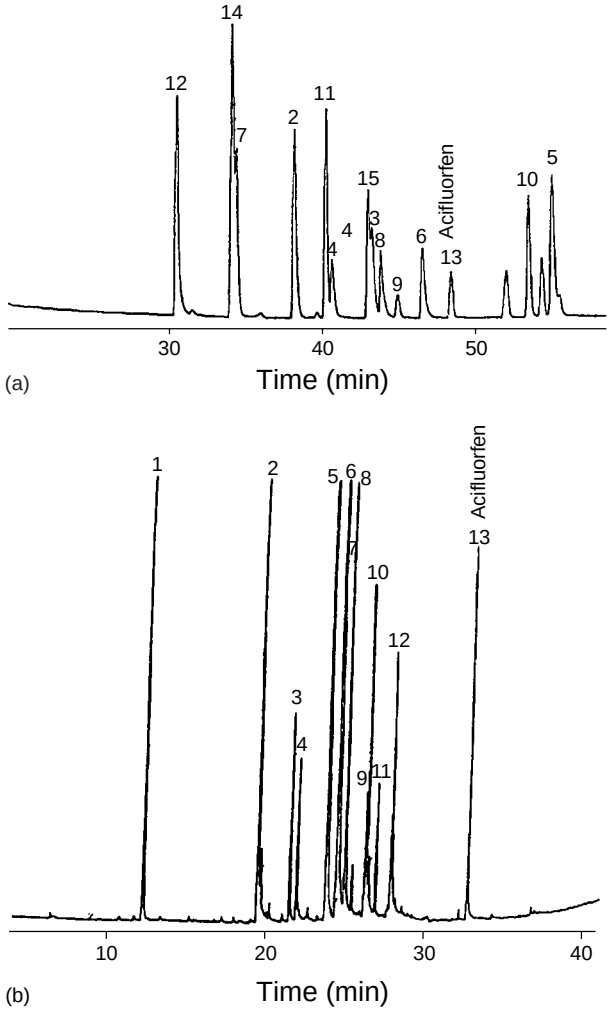

Figure 10 Chromatograms by (a) HPLC/UV and (b) GC/ECD separation of pesticides extracted from 1 L spiked water. (Reproduced by permission of Vieweg from F. Bruner, A. Berloni, P. Palma, *Chromatographia*, **43**, 279–284 (1996).)

Some derivatization methods have been developed to improve the sensitivity and the peak-tailing situation in the case of acifluorfen and fluoroglycofen. The most used derivatization procedures are with diazomethane,[85,90,91] 2,2,2-TFE,[37] pentafluorobenzyl bromide,[39,43] and methyl, ethyl, and butyl chloroformate.[41] As all DPhEs bear a chlorine heteroatom, the most used detection system is ECD.[37,41,59,85,86,90,91,93,96] Oxyfluorfen in water[86] is determined by GC with ECD and flame photometric detection (FPD). A GC/AED system[84] is used to analyze nitrogen-containing oxyfluorfen herbicides. The use of GC linked to AED offers an alternative to GC/MS for unambiguous determination of unknowns. The coupling technique was shown to be a very selective tool with element-characteristic chromatograms acquired using different element emission lines, thereby enhancing the selectivity of the method for environmental monitoring.

Figure 10 illustrates chromatograms by HPLC/UV and GC/ECD of a water sample spiked with acifluorfen.

Table 24 summarizes the column, mobile phase, detector, and DLs currently employed in routine DPhE residue analysis.

6.3.2 High-performance Liquid Chromatography

The usual way to obtain a good separation of acidic DPhE in GC is by chemical derivatization. HPLC is the method of choice for ionic compounds, acifluorfen, fluoroglycofen, fomesafen, and bifenox acid, which are not amenable to GC.

An RP/HPLC column containing C_{18} bonded silica sorbent[45,87,89,90,97–99] is used for analytical separation. Narrow-bore (3 mm ID) columns and 3 μm particles[87,99] have been applied in recent studies. For the elution, a mixture of acetonitrile and aqueous solution is used in isocratic mode.[45,97–99] Gradient elution with acetonitrile and water modified with TFA[90] or ammonium acetate[87,89] have also been utilized. Buffer solutions with the pH-adjusted to below the pK_a values of the acidic DPhE are prepared from TFA,[90] H_3PO_4,[97,98] and CH_3COOH.[45]

UV detection, in the range 206–296 nm, is the most popular detection method in the analysis of DPhE residues.[45,87,90,97–99] In the determination of bifenox from tap water[89] an increase of confirmation ability has been achieved with DAD technology. Zhou and Miles[99] have demonstrated that oxyfluorfen can be determined selectively in crops by PCD. PCD and UV liquid chromatographic detectors also are evaluated.

The analytical conditions for HPLC analysis are presented in Table 25. Figure 11 shows a chromatogram of a soil extract of bifenox and its main metabolite, bifenox acid, using an UV detector.

Table 24 Gas chromatographic determination of DPhE herbicides

Compound	Matrix	Column, dimension, temperature	Detector	DL	Ref.
Acifluorfen	water	SBP-5, 30 m × 0.25 mm ID, 60–300 °C	ECD	0.08 ng L^{-1}	90
Oxyfluorfen	water	ultra two-capillary column 5% phenylmethylsilicone, 25 m × 0.32 mm ID, 160–280 °C; HP-1 methylsilicone, 12.5 m × 0.22 mm ID, 45–270 °C	ECD, FPD	22 ng L^{-1}	86
Acifluorfen	groundwater	cross-bonded 65% dimethyl–35% diphenyl polysiloxane, 30 m × 0.53 mm ID, 130–260 °C	ECD	1000 ng L^{-1}	37
Oxyfluorfen	water	DB-5, 30 m × 0.32 mm ID, 75–320 °C	AED	0.22 ng L^{-1}	84
Acifluorfen	tap water, groundwater	SBP-608, 30 m × 0.25 mm ID, 50–280 °C	ECD	<100 ng L^{-1}	41
Acifluorfen	drinking water	5% phenyl–95% methylpolysiloxane, 30 m × 0.25 mm ID, 60–300 °C	ECD	–	85
Acifluorfen	reagent water, tap water, surface water, groundwater	DB-5, 30 m × 0.25 mm ID, 50–230 °C	ECD	250 ng L^{-1}	91
Oxyfluorfen	fruit and vegetables	DB-5, 60 m × 0.245 mm ID, 85–300 °C	ECD	10 ng g^{-1}	59
Bifenox Oxyfluorfen	crops	50% phenyl–50% methyl silicone, 15 m × 0.25 mm ID, 170–240 °C	ECD	–	96
Bifenox	river water	OV-17 CB-FSC, 25 m × 0.32 mm ID, 90–275 °C	ECD	5 ng L^{-1}	93

Table 25 HPLC Determination of DPhE herbicides

Compound	Matrix	Column	Mobile phase	Detector	Wavelength (nm)	DL	Ref.
Acifluorfen	water	C_{18} RP	solvent A: H_2O + 0.05% TFA solvent B: CH_3CN + 0.025% TFA gradient	UV	280	41.7 ng L^{-1}	90
Bifenox	tap water	LiChroCART 250–4, Superspher 100, RP-18	solvent A: ammonium acetate solvent B: CH_3CN gradient	DAD/UV	220	7000 ng L^{-1}	89
Bifenox	water	Hypersil ODS, 3 µm	solvent A: ammonium acetate solvent B: CH_3CN gradient	UV	220	61–147 ng L^{-1}	87
Bifenox Bifenox acid	soil	Hypersil ODS	CH_3CN–0.05 mol L^{-1} H_3PO_4 (65:35, v/v)	UV	206	4 ng g^{-1}	97
Fomesafen	groundwater	Ultrasphere ODS	CH_3CN–H_2O (50:50 v/v) + 0.4% acetic acid	UV	270	100 ng L^{-1}	45
Oxyfluorfen	Carbanzo beans	HS–C_{18}, 3 µm or OmiPac PAX-500	CH_3CN–H_2O (58:42, v/v); 1.8 mmol L^{-1} Na_2CO_3 and 1.7 mmol L^{-1} $NaHCO_3$	UV, PCD	277	<200 ng g^{-1}	99
Acifluorfen	soil, water	LiChrosorb RP-18	H_2O (pH 3 with H_3PO_4)–CH_3CN (30:70, v/v)	UV	296	6 ng g^{-1} 3 ng g^{-1}	98

6.3.3 Hyphenated Techniques

Although GC and HPLC are widely accepted for the residual analysis of DPhE, the need for compound identification cannot be overemphasized. This is particularly necessary for soil and agricultural products in which peak overlapping or unresolved peaks originating from matrix interference exist.

Table 26 summarizes the characteristics and the DLs of hyphenated techniques used for DPhE herbicide determination and confirmation in environmental and food samples.

GC/MS has become the preferred analytical technique. DPhE herbicides have been determined by GC/MS directly or after the preparation of derivatives.

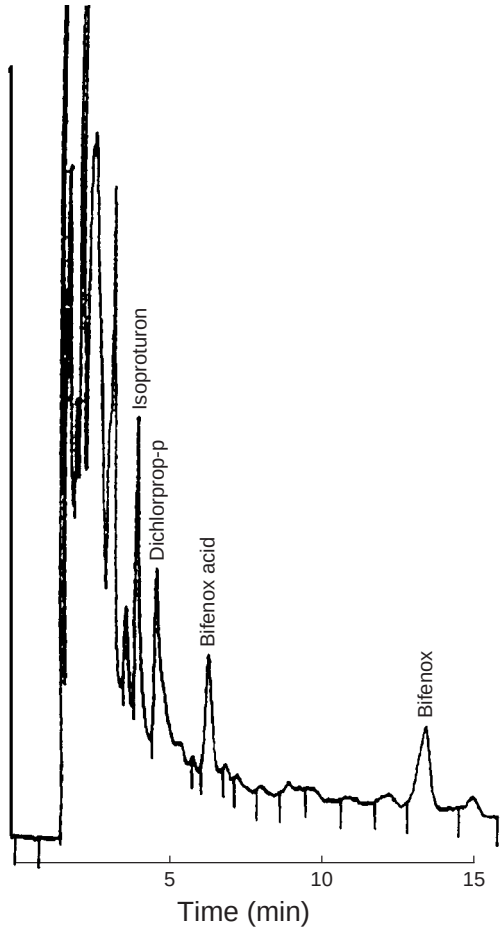

Figure 11 HPLC chromatogram for bifenox and bifenox acid in soil extract with cleanup; fortification level 0.1 mg kg^{-1}. (Reproduced from *Fresenius' J. Anal. Chem.*, 'Simultaneous Determination of the Herbicides Isoproturon, Dichlorprop-p and Bifenox in Soil Using RP/HPLC', H. Klöppel, J. Haider, C. Hoffmann, B. Lüttecke, **344**, 42–46, Figure 3, 1992, Copyright Springer-Verlag.)

GC/MS in EI mode in TIC was used for monitoring oxyfluorfen,[92] and acifluorfen[41] in water samples, bifenox and oxyfluorfen[67] and bifenox[53] in agricultural products. Capillary GC with MSD employing SIM with time window programming were developed for acifluorfen,[39,43] and chlornitrofen[88] in drinking water and surface water.

GC/MS of DPhE residue analysis has been largely restricted to quadrupole instruments. Another technique that has been used for DPhE quantitation is that of GC/IT/MS. The components of an IT system are much less complicated than the instrumentation previously mentioned and rely on electronic control to obtain mass spectra. Bao et al.[62] utilized IT/MS for testing oxyfluorfen in soil. The same technique was used by Cairns et al.,[51] and Tuinstra et al.[94,95] for bifenox and fluoroglycofen determination in agricultural products.

One potential approach to the determination of acifluorfen in water[90] is the use of HPLC/MS in conjunction with interface particle-beam. In this work, to improve the sensitivity, the PB target zone inside the ion source is coated with a Teflon layer. It is well known that when acidic pesticides enter the ion source, they undergo decomposition due to the high temperature of the ionization chamber (>200 °C). This inconveniently causes a reduction in sensitivity and an unwanted memory effect. Coating the ion source reduces adsorption and decomposition, and enhances the gas-phase conversion of analytes.

6.3.4 Other Techniques

Classical DC polarography and fast-scan differential pulse voltammetry were applied to study the reduction mechanism of fluoroglycofen-ethyl at a dropping mercury electrode (DME) and at a static mercury-drop electrode

Table 26 Determination of DPhE herbicides using hyphenated techniques

Compound	Matrix	Technique	MS conditions	DL	Ref.
Oxyfluorfen	soil	GC/MS	IT	4 ng g^{-1}	62
Acifluorfen	water	LC/PBMS	SIM	5.2 ng L^{-1}	90
Oxyfluorfen	water	GC/MS	EI, TIC	1 ng L^{-1}	92
Acifluorfen	drinking water, groundwater	GC/MS	EI, SIM	≤10 ng L^{-1}	39
Acifluorfen	tap water, groundwater	GC/MS	EI, TIC	<100 ng L^{-1}	41
Bifenox Fluoroglycofen	fruit and vegetables	GC/MS	IT	sub-μg g^{-1} level	51
Acifluorfen	drinking water, surface water, sewage water	GC/MS	EI, SIM	10 ng L^{-1}	43
Bifenox	vegetables	GC/MS	IT	50 ng g^{-1}	94
Chlornitrofen	river water	GC/MS	EI, SIM	50 ng L^{-1}	88
Bifenox Oxyfluorfen	fruit and vegetables	GC/MS	EI, TIC	150 μg g^{-1} 50 μg g^{-1}	67
Bifenox	agricultural products	GC/MS	EI, TIC	below 250 ng g^{-1}	53
Bifenox	agricultural products	GC/MS	IT, TIC	–	95

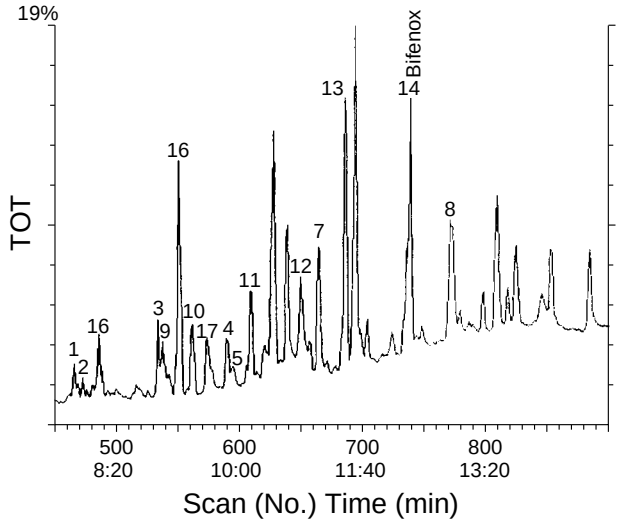

Figure 12 GC/MS reconstructed ion chromatogram of a combined endive, cucumber, and cauliflower extract spiked with pesticides (full-scan mode). (Reprinted from *J. Chromatogr.*, **552**, L.G.M.Th. Tuinstra, F.R. Povel, A.H. Roos, 'Multi-matrix Multi-pesticide Method for Agricultural Products', 259–264, Copyright 1991, with permission from Elsevier Science.)

(SMDE) in a medium of 47–49% (v/v) or 10–20% (v/v) dimethylformamide.[100] The nitro groups were reduced to hydroxylamine derivatives, or further to the corresponding amines, depending on the type of compound and the solution acidity. The determination of nitropesticide was performed using FSDVP and adsorptive stripping voltammetry (AdSV) in model samples. The DLs were found both experimentally and by calculation from regression analyses. According to the experimental condition their values were in the range 17–160 ng mL^{-1} or 0.1–2.8 ng mL^{-1} for FSDVP or AdSV, respectively. The nitropesticide in artificially contaminated soils was determined after two or three extractions with acetone.

Figure 12 shows a reconstructed ion chromatogram of an extract of crops spiked with bifenox. Selected-ion chromatograms of standards, blind, and recovery river water samples showing the ion trace for each selected masses are presented in Figure 13.

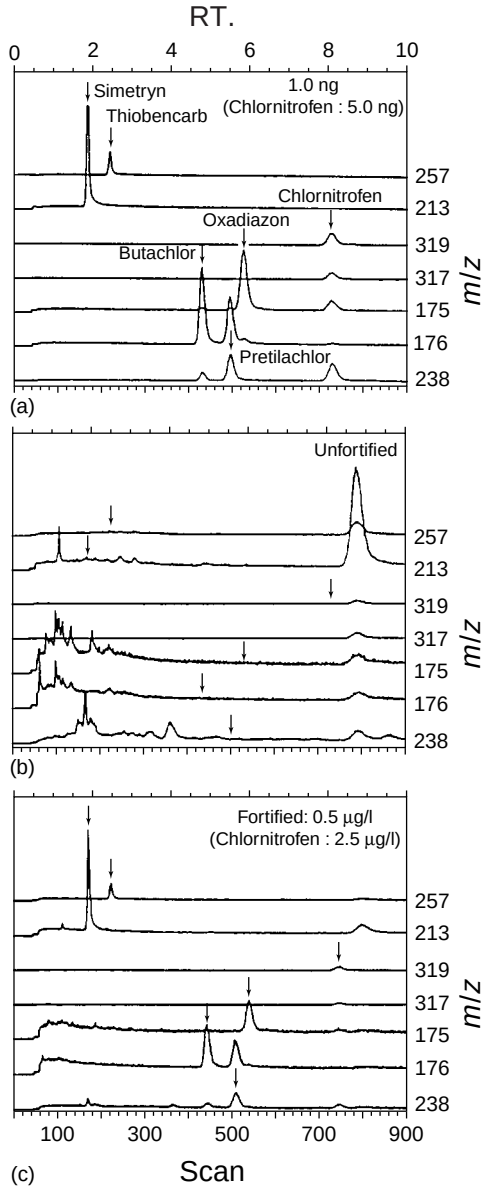

Figure 13 GC/MS fragmentograms: (a) a mixture of the pesticides examined; (b) blind river water sample; (c) recovery test sample. (Reprinted from *J. Chromatogr.*, **643**, H. Kobayashi, K. Ohyama, N. Tomiyama, Y. Jimbo, O. Matano, S. Goto, 'Determination of Pesticides in River Water by GC/MS/SIM', 197–202, Copyright 1993, with permission from Elsevier Science.)

ABBREVIATIONS AND ACRONYMS

AD	Amperometric Electrochemical Detection
AdSV	Adsorptive Stripping Voltammetry
AED	Atomic Emission Detection
AHAS	Acetohydroxy Acid Synthase
APCI/NI	Atmospheric Pressure Chemical Ionization with Negative Ion Mode of Operation
ArPP	Aryloxyphenoxypropionate
ATP	Adenosine Triphosphate
CI	Chemical Ionization
CoA	Coenzyme A
DAD	Diode Array Detector
DC	Direct Current
DL	Detection Limit
DME	Dropping Mercury Electrode

DPhE	Diphenyl Ether
DVB	Divinylbenzene
ECD	Electron Capture Detection
ECNCI	Electron Capture Negative Chemical Ionization
EI	Electron Impact
ELISA	Enzyme-linked Immuno-adsorbent Assay
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization/Mass Spectrometry
ESIMS/MS	Electrospray Ionization Tandem Mass Spectrometry
FDA	Food and Drug Administration
FI	Fluorimetric
FID	Flame Ionization Detection
FPD	Flame Photometric Detection
FSDVP	Fast Scan Differential Pulse Voltammetry
GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/IT/MS	Gas Chromatography Ion-trap Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
IA	Immunoassay
ID	Internal Diameter
IMI	Imidazolinone
IT	Ion-trap
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
LLP	Liquid–Liquid Partition
MAE	Microwave-assisted Extraction
MRL	Maximum Residue Limit
MS	Mass Spectrometry
MSD	Mass Selective Detection
NCI	Negative Chemical Ionization
NI	Negative Ion
NPD	Nitrogen–Phosphorus Detection
PAC	Polyacrylate
PB	Particle Beam
PBMS	Particle Beam Mass Spectrometry
PCD	Photoconductivity Detection
PCI	Positive Chemical Ionization
PDMS	Polydimethylsiloxane
PLRP-S	Copolymer Styrene Divinylbenzene
PoE	Post-emergence
PrE	Pre-emergence
PRP-1	Polymeric RP Material
PSDVB	Polystyrene–Divinylbenzene
RP	Reversed-phase
SAX	Strong Anion Exchange
SCE	Soil Column Extraction
SCX	Strong Cation Exchange
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SMDE	Static Mercury-drop Electrode
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
$TBAHSO_4$	Tetrabutylammonium Hydrogensulfate
TFA	Trifluoroacetic Acid
TFE	Trifluoroethanol
TIC	Total Ion Current
USDA	United States Department of Agriculture
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
XAD-2	Styrene–Divinylbenzene Copolymer

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of

Pesticides, Strategies for • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Pesticides (New Generation) and Related Compounds, Analysis of • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of

REFERENCES

1. 'EEC Drinking Water Guidelines', 80/779/EEC, EEC No L229/11-29, EEC, Brussels, 1980.
2. R.B. Maybury, 'Codex Alimentarius Approach to Pesticide Residue Standards', *J. Assoc. Off. Anal. Chem.*, **72**(3), 538–541 (1989).
3. R. Coscollá, *Residuos de plaguicidas en alimentos vegetales*, Mundi-Prensa, Madrid, 1996.
4. J.P. Frawley, 'Codex Alimentarius – Food Safety – Pesticides', *Food Drug Cosmet. Law J.*, **42**(2), 168–173 (1987).
5. R.T. Krause, 'Multiresidue Method for Determining *N*-Methylcarbamate Insecticides in Crops, Using High-performance Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **63**(5), 1114–1124 (1980).
6. P. Lombardo, 'The FDA Pesticide Program: Goals and New Approaches', *J. Assoc. Off. Anal. Chem.*, **72**(3), 518–520 (1989).
7. R.L. Ellis, 'Changing Pesticide Technology in Meat and Poultry Products', *J. Assoc. Off. Anal. Chem.*, **72**(3), 521–524 (1989).
8. H. Frehse, in 'Pesticide Science and Biotechnology (Proceedings of the 6th IUPAC Congress of Pesticide Chemistry, Ottawa 1986)', eds. R. Greenhalgh, T.R. Roberts, Blackwell, London, 293–330, 1987.
9. M. Zanco, G. Pfister, A. Kettrup, 'Determination of Fluazifop-butyl and Fluazifop with the Use of Disposable Solid-phase Extraction Columns for Selective Cleanup and Concentration of Soxhlet Soil Extract', *Fresenius' J. Anal. Chem.*, **344**(7–8), 345–349 (1992).
10. C.D.S. Tomlin (ed.), *The Pesticide Manual*, 11th edition, British Crop Protection Council, 1997.
11. D.L. Shaner, S.L. O'Connor (eds.), *The Imidazolinone Herbicides*, CRC Press, Boca Raton, FL, 1991.
12. D.L. Shaner, P.C. Anderson, M.A. Stidham, 'Imidazolinones Potent Inhibitors of Acetohydroxyacid Syntase', *Plant Physiol.*, **76**(2), 545–546 (1984).
13. P.C. Anderson, K.A. Hibberd, 'Evidence for Interaction of an Imidazolinone Herbicide with Leucine, Valine, and Isoleucine Metabolism', *Weed Sci.*, **33**(4), 479–483 (1985).
14. D.L. Shaner, N.M. Mallipudi, 'Mechanism of Selectivity of the Imidazolinone Herbicides', in *The Imidazolinone Herbicides*, eds. D.L. Shaner, S.L. O'Connor, CRC Press, Boca Raton, FL, 1991.
15. M.R. Van Ellis, D.L. Shaner, 'Mechanism of Cellular Absorption of Imidazolinones in Soybean (Glycine Max) Leaf Discs', *Pestic. Sci.*, **23**(1), 25–34 (1988).
16. F.D. Hess, 'Herbicide Absorption and Translocation and Their Relationship to Plant Tolerances and Susceptibility', in *Weed Physiology*, ed. S.O. Duke, CRC Press, Boca Raton, FL, Vol. 2, 1985.
17. J.B. Fischer, J.L. Michael, 'Use of ELISA Immunoassay Kits as a Complement to HPLC Analysis of Imazapyr and Triclopyr in Water Samples from Forest Watersheds', *Bull. Environ. Contam. Toxicol.*, **59**(4), 611–618 (1997).
18. A. Laganà, G. Fago, A. Marino, 'Simultaneous Determination of Imidazolinone Herbicides from Soil and Natural Waters Using Soil Column Extraction (SCE) and Off-line Solid-phase Extraction Followed by Liquid Chromatography with UV Detection or Liquid Chromatography/Electrospray–Mass Spectroscopy', *Anal. Chem.*, **70**(1), 121–130 (1998).
19. R.D. Mortimer, D.F. Weber, 'Determination of Residual Imazetapyr in Soybeans by Gas Chromatography/Nitrogen–Phosphorus Detection', *J. AOAC Int.*, **76**(2), 377–381 (1993).
20. W.H. Newsome, P.G. Collins, 'Determination of Imazamethabenz in Cereal Grain by Enzyme-linked Immunosorbent Assay', *Bull. Environ. Contam. Toxicol.*, **47**(2), 211–216 (1991).
21. S.J. Stout, A.R. Da Cunha, J.S. Fletcher, G.L. Picard, 'Rapid Determination of Imazapyr in Corn by Gas Chromatography/Mass Spectrometry with Electro Capture Negative Ion Chemical Ionization', *J. AOAC Int.*, **79**(5), 1220–1225 (1996).
22. J. Atienza, J.J. Jimenrez, A. Herguedas, J.L. Bernal, 'Comparative Study of Three Extraction Procedures for Imazamethabenz-methyl in Agricultural Soil', *J. Chromatogr. A*, **721**(1), 113–121 (1996).
23. C.S. Helling, M.A. Doherty, 'Improved Method for the Analysis of Imazapyr in Soil', *Pestic. Sci.*, **45**(1), 21–26 (1995).
24. O. Nováková, 'Determination of Imazethapyr and Imazapyr Residues in Soil by Coupled-column Liquid Chromatography', *Chromatographia*, **39**(1–2), 62–66 (1994).
25. W. Liu, A. Pusino, C. Gessa, 'High-performance Liquid Chromatographic Determination of the Herbicide Imazapyr Residues in Water and Soil', *Sci. Total Environ.*, **123/124**, 39–43 (1992).
26. W.S. Curran, R.A. Liebl, P.W. Simmons, 'Effects of Tillage and Application Method on Clomazone, Imazaquin and Imazethapyr Persistence', *Weed Sci.*, **40**(3), 482–489 (1992).
27. K. Reddy, M.A. Locke, 'Supercritical CO_2 Fluid Extraction of Imazaquin from Soil', *Weed Sci.*, **42**(2), 249–253 (1994).
28. S.J. Stout, A.R. Da Cunha, M.M. Safarpour, 'Simplified Determination of Imidazolinone Herbicides in

Soil at Part-per-billion Level by Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry', *J. AOAC Int.*, **80**(2), 426–432 (1997).
29. S.J. Stout, A.R. Da Cunha, D.G. Allardice, 'Microwave-assisted Extraction Coupled with Gas Chromatography/Electron Capture Negative Chemical Ionization Mass Spectrometry for the Simplified Determination of Imidazolinone Herbicides in Soil at the ppb Level', *Anal. Chem.*, **68**(4), 653–658 (1996).
30. S.J. Stout, A.R. Da Cunha, G.L. Picard, M.M. Safarpour, 'Microwave-assisted Extraction Coupled with Liquid Chromatography/Electrospray Ionization Mass Spectrometry for the Simplified Determination of Imidazolinone Herbicides and Their Metabolites in Plant Tissue', *J. Agric. Food Chem.*, **44**(11), 3548–3553 (1996).
31. S.J. Stout, A.R. Da Cunha, G.L. Picard, M.M. Safarpour, 'Rapid, Direct Determination of Imidazolinone Herbicide in Water at the 1 ppb Level by Liquid Chromatography/Electrospray Ionization Mass Spectrometry and Tandem Mass Spectrometry, *J. Agric. Food Chem.*, **44**(8), 2182–2186 (1996).
32. K.A. O'Bryan, B.J. Brecke, D.G. Shilling, D.L. Colvin, 'Comparison of Bioassay Techniques for Detecting Imazaquin in Soil', *Weed Technology*, **8**(2), 203–206 (1994).
33. R.B. Wong, Z.H. Ahmed, 'Development of an Enzyme-linked Immunosorbent Assay for Imazaquin Herbicide', *J. Agric. Food Chem.*, **40**(5), 811–816 (1992).
34. M. Lee, R.A. Durst, 'Determination of Imazethapyr Using Capillary Column Flow Injection Liposome Immunoanalysis', *J. Agric. Food Chem.*, **44**(12), 4032–4036 (1996).
35. K.A. Walker, S.M. Ridley, J.L. Harwood, 'Effects of the Selective Herbicide Fluazifop on Fatty Acid Synthesis in Pea (*Pisum sativum*) and Barley (*Hordeum vulgare*)', *Biochem. J.*, **254**(3), 811–817 (1988).
36. H. Nestler, in *Chemie der Pflanzenschutz und Schldädlings-bekämpfungsmittel*, ed. R. Wegler, Springer-Verlag, New York, 1–25, Vol. 8, 1982.
37. A.S. Lee, M.-K. Hong, A.E. Smith, 'Analysis of 2,2,2-Trifluoroethyl Derivatives of Carboxylic Acid Herbicides by Gas Chromatography with Mass-selective and Electron Capture Detection', *J. AOAC Int.*, **78**(6), 1459–1464 (1995).
38. M.F. Legrand, E. Costentin, A. Bruchet, 'Occurrence of 38 Pesticides in Various French Surface and Groundwater', *Environ. Technol.*, **12**(11), 985–996 (1991).
39. S. Butz, Th. Heberer, H.-J. Stan, 'Determination of Phenoxyalkanoic Acids and Other Acidic Herbicides at the Low ppt Level in Water Applying Solid-phase Extraction with RP-C_{18} Material', *J. Chromatogr. A*, **677**(1), 63–74 (1994).
40. A. Balinova, 'Solid-phase Extraction Followed by High-performance Liquid Chromatographic Analysis for Monitoring Herbicides in Drinking Water', *J. Chromatogr.*, **643**(1–2), 203–207 (1993).
41. S. Butz, H.-J. Stan, 'Determination of Chlorophenoxy and Other Acidic Herbicide Residues in Groundwater by Capillary Gas Chromatography of Their Alkyl Esters Formed by Rapid Derivatization Using Various Chloroformates', *J. Chromatogr.*, **643**(1–2), 227–238 (1993).
42. H.-J. Stan, E.T. Heberer, 'Automated Recognition of Target Compounds at Low Levels in Environmental Samples by Means of Capillary Gas Chromatography/Mass Spectrometry with Dedicated Mass Spectral Libraries and the Macro Program AUTARG', *J. Chromatogr. A*, **653**(1), 55–62 (1993).
43. Th. Heberer, S. Butz, H.-J. Stan, 'Analysis of Phenoxycarboxylic Acids and Other Acidic Compounds in Tap, Ground, Surface and Sewage Water at the Low ng L^{-1} Level', *Intern. J. Environ. Anal. Chem.*, **58**(1–4), 43–53 (1995).
44. A.M. Balinova, 'Analysis of Fenoxaprop-ethyl and Fenoxaprop in Drinking Water Using Solid-phase Extraction and Ion-pair HPLC', *Pestic. Sci.*, **48**(3), 219–223 (1996).
45. E. Bolygó, N.C. Atreya, 'Solid-phase Extraction for Multiresidue Analysis of Some Triazole and Pyrimidine Pesticides in Water', *Fresenius' J. Anal. Chem.*, **339**(6), 423–430 (1991).
46. E. Papadopoulou-Mourikidou, J. Patsias, 'Development of a Semi-automated High-performance Liquid Chromatographic/Diode Array Detection System for Screening Pesticides at Trace Levels in Aquatic Systems of the Axios River Basin', *J. Chromatogr. A*, **726**(1–2), 99–113 (1996).
47. S. Dupas, S. Guenu, V. Pichon, A. Montiel, B. Welté, M.-C. Hennion, 'Long-term Monitoring of Pesticides and Polar Transformation Products in Groundwater Using Automated On-line Trace-enrichment and Liquid Chromatography with Diode Array Detection', *Intern. J. Environ. Anal. Chem.*, **65**(1–4), 53–68 (1996).
48. K. Wuchner, R. Grob, 'Application de l'Extraction par Fluid Supercritique à l'Analyse de Polluants Aqueux', *Analusis*, **23**(5), 227–229 (1995).
49. A. Laganà, G. Fago, A. Marino, 'Determination of Aryloxyphenoxypropionic Acid Herbicides in Water Using Different Solid-phase Extraction Procedures and Liquid Chromatography/Diode Array Detection', *J. Chromatogr. A*, **796**(2), 309–318 (1998).
50. A. Gelsomino, B. Petrovicová, S. Tiburtini, E. Magnani, M. Felici, 'Multiresidue Analysis of Pesticides in Fruit and Vegetables by Gel Permeation Chromatography Followed by Gas Chromatography with Electron Capture and Mass Spectrometric Detection', *J. Chromatogr. A*, **782**(1), 105–122 (1997).
51. T. Cairns, K.S. Chiu, D. Navarro, E. Siegmund, 'Multiresidue Pesticide Analysis by Ion-trap Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**(11), 971–988 (1993).

52. H.-J. Stan, M. Linkerhägner, 'Capillary Gas Chromatography–Atomic Emission Detection: A Useful Instrumental Method in Pesticide Residue Analysis of Plant Foodstuffs', *J. High Resolut. Chromatogr.*, **16**(9), 539–548 (1993).
53. L.G.M.Th. Tuinstra, A.H. Roos, A.M. Matser, W.A. Traag, J.A. van Rhijn, 'Development of a Multi Residue/Multi-matrix method for Pesticide Analysis in Agricultural Products', *Fresenius' J. Anal. Chem.*, **339**(6), 384–386 (1991).
54. Z. Jehlicková, J. Hajšlová, J. Poustka, F. Pudil, J. Davidek, 'Determination of Aryloxyphenoxypropanoic Acid Derivatives in Crops Treated with Herbicidal Sprays', *Z. Lebensm. Unters. Forsch.*, **190**(5), 436–440 (1990).
55. B.S. Clegg, 'Gas Chromatographic Analysis of Fluazifop-butyl (Fusilade) in Potatoes, Soybeans, and Soil', *J. Agric. Food Chem.*, **35**(2), 269–273 (1987).
56. B.L. Worobey, J.B. Shields, 'Determination of Fluazifop-butyl and Fluazifop Acid in Soybeans and Soybean Oil Using Liquid Chromatography with Oxidative Amperometric Detection', *J. Assoc. Off. Anal. Chem.*, **72**(2), 368–371 (1989).
57. W. Liu, Z. Chen, H. Xu, Y. Shi, Y. Chen, 'Determination of Diclofop-methyl and Diclofop Residues in Soil and Crops by Gas Chromatography', *J. Chromatogr.*, **547**(1–2), 509–515 (1991).
58. Y.Y. Wigfield, M. Lanouette, 'Supercritical Fluid Extraction of the Fortified Residues of Fluazifop-P-butyl (Fusilade II) and Its Major Metabolite, Fluazifop-P, in Onions', *J. Agric. Food Chem.*, **41**(1), 84–88 (1993).
59. L. Kadenczki, Z. Arpad, I. Gardi, A. Ambrus, L. Gyorfi, G. Reese, W. Ebing, 'Column Extraction of Residues of Several Pesticides from Fruit and Vegetables: A Simple Multiresidue Analysis Method', *J. Assoc. Off. Anal. Chem.*, **75**(1), 53–61 (1992).
60. A. Laganà, G. Fago, A. Marino, M. Mosso, 'Soil Column Extraction Followed by Liquid Chromatography Electrospray Ionization Mass Spectrometry for the Efficient Determination of Aryloxyphenoxypropionic Herbicides in Soil Samples at ppb Levels', *Anal. Chim. Acta*, **375**, 107–116 (1998).
61. M. Zanco, G. Pfister, A. Kettrup, 'A New HPLC Method for the Simultaneous Determination of Fluazifop-butyl and Fluazifop in Soil Samples', *Fresenius' J. Anal. Chem.*, **344**(1–2), 39–41 (1992).
62. M.L. Bao, F. Pantani, K. Barbieri, D. Burrini, O. Griffini, 'Multi-residue Pesticide Analysis in Soil by Solid-phase Disk Extraction and Gas Chromatography/Ion-trap Mass Spectrometry', *Intern. J. Environ. Anal. Chem.*, **64**(4), 233–245 (1996).
63. W. Liu, Z. Chen, Q. Lu, Y. Shi, 'Analysis of Fluazifop-butyl and Fluazifop Residues in Soil and Crops by Gas Chromatography', *Analyst*, **116**(3), 273–276 (1991).
64. G.W. Bruns, S. Nelson, D.G. Erickson, 'Determination of MCPA, Bromoxynil, 2,4-D, Trifluralin, Triallate, Picloram, and Diclofop-methyl in Soil by GC/MS Using Selected Ion Monitoring', *J. Assoc. Off. Anal. Chem.*, **74**(3), 550–553 (1991).
65. L. Celi, M. Nègre, M. Gennari, 'HPLC Determination of Fenoxaprop and Fenoxaprop-ethyl in Different Soils', *Pestic. Sci.*, **38**(1), 43–47 (1993).
66. M. Patumi, C. Marucchini, M. Buselli, C. Vischetti, 'An Analytical Method for the Determination of Fluazifop-butyl and Fluazifop Residues in Soil', *Pestic. Sci.*, **21**(3), 193–201 (1987).
67. W. Liao, T. Joe, W.G. Cusick, 'Multiresidue Screening Method for Fresh Fruit and Vegetables with Gas Chromatographic/Mass Spectrometric Detection', *J. Assoc. Off. Anal. Chem.*, **74**(3), 554–565 (1991).
68. M. Schwalbe, E. Dorn, K. Beyermann, 'Enzyme Immunoassay and Fluoroimmunoassay for the Herbicide Diclofop-methyl', *J. Agric. Food Chem.*, **32**(4), 734–741, (1984).
69. D.E. Moreland, W.J. Blackmon, H.E. Todd, F.S. Farmer, 'Effects of Diphenylether Herbicides on Reactions of Mitochondria and Chloroplasts', *Weed Sci.*, **18**(5), 636–642 (1970).
70. J.F. Pereira, W.E. Splittstoesser, H.J. Hopen, 'Response of Plant Tissues to Nitrofen', *Weed Sci.*, **19**(5), 662–667 (1971).
71. E.F. Eastin, 'Selectivity of Fluorodifen Between Peanut and Cucumber', *Abstr. Weed Sci. Soc. Amer.*, No. 180, 1972.
72. P.J. Kruger, D. Ward, R.J. Thiessen, C.R. Downing, H.A. Kaufman, 'Bifenox: A Selective Weed Killer', Proc. 12th Brit, Weed Control Conf., 839, 1974.
73. T. Pollak, 'The Response of Beans to Fluorodifen as Affected by Light and Other Factors', *Diss. Abstr. Int.*, **B34**, 5752 (1974).
74. M.W. Bugg, J. Whitmarsch, C.E. Rieck, W.S. Cohen, 'Inhibition of Photosynthetic Electron Transport by Diphenyl Ether Herbicides', *Plant Physiol.*, **65**(1), 47–50 (1980).
75. S. Matsunaka, 'Acceptor of Light Energy in Photoactivation of Diphenyl Ether Herbicides', *J. Agric. Food Chem.*, **17**(2), 171–175 (1969).
76. S. Matsunaka, 'Utilization of Artificial Mutant Rice for Fundamental Researches of Herbicidal Mechanisms', *JAQR*, **6**, 189–192 (1972).
77. R.O. Fadayomi, G.F. Warren, 'The Light Requirement for Herbicidal Activity of Diphenyl Ethers', *Weed Sci.*, **24**(4), 598–600 (1976).
78. G.N. Prendeville, G.F. Warren, 'Effect of Four Herbicides and Two Oils on Leaf-cell Membrane Permeability', *Weed Res.*, **17**(4), 251–258 (1977).
79. D.E. Vanstone, 'Physiological Aspects of the Mode of Action of Nitrofluorfen and Oxyfluorfen', *Diss. Abst. Int.*, **B38**, 2974 (1978).

80. D.E. Vanstone, E.H. Stobbe, 'Light Requirement of the Diphenyl Ether Herbicide Oxyfluorfen', *Weed Sci.*, **27**(1), 88–91 (1979).
81. M.K. Pritchard, G.F. Warren, R.A. Dilley, 'Site of Action of Oxyfluorfen', *Weed Sci.*, **28**(5), 640–646 (1980).
82. G.L. Orr, F.D. Ess, 'Characterization of Herbicidal Injury by Acifluorfen-methyl in Excised Cucumber (*Cucumis sativus* L.) Cotyledons', *Pestic. Biochem. Physiol.*, **16**(3), 171–178 (1981).
83. Y. Hayashi, H. Kouji, 'Synthesis and Herbicidal Activity of Geometrical Isomers of Methyl [[[1-[5-[2-Chloro-4-(trifluoromethyl)phenoxy]-2-nitrophenyl]-2-methoxyethylidene]amino]oxy]Acetate (AKH-7088)', *J. Agric. Food Chem.*, **38**(3), 845–850 (1990).
84. N.L. Olson, R. Carrel, R. Cummings, R. Rieck, S. Reimer, 'Atomic Emission Detection for Gas Chromatographic Analysis of Nitrogen-containing Herbicides in Water', *J. AOAC Int.*, **78**(6), 1464–1473 (1995).
85. K.W. Edgell, E.J. Erb, R.J. Wesselman, J.E. Longbottom, 'Gas Chromatographic/Electron Capture Detection Method for Determination of Chlorinated Acids in Water: Collaborative Study', *J. AOAC Int.*, **76**(5), 1098–1112 (1993).
86. C. de la Colina, A. Peña, M.D. Mingorance, F. Sánchez-Rasero, 'Influence of Solid-phase Extraction Process on Calibration and Performance Parameters for the Determination of Pesticide Residues in Water by Gas Chromatography', *J. Chromatogr. A*, **733**(1 + 2), 275–281 (1996).
87. J. Schülein, D. Martens, P. Spitzauer, A. Kettrup, 'Comparison of Different Solid-phase Extraction Materials and Techniques by Application of Multiresidue Methods for the Determination of Pesticides in Water by High-performance Liquid Chromatography', *Fresenius' J. Anal. Chem.*, **352**(6), 565–571 (1995).
88. H. Kobayashi, K. Ohyama, N. Tomiyama, Y. Jimbo, O. Matano, S. Goto, 'Determination of Pesticides in River Water by Gas Chromatography/Mass Spectrometry/Selected-ion Monitoring', *J. Chromatogr.*, **643**(1 + 2), 197–202 (1993).
89. A. Junker-Buchheit, M. Witzenbacher, 'Pesticide Monitoring of Drinking Water with the Help of Solid-phase Extraction and High-performance Liquid Chromatography', *J. Chromatogr. A*, **737**(1), 67–74 (1996).
90. F. Bruner, A. Berloni, P. Palma, ' Determination of Acidic Chlorinated Pesticides in Water: Comparison of Official USEPA Method 505.1 and Liquid–Solid Extraction HPLC/UV and HPLC/PBMS Analyses', *Chromatographia*, **43**(5–6), 279–284 (1996).
91. J. Hodgeson, J. Collins, W. Bashe, 'Determination of Acid Herbicides in Aqueous Samples by Liquid–Solid Disk Extraction and Capillary Gas Chromatography', *J. Chromatogr. A*, **659**(2), 395–401 (1994).
92. A.A. Boyd-Boland, S. Magdic, J.B. Pawliszyn, 'Simultaneous Determination of 60 Pesticides in Water Using Solid-phase Microextraction and Gas Chromatography/Mass Spectrometry', *Analyst*, **121**(7), 929–938 (1996).
93. M. Ishibashi, M. Suzuki, 'Simultaneous XAD-2 Resin Extraction and High-resolution Electron Capture Gas Chromatography of Chlorine-containing Herbicides in Water Samples', *J. Chromatogr.*, **456**(2), 382–391 (1988).
94. L.G.M.Th. Tuinstra, P. Van De Spreng, P. Gaikhorst, 'Ion-trap Detection for Development of a Multiresidual Matrix Method for Pesticide Residues in Agricultural Products', *Intern. J. Environ. Anal. Chem.*, **58**(1–4), 81–91 (1995).
95. L.G.M.Th. Tuinstra, F.R. Povel, A.H. Roos, 'Multimatrix Multi-pesticide Method for Agricultural Products', *J. Chromatogr.*, **552**(1 + 2), 259–264 (1991).
96. R.-C. Hsu, I. Biggs, N.K. Saini, 'Solid-phase Extraction Cleanup of Halogenated Organic Pesticides', *J. Agric. Food Chem.*, **39**(9), 1658–1666 (1991).
97. H. Klöppel, J. Haider, C. Hoffmann, B. Lüttecke, 'Simultaneous Determination of the Herbicides Isoproturon, Dichlorprop-p and Bifenox in Soil Using RP-HPLC', *Fresenius' J. Anal. Chem.*, **344**(1–2), 42–46 (1992).
98. M. Gennari, M. Negre, A. Cignetti, 'Liquid Chromatography Determination of Acifluorfen in Soil and Water', *J. Assoc. Off. Anal. Chem.*, **73**(4), 599–601 (1990).
99. M. Zhou, C.J. Miles, 'Determination of Oxyfluorfen Herbicide and Oxyfluorfen Amine Residues in Carbanzo Beans by Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **74**(3), 546–550 (1991).
100. M. Kotoucek, M. Opravilová, 'Voltammetric Behavior of Some Nitropesticides at the Mercury Drop Electrode', *Anal. Chim. Acta*, **329**(1–2), 73–81 (1996).

High-performance Liquid Chromatography Methods in Pesticide Residue Analysis

Elbert A. Hogendoorn
National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands

1 Introduction **2**
2 Reversed-phase Liquid Chromatography with Ultraviolet Detection **4**
2.1 General Aspects 4
2.2 Practical Method Development 6
2.3 Systematic Method Development in Coupled Column Liquid Chromatography 10
2.4 Liquid Chromatography/Ultraviolet Methods 11
2.5 Solid-phase Extraction/Liquid Chromatography/Ultraviolet Methods 12
2.6 Coupled Column/Ultraviolet Methods 14
2.7 Coupled Column/Ultraviolet Methods Employing Off-line Solid-phase Extraction 16
3 Reversed-phase Liquid Chromatography with Fluorescence Detection **20**
3.1 Liquid Chromatography/Fluorescence Detection Methods 20
3.2 Solid-phase Extraction/Liquid Chromatography/Fluorescence Detection Methods 20
3.3 Coupled Column/Fluorescence Detection Methods 21
4 Reversed-phase Liquid Chromatography with Mass Spectrometric Detection **22**
4.1 General Aspects 22
4.2 Liquid Chromatography/Mass Spectrometry Methods 23
4.3 On-line Liquid Chromatography/Mass Spectrometry Methods 25
5 Liquid Chromatography Sample Pretreatment Methods prior to Gas Chromatography **26**
5.1 Gel Permeation Chromatography and Normal-phase Liquid Chromatography 26
5.2 Solid-phase Extraction Combined On-line with Gas Chromatography 27
6 Conclusions and Trends **27**
Acknowledgments **28**
Abbreviations and Acronyms **28**
Related Articles **29**
References **29**

The important role of high-performance liquid chromatography (HPLC) in pesticide residue analysis (PRA) is presented. Based on the experience and view of the author, a number of reliable and robust liquid chromatography (LC) approaches and methods have been selected to highlight the favorable features of this technique in this field of analysis.

Reversed-phase liquid chromatography (RPLC) with ultraviolet (UV) detection mode is most widely adopted, owing to its ability to perform efficient separation of very polar to apolar pesticides and the universal character of the detection mode. Therefore, a large part will deal with reversed-phase liquid chromatography/ultraviolet (RPLC/UV) applications showing its excellent separation performance and detectability of compounds not directly amenable to gas chromatography (GC).

Another important feature of RPLC is the possibility of including sample pretreatment steps such as sample concentration and/or cleanup in the analytical procedure. The very good compatibility with aqueous samples or extracts easily permits RPLC on-line sample loading (trace enrichment) by means of column switching, allowing fully automated procedures. It will be shown that, depending on whether the first column is a low- or a high-resolution column, two different modes of column switching, solid-phase extraction (SPE) to LC and LC to LC, are adopted in the determination of pesticides. The type of columns and materials, the advantages and the application range of both column-switching modes is discussed in detail. Attention will be paid to method development including both the straightforward trial-and-error approach and more systematic method development procedures.

The more selective/sensitive detection modes, fluorescence detection (FD) and mass spectrometry (MS) as applied in pesticides residue analysis will be discussed with emphasis on the increasing importance of MS for identification and quantification of polar pesticides in various types of samples.

Finally, the use of LC methods involving normal-phase liquid chromatography (NPLC), low-resolution RPLC and gel permeation chromatography (GPC), will demonstrate the powerful sample pretreatment feature of LC prior to gas chromatographic analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The major task of the analytical discipline PRA is to provide reliable and, moreover, cost-effective methods for the identification and quantification of hundreds of pesticides with large differences in physicochemical properties in very different types of matrices. The introduction of high-resolution capillary GC employing fused silica columns in the early 1980s, combined with the commercial availability of several sensitive and, moreover, selective GC detectors, contributed greatly to the rapidly increasing popularity of GC-based procedures for PRA. These favorable aspects make capillary GC, today, the major analytical and productive technique for this type of analysis. For example, in the Netherlands a multiresidue method (MRM) based on GC with MS detection used for regulatory purposes[1] covers the determination of 270 pesticides in foodstuffs from a total of 465 mentioned in the Dutch Regulation on Pesticides in Foodstuffs.[2] An overview of most pesticides which are not covered by the gas chromatography/mass spectrometry (GC/MS) method is given in Table 1. These pesticides and their corresponding chemical families are characterized by low volatility, high polarity and/or thermal instability making them not directly amenable for GC. For this type of analyte LC can be a good alternative. Most of these analytes can be nicely separated and efficiently detected by means of RPLC with UV detection, while GC analysis requires an often complicated derivatization procedure prior to separation. The availability of robust and more selective fluorescence and mass spectrometric detectors for LC further enhances the practical value of this separation technique.

It must be stressed that, because it is possible to combine high sensitivity and selectivity (confirmation), GC/MS procedures involving a derivatization step are still attractive and frequently applied for the thermostable compounds such as phenoxy acids. In fact, one must emphasize the complementary aspect of LC to GC. For example, LC can be a more viable way for the determination of thermolabile compounds such as carbamates or analytes difficult to derivatize, such as quaternary ammonium compounds.

Beside the advantageous use of HPLC for (very) polar pesticides, there are, as displayed in Figure 1, other important considerations for selecting HPLC in PRA. For example, in monitoring and/or indicative studies rendering a relatively low number of positive samples, the availability of a screening method(s), e.g. RPLC/UV for the determination of polar pesticides, can be cost-effective. The ability of RPLC to separate apolar and very polar analytes in one run makes it possible to determine pesticides simultaneously with their usually distinctly more polar metabolites.

RPLC is very compatible with aqueous samples, which enables sample pretreatment steps to be included by means of column switching in the analytical procedure, allowing fully automated (on-line) methods for the determination of pesticides in water samples. Furthermore, direct injection of a large volume of water makes RPLC attractive for other types of study, e.g. to study the formation of metabolites and transformation products (TPs) of pesticides at real-life trace levels (in-line analysis) or the at-site monitoring of pesticides in suspected water locations.

Other LC modes like GPC or NPLC have proven to be efficient for the cleanup of difficult matrices, e.g. fatty food products or soils. The high potential for automation provides semi and/or fully automated procedures effectively replacing the conventional laborious and less accurate open-column LC procedures.

Concerning methodology one can discriminate between single residue methods (SRMs) and MRMs. From an economical point of view, MRMs are highly desirable because the detection and quantitative determination of a series of pesticides in one run considerably reduces the cost of analysis.

However, multiresidue analysis of pesticides is faced with several problems. Firstly PRA has to cover the two disciplines including public health and environmental protection. Consequently, one has to deal with a large variety of sample types of widely diverging composition making it difficult to include the multimatrix aspect into an MRM. There is no guarantee that a method developed, for example, for the determination of chlorophenoxy acids in soils, will work for some foodstuffs, so different MRMs are required for the same group of pesticides.

Another difficulty is caused by regulations in relation to the different matrices. For example, maximum residue levels (MRLs) of pesticides in foodstuffs arranged by the Dutch legislation[2] are in agreement with the directives of the European Community (EC) and comply with the recommendations of the Codex Committee on Pesticide Residues (CCPR). These levels are based on scientific evaluations estimating the acceptable daily intake and the expected residue level in food when a pesticide is used according to good agricultural practice. As a result, for a single pesticide MRL ranges can be encountered which comprise several orders of magnitude posing a severe problem in multiresidue analysis. For most environmental samples and drinking water, the MRLs are generally much lower than for food samples. The EC directive for drinking water[3] states that the concentration of pesticides and related products should not exceed the level of $0.1\,\mu g\,L^{-1}$ (ppb) for individual compounds and $0.5\,\mu g\,L^{-1}$ for total pesticides. This essentially means that methods of analysis are required which are about 1000 times more sensitive than those for foodstuffs.

Table 1 Overview of pesticides not amenable to GC[a]

Pesticide	Chemical family	Pesticide	Chemical family
Abamectin	antibiotic	Fenoxaprop-ethyl	acid, phenoxy
Aclonifen	phenoxyaniline	Fenoxaprop-P-ethyl	acid, phenoxy
Aldicarb	carbamoyloxime	Fentin acetate	organotin
Amitraz	amidine	Fentin hydroxide	organotin
Amitrole	triazole	Fenuron	urea, phenyl
Azaconazole	triazole	Fluometuron	urea, phenyl
Benazolin	benzothiazole	Formetanate	carbamate
Bendiocarb	carbamate	Glufosinate	acid, phosphinic
Benomyl	benzimidazole	Glyphosate	acid, phosphinic
Bentazone	benzothiadiazole	Ioxynil	nitrile;
Benzthiazuron	urea, phenyl	Isoprocarb	carbamate
Bromoxynil	nitrile,	Isoproturon	urea, phenyl
Butocarboxim	carbamoyloxime	Linuron	urea, phenyl
Butoxycarboxim	carbamoyloxime	Maleic hydrazide	pyridazine
Carbaryl	carbamate	MCPA	acid, phenoxy
Carbendazim	benzimidazole	MCPB	acid, phenoxy
Carbetamide	carbamate	MCPP	acid, phenoxy
Carbofuran	carbamate	Mepiquat chloride	quaternary ammonium
Chlorbufam	carbamate	Metamitron	triazine
Chloridazon	pyridazinone	Methabenzthiazuron	urea, phenyl
Chlormequat	quaternary ammonium	Methiocarb	carbamate
Chlorobromuron	urea, phenyl	Methomyl	carbamate
Chlorotoluron	urea, phenyl	Metobromuron	urea, phenyl
Chloroxuron	urea, phenyl	Metoxuron	urea, phenyl
Chlorpropham	carbamate	Metribuzin	triazine
Chlorsulfuron	urea; sulfonyl	Metsulfuron-methyl	urea, sulfonyl
Chlorthiamid	thiobenzamide	Monolinuron	urea, phenyl
Cloethocarb	carbamate	Monuron	urea, phenyl
2,4-D	acid, phenoxy	Natamycin	antibiotic
2,4-DB	acid, phenoxy	Neburon	urea, phenyl
Desmedipham	carbamate	Oxamyl	carbamoyloxime
Dicamba	acid, benzoic	Paraquat	quaternary ammonium
2,4-DP	acid, phenoxy	Phenmedipham	carbamate
Dicloran	aniline (nitro-)	Promecarb	carbamate
Difenoxuron	urea, phenyl	Propamocarb	carbamate
Difenzoquat	quaternary ammonium	Propoxur	carbamate
Diflubenzuron	urea, phenyl	Pyridate	pyridazine
Dimethomorph	acid, cinnamic	Quinmerac	quinoline
Dinoseb	dinitrophenol	Rimsulfuron	urea, sulfonyl
Dinoseb-acetate	dinitrophenol	Rotenone	rotenoid
Dinoterb	dinitrophenol	Streptomycin	antibiotic
Dioxacarb	carbamate	2,4,5-T	acid, phenoxy
Diquat	quaternary ammonium	Teflubenzuron	urea, phenyl
Dithiocarb	carbamate	Thiabendazole	benzimidazole
Diuron	urea, phenyl	Thifensulfuron-methyl	urea, sulfonyl
DNOC	dinitrophenol	Thiodicarb	carbamate
Ethidimuron	carbamate	Thiofanox	carbamoyloxime
Ethiofencarb	carbamate	Thiram	dithiocarbamate
Fenobucarb	carbamate	Triclopyr	acid, pyridyloxy-acetic
2,4,5-TP	acid, phenoxy	Validamycin	antibiotic

2,4-D, 2,4-dichlorophenoxy acetic acid; 2,4-DB, 4-(2,4-dichlorophenoxy)butanoic acid; 2,4-DP, dichlorprop; 2,4,5-TP, fenoprop; DNOC, 4-6-dinitro-*o*-cresol; MCPA, (4-chloro-2-methylphenoxy)acetic acid; MCPB, 4-(4-chloro-2-methylphenoxy)butanoic acid; MCPP, mecoprop; 2,4,5-T, (2,4,5-trichlorophenoxy)acetic acid.

[a] Data from van Zoonen.[1]

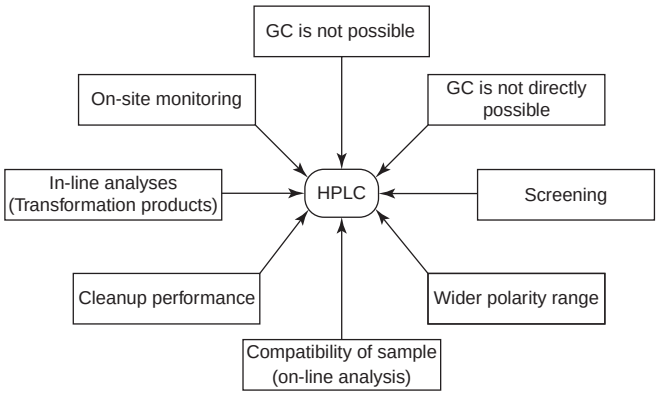

Figure 1 The involvement of HPLC in PRA.

Thus for the same group of compounds, different MRMs for different matrices and corresponding MRLs are sometimes encountered in literature.

The large variety of physico-chemical properties such as polarity, volatility, thermostability and, occasionally, molecular weight of the pesticides makes the simultaneous chromatographic separation and detection more difficult and counteracts the availability of MRMs.

Because of productivity most attention will be given to MRMs. However, the availability of fast SRMs can be attractive in this field of analysis. Especially, in survey programs based on use patterns and ad hoc requests for a certain pesticide/matrix combination, screening methods are efficient. For this type of analysis it is even better to have efficient method development strategies available.

Not all methods that have been published can be included and, hence, a comprehensive overview is not given. For such information, excellent textbooks reviewing the LC analysis of pesticides residues in foodstuffs[4] and environmental[5] samples are available.

Here, the versatility and feasibility of LC in PRA is demonstrated by making a careful selection of various types of current state methodologies encountered in PRA. After a general section on the LC analysis of pesticides, the further content is based on different LC modes. Within these modes further discrimination will be made as regards analytes and matrices.

2 REVERSED-PHASE LIQUID CHROMATOGRAPHY WITH ULTRAVIOLET DETECTION

2.1 General Aspects

RPLC is a suitable technique for the determination of polar analytes in water, since derivatization is usually superfluous and in the analysis of aqueous samples the mobile-phase system is fully compatible permitting automation of the analytical procedure. The wide application range, long-term stability, ease of use, low cost and improved selectivity (diode array) means that the combination of UV detection and RPLC is still widely used in residue analysis.

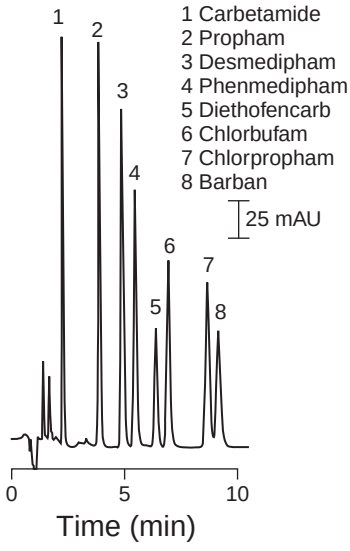

Figure 2 Performance of RPLC/UV (236 nm). Injection of 200 µL of standard carbamate pesticides in water–methanol (10 : 90, v/v) on a 100 × 4.6 mm internal diameter (ID) column packed with 3 µm Microspher C_{18}. Mobile phase, methanol–water (55 : 45, v/v) at a flow of 1 mL min^{-1}. Injected amount, about 10 ng of each pesticide on column.

The chromatogram displayed in Figure 2 nicely illustrates RPLC performance. Employing an injection volume of 200 µL, an efficient elution and separation of eight carbamate pesticides is obtained in less than 10 min employing a 100 × 4.6 mm column packed with 3 µm C_{18}, and an isocratic elution. The chromatogram of Figure 2 also shows limitations. Under these conditions the number of peaks that can be separated in one run is limited, increase of separation power by extending the length of the column will be restricted by column pressure, and the longer retention time inversely affects sensitivity, as shown in the chromatogram of Figure 2. Gradient elution enhances the total number of peaks that can be separated in one run and effectively controls band broadening of analytes. However, a severe baseline drifting usually occurs in the trace analysis employing nonselective UV detection at low wavelengths.

The problems usually encountered in RPLC/UV analyses of residues of pesticides in samples and extracts is shown by the simulated chromatogram in Figure 3. The most common problem is caused by abundantly present early eluting sample constituents indicated as S1. In many cases S1 can be largely removed via low-resolution cleanup such as off-line SPE or on-line with solid-phase

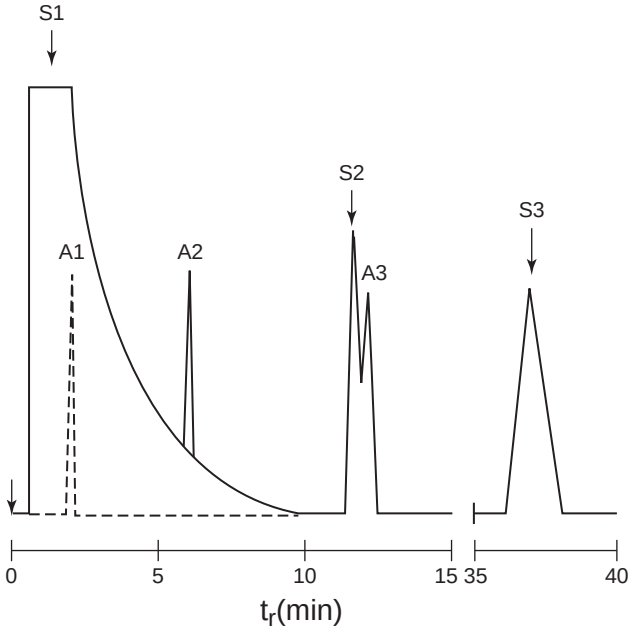

Figure 3 Simulated chromatogram of sample interferences occurring in the RPLC/UV trace analyses of pesticides. S1, S2 and S3, sample interferences; A1, A2 and A3, analytes.

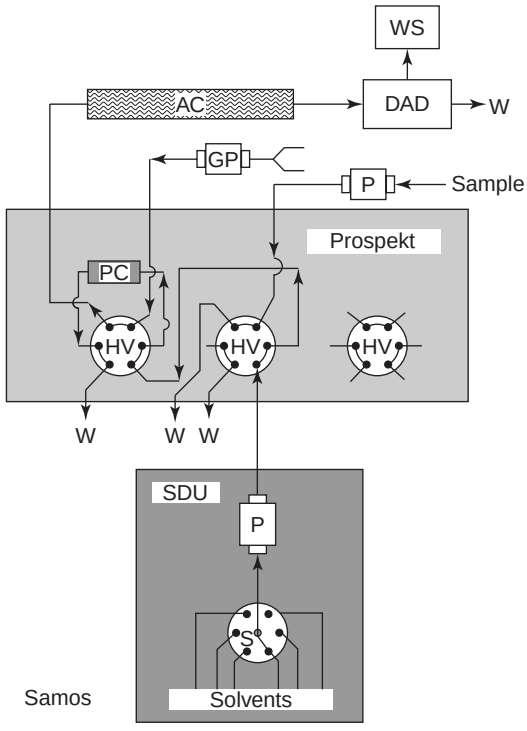

Figure 4 Scheme of SPE/LC/DAD (SAMOS). HV, high pressure valve; S, solvent selection valve; PC, trace enrichment (exchangeable) column; P, isocratic LC pump; GP, gradient LC pump; AC, analytical column; DAD, diode array detection, UV-diode array detector; WS, workstation; SDU, solvent delivery unit; PROSPEKT, valve and cartridge switching system.

extraction/liquid chromatography (SPE/LC). However, in the case of analytes with a polarity close to the range of the large excess of interfering constituents such a cleanup will not be sufficient.

Compounds with longer retention times than the analytes, S3, do not really interfere, but cause an unfavorable increase of the time of analysis. Tuning the selectivity by small changes in mobile-phase compositions, e.g. eluotropic strength or type of modifier, is usually applied to improve the separation caused by interferences of type S2.

The high compatibility between mobile phases and aqueous samples and extracts, and convenient instrumental operation provide an important feature of RPLC, that is, column switching. The use of more then one column offers the possibility to perform on-line sampling and/or cleanup which enhance sample throughput, sensitivity and/or selectivity in a cost-effective way.

In most cases, the sample preparation step is performed on a small precolumn (PC) installed in a column-switching system which permits the analytes to be desorbed automatically, after sample loading and a washing step, to the analytical LC column for separation and detection. Until the early 1990s this technique was usually named precolumn/liquid chromatography (PC/LC). Nowadays (and also here) the term SPE/LC is used for the various types of system using PCs, disposable cartridges and/or solid-phase extraction disks (SPEDs) coupled on-line to the analytical column.

In SPE/LC, PCs with dimensions of, typically, 5–10 mm (L) × 2–4.6 mm ID are usually employed. In order to enable fast sampling, PCs are packed in most instances with 10–40 μm material. The small size reduces cost, and favors a fast desorption to the analytical column. A typical setup of a fully automated solid-phase extraction/liquid chromatography/diode array detection (SPE/LC/DAD) system is shown in Figure 4. This SPE/LC set-up is called SAMOS (System for Automated Monitoring of Organic Compounds in Surface Water) and was specially developed as an early warning system for the pollution control of surface waters.[6,7]

It will be obvious that hydrophobic type SPE/LC is a highly efficient trace-enrichment procedure, but does not contribute much to increase selectivity. For the determination of some specific groups of pesticides, sorbents with a different retention mechanism such as ion exchange have been successfully applied to improve selectivity in SPE/LC.[8]

Coupled column reversed-phase liquid chromatography (LC/LC) employing a full size separation C_{18} column as the first column (C-1) is a good alternative to improve selectivity in hydrophobic column switching.[9–11] This

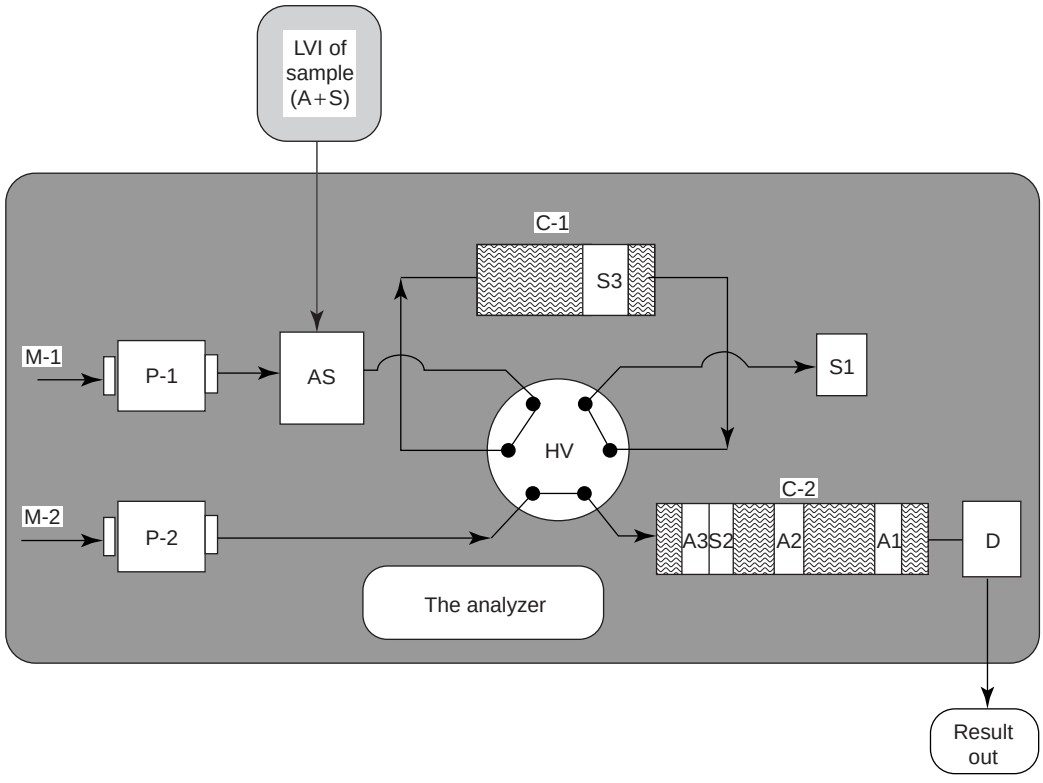

Figure 5 Coupled column reversed-phase liquid chromatography/ultraviolet (LC/LC/UV) analyzer. AS, autosampler; P-1 and P-2, LC pumps; C-1 and C-2, first and second separation column; HV, high-pressure valve; M-1 and M-2, first and second mobile phase; D, UV detector; A, analyte(s); S, sample interference(s), large volume injection (LVI) of aqueous sample.

is illustrated in Figure 5 which schematically shows the important steps involved in the heart-cutting technique. As well as using the separation power of the first column (C-1) to remove effectively interferences of types S1 and S3, this approach acts as an analyzer. The high sample throughput of LC/LC originates from the fact that sample injection is performed in the high-pressure mode by an autosampler equipped with a large volume loop and a syringe pump. The insertion by the injector of an air gap in front of and after the sample provides fast and accurate LVI with a minimal consumption of sample. This feature allows the use of rather small autosampler vials and facilitates the processing of large series of samples.

The major advantage of LC/LC is the high resolution of C-1, which contributes to the enhancement of the selectivity by removing a large part of the early eluting interferences, S1. Another crucial feature of LC/LC is the transfer volume, or the time that C-1 is coupled on-line to C-2. Obviously, the size of the transfer volume (A) will be inversely proportional to selectivity, making LC/LC most powerful in single-residue analysis. However, the key to success in the enhancement of the selectivity in the analysis of polar analytes, especially in environmental samples, is the effective removal of early eluting interferences. Therefore, the development of MRMs involving larger transfer volumes, which would lead to less selectivity, can still be quite attractive.

Starting with RPLC/UV method development for successful applications in trace analysis of pesticides, methods will be discussed highlighting the possibilities, features and limitations of RPLC, SPE/LC and LC/LC with or without off-line sample pretreatment procedures.

Next, LC methods in combination with more selective detection, FD and MS detection, will be discussed. Finally, with emphasis on cleanup, other types of LC applications will be presented.

2.2 Practical Method Development

The discussion of the important RPLC/UV parameters will be done by following the scheme of Figure 6 which is based on the author's experience, and used in LC and LC/LC methods developed for the determination of polar pesticides in environmental samples.[9–11]

The possibility of an RPLC/UV method is determined by the applicable UV absorption (wavelength, λ (nm) and molar extinction coefficient, ε ($L\,mol^{-1}\,cm^{-1}$)) and the retention (capacity factor, k) of the analyte(s) from pure water onto the C_{18}-type stationary phase. With insufficient

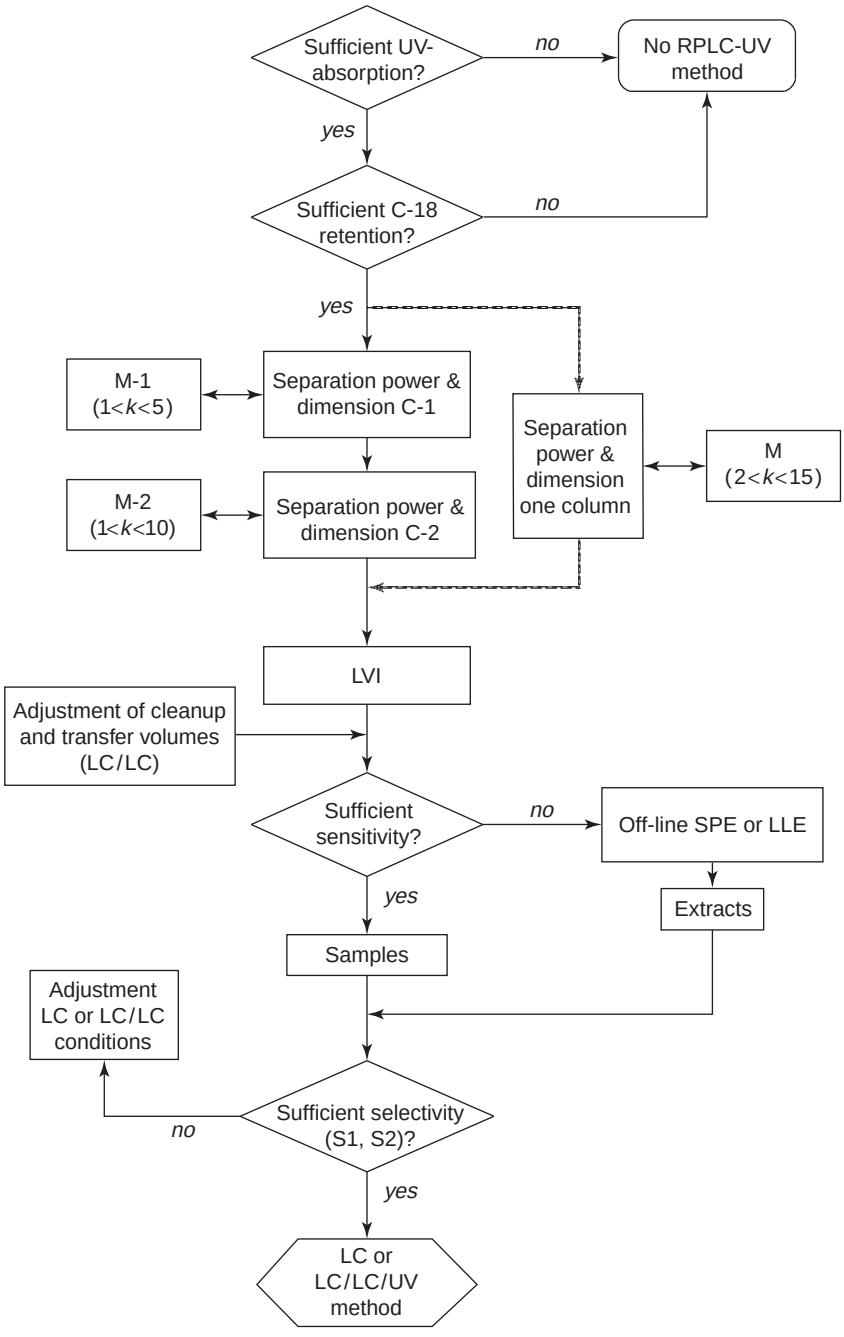

Figure 6 Scheme of straightforward LC- and LC/LC/UV method development in PRA.

ultraviolet/visible (UV/VIS) absorption/selectivity (ε values of less than ca. $1000\,\mathrm{L\,mol^{-1}cm^{-1}}$ and $\lambda < 200\,\mathrm{nm}$) and/or an essential absence of retention on a C_{18}-bonded phase (k values of less than ca. 1), the successful development of an RPLC/UV procedure for trace level analysis becomes highly unlikely.

RPLC/UV information on pesticides acquired at our laboratory is given in Table 2. In this Table the UV absorption maximum was selected as a good compromise between selectivity and sensitivity as encountered in this field of analysis. Beside UV sensitivity (ε), the peak elution volume of the analyte inversely contributes to the degree of analyte detectability. Therefore, information on the elution of the analyte, expressed as the column efficiency (N), is included in Table 2. The large difference in efficiency ranging from 280 (cypermethrin) to 5100 (chlorbufan) clearly emphasizes the importance of this parameter to sensitivity.

Table 2 RPLC/UV information on pesticides[a]

Pesticide	Modifier (%)	k	N[b]	λ(nm)	ε_0(L mol^{-1} cm^{-1})
Acephate	10[c]	2.0	1100	205	1000
Alachlor	50	5.9	1180	220	9000
Alloxydim-sodium	50	8.4	3400	260	8300
Alloxydim-sodium				300	4750
Amitraz	35	2.1	900	235	3000
Asulam	25	1.5	1000	269	15 500
Atrazine	35	3.2	2400	221	29 000
Azinphos-ethyl	50	6.2	5100	222	25 700
Azinphos-ethyl				286	7000
Azinphos-methyl	50	3.0	2900	221	24 500
Azinphos-methyl				285	6150
Barban	45[c]	9.8	4900	238	17 000
Benazolin	35	2.5	2300	217	35 000
Bentazone	35	2.6	2350	217	25 000
Benzoximate	65	5.9	4350	225	24 500
Bromacil	35	1.5	950	277	7000
Bromoxynil	40	4.8	2500	220	30 000
Carbetamide	45[c]	1.7	3000	234	17 000
Chlorbufan	45[c]	7.3	5100	237	19 000
Chlordimeform	35	2.7	1900	239	2000
Chloridazon	20	3.4	2500	229	27 000
Chloridazon				285	8500
CAAL	0	7.0	1100	205	10 000
Chlorpropham	45[c]	9.3	5100	238	17 000
Chlorpropham	50	4.5	3700	230	13 900
Chlorsulfuron	35	6.7	5100	224	25 800
Cinosulfuron	35	3.2	4600	221	24 500
Cyanazin	35	1.8	1250	220	28 500
Cypermethrin	80	4.0	280	230	30 000
2,4-D	50	1.6	700	228	8050
2,4-DB	50	2.8	1450	229	7150
Deltamethrin	80	4.5	700	230	23 300
Desmedipham	45[c]	4.8	4700	233	35 000
Dicamba	20	3.0	1350	230	7300
Dichlobenil	50	5.0	4450	230	7300
2,6-Dichlobenzamide	25	2.1	1500	220	7250
2,4-DP	35	7.1	2750	228	8500
Diflubenzuron	48	7.2	4500	260	16 000
Diethofencarb	45[c]	6.6	4500	244	18 000
Dinoseb	50	8.0	3630	271	11 600
Dinoterb	50	9.1	4000	267	11 750
Diuron	35	5.4	1260	249	21 500
DNOC	35	7.2	2200	266	12 500
Ethofumesate	55	3.2	5000	225	6000
ETU	0	1.6	1000	233	18 000
Fenfuram	40	4.1	2100	258	19 000
Fenpropathrin	80	3.7	680	230	11 350
Fenpropimorf	55[d]	13.5	2850	205	20 000
Ioxynil	40	6.4	2500	233	35 000
Iprodion	50	5.1	3750	220	21 300
Isoproturon	35	5.0	3250	240	23 700
Linuron	50	4.9	2350	248	17 500
MCPA	50	1.8	600	227	8000
MCPP	50	2.9	1400	228	7800
Metalaxyl	50	1.6	1250	220	9900
Metamitron	35	0.9	565	308	10 800
Methabenzthiazuron	35	3.6	2880	217	26 000
Methabenzthiazuron				267	13 450
Methamidophos	10[c]	1.3	1200	205	420
Methylisothiocyanate	40	2.8	2200	237	3000

Table 2 *(continued)*

Pesticide	Modifier (%)	k	N^b	λ(nm)	ε_0(L mol^{-1} cm^{-1})
Metobromuron	50	2.0	1550	250	20 700
Metoxuron	30	5.5	2500	240	16 000
Metribuzin			2350	294	9260
Metsulfuron-methyl	35	4.9	5000	226	25 500
Monuron	35	2.4	1800	245	16 400
2-Naphthoxyaceticacid	35	3.5	1850	224	44 000
1-Naphthylacetamide	35	1.6	760	220	46 000
Pencycuron	55[d]	7.0	3450	240	20 000
PCP	60	4.0	2550	220	22 000
Permethrin-I	80	5.3	1000	230	21 700
Permethrin-II	80	6.3	1340	230	21 000
Phenmedipham	45[c]	5.5	3300	237	39 000
Prochloraz	65	5.9	350	220	12 650
Prometryn	50	7.4	300	225	23 000
Propham	45[c]	3.6	5000	231	19 500
Quintozene	65	8.6	800	230	19 350
Sulfometuron-methyl	35	5.6	5100	235	20 600
2,4,5-T	50	2.8	1100	229	8700
2,4,5-TP	40	4.5	5000	230	7800
Thifensulfuron-methyl	35	4.4	4500	225	17 500
Trichlopyr	50	2.3	1200	232	8500
Vamidothion	10[c]	15.0	2000	205	5300

[a] Determined on a 50 × 4.6 mm ID packed with 3 µm C_{18} and mobile phase of acetonitrile (ACN) (modifier) −0.03 M phosphate buffer, pH 3, at a flow of 1 mL min^{-1}.
[b] Number of plates on the column measured as $(t_r/\sigma)^2$.
[c] Methanol in water.
[d] ACN in 0.2% ammonia water.
CAAL, chloroallylalcohol; ETU, ethylenethiourea; PCP, pentachlorophenol.

As indicated in Figure 6 the next step is to select adequate RPLC parameters concerning separation power and elution conditions. These parameters are determined by the dimension of the column (LC) or columns (LC/LC) and the mobile phases (M). It must be stressed that, from an efficiency point of view, one should not optimize the method beyond a certain degree: the LC system must be kept as simple as possible. For example, if sample extracts are clean enough, column-switching can be omitted. However, trace analysis requires in most cases an efficient cleanup of sample or extract.

For the on-line cleanup, C-1 must provide sufficient separation to remove a major part of the abundantly available early eluting matrix constituents, S1, prior to elution of the analyte. Therefore a column with high separation power is attractive. However, an unnecessary increase of separation power and dimension of the column must be avoided; the peak elution volume, which is inversely proportional to the sensitivity in detection of the analyte, should be kept as low as possible. An increase of the column dimension will also increase the reconditioning time (if necessary) of C-1 between analyses. In general, an eluotropic strength of M-1 should be selected such that the capacity factor of the analyte is in the range of $1 < k' < 5$, yielding a cleanup volume of at least two times the dead volume of the column (V_0) to remove a large part of the excess of early eluting interferences.

Especially in the analysis of compounds with little C_{18} retention, it is favorable to select a second column (C-2) with a higher separation power than C-1. This will provide some flexibility when optimizing the eluotropic strength of M-2 in order to reach a proper compromise between the required separation efficiency on the second column, and the desired peak compression (sensitivity of detection). It must be emphasized here, that the eluotropic strength, φ (fraction of organic of modifier), of M-1 must never exceed that of M-2 ($\varphi_{M-1} \leq \varphi_{M-2}$). During this part of the optimization one can, if required, tune selectivity by taking another type of hydrophobic column packing material or one from a different manufacturer or, in most cases firstly, by changing the mobile-phase constituents such as type of modifier, buffer and pH. The latter parameter can play a crucial role for the determination of acidic compounds in environmental water and soil samples.

In case of the on-line analysis of aqueous samples the attainable sensitivity is determined by LVI of standard solutions in HPLC-grade water. This is usually done by connecting C-1 to the UV detector which also provides the establishment of cleanup and transfer volumes. The maximum allowable sample injection volume, that is the volume that can be injected without excessive band

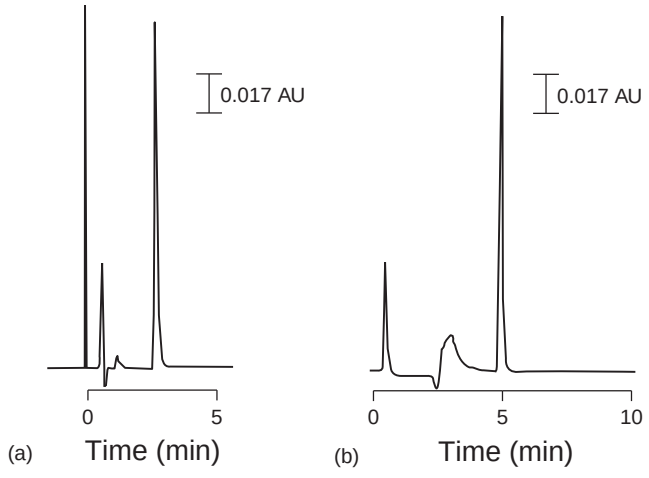

Figure 7 Performance of LVI in RPLC. Injection of standard bentazone (215 ng) in water containing 0.05% trifluoroacetic acid (TFA) (pH 2.4) on a 50 × 4.6 mm ID column packed with 3 µm Microspher C_{18} with a mobile phase of methanol–0.05% TFA in water, pH 2.4 (50 : 50, v/v) at a flow 1 mL min^{-1}. Injection volume: (a) 100 µL; (b) 2000 µL.

broadening of the analyte, largely depends on the C_{18} retention of analyte. For example, the small C_{18} retention of polar compounds such as CAAL and ETU (see Table 2) makes focusing of the analyte on the top of the column impossible, and because of migration of analyte during injection, band broadening rapidly starts to increase with increasing injection volume. In general, the start of peak deformation of the analyte upon injection on C-1 is selected as a criterion for the maximum allowable injection volume.

As demonstrated for bentazone in Figure 7, for analytes with sufficient retention ($k > 100$ in pure water as the mobile phase), RPLC allows LVI without additional band broadening. If sufficient sensitivity can not be obtained with a large volume loop injection below 8 mL, then an off-line concentration step such as SPE or liquid extraction will be necessary.

Finally, the total procedure is used to analyze real-life water samples and/or extracts. Insufficient selectivity can be improved by either re-adjustment of LC conditions or by changing the selectivity of the stationary and/or the mobile phase.

The important step to be considered in the development of on-line SPE/LC methods is the breakthrough volume of the analyte(s) on the small PC. During sample enrichment the mobile phase consists of pure aqueous sample and, therefore, breakthrough is mostly caused by insufficient retention. Breakthrough volumes can be accurately determined by measuring the UV signal of a water sample spiked with a pesticide and percolating through the PC connected to the detector.[12] A more practical way is to inject a small volume, e.g. 10 µL of concentrated standard of the pesticide in water onto the PC and determine the peak elution volume of the analyte using water as the mobile phase.

2.3 Systematic Method Development in Coupled Column Liquid Chromatography

The method development procedure discussed above is carried out in a rather straightforward way on the basis of trial-and-error experiments. The task of finding the adequate LC/LC conditions for separation between analytes and/or interference can be very difficult for the case of multiresidue analysis including compounds with a large and/or small difference in retention. For such cases, sophisticated method development procedures based on simulation and/or calculation were developed in order to avoid excessive experimental work.[13–15]

As a first step towards the development of dynamic MRMs for pesticides with LC/LC a simulation programme was developed.[13] It was based on a diffusion model; the main objective was an accurate prediction of retention times and peak volumes in order to formulate in a more effective way boundary conditions for development of MRMs using LC/LC.

Another step forward was to include the essential criteria involved in the optimization of step gradient elution in MRMs,[14] namely:

1. The first-eluting analyte must have a retention which is at least twice as large as that of the unretained compound (similar to SRM approach).
2. The total time of the chromatographic run should be kept relatively short (for sensitivity and sample throughput).
3. The resolution (Rs) between two adjacent peaks must be at least 1.2 to prevent problems due to UV wavelength switches and/or changes in the mobile-phase compositions.
4. The number of steps during gradient elution should be minimized to reduce baseline distortions and the complexity of the LC system.

Finally, a convenient and automated optimization procedure *OPTIME Version 3.1* has been developed.[15] *OPTIME Version 3.1* is based on the use of analytical equations for the prediction of chromatographic data (retention and peak volume) of analytes eluting under one- or two-step-gradient LC/LC conditions. The procedure is a constructed worksheet (Microsoft® Excel) and can be applied as a calculation and/or an optimization procedure.

In comparison to simulation, the developed calculation procedure enables a more rapid search for suitable

conditions for on-line cleanup and separation of analytes using coupled-column RPLC and is therefore highly productive and flexible.[15]

2.4 Liquid Chromatography/Ultraviolet Methods

The wide scope of both retention and detection of RPLC/UV enables fast separation and determination of many and various types of pesticides. Because of this favorable feature, RPLC/UV is frequently used in providing rather easy SRMs for the screening/monitoring of a limited number of target pesticides. In this context, RPLC/UV is attractive for use in the set-up of a screening method for recently developed polar pesticides.[16–20] The assay of the thermolabile, polar and relatively new insecticide imidacloprid in water and soil was rapidly performed with an RPLC/UV (270 nm) method using a 250 × 4.6 mm ID C_8 column with a mobile phase of methanol–water (20 : 80, v/v) at a flow rate of 1.5 mL min^{-1}. Extraction was performed with either SPE on 500 mg C_{18} columns (water) or with ACN (soil) and the analyte was determined in the uncleaned extracts to a level of 0.5 μg L^{-1} and 5 μg kg^{-1} in water and soil, respectively.[16] The simultaneous determination of six imidazolinone herbicides in natural waters was successfully carried out with off-line SPE (0.5 to 1 L of sample on 500 mg Carbograph-1) followed by RPLC/UV (240 nm); the analytes were separated on a 5 μm C_{18} column (250 × 4.6 mm) and a gradient elution with ACN, water and formic acid in water as mobile-phase constituents; the method includes the analysis of soils and confirmation of analytes with liquid chromatography/electrospray mass spectrometry (LC/ESMS).[17] A relatively new herbicide, thioazopyr and its mono-acid metabolite were assayed in groundwater with a simple method using SPE on graphitized carbon black (GCB) cartridges for sample treatment and RPLC/UV (220 nm) for instrumental analysis of the extracts obtained.[18]

Another relatively new class of herbicides are sulfonylurea herbicides characterized by low application rates and low toxicity to mammals. A number of RPLC/UV-based methods has been reported on the efficient determination of these herbicides in water and/or soil.[19,20] The combination of microwave-assisted solvent extraction (MASE) and RPLC/UV (226 nm) proved to be a convenient technique for the multiresidue analysis of sulfonylureas (5 tested) in soils down a level of at least 5 μg kg^{-1}.[19] After MASE (10 min), part (10 mL) of the organic solvent (dichloromethane–methanol, 90 : 10) was evaporated and the redissolved residues were analyzed with RPLC/UV in less than 15 min without additional cleanup. An important aspect was that samples with aged residues (60 days at 4 °C) were included in the method validation.[19]

A highly sensitive and rather selective high-performance liquid chromatography/ultraviolet (HPLC/UV) (230 nm) method for the determination of seven sulfonylureas in water was developed by Di Corcia et al.[20] Both selectivity and sensitivity were obtained by employing off-line SPE on 500 mg Carbograph 4 cartridges. This type of cartridge allows very large volume sampling (up to 4 L tested) without breakthrough of the sulfonylureas and stepwise elution of acid compounds. The latter feature avoids co-elution of humic acid interferences when using dichloromethane–methanol (80 : 20, v/v) acidified with acetic acid (10 mmol L^{-1}) as the solvent for desorption; confirmation of compounds was performed with electrospray mass spectrometry (ESMS) detection.[20] In the case of relatively large molecules, e.g. antibiotics, and/or very polar or ionic pesticides such as bipyridylium herbicides, RPLC provides an efficient separation. For example, efficient RPLC/UV methods have been developed for the residue analysis of warfarin in eggs[21] and spinosad and its metabolites in vegetables and fruits.[22]

Ion-pair/reversed-phase liquid chromatography (IP/RPLC) provides an efficient separation of ionic organic compounds. Despite the inherent LC pump wear of pistons and seals, IP/RPLC with UV detection is still the chosen technique for the determination of bipyridinium-type herbicides. Dedicated methods are available for the determination of diquat and/or paraquat in various type of matrices including mammal tissues[23] and biological fluids[24,25] and mammal tissue.[25] Because of their different UV spectra, detection is carried out at 257 and 308 nm for paraquat and diquat, respectively.

IP/RPLC with UV detection at 292 nm has been successfully applied for the simple, rapid and sensitive assay of 10 benzimidazole residues in milk[26] The method based on a pH-based liquid–liquid extraction (LLE) with ethyl acetate/buffer and the LC/UV analysis of uncleaned extracts enables determination of all compounds to a level of 5 μg L^{-1} with overall recoveries ranging between 70 and 100%.

Di Corcia et al.[27,28] developed a MRM for the monitoring of 89 pesticides in environmental waters. The group comprises base/neutral and acidic pesticides from various classes, e.g. carbamates, triazines, phenylureas, phenoxy acids and phenols. The method is based on the use of off-line SPE and HPLC/UV. In this method the role of the SPE on 300 mg Carbopack B cartridges is important. This GCB material allows sampling of water at a high-speed (120–130 mL min^{-1}) and provides selectivity by separating base/neutral from acidic analytes by means of a step-wise elution. Chromatographic analysis was performed on 250 × 4.5 mm ID columns packed with different 5 μm silica-bonded materials. For the base/neutral fraction (71 pesticides), the primary column contained a LC-C_{18}-DB (de-based) packing and the confirmation column contained a LC CN (cyano) packing. The acidic fraction (18 compounds) was separated on LC-C_{18}. The

separation of analytes was performed by optimized gradient elutions using ACN and buffered water as mobile constituents. UV detection was carried out at 220 nm and 230 nm for the determination of base/neutral and acidic pesticides, respectively. The limits of quantification (5 times the limit of detection LOD) were lower than 0.1 $\mu g\,L^{-1}$ for 84 pesticides and recoveries ranged between 80 and 105% with relative standard deviations (RSDs) below 10% for recoveries at low levels. The procedure was further modified by extending the analysis to very polar pesticides and metabolites by using 1 g GCB cartridges designed to perform re-extraction of the analytes in the back-flush mode.[29]

A general problem encountered in the RPLC/UV analysis of acidic pesticides in environmental samples is the presence of a broad hump in the chromatogram caused by the presence of co-extracted humic and fulvic acids. As will be discussed below, column-switching procedures employing sorbents with a different selectivity are available to improve significantly the chromatographic analysis. An effective off-line SPE procedure has been presented[30] employing cartridges packed with 200 mg of SDB-1 sorbent (styrene–diphenylbenzene). Beside the sufficient retention capacity of polar pesticides such as oxamyl and triazine metabolites deisopropylatrazine (DIA) and deethylatrazine (DEA) this material also retained the tested acidic pesticides (7 compounds) in their ionic form. This feature allowed sampling (500 mL) of water at pH 7 and eliminated largely the humic interferences making it possible to determine with RPLC/UV (220 nm) a group of 17 various pesticides in river water to a level of 0.1 $\mu g\,L^{-1}$.

Immuno-affinity-based extraction procedures offer high selectivity. This is clearly demonstrated by Pichon et al.[31,32] who developed immunosorbents for the selective SPE of phenylurea and triazine herbicides from environmental waters. The important advantage of the off-line immunosorbent is that it performs trapping of the majority of analytes of a chemical class (phenylurea, 9 out of 13; triazines, 6 out of 9), extending the scope of immunomethodology towards multiresidue analysis. The high selectivity is illustrated in Figure 8 showing the RPLC/UV (244 nm) analysis of phenylurea herbicides in extracts of surface water obtained after 200 mL of preconcentration on the immunosorbent cartridge.[31]

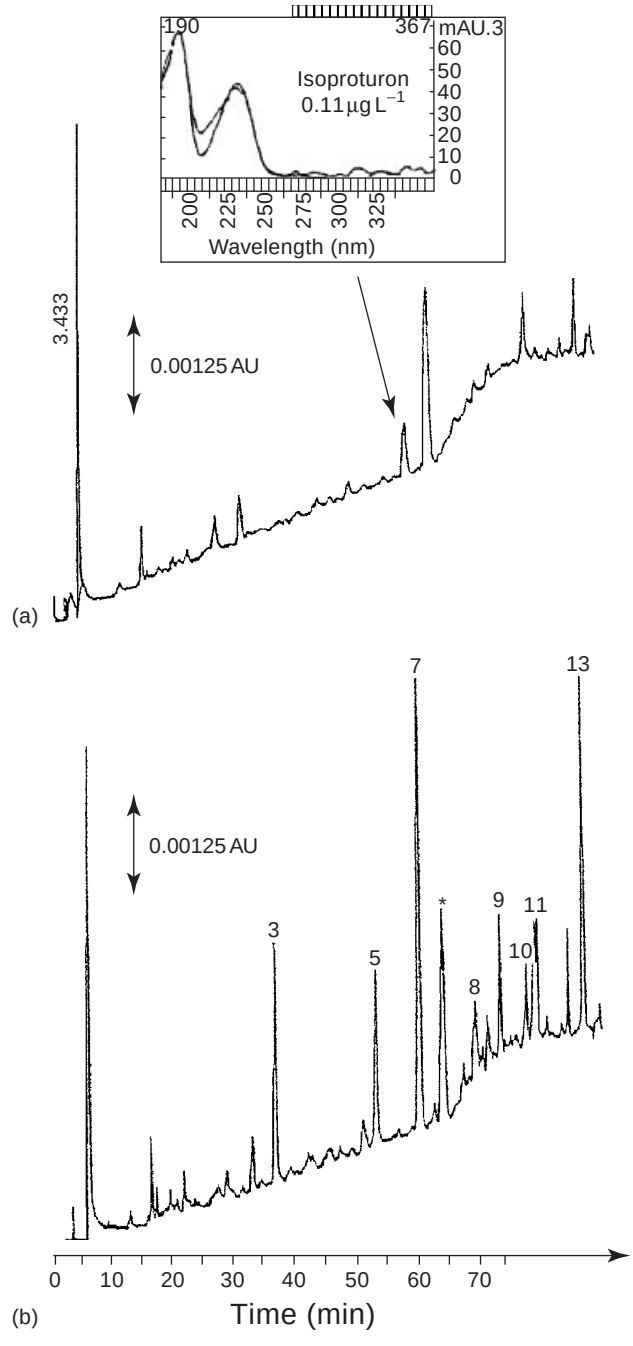

Figure 8 Off-line immuno-assay and RPLC/UV (244 nm) for the preconcentration of 200 mL of Seine River water (a) nonspiked and (b) spiked with 0.5 $\mu g\,L^{-1}$ of each of phenylurea herbicide on a cartridge containing silica bonded with anti-isoproturon antibodies. LC conditions: 250 × 4.6 mm ID 5 µm Supelcosil LC-18-DB column; ACN–water gradient elution consisting of 20 to 35% modifier from 0 to 52 min, 35 to 70% from 52 to 80 min, and 70 to 100% from 80 to 85 min. Peak assignment: (3) monuron, (5) chlortoluron, (7) isoproturon, (8) difenoxuron (9) buturon, (10) linuron, (11) chlorbromuron, (13) neburon. (Reprinted with permission from Pichon et al., *Anal. Chem.*, **67**, 2541–2460. Copyright © 1995, American Chemical Society.)

2.5 Solid-phase Extraction/Liquid Chromatography/Ultraviolet Methods

Fully automated SPE/LC methods using on-line trace enrichment, gradient elution and UV DAD were firstly directed to the multiresidue analyses of polar pesticides in various types of water.[6,33,34] Trace enrichment (100–150 mL of sample) was usually carried out with

an automated PC exchange system (Prospekt) using 10 × 2.0 mm ID disposable cartridges packed with 20 μm polymeric reversed-phase spherical packing (PLRP-S). Separation of analytes was accomplished on a 250 × 4.6 mm ID column packed with 5 μm C_{18} (DB) and gradient elution with ACN and 0.03 M phosphate buffer (pH 3) as mobile phase constituents. The effectiveness of DAD in providing very well matched UV spectra for confirmation of pesticides and or metabolites in European rivers, traced at levels between 0.05 and 1.4 μg L^{-1} has been clearly demonstrated.[6,7,33] Direct automated evaluation of data integrated in the fully automated LC-system resulted in an early warning system called SAMOS which has been successful in the determination of 27 organic micropollutants at concentrations of 1 μg L^{-1} in surface waters.[6,7] It is interesting to note here that for compounds amenable to GC the same automated sample pretreatment procedure was used, providing a solid-phase extraction/gas chromatography (SPE/GC) based SAMOS.[35]

The SPE/LC SAMOS approach was further optimized by a more efficient separation of analytes involving the use of a 150 × 4.6 mm ID column packed with 3.5 μm Zorbax SB-C_{18}, a column temperature at 43 °C;[36] buffering (pH of 7) of the mobile phase with ammonium acetate significantly improved the peak shape of some triazines.[36] The improved conditions allowed the automated SPE/LC/DAD analysis of triazine and phenylurea herbicides in raw and drinking water with LODs ranging between 0.002 and 0.012 μg L^{-1}. Confirmation with UV DAD was accomplished, even at levels near to the LODs. Requiring an analytical run time of 65 min for adequate separation of analytes and the analyses of standards, the system is able to process about 10 samples per day.

Up to about 110 compounds representing many classes of polar compounds were determined on-line with SPE/LC/DAD.[37] This study emphasized that beside important parameters such as sample pH, column temperature and elution conditions, the type of LC column was found to be the most important parameter. From the six columns tested the 150 × 4.6 mm ID column packed with 3.5 μm Zorbax SB-C_{18} provided the highest selectivity/sensitivity. The very good performance of DAD confirmation of phenolic compounds in case of base-line distortion caused by humic acid interferences is shown is Figure 9. Decrease of the base-line distortion in the first part of chromatogram was achieved by applying on-line dual PCs (both 10 × 2 mm ID, packed with 15–25 μm PLRP-S). With this approach, the highly polar compounds which showed breakthrough on the first PC were trapped on the second one while most of the humic substances are retained on the first one.

In comparison to PCs, hydrophobic membrane extraction disks (SPEDs) appeared to be a good alternative with respect to sorption, capacity, back-pressure and stability

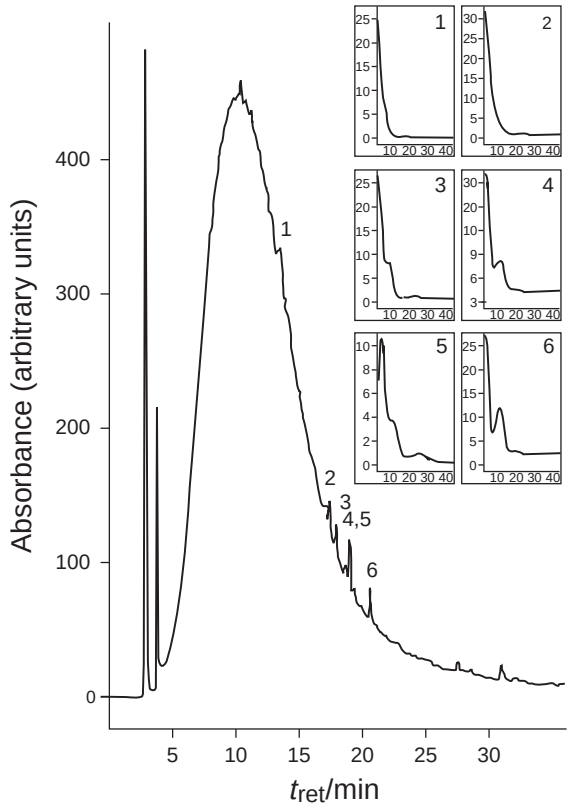

Figure 9 SPE/LC/DAD/UV (215 nm) chromatogram of a surface water sample spiked with dichlorobenzoic acids and bentazone at the 1 μg L^{-1} level. Peaks: (1) 2,6-dichlorobenzoic acid, (2) 2,3-dichlorobenzoic acid, (3) 2,5-dichlorobenzoic acid, (4) 2,4-dichlorobenzoic acid, (5) bentazone, (6) 3,4-dichlorobenzoic acid. Analytical conditions: 10 × 2 mm ID 15–25 μm PLRP-S SPE cartridge; sampling of 50 mL of water sample acidified with hydrochloric acid to pH 2; 250 × 4.6 mm ID 5 μm Zorbax-RX column; LC gradient of ACN–10 mM phosphate buffer of pH 2 from 10% to 80% modifier in 30 min; column oven temperature, 40 °C. Insets: UV absorbance spectra of analytes. (Reproduced with permission from Hogenboom et al., *Analyst*, **122**, 1371–1377. Copyright © 1997, The Royal Society of Chemistry.)

after repeated use.[38–40] A simple leak-tight membrane disk holder offering flexibility in choosing type, size and number of disks was developed to perform on-line operation.[38,39] Solid-phase extraction disk/liquid chromatography/ultraviolet/diode array detection (SPED/LC/UV/DAD) methods allowed the rapid trace level determination of three basic pesticides (carbendazim, chloridazon, simazine)[39] and nine neutral pesticides (carbamates, phenylureas)[40] in river and drinking water, respectively.

The simultaneous on-line determination of the diquat, paraquat and difenzoquat in water at levels of 0.1 μg L^{-1} was carried out with SPE/LC/UV.[41,42] Sampling and separation of analytes was performed on Hypercarb-GCB[41] or bare silica material[42] employing sampling volumes of 50 (on GCB) or 200 mL (on silica), respectively.

A distinct advantage of on-line SPE systems is to improve selectivity by using SPE columns packed with an analyte(s) trapping sorbents of a different separation mechanism. One must take care that the desorption kinetics are fast enough to avoid a large loss of efficiency. For example, this drawback occurs in combining carbon materials, which have the best enrichment properties, with a C_{18} analytical column.[43] Methods based on use of selective SPE sorbents have been developed for the fully automated analysis of chlorophenoxy acid herbicides in water by using either a PRP-1 and anion-exchange PC in series[8] or a PC in combination with pH step-gradient from 2.8 (sampling) to 8.8 (desorption) followed by ion-pair LC separation at pH 8.8.[44] Both methods allowed the determination of the acidic pesticides in tap water at levels between 15 and 30 ng L^{-1}. For the on-line SPE/LC determination of a group of phenolic and neutral pesticides, interference of humic acid was significantly decreased by using a modified polymeric resin (*o*-carboxybenzoyl moiety) in combination with addition of sodium sulfite salt to the water sample.[45]

A new development in improving selectivity of the on-line SPE/LC techniques is the use of pressure resistant immunosorbents for trace enrichment. Especially, the availability of material providing sufficient cross-reactivity to pesticides of a same chemical class makes this technique more attractive for use in multiresidue analysis. Based on this approach methods have been developed for the simultaneous on-line trace analysis of phenylurea herbicides in environmental water using a silica-based immuno-[46] or mixed-immuno silica-based sorbent.[47]

Recently, Kok and Hiemstra[48] presented an overview of their developed HPLC methods[1,48–51] used in daily practice for the efficient screening of polar and/or thermolabile pesticides in fruits and vegetables. The five SPE/LC based methods are schematically presented in Figure 10 and cover a total of about 100 pesticides including the classes: benzoylphenylureas,[49] benzimidazoles,[50] conazoles,[48] carbamates,[51] and phenylureas.[1,48] Study and validation demonstrated[48] that the efficient and miniaturized extraction (see Figure 10) used for the GC MRMs[1] was also applicable for extraction of the polar pesticides. Each SPE/LC procedure was fully automated by means of a commercially available SPE apparatus (ASPEC). Concerning carbamates and phenylureas more information about method performance will be given in section 3.2. The separation of benzimidazoles (carbendazim/thiabendazole) and conazoles (imazalil/prochloraz) were performed on a 150 × 8 mm ID column packed with Shodex Rspak DE-613 polymer material and a 250 × 4.6 mm ID column packed with Inertsil ODS-3, respectively. The mobile phase consisted of 70% ACN (conazoles) or methanol (benzimidazoles) in phosphate buffer pH 7. The SPE cleanup on the diol cartridges appeared to be very effective. It allowed UV detection at 210 nm of imazalil and prochloraz at a level of 0.1 mg kg^{-1} without the occurrence of interfering peaks from the various types of blank commodies (foodstuffs) investigated. Also for carbendazim and thiabendazole this cleanup provided high selectivity.[50]

Benzoylphenylureas were separated on a 150 × 3 mm ID column packed with 4 µm Superspher 60 RP-8 and using an ACN–water gradient of 15 min at flow of 0.5 mL min^{-1}. The good performance of the LC method is shown in Figure 11 for the analysis of an extract of a green pepper spiked with analytes at a level of 0.1 mg kg^{-1}. Additional confirmation was obtained with DAD and unambiguous identification was done with GC/MS using either electron impact (EI) and/or chemical ionization (CI) mode. The limit of determination of the method was 20 µg kg^{-1} for each compound. All LC methods allowed the determination of these pesticides far below the MRLs.

2.6 Coupled Column/Ultraviolet Methods

An overview of on-line LC/LC/UV methods[9–11,52–57] for the trace analysis of various pesticides and related compounds in water samples is made in Table 3. The simplicity and effectiveness of the LC/LC/UV analyzer (see Figure 5) provides selective methods with a high sample throughput (run times of about 10 min). As discussed above and illustrated by the data given in Tables 2 and 3 the LVI in case of very polar analytes is limited. Hence, for compounds like CAAL and ETU, LVI is limited to 0.2–0.5 mL yielding LODs in the range of 1 µg L^{-1}. By the use of relatively simple manual concentration steps the LODs can be lowered to the EC drinking-water limit of 0.1 µg L^{-1}.[52–54] For compounds with sufficient C_{18} retention and UV detection (see Table 2) LODs of 0.1 µg L^{-1} can be obtained employing sample injections below 8 mL (see data of Tables 2 and 3). Figure 12 clearly demonstrates the performance of LC/LC/UV analysis: the trace analysis of isoproturon in surface water in less than 8 min. It also shows the significant improvement in selectivity of LC/LC in comparison to the use of two coupled C_{18} columns without column switching, a phenomena also encountered in ion chromatography.[58]

A rather unexpected matrix-dependent retention time was reported in the simultaneous determination of bromacil and in diuron employing LVI LC/LC/UV.[55] Investigating LC grade, drinking and surface water differences in retention times of about 1 min were observed indicating that one must consider this effect in LC/LC method development. Nevertheless, the LC/LC method is distinctly more efficient in comparison to the method previously in use, which employed a laborious sample pretreatment step.[59]

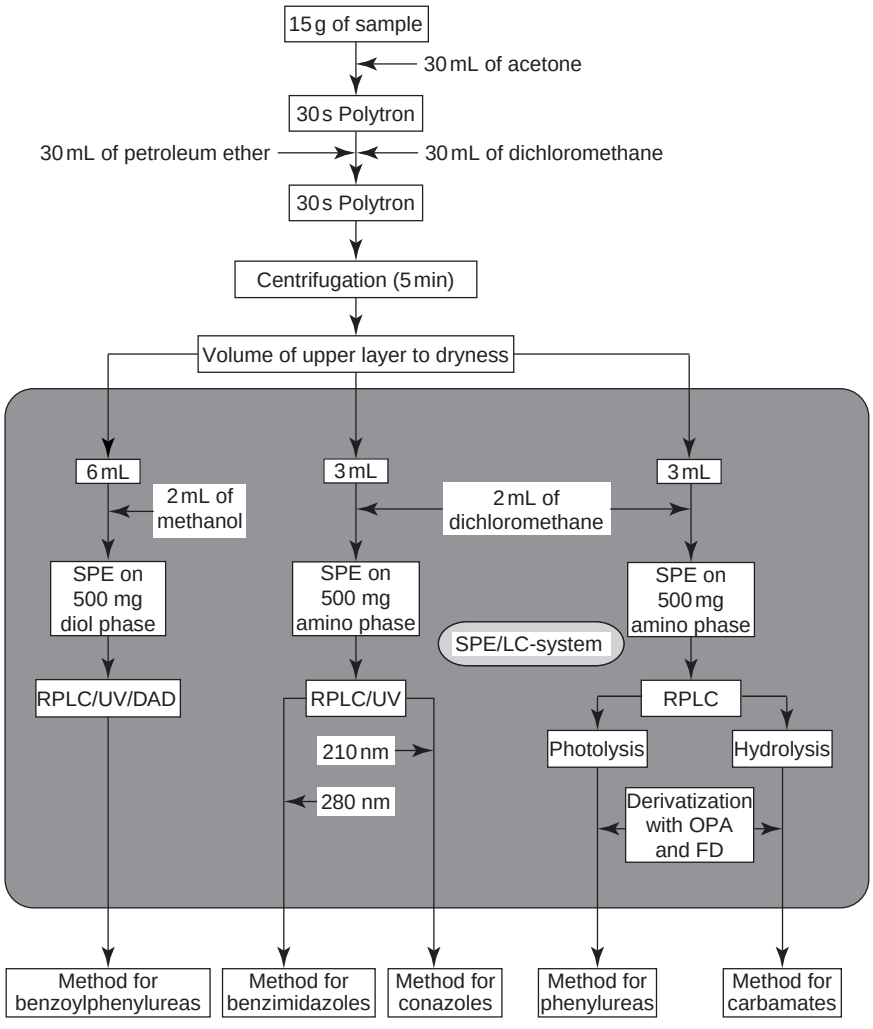

Figure 10 Scheme of MRMs as applied for the determination of polar pesticides in fruits and vegetables and according to refs. 48–51.

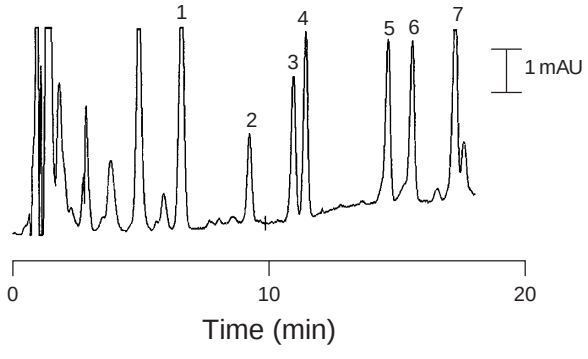

Figure 11 RPLC/UV (260 nm) of an extract of a green pepper sample spiked with benzoylphenylurea pesticides at a level of 0.1 mg kg^{-1}. Sample pretreatment is given in Figure 10. LC conditions: 150 × 3 mm ID 4 µm Superspher 60 RP-8 column with an ACN–water gradient elution from 25 to 50% in 15 min. Peaks: 1. diflubenzuron; 2. triflumuron; 3. triflubenzuron; 4. hexaflumuron; 5. lufenuron; 6. flufenoxuron; 7. chlorfuazuron.

The SRMs shown in Table 3 make use of small transfer volumes, typically 0.2–0.9 mL, in order to enhance selectivity. Although expected, the MRMs requiring larger transfer volumes[57] did not offer sufficient selectivity. This emphasizes that in LC/LC the performance of an efficient cleanup on C-1, which means the removal of S1, is a predominant selectivity parameter for this type of analysis. The performance of on-line multiresidue analysis of triazines in surface water at a level of 0.4 µg L^{-1} is given Figure 13.

The examples discussed above clearly demonstrate the main advantage of on-line LC/LC/UV: the fast screening of a wide variety of compounds in environmental water samples. For analytes with sufficient UV absorption/selectivity (ε and λ values of more than 15 000 L mol^{-1} cm^{-1} and 220 nm, respectively) and C_{18} retention ($k > 50$, in pure water) LODs in the range of 0.1 µg L^{-1} are feasible. However, owing to its maximum

Table 3 Conditions and performance of LC/LC/UV methods for the on-line analysis of pesticides in water

Analyte	C-1[a]	C-2[a]	M-1	M-2	Injection volume (mL)	Cleanup volume (mL)	Transfer volume (mL)	UV (nm)	LOD ($\mu g\,L^{-1}$)	Time of analysis (min)	Ref.
Bentazone	A	B	MeOH–buffer,[b] pH 2.3 (50:50)	MeOH–buffer,[b] pH 2.7 (50:50)	2.0	4.65	0.45	220	0.1	8	9
Isoproturon	A	B	ACN–water (48:52)	ACN–water (48:52)	4.0	5.85	0.40	240	0.1	10	9
Linuron	A	B	ACN–water (50:50)	ACN–water (50:50)	4.0	6.60	0.40	249	0.1	12	9
Methabenz-thiazuron	A	B	ACN–water (42:58)	ACN–water (42:58)	4.0	5.90	0.45	267	0.1	8	10
ETU	C	C	ACN–water[c] (1:99)	ACN–water[c] (1:99)	0.20	2.60	0.44	233	1.0	5	54
CAAL	A	B	Water	MeOH–water (5:95)	0.20	1.20	0.80	205	1.0	7	52
Methyliso-thiocyanate	A	B	ACN–water (40:60)	ACN–water (50:50)	0.77	1.90	0.40	237	1	7	53
Metamitron	C	C	ACN–water[c] (25:75)	ACN–water[c] (30:70)	6.0	10.4	0.15	310	0.1	15	11
Bromacil, diuron	D	B	MeOH–water (30:70)	MeOH–water (65:35)	2.0	5.00	0.90	277	0.1	10	55
Simazine, atrazine, terbuthylazine, terbutryn	D	B	ACN–water (40:60)	ACN–water (70:30)	2.0	2.90	0.70	223	0.1	7	56
Atrazine, DIA, hydroxy-atrazine and DEA	D	E	ACN–water (20:80)	ACN–water (35:65)	2.0	2.60	4.20	220	0.3	12	57

[a] Type of columns: A, 50 × 4.6 mm ID packed with 3 µm Microspher C_{18}; B, 100 × 4.6 mm ID packed with 3 µm Microspher C_{18}; C, 150 × 4.6 mm ID packed with 5 µm Hypersil ODS; D, 30 × 4.6 mm ID packed with 5 µm Spherisorb ODS-2; E, 125 × 4.6 mm ID packed with 5 µm Spherisorb ODS-2.
[b] 0.1% phosphoric acid in water.
[c] Containing 0.2% of ammonia.

sample injection volume of about 8 mL the on-line LC/LC/UV approach is not viable for pesticides with sufficient C_{18} retention but having ε values below about 14 000, e.g. chlorophenoxy acid herbicides (see Table 2).

2.7 Coupled Column/Ultraviolet Methods Employing Off-line Solid-phase Extraction

In order to employ the favorable LC/LC/UV features (efficient removal of S1) an off-line preconcentration step will be necessary for the trace analysis of compounds with insufficient UV sensitivity (see above).

Besides enhancement of sensitivity, off-line extraction procedures can offer distinct advantages above on-line procedures in this field of analysis. For example, organic extracts (concentrated) obtained by off-line procedures, e.g. LLE or SPE may be stored over a longer period of time so that they can be analyzed in batches. This is more efficient in case of monitoring studies involving small series of samples with a low frequency of sample collection over a long period of time.

In studies involving the determination of pesticides of different chemical families, operational MRMs as applied in monitoring programs are often applied in combination with GC-based MRMs. It is then often more convenient to use a unified extraction/concentration procedure with an organic solvent or (semi) automated SPE, so that the same extract can be used for all MRMs to be applied on a sample.

A comparison between LVI on-line LC/LC/UV and off-line SPE combined with LC/LC/UV for the trace analyses of triazines and some metabolites in water samples showed similar results: on-line LC/LC/UV favors speed of analysis while a simple preceding off-line SPE step provides LODs in the range of 0.02–0.05 $\mu g\,L^{-1}$.[57] Another comparative study included the use of hydrophobic on-line SPE/LC (PC/LC).[60] It appeared that in comparison to LVI LC/LC/UV and off-line SPE combined with LC/LC/UV, PC/LC favors sensitivity and speed of analysis at low levels ($<0.1\,\mu g\,L^{-1}$) but selectivity is considerably less.

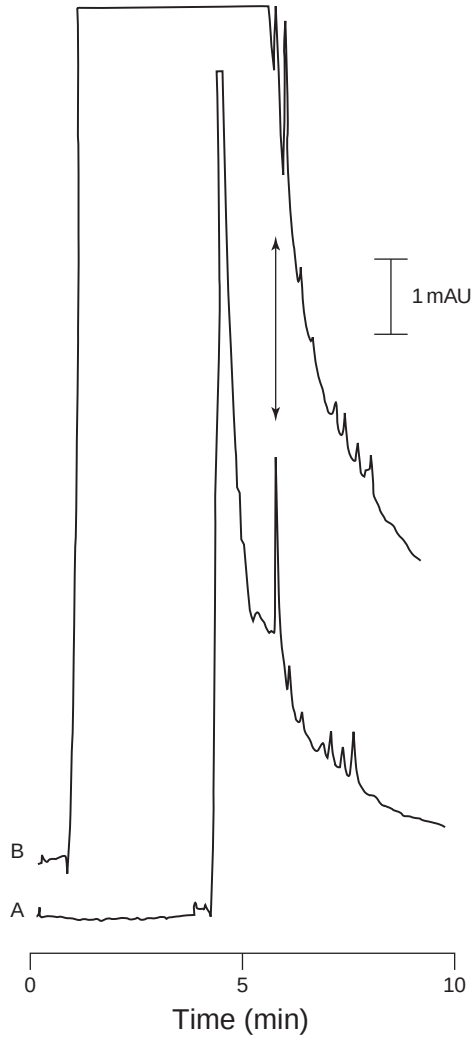

Figure 12 LC/LC/UV (244 nm) of surface water sample spiked with isoproturon at a level of 0.5 μg L^{-1}. LC/LC conditions, see Table 3.[9] A, chromatogram obtained with column switching; B, chromatogram using the two columns in series without column switching.

Off-line SPE and LC/LC/UV were successfully applied in an interlaboratory study involving the determination of phenylurea herbicides spiked at levels of about 0.1 μg L^{-1} to drinking water, groundwater and surface water. The SPE procedure involved the sampling of 250 mL of water on a 500 mg C_{18} SPE cartridge preconditioned before use with 3 mL of methanol, 3 mL of acetone, 3 mL of methanol and 6 mL of LC-grade water. After sample loading, the cartridges were dried by passing air for 30 min and the analytes were eluted with 2 mL of acetone. An aliquot of the acetone corresponding to 125 mL of sample was transferred in a tube and evaporated to dryness with a gentle stream of nitrogen. The residue was redissolved by adding first 0.20 mL of ACN followed by 1.8 mL of LC-grade water. The LC/LC/UV analysis displayed in

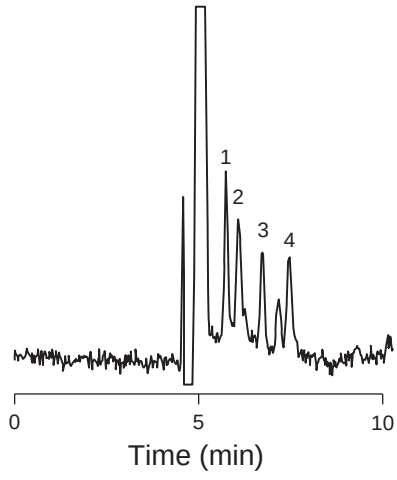

Figure 13 LC/LC/UV (233 nm) of a surface water sample spiked with triazine herbicides at a level of 0.4 μg L^{-1}. LC conditions, see Table 3.[56] Peaks: (1) simazine, (2) atrazine, (3) terbuthylazine, (4) terbutryn.

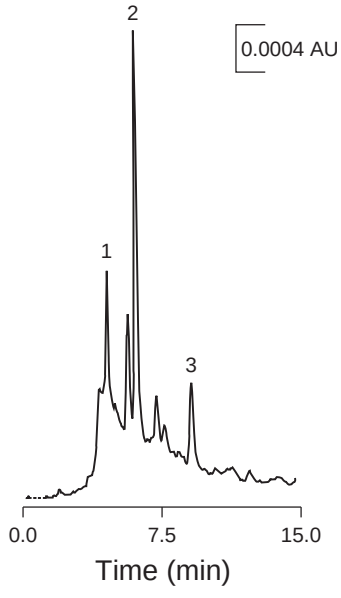

Figure 14 Off-line SPE and LC/LC/UV (244 nm) of an SPE extract of a surface water containing (1) monuron (0.18 μg L^{-1}), (2) monolinuron (0.70 μg L^{-1}) and (3) diuron (0.20 μg L^{-1}). LC conditions: C-1, 3 μm Microspher C_{18} (50 × 4.6 mm ID); C-2, 3 μm Microspher C_{18} (100 × 4.6 mm ID); M-1 and M-2, methanol–water (55 : 45, v/v) both at 1 mL min^{-1}; injection 150 μL of SPE extract (procedure, see Text); cleanup volume, 1.3 mL; transfer volume, 3.4 mL.

Figure 14 of a surface water sample containing monuron, monolinuron and diuron at a level of 0.18, 0.70 and 0.20 μg L^{-1}, respectively nicely demonstrates the good performance of this approach on both selectivity and sensitivity.

A new development in the determination of acidic compounds in environmental samples is the beneficial

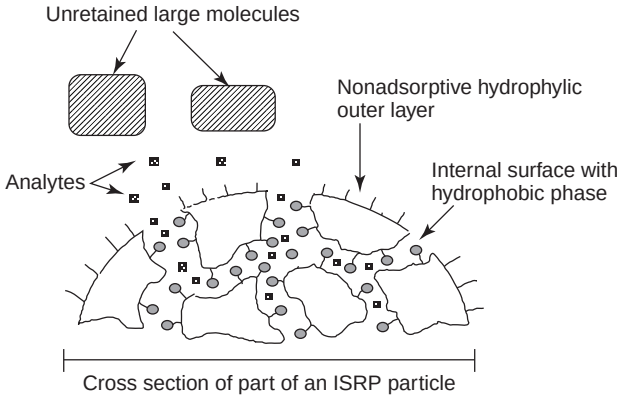

Figure 15 Schematic presentation of retention mechanism of the 5 µm GFF-II ISRP material. Outer layer, glycine (polar); Inner layer, glycine–phenylalanine (nonpolar).

use of analytical columns packed with restricted access material (RAM). As schematically presented in Figure 15 for internal surface reversed-phase (ISRP) material, efficient separation is obtained between small molecular analytes and large molecular substances. Originally, RAM columns were developed for the on-line processing of body fluids such as serum and blood providing retention for analytes but eluting large molecules, e.g. proteins, in the void column volume.[61–63]

The RPLC/UV analysis of acidic pesticides (small molecules) in environmental samples is usually severely hampered by co-extracted humic substances (large molecules) which show up as a broad hump in the chromatogram. Indeed, improved chromatographic performance was obtained for determination of a group of chlorophenoxy acid herbicides in environmental waters by SPE/LC/UV (220 nm) employing a PC packed with ISRP 5 µm GFF-II, RAM.[64] Unfortunately, the ISRP PCs of different batches showed poor reproducibility in the required retention of the analytes, hence providing insufficient selectivity.

Improved results were obtained when employing RAM columns in the LC/LC mode, that is, using analytical RAM columns. In a recent comprehensive study[65] the applicability of several commercially available analytical RAM columns in LC/LC/UV for the trace analysis of acidic herbicides in water samples was investigated. Different LC configurations were studied including the single RAM column mode and column switching LC/LC employing one RAM column in combination with an analytical C_{18} column or two RAM columns. Involving a group of acidic herbicides representing different chemical families, both the SRM and MRM approach were tested by analyzing SPE C_{18} extracts of reference water samples spiked with the analyte(s) at the 0.5–1.0 µg L^{-1} level and containing dissolved organic carbon (DOC) between 3 and 12 mg L^{-1}. Information on the type of RAM of columns studied and adequate coupled column conditions is given in Table 4.

As regards SRM the use of one analytical RAM or two different types of RAM columns sufficiently improved the resolution between the analyte and the interference of humic acids to perform trace analysis in the SPE extracts with a preceding cleanup. Figure 16 demonstrates the effectiveness of the LC/LC approach involving an ISRP column as the C-1 for the single residue analysis of the polar metsulfuron-methyl herbicide at a level of 1 µg L^{-1}.

Table 4 Overview of LC/LC methods using analytical RAM columns for the trace analysis of acidic herbicides in uncleaned SPE extracts[a] of DOC containing (6–12 mg L^{-1}) water samples

Method	Herbicides	C-1[b] and M-1[c]	C-2[b] and M-1[c]	Cleanup/transfer volume (mL)
SRM	MCPP	C_{18} (50 × 4.6 mm ID); 60%	SPS (150 × 4.6 mm ID); 60%	3.6/1.0
		SPS (50 × 4.6 mm ID); 60%	Hisep (50 × 4.6 mm ID); 60%	4.0/1.0
		Hisep (50 × 4.6 mm ID); 60%	C_{18} (100 × 4.6 mm ID); 60%	3.5/1.0
		ISRP (50 × 4.6 mm ID); 40%	C_{18} (100 × 4.6 mm ID); 60%	5.0/1.0
SRM	Metsulfuron methyl	Hisep (50 × 4.6 mm ID); 40%	C_{18} (50 × 4.6 mm ID); 40%	2.8/0.8
		ISRP (50 × 4.6 mm ID); 40%	C_{18} (50 × 4.6 mm ID); 40%	2.2/1.0
MRM	Bentazone, bromoxynil	C_{18} (50 × 4.6 mm ID); 35%	SPS (150 × 4.6 mm ID); 40%	2.9/0.9
MRM	Bentazone, bromoxynil, metsulfuronmethyl, MCPA, MCPP	ISRP (50 × 4.6 mm ID); 40%	C_{18} (50 × 4.6 mm ID); 60%	2.3/2.9
		ISRP (50 × 4.6 mm ID); 40%	SPS (50 × 4.6 mm ID); 55%	2.2/3.6
		C_{18} (50 × 4.6 mm ID); 52.5%	SPS (50 × 4.6 mm ID); 52.5%	1.9/5.9

[a] Procedure. Passing of 250 mL of water sample (brought to pH 2.2 with HCl) through a 500 mg C_{18} cartridge. Desorption with 2 mL of acetone. After evaporation acetone, residue is redissolved in 0.4 mL methanol and 1.6 mL of 0.1 TFA in water.
[b] Packing materials: C_{18}, 3 µm Microspher C_{18} (Chrompack); SPS, 5 µm SPS-5PM-S5-100-ODS (Regis); ISRP, 5 µm Pinkerton ISRP GFF-II-S5-80 (Regis); Hisep, 5 µm Hisep (Supelco).
[c] Modifier content in mobile phase consisting of methanol–0.03 M phosphate, pH 2.4 (v/v).

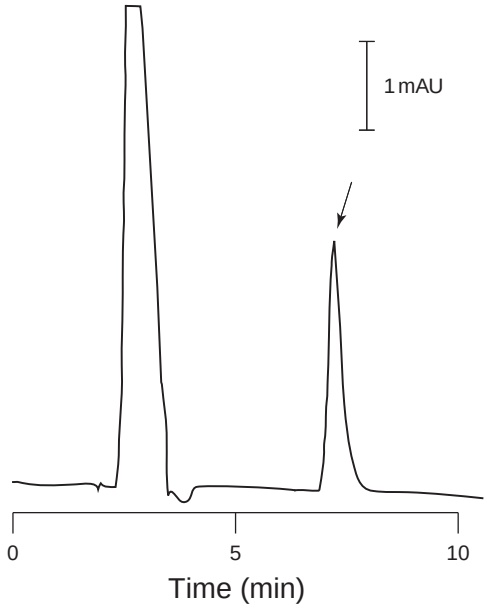

Figure 16 LC/LC/UV (220 nm) using the ISRP/C_{18} column combination for the analysis of a SPE extract of a water sample containing 12 mg L^{-1} DOC and spiked with metsulfuron-methyl at a level of 1 µg L^{-1}. Injection volume, 300 µL. SPE and LC/LC conditions, see Table 4.

Both ISRP and Hisep RAM columns provide a very efficient preseparation between acidic analytes and humic interferences. However, these columns showed, in comparison to C_{18}, a different separation mechanism which resulted in a reversed elution order and a significant increase of band broadening for some of the analytes. This feature can be disadvantageous in the MRM of pesticides of various classes. The good performance of LC/LC/UV involving the C_{18}/SPS analytical column combination is demonstrated in Figure 17 showing the simultaneous trace analysis of a group of acidic pesticides from different classes in water containing a DOC content of 6 mg L^{-1} and spiked with the analytes at the 0.5–1.0 µg L^{-1} level.

In comparison to water, soil will contain more interferences. Hence, cleanup will be even more important in the analysis of soil extracts, therefore LC/LC is an attractive technique to perform efficient on-line cleanup. In a monitoring program aimed at the occurrence of the more persistent pesticides, two LC/LC/UV based screening methods were developed for the analysis of the pesticides pencycuron and fenpropimorf in soil.[9] A simple sample extraction/concentration procedure was applied employing an overnight standing of soil in ACN followed by a filtration and a concentration step. The gain in selectivity provided by LC/LC is demonstrated in Figure 18 which shows the RPLC/UV (215 nm) analysis of soil spiked with 50 µg kg^{-1} of fenpropimorf.

As has been demonstrated recently, the use of an analytical RAM column in LC/LC/UV is also favorable

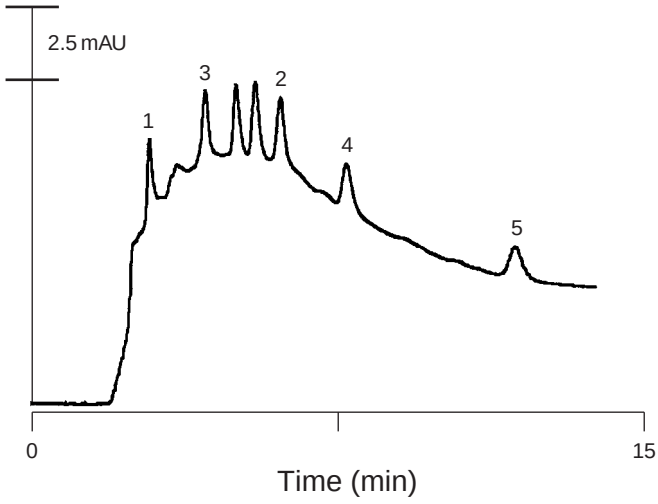

Figure 17 LC/LC/UV (220 nm) using the C_{18}/SPS column combination for the analysis of a SPE extract of a water sample containing 6 mg L^{-1} DOC and spiked with acidic pesticides at the 0.5–1.0 µg L^{-1} level. Injection volume, 300 µL. SPE and LC/LC conditions, see Table 4. Analytes (and levels): (1) metsulfuron-methyl (0.5 µg L^{-1}); (2) bromoxynil (0.5 µg L^{-1}); (3) bentazone (0.5 µg L^{-1}); (4) MCPA (1 µg L^{-1}); (5) MCPP (1 µg L^{-1}).

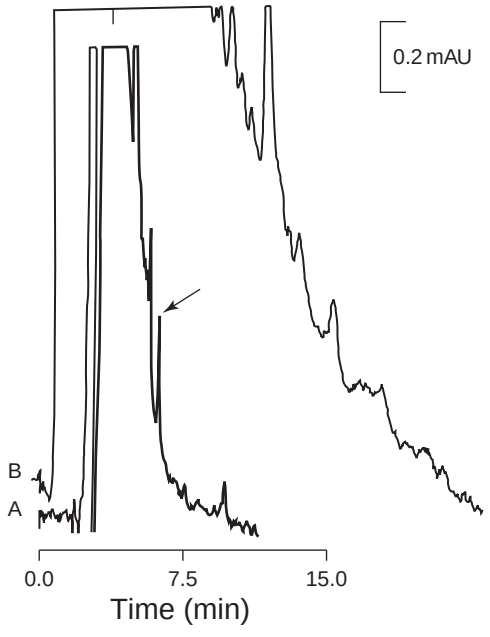

Figure 18 LC/LC/UV (215 nm) analysis of 100 µL of extract (corresponds to 50 mg of soil) of a soil sample spiked with fenpropimorf at a level of 50 µg kg^{-1}. LC/LC conditions: C-1, 5 µm Hypersil SAS (60 × 4.6 mm ID); C-2, 5 µm Hypersil ODS (150 × 4.6 mm ID); M-1, ACN–0.5% ammonia in water (50 : 50; v/v); M-2, ACN–0.5% ammonia in water (90 : 10; v/v); flow rates, 1 mL min^{-1}; cleanup volume, 5.9 mL; transfer volume, 0.45 mL. A, chromatogram obtained with column switching; B, chromatogram using the two columns in series without column switching and using M-2 as the mobile phase.

for the analysis of acidic compounds in soils.[66] For the screening of MCPP in soils a LC/LC/UV method has been developed using a 50 × 4.6 mm ID RAM column packed with 5 µm ISRP (Pinkerton) as C-1. It appeared that in comparison to an analytical C_{18} column the ISRP column substantially improved the separation between acidic analyte and co-extracted humic substances. Under the selected LC/LC conditions uncleaned soil extracts obtained after a rapid extraction/concentration procedure were directly processed with LC/LC/UV (220 nm) allowing the determination of MCPP in soils to a level of 0.02 mg kg^{-1}.[66]

3 REVERSED-PHASE LIQUID CHROMATOGRAPHY WITH FLUORESCENCE DETECTION

3.1 Liquid Chromatography/Fluorescence Detection Methods

FD measuring the emitted light after the absorption of light by a compound both at a specific wavelength is very selective and sensitive. Compounds that normally undergo fluorescence can have a 100–1000-fold increase in sensitivity over UV detection, making it attractive to be used in pesticide residue analyses. Unfortunately, very few pesticides have a favorable fluorophore. A good example is the antiparasitic agent ivermectin for which a selective and sensitive method has been developed for its determination in animal tissues at the low parts per billion level.[67]

Pre- or postcolumn derivatization techniques have substantially extended the use of FD in this field of analysis.[68,69] The most widely adopted technique is the postcolumn reactor for the detection of *N*-methylcarbamates. After the RPLC separation a two-stage postcolumn reaction system provides hydrolyses of analytes to methylamine and reaction of this product with *o*-phthalaldehyde (OPA) for the formation of a highly fluorescent derivative. This approach has been introduced by Moye et al.[70] and further modified and optimized by Krause[71] providing robust, selective and sensitive methods for the multiresidue determination of *N*-methylcarbamates in crop samples.[72–74] The SPE cleanup of the organic extracts on a cartridge packed with bonded aminopropyl-silica and carried out in the normal-phase mode appeared to be effective and rapid.[48,51] The robustness of the OPA postcolumn reaction detection is also emphasized by forming the basis of the United States Environmental Protection Agency (USEPA) Method 531.1[75] for the determination of *N*-methylcarbamate pesticides in environmental water samples.

Optimization of hydrolysis by means of a solid-phase reactor consisting of a column packed with anion-exchange material distinctly simplified the post-column derivatization technique and extended the analysis to 22 carbamate pesticides and 10 metabolites in crops with LODs in the range of 1–10 µg kg^{-1}.[76] The same approach including off-line SPE of 50 mL of sample on cartridges packed with 500 mg of low-carbon C_{18}-bonded silica (C_{18}/OH) for trace enrichment provided a method for determination of 20 carbamates and 12 metabolites in surface water at the low ng L^{-1} level;[77] a 250 × 4 mm ID column packed with 4 µm Superspher 60 RP-8 was used for the separation of carbamates and metabolites using ACN–water gradient of 30 min at flow of 1 mL min^{-1}.

Another approach to hydrolysis of carbamates and *N*-containing pesticides such as phenylureas prior to derivatization with OPA was introduced by Miles and Moye who employed a photolytic reactor consisting of a UV lamp inserted in the center of a woven coil of Teflon™ tubing.[78] The postcolumn on-line photochemical reactor in combination with fluorescence, electrochemical or conductivity detection provided MRMs for the assay of over 100 pesticides in the groundwater samples.[79]

PC derivatization using 1-methylimidazole and TFA was successfully applied for the LC/FD determination of the large molecular compound abamectin in forest matrices including water, soils and foliage.[80] Sample pretreatment consisted of an extraction with dichloromethane and cleanup with Florisil column chromatography. Limits of quantification were 0.2 µg kg^{-1} and 0.01 µg L^{-1} for solid and water samples respectively.

3.2 Solid-phase Extraction/Liquid Chromatography/Fluorescence Detection Methods

Hiemstra and Kok[81] transformed the method involving off-line SPE[76] into a completely automated method for the determination of carbamates in water. The study included the testing of two different, commercially available, automated trace enrichment devices (Prospekt and OSP-2). Both trace enrichment systems showed a good method performance and with small sampling volumes of 3 mL (OSP-2 system) or 5 mL (Prospekt system) detection limits (DLs) of the carbamates and their metabolites for surface water were between 30 and 50 ng L^{-1}; the sample throughput was about 30 samples per 24 h.

The same degree of automation was obtained for the processing of crop extracts by means of an automated SPE cleanup apparatus.[51] This system (ASPEC), performs cleanup on a SPE amino-bonded silica column and, after a solvent switch, the injection of part of the extract into the LC carbamate analysis system (see Figure 10). The good results on recovery and repeatability from 13 carbamates and 12 metabolites on 12 different types of

food products and the registration of control samples in Shewhart charts during several years of routine analysis have clearly demonstrated the good performance of this approach.[48,51]

In the analysis of *N*-methyl carbamates employing LC, hydrolysis, OPA derivatization and FD, on-line and off-line cleanup of vegetable extracts has been carried out by using activated carbon membranes.[82] The highly retentive property of the SPED material allowed direct injection of 20 µL of toluene-modified ACN extract and provided an effective retention of sample interferences. The method tested on green peppers spiked at a level of 0.25 mg kg^{-1} with 10 carbamates provided recoveries at levels between 56–93% and 87–99% for the on-line and off-line SPED cleanup, respectively.

The combination of postcolumn photolysis by means of UV irradiation at 254 nm, derivatization with OPA and FD[79,80] has been successfully applied for the monitoring of phenylurea herbicides in foodstuffs.[1,48] The instrumental analysis displayed in Figure 19 of an uncleaned extract of a carrot sample fortified with 0.1 mg kg^{-1} of phenylurea herbicides shows the good performance and the high selectivity/sensitivity of this approach; if necessary, an automated SPE sample cleanup can be included in the method (see Figure 10).

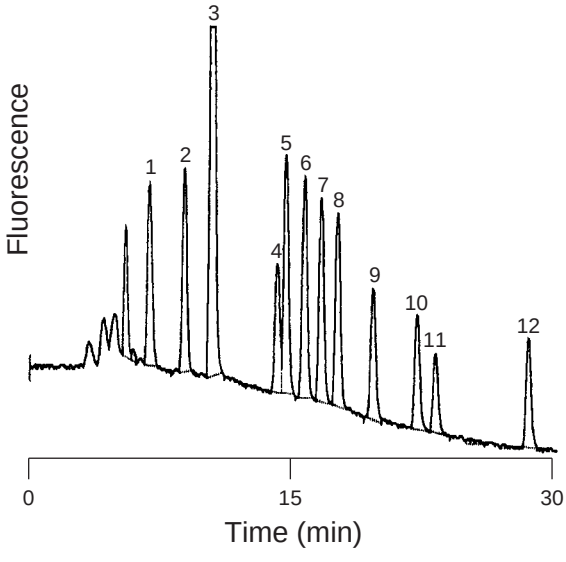

Figure 19 RPLC with FD (ex: 340 nm, em: 455 nm) with reaction detection involving post-column photolysis and derivatization with OPA of an uncleaned extract of a carrot sample spiked with phenylurea herbicides at a level of 0.1 mg kg^{-1}. Sample pretreatment is given in Figure 10. LC conditions: 250 × 4.6 mm ID 4 µm Superspher 60 RP-8 column with an ACN–water linear gradient elution from 35 to 60% of modifier in 30 min at flow of 1 mL min^{-1}; photolysis, knitted reaction coil (250 × 5 mm) and irradiation with UV at 254 nm. Peaks: 1. fenuron; 2. metoxuron; 3. monuron; 4. chlorotoluron; 5. flurometruron; 6. isoproturon; 7. difenoxuron; 8. metobromuron; 9. buturon; 10. linuron; 11. chlorbromuron; 12. neburon.

3.3 Coupled Column/Fluorescence Detection Methods

The LC/LC/FD technique has been successfully applied for the efficient determination of phosphinic acid herbicides glufosinate and glyphosate and glyphosate's metabolite amino methyl phosphonic acid (AMPA).[83–85] These herbicides are widely used in agriculture as nonselective contact herbicides. Because of their high polarity, amphoteric properties and poor detectability, methodology to determine these compounds at trace levels requires a large effort in sample preparation steps such as extraction, concentration, derivatization and cleanup. A study on the determination of glufosinate in environmental water samples[83] demonstrated that the combination of PC derivatization with 9-fluorenylmethylchloroformate (FMOC-Cl) and LC/LC/FD considerably improves the sample throughput. The method uses a C_{18} column coupled to an ion-exchange amino column with a mobile phase of ACN–0.05 M phosphate buffer, pH 5.5 (35 : 65, v/v) for both columns.

It must be emphasized that the cleanup performance here is obtained in a reversed way. Instead of removing an excess of early eluting interferences (S1, see also Figures 3 and 5) one advantageously makes use of the amphoteric property of FMOC–glyphosate. In the LC/LC approach the mobile phase ionizes the analyte which will result in little C_{18} retention on C-1 and an adequate retention on C-2. The efficient cleanup is obtained by transferring an almost unretained small-analyte-containing fraction from C-1 to C-2. All interferences with more C_{18} retention, e.g. the excess of FMOC-OH, are retained on C-1 and sent to waste by the rinsing mobile phase during the separation of the analyte on C-2.

It was established that LVI could be achieved by reducing the percentage of modifier (40% ACN is required for adequate derivatization with FMOC) with an aqueous borate buffer solution before injection. For example, a four-fold dilution step allowed the injection of 2.0 mL of sample without significant band broadening of the polar FMOC–glufosinate during injection on the C_{18} column (C-1). However, in the mobile phase (M-1) the C_{18} retention of an ionic FMOC-phosohic acid is very little and the analyte will elute then almost as an unretained compound. Hence, shortly after LVI the analyte will be transfered to the amino column (C-2) in a relatively small volume (approx. 250 µL) In this way an efficient preseparation between the large excess of the FMOC reagent and the FMOC analyte is realized.

The usefulness of this approach was also investigated for the determination of glyphosate and its main metabolite AMPA in environmental water samples.[84] It appeared that for glyphosate an eight-fold dilution step was necessary to avoid excessive band broadening of

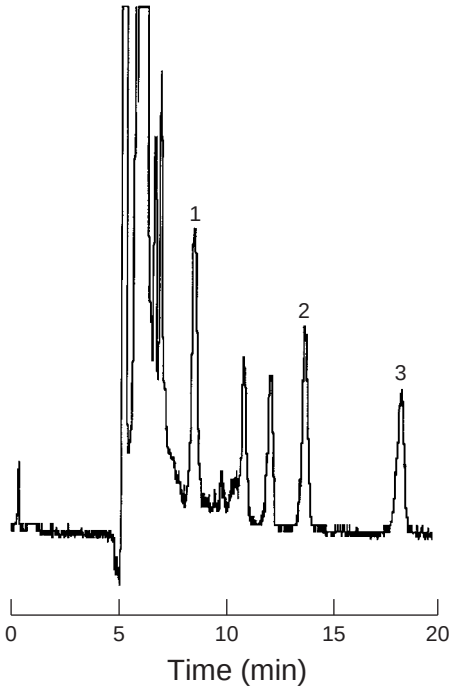

Figure 20 LC/LC/FD (ex: 263 nm, em: 317 nm) of a 2 mL solution obtained after derivatization with FMOC and 8-fold dilution with a borate buffer[84] of a surface water sample spiked with the analytes at a level of 4 μg L^{-1}. LC/LC conditions: C-1, 5 μm C_{18} (30 × 4.6 mm ID); C-2, 5 μm NH2 (250 × 4.6 mm ID); M-1 and M-2, ACN–0.05 M phosphate buffer, pH 5.5 (35 : 65, v/v) both at a flow of 1 mL min^{-1}; cleanup volume, 2.21 mL (injection volume included); transfer volume, 0.53 mL. Peaks: 1. AMPA; 2. glufosinate; 3. glyphosate. (Reprinted from Sancho et al., 'Rapid Determination of Glufosinate, Glyphosate and AMPA in Environmental Water Samples Using Precolumn Fluorogenic Labeling and Coupled-column Liquid Chromatography,' *J. Chromatogr. A.*, **737**, 75–83. Copyright © 1996 with permission from Elsevier Science.)

glyphosate–FMOC during LVI. Glufosinate, glyphosate and AMPA could indeed be assayed simultaneously in aqueous samples at a level of 1 μg L^{-1} with a high sample throughput. An example of this approach is given in Figure 20, showing the LC/LC/FD analysis of surface water sample spiked with three analytes at the level of 4 μg L^{-1}.

SRMs focused on the analysis of one analyte applying very accurately adjusted volumes for dilution, cleanup and transfer allowed the determination down to the 0.1 μg L^{-1} level.

Employing SRMs, glufosinate, glyphosate and AMPA were successfully recovered from water samples at 0.5–10 μg L^{-1} level fortification levels, with a sample throughput of at least 40 samples per day.[84]

The same approach has been applied for the determination of glyphosate and AMPA in soil samples.[85] After extraction in alkaline media an aliquot of the extract is neutralized and processed with the procedures mentioned above. Depending on the type of soil, whether it had high or low clay content and organic matter, limits of determination are 100 and 10 μg kg^{-1}, respectively for both analytes.

4 REVERSED-PHASE LIQUID CHROMATOGRAPHY WITH MASS SPECTROMETRIC DETECTION

4.1 General Aspects

The possibility of performing determination and unambiguous identification of polar pesticides simultaneously makes LC with MS detection an attractive and fast growing technique in the field of PRA. In comparison to the usually applied universal UV or photodiode array detectors, confirmation analysis with MS detection is usually not encountered with problems like (almost) identical UV spectra of compounds of a same class (e.g. phenylureas) or compounds lacking a chromophore (e.g. phosphonic acids). This results in liquid chromatography/mass spectrometry (LC/MS) having a high potential for analyzing a broad spectrum of pesticides and, hence, potentially a high productivity in multiresidue analyses.

The principles of the various LC/MS interfaces and the various types of interfaces as applied for the determination of polar pesticides have been described.[86] Until the mid-1990s most applications involved thermospray (TSP) or particle beam (PB) interfaces.[86] The advantage of obtaining EI spectra when using the PB interface makes liquid chromatography/particle beam mass spectrometry (LC/PBMS) very powerful in the detection and determination of polar pesticides. Several LC/PBMS based MRMs have been developed for the trace analysis of triazines, phenylureas and/or carbamates in water samples employing off-line[87,88] or on-line[89–92] solid extraction procedures. The possibility to obtain positive and/or negative CI spectra and to compare data to GC/MS data bases is an important feature of this technique, but evidently concerns only GC-amenable pesticides such as triazines and organophosphorus compounds.[92]

In comparison to the PB interface, liquid chromatography/thermospray mass spectrometry (LC/TSPMS) is more compatible with conventional HPLC-system mobile phases, including solvent additives, and offers in most cases higher sensitivity. Based on studies[93–95] investigating optimal TSP ionization conditions for pesticides, several LC/TSPMS MRMs including the determination of numerous pesticides of different classes, e.g. phenylureas, carbamates, phenoxy acids, triazines and oxime fungicides etc. in aqueous environmental samples have been developed.[96–98]

Unfortunately, TSP shows a number of experimental problems such as highly variable compound-dependent responses, the need for a critical control of relevant temperatures during analysis, thermal breakdown of labile analytes and the lack of structurally informative mass spectra. The last problem can be overcome by using liquid chromatography/tandem mass spectrometry (LC/MS/MS); identification can be achieved by using daughter ion tandem mass spectrometry (MS/MS) scans and collisionally induced dissociation (CID), most commonly on a triple quadrupole instrument. However, in comparison to LC/MS the instrumental cost of MS/MS systems are distinctly higher.

An important limitation encountered in both thermospray mass spectrometry (TSPMS) and particle beam mass spectrometry (PBMS) is the significant variation in sensitivity between the different classes of compounds or even for compounds from the same class. As regards this aspect, atmospheric pressure ionization (API) is most promising in combining high sensitivity and improved ionization stability by solving the problem of vaporized solvent entering the vacuum system of the mass spectrometer. API includes a group of interfaces usually addressed as electrospray (ES), ionspray (IS) and atmospheric pressure chemical ionization (APCI), which is the most flexible technique regarding conventional LC flow rates.

At present API interfaces are the most commonly used in LC/MS for target and nontarget analysis of pesticides from various matrices and the various mode applied in this field of analysis will discussed in more detail below.

4.2 Liquid Chromatography/Mass Spectrometry Methods

Volmer et al.[99] evaluated the use of LC/ESMS for multiresidue determination and confirmation of a large number of pesticides such as triazines, carbamates, phenyl- and sulfonylureas. Performing separation of analytes on 5 μm C_{18} column (150 × 3.2 mm ID) with a methanol–water gradient (20–95% methanol over 45 min) at 0.6 mL min^{-1}, post-column flow addition (0.4% formic acid in methanol, at 0.2 mL min^{-1}) and by means of splitting a 150 μL min^{-1} flow to the ESMS positive ionization (PI mode), time-scheduled single ion monitoring (SIM) yielded sensitive detection of most pesticides at the very low picogram level. Di Corcia et al.[100] developed a sensitive and specific method for the ultratrace analysis of atrazine and its major six metabolites in environmental water samples. The method makes use of SPE on cartridges packed with GCB followed by LC/ESMS. LC separation was carried on 5 μm C_{18} column (250 × 4.6 mm ID) with a gradient elution of methanol–water (both solvents contained 10^{-5} mol L^{-1} of formic acid) going from 0 to 84% methanol in 36 min. The flow rate of the mobile phase was 1 mL min^{-1}, and 40 μL of the column effluent was diverted to the ES source. Employing extraction volumes of 1 and 4 L of sample (river, ground- and drinking water) LODs of 3 ng L^{-1} were obtained (two-ion SIM acquisition) with a reproducibility of 2–7% and recoveries above 80%.

More or less the same approach concerning sample pretreatment and instrumental analysis was successfully applied by D'Ascenzo and co-workers[101] for the determination of five imidazolinone herbicides in various drinking water samples providing LODs of 2–5 ng L^{-1}.

The same group[102] developed an MRM for determination of 15 acidic herbicides (7 sulfonylureas, 4 imidazolinones, 4 aryloxyphenoxypropionic acids) in water by LC/ESMS in PI mode. A similar pretreatment procedure as mentioned above was used: passing of 2 L of groundwater and 4 L of drinking water samples, respectively, through a 0.5 g GCB extraction cartridge. A conventional 4.6 mm ID column operating with a methanol–water mobile phase (both solvents contained 10 mmol L^{-1} formic acid) at 1 mL min^{-1} was used to separate the analytes; from the column effluent, 200 μL min^{-1} were split to the ESI source. The very good performance of this approach is illustrated in Figure 21 showing total ion current (TIC) and time scheduled SIM chromatograms of the LC/ESMS analysis of a drinking water sample spiked with analytes at a level 25 ng L^{-1} each. The LODs calculated in SIM mode for drinking water were in the range of 0.5–4 ng L^{-1}. This method has also been applied to real groundwater samples in which imazathabenz, imazathabenz-methyl, haloxyfop, fluazifop and chlorsulfuron were found in concentrations ranging from 10 to 100 ng L^{-1}. Employing slightly different LC/ESMS conditions and SPE concentration on Porapak Rdx cartridges (Waters) comparable results were obtained for a group of 22 acidic pesticides of different classes when using SIM in negative ionization (NI) mode.[102]

The viability of off-line sampling of water samples on GCB cartridges and LC/ESMS towards multiresidue analysis of pesticides in environmental water samples is also illustrated by Crescenzi et al.[103] They developed a method for the determination of 45 base/neutral pesticides (carbamates, triazines, imidazole, phenylureas, anilides, triazoles, pyridazones) having a broad range in polarity. ESMS detection was optimized by adopting in-glass distilled methanol and 10 μmol L^{-1} TFA in water as mobile-phase constituents. For drinking water, LODs ranged between 1 (atrazine) and 9 (butoxycarboxim) ng L^{-1} from TIC chromatograms and were for most pesticides about 10 times lower when using SIM. The repeatability of the method of less then 10% for the analysis of an extract ($n = 7$) of a surface water sample spiked with the analytes at a level of 200 ng L^{-1}, emphasizes the ruggedness of the ESMS arrangement.

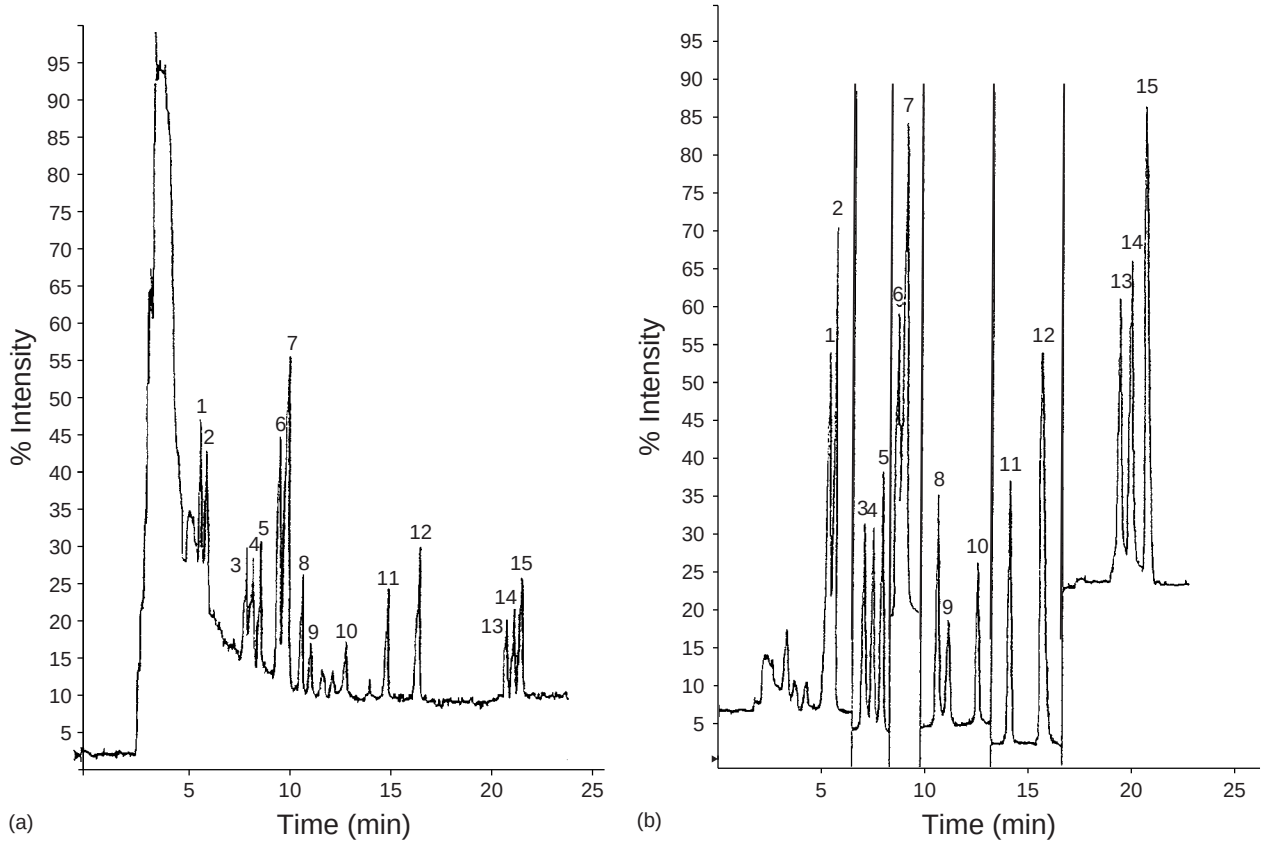

Figure 21 (a) TIC and (b) time scheduled SIM chromatogram of LC/ESMS (positive mode) of an extract of a drinking water sample spiked with the acidic herbicides at a level of $0.025\,\mu g\,L^{-1}$. Peaks: (1) imazapyr, (2) imazamethabenz, (3) triasulfuron, (4) thifensulfuron-methyl, (5) metsulfuron, (6) imazaquin, (7) imazamethabenz-methyl, (8) chlorsulfuron, (9) rimsulfuron, (10) tribenuron-methyl, (11) bensulfuronmethyl, (12) fluazifop, (13) fenoxaprop, (14) quizalofop, (15) haloxyfop. LC conditions: see text. (Reprinted with permission from D'Ascenzo et al., *Environ. Sci. Technol.*, **32**, 1340–1347. Copyright © 1998, American Chemical Society.)

The combination of a SPE cleanup on a GCB cartridge and LC/ESMS appeared also viable for the determination of 12 carbamate insecticides in fruits and vegetables to levels below $1\,\mu g\,kg^{-1}$ of crop.[104]

An intra- and interclass comparison of sensitivity, ionization and fragmentation for the determination of 17 pesticides in 5 chemical classes (triazines, phenylureas, organophosphorus, miscellaneous) in water with liquid chromatography/atmospheric pressure chemical ionization mass spectrometry (LC/APCIMS) was investigated by Doerge and Bajic.[105] Performing LC separation on a short C_{18} column (30×4.6 mm ID) with ACN–water mixtures at a flow rate of $0.5\,mL\,min^{-1}$ as the mobile phase the spectra obtained from most compounds consisted solely of $[M + H]^+$ in positive ion mode. Low nanogram LODs were observed for all compounds in scan mode and these were reduced to picogram levels by using SIM mode. For these diverse chemical classes variations in LODs were obtained of ca. 10-fold and 100-fold for scan and SIM mode respectively.

Barnes et al. developed LC/APCIMS based methods[106,107] for the determination of various pesticides in different types of foodstuffs. Diflubenzuron and clofentezine were determined in fruits employing a liquid organic solvent extraction and high-performance gel permeation chromatography (HPGPC) cleanup as sample pretreatment.[106] Applying RPLC with methanol–water (80 : 20; v/v) as the mobile phase ($1\,mL\,min^{-1}$) and atmospheric pressure chemical ionization mass spectrometry (APCIMS) in the PI mode, LODs were about $0.01\,mg\,kg^{-1}$. For the multiresidue analysis of 10 various pesticides (6 carbamates, thiabendazole, carbendazim, clofentezine, diflubenzuron) in fruits[107] samples were extracted with ethyl acetate. After solvent evaporation part of the diluted extract was directly processed with LC/APCIMS using PI/NI mode during each acquisition, which provided LODs in the range of $0.002–0.03\,mg\,kg^{-1}$ in the crop. The LODs obtained in the LC/APCIMS methods[106,107] clearly met the required MRLs for these compounds in foodstuffs. Despite the favorable features

it was observed that sensitivity is both compound- and matrix-dependent and will require next to solvent-based standards (check on column/instrument performance) additional calibration with matrix-matched standards.[107]

API applied in the LC/MS methods discussed above is a soft ionization technique that predominantly produces the protonated $(M + H)^+$ or deprotonated $(M - H)^-$ molecular ions in positive or negative mode, respectively. Increase in compound identification can be obtained by MS/MS based on detection of product ion(s) (daughter ions) formed by CID of the initially formed (de)protonated molecular ion (parent ion). For the determination of 22 acidic pesticides in water a comparison of LC/ES method performance by single MS and MS/MS was made employing a triple quadrupole instrument.[108] Despite a three-to four-fold decrease of the signal-to-noise ratio of analytes, the significant improvement in selectivity of MS/MS is clearly demonstrated in Figure 22 showing LC/ESMS analysis of a real-life groundwater sample containing three phenoxy acids at levels of about $10\,ng\,L^{-1}$.

The impressive improvement in selectivity of MS/MS was also demonstrated for the determination of imidazoline herbicides in soils.[109] This new class of low-use-rate herbicides must be monitored at low parts per billion levels, which could be achieved by using liquid chromatography/electrospray tandem mass spectrometry (LC/ESMS/MS).

4.3 On-line Liquid Chromatography/Mass Spectrometry Methods

Because of the possibility of automation, on-line procedures allow large screening and monitoring of environmental water samples. Therefore, besides or complementary to automated SPE/LC/UV/DAD systems the use of atmospheric pressure ionization mass spectrometry (APIMS) detection becomes more important in this field of analysis. The determination of 8 various types of acidic herbicides in environmental water was performed by automated on-line SPE on C_{18} cartridges and LC/MS using pneumatically assisted (or high flow) electrospray (PAES) for ionization in the NI mode.[110] The procedure required only 50 mL of water with LODs ranging between 10 and $30\,ng\,L^{-1}$ including the determination of the two major metabolites of bentazone.

SPE/LC (100 mL of sample) coupled to APCIMS, pneumatically assisted/electrospray mass spectrometry (PA/ESMS) and MS/MS detection was used for the trace analysis of a group of neutral pesticides (carbamates, phenylureas, triazines and organophosphorus) in water samples.[111] The interfaces provided comparable LODs for 15 pesticides (out of 17) in drinking water of $0.007–3\,\mu g\,L^{-1}$ and $0.1–200\,ng\,L^{-1}$ in TIC and SIM

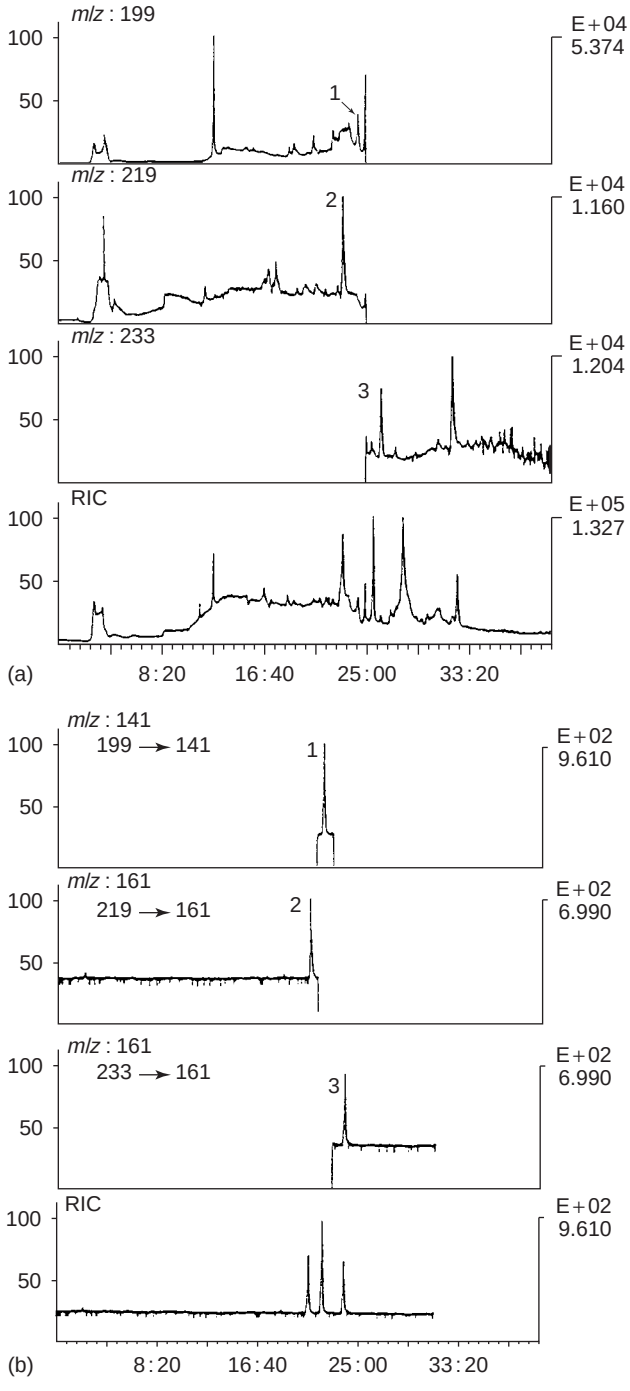

Figure 22 LC/ESMS analyses of a SPE extract of a real-life groundwater sample containing chlorophenoxy acid herbicides at a level of about $0.010\,\mu g\,L^{-1}$. The sample was analyzed using (a) single MS in SIM and by (b) MS/MS in selected reaction monitoring mode. Peaks: (1) MCPA, (2) 2,4-D, (3) 2,4-DP. (Reprinted from Koppen and Spliid, 'Determination of Acidic Herbicides Using Liquid Chromatography with Pneumatically Assisted Electrospray Ionization Mass Spectrometry and Tandem Mass Spectrometry Determination', *J. Chromatogr. A.*, **803**, 157–168. Copyright © 1998 with permission from Elsevier Science.)

mode, respectively, with an analysis time of 65 min. NI operation was found to be less sensitive for the majority of compounds tested (73 in total). Both ionization techniques gave similar product ion spectra from protonated molecules and a MS/MS library was built for more than 60 pesticides and successfully applied for searching product-ion spectra at low levels (10 ng L^{-1}) in tap water.

The simultaneous analysis of a smaller and less heterogeneous group of pesticides, usually enhances sensitivity in LC/MS. A nice example is given in the study by Baltussen et al.[112] A sample volume of 10 mL appeared to be sufficient for the on-line LC/MS determination of five phenylurea herbicides in river water to a level of 10 ng L^{-1}.

High sensitivity/selectivity was obtained by on-line immunosorbent phase extraction and LC/APCIMS (PI mode) for the analysis of 8 triazine and 6 phenylurea herbicides in environmental samples.[113] Sample volumes of 20 mL of groundwater provided LODs in the range 1–5 ng L^{-1} For obvious reasons such as removal of early eluting interferences and protection of analytical column, on-line methods usually make use of column-switching (see above) for the trace enrichment of analytes on a SPE column (PC).

For the reduction of expensive measurement time of the LC/MS system the use of a single short column (SSC) for both trace enrichment and separation in combination with selective MS/MS detection is a powerful approach. The compromise between the minimum time of analysis and the minimum required chromatographic separation for a given application is determined by the efficiency of the column (total number of plates, n) which in turn will depend on the dimensions of the column and efficiency of the packing material. In fact, suitable conditions for sorption and desorption will always include a washing step (removal of inorganic salts) prior to elution, and minimum required separation between analyte(s) and solvent front in order to avoid the unstable ionization conditions caused by the rapidly changing elution solvent composition.

The viability of the SSC approach is shown by a number of methods involving the rapid determination of organic pollutants in water.[114–116] The first report on the SSC approach involved the use of a home-packed (5–8 μm C_{18} or 10–15 μm PLRP-S material) 10 × 2 mm ID column in combination with thermospray/tandem mass spectrometry (TSPMS/MS) for the on-line determination of triazines in surface water.[114] Unfortunately, stability-system problems related to the reproducibility and stability of the experimental column packings did not provide LODs below 10 μg L^{-1}. Improved results were obtained when using a 20 × 4.6 mm ID column high-pressure packed with 5 μm C_{18} combined with atmospheric pressure chemical ionization tandem mass spectrometry (APCIMS/MS).[115] For the set of 15 various pesticides analyzed before with SPE/LC, the SSC approach involving the processing of sample volumes as small as 15 mL resulted in LODs between 0.03 and 5 μg L^{-1} in full-scan mode and between 2 and 750 ng L^{-1} by SIM, both recorded in the positive ion mode. Further optimization of the use of SSC with APCIMS/MS involving ion-trap MS/MS instead of triple quadrupole MS/MS allowed the trace analysis of six triazines in less then 20 min by taking only 4 mL of sample and, hence, providing a high sample throughput. Unfortunately, sensitivity was not sufficient for the 8 phenylurea herbicides tested; this was probably caused by several current instrumental problems. SSC in the instrumental set-up including both diode-array UV and MS detection is an efficient approach in guarding microorganic pollution of surface waters. In these types of samples usually few or no microcontaminants are present allowing a degree of selectivity to be sacrificed by using less separation power in order to obtain a cost-effective procedure.[116]

5 LIQUID CHROMATOGRAPHY SAMPLE PRETREATMENT METHODS PRIOR TO GAS CHROMATOGRAPHY

The availability of techniques, nowadays, to perform large-volume introduction GC,[117] has made the combination of LC and GC more accessible. As a result, one can easily combine the advantageous features of LC such as on-line trace enrichment and/or cleanup with the high separation power and selectivity/sensitivity of capillary GC.

Based on the type of sample/extract and the type of mobile phase, that is an aqueous solvent or an organic solvent, one can discriminate two different modes in LC methods used in combination with GC.

The LC mode employing organic solvents is focused on the efficient cleanup between macromolecules such as fatty substances (triglycerides) or plant material (chlorophyllic) and pesticides. In these types of separations GPC and NPLC are used because of their high degree of automation. The other LC-mode is based on coupling SPE columns to capillary GC providing on-line analysis of pesticides in environmental water samples.

Methods involving these LC modes will be briefly discussed below.

5.1 Gel Permeation Chromatography and Normal-phase Liquid Chromatography

For the determination of apolar pesticides with accumulating properties in fat, e.g. organochlorine pesticides, the first step is the isolation of the fat from the fatty

matrix. Usually this is done by extraction with an organic solvent yielding an extract in which the fat has to be separated from the analytes. As regards vegetable, human and animal fat, cleanup will be focused on the separation between pesticides (low-molecular-weight compounds) and triglyceride-esters (high-molecular-weight compounds).

Two powerful LC-based fully automated approaches are available using either GPC[1,118–124] or NPLC silica cleanup[125–129] to perform the separation between the fatty compounds and the pesticides. GPC is applied most frequently since its separation, based on molecular mass, results in the widest scope concerning both matrices and analytes.[1,118,119] The availability nowadays of HPGPC columns packed with small size particles (7–10 μm) providing small elution volumes of analytes,[121–124] typically 100–300 μL, allows the coupling of GPC to GC allowing automated procedures.[123,124]

The advantage/disadvantage of the NPLC silica cleanup approach is that the high separation power of analytical column packed with 5 μm silica favors selective cleanup but counteracts the range in polarity of analytes.[125–128] Productive methods have been developed for monitoring of organochlorine pesticides in milk[125] and various types of fatty matrices[126,127] and 1,1-dichloro-2,2-bis(4-chlorophenyl)ethane (DDE) and polychlorinated biphenyls (PCBs) in extracts of adipose tissues[128] employing NPLC off-line[125,126] or on-line,[127,128] respectively. For the determination of the more polar organophosphorus pesticides in mussels the separation between analytes and fat was successfully carried with NPLC by collecting the analyte in a fat-free fraction using different mixtures of hexane–ethyl acetate.[129]

5.2 Solid-phase Extraction Combined On-line with Gas Chromatography

The coupling of SPE to GC provides fully automated procedures for the determination of pesticides in environmental water samples. Small volumes of sample, typically 1–5 mL, are sufficient to obtain the required sensitivity. PCs (10 × 2 mm ID) packed with 15–25 μm PLRP-S copolymer have been used for the on-line determination of nitrogen- and phosphorus-containing pesticides in water at the low nanogram per litre level.[130–132] This PC material shows favorable properties for the required solvent-phase switch involving (i) adsorption of analytes during sampling, (ii) the removal of water by passing nitrogen and (iii) desorption of pesticides with an organic solvent, e.g. ethyl acetate.

An automated ''off-line'' SPE system (ASPEC) has been coupled to GC with electron capture detection (ECD) for the on-line analysis of pyrethroid insecticides in water.[133] This approach combines extraction and cleanup by adding 30% of modifier (methanol) to the water sample prior to SPE on a 100 mg C_{18} cartridge and a washing step of 1 mL of 30% methanol in water. After drying with air, the cartridge is eluted with 1.5 mL of toluene. GC analysis with ECD and on-column introduction of 100 μL of the extract provided LODs in the range of 1–8 ng L^{-1}.

6 CONCLUSIONS AND TRENDS

LC has proven to be an efficient technique in PRA making it very complementary to GC for the determination of polar and/or thermolabile pesticides. LC methods cover samples from both the disciplines of public health and environmental protection and the column switching mode easily provides semi- or fully automated procedures including on-line cleanup and/or trace enrichment. Future developments in LC will be focused on the enhancement of sample throughput, robustness, scope of analytes, selectivity, sensitivity and, even more importantly, reduction of cost.

Providing high selectivity/sensitivity for the determination of pesticides with widely different (physicochemical) properties, RPLC combined with APIMS is a powerful and attractive technique, and increasing use in this field of analysis can be expected. Especially in studies on the identity and concentration of formed TPs, always more polar then the orginal pesticides, LC/MS will be a highly desirable technique.

Despite the high selectivity of MS or MS/MS combined use with column switching techniques, viz. SPE/LC or LC/LC, can be attractive in order to eliminate matrix-dependent quantification of analytes in the processing of uncleaned extracts.

In comparison to conventional LC, capillary LC offers a number of favorable features such as reduction of solvent consumption, improvement of sensitivity and interfacing to more selective and sensitive detectors. Because of the susceptibility of the technique to clogging/contamination problems caused by the excess of matrix interferences the use of capillary LC in PRA is still limited. Here, column switching appeared to be favorable to enhance sample load and cleanup as shown for the coupling of capillary LC to infrared spectrometry for the unambiguous identification/determination of medium polar pesticides in surface water at the 2–5 μg L^{-1} level.[134]

An interesting development in obtaining high selectivity is the on-line coupling of RPLC and high-performance thin-layer chromatography (HPTLC) in the normal phase mode.[135] This two-dimensional low-cost separation system offers high peak capacity, an attractive feature in PRA.

ACKNOWLEDGMENTS

The assistance of Veronica 't Hart-de Klein is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
AMPA	Amino Methyl Phosphonic Acid
APCI	Atmospheric Pressure Chemical Ionization
APCIMS	Atmospheric Pressure Chemical Ionization Mass Spectrometry
APCIMS/MS	Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry
API	Atmospheric Pressure Ionization
APIMS	Atmospheric Pressure Ionization Mass Spectrometry
CAAL	Chloroallylalcohol
CCPR	Codex Committee on Pesticide Residues
CI	Chemical Ionization
CID	Collisionally Induced Dissociation
DAD	Diode Array Detection
DDE	1,1-Dichloro-2,2-bis(4-chloro-phenyl)ethane
DEA	Deethylatrazine
DIA	Deisopropylatrazine
DL	Detection Limit
DNOC	4-6-Dinitro-*o*-cresol
DOC	Dissolved Organic Carbon
EC	European Community
ECD	Electron Capture Detection
EI	Electron Impact
ES	Electrospray
ESMS	Electrospray Mass Spectrometry
ETU	Ethylenethiourea
FD	Fluorescence Detection
FMOC	9-Fluorenylmethyl-chloroformate
GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
HPGPC	High-performance Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
HPLC/UV	High-performance Liquid Chromatography/Ultraviolet
HPTLC	High-performance Thin-layer Chromatography
ID	Internal Diameter
IP/RPLC	Ion-pair/Reversed-phase Liquid Chromatography
IS	Ionspray
ISRP	Internal Surface Reversed-phase
LC	Liquid Chromatography
LC/APCIMS	Liquid Chromatography/ Atmospheric Pressure Chemical Ionization Mass Spectrometry
LC/ESMS	Liquid Chromatography/ Electrospray Mass Spectrometry
LC/ESMS/MS	Liquid Chromatography/ Electrospray Tandem Mass Spectrometry
LC/LC	Coupled Column Reversed-phase Liquid Chromatography
LC/LC/UV	Coupled Column Reversed-phase Liquid Chromatography/Ultraviolet
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/MS/MS	Liquid Chromatography/ Tandem Mass Spectrometry
LC/PBMS	Liquid Chromatography/ Particle Beam Mass Spectrometry
LC/TSPMS	Liquid Chromatography/ Thermospray Mass Spectrometry
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
LVI	Large Volume Injection
MASE	Microwave-assisted Solvent Extraction
MCPA	(4-Chloro-2-methylphenoxy)-acetic Acid
MCPB	4-(4-Chloro-2-methyl-phenoxy)butanoic Acid
MCPP	Mecoprop
MRL	Maximum Residue Level
MRM	Multiresidue Method
MS	Mass Spectrometry

MS/MS	Tandem Mass Spectrometry
NI	Negative Ionization
NPLC	Normal-phase Liquid Chromatography
OPA	*o*-Phthalaldehyde
PAES	Pneumatically Assisted Electrospray
PA/ESMS	Pneumatically Assisted/ Electrospray Mass Spectrometry
PB	Particle Beam
PBMS	Particle Beam Mass Spectrometry
PC	Precolumn
PCB	Polychlorinated Biphenyl
PC/LC	Precolumn/Liquid Chromatography
PCP	Pentachlorophenol
PI	Positive Ionization
PLRP-S	Polymeric Reversed-phase Spherical Packing
PRA	Pesticide Residue Analysis
RAM	Restricted Access Material
RPLC	Reversed-phase Liquid Chromatography
RPLC/UV	Reversed-phase Liquid Chromatography/Ultraviolet
RSD	Relative Standard Deviation
SAMOS	System for Automated Monitoring of Organic Compounds in Surface Water
SIM	Single Ion Monitoring
SPE	Solid-phase Extraction
SPED	Solid-phase Extraction Disk
SPED/LC/UV/DAD	Solid-phase Extraction Disk/ Liquid Chromatography/ Ultraviolet/Diode Array Detection
SPE/GC	Solid-phase Extraction/Gas Chromatography
SPE/LC	Solid-phase Extraction/Liquid Chromatography
SPE/LC/DAD	Solid-phase Extraction/Liquid Chromatography/Diode Array Detection
SRM	Single Residue Method
SSC	Single Short Column
TFA	Trifluoroacetic Acid
TIC	Total Ion Current
TP	Transformation Product
TSP	Thermospray
TSPMS	Thermospray Mass Spectrometry
TSPMS/MS	Thermospray/Tandem Mass Spectrometry
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
2,4-D	2,4-Dichlorophenoxy Acetic Acid
2,4-DB	4-(2,4-Dichlorophenoxy)-butanoic Acid
2,4-DP	Dichlorprop
2,4,5-T	(2,4,5-Trichlorophenoxy)acetic Acid
2,4,5-TP	Fenoprop

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants • Infrared Spectroscopy in Environmental Analysis

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Herbicide Residues in Biota, Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Liquid Chromatography **(Volume 13)**
Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Electron Ionization Mass Spectrometry

REFERENCES

1. P. van Zoonen (ed.) 'Analytical Methods for Pesticide Residues in Foodstuffs', 6th edition, General

Inspectorate for Health Protection, Ministry of Public Health, Welfare and Sport, Den Hague, 1996.
2. Regeling Residuen van Bestrijdingsmiddelen, Staatscourant van 4 juli 1995 (Stcrt. 126), laatstelijk gewijzing bij besluit van 2 juni 1995, Stcrt 116.
3. EC Directive 80/778/EEC (OJI 229 30.80).
4. R.J. Bushway, 'Analysis of Pesticide Residues in Food by HPLC', in *Food Analysis by HPLC*, ed. L.M.L. Nollet, Marcel Dekker, New York, 507–577, 1992.
5. D. Barceló, M.-C. Hennion, in 'Trace Determination of Pesticides and their Degradation Products in Water', *Techniques and Instrumentation in Analytical Chemistry*, eds. D. Barceló, M.-C. Hennion, Elsevier, Amsterdam, Vol. 19, 1997.
6. J. Slobodnik, M.G.M. Groenewegen, E.R. Brouwer, E.R.H. Lingeman, U.A.Th. Brinkman, 'Fully Automated Multi-residue Method for Trace Level Monitoring of Polar Pesticides by Liquid Chromatography', *J. Chromatogr.*, **642**, 359–370 (1993).
7. U.A.Th. Brinkman, H. Lingeman, J. Slobodnik, 'SAMOS LC: A Fully Automated Liquid Chromatographic Method for Trace-level Monitoring of Polar Pesticides', *LC-GC Int.*, **7**, 157–163 (1994).
8. V. Coquart, M.-C. Hennion, 'Determination of Phenoxyacid Herbicides in Drinking at the ppt-level by Liquid Chromatography and On-line Selective Preconcentration', *Sci. Tot. Environ.*, **132**, 349–360 (1993).
9. E.A. Hogendoorn, U.A.Th. Brinkman, P. van Zoonen, 'Coupled-column Reversed-phase Liquid Chromatography-UV Analyzer for the Determination of Polar Pesticides in Water', *J. Chromatogr.*, **644**, 307–314 (1993).
10. E.A. Hogendoorn, P. van Zoonen, 'Coupled-column Reversed-phase Liquid Chromatography in Environmental Analysis', *J. Chromatogr. A*, **703**, 149–166 (1995).
11. E.A. Hogendoorn, R. Hoogerbrugge, R.A. Baumann, H.D. Meiring, A.P.J.M. de Jong, P. van Zoonen, 'Screening and Analysis of Polar Pesticides in Environmental Monitoring Programmes by Coupled-column Liquid Chromatography and Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **754**, 49–60 (1996).
12. C.E. Werkhoven-Goewie, W.M. Boon, A.J.J. Praat, R.W. Frei, U.A.Th. Brinkman, C.J. Little, 'Preconcentration and LC Analysis of Chlorophenols, Using a Styrene-divinyl-benzene Copolymeric Sorbent and Photochemical Reaction Detection', *Chromatographia*, **16**, 53–59 (1982).
13. E.A. Hogendoorn, R. Hoogerbrugge, P. van Zoonen, C.E. Goewie, P.J. Schoenmakers, 'Development of a Rational Optimization Procedure for the Automated Sample Clean-up with Column-switching in Pesticide Residue Analysis', *J. Chromatogr.*, **552**, 113–135 (1991).
14. E.A. Hogendoorn, R. Hoogerbrugge, E. Dijkman, S.M. Gort, U.A.Th. Brinkman, P. van Zoonen, 'General Strategy for Multi-residue-analysis of Polar Pesticides Using Coupled-column RPLC and Step-gradient Elution', *J. Chromatogr. Sci.*, **31**, 433–439 (1993).
15. S.M. Gort, E.A. Hogendoorn, E. Dijkman, P. van Zoonen, R. Hoogerbrugge, 'The Optimization of Step-gradient Elution Conditions in Liquid Chromatography. Application to Pesticide Residue Analysis with Coupled-column Reversed-phase LC', *Chromatographia*, **42**, 17–24 (1996).
16. S. Baskaran, R.S. Kookan, R. Naidu, 'Determination of the Insecticide Imidacloprid in Water and Soil Using High-performance Liquid Chromatography', *J. Chromatogr. A*, **787**, 271–275 (1997).
17. A. Lagana, G. Fago, A. Marino, 'Simultaneous Determination of Imidazolinone Herbicides from Soil and Natural Waters Using Soil Column Extraction and Off-line Solid-phase Extraction Followed by Liquid Chromatography with UV Detection or Liquid Chromatography/Electrospray Mass Spectrometry', *Anal. Chem.*, **70**, 121–130 (1998).
18. M. Page, Y. Ma, H. Qi, E. Healy, J. Trolinger, 'Determination of Thiazopyr and its Metabolite in Florida Groundwater Using Graphitized Carbon Solid-phase Extraction and Liquid Chromatographic Determination', *J. Agric. Food Chem.*, **45**, 3095–3099 (1997).
19. N. Font, F. Hernández, E.A. Hogendoorn, R.A. Baumann, P. van Zoonen, 'Microwave-assisted Solvent Extraction and Reversed-phase Liquid Chromatography-UV Detection for Screening Soils for Sulfonylurea Herbicides', *J. Chromatogr. A*, **798**, 179–186 (1998).
20. A. Di Corcia, C. Crescenzi, R. Samperi, L. Scappaticcio, 'Trace Analysis of Sulfonylurea Herbicides in Water: Extraction and Purification by a Carbograph 4 Cartridge, Followed by Liquid Chromatography with UV Detection, and Confirmatory Analysis by an Electrospray/Mass Detector', *Anal. Chem.*, **69**, 2819–2826 (1997).
21. H. Pouliquen, V. Fauconnet, M. Morvan, L. Pinault, 'Determination of Warfarin in the Yolk and the White of Hen's Eggs By Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. B*, **702**, 143–148 (1997).
22. L.-T. Yeh, D.A. Schwedler, G.E. Schelle, J.L. Balcer, 'Application of Empore Disk Extraction for Trace Analysis of Spinosad and Metabolites in Leafy Vegetables, Peppers and Tomatoes by High-performance Liquid Chromatography with Ultraviolet Detection', *J. Agric. Food Chem.*, **45**, 1746–1751 (1997).
23. M.T. Corasaniti, M.C. Strongoli, G. Nisticó, 'Determination of Paraquat in Rat Brain Using Ion-pair Solid-phase Extraction and RPLC with Ultraviolet Detection', *J. Chromatogr.*, **527**, 189–195 (1990).
24. K. Ameno, C. Fuke, S. Ameno, H. Kinoshita, I. Ijiri, 'Simultaneous Quantitation of Diquat and its Two Metabolites in Serum and Urine by Ion-paired HPLC', *J. Liq. Chromatogr.*, **18**, 2115–2121 (1995).

25. C. Madhu, Z. Cregus, C.D. Klaasen, 'Simple Method for Analysis of Diquat in Biological Fluids and Tissues by High-performance Liquid Chromatography', *J. Chromatogr. B*, **674**, 193–196 (1995).
26. D. Fletouris, N. Botsoglou, I. Psomas, A. Mantis, 'Rapid Quantitative Screening Assay of Trace Benzimidazole Residues in Milk by Liquid Chromatography', *J. AOAC Int.*, **79**, 1281–1287 (1996).
27. A. Di Corcia, M. Marchetti, 'Multiresidue Method for Pesticides in Drinking Water Using a Graphitized Carbon Black Cartridge Extraction and Liquid Chromatographic Analysis', *Anal. Chem.*, **63**, 580–585 (1991).
28. A. Di Corcia, M. Marchetti, 'Method Development for Monitoring Pesticides in Environmental Waters: Liquid–Solid Extraction Followed by Liquid Chromatography', *Environ. Sci. Technol.*, **26**, 66–74 (1992).
29. A. Di Corcia, R. Samperi, A. Marcomini, S. Steluto, 'Graphitized Carbon Black Extraction Cartridges for Monitoring Polar Pesticides in Water', *Anal. Chem.*, **65**, 907–912 (1993).
30. V. Pichon, C. Cau Dit Coumes, L. Chen, S. Guenu, M.-C. Hennion, 'Simple Removal of Humic and Fulvic Acid Interferences Using Polymeric Sorbents for the Simultaneous Solid-phase Extraction of Polar Acidic, Neutral and Basic Pesticides', *J. Chromatogr. A*, **737**, 25–33 (1996).
31. V. Pichon, L. Chen, M.-C. Hennion, R. Daniel, A. Martel, F. Le Goffic, J. Abain, D. Barceló, 'Preparation and Evaluation of Immunosorbents for Selective Trace Enrichment of Phenylurea and Triazine Pesticides in Environmental Waters', *Anal. Chem.*, **67**, 2541–2460 (1995).
32. V. Pichon, H. Rogniaux, N. Fischer-Durand, S. Ben Rejeb, F. Le Goffic, M.-C. Hennion, 'Characteristics of Immunosorbents used as a New Approach to Selective Solid-phase Extraction in Environmental Analysis', *Chromatographia*, **45**, 289–295 (1997).
33. J. Slobodnik, E.R. Brouwer, R.B. Geerdink, W.H. Mulder, H. Lingeman, U.A.Th. Brinkman, 'Fully Automated On-line Liquid Chromatographic Separation System for Polar Pollutants in Various Types of Water', *Anal. Chim. Acta*, **268**, 55–65 (1992).
34. V. Pichon, M.-C. Hennion, 'Determination of Pesticides in Environmental Waters by Automated On-line Trace Enrichment and Liquid Chromatography', *J. Chromatogr. A*, **665**, 269–281 (1994).
35. U.A.Th. Brinkman, J. Slobodnik, J.J. Vreuls, 'Trace-level Detection and Identification of Polar Pesticides in Surface Water: the SAMOS Approach', *TrAC*, **13**, 373–381 (1994).
36. G.R. Mills, 'Application of an On-line Liquid Chromatographic System for the Determination of Polar Herbicides in Drinking Water Within a Routine Laboratory', *J. Chromatogr.*, **813**, 63–70 (1998).
37. A.C. Hogenboom, I. Jagt, R.J.J. Vreuls, U.A.Th. Brinkman, 'On-line Trace Level Determination of Polar Organic Microcontaminants in Water Using Various Pre-column-analytical Column Liquid Chromatographic Techniques with Ultraviolet Absorbance and Mass Spectrometric Detection', *Analyst*, **122**, 1371–1377 (1997).
38. E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'Use of Membrane Extraction Disks for On-line Trace Enrichment of Organic Compounds from Aqueous Samples', *Chromatographia*, **29**, 415–418 (1990).
39. E.R. Brouwer, D.J. van Iperen, I. Liska, H. Lingeman, U.A.Th. Brinkman, 'Liquid Chromatographic Determination of Polar Pollutants in Surface Water Using Membrane Extraction Disks for On-line Trace Enrichment', *Intern. J. Environ. Anal. Chem.*, **47**, 257–266 (1992).
40. S. Chiron, D. Barceló, 'Determination of Pesticides in Drinking Water by On-line Solid-phase Disk Extraction Followed by Various Liquid Chromatographic Systems', *J. Chromatogr.*, **645**, 125–134 (1993).
41. M. Ibáñez, Y. Picó, J. Mañes, 'On-line Determination of Bioyridinium Herbicides in Water by HPLC', *Chromatographia*, **45**, 402–407 (1997).
42. M. Ibáñez, Y. Picó, J. Mañes, 'On-line Liquid Chromatographic Trace Enrichment and High-performance Liquid Chromatographic Determination of Diquat, Paraquat and Difenzoquat in Water', *J. Chromatogr. A*, **728**, 325–331 (1996).
43. Liska, E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'Trace Enrichment of Polar Pollutants from Water and the Influence of Humic Substances', *Chromatographia*, **37**, 13–19 (1993).
44. R.B. Geerdink, S. van Tol-Wildenburg, W.M.A. Niessen, U.A.Th. Brinkman, 'Determination of Phenoxy Acid Herbicides from Aqueous Samples by Improved Clean-up on Polymeric Pre-columns at High pH', *Analyst*, **122**, 889–893 (1997).
45. N. Masqué, M. Galia, R. Marcé, F. Borull, 'New Chemically Modified Polymeric Resin for Solid-phase Extraction of Pesticides and Phenolic Compounds from Water', *J. Chromatogr. A*, **803**, 147–155 (1998).
46. V. Pichon, L. Chen, M.-C. Hennion, 'On-line Preconcentration and Liquid Chromatographic Analysis of Phenylurea Pesticides in Environmental Water Using a Silica-based Immunosorbent', *Anal. Chim. Acta*, **311**, 429–436 (1995).
47. A. Martin-Esteban, P. Fernández, D. Stevenson, C. Cámara, 'Mixed Immunosorbent for Selective On-line Trace Enrichment and Liquid Chromatography of Phenylurea Herbicides in Environmental Waters', *Analyst*, **122**, 1113–1117 (1997).
48. A. de Kok, and M. Hiemstra, 'A Multiresidue Method for the Determination of Pesticides in Fruits and Vegetables Using Acetone Extraction and Multiple Chromatography – Detection Systems. Part 2: Validation of HPLC Methods', presented at 2nd European

Pesticide Residue Workshop, May 1998, Almeria, Spain.
49. M. Hiemstra, A. Toonen, A. de Kok, 'Determination of Benzoylphenylurea Insecticides in Various Crops Using Chromatography with Diode Array Detection', presented at 2nd European Pesticide Residue Workshop, May 1998, Almeria, Spain.
50. M. Hiemstra, J.A. Joosten, A. de Kok, 'Fully Automated Solid-phase Extraction Cleanup and On-line Liquid Chromatographic Determination of Benzimidazole Fungicides in Fruits and Vegetables', *J. AOAC Int.*, **78**, 1267–1274 (1995).
51. A. de Kok, M. Hiemstra, 'Optimization, Automation and Validation of the SPE Clean-up and On-line Liquid Chromatographic Determination of *N*-methylcarbamate Pesticides in Fruits and Vegetables', *J. AOAC Int.*, **75**, 1063–1072 (1992).
52. E.A. Hogendoorn, A.P.J.M. de Jong, P. van Zoonen, U.A.Th. Brinkman, 'Reversed-phase LC Column-switching for the Sensitive and Selective Determination of Polar Compounds; Application to the Analysis of Chloroallylalcohol at Sub-ppb Level in Groundwater', *J. Chromatogr.*, **511**, 243–256 (1990).
53. E.A. Hogendoorn, C. Verschraagen, P. van Zoonen, U.A.Th. Brinkman, 'Coupled Column Liquid Chromatography for the Trace Analysis of Polar Pesticides in Water Using Direct Large Volume Injection: Method Development Strategy Applied to Methylisothiocyanate (MITC)', *Anal. Chim. Acta*, **268**, 205–215 (1992).
54. E.A. Hogendoorn, P. van Zoonen, U.A.Th. Brinkman, 'Column-switching Reversed-phase Liquid Chromatography for the Trace-level Determination of Ethylenethiourea (ETU) in Aqueous Samples', *Chromatographia*, **32**, 285–292 (1991).
55. J.V. Sancho, C. Hidalgo, F. Hernández, 'Direct Determination of Bromacil and Diuron Residues in Environmental Water Samples by Coupled-column Liquid Chromatography and Large-volume Injection', *J. Chromatogr. A*, **761**, 322–326 (1997).
56. C. Hidalgo, J.V. Sancho, F. Hernández, 'Trace Determination of Triazine Herbicides by Means of Coupled-column Liquid Chromatography and Large Volume Injection', *Anal. Chim. Acta*, **338**, 223–229 (1997).
57. F. Hernández, C. Hidalgo, J.V. Sancho, 'New Method for the Rapid Determination of Triazine Herbicides and Some of their Main Metabolites in Water by Using Coupled-column Liquid Chromatography and Large Volume Injection', *J. Chromatogr. A*, **778**, 171–181 (1997).
58. C. Umile, J.F.K. Huber, 'Significant Reduction of the Detection Limit in Ion Chromatography by Relative Analyte Enrichment with Column Switching', *J. Chromatogr. A*, **723**, 11–17 (1996).
59. C.E. Goewie, E.A. Hogendoorn, 'Analysis of Bromacil, Diuron and 3,4-Dichloroaniline in Contaminated Well Water, Using a High-performance Liquid Chromatographic Column-switching Procedure', *J. Chromatogr.*, **410**, 211–216 (1987).
60. F. Hernández, C. Hidalgo, J.V. Sancho, F.J. López, 'Couple-column Liquid Chromatography Applied to the Trace Level Determination of Triazine Herbicides and Some of their Metabolites in Water Samples', *Anal. Chem.*, **70**, 3322–3328 (1998).
61. I.H. Hagestam, T.C. Pinkerton, 'Internal Surface Reversed-phase Silica-supports for Liquid Chromatography', *Anal Chem.*, **57**, 1757–1763 (1985).
62. K.K. Unger, 'Packing Materials with Surface Barriers, Controlled Distribution and Topography of Ligands for Sample Clean-up and Analysis of Biologically Active Solutes in HPLC', *Chromatographia*, **31**, 507–511 (1991).
63. C.P. Desilets, M.A. Rounds, F.E. Regnier, 'Semi-permeable-surface Reversed-phase Media for High-performance Liquid Chromatography', *J. Chromatogr.*, **544**, 25–39 (1991).
64. J.V. Sancho-Llopis, F. Hernández-Hernández, E.A. Hogendoorn, P. van Zoonen, 'A Rapid Method for the Residue Analysis of Eight Chlorophenoxy Acids in Environmental Water Samples Using Off-line SPE and On-line Selective Precolumn Switching', *Anal. Chim. Acta*, **283**, 287–296 (1993).
65. E.A. Hogendoorn, E. Dijkman, R.A. Baumann, C. Hidalgo, J.V. Sancho, F. Hernandez, 'Strategy in Using Analytical Restricted Access Media (RAM) Columns for the Decrease of Humic Acid Interferences in the Trace Analysis of Acidic Herbicides in Water Samples Using Coupled Column Liquid Chromatography with UV Detection', *Anal. Chem.*, **71**, 1111–1118 (1999).
66. P. Parrilla, P. Kaim, E.A. Hogendoorn, R.A. Baumann, 'An Internal Surface Reversed-phase Column for the Effective Removal of Humic Acid Interferences in the Trace Analysis of Mecoprop in Soils with Coupled RPLC-UV', *Fresenius' J. Anal. Chem.*, **363**, 77–82 (1999).
67. C.D.C. Salisbury, 'Modified Method for the Determination of Ivermectin Residues in Animal Tissues', *J. AOAC Int.*, **76**, 1149–1151 (1993).
68. B.D. McGravey, 'Derivatization Reactions Applicable to Pesticide Determination by High-performance Chromatography', *J. Chromatogr. B*, **659**, 243–257 (1994).
69. R. Schuster, W. Haecker, 'Automated Precolumn Sample Preparation System', Hewlett Packard Application Note, 12-5954-8911, 1–9, 1987.
70. H.A. Moye, S.J. Scherer, P.A. St. John, 'Dynamic Fluorogenic Labelling of Pesticides for High-performance Liquid Chromatography: Detection of *N*-methylcarbamates with *o*-Phthalaldehyde', *Anal. Lett.*, **10**, 1049–1073 (1977).
71. R.T. Krause, 'Further Optimization and Refinement of an HPLC Post-column Fluorometric Labeling

Technique for the Determination of Carbamate Pesticides', *J. Chromatogr. Sci.*, **16**, 281–288 (1978).
72. R.T. Krause, 'Liquid Chromatographic Determination of *N*-methylcarbamate Insecticides and Metabolites in Crops. II Analytical Characteristics and Residue Findings', *J. Assoc. Off. Anal. Chem.*, **68**, 734–741 (1985).
73. A. de Kok, M. Hiemstra, C.P. Vreeker, 'Improved Cleanup Method for the Multiresidue Analysis of *N*-methylcarbamates in Grains, Fruits, and Vegetables by means of HPLC with Postcolumn Reaction and Fluorescence Detection', *Chromatographia*, **24**, 469–476 (1987).
74. M.J. Page, M. French, 'Determination of *N*-methylcarbamate Insecticides in Vegetables, Fruits and Feeds Using Solid-phase Extraction Cleanup in the Normal Phase', *J. AOAC Int.*, **75**, 1073–1083 (1992).
75. Method 531.1, 'Measurement of *N*-methylcarbomoyloximes and *N*-methylcarbamates in Water by Direct Aqueous Injection HPLC with Post-column Derivatization', US Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH, 1989.
76. A. de Kok, M. Hiemstra, C.P. Vreeker, 'Optimization of the Postcolumn Hydrolysis Reaction on Solid-phases for the Routine High-performance Liquid Chromatographic Determination of *N*-methylcarbamate Pesticides in Food Products', *J. Chromatogr.*, **507**, 459–472 (1990).
77. A. de Kok, M. Hiemstra, U.A.Th. Brinkman, 'Low ng/l-level Determination of Twenty *N*-methylcarbamate Pesticides and Twelve of their Polar Metabolites in Surface Water Via Off-line Solid-phase Extraction and High-performance Liquid Chromatography with Post-column Reaction and Fluorescence Detection', *J. Chromatogr.*, **623**, 265–276 (1992).
78. C.J. Miles, H.A. Moye, 'HPLC with Post-column Photolysis of Pesticides for Generation of Fluorophores', *Chromatographia*, **24**, 628–632 (1987).
79. C.J. Miles, 'Determination of National Survey of Pesticides Analytes in Groundwater by Liquid Chromatography with Post Column Reaction Detection', *J. Chromatogr.*, **592**, 283–290 (1992).
80. K.M.S. Sundaram, J. Curry, 'Determination of Abamectin in Some Forest Matrices by Liquid Chromatography with Fluorescence Detection', *J. Liq. Chrom. Relat. Technol.*, **20**, 1757–1772 (1997).
81. M. Hiemstra, A. de Kok, 'Determination of *N*-methylcarbamate Pesticides in Environmental Water Samples Using Automated On-line Trace Enrichment with Exchangeable Cartridges and High-performance Liquid Chromatography', *J. Chromatogr. A*, **667**, 155–166 (1994).
82. L.E. Sojo, 'The Study of the Feasibility of the Use of Activated Carbon Membranes for the On-line Cleanup of Vegetable Extracts in the Determination of *N*-methylcarbamate Pesticides by Liquid Chromatography', *J. Chromatogr. A*, **788**, 87–94 (1997).
83. J.V. Sancho, F.J. López, F. Hernández, E.A. Hogendoorn, P. van Zoonen, 'Rapid Determination of Glufosinate in Environmental Water Samples Using 9-fluorenylmethoxycarbonyl Precolumn Derivatization, Large-volume Injection and Coupled-column Liquid Chromatography', *J. Chromatogr. A*, **678**, 59–67 (1994).
84. J.V. Sancho, F. Hernández, F.J. López, E.A. Hogendoorn, E. Dijkman, P. van Zoonen, 'Rapid Determination of Glufosinate, Glyphosate and AMPA in Environmental Water Samples Using Precolumn Fluorogenic Labeling and Coupled-column Liquid Chromatography', *J. Chromatogr. A*, **737**, 75–83 (1996).
85. V. Sancho, C. Hidalgo, F. Hernández, F.J. López, E.A. Hogendoorn, E. Dijkman, 'Rapid Determination of Residues of Glyphosate and its Main Metabolite AMPA in Soil Samples by Liquid Chromatography', *Int. J. Env. Anal. Chem.*, **62**, 53–63 (1996).
86. J. Slobodnik, B.L.M. van Baar, U.A.Th. Brinkman, 'Column Liquid Chromatography/Mass Spectrometry: Selected Techniques in Environmental Applications for Polar Pesticides and Related Compounds', *J. Chromatogr. A*, **703**, 81–121 (1995).
87. A. Cappiello, G. Famiglini, F. Bruner, 'Determination of Acidic and Basic/Neutral Pesticides in Water with a New Microliter Flow Rate LC/MS Particle Beam Interface', *Anal. Chem.*, **66**, 1416–1423 (1994).
88. M. Honing, J. Riu, D. Barceló, B.L.M. van Baar, U.A.Th. Brinkman, 'Determination of Ten Carbamate Pesticides in Aquatic and Sediment Samples by Liquid Chromatography/Ionspray and Thermospray Mass Spectrometry', *J. Chromatogr. A*, **733**, 283–294 (1996).
89. R.M. Marcé, H. Prosen, C. Crespo, M. Calull, F. Borrull, U.A.Th. Brinkman, 'On-line Enrichment of Polar Pesticides in Environmental Waters by Reversed-phase Liquid Chromatography–Diode Array Detection–Particle Beam Mass Spectrometry', *J. Chromatogr. A*, **696**, 63–74 (1995).
90. C. Aguilar, F. Borrull, R.M. Marcé, 'Determination of Pesticides by On-line Trace Enrichment–Reversed-phase Liquid Chromatography–Diode-array Detection and Confirmation by Particle-beam Mass Spectrometry', *Chromatographia*, **43**, 592–598 (1996).
91. J. Slobodnik, S.J.F. Hoekstra-Oussoren, M.E. Jager, M. Honing, B.L.M. van Baar, U.A.Th. Brinkman, 'On-line Solid-phase Extraction–Liquid Chromatography/Particle Beam Mass Spectrometry and Gas Chromatography–Mass Spectrometry of Carbamate Pesticides', *Analyst*, **121**, 1327–1334 (1996).
92. C. Aguilar, F. Borrull, R.M. Marcé, 'Identification of Pesticides by Liquid Chromatography–Particle Beam Mass Spectrometry Using Electron Ionization and Chemical Ionization', *J. Chromatogr. A*, **805**, 127–135 (1998).
93. R.J. Vreeken, W.D. van Dongen, R.T. Gijsen, U.A.Th. Brinkman, 'The Use of Mobile Phase Additives in the

Determination of 55 (polar) Pesticides by Column Liquid Chromatography–Thermospray–Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **54**, 119–145 (1994).
94. D. Barceló, G. Durand, R.J. Vreeken, G.J. de Jong, H. Lingeman, U.A.Th. Brinkman, 'Evaluation of Eluents in Thermospray Liquid Chromatography–Mass Spectrometry for Identification and Determination of Pesticides in Environmental Samples', *J. Chromatogr.*, **553**, 311–328 (1991).
95. D. Volmer, K. Levsen, 'Mass Spectrometric Analysis of Nitrogen- and Phosphorus-containing Pesticides by Liquid Chromatography–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**, 655–675 (1994).
96. D. Volmer, K. Levsen, G. Wünsch, 'Thermospray Liquid Chromatographic–Mass Spectrometric Multiresidue Determination of 128 Polar Pesticides in Aqueous Environmental Samples', *J. Chromatogr. A*, **660**, 231–248 (1994).
97. S. Chrion, S. Dupas, P. Scribe, D. Barceló, 'Application of On-line Solid-phase Extraction Followed by Liquid Chromatography–Thermospray Mass Spectrometry to the Determination of Pesticides in Environmental Waters', *J. Chromatogr. A*, **665**, 295–305 (1994).
98. K.M. Moore, S.R. Jones, C. James, 'Multi-residue Analytical Method for Uron and Carbamate Pesticides in Water Using Solid-phase Extraction and Liquid Chromatographic–Mass Spectrometry', *Wat. Res.*, **29**, 1225–1230 (1995).
99. D.A. Volmer, D.L. Volmer, J.G. Wilkes, 'Multiresidue Analysis of Pesticides by Electrospray LC-MS and LC-MS-MS', *LC-GC International*, **14**, 216–224 (1996).
100. A. Di Corcia, C. Crescenzi, E. Guerriero, R. Samperi, 'Ultratrace Determination of Atrazine and Its Six Major Degradation Products in Water by Solid-phase Extraction and Liquid Chromatography–Electrospray Mass Spectrometry', *Environ. Sci. Technol.*, **31**, 1658–1663 (1997).
101. G. D'Ascenzo, A. Gentili, S. Marchese, D. Perret, 'Development of a Method Based on Liquid Chromatography–Electrospray Mass Spectrometry for Analyzing Imidazolinone Herbicides in Environmental Water at Part-per-trillion Levels', *J. Chromatogr. A*, **800**, 109–119 (1998).
102. G. D'Ascenzo, A. Gentili, S. Marchese, A. Marino, D. Perret, 'Multiresidue Method for Determination of Post-emergence Herbicides in Water by HPLC/ESI/MS in Positive Ionization Mode', *Environ. Sci. Technol.*, **32**, 1340–1347 (1998).
103. C. Crescenzi, A. Di Corcia, E. Guerriero, R. Samperi, 'Development of a Multiresidue Method for Analyzing Pesticide Traces in Water Based on Solid-phase Extraction and Electrospray Liquid Chromatography Mass Spectrometry', *Environ. Sci. Technol.*, **31**, 479–488 (1997).
104. A. Di Corcia, C. Crescenzi, A. Lagana, E. Sebastiani, 'Evaluation of a Method Based on Liquid Chromatography/Electrospray/Mass Spectrometry for Analyzing Carbamate Insecticides in Fruits and Vegetables', *J. Agr. Food Chem.*, **44**, 1930–1938 (1996).
105. D.R. Doerge, S. Bajic, 'Analysis of Pesticides Using Liquid Chromatographic/Atmospheric-pressure Chemical Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 663–666 (1992).
106. K.A. Barnes, R.J. Fussell, J.R. Startin, S.A. Thorpe, S.L. Reynolds, 'Determination of the Pesticides Diflubenzuron and Clofentezine in Plums, Strawberries and Blackcurrent-based Fruit Drinks by High-performance Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1441–1445 (1995).
107. K.A. Barnes, R.J. Fussell, J.R. Startin, M.K. Pegg, S.A. Thorpe, S.L. Reynolds, 'High-performance Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry with Ionization Polarity Switching for the Determination of Selected Pesticides', *Rapid Commun. Mass Spectrom.*, **11**, 117–123 (1997).
108. B. Koppen, N. Spliid, 'Determination of Acidic Herbicides Using Liquid Chromatography with Pneumatically Assisted Electrospray Ionization Mass Spectrometry and Tandem Mass Spectrometry Determination', *J. Chromatogr. A*, **803**, 157–168 (1998).
109. S. Stout, A.R. DaCunha, M. Safarpour, 'Simplified Determination of Imidazolinone Herbicides in Soil at Parts-per-billion Level by Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry', *J. AOAC Int.*, **80**, 426–432 (1997).
110. S. Chiron, S. Papilloud, W. Haerdi, D. Barceló, 'Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography/High-flow Pneumatically Assisted Electrospray MS for the Determination of Acidic Herbicides in Environmental Water', *Anal. Chem.*, **67**, 1637–1643 (1995).
111. J. Slobodnik, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Trace-level Determination of Pesticide Residues Using On-line Solid-phase Extraction-column Liquid Chromatography with Atmospheric Pressure Ionization Mass Spectrometric and Tandem Mass Spectrometric Detection', *J. Chromatogr. A*, **741**, 59–74 (1996).
112. E. Baltussen, H. Snijder, H.-G. Janssen, P. Sandra, C.A. Cramers, 'Determination of Phenylurea Herbicides in Water Samples Using On-line Sorptive Preconcentration and High-performance Liquid Chromatography with UV or Electrospray Mass Spectrometric Detection', *J. Chromatogr. A*, **802**, 285–295 (1998).
113. I. Ferrer, M.-C. Hennion, D. Barceló, 'Immunosorbents Coupled On-line with Liquid Chromatography/Atmospheric Pressure Chemical Ionization/Mass Spectrometry

for the Part per Trillion Level Determination of Pesticides in Sediments and Natural Waters Using Low Preconcentration Volumes', *Anal. Chem.*, **69**, 4508–4514 (1997).

114. W.A. Minnaard, A.C. Hogenboom, U.K. Malmqvist, P. Manini, W.M.A. Niessen, U.A.Th. Brinkman, 'Rapid Target Analysis by On-line Single Short-column Liquid Chromatographic/Thermospray Tandem Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 1569–1574 (1996).
115. A.C. Hogenboom, J. Slobodník, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Single Short-column Liquid Chromatography with Atmospheric Pressure Chemical Ionization-(tandem) Mass Spectrometric Detection for Trace Environmental Analysis', *Chromatographia*, **42**, 506–514 (1996).
116. A.C. Hogenboom, W.M.A. Niessen, U.A.Th. Brinkman, 'Rapid Target Analysis of Microcontaminants in Water by On-line Single-short-column Liquid Chromatography Combined with Atmospheric Pressure Chemical Ionization Ion-trap Mass Spectrometry', *J. Chromatogr. A*, **794**, 201–210 (1998).
117. K. Grob, *On-line Coupled LC-GC*, Heidelberg, Hüthig, 1991.
118. W. Specht, M. Tillkes, 'Gas-chromatografische Bestimmung von Rückständen an Pflazenbehandlungsmitteln nach Clean-up über Gel-chromatographie und Minikieselgel-säulen-chromatographie', *Fresenius' J. Anal. Chem.*, **322**, 443–455 (1985).
119. J.A. van Rhijn, L.G.M.Th. Tuinstra, 'Miniaturization of Size-exclusion Chromatography as a Powerful Clean-up Tool in Residue Analysis', *J. Chromatogr.*, **552**, 517–526 (1991).
120. D.M. Holstege, D.L. Scharberg, E.R. Tor, L.C. Hart, F.D. Galley, 'A Rapid Multiresidue Screen for Organophosphorus, Organochlorine, and *N*-methyl carbamate Insecticides in Plant and Animal Tissues', *J. AOAC Int.*, **77**, 1263–1274 (1994).
121. A. Khoshob, R. Taesdale, 'Application of High-performance Gel Permeation Chromatography Clean-up in the Determination of Hexaflumuron Residues in Soil', *J. Chromatogr. A*, **660**, 195–198 (1994).
122. D.A. Rimmer, P.D. Johnson, R.H. Brown, 'Determination of Phenoxy Acid Herbicides in Vegetation, Utilising High-resolution Gel Permeation Chromatographic Clean-up and Methylation with Trimethylsilyldiazomethane Prior to Gas Chromatographic Analysis with Mass-selective Detection', *J. Chromatogr. A*, **755**, 245–250 (1996).
123. M. De Paoli, M.T. Barbina, R. Mondini, A. Pezzoni, A. Valention, K. Grob, 'Determination of Organophosphorus Pesticides in Fruits by On-line Size-exclusion Chromatography–Liquid Chromatography–Gas Chromatography–Flame Photometric Detection', *J. Chromatogr.*, **626**, 145–150 (1992).
124. J.J. Vreuls, R.J.J. Swen, V.P. Goudriaan, M.A.T. Kerkhoff, G.A. Jongenotter, U.A.Th. Brinkman, 'Automated On-line Gel Permeation Chromatography–Gas Chromatography for the Determination of Organophosphorus Pesticides in Olive Oil', *J. Chromatogr. A*, **750**, 275–286 (1996).
125. E.A. Hogendoorn, G.R. van der Hoff, P. van Zoonen, 'Automated Sample Clean-up and Fractionation of Organochlorine Pesticides and Polychlorinated Biphenyls in Human Milk Using NP-HPLC with Column Column-switching', *J. High Res. Chromatogr.*, **12**, 784–790 (1989).
126. G.R. van der Hoff, A.C. van Beuzekom, U.A.Th. Brinkman, R.A. Baumann, P. van Zoonen, 'Determination of Organochlorine Compounds in Fatty Matrices: Application of Rapid Off-line Normal-phase Liquid Chromatographic Cleanup', *J. Chromatogr. A*, **754**, 487–496 (1996).
127. G.R. van der Hoff, R.A. Baumann, P. van Zoonen, U.A.Th. Brinkman, 'Determination of Organochlorine Compounds in Fatty Matrices: Application of Normal-phase LC Clean-up Coupled On-line to GC/ECD', *J. High Res. Chromatogr.*, **20**, 222–226 (1997).
128. S.M. Gort, R. van der Hoff, R.A. Baumann, P. van Zoonen, J.M. Martin-Moreno, P. van 't Veer, 'Determination of *p,p′*-DDE and PCBs in Adipose Tissues Using High-performance Liquid Chromatography Coupled On-line to Capillary GC – The EURAMIC Study', *J. High Res. Chromatogr.*, **20**, 138–142 (1997).
129. R. Serrano, F.J. López, A. Roig-Navarro, F. Hernández, 'Automated Sample Clean-up and Fractionation of Chlorpyrifos, Chlorpyrifos-methyl and Metabolites in Mussels Using Normal-phase Liquid Chromatography', *J. Chromatogr. A*, **778**, 151–160 (1997).
130. K.K. Verma, A.J.H. Louter, A. Jain, E. Pocurull, J.J. Vreuls, U.A.Th. Brinkman, 'On-line Solid-phase Extraction–Gas Chromatography–Ion Trap Tandem Mass Spectrometric Detection for the Nanogram per litre Analysis of Trace Pollutants in Aqueous Samples', *Chromatographia*, **44**, 372–380 (1997).
131. T.H.M. Noij, M.M.E. van der Kooi, 'Automated Analysis of Polar Pesticides in Water by On-line Solid-phase Extraction and Gas Chromatography Using the Co-solvent Effect', *J. High Res. Chromatogr.*, **18**, 535–539 (1995).
132. H.G.J. Mol, H.-G.M. Janssen, C.A. Cramers, J.J. Vreuls, U.A.Th. Brinkman, 'Trace Level Analysis of Micropollutants in Aqueous Samples Using Gas Chromatography with On-line Sample Enrichment and Large Volume Injection', *J. Chromatogr. A*, **703**, 265–276 (1995).
133. G.R. van der Hoff, F. Pelusio, U.A.Th. Brinkman, R.A. Baumann, P. van Zoonen, 'Automated Solid-phase Extraction Coupled to Gas Chromatography with

Electron Capture Detection: a Combination of Extraction and Clean-up of Pyrethroids in the Analysis of Surface Water', *J. Chromatogr. A*, **719**, 59–67 (1996).

134. G.W. Somsen, I. Jagt, C. Gooijer, N.H. Velthorst, U.A.Th. Brinkman, T. Visser, 'Identification of Herbicides in River Water Using On-line Trace Enrichment Combined with Column Liquid Chromatography/Fourier-transform Infrared Spectrometry', *J. Chromatogr. A*, **756**, 145–157 (1996).

135. H.-J. Stan, F. Schwarzer, 'On-line Coupling of Liquid Chromatography with Thin-layer Chromatography', *J. Chromatogr. A*, **819**, 35–44 (1998).

High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Achille Cappiello and Pierangela Palma
Università di Urbino, Urbino, Italy

1 Introduction 1
2 Instrumental Parameters Involved in Pesticide Analysis 2
2.1 Mass Range 2
2.2 Mass Resolution 2
2.3 Spectral Characteristics 3
2.4 Limits of Detection 3
2.5 Universality 3
3 High-performance Liquid Chromatography/Mass Spectrometry Coupling Techniques 3
3.1 Introduction 3
3.2 Thermospray 4
3.3 Electrospray 5
3.4 Atmospheric Pressure Chemical Ionization 6
3.5 Particle Beam 6
4 Mass Analyzer Alternatives 7
4.1 Introduction 7
4.2 Quadrupole 7
4.3 Ion Trap 8
4.4 Time-of-flight 9
4.5 Sector 9
5 Multiple Mass Spectrometry Detection 10
6 Application of High-performance Liquid Chromatography/Mass Spectrometry in Pesticide Analysis 11
6.1 Thermospray Interface 11
6.2 Electrospray Interface 12
6.3 Particle Beam Interface 14
6.4 Atmospheric Pressure Chemical Ionization Interface 15
Abbreviations and Acronyms 16
Related Articles 16
References 17

In this article, a detailed overview of the current high-performance liquid chromatography/mass spectrometry (HPLC/MS) instrumentation, suitable for the determination of a large number of pesticides, is reported. Particular emphasis is given to the role of the mass spectrometric apparatus involved in each interfacing technique with several types of information regarding the ionization processes, analyzer principles and system configurations. The article gives an up-to-date picture of the current instrumentation that is available and readers will find only brief references to some important historical approaches that are no longer in use.

1 INTRODUCTION

Can mass spectrometry (MS) do for high-performance liquid chromatography (HPLC) what it does for gas chromatography (GC)? This is probably the question that many analytical chemists, biochemists and pharmacologists asked themselves while arguing in front of their ultraviolet (UV) chromatograms. In fact, the use of an ultraviolet/visible (UV/VIS) spectrophotometric detector, which is the most widely used detector for liquid chromatography (LC), is restricted in sample selectivity by the sensitivity in a particular absorption band. Other detectors, such as fluorimetric or electrochemical, are even more limited and rarely employed. When compared with GC, for which a variety of excellent detectors are available, HPLC detectors cannot meet the demand.

MS is the most specific technique for the detection and identification of organic compounds. MS can provide not only molecular weight information but also a wealth of structural details that together give a unique fingerprint for each analyte. MS is the detector with by far the highest information output per unit sample weight. The dynamic combination of gas chromatography/mass spectrometry (GC/MS) has yielded the most specific and sensitive method for the characterization of the components of complex volatile mixtures.

Under this circumstance, the use of a mass spectrometer as an HPLC detector has been widely investigated. Since the majority of all known compounds are amenable to HPLC, the integration of HPLC and MS greatly extends the range of compound classes that can benefit from the particular characteristics of MS detection. With uninterrupted evolution over the last 25 years, it has produced amazing results for both chromatographers and mass spectrometrists to the point that HPLC/MS represents one of the most important tools in the characterization of all organic compounds. Unfortunately, while the evolution of the coupling of GC and MS has led to a single, relatively simple and inexpensive device, progress in HPLC/MS interfacing, far more challenging than that with

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

GC/MS, generated several highly developed interfaces with radically different designs. The reason for the complexity of HPLC/MS is that the domains of LC and MS are often seen as antagonists: the effort of achieving and maintaining the high vacuum required for MS operations conflicts with the intrinsic nature of LC, predominantly operating at high solvent flow rates. In addition, the low tolerance of mass spectrometers to nonvolatile mobile phase components contrasts with the LC dependence on nonvolatile buffers to achieve high-resolution separations. In the effort to overcome a series of evident difficulties, HPLC/MS interfaces use different strategies and, as consequence, they are usually expensive and complicated compared with the intrinsic simplicity of GC/MS.

Difficulties in HPLC/MS can be summarized by the following two points:

1. *Solvent restriction*: an HPLC mobile phase (solvent) is a liquid, and sample components are led to the detector together with the solvent. Because a high vacuum is required during MS operations, not much solvent should enter the MS when it functions as a detector. If the HPLC flow rate is $1\,mL\,min^{-1}$ (liquid) and the mobile phase is water, the gas volume generated in the mass spectrometer is as large as $1200\,mL\,min^{-1}$!
2. *Sample restriction*: LC analyzes all samples that can be dissolved in solvents. The analytes may vary dramatically in weight, polarity and stability and this poses a severe restriction in terms of hardware requirements for both the interface and mass spectrometer.

In this article, recent developments in the use of HPLC/MS methods in the analysis of pesticides are outlined. Particular attention is devoted to the various interfacing techniques currently used in this field, such as thermospray (TSP), electrospray (ES), atmospheric pressure chemical ionization (APCI) and particle beam (PB). A selection of the mass analyzers normally used in conjunction with HPLC/MS interfaces is also considered. In the discussion regarding the performance of the various systems, sensitivity, detection limits, response linearity and other parameters are critically evaluated.

2 INSTRUMENTAL PARAMETERS INVOLVED IN PESTICIDE ANALYSIS

A number of parameters are particularly useful to the analyst in selecting the most appropriate HPLC/MS system to identify and quantify the analytes, depending on the chemical properties and complexity of the sample. Far from being exhaustive, we feel that the criteria described below are particularly important for evaluating the alternatives to accomplish the needs when performing pesticide analysis.

2.1 Mass Range

The difference between the upper and the lower limits of the mass-to-charge ratio (m/z) is called mass range and defined according to Equation (1):

$$\Delta m_{\text{range}} = m_{\text{upper limit}} - m_{\text{lower limit}} \tag{1}$$

Depending on the mass analyzer, the mass ranges can span from 500 to 3000 for a quadrupole or an ion trap to no upper limits for a time-of-flight (TOF) mass analyzer. Since the molecular weight of most pesticides is <500, the higher cost of an extended mass range is not required.

2.2 Mass Resolution

Mass resolution is the capability of a mass analyzer to resolve one mass from the adjacent mass and is described by Equation (2):

$$R_{\text{m}} = \frac{m}{\Delta m_{\text{resolution}}} \tag{2}$$

where m is the measured mass and $\Delta m_{\text{resolution}}$ is the difference between two adjacent mass peaks. A resolving power of 100–1000 (1000–2000) is typical of low-resolution analyzers such as quadrupoles and ion traps and implies unit resolution. This mode of operation is particularly convenient when using HPLC/MS, since possible interferences can be removed chromatographically. The advantage offered by electron ionization (EI) is that the acquired spectra can be compared with those in libraries for the fast identification of unknown compounds. This interesting feature is currently available only with a PB interface. TOF analyzers can operate at both low and high resolution. Sectors operate at high resolution for accurate mass measurements with a resolving power of 5000–10 000. This allows the separation of more than one analyte with the same nominal mass and it represents an unsurpassed tool for the determination of elemental composition. It is particularly important to determine the increasing presence of pesticides and their metabolites in the environment. The performance of a sector can be enhanced, providing additional specificity, when more than one analyzer is placed in series either in four sectors or a hybrid configuration. On the other hand, the cost and the operating skills required for such an instrument are elevated. Triple quadrupoles and ion traps allow mass spectrometry/mass spectrometry (MS/MS) configurations at a lower price with competitive sensitivity.

2.3 Spectral Characteristics

The characteristics of a mass spectrum can vary widely, depending on the ionization technique, interface and analyzer used. Highly informative, fragmentation spectra are needed in those applications where structural elucidation and identification are under investigation. Fragmentation spectra are obtained when electrons coming from a hot filament (EI) ionize the analyte molecules. The PB is the only interface that provides fragmentation spectra in single analyzer configuration on both quadrupole and sector analyzers. Collisions within the interface [source–collision-induced dissociation (CID)], available with ES, or tandem mass analyzers (MS/MS) can be used to obtain fragmentation spectra.

It has been demonstrated that even polar or ionic pesticides can be analyzed under EI conditions.[1] In this case and when the molecular ion intensity is too low, alternative ionization techniques can provide simple spectra containing the molecular ion. This is the case with APCI, ES and TSP. CID can provide additional structural information.

2.4 Limits of Detection

When performing pesticide analysis, one should keep in mind their extremely low concentrations in the matrix and all the possible interferences. Selectivity can be achieved in any of different steps of the analysis, from sample preparation to the different alternatives of liquid chromatography/mass spectrometry (LC/MS) equipment. Instrument noise can originate either from the components of the system or from the chemical background and may severely complete with the very low signal of the analyte at ultratrace levels. APCI, ES and TSP allow the detection of most pesticides at the picogram level, whereas PB is less sensitive, with detection limits in the range of nanograms. Improving the detection limits depends not only on the interface and ionization technique chosen, but also on the analyzer and on the analyzer configuration in the case of MS/MS. With the latter, increasing the signal-to-noise ratio allows the detection of extremely low concentrations of substances.

2.5 Universality

When dealing with complex mixtures or real samples, it is very welcome to count on an instrument that can respond to all components, especially when unknown molecules need to be characterized. This is particularly true in pesticide analysis when metabolites or adducts need to be characterized. Universality is in this case a fundamental requirement to guarantee the complete identification and quantification of the sample components. Once again, one can play with the selectivity to restrict the field of interest. The polarity of the molecule plays a major role in choosing the ionization technique. EI is *the* universal ionization technique for nonpolar, polar nonionic and ionizable substances. The use of EI is limited to the PB, but this does not represent a limit in pesticide analysis since the molecular weight of most pesticides is far lower than 600 u. ES is preferable with ionic compounds and APCI and TSP produce a very good response with polar nonionic and ionizable compounds.

3 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY COUPLING TECHNIQUES

3.1 Introduction

In its infancy, HPLC/MS was considered as an adaptation of GC/MS towards thermolabile molecules carried by a liquid medium. In the mid-1970s, EI was like a flagship for MS. A reliable, reproducible, library-matchable EI spectrum was the key for a winning GC/MS technology, so its adaptation to a vaporizable eluate was considered a valid approach for the emerging LC/MS 'hyphenation'. At that time, the extension of the analyzable mass range was limited and so was the mass range of the analytes convertible into the gas phase. The HPLC eluate was therefore forced into a classical EI ion source directly or through a mechanical adapter. The first case, called direct liquid introduction (DLI), was first developed by Baldwin and McLafferty[2] and consisted in direct vaporization and ionization of a split HPLC column effluent. The reduced flow rate ($10–50\,\mu L\,min^{-1}$) allowed normal MS operation but, because of the solvent vapor residues, imposed chemical ionization (CI) as the only ionization choice. Simplicity was good but the sensitivity was only acceptable and did not match the performance of GC. The mechanical alternative to DLI into a conventional EI/CI source was the moving belt interface.[3] This innovative interface consisted of an auxiliary vacuum chamber through which a Kapton moving belt carried the column eluate and evaporated the solvents. The dry solutes were carried into the ion source, vaporized and ionized. After a cleaning process, the belt was ready for a new cycle. Solvents were removed completely and EI was achievable. Major drawbacks were related to the mechanical complexity of the interface, loss of chromatographic resolution and memory effects on the reused belt. Both interfaces were abandoned and were replaced by the much more efficient and reliable PB interface. The PB interface, first developed as monodisperse aerosol generator (MAGIC) LC/MS[4] to underline the monodisperse aerosol mechanism through

which the liquid effluent is introduced into the ion source, is the best EI-based LC/MS interface currently in production.

LC/MS interfacing has also introduced a new concept: the integration between coupling mechanism and sample ionization in one combined process. This innovative approach opened the way to new 'soft' ionization techniques, namely TSP, fast atom bombardment (FAB), and atmospheric pressure ionization (API), which have revolutionized MS with their ability to ionize high-molecular-weight molecules and, in some cases, allowing the detection of picograms or less of material. These techniques, producing poor or no fragmentation of the analytes, with the absolute predominance of the molecular ion in the mass spectrum, have their best attribute in a higher sensitivity for several compounds. Differently from EI spectra, which reproduce what was in the original sample, API encumbers interpretation with artifacts that were not present in the sample (adducts, polymers, etc.). For instance, aldicarb only produces significant $[M + Na]^+$ without a molecular ion $[M + H]^+$, which could cause difficulty for a novice. In addition, there is no capability for commercial library identification of unknowns or 'unsures' in the hands of novices. The lack of typical EI information capability was partially compensated by a contemporary evolution of high-performance, coupled MS analyzers (MS^n) that, although complex and expensive, ensured a valid identification criterion for the target compounds. The family of soft ionization techniques, readily embraced by protein chemists for their high-mass capabilities, were slowly converted also to the detection of low-mass analytes at ultratrace concentrations. Only FAB ionization, which opened the way for the use of MS in biomedical applications, was rarely applied to samples of environmental concern.[5]

In conclusion, whatever the interface principle being used, it must be subject to three conversion processes, always required when LC is coupled to MS:

1. An *evaporation process* is required to convert the analyte from a liquid phase to a rarefied gas phase.
2. A *pressure-reduction process* is required to convert the analyte, separated at atmospheric pressure, into high vacuum for mass analysis.
3. An *ionization process* is required to convert an often-neutral sample into detectable ions.

The occurrence of the three processes may vary in sequence in the various interfaces but they must be always present. On this basis, two groups can be highlighted: the API and TSP interfaces, in which the ionization step precedes the vacuum conversion, and the rest of the interfaces in which the ionization occurs at high vacuum.

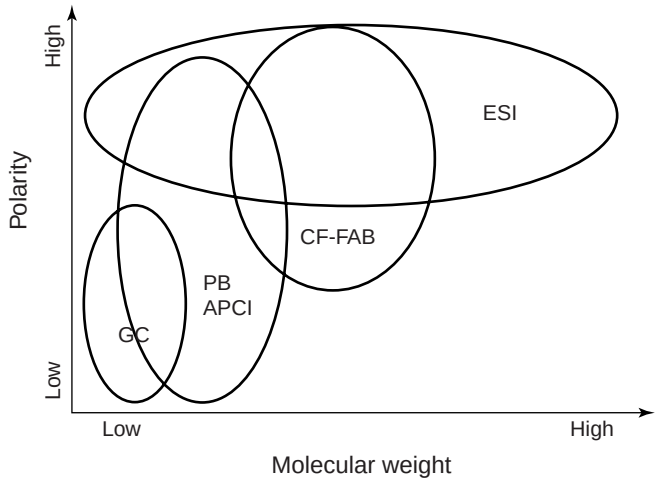

Figure 1 Domains of application of LC/MS interfaces.

Daring combinations of up-to-date ionization techniques and advanced analyzers keep the LC/MS market very unpredictable but arguably, API and PB (TSP having largely been superseded by APCI) are now well established to be considered as the foundations for future interface developments.[6] Figure 1 illustrates the domain of each interface with respect to GC/MS.

3.2 Thermospray

The TSP interface (Figure 2), developed in 1980 by Vestal et al.,[7] was one of the first attempts at LC/MS techniques using soft ionization. A TSP interface addresses two fundamental problems: it reduces the pressure of the residual HPLC solvents into the mass spectrometer and self-generates ions of the analyte. Both results are obtained by applying heat to the incoming HPLC eluate.

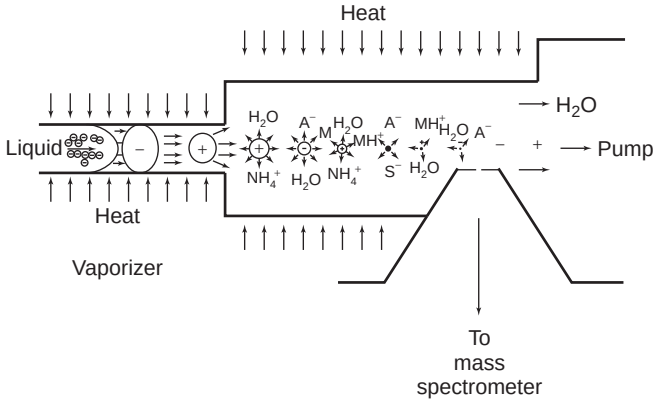

Figure 2 Schematic diagram of TSP interface. (Reproduced from Blakely et al.[7] with permission from John Wiley & Sons Inc.)

The liquid effluent is forced into a heated capillary tubing that ends inside a heated desolvation chamber. The heat vaporizes the liquid while it travels through the interface and the majority of the solvent vapor is removed by an auxiliary vacuum pump placed at the end of the path. Only a small fraction of vapor is admitted into the mass spectrometer through a small orifice in the chamber, thus achieving the required pressure-reduction step. When properly controlled, the TSP interface is capable of accepting aqueous HPLC effluents at rates up to $2\,mL\,min^{-1}$.

Nonvolatile molecules are retained inside the aerosol droplets while they diminish in size during the vaporization process. An electrolyte, added to the HPLC mobile phase, produces an initial electrostatic field on the droplet surface and, in a way that resembles ES ionization, promotes the analyte ionization: the rapidly desolvating charged droplets increase the surface field while diminishing their size, until the point at which the charged analytes are ejected into the vapor stream. Ions are extracted from the desolvation chamber and accelerated toward the mass analyzer. Ammonium acetate can be added postcolumn and is capable of enhancing ionization of target analytes.

TSP produces molecular and adduct ions of the nonvolatile analyte with poor fragmentation. Molecules that are neutral in solution may produce $[M + H]^+$ ions during the TSP process while adding postcolumn ionic additives to the HPLC mobile phase can generate negative ions (NIs). A filament, producing gas-phase electrons, can be added to the chamber, inducing CI reactions with the solvent used as the reagent gas.

TSP is limited by the use of aqueous mobile phases. Lack of structural information is not compensated by higher sensitivity, which is two orders of magnitude lower than in ES. TSP may produce more fragmentation than FAB but is not comparable to the superior identification capability of PB that shares the same applicability. TSP shows its best attribute for ionic and polar compounds with molecular weights <2000 u.

3.3 Electrospray

ES[8,9] has revolutionized MS with its ability to ionize high-molecular-weight molecules and detect even femtogram levels of material. The ES LC/MS interface relies on the API technique for ion generation. Similarly to TSP, the ionization of the analytes takes place into fast evaporating liquid droplets but, instead of using an electrolyte and a high temperature, an electric potential is applied to the incoming HPLC effluent at atmospheric pressure (Figure 3).

The intrinsic simplicity of the ES design makes it a successful tool for LC/MS interfacing. A strong electric

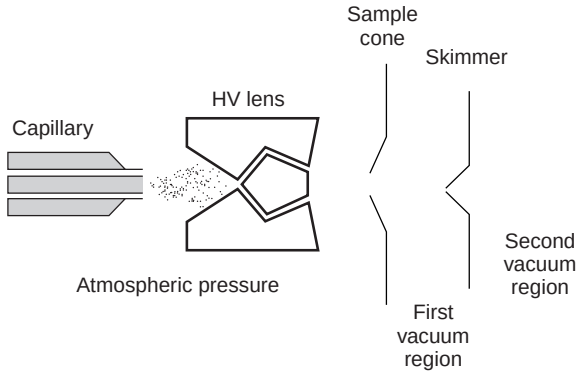

Figure 3 Schematic diagram of ES interface.

field (2–4 kV) is applied between a hollow needle, through which the HPLC effluent emerges, and a collector. The high field at the tip of the needle produces a cone-shaped liquid meniscus from which a spray of highly charged droplets is formed. During their travel from the needle to the collector, the droplets evaporate while diminishing their size, thus allowing the electric charge to concentrate at the droplet surface. At this point, a prerequisite for the final gas-phase ion generation is that the analyte exists as an ion in solution or could be easily ionized through a charge transfer when still in solution.[10] If a species is present with more than one ionizable site, then ES will also produce more than one charge in the gas-phase ions. The benefit is evident with high-molecular-weight substances where the multiple charged ions, generated by ES, shrink the mass spectrum into the mass range of a typical mass analyzer. The charge concentration, observed during the droplet evaporation, weakens the natural solvent cohesion up to the point at which electrohydrodynamic disintegration into many smaller droplets occurs. This mechanism ends when the droplet size reaches the Rayleigh limit (the point at which repulsive forces overcome the cohesive forces of the solvent) and eventually it results in the expulsion of the ionized analyte into the gas phase. This ion evaporation process is believed to be the major mechanism for ionizing smaller molecules.[11]

At atmospheric pressure, the solvated ions pass the collector and, following their momentum, travel toward the analyzer. A gas curtain (usually nitrogen at 1 Torr pressure), placed between the collector and the vacuum region, impacts the incoming ions quickly removing the last solvent molecules and generating low-internal-energy quasimolecular ions. In fact, the energy of those impacts, controlled by electric potential applied to the needle, can be a useful source of CID spectra of the analyte.[12] The CID spectrum, generated in the ES region, is qualitatively similar to the spectrum recorded on a tandem MS instrument and can provide

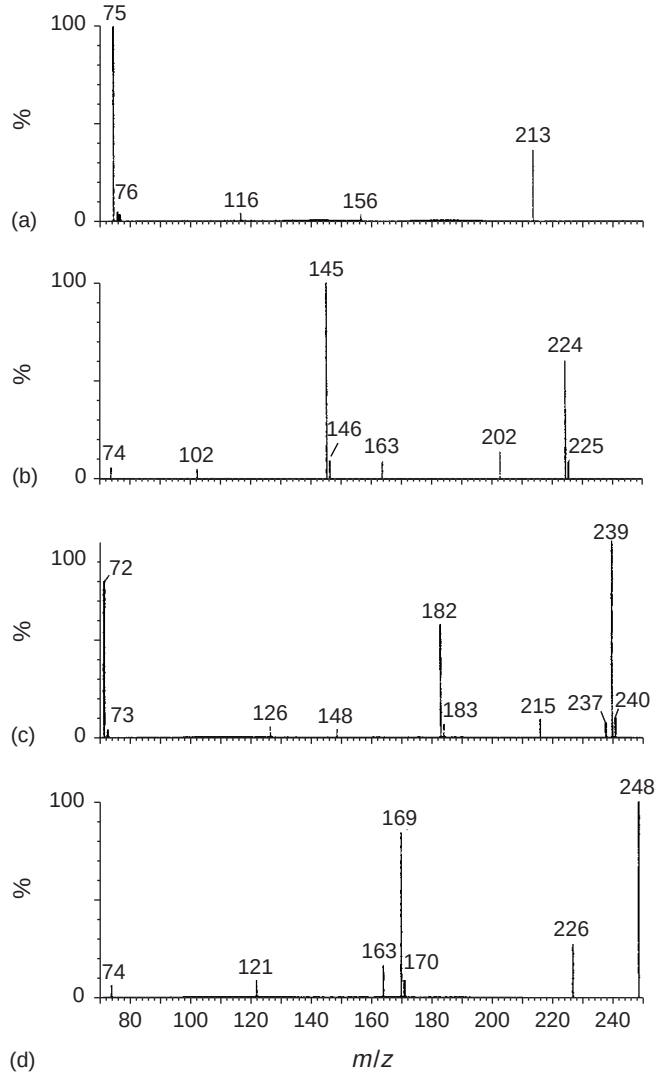

Figure 4 Full-scan background-subtracted spectra taken from the average of the peaks for (a) butocarboxim, (b) carbaryl, (c) primicarb and (d) mercaptodimethur acquired by on-column injection of 5 ng each of the analytes. (Reprinted with permission from A. Di Corcia et al., *J. Agric. Food Chem.*, **44**, 1930 (1996). Copyright 1996 American Chemical Society.)

the structural information necessary for the analyte confirmation (Figure 4).[13] However, the ion pattern does not show the high reproducibility of that with typical EI. Unlike the analysis by tandem MS, where the ions are separated before CID, in a source CID the isolation of any analyte ion relies only on the LC separation. ES is also very susceptible to the chemistry of solutions and all the equilibria involved may affect the final sample result.

Conventional ES usually operates between 1 and $10\,\mu L\,min^{-1}$. Pneumatically assisted aerosol generation [ionspray (IS)] allows the use of higher flow rates.

3.4 Atmospheric Pressure Chemical Ionization

ES ionization is best suited for high polar species or molecules with an acidic or basic site. An alternative for less polar molecules is APCI.[14] APCI relies on an API interface as in ES, but produces the ions directly in the gas phase after the solvent and solute have been completely evaporated. In order to promote vaporization, the HPLC eluate, at a flow rate of $0.5–2\,mL\,min^{-1}$, is converted into an aerosol and rapidly heated to 450–550 °C inside an ionization chamber. Ionization is accomplished with a source of electrons addressed on-axis with the heated spray. The electrons are usually supplied by a corona discharge source but a ^{63}Ni emitter can also be used. A plasma of reagent ions is produced at atmospheric pressure in the ion source by EI of source gases and solvents. At this pressure the chance of interactions between reagent ions and analyte molecules is very high and protonated ions are normally generated. The mass range is limited compared with ES but the applicability of APCI is wider, spanning from ionizable to nonpolar compounds. ES and APCI can be configured on the same instrument, enhancing the problem-solving capabilities of both.

3.5 Particle Beam

The PB interface[15] can be considered as the liquid-phase counterpart of a GC/MS interface. Its design is not new but, differently from other attempts, survived the tough competition with the emerging API techniques and now plays a key role in the solution of several problems. The PB can be considered as a 'transport' interface in which the majority of mobile phase solvents are removed and the analyte is transferred into the mass spectrometer for the final ionization. The breakthrough of the PB interface is represented by the conversion of the HPLC effluent into an aerosol of high-surface-area droplets and, after a rapid evaporation, into a beam of solute particles. Even though the development of PB took advantage of several detailed implementations involving different aspects of its functioning,[16] it still relies on three key components (Figure 5): (1) aerosol generator, (2) desolvation chamber and (3) momentum separator.

The production of a fine spray can be accomplished by either a pneumatic or a thermal aerosol generator designed to accommodate mobile phase flow rates up to $0.5\,mL\,min^{-1}$. A dispersion gas, usually helium, is driven coaxially to the liquid jet. A capillary-scale version of the interface enhances the overall performance at mobile phase flow rates of only a few microliters per minute.[17] The atmospheric pressure desolvation of the aerosol droplets produces a mixture of dispersion gas, solvent vapors and solute particles, but only the last (the high-momentum fraction) reaches the ion source, the rest being

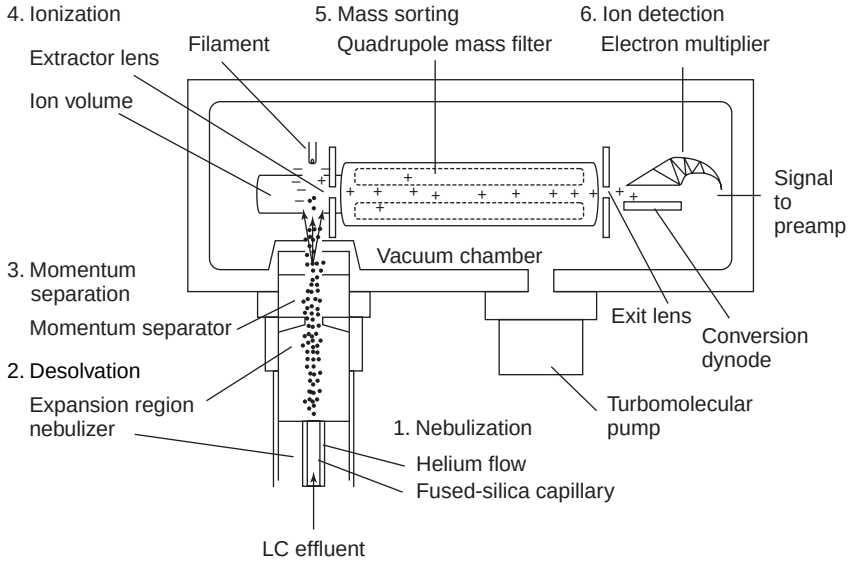

Figure 5 ThermaBeam LC/MS interface and vacuum chamber processes. (Reproduced by permission of Waters Corporation from *Waters ThermaBeam Mass Detector Installation Manual*.)

removed in the momentum separator. This effective and simple interfacing approach relies on a classical EI/CI ion source and generates reproducible, library-matchable spectra for easy compound identification. This is ideal for those applications where structure information for small (<1000 u), HPLC-amenable molecules are needed. Even though picogram-level sensitivity is hardly achieved, PB plays a unique role in the world of HPLC/MS interfaces.

4 MASS ANALYZER ALTERNATIVES

4.1 Introduction

A mass analyzer represents the heart of a mass spectrometer. Choosing among different mass analyzers is necessarily the first move and should precede any other consideration when approaching MS. They all do basically the same job, some just do it better than others and at a different cost. Mass analyzers come in a variety of shapes and sizes but they all require three conditions without which ion detection would be impossible: (1) high vacuum, (2) charged analytes and (3) gas-phase ions.

Besides these common features, mass analyzers differ significantly. They all use a combination of electric or magnetic fields or both in order to separate the ions in space or time. Each analyzer may excel in mass resolution, extension of the mass range, detection sensitivity or scan speed. Some of them may include the possibility of generating MS/MS or MS^n ($n = 1, 2, 3, \ldots$) spectra following a high- or low-energy CID event. All of them are suitable for coupling with chromatography but at a different cost and with different results. In conclusion, as highlighted in Table 1, where a comparison among the most common analyzers is reported, each one should balance advantages and limitations in order to suit better the LC/MS needs.

4.2 Quadrupole

A quadrupole mass analyzer, as its name implies, consists of a set of four parallel electrodes with either a circular or hyperbolic cross-section. The electrodes, placed at

Table 1 Key characteristics of different MS analyzers

Mass analyzer	Separation principle	Ion production	Resolution	Mass range (u)	Scan speed ($u\,s^{-1}$)
Quadrupole	Electric field (filter)	Continuous	Low	3000	4000
Ion trap	Electric field (stability)	Pulsed	Low	3000	4000
TOF	Electric field (velocity)	Pulsed	Low (or high)	Unlimited	10^6
Sector	Magnetic field (filter)	Continuous	High	20 000	Slow

Reproduced by permission of Chem-Space Associates from R. Willoughby, E. Sheehan, S. Mitrovich, *A Global View of LC/MS*, Global View Publishing, Pittsburgh, PA, 526, 1998.

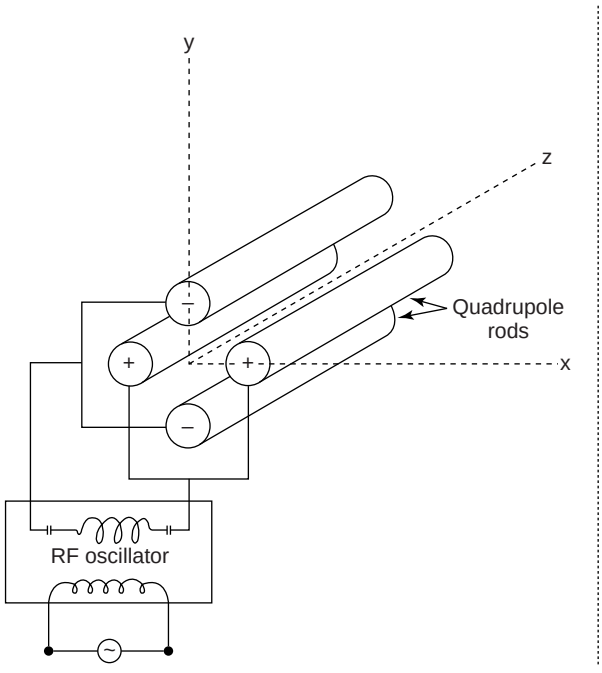

Figure 6 Schematic diagram showing arrangement of quadrupole rods and electrical connection to RF generator. (Reproduced by permission from J. Throck Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, New York, 1997, 74.)

the four angles of a square, are charged with two voltage components: a direct current (DC) and an alternating current (AC) oscillating in the radiofrequency (RF) range. Opposing sets of rods share the same voltage pattern but are charged with a different polarity (Figure 6). The electric field created in the space delimited by the four rods is called the quadrupole field. The quadrupole field shows some peculiar characteristics that can influence the path of a moving charged particle. When the quadrupole field is placed between the ion source and the detector, the stability of the ion trajectory is directly related to the RF/DC voltage ratio applied to the electrodes so that only a specific m/z value is allowed into the detector for a given voltage ratio. By ramping the electrode voltages, but keeping the RF/DC voltage ratio constant, a complete range of masses can be passed to the detector and the mass spectrum recorded. Therefore, the quadrupole analyzer is basically a mass filter whose mechanism can be explained by the superimposition of a high-pass mass filter and low-pass mass filter. The positively charged rods permit the transmission of all the ions above a critical m/z value while the negatively charged rods permit the transmission of those ions below a certain m/z value. The overlapped region created by the two filters forms a band-pass for a specific mass-to-charge ratio whose width can be varied by altering the RF/DC voltage ratio of the opposite sets of rods. Mass resolution is directly related to this bandwidth but it usually remains low (1000 at m/z 1000). The mass range is also limited, often being lower than 1000 u, but in some cases it can be extended up to 4000 u. This limit can be easily overcome when an ES LC/MS interface is in use.

A quadrupole analyzer is less costly than other analyzers and permitted the wide diffusion of MS detectors among the chromatographic community. Today, the quadrupole can be coupled to HPLC by the means of any available interface. Its typical sturdiness and reliability make it the optimum choice for high-throughput applications where dirty samples and high backgrounds may affect the quality of the results. It has evolved also as a tandem analyzer (MS/MS) in a configuration known as a triple quadrupole. The third quadrupole is an RF-only system and for its characteristics is used as collision region for CID experiments. Quadrupoles are also used in hybrid mass spectrometer configurations in which different analyzers are coupled together.

4.3 Ion Trap

The ion trap analyzer is very close in principle to a quadrupole analyzer in which the positively charged rods are bent and joined together to form a donut-shaped electrode (ring electrode) and the remaining set of rods are transformed in two hemispherical electrodes placed above and below the donut (end-caps). This configuration (Figure 7) is called a quadrupole ion trap. The similitude does not end here: in fact, as in the quadrupole analyzer, an RF oscillating voltage is applied to the ring electrode

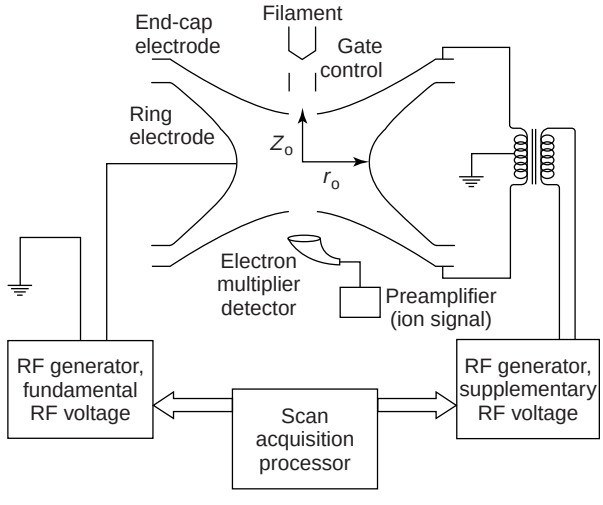

Figure 7 Schematic diagram of ion trap mass spectrometer. (Reproduced by permission from J. Throck Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, New York, 1997, 81.)

although, differently from a quadrupole, a DC potential is not required. The end-caps can be charged with a ground, DC or AC potential. While the quadrupole analyzer is basically a mass filter that temporarily traps the ions into a two-dimensional space only during their stable trajectory inside the quadrupole field, the ion trap analyzer forces all the ions together in three-dimensional space inside the trap until, depending on their m/z value, they are ejected towards the detector. The extraction of the ions from the trap is m/z specific and is obtained only when an RF potential is applied to the end-caps. A full mass spectrum of the analyte is recorded by ramping the ring electrode potential. As the ring electrode potential changes, the frequency of the ion stability also changes until the end-cap frequency matches the resonance frequency of an ion. All the ions sharing the same mass-to-charge ratio become excited into an expanding oscillating motion up to the point at which they are expelled from the trap. The mass resolution is comparable to that of a quadrupole analyzer and is obtained confining the ions with a given mass-to-charge ratio in similar spatial position inside the trap volume. A buffer gas such as helium, admitted into the trap at a pressure of 10^{-3} Torr, promotes this process by cooling the kinetic energy of the ions. This process reduces time dispersion when the ions are expelled for recording.

The trapping principle is highly efficient especially when compared with a quadrupole analyzer where all the ions, except those with a specific mass-to-charge ratio that reach the detector, are wasted. This may generate, in theory, very sensitive full spectra not far from the typical selected ion monitoring (SIM) performance. Unfortunately, the ion accumulation process cannot proceed indefinitely since the spatial focusing of the trap may promote space-charge disturbances among too close oscillating ions, with loss of resolution and errors in mass scale calibration. Thus, ion access to the trap is pulsed and an automatic gain controller (ACG) regulates their admission from an external ion source. It is noteworthy that samples that give good results in a quadrupole instrument may be too concentrated, by one or two orders of magnitude, for ion trap MS. In an ion trap analyzer, MS/MS experiments are spaced sequentially in time instead of space. The precursor ion is kept in place by applying an RF potential on the end-caps and selectively ejecting the remaining ion from the trap. A resonating frequency that corresponds to the isolated mass is then applied to the end-caps, causing the precursor ion to increase its oscillation. After numerous collisions with helium atoms, product ions are generated and recorded. Instead of recording all the fragment ions, a new precursor ion can be selected and with a new sequence of operations MS^n experiments can be performed.

4.4 Time-of-flight

The TOF mass analyzer is, in principle, very simple, being based on the fact that ion velocity is mass dependent. A 'linear' TOF consists of an ion source, a field-free flight tube and a detector. As soon as the ions are generated in the ion source they are accelerated along the flight tube by a burst of electric potential. For every pulse a group of ions will acquire the same kinetic energy and will start at the same time. Based on this fact, each mass entering the flight tube has a different velocity: the ions of lowest mass have the highest speed. The initial group of ions will separate as they proceed along the flight tube and arrive at the detector at different times. Hence, the arrival time is directly related to mass or, in this case, m/z. In this configuration though, a TOF instrument shows a poor mass resolution ($\sim$1000 at m/z 1000). Two factors are responsible for this effect: variable ion formation time and different initial velocity. Both factors will cause ions, of the same mass, to have different arrival times at the detector. A common variation of TOF, which allows a consistent improvement of the mass resolution, incorporates an electrostatic mirror (reflectron) into the field-free region to compensate for energy differences at the start. This field reflects the ions in a position close to the ion source where they are detected. The fastest ions, with the same mass-to-charge ratio, penetrate deeply inside the reflectron field before being reflected. The longer path to the detector is covered in the same time as the slower ions, thus increasing mass resolution.

TOF has characteristics that are ideal for structural elucidation problems, namely high sensitivity in the scan mode and high scan speed. The last feature makes it a valid contender for high-speed GC/MS applications and a sensitive detector for full-scan acquisition. It is noteworthy that SIM acquisition, although possible, does not improve the analyte response, the relative signals being extracted when all the ions have already been collected. Recent technology developments have also added high resolution and high mass accuracy to the list of TOF capabilities and it often used as second analyzer in MS/MS instruments (Q-TOF, EB-TOF, etc.). Newer Q-TOF instruments are appearing with a resolution of 15 000.

4.5 Sector

A sector analyzer is usually referred to as a sequence of electrostatic and magnetic fields located along the ion path with the purpose of selecting a highly resolved isomass beam (same m/z value). Sector instruments have been always considered the reference for high-performance, high-resolution MS, but their cost and complexity have often confined MS into an elitist club.

In a sector instrument, the ions, accelerated from the ion source at a voltage of 4–8 kV, are deflected in a circular trajectory inside a strong, sector-shaped magnetic field (B) (Figure 8). The radius of the trajectory is m/z dependent, being created by the balance between the centripetal Lorenz force and the ion momentum, so that ions of various mass-to-charge ratios are separated in space and exit the sector at different positions. The ion detector is located at the end of a single ion beam whose value depends on either the magnetic field or the accelerating potential. These fields can be varied opportunely for mass spectra recording. In this simple configuration, the mass resolution is heavily affected by the thermal energy dispersion at the moment of the ion formation. In order to increase the mass resolution power of the instrument, a second radial electrostatic sector (E) is added on the ion beam path before or after the magnetic sector, thus creating a variety of tandem configurations. In the electrostatic sector the energy dispersion of the ions is converted into a spatial dispersion of isomass ions at different velocities. Then, entering the magnetic field in different positions, the ions undergo compensating deflections with a final refocusing of the beam. Sector instruments separate ions simultaneously but, since only one beam is allowed to the detector, no more than one species can be collected at a time. A multichannel array detector allows simultaneous detection over a larger area, collecting several ions at the same time.

Sector MS is still the instrument of choice when high-resolution ($R \leq 100\,000$) mass spectra are required. MS/MS operations, either with a four-sector (EBEB, BEBE) or a hybrid configuration (EBQQ, BEQQ and BE-TOF), can take advantage of the high-resolution isolation of the precursor ion or the higher kinetic energy available for CID experiments.

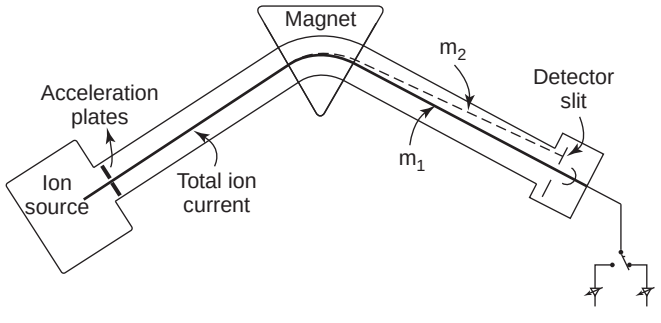

Figure 8 Schematic diagram of magnetic sector mass spectrometer. (Reproduced by permission from J. Throck Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, New York, 1997, 66.)

5 MULTIPLE MASS SPECTROMETRY DETECTION

A sequence of mass spectral detection in a chain of fragmentation events originating from an analyte ion is called MS^n ($n = 1, 2, 3, \ldots$), in order to highlight the number of stages of mass analysis involved in the process (Figure 9). MS^n is a very general term and does not define any particular instrument configuration. Generally, any process leading to the detection of fragment ions originating outside the ion source can be included in this category. The simplest form of MS^n detection is a sequence of two mass spectral stages (MS/MS). In this mode, the first stage of mass analysis (MS_1) is used to isolate a selected ion (*precursor ion*) into a collision region. The precursor ion is

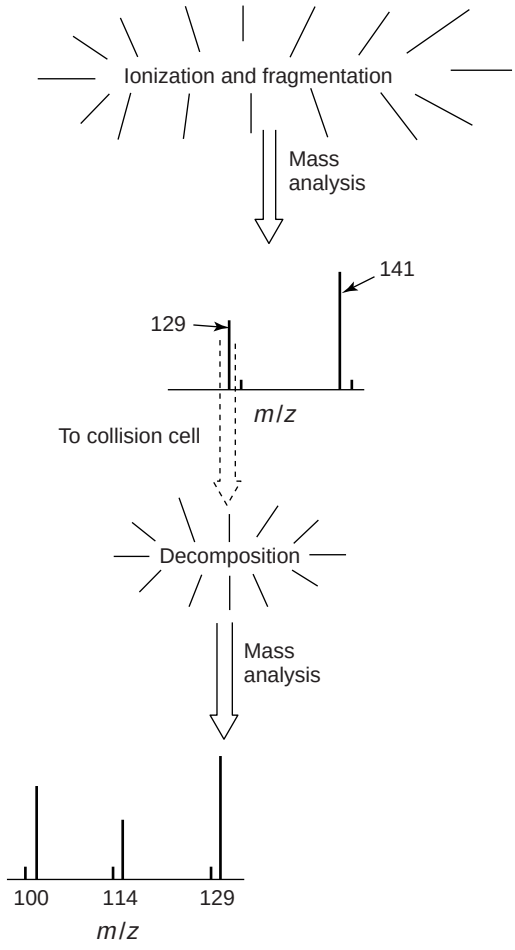

Figure 9 Conceptual representation of the technique of MS/MS. Ions with m/z 129 are selected by the first mass spectrometer; these ions are directed into a collision chamber and the decomposition products are analyzed by the second mass spectrometer to produce the product-ion spectrum of m/z 129. (Reproduced by permission from J. Throck Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, New York, 1997, 110.)

Table 2 MS/MS capabilities of different analyzers

Analyzer	MS/MS configuration	Product ion mode	Precursor ion mode	Neutral loss	Utility	Resolution	Collision energy
Quadrupole	QQQ	✓	✓	✓	Sensitive in SRM	Low	Low
	EBQQ	✓	✓	✓	High precursor selectivity	Mixed	Both
	BEQQ	✓					
	Q-TOF	✓			Sensitive product ion spectra	High	Low
Trap	MS^n	✓			Versatile/affordable	Low	Low
TOF	Q-TOF	✓			Sensitive product ion spectra	Mixed	Low
Sector	EB	✓				Low	High
	BE	✓				Low	High
	BEB	✓			High performance	Mixed	High
	EBE	✓			High performance	Mixed	High

Reproduced by permission of Chem-Space Associates from R. Willoughby, E. Sheehan, S. Mitrovich, *A Global View of LC/MS*, Global View Publishing, Pittsburgh, PA, 113, 1998.

excited, usually by interaction with a collision gas (CID), leading to the formation of fragment ions (*product ions*). Product ions are eventually mass resolved in a second stage of mass analysis (MS_2). The energy involved in the decomposition process influences the extent and quality of fragmentation. As a rule of thumb a higher collision energy, like those achieved when MS_1 is a sector analyzer, leads to greater decomposition of any analyte precursor ion and, ultimately, more structural information. Both sensitivity and specificity are greatly enhanced in the course of an MS/MS experiment. The coupling of multiple stages of mass analysis provides further structural information. In the selected reaction monitoring (SRM) mode, a number of selected product ions are continuously monitored while a precursor ion is admitted in the collision region by means of the first mass analysis stage. This mode of operation, which resembles the selected ion monitoring (SIM) mode in EI, yields the highest specificity for target analytes. The sharp increase in the signal-to-noise ratio after any stage of MS/MS operation may lower the analyte detection limit.

As shown in Table 2, MS analyzers can be arranged in different configurations to realize an MS/MS detector. Most of them can be optically coupled in a twin or hybrid configuration. In an ion trap mass spectrometer, where mass analysis stages are separated in time instead of space, the MS/MS process does not require an optically linked chain of analyzers and can, in theory, be repeated indefinitely. Double-focusing magnetic instruments, designed to focus or select ions based on both energy and momentum, can be used alone in a variety of different low-resolution scan modes to allow products of dissociation in the field-free region to be detected (energy scans, linked scans). A sort of CID, where the precursor ions can be only chromatographically resolved, is also available with the ES interface by means of the curtain gas region. MS^n detection is of a crucial importance in HPLC/MS interfacing where most of the ionization techniques are 'soft' and considered insufficient for structural elucidation or analyte confirmation.

6 APPLICATION OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY IN PESTICIDE ANALYSIS

The heterogeneous group of chemicals known as pesticides can be found everywhere in the environment and their occurrence, even at low concentration, can be a serious threat to human health. In many cases capillary GC combined with the appropriate selective detectors or a mass spectrometer is employed. However, many pesticides currently in use are not amenable by GC owing to their polarity, thermolability and/or low volatility. In this case a separation in a liquid phase is more appropriate and HPLC/MS plays an important role.[18–21]

6.1 Thermospray Interface

Several classes of pesticides have been analyzed with TSP, demonstrating the extensive use of this technique in environmental analysis. Even though it has been replaced by newer interfaces, its importance in the field is worthy of a detailed description. Although TSP ionization is known to be 'soft', producing a poor fragmentation, several techniques can be used to enhance the performance of this interface, increasing structural elucidation capabilities. These include the control of the vaporizer temperature, the use of MS/MS and the possibility of working in positive ion (PI) and NI modes.

The performance of TSP ionization over a broad range of pesticides has been investigated in a number of interesting studies. Bellar and Budde developed a broad-spectrum method for the determination of nonvolatile target compounds in aqueous samples.[22] Several classes

of pesticides and herbicides, such as *N*-substituted amides, carbamates, sulfonylureas, triazines, phenylureas and others, were used to spike aqueous samples at 2–50 and 20–500 ppb levels for liquid–liquid and liquid–solid extraction. The TSP source was operated with the filament off, the discharge off and the repeller off and the detection was obtained in the PI mode. The LC mobile phase contained ammonium acetate alone or mixed with ammonium formate as ionizing reagent to perform volatile salt ionization (VSI) producing protonated molecular ions and other adduct $[M + H]^+$ and $[M + NH_4]^+$ ions. The detection limits obtained using liquid–liquid extraction ranged from 0.2 ppb for cynazine to 18 ppb for linuron, while those obtained with liquid–solid extraction were higher owing to the small size and limited capacity of the cartridges used. Combining the information obtained from the LC separation and those from MS, this approach allowed the satisfactory identification of the target analytes.

Volmer et al. conducted an interesting study on the dependence of the temperature and salt concentration on the ion abundance in TSP mass spectra of several pesticides under CID conditions.[23] In many instances, raising the vaporizer or the gas-phase temperature can promote further fragmentation. They demonstrated that, for some pesticide classes, fragmentation is mainly due to chemical reactions that occur in the gas or liquid phase rather than thermal decomposition. In another interesting paper, the same group demonstrated applicability of TSP in the determination of 128 polar pesticides, with a wide range of structures and polarities, in aqueous environmental samples with a detection limit of 100 ppt.[24] Sensitivity and selectivity were enhanced by postcolumn addition of aqueous buffer solution and the use of external ionization media. The mass spectrometer was operating in both PI and NI modes. In all experiments discharge-assisted buffer ionization (1000 V) was used to obtain a higher signal intensity, and LC/TSP/MS/MS experiments were carried out on a triple-stage quadrupole mass spectrometer, choosing $[M + H]^+$ and $[M + NH_4]^+$ ions as precursor ions. In combination with time-scheduled SIM, this method permits the detection of all pesticides at the 10–100 ppt level.

In a further developments, Volmer et al. examined a series of nitrogen- and phosphorus-containing pesticides, polar pesticides and sulfonylureas, employing off-line and on-line solid-phase extraction (SPE) and LC/TSP/MS with time scheduled SIM.[25–28] The precision, linearity and instrumental detection limits obtained allow the determination in full-scan mode of many analytes in the low nanograms per liter range, according to European Community Water Guidelines. In some cases, confirmation of the fragmentation pathways and further structural elucidation was obtained by comparison with APCI, ES, FAB, ^{252}Cf plasma desorption and CID spectra. Sulfonylurea herbicides and their degradation products in soil were also monitored in an early work by Shalaby et al.[29] PI full-scan spectra, with the post-column addition of a solution of ammonium acetate, were generated, allowing the simultaneous determination of the compounds of interest at the 0.02 ppm level.

In another effort to determine multiresidue pesticides and their metabolites at trace levels, SPE either on Empore C_{18} extraction discs or on a conventional column packed with PRP-1 copolymer followed by LC/TSP with time-scheduled SIM was performed.[30] The TSP/LC/MS system was operating in both PI and NI filament-on modes. Ammonium formate was added to the LC mobile phase. This approach allowed the detection of the selected compounds at concentrations of 0.02–0.4 $\mu g\,L^{-1}$ in the water samples. Calibration plots, constructed using spiked estuarine water samples, show good linearity for 16 of the 34 pesticides considered. The same approach was used to determine the persistence of temephos and its metabolites in rice crop field waters[31] and to determine chlorotriazine and its photolysis products in water.[32]

Sixteen carbamate pesticides were used in a very interesting comparison, under standardized experimental conditions, among three different TSP interfaces and ion sources to evidence possible differences ion formation and sensitivity of detection.[33] Twelve of the investigated compounds showed no significant differences in their mass spectra, especially at low temperatures. Thermally labile molecules such as asulam, thiodicarb and thiofanox showed significant differences in the mass spectra, mainly due to thermally assisted hydrolysis of the parent compounds in the vaporizer probe, not necessarily reflecting differences in the ion source geometry.

6.2 Electrospray Interface

ES is one of the most widely used interfaces in pesticide analysis thanks to its high sensitivity and versatility. However, the use of this interface is limited to ionic and polar compounds. The introduction of ES has broadened the horizon in the analysis of thermolabile and polar compounds. This interface requires very low LC flow rates ($<10\,\mu L\,min^{-1}$), meaning that the conventional LC column effluent must be split. Different solutions have been devised to overcome this inconvenience, one of them being the development of a pneumatically assisted ES interface (IS) that allows flow rates up to $2\,mL\,min^{-1}$. This interface is particularly suitable for thermally labile analytes because it allows the nebulization of the column effluent at low temperature. The application of an extraction cone voltage adds the possibility of collisionally induced fragmentation, to help structural elucidation. This feature is common

to all API techniques. An interesting paper compared the performance of LC/TSP/MS and LC/IS/MS for the simultaneous determination of 10 carbamate pesticides in spiked tap water and river water and in two different types of spiked sediments which exhibit large differences in polarity.[34] The use of the TSP interface for carbamate analysis had already been tested by the same group in a previous study in which they obtained satisfactory structural information, although sensitivity was not the point of force of this interface.[35] The use of LC/IS/MS leads to detection limits between 10 and 50 pg in the SIM mode, whereas in TSP the detection limits for the same compounds are typically between 0.5 and 5 ng. Also, the repeatability in the ion signal intensities is better with the IS system. Coextracted compounds do not interfere with the analytes of interest in sediments down to the 50 μg kg^{-1} level (Figure 10a and b).

LC/IS/MS has been used in the determination of polar and/or thermally labile organophosphorus pesticides in water samples prior to SPE.[36] This method permitted the identification also of some phthalates present as interferences, and the characterization of *cis*- and *trans*-mevinphos. The limits of detection were in the range 0.01–0.20 μg L^{-1} in the SIM mode.

Henion et al. reported the quantitative and qualitative determination of eight sulfonylurea herbicides in soil.[37] The experiments were carried out with LC/IS/MS/MS in the SRM mode, providing high specificity and high sensitivity down to 0.05 ppb.

In a more recent paper, a multiresidue method for the determination of six pesticides belonging to three major pesticides classes, carbamates, triazines and ureas, in water, using an in IS LC/MS interface with a narrow-bore system, is described.[38] In order to provide structural information, CID combined with mass analysis can be obtained by increasing the orifice voltage (OR) in front of the mass analyzer. In this work, two different ORs were used: a low voltage of 20 V gives peaks for the protonated molecule $[M + H]^+$ and the sodium adduct $[M + Na]^+$, and a high voltage of 60 V generates fragment ions from these precursor ions. This approach allows increased sensitivity without the need for a tandem analyzer instrument.

The spectra obtained at 60 V indicate that fragmentation had occurred by the expulsion from the protonated or cationated molecule of small, stable, neutral fragments. The results obtained are similar to those obtained with conventional MS/MS, the only drawback of this approach being that the chemical noise is not removed from the parent and fragment ion signals. This system allows detection limits lower than those required by the European

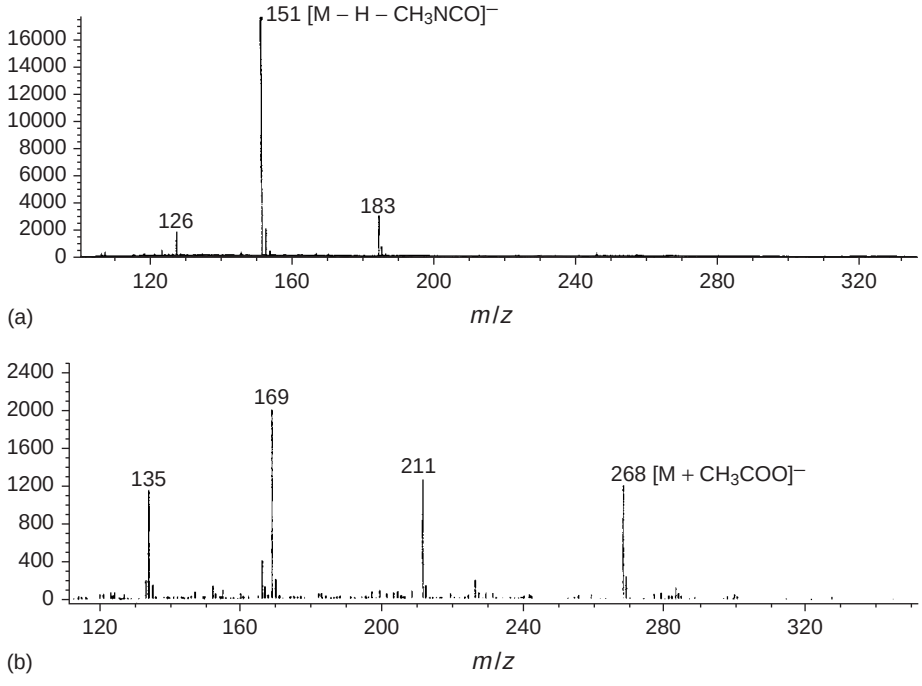

Figure 10 Filament-on NI TSP mass spectra of 0.5 μg of propoxur, with 50 : 50 acetonitrile–water, (a) without additives and (b) with 50 mM ammonium acetate as the carrier stream. (Reprinted by permission of Elsevier Science from M. Honing, D. Barceló, B.L.M. van Baar, R.T. Ghijsen, U.A.Th. Brinkmann, 'Ion Formation of *N*-Methylcarbamate Pesticides in Thermospray Mass Spectrometry: the Effects of Additives to the Liquid Chromatographic Eluent and of the Vaporizer Temperature', *J. Am. Soc. Mass Spectrom.*, **5**, 913 (1994). Copyright 1994 by American Society for Mass Spectrometry.)

Community Drinking Water Directive for the compounds analyzed.

Several other classes of pesticides were detected at ultratrace level using on-line liquid–solid and SPE LC/IS.[39,40] In the earlier study,[39] thermally labile and polar organophosphorus pesticides in water were detected at the $0.1\,\mu g\,L^{-1}$ level combining various SPE methods with IS. Detection limits were 100 times higher than those obtained with TSP, with no thermal degradation, which allows the detection of thermally labile compounds. Extraction of particularly delicate compounds was accomplished using off-line SPE followed by the MS analysis. In the later study,[40] several acidic pesticides such as benazolin, bentazone, 2,4-D, MCPA, MCPP, MCPB and 6- and 8-hydroxybentazone were detected in water samples at $0.01-0.03\,\mu g\,L^{-1}$ levels. With extraction voltages above 30 V it was possible to distinguish between the two isomeric products of bentazone, 6- and 8-hydroxybentazone. ES has been demonstrated to be a valid technique also in the determination of sulfonylurea herbicides in soil.[41] The sample was spiked at $6.0\,\mu g\,kg^{-1}$ with chlorsulfuron, metsulfuron-methyl, thifensulfuron-methyl and triasulfuron, cleaned up on a C_{18} cartridge after extraction, then analyzed using ES.

Di Corcia et al. have investigated in detail the possibility of determining several classes of pesticides in different matrices. Acidic and basic–neutral pesticides with a broad range of polarity were extracted from aqueous samples by means of graphitized carbon black (GCB) cartridges and analyzed by LC/IS.[42,43] The conventional column flow rate of $1\,mL\,min^{-1}$ was split and only $30-40\,\mu L\,min^{-1}$ were allowed into the ion source. The mass spectrometer was operating in the NI mode for the acidic compounds and a voltage of 20 V was applied to the skimmer cone, and in the PI mode with a voltage of 30 V for the basic–neutral pesticides. An ion-pair agent was added to the mobile phase in order to achieve the separation of the acidic pesticides. The ion signal was affected by the nature of the ion-pair and its counterion and also by the nature of the pesticide itself. The limits of sensitivity for the acidic pesticides were estimated to be $2-6\,ng\,L^{-1}$ in drinking water and $8-25\,ng\,L^{-1}$ in surface water, and for the basic–neutral pesticides they ranged from 1 to $9\,ng\,L^{-1}$ from total ion current (TIC) and from 0.06 to $1.5\,ng\,L^{-1}$ from extracted ion current profiles. With little difference in the procedures, Di Corcia et al. demonstrated that this approach could be applied to a wide range of pesticides. The same LC/ES system was used for the analysis of carbamate insecticides at the picograms per gram level in fruits and vegetables, operating in the PI mode with a skimmer cone voltage of 30 V and time-scheduled SIM.[16] Atrazine and its degradation products and sulfonylurea herbicides in water were also determined with the same instrumentation.[44,45]

6.3 Particle Beam Interface

The PB interface, in particular in its capillary-scale version,[17,46] is very competitive for the analysis of pesticides. Its major advantage relies on EI, which provides library-searchable, reproducible spectra of the analytes. Budde et al.[47] were among the first workers to evaluate the suitability of the PB as a major component of a general purpose, broad-spectrum analytical method for the determination of pollutants. The results were encouraging in terms of response linearity, reproducibility and detection limits for a wide selection of polar compounds. A comparison between PB and TSP in the analysis of phenoxy acid herbicides was reported in another paper.[48] TSP with NI detection showed better sensitivity, whereas the PB was more precise, particularly at the high concentration levels.

Phenoxy acid and ester herbicides were the target compounds in another method for their determination in soil and water.[49] The PB was used for the confirmation of UV detection results. Detection limits ranging between 12 and 80 ng, corresponding to 4.8–32 ppb in water and 20–133 ppb in soil, were obtained with this method. EI showed the molecular and the phenoxy (base) ions of the selected compounds for unambiguous identification.

The advantages offered by a capillary-scale interface, developed by Cappiello et al.,[50] were found particularly convenient for the analysis of a large number of basic–neutral and acidic pesticides. The improved sensitivity, especially in reversed-phase conditions, lowers the overall detection limits, allowing the low parts per billion level determination of the target analytes in water. However, the results relative to the acidic pesticides are still affected by thermal degradation and an apparent loss of chromatographic resolution when eluted from the column. The difference is noticeable when comparing the TIC profiles for basic–neutral and acidic pesticides.

A breakthrough in the analysis of these compounds is represented by the use of Teflon® as the target vaporization surface.[50] The validity of this approach is demonstrated by the analysis of 18 acidic pesticides with a Teflon®-modified ion source (Figure 11). The PI EI assay combines the advantages of a reduced mobile phase flow rate ($1\,\mu L\,min^{-1}$) with a more efficient solute vaporization surface. Under these conditions, the original chromatographic profile is reproduced more accurately, with no signs of analyte adsorption or thermal decomposition. Instrument detection limits are improved significantly (0.6–5 ng), exceeding the parts per billion limit for the determination of the pesticides in water.

Cappiello et al.[51] developed an HPLC method in which ion interaction chromatography was coupled with a capillary-scale PB, allowing the simultaneous detection of basic–neutral and acidic pesticides. This

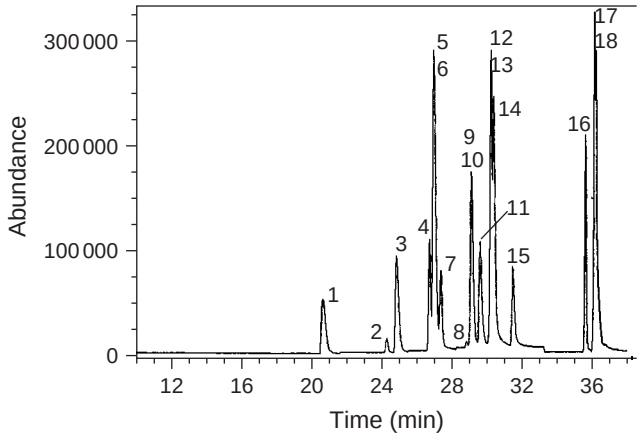

Figure 11 Reconstructed ion chromatogram for the separation of a standard solution of 18 acidic pesticides: (1) 4-nitrophenol, (2) 2,4-dinitrophenol, (3) dicamba, (4) bentazone, (5) 2,4-D, (6) MCPA, (7) bromoxynil, (8) 3,5-dichlorobenzoic acid, (9) mecoprop, (10) dichlorprop, (11) 2,4,5-T, (12) warfarin, (13) 2,4-DB, (14) MCPB, (15) 2,4,5-TP, (16) dinoseb, (17) dinoterb and (18) pentachlorophenol. (Reprinted with permission from A. Cappiello, G. Famiglini, P. Palma, A. Berloni, F. Bruner, *Environ. Sci. Technol.*, **29**(9), 2295 (1995). Copyright 1995 American Chemical Society.)

method was made possible by the ability of micro-PB to accept nonvolatile HPLC modifiers. Less than 0.1 ppb of several pesticides was detected in tap water using this procedure.

6.4 Atmospheric Pressure Chemical Ionization Interface

The applicability of APCI is similar to that offered by PB and has been demonstrated in several reports to be particularly suitable for pesticide analysis. The APCI interface was compared with other LC interfaces (TSP, IS, PB) in a study on the determination of *N*-methylcarbamates in pepper extracts.[52] TSP, working in the PI filament-on, mode provided good sensitivity, although some fluctuation of the ion signal perturbed the analyses. Further fragmentation could be induced for enhanced confirmatory capabilities. The PB interface used in this work was a conventional one and was mounted on the same mass spectrometer as utilized for TSP. In this configuration, the PB performance was not satisfactory, mainly because the transport of the analytes into the ion source was inefficient. The performance of IS was comparable to that of TSP, providing mainly protonated molecules, although more information can be obtained by varying the cone voltage. CID will allow further structure elucidation. APCI with a heated nebulizer permits both protonated molecules and fragment ions. API techniques, besides their ease of use, are complementary and readily interchangeable, allowing a wide range of compounds with different chemical properties to be analyzed (Figure 12a–g).

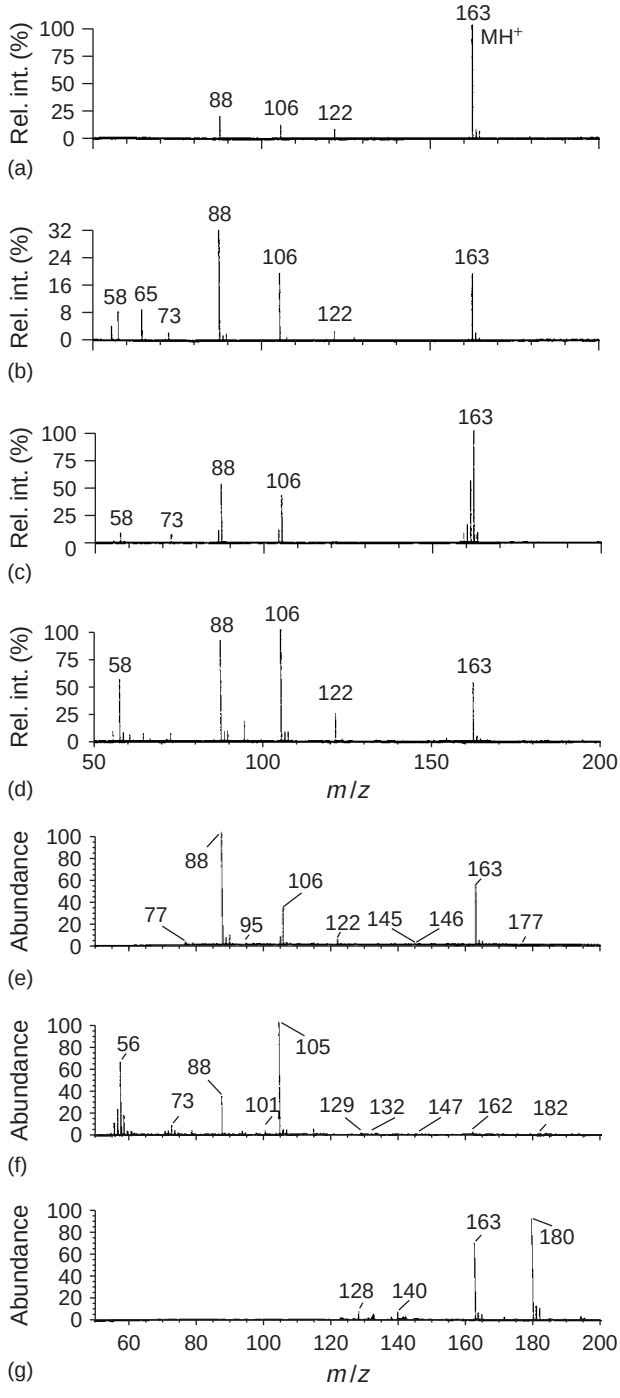

Figure 12 Background-subtracted flow injection analysis (FIA) spectra of methomyl obtained by IS/MS (250 ng) at OR (a) 50 V and (b) 70 V, (c) IS/MS/MS (250 ng), (d) APCI/MS (50 ng), (e) PB/CI (2 μg), (f) PB/EI (2 μg) and (g) TS/MS (250 ng). (Reprinted by permission of Elsevier Science from S. Pleasance, J.F. Anacleto, M.R. Bailey, D.H. North, 'An Evaluation of Atmospheric Pressure Ionization Techniques for the Analysis of *N*-Methylcarbamate Pesticides by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **3**, 378 (1992). Copyright 1992 by American Society for Mass Spectrometry.)

Ten different pesticides were determined in extracts of strawberries and plums. In order to select the most suitable LC/MS technique, an initial comparison between APCI and ES was made.[53] Spectra of the selected compounds were acquired using APCI and ES in both PI and NI modes, obtaining similar spectra, although for most compounds with ES the most significant ion was the sodium adduct ion. Limits of detection were comparable for most compounds with the two techniques, but APCI was more sensitive for aldicarb and its metabolites. For this reason and for greater flexibility with LC flow rates, APCI was the interface chosen. This method allowed detection limits as low as $0.02\,mg\,kg^{-1}$ to be achieved for most compounds, meeting Codex Alimentarius Commission requirements.

APCI/MS (APCI/MS/MS) coupled with single, short LC columns was successfully used for the determination of several pesticides and herbicides in tap and river water.[54,55] The short column allowed fast separation of the analytes, increasing sample throughput. The protonated molecules generated by the APCI interface were subjected to CID for structure elucidation and for the confirmation of target and unknown compounds. This instrumentation has been used to obtain CID spectral data in the PI mode with the purpose of building a spectral library. The same approach was used to analyze 17 different pesticides belonging to six different classes in aqueous samples. A comparison between APCI and IS was also performed. The two interfaces were operated under the same conditions, with no volatile acids added to the mobile phase. The two interfaces performed similarly in terms of limits of detection (0.1–50 ng in full-scan mode for 12 of the 17 pesticides. Two of the investigated compounds, fenchlorphos and bromophos-ethyl, were detected by neither interface. In order to facilitate interpretation of the spectra, an MS/MS library of the product-ion spectra of protonated molecules was produced. The recording of the spectra of 73 pesticides and their metabolites was accomplished first in the PI full-scan mode and then, with $[M + H]^+$ ions selected as precursor ions, in the MS/MS product-ion mode.

An automated on-line SPE/LC/APCI system was used for the analysis of organophosphorus pesticides in groundwater[56] and triazine and phenylurea herbicides in miscellaneous environmental matrices.[57]

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
ACG	Automatic Gain Controller
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
CI	Chemical Ionization
CID	Collision-induced Dissociation
DC	Direct Current
DLI	Direct Liquid Introduction
EI	Electron Ionization
ES	Electrospray
FAB	Fast Atom Bombardment
FIA	Flow Injection Analysis
GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
IS	Ionspray
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
MAGIC	Monodisperse Aerosol Generator
MS	Mass Spectrometry
MS/MS	Mass Spectrometry/Mass Spectrometry
NI	Negative Ion
OR	Orifice Voltage
PB	Particle Beam
PI	Positive Ion
RF	Radiofrequency
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SRM	Selected Reaction Monitoring
TIC	Total Ion Current
TOF	Time-of-flight
TSP	Thermospray
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VSI	Volatile Salt Ionization

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • Multiclass, Multiresidue Analysis

of Pesticides, Strategies for • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Gradient Elution Chromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. A. Cappiello, G. Famiglini, F. Bruner, 'Determination of Acidic and Base–Neutral Pesticides in Water with a New Micro Flow Rate LC/MS Particle Beam Interface', *Anal. Chem.*, **66**, 1416–1423 (1994).
2. M. Baldwin, F. McLafferty, 'Liquid Chromatography–Mass Spectrometry Interface. I. Direct Introduction of Liquid solutions into a Chemical Ionization Mass Spectrometer', *Org. Mass Spectrom.*, **7**, 1111–1112 (1973).
3. W.H. McFadden, H.L. Schwartz, S. Evans, 'Direct Analysis of HPLC Effluent', *J. Chromatogr.*, **122**, 389–396 (1976).
4. R.C. Willoughby, R.F. Browner, 'Monodisperse Aerosol Generation Interface for Combining Liquid Chromatography with Mass Spectrometry', *Anal. Chem.*, **56**, 2625–2632 (1984).
5. R.W. Reiser, A.C. Barefoot, R.F. Dietrich, A.J. Fogiel, W.R. Jonhson, M.T. Scott, 'Application of Microcolumn Liquid Chromatography–Continuous-flow Fast Atom Bombardment Mass Spectrometry in Environmental Studies of Sulfonylurea Herbicides', *J. Chromatogr.*, **554**, 91–101 (1991).
6. W.M.A. Niessen, A.P. Tinke, 'Liquid-chromatography–Mass Spectrometry. General Principles and Instrumentation', *J. Chromatogr.*, **703**, 37–57 (1995).
7. C.R. Blakely, J.J. Carmody, M.L. Vestal, 'A New Liquid Chromatography/Mass Spectrometry Interface Using Crossed-beam Techniques', *Adv. Mass Spectrom.*, **8B**, 1616–1623 (1980).
8. M. Dole, L.L. Mack, R.I. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice, 'Molecular Beams of Microions', *J. Chem. Phys.*, **49**, 2240–2249 (1968).
9. M. Yamashita, J.B. Fenn, 'Electrospray Ion-source. Another Variation on the Free-jet Theme', *J. Phys. Chem.*, **88**, 4451–4459 (1984).
10. K.M. Downard, K. Biemann, 'Charging Behavior of Highly Basic Peptides During ESI/MS', *Int. J. Mass Spectrom. Ion Processes*, **148**, 191–202 (1995).
11. J.B. Fenn, 'Ion Formation from Charged Droplets: Roles of Geometry, Energy and Time', *J. Am. Soc. Mass Spectrom.*, **4**, 524–535 (1993).
12. R.D. Voyksner, T. Pack, 'Investigation of Collisional-activation Decomposition Process and Spectra in the Transport Region of an Electrospray Single-quadrupole Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **5**, 263 (1991).
13. A. Di Corcia, C. Crescenzi, A. Laganà, E. Sebastiani, 'Evaluation of a Method Based on Liquid Chromatography/Electrospray/Mass Spectrometry for Analyzing Carbamate Insecticides in Fruits and Vegetables', *J. Agric. Food Chem.*, **44**, 1930–1938 (1996).
14. E.C. Horning, D.I. Carroll, I. Dzidic, K.D. Haegele, M.G. Horning, R.N. Stillwell, 'Atmospheric Pressure Ionization (API) Mass Spectrometry. Solvent Mediated Ionization of Sample Introduced in Solution and in a Liquid Chromatograph Effluent System', *J. Chromatogr. Sci.*, **12**, 725–729 (1974).
15. R.C. Willoughby, R.F. Browner, 'Monodisperse Aerosol Generation Interface for Coupling Liquid Chromatography with Mass Spectroscopy', *Anal. Chem.*, **56**, 2626–2631 (1984).
16. A. Cappiello, 'Is Particle Beam an Up-to-date Interface? State of the Art and Perspective', *Mass Spectrom. Rev.*, **15**, 283–296 (1996).
17. A. Cappiello, F. Bruner, 'Micro-flow Rate Particle Beam Interface for Capillary Liquid Chromatography–Mass Spectrometry', *Anal. Chem.*, **65**, 1281–1287 (1993).
18. M.H. Lamoree, R.T. Ghijsen, U.A.Th. Brinkman, 'LC/MS Interfaces Systems in Environmental Analysis: Application to Polar Pesticides', in *Environmental Analysis*, ed. D. Barceló, Elsevier, Amsterdam, Chapter 15, 1993.

19. D. Barceló, 'A Review of Liquid Chromatography in Environmental Pesticide Analysis', *Chromatographia*, **25**, 928–936 (1988).
20. I. Ferrer, D. Barceló, 'LC/MS Methods for Trace Determination of Pesticides in Environmental Samples', *Analusis*, **26**, 118–122 (1998).
21. M.P. Balogh, 'Commercialization of LC/MS 1987–1997: A Successful Decade in Review', *LC/GC Int.*, 728 (1997).
22. T.A. Bellar, W.L. Budde, 'Determination of Nonvolatile Organic Compounds in Aqueous Environmental Samples Using Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **60**, 2076–2083 (1988).
23. D. Volmer, A. Preiss, K. Levsen, G. Wünsch, 'Thermospray Mass Spectral Studies of Pesticides. Temperature and Salt Concentration Effects on the Ion Abundances in Thermospray Mass Spectra', *J. Chromatogr.*, **647**, 235–259 (1993).
24. D. Volmer, K. Levsen, G. Wünsch, 'Thermospray Liquid Chromatographic–Mass Spectrometric Multiresidue Determination of 128 Polar Pesticides in Aqueous Environmental Samples', *J. Chromatogr.*, **660**, 231–248 (1994).
25. D. Volmer, K. Levsen, 'Mass Spectrometric Analysis of Nitrogen- and Phosphorus-containing Pesticides by Liquid Chromatography–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**, 655–675 (1994).
26. D. Volmer, K. Levsen, W. Engewald, 'Analysis of Polar Pesticides in Aqueous Samples by Combined On-line Trace Enrichment and Thermospray Liquid Chromatography–Mass Spectrometry', *Wasser*, **82**, 335–364 (1994).
27. S. Sennert, D. Volmer, K. Levsen, G. Wünsch, 'Multiresidue Analysis of Polar Pesticides in Surface and Drinking Water by On-line Enrichment and Thermospray LC/MS', *Fresenius' J. Anal. Chem.*, **351**, 642–649 (1995).
28. D. Volmer, J.G. Wilkes, K. Levsen, 'Liquid Chromatography/Mass Spectrometry Multiresidue Determination of Sulfonylureas After On-line Trace Enrichment', *Rapid Commun. Mass Spectrom.*, **9**, 767–771 (1995).
29. L. Shalaby, F.Q. Bramble, Jr, P.W. Lee, 'Application of Thermospray LC/MS for Residue Analysis of Sulfonylurea Herbicides and Their Degradation Products', *J. Agric. Food Chem.*, **40**, 513–517 (1992).
30. S. Chiron, S. Dupas, P. Scribe, D. Barceló, 'Application of On-line Solid Phase Extraction Followed by Liquid Chromatography–Thermospray Mass Spectrometry to the Determination of Pesticides in Environmental Waters', *J. Chromatogr.*, **665**, 295–305 (1994).
31. S. Lacorte, N. Ehresmann, D. Barceló, 'Persistence of Temephos and Its Transformation Products in Rice Crop Field Waters', *Environ. Sci. Technol.*, **30**, 917–923 (1996).
32. G. Durand, D. Barceló, 'Determination of Chlorotriazines and Their Photolysis Products by Liquid Chromatography with Photodiode-array and Thermospray Mass Spectrometric Detection', *J. Chromatogr.*, **502**, 275–286 (1990).
33. D. Volmer, K. Levsen, M. Honing, D. Barceló, J. Abian, E. Gelpí, B. van Baar, U. Brinkman, 'Comparative Study of Different Thermospray Interfaces with Carbamate Pesticides: Influence of the Ion Source Geometry', *J. Am. Soc. Mass Spectrom.*, **6**, 656–667 (1995).
34. M. Honing, J. Riu, D. Barceló, B.L.M. van Baar, U.A.Th. Brinkman, 'Determination of Ten Carbamate Pesticides in Aquatic and Sediment Samples by Liquid Chromatography–Ionspray and Thermospray Mass Spectrometry', *J. Chromatogr.*, **733**, 283–294 (1996).
35. M. Honing, D. Barceló, B.L.M. van Baar, R.T. Ghijsen, U.A.Th. Brinkmann, 'Ion Formation of *N*-Methylcarbamate Pesticides in Thermospray Mass Spectrometry: The Effects of Additives to the Liquid Chromatographic Eluent and of the Vaporizer Temperature', *J. Am. Soc. Mass Spectrom.*, **5**, 913–927 (1994).
36. C. Molina, P. Grasso, E. Benfenati, D. Barceló, 'Automated Sample Preparation with Extraction Columns Followed by Liquid Chromatography–Ion Spray Mass Spectrometry. Interferences, Determination and Degradation of Polar Organophosphorus Pesticides in Water Samples', *J. Chromatogr.*, **737**, 47 (1996).
37. L.Y.T. Li, D.A. Campbell, P.K. Bennett, J. Henion, 'Acceptance Criteria for Ultratrace HPLC–Tandem Mass Spectrometry: Quantitative and Qualitative Determination of Sulfonylurea Herbicides in Soil', *Anal. Chem.*, **68**, 3397–3404 (1996).
38. D. Giraud, A. Ventura, V. Camel, A. Bermond, P. Arpino, 'Determination of Traces of Pesticides in Water by Solid-phase Extraction and Liquid Chromatography–Ionspray Mass Spectrometry', *J. Chromatogr.*, **777**, 115–125 (1997).
39. C. Molina, M. Honing, D. Barceló, 'Determination of Organophosphorus Pesticides in Water by Solid-phase Extraction Followed by Liquid Chromatography/High-flow Pneumatically Assisted Electrospray Mass Spectrometry', *Anal. Chem.*, **66**, 4444–4449 (1994).
40. S. Chiron, S. Papilloud, W. Haerdi, D. Barceló, 'Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography–High-flow Pneumatically Assisted Electrospray Mass Spectrometry for the Determination of Acidic Herbicides in Environmental Waters', *Anal. Chem.*, **67**, 1637–1643 (1995).
41. L.J. Marek, W.C. Koskinen, 'LC/MS Analysis of 4-Methoxy-6-methyl-1,3,5-triazin-2-yl-containing Sulfonylurea Herbicides in Soil', *J. Agric. Food Chem.*, **44**, 3878–3881 (1996).
42. C. Crescenzi, A. Di Corcia, S. Marchese, R. Marchese, 'Determination of Acidic Pesticides in Water by a Benchtop Electrospray Liquid Chromatography Mass Spectrometer', *Anal. Chem.*, **67**, 1968–1975 (1995).
43. C. Crescenzi, A. Di Corcia, E. Guerriero, R. Samperi, 'Development of a Multiresidue Method for Analyzing Pesticide Traces in Water Based on Solid-phase Extraction and Electrospray Liquid Chromatography Mass Spectrometry', *Environ. Sci. Technol.*, **31**, 479–488 (1997).

44. A. Di Corcia, C. Crescenzi, E. Guerriero, R. Samperi, 'Ultratrace Determination of Atrazine and Its Six Major Degradation Products in Water by Solid-phase Extraction and Liquid Chromatography–Electrospray/Mass Spectrometry', *Environ. Sci. Technol.*, **31**, 1658–1663 (1997).
45. A. Di Corcia, C. Crescenzi, R. Samperi, L. Scappaticcio, 'Trace Analysis of Sulfonylurea Herbicides in Water: Extraction and Purification by a Carbograph 4 Cartridge, Followed by Liquid Chromatography with UV Detection and Confirmatory Analysis by an Electrospray/Mass Detector', *Anal. Chem.*, **69**, 2819–2826 (1997).
46. A. Cappiello, G. Famiglini, 'Evaluation of the Performance of a Micro-flow Rate LC/MS Particle Beam Interface', *Anal. Chem.*, **66**, 3970–3976 (1994).
47. T.D. Behymer, T.A. Bellar, W.L. Budde, 'Liquid Chromatography/Particle Beam/Mass Spectrometry of Polar Compounds of Environmental Interest', *Anal. Chem.*, **62**, 1686–1690 (1990).
48. T.L. Jones, L.D. Betowski, B. Lesnik, T.C. Chiang, J.E. Teberg, 'Interlaboratory Comparison of Thermospray and Particle Beam Liquid Chromatography/Mass Spectrometry Interfaces: Evaluation of a Chlorinated Phenoxy Acid Herbicide Liquid Chromatography/Mass Spectrometry Analysis Method', *Environ. Sci. Technol.*, **25**, 1880–1884 (1991).
49. I.S. Kim, F.I. Sasinos, R.D. Stephens, J. Wang, M.A. Brown, 'Determination of Chlorinated Phenoxy Acid and Ester Herbicides in Soil and Water by Liquid Chromatography Particle Beam Mass Spectrometry and Ultraviolet Absorption Spectrophotometry', *Anal. Chem.*, **63**, 819–823 (1991).
50. A. Cappiello, G. Famiglini, 'Analysis of Thermally Unstable Compounds by a LC/MS Particle Beam Interface with a Modified Ion Source', *Anal. Chem.*, **67**, 412–419 (1995).
51. A. Cappiello, G. Famiglini, F. Mangani, S. Angelino, M.C. Gennaro, 'Simultaneous Determination of Acidic and Base–Neutral Pesticides by IIR–Micro HPLC/MS Particle Beam Interface', presented at the 45th ASMS Conference on Mass Spectrometry and Allied Topics, Palm Springs, CA, June 1–5, 1997.
52. S. Pleasance, J.F. Anacleto, M.R. Bailey, D.H. North, 'An Evaluation of Atmospheric Pressure Ionization Techniques for the Analysis of *N*-Methylcarbamate Pesticides by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **3**, 378–397 (1992).
53. K.A. Barnes, R.J. Fussel, J.R. Startin, M.K. Pegg, S.A. Thorpe, S.L. Reynolds, 'High Performance Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry with Ionization Polarity Switching for the Determination of Selected Pesticides', *Rapid Commun. Mass Spectrom.*, **11**, 117–123 (1997).
54. A.C. Hogenboom, J. Slobodník, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Single Short-column Liquid Chromatography with Atmospheric Pressure Chemical Ionization–(Tandem) Mass Spectrometric Detection for Trace Environmental Analysis', *Chromatographia*, **42**(9/10), 506–514 (1996).
55. A.C. Hogenboom, P. Speksnijder, R.J. Vreeken, W.M.A. Niessen, U.A.Th. Brinkman, 'Rapid Target Analysis of Microcontaminants in Water by On-line Single-short-column Liquid Chromatography Combined with Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry', *J. Chromatogr.*, **777**, 81–89 (1997).
56. S. Lacorte, D. Barceló, 'Determination of Parts per Trillion Levels of Organophosphorus Pesticides in Groundwater by Automated On-line Liquid–solid Extraction Followed by Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry Using Positive and Negative Ion Modes of Operation', *Anal. Chem.*, **68**, 2464–2470 (1996).
57. I. Ferrer, M.C. Hennion, D. Barceló, 'Immunosorbents Coupled On-line with Liquid Chromatography/Atmospheric Pressure Chemical Ionization/Mass Spectrometry for the Part per Trillion Level Determination of Pesticides in Sediments and Natural Waters Using Low Preconcentration Volumes', *Anal. Chem.*, **69**, 4508–4514 (1997).

Immunochemical Assays in Pesticide Analysis

Andrea Dankwardt
Sension GmbH, Augsburg, Germany

1 Introduction **1**
2 Antibodies **2**
2.1 Antibody Structure 2
2.2 Antibody Production 3
2.3 Immunogens 3
3 Immunoassay Types **5**
3.1 Assay Formats 5
3.2 Labels 5
3.3 Solid Phases 6
4 Properties of Competitive Immunoassays **6**
4.1 Dose–Response Curve 6
4.2 Quality Control 7
4.3 Cross-reactivity 8
4.4 Sample Preparation and Matrix Effects 8
5 Application to Environmental Samples **9**
5.1 Water 9
5.2 Soil 13
5.3 Food 13
5.4 Biomonitoring 14
6 New Developments **14**
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Immunochemical assays (immunoassays, IAs) are biochemical assays which work according to the law of mass action. They are based on the recognition of an antigen (Ag) or a hapten by antibodies (Abs). Abs are serum glycoproteins of the immunoglobulin (Ig) class and are produced by the vertebrate immune system against foreign material of high molecular mass. The result of the binding reaction between the Ab and an analyte is usually made visible by means of enzymatic, chemiluminescent, fluorescent or radioactive markers. According to the label used IAs can be classified into enzyme immunoassays (EIAs), radioimmunoassays (RIAs), fluorescence immunoassays (FIAs) or chemiluminescent immunoassays (CLIAs). The measuring range of most IAs for pesticides is in the parts per trillion to lower parts per billion range. A lot of samples can be analyzed within a short time, while only low sample volumes are necessary. In many cases (water, some liquid food samples) no extraction step and no cleanup are necessary. Not all assays are completely specific to one single compound. Cross-reactivities of the Abs with haptens similar to the analyte can be observed. In some cases, matrix effects may occur, especially with soil or colored food extracts. Therefore, validation of the assays for the matrix of interest should be carried out. As IAs are usually targeted at a single analyte or a group of analytes, multianalyte approaches using Ab arrays or a combination of immunochemical techniques with liquid chromatography (LC) are pursued.

1 INTRODUCTION

Interest in immunochemical assays for the determination of pesticides has been steadily increasing. IAs are now commonly applied for the analysis of contaminants in water, soil, food and body fluids.[1–9] The first immunological experiments had already been carried out as early as the late eighteenth century when Edward Jenner, an English physician, used cowpox to prevent infection with smallpox. Based on these studies Louis Pasteur developed the use of attenuated strains of microorganisms for successful vaccinations. Emile Roux and Alexandre Yersin then found that immunity is caused by soluble compounds of microorganisms, which they called toxins. These toxins induce specific compounds in the immunized animal, which were named ''antitoxins'' by Emil von Behring and Shibasaburo Kitasato (1890) and are now called Abs. The Ab ''generating'' compounds are known as Ags.

Around the turn of the century it was shown that Abs are not only produced against microorganisms and their toxins, but also by other substances such as milk, protein or plant-derived toxins. Paul Ehrlich was the first to carry out quantitative studies on Ag–Ab interactions. The great interest in this field led to the first book on immunochemistry, published by Svante Arrhenius in 1907.[10] Karl Landsteiner also belongs to the pioneers in immunochemistry. He systematically used small artificial molecules, which he called haptens, coupled to a carrier molecule for immunization. In 1923 Heidelberger and co-workers found polysaccharides to be antigenic as well.

Studies by Rodney Porter (1959) and Gerald Edelman (1961) have provided the chemical structure of the Ab molecule. The enormous variety of Abs was explained by Frank McFarlane Burnet in 1957, based on a hypothesis of Niels Jerne dating from 1955, the now widely accepted clonal-selection theory. It describes each Ab-producing

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

cell as carrying on its surface only one type of Ab as a receptor. The binding of a respective Ag to this receptor leads to a clonal expansion of this cell and to the maturation of Ab-producing cells.

Immunochemical methods have their origin in the medical field. The first IA, a RIA for the quantification of insulin in serum, was described by Yalow and Berson.[11] Later, radiolabels were replaced by enzymes in EIAs by Engvall and Perlmann[12] and Van Weeman and Schuurs.[13] Since then radiolabels have obtained broad application in medical diagnostics and environmental analysis.

IAs belong to the most common methodology in the field of immunoanalysis. Even though Abs are (still) produced by a biological process, IAs are nevertheless chemical analytical procedures. The basic principle applying to all immunoreactions is based upon the law of mass action. In the equilibrium reaction between a free Ag or a hapten, such as a pesticide, and the Ab forming the hapten–Ab complex HAb (=bound hapten), represented by Equation (1),

$$H + Ab = HAb \quad (1)$$

the affinity constant K determines the concentration ratio between the bound hapten and the free reaction partners, Equation (2):

$$K = \frac{[HAb]}{[H][Ab]} mol^{-1} \quad (2)$$

A low detection limit (DL) in an IA therefore requires a high affinity of the Ab toward the analyte, which is expressed by a high affinity constant. For further details refer to Hock et al.[14]

IAs are based upon the measurement of Ab binding-site occupancy by the analyte (Figure 1). This reflects the analyte concentration in the sample. Since the binding reaction does not produce a signal which can be detected by simple means, various markers, e.g. radioactivity, enzymes, or fluorescence, are employed for the detection of the immunoreaction (see section 3.2). However, more sophisticated techniques like some immunosensing methods do not rely on a label (see section 6).

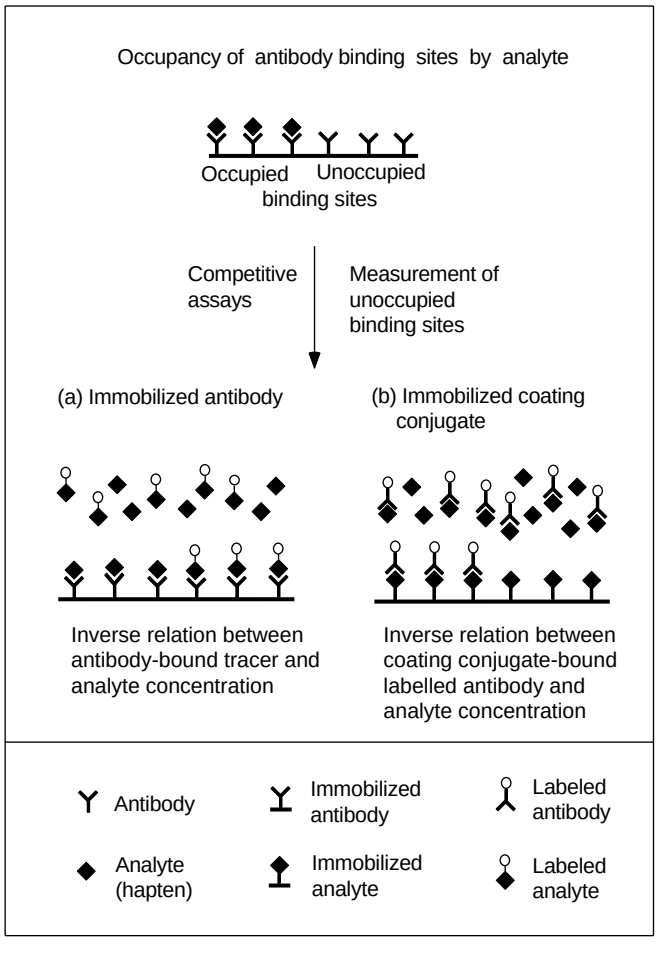

Figure 1 Principle of the competitive IA. In the first format with immobilized Ab (a) the plates are coated with Ab. Analyte and enzyme-labeled analyte compete for the Ab binding sites. In the second format a hapten–protein conjugate is immobilized in the solid phase (b). This protein conjugate and the free analyte compete for the binding sites of the Ab in solution. (Reproduced from Dankwardt and Hock[7] with permission from *Food Technology and Biotechnology*.)

2 ANTIBODIES

2.1 Antibody Structure

Immunochemical analysis is based upon the specific reaction between an Ab and its corresponding Ag or hapten. Abs are part of the vertebrate defense system (for more details refer to immunology textbooks, for example Golub[15] and Roitt[16]). They are serum glycoproteins of the Ig class produced by the immune system against foreign material such as pathogens or xenobiotics, and bind the target substance with high selectivity and affinity. Although there are five distinct classes of Ab in most higher mammals (IgA, IgD, IgE, IgG, IgM) IgG makes up approximately 80% of the total Ig in human serum. Most IAs rely upon IgG as the major Ig.

The basic structure of an Ab molecule is shown in Figure 2. It consists of two identical heavy (H) chains and two identical light (L) chains stabilized and linked by inter- and intrachain disulfide bonds. The H- and L-chains are organized into variable and constant regions. The Ag binding site (combining site) is formed by the association of parts of the variable regions of the H- and L-chains, located at the amino terminal end. The variable regions of both chains are organized into three

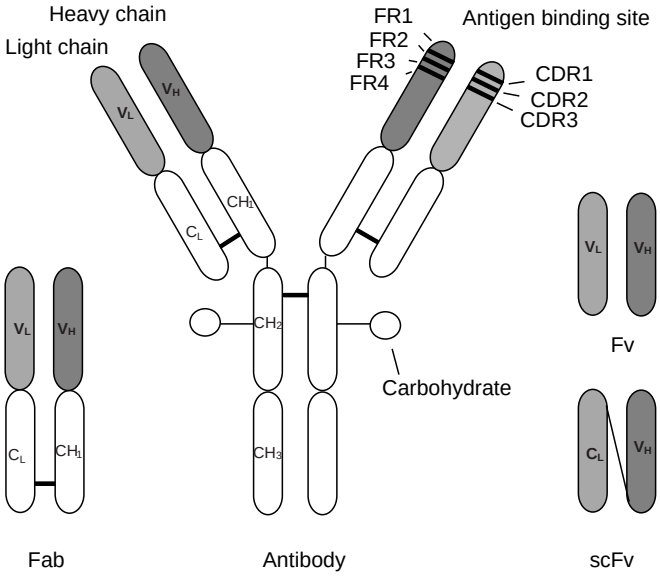

Figure 2 Structure of IgG Abs and their fragments (modified after Hock et al.[17]). scFv is the recombinant antibody fragment, a single chain fragment containing only the variable region, FR is the frame region, V_L is the variable region of light chain, V_H is the variable region of heavy chain, C_L is the constant region on the light chain, CH_1, CH_2, CH_3 is a constant region on the heavy chain.

hypervariable or complementarity determining regions (CDRs) separated by four framework regions. The greatest amino acid sequence variation occurs within the CDRs whereas the framework regions are more conserved. It is assumed that the association of the CDR regions forms the combining site. The lower part of the molecule, the Fc (antibody fragment containing the crystallizable fragment) is responsible for some important biological effector functions such as complement fixation and is not necessary for Ag or hapten binding. It contains the last heavy chain domains. The whole of the Ig molecule or Ab fragments, $F(ab)_2$ and Fab (antibody fragments containing the antigen binding site(s)) can be used in IAs.

A substance that after injection into the body of a vertebrate induces a specific Ab synthesis, is called an Ag. Ags are principally macromolecules, for instance proteins, polysaccharides or nucleic acids. Synthetic polymers also belong to the antigens, i.e. they can be used as, or act as, Ags. Small molecules (haptens) such as pesticides have to be coupled to a macromolecular carrier to elicit an Ab response (see section 2.3). The ability of an Ab molecule to bind an Ag or a hapten specifically is controlled by structural and chemical interactions between the ligand and the Ab at the combining site. The Ag–Ab interaction is reversible and does not involve formation of covalent bonds.[16] The binding is a result of a variety of interactions such as hydrophobic, ionic, H-bonding, π-π electron interaction, and van der Waals forces. The binding energy (relative affinity of the Ab) increases with the number of specific chemical interactions between the analyte and the amino acid residues in the Ab combining site. Therefore, the selectivity and sensitivity of an IA is controlled by the nature of the Ag–Ab binding process.

2.2 Antibody Production

Ab production is conveniently carried out in warm blooded animals, e.g. rabbits, sheep, mice or chickens.[17] Polyclonal antibodies (pAbs) are obtained from the serum and comprise a mixture of different Ab populations. Monoclonal antibodies (mAbs) consist of a single monospecific Ab population. These Abs are produced in cell culture by a single hybridoma cell derived from the fusion of B-lymphocytes with myeloma cells.[18] The hybridoma cells can then be propagated almost indefinitely in culture and will continue to produce the Ab of the lymphocyte parent. Since an individual lymphocyte produces only a single Ab type, all of the Ab molecules produced by a hybridoma cell line derived from a single hybrid cell are identical and have the same binding properties. Therefore, the hybridoma technology guarantees the unlimited production of mAbs with constant characteristics.[19] Owing to the great effort involved in mAb production many IAs still employ pAb. A third possibility for creating Abs has emerged, recombinant antibody (rAb) techniques. Here, Ig genes can be cloned, introduced and expressed in inexpensive and relatively simple host systems.[20,21] Although several nonmammalian host systems (yeast, plant and insect cells) have been used to produce rAbs, the most common vehicle is *Escherichia coli.*[22,23] The main properties of pAbs, mAbs and rAbs are listed in Table 1.

2.3 Immunogens

Most pesticides are of low molecular mass and therefore are not ordinarily antigenic. They have to be coupled to a carrier molecule, usually a protein, in order to induce an Ab response in the vertebrate immune system.[14] The site of coupling to the carrier, the coupling procedure as well as the number of haptens bound to one carrier molecule can be of major importance for the sensitivity and the selectivity of the resulting Ab (for reviews refer to Erlanger,[24,25] Goodrow et al.[26] and Szurdoki et al.[27]).

The protein carriers used in various laboratories include globulin fractions, serum albumins of different species, hemocyanin, ovalbumin, thyroglobulin, and fibrinogen. Also nonproteinaceous carriers have been used such as liposomes or dextran.[28,29] Keyhole limpet hemocyanin (KLH), a protein from mollusks, is often viewed

Table 1 Properties of polyclonal, monoclonal and recombinant Abs

Properties	pAb (Ab from serum)	mAb (Ab from hybridoma cells)	rAb (Ab produced by gene technology)
Supply	Limited and variable	Unlimited production possible	Unlimited production possible, immunization not mandatory
Uniformity	Changing properties with different sera and bleedings	Constant properties of a mAb	Constant properties of a rAb, can be changed by genetic manipulations
Affinity	Mixture of Ab with different affinities, affinity often higher with pAb	Uniformly high or low, can be selected by testing	Uniformly high or low, can be selected by testing and can be modified
Cross-reactivity	Results from different selectivities and low affinity interactions	Different, dependent upon the individual Ab	Different, dependent upon the individual Ab, can be modified
Classes and subclasses	Typical spectrum	One defined isotype	Different, depending on molecular design
Demands on Ag	High purity required for specific antisera	Impure Ags or mixture of Ags can be used for immunization, pure Ags necessary for screening	Impure Ags or mixture of Ags can be used for immunization, pure Ags necessary for screening, immunization not mandatory
Costs	Low	High	Once established, low

as a superior carrier because it is foreign to the vertebrate immune system.[30]

Another important issue concerns the optimal number of haptens bound to the carrier protein (i.e. optimal epitope density). Highly substituted carriers usually lead to the best results. For bovine serum albumin (BSA) molar ratios of 10 : 1 to 20 : 1 (hapten : carrier) are desirable; for larger molecules such as hemocyanin, ratios of 800 : 1 to 1000 : 1 should be obtained.[17] However, very high ratios may reduce immunogenicity because of either the changes in tertiary structure of the protein caused by masking of the essential free amino groups or the removal of critical determinant sites on the carrier by haptenic blocking.

For the production of pAbs the purity of the hapten used to prepare an immunoconjugate should be as high as possible. After synthesis of haptenic substances, closely related substances may be present in small quantities, leading to the production of nonspecific Abs and, consequently, to unwanted cross-reactivities of the antisera. This is not a problem with mAb, because a single cell line producing only one kind of Ab with the desired properties can be selected.

Another important consideration is the point of attachment on the hapten. Ab specificity is directed primarily at the part of the hapten molecule farthest away from the functional group that is linked to the protein carrier.[25] Even better specificity can be obtained with conjugates in which the hapten is coupled to the carrier via a spacer, thereby giving much better exposure of the hapten on the surface of the carrier. The spacer should be attached as far as possible from the unique determinant structures.[26] Usually C_3–C_6 spacers are used; if the spacer is too long, it may bend back to the carrier and the hapten will not be properly exposed. Strategies for IA hapten design for the triazine, arylurea and chloroacetanilide herbicides have been summarized by Goodrow et al.[26]

The functional groups of the hapten govern the selection of the method to be used to conjugate the hapten to the functional groups of the carrier. The functional groups of the protein carrier available for attachment of the haptens are the carboxyl group of the C terminal and of the aspartic and glutamic acid residues, the amino group of the N terminal and the lysine residues, the imidazol and phenolic functions of the histidine and tyrosine residues, respectively, and the sulfhydryl group of cysteine residues. General procedures for the preparation of conjugates can be found in Erlanger.[24,25]

After coupling, characterization of the conjugates can be carried out (see Erlanger[25]). Generally, the haptenic groups have an absorbance spectrum that can be differentiated from the protein carrier. Elemental analysis for the chlorine content can be carried out for some triazine conjugates. A more direct procedure is the incorporation of some radioactive hapten in the conjugation procedure. Another approach is quantitating the change in free amino groups as a result of conjugation. A recently applied technique is the determination of hapten density by matrix-assisted ultraviolet laser desorption/ionization mass spectrometry (MALDIMS)[31] and electrospray ionization mass spectrometry (ESIMS).[3] Application of energy-minimized molecular modeling methods to hapten design will help to choose the best derivatives and conjugation methods for successful Ab production.[32]

3 IMMUNOASSAY TYPES

3.1 Assay Formats

For low-molecular-mass analytes (haptens) such as pesticides in solution, competitive tests have to be employed, using limiting Ab concentrations. The tests can be performed as homogeneous assays without separation of the reactants,[33] but more common are heterogeneous tests where unreacted reagents are removed before evaluation. Two different formats are available, (1) with immobilized Ab (Figure 1a) and (2) with immobilized coating conjugate (Figure 1b). In variant (1) analyte and a labeled analyte (tracer) compete for the free Ab binding sites. After removal of unbound reactants the bound tracer yields a signal that is inversely proportional to the analyte concentration. The variant (2) employs an immobilized hapten-carrier conjugate on the solid phase to which analyte and Ab are added. The Ab binds to the free analyte or to the immobilized hapten according to the concentration of the reactants. If a labeled Ab is used, the amount of Ab bound to the solid phase can be directly determined after a washing step. Alternatively, a secondary labeled Ab may be used to detect the Ab which has bound to the solid phase. The signal is inversely proportional to the amount of free analyte in the sample. Very sensitive competitive IAs have been developed with DLs between 1 and 50 ng L^{-1}, for example for the triazines and urea herbicides.[34–36]

An example for a homogeneous assay system is the polarization fluoroimmunoassay (PFIA). PFIA measures the increased polarization of fluorescence when a fluorophore-labeled hapten (tracer) is bound by a specific Ab, and the decreased signal when free analyte competes with the tracer for binding.[37] While these assays are easy to carry out and very suitable for automation, they usually show a lower sensitivity than EIAs, e.g. for simazine a DL of 5 μg L^{-1} was observed.[37]

Noncompetitive assays can only be applied for high-molecular-mass analytes with more than one antigenic determinant (i.e. Ag) or low-molecular-mass analytes (haptens) bound to a solid phase, exposing the antigenic determinant. They work with an Ab excess. Noncompetitive IAs have been employed for the detection of soil-bound pesticides.[38,39] In this case the soil particles, to which the pesticide residues have bound, form the solid phase, and the residues can be detected by a labeled Ab specific to the analyte.

3.2 Labels

Depending on the label, IAs are classified in different groups. Radioisotopes are used in RIAs, enzymes in enzyme-linked immunosorbent assays (ELISAs) or EIAs, fluorophores in FIAs or PFIAs and chemiluminescent compounds in CLIAs. Additional types of IA exist, but are not very common in pesticide analysis. A more detailed description of these IAs can be found in Gosling.[40]

EIAs are most commonly used in pesticide analysis as they avoid the necessity of working with radioactive material and low DLs can be reached. Simple and cheap photometers which give an extremely rapid measurement capability and long-lasting stability of the colored product after the reaction has stopped make EIA superior to fluorimetry or luminometry, even though with these methods lower DLs may be reached.[33] Enzymes commonly used as labels in heterogeneous EIA are listed in Table 2.

Table 2 Enzyme systems commonly used for EIAs

Enzyme	Source	Molecular weight	pH optimum	Colorimetric substrates	Fluorometric substrates	Luminometric substrates
Alkaline phosphatase	Calf intestine	100 000	9–10	*p*-Nitrophenyl-phosphate	4-Methylumbelliferyl-phosphate	Adamantyl-1,2-dioxyethane Phenylphosphate-substituted dioxyethane
β-Galactosidase	*Escherichia coli*	540 000	6–8	*o*-Nitrophenyl-β-D-galactopyranoside Chlorophenolic red-β-D-galactopyranoside	4-Methylumbelliferyl-β-D-galactopyranoside	–
Peroxidase	Horseradish	40 000	5–7	2,2′-Azino-di(3-ethyl-benzthiazoline sulfonic acid-6) (ABTS)/H_2O_2 3,3′-5,5′-Tetramethyl-benzidine (TMB)/H_2O_2 *o*-Phenylendiamine (OPD)/H_2O_2	*p*-Hydroxyphenyl-acetic acid *p*-Hydroxyphenyl-propionic acid	Luminol

Table 3 Solid phases used for EIAs

Material	Form	Binding	Capacity
Polystyrene	Microtiter plates, tubes, pins, beads	Noncovalent	250–500 ng cm^2
Polyethylene	Tubes	Noncovalent	ca. 300 ng cm^{-2}
Polypropylene	Microtiter plates	Noncovalent	ca. 300 ng cm^{-2}
Polyvinylchloride and similar	Microtiter plates, membranes	Noncovalent	ca. 300 ng cm^{-2} (plates)
Polycarbonate	Beads, membranes	Noncovalent	ca. 300 ng cm^{-2} (beads)
Nitrocellulose	Microtiter plates, membranes	Noncovalent	ca. 100 µg cm^{-2}
Protein A coated	Microtiter plates, beads	Noncovalent	20 mg mL^{-1} (for Ab only)
Activated polymer, with amino or carboxyl groups	Microtiter plates, beads	Covalent, using bifunctional reagents	2×10^{13}–1×10^{14} reactive sites/cm^2
Magnetic	Beads	Depends on the surface of the beads	Varies

The following requirements are necessary for the use of an enzyme as a marker:

(1) high specific activity (turnover number) of free enzyme and after labeling,
(2) availability of soluble, purified enzyme at low cost and reproducible quality,
(3) high stability in free and conjugated form under storage and assay conditions,
(4) presence of reactive groups for covalent linkage to hapten,
(5) simple and gentle conjugation methods,
(6) inexpensive and stable nontoxic substrates with formation of stable chromogenic, fluorogenic and/or chemiluminogenic products.

3.3 Solid Phases

IAs are mainly carried out in 96-well polystyrene, polyethylene, polypropylene or polyvinyl microtiter plates, owing to the easy separation of the reactants in a washing step, but polystyrene tubes, beads or pins are also available (Table 3). The plastic plates are of comparatively low binding capacity and low surface area to volume ratio. High-binding supports include agarose and cellulose. Particulate solid phases are very efficient, because they become scattered throughout the reaction mixture and have a much higher surface area to volume ratio.[41] For example, many chemically different beads are available (e.g. polystyrene, latex, polycarbonate and copolymer beads). Immunological reagents are bound to the beads in a similar manner as they are to microtiter plates. Separation of bound and free reagents occurs by washing and centrifuging. IAs using magnetic beads employ a magnet for the separation step.[42] Abs and Ags may be immobilized to some solid phases via covalent binding. Those solid supports contain amino or carboxy groups on a modified surface through which the immunoreagents can be bound by water-soluble carbodiimides or bifunctional reagents such as glutaraldehyde.

Other solid-phase supports for IAs are membranes. They can be used for dip sticks, which are incubated for a short time in the solution[36] or for dot blots and immunofiltration tests. Here the reactants are filtered through the membrane.[43,44] The test principle is the same as for the microtiter plate tests but the reaction time is much shorter owing to the high surface area of the membrane and the short distance between reaction partners. Application of remission measurements yields a proportional relationship between analyte and remitted light. By using a pocket reflectometer, this set-up is ideally suited for field-monitoring purposes.[45,46]

4 PROPERTIES OF COMPETITIVE IMMUNOASSAYS

4.1 Dose–Response Curve

In IAs the signal produced is inversely correlated to the analyte concentration in the sample (Figure 3a). The typical dose–response curve is of sigmoidal shape when the signal is plotted versus the logarithm of the analyte concentration. A linear range is obtained around the middle of the test (IC_{50}, middle of assay, concentration of analyte that causes 50% inhibition), which should be used for determinations. Within this working range, the change in absorbance is linearly correlated to the analyte concentration. The linear part of the curve is confined by the upper and lower limits of quantification. These are the cut-off values above or below which quantitative results can be obtained with a stated relative precision, or specified degree of confidence in real samples.[47] The experimental errors increase toward these limits. Consequently, the most

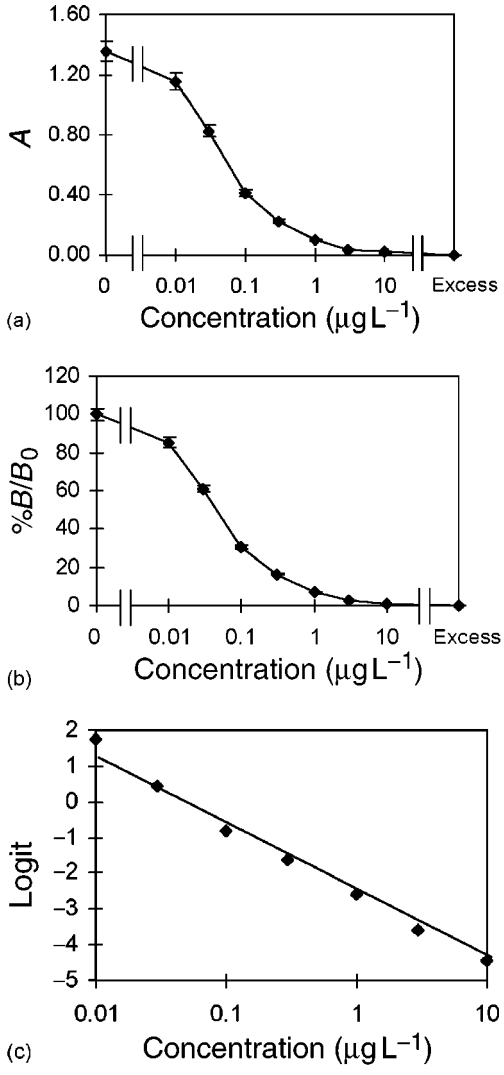

Figure 3 EIA for the determination of atrazine using pAb. (a) Absorption curve (means of three determinations ± standard deviations), (b) B/B_0 curve, and (c) logit/log transformation by the two-parameter fit.

precise measurements are obtained in the region close to the middle of the test.

The DL (or least detectable dose) is the smallest concentration of the analyte that produces a signal which can be significantly distinguished from zero for a given sample matrix with a stated degree of confidence. Very often a dose is selected which inhibits 10–20% of enzyme tracer from binding with the Ab or the dose calculated after subtraction of two or three times the standard deviation from the mean measurements of the zero dose signal.[3]

Linearization of the calibration curve is useful for many purposes, for instance, for the direct comparison of curves if matrix effects are evaluated. Absorbance curves can be normalized by converting the absorptions to $\%B/B_0$ values. These can be expressed as the ratio of bound tracer in the presence of hapten to bound tracer in the absence of hapten and lies between 100% ($=A_0$, the upper asymptote of the curve) and 0% ($= A_{\text{Excess}}$, the lower asymptote) (Figure 3b). They are calculated by Equation (3):

$$\frac{\%B}{B_0} = \frac{A - A_{\text{Excess}}}{A_0 - A_{\text{Excess}}} \times 100 \tag{3}$$

Linearization can be obtained by various mathematical transformations.[47] Usually, IAs are evaluated with commercial IA programs, often based on logistic models (cf. Rodgers[48] and Dudley et al.[49]), e.g. four-parameter models or the more simple logit-log transformation (two-parameter model, Figure 3c) which can also be carried out with a calculator (s), Equation (4):

$$\text{logit}\,\frac{\%B}{B_0} = \ln\frac{\%B/B_0}{100 - \%B/B_0} \tag{4}$$

4.2 Quality Control

Precision and accuracy of IA are important properties which deserve special attention. The quality and stability of the employed material (microtiter plates, pipettes) and reagents (e.g. Abs, enzyme tracer or buffers), play a crucial role.[50] The long-term stability of reagents has to be ensured, e.g. by freeze-drying of Abs and, if necessary, addition of stabilizing components to the test reagents such as the enzyme tracer.[51]

In spite of the simple handling of the assays, expert knowledge is required, especially to recognize and remove incident errors. Therefore, IAs should be performed by trained personnel. The development of simple and rapid assays, such as dip-stick assays or immunofiltration tests reduces the requirement for trained users, but one has still to be aware of potential problems such as interferences from the sample matrix.

The precision of an IA is defined as the extent to which replicate analyses of a sample agree with each other. The reproducibility is the ability to yield the same results within analyses, between analyses, and between operators. The investigation of the variability of an IA gives valuable information about the consistency of the test. Coefficients of variation (CV) of IA measurements are usually between 10 and 20% for an optimized assay,[52,53] although more precise results can be obtained.[54,55] Same-day and day-to-day CV of samples have been determined in different matrices.[53,56] Interlaboratory tests of the same IA as that carried out by Hock and the IA Study Group[57] and Hayes et al.[58] for the investigation of triazines help to evaluate the general applicability of a test. However, several conditions like exact description of the assay including calibration curves, DLs, cross-reactivities, a working range close to the middle of the

test, enough parallel measurements, etc. must be met (see also AOAC (Association of Official Analytical Chemists) criteria). Meanwhile, standardized procedures for IAs in water are adopted by AOAC International and have been established in Germany as a prenorm.[50,58]

A validation of the results obtained by IA should be carried out. To a limited extent this can be done by IA itself. Dilution of the samples as well as spiking of the authentic sample with known amounts of the contaminant can be used to check whether the matrix interferes with the IA.[59] However, spiked samples do not completely mimic real unknown samples. They do not contain potential metabolites of the contaminant nor residues from other compounds which may be present in real samples. Furthermore, spiked samples cannot be a model for aged residues which are more difficult to extract and detect because, for example, they may have bound to soil constituents. Therefore, an IA should also be validated by a different established method like high-performance liquid chromatography (HPLC), gas chromatography (GC) or GC/MS (gas chromatography/mass spectrometry). Many groups have used this approach and have usually obtained correlation coefficients of >0.9.[60–63] Often a slight overestimation of the IA in comparison with HPLC or GC is observed owing to cross-reactivities of the Ab or matrix effects.

4.3 Cross-reactivity

Depending on the conjugate used for immunization and the class of chemicals under investigation, cross-reactivities of the Ab with haptens similar to the analyte are frequently observed (see e.g. Hock[14] and Harrison et al.[64]). Therefore, it should be checked which compounds cross-react to what degree with the Ab. This is usually done by comparing the standard curves of the analyte under investigation with similar haptens, using analyte concentrations at 50% of the inhibition curve as the reference. However, cross-reactivity with a certain analyte is not the same over the whole range of a standard curve. Often higher cross-reactivities can be observed at low concentrations of the cross-reacting analyte.[3] Therefore, it has been recommended that cross-reactivities be measured at different concentrations over the range where the assay is suitable.[65]

If an Ab is selective for a single compound, it is regarded as monospecific[66] (Figure 4). An Ab that recognizes several compounds to the same extent (e.g. a group of s-triazines), can be used for the screening of a class of herbicides[67] (group-specific Ab, Figure 4). If cross-reacting compounds are not expected in the samples, because the compounds are not licensed (e.g. propazine in most European countries), a group-specific Ab can also be used for quantitative measurements of one compound.[68]

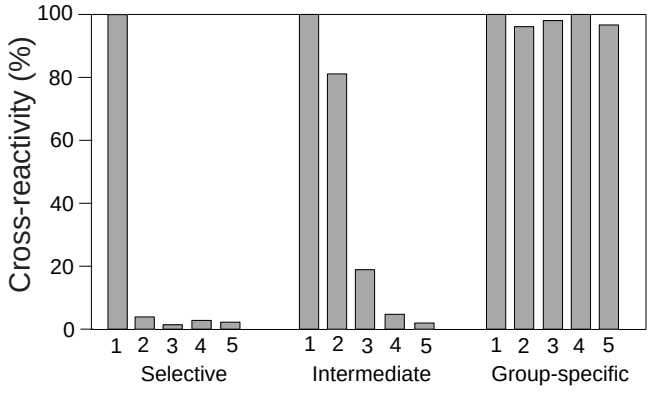

Figure 4 Selective, intermediate and group specific Abs. This example uses Abs which have been produced against hapten 1. Substances 2–5 are assumed to be cross-reacting haptens (modified after Hock et al.[14]).

Strong cross-reactivities of an Ab to unexpected metabolites, for example, can produce false positive values. An Ab for alachlor was found to react very strongly to the sulfonic acid metabolite using an alachlor screening kit.[69] This problem could be solved, however, by using solid-phase extraction (SPE) prior to IA and sequential elution of the two compounds with different organic solvents.

4.4 Sample Preparation and Matrix Effects

Samples can contain compounds in addition to the target analyte, which may interfere with the test. Several groups investigated the influence of ions on EIAs.[70–72] Ruppert et al.[70] observed an inhibition by several anions like azide, which inhibits the peroxidase by binding to the heme group of the enzyme. Most cations did not have an effect except for Ca^{2+}, which leads to an activation of the peroxidase. No interference by different ions such as nitrate, copper, magnesium etc. up to a concentration of 250 ppm was detected in an EIA for pentachlorphenol in water.[71] While ions may inhibit the enzyme used as a label or lead to precipitates by reacting with the buffer components, humic substances present in water or soil extracts may bind nonspecifically to the Ab and thereby interfere with the specific binding of the analyte.[73] These reactions may lead to false positive results. Water samples from forest stands or soil extracts particularly contain a high content of organic compounds such as humic acids (HAs).

Matrix effects in food samples frequently occur owing to colored extracts or to the content of lipids, proteins or polyphenols that may be coextracted during sample preparation.[74] As food samples usually have to be extracted prior to immunochemical analysis, the method of analyte extraction is of great importance. Analytes

that are water soluble and can be efficiently extracted in aqueous buffer will have the most direct extraction method and eliminate the need for organic solvents. However, many pesticides are not readily water soluble and must be extracted with an organic solvent.[5] For the extraction of pesticides from solid foods a variety of solvents have been tested, such as acetone, ether, petroleum ether, methanol, acetonitrile or hexane.[75] Direct analysis of extracts by IA requires the use of solvents that are miscible with water and (at low concentrations) are nondenaturing to proteins such as Ab. IAs are to a certain degree tolerant to a variety of solvents, but each system must be tested to determine which solvent can be accepted and to what extent (for example Hill et al.,[75] Nugent[76] and Schneider and Hammock[77]). Usually the extracts are further diluted with water prior to the EIA, but an EIA for parathion was developed, in which the analyte dissolved in hexane could be directly measured in the EIA without prior removal of the hexane. This was achieved by using Ab encapsulated in reverse micelles composed of Aerosol T with aqueous centers.[78] However, a 10^4-fold decrease in sensitivity was observed.

In some cases a cleanup step is introduced, in which the analyte of interested is separated from the matrix. This can be carried out by C_{18}-columns or immunoaffinity columns.[69,79,80] A very interesting approach is the application of supercritical fluid extraction (SFE) prior to immunoanalysis. These methods generally employ CO_2 or CO_2 containing various modifiers.[5,81]

Some problems with interfering ions can be solved by changing the buffer of the assay system so that no precipitates may be formed.[70] Addition of BSA to the plates prior to the addition of the standard and sample solutions[82] or to the enzyme tracer[73] greatly reduces the influence of humic and fulvic acids on the EIA. It may also be helpful to switch to a different batch of Abs or a different assay kit, as different Abs may show different sensitivities to interfering substances. The buffering capacity of the assay buffer should also be checked, as some water or food samples may show relatively low pH values. No effects, however, were observed between pH 3 and 10 by different investigators.[56,67,83]

5 APPLICATION TO ENVIRONMENTAL SAMPLES

IAs have been developed for many environmental contaminants during the 1990s. A list of several IAs described in the literature can be found in Table 4. Most of them have been developed in laboratories, showing the increasing importance of immunochemical methods in residue analysis. Not all of them are commercially available. Available commercial IAs have been listed in e.g. Dankwardt et al.,[2] Hennion and Barcelo,[3] Knopp,[8] but a lot of movement has been observed in environmental IA markets, leading to the disappearance of IA companies. At the moment IAs for environmental contaminants can be obtained for example from Strategic Diagnostics Inc. (Newark, DE, USA, sells former Ensys, Millipore and Ohmicron kits) and EnviroLogix (Westbrook, MA, USA).

5.1 Water

EIA have been used intensively for the determination of pesticides in surface and rainwater[56,60,69,211–217] and groundwater.[60,69,214,218,219] A substantial number of these studies were carried out for triazine herbicides.[56,60,211,212,214,216,218,219] This illustrates the widespread occurrence of these herbicides in the aquatic environment. Many groups have used commercial test kits, which allow the investigation of samples without time-consuming Ab production. Thurman et al.,[60] for example, used a Res-I-Mune kit (ImmunoSystems) for the investigation of triazines in surface and groundwater. The EIA was compared to GC/MS results obtained from samples that were extracted by SPE. Correlation coefficients between 0.91 and 0.95 were obtained after introducing cross-reactivity factors for each of the triazines in order to calculate a sum parameter for the GC. The majority of the samples contained only atrazine (up to $3\,\mu g\,L^{-1}$). Therefore, the EIA results corresponded well with the atrazine concentrations obtained by GC/MS.

Mouvet et al.[220] compared four commercially available test kits and one in-house developed assay for the determination of triazines in surface and groundwater. Operational characteristics, cross-reactivity, sensitivity, CV and agreement with GC/LC (gas chromatography/liquid chromatography) measurements were investigated. DLs were determined between 0.003 and $0.07\,\mu g\,L^{-1}$. Intra-assay CVs were below 7% for all tests, interassay CVs below 20%. Correlation studies between the EIA kits and GC/LC were carried out for samples from different water matrices. Depending on the water source, different levels of significance were observed with different tests. The best results were obtained for surface water, while not all kits showed a good agreement for lysimeter samples.

Apart from the triazines some other pesticides were investigated in water samples, also using commercial test kits. Alachlor was determined in ground and surface water using commercial tests.[213] SPE was carried out prior to EIA to remove interfering substances and to concentrate the analyte. Concentrations of up to $0.8\,\mu g\,L^{-1}$ were observed, and a comparison with GC/MS

Table 4 Pesticide IAs described in the literature

Pesticides	Test format	Ab	Range, DL, or middle of test (IC_{50})	Ref.
Herbicides				
Alachlor	EIA	p	0.2–8 µg L^{-1}	84
	EIA	p	0.1–10 µg L^{-1}	85
	EIA	p, m	0.2–8 µg L^{-1}	86
Amitrole	EIA	p	1.7–4200 µg L^{-1}	87
Atrazine	CLIA	p	25–500 ng L^{-1}	88
	EIA	p	0.5–10 µg L^{-1}	89
	EIA	p	0.01 µg L^{-1} (DL)	90
	EIA	m	0.03–1 µg L^{-1}	35
	EIA	p	0.2–100 µg L^{-1}	64
	EIA	p	0.011–33 µg L^{-1}	91
	EIA	m	0.05–3 µg L^{-1}	92
	EIA	p, m	0.1–100 µg L^{-1}	93
	EIA	p	0.5–10 µg L^{-1}	94
	EIA	m	0.05 µg L^{-1} (DL)	95
	EIA	m	0.01–10 µg L^{-1}	77
	EIA	p	0.03–3 µg L^{-1}	96
	EIA	p	1–1000 ng L^{-1}	34
Bentazon	EIA	p	2–24 µg L^{-1}	97
Bromacil	EIA	p	0.1–160 µg L^{-1}	98
	EIA	p	0.01–1 µg L^{-1}	99
Chlorodiamino-s-triazine	EIA	p	160–480 µg L^{-1}	100
Chlorsulfuron	EIA	p	0.1 µg L^{-1} (DL)	101
Clomazone	EIA	p	2–250 µg L^{-1}	102
	EIA	p	0.5–500 µg L^{-1}	103
Cyanazine	EIA	p	0.035–3 µg L^{-1}	104
	EIA	p	0.5 µg L^{-1} (DL)	105
	EIA	p	0.5 µg L^{-1} (DL)	106
Diethylatrazine	EIA	p	0.01–100 µg L^{-1}	107
Diclofop-methyl	EIA	p	10–75 µg L^{-1}	108
2,4-D	EIA	p	50–5000 µg L^{-1}	109
	EIA	m	2–20 µg L^{-1}	110
	RIA	p	0.1–10 mg L^{-1}	111
	EIA	p	0.05–10 mg L^{-1}	111
	RIA	p	5–250 µg L^{-1}	112
	PFIA	m	0.6 µg L^{-1} (DL)	113
	RIA	p	1–1000 µg L^{-1}	114
	EIA	m	0.096 µg L^{-1} (DL)	115
Dichlorprop	PFIA	p	0.01–100 µg mL^{-1}	116
Diuron	EIA	m	2 µg L^{-1} (IC_{50})	117
	EIA	p	0.05–1 µg L^{-1}	118
Hexazinone	EIA	p	0.22–17.6 µg L^{-1}	119
Hydroxyatrazine	EIA	m	0.03–1 µg L^{-1}	120
	EIA	m	0.05 µg L^{-1} (DL)	95
	EIA	p	0.01–10 µg L^{-1}	66
	EIA	p	3–300 µg L^{-1}	121
Imazamethabenz	EIA	p	0.5–32 µg L^{-1}	122
Imazaquin	EIA	p	0.45–25 µg L^{-1}	123
Isoproturon	EIA	p	0.01–10 µg L^{-1}	124
	EIA	m	20–250 µg L^{-1}	125
	EIA	NA	0.02–1 µg L^{-1}	126
Maleic hydrazide	EIA	m	0.01–11 µg mL^{-1}	127
MCPB	EIA	p	0.03–0.9 µg L^{-1}	128
Metazachlor	EIA	p	10–1000 ng L^{-1}	129
Methabenzthiazuron	EIA	p	0.05–10 µg L^{-1}	130

Table 4 (*continued*)

Pesticides	Test format	Ab	Range, DL, or middle of test (IC_{50})	Ref.
Metolachlor	EIA	m	0.05–10 μg L^{-1}	131
	EIA	p	0.05–5 μg L^{-1}	132
	EIA	p	6 μg L^{-1} (IC_{50})	133
Molinate	EIA	p	3–2000 μg L^{-1}	134
	EIA	p	10–500 μg L^{-1}	135
Monuron (Diuron)	EIA	p	0.08–5 μg L^{-1}	136
Monuron (Diuron, Linuron)	EIA	p	0.05–5 μg L^{-1}	36
Norflurazon	EIA	p	1–1000 μg L^{-1}	137
Paraquat	RIA	m	0.46–165 μg L^{-1}	138
	FIA	p	20–2000 μg L^{-1}	139
	RIA	p	1–100 μg L^{-1}	140
	EIA	p	20–200 ng L^{-1}	141
	RIA	p	50–1600 μg L^{-1}	142
	EIA	m	10–100 μg L^{-1}	143
	EIA	m	0.8–12 μg L^{-1}	144
	EIA	p	20 ng L^{-1} (DL)	145
	EIA	p	1–100 μg L^{-1}	146
	EIA	p	0.1–27 μg L^{-1}	147
Picloram	EIA	m	1–200 μg L^{-1}	148
	EIA	p	5–5000 μg L^{-1}	148
	RIA	p	0.05–5 mg L^{-1}	149
Propazine	EIA	m	0.02–3 μg L^{-1}	120
Simazine	EIA	p	0.1–10 μg L^{-1}	150
	PFIA	p	3–1000 μg L^{-1}	150
	EIA	p	0.05–10 μg L^{-1}	151
Terbuthylazine	EIA	m	0.14–10 μg L^{-1}	152
Terbutryn	EIA	m	0.05–1 μg L^{-1}	153
	EIA	p	0.1–600 μg L^{-1}	154
	EIA	m	0.1–10 μg L^{-1}	95
Thiobencarb	EIA	p	20–1000 μg L^{-1}	155
Triasulfuron	EIA	p	0.004–40 μg L^{-1}	156
	EIA	m	0.01–1 μg L^{-1}	157
2,4,5-T	RIA	p	1–1000 μg L^{-1}	158
Trifluralin	EIA	p	0.1–1 mg L^{-1}	159
Insecticides				
Aldicarb	EIA	p	0.3–40 μg L^{-1}	160
	EIA	p	1–100 μg L^{-1}	161
Aldrin	RIA	p	0.7–35 ng	162
Azinphos-methyl	EIA	m	0.4–20 μg L^{-1}	163
S-Bioallethrin	EIA	m	1–250 μg L^{-1}	164
	EIA	p	0.8–28 μg L^{-1}	165
	RIA	p	0.03–3 μg	166
	EIA	p	0.01–100 mg L^{-1}	167
Bioresmethrin	EIA	p	50–10 000 μg L^{-1}	168
Carbaryl	EIA	m	0.02–20 μg L^{-1}	169
	EIA	p	0.05–10 μg L^{-1}	170
Carbofuran	EIA	p	0.056–5 μg L^{-1}	72
Chlorpyriphos	EIA	m	2–45 μg L^{-1}	55
DDA	EIA	p	10–100 μg L^{-1}	171
DDT	EIA	m	2–11 nM L^{-1} (IC_{50})	172
Dieldrin	RIA	p	0.08–38 ng	162
Diflubenzuron	EIA	p	0.5–15 μg L^{-1}	173
Endosulfan	EIA	p	3–400 μg L^{-1}	174
	EIA	p	0.2–10 μg L^{-1}	175
	EIA	p	3–500 μg L^{-1}	176

(*continued overleaf*)

Table 4 (*continued*)

Pesticides	Test format	Ab	Range, DL, or middle of test (IC_{50})	Ref.
Fenitrothion	EIA	p, m	1–1000 μg L^{-1}	177
Chlorpyrifos	EIA	p	0.05–2.5 μg L^{-1}	178
Heptachlor	EIA	m	10–200 μg L^{-1}	179
Methoprene	EIA	p	0.1–100 μg L^{-1}	75
	EIA	p	1–10 μg L^{-1}	180
1-Naphthol	EIA	p	10–1000 μg L^{-1}	181
Parathion	RIA	p	100 μg L^{-1} (DL)	182
	EIA	p	0.1 μg L^{-1} (DL)	183
	EIA	p	0.1–10 μg L^{-1}	78
	EIA	p	0.2 μg L^{-1} (DL)	184
Paraoxon	EIA	m	10–100 μg mL^{-1}	185
	RIA	p	0.2–19 ng L^{-1}	186
	EIA	p	28–275 μg L^{-1}	186
	EIA	p	0.03 μg L^{-1}–275 mg L^{-1}	187
PCP	EIA	p	0.3–150 μg L^{-1}	188
	EIA	p	0.1–10 μg L^{-1}	71
	EIA	p	25 μg L^{-1} (DL)	189
Permethrin	EIA	p	10–6000 μg L^{-1}	190
	EIA	m	15–100 μg L^{-1}	191
	EIA	m	1.5–90 μg L^{-1}	192
Pyrimiphos-methyl	EIA	p	30 μg L^{-1} (DL)	193
Quassin	EIA	p	0.05–3 μg L^{-1}	194
3,5,6-Trichloro-2-pyridinol	EIA	m	0.25 μg L^{-1}	195
Fungicides				
Benomyl	EIA	m	0.1–1 μg L^{-1}	196
	PFIA	p	0.1–10 μg L^{-1}	197
	RIA	p	1.25 μg L^{-1} (DL)	198
	EIA	p	0.38 μg L^{-1} (DL)	199
	EIA	p	0.35 μg L^{-1} (DL)	200
Benzimidazole	EIA	m	1–20 μg L^{-1}	201
Captan	EIA	p	1–200 μg L^{-1}	202
Chlorothalonil	EIA	p	0.07 μg L^{-1} (DL)	203
Fenpropimorph	EIA	p	0.1–0.8 μg L^{-1}	204
Iprodione	EIA	p	0.1–10 mg kg^{-1}	205
Metalaxyl	EIA	p	0.06–1 μg L^{-1}	206
Myclobutanil	EIA	p	0.5–50 mg L^{-1}	207
Procymidone	EIA	p	1 μg L^{-1} (DL)	141
Thiabendazole	EIA	m	0.5–10 μg L^{-1}	208
	EIA	p	0.03–0.5 μg L^{-1}	141
Triadimefon	EIA	p	2 μg L^{-1} (DL)	209
Triazole	EIA	p	10–1200 μg L^{-1}	210

MCPB, 4-(4-chloro-2-methylphenoxy)butyric acid; 2,4,5-T, 2,4,5-trichlorophenoxyacetic acid; DDA, bis(*p*-chlorophenyl)acetic acid; DDT, 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane; 2,4-D, 2,4-dichlorophenoxyacetic acid; PCP, pentachlorophenol. NA, not available.

showed a correlation coefficient of 0.95 with a slight underestimation by EIA. The occurrence of carbaryl was determined by Marco et al.[221] in wellwater from Spain with their own assay and compared with a commercial test kit. Both IAs yielded a good agreement with conventional methods. Concentrations of 0.08–1.37 μg L^{-1} were observed. Two commercial test kits were used in the Netherlands to determine 2,4-D concentrations in the rivers Rhine and Meuse.[217] By diluting the kit standards with kit zero-buffer and calculating the DL on the basis of the error in the zero-standard, the DL as originally indicated by the manufacturer was significantly lowered. The water matrix substantially affected the recovery of 2,4-D with one assay kit, yielding unexpectedly low recoveries in demineralized and tap water. However, similar results were obtained by EIA and GC/MS for spiked samples from the river Rhine (with a slope of about 1 and $r = 0.99$). Routine samples were also analyzed

and yielded analyte concentrations mostly below the DL ($0.03–0.05\ \mu g\ L^{-1}$).

5.2 Soil

Many studies for pesticides in soil have been carried out for triazines.[38,73,222–227] Other analytes include carbofuran,[72] chlorpyrifos,[178] hexazinone,[119] metolachlor,[227] 4-nitrophenol,[184] PCP,[71] and triasulfuron.[157,228]

Usually, the pesticides are extracted from soil with organic solvents such as acetone, ethyl acetate or methanol. In some cases, for example in the investigation of the quite water-soluble atrazine, extraction with water was sufficient. In one study no difference was observed in EIA measurements after extraction of atrazine from corn fields with water or acetone.[225] Owing to the extraction with water it was very easy to apply the samples directly to the IA. However, in most cases the sensitivity of the assays is high enough to dilute organic solvent extracts sufficiently so that they do not interfere with the assay. The use of supercritical fluids for soil extraction is becoming increasingly popular. The low critical temperature (e.g. 31 °C for CO_2) means that low extraction temperatures can be used to recover thermally unstable solutes.[184,229] Wong et al.[184] compared SFE with solvent extraction using ethyl acetate for the investigation of parathion and 4-nitrophenol from soil. The extracts could be measured directly after SFE by IA or GC without solvent exchange. Similar concentrations were observed by both extraction methods.

The use of IA with magnetic particles has been described for the investigation of PCP, carbofuran, and triasulfuron in soil.[71,72,228] The dispersion of the covalently immobilized Ab on the surface of the magnetic beads results in a more even distribution in the sample and leads to a lower variability of measurements, which is usually higher in soil extracts than in water samples.

Immunochemical methods have also been applied to the detection of bound pesticide residues in soil. These are formed by binding of pesticides to the organic matter of the soil, mainly humic and fulvic acids, and cannot be analyzed using common extraction and assay methods. Hahn et al.[223] used Fab fragments labeled with a fluorescent dye to detect nonextractable residues of atrazine in soil from corn fields. The fluorescence signal obtained was related to the amount of bound atrazine in native soil samples determined by GC after supercritical methanol extraction. A noncompetitive sandwich IA for the analysis of bound residues based on HA–Ab and triazine Ab was developed by Ulrich et al.[38] HA was extracted from soil, bound to the plates by the HA–Ab and the nonextractable triazine residues were detected by a peroxidase-labeled Ab against triazines. Concentrations of bound atrazine were found in the region of 2 mg per kg soil in corn fields with a long history of atrazine application. A competitive EIA was used by Dankwardt et al.[225] for the investigation of bound residues. Free atrazine residues were removed from organic matter extracts by SPE and the extracts were then measured by EIA.

5.3 Food

A variety of pesticides has been determined in food samples by EIA. In many cases Abs against the target compound were developed in-house. Mainly pAbs were still used, but mAbs were applied for the determination of atrazine, benzimidazoles, thiabendazoles and carbaryl.[46,201,208,230]

Peel of apples, potatoes, oranges, grapefruits and bananas were investigated for thiabendazole residues employing an EIA with mAb.[231] Residues were extracted by soaking peel overnight in 80% methanol and filtering the decanted supernatants. Most of the thiabendazole was extracted within 1 h, but an extraction time of 16 h was chosen as uniformly high recoveries from all matrices were obtained. A 20-fold dilution eliminated significant matrix effects. The EIA had a DL of 0.1 ppm in peel samples, corresponding to 10–40 ppb in the whole fruit or tuber. Results obtained by EIA were compared with HPLC analyses. Although the EIA values were higher compared with the HPLC results, the two sets of data correlated well. The higher values by EIA were attributed to the loss of thiabendazole during sample work-up for HPLC and the high dilutions necessary for EIA.

Triadimefon was added to different food commodities such as apples, pears, pineapples and grapes. Fungicide residues were determined by EIA and GC after extraction with ethyl acetate or methanol.[209] Ethyl acetate was required to extract grapes for the EIA procedure, since methanol resulted in coextractives which gave too low values. Methanol, however, yielded good recoveries at 0.5 ppm and above for other commodities and is preferable to ethyl acetate because it avoids an evaporation step. Low recoveries were obtained at 0.1 ppm and were not improved using ethyl acetate. On the whole, recoveries obtained by EIA correlate well with those measured by GC. The middle of the test was observed at $2.4\ \mu g\ L^{-1}$.

A mAb-based EIA was applied to the determination of carbaryl in apple and grape juices.[230] The juices were used without any sample pretreatment and spiked with different carbaryl concentrations. The influence of matrix dilution was investigated using different dilutions of the samples. For proper analysis the samples should be diluted at least 1 : 5–1 : 10. With a dilution of 1 : 100 the most accurate and precise results were obtained.

Therefore 2–5 $\mu g\,L^{-1}$ are considered the lowest cabaryl concentrations in juices that can be reliably measured with the EIA. CVs ranged from 4 to 13%, with most of them below 8%.

A dip-stick IA using mAb immobilized on a membrane was used for the determination of atrazine in water and liquid food samples.[46] The measuring range was 0.3–10 $\mu g\,L^{-1}$ using reflectance detection. The total assay time was 25 min using precoated dip sticks. The atrazine concentrations could be determined directly in spiked water, milk and juice samples yielding satisfactory agreement with the spiking concentrations. The black tea samples, however, showed an overestimation due to the unspecific binding of the tannins to the membrane.

Paraquat was determined in milk, beef and potatoes using an EIA with pAb.[232] Potatoes were shredded with dry ice. The potato and the meat samples were extracted with HCl after spiking the samples. The acid extracts were evaporated to dryness and reconstituted for the EIA. The milk was diluted with phosphate buffer. The EIA was able to detect less than 1 ppb of paraquat in whole milk and down to 2.5 ppb in beef. The efficiency of HCl for extracting potato and ground beef was determined by using methyl-^{14}C paraquat. Recoveries between 60 and 70% were obtained. Since paraquat is known to bind tightly to many matrices, recoveries for both matrices were determined after storage of the spiked samples for several days at 21 °C.

Ibrahim et al.[233] investigated eggs for aldrin and dieldrin residues. An EIA with pAb was applied. The EIA detected only dieldrin, but aldrin is metabolized to dieldrin. The eggs were homogenized after removal of the shell. This solution was diluted 1 : 2 with wash buffer containing 2% BSA. Egg samples were collected in Egypt and assayed by EIA. The standard curve for dieldrin was prepared in egg solution of noncontaminated eggs. The egg samples showed concentrations of aldrin and dieldrin in eggs up to 0.7 $mg\,L^{-1}$. These are concentrations above the World Health Organisation average daily intake levels.

5.4 Biomonitoring

Human biomonitoring involves the measurement of a parent chemical and/or metabolites or a product of its reaction with cellular components (e.g. protein adduct, nucleic acid adduct) in selected tissues, body fluids such as blood, milk, urine or sweat, or expired breath of an exposed individual.[8,9] Most IAs for pesticides are sensitive enough for biomonitoring. Many analytes could be determined without any sample preparation other than dilution with water or buffer. In some cases the sample was filtered. Slightly reduced sensitivity or higher blank values due to matrix effects were sometimes found when the assay was performed directly in the biological sample. With increasing hydrophobicity of the analyte, sample preparation becomes more complicated.[8]

Validation experiments are carried out by fortification of samples with the analyte in question and comparison with independent control methods such as GC, GC/MS or LC. Samples from a number of individuals should be used in any validation experiment because of the inherent variability of biological matrices. Features such as pH, protein, sugar and salt concentration must be considered. Also the intake of drugs can interfere with the immunological determination of environmental pollutants.[8]

While many investigations were restricted to spiked samples only, some studies demonstrated the use of IA for surveillance of occupationally exposed individuals (e.g. for DDA, a metabolite of DDT, 2,4-D, paraquat and atrazine mercapturate).[140,171,234,235]

6 NEW DEVELOPMENTS

The strength of immunochemical methods lies in the screening of a large number of samples within a short time at low cost. Therefore, they can be valuable supplements to conventional analytical methods. Important applications are seen in the analysis of ground and drinking water, where matrix effects are seldom observed. Also, food commodities that turn over quickly are ideal targets for IA measurements. Owing to the low cost of one analysis more replicates from one site can be measured or special sites can be sampled more often, for example to obtain more information about variations of analyte concentrations depending on seasons or rainfall, etc.

Some restrictions are imposed by the fact that IAs are de facto single analyte methods. However, new approaches for multianalyte measurements are being undertaken, such as the integration of IA with LC. Here, Abs are used in conjunction with LC, e.g. to concentrate an analyte from a large volume of sample and separate it from an interfering matrix.[236–239] In this case an immunoadsorbent column is used before analysis by LC. The immunoadsorbent column contains immobilized specific Abs which bind the analyte, while interfering substances pass through. The analyte can be eluted by using a pH gradient[79] or an organic solvent.[80] Therefore, large sample volumes with low concentrations of the analyte can be reduced to small volumes with sufficiently high concentrations without coextracting interfering substances like HAs or food compounds. This raises the effective sensitivity of the analysis. Ab mixtures can be used to bind substances from different compound classes, e.g. the phenyl urea herbicides and the triazines.[239] In this case, the eluted compounds were injected into the LC, yielding a DL of 0.03–0.5 $\mu g\,L^{-1}$ from samples volumes as low as 25 or 50 mL.

When cross-reacting Abs are applied in IA, the signal obtained is not only related to the analyte, but also to similar compounds. This problem can be circumvented by the use of LC prior to the IA. LC/IA was applied by Krämer et al.[236] to determine 4-nitrophenols. The nitrophenols were separated by different LC systems and determined by IA. LC/IA was about 8–10 times more sensitive compared with LC with ultraviolet detection. Therefore, the integration of LC with IA combines the high separation quality of LC and the sensitivity of IA.[240,241]

Furthermore, multianalyte systems are under development. One concept is the microspot IA,[242] which uses many microspots with fluorescence-labeled Abs of different selectivity immobilized on a chip. After incubation with the analyte (Ag or hapten) a fluorescence-labeled tracer Ab is added. The tracer Ab is either directed against the Ag or consists of an anti-idiotype Ab directed against the binding site of the capture Ab. Sensor and tracer Ab carry different fluorescence labels. Therefore it is possible to determine the amount of analyte bound to the sensor Abs with optical scanning methods by measuring the signal ratio (ratiometric assay). Lately, a variety of noncompetitive and competitive microspot analysis systems have been developed, mainly related to the medical field,[243] but are clearly of particular importance in areas such as environmental monitoring.

Another possibility is the use of cross-reacting Abs for multianalyte detection. Known cross-reactivities of different Abs can be used to calculate the different concentrations of different analytes in a sample containing several contaminants.[244] The estimation of the individual concentrations is carried out by complex calculating procedures, e.g. by neural networks[245] or iterative procedures.[246]

Immunochemical analysis is a fast developing field with numerous possibilities for further improvement. Much effort is being put into the development of continuous measurements, such as flow-injection immunoanalysis (FIIA) and immunosensors.[247,248] A quasicontinuous FIIA of pesticides was developed by Krämer and Schmid[249] on the basis of a competitive IA. Here, the Abs are immobilized on a membrane. The reaction takes place in the membrane reactor, the central part of the flow injection system. All reagents are sequentially added to the reactor and the product is assayed with the aid of a flow fluorimeter. The measuring range of the flow injection analysis almost equals that of the EIA. Wittmann and Schmid[250] used an Ab column reactor filled with polystyrene or glass beads with the Ab immobilized via the avidin/biotin system. This system showed a stable Ab activity for a minimum of 500 measuring cycles. DLs for atrazine of about $1\,ng\,L^{-1}$ with pAb and $30\,ng\,L^{-1}$ with mAb could be reached.

Important progress is to be expected in the field of immunosensors where the detectors are based on Abs.[247] Some relatively simple devices such as dip sticks and immunofiltration assays have been mentioned before (see section 3.3). An interesting development is liposome-amplified immunomigration strips.[251,252] They employ liposome-encapsulated markers which act as signal enhancers of the competitive binding reaction instead of enzymes. These devices have been used for the determination of alachlor.[251] If a pesticide of interest is conjugated to a lipid it can also be incorporated into the liposome structure, leading to a competitive liposome IA.[253] In more complicated systems the immunological recognition system is immobilized in the direct vicinity of a transducer, an electrochemical, optical or gravimetric device. They respond to chemical compounds or ions and yield electrical signals which depend on the concentration of the analyte. Immunosensors with piezoelectric crystals as physical sensors are in a relatively advanced state of development.[254] They function as microbalances onto which Abs are immobilized. Other physical sensors use optical systems such as surface plasmon resonance (SPR), interferometry or grating couplers.[248,255,256] A biosensor employing SPR was used for the determination of atrazine.[256] A DL of $0.05\,\mu g\,L^{-1}$ for atrazine in water was reached with an analysis time of 15 min. Bier and Schmid[257] used a grating coupler immunosensor for the determination of terbutryn, a triazine herbicide. A DL of $15\,nmol\,L^{-1}$ (ca. $3.6\,\mu g\,L^{-1}$) was established. Interesting developments can be expected from Ab electrodes.[258]

New strategies for Ab production are also being developed. Genetically engineered Abs appear very attractive because their selectivity and affinity can be tailored by site-directed mutations without requiring new immunizations.[259] Methods are now provided to isolate desired clones rapidly from Ab libraries and to manipulate individual rAbs to match specific demands of environmental analysis. Binding proteins derived from Abs but consisting only of a part of their light or heavy chain (scFv, Figure 2) and rAb fragments, Fab, directed against different s-triazines, diuron and parathion have been produced.[260–264] In several cases the DL of the rAb was the same as with the parent mAb.[263,264]

A promising goal is the completely synthetic production of binding proteins or other synthetic receptors which are fitted to the structure of the analyte by molecular design. The use of libraries guarantees to close the bottleneck in Ab production. Abs with special properties such as resistance to matrix effects or organic solvent stability can also be selected from libraries, providing an important contribution to the analysis of water and food samples.

ACKNOWLEDGMENTS

I am grateful to Dr Armin Naß, Sension GmbH, Augsburg, for reading the manuscript and to Dr Sabine Pullen, Institute for Pharmacology and Toxicology, University of Erlangen, for providing some material on immunoassays for pesticides.

ABBREVIATIONS AND ACRONYMS

Ab	Antibody
Ag	Antigen
AOAC	Association of Official Analytical Chemists
BSA	Bovine Serum Albumin
CDR	Complementarity Determining Region
CLIA	Chemiluminescent Immunoassay
CV	Coefficients of Variation
DDA	Bis(*p*-chlorophenyl)acetic Acid
DDT	1,1,1-Trichloro-2,2-bis-(*p*-chlorophenyl)ethane
DL	Detection Limit
EIA	Enzyme Immunoassay
ELISA	Enzyme-linked Immunosorbent Assay
ESIMS	Electrospray Ionization Mass Spectrometry
Fab or $F(ab)_2$	Antibody Fragments Containing the Antigen Binding Site(s)
Fc	Antibody Fragment Containing the Crystallizable Fragment
FIA	Fluorescence Immunoassay
FIIA	Flow-injection Immunoanalysis
FR	Frame Region
GC	Gas Chromatography
GC/LC	Gas Chromatography/Liquid Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HA	Humic Acid
HPLC	High-performance Liquid Chromatography
IA	Immunoassay
IC_{50}	Middle of Assay, Concentration of Analyte that Causes 50% Inhibition
Ig	Immunoglobulin
KLH	Keyhole Limpet Hemocyanin
LC	Liquid Chromatography
mAb	Monoclonal Antibody
MALDIMS	Matrix-assisted Ultraviolet Laser Desorption/Ionization Mass Spectrometry
MCPB	4-(4-Chloro-2-methylphenoxy)butyric Acid
pAb	Polyclonal Antibody
PCP	Pentachlorophenol
PFIA	Polarization Fluoroimmunoassay
rAb	Recombinant Antibody
RIA	Radioimmunoassay
scFv	Recombinant Antibody Fragment, Single Chain Fragment Containing only the Variable Region
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction
SPR	Surface Plasmon Resonance
2,4-D	2,4-Dichlorophenoxyacetic Acid
2,4,5-T	2,4,5-Trichlorophenoxyacetic Acid

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Immunochemistry

Environment: Water and Waste **(Volume 3)**
Immunoassay Techniques in Environmental Analyses

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food

Forensic Science **(Volume 5)**
Immunoassays in Forensic Toxicology

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Herbicide Residues in Biota, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation

REFERENCES

1. J. Sherry, 'Environmental Immunoassays and Other Bioanalytical Methods: Overview and Update', *Chemosphere*, **34**, 5–7 (1997).
2. A. Dankwardt, S. Pullen, B. Hock, 'Immunoassays. Applications for the Aquatic Environment', in *Microscale Aquatic Toxicology – Advances, Techniques and Practice*, eds. P.G. Wells, K. Lee, C. Blaise, CRC Lewis Publishers, Florida, 13–29, 1998.
3. M.-C. Hennion, D. Barcelo, 'Strengths and Limitations of Immunoassays for Effective and Efficient Use for Pesticide Analysis in Water Samples: A Review', *Anal. Chim. Acta*, **362**, 3–34 (1998).

4. D.S. Aga, E.M. Thurman, 'Environmental Immunoassays: Alternative Techniques for Soil and Water Analysis', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, American Chemical Society, Washington, DC, 1–20, 1997.
5. V. Lopez-Avila, C. Charan, J. Van Emon, 'Supercritical Fluid Extraction–Enzyme-linked Immunosorbent Assay Applications for Determination of Pesticides in Soil and Food', in *Immunoassays for Residue Analysis: Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 439–449, 1996.
6. S.W. Jourdan, A.M. Scutellaro, M.C. Hayes, D.P. Herzog, 'Adapting Immunoassays for the Analysis of Food Samples', in *Immunoassays for Residue Analysis: Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 82–98, 1996.
7. A. Dankwardt, B. Hock, 'Enzyme Immunoassays for the Analysis of Pesticides in Water and Food,' *J. Food Technol. Biotechnol.*, **35**, 165–174 (1997).
8. D. Knopp, 'Application of Immunological Methods for the Determination of Environmental Pollutants in Human Biomonitoring. A Review', *Anal. Chim. Acta*, **311**, 383–392 (1995).
9. A.S. Harris, A.D. Lucas, P.M. Krämer, M.-P. Marco, S.J. Gee, B.D. Hammock, 'Use of Immunoassays for the Detection of Urinary Biomarkers of Exposure', in *New Frontiers in Agrochemical Immunoassay*, eds. D.A. Kurtz, J.H. Skerritt, L. Stanker, AOAC International, Arlington, VA, 217–235, 1995.
10. S. Arrhenius, *Immunchemie*, Akademische Verlagsgesellschaft, Leipzig, 1907.
11. R.S. Yalow, S.A. Berson, 'Assay of Plasma Insulin in Human Subjects by Immunological Methods', *Nature*, **184**, 1643–1644 (1959).
12. E. Engvall, P. Perlmann, 'Enzyme-linked Immunosorbent Assay (ELISA): Quantitative Assay of Immunoglobulin G', *Immunochemistry*, **8**, 871–874 (1971).
13. B.K. Van Weemen, A.H.W.M. Schuurs, 'Immunoassay Using Antigen–Enzyme Conjugates', *FEBS Lett.*, **15**, 232–236 (1971).
14. B. Hock, A. Dankwardt, K. Kramer, S. Pullen, 'Toxicity Assessment and On-line Monitoring: Immunoassays', *Environ. Toxicol. Water Qual.*, **9**, 243–262 (1994).
15. E.S. Golub, *Immunology: A Synthesis*, Sinaver Associates, Sunderland, MA, 1987.
16. I. Roitt, 'Essential Immunology', Blackwell Scientific, Oxford, 1991.
17. B. Hock, T. Giersch, K. Kramer, A. Dankwardt, 'Antibody Production and Progress in Hybridoma Technology by Immunomagnetic Screening', in *New Frontiers in Agrochemical Immunoassay*, eds. D.A. Kurtz, J.H. Skerritt, L. Stanker, AOAC International, Arlington, VA, 149–162, 1995.
18. G. Köhler, C. Milstein, 'Continuous Culture of Fused Cells Secreting Antibody of Predefined Specificity', *Nature*, **256**, 495–497 (1975).
19. B. Hock, A. Dankwardt, K. Kramer, A. Marx, 'Immunochemical Techniques: Antibody Production for Pesticide Analysis. A Review', *Anal. Chim. Acta*, **311**, 393–406 (1995).
20. P.V. Choudary, H.A. Lee, B.D. Hammock, M.R.A. Morgan, 'Recombinant Antibodies: New Tools for Immunoassays', in *New Frontiers in Agrochemical Immunoassay*, eds. D.A. Kurtz, J.H. Skerritt, L. Stanker, AOAC International, Arlington, VA, 171–186, 1995.
21. K. Kramer, B. Hock, 'Recombinant Antibodies for Pesticide Analysis', in *Residue Analysis in Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 471–484, 1996.
22. H.A. Lee, M.J.C. Alcocer, T.G. Lacarra, D.J. Jeenes, M.A. Morgan, 'Recombinant Antibodies: Expression in *Escherichia coli* Using Plasmid and Phagemid Vectors and Application to Food Analytes', in *New Frontiers in Agrochemical Immunoassay*, eds. D.A. Kurtz, J.H. Skerritt, L. Stanker, AOAC International, Arlington, VA, 187–196, 1995.
23. J.C. Hall, G.M. O'Brien, S.R. Webb, 'Phage-Display Technology for Environmental Analysis', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, American Chemical Society, Washington, DC, 22–37, 1997.
24. B.F. Erlanger, 'Principles and Methods for the Preparation of Drug Protein Conjugates for Immunological Studies', *Pharmacol. Rev.*, **25**, 271–280 (1973).
25. B.F. Erlanger, 'The Preparation of Antigenic Hapten–Carrier Conjugates: A Survey', *Methods Enzymol.*, **70**, 85–103 (1980).
26. M.H. Goodrow, J.R. Sanborn, D.W. Stoutmire, S.J. Gee, B.D. Hammock, 'Strategies for Immunoassay Hapten Design', in *Immunoanalysis of Agrochemicals: Emerging Technologies*, eds. J.O. Nelson, A.E. Karu, R.B. Wong, ACS Symposium Series 586, American Chemical Society, Washington, DC, 119–139, 1995.
27. F. Szurdoki, H.K.M. Bekheit, M.-P. Marco, M.H. Goodrow, B.D. Hammock, 'Important Factors in Hapten Design and Enzyme-linked Immunosorbent Assay Development', in *New Frontiers in Agrochemical Immunoassay*, eds. D.A. Kurtz, J.H. Skerritt, L. Stanker, AOAC International, Arlington, VA, 39–63, 1995.
28. M. Böcher, T. Giersch, R.D. Schmid, 'Dextran, a Hapten Carrier in Immunoassays for s-Triazines – a Comparison with ELISAs Based on Hapten–Protein Conjugates', *J. Immunol. Methods*, **151**, 1–8 (1992).
29. A.M.J. Buiting, N. Van Rooijen, E. Claassen, 'Liposomes as Antigen Carrier and Adjuvants in vivo', *44th Forum in Immunology*, 541–548 (1992).

30. M.B. Rittenberg, A.A. Amkraut, 'Immunogenicity of Trinitrophenyl–Hemocyanin: Production of Primary and Secondary Anti-hapten Precipitins', *J. Immunol.*, **97**, 421–430 (1966).
31. I. Wengatz, R.D. Schmid, S. Kreißig, C. Wittmann, B. Hock, A. Ingendoh, F. Hillenkamp, 'Determination of the Hapten Density of Immunoconjugates by Matrix-assisted UV Laser Desorption/Ionization Mass Spectrometry', *Anal. Lett.*, **25**, 1983–1997 (1992).
32. B.G. Rose, S.A. Buckley, C. Kamps-Holtzapple, R.C. Beier, L. Stanker, 'Molecular Modeling Studies of Ceftiofur: A Tool for Hapten Design and Monoclonal Antibody Production', in *Residue Analysis in Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 82–98, 1996.
33. T. Porstmann, S.T. Kiessig, 'Enzyme Immunoassay Techniques. An Overview', *J. Immunol. Methods*, **150**, 5–21 (1992).
34. C. Wittmann, B. Hock, 'Improved Enzyme Immunoassay for the Analysis of s-Triazines in Water Samples', *Food Agric. Immunol.*, **1**, 211–224 (1989).
35. T. Giersch, 'A New Monoclonal Antibody for the Sensitive Detection of Atrazine with Immunoassay in Microtiter Plate and Dipstick Format', *J. Agric. Food Chem.*, **41**, 1006–1011 (1993).
36. P. Schneider, M.H. Goodrow, S.J. Gee, B.D. Hammock, 'A Highly Sensitive and Rapid ELISA for the Arylurea Herbicides Diuron, Monuron and Liuron', *J. Agric. Food Chem.*, **42**, 413–422 (1994).
37. S.A. Eremin, 'Polarization Fluoroimmunoassay for Rapid Specific Detection of Pesticides', in *Immunoanalysis of Agrochemicals*, eds. J.O. Nelson, A.E. Karu, R.B. Wong, ACS Symposium Series 586, American Chemical Society, Washington, DC, 223–234, 1995.
38. P. Ulrich, R. Niessner, 'Development of a Sandwich Immunoassay for the Determination of Bound Residues', *Fresenius' J. Anal. Chem.*, **354**, 352–358 (1996).
39. A. Dankwardt, B. Hock, 'Determination of Bound Triazine Residues in Soil by Immunochemical Methods', *GIT Fachz. Lab.*, **39**, 721–722 (1995).
40. J.P. Gosling, 'A Decade of Development in Immunoassay Methodology', *Clin. Chem.*, **36**, 1408–1427 (1990).
41. S.L. Hefle, 'Immunoassay Fundamentals', *Food Technol.*, **49**, 102–107 (1995).
42. J.A. Itak, M.Y. Selisker, D.P. Herzog, J.R. Fleeker, E.R. Bogus, R.O. Mumma, 'Determination of Captan in Water, Peaches, and Apple Juice by a Magnetic Particle-based Immunoassay', *J. AOAC Int.*, **77**, 86–91 (1994).
43. M.E. Ploum, W. Haasnoot, R.J.A. Paulussen, G.D. Van Bruchem, A. Hames, R. Schilt, F.A. Hug, 'Test Strip Enzyme Immunoassays and the Fast Screening of Nortesterone and Clenbuterol Residues in Urine Samples at the Parts per Billion Level', *J. Chromatogr.*, **546**, 413–427 (1991).
44. A. Dankwardt, B. Hock, 'Rapid Immunofiltration Assay for the Detection of Atrazine in Water and Soil Samples', *Biosens. Bioelectron.*, **8**, XX–XXI (1993).
45. R. Niessner, 'Immunoassays in Environmental Analytical Chemistry: Some Thoughts on Trends and Status', *Anal. Methods Instrument.*, **1**, 134–144 (1994).
46. C. Wittmann, U. Bilitewski, T. Giersch, U. Kettling, R.D. Schmid, 'Development and Evaluation of a Dipstick Immunoassay Format for the Determination of Atrazine Residues On-site', *Analyst*, **121**, 863–869 (1996).
47. J. Brady, 'Interpretation of Immunoassay Data', in *Immunoanalysis of Agrochemicals*, eds. J.O. Nelson, A.E. Karu, R.B. Wong, ACS Symposium Series 586, American Chemical Society, Washington, DC, 266–287, 1995.
48. R.P.C. Rodgers, 'Data Analysis and Quality Control of Assays: A Practical Primer', in *Clinical Immunoassay: The State of the Art*, ed. W.R. Butt, Marcel Dekker, New York, 253–308, 1984.
49. R.A. Dudley, P. Edwards, R.P. Ekins, D.J. Finney, I.G. McKenzie, M. Raab, M.D. Rodbard, R.P.C. Rodgers, 'Guidelines for Immunoassay Data Processing', *Clin. Chem.*, **31**, 1264–1271 (1985).
50. German Standard Methods for the Examination of Water, Wastewater and Sludge – Sub-animal Testing (group T) – Part 2: Guideline for Selective Immunotest Methods (Immunoassays) for the Determination of Plant Treatment and Pesticide Agents (T2), DIN V 38415-2, Beuth Verlag, Berlin, 1995.
51. A. Dankwardt, J. Müller, B. Hock, 'Stabilization of Enzyme Immunoassays for Atrazine', *Anal. Chim. Acta*, **362**, 35–45 (1998).
52. S.J. Huber, 'Improved Solid-phase Enzyme Immunoassay Systems in the ppt Range for Atrazine in Fresh Water', *Chemosphere*, **14**, 1795–1803 (1985).
53. R.J. Bushway, L.B. Perkins, H.L. Hurst, 'Determination of Atrazine in Milk by Immunoassay', *Food Chem.*, **43**, 283–287 (1992).
54. J. Van Emon, B.D. Hammock, J.N. Seiber, 'Enzyme-linked Immunosorbent Assay for Paraquat and its Application to Exposure Analysis', *Anal. Chem.*, **58**, 1866–1873 (1986).
55. J.J. Manclus, A. Montoya, 'Development of an Enzyme Immunoassay for the Analysis of Chlorpyrifos and its Major Metabolite 3,5,6-Trichloro-2-pyridinol in the Aquatic Environment', *Anal. Chim. Acta*, **311**, 341–348 (1995).
56. A. Dankwardt, S. Wüst, W. Elling, E.M. Thurman, B. Hock, 'Determination of Atrazine in Rainfall and Surface Water by Enzyme Immunoassay', *Environ. Sci. Pollut. Res.*, **1**, 196–204 (1994).
57. B. Hock and The Immunoassay Study Group, 'Enzyme Immunoassays for the Determination of s-Triazines in Water Samples: Two Interlaboratory Tests', *Anal. Lett.*, **24**, 529–549 (1991).

58. M.C. Hayes, S.W. Jourdan, D.P. Herzog, 'Determination of Atrazine in Water by Magnetic Particle Immunoassay: Collaborative Study. *J. AOAC Int.*, **79**, 529–537 (1996).
59. W.L. Pengelley, 'Validation of Immunoassays', in *Plant Growth Substances*, ed. M. Bopp, Springer Verlag, Berlin, 35–43, 1985.
60. E.M. Thurman, M. Meyer, M. Pomes, C.A. Perry, P.A. Schwab, 'Enzyme-linked Immunosorbent Assay Compared with Gas Chromatography/Mass Spectrometry for the Determination of Triazine Herbicides in Water', *Anal. Chem.*, **62**, 2043–2048 (1990).
61. K.A. Mountfort, S.L. Reynolds, S.A. Thorpe, S.N. White, 'Comparison of ELISA and HPLC Techniques for the Analysis of Carbendazim and Thiabendazole Residues in Fruits and Vegetables', *Food Agric. Immunol.*, **6**, 17–22 (1994).
62. L. Lavin, B.S. Young, T.D. Spittler, 'Analysis of Benomyl Residues in Commodities by Enzyme Immunoassay', in *Residue Analysis in Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 150–166, 1996.
63. A. Dankwardt, E.M. Thurman, B. Hock, 'Terbuthylazine and Deethyl-Terbuthylazine in Rain and Surface Water–Determination by Enzyme Immunoassay and Gas Chromatography/Mass Spectrometry', *Acta Hydrochim. Hydrobiol.*, **25**, 5–10 (1997).
64. R.O. Harrison, M.H. Goodrow, B.D. Hammock, 'Competitive Inhibition ELISA for the s-Triazine Herbicides: Assay Optimization and Antibody Characterization', *J. Agric. Food Chem.*, **39**, 122–128 (1991).
65. C. Keuchel, L. Weil, R. Niessner, 'Enzyme-linked Immunosorbent Assay for the Determination of 2,4,6-Trinitrotoluene and Related Nitroaromatic Compounds', *Anal. Sci.*, **8**, 9–12 (1992).
66. C. Wittmann, B. Hock, 'Development of an Enzyme Linked Immunoassay for the Analysis of the Atrazine Metabolite Hydroxyatrazine', *Acta Hydrochim. Hydrobiol.*, **22**, 60–69 (1994).
67. J. Seifert, 'Baker RaPID Assay–Ein ELISA zur Bestimmung von Pestiziden mit magnetischen Partikeln als fester Phase', *Gewässerschutz-Wasser-Abwasser*, **134**, 129–144 (1992).
68. A. Dankwardt, J. Seifert, B. Hock, 'Magnetpartikelenzymimmunoassay als schnelle Screening-Methode zur Bestimmung von Atrazin in Umweltproben', *Acta Hydrochim. Hydrobiol.*, **21**, 110–113 (1993).
69. D.S. Aga, E.M. Thurman, M.L. Pomes, 'Determination of Alachlor and its Sulfonic Acid Metabolite in Water by Solid-phase Extraction and Enzyme-linked Immunosorbent Assay', *Anal. Chem.*, **66**, 1495–1499 (1994).
70. T. Ruppert, L. Weil, R. Niessner, 'Influence of Water Contents on an Enzyme Immunoassay for Triazine Herbicides', *Vom Wasser*, **78**, 387–401 (1992).
71. C.S. Hottenstein, S.W. Jourdan, M.C. Hayes, F.M. Rubio, D.P. Herzog, T.S. Lawruk, 'Determination of Pentachlorphenol in Water and Soil by a Magnetic Particle-based Enzyme Immunoassay', *Environ. Sci. Technol.*, **29**, 2754–2758 (1995).
72. S.W. Jourdan, A.M. Scutellaro, J.R. Fleeker, D.P. Herzog, F.M. Rubio, 'Determination of Carbofuran in Water and Soil by a Rapid Magnetic Particle-based ELISA', *J. Agric. Food Chem.*, **43**, 2784–2788 (1995).
73. A. Dankwardt, B. Hock, R. Simon, D. Freitag, A. Kettrup, 'Determination of Non-extractable Triazine Residues by Enzyme Immunoassay: Investigation of Model Compounds and Soil Fulvic and Humic Acids', *Environ. Sci. Technol.*, **30**, 3493–3500 (1996).
74. J. Skerritt, B.E.A. Rani, 'Detection and Removal of Sample Matrix Effects in Agrochemical Immunoassays', in *Residue Analysis in Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 29–43, 1996.
75. A.S. Hill, J.V. Mei, C. Yin, B.S. Ferguson, J.H. Skerritt, 'Detection of the Insect Growth Regulator Methoprene in Wheat Grain and Milling Fractions Using an EIA', *J. Agric. Food Chem.*, **39**, 1882–1886 (1991).
76. P. Nugent, 'Enzyme-linked Competitive Immunoassay', in *Emerging Strategies for Pesticide Analysis*, eds. T. Cairns, J. Sherma, CRC Press, Boca Raton, FL, 247–258, 1992.
77. P. Schneider, B.D. Hammock, 'Influence of the ELISA Format and the Hapten Enzyme Conjugate on the Sensitivity of an Immunoassay for s-Triazine Herbicides Using Monoclonal Antibodies', *J. Agric. Food Chem.*, **40**, 525–530 (1992).
78. J.M. Francis, D.H. Craston, 'Immunoassay for Parathion Without its Removal from Solution in Hexane', *Analyst*, **119**, 1801–1805 (1994).
79. A. Marx, T. Giersch, B. Hock, 'Immunoaffinity Chromatography of s-Triazines', *Anal. Lett.*, **28**, 267–278 (1995).
80. V. Pichon, L. Chen, M.-C. Hennion, 'On-line Preconcentration and Liquid Chromatographic Analysis of Phenylurea Herbicides in Environmental Water Using a Silica-based Immunosorbent', *Anal. Chim. Acta*, **311**, 429 (1995).
81. J.W. King, K.-S. Nam, 'Coupling Enzyme Immunoassays with Supercritical Fluid Extraction', in *Residue Analysis in Food Safety*, eds. R.C. Beier, L.H. Stanker, ACS Symposium Series 621, American Chemical Society, Washington, DC, 422–438, 1996.
82. C. Keuchel, L. Weil, R. Niessner, 'Development of an Enzyme Immunoassay for the Determination of 2,4,6-Trinitrotoluene – Probing the Influence of Humic Acids', *SPIE Proc. Ser.*, **1716**, 44–50 (1992).
83. J.A. Itak, E.G. Olson, J.R. Fleeker, D.P. Herzog, 'Validation of a Paramagnetic Particle-based ELISA for the Quantitative Determination of Carbaryl in Water', *Bull. Environ. Contam. Toxicol.*, **51**, 260–267 (1993).

84. P.C.C. Feng, S.R. Horton, C.R. Sharp, 'A General Method for Developing Immunoassays to Chloroacetanilide Herbicides', *J. Agric. Food Chem.*, **40**, 211–214 (1992).
85. T.S. Lawruk, C.S. Hottenstein, D.P. Herzog, F.M. Rubio, 'Quantification of Alachlor in Water by a Novel Magnetic Particle-based ELISA', *Bull. Environ. Contam. Toxicol.*, **48**, 643 (1992).
86. C.R. Sharp, P.C.C. Feng, S.R. Horton, E.W. Logusch, 'Development of Highly Specific Antibodies to Alachlor by Use of a Carboxy-alachlor Protein Conjugate', in *Pesticide Residue and Food Safety*, ed. B.G. Tweedy, American Chemical Society, Washington, DC, 87–95, 1991.
87. F. Jung, A. Székács, Q. Li, B.D. Hammock, 'Immunochemical Approach to the Detection of Aminotriazoles Using Selective Amino Group Protection by Chromophores', *J. Agric. Food Chem.*, **39**, 129–136 (1991).
88. G.W. Aherne, A. Hardcastle, P. Saleem, N. England, 'Enhanced Chemiluminescent Immunoassays for Environmental Monitoring', in *Bioluminescence and Chemiluminescence, Current Status*, eds. P.E. Stanley, L.J. Kricka, John Wiley and Sons, New York, 91–98, 1990.
89. R.J. Bushway, B. Perkins, S.A. Savage, S.J. Lekousi, B.S. Ferguson, 'Determination of Atrazine Residues in Water and Soil by Enzyme Immunoassay', *Bull. Environ. Contamin. Toxicol.*, **40**, 647–654 (1988).
90. B. Dunbar, B. Riggle, G. Niswender, 'Development of Enzyme Immunoassay for the Detection of Triazine Herbicides', *J. Agric. Food Chem.*, **38**, 433–437 (1990).
91. S.J. Huber, 'Improved Solid-phase Enzyme Immunoassay Systems in the ppt Range for Atrazine in Fresh Water', *Chemosphere*, **14**, 1795–1803 (1985).
92. A.E. Karu, R.O. Harrison, D.J. Schmidt, C.E. Clarkson, J. Grassman, M.H. Goodrow, A. Lucas, B.D. Hammock, J.M. van Emon, R.J. White, 'Monoclonal Immunoassay for Triazine Herbicides', in *Immunoassays for Trace Chemical Analysis*, eds. M. Vanderlaan, L.H. Stanker, B.E. Watkins, D.W. Roberts, ACS Symposium Series 451, American Chemical Society, Washington, DC, 59–77, 1991.
93. A.D. Lucas, P. Schneider, R.O. Harrison, J.N. Seiber, B.D. Hammock, J.W. Biggar, D.E. Rolston, 'Determination of Atrazine and Simazine in Water and Soil Using Polyclonal and Monoclonal Antibodies in Enzyme-linked Immunosorbent Assay', *Food Agric. Immunol.*, **3**, 155–167 (1991).
94. F.M. Rubio, J.A. Itak, A.M. Scutellaro, M.Y. Selisker, D.P. Herzog, 'Performance Characteristics of a Novel Magnetic-particle-based Enzyme-linked Immunosorbent Assay for the Quantitative Analysis of Atrazine and Related Triazines in Water Samples', *Food Agric. Immunol.*, **3**, 113–125 (1991).
95. J.-M. Schlaeppi, W. Föry, K.J. Ramsteiner, 'Hydroxyatrazine and Atrazine Determination in Soil and Water by Enzyme Linked Immunosorbent Assay Using Specific Monoclonal Antibodies', *J. Agric. Food Chem.*, **37**, 1532–1538 (1989).
96. S. Wuest, B. Hock, 'A Sensitive Enzyme Immunoassay for the Detection of Atrazine Based Upon Sheep Antibodies', *Anal. Lett.*, **25**, 1025–1037 (1992).
97. Q.X. Li, B.D. Hammock, J.N. Seiber, 'Development of an Enzyme-linked Immunosorbent Assay for the Herbicide Bentazon', *J. Agric. Food Chem.*, **39**, 1537–1544 (1991).
98. H.K.M. Bekheit, A.D. Lucas, F. Szurdoki, S.J. Gee, B.D. Hammock, 'An Enzyme Immunoassay for the Environmental Monitoring of the Herbicide Bromacil', *J. Agric. Food Chem.*, **41**, 2220–2227 (1993).
99. F. Szurdoki, H.K.M. Bekheit, M.-P. Marco, M.H. Goodrow, B.D. Hammock, 'Synthesis of Haptens and Conjugates for an Enzyme Immunoassay for Analysis of the Herbicide Bromacil', *J. Agric. Food Chem.*, **40**, 1459–1465 (1992).
100. M.T. Muldoon, R.-N. Huang, C.J. Hapeman, G.F. Fries, M.C. Ma, J.O. Nelson, 'Hapten Synthesis and Immunoassay Development for the Analysis of Chlorodiamino-s-Triazine in Treated Pesticide Waste and Rinsate', *J. Agric. Food Chem.*, **42**, 747–755 (1994).
101. M.M. Kelley, E.W. Zahnow, W.C. Petersen, S.T. Toy, 'Chlorsulfuron Determination in Soil Extracts by Enzyme Immunoassay', *J. Agric. Food Chem.*, **33**, 962–965 (1985).
102. R.V. Dargar, J.M. Tymonko, P. VanDerWerf, 'Clomazone Measurement by Enzyme-linked Immunosorbent Assay', *J. Agric. Food Chem.*, **39**, 813–819 (1991).
103. F.K. Koppatschek, R.A. Liebl, A.L. Kriz, L.L. Melhado, 'Development of an Enzyme-linked Immunosorbent Assay for the Detection of the Herbicide Clomazone', *J. Agric. Food Chem.*, **38**, 1519–1522 (1990).
104. T.S. Lawruk, C.E. Lachmann, S.W. Jourdan, J.R. Fleeker, D.P. Herzog, F.M. Rubio, 'Quantification of Cyanazine in Water and Soil by a Magnetic Particle-based ELISA', *J. Agric. Food Chem.*, **41**, 747–752 (1993).
105. K.M. Robotti, J.K. Sharp, P.R. Ehrmann, L.J. Brown, B.W. Hermann, 'An ELISA Method for the Detection of Cyanazine', Abstracts of Papers Presented at the 192nd ACS National Meeting, American Chemical Society, Washington, DC, AGRO 42, 1986.
106. M.J. Wraith, D.W. Britton, 'Immunochemical Methods for Pesticide Residue Analysis', Brighton Crop Protection Conference, British Crop Protection Council, London, 131–137, 1988.
107. C. Wittmann, B. Hock, 'Development of an ELISA for the Analysis of Atrazine Metabolites Deethylatrazine and Deisopropylatrazine', *J. Agric. Food Chem.*, **39**, 1194–1200 (1991).
108. M. Schwalbe, E. Dorn, K. Beyermann, 'Enzyme Immunoassay and Fluoroimmunoassay for the Herbicide Diclofop-methyl', *J. Agric. Food Chem.*, **32**, 734–741 (1984).

109. R. Fleeker, R.O. Mumma, 'Development of an Enzyme Linked Immunoassay for Aldicarb', Abstract of Papers Presented at the 194th ACS National Meeting, American Chemical Society, Washington, DC, AGRO 146, 1987.
110. M. Franek, V. Kolar, M. Granatova, Z. Nevorankova, 'Monoclonal ELISA for 2,4-Dichlorophenoxyacetic Acid: Characterization of Antibodies and Assay Optimization', *J. Agric. Food Chem.*, **42**, 1369–1374 (1994).
111. J.C. Hall, R.J.A. Deschamps, K.K. Krieg, 'Immunoassays for the Detection of 2,4-D and Picloram in River Water and Urine', *J. Agric. Food Chem.*, **37**, 981–984 (1989).
112. D. Knopp, P. Nuhn, H.-J. Dobberkau, 'Radioimmunoassay for 2,4-Dichlorophenoxyacetic Acid', *Arch. Toxicol.*, **58**, 27–32 (1985).
113. Y.V. Lukin, I.M. Dekuchaev, I.M. Polyakl, S.A. Eremin, 'Detection of 2,4-Dichlorophenoxyacetic Acid by Microtiter Particle Agglutination Inhibition Test and Polarization Fluoroimmunassay', *Anal. Lett.*, **27**, 2973–2982 (1994).
114. D.F. Rinder, J.R. Fleeker, 'A Radioimmunoassay to Screen for 2,4-Dichlorophenoxyacetic Acid and 2,4,5-Trichlorophenoxyacetic Acid in Surface Water', *Bull. Environ. Contam. Toxicol.*, **26**, 375–380 (1981).
115. A. Dzgoev, M. Mecklenburg, B. Xie, A. Miyabayashi, P.O. Larsson, B. Danielsson, 'Optimization of a Charge Coupled Device Imaging Enzyme Linked Immunosorbent Assay and Supports for the Simultaneous Determination of Multiple 2,4-D Samples', *Anal. Chim. Acta*, **347**, 87–93 (1997).
116. F.G. Sanchez, A. Navas, F. Alonso, J. Lovillo, 'Polarization Fluoroimmunoassay of the Herbicide Dichlorprop', *J. Agric. Food Chem.*, **41**, 2215 (1993).
117. A.E. Karu, M.H. Goodrow, D.J. Schmidt, B.D. Hammock, M.W. Bigelow, 'Synthesis of Haptens and Derivation of Monoclonal Antibodies for Immunoassays of the Phenylurea Herbicide Diuron', *J. Agric. Food Chem.*, **42**, 301 (1994).
118. N. Lee, J.H. Skerritt, M. Thomas, W. Korth, K.H. Bowmer, K.A. Larkin, B.S. Ferguson, 'Quantification of the Urea Herbicide, Diuron, in Water by Enzyme Immunoassay', *Bull. Environ. Contam. Toxicol.*, **55**, 479–486 (1995).
119. R.J. Bushway, L.E. Katz, L.B. Perkins, A.W. Reed, T.S. Fan, B.S. Young, 'Analysis of Hexazinone in Soil by Enzyme-linked Immunosorbent Assay', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, Washington, DC, 303–311, 1997.
120. J. Mangler, M.G. Weller, L. Weil, R. Niessner, H. Hämmerle, B. Schloßhauer, 'New Monoclonal Antibodies to Triazine Herbicides', *Fresenius' J. Anal. Chem.*, **349**, 346–348 (1994).
121. A.D. Lucas, H.K.M. Bekheit, M.H. Goodrow, A.D. Jones, S. Kullman, F. Matsumura, J.E. Woodrow, J.N. Seiber, B.D. Hammock, 'Development of Antibodies Against Hydroxyatrazine and Hydroxysimazine: Applications to Environmental Samples', *J. Agric. Food Chem.*, **41**, 1523–1529 (1993).
122. W.H. Newsome, P.G. Collins, 'Determination of Imazametabenz in Cereal Grain by Enzyme-linked Immunosorbent Assay', *Bull. Environ. Contam. Toxicol.*, **47**, 211 (1991).
123. R.B. Wong, Z.H. Ahmed, 'Development of an Enzyme-linked Immunosorbent Assay for Imazaquin Herbicide', *J. Agric. Food Chem.*, **40**, 811–816 (1992).
124. M.F. Katmeh, G. Frost, W. Aherne, D. Stevenson, 'Development of an Enzyme-linked Immunosorbent Assay for Isoproturon in Water', *Analyst*, **119**, 431 (1994).
125. E. Liegeois, Y. Dehon, B. Debrabant, P. Perry, D. Portetelle, A. Copin, 'ELISA Test, a New Method to Detect and Quantify Isoproturon in Soil', *Sci. Total Environ.*, **123**, 17–28 (1992).
126. J.M. Matt, K.A. Larkin, R.O. Harrison, 'Determination of Isoproturon in Ground and Surface Water by Enzyme Immunoassay', Book of Abstracts, Eighth IUPAC International Congress of Pesticide Chemistry, Washington, DC, Abstract 47, 1994.
127. R.O. Harrison, A.A. Brimfield, J.O. Nelson, 'Development of a Monoclonal Antibody Based Enzyme Immunoassay for Maleic Hydrazide', *J. Agric. Food Chem.*, **37**, 958–964 (1989).
128. W. Weber, K. Rubach, 'Entwicklung von Enzymimmunoassays zum Nachweis von Phenoxycarbonsäure–Herbiziden in Trink- und Grundwässern', *Acta Hydrochim. Hydrobiol.*, **2**, 53–59 (1994).
129. H. Scholz, B. Hock, 'Development of an Enzyme Immunoassay for the Determination of Metozachlor', *Z. Wasser-Abwasser-Forsch.*, **24**, 29–31 (1991).
130. S. Kreissig, B. Hock, R. Stoecker, 'An Enzyme Immunoassay for the Determination of Methabenzthiazuron', *Anal. Lett.*, **24**, 1729–1739 (1991).
131. J.-M. Schlaeppi, H. Moser, K. Ramsteiner, 'Determination of Metolachlor by Competitive Enzyme Immunoassay Using a Specific Monoclonal Antibody', *J. Agric. Food Chem.*, **39**, 1553 (1991).
132. T.S. Lawruk, C.E. Lachmann, S.W. Jourdan, J.R. Fleeker, D.P. Herzog, F.M. Rubio, 'Determination of Metolachlor in Water and Soil by a Rapid Magnetic Particle-based ELISA', *J. Agric. Food Chem.*, **41**, 1426–1431 (1993).
133. P.C.C. Feng, S.R. Horton, C.R. Sharp, 'A General Method for Developing Immunoassays to Chloroacetanilide Herbicides', *J. Agric. Food Chem.*, **40**, 211–214 (1992).
134. S.J. Gee, T. Miyamoto, M.H. Goodrow, D. Buster, B.D. Hammock, 'Development of an Enzyme-linked Immunosorbent Assay for the Analysis of the Thiocarbamate Herbicide Moliate', *J. Agric. Food Chem.*, **36**, 863–870 (1988).

135. F. Jung, S.H. Gee, R.O. Harrison, M.H. Goodrow, A.E. Karu, Q.X. Li, A.L. Braun, B.D. Hammock, 'Use of Immunochemical Techniques for the Analysis of Pesticides', *Pesticide Sci.*, **26**, 303–317 (1989).
136. W.H. Newsome, P.G. Collins, 'Development of an ELISA for Urea Herbicides in Foods', *Food Agric. Immunol.*, **2**, 75–84 (1990).
137. B. Riggle, B.D. Dunbar, 'Development of Enzyme Immunoassay for the Detection of the Herbicide Norflurazon', *J. Agric. Food Chem.*, **38**, 1922–1925 (1990).
138. M.R. Bowles, D.W. Eyles, C.G.M. Hampson, S.M. Pond, 'Quantitation of Paraquat in Biological Samples by Radioimmunoassay Using a Monoclonal Antibody', *Fundam. Appl. Toxicol.*, **19**, 375–379 (1992).
139. R.E. Coxon, C. Rae, G. Gallacher, J. Landon, 'Development of a Simple Fluoroimmunoassay for Paraquat', *Clin. Chim. Acta*, **175**, 297–306 (1988).
140. D. Fatori, W.M. Hunter, 'Radioimmunoassay for Serum Paraquat', *Clin. Chim. Acta*, **100**, 81–90 (1980).
141. B.S. Ferguson, D.E. Kelsey, T.S. Fan, R.J. Bushway, 'Pesticide Testing by Enzyme Immunoassay at Trace Levels in Environmental and Agricultural Samples', *Sci. Total Environ.*, **132**, 415–428 (1993).
142. T. Levitt, 'Radioimmunoassay for Paraquat', *Lancet*, **307**(8033), 358 (1977).
143. M. Nagao, T. Takatori, B. Wu, K. Terazawa, H. Gotouda, H. Akabane, 'Development and Characterization of Monoclonal Antibodies Reactive with Paraquat', *J. Immunoassay*, **10**, 1 (1989).
144. Z. Niewola, C. Hayward, B.A. Symington, R.T. Robson, 'Quantitative Estimation of Paraquat by an Enzyme Linked Immunosorbent Assay Using a Monoclonal Antibody', *Clin. Chim. Acta*, **148**, 149 (1985).
145. M.Y. Selisker, D.P. Herzog, R. Erber, J.R. Fleeker, J.A. Itak, 'Determination of Paraquat in Fruits and Vegetables by a Magnetic Particle Based Enzyme-linked Immunosorbent Assay', *J. Agric. Food Chem.*, **43**, 544–547 (1995).
146. C.A. Spinks, B. Wang, E.N.C. Mills, M.R.A. Morgan, 'Production and Characterization of Monoclonal Anti-idiotype Antibody Mimics for the Pyrethroid Insecticides and the Herbicide Paraquat', *Food Agric. Immunol.*, **5**, 13 (1993).
147. J. Van Emon, B.D. Hammock, J.N. Seiber, 'Enzyme-linked Immunosorbent Assay for Paraquat and its Application to Exposure Analysis', *Anal. Chem.*, **58**, 1866–1873 (1986).
148. R.J.A. Deschamps, J.C. Hall, M.R. McDermott, 'Polyclonal and Monoclonal Enzyme Immunoassays for Picloram Detection in Water, Soil, Plants and Urine', *J. Agric. Food Chem.*, **38**, 1881 (1990).
149. J.C. Hall, R.J.A. Deschamps, K.K. Krieg, 'Immunoassays for the Detection of 2,4-D and Picloram in River Water and Urine', *J. Agric. Food Chem.*, **37**, 981–984 (1989).
150. S.A. Eremin, J.V. Samsonova, 'Development of Polarization Fluoroimmunoassay for the Detection of s-Triazine Herbicides', *Anal. Lett.*, **27**, 3013–3025 (1994).
151. M. Wortberg, M.H. Goodrow, S.J. Gee, B.D. Hammock, 'Immunoassay for Simazine and Atrazine with Low Cross-reactivity for Propazine', *J. Agric. Food Chem.*, **44**, 2210–2219 (1996).
152. T. Giersch, K. Kramer, B. Hock, 'Optimization of a Monoclonal Antibody-based Enzyme Immunoassay for the Detection of Terbuthylazine', *Sci. Total Environ.*, **132**, 435–448 (1993).
153. T. Giersch, K. Kramer, M.G. Weller, B. Hock, 'Improvement of a Monoclonal Antibody-based Immunoassay for the Determination of Terbutryn', *Acta Hydrochim. Hydrobiol.*, **21**, 312–315 (1993).
154. S.J. Huber, B. Hock, 'A Solid-phase Enzyme Immunoassay for Quantitative Determination of the Herbicide Terbutryn', *J. Plant Disease Protection*, **92**, 147–156 (1985).
155. P.Y.K. Cheung, S.J. Gee, B.D. Hammock, 'Pesticide Immunoassay as a Biotechnology', in *The Impact of Chemistry on Biotechnology*, eds. M. Philips, R.M. Shoemaker, R. Ottenbrite, ACS Symposium Series 362, American Chemical Society, Washington, DC, 217, 1988.
156. R. Ghildyal, M. Kariofillis, 'Polyclonal Antibody-based ELISA for Triasulfuron', *Bull. Environ. Contam. Toxicol.*, **54**, 647 (1995).
157. J.-M. Schlaeppi, W. Meyer, K.A. Ramsteiner, 'Determination of Triasulfuron in Soil by Monoclonal Antibody-based Enzyme Immunoassay', *J. Agric. Food Chem.*, **40**, 1093–1098 (1992).
158. D.F. Rinder, J.R. Fleeker, 'A Radioimmunoassay to Screen for 2,4-Dichlorophenoxyacetic Acid and 2,4,5-Trichlorophenoxyacetic Acid in Surface Water', *Bull. Environ. Contam. Toxicol.*, **26**, 375 (1981).
159. B. Riggle, 'Development of a Preliminary Enzyme-linked Immunosorbent Assay for the Herbicide Trifluralin', *Bull. Environ. Contam. Toxicol.*, **46**, 404 (1991).
160. J.F. Brady, J.R. Fleeker, R.A. Wilson, R.O. Mumma, 'Enzyme-immunoassay for Aldicarb', in *Biological Monitoring of Pesticide Exposure, Measurement, Estimation and Risk Reduction*, eds. R.G.M. Wang, C.A. Franklin, R.C. Honeycut, J.C. Reinhert, ACS Symposium Series 382, American Chemical Society, Washington DC, 262–284, 1989.
161. J.A. Itak, M.Y. Selisker, D.P. Herzog, 'Development and Evaluation of a Magnetic Particle Based Enzyme Immunoassay for Aldicarb, Aldicarb Sulfone and Aldicarb Sulfoxide', *Chemosphere*, **24**, 11–21 (1992).
162. J.L. Langone, H. Van Vunakis, 'Radioimmunoassay for Dieldrin and Aldrin', *Res. Commun. Chem. Pathol. Pharmacol.*, **10**, 163 (1975).
163. J.V. Mercader, A. Montoya, 'A Monoclonal Antibody-based ELISA for the Analysis of Azinphos-methyl in Fruit Juices', *Anal. Chim. Acta*, **347**, 95–101 (1997).

164. S. Pullen, B. Hock, 'Development of Enzyme Immunoassay for the Detection of Pyrethroid Insecticides 1. Monoclonal Antibodies for Allethrin', *Anal. Lett.*, **28**, 765–779 (1995).
165. S. Pullen, B. Hock, 'Development of Enzyme Immunoassay for the Detection of Pyrethroid Insecticides 2. Polyclonal Antibodies for Pyrethroid Insecticides', *Anal. Lett.*, **28**, 781–795 (1995).
166. K.D. Wing, B.D. Hammock, 'Stereoselectivity of a Radioimmunoassay for the Insecticide S-bioallethrin', *Experientia*, **35**, 1619–1620 (1979).
167. J.M. Van Emon, J.N. Seiber, B.D. Hammock, 'Applications of Immunoassay to Paraquat and other Pesticides', in *Bioregulators for Pest Control*, ed. P.A. Hedin, ACS Symposium Series 76, American Chemical Society, Washington, DC, 307–316, 1985.
168. A.S. Hill, D.P. McAdam, S.L. Edward, J.H. Skerritt, 'Quantitation of Bioresmethrin, a Synthetic Pyrethroid Grain Protectant, by Enzyme Immunoassay', *J. Agric. Food Chem.*, **41**, 2011 (1993).
169. A. Abad, A. Montoya, 'Production of Monoclonal Antibodies for Carbaryl from a Hapten Preserving the Carbamate Group', *J. Agric. Food Chem.*, **42**, 1818–1823 (1994).
170. M.P. Marco, S.J. Gee, H.M. Cheng, Z.Y. Liang, B.D. Hammock, 'Development of an ELISA for Carbaryl', *J. Agric. Food Chem.*, **41**, 423–430 (1993).
171. B.D. Banerjee, 'Development of an Enzyme-linked Immunosorbent Assay for the Quantification of DDA (2,2-bis (*p*-Chlorophenyl) Acetic Acid) in Urine', *Bull. Environ. Contam. Toxicol.*, **38**, 798 (1987).
172. A. Abad, J.J. Manclus, F. Mojarrad, J.V. Mercader, M.A. Miranda, J. Primo, V. Guardiola, A. Montoya, 'Hapten Synthesis and Production of Monoclonal Antibodies to DDT and Related Compounds', *J. Agric. Food Chem.*, **45**, 3694–3702 (1997).
173. S.I. Wie, B.D. Hammock, 'Comparison of Coating and Immunizing Antigen Structure on the Sensitivity and Specificity of Immunoassays for Benzoylphenylurea Insecticides', *J. Agric. Food Chem.*, **32**, 1294–1301 (1984).
174. R.M. Dreher, B. Podratzki, 'Development of an Enzyme Immunoassay for Endosulfan and its Degradation Products', *J. Agric. Food Chem.*, **36**, 1072–1075 (1988).
175. N. Lee, J.H. Skerritt, D.P. McAdam, 'Hapten Synthesis and Development of ELISAs for Detection of Endosulfan in Water and Soil', *J. Agric. Food Chem.*, **43**, 1730 (1995).
176. B. Reck, J. Frevert, 'Competitive-and Inhibition-type Immunoassay for Determination of Endosulfan', in *Immunochemical Methods for Environmental Analysis*, eds. J.M. Van Emon, R.O. Mumma, ACS Symposium Series 442, American Chemical Society, Washington, DC, 193–198, 1990.
177. D.P. McAdam, A.S. Hill, H.L. Beasley, J H. Skerritt, 'Mono- and Polyclonal Antibodies to the Organophosphate Fenitrothion. 1. Approaches to Hapten–Protein Conjugation', *J. Agric Food Chem.*, **40**, 1466–1470 (1992).
178. A.S. Hill, J.H. Skerritt, R.J. Bushway, W. Pask, K.A. Larkin, M. Thomas, W. Korth, K. Bowmer, 'Development and Application of Laboratory and Field Immunoassays for Chlorpyrifos in Water and Soil Matrices', *J. Agric. Food Chem.*, **42**, 2051–2058 (1994).
179. L.H. Stanker, B. Watkins, M. Vanderlaan, R. Ellis, J. Rajan, 'Analysis of Heptachlor and Related Cyclodiene Insecticides in Food Products', in *Immunoassays for Trace Chemical Analysis*, eds. M. Vanderlaan, L.H. Stanker, B.E. Watkins, D.W. Roberts, ACS Symposium Series No. 451, American Chemical Society, Washington, DC, 108–123, 1990.
180. J.V. Mei, C.-M. Yin, L.A. Carpino, B.S. Ferguson, 'An Enzyme-linked Immunosorbent Assay for Residue Detection of Methoprene', *J. Agric. Food Chem.*, **39**, 2083–2090 (1991).
181. P.M. Krämer, M.-P. Marco, B.D. Hammock, 'Development of a Selective Enzyme-linked Immunosorbent Assay for 1-Naphthol – the Major Metabolite of Carbaryl (1-Naphthyl *N*-Methylcarbamate)', *J. Agric. Food Chem.*, **42**, 934–943 (1994).
182. C.D. Ercegovich, R.P. Vallejo, R.R. Gettig, L. Woods, E.R. Bogus, R.O. Mumma, 'Development of a Radioimmunoassay for Parathion', *J. Agric. Food Chem.*, **29**, 559–563 (1981).
183. B.S. Ferguson, K.A. Larkin, 'Monitoring Parathion and Parathion-methyl in Ground, Surface and Rice Paddy Water with a Rapid and Sensitive Immunoassay', in Book of Abstracts, Eighth IUPAC International Congress of Pesticide Chemistry, Washington DC, Abstract 43, 1994.
184. J.M. Wong, Q.X. Li, B.D. Hammock, J.N. Seiber, 'Method for the Analysis of 4-Nitrophenol and Parathion in Soil Using Supercritical Fluid Extraction and Immunoassay', *J. Agric. Food Chem.*, **39**, 1802–1807 (1991).
185. A.A. Brimfield, D.E. Lenz, C. Graham, K.W. Hunter, 'Mouse Monoclonal Antibodies Against Paroxon: Potential Reagents for Immunoassay with Constant Immunochemical Characteristics', *J. Agric. Food Chem.*, **33**, 1237–1242 (1985).
186. E. Heldman, A. Balan, O. Horowitz, S. Ben-Zion, M. Torten, 'A Novel Immunoassay with Direct Relevance to Protection Against Organophosphate Poisoning', *FEBS Lett.*, **180**, 243–248 (1985).
187. K.W. Hunter, D.E. Lenz, 'Detection and Quantification of the Organophosphate Insecticide Paroxon by Competitive Inhibition Enzyme Immunoassay', *Life Sci.*, **30**, 355–361 (1982).
188. L.T. Hall, J. Van Emon, V. Lopez-Avila, 'Development of Immunochemical Personal Exposure Monitors for

Pentachlorophenol', *Environ. Process Monitoring Technol.*, **1637**, 189–195 (1992).
189. J.P. Mapes, K.D. McKenzie, L.R. McClelland, S. Movassaghi, R.A. Reddy, R.L. Allen, S.B. Friedman, 'Penta RISC™ Soil–a Rapid, on-site Screening Test for Pentachlorophenol in Soil', *Bull. Environ. Contam. Toxicol.*, **49**, 334–341 (1992).
190. G.A. Bonwick, M. Putman, P.J. Baugh, C.J. Smith, R. Armitage, D. Davies, 'Immunoassay Development for Permethrin Residues', *Food Agric. Immunol.*, **6**, 341 (1994).
191. L.H. Stanker, C. Bigbee, J. Van Emon, B. Watkins, R.H. Jensen, C. Morris, M. Vanderlaan, 'An Immunoassay for Pyrethroids: Detection of Permethrin in Meat', *J. Agric. Food Chem.*, **37**, 834–839 (1989).
192. J.H. Skerritt, A.S. Hill, D.P. McAdam, L.H. Stanker, 'Analysis of the Synthetic Pyrethroids Permethrin and 1(R)Phenothrin in Grain Using a Monoclonal Antibody-based Test', *J. Agric. Food Chem.*, **40**, 1287–1292 (1992).
193. J.H. Skerritt, A.S. Hill, H.L. Beasley, S.L. Edward, D.P. McAdam, 'Enzyme-immunoassay for Quantitation of the Organophosphate Pesticides, Fenitrothion, Chlorpyrifos-methyl and Pirimifos-methyl in Wheat Grain and Flour-milling Fractions', *J. Assoc. Off. Anal. Chem.*, **75**, 519–528 (1992).
194. R.J. Robins, M.R.A. Morgan, M.J.C. Rhodes, J.M. Furze, 'Determination of Quassin in Picogram Quantities by an Enzyme-linked Immunosorbent Assay', *Phytochemistry*, **23**, 1119–1123 (1984).
195. J.A. Itak, W.A. Day, A. Montoya, J.J. Manclus, A.M. Phillips, D.A. Lindsay, D.P. Herzog, 'A Paramagnetic Particle-based Enzyme-linked Immunosorbent Assay for the Quantitative Determination of 3,5,6-Trichloro-2-pyridinol in Water', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, American Chemical Society, Washington, DC, 261–270, 1997.
196. D.A. Fitzpatrick, 'An Immunoassay for the Determination of Benomyl in Produce as MBC', in Book of Abstracts, Eighth IUPAC International Congress of Pesticide Chemistry, Washington, DC, Abstract 45, 1994.
197. H.R. Lukens, C.B. Williams, S.A. Levison, W.B. Dendliker, D. Murayama, R.L. Baron, 'Fluorescence Immunoassay Technique for Detecting Organic Environmental Contaminants', *Environ. Sci. Technol.*, **11**, 292 (1977).
198. W.H. Newsome, J.B. Shields, 'A Radioimmunoassay for Benomyl and Methyl 2-Benzimidazol Carbamate in Food Crops', *J. Agric. Food Chem.*, **29**, 220–222 (1981).
199. J.A. Itak, M.Y. Selisker, S.W. Jourdan, J.R. Fleeker, D.P. Herzog, 'Determination of Benomyl (as Carbendazim) and Carbendazim in Water, Soil, and Fruit Juice by a Magnetic Particle-based Immunoassay', *J. Agric. Food Chem.*, **41**, 2329–2332 (1993).
200. W.H. Newsome, P.G. Collins, 'Enzyme-linked Immunosorbent Assay of Benomyl and Thiabendazole in Some Foods', *J. Assoc. Off. Anal. Chem.*, **70**, 1025–1027 (1987).
201. D.L. Brandon, R.G. Binder, A.H. Bates, W.C. Montague, Jr, 'Monoclonal Ab for Multiresidue ELISA of Benzimidazole in Liver', *J. Agric. Food Chem.*, **42**, 1588–1594 (1994).
202. W.H. Newsome, J.M. Yeung, P.G. Collins, 'Development of EIA for Captan and its Degradation Product Tetrahydrophthalimide in Foods', *J. AOAC Int.*, **76**, 381–386 (1993).
203. T.S. Lawruk, A.M. Gueco, S.W. Jourdan, J.R. Fleeker, D.P. Herzog, F.M. Rubio, 'Determination of Chlorothalonil in Water and Agricultural Products by Magnetic Particle-based Enzyme Immunoassay', *J. Agric. Food Chem.*, **41**, 747–752 (1993).
204. F. Jung, H.H. Meyer, R.T. Hamm, 'Development of an Enzyme-linked Immunoasorbent Assay for the Fungicide Fenpropimorph', *J. Agric. Food Chem.*, **37**, 1183–1187 (1989).
205. W.H. Newsome, 'Determination of Iprodione in Foods by ELISA', in *Pesticide Science and Biotechnology*, eds. R. Greenhalgh, R. Roberts, Blackwell Scientific, Oxford, 341–352, 1987.
206. W.H. Newsome, 'An ELISA for Metalaxyl in Foods', *J. Agric. Food Chem.*, **33**, 528–530 (1985).
207. A. Szekacs, B.D. Hammock, 'Development of an Enzyme-linked Immunosorbent Assay for the Detection of the Triazole Fungicide Myclobutanil', *J. Agric. Food Chem.*, **43**, 2083–2091 (1995).
208. D.L. Brandon, R.G. Binder, A.H. Bates, W.C. Montague, Jr, 'Monoclonal Ab Based ELISA for Thiabendazole in Liver', *J. Agric. Food Chem.*, **40**, 1722–1726 (1992).
209. W.H. Newsome, 'Development of an ELISA for Triadimefon in Foods', *Bull. Environ. Contamin. Toxicol.*, **36**, 9–14 (1986).
210. F. Forlani, A. Arnoldi, S. Pagani, 'Development of an Enzyme-linked Immunosorbent Assay for Triazole Fungicides', *J. Agric. Food Chem.*, **40**, 328–331 (1992).
211. R.J. Bushway, B. Perkins, L. Fukal, R.O. Harrison, B.S. Ferguson, 'Comparison of Enzyme-linked Immunosorbent Assay and High-performance Liquid Chromatography for the Analysis of Atrazine in Water from Czechoslovakia', *Arch. Environ. Contamin. Toxicol.*, **21**, 365–370 (1991).
212. K.J. Huber, L. Weil, R. Niessner, 'Total Deposition Monitoring of the Triazine Herbicide Atrazine by Use of an Enzyme-linked Immunosorbent Assay (ELISA)', *Fresenius' J. Anal. Chem.*, **343**, 146 (1992).
213. D.S. Aga, E.M. Thurman, 'Coupling Solid-phase Extraction and Enzyme-linked Immunosorbent Assay for Ultratrace Determination of Herbicides in Pristine Water', *Anal. Chem.*, **65**, 2894–2898 (1993).
214. C. Wittmannn, B. Hock, 'Evaluation and Performance Characteristics of a Novel ELISA for the Quantitative

Analysis of Atrazine in Water, Plants and Soil', *Food Agric. Immunol.*, **2**, 65–74 (1990).
215. J. Gascon, G. Durand, D. Barcelo, 'Pilot Survey for Atrazine and Total Chlorotriazines in Estuarine Waters Using Magnetic Particle Based Immunoassay and Gas Chromatography Nitrogen/Phosphorus Detection', *Environ. Sci. Technol.*, **29**, 1551–1556 (1995).
216. B. Gruessner, N.C. Shambaugh, M.C. Watzin, 'Comparison of an Enzyme Immunoassay and Gas Chromatography/Mass Spectrometry for the Detection of Atrazine in Surface Waters', *Environ. Sci. Technol.*, **28**, 251–254 (1995).
217. E.P. Meulenberg, P.G. Stoks, 'Water Quality Control in the Production of Drinking Water from River Water. The Application of Immunological Techniques for the Detection of Chlorphenoxy Acid Herbicides', *Anal. Chim. Acta*, **311**, 407–414 (1995).
218. J.P. Sherry, A. Borgmann, 'Enzyme-immunoassay Techniques for the Detection of Atrazine in Water Samples – Evaluation of a Commercial Tube Based Assay', *Chemosphere*, **26**, 2173–2184 (1993).
219. J.F. Brady, G.S. Lemasters, R.K. Williams, J.H. Pittman, J.P. Daubert, M.W. Cheung, D.H. Skinner, J. Turner, M.A. Rowland, J. Lange, S.M. Sobek, 'Immunoassay Analysis and Gas Chromatographic Confirmation of Atrazine Residues in Water Samples from a Field Study Conducted in the State of Wisconsin', *J. Agric. Food Chem.*, **43**, 268–274 (1995).
220. C. Mouvet, S. Broussard, R. Jeannot, C. Maciag, R. Abuknesha, G. Ismail, 'Validation of Commercially Available ELISA Microtiter Plates for Triazines in Water Samples', *Anal. Chim. Acta*, **311**, 331–340 (1995).
221. M.-P. Marco, S. Chiron, J. Gascon, B.D. Hammock, D. Barcelo, 'Validation of Two Immunoassay Methods for Environmental Monitoring of Carbaryl and 1-Naphthol in Groundwater Samples', *Anal. Chim. Acta*, **311**, 319–330 (1995).
222. R.J. Bushway, B. Perkins, S.A. Savage, S.J. Lekousi, B.S. Ferguson, 'Determination of Atrazine Residues in Water and Soil by Enzyme Immunoassay', *Bull. Environ. Contamin. Toxicol.*, **40**, 647–654 (1988).
223. A. Hahn, F. Frimmel, A. Haisch, G. Henkelmann, B. Hock, 'Immunolabelling of Atrazine Residues in Soil', *Z. Pflanzenernähr. Bodenk.*, **155**, 203–208 (1992).
224. R.J. Schneider, L. Weil, R. Niessner, 'Screening and Monitoring of Herbicide Behavior in Soils by Enzyme Immunoassays', *Int. J. Environ. Anal. Chem.*, **46**, 129–140 (1992).
225. A. Dankwardt, B. Hock, R. Simon, D. Freitag, A. Kettrup, 'Determination of Non-extractable Triazine Residues by Enzyme Immunoassay: Investigation of Model Compounds and Soil Fulvic and Humic Acids', *Environ. Sci. Technol.*, **30**, 3493–3500 (1996).
226. K.S. Goh, J. Hsu, D.J. Weaver, P.J. Stoddard, J. White, 'Development and Use of an Enzyme-linked Immunosorbent Assay to Monitor Compliance with Prohibited Use of Atrazine, Simazine, and Prometon in California Soils', in *New Frontiers in Agrochemical Immunoassay*, eds. D.A. Kurtz, J.H. Skerritt, L. Stanker, AOAC International, Arlington, VA, 31–38, 1995.
227. G.K. Stearman, M.J.M. Wells, 'Enzyme Immunoassay Microtiter Plate Response to Atrazine and Metolachlor in Potentially Interfering Matrices', *Bull. Environ. Contam. Toxicol.*, **51**, 588–595 (1993).
228. J.M.A. Schlaeppi, A. Kessler, W. Föry, 'Development of a Magnetic Particle-based Automated Chemiluminescent Immunoassay for Triasulfuron', *J. Agric. Food Chem.*, **42**, 1914–1919 (1994).
229. G.K. Stearman, M.J.M. Wells, S.M. Adkisson, T.E. Ridgill, 'Supercritical Fluid Extraction Coupled with Enzyme Immunoassay Analysis of Soil Herbicides', *Analyst*, **120**, 2617–2621 (1995).
230. A. Abad, A. Montoya, 'Application of a Monoclonal Antibody-based ELISA to the Determination of Carbaryl in Apple and Grape Juice', *Anal. Chim. Acta*, **311**, 365–370 (1995).
231. D.L. Brandon, R.G. Binder, A.H. Bates, W.C. Montague, Jr, 'Comparative ELISA for Thiabendazole Residues in Produce Using Indirect Immobilized Monoclonal Antibodies', *Food Agric. Immunol.*, **7**, 99–108 (1995).
232. J. Van Emon, J. Seiber, B. Hammock, 'Application of an ELISA to Determine Paraquat Residues in Milk, Beef and Potatoes', *Bull. Environ. Contamin. Toxicol.*, **39**, 490–497 (1987).
233. A.M.A. Ibrahim, A.A. Ragab, M.A. Morsy, M.M. Hewedi, C.J. Smith, 'Application of an Aldrin and Dieldrin ELISA to the Detection of Pesticides in Eggs', *Food Agric. Immunol.*, **6**, 39–44 (1994).
234. D. Knopp, 'Assessment of Exposure to 2,4-Dichlorophenoxyacetic Acid in the Chemical Industry: Results of a Five Year Biological Monitoring Study, *Occup. Environ. Med.*, **51**, 152 (1994).
235. A.D. Lucas, A.D. Jones, M.H. Goodrow, S.G. Saiz, C. Blewett, J.N. Seiber, B.D. Hammock, 'Determination of Atrazine Metabolites in Human Urine: Development of a Biomarker of Exposure', *Chem. Res. Toxicol.*, **6**, 107–116 (1993).
236. P.M. Krämer, Q.X. Li, B.D. Hammock, 'Integration of Liquid Chromatography with Immunoassay: An Approach Combining the Strength of Both Methods', *J. AOAC Int.*, **77**, 1275–1287 (1994).
237. J.F. Lawrence, C. Menard, M.C. Hennion, V. Pichon, F. Le Goffic, N. Durand, 'Use of Immunoaffinity Chromatography as a Simplified Cleanup Technique for the Liquid Chromatographic Determination of Phenylurea Herbicides in Plant Material', *J. Chromatogr., A*, **732**, 277–281 (1996).
238. M.F. Katmeh, A.J.M. Godfrey, D. Stevenson, G.W. Aherne, 'Enzyme Immunoaffinity Chromatography – A Rapid Semi-quantitative Immunoassay Technique for Screening the Presence of Isoproturon in Water Samples', *Analyst*, **122**, 481–486 (1997).

239. V. Pichon, L. Chen, M.-C. Hennion, N. Durand, F. Le Goffic, 'Selective Multiresidue Analysis of Pesticides in Surface Water Using Immunosorbents', 5th Symposium on Chemistry and Fate of Modern Pesticides, Paris, September 6–8, International Association of Environmental Analytical Chemistry, Abstract L-9, 1995.
240. M. de Frutos, F.E. Regnier, 'Tandem Chromatographic-immunological Analyses', *Anal. Chem.*, **65**, 17A–25A (1993).
241. A.D. Lucas, S.J. Gee, B.D. Hammock, 'Integration of Immunochemical Methods with Other Analytical Techniques for Pesticide Residue Determination', *J. AOAC Int.*, **78**, 585–591 (1995).
242. P.R. Ekins, F.W. Chu, E.M. Biggart, 'Multianalyte Immunoassay: The Immunological "Compact Disk" of the Future', *J. Clin. Immunoassay*, **13**, 169–181 (1990).
243. F.W. Chu, P.R. Edwards, R. Ekins, 'Microarray-based Immunoassays', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, American Chemical Society, Washington, DC, 170–184, 1997.
244. M.T. Muldoon, G.F. Fries, J.O. Nelson, 'Evaluation of an ELISA for the Multianalyte Analysis of s-Triazines in Pesticide Waste and Rinsate', *J. Agric. Food Chem.*, **41**, 322–328 (1993).
245. C. Wittmann, R.D. Schmid, S. Löffler, A. Zell, 'Application of a Neural Network for Pattern Recognition of Pesticides in Water Samples by Different Immunochemical Techniques', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, American Chemical Society, Washington, DC, 343–360, 1997.
246. G. Jones, M. Wortberg, D.M. Rocke, B.D. Hammock, 'Immunoassay of Cross-reacting Analytes', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Symposium Series 657, American Chemical Society, Washington, DC, 331–342, 1997.
247. A. Rigo, B. Hock, 'On-line Analysis with Immunosensing Devices', *GIT Lab. J.*, **1**, 22–24 (1997).
248. P. Krämer, 'Biosensors for Measuring Pesticide Residues in the Environment: Past, Present, and Future', *J. AOAC Int.*, **79**, 1245–1254 (1996).
249. P.M. Krämer, R.D. Schmid, 'Automated Quasi-continuous Immunoanalysis of Pesticides with a Flow Injection System', *Pesticide Sci.*, **32**, 451–462 (1991).
250. C. Wittmann, R.D. Schmid, 'Development and Application of an Automated Quasi-continuous Immunoflow Injection System for the Analysis of Pesticide Residues in Water and Soil', *J. Agric. Food Chem.*, **42**, 1041–1047 (1994).
251. S.A. Siebert, S.G. Reeves, M.A. Roberts, R.A. Durst, 'Improved Liposome Immunomigration Strip Assay for Alachlor Determination', *Anal. Chim. Acta*, **311**, 309–318 (1995).
252. S.G. Reeves, S.A. Siebert, M.A. Roberts, R.A. Durst, 'Liposome Immunosensing Devices for Environmental Contaminant Screening', *Trends Anal. Chem.*, **14**, 351–355 (1995).
253. S.G. Reeves, S.A. Siebert, R.A. Durst, 'Liposome-amplified Immunoanalysis for Pesticides', in *Immunoanalysis of Agrochemicals, Emerging Technologies*, eds. J.O. Nelson, A.E. Karu, R.B. Wong, ACS Symposium Series, American Chemical Society, Washington, DC, Vol. 586, 210–222, 1995.
254. M. Minnunni, M. Mascini, G.G. Guilbaut, B. Hock, 'The Quartz Crystal Microbalance as Biosensor. A Status Report on its Future', *Anal. Lett.*, **28**, 749–764 (1995).
255. A. Brecht, J. Piehler, G. Lang, G. Gauglitz, 'A Direct Immunosensor for Atrazine', *Anal. Chim. Acta*, **311**, 289–300 (1995).
256. M. Minunni, M. Mascini, 'Determination of Pesticides in Drinking Water Using Real-time Biospecific Interaction Analysis (BIA)', *Anal. Lett.*, **26**, 1441–1460 (1993).
257. F.F. Bier, R.D. Schmid, 'Real Time Analysis of Competitive Binding Using Grating Coupler Immunosensors for Pesticide Detection', *Biosens. Bioelectron.*, **9**, 125–130 (1994).
258. K. Cammann, 'Selectivity Modulation of Ion-selective Membranes – The General Principle of a New Class of Immunosensors', *Biosensors*, **90**, Elsevier Scientific, London, 1991.
259. C.F. Barbas, D.R. Burton, 'Selection and Evolution of High-affinity Human Anti-viral Antibodies', *TibTech.*, **14**, 230–234 (1996).
260. V.K. Ward, P.G. Schneider, S.B. Kreissig, B.D. Hammock, P.V. Choudary, 'Cloning, Sequencing and Expression of the Fab Fragment of a Monoclonal Antibody to the Herbicide Atrazine', *Protein Eng.*, **6**, 981–988 (1993).
261. F.G. Byrne, S.D. Grant, A.J. Porter, W.J. Harris, 'Cloning, Expression and Characterization of a Single-chain Antibody Specific for the Herbicide Atrazine', *Food Agric. Immunol.*, **8**, 19–29 (1996).
262. K. Kramer, B. Hock, 'Recombinant Single-chain Antibodies Against s-Triazines', *Food Agric. Immunol.*, **8**, 97–109 (1996).
263. A.E. Karu, K.B.G. Scholthof, G. Zhang, C.W. Bell, 'Recombinant Antibodies to Small Analytes and Prospects for Deriving them from Synthetic Combinatorial Libraries', *Food Agric. Immunol.*, **6**, 277–286 (1994).
264. S.D. Garrett, D.J.A. Appleford, G.M. Wyatt, H.A. Lee, M.R.A. Morgan, 'Production of a Recombinant Anti-parathion-antibody (scFv), Stability in Methanolic Food Extracts and Comparison to an Anti-parathion Monoclonal Antibody', *J. Agric. Food Chem.*, **45**, 4183–4189 (1997).

Multiclass, Multiresidue Analysis of Pesticides, Strategies for

Steven J. Lehotay
US Department of Agriculture, Agricultural Research Service, USA

1 Introduction **1**
1.1 Reasons for Analysis 1
1.2 Matrices 3
1.3 Pesticides 5
1.4 Quality Assurance and Quality Control 9
1.5 Strategies 11
2 Sample Preparation **13**
2.1 Homogenization and Sample Size 13
2.2 Extraction 14
2.3 Cleanup 16
3 Analytical Separations and Detection **19**
3.1 Gas Chromatography 20
3.2 High-performance Liquid Chromatography 24
3.3 Capillary Electrophoresis 26
3.4 Immunochemical Methods 26
4 Conclusions **27**
Abbreviations and Acronyms **32**
Related Articles **33**
References **34**

Strategies for multiclass, multiresidue analysis of pesticides depend on several factors that all derive from three fundamental considerations: (1) the need for the results; (2) the form and composition of the matrix; and (3) the time, resources, and technology available to the analyst. First, the need for the data dictates the quality assurance (QA)/quality control (QC) guidelines and the desired analytical figures of merit, such as limits of detection (LOD) and the accuracy and precision of the results. These needs also contribute to the choice of pesticide analytes, which sets the polarity range for the analysis. No single method can determine all pesticides in diverse matrices, so an overall strategy is often used to compartmentalize a series of approaches in order best to detect the targeted pesticides in an effective and efficient process. Second, the various types of matrices that are commonly analyzed for pesticide residues may affect the sample preparation approach employed in the analytical strategy. Common matrices for pesticide analysis consist of solids, liquids, and gases in applications associated with food, agricultural, environmental, and other sample types. Varying degrees of pesticide–matrix interactions in different materials require more or less stringent extraction procedures to separate the pesticides from the sample. Sample preparation in multiclass, multiresidue strategies frequently utilizes liquid–solid, liquid–liquid and other partitioning processes including those, such as pressurized fluid extraction (PFE), in which pressure and temperature of different fluids are controlled. Extracts often require cleanup to remove interfering co-extracted matrix components prior to analysis. A variety of cleanup techniques based on a number of partitioning processes, such as liquid chromatography (LC), may be employed in an on-line or off-line fashion. Lastly, the sample extract is usually further separated and analyzed using one or more techniques. The majority of pesticides are analyzed by gas chromatography (GC) with mass spectrometry (MS) and/or element-selective GC detectors. High-performance liquid chromatography (HPLC) is often used in the analysis of thermally labile pesticides that are not amenable to GC analysis. Other alternatives for analysis of certain pesticide include capillary electrophoresis (CE) and immunoassays. Fundamentally, a high degree of selectivity helps in generating excellent analytical results, but the wide polarity range of analytes in multiclass, multiresidue methods (MRMs) for pesticides necessitates a great deal of effort to attain a high degree of selectivity. Trade-offs and compromises are omnipresent in multiclass, multiresidue analysis of pesticides, and a host of strategies are available depending on how the analytical chemist prioritizes the analytical and practical considerations.

1 INTRODUCTION

1.1 Reasons for Analysis

Pesticides are essential in agriculture to meet the food supply demands of the world. However, owing to notable transgressions that have occurred related to the dangerous use of pesticides harmful to human and ecological health, many people throughout the world have grown fearful of pesticides. Real and perceived concerns about pesticides have led to their strict regulation, and the need for a variety of analytical strategies in diverse applications.

The analytical strategy that is used for multiclass, multiresidue analysis of pesticides depends, in part, on the need for the results. Analytical methods are needed to monitor pesticide residue levels in food and other sample types, to help measure toxicological or other effects of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the pesticides on living organisms, and to discover the fate and transport of pesticides in plants, animals, and the environment. For quantitative risk assessment purposes, it is necessary to identify and quantify accurately a great number of pesticide residues at low levels (e.g. less than nanograms per gram). In other situations, it may only be necessary to screen rapidly for certain pesticides of concern at higher levels. Also, practical considerations, such as time, labor, instrumentation, and budget, although not formal analytical figures of merit, are frequently limiting factors in the type of approach chosen by a laboratory.

Many situations in the modern world require the screening, quantitation, and/or identification of pesticide residues in food, the environment, and other types of samples. Even before a potential pesticide product is developed and introduced into the market, manufacturers must have the means to isolate, analyze, and characterize the chemical in order to gain an understanding of its nature and effects. If the manufacturer decides to introduce the pesticide into the market, nearly all nations require that the pesticide be approved (or registered) for specified uses in their country or group of countries. One common facet of the pesticide registration process requires that acceptable analytical methods be available, or that new methods be developed, for the parent pesticide and potentially harmful transformation products or coformulants.

Pesticide manufacturers are usually required to test the utility of common MRMs when applying for registration of a new pesticide. In the USA, the MRMs for foods are published in the *Pesticide Analytical Manual*, Vol. I (PAM I).[1] However, for a new pesticide, the PAM I MRMs may fail to achieve adequate recovery (typically >70%) or detection limit (less than half of the regulatory limit) for reasons related to the nature of the pesticide, matrices of interest, the need to detect metabolites, and/or an inability to modify the PAM I methods. In this case, the manufacturer must develop a new method, which will likely be a single analyte method specific to the pesticide and commodity of interest. In the USA, these methods are published in the *Pesticide Analytical Manual*, Vol. II (PAM II).[2] Modification of a multiclass, MRM or the development of a new method to widen the analytical range may be possible, but the manufacturer has little incentive to develop such an approach which could take much time and effort, benefit its competitors, and lead to greater enforcement monitoring of the product once it reaches the market. In general, single analyte methods in PAM II are complicated, rapidly outdated, and infrequently used in monitoring of pesticide residues owing to the cost of conducting many separate analyses. However, the use of multiclass, MRMs usually provides the most efficient approach for pesticide analyses.

Individual governments and international bodies, such as the Codex Alimentarius, commonly set maximum residue limits (MRLs), which are known as tolerance levels in the USA, for pesticides in food commodities and other matrices, such as water, which may be exposed to pesticides. Proper usage of pesticides is enforced, in part, by periodic monitoring of pesticide residues in collected samples. Nearly all governments worldwide (national and often local) conduct multiresidue pesticide analysis for food and other matrix types. A stipulation in international trade is that imported foods must not contain violative levels of pesticide residues. Thus, food imports are often tested at a higher sampling rate than domestic products, as included in Table 1. Furthermore, it is not unusual for

Table 1 Number of food samples analyzed for multiple pesticide residues from selected monitoring and compliance programs

Country/program[a]	Year(s)	No. of samples	Commodities	Imports (%)[b]
US (Federal):				
USDA Pesticide Data Program	1997	8177	Produce, grains, milk	13
FDA Total Diet Study	1997	1036	Many types	–
FDA Surveillance and Monitoring Program	1997	9843	Many types	54
USDA Food Safety Inspection Service	1996	7705	Meat, poultry	–
Canada: Food Inspection Agency	1992–1994	~7300/year	Fruits and vegetables	73.7
UK: Ministry of Agriculture, Food, and Fisheries	1996	3665	Many types	39
Sweden: National Food Administration	1997	~3750	Fruits and vegetables	NR
Finland: National Food Administration	1997	2479	Fruits and vegetables	77
Norway: monitoring program	1996–1997	>5940	Fruits and vegetables	NR
Netherlands: Inspectorate of Health	1997	~10 000	Agricultural products	NR
Spain: Ministry of Agriculture and Fisheries	1995	3051	Produce	NR
Italy: monitoring programs	1992–1997	>10 400/year	Food products	NR
Switzerland: cantons	1994	7945	Many types	41
Germany: food monitoring program	1995	~5000	Many types	50

Information provided by national monitoring programs in reports or presentations at Pesticide Residue Workshops.
[a] USDA, US Department of Agriculture; FDA, US Food and Drug Administration.
[b] NR = not reported.

pesticide or other types of residues to be used as a trade barrier or hurdle to help protect domestic markets.

Many industrial and academic purposes also necessitate the analysis of pesticides. For example, food processors, or their contract laboratories, often perform testing to ensure that pesticides do not occur in their products at adulterating amounts. Toxicologists, ecologists, agricultural and environmental scientists, and several others also require the multiclass, multiresidue analysis of pesticides in many types of studies. Individuals, consumer groups, and environmental organizations may also hire contract laboratories to analyze pesticides in various types of samples for a number of reasons. Additionally, the continuing rapid progress being made in analytical instrumentation and techniques have led to a great deal of method development and evaluation studies which involve the analysis of a large number of samples. Analytical methods approved by the US Environmental Protection Agency (EPA) for pesticides and other contaminants in environmental samples have been published.[3]

Estimation of the number of multiclass, multiresidue analyses of pesticides that are performed yearly around the world is difficult. Essentially, only analyses performed by government agencies in their food monitoring and compliance programs are published, and Table 1 lists the number of analyses reported by selected countries. Many more samples are analyzed than those presented in Table 1, and it is safe to estimate that >150 000 multiresidue pesticide analyses are performed by governments and industry yearly worldwide.[4]

1.2 Matrices

A variety of matrices are regularly analyzed for pesticide residues. These matrices can be broadly divided into ''environmental'' and ''food and agricultural'' sample types. Environmental samples can be divided into air, water, and solids (soils, sediments, particles, sludges, nonanimal solid wastes, and surfaces). Plant, animal, and fungi samples, including bodily fluids, may be taken for clinical, environmental, or other purposes, but owing to their similar compositions as some foods, they will be classified as ''food and agricultural'' samples. Therefore, ''food and agricultural'' items essentially consist of all edible and nonedible organisms, tissues, and products from organisms. Examples include fruits, vegetables, grains, nuts, beans, fish, meats, meat byproducts, eggs, poultry, animal feed, dairy products, beverages, spices, honey, mushrooms, processed foods, dried fruits, tobacco, plant tissues, bodily fluids, animal waste, and a host of other materials (drinking water will be regarded as an environmental water sample rather than a beverage). For purposes of residue analysis, food and agricultural matrices can generally be divided based on their composition, particularly with regard to water and fat content.

1.2.1 Food and Agricultural Samples

Because agricultural applications make up the largest area of use for pesticides, and food safety is such an important concern, foods are a widely analyzed matrix type for pesticide residues. Raw commodities, such as fruits and vegetables, are often the most appropriate food items in which to monitor pesticides because they are generally ready for consumption nearest to the times of pesticide applications and typically have the highest residue concentrations.

Foods are often complex matrices with widely varying compositions, but many foods consist of four major components: carbohydrate, fat, protein, and moisture. A fifth component, ash (or mineral composition), is rarely >5%. Of the four major components, carbohydrates include sugars, starches, oligosaccharides, and fibers; fat includes mono-, di-, and triglycerides, oils, fatty acids, and sterols; protein includes all organonitrogen components (amino acids, enzymes, etc.), and moisture is essentially water content.[5] For nutritional labeling purposes and simplicity, food chemists have deleted water from consideration and developed a diagram, as shown in Figure 1, to help categorize foods based on the relative compositions of proteins, lipids, and carbohydrates.[6] Figure 1 contains nine smaller equilateral triangles to serve as representational sectors in which similar food types appear (some examples of the relative compositions of different foods are provided in the diagram). USDA provides a Nutrient Database that is accessible in searchable format on the internet and currently lists the compositions of about 6000 food items.[7]

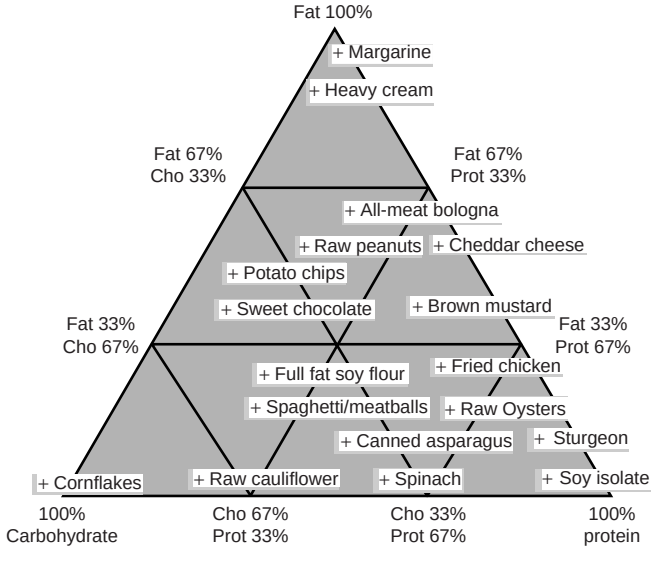

Figure 1 Major composition of foods. [Reprinted from Wolf.[6]]

The capability of analytical methods can be matrix dependent, but the great variety of foods limits practicality to validate new analytical methods for all food matrices. For this reason, scientists within the Association of Official Analytical Chemists (AOAC®) International have proposed that one type of food within a sector could serve as a reference for other types of foods within the same sector, particularly for methods to measure food composition.[8] For example, investigators in one study involving nonfatty foods chose 27 pesticides in six matrices (apple, carrot, spinach, strawberry, tomato, and wheat) to represent the broader scope of pesticides and commodities in a comparison of different multiclass, MRMs of analysis among 24 laboratories.[9,10]

For residue analysis and other purposes, the FDA has categorized and coded many different food products, which are summarized in PAM I.[1] Table 2 provides ranges of % fat, % water, and % sugars of selected commodities. In broader terms, the FDA separates analytical methods that are applicable to fatty (>2%) and nonfatty (<2% fat) foods. Further demarcations can be drawn based on % water or % sugars, but it is easy (and sometimes necessary) to add water[1] or sugars to the matrix during sample preparation, whereas fats are never added. There are instances when the addition of sugar helps in a liquid–liquid partitioning step, and the presence of water in an analytical procedure can be a critical benefit or detriment depending on the application. Thus, rather than devising a number of different strategies depending on the matrix, it is possible to limit the number of necessary methods by forcing the matrix to suit the method. Similarly, some analytical methods may include procedures, such as gel permeation chromatography (GPC), that are able to remove fats from sample extracts even though the samples may not contain fat. This concept is not necessarily the most efficient, but it does limit the number of different methods needed in the overall strategy for diverse matrices.

Table 2 Composition of selected food commodities [from Food and Drug Administration[1]]

Commodity	Fat (%)	Water (%)	Sugars (%)
Fruits	0.1–1	80–95	5–18
Vegetables	0.1–1	80–95	1–7
Rice	1.1	7.8	2.5
Lima beans	0.7	10.2	8.5
Whole milk (cow)	3.7	87.7	4.9
Soybeans	19.9	8.5	6.6
Peanuts	49.2	6.5	4.3
Mushrooms	0.4	91.8	1.8
Oyster	2.5	85.1	0
Fish, sturgeon	4.0	76.6	0
Avocado	17.3	72.6	0.9
Honey	0	17.2	81.9

Fatty foods are known for their greater difficulty in analysis owing to smaller differences in the polarity of lipids from many common pesticides. Furthermore, greater importance is often placed on the analysis of lipophilic pesticides, such as organochlorines, in fatty foods because those pesticides may be persistent, bioaccumulative, and/or pose higher human health or ecological risk. Nonlipophilic pesticides do not accumulate in animal tissues, hence there is less need to monitor for them in fatty matrices.

Nonfood plant and animal tissues can often be treated in a similar fashion to food items based on their composition; however, the nutrient compositions of tissues inedible to humans are not as well documented. Also, the cell walls and exoskeletons of plants and animals not generally eaten by humans often require more effort to rupture during extraction than the techniques used for food items. For example, longer extraction times, acid–base conditions, and/or higher temperatures may be needed. Furthermore, more complications in analysis can occur owing to the potential interference of natural compounds developed in defense and other evolutionary processes. For these reasons, in part, it is common for analysts to devise single analyte or single class residue methods for unique matrices. In some matrices, it may be impossible to achieve sufficiently low detection limits and high recoveries for a wide polarity range of pesticides with traditional multiclass, MRMs.

1.2.2 Environmental Samples

1.2.2.1 Water Environmental water samples consist of drinking water, surface and groundwater (salty or fresh), precipitation, ice, process streams, waste effluents, and other aqueous media. Water is generally considered an easy matrix for pesticide residue analysis because it often contains few interferences and is usually sampled as a homogeneous liquid. Water samples are often filtered to remove suspended particles which allows the separate determination of dissolved analytes from those associated with the particle phase. Very low detection limits can be achieved by concentrating extracts from large sample volumes. However, some water samples with high concentrations of organic substances, such as marsh water or waste streams, may require cleanup of the extracts. Solid-phase extraction (SPE) has become the primary means to extract and cleanup water samples.

1.2.2.2 Solids Soils, sediments, sludges, compost, particles separated from water and air, surfaces, and other solid environmental materials can be categorized as environmental solid sample types. Sediment is essentially

the same as soil except that it occurs underwater; sludge is the waste material that settles in waste treatment plants or other facilities. Soils and other environmental solids tend to have much stronger adhesion and/or other forms of chemical interactions than other types of matrices. An unsettled issue among pesticide residue chemists is whether special measures should be taken to extract bound residues that are covalently or otherwise strongly infused with the matrix.[11] In the pesticide registration process in the USA, the EPA requires that a ^{14}C-labeled analyte be used in method development studies to determine the total extractability of the pesticide in the targeted matrices. In the case of soils, the extraction process of pesticides often requires stronger measures (sonication, longer extraction times, acid–base conditions, repetitive steps, and/or higher temperatures) than in the case of food samples.

The variability of soils is very high, and just as nutritionists have characterized foods by protein, carbohydrate, and fat (Figure 1), soil scientists have developed a textural classification diagram for soils based on sand, silt, and clay content (Figure 2). The classification boundaries and nomenclature of soil types appear in the diagram (loam is a mixture of sand, silt, and/or clay). According to USDA definitions, sand (and gradations thereof) has particle sizes between 0.05 and 2 mm, silt particle sizes are between 0.002 and 0.05 mm, and clay consists of particles smaller than 0.002 mm.[12] Sand separates (very fine, fine, medium, coarse, and very coarse) are determined by sieving, and silt and clay are differentiated by their settling rates in water. Particles larger than 2 mm, which have little bearing on either soil properties or analytical residue methods, are considered as stones or gravel and are excluded from the textural classification.[12]

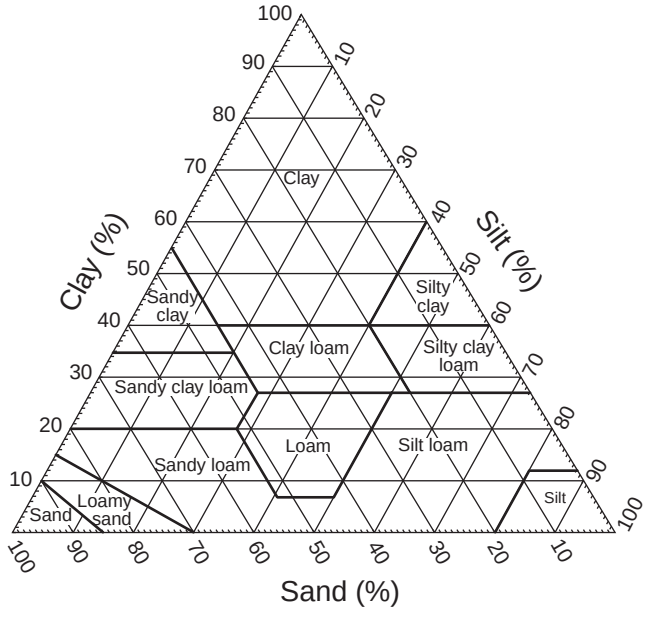

Figure 2 Textural diagram of soil. [Reprinted from Troeh, Thompson.[12]]

As in the case of the food classification diagram, water is not represented in the soil textural diagram even though it can be a major constituent. The organic content of soils, which can be likened to the many potential analytical interferents in food and agricultural samples, is also an important component that is left out of the diagram (the organic content is usually a low percentage overall). However, water, salts, pH, and organic content can all have major effects in the analysis of pesticides in soils, sediment, sludges, compost, and suspended particles.

Other types of solid samples that may be analyzed for multiple types of pesticides include exposed surfaces. Flooring, counters, and other surfaces may be sampled by rinsing or wiping with solvents to test for effectiveness and legality of pest control in homes and businesses, and clothing, skin, hair, and similar materials may be sampled in the study of occupational or other cause of exposure to pesticides. Also, it is sometimes necessary to test container surfaces to determine unknown types of pesticides that may have been present previously or to study the effectiveness of a cleaning process. Similarly, multiclass, MRMs are needed to identify unlabeled pesticide formulations or waste materials of unknown origin.

1.2.2.3 Air Owing to the nature of the gaseous medium, air samples require very different approaches, particularly in the sample collection process, to those employed for solids and liquids.[13] Sample volumes are determined by flow rates, time, temperature, and pressure, which must be known to convert volumes to standard conditions. Collection of fog and suspended particles in air also requires unique procedures, but once air or samples from the air have been collected, cleanup steps are rarely required. Typically, the sample extracts are concentrated in a liquid or solid-phase trap and injected directly into a GC column for analysis. LC is generally not used owing to the nature of typical pesticide analytes found in air and more complicated injection procedures for air samples. Constituents from very high volumes of air can be concentrated by trapping to achieve exceptionally low detection limits (below the nanograms per cubic meter level).

1.3 Pesticides

The diverse range of purposes and sample types for pesticide analysis is only matched by the diversity of chemicals to be analyzed. There are numerous sources of information on the chemistry, applications, and regulation of pesticides.[14–17] A brief review of fundamental aspects of pesticides is provided in the following paragraphs.

The term ''pesticide'', as defined by the EPA, broadly represents ''any substance or mixture of substances intended for preventing, destroying, repelling, or mitigating any pest''. These pests can be divided into categories such as weeds, insects, arachnids (mites and ticks), fungi, slugs, nematodes (a type of worm), rodents, microorganisms, and miscellaneous other pests, which form the basis for the more specific terms, ''herbicides'', ''insecticides'', ''acaricides'' (or ''miticides''), ''fungicides'', ''molluscicides'', ''nemiticides'', ''rodenticides'', and ''antimicrobial pesticides'' (and ''biocides'' or ''disinfectants''). Pesticides also generally include fumigants, sterilants, growth regulators, defoliants, repellents, and synergists (compounds that aid action of a pesticide). For the purpose of simplicity, the term ''insecticide'' is commonly used to include all pesticides used for control of animal pests, ''herbicide'' includes plant growth regulators, and ''fungicide'' can also refer to chemicals used in noninsecticidal postharvest applications (e.g. antiscalding).

In strict terminology, broader or narrower terms have been devised, such as ''agrochemical'' (to include fertilizers and other compounds of agricultural significance), or ''crop protection agent'' (to differentiate agricultural and nonagricultural uses of pesticides), but sometimes other terms, such as ''plant protection product'', are used to avoid the word ''pesticide'' for public perception reasons. Pesticides also include natural chemicals used for pest control, but these are sometimes called ''alternative'' or ''natural'' pesticides to distinguish them from synthetically derived pesticides.

Agricultural usage of pesticides constitutes the largest application sector, but home, business, and other pest control applications are also significant. The USDA National Agricultural Statistics Service (NASS) provides annual reports, which are accessible through the Internet, concerning pesticide usage on field crops in the USA.[18] Herbicides are the most heavily used pesticides in agriculture, followed by insecticides, then fungicides. Agricultural applications of pesticides can be further divided into preplanting, preemergent, preharvest, and postharvest in the case of plant products, or preslaughter and postslaughter in the case of livestock.

Pesticides are either inorganic or organic chemicals, but usually different analytical methods are used to detect inorganic compounds to those used for organic analytes. Inorganic pesticides, which can be herbicides, insecticides, and/or fungicides, include compounds and salts containing arsenic, copper, sulfur, mercury, or other metals (organometallic pesticides are considered organics). However, only organic species are generally analyzed in multiclass, MRMs for pesticides, and the inorganic pesticides are typically analyzed in foods and the environment separately using elemental detection methods. A few approaches, such as atomic emission detection (AED) after GC, may be able to analyze organometallic species in the same procedure as other classes of pesticides,[19] but it is more common to have separate methods for organometallics.

1.3.1 Herbicides

Table 3 presents major classes of herbicides, their solubility range in water, common examples of individual herbicides within each class, and their basic chemical structures. Many herbicides are applied early in the agricultural season and generally dissipate rapidly. Furthermore, more recently introduced herbicides, such as sulfonylureas and imidazolinones, tend to have selective weed-control properties, pose little toxicological threat to animals, have relatively low application rates, and degrade rapidly.[20] Few herbicides appear as residues in foods, and many are not commonly included in multiresidue, multiclass methods. Those which have longer environmental persistence than a few months, such as some triazines and acetamides, can typically be recovered and detected at low levels (nanograms per gram) using many current multiclass, MRMs, and are frequently included in analyses, but they are still rarely found in foods.[21] However, heavily applied and relatively stable herbicides, such as triazines, are frequently found in environmental water and soil/sediment samples.[22]

Owing to the relatively high polarity and often ionic nature of many herbicides, they are more prevalent in water and soil samples than in food and air matrices. Owing to the nature of the analytes, the development of multiclass, MRMs for polar/ionic herbicides is more difficult than in the case of less polar pesticides. Instrumental approaches, such as HPLC or CE, may permit the simultaneous analysis of multiple polar/ionic herbicides,[23,24] but few cleanup techniques allow the recovery of a wide range of polar analytes in complex matrices.[25] Hence, typical multiclass, MRMs for polar herbicides are generally only useful for relatively clean aqueous matrices.

1.3.2 Insecticides

Many insecticides are also used as acaricides, nematicides, and rodenticides, and vice versa. A few herbicides (dinitrophenols) and fungicides (organotins) may also have insecticidal activity, and the categorization of pesticides can depend as much on how a pesticide is used as its chemical class. Table 4 lists the major classes of insecticides on the basis of relatively similar chemical structures. In some classification schemes, the elemental composition of the pesticide determines its category, such as organophosphorus, organonitrogen, organosulfur, and

Table 3 Major classes of organic herbicides and examples

Class	Solubility range in water (mg L^{-1})	Examples	Basic structure
Acetamide	10^2–10^4	Alachlor, metolachlor	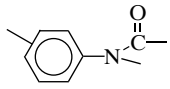
Chlorophenoxy	10^1–10^6	2,4-D, 2,4,5-T, MCPA, silvex	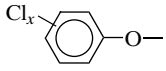
Dinitroaniline	10^{-1}–10^0	Pendimethalin, trifluralin, dinitramine	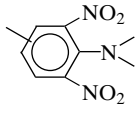
Imidazolinone	10^1–10^6	Imazaquin, imazapyr, imazethapyr	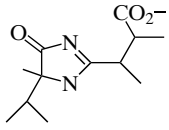
Phenylphenoxy	10^{-1}–10^6	Fomesafen, bifenox, fluazifop-butyl	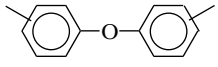
Phenylurea	10^0–10^3	Linuron, diuron, thidazuron, neburon	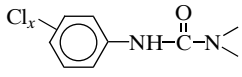
Sulfonylurea	10^2–10^4	Chlorsulfuron, chlorimuron-ethyl	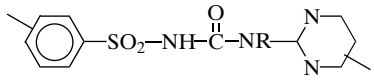
Thiocarbamate	10^0–10^6	Vernolate, asulam, butylate, thiobencarb	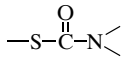
Triazine	10^0–10^4	Atrazine, ametryn, simazine, prometon	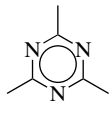

Table 4 Major classes of insecticides

Class	Solubility range in water (mg L^{-1})	Examples	Basic structure
Carbamate	10^1–10^5	Carbofuran, aldicarb, propoxur, oxamyl	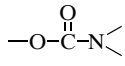
Organochlorine	10^{-3}–10^0	Methoxychlor, DDE, lindane, endosulfan	Insecticides containing chlorine
Organophosphorus	10^{-1}–10^6	Diazinon, chlorpyrifos, acephate, ethion	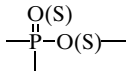
Pyrethroid	10^{-3}–10^{-1}	Permethrin, cyfluthrin, fenvalerate, bifenthrin	Similar to 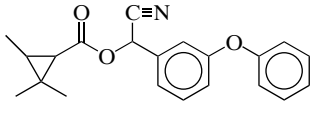

organochlorine (or organohalogen). This classification approach is useful when considering which type of elemental selective detector in GC can be used, but the pesticides within these categorizations are not necessarily recovered through the same extraction and cleanup approaches. Also, there is considerable overlap for many pesticides that contain more than one of these elements. Both of these circumstances are also true for chemical structural classification schemes, but there is less overlap.

Chlorinated hydrocarbon and cyclodiene insecticides, widely known as organochlorine pesticides (OCs), are the most widely analyzed pesticides in the most matrix types. They are most commonly detected in fatty and nonfatty foods, soils/sediment, water, and animal and plant tissue samples. Nearly all OCs were banned in industrialized nations during the 1970s, and essentially only endosulfan and dicofol, which are less persistent and pose less risk than the others, are still commonly used in these countries today. However, many of the banned OCs are still applied in several developing countries to combat malaria and other insect-borne diseases. In environmental and epidemiological studies, analytical methods often target OC insecticides along with polychlorinated biphenyls, polycyclic aromatic hydrocarbons, and other nonpolar contaminants in a similar or the same general approach.

Organophosphorus insecticides (OPs) are currently the most widely used insecticides, but they are less frequently monitored in environmental samples than the OCs owing to their lower persistence. In nonfatty foods, OPs are detected more frequently than OCs, and their residue concentrations are typically slightly higher, whereas the opposite is true in the case of fatty foods. The variety of OPs and their different uses are unmatched among all pesticide classes.

Carbamate insecticides, which are sometimes called *N*-methylcarbamates to distinguish them from (thio/phenyl) carbamate herbicides and (dithio)carbamate fungicides, are also frequently used in agricultural and home applications, but there are fewer carbamate insecticides than OPs. A few of the thermally stable carbamates (propoxur, carbofuran, and carbaryl) can be analyzed by GC methods, but a separate HPLC procedure is more commonly conducted for the simultaneous analysis of *N*-methylcarbamates.[26] HPLC can also be used for the simultaneous analysis of OPs along with carbamates and other pesticides, but capillary GC provides superior separations and more sensitive detectors for the numerous OPs. The use of single class, or selective, MRMs such as methods for carbamates, benzimidazole fungicides, phenylurea herbicides, and others are often incorporated into an overall multiclass, multiresidue strategy.[1,22]

Owing to recent studies on neurotoxic effects of cholinesterase-inhibiting insecticides (OPs and carbamates), especially with respect to infants and children,[28] there is currently much discussion among scientists, environmental groups, farmers, the chemical industry, regulators, and politicians about whether the use of OPs should be curtailed. For this reason, regulatory and risk assessment needs have further emphasized the detection of OP and carbamate pesticides at ultratrace levels in all matrices to which humans are exposed.

Pyrethroid, nicotinyl, alkyloid, and rotenone insecticides are either synthetically derived from naturally occurring compounds produced by plants in defense from pests or isolated from organisms themselves. The pyrethroids constitute the most widely used of these classes currently, and their use is expected to grow if the prevalence of OPs decreases. The human health risk of pyrethroid pesticides is lower than that of OPs but they commonly have higher acute toxicity to fish and other aquatic species. In the future, several other chemical, physical, and biological pesticide alternatives may help to reduce the reliance on many of the current commonly used insecticides.

1.3.3 Fungicides

Table 5 lists common fungicide classes and their basic structures. Some fungicides are used in preharvest applications to control fungi in the soil or above ground on plants. Other fungicides are used in postharvest applications to improve product quality or extend shelf-life. The fungicides that are used in postharvest applications are the most commonly found in fresh produce, and their concentrations can be much greater than those of other residues.[21] Therefore, multiclass, multiresidue strategies for foods often include several fungicides in the approach, but several organonitrogen fungicides (azoles, imides, etc.) can be problematic in the procedures. However, there is little toxicological risk of approved postharvest fungicides to animals. Tolerances or MRLs for several postharvest fungicides are $>5\,\mathrm{mg\,kg^{-1}}$ in fresh commodities, which can be more than 100-fold higher than some higher risk insecticides.[21]

Postharvest fungicides are rarely expected to occur in environmental samples, and few environmental methods look for them. However, some preharvest fungicides may be found both in the environment and in foods. OC fungicides, such as chlorothalonil and quintozene, and transformation products are sometimes also included in multiresidue, multiclass analyses of environmental samples.

The widely used dithiocarbamate fungicides are not analyzed with the same methods as other pesticides owing to their chemical instability and unique properties.

Table 5 Major classes of fungicides

Class	Solubility range in water (mg L^{-1})	Examples	Basic structure
Dithiocarbamate	10^{-1}–10^{5}	Mancozeb, maneb metriam, zineb	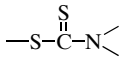
Imidazole	10^{0}–10^{4}	Carbendazim, imazalil, thiabendazole	
Phthalimide	10^{0}	Procymidone, captan, captafol, folpet	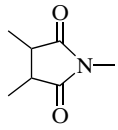
Triazole	10^{1}–10^{3}	Myclobutanil, propiconazole	

A common strategy is to digest the dithiocarbamates to form either ethylenethiourea (ETU) or CS_2 for detection.[29] The result represents a summation of all dithiocarbamates, but the chance of inaccurate results is substantial owing to natural sources of compounds that also produce ETU or CS_2 in the approach. Extraction with a basic buffer and ethylenediaminetetraacetic acid (EDTA) followed by ion-pairing HPLC can determine individual dithiocarbamates.[30]

1.3.4 Pesticide Properties

In any analysis, the physicochemical properties of the analyte(s) determine the type of possible approaches that lead to a successful extraction, cleanup and detection procedure. An important aspect is the polarity of the analyte, which can be estimated for the purposes of this article through its solubility in water and/or its octanol–water partition coefficient (K_{ow}). In the analysis of pesticides that are weak acids and bases, pH and ionic strength also become critical aspects in the approach. Other factors of note include volatility and stability, which indicate what precautions must be made to avoid analyte losses.

Several chemical properties of interest for many pesticides are listed in the USDA Pesticide Properties Database,[31] which is also accessible on the Internet (http://www.arsusda.gov/rsml/ppdb.html). Figure 3 includes the 353 pesticides in the database categorized by their solubilities in water which broadly indicate the number of pesticides with a particular polarity range. Tables 3–5 provide the water solubilty ranges of the different pesticide classes listed in the database. The majority of pesticides are considered to be semipolar, with solubilities in water of ~1–10 mg L^{-1}, as can be seen by the profile in Figure 3, but many pesticides, generally herbicides, tend to be rather polar with solubilities in water >100 mg L^{-1}. Polar pesticides are typically more

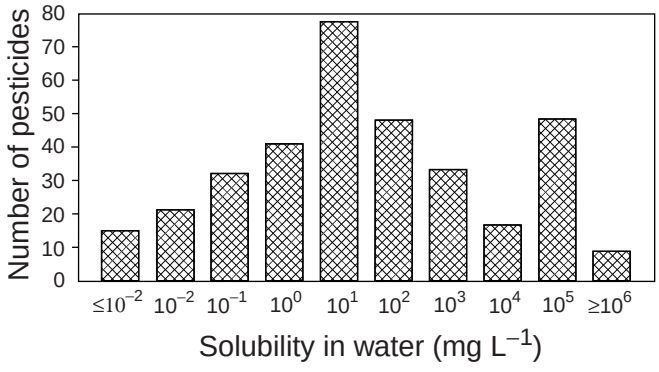

Figure 3 Polarities of pesticides in the Pesticide Properties Database as estimated by their solubilities in water. [Reprinted from Lehotay.[61]]

problematic in multiresidue analysis, but they also tend to be less toxic to humans and less persistent in the environment. However, analytical strategies for less polar pesticides are fairly straightforward. The inclusion of polar pesticides increases the difficulty in the analysis and, in practice, it is more common to have separate methods for the most polar pesticides owing to their more unique characteristics. Even polar pesticides with the same solubility in water at a given pH may behave much differently in a particular approach owing to different functional groups, which is a less frequent problem for nonpolar and semipolar pesticides.

1.4 Quality Assurance and Quality Control

QA and QC overviews have been published.[33,34] For pesticide residue monitoring of foods, a working group of the European Union (EU) recently published their recommended QA/QC guidelines.[35] QA is the process to help ensure the quality of results using any analytical

method, and QC represents the provisions taken for a particular set of samples, such as blanks and spikes, to help verify the quality of the results of that sample set.

QA/QC standards should be set to meet the needs for the results. For example, strict standards are used when the most accurate and precise results are needed, and looser standards may be used when rapid screening or general surveillance is being conducted. Very accurate and precise results meet the needs of all applications, but often the sample throughput, cost requirements, or analytical difficulties limit the ability to always obtain results of the highest quality. A balance in the worth of the results versus the cost of obtaining them should be sought.

In recent times, QA guidelines have been expanded to more holistically incorporate laboratory functions, and the term good laboratory practices (GLP) has emerged. GLP is a process to ensure that proper procedures are followed within the laboratory to achieve results of acceptable quality. GLP includes such factors as having a QA manual and officer, developing and following standard operating procedures (SOPs), training and certifying analysts, periodically checking and recording freezer temperatures, regularly calibrating balances, and several other functions.[36] GLP, in turn, has led to the next step for a laboratory, which is to obtain accreditation for a particular analysis by an independent accrediting body. The standards that are followed for laboratory accreditation have been debated among different organizations for many years, but recently, ISO 25 criteria devised by the International Standards Organization (ISO) have begun to be generally accepted.

1.4.1 Method Validation

Analytical methods must be validated through repetitive recovery studies of fortified samples and, depending on customer requirements and analytical needs, these studies may need to be performed in a minimum number of different laboratories. In the case of pesticide residue analysis, pesticide recoveries of 70–120% in matrix spikes using three replicates each at 2–3 different concentrations are generally acceptable. Depending on the concentration, the variability of the recoveries should be <20% relative standard deviation (RSD). Furthermore, methods must be capable of attaining the desired limit of quantitation (LOQ) which may be checked by spiking samples at this level. Some organizations require laboratories to submit only those results that have been obtained using prescribed methods that have undergone interlaboratory method validations.

The AOAC® International is an organization that is involved in the certification of analytical methods through interlaboratory collaborative studies in the analysis of shared samples. For a method to obtain AOAC® Official Methods or Peer Verified status, defined procedures are followed and certain statistical criteria for the results must be met.[37] Interlaboratory validations for any method can be time consuming, laborious, and expensive, and subsequently a good method may still give inconclusive results in the study for a variety of reasons. The effort required to carry out a multiclass, multiresidue collaborative study is substantial, and to consider all analytes and matrices is practically impossible. Since 1980, only a single multiclass, MRM for pesticide residues in foods has been evaluated and attained Official Method status.[38] The collaborative study, conducted in 1984, involved only six pesticides in three fruit and vegetable matrices.[39] Another collaborative study was attempted only recently and the results were acceptable for nearly all of the 13 fortified pesticides evaluated in 3 commodities.[40]

The rigorous requirements and great effort to perform interlaboratory validations have tended to stifle the implementation of new approaches for official monitoring purposes. It often takes many years for a new, advantageous technique to achieve general use, in part owing to stringent, impractical, or incompatible method acceptability parameters. For this and other reasons, the concept of performance-based measurement systems (PBMS) is common for multiclass, MRMs. PBMS relies on internal laboratory method validation results and QA/QC protocols to verify that the quality of analytical results is satisfactory. The proliferation in the use of proficiency testing, or blind check samples, has made it possible to determine the quality of results from a laboratory, analyst, and/or method in a situation reasonably close to real-world analysis. In some laboratories, check samples are prepared regularly by the QA officer, and more widespread programs have developed among a group of laboratories or in a fee-based service.[41]

The proficiency test samples may contain naturally incurred residues in addition to those spiked into the matrix. One of the problems with the use of only spiked samples in method validation is that the pesticides are not incorporated as strongly into the matrix, so higher recoveries may be achieved for the spiked samples than real-world samples. However, results from studies involving the analyses of proficiency test samples indicate that several different methods can achieve equally valid results for many pesticides in food matrices.[9,10,41]

The lack of a known concentration for incurred residues has made the determination of accuracy somewhat questionable, but comparison of results using different methods is useful. Standard reference materials are available for certain OCs in animal tissues[42] or environmental solids to help verify analytical methods, but none have been developed for other types of pesticides. An attempt to develop a reference material for pesticides in grain determined that inadequate stability of the pesticides,

inhomogeneity of the samples, and variability in the results made the attempt unsuccessful.[43] The comparison of results for shared samples has become a means by which new methods are commonly validated.

1.5 Strategies

Usually, the most efficient overall analytical strategy for a laboratory to reduce time, labor, and cost is to use the fewest possible methods for the most analytes. An inherent difficulty with multiclass, MRMs is that as the polarity range of analytes increases, more matrix co-extractives and potential interferences occur. No current method can analyze for all pesticides from all matrices with a single approach. Therefore, the overall analytical strategy depends on how the chemist decides to compartmentalize the range of analytes and matrices to be analyzed. The strategy is constrained by the laboratory budget/resources, among other factors, but it must still meet the QA criteria determined by the need for the data.

The majority of nonpolar and semipolar pesticides do not pose much trouble in analysis, and many techniques can achieve high-quality results for stable OCs, OPs, and similar compounds. The challenges and main differences in approaches occur for those problematic pesticides that are more volatile than the others, tend to degrade, possess matrix-dependent interactions, interfere in the detection of another pesticide, or stick to glassware, among other possibilities. For best results with problematic pesticides, single-analyte or single-class methods may be more appropriate,[1,27,44] but this is not commonly possible in laboratories owing to the added effort and cost. The multiclass, multiresidue strategy within a laboratory should be developed to minimize cost and maximize convenience for the pesticides and matrices of interest while still meeting the needs for the results.

Figure 4 illustrates the monitoring strategy used by laboratories of the General Inspectorate for Health Protection in The Netherlands for the analysis of pesticide residues in fruits and vegetables. The ''Dutch approach'' is probably the most streamlined and cost-effective strategy implemented for a wide range of pesticide residues in fruits and vegetables, and it has been extensively evaluated.[27]

However, the multiclass, multiresidue analysis of pesticides is a constantly evolving process. New pesticides, instruments, and techniques are frequently introduced, and continually changing priorities, budgets, and analytical needs compound the problem. Many chemists in many monitoring laboratories around the world are striving to devise the optimal approach. Even well-established methods are modified frequently, and few laboratories follow methods exactly as published in the literature, even those in which the published methods were originally developed.

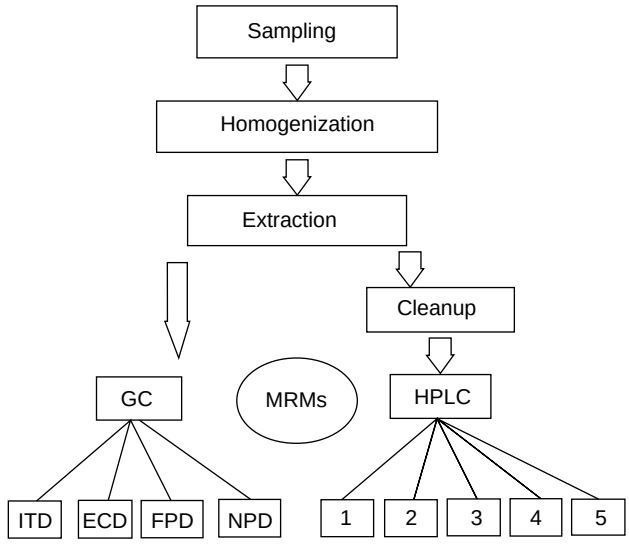

Figure 4 Current overall strategy of the Dutch multiclass, multiresidue analysis of pesticides in foods. HPLC methods: 1, *N*-methylcarbamate insecticides; 2, benzimidazole fungicides; 3, phenylurea herbicides; 4, benzoylphenylurea insecticides; 5, conazole fungicides. ITD, ion-trap detector. [Reprinted from A. de Kok, M. Hiemstra, presentation at the 2nd European Pesticide Residue Workshop, Almerìa, Spain, May 24–27, 1998.]

For example, Milton Luke of the FDA, the originator of the ''Luke'' method,[45] has spent many years altering and optimizing the method named after him. In another case, Fillion et al. altered their primary multiclass, MRM used in the Canadian Monitoring Program in order to utilize new SPE cartridges in the market.[46] Furthermore, the chemists in the Inspectorate of Health in The Netherlands took much time and effort to validate their current suite of methods as part of their overall analytical strategy, and they will likely spend several years in the future developing and validating new methods that take advantage of the improvements offered by recent advances in liquid chromatography/mass spectrometry (LC/MS) and other technologies.

Invariably, trade-offs and compromises are required in any multiclass, multiresidue strategy, and not all analytes will be recovered completely for detection at low levels in all matrices. Each approach may possess advantages and disadvantages with respect to certain pesticides, matrices, and practical aspects (convenience, costs, time, space, etc.). There are so many different variations in the analysis of pesticide residues that it is practically impossible to present them all. Thus, only major differences in the analytical process (sample preparation, analytical separations, and detection) will be presented in this article. Recent review articles for pesticide residue analysis can be found in a number of publications.[47–50] Table 6 lists and compares different methods of analysis

Table 6 Comparison of multiresidue extraction methods

Parameter	FDA[51]	CDFA[52]	Swedish[44]	Canadian[53]	Dutch[27]	SFE[40]
Sample size (g)	100	50	75	50	15	50 (3)
Solvents (mL):						
Acetone	367	2		40	30	3
MeCN		100		115		
EtOAc			225			
CH_2Cl_2	300	10			34	
Light petroleum	23				30	
Isooctane					4.5	
Toluene				5	0.5	
Cyclohexane			25			
Hexane		20				
Total	690	132	250	160	99	60 g CO_2
Estimated time (h)	3	2	2	2	1.5	1.5
Labor	Manual	Manual	Manual automated GPC	Manual	Manual automated SPE	Automated
Space	Lab.	Lab.	Lab.	Lab.	Bench + hood	Bench
Glassware	>10 items	>10 items	<5 items	<10 items	<5 items	jar, vial
Equipment	Blender, SPE set-up Steam-bath Filters	Blender Steam-bath SPE set-up, N_2 evap., Filters	Blender, GPC system, Rotovap N_2 evap.	Blender Centrifuge, N_2 evap., SPE set-up, Rotovap	Blender Centrifuge, SPE set-up	Blender, SFE
Chemicals	C_{18} SPE Na_2SO_4, 5 g NH_2 SPE QMA SPE	NaCl, phosphate Na_2SO_4, 20 g C_{18} SPE, NH_2 SPE	Na_2SO_4, 60 g 40 × 1 cm SX-3 GPC column	NaCl, 10 g Na_2SO_4, 10 g C_{18}, carbon, NH_2 SPE	NH_2 SPE	Hydromatrix, 50 g $MgSO_4$, 3 g ODS trap
Approx. material cost($)[a]	13	10	5	9	2	3

CDFA, California Department of Food and Agriculture; Current FDA, CDFA, and Canadian methods have been modified slightly from publications. SFE, supercritical fluid extraction; ODS, octadecylsilane; QMA, quaternary methylamine.

[a] Costs vary depending on vendors; other costs not presented include labor, instrumentation, and indirect charges.

for multiple pesticide residues fruits and vegetables. Minor modifications of these methods (except in the case of the SFE method) often permit the analysis of fatty foods, and these methods are often also applicable to other matrices. Analytical methods commonly used for water, soil, sludge, and other environmental matrices have been published.[3,54]

2 SAMPLE PREPARATION

All aspects in the analytical process are equally important, and no step can improve upon the quality of the previous step in the procedure. A selective, sensitive, and rugged detection system can ease or reduce the sample preparation procedures in a method, but not improve recoveries. Unfortunately, a highly sophisticated detection system cannot improve upon poor sample preparation procedures, nor can the achievement of very accurate and precise results by a laboratory extend beyond the scope of what the samples collected statistically represent. Therefore, the analytical strategy actually begins with devising an adequate statistically representative sample collection scheme, proper sample handling protocols, and appropriate QA/QC guidelines to meet the goals of the analysis. The USDA Pesticide Data Program serves as an example what is required to plan and implement a monitoring program to meet specified needs.[21]

Sample preparation consists of homogenization, extraction, and cleanup steps. In the case of multiclass, multiresidue pesticide analysis, different approaches can have substantially different sample preparation procedures, but employ the same determinative steps. Some detection methods may permit fewer sample preparation steps, but in some cases the quality of the results or ruggedness of the method suffers when short cuts are attempted. The manner in which sample preparation steps overcome the selectivity vs. polarity range conundrum inherent to multiclass, multiresidue analysis is often the most pivotal facet in the overall approach.

2.1 Homogenization and Sample Size

The choice of sample size and homogenization step begins the analytical strategy in the laboratory. However, sample collection, transport, and storage begin the overall process and as much care must be taken in these steps as any other. In applications that require the most care in the analysis, samples should be stored frozen, generally at the temperature of dry ice sublimation (−78.5 °C), or at −40 or −20 °C in freezers. For water samples, which may be too large to store in most laboratories, extraction is often performed as soon as possible after collection, and the concentrated extracts are stored. In many pesticide monitoring applications, even extremely large samples are extremely small in comparison with the commodity, environmental area, or other unit that the sample is meant to represent. It is often better to take several replicate samples of a manageable size, depending on the situation, rather than a larger sample that may be difficult to homogenize. The degree of variability of the sample type is determined by taking samples from different locations at the same time and/or samples at the same location over the course of time.[55]

Sample sizes in environmental applications generally range from very small (<1 g in particle phase samples) to very large (e.g. 400 m^3) in air analysis. For water, sample size is typically 1–4 L, but smaller or larger samples are sometimes adequate or necessary in certain applications. Soil samples range from ca. 3 to 100 g but, unlike many food samples, homogenization and subsampling procedures with large samples are not generally performed. Replicate samples from the monitoring area to be assessed are usually analyzed instead to improve the accuracy of the overall results.

In fresh food commodity applications, the FDA recommends taking samples of at least 20 lb. (9.1 kg) for comminution using an appropriate chopper.[1] However, many food sampling programs commonly take smaller sample sizes. In some cases, the variability of the individual items of the commodity is a goal of the study which involves the analysis of many individual items. Furthermore, the distribution of the pesticides within a single item may be desired, and appropriate samples (exposed surfaces, inner tissues, different sides, etc.) may be taken.

For extraction, current approaches do not conveniently extract solid samples sizes greater than 100 g, nor would larger sample sizes be desirable in most applications. Hence, subsampling of the larger composite sample is generally needed prior to extraction, and it is very important to homogenize the composite sample as well as possible. Miniaturization of the subsample size taken for analysis generally leads to a more cost-effective and convenient approach, and smaller amounts of potentially interfering matrix co-extractives have to be contended with or removed. In general, the smallest sample size should be taken that adequately represents what is intended and allows an adequate LOD. Increasing the sample size does not decrease the LOD if matrix co-extractives are the limiting source of noise in the analysis.

Simple precautions taken during the homogenization step can lead to great benefits in the steps that follow. Ideally, the food samples are comminuted in the frozen state to improve the quality and ease of homogenization, and to decrease pesticide losses.[56] Certain volatile (e.g. dichlorvos) and unstable pesticides (e.g. dithiocarbamates) are completely or partially lost during homogenization at room temperature.[57] The addition of dry ice is a common

approach to maintain cold temperatures and it also acts to introduce a dispersant which improves the homogenization process.[32] Other options are to perform the homogenization in a cold environment or to precool the mixing surfaces and work quickly. Otherwise, a special cryogenic mill may be used, but these devices are generally more applicable to small sample sizes.[58] The quality of the chopper and the use of an appropriate sample size in relation to the capacity of the chopper volume are also important factors in achieving the most homogeneous samples.[59] Even if all precautions are taken, certain matrix types and pesticides of interest may give more variable results than others. For example, Hemingway et al. found that cypermethrin homogenized in cabbage gave more reproducible results than in apple.[60]

2.2 Extraction

The extraction procedure begins the process to separate the analytes from the matrix and present the material in a form that can be more easily analyzed. Extraction strategies in multiclass, multiresidue analysis may be designed to (1) employ exhaustive extraction steps to achieve the desired wide range of analytes and rely on more postextraction cleanup steps to improve selectivity or (2) divide the analytes into different classes and utilize several selective extractions to cover the overall polarity range. Most laboratories opt for the former means to reduce effort in extraction, often at the expense of more effort in cleanup prior to analysis. However, cleanup steps can often achieve a higher degree of selectivity than most extraction steps because extraction frequently requires stronger measures to overcome analyte–matrix interactions and to reach small pores in the matrix. This reduces the degree of analyte/matrix selectivity that can be achieved in the extraction procedure.

The choice of solvent generally determines the selectivity of the extraction. With the introduction of instrumental techniques that control pressure and/or temperature of the solvent during extraction, a new way of extraction is emerging. The overall approach may be termed PFE, which consists of SFE,[61–64] microwave-assisted extraction (MAE),[65–67] and pressurized liquid extraction (PLE),[68–70] which is also referred by names such as accelerated solvent extraction (ASE™), pressurized solvent extraction (PSE), and enhanced solvent extraction (ESE). Of course, the extraction concept is the same independent of these approaches, but the use of these techniques to affect the dielectric constant, viscosity, and other properties of solvents can provide benefits in achieving the desired analyte solubilities, extraction kinetics, solvent-to-sample ratios, and a higher degree of selectivity. These techniques can also be useful to obtain more stringent conditions to help overcome matrix–analyte interactions.

Traditionally, liquid solvents at room temperature and pressure are used in the extraction of pesticide residues in solid samples. Table 7 provides some of the properties of typical solvents that are commonly used in the analytical laboratory for pesticide analysis. The important properties of the solvents in analytical applications include polarity (as estimated by the dielectric constant, ε), boiling point, density, miscibility with other solvents, viscosity, safety, and cost. The combination of these solvents at room temperature and pressure can provide a reasonable degree of control in the desired properties, but this ability is small when compared with the possibilities of adding the dimensions of pressure and temperature to the equation for the same solvents.

Pressure and temperature control also permits the inclusion of unconventional solvents, such as carbon dioxide and other substances which are gases at room temperature and pressure. A pressurized fluid that becomes a gas after extraction saves the trouble of performing time-consuming and wasteful solvent evaporation steps prior to analysis. Many instrumental techniques reduce labor and variability through automation. Furthermore, sample size reduction and a higher degree of selectivity in the extraction can reduce or eliminate the cleanup steps

Table 7 Properties of common solvents used in pesticide residue analysis

Solvent	Dielectric constant, ε	Polarity index	Viscosity ($mN\,s\,m^{-2}$)	Boiling point (°C)	Solubility in water (% w/w)
Water	78.5	9.0	1.00	100.0	–
Acetonitrile	37.5	5.8	0.37	81.6	100
Acetone	20.7	5.1	0.32	56.2	100
Methanol	32.6	5.1	0.60	64.6	100
Ethanol	24.3	5.2	1.20	78.4	100
Ethyl acetate	6.0	4.4	0.45	77.2	8.7
Diethyl ether	4.3	2.8	0.24	34.6	6.9
Dichloromethane	9.1	3.1	0.44	40.7	1.6
Toluene	2.4	2.4	0.52	110.8	0.05
Cyclohexane	2.0	0.2	0.98	80.7	0.01
Hexane	1.9	0.0	0.32	69.0	0.001

of the extract. The use of less solvent per fixed sample size by controlling pressure, temperature, and/or addition of gaseous co-solvents can also ease post extraction procedures (e.g. solvent evaporation and water removal in GC applications).

Despite the advantages of SFE, PLE, and MAE in some respects, traditional liquid-based methods at room temperature and pressure have advantages of their own. Traditional extraction methods are easy and do not require the added expense of purchasing and maintaining instrumentation. Fewer parameters to control and known expectations based on accumulated knowledge makes traditional method development less complicated. Liquids often serve as a good medium to homogenize samples and often limit analyte degradation and volatilization. In aqueous systems, the use of salts, acids, and bases often has a strong effect in extraction and partitioning. Unfortunately, the inherently wide polarity range of analytes in multiclass, MRMs handicaps techniques that permit a higher degree of selectivity. The optimal goal in any extraction is to match the polarity range of the analytes to be extracted, and not to exceed it.

At the present time, liquid-based extraction at room temperature and pressure remains the most common approach to extraction of pesticide residues. In the case of foods, a 2:1 liquid solvent-to-sample ratio is typical, and a blending device is employed in extraction. For environmental solid samples, Soxhlet extraction, sonication, and extended shaking with solvents are commonly used, and higher solvent-to-sample ratios are more common. Prior to the 1990s, water samples were more frequently extracted with liquid–liquid extraction (LLE) using an immiscible solvent, such as dichloromethane, but SPE has become the prevalent approach for water now. However, LLE can be adapted to use smaller volumes of organic solvents and increase convenience.[71]

The most common solvents used for extraction in multiclass, multiresidue analysis are acetone, acetonitrile, ethyl acetate, and methanol. For high-moisture samples, the water in the sample incorporates with miscible extraction solvents and helps to improve recoveries of many pesticides.[52] In the case of ethyl acetate (or dichloromethane and hexane), wet samples are frequently mixed with a drying agent, such as sodium sulfate, to remove water. In SFE, a similar procedure using diatomaceous earth or salt desiccants for wet samples is followed to prevent adverse effects of water[60,63,72] (supercritical CO_2 is only slightly miscible with water).

Water is considered the first choice as the extraction solvent for safety, cost, and environmental reasons, but water is not very useful by itself in typical multiclass, multiresidue applications. The less polar analytes are not extracted well with water, even at relatively high temperatures, but a host of problematic water-soluble matrix components often make cleanup more difficult.[1] Concentration steps are usually needed prior to analysis, and evaporation of water is very slow and relatively volatile pesticides are lost at conditions needed to evaporate water, even during lyophilization.[73,74] Also, many pesticides degrade more rapidly in water than in other solvents, and several other pesticides will either precipitate or otherwise interact with the glassware rather than remain in the water. It is possible to decrease the polarity of water by increasing the temperature and pressure, but then degradation is increased and polarity advantages of water are lost.

In extracts that contain water, however, the water should be concentrated or removed before analysis using conventional GC techniques (water is often not a problem in LC methods). Water can be separated from the mixture by exposing the solution to a drying agent,[75] addition of salt,[76] freezing-out the water phase,[77] and/or liquid–liquid partitioning with water-immiscible solvent(s). The organic phase is then usually concentrated by evaporation, which is why low-boiling solvents are desired to reduce the time of the extract concentration step. SPE is also utilized to separate the solutes from an aqueous solution. An advantage of SPE is that it concentrates the analytes as they partition out of the aqueous solution on to the solid sorbent. Thus, the concentration step after SPE is shortened or eliminated. Some polar analytes may break through SPE sorbents if the use of excessive aqueous volumes is attempted, but the problematic pesticides tend to be the same ones that only partially partition into the organic phase in LLE.[78]

The use of ethyl acetate as the extraction solvent avoids the water removal and partitioning steps, but ethyl acetate tends to be a less selective (or more exhaustive) extraction solvent than solvents that are completely miscible with water. Thus, ethyl acetate methods commonly require a solvent-intensive GPC step to remove co-extractives, even in the case of nonfatty matrices.[44] Less cleanup is usually needed for methods that use acetone and acetonitrile.[1,52] Acetonitrile is particularly useful in the case of fatty foods because fats have a lower solubility in acetonitrile than several other solvents.[79]

2.2.1 Solid-phase Microextraction

A novel approach using sorbents in relatively selective extractions is solid-phase microextraction (SPME).[80] Figure 5 shows a drawing of an SPME fiber. The key advantage of SPME is the simplicity of operation; all that is required is sequentially to place a fiber into a liquid or gaseous sample and then into a chromatographic injector. SPME works by coating a small fiber with a liquid polymer or solid phase and exposing this coated fiber to the sample. The analytes (and other chemicals)

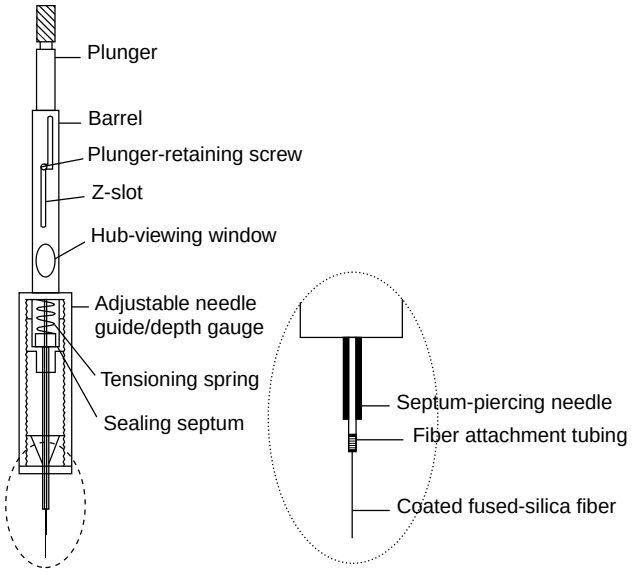

Figure 5 Diagram of an SPME fiber. [Reprinted with permission from Pawliszyn.[80]]

partition into the coating phase and, after a certain time, the SPME fiber is desorbed (thermally or with solvents) into a GC or HPLC system.[80] Liquid samples are nearly always water in SPME because pesticide analytes tend to remain in organic solvents rather than partition into the fiber (and the coating may be stripped off the fiber in strong solvents). Even in the case of aqueous samples, many pesticides do not generally partition completely into the fiber, and salts are typically added to "salt-out" the pesticides and speed the partitioning process.

Important factors to control in SPME (with fixed fiber parameters) include exposure times, temperature, volumes, and stirring (if necessary). The exposure time may be short (5 min) in some applications in which the analytes quickly partition into the sorbent, or longer times (>1 h) may be needed in an attempt to decrease the LOD and increase the range of analytes. Quantitation can be performed from fundamental thermodynamic equations if all parameters are known and the system is at equilibrium. This is also true if the kinetics of the system are well known, but it is easier and more common to quantify results empirically as in other approaches with the use of calibration standards. Matrix components can greatly affect partitioning and lead to inaccurate quantitation,[81] but the method of standard additions or use of standards in matrix may help overcome this problem.

In multiclass, multiresidue applications, SPME can be too selective and not possess a sufficiently wide polarity range to extract the targeted analytes. At the same time, SPME may not be selective enough because it may extract too many matrix interferents. At this time, automated SPME systems and a few different fiber coatings that cover different ranges in polarity are commercially available. SPME can be very useful in applications of relatively clean air and water matrices for a reasonable number of pesticides, but direct extraction of complex samples is often problematic with SPME. To reduce problems of nonvolatile matrix components in liquid samples, the fiber may be placed in the headspace of the liquid. The amount of pesticide analyte extracted will not be altered, but it may take an exceedingly long time for the equilibration process between the liquid, headspace, and fiber coating phases. Overly small sample volumes can also be a disadvantage in SPME, and consequently the LOD may also be too high in SPME for some applications. In the future, the use of a coated capillary, as in a GC column, rather than a coated rod as in SPME, will permit larger volumes and shorter analysis times by being able to flow the sample through the capillary.[82]

Owing the frequently less than 100% partitioning of analytes into the fiber, the recovery concept of extraction is not necessarily an appropriate indicator in SPME.[83] However, recovery has traditionally been an indicator of accuracy in a method. Therefore, SPME may only be used in limited quantitative applications for pesticide residues, but it may become a prominent screening approach owing to its ease of use and potentially fast operation.[84] SPME can also be very useful in determining properties of pesticides such as Henry's law constants and in process monitoring systems.[85]

2.2.2 Extraction of Gaseous Samples

Analytes are extracted from gases by precipitation under cold conditions (e.g. cryofocusing), adsorbed by a solid (e.g. SPE), or absorbed into a liquid (e.g. bubbling into a liquid). SPME (a liquid absorption or solid adsorption approach, depending on the fiber coating) can be useful in the extraction of small volumes of gases or in air screening applications, but traditional air sampling approaches remain prevalent in environmental air studies. The use of polyurethane foam (PUF) for large-volume air sampling is an effective approach for a variety of airborne pesticides.[13,86,87] Other common solid-phase adsorbents include Tenax®, carbon, silica-based sorbents, and polymers.[88] The combination of low temperature with adsorption or absorption extraction methods can be used to increase extraction efficiencies.

2.3 Cleanup

The separation of analytes from undesirable matrix components, or "cleanup" of sample extracts, can be accomplished through liquid–liquid partitioning, liquid–solid partitioning, size-exclusion chromatography (SEC) or GPC, dialysis, distillation, centrifugation, filtration, precipitation, immunoaffinity, or other techniques that take

advantage of differences in the physicochemical properties of the analytes and the interferents. Owing to the way in which analysts choose to compartmentalize the overall multiresidue, multiclass strategy, different laboratories that use the same extraction methods may diverge during cleanup for different pesticide groups.

2.3.1 Liquid–Liquid Partitioning

Liquid–liquid partitioning has been an effective cleanup technique for over a century, and is still commonly used today. However, its disadvantages include the typical use of large volumes of potentially hazardous solvents, the necessity for concentration steps afterwards, the generation of hazardous waste, labor-intensive and/or time-consuming procedures, potential occurrence of emulsions, the need for a laboratory hood and storage space, added external source of potential sample contamination, and the dishwashing requirements of the many pieces of glassware. The traditional wet chemistry methods are slowly being replaced by newer methods that are more easily automated, take less labor, space, and time, and use less solvent and glassware. Despite the encroachment of newer cleanup approaches, however, liquid–liquid partitioning will remain a viable approach.

Dichloromethane and ethyl acetate are particularly useful solvents for the cleanup of acid–base pesticides in aqueous extracts. Changing the pH of the aqueous phase can help either to separate ionic pesticides from neutral compounds or to separate neutral pesticides from ionic compounds.[25] The use of chlorinated solvents, such as dichloromethane, has become a concern among many laboratories for waste disposal and safety reasons, but its immiscibility with water, ease of evaporation, solubilizing power, and greater density than water (hence it settles as the lower phase in a separating funnel) make dichloromethane a difficult solvent to dismiss.

Liquid–liquid partitioning can help remove fats from water and/or acetonitrile solutions with hexane or immiscible ethers.[25,89] One of the drawbacks for removing lipids with liquid–liquid partitioning approaches in multiclass, MRMs, however, is that lipophilic pesticides also tend to partition into the same phase as the lipids. A partitioning factor, or *P* value, can be used to correct for the known percentages of affected pesticides that partition with the fats under predetermined conditions.[90] However, this can lead to higher detection limits (owing to the analyte losses) and greater margins of error. Also, methods should achieve pesticide recoveries >70% to meet most QA criteria.

2.3.2 Gel Permeation Chromatography

The development of GPC, also known as SEC, in the 1970s caused the use of liquid–liquid and liquid–solid partitioning to diminish.[91] GPC achieves fewer losses of pesticides when separating fats and other large molecules by taking advantage of the molecular size differences of the larger chemicals from the smaller pesticides. Because many pesticides tend to have relatively similar molecular sizes, they can generally be collected in a single fraction. Another advantage is that the GPC column can be re-used many times and is easily automated, unlike typical uses of adsorbent columns. Unfortunately, many matrix co-extractives of similar molecular size as the pesticides are not separated in GPC, and other cleanup techniques may still be required afterwards. Practical disadvantages of GPC include the very large volumes of potentially hazardous solvents involved, the generation of waste, the need for post-GPC concentration steps, and the costs associated with purchase and maintenance of instrumentation and columns.[27] In recent years, new gel materials, more benign solvent combinations, and miniaturized columns have improved practical aspects of GPC, and it remains the most common means to remove high- and very low-molecular-weight components from food and environmental samples.[92]

2.3.3 Distillation

Another way to remove fats and other relatively nonvolatile materials is sweep codistillation (SCD).[93] Essentially, SCD can be likened to preparative-scale GC in which the relatively volatile pesticides are separated from nonvolatile components. Nitrogen gas is generally the mobile phase and silanized glass beads or other materials may be used as the stationary phase. SCD was commercially developed in the 1960s, but owing to variability of results with the approach, few scientists continued to use the technique.

More advanced instruments were developed in the 1980s which regenerated interest in the approach.[85] Thermally labile pesticides may be lost at the temperatures involved (~250 °C) and separate cleanup may still be required in some applications. In the analysis of OCs in fatty matrices, SCD has been shown to provide equally valid results as GPC and Florisil adsorption. However, with the use of current commercial instrumentation, SCD gave a 10-fold higher sample throughput and 40-fold lower solvent consumption than the other two approaches.[94]

2.3.4 Solid-phase Extraction

The most common and diverse approach to cleanup (and extraction of water samples) in pesticide residue analysis is SPE.[95–97] For several decades, liquid–solid partitioning has been an option for analytical chemists to separate components from a solution, but prior to the introduction of commercial SPE cartridges in the 1970s, liquid–solid

partitioning was performed using columns prepared by the analyst in the laboratory. The choice of sorbents for cleanup at that time were essentially limited to normal-phase applications with alumina, Florisil, silica, and charcoal. The poor reproducibility of the sorbents from batch to batch, their need for activation/deactivation, and the manual column preparation procedures involved led to variable results.

The advent and growth of HPLC in the late 1960s and 1970s led to more choices and better reproducibility of stationary phases. Reversed-phase applications with C_8 and C_{18} stationary phases and the commercial introduction of premade disposable cartridges initially made SPE a useful technique for the extraction and cleanup of aqueous samples and extracts. Over the last 20 years, improvements and diversifications in SPE formats, sorbent types, and apparatus have made SPE a widely used approach for a variety of applications, including multiclass, multiresidue analysis of pesticides.

SPE cartridges or disks can be likened to very low-resolution HPLC columns in that similar stationary and mobile phases are used. However, HPLC is not always helpful in determining conditions for SPE, and vice versa. Typical particle sizes in SPE are 40 μm, and the plastic cartridges are generally packed with 0.1–1 g of sorbent in wide tubes, whereas higher density packing of 5-μm stationary phase particles in narrower columns is currently prevalent in HPLC. Larger LC columns and typical injection volumes > 1 mL can be very useful for the cleanup of extracts in semipreparative LC.[23,25,98] With automated systems, SPE can be likened to semipreparative LC because the procedures and materials can be similar. The greater separation power of semipreparative LC, however, provides higher potential for better selectivity by collecting heart-cut fractions for later analysis or direct analysis in coupled analytical systems.

The choice of reversed-phase, normal-phase, and ion-exchange media in SPE is very diverse, and Table 8 lists some of the more popular SPE applications for cleanup of multiple pesticides. The modes of operation involve all options of either retaining, or not retaining, the analytes and/or matrix interferences through the selection of appropriate sorbents and solvents. SPE formats can be cartridges or membrane extraction disks of various sizes designed for single- or bi-directional flow. Recently, SPE has even been commercialized in disposable pipet tips for convenient miniaturized applications.

Table 8 Common SPE sorbents used in multiclass, multiresidue analysis of pesticides

Normal phase	Reversed phase	Ion exchange
GCB	GCB	SAX
Alumina (acid, base, neutral)	ODS (C_{18})	PSA
Silica	Polymer (DVB, other)	Aminopropyl (NH_2)
Florisil	Cyanopropyl (CN)	SCX

GCB, graphitized carbon black; DVB, divinylbenzene; SAX, strong anion exchange; PSA, primary/secondary amine; SCX, strong cation exchange.

The introduction of GCB and polymer-based sorbents, such as DVB, has led to modifications of some multiclass, MRMs for pesticides in water, soils, and foods.[99,100] Carbon has been known as an excellent liquid–solid partitioning sorbent for many years, but carbon was notorious among chemists owing to the great number of charcoal types, sources, and irreproducibility. The advantages of certain polymer-based sorbents in reversed-phase applications over silica-based sorbents include higher possible flow rates, wider pesticide polarity range, wider pH range, and allowance of the cartridge to go dry without losses. This is useful in applications, because water can be removed from the extract by drying the cartridge with flowing gas prior to elution of the analytes with the chosen solvent(s).

The choice of sorbent–solvent combination in SPE for cleanup of extracts is much like choosing the extraction solvent for certain pesticide(s) in a particular matrix. A great deal of experience with SPE and knowledge about the chromatographic behavior of the analytes and matrix co-extractives can save much effort in method development using SPE. However, trial and error is the more common approach to method development using SPE because unpredictable results and/or subtle differences may occur between different stationary–mobile phase combinations. Also, in multiclass, multiresidue applications, the need to maintain a wide polarity range of analytes does not allow the analyst to achieve the best degree of selectivity for the analyte as is possible in SPE (unless the analyst chooses to divide the extract into a series of separate cleanup procedures).

This polarity range dilemma can lead to unconventional SPE methods. For example, the use of a carbon or alumina (traditionally normal-phase sorbents) with acetonitrile (a solvent more common in the reversed-phase format) may separate interferences from analytes reasonably well in multiclass, multiresidue applications.[100] Also, stationary phases used in ion-exchange approaches, which have the strongest interactions in water, are used with nonaqueous mobile phases in multiclass, multiresidue procedures.[51]

The use of liquid–solid partitioning with charcoal, silica, Florisil and/or alumina is also possible to aid in the removal of lipids, but semipolar pesticides require large elution volumes and many relatively polar pesticides are completely retained by these adsorbing phases. The elution volumes and recoveries of different pesticides

from different columns and mobile phases are outlined in PAM I.[1] The recoveries of a large number of pesticides using the ''Luke'' and other methods tested by the FDA are also listed in PAM I.

2.3.5 *Automation*

Off-line SPE approaches are generally conducted manually in batches of 8–12 using SPE vacuum manifolds. Laboratories that conduct routine analysis have begun to use automated approaches in general, and SPE is no exception. A major advantage of SPE is that it can be automated and conveniently combined with exaction and/or analysis in a single instrumental arrangement.[101] The use of on-line, or coupled, LC approaches (including SPE) permits automation of analyses to reduce labor and typically increase sample throughput.

In SFE, it is common to collect and concentrate the extracts on a solid-phase trap, and a higher degree of selectivity can be achieved by using a sorbent used in SPE applications, such as C_{18}, as the trapping material in SFE rather than an inert surface or solvent.[102] In pesticide and other analytical applications of SFE, it is often preferable to use an SPE sorbent to attain higher trapping efficiencies, lower solvent consumption, and greater selectivity. By eluting the analytes with a certain volume of appropriate solvent(s), a simple and convenient SFE/SPE procedure can be conducted. PLE instruments also may permit the automatic combination of extraction and cleanup. GPC has also been combined with SFE,[103] but automation and coupling of liquid–liquid partitioning methods are uncommon.

The approach to coupling techniques may be taken a step further by linking the SPE step with an analytical instrument.[104] The introduction of the extract into a GC or HPLC instrument immediately after SPE is possible, but a small concentration step is more commonly performed in the attempt to lower the LOD. Direct coupling of GPC to the analytical instrument is very rare in pesticide residue studies because large solvent volumes must usually be evaporated after GPC, and in the case of reversed-phase HPLC the types of solvents used are not typically amenable.

2.3.6 *Matrix Solid-phase Dispersion*

Another way to combine SPE with the extraction process is to utilize matrix solid-phase dispersion (MSPD).[105] In MSPD, the sample is mixed with a solid-phase material which separates matrix interferents from the pesticide analytes during extraction. Thus, cleanup is performed at the same time as the extraction.

A form of MSPD is nearly always conducted in SFE and PLE to disperse the sample, absorb water, and/or retain unwanted matrix components. A relatively inert sorbent, such as diatomaceous earth, is used for sample dispersion and water absorption, but Florisil, alumina, silica, or other sorbents may be used to retain matrix components. Conversely, the choice of sorbent–solvent combination and extraction conditions may be designed to retain the analytes and remove the interferents. This technique, dubbed inverse extraction, is followed with a second set of extraction conditions to extract the analytes after the interferents have been removed.

MSPD has been evaluated in pesticide applications,[106] but drawbacks include the small sample size involved ($\sim$0.5 g) and the high cost of sorbents, especially if the sample size is increased. MSPD in multiclass, multiresidue analysis of pesticides at most uses a 4 : 1 sample-to-sorbent ratio.[107] The sorbents are generally the most expensive material cost in multiclass, multiresidue approaches that involve SPE.

3 ANALYTICAL SEPARATIONS AND DETECTION

Any approach applicable to the analysis of organic molecules (and a few inorganics) may be useful in the analysis of pesticides, but MRMs (and even single analyte methods) nearly always require an analytical separation step prior to the detection of the different analytes. Certain screening techniques may not involve a separation step, but positive findings would then usually require a determinative method to quantify and identify the pesticide(s).[108] Techniques that involve optical spectroscopy[109] or electrochemistry[110] may provide a degree of selectivity, but such detection methods are rarely used in residue methods except as detection after chromatographic separations.[111]

At this time, nearly all analytical methods utilize GC or HPLC as the separation step, but other column chromatographic techniques and CE are also reasonable options for conducting separations. Planar or thin-layer chromatography (TLC) may be acceptable in some applications, but currently TLC is typically used as an analytical approach only when the analyst places a higher priority on sample throughput than on achieving the type of higher quality results possible with more modern methods. However, TLC can be very useful in screening applications owing to its potential for high sample throughput and ability to provide better separation than several other screening approaches.[112] Fast GC or flash GC, in which the GC conditions are optimized for speed, is another chromatographic option,[113] but the this type of approach has not been utilized in routine applications.

The development of supercritical fluid chromatography (SFC) and experimentation with mobile phases consisting of dense gases, supercritical fluids, and/or

liquids at high temperatures and pressure has led some analytical chemists to think in terms of ''unified chromatography''.[114] Fundamentally, there are no stationary phase boundaries between GC, HPLC, and SFC, and the mobile phases are actually a continuum of fluids. Just as pressure and temperature control have added new dimensions in extraction processes, the control of pressure and temperature in chromatography coupled with the diversity of mobile and stationary phase options may lead to extraordinary possibilities in separations. For example, SFC has been demonstrated to separate an exceptionally wide polarity range of pesticides using a single set of conditions.[115] Pesticides that are currently analyzed by either GC or HPLC may be determined with a single unified chromatographic system. At this time, the lack of rugged injection and sensitive analyte detection processes has been the limiting factor in pesticide residue applications in this type of approach.

For detection, MS is rapidly becoming the method of choice for multiclass, multiresidue analysis owing to its many advantages, recent improvements in technology, and availability of cost-effective commercial instrumentation. Detection systems in general are continually being improved and, in combination with the improvements in chromatographic instruments and techniques, exceptionally low LODs are possible for pesticide residues.

3.1 Gas Chromatography

GC, also sometimes referred to as gas–liquid chromatography (GLC), is the most common method for the analytical separation of pesticides in multiclass, MRMs. The possible number of theoretical plates in separations and the variety and capabilities of the detectors that can be coupled to GC usually make it the first choice for analysis. Some laboratories decide to limit the types of analyses conducted by using GC for nearly all analyses. For example, certain pesticides may be included in a multiresidue GC method despite the fact that HPLC may provide better results, such as in the case of thiabendazole.[9]

3.1.1 Derivatization

The analysis of thermally labile pesticides is not possible using GC methods unless the analytes are converted into a more volatile and stable form. Analysis of the analyte in its natural state is the preferred approach in any method, and derivatization leads to extra procedural steps and other potential complicating factors. However, until recent advances in LC/MS, chemical derivatization for GC analysis was essentially the only option in the case of thermally labile pesticides that did not give high absorbance or fluorescence signals. At this time LC/MS continues to become more sensitive, more rugged, and less expensive, but chemical derivatization remains a common approach for targeted thermally labile pesticides using GC analysis.

If derivatization is performed in multiresidue, multiclass strategies, a portion of the original extract is usually separated from the fraction that will be used in the analysis of the larger number of pesticides. The extracts may need separate cleanup steps, but it is sometimes possible to recombine the different extract fractions for analysis with the same GC method.[116] Common derivatization procedures for GC methods involve the alkylation of acids which may also serve to add functional groups that achieve highly sensitive detection with the chosen detector. Other examples include the use of pentafluorobenzyl bromide to link with carbendazim (degradation product of benomyl and thiophanate-methyl),[117] among other single-class pesticide analysis schemes.[118,119]

One problem with derivatization is the possible lack of selectivity in the reaction which can make a host of other derivatized compounds. In single analyte methods, this is not usually a problem, but in multiresidue separations, the other derivatized compounds are more likely to interfere with detection of other analytes. Highly resolved separations and/or universally selective detection methods, such as MS or AED, may overcome this problem, but it is more common to follow single-analyte methods when performing derivatization.

3.1.2 Gas Chromatography Separations

The selection of GC conditions for the analysis of pesticides mostly depends on the targeted pesticides and chosen detector(s). A variety of sources are available for relative retention indices of pesticides on different stationary phases and conditions.[1,27,120] The choice of the column may not make a significant difference in the final quantitation of most pesticides,[9] but relatively polar pesticides often give better peak shapes using relatively polar stationary phases, whereas nonpolar pesticides can give good peak shapes in a variety of columns. The most common GC stationary phases in pesticide analysis consist of dimethylpolysiloxane or methylpolysiloxane with 5–50% phenyl, cyanopropyl, or cyanopropylphenyl. The use of packed GC columns has virtually been universally replaced with the use of capillary columns, which may be narrow-bore (0.25 mm i.d.), wide-bore (0.32 mm i.d.) or mega-bore (0.53 mm), or of other sizes depending on the application. Capillary columns narrower than 0.25 mm i.d. are rarely used in food or environmental multiresidue applications, and guard columns (also known as retention gaps) are often useful in the analysis of complex extracts.

Injection of 1–3 μL of extracts is typical using split/splitless techniques, but on-column injection has

advantages for some pesticides. Larger injection volumes (0.01–1 mL) are now possible with some current GC injection devices, but large-volume and/or on-column/cold injection methods in GC typically require more column maintenance than split/splitless volume techniques. Recent advances in GC instruments in addition to large-volume injection consist of exceptionally fast separations, improved columns, electronic pressure control, retention time locking, and other capabilities. These topics are discussed in section 55 of this Encylopedia.

3.1.3 Matrix Enhancement Effect

An important issue in the multiresidue analysis of pesticide residues in GC, particularly for polar OPs, is the ''matrix enhancement effect''.[121,122] This effect occurs owing to polar functional groups on the pesticide analytes interacting with silanol, metal ions, and other active sites in the surfaces of GC systems (mainly from glass in the injection liner). In the presence of matrix co-extractives which tend to fill the active sites, the matrix enhancement effect results in an increased analyte injection efficiency (and consequently a higher signal) for certain pesticides in the presence of matrix than in solvent solutions which contain no matrix. Thus, if pesticide solutions only in solvent are used for calibration standards, the recoveries of matrix spikes and concentrations of the affected pesticides in sample extracts are systematically overestimated. For low-level spikes of affected pesticides, it is not unusual to observe enhancements of >150%. Table 9 provides results that show which pesticides are (and are not) affected by the matrix enhancement effect. These results are averaged from a number of food analyses over the course of time using different GC detectors as part of the Swedish MRM.[44] Owing to the nature of multiclass, multiresidue analysis, it is difficult to completely overcome this effect for polar OPs.[123] The use of deuterated internal standards, on-column or high-flow injection,[124] coating the injection liners with column phase, or method of standard additions can overcome or compensate for the matrix enhancement effect, but the most practical and common way to avoid the problem is to use calibration standards in blank matrix extracts.[27,34,44,122] Blank extracts are usually obtained from real samples in the monitoring program that have been shown to be free of residues, but blank field samples, which are often included in some studies anyway, may also be used as material for matrix-matched calibration standards.

3.1.4 Selective Detectors

For detection, the flame ionization detector (FID) and photoionization detector (PID) are rarely used in pesticide residue applications, in part, owing to the lack of selectivity, but the use of traditional selective detectors is a very common approach to multiresidue analysis of pesticides.[1] The use of a halogen-selective detector [ECD, electrolytic conductivity detector (ELCD), or halogen-specific detector (XSD™)] in conjunction with a phosphorus- [flame photometric detector (FPD), pulsed flame photometric detector (PFPD), NPD], sulfur- [FPD, PFPD, sulfur chemiluminescence detector], and/or nitrogen-selective detector (NPD, PFPD) is a common approach to cover a wide range of pesticides and to reduce sample cleanup. The AED is another element-selective detector, but it is useful for a number of elements simultaneously.[125,126] The PFPD is also useful for the selective detection of 28 elements (up to two simultaneously), and can provide lower LODs at a lower cost than an AED.[127]

Some multiclass, multiresidue schemes use up to five detectors in separate GC analyses which are selective for halogens, nitrogen, phosphorus, sulfur, and other elements, plus MS for confirmation. Thus, each sample extract requires five GC injections, plus HPLC methods, to analyze and identify the wide range of pesticides targeted for analysis. Some analysts choose to split the flow of a GC separation to two detectors which minimizes the number of injections, but this approach also reduces the amount of analyte that reaches the detector and complicates separations by using the same column for all analytes.[128]

3.1.5 Gas Chromatography/Mass Spectrometry

The most widely regarded approach to accomplish the analysis of as many pesticides as possible in as few steps as possible is to use MS detection.[129–132] MS is considered a universally selective detection method because it detects all compounds independent of elemental composition and further separates the signal into mass spectral scans to provide a high degree of selectivity. Unlike gas chromatography/atomic emission detection (GC/AED), gas chromatography/mass spectrometry (GC/MS) may provide acceptable confirmation of the identity of analytes without the need for further information. This reduces the need to reinject a sample into a separate GC system for pesticide confirmation. Through the use of selected ion monitoring (SIM), efficient ion-trap or quadrupole devices, and/or tandem mass spectrometry (MS/MS), modern GC/MS instruments provide LODs similar to or lower than those with selective detectors, depending on the analytes, methods, and detectors.

MS detection does not necessarily require as highly resolved GC separations as in the case of selective detectors because the likelihood of an overlapping mass spectral peak among pesticides with the same retention time is less than the likelihood of an overlapping peak from

Table 9 Average matrix effect (% recovery) for pesticides in food commodities using the Swedish method (from A. Andersson, H. Pålsheden, B. Arén, 'Matrix Effects in Pesticide Multiresidue Analysis', 1st European Pesticide Residue Workshop, Alkmaar, The Netherlands, June 10–12, 1996, P-055)

Pesticide	Detection	Normal matrix conc.		Matrix diluted 5×		Pesticide	Detection	Normal matrix conc.		Matrix diluted 5×	
		SE-30	OV-1701	SE-30	OV-1701			SE-30	OV-1701	SE-30	OV-1701
Acephate[a]	NPD	152	130	130	118	Fenson	ECD	106	107	103	105
Azinphos-ethyl[b]	NPD	114	115	109	113	Fensulfothion[a]	NPD	127	122	114	123
Azinphos-methyl[a]	NPD	135	128	121	116	Fenvalerate[a]	ECD	115	118	105	103
Bitertanol[a]	NPD	120	134	113	120	Folpet[a]	ECD	135	136	110	105
Bromophos[b]	NPD	116	106	105	105	α-HCH	ECD	96	93	99	99
Bromopropylate[a]	ECD	118	120	109	112	γ-HCH	ECD	98	97	100	103
Bupirimate	NPD	87	97	102	106	Heptenophos	NPD	108	100	105	100
Captan[a]	ECD	136	152	111	130	Hexachloro-benzene	ECD	95	100	99	104
Carbaryl[b]	NPD	120	109	113	108	Hexazinone[a]	NPD	116	125	114	116
Carbofuran[a]	NPD	133	108	113	108	Imazalil	NPD	91	96	105	107
Carbo-phenothion[a]	NPD	123	112	109	108	Iprodione[b]	ECD	129	158	107	125
Chlorfenvinphos	NPD	101	100	105	105	Leptophos[b]	ECD	112	120	106	113
Chlormephos	ECD	110	108	102	101	Malathion	NPD	103	102	104	103
Chlormephos	NPD	102	101	102	102	Mecarbam[b]	NPD	114	109	108	106
Chlorothalonil	ECD	102	106	102	102	Metalaxyl	NPD	106	110	104	110
Chlorpropham	NPD	111	107	106	106	Methamidophos[a]	NPD	139	123	118	116
Chlorpyrifos-methyl	NPD	106	107	105	102	Methidathion[a]	NPD	123	112	111	108
Chlorpyrifos	NPD	103	108	103	102	Parathion	ECD	103	105	100	102
Coumaphos[a]	ECD	122	144	110	114	Parathion	NPD	109	108	105	104
Coumaphos[a]	NPD	121	122	110	114	Parathion-methyl[a]	NPD	117	110	106	106
Cyanophos	NPD	109	106	106	104	Pentachloro-aniline	ECD	106	99	102	101
Cypermethrin[b]	ECD	113	117	106	106	Permethrin	ECD	105	112	101	103
p,p′-DDD	ECD	102	106	104	102	Phosmet[a]	NPD	121	118	112	114
p,p′-DDE	ECD	94	99	99	100	Pirimicarb	NPD	85	90	96	99
o,p-DDT	ECD	108	100	103	98	Pirimiphos-methyl	NPD	114	105	104	103
p,p′-DDT[b]	ECD	112	114	105	101	Procymidone	ECD	92	104	99	101
Deltamethrin[a]	ECD	122	119	109	106	Promecarb[a]	NPD	116	108	109	107
Diazinon	NPD	109	104	102	102	Propham	NPD	107	108	104	106
Dichlofluanid[b]	ECD	116	106	103	103	Propiconazole[b]	ECD	156	139	108	107
Dichloran	ECD	104	102	100	100	Pyrazophos[b]	NPD	115	117	106	116
Dicrotophos[b]	NPD	115	106	117	112	Quintozene	ECD	101	103	102	103
Dieldrin	ECD	98	93	101	101	1,2,5,6-TCA	ECD	93	94	97	101
Dimethoate[b]	NPD	126	117	110	120	Tecnazene	ECD	101	100	100	105
Diphenylamine	NPD	110	110	106	109	Tetradifon	ECD	108	109	105	102
Ditalimfos	ECD	105	106	106	102	Thiabendazole[a]	NPD	120	128	110	118
Ditalimfos	NPD	105	105	104	105	Tolyfluanid	ECD	107	112	104	106
Endosulfan I	ECD	98	99	100	100	Triademefon	NPD	106	104	105	107
Endosulfan II	ECD	103	101	100	101	Triademenol[a]	NPD	119	119	111	109
Endosulfan sulfate[a]	ECD	109	116	101	102	Triazophos	NPD	115	112	111	111
Ethion[b]	NPD	112	110	105	108	Vinclozolin	ECD	100	103	101	102
Fenchlorphos	NPD	110	106	103	103						

NPD, nitrogen–phosphorus detector; ECD, electron-capture detector.
[a] Mean >115%.
[b] Mean 111–115%.

the same element. Unfortunately, this advantage cannot always be optimized because in SIM and current gas chromatography/tandem mass spectrometry (GC/MS/MS) methods, the ability to detect two analytes at once increases the LOD of each analyte. Overlapping full-scan mass spectra can also pose problems with confirmation, so baseline resolution of as many pesticides as possible is still desirable. Low-bleed capillary columns are typically used in GC/MS systems to reduce background noise, but until recently, only nonpolar phases had low-bleed properties. The commercial introduction of relatively polar low-bleed phases has improved the capability to achieve good separations and peak shapes for the relatively polar pesticides at the same conditions for nonpolar pesticides.

Another reason that it is difficult in practice to achieve the lowest LOD possible for more than about 100 pesticides in a single injection by MS and MS/MS is that some pesticides provide lower LODs when using positive ion chemical ionization (PCI) while others are better detected in electron impact ionization (EI) or negative ion chemical ionization (NCI) modes.[27] It is not uncommon for analysts to use a series of injections in GC/MS to best cover the large number of pesticides (>300) possible with the approach.[129,133] Sequential EI and PCI modes used in the same run is possible with some instruments, but this approach is seldom used.

The precision of GC/MS for analysis of pesticide residues generally varies from 5 to 20% RSD, which is similar to that in approaches using other GC detectors.[9] Quantitation with modern GC/MS instruments suitably meet the needs of most quantitative applications. However, some chemists have decided to use MS only for applications involving the determination of unknowns, confirmation of pesticide identities, and/or rapid screening.[134]

3.1.6 Pesticide Identification

Confirmation criteria in pesticide residue analysis are a subject of debate, depending on the application, but it is generally recognized that human judgment is required in the final decision. Criteria to consider in MS include: (1) proper GC retention time; (2) three ions of correct mass/charge (m/z) ratios; (3) adequate signal-to-noise ratios; and (4) absence of similar signal in blank samples (to indicate no contamination or carryover).[135] The definitions of ''proper'' ''correct'' and ''adequate'' are based on the need and importance of the result. In any case, some pesticides do not provide three ions in their spectra, or the ion ratios for some spectra may be more variable than deemed acceptable if strict definitions are used. Other factors are sometimes taken into account, such as chromatographic peak shape, sample history, and/or analyst experience. In these difficult cases, information from other analyses, perhaps using other ionization modes in MS or even infrared spectroscopy,[111] are sometimes performed to gather more evidence. The combination of an element-selective detector and MS can provide excellent information for determining the presence of an analyte, or just as important, eliminating the chance that a pesticide is present.[126] The added selectivity gained in the use of multiple mass spectrometry (MS^n) can also weigh heavily in confirmation of pesticide identity despite the possible lack of three ions of defined ratios.[132,133] Some older methods of confirmation, such as the use of two different stationary phases in GC, are no longer acceptable in many cases, especially if the results will be used for regulatory purposes.

3.1.7 Direct Sample Introduction

Conventional wisdom maintains that elimination of cleanup steps in an analysis reduces the quantitative capabilities and/or ruggedness of the analysis. The effects of matrix and/or lack of ruggedness are often the biggest limitation of using the modern sensitive and selective detectors in GC analysis of complex extracts. The detectors may be capable of discriminating low-level analytes from potentially high-background interferents, but the GC injector can become contaminated with nonvolatile components after only a few injections unless cleanup is performed. A recent advance, direct sample introduction or ''dirty sample'' injection (DSI), has addressed these factors.[136,137]

DSI improves the ruggedness of the approach because nonvolatile components that can contaminate GC columns and MS sources remain in the disposable sample vials used for injection. Figure 6 shows a drawing of a DSI device inside a GC injection port. In DSI, samples are extracted by the means available, and up to 30 μL of the liquid extracts are placed in a disposable vial. Essentially, DSI operates by thermal desorption of sample extracts from the vial placed in the injection port into the GC column. This is not unlike SPME, which also does not contaminate the GC system (because nonvolatile components remain in the fiber coating during injection), but extraction methods in DSI can provide a wider analyte polarity range than SPME. This is because DSI operates solely on analyte volatility whereas SPME depends on partitioning ratios into the fiber coating. Furthermore, the expensive, reusable SPME fiber remains contaminated with nonvolatile materials after injection whereas the inexpensive microvial in DSI is disposable.

The DSI approach for complex extracts requires the use of selective detection to discriminate the pesticide analytes from the other semivolatile matrix components in the extract, but the use of selective detectors, such as PFPD,[137] and also MS[136] and MS^n,[138] can meet this challenge. The quantitative aspects of DSI

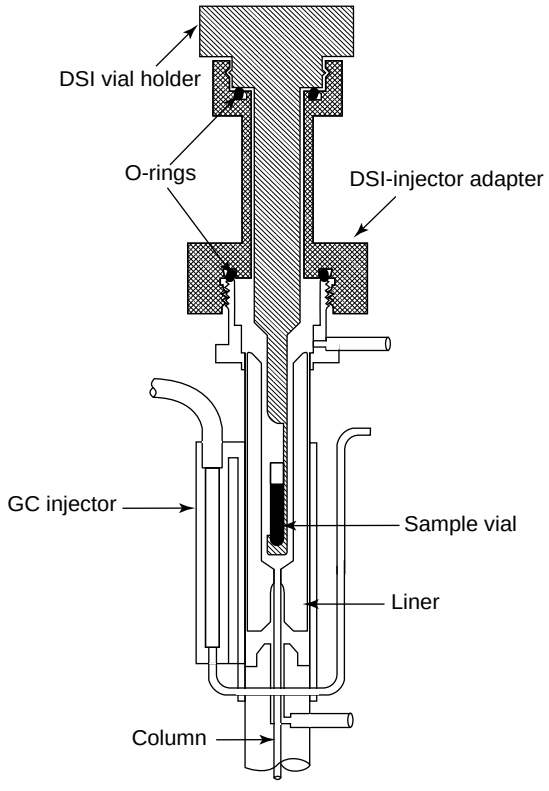

Figure 6 Diagram of a DSI device. [Reprinted with permission from A. Amirav, S. Dagan, *Eur. Mass Spectrom.*, **3**, 105–111 (1997).]

with GC/MS/MS has been demonstrated for 23 diverse pesticides in fruits and vegetables.[138] Recoveries were near 100% (ca. 10% RSD) with the approach, and the LODs were $<1\ \text{ng g}^{-1}$ for the pesticides.

3.2 High-performance Liquid Chromatography

The common need to achieve separation of so many analytes in multiclass, multiresidue applications tends to preclude HPLC as the primary means of analytical separations. However, HPLC is often an integral part of an overall analytical scheme in which the approach is used for the analysis of single classes of pesticides not amenable to GC.[1,27,53] Typically, reversed-phase high-performance liquid chromatography (RPHPLC) is employed for the thermally labile, semipolar and polar pesticides. Normal-phase HPLC is capable of separating nonpolar and semipolar pesticides,[139] but these types of pesticides are not typically thermally labile and the GC methods often give better separations and lower detection limits. Some laboratories may choose to derivatize thermally labile pesticides for an overall method relying on GC analysis,[140] but it is generally preferable to detect the analytes as they occur in the sample. Thus, HPLC is a common alternative to GC derivatization methods. Liška and Slobodník reviewed and compared the use of GC and LC methods for the analysis of polar pesticides in water.[141]

HPLC is a more rugged approach than GC owing to the continual rinsing of the system with liquid and simple means of liquid sample introduction. HPLC permits the injection of larger sample volumes (typically 25–250 μL) than GC (typically 1–3 μL), which also helps to improve the sensitivity of HPLC methods. Despite this aspect, LODs in GC methods with selective detectors may be lower than those in HPLC methods (depending on the analysis), despite the much larger HPLC injection volume. A disadvantage of HPLC is the potentially large amount of mobile phase solvent needed and waste generated. However, solvent recycling, newer and smaller packing phases, narrower columns, and slower flow rates are reducing this potential drawback. The articles on HPLC (section 59 of this Encyclopedia) should be consulted for a discussion of these and other aspects of HPLC.

3.2.1 Detectors

HPLC detectors are not as varied as GC detectors. Common HPLC detectors used in pesticides residue analysis include ultraviolet/visible (UV/VIS), fluorescence, and MS detectors, whereas electrochemical detectors are less commonly used. Currently, single-wavelength UV/VIS absorbance is a common approach to detection in HPLC, but this approach is well noted for its general lack of selectivity, especially at low wavelengths, and inability to detect analytes that do not contain chromophores. The diode-array detector (DAD), also called photodiode array (PDA), provides a higher degree of selectivity, but it is also limited to the detection of analytes with chromophores and is usually associated with higher LODs than single-wavelength detectors. Fluorescence is a much more selective and sensitive option for HPLC detection, but few pesticides fluoresce strongly enough to warrant a separate method for their detection, especially in multiclass, MRMs.[109] A beneficial feature of spectrophotometry and spectrofluorimetry, however, is their nondestructive nature, which enables them to be combined in series with themselves[142] and/or other detection methods (such as MS) to increase the overall selectivity.

3.2.2 Post-column Derivatization

Even though certain thermally labile and nonvolatile pesticides may be amenable to separation in HPLC, their detection can be difficult with traditional detectors if they do not provide a sufficient absorbance, fluorescence, or electrochemical signal. Derivatization of the analyte to improve detection (with any detector) is sometimes the only option to meet desired the LOD. If the pesticide

analyte can be derivatized in a rapid reaction, then postcolumn derivatization can be a useful method for detection in HPLC. Chemical derivatization steps during the sample preparation are an option prior to HPLC analysis, and have the added benefit of permitting longer reaction times, but traditionally GC would be used if HPLC postcolumn derivatization is not feasible. Postcolumn derivatization eases the sample preparation process because separate derivatization steps are not needed, and the pesticides have already been separated by the analytical HPLC column. Because relatively few other reactive compounds are likely to be present at the analyte retention times, the selectivity of the overall approach is further increased.

Fluorescence detection is the most common postcolumn derivatization approach owing to the high sensitivity and selectivity of the detector. A common approach in the postcolumn derivatization process is the reaction of *o*-phthaldialdehyde with amines. Heat or ultraviolet (UV) radiation applied to a reaction bed or length of narrow tubing may be used to drive the reaction. Postcolumn derivatization fluorescence detection in HPLC is useful in the analysis of several amine-containing pesticides, such as carbamates, ureas, and certain organophosphates.[143] In particular, postcolumn derivatization fluorescence detection has become the standard approach for the analysis of *N*-methylcarbamate insecticides.[1,26,27,53,143]

3.2.3 Liquid Chromatography/Mass Spectrometry

Owing to the rapid pace of major advances in LC/MS, a number of reviews of LC/MS and its use in the analysis of pesticides has been published in recent years.[144–146] There has been a strong need for a universal and selective detector such as MS in LC (and other) applications since the advent of the technique. In the 1980s, moving belt, particle beam (PB), and thermospray (TSP) interfaces were used with limited success in the analysis of pesticides.[147] With the improvements and commercialization of liquid nebulization interfaces and atmospheric pressure ionization (API) for MS, the need for derivatization of thermally labile pesticides for GC or HPLC detection is expected to wane.[148] Applications of liquid chromatography/particle beam mass spectrometry (LC/PBMS) and liquid chromatography/thermospray mass spectrometry (LC/TSPMS) for pesticides have essentially been replaced with techniques using atmospheric pressure chemical ionization (APCI) and electrospray (ESP) ionization, also known as ionspray (ISP), techniques.[149] Modern LC/MS instruments capable of ESP and APCI modes, and that now fit on a benchtop, provide lower LODs, more facile and rugged operation, and lower costs than ever before. LC/MS instruments are still several times the price of a traditional system, but the identification and universal detection capabilities of MS (or MS^n) make it a powerful analytical tool.

Many of the same general advantages and disadvantages of GC/MS are prevalent in high-performance liquid chromatography/mass spectrometry (HPLC/MS), but a higher degree of complexity and more limitations are associated with HPLC/MS. Most notably, the current main ionization modes in LC/MS (ESP, APCI) can give very different ionization levels and responses for different pesticides.[150] This is further complicated by the choice of positive or negative ionization in each of the two ionization modes which leads to a choice of four possible LC/MS operating conditions (+/− ESP, +/− APCI), none of which can currently be interchanged during the course of a run. This further limits the utility of LC/MS for multiclass, multiresidue pesticide methods in a single run. Furthermore, the use of the soft ESI and APCI ionization techniques and easy formation of analyte–salt adducts mean that separate injections from the quantitation method may still be needed to conduct confirmation through collision-induced dissociation (CID) of the analytes. Also, owing to salt build-up in the interface, high concentrations of nonvolatile salts should not be added to the mobile phase in LC/MS, thereby diminishing the ability to optimize buffers and ion-pairing reagents which are helpful in some pesticide applications.[142] Also, lower flow rates are often used in LC/MS, particularly with ESP unless split flow or other precautions are taken, which often necessitates the use of more precise pumps and smaller columns than in traditional HPLC applications. However, this aspect may in fact be advantageous because narrow columns, smaller stationary phase packings, and low flow rates in HPLC separations may lead to improved LC/MS applications for pesticides.

3.2.4 Signal Suppression Effect

Another drawback in LC/MS is the signal suppression effect that is prevalent when an analyte co-elutes with a matrix component.[151] MS detection can isolate the analyte from the co-eluting peak through the proper choice of mass(es), but the quantitation can be significantly affected by a reduction in the ionization efficiency of the analyte prior to MS detection. Unlike the matrix enhancement effect in GC, which can lead to overestimation of the calculated concentration for affected pesticides, the signal suppression effect in LC/MS leads to lower resulting calculated concentrations. The methods of reduction or control of the effect may be possible through one or more of the following approaches: (1) further cleanup of the extracts; (2) alteration of the separation conditions; (3) utilization of a different ionization technique; (4) addition of deuterated internal standards; (5) use of the method of standard additions;

and (6) use of calibration standards in a blank matrix. Another approach is to correct results for recoveries, but this is not acceptable in many QA protocols, and the results are typically inaccurate when the effect is concentration dependent.

3.2.5 Liquid Chromatography/Tandem Mass Spectrometry

The use of liquid chromatography/tandem mass spectrometry (LC/MS/MS) (or MS^n) does not overcome the signal suppression effect in the presence of interferents, but it can be very useful in avoiding interfering masses and reducing the LOD when two peaks co-elute.[152] There are two types of instruments for MS/MS detection, either tandem quadrupole or ion-trap, that each have advantages and disadvantages with respect to each other. Triple-stage quadrupole MS instruments generally have better precision, ability to determine neutral losses (which is useful in making identifications), and higher mass resolution. The ion-trap instruments are generally smaller, cheaper, and provide MS^n capability. LC/MS/MS (or MS^n) is considered the state-of-the-art approach for many of the modern pesticides and their metabolites, and pesticide manufacturing companies have relied heavily on the use of the LC/MS approaches in the development and analysis of pesticides. As single-analyte methods are developed for pesticide registrations using LC/MS approaches, the application of these techniques in multiclass, MRMs is expected to increase.

3.3 Capillary Electrophoresis

CE is a technique developed in the early 1980s, but it has only recently made its way into pesticide residue applications.[153–156] A variety of formats and options for different types of applications are possible in CE, such as micellar electrokinetic chromatography (MEKC), isotachophoresis, capillary gel electrophoresis (CGE), and capillary electrochromatography (CEC), but biochemical and pharmaceutical applications, not pesticide residues, remain the major focus in studies that use CE.[157] The main problem with CE for residue analysis of small molecules has been the low sensitivity of detection in the narrow capillary used in the separation. With the development of extended detection path lengths and special optics, absorbance detection can give reasonably low detection limits with clean samples.[156] However, complex samples can be very difficult to analyze using CE/UV. CE coupled with laser-induced fluorescence detection can provide extraordinarily low LODs[158] but the analytes must be fluorescent with excitation peaks at common laser wavelengths for this approach to work. Derivatization of the analytes with appropriate fluorescent labels may be possible, as is done in biochemical applications, but pesticide analysis has not been such an important application to utilize such an approach.

As in HPLC, the coupling of MS detection with CE has provided an excellent opportunity for more selective analysis, but the much reduced flow rates, small injection volumes, limitations in the types of buffers used, and the need to maintain a large voltage potential along the capillary make CE/MS (capillary electrophoresis/mass spectrometry) more complicated than LC/MS.[155] Furthermore, 1000-fold lower injection volumes in CE lead to higher LODs in CE/MS than in LC/MS. Techniques such as stacking multiple CE injections can increase injection volumes to possibly overcome the sample injection volume limitation in CE.[153,156,157]

The main advantages of CE over HPLC relate to the greater number of theoretical plates, reduction in the use of hazardous solvents, lack of waste, utility in sample limited applications, and general capability for lower LODs with UV and fluorescence detectors (owing to increased resolution in separations and higher signal-to-noise ratio despite the smaller injection volume).[25] Unlike the laminar flow profile of chromatographic techniques, plug flow occurs in CE because flow originates along the capillary walls. This generally translates into sharper peaks in CE and better separations.

A common criticism of CE is the lower consistency of migration times for peaks than in HPLC or GC. The effects of matrix components and small differences in ionic strength and pH can have significant effects in the separation. The use of a migration time marker provides a consistent relative migration time that may be used to identify peaks of interest, especially in complex electropherograms.[25] As in HPLC, it is often difficult to separate a diverse range of pesticides in CE with the same set of conditions. Hence, CE applications are most likely to be limited to single-class MRMs.[154] CE is most useful for ionic pesticides, but neutral analytes are also possible targets using CEC and MEKC.[159]

3.4 Immunochemical Methods

Immunochemical techniques may be useful for fairly rapid and reasonably inexpensive techniques for selected pesticides, or perhaps a selected class of pesticides, but the application of immunochemistry in multiclass, MRMs is limited.[160–163] The advantages of the high selectivity, reasonable ease, and relatively low costs are lost when several antibodies are needed to increase the analytical range of the technique. As in nearly all approaches, the selectivity of immunochemical techniques decreases as the analytical range increases, and costs and complexity concomitantly increase. Furthermore, the results of immunoassays are often only semiquantitative, and

separate methods may be needed for quantitation and identification. Furthermore, immunochemical techniques are only useful in aqueous solutions, and pH, ionic strength, matrix components, temperature, and other factors can have large effects in the assays. However, some enzyme-linked immunosorbent assays (ELISAs) are satisfactorily quantitative, rugged, and capable of detecting very low levels (<1 ppb) of targeted pesticides. For example, commercial ELISAs for triazine (e.g. atrazine) and acetanilide (e.g. alachlor) herbicides have performed particularly well in environmental monitoring applications.

In ELISA, relating signal to the concentration is more complicated than traditional means because the calibration plot is inversely proportional and semilogarithmically correlated to the analyte concentration. Also, the signal in ELISA consists of the absorbance of a sample divided by the absorbance from blank sample(s). Furthermore, the analytical range for quantitation in ELISA is generally less than two orders of magnitude. For these reasons, ELISA is not especially useful for quantitative analysis, especially with complex extracts, but its capabilities can be useful in screening applications for individual pesticides. Cleanup of complex extracts has been performed prior to ELISA to improve reliability and quantitation, but in this case it is often less expensive and more reliable to use a traditional method in the laboratory.[164]

At this time, ELISA kits are commercially available for approximately 25 pesticides, but several other antibodies have been developed for other pesticides in noncommercial applications. Depending on the format, ELISA kits typically cost \$5–10 per assay, which is similar to the cost of materials in typical quantitative and confirmatory multiclass, MRMs (Table 6). ELISA tests are often performed in batches and can take from 15 to 120 min, depending on the incubation times for the competitive reaction and enzymatic color change. A batch can consist of 96 samples in the microtiter plate format, but usually fewer samples are analyzed owing to the need for replicates. In the commercial magnetic bead format, replicates are less necessary owing to lower variability, and up to 60 samples may be analyzed in a batch. In some cases, portable ELISA kits can be applied in the field using tube formats with detection by eye or with tube photometers.

Interestingly, many of the disadvantages with ELISA in residue analysis are also advantages, and vice versa, depending on the application. For example, the high selectivity of detection in ELISA is a disadvantage in multiresidue applications. The costs and labor of ELISA increase significantly if more than one kit is to be used to detect a number of pesticides. Also, the use of ELISA can be considered slow and tedious compared with SPME, DSI, or MS screening applications, but it is generally much easier and faster than single-analyte methods. However, the semiquantitative and nonconfirmatory nature of ELISA can potentially lead to a high rate of false positives and false negatives. Even if the results are reliable, a determinative and/or confirmatory method is often required in the case of a positive response.

The capabilities of immunochemical methods are highly dependent on the antibody. A rugged antibody of high selectivity can provide exceptional results for the targeted analytes in complex extracts that may even contain a small percentage of organic solvents in the aqueous solutions. In other cases, the antibodies may have poor stability and be affected strongly by pH, salt content, temperature, and matrix components. Some antibodies are specific to a single antigen, whereas others are selective to a chemical moiety, and still others can interact rather nonselectively with a variety of compounds. By altering incubation times and antibody-to-sample-to-enzyme–antigen ratios, the sensitivity of analysis can be tailored to achieve very low, or not as low, detection limits.

Investigations of immunochemical or biochemical formats that are potentially faster, more quantitative, easier to perform, less expensive, more rugged, and suitable for a wider range of pesticides than ELISA are being conducted. These approaches involve enzymes, biosensing, flow injection, optical fibers, molecular imprinting, and other methods.[165–167] Biosensing methods of analysis that take advantage of modes of action of different pesticides, e.g. cholinesterase inhibition, can be very effective but, unfortunately, current cholinesterase-based detection kits for OPs and carbamates are not sensitive enough for most monitoring applications.[168] Sensitive, specific, rapid, and reliable screening assays based on immunochemical or biosensing concepts are commercially available and widely used in clinical and veterinary testing applications. However, the analysis of pesticides has not warranted such attention from manufacturers, mainly owing to the need for multiclass, multiresidue results rather than those for specific analytes. Unless these type of techniques for pesticide residues become commercially available, they will only rarely be used because the process to prepare the biosensing or immunological materials is often expensive, time-consuming, and complicated. Screening methods are often desired to increase the number of monitored samples, but fast and accurate analytical methods other than biosensing or immunological techniques are more likely to be developed for a wide range of pesticides to meet the analytical goals.

4 CONCLUSIONS

Owing to the introduction of a number of sophisticated technologies and instruments, tremendous improvements

Table 10 CAS numbers, usage, classes, and solubilities in water [from Hornsby et al.[31]] for many pesticides

Common name	CAS No.	Use[a]	Class	Solubility (mg L^{-1})
Acephate	30560-19-1	I	OP	818000
Acifluorfen	62476-59-9	H	Phenoxy	250000
Alachlor	15972-60-8	H	Acetamide	240
Aldicarb	116-06-3	I	Carbamate	6000
Aldicarb sulfone	15972-60-8	I	Carbamate	10000
Aldrin	309-00-2	I	OC	0.027
Allidochlor	93-71-0	H	Acetamide	20000
Ametryn	834-12-8	H	Triazine	185
Aminocarb	2032-59-9	I	Carbamate	915
Amitraz	33089-61-1	I	Imidamide	1
Amitrole	61-82-5	H	Triazole	360000
Anilazine	101-05-3	F	Triazine	8
Atrazine	1912-24-9	H	Triazine	33
Azinphos-methyl	86-50-0	I	OP	29
Bendiocarb	22781-23-3	I	Carbamate	40
Benfluralin	1861-40-1	H	Nitroaniline	0.1
Benodanil	15310-01-7	F	Anilide	20
Benomyl	17804-35-2	F	Imidazole	2
Bensulfuron-methyl	83055-99-6	H	Sulfonylurea	120
Bensulide	741-58-2	H	Sulfonamide	5.6
Bifenox	42576-02-3	H	Phenoxy	0.398
Bifenthrin	82657-04-3	I	Pyrethroid	0.1
Bromacil	314-40-9	H	Uracil	700
Butachlor	23184-66-9	H	Acetanilide	23
Butylate	2008-41-5	H	Thiocarbamate	44
Captafol	2425-06-1	F	Imide	1.4
Captan	133-06-2	F	Imide	5.1
Carbaryl	63-25-2	I	Carbamate	120
Carbendazim	10605-21-7	F	Imidazole	8
Carbofuran	1563-66-22	I	Carbamate	351
Carbophenothion	786-19-6	I	OP	0.34
Carboxin	5234-68-4	F	Anilide	195
Chlorbromuron	13360-45-7	H	Phenylurea	35
Chlordane	57-74-9	I	OC	0.06
Chlordimeform HCl	6164-98-3	I	Imidamide	500000
Chlorimuron-ethyl	90982-32-4	H	Sulfonylurea	1200
Chlorbenzilate	510-15-6	I	Benzilate	13
Chloroneb	2675-77-6	F	Thiocarbamate	8
Chlorothalonil	1897-45-6	F	OC phthalonitrile	0.6
Chloroxuron	1982-47-4	H	Phenylurea	2.5
Chlorpropham	101-21-3	H	Phenylcarbamate	89
Chlorpyrifos	2921-88-2	I	OP	0.4
Chlorpyrifos-methyl	5598-13-0	I	OP	4
Chlorsulfuron	64902-72-3	H	Sulfonylurea	7000
Chlozolinate	84332-86-5	F	Thiocarbamate	1
Clofentezine	74115-24-5	I	Tetrazine	0.1
Clomazone	81777-89-1	H	Oxazolidinone	1100
Cryolite	15096-52-3	I	Aluminofluoride	420
Cyanazine	21725-46-2	H	Triazine	170
Cycloate	1134-23-2	H	Thiocarbamate	95
Cyfluthrin	68359-37-5	I	Pyrethroid	0.002
λ-Cyhalothrin	91465-08-6	I	Pyrethroid	0.005
Cyhexatin	13121-70-5	I	Organotin	1
Cypermethrin	52315-07-8	I	Pyrethroid	0.004
Cyromazine	66215-27-8	I	Triazine	136000
2,4-D	94-75-7	H	Phenoxy	890
Dacthal	1861-32-1	H	OC carboxylate	0.5
2,4-DB	94-82-6	H	Phenoxy	46

Table 10 (*continued*)

Common name	CAS No.	Use[a]	Class	Solubility (mg L^{-1})
DDD	72-51-8	I	OC	0.02
DDE	72-55-9	I	OC	0.1
DDT	50-29-3	I	OC	0.0055
Demeton	8086-48-3	I	OP	60
Desmedipham	13684-56-5	H	Carbamate	8
Diazinon	333-41-5	I	OP	60
Dichlobenil	1194-65-6	H	Benzonitrile	21.2
Dichlone	117-80-6	F	Quinone	0.1
Dichlorvos	62-73-7	I	OP	10000
Diclofop-methyl	51338-27-3	H	Phenoxy	0.8
Dicloran	99-30-9	F	Nitroaniline	7
Dicofol	115-32-2	I	OC	0.8
Dicrotophos	3738-78-3	I	OP	1000000
Dieldrin	60-57-1	I	OC	0.2
Diflubenzuron	35367-38-5	I	Benzamide	0.08
Dimethoate	60-51-5	I	OP	39800
Dinitramine	29091-05-2	H	Nitroaniline	1.1
Dinoseb	88-85-7	H	Nitrophenol	52
Dioxacarb	6988-21-2	I	Carbamate	6000
Diphenamid	957-51-7	H	Acetamide	260
Disulfoton	298-04-4	I	OP	25
Diuron	330-54-1	H	Phenylurea	42
Endosulfan	115-29-7	I	OC	0.32
Endrin	72-20-8	I	OC	0.23
EPN	2104-64-5	I	OP	0.5
EPTC	759-94-4	H	Thiocarbamate	344
Esfenvalerate	66230-04-4	I	Pyrethroid	0.002
Ethalfluralin	55283-68-6	H	Nitroaniline	0.3
Ethion	563-12-2	I	OP	1.1
Ethoprop	13194-48-4	I	OP	750
Etridiazole	2593-15-9	F	Thiadiazole	50
Fenamiphos	22224-92-6	I	OP	400
Fenbutatin oxide	13356-08-6	I	Organotin	0.0127
Fenfuram	24691-80-3	F	Furanilide	100
Fenitrothion	122-14-5	I	OP	30
Fenoxaprop-ethyl	66441-23-4	H	Phenoxy	0.8
Fenoxycarb	79127-80-3	I	Carbamate	6
Fenpropathrin	39515-41-8	I	Pyrethroid	0.33
Fensulfothion	115-90-2	I	OP	1540
Fenthion	55-38-9	I	OP	4.2
Fenuron	101-42-8	H	Phenylurea	3850
Fenvalerate	51630-58-1	I	Pyrethroid	0.002
Ferbam	14484-64-1	F	Thiocarbamate	120
Fluazifop-butyl	69806-50-4	H	Phenoxy	2
Fluazifop-*P*-butyl	79241-46-6	H	Phenoxy	2
Fluchloralin	33245-39-5	H	Nitroaniline	0.9
Flucythrinate	70124-77-5	I	Pyrethroid	0.06
Fluometuron	2164-17-2	H	Phenylurea	110
Fluridone	59756-60-4	H	Pyridinone	10
Fluvalinate	102851-06-9	I	Pyrethroid	0.005
Fomesafen	72178-02-0	H	Phenoxy	700000
Fonofos	944-22-9	I	OP	16.9
Formetanate HCl	23422-53-9	I	Imidamide	500000
Haloxyfop-methyl	69806-40-2	H	Phenoxy	43
Heptachlor	76-44-8	I	OC	0.056
Hexachlorobenzene	118-74-1	F	OC	0.005
Hexazinone	51235-04-2	H	Triazine	33000
Hexythiazox	78587-05-0	I	Carboxamide	0.5
Hydramethylnon	67485-29-4	I	Pyrimidinone	0.006

(*continued overleaf*)

Table 10 (*continued*)

Common name	CAS No.	Use[a]	Class	Solubility (mg L^{-1})
Imazalil	35554-44-0	F	Imidazole	1400
Imazapyr	81334-34-1	H	Imidazolinone	11000
Imazaquin	81335-37	H	Imidazolinone	60
Imazethapyr	101917-66-2	H	Imidazolinone	200000
Iprodione	36734-19-7	F	Imidazoline	13.9
Isazofos	42509-80-8	I	OP	69
Isofenphos	25311-71-1	I	OP	24
Isopropalin	33820-53-0	H	Nitrocumidine	0.1
Isoxaben	82558-50-7	H	Benzamide	1
Lactofen	77501-63-4	H	Phenoxy	0.1
Lindane	58-89-9	I	OC	7
Linuron	330-55-2	H	Phenylurea	75
Malathion	121-75-5	I	OP	130
Mancozeb	8018-01-7	F	Thiocarbamate	6
Methamidophos	10265-92-6	I	OP	1000000
Methiazole	20354-26-1	H	Oxadiazolidine	1.5
Methidathion	950-37-8	I	OP	220
Methiocarb	2032-65-7	I	Carbamate	24
Methomyl	16752-77-5	I	Carbamate	58000
Methoxychlor	72-43-5	I	OC	0.1
Metiram	9006-42-2	F	Thiocarbamate	0.1
Metolachlor	51218-45-2	H	Acetamide	530
Metribuzin	21087-64-9	H	Triazinone	1220
Metsulfuron-methyl	74223-64-6	H	Sulfonylurea	9500
Mevinphos	7786-34-7	I	OP	600000
Mexacarbate	315-18-4	I	Carbamate	100
Mirex	2385-85-5	I	OC	0.00007
Molinate	2212-67-1	H	Thiocarbamate	970
Monocrotophos	6923-22-4	I	OP	1000000
Monolinuron	1746-81-2	H	Phenylurea	735
Monuron	150-68-5	H	Phenylurea	230
Myclobutanil	88671-89-0	F	Triazole	142
Naled	300-76-5	I	OP	2000
Neburon	555-37-3	H	Phenylurea	5
Nicosulfuron	111991-09-4	H	Sulfonylurea	22000
Nitrofen	1836-75-5	H	Phenyl ester	1
Norflurazon	27314-13-12	H	Pyridazinone	28
Oryzalin	19044-88-3	H	Nitroaniline	2.5
Oxadiazon	1966-30-9	H	Oxadiazolone	0.7
Oxamyl	23135-22-0	I	Carbamate	282000
Oxycarboxin	5259-88-1	F	Carboxamide	1000
Oxydemeton-methyl	301-12-2	I	OP	1000000
Oxyfluorfen	42874-03-3	H	Nitrophenoxy	0.1
Parathion	56-38-2	I	OP	24
Parathion-methyl	298-00-0	I	OP	60
Pebulate	1114-71-2	H	Thiocarbamate	100
Pendimethalin	40487-42-1	H	Nitroaniline	0.275
Perfluidone	37924-13-3	H	Sulfonimide	500000
Permethrin	52645-53-1	I	Pyrethroid	0.006
Phenmedipham	13684-63-4	H	Carbamate	4.7
Phenthoate	2597-03-7	I	OP	11
Phorate	298-02-2	I	OP	22
Phosalone	2310-17-0	I	OP	3
Phosmet	732-11-6	I	OP	20
Posphamidon	13171-21-6	I	OP	1000000
Pirimicarb	23103-98-2	I	Carbamate	2700
Pirimiphos	23505-41-1	I	OP	93
Pirimiphos-methyl	2923-93-7	I	OP	9
Primisulfuron-methyl	86209-58-0	H	Sulfonylurea	70
Prochloraz	67747-09-5	F	Carbamate	34

Table 10 (*continued*)

Common name	CAS No.	Use[a]	Class	Solubility (mg L^{-1})
Procymidone	32809-16-8	H	Imide	4.5
Prodiamine	29091-21-2	H	Nitroaniline	0.013
Profenofos	41198-08-7	I	OP	28
Profluralin	26399-36-0	H	Nitroaniline	0.1
Promecarb	2631-37-0	I	Carbamate	91
Prometuron	1610-18-0	H	Triazine	720
Prometryn	7287-19-6	H	Triazine	33
Pronamide	23950-58-5	H	Benzamide	15
Propachlor	1918-16-7	H	Acetamide	613
Propamocarb HCl	24579-73-5	F	Carbamate	1000000
Propanil	709-98-8	H	Propanamide	200
Propargite	2312-35-8	I	Sulfite ester	0.5
Propazine	139-40-2	H	Triazine	8.6
Propham	122-42-9	H	Carbamate	250
Propiconazole	60207-90-1	F	Triazole	110
Propoxur	114-26-1	I	Carbamate	1800
Pyrazon	1698-60-8	H	Carbamate	400
Quintozene	82-68-8	F	OC	0.44
Quizalofop-ethyl	76578-14-8	H	Phenoxy	0.31
Resmethrin	10453-86-8	I	Pyrethroid	0.01
Siduron	1982-49-6	H	Phenylurea	18
Silvex	93-72-1	H	Phenoxy	140
Simazine	122-34-9	H	Triazine	6.2
Simetryn	1014-70-6	H	Triazine	450
Sulfometuron-methyl	74222-97-2	H	Sulfonylurea	70
Sulprofos	35400-43-2	I	OP	0.31
2,4,5-T	93-76-5	H	Phenoxy	278
Tebuthiuron	34014-18-1	H	Methylurea	2500
Temephos	3383-96-8	I	OP	0.001
Terbacil	5902-51-2	H	Methyluracil	710
Terbufos	13071-79-9	I	OP	5
Terbutryn	886-50-0	H	Triazine	22
Tetrachlorvinphos	961-11-5	I	OP	11
Thiabendazole	148-79-8	F	Imidazole	50
Thidiazuron	51707-55-2	D	Phenylurea	20
Thifensulfuron-methyl	79277-27-3	H	Sulfonylurea	2400
Thiobencarb	28249-77-6	H	Thiocarbamate	28
Thiodicarb	59669-26-0	I	Carbamate	19.1
Thiophanate-methyl	23564-05-8	F	Imidazole	3.5
Thiram	137-26-8	F	Diamide	30
Tolclofos-methyl	57018-04-9	F	OP	0.3
Tralomethrin	66841-35-2	I	Pyrethroid	0.001
Triadimefon	43121-43-3	F	Triazole	71.5
Triadimenol	55219-65-3	F	Triazole	47
Triallate	2303-17-5	H	Thiocarbamate	4
Tribenuron-methyl	101200-48-0	H	Sulfonylurea	280
Tribufos	78-48-8	D	OP	2.3
Trichlorfon	52-68-6	I	OP	120000
Trichloronat	327-98-0	I	OP	50
Triflumizole	99387-89-0	F	Imidazole	12500
Trifluralin	1582-09-8	H	Nitroaniline	0.3
Trimethacarb	2655-15-4	I	Carbamate	58
Vernolate	1929-77-7	H	Thiocarbamate	108
Vinclozolin	50471-44-8	F	Oxazolidine	1000
Zineb	12122-67-7	F	Thiocarbamate	10
Ziram	137-30-4	F	Thiocarbamate	65

[a] D = defoliant; F = fungicide; H = herbicide; I = insecticide/acaricide.

in the ability to analyze multiple pesticides for multiple classes in a variety of sample matrices have occurred in recent years. A growing number of techniques are available to the analytical chemist, and many strategies are possible to meet the purpose of analyses. In general, the use of the fewest analytical steps that provides reliable results in a rugged approach serves as the best overall approach to determining pesticide residues in food, environmental, and other types of samples. An inherent difficulty in multiclass, multiresidue analysis is that as the range of analytes increases, the overall selectivity of the sample preparation decreases. Therefore, trade-offs and compromises must often be made in obtaining high recoveries of a wide range of analytes while minimizing time, effort, and cost of the procedure(s).

Ultimately, these fundamental and practical aspects may limit the ability of the strategy to meet the needs of the analysis. As this article describes, a host of strategies are available to the chemist, but practical concerns in the laboratory, such as time, budgets, available instruments, and personnel, limit the amount of effort and resources that can be devoted to the analysis. A clearly superior, efficient and all-encompassing strategy has yet to be devised, which has made method development in multiclass, multiresidue analysis of pesticides an active area of study. The number of different pesticides and sample types leads to unique challenges in the determination of multiclass, multiresidue pesticides that few other areas of analysis match.

Finally, for reference, a list of a wide range of pesticides, with CAS numbers, uses, classes and solubilities in water are given in Table 10.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
AOAC®	Association of Official Analytical Chemists
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
ASE™	Accelerated Solvent Extraction
CDFA	California Department of Food and Agriculture
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CGE	Capillary Gel Electrophoresis
CID	Collision-induced Dissociation
DAD	Diode-array Detector
DSI	Direct Sample Introduction or ''dirty sample'' injection
DVB	Divinylbenzene
ECD	Electron-capture Detector
EDTA	Ethylenediaminetetraacetic Acid
EI	Electron Impact Ionization
ELCD	Electrolytic Conductivity Detector
ELISA	Enzyme-linked Immunosorbent Assay
EPA	US Environmental Protection Agency
ESE	Enhanced Solvent Extraction
ESP	Electrospray
ETU	Ethylenethiourea
EU	European Union
FDA	US Food and Drug Administration
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
GC	Gas Chromatography
GC/AED	Gas Chromatography/Atomic Emission Detection
GCB	Graphitized Carbon Black
GC/MS	Gas Chromatography/Mass Spectrometry
GC/MS/MS	Gas Chromatography/Tandem Mass Spectrometry
GLC	Gas–Liquid Chromatography
GLP	Good Laboratory Practices
GPC	Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry
ISO	International Standards Organization
ISP	Ionspray
ITD	Ion-trap Detector
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LC/PBMS	Liquid Chromatography/Particle Beam Mass Spectrometry
LC/TSPMS	Liquid Chromatography/Thermospray Mass Spectrometry
LLE	Liquid–Liquid Extraction
LOD	Limits of Detection
LOQ	Limit of Quantitation
MAE	Microwave-assisted Extraction
MEKC	Micellar Electrokinetic Chromatography
MRL	Maximum Residue Limit
MRM	Multiresidue Method
MS	Mass Spectrometry
MS^n	Multiple Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MSPD	Matrix Solid-phase Dispersion
NASS	National Agricultural Statistics Service

NCI	Negative Ion Chemical Ionization
NPD	Nitrogen–Phosphorus Detector
OC	Organochlorine Pesticide
ODS	Octadecylsilane
OP	Organophosphorus Insecticide
PAM I	Pesticide Analytical Manual, Vol. I
PAM II	Pesticide Analytical Manual, Vol. II
PB	Particle Beam
PBMS	Performance-based Measurement Systems
PCI	Positive Ion Chemical Ionization
PDA	Photodiode Array
PFE	Pressurized Fluid Extraction
PFPD	Pulsed Flame Photometric Detector
PID	Photoionization Detector
PLE	Pressurized Liquid Extraction
PSA	Primary/Secondary Amine
PSE	Pressurized Solvent Extraction
PUF	Polyurethane Foam
QA	Quality Assurance
QC	Quality Control
QMA	Quaternary Methylamine
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RSD	Relative Standard Deviation
SAX	Strong Anion Exchange
SCD	Sweep Codistillation
SCX	Strong Cation Exchange
SEC	Size-exclusion Chromatography
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SOP	Standard Operating Procedure
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TLC	Thin-layer Chromatography
TSP	Thermospray
USDA	US Department of Agriculture
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
XSD™	Halogen-specific Detector

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Organic Analysis in Environmental Samples by Electrochemical Methods • Organometallic Compound Analysis in Environmental Samples • Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Waste Extraction Procedures

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Liquid Chromatography in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food • Sample Preparation for Food Analysis, General

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicide Residues in Biota, Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Antibiotics, Pharmaceutical Analysis of • Chemical Reagents and Derivatization Procedures in Drug Analysis • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Micellar Electrokinetic Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Literature Searching Methodology • Microwave Techniques • Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. Food and Drug Administration, *Pesticide Analytical Manual, Vol. I: Multiresidue Methods*, 3rd edition, US Department of Health and Human Services, Washington, DC, 1994.
2. Food and Drug Administration, *Pesticide Analytical Manual*, Vol. II, US Department of Health and Human Services, Washington, DC, 1987.
3. *Guide to Environmental Analytical Methods*, 4th edition, Genium Publishing, Schenectady, NY, 1998.
4. J.H. Hotchkiss, 'Pesticide Residue Controls to Ensure Food Safety', *Crit. Rev. Food Sci. Nutr.*, **31**, 191–203 (1992).
5. E.E. Ziegler, L.J. Filer, Jr (eds.), *Present Knowledge in Nutrition*, 7th edition, International Life Sciences Institute Press, Washington, DC, 1996.
6. W.R. Wolf, 'Reference Materials', in *Methods of Analysis for Nutrition Labeling*, eds. D.M. Sullivan,

D.E. Carpenter, AOAC® International, Arlington, VA, 111–122, 1993.
7. Nutrient Data Laboratory, 'USDA Food Composition Data' [computer file], Agricultural Research Service, Riverdale MD, 1996 (http://www.nal.usda.gov/fnic/foodcomp/).
8. J.T. Tanner, W. Horwitz, 'The Importance of Reference Materials to the Food and Drug Administration, USA', ISO/REMCO Report N 317, June 1994, International Organization for Standardization, Geneva.
9. S.L. Reynolds, R. Fussel, M. Caldow, R. James, S. Nawaz, C. Ebden, D. Pendlington, T. Stijve, H. Diserens, 'Intercomparison Study of Two Multi-residue Methods for the Enforcement of EU MRLS for Pesticides in Fruit, Vegetables and Grain, Phase I', Report EUR 17870 EN, European Commission, Brussels, 1997.
10. S.L. Reynolds, R. Fussel, M. Caldow, R. James, S. Nawaz, C. Ebden, D. Pendlington, T. Stijve, H. Diserens, 'Intercomparison Study of Two Multi-residue Methods for the Enforcement of EU MRLS for Pesticides in Fruit, Vegetables and Grain, Phase II', Report EUR 118639 EN, European Commission, Brussels, 1998.
11. L.E. Sojo, D.S. Gamble, D.W. Gutzman, 'Sorption and Bound Residue Formation of Linuron, Methylparathion, and Metolachlor by Carrot Tissues: Kinetics by On-line HPLC Microextraction', *J. Agric. Food Chem.*, **45**, 3634–3641 (1997).
12. F.R. Troeh, L.M. Thompson, *Soils and Soil Fertility*, 5th edition, McGraw-Hill, New York, 1993.
13. M.S. Majewksi, P.D. Capel, *Pesticides in the Atmosphere*, Ann Arbor Press, Chelsea, MI, 1995.
14. C.D.S. Tomlin (ed.), *The Pesticide Manual: A World Compendium*, 11th edition, British Crop Protection Council, Farnham, Surrey, UK, 1997.
15. R.T. Meister (ed.), *Farm Chemicals Handbook '96*, Meister Publishing, Willoughby, OH, 1996.
16. K.H. Büchel (ed.), *Chemistry of Pesticides*, John Wiley & Sons, New York, 1983.
17. L.G. Copping, H. Kidd, C.D.S. Tomlin (eds.), *The Pesticide Index*, 3rd Edition, British Crop Protection Council, Farnham, Surrey, UK, 1995.
18. 'Agricultural Chemical Usage, Summary Reports on Field Crops, Fruits, Vegetables', USDA, National Agricultural Statistics Service (http://www.usda.gov/nass/pubs/pubs.htm), 1999.
19. M. Kawasaki, K. Fukuhara, R. Katsumura, N. Takasaka, S. Uchiyama, 'GC/MS (SIM) Determination of 14 Pesticides Including Cyhexatin and 2,4,5-T in Nuts', *J. Food Hyg. Soc. Jpn.*, **38**(3), 161–169 (1997).
20. A.M. Blair, T.D. Martin, 'A Review of the Activity, Fate, and Mode of Action of Sulfonylurea Herbicides', *Pestic. Sci.*, **22**, 195–219 (1988).
21. 'Pesticide Data Program Annual Summary', Agricultural Marketing Service, US Department of Agriculture, 1992–1997 Reports.
22. S.J. Lehotay, J.A. Harman-Fetcho, L.L. McConnell, 'Agricultural Pesticide Residues in Oysters and Water from Two Chesapeake Bay Tributaries', *Mar. Pollut. Bull.*, **37**(1–2), 32–44 (1998).
23. E.A. Hogendoorn, R. Hoogerbrugge, R.A. Baumann, H.D. Meiring, A.P.J.M. de Jong, P. van Zoonen, 'Screening and Analysis of Polar Pesticides in Environmental Monitoring Programmes by Coupled-column Liquid Chromatography and Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **754**, 49–60 (1996).
24. M. Knutsson, G. Nilvé, L. Mathiasson, J.Å. Jönsson, 'Supported Liquid Membrane Technique for Time-integrating Field Sampling of Acidic Herbicides at Sub Parts per Billion Level in Natural Waters', *J. Agric. Food Chem.*, **40**, 2413–2417 (1992).
25. S. Nemoto, S.J. Lehotay, 'Analysis of Multiple Herbicides in Soybeans Using Pressurized Liquid Extraction and Capillary Electrophoresis', *J. Agric. Food Chem.*, **46**(6), 2190–2199 (1998).
26. A. de Kok, M. Heimstra, 'Optimization, Automation, and Validation of the Solid-phase Extraction Cleanup and On-line Liquid Chromatographic Determination of *N*-Methylcarbamate Pesticides in Fruits and Vegetables', *J. AOAC® Int.*, **75**(6), 1063–1072 (1992).
27. *Analytical Methods for Pesticide Residues in Foodstuffs*, 3rd edition, General Inspectorate for Health Protection, Ministry of Public Health, Welfare and Sport, Amsterdam, The Netherlands, 1996.
28. National Research Council, *Pesticides in the Diets of Infants and Children*, National Academy Press, Washington, DC, 1993.
29. H.D. Meiring, A.P.J.M. de Jong, 'Determination of Ethylenethiourea in Water by Single-step Extractive Derivatization and Gas Chromatography–Negative Ion Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **683**, 157–165 (1994).
30. H. van Lishaut, W. Schwack, 'Trace Determination of Dithiocarbamates: Comparison of Ion-pair Chromatography with Acid Digestion', in 2nd European Pesticide Residue Workshop, Almería, Spain, May 24–27, 1998, abstract A26.
31. A.G. Hornsby, R.D. Wauchope, A.E. Herner, *Pesticide Properties in the Environment*, Springer-Verlag, New York, 1996.
32. E.A. Bunch, D.M. Altwein, L.E. Johnson, J.R. Farley, A.A. Hammersmith, 'Homogeneous Sample Preparation of Raw Shrimp Using Dry Ice', *J. AOAC® Int.*, **78**(3), 883–887 (1995).
33. F.M. Garfield, *Quality Assurance Principles for Analytical Laboratories*, 2nd edition, Association of Official Analytical Chemists, Arlington, VA, 1991.
34. *Environmental Laboratory Data Evaluation*, Genium Publishing, Schenectady, NY, 1998.
35. A.R.C. Hill, 'Quality Control Procedures for Pesticide Residues Analysis', European Commission Document 7826/VI/97, European Commission, Brussels, 1997.

36. W.Y. Garner, M. Barge (eds.), *Good Laboratory Practices: An Agrochemical Perspective*, ACS Symposium Series, Vol. 369, American Chemical Society, Washington, DC, 1988.
37. 'AOAC® Official Methods Program Associate Referee's Manual on Development, Study, Review, and Approval Process', AOAC® International, Gaithersburg, MD, 1998.
38. AOAC® Official Method 985.22, 'Organochlorine and Organophosphate Pesticides, Gas Chromatographic Method', in *Official Methods of Analysis*, 16th edition, Revision 3, AOAC® International, Gaithersburg, MD, 1998.
39. L.D. Sawyer, 'The Luke et al. Method For Determining Multipesticide Residues in Fruits and Vegetables: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **68**(1), 64–71 (1985).
40. S.J. Lehotay, 'Determination of Pesticide Residues in Nonfatty Foods by SFE and GC/MS: Collaborative Study', *J. AOAC® Int.*, in preparation.
41. P.E. Key, A.L. Patey, S. Rowling, A. Wilbourn, F.M. Worner, 'International Proficiency Testing of Analytical Laboratories for Foods and Feeds From 1990 to 1996: the Experiences of the United Kingdom Food Analysis Performance Assessment Scheme', *J. AOAC® Int.*, **80**(4), 895–899 (1997).
42. R. Armishaw, J.M. Majewski, P.J. McLay, 'Development and Certification of Reference Materials for Residues of Organochlorine and Organophosphorus Pesticides in Beef Fat ACSL CRM 1 and 2', *Fresenius' J. Anal. Chem.*, **360**, 630–639 (1998).
43. A. Boenke, 'Pesticide Residue Measurements: Activities and Current Research from the EC Standards, Measurements and Testing Programme (SMT). *Trends Anal. Chem.*, **18**(7), 440–448 (1999).
44. *Pesticide Analytical Methods in Sweden, Part I*, National Food Administration, Uppsala, 1998.
45. M.A. Luke, H.T. Masumoto, 'Pesticide Residue Analysis of Foods', in *Analytical Methods for Pesticides and Plant Growth Regulators*, Academic Press, New York, 161–200, Vol. 15, 1989.
46. J. Fillion, L. Nolan, 'SPE: an Effective Way to Clean Up Pesticide Residues', *Today's Chemist at Work*, September, 14–24 (1996).
47. W.G. Fong, H.A. Moye, J.N. Seiber, J.P. Toth, *Pesticide Residues in Foods: Methods Techniques and Regulations*, John Wiley & Sons, New York, 1999.
48. T. Cairns, J. Sherma (eds.), *Emerging Strategies for Pesticide Analysis*, CRC Press, Boca Raton, FL, 1–528, 1992.
49. J. Sherma (ed.), 'Chromatography and Electrophoresis in Environmental Analysis: Pesticide Residues', *J. Chromatogr. A*, **754**, (1996).
50. L.M.L. Nollett (ed.), *Handbook of Food Analysis, Volume 2, Residues and Other Food Component Analysis*, Marcel Dekker, New York, 1996.
51. T. Cairns, M.A. Luke, K.S. Chiu, D.N. Navarro, E.G. Siegmund, 'Multiresidue Pesticide Analysis by Ion-trap Technology; a Cleanup Approach for Mass Spectral Analysis', *Rapid Commun. Mass Spectrom.*, **7**, 1070–1076 (1993).
52. S.M. Lee, M.L. Papathkis, H.C. Feng, G.F. Hunter, J.E. Carr, 'Multipesticide Residue Method for Fruits and Vegetables: California Department of Food and Agriculture', *Fresenius' J. Anal. Chem.*, **339**, 376–383 (1991).
53. J. Fillion, R. Hindle, M. Lacroix, J. Selwyn, 'Multiresidue Method for the Determination of Pesticides in Fruit and Vegetables by GC/MS and HPLC/Fluorescence', *J. AOAC® Int.*, **78**(5), 1252–1266 (1995).
54. D. Barceló (ed.), *Environmental Analysis Techniques, Applications and Quality Assurance*, Elsevier, New York, 1993.
55. A.M. Kern, J.D. Evans, S.L. Hayes, K.G. Partymiller, G.R. Swanson, 'Representative Sampling and Analysis of Heterogeneous Soils', *Environ. Testing Anal.*, March/April, 20–27 (1997).
56. S.J. Lehotay, N. Aharonson, E. Pfeil, M.A. Ibrahim, 'Development of a Sample Preparation Technique for Supercritical Fluid Extraction for Multiresidue Analysis of Pesticides in Produce', *J. AOAC® Int.*, **78**(3), 831–840 (1995).
57. J. Vuik, R. van Dinter, R.H. de Vos, 'Improved Sample Pretreatment of the Carbon Disulfide Evolution Method for the Determination of Dithiocarbamate Residues in Lettuce', *J. Agric. Food Chem.*, **40**, 604–606 (1992).
58. R.H. Obenhauf, R. Bostwick, D.J. Burger, J.D. McCormack, D. Selem (eds.), *SPEX Handbook of Sample Preparation and Handling*, SPEX Industries, Edison, NJ, 1991.
59. S.J.V. Young, C.H. Parfitt, 'Homogeneity of Fruits and Vegetables Comminuted in a Vertical Cutter Mixer', *J. AOAC® Int.*, **79**(4), 976–980 (1996).
60. R.J. Hemingway, N. Aharonson, P.A. Greve, T.R. Roberts, H.P. Thier, 'Improved Cost-effective Approaches to Pesticide Residues Analysis', *Pure Appl. Chem.*, **56**(8), 1131–1152 (1984).
61. S.J. Lehotay, 'Supercritical Fluid Extraction of Pesticides in Foods', *J. Chromatogr. A*, **785**, 289–312 (1997).
62. I.A. Stuart, J. MacLachlan, A. McNaughtan, 'Compounds of Agricultural Significance Using Environmental Analytical Supercritical Fluid Extraction', *Analyst*, **121**(3), 11R–28R (1996).
63. V. Camel, 'The Determination of Pesticide Residues and Metabolites Using Supercritical Fluid Extraction', *Trends Anal. Chem.*, **16**(6), 351–369 (1997).
64. T.L. Chester, J.D. Pinkston, D.E. Raynie, 'Supercritical Fluid Chromatography and Extraction', *Anal. Chem.*, **70**(12), 301R–319R (1998).
65. V. Lopez-Avila, R. Young, N. Teplitsky, 'Microwave-assisted Extraction as an Alternative to Soxhlet, Sonication, and Supercritical Fluid Extraction', *J. AOAC® Int.*, **79**(1), 142–156 (1996).

66. H.M. Pylypiw, Jr, T.L. Arsenault, C.M. Thetford, M.J.I. Mattina, 'Suitability of Microwave-assisted Extraction for Multiresidue Pesticide Analysis of Produce', *J. Agric. Food Chem.*, **45**(9), 3522–3528 (1997).
67. S.J. Stout, A.R. daCunha, G.L. Picard, M.M. Safarpour, 'Microwave-assisted Extraction Coupled with Liquid Chromatography/Electrospray Ionization Mass Spectrometry for the Simplified Determination of Imidazolinone Herbicides and Their Metabolites in Plant Tissue', *J. Agric. Food Chem.*, **44**(11), 3548–3553 (1996).
68. B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: a Technique for Sample Preparation', *Anal. Chem.*, **68**(6), 1033–1039 (1996).
69. S.J. Lehotay, C.-H. Lee, 'Evaluation of Fibrous Cellulose Drying Agent in Supercritical Fluid Extraction and Pressurized Liquid Extraction of Diverse Pesticides', *J. Chromatogr. A*, **785**, 313–327 (1997).
70. H. Obana, K. Kikuchi, M. Okihashi, S. Hori, 'Determination of Organophosphorus Pesticides in Foods Using an Accelerated Solvent Extraction System', *Analyst*, **122**(3), 217–220 (1997).
71. A. Fernandez-Gutierrez, J.L. Martinez-Vidal, F.J. Arrebola-Liebanas, A. Gonzalez-Casado, J.L. Vilchez, 'Determination of Endosulfan and Some Pyrethroids in Waters by Micro Liquid–Liquid Extraction and GC/MS', *Fresenius' J. Anal. Chem.*, **360**, 568–572 (1998).
72. J.L. Snyder, R.L. Grob, M.E. McNally, T.S. Oostdyk, 'The Effect of Instrumental Parameters and Soil Matrix on the Recovery of Organochlorine and Organophosphate Pesticides from Soils Using Supercritical Fluid Extraction', *J. Chromatogr. A*, **31**, 183–191 (1993).
73. A. Martín-Esteban, P. Fernández, C. Cámara, G.N. Kramer, E.A. Maier, 'Preparation, Homogeneity and Stability of Polar Pesticides in Freeze-dried Water Interlaboratory Exercise', *Int. J. Environ. Anal. Chem.*, **67**, 125–141 (1997).
74. S.J. Lehotay, M.A. Ibrahim, 'Supercritical Fluid Extraction and Gas Chromatography/Ion Trap Mass Spectrometry of Pentachloronitrobenzene Pesticides in Vegetables', *J. AOAC® Int.*, **78**(2), 445–452 (1995).
75. K. Iijima, M. Saka, Y. Odanaka, O. Matano, 'Multiresidue Analytical Method of Pesticides by GC/MS: Application of Macroporous Diatomaceous Earth Column and Silica Gel Cartridge', *J. Pestic. Sci.*, **22**(1), 17–26 (1997).
76. H. Steinwandter, 'Universal Extraction and Cleanup Methods', *Anal. Methods Pestic. Plant Growth Regul.*, **17**, 35–89 (1989).
77. C.H. Parfitt, Jr, 'Miniaturized Multiresidue Approach to Determine Pesticide Residues in Fresh Fruits and Vegetables', *FDA Lab. Inf. Bull. No. 3616*, Nov. (1991).
78. J.J. Ridal, M.E. Fox, C.A. Sullivan, R.J. Maguire, A. Mazumder, D.R.S. Lean, 'Evaluation of Automated Extraction of Organochlorine Contaminants from Freshwater', *Anal. Chem.*, **69**(4), 711–717 (1994).
79. D.A. Bennett, A.C. Chung, S.M. Lee, 'Multiresidue Method for Analysis of Pesticides in Liquid Whole Milk', *J. AOAC® Int.*, **80**(5), 1065–1077 (1997).
80. J. Pawliszyn, *Solid Phase Microextraction*, Wiley-VCH, New York, 1997.
81. A.A. Boyd-Boland, S. Magdic, J.B. Pawliszyn, 'Simultaneous Determination of 60 Pesticides in Water Using Solid-phase Microextraction and Gas Chromatography–Mass Spectrometry', *Analyst*, **121**, 929–938 (1996).
82. A. Amirav, personal communication.
83. W.Q. Chen, K.F. Poon, M.H.W. Lam, 'The Application of Solid Phase Microextraction in the Analysis of Organophosphorus Pesticides in a Food Plant', *Environ. Sci. Technol.*, **32**(23), 3816–3820 (1998).
84. R.W. Hu, B. Hennion, L. Urruty, M. Montury, 'Solid Phase Microextraction of Pesticide Residues from Strawberries', *Food Addit. Contam.*, **16**(3), 111–117 (1999).
85. M. Chai, C. Arthur, J. Pawliszyn, 'Determination of Volatile Chlorinated Hydrocarbons in Air and Water with Solid-phase Microextraction', *Analyst*, **118**, A1501 (1993).
86. L.L. McConnell, E. Nelson, C.P. Rice, J.E. Baker, W.E. Johnson, J.A. Harman, K. Bialek, 'Chlorpyrifos in the Air and Surface Water of Chesapeake Bay: Predictions of Atmospheric Deposition Fluxes', *Environ. Sci. Technol.*, **31**(5), 1390–1398 (1997).
87. M.S. Majewski, W.T. Foreman, D.A. Goolsby, N. Nakagaki, 'Airborne Pesticide Residues Along the Mississippi River', *Environ. Sci. Technol.*, **32**(23), 3689–3698 (1998).
88. M. Millet, H. Wortham, A. Sanusi, P. Mirabel, 'A Multiresidue Method for Determination of Trace Levels of Pesticides in Air and Water', *Arch. Environ. Contam. Toxicol.*, **31**, 543–556 (1996).
89. R.J. Argauer, S.J. Lehotay, R.T. Brown, 'Determining Lipophilic Pyrethroids and Chlorinated Hydrocarbons in Fortified Ground Beef Using Ion-trap Mass Spectrometry', *J. Agric. Food Chem.*, **45**(10), 3936–3939 (1997).
90. M.C. Bowman, M. Beroza, 'Extraction *P*-Values of Pesticides and Related Compounds in Six Binary Solvent Systems', *J. Assoc. Off. Anal. Chem.*, **48**, 943–952 (1965).
91. von W. Specht, M. Tillkes, 'Gaschromatographische Bestimmung von Rückständen an Pflanzenbehandlungsmitteln nach Cleanup über Gelchromatographie und Mini-Kieselgel-Säulenchromatographie', *Beitr. Tabakforsch. Int.*, **10**(1), 73–79 (1979).
92. G.G. Rimkus, M. Rummler, I. Nausch, 'Gel Permeation Chromatography–High-performance Liquid Chromatography Combination as an Automated Cleanup Technique for the Multiresidue Analysis of Fats', *J. Chromatogr. A*, **737**, 9–14 (1996).
93. B.G. Luke, 'Sweep Codistillation: Recent Developments and Applications', *Anal. Methods Pestic. Plant Growth Regul.*, **17**, 75–99 (1989).
94. P. Armishaw, R.G. Millar, 'Comparison of Gel Permeation Chromatography, Sweep Codistillation, and

Florisil Column Adsorption Chromatography as Sample Cleanup Techniques for the Determination of Organochlorine Pesticide Residues in Animal Fats', *J. AOAC® Int.*, **76**(6), 1317–1322 (1993).
95. Y. Odanaka, O. Matano, S. Goto, 'The Use of Solid Bonded-phase Extraction as Alternative to Liquid–Liquid Partitioning for Pesticide Residue Analysis of Crops', *Fresenius' J. Anal. Chem.*, **339**, 368–373 (1991).
96. F.J. Schenck, R. Wagner, M.K. Hennessy, J.L. Okrasinski, Jr, 'Screening Procedure for Organochlorine and Organophosphorus Pesticide Residues in Eggs Using a Solid-phase Extraction Cleanup and Gas Chromatographic Detection', *J. AOAC® Int.*, **77**(4), 1036–1040 (1994).
97. G. Font, J. Mañes, J.C. Molto, Y. Pico, 'Solid-phase Extraction in Multi-residue Analysis of Water', *J. Chromatogr.*, **642**, 135–161 (1993).
98. K.-C. Ting, G.S. Tamashiro, 'Off-line High-performance Liquid Chromatography and Solid-phase Extraction Cleanup for Confirmation of Pesticide Residues in Fresh Produce by Gas Chromatography–Mass Spectroscopy', *J. Chromatogr. A*, **754**, 455–462 (1996).
99. C. Crescenzi, A. Di Corcia, G. Passariello, R. Samperi, M.I.T. Carou, 'Evaluation of Two New Examples of Graphitized Carbon Blacks for Use in Solid-phase Extraction Cartridges', *J. Chromatogr. A*, **733**, 41–55 (1996).
100. L.E. Sojo, A. Brocke, J. Fillion, S.M. Price, 'Application of Activated Carbon Membranes for On-line Cleanup of Vegetable and Fruit Extracts in the Determination of Pesticide Multiresidue by Gas Chromatography with Mass Selective Detection', *J. Chromatogr. A*, **788**, 141–154 (1997).
101. Th. Hankemeier, P.C. Steketee, J.J. Vreuls, U.A.Th. Brinkman, 'Automated At-line Solid-phase Extraction–Gas Chromatographic Analysis of Micropollutants in Water Using the PrepStation', *J. Chromatogr. A*, **750**, 161–174 (1996).
102. S.J. Lehotay, A. Valverde-García, 'Evaluation of Different Solid-phase Traps for Automated Collection and Cleanup in the Analysis of Multiple Pesticides in Fruits and Vegetables After Supercritical Fluid Extraction', *J. Chromatogr. A*, **765**, 69–84 (1997).
103. D.L. Stalling, S. Saim, K.C. Kuo, J.J. Stunkel, 'Integrated Sample Processing by On-line Supercritical Fluid Extraction–Gel Permeation Chromatography', *J. Chromatogr. Sci.*, **30**, 486–490 (1992).
104. M. Heimstra, A. de Kok, 'Determination of *N*-Methylcarbamate Pesticides in Environmental Water Samples Using Automated On-line Trace Enrichment with Exchangeable Cartridges and High-performance Liquid Chromatography', *J. Chromatogr. A*, **667**, 155–166 (1994).
105. C.C. Walker, H.M. Lott, S.A. Barker, 'Matrix Solid Phase Dispersion Extraction and the Analysis of Drugs and Environmental Pollutants in Aquatic Species', *J. Chromatogr.*, **642**, 225–242 (1993).
106. E. Viana, J.C. Molto, G. Font, 'Optimization of a Matrix Solid-phase Dispersion Method for the Analysis of Pesticides Residues in Vegetables', *J. Chromatogr. A*, **754**, 437–444 (1996).
107. F.J. Schenck, R. Wagner, 'Screening Procedure for Organochlorine and Organophosphorus Pesticide Residues in Milk Using Matrix Solid Phase Dispersion (MSPD) Extraction and Gas Chromatographic Determination', *Food Addit. Contam.*, **12**(4), 535–541 (1995).
108. A.C. Warwick, S.A. Anderson, L.M. Thorpe, S.E. Owen, H.M.C. Anderson, S.L. Reynolds, 'The Analysis of Cyhexatin Residues in Apples, Pears, and Kiwi Fruit Using Inductively Coupled Plasma Mass Spectrometry as an Initial Screen for Total Tin, with Confirmation by Gas Chromatography–Mass Spectrometry', *Food Addit. Contam.*, **15**(3), 288–292 (1998).
109. T.C. Mueller, T.B. Moorman, M.A. Locke, 'Detection of Herbicides Using Fluorescence Spectroscopy', *Weed Sci.*, **40**, 270–274 (1992).
110. N.Y. Seedhar, P.R.K. Reddy, G.V.S. Reddy, S.J. Reddy, 'Differential Pulse Polarographic Determination of Dicrotophos, Crotoxyphos, and Chlorfenviphos in Grains and Soils', *Talanta*, **44**, 1859–1863 (1997).
111. T. Visser, M.J. Vredenbregt, A.P.J.M. de Jong, G.W. Somsen, Th. Hankemeier, N.H. Velthorst, C. Gooijer, U.A.Th. Brinkman, 'Improvements in Environmental Trace Analysis by GC/IR and LC/IR', *J. Mol. Struct.*, **408/409**, 97–103 (1997).
112. S. Butz, H.-J. Stan, 'Screening of 265 Pesticides in Water by Thin-layer Chromatography with Automated Multiple Development', *Anal. Chem.*, **67**(3), 620–630 (1995).
113. S. Dagan, A. Amirav, 'Fast, Very Fast and Ultra Fast GC/MS of Thermally Labile Steroids, Carbamates and Drugs in Supersonic Molecular Beams', *J. Am. Soc. Mass. Spectrom.*, **7**, 737–752 (1996).
114. T.L. Chester, 'Chromatography from the Mobile-phase Perspective', *Anal. Chem.*, **69**, 165A–169A (1997).
115. T. Berger, 'Feasibility of Multifamily Pesticide Screening by High-efficiency Packed-column Supercritical Fluid Chromatography', Hewlett-Packard Application Note 228-186, Hewlett-Packard, Avondale, PA, June 1992.
116. P.D. Johnson, D.A. Rimmer, R.H. Brown, 'Adaptation and Application of a Multi-residue Method for the Determination of a Range of Pesticides, Including Phenoxy Acid Herbicides in Vegetation, Ased on High-resolution Gel Permeation Chromatographic Cleanup and Gas Chromatographic Analysis with Mass-selective Detection', *J. Chromatogr. A*, **765**, 3–11 (1997).
117. M. Anastassiades, W. Schwack, 'Analysis of Carbendazim/Benomyl, Thiophanate Methyl and 2,4-Dichlorophenoxyacetic Acid in Fruits and Vegetables After Supercritical Fluid Extraction', *J. Chromatogr. A*, **825**(1), 45–54 (1998).

118. M.L. Hopper, B. McMahon, K.R. Griffitt, K. Cline, M.E. Fleming-Jones, D.C. Kendall, 'Analysis of Fatty and Nonfat Foods for Chlorophenoxy Alkyl Acids and Pentachlorophenol', *J. AOAC® Int.*, **75**(4), 707–713 (1992).
119. C.L. Deyrup, S.-M. Chang, R.A. Weintraub, H.A. Moye, 'Simultaneous Esterification and Acylation of Pesticides for Analysis by Gas Chromatography. 1. Derivatization of Glyphosate and (Aminomethyl)phosphonic Acid with Fluorinated Alcohols–Perfluorinated Anhydrides', *J. Agric. Food Chem.*, **33**, 944–947 (1985).
120. W.L. Saxton, 'Emergence Temperature Indices and Relative Retention Times of Pesticides and Industrial Chemicals Determined by Linear Programmed Temperature Gas Chromatography', *J. Chromatogr.*, **393**, 175–194 (1987).
121. D.R. Erney, A.M. Gillespie, D.M. Gilvydis, C.F. Poole, 'Explanation of the Matrix-induced Chromatographic Response Enhancement of Organophosphorus Pesticides During Open Tubular Column Gas Chromatography with Splitless or Hot On-column Injection and Flame Photometric Detection', *J. Chromatogr. A*, **638**, 57–63 (1993).
122. J. Hajslova, K. Holadova, V. Kocourek, J. Poustka, M. Godula, P. Cuhra, M. Kempny, 'Matrix-induced Effects: a Critical Point in the Gas Chromatographic Analysis of Pesticide Residues', *J. Chromatogr. A*, **800**, 283–295 (1998).
123. F.J. Schenck, S.J. Lehotay, 'Does Further Cleanup Reduce the Matrix Enhancement Effect in GC Analysis of Pesticide Residues in Food?', *J. Chromatogr. A*, in press.
124. P.L. Wylie, K. Uchiyama, 'Improved Gas Chromatographic Analysis of Organophosphorus Pesticides with Pulsed Splitless Injection', *J. AOAC® Int.*, **79**(2), 571–577 (1996).
125. H.-J. Stan, M. Linkerhäger, 'Pesticide Residue Analysis in Foodstuffs Applying Capillary Gas Chromatography with Atomic Emission Detection State-of-the-art Use of Modified Multimethod S19 of the Deutsche Forschungsgemeinschaft and Automated Large-volume Injection with Programmed-temperature Vaporization and Solvent Venting', *J. Chromatogr. A*, **750**, 369–390 (1996).
126. J. Cook, M. Engel, P. Wylie, B. Quimby, 'Multiresidue Screening of Pesticides in Foods Using Retention Time Locking, GC/AED, Database Search, and GC/MS Identification', *J. AOAC® Int.*, **82**(2), 313–326 (1999).
127. H. Jing, A. Amirav, 'Pulsed Flame Photometer Detector – A Step Forward Towards Universal Heteroatom Selective Detection', *J. Chromatogr. A*, **805**, 177–215 (1998).
128. Y. Oh-Shin, M. Ko, H. Shin, 'Simultaneous Quantification of Insecticides Including Carbaryl in Drinking Water by Gas Chromatography Using Dual Electron-capture and Nitrogen–Phosphorus Detection', *J. Chromatogr. A*, **769**(2), 285–291 (1997).
129. T. Cairns, K.S. Chiu, D. Navarro, E. Siegmund, 'Multiresidue Pesticide Analysis by Ion-trap Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 971–988 (1993).
130. T. Cairns, 'Regulatory Mass Spectrometry for Pesticide Residue Analysis: Past, Present, and Future', *J. AOAC® Int.*, **75**(4), 591–593 (1992).
131. H. Jobst, V. Deutscher, 'Investigation on the Quantification of Pesticides with Mass Selective Detectors', *Fresenius' J. Anal. Chem.*, **349**, 298–300 (1994).
132. S. Schacterle, C. Feigel, 'Pesticide Residue Analysis in Fresh Produce by Gas Chromatography–Tandem Mass Spectrometry', *J. Chromatogr. A*, **754**, 411–422 (1996).
133. R.S. Sheridan, J.R. Meola, 'Analysis of Pesticide Residues in Fruits, Vegetables, and Milk by Gas Chromatography/Tandem Mass Spectrometry', *J. AOAC® Int.*, **82**, 982–990 (1999).
134. W. Liao, T. Joe, W.G. Cusick, 'Multiresidue Screening Method for Fresh Fruits and Vegetables with Gas Chromatographic/Mass Spectrometric Detection', *J. AOAC® Int.*, **74**(3), 554–565 (1991).
135. T. Cairns, 'Developing Guidelines for Detection, Confirmation and Quantification by MS', *Food Test. Anal.*, April/May, 8–17 (1997).
136. A. Amirav, H. Jing, 'Simultaneous PFPD and MS Detection for Enhanced Pesticide Analysis Capabilities', *J. Chromatogr. A*, **814**, 133–150 (1998).
137. H. Jing, A. Amirav, 'Pesticide Analysis with the Pulsed-flame Photometer Detector and a Direct Sample Introduction Device', *Anal. Chem.*, **69**(7), 1426–1435 (1997).
138. S.J. Lehotay, 'Analysis of Pesticide Residues in Fruit and Vegetable Extracts by Direct Sample Introduction/Gas Chromatography/Tandem Mass Spectrometry (DSI/GC/MS/MS)', *J. AOAC® Int.*, in press.
139. G.R. van der Hoff, A.C. van Beuzekom, U.A.Th. Brinkman, R.A. Baumann, P. van Zoonen, 'Determination of Organochlorine Compounds in Fatty Matrices. Application of Rapid Off-line Normal-phase Liquid Chromatographic Cleanup', *J. Chromatogr. A*, **754**, 487–496 (1996).
140. M. Anastassiades, E. Scherbaum, 'Multimethode zur Bestimmung von Pflanzenschutz- und Oberflächenbehandlungsmittel-Rückständen in Zitrusfrüchten mittels GC/MSD', *Dtsch. Lebensml.-Rundsch.*, **93**(10), 316–327 (1997).
141. I. Liška, J. Slobodník, 'Comparison of Gas and Liquid Chromatography for Analysing Polar Pesticides in Water Samples', *J. Chromatogr. A*, **733**, 235–258 (1996).
142. D.M. Gilvydis, S.M. Walters, 'Ion-pairing Liquid Chromatographic Determination of Benzimidazole Fungicides in Foods', *J. Assoc. Off. Anal. Chem.*, **73**(5), 753–761 (1990).
143. C.J. Miles, 'Determination of National Survey of Pesticides Analytes in Groundwater by Liquid Chromatography with Postcolumn Reaction Detection', *J. Chromatogr. A*, **592**, 283–290 (1992).

144. W.M.A. Niessen, *Liquid Chromatography–Mass Spectrometry*, 2nd edtion, Marcel Dekker, New York, 1998.
145. J. Slobodník, B.L.M. van Baar, U.A.Th. Brinkman, 'Column Liquid Chromatography–Mass Spectrometry: Selected Techniques in Environmental Applications for Polar Pesticides and Related Compounds', *J. Chromatogr. A*, **703**, 81–121 (1995).
146. W.M.A. Niessen, 'Advances in Instrumentation in Liquid Chromatography–Mass Spectrometry and Related Liquid-introduction Techniques', *J. Chromatogr. A*, **794**, 407–435 (1998).
147. C.-H. Liu, G.C. Mattern, X. Yu, R.T. Rosen, J.D. Rosen, 'Multiresidue Determination of Nonvolatile and Thermally Labile Pesticides in Fruits and Vegetables by Thermospray Liquid Chromatography/Mass Spectrometry', *J. Agric. Food Chem.*, **39**(4), 718–723 (1991).
148. D.A. Volmer, D.L. Vollmer, J.G. Wilkes, 'Multiresidue Analysis of Pesticides by Electrospray LC/MS and LC/MS/MS', *LC/GC*, **14**(3), 216–224 (1996).
149. M. Honing, J. Riu, D. Barceló, B.L.M. van Baar, U.A.Th. Brinkman, 'Determination of Ten Carbamate Pesticides in Aquatic and Sediment Samples by Liquid Chromatography–Ionspray and Thermospray Mass Spectrometry', *J. Chromatogr. A*, **733**, 283–294 (1996).
150. J. Slobodník, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Trace-level Determination of Pesticide Residues Using On-line Solid-phase Extraction–Column Liquid Chromatography with Atmospheric Pressure Ionization Mass Spectrometric and Tandem Mass Spectrometric Detection', *J. Chromatogr. A*, **741**, 59–74 (1996).
151. D. Volmer, K. Levson, G. Wünch, 'Thermospray Liquid Chromatographic–Mass Spectrometric Multi-residue Determination of 128 Polar Pesticides in Aqueous Environmental Samples', *J. Chromatogr. A*, **660**, 231–248 (1994).
152. S.J. Stout, A.R. DaCunha, M.M. Safarpour, 'Simplified Determination of Imidazolinone Herbicides in Soil at Parts-per-billion Level by Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry', *J. AOAC® Int.*, **80**(2), 426–432 (1997).
153. Z. El Rassi, 'Capillary Electrophoresis of Pesticides', *Electrophoresis*, **18**, 2465–2481 (1997).
154. A.J. Krynitsky, D.M. Swineford, 'Determination of Sulfonylurea Herbicides in Grains by Capillary Electrophoresis', *J. AOAC® Int.*, **78**(4), 1091–1096 (1995).
155. F. Garcia, J. Henion, 'Fast Capillary Electrophoresis–Ion Spray Mass Spectrometric Determination of Sulfonylureas', *J. Chromatogr. A*, **606**, 237–247 (1992).
156. G.M. McGaughlin, A. Weston, K.D. Hauffe, 'Capillary Electrophoresis Methods Development and Sensitivity Enhancement Strategies for the Separation Industrial and Environmental Chemicals', *J. Chromatogr. A*, **744**(1–2), 123–134 (1996).
157. S.C. Beale, 'Capillary Electrophoresis', *Anal. Chem.*, **70**(12), 279R–300R (1998).
158. D.Y. Chen, K. Adelhelm, X.L. Chen, N.J. Dovichi, 'A Simple Laser-induced Fluorescence Detector for Sulforhodamine 101 in a Capillary Electrophoresis System: Detection Limits of 10 Yoctomoles or Six Molecules', *Analyst*, **119**, 349–352 (1994).
159. A. Farran, S. Ruiz, C. Serra, M. Aguilar, 'Comparative Study of High-performance Liquid Chromatography and Micellar Electrokinetic Capillary Chromatography Applied to the Analysis of Different Mixtures of Pesticides', *J. Chromatogr. A*, **737**(1), 109–116 (1996).
160. A.D. Lucas, S.J. Gee, B.D. Hammock, J.N. Seiber, 'Integration of Immunochemical Methods with Other Analytical Techniques for Pesticide Residue Determination', *J. AOAC® Int.*, **78**(3), 585–591 (1995).
161. S.J. Gee, B.D. Hammock, J.M. Van Emon, *Environmental Immunochemical Analysis for Detection of Pesticides and Other Chemicals: a User's Guide*, Noyes Publications, Westwood, NJ, 1996.
162. D. Barceló, M.-C. Hennion, *Trace Determination of Pesticides and their Degradation Products in Water*, Elsevier, New York, 1997.
163. B.M. Kaufman, M. Clower, Jr, 'Immunoassay of Pesticides: an Update', *J. AOAC® Int.*, **78**(4), 1079–1090 (1995).
164. G.S. Nunes, D. Barceló, 'Analysis of Carbamate Insecticides in Foodstuffs Using Chromatography and Immunoassay Techniques', *Trends Anal. Chem.*, **18**(2), 99–107 (1999).
165. D.A. Kurtz, J.H. Skerritt, L.H. Stanker (eds.), *New Frontiers in Agrochemical Immunoassay*, AOAC® International, Arlington, VA, 1995.
166. D. Barceló (ed.), 'Molecular Imprints and Related Approaches for Solid-phase Extraction and Sensors in Chemical Analysis', *Trends Anal. Chem.*, **18**(3), 137–225 (1999).
167. P.M. Krämer, 'Biosensors for Measuring Pesticide Residues in the Environment', *J. AOAC® Int.*, **79**(6), 1245–1254 (1996).
168. J.-L. Marty, D. Garcia, R. Rouillon, 'Biosensors: Potential in Pesticide Detection', *Trends Anal. Chem.*, **14**(7), 329–333 (1995).

Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of

Alfonso Di Muccio
Istituto Superiore di Sanità, Rome, Italy

1 Introduction **2**
1.1 Dichlorodiphenyltrichloroethane and Related Compounds 2
1.2 Hexachlorocyclohexanes 2
1.3 Chlorocyclodienes 5
1.4 Camphechlor 5
1.5 Other Relevant Organochlorinated Compounds 6
1.6 Pyrethrins and Pyrethroid Insecticides 6
1.7 Analytical Methods 6
2 Sample Preparation **6**
3 Extraction **9**
3.1 Nonfatty Samples 9
3.2 Fatty Samples 9
3.3 Matrix Solid-phase Dispersion Extraction 10
3.4 Soil and Sediments 10
3.5 Water 11
3.6 Pressurized Liquid Extraction 11
3.7 Supercritical Fluid Extraction 12
4 Cleanup **12**
4.1 Liquid–Liquid Partitioning 12
4.2 Size-exclusion Chromatography 14
4.3 Sweep Codistillation – Assisted Distillation 16
4.4 Column Chromatography 16
4.5 Solid Phase Extraction Chromatography 18
4.6 High-performance Liquid Chromatography 18
4.7 Thin-layer Chromatography 19
4.8 Degradative Cleanup 19
5 Identification and Determination **20**
5.1 Gas Chromatography 20
5.2 High-performance Liquid Chromatography 27
5.3 Thin-layer Chromatography 28
Abbreviations and Acronyms **28**
Related Articles **29**
References **30**

Organochlorine (OC) pesticides is a general term for a broad class of compounds including some well-known compounds such as dichlorodiphenyltrichloroethane (DDT), lindane, dieldrin, endrin, heptachlor epoxide (HEPO), endosulfan, and chlordane. OC pesticides are the first class of compounds introduced in agricultural and civil uses to counteract noxious insects and insect-borne diseases. In general they are lipophilic compounds with noticeable chemical and environmental stability. Although most OC pesticides have been progressively restricted and then banned in the 1970s in most industrialized countries, a widespread environmental pollution has resulted from their use in agriculture and civil uses.

Pyrethrins are natural insecticidal compounds found in the extract of pyrethrum flowers. Also, a number of pesticides have been synthesized that share the biologically relevant chemical moiety with the pyrethrins and are referred to as pyrethroids (PYR). These compounds possess equal or better biological activity and better chemical stability that allow their use in agriculture, husbandry, and civil applications. Both OC and PYR pesticides can occur in fruits and vegetables as well as in food of animal origin and environment. The levels can be different as a result of direct application or indirect contamination. Such different matrices and levels of contamination require different analytical approaches and resources.

A review is presented of the array of the analytical techniques most frequently applied for the extraction, cleanup and identification/determination. Extraction techniques include solvent extraction, supercritical fluid extraction (SFE), solid-phase extraction (SPE), and solid-phase microextraction (SPME). Solvent extraction is carried out in different ways depending on the type of matrix (fatty and nonfatty samples, soil, sediments, and water), such as Soxhlet, liquid–liquid extraction (LLE), pressurized liquid extraction (PLE), and matrix solid-phase dispersion (MSPD).

Different cleanup techniques are presented that can be applied to fatty and nonfatty samples. They include liquid–liquid partitioning (LLP), size-exclusion chromatography (SEC), sweep codistillation and assisted distillation, column adsorption chromatography, chromatography on SPE cartridges, high-performance liquid chromatography (HPLC), thin-layer chromatography (TLC) and degradative cleanup. The techniques are presented as modular steps that can be arranged in different ways to cope with cleanup requirements posed by different sample composition and different selectivity/sensitivity of the identification/determinative techniques. The performances of the cleanup steps are discussed in terms of matrix removal

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and amount of sample that can be handled. Also the possibilities of linking different steps and the chance for automation of the cleanup process are indicated. Some environmental pollutants, such as polychlorinated biphenyl (PCB), have similar analytical behavior to some OC pesticides and can occur in the same environmental matrices. Some cleanup steps useful for the group separation of PCB and OC pesticides are presented.

The techniques for identification and determination include gas chromatography (GC) with selective detectors, HPLC, and their combination with mass spectrometry (MS). The importance of the combination of responses from different analytical techniques to obtain reliable identification at trace levels is underlined.

1 INTRODUCTION

Pesticides are chemicals intended to combat unwanted or noxious organisms in a variety of matrices in different fields of application such as agriculture, husbandry, civil, and industrial premises. Pesticides belong to a wide range of chemical classes and are divided into categories of organisms they counteract (insecticides for insects, acaricides for mites, nematocides for nematodes, fungicides for fungi, and herbicides for weeds).

The modern history of pesticides dates back to World War II when for the first time the insecticidal properties of DDT were recognized. DDT was first introduced on a large scale to fight fleas, lice, flies, and mosquitoes and reduce the spread of insect-borne diseases such as typhus (body lice), malaria, and yellow fever (mosquitoes) and plague (fleas).[1]

The first synthetic pesticides belong to the broad class of OC pesticides which includes the chloro diphenyl aliphatics (DDT, dichlorodiphenyldichloroethane (DDD), methoxychlor), the hexachlorocyclohexanes (HCHs), chlorocyclodienes (heptachlor, aldrin, chlordane, dieldrin, endrin, and mirex).[2,3] Table 1 gives chemical formulas of the cited OC pesticides.

Pesticides as well as other chemicals released into the environment tend to distribute into different compartments and their distance from the site of application depends on the chemico-physical properties and chemical stability of the pesticide and the environmental conditions. Pesticide degradation processes involve biotic (vegetable and animal metabolism, microorganism and enzymes) and abiotic (oxidation, hydrolysis, breakdown by heat and ultraviolet (UV) light) transformations. OC pesticides are generally recognized as highly stable lipophilic compounds. Thus, they have a tendency for bioaccumulation in fatty tissues of animals along the food chain.

1.1 Dichlorodiphenyltrichloroethane and Related Compounds

DDT belongs to the class of diphenyl aliphatics, which consists of an aliphatic chain with two phenyl rings attached. DDT was produced as a technical material comprising the *p*,*p*′-DDT and *o*,*p*′-DDT isomers and the related compounds *p*,*p*′-DDD (TDE), *o*,*p*′-DDD (TDE), *p*,*p*′-DDE (dichlorodiphenyldichloroethylene) and *o*,*p*′-DDE. A typical sample of technical DDT comprised 77.1% *p*,*p*′-DDT, 14.9% *o*,*p*′-DDT, 0.3% *p*,*p*′-DDD, 0.1% *o*,*p*′-DDD, 4.0% *p*,*p*′-DDE, 0.1% *o*,*p*′-DDE, with the remaining 3.5% being unidentified compounds. Besides being present in the technical product, some of these compounds are also metabolites of DDT. Two main pathways of DDT metabolism exist in mammals, i.e. dehydrochlorination to DDE and stepwise conversion to DDA (di(*p*-chlorophenyl)acetic acid) via DDD. DDE and DDD are also formed in insects, soil, and to a lesser extent in vegetables. In many industrialized countries DDT has been progressively restricted and then banned in the 1970s. However, it is still used in some countries to control malaria. Therefore, from an analytical perspective, the determination of the complex of isomers and metabolites is important at residue level.[4] Related pesticides include TDE (DDD) (which has not been used to a great extent), methoxychlor and dicofol.

Methoxychlor is an insecticide. Owing to the presence of two methoxy groups at *para* positions on the phenyl ring, methoxychlor is less environmentally stable than DDT and it is degraded via *O*-dealkylation and dehydrochlorination. Thus, methoxychlor has a poor tendency to bioaccumulation.

Dicofol is a nonsystemic acaricide used to control phytophagus mites. It is degraded in plants and animals to dichlorobenzophenone (DCBP).[5] Its synthesis and degradative pathways have some steps in common with DDT and its metabolites.[6]

1.2 Hexachlorocyclohexanes

HCH is made by chlorination of benzene, which results in the formation of five isomers differing in the spatial position of chlorine atoms and named using the Greek letters α, β, γ, δ, and ε. Only the γ-isomer (lindane) has insecticidal activity and is present at about 14–15% in the technical material, which also typically contains 65–70% α-HCH, 7–10% β-HCH, and ca. 10% of δ- and ε-isomers. Technical material was initially used for insect control, but after recognizing its undesirable characteristics (musty odor and flavor), only lindane continues to be used for a limited type of application, such as seed treatment, in industrialized countries.[4] However, the other isomers are more persistent than lindane and

Table 1 Chemical formulas of some OC pesticides

Compound	CAS reg. no.	Molecular formula	Molecular weight	Chemical formula
HCB	118-74-1	C_6Cl_6	284.8	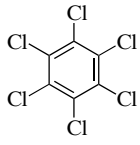
Quintozene	82-68-8	$C_6Cl_5NO_2$	295.3	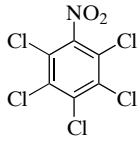
γ-HCH	58-89-9	$C_6H_6Cl_6$	290.8	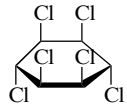
cis-Chlordane (α)	5103-71-9	$C_{10}H_6Cl_8$	409.8	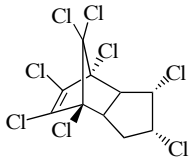
trans-Chlordane (γ)	5103-74-2	$C_{10}H_6Cl_8$	409.8	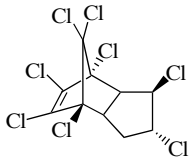
Oxychlordane	27304-13-8	$C_{10}H_4Cl_8O$	423.8	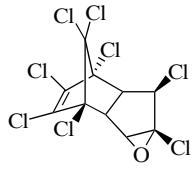
cis-Chlordene (α)	56534-02-2	$C_{10}H_6Cl_6$	338.9	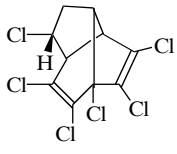
trans-Chlordene (γ)	56641-38-4	$C_{10}H_6Cl_6$	338.9	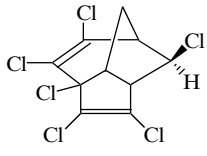
Heptachlor	76-44-8	$C_{10}H_5Cl_7$	373.3	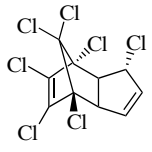
cis-HEPO (β)	1024-57-3	$C_{10}H_5Cl_7O$	389.3	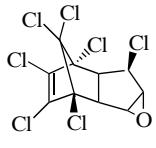

(*continued overleaf*)

Table 1 (*continued*)

Compound	CAS reg. no.	Molecular formula	Molecular weight	Chemical formula
α-Endosulfan (I)	959-98-8	$C_9H_6Cl_6O_3S$	406.9	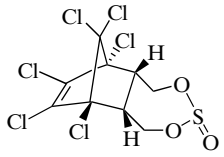
β-Endosulfan (II)	33213-65-9	$C_9H_6Cl_6O_3S$	406.9	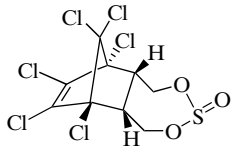
Endosulfan sulfate	1031-07-8	$C_9H_6Cl_6O_4S$	422.9	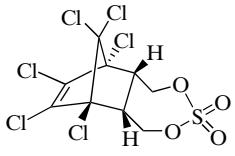
Dieldrin	60-57-1	$C_{12}H_8Cl_6O$	380.9	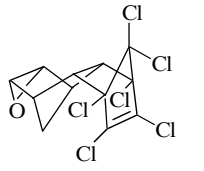
Endrin	72-20-8	$C_{12}H_8Cl_6O$	380.9	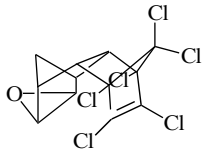
Aldrin	309-00-2	$C_{12}H_8Cl_6$	364.9	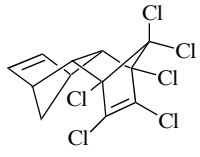
o,p′-DDE	3424-82-6	$C_{14}H_8Cl_4$	318.0	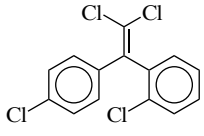
p,p′-DDE	72-55-9	$C_{14}H_8Cl_4$	318.0	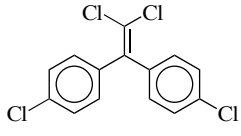
o,p′-DDD	53-19-0	$C_{14}H_{10}Cl_4$	318.0	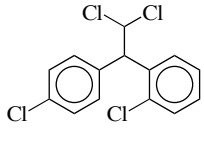
p,p′-DDD	72-54-8	$C_{14}H_{10}Cl_4$	320.0	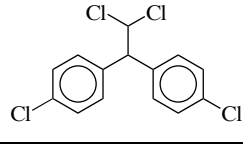

Table 1 (*continued*)

Compound	CAS reg. no.	Molecular formula	Molecular weight	Chemical formula
o,p′-DDT	789-02-6	$C_{14}H_9Cl_5$	354.5	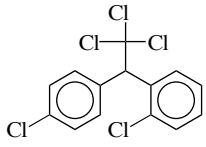
p,p′-DDT	50-29-3	$C_{14}H_9Cl_5$	354.5	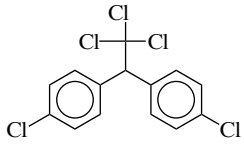
Dicofol	115-32-2	$C_{14}H_9Cl_5O$	370.5	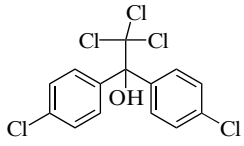
Methoxychlor	72-43-5	$C_{16}H_{15}Cl_3O_2$	345.7	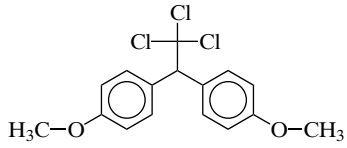
4,4′-DCBP	90-98-2	$C_{13}H_8Cl_{20}O$	250	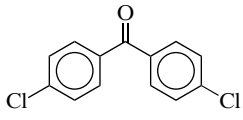
Mirex	2385-85-5	$C_{10}Cl_{12}$	546	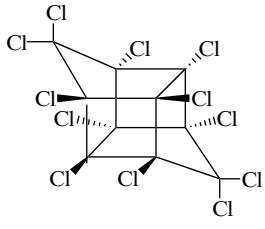
Camphechlor (toxaphene)	8001-35-2	$C_{10}H_{10}Cl_8$	413.8	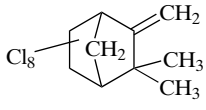

may still be found at residue levels in the environment and animal tissues.

1.3 Chlorocyclodienes

The main compounds in this class are chlordane, aldrin and dieldrin, heptachlor, endrin, mirex, α-endosulfan and β-endosulfan. They are persistent insecticides which may be used in some countries mostly in soil applications. Because of their persistence, the use of chlorocyclodienes on crops, except endosulfan, was restricted and/or banned between 1975 and 1980.

Technical endosulfan consists of a mixture of α- and β-isomers in the approximate ratio of 70:30 (w/w). Endosulfan is less persistent than the other chlorocyclodienes, and oxidizes into endosulfan sulfate. Dieldrin, besides being an insecticide per se, is the oxidation product of aldrin. Heptachlor and chlordane also oxidize into heptachlorepoxide and oxychlordane, respectively. Other compounds such as *trans*-nonachlor are present at residue level as byproducts of the synthesis of heptachlor, while hexachlorobutadiene is a common byproduct of the chlorination process. Chlordane has two isomers, α and γ, and is also present in technical heptachlor.[1–4]

1.4 Camphechlor

Camphechlor (toxaphene) is manufactured from the chlorination of camphene and consists of more than

177 polychlorinated derivatives, most of them being isomeric C_{17}, C_{18}, and C_{19} bornanes. Toxaphene is an insecticide mainly used in cotton fields and for the control of termites. Toxaphene is persistent in soil but is fairly easily metabolized by mammals and birds, and does not accumulate in body fat to such a great extent as do DDT, HCH, or chlorocyclodienes. Despite its low toxicity to mammals, insects, and birds, fishes are highly susceptible to toxaphene poisoning.[1–4]

These compounds as well as PCBs occur in the environment as complex mixtures at trace level and require special analytical methodologies which are not specifically dealt with in this article.[7]

1.5 Other Relevant Organochlorinated Compounds

Hexachlorobenzene (HCB) is actually a fungicide used in the past in the treatment of stored grains. It is highly lipophilic and is found ubiquitously in environmental fatty matrices.[4] It is also found as a contaminant of quintozene (pentachloronitrobenzene), a fungicide used in soil treatments.[4]

PCBs are materials used mostly in the past years in industrial applications such as dielectrics, heat exchangers, and brake and hydraulic fluids. Accidental release and/or inadequate waste disposal has lead to widespread environmental pollution. Because of their stability and lipophilicity they pose almost the same problems as OC insecticides and behave similarly in most analytical procedures.[3]

A number of other polychlorinated compounds are of concern in environmental pollution including polychloronaphthalenes, polychloroterphenyls,[4] polychlorodibenzodioxins and polychlorodibenzofurans.[4]

1.6 Pyrethrins and Pyrethroid Insecticides

With the development of pesticide chemistry other classes of insecticides have been introduced. These include organophosphates and carbamates that are generally much more acutely toxic to vertebrates than OCs insecticides. They are less chemically stable and persistent than OCs and do not give rise to environmental bioaccumulation.

Pyrethrins are insecticidally active compounds present in the extract of pyrethrum flowers. Pyrethrin extract contains six related compounds, but the analysis is focused on the main and more active compounds, pyrethrins I and II, which are present at about 33–35% each in the extract. Pyrethrins are normally used in formulations containing the synergist piperonyl butoxide, which is also of analytical interest. Pyrethrins have rapid knock-down activities on insects, but are not suitable for agricultural uses because of their instability in sunlight.

One of the last classes of insecticides introduced is PYR. In the course of several years a number of compounds have been synthesized that possess some of the characteristics of pyrethrins combined with sufficient environmental stability and residual activity for agricultural uses. Most PYR pesticides were first introduced on the market as a mixture of stereoisomers resulting from the synthesis. Individual enantiomers have different biological activity and are increasingly used in formulations. Also the degradation of the enantiomers is different in the environment and at residue level.[4]

Some of the above-mentioned compounds are shown in Table 2.[5]

Depending on the chemical structure, stability, chemicophysical properties, environmental conditions, and time and mode of application, pesticides used in agriculture and in civil applications or sanitary programs may give rise to residues in crops at harvesting and/or in different environmental matrices. In general terms, residues occur as high as a few milligrams per kilogram (ppm, parts per million) for pesticides used in agriculture, but indirect, long-lasting contamination by persistent compounds such as OC pesticides may be found ubiquitously at microgram per kilogram (ppb, parts per billion; 1 in 10^9), nanogram per kilogram (ppt, parts per trillion; 1 in 10^{12}), and femtogram per kilogram (ppq, parts per quadrillion; 1 in 10^{15}), levels.

1.7 Analytical Methods

It is obvious that analytical determinations at such widely different levels pose different difficulties. However, a tremendous number of publications have been devoted to the development of methods for pesticide residue analysis in foods and environmental matrices.

The determination of OC and PYR pesticides follows some common approaches applicable in general to the determination of trace amounts of organic compounds in complex matrices.[8–17]

2 SAMPLE PREPARATION

The analytical method is generally considered to begin with sample preparation. Sample preparation involves the isolation of the sample portions that are indicated in the definition of the legal limit, or maximum residue level (MRL), for foods, or the portion of the substrate for which the analysis is intended. Afterwards, this sample portion (*analytical sample*) is taken for analysis. The analytical sample is often larger than the amount strictly needed for the analysis from a technical point of view, but the larger size may be more convenient to work with or lead to a more representative sampling procedure. Otherwise,

Table 2 Chemical formulas of some PYR pesticides

Compound	CAS reg. no.	Molecular formula	Molecular weight	Chemical formula
Pyrethrin I	121-21-1	$C_{21}H_{28}O_3$	328.4	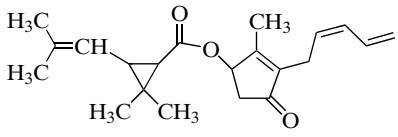
Pyrethrin II	121-29-9	$C_{22}H_{28}O_5$	372.4	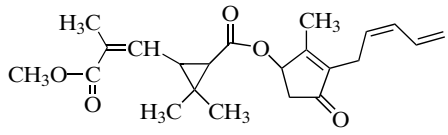
Piperonyl butoxide	51-03-6	$C_{19}H_{30}O_5$	338.4	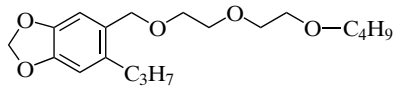
Tefluthrin	79538-32-2	$C_{17}H_{14}ClF_7O_2$	418.7	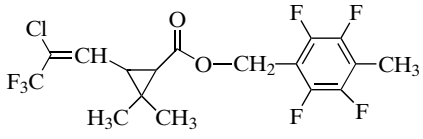
Tetramethrin	7696-12-0	$C_{19}H_{25}NO_4$	331.4	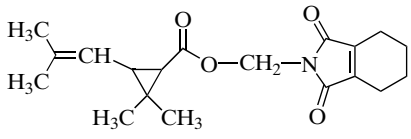
Cyphenothrin	39515-40-7	$C_{24}H_{25}NO_3$	375.5	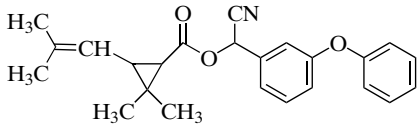
Cyfluthrin	68359-37-5	$C_{22}H_{18}Cl_2FNO_3$	434.3	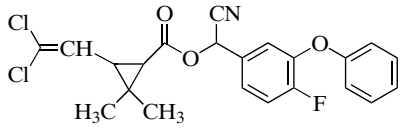
Flucythrinate	70124-77-5	$C_{26}H_{23}F_2NO_4$	451.4	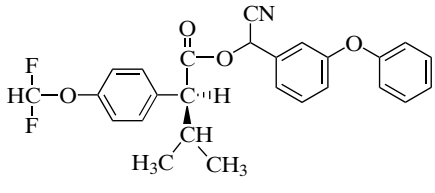
Fluvalinate-τ	102851-06-9	$C_{26}H_{22}ClF_3N_2O_3$	502.9	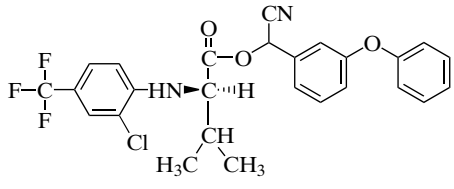

(*continued overleaf*)

Table 2 (*continued*)

Compound	CAS reg. no.	Molecular formula	Molecular weight	Chemical formula
Deltamethrin	52918-63-5	$C_{22}H_{19}Br_2NO_3$	505.2	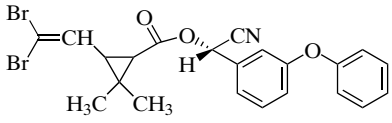
Bioallethrin	584-79-2	$C_{19}H_{26}O_3$	302.4	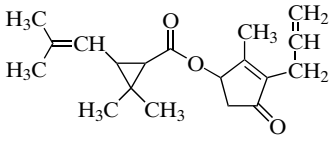
Fenpropathrin	64257-84-7	$C_{22}H_{23}NO_3$	349.4	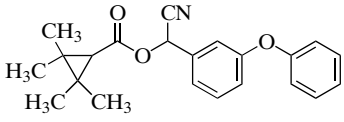
λ-Cyhalothrin	91465-08-6	$C_{23}H_{19}ClF_3NO_3$	449.9	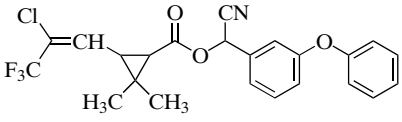
Permethrin	52645-53-1	$C_{21}H_{20}Cl_2O_3$	391.3	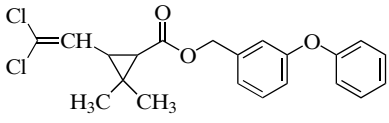
Cypermethrin	52315-07-8	$C_{22}H_{19}Cl_2NO_3$	416.3	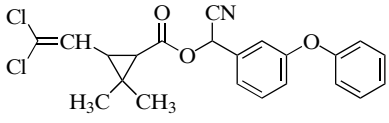
Esfenvalerate	66230-04-4	$C_{25}H_{22}ClNO_3$	419.9	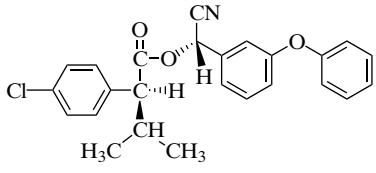
Tralomethrin	66841-25-6	$C_{22}H_{19}Br_4NO_3$	665	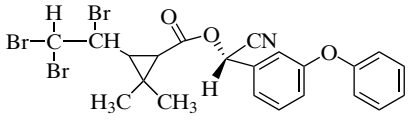
Bifenthrin	82657-04-3	$C_{23}H_{22}ClF_3O_2$	422.9	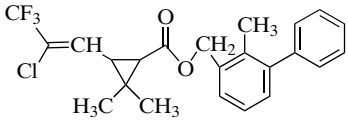

a procedure should be applied to derive the *aliquot of analysis*, i.e. the portion of the sample brought to the extraction step. In order to have a representative aliquot, the laboratory/analytical sample should be thoroughly comminuted (chopped, or milled) and mixed to obtain an homogeneous material. If possible, the entire laboratory/analytical sample should be processed. If this is not possible, a reducing procedure such as quartering (division into quarters and taking opposite corners), is applied. When applying a quartering procedure to large

units, they should be divided according to their symmetry, taking into account that the distribution of pesticide residues may not be even.

The differences in the sample structure may also require different procedures and mechanical equipment to get a homogeneous analytical sample. Most biological samples are better comminuted while frozen. For this purpose, liquid nitrogen can be useful (e.g. cryogenic milling). The addition of dry ice keeps the sample frozen and increases the mechanical shredding. Comminution at low temperature is also recommended to avoid decomposition of heat-labile compounds. Lyophilization is sometimes used to remove the water from the samples in order to increase the aliquot taken for the analysis and/or to facilitate the extraction, but may give rise to losses by evaporation.

The action of chopping, cutting, blending, or milling machines should be evaluated with representative crops or goods to be analyzed in order to check the performance with samples of different texture and to assess the minimum size of the aliquot of analysis that can be still considered representative of the laboratory sample.[18] Homogenization is particularly important when using methods that require a small aliquot for analysis. The variance of the sample preparation adds to the variance of the analytical procedure and can contribute significantly to the overall quality (accuracy and precision) of the results.[19]

3 EXTRACTION

3.1 Nonfatty Samples

Extraction is the step aimed at separating the analytes from the insoluble part of the matrix and transferring them into a solvent amenable to the subsequent steps of the method. The use of different solvents and solvent mixtures has been reported, depending on the analytes and matrices. In multiresidue methods, the extraction solvent should extract polar and apolar analytes because it is of interest to extract as many analytes as possible. Increasing the polarity of the solvent improves the extraction of polar analytes, but generally increases the amount of coextractives.

For nonfatty samples (commonly those containing less than 2–5% fat), water-miscible solvents such as acetonitrile,[8,9,20,21] acetone,[22–29] ethanol,[30] and methanol,[31,32] or polar, partially water-miscible solvents such as ethyl acetate,[33–35] are among the most frequently used. Most methods require 2 mL solvent per gram of sample. Although acetonitrile is known to give cleaner (less waxes and fats) extracts than acetone it has some drawbacks such as higher toxicity, higher cost, high boiling point, and the formation of a two-phase system with water when fruits or vegetables with low water and/or high sugar content are analyzed.[36] Alternatively, acetone is less toxic and expensive, and has a lower boiling point. Methanol also has been less frequently used because of its toxicity, although it is indicated as a better extracting solvent for some categories of pesticides such as carbamates,[9] but does not work as well for nonpolar analytes.

Low-moisture foods, such as grains, are better extracted with solvent mixtures such as acetonitrile–water (65 : 35 v/v),[8,37] acetone–methanol (1 : 1 v/v),[38] or acetone–water (8 : 1 v/v).[39] Water helps to swell the dried material, thus allowing a better penetration of the water-miscible solvent. Both water and methanol generally improve the extraction by reducing the adsorption of the analytes on insoluble material and container surface, but this is not necessarily the case for nonpolar compounds. The process of extraction is carried out by homogenizing a portion of the previously comminuted sample in high-speed blenders/homogenizers. The solvent is separated by filtration after the addition of filter aids, such as diatomaceous materials, or after centrifugation. In the extraction with ethyl acetate or other solvents immiscible with water, anhydrous sodium sulfate is added to remove the water.

3.2 Fatty Samples

Most methods for fatty foods have been developed for PYR, OC and organophosphorus (OP) pesticides due to the lipophilic nature of these pesticides. The most common approach entails thorough extraction of the fat to completely extract lipophilic pesticides. Therefore, extraction procedures for pesticides follow the classical methods for the determination of total fat content. For instance, animal tissues and dairy products are thoroughly mixed with anhydrous sodium sulfate to remove water and help disintegrate the tissues. Then the mixture is extracted with a Soxhlet apparatus with petroleum ether,[40] or *n*-hexane–acetone (59 : 41 v/v) mixture,[41] or *n*-hexane–dichloromethane (DCM) (4 : 1 v/v),[42] or blended at high speed.[43] The solvent is evaporated and the raw fatty extract is obtained after evaporation of the extracting solvent,[40,42] or after partition into *n*-hexane and evaporation.[43]

In some cases, an extraction step may not be needed. For example, butter fat is obtained from butter by melting and filtering to remove the solids. Fats and oils can be subjected directly to one of the cleanup steps.

Milk and dairy products are usually extracted with the total fat extraction approach using petroleum ether–diethyl ether (1 : 1 v/v), after the addition of denaturing agents such as potassium oxalate, which solubilizes the proteins, and methanol or ethanol, that avoids the formation of emulsions. Also human serum has been

extracted with petroleum ether–diethyl ether in the presence of methanol.[44] Extraction of total milk fat and lipophilic pesticides is also possible with a combination of apolar and polar solvents, such as *n*-hexane–acetone (50 : 50 v/v),[40,45] ethyl acetate–ethanol (95 : 9 v/v),[46] or ethyl acetate–acetone–methanol (2 : 4 : 4 by vol.),[47] or with acetone alone[48] or acetonitrile.[49]

Eggs have been extracted according to the total fat extraction approach by drying with anhydrous sodium sulfate and blending with a mixture of *n*-hexane–acetonitrile (2 : 1 v/v),[50] which has been proved to extract more fat than extraction with petroleum ether–diethyl ether.[8]

In a different approach, fatty tissues can be extracted with acetonitrile.[49,51] In this case, a selective extraction of pesticide residues including OC and PYR pesticides is accomplished. Pesticide residues are extracted efficiently while most of the fatty matrix is left, with great advantages in the subsequent cleanup process.

Also, for milk a selective extraction of pesticide residues is possible by mixing the milk with acetonitrile and ethanol and partitioning into *n*-hexane,[52] or by mixing with acetonitrile and passing the resulting solution through an SPE C_{18} cartridge (SPE with octadecyl-silica), which traps the OC pesticides.[49] In both cases OC pesticides are recovered efficiently accompanied by a substantially lower amount of fat than in the conventional total fat extraction approach. This amount of fat allows a small-scale cleanup with saving of solvents, adsorbents, and time.

The above-mentioned extraction processes are carried out by putting into intimate contact the sample and the solvent. Mechanical blenders/homogenizers are most commonly used, and ultrasound is often applied to this end. The extracting solvent is separated by centrifugation or by filtration through paper or glass fiber disks, often after the addition of filtration aids, usually diatomaceous earth. Not infrequently emulsions and/or jelly suspensions are formed that complicate the separation of the extracting solvent. Addition of small amounts of polar solvents or saline solution and centrifugation are often useful to overcome this problem.

3.3 Matrix Solid-phase Dispersion Extraction

In a different approach, MSPD (sometimes referred to as on-column extraction), the intimate contact between the sample and the solvent is realized by thoroughly mixing the finely divided sample with solid sorbents. Common sorbents used in MSPD include diatomaceous earth material (Celite, Extrelut, Hydromatrix), synthetic magnesium silicate (Florisil), or silica. The sample plus material mixture is packed in a column and eluted with the solvent under gravity at ambient temperature. This column extraction has been reported for vegetables mixed with Extrelut (10 g vegetable + 12 g Extrelut eluted with 100 mL DCM)[53,54] or Hydromatrix (25 g sample + 25 g Hydromatrix eluted with 250 mL ethyl acetate),[55] or Florisil (5 g vegetable + 8 g Florisil eluted with 70 mL DCM–acetone (9 : 1 v/v) or ethyl acetate).[56] In a similar application 5 g vegetable was mixed with 8 g Florisil and extracted on-column with 60 mL *n*-hexane–acetone (9 : 1 v/v).[57]

Milk and dairy products, and eggs, were mixed with silica gel (5 g sample + 15 g silica gel), while fish tissue was mixed with sodium sulfate and clean sand (1 g fish + 3 g sea-sand and 10 g sodium sulfate). In both cases, the sample mixture was loaded onto the top of a 10% water-deactivated silica gel column, and the bilayer column eluted overnight with 150 mL petroleum ether–DCM (80 : 20 v/v).[58,59] In these applications, the on-column extraction performs both the extraction and cleanup, because the extract is passed in-line through the silica gel adsorption column. In similar applications, OC pesticides can be extracted from fish tissue homogeneously mixed with C_{18} powder (0.5 g tissue + 2 g C_{18}). The mixture is put on the top of a small column of Florisil. The bilayer column is eluted with 8 mL acetonitrile.[60–62] The MSPD takes advantage of the sample shredding generated while grinding the sample with irregularly shaped solid particles, and lipid-retaining ability of a lipophilic support (C_{18}). Thus, the bilayer column carries out both a selective extraction and a cleanup step (Florisil) in a single process. MSPD uses mild conditions for extraction (ambient temperature and pressure) and has been found useful in avoiding the degradation of dicofol to DCBP during Soxhlet extraction of avian eggs.[63] Preferential flow and sample that has not been carefully mixed are the principal drawbacks that can result in poor recovery.

Also the selective extraction of OC, OP, and PYR pesticides from milk and milk powder can be carried out by loading a mixture of milk or reconstituted milk with acetonitrile and ethanol (5 mL of a mixture of 10 mL milk + 5 mL acetonitrile and 1 mL ethanol) into a cartridge filled with Hydromatrix. The OC, OP, PYR pesticides are selectively recovered from the material by eluting the cartridge with 20 mL *n*-hexane saturated with acetonitrile and ethanol. Milk fat is carried over to a limited extent (ca. 5 mg mL^{-1} of milk) compared to the nominal fat content (ca. 180 mg mL^{-1}) of the whole milk and the subsequent cleanup can be miniaturized to save solvents, reagents, and time.[64–66]

3.4 Soil and Sediments

OC pesticides are extracted from soil by shaking the sample overnight with *n*-hexane–acetone (1 : 1 v/v) (typically 10 g with 100 mL) after the addition of ammonium chloride solution.[8] Fenpropathrin and fluvalinate have

been extracted from soil mixed with charcoal (25 g soil + 0.5 g charcoal) by column extraction with 100 mL *n*-hexane–acetone (9 : 1 v/v).[67] Lower solvent/soil ratio (20 mL/10 g) with methanol and sonication for 15 min proved more efficient and rapid than the conventional shaking method or Soxhlet extraction in extracting some PYR pesticides from soil.[68] Sonication provides a more efficient contact between the solid and the solvent, but longer sonication caused a decrease in recovery, probably due to some degradation of the compounds. Sonication for 5 min and high solvent/soil ratio (150 + 100 mL/20 g) were found necessary in a multiresidue method for extracting 120 pesticides, including OC and PYR, from high organic matter soil.[69]

To improve the efficiency of the extraction, heat can be supplied by microwave energy. The principle involves heating both the solvent and the matrix by wave–matter interaction. The process is carried out in a special microwave oven (typical conditions: 10 g soil and 30 mL *n*-hexane–acetone (1 : 1 v/v) at 110 °C for 10 min). As a result both temperature and pressure are increased and there is a chance of degradation of sensitive compounds by exposure to high temperature and catalytic sites in the soil.[70–72]

3.5 Water

Water or wine is extracted by classical LLE in a separatory funnel by shaking it with an immiscible solvent such as DCM.[73,74] To increase the sensitivity and avoid concentration of the extracting solvent, OC and PYR pesticides can be extracted from water with a small volume of *n*-hexane (1 mL). After the extraction, the hexane layer is raised up to the bottle neck of the separatory funnel by raising the communicating vessel filled with deionized water.[75]

Water and water-diluted liquid samples can be passed through lipophilic SPE cartridges to extract OC and PYR pesticides. The most widely used sorbents are C_{18}, polystyrene–divinylbenzene (PS–DVB), and graphitized carbon.[76] A typical application includes extraction of 1 L of water,[77,78] or 50 mL of wine diluted with 200 mL of water,[79] both with a 500-mg C_{18} cartridge. Centrifuged human plasma (1 mL diluted with 3 mL water) was passed through a 100 mg C_{18} cartridge.[80]

Filter disks of polytetrafluoroethylene with C_{18} particles embedded have been used to speed up the extraction of large samples (about 20 min with a 47 mm diameter disk compared to about 1–2 h with a C_{18} cartridge, for a 1 L sample).[81,82] Automated SPE can be combined with large-volume GC introduction techniques for the automated extraction and determination of PYR pesticides in water.[76,83]

SPME is a solventless extraction technique in which the analytes are extracted directly from the aqueous samples onto a polymeric stationary phase that coats a fused silica fiber. The fiber is dipped into the aqueous sample under stirring (in certain applications the fiber is exposed to the headspace above the sample) allowing the adsorption of the analytes, according to their partition coefficient, into the polymeric phase. The analytes are thermally desorbed directly into a gas chromatograph by inserting the fiber into the hot injector. After a few minutes the fiber is withdrawn and reused for subsequent analysis. Different fiber coatings provide different adsorption properties for different compounds. The most commonly used fiber coating is poly(dimethylsiloxane), which is relatively nonpolar, and useful for nonpolar compounds such as OC pesticides. For more polar analytes a polyacrylate coating has been used. Fiber coating thickness is in the range 15–150 μm. Temperature and time of absorption and desorption have to be optimized in relation to the sample characteristics and analytes. The typical temperature for absorption is from ambient to 45 °C for 20–90 min, while the desorption temperature is in the range 250–260 °C for a few minutes. Elimination of solvents, easy automation, and the possibility of on-site sampling are among the advantages of SPME.[84–90] On-line solvent desorption into a liquid chromatographic column has been developed for compounds that require HPLC for analysis.[85]

3.6 Pressurized Liquid Extraction

PLE is a semiautomated technique in which extraction can be enhanced and speeded up, and solvent consumption reduced, by increasing temperature and pressure. As in MSPD, the sample is mixed and dispersed with macroporous materials such as Hydromatrix and/or anhydrous sodium sulfate. This acts to disperse the matrix, improve homogeneity, and ease flow of solvent through the sample. The mixture is loaded into a closed vessel through which the solvent is passed under increased temperature and pressure. The solvents are those normally used for conventional extraction procedures. The technique basically uses a static extraction step of 5 min at elevated temperature (50–200 °C) and pressure (7–20 MPa) that enhances solubilization and desorption of the analytes from the matrix, followed by a washing of the cell with fresh solvent for total recovery of the extracted analytes.[91] Dried materials such as soil are mixed with anhydrous sodium sulfate in about a 1 : 1 w/w ratio and extracted with *n*-hexane–acetone (1 : 1 v/v)[92] or DCM[93] for 5 min static extraction at 100 °C and 13.8 MPa under conditions similar to those of EPA Method 3545.[94] Other pesticides have been extracted from soil mixed with Hydromatrix (1 : 0.1 w/w) by acetonitrile under the same conditions.[95] PLE gives excellent recoveries in a short time and has been compared favorably with Soxhlet, sonication, microwave-assisted extraction (MAE) and

SFE.[71] Vegetable samples were mixed with Hydromatrix or Extrelut in a 1:1 w/w ratio and extracted with ethyl acetate at temperatures of 50–100 °C and a pressure of 143.8 MPa for 5 min. Typical sample sizes are 5–10 g mixed with inert material in a 11- or 22-mL vessel eluted with 20–40 mL. Lower temperature reduces the amount of water in the extract that has to be removed by anhydrous sodium sulfate. Furthermore, under increased temperature and pressure there is a chance of degradation as in MAE in a closed vessel. Instruments are commercially available under the trade names ASE, ESE, and PSE for processing multiple samples.

3.7 Supercritical Fluid Extraction

In SFE sample preparation is quite similar to that in PLE; the sample is extracted by passing a supercritical fluid (typically CO_2) through the cell. In some cases, a small amount of polar solvent modifiers (e.g. methanol, ethanol, or acetonitrile) is added to improve extraction, depending on the analytes and matrix. SFE is a relatively selective extraction technique which has been applied to the extraction of pesticide residues from vegetables, soils, and fatty tissues with a reduced removal of coextractives. However, expensive apparatus and selection of appropriate conditions are needed, and, in the case of fatty foods, further cleanup for OC pesticides is needed prior to determination. Typical sample sizes are 2–5 g in 10-mL vessels.

Fatty and nonfatty samples are mixed with Hydromatrix in a 1:1 to 1:2 w/w ratio to retain the water that would be extracted by supercritical CO_2.[96–99] Typical conditions are: 2 min static extraction followed by dynamic extraction at 20–32 MPa and 40–60 °C with CO_2 flow rates in the range 1.6–3 mL min^{-1} for six extraction vessel volumes. The extract is collected in a trap filled with chromatographic material such as C_{18},[97,98] alumina, Florisil or glass beads,[99] usually at ambient or subambient temperature. For lyophilized fish powder the trap was Florisil.[100] Inclusion of alumina in the extraction vessel served to reduce the extraction of fat from lyophilized fish. However, while PCBs were satisfactorily recovered, some degradation of DDT was observed.[101,102] A cartridge of chromatographic material such as silica gel, aminopropyl silica gel, or Florisil connected in-line downstream to the extraction vessel adequately separated PCBs from lipids when eluted with supercritical CO_2.[103] In another application to fatty samples, the extracted material is cleaned up by passing the supercritical fluid in-line through a preparative column of methyl silica (C-1) connected downstream to the extraction vessel.[104] A multivessel SFE instrument has been evaluated for the simultaneous extraction of up to six fatty samples.[105]

4 CLEANUP

Cleanup is the part of the analytical method aimed at selectively removing as much coextracted material as possible before the identification/determination step. The amount of coextracted material that needs to be removed is often dictated by the selectivity, sensitivity, and robustness of the determinative technique and by the concentrations at which the analytes are measured. Therefore, cleanup may consist of a single step or a combination of steps, some of which are essential and some of which can be omitted depending on the type of sample, extraction method, and the technique and instrumentation used for identification/determination. In some instances, the raw extract can be analyzed without cleanup. In any case, cleanup is a compromise involving efficiency, time, and cost. Although published methods report specific sequences of cleanup procedures (blocks), most cleanup steps can be arranged in desired sequences to separate particular types of chemicals, or deal with particular problems.

Simple operations can be used to link different cleanup blocks by tailoring the approach to most conveniently take into account the amount of coextracted material and type of solvent each block can accept. Solvent exchange, i.e. concentration to dryness and redissolution in a suitable solvent, is the most common operation between different steps of the analytical method.

Cleanup includes an array of chemical separations aimed at removing unwanted coextractives by taking advantage of physico-chemical differences in polarity, dissociation constants, volatility, molecular size, chemical stability, solubility in different solvents, and other factors.

4.1 Liquid–Liquid Partitioning

4.1.1 Nonfatty Samples

When fruits, vegetables, and other nonfatty samples are extracted with water-miscible solvents such as acetone, acetonitrile, and methanol the usual subsequent step in the approach involves the dilution of the extract with saline solution and the extraction of pesticide residues from the diluted raw extract with an immiscible solvent or solvent mixture by shaking in a separatory funnel. Hydrophilic coextractives remain preferentially in the watery phase, while nonpolar and semipolar pesticide residues partition into a solvent that is more easily evaporated and amenable to the subsequent cleanup steps. The more polar the extracting solvent, the wider the range of the analytes that can be recovered, but the amount of the coextractives generally increases. For recovery of low to medium polarity analytes, such as OC and PYR pesticides from aqueous acetone

or aqueous acetonitrile, a relatively nonpolar solvent such as petroleum ether achieves high recovery of the analytes and fewer nonfatty matrix components partition into the organic phase.[8,9,27,37,106,107] DCM, or DCM–petroleum ether (1:1 v/v), are more appropriate to recover pesticides with a wider polarity range.[22–26,38] With acetone as the extracting solvent of fruits and vegetables, after the partition into DCM about 0.5–2 mg of coextractives per gram of sample remain in the extract. Adjusting the pH to 7, generally with phosphate buffer, in the extraction or partition step prevents the partitioning of endogenous acids and is useful to reduce coextractives from fruits and vegetables, especially citrus fruits.[21] For vegetable samples extracted with methanol or ethanol, the analytes are usually partitioned into toluene.[30–32]

The LLP procedure is normally carried out in a separatory funnel, in which the solvent of higher density forms the lower layer. Only one separatory funnel is needed to perform multiple extractions, while solvents less dense than water (e.g. ethyl acetate) require multiple separatory funnels to perform multiple extractions. Emulsions (especially with sugar- and pectin-rich vegetables), manual labor, high glassware needs, and lengthy operation are the main drawbacks of traditional LLP.

Other ways of getting rid of the aqueous phase include the ''salting out'' process to separate the acetonitrile,[20,21] or the addition of water-immiscible solvents, such as DCM–*n*-hexane (1:1 v/v) or ethyl acetate, or a mixture thereof, to separate a mixture of solvents containing the analytes.[15,108] This type of partition leads to the separation of polar solvent mixtures, so that OC, PYR, and more polar compounds can be recovered. The operation is normally carried out in the same vessel used for the extraction. The volume of solvent separated should be recorded and taken into account in the calculation of the results.

Instead of the labor- and time-consuming operation using separatory funnels, the same type of partition can be advantageously carried out by absorbing a portion of the aqueous acetone extract into cartridges filled with macroporous diatomaceous material.[109] The material acts as a support where the partition takes place. Then, acetone is partially evaporated by an upward stream of nitrogen, and the cartridge is eluted with petroleum ether, DCM, or mixtures thereof. With different combinations of nitrogen flow and time, different amounts of acetone remain on the columns, and, combined with the selected solvent, the polarity range of the compounds that can be recovered is controlled. This also provides some control over the amount of coextractives removed from the cartridge. Low to medium polarity compounds, such as OC and PYR pesticides, can be recovered by reducing the acetone content to ca. 50% of the original (2:1 acetone–sample to 1:1 acetone–sample) and eluting with low polarity solvents, such as petroleum ether or petroleum ether–DCM (75:25 v/v).[109] Unlike using the separatory funnel, this approach uses disposable items, is simple and straightforward, and avoids emulsions.

Another way of recovering pesticide residues from aqueous acetone or aqueous acetonitrile is by SPE. The solution resulting from the extraction is diluted with water to bring the solvent concentration down to ca. 5–10% and, then, is passed through a lipophilic cartridge, such as PS–DVB[29] or C_{18},[110] or a C_{18} extraction disk.[111] Pesticide residues are retained in the solid phase while the aqueous phase is discarded. Pesticide residues are eluted from the cartridge with a solvent, for instance, ethyl acetate[29] or acetone–ethyl acetate (1:1 v/v).[111]

4.1.2 Fatty Samples

The extraction of fatty samples, which can include both plant (e.g. avocado) or animal tissues, commonly results in significant amounts of coextracted lipidic material. The separation of pesticide residues from the bulk of the coextracted fat is a mandatory step before an instrument-based identification/determination step. Classically, multiple separatory funnel partitions between immiscible solvents are used. The most commonly used solvent pair is *n*-hexane–acetonitrile,[8,9] whilst *n*-hexane–dimethylformamide (DMF)[112] and *n*-hexane–dimethyl sulfoxide (DMSO)[113] are also useful. Most of the lipidic material remains in the nonpolar solvent (e.g. *n*-hexane) while pesticide residues are partitioned into the polar solvent (e.g. acetonitrile) of the above-mentioned pairs. Unfortunately some of the most nonpolar pesticides such as HCB, mirex, and highly chlorinated PCBs also partition into the nonpolar phase. The so-called p-value, defined as the fraction of the analyte partitioning into the nonpolar phase of an equi-volume two-phase system, gives a measure of the ability of the polar solvent to extract the analyte of interest.[114] The higher the p-value, the lesser the partitioning into the polar solvent. OC and PYR pesticides show a wide range of p-values in the *n*-hexane–acetonitrile partition system, ranging from 0.91 (mirex) (lipophilic compound) to 0.035 (endosulfan sulfate) (more polar compound).[114] The p-values of some of the OC and PYR pesticides are reported in Table 3. The recovery into acetonitrile is almost complete after four partition stages using a 1:2 nonpolar–polar solvent ratio as it is normally applied.[8,9] However, the recovery of HCB and mirex (the most lipophilic OC pesticides) is barely satisfactory (ca. 70%). Being more polar than OC, PYR pesticides strongly partition into acetonitrile. Up to 3 g of fat dissolved in 15 mL

Table 3 p-Values of some OC and PYR pesticides between *n*-hexane, or isooctane, and acetonitrile

OC *n*-Hexane–acetonitrile[a]		PYR Isooctane–acetonitrile[b]	
Mirex	0.91	Bifenthrin	0.35
Aldrin	0.73	*trans*-Permethrin	0.21
p,p′-DDE	0.56	Cypermethrin	0.06
Endosulfan	0.39	Fluvalinate	0.02
p,p′-DDT	0.38	Deltamethrin	0.07
Lindane	0.13		
Endosulfan sulfate	0.035		

[a] p-Values from Beroza et al.[114]
[b] p-Values from Argauer et al.[115]

of *n*-hexane can be extracted with four 30-mL fractions of acetonitrile.[8,9] However, the removal of lipidic material is largely dependent on the type of fat. The removal is better with saturated fats, while it is less satisfactory with unsaturated oils such as fish oils. Furthermore, secondary components of fats, such as sterols and tocopherols, are extracted into the polar solvent and should be removed by further cleanup steps before GC with electron capture detection. MS detectors may avoid this problem.[115]

In this approach, once pesticides have been extracted into acetonitrile, this solvent is generally exchanged because it is not amenable to most of the subsequent analytical steps, and interferes with electron capture detection. Three main methods have been used.

First, acetonitrile can be concentrated to dryness by typical rotary evaporation techniques under reduced pressure. The azeotropic distillation of acetonitrile after the addition of methanol (acetonitrile–methanol, 4 : 1 v/v)[66] or addition of *n*-hexane (acetonitrile–*n*-hexane, 1 : 3.5 v/v)[21] helps to lower the boiling point and to reduce losses of volatile analytes (e.g. HCB). The small amount of lipidic material present in the acetonitrile acts as a ''keeper'', i.e. reduces or avoids the loss of lipophilic analytes by solubilizing them during the concentration process. In this way, all the components present in the acetonitrile remain in the extract and a different solvent is used for the next cleanup step.

In a classical method, acetonitrile is diluted with saline solution until its concentration in the solution is about 15% and pesticide residues are extracted from the aqueous acetonitrile solution into a low-boiling solvent by multiple partitioning steps. This type of partitioning is termed ''back'' extraction. For low polarity lipophilic compounds, such as OC and PYR pesticides, low polarity petroleum ether is appropriate, because it removes less coextractives from the aqueous phase.[8,9] For more general multiclass methods, DCM is more suitable. The petroleum ether solution can be easily concentrated and subjected to further cleanup steps after solvent exchange, if necessary. In this way, some more polar components present in the acetonitrile extract could be removed by the aqueous phase.

In a third method, acetonitrile is diluted with water until its concentration is about 30% in water and the solution is passed through a reverse-phase cartridge (usually C_{18}) that retains lipophilic pesticides (and other compounds). After drying the cartridge, OC and PYR pesticides can be recovered by eluting with a low-polarity solvent mixture such as toluene–*n*-hexane (3 : 97 v/v), which is suitable to be transferred on-line to an adsorption chromatography cartridge.[51]

As in the case of nonfatty samples, a separatory funnel can be avoided in the *n*-hexane–acetonitrile partitioning step by absorbing the *n*-hexane solution into a macroporous diatomaceous material.[116–118] This acts as a support for the partition process, which is carried out by eluting the bed with acetonitrile saturated with *n*-hexane. Compared with the separatory funnel, the solid matrix partition process is simple and rapid and uses small volumes of solvent (typically 1 g of fat requires 15 mL of acetonitrile and 4 mL of methanol for azeotropic distillation). It is also able to satisfactorily recover from relatively nonpolar compounds such as OC and PYR to polar OP pesticides, while HCB shows a slightly lower recovery (ca. 60%) than the separatory funnel operation.[116] In these applications, a C_{18} cartridge is used downstream to the partition cartridge for in-line further removal of fatty matrix.[116–118] The recovery of 15 OC pesticides from 1 g of soya oil spiked at different levels is reported in Table 4.

In other methods the partition process is carried out by dispersing the extracted fat or fat solution over a chromatographic material that is used to retain the fat and from which OC pesticides are recovered by eluting with acetonitrile or water–acetonitrile mixtures. Alumina eluted with water–acetonitrile (20 : 80 v/v),[119] Florisil eluted with water–acetonitrile (10 : 90 v/v)[48] (both followed by back extraction of pesticide residues into petroleum ether), Celite eluted with DMSO,[40] C_{18} eluted with acetonitrile,[62] or silica gel over PS–DVB eluted with acetonitrile[29] have been found to be useful for this purpose.

4.2 Size-exclusion Chromatography

SEC (often referred to as gel permeation chromatography) is one of the most commonly used cleanup steps to separate pesticide residues from high molecular weight coextractives and fats. The most popular column material in SEC is a cross-linked styrene divinylbenzene (DVB) resin which is eluted with solvent mixtures such as toluene–ethyl acetate or cyclohexane–ethyl acetate

Table 4 Recovery values of 15 OC pesticides extracted from ca. 1 g soya oil spiked at different levels and analyzed after Extrelut-3/C_{18} cleanup followed by Florisil adsorption chromatography, according to the method in Di Muccio et al.[116]

Compound	Spiking level 1 ($mg\,kg^{-1}$)	Recovery (%) ($n = 6$)		Spiking level 2 ($mg\,kg^{-1}$)	Recovery (%) ($n = 6$)		Spiking level 3 ($mg\,kg^{-1}$)	Recovery (%) ($n = 4$)	
		Mean	SD		Mean	SD		Mean	SD
1. HCB	0.021	58	8	0.040	60	5	0.110	62	7
2. α-HCB	0.021	90	8	0.040	90	5	0.110	89	8
3. β-HCH	0.021	89	7	0.040	90	5	0.110	80	8
4. γ-HCH	0.021	88	7	0.040	91	8	0.110	85	7
5. Dieldrin	0.021	85	12	0.040	90	8	0.111	88	8
6. Endrin	0.022	65	12	0.042	78	7	0.116	70	7
7. *cis*-Chlordane (α)	0.020	90	5	0.038	91	7	0.104	85	8
8. *trans*-Chlordane (γ)	0.018	90	8	0.035	90	8	0.095	80	8
9. Oxychlordane	0.023	90	10	0.045	89	10	0.123	88	9
10. *trans*-Nonachlor	0.021	84	6	0.039	85	5	0.108	87	7
11. *cis*-HEPO (β)	0.020	90	7	0.038	87	7	0.106	86	8
12. *p,p′*-DDE	0.020	91	8	0.039	82	8	0.107	80	5
13. *p,p′*-TDE	0.021	90	7	0.040	83	10	0.111	77	5
14. *o,p′*-DDT	0.021	85	5	0.040	88	5	0.111	92	5
15. *p,p′*-DDT	0.021	82	10	0.040	86	10	0.109	95	8

Unpublished data from the laboratory of the author.

for vegetable extracts, and cyclohexane–DCM or *n*-hexane–DCM for fats.

The predominant mechanism of separation relates to the difference in the molecular size between many of the coextractives and the pesticide residues, which generally have molecular weights lower than 500 amu. The resin also exhibits an adsorption mechanism which is mostly masked by the polar solvent used in the mixture. The larger compounds such as fats, oils, and the major part of vegetable coextractives elute earlier (''dump'' fraction) than the pesticide residues (''collect'' fraction). One of the major features of SEC is that it is a nondiscriminative cleanup very useful in multiclass multiresidue methods. Large columns (ca. 45 cm × 2.5 cm ID) have been used that are eluted at $5\,mL\,min^{-1}$ and require about 350–400 mL per sample (''dump'' fraction + ''collect'' fraction + column wash). The collect fraction is about 100–150 mL.[8,9,26,120,123] Large columns (300 mm × 19 mm ID) eluted with DCM at $5\,mL\,min^{-1}$ and with ''collect'' fraction of 20 mL are used in the cleanup of OC and PYR pesticides by high-performance size-exclusion chromatography (HPSEC).[28] The recovery of 15 OC pesticides from 500 mg of spiked soya oil through HPSEC cleanup is reported in Table 5. Also, smaller columns (10 mm ID) for low-pressure chromatography[33] or 7.8 mm ID for HPSEC have been used both for OC and PYR pesticide cleanup and allow a substantial 10-fold saving of time and solvent.[65,124] The column is also reusable, and SEC lends itself to automation. Apart from requiring expensive instruments, SEC is a sequential cleanup procedure, limiting the number of samples that can be processed in a given period of time.

Table 5 Recovery values of 15 OC pesticides from 0.5 g of soya oil spiked at different levels analyzed after HPSEC[a] macro cleanup followed by Florisil adsorption chromatography;[b] spiking levels $0.05\,mg\,kg^{-1}$ for each compound

Compound	Recovery (%) ($n = 7$)	
	Mean	SD
HCB	99.3	5.47
α-HCH	89.3	3.85
β-HCH	92.1	3.00
γ-HCH	90.2	2.93
cis-Chlordane (α)	108.2	5.11
trans-Chlordane (γ)	103.1	1.48
Oxychlordane	105.8	1.58
trans-Nonachlor	104.5	1.50
p,p′-DDE	100.6	7.71
p,p′-DDD	103.3	2.14
o,p′-DDT	102.8	2.55
p,p′-DDT	103.5	3.83
cis-HEPO (β)	102.3	4.52
Dieldrin	96.0	3.91
Endrin	80.2	5.61

[a] Conditions: Phenogel 10–100 Å, 300 mm × 21.2 mm ID eluted by DCM at $5\,mL\,min^{-1}$.
[b] According to the method in Suzuki et al.[52] Unpublished data from the laboratory of the author.

After SEC cleanup (and a concentration step), the nonfatty sample extracts are normally suitable for GC analysis using a selective detector such as an electron capture detector (ECD), a thermionic nitrogen–phosphorus detector (NPD), and/or a flame photometric detector (FPD) without further cleanup. However, for the

determination of OC pesticides by GC/ECD at the lowest levels a further cleanup by adsorption chromatography is often necessary.[121]

SEC offers the opportunity of diverting the ''collect'' fraction directly to a cartridge of chromatographic material for further cleanup on-line. For instance, when low polarity solvent mixtures are used in the SEC of fats, the collect fraction can be purified by adsorption chromatography by passing it through a silica gel cartridge.[125] This avoids the usual sequence of operations, such as collection of the fraction, concentration, solvent exchange, and quantitative transfer to the adsorption column. In a similar application for vegetables, the collect fraction can be passed through a cartridge of active charcoal for on-line removal of pigments.[126] The removal of high molecular weight components by SEC, that are not necessarily removed by adsorption chromatography cleanup, is beneficial to the subsequent GC determination in that it avoids contamination of the injection port and column. Furthermore, micro SEC can be coupled on-line to a gas chromatograph by means of valve switching technology and using a gas chromatograph equipped with a large volume injection system.[127]

4.3 Sweep Codistillation – Assisted Distillation

Sweep codistillation is a technique of cleanup based on thermal degradation of the coextractives including lipidic material and distillation/entrainment of the pesticide residues. In classical operation, the solvent solution of the raw extract (maximum 1 g of fat) is injected into a hot tube filled with glass beads in a stream of nitrogen. Most of the coextractives and the lipidic material adhere to the filling material. Further injections of solvent into the tube elute the pesticide residues via distillation with the hot stream of the nitrogen (sweep codistillation). In more recent developments, the fat is not degraded, but rather deposited in a special void tube held at high temperature. The fat spreads as a thin film over the inner walls of the tube, which is swept by nitrogen, and no injections of solvent are necessary (assisted distillation).[128–130]

To recover the pesticide residues, the effluent is passed through a cold trap and then transferred off-line for a further cleanup. In commercially available systems designed for OC pesticides, the effluent is passed on-line through a cartridge of Florisil for further adsorption cleanup prior to the GC/ECD analysis.

The technique is mainly suited to OC and other fairly stable compounds due to the high temperature involved. Temperatures are in the range of 200–250 °C and typical nitrogen flow is in the range of 200–250 mL min^{-1}. Apparatus is commercially available for processing several samples in parallel.

4.4 Column Chromatography

Adsorption chromatographic cleanup separates the coextractives on the basis of their polarity. Thus, it is appropriate for separation of low polarity compounds such as OC pesticides from the more polar coextractives present in the extracts. Adsorption chromatography under gravity alone was the cleanup of choice in the early days of pesticide residue analysis and it is still widely used. When instruments were not so sensitive as today and large analytical samples were extracted, a large amount of coextractives had to be removed. Consequently columns consisted of 10–20 g of adsorption material which required two to three fractions of 100–200 mL solvent each.

Now, instrumental sensitivity satisfactorily allows the use of adsorption cleanup on smaller columns (1–5 g of adsorbent eluted with fractions of 10–50 mL) as the sole cleanup when a small sample aliquot is extracted, or as a supplemental cleanup after the large-capacity cleanup steps already described.

Column chromatography as cleanup can also be used to collect a single fraction containing a wide range of pesticides with appropriate solvent(s). This approach is mostly useful in multiresidue methods to analyze for as many pesticides as possible. Of course, this approach requires sufficient resolving power in the identification step. However, the fractionation of the extract simplifies analyte identification and reduces the number of potential interferences in the collected fractions. Furthermore, the appearance of tentatively identified analytes in the expected fraction adds further evidence to the analyte confirmation. However, fractionation of the extract increases the number of identification processes. With adsorbents such as Florisil, alumina, and silica gel the reliability of the fractionation mostly relies on controlling and maintaining the right degree of activation/deactivation.[9,45,131] For this reason, adsorption columns and cartridges are used for just a single sample and are not reused.

In the case of OC and PYR pesticides, the most frequently used adsorbents are Florisil, alumina, and silica gel. Depending on the particular application, these sorbents can be used either ''activated'', i.e. heated at a temperature high enough to release most of the moisture, or ''deactivated'' with addition of 1–10% (w/w) water. The more active the material, the more coextractives and fat that are retained. Partial deactivation allows the use of a less polar eluent, while maintaining sufficient retention of the fat. OC pesticides are eluted with hydrocarbon solvents or low polarity mixtures, but more polar eluting mixtures are required to recover oxygenated OC pesticides, such as dieldrin, endrin, heptachloroepoxide, and oxychlordane, while PYR pesticides require significantly more polar

eluents. However, more polar eluents generally lead to "dirtier" fractions. Apolar fractions are almost free of lipidic material, while a few milligrams may appear in the polar ones. Water deactivation allows the use of less polar eluents. Hence 10 g of 5% water-deactivated Florisil, eluted with 200 mL of diethyl ether–petroleum ether (6 : 94 v/v), has been used to clean up PYR extracted from wheat and vegetables.[37]

With Florisil macrocolumns (ca. 20 g) two eluting schemes are well known and are continuously updated with a number of new pesticides. In the first scheme, three fractions are collected eluting with 6%, 15%, and 50% ethyl ether in petroleum ether, while in the second scheme three fractions, consisting of 20% DCM in *n*-hexane, 50% DCM + 0.35% acetonitrile in hexane, and 50% DCM + 1.5% acetonitrile in *n*-hexane, are used.[8,9]

Examples of Florisil minicolumns include: (1) a 2.5-g column eluted with *n*-hexane–benzene–ethyl acetate (180 : 19 : 1 by vol.) for OC pesticides,[52] and with *n*-hexane–benzene–ethyl acetate (176 : 19 : 5 by vol.) for PYR pesticides;[66] (2) a 4-g column eluted with one fraction of 50% DCM + 1.5% acetonitrile in *n*-hexane for OC and PYR;[9] (3) a 5-g column of 10% water-deactivated Florisil eluted with diethylether–*n*-hexane (4 : 96 v/v) for OC.[132] The fractionation of 28 OC pesticides and 20 PCBs from a 2.5-g Florisil column eluted with a three-eluent scheme is reported in Table 6. Application of Florisil cleanup to PYR pesticides also include: (1) a 10-g column eluted with 100 mL of ethyl acetate–*n*-hexane (3 : 7 v/v);[133] (2) a 6-g column eluted with 6 mL of *n*-hexane–diethyl ether (7 : 3 v/v);[134] (3) a 0.5-g minicolumn eluted with 5 mL of toluene;[30] (4) a minicolumn of 0.5 g of Florisil + 0.04 g of active charcoal eluted with 5 mL of toluene.[31] A minicolumn of 4 g of Florisil + 1 g of active charcoal eluted with 40 mL of toluene–acetonitrile (98 : 1 v/v) has been used for the cleanup of extracts from vegetables.[32] A similar, smaller column (1.6 g Celite + 0.4 g charcoal eluted with 50 mL of acetonitrile–toulene (3 : 1 v/v)) is the cleanup used in a multiclass, multiresidue method for fruits and vegetables.[20]

Alumina in large columns (15 g of acidic alumina eluted with 160 mL of DCM) has been used as a very common cleanup step before the GC/ECD determination of OC and PYR pesticides in grains.[38] A minicolumn of 1 g of 10% water-deactivated alumina has been used for the separation of OC pesticides from 70–100 mg of fat.[135] However, the control of the deactivation process is done by recovery trials of critical compounds and, due to poor batch-to-batch reproducibility, the deactivation has to be repeated with new batches.[76]

Popular multiresidue schemes with silica gel include: (1) a 5-g column eluted with five fractions of *n*-hexane, *n*-hexane–benzene (4 : 6 v/v), benzene, benzene–ethyl acetate (1 : 1 v/v), and ethyl acetate to elute a diversity of

Table 6 Fractionation of OC pesticides and PCB from a column of 2.5 g activated Florisil eluted with a three-solvent elution scheme in the presence of 50 mg soya oil

Compound	Mass loaded (μg)	Recovery (%)		
		1st fraction	2nd fraction	3rd fraction
PCB-TriCl (31)	0.05	110		
PCB-TriCl (28)	0.05	102		
PCB-TetraCl (66)	0.05	94		
PCB-TetraCl (52)	0.05	98		
PCB-PentaCl (101)	0.05	99		
PCB-PentaCl (104)	0.05	117		
PCB-PentaCl (110)	0.05	91		
PCB-PentaCl (118)	0.05	102		
PCB-PentaCl (128)	0.05	89		
PCB-HexaCl (132)	0.05	93		
PCB-HexaCl (138)	0.05	105		
PCB-HexaCl (151)	0.05	99		
PCB-HexaCl (153)	0.05	94		
PCB-HexaCl (156)	0.05	93		
PCB-HeptaCl (170)	0.05	90		
PCB-HeptaCl (180)	0.05	100		
PCB-HeptaCl (189)	0.05	97		
PCB-OctaCl (194)	0.05	97		
PCB-NonaCl (206)	0.05	98		
PCB-DecaCl (209)	0.05	100		
Octachlorostyrene	0.09	94		
HCB	0.05	79	10	
α-HCH	0.05		100	
β-HCH	0.10		93	9
γ-HCH	0.05		104	
δ-HCH	0.05		103	
cis-Chlordane (α)	0.10		109	
trans-Chlordane (γ)	0.10		113	
Oxychlordane	0.10		107	
cis-Chlordene (α)	0.10	91	3	
trans-Chlordene (γ)	0.10	90	5	
trans-Nonachlor	0.07	108	9	
o,p′-DDE	0.10	82	15	
p,p′-DDE	0.10	112	5	
o,p′-DDD	0.15		112	
p,p′-DDD	0.10		110	
o,p′-DDT	0.15	66	35	
p,p′-DDT	0.15		114	
Heptachlor	0.05	60	22	
cis-HEPO (β)	0.05			92
α-Endosulfan	0.05			89
β-Endosulfan	0.07	NR	NR	NR
Endosulfan sulfate	0.18	NR	NR	NR
Dieldrin	0.10			102
Endrin	0.10			75
Aldrin	0.05	117		
Mirex	0.16	119		
Quintozene	0.05		29	48
Methoxychlor	0.35			103

1st fraction: (1 mL sample application + 3 × 0.5 mL washings + 30 mL) *n*-hexane.
2nd fraction: 25 mL *n*-hexane–benzene (80 : 20 v/v).
3rd fraction: 30 mL *n*-hexane–benzene–ethyl acetate (180 : 19 : 1 by vol.).
NR, not recovered.
Unpublished data from the laboratory of the author.

pesticides;[23] (2) a 1-g column of 1.5% water-deactivated silica eluted with five fractions of increasing polarity (*n*-hexane–toluene (65:35 v/v), toluene, toluene–acetone (95:5 v/v), toluene–acetone (80:20 v/v), and acetone) for over 400 compounds including OC and PYR;[10,26] and (3) a 1-g column eluted with four fractions of *n*-hexane, benzene–*n*-hexane (60:40 v/v), benzene, and ethyl acetate–benzene (50:50 v/v), to separate 50 OC pesticides and PCB.[136] An application for PYR pesticide residues from fruits and vegetables is based on a 5-g silica column eluted with 25 mL of *n*-hexane–DCM (4:1 v/v) and 60 mL of DCM (contains the analytes).[137] An application of silica gel to OC, OP, and triazines in wine is based on a 15-g column eluted with 40 mL of *n*-hexane, 16 mL *n*-hexane–benzene (1:1 v/v) and 50 mL of acetone.[74] Sequential application of Florisil cleanup and silica fractionation has been used for PCBs in human serum.[44]

4.5 Solid Phase Extraction Chromatography

Currently, ready-to-use cartridges filled with 0.1–1 g of classical adsorbents (alumina, silica, Florisil, carbon), the so-called SPE cartridges, are increasingly used to miniaturize the cleanup; this saves solvents and time, and allows automation of the cleanup step.

C_{18} cartridges are used to reduce or retain lipophilic coextractives when acetonitrile or aqueous acetonitrile extracts from either fatty[46,49,51,116,117] or nonfatty samples[21,110] are passed through. It has been noted that most pesticides elute off a C_{18} cartridge with an eluent composition of acetonitrile–water (75:25 v/v) (which is about the composition of the aqueous acetonitrile from the extraction of vegetable samples), while lipids and other highly lipophilic coextractives are retained.[21] In a different way, aqueous acetone,[29] or aqueous acetonitrile,[110] resulting from the extraction of vegetables, can be further diluted with water and passed through a lipophilic C_{18}[110] or PS–DVB[29] cartridge to retain the analytes. After retention, C_{18} and PS–DVB cartridges are dried with a nitrogen stream and the analytes are recovered with solvents such as ethyl acetate[29] or toluene–petroleum ether (3:97 v/v).[49,51] The latter, being of low polarity, can be passed directly to a Florisil cartridge for subsequent adsorption cleanup,[49,51] avoiding the usual sequence of linking steps (collection, concentration, solvent exchange, quantitative transfer).

Florisil is used to remove polar coextractives in the analysis of both fatty[49,51] and nonfatty[29] sample extracts. In an application of Florisil SPE cartridges to OC in fish, crab, and milk the eluting mixture was diethyl ether–petroleum ether (2:98 v/v),[49,51] while it was acetone–*n*-hexane (15:85 v/v) for PYR pesticide residues in paddy rice.[138] Alumina SPE has also been used for adsorption cleanup of fatty extracts.[118]

SPE-NH_2 has also been used in normal phase cleanup in multiresidue methods for nonfatty[15,21] and milk samples.[46] It is usually eluted with DCM–methanol (97:3 v/v)[46] or DCM–methanol (99:1 v/v).[15]

In a multiresidue method based on acetonitrile extraction, after separating the water with the "salting out" procedure, a portion of the acetonitrile extract is cleaned up with a minicolumn of 2 g of charcoal–Celite (1:4 w/w) or on a carbon SPE cartridge of 500 mg. The pigments are mostly retained while the pesticides are eluted with acetonitrile– toluene (3:1 v/v).[20] It is noteworthy that HCB is poorly or not recovered through carbon, depending on the type of material. Such a cartridge has been used also downstream from an SEC column eluted with cyclohexane–DCM (1:1 v/v) to trap pesticides for subsequent SPE cleanup.[126] Another way of removing pigments from fruits and vegetable extracts is the use of an SAX (strong anion exchanger) or an SAX/PSA (propylethylenediamine – silica): the extract is applied in DCM and eluted with DCM.[54]

The great advantage of cleanup on an SPE cartridge is that it is amenable to robotic manipulation for unattended processing of large sequences of extracts. Also SPE cartridges are ready-made, disposable, and reduce solvent consumption, time, and labor.

4.6 High-performance Liquid Chromatography

Semipreparative HPLC on a column of 250 mm × 9.2 mm ID filled with porous spherical silica particles and eluted with 20% DCM in *n*-hexane has been used to separate five OC pesticides and PCB from 300 mg of butter fat.[139] By using column switching technology, fully automated sample cleanup and fractionation of 10 OC pesticides and PCB from 45 mg of fat was achieved by normal phase chromatography using a short column and an analytical column of silica gel.[45] A rapid automated off-line cleanup on a silica gel HPLC column has been reported for the cleanup of 45 mg of fat in 32 min per sample.[140] Advantages include high separation power and reproducibility, and reusability of the column.

An automated sample cleanup and fractionation has been used for the analysis of mussel extract using normal phase high-performance liquid chromatography (NPHPLC) with a column of silica (150 mm × 3.9 mm ID) eluted stepwise with 4 mL each of *n*-hexane, *n*-hexane–ethyl acetate (99:1 v/v), and ethyl acetate for the separation and collection of compounds of different polarity.[141] NPHPLC on aminopropyl silica has been used after lipid separation by SEC for the cleanup and group separation of PCB and OC pesticides in cod liver oil. The column was eluted with two fractions: the first (10 mL of *n*-hexane) elutes the nonpolar PCB,

HCB, DDE, and mirex, and the second (32 mL of *n*-hexane–DCM (90 : 10 v/v)) elutes the more polar HCH, DDD, DDT, and chlordane.[142]

4.7 Thin-layer Chromatography

Due to its high separation capabilities, TLC can also be used as a mini preparative, highly selective cleanup. Crude or partially cleaned up extracts equivalent to 2–10 g of product (acetone extraction and partition into DCM) can be applied bandwise on a standard 20 mm × 20 mm × 0.25 mm silica gel plate. After the development, compounds of interest can be recovered from the plate by scraping the band at the appropriate retention factor (R_f) (identified by running the target compound(s) alongside at higher concentration) and eluting the silica material with a suitable solvent. Compound(s) recovered are highly cleaned up and are suitable for identification/determinative techniques. The agreement between the R_f and the response from the identification techniques adds further evidence for confirmation of the tentatively identified compounds at residue level. As a cleanup step, TLC has high separation power but low capacity and, therefore, complements a high-capacity, low resolving power cleanup, such as SEC or partition processes.[8,143] TLC is inconvenient but can increase throughput due to parallel operation.

4.8 Degradative Cleanup

Degradative cleanup for samples of plant and animal origin can be accomplished by chemical treatment which converts many coextractives into chemical species that are no longer extractable, or permits their easy removal with a selective process.

Due to their relatively high chemical stability, strong degradative cleanup has often been applied to extracts for the analysis of OC pesticides. Relatively mild degradative cleanup is conceivable for PYR and other classes of pesticides.

Concentrated sulfuric acid treatment is a classical method to destroy organic matter in raw extracts. It has been applied either by shaking a solution of the extract with sulfuric acid,[144,145] or by passing the solution through a column of inert diatomaceous material.[146] The recoveries of 22 OC compounds through a sulfuric acid impregnated Extrelut cartridge are reported in Table 7. A small amount of silica gel impregnated with concentrated sulfuric acid has been used as the bottom layer in a bilayer MSPD column for the extraction of PCBs in fish.[147]

Alkaline treatments have been used to saponify and remove the fatty material. They can be given either by passing the solvent solution through a column of MgO mixed with diatomaceous material,[8] or by treatment with 10% KOH in ethanol.[43,145] In the first case the solvent solution is directly amenable to the next determinative step, while in the second case an extraction into an organic solvent is carried out. An example is the saponification applied to the analysis of PCBs in fats and oils.

Table 7 Behavior of 22 OC compounds through an Extrelut cartridge charged with 1 mL of either 95% or 90% sulfuric acid, eluted with 10 mL of light petroleum, according to the method in Di Muccio et al.[146]

Compounds	Amount (μg)	Average recovery ± SD ($n = 6$) (%)	
		95% H_2SO_4	90% H_2SO_4
HCB	0.01	74 ± 17	86 ± 8
α-HCH	0.01	88 ± 15	95 ± 4
β-HCH	0.02	85 ± 15	90 ± 10
γ-HCH	0.01	88 ± 9	95 ± 10
δ-HCH	0.01	86 ± 10	110 ± 15
cis-Chlordane (α)	0.03	93 ± 7	85 ± 10
trans-Chlordane (γ)	0.03	91 ± 9	87 ± 8
Oxychlordane	0.02	91 ± 10	–[b]
cis-Chlordene (α)	0.02	70 ± 12	–[b]
trans-Chlordene (γ)	0.02	63 ± 15	–[b]
trans-Nonachlor	0.02	79 ± 10	–[b]
p,p′-DDE	0.03	95 ± 8	91 ± 11
o,p′-DDD	0.05	90 ± 7	90 ± 10
p,p′-DDD	0.05	92 ± 6	89 ± 9
o,p′-DDT	0.05	85 ± 8	85 ± 10
p,p′-DDT	0.05	83 ± 10	86 ± 10
Dieldrin	0.05	NR	NR
Endrin	0.05	–[a]	–[a]
Aldrin	0.01	87 ± 12	97 ± 15
cis-HEPO (β)	0.01	88 ± 10	92 ± 15
Methoxychlor	0.15	21 ± 20	45 ± 15
Mirex	0.11	90 ± 6	92 ± 10

[a] Recovered as endrin ketone, which appears as a peak with a retention time longer than that of the parent compound.
[b] Not investigated.
NR, not recovered.
Unpublished data from the laboratory of the author.

Both sulfuric acid and alkaline treatments have been used either as the sole cleanup for the analysis,[42,43] or as a supplemental cleanup after some of the main cleanup steps described above.

Degradative cleanup, however, can introduce some transformation of the analyte(s) of interest. Most OC pesticides including PCNB and PCB are stable under acidic conditions except dieldrin, which is destroyed; endrin, which is converted into endrin ketone; and methoxychlor, which is partially destroyed.[146]

HCB, cyclodienic pesticides (heptachlor, heptachlor-epoxide, aldrin, dieldrin, endrin, and mirex) and DDE are stable under alkaline conditions at ambient temperature, while PCNB is slowly degraded, and DDT, DDD, and methoxychlor are dehydrochlorinated to the corresponding olefins.[145] Dicofol is fairly stable at ambient temperature,[145] but it is converted to DCBP in

strong alkaline conditions.[5] HCHs are converted into trichlorobenzenes.[148] As esters, PYR pesticides are degraded under alkaline conditions.

Silica gel, alumina, or Florisil impregnated with silver nitrate have been used for the removal of some sulfur-containing compounds from extracts of onions and kale.[149] By the addition of $NaBH_4$ prior to sample extraction with acetonitrile, nearly all matrix interferences (organosulfur compounds) can be eliminated from green onion samples.[21]

Oxidative cleanup with $KMnO_4$ solution has been used in the analysis of vegetables.[150]

The alteration of some analyte(s) may be used to remove or separate them from other analytes, for instance, the oxidation of DDE to DCBP and its separation by adsorption chromatography can eliminate this interference in the analysis of PCB.[6] Also this approach can be used to convert the analytes into species with better chromatographic, sensitivity, and selectivity properties. Derivatization of the analytes into species whose detection is a confirmation of identity is also sometimes necessary. For instance, detection of endrin ketone after acid treatment is diagnostic for endrin, and detection of a specific late-eluting peak after reaction with a hydrobromic acid–acetic anhydride mixture is diagnostic for dieldrin, whose epoxide ring is opened under these conditions.[151] HCB can be converted into later-eluting monoisopropoxypentachlorobenzene and diisopropoxytetrachlorobenzene for confirmation of identity.[152] The formation of compounds with longer retention times in GC is beneficial because usually most of the matrix interference appears in the early portion of the chromatogram.

As stated above for dicofol,[63] degradation of analytes during the analytical process is also possible.

In the analysis of PCB and OC pesticides in environmental samples, especially soil and sediments, it is necessary to remove elemental sulfur to obtain a good GC/ECD chromatogram. The removal is carried out by activated copper powder,[71] with tetrabutylammonium sulfite, or with SEC.[7]

5 IDENTIFICATION AND DETERMINATION

5.1 Gas Chromatography

OC and PYR, as well as other classes of pesticides which have relatively low molecular weight with sufficient volatility, are most commonly analyzed by GC. A great number of developments and improvements in GC and related detectors has been made in connection with the analysis of pesticide residues in food and environment. The relative simplicity and ease of operation, together with the high separation power and an array of sensitive and selective detectors, make GC the technique of choice for determination of thermally stable pesticides.

5.1.1 Packed Columns

The first GC applications involving pesticide analysis were developed on packed columns. Early columns were generally glass tubes 1–2 m long, 2–4 mm ID, filled with a solid, high surface material covered with 1–5% (w/w) of thermally stable, low volatility polymers. A variety of stationary phases including methyl silicone (DC-200, OV-1, OV-101, SE-30), methyl phenyl silicone (OV-17, DC-710), fluoropropyl silicone (QF-1, OV-210), cyanopropyl silicone (XE-60, OV-225), polyethers (Carbowax 20 M), and polyesters (DEGS, NPGS) allowed a diversification of selectivity, which can be further expanded by packing the columns with mixed packings, such as 5% DC-200 + 7.5% QF-1, 1.5% OV-17 + 1.95% QF-1, 1.5% SP-2250 + 1.95% SP-2401, 5% OV-101 + 3% OV-225, and 5% OV-61 + 3% XE-60 + 5% QF-1.[8,9,11,151–153] The chemical composition of the main stationary phases and trade codes are reported in Table 8.

Although packed columns have almost been entirely superseded by capillary columns, they could have useful application in the screening of a few target analytes in dirty extracts. Indeed, the injection of coextractives into the first, hot portion of the packing material provides an extra cleanup by irreversible adsorption or thermal degradation. However, thermally labile compounds suffer from the activity of the packing material.

5.1.2 Capillary Columns

Nowadays, pesticide residue analysis is almost always carried out on fused silica capillary columns. In this approach, columns from 10 to 60 m with IDs of 0.18–0.53 mm (0.18 mm, ''microbore''; 0.25 mm, ''narrow-bore''; 0.32 mm, ''wide-bore''; 0.53 mm, ''megabore'') are commercially available. Smaller diameter columns have greater separation capabilities, but also have a smaller sample capacity and limited lifespan. The liquid stationary phase is normally chemically bonded to the inner surface of the columns, and film thicknesses used in pesticide residue analysis are typically in the range 0.25–1.5 μm. The special preparation of the inner surface and the coverage of the film give a high degree of inertness that allows the analysis of relatively labile and/or polar compounds. A variety of polymers of similar nature to the packed columns are available for capillary columns, and even specially prepared columns are available to suit specific separations, such as the ''608'' series columns for OC pesticides listed in EPA Method 608.[154]

Table 8 Chemical composition and trade codes of the most common GC stationary phases

Chemical composition	Trade code
100% Dimethyl polysiloxane	007-1, CP-Sil 5CB, DB-1, DB-lht, HP-1, HP-101, OV-1, RSL-150, RSL-160, Rtx-1, SE-30, SPB-1, SPB-Sulfur, ULTRA-1, SP-2100, BP-1, DC-200, AT-1, ATS-1, EC-1
(5% Diphenyl) 95% dimethyl polysiloxane	007-2, CP-Sil 8CB, DB-5, DB-5.625, DB-5ht, DB-5ms, HP-5, HP-5ms, OV-5, PTE-5, PTE-5QTM, PAS-5, RSL-200, Rb-5, Rtx-5ms, SAC-5, SE-54, SPB-5, ULTRA-2, XTI-5, SE-52, BP-5, AT-5, ATS-5, EC-5
(20% Diphenyl) 80% dimethyl polysiloxane	007-7, Rtx-20, SPB-20, VOCOL, AT-20, EC-20
(35% Diphenyl) 65% dimethyl polysiloxane	007-11, DB-35, DB-35ms, HP-35, RSL-300, Rtx-35, SPB-35, SPB-608, OV-11, AT-35
(50% Diphenyl) 50% dimethyl polysiloxane	007-17, DB-17, DB-17ht, HP-17, HP-50+, OV-17, Rtx-50, SP-50, SP-2250, SPB-50, AT-50
(6% Cyanopropylphenyl) 94% dimethyl polysiloxane	007-502, DB-624, DB-1301, HP-1301, HP-624, Rtx-1301, Rtx-624, AT-624, AT-1301
(14% Cyanopropylphenyl) 86% dimethyl polysiloxane	007-1701, CP-Sil 19CB, DB-1701, HP-1701, OV-1701, PAS-1701, Rtx-1701, SPB-1701, AT-1701
Trifluoropropylmethyl polysiloxane	DB-210, RSL-400, Rtx-200, OV-202, OV-210, OV-215, QF-1, SP-2401, AT-210
(50% Cyanopropylphenyl) 50% phenylmethyl polysiloxane	007-225, CP-Sil43CB, DB-225, HP-225, OV-225, RSL-500, Rtx-225, AT-225
(50% Cyanopropyl) 50% methyl polysiloxane	DB-23, AT-SILAR
Biscyanopropyl, cyanopropylphenyl polysiloxane	CP-Sil 84, Rtx-2330, Rt-2330, SP-2330, SP-2331, SP-2560, SP-2380, AT-SILAR
Poly(ethylene glycol)	007-CW, Carbowax 20M, CP-Wax 52CB, DB-WAX, HP-20M, HP-Wax, lnnowax, Omegawax, Stabilwax, AT-WAX, EC-WAX, SUPELCOWAX-10, SUPEROX II, BP-20, ATS-WAX
Poly(ethylene glycol) modified with acids	007-FFAP, CP-Wax 58CB, DB-FFAP, HP-FFAP, Nukol, OV-351, SP-1000, Stabilwax-DA, SUPEROX-FA, BP-21, Stabilwax-DB, CAM, Carbowax Amine, AT-1000, EC-1000

The type and dimension of the column depend on the type of analysis and analyte/matrix combination. In routine pesticide analysis of fruits and vegetables, megabore columns of 15 m × 0.53 mm ID × 0.8–1.5 μm film thickness provide sufficient separation in a short time combined with a good capacity and ease of operation.[30,66] In large, multiclass, multiresidue methods for monitoring pesticide residue in foods, columns of 30 m × 0.25 mm ID × 0.25 μm film thickness are most commonly used.[20,28,54,57,153,155] For highly demanding environmental analysis regarding the analysis of complex mixtures of environmental contaminants such as PCBs and camphechlor (toxaphene), columns of 60 m × 0.18–0.25 mm ID × 0.18–0.25 μm film thickness are used.[44,156] The most widely used stationary phases are the 100% dimethyl polysiloxane (''code-1''), the 5% diphenyl–95% dimethyl polysiloxane (''code-5''), the 35% diphenyl–65% dimethyl polysiloxane (''code-35'' and ''code-608''), the 50% diphenyl–50% dimethyl polysiloxane (''code-50'' and ''code-17''), and the 14% cyanopropylphenyl–86% dimethyl polysiloxane (''code-1701'') phases. Each phase has a particular selectivity and is better suited to particular separations. The most commonly sought OC pesticides are better separated on a ''608'' phase, including the critical pair DDE–dieldrin. The group of early eluting HCB and HCH isomers is better separated on a ''1701'' phase. Some pairs of OC pesticides, including DDE–dieldrin, *cis*-heptachloroepoxide–oxychlordane, *cis*-chlordane–*trans*-nonachlor, *o*,*p*′-DDT–endrin, and β-endosulfan–*p*,*p*′-DDD, are considered critical. A number of critical pairs are found in complex PCB mixtures, for instance, PCB 55/60, PCB 138/178.[44] Analyses on columns of different polarity are usually carried out to resolve critical pairs.

Continuing research is dedicated to the development of new phases to meet the stringent requirements of the separations needed for complex mixtures of OC compounds such as PCB and toxaphene in environmental samples.[7,142,156] Toxaphene consists of several compounds with a different degree of chlorination and most of them are chiral. Enantiomers often have different toxic and degradative properties. Enantiomer separation has been carried out on chiral capillary columns.[157] Tandem columns, i.e. systems made of a combination of a column that gives isomer-selective separation and a column that gives enantiomer-selective separation, can be optimized for the isomer- and enantiomer-selective separation of toxaphene.[158] PYR are currently analyzed by using conventional GC phases where only the separation of the stereoisomers occurs.[30,31] For instance, cypermethrin

(with three asymmetric carbon centers giving eight enantiomers) shows only four peaks corresponding to the four pairs of enantiomers, eluting in the order *cis*, *trans*, *cis*, *trans* from a low-polarity column.[153]

Being chemically bonded to the surface, the stationary phase is more stable and, after being subjected to high-temperature conditioning, releases a low ''bleeding'' even at high temperature, so that detectors can be operated at higher sensitivity and programmed temperature operations are possible and still maintain a flat baseline. Temperature programming and the high resolving power of the capillary columns allow the analysis of a wide variety of analytes in complex mixtures in a single run.

5.1.3 Injectors

While injection into packed columns is quite simple, special injectors and injection techniques have been developed to fully exploit the separation power of capillary columns. Injection can be carried out either in a ''liner'' (an injector chamber held in the injector oven to which the column is connected) or ''on-column''. Injection into a liner is carried out with conventional microsyringes through a polymeric rubber septum. On-column injection into narrow- and wide-bore columns is carried out through special valves and requires tiny needles and special alignment systems. However, ''liner''-type on-column injection is possible with conventional techniques into the first portion of a megabore column held in the injector oven. Autoinjectors also work more simply and reproducibly when injection is made into liners or on megabore columns.

The site of injection, either the liner or the beginning of the column in on-column operation, is located in the injector oven. At the time of injection the temperature can be kept either high enough to vaporize all the components (hot or flash vaporization) or suitably low to allow the vaporization of the solvent only, with the analytes being vaporized on temperature programming (temperature programmable injectors). The injection chamber (the liner) and the column can be connected so that all the sample injected is transferred to the column, or part of the vapor produced upon injection can be ''split'' and vented outside (split injectors). In pesticide residue analysis, as in general in trace analysis, one is interested in maximum sensitivity. Thus, split injectors are operated in ''splitless'' mode. In this mode, the split vent of the injection chamber is closed for about a minute to allow the almost quantitative transfer of the sample into the column, and then the split vent is opened to sweep the residual vapor from the injection chamber.

Regardless of the type of injection, the vapor produced by the usual injection volume (1–3 μL) would flood the column and spread the analytes along a great length of the column unless special operating techniques were used. To refocus the analytes into a narrow band at the beginning of the column, the temperature of the column oven is kept relatively low (cold trapping) while the analytes are transferred from the injection site into the column. Furthermore, with proper selection of the column temperature with respect to the boiling point of the solvent injected, solvent vapors are temporarily condensed, which leads to a temporary increase of the film thickness of the stationary phase. This leads to the enhancement of the retention properties of the column in that zone and refocusing effects on the analytes (solvent effect).

Injection is usually done through a septum made of special silicone polymers. At high temperature, a certain bleeding occurs and possibly leads to ghost peaks in the chromatogram. Thus, septum injectors in trace analysis are normally equipped with a septum purge line, i.e. a small stream of the carrier gas that sweeps the inner face of the septum to vent much of the bleeding.

Usually volumes of 1–3 μL are introduced, but this depends on the type of injector. Hot vaporizing injectors should have liners with sufficient volume to accommodate the vapor, or else contamination of the carrier gas line can occur due to backflushing. Larger volumes give longer vapor residence time in the liner and contact with hot surfaces can enhance degradative effects on the analytes. To decrease the residence time, hot vaporizing injectors are better operated with megabore columns with flow in the range 5–10 mL min^{-1}. To increase the volume available for the expansion of the vapor and to prevent or reduce the injected coextractives from entering the analytical column, a ''retention gap'' is usually inserted between the injector and the column. A retention gap (or guard column) is a length of deactivated fused silica column without stationary phase, and thus without appreciable retention. Retention gaps of 1–5 m are used. About 30–40 cm of the retention gap are removed after a certain number of injections, depending on the type of matrix, extraction, and cleanup. Multiclass, multiresidue methods that use a minimal cleanup require more frequent cuttings.[15,21,25]

Injection liners are made of glass that is often deactivated by silanization. The inertness of the injection chamber can be monitored through the behavior of indicator compounds. For instance, appearance of DDE and DDD upon injection of DDT, and/or endrin aldehyde and endrin ketone upon injection of endrin, together with the shape of the peaks, is an indication of degradative activity. However, activity may be useful in certain cases. For instance, conversion of tralomethrin into deltamethrin in the injection chamber is part of the analytical method for the determination of

tralomethrin. As with packed columns, injection into hot vaporizing injectors introduces some degradation of the coextractives, which can lead to beneficial effects on the life of the column and quality of the chromatograms. Injection into cold injectors is preferable for heat-labile compounds, but often gives rise to more complex chromatograms.

When high sensitivity is desirable, large volumes of solvent can be injected with special techniques. In one approach, the injector can be kept at subambient temperature while repeatedly injecting the solution, or, with the "cryo focusing" technique, a short length of the column can be cooled to focus the analytes. In a special apparatus, the large volume of vapor generated by the injection is vented outside by activating an "early exit" valve during the concurrent evaporation of the solvent in the injector and retention gap.[83] Increasing the portion of the analytical sample injected allows a reduction in size of the aliquot of analysis and the analytical process, which is especially useful for instance when a large volume of water has to be extracted to attain low limits of determination.

Capillary columns are usually operated with helium as the carrier gas (less frequently hydrogen) since the optimum HETP (height equivalent theoretical plate) is maintained over a wide range of linear gas velocity. However, with conventional pressure- or flow-regulated systems, the change in flow rate due to column temperature programming is not taken into account, and the retention times and peak widths of low-volatility compounds increase. Most GC instruments now operate with electronic pressure control (EPC), which is a system able to maintain constant flow over the entire temperature program. This acts to reduce the analysis time and improve separation of compounds eluting at high temperatures, which is the case for some PYR pesticides. Pressure programming offered by EPC has been used to carry out pulsed splitless injection in conjunction with megabore columns. In this way, a high flow rate is used during injection to sweep the analytes out of the liner rapidly, thus reducing analyte loss due to adsorption or thermal degradation. This is useful for DDT/DDE, PYR, and polar OP pesticides. Also, pulsed splitless injection strongly reduces the chromatographic response enhancement caused by the so-called matrix effect.[7,159]

5.1.4 Selective Detectors

The great advantage of GC separation compared to other techniques in trace analysis in general and in pesticide residue analysis in particular is the availability of an array of simple-to-operate, sensitive, and selective detectors, such as the ECD, electroconductivity detector

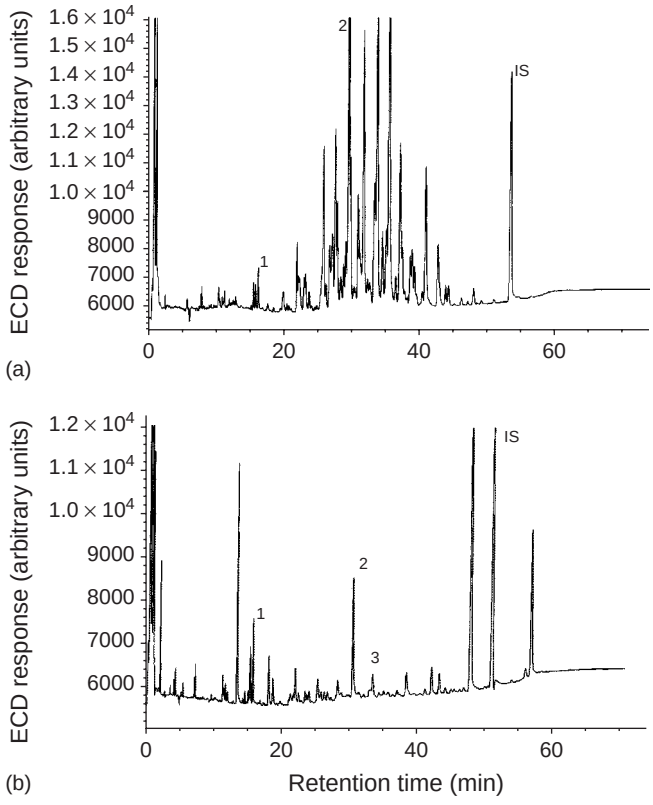

Figure 1 GC/ECD chromatograms of *Scomber scombrus* fish. (a) 11.37 g + IS (internal standard) PCB 209, 0.06 µg, 1 µL injected, 1st fraction (*n*-hexane) of Florisil cleanup: 1, HCB, 0.4 µg kg^{-1}; 2, *p*,*p*′-DDE, 0.6 µg kg^{-1}; other peaks, PCB congeners, 0.28–12.45 µg kg^{-1}. (b) 11.37 g + IS PCB 209, 0.06 µg, 1 µL injected, 2nd fraction (*n*-hexane–benzene, 80 : 20 v/v) of Florisil cleanup: 1, γ-HCH, 0.96 µg kg^{-1}; 2, *p*,*p*′-TDE, 3.0 µg kg^{-1}; 3, *p*,*p*′-DDT, 0.42 µg kg^{-1}. Column: DB-1701, 15 m × 0.53 mm ID × 1.0 µm.

(ElCD), halogen-specific detector (XSD), FPD, NPD, and chemiluminescent detector (CLD).[46,160]

The ECD is sensitive to compounds containing electron acceptor atoms or groups, such as halogen atoms, nitro groups, and unsaturated systems. OC pesticides by definition contain chlorine, while most PYR pesticides contain either chlorine, bromine, or fluorine. OC pesticides are usually analyzed by GC/ECD in a working range of 1–100 pg, while PYR pesticides are less sensitively measured in working ranges of 20–150 pg up to 50–500 pg, depending on the number and type of halogen in the molecules.[116,117] Figure 1 shows typical GC/ECD chromatograms of a fish extract (analyzed with acetone–*n*-hexane extraction, solid-matrix partition for separation of fat and cleanup by Florisil adsorption chromatography). Standard mixtures of OC pesticides are shown for comparison in Figure 2. Typical GC/ECD chromatograms of two soya oil samples spiked with two different standard mixtures of PYR pesticides together

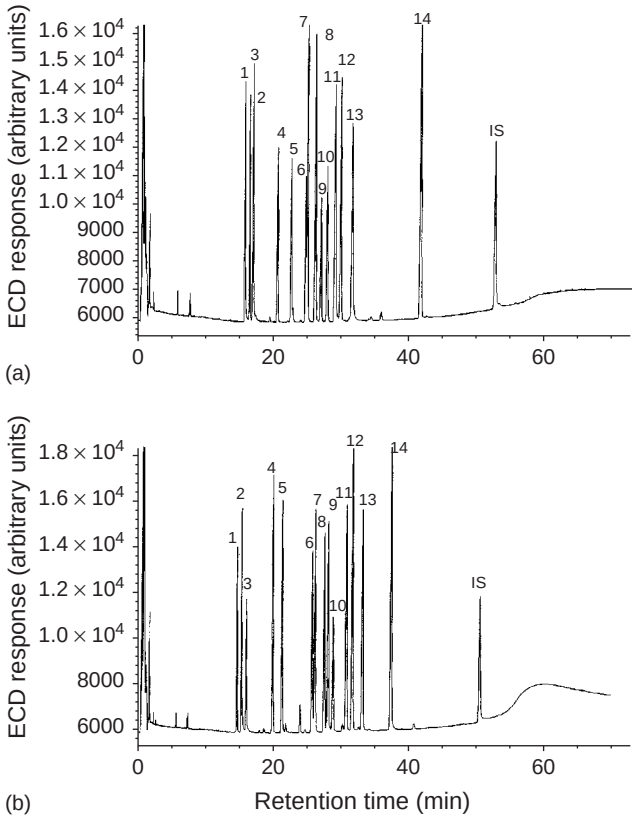

Figure 2 GC/ECD chromatograms of OC standard mixtures 1 (a) and 2 (b). (a) 1, β-HCH, 0.05 ng; 2, γ-HCH, 0.025 ng; 3, quintozene, 0.025 ng; 4, heptachlor, 0.025 ng; 5, aldrin, 0.025 ng; 6, β-HEPO, 0.025 ng; 7, oxychlordane, 0.05 ng; 8, γ-chlordane, 0.05 ng; 9, α-endosulfan, 0.026 ng; 10, *trans*-nonachlor, 0.035 ng; 11, *p,p′*-DDE, 0.05 ng; 12, endrin, 0.05 ng; 13, *p,p′*-TDE, 0.05 ng; 14, mirex, 0.082 ng; IS = PCB 209, 0.03 ng. (b) 1, α-HCH, 0.025 ng; 2, HCB, 0.025 ng; 3, δ-HCH, 0.025 ng; 4, α-chlordene, 0.05 ng; 5, γ-chlordene, 0.05 ng; 6, *o,p′*-DDE, 0.05 ng; 7, α-chlordane, 0.05 ng; 8, dieldrin, 0.05 ng; 9, *o,p′*-TDE, 0.075 ng; 10, β-endosulfan, 0.03 ng; 11, *o,p′*-DDT, 0.052 ng; 12, endosulfan sulfate, 0.09 ng; 13, *p,p′*-DDT, 0.075 ng; 14, methoxychlor, 0.175 ng; IS, PCB 209, 0.03 ng. Column: DB-1701, 15 m × 0.53 mm ID × 1.0 μm.

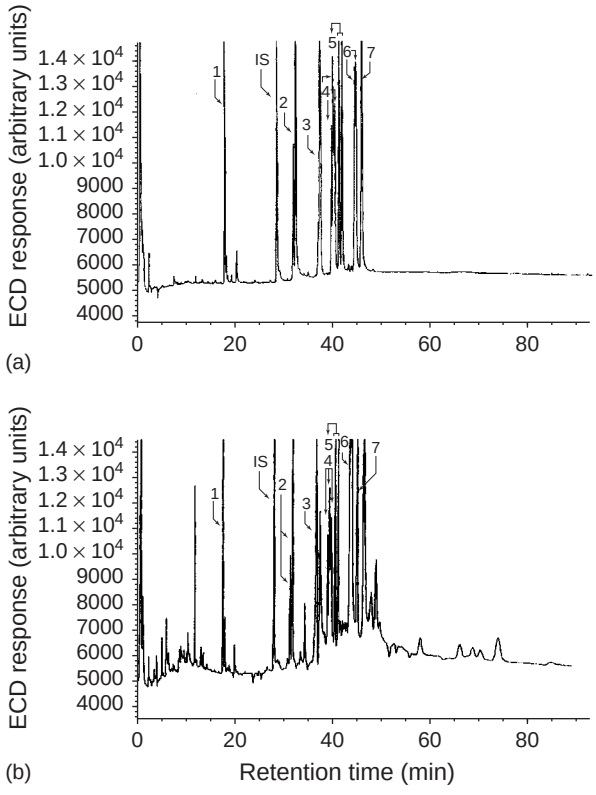

Figure 3 GC/ECD chromatograms. (a) Standard mixture PYR-A, 1 μL injected. IS, PCB 153, 0.05 ng; 1, tefluthrin, 0.12 ng; 2, tetramethrin, 0.51 ng; 3, cyphenothrin, 0.25 ng; 4, cyfluthrin, 0.20 ng; 5, flucythrinate, 0.25 ng; 6, fluvalinate, 0.30 ng; 7, deltamethrin, 0.15 ng. (b) Soya oil spiked at the following levels (mg kg^{-1}) with PYR-A, matrix concentration 1 g per 5 mL, 1 μL injected. IS, PCB 153, 0.05 ng; 1, tefluthrin, 0.61; 2, tetramethrin, 2.57; 3, cyphenothrin, 1.27; 4, cyfluthrin, 0.98; 5, flucythrinate, 1.26; 6, fluvalinate, 1.52; 7, deltamethrin, 0.76. Column: DB-1, 15 m × 0.53 mm ID × 1.5 μm.

with the corresponding standard mixtures are shown in Figures 3 and 4, respectively. For nonhalogen-containing pesticides, halogenated derivatives can be prepared to improve the detection with ECD.[95] A recent development is the nonradioactive ECD. Conventional ECD uses a ^{63}Ni foil as a source of electrons. In nonradioactive ECD, a pulsed discharge in pure helium generates high-energy photons that are used to photoionize a dopant gas added to the cell. The resulting electrons are used in the electron capture process. The pulsed discharge ECD offers several advantages over the ECD: (1) cleanliness, because the GC effluent does not come into contact with the ionization source; (2) linearity of four orders of magnitude; (3) sensitivity equal or better than the ECD; (4) low internal volume; and (5) elimination of the radioactive foil and associated safety measures.[161,162]

The ElCD (or Hall detector) is based on the catalytic reduction of halogen-containing compounds that produce hydrohalic acids which are detected and measured through the increase of electrolytic conductivity in a cell flushed by propanol. It is more selective than the ECD for halogens and is used for the analysis of halogen-containing PYR and OC pesticides.[25,46,104,105,111,163] Being halogen-selective, the ElCD is less sensitive to nonhalogen, electron-capturing interferences than the ECD. A number of fatty foods were analyzed for OC and PYR pesticides after SFE extraction and cleanup, without further cleanup.[164]

The XSD has been specifically developed for the selective detection of halogen-containing compounds in GC. Unlike the ElCD, the XSD does not require catalyst tubes, solvents, resin cartridges, a pump, or a transfer

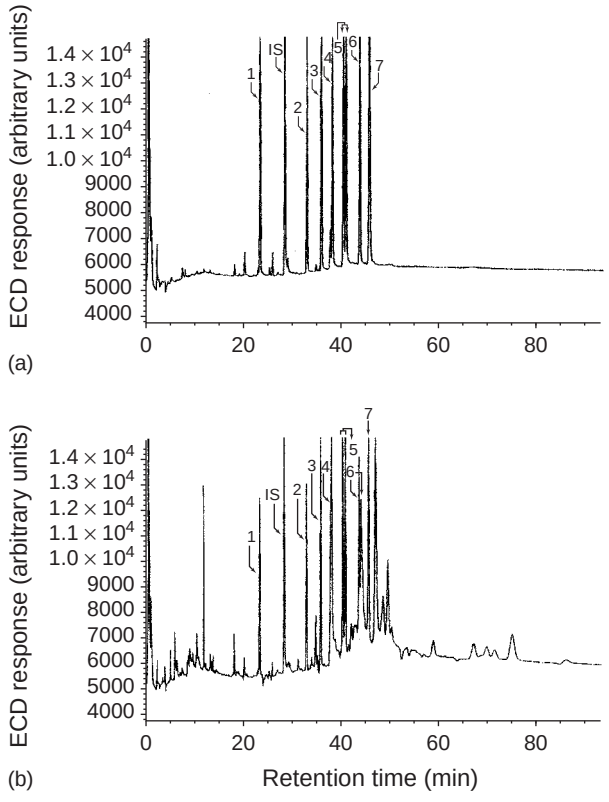

Figure 4 GC/ECD chromatograms. (a) Standard mixture PYR-B, 1 µL injected. IS, PCB 153, 0.05 ng; 1, bioallethrin, 0.10 ng; 2, fenpropathrin, 0.10 ng; 3, λ-cyhalothrin, 0.11 ng; 4, permethrin, 0.20 ng; 5, cypermethrin, 0.25 ng; 6, esfenvalerate, 0.10 ng; 7, tralomethrin, 0.16 ng. (b) Soya oil spiked at the following levels (mg kg^{-1}) with PYR-B, matrix concentration 1 g per 5 mL, 1 µL injected. IS, PCB 153, 0.05 ng; 1, bioallethrin, 0.49; 2, fenpropathrin, 0.49; 3, λ-cyhalothrin, 0.51; 4, permethrin, 0.97; 5, cypermethrin, 1.23; 6, esfenvalerate, 0.49; 7, tralomethrin, 0.76. Column: DB-1, 15 m × 0.53 mm ID × 1.5 µm.

line. The XSD does not contain a radioactive source. The reactor is operated in an oxidative mode, which pyrolyzes the effluent from a GC column. This oxidative pyrolysis efficiently converts compounds containing halogen to their oxidation products and free halogen atoms. The cathodic surface is activated by neutralization of alkali ions emitted from the anodic surface. The adsorption and reaction of free chlorine atoms with this alkali-sensitized cathodic surface yields an increased thermionic emission comprised of free electrons and halogen ions. The total cathodic current is measured. XSD has a linear range of 10^4 and a selectivity Cl/HC $>10^4$.

The FPD, NPD, and CLD are element-selective detectors for compounds containing P or S, N and P, and N or S, respectively, but are not necessarily of interest in trace analysis for OC and PYR pesticides. Although some PYR contain N, they are not sensitively detected by N-selective detectors and interferences of N are greater than S or Cl. Endosulfan can be detected via S with pulsed FPD (PFPD) or CLD. The new PFPD is based on a pulsed flame and combustible gas flow rate that cannot sustain a continuous flame operation, and a pulsed emission of light is obtained. The pulsed flame operation shows some interesting features in pesticide residue analysis, including increased sensitivity for P- and S-containing compounds along with the possibility of distinguishing between the two signals and selectivity for other elements such as N, Sn, that are of interest in pesticide residue analysis.[165,166] The increased sensitivity for S would allow the detection of endosulfan and especially endosulfan sulfate. Indeed endosulfan sulfate, being more polar, usually occurs in dirtier fractions accompanied by more matrix interferences.

The coupling of a microwave-induced helium plasma atomic emission detector (AED) to GC has introduced multiple element selectivity for helping to identify organic chemicals in GC. Depending on the spectral lines selected, GC/AED can acquire signals for different elements in a single run. For example, Cl, Br, F, and N atoms for PYR and Cl for OC pesticides can be monitored.[167] This technique is less popular than GC/MS because of expense, reduced sensitivity and a lesser degree of confirmation.[168] AED has the potential in multiresidue analysis to detect compounds with metallic elements, for instance tin,[169] and to complement other element-selective detectors or MS. Indeed assembling a matrix of retention time parameters and responses to selective detectors is a common approach to identify compounds in screening samples for multiple pesticides in fruits and vegetables.[163,166,167]

5.1.5 Identification and Determination Procedures

For compounds analyzed by GC, tentative identification is based on matching the retention times or the relative retention times to those of standards. The minimum criterion for a tentative identification with selective detectors is the correspondence of the retention parameters on at least two columns of different polarity. Several compilations are available that give relative retention times on GC capillary columns of different polarity useful in multiclass, multiresidue analysis.[9–11,15,20,46,142] To reduce time for identification, GC instruments are often equipped with twin injectors (or a single injector connected to two columns with a Y-splitter) for simultaneous injection into two columns of different polarity connected to two ECDs,[44,142,155,156,170] or two different detectors, for instance an ECD and an ElCD in the determination of OC and PCB in foods and environmental samples.[105,171] In multiclass, multiresidue methods an ECD or an ElCD is combined with an NPD or an FPD.[46,54] The two chromatograms are acquired simultaneously. Cross matching the retention times helps the identification of the analytes.

Less frequently used is bidimensional GC in which a ''heart cut'' from the first column is transferred to a second column of different polarity.[7,156,172,173] Comprehensive orthogonal GC on two short columns of different polarity connected through a two-stage thermal desorption modulator has been applied to analyze OC pesticides from human serum in a few minutes with powerful separation capabilities.[156,174] In this case a complex bidimensional chromatogram is obtained that requires unconventional techniques for interpretation.

Further evidence for identification is provided by matching the analytical behavior through a combination of techniques based on different mechanisms of separation. Target compounds can be separated by adsorption TLC and transferred off-line to a gas chromatograph (matching R_f and retention time). Compounds separated by HPLC can be transferred to GC off-line or on-line. In on-line operation the peak of interest is trapped in a loop and by valve-switching transferred to a gas chromatograph equipped with a large-volume injection system. However these operations are useful for confirmation of target analyte(s) step by step.[175]

In GC- and HPLC-based techniques quantitation is usually carried out through peak area comparison with solutions of the standard compound at known concentration and either using a multilevel calibration curve, or interpolating between two standard solutions whose responses ''bracket'' the response of the sample (bracketing standards).[46] In some circumstances, single-point comparison is permissible.[37] The use of an IS is preferred to control final volume and injected volume variations. Not infrequently a ''matrix effect'' is observed in quantitation of pesticide residues at low levels. The matrix effect describes the difference of the response of the analytes injected in pure solvent (the usual ''solvent standards'') or mixed with the matrix components as they occur in the final extract (''matrix-matched standards''). Usually an increase in response is observed when the standard is injected with the matrix. Thus, if a sample component is quantitated against a standard in pure solvent, an overestimation of the component in the sample can often be observed. The matrix coextractives have a protective effect on the analyte during the injection and transfer into the column. The matrix effect depends on the nature of the compounds, and the performances of the GC injector. It is mostly apparent with polar, easily degradable compounds such as phosphoramides. Thus, the use of matrix-matched standards is recommended. The matrix effect has been observed with PYR,[125] and to a lesser extent with OC pesticides.[16,28,46,79,97,176] In a separate mechanism, in both GC- and HPLC-based quantitation the coelution of matrix components with the analyte, even if not detected by selective detectors, may alter the detector response. So, although certain analyses can be carried out with minimal cleanup, a more rigorous separation of matrix interferences may be beneficial from the quantitation and ruggedness points of view.

5.1.6 Gas Chromatography/Mass Spectrometry and Gas Chromatography/Fourier Transform Infrared Spectrometry

For on-line fast confirmation in multiresidue methods, GC takes advantage of the coupling to spectrometric identification techniques, such as MS and infrared (IR) spectrometry.

GC/MS is the most widely used technique for identification of organic GC-amenable compounds.[177] GC/MS is of paramount importance in analyzing PCB and toxaphene in environmental samples, and for the identification of metabolites and degradation products, for instance, of chlordane,[157] pentachloronitrobenzene,[178,179] and cypermethrin.[180] With modern benchtop GC/MS systems operating with capillary columns, the column is inserted directly into the ion source. Benchtop GC/MS instruments are based on either a quadrupole mass filter[20,29,46,90,134,163,168,181,182] or an ion trap detector (ITD).[69,97,99,115,183,184] The ITD also allows simple MS/MS operation, i.e. obtaining a mass spectrum from a previously formed fragment, that strongly increases the compound specificity of the technique.[185,186] Different ionization techniques are available, such as electron impact (EI) and chemical ionization (CI) in positive or negative mode. EI ionization mode gives extensive fragmentation that is characteristic of the molecular structure of the compound. Most frequently used is GC/EIMS, for which large electronic libraries of spectra are available that can be searched by appropriate software available on the data station. This search can often be carried out in combination with other parameters such as retention time and elemental information to restrict the number of candidate identifications.[166] Full spectra can be acquired with fractions of a nanogram, depending on the compound injected. Comparison of a full spectrum with the library spectrum and the spectrum of the authentic compound at the appropriate retention time gives almost unequivocal identification.

In selected ion monitoring (SIM) mode, also named multiple ion detection, MS is a sort of compound-selective detector tunable to different sets of ions. Decreasing the number of ions to be collected usually increases the sensitivity. Thus, SIM mode provides the highest sensitivity.

In multiresidue methods, for screening purposes the SIM mode is used, i.e. some diagnostic ions are followed in time windows appropriate to the detection of target compounds.[20,46,90,123,134,181,187] For reliable identification, the appearance of one to three ions at the correct

retention time and with the correct intensity ratios is considered necessary.[184] Several methods report a list of ions (m/z) that may be used to select diagnostic ions for SIM analysis.[15,16,17,20,46] However, a full EI spectrum should be obtained for unequivocal confirmation. In this respect the ITD is very sensitive in full scan mode and full scan spectra can be obtained with sub-nanogram amounts.

CI (more commonly with methane) gives less fragmentation than the EI mode and spectra with only a few ions are recorded, with $M+1$ or $M+CH_4$ being normally the most abundant in positive CI. The lesser fragmentation results in increased sensitivity. Although less definitive than the EI spectrum, the CI spectrum together with retention time data provides sufficient evidence for the presence of an analyte.[69,184] Negative ion CIMS takes advantage of the electronegative character of polychlorinated compounds. Negative ion CIMS has been used to detect PCB and OC pesticides with high sensitivity in plant materials,[29] human blood,[182] and the environment.[156] Negative ion CIMS with methane has also been used for the determination of four PYR in soil,[134] and offers sensitivity of the same order of GC/ECD.[153]

High-resolution MS is used as the ultimate identification technique at very low levels in food,[46] and in environmental studies.[71,156]

The gas chromatograph coupled to a Fourier transform infrared (FTIR) spectrometer is also a powerful identification technique but has not achieved widespread routine application in pesticide residue analysis.[188]

5.2 High-performance Liquid Chromatography

With the development of high-performance columns and sensitive detectors the applications of HPLC to residues analysis are paralleling the GC applications. HPLC is primarily suited to the analysis of polar, low-volatility, and thermally labile compounds. This is particularly the case for the new classes of pesticides having more complex structures and increasing molecular weight. A limitation of HPLC compared to GC is the fewer types of detectors. Traditionally, the variable-wavelength ultraviolet/visible (UV/VIS) detector usually operated in the 200–350 nm range and the fluorescence detector are the main choices. Most OC pesticides are poorly detected by either technique and for them HPLC has been used mainly as a separative cleanup, as mentioned previously. PYR can be analyzed at residue level by the UV detector.[32,38,138,139] Particularly for nonhalogen-containing PYR, such as allethrin, fenpropathrin, phenothrin, resmethrin, and tetramethrin, HPLC can provide the only determinative technique at residue level apart from GC/MS.[153] Reverse-phase chromatography either on octyl-silica (C_8) or octadecyl-silica (C_{18}) is the most popular separation mechanism, where low polarity compounds, such OC and PYR pesticides, are strongly retained, while more polar compounds elute earlier.

Fluorescence detection is useful for naturally fluorescent compounds or compounds that can be derivatized to fluorescent compounds. HPLC with fluorescence detection has been used for the determination of residues of piperonyl butoxide, a synergist used in PYR formulations, in rice, grains, and beans.[153]

The analysis of single enantiomers is important not only to characterize the formulated products, but also for appropriate assignment of the MRL value in residue analysis.[153] Chiral phases are available as HPLC columns that permit the separation of some of these enantiomers.[153,189]

As in GC, HPLC coupled to spectral techniques is necessary for a reliable identification at trace levels. Fast scanning or photodiode array (PDA) spectrophotometers allow the identification of compounds by retention time and spectrum comparison with authentic compounds under the same conditions. HPLC with PDA instrumentation has become common in the residue laboratory.

HPLC, especially with minibore columns, can be coupled to GC. Selected peaks from HPLC can be transferred to GC by valve switching technology and large-volume injection systems. HPLC/GC can, thus, take advantage of the sensitivity and selectivity of the array of detectors developed for GC, including GC/MS, and GC takes advantage of the selectivity options offered by liquid chromatography. Thus, the identification can be supported by two mechanisms of separation and a spectral technique.[175]

After years of continuous developments and improvements, HPLC/MS coupling has almost reached the simplicity and reliability of GC/MS, at least for benchtop, quadrupole or ion trap mass analyzers. Nowadays the most used interfaces are electrospray and atmospheric pressure CI. Thus, mass spectral confirmation can be obtained also for polar, heat-labile, low volatility, high molecular weight compounds, which include PYR and other new pesticides. HPLC/MS gives CI-type mass spectra which typically have peaks for structural elucidation or confirmation purposes. Both positive and negative ion extraction mode is possible. The choice between positive and negative ion mode depends on the compounds to be analyzed. In the positive ion mode the base peaks are normally the protonated forms of the compounds or their adducts with the cations used in the HPLC eluent. The negative ion mode is more sensitive for electronegative compounds and the base peaks are normally the adducts with the

anions used in the eluent buffer. Also, apparatus is available for HPLC/MS that can perform the MS/MS operation.[190–193]

5.3 Thin-layer Chromatography

Adsorption TLC on alumina has been used for OC pesticides.[8] The eluting solvent was *n*-hexane or its mixtures with a small amount of polar modifiers, such as DCM or acetone. Detection of OC pesticides on the plates was carried out by spraying a chromogenic reagent (a solution of silver nitrate in 2-phenoxyethanol–acetone) and exposing the plate to UV light. Quantities as low as 5 ng of most OC pesticides can be detected. Separation and identification on silica gel plates have been reported.[194,195] Although almost superseded by other techniques, in pesticide residue analysis TLC is gaining a renewed interest after the introduction of high-performance thin-layer chromatography (HPTLC) plates and automated multiple development (AMD) with inverse gradient elution, especially in the analysis of environmental samples.[196–198] Very small bands with reproducible R_f are obtained so that the separation of a large number of compounds in a single run is possible. Identification can be obtained through comparison of R_f and UV or fluorescence spectra acquired with spectrophotometric scanners. The quantitation is possible at very low levels and is carried out by peak areas comparison in multiple-wavelength scanning.

Micro liquid chromatography can be coupled to HPTLC by nitrogen-assisted spraying of different portions of the HPLC eluent onto different lanes of the HPTLC plate. Afterward the plate is developed with the AMD mode and each lane is scanned for peak identification and quantitation. HPLC/HPTLC is a powerful identification tool which combines the high resolution power of HPLC and HPTLC with two different mechanisms of separation (for instance reverse-phase HPLC and adsorption TLC) in an orthogonal high peak capacity separation system, particularly suited for polar heat-labile compounds.[198]

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detector
AMD	Automated Multiple Development
CI	Chemical Ionization
CLD	Chemiluminescent Detector
DCBP	Dichlorobenzophenone
DCM	Dichloromethane
DDA	Di(*p*-chlorophenyl)acetic Acid
DDD	Dichlorodiphenyldichloroethane
DDE	Dichlorodiphenyldichloroethylene
DDT	Dichlorodiphenyltrichloroethane
DMF	Dimethylformamide
DMSO	Dimethyl Sulfoxide
DVB	Divinylbenzene
ECD	Electron Capture Detector
EI	Electron Impact
ElCD	Electroconductivity Detector
EPC	Electronic Pressure Control
FPD	Flame Photometric Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HCB	Hexachlorobenzene
HCH	Hexachlorocyclohexane
HEPO	Heptachlor Epoxide
HETP	Height Equivalent Theoretical Plate
HPLC	High-performance Liquid Chromatography
HPSEC	High-performance Size-exclusion Chromatography
HPTLC	High-performance Thin-layer Chromatography
IR	Infrared
IS	Internal Standard
ITD	Ion Trap Detector
LLE	Liquid–Liquid Extraction
LLP	Liquid–Liquid Partitioning
MAE	Microwave-assisted Extraction
MRL	Maximum Residue Level
MS	Mass Spectrometry
MSPD	Matrix Solid-phase Dispersion
NPD	Nitrogen–Phosphorus Detector
NPHPLC	Normal Phase High-performance Liquid Chromatography
OC	Organochlorine
OP	Organophosphorus
PCB	Polychlorinated Biphenyl
PDA	Photodiode Array
PFPD	Pulsed FPD
PLE	Pressurized Liquid Extraction
PSA	Propylethylenediamine – Silica
PS–DVB	Polystyrene–Divinylbenzene
PYR	Pyrethroids
R_f	Retention Factor
SAX	Strong Anion Exchanger
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TLC	Thin-layer Chromatography
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
XSD	Halogen-specific Detector

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Laser Mass Spectrometry in Trace Analysis

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Industrial Waste Dumps, Sampling and Analysis • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Polychlorinated Biphenyls Analysis in Environmental Samples • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Liquid Chromatography in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General • Sample Preparation, Headspace Techniques

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicide Residues in Biota, Analysis of • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation • *s*-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy • Liquid Chromatography/Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. G.W. Ware, *Pesticides. Theory and Application*, Freeman and Company, San Francisco, 1978.
2. R.L. Metcalf, *Organic Insecticides. Their Chemistry and Mode of Action*, Interscience Publishers, New York, 1955.
3. G. Matolcsy, M. Nádasy, V. Andriska, *Pesticide Chemistry*, Elsevier, Amsterdam, Vol. 32, 1988.
4. WHO, *Environmental Health Criteria*, World Health Organization, Geneva, Vols. 2, 9, 34, 38, 40, 41, 44, 45, 82, 83, 87, 88, 91, 92, 94, 95, 96, 97, 98, 99, 123, 124, 130, 140, 195, 1976–1997.
5. C. Tomlin (ed.), *The Pesticide Manual*, 10th edition, The Royal Society of Chemistry, Cambridge, 1995.
6. L. Fishbein, 'Chromatography and Biological Aspects of DDT and its Metabolites', *J. Chromatogr.*, **98**, 177–251 (1974).
7. J. de Boer, 'Capillary Gas Chromatography for the Determination of Halogenated Micro-contaminants', *J. Chromatogr. A*, **843**, 179–198 (1999).
8. P. Cunniff (ed.), *Official Methods of Analysis of the Association of Official Analytical Chemists International*, AOAC International, Gaithersburg, Vol. 1, 1996.
9. B. McMahon, N.F. Hardin (eds.), *Pesticide Analytical Manual*, US Department of HHS, Food and Drug Administration, Rockville, Vol. 1, 1994.
10. H.-P. Thier, H. Zumer, J. Kirchoff (eds.), *Deutsche Forschung Gemeinschaft – Manual of Pesticide Residue Analysis*, VCH Verlaggesellschaft, Weinheim, Vol. I, 1987 and Vol. II, 1992.
11. J. Sherma (ed.), *Manual of Analytical Methods for the Analysis of Pesticides in Human and Environmental Samples*, US Environmental Protection Agency, 600/8-80-038, NTIS, Atlanta, 1980.
12. G. Zweig, J. Sherma (eds.), *Analytical Methods for Pesticides and Plant Growth Regulators*, Academic Press, New York, Vols. 1–17, 1963–1989.
13. J. Sherma, T. Cairns (eds.), *Comprehensive Analytical Profiles of Important Pesticides*, CRC Press, Boca Raton, 1992.
14. J. Sherma, T. Cairns (eds.), *Emerging Strategies for Pesticide Analysis*, CRC Press, Boca Raton, 1992.
15. P. ven Zoonen (ed.), *Analytical Methods for Pesticide Residues in Foodstuff*, 6th edition, General Inspectorate for Health Protection, The Netherlands, 1996.
16. 'Pesticide Analytical Methods in Sweden', Part I, Report 17/98 (ISSN 1104-7089), NFA – National Food Administration, Uppsala, 1998.
17. 'Multiresidue Method for the Analysis of Pesticide Residues in Vegetable Products' (in Italian), Rapporti ISTISAN 97/23 (ISSN 1123-3117), ISS – Istituto Superiore di Sanità, Rome, 1997.
18. S.J.V. Young, C.H. Parfitt, Jr, R.F. Newell, T.D. Spittler, 'Homogeneity of Fruits and Vegetables Comminuted in a Vertical Cutter Mixer', *J. AOAC Int.*, **79**, 976–980 (1996).
19. Á. Ambrus, E.M. Solymosné, I. Korsós, 'Estimation of Uncertainty of Sample Preparation for the Analysis of Pesticide Residues', *J. Environ. Sci. Health*, **B31**, 443–450 (1996).
20. J. Fillion, R. Hindle, M. Lacroix, J. Selwyn, 'Multiresidue Determination of Pesticides in Fruit and Vegetables by Gas Chromatography – Mass Selective Detection and Liquid Chromatography with Fluorescence Detection', *J. AOAC. Int.*, **78**, 1252–1266 (1995).
21. S.M. Lee, M.L. Papathakis, H.-M.C. Feng, G.F. Hunter, J.E. Carr, 'Multipesticide Residue Method for Fruits and Vegetables: California Department of Food and Agriculture', *Fresenius' J. Anal. Chem.*, **339**, 376–383 (1991).
22. G. Becker, 'Eine Multimethode zur gleichzeitigen Erfassung von 75 Pflanzenbehandlungsmitteln auf pflanzichen Material', *Dtsch. Lebensmittel. Rdsch.*, **75**, 148–152 (1979).
23. Á. Ambrus, J. Lantos, É. Visi, I. Csatlós, L. Sárvári, 'General Method for Determination of Pesticide Residues in Samples of Plant Origin, Soil and Water. I. Extraction and Cleanup', *J. Assoc. Off. Anal. Chem.*, **64**, 733–742 (1981).
24. M.A. Luke, J.E. Froberg, H.T. Masumoto, 'Extraction of Organochlorine, Organophosphate, Organonitrogen, and Hydrocarbon Pesticide in Produce for Determination by Gas–Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **58**, 1020–1026 (1975).
25. M.A. Luke, J.E. Froberg, G.M. Doose, H.T. Masumoto, 'Improved Multiresidue Gas Chromatographic Determination of Organophosphorus, Organonitrogen, and Organohalogen Pesticides in Produce using Flame Photometric and Electrolytic Conductivity Detectors', *J. Assoc. Off. Anal. Chem.*, **64**, 1187–1195 (1981).
26. W. Specht, M. Tilkes, 'Gas-chromatographische Bestimmung von Rückständen an Pflanzenbehandlungsmitteln nach Cleanup über Gel-chromatographie und Minikieselgel–Säulen Chromatographie', *Fresenius' Z. Anal. Chem.*, **322**, 443–455 (1985).
27. P. Sukul, 'Extraction and Cleanup Procedures for the Analysis of Permethrin, Cypermethrin, Deltamethrin and Fenvalerate in Crops by GLC–ECD', *Toxicol. Environ. Chem.*, **44**, 217–223 (1994).
28. P.D. Johnson, D.A. Rimmer, R.H. Brown, 'Adaptation and Application of a Multiresidue Method for the Determination of a Range of Pesticides, Including Phenoxy Acid Herbicides in Vegetation, Based on High-resolution Gel Permeation Chromatographic Cleanup and Gas Chromatographic Analysis with Mass-selective Detection', *J. Chromatogr. A*, **765**, 3–11 (1997).
29. G. Niessner, W. Buchberger, R. Eckerstorfer, 'Multiresidue Screening Methods for the Determination of Pesticides in Plant Materials', *J. Chromatogr. A*, **846**, 341–348 (1999).

30. H.B. Wan, M.K. Wong, P.Y. Lim, C.Y. Mok, 'Small Scale Multiresidue Method for the Determination of Organochlorine and Pyrethroid Pesticides in Vegetables', *J. Chromatogr. A*, **662**, 147–152 (1994).
31. G.-F. Pang, C.-L. Fan, Y.-Z. Chao, T.-S. Zhao, 'Rapid Method for the Determination of Multiple Pyrethroid Residues in Fruits and Vegetables by Capillary Column Gas Chromatography', *J. Chromatogr. A*, **667**, 348–353 (1994).
32. G.-F. Pang, Y.-Z. Chao, X.-S. Liu, C.-L. Fan, 'Multiresidue Liquid Chromatographic Method for Simultaneous Determination of Pyrethroid Insecticide in Fruits and Vegetables', *J. AOAC Int.*, **78**, 1474–1480 (1995).
33. A.H. Roos, A.J. Van Munsteren, F.M. Nab, L.G.M.Th. Tuinstra, 'Universal Extraction/Cleanup Procedure for Screening of Pesticides by Extraction with Ethyl Acetate and Size Exclusion Chromatography', *Anal. Chim. Acta*, **196**, 95–102 (1987).
34. A. Andersson, H. Pålsheden, 'Comparison of the Efficiency of Different GLC Multiresidue Methods on Crops Containing Pesticide Residues', *Fresenius' J. Anal. Chem.*, **339**, 365–367 (1991).
35. G.P. Molinari, S. Cavanna, B. Ferroni, 'Multiresidue Method for Determination of Organophosphorus Pesticides in Vegetables', *Food Add. Contam.*, **15**, 510–517 (1998).
36. A. Ambrus, H.-P. Thier, 'Application of Multiresidue Procedures in Pesticides Residues Analysis', *Pure Appl. Chem.*, **58**, 1035–1062 (1986).
37. G.F. Pang, Y.-Z. Cao, C.-L. Fan, J.-J. Zhang, X.-M. Li, 'Multiresidue Gas Chromatographic Method for Determining Synthetic Pyrethroid Pesticides in Agricultural Products: Collaborative Study', *J. AOAC Int.*, **82**, 186–212 (1999).
38. P. Bottomley, P.G. Baker, 'Multiresidue Determination of Organochlorine, Organophosphorous and Synthetic Pyrethroid Pesticides in Grain by Gas–Liquid and High-performance Liquid Chromatography', *Analyst*, **109**, 85–90 (1984).
39. Committee for Analytical Methods for Residues of Pesticides and Veterinary Products in Foodstuffs and the Working Party on Pesticide Residues of the Ministry of Agriculture, Fisheries and Food, 'Determination of a Range of Organophosphorous Pesticide Residues in Grain', *Analyst*, **110**, 756–758 (1985).
40. N.F. Wood, 'Extraction and Cleanup of Organochlorine Pesticide Residues by Column Chromatography', *Analyst*, **94**, 399–405 (1969).
41. G.D. Veith, D.W. Kuehl, J. Rosenthal, 'Preparative Method for Gas Chromatographic/Mass Spectral Analysis of Trace Quantities of Pesticides in Fish Tissue', *J. Assoc. Off. Anal. Chem.*, **58**, 1–5 (1975).
42. L. Berdié, J.O. Grimalt, 'Assessment of the Sample Handling Procedures in a Labor-saving Method for the Analysis of Organochlorine Compounds in a Large Number of Fish Samples', *J. Chromatogr. A*, **823**, 373–380 (1998).
43. A. Bentabol, M. Jodral, 'Determination of Organochlorine Pesticides in Cheese', *J. AOAC Int.*, **78**, 94–98 (1995).
44. A.R. Najam, M.P. Korver, C.C. Williams, V.W. Burse, L.L. Needham, 'Analysis of a Mixture of Polychlorinated Biphenyls and Chlorinated Pesticides in Human Serum by Column Fractionation and Dual-column Capillary Gas Chromatography with Electron Capture Detection', *J. AOAC Int.*, **82**, 177–185 (1999).
45. E.A. Hogendoorn, G.R. van der Hoff, P. van Zoonen, 'Automated Sample Cleanup and Fractionation of Organochlorine Pesticides and Polychlorinated Biphenyls in Human Milk using NP–HPLC with Column Switching', *J. High Res. Chromatogr.*, **12**, 784–789 (1989).
46. D.A. Bennett, A.C. Chung, S.M. Lee, 'Multiresidue Methods for Analysis of Pesticides in Liquid Whole Milk', *J. AOAC Int.*, **80**, 1065–1077 (1997).
47. T. Prapamontol, D. Stevenson, 'Rapid Method for the Determination of Organochlorine Pesticides in Milk', *J. Chromatogr.*, **552**, 249–257 (1991).
48. L. Giuffrida, D.C. Bostwick, N.F. Ives, 'Rapid Cleanup Techniques for Chlorinated Pesticide Residues in Milk, Fats and Oils', *J. Assoc. Off. Anal. Chem.*, **49**, 634–638 (1966).
49. F.J. Schenck, L. Calderon, D.E. Sudang, 'Florisil Solid-phase Extraction Cartridges for Cleanup of Organochlorine Pesticide Residues in Foods', *J. AOAC Int.*, **79**, 1454–1458 (1996).
50. N. Furusawa, A. Ozaki, M. Nakamura, Y. Morita, K. Okazaki, 'Simple and Rapid Extraction Method of Total Egg Lipids for Determining Organochlorine Pesticides in the Egg', *J. Chromatogr. A*, **830**, 473–476 (1999).
51. F.J. Schenck, L. Calderon, L.V. Podhorniak, 'Determination of Organochlorine Pesticide and Polychlorinated Biphenyl Residues in Fatty Fish by Tandem Solid-phase Extraction Cleanup', *J. AOAC Int.*, **79**, 1209–1214 (1996).
52. T. Suzuki, K.E. Ishikawa, N. Sato, K.-I. Sakai, 'Determination of Chlorinated Pesticide Residues in Food. I. Rapid Screening Method for Chlorinated Pesticides in Milk', *J. Assoc. Off. Anal. Chem.*, **62**, 681–684 (1979).
53. L. Dagna, G. Gasparini, M.L. Icardi, E. Sesia, 'Metodo Multiresiduo di Estrazione su Colonna per la Determinazione di Residui di Fitofarmaci in Matrici Vegetali e Liquide', *Boll. Chim. Igien.*, **44**, 383–397 (1993).
54. C. Bicchi, A. D'Amato, C. Balbo, 'Multiresidue Method for Quantitative Gas Chromatographic Determination of Pesticide Residues in Sweet Cherries', *J. AOAC Int.*, **80**, 1281–1286 (1997).
55. F. Sicbaldi, A. Sarra, G.L. Copeta, 'Diatomaceous Earth-assisted Extraction for the Multiresidue Determination of Pesticides', *J. Chromatogr. A*, **765**, 23–30 (1997).

56. L. Kadenczki, Z. Arpad, I. Gardi, Á. Ambrus, L. Gyorfi, G. Reese, W. Ebing, 'Column Extraction of Residues of Several Pesticides from Fruits and Vegetables: A Simple Multiresidue Analysis Method', *J. AOAC Int.*, **75**, 53–61 (1992).
57. Y.-C. Ling, I.-P. Huang, 'Multiresidue Matrix Solid-phase Dispersion Method for Determination of Six Synthetic Pyrethroids in Vegetables Followed by Gas Chromatography with Electron Capture Detection', *J. Chromatogr. A*, **695**, 75–82 (1995).
58. H. Steinwandter, 'Contributions to Silica Gel Application in Pesticide Residue Analysis. III. An On-line Method for Extracting and Isolating Chlorinated Hydrocarbon Pesticides and Polychlorinated Biphenyls (PCBs) from Milk and Dairy Products', *Fresenius' Z. Anal. Chem.*, **312**, 342–345 (1982).
59. H. Steinwandter, 'Contributions to Silica Gel Application in Pesticide Residue Analysis. IV. On-line Method for Extracting and Isolating Chlorinated Hydrocarbon Pesticides and Polychlorinated Biphenyls (PCBs) from Meat and Fish Products', *Fresenius' Z. Anal. Chem.*, **314**, 129–130 (1983).
60. H.M. Lott, S.A. Barker, 'Extraction and Gas Chromatographic Screening of 14 Chlorinated Pesticides in Crayfish (*Procambarus clarkii*) Hepatopancreas', *J. AOAC Int.*, **76**, 663–668 (1993).
61. C.C. Walker, H.M. Lott, S.A. Barker, 'Matrix Solid-phase Dispersion Extraction and the Analysis of Drugs and Environmental Pollutants in Aquatic Species', *J. Chromatogr.*, **642**, 225–242 (1993).
62. S.A. Barker, A.R. Long, M.E. Eines, II, 'Disruption and Fractionation of Biological Materials by Matrix Solid-phase Dispersion', *J. Chromatogr.*, **629**, 23–34 (1993).
63. A.J. Krynitsky, C.J. Stafford, S.N. Wiemeyer, 'Combined Extraction-cleanup Column Chromatographic Procedure for Determination of Dicofol in Avian Eggs', *J. Assoc. Off. Anal. Chem.*, **71**, 539–542 (1988).
64. A. Di Muccio, M. Rizzica, A. Ausili, I. Camoni, R. Dommarco, F. Vergori, 'Selective, On-column Extraction of Organochlorine Pesticide Residues from Milk', *J. Chromatogr.*, **456**, 143–148 (1988).
65. A. Di Muccio, P. Pelosi, I. Camoni, D. Attard Barbini, R. Dommarco, T. Generali, A. Ausili, 'Selective, Solid-matrix Dispersion Extraction of Organophosphate Pesticide Residues from Milk', *J. Chromatogr. A*, **754**, 497–506 (1996).
66. A. Di Muccio, P. Pelosi, D. Attard Barbini, T. Generali, A. Ausili, F. Vergori, 'Selective Extraction of Pyrethroid Pesticide Residues from Milk by Solid-matrix Dispersion', *J. Chromatogr. A*, **765**, 51–60 (1997).
67. N. Kaushik, S.K. Handa, 'New Clean-up Method for Gas Chromatographic Analysis of Pyrethroid Residues', *Chromatographia*, **46**, 209–212 (1997).
68. S. Babić, M. Petrović, M. Kaštelan-Macan, 'Ultrasonic Solvent Extraction of Pesticides from Soil', *J. Chromatogr. A*, **823**, 3–9 (1998).
69. P. Mogadati, J.B. Louis, J.D. Rosen, 'Multiresidue Determination of Pesticides in High-organic-content Soils by Solid-phase Extraction and Gas Chromatography/Mass Spectrometry', *J. AOAC Int.*, **82**, 705–715 (1999).
70. V. Lopez-Avila, R. Young, N. Teplitsky, 'Microwave-assisted Extraction as an Alternative to Soxhlet, Sonication, and Supercritical Fluid Extraction', *J. AOAC Int.*, **79**, 142–156 (1996).
71. J.R. Donnelly, A.H. Grange, N.R. Herron, G.R. Nichol, J.L. Jeter, R.J. White, W. Brumley, J. van Emon, 'Modular Methodology for Determination of Polychlorinated Biphenyls in Soil as Aroclors and Individual Congeners', *J. AOAC Int.*, **79**, 953–961 (1996).
72. V. Lopez-Avila, J. Benedicto, K.M. Bauer, 'Stability of Organochlorine and Organophosphorus Pesticides when Extracted from Solid Matrixes with Microwave Energy', *J. AOAC Int.*, **81**, 1224–1232 (1998).
73. V. Lopez-Avila, R. Wesselman, K. Edgell, 'Gas-chromatographic–Electron Capture Detection Method for Determination of 29 Organochlorine Pesticides in Finished Drinking Water: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **73**, 276–286 (1990).
74. R. Garcia-Repetto, I. Garrido, M. Repetto, 'Determination of Organochlorine, Organophosphorus, and Triazine Pesticide Residues in Wine by Gas Chromatography with Electron Capture and Nitrogen–Phosphorus Detection', *J. AOAC Int.*, **79**, 1423–1427 (1996).
75. A. Fernández-Guetiérrez, J.L. Martínez-Vidal, F.J. Arrebola-Liébanas, A. Gonzalez-Casado, J.L. Vílchez, 'Determination of Endosulfan and Some Pyrethroids in Waters by Micro Liquid–Liquid Extraction and GC/MS', *Fresenius' J. Anal. Chem.*, **360**, 568–572 (1998).
76. G.R. van der Hoff, P. van Zoonen, 'Trace Analysis of Pesticides by Gas Chromatography', *J. Chromatogr. A*, **843**, 301–322 (1999).
77. J.C. Moltó, Y. Picó, J. Mañes, G. Font, 'Analysis of Polychlorinated Biphenyls in Aqueous Samples using C_{18} Glass Column Extraction', *J. AOAC Int.*, **75**, 714–719 (1992).
78. I. Vassilakis, D. Tsipi, M. Scoullos, 'Determination of a Variety of Chemical Classes of Pesticides in Surface and Ground Waters by Off-line Solid-phase Extraction, Gas Chromatography with Electron-capture and Nitrogen–Phosphorus Detection, and High-performance Liquid Chromatography with Post-column Derivatization and Fluorescence Detection', *J. Chromatogr. A*, **823**, 49–58 (1998).
79. J.L. Bernal, M.J. del Nozal, J.J. Jiménez, J.M. Rivera, 'Matrix Effects in the Determination of Acaricides and Fungicides in Must by Gas Chromatography with Electron-capture and Nitrogen–Phosphorus Detection', *J. Chromatogr. A*, **778**, 111–117 (1997).
80. R. Wintersteiger, B. Ofner, H. Juan, M. Windisch, 'Determination of Traces of Pyrethrins and Piperonyl Butoxide in Biological Material by High-performance

Liquid Chromatography', *J. Chromatogr. A*, **660**, 205–210 (1994).
81. P.H. Tang, J.S. Ho, J.W. Eichelberger, 'Determination of Organic Pollutants in Reagent Water by Liquid–Solid Extraction Followed by Supercritical Fluid Elution', *J. AOAC Int.*, **76**, 72–82 (1993).
82. B.A. Tomkins, R. Merriweather, R.A. Jenkins, C.K. Bayne, 'Determination of Eight Organochlorine Pesticides at Low Nanogram/Liter Concentrations in Groundwater using Filter Disk Extraction and Gas Chromatography', *J. AOAC Int.*, **75**, 1091–1099 (1992).
83. G.R. van der Hoff, F. Pelusio, U.A.Th. Brinkman, R.A. Baumann, P. van Zoonen, 'Automated Solid-phase Extraction Coupled to Gas Chromatography with Electron-capture Detection: A Combination of Extraction and Clean-up of Pyrethroid in the Analysis of Surface Water', *J. Chromatogr. A*, **719**, 59–67 (1996).
84. C.L. Arthur, J. Pawliszyn, 'Solid Phase Microextraction with Thermal Desorption using Fused Silica Optical Fibers', *Anal. Chem.*, **62**, 2145–2148 (1990).
85. J. Chen, J.B. Pawliszyn, 'Solid Phase Microextraction Coupled to High-performance Liquid Chromatography', *Anal. Chem.*, **67**, 2530–2533 (1995).
86. S. Magdic, A. Boyd-Boland, K. Jinno, J.B. Pawliszyn, 'Analysis of Organophosphorus Insecticides from Environmental Samples using Solid-phase Microextraction', *J. Chromatogr. A*, **736**, 219–228 (1996).
87. R. Young, V. Lopez-Avila, W.F. Beckert, 'On-line Determination of Organochlorine Pesticides in Water by Solid-phase Microextraction and Gas Chromatography with Electron Capture Detection', *J. High. Resol. Chromatogr.*, **19**, 247–256 (1996).
88. S. Magdic, J.B. Pawliszyn, 'Analysis of Organochlorine Pesticides using Solid-phase Microextraction', *J. Chromatogr. A*, **723**, 111–122 (1996).
89. R. Eisert, K. Levsen, 'Solid-phase Microextraction Coupled to Gas Chromatography: A New Method for the Analysis of Organic in Water', *J. Chromatogr. A*, **733**, 143–157 (1996).
90. C. Aguilar, S. Peñalver, E. Pocurull, F. Borrull, R.M. Marcé, 'Solid-phase Microextraction and Gas Chromatography with Mass Spectrometric Detection for the Determination of Pesticides in Aqueous Samples', *J. Chromatogr. A*, **795**, 105–115 (1998).
91. B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: A Technique for Sample Preparation', *Anal. Chem.*, **68**, 1033–1039 (1996).
92. E. Björklund, S. Bøwadt, T. Nilsson, L. Mathiasson, 'Pressurized Fluid Extraction of Polychlorinated Biphenyls in Solid Environmental Samples', *J. Chromatogr. A*, **836**, 285–293 (1999).
93. M.M. Schantz, J.J. Nichols, S.A. Wise, 'Evaluation of Pressurized Fluid Extraction for the Extraction of Environmental Matrix Reference Materials', *Anal. Chem.*, **69**, 4210–4219 (1997).
94. EPA Method 3545, 'Pressurized Fluid Extraction. Test Method for Evaluating Solid Waste', 3rd edition, Update III, EPA SW-846, US-GPO, Washington, DC, 1995.
95. E. Conte, R. Milani, G. Morali, F. Abballe, 'Comparison Between Accelerated Solvent Extraction and Traditional Extraction Methods for the Analysis of the Herbicide Diflufenican in Soil', *J. Chromatogr. A*, **765**, 121–125 (1997).
96. M.L. Hopper, J.W. King, 'Enhanced Supercritical Fluid Carbon Dioxide Extraction of Pesticides from Foods using Pelletized Diatomaceous Earth', *J. Assoc. Off. Anal. Chem.*, **74**, 661–666 (1991).
97. S.J. Lehotay, K.I. Eller, 'Development of a Method of Analysis for 46 Pesticides in Fruits and Vegetables by Supercritical Fluid Extraction and Gas Chromatography/Ion Trap Mass Spectrometry', *J. AOAC Int.*, **78**, 821–830 (1995).
98. S.J. Lehotay, N. Aharonson, E. Pfeil, M.A. Ibrahim, 'Development of a Sample Preparation Technique for Supercritical Fluid Extraction for Multiresidue Analysis of Pesticides in Produce', *J. AOAC Int.*, **78**, 831–840 (1995).
99. S.J. Lehotay, M.A. Ibrahim, 'Supercritical Fluid Extraction and Gas Chromatography/Ion Trap Mass Spectrometry of Pentachloronitrobenzene Pesticides in Vegetables', *J. AOAC Int.*, **78**, 445–452 (1995).
100. S. Bøwadt, B. Johansson, P. Fruekilde, M. Hansen, D. Zilli, B. Larsen, J. de Boer, 'Supercritical Fluid Extraction of Polychlorinated Biphenyls from Lyophilized Fish Tissue', *J. Chromatogr. A*, **675**, 189–204 (1994).
101. H.-B. Lee, T.E. Peart, A.J. Nimi, C.R. Knipe, 'Rapid Supercritical Carbon Dioxide Extraction Method for Determination of Polychlorinated Biphenyls in Fish', *J. AOAC Int.*, **78**, 437–444 (1995).
102. W.J. Yoo, L.T. Taylor, 'Supercritical Fluid Extraction of Polychlorinated Biphenyls and Organochlorine Pesticides from Freeze-dried Tissues of Marine Mussel (*Mytilus edulis*)', *J. AOAC Int.*, **80**, 1336–1345 (1997).
103. E.G. Alley, G. Lu, 'Analysis of Polychlorinated Biphenyls in Fatty Biological Matrices by On-line Supercritical Fluid Extraction and Supercritical Fluid Cleanup', *J. AOAC Int.*, **78**, 1051–1054 (1995).
104. M.L. Hopper, 'Extraction and Cleanup of Organochlorine and Organophosphorous Pesticide Residues in Fats by Supercritical Fluid Techniques', *J. AOAC Int.*, **80**, 639–646 (1997).
105. M.L. Hopper, J.W. King, J.H. Johnson, A.A. Serino, R.J. Butler, 'Multivessel Supercritical Fluid Extraction of Food Items in Total Diet Study', *J. AOAC Int.*, **78**, 1072–1079 (1995).
106. G.-F. Pang, Y.-Z. Chao, C.-L. Fan, J.-J. Zhang, X.-M. Li, T.-S. Zhao, 'Modification of AOAC Multiresidue Method for Determination of Synthetic Pyrethroid Residues in Fruits, Vegetables, and Grains. Part I: Acetonitrile Extraction System and Optimization of Florisil

Cleanup and Gas Chromatography', *J. AOAC Int.*, **78**, 1481–1488 (1995).
107. G.-F. Pang, Y.-Z. Chao, C.-L. Fan, J.-J. Zhang, X.-M. Li, Y.M. Liu, T.-S. Zhao, 'Modification of AOAC Multiresidue Method for Determination of Synthetic Pyrethroid Residues in Fruits, Vegetables, and Grains. Part II: Acetone Extraction System', *J. AOAC Int.*, **76**, 1489–1496 (1995).
108. E. Lacassie, M.-F. Dreyfuss, J.L. Daguet, M. Vignaud, P. Marquet, G. Lachâtre, 'Multiresidue Determination of Pesticides in Apples and Pears by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **805**, 319–326 (1998).
109. A. Di Muccio, D. Attard Barbini, T. Generali, P. Pelosi, A. Ausili, F. Vergori, I. Camoni, 'Clean-up of Aqueous Acetone Vegetable Extracts by Solid-matrix Partition for Pyrethroid Residue Determination by Gas Chromatography–Electron-capture Detection', *J. Chromatogr. A*, **765**, 39–49 (1997).
110. R.-C. Hsu, I. Biggs, N.K. Saini, 'Solid-phase Extraction Cleanup of Halogenated Organic Pesticides', *J. Agric. Food Chem.*, **39**, 1658–1666 (1991).
111. J.A. Casanova, 'Use of Solid-phase Extraction Disks for Analysis of Moderately Polar and Nonpolar Pesticides in High-moisture Foods', *J. AOAC Int.*, **79**, 936–940 (1996).
112. M.J. de Faubert Maunder, H. Egan, E.W. Godly, E.W. Hammond, J. Roburn, J. Thomson, 'Clean-up of Animal Fats and Dairy Products for the Analysis of Chlorinated Pesticide Residues', *Analyst*, **89**, 168–174 (1964).
113. H.O. Friestad, 'Evaluation of Rapid Partitioning Cleanup Method for Pesticides', *J. Assoc. Off. Anal. Chem.*, **60**, 268–271 (1977).
114. M. Beroza, M.N. Inscoe, M.C. Bowman, 'Distribution of Pesticides in Immiscible Binary Solvent Systems for Cleanup and Identification and its Application in the Extraction of Pesticides from Milk', *Res. Rev.*, **30**, 1–61 (1969).
115. R.J. Argauer, S.J. Lehotay, R.T. Brown, 'Determining Lipophilic Pyrethroids and Chlorinated Hydrocarbons in Fortified Ground Beef using Ion-trap Mass Spectrometry', *J. Agric. Food Chem.*, **45**, 3936–3939 (1997).
116. A. Di Muccio, T. Generali, D. Attard Barbini, P. Pelosi, A. Ausili, F. Vergori, S. Girolimetti, 'Single-step Separation of Organochlorine Pesticide Residues from Fatty Materials by Combined Use of Solid-matrix Partition and C_{18}-cartridges', *J. Chromatogr. A*, **765**, 61–68 (1997).
117. A. Di Muccio, P. Pelosi, D. Attard Barbini, T. Generali, S. Girolimetti, P. Stefanelli, A. Leonelli, G. Amendola, L. Vergori, E. Vìana Fresquet, 'Determination of Pyrethroid Pesticide Residues in Fatty Materials by Solid-matrix Dispersion Partition, Followed by Mini-column Size-exclusion Chromatography', *J. Chromatogr. A*, **833**, 19–34 (1999).
118. A.M. Gillespie, S.L. Daly, D.M. Gilvydis, F. Schneider, S.M. Walters, 'Multicolumn Solid-phase Extraction Cleanup of Organophosphorus and Organochlorine Pesticide Residues in Vegetable Oils and Butterfat', *J. AOAC Int.*, **78**, 431–437 (1995).
119. A.M. Gillespie, S.M. Walters, 'Alumina Blending Technique for Separation of Pesticides from Lipids', *J. Assoc. Off. Anal. Chem.*, **67**, 290–294 (1984).
120. L.D. Johnson, R.H. Waltz, J.P. Ussary, F.E. Kaiser, 'Automated Gel Permeation Chromatographic Cleanup of Animal and Plant Extracts for Pesticide Residue Determination', *J. Assoc. Off. Anal. Chem.*, **59**, 174–187 (1976).
121. M.L. Hopper, 'Automated Gel Permeation System for Rapid Separation of Industrial Chemicals and Organophosphate and Chlorinated Pesticides from Fats', *J. Agric. Food Chem.*, **30**, 1038–1041 (1982).
122. M.M. Krahn, L.K. Moore, R.G. Bogar, C.A. Wigren, S.-L. Chan, D.W. Brown, 'High-performance Liquid Chromatographic Method for Isolating Organic Contaminants from Tissue and Sediment Extracts', *J. Chromatogr.*, **437**, 161–175 (1988).
123. A. Sannino, P. Mambriani, M. Bandini, L. Bolzoni, 'Multiresidue Method for Determination of Organochlorine Insecticides and Polychlorinated Biphenyls Congeners in Fatty Processed Foods', *J. AOAC Int.*, **79**, 1434–1446 (1996).
124. M.P. Seymour, T.M. Jefferies, L.J. Notarianni, 'Large Scale Separation of Lipids from Organochlorine Pesticides and Polychlorinated Biphenyls using a Polymeric High-performance Liquid Chromatographic Column', *Analyst*, **111**, 1203–1205 (1986).
125. D.M. Holstege, D.L. Scharberg, E.R. Richardson, G. Möller, 'Multiresidue Screen for Organophosphorous Insecticides using Gel Permeation Chromatography–Silica Gel Cleanup', *J. Assoc. Off. Anal. Chem.*, **74**, 394–399 (1991).
126. D. Chaput, 'Simplified Multiresidue Method for Liquid Chromatographic Determination of *N*-Methyl Carbamate Insecticides in Fruits and Vegetables', *J. Assoc. Off. Anal. Chem.*, **71**, 542–546 (1988).
127. K. Grob, I. Kaelin, 'Attempt for an On-line Size Exclusion Chromatography–Gas Chromatography Method for Analyzing Pesticide Residues in Foods', *J. Agric. Food Chem.*, **39**, 1950–1953 (1991).
128. E. Neidert, P.W. Saschenbrecker, 'Improved Storherr Tube for Assisted and Sweep Co-distillation Cleanup of Pesticides, Chlorinated Biphenyls, and Pentachlorophenol from Animal Fats', *J. Assoc. Off. Anal. Chem.*, **67**, 773–775 (1984).
129. B.G. Luke, J.C. Richards, E.F. Dawes, 'Recent Advances in Cleanup of Fats by Sweep Co-distillation. Part I: Organochlorine Residues', *J. Assoc. Off. Anal. Chem.*, **67**, 295–298 (1984).
130. A.B. Heath, R. Black, 'Assisted Distillation Cleanup of Pesticide Residues in Animal Fats: New Beadless Distillation Tube', *J. Assoc. Off. Anal. Chem.*, **70**, 862–865 (1987).

131. M. Miyahara, M. Murayama, T. Suzuki, Y. Saito, 'Silica Gel Chromatographic Cleanup Procedure for Organochlorine Pesticide Analysis with Capillary Gas Chromatography', *J. Agric. Food Chem.*, **41**, 221–226 (1993).
132. H. Diserens, 'Simplified Extraction and Cleanup for Multiresidue Determination of Pesticides in Lanolin', *J. Assoc. Off. Anal. Chem.*, **72**, 991–993 (1989).
133. Y. Nakamura, Y. Tonogai, Y. Sekiguchi, Y. Tsumura, N. Nishida, K. Takakura, M. Isechi, K. Yuasa, M. Nakamura, N. Kifune, K. Yamamoto, S. Terasawa, T. Oshima, M. Miyata, K. Kamakura, Y. Ito, 'Multiresidue Analysis of 48 Pesticides in Agricultural Products by Capillary Gas Chromatography', *J. Agric. Food Chem.*, **42**, 2508–2518 (1994).
134. M. Yasin, P.J. Baugh, G.A. Bonwick, D.H. Davies, P. Hancock, M. Leinoudi, 'Analytical Method Development for the Determination of Synthetic Pyrethroid Insecticides in Soil by Gas Chromatography–Mass Spectrometry Operated in Negative-ion Chemical Ionization Mode', *J. Chromatogr. A*, **754**, 225–243 (1996).
135. P.A. Greve, W.B.F. Grevenstuk, 'A Convenient Small-scale Clean-up Method for Extracts of Fatty Samples with Basic Alumina before GLC Analysis of Organochlorine Pesticide Residues', *Medd. Fak. Landb. Gent.*, **40**, 1115–1124 (1975).
136. V. Leoni, 'The Separation of Fifty Pesticides and Related Compounds and Polychlorobiphenyls into Four Groups by Silica Gel Microcolumn Chromatography', *J. Chromatogr.*, **62**, 63–71 (1971).
137. P.G. Baker, P. Bottomley, 'Determination of Residues of Synthetic Pyrethroids in Fruit and Vegetables by Gas–Liquid and High-performance Liquid Chromatography', *Analyst*, **107**, 206–212 (1982).
138. P.R. Haddad, J.G. Brayan, G.J. Sharp, S. Dilli, J.M. Desmarchelier, 'Determination of Pyrethroid Residues on Paddy Rice by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. A*, **461**, 337–346 (1989).
139. A.M. Gillespie, S.M. Walters, 'HPLC Silica Column Fractionation of Pesticides and PCB from Butterfat', *J. Liq. Chromatogr.*, **9**, 2111–2141 (1986).
140. G.R. van der Hoff, A.C. van Beuzekom, U.A. Th Brinkman, R.A. Baumann, P. van Zoonen, 'Determination of Organochlorine Compounds in Fatty Matrices. Application of Rapid Off-line Normal-phase Liquid Chromatographic Clean-up', *J. Chromatogr. A*, **754**, 487–496 (1996).
141. R. Serrano, F.J. López, A. Roig-Navarro, F. Hernández, 'Automated Sample Clean-up and Fractionation of Chlorpyrifos, Chlorpyrifos-methyl and Metabolites in Mussels using Normal-phase Liquid Chromatography', *J. Chromatogr. A*, **778**, 151–160 (1997).
142. R. Looser, K. Ballschmiter, 'Gas Chromatographic Separation of Semivolatile Organohalogen Compounds on the New Stationary Phase Optima δ-3', *J. Chromatogr. A*, **836**, 271–284 (1999).
143. S.N. Tewari, I.C. Sharma, 'Isolation and Determination of Chlorinated Organic Pesticides by Thin-layer Chromatography and the Application to Toxicological Analysis', *J. Chromatogr.*, **131**, 275–284 (1977).
144. S.M. Waliszewski, G.A. Szymczyński, 'Simple Low-cost Method for Determination of Selected Chlorinated Pesticides in Fat Samples', *J. Assoc. Off. Anal. Chem.*, **65**, 677–679 (1982).
145. F. Hernández Hernández, F.J. López Benet, J. Medina Escriche, J.C. Barberá Ubeda, 'Sulfuric Acid Cleanup and KOH–Ethanol Treatment for Confirmation of Organochlorine Pesticides and Polychlorinated Biphenyls: Application to Wastewater Samples', *J. Assoc. Off. Anal. Chem.*, **70**, 727–733 (1987).
146. A. Di Muccio, A. Santilio, R. Dommarco, M. Rizzica, L. Gambetti, A. Ausili, F. Vergori, 'Behaviour of 23 Persistent Organochlorine Compounds During Sulphuric Acid Cleanup on a Solid-matrix Column', *J. Chromatogr.*, **513**, 333–337 (1990).
147. Y.-C. Ling, M.-Y. Chang, I.-P. Huang, 'Matrix Solid-phase Dispersion Extraction and Gas Chromatographic Screening of Polychlorinated Biphenyls in Fish', *J. Chromatogr. A*, **669**, 119–124 (1994).
148. W.P. Cochrane, R.B. Maybury, 'Chemical Confirmation of BHC Isomers: Comparison of Alkaline Reactions in Solution and by Gas Chromatographic Pre-column', *J. Assoc. Off. Anal. Chem.*, **56**, 1324–1328 (1973).
149. T. Suzuki, K. Ishikawa, N. Sato, K.-I. Sakai, 'Determination of Chlorinated Pesticides Residues in Foods. III. Simultaneous Analysis of Chlorinated Pesticides and Phthalate Ester Residues by using $AgNO_3$-coated Florisil Column Chromatography for Cleanup of Various Samples', *J. Assoc. Off. Anal. Chem.*, **62**, 689–694 (1979).
150. T. Suzuki, K. Ishikawa, N. Sato, K.-I. Sakai, 'Determination of Chlorinated Pesticide Residues in Foods. II. Potassium Permanganate Oxidation for Cleanup of Some Vegetable Extracts', *J. Assoc. Off. Anal. Chem.*, **62**, 685–688 (1979).
151. J.H. Hamence, P.S. Hall, D.J. Caverly, 'The Identification and Determination of Chlorinated Pesticides Residues', *Analyst*, **90**, 649–665 (1965).
152. H.L. Crist, R.F. Moseman, J.W. Noneman, 'Rapid Determination and Confirmation of Low Levels of Hexachlorobenzene in Adipose Tissues', *Bull. Environ. Contam. Toxicol.*, **14**, 273–275 (1975).
153. Z.M. Chen, Y.-H. Wang, 'Chromatographic Methods for the Determination of Pyrethrin and Pyrethroid Pesticide Residues in Crops, Foods and Environmental Samples', *J. Chromatogr. A*, **754**, 367–395 (1996).
154. Federal Register, 'Rules and Regulations, Method 608, Organochlorine Pesticides and PCBs', Vol. 29, No. 209, 43321. US-GPO, Washington, DC, 1984.
155. F. Sicbaldi, A. Sarra, D. Muti, P.F. Bo, 'Use of Gas–Liquid Chromatography with Electron-capture and Thermionic-sensitive Detection for the Quantitation and

Identification of Pesticide Residues', *J. Chromatogr. A*, **765**, 13–22 (1997).
156. J.W. Cochran, G.M. Frame, 'Recent Developments in the High-resolution Gas Chromatography of Polychlorinated Biphenyls', *J. Chromatogr. A*, **843**, 323–368 (1999).
157. H.R. Buser, M.D. Müller, 'Enantiomer Separation of Chlordane Components and Metabolites using Chiral High-resolution Gas Chromatography and Detection by Mass Spectrometric Techniques', *Anal. Chem.*, **64**, 3168–3175 (1992).
158. R. Baycan-Keller, M. Oehme, 'Optimization of Tandem Columns for the Isomer and Enantiomer Selective Separation of Toxaphenes', *J. Chromatogr. A*, **837**, 201–210 (1999).
159. P.L. Wylie, K. Uchiyama, 'Improved Gas Chromatographic Analysis of Organophosphorus Pesticides with Pulsed Splitless Injection', *J. AOAC Int.*, **79**, 571–577 (1996).
160. J. Tekel′, Š. Hatrìk, 'Pesticide Residue Analyses in Plant Material by Chromatographic Methods: Clean-up Procedures and Selective Detectors', *J. Chromatogr. A*, **754**, 397–410 (1996).
161. W.E. Wentworth, J. Huang, K. Sun, Y. Zhang, L. Rao, H. Cai, S.D. Stearns, 'Non-radioactive Electron-capture Detector', *J. Chromatogr. A*, **842**, 229–266 (1999).
162. H. Cai, S.D. Stearns, W.E. Wentworth, 'Pulsed Discharge Electron Capture Detector Operating in the Constant-current Mode by Means of Feedback DC Bias Voltage', *Anal. Chem.*, **70**, 3770–3776 (1998).
163. J. Cook, M. Engel, P. Wylie, B. Quimby, 'Multiresidue Screening of Pesticides in Foods using Retention Time Locking, GC–AED, Database Search, and GC/MS Identification', *J. AOAC Int.*, **82**, 313–326 (1999).
164. M.L. Hopper, 'Automated One-step Supercritical Fluid Extraction and Clean-up System for the Analysis of Pesticide Residues in Fatty Matrices', *J. Chromatogr. A*, **840**, 93–105 (1999).
165. S. Cheskis, E. Atar, A. Amirav, 'Pulsed-flame Photometer: A Novel Gas Chromatography Detector', *Anal. Chem.*, **65**, 539–555 (1993).
166. A. Amirav, H. Jing, 'Simultaneous Pulsed Flame Photometric and Mass Spectrometric Detection for Enhanced Pesticide Analysis Capabilities', *J. Chromatogr. A*, **814**, 133–150 (1998).
167. K.-C. Ting, P. Kho, 'GC/MIP/AED Method for Pesticide Residue Determination in Fruits and Vegetables', *J. Assoc. Off. Anal Chem.*, **74**, 991–998 (1991).
168. M. Miyahara, T. Suzaki, Y. Saito, 'Multiresidue Method for Some Pesticides in Lanolin by Capillary Gas Chromatography with Detection by Electron Capture, Flame Photometric, Mass Spectrometric, and Atomic Emission Techniques', *J. Agric. Food Chem.*, **40**, 64–69 (1992).
169. Y. Liu, V. Lopez-Avila, M. Alcaraz, W.F. Beckert, 'Capillary Gas Chromatography–Atomic Emission Detection Method for the Determination of Pentylated Organotin Compounds: Interlaboratory Study', *J. Assoc. Off. Anal. Chem.*, **78**, 1275–1285 (1995).
170. G.S. Durrell, T.C. Sauer, 'Simultaneous Dual-column, Dual-detector Gas Chromatographic Determination of Chlorinated Pesticides and Polychlorinated biphenyls in Environmental Samples', *Anal. Chem.*, **62**, 1867–1871 (1990).
171. M.L. Hopper, 'Analysis of Organochlorine Pesticide Residue using Simultaneous Injection of Two Capillary Columns with Electron Capture and Electrolytic Conductivity Detectors', *J. Assoc. Off. Anal. Chem.*, **74**, 974–981 (1991).
172. H.-J. Stan, D. Mrowetz, 'Residue Analysis of Organophosphorus Pesticides in Food with Two-dimensional Gas Chromatography using Capillary Columns and Flame Photometric Detection', *J. High Res. Chromatogr. Chromatogr. Commun.*, **6**, 255–263 (1983).
173. H.-J. Stan, 'Two Dimensional Capillary Gas Chromatography of Residue Analysis', in *Emerging Strategies for Pesticide Analysis*, eds. J. Sherma, T. Cairns, CRC Press, Boca Raton, 175–211, 1992.
174. Z. Liu, S.R. Sirimanne, D.G. Patterson, Jr, L.L. Needham, J.B. Phillips, 'Comprehensive Two-dimensional Gas Chromatography for the Fast Separation and Determination of Pesticides Extracted from Human Serum', *Anal. Chem.*, **66**, 3086–3092 (1994).
175. D. Barcelò, 'Hyphenated Methods for Pesticide Residue Analysis', in *Emerging Strategies for Pesticide Analysis*, eds. J. Sherma, T. Cairns, CRC Press, Boca Raton, 273–294, 1992.
176. D.R. Erney, A.M. Gillespie, D.M. Gilvydis, C.F. Pode, 'Explanation of the Matrix-induced Chromatographic Response Enhancement of Organophosphorus Pesticides during Open Tubular Column Gas Chromatography with Splitless or Hot On-column Injection and Flame Photometric Detection', *J. Chromatogr.*, **638**, 57–63 (1993).
177. P.T. Holland, 'Mass Spectrometric Determination of Pesticide Residues', *Pure Appl. Chem.*, **62**, 317–336 (1990).
178. D.L. Heikes, 'Residues of Pentachloronitrobenzene and Related Compounds in Peanut Butter', *Bull. Environm. Contam. Toxicol.*, **24**, 338–343 (1980).
179. T. Cairns, E.G. Siegmund, F. Krick, 'Identification of Several New Metabolites from Pentachloronitrobenzene by Gas Chromatography/Mass Spectrometry', *J. Agric. Food Chem.*, **35**, 433–439 (1987).
180. T.J. Class, 'Environmental Analysis of Cypermethrin and its Degradation Products after Forestry Applications', *Int. J. Environ. Anal. Chem.*, **49**, 189–205 (1992).
181. J. Hong, Y. Eo, J. Rhee, T. Kim, K. Kim, 'Simultaneous Analysis of 25 Pesticides in Crops using Gas Chromatography and their Identification by Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **639**, 261–271 (1993).

182. R. Chaler, R. Vilanova, M. Santiago-Silva, P. Fernandez, J.O. Grimalt, 'Enhanced Sensitivity in the Analysis of Trace Organochlorine Compounds by Negative-ion Mass Spectrometry with Ammonia as Reagent Gas', *J. Chromatogr. A*, **823**, 73–79 (1998).
183. L.G.M. Th. Tuinstra, P. van de Spreng, P. Gaikhorst, 'Ion Trap Detection for Development of a Multi Residue/Multi Matrix Method for Pesticide Residues in Agricultural Products', *Int. J. Environ. Anal. Chem.*, **58**, 81–91 (1995).
184. G.C. Mattern, G.M. Singer, J. Louis, M. Robson, J.D. Rosen, 'Determination of Several Pesticides with a Chemical Ionization Ion Trap Detector', *J. Agric. Food. Chem.*, **38**, 402–407 (1990).
185. H.M. Chan, J. Zhu, F. Yeboah, 'Determination of Toxaphene in Biological Samples using High Resolution GC Coupled with Ion Trap MS/MS', *Chemosphere*, **36**, 2135–2148 (1998).
186. S. Schachterle, C. Feigel, 'Pesticide Residue Analysis in Fresh Produce by Gas Chromatography–Tandem Mass Spectrometry', *J. Chromatogr. A*, **754**, 411–422 (1996).
187. W.C. Brumley, E. Latorre, V. Kelliher, A. Marcus, D.E. Knowles, 'Determination of Chlordane in Soil by LC/GC/ECD and LC/GC/EC–NIMS with Comparison of ASE, SFE, and Soxhlet Extraction', *J. Liq. Chrom. Rel. Technol.*, **21**, 1199–1216 (1998).
188. K.S. Kalasinsky, 'Pesticide Determination by GC/FT-IR Spectrometry', in *Analytical Methods for Pesticides and Plant Growth Regulators*, eds. G. Zwerg, J. Sherma, Academic Press, New York, 101–117, Vol. 17, 1989.
189. E. Papadopoulou-Morkidou, 'Recent Advances in Pyrethroid Determination', in *Analytical Methods for Pesticides and Plant Growth Regulators*, eds. G. Zweig, J. Sherma, Academic Press, New York, 179–206, Vol. 16, 1988.
190. R.J. Vreeken, W.D. van Dongen, R.T. Ghijsen, U.A.Th. Brinkman, 'The Use of Mobile Phase Additives in the Determination of 55 (Polar) Pesticides by Column Liquid Chromatography–Thermospray–Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **54**, 119–145 (1994).
191. J. Slobodník, A.C. Hogenboom, J.J. Vreuls, J.A. Rontree, B.L.M. van Blaar, W.M.A. Niessen, U.A.Th. Brinkman, 'Trace Level Determination of Pesticide Residues using On-line Solid-phase Extraction–Column Liquid Chromatography with Atmospheric Pressure Ionization Mass Spectrometric and Tandem Mass Spectrometric Detection', *J. Chromatogr. A*, **741**, 59–74 (1996).
192. M. Careri, A. Mangia, M. Musci, 'Applications of Liquid Chromatography–Mass Spectrometry Interfacing Systems in Food Analysis: Pesticide, Drug and Toxic Substance Residues', *J. Chromatogr. A*, **727**, 153–184 (1996).
193. A.C. Hogenboom, J. Slobodník, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A.Th. Brinkman, 'Single Short-column Liquid Chromatography with Atmospheric Pressure Chemical Ionization–(Tandem) Mass Spectrometric Detection for Trace Environmental Analysis', *Chromatographia*, **42**, 506–514 (1996).
194. M.D. Osselton, R.D. Snelling, 'Chromatographic Identification of Pesticides', *J. Chromatogr.*, **368**, 265–271 (1986).
195. A. Ambrus, E. Argitai, G. Karoly, A. Fulop, J. Lantós, 'General Method for Determination of Pesticide Residues in Samples of Plant Origin, Soil, and Water. II. Thin Layer Chromatographic Determination', *J. Assoc. Off. Anal Chem.*, **64**, 743–748 (1981).
196. U. de la Vigne, D.E. Jaenchen, W.H. Weber, 'Application of High-performance Thin-layer Chromatography and Automated Multiple Development for the Identification and Determination of Pesticides in Water', *J. Chromatogr.*, **553**, 489–496 (1991).
197. J. Sherma, 'Recent Advances in Thin-layer Chromatography of Pesticides', *J. AOAC Int.*, **82**, 48–53 (1999).
198. R.J. van de Nesse, G.J.M. Hoogland, J.J.M. De Moel, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, 'On-line Coupling of Liquid Chromatography to Thin-layer Chromatography for the Identification of Polycyclic Aromatic Hydrocarbons in Marine Sediment by Fluorescence Excitation and Emission Spectroscopy', *J. Chromatogr.*, **552**, 613–623 (1991).

Organophosphorus Pesticides in Water and Food, Analysis of

Sílvia Lacorte and Damià Barceló
CID–CSIC, Barcelona, Spain

1 Introduction 1
1.1 Sources and Uses of Organophosphorus Pesticides 1
1.2 Distribution and Fate of Pesticides 2
1.3 Physical and Chemical Properties 2
1.4 Regulations 4
2 Outline of the Analytical Method 5
3 Water Sample Handling 6
3.1 Liquid–Liquid Extraction 6
3.2 Solid-phase Extraction 6
3.3 Types of Sorbent Materials 9
3.4 Parameters Involved in Solid-phase Extraction 10
4 Collection and Preparation of Food Samples 11
4.1 Collection of Food Samples 11
4.2 Analytical Preparation of Food Samples 11
5 Gas Chromatographic Determination 13
5.1 Gas Chromatographic Detectors 13
5.2 Gas Chromatography/Mass Spectrometry with Electron Ionization 13
5.3 Gas Chromatography/Mass Spectrometry with Chemical Ionization 14
6 High-performance Liquid Chromatography 15
6.1 Liquid Chromatography with Diode-array Detection 16
6.2 Liquid Chromatography and Mass Spectrometry with Thermospray Interface 16
6.3 Liquid Chromatography and Mass Spectrometry with Electrospray/Ionspray Interfaces 18
6.4 Liquid Chromatography and Mass Spectrometry with Atmospheric Pressure Chemical Ionization 19
7 Validation of Analytical Methods 22
8 Conclusions 22
Abbreviations and Acronyms 23
Related Articles 24
References 25

The purpose of this article is to provide an overview of the main characteristics of organophosphorus pesticides (OPPs), such as their physicochemical characteristics, sources, fate and distribution in the environment. It covers regulatory compliance from the Commission of the European Communities (CEC) and the United States National Pesticide Survey (USNPS), regarding the maximum permissible levels of the most toxic and persistent OPPs and their transformation products (TPs) in water. This article also describes the analytical methods currently used to monitor pesticides and their degradation products in water and food matrices. A major part concerns sample preparation techniques and includes traditional liquid–liquid extraction (LLE) and novel techniques based in automated solid-phase extraction (SPE), either on-line or off-line. New sorbent materials, their applications and the parameters involved in the extraction are described. Chromatographic methods used for their determination are discussed in detail. Much emphasis has been given to gas chromatography (GC) and liquid chromatography (LC) coupled to mass spectrometry (MS) as a confirmation technique and mass spectral information on several pesticides has been included. The final section is devoted to method validation, with the aim of providing a basis and a framework of information for quality assurance (QA) and interlaboratory calibration studies.

1 INTRODUCTION

1.1 Sources and Uses of Organophosphorus Pesticides

Much concern is being given to the determination of pesticides in the different environmental matrices, owing to the increased number of pesticides detected and to the severe rules imposed by legislation aimed at protecting natural resources.[1,2] OPPs were introduced on to the market in the period 1945–55, to replace the toxic and persistent organochlorine insecticides, mainly DDT. They represent a family of about 140 different compounds, which are classified in different groups depending on their chemical structure. They are esters, amides or thiols which in general derive from phosphoric or phosphonic acid, where one or more atoms of hydrogen are substituted by organic groups. Depending on the functional groups, OPPs are divided into six groups or families: phosphates, phosphonates, phosphorothionates, phosphorodithioates, phosphorothiolate and phosphoroamidate. Table 1 summarizes the chemical structure

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Chemical structures of different families (groups) of OPPs

Group	Structure[a]	Examples
Phosphate	$(R{-}O)_2\overset{\overset{O}{\|}}{P}{-}O{-}X$	Chlorfenvinfos, dichlorvos, mevinphos
Phosphonate	$(R{-}O)_2\overset{\overset{O}{\|}}{P}{-}H{-}X$	Trichlorfon
Phosphorothioate	$(R{-}O)_2\overset{\overset{S}{\|}}{P}{-}O{-}X$	Chlorpirifos, coumaphos, diazinon, fenitrothion, fenthion, parathion, temephos
Phosphorothiolate	$(R{-}O)_2\overset{\overset{O}{\|}}{P}{-}S{-}X$	Demeton-*S*-methyl
Phosphorodithioate	$(R{-}O)_2\overset{\overset{S}{\|}}{P}{-}S{-}X$	Azinphos, disulfoton, malathion, fosmet
Phosphoramidate	$X/R{-}O{-}\overset{\overset{O}{\|}}{P}{-}NH{-}R$	Fenamiphos

[a] R = alkyl or aryl groups; X = aliphatic, aromatic or heterocyclic groups.

of each group. Because the P=S bond is the most stable, many OPPs are synthesized with such a configuration. The characteristics of each group of OPPs can vary substantially, and they confer the specificity to react against one type of organism or another. In general terms, the mode of action of OPPs is based on the inhibition of the enzyme acetylcholine esterase of the organisms, which blocks the nervous transmission and produces death normally by respiratory depression. Owing to this specificity and because they are rapidly degraded after application and are rarely accumulated in the trophic chain, OPPs are extensively used to combat different pests. The main uses are in agricultural crop production, industrial activities and in the tertiary sector (domestic use, mosquito and rodent control, greenhouses, cemeteries, aquaculture, golf courses, etc.). In such practices, OPPs are applied as formulations to enhance their effectiveness upon application, absorption, translocation, immobilization and detoxification. Formulations contain additives and adjuvants such as mixtures of surfactants, mineral and vegetable oils, emulsifiers and salts.

1.2 Distribution and Fate of Pesticides

The distribution and fate of pesticides in the environment depends on whether the pesticide is locally applied or is thoroughly spread over large crop field extensions. Local applications are normally performed by manual spraying or from tractor-trailers, and represent the utmost controlled and optimal conditions since only target organisms are eliminated and pesticide spreading and dissipation are limited. Helicopters or aeroplanes are used to spray large extensions, and dispersion and especially volatilization often occur; as a consequence, pesticides can be found in the atmosphere and in areas far away from the focus treatment point. Once the pesticides have been introduced onto the water or soil surface, they can be transported from the point-source by drift, runoff or soil migration. Moreover, their concentration starts to decrease as a consequence of dispersion, volatilization, degradation and lixivation, among others. The importance of each process greatly depends on the physicochemical properties of each compound, and also on the water and soil properties, the environmental conditions and the applied concentration and mode of application. During the time a pesticide remains in the environment, chemical (mainly photolysis and hydrolysis) and biological degradation can occur, with the subsequent formation of a TP[3] such as oxo derivatives, sulfones and sulfoxides. Moreover, chloro- and nitrophenols have been reported as the main degradation products of many OPPs.[4,5] Parent pesticides can remain in the soil or sediment whereas TPs, which are more polar than the parent compounds, have a higher capacity to lixiviate and migrate through soil and sediment. As a result, TPs have been encountered in places far away from the point of application and contaminating groundwaters.[6] Groundwater pollution represents a serious environmental problem since in many places these waters are used for drinking and irrigation purposes. In addition, it is known that many TPs are more toxic than their parent compounds[7] so that they can produce deleterious effects towards flora and fauna of surface and groundwaters and in sediment organisms.

It is impossible to define a unique fate of pesticides and TPs in the environment since distribution, transport, degradation, dispersion, soil retention and lixiviation will depend on the environmental conditions and on the physicochemical properties of each pesticide. Hence it is clear that to assess the environmental impact of pesticides and their TPs, one should consider all environmental matrices (surface waters, groundwater, soil, vegetables and fruits, atmosphere, etc.). The following section is intended to summarize the most important physicochemical parameters of OPPs and the role they play under environmental conditions.

1.3 Physical and Chemical Properties

OPPs have characteristic chemical structures which confer different physicochemical properties. For instance,

polarity, water solubility, vapor pressure, water–octanol partition coefficient, half-life and acid–base ionization are the main aspects that will influence the persistence or degradation of each OPP in the environment. These physicochemical parameters are generally measured under controlled laboratory conditions, and under similar conditions comparable values are obtained in the literature. Some of these parameters, such as half-life or Ground ubiquity score (mobility index) (GUS), to mention the most important, depend considerably on the specific conditions set up for the experiment and can be measured under either laboratory or environmental conditions. Experiments performed under environmental conditions provide very realistic data which permit one to determine the main degradation processes that occur in natural conditions. However, since they are affected by the temperature, humidity, soil characteristics, water pH, etc., very disparate data are obtained from different authors. *The Pesticide Manual*[8] and a book on trace determination of pesticides in water[9] list several physicochemical parameters of OPPs. In evaluating the behavior of pesticides in the environment, one should consider the physicochemical properties of each specific pesticide with the objective of establishing the main pathways of the pesticides and the routes of degradation. On the other hand, knowledge of the physicochemical properties of each compound is useful in defining the analytical method for its determination.

Water solubility, defined as the concentration of a chemical dissolved in water, is a parameter that both affects the extraction of pesticides from water, food and soil samples and influences the fate of pesticides in the environment. The water solubilities of OPPs are very variable, with compounds which are almost completely soluble, such as mevinphos ($600\,g\,L^{-1}$), and others which are totally insoluble, such as temephos ($0.00003\,g\,L^{-1}$). In general, distant water solubilities affect the extraction procedure, since the most soluble ones (water solubilities above the grams per liter level) will be difficult to trap and the most insoluble (water solubilities lower than $0.5–1\,mg\,L^{-1}$) will easily remain adsorbed upon the glassware and other laboratory material, which leads to lower recovery values.

After application of an OPP, this parameter is essential for determining whether a specific compound will remain dissolved in water, is a potential leacher, will be adsorbed on sediment particles or will precipitate at the surface soil. In general terms, it gives a clue to the distribution of a pesticide in the environment. However, this parameter cannot be used solely to predict the distribution of OPPs in the environment.

A parameter related to the water solubility is the water–octanol partition coefficient (reported as a logarithm, usually $\log K_{ow}$ or $\log P_{ow}$), which is an indicator of the degree of hydrophobicity of a pesticide and its tendency to accumulate in living cells or organisms or to be adsorbed on the organic matter of soils and sediments (humic and fulvic substances). As a rule of thumb, pesticides are classified as apolar if $\log K_{ow}$ values are above 4–5 and polar if they are below 1.5. Between these values, they are considered as ''moderately polar''. OPPs with a high $\log K_{ow}$ value (above 4) have a high bioaccumulation potential and therefore they are compounds that if applied at high concentrations can be transferred along the trophic chain. In addition, the soil organic carbon sorption coefficient (K_{oc}), expressed as $cm^3\,g^{-1}$, is used as an indicator of the degree of attraction and retention of the pesticides in the particles of the soil. The retention will basically depend on the type of soil, and will affect the behavior of the OPP in the field, since an adsorbed molecule can be difficult to solubilize by water, and its availability will decrease. This parameter will also indicate the leaching potential of a pesticide. In general, pesticides with K_{oc} values below 50 are highly mobile, values between 150 and 500 represent moderately mobile compounds and values higher than 2000 are slightly mobile compounds. The K_{oc} values of OPPs are very variable, and vary from 44 for mevinphos to 498 for chlorpyriphos.

The solubility, mobility and volatilization of OPPs in the environment will largely depend on their acid–base ionization constants (pK_a) and whether they are ionized at the pH of each compartment. Volatilization depends on the vapor pressure of each pesticide and on the climatic conditions where the pesticide has been applied, mainly the ambient temperature. OPPs are medium-volatility compounds, with vapor pressures which vary between 0.13 and 7000 mPa. The most volatile compounds will be lost to the atmosphere, especially during application. However, once the OPPs have settled upon the aquatic medium, volatilization will depend on their solubility in water. The Henry's law constant, denoted H or K_H (expressed in $Pa\,m^3\,mol^{-1}$), is defined as the coefficient between the vapor pressure of a pesticide and its solubility in water, and indicates the volatilization of a compound in solution. To give an example, chlorpyrifos, with a low water solubility ($0.0014\,g\,L^{-1}$) will be easily lost to the atmosphere. It is generally accepted that compounds with H values below $10^{-5}\,Pa\,m^3\,mol^{-1}$ will have little tendency to volatilize.

A common way to evaluate the persistence of a pesticides in the environment is the half-life ($t_{1/2}$). The half-life is defined as the time required for a pesticide to reduce the initial concentration by 50%, and it is calculated from the first-order Equation (1):

$$t_{1/2} = \frac{\ln 2}{k} \tag{1}$$

where k is the time constant. Half-lives can be calculated under controlled laboratory conditions using Milli-Q water and artificial radiation, simulating environmental conditions,[10] or under real environmental conditions. The last case represents the most realistic approach since many different factors and processes play a role in pesticide degradation and dissipation.[6,11] Generally, one takes advantage of phytosanitary treatments to calculate the half-life of a pesticide after it has been applied. The half-lives of OPPs in the environment vary from a few hours to several days, but this value depends on the intrinsic characteristics of each pesticide and on the experimental conditions (climatic conditions, microbial activity, soil and water type, etc.).

Another parameter of environmental interest is the mobility index or GUS, which provides a leaching classification depending on the soil half-life and the K_{oc} value Equation (2):

$$\mathrm{GUS} = \log T_{0.5}(4 - \log K_{oc}) \qquad (2)$$

where $T_{0.5}$ is the soil half-life of the pesticide. A GUS higher than 2.8 indicates that the pesticide can be considered as a leacher. Values lower than 1.8 indicate nonleacher pesticides.

In general terms, it can be concluded that the stability of OPPs in the environment depend on several parameters, many of them correlated, in such a way that it is difficult to define a unique parameter as responsible for the fate of OPPs in the environment.

1.4 Regulations

Nowadays, hundreds of different OPPs are being used for different purposes. Since 1970, authorities have started to appreciate the danger of water pollution and the need to check the water quality to protect natural drinking sources. Strict directives have been set up and monitoring the concentration levels of OPPs and their TPs in water and other environmental matrices is compulsory. The European Union (EU) has elaborated a ''Black List'' which includes the 132 most toxic substances for the environment.[12] Among other families of compounds, 21 toxic, persistent and widely used OPPs are included. These are designated as priority OPPs and are listed in Table 2. Effort is directed at determining these compounds in monitoring programs.

Moreover, the EU Directive on the Quality of Water Intended for Human Consumption (CEC 76/464/EEC) has elucidated maximum permissible levels of pesticides in drinking water which should be below $0.1\,\mu g\,L^{-1}$ for each individual pesticide and that the sum of all pesticides should be below $0.5\,\mu g\,L^{-1}$.[2] In surface water, the concentration of pesticides should be between 1 and $10\,\mu g\,L^{-1}$. Moreover, the USNPS has published a list which includes all pesticides and their TPs, mainly sulfoxides, sulfones and oxo derivatives of OPPs, which were detected in US groundwaters and are considered polar, toxic and persistent.[13,14] An overview on the use, consumption and regulatory activities and actions of both the EU and USNPS has been described by Barceló.[15] Furthermore, a recent EU Directive appeared in 1996[16] that expands the number of contaminants to be monitored in surface and waste effluent waters. The objective is to determine the potential toxicity and environmental impact of newly identified substances (e.g. TPs of pesticides and very polar compounds, which are difficult to analyze using conventional techniques). As a result, most countries control the levels of pesticides in different types of waters with the aim of providing ways and solutions to reduce the level of pollutants in surface and groundwaters. To accomplish such a purpose, it is necessary to use analytical methods which permit the

Table 2 Pesticides included in the ''Black List'' of the 132 most toxic substances according to the EU, directive 76/464 on pollution of the aquatic environment[12]

Aldrine (2)	Disulfoton (1)	Monolinuron (4)
Atrazine (5)	Endosulfan (1)	Ometoate (1)
Azinphos-ethyl (1)	Endrin (2)	Oxidemeton-methyl (1)
Azinphos-methyl (1)	Fenitrothion (1)	Parathion-ethyl (1)
Chlordane (2)	Fenthion (1)	Parathion-methyl (1)
Coumaphos (1)	Heptachlor (2)	Foxim (1)
2,4-D (3)	Hexachlorbenzene (2)	Propanil (4)
DDT (2)	Linuron (4)	Pyrazon (1)(2)
Demeton (1)	Malathion (1)	Simazine (6)
Dichlorprop (3)	MCPA (3)	2,4,5-T (3)
Dichlorvos (1)	Mecoprop (3)	Triazophos (1)
Dieldrin (2)	Metamidophos (1)	Trichlorfon (1)
Dimethoate (1)	Mevinphos (1)	Trifluralin (6)

Pesticide types: (1) Organophosphorus, (2) organochlorine, (3) phenoxyacetic acid, (4) amide, urea, (5) triazine, (6) toluidine.

unequivocal identification and confirmation of pesticides along with their TPs in the environment in a quick and reliable way.

This article describes first the different techniques used for the extraction of pesticides in water and food samples, showing the advantages and disadvantages of each, and discussing the parameters which influence the extraction step. Subsequently, an overview of the most common detection techniques used for OPPs is given. This will permit one to evaluate which is the preferred technique according to the chemical nature of each compound, in which cases it can be used and the information that one will gather. The final section consists of a review of the different validation steps that have to be performed to evaluate any analytical technique. This is specially relevant since all techniques have to be validated for qualitative parameters before they can be routinely applied for monitoring purposes.

2 OUTLINE OF THE ANALYTICAL METHOD

In general terms, determinations of OPPs include four main steps: (i) sampling, (ii) extraction and cleanup of the analytes from the sample, (iii) concentration of the analytes and (iv) detection. The sampling step has to be performed so that the sample is a representative part of a whole area, recovery is maximized and sample contamination is avoided. Barceló and Hennion[17] have reviewed different sampling strategies, which depend on the type of water to be analyzed and on the sample location.

Many different extraction procedures have been described to isolate pesticides from water. Dichloromethane LLE methods have been reported in the literature, covering several groups of pesticides and being also recommended by the United States Environmental Protection Agency (USEPA) through its National Pesticide Survey.[18,19] As an alternative to LLE, SPE procedures, also known as liquid–solid extraction (LSE), have been implemented in the last few years, and can be performed both off-line and on-line. SPE methods can be easily converted into fully automated on-line systems coupled to high-performance liquid chromatography (HPLC), in brief LC or GC. Such systems, also referred to as precolumn technology, show additional advantages such as lower detection limits (DLs) (for a given amount of water volume preconcentrated) and that sample manipulation is minimized, which is translated to no evaporation losses, no contamination, no need to concentrate the sample, easy automation and use of small amounts of toxic solvents. There is a great availability of sorbents and configurations which permit the efficient trapping of compounds of diverse polarities with the percolation of only 50–100 mL of water. Recently, besides the well-known C_{18} and polymeric sorbents, immunoaffinity precolumns have appeared as an alternative to preconcentrate selectively the analytes of interest.[20] To isolate pesticides from food samples, Soxhlet extraction followed by a cleanup step is the most common approach.

As regards to the analysis of pesticides belonging to different classes, the traditional detection technique has been GC using a selective detector such as the nitrogen-phosphorus detector (NPD)[21] or flame photometric detection (FPD),[22] among others. Classically, identification of the analytes is done through retention time comparison with a standard solution. Errors and lack of reproducibility can often appear owing to variations in the response of the detector and to retention time shifts. This is specially relevant when GC/NPD or gas chromatography/flame ionization detection (GC/FID) is used, and therefore duplicate analysis using two columns of different polarity, or otherwise the addition of internal standards (added at the beginning of the analysis) or surrogates (preferably a labeled standard added prior to injection), is needed in order to meet quality control (QC) requirements.[19] The use of GC with MS detection is a valid alternative to confirm the results obtained by other techniques. However, GC-based techniques are not suitable for the analysis of polar, non-volatile and thermolabile compounds, since for these compounds a derivatization step prior to GC analysis is compulsory. The use of two different detection techniques (or the more expensive MS detector) and the inability to analyze both the polar TPs along with the parental pesticides limit the applicability of GC-based techniques for OPPs' analysis.

LC techniques are preferred for the analysis of organic pollutants since they cover a wide range of polarities and permit the identification of pesticides along with their TPs. Liquid chromatography with diode-array detection (LC/DAD) has successfully been used to monitor OPPs.[23] The main advantage of LC/DAD is related to its easy use and to the fact that it offers an absorbance spectra that can be used to identify pesticides through spectral comparison. However, limitations arise for nonultraviolet (UV)-absorbing compounds and because of the impossibility of identifying unknown pesticides (e.g. TPs) in real water samples. Moreover, typical problems related to the analysis of surface waters with LC/DAD are the co-elution of two or more compounds and the presence of interferences, mainly due to humic and fulvic material. To overcome these problems, liquid chromatography with mass spectrometric detection (LC/MS) is used. The higher selectivity and the possibility of obtaining spectral information, together with sufficient sensitivity to use it for quantification purposes, are the main advantages

of this technique. LC/MS with a thermospray (TSP) interface[24–26] has probably been the most widely used technique for pesticide monitoring. In general, it produces molecular information and scarce fragmentation. More structural information can be obtained by using filament-on or discharge ionization. LC/MS with a particle beam (PB) interface[27] has also been extensively applied since it offers abundant fragmentation and can be compared with electron ionization (EI) mass spectra. The high DLs and poor reproducibility are the main problems related to this technique. The advent of atmospheric pressure ionization (API) LC/MS interfaces has overcome the disadvantages of GC/MS and other LC/MS interfacing devices such as TSP, because they can provide structural information similar to chemical ionization (CI) and very good sensitivity. Electrospray (ESP), ionspray (ISP) and atmospheric pressure chemical ionization (APCI) have been applied for pesticide monitoring[28–30] and a DL of few nanograms per liter and a linear response range over two orders of magnitude have been reported.

3 WATER SAMPLE HANDLING

After application, pesticides are generally found in the environment at low concentrations, generally around parts per billion or trillion (ppb–ppt level respectively). The trace analysis of pesticides in drinking, surface and wastewaters require the preconcentration of pesticides from the water matrix prior to the chromatographic analysis. The preconcentration step, also referred to as sample pretreatment, involves not only the extraction of pesticides from the water matrix but also the isolation or sample cleanup and generally account for the major part of sample analysis. Traditionally the preconcentration of pesticides was carried out by LLE but nowadays there is a general trend to change to SPE procedures. Both approaches will be discussed below.

3.1 Liquid–Liquid Extraction

LLE is in many cases the sample preparation technique preferred by the environmental analyst because of its extensive implementation in official methods. LLE is based on the partitioning of different components between aqueous solution and an immiscible organic solvent. Most of the current USEPA methodologies for the analysis of priority pollutants, including pesticides, in water samples involve extraction with various solvents such as dichloromethane, either in acidic conditions to avoid hydrolysis of pesticides[31] or at neutral pH,[32] with butyl acetate,[33] chloroform[34] and hexane.[35] Other approaches have been described using ion pair-forming agents by adjusting the pH to a basic value and adding quaternary ammonium salts.[36] The selectivity of LLE depends on the solvent used, extraction pH, ionic strength, water to solvent ratio, number of extractions and the chemical nature and concentration of the analytes. However, LLE has some drawbacks, e.g. it requires large amounts of generally toxic and inflammable organic solvents, emulsions can be generated and further cleanup is often necessary. The procedure is tedious and it involves considerable sample manipulation, which may lead to losses during the concentration step and result in many cases in poor recoveries and bad reproducibility, with the risk of sample contamination. In addition, evaporative losses may occur during removal of the large solvent volumes, thus making the analysis inaccurate and scarcely reproducible. Moreover, its automation requires the use of expensive robots. Hence there is a general trend to change current LLE procedures to SPE protocols.

3.2 Solid-phase Extraction

SPE has been successfully applied to extract and store pesticides from environmental water matrices and as a cleanup procedure for sediments and food samples.[37] The procedure is based on the use of an adsorbent material which is capable of selectively retaining the pesticides from the matrix. Briefly, the mode of action consists in percolating the sample through the sorbent where the pesticides are retained; subsequently, the pesticides are selectively eluted with an organic solvent. Nowadays there exist a wide variety of SPE methods, which may be divided into off-line and on-line (or precolumn technology) methods. The chemistry and principles of these two approaches are identical. The method involves different steps: (i) washing the sorbent with an organic solvent, often the same solvent as will be used for the elution step; (ii) activation of the sorbent (wetting) with a water-miscible solvent (normally methanol or acetonitrile); (iii) removal of the excess of solvent with pure water; (iv) percolation of the water sample; (v) removal of interferences (cleanup) by flowing water or solvent through the sorbent bed at a high flow rate; and (vi) elution of trapped analytes with a selective solvent, normally compatible with further chromatographic detection system. The different SPE procedures will be described in detail.

3.2.1 Off-line Solid-phase Extraction Methods

In off-line SPE methods, the water sample is percolated through the sorbent bed which is packed in a disposable cartridge or enmeshed in a membrane extraction disk. Samples can move through the sorbent bed by gravity or by application of a positive (syringe) or negative (vacuum manifold) pressure. In general, off-line practices

are well accepted owing to the large variety of packing materials available, C_8, C_{18}, polymeric phases and graphitized carbon black (GCB) being the most common, with different sizes and volumes, which allows great operational flexibility. In the cartridge format, 30 mg–2 g or more of sorbent (normally of 40–150-μm particle size) are usually packed in a polypropylene, Teflon or glass syringe to avoid sample contamination. The off-line procedure using cartridges is illustrated in Figure 1 and consists in conditioning the sorbent, application of the sample and elution. When cartridges are used, the sample normally moves at flow rates of 5–10 mL min^{-1} to avoid the formation of preferential channels in the sorbent bed that could reduce the adsorption of the analytes towards the sorbent. After sample application, the sorbent is washed with water or solvent to remove impurities, then dried for 5–30 min to displace the excess of water or solvent. Depending on the application, especially if GC analysis is performed, complete drying of the sorbent (under vacuum or with nitrogen) is required before proceeding to the elution step. This drying step may enhance losses of more polar and volatile compounds. The trapped analytes are eluted with 5–20 mL of solvent, generally methanol, acetonitrile, dichloromethane, ethyl acetate, hexane, etc., or mixtures of those, which are applied in two or more steps. A common practice to enhance recoveries is to let the elution solvent soak the sorbent for several minutes before elution. After elution, the excess of elution solvent is evaporated either by rotary evaporation or under a gentle stream of nitrogen to a final extract volume of 0.1–1 mL, depending on the initial concentration of pesticides. A mild evaporation process is usually recommended and dryness should be avoided to prevent important analyte losses. From this extract, an aliquot is injected into the chromatograph.

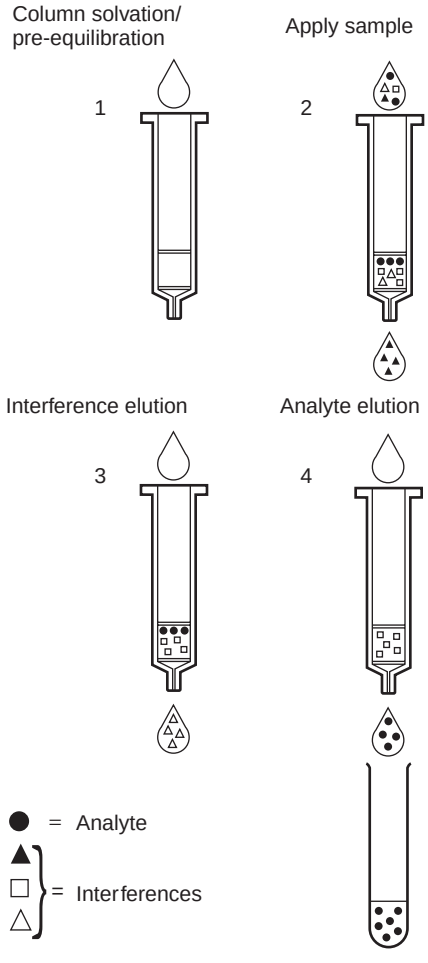

Figure 1 Off-line SPE process using SPE cartridges. The procedure includes (1) conditioning of the precolumn, (2) sample loading or preconcentration step, (3) elimination of the interferences and (4) elution of the analytes. (Reprinted from 'Handbook of HPLC' eds. E. Katz, R. Eksteen, P. Schoenmakez, N. Miller, with permission of Marcel Dekker Inc.)

Application of SPE cartridges for OPPs extraction include the use of Sep-Pak C_8 and C_{18}.[38,39] XAD resins have been used for extraction and stabilization of pesticides from water.[40] However, this sorbent needs an extensive cleanup prior to the preconcentration step.[41]

The main advantages of using cartridges is that manipulation, solvent disposal and analysis time are reduced in comparison with conventional LLE techniques. Percolation of water samples can be done manually, but nowadays automated devices such as the ASPEC XL from Gilson (France) or the Zymark system (USA) permit the automation of the technique, making it useful for routine analysis. In environmental applications, it is interesting to obtain high enrichment factors so that one can achieve high sensitivity. This can be accomplished with off-line SPE, since there is a great choice of SPE sorbents, with different capacities, and therefore high sample volumes can be percolated. Prior to SPE, surface water samples have to be filtered through 0.45 μm filters to remove sample particles and to avoid clogging of the SPE cartridges. However, filtration may result in poor recoveries of the most apolar compounds, such as temephos, trifluralin and others, which, being preferentially adsorbed on particles, are removed from the aqueous sample together with the particles. In SPE, the extraction efficiency depends on the type of water matrix concerned. In general, drinking or groundwaters produce better recoveries and DLs, since the amounts of interferences are very low. In contrast, when preconcentrating surface or wastewaters, impurities are also preconcentrated, so that worse quality parameters are achieved.

An alternative to off-line SPE with cartridges is the use of Empore extraction disks and Speedisks (Baker, The Netherlands), where the sorbent particles are meshed in Teflon fibrils to form strong sheets or

membranes which can be used in standard filtration devices. Since the channeling phenomenon does not occur in the disk format, higher sample flow rates can be applied in comparison with cartridges. The main drawback is that automation is not straightforward. Recently, a new disk format has appeared on the market, the Speedisk, and it is provided with, in addition to the active sorbent, a mesh which acts as a filter and, therefore, filtration and extraction are performed in a single step. With these disks, the sample flow rate can be increased to 200 mL min^{-1}, which is interesting from the viewpoint of routine analysis of industrial and wastewater samples, which normally contain large amounts of particulate material. The overall conditioning and extraction time using Speedisks is 10 min. Moreover, racks of six positions are available, which means that six samples can be processed simultaneously in 10 min. Barceló et al.[42,43] have applied Empore extraction disks in OPPs analysis and more recently Speedisks have been used to extract several pesticides from surface waters.[44]

3.2.2 *On-line Solid-phase Extraction Methods*

A different approach for SPE is the use of on-line systems. In this case, the sorbent is packed in a stainless-steel or Teflon cylindrical precolumn, which can be used under pressure, and is connected on-line with LC or GC. Typical on-line methods for extraction of pesticides were described by Nielen et al.[45] and Brinkman.[46] The set-up is very simple and is illustrated in Figure 2. The water sample containing the pesticides is pumped, via an inexpensive pump, to the precolumn, which is placed at the loop position of a six-port valve. After preconcentration of the water sample, the valve is electrically switched from the preconcentration position to the elution position and the pesticides are desorbed by the mobile phase and are directed to the analytical column where separation will take place. The main advantage of on-line systems is that they can be fully automated, which makes them specially suitable for routine monitoring programs. The Prospekt (Spark Holland, The Netherlands) and OSP-2 (Merck, Germany) are the two systems commercially available for automated on-line preconcentration. The two systems have their own precolumn design, which covers a large variety of packing materials and includes the possibility of elution in normal and backflush modes.

Parameters that can be optimized for on-line SPE are the volume of water to be preconcentrated, the flow rate, normal or backward elution and type of sorbent used. However, in order to achieve optimal performance with on-line SPE, the sorbent of the precolumn should be as similar as possible to the analytical column packing

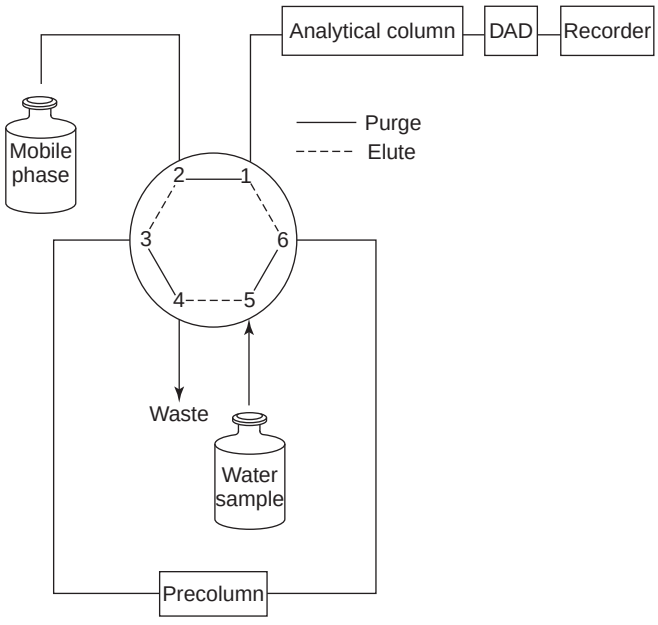

Figure 2 General diagram of the system used for on-line preconcentration and determination of OPPs in water samples. (Reprinted from 'Handbook of HPLC', eds. E. Katz, R. Eksteen, P. Schoenmakez, N. Miller, with permission of Marcel Dekker Inc.)

in terms of type of packing, particle size, etc., to avoid band broadening which can appear specially when normal flush elution is used. Owing to band broadening, quantification of on-line SPE samples with direct injection is not advisable and Pichon and Hennion[47] indicated a 5% error if quantification of on-line samples was compared with direct injection. Band broadening can be minimized by using a suitable gradient which causes peak compression on the top of the analytical column. The size of the precolumn is also of importance because the elution profile of the analytes should be as narrow as possible, especially at the beginning of the separation, where the high water content tends to cause peak distortion. Precolumns are generally 2 mm long and 2–3 mm i.d., and are packed with 10–60-μm sorbent material which efficiently traps the analytes while precolumn clogging is prevented. This design is adapted for classical analytical columns of 15–25 cm × 0.46 cm i.d.

Recently, a prototype has been developed by Hewlett-Packard where on-line SPE with LC/DAD has been coupled on-line with a filtration device and refrigerated reservoirs. Such a system is controlled by a unique software and it is totally automated, from sample collection and filtration to the final chromatographic analysis.[48] This is a useful approach for the routine monitoring of pesticides in situations where it is necessary to control the water quality over a long time (e.g. depuration plants).

An alternative to SPE coupled with LC is the coupling of SPE on-line with GC.[49] This system, known as SAMOS GC, includes the preconcentration unit Prospekt (Spark Holland), an HP 6890 gas chromatograph and an HP Chemstation. Interfacing SPE with GC is the most critical part of the system and it includes a retention gap and a retaining column prior to the analytical column. The different interfaces have been discussed by Brinkman et al.[46,50,51] SPE with GC has been coupled to flame ionization detection (FID), NPD and MS with an HP 5973 mass-selective detector. In general, the preconcentration of 1–10 mL of river water is sufficient to quantify pesticides at concentrations of 30–100 ng L^{-1}.[52]

On-line methods have the advantage that they can be easily coupled to both LC and GC, making these techniques suitable for the analysis of pesticides in water. In relation to these techniques, the use of small precolumns permits the miniaturization and total automation of the preconcentration technique. DLs at the nanograms per liter level and excellent reproducibility can be obtained by using the on-line approach because the entire sample is transferred to the analytical column and losses during sample manipulation are minimized. As a result, degradation studies can be followed without risk of degradation of pesticides during the extraction procedure. An additional advantage is that by using different types of sorbent it is possible to fractionate the extraction of compounds of different chemical classes.

Another methodology designed for a fast extraction and detection method has been developed by Brinkman's group.[54] The technique, known as single short column (SSC), consists in a unique column 3 cm long in which extraction of the analytes is done at the same time as they are separated. These columns are packed with C_{18} or polymeric sorbent and differ from that of a conventional analytical column because they have a higher particle and pore size. As a result, a slight separation of the analytes is performed and, upon elution, the analytes are detected. Similarly to conventional on-line systems, this unique column is coupled to the chromatographic system, and after pumping the water through the SSC desorption takes place on passing the mobile phase through the column. However, ideally an MS detector is necessary to achieve selectivity and unequivocal identification. Hogenboom et al.[55,56] detailed the application of the method for water monitoring and indicated that it can be used to identify and even to quantify the analytes. The main advantages of the system are its simplicity and that it consists of a cheap method, which is useful as a screening method for samples that, often, contain small amounts or no trace of pesticides.

Similarly to cartridges, disposable SPE precolumns can also be used for the storage of pesticides. Lacorte et al.[53] have used the precolumns of the Prospekt (Spark Holland) for stabilizing 19 OPPs extracted from water samples. Table 3 reports the percentage recoveries of selected OPPs after storing the loaded precolumns at room temperature for 1 month, at 4 °C for 3 months and at −20 °C for 8 months. It can be seen that −20 °C is the best option to avoid degradation. The main advantages reported are the easy transport of the sample and the simplicity and low cost of storing water samples.

Table 3 Average percentage recovery ($n = 3$) of OPPs stored at room temperature (RT) for 1 month, 4 °C for 3 months and −20 °C for 8 months

Compound	λ (nm)	RT (1 month)	4 °C (3 months)	−20 °C (8 months)
Mevinphos-*cis*	220	n.f.	35	22
Mevinphos-*trans*	220	n.f.	19	58
Dichlorvos	220	59	n.f.	55
Fensulfothion	254	123	74	89
Azinphos-methyl	220	115	100	111
Fenamiphos	254	n.f.	n.f.	101
Phosmet	220	39	61	34
Pyridafenthion	254	84	n.c.	100
Parathion-methyl	280	117	109	105
Malathion	220	116	n.c.	124
Fenitrothion	254	112	91	108
Azinphos-ethyl	254	112	91	108
Chlorfenvinphos	254	114	90	106
Fenthion	254	89	85	103
Parathion-ethyl	280	111	104	109
Coumaphos	280	117	103	113
Fonofos	254	n.f	n.f.	107
EPN	220	99	103	83
Chlorpyrifos	220	88	104	133

Analyses were performed using on-line LSE with LC/DAD. Quantification wavelength is indicated (λ). Volume of water preconcentrated = 26 mL, spiked at 10 ng mL^{-1}. n.f. = not found; n.c. = not calculated/impossible to quantify.

3.3 Types of Sorbent Materials

A wide range of sorbent materials are available today in different configurations. Even though bonded silicas such as C_{18} or C_8 are still the most commonly used materials for environmental analysis, polymeric sorbents have become a good alternative for the effective preconcentration of moderately to polar organic compounds.[57,58] Polymeric sorbents are available from most manufacturers under various trade names such Envi-chrom, PLRP-S, Lichrolut EN, Isolute ENV, Oasis, etc., the main differences among them being the pore and particle size distribution. Modifications such as acetylation[59] or sulfonation[60] have a mixed retention mechanism consisting in hydrophobic and ionic interactions which facilitate the trapping of the most polar pesticides and increase recovery values. DiCorcia et al.[61,62] described the use of GCB for analyzing phenols, but it has not been widely applied

for the analysis of OPPs owing to excessive retention of the most apolar compounds. The poor mechanical stability of GCB has prevented its routine application in the on-line approach. This problem was solved by the development of porous graphite carbon (PGC), where the graphite is immobilized on a silica structure and the phase is pressure-resistant and shows better stability.[63] In general, PGC precolumns should be coupled with a PGC analytical column to prevent band broadening.

Anion exchangers are also applied for the analysis of very polar analytes which are not trapped using polymeric phases. Combination of various sorbent materials such C_{18}, GCB or PLRP-S with an anion exchanger are useful for extracting analytes of different polarities and can be used with both on-line and off-line procedures. Connection of two precolumns can also be used for sample cleanup, since one adsorbent is used to trap the analytes of interest while the other retains the matrix impurities.

Some limitations of conventional SPE materials are related to their non-selectivity because they are based on polar and hydrophobic interactions capable of retaining humic and fulvic acids and other water matrix compounds and the incapacity of retaining very polar compounds. These problems have led to the development of new sorbent materials, which are based on specific interactions between the target analyte and the sorbent, and have the property of specifically retaining the analytes for which they have been designed. Really promising is the development of different types of immunoaffinity sorbent (IS) or immunosorbent[20] and molecular imprinted polymer (MIP)[64] for SPE. In the case of IS, by taking advantage of cross reactivity, trapping capacity and selectivity, single analytes or families of compounds of equal chemical nature can be extracted, thus avoiding matrix interferences. This approach has recently been applied for pesticide analysis and Pichon et al.[20,65] and Ferrer et al.[66] reported the breakthrough volumes of several herbicides obtained with immunosorbents. So far, immunosorbents have not been developed for OPPs.

The other newly developed alternative to conventional SPE is the use of MIPs, which act in a similar way to antibodies since the interaction between the analyte and the polymer is also highly selective. In contrast to IS, MIPs are synthesized chemically. They consist of synthetic polymers obtained by polymerizing a monomer with a cross-linking agent around a template (analyte). The sorbent is packed in cartridge format. The preconcentration takes place owing to the cavities on the polymer left after washing the template. These cavities will selectively bind the target analyte and structurally related compounds. As in the case of IS, it consists in a selective preconcentration where no further cleanup step is required. MIPs have started to be used for analyte SPE mainly in the pharmaceutical area. The principal application of MIPs is in areas where a specific recognition of particular molecules is needed and antibodies are not available. In these cases, conventional sorbents are not useful since an efficient cleanup is necessary to avoid matrix interferences.[67] In environmental applications, MIPs have been developed for atrazine, which is a well-known herbicide, as an ideal compound. In this area of application, only model conditions have been used, such as preconcentration of distilled water, and the extraction procedure is still under development, since the extraction efficiency depends on parameters such as solvent composition and pH.[64] The development of MIPs for OPPs is a promising approach which will permit selective extraction from both water and food samples.

The main advantages of IS and MIPs over conventional SPE sorbents is that since there is a selective molecular recognition to capture the target analyte, they permit a selective extraction of the sample without the need for further cleanup. Therefore, a high degree of sample preconcentration and purification can be achieved with IS and MIPs. Straight benefits will be for samples in which the impurities co-extracted with a conventional reversed-phase sorbent interfere in the chromatographic analysis, e.g. with diode-array detection (DAD), or suppress the ionization process in the case of MS detection.

3.4 Parameters Involved in Solid-phase Extraction

Owing to the large variety of packing materials, sizes and formats, the extraction efficiency for both off- and on-line modes will vary depending on the nature of the analytes and on the sample matrix. The most critical parameter in SPE efficiency is the breakthrough volume, or the maximum volume of water that can be preconcentrated upon an adsorbent without producing losses of a particular analyte. The breakthrough volume depends on the polarity of each analyte, on its affinity for a particular type of sorbent and also on the amount of sorbent, and therefore will vary from pesticide to pesticide. In on-line methods, breakthrough of the more polar pesticides is originated by the restricted size of the precolumns. This does not happen with off-line procedures since the amount of sorbent is not a limitation and, therefore, higher breakthrough volumes are obtained. Different methods for calculating the breakthrough volumes are described in a review by Barceló and Hennion[68] and involve the study of breakthrough curves or empirical estimation from elution chromatography.[69,70]

In addition to the physicochemical properties of each pesticide, the breakthrough volume will also depend on the type of water. The analysis of environmental waters gives, in general, lower breakthrough values than the

analysis of drinking water. This is due to the presence of humic and fulvic acids, which can bind organic contaminants and decrease the recoveries since only the dissolved fraction will be enriched.[71] As mentioned by Senseman et al.,[72] the type of water can notably affect analytical performance. An important parameter that affects the extraction efficiency is the pH, and the sample pH should be optimized to maximize the analyte recovery and DLs. Acidification of the sample is a common practice in environmental analysis because it can prevent the deprotonation of the most acidic compounds and inhibit microbiological degradation. However, acidification can raise the hydrophobic character of humic material and lead to higher absorption of the humic material into the sorbent. Those analytes which lack a strong affinity for the sorbent will have lower breakthrough volumes since there is strong competition between the analytes and the humic substances. The humic acids appear in the chromatogram as a large unidentified complex matrix (UCM) interfering peak.[57] Humic material can be partially removed from the sorbent by passing a small amount of water through the sorbent bed after percolating the sample. However, it should be taken into consideration that the washing step can lead to a decrease in recovery because when removing humic material analytes will also be removed. Therefore, the water volume should be carefully optimized, especially when working with low breakthrough volume analytes.

4 COLLECTION AND PREPARATION OF FOOD SAMPLES

Owing to food QC and regulations on QA, OPPs have to be monitored in all agricultural food samples before they are introduced to the market.[73] The major input of pesticides on agricultural products is attributed to direct pesticide application for crop protection and contamination through polluted groundwater irrigation. Food poisoning by OPPs in different food matrices has been reported by Lee et al.[74] and Miyamara et al.[75] The literature regarding the analysis of OPPs in milk and cheese,[76,77] oils and butterfat,[78,79] vegetables,[80,81] eggs[82] and grain,[83] among a long list, refers to different extraction procedures. A comparison of several detection techniques was made by Lee and Wylie.[84] The analysis of food samples differs considerably from that of water samples, and will be discussed below.

4.1 Collection of Food Samples

Meaningful residue data in food samples can only be obtained if careful sampling and collection have been performed. Therefore, the sampling plan should be well studied in relation to the type of pesticide considered and the matrix to be studied. The sample should be representative of a whole population, and should be gathered according to the level of pesticides that have to be determined. Special care should be taken during sample collection to avoid nonrepresentative or spoilt samples. The correct procedure is to collect random grab samples of a whole population, mix the grab samples and reduce the sample size to a final laboratory-size proportion. However, the nature of the matrix will influence the sampling plan. The *Manual of Pesticide Residue Analysis*[85] reviews the sampling procedures for and analysis of OPPs in different types of food. It is established that for homogeneous products, 0.5–1 kg is sufficient for analysis. For products consisting of small units, sampling should be performed depending on the weight of each unit, e.g. 50 units (1 kg) for items of less than 25 g, 30 units (1–3 kg) for items of 25–100 g, 15 units (2–5 kg) for items of 100–250 g and at least 10 units for items larger than 250 g. After collection, food samples should be immediately transported to the laboratory in suitable conditions (normally refrigerated) in order to avoid degradation of the pesticides or spoilage of the sample matrix.

4.2 Analytical Preparation of Food Samples

Upon receipt, samples should be processed immediately, or otherwise stored at 4 °C or preferably at −20 °C. Before analysis, it is common practise to check the food sample and note the appearance, aroma, possible deposits, etc. All the physical characteristics should be described, which will aid in interpretation of the final result. Moreover, at this stage, impurities such as stones, insects or rotten parts should be removed. The sample should never be washed before processing. The sample should subsequently be weighed and the number of units should be recorded. In some cases, non-edible parts, e.g. stems, peel, leaves and roots, should be removed and the concentration of residues is then established for the weight of the edible portion. In other cases, it is important to analyze the residue levels of OPPs in the outer sample surface. Of course, the proportion of the sample analyzed will depend on the aim of each study.

At this stage, the gross laboratory sample should first be composited in such a way that representative analytical size samples are obtained from it. This step will depend on the type of sample: whenever possible, material should be homogenized, ground and mixed; products consisting of small units should be quartered and two opposite quarters should be homogenized. For large samples units (cheese, melons, etc.) which cannot be homogenized entirely, aliquots should be gathered and finally homogenized. In any case, from the gross laboratory sample it should be possible to obtain several analytical-size samples, from

which one or two will be analyzed and the rest will be kept as a reserve. These samples should be kept at −20 °C and wrapped in aluminum foil or kept in glassware, to prevent further decomposition or contamination.

Once the sample has been homogenized and mixed, the pesticides should be extracted. Food samples are more complex to analyze than water samples since the former are far more diverse and differ in the matrix composition (presence of fats, sugars, etc.). The *Manual of Pesticide Residue Analysis*[85] describes extensively the different methods used to extract OPPs from different types of food matrices. In this section, only the most common extraction procedures will be mentioned. Since for food samples extraction is influenced by the matrix, the application of one extraction method or another will be basically dependent on the type of pesticide to be analyzed (polarity, K_{oc}, volatility, etc.) and on the type of matrix. In contrast to water samples, in most food samples it is often necessary to perform a cleanup step to remove impurities which would interfere or even suppress the analytical determination. For this reason, it is important to use a selective extraction procedure which will maximize sample cleanup and will permit the use of conventional chromatographic techniques for their analysis.

Freeze-drying is a common method for conditioning biota samples for the extraction of pesticides.[86] Freeze-drying of biological material is used before extracting the pesticides since it destroys cell membranes and thereafter the contact between the cell tissue and the extraction solvent is increased. Soxhlet extraction is the most usual method to extract pesticides from food samples, despite the sample manipulation and tedious work involved. The most common solvents used for this purpose are ethyl acetate,[32] acetone and methanol,[87] acetonitrile[88] and hexane.[89]

Nowadays, microwave energy is sometimes used to extract various types of organic compounds from sediment and food samples. However, a recent trial to extract 64 OPPs from onions using a microwave oven indicated that extraction and recovery are affected by microwave heating.

Many extraction methods have been developed to extract pesticides from food samples and the application of one or another depends on the food sample matrix. For instance, strong acids or alkalis are used to facilitate the extraction of pesticides from foods with a high lipid content. However, many OPPs are decomposed in such media. As a result, the sample is extracted at neutral pH and the fatty matrix is subsequently removed by a suitable and nondestructive cleanup step, usually column chromatography with silica gel, alumina or Florisil.[90,91] The last method is that adopted by the USEPA.[92]

Gel permeation chromatography (GPC) is another cleanup alternative, and can be used with low-pressure chromatographic columns in such a way that it permits the discrimination of compounds depending on their molecular size. Large biogenic compounds, such as lipids, are excluded from the pores of the polymeric material and are eluted before smaller analytes, which are retained in the column pores. The separation mechanism is a combination of size exclusion, adsorption and partition. The prevalence of one mechanism over another depends on the mobile phase and the pore size of the polymeric material chosen. Common elution solvents used in GPC are ethyl acetate–cyclohexane,[32] acetone–cyclohexane, acetone–light petroleum[93] and cyclohexane–dichloromethane.[94] Thirty-nine pesticides (including their major metabolites) were determined in fatty processed food using dichloromethane extraction and GPC with a Biobeads SX3 column and dichloromethane–cyclohexane as eluent, giving recoveries between 51 and 185%.[95] The same column has been used to clean up carrot extracts, indicating that the high pigment content of samples produced band broadening on the column.[96] A different approach consisting of accelerated solvent extraction and cleanup with GPC has been used to analyze OPPs, among others, in foods, indicating recoveries from 80 to 90% and precision below 10%.[97] GPC has been also coupled on-line with GC and has been applied to extract OPPs in olive oil.[98] Other authors have indicated an efficient extraction of olive oil with acetonitrile, with no need for sample cleanup prior to GC, and reported recoveries ranging from 74 to 118%, indicating that the column efficiency is not affected by co-extracted oil impurities.[99]

A less common approach for the cleanup of food samples is the use of normal-phase LC, with silica columns and dichloromethane–hexane or dichloromethane–pentane as eluent.[100] With this system, pesticides are the first compounds to elute, and the lipidic components elute at the end of the chromatogram.

SPE with extraction disks has also been used to extract OPPs in high-moisture foods (<2% fat) with no need for further cleanup, and better recoveries than using conventional LLE procedures were reported.[101] SPE has been used as a cleanup technique to extract OPPs from eggs[82] and from cereal flours[102] by using minicolumns prepacked with Florisil. A promising analytical technique to extract OPPs selectively from food samples is the use of IS or MIPs, but application in such matrices is still under development.

Other techniques to extract OPPs selectively without further cleanup consist of matrix solid-phase dispersion, which has been applied to extract OPPs from milk[103] and oranges,[104] while supercritical fluid extraction (SFE) permits better recoveries of several OPPs in comparison with Florisil cleanup.[105]

5 GAS CHROMATOGRAPHIC DETERMINATION

5.1 Gas Chromatographic Detectors

GC plays an important role in trace analysis. It is a well-known and widely established analytical technique and it is the official technique adopted by the USEPA (EPA Method 8140).[106] GC coupled to an NPD, FID or electron-capture detector (ECD), atomic emission detector (AED) or pulsed-FPD is still the most common technique for OPP analysis in different environmental samples. The main reasons are the low DLs that can be achieved and high selectivity which permits multicomponent analysis and minimizes the effect of the sample matrix. In general, GC detectors provide a linear range over various orders of magnitude and DLs at the low nanograms per liter level in water and nanograms per gram level in food samples. However, when using these detectors, the USEPA indicate the need to confirm the results by using two different GC columns of different polarity, usually a DB-5 and a DB-1701.[107] The use of different columns solves the risk of co-elution and avoids false-positive determinations in environmental water analysis. Lacorte and Barceló[21] carried out a systematic study with the aim of studying the retention times of 26 priority OPPs and various TPs in three different capillary GC columns. Another way to confirm the results is the use of GC with MS as the detection technique[14] for both qualitative (it provides mass spectral information) and quantitative purposes. GC/MS has been performed either with the EI technique, positive chemical ionization (PCI), negative chemical ionization (NCI) and tandem mass spectrometry (MS/MS) and has been widely applied for pesticide analysis.[86,108,109] Information that can be obtained with GC/MS is summarized below.

5.2 Gas Chromatography/Mass Spectrometry with Electron Ionization

One of the main advantages of GC/MS with EI is the existence of libraries which contain more than 120 000 spectra. EI gives very good reproducibility of the spectral data, even when using different mass spectrometers, so that identification and confirmation of the presence of OPPs in an environmental sample are performed by spectral matching. At the same time, GC/MS gives good sensitivity and can also be used for quantification purposes. EI is normally coupled with quadrupole mass analyzers, but magnetic sectors and ion traps can also be used. Different analyzers can lead to variations in the relative abundances in the EI mass spectrum. Various references with GC/MS data using EI for the analysis of OPPs have been published.[110–114] EI provides abundant fragmentation of the analytes but molecular weight information is scarce, since the molecular ion species, $M^{\bullet+}$ is not always observed. For this reason, the USEPA recommends the identification for one pesticide of at least three different ions. Upon fragmentation, OPPs produce diagnostic or characteristic ions which correspond to the functional group structure of each family of pesticides and permit their identification. Other fragment ions formed correspond to specific ions, and are typical of each compound. An extensive study reported by Wilkins et al.[115] was based on the identification of various OPP compounds by five typical rearrangements, with diagnostic ions corresponding to m/z values of 93, 97, 109, 121 and 125. A more recent study,[21] also concerning OPPs, gave information on diagnostic fragment ions obtained by GC/MS with EI as well as compound-specific ions, as indicated in Table 4. Various possibilities of tentative ion identification for the same fragment ion, e.g. m/z 97, 109, 127 and 137, are given. This qualitative information is valuable since it can be used for identifying OPPs and their TPs from complex environmental matrices. It was concluded that (i) diagnostic fragment ions reported in Table 4 usually correspond to a class of OPP compounds. In this sense, m/z at 109 and 125 are base peaks for the dimethylphosphates and dimethylphosphorothioates, respectively. (ii) In some cases, diagnostic fragment ions depend on the type of molecule and it is possible that two

Table 4 Typical fragment ions of OPPs in GC/MS with EI

m/z	Ion	Other possibilities
47	CH_3S^+	
62	CH_3OP^+	
63	CH_3OPH^+	
75	$CH_2SC_2H_5^+$	
77	$C_2H_5OPH^+$	
79	CH_3OPOH^+	
93	$(CH_3O)_2P^+$	
96	$CH_3OP(OH)_2^+$	
97	$(HO)_2PS^+$	$P(SH_2)^+$
109	$(CH_3O)_2PO^+$	$(C_2H_5O)OHPO^+$, $(CH_2H_5O)PSH^+$
121	$(C_2H_5O)_2P^+$	
125	$(CH_3O)_2PS^+$	
126	$(C_2H_5O)OHPSH^+$	
127	$(CH_3O)_2P(OH)_2^+$	$(C_2H_5O)P(OH)_3^+$
137	$(C_2H_5O)(C_2H_5)PS^+$	$(C_2H_5O)_2PO^+$
141	$(C_2H_5O)OHPS^+$	
142	$(CH_3O)_2POHS^+$	
153	$(C_2H_5O)_2PS^+$	
157	$(CH_3O)_2PS_2^+$	
158	$(CH_3O)_2PS_2H^+$	
169	$(C_2H_5O)_2POS^+$	
171	$(C_2H_5O)_2POHSH^+$	
185	$(C_2H_5O)_2PS_2^+$	

Reprinted from S. Lacorte, C. Molina, D. Barceló, 'Screening of OPP in Environmental Matrices by Various Gas-Chromatographic Techniques', *Anal. Chim. Acta*, **281**, 2451–2460, [21] Copyright (1995), with permission from Elsevier Science.

compounds will have an ion at the same m/z value but have different chemical compositions (indicating different families). This is the case of m/z 109, corresponding either to $(CH_3O)_2PO^+$ (methylphosphates), $(C_2H_5O)OHPO^+$ (ethylphosphates) or $(C_2H_5O)PSH^+$ (ethylphosphorothioates). The rearrangement ion at m/z 97 usually corresponds to $(OH)_2PS^+$ (phosphorothioate), but in the case of phosphorodithioates it corresponds to $(SH)_2P^+$. It is therefore useful to have the chemical structures of the different compounds since some of the fragment ions can then be predicted. (iii) The abundance of molecular ion species M^+ increases with its ability to sustain a positive charge and is related to the ionization energy, which is lower for S than O. Examples of this case include parathion-ethyl and fenitrothion, which show higher relative abundances of M^+ as compared with the oxo derivatives. (iv) The chlorinated compounds exhibit as main peaks the corresponding fragment ions with chlorine losses. (v) Specific fragments which are typical of some compounds are also observed, e.g. McLafferty rearrangement involving proton extraction by an oxygen of the nitro group indicating the presence of CH_3 adjacent to an NO_2 group, as for fenitrothion and fenitrooxon.

The sensitivity of GC/MS can be increased by one order of magnitude by using selected ion monitoring (SIM), where only a few ions are recorded. By selecting appropriate diagnostic and specific fragment ions of OPPs, quantitation of environmental samples can be performed at the parts per billion level and the technique can be routinely applied in monitoring studies. In general, the use of higher m/z values is preferred since better selectivity is obtained and matrix interferences are minimized.

5.3 Gas Chromatography/Mass Spectrometry with Chemical Ionization

GC/MS with either PCI or NCI consists in a soft ionization technique which gives molecular weight information that EI spectra rarely give. In CI a reagent gas, normally methane, isobutane or ammonia, is used for the ionization process. CI mass spectra are not as reproducible as with EI, since many parameters are involved in the ionization process. The reagent gas used, the reagent gas pressure applied and the source temperature affect the fragmentation pattern and the sensitivity. It is for this reason, and because mass spectra are very much dependent on the instrument type, that there are no libraries of spectra available for CI mass spectra. Of the two modes of ionization, NCI is more selective and sensitive than PCI and EI and it is the preferred option for the determination of OPPs with electron-withdrawing groups (nitro- or chloroaromatic groups).[108,109,116,117] The molecules of the so-called parathion-like structure offered high sensitivity under the negative ion conditions owing to the presence of an aromatic moiety which is accompanied by a nitro group, thus forming a kind of pseudoacid by electron attachment which stabilizes the negative charge.

When this aromatic moiety contains other electron-withdrawing groups, e.g. chlorine atoms, the stabilization of the negative charge and sensitivity are also enhanced. The sensitivity can be increased by two orders of magnitude compared with GC/MS with EI.[108,116] Table 5 lists the diagnostic ions of the different OPPs, which are characteristic of the different chemical groups. The following general remarks can be made. (i) The diagnostic ions formed under NCI are, in many cases, the base peaks of the analytes and, in some cases, the only spectral information obtained. This is a typical characteristic on MS with NCI of OPPs, so their unequivocal identification should always be performed in combination with the retention time. (ii) The formation of M^{*-} has a 100% relative abundance for the OPPs of the so-called parathion group, and compounds having an aromatic moiety (with the exception of chlorinated OPPs). This is due to the fact that M^{*-} is fairly stable under NCI conditions when an aromatic moiety exists in the molecule, and is even more important when a nitro group is present. (iii) The chlorinated OPPs with an aromatic moiety have intense or base peaks corresponding to losses of Cl, such as CH_3Cl or HCl, owing to the facility of such processes under NCI conditions. (iv) The formation of thiofenolate versus fenolate ions for compounds such as parathion, fenitrothion, coumaphos and fenchlorfos is due to their stronger acidity in the gas phase and there is a transfer from the aromatic moiety from the oxygen to the sulfur atom.

Complementing GC/MS with EI and NCI is useful in environmental analysis because some compounds will give a better response under one type of ionization mode whereas others, with different structural characteristics, give a better signal under the another. Moreover, the use of two GC/MS techniques is a means of confirming the presence of pesticides in environmental matrices

Table 5 Diagnostic ions of OPPs with GC/MS and NCI

m/z	Ion	Group
125	$(CH_3O)_2PO_2$	Dimethylphosphates
141	$(CH_3O)_2POS$	Dimethylphosphorothioates
153	$(C_2H_5O)_2PO_2$	Dimethylphosphorothionates
157	$(CH_3O)_2PS_2$	Dimethylphosphorodithioates
169	$(C_2H_5O)_2POS$	Diethylphosphorothioates
185	$(C_2H_5O)_2PS_2$	Diethylphosphorodithioates

Reprinted from S. Lacorte, C. Molina, D. Barceló, 'Screening of OPPs in Environmental Matrices by Various Gas-Chromatographic Techniques', *Anal. Chim. Acta*, **281**, 2451–2460,[21] Copyright (1995), with permission from Elsevier Science.

and, more importantly, it permits the unequivocal identification of TPs that might be formed in environmental samples. Such an approach was used to follow the degradation kinetics of fenitrothion after being applied in the environment for eradicating the crab *Procamburus clarkii*. Fenitrooxon, the *s*-methyl isomer of fenitrothion and 3-methyl-4-nitrophenol were detected as the main TPs of fenitrothion by GC/MS with both EI and NCI, and trace levels were found in water 4 days after application.[6] Such techniques permitted the concentration of fenitrothion to be monitored for 7 days, which was useful for determining the half-life. Under the field conditions and at water temperatures of 10–13 °C, the half-life of fenitrothion was estimated to be 13 h. The evolution of the levels of fenitrothion under environmental conditions are shown in Figure 3. Figure 4 shows the degradation pathway of fenitrothion under these conditions. It can be concluded, therefore, that GC is a powerful technique in environmental water analysis, owing to the separation efficiency, the wide availability of detectors, including MS, and the possibility of performing degradation studies. However, one of the limitations of GC/MS is that polar, non-volatile or thermolabile pesticides cannot be directly analyzed without a derivatization step.[118] Derivatization increases sample manipulation and time of analysis and introduces new sources of errors. For such reasons, there is a general tendency to switch to LC, which can overcome the aforementioned limitations.

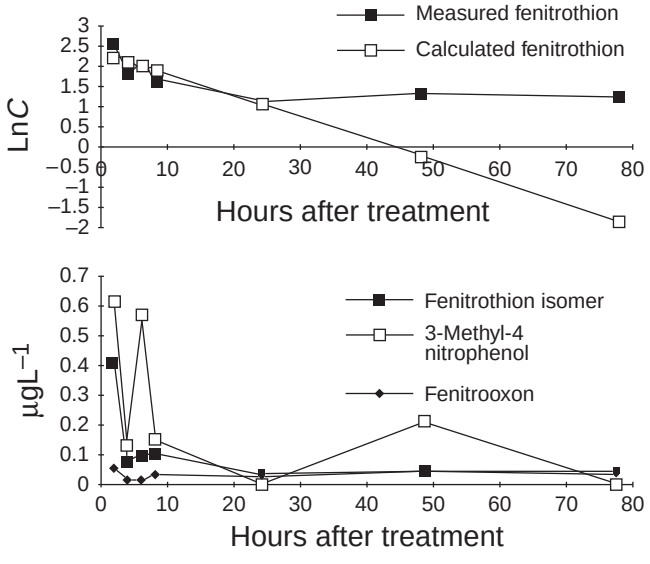

Figure 3 Evolution of the levels of the concentration of fenitrothion [plotted as logarithm of concentration versus time, giving both measured values and calculated values, according to the first-order rate Equation (1)] and its TPs after application of fenitrothion in the field at a concentration of 200 µg L^{-1}. (Reprinted from 'Rapid Degradation of Fenitrothion in Estuarine Waters.' *Environ. Sci. Technol.*, **28**, 1159–1163. Copyright (1994) American Chemical Society.)

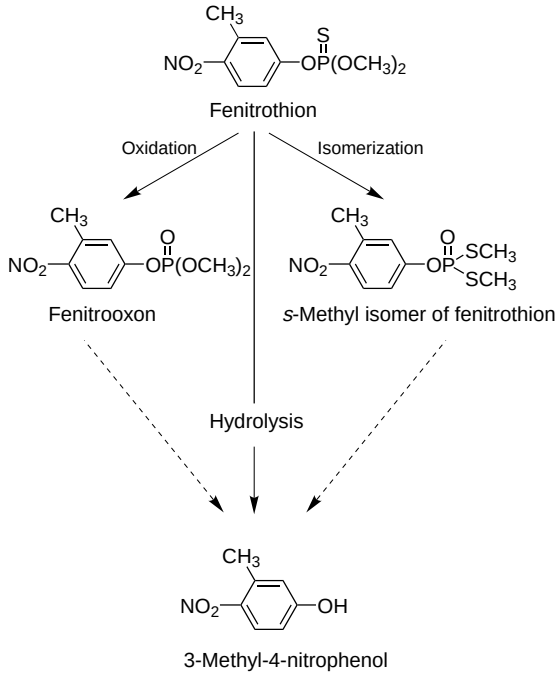

Figure 4 Decomposition pathways of fenitrothion in estuarine waters. (Reprinted from 'Rapid Degradation of Fenitrothion in Estuarine Waters.' *Environ. Sci. Technol.*, **28**, 1159–1163. Copyright (1994) American Chemical Society.)

6 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

HPLC has become popular for the analysis of compounds that are not amenable to GC. The increasing development of HPLC techniques has led to the sensitivity needed to meet EU legislation for the analysis of polar pesticides in surface and food samples. In HPLC the mobile phase flows through the stationary phase and separation of the analytes takes place owing to the interactions between the solute, the mobile phase and the sorbent. In environmental applications, reversed-phase chromatography is the most common separation mechanism used in HPLC, although other approaches such as adsorption, ion-exchange and size-exclusion chromatography (SEC) can also be used. Stationary phases used in reversed-phase chromatography are alkyl- or phenyl-bonded silicas, polymeric materials or carbon-based sorbents. By using methanol, acetonitrile, water and tetrahydrofuran (THF) or combinations of them, it is possible to separate pesticides over a wide range of polarities, including their TPs. Since HPLC techniques can be applied to non-volatile, thermolabile and polar compounds, it is becoming the favored technique for OPP analysis. An additional advantage over GC is that on-line SPE and postcolumn systems are easily compatible with

LC. This feature combines the extraction step and chromatographic analysis in a single, totally automated step, which is an interesting property from the viewpoint of routine analysis and reduction of sample manipulation.

HPLC can be coupled with various types of detectors, such as UV and fluorescence detectors, DAD and mass spectrometers, which make the techniques robust and adapted to each experimental requirement. Coupling LC with MS is not as obvious as with GC/MS, since the mobile phase has to be eliminated before entering the mass spectrometer and at the same time ions have to be generated. A book[119] has been published that provides an overview of the different LC/MS applications for environmental analysis. Several interfaces have been designed to overcome the problem: trace enrichment (PB and moving belt) and nebulization (TSP and API interfaces). Trace enrichment interfaces have the advantage that they can provide EI mass spectra but their application in the environmental field has been reduced owing to their low sensitivity and poor reproducibility. For this reason, these interfaces will not be treated in this article and all the discussion will be focused towards DAD and nebulization devices.

6.1 Liquid Chromatography with Diode-array Detection

Recent developments in DAD permit the achievement of DLs in the nanograms range and a UV visible spectra can be recorded so that the analytes can be identified by spectral comparison against a standard previously injected. Pesticides susceptible to absorbing UV radiation are those which present a suitable chromophore and in general they present a typical UV spectrum with a maximum between 230 and 280 nm. Modern instruments can monitor between 190 and 800 nm, at a rate up to one spectrum per second, which permits very good spectral resolution. Moreover, most software includes the peak purity analysis of each detected compound which indicate potential co-elution problems with another pesticide or with matrix interferents. These two characteristics are very useful since it is possible to distinguish compounds of the same family that can have very similar UV spectra. At the same time, pesticide TPs have spectral similarities with their parent compounds, and there is only a slight shift in the absorbance spectra. By setting the acquisition parameters to obtain good resolution, it is possible to identify chemically related compounds. This characteristic makes HPLC with DAD the most common detection method for pesticide analysis in water and food samples. Peak identification with LC/DAD is usually carried out by means of the retention time in comparison with a standard and by spectral matching through a library search, an option which is nowadays included in most software. Selectivity can be enhanced by choosing an appropriate UV wavelength which permits the identification and quantification of pesticides with very close retention times and even a reduction in the contribution of matrix interferences. Therefore, it is recommended to perform quantitative analysis at the wavelength which provides better calibration adjustment. Peak purity analysis is also important to achieve good accuracy.

A review describes the applications of LC/DAD and other detection techniques for the determination of polar pesticides and related compounds in aquatic samples.[120] Other papers report the use of LC/DAD to monitor OPPs in surface waters, either using on-line SPE[121–123] or off-line methods.[124] The main drawbacks related to DAD are due to the poor selectivity when analyzing complex environmental samples which contain large amounts of humic and fulvic material or other interferents. At this point, sample pretreatment should be optimized, especially the cleanup step to remove interferences which appear especially when working at low wavelengths. At the same time, DLs are increased by this matrix effect.

6.2 Liquid Chromatography and Mass Spectrometry with Thermospray Interface

Liquid chromatography coupled to mass spectrometry with a thermospray interface (LC/TSPMS) has been widely used to determine organic contaminants in water matrices[125,126] and has been adopted by the USEPA (Method 8321)[127] for the analysis of OPPs in different types of water matrices. The LC/TSPMS interface consists of sample nebulization and ionization through a heated capillary. It is a soft ionization technique which produces molecular peak spectra and scarce fragment ions. The spectral pattern depends on the additives added to the mobile phase, which are mainly ammonium acetate or formate. Ionization is due to gas-phase protonation reactions between the analyte and the ammonium.[26,128] The main disadvantage of LC/TSPMS is that it gives poor structural information. This can be solved by using filament-on or discharge assisted ionization which generates solvent mediated ionization similar to direct TSP ionization (without additives) and sensitivity is enhanced.[129] The use of liquid chromatography coupled to tandem mass spectrometry with a thermospray interface (LC/TSPMS/MS) is the best choice to enhance fragmentation of quasimolecular ions[130] and it offers collision-induced dissociation (CID) spectra and is used for both identification and quantification purposes. However, this is an expensive instrument which cannot be afforded by most laboratories.

Ion formation in TSP in the positive ionization (PI) mode generates $[M + H]^+$ and the ammonium adduct $[M + NH_4]^+$ as base peaks for pesticides, depending on the proton affinity of each compound. In the negative ionization (NI) mode, $[M - H]^-$ or acetate or formate adducts are formed. The NI mode is more sensitive and

selective for the analysis of electronegative compounds, such as chlorinated OPPs.

The characterization of several pesticides by LC/TSPMS in PI and NI modes using filament-on ionization was detailed by Lacorte and Barceló,[126] who remarked that little or no fragmentation is observed for many of the pesticides. This represents a drawback since the identification of unknowns in environmental samples (e.g. TPs of OPPs) becomes difficult. The requirement set by the USNPS, in which at least three different ions are needed to confirm unequivocally the presence of a pesticide in a real water sample, generally is not accomplished by using LC/TSPMS. Rather than compound identification, LC/TSPMS is more useful for target compound analysis. Under full-scan conditions, instrument detection limits (IDLs) are around 1 ng for pesticides.[131]

Sensitivity can be enhanced by coupling SPE on-line with LC/TSPMS. This approach has been used to quantify and identify OPPs and their TPs in environmental matrices.[126,132,133] On-line SPE coupled with LC/TSPMS and on-line SPE coupled with LC/DAD have been used to follow the degradation of several OPPs in seminatural conditions (river water spiked at a low concentration level and exposed outdoors for 4 weeks).[10] The half-life measured as $t_{1/2} = \ln 2/k$ indicated that pesticides persisted in such conditions for 6–12 days. However, when degradation occurred, confirmatory analysis of pesticide TPs was carried out by means of MS detection with a TSP interface. The oxo derivatives of chlorpyrifos-methyl, diazinon, isofenphos, pyridafenthion and temephos were detected 4 weeks after sunlight exposure. In general, the oxygen analogues could be identified since they followed the same adduct formation as the parent pesticide and eluted before the parent compound. Figure 5(a–c) shows the LC/DAD and LC/TSPMS chromatograms of a water sample spiked with chlorpyrifos-methyl and analyzed after 4 weeks. Whereas LC/DAD detected chlorpyrifos and a TP not unequivocally identified, LC/TSPMS in the PI and NI modes confirmed the presence of chlorpyrifos (*m*/*z* 323), its oxo derivative (*m*/*z* 324) and 3,5,6-trichloro-2-pyridinol (*m*/*z* 198). Chlorpyrifos and the two TPs identified exhibited losses of chlorine atoms and proton abstraction, similarly to the parent compound. The toxicity of 3,5,6-trichloro-2-pyridinol is greater than that of chlorpyrifos, with EC_{50} values of 18.6 and 46.3 $\mu g\,mL^{-1}$, respectively, calculated with the Microtox system.[134] Also, it has been reported that this compound is toxic to soil microorganisms and results in less mineralization and therefore enhanced persistence of chlorpyrifos in soil. This example indicates the importance of analyzing TPs and the need to include them in the analytical methods.

Although LC/TSPMS affords some structural information, often more fragment ions are needed for confirmatory purposes. LC/TSPMS/MS techniques or API interfaces can solve the problem, since they can produce spectra by CID. An additional drawback of LC/TSPMS is the high intraday signal variations, which lead to poor reproducibility. For these reasons, many laboratories tend to switch to the new API interfaces which nowadays is the preferred option for the analysis of pesticides in environmental samples since it is a more robust technique. API includes several interfaces, i.e. ESP, the high-flow pneumatically assisted ESP, commonly named ISP, the heated pneumatic nebulizer (HPN) and APCI. The main

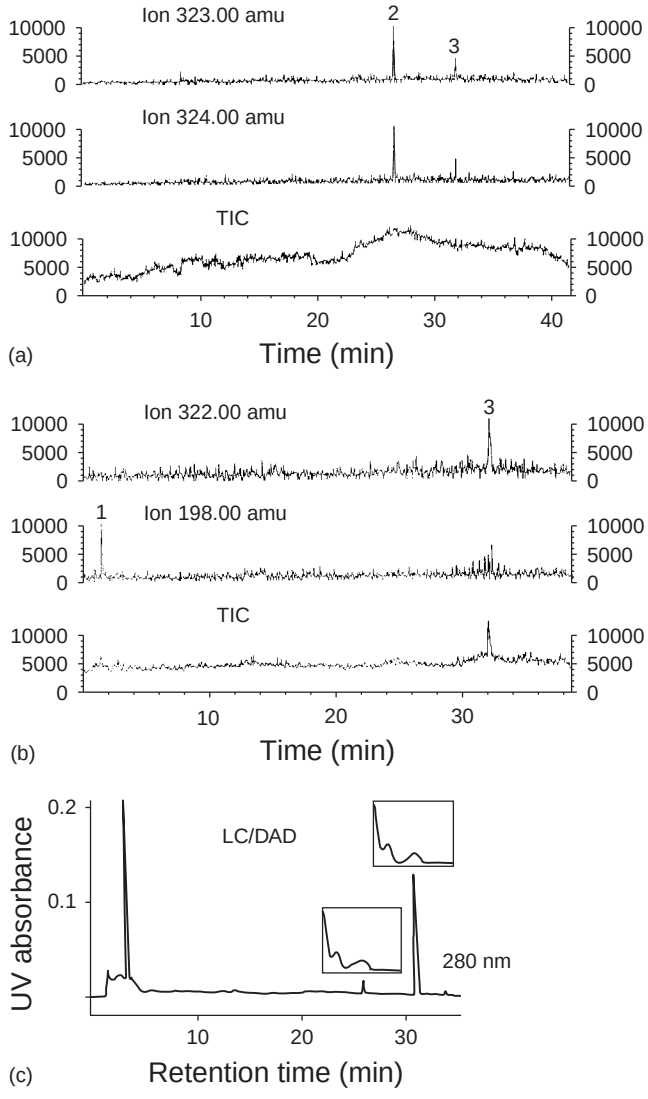

Figure 5 Reconstructed ion chromatograms obtained with on-line LSE with LC/TSPMS with (a) PI and (b) NI mode of ionization of an Ebre river water sample spiked with chlorpyrifos-methyl at 50 $\mu g\,L^{-1}$, and analyzing the water sample 4 weeks after spiking. Peak identification: (1) 3,5,6-trichloro-2-pyridinol; (2) chlorpyrifos-methyl oxone; (3) chlorpyrifos-methyl. (c) Same water sample analyzed by on-line SPE with LC/DAD at 280 nm. (Reprinted with permission from Elsevier Science.)

advantage of these interfaces is their higher sensitivity for OPP analysis compared with TSP or PB.[135–137]

6.3 Liquid Chromatography and Mass Spectrometry with Electrospray/Ionspray Interfaces

Electrospray/ionspray (ESP/ISP) consists in applying a voltage to a capillary that ionizes the solvent extract with the pesticides, which emerge from the capillary as a spray. This spray contains microdroplets which are positively or negatively charged, depending on the difference of potential between the capillary and the counterelectrode. In the desolvation chamber, the droplets undergo ion evaporation and ions escape to the gas phase. Such ions are then focused towards the quadrupole of the mass spectrometer. In electrospray/ionspray mass spectrometry (ESP/ISPMS), 90% of the ions are formed in solution and at atmospheric pressure. The main advantage of ESP/ISP techniques is that no heat is applied to ionize the samples, eliminating thermally assisted degradation, as happens when using a GC or TSP interface. This allows the LC/MS analysis of thermolabile analytes such as trichlorfon,[138] which is a compound that has been shown to suffer thermal degradation under LC/TSPMS caused by the probe tip and gas-phase temperatures higher than 200 °C. The major differences between these two interfaces is that in ESP only 10–80 μL min^{-1} can be directed to the source, whereas in ISP flow rates from 300 to 1000 μL min^{-1} can be applied. A few applications using ESP have been undertaken for OPPs.[29,139] In ESP it is necessary to use syringe LC pumps, or otherwise when using conventional LC pumps a splitting device should be used.[140] In ISP, higher flow rates (up to 1500 μL min^{-1} in the new-generation interfaces) can be handled; a gas flow, normally nitrogen, is applied in the interface and acts as a nebulizer. After nebulization of the liquid, a heated drying gas is applied countercurrent and droplet evaporation is facilitated by breaking the analyte–solvent clusters at the desolvation chamber. The drying gas also prevents clogging of the sample orifice. With ISP, ions emerge from the capillary owing to the simultaneous action of the nitrogen and the capillary voltage set at 2–4 kV.

Ion formation in ESP/ISP is due to the evaporation of ions from charged droplets, which results in ions with very little internal energy and, therefore, molecular or quasimolecular ions are formed with little or no fragmentation. To enhance fragmentation, an extraction potential of 10–200 V, depending on the particular design of the interface, is applied at the extraction cone. This potential focuses the spray towards the sample orifice and into the high vacuum of the mass spectrometer. An increase in this potential accelerates the ions and the collisions between them induce fragmentation via CID. In some cases, an increase in the cone voltage can lead to a decrease in sensitivity. Therefore, careful optimization of the cone voltage is required and generally a compromise between structural information and sensitivity has to be achieved.

In ESP/ISPMS, methanol or acetonitrile and water are generally used as mobile phase. When working in PI and at low extraction voltage (e.g. 20 V), the formation of $[M + H]^+$ and $[M + Na]^+$ is favored and these are generally the base peaks. The generation of $[M + Na]^+$ is more common when methanol is used in the mobile phase since the Na^+ ions are due to impurities in the methanol solution. On increasing the extraction voltage, fragmentation occurs at the expense of $[M + H]^+$ ion and results in diagnostic or compound-specific fragment ions.

ISP under PI permitted DLs at the picogram level to be achieved for selected pesticides. Sensitivity problems for the parathion group (parathion-methyl, parathion-ethyl, fenitrothion and fenitrooxon) were noticed when using ISP or conventional ESP.[29] The ionization of these compounds under ESP conditions offers problems and generally it is difficult to achieve low nanograms per liter DLs in water samples. In general, the repeatability and the long-term reproducibility ($n = 9$) for OPPs vary from 12–17 to 22–30%, respectively, on injecting 1 ng of each compound, which is better than that obtained by using LC/TSPMS.

A remarkable fact is that typical diagnostic ions of OPPs formed in liquid chromatography coupled to electrospray/ionspray mass spectrometry (LC/ESP/ISPMS) are similar to those in EI. The formation of diagnostic ions at m/z 109, 110, 125 and 141, to mention some, permits the identification of each family of pesticides. This is an important feature, which is not accomplished by TSP, and therefore these techniques can be used for the identification of pesticides and their TPs in environmental samples. Moreover, more structural information is obtained at 40 V or higher voltages (compared with 20 V) and permits the easy identification of nontarget pesticides in environmental samples.

The use of ISP in NI has been scarce owing to the poor ion stability and the problems of electric (corona) discharge, presumably caused by electrons emanating from the sharp edges of the ISP capillary needle held at a few thousand volts negative relative to a counterelectrode. The signal instability of electrospray ionization (ESI) can be reduced by adding chloroacetonitrile to the LC eluent, which acts as an electron scavenger, and there is a suppression of the electrical (corona) discharge phenomena leading to a stable ESI with finer charged liquid droplets. Nevertheless, the improvement in ion stability is obtained at the expense of sensitivity (one order of magnitude less) owing to the $[M + Cl]^-$ adduct ion formation and to the higher cluster ions intensity. Since the major analytical requirement of environmental analysis is to achieve a low DL, chloroacetonitrile is not commonly used.

6.4 Liquid Chromatography and Mass Spectrometry with Atmospheric Pressure Chemical Ionization

In APCI, the column eluent is pneumatically and thermally nebulized into an atmospheric pressure ion source. The APCI technique differs from ESP and ISP in that the spray is generated as a result of applying nitrogen gas and heat at the interface. The spray, which contains the mobile phase and the analytes, is subsequently ionized at the nebulizing chamber after applying a corona discharge of 2–4 kV. The analytes are ionized by gas-phase ion–molecule reactions (e.g. proton transfer). The ions formed, together with solvent vapor and nebulizing gas, are sampled towards the sample cone and into the mass spectrometer.

The main advantages of liquid chromatography coupled to mass spectrometry with atmospheric pressure chemical ionization (LC/APCIMS) over ESP/ISP is that ions do not have to be formed in solution and, therefore, compounds with a wider range of polarities and physicochemical properties can be determined. Another advantage is that more fragmentation is obtained with APCI than ESP/ISP and that sodium clusters rarely appear. As heat is applied in addition to pneumatic nebulization, some OPPs can be determined with APCI whereas they gave no signal under ESP/ISP. Moreover, in APCI, higher flow rates can be handled (up to 2 mL min^{-1}), which makes it easy to couple with conventional columns and with SPE switching techniques. In APCI, no buffer addition is required and, in contrast to TSP, no significant differences in sensitivity are observed when working at different ratios of water and organic modifier.

From the practical point of view, there are two main parameters to be optimized in APCI: the corona discharge voltage and the cone extraction voltage. For the corona discharge, a voltage between 2.5 and 3 kV is usually chosen, as at lower discharge voltages weak ionization is achieved and above 3 kV the ionization efficiency decreases dramatically. Similar to the ISP interface, by raising the cone voltage structural information can be obtained via CID. The advantage of APCI techniques is that analytical parameters can be varied to gain both in structural information and sensitivity. The pressure of the nebulizing and drying gas basically affects sensitivity. A study[141] has been published on the probe temperature and extraction voltage effect of the APCI interface on the fragmentation and sensitivity of several OPPs. Whereas the interface temperature has a major role in sensitivity and does not affect fragmentation, an increase in the extraction voltage produces more fragmentation, at the expense of sensitivity. It was found that at a 30 V extraction voltage, the best sensitivity was obtained for the majority of the OPPs studied, whereas 40 V produced the highest fragmentation, giving in most cases fragmentation patterns similar to those obtained with MS/MS. By increasing the cone voltage to 60 V, pseudomolecular ions of pesticides underwent strong fragmentation to the point that only diagnostic fragment ions were present in the CID spectra.[141] Pesticides of the parathion group underwent strong fragmentation at low extraction voltages, giving fragment ions with low m/z as base peaks. For these pesticides, there were no differences in the formation of fragment ions under cone voltages of 20 and 40 V, probably because an extraction voltage of 20 V was sufficient to fragment the molecules. Table 6 indicates the fragment

Table 6 Characterization of OPPs with LC/APCIMS and PI at extraction voltages of 40 and 20 V, with percentage of abundance of each ion and tentative identification

Compound	m/z	Abundance (%)		Tentative identification
		40 V	20 V	
Acephate	49	100	0	$[CH_3SH + H]^+$ or $[HPO + H]^+$
(MW 183)	95	68	0	$[(CH_3S)POH]^+$
	125	20	0	$[(CH_3S)(CH_3O)PO]^+$
	143	38	100	$[(CH_3S)(CH_3O)PO(OH) + H]^+$
	184	0	10	$[M + H]^+$
	206	0	16	$[M + Na]^+$
Azinphos-ethyl	77	67	0	$[C_6H_4 + H]^+$
(MW 345)	105	35	8	$[C_6H_4N_2 + H]^+$ or $[C_6H_4CO + H]^+$
	132	100	58	$[C_6H_4N_3CH_2 + H]^+$
	160	15	100	$[M - ((OCH_2CH_3)_2PS_2)]^+$
	368	0	11	$[M + Na]^+$
Fenitrothion	109	0	25	$[(CH_3O)_2PO]^+$
(MW 277)	122	72	30	$[C_6H_4CH_3S]^+$
	125	70	62	$[(CH_3O)_2PS]^+$
	138	100	100	$[C_6H_4CH_3NO_2 + H]^+$
	248	11	78	$[M - NO + H]^+$

(*continued overleaf*)

Table 6 (*continued*)

Compound	*m/z*	Abundance (%)		Tentative identification
		40 V	20 V	
	262	0	12	$[M - CH_3]^+$
	278	0	7	$[M + H]^+$
	300	0	69	$[M + Na]^+$
Fensulfoton	94	6	0	$[(CH_3CH_2O)P(OH) + H]^+$
(MW 308)	140	23	0	$[C_6H_4(CH_3S)OH]^+$
	157	100	14	$[C_6H_4(CH_3S)(OH)_2]^+$
	173	57	0	$[C_6H_4(CH_3S)(OH)(SH)]^+$
	219	38	0	$[C_6H_4(CH_3S)(OH)(POS)]^+$
	235	75	0	$[C_6H_4(CH_3S)(OH)(POS)O)]^+$
	281	25	14	$[M - (CH_2)_2]^+$
	309	9	100	$[M + H]^+$
	331	0	21	$[M + Na]^+$
Fenthion	153	68	0	$[M - (CH_3O)_2PS]^+$
(MW 278)	216	100	0	$[C_6H_3CH_3SCH_3POS]^+$ or $[M - (CH_3)_2S]^+$
	231	93	21	$[M - CH_3S]^+$
	279	0	100	$[M + H]^+$
	301	0	15	$[M + Na]^+$
Metamidophos	47	11	0	$[PO]^+$ or $[CH_3S]^+$
(MW 141)	79	5	0	$[POS]^+$
	94	100	68	$[(CH_3S)PO]^+$ or $[M - CH_3S]^+$
	125	25	28	$[(CH_3S)(CH_3O)PO]^+$
	142	0	100	$[M + H]^+$
Naled	127	n.d.	100	$[(CH_3O)_2P(OH)_2]^+$
(MW 380)	177	n.d.	16	n.i.
Paraoxon-methyl	65	27	0	$[P(OH)_2]^+$
(MW 247)	80	45	0	$[(CH_3O)_2OH + H]^+$
	93	100	24	$[(CH_3O)_2P]^+$
	109	33	100	$[(CH_3O)_2PO]^+$
	248	6	30	$[M + H]^+$
Parathion-methyl	65	40	0	$[P(OH)_2]^+$
(MW 263)	93	59	0	$[(CH_3O)_2P]^+$
	109	79	100	$[(CH_3O)_2PO)]^+$
	125	100	0	$[(CH_3O)_2PS]^+$
	154	15	17	$[C_6H_4NO_2S]^+$
	234	8	62	$[M - NO + H]^+$
	264	0	19	$[M + H]^+$
Trichlorfon	79	12	9	$[(CH_3O)POH]^+$
(MW 257)	109	100	45	$[(CH_3O)_2PO]^+$
	111	0	100	$[(CH_3O)_2PO + H]^+$
	221	0	35	$[M - Cl]^+$
	257	0	46	$[M]^+$
Vamidothion	58	41	0	$[CH_3NHCO]^+$
(MW 287)	87	59	0	$[CH_3NHCO(CH_2CH_3)]^+$
	118	100	0	$[CH_3NHCO(CHSCH_3)]^+$
	146	89	100	$[M - (CH_3O)_2POS]^+$
	288	0	6	$[M + H]^+$
	310	25	38	$[M + Na]^+$
Vamidothion sulfoxide	58	27	5	$[CH_3NHCO]^+$
(MW 303)	86	0	100	$[CH_3NHCO(CH_2CH_3) + H]^+$
	87	38	0	$[CH_3NHCO(CH_2CH_3)]^+$
	109	54	0	$[(CH_3O)_2PO]^+$
	125	8	0	$[(CH_3O)_2PS]^+$
	162	0	30	$[M - (CH_3O)_2POS]^+$
	169	55	41	$[(CH_3O)_2POS(CH_2)_2]^+$
	201	8	33	$[(CH_3O)_2POS(CH_2)_2SO]^+$
	241	100	27	$[M - (CH_3O)_2]^+$
	304	0	10	$[M + H]^+$
	326	0	68	$[M + Na]^+$

n.i. = not identified.

ions of several OPPs obtained by LC/APCIMS in the PI mode of operation with extraction voltages of 20 and 40 V. It can be observed that the fragmentation pattern differs in the two conditions and, in general, fragment ions of lower m/z values are formed at a 40-V extraction voltage.

Previous studies on OPP analysis with LC/APCIMS include those of Kawasaki et al.,[142] who obtained structural information under APCI conditions and DLs varying between 2 and 50 ng under SIM conditions. The optimization and validation of on-line SPE with LC/APCIMS for OPP analysis has been performed by Lacorte et al.,[30] who report DLs at the low nanograms per liter level and good recovery and reproducibility values. Figure 6(a–c) shows a typical chromatographic profile of a 100-mL groundwater sample spiked with pesticides at $1\,\mu g\,L^{-1}$ and analyzed under full-scan and SIM in PI and NI modes of ionization. An interesting feature of on-line SPE with LC/APCIMS is that it can be operated under full-scan conditions as an ''alarm'' system to detect organic contaminants in water samples, since pesticides are found in natural waters at levels between 1 and $10\,\mu g\,L^{-1}$. In general, DLs of OPPs are around $500\,ng\,L^{-1}$ in full-scan and $1–60\,ng\,L^{-1}$ in the SIM mode. The excellent sensitivity, even in the full-scan mode, combined with the abundant structural information provided by the APCI interface makes the technique suitable for the screening of pesticides in environmental samples in the parts per trillion to billion range. The APCI interface can also be used for structural elucidation studies. Several TPs of temephos and fenthion[143] were identified by increasing the extraction voltage of the APCI interface. At a 40-V extraction voltage, oxo analogues, sulfoxides and the isomeric forms of the above-mentioned pesticides were unequivocally identified. This means that in the near future, this technique can be implemented in routine work for the analysis of real water samples, since it can detect both parents and TPs of pesticides at the nanograms per liter level. However, more research is needed on the characterization of oxo analogues and other metabolites before optimal analytical conditions can be defined.

LC/APCIMS under NI normally generates $[M - H]^-$ ion or specific fragment ions as the base peak. The use of the NI mode permits the detection of most electronegative compounds and better selectivity and sensitivity are obtained. However, many OPPs can undergo strong fragmentation, even at a 20-V extraction voltage, and produce in some cases a single ion, which is often a diagnostic ion, so that compound elucidation is difficult.

In summary, future trends will lead to the more extensive use of API interfaces for environmental applications owing their high sensitivity and the possibility of obtaining structural information. However, the main operational problem with APCI and in general all API

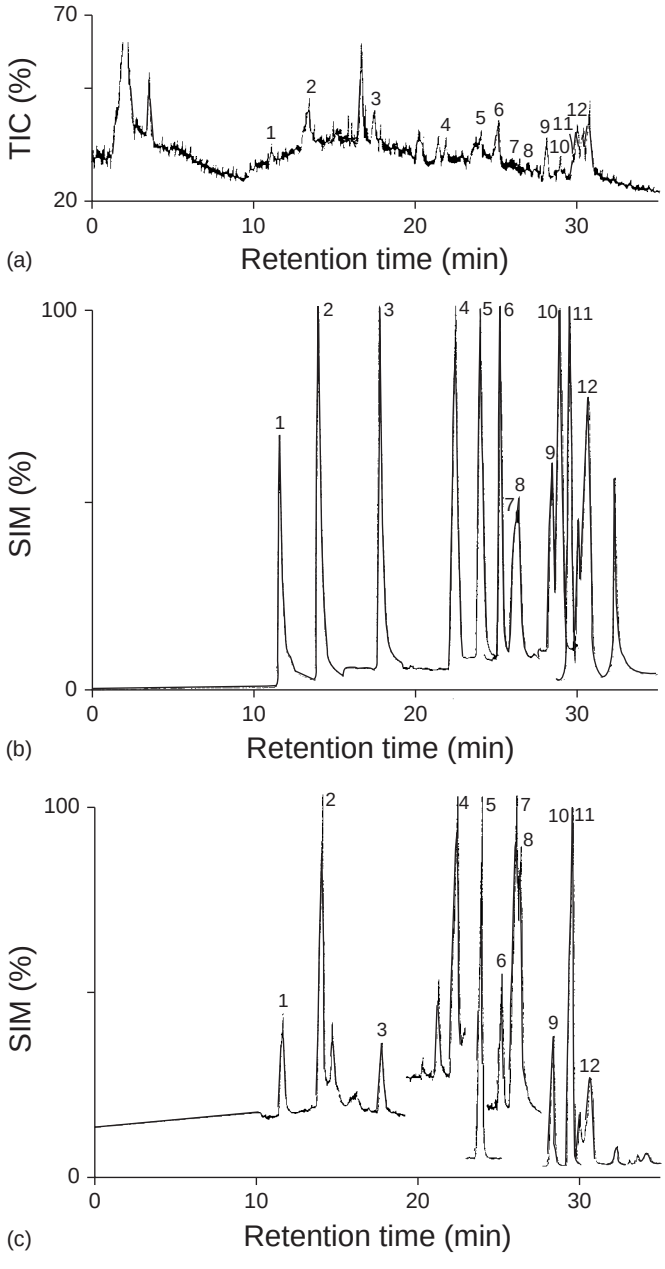

Figure 6 On-line SPE using the OSP-2 followed by LC/APCI/MS of 100 mL of groundwater sample spiked with OPPs at the $1\,\mu g\,L^{-1}$ level in (a) PI mode and full-scan conditions, (b) PI mode and time-scheduled SIM and (c) NI mode and time-scheduled SIM. Peak identification: (1) mevinphos-*cis*; (2) mevinphos-*trans*; (3) dichlorvos; (4) azinphos-methyl; (5) parathion-methyl; (6) malathion; (7) fenitrothion; (8) azinphos-ethyl; (9) parathion-ethyl; (10) fenthion; (11) chlorfenvinfos; and (12) diazinon. (Reprinted from S. Lacorte, D. Barceló, 'Determination of Parts Per Trillon Levels of Organophosphorus Pesticides in Groundwater by Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry Using Positive and Negative Ion Modes of Operation', *Anal. Chem.*, **68**, 2464–2470 (1996). Copyright (1996) American Chemical Society.)

interfaces arises from the high cleaning requirements of the sample cone and skimmer assembly, especially when dirty water samples are processed. However, it should be remembered that this is a common problem with most of the LC/MS interfaces currently available. To achieve maximum sensitivity in API techniques, the system should be calibrated very accurately, and therefore it is essential to have a clean interface. In API interfaces, the system response should be checked after every cleaning step, but even with a clean interface, reproducibility values above 25% are common. Better values are obtained if an internal standard is used.

7 VALIDATION OF ANALYTICAL METHODS

As a result of variations in the results reported by various laboratories, method testing, based on repeatability, reproducibility and calibration data, should be done routinely to evaluate the performance of an analytical method. Any analytical method should be precisely developed to provide reliable quantitative results. Therefore, validation becomes a key issue before method approval and implementation can be done. Any analytical method should provide statistical documentation of its capacity for rendering accurate results. In environmental chemistry, the achievement of an accurate determination is a challenge since (i) the analytes of interest are often at a very low concentration level, normally in the parts per trillion to billion range, (ii) environmental matrices can be very diverse and complex, from drinking and groundwater samples, which represent the easiest matrix, to surface or wastewater, which pose some problems of analysis owing to the presence of humic and fulvic material and other interferences; moreover, food and sediment samples also represent a problem in analysis, owing to the interfering matrix or sugar and fat content; and (iii) a highly efficient separation and detection technique is required to determine numerous isomers or congeners and compounds with different polarities and physicochemical properties. For such reasons, it is obligatory to use validated methods since more and more different laboratories need to compare the levels of OPPs in different matrices for QA purposes. To compare results, one has to prove that the method used is accurate and precise and that these parameters hold with time.

To test the performance and validate a method, certified reference materials are used. There are two main types of reference materials. The first type consists of high-purity mixtures of compounds of one chemical class dissolved in a pure organic solvent, and are used for calibration of the analytical system, for identification of the analytes and for spiking the samples, in other words, for method performance. The second type is used to asses the whole analytical procedure, from the extraction of the analytes to the final analysis, which is performed by participating in interlaboratory calibration exercises. The advantage of participating in interlaboratory exercises is that each laboratory can validate a new-sprung extraction and detection method by comparing it with well-known and established techniques, such as GC. Lacorte et al. reported the validation of SPE coupled on-line with LC/DAD, LC/TSPMS and LC/APCIMS for the analysis of OPPs in ground and wastewaters.[30,144] The established procedure for validating an analytical method includes the evaluation of several quality parameters, such as: (i) linearity of the detector for establishing the minimum and maximum concentrations that can be handled; (ii) DL, measured from the calibration curve and taking a signal-to-noise ratio (S/N) of 3 (it should be verified experimentally); (iii) the limit of quantification (LOQ), which is calculated taking an S/N of 10; (iv) precision or repeatability of the technique, which indicates the percentage error made in consecutive analyses, without changing any experimental conditions; for a standard injection, it should be 1–2%, and for a whole analytical procedure (e.g. analysis of waters) it can rise to 10%, which is the allowed error according to the USEPA; (v) reproducibility, or variation of the analytical response between-day precision, keeping all the experimental conditions equal; under controlled conditions (column oven, mobile phase, pH, etc.), the reproducibility should not be much higher than the repeatability, but in practice, it is seen that many MS detectors do not respond in the same way since the ionization efficiency can vary substantially in different days, and to avoid this problem one should work with internal standards; (vi) recoveries, which should be between 70 and 130% for the tested analytes, according to the USEPA; and (vii) accuracy or determination of the relative difference between the experimental value and the true value, which should be checked in interlaboratory calibration exercises.

To summarize, validation of analytical methods is important not only to prove the accuracy and reliability of an analytical method, but also to be able to compare results obtained in different laboratories. Validation is tedious but needs to be done.

8 CONCLUSIONS

The analysis of OPPs in water and food samples include three main steps: extraction, cleanup and detection. The present trend is the use of multiresidue, automated methods for the surveillance of pesticides and their degradation products in environmental samples. For water samples, LLE, which used to be the most

popular method, has been replaced by SPE, which has the advantage that it is less laborious, quicker, diminishes solvent consumption, can be used as storage media and produces more reproducible results. Both on-line and off-line SPE can be nowadays automated by sophisticated equipment such as the Prospekt, the OSP-2 and the ASPEC or Zymark, respectively. Since manipulation of the water sample is minimized, these techniques produce extremely precise and accurate results. Different types of packing materials for SPE (e.g. silica-based C_8 and C_{18}, styrene–divinylbenzene copolymeric materials, immunosorbents and MIPs) provide an extremely selective way to trap compounds of diverse physicochemical properties. The advent of immunosorbents or MIPs will certainly increase the efficiency of a selective extraction both from water and food samples and will minimize the sample cleanup.

Food samples represent a more complex matrix owing to their great versatility and the presence of fats and sugars. Food samples are freeze-dried and pesticides are extracted by Soxhlet or ultrasonic methods. Sample cleanup using GPC or column chromatography with Florisil, alumina or other sorbents is essential to eliminate matrix impurities. Generally, the analysis of food samples requires a high degree of manipulation and there is a need to optimize the analytical technique for each type of food matrix.

In all cases, sampling, transport and storage of the water or food samples should be carefully performed to avoid sample deterioration or degradation of the pesticides. Normally, samples are transported refrigerated and kept preferably at $-20\,^{\circ}C$.

Most commonly, OPPs are analyzed by GC, which is a well known and established technique for OPP monitoring and is the official method adopted by the USEPA. The high sensitivity of GC methods, together with the high separation efficiency and speed of analysis, favors multidimensional analysis and routine monitoring. GC with NPD and ECD are the most common techniques, and analysis has to be performed with two columns of different polarity or using MS detection. In this way, it is possible to confirm the presence of a pesticide in water or food samples. GC/MS methods are sensitive, selective and reproducible and offer structural information. GC/MS with EI produces abundant fragment ions and the analytes can be identified by library spectral matching whereas NCI gives molecular information and is more sensitive and selective for electronegative pesticides. The main disadvantage of GC is that it cannot handle polar, nonvolatile or thermolabile compounds without a previous derivatization step.

In the last few years, new technologies based on HPLC systems have been developed to determine organic pollutants in the environment. UV and DAD are the most common detection devices used for determining priority pesticides in water and food matrices. LC/DAD has the advantage that it gives a UV spectrum that can be used for identification and confirmation purposes. LC can be coupled to MS with different types of interfaces, such as PB, TSP and the recent API techniques. API interfaces have slowly replaced the others since they can produce abundant fragmentation by increasing the electrical field in the intermediate region of the mass analyzer so that CID spectra can be obtained by a single quadrupole. Compared with PB and TSP, API techniques are highly sensitive, with DLs at the picograms level, reproducible and have a linear range over various orders of magnitude. Moreover, provided an efficient preconcentration technique is applied, it is possible to quantify the analytes at the nanograms per liter level, similar to what is obtained with GC. As a consequence, LC/MS is becoming a highly robust technique and the preferred option for the identification and confirmation of OPPs in routine monitoring programs.

With regard to qualitative assurance, all the analytical techniques have to be validated in order to provide reliability of the results. This is important for qualitative control and to allow the comparison of the results provided by different laboratories. The analysis of certified samples is important in order to establish the accuracy of each method and it is necessary before a technique can be implemented in routine monitoring programs concerning OPPs.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detector
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
CEC	Commission of the European Communities
CI	Chemical Ionization
CID	Collision-induced Dissociation
DAD	Diode-array Detection
DL	Detection Limit
ECD	Electron-capture Detector
EI	Electron Ionization
ESI	Electrospray Ionization
ESP	Electrospray
ESP/ISP	Electrospray/Ionspray
ESP/ISPMS	Electrospray/Ionspray Mass Spectrometry
EU	European Union
FID	Flame Ionization Detection
FPD	Flame Photometric Detection

GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/FID	Gas Chromatography/Flame Ionization Detection
GPC	Gel Permeation Chromatography
GUS	Ground Ubiquity Score (Mobility Index)
HPLC	High-performance Liquid Chromatography
HPN	Heated Pneumatic Nebulizer
IDL	Instrument Detection Limit
IS	Immunoaffinity Sorbent
ISP	Ionspray
LC	Liquid Chromatography
LC/APCIMS	Liquid Chromatography Coupled to Mass Spectrometry with Atmospheric Pressure Chemical Ionization
LC/DAD	Liquid Chromatography with Diode-array Detection
LC/ESP/ISPMS	Liquid Chromatography Coupled to Electrospray/Ionspray Mass Spectrometry
LC/MS	Liquid Chromatography with Mass Spectrometric Detection
LC/TSPMS	Liquid Chromatography Coupled to Mass Spectrometry with a Thermospray Interface
LC/TSPMS/MS	Liquid Chromatography Coupled to Tandem Mass Spectrometry with a Thermospray Interface
LLE	Liquid–Liquid Extraction
LOQ	Limit of Quantification
LSE	Liquid–Solid Extraction
MIP	Molecular Imprinted Polymer
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NCI	Negative Chemical Ionization
NI	Negative Ionization
NPD	Nitrogen-phosphorus Detector
OPP	Organophosphorus Pesticide
PB	Particle Beam
PCI	Positive Chemical Ionization
PGC	Porous Graphite Carbon
PI	Positive Ionization
QA	Quality Assurance
QC	Quality Control
SEC	Size-exclusion Chromatography
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
S/N	Signal-to-noise Ratio
SPE	Solid-phase Extraction
SSC	Single Short Column
THF	Tetrahydrofuran
TSP	Thermospray
TP	Transformation Product
UCM	Unidentified Complex Matrix
USEPA	United States Environmental Protection Agency
USNPS	United States National Pesticide Survey
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Dyes, Environmental Analysis of

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Sampling Considerations for Biomonitoring • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Liquid Chromatography in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food • Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Chemometrics **(Volume 11)**
Soft Modeling of Analytical Data

Liquid Chromatography **(Volume 13)**
Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. E. Funari, L. Donati, D. Dandroni, M. Vighi, in *Pesticide Risk in Groundwater*, eds. M. Vighi, E. Funari, CRC Press, Boca Raton, FL, 4–44, 1995.
2. M. Fielding, D. Barceló, A. Helweg, S. Galassi, L. Tortensson, P. Van Zoonen, R. Wolter, G. Angeletti, *Pesticides in Ground and Drinking Water*, Water Pollution Research Report 27, Commission of the European Communities, Brussels, 1–136, 1992.
3. P.H. Pritchard, C.R. Cripe, W.W. Walter, A.W. Bourquin, 'Biotic and Abiotic Degradation Rates of Methyl Parathion in Freshwater and Estuarine Water and Sediment Samples', *Chemosphere*, **16**(7), 1509–1520 (1987).
4. L. Marcheterre, G.G. Choudry, G.R.B. Webster, 'Environmental Photochemistry of Herbicides', *Rev. Environ. Contam. Toxicol.*, **103**, 61–126 (1988).
5. S. Lacorte, D. Barceló, 'Rapid Degradation of Fenitrothion in Estuarine Waters', *Environ. Sci. Technol.*, **28**, 1159–1163 (1994).
6. R. Kimbrough, D.W. Litke, 'Pesticides in Streams Draining Agricultural and Urban Areas in Colorado', *Environ. Sci. Technol.*, **30**, 908–916 (1996).
7. J. Miyamoto, N. Mikami, K. Mihara, Y. Takimoto, H. Kohda, H. Suzuki, 'Biological Activity of Fenitrothion and its Degradation Products', *Pestic. Sci.*, **3**(1), 35–41 (1978).
8. C. Tomlin (ed.), *The Pesticide Manual*, 10th edition, Crop Protection Publication and Royal Society of Chemistry, Cambridge, 1994, reprinted 1995.
9. D. Barceló, M.C. Hennion, *Trace Determination of Pesticides and Their Degradation Products in Water*, Elsevier, Amsterdam, 1997.
10. S. Lacorte, S. Lartiges, P. Garrigues, D. Barceló, 'Degradation of Organophosphorus Pesticides and Their Transformation Products in Estuarine Waters', *Environ. Sci. Technol.*, **29**, 431–438 (1995).
11. S. Lacorte, N. Ehersmann, D. Barceló, 'Persistence of Temephos and its Transformation Products in Rice Crop Field Waters', *Environ. Sci. Technol.*, **30**, 917–923 (1996).
12. Directive 76/464/CEE of the European Union, 1986.
13. D.J. Munch, R.L. Graves, R.A. Maxey, T.M. Engel, 'Methods Development and Implementation for the National Pesticide Survey', *Environ. Sci. Technol.*, **24**(10), 1446–1451 (1990).
14. D.J. Munch, Ch.P. Frebis, 'Analyte Stability Studies Conducted During the National Pesticide Survey', *Environ. Sci. Technol.*, **26**, 921–925 (1992).
15. D. Barceló, 'Environmental Protection Agency and Other Methods for the Determination of Priority Pesticides and Their Transformation Products in Water', *J. Chromatogr.*, **643**, 117–143 (1993).
16. IPPC Directive 96/61/EC, OJ No. L 257, 1996.
17. D. Barceló, M.C. Hennion, 'Sampling of Polar Pesticides from Water Matrices', *Anal. Chim. Acta*, **338**, 3–18 (1997).
18. *EPA Method 553*, US Environmental Protection Agency, Cincinnati, OH, 1985.
19. D. Barceló, in *Environmental Analysis: Techniques, Application and Quality Assurance*, Elsevier, Amsterdam, 1993.
20. V. Pichon, L. Chen, M.C. Hennion, R. Daniel, A. Martel, F.L. Goffic, J. Abian, D. Barceló, 'Preparation and Evaluation of Immunosorbents for Selective Trace Enrichment of Phenylurea and Triazine Herbicides in Environmental Waters', *Anal. Chem.*, **67**, 2451–2460 (1995).
21. S. Lacorte, C. Molina, D. Barceló, 'Screening of Organophosphorus Pesticides in Environmental Matrices by Various Gas-chromatographic Techniques', *Anal. Chim. Acta*, **281**, 71–84 (1993).
22. D.W. McLeese, V. Zitko, D.B. Sergeant, 'Uptake and Excretion of Fenitrothion by Clams and Mussels', *Bull. Environ. Contam. Toxicol.*, **22**(6), 800–806 (1979).
23. S. Lacorte, D. Barceló, 'Validation of an Automated Precolumn Exchange System (PROSPEKT) Coupled to Liquid Chromatography with Diode Array Detection. Application to the Determination of Pesticides in Natural Waters', *Anal. Chim. Acta*, **296**, 223–234 (1994).
24. D. Barceló, 'Application of Thermospray Liquid Chromatography/Mass Spectrometry for Determination of Organophosphorus Pesticides and Trialkyl and Triaryl Phosphates', *Biomed. Environ. Mass Spectrom.*, **17**(5), 363–369 (1988).
25. D. Barceló, 'Occurrence, Handling and Chromatographic Determination of Pesticides in the Aquatic Environment', *Analyst*, **116**(7), 681–689 (1991).
26. D. Volmer, K. Levsen, 'Mass Spectrometric Analysis of Nitrogen and Phosphorus Containing Pesticides by Liquid Chromatography–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**, 655–675 (1994).
27. R.D. Voyksner, C.S. Smith, P.C. Knox, 'Optimization and Application of Particle Beam High Performance Liquid Chromatography/Mass Spectrometry. Two Compounds of Pharmaceutical Interest', *Biomed. Environ. Mass Spectrom.*, **19**(9), 523–534 (1990).
28. C. Cresenzi, A. Di Corcia, S. Marchese, R. Samperi, 'Determination of Acidic Pesticides in Water by a Benchtop Electrospray Liquid Chromatography Mass Spectrometer', *Anal. Chem.*, **67**, 1968–1975 (1995).

29. C. Molina, M. Honing, D. Barceló, 'Determination of Organophosphorus Pesticides in Water by Solid Phase Extraction Followed by Liquid Chromatography/High Flow Pneumatically Assisted Electrospray Mass Spectrometry', *Anal. Chem.*, **66**, 4444–4449 (1994).
30. S. Lacorte, D. Barceló, 'Determination of Parts Per Trillon Levels of Organophosphorus Pesticides in Groundwater by Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry Using Positive and Negative Ion Modes of Operation', *Anal. Chem.*, **68**, 2464–2470 (1996).
31. T.C. Wang, R.A. Lenahan, J.W. Tucker, Jr, 'Deposition and Persistence of Aerially-applied Fenthion in a Florida Estuary', *Bull. Environ. Contam. Toxicol.*, **38**(2), 226–231 (1987).
32. D. Barceló, C. Porte, J. Cid, J. Albaigés, 'Determination of Organophosphorus Compounds in Mediterranean Coastal Waters and Biota Samples using Gas Chromatography with Nitrogen Phosphorus and Chemical Ionization Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **38**, 199–209 (1990).
33. B.K. Afgan, P.E. Belliveau, R.H. Larose, J.F. Ryan, 'Improved Methods for Determination of Trace Quantities of Phenols in Natural Waters', *Anal. Chim. Acta*, **71**, 355–366 (1974).
34. A. Neicheva, E. Kovacheva, G. Marudov, 'Determination of Organophosphorus Pesticides in Apples and Water by Gas–Liquid Chromatography with Electroncapture Detection', *J. Chromatogr.*, **437**(1), 249–253 (1988).
35. K. Abrahamsson, T.M. Xie, 'Direct Determination of Trace Amounts of Chlorophenols in Fresh Water, Wastewater and Seawater', *J. Chromatogr.*, **279**, 199–208 (1983).
36. P.J.M. Kwakman, D.A. Kamminga, U.A.Th. Brinkman, G.J. de Jong, 'Sensitive Liquid Chromatographic Determination of Alkylphenols, Nitrophenols and Chlorophenols by Precolumn Derivatization with Dansyl Chloride Postcolumn Photolysis and Peroxyoxalate Chemioluminiscence Detection', *J. Chromatogr.*, **553**(1/2), 345–356 (1991).
37. D.R. Green, D.L. Le Pape, 'Stability of Hydrocarbons Samples on Solid Phase Extraction Columns', *Anal. Chem.*, **59**, 699–703 (1987).
38. R. Bagnati, E. Benfenati, E. Davoli, R. Fanelli, 'Screening of 21 Pesticides in Water by Single Extraction With C_{18} Silica Bonded Phase Columns and HRGC–MS', *Chemosphere*, **17**(1), 59–66 (1988).
39. H.J. Borburgh, W.E. Hammers, 'Extraction of Organophosphorus Pesticides from Water Samples on Bakerbond SPE Octadecyl Cartridges at the ng/L Level', *Toxicol. Environ. Chem.*, **35**, 79–86 (1992).
40. K. Berkane, G.E. Caissie, V.N. Mallet, 'The Use of Amberlite XAD-2 Resin for the Quantitative Recovery of Fenitrothion from Water. A Preservation Technique', *J. Chromatogr.*, **139**(2), 386–390 (1977).
41. H.A. James, C.P. Steel, Y. Wilson, 'Impurities Arising from the Use of XAD-2 Resin for the Extraction of Organic Pollutants in Drinking Water', *J. Chromatogr.*, **208**(1), 89–95 (1981).
42. G. Durand, S. Chiron, V. Bouvot, D. Barceló, 'Use of Extraction Disks for Trace Enrichment of Various Pesticides from River and Seawater Samples', *Int. J. Environ. Anal. Chem.*, **49**, 31–42 (1992).
43. D. Barceló, S. Chiron, S. Lacorte, E. Martínez, J.S. Salau, M.C. Hennion, 'Solid-phase Sample Preparation and Stability of Pesticides in Water Using Empore Disks', *Trends Anal. Chem.*, **13**, 352–361 (1994).
44. V. Pichon, M. Charpak, M.C. Hennion, 'Multiresidue Analysis of Pesticides Using New Laminar Extraction Disks and Liquid Chromatography and Application to the French Priority List', *J. Chromatogr. A*, **795**, 83–92 (1998).
45. M.W. Nielen, R.W. Frei, U.A.Th. Brinkman, in *Selective Sample Handling and Detection in High Performance Liquid Chromatography*, eds. R.W. Frei, K. Zech, Elsevier, Amsterdam, Chapter 1, 5, 1988.
46. U.A.Th. Brinkman, 'On-line Sample Treatment for or via Column Liquid Chromatography', *J. Chromatogr. A*, **665**, 217–234 (1994).
47. V. Pichon, M.C. Hennion, 'Determination of Pesticides in Environmental Waters by Automated On-line Trace Enrichment and Liquid Chromatography', *J. Chromatogr. A*, **665**, 269–281 (1994).
48. S. Lacorte, J.J. Vreuls, J.S. Salau, D. Barceló, 'Strategies for Pesticide Sampling and Analysis', *J. Chromatogr. A*, **795**, 71–81 (1998).
49. P.M.J. Kwakman, J.J. Vreuls, U.A.Th. Brinkman, R.T. Ghijsen, 'Determination of Organophosphorus Pesticides in Aqueous Samples by Online Membrane Disk Extraction and Capillary Gas Chromatography', *Chromatographia*, **34**(1/2), 41–47 (1992).
50. J.J. Vreuls, R.T. Ghijsen, G.J. de Jong, U.A.Th. Brinkman, 'Drying Step for Introduction of Water-free Desorption Solvent into a Gas Chromatograph after Online Liquid Chromatographic Trace Enrichment of Aqueous Samples', *J. Chromatogr.*, **625**(2), 237–245 (1992).
51. Y. Picó, J.J. Vreuls, R.T. Ghijsen, U.A.Th. Brinkman, 'Drying Agents for Water-free Introduction of Desorption Solvents into a GC after Online SPE of Aqueous Samples', *Chromatographia*, **38**(7/8), 461–469 (1994).
52. A.J. Bullterman, J.J. Vreuls, R.T. Ghijsen, U.A.Th. Brinkman, 'Selective and Sensitive Detection of Organic Contaminants in Water Samples by On-line Trace Enrichment–Gas Chromatography–Mass Spectrometry', *J. High Resolut. Chromatogr.*, **16**(7), 397–403 (1993).
53. S. Lacorte, N. Ehresmann, D. Barceló, 'Stability of Organophosphorus Pesticides on Disposable Solid-phase Extraction Precolumns', *Environ. Sci. Technol.*, **29**, 2834–2841 (1995).

54. W.A. Minnaard, A.C. Hogenboom, U.K. Malmqvist, W.M.A. Niessen, U.A.Th. Brinkman, 'Rapid Target Analysis by On-line Single Short Column Liquid Chromatography/Thermospray Tandem Mass Spectrometry', *Rapid. Commun. Mass Spectrom.*, **10**, 1569–1574 (1996).
55. A.C. Hogenboom, U.K. Malmqvist, K. Nolkranz, J.J. Vreuls, U.A.Th. Brinkman, 'Rapid Analysis of Organic Micropollutants in Environmental Samples by Trace Enrichment and Liquid Chromatography on a Single Short Column', *J. Chromatogr. A*, **759**, 55–74 (1997).
56. A.C. Hogenboom, P. Speksnijder, R.J. Vreeken, W.M. A. Niessen, U.A.Th. Brinkman, 'Rapid Target Analysis of Microcontaminants by On-line Single Short Column Liquid Chromatography Combined with Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry', *J. Chromatogr. A*, **777**, 81–90 (1997).
57. I. Liska, E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'Trace Enrichment of Polar Pollutants from Water and the Influence of Humic Substances', *Chromatographia*, **37**(1/2), 13–19 (1993).
58. M.C. Hennion, V. Pichon, 'Solid-phase Extraction of Organic Pollutants from Water', *Environ. Sci. Technol.*, **28**(13), 576A–583A (1994).
59. M.W. Powell, 'Development of a Binary Solid Phase Extraction Cartridge for Use in Screening Water Samples for Organic Pollutants', *J. Chromatogr. A*, **697**(1/2), 101–105 (1995).
60. J.S. Fritz, P.L. Dumont, L.W. Smith, 'Methods and Materials for Solid-phase Extraction', *J. Chromatogr.*, **691**(1/2), 133–140 (1995).
61. A. Di Corcia, M. Marchetti, 'Multiresidue Method for Pesticides in Drinking Water Using a Graphitized Carbon Black Cartridge Extraction and Liquid Chromatographic Analysis', *Anal. Chem.*, **63**, 580–585 (1991).
62. A. Di Corcia, S. Marchese, R. Samperi, 'Selective Determination of Phenols in Water by a Two-trap Tandem Extraction System followed by Liquid Chromatography', *J. Chromatogr.*, **642**(1/2), 175–184 (1993).
63. E. Forgacs, T. Cserháti, 'Retention Strength and Selectivity of Porous Graphitized Carbon Columns. Theoretical Aspects and Practical Applications', *Trends Anal. Chem.*, **14**(1), 23–29 (1995).
64. J. Matusi, M. Okada, M. Tsurokoa, T. Takeuchi, 'Solid Phase Extraction of Triazine Herbicides using a Molecular Imprinted Receptor', *Anal. Commun.*, **34**, 85–87 (1997).
65. V. Pichon, L. Chen, N. Durand, F. Le Goffic, M.C. Hennion, 'Selective Trace Enrichment on Immunosorbents for the Multiresidue Analysis of Phenylurea and Triazine Herbicides', *J. Chromatogr. A*, **725**, 107–119 (1996).
66. I. Ferrer, V. Pichon, M.C. Hennion, D. Barceló, 'Automated Sample Preparation with Extraction Columns by Means of Anti-isoproturon Immunosorbents for the Determination of Phenylurea Herbicides in Water Followed by Liquid Chromatography–Diode Array Detection and Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **777**, 91–98 (1997).
67. M. Muldoon, L.H. Stanker, 'Moleculary Imprinted Solid-phase Extraction of Atrazine from Beef Liver Extracts', *Anal. Chem.*, **69**, 803–808 (1997).
68. D. Barceló, M.C. Hennion, 'On-line Sample Handling Strategies for the Trace Level Determination of Pesticides and Their Degradation Products in Environmental Waters', *Anal. Chim. Acta*, **318**, 1–41 (1995).
69. P. Subra, M.C. Hennion, R. Rosset, R.W. Frei, 'Recovery of Organic Compounds from Large Volume Aqueous Samples Using On-line Liquid Chromatographic Preconcentration Techniques', *J. Chromatogr.*, **456**, 121–141 (1988).
70. M.C. Hennion, V. Coquart, 'Comparison of Reverse-phase Extraction Sorbents for the On-line Trace Enrichment of Polar Organic Compounds in Environmental Aqueous Samples', *J. Chromatogr.*, **642**(1–2), 211–224 (1993).
71. J. Poerschmann, U. Stottmeister, 'Methodical Investigation of Interactions Between Organic Pollutants and Humic Organic Material in Coal Wastewaters', *Chromatographia*, **36**, 207–211 (1993).
72. S.A. Senseman, T.L. Lavy, J.D. Mattice, 'Influence of Dissolved Humic Acid and Ca-Montmorillonte Clay on Pesticide Extraction Efficiency from Water Using Solid-phase Extraction Disks', *Environ. Sci. Technol.*, **29**(10), 2647–2653 (1995).
73. R.A. Lovell, D.G. McChesney, W.D. Price, 'Organohalogen and Organophosphorus Pesticides in Mixed Feed Rations: Findings from the FDA's Domestic Surveillance During Fiscal Years 1989–1994', *J. AOAC Int.*, **79**(2), 544–548 (1996).
74. W.O. Lee, M.L.M. Law, S.K. Wong, 'Determination of Metamidophos Residues in Food Remnants', *Food Addit. Contam.*, **13**(6), 687–694 (1996).
75. M. Miyahara, Y. Okada, H. Takeda, G. Aoki, A. Kobayashi, 'Multiresidue Procedures for the Determination of Pesticides in Food Using Capillary Gas Chromatography, Flame Photometric and Mass Spectrometric Techniques', *J. Agric. Food. Chem.*, **42**(12), 2795–2802 (1994).
76. H. Mallatou, C.P. Pappas, E. Kondyli, T.A. Albanis, 'Pesticide Residues in Milk and Cheese from Greece', *Sci. Total Environ.*, **196**(2), 111–117 (1997).
77. A. DiMuccio, P. Pelosi, I. Camoni, D. Attard-Barbini, R. Dommarco, T. Generali, A. Ausili, 'Selective, Solid Matrix Dispersion Extraction of Organophosphate Pesticides from Milk', *J. Chromatogr. A*, **754**(1/2), 497–506 (1996).
78. A.M. Gillespie, S.L. Daly, D.M. Gilvydis, F. Schneider, S.M. Walters, 'Multicolumn Solid Phase Extraction Cleanup of Organophosphorus and Organochlorine Pesticide Residues in Vegetable Oils and Butterfat', *J. AOAC Int.*, **78**(2), 431–437 (1995).

79. C. Rizos-Lentza, E.J. Avramides, 'Organophosphorus Insecticides Residues in Virgin Greek Olive Oil, 1988–1990', *Sci. Essex*, **32**(2), 161–171 (1990).
80. E. Viana, J.C. Molto, G. Font, 'Optimization of Matrix Solid Phase Dispersion Method for the Analysis of Pesticides Residues in Vegetables', *J. Chromatogr. A*, **754**(1/2), 437–444 (1996).
81. A. Aguera, M. Contreras, R.A. Fernandez-Alba, 'Gas Chromatographic Analysis of Organophosphorus Pesticides of Horticultural Concern', *J. Chromatogr. A*, **655**(2), 293–300 (1993).
82. F.J. Schenick, R. Wagner, M.K. Hennessy, J.L. Okrasinski, 'Screening Procedure for Organochlorine and Organophosphorus Pesticide Residues in Milk Using Matrix Solid Phase Dispersion (MSPD) Extraction and Gas Chromatographic Determinations', *J. AOAC Int.*, **77**(4), 1036–1040 (1994).
83. H.G. Beasley, J.H. Skerritt, A.S. Hill, J.M. Desmarchelier, 'Rapid Field Test for the Organophosphorus Pesticides, Fenitrothion and Pirimiphos-methyl. Reliable Estimates of Residues in Stored Grain', *J. Stored Prod. Res.*, **29**(4), 357–369 (1993).
84. S.M. Lee, P.L. Wylie, 'Comparison of the Atomic Emission Detector to Other Element Sensitive Detectors for the Gas Chromatographic Analysis of Pesticide Residues', *J. Agric. Food. Chem.*, **39**(12), 2192–2199 (1991).
85. H.-P. Thier, H. Zeumer (eds.), *Manual of Pesticide Residue Analysis*, VCH, Weinheim, Vol. I, 1987.
86. G. Durand, R. Forteza, D. Barceló, 'Determination of Chlorotriazine Herbicides, Their Dealkylated Degradation Products and Organophosphorus Pesticides in Soil Samples by Means of two Different Clean-up Procedures', *Chromatographia*, **28**, 597–604 (1989).
87. A.S. Chau, B.K. Afghan, *Analysis of Pesticides in Water*, CRC Press, Boca Raton, FL, Vols. I, II and III, 1982.
88. A. Chovelon, G. Lee, C. Gulayets, Y. Hoyano, E. McGuiness, J. Moore, S. Ramamoorthy, P. Singer, 'Pesticide and PCB levels in Fish from Alberta (Canada)', *Chemosphere*, **13**(1), 19–32 (1984).
89. *Organophosphorus Pesticides in River and Drinking Water, 1980, Tentative Method*, HM Stationery Office, London, 1–17, 1983.
90. G.J. Sharp, J.G. Brayan, S. Dilli, P.R. Haddad, J.M. Desmarchieler, 'Extraction, Clean-up and Chromatographic Determination of Organophosphate, Pyrethroid and Carbamate Insecticides in Grain and Grain Products', *Analyst*, **113**(10), 1493–1507 (1988).
91. H.S. Dorea, J.L. Tadeo, C. Sanchez-Brunete, 'Determination of Organophosphorus Pesticides Residues in Fruits by Gas Chromatography with ITD and NPD Detection', *Chromatographia*, **43**(7/8), 380–386 (1996).
92. T.A. Pressley, J.E. Longbottom, 'The Determination of Organophosphorus Pesticides in Industrial and Wastewaters', *Method 614*, US Environmental Protection Agency, Cincinnati, OH, 1–24, 1982.
93. H. Steinwandter, 'Contributions to the Application of Gel Chromatography in Residue Analysis', *Fresenius' J. Anal. Chem.*, **331**(5), 499–502 (1988).
94. J.F. Lawrence, 'Analytical Methodology for Organophosphorus Pesticides Used in Canada', *Int. J. Environ. Anal. Chem.*, **29**(4), 289–303 (1987).
95. A. Sannino, P. Mambriani, M. Bandini, L. Bolzoni, 'Multiresidue Method for Determination of Organophosphorus Insecticide Residues in Fatty Processed Foods by Gel Permeation Chromatography', *J. AOAC Int.*, **78**(6), 1502–1512 (1995).
96. U. Ugyun, 'Determination of Organophosphorus Pesticide Residues in Carrot Using Gel Permeation Chromatography', *J. Liq. Chromatogr. Relat. Technol.*, **20**(5), 771–777 (1997).
97. H. Obana, K. Kikuchi, M. Okihashi, S. Hori, 'Determination of Organophosphorus Pesticides in Foods Using Accelerated Solvent Extraction System', *Analyst*, **122**(3), 217–220 (1997).
98. J.J. Vreuls, R.J.J. Swen, V.P. Goudriaan, M.A.T. Kerkhoff, G.A. Jongenotter, U.A.Th. Brinkman, 'Automated On-gel Permeation Chromatography for the Determination of Organophosphorus Pesticides in Olive Oil', *J. Chromatogr. A*, **750**(1/2), 275–286 (1996).
99. A. Angioni, M. Melis, E. Minelli, F. Pirisi, 'Simplified Multiresidue Method for the Determination of Organophosphorus Insecticides in Olive Oil', *J. Chromatogr. A*, **761**(1/2), 327–331 (1997).
100. A.M. Gillespie, S.M. Walters, 'HPLC Silica Column Fractionation of Pesticides and PCB from Butterfat', *J. Liq. Chromatogr.*, **9**(10), 2111–2141 (1986).
101. J.A. Casanova, 'Use of Solid Phase Extraction Disks for Analysis of Moderately Polar and Non-polar Pesticides in High Moisture Foods', *J. AOAC Int.*, **79**(4), 936–940 (1996).
102. G. Pavoni, E. Errani, 'Rapid Solid-phase Extraction/Gas Chromatography Method for Determination of Organophosphorus Pesticides Residues in Cereal Flours', *Boll. Chim. Ig. Parte Sci.*, **47**(S3), 143–145 (1996).
103. F.J. Schenck, R. Wagner, 'Screening Procedures for Organochlorine and Organophosphorus Pesticide Residues in Eggs Using a Solid-phase Extraction Cleanup and Gas Chromatographic Detection', *Food Addit. Contam.*, **12**(4), 535–541 (1995).
104. C.M. Torres, Y. Picó, M.J. Redondo, J. Mañes, 'Matrix Solid-phase Dispersion Extraction Procedure for Multiresidue Pesticide Analysis in Oranges', *J. Chromatogr. A*, **719**(1), 95–103 (1996).
105. M.L. Hopper, 'Extraction and Clean-up of Organochlorine and Organophosphorus Pesticide Residues in Fats by Supercritical Fluid Techniques', *J. AOAC Int.*, **80**(3), 639–646 (1997).
106. 'Organophosphorus Pesticides', *EPA Method 8140*, US Environmental Protection Agency, Cincinnati, OH, Revision, 1986.

107. L. Graves, 'Determination of Nitrogen and Phosphorus Containing Pesticides in Water by Gas Chromatography with a Nitrogen and Phosphorus Detector', *Method 507, Revision 2.0*, US Environmental Protection Agency, Cincinnati, OH, 1989.
108. D. Barceló, G. Durand, R.J. Vreeken, G.J. de Jong, H. Lingeman, U.A.Th. Brinkman, 'Evaluation of Eluents in Thermospray Liquid Chromatography–Mass Spectrometry for Identification and Determination of Pesticides in Environmental Samples', *J. Chromatogr.*, **553**(1/2), 311–328 (1991).
109. G. Durand, F.S. Banchez-Baeza, A. Messeguer D. Barceló, 'Thermospray Mass Spectrometry of Phosphorothionate Pesticides and Their Oxygen Analogs', *Biol. Mass Spectrom.*, **20**(1), 3–10 (1991).
110. H.J. Stan, S. Heil, 'Two-dimensional Capillary Gas Chromatography with Three Selective Detectors as a Valuable Tool in Residue Analysis – State-of-Art', *Fresenius' J. Anal. Chem.*, **339**(1), 34–39 (1991).
111. H.J. Stan, B. Abraham, J. Jung, M. Keller, K. Teinland, 'Detection of Organophosphorus Insecticides by GLC/MS', *Fresenius' J. Anal. Chem.*, **287**(4/5), 271–285 (1977).
112. G. Durand, M. Mansour, D. Barceló, 'Identification and Determination of Fenitrothion Photolysis Products in Water–Methanol by Gas Chromatography–Mass Spectrometry', *Anal. Chim. Acta*, **262**(1), 167–178 (1992).
113. J.M. Desmarchelier, M.J. Lacey, in *Mass Spectrometry in Environmental Sciences*, eds. F.W. Karasek, O. Hutzinger, S. Safe, Plenum Press, New York, 455–474, 1985.
114. A.K. Singh, D.W. Hewetson, K.C. Jordon, M. Ashraf, 'Analysis of Organophosphorus Insecticides in Biological Samples by Selective Ion Monitoring Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **369**(1), 83–96 (1986).
115. J.P.G. Wilkins, A.R.C. Hill, D.F. Lee, 'Organophosphorus Sulfides, Sulfoxides and Sulfones. Characterization by Gas Chromatography–Mass Spectrometry', *Analyst*, **110**(9), 1045–1051 (1985).
116. H.J. Stan, G. Kellner, 'Negative Chemical Ionization Mass Spectrometry of Organophosphorus Pesticides', *Biomed. Mass Spectrom.*, **9**(11), 483–492 (1982).
117. H.J. Stan, G. Kellner, 'Confirmation of Organophosphorus Pesticides Residues in Food Applying Gas Chromatography/Mass Spectrometry with Chemical Ionization and Pulsed Positive Negative Detection', *Biomed. Environ. Mass Spectrom.*, **18**, 645–651 (1989).
118. J. Haslová, V. Kocourek, I. Zemanová, F. Pudil, J. Davídek, 'Gas Chromatographic Determination of Chlorinated Phenols in the Form of Various Derivatives', *J. Chromatogr.*, **439**(2), 307–316 (1988).
119. D. Barceló (ed.), *Application of LC–MS in Environmental Chemistry*, Journal of Chromatography Library, Vol. 59, Elsevier, Amsterdam, 1996.
120. E.R. Brower, S. Kofman, U.A.Th. Brinkman, 'Selected Procedures for the Monitoring of Polar Pesticides and Related Microcontaminants in Aquatic Samples', *J. Chromatogr. A*, **703**, 167–190 (1995).
121. U.A.Th. Brinkman, J. Slobodnik, J.J. Vreuls, 'Trace-level Detection and Identification of Polar Pesticides in Surface Waters: the SAMOS Approach', *Trends Anal. Chem.*, **13**(9), 373–381 (1994).
122. J. Slobodnik, M.G.M. Groenewegelt, E.R. Brower, H. Lingeman, U.A.Th. Brinkman, 'Fully Automated Multi-residue Method for Trace Level Monitoring of Polar Pesticides by Liquid Chromatography', *J. Chromatogr.*, **642**(1/2), 359–370 (1993).
123. P.J.M. Van Hout, U.A.Th. Brinkman, 'The Rhine Basin Program', *Trends Anal. Chem.*, **13**(9), 382–388 (1994).
124. P. Parrilla, J.L. Martinez Vidal, M. Martinez Galera, A.G. Frenich, 'Simple and Rapid Screening Procedure for Pesticides in Water Using SPE and HPLC/DAD Detection', *Fresenius' J. Anal. Chem.*, **350**, 633–637 (1994).
125. D. Barceló, G. Durand, V. Bouvot, M. Nielen, 'Use of Extraction Disks for Trace Enrichment of Various Pesticides from River Water and Simulated Seawater Samples followed by Liquid Chromatography–Rapid-scanning UV–Visible and Thermospray–Mass Spectrometry Detection', *Environ. Sci. Technol.*, **27**, 271–277 (1993).
126. S. Lacorte, D. Barceló, 'Determination of Organophosphorus Pesticides and Their Transformation Products in River Waters by Automated On-line Solid-phase Extraction Followed by Thermospray Liquid Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **712**, 103–112 (1995).
127. 'Solvent Extractable Non-volatile Compounds by High Performance Liquid Chromatography/Thermospray/Mass Spectrometry (HPLC–TS/MS) or Ultraviolet (UV) Detection', *EPA Method 8321A*, *Revision 1*, Environmental Protection Agency, Cincinnati, 1995.
128. M. Honing, J. Riu, D. Barceló, B.L.M. Van Baar, U.A.Th. Brinkman, 'Determination of Ten Carbamate Pesticides in Aquatic and Sediment Samples by Liquid Chromatography–Ionspray and Thermospray Mass Spectrometry', *J. Chromatogr.*, **733**, 283–294 (1996).
129. D. Volmer, A. Preiss, K. Levsen, G. Wünsch, 'Thermospray Mass Spectral Studies of Pesticides. Temperature and Salt Concentration Effects on the Ion Abundances in Thermospray Mass Spectra', *J. Chromatogr.*, **647**, 235–259 (1993).
130. J. Abián, G. Durand, D. Barceló, 'Analysis of Chlorotriazine and Their Degradation Products in Environmental Samples by Selecting Various Operating Modes in Thermospray HPLC/MS/MS', *J. Agric. Food Chem.*, **41**, 1264–1273 (1993).
131. R.D. Voyksner, J.T. Bursey, E.D. Pellizzari, 'Postcolumn Addition of Buffer for Thermospray Liquid Chromatography/Mass Spectrometry Identification of Pesticides', *Anal. Chem.*, **56**(8), 1507–1514 (1984).

132. H. Bagheri, E.R. Brower, R.T. Ghijsen, U.A.Th. Brinkman, 'Online Low-level Screening of Polar Pesticides in Drinking and Surface Waters by Liquid Chromatography–Thermospray Mass Spectrometry', *J. Chromatogr.*, **647**(1), 121–129 (1993).
133. S. Sennert, D. Volmer, K. Levsen, G. Wunsch, 'Multiresidue Analysis of Polar Pesticides in Surface and Drinking Water by Online Enrichment and Thermospray LC–MS', *Fresenius' J. Anal. Chem.*, **351**(7), 642–649 (1995).
134. L. Somasundaram, J.R. Coats, K.D. Racke, H.M. Stahr, 'Application of the Microtox System to Assess the Toxicity of Pesticides and Their Hydrolysis Metabolites', *Bull. Environ. Contam. Toxicol.*, **44**(2), 254–259 (1990).
135. R. Straub, R.D. Voyksner, J.T. Keever, 'Thermospray, Particle Beam and Electrospray Liquid Chromatography– Mass Spectrometry of Azo Dyes', *J. Chromatogr.*, **627**(1–2), 173–186 (1992).
136. S. Pleasance, J.F. Anacleto, M.R. Bailiey, 'An Evaluation of Atmospheric Pressure Ionization Technique for the Analysis of *N*-Methyl Carbamate Pesticides by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **3**, 378–397 (1992).
137. C. Aguilar, I. Ferrer, F. Borrull, R.M. Marcé, D. Barceló, 'Comparison of Automated On-line Solid Phase Extraction Followed by Liquid Chromatography–Mass Spectrometry with Atmospheric Pressure Chemical Ionization and Particle Beam Mass Spectrometry for the Determination of a Priority Group of Pesticides in Environmental Waters', *J. Chromatogr. A*, **794**, 147–163 (1998).
138. L.D. Betowski, T.L. Jones, 'The Analysis of Organophosphorus Pesticide Samples by High Performance Liquid Chromatography/Mass Spectrometry and High Performance Liquid Chromatography/Mass Spectrometry/Mass Spectrometry', *Environ. Sci. Technol.*, **22**, 1430–1434 (1988).
139. C. Molina, P. Grasso, E. Benfenatti, D. Barceló, 'Automated Sample Preparation with Extraction Columns Followed by Liquid Chromatography–Ionspray Mass Spectrometry. Interferences, Determination and Degradation of Polar Organophosphorus Pesticides in Water Samples', *J. Chromatogr. A*, **737**, 47–58 (1996).
140. C. Crescenzi, A. Di Corcia, E. Guerriero, R. Samperi, 'Development of a Multiresidue Method for Analyzing Pesticide Traces in Water Based on Solid-phase Extraction and Electrospray Liquid Chromatography Mass Spectrometry', *Environ. Sci.Technol.*, **31**, 479–488 (1997).
141. S. Lacorte, C. Molina, D. Barceló, 'Temperature and Extraction Voltage Effect on Fragmentation of Organophosphorus Pesticides in Liquid Chromatography/Atmospheric Pressure Chemical Ionization', *J. Chromatogr. A*, **795**, 13–26 (1997).
142. S. Kawasaki, H. Ueda, H. Itoh, J. Tadano, 'Screening of Organophosphorus Pesticides Using Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr.*, **595**, 193–202 (1992).
143. S. Lacorte, G. Jeanty, J.L. Marty, D. Barceló, 'Identification of Fenthion and Temephos and Their Transformation Products in Water by High Performance Liquid Chromatography with Diode Array Detection and Atmospheric Pressure Chemical Ionization Mass Spectrometric Detection', *J. Chromatogr. A*, **777**, 99–114 (1997).
144. S. Lacorte, D. Barceló, 'Improvements in the Determination of Organophosphorus Pesticides in Ground and Wastewater Samples from Interlaboratory Studies by Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography–Diode Array Detection', *J. Chromatogr. A*, **725**, 85–92 (1995).

Pesticides (New Generation) and Related Compounds, Analysis of

Y. Picó, G. Font, J.C. Moltó, and J. Mañes
Universitat de València, València, Spain

1 Introduction 1
2 Description 2
3 Characteristics 2
3.1 Physical and Chemical Properties 2
3.2 Toxicology 7
3.3 Regulations 8
4 Isolation of the Sample 8
4.1 Macrocyclic Lactones 8
4.2 Chloronicotinyls 10
4.3 Tetranortriterpenoids 11
4.4 Ammonium Quaternary Salts 11
4.5 Dinitroanilines 11
4.6 Acetamides 11
4.7 Oximes 11
4.8 Triazoles 12
4.9 Pyridine-based Molecules 12
5 Identification and Determination 13
5.1 Gas Chromatography 13
5.2 High-performance Liquid Chromatography 18
5.3 Immunoassay 22
5.4 Other Techniques 23
6 Food and Environmental Applications 24
6.1 Macrocyclic Lactones 24
6.2 Chloronicotinyls 25
6.3 Tetranortriterpenoids 25
6.4 Ammonium Quaternary Salts 25
6.5 Dinitroanilines 26
6.6 Acetamides 26
6.7 Oximes 28
6.8 Triazoles 28
6.9 Pyridine-based Molecules 28
7 Conclusions 28
Acknowledgments 28
Abbreviations and Acronyms 28
Related Articles 29
References 29

Miscellaneous pesticides and related compounds include the compounds that can neither be included in the classical chemical category of pesticides nor classified by their class of action. They are a heterogeneous group of substances like macrocyclic lactones, chloronicotinyls, tetranortriterpenoids, ammonium quaternary salts, dinitroanilines, acetamides, oximes, triazoles and pyridine-based molecules. Their common characteristic is that they are recently developed pesticides. Sample preparations in environmental matrices: air, water, soil, food and biological samples are described. Analytical techniques for characterizing these compounds include gas chromatography (GC), high-performance liquid chromatography (HPLC), and a wide variety of other emerging techniques such as capillary electrophoresis (CE), supercritical fluid chromatography (SFC), voltammetric and immunoassay techniques. Here, those applied to the field of real samples are presented.

1 INTRODUCTION

Environmental management depends on information on a wide range of compounds. Unreliable or inaccurate information may lead to large economic losses which are caused, for example, by sanitation measures that are not needed, or to unacceptable risk to the environment or human health if necessary measures are not taken.

Method development within environmental areas is of great importance in helping to obtain the selectivity and sensitivity necessary for monitoring low levels of toxic substances as required by environmental regulations. Analytical techniques that fulfil these requirements can be achieved by the use of chromatographic separation steps coupled with selective sample handling techniques. In these steps, the pesticides are separated from the many matrix substances and simultaneously enriched. Goals are also reached by striving for more selective detection principles and techniques, such as mass spectrometry (MS) combined with GC or HPLC. Immunoassays offer certain advantages over conventional instrumental methods for the analysis of pesticide residues and are ideally suited for screening a large number of samples for low levels of specific analytes.

New groups of pesticides have been invented for plant protection at low doses, with a reversible mode of action, less persistence, systemic action and adequate potency against crop pests. A critical review of the analytical methodology used to determine these newly developed pesticides is required so that the methods of analysis evolved for the generation of data can be used as standards.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 DESCRIPTION

There are many registered chemicals that can be included under the heading of miscellaneous pesticides and related compounds which comprise a large group of substances with very different characteristics and applications.[1] Their common attribute is that they do not belong to the classical chemical categories of pesticides, such as organochlorines, organophosphates, carbamates or triazines.[2] However, depending on their chemical structure they can be classified into several chemical classes, the most important of which are macrocyclic lactone, chloronicotinyl, tetranortriterpenoid, ammonium quaternary salt, dinitroaniline, acetamide, oxime, triazole and pyridine-based molecules. Table 1 shows the chemical structures and some examples of the major classes that could be considered to be miscellaneous pesticides.

The avermectins are macrocyclic lactones produced by the soil microorganisms *Streptomyces avermitilis*. Two avermectins have been commercialized to date. The semisynthetic 2,2,2,2,3-dihydroavermectin B_1 known as ivermectin is used as an antiparasitic drug in animals and man; avermectin B_1, known as abamectin, is widely used as an insecticide and miticide for agricultural crops.[3]

The chloronicotinyl insecticides act on cholinergic receptors and have good contact properties and powerful systemic action after uptake through the root system. They are effective in controlling aphids, whiteflies, thrips, scales, psillids, plant bugs, leafhoppers, planthoppers, and other harmful pest species including resistant strains. Owing to their insecticidal effectiveness at low concentrations and their safety in the environment, the chloronicotinyls are seen as promising newcomers.[4]

Azadirachtin is a tetranortriterpenoid (limonoid) present in neem seeds (*Azadirachta indica* A. Juss). Azadirachtin has gained worldwide attention for its insect antifeedant and ecdysis-inhibiting properties. This compound is highly potent at low concentration against more than 200 agricultural pests and it is ecofriendly. It is therefore a potentially safe alternative to the toxic synthetic pesticides and a number of commercial formulations have been introduced worldwide.

The bipyridylium herbicides DQ, PQ and DF are important cationic pesticides used in agriculture. They are usually supplied as dibromide, dichloride and methyl sulfate salts, respectively. They are nonselective contact herbicides that are absorbed by foliage with some translocation though the xylem. They are used for the broad-spectrum control of broad-leaved weeds and grasses in fruit orchards, plantation crops, ornamentals and shrubs. They are also applied for general weed control on noncrop land and have been recommended for the control of aquatic weeds. They have been applied as defoliants on crops such as cotton and as desiccants for pineapples, potatoes, sugarcane and sunflower. CQ and MQ chloride are cationic plant growth regulators, structurally related to the bipyridylium herbicides, which are mainly used to prevent loging in barley and rye and also to increase the yield of cotton.

Dinitroanilines are selective pre-emergence herbicides used to control some broad-leaved weeds and the major annual grasses in a wide variety of agronomic crops. They disrupt mitosis in the meristematic cell of seedling plants by inhibiting the formation of microtubules.

The group of acetamide pesticides (some also known as chloroacetamides) encompasses a considerable number of herbicides and fungicides used to control weeds and fungi in crops. The compounds are widely used to control annual grasses and certain broad-leaved weeds in corn, soybeans and peanuts and to control phytopathogenic fungi (*Peronosporales*) in potatoes, sugar beets, and other crops. They act by inhibiting protein synthesis by reaction of the activated Cl atom of the chloroacetyl group with reactive sites in proteins.

Oximes are bioactive compounds, originally discovered in insects, that have recently been synthetized, and some are effective herbicides.[5]

Triazole derivatives are aromatic heterocycles widely used as weed killers, fungicides, insecticides, plant growth regulators and antimicrobial agents.[6] They are nonselective systematic herbicides used against a wide variety of plants.

Pyridine-based molecules are a group of substances that include pyridazines, pyridazones and pyridones. All of them are herbicides but their applications vary. For example, Pyridate is a pyridazine that acts by contact while pyridazone derivatives such as norflurazon and cloridazon are soil-applied herbicides and fluridone, a pyridone, is an experimental herbicide developed for aquatic plant management systems.

3 CHARACTERISTICS

3.1 Physical and Chemical Properties

In order to develop the best residue analytical methods with the most appropriate instruments, a knowledge of the physical and chemical properties of the pesticides is essential. For example, it is impossible to develop a GC residue method for imidacloprid without derivatization because of the very polar and nonvolatile properties displayed by this molecule. Similarly, a direct GC method for DQ, PQ or azadirachtin is also impossible because of the polarity of these compounds. The physical and chemical properties of some typical herbicides are shown in Table 2.

Table 1 Class of compounds, typical examples and structures included as miscellaneous compounds

Class	Typical examples	Chemical structure
Macrocyclic lactone	Avermectin Ivermectin Emamectin	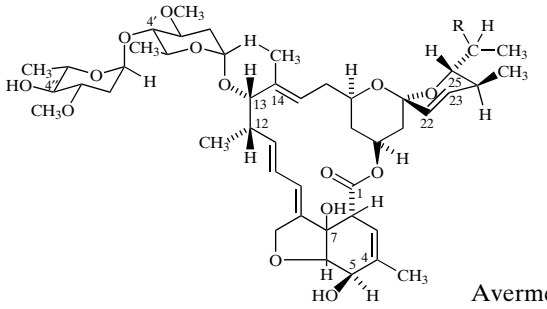 Avermectin
Chloronicotinyl	Imidacloprid Acetamiprid	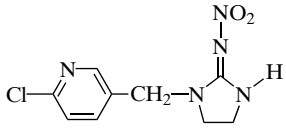 Imidacloprid
Tetranortriterpenoid	Azadirachtin	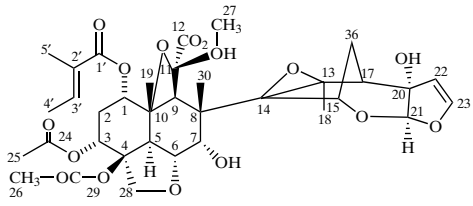
Ammonium quaternary salt	DQ PQ DF CQ MQ	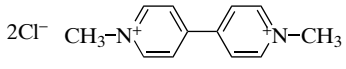 Paraquat
Dinitroaniline	Butralin Dinitramine Ethalfluralin Oryzalin Pendimethalin Trifluralin	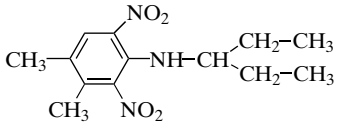 Pendimethalin
Acetamide	Alachlor Acetochlor Metalaxyl Metolachlor	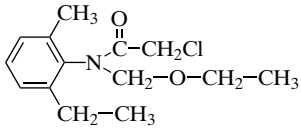 Acetochlor
Oxime	Alloxydim Clethodim Sethoxydim	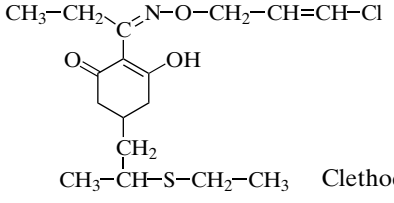 Clethodim
Triazole	Amitrole Carfentrazone ethyl Fenchlorazol ethyl Flupoxam	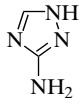 Amitrole
Pyridine-based molecule	Chloridazon Thiazopyr Norflurazon Fluridone Pyridate	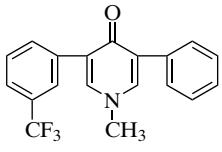 Fluridone

DQ, diquat; PQ, paraquat; DF, difenzoquat; CQ, chlormequat; MQ, mepiquat.

Table 2 Physical and chemical properties of some miscellaneous compounds

CAS number	Pesticide	Molecular weight	Appearance	Mp (°C)	Vapor pressure (20–25 °C)	Solubility				LogP_{ow}
						Water (mg L^{-1})	Toluene (g L^{-1})	Acetone (g L^{-1})	Methanol (g L^{-1})	
71751-41-2	Avermectin	873.11	White solid	150–155	1.5 mPa	<0.005	350	100	19.5	3.99
137512-74-4	Emamectin	1008.26	White solid	141–146	–	300	–	Soluble	Soluble	–
138261-41-3	Imidacloprid	255.66	Pale solid	143.8	0.2 µPa	510	0.5	–	10	0.57
	Acetamiprid	222.68	Pale solid	98.9	<1 µPa	4250	–	Soluble	Soluble	0.8
11141-17-6	Azadirachtin	720.7	Yellow solid	180	–	50	–	–	–	–
2764-72-9 85-00-7 6385-62-2	DQ	344.06	Pale yellow solid	>300	<13 µPa	700 000	Insoluble	–	Slightly soluble	−4.6
4685-14-7 1910-42-5	PQ	257.2	White solid	300	<133 nPa	700 000	Insoluble	–	Slightly soluble	
43222-58-6	DF	360.44	Pale liquid	155–157	13.33 µPa	765 000	Insoluble	9.8	620	0.2838
999-81-5	CQ	158.07	Pale yellow liquid	245	<10 µPa	>1 kg kg^{-1}	–	0.3	–	−1.58
24307-26-5	MQ	149.7	White solid	285	<10 µPa	>1 kg kg^{-1}	–	20	–	−2.82
33629-47-9	Butralin	295.3	Orange solid	60–61/ 134–136	1.7 mPa	1	–	–	–	–
29091-05-2	Dinitramine	332.2	Yellow solid	98–99/ >200	479 µPa	1	–	1040	–	
55283-68-6	Ethalfluralin	333.26	Orange solid	57–59	110 µPa	0.3	>500	>500	82–100	–
19044-88-3	Oryzalin	346.36	Orange solid	141–142	13.3 µPa	2.5	Insoluble	>500	50	
40487-42-1	Pendimethalin	281.31	Yellow solid	54–58/ 330	4 mPa	0.275	–	700–800	–	5.18

1582-09-8	Trifluralin	335.28	Orange solid	48–49/ 96–97	140 µPa	<1	Soluble	400	–	5.07
15972-60-8	Alachlor	269.77	White solid	41/100	2 mPa	170	Soluble	–	–	–
34256-82-1	Acetochlor	269.77	Liquid	<0	45.3 nPa	223	Soluble	Soluble	–	2.47
57827-19-1	Metalaxyl	279.34	Pale solid	71–72	293 µPa	7100	9.1	–	650	1.75
51218-45-2	Metolachlor	283.8	Oily liquid	282	1.7 mPa	530	Soluble	Soluble	Soluble	3.45
66003-55-2	Alloxydim-sodium	345.5	White solid	185.5	133 µPa	2 000 000	–	14	619	0.63
99129-21-2	Clethodim	359.92	Clear liquid	–	13 µPa	–	Soluble	Soluble	Soluble	4.18
74051-80-2	Sethoxydim	327.5	Oily liquid	90	133.8 µPa	4700	Soluble	Soluble	Soluble	4.51
61-82-5	Amitrole	84.08	Pale solid	157–159	<1 mPa	280 000	Insoluble	Insoluble	Moderately soluble	0.85
103112-35-2	Fenchlor-azole-ethyl	403.48	Pale yellow solid	114–116	890 nPa	0.9	270	360	27	3.998
122836-35-5	Sulfentrazone	387.19	Dark brown solid	75–78	133.28 nPa	0.11	–	Slightly soluble	Slightly soluble	–
1698-60-8	Chloridazon	221.65	Pale solid	205–206	<10 µPa	400	–	28	34	1.18
117718-60-2	Thiazopyr	396.4	Brown solid	77.3–79	266 µPa	2.5	–	–	287	3.89
27314-13-2	Norflurazon	303.67	White solid	174–180	2.7 µPa	28	–	50	–	2.4–2.5
59756-60-4	Fluoridone	329.3	White solid	151–155	<13.33 µPa	12	–	–	10–20	1.87
55512-33-9	Pyridate	378.9	Brown solid	20–25	133.28 nPa	Insoluble	Soluble	Soluble	Soluble	3.01

The avermectins are a family of macrocyclic lactones. Abamectin is a mixture of two components. The major one is avermectin B_{1a}, which makes up 80% or more of the mixture, and the minor component is avermectin B_{1b}, which supplies 20% or less; the two components differ in a single methylene group.[7] Ultraviolet (UV) light below 280 nm rapidly isomerizes the *E* (*trans*) 8,9 and 10,11 double bonds of avermectin to 8,9- and 10,11-*Z* isomers. However, solutions of avermectin in Pyrex flasks are generally stable because Pyrex excludes most of the UV light below 280 nm. Avermectin residues degrade rapidly by both oxidative and photochemical pathways, forming a variety of products when applied to a number of different crops. Therefore, the only residues of toxicological significance are avermectin and 8,9-*Z* avermectin.[8]

Ivermectin is synthesized from avermectin by reducing the 22,23-double bond, thus producing dihydroavermectin B_{1a} and dihydroavermectin B_{1b}. These compounds are notable for their potent anthelmintic and insecticidal activities.

Emamectin benzoate (MK-224, MAB_1 benzoate, 4-deoxy-4-epimethylamino avermectin B_1) is a derivative of abamectin and consists of two avermectin homologs, each with a molecular weight of approximately 900. By specification it consists of at least 90% 4″-deoxy-4″-epimethylaminoavermectin B_{1a} benzoate and not more than 10% of 4″-deoxy-4″ epimethylaminoavermectin B_{1b} benzoate. These components differ only in a methylene group on the C-25 side chain. Emamectin benzoate is an effective insecticide against lepidopteran larvae and is currently under development for use on a number of crops including celery, lettuce, cabbage and tomato.[9] Emamectin benzoate is an ionizable salt and displays higher water solubility than other avermectins,[10] but binds very tightly to topsoil and is, therefore, immobile.[9]

Soil biodegradation studies on the chloronicotinyl imidacloprid have demonstrated that its major metabolite is 6-chloronicotinic acid, and that it leaves no soil residue after the first 3 months.[11] Acetamiprid is stable in solution at pH between 4.5 and 7. It is slowly degraded at pH 9 and 45 °C.

Azadirachtin is nonvolatile and highly polar. It has UV absorption due to the α,β-unsaturated carbonyl chromophore in the ligate ester and the vinyl ether, but the absorption maximum is at a very short wavelength.[12] Azadirachtin hydrolyzes readily at 35 °C, and its disappearance follows simple pseudo-first-order kinetics. Its hydrolysis is faster at a basic than an acidic pH.[13]

DQ is marketed as a dibromide and dibromide monohydrate salts and as a base. PQ is marketed as a chloride and dimethyl sulfate salts and DF is marketed as methyl sulfate salts.[14] The PQ salts are hygroscopic, not volatile and very soluble in water. PQ is extremely stable in the presence of acids but it is destroyed quickly in a basic medium. It is soluble in water and insoluble in organic solvents. DQ is also very thermostable and not volatile.[15] The herbicide activity of PQ and DQ is related to the planar structure of their molecules and the 18 possible resonant structures that stabilize the radical formed in reduction reactions in which they are involved.[16] DF salts are very soluble in water and are stable to hydrolysis.[17] DF is stable at weakly acid pH but is degraded in strong acids and in the presence of oxidizers. The three herbicides present an absorption band in the UV at 260, 310 and 255 nm, respectively, due to the presence of aromatic rings in their structures.

Although the dinitroaniline herbicides are chemically related, they differ in volatility, persistence in soil and absorption by crops, and for this reason may differ in their effects on soil, plants and air. These compounds are among the least mobile herbicides, and therefore runoff is the principal route of the contamination of surface waters.[18] Dinitroaniline herbicides are water insoluble, relatively volatile and strongly adsorbed to soil colloids. In soil, both chemical reactions and biological processes degrade pendimethalin. In general, dinitroaniline herbicide degradation is more rapid under flooded anaerobic conditions than under aerobic ones.

Acetamide herbicides have moderate water solubility and are rapidly absorbed into plants. In susceptible plants, the herbicides act by inhibiting protein synthesis, whereas insensitive plants rapidly inactivate these herbicides via gluthathion conjugation. In sensitive fungi, the structurally related fungicides have effects on RNA synthesis. As these reactions involve various chiral structures, some stereoselectivity is expected in the activity of these compounds.

Oximes are bioactive compounds originally discovered in insects. Field trials with these compounds showed that they were not consistently effective, perhaps because of the instability of this active ingredient. Sethoxydim undergoes degradation, including photodegradation. Clethodim is also degraded in aqueous solution by acid medium and light. Clethodim degradation increases as solution pH decreases and photolysis is more rapid and more complete than hydrolysis.[5]

Amitrole belongs to the triazole group[19] and is soluble in water, methanol, ethanol and chloroform, slightly soluble in ethyl acetate and insoluble in ether and acetone. Aqueous solutions are neutral.

Pyridate is a colorless crystalline solid that melts at 27 °C and boils at 220 °C under 10^{-6} mbar vacuum. Its vapor pressure is 1.3×10^{-9} mbar at 20 °C. It is stable in neutral medium, but is hydrolyzed in strong acid and strong alkali media.[20]

3.2 Toxicology

Pesticide residues are regulated at the international and national level according to the toxicity of the pesticide and the human exposure to a particular substance. Table 3 lists the acute oral and dermal toxicity for different animal species, chronic toxicity for dogs, and the ADI.

The macrocyclic lactones are neurotoxins that manifest their action by disrupting the normal function of γ-aminobutyric acid (GABA), an important neurotransmitter in the central nervous system of vertebrates and in the peripheral nervous system of invertebrates. Because mammals only have GABA ergic synapses in the central nervous system, the mammalian blood–brain barrier ensures a degree of specificity. A notable feature of this group of compounds is their low LD_{50} (lethal dose 50) values, but they are not usually highly toxic by the dermal route on account of their large molecules and poor transdermal absorption. In vitro studies with preparations of rat brain have shown that avermectin B_{1a} stimulates presynaptic binding of GABA and enhances postsynaptic binding of GABA; the action of avermectin B_{1a} is antagonized by bicuculline and picrotoxin.[21]

Avermectin and ivermectin are metabolized in a qualitatively similar way among different species. The major metabolites of both in cattle, sheep, swine and rats are either 24-hydroxymethyl or 3″-*O*-desmethyl derivatives. However, the enzymes responsible for the metabolism have not been identified in any species.[7] These compounds undergo little metabolism and most of the dose given to the animal is excreted relatively unaltered, primarily in the feces.[3] Abamectin induces teratogenic effects such as cleft palate.

The chloronicotinyl insecticides interfere with neuronal functions as do organophosphate, carbamate and pyrethroid insecticides. Unlike the latter pesticides, they act on nicotinergic acetylcholine receptors on the

Table 3 Toxicologic characteristics of miscellaneous pesticides

Pesticide	Oral LD_{50} rat ($mg\,kg^{-1}$)	Oral LD_{50} mice ($mg\,kg^{-1}$)	Dermal LD_{50} rabbit ($mg\,kg^{-1}$)	Chronic toxicity NOEL in dog	ADI ($mg\,kg^{-1}$)
Abamectin	10.6–11.3	13.6	>2000	0.25 $mg\,kg^{-1}\,day^{-1}$	0.0025
Emamectin	70	–	>2000	–	–
Imidacloprid	410–440	98–100	Non irritant	500 $mg\,kg^{-1}$ diet	0.057
Acetamiprid	146–217	184–198	Non irritant	–	–
Azadirachtin	>5000	–	>5000	–	–
DQ	215–235	125	>400	50 $mg\,kg^{-1}$ diet	0.008
PQ	100–150	104	Irritant	15 $mg\,kg^{-1}$ diet	0.004
DF	270–470	31–44	470	2500 $mg\,kg^{-1}$	0.1
CQ	807–966	–	>2000	–	0.05
MQ	1490	–	Non irritant	–	0.05
Butralin	12 600	–	10 200	–	0.015
Dinitramine	3000	–	>6800	2000 $mg\,kg^{-1}$ diet	0.05
Ethalfluralin	10 000	10 000	>2000	5000 $mg\,kg^{-1}\,day^{-1}$	0.05
Oryzalin	>10 000	>10 000	>2000	–	0.15
Pendimethalin	1050–1250	1340–1620	>5000	12.5 $mg\,kg^{-1}\,day^{-1}$	0.0125
Trifluralin	>10 000	500	>2000	1000 $mg\,kg^{-1}\,day^{-1}$	0.25
Alachlor	930–1350	–	>2000	200 $mg\,kg^{-1}$ diet	0.005
Acetochlor	2148	–	4166	12 $mg\,kg^{-1}\,day^{-1}$	0.01
Metalaxyl	669	788	>6000	62.5 $mg\,kg^{-1}\,day^{-1}$	0.03
Metolachlor	2780	894	>10 000	2.5 $mg\,kg^{-1}\,day^{-1}$	0.015
Alloxydim	2260–2322	3000–3200	>5000	40 $mg\,kg^{-1}\,day^{-1}$	0.125
Clethodim	1360–1630	–	>5000	–	0.01
Sethoxydim	3200–3500	5600–6300	–	–	0.15
Amitrole	1100–24 600	11 000	10 000	–	0.000 03
Carfentrazone ethyl	5143	–	–	–	–
Fenchlorazol ethyl	>5000	>2000	>2000	4.5–5.4 $mg\,kg^{-1}\,day^{-1}$	–
Flupoxam	>5000	–	>5000	–	–
Sulfentrazone	2855	–	>2000	–	–
Chloridazon	1100–3830	2500	>2500	–	0.075
Thiazopyr	>5000	–	>5000	–	0.0025
Norfluorazon	8400	–	>20 000	150 $mg\,kg^{-1}\,day^{-1}$	0.0038
Fluridone	>10 000	>10 000	>500	150 $mg\,kg^{-1}\,day^{-1}$	0.08
Pyridate	1800–2100	>10 000	>10 000	7.9 $mg\,kg^{-1}\,day^{-1}$	0.04

NOEL, no observable effect level; ADI, admissible daily intake.

postsynaptic membrane of an identified insect motor neuron.[11]

Azadirachtin is nonmutagenic and does not appear to exhibit mammalian toxicity.[22]

PQ toxicity in both experimental animals and humans targets primarily the lung, whereas DQ does not. It is believed that the differences between PQ and DQ are a result of the fact that the lung selectively accumulates PQ and not DQ. PQ-induced pulmonary injury takes place in two phases: destructive and proliferative. It has been suggested that the biochemical reactions that lead to the destructive effect of PQ are analogous to its toxic action on plant cells. NADPH (nicotinamide adenine dinucleotide phosphate (reduced form)) is the donor for the single-electron reduction of PQ. Reduced PQ is reoxidized rapidly by molecular oxygen and superoxide radicals are formed. The superoxide radicals initiate a chain of reactions that produce toxic reactive intermediates that include hydrogen peroxide and hydroxy radicals and also produce lipid peroxidation. They are responsible for the disruption of cellular membranes. In addition, PQ competes for and deprives other systems of essential NADPH and compromises their cellular integrity.[23,24]

DQ poisoning differs from that of PQ in that the renal effects are more prominent and lung changes do not generally occur. An effect of DQ that has been extensively investigated is its ability to produce cataracts in experimental animals.[23] PQ produces chronic effects such as costal cartilage malformation in rats when injected during the gestation. It appears that the teratogenicity of PQ results from its effect on collagen biosynthesis.[24]

CQ is known to be a competitive inhibitor of cholinesterase in animals. This anticholinesterase chemical causes acetylcholine accumulation at cholinergic receptor sites and thus is capable of producing effects equivalent to excessive stimulation of cholinergic receptors through the central and peripheral nervous system.[25]

Substituted anilines have the general property of causing methemoglobinemia, as do many other aniline derivatives. The probable mechanism of methemoglobinemia is N-hydroxylation to the corresponding hydroxylamine, which then takes part in an intraerythrocytic cycle with the corresponding nitroso derivative generating methemoglobin at the same time. Alachlor is classed as a probable human carcinogen by the USEPA (United States Environmental Protection Agency)[26] because of its carcinogenic effect on rodents, where it produces posterior nasal and stomach tumors, possibly by a nongenotoxic mechanism.

Little is known of the human metabolism of amitrole but experiments with [^{14}C]amitrole orally administered to rats show that 80–90% of the compound is excreted unchanged in the urine.[27] The International Association for Research on Cancer (IARC) classifies amitrole as a possible human carcinogen that produces thyroid tumors apparently through an indirect mechanism. Amitrole inhibits thyroid peroxidase, thus lowering the thyroxin level. The thyroid gland becomes hyperplasic and eventually forms tumors.

The literature is scarce on toxicity of the oximes. There are only a few studies demonstrating that sethoxydim produces lesions in bone marrow and the liver of dogs.[28]

In laboratory metabolism studies with ^{14}C-fluridone, a major fluridone metabolite, 1-methyl-3-(4-hydroxyphenyl)-5-[3-(trigluoromethyl)phenyl]-4-(1*H*)-pyridinone was present in bluegill and fathead minnows at levels equivalent to 50% and 15% of the total residue in fish, respectively.

3.3 Regulations

A legal tolerance is defined as the maximum level or concentration of a pesticide, or its metabolites or derivatives, that is permitted in or on a particular food crop or food product. The amount is generally no higher than the concentration present in or on a particular food crop or food product as a result of good agricultural practices.

The European Union (EU) has proposed rigid limits for pesticides in drinking water, especially 0.1 $\mu g\,L^{-1}$ (0.1 ppb) for a single pesticide and 0.5 $\mu g\,L^{-1}$ (0.5 ppb) for total pesticides including their degradation products.[29] In the United States, USEPA has set values for maximum contaminant levels for pesticides individually. The USEPA values range approximately from 1 $\mu g\,L^{-1}$ to 1 mg L^{-1} (1 ppb to 1 ppm).[30,31] Such levels are more specific than EU levels, which are fixed for all the individual pesticides without making any distinction between individual pesticides of different toxicity.[32] However, of all the pesticides covered in this article, USEPA has only established a health advisory level of 20 $\mu g\,L^{-1}$ for DQ.[31]

The maximum levels of these miscellaneous pesticides in fruit and vegetables allowed in the EU are shown in Table 4.[33–36] Maximum levels in animal tissues, other biological samples or different environmental compartments have not been established. These established tolerances are revised periodically and the tendency is to diminish the permitted levels. This fact poses substantial analytical problems and demands a methodology of greater sensitivity.

4 ISOLATION OF THE SAMPLE

4.1 Macrocyclic Lactones

Abamectin has been determined in ambient air and on surfaces following their treatment with the pesticide. Surface sampling was performed with cotton dosimeters

with or without wetting solutions such as 2-propanol. Air sampling is achieved in Teflon™ filters or in Orbon™ 42 tubes. Teflon™ air filters and cotton dosimeters were extracted with acetonitrile using amber jars or Soxhlet™. Cleanup was done by diluting with the ionized water passing through octyl (bonded C_8) and amipropyl ($-NH_2$) solid-phase extraction (SPE) columns.[37,38]

Emamectin benzoate was concentrated from water on C_8 SPE cartridges and eluted with acetate ammonium in methanol.[10]

The methods described for the determination of macrocyclic lactones from vegetables require complicated homogenization/extraction schemes. Fruit and vegetables are mainly extracted with acetonitrile/water/hexane (1 : 1 : 5). Abamectine and its δ 8,9-isomer are concentrated by passing the hexane layer through an aminopropyl SPE cartridge.[39–43] Lettuce and cucumber are extracted with ethyl acetate and the extract is purified by SPE using a Sep-Pack™ silica cartridge.[44] Emamectin benzoate is extracted from lettuce and celery with methanol, and the extract is cleaned up on a C_8 cartridge followed by LLE (liquid–liquid extraction) with ethyl acetate before passing through a propyl sulfonyl cation-exchange cartridge.[45]

Abamectin residues from soil samples have been extracted with acetonitrile.[9,21] The insecticide abamectin is extracted from soil and animal tissues using supercritical-fluid extraction (SFE).[46] Supercritical carbon dioxide

Table 4 Miscellaneous pesticides tolerance in fruit and vegetables[33–36]

Pesticide	Residue	MRL (ppm)
Abamectin	Total toxic residue includes parent abamectin and its δ 8,9-isomers	0.01
Imidacloprid	Sum of imidacloprid and its major metabolite 6-chloronicotinic acid expressed as 6-chloronicotinic acid.	Citrus fruits 1 Seed fruits, apricot, peach, nectarine and paprika 0.5 Tomato, cucumber, gherkin and marrow 0.1 Other products 0.05
DQ	Sum of DQ and its dibromide expressed as ion DQ	Cotton seeds 0.5 Vegetables and sunflower seeds 0.1 Other products 0.05
PQ	Sum of PQ and PQ dichloride expressed as ion PQ	Tea 0.1
DF	–	Grain straw 3 Other products 0.1
CQ	Sum of CQ and its chloride expressed as ion CQ	Grain straw 10 Pear, rye and oats 3 Grape, raisin and wheat 1 Other products 0.05
MQ	–	Cotton seeds 0.5 Grain 0.3 Garlic 0.1 Other products 0.05
Butralin	–	Tobacco 5 Citrus fruit, seed fruits, artichoke, garlic, eggplant, onion, strawberry, soya beans, bean, paprika, tomato, carrot and cotton seed 0.1 Other products 0.1
Dinitramine	–	All products 0.01
Ethalfluralin	–	Bean, soya bean, paprika and corn 0.05 Other products 0.02
Oryzalin	–	All products 0.01
Pendimethalin	–	All products 0.05
Trifluralin	–	Carrot 1 Other products 0.05
Alachlor	Sum of alachlor, 2,6-diethylaniline and its 1-hydroxyethyl, expressed as alachlor	Corn and sorghum foragers 0.5 Onion, species of genus Brassica, oleaginous seeds, and corn 0.1 Other products 0.05

(*continued overleaf*)

Table 4 (*continued*)

Pesticide	Residue	MRL (ppm)
Acetochlor	–	All products 0.01
Metolachlor	–	All products 0.05
Alloxydim-sodium	Sum of Alloxydim, its sodium salt and its analogous deallyloxydate (CM-l), expressed as alloxydim	Fruit and pepo vegetables, species of genus Brassica, lettuce, spinach and similars, watercress and endive, other aromatic herbs, fresh green leguminous, other young stem and mushrooms 0.5 Strawberries and beet leaves 0.2 Tubers and bulbs, onion seedling, fennel, leek, seed of sesame and sunflower, potato, alfalfa and other leguminous and legumes 0.05 Roots and sugar beet 0.02 Other products 0.01
Clethodim	Sum of clethodim and its hydroxylate metabolites, expressed as clethodim	Soya bean 0.3 Tomato, fresh green leguminous and seeds of flax, colza and cotton 0.1 Other products 0.05
Sethoxydim	Sum of sethoxydim and its main metabolites expressed as sethoxydim	Roots (except carrots), tubers, bulbs, fruits and pepo vegetables, Brassica, leaf vegetable and fresh aromatic herbs, fresh green leguminous, young stem and mushroom 0.5 Other products 0.1
Amitrole	–	Tea and dry hop 0.1 Other infusions 0.1 Fresh, desiccated or raw, frozen and without sugar fruits, nuts, fresh or raw, frozen or desiccated, legumes, oleaginous seeds and potato 0.05 Other products 0.05
Fenchlorazol ethyl	–	All products 0.05
Chloridazon	–	Beet leaves 0.5 Beet, spinach and similars 0.1 Other products 0.05
Thiazopyr	Sum of thiazopyr and its nitrile and amide derived thiazolic ring	Seed fruits 0.2 Other products 0.02
Norflurazon	Sum of norflurazon and demethylnorflurazon, expressed as norflurazon	Citrus fruits, seed fruit 0.1 Other products 0.05
Pyridate	–	All products 0.2

MRL, maximum residue limit; CM-l, methyl 5-(l-aminobutylidene)-2,2-dimethyl-4,6-dioxocyclohexanecarboxylate.

modified with 9% 2-methoxyethanol extracts abamectin from soil samples without the need for cleanup. Abamectin is also easily extracted from animal tissues without interfering coextractives.

Fish tissues have been extracted with acetone/methylene chloride and cleaned up with aminopropyl cartridges.[47] A simple specific procedure for determining avermectin B_1 in biological samples has been described[48] and is based on the development of a specific antiavermectin B_1 immunosorbent. Avermetin B_1 is extracted from cattle plasma, cattle meat and pear samples. The extracts are cleaned up on a SPE column (C_{18}) and an immunoaffinity column. Recoveries of avermectin B_1 from samples spiked at levels of 6 and 60 ppb ranged from 80–86%, with coefficients of variation of 5–14%. Feces samples are extracted with an acetone–water mixture, partitioned with isooctane and cleaned up with a C_{18} cartridge.[49]

4.2 Chloronicotinyls

To quantify imidacloprid in liquid and solid formulations samples were extracted in solvent or dissolved in acetonitrile.[50] Water samples were extracted with C_{18} SPE.[51] For extraction in vegetables, the acetone–dichloromethane extraction followed by a cleanup step on a C_{18} cartridge system is selected owing to the good recovery values.[52] Rice and cucumber are extracted with acetonitrile–water (80 : 20) and silica gel

column chromatography.[4] Soil samples are shaken with acetonitrile–water,[4,51] acetonitrile–methanol–water[11] or with chloroform.[53]

Acetamiprid in crops has been extracted with methanol, cleaned up by liquid–liquid partition with *n*-hexane or dichloromethane, and Florisil™ column chromatography used to remove interferences.[54,55]

4.3 Tetranortriterpenoids

Azadirachtin has only been determined in insecticidal formulations of kernel extract. Neem seed powder has been extracted by chloroform in a percolator[56] or by hexane in a Soxhlet™.[57] Insecticide formulations were dissolved in methanol–water (90 : 10) and C_{18} SPE was used to retain unwanted material.[22] SFC has also proven to be a suitable system for determining azadirachtin in crude extracts of neem.[12,58]

4.4 Ammonium Quaternary Salts

Water samples containing more than $1\,\mu g\,mL^{-1}$ can be analyzed directly.[14,59,60] The determination of trace amounts of DQ, PQ and DF in natural and drinking waters requires a preconcentration step before the HPLC determination. SPE techniques are frequently used with C_{18},[61] C_8,[62] silica,[17,63–65] graphitized carbon black (GCB)[66] and ion-exchange column chromatography.[59,67,68] Water sample treatment by isotachophoresis,[69] electrodialysis[70] and flow-injection analysis (FIA)[71–73] has also been reported.

Traditional methods of determination in fruit and vegetable crops include matrix maceration with acid solution and isolation with silica columns[16,74–77] or cation exchange resin.[78] DQ and PQ can also be isolated after derivatization and LLE with diethyl ether.[79] The separation of MQ after acid maceration can be carried out by partition with hexane followed by a cation exchange column.[80] DQ and PQ can be analyzed in edible oil after sonication with a phosphate buffer and cleanup with C_{18} cartridges.[81] CQ has been extracted from crops by maceration with methanol and after concentration, partitioned with hexane prior to alumina cleanup.[82]

Methods of determination in soil are similar to those reported for fruit and vegetables and involve matrix maceration with acid solution and isolation with cation exchange resin[78] or by partitioning with dichloromethane.[15]

Several methods are available for measuring DQ and PQ in biological fluids. DQ determination in blood serum and urine is performed directly after precipitating out the proteins and filtration.[78,83–85] DQ and PQ have been extracted using ion-pair extraction into organic solvents[86–90] and ion-pair SPE, using disposable cartridges of C_{18}.[91] The latter method is rapid, easy to perform and provides high recovery values. Cianopropyl extraction columns[92] or silica columns[93] have been applied to determine DQ and PQ in human samples (blood, urine, milk) by direct SPE. Simultaneous determination of PQ and DQ in tissues is achieved after homogenization with perchloric acid.[94] If further cleanup is required, the mixture is centrifuged and after adjusting the pH of the supernatant is passed through C_{18} cartridges.[95,96]

4.5 Dinitroanilines

Formulations containing benefin and trifluralin have been extracted in hexane.[97] Dinitroaniline herbicide concentrations in soil and plants have usually been measured after sample extraction with an organic solvent and purification using Florisil™ cartridges[98] or glass chromatography columns packed with Florisil™.[99,100]

A method based on the extraction of soil samples in small columns has been used to determine dinitroanilines. Acetone/water (90 : 10) was used for the extraction of simazine. Pendimethalin is extracted from soil with an extraction solution containing acetonitrile, water and glacial acetic acid and partitioning with *n*-hexane.[101]

Simultaneous determination of the dinitroaniline herbicides dinitramine, ethalfluralin, trifluralin, pendimethalin and isopropalin in soil and surface water has been reported.[18] The soil was extracted with diethyl ether and analyzed without any cleanup. The water was analyzed after purification and concentration on C_{18} cartridges.

4.6 Acetamides

Acetamide herbicides have been isolated from river water and rain samples by SPE with C_{18},[102–111] styrene–divinyl benzene copolymer,[112] XAD-2 resin,[113] carbon black[114] or by LLE with dichloromethane[115–117] or hexane.[26]

Acetamides in crops have been determined after extraction of metabolites obtained by hydrolysis. The anilines were isolated by steam distillation and partitioned with Superclean™ ENVI-Chrom P.[118–120] Soil was extracted with ethyl acetate[120] or with methanol–water (9 : 1) followed by SPE with C_{18}.[111] SFE was applied for the detection of herbicides in meat samples[121] and soil samples.[122,123]

4.7 Oximes

Few analytical methods have been reported for extracting and concentrating oxime herbicides from different matrices. An analytical method for the total content of alloxydim-sodium herbicide and its degradation products

in groundwater has been developed based on concentration by evaporation of water and ethyl acetate partitioning after derivatization of the analytes with H_2O_2.[124] Residue analysis of the herbicide sethoxydim and its major active metabolite was also performed in rapeseed.[125] The extraction procedure was based on celite–methanol extraction, which requires complicated partitioning mainly with $CHCl_3$. Recoveries ranged from 83 to 102%.

4.8 Triazoles

Commercial samples of amitrole have been analyzed.[19] The samples used were ''Compo-Total-Unkrautvernichter'' from Compo GmbH (composition: 10% amitrole, 25% simazine, 25% diuron) and ''Domatol Speziall'' (composition: 38% amitrole, 16% MCPA (4-chloro-2-methyl(phenoxy)acetic acid), 19.5% simazine) obtained from a plant protection office in Trier, Germany. The determination can be performed directly with solutions of the commercial herbicide formulations.

The difficulty of determining aminotriazole at trace levels in water is linked to the necessary extraction and concentration from the aqueous media. LLE or SPE cannot be performed with the classical sorbents owing to the low solubility of aminotriazole in organic solvents and its high solubility in water.[126,127] Since aminotriazole is ionized at a low pH, an ion exchanger can be used to extract it from water.[126,128] However, the effect of the inorganic cations causes rapid breakthrough in drinking water. It is also possible to use an ion-pairing preconcentration.[126,128] However, the most reported method for concentrating this compound is evaporation of water.[126,129,130]

In spite of all the above, a fast method for the determination of amitrole in drinking and groundwater has been reported.[131] Amitrole is separated from other substances by HPLC with a coulometric electrode array detector. Tap and well water with concentrations down to 0.1 μg amitrole L^{-1} can be determined without any enrichment step using this method. An HPLC method for direct determination of amitrole in urine has also been described[27] based on derivative formation.

Determination of amitrole in tissues has been performed after extraction of the material with methanol. The cleanup procedure includes chromatography on aluminum oxide and gel permeation chromatography to eliminate hydrophilic substances.[132] The average recovery is 80%.

Amitrole has been extracted from plant tissues and sandy soils with ethanol, adsorbed on a resin and desorbed with ammonia.[133] After acetylation with acetic anhydride and cleanup over a Sep-Pack™ silica cartridge, the amitrole-acetyl derivative is determined by GC. The average recoveries are between 81 and 96%.

4.9 Pyridine-based Molecules

A study comparing two SPE extraction methods for the analysis of thiazopyr and its monoacid metabolite in groundwater has been reported.[134] One of the methods is based on SPE with C_{18} extraction disks and the other with GCB. Recovery studies indicate excellent recovery percentages for thiazopyr and thiazopyr monoacid by both methods.

Thiazopyr residues have also been determined in soil, grass weed and seedlings of woody plants by GC. The herbicide was extracted from soil and plant samples with ethyl acetate and plant extracts were cleaned up on a Florisil™ column.[135] The average recoveries were always higher than 97% with a relative standard deviation between 1 and 6%.

A method has been presented for determining the aquatic herbicide fluridone.[136] Samples are passed through a Sep-Pack™ C_{18} cartridge to extract fluridone and eluted with methanol. Recoveries averaged 98%.

Residue methods for determination of fluridone in meat, milk, eggs, crops, fish and crayfish have also been described.[137,138] Fluridone is extracted from meat, eggs, crops, fish or crayfish with methanol, and the herbicide is extracted from milk with a disposable C_{18} cartridge. Extracts are purified by liquid partitioning and/or alumina column chromatography.[137,138] When the metabolite of fluridone is determined together with the parent compounds, the methanolic extracts are subjected to acidic hydrolysis to release the conjugated forms of fluridone and the metabolite.[138]

Norflurazon and its initial metabolite in soil have been determined. The method consists of soil sample extraction with methanol. Recoveries from fortified soils were >90% for norflurazon and >80% for desmethylnorflurazon.[139]

The conditions for recovering the herbicide chloridazon from the soil were studied using acetone as the extractant reagent, and the optimum conditions were a time of 15 min for a single extraction. Results showed that 85.7% of the chloridazon was recovered from soil samples.

A metabolite of the herbicide pyridate (3-phenyl-4-hydroxy-6-chloropyridazine, known as CL9673) has been determined in drinking and groundwater.[140] A recovery study using tap water samples spiked with CL9673 at a concentration of 0.1 μg L^{-1} showed a recovery of 84.8%.

Pyridate and its metabolite have also been determined in green cereals, grain and straw.[20] Samples were macerated with acetone–celite and filtered. The acetone extracts were evaporated. The aqueous residue was transferred to a separating funnel with sodium acetate and dichloromethane. The dichloromethane phase contains the herbicide pyridate, and the sodium acetate phase contains the metabolite. Each phase was then extracted further. In the 0.05–1.0 mg kg^{-1} range, recoveries are

between 73.7 and 109% from green plants, between 77 and 102% from grains and between 77.3 and 87.5% from straw.

5 IDENTIFICATION AND DETERMINATION

The wide range of physicochemical properties covered by these compounds explains why many analytical techniques have been used to determine them. Table 5 summarizes the techniques employed for the compounds' determination. As can be observed in the table, HPLC is the procedure most employed to separate and detect miscellaneous pesticides. This may be attributed to general guidelines that have been established to generate new pesticides, less hazardous for human beings. Pesticides are being made more water soluble, more polar and less volatile. This group of characteristics makes HPLC an ideal technique for determining them.

5.1 Gas Chromatography

Direct determination of imidacloprid by GC is not possible because of its thermolabile and polar *N*-nitroguanidyl moiety. However, a method for determining imidacloprid in water and soil samples, after hydrolysis in a basic medium, followed by chromatography–MS and selected ion monitoring (SIM) has been reported.[53]

In contrast, acetamiprid is directly determined by GC with electron-capture detection (ECD).[54] The operating conditions of GC include the use of a packed column. The inlet and detector temperatures were both 320 °C, and the column temperature was 260 °C. In the same context a method for determining acetamiprid and its metabolites was developed by GC/ECD. Acetamiprid and its metabolites in crops were derivatized to methyl 6-chloronicotinate through alkali hydrolysis, potassium permanganate oxidation and then esterification with diazomethane.[55]

Bipyridylium herbicides cannot be determined by GC unless they are converted into volatile compounds. Reduction of DQ and PQ with $NaBH_4$ in the presence of a catalyzer such as $NiCl_2$ makes it possible to obtain the hydrogenated derivatives that could be detected by a NPD (nitrogen phosphorus detector).[79,84]

CQ is also determined by GC.[82] The technique is based on an in vitro multistep reaction by which CQ is initially N-demethylated with a potassium pentafluorothiophenolate followed by a further reaction with an excess of reagent to produce the pentafluorothiophenyl derivative. ECD and MS were used for quantitative determination and identification. However, Mortimer and Weber[192] reported that the method contains a significant error and it is not specific for CQ.

Determination of dinitroaniline herbicide concentrations in soil, air and plants has usually been carried out by GC with ECD,[98,101] NPD[99,153–155] or MS.[98,99,153,155] The concentrations of these herbicides were determined by both NPD and ECD and confirmed by their mass spectra. In this way, benefin, butralin, dinitramine, ethalfluralin, isopropalin, pendimethalin and trifluralin have been determined, often together with herbicides of other types in several matrices. Figure 1 shows the GC/MS chromatogram at the detection limit of $0.01\,\mu g\,g^{-1}$ of pendimethalin and three other herbicides in soil samples.

However, oryzalin requires a general derivatization method, since the intact chemical does not possess good gas chromatographic properties. The di-*N*-methyl derivative has been well recognized and conversion of oryzalin to this derivative has been used in determination and characterization by GC/MS.[157]

Analysis of chloroacetamides has also usually been carried out by GC with different detectors.[26,103,105–108,110–114,116,117,119,120,158–161,163,164,183,184,189] With nonspecific detection such as FID (flame ionization detection) the sensitivity of the determination is relatively low and laborious purification of the samples is needed.[159,183,184,189] To avoid these drawbacks, a comprehensive two-dimensional gas chromatograph with FID was constructed and evaluated for fast separation and analysis of different pesticides, among them alachlor, metalachlor and trifluralin.[160]

GC/NPD has been applied to the analysis of alachlor, propalachlor, butalachlor, metolachlor and two alachlor degradation product residues.[26,106,116,120,158] ECD is a highly sensitive detector for these compounds. It has been used in the analysis of acetochlor, alachlor, butachlor, metolachlor and propachlor.[110,113,154,161] Other authors have used ECD to analyze alachlor and its metabolite after derivatization to the heptafluorobutyryl derivative.[119] This method is highly sensitive but the analysis is complicated. The detection of chloroacetamides also has been performed using ECD and NPD in parallel.[105]

GC/MS has been used as a confirmatory technique.[26,108,110,116,120] GC/MS also allows routine determination of chloroacetamides simultaneously with other herbicides such as triazines, carbamates and phenoxyacetic acids.[103,107,112,114,117,163] Isotope dilution GC/MS has been used to determine alachlor, metolachlor, atrazine and simazine in water and soil.[111] Known amounts of ^{15}N, ^{14}C-alachlor and $^{2}H_5$-atrazine are added to each sample as internal standards. The GC/MS system with data acquisition in SIM mode was used to quantify herbicides in the extract.

The chromatographic separation of enantiomers and diastereoisomers of some acetamide pesticides was investigated using achiral and chiral GC/MS.[164] The

Table 5 Determination techniques employed for miscellaneous pesticides and related compounds

Class	Pesticide	Determination	Samples	Refs.
Macrocyclic lactone	Avermectin	HPLC fluorescence	Air, surfaces, lettuce, bread, wine, apples, cattle feces, soil, hops tomatoes, bovine tissue, orange and celery	37–43, 46, 49
		HPLC/UV	Lettuce, cucumbers, biological samples	44, 48
	^{3}H-avermectin	LSC	Citrus fruit and foliage, rats, soil, plants	7, 21, 141
		HPLC/UV and LSC TLC autofluorography		
	^{14}C-avermectin	LSC	Soil, plants	21
	Abamectin-8,9-oxide	HPLC Visible	Standard	142
	Emamectin	HPLC fluorescence	Water, celery and lettuces	10, 45
	^{3}H-emamectin	HPLC/LSC	Fish sample	47
	^{14}C-emamectin	HPLC/UV HPLC/LSC HPLC/MS/MS	Soil	9
	Ivermectin	HPLC fluorescence	Cattle feces	49
	^{3}H-ivermectin	HPLC/UV and LSC	Rats	143
Chloronicotinyl	Imidacloprid	HPLC/UV	Liquid and solid formulations, water, soil, cucumber, rice	4, 50, 51
		HPLC/DAD	Pepper, tomato, cucumber,	52
		HPLC/PAD	Soil	11
		GC/MS	Soil, water	53
	Acetamiprid	GC/ECD	Fruit and vegetables, water	54, 55
Tetranortri-terpenoid	Azadirachtin	HPLC/UV	Stability studies, insecticidal formulations, waters	13, 22, 56, 58
		HPLC/DAD	Neem extracts	57
		SFC/UV	Neem seeds extracts	12, 58
Ammonium quaternary salt	DQ	HPLC/UV	Serum, urine, plasma, vitreous humor, blood, bile, liver, kidney, water, crops, oil	16, 17, 60, 62–64, 66, 68, 70, 76, 81, 83, 89, 94, 96
		HPLC/DAD	Crops	77
		HPLC/MS	Water, soil	15, 61, 67, 144, 145
		GC/NPD	Crops, blood and urine	79, 84
		CE/UV	Serum, water, soil, potatoes	59, 60, 69, 75, 91, 146, 147
		CE/MS	Water	148, 149
		SWV	Water	150
		Spectrophotometric	Plasma, water, soil, potatoes, biological materials	59, 78, 85, 87, 88
		Spectrofluorimetric	Water, potatoes, flowers and soil	71
	PQ	HPLC/UV	Plasma, urine, vitreous humor, brain, water, oil	16, 17, 60, 63, 64, 66, 68, 70, 76, 81, 89, 95, 96
		HPLC/DAD	Crops	77
		HPLC/MS	Water, soil	15, 61, 67, 144, 145
		GC/NPD	Crops, biological materials	79, 84
		ELISA	Fruits	74
		CE/UV	Serum, water	14, 60, 69, 75, 91, 146, 147
		CE/MS	Water	148, 149
		SWV	Water	150, 151

Table 5 (*continued*)

Class	Pesticide	Determination	Samples	Refs.
		Spectrophotometric	Water, grain and plant materials, plasma, urine	65, 72, 73, 85, 87, 90, 92, 93
	DF	HPLC/UV	Water	60, 63, 64, 66
		HPLC/MS	Water, soil	15, 152
		CE/UV	Water	14, 60, 146
		CE/MS	Water	148
	CQ	GC/ECD and GC/MS	Cotton seed	82
		HPLC/MS	Water, soil	15
		CE/UV	Water	146
	MQ	HPLC/CD	Animal and plant matrices	80
		HPLC/MS	Water, soil	15
		CE/UV	Water	146
Dinitroaniline	Benefin	GC/MS	Soil	98
		GC/ECD	Soil	98
	Butralin	GC/NPD	Soil, plant, air	99
		GC/MS	Soil, plant, air	99
	Dinitramine	GC/NPD	Soil, plant, air	99
		GC/MS	Soil, plant, air	99
		HPLC/UV	Soil, water	18
	Isopropalin	GC/MS	Soil	98
		GC/ECD	Soil	98
		HPLC/UV	Soil, water	18
	Pendimethalin	GC/NPD	Soil, plant, air, rainwater	99, 120, 153, 154
		GC/MS	Soil, plant, air, water	99, 112, 120, 153, 155
		GC/ECD	Soil	101
		HPLC/UV	Soil, water	18, 156
	Ethalfluoralin	GC/NPD	Soil, plant, air	99
		GC/MS	Soil, plant, air	98, 99
		GC/ECD	Soil	98
		HPLC/UV	Soil, water	18
	Oryzalin	HPLC/UV	Soil	100
		GC/MS	Water	157
	Prosulfalin	HPLC/UV	Soil	100
	Trifluralin	GC/NPD	Soil, plant, air, estuarine water	99, 116
		GC/MS	Soil, plant, air, water	98, 99, 112, 116
		GC/ECD	Soil	98
		HPLC/UV	Soil, water	18
		HPLC/DAD	Estuarine waters	116
Acetamide	Alachlor	GC/NPD	Rainwater, soil, groundwater, well water	105, 106, 116, 120, 154, 158
		GC/FID	Formulations, human serum	159, 160
		GC/ECD	Corn, soybean, potato, groundwater, plants, soil	105, 110, 119, 161
		GC/MS	Well water, groundwater, soil	26, 107, 108, 110–112, 114, 116, 117, 120, 162–164
		HPLC/UV	Formulations, urine, water	143, 164, 165
		HPLC/DAD	Well water	116, 166
		HPLC/ED	Crops	118
		HPLC/MS	Water, urine	104, 109, 167
		ELISA	Water, urine, meat, soil	108, 121–123, 143, 162, 166, 168–175
		Biosensor	Water	176–178
		CE/UV	Water	179
		Optic sensor	Water	180
		FABMS	Water	181
		NMR	Water	181

(*continued overleaf*)

Table 5 (*continued*)

Class	Pesticide	Determination	Samples	Refs.
	Acetochlor	HPLC/DAD	River water and rain	102
		HPLC/ED	Crops	118
		HPLC/MS	Water	109
		GC/MS	River water and rain	102
		GC/ECD	Plant, soil	161
		NMR	Water	181
		FABMS	Water	181
	Amidochlor	ELISA	Water	182
	Butachlor	GC/FID	Formulations	183, 184
		GC/ECD	Water	113
		GC/MS	River water, groundwater	26, 103, 112
		HPLC/UV	Formulations	183
		HPLC/ED	Crops	118
		ELISA	Water	182
		FABMS	Water	181
		NMR	Water	181
	Dimethenamid	GC/MS	Water	114
	Metazachlor	HPLC/UV	Water	115
		HPLC/ED	Water	115
	Metolachlor	GC/NPD	Rainwater, soil groundwater, estuarine water	26, 116, 120, 154
		GC/ECD	Plant, soil	161
		GC/MS	Wellwater, soil, estuarine water	107, 108, 111, 114, 116, 117, 120, 164
		HPLC/MS	Water	109
		HPLC/UV	Carrot, water	164, 185
		ELISA	Water, soil	108, 122, 173, 182, 186, 187
		CE/UV	Water	179
		FABMS	Water, soil	181, 188
		NMR	Water	181
	Pretilachlor	GC/MS	River water	103
	Propachlor	GC/FID	Formulations	189
		GC/NPD	Groundwater	26
		GC/ECD	Plant, soil	161
		FABMS	Water	181
		NMR	Water	181
Oxime	Clethodim	HPLC/UV	Water solutions	5
	Alloxydim	GC/MS	Water	124
	Sethoxydim	HPLC/UV	Rapeseed	125
Triazole	Aminotriazole	HPLC/EAD	Drinking and groundwater	131
		HPLC/ED	Drinking and groundwater, animal tissues	19, 27, 126, 128, 129, 132
		HPLC/UV	Drinking water	126, 128
		GC/NPD	Plant tissues, sandy soils, water	130, 133
		TLC	Water, formulations	6, 127
Pyridine-based molecule	Thiazopyr	GC/NPD	Soil	135, 153
		GC/MS	Soil	153
		GC/ECD	Groundwater	134
		HPLC/DAD	Groundwater	134
	Fluridone	HPLC/UV	Water, meat, eggs, crops, fish and crayfish	136–138
	Norflurazon	HPLC fluorescence detection	Soil	139
	Chloridazon	HPLC/DAD	Formulations	190
		Voltammetric	Soil	191
	Pyridate	HPLC/UV	Plant material	20
		HPLC/ECD	Water	140

CD, conductivity detection. See text for explanation of other acronyms.

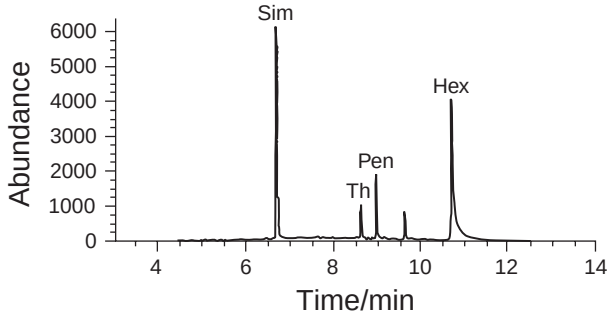

Figure 1 GC/MS detection limit of simazine (Sim), thiazopyr (Th), pendimethalin (Pen) and hexazinone (Hex) in soil samples in SIM ($0.01\,\mu g\,g^{-1}$). (Reproduced with permission from R.A. Pérez, C. Sánchez-Brunete, E. Miguel, J.L. Tadeo, *J. Agric. Food Chem.*, **46**, 1864–1869. Copyright (1998) American Chemical Society.[153])

compounds studied were alachlor, acetochlor, metalaxyl, metolachlor, and dimethenamid. Whereas alachlor is achiral, all the other compounds are axial- and/or C-chiral and consist of two or four stereoisomers, as is shown in Figure 2.

Alloxydim-sodium herbicide and its degradation product have been determined by GC/MS. Residues were derivatized and transformed into 2-methoxycarbonyl-3,3-dimethylpentane-1,5-dioic acid, which was then determined.[124]

A GC method with NPD is described for the determination of amitrole in water, plant and soil samples. The method is based on the acetylation of amitrole with acetic anhydride.[130,133]

Pyridine-based molecules have seldom been determined by GC. Thiazopyr has been the one most often determined by GC and different detectors, such as

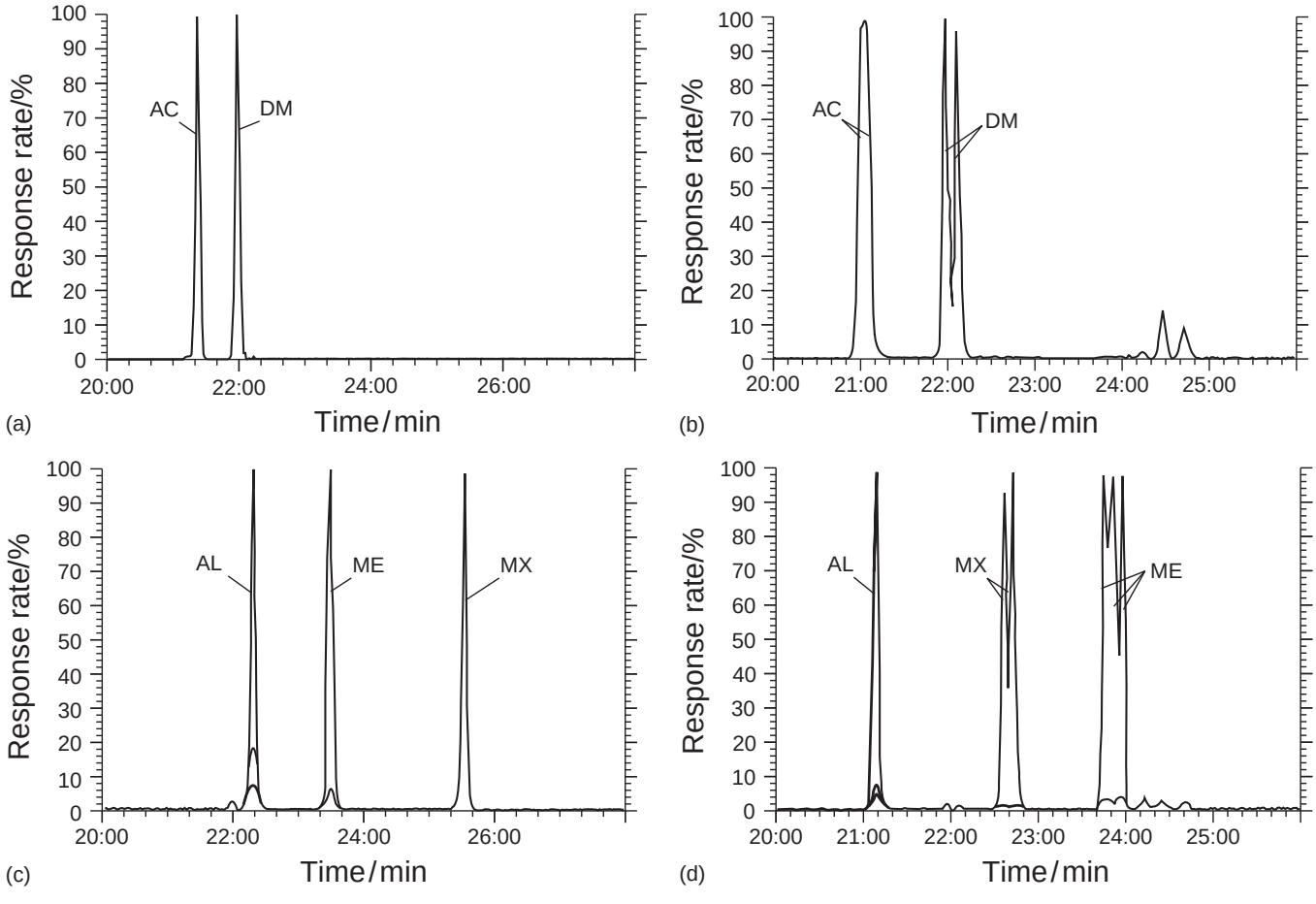

Figure 2 EI (electron impact) mass and SIM chromatograms (m/z 223 + 230, (a), (b); m/z 206 + 237 + 238, (c), (d)) showing elution of acetamide pesticides on: (a), (c) the achiral SILOV (silar 10c/ov1701) and (b), (d) the chiral OV1701-BSCD (*ter*-butyldimethylsilyl-β-cyclodextrin) HRGC (high-resolution gas chromatography) columns. Note the elution of single peaks from the achiral column. Note the marginal resolution of acetochlor into the atropisomers on the chiral column. Note the resolution of metalaxyl, metolachlor and dimethenamid into two or three isomers on the chiral HRGC column. Abbreviations: AL, alachlor; AC, acetochlor; MX, metalaxyl; ME metolachlor; DM, dimethenamid. Retention times in minutes. (Reproduced with permission from H.R. Buser, M.D. Müller, *Environ. Sci. Technol.*, **29**, 2023–2030. Copyright (1995) American Chemical Society.[164])

ECD,[134] NPD and MS[135,153] have been used. The derivatization of fluridone with phosphorus tribromide to determine it by GC/ECD has been proposed as an alternative or confirmatory procedure.[137]

5.2 High-performance Liquid Chromatography

Several liquid chromatographic methods have been reported for determining macrocyclic lactone residues in various matrices. HPLC-fluorescence detection after dehydration of the residues with trifluoroacetic anhydride[10,37–43,45,46,49] is the most widely applied determination technique. The major advantages of this method are instantaneous fluorescence derivatization, high sensitivity, improved selectivity and the ability to quantify ivermectin, avermectin B_1, emamectin and their 8,9-δ degradation product. Figure 3 shows a typical chromatogram for the derivatized avermectin B_1 containing a B_{1a} : B_{1b} ratio of about 93 : 7 in untreated whole apples fortified with avermectin B_1 at the 2.1 ppb level.

A more rapid HPLC method has been described[44] for the detection of abamectin in cucumber and lettuce. The method employs UV detection based on the conjugated-diene chromophore of abamectin, which has an absorbance maximum at 245 nm. This method, however, lacks sensitivity and the δ 8,9-isomer is not measured. In addition, when the method is employed with other samples, it shows matrix peaks that interfere with it and prevent quantification of abamectin. Junsuo and Qian[48] use immunoaffinity chromatography for cleanup and concentration followed by UV detection.

Several methods have been described for the determination of abamectin and emamectin benzoate using radiolabeled compounds (^3H and ^{14}C) in a variety of samples.[9,21,47,141] The samples were analyzed by HPLC and the compounds determined on a direct liquid scintillation counter (LSC).

The chloronicotinyl insecticide imidacloprid has been extensively determined by reversed-phase high-performance liquid chromatography (RPHPLC) with UV detection at 270 nm using a mobile phase of acetonitrile–water.[4,50,51] The limits of detection ranged from 0.01 to 1.0 mg kg^{-1} depending on the material analyzed. The methodology for determining imidacloprid residues in vegetables[52,193] by HPLC diode array detection (DAD) has been developed owing to its suitability for selecting the optimum wavelength for maximum sensitivity and confirmation from UV spectral information.

HPLC with pulse amperometric detection (PAD) has been used to determine the insecticide imidacloprid and its major metabolite 6-chloronicotinic acid in soil. Electrochemical study of the compounds demonstrates the reduction and adsorption processes on the mercury electrodes. PAD overcomes electrode fouling problems by repeated and continuously applying cleaning potentials to the working electrode. The measurement potential was −1300 mV.[11] Figure 4 gives the chromatograms obtained with amperometric detection in direct current (DC) mode and pulse mode.

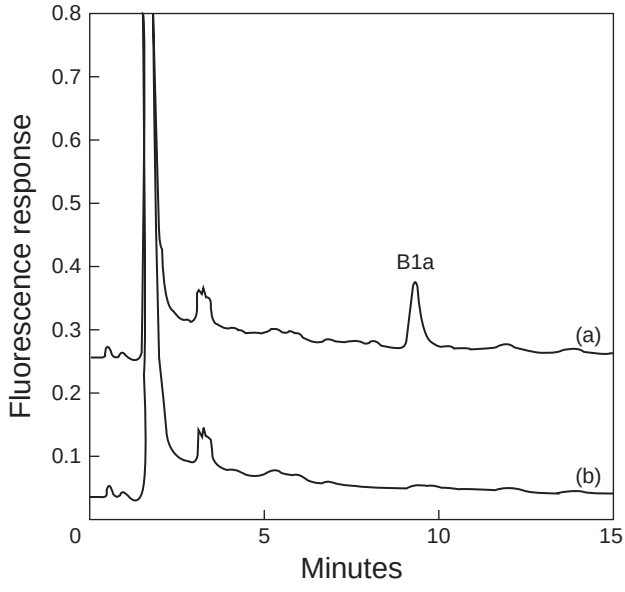

Figure 3 Typical LC chromatograms for apple fortified with 2 ppb avermectin B_1 (B1a) (a) and apple control, 0 ppb avermectin B_1 (b). (Reproduced with permission from J.A. Cobin, N.A. Johnson, *J. AOAC Int.*, **78**(2), 419–423 Figure 3. Copyright (1995) AOAC International.[41])

Azadirachtin is nonvolatile and highly polar. It is unsuitable therefore for GC and only has UV absorption at a very short wavelength. RPHPLC has been successfully used for separation and quantification of azadirachtin.[13,22,56–58] The problem is that common aliphatic and aromatic solvents and surfactants used to formulate the products interfere in the region of the azadirachtrin-A peak.[57]

HPLC is the analytical instrument of choice for ammonium quaternary salts because they are readily soluble in water and in another polar solvent and are not volatile. The most commonly used stationary phases are reversed-phases (RP), comprising silica bonded to carbonate chains of several lengths, mainly C_8 and C_{18}. Ion-pair formation in LC (liquid chromatography) with a UV detector is a common technique for determining quaternary ammonium compounds, for example by using hexanesulfonate,[70] heptanesulfonate,[60,95] octanesulfonate,[17,83,89,94] ammonium formate,[152] ammonium acetate[67,145] or potassium bromide.[96] as counter ions.

C_{18}- and C_8-bonded silica stationary phases always have free silanol groups that complicate the analysis of quaternary amines like bipyridylium herbicides, provoking the

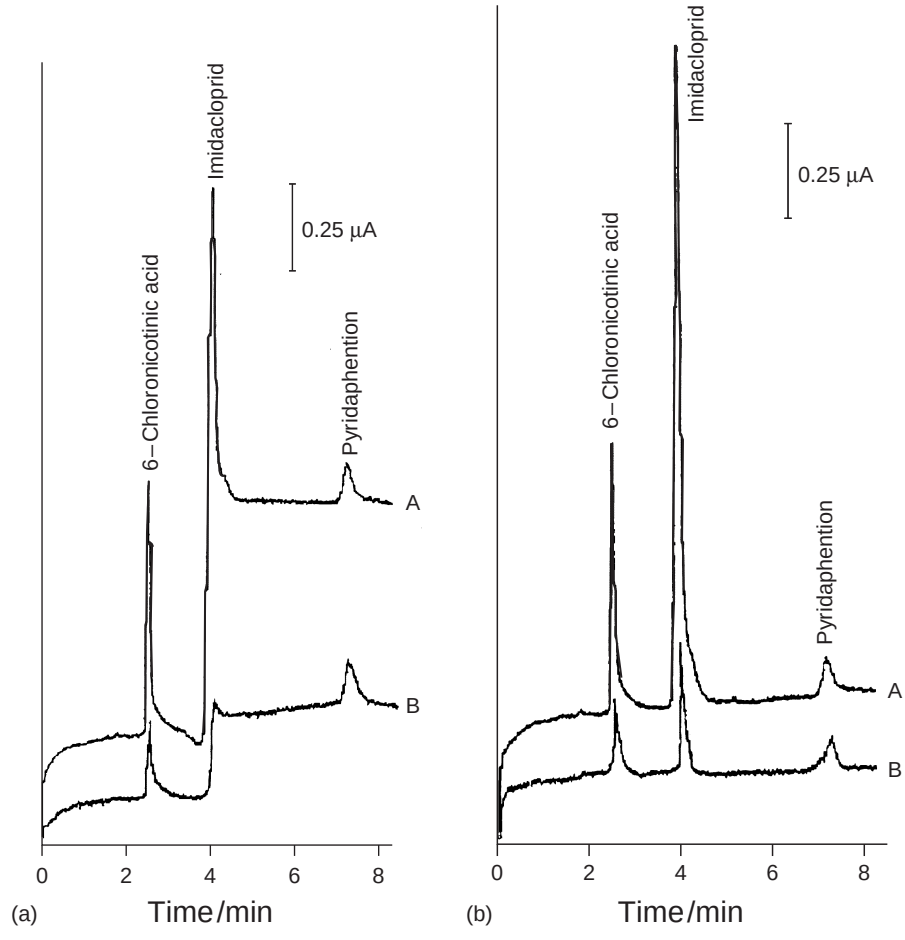

Figure 4 Chromatograms in (a) DC detection mode, E = −1300 mV, (b) pulse mode. Flow rate: 1.0 mL min^{-1}, injection volume: 25 μL, drop size 2.5 mm^2, amount injected: (A) 500 ng and (B) 100 ng of each substance. (Reproduced from N. Ruíz de Erenchum, Z. Gómez de Balugera, M.A. Goicolea, R.J. Barrio, *Anal. Chim. Acta*, **349**, 199–206. Copyright (1997) with permission of Elsevier Science B.V.[11])

appearance of tailing peaks and increasing their width, and originating a variation in the retention times even when the same column is used. To avoid this one of four actions can be taken:

1. Choose another type of column.
2. Reduce the ionization of the silanol groups by employing mobile phases of pH < 4.
3. Add a silanol group blocking agent to the mobile phase.
4. Increment the ionic strength of the mobile phase.

In addition to the counter ion, the mobile phases reported contain an acid, such as phosphoric acid,[15,17,61,62,70,81,89,94–96,152,194] glacial acetic acid[67] or trifluoroacetic acid,[145] together with a silanol group blocking agent, such as diethylamine[17,70,89,94,194] or triethylamine.[95,96] Because of the basic character of the latter, they also help to maintain the pH around 3.

In order to avoid the free silanol group effect, analytical columns are being substituted for the base made from silica particles for polymeric packings such as polystyrene–divinylbenzene. These columns are also more resistant to the acid pHs.[16,68] The mobile phases contain phosphoric acid and triethylamine.

The utility of silica as a stationary phase for HPLC analysis of quaternary ammonium compounds has been discussed.[63,64,76,77] A silica column with sodium chloride as an ion-pair reagent in acetonitrile–water as the mobile phase provided rapid efficient selective separation of PQ and DQ.[76,77] The mechanism involved in this chromatographic process has not been well defined and the interactions between the normal phase (NP) and the herbicides may include ionic interactions between the positive charge of the ammonium quaternary group and the negative one of the silanol group. The herbicide fixed to the silica can be liberated using displacing agents in the mobile phase such as tetramethyl ammonium

hydroxide.[63,64] Using silica columns, PQ, DQ and DF were chromatographed as symmetrical well-resolved peaks whose retention time can be controlled by the concentration of added salts.

An analytical column of GCB has also been tested for chromatographic determination of DQ, PQ and DF.[66] The elution was performed with a gradient of pH 3 aqueous solution of tetramethylammonium hydroxide and ammonium sulfate, and methanol. A hypercarb column was found to give a low probability of false positives for bipyridylium herbicides because it is selective for polar compounds.

Conventional detection of ammonium quaternary compounds in HPLC is usually achieved with UV and DAD detectors. The most commonly used is the UV detector, monitoring wavelengths of 310 nm for DQ, 255 nm for PQ and 258 nm for DF.[17,60,63,64,66,70,81,83,89,94–96] In combination with DAD, an unambiguous identification in environmental samples is sometimes possible.[16,67,68,76,77]

However, the international agencies prefer methodologies which confirm identity using MS. The direct flow injection thermospray (TSP) MS of DQ and PQ demonstrates that the addition of a volatile buffer salt to the mobile phase is necessary because it provides a strong increase in sensitivity.[61,144]

The problem of coupling HPLC with MS detection to determine ammonium quaternary compounds is that nonvolatile ion-pairing reagents are used to facilitate chromatography. These reagents contain insoluble sodium and potassium salts, phosphate and sulfonates, which are not amenable to MS confirmation of ammonium quaternary compounds, because these salts quickly precipitate and clog the HPLC/MS interface.

Yoshida et al.[145] and Barcelo et al.[15] characterize PQ, DF, DQ, MQ and CQ by HPLC/TSP/MS. Under RP conditions and using ammonium formate or ammonium acetate as ionizing additives, 1–2 μg of each compound was positively identified under full scan conditions. For the characterization of DF, a postcolumn ion-pairing extraction system was used. An aqueous mobile phase with sulfonate-type counter ion was applied and an extraction solvent containing cyclohexane–dichloromethane–*n*-butanol was used in HPLC/TSP/MS.[152] An HPLC method compatible with particle beam (PB) MS for DQ and PQ confirmation has also been developed; it ensures correct identification and quantification of both compounds.[67]

An analytical method for the determination of MQ residues has also been reported.[80] The final separation and determination is based on ion-pair chromatography using a column switching system and CD.

An HPLC chromatographic method for the simultaneous determination of the dinitroaniline herbicides dinitramine, ethalfluralin, trifluralin, pendimethalin and isopropalin has been proposed. In order to achieve the separation of dinitroaniline herbicides, different RP (C_1,C_6,C_8, (octadecylsilica) ODS-1 and ODS-2) columns were employed.[18] Each column allowed separation of the five herbicides with different water–acetonitrile mixtures. The elution order was the same (first and last, respectively) for dinitramine and isopropalin on all the columns tested, whereas ethalfluralin, trifluralin and pendimethalin were eluted with different orders in the different columns.

Analytical procedures have also been described for the trace analysis of oryzalin and prosulfalin residues in soil using an optimized HPLC system.[100] The method requires no derivatization.

HPLC has recently been shown to be useful for trace analysis of a number of pesticide residues. Pendimethalin, chlorpyriphos, and the two esters 2,4-DIOE (isooctyl ester of (2,4-dichlorophenoxy)acetic acid) and 2,4-DPBEE (butoxyethanol ester of 2-(2,4-dichlorophenoxy)propionic acid) were quantified by HPLC using a C_{18} column, a methanol–water (82 : 18) mobile phase and UV detection.[156]

HPLC methods allow direct determination of chloroacetamides and their metabolites by UV and/or ED (electrochemical detection).[115,118,165,183,185] As a complementary technique, HPLC/DAD can be employed for confirmation of these herbicides only if there is sufficient quantitative enrichment.[102,116,165]

HPLC/MS/MS has been used to detect and identify alachlor, its ethanesulfonate metabolite, and its hydrolysis product 2,6-diethylaniline in water and urine.[104,167] One relatively new method for analyzing herbicides in water involves using HPLC coupled with MS by means of electrospray ionization (ESP). The technique has been applied to the trace analysis of oxanilic and sulfonic acids of acetochlor, alachlor and metolachlor.[109]

Chiral HPLC using modified cellulose and phenylglycine columns also showed some isomer resolution and was used for the isolation of acetochlor and metolachlor. The (+) and (−) enantiomers were assigned to acetochlor and metalaxyl using chiral HPLC in combination with chiroptical detection.[164]

Clethodim herbicide has been separated from acid and photodegradation products (PDPs) by HPLC.[5] Using an adequate mobile phase gradient, 31 peaks were separated following photocatalysis, as illustrated in Figure 5. Addition of acetic acid to the mobile phase improved peak symmetry and HPLC separation. Sethoxydim and its major active metabolite, sethoxydim sulfoxide, has been determined by normal-phase high-performance liquid chromatography (NPHPLC) using a silica column and methanol–dichloromethane containing acetic acid.[125]

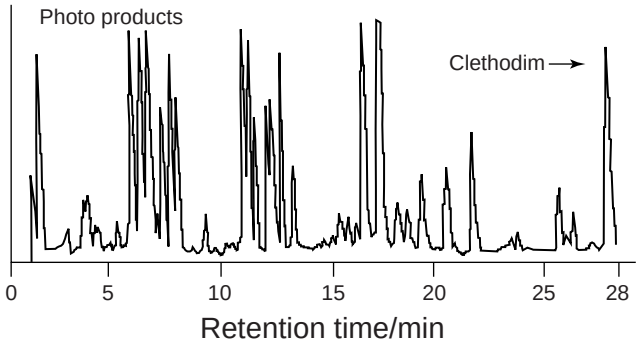

Figure 5 Chromatogram of PDPs of clethodim showing separation by LC with UV detection at 254 nm. (Reproduced with permission from L.N. Falb, D.C. Bridges, A.L. Smith, *J. Assoc. Off. Anal. Chem. Int.*, **74**(6) 999–1002, Figure 2. Copyright (1991), AOAC International.[5])

The excellent water solubility and the high polarity of amitrole makes it impossible to determine by RPHPLC with the usual aqueous eluent. The retention behavior shows no separation between amitrole and the solvent peaks. It is well known that satisfactory HPLC separations can be achieved in presence of ion-pairing reagents like sodium dodecyl sulfate (SDS),[128,129,131,132] sodium butanesulfate[126] and tetrabutylammonium sulfate.[27] Figure 6(a) is a chromatogram corresponding to an injection of a solution containing $50\,\mu g\,L^{-1}$ of aminotriazole, and Figure 6(b) of a solution containing $0.5\,mg\,L^{-1}$, where the separation between solvent peak and amitrole is accomplished with SDS. The separation has also been done by NPHPLC on a polar cianopropyl-column using hexane, propanol, water with $LiClO_4$ and trichloroacetic acid as electrolytes[19] or using ion exchange chromatography.[126]

The direct UV detection of amitrole suffers from a fairly low absorption coefficient and cannot be used.[126] For greater sensitivity, derivatives of amitrole have to be formed by diazotization by coupling with 1-naphthol-3,6-disulfonic acid.[27,126] Without a prior derivatization step, amitrole is directly detectable by an amperometric technique following ion-pair HPLC separation[19,128,129,132] and coulometric electrode array detection (EAD).[131] HPLC with UV after derivatization and ED required enrichment steps for the determination of concentrations in the range of $0.1\,\mu g\,L^{-1}$. Samples with concentration up to $1\,\mu g\,L^{-1}$ can be directly injected. Using EAD, tap and well water with concentrations down to $0.1\,\mu g$ amitrole L^{-1} can be determined without any enrichment steps.

Thiazopyr was determined by HPLC/DAD in water.[134] It was hoped that DAD would provide spectral information to augment confirmation of incurred thiazopyr and its metabolite residues. Unfortunately, there

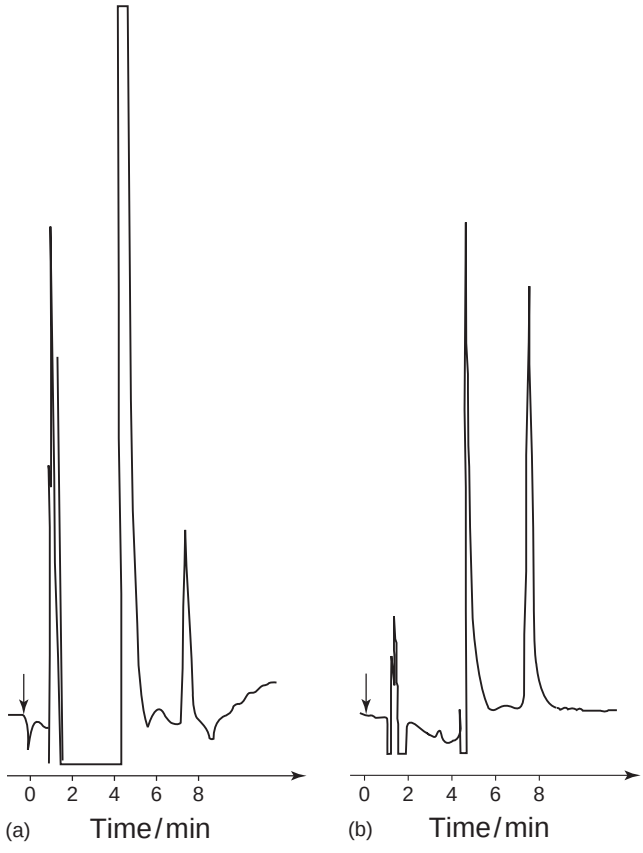

Figure 6 Detection limit and system peaks for aminotriazole. Column: C_{18} silica, 5 µm, 15 cm × 0.46 cm internal diameter; mobile phase: 65% acetonitrile 0.05 M SDS at pH 3 with perchloric acid and 35% acetonitrile; ED 0.85 V (vs Ag/AgCl). (a) Injection of 100 µL of $0.05\,mg\,L^{-1}$ solution, 1 µA full scale and (b) injection of 100 µL of a $0.5\,mg\,L^{-1}$ solution, 5 µA full scale. (Reproduced from V. Pichon, M.C. Hennion, *Anal. Chim. Acta.*, **284**, 317–326. Copyright (1993) with permission of Elsevier Science B.V.[128])

were no distinctive wavelength maxima that could be exploited for monitoring purposes.

Several methods for determining of underivatized fluridone in milk, animal tissues and crops by HPLC have been described.[136–138] HPLC methods are advantageous in that they are rapid and avoid the use of hazardous derivative reagents.

Norflurazon and its initial soil metabolite can be determined using HPLC with fluorescence detection.[139] The fluorescence properties of norflurazon dissolved in acetonitrile were examined previously. Adding water or acidifying the solution had no effect on the observed fluorescence. A method for measuring chloridazon with an HPLC microbore column and DAD has been reported.[190]

The application of a simple RPHPLC/UV method for the determination of the contact herbicide pyridate and its main metabolite CL9673 has been described.[20]

Amperometric detection has also been reported for the metabolite.[140]

5.3 Immunoassay

Immunoassay offers certain advantages over conventional instrumental methods for the analysis of pesticide residues. Inmunoassays are highly efficient and cost effective and are ideally suited for screening large numbers of samples for low levels of specific analytes.[195] An enzyme immunoassay uses either a labeled antibody or a labeled analog of the target compound as the detection marker and can be performed in different formats. The most common format is the competive enzyme-linked immunosorbent assay (ELISA). In competitive ELISA, the antibodies specific to the analyte are immobilized onto a solid phase. The enzyme–hapten conjugate and the analyte are added to the antibody-coated tube or plate for incubation. After incubation, the unreacted material is removed, and a substrate and chromogen are added and allowed to react for a short period of time during which the enzyme converts the substrate-chromogen to a colored product. The absorbance can either be measured with a spectrophotometer or estimated visually.

In some ELISAs the solid phases employed are polystyrene wells, balls or tubes, on which the antibody or hapten–protein conjugate are passively adsorbed. The desorption or leaching off of antibody or other proteins which have been passively adsorbed is a major factor that adversely affects sensitivity and precision assays. Variability of wells within microtiter plates has been shown to be the greatest contributor to total assay imprecision. A particle-based ELISA eliminates these imprecision problems through covalent coupling of antibody to the solid phase. The uniform dispersion of particles through the reaction mixture allows for rapid reaction kinetics and precise addition of antibodies.[169,170]

A competitive ELISA for the quantification of free PQ cation was developed using polyclonal rabbit antiparaquat antibodies bound to a magnetic particle solid phase.[74] The immunoassay described utilizes a competitive format. Since enzyme-labeled PQ competes with the unlabeled (sample) PQ cation for the antibody sites, the color developed is inversely proportional to the concentration of PQ cation on the sample. The speed and sensitivity of the method makes it suitable for large numbers of samples in short periods of time.

A carboxylic derivative of metolachlor, *N*-(chloroacetyl)-*N*-(1-methyl-2-methoxyethyl)-2-methyl-4-(4′-carboxybutoxy)-6-ethylaniline, was synthesized and covalently linked to a carrier protein to prepare monoclonal antibodies. An antibody with high binding affinity for metolachlor was used to develop two competitive ELISA procedures, a direct one with enzyme-labeled antigen and an indirect one with an enzyme-labeled second antibody.[186] No cross-reactivity was observed with alachlor, furalaxyl, metalaxyl, and most of the metabolites of metolachlor and metalaxyl.

Antibodies against alachlor were generated using an alachlor–protein conjugate. Alachlor was conjugated to the protein carrier through the chlorine-bearing carbon via a thioether bond. This type of linkage was easily obtained because of the reactivity of the chlorine-bearing carbon with nucleophiles such as glutathione. Furthermore, the sulfur atom facilitates the formation of antibodies that recognize alachlor. Conjugation of alachlor to proteins via the thioether linkage is a simple one-step reaction requiring no synthesis of novel haptens.[123,172] Rabbit polyclonal antibodies generated using the thioether conjugate were useful in the development of a sensitive alachlor ELISA. The ELISA discriminated alachlor from other chloroacetanilide herbicides, although strong cross-reactivity was detected with some thioether metabolites.[171,175] Subsequently an analog of alachlor was conjugated to proteins through a carboxyl group in the methoxymethyl side chain. Antibodies from the carboxy–alachlor conjugate showed even greater specificity towards alachlor and discriminated against other chloroacetamides as well as alachlor thioether metabolites.

Using the same thioether conjugation method, ELISAs for other chloroacetanilines herbicides, metolachlor, amidochlor and butachlor, have been developed.[182] These chloroacetanilide herbicides were easily conjugated to proteins via a thioether linkage.

The ELISA method was evaluated for the detection of alachlor in environmental water samples and was compared with an established GC/MS method. The results indicate that the alachlor ELISA can be effectively utilized as a primary screen to select environmental samples for confirmatory instrumental verification of the presence of alachlor at low parts per billion levels.[168,195] However, an ELISA for the detection of alachlor or putative alachlor metabolites in human urine has been described and compared with HPLC analysis, and the two analytical methods gave results with statistically significant differences.[143]

An automated flow-injection liposome immunoanalysis system was developed using alachlor as a model compound.[176–178] The experimental system consists of an autosampler attached to an immunoreactor column containing antialachlor antibody covalently coupled to glass beads.

The utility of immunoassay techniques as part as groundwater monitoring/compliance programs has been evaluated.[166] There are several commercially available kits for determining chloroacetamides.[121,122,174] The

results provide evidence that immunoassay techniques are simple and suitable. However, false positives should be tested for possible metabolites that may cross-react with the antibody.[173,187]

5.4 Other Techniques

5.4.1 Spectrometric and Spectrofluorimetric Methods

Methods for the determination of PQ and DQ by spectrophotometry have been reported. The spectrophotometric methods most often used are based on their reduction in alkaline solution. Reduction of PQ with sodium dithionite forms a blue reaction product[73,90,92] and reduction of DQ forms a green[71] or red one.[88] Ascorbic acid[65,72] or sodium borohydride[93] also form a blue free radical ion with an absorbance maxima at 600 nm with PQ. PQ and DQ can be determined simultaneously using second-derivate spectroscopy after reduction with NaOH dithionite solution.[85,87]

It should be noted that the reaction product formed with DQ and dithionite is also fluorescent and can be measured using an excitation wavelength of 428 nm and an emission wavelength of 497 nm.[71] Determination of DQ by photokinetic methods has been described;[59,78] it is based on the rate of photoreduction of DQ by ethylenediaminetetraacetic acid (EDTA) sensitized by acridine yellow in absence of oxygen[78] or on the formation of a charge-transfer complex between DQ and cysteine.[59]

5.4.2 Thin-layer Chromatography

The relationships between the structure of N_1- or N_4- substituted 1,2,4-triazoles and their behavior in thin-layer chromatography (TLC) on silica gel were studied. Linear correlations were obtained between the retention parameter, R_m, and the logarithm of the molar fraction of a strong solvent in eight binary solvent systems for five pairs of positional isomers.[6] Detection using 2-dichloromethylbenzoimidazole as the reagent on TLC has been described.[132] A study on the adsorption and dissociation of the herbicide amitrole has been performed using radiolabeled compounds and TLC.[127]

5.4.3 Supercritical Fluid Chromatography

SFC is a suitable technique for determining azadirachtin in crude extracts of neem. Almost all the accompanying substances are less polar than azadirachtin and are eluted quickly. On the RPHPLC they elute very slowly and must be removed by pretreatment. The UV absorbing peak for azadirachtin can be separated from the other substances present in a relatively short time.[12,56] Figure 7 shows the chromatograms obtained by SFC and HPLC. The chief differences between SFC and HPLC were the much longer time required for an HPLC separation and the much larger volume of solvent required.

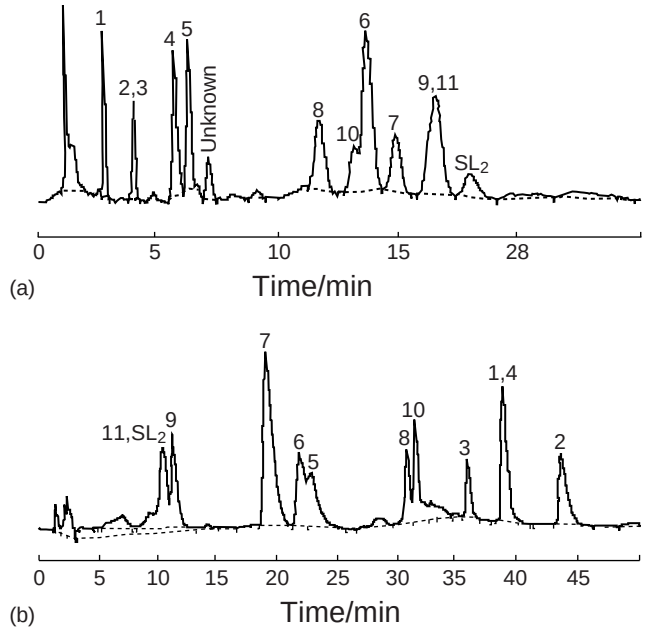

Figure 7 Chromatograms of triterpenoids found in neem seeds. (a) SFC and (b) HPLC. Peak identification: (1) nimbin, (2) salamin, (3) 6-desacetylnimbin, (4) 3-desacetylsalamin, (5) 3-tigloylazarachtol, (6) azadirachtin, (7) azadirachtin-D, (8) 3-acetyl-1-tigloylazadirachtinin, (9) azadirachtin H, (10) salannolide, (11) azadirachtin I. The structure of the compound SL_2 has not been determined. The peak areas do not represent their natural abundance in seeds, as the minor components have been enriched to make them more easily visible. (Reproduced from S. Johnson, D. Morgan, *J. Chromatogr. A.*, **761**, 53–63. Copyright (1997) with permission of Elsevier Science B.V.[58])

5.4.4 Capillary Electrophoresis

CE techniques have proven to be promising alternatives to the separation and/or analysis of various ionogenic pesticides. Separations in CE are based on differences in the electrophoretic mobility of the injected ions. Tomita et al.[91] have described the use of CE to determine PQ and DQ. The optimal conditions for CE separation of DQ, PQ and DF are well established[14,69,75,147] and have been compared with HPLC determination.[60] CE was simpler and cheaper, but HPLC offered a higher sensitivity for real samples and was independent of the matrix effects. Investigation into the application of indirect UV detection in CE published by Galceran et al.[146] showed the suitability of an absorbing carrier electrolyte for the determination of the nonabsorbing quats, CQ and MQ, and also the simultaneous detection of the

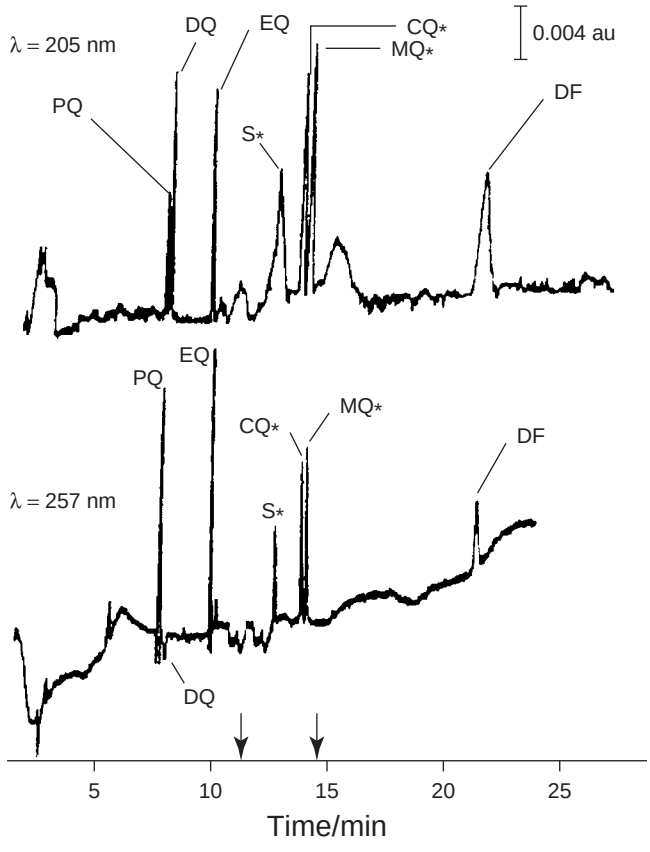

Figure 8 Electroferograms obtained under optimal conditions, 10 mM PP (pH 2.5) 10% MeOH. Applied potential, +20 kV; hydrodynamic injection, 5 s, at different detection wavelengths. Standard solution: PQ, 19.6 μg mL^{-1}; DQ, 37.3 μg mL^{-1}; EQ (ethyl viologen), 38.0 μg mL^{-1}; CQ, 25.2 μg mL^{-1}, MQ, 25.5 μg mL^{-1}, DF 27.6 μg mL^{-1}. * Negative peak (the arrows indicate a data acquisition interval time during which a negative peak is reversed and plotted as a positive peak). (Reproduced from M.T. Galceran, M.C. Carneiro, M. Díez, L. Puignou, *J. Chromatogr. A.*, **782**, 289–295. Copyright (1997) with permission of Elsevier Science B.V.[146])

absorbing quats. Figure 8 shows the electropherograms obtained at the optimal conditions, 10 mM PP (1-(4-pyridyl)pyridinium chloride hydrochloride) (pH 2.5) and 10% methanol, at two different wavelengths, 205 nm and 257 nm. CE/MS have also been proposed for the characterization of these herbicides.[148,149]

CE/UV has been used for analysis of alachlor and metolachlor together with atrazine and simazine. The reported method is simple, rapid and efficient and the compounds can be separated in a relatively short time.[179]

5.4.5 Voltammetry

PQ and DQ have been determined by square wave voltammetry (SWV).[150,151] This method is nonselective because the PQ and DQ species were reduced at the same potential values. However, these compounds present similar toxicities and are often used in mixtures, so that cumulative residue analysis is needed.

An electrochemical study of the chloridazon oxidative process in a silica-modified carbon paste electrode in Britton–Robinson buffer (pH = 11.30) using different voltametric techniques has been carried out.[191] An oxidation wave is obtained at 900 mV and the system is irreversible and controlled mainly by adsorption. No interference of other herbicides has been observed and good results were obtained when this method was applied to the determination of chloridazon is soils.

5.4.6 Other Characterization Methods

The chloroacetanilide herbicides acetochlor, alachlor, butachlor, propachlor and metolachlor and their sulfonic acid metabolites were studied by nuclear magnetic resonance (NMR) spectroscopy and fast atom bombardment mass spectrometry (FABMS).[181,188] Although similar in structure, the sulfonic acid metabolites were successfully characterized.

Fiber optic evanescent wave spectroscopy (FEWS) is a technique which allows in situ mid-infrared (MIR) absorption spectroscopy to be carried out in aqueous environments. The potential of this technique for monitoring alachlor at parts per million levels demonstrated that it can be employed as an alarm level indicator.

6 FOOD AND ENVIRONMENTAL APPLICATIONS

6.1 Macrocyclic Lactones

Zeng et al.[7] studied cytochrome P-450 isoforms responsible for the metabolism of abamectin and ivermectin in the rat. The conclusions presented are based on results from incubation with liver microsomes from induced rats. The enzymatic activity of the cytochrome P-450 had been induced to increase the metabolism of the abamectine because of the limited quantities of metabolites produced by liver microsomes from untreated rats. The study was performed using radiolabeled compounds and metabolite profiles were determined by RPHPLC. One-minute fractions of the column eluate were collected into miniscintillation vials and mixed with a scintillation cocktail. The radioactivity of these samples was determined by scintillation spectrometry.

Information on residue levels, dissipation rates, and translocation resulting from a field treatment with ^{3}H-avermectine B_1 was applied to orange and lemon fruits and lemon leaves. Whole fruit residues of avermectin at harvest are less than 0.001 ppm.[141] Translocation from

the treatment sites is so low as to be undetectable. Because of short persistence, avermectin should cause minimal environmental and food contamination.

Aerobic and anaerobic incubation of microbially active soil with [^{14}C]emamectin benzoate resulted in measurable evolution of $^{14}CO_2$. Chromatographic and mass spectral analysis of extractable residues demonstrated the degradation of the compounds.[9] The degradation in soils follows biphasic kinetics and is relatively rapid between days 0 and 60, becoming less rapid between days 120 and 360. The results of this study demonstrated that emamectin benzoate is biodegradable in soil, forming multiple residual products, and is eventually incorporated into soil components and mineralized to CO_2. Laboratory studies of the fate of ^{3}H-labeled avermectin B_{1a} in Lufkin fine sandy load, Houston clay and coarse sand demonstrated that under aerobic conditions the compound was degraded at a fairly rapid rate to at least 13 radioactive products.[21]

Emamectin benzoate and its δ isomers were determined in a lettuce and celery field trial.[45] The emamectin benzoate level dissipated rapidly on days 0–1 of the postapplication period and then dissipated thereafter at a slower rate. At seven days post-treatment, emamectin benzoate residues were not detected.

Rapid uptake of ^{3}H-emamectin benzoate by bluegill sunfish was observed. The steady-state bioconcentration factors indicate that emamectin benzoate neither bioconcentrates in individual aquatic organisms nor biomagnifies in the food chain.[47]

6.2 Chloronicotinyls

Determination of imidacloprid in water, after hydrolysis in basic medium, followed by GC/MS in SIM mode has been applied to groundwater samples from Santa Maria farm, near Granada (Spain) and in tap water from the city of Granada itself. Imidacloprid was not found above of detection limit ($0.16\,\mu g\,L^{-1}$).[53]

Imidacloprid residues have been extracted from vegetables and determination was performed by HPLC/DAD. Fifty-four vegetable samples (peppers, tomatoes and cucumbers) with incurred residues from three different greenhouse plantations, treated with Confidor™ 20 LS (Bayer; 20% (w/v) of imidacloprid) at different doses and collected during the week after treatment, were selected to evaluate the effectiveness of the proposed method. Residues of imidacloprid were found in all samples in the $0.01-0.3\,mg\,kg^{-1}$ range.[52]

The behavior of insecticide imidacloprid in rice and cucumbers was determined by HPLC/UV. The method was applied to determine the residues and rate of disappearance of imidacloprid from rice (nursery box treatment, 1 g of active ingredient/1800 cm^2) and cucumber plants (stem injection, 20 μg of active ingredient per plant). The insecticide incorporated into the plants decreased rapidly with a half-life of less than 3 days. In the rice, however, it controlled the brown planthopper even 60 days after application, with the marginal concentration of $0.01\,mg\,kg^{-1}$.[4]

6.3 Tetranortriterpenoids

The neem tree, originally from India and Pakistan, has received much attention, thanks to the many valuable insecticide properties of azadirachtin. The compound is not easily separable from the many compounds of similar structure and polarity, of which over 100 have been isolated. The quantitative determination of the natural insecticide azadirachtin in crude extracts of neem seeds can be carried out by packed column SFC, with detection by UV absorption at 210–220 nm using CO_2–methanol as the mobile phase.[12] Almost all the accompanying substances are less polar than azadirachtin and elute quickly.

The HPLC procedures for the quantitative determination of azadirachtin in an insecticidal formulation utilize detection with UV absorption at 215 nm. Formulations from neem kernels may contain additional constituents such as triglycerides, fatty acids and surfactants which can interfere with chromatographic analysis. Separation of the interferences is a prerequisite for analysis.[22,57,58]

The hydrolysis of azadirachtin was studied in several natural waters. Azadirachtin appeared to be more susceptible to hydrolysis than synthetic organophosphates. Accordingly, azadirachtin is expected to be nonpersistent in water.[13]

The stability of azadirachtin was examined in six common organic solvents (chloroform, acetone, ethyl acetate, acetonitrile, methanol and ethanol) to help in the development of formulations.[56]

6.4 Ammonium Quaternary Salts

Both CE and ion-pair RPHPLC have been used for the determination of PQ, DQ and DF. Ion pair RPHPLC offers greater sensitivity for determination of herbicides in crop waters. The results obtained show that three samples contained PQ.[60]

Analysis of DQ, PQ and DF in water has been performed using SPE on silica cartridges followed by NPHPLC determination. To show the applicability of the method in routine analysis of natural waters, three samples were taken from three different places in the Valencian Community (Spain): from irrigation channels of Ribarroja del Turia, El Palmar and Puçol with chemical oxygen demand (COD) values of 1.4, 3.9 and 12.0 mg $O_2\,L^{-1}$, respectively. DQ was found in the El Palmar and Ribarroja del Turia samples at concentrations of

6 and 13 $\mu g\,L^{-1}$, respectively.[63] An automatic sample preparation device (OSP (on-line sample preparatory), Merck) was employed to determine these pesticides. The procedure was applied to monitor the three herbicides in drinking and surface water. DQ, PQ and DF were not detected in drinking water. Twelve surface water samples from the natural park of L'Albufera (València, Spain) were analyzed. The pesticide presence was detected in four samples at concentrations ranging from 0.4 to 0.8 $\mu g\,L^{-1}$.[64]

An on-line SPE coupled with an HPLC procedure has been developed to determine DQ, PQ and DF in water, using GCB for extraction and determination.[66] Four surface waters from Torreblanca Natural Park (Castelló, Spain) were analyzed. Pesticides were found in three of the four samples at levels between 0.1–0.25 $\mu g\,L^{-1}$ after passing 50 mL of water into the extraction cartridges.

Various techniques are used to improve contact of aquatic herbicides with submerged weeds to facilitate control. One commonly used method is to add one of a group of adjuvants, known as polymers to the spray tank mixture. Information on how these adjuvants affect herbicide distribution will be helpful for predicting movement of DQ from treated areas. DQ was applied in a lake to three 1.6 ha plots either with a polymer, which reportedly aids in sinking and confinement of aquatic herbicides, or without a polymer.[62] DQ was analyzed by extraction with C_8 and ion-pair RPHPLC. Forty-six percent of the samples collected at the edges of the plot did not contain detectable DQ residues and only 66% of the sample with detectable DQ contained more than the drinking water tolerance (10 ppb).

A method for measuring CQ based on an in vitro multistep reaction by which CQ is initially N-demethylated and then derivatized was proposed.[82] ECD and MS were used to analyze cotton seed harvested from cotton plants treated with various concentrations of CQ. Concentrations between 0.01 and 1.0 $\mu g\,g^{-1}$ were determined with a limit of detection of 0.003 $\mu g\,g^{-1}$.

The herbicides DQ and PQ can be analyzed in edible oils by HPLC/UV at 210 nm using an ODS column. Recoveries from 90 to 93% can be obtained with a pH 7 aqueous phosphate buffer extraction procedure and cleanup with C_{18} cartridges. Only two of the 12 analyzed virgin oil samples from olive crops sprayed with these herbicides contained DQ residues.[81]

A photokinetic method based on photoreduction with EDTA and acridine yellow using polarography has been applied to DQ in potatoes. The concentration of DQ found in the potatoes ranged from 13.74 to 24.51 $\mu g\,g^{-1}$.[78]

A spectrophotometric method for the determination of PQ using a reducing agent was described[93] using sodium borohydride and alkaline medium to give a blue radical ion with an absorbance maximum at 600 nm. The method was applied to the determination of PQ in human samples, such as blood, urine and mother's milk compared to food and environmental samples.

6.5 Dinitroanilines

Studies were undertaken to evaluate the dissipation of pendimethalin in two Canadian field soils following application of an emulsifiable concentrate formulation in the spring. Residues of pendimethalin were determined by extraction with acetonitrile and GC/ECD. Over a longer period there was no leaching of the herbicide below 10 cm.[101] Residues of pendimethalin were carried over in soils to the next crop year in amounts between 12 and 28% of that applied (1.11 $kg\,ha^{-1}$).

Pendimethalin, alachlor and metolachlor were often found to be present in rainfall in Indiana and Ohio. Analysis was performed by GC with dual-column using NPD.[154] Concentrations were in the range 0.1–1.0 $\mu g\,L^{-1}$. During periods of frequent or prolonged rainfall, concentrations decreased sharply. Concentrations in the next rainfall generally were higher that those at the end of these periods, consistent with a scavenging effect of precipitation on the local atmospheric load of these compounds.

A method based on the extraction of soil samples in small columns has been used for the determination of pendimethalin and other herbicides in soil. Soil samples from experimental fields taken at different times after treatment from the low layer were analyzed and the values were similar to those obtained using a conventional LLE method.[153] Soil from experimental fields treated with pendimethalin was sampled at various times. A concentration of about 3 $\mu g\,g^{-1}$ was found at the beginning of the application. One month later the concentration decreased to 1.2 $\mu g\,g^{-1}$.

6.6 Acetamides

Determination of the ethanesulfonate metabolite of alachlor was performed by RPHPLC/UV.[104] Eleven water samples (ground and surface) collected from Indiana and Ohio were analyzed. Of these 11 samples, 9 were shown by HPLC to contain ethanesulfonate varying in concentration from 0.6 to 74 $ng\,mL^{-1}$.

SPE was combined with HPLC and electrospray ionization mass spectrometry (ESIMS) for the trace analysis of oxanilic and sulfonic acids of acetochlor, alachlor and metolachlor. The isolation procedure separated the chloroacetanilide metabolites from the parent herbicides during the elution from C_{18} cartridges using ethyl acetate for parent compounds followed by methanol for the anionic metabolites.[109] Concentrations in the nanogram per liter range were found in surface and groundwater

samples. Extraction of butachlor with XAD-2 resin and determination with GC/ECD was described.[113] Pesticide residues in river and surface waters can be identified and determined. Residues of butachlor were frequently detected when it was applied to control the weeds in flooded paddies but the results indicate that the herbicide was not persistent in the aquatic environment.

A method for determining trace levels of butachlor, pretilachlor, and other herbicides in river water has been developed. The method is based on C_{18} SPE, followed by GC/MS/SIM.[103] Recoveries at the 0.5 or 2.5 $\mu g\,L^{-1}$ fortification level were between 79 and 98%. The detection limits were 0.01 $\mu g\,L^{-1}$. The analytical findings on surface water samples collected from rivers indicated that most of the pesticides were detected in river water at 1–10 $\mu g\,L^{-1}$ levels during the first 2–4 weeks after application in the fields. A similar method was reported for determining alachlor, butachlor and other herbicides in a dissolved and suspended phase of river water.[112] This method was applied to monitor herbicides in river water from the Shinano River (Japan). Butachlor was detected in some filtered water samples.

A method for the simultaneous identification of neutral and acidic pesticides in natural water[114] includes enrichment of the compounds by SPE on GCB, followed by sequential elution of the neutral and acidic pesticides and derivatization of the latter fraction with diazomethane. Identification and quantification of the compounds was performed by GC/MS. The method has been used to analyze various natural waters including rainwater, roof runoff, surface water and groundwater. Alachlor was frequently detected in rainwater and runoff water during the application period of pesticides. Concentrations were 16 $ng\,L^{-1}$ and 19 $ng\,L^{-1}$, respectively.

Heyer et al.[105] describe SPE with RP-C_{18} adsorbent for the analysis of alachlor and the development of a micro LLE method following GC/ECD and NPD determination of alachlor and its metabolites DEA (2,6-diethylamine) and CDEA (2-chloro-2′,6′-diethylacetanilide). Both methods were applied to the analysis of more than 200 groundwater samples during a field study and gave similar results. The concentration of alachlor was around 0.5 $ng\,L^{-1}$. Degradation products of alachlor were not found in the field study.

Atrazine, propachlor, butachlor, metolachlor, alachlor and two alachlor products were determined in water, also using a micro LLE.[26] The method involves a one-step extraction of water samples with *n*-hexane followed by direct analysis of extracts using GC/NPD or GC/MS. The analytes detected in the five groundwater samples included alachlor (0.1–1.2 $\mu g\,L^{-1}$) and metolachlor (0.1–2.5 $\mu g\,L^{-1}$). The target alachlor degradation products were not detected in these analyses.

The corn herbicide acetochlor was detected and analyzed by GC/MS. The concentrations and fluxes of acetochlor in rain and in the Blue Earth River in Minnesota were reported. Acetochlor was consistently observed and the concentrations in both rainwater and the river were in the 10–250 $ng\,L^{-1}$ range in most samples.[102]

A combination of HPLC/DAD and GC/NPD with UV and MS confirmation, respectively, has been applied to the determination of trace levels of herbicides in relatively clean water samples by LLE. Illustrative examples of the determination of several herbicides in estuary water samples from the Ebro Delta (Tarragona, Spain) are shown. Alachlor and metolachlor were the most commonly found acetanilides, with concentrations levels varying from 5 $ng\,L^{-1}$ to 5 $\mu g\,L^{-1}$.

Analysis of alachlor was performed by immunoassay using EnviroGard™, and 67 well samples were analyzed. Alachlor was contained in many samples that are positive by immunoassay. In fact, of the 18 positive well samples by immunoassay, just six were found to have alachlor by HPLC.[166]

SPE and ELISA were combined for the trace analysis of the herbicide alachlor and its major soil metabolite ethane sulfonic acid. These compounds were isolated from water by SPE on C_{18} and eluted sequentially with ethyl acetate and methanol. Analyses of surface and groundwater samples were confirmed by GC/MS and HPLC/DAD. Results showed widespread occurrence of ethane sulfonic acid in the midwestern USA, with concentrations ranging from <0.10 to $>10\,\mu g\,L^{-1}$.[171]

The ELISA method was evaluated for the detection of alachlor in environmental water samples and was compared with a GC/MS method. In analyses of environmental water samples ELISA was less accurate and precise than GC/MS.[162]

Unexpectedly high frequencies of positive responses were observed in ELISA screening for alachlor in private rural wells, when samples were analyzed using the EnviroGard™ Alachlor Plate Kit. The majority of the positive responses were false positives with respect to alachlor. In the false positive samples, most appear to be due to the ethanesulfonate metabolite of alachlor. Confirmation testing for alachlor used SPE and analysis by GC/MS. A method using SPE and HPLC/UV was developed for the analysis of the metabolite.[170,175]

An immunoassay for detecting alachlor or putative alachlor metabolites in human urine has been developed and compared with measurements obtained by HPLC analysis of urine hydrolyzate for DEA.[143] Twenty applicators and seven hauler/mixers participated in the study. Also, eight employees of the application companies who were thought to have limited exposure to pesticides submitted urine samples for alachlor dose estimation. The ELISA and HPLC analytical methods gave different

results because the metabolism and excretion of alachlor is not well defined.

Human alachlor exposure has commonly been estimated by quantifying the hydrolysis product DEA, of alachlor metabolites in urine.[167] The alachlor metabolite was identified by using HPLC/MS/MS to analyze extracts of urine samples from subjects who were occupationally exposed to alachlor.

Several herbicides widely used to control weeds in corn (atrazine, alachlor, metolachlor and pendimethalin), have been determined in soil and water.[120] Analysis of herbicides was performed by GC/NPD and GC/MS. Soil samples from corn fields after harvest were taken in several fields located in Spain and water was taken from wells located in the fields. Herbicides were not detectable in water samples from one well, but in another well alachlor was detected at levels higher than 0.1 ppb. In soil, residues varied in the different fields and were in some cases higher than $0.1\,\mu g\,g^{-1}$.

An ethanesulfonic acid metabolite of metolachlor was identified in soil sample extracts by FABMS and FABMS/MS. A dissipation study revealed that ethanesulfonic acid is formed in soil under field conditions corresponding to a decrease in the concentration of the metolachlor.[188]

6.7 Oximes

Several analytical methods have been described for determination of alloxydim sodium, sethoxydim and its major active metabolites and clethoxidim.[5,124,125] Results obtained after application to residue determination in real samples are not reported.

6.8 Triazoles

An HPLC method with amperometric detection was described for determining residue levels of amitrole in drinking and groundwater. Samples with concentrations up to 1 ppb can be injected directly. When a simple evaporation enrichment step is used, the limit of determination is 0.1 ppb. If the method is applied to groundwater samples, no interference from co-eluting constituents appears in the chromatograms.[129] Another method using a coulometric array detector was also applied to tap and well water samples. Concentrations of amitrole down to $0.1\,\mu g\,L^{-1}$ can be determined without any enrichment steps. Several well water samples were investigated. Amitrole was not detectable.[131]

Amitrole can be extracted with ethanol, adsorbed on resin and desorbed with ammonia. After acetylation with acetic anhydride and cleanup over a Sep-Pack™ silica cartridge, the amitrole-acetyl derivative was determined by GC with FPD (flame photometric detection) in plant tissues and sandy soils. The method was used for screening the abuse of amitrole on vegetation growing on sandy soils. Over a period of more than one year a wide variety of weeds was analyzed.[133]

6.9 Pyridine-based Molecules

Thiazopyr residues have been determined in soil, grass weed and seedlings of woody plants after extraction with ethyl acetate, cleanup on a Florisil™ column and determination by GC/NPD and confirmation by GC/MS with ion trap detection. The proposed method was applied to the determination of the herbicide level in treated soil and plant samples.[135] The concentration range was $0.2-1\,ng\,g^{-1}$.

7 CONCLUSIONS

Miscellaneous pesticides are mainly polar compounds. For this reason, HPLC is the preferred technique for their separation and determination. However, complicated derivatization schemes are still used to enhance sensitivity and selectivity. These derivatization reactions present several problems when environmental analyses are performed or many samples are processed because they are time-consuming and give rise to interferences caused by the reagents and endogenous compounds of the matrices.

Best extraction methods should be used in order to obtain better separation of the endogenous interference compounds from the matrix and better sensitivity. Extraction methods using SPE are rarely employed in the analysis of miscellaneous pesticides. The use of these methods would make it possible to incorporate on-line extraction procedures, and to include these pesticides in the general multiresidue systems employed to determine a wide range of pesticide varieties from several matrices.

The development of modern techniques will facilitate application to real samples, monitor the presence of the pesticides in the environment and establish their persistence, fate, degradation and risk on the basis of real field data.

ACKNOWLEDGMENTS

The authors thank to the Generalitat Valenciana (GV-CAPA-97-10) for financial support of this study.

ABBREVIATIONS AND ACRONYMS

ADI	Admissible Daily Intake
BSCD	*ter*-Butyldimethylsilyl-β-cyclodextrin
CD	Conductivity Detection
CDEA	2-Chloro-2′,6′-diethylacetanilide

CE	Capillary Electrophoresis
CL9673	3-Phenyl-4-hydroxy-6-chloropyridazine
CM-l	Methyl 5-(l-aminobutylidene)-2, 2-dimethyl-4,6-dioxocyclo-hexanecarboxylate
COD	Chemical Oxygen Demand
CQ	Chlormequat
DAD	Diode Array Detection
DC	Direct Current
DEA	2,6-Diethylamine
DF	Difenzoquat
DQ	Diquat
EAD	Electrode Array Detection
ECD	Electron-capture Detection
ED	Electrochemical Detection
EDTA	Ethylenediaminetetraacetic Acid
EI	Electron Impact
ELISA	Enzyme-linked Immunosorbent Assay
EQ	Ethyl Viologen
ESIMS	Electrospray Ionization Mass Spectrometry
ESP	Electrospray
EU	European Union
FABMS	Fast Atom Bombardment Mass Spectrometry
FEWS	Fiber Optic Evanescent Wave Spectroscopy
FIA	Flow-injection Analysis
FID	Flame Ionization Detection
FPD	Flame Photometric Detection
GABA	γ-Aminobutyric Acid
GC	Gas Chromatography
GCB	Graphitized Carbon Black
HPLC	High-performance Liquid Chromatography
HRGC	High-resolution Gas Chromatography
IARC	International Association for Research on Cancer
LC	Liquid Chromatography
LD_{50}	Lethal Dose 50
LLE	Liquid–Liquid Extraction
LSC	Liquid Scintillation Counter
MCPA	4-Chloro-2-methyl(phenoxy)acetic Acid
MIR	Mid-infrared
MQ	Mepiquat
MRL	Maximum Residue Limit
MS	Mass Spectrometry
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
NMR	Nuclear Magnetic Resonance
NOEL	No Observable Effect Level
NPD	Nitrogen Phosphorus Detector
NPHPLC	Normal-phase High-performance Liquid Chromatography
ODS	Octadecylsilica
OSP	On-line Sample Preparatory
PAD	Pulse Amperometric Detection
PB	Particle Beam
PDPs	Photodegradation Products
PP	1-(4-Pyridyl)pyridinium Chloride Hydrochloride
PQ	Paraquat
RP	Reversed-phase
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SDS	Sodium Dodecyl Sulfate
SFC	Supercritical Fluid Chromatography
SFE	Supercritical-fluid Extraction
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SWV	Square Wave Voltammetry
TLC	Thin-layer Chromatography
TSP	Thermospray
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
2,4-DIOE	Isooctyl Ester of (2,4-Dichloro-phenoxy)acetic Acid
2,4-DPBEE	Butoxyethanol Ester of 2-(2,4-dichloro-phenoxy)propionic Acid

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food

REFERENCES

1. C. De Liñán y Vicente, *Vademecum de Productos Fitosanitarios y Nutricionales*, Agrotécnicas, Madrid, 1997.

2. C. De Liñán y Vicente, *Farmacologta Vegetal. Compedio de las Materias Activas más Interesantes en Fitopatologta. Su Toxicologta y Empleo*, Agrotécnicas, Madrid, 1997.
3. B.A. Halley, J.A. VandenHeuvel, P.G. Wislocki, 'Environmental Effects of the Usage of Avermectins in Livestock', *Vet. Parasitol.*, **48**, 109–125 (1993).
4. Y. Ishii, I. Kobori, Y. Araki, S. Kurogochi, I. Koji, S. Kagabu, 'HPLC Determination of the New Insecticide Imidacloprid and its Behavior in Rice and Cucumber', *J. Agric. Food Chem.*, **42**, 2917–2921 (1994).
5. L.N. Falb, D.C. Bridges, A.E. Smith, 'Separation of Clethodim Herbicide from Acid and Photodegradation Products by Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **74**(6), 999–1002 (1991).
6. J. Thomas, 'Relationships Between Retention Behavior and Structure of N-substituted 1,2,4-triazoles in Thin Layer Chromatography on Silica Gel', *J. Chromatogr. A*, **404**, 73–85 (1987).
7. Z. Zeng, N.W. Andrew, J.M. Woda, B.A. Halley, L.S. Crouch, R.W. Wang, 'Role of Cytochrome P450 Isoforms in the Metabolism of Abamectin and Ivermectin in Rats', *J. Agric. Food Chem.*, **44**, 3374–3378 (1996).
8. J.A. Cobin, N.A. Johnson, 'Liquid Chromatographic Method for Determination of Total Avermectin B_1 and 8,9-*Z*-avermectin B_1 Residues in Hops', *J. AOAC Int.*, **79**(2), 503–507 (1996).
9. A.C. Chukwudebe, R.H. Atkins, P.G. Wislocki, 'Metabolic Fate of Emamectin Benzoate in Soil', *J. Agric. Food Chem.*, **45**, 4137–4146 (1997).
10. M.B. Hicks, L.D. Payne, S.V. Prabhu, T.A. Wehner, 'Determination of Emamectin Benzoate in Freshwater and Seawater at Picogram-per Millilitre Levels by Liquid Chromatography with Fluorescence Detection', *J. AOAC Int.*, **80**(5), 1098–1103 (1997).
11. N. Ruíz de Erenchun, Z. Gómez de Balugera, M.A. Goicolea, R.J. Barrio, 'Determination of Imidacloprid and its Major Metabolite in Soils by Liquid Chromatography with Pulsed Reductive Amperometric Detection', *Anal. Chim. Acta*, **349**, 199–206 (1997).
12. H.P. Huang, D. Morgan, 'Analysis of Azadirachtin by Supercritical-fluid Chromatography', *J. Chromatogr.*, **519**, 137–143 (1990).
13. S.Y. Szeto, M.T. Wan, 'Hydrolysis of Azadirachtin in Buffered and Natural Waters', *J. Agric. Food Chem.*, **44**(4), 1160–1163 (1996).
14. M.T. Galceran, M.C. Carneiro, L. Puignou, 'Capillary Electrophoresis of Quaternary Ammonium Ion Herbicides: Paraquat, Diquat and Difenzoquat', *Chromatographia*, **39**(9/10), 581–586 (1994).
15. D. Barceló, G. Durand, R.J. Vreeken, 'Determination of Quaternary Amine Pesticides by Thermospray Mass Spectrometry', *J. Chromatogr.*, **647**, 271–277 (1993).
16. B.L. Worobey, 'Liquid Chromatographic Method for Determination of Diquat and Paraquat Herbicides in Potatoes: Collaborative Study', *J. AOAC Int.*, **76**(4), 881–887 (1993).
17. O. Huschens, 'Bestimmung Pestizidwirksamer Quaternärer Ammonium-verbindungern in Wäßriger Phase, Beispielhaft Ausgeführt für die Wirkstoffe Paraquat, Diquat un Difenzoquat Mittels HPLC', *Wasser Abwasser*, **134**(10), 620–622 (1993).
18. P. Cabras, M. Melis, L. Spanedda, C. Tuberoso, 'High-performance Liquid Chromatographic Determination of Dinitroaniline Herbicides in Soil and Water', *J. Chromatogr. A*, **585**, 164–167 (1991).
19. A. Meyer, G. Henze, 'HPLC with Amperometric Detection for the Determination of Pesticides of High Polarity', *Fresenius' J. Anal. Chem.*, **349**, 650–653 (1994).
20. M.A. Alawi, 'Determination of Pyridate and its Main Metabolite CL-9673 in Plant Materials', *Chromatographia*, **22**, 1–6 (1986).
21. D.L. Bull, G.W. Ivie, J.G. MacConnell, V.F. Gruber, C.C. Ku, B.H. Arison, J.M. Stevenson, J.A. VandenHeuvel, 'Fate of Avermectin B_1 in Soil and Plants', *J. Agric. Food Chem.*, **32**, 94–102 (1984).
22. C.J. Hull, W.R. Dutton, B.S. Switzer, 'Quantitation of Azadirachtins in Insecticidal Formulations by High-performance Liquid Chromatography', *J. Chromatogr.*, **633**, 300–304 (1993).
23. T.C. Marrs, 'Toxicology of Pesticides', in *General and Applied Toxicology*, eds. B. Ballantyne, T. Marrs, P. Turner, M Stockton Press, Wimbledon, 1329–1341, 1994.
24. D.S. Saunders, C. Harper, 'Pesticides', in *Principles and Methods of Toxicology*, ed. A.W. Hayes, Raven Press, New York, 389–415, 1994.
25. C.L. Winek, W.W. Wahba, J.M. Edelstein, 'Sudden Death following Accidental Ingestion of Chlormequat', *J. Anal. Toxicol.*, **14**, 257–258 (1990).
26. T.L. Potter, T. Carpenter, R. Putnam, K. Reddy, J.M. Clark, 'Rapid Method for Analysis of Atrazine and Acetanilide Herbicides in Groundwater by Micro Liquid/Liquid Extraction', *J. Agric. Food Chem.*, **39**, 2184–2187 (1991).
27. A.W. Archer, 'Determination of 3-amino-1,2,4-triazole (Amitrole) in Urine by Ion-pair High-performance Liquid Chromatography', *J. Chromatogr. A*, **303**, 267–271 (1984).
28. 'Discussion and Principal Supporting Studies of Sethoxydim', 152669, BASF Corporation Chemical Division, Washington, DC 20460, 1997.
29. 'Quality of Water for Human Consumption', EC, 80/778/EEC, Brussels, 1980.
30. 'National Pesticide Survey Phase I', EPA, PB91-125765, Springfield, VA, 1991.
31. 'National Survey of Pesticides in Drinking Water Wells Phase II', EPA, EPA 570/9-91-020, Springfield, VA, 1992.

32. D. Barceló, 'Official Methods of Analysis of Priority Pesticides in Water Using Gas Chromatographic Techniques', in *Environmental Analysis. Techniques, Applications and Quality Assurance*, ed. D. Barceló, Elsevier, Amsterdam, 149–180, 1993.
33. 'Council Directive Relating to the Fixing of Maximum Level of Pesticide Residues in and on Fruit and Vegetables', EC, 76/895/EEC, Brussels, 1976.
34. 'Council Directive Relating to the Fixing of Maximum Levels of Pesticide Residues in and on Certain Products of Plant Origin, Including Fruit and Vegetables, and Providing for the Establishment of a List of Maximum Levels', EC, 90/642/EEC, Brussels, 1990.
35. 'Council Directive Amending Annex II to Directive 76/895/EEC and Annex II to Directive 90/642/EEC', EC, 96/32/EC, Brussels, 1996.
36. 'Council Directive Amending Directives 76/895/EEC, 86/362/EEC, 86/363/EEC and 90/642 EEC Relating to the Fixing of Maximum Levels for Pesticide Residues in and on, Respectively, Fruit and Vegetables, Cereals, Foodstuffs of Animal Origin, and Certain Products of Plant Origin, Including Fruit and Vegetables', EC, Brussels, 1998.
37. K.A. Krieger, L.A. Rosenheck, L.L. Schuester, 'Adult and Infant Abamectin Exposure following Avert 310 and Pressurized Gel Crack and Crevice Treatment', *Bull. Environ. Contam. Toxicol.*, **58**, 681–687 (1997).
38. C.G. Wright, R.B. Leidy, H.E. Dupree, 'Abamectin in the Ambient Air, on Surfaces and in Food of Dining Facilities Treated for Cockroaches', *Bull. Environ. Contam. Toxicol.*, **49**, 171–178 (1992).
39. S.V. Prabhu, R.J. Varsolona, T.A. Wehner, R.S. Egan, P.C. Tway, 'Rapid and Sensitive High-performance Liquid Chromatographic Method for the Quantitation of Abamectine and its Delta 8,9-isomer', *J. Agric. Food Chem.*, **40**, 622–625 (1992).
40. N. Chamkasem, M.L. Papathakis, S.M. Lee, 'Liquid Chromatographic Determination of Abamectin in Fruits and Vegetables', *J. AOAC Int.*, **76**(3), 691–694 (1993).
41. J.A. Cobin, N.A. Johnson, 'Liquid Chromatographic Method for Rapid Determination of Total Avermectin B_1 and 8,9-*Z*-Avermectin B_1 Residues in Apples', *J. AOAC Int.*, **78**(2), 419–422 (1995).
42. J.A. Cobin, N.A. Johnson, 'Determination of Total Avermectin B_1 and 8,9-*Z*-avermectin B_1 Residues in Wine by Liquid Chromatography', *J. AOAC Int.*, **79**(5), 1158–1161 (1996).
43. J.A. Cobin, N.A. Johnson, 'Liquid Chromatographic Method for Determination of Total Avermectin B_1 and 8,9-*Z*-Avermectin B_1 Residues in Hops', *J. AOAC Int.*, **79**, 503–507 (1996).
44. J. Vuik, 'Rapid Determination of Abamectin in Lettuce and Cucumber Using High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **39**, 303–305 (1991).
45. S.V. Prabhu, T.A. Wehner, R.S. Egan, P.C. Tway, 'Determination of 4″-deoxy-4″-(epimethylamino) Avermectin B_1 Benzoate (MK-0244) and its Delta 8,9-isomer in Celery and Lettuce by HPLC with Fluorescence Detection', *J. Agric. Food Chem.*, **39**, 2226–2230 (1991).
46. M.W. Brooks, P.C. Uden, 'The Determination of Abamectin from Soil and Animal Tissue by Supercritical Fluid Extraction and Fluorescence Detection', *Pesticide Sci.*, **43**, 141–146 (1995).
47. A.C. Chukwudebe, N. Andrew, K. Drottar, J. Swigert, P.G. Wislocki, 'Bioaccumulation Potential of 4″-*epi*-(methylamino)-4″-deoxyavermectin B_1 a Benzoate (Emamectin Benzoate) in Bluegill Sunfish', *J. Agric. Food Chem.*, **44**, 2894–2899 (1996).
48. L. Junsuo, C. Qian, 'Determination of Avermectin B_1 in Biological Samples by Immunoaffinity Column Cleanup and Liquid Chromatography with UV Detection', *J. AOAC Int.*, **79**(5), 1062–1067 (1996).
49. L.D. Payne, M.B. Hicks, T.A. Wehner, 'Determination of Abamectin and/or Ivermectin in Cattle Feces at Low Parts per Billion Levels Using HPLC with Fluorescence Detection', *J. Agric. Food Chem.*, **43**, 1233–1237 (1995).
50. M.M. Macke, 'Quantitation of Imidacloprid in Liquid and Soil Formulations by Reversed-phase Liquid Chromatography: Collaborative Study', *J. AOAC Int.*, **81**(2), 344–348 (1998).
51. S. Baskaran, R.S. Kookana, R. Naidu, 'Determination of the Insecticide Imidacloprid in Water and Soil Using High-performance Liquid Chromatography', *J. Chromatogr. A*, **787**, 271–275 (1997).
52. A.R. Fernández-Alba, A. Valverde, A. Agüera, M. Contreras, S. Chiron, 'Determination of Imidacloprid in Vegetables by High-performance Liquid Chromatography with Diode-array Detection', *J. Chromatogr. A*, **721**, 97–105 (1996).
53. J.L. Vílchez, R.E. Khattabi, J. Fernández, A. González-Casado, A. Navalón, 'Determination of Imidacloprid in Water and Soil Samples by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **746**, 289–294 (1996).
54. M. Tokieda, K. Iiyoshi, K. Sugioka, T. Gomyo, 'Residue Determination Method for the Insecticide Acetamiprid in Crops by Gas Chromatography', *J. Pesticide Sci.*, **22**, 129–132 (1997).
55. M. Tokieda, M. Ozawa, S. Kobayashi, T. Gomyo, 'Method to Determination of Total Residue of the Insecticide Acetamiprid and its Metabolites in Crops by Gas Chromatography', *J. Pesticide Sci.*, **22**, 77–83 (1997).
56. R. Thejavathi, R.S. Yakkundi, B. Ravindranath, 'Determination of Azadirachtin by Reversed-phase High-performance Liquid Chromatography Using Anisole as Internal Standard', *J. Chromatogr. A*, **705**, 374–379 (1995).
57. M.E. Azam, S. Rengasamy, B.S. Parmar, 'Estimation of Azadirachtin-A Content of Emulsifiable and Solution Concentrates of Neem', *J. AOAC Int.*, **78**(4), 893–896 (1995).

58. S. Johnson, E.D. Morgan, 'Comparison of Chromatographic Systems for Triterpenoids from Neem (*Azadirachta Indica*) Seeds', *J. Chromatogr. A*, **761**, 53–63 (1997).
59. T. Pérez Ruíz, C. Martínez Lozano, V. Tomás, R. Casajus, 'Kinetic–Spectrophotometric Determination of Diquat Based on a Charge-transfer Complex with Cysteine', *Analyst*, **119**(7), 1557–1560 (1994).
60. M.C. Carneiro, L. Puignou, M.T. Galceran, 'Comparison of Capillary Electrophoresis and Reversed-phase Ion-pair High-performance Liquid Chromatography for the Determination of Paraquat, Diquat and Difenzoquat', *J. Chromatogr. A*, **669**, 217–224 (1994).
61. D. Volmer, K. Levsen, G. Wünsch, 'Thermospray Liquid Chromatographic–Mass Spectrometric Multi-residue Determination of 128 Polar Pesticides in Aqueous Environmental Samples', *J. Chromatogr. A*, **660**, 231–248 (1994).
62. K.A. Langeland, A.M. Fox, F.B. Laroche, B.B. Martín, D.F. Martín, C.D. Norris, C. Wang, 'Diquat Distribution in Water after Application to Submerged Weeds', *Water Resources Bull.*, **30**(1), 93–97 (1994).
63. M. Ibáñez, Y. Picó, J. Mañes, 'Influence of Organic Matter and Surfactants on Solid-phase Extraction of Diquat, Paraquat and Difenzoquat from Waters', *J. Chromatogr. A*, **727**, 245–252 (1996).
64. M. Ibáñez, Y. Picó, J. Mañes, 'On-line Liquid Chromatographic Trace Enrichment and High Performance Liquid Chromatographic Determination of Diquat, Paraquat and Difenzoquat in Water', *J. Chromatogr. A*, **728**, 325–331 (1997).
65. P. Shivhare, V.K. Gupta, 'Spectrophotometric Method for the Determination of Paraquat in Water, Grain and Plants Materials', *Analyst*, **116**, 391–393 (1991).
66. M. Ibáñez, Y. Picó, J. Mañes, 'On-line Determination of Bipyridylium Herbicides in Water by HPLC', *Chromatographia*, **45**, 402–407 (1997).
67. I. Kambhampati, K.S. Roinestad, T.G. Hartman, J.D. Rosen, E.K. Fukuda, R.L. Lippincott, R.T. Rosen, 'Determination of Diquat and Paraquat in Water Using High-performance Liquid Chromatography with Confirmation by Liquid Chromatography–Particle Beam Mass Spectrometry', *J. Chromatogr. A*, **688**, 67–73 (1994).
68. 'Methods for Determination of Organic Compounds in Drinking Water, No 549', US Environmental Protection Agency, Springfield, VA, 1990.
69. D. Kaniansky, F. Iványi, F.I. Onuska, 'On-line Isotachophoretic Sample Pretreatment in Ultratrace Determination of Paraquat and Diquat in Water by Capillary Zone Electrophoresis', *Anal. Chem.*, **66**, 1817–1824 (1994).
70. A.J.J. Debets, K.P. Hupe, W.T. Kok, U.A.Th. Brinkman, 'Electrodialytic Sample Treatment Coupled On-line with Column Liquid Chromatography for the Determination of Basic and Acidic Compounds in Environmental Samples', *J. Chromatogr.*, **600**, 163–173 (1992).
71. T. Pérez Ruíz, C. Martínez Lozano, V. Tomás, 'Spectrofluorimetric Determination of Diquat by Manual and Flow-injection Methods', *Anal. Chim. Acta*, **244**, 99–104 (1991).
72. A. Jain, K.K. Verma, A. Townshend, 'Determination of Paraquat by Flow-injection Spectrophotometry', *Anal. Chim. Acta*, **284**, 275–279 (1993).
73. M. Agudo, A. Ríos, M. Valcárcel, 'Automatic Continuous-flow Determination of Paraquat at the Subnanogram per millilitre Level', *Anal. Chim. Acta*, **281**, 103–109 (1993).
74. M.Y. Selisker, D.P. Herzog, R.D. Erber, J.R. Flecker, J.A. Itak, 'Determination of Paraquat in Fruits and Vegetables by a Magnetic Particle Based Enzyme-linked Immunosorbent Assay', *J. Agric. Food Chem.*, **43**(2), 544–547 (1995).
75. Y.Y. Wigfield, K.A. McCormack, R. Grant, 'Simultaneous Determination of Residues of Paraquat and Diquat in Potatoes Using High-performance Capillary Electrophoresis with Ultraviolet Detection', *J. Agric. Food Chem.*, **41**, 2315–2318 (1993).
76. T.M.P. Chichila, D.M. Gilvydis, 'Determination of Paraquat and Diquat in Low-moisture Food Crops Using Silica Column Cleanup and Liquid Chromatography with UV Detection', *J. AOAC Int.*, **76**(6), 1323–1328 (1993).
77. T.M. Chichila, S.M. Walters, 'Liquid Chromatographic Determination of Paraquat and Diquat in Crops Using a Silica Column with Aqueous Ionic Mobile Phase', *J. Assoc. Off. Anal. Chem.*, **74**(6), 961–967 (1991).
78. T. Pérez Ruíz, C. Martínez Lozano, V. Tomás, E. Yagüe, 'Determination of Diquat by a Photokinetic Method', *Analyst*, **115**, 783–786 (1990).
79. J. Hajslová, P. Cuhra, T. Davídek, J. Davídek, 'Gas Chromatographic Determination of Diquat and Paraquat in Crops', *J. Chromatogr.*, **479**, 243–250 (1989).
80. A. Fegert, U. Schepers, B. Schwarz, 'Determination of Mepiquatchloride in Animal and Plant Matrices by Ion Chromatography with Conductivity Detection', *Fresenius' J. Anal. Chem.*, **339**, 441–443 (1991).
81. Y. Picó, J.C. Moltó, G. Font, J. Mañes, 'Determinación de Resíduos de Dicuat, Paracuat y Glifosato en Aceites', *Rev. Toxicol.*, **7**(3), 293–299 (1990).
82. W.J. Allender, 'Determination of Chlormequat Residues in Cotton Seed by Gas Chromatography', *Pesticide Sci.*, **35**, 265–269 (1992).
83. K. Ameno, C. Fuke, S. Ameno, H. Kinoshita, I. Ijiri, 'Simultaneous Quantitation of Diquat in its Two Metabolites in Serum and Urine by Ion Paired-HPLC', *J. Liq. Chromatogr.*, **18**(11), 2115–2121 (1995).
84. S. Kawase, S. Kanno, S. Ukai, 'Determination of the Herbicide Paraquat and Diquat in Blood and Urine by Gas Chromatography', *J. Chromatogr.*, **283**, 231–240 (1984).
85. C. Fuke, K. Ameno, T. Kiriu, T. Shinohara, K. Sogo, I. Ijiri, 'A Rapid, Simultaneous Determination of

Paraquat and Diquat in Serum and Urine Using Second-derivative Spectroscopy', *J. Anal. Toxicol.*, **16**(4), 214–216 (1992).
86. T. Lukaszewski, 'The Extraction and Analysis of Quaternary Ammonium Compounds in Biological Material by GC and GC/MS', *J. Anal. Toxicol.*, **9**, 101–108 (1985).
87. I. Sánchez Serello, M. López Rivadulla, A. Cruz, A. Bermejo, P. Fernández, 'A Sequential Spectrophotometric Method for the Determination of Paraquat and Diquat in Plasma', *Anal. Lett.*, **26**(9), 1891–1904 (1993).
88. K. Minakata, O. Suzuki, M. Asano, 'Extraction of Diquat with 1-Butanol from Biological Materials', *Forensic Sci. Int.*, **44**, 27–35 (1990).
89. I. Sánchez Serello, A. Cruz, P. Fernández, M. López Rivadulla, 'Liquid Chromatographic Method for Simultaneous Determination of Paraquat and Diquat in Plasma, Urine and Vitreous Humor', *J. Liq. Chromatogr. Relat. Technol.*, **19**(12), 2009–2024 (1996).
90. M. Akerblom, 'Rapid Determination of Paraquat in Urine with Ion-pair Extraction and Spectrophotometry', *Bull. Environ. Contam. Toxicol.*, **45**, 157–164 (1990).
91. M. Tomita, T. Okuyama, Y. Nigo, 'Simultaneous Determination of Paraquat and Diquat in Serum Using Capillary Electrophoresis', *Biomed. Chromatogr.*, **6**, 91–94 (1992).
92. N.B. Smith, S. Mathialagan, K.E. Brooks, 'Simple Sensitive Solid-phase Extraction of Paraquat from Plasma Using Cyanopropyl Columns', *J. Anal. Toxicol.*, **17**, 143–145 (1992).
93. M.K. Rai, J.V. Das, V.K. Gupta, 'A Sensitive Determination of Paraquat by Spectrometry', *Talanta*, **45**, 347–348 (1997).
94. C. Madhu, Z. Gregus, C.D. Klaassen, 'Simple Method for Analysis of Diquat in Biological Fluids and Tissues by High-performance Liquid Chromatography', *J. Chromatogr. B*, **674**, 193–196 (1995).
95. M.T. Corasaniti, M.C. Strongoli, G. Nisticò, 'Determination of Paraquat in Rat Brain Using Ion-pair Solid-phase Extraction and Reversed-phase High-performance Liquid Chromatography with Ultraviolet Detection', *J. Chromatogr. Biomed. Appl.*, **527**, 189–195 (1990).
96. S. Ito, T. Nagata, K. Kudo, K. Kimura, T. Imamura, 'Simultaneous Determination of Paraquat and Diquat in Human Tissues by High-performance Liquid Chromatography', *J. Chromatogr.*, **617**, 119–123 (1993).
97. T.M. Fitzpatrick, C.C. Gehrig, 'Simultaneous Determination of Benefin and Trifluralin in Pesticide Formulations by Gas Chromatography', *J. Assoc. Off. Anal. Chem.*, **73**(6), 981–982 (1990).
98. S.D. West, J.H. Weston, E.W.J. Day, 'Gas Chromatographic Determination of Residue Levels of the Herbicides Trifluralin, Benefin, Ethalfluralin, and Isopropalin in Soil with Confirmation by Mass Selective Detection', *J. Agric. Food Chem.*, **71**(6), 1082–1085 (1988).
99. A.I. García-Valcárcel, C. Sánchez-Brunete, L. Martínez, J.L. Tadeo, 'Determination of Dinitroaniline Herbicides in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **719**, 113–119 (1996).
100. T.D. Macy, A. Loh, 'Determination of Oryzalin and Prosulfalin Residues in Soil by High Pressure Liquid Chromatography', *Anal. Chem.*, **52**, 1381–1383 (1980).
101. A.E. Smith, A.J. Aubin, T.C. McIntosh, 'Field Persistence Studies with Emulsifiable Concentrate and Granular Formulation of Herbicide Pendimethalin in Saskatchewan', *J. Agric. Food Chem.*, **43**(11), 2988–2991 (1995).
102. P.D. Capel, L. Ma, B.R. Schroyer, S.J. Larson, T.A. Gilchrist, 'Analysis and Detection of the New Corn Herbicide Acetochlor in River Water and Rain', *Environ. Sci. Technol.*, **29**(6), 1702–1705 (1995).
103. H. Kobayashi, K. Ohyama, N. Tomiyama, Y. Jimbo, O. Matamo, S. Goto, 'Determination of Pesticides in River Water by Gas Chromatography–Mass Spectrometry-selected-ion Monitoring', *J. Chromatogr. A*, **643**, 197–202 (1993).
104. C. Macomber, R.J. Bushway, L.B. Perkins, D. Baker, T.S. Fan, B.S. Ferguson, 'Determination of the Ethanesulfonate Metabolite of Alachlor in Water by High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **40**(8), 1450–1452 (1992).
105. R. Heyer, A. Zapf, H.J. Stan, 'Determination of Alachlor and Two Metabolites in Groundwater Using Solid Phase and a New Micro-liquid–Liquid Extraction Method', *Fresenius' J. Anal. Chem.*, **351**, 752–757 (1995).
106. R.G. Nash, 'Solid-phase Extraction of Carbofuran, Atrazine, Simazine, Alachlor, and Cyanazine from Shallow Well Water', *J. Assoc. Off. Anal. Chem.*, **73**(3), 438–442 (1990).
107. S.A. Schuette, R.G. Smith, L.R. Holden, J.A. Graham, 'Solid-phase Extraction of Herbicides from Well Water for Determination by Gas Chromatography–Mass Spectrometry', *Anal. Chim. Acta*, **236**, 141–144 (1990).
108. E.M. Thurman, M. Meyer, M. Pomes, C.A. Perry, A.P. Schwab, 'Enzyme-linked Immunosorbent Assay Compared with Gas Chromatography/Mass Spectrometry for the Determination of Triazine Herbicides in Water', *Anal. Chem.*, **62**, 2043–2048 (1990).
109. I. Ferrer, E.M. Thurman, D. Barceló, 'Identification of Ionic Chloroacetanilide–Herbicide Metabolites in Surface Water and Groundwater by HPLC/MS Using Ion Negative Ion Spray', *Anal. Chem.*, **69**, 4547–4553 (1997).
110. G.A. Peñuela, D. Barceló, 'Comparative Degradation Kinetics of Alachlor in Water by Photocatalysis with $FeCl_3$, TiO_3 and Photolysis, Studied by Solid-phase Disk Extraction followed by Gas Chromatographic Techniques', *J. Chromatogr. A*, **754**, 187–195 (1996).
111. L.Q. Huang, 'Simultaneous Determination of Alachlor, Metolachlor, Atrazine, and Simazine in Water and Soil by Isotope Dilution Gas Chromatography/Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **72**(2), 349–354 (1989).

112. A. Tanabe, H. Mitobe, K. Kawata, M. Sakai, 'Monitoring of Herbicides in River Water by Gas Chromatography–Mass spectrometry and Solid Phase Extraction', *J. Chromatogr. A*, **754**, 159–168 (1996).
113. = M. Ishibashi, M. Suzuki, 'Simultaneous XAD-2 Resin Extraction and High-resolution Electron-capture Gas Chromatography of Chlorine-containing Herbicides in Water Samples', *J. Chromatogr.*, **456**, 382–391 (1988).
114. T.D. Bucheli, F.C. Grüebler, S.R. Müller, R.P. Schwarzenbach, 'Simultaneous Determination of Neutral and Acidic Pesticides in Natural Waters at the Low Nanogram per Liter Level', *Anal. Chem.*, **69**, 1569–1576 (1997).
115. W. Schüssler, 'Automatic Measurement of Some Common Herbicides by LC Using Simultaneous UV Detection and Electrochemical Detection in Series', *Chromatographia*, **27**(9/10), 431–435 (1989).
116. G. Durand, V. Bouvot, D. Barceló, 'Determination of Trace Levels of Herbicides in Estuarine Water by Gas and Liquid Chromatographic Techniques', *J. Chromatogr.*, **607**, 319–327 (1992).
117. W.E. Pereira, C.E. Rostad, T.J. Leiker, 'Determination of Trace Levels of Herbicides and Degradation Products in Surface and Ground Waters by Gas Chromatography/Ion-trap Mass Spectrometry', *Anal. Chim. Acta*, **228**, 69–75 (1990).
118. D.A. Nortrup, 'Determination of Residues of Alachlor and Related Herbicides in Crops by Liquid Chromatography with Electrochemical Detection', *J. AOAC Int.*, **80**(5), 1104–1110 (1997).
119. W.H. Newsome, P. Collins, D. Lewis, 'Determination of Residues of Alachlor and its Metabolite in Corn, Soybeans, and Potatoes as Heptafluorobutyryl-2,6-diethylaniline', *J. Assoc. Off. Anal. Chem.*, **69**(4), 591–593 (1986).
120. C. Sánchez-Brunete, L. Martínez, J.L. Tadeo, 'Determination of Corn Herbicides by GC/MS and GC-NPD in Environmental Samples', *J. Agric. Food Chem.*, **42**, 2210–2214 (1994).
121. K.S. Nam, J.W. King, 'Supercritical Fluid Extraction and Enzyme Immunoassay for Pesticide Detection in Meat Products', *J. Agric. Food Chem.*, **42**, 1469–1474 (1994).
122. V. Lopez-Avila, C. Charan, 'Using Supercritical Fluid Extraction and Enzyme Immunoassays to Determine Pesticides in Soil', *Trends Anal. Chem.*, **13**(3), 118–125 (1994).
123. G.K. Stearman, J.M. Wells, S.M. Adkisson, T.E. Ridgill, 'Supercritical Fluid Extraction Coupled with Enzyme Immunoassay Analysis of Soil Herbicides', *Analyst*, **120**, 2617–2621 (1995).
124. T. Gomyo, S. Kobayashi, Y. Soeda, 'Analytical Method for the Total Content of Alloxydim-sodium Herbicide and its Degradation Products in Ground Water by Gas Chromatography Mass Spectrometry', *Anal. Sci.*, **7**, 23–27 (1991).
125. W.X. Chen, Z.B. Lu, Y.L. Liu, 'Residue Analysis of Herbicide Sethoxydim and its Major Active Metabolite in Rapeseed', *Sepu.*, **10**, 371–338 (1992).
126. J. Dugay, M.C. Hennion, 'Evaluation of the Performance of Analytical Procedures for the Trace-level Determination of Aminotriazole in Drinking Waters', *Trends Anal. Chem.*, **14**(8), 407–414 (1995).
127. W. Dedek, K.D. Wenzel, H. Oberländer, B. Mothes, J. Männig, 'Preconcentration of Hydrophilic and Hydrophobic Pesticides from Aqueous Solutions and Extraction of Residues Using the Polymeric Sorbent Wofatit Y77', *Fresenius' J. Anal. Chem.*, **339**, 201–206 (1991).
128. V. Pichon, M.C. Hennion, 'Comparison of On-line Enrichment Based on Ion-pair and Cation-exchange Liquid Chromatography for the Trace-level Determination of 3-amino-1,2,4-triazole (Aminotriazole) in Water', *Anal. Chim. Acta*, **284**, 317–326 (1993).
129. A. Pachinger, E. Eisner, H. Begutter, H. Klus, 'A Simple Method for the Determination of Amitrole in Drinking and Ground Water', *Fresenius' J. Anal. Chem.*, **342**, 413–415 (1992).
130. J.M. Van der Poll, M. Vink, J.L. Quirijns, 'Capillary Gas Chromatographic Determination of Amitrole in Water with Alkali Flame Ionization Detection', *Chromatographia*, **25**(6), 511–514 (1988).
131. I. Bernwieser, G. Sontag, 'HPLC with Electrode Array Detection. Determination of Amitrole in Drinking and Ground Water', *Fresenius' J. Anal. Chem.*, **347**, 499–501 (1993).
132. W. Ternes, H.A. Rüssel-Sinn, 'Determination of 3-amino-1,2,4-triazole (Amitrole) in Tissues with Electrochemical Detection by Ion-pair High-performance Liquid Chromatography', *Fresenius' J. Anal. Chem.*, **326**, 757–759 (1987).
133. J.M. Van der Poll, M. Vink, J.L. Quirijns, 'Determination of Amitrole in Plant Tissues and Sandy Soils by Capillary Gas Chromatography with Alkali Flame Ionization Detection', *Chromatographia*, **30**(3/4), 155–158 (1990).
134. M.J. Page, Y. Ma, H. Qi, E. Healy, J. Trolinger, 'Determination of Thiazopyr and its Metabolite in Florida Groundwater Using Graphitized Carbon Solid Phase Extraction and Liquid Chromatographic Determination', *J. Agric. Food Chem.*, **45**(8), 3095–3099 (1997).
135. R.A. Pérez, C. Sánchez-Brunete, J.L. Tadeo, 'Determination of Thiazopyr Residues in Soil and Plants by Gas Chromatography with Nitrogen–Phosphorus Detection and Confirmation by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **778**, 193–199 (1997).
136. S.D. West, L.G. Turner, 'Residue Level Determination of the Aquatic Herbicide Fluridone and a Potential Photoproduct (*N*-methylformamide) in Water', *J. Assoc. Off. Anal. Chem.*, **71**(5), 1049–1053 (1988).
137. S.D. West, E.W.J. Day, 'Determination of Fluridone Residues in Meat, Milk, Eggs, and Crops by

High-performance Liquid Chromatography or Gas Chromatography', *J. Agric. Food Chem.*, **36**, 53–56 (1988).
138. S.D. West, E.W.J. Day, 'Liquid Chromatographic Determination of Fluridone Aquatic Herbicide and its Metabolite in Fish and Crayfish', *J. Assoc. Off. Anal. Chem.*, **69**(5), 856–859 (1986).
139. W.T. Willian, T.C. Mueller, 'Liquid Chromatographic Determination of Norflurazon and its Initial Metabolite in Soil', *J. AOAC Int.*, **77**(3), 752–755 (1994).
140. A. Pachinger, E. Eisner, H. Begutter, H. Klus, 'Determination of a Metabolite of the Herbicide Pyridate in Drinking and Groundwater Using High-performance Liquid Chromatography with Amperometric Detection', *J. Chromatogr. A*, **558**, 369–373 (1991).
141. Y. Iwata, J.G. MacConnell, J.E. Flor, I. Putter, T.M. Dinoff, 'Residues of Avermectin B_{1a} on and in Citrus Fruits and Foliage', *J. Agric. Food Chem.*, **33**, 467–471 (1985).
142. R.J. Demchak, J.G. MacConnell, 'Normal-phase Chromatography and Post-column Colorimetric Detection of Abamectin-8,9-oxide', *J. Chromatogr.*, **522**, 353–358 (1990).
143. R.E. Biagini, W. Tolos, W.T. Sanderson, G.M. Henningsen, B. MacKenzie, 'Urinary Biomonitoring for Alachlor Exposure in Commercial Pesticide Applications by Immunoassay', *Bull. Environ. Contam. Toxicol.*, **54**, 245–250 (1995).
144. D. Volmer, A. Preiss, K. Levsen, G. Wünsch, 'Thermospray Mass Spectral Studies of Pesticides. Temperature and Salt Concentration Effects on the Ion Abundances in Thermospray Mass Spectra', *J. Chromatogr.*, **647**, 235–259 (1993).
145. M. Yoshida, T. Watabiki, T. Tokiyasu, N. Ishida, 'Determination of Paraquat and Diquat by Liquid Chromatography–Thermospray Mass Spectrometry', *J. Chromatogr.*, **628**, 235–239 (1993).
146. M.T. Galceran, M.C. Carneiro, M. Díez, L. Puignou, 'Separation of Quaternary Ammonium Herbicides by Capillary Electrophoresis with Indirect UV Detection', *J. Chromatogr. A*, **782**, 289–295 (1997).
147. J. Cai, Z. El Rassi, 'Capillary Zone Electrophoresis of Two Cationic Herbicides, Paraquat and Diquat', *J. Liq. Chromatogr.*, **15**(6–7), 1193–1200 (1992).
148. E. Moyano, D.E. Games, M.T. Galceran, 'Determination of Quaternary Ammonium Herbicides by Capillary Electrophoresis–Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**(11), 1379–1385 (1996).
149. X.B. Song, W.L. Budde, 'Capillary Electrophoresis–Electrospray Mass Spectra of the Herbicides Paraquat and Diquat', *J. Am. Soc. Mass Spectrom.*, **7**(9), 981–986 (1996).
150. A. Walcarius, L. Lamberts, 'Square Wave Voltammetric Determination of Paraquat and Diquat in Aqueous Solution', *Electroanal. Chem.*, **406**(1–2), 59–68 (1996).
151. J.M. Zen, S.H. Jeng, H.J. Chen, 'Determination of Paraquat by Square Wave Voltammetry at a Perfluorosulfonated Ionomer/Clay-modified Electrode', *Anal. Chem.*, **68**(3), 498–502 (1996).
152. D. Barceló, G. Durand, R.J. Vreeken, G.J. De Jong, H. Lingeman, U.A.Th. Brinkman, 'Evaluation of Eluents in Thermospray Liquid Chromatography–Mass Spectrometry for Identification and Determination of Pesticides in Environmental Samples', *J. Chromatogr.*, **553**, 311–328 (1991).
153. R.A. Perez, C. Sanchez-Brunete, E. Miguel, J.L. Tadeo, 'Analytical Method for the Determination in Soil of Herbicides Used in Forestry by GC-NPD and GC/MS', *J. Agric. Food Chem.*, **46**, 1864–1869 (1998).
154. R.P. Richards, J.W. Kramer, D.B. Baker, K.A. Krieger, 'Pesticides in Rainwater in the Northeastern United States', *Nature*, **327**, 129–131 (1987).
155. A. Copin, R. Deleu, 'Application de la Chromatographie en Phase Gazeuse et de la Spectrométrie de Masse á la Recherche des Residus de Pendimethaline dans les Eaux', *Anal. Chim. Acta*, **208**, 331–336 (1988).
156. E.R. Bogus, T.L. Watschke, O.R. Mumma, 'Utilization of Solid-phase Extraction and Reversed-phase and Ion-pair Chromatography in the Analysis of Seven Agrochemicals in Water', *J. Agric. Food Chem.*, **38**(1), 142–144 (1990).
157. P.C. Bardalaye, W.B. Wheeler, J.L. Templeton, 'Derivatization of Oryzalin for Characterization by Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **314**, 450–454 (1984).
158. W.C. Koskinen, L.J. Jarvis, R.H. Dowdy, D.L. Wyse, D.D. Buhler, 'Automation of Atrazine and Alachlor from Soil Using a Laboratory Robotic System', *Soil Sci. Soc. Am. J.*, **55**, 561–562 (1991).
159. D.F. Tomkins, 'Gas Chromatographic Determination of Alachlor in Microcapsulated Formulations: Minicollaborative Study', *J. Assoc. Off. Anal. Chem.*, **70**(6), 1056–1058 (1987).
160. Z. Llu, S.R. Sirimanne, D.G. Patterson, L.L. Needham, 'Comprehensive Two-dimensional Gas Chromatography for the Fast Separation and Determination of Pesticides Extracted from Human Serum', *Anal. Chem.*, **66**, 3086–3092 (1994).
161. A. Balinova, 'Gas Chromatographic Determination of Chloroacetamide Herbicides in Plants and Soil', *J. Chromatogr. A*, **455**, 391–395 (1988).
162. R.J. Bushway, L.R. Paradis, L.B. Perkins, T.S. Fan, B.E.S. Young, P.E. Walser, 'Determination of Methyl 2-Benzimidazolecarbamate in Wine by Competitive Inhibition Enzyme Immunoassay', *J. AOAC Int.*, **76**(4), 851–857 (1993).
163. US Environmental Protection Agency, 'Analyte-specific Method of Acetochlor Detection', USEPA, Washington, DC, 1998.
164. H.-R. Buser, M.D. Müller, 'Environmental Behavior of Acetamide Pesticide Stereoisomers. 1-Stereo- and

Enantioselective Determination Using Chiral High-resolution Gas Chromatography and Chiral High-performance Liquid Chromatography', *Environ. Sci. Technol.*, **29**, 2023–2030 (1995).
165. J.W. Kuan, A.B. Pepperman, 'High-performance Liquid Chromatographic Determination of Alachlor in Alginate-based Formulations', *J. Liq. Chromatogr. Relat. Technol.*, **19**(4), 645–659 (1998).
166. R.J. Bushway, H.L. Hurst, L.B. Perkins, L. Tian, C.G. Cabanillas, B.E.S. Young, B.S. Ferguson, H.S. Jennings, 'Atrazine, Alachlor and Carbofuran Contamination of Well Water in Central Maine', *Bull. Environ. Contam. Toxicol.*, **49**, 1–9 (1992).
167. W.J. Driskell, R.H. Hill, D.B. Shealy, R.D. Hull, C.J. Hines, 'Identification of a Major Human Urinary Metabolite of Alachlor by LC-MS/MS', *Bull. Environ. Contam. Toxicol.*, **56**, 853–859 (1996).
168. C.J. McMorris, J.V. Lamont, S.P. Fithgerald, 'Detection of Alachlor by a Sensitive Enzyme-linked Immunoassay', *Food Agric. Immunol.*, **6**, 395–399 (1994).
169. T.S. Lawruk, C.S. Hottenstein, 'Quantification of Alachlor in Water by Novel Magnetic Particle-based ELISA', *Bull. Environ. Contam. Toxicol.*, **48**, 643–650 (1992).
170. D.S. Aga, E.M. Thurman, 'Coupling Solid-phase Extraction and Enzyme-linked Immunosorbent Assay for Ultratrace Determination of Herbicides in Pristine Water', *Anal. Chem.*, **65**, 2894–2898 (1993).
171. D.S. Aga, E.M. Thurman, M. Pomes, 'Determination of Alachlor and its Sulfonic Acid Metabolite in Water by Solid-phase Extraction and Enzyme-linked Immunosorbent Assay', *Anal. Chem.*, **66**, 1495–1499 (1994).
172. J. Gascón, E. Martínez, D. Barceló, 'Determination of Atrazine and Alachlor in Natural Waters by a Rapid-magnetic Particle Based ELISA Influence of Common Cross-reactants: Deethylatrazine, Deisopropylatrazine, Simazine and Metolachlor', *Anal. Chim. Acta*, **311**, 357–364 (1995).
173. T.L. Fare, J.A. Itak, T.S. Lawruk, F.M. Rubio, D.P. Herzog, 'Cross-reactivity Analysis Using a Four Parameter Model Applied to Environmental Immunoassays', *Bull. Environ. Contam. Toxicol.*, **57**, 367–374 (1996).
174. J.E. France, J.W. King, 'Supercritical Fluid Extraction /Enzyme Assay: A Novel Technique to Screen for Pesticide Residues in Meat Products', *J. Assoc. Off. Anal. Chem.*, **74**(6), 1013–1016 (1991).
175. D.B. Baker, R.J. Bushway, S.A. Adams, C. Macomber, 'Immunoassay Screens for Alachlor in Rural Wells: False Positive and Alachlor Soil Metabolite', *Environ. Sci. Technol.*, **27**, 562–564 (1993).
176. S.T.A. Siebert, S.G. Reeves, R.A. Durst, 'Liposome Immunomigration Field Assay Device for Alachlor Determination', *Anal. Chim. Acta*, **282**, 297–305 (1993).
177. S.G. Reeves, G.S. Rule, M.A. Roberts, A.J. Edwards, R.A. Durst, 'Flow-injection Liposome Immunoanalysis (FILIA) for Alachlor', *Talanta*, **41**(10), 1747–1753 (1998).
178. S.T.A. Siebert, S.G. Reeves, M.A. Roberts, R.A. Durst, 'Improved Liposome Immunomigration Strip Assay for Alachlor Determination', *Anal. Chim. Acta*, **311**, 309–318 (1995).
179. K.V. Penmetsa, R.B. Leidy, D. Shea, 'Herbicide Analysis by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **745**, 201–208 (1996).
180. J.E. Walsh, B.D. MacCraith, M. Meaney, J.G. Vos, F. Regan, A. Lancia, S. Artjushenko, 'Sensing of Chlorinated Hydrocarbons and Pesticides in Water Using Polymer Coated Mid-infrared Optical Fibers', *Analyst*, **121**, 789–792 (1996).
181. M.D. Morton, F.H. Walters, D.S. Aga, E.M. Thurman, C.K. Larive, 'Nuclear Magnetic Resonance Identification of New Sulfonic Acid Metabolites of Chloroacetanilide Herbicides', *J. Agric. Food Chem.*, **45**(4), 1240–1243 (1997).
182. P.C.C. Feng, S.R. Horton, C.R. Sharp, 'A General Method for Developing Immunoassays to Chloroacetanilide Herbicides', *J. Agric. Food Chem.*, **40**, 211–214 (1992).
183. S. Ramanathan, K.O. Jayaseelan, K.K. Sonawane, P.S. Bhate, 'Chromatographic Approaches for the Assay of Butalchlor in Technical Grade Materials', *Res. Ind.*, **35**, 17–22 (1990).
184. D.F. Tomkins, 'Gas Chromatographic Determination of Butachlor in Formulation: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **69**(4), 721–722 (1986).
185. L.E. Sojo, D.S. Gamble, D.W. Gutzman, 'Sorption and Bound Residue Formation of Linuron, Methylparathion, and Metolachlor by Carrot Tissues: Kinetics by On-line HPLC Microextraction', *J. Agric. Food Chem.*, **80**(45), 3634–3641 (1997).
186. J. Schlaeppi, H. Moser, K. Ramsteiner, 'Determination of Metolachlor by Competitive Enzyme Immunoassay Using a Specific Monoclonal Antibody', *J. Agric. Food Chem.*, **39**, 1533–1536 (1991).
187. G.K. Stearman, J.M. Wells, 'Enzyme Immunoassay Microtiter Plate Response to Atrazine and Metolachlor in Potentially Interfering Matrices', *Bull. Environ. Contam. Toxicol.*, **51**, 588–595 (1993).
188. D.S. Aga, E.M. Thurman, M.E. Yockel, L.R. Zimmerman, T.D. Williams, 'Identification of a New Sulfonic Acid Metabolite of Metolachlor in Soil', *Environ. Sci. Technol.*, **30**, 592–597 (1996).
189. D.F. Tomkins, 'Gas Chromatographic Determination of Propachlor in Pesticide Formulations: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **69**(4), 723–725 (1986).
190. F. Sánchez-Rasero, H. Peña, 'Liquid Chromatographic Method for Determination of Chloridazon in Technical Concentrates and Formulated Products: Comparison with CIPAC Method', *J. Assoc. Off. Anal. Chem.*, **71**(5), 1064–1067 (1988).
191. S. Fernández de Betoño, A. Arranz, A. Cid, J.M. Moreda, J.F. Arranz, 'Voltammetric Determination of

Chloridazone with a Silica Modified Carbon Paste Electrode', *Qutm. Anal.*, **14**, 112–116 (1995).

192. R.D. Mortimer, D.F. Weber, 'Some Comments on a Recently Proposed Method of Determining Chlormequat Residues by Derivatization with Pentafluorothiophenol', *Pesticide Sci.*, **40**, 31–35 (1994).
193. A. Valverde-García, A.R. Fernández-Alba, M. Contreras, A. Agüera, 'Supercritical Fluid Extraction of Pesticides from Vegetables Using Anhydrous Magnesium Sulfate for Sample Preparation', *J. Agric. Food Chem.*, **44**, 1780–1784 (1996).
194. M. Chiba, D.F. Veres, 'High Performance Liquid Chromatographic Method for Simultaneous Determination of Residual Benomyl and Methyl 2-Benzimidazole Carbamate on Apple Foliage without Cleanup', *J. Assoc. Off. Anal. Chem.*, **63**(6), 1291–1295 (1980).
195. P.C.C. Feng, S.J. Wratten, S.R. Horton, C.R. Sharp, E.W. Logusch, 'Development of a Enzyme-linked Immunosorbent Assay for Alachlor and its Application to the Analysis of Environmental Water Samples', *J. Agric. Food Chem.*, **38**, 159–163 (1990).

Pesticides in Water: Sampling, Sample Preparation, Preservation

Igor Liška
Water Research Institute, Bratislava, Slovakia

1 Introduction **1**
2 Sampling and Sample Preservation **2**
2.1 Sampling Site 2
2.2 Sample Volume 2
2.3 Sampling 2
2.4 Preservation of Water Samples 3
3 Sample Handling **5**
3.1 Liquid–Liquid Extraction 6
3.2 Solid-phase Extraction 7
3.3 Other Sample-handling Methods 11
4 Conclusions **12**
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**

Contamination of water by pesticides is an important issue in many regions, posing problems in the environmental, water management and health sectors. To assess the extent of contamination of water, effective and properly designed analytical methods having sufficient sensitivity and accuracy are needed. However, no matter how excellent are the state-of-the-art analytical instrumentation and techniques applied, the data on trace concentrations of pesticides provided by the analytical system will be useless unless sufficient attention is given to sampling and sample preparation.

To collect a representative water sample for pesticide analysis, all sampling parameters must be selected properly. This refers predominantly to the selection of the appropriate sampling site, sampling technique and volume of the water sample. To prevent the sample matrix from any undesirable alterations during its transport from the sampling site to an analytical laboratory, physical (e.g. temperature, light intensity) and chemical/biochemical (e.g. pH, microbial growth) conditions must be under control, i.e. the sample must be preserved carefully. An efficient alternative to the transport of liquid water samples from the sampling point to the laboratory is on-site sorption of analytes on a solid phase. This minimizes the weight of transported samples and increases the stability of analytes.

The determination of trace concentrations of pesticides in water samples and the complexity of environmental matrices require the application of an efficient sample-handling procedure prior to separation and detection of these analytes. At present, liquid–liquid extraction (LLE) and solid-phase extraction (SPE) are the most frequently applied sample-handling techniques in the determination of pesticides in water. LLE is a traditional method that is still widely used in standardized methods. The increasing popularity of SPE is a result of its versatility and effectiveness and SPE is also preferred owing to the lower health risks.

The introduction of new sorbent materials, the development of automated sample-handling systems allowing unattended operation and newly emerging sample-handling techniques such as solid-phase microextraction (SPME) and membrane separation methods guarantee that sample preparation in pesticide analysis will remain a fertile area of development in the future.

1 INTRODUCTION

The extent of the application of pesticides reaches many thousands of tons per pesticide annually in many countries and the diversity of the insecticides, herbicides, growth regulators, fungicides and other applied substances is increasing rapidly. This extensive application has a negative impact on the quality of the environment. Because of their widespread use, pesticides are currently detected by determination of their residues in various environmental matrices, especially water.

To prevent water pollution by pesticides, precise information on their concentration levels in the water environment is needed. Owing to the diversity of pesticides and their low concentrations in water, the acquisition of this information requires the development of effective and properly designed analytical methods. The initial part of any of these methods comprises sampling and sample preservation and preparation. While in the past the importance of these initial steps has often been neglected, it is now clear that any mistake that occurs in collecting and processing a water sample can lead to substantial errors in the final data, regardless of the excellent performance of the state-of-the-art analytical technique that may be applied subsequently. Moreover, the results of a survey among analysts indicated that, for all respondents, two-thirds of the analysis time was spent on collection and preparation of the sample.[1] Therefore, it is clear that sample collection and handling deserves the attention of any water analyst who is trying to increase the throughput of methods used and to improve the overall performance of the analyses performed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 SAMPLING AND SAMPLE PRESERVATION

2.1 Sampling Site

The sampling site must be selected with respect to the objectives of a monitoring program or a survey. Thus, the overall monitoring strategy predetermines the choice of profiles in water streams or the number and/or the depth of the monitoring wells.

For instance, if the objective is to monitor transboundary pollution of the water stream, the location should be as close as possible to the border. If the objective is to assess the quality of water used for drinking-water production, the sampling site should be positioned close to the intake. The number and location of sampling stations for groundwater monitoring is a function of the objectives and scale of assessment (background or trend monitoring, emergency surveys around a spill, operational surveillance of the quality of potable water resources), the hydrogeological complexity and economic considerations.

In all those cases it must be assured that the location is suitable for taking representative samples with due regard for the objective of the sampling. Furthermore, to assure representativeness of the sample, it is recommended that replicate samples be taken occasionally to determine temporal and spatial variability. A thorough discussion on monitoring strategies is beyond the scope of this article and can be found in various manuals.[2–4]

2.2 Sample Volume

Sample volume depends on the number of analyses to be performed and on the technical requirements of a particular analysis. The volume of the water sample required for a single analysis of pesticides usually ranges from tens of milliliters up to 1–2 L, depending on the methodology applied. Traditional gas chromatography (GC) or high-performance liquid chromatography (HPLC) methods, in which an aliquot of the extract of the water sample is analyzed, require processing of a large volume of water (500 mL to 2 L) in order to reach the desired sensitivity. For modern SPE procedures, several hundred milliliters of water sample usually suffice. Introduction of highly sensitive detectors, on-line SPE techniques and large-volume injection possibilities permitted the reduction of the sample volume required for one analysis down to tens or even units of milliliters.

The total volume of water sample to be collected from the sampling site depends on the total number of analytes to be analyzed. Hence, theoretically, it can vary from 10 mL when, e.g. atrazine in water is to be determined using an on-line SPE/HPLC method to 10 L and more when a complete pesticide survey is to be carried out using standard methods employing LLE. To cover a large range of pesticides requires the application of a large number of different target methods, which makes complex screening laborious. A solution is the use of a suitable combination of methods covering as much of the pesticide range as possible. This approach has been used in a National Pesticide Survey performed in the USA where seven analytical methods were used to cover more than 100 pesticides of interest.[5]

2.3 Sampling

In general, the type of sampling depends on the goal of the monitoring program and differs for river water, reservoirs, groundwater, rainwater, wastewater, drinking water or pore water from the unsaturated zone. In river waters the sampling methods are usually based on bottle collection or water pump systems. When a homogeneous reach of a stream is monitored, the collection of samples in a single vertical mode may be sufficient. For small streams a grab sample taken at the centroid of flow is usually adequate.

For sampling profiles located on a nonhomogeneous reach of a river or stream, it is necessary to sample the cross-section of the river or stream at a specified number of points and depths. In principle, the cross-section has to be checked for homogeneity, analyzing a minimum of three locations within the cross-section, the left and right banks and the center of the river. Vertical homogeneity should also be checked. Sampling in streams should be normally performed at a depth of 0.3–1.0 m. Surface sampling is employed when there is a need to collect the heterogeneous matter floating on the river surface.

A similar situation is encountered in reservoirs, where the sampling design must take into account the vertical stratification. The presence of anaerobic conditions in the lower layers and algal growth in the surface layers leads to the establishment of specific distribution of physicochemical parameters in water influencing the stability and behavior of dissolved pesticides. Therefore, along with the pesticide analysis, data on basic physicochemical parameters such as temperature, pH, conductivity, nitrates, ammonia and oxygen should also be collected.

For sampling of groundwater, different kinds of pumps withdrawing the water from the boreholes or wells are used. Such discharge samples are normally collected in bottles from a well-head tap or directly from a pump outflow. In some cases grab sampling from nonpumping boreholes is also possible. The most common method of unsaturated zone sampling is by suction or tension. When taking samples from pump discharge, it must be ensured that the sample is representative of hydrochemical conditions in the aquifer, rather than those in the well.

In the case of public supply or irrigation wells which are operating continuously (or at least for long periods) at high discharge rates, it is relatively easy to collect a sample after a sufficiently long period of pumping to be certain that the water has been drawn from the aquifer rather than from within the well itself. If, however, specialist sampling pumps with low discharge rates are used to sample nonpumping observation wells, then purging the well of standing water is required before sampling.[6] It is often recommended to purge the water by removing 3–10 well volumes. It is also advised to measure continuously selected physicochemical parameters (pH, conductivity, temperature) to detect when stable conditions have been reached, indicating that the required high proportion of aquifer water is being drawn.

Even for one particular matrix the type of sampling can vary depending on the character of the information that is to be obtained from the monitoring. There are principally two types of samples: grab samples and composite samples. A grab sample is taken at a selected location and time, and then analyzed for pesticides. The collection of grab samples is appropriate when it is desired (i) to characterize water quality at a particular time and location, (ii) to provide information about minima and maxima and (iii) to analyze parameters which can be subject to change. A composite sample is obtained by mixing several discrete samples of equal or weighted volumes collected at regular time intervals in one container, which is subsequently analyzed for the parameters of interest. Instead of discrete sampling the water sample can be pumped continuously and required volumes of the sample for the analysis can be abstracted from the total bulk. Continuous sampling is advantageous in the case of automated monitoring systems. The selection of an appropriate sampling strategy has a considerable effect on the information output from a monitoring campaign. Especially for pesticides with seasonal occurrence it is very important to determine a suitable frequency and overall scheme of sampling, otherwise large deviations in the assessment of annual mean concentrations can result.[7]

To ensure that the collected water sample is a real representative of the sampled site and to avoid the detection of false positives, strict quality-control principles should be followed. These principles include qualification of sampling staff, checking for purity of sampling containers and collecting field blanks and field check samples. The use of duplicate samples for checking of sample stability and for eliminating random sampling errors is highly advisable.

2.4 Preservation of Water Samples

Preservation of the sample during transport and storage depends on the type of pesticides to be analyzed. In principle, we have to prevent any potential means of losses of analytes from the liquid phase, such as transfer to the gaseous phase (evaporation), transfer to the solid phase (adsorption on the container walls), hydrolysis, photodegradation and biodegradation. The losses due to analyte evaporation are not a critical issue for most pesticides. In any case, the storage of the water sample at 4 °C, minimization of the volume of the gaseous phase in the container and the use of gas-tight caps is recommended. Hydrophobic pesticides can be easily adsorbed on polymer surfaces. From an aqueous sample containing highly lipophilic organic compounds that are stored in a common plastic bottle, more than 90% of these compounds can be adsorbed within 24 h. Therefore, the storage of water samples in plastic containers must be avoided and the use of glass vessels only is recommended. Many modern polar pesticides can easily hydrolyse when the pH reaches a certain critical value. Hence it is necessary to maintain the pH at a desired value using a buffer solution or, usually in the case of acidic compounds, simply by acidifying the water sample. Keeping the samples in the dark and using amber-glass sample containers when available should prevent photolysis of the analytes. The presence of bioorganisms in water leads to biodegradation of dissolved pesticides. To suppress the biological activity of the aqueous environment, biodegradation inhibitors such as $HgCl_2$ are used.

Most standardized methods recommend reducing the time of storage of the water sample to a minimum and performing the extraction as soon as possible. Storage periods that are usually officially suggested range from 7 to 14 days. In the United States Environmental Protection Agency (USEPA) National Pesticide Survey, thorough analyte stability studies of both well water samples and final extracts were performed.[5] The primary focus of these studies was to demonstrate the stability of analytes for the maximum sample and extract holding times that would be permitted during the survey. Up to 100% loss was observed for 26 out of the 147 pesticides spiked into well water samples which had been biologically inhibited and stored at 4 °C for 14 days. The same pesticides generally remained stable in stored sample extracts.

An alternative to the transport of liquid samples is the on-site sorption of analytes on the solid phase, which minimizes the weight of transported samples and increases the stability of analytes. In many cases, the transfer of groundwater samples from an anaerobic aquifer to an aerobic environment may initiate oxidation or biodegradation of organic compounds, which may continue during the transport. Also, when taking a sample from an aerobic environment, photodegradation and biodegradation processes can alter the composition of the aqueous sample irreversibly. Using sorption

on a sorbent, analyte evaporation is minimized and adsorption on the container walls is avoided. Binding of molecules of the organic compounds to the active sites on the sorbent surface slows undesirable changes in their quality and quantity. Green and LePape[8] observed that XAD-2 macroreticular resin and octadecyl-bonded silica (C_{18}) had a preservative effect, which prevented the breakdown of sorbed hydrocarbons by bacteria. Hydrocarbons stored on these solid phases for periods of up to 100 days in the presence of an oleophilic bacterial population showed no evidence of biological degradation. In contrast, hydrocarbons stored in water samples containing the same bacteria showed pronounced degradation over much shorter storage periods. These authors suggested that the preservative effect resulted from trapping the organic compounds in the adsorbent lattice structure. The pores of XAD-2 or silica gel were smaller than bacteria, hence, the hydrocarbons were protected from bacterial attack. The stability of 34 selected pesticides extracted from water on to a graphitized carbon black (GCB) surface was evaluated under various storage conditions.[9] The best results were obtained by first minimizing the water content in the GCB extraction cartridge by a suitable methanol washing and freezing the cartridge. Under these conditions and over a storage period of 3 weeks the stability of pesticides extracted from four river water samples on to the GCB surface was assessed and compared with that in water at 4 °C with and without an inhibitor of biological degradation such as mercuric chloride. The results indicated that storage on the GCB material was a far better preservation procedure than storage in water at 4 °C. Several of the pesticides considered were completely degraded when stored in water in the presence of $HgCl_2$. Similarly to disposable off-line SPE cartridges, SPE precolumns were also tested for analyte storage possibilities. The stability of 19 organophosphorus pesticides was investigated using precolumns from the Prospekt (automated on-line SPE) system packed with C_{18} and stored under different conditions.[10] This study demonstrated that many organophosphorus pesticides that showed instability problems in water matrices were stable in disposable SPE precolumns for a period of up to 8 months at −20 °C. Similar cartridges were also used for the investigation of the stability of different groups of polar pesticides (triazines, phenylureas, carbamates and chlorinated phenols) sorbed on polymer sorbents from water samples.[11] The results showed the stability of most of the tested pesticides stored in cartridges for 7 weeks at the laboratory temperature or in a refrigerator. In Figure 1 it can be seen that except for carbaryl the recoveries of all tested pesticides were in the range 80–120% within 7 weeks of storage, which is an acceptable variance in organic trace analysis. The differences between stability of analytes when stored on a wet or silica gel-dried sorbent were negligible. The pH value of the water sample had a significant influence on the stability of only those compounds

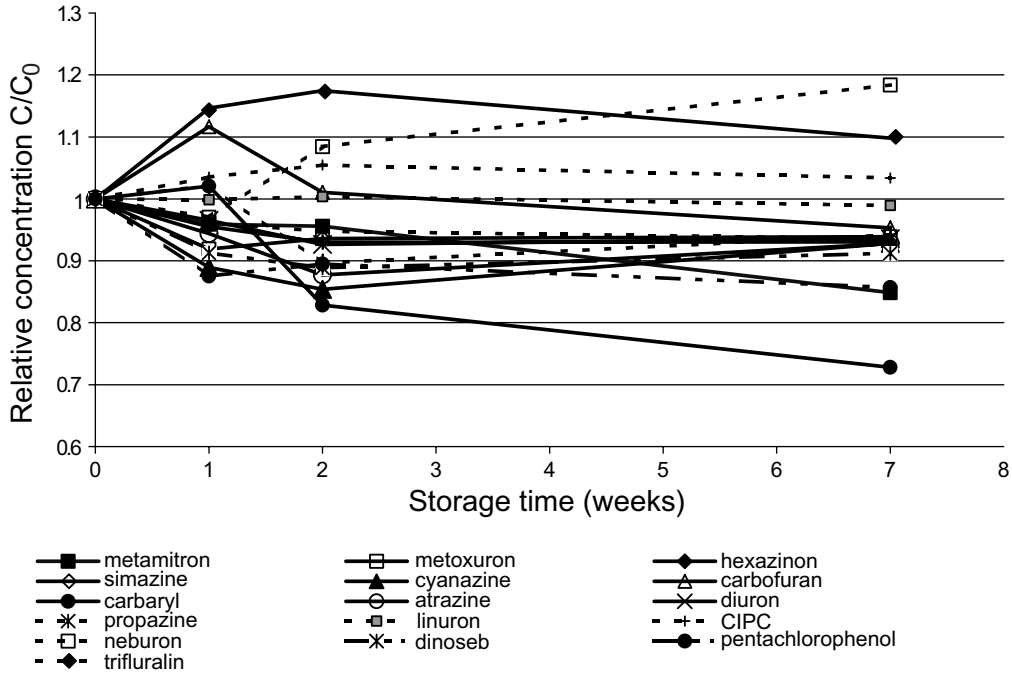

Figure 1 Stability of pesticides sorbed from spiked tap water on to polymer cartridges that were stored wet in a refrigerator at 4–5 °C. Stability is expressed as the change of chromatographic peak area with time.

which had acid–base properties. The results obtained with surface water were similar to those obtained with tapwater. Also, a new SPE material, membrane extraction disks, was found to be a good preservation medium for pesticides.[12] The stability of triazines and acetanilides on these disks at −20 °C for 3 months, during which there was no evidence of degradation, indicated the feasibility of storage under these conditions. In general, in all of these stability investigations SPE was confirmed to be a suitable sample-handling technique when long-term preservation of analytes prior to the analysis was required.

3 SAMPLE HANDLING

Trace concentrations of organic analytes in water samples and the complexity of environmental matrices require the application of an appropriate sample-handling procedure prior to the separation and detection of these analytes. As has already been mentioned earlier, for contemporary environmental analysis it is characteristic that highly efficient separation and detection systems are often coupled with laborious and time-consuming sample-handling procedures which limit the sample throughput and hence the overall performance of the method.[1] Hence, at present, sample handling is considered to be the weakest point in environmental chromatographic analysis. The methods of and approaches to the sample handling of pesticides in the aquatic environment have already been thoroughly reviewed by Barceló and Hennion[13] and related reviews can also be found elsewhere.[14–17]

The major reasons for the preparation of water samples prior to pesticide analysis are

- preconcentration of analytes in order to enrich the sample with traces of the analytes;
- removal or reduction of compounds that interfere in the separation and/or detection steps;
- transfer of analytes into a homogeneous liquid matrix amenable to chromatographic analysis.

In most cases the preconcentration factors required for the efficient analysis of pesticides at the environmentally relevant levels are 100–10 000. In addition, a sample-handling procedure must be sufficiently robust to provide reproducible data for different aqueous matrices with various properties. The selected sample-handling strategy usually depends on the aim of the analysis and on the information expected, i.e. whether the analyst wants to detect one or several particular compounds of interest (target analysis) or if a multiresidue screening of a wide range of pesticides is required. In target analysis, the sample-handling step employs more selective procedures and media and more attention is paid to cleanup techniques. In screening analysis, nonselective procedures are usually preferred and, when necessary, a combination of several sample-handling techniques can be applied simultaneously. For example, during a long-term international program on river pollution identification that was performed in the Nitra river basin,[18] the simultaneous use of two LLE procedures along with an SPE method was performed to increase the ''analysis window'' as much as possible so that most of the organic pollutants present in the river water could be detected.

The complexity of a sample-handling procedure to be employed depends also on the type and level of the separation and detection system applied for this analysis. The use of a relatively nonselective detector or a chromatographic system with a low separation power in a target method requires the application of a selective trace enrichment procedure. On the other hand, a selective and sensitive detector used with a good separation system provides acceptable results even when a very simple sample-handling scheme is applied.

Nowadays, LLE and SPE are the most frequently applied sample-handling techniques in the determination of pesticides in water. The prevalence of LLE applications in environmental analysis in the past resulted from the hydrophobic character of pesticides listed on the priority pollutant lists at that time and also from the absence of an equally suitable and efficient technique. At present, LLE is still widely used in standardized methods for pesticide analysis, primarily those used for nonpolar pesticides. It is often preferred owing to the availability of standardized procedures and good descriptions in the analytical literature.

The increasing popularity of SPE is due not only to its obvious advantages over LLE (minimized consumption of organic solvents, no emulsion formation, reduced contact of the analyst with potentially toxic substances) but also to gradual refinements of this procedure which minimize its original drawbacks. The introduction of membrane extraction disks increased the speed of the preconcentration and the growing production of SPE materials lowered their prices. A wide choice of newly developed materials makes SPE perfectly suitable to cope with the growing variability of modern pesticides. At present, the major factors stimulating the preference of the use of SPE in water analysis are the possibility of automation, sample storage and preservation ability and a high potential in offering solutions for complex preseparation schemes.

The theoretical background of LLE and SPE is based on similar principles of the interphase distribution of a solute in a two-phase system. The distribution of a solute in such a two-phase system (liquid–liquid or liquid–solid)

is governed by the affinities of the solute to these phases and it is characterized by a distribution constant. In LLE, the transport of solutes from an aqueous phase into an organic solvent immiscible with water depends mostly on the hydrophobicity of the solute and its solubility in the solvent. In this case the preconcentration is carried out as a one-step transfer of the solute from water to an organic solvent. After equilibrium has been reached, an aliquot of the organic solvent is introduced into an analyzer. However, when needed, the solutes can be redissolved in another suitable solvent.

On the other hand, most SPE procedures comprise two-step transfer of the solute. First, the solute is sorbed from water on the solid surface of a sorbent or in the stationary phase bound to this surface and, subsequently, the solute is eluted with an appropriate solvent (or thermally desorbed into the gas chromatograph). Here, the choice of a suitable sorbent (i.e. the selection of strength of interactions responsible for binding the solute on the solid phase) and the choice of eluting solvent or thermal desorption conditions (i.e. the energy needed for breaking of the bonds) must be properly tuned to obtain a quantitative preconcentration.

3.1 Liquid–Liquid Extraction

In the 1960s and 1970s, the attention of environmental scientists was focused mostly on nonpolar pesticides which, thanks to the extensive implementation of electron-capture detection (ECD) in analytical practice, were found to be ubiquitous and owing to their accumulation potential they were recognized as a serious threat to the environment. Despite the excellent sensitivity of ECD, the preconcentration step was needed to reach low nanogram per liter levels, as was required by environmental standards. The strongly hydrophobic character of most of those compounds was a prerequisite for the choice of LLE as a preferred preconcentration technique that gave satisfactory recoveries for nonpolar pesticides.

Most of the classical preconcentration methods employing LLE have a relatively similar pattern: a water sample is extracted two or three times with a small volume (usually 10–50 mL) of an organic solvent (hexane, heptane, cyclohexane, methylene chloride, Freon, etc.) and the extracts (plus the solvent used for rinsing the internal surface of the glassware used for the extraction) are dried with purified sodium sulfate, filtered, evaporated either to dryness or to a volume usually under 1 mL and redissolved in a small volume of the solvent compatible with a selected chromatographic technique. This general pattern can, naturally, differ in various modifications and/or refinements of original methods but, in principle, three consecutive extraction steps with extraction volume ratios of approximately 1 : 30–40 are sufficient for removal of at least 90% of the hydrophobic compounds from water. To remove interfering compounds, the extract can be cleaned by percolation through a column packed with Florisil, alumina or another suitable sorbent.

Many different methods for pesticide analysis involving LLE have been published in the past but the actual strength of LLE is the availability of a number of standard procedures. Examples of standard methods for the determination of different groups of pesticides in water using LLE are shown in Table 1. Moreover, LLE is used not only in standardized methods for the determination of specific groups of pesticides but also in broad-range methods covering a large number of priority pollutants.[19] A great strength of common standardized LLE methods is that in addition to their standard description needed to achieve status as an official method, also the results of their interlaboratory validation can be found in the literature.[20,21] These collaborative studies help in judging the acceptability of a particular method by characterizing its multilaboratory performance in terms of recovery, overall and single analyst precision and the effect of water type on recovery and precision. The availability of the results from an interlaboratory validation supports the reliability of a tested method.

Even though it may seem that in the field of LLE everything has been investigated and solved, there are still new approaches to be studied. The use of a Goulden large-sample extractor was tested to reach DLs in the low nanogram per liter range[22] and this technique was presented as an alternative for SPE. The general strategy for multiresidue analysis of polar pesticides in groundwater using LLE with methylene chloride was published by Hoogendoorn et al.[23,24] The extracts were cleaned up on-line by means of column switching and subsequently analyzed by step-gradient elution. This approach was found to be powerful for the simultaneous assay of polar pesticides.

The use of LLE also simplifies the problem of the determination of the total content of a hydrophobic organic compound in a water environment. This problem is relevant especially for surface-water samples with a high content of suspended solids. In an LLE procedure, an organic solvent extracts hydrophobic analytes from both the aqueous solution and solid particles suspended in the water. Rinsing of the glassware minimizes the losses of analytes caused by adsorption on the glassware surface. Thus, the analytes present in the water sample are quantitatively recovered into the extract. Using an SPE procedure for handling of such heterogeneous water samples usually requires filtration through a membrane filter prior to an extraction. Therefore, the result of such SPE analysis provides information only on the content of an analyte dissolved in the aqueous phase. Also, if no filtration was applied prior to SPE preconcentration, the analytes

Table 1 Examples of preservation and handling of water samples in standardized methods for pesticide analysis employing LLE

Method[a]	Sample preservation	Extraction	Separation and detection[b]	Method DL ($\mu g L^{-1}$)
EPA 505 Organohalide pesticides	14 days (heptachlor 7 days), 4 °C	Extraction of sample with 2 mL of hexane	GC/ECD	0.002–6.8
EPA 507 Nitrogen- and phosphorus-containing pesticides	10 mg L^{-1} $HgCl_2$, 4 °C in the dark, 14 days only for some analytes, immediate analysis recommended	Extraction with methylene chloride, concentration of extract to 5 mL after substitution with methyl *tert*-butyl ether	GC/NPD	0.075–5
EPA 508 Chlorinated pesticides	10 mg L^{-1} $HgCl_2$, 4 °C, 7 days, immediate analysis recommended	Extraction with methylene chloride, concentration of extract to 5 mL after substitution with methyl *tert*-butyl ether	GC/ECD	0.0015–5
EPA 515.1 Chlorinated acids	10 mg L^{-1} $HgCl_2$, 4 °C in the dark, 14 days	Hydrolysis by adjustment of water sample to pH 12, solvent wash, acidification, extraction with diethyl ether, esterification using diazomethane	GC/ECD	0.02–1.3
EN ISO 6468 Organochlorine insecticides	4 °C in the dark, 24 h	Extraction with an organic solvent (hexane, heptane or petroleum ether), concentration of the extract to 1 mL	GC/ECD	0.001–0.01
USGS O-1104-83 Organochlorine and organo-phosphorus compounds	4 °C	Hexane, evaporation to 0.5 mL	GC/ECD	0.01
USGS O-1105-83 Chlorophenoxy acids	pH 2, 4 °C	Extraction with diethyl ether or methyl *tert*-butyl ether. Hydrolysis to free acids, methylation with boron trifluoride–methanol and purification using adsorption chromatography	GC/ECD	0.01
USGS O-3106-83 Triazines	–	Extraction with methylene chloride, adjustment to pH 7 to 9. Optional cleanup on alumina	GC/NPD	0.1
USGS 3107-83 Carbamates	–	Extraction with methylene chloride, evaporation to 1 mL, redissolution in acetonitrile	HPLC/UV	2.0
ASTM 6630 Organochlorine pesticides	4 °C, 7 days	Extraction with a mixture of methylene chloride and hexane or diethyl ether and hexane	GC/ECD	0.003–0.24
ASTM 6640 Chlorophenoxy acids	–	Extraction of acidified sample with diethyl ether, hydrolysis, esterification	GC/ECD	0.002–0.01

[a] ISO, International Organization for Standardization; EPA, United States Environmental Protection Agency; ASTM, Standard Methods for the Examination of Water and Wastewater, 18th Edition, American Public Health Association, Washington, DC, 1992; USGS, Methods for the Determination of Organic Substances in Water and Fluvial Sediments, United States Geological Survey, Washington, DC, 1987.
[b] DL, detection limit; NPD, nitrogen–phosphorus detector.

bound to suspended solids can hardly be trapped on the sorbent in the SPE cartridge and, therefore, they are lost for the analysis. The distribution of the analyte between the solid and aqueous phases depends on its hydrophobicity and its loss will increase with decreasing polarity.

3.2 Solid-phase Extraction

SPE, which has been developed intensively in recent decades, has become a powerful alternative technique to LLE owing to its simplicity, flexibility and high sample

throughput. Additional advantages include reduction of toxic solvent consumption and greater health safety. An overview of the advantages of SPE over LLE in water analysis can be found in the literature.[13–15,25] The results of experimental comparative tests have shown that a performance comparable to that of LLE could be achieved using SPE.[26–31]

The basic principle of SPE is the transfer of analytes from the aqueous phase to the active sites of the adjacent solid phase. This transfer is stimulated by the selection of appropriate operational conditions in the system of three major components, water (liquid phase)–sorbent–analyte. After the replacement of the water by a suitable liquid phase in this system, the analyte can be desorbed from the sorbent and further analyzed. Usually, the SPE process is carried out in the column and is often referred to as low-performance liquid chromatography.

Many factors influence the efficiency of the SPE process, the two most important being capacity and retention. An insufficient capacity of the sorbent surface can cause its overloading and, consequently, earlier breakthrough of analytes. However, this situation is considered not to be very likely owing to the low concentrations of organic compounds in treated water samples and the relatively high sorption capacities of applied sorbents. The more critical factor is the retention of analytes, which should be a maximum in the water–sorbent–analyte system and a minimum in the eluent–sorbent–analyte system. The existence of these two contradictory demands on the strength of the sorbent–analyte interactions leads to the necessity to make a compromise during the selection of working conditions for sorption and desorption so as to obtain an optimum preconcentration.

A typical sequence of SPE includes the following steps: activation of the sorbent (wetting with a suitable solvent), conditioning (replacing of the activation solvent by the aqueous phase), percolation of the water sample, cleanup (removal of interfering components), drying of the sorbent bed, elution of accumulated analytes and regeneration of the sorbent [usually not recommended for disposable cartridges because of memory effects[32] and/or hysteresis effects[33]].

With respect to the system approach, two modes of SPE can be distinguished: off-line and on-line. In the on-line configuration, the SPE column output is connected directly to the analytical column, so that elution and separation of analytes are performed in one step. In the off-line configuration, the elution and separation of analytes are two separate steps. The chemistry and general principles are the same for both of these variants, but, the differences in their methodology are sometimes the reason for certain drawbacks or advantages of one or other of these approaches in a particular application.[34]

3.2.1 Off-line Solid-phase Extraction

The off-line approach is very simple and even though it can be easily performed using a home-made system, a number of vacuum manifolds for sample extraction and analyte elution are commercially available. Similarly to traditional LLE, off-line SPE is also becoming a part of standardized protocols.[35] In general, using an off-line configuration, high operational flexibility with regard to the selection of different materials and elution solvents is available. This approach allows the selective fractionation of the original sample and the simultaneous use of different separation and detection techniques for one sample extract. Asafu-Adjaye et al.[36] separated kepone from other pesticides by flushing the loaded SPE column with hexane, which eluted DDT (1,1,1-trichloro-2,2-bis[*p*-chlorophenyl]-ethane), DDE (1,1-dichloro-2,2-bis[*p*-chlorophenyl]-ethylene) and HCB (hexachlorobenzene). Kepone and its metabolites were subsequently eluted with a mixture of hexane and diethyl ether. Wells et al.[37] discussed SPE as the chromatographic technique to be used for the selective fractionation of wastewater effluents. An elution scheme incorporating water, methanol and methylene chloride for the fractionation of nonpolar organic toxicants using SPE was found to be effective for the separation and isolation of compounds with log K_{ow} values ranging from 2.5 to 7 from aqueous solutions and for toxicants from a sediment pore water sample.[38]

In addition to the fractionation possibility, off-line SPE is a good substitute for LLE in the preconcentration of pesticides with a large range of polarity. The efficiency of the preconcentration depends substantially on the selection of a suitable solid phase. Bonded silicas, polymers and carbon sorbents are the most frequently applied SPE materials in newly developed sample-handling procedures used for the determination of pesticides in water.[39,40] When needed, the combination of several SPE columns and the use of several eluting solvents can increase the efficiency of the SPE procedure. DiCorcia et al.[41] extracted phenoxyacetic acid herbicides from water using a miniaturized cartridge containing GCB in the top part and a silica-based strong anion exchanger in the bottom part. After the percolation of the water sample through this cartridge, the anion exchanger was activated with sodium acetate solution. The sorbed analytes were then transferred from the GCB to the anion exchanger with methylene chloride–methanol solvent mixture basified with sodium hydroxide. After washing, herbicides were desorbed from the anion exchanger with water–methanol containing trifluoroacetic acid and

potassium chloride. The DLs were in the parts per trillion (nanograms per liter) range. The use of off-line SPE in screening for pesticides in water requires simple instrumentation and provides low DLs owing to the possibility of partial evaporation of the eluate from the SPE cartridge. DiCorcia et al.[42] presented an off-line approach to monitoring a large group of pesticides in ground- and river water. The method incorporated a fractionation of analytes into basic and neutral compounds and acidic compounds, based on two different interaction mechanisms of GCB. Processing of large volumes of water (0.5–2 L) and evaporation of the eluates led to DLs lower than 0.1μg L^{-1} for most of the pesticides in this case.

The ease of manipulation with disposable SPE media is being further enhanced. In the 1990s, membrane extraction disks were introduced as an attractive kind of SPE material. These membranes consist of a fibrillated polytetrafluoroethylene (PTFE) matrix in which sorbents such as bonded silicas, polymers or ion exchangers are enmeshed. Owing to the internal structure of the disks, high flow rates of water sample can be achieved and a reduction of the recovery due to channelling is avoided. The easy manipulation with disks makes them suitable for quick testing methods. Owing to obvious benefits of their use, disks became the preferred media for Drinking Water Test Methods as defined by the USEPA.[43] In the determination of pesticides in water, the eluate from the disk can be injected into a GC system and, if necessary, derivatization of analytes can be performed prior to the injection.[44] For screening of a wide range of pesticides, the eluate can be analyzed by both GC and liquid chromatography (LC) systems simultaneously.[45]

3.2.2 On-line Solid-phase Extraction

The principle of the on-line approach is the direct transfer of adsorbed analytes from the SPE column to the analytical column by the mobile phase. Hence this approach is also called precolumn switching or precolumn technology. The principles and technical aspects of the on-line approach, especially for HPLC analyses, have been described thoroughly[13,34,46] and many applications have been published in environmental and in biomedical analysis.

A typical simple on-line system is shown in Figure 2. It consists of two circuits connected together by the switching valve. The aqueous sample is introduced to the precolumn by the pump in the low-pressure (preconcentration) circuit and, subsequently, the precolumn is switched to the high-pressure circuit where the analytes are eluted by the mobile phase directly into the analytical column. The sample volume applied usually ranges

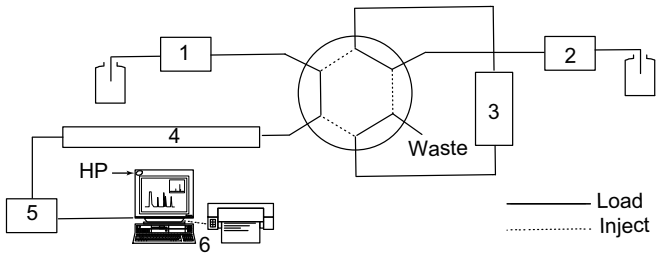

Figure 2 Basic on-line SPE/HPLC set-up. 1 = HPLC pump for analytical mobile phase; 2 = sample pump; 3 = SPE precolumn; 4 = analytical column; 5 = HPLC detector; 6 = workstation with printer.

from 10 to 200 mL depending on the analytes and on the total organic load of the water sample. The internal volume of the precolumn ranges from several tens to several hundreds of microliters. In principle, sorbent material in the SPE precolumn should be identical with the packing material in the analytical column to prevent broadening of chromatographic peaks (band broadening). Therefore, owing to the prevalence of octadecylsilica LC columns, C_{18} is frequently used as a precolumn packing. C_{18} precolumns are sufficient for the preconcentration of compounds with low polarity but problems occur in case of polar analytes. In such case, a sorbent having stronger interactions with these analytes than a conventional C_{18} material should be used in the precolumn. If the precolumn containing more hydrophobic material is connected to a C_{18} analytical column, this will lead to additional band broadening. Hence, a compromise between sufficient retention in the precolumn and band broadening in the analytical column is needed. Styrene–divinylbenzene polymeric sorbent was found to be a good solution to this problem.[47–51] The particle size of the precolumn material should be the same as used in the analytical column; however, it is possible to use particles with a larger diameter when processing water samples with a high concentration of suspended solids to prevent problems with high back-pressure.

The need for continuous monitoring of water quality initiated the development of rapid screening on-line SPE/HPLC methods with diode-array ultraviolet detection (DAD/UV). The use of on-line systems for this purpose allowed rapid access to information on water quality and a relatively high frequency of sampling. Within the framework of an international Rhine Basin Program, an on-line SPE/LC/DAD/UV system capable of screening for a large range of pesticides with different polarities was tested for more than 50 pesticides[48] and then was fully automated and validated for the determination of 27 pesticides.[51] Similar on-line systems for broad-range determinations of pesticides in water were developed using membrane extraction disks as packing

of the precolumn.[52,53] An interesting option in this kind of application is the use of a bifunctional membrane extraction disk cartridge containing both C_{18} and cation-exchange disks.[54] The combination of SPE materials having different functionalities allowed the simultaneous determination of basic, neutral and acidic pesticides. On-line systems can be used also for preconcentration of more ''problematic'' compounds. To improve the preconcentration efficiency for pyrethroid insecticides having both hydrophobic and hydrophilic behavior, both breakthrough on the precolumn and adsorption on the inner walls and surfaces were prevented by adding a neutral surfactant to the aqueous sample. The developed method involving micelle-mediated preconcentration was tested for surface water analysis and the whole system was found to be robust and to give good and repeatable results.[55]

To increase the functionality of the on-line SPE system, a more sophisticated scheme than that displayed in Figure 2 can be designed. For the preconcentration and fractionation of organic pollutants in industrial effluents, Nielen et al.[56] used three precolumns in series (C_{18}, PRP-1 and a cation exchanger) and eluted all precolumns separately. Brouwer et al.[57] connected two polymer (PLRP-S) precolumns in series and the outlet of each precolumn was directed on-line to a separate PLRP-S analytical column. While the first precolumn was operated in the reversed-phase mode, trace enrichment in the second precolumn was based on an ion-pairing mechanism. This allowed the preconcentration of acidic and basic compounds within one analysis. The same authors developed a similar system with one HPLC analytical column. In this system, they used as precolumns specially designed holders packed with membrane extraction disks.[58] A top state-of-the-art level in the development of sophisticated on-line systems was achieved by Slobodník et al.,[59] who developed an integrated system combining on-line SPE with LC and GC separation and detection by mass spectrometry (MS). The trace enrichment procedure was automated by a Prospekt cartridge-exchange/solvent-selection/valve-switching unit. After loading of the water sample, the precolumn was eluted on-line in two subsequent runs, first on to the gas chromatography/mass spectrometry (GC/MS) system and, next, onto the LC/DAD/UV/MS system using a particle beam interface (Figure 3). With this system, GC/MS, LC/DAD/UV and liquid chromatography/mass spectrometry (LC/MS) data on the same water sample could be obtained within 3 h, providing a large amount of structural information on unknown organic compounds present in the sample.

On-line coupling of SPE to GC is more complicated than to HPLC owing to problems with the liquid-phase/gas-phase interface. Even though this approach was already applied to the determination of pesticides in water in the 1980s,[60] more frequent use of on-line SPE/GC can be observed in the 1990s. An automatic SPE sample

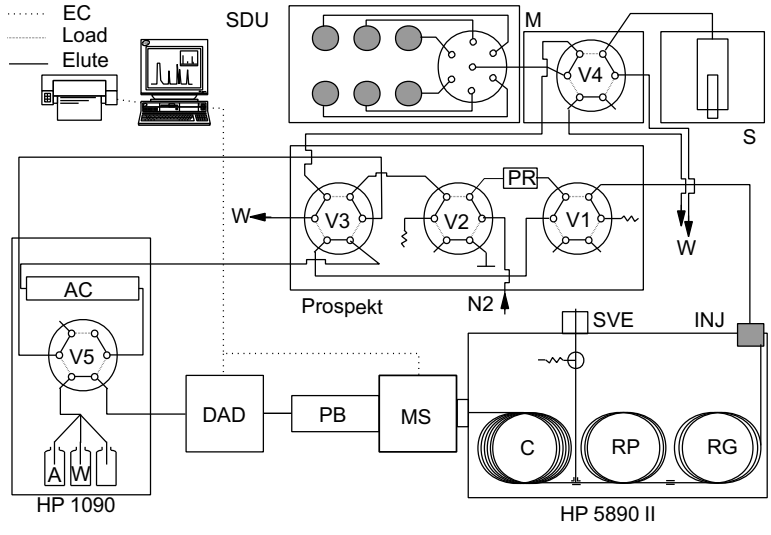

Figure 3 Schematic representation of the Multianalysis system which combines SPE/GC/MS and SPE/LC/DAD/PB/MS in one set-up. HP 1090 = liquid chromatograph; HP 5890 II = gas chromatograph; MS = mass spectrometer; PB = particle beam interface; DAD/UV = diode-array detector; Prospekt = automated valve-switching, solvent selection and cartridge-exchange unit; SDU = solvent delivery unit of Prospekt equipped with six-port solvent selection valve, pulse damper and LC pump; M = MUST automatic six-port switching valve unit; S = syringe pump; PR = precolumn packed with PLRP-S material; AC = LC analytical column; C = GC analytical column; RP = retaining precolumn; RG = retention gap; SVE = solvent vapor exit; INJ = on-column injector; N_2 = nitrogen; W = waste; V1–5 = six-port switching valves; EC = electronic connections, load/elute = positions of automatic six-port switching valves, computer and printer. (Reprinted from J. Slobodník, A.C. Hogenboom, A.J.H. Louter, U.A.Th. Brinkman, *J. Chromatogr. A*, **730**, 353–371 (1996),[59] Copyright 1996, with permission from Elsevier Science.)

preparation system was coupled on-line to capillary GC/ECD by means of a loop-type interface equipped with a solvent vapor exit.[61] Using SPE cartridges packed with silica for cleanup of the hexane extract of a water sample, it was possible to analyze concentrated surface water extracts for a group of organochlorine and pyrethroid insecticides present in the water at parts per trillion levels. This approach can be considered as a combination of LLE and SPE. If, in an SPE/GC system, the SPE cartridge is used for the direct accumulation of analytes from water, drying of the eluent prior to its injection into the GC system is desirable. For this purpose a small cartridge containing a suitable drying agent can be inserted between the SPE column and the GC system.[62] Introduction of programmable temperature vaporization (PTV) injectors in GC permitted the use of a GC liner packed with a solid support as an alternative to a SPE column. Thus, large volumes of water samples (up to 1 mL) can be introduced into a GC injector, water is separated in the liner and, after switching of a valve, compounds are thermally desorbed in the splitless mode by heating the PTV unit into the GC column.[63] The DLs obtained with such systems are in the sub-parts per billion range.

The application of a sophisticated cleanup procedure is not the only possibility for achieving a high selectivity. An alternative way is the use of a highly selective solid phase such as ion exchangers, metal-loaded sorbents or immunosorbents. The procedures utilizing these materials are rather complicated and require a skilled operator, but their efficiency and suitability for special target analyses have been demonstrated by different authors.[64–66]

A new development in the use of a solid phase for the preconcentration of pesticides from water is SPME, introduced and thoroughly developed by Pawliszyn et al.[67] SPME is based on a fiber coated with the stationary phase (e.g. polysiloxane or polyacrylate polymers) that is used for the direct extraction of organic compounds from water by dipping the fiber into the water sample. After sorption, the fiber is transferred into the heated injector of a gas chromatograph where the organic compounds are thermally desorbed from the stationary phase. This technique is becoming popular, as it is fast, sensitive, inexpensive, portable and solvent-free. Numerous applications of SPME in water analysis were reviewed by Eisert and Levsen[68] and SPME was found to be very attractive for routine use even though it still represents an exciting field of research.

3.3 Other Sample-handling Methods

Apart from LLE and SPE, only a few techniques based mostly on membrane and electromigration processes have received the attention of environmental analysts for handling of aqueous samples.

A supported liquid membrane (SLM) is typically constructed by immobilizing an immiscible liquid within the pores of a microporous film. In principle, the immobilized liquid can be aqueous or organic and various materials such as PTFE, cellulose acetate and polypropylene can be used as a support. The sample preparation set-up can be coupled on-line to the chromatograph. The use of closed systems minimizes the risk of contamination and losses during sample-handling. Since the volumes of organic solvent used in such a system are very small ($<500\,\mu L$) and the same solvent volume is used for a large number of samples, it is possible to use solvents which are expensive and also considered hazardous in batch extractions. Dihexyl ether was used in a liquid membrane applied for sample preparation for the determination of sulfonylurea herbicides in aqueous samples.[69] The technique studied in this work utilized extraction and back-extraction in an automated flow system and was coupled on-line to an LC system. The extractor unit consisted of an immobilized liquid membrane, separating two aqueous phases. From the acidified donor phase the analytes were extracted into the organic solvent of the membrane. After traversing the membrane, they were back-extracted into an alkaline/neutral aqueous acceptor phase. The use of the precolumn technique made it possible to inject the total amount of the extracted analytes. A liquid membrane device was also used for continuous sampling from a water stream in an integrated field sampling method for MCPA. MCPA was transferred as an uncharged species from a donor phase to a stagnant acceptor phase where it was trapped in a buffer, permitting a high degree of enrichment.[70] The final determination was made by HPLC utilizing a precolumn instead of a sample loop in the injector. The method has been found to give values in good agreement with a technique based on batch extraction and GC with a DL of ca. 0.03 ppb.

Membrane-based techniques can be efficient not only for trace enrichment purposes but also for the removal of interferences. The potential of the on-line combination of dialysis for the removal of interfering humic substances from environmental samples and trace enrichment on a precolumn packed with C_{18} for the determination of polar pesticides by column LC with UV detection was investigated by van de Merbel et al.[71] using six phenylurea herbicides as model compounds. Using a planar dialysis membrane with a molecular weight cut-off value of 3.5 kDa and a sample of 1.2 mL, interfering humic substances were sufficiently removed to allow the automated determination of selected phenylureas at the 1 ppb level. The transport of ionizable compounds through membranes can be supported by application of a voltage. Basic considerations for analyte enrichment and recovery obtainable by electrodialysis as a sample treatment

method were presented by Debets et al.[72] A practical application of on-line electrodialysis for the treatment of natural waters prior to the column LC determination of polar pollutants was described by Groenewegen et al.[73] Another approach to sample-handling of water samples containing ionizable compounds is the use of electromigration techniques. On-line coupling of isotachophoretic pretreatment to capillary zone electrophoresis (CZE) allowed the determination of paraquat and diquat at nanomoles per liter levels in tap and surface water.[74] Isotachophoresis was also presented as a selective sample pretreatment technique for ionic substances prior to HPLC analysis.[75]

Supercritical fluid extraction (SFE) has been demonstrated to be a good method for preparation of solid samples. However, there is only a very limited number of papers dealing with direct SFE of aqueous samples and owing to experimental problems this approach does not seem very promising for the future.[76]

4 CONCLUSIONS

Sample preparation, for reasons mentioned in the Introduction, will remain a fertile area of development in the future. Reduction of time and costs needed for sample handling and reduction of the consumption of toxic solvents will be the leading factors supporting any progress in this area. Automation will save time and reduce errors. New approaches and new techniques will make sample handling safer and more effective. There will be the simultaneous development of both simple and fast preconcentration schemes for routine use in rapid standard methods and sophisticated procedures coping with the continuous development of modern polar pesticides. The growing interest in ecotoxicological assessment and biomonitoring will influence the prevailing ''chemical'' character of the preparation of water samples. Also, obviously, research in environmental pesticide analysis will be tightly bound to regulatory policies in environmental protection and any pushing of regulatory levels to lower values or introduction of new parameters to the ''Black Lists'' will further stimulate the development of sample-handling techniques, especially for polar hydrophilic analytes.

ABBREVIATIONS AND ACRONYMS

C_{18}	Octadecyl-bonded Silica
CZE	Capillary Zone Electrophoresis
DAD/UV	Diode-array Ultraviolet Detection
DDE	1,1-Dichloro-2,2-bis[*p*-chlorophenyl]-ethylene
DDT	1,1,1-Trichloro-2,2-bis[*p*-chlorophenyl]-ethane
DL	Detection Limit
ECD	Electron-capture Detection
GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/MS	Gas Chromatography/Mass Spectrometry
HCB	Hexachlorobenzene
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LLE	Liquid–Liquid Extraction
MS	Mass Spectrometry
NPD	Nitrogen–Phosphorus Detector
PTFE	Polytetrafluoroethylene
PTV	Programmable Temperature Vaporization
SFE	Supercritical Fluid Extraction
SLM	Supported Liquid Membrane
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicide Residues in Biota, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Pesticides (New Generation) and Related Compounds, Analysis of

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

REFERENCES

1. R.E. Majors, 'An Overview of Sample Preparation', *LC/GC Int.*, **4**(2), 10–14 (1991).
2. D. Chapman (ed.), *Water Quality Assessments*, Chapman and Hall, London, 1992.
3. *General Principles of Sampling Waters and Associated Materials*, 2nd edition, HMSO Books, London, 1996.
4. *Monitoring Water Quality in the Future*, Ministry of Housing, Spatial Planning and the Environment, The Hague, Vols 1–5, 1995.
5. D.J. Munch, C.P. Frebis, 'Analyte Stability Studies Conducted During the National Pesticide Survey', *Environ. Sci. Technol.*, **26**(5), 921–925 (1992).
6. D. Chapman (ed.), *Water Quality Assessments*, Chapman and Hall, London, 466, 1992.
7. W.A. Battaglin, L.E. Hay, 'Effects of Sampling Strategies on Estimates of Annual Mean Herbicide Concentrations in Midwestern Rivers', *Environ. Sci. Technol.*, **30**(3), 889–896 (1996).
8. D.R. Green, D. LePape, 'Stability of Hydrocarbon Samples on Solid Phase Extraction Columns', *Anal. Chem.*, **59**, 699–703 (1987).
9. C. Crescenzi, A. DiCorcia, M. Madbouly, R. Samperi, 'Pesticides Stability Studies Upon Storage in a Graphitized Carbon Black Extraction Cartridge', *Environ. Sci. Technol.*, **29**(9), 2185–2190 (1995).
10. S. Lacorté, N. Ehresmann, D. Barceló, 'Stability of Organophosphorus Pesticides on Disposable Solid Phase Extraction Precolumns', *Environ. Sci. Technol.*, **29**, 2834–2841 (1995).
11. I. Liška, K. Bíliková, 'Stability of Polar Pesticides on Disposable Solid Phase Extraction Precolumns', *J. Chromatogr. A*, **795**, 61–69 (1998).
12. E. Martinez, D. Barceló, 'The Stability of Selected Herbicides Preconcentrated from Estuarine River Waters on Solid Phase Extraction Disks', *Chromatographia*, **42**(1/2), 72–76 (1996).
13. D. Barceló, M.C. Hennion, *Trace Determination of Pesticides and Their Degradation Products in Water*, Elsevier, Amsterdam, 1997.
14. D. Barceló, 'Occurrence, Handling and Chromatographic Determination of Pesticides in Aquatic Environment', *Analyst*, **116**, 681–689 (1991).
15. M.C. Hennion, 'Sample-handling Strategies for the Analysis of Non-volatile Organic Compounds from Environmental Water Samples', *Trends Anal. Chem.*, **10**(10), 317–323 (1991).
16. I. Liška, J. Slobodník, 'Comparison of Gas and Liquid Chromatography for Analysing Polar Pesticides in Water Samples', *J. Chromatogr. A*, **733**, 235–258 (1996).
17. F.I. Onuska, 'Pesticide Residue Analysis by Open Tubular Column Gas Chromatography: Trials, Tribulations and Trends', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **7**, 660–670 (1984).
18. I. Liška, D. Barceló, M. Grasserbauer, 'Strategy for the Screening of Organic Pollutants in a River Basin – an Overview of the Nitra River Monitoring Programme', *Trends Anal. Chem.*, **15**(8), 326–334 (1996).
19. W.A. Telliard, 'Broad-range Methods for Determination of Pollutants in Wastewater', *J. Chromatogr. Sci.*, **28**(9), 453–459 (1990).
20. V. Lopez-Avila, R. Wesselman, K. Edgell, 'Gas Chromatographic–Electron Capture Detection Method for Determination of 29 Organochlorine Pesticides in Finished Drinking Water: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **73**(2), 276–286 (1990).
21. K.W. Edgell, E.J. Erb, R.J. Wesselman, J.E. Longbottom, 'Gas Chromatographic–Electron Capture Detection Method for Determination of Chlorinated Acids in Water: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **76**(5), 1098–1112 (1993).
22. G.D. Foster, P.F. Rogerson, 'Enhanced Preconcentration of Pesticides from Water Using the Goulden Large-sample Extractor', *Int. J. Environ. Anal. Chem.*, **41**, 105–117 (1990).
23. E.A. Hogendoorn, C.E. Goewie, P. van Zoonen, 'Application of HPLC Column Switching in Pesticide Residue Analysis', *Fresenius' J. Anal. Chem.*, **339**, 348–356 (1991).
24. E.A. Hoogendoorn, E. Dijkman, S.M. Gort, R. Hoogerbrugge, P. van Zoonen, U.A.Th. Brinkman, 'General Strategy for Multiresidue Analysis of Polar Pesticides in Groundwater Using Coupled Column Reversed-phase LC and Step-gradient Elution', *J. Chromatogr. Sci.*, **31**(10), 433–439 (1993).
25. I. Liška, J. Krupčík, P.A. Leclercq, 'The Use of Solid Sorbents for Direct Accumulation of Organic Compounds from Water Matrices – a Review of Solid Phase Extraction Techniques', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **12**, 577–589 (1989).

26. E. Chladek, R.S. Marano, 'Use of Bonded Phase Silica Sorbents for the Sampling of Priority Pollutants in Wastewaters', *J. Chromatogr. Sci.*, **22**, 313–320 (1984).
27. S. Fingler, V. Drevenkar, Z. Vasilic, 'A Comparative Study of Trace Enrichment of Chlorophenols in Water by Extraction with C-6 and C-8 Alkanes and by C-18 Reversed-phase Adsorption', *Mikrochim. Acta*, **II**, 163–175 (1987).
28. J.J. Richard, G.A. Junk, 'Solid Phase Versus Solvent Extraction of Pesticides from Water', *Mikrochim. Acta*, **I**, 387–394 (1986).
29. S. Chiron, A. Fernandez Alba, D. Barceló, 'Comparison of On-line Solid Phase Extraction for Monitoring Selected Pesticides in Environmental Waters', *Environ. Sci. Technol.*, **27**(12), 2352–2359 (1993).
30. J.C. Moltó, C. Albelda, G. Font, J. Manes, 'Solid Phase Extraction of Organochlorine Pesticides from Water Samples', *Int. J. Environ. Anal. Chem.*, **41**, 21–26 (1990).
31. G.H. Tan, 'Comparison of Solvent Extraction and Solid Phase Extraction for the Determination of Organochlorine Pesticide Residues in Water', *Analyst*, **117**(7), 1129–1132 (1992).
32. R.D. McDowall, J.C. Pearce, G.S. Murkitt, 'Liquid–Solid Sample Preparation in Drug Analysis', *J. Pharm. Biomed. Anal.*, **4**, 3–21 (1986).
33. Z. Voznakova, M. Popl, M. Kovar, 'Sorbents for the Preconcentration of Herbicides from Water', *Scientific Papers of the VSCHT*, H 19, 85–96, Statni Pedagogicke Nakladatelstvi, Prague, 1984.
34. I. Liška, 'On-line Versus Off-line Solid Phase Extraction in the Determination of Organic Contaminants in Water, Advantages and Limitations', *J. Chromatogr. A*, **655**, 163–176 (1993).
35. DIN 38 407, *Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung, Gemeinsam erf aßbare Stoffgruppen (Gruppe F)*, parts 6, 12, 14, Deutsches Institut fur Normung, Berlin, 1993–95.
36. E.B. Asafu-Adjaye, P.W. Kohler, S.Y. Su, 'Retention Behavior and Solvent Optimization of Organochlorine Pesticides in Solid Phase Extraction', *Chromatographia*, **21**(8), 435–438 (1986).
37. M.J.M. Wells, A.J. Rosano, E.C. Roberts, 'Solid Phase Extraction for Toxicity Reduction Evaluations of Industrial Wastewater Effluents', *Anal. Chim. Acta*, **236**, 131–139 (1990).
38. E. Durhan, M. Lukasewycz, S. Baker, 'Alternatives to Methanol–Water Elution of Solid Phase Extraction Columns for the Fractionation of High Log K_{ow} Organic Compounds in Aqueous Environmental Samples', *J. Chromatogr.*, **629**, 67–74 (1993).
39. I. Tolosa, J.W. Readman, L.D. Mee, 'Comparison of the Performance of Solid Phase Extraction Techniques in Recovering Organophosphorus and Organochlorine Compounds from Water', *J. Chromatogr. A*, **725**, 93–106 (1996).
40. A. DiCorcia, C. Crescenzi, E. Guerriero, R. Samperi, 'Ultratrace Determination of Atrazine and its Six Major Degradation Products in Water by Solid Phase Extraction and Liquid Chromatography/Electrospray/Mass Spectrometry', *Environ. Sci. Technol.*, **31**, 1658–1663 (1997).
41. A. DiCorcia, M. Marchetti, R. Samperi, 'Extraction and Isolation of Phenoxy Acid Herbicides in Environmental Waters Using Two Adsorbents in One Minicartridge', *Anal. Chem.*, **61**, 1363–1367 (1989).
42. A. DiCorcia, M. Marchetti, 'Method Development for Monitoring Pesticides in Environmental Waters: Liquid–Solid Extraction Followed by Liquid Chromatography', *Environ. Sci. Technol.*, **26**, 66–74 (1992).
43. D. Wells, G. Goken, 'Chemical Separation Technology Reduces Hazardous Solvent Usage', *Pollut. Preve.*, **2**, 42–44 (1992).
44. T.S. Thompson, B.D. Miller, 'Use of Solid Phase Extraction Disks for the GC/MS Analysis of Acidic and Neutral Herbicides in Drinking Water', *Chemosphere*, **36**(14), 2867–2878 (1998).
45. J.S. Salau, R. Alonso, G. Batlló, D. Barceló, 'Application of Solid Phase Disk Extraction Followed by Gas and Liquid Chromatography for the Simultaneous Determination of the Fungicides: Captan, Captafol, Carbendazim, Chlorothalonil, Ethirimol, Folpet, Metalaxyl and Vinclozolin in Environmental Waters', *Anal. Chim. Acta*, **293**, 109–117 (1994).
46. U.A.Th. Brinkman, 'On-line Sample Treatment for or via Column Liquid Chromatography', *J. Chromatogr. A*, **665**, 217–231 (1994).
47. I. Liška, E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'Trace Enrichment of Polar Pollutants from Water and the Influence of Humic Substances', *Chromatographia*, **37**(1/2), 13–19 (1993).
48. I. Liška, E.R. Brouwer, A.G.L. Ostheimer, H. Lingeman, U.A.Th. Brinkman, R.B. Geerdink, W.H. Mulder, 'Rapid Screening of a Large Group of Polar Pesticides in River Water by On-line Trace Enrichment and Column Liquid Chromatography', *Int. J. Environ. Anal. Chem.*, **47**, 267–291 (1992).
49. P. Subra, M.C. Hennion, R. Rosset, R.W. Frei, 'Recovery of Organic Compounds from Large Volume Aqueous Samples Using On-line Liquid Chromatographic Preconcentration Techniques', *J. Chromatogr.*, **456**, 121–141 (1988).
50. P. Subra, M.C. Hennion, R. Rosset, R.W. Frei, 'Applications of On-line Precolumn Technology to Water Quality Control', *Int. J. Environ. Anal. Chem.*, **37**, 45–62 (1989).
51. J. Slobodník, M.G.M. Groenewegen, E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'Fully Automated Multi-residue Method for Trace Level Monitoring of

Polar Pesticides by Liquid Chromatography', *J. Chromatogr.*, **642**, 359–370 (1993).
52. S. Chiron, D. Barceló, 'Determination of Pesticides in Drinking Water by On-line Solid Phase Disk Extraction Followed by Various Liquid Chromatographic Systems', *J. Chromatogr.*, **645**, 125–134 (1993).
53. S. Chiron, S. Dupas, P. Scribe, D. Barceló, 'Application of On-line Solid Phase Extraction Followed by Liquid Chromatography/Thermospray Mass Spectrometry to the Determination of Pesticides in Environmental Waters', *J. Chromatogr. A*, **665**, 295–305 (1994).
54. E.H.R. van der Wal, E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'On-line Trace Enrichment–Column Liquid Chromatography of Polar Pollutants in Surface Water Using Bifunctional Membrane Based Extraction Disks', *Chromatographia*, **39**(3/4), 239–245 (1994).
55. E.R. Brouwer, E.A. Struys, J.J. Vreuls, U.A.Th. Brinkman, 'Automated Determination of Pyrethroid Insecticides in Surface Water by Column Liquid Chromatography with Diode-array UV Detection Using On-line Micelle Mediated Sample Preparation', *Fresenius' J. Anal. Chem.*, **350**, 487–495 (1994).
56. M.W.F. Nielen, U.A.Th. Brinkman, R.W. Frei, 'Industrial Wastewater Analysis by Liquid Chromatography with Precolumn Technology and Diode-array Detection', *Anal. Chem.*, **57**, 806–810 (1985).
57. E.R. Brouwer, I. Liška, R.B. Geerdink, P.C.M. Frintrop, W.H. Mulder, H. Lingeman, U.A.Th. Brinkman, 'Determination of Polar Pollutants in Water Using an On-line Liquid Chromatographic Preconcentration System', *Chromatographia*, **32**(9/10), 445–452 (1991).
58. E.R. Brouwer, D.J. van Iperen, I. Liška, H. Lingeman, U.A.Th. Brinkman, 'Liquid Chromatographic Determination of Polar Pollutants in Surface Water Using Membrane Extraction Disks for On-line Trace Enrichment', *Int. J. Environ. Anal. Chem.*, **47**, 257–266 (1992).
59. J. Slobodník, A.C. Hogenboom, A.J.H. Louter, U.A.Th. Brinkman, 'Integrated System for On-line Gas and Liquid Chromatography with a Single Mass Spectrometric Detector for the Automated Analysis of Environmental Samples', *J. Chromatogr. A*, **730**, 353–371 (1996).
60. E. Noroozian, F.A. Maris, M.W.F. Nielen, R.W. Frei, G. de Jong, U.A.Th. Brinkman, 'Liquid Chromatographic Trace Enrichment with On-line Capillary Gas Chromatography for the Determination of Organic Pollutants in Aqueous Samples', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **10**, 17–24 (1987).
61. G.R. van der Hoff, S.M. Gort, R.A. Baumann, P. van Zoonen, U.A.Th. Brinkman, 'Cleanup of Some Organochlorine and Pyrethroid Insecticides by Automated Solid Phase Extraction Cartridges Coupled to Capillary GC/ECD', *J. High Resolut. Chromatogr.*, **14**, 465–470 (1991).
62. Y. Picó, J.J. Vreuls, R.T. Ghijsen, U.A.Th. Brinkman, 'Drying Agents for Water-free Introduction of Desorption Solvent into a GC After On-line SPE of Aqueous Samples', *Chromatographia*, **38**(7/8), 461–469 (1994).
63. H.G.J. Mol, H.G.M. Janssen, C.A. Cramers, U.A.Th. Brinkman, 'Use of a Temperature Programmed Injector with a Packed Liner for Direct Water Analysis and On-line Reversed Phase LC/GC', *J. High Resolut. Chromatogr.*, **16**, 459–463 (1993).
64. V. Pichon, L. Chen, M.C. Hennion, 'On-line Preconcentration and Liquid Chromatographic Analysis of Phenylurea Pesticides in Environmental Water Using a Silica-based Immunosorbent', *Anal. Chim. Acta*, **311**, 429–436 (1995).
65. V. Pichon, M.C. Hennion, 'Comparison of On-line Enrichment Based on Ion-pair and Cation-exchange Liquid Chromatography for the Trace-level Determination of 3-Amino-1,2,4-triazole (Aminotriazole) in Water', *Anal. Chim. Acta*, **284**, 317–326 (1993).
66. C.E. Goewie, P. Kwakman, R.W. Frei, U.A.Th. Brinkman, W. Maasfeld, T. Seshadri, A. Kettrup, 'Precolumn Technology in High Performance Liquid Chromatography for the Determination of Phenylurea Herbicides in Water in the Presence of Their Anilines', *J. Chromatogr.*, **284**, 73–86 (1984).
67. C.L. Arthur, D.W. Potter, K.D. Buchholz, S. Motlagh, J. Pawliszyn, 'Solid Phase Microextraction for the Direct Analysis of Water: Theory and Practice', *LC/GC Int.*, **5**(10), 8–14 (1992).
68. R. Eisert, K. Levsen, 'Solid Phase Microextraction Coupled to Gas Chromatography: a New Method for the Analysis of Organics in Water', *J. Chromatogr. A*, **733**, 143–157 (1996).
69. G. Nilvé, R. Stebbins, ' Automated Sample Preparation Using Supported Liquid Membranes for Liquid Chromatographic Determination of Sulfonylurea Herbicides', *Chromatographia*, **32**(5/6), 269–277 (1991).
70. L. Mathiasson, G. Nilvé, B. Ulén, 'A Liquid Membrane Enrichment Technique for Integrating Field Sampling in Water Applied to MCPA', *Int. J. Environ. Anal. Chem.*, **45**, 117–125 (1991).
71. N.C. van de Merbel, F.M. Lagerwerf, H. Lingeman, U.A.Th. Brinkman, 'On-line Dialysis–Column Liquid Chromatography for the Determination of Polar Pesticides in Water Samples Containing Humic Substances', *Int. J. Environ. Anal. Chem.*, **54**, 105–118 (1994).
72. A.J.J. Debets, P.v. Wier, K.P. Hupe, U.A.Th. Brinkman, W.Th. Kok, 'Theoretical Models for Electrodialytic Sample Treatment in Trace Analysis by Liquid Chromatography', *Chromatographia*, **39**(7/8), 460–468 (1994).
73. M.G.M. Groenewegen, N.C. van de Merbel, J. Slobodník, H. Lingeman, U.A.Th. Brinkman, 'Automated Determination of Weakly Acidic and Basic Pollutants in

Surface Water by On-line Electrodialysis Sample Treatment and Column Liquid Chromatography', *Analyst*, **119**, 1753–1758 (1994).

74. D. Kaniansky, F. Ivanyi, F.I. Onuska, 'On-line Isotachophoretic Sample Pretreatment in Ultratrace Determination of Paraquat and Diquat in Water by Capillary Zone Electrophoresis', *Anal. Chem.*, **66**, 1817–1824 (1994).
75. P.J.M. Hendriks, H.A. Claessens, T.H.M. Noij, F.M. Everaerts, C.A. Cramers, 'On-line Isotachophoresis. A Selective Sample Pretreatment Prior to Column Liquid Chromatographic Analysis', *Chromatographia*, **33**(11/12), 539–544 (1992).
76. V. Janda, M. Mikešová, J. Vejrosta, 'Direct Supercritical Fluid Extraction of Water-based Matrices', *J. Chromatogr. A*, **733**, 35–40 (1996).

Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of

Thomas Heberer
Institute of Food Chemistry, Technical University of Berlin, Germany

1 Introduction **1**
2 Agricultural Use and Importance **2**
3 Instrumental Analysis **5**
3.1 Capillary Gas Chromatography 5
3.2 High-performance Liquid Chromatography 23
3.3 Other Techniques 24
4 Environmental Applications **27**
4.1 Analysis of Water Samples 28
4.2 Analysis of Soil Samples 31
4.3 Interferences 33
4.4 Some Practical Applications 35
Acknowledgments **37**
Abbreviations and Acronyms **37**
Related Articles **38**
References **39**

Owing to their widespread use in agriculture and urban weed control and because of their high mobility in the subsoil, polar pesticides such as phenoxy acids and other acidic pesticides have been especially considered as a potential source of groundwater contamination. In Europe, pesticide residues tolerances in drinking water were set by the European Union Commission to 100 ng L^{-1} for an individual compound and 500 ng L^{-1} for the sum of pesticide residues. These maximum tolerances constitute a real challenge to analysts working in the field of pesticide residue monitoring. In the following article, analytical methods for the multiresidue analysis of phenoxy acids and other acidic pesticides in environmental samples at trace level concentrations are presented. Analytical methods using capillary gas chromatography (GC), high-performance liquid chromatography (HPLC), capillary zone electrophoresis (CZE) and automated multiple development thin-layer chromatography (AMDTLC) are described. GC methods are most often used in routine analysis of acidic pesticides owing to their unrivaled advantages in separation power. GC methods can be used in multiresidue analysis of up to more than 50 analytes at trace level concentrations. Owing to their high polarity and low volatility, acidic pesticides are not directly amenable to GC analysis, thus suitable derivatization methods are required. The applicability and drawbacks of several derivatization methods are described in detail. GC can be used with various detection methods such as electron capture detection (ECD), nitrogen–phosphorus detection (NPD), atomic emission detection (AED) or mass spectrometry (MS). GC/MS (gas chromatography/mass spectrometry) or GC/MS/MS (tandem mass spectrometry) detection provides the highest possible level of confidence independent of the complexity of the environmental matrix from water or even soil samples. The highest selectivity and sensitivity is achieved in selected ion monitoring (SIM) or in selected reaction monitoring (SRM) mode. Progress in coupling of HPLC to MS has improved the possibilities of identification and confirmation of analytes at trace level concentrations using HPLC methods and new promising analytical approaches such as HPLC coupled to atmospheric pressure ionization (API) MS/MS will gain much importance in the future. Enantioselective separation of different chiral isomers of phenoxy acid pesticides can be achieved applying CZE or by applying GC or HPLC with special chiral phases.

In water analysis, conventional liquid–liquid extraction (LLE) has mostly been replaced by solid-phase extraction (SPE) methods to extract and enrich the analytes from the samples. Two standard operating procedures (SOPs) are presented as examples for the analysis of phenoxy acids and other acidic pesticides in environmental samples (water and soil). Detection limits down to the low nanogram per liter level or down to the low microgram per kilogram level can be achieved for water samples or soil samples, respectively. Several examples for the environmental analysis of actual samples show the performance and sensitivity of today's trace level multiresidue analysis.

1 INTRODUCTION

The occurrence of polar pesticides in the environment has become a subject of increased public concern. Owing to their widespread use in agriculture and urban weed control and because of their high mobility in the subsoil, polar pesticides such as phenoxy acids and other acidic pesticides have been especially considered as a potential source of groundwater contamination. Consequently, in many environmental investigations these pesticides are analyzed to check their concentration against the maximum tolerances of pesticide residues in drinking water.

In Europe, these tolerances were set by the European Union Commission[1] to 100 ng L^{-1} for an individual compound and 500 ng L^{-1} for the sum of pesticide

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

residues. These maximum tolerances constitute a real challenge to analysts working in the field of pesticide residue monitoring. Analytical routine methods have to guarantee high accuracy and reliability at trace level concentrations independent of the amount and composition of the environmental matrix. Because of their polar structures, extraction from environmental samples and their analysis are much more difficult for acidic pesticides than for nonpolar pesticides such as organochlorine or triazine pesticides such as DDT (2,2-bis(chlorophenyl)-1,1,1-trichloroethane) or atrazine. To lower the costs of routine analysis it is also desirable to determine as many analytes as possible at one time in one procedure applying only a few so-called multimethods.

In this article, analytical procedures are presented which fulfill all stated requirements enabling multiresidue analysis of up to 40 acidic pesticides at trace level concentrations in water or soil.

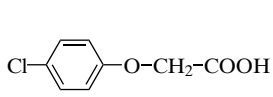

4-CPA
(1)

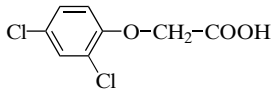

2,4-D
(2)

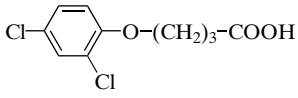

2,4-DB
(3)

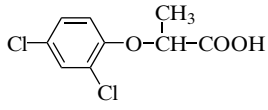

Dichlorprop
(4)

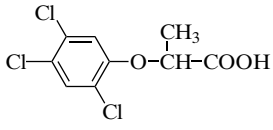

Fenoprop
(5)

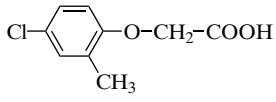

MCPA
(6)

2 AGRICULTURAL USE AND IMPORTANCE

Phenoxy acid pesticides, also called phenoxy alkanoic acid pesticides, are selective herbicides developed in the 1940s in the course of attempts to synthesize analogs of the auxin, β-indoleacetic acid, which had been recognized as a natural plant growth regulator.[2] The common mode of action of auxin-type herbicides is to mimic natural auxin, β-indoleacetic acid, thereby producing an abnormal lethal growth. On the molecular scale, they influence the levels of RNA and DNA polymerase and enzymes that regulate

Table 1 Common names, IUPAC names, CAS numbers and properties of acidic pesticides mentioned in this chapter

Common name (**structure no.**)	CAS no.	Chemical name (IUPAC)	Compound class/use
Acifluorfen **(20)**	50594-66-6	5-(2-Chloro-α,α,α-trifluoro-*p*-tolyloxy)-2-nitrobenzoic acid	Diphenyl ether herbicide, protoporphyrinogen oxidase inhibitor
Benazolin **(21)**	3813-05-6	4-Chloro-2,3-dihydro-2-oxo-1,3-benzothiazol-3-yl-acetic acid	Arylacetic acid herbicide, auxin type
Bentazone **(10)**	25057-89-0	3-Isopropyl-1*H*-2,1,3-benzothiadiazin-4(3*H*)-one 2,2-dioxide	Benzothiadiazinone herbicide, photosynthetic electron transport inhibitor
Bromoxynil **(13)**	1689-84-5	3,5-Dibromo-4-hydroxybenzonitrile	Hydroxybenzonitrile herbicide, photosynthetic and respiratory electron transport inhibitor
Chloramben **(22)**	133-90-4	3-Amino-2,5-dichlorbenzoic acid	Benzoic acid herbicide, auxin type
Chlorfenac **(23)**[a]	20782-58-5	(2,3,6-Trichlorophenyl)acetic acid	Aryloxyalkanoic acid herbicide, plant growth regulator
Chloroxynil **(12)**[a]	1891-95-8	3,5-Dichloro-4-hydroxybenzonitrile	Hydroxybenzonitrile herbicide, photosynthetic and respiratory electron transport inhibitor
4-CPA **(1)**	122-88-3	*p*-Chlorophenoxyacetic acid	Aryloxyalkanoic acid herbicide, plant growth regulator
Clopyralid **(24)**	1702-17-6	3,6-Dichloropyridine-2-carboxylic acid	Pyridinecarboxylic acid herbicide, auxin type
2,4-D **(2)**	94-75-7	(2,4-Dichlorophenoxy)acetic acid	Aryloxyalkanoic acid herbicide, auxin type
2,4-DB **(3)**	94-82-6	(2,4-Dichlorophenoxy)butyric acid	Aryloxyalkanoic acid herbicide, auxin type
Dicamba **(25)**	1918-00-9	3,6-Dichloro-*o*-anisic acid	Arenecarboxylic acid herbicide, auxin type
Dichlorophen **(17)**	97-23-4	4,4′-Dichloro-2,2′-methylenediphenol	Chlorophenol fungicide

Table 1 (*continued*)

Common name (**structure no.**)	CAS no.	Chemical name (IUPAC)	Compound class/use
Dichlorprop **(4)**	120-36-5	2-(2,4-Dichlorophenoxy)propionic acid	2-Aryloxyalkanoic acid herbicide, auxin type
Dinoseb **(15)**	88-85-7	2-*sec*-Butyl-4,6-dinitrophenol	Dinitrophenol herbicide, respiratory electron transport inhibitor
DNOC **(16)**	543-52-1	4,6-Dinitro-*o*-cresole	Dinitrophenyl insecticide and herbicide, respiratory electron transport inhibitor
Endothal **(26)**	145-73-3	7-Oxabicyclo[2,2,1]heptane-2,3-dicarboxylic acid	Cyclohexane-1,2-dicarboxylic acid herbicide
Fenoprop **(5)**[a]	93-72-1	2-(2,4,5-Trichlorophenoxy)propionic acid	2-Aryloxyalkanoic acid herbicide, auxin type
Flamprop **(27)**	58667-63-3	*N*-benzoyl-*N*-(3-chloro-4-fluorophenyl)-DL-alanine	Arylalanine herbicide, fatty acid synthesis inhibitor
Fluazifop **(28)**	69335-91-7	2-[4-(5-Trifluoromethyl-2-pyridyloxy) phenoxy]propionic acid	2-(4-Aryloxyphenoxy) alkanoic acid herbicide, fatty acid synthesis inhibitor
Flurenol **(29)**	467-69-6	9-Hydroxyfluorene-9-carboxylic acid	Fluorene-9-carboxylic acid herbicide, plant growth regulator
Fluroxypyr **(30)**	69377-81-7	4-Amino-3,5-dichloro-6-fluoro-2-pyridyloxyacetic acid	Pyridyloxyacetic acid herbicide, auxin type
Fuberidazole **(11)**	3878-19-1	2-(2-Furyl)benzimidazole	Benzimidazole fungicide
Haloxyfop **(31)**	69806-34-4	2-[4-(3-Chloro-5-trifluoromethyl-2-pyridyloxy)phenoxy]propionic acid	2-(4-Aryloxyphenoxy) alkanoic acid herbicide, fatty acid synthesis inhibitor
Imazapyr **(36)**	81334-34-1	2-(4-Isopropyl-4-methyl-5-oxo-2-imidazolin-2-yl)nicotinic acid	Imidazolinone herbicide, branched chain amino acid synthesis inhibitor
4-Indol-3-yl-butyric acid **(35)**	133-32-4	4-Indol-3-yl-butyric acid	4-Indol-3-yl-butyric acid plant growth regulator
Indol-3-yl-acetic acid **(34)**	87-51-4	β-Indoleacetic acid	Indol-3-yl-acetic acid plant hormone
Ioxynil **(14)**	1689-83-4	4-Hydroxy-3,5-di-iodobenzonitrile	Hydroxybenzonitrile herbicide, photosynthetic and respiratory electron transport inhibitor
MCPA **(6)**	94-74-6	(4-Chloro-2-methylphenoxy)acetic acid	Aryloxyalkanoic acid herbicide, auxin type
MCPB **(7)**	94-81-5	4-(4-Chloro-2-methylphenoxy)butanoic acid	Aryloxyalkanoic acid herbicide, auxin type
Mecoprop **(8)**	7085-19-0	2-(4-Chloro-2-methylphenoxy)propionic acid	Aryloxyalkanoic acid herbicide, auxin type
2-(1-Naphthyl)acetic acid **(32)**	86-87-3	2-(1-Naphthyl)acetic acid	Arylacetic acid plant growth regulator
PCP **(18)**	608-93-5	Pentachlorophenol	Phenolic herbicide, fungicide and insecticide
2-Phenylphenol **(19)**	90-43-7	Biphenyl-2-ol	Phenolic fungicide
Picloram **(33)**	1918-02-1	4-Amino-3,5,6-trichloropyridine-2-carboxylic acid	Pyridinecarboxylic acid herbicide, auxin type
Quinchlorac **(37)**	84087-01-4	3,7-Dichloroquinoline-8-carboxylic acid	Quinolinecarboxylic acid herbicide, auxin type
Quinmerac **(38)**	90717-03-6	7-Chloro-3-methylquinoline-8-carboxylic acid	Quinolinecarboxylic acid herbicide, auxin type
2,4,5-T **(9)**	93-76-5	(2,4,5-Trichlorophenoxy)acetic acid	Aryloxyalkanoic acid herbicide, auxin type
2,3,6-TBA **(39)**	50-31-7	2,3,6-Trichlorobenzoic acid	Benzoic acid herbicide, auxin type
Triclopyr **(40)**	55335-06-3	3,5,6-Trichloro-2-pyridyloxyacetic acid	Pyridyloxyacetic acid herbicide, auxin type

[a] compounds which are believed to be currently of little commercial interest.
4-CPA, *p*-chlorophenoxyacetic acid; 2,4-DB, (2,4-dichlorophenoxy)butyric acid; DNOC, 4,6-dinitro-*o*-cresol; MCPB, 2-(4-chloro-2-methylphenoxy) propionic acid; PCP, pentachlorophenol; 2,4,5-T, (2,4,5-trichlorophenoxy)acetic acid; 2,3,6-TBA, 2,3,6-trichlorobenzoic acid.

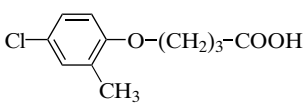
MCPB
(7)

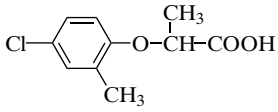
Mecoprop
(8)

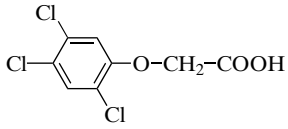
2,4,5-T
(9)

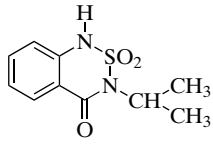
Bentazone
(10)

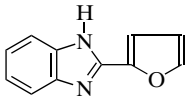
Fuberidazole
(11)

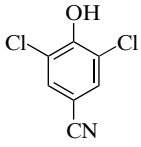
Chloroxynil
(12)

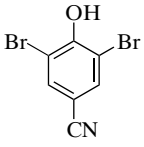
Bromoxynil
(13)

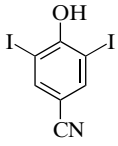
Ioxynil
(14)

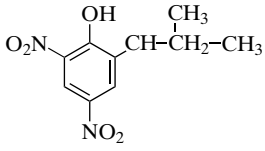
Dinoseb
(15)

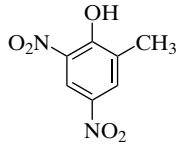
DNOC
(16)

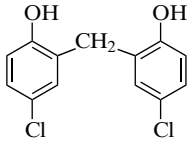
Dichlorophen
(17)

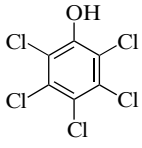
PCP
(18)

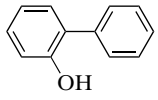
2-Phenylphenol
(19)

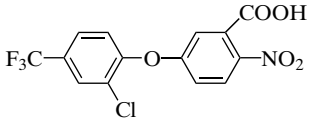
Acifluorfen
(20)

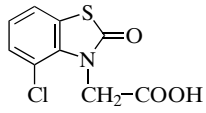
Benazolin
(21)

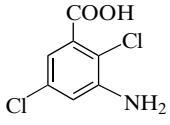
Chloramben
(22)

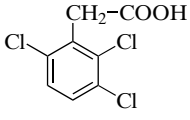
Chlorfenac
(23)

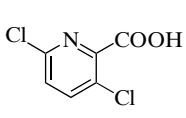
Clopyralid
(24)

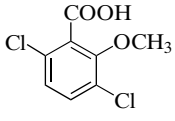
Dicamba
(25)

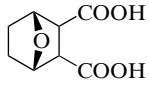
Endothal
(26)

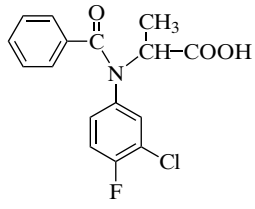
Flamprop
(27)

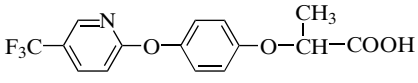
Fluazifop
(28)

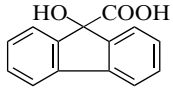
Flurenol
(29)

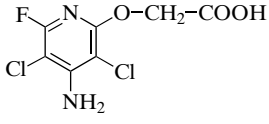
Fluroxypyr
(30)

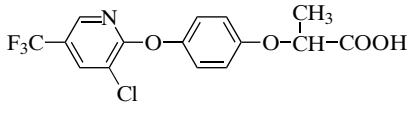
Haloxyfop
(31)

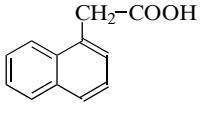
2-(1-Naphthyl)acetic acid
(32)

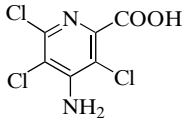
Picloram
(33)

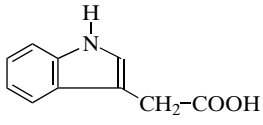
Indol-3-ylacetic acid
(34)

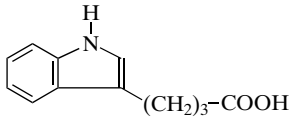
4-Indol-3-yl-butyric acid
(35)

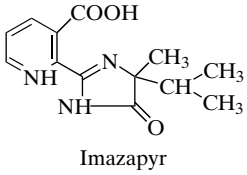
Imazapyr
(36)

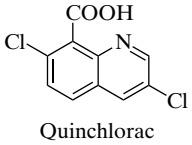
Quinchlorac
(37)

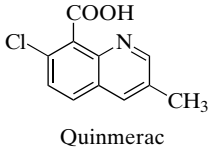
Quinmerac
(38)

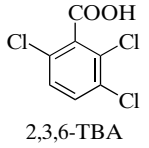
2,3,6-TBA
(39)

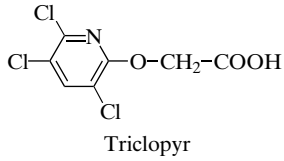
Triclopyr
(40)

normal growth and development processes.[2] Although phenoxy acids are the oldest group among the synthetic herbicides[3] they are still frequently used worldwide. In Germany, the phenoxy acids dichlorprop **(4)**, mecoprop **(8)**, 2,4-D ((2,4-dichlorophenoxy)acetic acid) **(2)** and MCPA ((4-chloro-2-methylphenoxy)acetic acid) **(6)** are among the most important agricultural pesticides.[4] In total, more than 1000 tonnes of phenoxy acid herbicides are applied annually in Germany.[4] Additionally, new polar herbicides have been developed and introduced in agriculture.

Some acidic pesticides can be determined in routine multimethods together with the phenoxy acids. The common names of all pesticides that can be analyzed with the analytical methods presented in this article are compiled in Table 1 together with their IUPAC (International Union of Pure and Applied Chemistry) names, CAS (Chemical Abstracts Service) numbers and their mode of action as pesticides. The structural formulae of all compounds are shown; **(1)** to **(9)** are phenoxyalkanoic acid herbicides, **(10)** to **(19)** are amide and phenolic pesticides and **(20)** to **(40)** are other acidic pesticides. The range of analytes includes many carboxylic acids and also two amides, namely bentazone **(10)** and fuberidazole **(11)**, and some phenolic pesticides such as bromoxynil **(13)** and ioxynil **(14)** which are frequently used in combination with phenoxy acid herbicides to broaden the range of weeds controlled.[5] As a whole, more than 40 acidic pesticides can be analyzed from all different kinds of water and soil samples using the methods described in this article.

3 INSTRUMENTAL ANALYSIS

Separation of the analytes from each other and from interfering matrix compounds is a prerequisite in environmental multiresidue analysis. Therefore, HPLC and capillary GC are the two favorite separation techniques in multiresidue analysis of acidic herbicides in water or soil samples. As will be discussed in detail in the following sections, both techniques have advantages and drawbacks when analyzing acidic pesticides. At the moment, GC methods in combination with MS have some advantages compared to HPLC or other GC methods and thus they are most frequently applied in all modern environmental laboratories. Nevertheless, there are some very promising new instruments, such as HPLC coupled to API/MS/MS. As prices for these sophisticated but still easy to use instruments are decreasing they will probably gain much more importance in environmental analysis of polar pesticides. Apart from GC and HPLC methods there are also other techniques which are used for the analysis of acidic pesticides. A modern variation of thin-layer chromatography (TLC) is the so-called AMDTLC which offers much higher separation power than conventional TLC methods. In combination with HPLC the separation power of both techniques can be enhanced remarkably,[6] but has not yet been applied to acidic pesticides. Methods applying CZE have some importance, especially when information about the distribution of different enantiomers of acidic pesticides in the environment is required. There are also some immunochemical methods which have been developed for the analysis of different phenoxy acid herbicides, such as 2,4-D **(2)**, 2,4,5-T **(9)**, mecoprop **(8)** and dichlorprop **(4)**.[7–20] Owing to their selectivity, the immunochemical assays are only suitable for single compound analysis. These methods often show some cross reactivities to chemically related compounds whenever these occur at higher concentrations in the samples. The cross reactivities are, however, not normally desirable and have to be suppressed to avoid false positive results. Thus, immunochemical methods may be used as rapid screening methods in environmental analysis but have not been included in this article describing multiresidue methods.

3.1 Capillary Gas Chromatography

Capillary GC is the analytical method with the greatest separation power and therefore the best suited to multiresidue analysis of pesticides from environmental samples. Automated GC analysis has become routine for environmental samples, with autosamplers carrying out automatic injection. There are many monographs about GC and GC/MS which describe the hardware, the techniques, potential interferences and applications in environmental analysis.[21–29] A major drawback of conventional GC analysis using hot split/splitless injection is that only small proportions of the samples (1–5 μL) can be introduced into the GC. Therefore, new large volume injection techniques have been developed[30–32] to enable the direct analysis of up to 1 mL of sample extracts to improve, if necessary, the detection limits of GC analysis. One technique, the programmed temperature vaporization (PTV) has been used for the analysis of acidic pesticides, as will be described in section 3.1.4.2.

The ease of coupling GC to highly sensitive detectors such as the ECD, the NPD or the flame photometric detector (FPD), and most importantly the subsequent coupling of GC to MS, further increased the prevalence of GC over other analytical techniques in trace level analysis of pesticides.[33] As will be described in sections 3.1.3.2, 3.1.3.3 and 3.1.3.4, the use of MS applying special acquisition and/or ionization techniques such as SIM, chemical ionization (CI) or MS/MS acquisition enables the analysis of pesticide residues at parts per billion (ppb)

or even parts per trillion (ppt) levels. GC/MS detection also provides the highest possible level of confidence independent of the complexity of the matrix from water or even soil samples. Another new and very interesting GC detector is the AED, which enables the element specific analysis of pesticide residues. Its performance in the determination of acidic pesticides will be described in section 3.1.4.2.

Because of their high polarity and low volatility, acidic pesticides are not directly amenable to GC analysis and have to be derivatized using suitable derivatizing agents. Many different derivatization methods have been described and will be discussed in the following section. The determination of these derivatives applying GC with ECD, NPD, AED or MS detection will be demonstrated in sections 3.1.2 to 3.1.4.

3.1.1 Derivatization Techniques

Methylation by diazomethane and derivatization using pentafluorobenzyl bromide (PFBBr) are most commonly used for the determination of phenoxy acids and other acidic pesticides by GC. Depending upon the individual analytical problem there are some other very interesting alternative derivatization techniques whose benefits, drawbacks and suitability for the different compound classes will be described in this section.

3.1.1.1 Alkylation to Methyl, Ethyl and Butyl Esters Methylation of phenoxy acid herbicides with methanol/sulfuric acid[34–36] is one of the classical and very reliable derivatization methods. Unfortunately, methylation of phenolic or amide pesticides is not possible using this procedure. Therefore, it has for the most part been substituted by diazomethane derivatization[37–50] which is also able to methylate compounds such as bromoxynil **(13)**, ioxynil **(14)** or bentazone **(10)** which are often integrated in multiple-residue methods in water analysis. *Caution: Diazomethane has to be handled carefully, because it is toxic and known to be a potent carcinogen and mutagen. Diazomethane may also explode when used inappropriately, especially at higher concentrations.* Recently, Rimmer et al.[51,52] described the analysis of six phenoxy acid herbicides using trimethylsilyldiazomethane, which is commercially available. Derivatization with trimethylsilyldiazomethane provides a less hazardous method for preparing the phenoxy acid methyl esters under milder conditions than with diazomethane.[51]

As long as only carboxylic pesticides are to be analyzed, alkylation with chloroformates[53] is an interesting alternative. Using commercially available methyl, ethyl or butyl chloroformates, it is very easy to produce the corresponding methyl, ethyl or butyl esters. It is thus possible to identify and confirm acidic pesticides using only one GC column and a conventional ECD as shown in section 3.1.2. Methyl esters may also be formed by derivatization with boron trifluoride–methanol[54–56] or methyl iodide.[57] Another interesting technique for the methylation of acidic pesticides is the intra-injector derivatization using trimethylsulfonium hydroxide (TMSH)[58] or trimethylanilinium hydroxide (TMAH).[59] These derivatizations proceed very quickly and easily, but may be problematic in quantification, especially when interfering matrix compounds are involved.

There is one common drawback with all the above alkyl derivatives: they show some problems in ECD detection of lower chlorinated analytes such as mecoprop **(8)**, MCPA **(6)** or MCPB **(7)**. Other acidic pesticides such as bentazone **(10)** or 2-(1-naphthyl)acetic acid **(32)** are not detected by the ECD at all. Thus, many recent methods recommend GC/MS detection for the analysis of the methyl derivatives. Some typical alkylation procedures are described in detail in the following sections.

Methylation with Methanol/Sulfuric Acid. This derivatization procedure is according to the German standard method.[34] An aliquot of a standard solution or sample extract dissolved in methanol (about 2 mL) is combined with 1 mL of sulfuric acid ($\rho(H_2SO_4) = 1.84\,g\,L^{-1}$) in a 10-mL volumetric flask. The reaction mixture is allowed to stand for 10 min. Then the volume is adjusted to 10 mL by adding distilled water. One milliliter of hexane is added to the mixture which is then vigorously shaken. A portion of 1–2 µL of the hexane layer is injected into the GC.

For suitable GC conditions see sections on Methylation with Diazomethane and Alkylation using Chloroformates below.

Methylation with Diazomethane. This derivatization procedure is according to the EPA (Environmental Protection Agency) and the German standard method.[47,37] A diazomethane solution in methyl *tert*-butyl ether (MTBE) is produced by using a diazomethane solution generator under a fume hood as reported in detail elsewhere.[37,60] *(Caution: Wear appropriate safety equipment to minimize skin contact and inhalation hazards. Avoid heat, because diazomethane is explosive near 90 °C.)* Before use, the diazomethane solution can be stored for a few days or weeks at 0–5 °C[47] or −18 °C.[37]

An aliquot of a standard solution or sample extract containing the acidic pesticides (e.g. 1 mL) is dried under a gentle stream of nitrogen in a reaction tube. A portion of 0.5 mL of the diazomethane solution is added to the extract. After addition of diazomethane solution the analyte extracts should turn and remain yellow for more than 2 min, otherwise the methylation procedure should be repeated. The reaction tubes are then sealed with stoppers and held at room temperature for 30 min under a fume hood. The unreacted diazomethane is destroyed by adding 0.1–0.2-g of silicic acid or by evaporation to

a defined volume >200 μL. The reaction tubes are then left to stand for approximately 20 min until evolution of nitrogen gas ceases. The extracts are adjusted to a defined volume by adding MTBE and injected (2 μL) directly or after a cleanup procedure using Florisil[47] into GC.

Suitable GC conditions[37,45] are as follows. Instrument: GC/ECD or GC/MS. Capillary column: HT-5 (SE-54) fused-silica capillary column, 25 m × 0.25 mm ID (inner dimensions), with a film thickness of 0.1 μm and hot splitless injection. Carrier gas: helium (0.8 bar). The oven temperature is held at 80 °C for 2 min following injection, then programmed at 20 °C min^{-1} to 160 °C, which is held for 4 min, then increased at 4 °C min^{-1} to 200 °C, followed by an increase of 25 °C min^{-1} to 260 °C and finally held for 3 min. Injection port, detector (ECD) or interface (MS) temperatures are set to 250 °C, 300 °C and 270 °C, respectively.

Alkylation using Chloroformates. This derivatization procedure is according to Butz and Stan.[53] An aliquot of a standard solution or sample extract (e.g. 1 mL) is dried under a gentle stream of nitrogen in a reaction tube. To obtain the ethyl esters the dried residue is dissolved in 100 μL of a reaction mixture containing acetonitrile, ethanol, water and pyridine (5:2:2:1 v/v) followed by 7 μL of ethyl chloroformate. To obtain the methyl esters the dried residue is dissolved in 100 μL of a reaction mixture containing acetonitrile, methanol, water and pyridine (2:2:7:1 v/v) followed by 7 μL of methyl chloroformate. To obtain the butyl esters the dried residue is dissolved in 100 μL of a reaction mixture containing acetonitrile, butanol, water and pyridine (2:2:6:1 v/v) followed by 7 μL of butyl chloroformate. Then the reaction tube is shaken gently against a pad for about 5 s to initiate the gas evolution. The solution is dried under a gentle stream of nitrogen and the residue is dissolved in 100 μL of toluene. A 1-μL portion of this solution is injected into the GC.

Suitable GC conditions[53] are as follows. Instrument: GC/ECD or GC/MS. Capillary column: HP-5 fused-silica capillary column, 25 m × 0.2 mm ID, with a film thickness of 0.32 μm and hot splitless injection with the split closed for 1 min. Carrier gas: helium. The oven temperature is held at 100 °C for 1 min following injection, then programmed at 30 °C min^{-1} to 150 °C, which is held for 2 min, then increased at 3 °C min^{-1} to 205 °C, followed by an increase of 10 °C min^{-1} to 260 °C and finally held for 25 min. Injection port, detector (ECD) or interface (MS) temperatures are set to 220 °C, 300 °C and 190 °C, respectively.

3.1.1.2 Alkylation Introducing Hetero Atoms Derivatization methods to form alkyl esters containing hetero atoms such as fluorine or chlorine have been developed to overcome the sensitivity problem of the ECD. Some of the resulting derivatives also have excellent mass spectrometric properties for trace level multiresidue analysis applying GC/MS with SIM programming (see section 3.1.3.2).

The formation of 2,2,2-trifluoroethyl,[61–63] 2-chloroethyl,[61] 2,2,2-trichloroethyl[61,64] and 2,2,3,3,3-pentafluoropropyl esters[64] of phenoxy acids has been described. In environmental analysis the formation of pentafluorobenzyl (PFB) esters by reaction with PFBBr has gained much importance.[28,64–79] PFB esters increase the sensitivity of ECD detection tremendously and they allow the determination of acidic pesticides independent of their substitution with chlorine. Derivatization with PFBBr enables the determination of phenoxy acid herbicides, other carboxylic acid pesticides, some phenolic pesticides and even amide pesticides such as bentazone or fuberidazole.[28,75,76]

Three different procedures for the formation of PFB esters have been reported. The classical procedure is performed using acetone as solvent and potassium carbonate as catalyst.[64–71,78] Heberer et al.[28,75,76] described an alternative procedure using toluene as solvent and triethylamine as basic catalyst. This procedure has some advantages when analyzing environmental samples because many interfering matrix compounds are excluded from GC before analysis.[28,76] Meiring et al.[73] described the phase-transfer-catalyzed derivatization of polar pesticides using tetrahexylammonium hydrogensulphate as complexing agent which transports the analytes from the aqueous layer into a dichloromethane layer where they are derivatized with PFBBr.

As shown in section 3.1.2, a major drawback of the PFB ester procedure is the large number of interfering substances in the extract from the environmental matrix when the extract is analyzed by GC/ECD.[28,64,75] The analysis of derivatized sample extracts by GC/ECD often requires time-consuming cleanup procedures.[64] Thus, the screening of these derivatives using GC/MS with SIM (see section 3.1.3.2) or using GC with AED, as shown in section 3.1.4.2, has great advantages compared to conventional GC/ECD methods. GC/MS with SIM enables the analysis of more than 40 acidic pesticides down to the low microgram or even nanogram per kilogram level in soil samples and down to the low nanogram per liter level in water samples, as will be shown in section 4.

Pentafluorobenzylation. A first derivatization procedure is according to Heberer et al.[28,75,76] An aliquot of a standard solution or sample extract (e.g. 1 mL) is dried under a gentle stream of nitrogen in a sample vial. A 200-μL portion of a PFBBr solution (100-μL PFBBr dissolved in 4.9-mL toluene) and 2-μL triethylamine are added to the residue. The vial is capped with a Teflon®-lined septum and the reaction mixture is heated for 1 h at 110 °C

simply by placing in a drying cabinet. The mixture is then cooled to room temperature, the vials are decapped and the solution containing the derivatized compounds is dried under a gentle stream of nitrogen. The residue is dissolved in 100 μL of toluene and 2 μL of this solution is injected into the GC/MS.

The second derivatization procedure is according to Gurka et al.[65] An aliquot of a standard solution or sample extract (e.g. 1 mL) is dried under a gentle stream of nitrogen in a sample tube. Four milliliters acetone, 30 μL of 10% K_2CO_3 in distilled water and 200 μL of 3% PFBBr in acetone are added to the residue. The tube is closed with a glass stopper and mixed on a vortex mixer. The tube is then heated for 3 h at 60 °C in a tube heater. The mixture is evaporated to 0.5 mL under a gentle stream of nitrogen. A 2-mL portion of hexane is added to the solution and evaporation is repeated to dryness at room temperature. The residue is dissolved in toluene–hexane (1 + 6) for column cleanup using silica columns topped with anhydrous Na_2SO_4. The column is prewetted with 5-mL hexane (let drain to top of sorbent) and the reaction mixture is transferred quantitatively to the top of the column. The column is then washed with 8 mL of toluene–hexane mixture (1 + 6) and the PFB derivatives are eluted with toluene–hexane (9 + 1) to collect 8.0 mL. An aliquot of this fraction is analyzed by GC/ECD or GC/MS.

Suitable GC conditions[28,75,76] are as follows. Instrument: GC/ECD, GC/AED or GC/MS. Capillary column: 25 m × 0.2 mm ID × 0.33 μm HP-5 capillary column fitted with a 1.5 m × 0.32 mm ID × 0.33 μm HP-5 precolumn. Carrier gas is helium (purity: 99.999%) set to a flow of 28 cm s^{-1} at the initial temperature. The oven temperature is held at 100 °C for 1 min following injection, then programmed at 30 °C min^{-1} to 150 °C, which is held for 1 min, then at 3 °C min^{-1} to 205 °C followed by 10 °C min^{-1} to 260 °C and finally held for 23 min. Injection port, detector (ECD) or interface (MS) temperatures are set to 230 °C, 300 °C and 250 °C, respectively. Samples are injected using hot splitless injection with the split closed for 0.9 min.

Trifluoroethylation. This derivatization procedure is according to Lee et al.[63] An aliquot of a standard solution or sample extract (e.g. 1 mL) is dried under a gentle stream of nitrogen in a Teflon®-capped vial. A 0.5-mL portion of sulfuric acid (analytical grade) and 1.0 mL of 2,2,2-trifluoroethanol are added to the vial. The vial is capped tightly, the mixture is gently swirled. The reaction is carried out in a water bath at 90 °C for 4 h. Then the vial is cooled down under flowing tap water. Deionized water (6 mL) and 2 mL of hexane are added to the vial which is then shaken vigorously. After separation of the two liquid phases, 1 mL of the hexane layer is transferred to a 10-mL volumetric flask and diluted to volume with hexane. A 2-μL portion of this solution is injected into the GC.

Suitable GC conditions[63] are as follows.

1. Instrument: GC/ECD. Capillary column: Rtx-35 fused-silica capillary column, 30 m × 0.53 mm ID, with a film thickness of 0.5 μm. Samples are injected using splitless injection. Carrier gas is helium, make-up gas 5% methane in argon. The oven temperature is held at 130 °C for 6 min following injection, then programmed at 30 °C min^{-1} to 160 °C, which is held for 5 min, then increased to 210 °C, which is held for 6 min followed by an increase to 260 °C and finally held for 3 min. Injection port and detector temperatures are set to 265 °C and 300 °C respectively.

2. Instrument: GC/MS. Capillary column: HP-5 fused-silica capillary column, 30 m × 0.25 mm ID, with a film thickness of 0.25 μm and splitless injection. Carrier gas is helium set to a head pressure of 90 kPa. The oven temperature is held at 80 °C for 3 min following injection, then programmed at 30 °C min^{-1} to 140 °C, which is held for 8 min, then increased to 210 °C, which is held for 5 min followed by an increase to 260 °C and finally held for 5 min. Injection port and interface temperatures are set to 230 °C and 280 °C respectively.

3.1.1.3 Silylation Silylation is probably the most versatile derivatization technique currently available to enhance GC performance by blocking protic sites.[80] Therefore, it is frequently used especially in clinical chemistry. The use of silylation reagents has, however, one major drawback; almost all of the reagents and the resulting derivatives are sensitive to the hydrolytic effects of moisture. Thus, for a long time silylation with (2-cyanoethyl)dimethyl(diethylamino)silane (CEDMSDEA) was the only described silylation procedure for the analysis of acidic pesticides.[81,82]

Another promising method is the derivatization of acidic pesticides with *N*-(*tert*-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (MTBSTFA) forming their *tert*-butyldimethylsilyl (TBDMS) derivatives. The derivatization procedure is simple and rapid and, as shown in section 3.1.3.1, the resulting TBDMS derivatives exhibit characteristic EI (electron ionization) MS spectra. In contrast to other silylated derivatives they have proven to be stable for several days, even in the presence of moisture. The remarkable stability of the derivatives to hydrolysis can be explained by the sterical hindrance caused by the large tertiary butyl moiety which protects the silicon–oxygen bond from hydrolytic attack.[83] This method was successfully applied by Heberer and Stan[83] for the analysis of more than 50 substituted phenols including also six phenolic pesticides in water analysis. As presented here, this derivatization technique is also suitable

for the determination of other acidic pesticides, including all pesticides containing carboxylic acids. Unfortunately, amides such as bentazone **(10)** or fuberidazole **(11)** are not derivatized by this method.

Silylation with (2-Cyanoethyl)dimethyl(diethylamino)-silane. This derivatization procedure is according to Bertrand et al.[81,82] The preparation of the silylating reagent CEDMSDEA was published in 1986 by Bertrand et al.[84] A 20-μL portion of a standard solution or a sample extract in ethyl acetate is put into a 0.5-mL Reacti-Vial pretreated with dichlorodimethylsilane for reproducibility. Then 60 μL of CEDMSDEA reagent and 45 μL of ethyl acetate are added. The solution is shaken briefly and within 5 min, 1 μL of the extract is injected into the gas chromatograph.

Suitable GC conditions[81] are as follows. Instrument: GC/NPD or GC/MS. Capillary column: 15 m × 0.23 mm ID DB-1 (SE-30) capillary column. Carrier gas is helium and the injector is used with an initial split ratio of 32 : 1. The oven temperature is programmed from 180 °C to 208 °C at 3 °C min^{-1} followed by 10 °C min^{-1} to 250 °C. Injection port, detector (NPD) and source (MS) temperatures are set to 275 °C, 300 °C and 220 °C, respectively. When using NPD detection, helium is used as make-up gas at a flow of 30 mL min^{-1} and the flow rates of hydrogen and air at the detector are 2.6 and 100 mL min^{-1}, respectively.

Silylation with N-*(*tert-*Butyldimethylsilyl)*-N-*methyltrifluoroacetamide.* This derivatization procedure is according to Heberer and Stan.[83] An aliquot of a standard solution or sample extract (about 1 mL) is dried under a gentle stream of nitrogen in a sample vial. The residue is dissolved in 50 μL of acetonitrile, then 50 μL of MTBSTFA are added. The derivatization is carried out in sampling vials which are capped with a Teflon®-lined sampler cap and placed in a drying cabinet at 80 °C for 1 h. The mixture is cooled to room temperature and the vials are decapped. The solution containing the derivatives is transferred to 200-μL vial inserts and 1 μL is injected directly in the gas chromatograph.

Suitable GC conditions are as follows. Instrument: GC/MS. Capillary column: 25 m × 0.2 mm ID × 0.33 μm HP-5 capillary column fitted with a 1.5 m × 0.32 mm ID × 0.33 μm HP-5 precolumn. The carrier gas is helium (purity: 99.999%) set to a flow of 28 cm s^{-1} at the initial temperature. The oven temperature is held at 100 °C for 1 min following injection, then programmed at 30 °C min^{-1} to 150 °C, which is held for 1 min, then at 3 °C min^{-1} to 205 °C followed by 10 °C min^{-1} to 260 °C and finally held for 23 min. Injection port and interface temperatures are set to 230 °C and 250 °C, respectively. Samples are injected using hot splitless injection with the split closed for 0.9 min.

3.1.2 Gas Chromatography with Electron Capture Detection

GC with ECD has been used successfully for many years, especially, in trace level analysis of chlorinated pesticides. Analysis by GC/ECD still belongs to the most sensitive detection methods and is the yardstick for other analytical instruments. Many acidic pesticides are substituted with hetero atoms, mostly chlorine, which make them suitable for ECD. Thus, even the methyl esters can be analyzed by GC/ECD. When using methyl, ethyl or butyl chloroformates for derivatization, it is possible to identify and confirm acidic pesticides using only one GC column as shown in Figure 1. In this case, the shift of the retention times caused by introducing a methyl, ethyl or butyl moiety into the individual molecule is used for analyte identification.

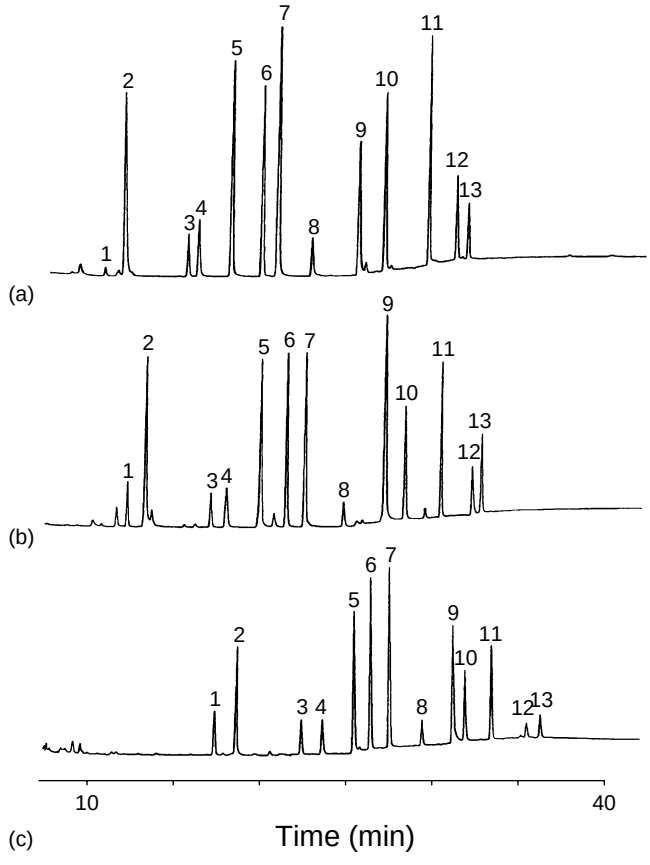

Figure 1 ECD chromatograms of a standard mixture containing 13 acidic pesticides as their (a) methyl, (b) ethyl and (c) butyl esters. One nanogram of each compound has been analyzed. 1, dichlorobenzoic acid; 2, clopyralid **(24)**; 3, dichlorprop **(4)**; 4, 2,4-D **(2)**; 5, triclopyr **(40)**; 6, fenoprop **(5)**; 7, 2,4,5-T **(9)**; 8, 2,4-DB **(3)**; 9, picloram **(33)**; 10, fluazifop **(28)**; 11, haloxyfop **(31)**; 12, flamprop **(27)**; 13, acifluorfen **(20)**. (Reprinted from Butz, Stan, *J. Chromatogr.*, copyright (1993),[53] with permission from Elsevier Science.)

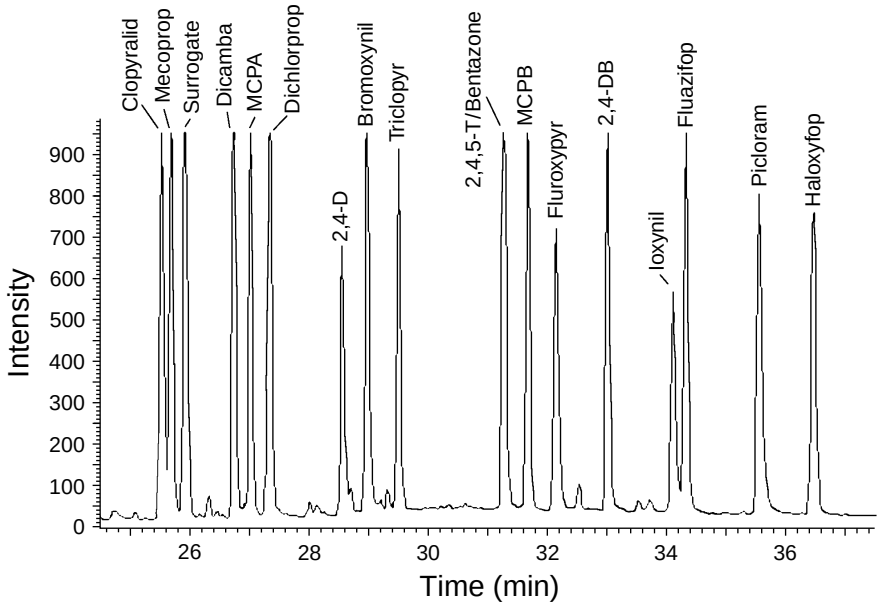

Figure 2 ECD chromatogram of a pentafluorobenzylated standard mixture containing 17 acidic pesticides and the surrogate standard compound 2-(4-chlorophenoxy)butyric acid. A sample of 2.5 ng per compound was injected.

As mentioned before, there is one drawback that all alkyl derivatives have in common; they present some problems in ECD detection of lower chlorinated analytes, e.g. mecoprop **(8)**, MCPA **(6)** or MCPB **(7)**. Alkyl esters of acidic pesticides such as bentazone **(10)** or 2-(1-naphthyl)acetic acid **(32)** are not detected by GC/ECD at all.

Lee et al.[64] compared the relative response factors (RRFs) of methyl, trichloroethyl (TCE), pentafluoropropyl (PFP) and PFB derivatives of 13 acidic herbicides applying GC/ECD detection. The RRFs of the PFB esters were found to be twice as large as those of the corresponding TCE esters and much higher than those of the PFP or even the methyl esters. The RRFs of the MCPA methyl or PFP ester were very low, thus both derivatives were found to be inapplicable to the determination of MCPA and other monochlorinated herbicides by GC/ECD at residues levels.[64] On the other hand, trichloroethanol did not react with bromoxynil **(13)**, and the yield of the 2,3,6-TBA **(39)** TCE ester was poor. Therefore, Lee et al.[64] concluded that the PFB esters were the derivatives best suited to the simultaneous determination of all 13 herbicides at trace levels applying GC/ECD. PFB esters increase the sensitivity of ECD detection tremendously and they allow the determination of acidic pesticides independent of their substitution by chlorine. As an example for multiresidue analysis, Figure 2 shows an ECD chromatogram of a standard mixture of 17 acidic pesticides as their PFB esters, amides and ethers.

A major drawback of the PFB ester procedure is the large number of interfering substances in the extract from the environmental matrix when the extract is analyzed by GC/ECD.[28,64,75] Figure 3 shows the ECD chromatogram of a derivatized extract of a drinking water sample measured without additional cleanup procedures.

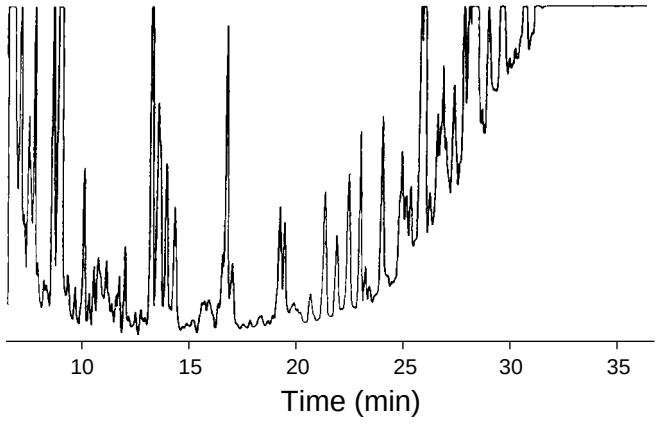

Figure 3 ECD chromatogram of a pentafluorobenzylated extract of a groundwater sample spiked with nine phenoxy acid herbicides. All analyte peaks are covered by matrix compounds. (Reproduced from Butz[85] by permission of Wissenschaft & Technik Verlag.)

As described by Lee et al.,[64] the analysis of derivatized sample extracts by GC/ECD often needs time-consuming cleanup procedures using gel permeation chromatography (GPC) and/or silica-gel columns. Therefore, the screening of PFB derivatives using GC/MS or using GC/AED has great advantages compared to conventional GC/ECD methods.

3.1.3 *Gas Chromatography with Mass Spectrometry and Gas Chromatography with Tandem Mass Spectrometry*

The coupling of GC to MS has greatly benefited from the development of fused silica capillary columns and the development of small, relatively inexpensive mass spectrometers as dedicated GC detectors. GC/MS instruments which are easy to use and easy to maintain are now readily available to residue chemists. Thus, specialists in MS are no longer a prerequisite for GC/MS operation. MS is the most sensitive method of molecular analysis yielding information on the molecular weight (MW) as well as the structure of the analytes. When GC and MS are directly combined into one GC/MS system, the capabilities of that system are not merely the sum of the capabilities of these two outstanding analytical methods; the increase in analytical information is exponential. Extreme selectivities can be obtained which are of great importance in screening analysis of target analytes in environmental samples.

Positive identification of pesticide residues in environmental samples (soil or water) at trace level concentrations confronts the analyst with a number of problems. The high sensitivity and selectivity of modern GC/MS instruments enable the detection of the analytes at the parts per billion ($\mu g\,L^{-1}$ or $\mu g\,kg^{-1}$) or even at the parts per trillion or parts per quadrillion ($pg\,L^{-1}$ or $pg\,kg^{-1}$) ($ng\,L^{-1}$ or $ng\,kg^{-1}$) level, depending on the matrix and in particular on the chemical structure of the pesticide.

In the beginning, GC/MS was mostly used for the confirmation of results positively monitored with screening methods using less selective detectors such as ECD or NPD. Nowadays, screening for pesticides in environmental samples with full scanning as well as target compound analysis applying SIM programming is becoming more and more popular for the multiresidue analysis of acidic pesticides. GC/MS analysis is sometimes unavoidable because some analytes which do not contain chlorine, fluorine or bromine atoms, such as the methyl esters of bentazone **(10)** or 2-(1-naphthyl)acetic acid **(32)**, are not even applicable to common GC/ECD analysis and often GC/ECD analysis of environmental samples is heavily disturbed by matrix interferences which mostly do not interfere with GC/MS analysis.

3.1.3.1 Electron Ionization Mass Spectrometry Measurements With most modern instruments, full scan spectra can be obtained at the low nanogram or even at the picogram level. Combined with suitable sample preparation methods, as shown in section 4.1.1, these detection limits are sufficient for the analysis of trace level concentrations of pesticides in water samples at the microgram per liter or even down to the nanogram per liter level. Using GC/MS with SIM programming (see section 3.1.3.2) acidic pesticides can be analyzed at much lower concentrations comparable to ECD detection. A prerequisite for GC/MS analysis is, however, knowledge of the mass spectral properties of the analytes, i.e. knowledge about important peaks (molecular ions and dominant fragments) and their corresponding intensities in the individual mass spectrum. Commercially available mass spectral libraries contain the EI mass spectra of up to 275 000 compounds, for example the WILEY library,[86] and there are some smaller, more specialized libraries which contain mass spectra of pesticides, e.g. the HPPEST or the Pfleger, Maurer, Weber library.[87,88] The MS libraries contain, however, only few mass spectra of pesticide derivatives. In this section, some GC/MS chromatograms of derivatized multicomponent standard mixtures of acidic pesticides are shown. Additionally, mass spectral data have been collected for many acidic pesticide derivatives and important mass spectral features of these derivatives will be discussed.

Methyl Derivatives. Figure 4 shows the GC/MS chromatogram of a standard mixture of 17 acidic pesticides which have been methylated with diazomethane. Even the coeluting derivatives of fluroxypyr **(30)** (mass to charge ratio, m/z 209, 268, 270) and 2,4-DB **(3)** (m/z 59, 101, 262) can be identified and quantified by their characteristic ion traces as demonstrated in Figure 5.

This methylation shifts the MW of the analytes by 14 Da. Many of the methyl derivatives exhibit characteristic ions in their EI mass spectra. Exceptions are given by the mass spectra of the methyl derivatives of MCPB **(7)** and 2,4-DB **(3)** which exhibit only noncharacteristic ions with m/z 59 and 101 at high intensities, whereas the molecular ions are of low intensity not suitable for trace level analysis. Table 2 shows the three most indicative ions in the EI mass spectra of individual pesticide derivatives together with their relative intensities. This information, together with the retention times of the analytes, facilitates creation of time scheduled SIM programs for trace level target compound analysis as shown in section 3.1.3.2.

Pentafluorobenzyl Derivatives. Derivatization with PFBBr shifts the MWs of the acidic pesticides by 180 Da. The typical fragment ion in the EI mass spectra of all PFB derivatives is the resonance-stabilized pentafluorobenzylium ion at m/z 181, which is always accompanied by a smaller fragment ion at m/z 161 resulting from the cleavage of HF. As can be seen from Table 3 the characteristic pentafluorobenzylium ion often appears as a base peak in the mass spectra, especially with phenolic pesticides. It may be used as a ''monitoring ion'' for PFB derivatives in SIM programming (section 3.1.3.2),

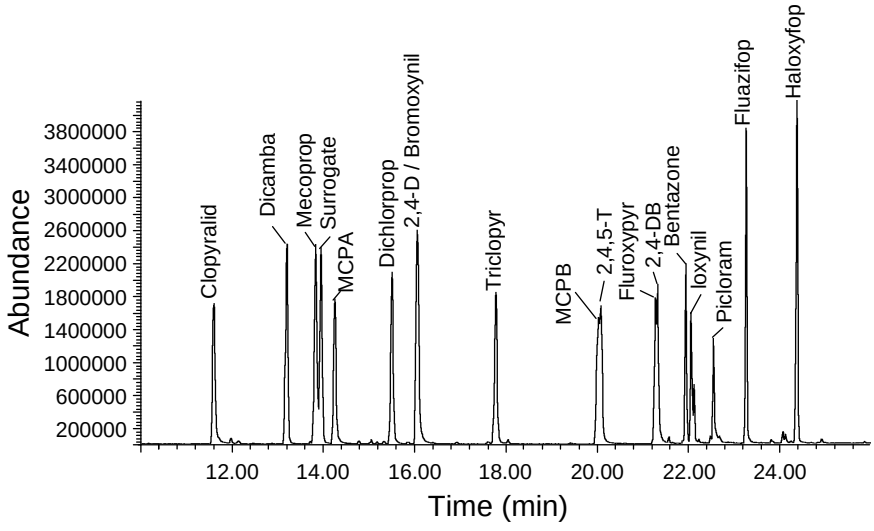

Figure 4 TIC (total ion current) chromatogram of a methylated standard mixture containing 17 acidic pesticides and the surrogate standard compound 2-(4-chlorophenoxy)butyric acid.

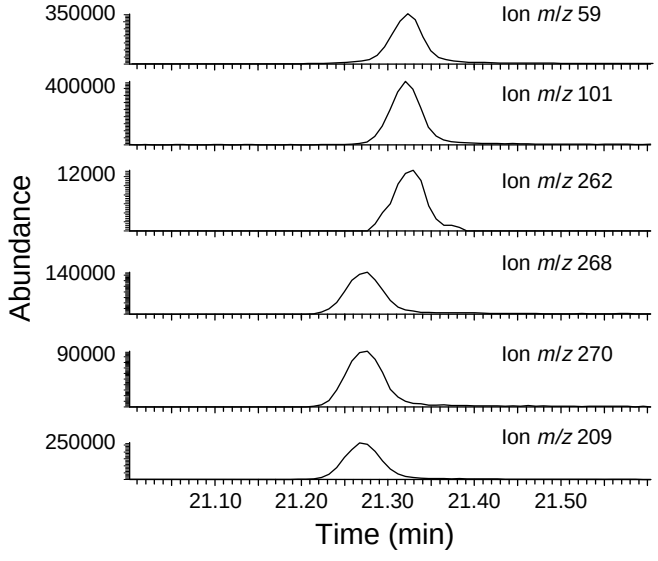

Figure 5 Indicative ion traces for fluroxypyr-methyl (*m*/*z* 268, 270 and 209) and 2,4-DB-methyl (*m*/*z* 59, 101 and 262) extracted from the TIC chromatogram in Figure 4.

but it is not selective and should not be used for identification and quantification of the pesticides. Most of the acidic pesticides exhibit other characteristic fragments or even high intensity molecular ions,[75] which enable their analysis at trace level concentrations. Table 3 shows a compilation of the three most indicative ions in the EI mass spectra of the individual pesticide derivative together with their relative intensities.

Figure 6 shows the GC/MS chromatograms of a derivatized standard mixture and the derivatized extract of a fortified sewage farm groundwater sample. Both chromatograms were recorded by GC/MS applying SIM programming, as described in detail in section 3.1.3.2. Although the fortified sewage farm groundwater sample contains a bulk of matrix compounds, the sample chromatogram looks as transparent as the standard chromatogram. Thus, the identification and quantitation of all 16 acidic pesticides was possible at a concentration level of only 200 ng L^{-1} by using the indicative ion traces.

Trifluoroethyl Esters. Table 4 shows the three most indicative ions in the EI mass spectra of the TFE (trifluoroethyl) esters of the individual pesticide together with their relative intensities.

tert-*Butyldimethylsilyl Derivatives.* Figure 7 shows the GC/MS chromatogram of a standard mixture of 15 acidic pesticides as their TBDMS derivatives. Derivatization with MTBSTFA shifts the MW of the analytes by 114 Da, except where the analytes are also derivatized a second functional group such as fluroxypyr **(30)** and picloram **(33)** (both +228 Da). The molecular ion is mostly weak or absent, but the EI mass spectra of the TBDMS derivatives exhibit very characteristic $(M\text{-}57)^+$ ions resulting from the cleavage of the *tert*-butyl moiety. This very intense peak also enables the determination of the MW for unknown compounds. Other characteristic mass fragments are $(M\text{-}15)^+$, from the cleavage of a methyl group, $(M\text{-}114)^+$, formed by the cleavage of the TBDMS group and $(M\text{-}101)^+$, resulting from the cleavage of the *tert*-butyl moiety and the additional loss of CO_2. Table 5 shows the three most indicative ions in the EI mass spectra of the individual pesticide derivatives together with their relative intensities.

Table 2 Retention times on an HP-5 capillary column, MWs and MS data of some acidic pesticides as their methyl derivatives

Methyl derivative	MW ($g\,mol^{-1}$)	Retention time (min)	Mass 1 (rel. int.)[a]	Mass 2 (rel. int.)[a]	Mass 3 (rel. int.)[a]
Bromoxynil **(13)**	291	16.05	291 (100)	289 (52)	276 (54)
2-(4-Chlorophenoxy)butyric acid (Surrogate)	228	13.94	128 (100)	130 (34)	228 (47)
Clopyralid **(24)**	205	11.58	147 (100)	174 (31)	205 (16)
2,4-D **(2)**	234	16.05	199 (100)	234 (55)	236 (35)
2,4-DB **(3)**	262	21.32	101 (100)	162 (23)	164 (15)
Dicamba **(25)**	234	13.19	203 (100)	205 (64)	234 (29)
Dichlorprop **(4)**	248	15.50	162 (100)	164 (65)	248 (57)
Fluazifop **(28)**	341	23.27	341 (100)	282 (94)	254 (72)
Fluroxypyr **(30)**	268	21.28	209 (100)	211 (64)	268 (51)
Haloxyfop **(31)**	375	24.38	316 (100)	375 (95)	377 (33)
Ioxynil **(14)**	385	22.06	385 (100)	370 (35)	243 (26)
MCPA **(6)**	214	14.24	141 (100)	214 (94)	216 (32)
MCPB **(7)**	242	20.03	101 (100)	59 (71)	242 (9)
Mecoprop **(8)**	228	13.82	169 (100)	142 (91)	228 (74)
Picloram**(33)**	254	22.55	196 (100)	198 (98)	254 (26)
2,4,5-T **(9)**	268	20.08	233 (100)	235 (58)	268 (49)
Triclopyr **(40)**	269	17.77	210 (100)	212 (98)	269 (35)

[a] Relative intensity of the ions in the full scan EI spectrum of the individual compound (base peak = 100).

3.1.3.2 Selected Ion Monitoring Since GC/MS with full scan often does not provide the sensitivity necessary to meet the maximum tolerances in drinking water or soil, GC/MS in SIM mode is applied to achieve lower detection limits. The increase in sensitivity is caused by the increase of the relative individual scan times (also called dwell times) which are spent detecting only indicative ions of the analytes.

GC/MS with SIM is commonly used for the confirmation of positive results obtained with conventional detectors such as GC/ECD. Following decreasing prices of GC/MS instruments, GC/MS with SIM is already in widespread use for screening analysis of pesticides in water.[27] Multiresidue methods using SIM have gained much importance for trace level analysis of pesticides in environmental samples, especially with acidic pesticides.[27] This may also be a consequence of the necessity to derivatize polar compounds prior to GC. Thus, only a limited number of derivatizable compounds has to be targeted. GC/MS with SIM acquisition is often a prerequisite for environmental trace level analysis because derivatized matrix compounds or by-products of the derivatization procedure often interfere with detection by conventional GC detectors.

Heberer et al.[75] reported a method for the detection of more than 30 acidic pesticides as their PFB esters in a single GC run, applying GC/MS with SIM time window programming. In this method, three indicative ions from the mass spectrum were selected for each analyte and an appropriate retention time window for recording those ions was set.[75] Table 6 shows the resulting SIM program used to produce the MID chromatogram, a summation of all ions in each corresponding time window, as shown in Figure 8. In this case, only 20 pg of each analyte were injected and analyzed by GC/MS with SIM. This amount is equivalent to a concentration of $1\,ng\,L^{-1}$ in water analysis presuming 100% recovery of the analytes and a concentration factor of 10 000 being obtained when using the sample preparation procedure described in section 4.1.1. The identity of the compounds can be easily confirmed by extracting the ion traces of the indicative analyte ions.

In general, SIM programming can be used for the analysis of any compound which is amenable to GC. The only prerequisite for SIM programming is knowledge of the individual GC retention times, the indicative masses and their relative abundances in the mass spectra. All this information, including the indicative ions labeled as mass 1 to 3, is given in Tables 2 to 5 for the different derivatives of acidic pesticides. Using this information and recording a full scan chromatogram with a derivatized standard mixture to determine the retention times, the analyst is able to create a suitable GC/MS SIM program for multicompound trace level analysis. To take full advantage of the increase in detection sensitivity provided by the SIM technique, the indicative ions are only recorded in a time window in which the target compound is expected. The only problem with SIM programming is to find intervals in the chromatogram where switching from one window to the other can be performed without missing any of the target compounds. When creating a SIM method for the analysis of so many compounds in one GC run a compromise has to be found between the wish to have the smallest number of target compounds

Table 3 Retention times on an HP-5 capillary column, MWs and MS data of some acidic pesticides as their PFB derivatives according to Heberer et al.[28,75,76]

Common name	MW ($g\,mol^{-1}$)	Retention time (min)	*m/z* 181 (rel. int.)[a]	Mass 1 (rel. int.)[a]	Mass 2 (rel. int.)[a]	Mass 3 (rel. int.)[a]
Acifluorfen PFB ester	541	42.96	100	345 (77)	511 (16)	541 (10)
Benazolin PFB ester	423	35.22	46	170 (100)	423 (88)	198 (41)
Bentazone PFB amide	420	31.17	100	378 (33)	420 (15)	341 (11)
Benzoic acid PFB ester	302	16.15	100	302 (65)	257 (40)	258 (23)
Bromoxynil PFB ether	455	29.07	100	161 (6)	457 (3)	459 (1)
Chloramben PFB ester	385	31.75	77	385 (100)	387 (66)	161 (27)
Chlorfenac PFB ester	418	28.93	100	195 (60)	383 (17)	418 (11)
Chloroxynil PFB ether	367	25.58	100	367 (1)	369 (0.7)	
2-(4-Chlorophenoxy)butyric acid PFB ester	394	26.63	100	128 (93)	394 (77)	169 (70)
4-CPA PFB ester	366	26.28	53	366 (100)	141 (52)	368 (33)
Clofibric acid PFB ester	394	25.42	38	128 (100)	130 (32)	394 (13)
Clopyralid PFB ester	371	26.18	100	147 (66)	175 (40)	308 (9)
2,3-D PFB ester	400	29.74	100	400 (39)	402 (26)	175 (31)
2,4-D PFB ester	400	28.72	100	175 (48)	400 (42)	402 (27)
3,4-D PFB ester	400	29.60	100	400 (80)	402 (52)	175 (57)
2,4-DB PFB ester	428	32.87	100	267 (24)	162 (8)	428 (1)
p, *p*′-DDA[c] PFB ester	460	–	20	235 (100)	237 (65)	460 (4)
Dicamba PFB ester	400	27.28	100	203 (79)	400 (31)	402 (20)
2,3-Dichlorobenzoic acid PFB ester	370	26.35	100	370 (24)	372 (16)	335 (26)
2,4-Dichlorobenzoic acid PFB ester	370	25.65	100	173 (48)	370 (42)	372 (28)
2,5-Dichlorobenzoic acid PFB ester	370	25.44	100	370 (14)	372 (9)	335 (44)
3,4-Dichlorobenzoic acid PFB ester	370	26.58	100	335 (72)	370 (32)	372 (21)
Dichlorophen PFB mono-ether	448	44.31	100	448 (23)	450 (15)	413 (14)
Dichlorprop PFB ester	414	27.79	100	162 (65)	414 (46)	416 (32)
Endothal-methyl PFB ester[b]	380	27.99	100	155 (16)	321 (6)	293 (6)
Fenoprop PFB ester	448	30.03	100	196 (49)	448 (22)	450 (21)
Flamprop PFB ester	501	39.64	5	105 (100)	276 (11)	501 (4)
Fluazifop PFB ester	507	34.34	22	282 (100)	254 (58)	507 (54)
Flurenol PFB ester	406	32.14	100	152 (31)	406 (12)	429 (2)
Fluroxypyr PFB ester	434	31.94	100	434 (36)	209 (28)	436 (23)
Fuberidazole PFB amide	364	31.60	13	364 (100)	183 (98)	90 (27)
Haloxyfop PFB ester	541	36.85	30	316 (100)	288 (63)	541 (44)
Imazapyr PFB ester	441	38.54	100	399 (73)	398 (36)	441 (20)
4-Indol-3-yl-butyric acid PFB ester	383	36.91	8	130 (100)	383 (57)	384 (11)
Indol-3-yl-acetic acid PFB ester	355	31.45	11	130 (100)	355 (57)	356 (11)
Indol-3-yl-propionic acid PFB ester	369	33.28	2	130 (100)	369 (60)	370 (21)
Ioxynil PFB ether	551	33.85	100	243 (4)	267 (3)	551 (0.5)
MCPA PFB ester	380	27.48	100	380 (95)	141 (52)	382 (32)
MCPB PFB ester	408	31.54	100	267 (25)	142 (5)	408 (3)
Mecoprop PFB ester	394	26.43	82	169 (100)	394 (94)	142 (59)
2-(1-Naphthyl)acetic acid PFB ester	366	29.17	18	141 (100)	366 (83)	115 (19)
Pentafluorophenoxyacetic acid PFB ester	422	17.12	100	197 (25)	422 (12)	
PCP PFB ester	444	–	100	444 (2)	446 (3)	448 (2)
Picloram PFB ester	420	35.48	96	196 (100)	198 (96)	420 (6)
Quinchlorac PFB ester	421	35.70	52	197 (100)	199 (66)	224 (43)
Quinmerac PFB ester	401	35.14	27	177 (100)	204 (37)	179 (32)
2,4,5-T PFB ester	434	31.12	100	434 (27)	436 (26)	211 (26)
2,3,6-TBA PFB ester	404	27.94	100	207 (37)	209 (37)	369 (19)
Triclopyr PFB ester	435	29.56	100	210 (30)	435 (20)	437 (19)

[a] Relative intensity of the ions in the full scan EI spectrum of the individual compound (base peak = 100).
[b] Rearrangement reactions result in the mono methyl PFB ester.
[c] DDA, bis-(chlorophenyl)acetic acid.

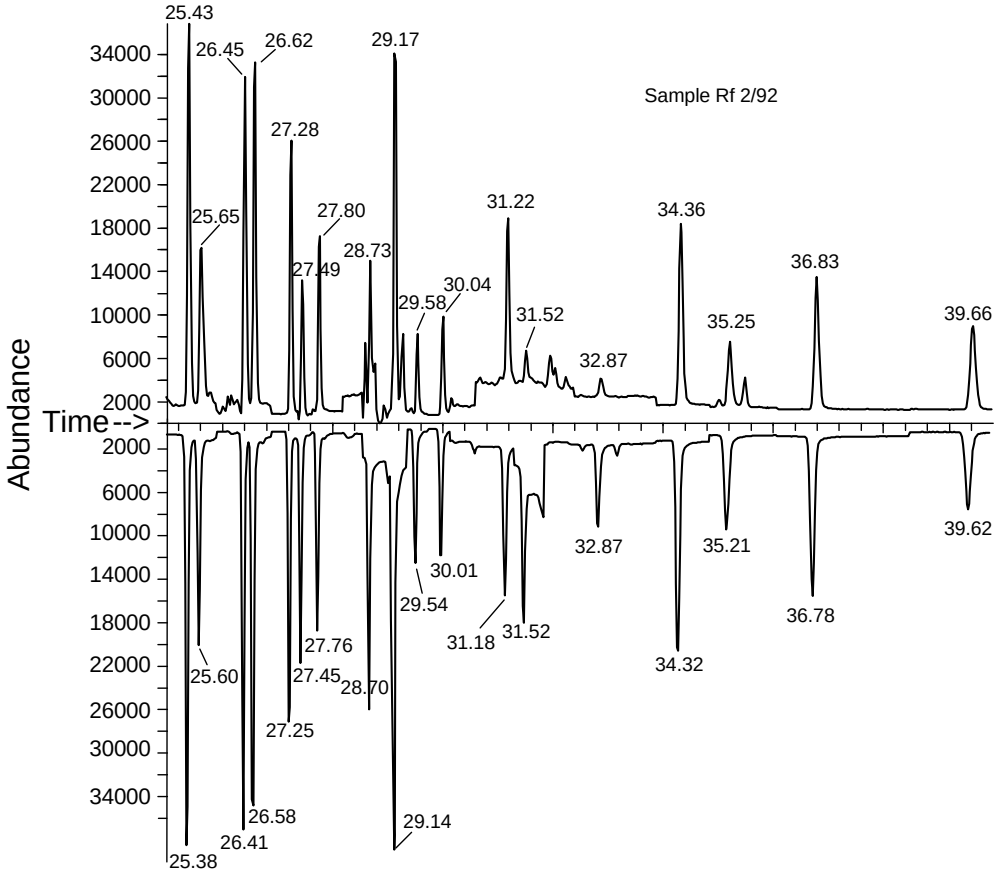

Figure 6 MID (multiple ion detection) chromatograms recorded with GC/MS applying SIM of the PFB esters of an acidic herbicide mixture (2 ng injected) and a derivatized extract of a sewage farm groundwater sample. For peak identification refer to Table 6. Sample spiked with 200 ng L^{-1}. (Reproduced from Stan and Heberer[27] by permission of Springer-Verlag (1995).)

Table 4 Retention times on an Rtx-35 capillary column and MS data of some acidic pesticides as their TFE derivatives according to Lee et al.[63]

Common name	Retention time (min)	Molecular peak (rel. int.)[a]	Mass 1 (rel. int.)[a]	Mass 2 (rel. int.)[a]	Mass 3 (rel. int.)[a]
Acifluorfen TFE ester	15.91	443 (100)	413 (43)	344 (22)	223 (37)
Clopyralid TFE ester	3.97	273 (22)	174 (65)	147 (100)	110 (35)
2,4-D TFE ester	7.26	302 (100)	267 (18)	175 (72)	161 (68)
2,4-DB TFE ester	11.46	330 (8)	231 (11)	169 (100)	127 (65)
Dicamba TFE ester	4.38	302 (41)	203 (100)	188 (26)	175 (14)
Diclofop TFE ester	18.78	408 (54)	281 (14)	253 (100)	162 (20)
Fenoprop TFE ester (Silvex)	8.53	350 (26)	225 (27)	196 (100)	83 (27)
Haloxyfop TFE ester	14.38	443 (36)	316 (35)	288 (100)	180 (28)
MCPA TFE ester	5.67	282 (100)	155 (39)	141 (83)	124 (31)
Mecoprop TFE ester	4.40	296 (97)	169 (87)	141 (100)	107 (50)
Picloram TFE ester	13.91	322 (15)	223 (25)	196 (100)	170 (12)
2,4,5-T TFE ester	9.89	336 (14)	301 (34)	209 (100)	195 (68)
Triclopyr TFE ester	8.13	337 (30)	210 (100)	182 (26)	146 (23)

[a] Relative intensity of the ions in the full scan EI spectrum of the individual compound (base peak = 100).

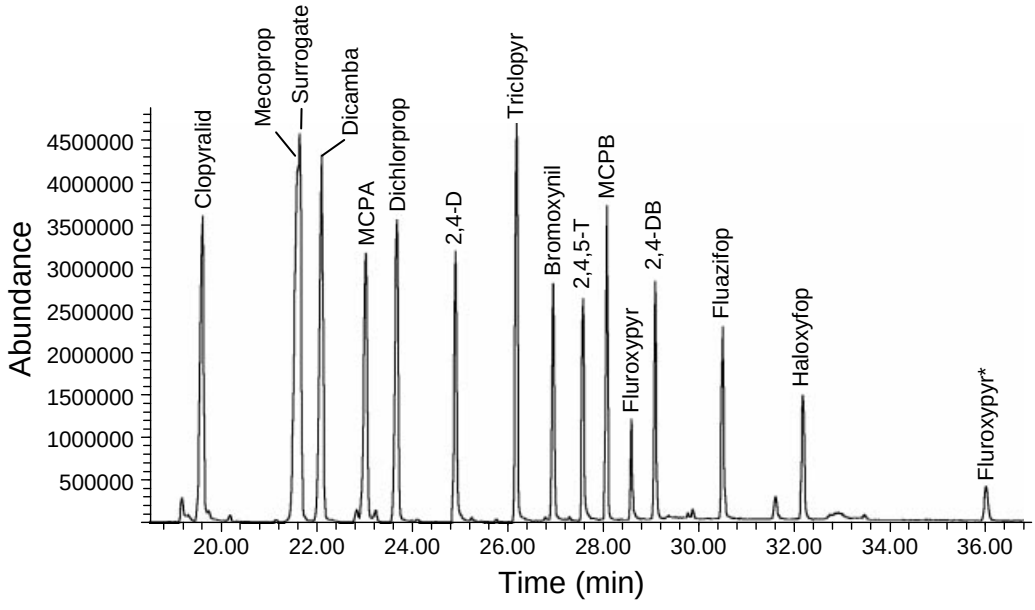

Figure 7 TIC chromatogram of a standard mixture of 15 acidic pesticides and the surrogate standard compound 2-(4-chlorophenoxy)butyric acid as their TBDMS derivatives. A 2.5-ng sample of each compound was injected and recorded applying GC/MS in full scan mode (mass range: m/z 50–500). * bis-silylated derivative.

Table 5 Retention times on an HP-5 capillary column, MWs and MS data of some acidic pesticides as their TBDMS derivatives

TBDMS derivative	MW ($g\,mol^{-1}$)	Retention time (min)	Mass 1 (rel. int.)[a]	Mass 2 (rel. int.)[a]	Mass 3 (rel. int.)[a]
Bromoxynil **(13)**	389	27.04	334 (100)	336 (52)	332 (49)
2-(4-Chlorophenoxy)butyric acid (Surrogate)	328	21.68	185 (100)	187 (37)	271 (40)
Clopyralid **(24)**	305	19.70	248 (100)	250 (68)	252 (14)
Chloroxynil **(12)**	301	–	244 (100)	246 (70)	93 (34)
2,4-D **(2)**	334	24.98	277 (100)	279 (69)	249 (46)
2,4-DB **(3)**	362	29.20	219 (100)	221 (75)	305 (28)
p, *p*′-DDA	394	–	337 (100)	339 (69)	165 (33)
Dicamba **(25)**	334	22.14	277 (100)	279 (73)	203 (39)
Dichlorprop **(4)**	348	23.73	291 (91)	293 (62)	245 (78)
Dinoseb **(15)**	354	–	297 (100)	281 (8)	
DNOC **(16)**	312	–	255 (100)	209 (14)	239 (9)
Fluazifop **(28)**	441	30.62	384 (100)	338 (36)	356 (14)
Fluroxypyr **(30)**	368	28.69	311 (100)	313 (70)	253 (88)
Fluroxypyr[b]	482	36.21	425 (100)	427 (76)	367 (27)
Haloxyfop **(31)**	475	32.31	418 (100)	372 (69)	420 (39)
Ioxynil **(14)**	485	31.73	428 (100)	301 (23)	
MCPA **(6)**	314	23.07	257 (100)	259 (41)	229 (54)
MCPB **(7)**	342	28.18	199 (100)	201 (86)	285 (34)
Mecoprop **(8)**	328	21.61	225 (100)	227 (41)	199 (36)
PCP **(18)**	378	–	323 (100)	325 (66)	93 (39)
2-Phenylphenol **(19)**	284	–	227 (100)	211 (73)	152 (15)
2,4,5-T **(9)**	368	27.67	311 (100)	313 (98)	285 (61)
Triclopyr **(40)**	369	26.27	314 (100)	312 (98)	254 (84)

[a] Relative intensity of the ions in the full scan EI spectrum of the individual compound (base peak = 100).
[b] Silylation of fluroxypyr resulted in a second bis-silylated product, because it was also derivatized at the amino group.

Table 6 SIM programming parameters and retention times for the determination of acidic pesticides as their PFB derivatives on an HP-5 capillary column applying GC/EIMS. Chromatographic conditions as described in section on Pentafluorobenzylation

Time window	PFB derivative	Peak no.	Retention time (min)	Time window (min)	Selected ions per window (*m*/*z*)	Dwell time per ion (ms)
1	Clofibric acid 2,4-Dichlorobenzoic acid	1 2	25.42 25.65	25–25.89	128, 130, 173, 181, 370, 372, 394	150
2	Clopyralid **(24)** 4-Chlorophenoxyacetic acid	3 4	26.18 26.28	25.90–26.34	141, 147, 175, 181, 308, 366, 368	150
3	Mecoprop **(8)**	5	26.43	26.35–26.89	142, 169, 181, 394	200
4	Dicamba **(25)** MCPA **(6)** Dichlorprop **(4)**	6 7 8	27.28 27.48 27.79	26.90–28.19	141, 162, 181, 203, 380, 382, 400, 402, 414, 416	100
5	Chlorflurenol-methyl Flurenol-butyl	9 10	28.31 28.40	28.20–28.55	152, 153, 181, 215, 217, 282	200
6	2,4-D **(2)** Chlorfenac **(23)** Bromoxynil **(13)** 1-Naphthyl acetic acid	11 12 13 14	28.72 28.93 29.07 29.17	28.56–29.35	115, 141, 161, 175, 181, 195, 366, 383, 400, 402, 418, 455, 457	100
7	Triclopyr **(40)**	15	29.56	29.36–29.74	181, 210, 435, 437	200
8	Fenoprop **(5)** Fluazifop-*p*-butyl	16 17	30.03 30.28	29.75–30.59	181, 196, 254, 282, 383, 448, 450	150
9	Flamprop-isopropyl 2,4,5-T **(9)** Bentazone **(10)**	18 19 20	30.92 31.12 31.17	30.60–31.30	105, 181, 211, 276, 341, 363, 378, 420, 434, 436	100
10	Indol-3-acetic acid MCPB **(7)** Chloramben **(22)**	21 22 23	31.45 31.54 31.75	31.31–31.84	130, 142, 161, 181, 267, 355, 356, 385 387, 408	100
11	Fluroxypyr **(30)** Flurenol **(29)**	24 25	31.94 32.14	31.85–32.39	152, 153, 181, 209, 406, 434, 436	150
12	2,4-DB **(3)** Indol-3-propionic acid	26 27	32.87 33.28	32.40–33.89	130, 162, 181, 267, 369, 370, 428	150
13	Fluazifop **(28)**	28	34.34	33.90–34.86	181, 254, 282, 507	200
14	Benazolin **(21)** Picloram **(33)**	29 30	35.22 35.48	34.87–36.09	170, 181, 196, 198, 420, 423	150
15	Haloxyfop **(31)** Indol-3-butyric acid	31 32	36.85 36.91	36.10–38.49	130, 181, 288, 316, 383, 384, 541	150
16	Flamprop **(27)**	33	39.64	38.50–40.99	105, 181, 276, 501	200
17	Acifluorfen **(20)**	34	42.96	41.00–45	181, 345, 511, 541	200

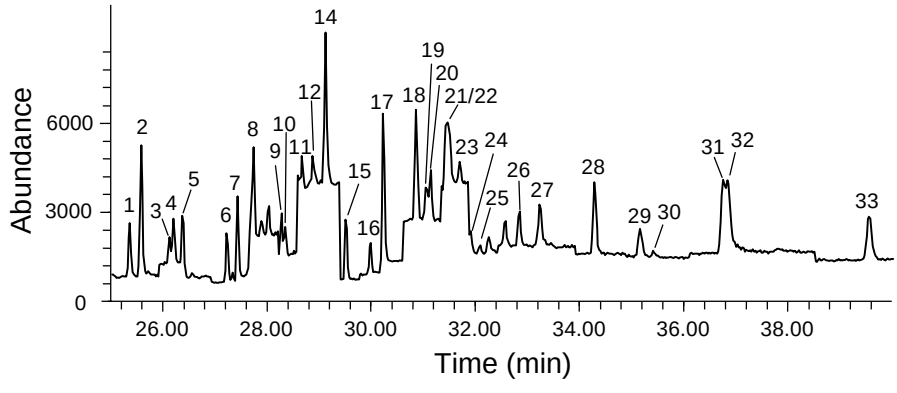

Figure 8 MID chromatogram recorded with GC/MS applying SIM programming of 34 acidic pesticides as their PFB derivatives (20 pg of each compound). For peak identification refer to Table 6. (Reprinted from Heberer et al.,[75] copyright (1994), with permission from AOAC International.)

in each window and the necessity of finding a sufficient gap in the chromatogram to enable window switching.[75] When using uniform dwell times within the individual SIM time windows the relative intensities of the various ions are comparable with the relative abundances in the full scan mass spectra as compiled in Tables 2–5. For analyte identification in environmental analysis, the peak area ratios of the indicative ions should match within ±20% with the relative abundances of the standard compounds.[41]

In trace analysis, the feature of an adjustable selectivity, as is possible using GC/MS with SIM acquisition, is very valuable and is not provided by a less selective detector such as an ECD.[75] This feature is shown in Figure 9 with the chromatogram of a pentafluorobenzylated derivatization blank. The MID chromatogram of the indicative ions (top) shows a flat baseline with only one small peak in the entire chromatogram, whereas the ion trace of the pentfluorobenzylium ion (*m*/*z* 181) which appears in the mass spectra of all derivatized compounds (analytes and matrix as well) shows strong interference. These peaks represent impurities from the reagents and by-products formed during derivatization. Thus, this ion should not be included in trace level analysis of target analytes. It may, however, be additionally recorded and used as a ''monitor ion'' to obtain an overview of the matrix content or it might help to avoid missing another important peak from a derivatized nontarget analyte.

3.1.3.3 Negative Chemical Ionization Negative chemical ionization (NCI) has been successfully used in environmental analysis for the selective and sensitive detection of electron capturing analytes such as organochlorine pesticides, polychlorinated biphenyls (PCBs) and dioxins.[89,90] Applying GC/MS with NCI appears to be especially promising for the analysis of PFB esters of acidic pesticides. Meiring et al.[73] described the analysis of 10 acidic pesticides as their PFB esters by applying GC with high resolution (HR) NCI/MS and SIM. Lee et al.[64] used GC/MS applying EI and NCI in water analysis to confirm 14 acidic herbicides as their PFB derivatives, and Vink and van der Poll[78] used NCI/MS to confirm eight acidic herbicides after GC/ECD analysis. Stan et al.[91] analyzed more than 30 PFB esters of acidic herbicides applying electron capture NCI using ammonia as reagent gas.

NCI is a so-called ''soft ionization'' technique which in most cases produces intensive molecular anions and shows only little fragmentation. This ionization behavior has some advantages because the molecular anions often

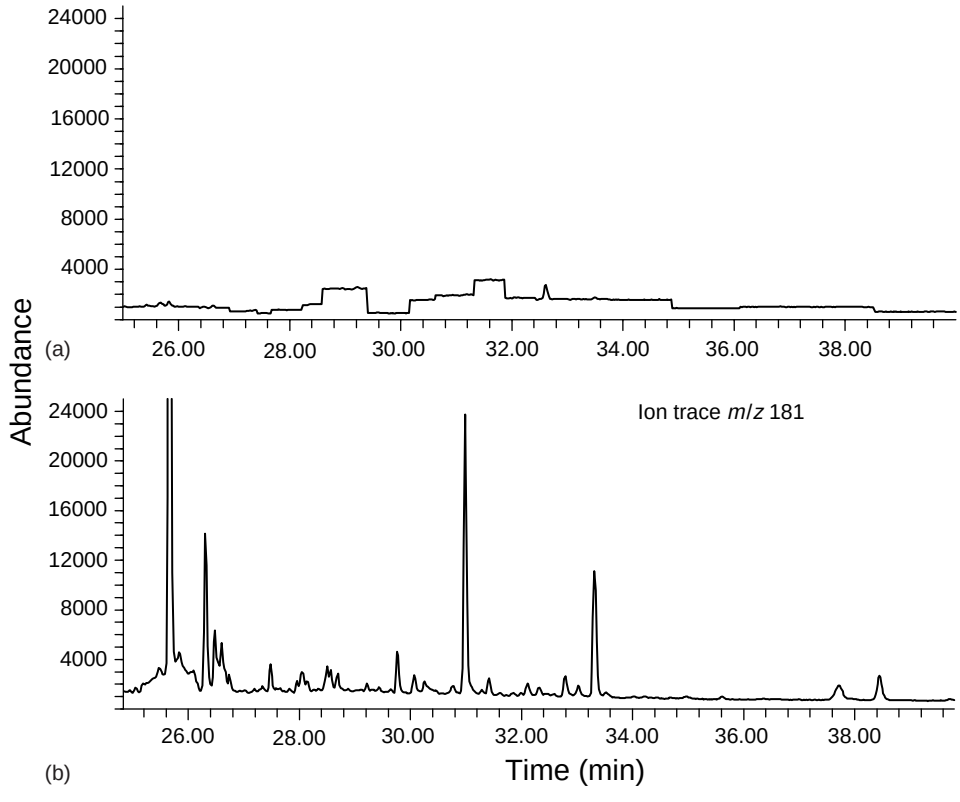

Figure 9 Derivatization blank of the pentafluorobenzylation. (a) MID chromatogram of the indicative ions. (b) Ion trace of *m*/*z* 181. (Reprinted from Heberer et al.,[75] copyright (1994), with permission from AOAC International.)

have an excellent signal-to-noise ratio. On the other hand, low fragmentation may become a problem in trace level analysis because there are no suitable confirmational ions. However, many of the acidic pesticides are substituted by chlorine atoms and therefore show the characteristic chlorine cluster ions, which may also be used to confirm the individual compound.

In the case of the PFB esters, the molecular anions are missing in the NCI mass spectra because the PFB moiety easily cleaves from the molecule during the dissociative electron capture process.[73,91] Thus, the base peaks in the mass spectra are formed by the anions of the original acidic pesticides $(M\text{-}181)^-$.[73,91] Table 7 shows the prominent anions in the ammonia NCI mass spectra of the PFB esters for more than 30 acidic herbicides.

When measuring standard compound solutions, GC/NCI/MS was found to be on average ten times more sensitive than conventional GC/EI/MS. In full scan mode, detection limits between 10 and 100 pg per compound injected were determined.[91] In SIM mode, mecoprop PFB ester could be detected down to 500 fg. For all compounds compiled in Table 7, the detection limits in SIM mode were estimated to be between 0.5 and 5 pg injected.[91]

Unfortunately, these results could not be confirmed when measuring real water samples. The detection limits of GC/NCI/MS with SIM were found to be more than ten times higher than those determined with GC/EI/MS with SIM.[91] In full scan mode the detection limit was on average 100 ng L^{-1}, which is similar to or slightly better than that of full scan mode applying EI/MS. However, in

Table 7 Analysis applying GC/NCI/MS using ammonia as collision gas. Retention times on an HP-5 capillary column, MWs and MS data of some acidic pesticides as their PFB derivatives according to Stan et al.[91]

Common name	MW ($g\,mol^{-1}$)	$[M\text{-}181]^-$ (rel. int.)[a]	Other fragments (rel. int.)[a]
Acifluorfen PFB ester	541	360 (100), 362 (33)	330 (60), 322 (22), 316 (30), 318 (10)
Benazolin PFB ester	423	242 (100)	198 (50)
Bentazone PFB amide	420	239 (100), 240 (12)	–
Bromoxynil PFB ether	455	274 (55), 276 (100), 278 (50)	198 (22), 196 (23), 118 (13), 79 (11), 81 (10)
Chloramben PFB ester	385	204 (100), 206 (66), 208 (11)	–
Chlorfenac PFB ester	418	–	193 (100), 195 (98), 197 (33)
Chlorflurenol PFB ester	440	–	214 (23), 215 (100), 217 (33), 199 (12)
2-(4-Chlorophenoxy)butyric acid PFB ester	394	213 (100), 215 (33)	127 (22)
Clofibric acid PFB ester	394	213 (100), 215 (33)	–
Clopyralid PFB ester	371	190 (100), 192 (66)	–
2,4-D PFB ester	400	219 (100), 221 (66), 223 (11)	–
2,4-DB PFB ester	428	247 (45), 249 (30)	161 (100), 163 (68), 165 (11)
Dalapon PFB ester	322	141 (100), 143 (66), 145 (11)	107 (60), 109 (20)
Dicamba PFB ester	400	219 (100), 221 (66), 223 (11)	175 (55), 177 (33)
2,4-Dichlorobenzoic acid PFB ester	370	189 (100), 191 (66), 193 (11)	–
Dichlorprop PFB ester	414	233 (100), 235 (66), 237 (11)	–
Endothal-methyl PFB ester[b]	380	199 (100)	–
Fenoprop PFB ester	448	267 (100), 269 (98), 271 (33)	–
Flamprop PFB ester	501	320 (100), 322 (33)	–
Fluazifop PFB ester	507	326 (100)	–
Flurenol PFB ester	406	–	181 (100)
Fluroxypyr PFB ester	434	253 (100), 255 (66), 257 (11)	–
Haloxyfop PFB ester	541	360 (100), 362 (33)	–
4-Indol-3-yl-butyric acid PFB ester	383	202 (100)	–
Indol-3-yl-acetic acid PFB ester	355	174 (100)	–
Indol-3-yl-propionic acid PFB ester	369	188 (100)	–
MCPA PFB ester	380	199 (100), 201 (33)	–
MCPB PFB ester	408	227 (100), 229 (33)	141 (53), 143 (17)
Mecoprop PFB ester	394	213 (100), 215 (66)	–
2-(1-Naphthyl)acetic acid PFB ester	366	185 (100)	141 (51)
Picloram PFB ester	420	239 (100), 241 (97), 243 (60)	384 (30), 386 (21), 320 (12), 322 (8)
2,4,5-T PFB ester	434	253 (100), 255 (98), 257 (33)	–
2,3,6-TBA PFB ester	404	223 (100), 225 (98), 227 (33)	179 (30), 181 (29), 183 (10)
Triclopyr PFB ester	435	254 (100), 256 (98), 258 (33)	–

[a] Relative intensity of the ions in the full scan EI spectrum of the individual compound (base peak = 100).
[b] Rearrangement reactions result in the mono methyl PFB ester.

contrast to EI/MS with SIM, no significant improvement of the detection limits was observed for NCI/MS with SIM.[91] The reason for this phenomenon lies in the selection of suitable anions for SIM programming. The masses of the prominent NCI anions of the PFB esters are much lower than those of the EI ions. Unfortunately, in environmental sample extracts the mass spectral background between m/z 200 and m/z 340 was found to be rather high in NCI/MS.[91] The effective signal-to-noise ratio of the SIM anions was therefore much worse than that observed for the EI/SIM ions of the PFB esters. Thus, Lee et al.[64] recommended extensive cleanup procedures in sample preparation and Meiring et al.[73] recommended the use of GC with HR NCI/MS for low-level analysis of dirty samples after derivatization with PFBBr. Owing to its selectivity HR/NCI/MS is not disturbed by coeluting matrix compounds. However, HR/MS instruments are rather expensive and need well-trained personnel. These instruments are therefore unlikely to be used for routine analysis of environmental samples. Another less expensive alternative is probably provided by low resolution NCI/MS/MS detection using ion-trap MS instruments. This technique has, however, not yet been described for the analysis of acidic pesticides.

3.1.3.4 Tandem Mass Spectrometry Measurements

Another promising analytical approach for the sensitive and highly selective determination of acidic pesticides in environmental samples is the application of GC with MS/MS. For a long time, MS/MS was only possible using highly sophisticated, very expensive MS instruments such as triple stage quadrupole mass spectrometers. Today, MS/MS recording is, with a few restrictions, also applicable using relatively inexpensive ion trap mass spectrometers.

In MS/MS acquisition a ''parent ion'' from the original mass spectrum is selected and fragmented again by collision induced dissociation (CID) forming a daughter ion spectrum. When a suitable parent ion for the individual analyte is selected, the resulting daughter ion spectra are highly selective and unlikely to be disturbed by interfering compounds. The daughter ions mostly exhibit better signal-to-noise ratios, further enhancing the sensitivity of GC/MS analysis. In SRM mode only one or a few daughter ions are monitored for the individual compound. Thus, SRM gives maximum sensitivity and selectivity and, similar to SIM, a time-scheduled SRM programming for multiresidue analysis may be developed. The hardware and operation of MS/MS acquisition are described in more detail elsewhere.[24,25]

MS/MS is already used in routine analysis to confirm positive analytical results and a few laboratories also make attempts to use MS/MS programming for multiresidue analysis of acidic pesticides using GC with ion trap MS. Such methods have not yet been validated over a longer time period in routine analysis, nor have they been published.

3.1.4 Other Detection Methods

3.1.4.1 Gas Chromatography with Nitrogen–Phosphorous Detection GC/NPD is frequently applied in environmental analysis to the determination of nitrogen or phosphorus containing pesticides such as triazine herbicides or organophosphorous insecticides. However, the use of GC/NPD for the determination of acidic pesticides has only been described by Bertrand et al.[81,82] Acidic

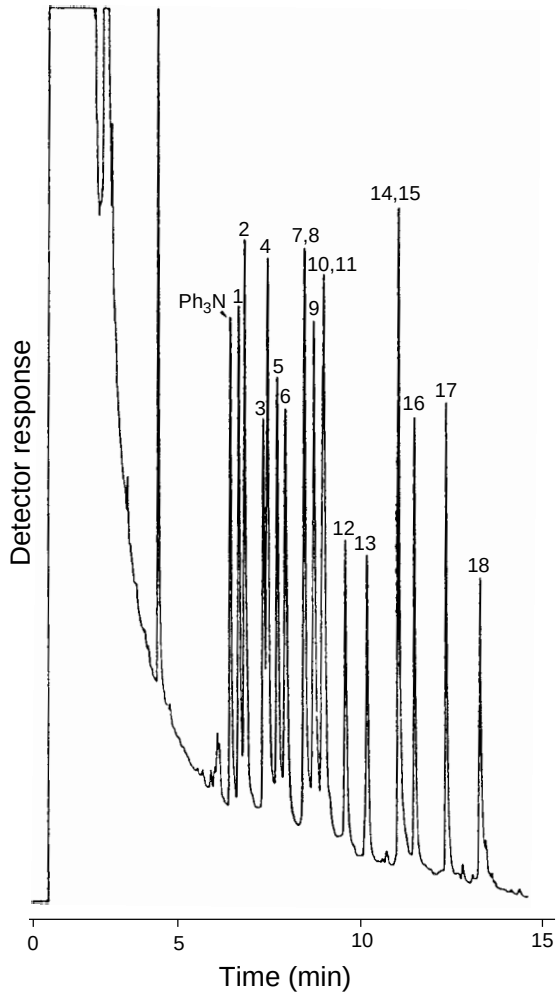

Figure 10 NPD chromatogram of a mixture of acidic herbicides derivatized with CEDMSDEA. Compounds: 1, mecoprop **(8)**; 2, dicamba **(25)**; 3, MCPA **(6)**; 4, 2,3,6-TBA **(39)**; 5, dichlorprop **(4)**; 6, 2,6-D; 7, 2,4-D **(2)**; 8, 2,5-D; 9, 2-(1-naphthyl)acetic acid **(32)**; 10, 2,3-D; 11, 3,4-D; 12, (4,6-dichloro-2-methylphenoxy) acetic acid; 13, fenoprop **(5)**; 14, chloramben **(22)**; 15, 2,4,5-T **(9)**; 16, MCPB **(7)**; 17, benazolin **(21)** and 18, 2,4-DB **(3)**. (Reprinted with permission from Bertrand et al., *Anal. Chem.*,[81] copyright (1987), by American Chemical Society.)

pesticides are originally unsuitable for NPD analysis, but derivatization with CEDMSDEA (see section on Silylation with (2-cyanoethyl)dimethyl(diethylamino)silane) introduces nitrogen into the molecules which renders the derivatives suitable for NPD analysis. Figure 10 shows the resulting GC/NPD chromatogram of a derivatized standard mixture.

3.1.4.2 Gas Chromatography with Atomic Emission Detection AED is a highly sophisticated and still very expensive technique which provides high element specific detection of all compounds amenable to GC. In theory, a GC/AED can detect any element in the periodic table except for helium.[92] Unlike the ECD, the AED can also distinguish between fluorine, chlorine, bromine, and iodine. As most modern pesticides contain hetero atoms such as sulfur, phosphorus, nitrogen, chlorine, bromine or fluorine, GC/AED is most suitable for pesticide analysis. However, for practical reasons, it is not always possible to determine all the element traces in one single GC run. Thus, carbon, hydrogen, chlorine, and bromine may be detected in a single GC run or alternatively carbon, nitrogen and sulfur can be monitored simultaneously.[92] Since molecules are fragmented into their constituting elements, these can be quantified individually by external calibration. Thus it is possible to estimate the sum formula of an unknown compound.[92] Several excellent reviews have been written which describe the hardware, the techniques, some applications and limitations of the AED.[92–98]

A major drawback of the AED in trace level analysis is the lack of sensitivity, especially for the important nitrogen or oxygen traces.[93] This drawback may be overcome by using large volume injection techniques such as PTV. Stan and Linkerhägner[93] demonstrated the analysis of 22 acidic compounds including 19 pesticides as their PFB esters at trace level concentrations applying GC/AED. Table 8 compiles the retention times, molecular formulae and detection limits in water analysis for the individual element traces applying GC/AED with PTV injection of 12.5 µL of a derivatized sample extract. The samples (1 L each) have been prepared according to the procedure described in section 4.1.1, therefore the amount injected using PTV corresponds to an aliquot of 125 mL of a water sample. Figure 11 shows a set of element-specific chromatograms obtained for a derivatized standard solution containing the analytes. All analytes are detected with the element traces reflecting their elemental composition. The element traces

Table 8 Compilation of retention times, elemental formulae and LODs in the individual element hetero atom traces of some PFB derivatives of acid herbicides according to Stan and Linkerhägner[93]

Compound name	Peak no.	Retention time (min) PTV	Elemental formula	LOD (ppt = ng L^{-1})[a]					
				N	O	S	F	Cl	Br
Clofibric acid PFB ester	1	20.61	$C_{17}H_{12}ClF_5O_3$		40		30	30	
2,4-Dichlorobenzoic acid PFB ester	2	20.95	$C_{14}H_5Cl_2F_5O_2$		40		30	15	
Clopyralid PFB ester	3	21.80	$C_{13}H_4Cl_2F_5NO_2$	100	40		20	15	
Mecoprop PFB ester	4	21.97	$C_{17}H_{12}ClF_5O_3$		40		30	30	
2-(4-Chlorophenoxy)butyric acid PFB ester	5	22.33	$C_{17}H_{12}ClF_5O_3$		40		30	30	
Dicamba PFB ester	6	23.33	$C_{15}H_7Cl_2F_5O_3$		40		30	15	
MCPA PFB ester	7	23.67	$C_{16}H_{10}ClF_5O_3$		40		30	30	
Dichlorprop PFB ester	8	24.14	$C_{16}H_9F_5Cl_2O_3$		50		30	30	
2,4-D PFB ester	9	25.93	$C_{15}H_7Cl_2F_5O_3$		40		30	15	
Bromoxynil PFB ether	10	26.30	$C_{14}H_4F_5Br_2NO$	100	150		50		20
2-(1-Naphthyl)acetic acid PFB ester	11	26.50	$C_{19}H_{11}F_5O_2$		40		30		
Triclopyr PFB ester	12	27.22	$C_{14}H_5Cl_3F_5NO_3$	100	40		30	10	
Fenoprop PFB ester	13	27.78	$C_{16}H_8Cl_3F_5O_3$		40		30	15	
Bentazone PFB amide	14	29.04	$C_{17}H_{13}F_5N_2O_3S$	60	40	10	40		
MCPB PFB ester	15	29.53	$C_{18}H_{14}ClF_5O_3$		40		30	25	
Fluroxypyr PFB ester	16	29.87	$C_{14}H_6Cl_2F_6N_2O_3$	60	70		30	25	
2,4-DB PFB ester	17	30.61	$C_{17}H_{11}Cl_2F_5O_3$		50		80	15	
Fluazifop PFB ester	18	31.84	$C_{22}H_{13}F_8NO_4$	150	70		30		
Benazolin PFB ester	19	32.13	$C_{16}H_7ClF_5NO_3S$	150	70	10	30	35	
Haloxyfop PFB ester	20	33.51	$C_{22}H_{12}ClF_8NO_4$	200	60		40	45	
Flamprop PFB ester	21	35.13	$C_{23}H_{14}ClF_6NO_3$	200	100		50	50	
Acifluorfen PFB ester	22	37.31	$C_{21}H_8ClF_8NO_5$	250	100		50	60	

[a] LOD (limit of detection) of the multiresidue methods using SPE with RP-C18 (reversed-phase octadecyl) adsorbent as described in section 4.1.1.

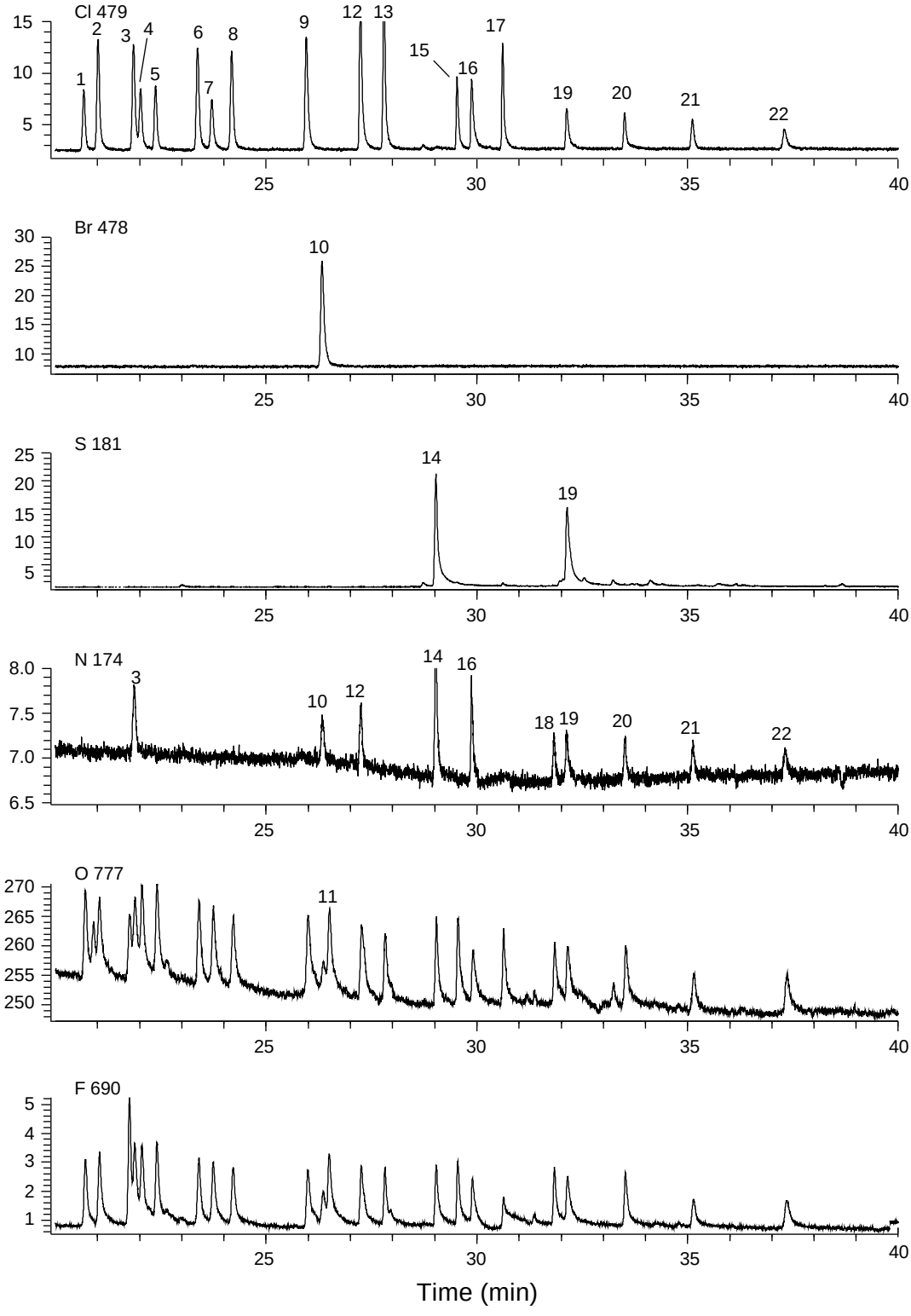

Figure 11 GC/AED chromatogram of the hetero element traces of an acidic herbicide standard mixture (1 ng µL^{-1}) derivatized with PFBBr; 12.5-µL cold split PTV-injection with solvent venting. Peak numbering as given in Table 8. (Reproduced from Linkerhägner[98] by permission of Shaker Verlag.)

of chlorine, bromine and sulfur show the AED to be much more sensitive to these than to nitrogen, oxygen or fluorine.[93] Thus the AED seems to be well suited to trace level analysis of compounds containing these hetero atoms, including most acidic pesticides. Other compounds such as 2-(1-naphthyl)acetic acid **(32)** are much more difficult to analyze as can also be seen in Figure 11. Here, 2-(1-naphthyl)acetic acid PFB-ester (peak no. 11) can only be detected in the oxygen or fluorine trace, both of which are heavily disturbed by matrix peaks when analyzing pentafluorobenzylated extracts of real water samples.

3.2 High-performance Liquid Chromatography

HPLC is well suited to the direct and nondestructive analysis of contaminants which are originally not amenable to GC analysis such as thermally unstable and highly polar pesticides or conjugated metabolites.[33] HPLC analysis does not require any preliminary derivatization steps for the separation of acidic pesticides, and pesticide metabolites can often easily be included in the analytical method,[99] see section 3.2.1. HPLC indeed appears to be the most appropriate technique for the determination of these analytes. HPLC is, however, less useful for screening purposes in multiresidue analysis than GC, mainly because of its relatively low separation efficiency.[73] Matrix compounds such as humic acids often interfere because they co-elute with the polar analytes.[100]

In the past, a major drawback of HPLC was the difficulty in coupling LC (liquid chromatography) to sensitive and selective detectors. Progress in coupling LC to MS has improved the possibilities of identification and confirmation of analytes at trace level concentrations.[33] Some applications of HPLC to the analysis of acidic pesticides will be described in the following sections.

3.2.1 Conventional Detection Methods (High-performance Liquid Chromatography with Ultraviolet, Diode Array Detection, Fluorescence Detection and Electrochemical Detection)

HPLC with conventional ultraviolet (UV) or diode array detection (DAD) is one of the most frequently applied methods for the analysis of acidic pesticides in water samples.[56,101–115] Additionally, fluorescence detection (FLD)[106,110,116,117] and electrochemical detection (ELCD)[106] have been used to lower the detection limits of HPLC analysis. Figure 12 shows the separation of a standard mixture of acidic pesticides, including bentazone **(10)** and two of its hydroxylized metabolites using HPLC with DAD.

A major drawback of using UV detection or DAD is that both are often prone to matrix interferences.

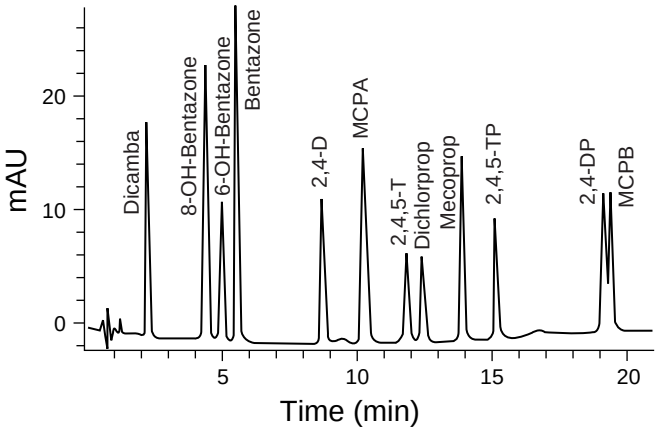

Figure 12 HPLC/DAD chromatogram of a standard mixture containing acidic herbicides and 6- and 8-hydroxy-bentazone. (Reproduced from Schuster and Gratzfeld-Hüsgen[118] by permission of Hewlett Packard.)

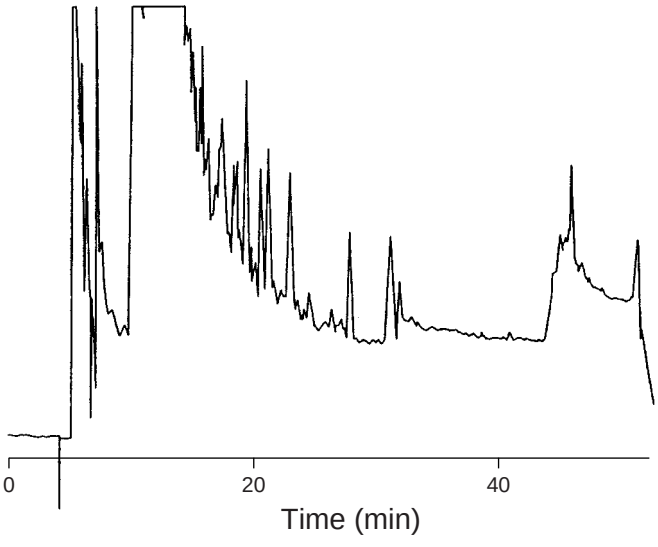

Figure 13 Typical HPLC/UV chromatogram of an extract of a surface water sample. The peaks correspond to solute concentrations up to the micrograms per liter level. (Reproduced from Noij and Brandt[100] by permission of Springer-Verlag (1995).)

At wavelengths below 300 nm a huge background level from humic substances and other naturally occurring dissolved organic compounds hampers trace analysis of polar analytes.[100] Figure 13 shows a typical HPLC chromatogram of a water sample where the early eluting polar analytes are completely masked by equally polar humic substances. Additionally, in HPLC analysis with UV detection, there is a tremendous risk of false positive results because peak identification is virtually based only on retention times.[100] This latter statement also applies to HPLC analysis using DAD spectra, because these often suffer from interference at lower wavelengths. In

general, the information content of DAD spectra is not comparable to MS spectra. In conclusion, HPLC with UV or DAD detection often does not provide the sensitivity and the degree of confidence which is absolutely necessary in acidic pesticide monitoring.

3.2.2 High-performance Liquid Chromatography with Mass Spectrometry and Tandem Mass Spectrometry

A variety of HPLC/MS techniques using thermospray (TSP) and particle beam (PB) interfaces have been examined in order to develop a rugged system with the necessary sensitivity and selectivity for trace level analysis of acidic pesticides in the environment.[110,119–123] However, it was found that the behavior and sensitivity of these techniques may vary considerably depending not only on chemical and physical properties of the individual pesticide, but also on instrumental and analytical conditions.[73] Another major limitation of LC/MS methods such as TSP/MS using CI processes is the lack of information for MS confirmation. Thus, Geerdink et al.[122] used TSP MS/MS applying the SRM mode to overcome this problem and to lower detection limits. PB interfaces allow both chemical and electron ionization. As discussed in the GC section, EI mass spectra often contain many fragments which are suitable for analyte confirmation purposes. A major drawback of PB interfaces is, however, the lack of sensitivity, which has been overcome to a certain degree using specially designed micro HPLC/PB/MS instruments.[123] Figure 14 shows the total ion PB/MS chromatogram recorded with EI for a standard and a soil sample spiked with acidic pesticides at a concentration of 33.3 ppm. It should be noted that both ionization techniques (PB- and TSP/MS) require much experience, method validation and time for instrument maintenance.

Newer LC/MS interface types, especially API interfaces increased sensitivity and made instrument maintenance and analysis much easier. API is a term covering two different principles of ionization, electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI). ESI is a soft ionization technique which typically generates only ionized molecular ions $[M + H]^+$ (in positive mode) and $[M - H]^-$ (in negative mode). The latest developments, fundamentals, instrumentation and applications of ESI/MS have been described in a monograph by Cole.[124] The first applications of ESI- and APCI-interfaces for the analysis of acidic pesticides applying HPLC/MS with time-scheduled SIM have been described.[125–129] Figure 15 shows a TIC chromatogram recorded with HPLC/ESI/MS of a river water sample spiked with a mixture of acidic pesticides. In Figure 16 the selected ion current chromatograms for selected

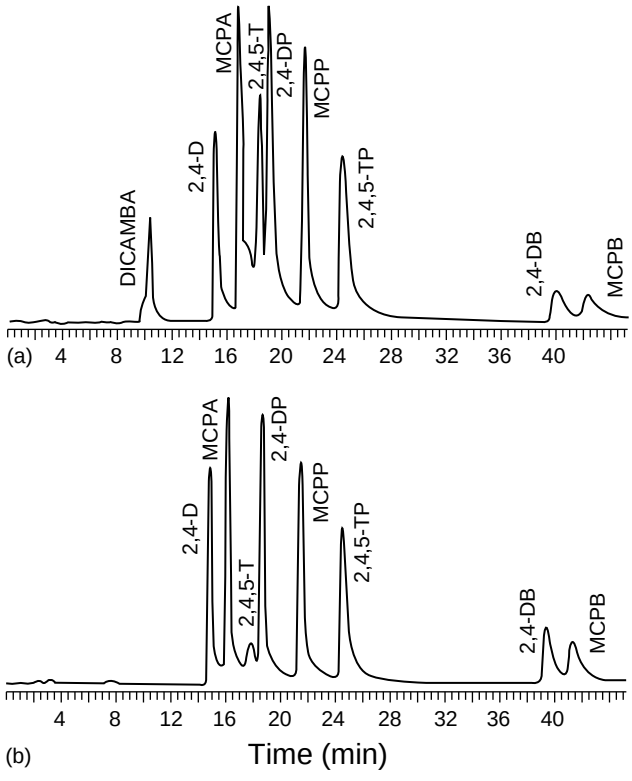

Figure 14 TIC chromatogram recorded with HPLC/PB/EI/MS showing (a) the separation of nine acidic herbicides, chlorophenoxy acid standards (125 ppm injected) and (b) the extract of a soil sample spiked at 33.3 ppm with eight phenoxy acid herbicides. (Reprinted with permission from Kim et al., *Anal. Chem.*,[121] copyright (1991), by American Chemical Society.)

ions extracted from the chromatogram in Figure 15 are shown.

Aguilar et al.[129] compared the performance of PB and APCI interfaces for the analysis of several pesticides including some acidic pesticides and concluded that HPLC/APCI/MS is much more sensitive than HPLC/PB/MS.

However, the HPLC/MS instruments are still much more expensive than conventional GC/MS instruments and simple LC/MS is often not sufficient for totally reliable identification of the analytes using low resolution MS. Thus, the use of hyphenated and expensive LC/MS/MS instruments is recommended[130] but not reasonable in routine analysis as long as cheaper alternatives are available.

3.3 Other Techniques

3.3.1 Capillary Zone Electrophoresis

Several of the phenoxy acid herbicides are chiral, e.g. those with the phenoxy substituent on the 2-position of propionic acid, and therefore exist as a pair (racemic

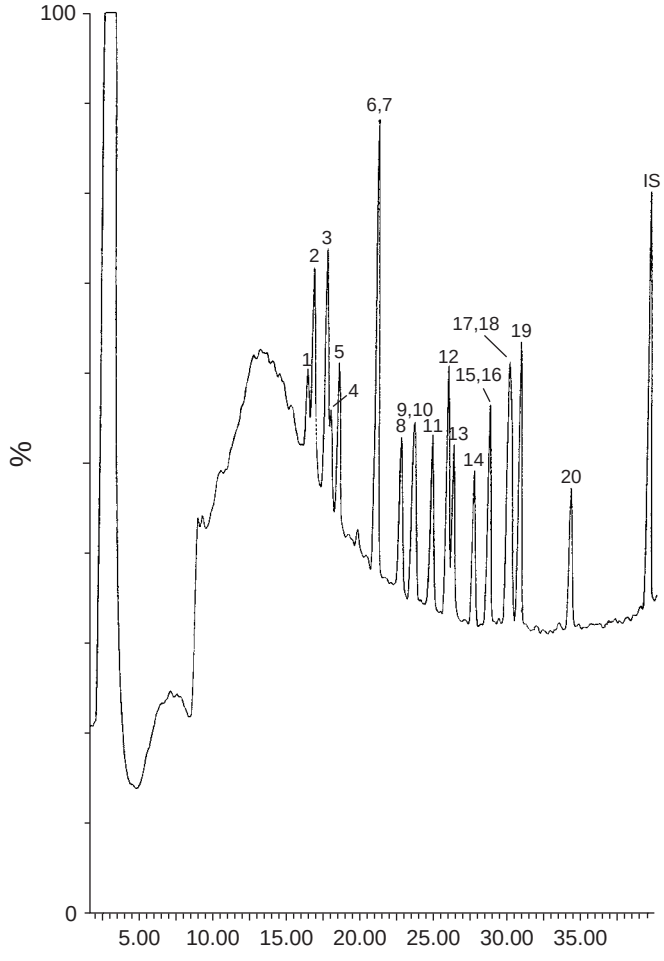

Figure 15 TIC chromatogram recorded with HPLC/ESI/MS of an extract of a river water sample fortified with acidic herbicides at a concentration of 0.5 μg L^{-1}: 1, dicamba **(25)**; 2, bentazone **(10)**; 3, 2,4-dinitrophenol; 4, benazolin **(21)**; 5, coumafuryl; 6, bromoxynil **(13)**; 7, DNOC **(16)**; 8, warfarin; 9, MCPA **(6)**; 10, 2,4-D **(2)**; 11, ioxynil **(14)**; 12, mecoprop **(8)**; 13, dichlorprop **(4)**; 14, 2,4,5-T **(9)**; 15, MCPB **(7)**; 16, 2,4-DB **(3)**; 17, dinoseb **(15)**; 18, fenoprop **(5)**; 19, dinoterb; 20, PCP. (Reprinted with permission from Crescenzi et al., *Anal. Chem.*,[125] copyright (1995), by American Chemical Society.)

mixture) of optical isomers called enantiomers.[131] It has been recognized that only the *R* enantiomers show herbicidal activity and since the 1980s enantiopure products have been developed, registered and now replace the earlier used racemic mixtures in many countries.[50] On the other hand, it has been shown that enantiomeric selectivity is also involved in biotic pesticide degradation processes.[50,79,131] Using high-performance capillary electrophoresis (HPCE) or CE it is possible to distinguish between the different enantiomers of phenoxy acid herbicides. Figure 17 shows the electropherograms of the enantiomeric separation of some acidic herbicides.

Garrison et al.[131] called the phenoxy acids "excellent candidates" for separation by CE, because their pK_a values allow separation by the simplest form of CE called capillary zone electrophoresis (CZE) which separates only charged species. However, enantioselective separation of different acidic pesticide derivatives has also been described applying HR GC/MS[50,56,79] and the enantiomers of mecoprop **(8)** and dichlorprop **(4)** have also been separated by HPLC using columns with chiral solid phases.[56,132]

Several applications of CE for the analysis of acidic herbicides have been reported.[131,133–141] A major drawback of CE analysis is, however, the insufficient sensitivity of commonly used CE techniques applying only to small sample volumes and UV detection[137] which make them unsuitable for trace level analysis of acidic pesticides in environmental samples. Nielen[134] applied CE with field-amplified injection after SPE for the analysis of phenoxy acid herbicides in water at sub parts per billion levels. Jung and Brumley[137] as well as Mechref and Rassi[139] produced fluorescent derivatives of phenoxy acid herbicides and applied CE with laser-induced fluorescence (LIF) detection for trace level analysis down to 0.2 ppb in water samples. Although the detection limits of these CE methods are still much higher than those of other methods, e.g. GC/MS, CE adds an interesting separation tool to the more conventional chromatographic instrumentation. Furthermore, the coupling of CE to API/MS (MS/MS) is a new and promising approach for the trace level analysis of acidic pesticides. Compared to UV detection, MS/MS detection is not impaired by matrix interferences and will also enhance the level of confidence in analyte identification.

3.3.2 Automated Multiple Development Thin-layer Chromatography

AMDTLC is a modern variation of classical TLC which offers much higher separation power than conventional TLC methods. In several reviewing articles the principles, hardware and some applications of AMDTLC in environmental analysis have been described.[142–144] Conventional TLC is not sufficient to separate pesticides from complex matrix compounds as found in environmental samples.[142] AMDTLC instruments are fully automated to process several chromatographic runs consecutively. Drying steps are included between the different TLC runs to remove the mobile phases of the previous runs from the TLC chamber and plate. Thus, it is possible to change the solvents from one run to the other slowly to produce a solvent gradient comparable to HPLC, but in contrast to HPLC it is also possible to change the conditions dramatically, e.g. from acidic to basic eluents. In AMDTLC, the efficiency of the separation

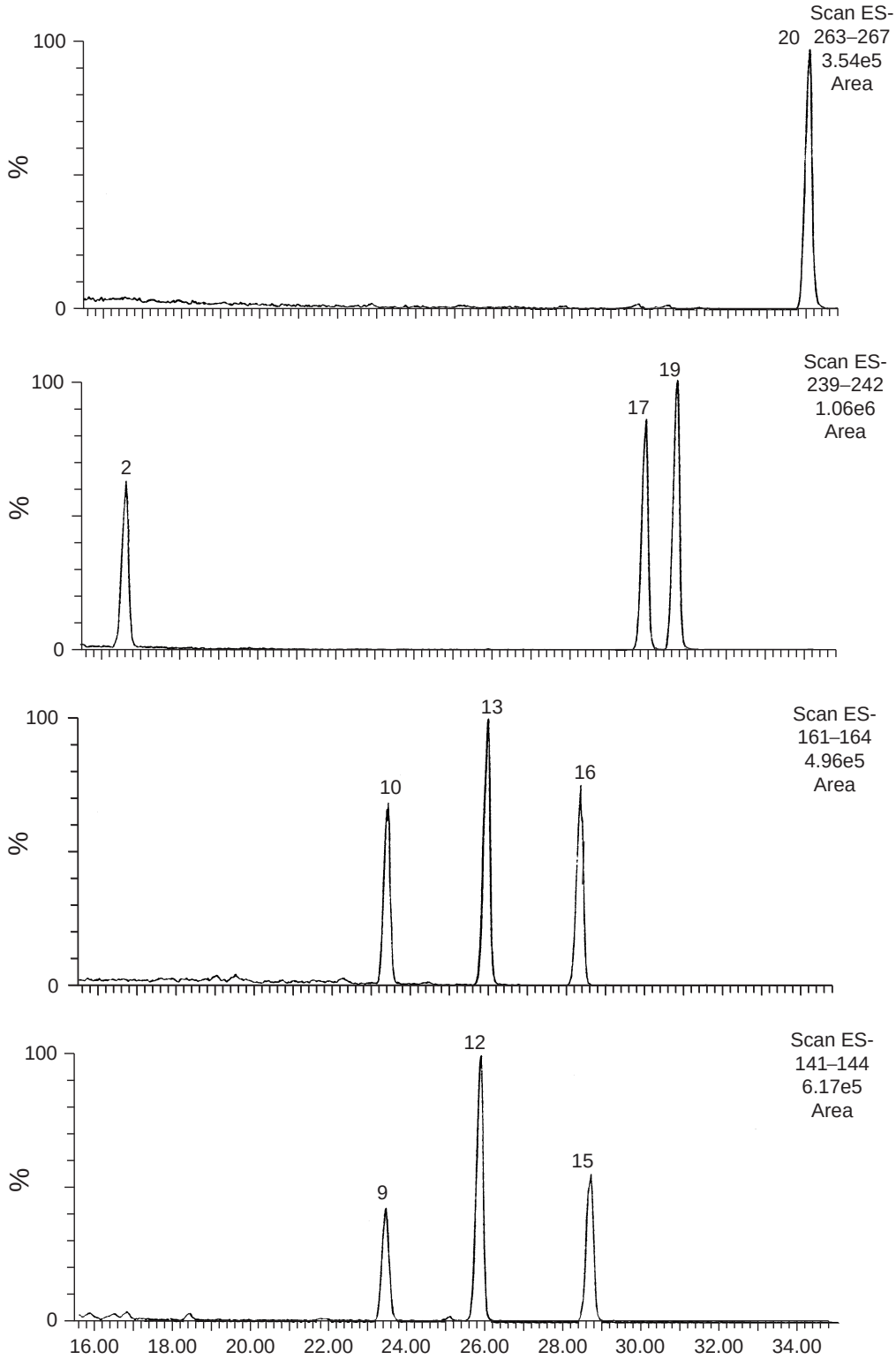

Figure 16 Selected ion current chromatograms for selected *m*/*z* values extracted from the TIC chromatogram in Figure 15. (Reprinted with permission from Crescenzi et al., *Anal. Chem.*,[125] copyright (1995), by American Chemical Society.)

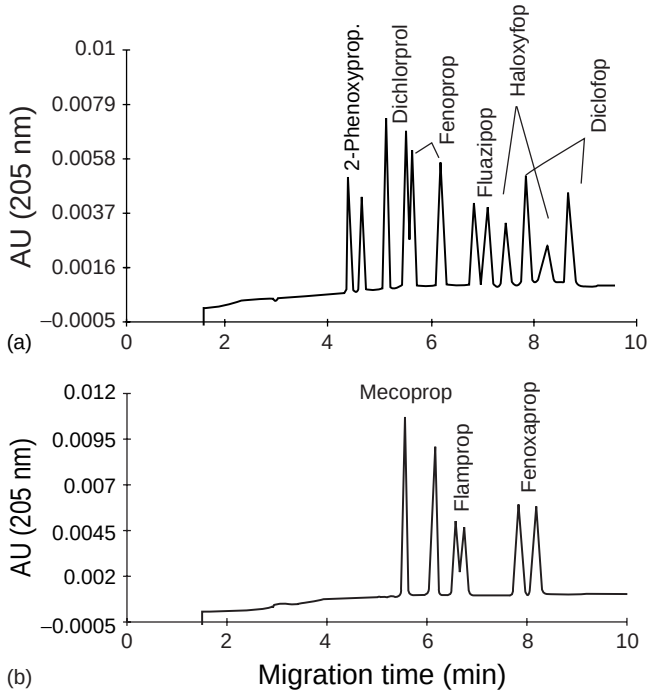

Figure 17 Electropherograms recorded using CE of the enantiomeric separation of some acidic herbicides. (Reprinted from Desiderio et al., *J. Chromatogr., A*, copyright (1997),[141] with permission from Elsevier Science.)

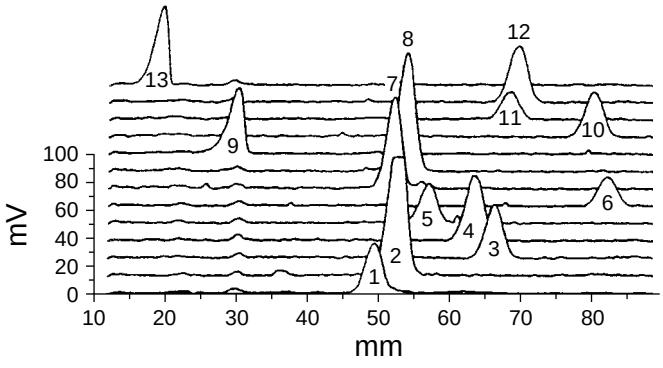

Figure 18 AMDTLC chromatogram showing the separation of 13 acidic herbicides. 1, 2,4-D **(2)**; 2, 2,4,5-T **(9)**; 3, fenoprop **(5)**; 4, dichlorprop **(4)**; 5, MCPA **(6)**; 6, MCPB **(7)**; 7, acifluorfen **(20)**; 8, fluroxypyr **(30)**; 9, clopyralid **(24)**; 10, 2,4-DB **(3)**; 11, mecoprop **(8)**; 12, triclopyr **(40)**; 13, picloram **(33)**. (Reproduced from Butz[85] by permission of Wissenschaft & Technik Verlag.)

is increased remarkably compared to common isocratic elution because the total distance migrated by the solvent front is in the meter range when all individual TLC runs are summed up.[142] The most important feature of AMDTLC which enhances the separation power of this technique tremendously is the so-called ''focusing effect''. During each run, the lower edge of a chromatographed spot is reached first by the solvent front and it begins to migrate to the upper edge of the spot which cannot move until it is wetted by the solvent. Thus, diffusion of the compound spots is suppressed and the analytes are concentrated in sharp peak bands on the TLC plate. It is a prerequisite that the AMD elution gradient always starts with a polar eluent (e.g. methanol) and ends with a nonpolar eluent (e.g. hexane) when using normal-phase silica gel TLC plates. The developed TLC plates are analyzed using a TLC scanner which allows the recording of UV spectra that can be used for confirmation of the analytes.[143]

Butz and Stan[144] applied AMDTLC for the screening in water of 265 pesticides including some acidic herbicides. For the analysis of acidic pesticides, detection limits between 5 and 15 ng have been reported for this method.[143] Figure 18 shows the AMDTLC chromatogram of a pesticide mixture. Using sample preparation techniques with concentration factors up to 10 000, as will be described in section 4.1.1, it is also possible to meet the EU (European Union) maximum tolerances of 100 ng L^{-1} per pesticide in drinking water.[1]

The AMDTLC technique has, however, some drawbacks:

1. The separation power is not comparable to GC.
2. UV detection is not as sensitive as other detection methods, e.g. GC/ECD, GC/MS or HPLC/MS/MS.
3. UV detection is prone to matrix interferences, e.g. by humic acids (as described in section 3.2.1).
4. UV spectra are not comparable to EI mass spectra with respect to information content or selectivity.
5. In contrast to GC, HPLC or CE columns, TLC plates represent an open chromatographic system which requires special care in handling.
6. AMDTLC equipment is, despite its apparent simplicity, relatively expensive.

A new approach is the on-line coupling of HPLC and AMDTLC[6] which has not yet been applied to acidic pesticides. In combination, the separation power of both techniques can be enhanced remarkably, but some of the drawbacks mentioned, such as problems in analyte detection, remain.

4 ENVIRONMENTAL APPLICATIONS

Environmental analysis always consists of three important steps, namely sample collection and handling, sample preparation and finally instrumental analysis. All these steps are equally important and errors in sample collection cannot be corrected in the laboratory. Thus, proper collection and handling of samples is an essential part of producing meaningful analytical results, especially at

trace level analyte concentrations.[145] As far as sample preparation is concerned, it has two functions in environmental analysis: (1) extraction of the analytes from the sample matrix and (2) enrichment of the analytes to a concentrated extract. All parts of the analytical procedure have to be combined, optimized and validated to achieve good laboratory practice (GLP). Sample preparation procedures which can be combined with any of the above instrumental analytical methods are described in the following sections. Nevertheless, some recommendations will help to enable the trace level determination of the acidic pesticides with the highest possible level of confidence, independent of the sample matrix.

4.1 Analysis of Water Samples

In Europe, the identification and quantitative analysis of acidic pesticides at a concentration level of 100 ng L^{-1} in drinking water for a single compound is required to meet the maximum tolerances set by the European Community's drinking water regulation.[1] In order to determine acidic pesticides at such low tolerance levels, the target compounds have to be extracted and enriched from the water sample with high recovery rates and concentration factors up to 10 000. Åkerblom[146] gave an excellent overview of recent techniques and procedures for the extraction and cleanup of water samples in pesticide analysis.

In the past, several analytical methods were described using conventional LLE with dichloromethane,[47,55,63,64,67,68,106] diethyl ether,[39,62] ethyl acetate[35,119,121] or benzene.[69] Recently, SPE has been gaining much popularity in environmental water analysis. SPE has some striking advantages:

(1) It enables concentration of high sample volumes.
(2) It may also be used for sample cleanup.
(3) It needs only small solvent volumes for analyte elution.
(4) It is less toxic.
(5) A high degree of automation is possible.
(6) It saves labor and time, because some samples can be prepared in parallel.
(7) It is usually less expensive.

A variety of SPE materials are commercially available and have been successfully used for the extraction of acid pesticides. Most publications describe the use of modified silica gels, especially RP-C18 adsorbents.[34,37,40,53,72,74–76,78,102,110,117,120] The use of the polymeric materials Amberlite XAD-4,[61] PLRP-S, PRP-1,[104,108,110,114] ion exchanging adsorbents,[38,57] graphitized carbon black adsorbents[115,125,127,128] or even a combination of Carbopack B/SAX[103] has also been described. The flow chart in Figure 19 gives an overview of possible combinations of sample preparation, derivatization and instrumental analysis which can be applied for the determination of acidic pesticides in water.

At present, SPE using RP-C18 seems to be state of the art for the extraction of acidic pesticides from water samples. Therefore, it will be part of the application example described in the following section.

4.1.1 Sample Preparation

The flow chart in Figure 20 depicts an example of a sample preparation procedure for the determination of acidic pesticides in water. This procedure applies SPE with RP-C18 adsorbent, derivatization with PFBBr and GC/MS applying full scan and/or SIM acquisition for analyte detection. Derivatization with PFBBr is described in detail in the section on Pentafluorobenzylation. GC/MS detection applying SIM is carried out using the parameters described in sections on Pentafluorobenzylation, Pentafluorobenzyl Derivatives and section 3.1.3.2 on SIM.

Sample concentration factors up to 10 000 are possible, but sample volumes should be reduced at higher matrix loads to avoid blocking of the SPE cartridges, especially when analyzing surface or even sewage samples. As discussed in section 3.1.1, the choice of the derivatization procedure, as prerequisite for GC determinations applying GC, depends on individual requirements such as the range of compounds which have to be analyzed, the GC detection method and the detection limits required. The use of GC/MS or GC/MS/MS is not always necessary but is recommended when analyzing environmental samples. In general, SPE can be combined with any instrumental method, but the use of GC/MS (or MS/MS) or HPLC/MS/MS for analyte detection is superior in achieving low detection limits, lower matrix interferences and a higher degree of confidence in the analytical results.

Some key conditions have to be considered in applying SPE to the extraction of acidic pesticides from water samples. Samples need to be acidified to a pH < 2 to obtain the undissociated form of the acidic pesticides, otherwise the acidic analytes will not be retained significantly by the nonpolar reversed-phase adsorbent and recovery rates will be low. Nowadays, a variety of RP-C18 materials is commercially available, nonendcapped materials with lower carbon load and some residual active silanole groups have proven to be best suited for the extraction of acidic pesticides. Other important conditions in sample preparation are the conditioning of the SPE cartridges, the percolation velocity, the drying step and the choice and volume of the elution solvent. Some acidic pesticides are applied as their esters in agricultural use. Most of them are easily hydrolyzed into the active acid form in

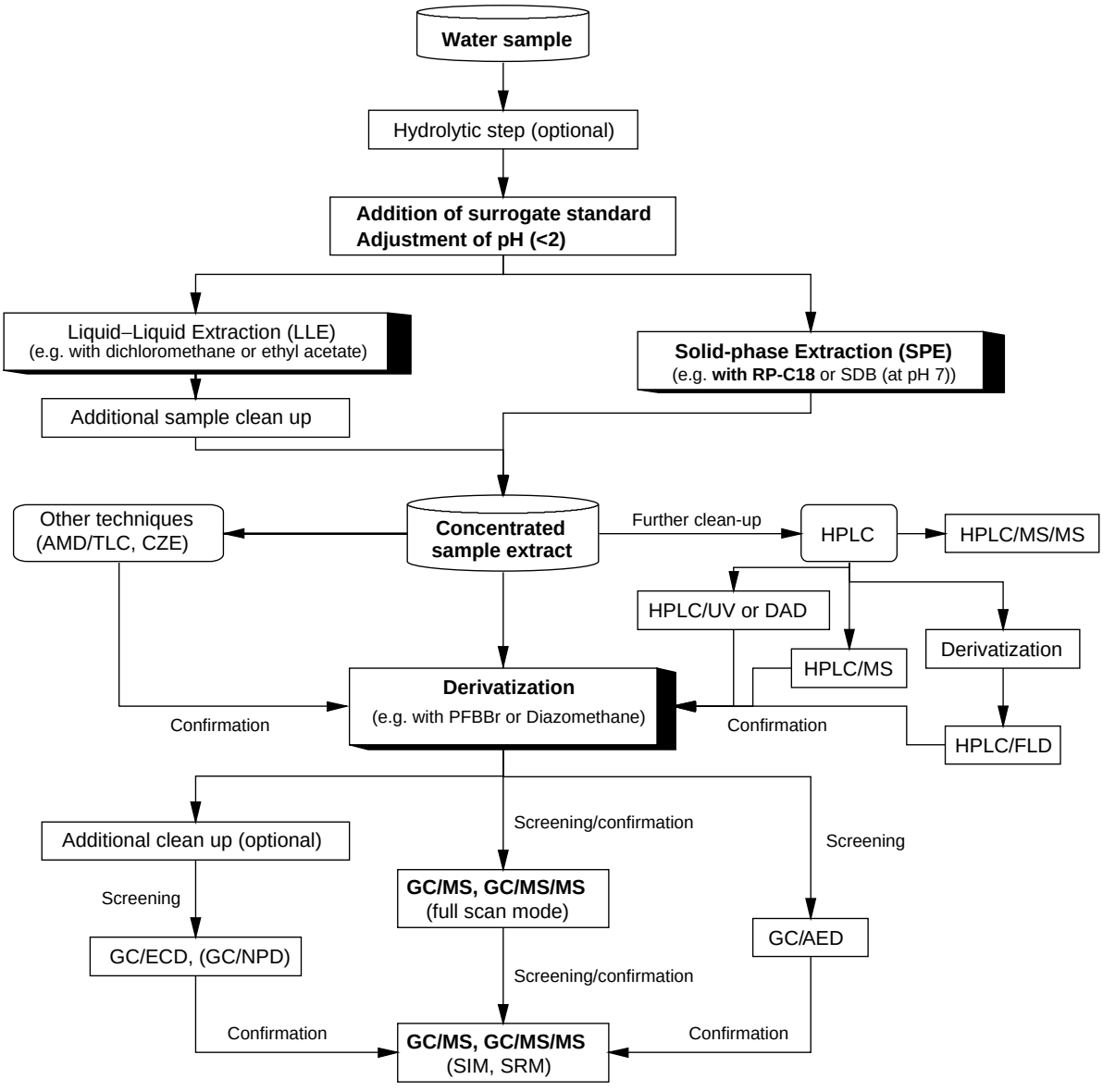

Figure 19 Possible analytical instrumentation for the multiresidue analysis of acidic pesticides in water. In bold: principle of the analytical procedure described in section 4.1.1. SDB, styrenedivinylbenzene.

the environment, but to include the parent compounds it may also be necessary to introduce a hydrolytic step such as alkaline hydrolysis before extraction.[146] According to Åkerblom[146] phenoxy acids, ioxynil **(14)**, flamprop **(27)**, and fluroxypyr **(30)** are gently cleaved at pH 12 by standing overnight at ambient temperature.

4.1.2 Recoveries, Limits of Detection and Limits of Quantitation

As compiled in Table 9, most acidic pesticides can be extracted and determined at trace level concentrations down to the low nanograms level from water samples applying SPE and GC/MS detection with SIM.[75,76] Low recoveries were only observed for the indolylalkanoic acids, picloram **(33)** and flurenol **(29)**. Fortunately, apart from picloram the other three compounds have no agricultural importance as pesticides.

Figure 21 shows the MID chromatogram of a drinking water sample extract spiked with a mixture of more than 30 acidic pesticides at a concentration level of only $10\,ng\,L^{-1}$ each. All analytes could be detected unequivocally although this concentration is far below all maximum tolerance levels set for pesticides in drinking water. This example shows that multiresidue analysis of acidic pesticides can meet the European maximum

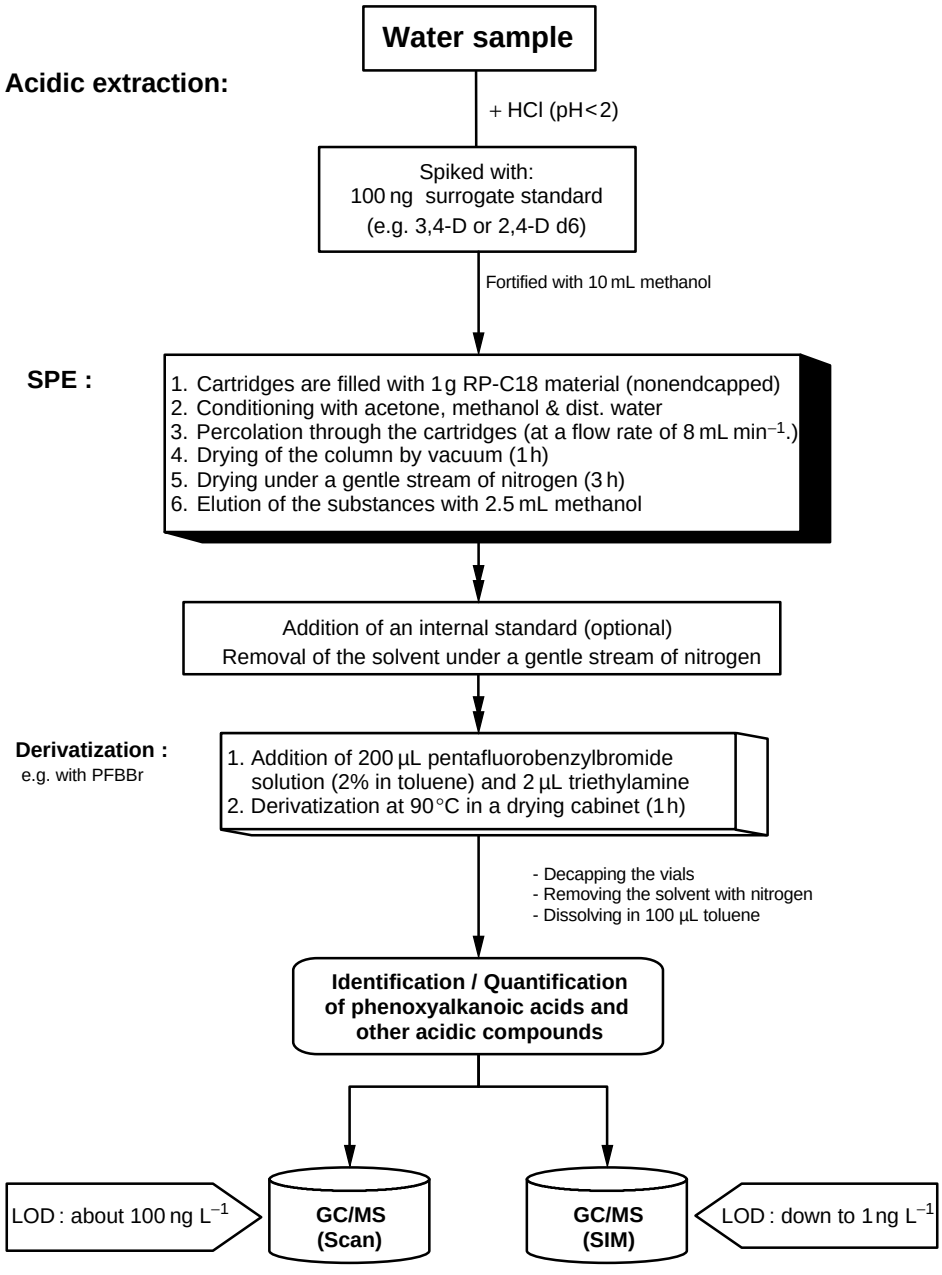

Figure 20 Flowchart showing a suitable procedure for the trace level analysis of acidic herbicides in water.

tolerance levels of 100 ng L^{-1} in drinking water.[1] In modern pesticide analysis, it is possible to achieve limits of detection (LODs) and also limits of quantitation (LOQs) clearly below 100 ng L^{-1} in many different kinds of matrices such as drinking, ground, surface or even sewage water[75,76] as will be described in section 4.4.

The use of new, more ''polar'' RP-C18 adsorbents further enhanced the recovery rates and accuracy of SPE for acidic pesticides. These adsorbents are so-called nonendcapped materials (e.g. BAKERBOND C-18 Polar Plus or LiChrolut® RP-select B) with lower carbon load and some residual active silanole groups. They have proven to be best suited for the extraction of acidic pesticides. Table 10 shows a comparison between different adsorbents for the extraction of some acidic analytes. An alternative to the commonly used reversed-phase adsorbents are polymeric adsorbents which have gained much importance for the extraction of polar analytes. Pichon et al.[114] achieved high recoveries using polymeric adsorbents for the extraction of acidic pesticides at pH 7. The advantage of an extraction at neutral pH is that the fraction of naturally occurring humic acids

Table 9 LODs and recovery rates of some acidic pesticides as their PFB derivatives applying GC/MS with SIM programming. Data taken from Heberer et al.[75]

Compound	LOD (STD)[a] (pg)	LOD (SPE)[b] (ng L^{-1})	Recovery[c] (%)	RSD[d] (%)
Acifluorfen PFB ester	100	≤10	77	13
Benazolin PFB ester	10	≤10	89	11
Bentazone PFB amide	20	≤10	71	24
Bromoxynil PFB ether	20	≤10	75	27
Chloramben PFB ester	20	≤25	58	6
Chlorfenac PFB ester	20	≤10	92	9
Chlorflurenol-methyl	10	≤10	109	2
4-Chlorophenoxyacetic acid PFB ester	10	≤10	72	11
Clofibric acid PFB ester	2	≤1	106	7
Clopyralid PFB ester	20	≤10	79	5
2,4-D PFB ester	20	≤5	79	16
2,4-DB PFB ester	10	≤10	53	5
Dicamba PFB ester	2	≤1	101	10
2,4-Dichlorobenzoic acid PFB ester	4	≤1	90	9
Dichlorprop PFB ester	1	≤1	94	10
Fenoprop PFB ester	2	≤5	100	7
Flamprop PFB ester	2	≤1	88	8
Flamprop-isopropyl	3	≤1	109	14
Fluazifop PFB ester	5	≤1	91	11
Fluazifop-*p*-butyl	1	≤1	101	6
Flurenol PFB ester	30	≤25	33	1
Flurenol-butyl	10	≤10	109	2
Fluroxypyr PFB ester	30	≤10	87	10
Haloxyfop PFB ester	10	≤1	88	10
Indol-3-acetic acid PFB ester	10	–[e]	(42)	–[e]
Indol-3-butyric acid PFB ester	10	–[e]	(10)	–[e]
Indol-3-propionic acid PFB ester	10	–[e]	(56)	–[e]
MCPA PFB ester	2	≤1	104	6
MCPB PFB ester	20	≤10	85	10
Mecoprop PFB ester	10	≤1	69	6
1-Naphthyl acetic acid PFB ester	2	≤5	89	8
Picloram PFB ester	30	–[e]	(9)	–[e]
2,4,5-T PFB ester	20	≤10	102	14
Triclopyr PFB ester	2	≤1	107	5

[a] LOD with target compounds injected as test standard mixtures.
[b] LOD with acidic compounds spiked to 1 L of tap water: 100, 50, 25, 10 ng L^{-1} extrapolated from the lowest concentration level.
[c] Recovery of six parallel analyses applying SPE (partly endcapped RP-C18 material) of water samples fortified with 100 ng L^{-1} of each compound.
[d] Relative standard deviation.
[e] No detection limits and standard deviations reported owing to unsatisfactory recoveries or great variations.

which is not soluble at pH values >2 does not interfere with sample analysis, especially when applying HPLC. Fulvic acids are, however, also soluble at neutral pH, thus cleanup of sample extracts may still be necessary.

4.2 Analysis of Soil Samples

Several analytical methods have been reported for the analysis of acidic pesticides in soil samples.[38,41,49,54,66,77,82,121] In particular, the methods of Lee et al.,[66] Bruns et al.[41] and Heberer and Stan[77] have been developed for trace level analysis with LODs of less than 10 µg kg^{-1} (ppb). Lee et al.[66] described the extraction of ten acidic herbicides from soil samples acidified with H_2SO_4 to a pH < 1. In this method a mixture of acetone and hexane (1 + 1) was used for extraction followed by liquid/liquid partitioning with dichloromethane

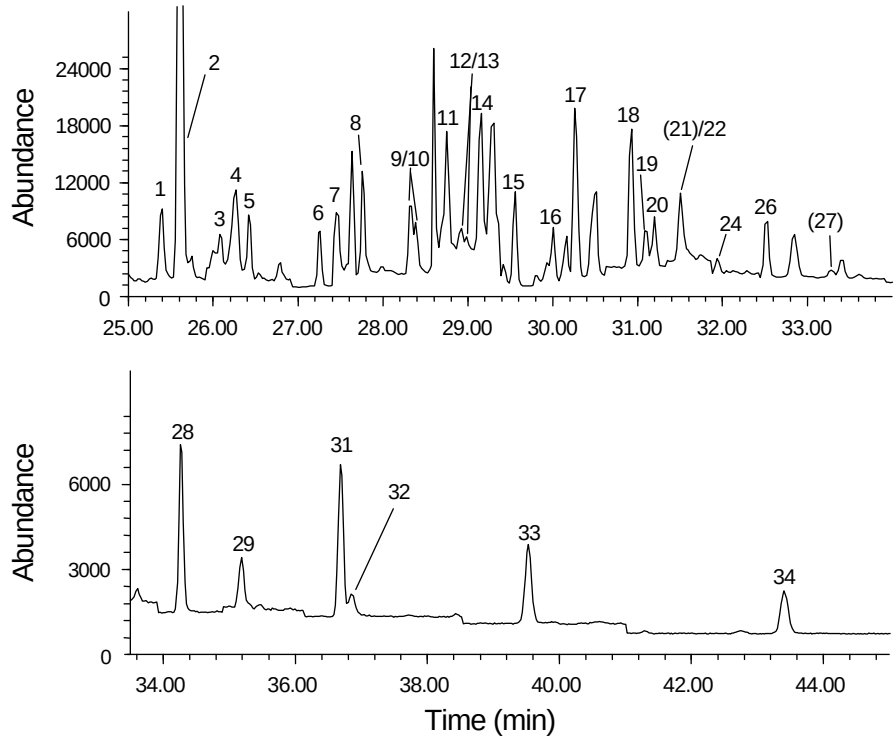

Figure 21 MID chromatogram of an extract of a water sample fortified with a mixture of acidic pesticides at 10 ng L^{-1}. Internal standard (2,4-dichlorobenzoic acid) added at a concentration level of 200 ng L^{-1}. For peak identification numbers refer to Table 6. (Reprinted from Heberer et al., copyright (1994),[75] with permission from AOAC International.)

Table 10 Comparison of recovery rates of different RP-C18 adsorbents (endcapped, partly endcapped, nonendcapped) from 1 L of spiked tap water samples (concentration: 100 ng L^{-1} of each analyte) applying SPE

Analyte	Endcapped[a]		Partly endcapped[a]		Nonendcapped[b]	
	Recovery (%)	RSD[c] (%)	Recovery (%)	RSD[c] (%)	Recovery (%)	RSD[c] (%)
Clofibric acid	96	5	106	7	97	6
Mecoprop **(8)**	64	4	89	8	99	10
2-(4-Chlorophenoxy)butyric acid (surrogate standard)	–	–	102	6	97	7
MCPA **(6)**	55	5	85	10	102	7
Dichlorprop **(4)**	67	6	94	10	100	6
2,4-D **(2)**	51	5	79	16	94	15

[a] $n = 6$.
[b] $n = 12$.
[c] RSD, relative standard deviation.

and evaporation to dryness. The analytes were derivatized with PFBBr (see section on Pentafluorobenzylation), the resulting extract was cleaned up by silica gel column fractionation and determined by GC/ECD (see section 3.1.2). In their method for the extraction of acidic herbicides from soil samples, Bruns et al.[41] used a mixture of acetone, distilled water and acetic acid (80 + 19 + 1) and a second extraction with 0.1 N KOH. This was followed by liquid/liquid partitioning with dichloromethane, evaporation, derivatization with diazomethane (see section on Methylation with Diazomethane) and GC/MS detection with SIM (see section on Methyl Derivatives and section 3.1.3.2 on SIM). Bruns et al.[41] found GC/MS detection applying SIM to be indispensable in achieving the necessary reliability of the analytical results for trace level concentrations. Heberer and Stan[77] tried to combine the advantages of both methods and invented SPE for analyte enrichment and sample cleanup as well. The resulting method allowed the analysis of dichlorprop **(4)**, mecoprop **(8)**, clofibric acid and naphthylacetic acid

(32) in soil samples at the low parts per billion level. This method, which will be described here, may also be extended to the analysis of some other acidic pesticides.

4.2.1 Sample Preparation and Cleanup Procedures

The method for the analysis of acidic pesticides applies acidic extraction, SPE, derivatization with PFBBr and GC/MS detection with SIM. An additional cleanup step by GPC was found to be useful to protect the gas chromatographic system from matrix contamination.[77] It was, however, found not to be necessary because of the high selectivity of GC/MS detection with SIM.

4.2.1.1 Sample Extraction A 50-g portion of wet soil is weighed into an Erlenmeyer flask and 200 ng of the surrogate standard (e.g. 2,4-D d_6 or 3,4-D) are added to the sample. Then 120 mL of the extraction solvent acetone–water–acetic acid (80 + 19 + 1; v/v/v) is added and the top of the Erlenmeyer flask is covered with aluminum foil and shaken overnight. After centrifuging the sample at 3000 rpm for 5 min, the supernatant is filtered into a round flask. Then 50 mL of the extraction solvent is added and the sample is re-extracted by ultrasound, centrifugation and filtration of the supernatant into the round flask again. The combined extracts are evaporated to approximately 20 mL by means of a rotary evaporator and approximately 180 mL of distilled water is added to this extract.[77]

4.2.1.2 Solid-phase Extraction SPE is carried out according to the procedure described in section 4.1.1.

4.2.1.3 Derivatization Derivatization is carried out according to the procedure described in the section on Pentafluorobenzylation. The resulting extract is analyzed directly by GC/MS. Alternatively, the dried sample residue is dissolved in 1.5 mL of cyclohexane/ethylacetate (50/50; v/v) and this extract is used for further cleanup by GPC.

4.2.1.4 Gel Permeation Chromatography (Optional) A 1-mL portion of the sample extract is transferred to the GPC column by means of a 1-mL injection loop. The first 20 mL of the GPC eluate (dump volume) is discarded, the fraction between 20 and 38 mL (collect volume) is collected and the GPC column is washed with another 25 mL. The collected fraction is evaporated to dryness by means of a rotary evaporator. Then the remaining residue is dissolved in 100 µL of toluene, transferred into a microvial and directed to GC/MS analysis.[77]

4.2.1.5 Gas Chromatography/Mass Spectrometry Detection GC/MS detection applying SIM is carried out using the parameters described in sections on Pentafluorobenzylation and PFB Derivatives and section 3.1.3.2 on SIM.

4.2.2 Recoveries, Limits of Detection and Limits of Quantitation

The recovery rates, LODs and LOQs of a few acidic analytes including dichlorprop **(4)** and mecoprop **(8)** are compiled in Table 11. The LOQs are all below 1 µg kg^{-1} (ppb) soil for these analytes and the recovery rates were all better than 90% even at the lowest of the spiked concentrations of only 1 ppb.[77] Figure 22 shows the resulting MID chromatogram of a derivatized extract of a soil sample fortified with 1 ppb of each analyte. The method may also be extended for the analysis of some other acidic pesticides.[77]

4.3 Interferences

Common interferences in environmental analysis of polar contaminants often derive from natural compounds such as fatty acids or humic acids which occur in water or soil samples at concentrations much higher than those of the analytes. As a consequence, some analytical methods require extended cleanup procedures for sample extracts.[64] In particular, analysis of matrix prone ground or surface water by HPLC with UV or DAD (section 3.2.1) or GC/ECD (section 3.1.2) may suffer heavy interference. Thus, the use of more selective

Table 11 Retention times, indicative EI/MS ions and detection limits in soil

Compound	Retention time (min)	Average recovery (%)[a]	RSD[b] (%) (n = 36)	LOD (ppb)	LOQ (ppb)
Clofibric acid	25.42	97	7	<0.1	<1
2,4-Dichlorobenzoic acid	25.65	–	–	–	–
Mecoprop **(8)**	26.43	91	7	<0.1	<1
2-(4-Chlorophenoxy)-butyric acid (surrogate standard)	26.63	95	6	<0.1	<1
Dichlorprop **(4)**	27.79	94	8	<0.1	<1
1-Naphthylacetic acid **(32)**	29.17	93	7	<0.1	<1

[a] Average recovery rates obtained for the analysis of different soil types fortified at concentrations between 1 and 200 µg kg (ppb).
[b] Relative standard deviation.

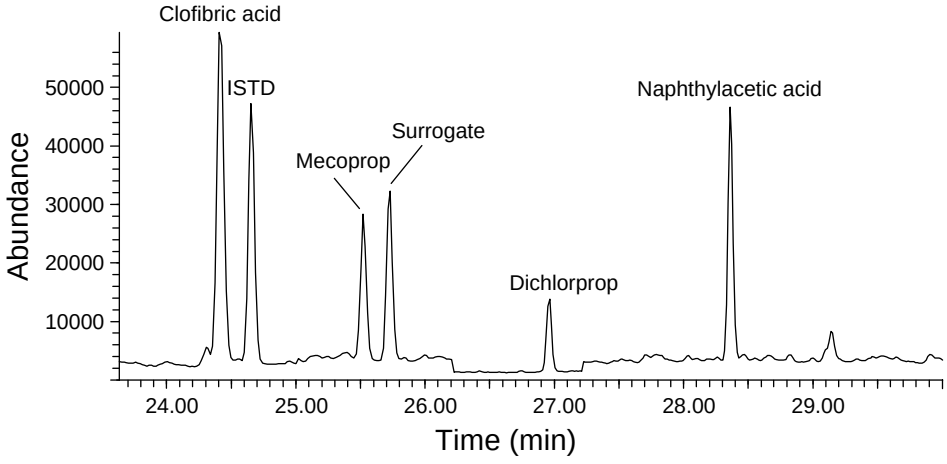

Figure 22 MID chromatogram recorded with GC/MS applying SIM programming of a derivatized extract derived from a sediment sample fortified with $1\,\mu g\,kg^{-1}$ of each analyte. (Reprinted from Heberer and Stan,[77] copyright (1996), with permission from AOAC International.)

detection methods such as MS, MS/MS or AED is often necessary to produce reliable results. Nevertheless, matrix interferences should also be avoided to protect the chromatographic system.

Although MS detection caused by its selectivity is not as prone to interferences as nonselective detectors, some considerations must be taken into account in method development, especially when applying SIM acquisition. Ions selected for SIM must be intense in the mass spectrum and must be indicative, meaning that all ions prominent in the mass spectrometric background, e.g. column bleed or common environmental contaminants such as phthalates and hydrocarbons, should be avoided. Interferences may occur in the GC chromatogram as discrete compound peaks (e.g. phthalates), in certain segments of the chromatogram or during the whole GC run. A compilation of possible background ions that may interfere in the analysis of the analytes is given in Table 12. Several authors[24,147,148] tabulated ions, or published mass spectra of possible background interferences, e.g. solvents, various column coatings, silicon oils or hydrocarbons. Dependent on individual instrumental conditions, such sources of interference should be taken into account and avoided as far as possible.

Interfering matrix compounds can often be separated from the analytes prior to or during gas chromatographic analysis without introducing special time-consuming cleanup steps. Thus, the application of SPE was found to be more selective in separating the target analytes from matrix compounds than LLE.[75] When using SPE for sample preparation, the choice of a suitable material is probably the first crucial step, other important details have already been discussed in section 4.1. In the past, SPE adsorbents of the same type were found to vary significantly in their ability to retain the analytes. Nowadays, most of the suppliers offer different

Table 12 List of ions which may also originate from common background interferences

Ions (m/z)	Possible source of interfering masses
149, 167, 279	Plasticizers (phthalic esters)
129, 185, 259, 329	Plasticizers (tributyl acetyl citrate)
99, 155, 211, 266	Plasticizers (tributyl phosphate)
91, 165, 198, 261, 368	Plasticizers (tricresyl phosphate)
73, 133, 147, 207, 221, 281, 355, 429, 503	Silicon oil, column coating (SE-30, SE-54, OV-101, OV-1, SF-96)
135, 197, 209, 259, 333, 345, 408, 465, 527	Column coating (OV-17, OV-11)
75, 91, 135, 156, 169, 183, 253, 352, 389, 449, 458, 502, 511, 520	Column coating (OV-225)
29, 43, 57, . . .; 41, 55, 69, . . .; 53, 67, 81, . . .	Hydrocarbons (homologs)
137, 165, 180; 205, 220	BHA, BHT (antioxidant)
64, 96, 128, 160, 192, 224, 256	Sulfur (S_8)
39, 45, 51, 65, 91, 92	Solvent (toluene)

BHA, butylated hydroxyanisole; BHT, butylated hydroxytoluene.

adsorbents of the same type which are characterized as endcapped, nonendcapped and/or by different carbon loads (see also section 4.1.1). Sometimes considerable batch to batch variations in recoveries have also been reported with materials from the same supplier.[74] Thus, in routine analysis or whenever larger sample series have to be analyzed it may be favorable to buy larger charges of adsorbents from the same batch to assure constant analytical quality.

When applying PFBBr for the derivatization of the acidic analytes, the use of toluene was found to have some advantages compared to commonly used acetone.[75] The PFB derivatives of the acidic pesticides are easily dissolved in toluene while many coextracted matrix compounds such as humic acids remain underivatized and undissolved.[75] Additionally, GC/MS analysis takes advantage of some properties of the PFB derivatives compared to the widely used methyl esters. By increasing the MWs of the analytes significantly (+180 Da) a better separation from coextracted matrix compounds is achieved which results in longer retention times.[75] By recording ions with higher *m*/*z* values, a better detection selectivity is achieved when using EI/MS.[75] Moreover, on a nonpolar GC-column (e.g. SE-54) molecules containing fluorine, such as the PFB derivatives, elute earlier than other molecules of similar molecular mass but without fluorine substitution.[75]

Varying recoveries are often observed when a large amount of matrix is present in environmental samples, thus demanding better control of the sample preparation procedure. The use of suitable internal and/or surrogate standards is almost essential in environmental analysis of pesticides at trace level concentrations. The internal standard may be used to control the reproducibility of the derivatization and the operation of the gas chromatographic system which is also crucial for SIM programming when working with narrow time windows.[76] The surrogate standard is added before analysis and may be used to control the whole analytical procedure including recovery control.[76] Heberer et al.[76] summarized the requirements for the properties of the surrogate standard as follows: (1) it should not be expected in environmental samples, (2) it should give high recoveries from fortified water samples and (3) it should be structurally similar to the most important target compounds. Thus, a suitable surrogate may be an isomer (e.g. 3,4-D or 2,3-D) or even better an isotopically labeled compound such as 2,4-D d_6.

4.4 Some Practical Applications

A few examples of positive findings of acidic pesticides and related polar contaminants are presented in this section to demonstrate the performance, sensitivity and potentials of multiresidue methods in environmental water analysis. The sensitivity of modern environmental analysis is demonstrated in Figure 23, showing the indicative ion traces (*m*/*z* 162, 414, 416) for dichlorprop PFB ester which were recorded for a derivatized sample extract applying GC/MS with SIM. In this case, dichlorprop **(4)** could be identified at a concentration level of only 0.5 ng L^{-1} in a sewage water sample with a dissolved organic carbon content of about 70 mg L^{-1}.[75] Even this small amount of dichlorprop extracted from one liter of sewage with a high load of organic matrix was identified by retention time and with the peak area ratios of the indicative ions matching the abundances listed in Table 3.

The performance and the potentials of multiresidue analysis of acidic pesticides applying GC/MS detection are demonstrated with the example of a derivatized river water sample shown in Figures 24 to 26.

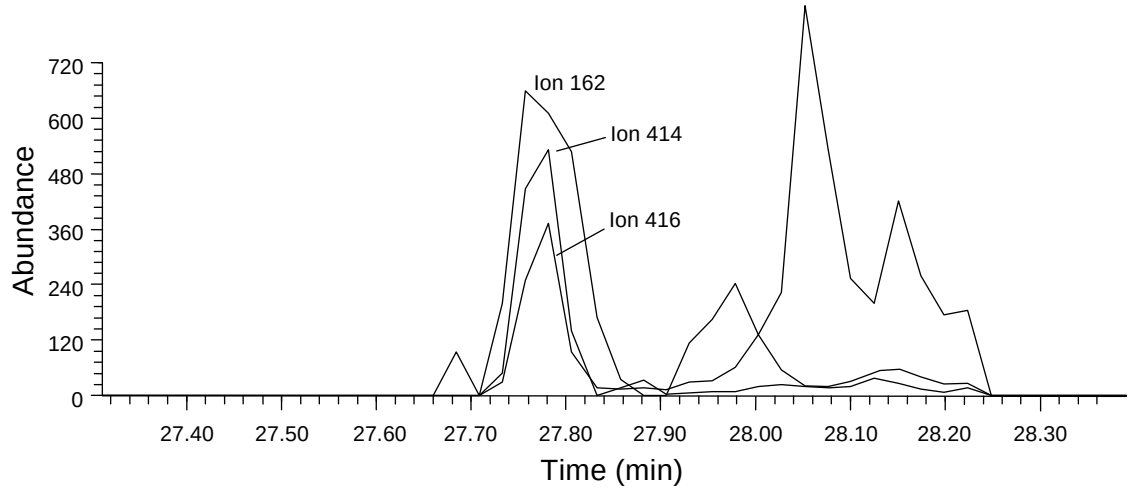

Figure 23 Ion traces *m*/*z* 416, 416, 162 for dichlorprop **(4)** found at a concentration of only 0.5 ng L^{-1} in a sewage water sample. (Reprinted from Heberer et al.,[75] copyright (1994), with permission from AOAC International.)

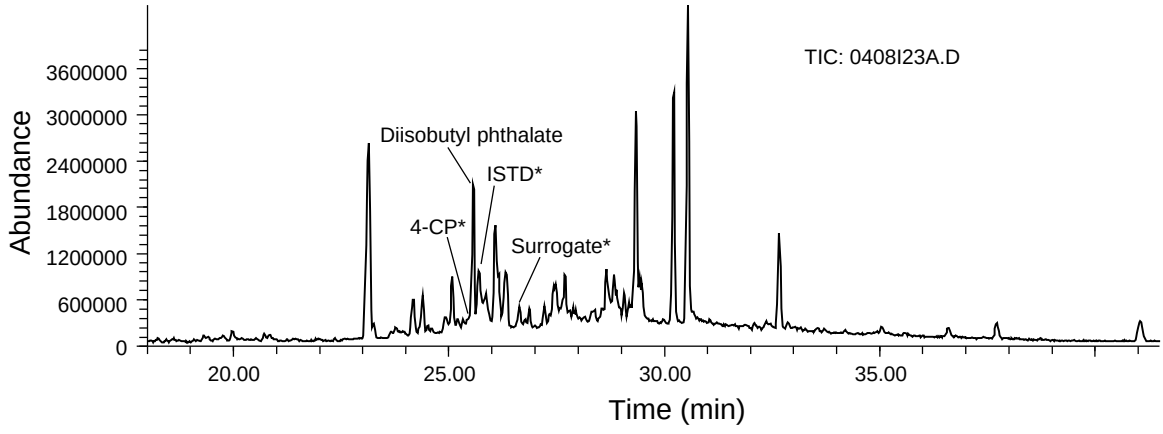

Figure 24 TIC chromatogram of a river water sample from the Havel near Berlin recorded with GC/MS in full scan mode (*m*/*z* 50–550). Sample spiked with 200 ng surrogate standard and 400 ng 2,4-dichlorobenzoic acid (ISTD) per liter. *Peaks resulting from matrix compounds overlapping the indicated target and standard compounds. (Reproduced from Heberer et al.[76] by permission of Gordon and Breach Science Publishers.)

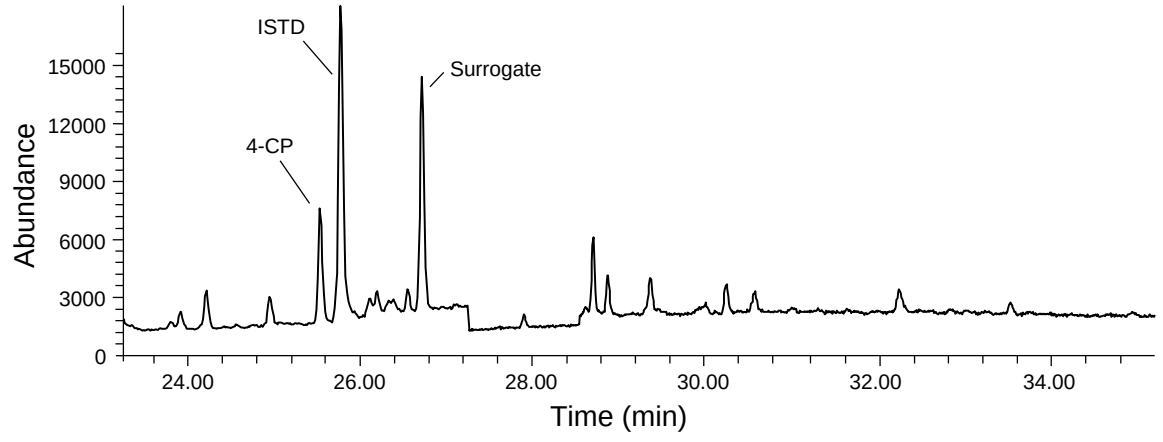

Figure 25 MID chromatogram recorded with GC/MS applying SIM programming of the river water sample from Figure 24. Sample with 74% recovery for the surrogate standard, containing 150 ng clofibric acid per liter. (Reproduced from Heberer et al.[76] by permission of Gordon and Breach Science Publishers.)

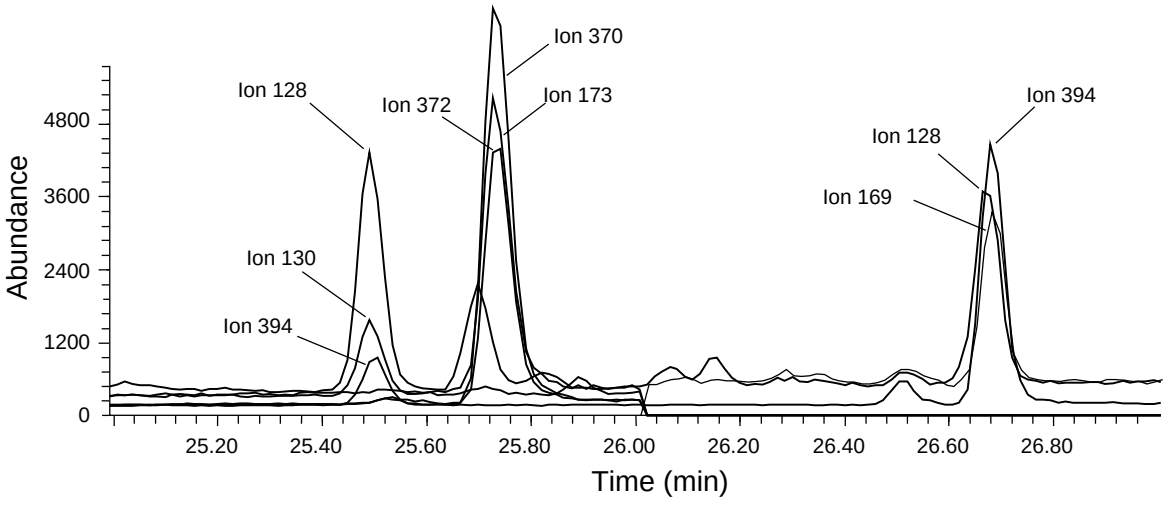

Figure 26 Ion traces extracted from the MID chromatogram of the river water sample shown in Figure 25. (Reproduced from Heberer et al.[76] by permission of Gordon and Breach Science Publishers.)

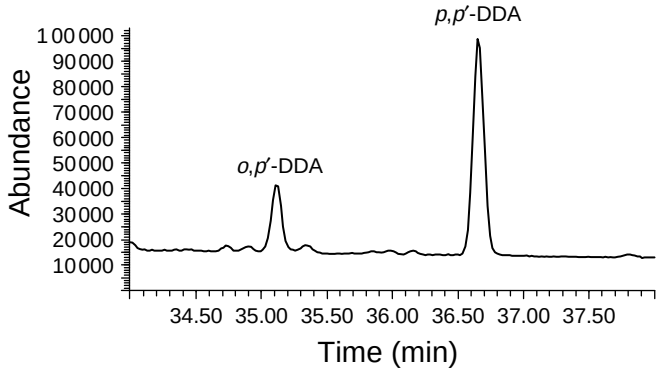

Figure 27 MID chromatogram of a derivatized extract of a bank filtrate sample recorded with GC/MS applying SIM. This sample contained 65 ng L^{-1} of *o*,*p*′-DDA and 220 ng L^{-1} of *p*,*p*′-DDA.

In the example shown in Figure 24, a TIC chromatogram of a typical river water sample recorded with GC/MS in full scan is presented, demonstrating how a high load of matrix gives a very complex chromatogram. The analyte clofibric acid labeled as 4-CP as well as the surrogate and the internal standard are overlapped by matrix compounds, although the occurrence of the peaks at the expected retention times already suggests their presence. In Figure 25, the corresponding MID chromatogram of the same river water sample recorded using GC/MS with SIM programming (section 3.1.3.2) is presented. The complex chromatogram resulting from full scan has turned into a clear chromatogram with the analyte peaks dominating. When examining the ion traces, as demonstrated in Figure 26, the signal-to-noise ratio is found to be excellent and the detection of the target compounds is not interfered with by matrix compounds. Thus, the 150-ng L^{-1} concentration of the analyte clofibric acid in the river water sample could be easily quantified. Clofibric acid is a phenoxy acid and a structural isomer of the herbicide mecoprop **(8)**. Clofibric acid is, however, not used as a pesticide. It is the biologically active metabolite of blood lipid regulating pharmaceuticals.[149] It is discharged as a persistent drug residue by sewage treatment effluents into the surface water.[149] This example shows clearly the potential of multiresidue analysis of acidic pesticides as it is also applicable to the analysis of other polar contaminants such as drug residues. The drug clofibric acid and another polar contaminant *N*-(phenylsulfonyl)-sarcosine, also originating from sewage effluents, have been found unexpectedly in groundwater samples from the Berlin area during the monitoring of acidic pesticides![150–152] In 1994, both contaminants were also detected at concentrations of up to 270 ng L^{-1} in Berlin drinking water.[153] The findings for clofibric acid in Berlin initiated many drug monitoring investigations. More than 40 pharmaceuticals have been found up to the microgram per liter level in surface water and a few even in groundwater samples.[154,155]

The last example is that of DDA which was also found during the analysis of other acidic pesticides.[156] The MID chromatogram in Figure 27 shows the detection of *o*,*p*′- and *p*,*p*′-DDA as their PFB esters in an extract of a groundwater sample from a bank filtration area. DDA is the long known polar metabolite of the nonpolar insecticide DDT. Although it was known as one key metabolite in the biological degradation of DDT residues, it was never included in any monitoring program for DDT residues in the aquatic environment.[157] Despite other DDT metabolites such as 2,2-bis(chlorophenyl)-1,1-dichloroethane (DDD), 2,2-bis(chlorophenyl)-1,1-dichloroethylene (DDE) or 2,2-bis(chlorophenyl)-1-chloroethylene (DDMU) it was neglected because it could not be analyzed together with the other nonpolar DDT derivatives. DDA can, however, be analyzed together with other acidic pesticides in one multiresidue method. In recent investigations of Berlin groundwater and surface water samples it was found as the major DDT derivative in the surface water of a DDT Superfund site.[157]

ACKNOWLEDGMENTS

The author thanks Mrs K. Schmidt-Bäumler and Mrs G. Fricke for their practical assistance and Mr B. Hatton for his review.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
AMDTLC	Automated Multiple Development Thin-layer Chromatography

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
BHA	Butylated Hydroxyanisole
BHT	Butylated Hydroxytoluene
CAS	Chemical Abstracts Service
CEDMSDEA	(2-Cyanoethyl)dimethyl-(diethylamino)silane
CI	Chemical Ionization
CID	Collision Induced Dissociation
CZE	Capillary Zone Electrophoresis
DAD	Diode Array Detection
DDA	bis-(chlorophenyl)acetic Acid
DDD	2,2-bis(chlorophenyl)-1,1-dichloroethane
DDE	2,2-bis(chlorophenyl)-1,1-dichloroethylene
DDMU	2,2-bis(chlorophenyl)-1-chloroethylene
DDT	2,2-bis(chlorophenyl)-1,1,1-trichloroethane
DNOC	4,6-Dinitro-*o*-Cresol
ECD	Electron Capture Detection
EI	Electron Ionization
ELCD	Electrochemical Detection
EPA	Environmental Protection Agency
ESI	Electrospray Ionization
EU	European Union
FLD	Fluorescence Detection
FPD	Flame Photometric Detector/Detection
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GLP	Good Laboratory Practice
GPC	Gel Permeation Chromatography
HPCE	High-performance Capillary Electrophoresis
HPLC	High-performance Liquid Chromatography
HR	High Resolution
ID	Inner Dimensions
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LIF	Laser-induced Fluorescence
LLE	Liquid–Liquid Extraction
LOD	Limit of Detection
LOQ	Limits of Quantitation
MCPA	(4-Chloro-2-methylphenoxy)acetic Acid
MCPB	2-(4-Chloro-2-methylphenoxy)-propionic Acid
MID	Multiple Ion Detection
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MTBE	Methyl *tert*-butyl Ether
MTBSTFA	*N*-(*tert*-Butyldimethylsilyl)-*N*-methyltrifluoroacetamide
MW	Molecular Weight
NCI	Negative Chemical Ionization
NPD	Nitrogen–Phosphorus Detection
PB	Particle Beam
PCB	Polychlorinated Biphenyl
PCP	Pentachlorophenol
PFB	Pentafluorobenzyl
PFBBr	Pentafluorobenzyl Bromide
PFP	Pentafluoropropyl
PTV	Programmed Temperature Vaporization
RP-C18	Reversed-phase Octadecyl
RRF	Relative Response Factor
SDB	Styrenedivinylbenzene
SIM	Selected Ion Monitoring
SOP	Standard Operating Procedure
SPE	Solid-phase Extraction
SRM	Selected Reaction Monitoring
TBDMS	*tert*-butyldimethylsilyl
TCE	Trichloroethyl
TFE	Trifluoroethyl
TIC	Total Ion Current
TLC	Thin-layer Chromatography
TMAH	Trimethylanilinium Hydroxide
TMSH	Trimethylsulfonium Hydroxide
TSP	Thermospray
UV	Ultraviolet
2,3,6-TBA	2,3,6-Trichlorobenzoic acid
2,4-D	(2,4-Dichlorophenoxy)acetic Acid
2,4-DB	(2,4-Dichlorophenoxy)butyric Acid
2,4,5-T	(2,4,5-Trichlorophenoxy)acetic Acid
4-CPA	*p*-Chlorophenoxyacetic Acid

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Dioxin-like Compounds, Screening Assays • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Quality Assurance in Environmental Analysis • Sample Preparation for

Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soxhlet and Ultrasonic Extraction of Organics in Solids • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Pesticides (New Generation) and Related Compounds, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography • Reversed Phase Liquid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. EEC Drinking Water Guideline, 80/779/EEC, EEC No. L229/11-29, EEC, Brussels, August 30th, 1980.
2. T.R. Roberts (ed.), *Metabolic Pathways of Agrochemicals. Part 1: Herbicides and Plant Growth Regulators*, The Royal Society of Chemistry, Cambridge, UK, 1998.
3. S. Hatrik, J. Tekel, 'Extraction Methodology and Chromatography for the Determination of Residual Pesticides in Water – Review', *J. Chromatogr. A*, **733**, 217–233 (1996).
4. Industrieverband Agrar, 'Mengenmäßig bedeutende Wirkstoffe in der Bundesrepublik Deutschland 1997', Industrieverband Agrar e.V., Frankfurt, Germany, 1998.
5. J.L. Tadeo, C. Sanchez-Brunete, A.I. García-Valcarcel, L. Martínez, R.A. Pérez, 'Determination of Cereal Herbicide Residues in Environmental Samples by Gas Chromatography', *J. Chromatogr. A*, **754**, 347–365 (1996).
6. H.-J. Stan, F. Schwarzer, 'On-line Coupling of Liquid Chromatography with Thin-layer Chromatography', *J. Chromatogr. A*, **819**, 35–44 (1998).
7. M.-P. Marco, D. Barceló, 'Environmental Applications of Analytical Biosensors – Review Article', *Meas. Sci. Technol.*, **7**, 1547–1562 (1996).
8. D. Knopp, M. Skerswetat, M. Schmidt, R. Niesser, 'Preparation and Characterization of Polyclonal Antibodies Against the Herbicide 2-(2,4-Dichlorophenoxy) propionic Acid (Dichlorprop)', *Fresenius Environ. Bull.*, **2**, 274–280 (1993).
9. F. García Sánchez, A. Navas, F. Alonso, J. Lovillo, 'Polarization Fluoroimmunoassay of the Herbicide Dichlorprop', *J. Agric. Food. Chem.*, **41**, 2215–2219 (1993).
10. M. Fránek, V. Kolár, M. Granátová, Z. Nevoránková, 'Monoclonal ELISA for 2,4-Dichlorophenoxyacetic Acid: Characterization of Antibodies and Assay Optimization', *J. Agric. Food Chem.*, **42**, 1369–1374 (1994).
11. W. Weber, K. Rubach, 'Entwicklung von Enzymimmunoassays zum Nachweis von Phenoxycarbonsäure-Herbiziden in Trink- und Grundwässern', *Acta Hydrochim. Hydrobiol.*, **22**, 53–59 (1994).
12. B.B. Dzantiev, A.V. Zherdev, M.F. Yulaev, R.A. Sitdikov, N.M. Dmitrieva, I. Yu Moreva, 'Electrochemical Immunosensors for Determination of Pesticides 2,4-Dichlorophenoxyacetic and 2,4,5-Trichlorophenoxyacetic Acids', *Biosens. Bioelectron.*, **11**, 179–185 (1996).
13. A. Dzgoev, M. Mecklenburg, P.-O. Larsson, B. Danielsson, 'Microformat Imaging ELISA for Pesticide Determination', *Anal. Chem.*, **68**, 3364–3369 (1996).
14. G. Matuszcyk, D. Knopp, R. Nießner, 'Development of an ELISA for 2,4-D: Characterization of Two Polyclonal Antisera', *Fresenius J. Anal. Chem.*, **354**, 41–47 (1996).
15. Ch. Wittmann, F.F. Bier, S.A. Eremin, R.D. Schmid, 'Quantitative Analysis of 2,4-Dichlorophenoxyacetic Acid in Water Samples by Two Immunosensing Methods', *J. Agric. Food Chem.*, **44**, 343–350 (1996).
16. A. Navas Díaz, F. García Sánchez, J. Lovillo, J.A. Ganzález García, 'Enhanced Chemiluminescence Kinetic ELISA of Dichlorprop Methyl Ester', *Anal. Chim. Acta*, **321**, 219–224 (1996).

17. J. Horácek, P. Skládal, 'Improved Direct Piezoelectric Biosensors Operating in Liquid Solution for the Competitive Label-free Immunoassay of 2,4-Dichlorophenoxyacetic Acid', *Anal. Chim. Acta*, **347**, 43–50 (1997).
18. E.P. Medyantseva, M.G. Vertlib, M.P. Kutyreva, E.I. Khaldeeva, G.K. Budnikov, S.A. Eremin, 'The Specific Immunochemical Detection of 2,4-Dichlorophenoxyacetic Acid and 2,4,5-Trichlorophenoxyacetic Acid Pesticides by Amperometric Cholinesterase Biosensors', *Anal. Chim. Acta*, **347**, 71–78 (1997).
19. A. Dzgoev, M. Mecklenburg, B. Xie, A. Miyabayashi, P.-O. Larsson, B. Danielsson, 'Optimization of a Charge Coupled Device Imaging Enzyme-linked Immunosorbent Assay and Supports for the Simultaneous Determination of Multiple 2,4-D Samples', *Anal. Chim. Acta*, **347**, 87–93 (1997).
20. T. Kaláb, P. Skládal, 'Disposable Multichannel Immunosensors for 2,4-Dichlorophenoxyacetic Acid Using Acetylcholinesterase as an Enzyme Label', *Electroanalysis*, **9**, 293–297 .
21. H.M. McNair, J.M. Miller, *Basic Gas Chromatography*, Wiley, Chichester, 1997.
22. M.C. McMaster, C. McMaster, *GC/MS – A Practical User's Guide*, Wiley, Chichester, 1998.
23. H.F. Linskens, J.F. Jackson (eds.), *Gas Chromatography/Mass Spectrometry*, Springer, Berlin, 1986.
24. H.-J. Hübschmann, *Handbuch der GC/MS – Grundlagen und Anwendung*, VCH, Weinheim, 1996.
25. H.-J. Stan, 'GC Analysis Based on Mass Spectrometry', in *Analysis of Pesticides in Ground and Surface Water I – Progress in Basic Multiple Residue Methods*, ed. H.-J. Stan, Springer Verlag, Berlin, 67–112, 1995.
26. P.T. Holland, 'Capillary GC with Selective Detectors (ECD, PND, FPD)', in *Analysis of Pesticides in Ground and Surface Water I – Progress in Basic Multiple Residue Methods*, ed. H.-J. Stan, Springer Verlag, Berlin, 141–184, 1995.
27. H.-J. Stan, Th. Heberer, 'Identification and Confirmatory Analysis based on Capillary GC-Mass Spectrometry', in *Analysis of Pesticides in Ground and Surface Water I – Progress in Basic Multiple Residue Methods*, ed. H.-J. Stan, Springer Verlag, Berlin, 141–184, 1995.
28. Th. Heberer, 'Identification and Quantification of Pesticide Residues and Environmental Contaminants in Ground and Surface Water Applying Capillary Gas Chromatography–Mass Spectrometry', TU Berlin thesis (in German), Wissenschaft & Technik Verlag, Berlin, 1995.
29. L. Matter, *Food and Environmental Analysis by Capillary Gas Chromatography – Hints for Practical Use*, Wiley-VCH, Weinheim, 1997.
30. K. Grob, 'Injection Techniques in Capillary GC', *Anal. Chem.*, **66**, 1009A–1019A (1994).
31. K. Grob, *Split and Splitless Injection in Capillary Gas Chromatography – With Some Remarks on PTV Injection*, VCH, Weinheim, 1995.
32. G. Schomburg, *Gas Chromatography – A Practical Course*, VCH, Weinheim, 1990.
33. I. Liška, J. Slobodník, 'Comparison of Gas and Liquid Chromatography for Analyzing Polar Pesticides in Water Samples – Review', *J. Chromatogr. A*, **733**, 235–258 (1996).
34. Fachgruppe Wasserchemie der GdCh (ed.), 'Determination of Phenoxyalkyl Carbonic Acids by Gas Chromatography and Mass Spectrometric Detection after Solid–Liquid Extraction and Derivatization (F14). German Methods for the Examination of Water, Wastewater, and Sludge; Jointly Determinable Substances (Group F), DIN 38 407, Part 14', 1994.
35. W. Weber, 'Ergebnisse und Tendenzen der Analytik von Pflanzenbehandlungsmitteln und ähnlichen Stoffen sowie deren Metaboliten in Grund- und Trinkwässern', in *Schr.-Reihe Verein WaBoLu 68*, Gustav Fischer Verlag, Stuttgart, 109–134, 1987.
36. H.A. Meemken, P. Rudolph, P. Fürst, 'Nachweis und Bestimmung von Chlorophenoxycarbonsäuren durch Kapillar-GC/MS', *Deutsche Lebensmittel-Rundschau*, **83**, 239–245 (1987).
37. Fachgruppe Wasserchemie der GdCh (ed.), 'Determination of Bentazone, Bromoxynil, and 8 Selected Phenoxyalkyl Carbonic Acids by Gas Chromatography and Mass Spectrometric Detection after Solid–Liquid Extraction and Derivatisation (F20). German Methods for the Examination of Water, Wastewater, and Sludge; Jointly Determinable Substances (Group F), DIN 38 407, Part 20', 1996.
38. L. Renberg, 'Ion Exchange Technique for the Determination of Chlorinated Phenols and Phenoxy Acids in Organic Tissue, Soil, and Water', *Anal. Chem.*, **46**, 459–461 (1974).
39. N.C. Jimenez, Y.H. Atallah, Th.R. Bade, 'Capillary Column Gas Chromatographic Determination of Dicamba in Water, Including Mass Spectrometric Confirmation', *J. Assoc. Off. Anal. Chem.*, **72**, 840–844 (1989).
40. C. Schlett, 'Gaschromatographische Bestimmung Polarer Pflanzenschutzmittel in Trink- und Rohwässern', *Z. Wasser–Abwasser-Forsch.*, **23**, 32–35 (1990).
41. G.W. Bruns, S. Nelson, D.G. Erickson, 'Determination of MCPA, Bromoxynil, 2,4-D, Trifluralin, Triallate, Picloram, and Diclofop-Methyl in Soil by GC/MS Using Selected Ion Monitoring', *J. Assoc. Off. Anal. Chem.*, **74**, 550–553 (1991).
42. S. Peldszus, G. Gerhard, J. Brodesser, H.F. Schöler, 'Nachweis von Chlorphenoxyalkancarbonsäuren in gering Belasteten Wässern nach Flüssig/Flüssig-Extraktion', *Vom Wasser*, **75**, 35–45 (1990).
43. L. Clark, J. Gomme, S. Hennings, 'Study of Pesticides in Waters from a Chalk Catchment, Cambridgeshire', *Pestic. Sci.*, **32**, 15–33 (1991).

44. R. Infante, C. Pérez, 'Analysis of Semi-volatile Aromatic Chlorinated Acids in Drinking Water by Liquid–Solid Extraction GC/MS', *Int. J. Environ. Anal. Chem.*, **43**, 165–176 (1991).
45. H.J. Brauch, S. Schullerer, 'Bestimmung von polaren PBSM-Wirkstoffen in Grund- und Trinkwasser mittels GC/MSD nach Derivatisierung', in *Buch der Umweltanalytik Bd.4. Methodik und Applikation in der Kapillargaschromatographie. W3*, eds. R. Weber, E.Weber, GIT Verlag, Darmstadt, 83–85, 1992.
46. H. Clemens, H.F. Schöler, 'Determination of Halogenated Acetic Acids and 2,2-Dichloropropionic Acid in Water Samples', *Fresenius' J. Anal. Chem.*, **344**, 47–49 (1992).
47. K.W. Edgell, E.J. Erb, R.J. Wesselman, J.E. Longbottom, 'Gas Chromatographic/Electron Capture Detection Method for Determination of Chlorinated Acids in Water: Collaborative Study', *J. AOAC*, **76**, 1098–1112 (1993).
48. J. Hodgeson, J. Collins, W. Bashe, 'Determination of Acidic Herbicides in Aqueous Samples by Liquid–Solid Disk Extraction and Capillary Gas Chromatography', *J. Chromatogr. A*, **659**, 395–401 (1994).
49. L.K. Tan, D. Humphries, P.Y.P. Yeung, L.Z. Florence, 'Determinations of Clopyralid, Picloram, and Silvex at Low Concentration in Soils by Calcium Hydroxide–Water Extraction and Gas Chromatography Measurement', *J. Agric. Food Chem.*, **44**, 1135–1143 (1996).
50. M.D. Müller, H.-R. Buser, 'Conversion Reactions of Various Phenoxyalkanoic Acid Herbicides in Soil. 1. Enantiomerization and Enantioselective Degradation of the Chiral 2-Phenoxypropionic Acid Herbicides', *Environ. Sci. Technol.*, **31**, 1953–1959 (1997).
51. D.A. Rimmer, P.D. Johnson, R.H. Brown, 'Determination of Phenoxy Acid Herbicides in Vegetation, Utilizing High-resolution Gel Permeation Chromatographic Cleanup and Methylation with Trimethylsilyldiazomethane Prior to Gas Chromatographic Analysis with Mass-selective Detection', *J. Chromatogr. A*, **755**, 245–250 (1996).
52. P.D. Johnson, D.A. Rimmer, R.H. Brown, 'Adaption and Application of a Multiresidue Method for the Determination of a Range of Pesticides, Including Phenoxy Acid Herbicides in Vegetation, Based on High-resolution Gel Permeation Chromatographic Cleanup and Gas Chromatographic Analysis with Mass-selective Detection', *J. Chromatogr. A*, **765**, 3–11 (1997).
53. S. Butz, H-J. Stan, 'Determination of Chlorophenoxy and other Acidic Herbicide Residues in Groundwater by Capillary Gas Chromatography of their Alkyl Esters Formed by Rapid Derivatization Using Various Chloroformates' *J. Chromatogr.*, **643**, 227–239 (1993).
54. C. Sánchez-Brunete, A.I. García-Valcárcel, J.L. Tadeo, 'Determination of Residues of Phenoxy Acid Herbicides in Soil and Cereals by Gas Chromatography–Ion Trap Detection', *J. Chromatogr. A*, **675**, 213–218 (1994).
55. A.L.S. Pena, M.I.N. Silveira, 'Determination of Chlorophenoxy Herbicides in Waters by Capillary Gas Chromatography with Ion Trap Detection', *J. Chromatogr. A*, **758**, 324–331 (1997).
56. F. Sánchez-Rasero, M.B. Matallo, G. Dios, E. Romero, A. Peña, 'Simultaneous Determination and Enantiomeric Resolution of Mecoprop and Dichlorprop in Soil Samples by High-performance Liquid Chromatography and Gas Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **799**, 355–360 (1998).
57. J.A. Field, K. Monohan, 'Chlorinated Acid Herbicides in Water by Strong Anion-exchange Disk Extraction and In-vial Elution and Derivatization', *J. Chromatogr. A*, **741**, 85–90 (1996).
58. H. Färber, S. Peldszus, H.F. Schöler, 'Gaschromatographische Bestimmung von Aciden Pestiziden in Wasser nach Methylierung mit Trimethylsulfoniumhydroxid', *Vom Wasser*, **76**, 13–20 (1991).
59. I. Brondz, I. Olsen, 'Intra-injector Formation of Methyl Esters from Phenoxy Acid Pesticides', *J. Chromatogr.*, **598**, 309–312 (1992).
60. T.H. Black, 'The Preparation and Reactions of Diazomethane', *Aldrichimica Acta*, **16**, 3–10 (1983).
61. S. Mierzwa, S. Witek, 'Gas–Liquid Chromatographic Method with Electron Capture Detection for the Determination of Residues of Some Phenoxyacetic Acid Herbicides in Water as their 2,2,2-Trichloroethyl Esters', *J. Chromatogr.*, **136**, 105–111 (1977).
62. T. Tsukioka, R. Takeshita, T. Murakami, 'Gas Chromatographic Determination of Triclopyr in Environmental Waters', *Analyst*, **111**, 145–149 (1986).
63. A.S. Lee, M.-K. Hong, A.E. Smith, 'Analysis of 2,2,2-Trifluoroethyl Derivatives of Carboxylic Acid Herbicides by Gas Chromatography with Mass-selective and Electron Capture Detection', *J. AOAC Int.*, **78**, 1459–1464 (1995).
64. H.B. Lee, Th.E. Peart, J.M. Carron, H. Tse, 'Chemical Derivatization Analysis of Pesticide Residues. Part XI. An Improved Method for the Determination and Confirmation of Acidic Herbicides in Water', *J. Assoc. Off. Anal. Chem.*, **74**, 835–842 (1991).
65. D.F. Gurka, F.L. Shore, S.-T. Pan, 'Pentafluorobenzylation Derivatization for Determination of Chlorinated Herbicide Acids', *J. Assoc. Off. Anal. Chem.*, **70**, 889–891 (1987).
66. H.-B. Lee, A.S.Y. Chau, 'Analysis of Pesticide Residues by Chemical Derivatization. VI. Analysis of Ten Acid Herbicides in Sediment', *J. Assoc. Off. Anal. Chem.*, **66**, 1023–1028 (1983).
67. H.-B. Lee, Y.D. Stokker, A.S.Y. Chau, 'Chemical Derivatization Analysis of Pesticide Residues X. Analysis of Ten Acid Herbicides in Natural Water', *J. Assoc. Off. Anal. Chem.*, **69**, 557–560 (1986).
68. H. Agemian, A.S.Y. Chau, 'Determination of Pesticides by Derivative Formation. Part IV. A Sensitive Gas Chromatographic Method for the Determination of

MCPA and MCPB Herbicides after Esterification with 1-Bromomethyl-2,3,4,5,6-Pentafluorobenzene', *Analyst*, **101**, 732–737 (1976).
69. J. de Beer, C.H. van Peteghem, A.M. Heyndrickx, 'Pentafluorobenzyl Ester Derivatives of Nine Chlorophenoxy Acid Herbicides: Spectrometric Properties and Utility for Identification Purposes', *J. Assoc. Off. Anal. Chem.*, **61**, 1140–1154 (1978).
70. A.S.Y. Chau, K. Terry, 'Analysis of Pesticides by Chemical Derivatization. III. Gas Chromatographic Characteristics and Conditions for the Formation of Pentafluorobenzyl Derivatives of Ten Herbicidal Acids', *J. Assoc. Off. Anal. Chem.*, **61**, 633–636 (1976).
71. R.E. Cline, G.D. Todd, D.L. Ashley, J. Grainger, J.M. McCraw, C.C. Alley, R.H. Hill, 'Gas Chromatographic and Spectral Properties of Pentafluorobenzyl Derivatives of 2,4-Dichlorophenoxyacetic Acid and Phenolic Pesticides and Metabolites', *J. Chromatogr. Sci.*, **28**, 167–172 (1990).
72. H.-J. Stan, Th. Heberer, 'Automated Recognition of Target Compounds at Low Levels in Environmental Samples by Means of Capillary Gas Chromatography–Mass Spectrometry with Dedicated Mass Spectral Libraries and the Macro Program AUTARG. II. Application to Pesticides in Groundwater Samples', *J. Chromatogr.*, **653**, 55–62 (1993).
73. H.D. Meiring, G. den Engelsman, A.P.J.M. de Jong, 'Determination of Polar Pesticides by Phase-transfer Catalysed Derivatization and Negative-ion Chemical Ionization Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **644**, 357–365 (1993).
74. S. Butz, Th. Heberer, H.-J. Stan, 'Determination of Phenoxyalkanoic Acids and other Acidic Herbicides at the Low ppt Level in Water Applying Solid-phase Extraction with RP-C18 Material', *J. Chromatogr. A*, **677**, 63–74 (1994).
75. Th. Heberer, S. Butz, H.-J. Stan, 'Detection of 30 Acidic Herbicides and Related Compounds as their Pentafluorobenzylic Derivatives Using Gas Chromatography/Mass Spectrometry', *J. AOAC Int.*, **77**, 1587–1604 (1994).
76. Th. Heberer, S. Butz, H.-J. Stan, 'Analysis of Phenoxycarboxylic Acids and Other Acidic Compounds in Tap, Ground, Surface and Sewage Water at the Low ppt-Level', *Int. J. Environ. Anal. Chem.*, **58**, 43–54 (1995).
77. Th. Heberer, H.-J. Stan, 'Determination of Dichlorprop, Mecoprop, Clofibric Acid and Naphthylacetic Acid in Soil at the Low ppb-Level by GC/MS Using Selected Ion Monitoring', *J. AOAC Int.*, **79**, 1428–1433 (1996).
78. M. Vink, J.M. van der Poll, 'Gas Chromatographic Determination of Acidic Herbicides in Surface Water Samples with Electron Capture Detection and Mass Spectrometric Conformation', *J. Chromatogr. A*, **733**, 361–366 (1996).
79. Ch. Zipper, M.J.-F. Suter, S.B. Haderlein, M. Gruhl, H.-P.E. Kohler, 'Changes in the Enantiomeric Ratio of (*R*)- to (*S*)-Mecoprop Indicate in Situ Biodegradation of this Chiral Herbicide in a Polluted Aquifer', *Environ. Sci. Technol.*, **32**, 2070–2076 (1998).
80. R.P. Evershed, 'Advances in Silylation' in *Handbook of Derivatives for Chromatography*, eds. K. Blau, J.M. Halket, John Wiley & Sons, Chichester, 51–108, 1993.
81. M.J. Bertrand, A.W. Ahmed, B. Sarrasin, V.N. Mallet, 'Gas Chromatographic and Mass Spectrometric Determination of Chlorophenoxy Acids and Related Herbicides as their (Cyanoethyl)dimethylsilyl Derivatives', *Anal. Chem.*, **59**, 1302–1306 (1987).
82. A.W. Ahmed, V.N. Mallet, M.J. Bertrand, 'Determination of Acidic Herbicides and Related Compounds in Water and Soil by Capillary Gas Chromatography Using a Nitrogen–Phosphorus Detector', *J. Assoc. Off. Anal. Chem.*, **72**, 365–366 (1989).
83. Th. Heberer, H.-J. Stan, 'Detection of More than 50 Substituted Phenols as their *tert*-Butyldimethylsilyl Derivatives Using Gas Chromatography–Mass Spectrometry', *Anal. Chim. Acta*, **341**, 21–34 (1997).
84. M.J. Bertrand, S. Stefanidis, B. Sarrasin, '2-Cyanoethyldimethyl(diethyl)aminosilane, a Silylating Reagent for Selective Gas Chromatographic Analysis Using a Nitrogen–Phosphorus Detector', *J. Chromatogr.*, **351**, 47–56 (1986).
85. S. Butz, *Bestimmung von Pflanzenschutzmittelrückständen in Wasser – Entwicklung einer Multimethode*, Wissenschaft & Technik Verlag, Berlin, 1994.
86. WILEY Mass Spectral Library, 'Wiley 6th Edition of 275,821 Spectra (full + condensed) – PBM format Re. C 00.00, 143 MB', 1995.
87. H.-J. Stan et al., 'HPPEST–HP Pesticide Library, Rev. B 00.00, 696 Spectra', Hewlett Packard Company, Palo Alto, USA, 1998.
88. K. Pfleger, H.H. Maurer, A. Weber, *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and their Metabolites*, VCH, Weinheim, 1992.
89. F.W. Karasek, O. Hutzinger, S. Safe (eds.), *Mass Spectrometry in Environmental Sciences*, Plenum Press, New York, 1985.
90. A.G. Harrison, *Chemical Ionization Mass Spectrometry*, CRC Press, Boca Raton, Florida, 1983.
91. H.-J. Stan, Th. Döring, Th. Heberer, S. Butz, 'Detektion von mehr als 30 sauren Herbiziden als Pentafluorbenzylderivate mittels Kapillargaschromatographie und Electron Capture NCI/MS' (unpublished results).
92. B.F. Scott, P.L. Wylie, 'Capillary GC with Atomic Emission Detection (AED)', in *Analysis of Pesticides in Ground and Surface Water II – Progress in Basic Multiple Residue Methods. Chemistry of Plant Protection*, ed. H.-J. Stan, Springer Verlag, Berlin, 33–57, Vol. 12, 1995.
93. H.-J. Stan, M. Linkerhägner, 'Multimethod Applying GC–AED to Water Samples', in *Analysis of Pesticides in Ground and Surface Water II–Progress in Basic Multiple Residue Methods. Chemistry of Plant Protection*, ed. H.-J. Stan, Springer Verlag, Berlin, 59–89, Vol. 12, 1995.

94. B.D. Quimby, J.J. Sullivan, 'Evaluation of a Microwave Cavity, Discharge Tube, and Gas Flow System for Combined Gas Chromatography–Atomic Emission Detection', *Anal. Chem.*, **62**, 1027–1034 (1990).
95. J.J. Sullivan, B.D. Quimby, 'Characterization of a Computerized Photodiode Array Spectrometer for Gas Chromatography–Atomic Emission Detection', *Anal. Chem.*, **62**, 1034–1043 (1990).
96. P.L. Wylie, R. Oguchi, 'Pesticide Analysis by Gas Chromatography with a Novel Atomic Emission Detector', *J. Chromatogr.*, **517**, 131–142 (1990).
97. S.M. Lee, P.L. Wylie, 'A Comparison of the Atomic Emission Detector to Other Element-selective Detectors for the Gas Chromatographic Analysis of Pesticide Residues', *J. Agric. Food Chem.*, **39**, 2192–2199 (1991).
98. M. Linkerhägner, *Bestimmung von Pflanzenschutzmittelrückständen mittels Kapillargaschromatographie und Atomemissionsdetektion*, Shaker, Aachen, 1994.
99. M.K. Fayyad, M.A. Alawi, T.J. El-Ahmed, 'HPLC Determination of the Phenolic Metabolites of Phenoxy Alkanoic Acid Herbicides', *Chromatographia*, **28**, 465–468 (1989).
100. Th.H.M. Noij, A. Brandt, 'HPLC with On-line Solid-phase Extraction for the Trace Analysis of Polar Pesticides', in *Analysis of Pesticides in Ground and Surface Water II – Progress in Basic Multiple Residue Methods. Chemistry of Plant Protection*, ed. H.-J. Stan, Springer Verlag, Berlin, 92–108, Vol. 12, 1995.
101. M. Åkerblom, 'Simple Precolumn Sample Enrichment in High-performance Liquid Chromatography for Determination of Phenoxy Acid Herbicides in Water Samples from Exposure Studies', *J. Chromatogr.*, **319**, 427–431 (1985).
102. S.H. Hoke, E.E. Brueggemann, L.J. Baxter, T. Trybus, 'Determination of Phenoxy Acid Herbicides Using Solid-phase Extraction and High-performance Liquid Chromatography', *J. Chromatogr.*, **357**, 429–432 (1986).
103. A. Di Corcia, M. Marchetti, R. Samperi, 'Extraction and Isolation of Phenoxy Acid Herbicides in Environmental Waters Using Two Adsorbents in One Minicartridge', *Anal. Chem.*, **61**, 1363–1367 (1989).
104. R.B. Geerdink, A.A. van Balkom, H.J. Brouwer, 'Determination of Phenoxyacid Herbicides in Water. Polymeric Pre-column Preconcentration and Tetrabutylammonium Ion-pair Separation on a PRP-1 Column', *J. Chromatogr.*, **481**, 275–285 (1989).
105. A. Betti, G. Lodi, S. Coppi, 'Separation and Determination of Acidic Herbicides on a PRP-1 Polymeric Column', *J. Chromatogr.*, **513**, 219–225 (1990).
106. W. Schüssler, 'Automatic Measurement of Bentazone and Phenoxy Acid Herbicides by HPLC with Three Different Detectors', *Chromatographia*, **29**, 24–30 (1990).
107. C. Sánchez-Brunete, S. Pérez, J.L. Tadeo, 'Determination of Phenoxy Ester Herbicides by Gas and High-performance Liquid Chromatography', *J. Chromatogr.*, **552**, 235–240 (1991).
108. I. Liska, E.R. Brouwer, A.G.L. Ostheimer, H. Lingeman, U.A.Th. Brinkman, 'Rapid Screening of a Large Group of Polar Pesticides in River Water by On-line Trace Enrichment and Column Liquid Chromatography', *Int. J. Environ. Anal. Chem.*, **47**, 267–291 (1992).
109. E.H.R. van der Wal, E.R. Brouwer, H. Lingeman, U.A.Th. Brinkman, 'On-line Trace Enrichment–Column Liquid Chromatography of Polar Pollutants in Surface Water Using Bifunctional Membrane-based Extraction-Disk Cartridges', *Chromatographia*, **39**, 239–245 (1994).
110. S. Chiron, E. Martinez, D. Barceló, 'On-line and Off-line Sample Preparation of Acidic Herbicides and Bentazone Transformation Products in Estuarine Waters', *J. Chromatogr.*, **665**, 283–293 (1994).
111. F.Th. Lange, U. Meier, M. Wenz, H.-J. Brauch, 'Automated Ion-pair Extraction and Ion-pair Chromatography–a Versatile Method in Water Analysis for the Determination of Highly Polar Micropollutants', *Acta Hydrochim. Hydrobiol.*, **23**, 6–12 (1995).
112. M.C. Gennaro, D. Giacosa, C. Abrigo, S. Simonetti, 'Separation of a Mixture of Hydrophilic and Hydrophobic Herbicides by Ion-interaction Reversed-phase HPLC', *J. Chromatogr. Sci.*, **34**, 334–340 (1996).
113. L.E. Vera-Avila, P.C. Padilla, M.G. Hernandez, J.L.L. Meraz, 'On-line Preconcentration, Cleanup and High-performance Liquid Chromatographic Determination of Chlorophenoxy Acid Herbicides in Water', *J. Chromatogr. A*, **731**, 115–122 (1996).
114. V. Pichon, C. Cau Dit Coumes, L. Chen, S. Genu, M.-C. Hennion, 'Simple Removal of Humic and Fulvic Acid Interferences Using Polymeric Sorbents for the Simultaneous Solid-phase Extraction of Polar Acidic, Neutral and Basic Pesticides', *J. Chromatogr. A*, **737**, 25–33 (1996).
115. A. Laganà, G. Fago, A. Marino, 'Determination of Aryloxyphenoxypropionic Acid Herbicides in Water Using Different Solid-phase Extraction Procedures and Liquid Chromatography–Diode Array Detection', *J. Chromatogr. A*, **796**, 309–318 (1998).
116. C. de Ruiter, W.A. Minnaard, H. Lingeman, E.M. Kirk, U.A.Th. Brinkman, R.R. Otten, 'Sensitive and Selective Fluorescence Detection of Phenoxyacid Herbicides by Liquid Chromatography with On-line Ion-Pair Extraction', *Int. J. Environ. Anal. Chem.*, **43**, 79–90 (1991).
117. T. Suzuki, S. Watanabe, 'Liquid Chromatographic Screening Method for Fluorescent Derivatives of Chlorophenoxy Acid Herbicides in Water', *J. AOAC*, **75**, 720–724 (1992).
118. R. Schuster, A. Gratzfeld-Hüsgen, 'Analysis of Phenoxyacid Herbicides and Bentazone by HPLC with Diode-array Detection', Application Note, Hewlett Packard, 1990.

119. A. Brown, R.D. Stephens, I.S. Kim, 'Liquid Chromatography–Mass Spectrometry – A New Window for Environmental Analysis', *Trends Anal. Chem.*, **10**, 330–335 (1991).
120. M.J.I. Mattina, 'Determination of Chlorophenoxy Acids Using High-performance Liquid Chromatography–Particle Beam Mass Spectrometry', *J. Chromatogr.*, **542**, 385–395 (1991).
121. I.S. Kim, F.I. Sasinos, R.D. Stephens, J. Wang, M.A. Brown, 'Determination of Chlorinated Phenoxy Acid and Ester Herbicides in Soil and Water by Liquid Chromatography Particle Beam Mass Spectrometry and Ultraviolet Absorption Spectrophotometry', *Anal. Chem.*, **63**, 819–823 (1991).
122. R.B. Geerdink, P.G.M Kienhuis, U.A.Th. Brinkman, 'Fast Screening Method for Eight Phenoxyacid Herbicides and Bentazone in Water. Optimization Procedures for Flow-injection Analysis–Thermospray Tandem Mass Spectrometry', *J. Chromatogr.*, 329–339 (1993).
123. A. Cappiello, G. Famiglini, A. Berloni, 'Large Volume Injection of Acidic Pesticides by Reversed-phase Micro High-performance Liquid Chromatography', *J. Chromatogr. A*, **768**, 215–222 (1997).
124. R.B. Cole (ed.), *Electrospray Ionization Mass Spectrometry – Fundamentals, Instrumentation and Applications*, John Wiley & Sons, New York, 1997.
125. C. Crescenzi, A. Di Corcia, S. Marchese, R. Samperi, 'Determination of Acidic Pesticides in Water by a Benchtop Electrospray Liquid Chromatography Mass Spectrometer', *Anal. Chem.*, **67**, 1968–1975 (1995).
126. S. Chiron, S. Papilloud, W. Haerdi, D. Barceló, 'Automated On-line Liquid–Solid Extraction Followed by Liquid Chromatography–High-flow Pneumatically Assisted Electrospray Mass Spectrometry for the Determination of Acidic Herbicides in Environmental Waters', *Anal. Chem.*, **67**, 1637–1643 (1995).
127. G. D'Ascenzo, A. Gentili, S. Marchese, D. Perret, 'Determination of Arylphenoxypropionic Herbicides in Water by Liquid Chromatography–Electrospray Mass Spectrometry', *J. Chromatogr. A*, **813**, 285–297 (1998).
128. G. D'Ascenzo, A. Gentili, S. Marchese, A. Marino, D. Perret, 'Multiresidue Method for Determination of Post-emergence Herbicides in Water by HPLC/ESI/MS in Positive Ionization Mode', *Environ. Sci. Technol.*, **32**, 1340–1347 (1998).
129. C. Aguilar, I. Ferrer, F. Borrull, R.M. Marcé, D. Barceló, 'Comparison of Automated On-line Solid-phase Extraction Followed by Liquid Chromatography–Mass Spectrometry with Atmospheric Pressure Chemical Ionization and Particle Beam Mass Spectrometry for the Determination of a Priority Group of Pesticides in Environmental Waters', *J. Chromatogr. A*, **794**, 147–163 (1998).
130. J. Rontree, P. Ackermans, F. van Hoof, J. Claereboudt, 'The LC/MS and LC/MS/MS Analysis of Pesticides in Drinking Water using a High Flow Rate Atmospheric Pressure Ionisation Source', Micromass, Application Note No. 211, Altrincham, UK, 1994.
131. A.W. Garrison, P. Schmitt, D. Martens, A. Kettrup, 'Enantiomeric Selectivity in the Environmental Degradation of Dichlorprop as Determined by High-performance Capillary Electrophoresis', *Environ. Sci. Technol.*, **30**, 2449–2455 (1996).
132. R. Ludwig, W. Gunkel, H. Hühnerfuss, 'Chromatographic Separation of the Enantiomers of Marine Pollutants. Part 5: Enantioselective Degradation of Phenoxycarboxylic Acid Herbicides by Marine Microorganisms', *Chemosphere*, **24**, 1423–1429 (1992).
133. M.W.F. Nielen, '(Enantio-)separation of Phenoxy Acid Herbicides Using Capillary Zone Electrophoresis', *J. Chromatogr.*, **637**, 81–90 (1993).
134. M.W.F. Nielen, 'Trace Enrichment of Environmental Samples in Capillary Zone Electrophoresis', *Trends Anal. Chem.*, **12**, 345–356 (1993).
135. M. Aguilar, A. Farran, V. Marti, 'Analysis of Phenoxyalkyl Acid Herbicides and Chlorophenols by Capillary Zone Electrophoresis', *Sci. Total Environ.*, **132**, 133–140 (1993).
136. A.W. Garrison, P. Schmitt, A. Kettrup, 'Separation of Phenoxy Acid Herbicides and their Enantiomers by High-performance Capillary Electrophoresis', *J. Chromatogr.*, **688**, 317–327 (1994).
137. M. Jung, W.C. Brumley, 'Trace Analysis of Fluorescein-derivatized Phenoxy Acid Herbicides by Micellar Electrokinetic Chromatography with Laser-induced Fluorescence Detection', *J. Chromatogr. A*, **717**, 299–308 (1995).
138. A. Farran, S. Ruiz, C. Serra, M. Aquilar, 'Comparative Study of High-performance Liquid Chromatography and Micellar Electrokinetic Capillary Chromatography Applied to the Analysis of Different Mixtures of Pesticides', *J. Chromatogr. A*, **737**, 109–116 (1996).
139. Y. Mechref, Z.E. Rassi, 'Capillary Electrophoresis of Herbicides. 1. Precolumn Derivatization of Chiral and Achiral Phenoxy Acid Herbicides with a Fluorescent Tag for Electrophoretic Separation in the Presence of Cyclodextrins and Micellar Phases', *Anal. Chem.*, **68**, 1771–1777 (1996).
140. Y. Mechref, Z.E. Rassi, 'Capillary Electrophoresis of Herbicides. II. Evaluation of Alkylglucoside Chiral Surfactants in the Enantiomeric Separation of Phenoxy Acid Herbicides', *J. Chromatogr. A*, **757**, 263–273 (1997).
141. C. Desiderio, C.M. Polcaro, P. Padiglioni, S. Fanali, 'Enantiomeric Separation of Acidic Herbicides by Capillary Electrophoresis Using Vancomycin as Chiral Selector', *J. Chromatogr. A*, **781**, 503–513 (1997).
142. K. Burger, 'Thin-layer Chromatography with Automated Multiple Development (AMD–TLC)', in *Analysis of Pesticides in Ground and Surface Water*

II – Progress in Basic Multiple Residue Methods, ed. H.-J. Stan, Springer Verlag, Berlin, 181–195, 1995.

143. H.-J. Stan, S. Butz, 'Multimethod Applying AMD–TLC Analysis to Drinking Water', in *Analysis of Pesticides in Ground and Surface Water II – Progress in Basic Multiple Residue Methods*, ed. H.-J. Stan, Springer Verlag, Berlin, 197–216, 1995.
144. S. Butz, H.-J. Stan, 'Screening of 265 Pesticides in Water by Thin-layer Chromatography with Automated Multiple Development (AMD–TLC) – A Powerful Tool for Residue Analysis in Drinking Water', *Anal. Chem.*, **67**, 620–630 (1995).
145. R.L. Jones, 'Sampling, Sample Preparation and Preservation', in *Analysis of Pesticides in Ground and Surface Water I – Progress in Basic Multiple Residue Methods*, ed. H.-J. Stan, Springer Verlag, Berlin, 3–18, 1995.
146. M. Åkerblom, 'Extraction and Cleanup', in *Analysis of Pesticides in Ground and Surface Water I – Progress in Basic Multiple Residue Methods*, ed. H.-J. Stan, Springer Verlag, Berlin, 19–66, 1995.
147. M. Spiteller, G. Spiteller, *Massenspektrensammlung von Lösungsmitteln, Verunreinigungen und einfachen aliphatischen Verbindungen*, Springer-Verlag, Wien, 1973.
148. H. Budzikiewicz, *Massenspektrometrie – Eine Einführung*, VCH, Weinheim, 1992.
149. Th. Heberer, H.-J. Stan, 'Determination of Clofibric Acid and *N*-(Phenylsulfonyl)-sarcosine in Sewage, River and Drinking Water', *Int. J. Environ. Anal. Chem.*, **67**, 113–124 (1997).
150. H.-J. Stan, M. Linkerhägner, 'Identifizierung von 2-(4-Chlorphenoxy)-2-methyl-propionsäure im Grundwasser mittels Kapillar-Gaschromatographie mit Atomemissionsdetektion und Massenspektrometrie', *Vom Wasser*, **79**, 85–88 (1992).
151. H.-J. Stan, Th. Heberer, M. Linkerhägner, 'Vorkommen von Clofibrinsäure im aquatischen System – Führt die therapeutische Anwendung zu einer Belastung von Oberflächen-, Grund- und Trinkwasser?', *Vom Wasser*, **83**, 57–68 (1994).
152. Th. Heberer, H.-J. Stan, '*N*-(Phenylsulfonyl)-sarcosine a New Contaminant in Sewage Farm Groundwater', *Fresenius Environ. Bull.*, **3**, 639–643 (1994).
153. Th. Heberer, H.-J. Stan, 'Vorkommen von Polaren Organischen Kontaminanten im Berliner Trinkwasser', *Vom Wasser*, **86**, 19–31 (1996).
154. H.-J. Stan, Th. Heberer, 'Pharmaceuticals in the Aquatic Environment – M.J.F. Suter (coordinator): Dossier Water Analysis', *Analusis*, **25**, M20–23 (1997).
155. Th. Heberer, H.-J. Stan, *Arzneimittelrückstände im aquatischen System*, Wasser & Boden, 20–25, Vol. 50, 1998.
156. U. Dünnbier, Th. Heberer, Ch. Reilich, 'Occurrence of Bis(chlorophenyl)–acetic acid (DDA) in Surface and Groundwater in Berlin', *Fresenius Environ. Bull.*, **6**, 153–159 (1997).
157. Th. Heberer, U. Dünnbier, 'DDT Metabolite bis(chlorophenyl)acetic acid (DDA): The Neglected Environmental Contaminant', *Environ. Sci. Technol.*, **33**, 2346–2351 (1999).

Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Marco Trevisan
Università di Perugia, Italy

Costantino Vischetti
Centro di Studio sulla Chimica e Biochimica dei Fitofarmaci CNR, Perugia, Italy

1 Introduction 1
2 Phenylurea Herbicides 2
3 Sulfonylurea Herbicides 8
3.1 Environmental Samples 8
Abbreviations and Acronyms 13
Related Articles 13
References 14

A review of analytical methods for phenyl and sulfonylurea herbicides is reported. Extraction from environmental samples can be carried out using different techniques like liquid–liquid partition, solid-phase extraction (SPE), microwave-assisted solvent extraction, and so on. Both herbicide classes are analyzed mainly using high-performance liquid chromatography (HPLC) coupled with a mass detector. Quantification limits are around the parts per billion level for many compounds. Multiresidue determination is easy for phenylurea compounds, but for sulfonylureas is limited by the different pK_a values of the compounds.

1 INTRODUCTION

Phenylureas and sulfonylureas are two distinct classes of herbicide developed over different periods.

Phenylureas were developed in the 1960s for use as soil-applied herbicides during the relatively short time interval between sowing the crop and its emergence. During this time the weeds are delicate and shallow-rooted and it is difficult to determine what proportion of a toxic dose of herbicide enters through the roots rather than the shoots.

In most members of the substituted urea herbicides, urea **(1)** is trisubstituted as phenylurea **(2)**. One of the amino groups carries either two methyl groups or one methyl and one methoxy group. The other amino group is substituted with a benzene ring which, in most cases, contains halogen atoms.

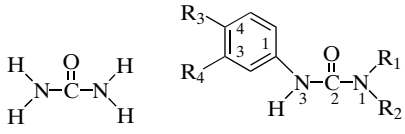

Urea **(1)** Phenylurea **(2)**

In Table 1 some of the most common phenylurea herbicides are reported.

Phenylureas are solids with a low vapor pressure (10^{-6}–10^{-8} mm Hg or 10^{-4}–10^{-6} Pa) at room temperature and their aqueous solubility ranges from 4 to 700 mg L^{-1}.

Hassall[1] classified ureas as herbicides that are mainly soil acting against seedling weeds and are inhibitors of photosynthesis. Some of the nonselective, preemergence and postemergence uses of phenylurea herbicides are listed in Table 2.

Urea herbicides in solution or suspension readily enter plant roots but the extent of translocation of different substances varies greatly. Chloroxuron and neburon tend to remain in the root system whereas monolinuron, fluometuron and metobromuron are largely translocated to the leaves. Once within leaves, some tend to stay near the leaf veins whereas others enter the mesophyll. Since movement is apoplastic, it is accelerated by factors that increase the rate of transpiration, including high temperatures, lower moisture and opening of stomata.

The principal mode of action of ureas is by disruption of the light reaction of photosynthesis, particularly by inhibition at one or both ends of photosystem II at the level of electron acceptor Q (quinone).

Sulfonylurea herbicides were developed in the mid-1970s for broad-spectrum weed control of a number of crops at very low use ratios (2–75 g active ingredient ha^{-1}). They have the advantage of good crop selectivity and very low acute and chronic animal toxicity.

The chemical structure of a generic sulfonylurea herbicide **(3)** is shown, while Table 3 reports some of the most common sulfonylurea herbicides.

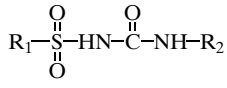

(3)

Hassall[1] classified sulfonylurea herbicides into those that act either at or before cell division.

The site at which the highly specific action of these herbicides is manifest is acetolactate synthase (ALS). This

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Phenylurea herbicides. R_1, R_2, R_3 and R_4 are shown in **(2)**

Name	R_1	R_2	R_3	R_4	Chemical name
Chlorotoluron	$-CH_3$	$-CH_3$	CH_3-	Cl–	3-(3-Chloro-*p*-tolyl)-1,1-dimethylurea
Diuron	$-CH_3$	$-CH_3$	Cl–	Cl–	3-(3,4-Dichlorophenyl)-1,1-dimethylurea
Isoproturon	$-CH_3$	$-CH_3$	$(CH_3)_2CH-$	H–	3-(Isopropylphenyl)-1,1-dimethylurea
Linuron	$-CH_3$	$-O-CH_3$	Cl–	Cl–	3-(3,4-Dichlorophenyl)-1-methoxy-1-methylurea
Metobromuron	$-CH_3$	$-O-CH_3$	Br–	H–	3-(4-Bromophenyl)-1-methoxy-1-methylurea
Neburon	$-CH_3$	$-CH_2(CH_3)_3$	Cl–	Cl–	1-Butyl-3-(3,4-dichlorophenyl)-1-methylurea

Table 2 Uses of some urea herbicides

Herbicide	Uses
Diuron	Total weed control on paths and other noncrop situations. Annual weeds in alfalfa, maize, cotton, pineapple, sugar cane, sorghum
Fluometuron	Annual weeds in cotton and sugar cane
Linuron	Annual weeds pre-emergence in potatoes, carrots, sorghum, soybean. Kills corn marigold, pre-emergence of spring cereals. Kills some grasses until 2 inches (5 cm) high, some dicots until 5 inches (12.5 cm) high. Seedling weeds pre-emergence and postemergence of ornamental bulbs, corms
Chlorbromuron	Annual weeds in potatoes, soybean, carrots, parsnips
Chlortoluron	Wild oats, blackgrass, dicot seedlings in some winter cereals
Fenuron	As pellets, for woody plants control in noncrop situations
Metobromuron	Annual grasses, dicot in potatoes
Metoxuron	Blackgrass, annual grasses, mayweed in some winter cereals
Monolinuron	Many seedling weeds in potatoes and postplanting of leeks. Annual grasses and seedling dicots in carrots, onions
Chloroxuron	Annual weeds postharvest in strawberries; seedling weeds in container plants

is the enzyme that catalyzes the first step in the synthesis of the branched-chain amino acids valine, leucine and isoleucine. The inhibition leads to the rapid cessation of plant cell division and growth.

Sulfonylurea herbicides are one of the most important classes of herbicides and their use is widespread owing to their selectivity towards a great variety of crops. A list of the uses and doses of the main sulfonylurea herbicides is reported in Table 4.

Analytical methods have been developed since the urea and sulfonylurea herbicides were first introduced. There are two distinct approaches to the development of an analytical method, namely single and multiresidue analyses. In both procedures the steps for determination are essentially the same, but there are differences in the solvents used for extraction and in the equipment. Differences were also observed in the dependence of the matrix by which the chemical must be extracted. The analysis of the two classes, urea and sulfonylurea herbicides, will be described separately.

2 PHENYLUREA HERBICIDES

Since the introduction of phenylurea herbicides, many analytical methods have been developed for their determination in water, soil and food samples. Initially, a direct gas-liquid chromatographic (GLC) analysis was used by a number of authors. Katz and Strusz[2] report a gas chromatographic (GC) method for separation of several urea herbicides and their metabolites. The main drawback in using residue analytical procedures for the substituted urea herbicides was the long and tedious procedure required for qualitative identification and for quantitative measurements after alkaline hydrolysis to the corresponding aniline and the determination of the aniline either directly or as a halogenated derivative.[3–5] This was necessary because of the ease of thermal decomposition of many phenylurea herbicides. Using the programmed technique, separation of the parent compounds can be accomplished using a 1.2 m × 6.35 mm XE-60 column or under isothermal conditions at 130 °C using a 5% SE-30 liquid phase on Chromosorb® W 60 to 80 mesh solid support. The separation of parent herbicides under these conditions is illustrated in Table 5. With the programmed technique, metobromuron and fluometuron cannot be separated because their retention times are almost identical. However, fenuron, linuron, chlorobromuron and monuron are easily separated while diuron is barely detectable. Neither neburon nor chloroxuron can be determined using the programmed technique. Under isothermal conditions, metobromuron and fluometuron were easily separated, and chloroxuron was easily distinguishable. Monuron, diuron and neburon all had a poor response and were essentially inseparable. The limit of detection (LOD) for the compounds was around 50 ng.

In the direct GC analysis, for good separations and reproducible analyses of such compounds, operational parameters must be rigidly controlled. Unfortunately,

Table 3 Sulfonylurea herbicides. R_1 and R_2 are shown in structure of sulfonylureas **(3)**

Name	Formula	Chemical name
A: Sulfonylureas characterized by an aliphatic group in R_1 and by a pyridinic ring (diazine) in R_2		
Amidosulfuron	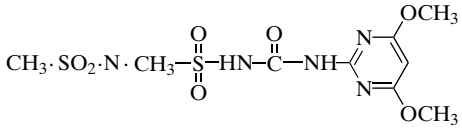 **(4)**	1-(4,6-Dimethoxypyrimidin-2-yl)-3-mesyl(methyl)sulfamoylurea
B: Sulfonylureas characterized by a phenylic ring in R_1 and by a pyridinic ring (diazine) in R_2		
Primisulfuronmethyl	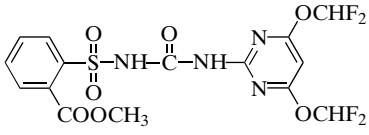 **(5)**	Methyl 2-[4,6-bis-(difluoromethoxy) pyrimidin-2-yl]carbamoyl sulfamoyl benzoate
C: Sulfonylureas characterized by a phenylic ring in R_1 and by a triazinic symmetric ring (1,3,5 triazine) in R_2		
Chlorsulfuron	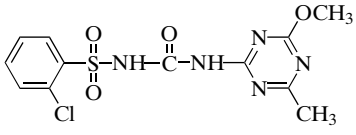 **(6)**	1-(2-Chlorophenylsulfonyl)-3-(4-methoxy-6-methyl-1,3,5-triazin-2-yl) urea
Cinosulfuron	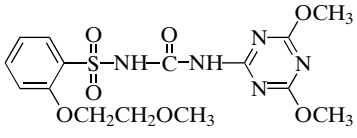 **(7)**	3-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-1-[2-(2-methoxyethoxy)-phenyl-sulfonyl] urea
Metsulfuronmethyl	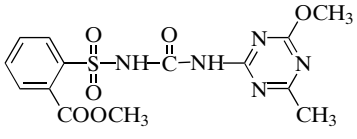 **(8)**	Methyl 2-[3-(4-methoxy-6-methyl-1,3,5-triazin-2-yl)ureidosulfonyl]benzoate
Triasulfuron	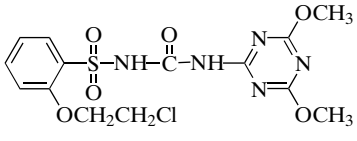 **(9)**	1-[2-(2-Chloroethoxy)phenylsulfonyl]-3-(4-methoxy-6-methyl-1,3,5-triazin-2-yl) urea
Triflusulfuronmethyl	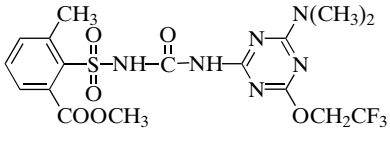 **(10)**	Methyl 2-[4-dimethylamino-6-(2,2,2-trifluoroethoxy)-1,3,5-triazin-2-yl-carbamoylsulfamoyl]-*m*-toluate

(*continued overleaf*)

Table 3 (*continued*)

Name	Formula	Chemical name
D: Sulfonylurea characterized by a pyridinic ring in R_1 and by a pyrimidinic ring (diazine) in R_2		
Nicosulfuron	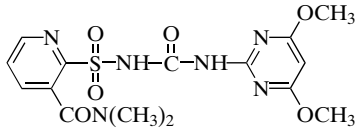 **(11)**	2-(4,6-Dimethoxypyrimidine-2-yl)-carbamoylsulfamoyl-*N*,*N*-dimethylnicotinamide
Rimsulfuron	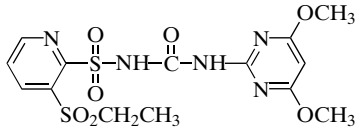 **(12)**	*N*-[(4,6-Dimethoxypyrimidin-2-yl) aminocarbonyl]-3-(ethylsulfonyl)-2-pyridine-sulfonamide
E: Sulfonylureas characterized by different chemical structures		
Bensulfuronmethyl	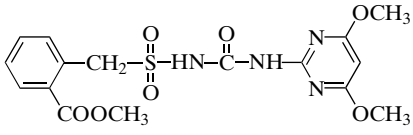 **(13)**	Methyl 2-[[[[[(4,6-dimethoxypyrimidin-2-yl) amino] carbonyl] amino] sulfonyl] methyl] benzoate
Thifensulfuronmethyl	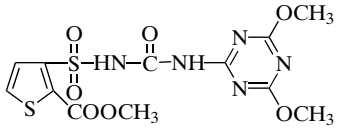 **(14)**	Methyl 3-[3-(4-methoxy-6-methyl-1,3,5-triazin-2-yl)ureidosulfonyl] thiophene-2-carboxylate
Tribenuronmethyl	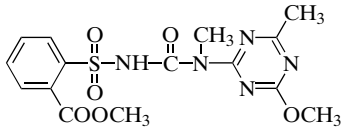 **(15)**	Methyl 2-[3-(4-methoxy-6-methyl-1,3,5-triazin-2-yl)-*N*-methyl-amino] carbonyl amino sulfonyl benzoate

during sample analysis such a restriction often makes direct determination impossible owing to unavoidable interferences. Direct alkylation of urea herbicides was first used in 1973[6] for sample analysis of urea herbicides. Lawrence and Laver[7] reported a GC method for the quantitative analysis of a number of urea herbicides (fluometuron, fenuron, monuron, linuron, diuron, chlorbromuron and chloroxuron) in selected foods (potatoes, carrots, oranges, peas, corn, pineapple, asparagus and spinach) by GLC after alkylation. The compounds were extracted from the crops with ethanol, which was subsequently partitioned between water and chloroform. The organic extract was evaporated to dryness and alkylated with sodium hydride–methyl iodide. The products were extracted with hexane and directly analyzed by GC using a Coulson conductivity detector in the nitrogen mode and a 2 m × 6 mm OD (outside diameter) coiled glass column packed with 4% SE 30/6% QF1 on Chromosorb® WHP (80–100 mesh). No column cleanup was required for any of the samples down to the 0.01-ppm level. Detection limits were about 0.005 ppm. Recoveries were usually 60–100% depending upon the herbicide.

Lawrence[8] reported a comparison of GLC and HPLC for the analysis of some herbicides in foods. This paper is noteworthy in that it is one of the first to propose detection of a urea herbicide (linuron) in foods as well by HPLC with a UV (ultraviolet) detector. Extraction was performed with acetone; extracts were partitioned into petroleum ether–methylene chloride, and cleaned up on a 2% deactivated Florisil® column

Table 4 Uses and doses of sulfonylurea herbicides

Name	Year	Use	Dose (g ha^{-1})
Chlorsulfuron	1982	wheat, barley	10–15
Metsulfuron	1984	wheat, rice	2–8
Sulfumeturon	1984	total	30–40
Chlorimuron	1985	soybean	8–15
Triasulfuron	1988	wheat	8–15
Thifensulfuron	1989	maize, wheat, soybean	7–12
Bensulfuron	1989	rice	40–60
Tribenuron	1989	wheat	8–30
Ethametsulfuron	1990		15–20
Cinosulfuron	1990	rice	40–60
Primisulfuron	1991	maize	10–30
Rimsulfuron	1992	maize, potato	5–15
Nicosulfuron	1992	maize	15–20
Imazosulfuron	1992	rice	20–40
Pyrazosulfuron	1992	rice	10–40
Amidosulfuron	1992	wheat	15–20
Triflusulfuron	1994	sugar beet	30–40
Flazasulfuron	1994	pasture grasses	20–100
Prosulfuron	1994	maize	10–20
Halosulfuron	1995	maize	18–35
Sulfosulfuron	–	wheat	10–20
Azimsulfuron	1996	rice	20–25
Ethoxysulfuron	–	wheat, rice	15–30
NC-330	–	cereal grains	15–20
DPX-T5975	–	cotton	10–25
Oxasulfuron	–	soybean	60–90
Flupyrsulfuron methyl sodium	–	wheat	10

Table 5 Retention times (minutes) of urea herbicides under programmed and isothermal conditions

Compound	Retention time – programmed (min)	Retention time – isothermal (min)
Fenuron	12.60	5.90
Metobromuron	13.95	15.15
Fluometuron	13.95	7.30
Linuron	16.15	23.65
Chlorobromuron	16.95	26.40
Monuron	17.75	1.00
Diuron	20.30	2.35
Neburon	–	1.55
Chloroxuron	–	20.90

before direct chromatographic analysis. Linuron was tested in cabbage and wheat samples both at 2 ppm and 0.2 ppm, with GLC (electron capture detector (ECD) and electrolytic-conductivity detector) and HPLC with a UV detector. All detectors could detect linuron at 0.2-ppm level. However, for electrolytic-conductivity analysis much larger quantities of samples had to be injected because of the low detector sensitivity for linuron. Electron-capture GLC and HPLC gave similar results.

Farrington et al.[9] reported an HPLC analytical method for the determination of several phenylurea herbicides (chlorbromuron, chlortoluron, chloroxuron, diuron, linuron, metobromuron, monolinuron and monuron) in grain, soil and river water. The herbicides were extracted from grain and soil samples with methanol and from water samples with dichloromethane. Organic extracts were chromatographed on microparticulate silica bonded with octadecyltrichlorosilane using a mixture of methanol, water and ammonia as mobile phase. Recoveries were satisfactory; the lower recoveries from wheat may be attributable to residual oil, which remains after the evaporation of the dichloromethane, increasing the volume of the solvent, as this oil is miscible with methanol. The lower limits of detection were estimated to be 0.2 ppm for wheat and soil and 0.01 ppm for river water. Below these levels both coextractives and signal noise interfere.

Numerous HPLC methods for determining phenylurea herbicides in samples were later reported.[10,11]

Over the years, analytical techniques have improved, especially concerning the sensitivity of the methods, which in the 1980s reached parts per billion and parts per trillion levels.

De Kok et al.[12] described several chromatographic procedures for the determination of 15 phenylurea herbicides and their aniline degradation products. Analysis of the separate classes of compounds was achieved by reverse- and normal-phase HPLC followed by determination by capillary GC with an ECD. The anilines were formed by catalytic hydrolysis of the herbicides on silica columns (impregnated with 1 mmol g^{-1} of dimethylamine) and were derivatized with heptafluorobutyric anhydride for determination by GC with ECD (for nanogram levels of herbicides). A range of 14 heptafluorobutyramides could be separated on CP-Sil 5 fused-silica columns, with detection limits (by ECD) of almost 0.2 pg. These and other techniques can be combined as complete schemes for the analysis of mixtures of the herbicides and the corresponding anilines.

Luchtefeld[13] reported an original method for the simultaneous determination of six phenylurea herbicides (chlorbromuron, chloroxuron, diuron, fluometuron, linuron and metobromuron) in eight different food products (asparagus, carrots, celery, corn, grapes, onions, potatoes, strawberries) by liquid chromatography (LC). The samples were extracted twice with methanol, by homogenizing the sample in the centrifuge bottle, and then centrifuging instead of filtering, so less sample manipulation was involved and solvent evaporation was minimized. The partition step was included to provide some cleanup as well as a means of transferring the residues to methylene chloride for further chromatographic cleanup. A single hexane extraction was sufficient to remove some oils that are found in products such as corn. Column

chromatography was included as a final cleanup prior to LC analysis. The Florisil® column was selected since it revealed no inconsistencies between batches. The extracts were analyzed using LC with postcolumn photodegradation, chemical derivatization with orthophthalaldehyde and spectrofluorometry. The LC procedure using postcolumn photolysis and chemical derivatization proved to be very selective for determining phenylurea herbicides in the presence of some food excipients which remain after partition and column cleanup. When UV detection alone was used, these excipients interfered with the analysis. Gradient elution was used to aid the separation of individual herbicides, to help separate any coextractives from herbicides and to remove coextractives from the column in each assay. Two mobile phase systems were used for the gradients: methanol–water (m-w) and acetonitrile–water (a-w). Recoveries were determined by spiking eight different food products with six phenylureas at 0.05 and 0.5 ppm. Average recovery at 0.05 ppm was 95% ± 7.9 and at 0.5 ppm, was 98% ± 6.9. Table 6 reports the LOD and the LOQ for the six phenylureas. The LOD and LOQ were determined by calculating the amount of phenylurea required to give a response that was three times the baseline noise for LOD and ten times the baseline for LOQ. As indicated by the data in Table 6, the detection and quantification limits were lower for the mobile phase system of acetonitrile and water. The described LC procedure with gradient elution, postcolumn photolysis and chemical derivatization has resulted in an unprecedented selectivity for the analysis of phenylureas.

Kuehne et al.[14] presented an analytical method for the determination of phenylurea herbicides in crops by column-switching HPLC, which allowed an improvement in residue analysis of phenylurea herbicides in terms of sensitivity. The method involves hydrolysis of a sample (5 g) of crops with aqueous NaOH for 6 h under reflux, partition of extract by steam distillation into hexane, partition into methanol–0.2 M HCl (33 : 67) and dilution with a mobile phase. Analysis was by column-switching HPLC on a column of Nucleosil® SA (10 μm), followed by a column of Nucleosil® C (10 μm), with methanol–aqueous phosphate buffer of pH 7 (33 : 67 and 3 : 2, respectively) as mobile phase (1 mL min^{-1}), and detection on the first column at 280 nm and on the second column at a vitreous carbon electrode at 0.9 V. Recoveries were 80–110% and detection limits were 0.02 mg kg^{-1} for each herbicide.

Table 6 LOD and limit of quantitation (LOQ) for the six phenylurea using two mobile phases

Herbicide	LOD (ppm)		LOQ (ppm)	
	m-w	a-w	m-w	a-w
Chlorbromuron	0.006	0.003	0.022	0.011
Chloroxuron	0.002	0.001	0.008	0.003
Diuron	0.002	0.001	0.007	0.003
Fluometuron	0.002	0.001	0.006	0.003
Linuron	0.004	0.005	0.014	0.017
Metobromuron	0.004	0.004	0.015	0.014

Di Corcia and Marchetti[15] presented a rapid and sensitive method for the determination of phenylurea herbicides in environmental aqueous samples by extraction with a Carbopack® cartridge followed by HPLC. The water sample is preconcentrated by passage at a flow rate of 150 mL min^{-1} through a 250-mg graphitized carbon black (Carbopack® B) cartridge. After washing with 0.6 mL of methanol, the Carbopack® B trap is connected to a cartridge containing a strong cation exchanger. Organic matter trapped by the Carbopack® cartridge was eluted by passage of 6 mL of methylene chloride–methanol (95 : 5, v/v). Anilines and other basic compounds which can complicate the analysis of phenylureas were readsorbed on the exchanger from the methylene chloride–methanol mixture flowing through it, while the phenylureas passed through completely unretained. After evaporation and redissolution, the sample is subjected to reversed-phase gradient elution HPLC with a 25 cm × 4.6 mm ID (internal diameter) column filled with 5-μm LC-18 reversed-phase packing. The organic modifier was methanol–acetonitrile (85 : 15, v/v). Gradient elution of phenylureas was performed by linearly increasing the percentage of organic modifier from 47–70% in 20 min. It was shown that, in order to fractionate the 14 phenylureas considered, 15% acetonitrile in methanol gave the best results. Phenylureas were detected with the UV detector set at 250 nm. Recoveries of the phenylureas added to water at levels between 30 and 300 ng L^{-1} were always over 92%. LOD (signal-to-noise ratio = 3) was estimated to be about 1 ng L^{-1} for a 2-L sample. One of the advantages of the liquid–solid extraction (LSE) over liquid–liquid partitioning is that sampling and extraction of a water sample can be done simultaneously by passing the water through a sorbent trap as it is pumped at the sampling site. The small volume cartridge can then be transported to the laboratory for desorption and HPLC analysis.

Hatrik and Tekel[16] reported a review of the methods for the determination of phenylurea herbicides, among other pesticide classes, in water samples.

The introduction in the 1990s of new accessories for the classic HPLC (diode array detectors, microcolumns) has led to new efforts in the development of original analytical methods for the determination of phenylurea herbicides in samples.

Brandsteterova et al.[17] described a method for the determination of linuron and monolinuron in milk samples using micro-HPLC. Milk (50 mL) spiked with linuron and monolinuron at 0.01–1 mg kg^{-1}, was mixed

with 50 mL of methanol and 10 mL of 0.1 M sodium acetate. After cooling in ice water for 10–15 min, the precipitate was filtered off and this process was repeated. The combined filtrates were extracted with $CHCl_3$ (3×30 mL) and the combined extracts were dried over Na_2SO_4 and concentrated to 2 mL at 40 °C. The residue was applied to a cartridge containing 3 g of Florisil® topped with 1 g of Na_2SO_4; elution was performed with 10 mL of $CHCl_3$. The eluate was evaporated to dryness and the residue was dissolved in 500 µL of methanol. The solution was filtered and the filtrate was analyzed by HPLC on a column (25 cm × 1 mm) of LiChrosorb® RP-18 (5 µm) with a mobile phase (60 µL min^{-1}) of aqueous 70% methanol and detection at 245 nm. Recoveries were 78.6–91.3% for linuron and 74.9–86.2% for monolinuron: the corresponding coefficients of variation were 0.8–1.8% and 0.7–1.2%.

Boussenadji et al.[18] reported a method for the determination of phenylureas in water using microcolumn HPLC. Water (1 L) was mixed with 10 mL of 0.1 M sodium acetate, filtered and extracted with CH_2Cl_2 or $CHCl_3$ (3×50 mL). The combined extracts were dried using anhydrous Na_2SO_4 and evaporated to dryness at 40 °C. The residue was dissolved in 100 µL of methanol and a 0.2 µL portion of the solution was analyzed by HPLC on a column (30 cm × 1 mm) of Kromasil® C18 (100 Å; 5 µm) with a mobile phase (20–40 µL min^{-1}) of water–methanol (3 : 1, v/v) in 0.01 M $LiClO_4$ adjusted to pH 5.5 using 1% H_3PO_4, UV detection at 254 nm and electrochemical detection at 1.35 V vs Ag–AgCl. Under these conditions, UV detection was more sensitive for linuron, whereas electrochemical detection exhibited a higher sensitivity for metoxuron, diuron and neburon. The detection limit for each herbicide was 0.05 ng.

Lagana et al.[19] reported a method for multiresidue analysis of phenylurea herbicides in crops by HPLC with diode array detector. Crops (20 g) were chopped and blended with 80 mL of methanol. The mixture was filtered and the filtrate was diluted to 100 mL with methanol. A 5-mL portion was passed through a LC-18 cartridge (6 cm × 8 mm) and the cartridge was washed with 5 mL of aqueous 80% methanol. This solution was diluted with 40 mL of H_2O and applied to a Carbopack® B cartridge (6 cm × 12 mm). The phenylureas were eluted with 6 mL of methanol–dichloromethane (1 : 19, v/v) and the eluate was evaporated under N_2 at 40 °C. The residue was dissolved in 250 µL of aqueous 40% methanol and a 50-µL portion was analyzed on a column (25 cm × 4.6 mm) of C1 (5 µm) equipped with a guard column (2 cm × 4.6 mm) of Supelguard® LC-18, with gradient elution (1.5 mL min^{-1}) with acetonitrile–methanol (3 : 17, v/v) in water and detection at 242 nm. Thirteen herbicides were separated using this approach. Recoveries of 0.05–0.2 ppm of each, added to nine different crops, ranged from 76 to 100%, with no interference from the matrix. The detection limit was 0.01 ppm. Confirmation of the identity of each phenylurea was obtained from complete spectra with a diode array detector.

The immunoassay technique has also been developed for analysis of phenylureas in environmental samples.

Newsome and Collins[20] presented the development of an ELISA (enzyme-linked immunosorbent assay) for urea herbicides in foods. Four haptens analogous to phenylurea herbicides were coupled to human serum albumin and ovalbumin in order to prepare immunogens and coating proteins, respectively. Antisera raised in rabbits immunized with three of the immunogens bound strongly to microtiter plates coated with a homologous hapten such that competitive assays were developed with the use of heterologous coating proteins. The specificity of the assays to 13 related compounds was studied. The *N*′-(4-chlorophenyl)-*N*-dimethylurea hapten system was used to determine monolinuron, diuron and linuron in foods. Methanol extracts (25 µL) were incubated at 4 °C for 30 min with antiserum diluted 1 : 4000 with bovine serum albumin, and 200-µL portions were added to microtiter plates which were then incubated for a further 15 min. The plate wells were emptied, washed and incubated at room temperature with a 1 : 1000 dilution of goat anti-rabbit IgG conjugate (210 µL). The wells were again emptied and washed, 210 µL of substrate (*O*-phenylenediamine dihydrochloric acid and H_2O_2 in citrate buffer) were added and after 20 min at room temperature, 2.5 M H_2SO_4 was added before optical density measurement at 492 nm. Satisfactory recoveries of monolinuron and diuron were obtained from food fortified at the 40–160 ng g^{-1} and 0.25–2.0 µg g^{-1} levels, respectively.

Schneider et al.[21] described a highly sensitive and rapid ELISA for the phenylurea herbicides diuron, monuron and linuron. Antisera, obtained from rabbits immunized with monuron–BSA conjugates which contained a C3 or C5 methylene group at terminal nitrogen were adsorbed on microtiter plates overnight at 4 °C in 0.5 M carbonate buffer of pH 9.6 and washed. Monuron, diuron and linuron samples (50 µL) and enzyme tracer (horseradish peroxidase–hapten conjugate; 50 µL) diluted in 0.2 M phosphate buffer in 8% saline, supplemented with 0.1% Tween 20 of pH 7.8 were incubated on the antibody-coated plate for 15 min. A 3,3′,5,5′-tetramethylbenzidine substrate solution (100 µL) was added, the reaction stopped after 15 min and absorbances were measured at 450 nm. Analyses were class or compound selective for individual herbicides dependent on the antisera used. Monuron, diuron and linuron demonstrated 50% inhibition values at 0.4, 0.5 and 0.8 µg L^{-1} and detection limits of 0.04, 0.05 and

0.08 μg L^{-1}, respectively, in buffer. The method was applied to the analysis of water, orange juice and milk.

Katmeh et al.[22] described a competitive ELISA for the determination of the phenylurea herbicide chlortoluron in water and biological fluids. Antibodies were raised in sheep against a *N*′-(3-chloro-4-methylphenyl)-*N*-methyl-*N*-carboxypropyl urea/bovine thyroglobulin conjugate and the resulting antiserum was used in an ELISA for chlortoluron in natural and potable water and in plasma and urine. A microtiter plate was coated with 200 μL per well of the chlortoluron antibody diluted 8000-fold with 0.07 M sodium barbitone buffer of pH 9.6. After overnight incubation at 4 °C, 100 μL of the sample and 100 μL horseradish peroxidase-labeled chlortoluron, diluted 16 000-fold in PBS (phosphate buffered saline) containing 2.5% normal sheep serum, were added to each well and the plate was incubated for 1 h at 37 °C. Then, 200 μL chromagen substrate were added to each well. After a further 30-min incubation, the absorbance was measured at 450 nm. The calibration graph was linear up to 100 μg L^{-1} of chlortoluron and the detection limit was 0.015 μg L^{-1}. The intra- and interassay RSDs (relative standard deviations) ($n = 6$) were 2.6–6.8% and 6.9–14.6%, respectively. Recoveries of chlortoluron added to water and biological fluids were 78–127% and 87–113%, respectively.

Guihot et al.[23] reported the application of immunoassays for molinate and diuron herbicides to field water analysis. The two ELISA described can be carried out either on microwell plates in the laboratory or in polystyrene tubes in the field. Both assays are based on competition with a horseradish peroxidase conjugate of the herbicide for immobilized antibodies against the herbicide, although the assay for diuron requires prior coating of the plate or tube with protein A before coating with the antibody. The detection limits of the laboratory and field assays were 0.5 and 1 ppb for molinate and 0.02 and 0.15 ppb for diuron. There were no appreciable cross-reactions in the assays for molinate, but other phenylurea herbicides cross-reacted in the diuron assays.

Other work helps to present a complete overview of all analytical methods for phenylurea herbicides utilized through the 1980s and 1990s.[24–39]

3 SULFONYLUREA HERBICIDES

Owing to their thermal and chemical instability, the determination of sulfonylurea herbicides at low levels is very challenging and many different approaches have been reported. The earliest approaches utilized normal-phase LC with photoconductivity detection, however this detector has unacceptably long equilibration times. Sulfonylureas are not directly amenable to GC because of their extremely low volatility and thermal instability. Nevertheless, GC has been used in conjunction with diazomethane derivatization, pentaflurobenzyl bromide derivatization and hydrolysis followed by analysis of the aryl sulfonamides. These approaches have not been widely accepted, most likely because none of them perform well for the entire family of sulfonylureas, and while capillary electrophoresis has been evaluated for water and soil, the low injection volumes required may not yield the required sensitivity for certain applications. Enzyme immunoassay has been used for chlorsulfuron and triasulfuron, with LODs ranging from 20 to 100 ppt in soil and water. The most common approaches to sulfonylurea determinations currently involve reversed-phase HPLC with either UV or mass spectrometric (MS) detection. HPLC/MS has been used in conjunction with thermospray, fast-atom bombardment and direct liquid introduction. Electrospray and tandem mass spectrometry (MS/MS) have been combined.

The first sulfonylurea herbicide (chlorsulfuron **(6)**, see Table 3) was introduced in 1982 and since then analytical methods for the determination of sulfonylurea herbicides in different matrices have appeared in journals worldwide.

3.1 Environmental Samples

Sulfonylurea analysis of both soil and water samples can be performed using quite different extraction and determination methods.

3.1.1 Extraction Techniques

Several methods have been proposed using liquid extraction for soil and water samples. Both matrices are extracted by modifying the pH of the medium and using different cleanup techniques.

For chlorsulfuron determination in soil, samples were extracted[40] with aqueous 0.1 M Na_2CO_3–0.1 M $NaHCO_3$ (pH 10). The aqueous solution was washed three times with chloroform. Because the pK_a of chlorsulfuron is about 3.8 the compound remained in the aqueous solution in its anionic form. The chloroform layers were then discarded. The pH of the aqueous solution was adjusted to 3–4 by adding 10% HCl and the solution was extracted three times with toluene and evaporated to dryness by adding 1% glacial acetic acid. The residue was dissolved in the mobile phase and purified in a silica Sep-PAK®.

Soil samples were extracted[41] for nicosulfuron **(11)** and rimsulfuron **(12)** (see Table 3) determination by vortex mixing with aqueous 80% acetonitrile (Table 7). The extract was filtered and the filtrates were concentrated to a reduced volume with nitrogen and directly analyzed

Table 7 Recovery results for multiresidue analysis of nicosulfuron, rimriduron and metabolites

Test compound	Fortification (ppm)	% Recovery	% Standard deviation
Nicosulfuron	0.02–0.20	81–96	7–22
n-Metabolite	0.02–0.20	83–97	8–14
Rimsulfuron	0.02–0.20	74–88	7–11
r-Metabolite	0.02–0.20	80–101	8–22

by a LC/MS (liquid chromatography/mass spectrometry) technique.

Tribenuron-methyl **(15)** was extracted[42] from 500-mL filtered sample of water on a C18 SPE column and cleaned up on a silica column. Chlorsulfuron **(6)**, metsulfuron-methyl **(8)** and thifensulfuron-methyl **(14)** (see Table 3) were extracted and cleaned up similarly after pH adjustment of the sample to 4.5 with aqueous 8.5% H_3PO_4. Each extract was evaporated to dryness in a nitrogen stream and the residues were dissolved in 1 mL aqueous 10% acetonitrile. The method could be applied to other water types, e.g. ground, runoff and surface water.

An analytical screening method for the determination of nine sulfonylurea herbicides (see Table 8) in soil and water was reported by Powley and de-Bernard.[43] Fifty grams of soil were extracted with 100 mL of 80 : 20 (v/v) 0.1 M ammonium carbonate–acetone solution while 200 mL of water sample was directly extracted as indicated below. Soil extract and water samples were passed through a preconditioned 1 g C18 SPE cartridge. The pH of water samples was adjusted to 3.0–3.5 using dilute phosphoric acid. Elution was performed with 10 mL of 0.1% (v/v) glacial acetic acid in ethyl acetate. The eluates were evaporated down and reconstituted in ethyl acetate. After further passage through a 1 g silica SPE cartridge the sample was analyzed by HPLC. This method is not applicable to the determination of nicosulfuron **(11)** and rimsulfuron **(12)** (see Table 3) because these compounds are not quantitatively eluted from the silica cartridge.

Rodriguez and Orescan[44] analyzed 12 different compounds (see Table 9) in water samples using a similar approach. Water (250 mL) was treated with 2.5 mL of acetic acid and drawn through a preconditioned RP-102 SPE cartridge. Elution was performed with 10 mL of methanol and the eluate cleaned up using an anion-exchange cartridge (SAX) stacked on top of an alumina cartridge. The elution was performed with 17 mL of 0.5% acetic acid in dichloromethane with final suction. The final extracts were analyzed using electrospray LC/MS.

In another study[45] nicosulfuron **(11)**, thifensulfuron-methyl **(14)**, metsulfuron-methyl **(8)**, sulfumeturon-methyl, chlorsulfuron **(6)**, bensulfuron-methyl **(13)**, tribenuron-methyl **(15)** (see Table 3) and chlorimuron-methyl recovered from soils were dissolved in 100 µL 90% acetonitrile/5 mM ammonium acetate at pH 3.5 and a 10 µL sample was transferred to a Betasil® C18 (5 µm) column (15 cm × 2 mm ID) with gradient elution (200 µL min^{-1}) with 90% acetonitrile/5 mM ammonium acetate. Chlorsulfuron **(6)**, metsulfuron-methyl **(8)**, thifensulfuron-methyl **(14)** and triasulfuron **(9)**[46] at a concentration of 6 µg kg^{-1} were extracted for 20 min from 20 g of soil using 120 mL of 0.1 M ammonium carbonate at pH 7. After centrifugation for 20 min, the supernatant

Table 9 Recovery results for 12 sulfonylurea herbicides in water. Fortification levels were 0.1–1.0 ppb ($n = 17$)

Test compound	% Recovery	% Standard deviation
Thifensulfuron-methyl	98–104	3–12
Metsulfuron methyl	97–104	3–9
Sulfumeturon methyl	97–104	4–7
Chlorsulfuron	97–103	6–11
Triasulfuron	99–109	5–8
Bensulfuron methyl	98–106	4–5
Prosulfuron	93–95	5–8
Halosulfuron methyl	85–898	10–12
Chlorimuron ethyl	95–96	5–9
Triflusulfuron methyl	92–95	6–7
Primisulfuron methyl	87–89	7–10

Table 8 Recovery results for multiresidue analysis of nine sulfonylureas in soil and water. Soil fortification was 1–10 ppb, water 0.1–0.5 ppb

Test compound	% Soil recovery	% Standard deviation	% Water recovery	% Standard deviation
Azimsulfuron	84–95	5–15	73–79	3–10
Chlorimuron-ethyl	93–107	4–11	88–95	2–8
Chlorsulfuron	80–92	2–10	92–98	6–8
Ethametsulfuron-methyl	92–106	5–9	90–95	1–12
Flupyrsulfuron-methyl	79–94	5–21	81–99	12–15
Metsulfuron-methyl	81–98	5–11	91–98	4–11
Sulfumeturon-methyl	91–101	6–13	85–90	3–8
Thifensulfuron-methyl	76–91	1–13	93–104	2–13
Tribenuron-methyl	81–92	6–17	92–101	8–18

was acidified to pH 3–3.5 using dilute H_3PO_4 and applied to a SPE column. The herbicides were eluted with acidic ethyl acetate, dried at 40 °C, reconstituted in methanol, agitated and mixed, evaporated to <0.4 mL and made up to 1 mL with aqueous 5% acetonitrile.

An SPE method based on polymeric sorbents was developed for the extraction of water and soil samples. SPE was performed[47] on cartridges preconditioned with methanol and water. Water samples were adjusted to pH 3.5 with 10% H_3PO_4 before application. Elution was effected with dichloromethane. The eluent was dried over sodium sulfate, evaporated under nitrogen and dissolved in 30% acetonitrile in high-purity water.

Water was acidified[48] to pH 3 ± 0.2 with anhydrous acetic acid, extracted by reversed-phase SPE eluted with methanol and cleaned up on a tandem system comprising a strong-anion exchange SPE cartridge stacked on an alumina SPE cartridge. The sample was eluted with 0.5% methylene acetic acid in dichloromethane, dried under nitrogen at 30 °C and reconstituted with 0.05 M acetic acid. Water (1 L) was acidified to pH 2 with HCl, mixed with 5 mL methanol and subjected to SPE on Empore® C18 disks using ethyl acetate for elution. The eluate was dried over anhydrous sodium sulfate, evaporated to near dryness and reconstituted in ethyl acetate prior to derivatization with diazomethane solution for 30 min. The solvent was removed under nitrogen and the sample dissolved in toluene for GC analysis. Soil (100 g) was extracted with 0.1 M $NaHCO_3$ with sonication and centrifuged. The extract was filtered, adjusted to pH 2 and mixed with 2 mL methanol prior to SPE and analysis as described above, but using a glass fiber filter on top of the extraction disk.

Seven sulfonylurea herbicides,[49] thifensulfuronmethyl **(14)**, metsulfuron-methyl **(8)**, triasulfuron **(9)**, chlorsulfuron **(6)**, rimsulfuron **(12)**, bensulfuronmethyl **(13)** and tribenuron-methyl **(15)** (see Table 3) were isolated from water by off-line SPE on a Carbograph 4 cartridge (0.5 g, 120–400 mesh). Water was forced through the cartridge at 70–100 mL min^{-1}. The cartridge was flushed with 20 mL of water at <10 mL min^{-1}, with air for 1 min, with 2 mL of methanol and with 12 mL of dichloromethane–methanol (4 : 1) at about 8 mL min^{-1}. Herbicides were recovered by backflushing the cartridge with 8 mL of 10 mM acetic acid in dichloromethane–methanol (4 : 1) at 6 mL min^{-1}. The eluate was evaporated to dryness and the residue was dissolved in 200 µL of 2 mM acetic acid in H_2O–acetonitrile (2 : 1) and analyzed by LC.

Cinosulfuron **(7)**, thifensulfuron-methyl **(14)**, metsulfuron-methyl **(8)**, sulfumeturon-methyl and chlorsulfuron **(6)**[50] (see Table 3) were extracted from soil using microwave-assisted solvent extraction. A mixture of 10 g of soil and 20 mL of CH_2Cl_2–methanol (9 : 1) was heated at 60 °C for 10 min in a microwave extraction system operated at 50% power and with a pressure limit of 100 psi (689 500 Pa). The extract was dried over sodium sulfate and a 5 mL volume was evaporated to dryness at 40 °C. The residue was dissolved in 100 µL of acetonitrile and the solution was diluted with 900 µL of water for HPLC analysis.

Supercritical fluid extraction was optimized for the extraction of metsulfuron-methyl **(8)**, sulfumeturon-methyl and nicosulfuron **(11)** (see Table 3) from soils.[51] Quantitative extraction was only achieved after 30 min at a high CO_2 density (0.9 g mL^{-1}) and with a low methanol modifier concentration (2% w/w). Good recoveries of metsulfuron-methyl **(8)** and sulfumeturon-methyl (75–89%) were obtained, but recoveries of nicosulfuron **(11)** were poor (1–4%).

3.1.2 Determination Techniques

3.1.2.1 Liquid Chromatography Purified samples[40,52] were analyzed by HPLC with a Du Pont Zorbax® SIL column (25 cm × 4.6 mm) set at 35 °C and a photoconductivity detector. This detector is necessary in order to obtain adequate sensitivity and selectivity. A mercury lamp was used in the detector as it provided much greater sensitivity than a zinc lamp. Under these conditions, it was essential that the chromatographic system provided good temperature control of the column and reasonably pulse-free delivery of the mobile phase to minimize baseline fluctuations. Recoveries were measured on 16 spiked samples over a concentration range of 0.2–2.0 ppb for chlorsulfuron **(6)**, rimsulfuron **(12)** (see Table 3) and sulfumeturon-methyl. An average value of 80% was found with a standard deviation of ±16%. Coupled with extraction, cleanup, and isolation procedures, the method provided a means of determining sulfonylureas in soil and water at levels as low as 200 pg g^{-1} (0.2 ppb).

Powley and de-Bernard[43] report an analytical screening method for the determination of nine sulfonylurea herbicides in soil and water by LC with UV detection. The LOQs were 1 and 0.1 ppb in soil and water, respectively, for all herbicides. The samples were analyzed on a Zorbax® SB-Phenyl 4.6 mm × 250 mm column with gradient elution with acetonitrile, pH 2.7 phosphate buffer and pH 6.2 phosphate (1.5 mL min^{-1}) and UV detection at 245 mm. The soil and water recoveries are reported in Table 8.

Cinosulfuron **(7)**, thifensulfuron-methyl **(14)**, metsulfuron-methyl **(8)**, sulfumeturon-methyl and chlorsulfuron **(6)**[50] (see Table 3) were analyzed by HPLC using a 2-µm Microspher® C18 column (10 cm × 4.6 mm ID) operated at 30 °C with methanol–0.1% phosphoric acid at pH 3 (9 : 11) as the mobile phase, an injection volume

of 200 μL and UV detection at 226 nm. Linear calibration graphs were obtained for 5–250 $\mu g\,L^{-1}$. The recoveries of herbicides from freshly spiked soil samples and samples with aged residues at levels between 20 and 1000 $\mu g\,kg^{-1}$ were between 70–100% with RSD between 1–10% ($n = 4$). The detection limits in soils were ~5 $\mu g\,kg^{-1}$.

LC was performed[47] on a 5-μm waters symmetry C18 column (15 cm × 3.9 mm ID) with gradient elution (1 $mL\,min^{-1}$) with 20 mM phosphate buffer of pH 3.4 (solvent A) and acetonitrile (solvent B) from 30% B (held for 4 min) to 80% B over 25 min and detection at 220–230 nm. The primary wavelength for quantification was 225 nm. For some analyses, tribenuron-methyl **(15)** and thifensulfuron-methyl **(14)** were also determined at 262 nm. Recoveries were >85% except for tribenuron-methyl **(15)** (75%) and detection limits were <100 $ng\,L^{-1}$ for water samples.

Seven sulfonylurea herbicides,[49] thifensulfuron-methyl **(14)**, metsulfuron-methyl **(8)**, triasulfuron **(9)**, chlorsulfuron **(6)**, rimsulfuron **(12)**, bensulfuron-methyl **(13)** and tribenuron-methyl **(15)** (see Table 3) were analyzed by LC with UV detection and confirmatory analysis with an electrospray/mass detector. Volumes of 50 μL were applied onto a 5-μm C18 column (25 cm × 4.6 mm ID) with gradient elution (1 $mL\,min^{-1}$) with 3-mM acetic acid in acetonitrile or water (32–62% A in 40 min) and UV detection at 230 nm. Detection limits depended on the particular herbicides and were 0.6–2 $ng\,L^{-1}$ in drinking water (4-L sample), 2–9 $ng\,L^{-1}$ in groundwater (2-L sample) and 13–40 $ng\,L^{-1}$ in river water (0.2-L sample). The recoveries of 10–250 $ng\,L^{-1}$ compound from various water samples were >94%. Confirmatory analysis of the isolated material was performed by LC/MS using the same chromatographic column. The MS data were collected using a time-scheduled three-ion SIM (selected ion monitoring) program. Detection limits for herbicides by LC/MS were 0.5–3 ng.

3.1.2.2 Liquid Chromatography/Mass Spectrometry

Applications involving the determination of sulfonylurea herbicides have included thermospray LC/MS techniques, which have shown a lower LOQ in the range of 20 ppb. Continuous-flow fast atom bombardment (CFAB) LC/MS techniques using packed capillaries have also been used. MS/MS techniques provide additional structural information as well as higher specificity than single mass analyzers equipped with mild ionized ion sources. All these techniques are described below for several sulfonylurea herbicides.

Reiser et al.[53] proposed the application of microcolumn LC/CFAB/MS in environmental studies of sulfonylurea herbicides.

Shalaby et al.[41] introduced the use of thermospray LC/MS for residue analysis of sulfonylurea herbicides and their degradation products. Thermospray LC/MS has emerged as a versatile and selective technique that offers chromatographic separation, positive structural confirmation, quantification and sensitivity, which meet the criteria for multiresidue methods, especially for water-soluble and polar pesticides such as sulfonylureas. The method has been developed for two sulfonylurea herbicides (nicosulfuron **(11)** and rimsulfuron **(12)**) and a major metabolite of each, involves a simple extraction scheme and requires no specific sample cleanup prior to chromatographic analysis. The sample solutions were analyzed by HPLC on a Whatman Partisil C8 column (25 cm × 4.6 mm) with gradient elution (1 $mL\,min^{-1}$) with acetonitrile–0.1 M acetic acid followed by postcolumn addition of 0.5 M ammonium acetate buffer solution and thermospray ionization MS detection at m/z of 143–650. Recoveries for all compounds over four concentration levels are shown in Table 7. The average herbicide recovery was 88.3%, with an average standard deviation of 14.8% for the data population of 96 data points from 24 sample analysis. The LC/MS total ion chromatogram provides a typical profile of the soil studied and shows no interferences at the retention times for the four test compounds. The LC/MS ion chromatogram of a 0.02 ppm fortified soil, which is the lowest fortification level, showed good signal-to-noise ratio for the four test compounds at the method quantitation level.

Bossi et al.[42] report an LC/MS method with an electrospray ionization interface to analyze four sulfonylurea herbicides: chlorsulfuron, metsulfuron-methyl, thifensulfuron-methyl and tribenuron-methyl. MS analysis was accomplished by SIM of the deprotonated molecular ion in negative-ion mode. The four compounds were quantitated with the external standard technique based on the integrated abundance of the molecular ion $[M - H]^-$. Good reproducibility and high sensitivity (detection limits for the four compounds were between 0.002 and 0.005 $\mu g\,L^{-1}$) were achieved. MS/MS in selected reaction monitoring mode on a characteristic product ion (m/z 139 for chlorsulfuron **(6)**, metsulfuron-methyl **(8)**, and thifensulfuron-methyl **(14)**; m/z 153 for tribenuron-methyl **(15)**) was used as a confirmatory method. High sensitivity was also maintained with MS/MS, making it possible to identify sulfonylurea herbicides at very low concentrations.

Rodriguez and Orescan[44] analyzed 12 different compounds in water samples using electrospray LC/MS on a Zorbax® RX-C8 (5 μm) column (25 cm × 4.6 mm) fitted with a guard column (1.25 cm × 4.6 mm) of the same material with gradient elution from 20% (held for 10 min) to 36.8% of acetonitrile in 0.15% acetic acid during the following 21 min and then to 50% (held for 12 min) in

1 min and electrospray MS detection with SIM. The limit of determination was 0.1 ng mL^{-1} for all herbicides, and recoveries are reported in Table 9.

Determination of 18 acidic herbicides was studied with use of HPLC coupled via electrospray ionization with MS or MS/MS.[54] Analytes, chlorsulfuron **(6)**, metsulfuron-methyl **(8)**, thifensulfuron-methyl **(14)** and triasulfuron **(9)** were separated in a column (25 cm × 2 mm ID) of Hypersil® BDS C18 (5 μm). The gradient mobile phase (0.2 mL min^{-1}) contained 1 part acetic acid in 900 parts of methanol–H_2O (1 : 9 to 100 : 0 in 45 min). The electrospray voltage was 5 kV at a temperature of 250 °C and in both MS and MS/MS. SIM was from *m/z* 50 upwards. For the latter the collision gas was Ar at 1 mTorr and 10–30 eV. Detection limits using MS in the SIM mode were generally 0.4–0.5 μg L^{-1}, whereas detection limits were ten times higher using MS/MS detection.

Chlorsulfuron **(6)**, metsulfuron-methyl **(8)**, thifensulfuron-methyl **(14)** and triasulfuron **(9)**[46] were analyzed on a 5-μm ODS Hypersil® column (10 cm × 2.1 mm ID) with gradient elution (0.2 mL min^{-1}) with 10–80% formic acid (0.1%)/acetonitrile over 28 min and SIM/MS detection. Calibration was carried out at four concentrations over 15–170 μg mL^{-1}. The detection limit was <0.1 μg kg^{-1} and recoveries were 80–100%.

Li et al.[45] reported a determination using a triple quadrupole mass spectrometer operated in positive ion mode, via an ion spray under SIM conditions. The nebulization gas (60 psi or 413 700 Pa) was nitrogen and the curtain gas was (1.2 L min^{-1}) nitrogen. The collision gas, argon, was set at an approximate collision thickness of 250 × 10^{12} atoms cm^{-2}. Calibration graphs were linear for 0.05–10 ppb of each sulfonated urea. Suggested acceptance criteria for forensically valid data include a HPLC retention time reproducibility of ±2% and the selection of at least two and preferably three precursor product ions and an absolute abundance of ±20% for selected ions.

3.1.2.3 Capillary Electrophoresis Chlorsulfuron **(6)**, chlorimuron and metsulfuron[55] were analyzed using the capillary electrophoresis technique with an apparatus equipped with an untreated fused-silica capillary (50 cm effective length × 75 μm ID) operated at 25 kV and 35 °C. The electrolyte buffer was 30 mM sodium borate/80 mM SDS (sodium dodecyl sulfate) containing 14% aqueous methanol/20% aqueous isopropanol at pH 7 and detection was at 214 nm. The peak area was used for the quantitation of chlorsulfuron **(6)**, metsulfuron and chlorimuron. Calibration graphs were linear from 0.6–10 ppm of sulfonylurea herbicides with detection limits of 0.63 ppm for chlorimuron and 0.5 ppm for metsulfuron and chlorsulfuron **(6)**. Inter- and intraday RSDs ($n = 10$) were ≤4% and ≤2.4%, respectively. The mean overall recovery for the three herbicides was 95.4 ± 16.1%. For concentrations < 10 ppb the concentration was a limiting factor and HPLC was easier and more accurate than micellar electrokinetic capillary chromatography (MECC). Metabolites[56] of eight sulfonylurea herbicides (chlorsulfuron **(6)**, bensulfuron-methyl **(13)**, tribenuron, triasulfuron, ethametsulfuron, rimsulfuron, chlorimuron and metsulfuron methyl **(8)**) in aqueous solution were separated using the same apparatus. The only difference was the running buffer of 50 mM sodium borate/22 mM SDS/10% methanol at pH 8.

Capillary electrophoresis[48] was performed on a column (122 cm × 75 μm ID; effective length 100 cm) operated at 30 °C and 30 kV with a running buffer of 50 mM ammonium acetate of pH 4.75 in 12% acetonitrile and detection at 240 nm. LC was performed on a 5 μm Zorbax® RX-C8 column (15 cm × 2.7 mm ID) with gradient elution (0.25 mL min^{-1}) with acetonitrile/acetic acid and electrospray MS detection using a time-scheduled SIM (positive mode) of the M + H ions for each compound. The method was used to determine 12 sulfonylurea herbicides in water. Detection limits for LC/MS ranged from 0.01–0.12 ppb.

Bensulfuron-methyl **(13)**, sulfumeturon-methyl, nicosulfuron **(11)**, chlorimuron-ethyl, thifensulfuron-methyl **(14)**, metsulfuron and chlorsulfuron **(6)**[57] were separated, in the order cited, in a 57 cm × 75 μm ID capillary (effective length 50 cm) at 30 kV with (50 mM ammonium acetate buffer of pH 5.0)–acetonitrile (3 : 1) as running buffer and detection at 214 nm. On-column detection limits were 80–100 fM (injection volume 20 nL).

3.1.2.4 Gas Chromatography Thermally stable *N*,*N*′-dimethyl derivatives[58] for the GC analysis of chlorsulfuron **(6)** in Table 3 and metsulfuron-methyl **(8)** in Table 3, both model sulfonylurea herbicides, were produced by methylation of a 1 mL sample in ethyl acetate with 0.5 mL of diazomethane/ethyl acetate for 30 min. The solvent was evaporated under nitrogen and the residue was dissolved in 1 mL of toluene for GC analysis on a column (25 m × 0.2 mm ID) coated with HP-5 (0.33 μm) with temperature programming from 85 °C (held for 1 min) to 150 °C (held for 2 min) at 40 °C min^{-1}, then to 250 °C (held for 20 min) at 5 °C min^{-1} and ECD or nitrogen-phosphorus (N-P) detection. Peak identification was confirmed by GC/MS. The RSD were ≤3% and the calibration graphs were linear for 0.1–10 μg mL^{-1} **(6)** and **(8)** in Table 3. The detection limits were 50 and 100 pg for **(6)** and **(8)** in Table 3, respectively. Quantification limits were <0.1 μg L^{-1} for water and <1 μg kg^{-1} for soil. Recoveries in water were 94–120% and 84–114% for **(6)** and **(8)** in Table 3, respectively, with an RSD of 2–15%. The corresponding recoveries for soil were 59–104% and 65–121% (RSD 8–21%).

Crop determination has been reported only by Krynitsky and Swineford.[59] Grain samples were extracted with acetonitrile and partitioned with hexane. The combined acetonitrile layers were evaporated to dryness at 40 °C and the residue was dissolved in 0.05 M citric acid with sonicating. The resulting solution was cleaned up on an Isolute 5 CX cartridge. The sulfonylureas were eluted with 0.1 M ammonium acetate. The sample was collected and extracted with acetonitrile–CH_2Cl_2 (1 : 19, v/v). The organic extract was dried at 40 °C and dissolved in water. A solution was analyzed using a Model 270A-Ht instrument equipped with a fused-silica capillary (90 cm × 75 μm ID; 68 cm to detector) operated at 35 °C with an applied potential of 25 kV and UV detection. The running buffer comprised 50 mM SDS/25 mM NaH_2PO_4 adjusted to pH 6.15 with 25 mM Na_2HPO_4. The lower detection limit was ~0.02 ppm, except for rimsulfuron **(12)** and tribenuron-methyl **(15)** (0.035 ppm) (see Table 3). Recoveries were 72.9–118.5%.

3.1.2.5 Enzyme Immunoassay Kelley et al.[60] proposed a method for determination of chlorsulfuron **(6)** (see Table 3) in soil by enzyme immunoassay. They describe ELISA, a sensitive immunological technique that can monitor nanogram quantities of chlorsulfuron **(6)** in crude soil extracts. The ELISA technique is based on the ability of animals to produce highly specific antibodies to foreign materials. Antibody is collected from rabbit serum and becomes one reagent in a rapid solid-phase assay that is both specific and sensitive. The assay detects other sulfonylurea herbicides, which are chemically similar to chlorsulfuron **(6)**, but it fails to detect sulfonylurea with different chemical structures. The detection limit in soil was 0.4 ppb.

Other work helps to present a complete overview of all analytical methods for sulfonylurea herbicides utilized through the 1990s.[61–76]

ABBREVIATIONS AND ACRONYMS

ALS	Acetolactate Synthase
CFAB	Continuous-flow Fast Atom Bombardment
ECD	Electron Capture Detector
ELISA	Enzyme-linked Immunosorbent Assay
GC	Gas Chromatographic
GLC	Gas-liquid Chromatographic
HPLC	High-performance Liquid Chromatography
ID	Internal Diameter
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LOD	Limit of Detection
LOQ	Limit of Quantitation
LSE	Liquid–Solid Extraction
MECC	Micellar Electrokinetic Capillary Chromatography
MS	Mass Spectrometric
MS/MS	Tandem Mass Spectrometry
OD	Outside Diameter
PBS	Phosphate Buffered Saline
Q	Quinone
RSD	Relative Standard Deviation
SDS	Sodium Dodecyl Sulfate
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Organic Analysis in Environmental Samples by Capillary Electrophoresis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Reversed Phase Liquid Chromatography • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Chemical Ionization Mass Spectrometry: Theory and Applications • Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. K.A. Hassall, *The Biochemistry and Uses of Pesticides*, 2nd edition, Wiley-VCH, Weinheim, New York, 1990.
2. S.E. Katz, R.F. Strusz, 'Gas Chromatographic Separation of Several Urea Herbicides and their Metabolites', *J. Agric. Food Chem.*, **17**(6), 1409–1411 (1969).
3. J.J. Kirkland, 'Trace Analysis by Programmed Temperature Gas Chromatography', *Anal. Chem.*, **34**, 428–433 (1962).
4. W.H. Gutenmann, D.J. Lisk, 'Electron Affinity Residue Determination of CIPC, Monuron, Diuron, and Linuron by Direct Hydrolysis and Bromination', *J. Agric. Food Chem.*, **12**, 45–48 (1964).
5. S.E. Katz, 'Determination of Substituted Urea Herbicides Linuron, Monuron, Diuron, Neburon and Fenuron in Surface Waters', *J. Assoc. Off. Anal. Chem.*, **49**, 452–456 (1966).
6. F.S. Tanaka, R.G. Wien, 'Gas Chromatography of Substituted Phenylureas by Flash-heater Methylation with Trimethylanilinium Hydroxide', *J. Chromatogr.*, **87**, 85–93 (1973).
7. J.F. Lawrence, G.W. Laver, 'Analysis of Some Carbamate and Urea Herbicides in Foods by Gas–Liquid Chromatography after Alkylation', *J. Agric. Food Chem.*, **26**(6), 1106–1109 (1975).
8. J.F. Lawrence, 'A Comparison of Electron-capture GLC, Electrolytic-conductivity GLC and UV-absorption HPLC for the Analysis of Some Herbicides in Foods', *J. Chromatogr. Sci.*, **14**, 557–559 (1976).
9. D.S. Farrington, R.G. Hopkins, J.H.A. Ruzicka, 'Determination of Residues of Substituted Phenylurea Herbicides in Grain, Soil and River Water by Use of Liquid Chromatography', *Analyst*, **102**, 377–381 (1977).
10. A.E. Smith, K.A. Lord, 'Method for Determining Trace Quantities of the Herbicide Chlortoluron in Soil by Liquid Chromatography', *J. Chromatogr.*, **107**, 407–410 (1975).
11. J.D. Mattice, T.L. Lavy, 'Rapid High-performance Liquid Chromatographic Method for Determining Trace Levels of Fluometuron in Soil', *J. Chromatogr.*, **250**, 109–112 (1982).
12. A. De Kok, Y.J. Vos, C. Van Garderen, T. De Jong, M. Van Opstal, R.W. Frei, R.B. Geerdink, U.A.T. Brinkman, 'Chromatographic Determination of Phenylurea Herbicides and their Corresponding Aniline Degradation Products in Environmental Samples', *J. Chromatogr.*, **288**(1), 71–89 (1984).
13. R.G. Luchtefeld, 'Multiresidue Method for Determining Substituted Urea Herbicides in Foods by Liquid Chromatography', *J. Assoc. Off. Anal. Chem.*, **70**(4), 740–745 (1987).
14. R.O. Kuehne, H. Egli, G. Heinemann, 'Improved Residue Analytical Method for Phenylurea Herbicides', *Fresenius J. Anal. Chem.*, **339**(6), 374–375 (1991).
15. A. Di Corcia, M. Marchetti, 'Rapid and Sensitive Determination of Phenylurea Herbicides in Water in the Presence of their Anilines by Extraction with a Carbopack® Cartridge Followed by Liquid Chromatography', *J. Chromatogr.*, **541**, 365–373 (1991).
16. S. Hatrik, J. Tekel, 'Extraction Methodology and Chromatography for the Determination of Residual Pesticides in Water', *J. Chromatogr. A*, **773**, 217–233 (1996).
17. E. Brandsteterova, V. Chovancova, E. Matisova, J. Holeckova, J. Tekel, 'Application of Micro-HPLC for the Determination of Phenylurea Herbicides in Milk', *J. High Resolut. Chromatogr.*, **14**(10), 696–698 (1991).
18. R. Boussenadji, P. Dufek, M. Porthault, 'Determination of Phenylurea Herbicides in Water Using Micro-column High-performance Liquid Chromatography with UV and Electrochemical Detection', *LC-GC-Int.*, **5**(10), 40–43 (1992).
19. A. Lagana, A. Marino, G. Fago, B.P. Martinez, 'Multiresidue Analysis of Phenylurea Herbicides in Crops by HPLC with Diode Array Detector', *Analusis*, **22**(2), 63–69 (1994).
20. W.H. Newsome, P.G. Collins, 'Development of an ELISA for Urea Herbicides in Foods', *Food Agric. Immunol.*, **2**(2), 75–84 (1990).
21. P. Schneider, M.H. Goodrow, S.J. Gee, B.D. Hammock, 'Highly Sensitive and Rapid ELISA for the Arylurea Herbicides Diuron, Monuron and Linuron', *J. Agric. Food Chem.*, **42**(2), 413–422 (1994).
22. M.F. Katmeh, G.W. Aherne, D. Stevenson, 'Competitive Enzyme-linked Immunosorbent Assay for the Determination of the Phenylurea Herbicide Chlortoluron in

Water and Biological Fluids', *Analyst*, **121**(11), 1699–1703 (1996).
23. S.L. Guihot, J.H. Skerrit, M. Thomas, W. Korth, 'Application of Immunoassay for the Herbicides Molinate and Diuron to Field Water Analysis', *Food Agric. Immunol.*, **8**(3), 169–183 (1996).
24. H.D. Ha, K.G. Bergner, 'Thin-layer Chromatographic Study of Phenylurea and Phenylcarbamate Herbicide Residues in Carrots, Potatoes and Wine I. Detection and Identification of Herbicides', *Dtsch Lebensm. Rundsch.*, **76**(11), 390–394 (1980).
25. H.D. Ha, K.G. Bergner, 'Thin-layer Chromatographic Determination of Phenylurea and *N*-phenylcarbamate Residues. II. Determination in Carrots, Potatoes and Wine', *Dtsch Lebensm. Rundsch.*, **77**(3), 102–106 (1981).
26. U.A.T. Brinkman, A. De Kok, R.B. Geerding, 'Determination of Phenylurea Herbicides via Direct Derivatization with Heptafluorobutyric Anhydride', *J. Chromatogr.*, **283**, 113–126 (1984).
27. J. Lantos, U.A.T. Brinkman, R.W. Frei, 'Residue Analysis of Metoxuron and its Breakdown Product (3-chloro-4-methoxyaniline) by Thin-layer and High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **292**(2), 117–127 (1984).
28. E.Yu. Lopatukhin, S.A. Puzakos, V.A. Popkov, 'Gas-chromatographic Analysis of Herbicide Residues in Medicinal Raw Materials', *Gig. Sanit.*, **6**, 47–49 (1989).
29. S. Perez, J.M. Garcia-Baudin, J.L. Tadeo, 'Determination of Chorotoluron, Isoproturon and Metoxuron in Soil by GLC–NPD (Nitrogen-phosphorus Detection) and Confirmation Using GLC–MS', *Fresenius' J. Anal. Chem.*, **339**(6), 413–416 (1991).
30. H.-J. Stan, P. Klaffenbach, 'Determination of Urea Pesticides by Derivatization and High-resolution GC–MS', *Lebensmittelchemie*, **45**(2), 37–39 (1991).
31. J. Tekel, K. Schultzova, J. Kovacicova, E.Brandsterova, 'Capillary GC with Nitrogen-selective Detection for Online Analysis of Substituted Anilines Formed from Phenylurea Herbicides', *J. High Resolut. Chromatogr.*, **16**(2), 126–128 (1993).
32. R. Boussenadji, P. Dufek, M. Porthault, 'Determination of Phenylurea Herbicides in Water Using Microcolumn High-performance Liquid Chromatography with UV and Electrochemical Detection', *LC GC*, **11**(6), 450, 452, 454 (1993).
33. A. Lagana, G. Fago, A. Marino, B. Pardo-Martinez, 'Rapid Method for Determination of Phenylurea Herbicides in Milk', *Chromatographia*, **38**(1–2), 88–92 (1994).
34. E. Brandsteterova, D. Kralikova, J. Lehotay, J. Tekel, 'HPLC Analysis of Some Substituted Anilines and Phenylurea Herbicides in Milk', *Nahrung*, **38**(4), 439–441 (1994).
35. S. Hatrik, J. Lehotay, J. Tekel, 'Simultaneous HPLC Determination of Phenylurea Herbicides and their Corresponding Anilines in Water after Online Preconcentration', *J. High Resolut. Chromatogr.*, **17**(11), 756–758 (1994).
36. M.C. Gennaro, D. Giacosa, C. Baglietto, M. Gennari, M. Negre, 'Simultaneous Separation of Phenylurea-, Triazine- and Phenoxy-acid Herbicides by Reverse-phase Ion-interaction HPLC. Application to Soil Analysis', *J. Liq. Chromatogr. Relat. Tech.*, **19**(6), 911–924 (1996).
37. S. Hatrik, J. Tekel, 'Extraction Methodology and Chromatography for the Determination of Residual Pesticides in Water', *J. Chromatogr. A*, **733**(1–2), 217–233 (1996).
38. J.-P. Lautie, V. Stankovic, 'Automated Multiple Development TLC of Phenylurea Herbicides in Plants', *J. Planar chromatogr. Mod. TLC*, **9**(2), 113–115 (1996).
39. I. Ferrer, V. Pichon, M.-C. Hennion, D. Barcelo, 'Automated Sample Preparation with Extraction Columns by Means of Anti-isoproturon Immunosorbents for the Determination of Phenylurea Herbicides in Water Followed by Liquid Chromatography–Diode Array Detection and Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **777**(1), 91–98 (1997).
40. E.W. Zanhow, 'Analysis of Herbicide Chlorsulfuron in Soil by Liquid Chromatography', *J. Agric. Food Chem.*, **30**, 854–857 (1982).
41. L.M. Shalaby, F.Q. Bramble, Jr, P.W. Lee, 'Application of Thermospray LC/MS for Residue Analysis of Sulfonylurea Herbicides and their Degradation Products', *J. Agric. Food Chem.*, **40**, 513–517 (1992).
42. R. Bossi, B. Koppen, N.H. Spliid, J.C. Streibig, 'Analysis of Sulfonylurea Herbicides in Soil Water at Sub-part-per-billion Levels by Electrospray Negative Ionization Mass Spectrometry Followed by Confirmatory Tandem Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **81**(4), 775–784 (1998).
43. C.R. Powley, P.A. de-Bernard, 'Screening Method for Nine Sulfonylurea Herbicides in Soil and Water by Liquid Chromatography with Ultra-violet Detection', *J. Agric. Food Chem.*, **46**(2), 514–519 (1998).
44. M. Rodriguez, D.B. Orescan, 'Confirmation and Quantitation of Selected Sulfonylurea, Imidazolinone and Sulfonamide Herbicides in Surface Water Using Electrospray LC–MS', *Anal. Chem.*, **70**(13), 2710–2717 (1998).
45. L.Y.T. Li, D.A. Campbell, P.K. Bennett, J. Henion, 'Acceptance Criteria for Ultratrace HPLC–Tandem Mass Spectrometry: Quantitative and Qualitative Determination of Sulfonylurea Herbicides in Soil', *Anal. Chem.*, **68**(19), 3397–3404 (1996).
46. L.J. Marek, W.C. Koskinen, 'LC–MS Analysis of 4-methoxy-6-methyl-1,3,5-triazin-2-yl Containing Sulfonylurea Herbicides in Soil', *J. Agric. Food Chem.*, **44**(12), 3878–3881 (1996).
47. M.S. Young, 'Preparation of Environmental Samples for Determination of Sulfonylurea Herbicides by Solid-phase

Extraction Using a Polymeric Sorbent', *J. Assoc. Off. Anal. Chem.*, **81**(1), 99–104 (1998).
48. A.J. Krynitsky, 'Determination of Sulfonylurea Herbicides in Water by Capillary Electrophoresis and by Liquid Chromatography–Mass Spectrometry', *J. Assoc. Off. Anal. Chem.*, **80**(2), 392–400 (1997).
49. A. Di-Corcia, C. Crescenzi, R. Samperi, L. Scappaticcio, 'Trace Analysis of Sulfonylurea Herbicides in Water: Extraction and Purification by a Carbograph 4 Cartridge, Followed by Liquid Chromatography with UV Detection, and Confirmatory Analysis by an Electrospray Mass Detector', *Anal. Chem.*, **69**(14), 2819–2826 (1997).
50. N. Font, F. Hernandez, E.A. Hogendoorn, R.A. Baumann, P. van-Zoonen, 'Microwave-assisted Solvent Extraction and Reversed-phase Liquid Chromatography–UV Detection for Screening Soils for Sulfonylurea Herbicides', *J. Chromatogr. A*, **798**(1–2), 179–186 (1998).
51. T. Berglof, W.C. Koskinen, H. Kylin, 'Supercritical Fluid Extraction of Metsulfuron Methyl, Sulfometuron Methyl and Nicosulfuron from Soils', *Int. J. Environ. Anal. Chem.*, **70**(1–4), 37–45 (1998).
52. E.W. Zanow, 'Analysis of the Herbicide Sulfometuron Methyl in Soil and Water by Liquid Chromatography', *J. Agric. Food Chem.*, **33**, 479–483 (1985).
53. R.W. Reiser, A.C. Barefoot, R.F. Dietrich, A.J. Fogiel, W.R. Johnson, M.T. Scott, 'Application of Microcolumn Liquid Chromatography–Continuous-flow Fast Atom Bombardment Mass spectrometry in Environmental Studies of Sulfonylurea Herbicides', *J. Chromatogr.*, **554**, 91–101 (1991).
54. B. Koppen, N.H. Spliid, 'Determination of Acidic Herbicides Using Liquid Chromatography with Pneumatically-assisted Electrospray Ionization Mass Spectrometric and Tandem Mass Spectrometric Detection', *J Chromatogr. A*, **803**(1–2), 157–168 (1998).
55. G. Dinelli, A. Vicari, V. Brandolini, 'Detection and Quantitation of Sulfonylurea Herbicides in Soil at the ppb Level by Capillary Electrophoresis', *J. Chromatogr. A*, **700**(1–2), 201–207 (1995).
56. G. Dinelli, A. Vicari, A. Bonetti, 'Separation of Sulfonylurea Metabolites in Water by Capillary Electrophoresis', *J. Chromatogr. A*, **700**(1–2), 195–200 (1995).
57. W.H. Matchett, W. Winnik, W.C. Brumley, 'Capillary Electrophoretic Behaviour of Seven Sulfonylureas', *J. Capillary Electrophoresis*, **3**(4), 199–204 (1996).
58. P. Klaffenbach, P.T. Holland, D.R. Lauren, 'Analysis of Sulfonylurea Herbicides by Gas–Liquid Chromatography. 1. Formation of Thermostable Derivatives of Chlorsulfuron and Metsulfuron-methyl', *J. Agric. Food Chem.*, **41**(3), 388–395 (1993).
59. A.J. Krynitsky, D.M. Swineford, 'Determination of Sulfonylurea Herbicides in Grains by Capillary Electrophoresis', *J. Assoc. Off. Anal. Chem.*, **78**(4), 1091–1096 (1995).
60. M.M. Kelley, E.W. Zahnow, W.C. Petersen, S.T. Toy, 'Chlorsulfuron Determination in Soil by Enzyme Immunoassay', *J. Agric. Food Chem.*, **33**, 962–965 (1985).
61. A.L. Howard, L.T. Taylor, 'Quantitative Supercritical Fluid Extraction of Sulfonylurea Herbicides from Aqueous Matrices via Solid-phase Extraction Discs', *J. Chromatogr. Sci.*, **30**(9), 374–382 (1992).
62. F. Garcia, J. Henion, 'Fast Capillary Electrophoresis–Ion-spray Mass Spectrometric Determination of Sulfonylureas', *J. Chromatogr.*, **606**(2), 237–247 (1992).
63. D.G. Thompson, L.M. MacDonald, 'Trace-level Quantitation of Sulfonylurea Herbicides in Natural Water', *J. Assoc. Off. Anal. Chem.*, **75**(6), 1084–1090 (1992).
64. A.L. Howard, L.T. Taylor, 'Considerations for Analytical Supercritical Fluid Extraction of Sulfonylureas Employing a Modified Fluid', *J. High Resolut. Chromatogr.*, **16**(1), 39–45 (1993).
65. P. Klaffenbach, P.T. Holland, 'Analysis of Sulfonylurea Herbicides by Gas–Liquid Chromatography. II. Determination of Chlorsulfuron and Metsulfuron-methyl in Soil and Water Samples', *J. Agric. Food Chem.*, **41**(3), 396–401 (1993).
66. J. Tekel, J. Kovacicova, 'Chromatographic Methods in the Determination of Herbicide Residues in Crops, Food and Environmental Samples', *J. Chromatogr.*, **643**(1–2), 291–303 (1993).
67. A.L. Howard, W.J. Yoo, L.T. Taylor, 'Supercritical Fluid Extraction of Environmental Analytes Using Trifluoromethane', *J. Chromatogr. Sci.*, **31**(10), 401–408 (1993).
68. P. Klaffenbach, P.T. Holland, 'Analysis of Sulfonylurea Herbicides by Gas–Liquid Chromatography. III. Mass Spectrometry and Multi-residue Determination', *Biol. Mass Spectrom.*, **22**(10), 565–578 (1993).
69. D. Volmer, J.G. Wilkes, K. Levsen, 'Liquid Chromatography–Mass Spectrometry Multiresidue Determination of Sulfonylureas after Online Trace Enrichment', *Rapid Commun. Mass Spectrom.*, **9**(9), 767–771 (1995).
70. H.Z. Lianz, W.B. Zhang, Q. Jiang, J. Miao, 'Simultaneous Determination of Metsulfuron-methyl, Chlorsulfuron and Bensulfuron-methyl in Various Formulations of Sulfonylurea Herbicides by HPLC–UV Detection', *J. Liq. Chromatogr. Relat. Tech.*, **19**(2), 207–216 (1996).
71. B.M. Berger, N.L. Wolfe, 'Multiresidue Determination of Sulfonylurea Herbicides by Capillary Electrophoresis for Hydrolysis Studies in Water and Sediments', *Fresenius' J. Anal. Chem.*, **356**(8), 508–511 (1996).
72. M.E. Roche, R.P. Oda, D. Machacek, G.M. Lawson, J.P. Landers, 'Enhanced Throughput with Capillary Electrophoresis via Continuous Sequential Sample Injection', *Anal. Chem.*, **69**(1), 99–104 (1997).
73. F. Susanto, H. Reinauer, 'Screening and Simultaneous Quantitative Measurement of Six Sulfonylureas in Serum by Liquid Chromatography/Mass Spectrometry with Atmospheric-pressure Chemical-ionization (APCI LC–

MS)', *Fresenius' J. Anal. Chem.*, **357**(8), 1202–1205 (1997).

74. D. Volmer, K. Levsen, 'Mass Spectrometric Analysis of Nitrogen- and Phosphorus-containing Pesticides by Liquid Chromatography–Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**(7), 655–675 (1994).
75. G. Nilve, M. Knutsson, J.A. Jonsson, 'Liquid Chromatographic Determination of Sulfonylurea Herbicides in Natural Waters after Automated Sample Pretreatment Using Supported Liquid Membranes', *J. Chromatogr. A*, **688**(1–2), 75–82 (1994).
76. T.A. Berger, 'High Efficiency Packed Column Supercritical-fluid Chromatography of Sulfonylurea Herbicides and Metabolites from Large Water Samples', *Chromatographia*, **41**(3–4), 133–140 (1995).

Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

T.R. Steinheimer
USDA–ARS National Soil Tilth Laboratory, Ames, Iowa, USA

1 Introduction 1
2 The Occurrence of Pesticide Residues in Soils and Sediment 2
3 Field Sampling, Preparation, and Preservation 3
4 General Considerations for Extraction 5
5 The Five Most Widely Used Techniques 7
6 Soxhlet Extraction and Solvent-shake Extraction 7
6.1 Principle 7
6.2 Applications 8
6.3 Advantages 8
7 Pressurized Fluid Extraction 9
7.1 History 9
7.2 Principle 9
7.3 Applications 9
8 Supercritical Fluid Extraction 10
8.1 History 10
8.2 Principle 10
8.3 Applications 11
9 Microwave Assisted Extraction 12
9.1 History 12
9.2 Principle 12
9.3 Applications 12
10 Extraction by Sonication 13
10.1 Principle 13
10.2 Applications 13
11 Pesticide Biogeochemistry in Soil: Formation and Extraction of Pesticide-derived Compounds 13
12 Author Statement 15
12.1 Disclaimer 15
Acknowledgments 15
Abbreviations and Acronyms 15
Related Articles 16
Further Reading 16
References 16

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

For pesticide residue determination, field sampling and analysis of agricultural soils and sediments requires the consideration of several seemingly interrelated factors. The mechanisms by which pesticides are adsorbed, absorbed, and desorbed from soils are not well understood for most families of chemicals or for most soils. In addition, soil properties are both spatially and temporally variable, so that the question of sampling methodology becomes important in order to insure that the samples taken are statistically representative of the larger population being examined. Thus, the investigator must weigh the attributes and limitations of each and then choose the best overall compromise among sample size and numbers, analyte specificity, and experimental designs. Following proper collection, the laboratory analyst must then choose among several options for selective extraction techniques for removal of the pesticide analytes of interest. For most applications, one of the following will be used: a classical approach employing a suitable solvent or cosolvent mixture in either a batch shake-out or Soxhlet extraction apparatus. Other more recently developed options include the use of extractant solvents or cosolvent mixtures held under elevated pressures and temperatures (pressurized fluid extraction (PFE)) or pure substances amended with cosolvent modifiers and held above their respective critical points (CPs) (supercritical fluid extraction (SFE)). Another option is the use of ultrasonic energy (sonication) to facilitate extraction. Each of these approaches relies upon an input of sufficient kinetic energy from an externally applied source to the sample slurried with extractant. As the binding energies responsible for the retention of pesticides on soil are overcome, sorbed solutes are released into the solution phase. Studies which address pesticide fate issues often also require determination of soil metabolite and degradate compounds along with the registered parent chemical. For many families of insecticides and herbicides, these same extraction protocols can be used to quantitatively remove the major daughter compounds from soils without requiring an additional extraction step.

1 INTRODUCTION

Soil is the basis of mankind's interactions with his environment because it is the interface between aquatic, atmospheric, and terrestrial ecosystems. Man's ability to manage soil for food and fiber production is a principal tool in combating global hunger and avoiding starvation. On many soils around the world food and fiber production is increased by the effective use of agricultural chemicals to combat a host of yield-reducing pests which attack or compete with crops and/or invade the soil volume. Beginning about 50 years ago,

numerous chemicals have been developed and marketed to farmers for the purpose of mitigating or eliminating pests, primarily insects, weeds, and fungi, which are responsible for reducing both the quality and quantity of harvested crops.

The use of pesticides to maximize crop production has grown steadily over four decades and the development and registration of new chemicals is continuing. In many areas of the world today it is common practice for producers to use combinations of chemicals on each field to control a variety of insects and weeds, either as a preplanting, at planting, or post-emergence application, all applied directly upon or into the soil. These farming operations, while producing a more abundant and higher quality food supply, also carry several unintended and potentially undesirable environmental consequences. For years many public and private institutions and agencies have sponsored research on the environmental fate and transport of the chemical legacy of pesticides used in agricultural production. As a result, it is now apparent that longterm use also yields several unintended outcomes. These include contamination of surface water (via runoff) and shallow groundwater (via soil profile leaching) resources in many areas, and diminished air quality wherever farming operations are conducted (via volatilization from plant and/or soil surfaces). Scientists studying soils have also suggested that the effects of longterm usage of pesticides may also include diminished soil fertility resulting from reduced or unintended shift in biodiversity within soil microorganism and macrofauna populations. Since pesticides are very often applied directly to soils it then becomes imperative to sample and analyze these materials when studying the environmental impacts of farming systems. If the goal is to quantify environmental water quality end-points, then some systematic soil sampling should be conducted parallel with the obligatory water sampling.

The purpose of any soil sample is to obtain information about a particular soil. The sample itself virtually is never the entire soil volume in which the investigator is interested. Therefore, this information from the sample is of interest only insofar as it provides information concerning the larger soil volume. For any soil sample, there are certain properties which describe it. In the broadest context, the purpose of sampling and analysis is to estimate these characteristics with an accuracy which will meet the data needs at the lowest cost. Sampling and analysis of agricultural soils following pesticide applications is a powerful tool through which to gain insight into the nature of the processes which result in both target pest control and potential environmental pollution. Soil sampling and analysis for pesticide content effectively supports many research goals including:

- Enhanced understanding of the cause-and-effect relationship between pesticide use for agricultural production at a given locale and observation of pesticide contamination of the surrounding environment.
- Enhanced understanding of the issues attendant to efficacy of the selected pesticide in controlling intended target pests and how soil properties and conditions might influence the response for a particular chemical.
- Delineation of the microbiological transformation processes which govern the environmental legacy in the soil for each pesticide.
- Support of transformation and transport studies in the field by delineation of soil residence time (half-life) for the parent chemical or family of chemicals within a soil series or combinations of materials.
- Experimental determination of soil affinity coefficients normalized for a soil series or to specific soil properties and useful to modellers as input parameters for model validation as well as pesticide transport simulations.
- Determination of specific potential for 'leaching to groundwater' by examination of the concentration gradient throughout the soil profile by analysis of small increments cut from continuous cores.
- Determination of surface loading from application in order to estimate losses to the atmosphere through volatilization from soil surfaces.
- Evaluation of agricultural soil as a test matrix for analytical chemistry applications; e.g. development of residue methods for new chemicals or for improvements of existing methods or for standardizing techniques against authentic reference materials.

2 THE OCCURRENCE OF PESTICIDE RESIDUES IN SOILS AND SEDIMENT

Soil is a complex living dynamic assemblage of organic and inorganic chemical components. The mineral matrix for most soils is a complex assembly of layered 2:1 phyllosilicate minerals bound together with interlayer metal bridges, most notably aluminum, iron, or manganese. Organic matter in soil is best viewed as a mixture of organic compounds which (a) vary in molecular weight from small monomeric electrolytes [such as the hexose monosaccharide glucose] to very large polysaccharadic electrolytes [such as plant starches or bacterial dextrans], (b) vary in stereochemistry and/or configuration, (c) vary widely in the functionality imposed upon both aliphatic and aromatic carbon skeletons, (d) vary greatly in their potential for electrostatic interactions,

and (e) lack the regularity of structure characteristic of many biopolymers, such as proteins, nucleic acids, and polysaccharides. Adsorption–desorption plays a paramount role in each physical process affecting the behavior of residues in soil. Pesticides are adsorbed by soil organic matter as well as by inorganic soil fractions, and the relationship between adsorption at both possible sites depends upon the specific properties of the soil (sorbent) and the pesticide (sorbate). Retention does not occur by a single mechanism, but rather is best described by a continuum of interactions. In general, binding energies for pesticide–soil associations fall within the 2–10 kcal mole range (van der Waals forces), much weaker than 50–100 kcal $mole^{-1}$ range for covalent bonds. This relatively weak interaction makes possible the extraction of residues by the techniques described herein.

As a class of commercially important synthetic chemicals, pesticides have come under scrutiny due to their wide-spread distribution throughout our terrestrial environment. Today more than 30 classes of chemicals with pesticidal properties are now widely acknowledged by most scientists and regulators, alike. Illustrating this diversity are the variety of chemicals registered for a broad spectrum of problems in weed, insect, and fungal control (Table 1). Nearly all of these chemicals represent post-World War II developments occurring in the last forty years.

Globally, the development of chemicals for control of weeds has proceeded much more rapidly than for other families of chemicals targeting insects, fungi, or other types of pests. In recent years, herbicides have accounted for approximately two-thirds of the global market for synthetic organic pesticides.

Table 1 Chemicals used for weed, insect, and fungus control

Herbicides	Insecticides	Fungicides
Arylanilines	Carbamates	Azoles
Benzoic acids	Organochlorines	Benzimidazoles
Bipyridyliums	OPs	Carboxamides
alpha-Chloroacetamides	Organotins	Dithiocarbamates
Cyclohexadione oximes	Oximino-carbamates	Morpholines
Dinitroanilines	Pyrethroids	OPs
Diphenyl ethers and esters		Phenylamides
Hydroxybenzonitriles		Strobilurine analogs
Imidazolinones		
OPs		
Phenoxyacetic acids		
SUs		
Thiocarbamates		
sym-Triazines		
unsym-triazinones		
Uracils		
Ureas		

SU, sulfonylurea; OP, organophosphate.

The nature of the association of a specific pesticide to a specific soil surface is largely governed by the interactions between functional groups. The major functional groups on inorganic surfaces which contribute to the adsorptive reactivity of pesticides are siloxane ditrigonal cavities contained in phyllosilicate clays and inorganic hydroxyl groups generally associated with metal (hydrous) oxides. Siloxane cavities are Lewis bases which create a net negative charge. Inorganic hydroxyl groups are the most abundant and reactive functional groups found on soil clays and are associated with the surfaces. Their specific reactivity varies depending upon both the number and type of coordination to metal ions. Soil organic matter includes polymeric organic materials, decomposing plant and animal residues, and soil microorganisms. A wide variety of functional groups are represented by the humic substances of soil. While the exact structure of humic substances remains elusive, most researchers agree that among its component chemical-types are large aromatic polymers composed of *N*-heterocycles, quinones, phenols, and benzoic acid derivatives. Recent studies employing infrared and magnetic resonance spectroscopic techniques have also identified carboxyl, carbonyl, phenylhydroxyl, amino, imidazole, sulfhydryl, and sulfonic acid functionality in soil humic materials. Soil humic materials also contain relatively high concentrations of stable free-radicals. The variety of functional groups identified in soil organic fractions along with the sterically and electrostatically controlled interactions between them results in a complex and continuous range of reactivities of soil organic matter toward pesticides. Thus, for any given pesticide, an increase in its polarity, number of functional groups, and ionic nature will all increase the number of potential adsorption mechanisms for that chemical. This is illustrated for the triazine herbicides, in which it has been postulated that van der Waals forces, charge transfer complexes, hydrophobic bonding, cation exchange, and cation bridging, all play a significant role in binding to soil.

3 FIELD SAMPLING, PREPARATION, AND PRESERVATION

Soil sampling is done in support of some types of field survey, which, in turn, are designed with a specific objective(s) in mind. There are five fundamental and interconnected essential steps to consider. First, an unambiguous objective for the sampling should be defined. A clear hypothesis needs to be developed from the question or issue to be researched. All ancillary

information deemed necessary for successful completion of the project should be assembled prior to the start of any experiments.

Second, a clear definition of what constitutes the variables characteristic of the population to be sampled across the field site. The selection of technique for determination is dictated by both the data requirements implied by the objective and the requisite minimal level of detection necessary for each analyte. As an example, for the location-in-space spatial aspect of soil sampling, is centimeter accuracy or minute of angle in latitude and longitude accuracy in the X and Y coordinate axes sufficient, and over what range or distance? For the chemical analysis aspect of soil sampling, what is the optimum sample size to be taken relevant to the minimum detection limit for the analytical method selected? Is a 50-g sample sufficient to achieve a chromatographic instrumental response distinguishable above background on an injection equivalent to 20 pg of pesticide analyte, or is a 2 kg sample required? Typically, in processing soil for pesticide residue determination, analysts have alternative means for adjusting instrumental responses for sample extracts which do not fall within the expected normal ranges. These include phase transfer, concentration, or dilution with solvent.

Third, the selection of an appropriate sampling plan designed in consultation with a skilled statistician is essential. Effective and thorough sampling regimes which indicate how many observations are necessary and at which points samples need to be taken can be categorized in two types. Classical, or design-based, sampling aims at estimating how much pesticide contamination is present. Alternatively, the geostatistical, or model-based, sampling aims to predict as accurately as possible where the contamination is present. In classical *systematic* sampling, sample units are selected at some equally spaced interval, resulting in a uniform distribution of observations over the entire population. Shortcomings to this approach lie in statistical probability in that not all locations within the population have an equal likelihood of being chosen for sampling. Another classical method, simple *random* sampling, allows for each location in the population to have an equal chance of being sampled.[1,2] Simple random sampling tends to undersample small, but potentially important, locations within the population.[3] In the combination of the two approaches, *random systematic* sampling, the locations within the population are arranged in some order, the first sample then randomly taken, and each subsequent sample then taken around that first sample using some fixed interval.[4] *Stratified random systematic* sampling allocates elements of the population into sub-populations based upon some chemical or physical property of the soil prior to sampling, then each stratum is randomly sampled.[5] Strata criteria might be based upon soil classification, slope of the terrain, or erosional class. This approach is considered to be the best option for geostatistical treatment of data because it avoids the possibility of bias resulting from periodicity in sampling locations.[6] Any test of the reliability of spatially distributed data requires assumptions regarding that distribution. Several researchers have criticized the use of classical statistics for treatment of spatially-distributed data, arguing that the conclusions are based upon assumptions not necessarily valid for spatially-dependent variables. The practice of using only a few sampling locations to characterize a spatially complex study area can be a major source error in spatial statistics. This issue is extensively discussed by Uehara et al.,[7] Lascano and Hatfield,[8] and Curran and Williamson.[9] Furthermore, many other sampling regimes exist.

Fourth, implementation of the survey with set procedures and protocols. All sample properties ancillary to the pesticide content and relevant to the objective must be sampled and determined as well and none should be omitted due to cost constraints. If necessary, fewer sampling sites is the best way to accommodate costs. The essential statistical support should be dictated by the sampling objective. Spatial uncertainty in the data is a consequence of taking an unnecessarily limited number of observations, and it can be avoided by taking more samples. Non-spatial uncertainty, such as that resulting from measurement error, can be reduced by repetition of the measurement or by increasing the accuracy of the measurement system. No two locations in any soil sampling regime are both temporally and spatially identical; therefore, a temporal variability experiment must be conducted in parallel time blocks in order for environmental data comparisons to be valid.

Fifth, performing an appropriate analysis(ses) with the given data. Visualization of pesticide residue data in three-dimensional formats is necessary to clearly observe the distribution of results. The extent of the distribution of pesticide concentrations is best presented in the context of soil volume data. In addition, linking the data to hydrologic models can provide insight into the mobility of pesticidal compounds through the soil volume.[10] A geostatistically based multidimensional visualization tool should be used to estimate and display the error or uncertainty associated with the spatial data. The use of the 'fill-in' tools, krieging or weighting, for estimating non-measured values by smoothing areas between sampled locations should be approached with caution, as they frequently can be taken to overrepresent the sampled information.

Depending upon the data requirements, a substantial saving of laboratory resources can be achieved through compositing of samples. A number of field samples are

thoroughly mixed to form a composite which is then subsampled for analysis. Composite sampling can be used in conjunction with stratification wherein each field unit is sampled and composited separately by stratum where it is known that the variabilities within the stratum are inconsequential. However, compositing of samples may result in a loss of spatial and temporal resolution of information.

Soil samples taken for pesticide analysis should be immediately placed in glass or Teflon® containers and sealed under screw-cap such that headspace is minimized. For pesticide determinations which commence with extraction, it is essential that samples be refrigerated as soon as possible, maintained under refrigeration until analysis, and analyzed as soon as possible. The most frequent changes in composition of soil samples during the time between collection and extraction are loss of volatile constituents, or biodegradation, oxidation or reduction of the specific pesticide residues of interest. Low temperatures reduce biodegradation and loss of volatile constituents, but freezing of water-containing samples can result in profound changes in the physical and/or chemical composition. Soil samples pulled from anaerobic zones must not be exposed to air. Typically, soil samples taken for pesticide extraction are not air-dried, nor are they ground to uniform particle size, nor are they screened or sieved, as is frequently done for determination of inorganic constituents, especially total elemental composition. Instead, a representative small portion of the thoroughly mixed bulk sample is taken for moisture determination and that value is then used to calculate a dry weight equivalent for the soil at its field moisture content. In the example, Equation (1) is used in the final calculation to compute analyte concentrations back to a dry-weight basis, and Equation (2) for calculating percent moisture in soil.

$$\text{Wt, dry} = \text{Wt, wet}\ \left(1 - \frac{\%\text{Moisture}}{100}\right) \qquad (1)$$

$$\%\text{Moisture} = 100\ \frac{\text{Wt, wet} - \text{Wt, dry}}{\text{Wt, Wet}} \qquad (2)$$

4 GENERAL CONSIDERATIONS FOR EXTRACTION

In order to extract a pesticide residue from a soil matrix and into an organic solvent phase, several interactions must be overcome and replaced with alternative interactions. The hydrogen bonding, van der Waals, and dipolar interactions between the pesticide molecule and its neighboring pesticide molecules and/or solvent molecules must be broken. The energy required to break these interactions generally increases with increasing polarity of the pesticide molecule. Once the pesticide is released from the matrix, it must be solubilized into the liquid solvent, which requires the breaking of interactions between solvent molecules. This can result from (a) weakening of the solvent–solvent molecule interactions so that they wet the soil surface sufficiently to disrupt the pesticide–soil interactions, or (b) weakening the solvent molecule network sufficiently to allow the pesticide to be accommodated into the solvent molecule environment. For polar solvents commonly used for extractions, the energy associated with this is larger than for less polar solvents, and can be estimated from surface tension and boiling point data alone.

Once solubilized into the solvent, the pesticide is enjoined into new interactions with solvent molecules as a stable solvent–molecule cage is formed. The stability of this interaction results in the release of binding energy, which is largest when the interaction of the pesticide and solvent is a more polar interaction. In order to achieve the maximum solubility in the shortest time, the energy barriers to solvation must be overcome. Increasing the temperature will result in increased rates for the interactions necessary for extraction to occur. Higher temperatures decrease hydrogen-bonding and dipolar interactions occurring on the relatively high energy surfaces of the soil. In addition, thermal energy reduces both surface tension and viscosity of the solvent, thereby enhancing diffusion. Heat will also disrupt stable lattice structures of the soil and impart a degree of kinetic energy to the pesticides. It will also decrease surface tension of the solvent and enhance diffusion. This results in a general decrease in solvent–solvent interactions, thereby accommodating pesticide molecules more easily into the solvent environment.

In analytical applications of soil extraction for pesticide residues, one important indicator characteristic for each chemical is its net molecular polarity. For most nonionic pesticides, polarity can be estimated from water solubility data. While pesticide association with soils material is mechanistically very complex, a useful rule-of-thumb for analysts choosing one among several extraction techniques is that water solubility and the soil-affinity coefficient usually enjoy an inverse relationship. The more strongly bound to soil are the pesticides, the more nonpolar and insoluble they appear to behave. Generally, more polar pesticides are best extracted from soil with more polar solvents or solvent mixtures. Table 2 presents water solubility range data for the most common families of herbicides, insecticides, and fungicides. Only the backbone structures which define the family are shown, as the specific substitution pattern for each chemical will also exert a great influence on extraction efficiency. Numerous resource materials are available which provide this type of information. The most widely used are the *Farm Chemicals Handbook*, updated annually and distributed by the Meister Publishing Co., Willoughby, Ohio; the

Table 2 Guideline to polarity, by core structures for herbicide, insecticide and fungicide family, estimated from water solubility range

	WS Range[a]	Examples	Core structure
Herbicides			
Sulfonylurea	10^2–10^4	chlorsulfuron, nicosulfuron, halosulfuron, primisulfuron	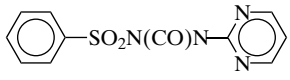
Imidazolinone	10^1–10^4	imazethapyr, imazaquin, imazapyr	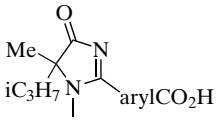
Chloroacetamide	10^2–10^4	metolachlor, alachlor, dimethenamid	Aryl–N–C(O)–CH_2Cl
Triazine	10^0–10^4	simazine, ametryne, prometone	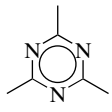
Thiocarbamate	10^0–10^6	butylate, thiobencarb, vernolate	–S–C(O)–N
N-phenylurea	10^0–10^3	linuron, diuron, monuron	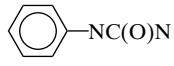
Dinitroaniline	10^{-1}–10^0	trifluralin, pendamethalin, benefin, oryzalin	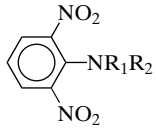
Diphenylether	10^{-1}–10^6	bifenox, fluazifop-butyl	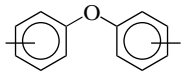
Chlorophenoxy	10^1–10^6	2,4-D, 2,4,5-T, silvex, MCPA	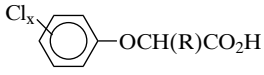
Benzamide	10^0–10^2	isoxaben, flamprop	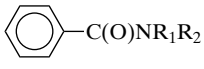
Insecticides			
Organophosphate	10^1–10^6	malathion, diazinon, acephate, chlorpyrifos, parathion, ethion	R_1(O or S)P=(O or S)(O or S)R_2R_3
Carbamate	10^1–10^5	carbofuran, aldicarb, oxamyl	–OC(O)N
Organochlorine	10^{-3}–10^0	DDE, lindane, methoxychlor, aldrin, dieldrin, endrin, mirex	$(Cl_xAryl)_2CHCCl_3$
Pyrethroid	10^{-3}–10^{-1}	fenvalerate, cyfluthrin, permethrin	cyclopropyl–C(O)O–CH–diphenylether
Fungicides			
Imidazole	10^0–10^4	carbendazim, thiabendazole	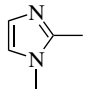
Dithiocarbamate	10^1–10^5	mancozeb, zineb, metiram	–S–C(S)–N
Triazole	10^1–10^3	propiconazole	
Phthalimide	10^0–10^1	captan, folpet	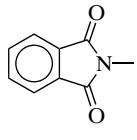

[a] Water solubility, mg L^{-1}, at 293–298 K.

Herbicide Handbook, published by the Weed Science Society of America, and also distributed by Meister; the *Pesticide Properties Database*, compiled and distributed by USDA; and the *Pesticide Manual*, distributed by the British Crop Protection Council, Surrey, United Kingdom. For the nonionic pesticides, most have water solubility in the range of 5–500 ppm at 273–293 K. For ionic pesticides capable of forming stable cations or anions, or those which are formulated as salts, pH-dependent higher solubilities prevail.

Quantifying the efficiency of extraction for pesticide residues incurred in agricultural operations generally requires spike recovery studies. In the most widely practiced approach soils are amended with known concentrations of each target analyte, extracted, and the concentration in the recovered extract is determined by a suitable method, typically gas or liquid chromatography. Alternatively, radiometric analysis of soils amended with radioisotopically enriched congeners of the pesticide being studied is an often used technique in biodegradation studies. Percent recovery is computed as the measure of extraction efficiency. Underlying the validity of this approach is the assumption that the spiking procedure accurately and completely simulates the processes in the agricultural field by which pesticides, their degradates, and metabolites, become bound to soil. Native pesticide is probably retained by combinations of many binding mechanisms, some of which are not duplicated by most spiking techniques. The retention process, both adsorption and absorption is very complex, and may involve different mechanisms of association and all of the components of the soil: the organic and inorganic solid phases, the water phase, and the gaseous phase. For most pesticides, there is likely a continuum of mechanisms responsible for sorption onto soil. The spiking technique recommended here is less disruptive than most of the other recommended options because it does not involve elevated temperature, reduced pressure, mechanical agitation, or the use of organic solvents. For example, the use of large volumes of methanol may promote methanolysis reactions occurring either on the soil components or with the pesticide itself.

> ''Each sample used for recovery studies is spiked with a mixture containing each of the target analytes, by weighing 10.0 gm portions of a well-mixed 60-mesh dry-sieved soil into a 250-mL flat-bottom centrifuge bottle. An appropriate volume of spiking solution, 5–500 μL of acetonitrile solution prepared at a stock concentration of 100–200 mg L^{-1} for each target analyte is diluted with 25 mL of organic-free, distilled, and deionized water. The mixture is transferred quantitatively to the sample bottle. It is then stirred magnetically for several hours and evaporated to dryness (48 h) in an air stream, uncovered, in a fume hood. These air-dried samples are extracted directly without additional pre-treatment.''

5 THE FIVE MOST WIDELY USED TECHNIQUES

Classical methods for extraction of pesticides from soils and sediment materials generally involve the use of a solvent selected because of its optimum solubility with respect to the analytes(s) and its minimum solubility with respect to the sample matrix. In the simplest application, solvent is added to the soil sample while it is agitated to allow the analytes to dissolve into the surrounding liquid until all have been removed. This technique works well when the analyte(s) of interest are very soluble and the sample matrix is in a finely-divided and uniform physical state, or is very porous, thereby permitting closer association between solvent and solute molecules. When better contact between analyte and solvent is required, additional steps can be taken to insure maximum removal. The most popular of these for pesticide residue extractions are additional heating, additional pressure, and sonication, all of which require application of externally-applied forces to the soil–solvent mixture.

6 SOXHLET EXTRACTION AND SOLVENT-SHAKE EXTRACTION

6.1 Principle

Traditionally, solvent extraction has been the most widely used approach to removal of pesticides from soils. Techniques range from simple batch extractions using mechanical shakers to sophisticated devices which continuously recycle extracting solvent. By far the most widely used apparatus is the Soxhlet extractor (Figure 1). It was conceived and developed by German chemist Franz von Soxhlet (1848–1913) more than 100 years ago. In this device a soil sample is contained in a thimble of some porous fiber composition and positioned over a boiling solvent. More recently, glass extraction thimbles with fritted disk sealed in the base have become popular for soil–pesticide extractions. Any desired volatile solvent may be used, either water-miscible or water-immiscible. The solvent is heating to boiling, where it vaporizes. When it condenses, it drops onto the soil contained in the thimble, and extracts soluble compounds. When the liquid level fills the body of the extractor, the solvent containing the extracted pesticide is automatically siphoned back into the boiling solvent. After many cycles and many hours, the solvent is evaporated, leaving a residue for either further purification treatment or direct instrumental analysis. Over the years several modifications have been described, the best known being the Goldfisch technique. In this procedure, the

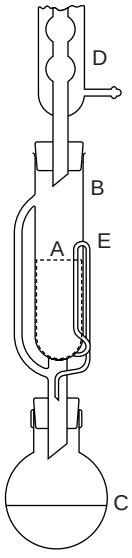

Figure 1 Soxhlet extractor. A, thimble; B, extractor body; C, solvent flask; D, condenser; E, siphon tube.

condensed solvent is allowed to drip into and through the sample thimble, immediately returning to the boiling solvent. It has no filling and siphoning action as in the original Soxhlet apparatus. Soxhlet extraction is simple, effective, and relatively inexpensive, other than the costs of solvent disposal. For many pesticides, it provides good analyte recovery. However, its traditional operation is slow and it requires large volumes of solvents. While extractions may last for 24 hours or more, they require very little user intervention once the process has begun. Small-volume Soxhlet systems are available, but the sample size often is dictated by the pesticide residue concentration, the minimum mass of analyte necessary for a representative sampling, and the chromatographic technique chosen for the determination. All of these often combine to require a larger soil sample size.

Through the years and through many improvements the original Soxhlet techniques remained time-consuming and prone to variability. By the mid-1970s a new accelerated solvent extraction (ASE) technique was described which reduced extraction time to as little as 30 min. Known as the Randall[11] modification, the sample thimble is totally immersed in the boiling solvent. The underlying principle is that the pesticide residue to be extracted is more soluble in hot solvent than in cold solvent at room temperature. This method has become the basis for many commercial automated solvent extraction systems derived from the original Soxhlet. Its limitations notwithstanding, Soxhlet remains the most exhaustive option available to the analyst.

6.2 Applications

Table 3 compares extraction efficiency, as determined by percent recovery, for a series of representative CHIs spiked onto clay samples for both conventional Soxhlet and automated Soxhlet techniques.[12] In every case, the Randall technique yields higher recoveries in a fraction of the time. While these data apply only to the nonpolar, very hydrophobic, low water-solubility insecticides, similar results are obtained with many classes of herbicides, including chlorophenoxy acids, triazines, chloroacetamides, and dinitroanilines. In recent years, automated extraction techniques have gained global acceptance as well as regulatory approval from the United States Environmental Protection Agency (USEPA). An exemplary USEPA approach can be found in the solid waste series, SW-846. Method 3541 describes the extraction of organic analytes, including many pesticides, from soil, sludge, and sediment by automated solvent extraction.[13] In addition to semivolatile organic compounds, which includes most pesticides, this method is applicable to polychlorinated biphenyls (PCBs) as well.

6.3 Advantages

Modern automated Soxhlet extraction offers several benefits over more traditional systems. The Randall boiling

Table 3 Comparison of extraction efficiency, as determined by percent recovery, for a series of representative chlorinated hydrocarbon insecticides (CHIs) spiked onto clay samples for both conventional Soxhlet and automated Soxhlet techniques[a]

Compound	Soxhlet[b]		Automated Soxhlet[c]	
	Mean recovery	RSD	Mean recovery	RSD
alpha-BHC	57.4	47.5	94.9	5.5
Hexachloro-benzene	55.4	24.5	81.7	7.1
delta-BHC	65.0	27.1	104.0	9.7
Heptachlor	59.6	34.1	87.1	5.4
Aldrin	69.8	8.8	78.2	5.7
Heptachlor epoxide	72.0	20.8	92.4	0.6
trans-Chlordane	75.6	12.5	85.8	2.2
Endosulfan I	76.4	5.5	90.5	2.0
Dieldrin	74.4	20.0	68.8	2.6
Endrin	81.0	3.9	112.0	4.4
Endosulfan II	78.5	6.7	90.3	10.0
p,p′-DDT	73.6	38.5	61.4	6.5
Mirex	75.5	15.0	84.0	5.1

[a] Reprinted from Lopez-Avila.[12]
[b] Extraction with hexane–acetone (1 : 1 v/v); 16-hours; mean of triplicate determinations.
[c] Extraction with hexane–acetone (1 : 1 v/v); 60-minute boiling and 60-minute rinsing times; quadruplicate determinations.
DDT, 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane.

step accelerates the extraction three- to five-fold over conventional Soxhlet. Automated systems allow users to extract multiple samples simultaneously, thereby saving time and money. For many pesticides, especially hydrophobic nonpolar compounds, repeatability of recovery and efficiency are improved, often by more than 20%. Commercially available extractors use less solvent than conventional systems, require less fume hood space, and include built-in safety features with mechanisms designed to minimize unsafe operations.

The traditional Soxhlet procedure also offers several benefits for some users. It is often the method of choice for smaller laboratories due to its relatively modest capital investment. In addition, USEPA considers it to be the exhaustive extraction technique for removal of organic compounds from solid matrices, so that it has become the *de facto* benchmark for comparisons.[14]

Many of the newest generation of herbicides and insecticides are nonionic, very polar, and highly water soluble. Consequently their residues are easily removed from cultivated soils merely by contact with various aqueous solutions. Such is the case for the SU metsulfuron-methyl (CAS #74223-64-6), which is readily extracted from cultivated soils by frequent shaking of a slurry with a dilute solution of PBS, (phosphate buffered saline) solution for 30 minutes. Following centrifugation, the aqueous extract is acidified to pH = 3, and the metsulfuron extracted by SPE on polymeric-resin cartridges.[15] After elution with methylene chloride and solvent phase transfer, metsulfuron methyl is determined by high-performance liquid chromatography/tandem mass spectrometry (HPLC/MS/MS). Using this very simple and direct approach to extraction, a minimum quantitation limit of 200 parts-per-trillion is reported.

7 PRESSURIZED FLUID EXTRACTION

7.1 History

PFE is also known as pressurized solvent extraction (PSE), ASE, pressurized accelerated solvent extraction (PASE), and enhanced solvent extraction (ESE). In its rapid adoption of several new methods employing this approach, the USEPA recognizes PFE as its official name for this technique.

7.2 Principle

PFE[16] is carried out similarly to other techniques which utilize solvents well above their temperature and pressure at standard condition in that the solvent and the sample are placed in a closed container and heated. The heat is provided by an external oven and the solvent is pumped into and out of the extraction chamber. In this approach extractions are performed at temperatures of 373 to 453 K. In order to use common water-miscible organic solvents at these temperatures, pressures of 100 to 135 atmospheres are required to maintain the solvents in their liquid state. Under these conditions, analyte recoveries equivalent to those obtained from traditional Soxhlet extraction are achieved using much smaller solvent volumes for much shorter extraction times. Increasing temperature results in increased solubility, diffusion rate, and mass transfer, while viscosity and surface tension decrease. At elevated temperatures, the activation energy associated with desorption is more easily achieved and the kinetics of desorption and solubilization become more favorable.

The sample is enclosed in a high pressure stainless steel cylindrical cell. Following an initial heating period, it is allowed to interact statically with the pressurized solvent for a predetermined time, after which it is purged from the cell with nitrogen gas and collected in a vial. Since the extraction can be fully automated, samples can be run sequentially. Typical sample sizes for soils are in the 10–30 g range, and the total volume of solvent required for extraction and rinsing is in the 20–40 mL range. Extraction times for most pesticides are in the 20-minute range.

7.3 Applications

PFE is the most recently developed of the options analysts have for extracting pesticide residues from soils, with no published applications reports prior to 1995, and no commercial instruments available prior to 1996. Since then, however, a few applications have appeared in the research literature. Fisher et al.[17] used PFE to extract semivolatile organic priority pollutants from a red clay loam soil, both spiked in the laboratory as well as contaminated in the field. The analyte list includes two CHIs *p*,*p*′-DDT (CAS #50-29-3) and endrin (CAS #72-20-8). Spiking levels ranged from 400 to 20 000 ppb. At spike levels between 4000 and 20 000 ppb, recoveries approached 100%. At the 400-ppb level, recoveries were lower owing primarily to matrix interference. Conte et al.[18] report the extraction of diflufenican (CAS #83164-33-4), a pyrydylcarboxamide herbicide, from soil in 10 minutes using acetonitrile at 373 K at 30 atmospheres. Recoveries average 96% for soil spiked at 100–400 ppb. Four samples per hour are easily processed by this technique. While the high-performance liquid chromatography separation clearly shows coextracted material, presumably organic materials from the soil itself, none seemed to interfere with the retention time or shape of the herbicide peak. Pyle and Marcus[19] extracted seven CHIs from a standard reference material (SRM) marine sediment using (1 : 1) hexane–acetone at 373 K and 135 atmospheres in 20 minutes.

The insecticides included both *cis*-(CAS #5103-71-9) and *trans*-chlordane (CAS #5103-74-2), *trans*-nonachlor (CAS #3734-49-4), and the *p*,*p*′-isomers of 1,1-dichloro-2,2-bis(4-chlorophenyl)ethylene (DDE) (CAS #72-55-9), 1,1-dichloro-2,2-bis(4-chlorophenyl)ethane (DDD) (CAS #72-54-8), and DDT (CAS #50-29-3). Recoveries indicated that the combination of PFE together with quantitation by gas chromatography/tandem mass spectrometry (GC/MS/MS) is a complementary technique which is comparable to or better than existing techniques in terms of speed and sensitivity. In another report addressing chlordane extraction from soil, Brumley et al.[20] used (1 : 1) methylene chloride–acetone at both 373 K and 423 K for extraction completed in 20 minutes. Recoveries from soils spiked with between 2000 and 20 ppb ranged between 80 and 120% at 373 K.

PFE approaches have not been limited to insecticidal agents. The fungicide hexaconazole (CAS #79983-71-4) is effectively removed from weathered soils using acetone at 133 atmospheres and 373 K. Frost et al.[21] report that PFE furnished 'cleaner' extracts requiring no further pretreatment prior to chromatographic analysis than did either of two other high pressure systems. The nature of the hexaconazole–soil interaction is very important, although its ability to be recovered depends on extraction conditions and the efficiency of the system. Within the herbicide families of pesticides, triazines and chloroacetamides are two of the most widely used for weed control in corn and soybean. Gan et al.[22] have shown that PFE efficiently removes both atrazine (A) (CAS #1912-24-9) and alachlor (CAS #15972-60-8) from aged and freshly spiked soils. Following selection of a solvent system, a central composite statistical design was used to optimize extraction conditions by evaluating the effect of varying temperature, pressure, and static time on PFE efficiency. Temperature varied from 333 K to 413 K at 20 degree intervals, pressure from 33 to 166 atmospheres at 33 atmosphere intervals, and static time from 5 to 25 minutes at 5 minute intervals. A quadratic model was used to describe the interactions between each variable and the recovery. Following eight weeks of incubation the efficiency of PFE for both A and alachlor was consistently better than either Soxhlet or solvent extractions. As the herbicide on the soil ages, harsher conditions are needed to separate the residue from the matrix.

8 SUPERCRITICAL FLUID EXTRACTION

8.1 History

Experimental observations of a supercritical (SC) phase from which the property of CP for a pure substance was discovered date back to the early 1820s[23] when it was first observed that there is a critical temperature above which a pure substance can only exist as a fluid and not as either a liquid or gas. Appreciation of the solvating properties of pure substances above their CP date to 1879.[24] Early in the 1900s, it was observed that solubilities of nonvolatile organic compounds in carbon dioxide held under SC conditions were several orders of magnitude greater than that which could be explained from vapor pressure considerations alone.[25] Extraction with supercritical fluids (SFs) raised little interest during the first half of this century, and the interest shown focused on industrial process-based applications and not analytical chemistry. A new era for SFE was ushered in with a 1970 patent which described the decaffeination of green coffee beans by immersing the wetted beans in SC carbon dioxide.[26] Since 1980, there has been rapid development of applications of SFE within the food industry. At the same time applications in environmental analysis began to appear, although the initial reception by the analytical community was one of skepticism. Prior to mid-1989, fewer than 30 citations on analytical SFE could be found in *Chemical Abstracts*.[27] In the 1990s, the situation has reversed itself to such an extent that now more than 300 citations referencing SFE applications can be found in *Chemical Abstracts* for 1993 and 1994 alone.

8.2 Principle

In order to understand SFE it is helpful to review the properties of SFs. The physical state of a substance of fixed composition is described by a phase diagram, a pressure–temperature plot showing the relationship between the processes of sublimation, melting, and boiling. Shown in Figure 2 is the phase diagram for carbon dioxide. It defines the regions corresponding to the gaseous, liquid, and solid states. Points along the lines separating the phases define the equilibrium existing between the two phases. The vapor pressure of the substance begins at the triple point (TP) and ends at the CP. The critical region has its origin at the CP. Thus, an SF is any substance which is held above its critical temperature (T_C) and critical pressure (P_C). The critical temperature, therefore, is the highest temperature at which a gas can be converted into a liquid by increasing its pressure. Correspondingly, the critical pressure, therefore, is the highest pressure at which a liquid can be converted into a gas by increasing its temperature. At the CP, the densities of the two phases become identical. Thus, for pure substances held above their CP, there is only one phase and it expresses some of the physicochemical properties intermediate between those of a liquid and those of a gas. Viscosity is low and diffusivity is high, resulting in solvating power approaching polar liquids.

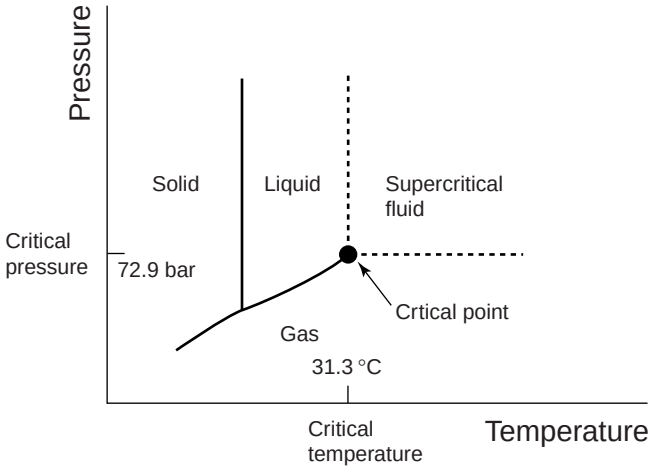

Figure 2 Phase diagram for carbon dioxide.

Using SFs for extraction of pesticides from soils offers several advantages. The solubilizing power, and thus the selectivity, can be adjusted in SFE by varying the density (changing temperature and pressure) of the SF. Often this results in cleaner extracts than liquid solvent extraction. This, in turn, reduces or eliminates the need for cleanup and facilitates quantitation by reducing analysis time while minimizing errors. By far, the most frequently used SF for pesticide extractions is carbon dioxide due to its relatively low T_C and P_C, low toxicity and reactivity, and availability in high purity at low cost. It has proven to be an excellent extractant for nonpolar pesticides, a reasonably good agent for moderately polar pesticides, and a poor extractant for polar pesticides. However, its effectiveness for the very polar analytes can be improved by the addition of polar organic solvents, known as cosolvents, as modifiers.

8.3 Applications

The use of SFE for extraction of many classes of pesticides from soils has received much attention in recent years. Most of this recent work has focused on current use insecticides and herbicides. Koinecke et al.[28] reported that the best efficiency for SFE of the insecticides fenpropimorph (CAS #67564-91-4), pirimicarb (CAS #23103-98-2), ethyl parathion (CAS #56-38-2), triallate (CAS #2303-17-5), and fenvalerate (CAS #51630-58-1) from soil was achieved using carbon dioxide with 5% methanol cosolvent at 330 K at 375 atmospheres pressure. Complete extraction of the insecticide chlorpyrifos (CAS #2921-88-2) and its major metabolite from soil using only subcritical water was reported by Jiminez–Carmona et al.[29]

Herbicides have also been a focus of SFE studies. Steinheimer et al.[30] removed triazine and chloroacetamide herbicides from two Iowa soils spiked at 100–2000 ppb range using cosolvent-modified carbon dioxide. A single extraction carried out at relatively high fluid density in both static and dynamic modes sufficed for all analytes. Principal component analysis was used to correlate the response of each analyte within a matrix of dependent variables influencing extraction efficiency. Polar pesticides, including the carboxylic acid herbicides, dicamba (CAS #1918-00-9) and (2,4-dichlorophenoxy)acetic acid (2,4-D) (CAS #94-7-7), are extracted from sand using in-situ derivatization under SFE conditions, followed by direct extraction of the derivatives. However, this technique is reported to be very sensitive to matrix effects.[31] Sand and a certified reference soil matrix were used to evaluate the extraction efficiency of carbon dioxide for 36 nitroaromatic compounds, 19 halogenated diphenyl ethers, and 19 insecticides.[32] SFE with methanol-modified CO_2 was compared with Soxhlet extraction for twelve pesticides, six CHIs and six OPs on four soil type materials: clay, sand, topsoil, and river sediment.[33] SFE was shown to have the best overall precision. The influence of both temperature and pressure on the extraction efficiencies of two *N*-phenylurea herbicides, diuron (CAS #330-54-1) and linuron (CAS #330-55-2), and a common metabolite from a sandy loam soil has been studied.[34]

SFE has been evaluated for extraction of SU herbicides from agricultural soils. Metsulfuron methyl (CAS #74233-64-6), sulfometuron methyl (CAS #74222-97-2), and nicosulfuron (CAS #111991-09-4) were studied on three Des Moines Lobe soils from the midwestern USA spiked at 400 and 4000 ppb. The extraction procedure was optimized for metsulfuron methyl with SF density, extraction time, and cosolvent addition, as the parameters studied.[35] Good recoveries (75–90%) for both metsulfuron methyl and sulfometuron methyl were obtained; however, only 1–4% of the added nicosulfuron was recovered. The authors concluded that SFE method optimization for specific compounds cannot always be extrapolated to similar compounds within the same pesticide family or to a different matrix, with respect to soil texture and composition, without additional method development and optimization.

The successes of SFs, amended or unamended, as extractants for trace organic analysis of soils and sediments has led some researchers to investigate the solvating power of liquids at elevated temperature and pressure, but well below their respective CPs.[36,37] Known as subcritical fluids, they have recently been applied to some pesticide applications. Crescenzi et al.[38] extracted eighteen herbicides from soil using 25 mL of subcritical water held at 363 K. The flow rate through the stainless steel cartridge containing 3 g of soil admixed with sand was increased from an initial 0.4 mL min^{-1} to 1.0 mL min^{-1}. The analytes consisted of both acidic

and nonacidic herbicides; the nonacidic chemicals largely from the *sym*-triazine and *N*-phenylurea families, and the acidic chemicals largely from the chlorophenoxy acid and benzoic acid families. Recoveries from both sand and silt loam soils spiked at 100 ppb ranged between 60 and 90%. This technique of subcritical water extraction is very similar to ASE with two distinct differences; first, pressures significantly above ambient are not required, and secondly, organic solvent use is either very small or not required at all. In an extraction method comparison with 12-h methanol Soxhlet versus a 2-h sonication in a methanol–water–acetic acid mixture, the subcritical water method yielded equal or greater recoveries than both other methods for both the acidic and the nonacidic herbicide analytes.

9 MICROWAVE ASSISTED EXTRACTION

9.1 History

In the 1980s, Ganzler et al.[39] conducted early experiments using household microwave ovens to aid in extraction conducted at atmospheric pressure. Since then, the process has been further investigated, refined, and patented. As a result, the technique has been successfully applied to extraction of organic compounds, priority pollutants, and volatile organics. That early development has resulted in specialized microwave instruments designed to provide temperature-controlled, closed-system operation.[40] These modern units permit heating of many samples simultaneously in closed, pressurized vessels which allow for heating of polar solvents to temperatures approximately 373 K above their standard conditions boiling points. Inertness of extraction vessel surfaces in contact with the sample is assured through the use of perfluoroalkoxy (PFA) liners.

9.2 Principle

The microwave portion of the electromagnetic spectrum is defined as energy of wavelength between 3 and 300 mm corresponding to a frequency range of 1–100 GHz. Microwave assisted extraction (MAE) is the process of using this energy to heat solid sample–solvent mixtures and results in the partitioning of the compounds of interest from the solid matrix into the solvent. Typically, extraction is done in a sealed vessel under temperature-controlled conditions. This results in extraction temperatures being raised above the atmospheric conditions boiling point for the solvent selected, thereby accelerating the extraction process. The closed-vessel microwave heating systems best suited for pesticide extraction from soil usually are outfitted with a turntable to rotate the sample within the cavity, thus insuring even distribution of thermal energy.

Microwave heating is significantly different from conductive heating methods. Conductive heating is sample independent. All samples placed inside a conduction heating oven will equilibrate to the preselected temperature, a process which can be quite slow. Microwave heating is much more solute-to-solvent ratio dependent. The rate of increase in the temperature depends on the microwave-absorbing characteristics of the sample. The cavity design provides the ability to heat several samples simultaneously and uniformly in a short time with the proper oven reflectile design.

9.3 Applications

MAE of CHIs has been described by Fish and Revesz.[41] Eleven compounds, including both cyclodiene derivatives and DDT-congeners, have been successfully recovered from certified reference soils spiked with each compound. A 1 : 1 acetone–hexane solvent mixture was used for extractions at selected temperatures between 363 K and 423 K with preset hold times of 15–20 minutes. Pressures ranged between 2.7 and 5.5 atmospheres. Results indicated that an extraction temperature of 293 K was the lowest at which recoveries were consistently at or above 95% for the reference soil spiked between 150 and 425 ppb for each CHI. Thermally labile CHIs, such as endrin and *p*,*p*′-DDT, showed no evidence of breakdown during the heating of the MAE process. Similarly, no loss of analyte could be attributed to the use of the PFA liners.

Successful application of MAE to pesticide removal in agricultural soils has not been limited to the hydrophobic and nonpolar end of the pesticide spectrum. A method for extraction of A together with several of its major soil metabolites has been reported by Steinheimer.[42] In this approach, A (CAS #1912-24-9), desethylatrazine (DEA) (CAS #6190-65-4), and desisopropylatrazine (DIA) (CAS #1007-28-9) are extracted simultaneously by subjecting a soil–aqueous phase mixture to rapid microwave heating, initially in water, and subsequently with dilute hydrochloric acid. The extracts are then recovered, combined, and taken to SPE (solid-phase extraction) clean-up and prepared for instrumental determination. Recoveries from two Iowa soils, a Des Moines Lobe glacial till and a Loess Hills silt, spiked at 100–10 000 ppb ranged from 60% for A to 80%–120% for both metabolites. MAE reduced extraction times from many hours to a very few minutes while conserving costly and disposal problematic organic solvents. In another study focusing on triazine herbicides, Molins et al.[43] report that MAE works very well for extraction of these residues from both freshly spiked soils as well as soils with spikes aged for up to 300 days. In general, the

freshly spiked soils yielded about 10% higher recovery than did the aged soils for DEA when using methylene chloride–methanol (90 : 10) as solvent, while A and DIA recoveries were nearly unchanged.

Among the most widely used newer classes of herbicides (post-1985 initial registration) used for both grass and broadleaf control in cash grain rowcrops are the imidazolinones and the SU compounds. Chemically, they are polar hydrophilic weak acids ($pK_a = 4$) with pH-dependent high water solubility. Extraction from soils by conventional methods is laborious and time-consuming. Stout et al.[44] report that imidazolinone herbicides are extracted from agricultural soil by MAE using 0.1 M NH_4 OAc in NH_4 OH as solvent at 398 K for 3 minutes. Average recoveries of 92% with a standard deviation of 13% are reported for soils spiked at the 1–50 ppb level. MAE together with solid-phase cartridge clean-up reduces the time required to prepare a set of twelve samples for instrumental determination to <10 minutes per sample. Better extractabilities are achieved both for surface spikes and for aged field-treated soils. Liquid chromatography/mass spectrometry (LC/MS) provides a 1 ppb limit of quantitation.

MAE is reported to be very effective in addressing the issue of 'bound pesticide residues' and their extractability. Nicollier and Donzel[45] have reported that MAE using 80% methanol in water at temperatures from 363 to 473 K effectively removed more of the ^{14}C residue, both parent and metabolites, than did cold extraction in both soils and biological tissue matrices. The addition of tetramethylammonium hydroxide (TMAH) drastically improved removal from biological samples, presumably as a result of hydrolytic breakdown of polymeric matrices or cleavage of pesticide adducts, releasing residue. Herbicides studied include dimethachlor (CAS #50563-36-5), metobromuron (CAS #3060-89-7), propiconazole (CAS #60207-90-1), cyprodinil (CAS #121552-61-2), and others.

10 EXTRACTION BY SONICATION

10.1 Principle

Principles underlying ultrasonic energy form the basis for the utility of sonication to enhance solvent extraction. Ultrasound is composed of sound waves transmitted at frequencies higher than those detected and discerned by the human ear. Ultrasonic energy generated by a piezoelectric transducer oscillating at a frequency of 20 000–40 000 Hz creates cavitation in liquids. Cavitation is the formation and collapse of countless tiny cavities or vacuum bubbles within a liquid. This energy causes alternating high- and low-pressure waves within the liquid as the waves alternately expand and compress the liquid. During expansion, tiny cavities grow at a microscopic level to a maximum critical diameter and then implode with release of the energy of the sonic wave. Localized vigorous agitation occurs at the soil surface during the contraction–implosion cycles, thus releasing sorbed pesticide from the surface and into the solvent.

Many parameters affect the strength of cavitation, including frequency and amplitude of the ultrasound, temperature, surface tension, vapor pressure, viscosity, and liquid density. Direct sonication uses a specially designed acoustic tool known as the horn or probe. The entire assembly, known as the sonotrode, is constructed of an inert material, such as a titanium alloy. The sonotrode is placed directly into a soil–solvent mixture. Horns and probes are available in different cross-sectional areas and lengths depending upon the solution volume to be processed and the desired intensity. Most of the energy associated with the sonotrode occurs at its tip, and the efficiency of energy transfer to this point is critical to the performance of the probe. The agitation also generates heat, and this heat can speed sonication extractions.

10.2 Applications

USEPA Method SW-846-3350 specifies sonication as one method for the extraction of nonvolatile and semivolatile organic compounds, including pesticides, from soils. The method is based on the expected concentration of organic compounds in the sample. A high-concentration method in which the concentrations of individual organic compounds are greater than 20 parts-per-million is simple, while the low-concentration method for components with concentrations less than 20 parts-per-million uses a larger sample size, and is more complicated. The sample (2 or 30 g, depending on the expected concentration) is mixed with anhydrous granular sodium sulfate to form a free-flowing powder that can be extracted three times or more using sonication. A solvent mixture of 1 : 1 acetone–methylene chloride is recommended for use with a specified horn-type sonotrode equipped with a titanium tip.[46] Sonication is a method of choice for extraction of polycyclic aromatic hydrocarbons (PAHs) and their hetero-atom analogs from sediments and soils but has enjoyed little adoption for pesticides. This is due to the greater popularity of alternatives, specifically, PFE and SFE, and ASE.

11 PESTICIDE BIOGEOCHEMISTRY IN SOIL: FORMATION AND EXTRACTION OF PESTICIDE-DERIVED COMPOUNDS

There are many different types of chemical reactions responsible for the conversion of pesticides to stable

daughter compounds in agricultural soils. Some are catalyzed primarily by consortia of indigenous microorganisms, including bacteria, fungi, and actinomycetes. Others are catalyzed by the surface functionality (such as dissociated silanols) found on clay mineral surfaces or the functionality of organic matter (such as carboxylic or phenolic acidity) present in the soil. In most cases, similar reactions occur for many members within a pesticide class. Within the CHI family of chemicals are both the DDT-congeners and the cyclodiene adducts. For the DDT-congeners, initial degradation in the soil proceeds through either of two breakdown pathways: hydrolytic dechlorination or reductive dechlorination. In the former, chlorine is replaced with hydroxyl, and in the latter, with hydrogen. The same type of pathway is observed for both DDD and DDE. In addition, DDT-congeners may undergo dehydrochlorination as well, as is observed in the conversion of DDT to DDE or methoxychlor (CAS #72-43-5) to methoxychlor olefin. For the cyclodiene adduct insecticides, those produced from Diels-Alder type cycloaddition of various olefins to hexachlorocyclopentadiene, initial soil breakdown usually proceeds through oxidation or dechlorination. In the case of aldrin (CAS #309-00-2) and isodrin (CAS #465-73-6), epoxidation occurs first, followed by dechlorination. In the case of dieldrin (CAS #60-57-1) and endrin (CAS #72-20-8), dechlorination occurs directly. A similar type scenario exists for the OP insecticides. All of the chemicals in this family of pesticides are derivatives of either ortho-phosphoric acid itself or one of its thiono- or thiolo-analogs. This circumstance is a consequence of the isoelectronic relationship between oxygen and sulfur, generally with either alkyl or aryl substitution. Initial degradation of OPs usually occurs via hydrolysis of one or more of these linkages giving rise to carbinol derivatives or their corresponding sulfur analogs. Thus, chlorpyrifos (CAS #2921-88-2) yields 3,5,6-trichloro-2-pyridinol as its major initial soil degradate, and parathion yields *p*-nitrophenol as its major initial soil degradate. Both carbamate herbicides and insecticides undergo hydrolysis in the soil, each yielding their corresponding carbinol or amine.

For the other major classes of herbicides, different types of breakdown reactions occur. *N*-phenylurea herbicides undergo hydrolysis to yield their corresponding aniline derivatives. In a report focused on fluometuron (CAS #2164-17-2), Mueller and Moorman[47] illustrate the degradation biochemistry and extraction protocol for this family. Degradation produces in sequence two *N*-demethyl derivatives in addition to the aniline derivative and all are efficiently extracted from soil by shaking with methanol for 30 minutes on a reciprocating shaker operating at 240 cycles per minute. Recoveries for a silt loam ranged between 30% and 90%, with the most polar metabolite, trifluoromethylaniline (CAS #87-12-7) yielding the poorest recovery. Rutherford[48] has extracted both parent and a soil metabolite of the benzamide herbicide isoxaben (CAS #82558-53-2) by refluxing a 50 g sample with 80 : 20 methanol–water for 1 hour. Clay loam and sandy loam soils spiked at 5–1000 ppb yielded recoveries between 70 and 95%. Both *unsym*- and *sym*-triazine herbicides undergo initial *N*-dealkylation reactions as well as hydrolytic dechlorination reactions. Chloroacetamide herbicides undergo both hydrolytic dechlorination to carbinol derivatives and hydrolysis to their corresponding anilino-derivatives. In addition, this family of herbicides undergoes conjugation reactions with soil components which lead to both sulfonic acid and oxanilic acid conjugates. SU and imidazolinone herbicides undergo hydrolysis reactions. For the SUs, cleavage of the sulfonated urea backbone produces aryl- or azaaryl-derivatives. Imidazolinones undergo a hydrolytic ring cleavage of the parent imidazolinone ring with the formation of carbamoyl carbamate derivatives.

Generally, and across most families of herbicides and insecticides, soil degradation initially produces daughter compounds which are smaller (lower in molecular weight), more asymmetric in structure, and more hydrophylic (owing to the incorporation of oxygen). As a consequence, they are efficiently extracted from soils by methods which can easily accommodate polar and water-miscible organic solvents. Depending upon the ionic nature of the degradates, dilute solutions of mineral acids or bases may enhance extraction efficiency by exploiting ion-exchange properties. Research papers specifically addressing extraction of soil metabolites and degradates mostly date from 1990 and include several of the techniques previously discussed. Exhaustive methods employed for extraction of CHIs from soils also remove structurally similar degradates with equal efficiency. Both DDE, the dehydrohalogenation degradate of DDT, and DDD, the dehydrochlorination degradate of DDT, are removed from soil along with DDT by any of the five highlighted approaches. Generally, if exhaustive extraction is the primary goal, without consideration of sample throughput, time, or solvent consumption and disposal, a variant of the Soxhlet technique remains the method of choice. Generally, for the OP insecticides, the same relationship prevails; techniques suitable for extraction of the parent pesticide also remove the initial soil metabolites and degradates with approximately the same efficiency. Thus, chlorpyrifos and 3,5,6-trichloro-2-pyridinol would be extracted simultaneously by any of the highlighted techniques.

For the major classes of herbicides, consideration of chemical properties of the target degradates makes the choice of extraction protocol somewhat more selective. In their report on SFE in which triazine herbicide

metabolites were extracted from two Iowa soils with methanol-modified CO_2, Steinheimer et al.[30] demonstrated that the process variables which exerted the greatest influence on the recoveries of the two A metabolites were carbon content of the soil and water content of the methanol cosolvent amendment. In the principal component analysis, these two parameters showed highest negative correlation with recovery for both DEA and, especially, DIA. Temperature, pressure, and both static and dynamic time intervals revealed lesser correlations with recovery. DIA recoveries from loam and silt loam soils averaged about 35%, while DEA recoveries from those soils averaged about 80%.

The application of subcritical water extraction to herbicide metabolites has recently been described by Di Corcia et al.[49] for the *sym*-triazine terbuthylazine (CAS #5915-41-3). This herbicide, which is not registered for agricultural use in the United States, is used as an alternative to A in Europe and elsewhere. Metabolites and degradates included all three desalkyl-metabolites: desethyl-, dest-butyl-, and desethyl,dest-butyl-. In addition, the corresponding hydroxyl derivatives of each of the derivatives were also included in the study of extraction responses. Both parent and daughter compounds are extracted together from 3 g of soil with phosphate-buffered water at 373 K. Extractant flow rate was increased from 0.3 mL min^{-1} initially to 1 mL min^{-1}. Analytes are trapped on a graphitized carbon black cartridge and recovered by elution. The subcritical water extraction was compared to two other methods proposed for the extraction of A and its degradation products from agricultural soils: 24-h Soxhlet extraction with methanol, and the room temperature mixed-mode extractant double extraction proposed by Lerch et al.[50] In comparison with the 24-h hot methanol extraction, the mixed-mode procedure gave the better recoveries, in spite of the fact that the soil was extracted only twice at room temperature. In contrast, the subcritical water procedure proved to be more effective in recovering the hydroxytriazine derivatives than was the mixed-mode protocol of Lerch et al.[50]

12 AUTHOR STATEMENT

With the exception of the Soxhlet approach all of these techniques have come into prominence only in the 1990s. Accordingly, many new applications are yet to be discovered and reported. As a consequence, the author views this discussion primarily as an educational tool and not as a reference work. He has, therefore, selected only citations which illustrate principles while providing well-documented examples referencing current literature at the time of publication. Accordingly, it is the author's wish that all readers will benefit from use not only of the references cited herein, but also of the references cited within those citations listed. Many of the listed references cite research reports in which the various extraction options have been evaluated comparatively against each other for a given extraction application. These are very helpful to the reader confronted with choosing one over the others.

12.1 Disclaimer

Mention of specific products, suppliers, or vendors is for identification purposes only and does not constitute an endorsement by the US Department of Agriculture to the exclusion of others.

ACKNOWLEDGMENTS

The author wishes to thank the following individuals for their support in providing resource materials for this article: Ms Shirley Anderson, Foss North America, Edin Prairie, MN; and Dr Greg LeBlanc, CEM Corporation, Matthews, NC. The author also extends his appreciation for their review comments to the following colleagues: Dr David Meek; Dr Thomas Moorman; Dr Richard Pfeiffer; and Mr Kenwood Scoggin.

ABBREVIATIONS AND ACRONYMS

A	Atrazine
ASE	Accelerated Solvent Extraction
CHI	Chlorinated Hydrocarbon Insecticide
CP	Critical Point
DDD	1,1-Dichloro-2,2-bis-(4-chlorophenyl)ethane
DDE	1,1-Dichloro-2,2-bis-(4-chlorophenyl)ethylene
DDT	1,1,1-Trichloro-2,2-bis-(4-chlorophenyl)ethane
DEA	Desethylatrazine
DIA	Desisopropylatrazine
ESE	Enhanced Solvent Extraction
GC/MS/MS	Gas Chromatography/Tandem Mass Spectrometry
HPLC/MS/MS	High-performance Liquid Chromatography/Tandem Mass Spectrometry
LC/MS	Liquid Chromatography/Mass Spectrometry
MAE	Microwave Assisted Extraction
OP	Organophosphate

PAH	Polycyclic Aromatic Hydrocarbon
PASE	Pressurized Accelerated Solvent Extraction
PBS	Phosphate Buffered Saline
PCB	Polychlorinated Biphenyl
PFA	Perfluoroalkoxy
PFE	Pressurized Fluid Extraction
PSE	Pressurized Solvent Extraction
SC	Supercritical
SF	Supercritical Fluid
SFE	Supercritical Fluid Extraction
SRM	Standard Reference Material
SU	Sulfonylurea
TMAH	Tetramethylammonium Hydroxide
TP	Triple Point
USEPA	United States Environmental Protection Agency
2,4-D	(2,4-Dichlorophenoxy)acetic Acid

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Soil Sampling for the Characterization of Hazardous Waste Sites • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicides (New Generation): Imidazolinones, Aryloxyphenoxypropionic Acids/esters, and Diphenylethers, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

FURTHER READING

W.H. Ahrens (ed.), *Herbicide Handbook*, 7th edition, Weed Science Society of America, 1994.
H. Aizawa (ed.), *Metabolic Maps of Pesticides*, Academic Press, New York, 1982.
M.R. Carter (ed.), *Soil Sampling and Methods of Analysis*, Lewis Publishers, Boca Raton, FL, 1993.
M. Caude, D. Thiébaut (eds.), *Practical Supercritical Fluid Chromatography and Extraction*, Harwood Academic Publishers, Amsterdam, The Netherlands, 1999.
H.H. Cheng (ed.), *Pesticides in the Soil Environment: Processes, Impacts, and Modeling*, Soil Science Society of America Series, Number 2, Soil Science Society of America, Madison, WI, 1990.
R.O. Gilbert, *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York, 1987.
A.J. Handley (ed.), *Extraction Methods in Organic Analysis*, CRC Press, Boca Raton, FL, 1999.
A. Klute (ed.), *Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods*, 2nd edition, Soil Science Society of America, Madison, WI, 1986.
R.T. Meister (ed.), *Farm Chemicals Handbook*, Meister Publishing Co., Willoughby, OH, 1999.
R.G.D. Steel, J.H. Torrie, *Principles and Procedures of Statistics: A Biometrical Approach*, 2nd edition, McGraw-Hill, New York, 1980.
M.E. Sumner (editor-in-chief), *Handbook of Soil Science*, CRC Press, Boca Raton, FL, 1999.
C.D.S. Tomlin (ed.), *Pesticide Manual*, 11th edition, British Crop Protection Council, Surrey, 1997.
USDA, 'Pesticide Properties Database', USDA-ARS, Remote Sensing and Modelling Laboratory, Beltsville, MD. Website: http://www.ars.usda.gov/Services/docs.htm?docid≡6433.
B. Wenclawiak (ed.), *Analysis with Supercritical Fluid: Extraction and Chromatography*, Springer-Verlag, New York, 1992.
S.A. Westwood (ed.), *Supercritical Fluid Extraction and Its Use in Chromatographic Sample Preparation*, CRC Press, Boca Raton, FL, 1992.

REFERENCES

1. W.G. Cochran, *Sampling Techniques*, 3rd edition, John Wiley and Sons, Inc., New York, 1977.
2. R.G. Petersen, L.D. Calvin, 'Sampling', in *Methods of Soil Analysis, Part 3 – Chemical Methods*, ed. D.L. Sparks, SSSA Book Series: 5, Soil Science Society of America, Madison, WI, 1996.

3. R.G. Congalton, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, Boca Raton, FL, 1999.
4. W.A.V. Clark, P.L. Hosking, *Statistical Methods for Geographers*, John Wiley and Sons, Inc., 1986.
5. A.M. Hay, 'Sampling Designs to Test Land-use Map Accuracy', *Photogrammetric Engineering and Remote Sensing*, **45**, 529–533 (1979).
6. B.J.L. Berry, A.M. Baker, 'Spatial Analysis; A Reader in Statistical Geography', in *Geographic Sampling*, eds. B.J.L. Berry, D.F. Marbles, Prentice-Hall, Englewood Cliffs, NJ, 1968.
7. G. Uehara, B.B. Trangmar, R.S. Yost, 'Spatial Variability of Soil Properties', in *Soil Spatial Variability*, eds. D.R. Nielsen, J. Bouma, Proc. Workshop ISSS and SSSA, Las Vegas, NV, 30 Nov–1 Dec, 61–92, 1984. PUDOC, Wageningen, The Netherlands.
8. R.J. Lascano, J.L. Hatfield, 'Spatial Variability of Evaporation along Two Transects of a Bare Soil', *Soil. Sci. Soc. Am. J.*, **56**(2), 341–346 (1992).
9. P.J. Curran, H.D. Williamson, 'Sample Size for Ground and Remotely Sensed Data', *Remote Sensing of Environment*, **20**(1), 31–41 (1986).
10. R.M. Hirsch, D.R. Helsel, T.A. Cohn, E.J. Gilroy, 'Statistical Treatment of Hydrologic Data', in *Handbook of Hydrology*, editor-in-chief D.R. Maidment, McGraw-Hill, New York, Part 3, Chapter 17, 1993.
11. E.L. Randall, 'Improved Method for Fat and Oil Analysis by a New Process of Extraction', *J. AOAC.*, **57**(5), 1165–1198 (1974).
12. V. Lopez-Avila, 'Development of a Soxtec Extraction Procedure for Extracting Organic Compounds from Soils and Sediments', EPA 600/X-91/140 (US Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada, October, 1991).
13. *Test Methods for Evaluation of Solid Waste*, Physical/Chemical Methods, SW-846, US Environmental Protection Agency, Office of Solid Waste, Washington, DC, December, 1996.
14. R.-K. Smith (ed.), *Handbook of Environmental Analysis*, 3rd edition, Genium Publishing Corporation, Schenectady, NY, 1997.
15. R. Bossi, P. Seiden, S.M. Andersen, C.S. Jacobsen, J.C. Streibig, 'Analysis of Metsulfuron-methyl in Soil by Liquid Chromatography/Tandem Mass Spectrometry. Application to a Field Dissipation Study', *J. Agric. Food Chem.*, **47**(10), 4462–4468 (1999).
16. B.E. Richter, B.A. Jones, J.L. Ezzell, N.A. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: A Technique for Sample Preparation', *Anal. Chem.*, **68**, 1033–1039 (1996).
17. J.A. Fisher, M.J. Scarlett, A.D. Stott, 'Accelerated Solvent Extraction: An Evaluation for Screening of Soils for Selected USEPA Semivolatile Organic Priority Pollutants', *Environ. Sci. Technol.*, **31**(4), 1120–1127 (1997).
18. E. Conte, R. Milani, G. Morali, F. Abballe, 'Comparison Between Accelerated Solvent Extraction and Traditional Extraction Methods for the Analysis of the Herbicide Diflufenican in Soil', *J. Chromatogr., A*, **765**, 121–125 (1997).
19. S.M. Pyle, A.B. Marcus, 'Rapid and Sensitive Determination of Pesticides in Environmental Samples by Accelerated Solvent Extraction and Tandem Mass Spectrometry', *J. Mass Spectrom.*, **32**, 897–898 (1997).
20. W.C. Brumley, E. Latorre, V. Kelliher, 'Determination of Chlordane in Soil by LC/GC/ECD and LC/GC/EC NIMS with Comparison of ASE, SFE, and Soxhlet Extraction', *J. Liq. Chrom. & Rel. Technol.*, **21**(8), 1199–1216 (1998).
21. S.P. Frost, J.R. Dean, K.P. Evans, K. Harradine, C. Cary, M.H.I. Comber, 'Extraction of Hexaconazole from Weathered Soils: a Comparison Between Soxhlet Extraction, Microwave-assisted Extraction, Supercritical Fluid Extraction and Accelerated Solvent Extraction', *Analyst*, **122**, 895–898 (1997).
22. J. Gan, S.K. Papiernik, W.C. Koskinen, S.R. Yates, 'Evaluation of Accelerated Solvent Extraction (ASE) for Analysis of Pesticide Residues in Soil', *Environ. Sci. Technol.*, **33**, 3249–3253 (1999).
23. C. Cagniard de la Tour, 'Exposé de quelques Résultats Obtenus par l'Action Combinée de la Chaleur et de la Compression sur Certains Liquides, Tels que l'Eau, l'Acool, l'Éther Sulfurique et l'Essence de Pétrole Rectifée', *Ann. Chim. Phys.*, **21**(2), 127 (1822).
24. J.B. Hannay, J. Hogarth, 'On the Solubility of Solids in Gases (Preliminary Notice)', *Proc. Roy. Soc. (London)*, **29**, 324 (1879); 'On the Solubility of Solids in Gases', *ibid*, **30**, 178 (1880).
25. E.H. Büchner, *Z. Phys. Chem.*, **54**, 665 (1906).
26. K. Zosel, *Decaffeinizing coffee*, French Patent 2 079 261, 1971.
27. S.B. Hawthorne, 'Analytical-scale Supercritical Fluid Extraction', *Anal. Chem.*, **62**, 633A–642A (1990).
28. A. Koinecke, R. Kreuzig, M. Bahadir, 'Effects of Modifiers, Adsorbents and Eluents in Supercritical Fluid Extraction of Selected Pesticides in Soil', *J. Chromatogr., A.*, **786**(1), 155–161 (1997).
29. M.M. Jiminez-Carmona, J.J. Manclus, A. Montoya, M.D. Luque de Castro, 'Sub- and Supercritical Fluid Extraction of Trichloropyridinol from Soil Prior to Immunoassay', *J. Chromatogr., A.*, **785**(1–2), 329–336 (1997).
30. T.R. Steinheimer, R.L. Pfeiffer, K.D. Scoggin, 'Extraction of Atrazine, Cyanazine, Desethylatrazine, Desisopropylatrazine, and Metolachlor from Fortified Western Cornbelt Soils by SFE with CO_2', *Anal. Chem.*, **66**(5), 645–650 (1994).
31. S.B. Hawthorne, D.J. Miller, J.J. Langenfeld, in *Supercritical Fluid Technology*, eds. F.V. Bright, M.P. McNally, ACS Symposium Series 488, American Chemical Society, Washington, DC, Chapter 13, 1992.
32. V. Lopez-Avila, W.F. Beckert, in *Supercritical Fluid Technology*, eds. F.V. Bright, M.P. McNally, ACS

Symposium Series 488, American Chemical Society, Washington, DC, Chapter 14, 1992.
33. J.L. Snyder, R.L. Grob, M.P. McNally, T.M. Oostdyk, 'Comparison of Supercritical Fluid Extraction with Classical Sonication and Soxhlet Extractions for Selected Pesticides', *Anal. Chem.*, **64**, 1940–1946 (1992).
34. M.P. McNally, J.R. Wheeler, 'Increasing Extraction Efficiency in Supercritical Fluid Extraction from Complex Matrices – Predicting Extraction Efficiency of Diuron and Linuron in Supercritical Fluid Extraction using Supercritical Fluid Chromatographic Retention', *J. Chromatogr.*, **447**, 53–63 (1988).
35. T. Berglöf, W.C. Koskinen, H. Kylin, 'Supercritical Fluid Extraction of Metsulfuron Methyl, Sulfometuron Methyl, and Nicosulfuron from Soils', in *Supercritical Fluid Extraction (SFE) in Environmental Analysis*, Tomas Berglöf, Swedish University of Agricultural Sciences, Doctoral Thesis, 1997.
36. S.B. Hawthorne, Y. Yang, D.J. Miller, 'Extraction of Organic Pollutants from Environmental Solids with Sub- and Supercritical Water', *Anal. Chem.*, **66**(18), 2912–2920 (1994).
37. Y. Yang, S.B. Hawthorne, D.J. Miller, 'Class-selective Extraction of Polar, Moderately Polar, and Nonpolar Organics from Hydrocarbon Wastes using Subcritical Water', *Environ. Sci. Technol.*, **31**(2), 430–437 (1997).
38. C. Crescenzi, G. D'Ascenzo, A. DiCorcia, M. Nazzari, S. Marchese, R. Samperi, 'Multiresidue Herbicide Analysis in Soil: Subcritical Water Extraction with an On-line Sorbent Trap', *Anal. Chem.*, **71**(11), 2157–2163 (1999).
39. K. Ganzler, A. Salgo, K. Valko, 'Microwave Extraction, a Novel Sample Preparation Method for Chromatography', *J. Chromatogr.*, **371**, 299–306 (1986).
40. H.M. Kingston, L.B. Jassie, *Introduction to Microwave Sample Preparation, Theory and Practice*, American Chemical Society, Washington, DC, 1988.
41. J.R. Fish, R. Revesz, 'Microwave Solvent Extraction of Chlorinated Pesticides from Soil', *LC-GC*, **14**(3), 230–234 (1996).
42. T.R. Steinheimer, 'HPLC Determination of Atrazine and Principal Degradates in Agricultural Soils and Associated Surface and Groundwater', *J. Agric. Food Chem.*, **41**(4), 588–595 (1993).
43. C. Molins, E.A. Hogendoorn, H.A.G. Heusinkveld, D.C. Van Harten, P. Van Zoonen, R.A. Baumann, 'Microwave Assisted Solvent Extraction (MASE) for the Efficient Determination of Triazines in Soil Samples with Aged Residue', *Chromatographia*, **43**(9/10), 527–532 (1996).
44. S.J. Stout, A.R. daCunha, D.G. Allardice, 'Microwave-assisted Extraction Coupled with Gas Chromatography/Electron Capture Negative Chemical Ionization Mass Spectrometry for the Simplified Determination of Imidazolinone Herbicides in Soil at the ppb Level', *Anal. Chem.*, **68**(4), 653–658 (1996).
45. G. Nicollier, B. Donzel, 'Release of Bound ^{14}C-Residues from Plants and Soil using Microwave Extraction', Ciba Crop Protection, CIBA-GEIGY Ltd., Basle, Switzerland.
46. 'US EPA SW-846, Update III, Test Methods for Evaluating Solid Wastes, Method 3545', Fed. Reg., 62(114), 32,451, 13 June 1997.
47. T.C. Mueller, T.B. Moorman, 'Liquid Chromatographic Determination of Fluometuron and Metabolites in Soil', *J. AOAC.*, **74**(4), 671–673 (1991).
48. B.S. Rutherford, 'Liquid Chromatographic Determination of the Herbicide Isoxaben and Its Soil Metabolite in Soil and Soil-turf Samples', *J. AOAC.*, **73**(2), 287–289 (1990).
49. A. Di Corcia, A.B. Caracciolo, C. Crescenzi, G. Giuliano, S. Murtas, R. Samperi, 'Subcritical Water Extraction Followed by Liquid Chromatography Mass Spectrometry for Determining Terbuthylazine and Its Metabolites in Aged and Incubated Soils', *Environ. Sci. Technol.*, **33**(18), 3271–3277 (1999).
50. R.N. Lerch, E.M. Thurman, E.L. Kruger, 'Mixed-mode Sorption of Hydroxylated Atrazine Degradation Products to Soil: A Mechanism for Bound Residue', *Environ. Sci. Technol.*, **31**(5), 1539–1546 (1997).

s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Carlo Crescenzi
Università ''La Sapienza'', Rome, Italy

1 Introduction **1**
1.1 General Analytical Considerations 3
2 Sample Preparation **6**
2.1 Liquid Sample 6
2.2 Solid Sample 10
3 Identification and Quantitation **11**
3.1 Gas Chromatography with Conventional Detector 12
3.2 Gas Chromatography/Mass Spectrometry 12
3.3 Liquid Chromatography 13
3.4 Liquid Chromatography/Mass Spectrometry 15
3.5 Immunochemical Assay 17
3.6 Electrochemical Biosensor 17
4 Recent Applications **19**
Abbreviations and Acronyms **20**
Related Articles **20**
References **21**

Triazine herbicides and some of their transformation products (TPs) are included in both the European Communities and United States Environmental Protection Agency (USEPA) priority list of pesticides. Owing to their widespread use and toxicity profile this family of herbicides is one of the classes of chemical pollutants that are more intensively monitored by water authorities. Many analytical methods for their determination have been published. A critical overview of these methods is presented in this paper. Conventional extraction techniques are generally nonselective and can lead to analytical interferences caused by coextracted components present in the environmental matrices. Parameters affecting extraction and problems in isolating s*-triazines, from water, soil, urine and food are discussed. The solid phase is particularly attractive because it may be coupled on-line with chromatographic systems. Many extraction systems, in sorbent cartridge or disk format, have been developed in the last few years. Recently, interest has been growing in employing new highly selective materials such as molecularly imprinted polymers (MIPs) and immunosorbents to extract triazines from complex matrices. Strong cationic exchangers allow an effective clean-up and are useful to improve the sensitivity of a method. Solid-phase microextraction (SPME) is a new technique that can also be used for extracting analytes from water. This technique was successfully automated and coupled with liquid chromatography (LC). For solid matrices Soxhlet extraction is still widely used. Blending the sample in the presence of the solvent in high-speed homogenizer machines or in an ultrasonic bath ensures complete sample disruption and a better extraction of pesticides. Microwave-assisted solvent extraction, supercritical fluid extraction (SFE) and accelerated solvent extraction (ASE) are alternatives for conventional extraction procedures.*

Both, gas chromatography (GC) and LC were successfully coupled with sensitive and selective detectors, but because of the legal implications, the analytical methods need sufficient specificity for testifying contamination. It has been recommended that analytical results indicating the presence of relevant concentrations should be confirmed by mass spectrometric methods. The use of chromatography with mass spectrometric detector and the most common interfaces (thermospray (TSP), electrospray (ES) and atmospheric pressure chemical ionization (APCI)) are illustrated. Rapid screening based on immunoassay and biosensor technologies are described. Finally some examples of recent applications are summarized.

1 INTRODUCTION

The herbicidal activity of *s*-triazine was discovered in 1952 by a Geigy Co. research group in Bern (Switzerland).[1] The activity of this class of herbicides is based on the inhibition of the Hill reaction. The Hill reaction is one of a cascade of biochemical reactions which constitute the process of photosynthesis in chloroplasts. Some plants (corn) can degrade the Cl-triazine to an inactive compound by a reaction with a natural compound present in the plant (2,4-dihydroxy-7-methoxy-1,4-benzoxazine); some other plants (cotton) can store it in the lysigenous gland.[2] *s*-Triazines are predominantly applied in agriculture as selective pre- and post-emergence weed control of corn, wheat, barley, sorghum and sugar cane but are also widely used for nonagricultural uses (railways and roadside verges). The estimated worldwide annual consumption of atrazine[3] in 1980 amounted to 24×10^6 kg and the annual consumption of *s*-triazine herbicides[4] in USA (1992) exceeded 48×10^6 kg. Since it first entered the market in 1957, atrazine has been a very popular herbicide and in the last two decades has

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

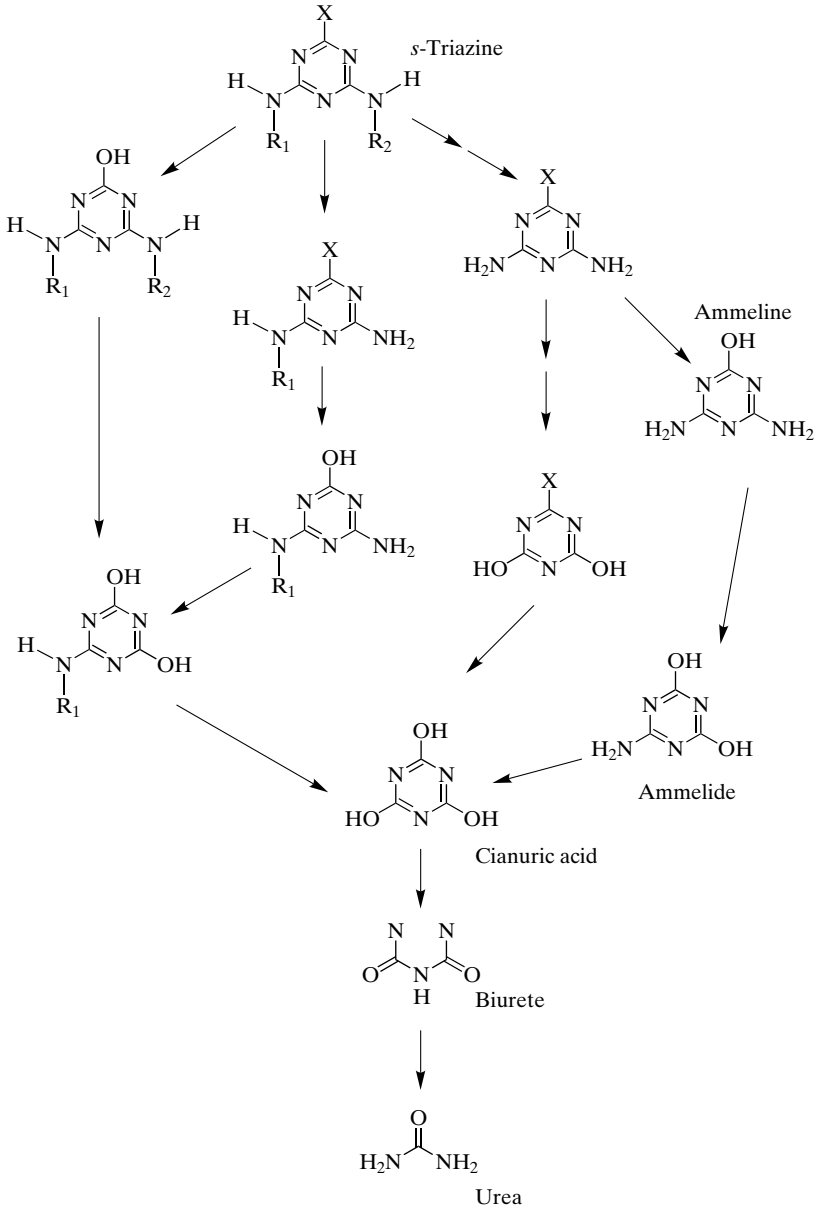

Scheme 1 *s*-Triazine herbicide structures and degradation pathways.

become the most frequently detected pesticide in surface and groundwater.[5] Because of the occurrence in the environment Italy banned atrazine in 1990 and Germany followed in 1991. In the past few years common application rates for atrazine on crops have been 2–5 kg ha^{-1} of active ingredient, in 1992 maximum application rates were reduced to 0.25–0.6 kg ha^{-1}.[6] Since then, maximum label-recommended application rates have been reduced further.

Triazine herbicides are rather soluble and relatively stable in water: it has been estimated that between 0.3 and 1.9% of the applied amount of atrazine on farmland has found its way into stream water.[7] Their metabolites (Scheme 1) are frequently detected in groundwater. Differences in chemical structures are reflected in the physicochemical properties. *s*-Triazine characteristics are primarily determined by the substituent in position 2 (R_1 in Scheme 1). Their basicity increases in the order Cl < SCH_3 < OCH_3 (Table 1); pK_a ranges between 1.65 (simazine) and 4.46 (terbuton). The groups in positions 4 and 6 have smaller but pronounced effects on the basicity: the greater the number of alkyl groups, the more basic are the compounds.

Triazine herbicides and some of their TPs are included in both the European Communities and USEPA priority list of pesticides.[8] In the USA the new Food Quality

Table 1 Most common *s*-triazine herbicides

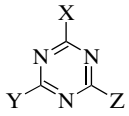

Substituent group			Common name	Chemical name
X	Y	Z		
CH_3S	$NHCH_2CH_3$	$NHCH(CH_3)_2$	Ametryne	2-Methylthio-4-ethylamino-6-isopropylamino-1,3,5-triazine
CH_3O	$NHCH_2CH_3$	$NHCH(CH_3)_2$	Atratone	2-Methoxy-4-ethylamino-6-isopropylamino-1,3,5-triazine
Cl	$NHCH_2CH_3$	$NHCH(CH_3)_2$	Atrazine	2-Chloro-4-ethylamino-6-isopropylamino-1,3,5-triazine
Cl	$NHCH_2CH_3$	$NHC{\cdot}CN{\cdot}(CH_3)_2$	Cyanazine	2-Chloro-4-ethylamino-6-(l-cyano-1-methylethylamino)-1,3,5-triazine
CH_3O	$NHCH(CH_3)_2$	$NHCH(CH_3)_2$	Prometone	2-Methoxy-4,6-bis(isopropyl-amino)-1,3-5-triazine
CH_3S	$NHCH(CH_3)_2$	$NHCH(CH_3)_2$	Prometryne	2-Methylthio-4,6-bis(isopropyl-amino)-1,3-5-triazine
Cl	$NHCH(CH_3)_2$	$NHCH(CH_3)_2$	Propazine	2-Chloro-4,6-bis(isopropylamino)-1,3,5-triazine
Cl	$NHCH_2CH_3$	$NHCH_2CH_3$	Simazine	2-Chloro-4,6-bis(ethylamino)-1,3,5-triazine
CH_3O	$NHCH_2CH_3$	$NHCH_2CH_3$	Simetone	2-Methoxy-4,6-bis(ethylamino)-1,3,5-triazine
CH_3S	$NHCH(CH_3)_2$	$NHCH(CH_3)_2$	Simetryne	2-Methylthio-4,6-bis(ethylamino)-1,3,5-triazine
CH_3S	$NHCH(CH_3)_2$	$NHCH(CH_3)_2$	Terbutryne	2-Methylthio-4-ethylamino-6-terbutylamino-1,3,5-triazine
Cl	$NHCH(CH_3)_2$	$NHC(CH_3)_3$	Terbutylazine	2-Chloro-4-ethylamino-6-terbutylamino-1,3,5-traizine

Protection Act took water into account to determine total human exposure to pesticides.[9] In order to evaluate the risk of cancer to general population by long-term exposure, in 1994 the USEPA placed the triazine herbicides into a ''special review''.[10]

Although in 1995 DuPont announced the voluntary phaseout of cyanazine (completed in 1999), a decline of the use of this family of herbicides is not predictable.

1.1 General Analytical Considerations

In Europe the maximum allowable concentration (MAC) of pesticides and ''related compounds'' in drinking water is set by the European Union Commission directive 80/778/ECC.[11,12] The directive states that environmental contaminants are not acceptable in water intended for human consumption and set MAC at $0.1\,\mu g\,L^{-1}$ per substance and at $0.5\,\mu g\,L^{-1}$ for the sum of compounds. In order to determine contaminants at the $0.1\,\mu g\,L^{-1}$ level analytical methods having a limit of detection of $0.02\,\mu g\,L^{-1}$ are needed.[8]

Trace analysis of polar compounds such as *s*-triazines in complex matrices poses serious problems:

1. When determining *s*-triazines and their TPs, poor recoveries from water by using either liquid–liquid extraction (LLE) or solid-phase extraction (SPE) with C18 bonded silica cartridges can be observed.[13–15]
2. Conventional extraction techniques are generally nonselective and can lead to analytical interferences caused by coextracted components present in the environmental matrices.
3. The analytical method needs sufficient specificity for testifying contamination. Because of the legal implications it has been recommended that analytical results indicating the presence of relevant concentrations be confirmed by mass spectrometric methods.[8]
4. A large-scale study will generate hundreds, if not thousands, of samples.

Table 2 Summary of analytical procedures for determining *s*-triazine herbicides in different matrices by various techniques

Matrices	Compound[a]	Sample treatment	Analysis	Ref.
Rain	Cl- RO- M.	–	ELISA	118
Soil	Cl- M.	–	ELISA	140
Soil, water	Cl-	–	ELISA	120
Water	Cl- RS- RO- M.	SPE (C18)	ELISA	119
Water	Cl-	–	ELISA	121
Water	Cl-	–	ELISA	122
Water	Cl-	–	ELISA	124
Water	Cl-	–	ELISA	126
Water, liquid food	Cl-	–	ELISA	125
Water	Cl-	SPE	ELISA/GC/MS	123
Water	Cl- RO- M.	SPE (C18)	ELISA/GC/MS	16
Water	Cl-	SPE (PS-DVB/C18)	GC/NPD	29
Crop	Cl-	Derivative (HFB)	GC/ECD	104
Water	Cl-	SPE (C18)	GC/ECD/LC/UV	99
Soil	Cl-	Soxhlet	GC/ECD/GC/CD	75
Water, soil, crop	Cl-	LLE, Shake extraction	GC/ECD/GC/CCD	79
–	Cl- RS- RO-	–	GC/FID	96
–	–	–	GC/FID	97
–	Cl- RS- RO-	–	GC/FID	98
–	Cl- M.	Derivative (MSTFA)	GC/FID	105
–	Cl- RS- RO- M.	–	GC/FID	106
Urine	Cl- RS- RO-	LLE	GC/HECD	141
Water	Cl-	SPE (C18)	GC/HRMS	110
–	Cl- RS- RO	–	GC/MS (EI, CI)	108
–	M.	TLC derivative (MBTFA)	GC/MS	142
Soil	–	MAE/Soxhlet/Ultrasonic/SFE	GC/MS	143
Soil	Cl- M.	Shake extraction	GC/MS	144
Soils	Cl- M.	Shake extraction – On-line SPE (C18)	GC/MS	145
Urine	Cl- M.	SPE (SAX-silica)	GC/MS	139
Water	Cl-	SPE disks (C18/PS-DVB)	GC/MS	25
Water	Cl- RS-	SPME	GC/MS	66
Water	Cl-	SPME	GC/MS	67
Water	Cl- RS- RO- M.	SPE (C18)	GC/MS	131
Water	Cl- M.	SPE (GCB)	GC/MS	132
Water	Cl- RO- M.	SPE (GCB)	GC/MS	146
Water, soil	Cl-	LLE, Shake extraction – Ultrasonic	GC/MS	109
Water	Cl- M.	SPE (C18)	GC/MS/LC/UV (DAD)	133
Soil	Cl- M.	MAE	GC/NPD	81
Soil	Cl-	ASE	GC/NPD	147
Water	Cl- RS- RO-	LLE	GC/NPD	19
Water	Cl- RS-	LLE	GC/NPD	21
Water	Cl-	LLE	GC/NPD	22
Water	Cl-	SPE (C18)	GC/NPD	28
Water	Cl-	SPE PS-DVB/C18	GC/NPD/GC/MS	30
Water	Cl-	SPME	GC/NPD	62
Water	Cl- RS- RO-	SPME	GC/NPD	65
Water	Cl-	LLE/SFE (freeze dried water)	GC/NPD/GC/MS	88
Water	Cl- RS-	SPE Separon SE50/50 and ground silica 1 : 2	GC/NPD/LC/UV	103
Water	Cl-	LLE	GC/NPD/GC/MS/LV/UV	135
Water, sediment-bound	Cl-	SPE (C18) Shake extraction	GC/NPD/GC/ITD	134
Crop	Cl- RS- RO-	LLE	GC/NPD/GC/CCD	107
Water	Cl- RS- M.	CLLE/SPE (GCB)	GC/NPD/LC/MS (APCI)	24
Soil	Cl- RS- RO-	Soxhlet	GC/NPD/LC/UV	73
Water	Cl- RS-	SDE	GC/PID	100
Soil	Cl- M.	Soxhlet	LC/MS (TSP)	76

Table 2 (*continued*)

Matrices	Compound[a]	Sample treatment	Analysis	Ref.
Soil	Cl- M.	SPE (Immunosorbent)	LC/MS (TSP)	87
Soil	Cl-	SPE (GCB)	LC/MS	94
Soil	Cl- M.	Subcritical water extraction/SPE (GCB)	LC/MS	95
Water	Cl- M.	SPE (GCB)	LC/MS (FAB)	13
Water	Cl- RS- M.	LLE	LC/MS (ESI/TSP)	23
Water	Cl- M.	SPE (C18)	LC/MS (ESI)	27
Water	Cl- RS- M.	SPE (C18)	LC/MS (TSP)	32
Water	Cl- M.	SPE (GCB)	LC/MS (ESI)	115
Water	Cl- M.	SPE (GCB)	LC/MS (ESI)	148
Water, sediment	Cl- M.	SPE (immunosorbent)	LC/MS (APCI)	38
Water	Cl- M.	SPE (GCB)	LC/MS (ESI)	149
Water	Cl-	LLE/SPE	LC/MS (TSP)	113
Water	Cl-	On-line (single short column) SPE (C18)	LC/MS/MS (APCI)	117
–	Cl- RS- RO-	–	LC/UV	150
–	Cl- RS- RO-	–	LC/UV	151
Humic acid, fruit, urine	Cl-	On-line SPE (MIP)	LC/UV	71
Milk	Cl- RS- RO-	SPE (GCB/SAX)	LC/UV	57
Plant, soil	Cl- M.	SFE/Shake extraction	LC/UV	86
Soil	Cl- RS- RO-	SPE (GCB/SCX)	LC/UV	56
Soil	Cl- M.	Ultrasonic	LC/UV	78
Soil	Cl- M.	Shake extraction	LC/UV	80
Soil	Cl- M.	SFE	LC/UV	82
Soil	Cl- M.	SFE	LC/UV	85
Soil	Cl- M.	SPE (CN, SCX, C18)	LC/UV	152
Soil	Cl- M.	MAE	LC/UV	153
Water	M.	LLE/CLLE/SPE (C18, GCB)	LC/UV	14
Water	Cl- M.	LLE	LC/UV	20
Water	Cl- RS-	SPE (C18)	LC/UV	31
Water	Cl-	SPE (GCB)	LC/UV	36
Water	Cl-	SPE (GCB)	LC/UV	37
Water	Cl- RS- RO- M.	On-line SPE (PS-DVB)	LC/UV (DAD)	40
Water	Cl- RS- RO-	Off-line/on-line SPE (immunosorbent)	LC/UV (DAD)	41
Water	Cl- RS- RO-	Off-line/on-line SPE (immunosorbent)	LC/UV (DAD)	42
Water	RO-	SLM	LC/UV	52
Water	RS-	SLM	LC/UV	53
Water	Cl-	SLM	LC/UV	54
Water	Cl-	On-line SPE/PS-DVB/C18	LC/UV	68
Water	Cl- M.	On-line SPE (PS-DVB)	LC/UV	69
Water	Cl-	On-line SPE/PS-DVB	LC/UV	70
Water	Cl-	SPE (C18/GCB)	LC/UV	154
Water	Cl- M.	SPE (PS-DVB/GCB)	LC/UV	155
Water	Cl- M.	SPE disks (C18 + SAX)	LC/UV (DAD)	156
Water, vegetables	Cl- RS- RO-	SPE (GCB-SCX)	LC/UV	55
Water	Cl- RS- M.	SPE (GCB)	LC/UV (DAD)	35
Water	Cl- M.	On-line SPE (C18)	LC/UV (DAD)	157
Water	Cl- M.	SPE disks (C18)	LC/UV/LC/MS (TSP)	26
Beef liver	Cl-	SPE (MIP)	LC/UV/ELISA	46
Soil	Cl-	Shake extraction	LC/UV/GC/NPD	138
Water	Cl-	–	RIFS	158
Water, milk	Cl-	–	Sensor	159

[a] M, transformation product (also referred to as metabolite); X, substituent (see Scheme 1).

ELISA, enzyme-linked immunosorbent assay; GC/MS, gas chromatography/mass spectrometry; PS-DVB, polystyrene-divinylbenzene; GC/NPD, gas chromatography/nitrogen–phosphorus detection; HFB, heptafluorobutyryl; ECD, electron capture detection; LC/UV, liquid chromatography/ultraviolet; FID, flame ionization detection; HRMS, high-resolution mass spectrometry; MAE, microwave-assisted extraction; GCB, graphitized carbon black; DAD, diode array detection; LC/MS, liquid chromatography/mass spectrometry; PID, photoionization detection; MS, mass spectrometry; SLM, supported liquid membrane.

Accurate analysis of all the samples would be prohibitive, so a low-cost, easily accomplished screening method should be used to select the sample for analysis with the confirmation method. This approach was tested in a large-scale study on herbicides water contamination and their metabolite transport in the ''Corn Belt''.[16] Due to their widespread use and toxicity profile this family of herbicides is one of the classes of chemical pollutants that are intensively monitored by water authorities. Many analytical methods for their determination have been published. An overview of these methods is presented in Table 2.

2 SAMPLE PREPARATION

2.1 Liquid Sample

2.1.1 Liquid–Liquid Extraction

Solvent extraction is usually carried out in a separation funnel, which is vigorously shaken to increase the contact area between the two liquids. The pH of the water sample is usually adjusted to 7 by adding phosphate buffer and, if necessary, H_2SO_4 or NaOH. The extraction efficiency is improved by adding salt (usually NaCl), which has the effect of weakening water–analyte interactions. In extracting *s*-triazines the preferred solvents are dichloromethane,[3,17] ethyl acetate[17] and a mixture of dichloromethane and ethyl acetate and ammonium formate.[18] After the extraction and phase separation, the organic phase is often dried with anhydrous Na_2SO_4. The volume of the extract is reduced using a Kuderna–Danish evaporator in a water bath at 65–70 °C. Often, solvent substitution is needed to make the final extract compatible with GC analysis.[19–24] The drawbacks of LLE are that it is labor-intensive, time-consuming and that it requires the use of relatively large volumes of pesticide-grade solvents, which are expensive, flammable and toxic. Even by using pesticide-grade solvents, concentration by a factor of 1000 or more can introduce analyte interferences by residual solvent impurities. Before use, NaCl and Na_2SO_4 should be heated at 400–500 °C for hours to eliminate potential interferences. Vigorous shaking of solvents and water, especially surface water, may create serious problems of emulsions because of the presence of natural or synthetic surfactants in the sample. Emulsions can be eliminated only by additional time-consuming operations. Finally, *s*-triazine herbicide TPs are less amenable than their parent compounds to extraction with organic solvents.

Large volumes of water are better handled by a continuous liquid–liquid extractor.[14] A drawback of this technique is that the extraction step takes several hours. Especially for more polar compounds – such as *s*-triazine metabolites – LLE is not the best choice.

2.1.2 Solid-phase Extraction

SPE is becoming the primary extraction technique for pesticides. Besides eliminating many problems associated with LLE, SPE is particularly attractive because it may be coupled on-line with chromatographic systems. Another feature of SPE is its adaptability to field extraction, by doing this organic pollutants can be isolated immediately from an aqueous matrix so that chemical and biological degradation can be avoided. Off-line SPE is commonly accomplished by forcing a water sample through the sorbent material in a small plastic cartridge by applying positive pressure or vacuum (Figure 1).

2.1.2.1 Nonspecific Sorbent In the last few years, filter disks consisting of C18 or polystyrene-divinylbenzene (PS-DVB) sorbents imbedded in an inert material have become commercially available for SPE. Advantages claimed for disk design over cartridge one are: shorter sample processing and decreased plugging by particulate matter due to the large cross-sectional area; reduced channeling effect; and a background cleaner from interferences (Figure 2). The latter advantage derives from the fact that, unlike SPE with cartridges, the extraction apparatus with disks can consist of glass.

The performances of these two types of materials on extracting *s*-triazine herbicides are compared by using gas chromatography/mass spectrometry (GC/MS) detection; this method allows a limit detection between 0.06 and 0.2 $\mu g\,L^{-1}$ in the full-scan mode.[25]

It has been shown that nonselective sorbents, as derivatized silica and apolar copolymers, allow determination of many pesticides at the 0.1 $\mu g\,L^{-1}$ level.[26–32] However, the main problem is the occurrence of many coextracted and coeluting interfering compounds. Determination of pesticides at nanogram

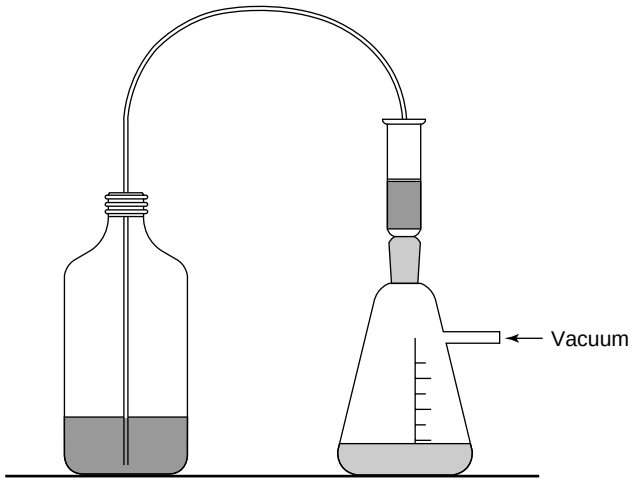

Figure 1 Schematic representation of an apparatus for off-line SPE.

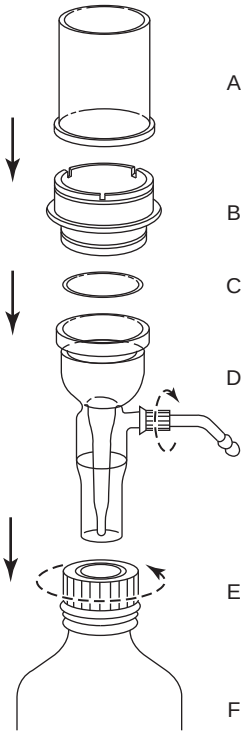

Figure 2 Schematic representation of the Baker Separex extraction apparatus: (A) sample reservoir; (B) Teflon connector; (C) Empore 47-mm-diameter disk; (D) filter holder; (E) screwcap; (F) filter funnel.

levels in surface water often requires an effective clean-up step. There is a trend to develop and apply more selective sorbents. A new sorbent material is graphitized carbon black (GCB). Due to the presence of chemical heterogeneities bearing a positive charge on its surface,[33] GCB can both be used as a reversed-phase sorbent and as an anion exchanger.[34] *N*-substituted *s*-triazines are adsorbed on the GCB surface by nonspecific interaction, but, by exploiting the GCB double nature, separation of triazines from coextracted acidic compounds (humic acids) can be easily performed. In this way extraction, preconcentration and clean-up can be performed with a single SPE cartridge.[14,35–37]

2.1.2.2 Immunosorbents Recently, interest has been growing in employing the highly selective analyte-antibody achieved by immunosorbents. In the immunosorbent, the antibody is immobilized onto a solid support and used as an affinity ligand to extract the target analytes and other compounds with similar structure from the aqueous sample. In this way, any component which is not recognized by the antibody is not retained, provided that there is no interaction with the solid support while the target analyte is bound to the antibody, leading to a high selectivity (Figure 3). Immunosorbent columns (with polyclonal and monoclonal antibodies) were used

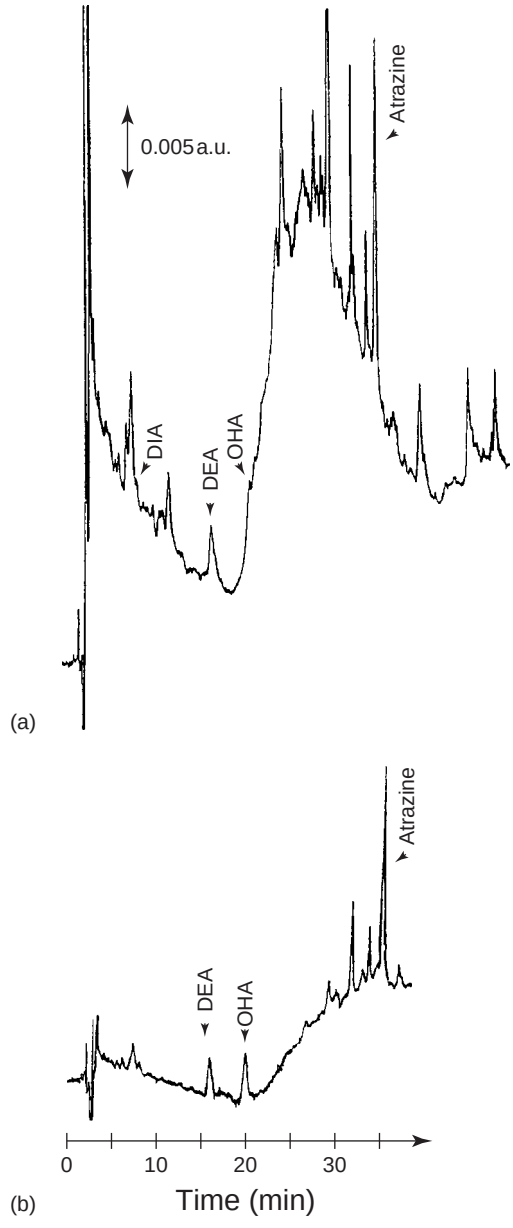

Figure 3 Analysis of the residue of a nonspiked soil sample percolated onto: (a) nonselective sorbent (apolar copolymers PS-DVB); and (b) an anti-triazine immunosorbent. Reversed-phase LC of a water–acetonitrile gradient and UV/DAD at 220 nm. De-ethylatrazine (DEA), hydroxyatrazine (OHA) and atrazine identified at respective concentration of 7, 16 and 23 ng L^{-1}.[160]

to detect *s*-triazines in environmental water samples by liquid chromatography/ultraviolet (LC/UV) (diode array detection, DAD)[38,39] or liquid chromatography/mass spectrometry (LC/MS).[38,40–42] This technique allows high recovery (except for dealkylated triazines) of many *s*-triazines at levels of 0.01–0.2 μg L^{-1} (Table 3). Baker's yeast (*Saccaromyces cerevisae*) cells were immobilized on

Table 3 Potential for class-specific immunosorbents as measured by recoveries obtained for samples spiked with a mixture of *s*-triazines. Recoveries measured by percolation of 50 mL of water spiked with 3 μg L^{-1} of each analyte. Mean values of three experiments: averaged relative standard deviation values in the range of 3–9%

Triazines	Anti-atrazine	Anti-simazine
De-isopropylatrazine	<5	56
Hydroxyatrazine	60	<5
De-ethylatrazine	98	30
Simazine	99	93
Cyanazine	91	74
Simetryn	63	17
Atrazine	99	88
Prometon	65	<5
Sebutylazine	88	91
Propazine	101	57
Terbutylazine	98	85

silica gel for the selective preconcentration of *s*-triazines and their TPs.[39]

2.1.2.3 Molecularly Imprinted Polymers Production of antibodies is laborious, time-consuming and expensive. Selective polymeric material can be prepared by the synthesis of highly cross-linked polymers in the presence of a template or ''print'' molecule followed by removal of the template. The remaining polymer will contain binding sites, which are complementary to the template molecule. Single or multiple interactions can describe the linking of the analyte and/or template molecules to the monomers (Figure 4). The types of interaction can be hydrogen bonds, ion pairing, $\pi-\pi$ interaction or hydrophobic-effect-driven.[43] MIPs are chemically more stable than antibodies and can cope with higher temperature and a wide range of solvents and pH. Molecularly imprinted materials have been found to exert antibody-like affinities towards the target substances and can be used for selective extraction. Highly selective MIPs can be prepared to extract triazine from complex matrices.[44–49]

2.1.2.4 Supported Liquid Membrane This sample handling technique was developed by Audunsson in the mid-1980s.[50] It utilizes a porous polytetrafluoroethylene membrane, which is impregnated with organic solvent immiscible in water. The organic solvent forms a barrier between the sample and the aqueous acceptor, which the analytes have to pass to be extracted (Figure 5). The supported liquid membrane (SLM) technique can be described as a two-step LLE procedure in a flow system. It is typically applied to acidic or basic compounds, but recently also extraction of permanently charged components has been demonstrated.[51] A comparative study for the determination of methoxy *s*-triazines in

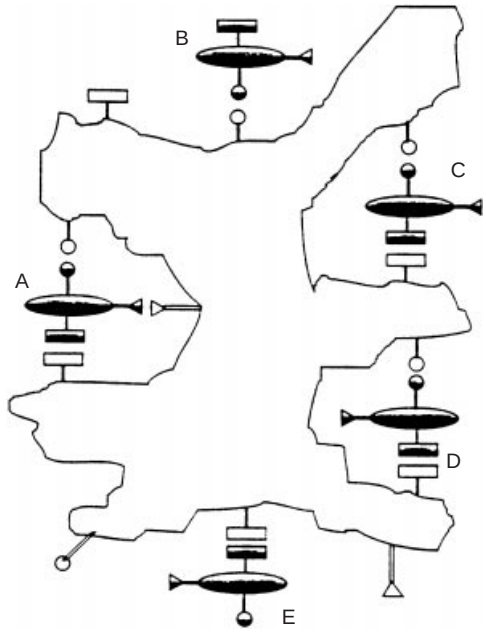

Figure 4 Schematic representation of the interaction of the analyte and synthesized polymer. The polymer contains several binding sites. The most favorable energetic interaction is shown at location A, where three binding sites result in the most specific interaction of the analyte with polymer. Two point interactions, which are energetically less favorable, are depicted at locations C and D, where the analyte is still allowed to rotate and, therefore, is less specific in its interaction with the polymer. Locations B and E show the least specific interaction possible and are, energetically, the least favorable.[43]

natural water by SPE and SLM has shown that a considerably cleaner extract can be obtained by the latter technique (Figure 6).[52] Various alkyl-*s*-triazines were enriched about 80 times from natural water.[53] However, the SLM procedure is slow and further improvements on this technique are necessary, especially the trapping of the analyte in the acceptor solution should be enhanced.[54]

2.1.2.5 Ion Exchanger Isolation of *s*-triazines, from water,[55] soil,[56] milk[57] and vegetable[55] extracts can be performed by using a strong cationic exchanger. Sorption of weak bases on a strong acid exchanger under strictly anhydrous conditions has been established to occur via

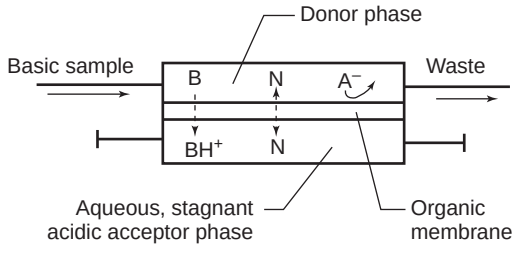

Figure 5 Schematic representation of the SLM principle. (Reproduced by permission from Jönsson et al.[51])

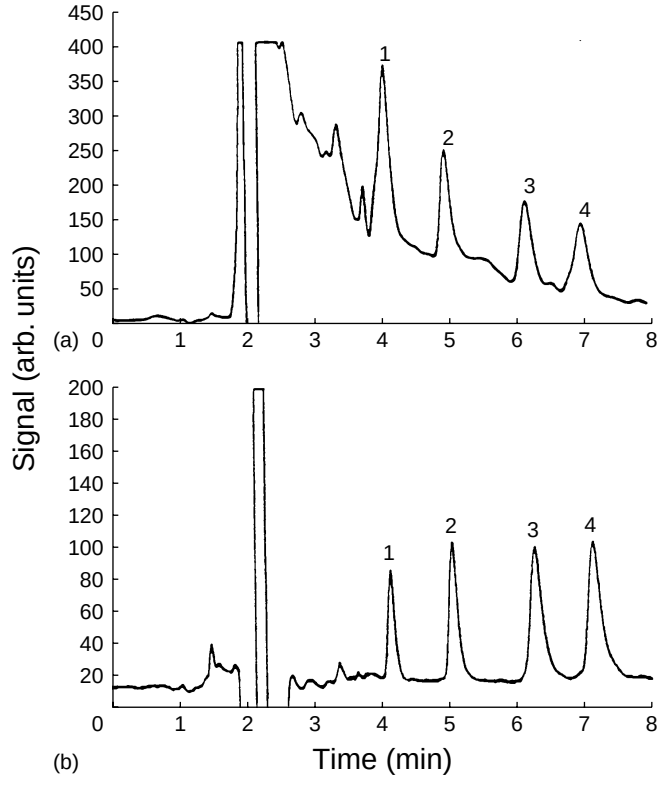

Figure 6 LC/UV chromatogram of methoxy-*s*-triazine herbicides. (a) SPE extract of spiked river water (1.0 μg L^{-1} of each analyte). (b) SLM extract of spiked river sample (0.5 μg L^{-1} of each analyte). Peak designations: (1) simetone; (2) atratone; (3) secbumetone; (4) terbumetone.[161]

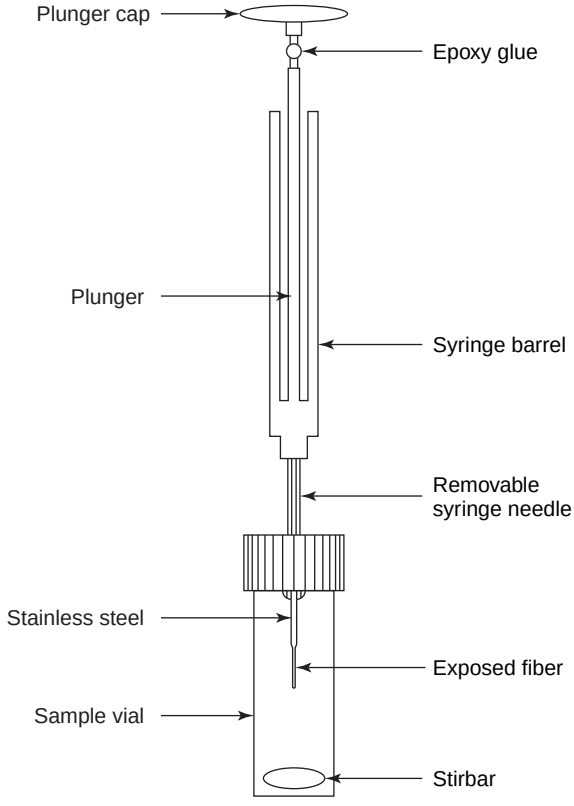

Figure 7 SPME device. (Reproduced by permission from Di Corcia.[162])

salt formation or hydrogen bonding.[58,59] This technique allows an effective clean-up and is useful to improve the sensitivity of a method.

2.1.3 Solid-phase Microextraction

SPME is a new technique used also for extracting analytes from water.[60] Figure 7 shows an SPME device. A 0.01–1 mm coated or uncoated fiber is immersed into a continuously stirred water sample. After equilibrium is reached (on average exposure times are 15–25 min), the fiber is introduced into the injection port of a gas chromatograph, where analytes are thermally desorbed and analyzed. Recently, SPME was successfully coupled with high-performance liquid chromatography (HPLC)[61,62] and automated systems have been developed coupling SPME with GC.[63] Parameters affecting SPME of triazine herbicides are salt effect, presence of humic acids, competition of major and minor components during the extraction process. In this experiment the polyacrylate phase was used and the sample was extracted over 30 min, followed by gas chromatography/nitrogen–phosphorus detection (GC/NPD) or GC/MS[64,65] (Figure 8a and b). Six *s*-triazine herbicides have been identified in groundwater well samples by an 8 μm polyacrylate fiber coating and GC/NPD.[62] The efficiency of this phase (85 μm) was also demonstrated by the time of extraction/desorption, pH and salt concentration study.[66] A multi-residue method for the simultaneous determination of 60 pesticides (atrazine, propazine and simazine included) was developed by using commercially available 100 μm polydimethylsiloxane fibers and GC/MS.[67]

2.1.4 On-line Solid-phase Extraction

In recent years, fully automated analysis of contaminants in water by on-line coupling of SPE to LC (Figure 9) or GC instrumentation has received increasing attention. Besides allowing rapid analysis, additional positive features of on-line SPE are that analyte loss due to evaporation does not occur and that the entire sample is introduced into the chromatographic instrument, instead of a fraction as with off-line procedures. In this way, the sampled volume can be reduced. In LC, when using a precolumn with a sorbent having a higher affinity for analytes than does the analytical column filling, broad peaks for the last eluted analytes

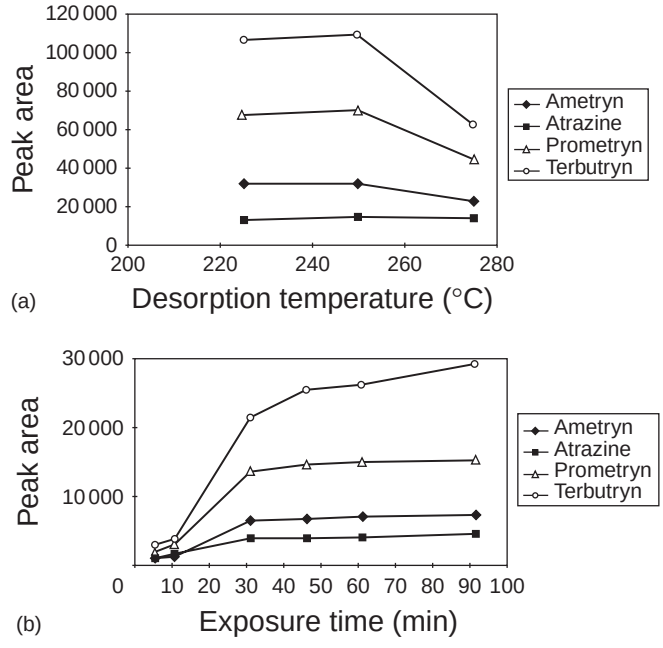

Figure 8 (a) Time dependence for the equilibration of the *s*-triazine herbicides between the aqueous and the polyacrylate phase. (b) Effect of desorption temperature.[66]

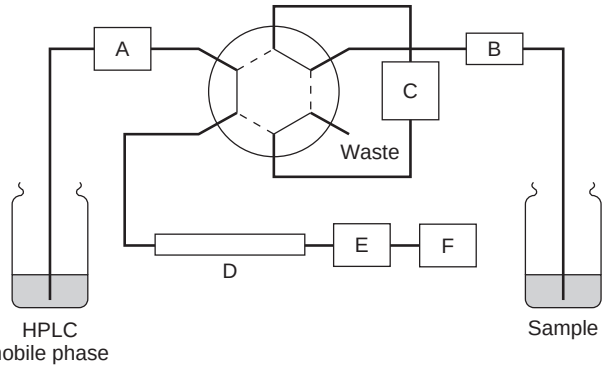

Figure 9 Set-up for the on-line trace enrichment of water samples where (-) represents load and (--) inject: (A, B) high pressure pumps; (C) extraction cartridge; (D) LC column; (E) detector; (F) recorder.

are obtained. In this case, backward elution of the analytes from the precolumn with the LC mobile phase can reduce peak broadening. SPE of *s*-triazine herbicides, using conventional reversed-phase materials (C18[68] or PS-DVB[69,70]), immunosorbent,[38] MIPs,[71] was successfully coupled with GC or HPLC with UV/DAD[40,41,68–70] and mass spectrometry (MS).[38]

2.2 Solid Sample

2.2.1 Soxhlet Extraction

Soxhlet extraction is still widely used, as it allows efficient isolation of pesticides from complex solid environmental samples. Adequate analyte recoveries depend on a suitable choice of the solvent. The extraction yields also strongly depend on the nature of the matrix. Weak basic triazine herbicides can be specifically adsorbed on soil and sediment particles, thus hindering efficient extraction of residues. To obtain an efficient extraction of organics from sediment and soil samples, the solvent should be able to penetrate into the pores of the matrix. Therefore, solvents having a relatively high surface tension should not be employed. Nonpolar solvents are often not suitable for extracting pesticides trapped in the core of soils and sediments, due to a poor wettability and limited miscibility with residual water present in the samples.[72–74] The wettability of a sample can be improved by just adding methanol and then the extracting solvent, for example ethyl acetate.[75] In a more recent study, methanol was used to extract a freeze-dried soil sample for 12 h.[76] The main drawbacks of this technique are the long time required by the extraction step and the large volume of solvent used. As an example, EPA method 3540 for extracting semivolatile organic priority pollutants, including some classes of pesticides, from soil or solid waste involves Soxhlet extraction with 300 mL of a solvent mixture for 16–24 h.[77]

2.2.2 Batch Extraction, Ultrasonic and Microwave-assisted Extraction

Blending the sample with a suitable solvent can be effective in many cases, blending the sample in the presence of the solvent in high-speed homogenizer machines or in an ultrasonic bath[78] ensures complete sample disruption and a better extraction of systemic pesticides.

Batch extraction of soil and corn samples with different mixtures of methanol, acetonitrile and acetone in a mechanical shaker for 90 min showed good recoveries.[79] Chloroform, dichloromethane, diethyl ether or methanol were used to extract terbutylazine and related TPs.[80]

Microwave-assisted extraction (MAE) is an alternative for the use of conventional extraction procedures. Water can be used to extract *s*-triazine herbicides and their TPs in a microwave oven while the sample is heated to 95–98 °C. This comparative study indicated that MAE with a mixture of dichloromethane–methanol (90 : 10, v/v) yields recoveries at least as good as those obtained by the conventional method. Moreover, the MAE procedure provides low solvent consumption in combination with a high sample throughput.[81]

2.2.3 Supercritical Fluid Extraction

SFE has been used in industry for many years, but only recently has it been employed in the analytical field for extracting a variety of pesticides from solid

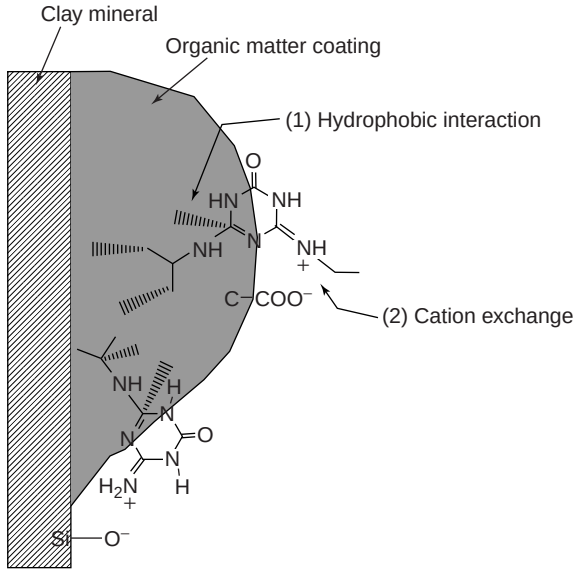

Figure 10 Mixed-mode model for adsorption of hydroxyatrazine to soil. (Reproduced by permission from the American Chemical Society.)

matrices, such as soil, sediment, vegetables and animal tissues. SFE emerged as an analytical tool in the mid-1980s (Figure 10). Carbon dioxide modified with polar cosolvents, such as methanol or water, gives a high recovery of the polar analyte.[82] Chromatograms of SFE extracts reveal fewer coextractants than seen in conventional solvent extraction. Optimal conditions for extraction are dependent on both the analyte and the soil matrix. SFE has been applied for the analysis of atrazine and its dealkylated TPs.[83–87] SFE of *s*-triazines was also applied to freeze-dried water samples as pretreatment for GC/NPD analysis.[88]

2.2.4 Accelerated Solvent Extraction and Subcritical Water Extraction

ASE has been proposed as an alternative for Soxhlet procedures. With this technique, a solid sample is

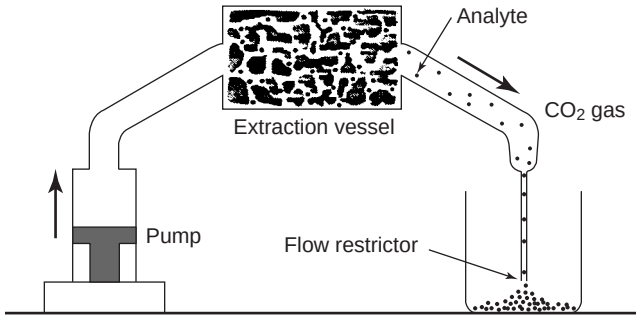

Figure 11 Schematic representation of an SFE device.

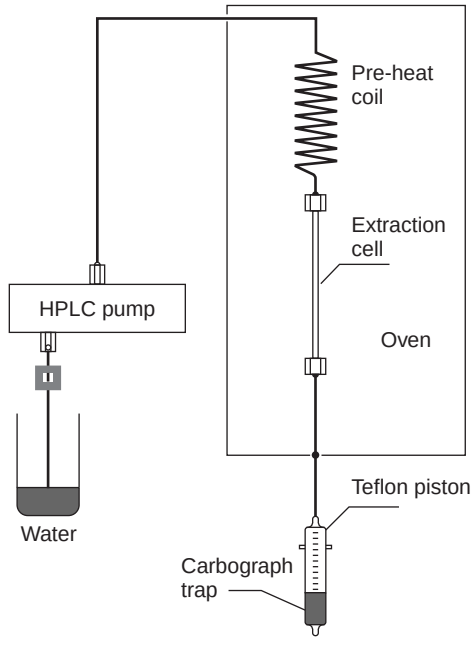

Figure 12 Schematic representation of the laboratory-made subcritical water extraction device.

packed into an extraction cartridge and analytes are extracted from the matrix with conventional low-boiling solvents or solvent mixtures at elevated temperature (up to 200 °C) and pressure (up to 20 MPa) to maintain the solvent in the liquid state. ASE allows extraction within a short time and with a small amount of solvent. Water has been used at ambient temperature to extract sulfonylurea herbicides[89,90] and phenoxy acid herbicides[91] from soil. Very recently, Hawthorne et al.[92,93] showed the effectiveness of subcritical water in extracting chlorophenols, alkylbenzenes, polycyclic aromatic hydrocarbons and *n*-alkanes from a solid matrix. Class-selective extraction of these organics was achieved by adjusting the water temperature (50–300 °C). This finding was explained by the fact that the polarity of water decreases as the temperature is increased, thus making water more compatible with nonpolar organics for adsorption on soil particles or soil organic matter (Figure 11).[94] The hot-water extraction procedure can also be called accelerated water extraction. The extraction device used to extract terbutylazine and TPs from soil by using phosphate buffer (0.5 mol L^{-1}, pH 7.5) in subcritical water condition[95] is shown in Figure 12.

3 IDENTIFICATION AND QUANTITATION

The most frequently used chromatographic techniques for the analysis of the final extract are GC with

conventional and mass spectrometric detectors. Except for hydroxylated TPs, GC analysis of the *s*-triazine compounds can be performed without a derivatization step. In recent years, several multiresidue methods have been developed for identifying and measuring triazine herbicides by HPLC with UV or mass spectrometric detectors.

3.1 Gas Chromatography with Conventional Detector

Gas–liquid chromatography (GLC) on a packed column has been used extensively for the analysis of triazines with various detection methods such as flame ionization detection (FID),[96–98] electron capture detection (ECD),[75,79,99] photoionization detection (PID),[100] Hall electrolytic conductivity detection,[101] Coulson electrolytic conductivity detection[75,79] and thermoionic (or nitrogen–phosphorus) detection.[19,21,102,103]

Nitrogen–phosphorus detection (NPD) is particularly suitable for the determination of triazine herbicides (Figure 13), all of which contain five nitrogen atoms per molecule.The nitrogen–phosphorus detector is a modification of the conventional flame ionization detector in which alkali salt or alkaline-earth salt is placed in the flame (alkali flame ionization detector). This gave a response for N- and P-containing molecules comparable to that obtained with FID, but the response for other molecules was negligible. For a compound containing an N or P atom, the response may be 1000–10 000 times higher than that for the corresponding hydrocarbon. The NPD detector was also found to be about 10 times more sensitive for triazines than the electrolytic conductivity detector. To improve the sensitivity using other detectors many derivatization methods have been developed. Conversion of *s*-triazines into their heptafluorobutyryl (HFB) derivatives[104] improves the sensitivity 300-fold to several thousands fold with ECD and 5- to 10-fold with electrolytic conductivity detection in the halogen mode. Trimethylsilyl derivatives of hydroxy and amino derivatives, cyanuric acid, ammelide, ammeline and melamine were separated by GLC and detected by FID.[105] Although several USEPA methods still require the use of packed columns, GC with capillary columns, commonly referred to as high-resolution GC (HRGC) has become the standard for analyzing pesticide residues. Twenty-nine *s*-triazine herbicides and *N*-dealkylated TPs were separated on a high-resolution glass capillary coated with Carbowax 20M glass capillary column at 200 °C.[106] In a comparative evaluation with a capillary column the sensitivity was found to be five times higher than with a packed column.[107]

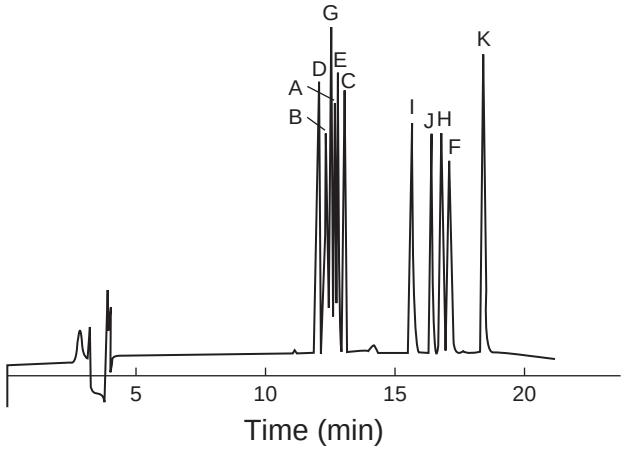

Figure 13 Analysis of triazine standards (1 ng each) by high-resolution gas chromatography (HRGC)/NPD: injection 250 °C, column, 30 m × 0.25 mm DB-1, 0.25 µm film; 70 °C, 0.5 min, 25 °C min^{-1} to 195 °C, hold 8 min; 10 °C min^{-1} to 200 °C: (A) prometone; (B) atratone; (C) propazine; (D) simetone; (E) atrazine; (F) prometryne; (G) simazine; (H) ametryne; (I) cyprazine; (J) simetryne; (K) cyanazine.[163]

3.2 Gas Chromatography/Mass Spectrometry

GC/MS is a well-established technique for the trace analysis of pesticides in environmental samples. This technique has benefited from the development of fused silica capillary columns, small relatively inexpensive mass spectrometers (quadrupoles) designed to be used as dedicated GC detectors, and sophisticated computer programs to perform computerized library searches.

Compounds eluted from the GC column enter the MS ion source where ionization of molecules occurs by impact with an electron beam. As a consequence, a series of ion fragments having characteristic mass-to-charge (m/z) ratios are formed. By scanning the MS over a defined m/z range, these ions are recorded by a photomultiplier or an electron multiplier and a resulting mass spectrum is obtained which displays m/z versus relative abundance. By summing the currents produced by each fragment ion and plotting vs time a total ion current (TIC) chromatogram is obtained (Figure 14). This chromatogram does not differ from those obtained with conventional chromatographic detectors. A much more sensitive method of quantitation of a target compound is that of plotting only one or few related ion masses from the TIC chromatogram. This method is called extracted ion chromatographic profiling (EICP). By EICP, accurate and sensitive analyte quantitation can be performed in the passage of a coeluted component, provided that two conditions are met. The first one is that the coextracted component does not produce the same selected ion. The second condition to be met is that the coeluted component does not affect the ionization process of the analyte. The maximum sensitivity with MS detection can be achieved by employing the selected ion monitoring (SIM) mode.

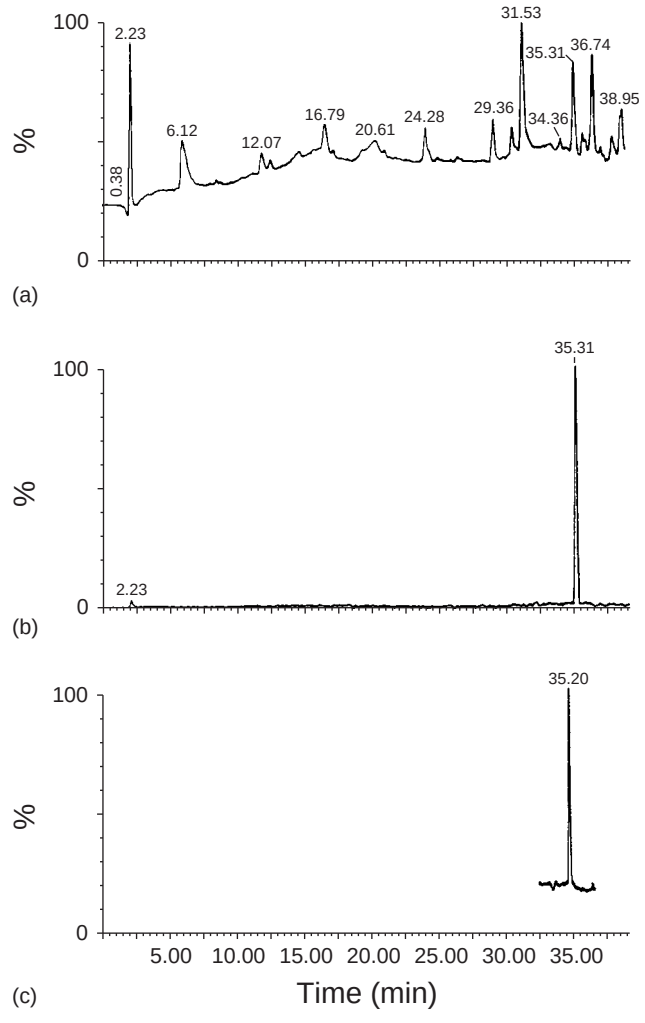

Figure 14 LC/ES/MS analysis of a spiked river water sample: (a) TIC chromatogram obtained by scanning the quadrupole filter from 40 to 220 m/z with 2-s scan time (sample spiked with 50 ng L^{-1} of *s*-triazine herbicides; (b) extracted ion chromatographic profiling (EICP) (same sample) of m/z 216 (molecular ion of atrazine); (c) selected ion monitoring (SIM) of two masses (m/z 174, 216) relative to the parent ion and the fragment ion from atrazine (sample spiked with 1 ng L^{-1}). The SIM and EICP traces are similar, but the ion signal of (c) was obtained by injecting 50-fold diluted sample.

Apparently, SIM looks like EICP. The former mode differs substantially from the latter one in that one m/z value is selected before the chromatographic run with SIM, the sensitivity of the analysis can be increased by a factor of 100 or more because the MS does not waste time to collect any other ion formed. Provided that the GC is able to separate all target components from each other and provides highly reproducible retention times, simultaneous SIM analysis for a number of analytes can be obtained by a time-scheduled program within a single chromatographic run, thus fully exploiting the potential of the MS detector. The weakness of the SIM operation mode by selection of one single ion for each analyte is that false positives can occur when analyzing pesticides in complex matrices. To avoid these false positives usually the molecular ion plus one or two characteristic fragment ions are selected for each analyte. In this way, quantitation can be performed by EICP of that ion which is not generated by the interfering compound. SIM by two or three ions can lower the sensitivity of the analysis to some extent, but greatly increases the accuracy. Analysis by time-scheduled SIM is the best compromise between sensitivity and specificity of the analysis. It is particularly useful when trace analysis of pesticides down to 100 ng L^{-1} in drinking water samples has to be performed for compliance with European standards. In this case it can be deduced that, considering that conventional GC injection devices permit introduction of 1 µL of a 500-µL final extract and that an easily interpretable background-subtracted spectrum with most MS detectors can be obtained by injecting 1–2 ng of an analyte, at least, the goal of obtaining highly diagnostic mass spectra cannot be reached by the full-scan acquisition mode. This situation however can be substantially improved by adopting large volume injector devices.

A complete study[108] on retention behavior and fragmentation under electron-impact and chemical ionization conditions shows that: (1) the differences in retention indices of the same triazines on a column with different polarity are proportional to pK_a values and to the dipole moments of these compounds; (2) under electron-impact conditions the stability of the compounds increases in the order chloro < methoxy < methylthio derivatives. Stable isotope dilution GC/MS was used for determination of low parts per billion concentrations of atrazine in water and soil.[109] An improved method (by high-resolution MS, HRMS) with lower detection limits allows the determination of triazine herbicides in parts per quadrillion range in groundwaters.[110]

3.3 Liquid Chromatography

Since several classes of pesticides cannot be analyzed by gas chromatographic techniques, official multi-residue methods based upon LLE/LC[20] or SPE/LC[111] with UV detection have been developed in recent years for a variety of pesticides including triazine herbicides and their metabolites in aqueous matrices. An additional feature of LC is that it can be easily coupled to both off- and on-line SPE. For analyzing *s*-triazines, the stationary phase is almost invariably silica chemically modified with reagents having an octadecyl alkyl chain. Commonly, this stationary phase is called simply C18. However, when analyzing basic pesticides, unreacted slightly acidic silanol

groups can play an important role in the fractionation process. The mobile phase is constituted by buffered water mixed with an organic solvent (usually methanol or acetonitrile). The triazine ring is a chromophore able to efficiently absorb radiations in the low UV region, that is, 210–240 nm. When operating in this region, the main drawback of the UV detector is that it can be classified as a quasi-universal detector. The main positive features of the UV detector are the high reproducibility and ruggedness, a broad linear dynamic range (10 000), a response virtually unaffected by the temperature and by composition of the mobile phase.

The most recent development in spectrophotometric detection has been the introduction of photodiode array detectors. Briefly, all the light from a deuterium lamp is passed through the sample cell onto the monochromator, which spreads out the beam into a spectrum. This falls across an array of 230–350 photodiodes mounted on a silicon chip (Figure 15). These can be read simultaneously by computer software to provide the full absorption spectrum from 200 to 700 nm every 0.1 s. In a simpler mode, the ratio of the absorption at two wavelengths can be used for analyte identification. This type of detector has a tenfold lower sensitivity in comparison to conventional UV detectors. Figure 16 shows a chromatogram obtained by analyzing 150 mL of a Seine River water sample by on-line SPE/LC/DAD. Apparently, peaks having retention times similar to those for simazine and atrazine were recorded. By comparing DAD/UV spectra, atrazine, but not simazine, was identified.

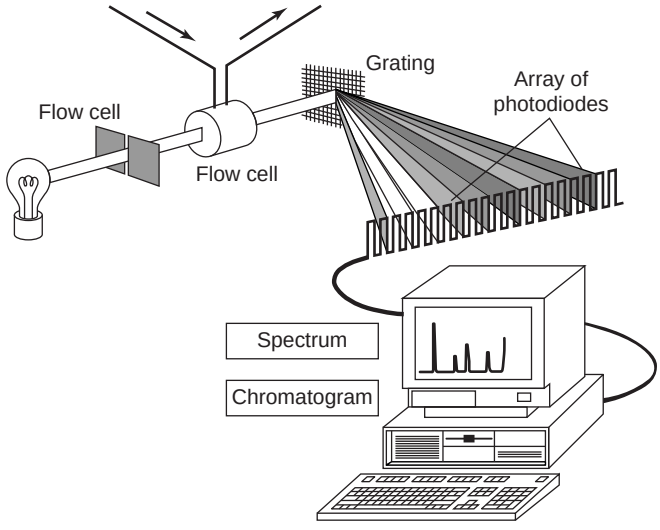

Figure 15 Schematic representation of photodiode array detector. (Reproduced by permission from Di Corcia.[162])

Using conventional DAD, the extremely wide range of pesticides can pose a difficult identification problem due to peak overlap. In addition, pesticides may coelute with other unknown sample constituents giving problems with quantitation. MCA, a technique that deconvolutes and quantifies known UV-absorbing substances in an unknown solution, can be used to examine the presence of a large number of pesticides from different classes in a single run.[70] With an on-line preconcentration of 200 mL of drinking water, detection limits ranged between

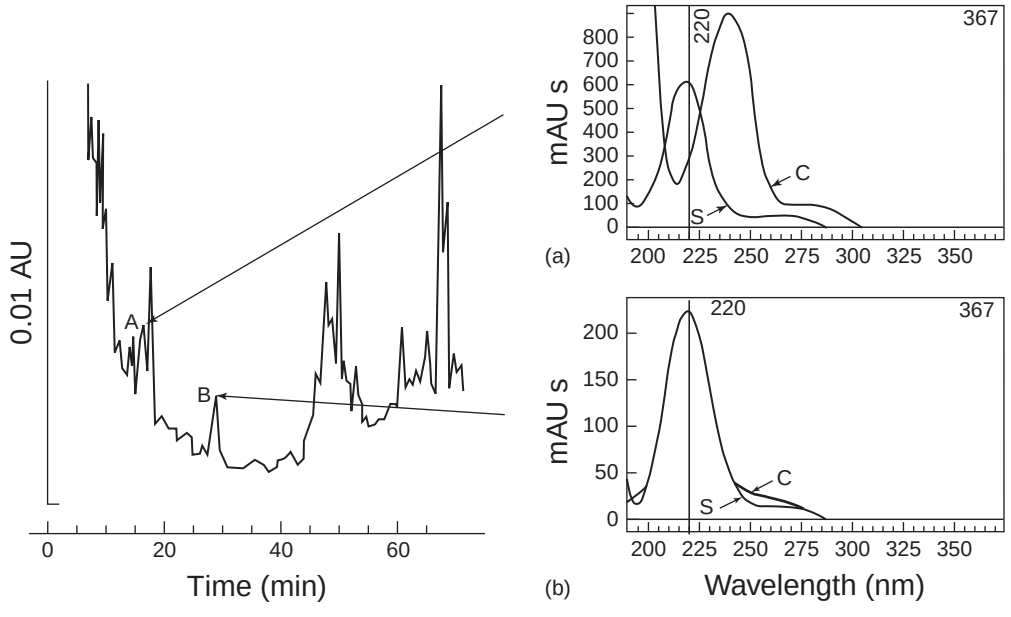

Figure 16 LC chromatogram and UV spectra obtained by analyzing a 150-mL sample from the Seine River. Comparison of standard spectra and unknown peak spectra for (a) simazine and (b) atrazine, where S refers to the standard and C to the peak of the unknown compound. (Reproduced by permission from Pichon and Hennion.[40])

10–50 ng L^{-1}. Because the UV spectra of *s*-triazine are not characteristic, DAD cannot always discriminate between triazines and matrix components. Selectivity problems in the LC analysis of *s*-triazines in real-life matrices can be resolved only by coupling LC to a mass spectrometer.

3.4 Liquid Chromatography/Mass Spectrometry

Nowadays LC/MS is applied in biomedical, pharmaceutical and environmental chemistry. For polar pesticides most of the technical problems of LC/MS have been solved and for many analytes the specificity and sensitivity with this technique are similar to GC/MS. Currently, the most popular LC/MS interfaces are TSP, particle beam, APCI and ES.

3.4.1 Thermospray

Besides separating solutes from the solvent mixture, the TSP interface is able to ionize compounds in solution at atmospheric pressure (Figure 17). Nonvolatile compounds can be ionized with TSP and detected with sensitivities comparable to GC/MS. Before entering the interface, the liquid flow from the LC column passes an electrically heated capillary with an internal diameter of 50–100 µm. Because of the partial evaporation in the capillary, a supersonic jet is formed that contains, in the presence of ionizing solvents, some charged droplets. Droplet charging occurs as a result of a statistical distribution of ions over the micrometer-size droplets.

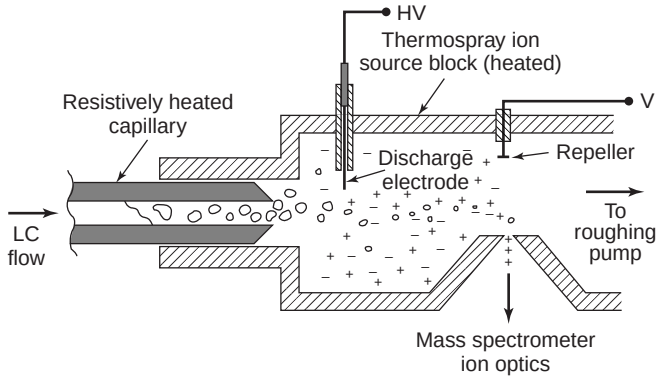

Figure 17 Schematic representation of a TSP interface. V, voltage; HV, high voltage. (Reproduced by permission from Di Corcia.[162])

While passing the heated TSP source, the size of the droplets are more and more reduced, because of solvent evaporation. As a result, ions are expelled from the highly charged surface droplets, or ions are formed by gas-phase reactions. These ions enter the MS region through a skimmer. An electrode (the repeller), with a polarity equal to that of the ions of interest and set in front of the cone, forces ions into the MS system. Ion desorption from the surface droplet is achieved by the addition of a volatile buffer salt, typically ammonium acetate, added to the LC mobile phase. In most cases, ion evaporation shows $[M + H]^+$ ions for compounds with high proton affinity, such as *s*-triazine (Figure 18). Otherwise, $[M + NH_4]^+$

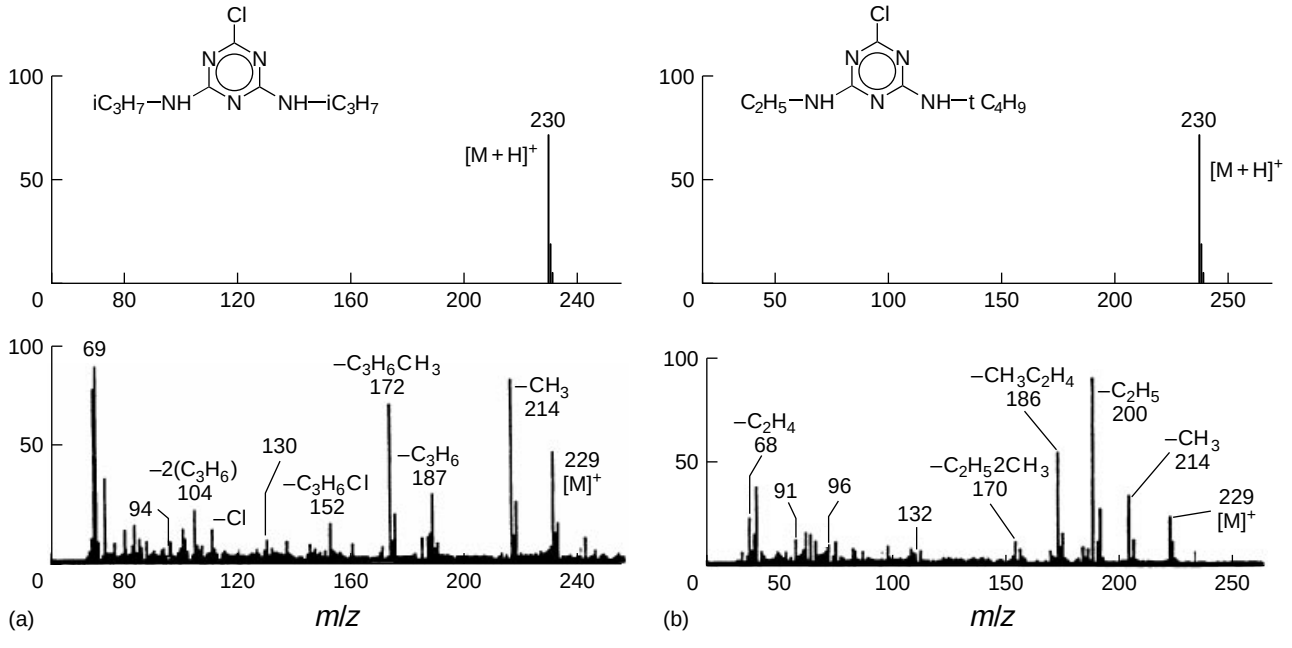

Figure 18 Mass spectra of (a) propazine and (b) terbutylazine. The top graphs represent TSP spectra, and the bottom graphs, particle beam EI spectra. (Reproduced by permission from Di Corcia.[162])

ions are formed. Owing to the soft ionization process, TSP spectra often do not display fragment ions. In some cases, fragmentation can be induced by using the discharge assisted TSP technique or by using higher repeller voltages. The device increases the kinetic energy of the ions; and the collision-induced decomposition (CID) process generates fragment ions. The specificity of methods involving TSP interface can be greatly enhanced by using tandem MS. TSP has been successfully employed in the analysis of *s*-triazines.[26,32,112,113]

3.4.2 Electrospray

The ES interface enables the LC/MS analysis of compounds with molecular weights up to 5 000 000 Da. Eluate of the LC column enters a chamber in which a high electrical field (3.5–4.0 kV) is imposed by a power supply. This electrical field causes an enrichment of positive electrolyte ions at the meniscus at the metal capillary tip. This net charge is pulled down field, expanding the meniscus into a cone (the so-called Taylor cone). If the electrical field is sufficiently high, the cone becomes unstable and filaments are formed from which positively charged droplets (positive ionization ES mode) detach and migrate towards the counter electrode. During migration, droplets shrink because of solvent evaporation caused by moderate heating and by a counterflowing gas (drying gas). It has been postulated that, when the surface concentration of the positive charges becomes sufficiently high, repulsion forces overcome cohesive forces and droplets explode generating daughter droplets (Figure 19). This process proceeds until the droplet radius reaches the Rayleigh limit. At this point, ion evaporation competes with further reduction of the droplet size. Any analyte capable of forming adducts with protons or other inorganic cations via ion-dipole forces can form ions in the gas phase in a variety of solvents. Moreover, the ionization process appears to be nearly 100% efficient, although not all of the ions formed are sampled. A cluster of ions with solvent molecules are sampled by a cone with a small orifice (100 μm) and pass into a first vacuum chamber (ca. 1 torr), called the dissolving chamber, where a supersonic expansion occurs. There, ions are declustered by collision with the residual molecules of drying gas. The core of the expansion is sampled by two skimmers, the second one set at about 20 V, and transported to the mass analyzer region (Figure 20).

Pneumatically assisted ES uses a nebulizer gas flowing coaxially with the liquid phase to create shrink droplets. With this device, the ES source can accept more than 1 mL min^{-1} of liquid phase (including water). Because the ES/MS arrangement is a concentration-sensitive

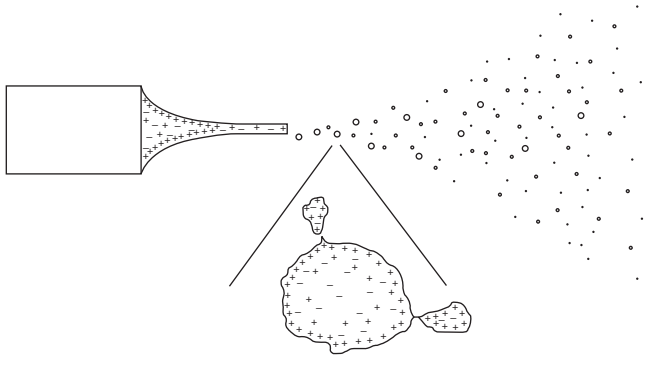

Figure 19 Schematic representation of the coulombian explosions of charged droplets in the ES source.

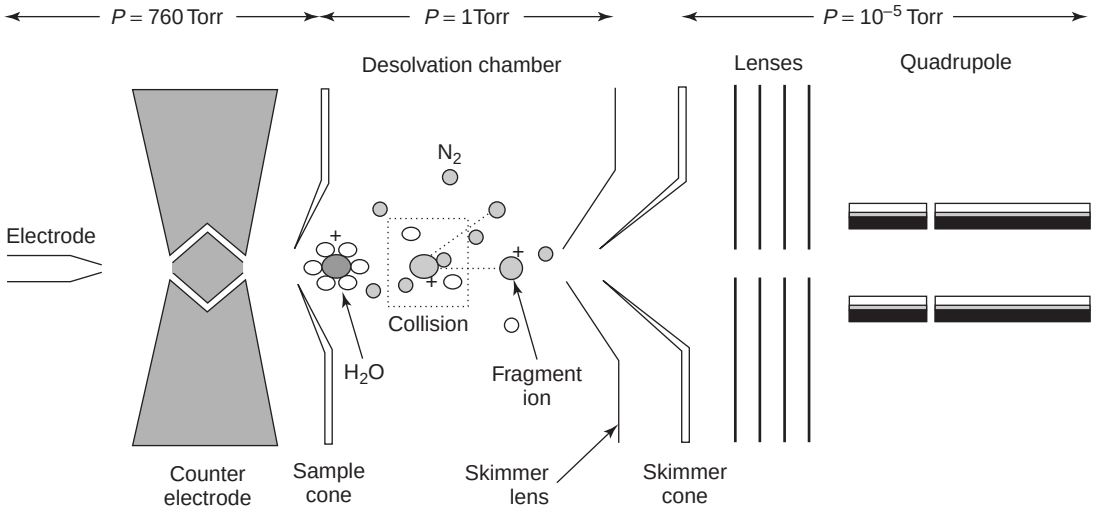

Figure 20 Schematic representation of the formation of fragment ions as a result of the collisions between $[M + H]^+$ ions and residual drying gas molecules.

detector,[114] diverting a fraction of the LC mobile phase to another detector does not affect the sensitivity. The problems arising from the presence of relatively high amounts of salts in the LC effluent have not yet been completely solved.

In the ES process, ionization is soft, leading to the formation of $[M + H]^+$ (positive ionization) or $[M - H]^-$ (negative ionization), even for most thermally labile and nonvolatile substances. Furthermore by increasing the electrical field in the dissolving chamber, the quasi-molecular ions can be accelerated so that multiple collisions with residual molecules from the drying gas occur generating characteristic fragment ions (Figure 21). The resulting CID spectra closely resemble those obtained by tandem MS system. In *s*-triazine LC/MS analysis addition of a low concentration (10^{-5} mol L^{-1}) of formic acid resulted in a 60% increase of ion signal.[115]

3.4.3 Atmospheric Pressure Chemical Ionization

Analyte ionization can also be induced by gas phase ion-molecule reactions under atmospheric pressure conditions. Reactant ion formation is achieved by the introduction of electrons from a corona discharge located in the chamber at atmospheric pressure. The discharge, occurring at the tip of a needle held at high voltage (4–6 kV), is routinely used in combination with both pneumatic and heated nebulization. This process allows formation of ions similar to those achieved with an ES source also for less polar components. Figure 22 shows an LC/APCI/MS chromatogram and spectra of a river water sample spiked with 1 μg L^{-1} of atrazine. A number of papers have been published on the determination of *s*-triazine using LC/APCI/MS.[24,116,117]

3.5 Immunochemical Assay

Immunoassays are detection methods based on a reaction between a target analyte and a specific antibody. Quantitation can be performed by monitoring a color change or by measuring radioactivity, fluorescence or other physicochemical properties. The recent rapid growth in immunochemical methods is attributed in part to the availability of polyclonal and monoclonal antibodies for a variety of components of environmental significance. Another driving force in the development of immunochemical methods is the need for rapid and simple tests that can be performed on-site without requiring sample transfer to an analytical laboratory. Many immunoassay-based field-portable methods enable rapid determination of target components by personnel without an analytical chemical training.

In a typical enzyme-linked immunosorbent assay (ELISA), a solid phase (beads or the wells of microtiter plates) is coated with a hapten–protein conjugate (Figure 23). A solution of a specific antibody that has been mixed with a calibration standard or sample is then added. During an incubation period, the analyte in solution and the hapten–protein conjugate immobilized on the solid phase compete for binding to the specific antibody. Washing with buffer removes excess antibody remaining in the solution and any antibody–analyte complex that has been formed. Then an enzyme-labeled secondary antibody is added, which will bind to the antibody captured on the solid phase. Alkaline phosphatase and horseradish peroxidase are commonly used as enzyme labels. An enzyme substrate is added, causing the development of a color. Absorbance or the rate of a color development can be measured with a spectrophotometer. Quantitation is based on competition for antibody-binding sites. In this assay format, color intensity is inversely related to the concentration of analyte in the sample. A single immunoassay takes 2–24 h and many samples can be processed in parallel. Also immunoassays can be easily automatized.[118] The disadvantage of this method is that the enzyme can be sensitive to matrix interferences. Control samples are necessary for quantitation of matrix effects and antibody cross-reactions.

Comparative studies can be carried out, using environmental samples, between immunochemical and conventional methods such as GC or GC/MS[16,118,119] or HPLC.[120] Time-resolved fluorescence immunoassay utilizes a fluorescent label and eliminates the need for an enzyme, thus making this method potentially less susceptible to matrix interferences.[121] A more rapid immunoassay[122] utilizes a specific antibody covalently coupled to a magnetic particle solid phase that has been shown to be more precise than methods using polystyrene-coated wells and tubes.[123] A screening test for determination of 0.15–4.00 ng L^{-1} of atrazine in drinking-, well- and surface water by magnetic particle immunoassay has been approved by the Association of Official Analytical Chemists.[124] A dipstick immunoassay format for atrazine[125] and terbutylazine[126] determination, based on the use of polystyrol strips as antibody coating support, can be very useful as a qualitative/semiquantitative "field test" for identifying "positive" samples.

3.6 Electrochemical Biosensor

A biosensor (or immunosensor) is a device that contains a biological component (e.g. enzymes, antibodies, or cell receptors) intimately connected to a transducer (e.g. electrochemical, optical, piezoelectric, or thermistor solid-state device). Whenever an interaction occurs between the target analyte and the biological component of the biosensor, some properties of the transducer

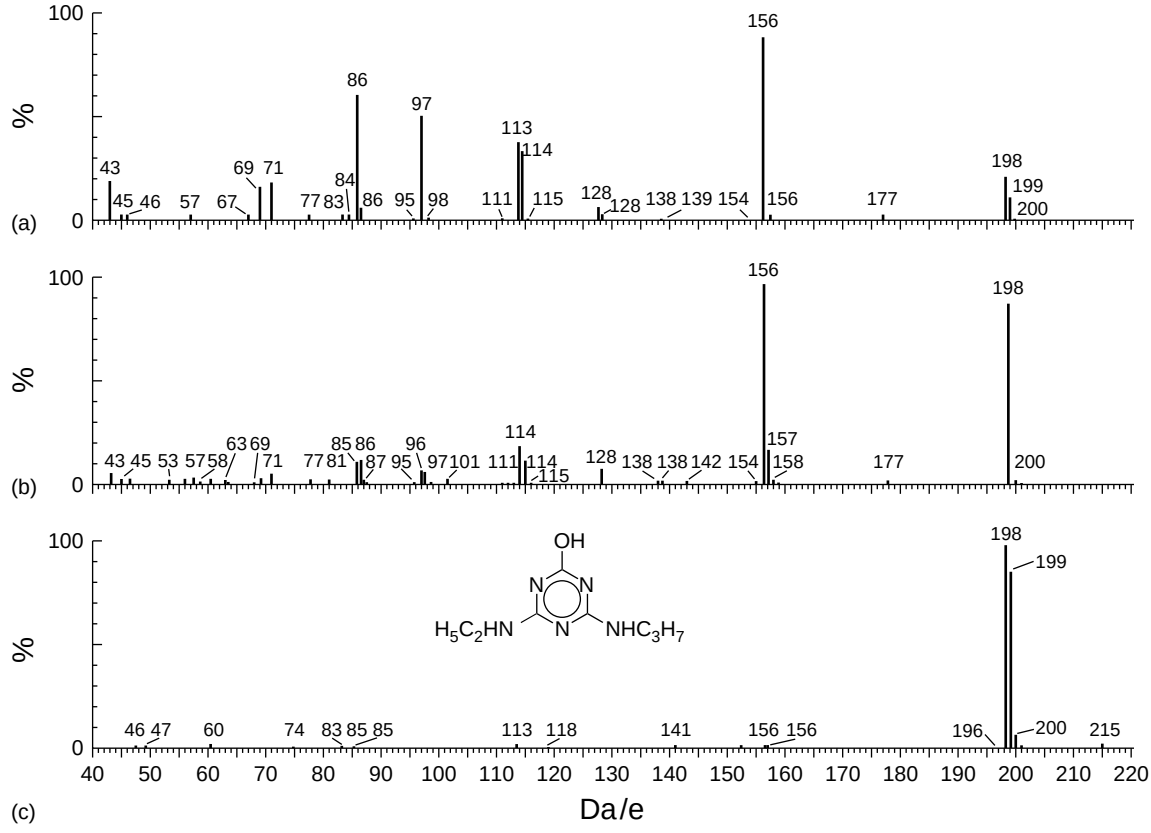

Figure 21 Background-subtracted CID spectra showing production of fragment ions from the $[M + H]^+$ ion for hydroxyatrazine by progressively increasing the electrical field in the desolvation chamber from (c) 20 V, to (b) 40 V, to (a) 50 V.

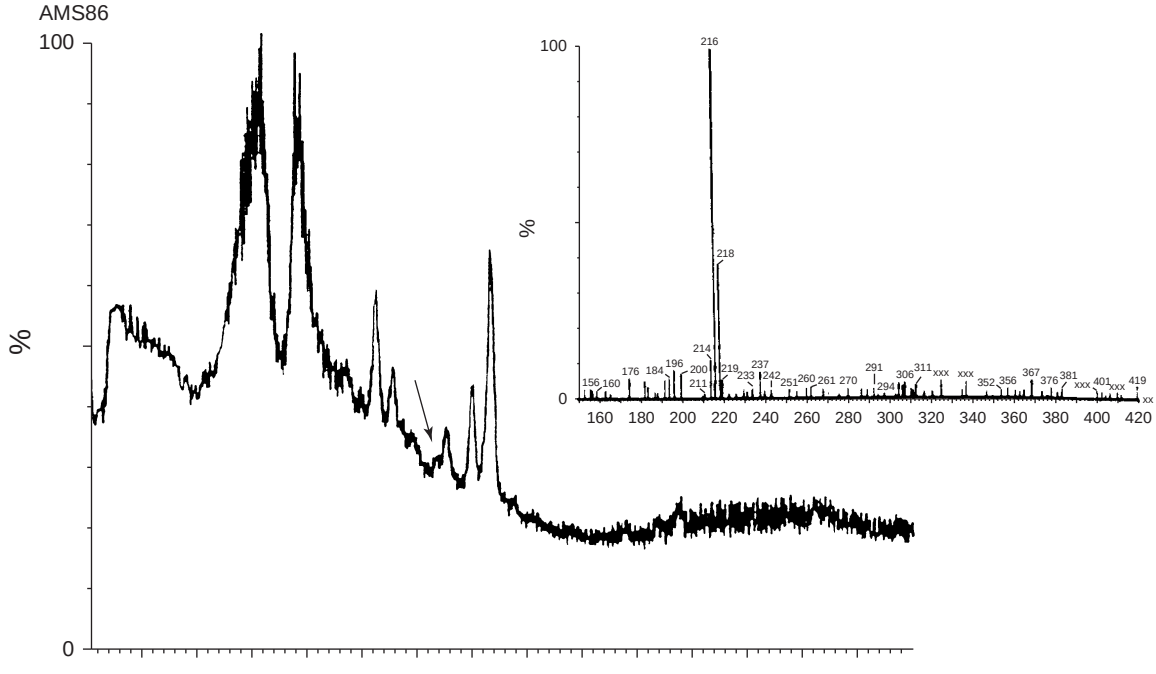

Figure 22 LC/APCI/MS spectrum and TIC chromatogram of 15 mL river water containing atrazine. Sample was preconcentrated with a short column system and acquired in full-scan mode. (Reproduced by permission from Springer-Verlag.)

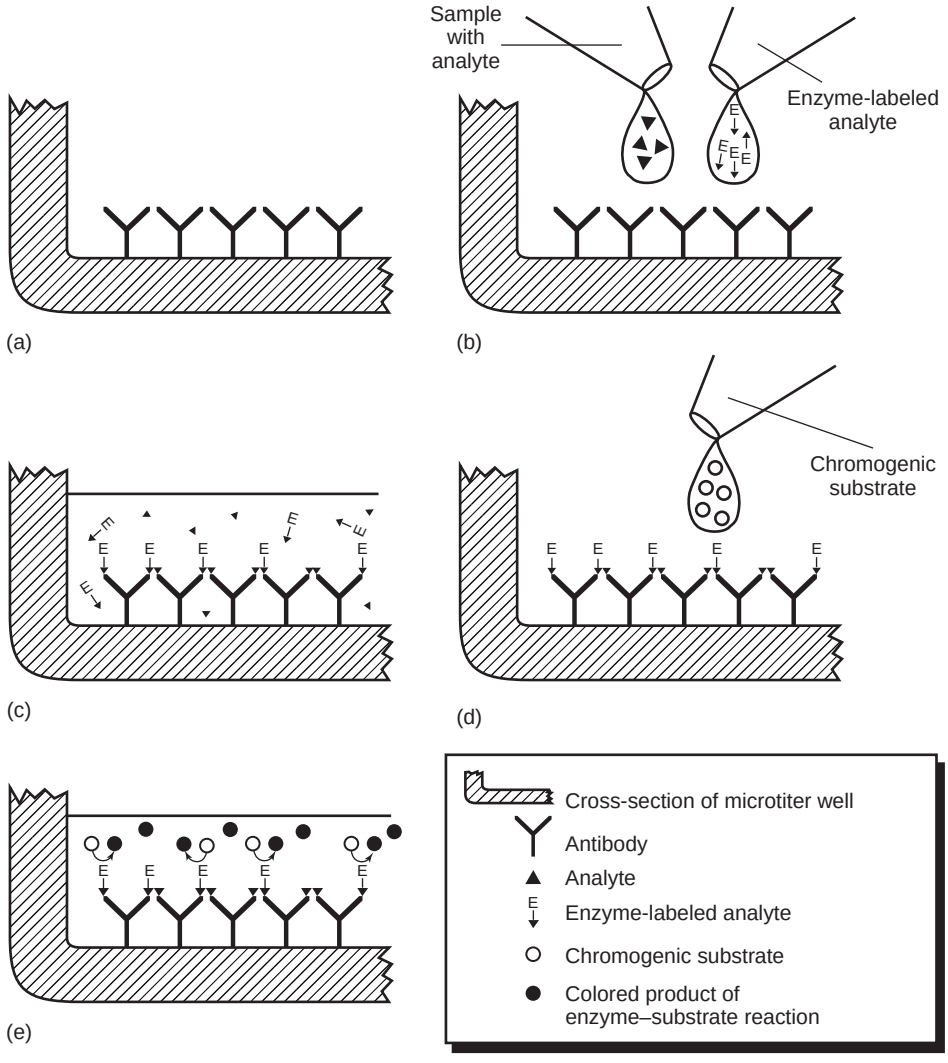

Figure 23 ELISA format of a field-portable immunoassay. (a) Antibody is adsorbed onto a microtiter plate or test tube. (b) The sample and a known amount of enzyme-labeled analyte are added. (c) The analyte and the enzyme-labeled analyte compete for antibody-binding sites. (d) A chromogenic substrate is added for color production. (e) The color intensity, which is inversely related to the amount of analyte in the sample, can be estimated visually or measured with a spectrophotometer.

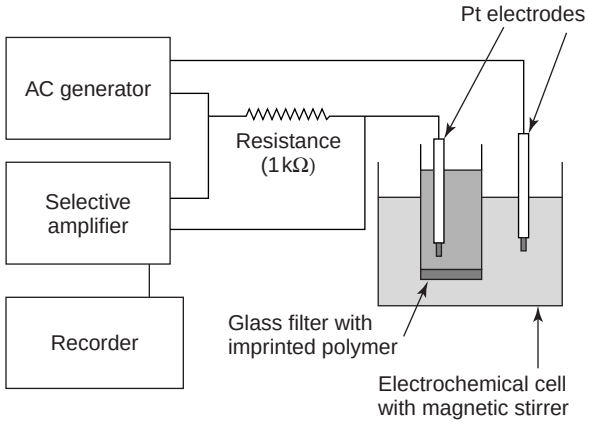

Figure 24 Schematic representation of an atrazine sensor.[128]

are altered and subsequently measured[127] (Figure 24). The substitution of membranes by MIPs of biological molecules in biosensors enables the development of a highly stable sensor that can be used in harsh environments.[128] Stable polymeric selective membranes have been used for conductometric[129] and fluorescence detection systems.[130]

4 RECENT APPLICATIONS

The occurrence of pesticides and other chemicals forced authorities to organize monitoring programs in order to evaluate risks from groundwater pollution. In these

programs *s*-triazines play a relevant role. In 1991, the US Geological Survey started the National Water Quality Assessment program (to be completed in 2002) to assess nationwide streams and groundwater quality in 20 of the USA major watersheds.[5] In this project (including analyses of 76 pesticides and 7 selected TPs) *s*-triazines were monitored by SPE followed by GC/MS determination. These data are the most extensive ever collected for such a wide range of pesticides and locations. A new computer tool ''Geographic Information System'' can easily map all these analytical data. Many other regional-scale monitoring programs were performed to assess the environmental water contamination by *s*-triazines. To curtail expenditure, analyses can be performed with low cost methods and confirmed by a mass spectrometric technique. This approach was used in a study on pesticides and TP transport. A rapid and inexpensive screening method such as ELISA was chosen because of its low cost per sample and GC/MS was used as the confirmatory technique.[16]

Occurrence and behavior of *s*-triazines in closed aquatic systems such as Swiss lake basins were thoroughly studied by SPE and GC/MS[131,132] or HPLC/UV[133] (DAD) analysis. The impact of *s*-triazine herbicides on open aquatic systems, such as estuaries, were performed by analyzing water and sediments in estuarine water of the Rivers Thames, Tamar and Mersey by SPE and HRGC with nitrogen-selective detection.[134] Another study was performed on River Ebro Delta water by LLE followed by GC/NPD and GC/MS as confirmatory technique.[135]

Groundwater can be contaminated by the leaching of pesticides. Many researchers studied the process that governs the accumulation and transport[136] or biodegradation[137,138] processes of *s*-triazines in soil.

Personnel that handle, mix or apply atrazine to corn and other crops face potential dermal and/or inhalation exposure. To estimate this potential exposure, excretion kinetics have to be known. Residues in urine of *s*-triazine and its dealkylated metabolites can be determined in urine at the parts per billion level by GC/MS.[139]

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
ASE	Accelerated Solvent Extraction
CID	Collision-induced Decomposition
DAD	Diode Array Detection
ECD	Electron Capture Detection
EICP	Extracted Ion Chromatographic Profiling
ELISA	Enzyme-linked Immunosorbent Assay
ES	Electrospray
FID	Flame Ionization Detection
GC	Gas Chromatography
GCB	Graphitized Carbon Black
GC/MS	Gas Chromatography/Mass Spectrometry
GC/NPD	Gas Chromatography/ Nitrogen–Phosphorus Detection
GLC	Gas–Liquid Chromatography
HFB	Heptafluorobutyryl
HPLC	High-performance Liquid Chromatography
HRGC	High-resolution Gas Chromatography
HRMS	High-resolution Mass Spectrometry
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/UV	Liquid Chromatography/Ultraviolet
LLE	Liquid–Liquid Extraction
MAC	Maximum Allowable Concentration
MAE	Microwave-assisted Extraction
MIP	Molecularly Imprinted Polymer
MS	Mass Spectrometry
NPD	Nitrogen–Phosphorus Detection
PID	Photoionization Detection
PS-DVB	Polystyrene-divinylbenzene
SFE	Supercritical Fluid Extraction
SIM	Selected Ion Monitoring
SLM	Supported Liquid Membrane
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
TIC	Total Ion Current
TPs	Transformation Products
TSP	Thermospray
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Urinalysis and Other Bodily Fluids

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Gas Chromatography with Selective Detectors for Amines • Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and

Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Liquid Chromatography in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Immunoassays in Forensic Toxicology • Mass Spectrometry for Forensic Applications

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction • Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Herbicide Residues in Biota, Analysis of • High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Immunochemical Assays in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. E. Knuesli, D. Berrer, G. Dupuis, H. Esser, '*s*-triazine', in *Degradation of Herbicides*, eds. P.C. Kearney, D.D. Kaufman, Marcel Dekker, New York, 51–78, 1969.
2. V. Andriska, 'Herbicides', in *Pesticides Chemistry, Environmental Studies*, eds. G. Matolcsy, M. Nádasy, V. Andriska, Elsevier, Amsterdam, Vol. 32, 694–700 1988.
3. C.D. Watts, L. Clark, S. Hennings, K. Moore, C. Parker, in 'Pesticides: Analytical Requirement for Compliance with EEC Directives (Water Pollution Research Report, No.11)', eds. B. Crathorne, G. Augelletti, Commission of the European Communities, Brussels, 16–34, 1989.
4. J.E. Barbash, G.P. Thelin, D.W. Kolpin, R.J. Gilliom, 'Distribution of Major Herbicides in Ground Water of the United States', Water-resources Investigations Report 98-4245, US Geological Survey, Sacramento, CA, 1999.
5. R.J. Gilliom, J.E. Barbash, D.W. Kolpin, S.J. Larson, *Environ. Sci. Technol.*, **26**, 2440–2447 (1992).
6. *Crop protection Chemicals Reference*, 10th edition, Chemical & Pharmaceutical Press, 1994.
7. R. Frank, G.J. Sirons, *Sci. Total Environ.*, **12**, 223–239 (1979).
8. M. Fielding, D. Barceló, A. Helweg, S. Galassi, L. Torstenson, P. van Zoonen, R. Wolter, G. Angeletti, Water Pollution Research Report, 27, Commission of the European Communities, Brussels, 1–136, 1992.
9. Food Quality Protection Act of 1996, 104th Congress, House of Representatives Report 104 669.
10. Environmental Protection Agency, Atrazine, Simazine and Cyanazine: Notice of Initiation of a Special Review, OPP-300 000-60, November 10, 1994.
11. EEC Drinking Water Guidelines, 80/778/ECC, ECC no. L229/11-29, Brussels, August 30, 1980.
12. 76/464/EEC, Brussels, 1976.
13. Z. Cai, V.M.S. Ramanujam, M.L. Gross, S.J. Monson, D.A. Cassada, R.F. Spalding, *Anal. Chem.*, **66**, 4202–4209 (1994).
14. A. Di Corcia, R. Samperi, A. Marcomini, S. Stelluto, *Anal. Chem.*, **65**, 907–912 (1993).

15. R.N. Lerch, W.W. Donald, *J. Agric. Food Chem.*, **42**, 922–927 (1994).
16. E.M. Thurman, D.A. Goolsby, M.T. Meyer, M.S. Mills, M.L. Pomes, D.W. Kolpin, *Environ. Sci. Technol.*, **26**, 2440–2447 (1992).
17. A.S.Y. Chau, B.K. Afghan, *Analysis of Pesticides in Water*, CRC Press, Boca Raton, FL, Vol. 3, 1982.
18. G. Durand, D. Barceló, *Toxicol. Environ. Chem.*, **25**, 1–7 (1989).
19. AOAC Official Method 991.07 final action 1993. Revision March, 1998.
20. AOAC Official Method 992.14 final action 1997. Revision March, 1998.
21. H. Lee, Y.D. Stokker, *J. Assoc. Off. Anal. Chem.*, **69**(4), 568–572 (1986).
22. T. Suzuki, K. Yaguchi, K. Ohnishi, T. Yamagashi, *J. Chromatogr. A*, **662**, 139–146 (1994).
23. C. Molina, G. Durand, D. Barceló, *J. Chromatogr. A*, **712**, 113–122 (1995).
24. H. Sabik, R. Jeannot, *J. Chromatogr. A*, **818**, 197–207 (1998).
25. C. Crespo, R.M. Marce, F. Borrul, *J. Chromatogr. A*, **670**, 135–144 (1994).
26. D. Barceló, G. Durand, V. Bouvot, M. Nielen, *Environ. Sci. Technol.*, **27**, 217–277 (1993).
27. D. Giraud, A. Ventura, V. Camel, A. Bermond, P. Arpino, *J. Chromatogr. A*, **777**, 115–125 (1997).
28. J.J. Jiménez, J.L. Bernal, Mª.J. del Nozal, J.Mª. Rivera, *J. Chromatogr. A*, **778**, 289–300 (1997).
29. G.A. Junk, J.J. Richard, *Anal. Chem.*, **60**, 451–454 (1988).
30. M. Psathaki, E. Manoussaridou, E.G. Stephanou, *J. Chromatogr. A*, **667**, 241–248 (1994).
31. P. Vitali, E. Venturini, C. Bonora, R. Calori, R. Raffaeli, *J. Chromatogr. A*, **660**, 219–222 (1994).
32. D. Volmer, K. Levsen, G. Wünsh, *J. Chromatogr. A*, **660**, 231–248 (1994).
33. L. Campanella, A. Di Corcia, A. Gambacorta, R. Samperi, *Mater. Chem.*, 429–438 (1982).
34. A. Di Corcia, M. Marchese, R. Samperi, *J. Chromatogr.*, **642**, 163–168 (1993).
35. M. Berg, S.R. Müller, R.P. Schwarzenbach, *Anal. Chem.*, **67**, 1860–1865 (1995).
36. A. Di Corcia, S. Marchetti, R. Samperi, *J. Chromatogr.*, **405**, 357–363 (1987).
37. A. Di Corcia, M. Marchetti, *Environ. Sci. Technol.*, **26**, 66–74 (1992).
38. I. Ferrer, M. Hennion, D. Barceló, *Anal. Chem.*, **69**, 4508–4514 (1997).
39. A. Martín-Esteban, P. Fernadez, C. Cámera, *Anal. Chem.*, **69**, 3267–3271 (1997).
40. V. Pichon, M.C. Hennion, *J. Chromatogr. A*, **665**, 269–281 (1994).
41. V. Pichon, L. Chen, N. Durand, F. Le Goffic, M.C. Hennion, *J. Chromatogr. A*, **725**, 107–119 (1996).
42. V. Pichon, L. Chen, M.C. Hennion, R. Daniel, A. Martel, F. Le Goffic, J. Abian, D. Barceló, *Anal. Chem.*, **67**, 2451–2460 (1995).
43. K. Ensing, T. de Boer, *Trends Anal. Chem.*, **18**, 138–145 (1999).
44. I. Ferrer, D. Barceló, *Trends Anal. Chem.*, **18**, 180–192 (1999).
45. C. Dauwe, B. Sellergreen, *J. Chromatogr. A*, **753**, 191–200 (1996).
46. M.T. Muldoon, L.H. Stanker, *Anal. Chem.*, **69**, 803–808 (1999).
47. M.T. Muldoon, L.H. Stanker, *J. Agric. Food Chem.*, **43**, 1424–1427 (1995).
48. J. Matsui, Y. Miyoshi, O. Doblhoff-Dier, T. Takeuchi, *Anal. Chem.*, **67**, 4404–4408 (1995).
49. M. Siemann, L.I. Andersson, K. Mosbach, *J. Agric. Food Chem.*, **44**, 141–145 (1996).
50. G. Audunsson, *Anal. Chem.*, **58**, 2714–2723 (1986).
51. J.Å. Jönsson, L. Mathiasson, *Trends Anal. Chem.*, **18**, 318–334 (1999).
52. N. Megersa, T. Solomon, J.Å. Jönsson, *J. Chromatogr. A*, **830**, 203–210 (1999).
53. N. Megersa, J.Å. Jönsson, *Analyst*, **123**, 225–231 (1998).
54. M. Knuttson, G. Nilvé, L. Mathiason, J.Å. Jönsson, *J. Chromatogr. A*, **754**, 197–205 (1995).
55. M. Battista, A. Di Corcia, M. Marchetti, *Anal. Chem.*, **61**, 935–939 (1989).
56. M. Battista, A. Di Corcia, M. Marchetti, *J. Chromatogr.*, **454**, 233–242 (1988).
57. A. Laganà, A. Marino, G. Fago, *Chromatographia*, **41**, 178–182 (1995).
58. J.E. Gordon, *J. Chromatogr.*, **18**, 542–555 (1965).
59. W. Funasaka, T. Hanai, K. Fujimura, T. Ando, *J. Chromatogr.*, **72**, 187–191 (1972).
60. C.L. Arthur, J. Pawliszyn, *Anal. Chem.*, **62**, 2145–2148 (1990).
61. R. Eisert, J. Pawliszyn, *Anal. Chem.*, **69**, 3140–3147 (1997).
62. R. Eisert, K. Levsen, *Fresenius, J. Anal. Chem.*, **351**, 555–562 (1995).
63. R. Eisert, K. Levsen, *J. Chromatogr. A*, **737**, 59–65 (1996).
64. R. Eisert, K. Levsen, *J. Am. Soc. Mass Spectrom.*, **6**, 1119–1130 (1995).
65. J. Dugay, C. Miège, M.C. Hennion, *J. Chromatogr.*, **795**, 27–42 (1998).
66. C. Aguilar, S. Peñalver, E. Pocurull, F. Borrul, R.M. Marcè, *J. Chromatogr.*, **795**, 105–115 (1998).
67. A.A. Boy-Boland, S. Magdic, J.B. Pawliszyn, *Analyst*, **121**, 929–938 (1996).
68. S. Lacorte, D. Barceló, *Anal. Chim. Acta*, **296**, 223–234 (1994).

69. J.M. Huen, R. Gillard, A.G. Mayer, B. Baltensperger, H. Kern, *Fresenius' J. Anal. Chem.*, **348**, 606–614 (1994).
70. J. Slobodník, E.R. Brouwer, R.B. Geerdink, W.H. Mulder, H. Lingeman, U.A.Th. Brinkman, *Anal. Chim. Acta*, **268**, 55–56 (1992).
71. B. Bjarnason, L. Chimuka, O. Ramström, *Anal. Chem.*, **71**, 2152–2156 (1999).
72. G.J. Sirons, R. Frank, T. Sawyer, *J. Agric. Food Chem.*, **21**, 1016–1020 (1973).
73. Y. Xu, W. Lrenz, G. Pfister, M. Bahadir, M. Korte, *Fresenius Z. Anal. Chem.*, **325**, 377–380 (1986).
74. M.D. McGlamery, F.W. Slife, H. Butler, *Weeds*, **15**, 35–38 (1967).
75. H.Y. Young, A. Chu, *J. Agric. Food Chem.*, **21**, 711–713 (1973).
76. J. Abian, G. Durand, D. Barceló, *J. Agric. Food Chem.*, **41**, 1264–1273 (1993).
77. *Test Methods for Evaluating Solid Wastes*. 3rd edition, USEPA SW-US GPO, Washington, DC, July, 1995; Method 3540, Update III.
78. N.M.J. Vermeulen, Z. Apostolides, D.J.J. Potgieter, P.C. Nel, N.S.H. Smit, *J. Chromatogr.*, **240**, 247–253 (1982).
79. R. Purkayastha, W.P. Cochrane, *J. Agric. Food Chem.*, **21**, 93–98 (1973).
80. P. Cabras, L. Spanedda, M. Pellecchia, M. Gennari, *J. Chromatogr.*, **472**, 411–415 (1989).
81. C. Molins, E.A. Hogendoorn, H.A.G. Heusinkveld, D.C. van Harten, P. van Zoonen, R.A. Baumann, *Chromatographia*, **43**, 527–532 (1996).
82. T.R. Steinheimer, R.L. Pfeiffer, K.D. Scoggin, *Anal. Chem.*, **66**, 645–650 (1994).
83. R. Robertson, *Environ. Sci. Technol.*, **30**, 1822–1826 (1996).
84. V. Lopez-Avila, N.S. Dodhiwala, W.F. Beckert, *J. Agric. Food Chem.*, **41**, 2035–2041 (1993).
85. S. Papilloud, W. Heardi, *Chromatographia*, **38**, 514–519 (1994).
86. S. Papilloud, W. Heardi, *Chromatographia*, **40**, 705–711 (1995).
87. S. Papilloud, W. Heardi, S. Chiron, D. Barceló, *Environ. Sci. Technol.*, **30**, 1822–1826 (1996).
88. R. Alzaga, G. Durand, D. Barceló, J.M. Bayona, *Chromatographia*, **38**, 502–508 (1994).
89. G.C. Galletti, A. Bonetti, G. Dinelli, *J. Chromatogr.*, **692**, 27–37 (1995).
90. L.Y. Li, D.A. Campbell, P.K. Bennet, J. Henion, *Anal. Chem.*, **66**, 3397–3404 (1996).
91. I.S. Kim, F.I. Sasinos, R.D. Stephens, J. Wang, M.A. Brown, *Anal. Chem.*, **63**, 819–823 (1991).
92. S.B. Hawthorne, Y. Yang, D.J. Miller, *Anal. Chem.*, **66**, 2912–2920 (1994).
93. Y. Yang, S.B. Hawthorne, D.J. Miller, *Environ. Sci. Technol.*, **31**, 430–437 (1997).
94. C. Crescenzi, G. D'Ascenzo, A. Di Corcia, M. Nazzari, S. Marchese, R. Samperi, *Anal. Chem.*, **71**, 2157–2163 (1999).
95. A. Di Corcia, C. Crescenzi, S. Murtas, R. Samperi, A. Barra Caracciolo, G. Giuliano, *Environ. Sci. Technol.*, **33**, 3271–3277 (1999).
96. E. Matisova, J. Krupcik, O. Liska, *J. Chromatogr.*, **169**, 261–269 (1979).
97. E. Matisova, J. Krupcik, O. Liska, N. Szentivanyi, *J. Chromatogr.*, **173**, 139–146 (1979).
98. K. Lekova, L. Kardjieva, N. Hlebarova, V. Natan, *J. Chromatogr.*, **212**, 85–91 (1981).
99. M.S.M. Wells, D.D. Riemer, M.C. Wells-Knecht, *J. Chromatogr. A*, **659**, 337–348 (1994).
100. V. Janda, K. Marha, *J. Chromatogr.*, **329**, 186–188 (1985).
101. M.D. Erickson, C.W. Frank, D.P. Morgan, *J. Agric. Food Chem.*, **27**, 740–743 (1979).
102. EPA Methods 505, 507 and 525.
103. M. Popl, Z. Voznakova, V. Tatar, J. Strnadova, *J. Chromatogr. Sci.*, **21**, 39–42 (1983).
104. R. Bailey, G. Le Bel, A.W. Schwartz, *J. Chromatogr.*, **161**, 251–257 (1978).
105. P.G. Stoks, A.W. Schwartz, *J. Chromatogr.*, **168**, 455–460 (1979).
106. E. Matisova, J. Krupcik, *J. Chromatogr.*, **205**, 464–469 (1981).
107. H. Roseboom, H.A. Herbold, *J. Chromatogr.*, **202**, 431–438 (1980).
108. P.A. Leclercq, V. Pacakova, *J. Chromatogr.*, **178**, 193–207 (1979).
109. V. Lopez-Avila, P. Hirata, S. Kraska, J. Taylor, S.C. Hern, *Anal. Chem.*, **57**, 2797–2801 (1985).
110. Z. Cai, V.M.S. Ramanujam, D.E. Giblin, M.L. Gross, R.F. Spalding, *Anal. Chem.*, **65**, 21–26 (1993).
111. S.L. Werner, M.R. Burkhardt, S.N. De Russeau, US Geological Survey Open-file Report 96–216, Denver, CO, 1996.
112. R.D. Voyksner, J.D. Rosen, (eds.), *Applications of New Mass Spectrometry Techniques in Pesticides Chemistry*, Wiley, New York, 247, 1987.
113. T.A. Bellar, W.L. Budde, *Anal. Chem.*, **60**, 2076–2083 (1988).
114. G. Hopfgartner, K. Bean, J. Henion, R. Henry, *J. Chromatogr. A*, **647**, 51–61 (1993).
115. A. Di Corcia, C. Crescenzi, E. Guerriero, R. Samperi, *Environ. Sci. Technol.*, **31**, 1658–1663 (1997).
116. I. Ferrer, M. Hennion, D. Barceló, *Anal. Chem.*, **69**, 4508–4514 (1997).
117. A.C. Hogenboom, J. Slobodník, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A. Brinkman, *Chromatographia*, **42**, 506–514 (1996).
118. M.L. Pomes, E.M. Thurman, D.S. Aga, D.A. Goolsby, *Environ. Sci. Technol.*, **32**, 163–168 (1998).

119. E.M. Thurman, M.T. Meyer, M.L. Pomes, C.A. Perry, A.P. Schwab, *Anal. Chem.*, **62**, 2043–2048 (1990).
120. P.L. Del Valle, M.T. Muldoon, J.S. Karns, J.O. Nelson, W.W. Mulbry, *J. Agric. Food Chem.*, **44**, 945–952 (1996).
121. G.J. Reimer, J.G. Shirley, B.D. Hammock, *J. Agric. Food Chem.*, **46**, 3353–3358 (1998).
122. C.S. Hottenstein, F.M. Rubio, D.P. Herzog, J.R. Flleker, T.S. Lawruk, *J. Agric. Food Chem.*, **44**, 3576–3581 (1996).
123. D.S. Aga, E.M. Thurman, *Anal. Chem.*, **65**, 2894–2898 (1993).
124. AOAC Official Method No. 995.08; 10.3.06, first action, 1995.
125. C. Wittmann, U. Bilitewski, T. Giersch, U. Kettling, R.D. Schmid, *Analyst*, **121**, 863–869 (1996).
126. L. Mosiello, C. Cremosini, L. Segre, S. Chiavarini, M. Spanò, T. Kimmel, A.J. Bäumner, R.D. Schmid, *J. Agric. Food Chem.*, **46**, 3847–3851 (1998).
127. M.A. Gonzáles-Martinez, R. Puchades, A. Maquira, *Trends Anal. Chem.*, **18**, 204–218 (1999).
128. K. Yano, I. Karube, *Trends Anal. Chem.*, **18**, 199–204 (1999).
129. S.A. Piletsky, E.V. Piletskaya, A.V. Elgersma, K. Yano, I. Karube, Y.P. Parhometz, A.V. El'skaya, *Biosens. Bioelectron.*, **10**, 959–964 (1995).
130. S.A. Piletsky, E.V. Piletskaya, A.V. El'skaya, R. Levi, K. Yano, I. Karube, *Anal. Lett.*, **30**, 445–455 (1997).
131. H.R. Buser, *Environ. Sci. Technol.*, **24**, 1049–1058 (1990).
132. T.D. Bucheli, S.R. Müller, S. Heberle, R.P. Schwarzenbach, *Environ. Sci. Technol.*, **32**, 3457–3464 (1998).
133. S.R. Müller, M. Berg, M.M. Ulrich, R.P. Schwarzenbach, *Environ. Sci. Technol.*, **31**, 2104–2113 (1997).
134. M. Ahel, K.M. Evans, T.M. Fileman, R.C.F. Mantoura, *Anal. Chim. Acta*, **268**, 195–204 (1992).
135. G. Durand, V. Bouvot, D. Barceló, *J. Chromatogr.*, **607**, 319–327 (1992).
136. U. Dörfler, E.A. Feicht, I. Scheunert, *Chemosphere*, **35**, 99–106 (1997).
137. N. Shapir, R.T. Mandelbaum, C.S. Jacobsen, *Environ. Sci. Technol.*, **32**, 3789–3792 (1998).
138. L.J.T. Van der Pas, M. Leistra, J.J.T.I. Boesten, *Pestic. Sci.*, **53**, 223–232 (1998).
139. L.H. Bodalbhai, R.A. Yokley, M.W. Cheung, *J. Agric. Food Chem.*, **46**, 161–167 (1998).
140. T.D. Bucheli, F.C. Grübler, S.R. Müller, R.P. Schwarzenbach, *Anal. Chem.*, **69**, 1569–1576 (1997).
141. Z. Cai, R.L. Cerny, R.F. Spalding, *J. Chromatogr. A*, **753**, 243–251 (1996).
142. C. Crescenzi, A. Di Corcia, E. Guerriero, R. Samperi, *Environ. Sci. Technol.*, **31**, 479–488 (1997).
143. C. Crescenzi, A. Di Corcia, M.D. Madbouly, R. Samperi, *Environ. Sci. Technol.*, **29**, 2185–2190 (1995).
144. C. Crescenzi, A. Di Corcia, G. Passariello, R. Samperi, M.I. Turnes Carou, *J. Chromatogr. A*, **733**, 41–55 (1996).
145. A. Dankwardt, B. Hock, S. Rupert, D. Freitag, A. Kettrup, *Environ. Sci. Technol.*, **30**, 3493–3500 (1997).
146. P. Dufek, V. Pacakova, *J. Chromatogr.*, **187**, 341–349 (1980).
147. P. Dufek, V. Pacakova, E. Tesarova, *J. Chromatogr.*, **191**, 115–120 (1980).
148. M.D. Erickson, C.W. Frank, D.P. Morgan, *J. Agric. Food Chem.*, **27**, 740–743 (1979).
149. I. Ferrer, D. Barceló, E.M. Thurman, *Anal. Chem.*, **71**, 1009–1015 (1999).
150. R. Frassanito, G. De Socio, D. Laura, D. Rotilio, *J. Agric. Food Chem.*, **44**, 2282–2286 (1996).
151. J. Gan, S.K. Papiernik, W.C. Koskinev, S.R. Yates, *Environ. Sci. Technol.*, **33**, 3249–3253 (1999).
152. F. Hernandez, C. Hidalgo, J.V. Sancho, F.J. Lopez, *Anal. Chem.*, **70**, 3322–3328 (1998).
153. R.N. Lerch, E.M. Thurman, E.L. Kruger, *Environ. Sci. Technol.*, **31**, 1539–1546 (1997).
154. V. Lopez-Avila, R. Young, N. Teplinsky, *J. Assoc. Off. Anal. Chem.*, **79**, 142–156 (1996).
155. M.S. Mills, E.M. Thurman, *Anal. Chem.*, **64**, 1985–1990 (1992).
156. M.S. Mills, E.M. Thurman, *Environ. Sci. Technol.*, **28**, 73–79 (1994).
157. C. Mouvet, L. Amalric, S. Broussard, G. Lang, A. Brecht, G. Gauglitz, *Environ. Sci. Technol.*, **30**(6), 1846–1851 (1996).
158. C. Siontorou, D.P. Nikolelis, U.J. Krull, K. Chiang, *Anal. Chem.*, **69**, 3109–3114 (1997).
159. T.R. Steinheimer, *J. Agric. Food Chem.*, **41**, 588–595 (1993).
160. V. Pichon, M. Bouzige, C. Miege, M.C. Hennion, *Trends Anal. Chem.*, **18**, 219–235 (1999).
161. N. Megersa, T. Soloman, J.Å. Jönsson, *J. Chromatogr. A.*, **830**, 203–210 (1999).
162. A. Di Corcia, ''Pesticides: Trace Analysis'', in *Encyclopedia of Environmental Pollution and Cleanup*, ed. R.A. Meyers, New York, Wiley–Interscience, Vol. 6, 3555–3598, 1999.
163. A.C. Hogenboom, J. Slobodník, J.J. Vreuls, J.A. Rontree, B.L.M. van Baar, W.M.A. Niessen, U.A. Brinkman, *Chromatographia*, **42**, 506–514 (1996).

Hydrocarbons Analysis: Introduction

Peter A. Nick
Process Systems Engineering, Yorba Linda, USA

The progress of inorganic chemical analysis technology has historically been dependent upon profit motives. Medieval alchemy was driven by the need to turn base materials into treasure. During the Enlightenment, scientific chemistry work was somewhat less commercial but still typically funded by a patron intent on deriving some specific benefit. Often the true benefit(s) discovered was not what was originally intended or expected. This is a normal occurrence in science and is what keeps most of us motivated to continue to work and study in this field. Still, economic justifications are the historic motivators for most practical advances. The Industrial Age, for example, demanded the implementation of efficient ore discovery, analysis, and conversion to metal for construction and fabrication purposes. Many of the basic quantitative wet chemistry methods owe their genesis to the late 19th century when large-scale mineral mining efforts became a matter of routine and chemical analysis of the ore a part of the process.

Another significant area of fundamental chemistry also has its root in commercial opportunism. Read any organic chemistry text and one finds that the first half of the book will discuss the wide and varied chemistry of the hydrocarbon – particularly those of petroleum-derived origin.

Although originally thought of as a medicinal base, wick-lamp fuel, or coating material, crude oil and its derivatives were completely unappreciated until two somewhat independent events occurred. First, a vast array of crude oil and gas hydrocarbon reservoirs were found in the time period of the mid-1800s (Pennsylvania) through the first decade of the 20th century (east Texas and other locations). The main products distilled from this oil were fairly low-volume and low-profit retail products for consumer consumption, of the types noted above. The gas associated with the oil was sometimes burned in municipal lighting facilities but only if it did not have to be piped very far from the source. As a result, it was often flared and burned for disposal. The big nuisance product of early petroleum processing consisted of a lighter-than-water, clear liquid phase, called petroleum naphtha or ''white oil'', which had a disturbing tendency to catch fire too easily and even explode violently under certain conditions. It made a wonderful solvent but was far too dangerous to be used in any significant amount by consumers and manufacturers. It usually was burned, like the gas, for disposal and was worth next to nothing, its supply far outweighing demand.

The second key development involved a technology which when commercialized would revolutionize the world and provide one of the biggest economic development engines in the history of mankind. The combustion engine, in its varied forms, was this invention, specifically the higher pressure petroleum naphtha injection-fueled version we know today. Initial engine schemes for mechanical land transport developed independently of the oil product gluts. They would utilize any type of fuel such as wood, bagasse, coal, or nearly anything else that could burn. Later versions used the heavier kerosene and gas oil streams from crude oil refining as these were reasonably abundant and relatively safe to handle. For several years overall demand for these first ''mechanical horses'' was limited and competition among engine types flourished.

It was when automobile makers discovered that high-compression ignition engines optimally combusted this cheap abundant material that events moved rapidly. With a growing worldwide consumer class wanting the freedom of personal transportation and two world wars demonstrating the military value of this engine technology, crude oil was soon being produced solely for its naphtha value, with the more traditional products being relegated to side-stream status. More and more research effort was put into understanding the chemistry of fuel hydrocarbons and finding ways of increasing this naphtha material production. For geographical areas deficient in petroleum, Europe and Germany in particular, derivation of naphtha-like material from much more abundant coal resources proceeded on a priority basis. The wartime fuel production efforts, especially the work to discover a means to synthesize rubber from oil, catalyzed the analysis work in all areas of hydrocarbon processing. The global economic booms in the latter half of the 20th century provided even more extreme incentives for the petroleum refining market to grow. Refinery process needs for base fuel production and requirements for increasingly stringent fuel performance specifications made analysis of these pure hydrocarbon systems a priority industrial activity. A host of derivative petrochemical and oxygenated organic chemical industries sprang up, in turn, from the roots of petroleum refining. Analytical procedures for each industry started from those used in the laboratories of the refining plants that fed them and were quickly extended into their own specific domains of need. Many of the newest analytical techniques still rely on this legacy of data to extend and validate the measurement results they produce today.

The vast portion of this technology growth occurred in the 30 years after the Second World War. During this

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

''golden age'' of refining, a massive knowledge bank of hydrocarbon chemistry information and analytical techniques was formed, much of which has been reported in the public domain. The detail with which we know pure hydrocarbon organic chemistry is due largely to the incentives generated by this unique set of historical and economic conditions. Today, this boom-time research and development activity appears to be at work in the biopharmaceutical industries.

The industrial hydrocarbon production and processing sector is now considered a mature entity. It is in commodity-production mode with low relative margins but its products are still demanded in huge quantities. Like all such low-profit (taxes are the biggest pump cost, not refinery margin, for those readers who dispute this contention), high-volume industries, cost reduction becomes a major business strategy and innovation takes a back seat to the needs of meeting revenue-production quotas. Downsizing and mergers of key industrial participants have greatly reduced the number and funding of scientists engaged with primary research into hydrocarbon systems. This malaise has affected both the industrial and academic sectors. Oil is no longer ''sexy'', as it were.

A new concern – and cost generator – is that of environmental remediation. Once the public domain was thought of as a free sink in which to put wastes. As a society, we have made waste minimization and elimination just another cost of doing business – with public health issues and the attendant costs of such rightly being of considerable concern in the value-accounting process. The analytical needs of this ''green'' technology protocol are being driven by cost factors once again, although the current incentives are essentially negative ones (do it or else get fined!).

At the same time, analytical reporting accuracy needs went from parts per million levels at the extremes of analysis to our current parts per billion and trillion exposure levels. Many of the initial techniques regulators and industry used for environmental analytical work were developed by refiners during the ''golden years'' alluded to above. Improvements to these methods, particularly for wet chemistry and heavy hydrocarbon analysis, were inevitable and today a whole new discipline of environmental analytical chemistry has evolved. This encyclopedia has a major subsection devoted to the discussion of environmental and regulatory analyses of industrial wastes and the processes that produce them. Examination of these volumes by readers would provide an excellent auxiliary source of analytical techniques to complement what is presented herein. In fact, once you have absorbed what is presented in this volume, examine the remainder of the encyclopedia. Many of those techniques can be and have been successfully used in laboratory work dealing with any organic material.

While mature, the hydrocarbon production and refining sector, particularly the fuels sector, is hardly dead, despite press reports to the contrary. The unique chemical paradigm of crude oil and traditional products, i.e. hundreds of different species distributed in a wide range of molecular size, has long been a constraining factor on the type and complexity of petroleum and petroleum product analyses. Traditional analyses usually measure quantities specified in terms of some easily defined inspection parameters, boiling point and gravity being the most common.

Most newer methods of analysis that can define chemical species concentrations of the order of parts per trillion must take advantage of higher resolution analytical instruments that the computer and embedded control chip revolution has spawned. The coupling of the primary measurement hardware with powerful data and calculation servers provides a wealth of information about any material quickly. This new technology has allowed commercial hydrocarbon stocks now to be speciated and type-categorized down to the heaviest reaches of the molecular weight distribution curve. This, in turn, has opened up new vistas for the production facilities to monitor and control the processing of many potentially hazardous compounds, either to limit their production or to replace them in the process altogether. In a turn of poetic justice, high-performance liquid chromatography (HPLC) techniques originally developed for regulatory analysis of aromatics in wastewater are now used to quantify the same species in refinery conversion units such as catalytic cracking and naphtha reforming.

This symbiotic approach across industrial areas is a recurring theme in all sectors of the chemical process industries, not just the hydrocarbon sector. Knowledge begets knowledge, as one saying goes; thus we owe a debt of gratitude to those who expanded this once nascent field of knowledge into the vast array of analysis techniques available to chemists today. Just read the contents of this encyclopedia to see how much progress has been made. One cannot hope to cover the wide array of methods and techniques in use or in research today. We have tried to retain a representative sample in this volume of what currently seems to be valuable to the hydrocarbon production and refined product industry. If there is a bias towards fuel production, it is because this is still a huge economic factor in our world and will continue to be so long after the pages of the volume you read here have disintegrated into pieces. Economics still drives the development of analytical chemistry technology. Thus the reader will find the present emphasis on petroleum-based analytical environments. Refiners still pay most of the bills!

By way of categorizing the contents of this section, we include five articles on specific methodologies as applied more to product analyses of various hydrocarbon intermediate or product cuts in general.

Dr Klawun's **Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis** extends the discussion of gas chromatography (GC) methodology to the analysis of hydrocarbon mixtures originating from mineral oil sources as well as providing an overview of GC and detectors, including hybrid instrumentation.

The article **Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices**, by Drs Nero and Drinkwater, focuses on some currently nonstandard methods which can be used to determine the composition of individual components or classes of components within various petroleum matrices. GC coupled with low-resolution mass spectrometry – gas chromatography/mass spectrometry (GC/MS) – is the most useful method for *quickly* identifying and precisely measuring specific components or classes of components in a petroleum matrix. Direct insertion probe tandem mass spectrometry (DIP/MS/MS), used for less volatile and thermally labile petroleum components, is covered is some detail.

A ''one-stop'' laboratory methodology is presented by Dr S.E. Scheppele using similar but more sophisticated and ''intelligent'' technology. His article entitled **Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams** describes what may well be the preferred methodology of the future. Several practitioners of this type of analytical system currently exist. Most are developing their homolog databanks in confidentiality. Dr Scheppele is to be commended for his willingness to discuss the University of Illinois efforts here in our compendium.

Dr Mizaikoff's contribution on **Fuel Performance Specifications, Mid-infrared Analysis of** covers an old technology with a fairly new spectrophotometic frequency range – at least as applied to fuels analysis. Most spectrophotometric work with hydrocarbons today covers the near-infrared frequency range. It is often adequate for quick or even ''on-line'' results but usually not too accurate unless repeatedly recalibrated or used in conjunction with one or more alternative analyses such as GC or nuclear magnetic resonance (NMR) spectroscopy. The authors here present a case otherwise – at least for the mid-infrared spectrum methods.

A current ''hot topic'' for refinery process control is use of the near-infrared spectrophotometric range analyses. They are currently proving useful in control of octane for gasoline production and blending processes, primarily because they capture the spectral nuances of chain branching fairly well. Chain branching has a strong link to preignition tendencies (octane) in a naphtha fuel. The technology, by itself, has been successful for other analyses, too. Dr Baughman's article **Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels** will elaborate in this still-developing technology.

Finally, **Nuclear Magnetic Resonance Characterization of Petroleum** by Dr F.P. Miknis provides a detailed overview of this old but sophisicated technology as applied to petroleum hydrocarbons.

If the reader is curious about any analysis methodology, several articles in general are available in this encyclopedia and the instrument vendors will always be willing to supply information that may be of use.

On the theory that old is not necessarily bad, this Section Editor has composed a piece involving the utilization of the more classical petroleum analysis methods. As Dr Nero states so well in his Introduction, ''Generally accepted industrials analytical procedures do not adequately define chemical composition of various petroleum products and mixtures, but these methods provide little or no information concerning specific chemical components. These industrial procedures only determine the overall physical or chemical characteristics of a petroleum product, which are necessary to market the product. Physical properties such as boiling range, density, viscosity, flash point and vapor pressures are determined. General chemical characteristics such as carbon, hydrogen, and nitrogen content, total sulfur, average molecular weight, octane number and cetane number are also determined''.

The Editor heartily agrees but these are precisely the values that commercial petroleum refiners, particularly for fuel production facilities, are most concerned about right now. They control their processes to these items and must meet certain product specifications, again involving these same quantities. The amount of quality historical data available for estimation and cross-correlation of these properties is impressive and is still being added to. Since these procedures will not die soon, we decided to discuss them for some measure of topical completeness.

Having covered some of the more important analytical techniques necessary for ''adequate'' characterization, our remaining articles cover the specific analytical needs of various feed and product streams themselves.

Two articles deal with the analysis of full-range crudes. Dr M.A. Ali's contribution, **Full Range Crudes, Analytical Methodology of**, covers the spectrum of classical and the latest methodologies for characterizing and speciating broad boiling range hydrocarbons. The article entitled **High-temperature Simulated Distillation Applications in Petroleum Characterization** by D.C. Villalanti et al. describes one of the most important tools for such analyses. (O.K., if you are a purist, this article might better have been grouped with those mentioned above but

it is an increasingly utilized technique in crude oil assay methods.)

Dr C.S. Hsu's contribution, **Diesel Fuels Analysis** describes physical properties and chemical composition of diesel fuels. It addresses various issues including performance, storage stability, evaluation of refinery process efficiency and product quality, and assessment of environmental impact. Various methods of physical property measurement, performance tests, contaminant analyses, and analytical techniques for chemical composition are described.

The Editor has reprised and updated an earlier Wiley encyclopedia covering similar analytical considerations required for satisfying mandated regulatory analysis of transportation fuels. With the main emphasis of this latter article being on light naphtha (gasoline) stocks, the previous article provides a good balance for our coverage of transport fuels. It was somewhat difficult to be authoritative on actual requirements since little new legislation has been formally enacted since the mid-1990s. Many refiners are just now gearing up to meet these fiats. Much is currently being discussed, however, and we try to treat these ''suggestions'' as what probably will be the driving legislation to come within the next few years, perhaps even by the time this volume is published.

Dr Stanley Sandler is a world-renowned thermodynamicist and his discussion of oxygenate vapor–liquid equilibrium analysis, **Oxygenate Vapor–Liquid Equilibrium in Gasolines**, could not be adequately covered in any other way but directly from the source himself. In his article, he reviews the equipment and measurement methods commonly used to study the vapor–liquid equilibrium of hydrocarbon–oxygenate and gasoline–oxygenate mixtures. Detailed tables indicate those oxygenate–hydrocarbon mixtures for which vapor–liquid equilibrium data are available.

Dr W.K. Robbins' **Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of** considers contaminants in the residual fraction drawn off the bottom of towers or flash drums. These residues have an increased concentration of minor impurities, including sulfur-, nitrogen-, and oxygen-bearing heterocyclics, soluble nickel and vanadium compounds, and a variety of trace metals. These impurities both dominate the unique properties of the residue and play a major role in determining the refining processes necessary for its conversion to marketable products. This article introduces the common analytical techniques for the measurement of sulfur, nitrogen, nickel, vanadium, and trace metals in this challenging organic matrix.

Dr Bacaud's **Petroleum Residues, Characterization of** deals with the analytical classification of speciation types and prediction of properties of residue streams in general. Distinctions are drawn between atmospheric and vacuum residues. Since the simplest, lighter constituents of petroleum have been separated by distillation, residues concentrate the most complex fraction of crude oil and a detailed inventory of the individual compounds is impossible. Thus characterization methods will aim at giving an adequate description of the types of compounds contained therein.

A lubricant is a substance that reduces friction and wear between surfaces. Base oils typically comprise 80–99% of the finished lubricant. They are refined from selected crude oils, re-refined from used oil, synthesized from specific organic compounds, or extracted from renewable resources, such as seed oils. Base oils formulated into finished lubricants help to reduce friction and wear between moving surfaces, remove heat, seal, cushion against shock loads, and remove abrasives and contaminants from the lubricated area. Because of the special handling and processing required for lubricant stocks, we have three articles dealing with their analysis and characterization.

Dr Shugarman's article, **Lubricant Base Oils: Analysis and Characterization of**, is concerned with the lubricant base oils, or base stock. These stocks are combined with performance additives to make finished industrial and automotive lubricants. Base oil analysis and characterization is important for understanding how the physical and chemical properties of these materials relate to performance in finished lubricants. It is also essential for maintaining quality control over the manufacturing process, quality assurance of the finished product, and for setting specifications between buyer and seller.

Lube Products. Molecular Characterization of Base Oils, by Dr Montanari, covers the other end of lube technology. Finished oil considerations include final product specifications which are nearly as varied as those of transport fuel. The author shares some of his career's work in this area.

Dr J.G. Groetsch has written **Refractive Index Technology as a Real Time Viscosity Technique**, which describes the use of an online digital process refractometer to determine the viscosity index of lube oil stocks. Since lube oils are Newtonian fluids, refractive index can be related to viscosity index. Real-time continuous measurement of the lube stock refractive index has eliminated sampling, improved product consistency, reduced waste, and helped to optimize the process.

Oil shale is an alternative source of liquid hydrocarbons. The organic matter in an oil shale is mostly in the form of kerogen, which is defined as that fraction of the organic matter in a sedimentary rock that is insoluble in common petroleum solvents. The insolubility of kerogen is the main reason why an oil shale must be heated to produce liquid products. As with petroleum, standard assays do not provide information

about what chemical properties of oil shales are important for producing liquids. Consequently, a number of alternative procedures have been developed to analyze oil shales. These procedures, along with methods of analyzing shale oils, are described in the article **Oil Shale and Shale Oil Analysis**. The Editor feels that most of these techniques would be appropriate, along with aromatics analysis techniques presented elsewhere in this volume, for the characterization of coal-derived liquid hydrocarbons.

With that, we have covered pretty much the whole range of hydrocarbons. Some of these methods work well for any multicomponent organic liquid but this is something for you the reader to decide. We hope we have been able to present the current state of hydrocarbon analysis in its many flavors.

ABBREVIATIONS AND ACRONYMS

DIP/MS/MS	Direct Insertion Probe Tandem Mass Spectrometry
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
NMR	Nuclear Magnetic Resonance

Diesel Fuels Analysis

Chang S. Hsu
Exxon Research & Engineering Co., Annandale, NJ, USA

1 Introduction 1
2 Methods for Physical Properties Measurement 2
2.1 Boiling Range 2
2.2 Flash Point 2
2.3 Low-temperature Flow Characteristics 2
2.4 Density/Relative Density (or American Petroleum Institute Gravity) 2
2.5 Viscosity 3
2.6 Aniline Point 3
2.7 Color 3
3 Performance Tests 3
3.1 Cetane Number 3
3.2 Cetane Index 3
3.3 Heating Value 4
3.4 Lubricity 4
3.5 Corrosivity 4
3.6 Stability 4
4 Analyses of Fuel Contaminants 4
4.1 Water and Sediment 5
4.2 Ash Content 5
4.3 Carbon Residue 5
5 Analytical Techniques for Chemical Composition 5
5.1 Elemental Analysis 5
5.2 Chromatography 5
5.3 Mass Spectrometry 7
5.4 Nuclear Magnetic Resonance and Molecular Spectroscopies 7
6 Molecular Composition of Diesel Fuels 7
Acknowledgments 8
Abbreviations and Acronyms 8
Related Articles 8
References 9

Diesel fuels are analyzed for their physical properties and chemical composition to address various issues including their performance, storage stability, evaluation of refinery process efficiency and product quality, and assessment of environmental impact. Various methods of physical property measurement, performance tests, contaminant analyses, and analytical techniques for chemical composition are described.

1 INTRODUCTION

Diesel fuels are the fuels used in diesel or compression ignition engines. In diesel engines, the air is compressed to high temperature before the fuel is injected into the cylinder to ignite or detonate. Compared to spark ignition engines, a diesel engine is cost effective because of its operating advantages, greater efficiency, high power output, and fuel economy under all loads. However, it has disadvantages of noise and emissions of particulates and nitrogen oxides (NO_x). Three grades of diesel fuels are in common use: (1) land diesel fuels, used in trucks, buses, trains, or other land transportation vehicles that have high variation of speed and load; (2) marine diesel fuels, used in ships that have variable speed but relatively high and uniform load; and (3) plant diesel fuels, used in electric power generation plants that have low or medium speed with heavy load. Hence, the quality of the diesel fuels depends on the performance requirements of the engines.

Today, the majority of diesel fuels are derived from petroleum, although some may come from other fossil fuel resources such as coal liquefaction fractions and through synthesis, such as Fischer–Tropsch distillates. The quality of diesel fuels can vary significantly owing to differences in the crude oils from which the fuels are derived and the refinery process streams from which they are blended. For example, there are significant differences in composition between straight run and hydrotreated cracked diesel streams. Diesel fuels taken from refinery streams for land transportation are the lightest and narrowest distillation cuts, while those for power plants are the heaviest and broadest cuts (fuel oils). The plant diesel fuels also contain the highest concentrations of sulfur and other heteroatoms.

The main sources of diesel fuels from refinery streams are straight run middle distillates from atmospheric and vacuum pipestills (distillation towers). Other streams that may be blended into the diesel pool include heavy cracker naphtha/oils and light coker oils. These streams are often hydrotreated to improve stability and to reduce sulfur and aromatics content. The most important properties of refinery streams for diesel blending include cetane number, initial boiling point (IBP), 95% boiling point, American Petroleum Institute (API) gravity (density), and viscosity. Cetane number is governed by industrial specifications. Among refinery streams that are blended into diesel pools, virgin middle

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

distillates and hydrocracker bottoms have the highest cetane number (35–60). Light catalytic cracker oil (LCCO) has the lowest cetane number (<20) and is generally blended with high cetane number streams to reach a cetane number >40 for land and marine (transportation) diesel fuels and >30 for plant (industrial) diesel fuels.

Diesel fuels are analyzed for their physical properties and chemical composition to address various issues including their performance (combustion efficiency), storage stability, evaluation of refining process efficiency and product quality, and assessment of environmental impact.

2 METHODS FOR PHYSICAL PROPERTIES MEASUREMENT

Physical properties of diesel fuels are measured to determine the ease of volatilization and flow in fuel lines and the storage stability of the fuels. The properties to be measured generally include boiling point, flash point, cloud point, pour point, density, viscosity, and color.

There are many established methods for measuring physical properties of diesel fuels. The American Society for Testing and Materials (ASTM) and the Institute of Petroleum (IP) have published standard methods for such measurements.[1,2] Some of the important tests that are generally included in fuel specifications are briefly discussed below.

2.1 Boiling Range

Boiling range (or distillation limits) is measured to ensure that fuel products have suitable volatility. The boiling range is defined as the temperatures between the IBP and the final boiling point (FBP). The temperature at which 95% of the fuel has evaporated (T95) is frequently specified. The boiling range of diesel fuels can be obtained by the low efficiency ASTM D 86 (IP 123) test method or by ASTM D 2892, which gives true boiling point measurements using a 15-theoretical plate column. However, both ASTM D 86 and D 2892 require a rather large volume of sample. Boiling range distributions (distillation curves) of diesel fuels have therefore been more conveniently determined using gas chromatography (GC), ASTM D 2887. This so-called ''GC simulated distillation'' method uses a GC column coated with nonpolar methyl silicone as a stationary phase. After calibration, the retention times of the components are essentially equivalent to the true boiling points determined by ASTM D 2892. The boiling range for typical land transportation diesel fuels is between 350 °F (177 °C) and 650 °F (343 °C), although for some diesel fuels the IBP and/or FBP are slightly out of this range.

2.2 Flash Point

Flash point is used to determine the temperature at which the vapor space above the fuel can be ignited by a spark or other ignition source. The flash point of a diesel fuel can be determined by ASTM D 93 (IP 34) using a Pensky–Martens closed cup tester or by ASTM D 56 using a Tag closed tester. In the test, the sample is placed in a cup of the tester with the lid closed. The lowest temperature at which a small flame of a specified size causes the vapor above the sample to ignite is recorded as the flash point. Flash points of transportation diesel fuels are generally >40 °C. Fuel oils used for low- and medium-speed engines have a minimum flash point of 55 °C.

2.3 Low-temperature Flow Characteristics

Either ASTM D 2500 (IP 219) or D 3117 test can be used to determine the temperature at which a cloud of wax crystals first appears in fuel when it is cooled. The temperature is recorded as cloud point. The cloud point of a diesel fuel is typically +10 to −40 °C.

The lowest temperature at which a fuel stays fluid is recorded as its pour point, as determined by ASTM D 97 (IP 15). A typical diesel fuel has a pour point between −10 and −40 °C.

There are a number of filterability tests. In some cases they supplement the measurements of pour point and cloud point, and are particularly useful for evaluating fuels containing flow improver additives. In the ASTM D 4539 test, the temperature of the test fuel is lowered at a controlled cooling rate. The lowest temperature at which a fixed volume of the fuel can be filtered through a specific mesh (17 μm) of screen in less than a given period of time (60 s) is recorded as minimum low-temperature flow test (LTFT) pass temperature. In Europe, the standard test method for cold-flow performance of diesel fuels is the cold filter plugging point (CFPP), designated as EN116 (identical with IP 309).[2,3] CFPP is the highest temperature at which a given volume of fuel fails to pass through a standardized wire mesh filter under a controlled vacuum in a specified time, when cooled under standardized conditions.

Cloud point, LTFT, and CFPP are measured for vehicle operability. Pour point is for fuel distribution system in the vehicle.

2.4 Density/Relative Density (or American Petroleum Institute Gravity)

There are two standard methods adopted by ASTM for density measurement. The D 1298 (IP 160) method uses

a hydrometer that is immersed in a test fuel at a specific temperature, and the D 4052 (IP 365) method uses a digital analyzer consisting of a U-shaped, oscillating sample tube and a system for electronic excitation, frequency counting, and display. The relative density (specific gravity) at 60 °F can also be represented by API gravity.

$$\text{API gravity, deg} = \frac{141.5}{\text{specific gravity at } 60\,^{\circ}\text{F}} - 131.5$$

Typical densities of transportation diesel fuels at 15 °C (60 °F) are 0.80–0.86 g mL^{-1}, or 33–45° API gravity.

2.5 Viscosity

The kinematic viscosity of diesel fuels is important to their effective use, e.g. for flow of fuels through fuel lines, injection nozzles, and orifices. In the ASTM D 445 (IP 71) test, a fixed volume of fuel flows through a calibrated viscometer under a reproducible driving head and at a closely controlled temperature. The kinematic viscosity is the product of measured flow time and the calibration constant of the visocometer. Dynamic viscosity is obtained by multiplying the measured kinematic viscosity by the density of the fuel. The kinematic viscosity range of transportation diesel fuels at 40 °C is 1.3–4.1 centistokes (cSt or $mm^2\,s^{-1}$) with most being between 1.9 and 4.1 cSt. The kinematic viscosity of fuels used in heavy-duty low-speed engines is in the range 5.5–24 cSt.

2.6 Aniline Point

Aniline point is not generally used any more for estimating the aromatics content of a fuel. To determine the aniline point, a specific volume of aniline is mixed with the fuel, as in the ASTM D 611 (IP 2) test method. The mixture is cooled at a controlled rate. The temperature at which two phases separate is recorded as the aniline point. Aniline points are high for paraffins and low for aromatics with cycloparaffins and olefins in between. The aniline point has been found to be highly correlated with the cetane number of diesel fuels.

2.7 Color

Color is used as an indication of fuel quality. A variation outside the established range may indicate degradation or contamination of the fuel with other products. It is not, however, a reliable predictor of stability. A scale ranging from 0.5 to 8.0 is used in the ASTM D 1500 (or IP 196) test, which employs a standard light source and a comparison of colored glass disks.

3 PERFORMANCE TESTS

One of the most important tests for diesel fuel performance is cetane number, a measurement of the ignition quality of fuels in a diesel engine. For usage and storage, standard tests for heating value, lubricity, filterability, corrosivity, and oxidation stability of the fuel have also been established.

3.1 Cetane Number

The cetane number of a fuel is measured on a single-cylinder rating engine by comparing the fuel performance with blends of *n*-hexadecane and α-methylnaphthalene or heptamethylnonane. A diesel fuel with cetane number 55, for example, matches the performance of a blend of 55% of *n*-hexadecane and 45% of α-methylnaphthalene in the cetane engine.

Methylene groups (CH_2) not incorporated in naphthenic rings and more than three carbon atoms from any other structures (such as aromatic rings, methyl groups, double bonds, or naphthenic rings) are the single most important molecular features for cetane number. The molecules with highest cetane number are the straight-chain normal paraffins. The molecules with lowest cetane number are those having few methylene groups. The presence of double bonds in molecules will lower the cetane number. Hence, the general trend of cetane number among the molecular types is *n*-paraffins > isoparaffins > cycloparaffins > aromatics. Within a homologous series, for example the paraffins, the cetane number increases with molecular weight.

Typical cetane number ranges of refinery process streams are: virgin distillate streams–light (jet), 35–46, mid (diesel), 35–60, heavy (light virgin gas oil), 46–56; LCCO, 14; heavy coker naphtha (HKN), 37; light coke gas oil (LKGO), 40–45; and hydrocracker fractionator bottoms, 42–58. These streams may be blended into diesel fuel pools.

3.2 Cetane Index

Various correlations between cetane number and other measured properties are in use. Frequently these correlations are proprietary or limited to restricted groups of diesel fuels. For the determination of cetane number, an engine performance test uses a large volume of fuel over a rather long period of time. In addition, the engine requires extensive and continuous maintenance to assure the quality of cetane ratings. Hence, ASTM has developed two cetane indices that refineries generally use for estimating the cetane number of diesel fuels. In ASTM D 976 (IP 264), the cetane index is calculated from API gravity and ASTM D 86 mid boiling point (T50). An

alternative method for estimating cetane number of a transportation diesel fuel is the ASTM D 4737 (IP 380), a four-variable method that uses density and 10% (T10), 50% (T50), and 90% (T90) recovery temperatures of the fuel as variables.

Cetane index does not reflect the use of cetane improver, but correlates with the aromatics of the fuel.

3.3 Heating Value

The heating value or energy content of diesel fuel affects fuel economy and power output. Standard methods for heating value measurement include ASTM D 4809 for transportation diesel fuels and ASTM D 4868 for plant diesel fuels. Typical heating value for land diesel fuels is 134 000 Btu gal^{-1} (or 19 700 Btu lb^{-1}); for marine diesel fuels, 138 000 Btu gal^{-1}; and for plant diesel fuels, 148 000 Btu gal^{-1}.

3.4 Lubricity

Lubricity of diesel fuel can affect the life of engine components such as fuel injection pumps. There are many methods for evaluating diesel lubricity. ASTM has standardized two methods: high-frequency reciprocating rig (HFRR) as ASTM D 6079 and scuffing load ball on cylinder lubricity evaluator (SBOCLE) as ASTM D 6078. The HFRR method has also been standardized by ISO (International Organization for Standardization) as CEC F-06-A-96. No absolute correlation has been developed between these two methods or between these methods and field performance.

In the ASTM D 6079 (HFRR) test, a fixed volume of the fuel is placed in the test reservoir. During the test, a vibrator holding a nonrotating steel ball loaded with a specified mass (e.g. 200 g) is lowered to contact a disk that is completely submerged in the fuel. At a specific test temperature, the ball rubs against the disk with a 1-mm stroke at a frequency of 50 Hz for 75 min. The wear scars on the ball are measured under a microscope with a 100 × magnification. Typical results are 250 (good) and 600 (bad) μm. The target is currently set at <460 μm.

In the ASTM D 6078 SBOCLE test, a load arm holding a nonrotating chrome alloy steel ball is lowered to contact a partially fuel-immersed polished steel test ring rotating at a specific speed (e.g. 525 rpm). The ball rubs against the test ring for a given period of time, and the tangential friction force is recorded. Tests can be performed with a single load or incremental load following prescribed procedures. Typical incremental load results are 2500 to 5000 g, with >3100 g considered as good lubricity.

3.5 Corrosivity

The corrosivity of diesel fuels can be measured by the ASTM D 130, jointly with IP 154, copper strip tarnish test. A polished copper strip is immersed in a given quantity of fuel sample and heated at a specific temperature for a given amount of time. The copper strip used in the test is then compared with an ASTM Copper Strip Corrosion Standard to classify the corrosivity of the fuel.

3.6 Stability

The stability and the quantity of insoluble material formed during storage of diesel fuels under air over a given period of time can be tested by ASTM D 4625 (or IP 378). In this test, fixed volumes (400 mL) of filtered fuel are aged in borosilicate glass containers at 43 °C for periods of 0, 4, 8, 12, 18, and 24 weeks. After aging for a selected time period, the total amount of filterable and adhesive insolubles is measured after a sample is removed from storage and cooled to room temperature. This method is not suitable for quality control testing because of long storage periods, but it can be used as a tool for studies of fuel storage properties.

There are two accelerated oxidation tests adopted by ASTM. In the method of D 2274, a fixed volume of fuel is bubbled with oxygen at a specific rate and temperature for a given period of time. The total amounts of filterable and adhesive insolubles formed during the test are measured as mg per 100 mL. The second acceleration test, ASTM D 5304, uses an oxygen overpressure. In this test, 100 mL of fuel, with or without stabilizer additives, is placed in a borosilicate container that is then placed in a pressure vessel preheated at 90 °C. The vessel is then pressurized with oxygen to 100 psig (pounds per square inch) (800 kPa) and placed in a forced air oven at 90 °C for 16 h. The sample is cooled at the end of test to determine gravimetrically the amount of fuel insolubles. This method achieves acceleration of oxidation without resorting to higher temperatures.

A thermal stability test method of diesel fuels is under development in ASTM, which is based on Octel/Dupont F21-61 Accelerated Fuel Oil Stability test. In this test, 50 mL of fuel is passed through a 200-mesh filter at 150 °C for 90 or 180 min. The reflectivity of the filter under a standardized light is measured to determine the thermal stability of the fuel.

4 ANALYSES OF FUEL CONTAMINANTS

Water and ash contents are contaminants in diesel fuels. They are measured to meet specifications that are agreed upon between purchasers and suppliers at the time and place of delivery.

4.1 Water and Sediment

Water and sediment in diesel fuel can cause increased corrosion, filter plugging, and engine damage. In ASTM D 1796 (IP 75), a mixture of equal volumes of fuel and water-saturated toluene is centrifuged, after which the volumes of the water and sediment layers are read.

4.2 Ash Content

The amount of ash-forming material in diesel fuel is measured to determine whether the fuel might have unacceptable deposits on combustion. Ash can come from oil- or water-soluble metallic compounds or from extraneous solids such as dirt and rust. In the standard ASTM D 482 (IP 4) test, a weighed sample is burned in a vessel until ash and carbon remain. The carbonaceous residue is then placed in a muffle furnace at 775 °C to reduce it to a carbon-free ash; it is weighed after it has cooled to room temperature.

4.3 Carbon Residue

The carbon residue of base fuel without additives can indicate the coke-forming propensity of the fuel when used in pot-type and sleeve-type burners. In the ASTM D 524 (IP 14) test, a weighed sample is placed in a glass bulb with a capillary opening; the bulb is heated in a metal furnace at 550 °C. After volatile matter has evaporated, a small amount of air is introduced into the bulb for slight oxidation. The residue remaining in the bulb is reported, as a percentage of the original sample, as the Ramsbottom carbon residue.

5 ANALYTICAL TECHNIQUES FOR CHEMICAL COMPOSITION

Diesel fuels are mainly composed of hydrocarbons, with minor or trace amounts of sulfur-, nitrogen-, and oxygen-containing compounds. The determination of chemical composition of diesel fuels involves both standard tests and nonstandard instrumental analysis.[4]

5.1 Elemental Analysis

In addition to carbon and hydrogen, diesel fuels contain trace amounts of sulfur, nitrogen, oxygen, and other elements. The hydrogen-to-carbon ratio can be used to assess fuel characteristics and refining process performance. Sulfur and nitrogen contents are environmental concerns. Like the hydrogen-to-carbon ratio, they can also be used to assess the efficiency of upgrading processes at refineries. Carbon, hydrogen, and nitrogen contents (CHN analysis) can be simultaneously determined using instruments, as described in ASTM D 5291. In this method, the sample is combusted at 950–1020 °C. The combustion product is passed through calcium oxide to remove sulfur oxides. The remaining combustion gas is then passed over a heated copper train to remove excess oxygen and reduce nitrogen oxides (NO_x) to N_2. The final gas mixture of N_2, CO_2, and H_2O is separated by a heated gas chromatographic column and measured by a thermal conductivity detector. Alternatively, the gas mixture is passed over NaOH to remove CO_2 and then over magnesium perchlorate to remove H_2O. A thermal conductivity cell measures the remaining N_2. Simultaneously, but separately from the nitrogen measurement, the carbon and hydrogen are determined by measuring the CO_2 and H_2O levels using infrared (IR) cells.

There are several standard methods for the sulfur measurement. In a combustion method for samples with >0.1% sulfur (ASTM D 129), the sample is combusted in a closed bomb containing oxygen. The sulfur content is determined by weighing barium sulfate in the bomb washings after reaction with a known amount of barium chloride solution. Another method adopted by ASTM, D 1552, is to combust the samples that contain at least 0.06% sulfur in a stream of oxygen. The combustion product is continuously absorbed in an acidic solution of potassium iodide with starch indicator. Potassium iodate is added to the absorber solution to maintain the solution as a slightly blue color. The amount of potassium iodate consumed during combustion is used to determine the sulfur content.

The sulfur content can also be determined by instrumental methods. For samples containing 0.05 to 0.5 mass% sulfur, ASTM D 4294, using energy dispersive X-ray fluorescence (EDXRF) spectroscopy, can be employed. For samples with less sulfur, ASTM D 2622, a method that is based on the measurement of the Kα line at 5.373 Å ($1\ \text{Å} = 10^{-10}$ m) by wavelength-dispersive X-ray spectrometry after background subtraction, can be used. However, care should be taken to avoid or to correct for interference from other elements as impurities in the sample.

The oxygen content of the fuel is usually of little concern. It can be obtained by taking the balance of the CHNS analysis, i.e. 100% minus the percentages of the four elements. More accurate measurement can be obtained by neutron activation analysis (NAA), in which the sample is bombarded by high-energy (~14 MeV) neutrons. The intensity of the γ-rays emitted from the nuclei at a known time after the sample has been neutron irradiated is measured to determine the concentration of the element.

5.2 Chromatography

Diesel fuels can be separated by chromatography according to polarity and boiling point. Polarity separation is

normally carried out in liquid phase by liquid chromatography (LC), while boiling point separation is carried out in gas phase by GC. In general, liquid-phase separation has low resolving power that can separate components only into compound type groupings. High resolving power to resolve individual components is achievable with gas-phase separation. However, because of the large numbers of components present in diesel fuels, co-elution of components is common even with high-resolution GC using capillary columns.

5.2.1 Liquid Chromatography

Several standard liquid chromatographic methods have been established for diesel analysis. A simple test for determining the concentrations of aromatics and olefins in diesel fuels uses a fluorescent indicator adsorption (FIA) method, ASTM D 1319. The sample is introduced into a glass adsorption column packed with activated silica gel, with a small layer of the gel containing a mixture of fluorescent dyes. After all the sample is adsorbed onto the gel, alcohol is added to desorb the sample down the column. Aromatic, olefinic, and saturated hydrocarbons are separated according to their affinity with the gel. The dyes are also separated with hydrocarbon types. The boundaries of different hydrocarbon zones can be made visible under ultraviolet light. The volume percentages of aromatics, olefins, and saturates are calculated from the length of the zones.

A standard high-performance liquid chromatography (HPLC) method, IP 391, has been adopted by the European Community. It utilizes a polar column and refractive index detection to measure the amounts of saturates (nonaromatic hydrocarbons) and mono-, di-, and polynuclear aromatic hydrocarbons.[2] In this method, a known mass of the sample is diluted in a nonpolar solvent, such as heptane, and a fixed volume of the solution is injected into a HPLC set-up with a polar column (normal phase HPLC) to separate components according to aromatic ring types. Nonaromatic hydrocarbons that have little affinity with the column elute first, followed by mono- and diaromatic hydrocarbons. Then, the column is backflushed to elute the polynuclear aromatic hydrocarbons as a single peak. The column temperature is maintained constant ($\pm 1\,^{\circ}C$) within the range 20–40 °C. However, the responses of the refractive index detector to hydrocarbon molecules depend on the structure, and they vary greatly with different types of molecules. Hence, careful calibration in response factors is necessary to ensure accuracy of the quantitative measurement. Photodiode array (PDA) detection can be used for distinguishing and identifying the aromatic type of the components. However, its use for quantitation is also limited by nonuniform responses for various types of aromatic hydrocarbons.

5.2.2 Supercritical Fluid Chromatography

Supercritical fluid chromatography (SFC) with flame ionization detection (FID) has proven to be quite effective for the determination of the total amount of aromatics in diesel fuels.[5,6] It has become a standard ASTM method (ASTM D 5196) for the determination of aromatic content and polynuclear aromatic content of diesel fuel. A small volume of the sample is injected onto a packed silica column and eluted using supercritical carbon dioxide as the mobile phase. Monoaromatic and polynuclear aromatic hydrocarbons are separated from nonaromatic hydrocarbons (saturates and mono-olefins, if present). The resolution among different aromatic types is poor, especially for the distinction between di- and triaromatics. However, the relatively uniform response of FID for different hydrocarbon types is an advantage of this method.

A more elaborate SFC separation scheme has been suggested by Andersson et al.[7] It uses multiple columns and supercritical carbon dioxide as a mobile phase to determine saturated compounds (paraffins and naphthenes), olefins, mono-, di-, and triaromatics, and polar compounds in diesel fuel distillates. Using a valve switching technique, compounds are retained and then eluted from the multiple columns at different times. During the compound retention period, polar compounds are retained in a cyano-bonded column, aromatics in a silica column, and paraffins and olefins in a silver-impregnated column. During the elution period, aromatics are eluted from the silica column first, after which the polars are back-flushed from the cyano-bonded column. Saturates are then eluted from the silver-impregnated column, followed by back-flushing the olefins from the silver-impregnated column. A mass spectrometer coupled with SFC is used to provide the basis of column switching, to select integration limits, and to obtain more detailed distributions of compound types.

5.2.3 Gas Chromatography

GC has been used for routine analysis of volatile complex organic mixtures. Although many different stationary phases can be used, the most common columns in GC analysis are coated with nonpolar methyl silicone. With a nonpolar GC column, components are separated essentially by boiling point. This correlation between boiling point and GC retention time is the basis for using GC simulated distillation, described earlier, to estimate yields of petroleum distillates, including diesel fuel streams. For general quantification, FID is used because of its uniform response across various hydrocarbon types.

GC with element-specific detectors can conveniently measure small or trace amounts of sulfur- and nitrogen-containing components in diesel fuels. Specific detection is important not only in quantifying trace amounts of

heteroatom-containing compounds without interference from the overwhelming hydrocarbon components, but also in helping combined gas chromatography/mass spectrometry (GC/MS) to focus on retention time regions in identifying heterocompounds of interest.

Atomic emission detectors (AED) can simultaneously detect several elements including C, H, N, S, and O.[8] The GC effluent is introduced into a microwave plasma cavity where the compounds are atomized and excited by plasma energy. The light emitted from the excited atoms is separated by PDA and sorted by computer to produce chromatograms specific to individual elements. Another specific detection is based on chemiluminescence to detect light emitted from chemical reactions. Unlike AED, in which the light comes from excited atoms, the light in chemiluminescence comes from excited molecules produced by reaction with a strong oxidizing agent such as ozone. Sulfur and nitrogen compounds can be detected using chemiluminescence as a means of specific detection.[9] Other sulfur- and nitrogen-specific detectors include flame photometric detectors (FPD)[10] and nitrogen/phosphorus detectors (NPD),[11] respectively. Oxygen compounds can also be detected by oxygen-selective flame ionization detectors (OFID).[1] The oxygenates are cracked by flowing the GC effluent through a capillary tube containing Pt/Rh catalysts at 850–1300 °C to form carbon monoxide, which is then converted into methane as it flows through a hydrogenator following the cracker. A flame ionization detector detects the resulting methane.

5.3 Mass Spectrometry

Mass spectrometry alone has also been used for hydrocarbon type analysis of diesel fuels and other middle distillates.[12] The concept of hydrocarbon type analysis is to use mass spectral fragment peak summations, each respectively characteristic of compound types to be determined.[13] Numerous components are grouped into a few compound types as paraffins, cycloparaffins, monoaromatics, diaromatics, polynuclear aromatics (3+-ring aromatics), and sulfur compounds. Sulfur compounds (benzothiophenes, dibenzothiophenes, and benzonaphthothiophenes) are the only heteroatom-containing compound types included in the analysis. These are the major compound types typically found in diesel fuels. Due to overlaps of mass series of certain saturated and aromatic hydrocarbons, the sample is preferably separated by elution chromatography, e.g. modified ASTM D 2549, into saturate and aromatic fractions before the mass spectrometric analysis. The separated fractions are analyzed by mass spectrometry using ASTM D 2425 for both or ASTM D 2786 for saturates and ASTM D 3239 for aromatics.[1]

The most detailed compositional analysis of diesel fuels can be obtained by GC/MS, in which a gas chromatograph separates components before their elution into a mass spectrometer for characterization. However, because of the numerous components in diesel fuels, complete or near-complete separation of individual components by GC is impossible. Hence, for trace components, open column, HPLC or other types of chemical separations are often used to concentrate components of interest (typically heteroatom-containing compounds) for GC/MS analysis with minimal interference from the predominant components.[14]

For further analysis, relatively pure fractions can be obtained by chromatographic procedures that fractionate the sample into saturates, monoaromatics, diaromatics, polyaromatics, and nitrogen/oxygen compounds (polars). Each fraction can then be characterized by GC/MS for the identification of isomers. Many of these fractions can be further separated by thermal diffusion for greater enrichment of various subtypes, i.e. paraffins, alkylbenzenes, naphthalenes, etc.

The carcinogenic, teratogenic, and/or mutagenic activities of polycyclic aromatic hydrocarbons (PAHs) are isomer specific. Diesel fuel can be separated by solid phase extraction into saturates, mono-, di-, and polyaromatic hydrocarbons. The 2- and 3-ring aromatics are identified by capillary GC and GC/MS.[15]

5.4 Nuclear Magnetic Resonance and Molecular Spectroscopies

Unlike chromatography and mass spectrometry, which can analyze components at a molecular level in a complex mixture, nuclear magnetic resonance (NMR) and molecular (IR and Raman) spectroscopies are used to determine the average structure of a mixture. However, the specific chemical shifts of functional groups in NMR and their unique vibrational frequencies in IR and Raman spectroscopies provide valuable information about their distribution in the mixture. For example, ^{13}C-NMR has been used to effectively determine the content of aromatic and aliphatic carbons in diesel fuels.[16]

6 MOLECULAR COMPOSITION OF DIESEL FUELS

Analysis of a representative diesel fuel by the methods described here shows that the main components of diesel fuels are saturated and aromatic hydrocarbons. The saturated hydrocarbons dominant in diesel fuels include normal paraffins, isoparaffins, and cycloparaffins (naphthenes). Normal paraffins have high cetane numbers and are desirable molecules in diesel fuels

for combustion. High-boiling normal paraffins, however, cause cold-flow problems if present in excess. Other than isoprenoid hydrocarbons such as pristane and phytane, most of the isoparaffins in diesel fuels are mono- and dimethyl alkanes, predominantly with branch points near the ends of the alkyl backbone. The remaining saturated hydrocarbons are mainly 1-ring to 3-ring cycloparaffins. The 4- and 5-ring cycloparaffins, such as steranes and hopanes, are rarely present. Aromatic compounds are mainly benzenes, indans, tetralins, indenes, naphthalenes, biphenyls, acenaphthenes, fluorenes, acenaphthalenes, phenanthrenes, anthracenes, and naphthenophenanthrenes. Diaromatic hydrocarbons are generally the most abundant aromatic components in diesel fuels. Trace amounts of PAHs (3+-ring aromatic compounds) such as chrysenes, pyrenes, benzanthracenes, and perylenes can also be present.

The most common sulfur compounds are benzothiophenes and dibenzothiophenes. Trace amounts of nitrogen compounds include indoles, carbazoles, quinolines, acridines, and phenanthridines. The oxygen compounds most often reported in diesel fuels are phenols and dibenzofurans.

Formulated diesel fuels also contain small amounts of additives, such as cetane improvers, antioxidants, corrosion inhibitors, metal deactivators, dispersants, detergents, lubricity agents, demulsifiers, and cold-flow improvers, to improve storage stability and performance.

Many physical and physico-chemical properties of diesel fuel are related to its composition. Linear regression and principal component analysis have been used for statistical analysis and correlation studies of chemical composition and physical characteristics (including cetane number and cetane index) of diesel fuels.[16–18]

ACKNOWLEDGMENTS

The author would like to acknowledge the valuable discussions with Dr Donna Hoel of Esso Research Centre, Abingdon, UK.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detectors
API	American Petroleum Institute
ASTM	American Society for Testing and Materials
CFPP	Cold Filter Plugging Point
EDXRF	Energy Dispersive X-ray Fluorescence
FBP	Final Boiling Point
FIA	Fluorescent Indicator Adsorption
FID	Flame Ionization Detection
FPD	Flame Photometric Detectors
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HFRR	High-frequency Reciprocating Rig
HKN	Heavy Coker Naphtha
HPLC	High-performance Liquid Chromatography
IBP	Initial Boiling Point
IP	Institute of Petroleum
IR	Infrared
ISO	International Organization for Standardization
LC	Liquid Chromatography
LCCO	Light Catalytic Cracker Oil
LKGO	Light Coke Gas Oil
LTFT	Low-temperature Flow Test
NAA	Neutron Activation Analysis
NMR	Nuclear Magnetic Resonance
NPD	Nitrogen/Phosphorus Detectors
OFID	Oxygen-selective Flame Ionization Detectors
PAH	Polycyclic Aromatic Hydrocarbon
PDA	Photodiode Array
SBOCLE	Scuffing Load Ball on Cylinder Lubricity Evaluator
SFC	Supercritical Fluid Chromatography

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Fuels Analysis, Regulatory Specifications for • Full Range Crudes, Analytical Methodology of • High-temperature Simulated Distillation Applications in Petroleum Characterization • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Nuclear Magnetic Resonance Characterization of Petroleum

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Normal-phase Liquid Chromatography • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy

REFERENCES

1. *Annual Book of ASTM Standards*, American Society for Testing and Materials, West Conshohocken, PA, 1999.
2. *IP Standard Methods for Analysis and Testing of Petroleum Products*, Institute of Petroleum, London, 1999.
3. 'Diesel and Domestic Heating Fuels–Determination of Cold Filter Plugging Point', British Standards Institution, London, BS 2000: Part 309, 1998.
4. C.S. Hsu, M.W. Genowitz, G.J. Dechert, D.J. Abbott, R. Barbour, 'Molecular Characterization of Diesel Fuels by Modern Analytical Techniques', Symposium on Chemistry of Diesel Fuels, 216th American Chemical Society National Meeting, Boston, MA, 23–27, August, 1998. *Prepr. Am. Chem. Soc., Div. Petrol. Chem.*, **43**, 506 (1998).
5. E.N. Chen, Jr, P.D. Cusatis, E.J. Popiel, 'Validation of the Aromatic Ring Distribution in Diesel Fuel Refinery Streams by Supercritical Fluid Chromatography and Mass Spectrometry', *J. Chromatogr.*, **637**, 181–186 (1993).
6. B.E. Richter, B.A. Jones, N.L. Porter, 'Optimized Supercritical Fluid Chromatographic Instrumentation for the Analysis of Petroleum Fractions', *J. Chromatogr. Sci.*, **36**, 444–448 (1998).
7. P.E. Andersson, M. Demirbüker, L.G. Blomberg, 'Characterization of Fuels by Multi-dimensional Supercritical Fluid Chromatography and Supercritical Fluid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **641**, 347–355 (1993).
8. B.D. Quimby, D.A. Grudoski, V. Giarrocco, 'Improved Measurement of Sulfur and Nitrogen Compounds in Refinery Liquids Using Gas Chromatography–Atomic Emission Detection', *J. Chromatogr. Sci.*, **36**, 435–443 (1998).
9. S.E. Eckert-Tilotta, S.B. Hawthorne, D.J. Miller, 'Comparison of Commercially Available Atomic Emission and Chemiluminescence Detectors for Sulfur-selective Gas Chromatographic Detection', *J. Chromatogr.*, **591**, 313–323 (1992).
10. J.N. Driscoll, A.W. Berger, 'Improved Flame Photometric Detector for the Analysis of Sulfur Compounds by Gas Chromatography', *J. Chromatogr.*, **468**, 303–308 (1989).
11. R. Buffington, M.K. Wilson, *Detectors for Gas Chromatography*, Hewlett Packard, Avondale, PA, 1991.
12. C.J. Robinson, 'Low Resolution Mass Spectrometric Determination of Aromatics and Saturates in Petroleum Fractions', *Anal. Chem.*, **43**, 1425–1434 (1971).
13. M.E. Fitzgerald, V.A. Cirillo, 'Mass Spectrometric Method for Analysis of Petroleum Distillates in the Furnace Oil–Kerosene Boiling Range', *Anal. Chem.*, **34**, 1276–1280 (1962).
14. C.S. Hsu, K. Qian, W.K. Robbins, 'Nitrogen Speciation of Petroleum Polars by Compound Class Separation and Online LC/MS', *J. High Resolut. Chromatogr.*, **17**, 271–276 (1994).
15. J. Bundt, W. Herbel, H. Steinhart, S. Franke, W. Francke, *J. High Resolut. Chromatogr.*, **14**, 91–98 (1991).
16. D.J. Cookson, B.E. Smith, 'Calculation of Jet and Diesel Fuel Properties Using ^{13}C NMR Spectroscopy', *Energy Fuels*, **4**, 152–156 (1990).
17. D.J. Cookson, C.P. Loydd, B.E. Smith, 'Investigation of Chemical Basis of Diesel Fuel Properties', *Energy Fuels*, **2**, 854–860 (1988).
18. M. Sjögren, H. Li, U. Rannug, R. Westerholm, 'A Multivariate Statistical Analysis of Chemical Composition and Physical Characteristics of Ten Diesel Fuels', *Fuel*, **74**, 983–989 (1995).

Fuel Performance Specifications, Mid-infrared Analysis of

Boris Mizaikoff
Vienna University of Technology, Austria

1 Introduction 1
2 Conventional Fuel Performance Specification 2
2.1 Fuel Parameters and Properties 2
2.2 United States Environmental Protection Agency Regulations 2
3 Mid-infrared Spectroscopy for the Determination of Fuel Performance Specifications 4
3.1 Short Introduction to Mid-infrared Spectroscopy 4
3.2 Fourier Transform Infrared-based Instrumentation 4
3.3 Infrared Filter-based Instrumentation 7
3.4 Gas Chromatography/Fourier Transform Infrared 7
3.5 Chemometrics 8
3.6 Mid-infrared Spectroscopy-based American Society for Testing and Materials Methods for Fuel Characterization 8
3.7 Comparison with Other Spectroscopic Techniques 9
4 The Way into the Market – Selected Examples 10
5 Conclusions 10
Abbreviations and Acronyms 10
Related Articles 11
References 11

The release of the Clean Air Act (CAA) Amendments of 1990 introduced reformulated gasoline legislation regulating the composition of fuels and particularly gasoline in order to improve the air quality by a significant reduction in vehicle exhaust emission levels. With the introduction of certain amounts of nonfuel oxygenates into fuel including various alcohols such as methanol, ethanol, t*-butanol and ethers like methyl* tert*-butyl ether (MTBE), ethyl* tert*-butyl ether (ETBE),* tert*-amyl methyl ether (TAME) and diisopropyl ether (DIPE) a reduction of carbon monoxide (CO) and unburned hydrocarbon emissions is envisaged. Furthermore, a reduction in aromatic compounds and benzene is demanded providing a decreased release of volatile organic compounds (VOCs).*

With the implementation of this program, the refining industry and their petrochemical laboratories have been forced to control accurately and to monitor continuously their gasoline blending processes. Furthermore, regulatory agencies are required to implement the use of portable analytical techniques in order to ensure the enforcement of such novel threshold values.

Besides the conventional techniques using separation-based methods such as gas chromatography (GC), the mid-infrared (MIR) spectral range has gained significant importance. Owing to the molecule specific absorption patterns of ethers, alcohols and aromatics, the distinct determination of fuel additives and subsequent derivation of general fuel parameters is enabled. The performance of FTIR (Fourier transform infrared) and filter-based spectroscopic techniques will be reviewed and compared to other spectroscopic approaches. Furthermore, data evaluation procedures and ASTM (American Society for Testing and Materials) approval of these novel methods will be highlighted.

1 INTRODUCTION

The improvement of the air quality by reduction of emissive pollution has been a major concern of environmental protection efforts in the 1990s. Since gasoline and diesel fuels together power more than 95% of the vehicle fleet in the USA, exhaust vehicle emissions, resulting from combustion products and contributing to the greenhouse effect as well as to the ozone problem, are considered a target issue.

With the release of the CAA Amendments of 1990, reformulated gasoline legislation was introduced, regulating the composition of fuels and particularly gasoline in order to improve the air quality by a significant reduction of vehicle exhaust emission levels.[1] In accordance with the act, the refining industry and their petrochemical laboratories are required to control accurately and to monitor continuously their gasoline blending processes. Added oxygenates include alcohols such as methanol, ethanol, *tert*-butanol and ethers like MTBE, ETBE, TAME and DIPE, which progressively reduce carbon monoxide (CO) and unburned hydrocarbon emissions with increasing oxygen content. Furthermore, the CAA enforces a reduction in content of aromatic compounds and benzene is reduced providing a decreased release of VOCs.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The standard determination method for such compounds is given in detail in section 2 and is based on gas chromatography with oxygen-selective flame ionization detection (GC/O-FID) as mandated by ASTM and USEPA (United States Environmental Protection Agency).[2,3] The increasing demand claimed by regulatory agencies as well as petroleum companies for rapid and reliable analytical techniques capable of assessing these parameters in the field, has initiated the search for novel methods to supplement and possibly replace conventionally used separation techniques.

Infrared (IR) spectroscopy and particularly the MIR spectral range considered from 4000 to 400 cm^{-1} (2.5–25 μm) represents an analytical technique of constantly increasing importance. The introduction of rapid and powerful microprocessors enabling the implementation of fast Fourier transform calculation algorithms, especially, have established FTIR spectroscopy as one of the most commonly applied analytical methods since the late 1970s.[4] While chromatographic techniques rely on the separation of the analyte molecules in the column in relation to their mass differences, IR spectroscopy enables the selective recognition of different species in relation to their specific molecular vibrations resulting in distinctive absorption bands in the corresponding spectral region. In contrast to NIR (near infrared) methods, MIR spectroscopy provides access to the comparatively strong fundamental vibrational modes of organic molecules, and individual fuel components can be differentiated by their characteristic pattern of absorption bands. Hence, the advantages of IR absorption spectroscopy compared to chromatographic systems particularly for a field applicable analytical method are evident: (i) usually no preparation of the sample is required, (ii) the potential exists for continuous operation providing on-line and real-time information because (iii) no preparation of the instrument prior to the measurement or after the determination is necessary, (iv) full analysis in typically less than 5 min, (v) the measurement and data evaluation procedure can be fully automated (chemometric methods), (vi) no extra trained personnel are needed.

Based on these considerations, several studies have proven that MIR FTIR spectroscopy is a versatile, efficient and accurate analytical technique for the determination of fuel additives and the estimation of key gasoline properties.[5] Besides supporting the control and implementation of environmental regulations, an alternative tool for petroleum laboratories is provided, which could considerably increase the sample throughput, minimize the sample preparation and volume requirements and reduce the costs per sample compared to conventional analysis.

2 CONVENTIONAL FUEL PERFORMANCE SPECIFICATION

2.1 Fuel Parameters and Properties

With respect to regulatory measures, the most important parameters of fuels which have to be addressed for sufficient fuel specification are listed in Table 1. The particular substances of interest, which have to be measured selectively, are given in Table 2.

With the introduction of oxygenate additives for gasoline, a reduction of ambient CO by about 10% is envisaged. Reformulated gasoline will reduce VOCs and air toxics by approximately 15% requiring a content of at least 2 wt% oxygen and a maximum content of 1 wt% benzene.

Besides this positive environmental impact it is of major interest to investigate potential human health risks related to these new additives, especially MTBE as the most common representative of the oxygenates. Since MTBE is a widely used fuel blending component its environmental fate, especially in the aquatic groundwater phase, as well as its adverse human effects are currently the subject of intensive investigations with particular focus on long-term studies.[6]

2.2 United States Environmental Protection Agency Regulations

Gasoline and diesel fuel produced from petroleum (crude oil) power more than 95% of the combustion engine motored vehicle fleet in the USA. With the release

Table 1 Standard fuel parameters and properties

Fuel property	Procedure
Density	ASTM D 4052, D 1298
Viscosity (cSt)	ASTM D 445
Vapor pressure (mbar)	ASTM D 5191, USEPA 40 CFR part 80, app. E, meth. 3
Distillation (°C) (initial/final boiling point)	ASTM D 86
Freezing point (°C)	ASTM D 823866
Flash point (°C)	IP 170
Motor octane number (MON)	ASTM D 2700
Research octane number (RON)	ASTM D 2699
Aromatic hydrocarbons (vol%)	ASTM D 1319
Olefinic hydrocarbons (vol%)	ASTM D 1319
Saturated hydrocarbons (vol%)	ASTM D 1319
Sulfur (wt%)	ASTM D 2622
Benzene (vol%)	ASTM D 3606, D 4420
Alcohols and ethers (vol%)	ASTM D 4815, D 5599, USEPA GC/O-FID
Total oxygen (wt%)	ASTM D 4815, D 5599, USEPA GC/O-FID

IP, Institute of Petroleum.

Table 2 Properties of the most important fuel additives

Substance	Abbreviation	CAS number	Formula	Structure	Molecular weight ($g\ mol^{-1}$)	Boiling point (°C)	Water solubility (% at 25 °C)	Vapor pressure (kPa)
Methyl *t*-butyl ether	MTBE	1634-04-4	$C_5H_{12}O$	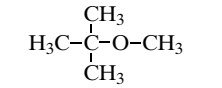	88.15	55.20	3.620	10 at -2 °C 100 at 54.8 °C
Ethyl *t*-butyl ether	ETBE	637-92-3	$C_6H_{14}O$	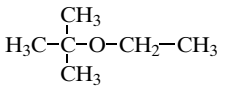	102.17	73.10	–	1 at –24.6 °C 10 at 14.4 °C 100 at 72.6 °C
t-Amyl methyl ether	TAME	994-05-8	$C_6H_{14}O$	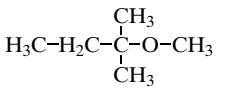	102.17	86.30	–	–
Diisopropyl ether	DIPE	108-20-3	$C_6H_{14}O$	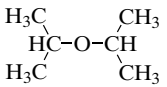	102.17	68.51	1.200	19.9 at 25 °C
Methanol	MeOH	67-56-1	CH_4O	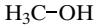	32.04	65.00	–	16.9 at 25 °C
Ethanol	EtOH	64-17-5	C_2H_6O	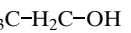	46.07	78.29	–	7.87 at 25 °C
t-Butanol	TBA	75-65-0	$C_2H_{10}O$	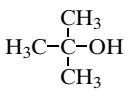	74.12	82.40	–	5.52 at 25 °C
Benzene		71-43-2	C_6H_6		78.11	80.09	0.177	12.7 at 25 °C

TBA, tertiary butyl alcohol.

of the CAA in 1990, strategies to reduce and prevent air pollution caused by vehicle exhaust emissions have refocused to emphasize improvements in fuel properties and thus fuel composition, rather than relying mainly on engine/vehicle emission control such as catalytic converters. This initiative led to provisions forcing modifications in the content in gasoline of nonpetroleum additives and, correspondingly, more stringent emission standards for combustion engines.

The pollution reduction programs implemented by the USEPA within the clean air legislation can be condensed into two main sections:[3]

(i) Oxygenated fuels program: during the winter months from November 1 to March 1 gasolines sold in metropolitan areas defined as CO nonattainment zones have to contain at least 2.7% (w/w) oxygen.

(ii) Reformulated gasoline program: to reduce the ozone and smog levels gasolines available in nine extended city areas with most severe ozone pollution have to be reformulated in the first phase of this program, launched in the beginning of 1995, to an oxygen content of at least 2.0% (w/w). A reformulation of their contents of aromatics, benzene, heavy metals, sulfur and detergents is also included in the program.

With the implementation and compliance of these regulations, an increasing demand for the accurate determination of the oxygen content of fuels and for appropriate analytical techniques is obvious.

Currently applied standard methods for the determination of fuel components including oxygenates and aromatics mainly involve separation techniques such as GC/AED (atomic emission detection), GC/FID (flame ionization detection) and GC/MS (mass spectrometry).[7–9]

3 MID-INFRARED SPECTROSCOPY FOR THE DETERMINATION OF FUEL PERFORMANCE SPECIFICATIONS

3.1 Short Introduction to Mid-infrared Spectroscopy

The MIR spectral range is generally considered to be from 4000 cm^{-1} to 400 cm^{-1} corresponding to 2.5–25 μm. Probing fuels and fuel mixtures with MIR radiation enables individual components to be detected and differentiated by their specific absorption pattern in this spectral region. Because the vibrational motions of the molecules occur at distinctive frequencies depending on their atomic masses, their relative bond strengths and their neighboring atoms and bonds, functional group assignment enables individual components to be distinguished in a fuel sample. This unique set of frequencies is directly associated with the molecular structure of the respective component and is represented in the so-called MIR spectrum of the analyte. Since MIR spectroscopy is based on the evaluation of the fundamental resonances, usually well-defined, sharp and specific absorption bands are available. The amplitude of the molecular vibrations is related to the concentration of the molecule within the sample and is represented in the peak height of an absorption spectrum. Since similar molecular structures partially result in absorption peaks located at the same frequency, the whole spectrum has to be evaluated for accurate data analysis.

The usefulness of applying MIR spectroscopy for fuel characterization becomes particularly evident when relating chemical gasoline features to the respective components influencing those properties.[10–12] As an example, the octane number is strongly influenced by the aromatic hydrocarbon concentration: an increase in these components increases the octane number and simultaneously decreases the cetane number for diesel fuels. As detailed later, aromatic components such as benzene can be detected directly in the MIR region. The investigation of so-called *oxygenated* fuels especially is facilitated using MIR spectroscopic methods.[13] Oxygen-containing fuel additives – mainly alcohols and ethers – improve the performance of combustion engines towards a more complete burning process thus reducing the CO emissions of vehicles by 15–20%. Because the characteristic C–O–H bonds of alcohols and the C–O–C bonds of ethers are represented by distinct absorption bands in the MIR region, analytical tools operating in that particular spectral region are of significant interest.

Examples of IR spectra of fuel additives are given in Figure 1 and demonstrate the suitability of detecting individual components using MIR spectroscopy as well as revealing the difficulty of differentiating between individual substances within a complex mixture.

Generally, the reported instrumentation can be divided into (i) FTIR spectrometer-based instruments that evaluate the whole spectral range using either attenuated total reflection (ATR) or micropath transmission cells, (ii) filter instruments that use a set of bandpass filters and a chopped light source to evaluate only the MIR spectral region of interest and (iii) hyphenated techniques that combine MIR spectroscopic detection with GC.

The most important techniques based on MIR spectroscopy, which can be applied to determine fuel specifications, are discussed below.

3.2 Fourier Transform Infrared-based Instrumentation

Various studies have shown that FTIR spectroscopy in the MIR region is the most versatile, efficient and accurate spectroscopic technique for the simultaneous assessment of key fuel parameters, because the whole spectral range

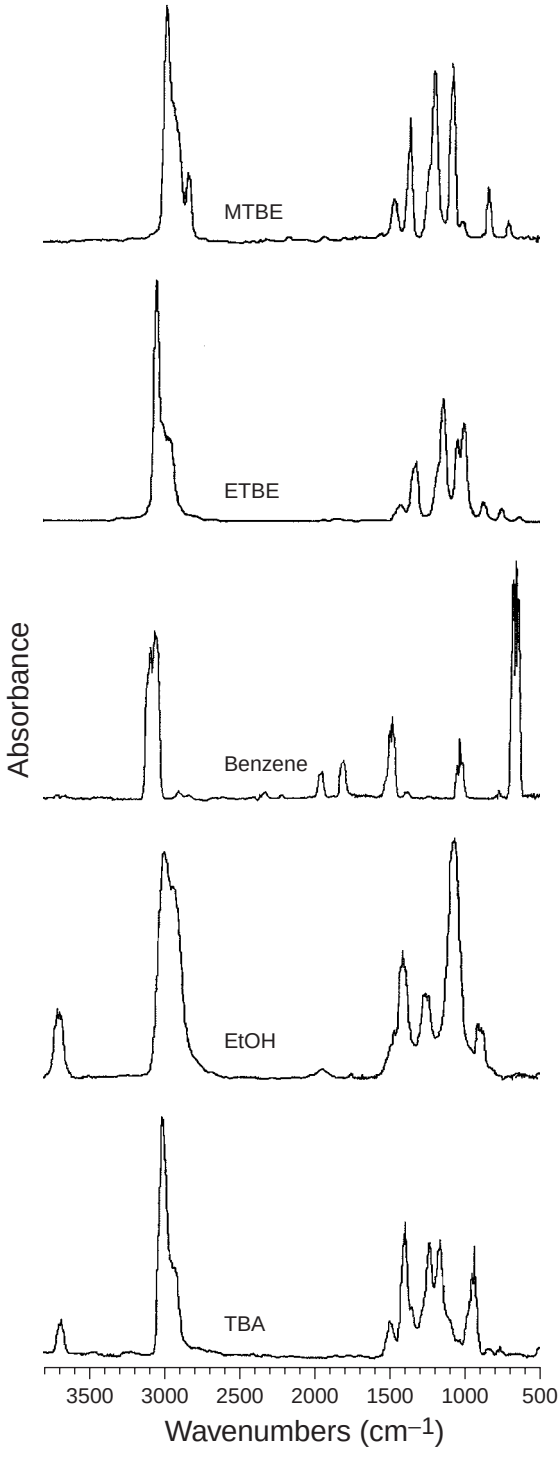

Figure 1 Examples of IR gas phase spectra of selected fuel additives.

from 4000 to 400 cm^{-1} can be used to gain molecule-specific information.

The main FTIR spectroscopic work on middle distillate fuels is focused on the direct investigation of gasoline properties, and particularly of regulated components such as benzene and oxygenates,[14,15] as well as on deriving properties like the carbon/hydrogen ratio, aromatic hydrocarbon content, cetane index, refractive index, heat of combustion, viscosity and density. Particularly, in combination with multivariate (chemometric) data evaluation procedures, even derived values of gasoline properties such as the motor and research octane number (MON, RON), total aromatic, olefinic and saturated hydrocarbon content or total oxygen content are accessible.

In general, all methods collect an FTIR spectrum of a fuel sample in order subsequently to relate the resulting spectra either to a particular fuel additive/component or to a correspondingly selected fuel property value. Most methods involve additional multivariate (chemometric) data evaluation strategies enabling full automation of data processing, as well as the possibility of detecting the smallest changes in the IR spectrum.

Two main sampling techniques can be distinguished using either ATR or micropath transmission cells, the latter frequently combined with a flow injection analysis (FIA) arrangement for sample handling.

3.2.1 Fourier Transform Infrared/Attenuated Total Reflection

ATR has become a widespread sampling technique in FTIR analysis since the introduction of internal reflection spectroscopy (IRS) by Harrick in late 1960s.[16] The evanescent field can be used to probe solid and liquid samples adjacent to the surface of a so-called ATR crystal, usually made from materials such as zinc selenide (ZnSe) or germanium (Ge). While circular ATR cells are also available, less expensive horizontal ATR configurations using conventional planar crystals with approximately 12 internal reflections are applied in the main, ensuring reproducible baselines and cross-contamination-free spectra following successive sampling.[5,17–19] The top side of the ATR crystal is usually enclosed by a gasketed cover in order to reduce the evaporation and allow fuel samples to flush through as schematically illustrated in Figure 2a.

The penetration depth (d_P) of the IR radiation depends on the wavelength of the incident light (λ), the refractive indices of the waveguide (n_1) and the surrounding medium ($n_2, n_2 < n_1$) and on the angle of incidence (θ), Equation (1)

$$d_P = \frac{\lambda}{2\pi\sqrt{n_1^2 \sin^2\theta - n_2^2}} \quad (1)$$

Typically, the penetration depth lies around 2 μm within the fingerprint region of the MIR range.

Following FTIR/ATR measurements evaluated with a partial least squares (PLS) spectral decomposition technique, Fodor et al. report reproducibility values for the

parameters listed in Table 1, including the boiling point at 50/90 vol% condensate and the end boiling point, that match or improve upon the ASTM reproducibility values of the respective standard procedures.[5] Furthermore, the measurement time of a single sample could be reduced to 1 min and the sample volume reduced to 2 mL, attesting to the potential of FTIR/ATR measurements combined with multivariate data evaluation.

3.2.2 Fourier Transform Infrared/Transmission Cell

The use of inexpensive flow cells with a very small optical pathlength, typically between 0.1 and 0.015 mm, enables to determine the concentration of various fuel components by a simple transmission measurement as schematically depicted in Figure 2b. The windows of such cells are usually made from IR-transparent materials such as potassium bromide (KBr), calcium fluoride (CaF) or zinc selenide (ZnSe) providing access to a spectral range down to 400 cm^{-1}.[20] Coupling such a measurement cell with an FIA system enables automated, simple and rapid sampling with the possibility of continuous monitoring.[21,22] Following the Lambert–Beer law, quantitative determination of fuel components is possible relating the absorbance (A) to the molar extinction coefficient (ε), the optical pathlength (l) and the concentration (c) by, Equation (2)

$$A = \varepsilon c l \tag{2}$$

Gallignani et al. report the detection limits of 0.02 vol% benzene with gasoline samples diluted 1 : 9 in hexane by

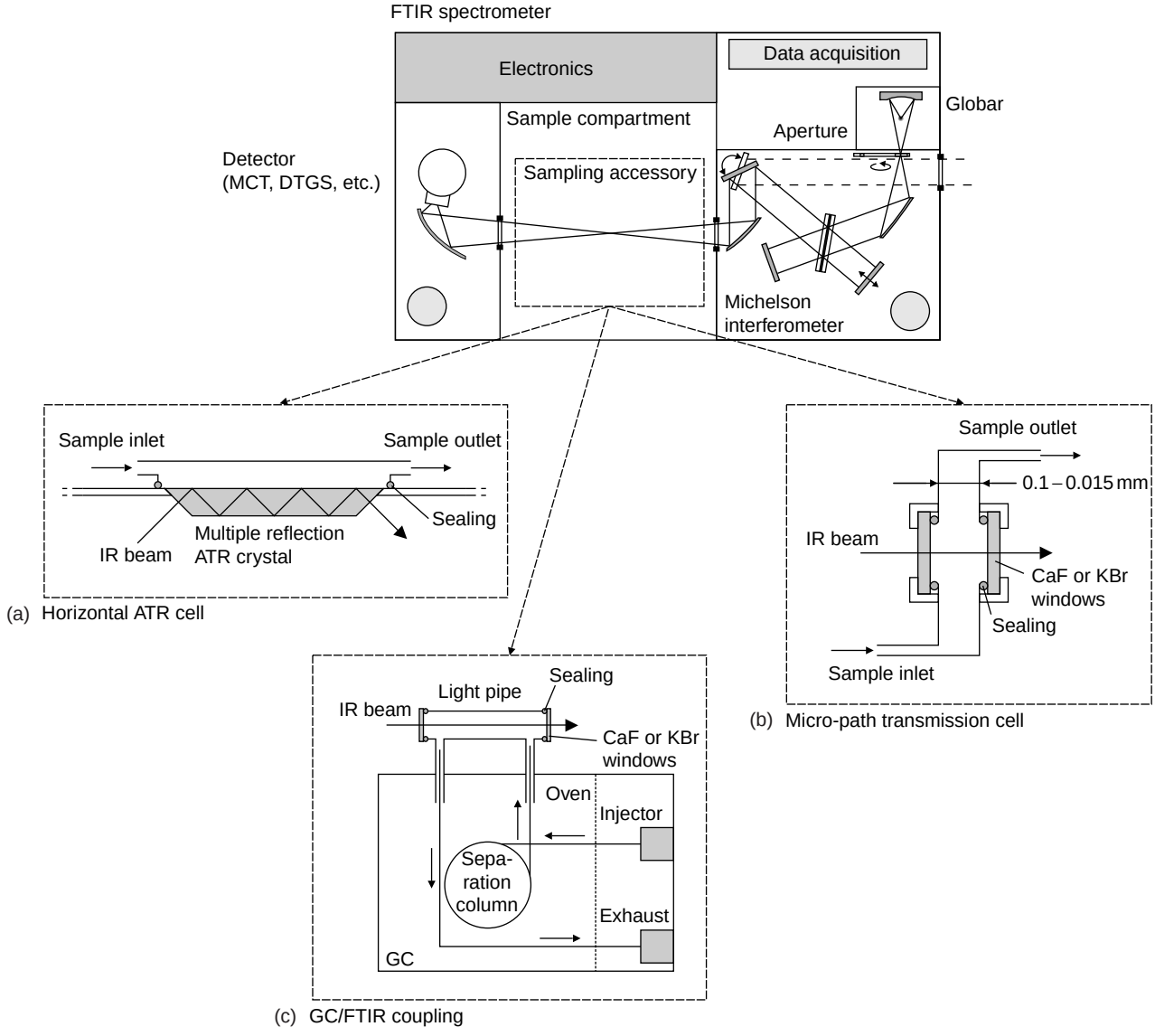

Figure 2 Combination of FTIR spectroscopy with (a) an ATR system, (b) a micropath transmission cell or (c) GC.

simply evaluating the peak height of the characteristic benzene absorption band at 675 cm^{-1}.[21] Coupling with a flow injection manifold ensures reproducible sampling and dilution conditions as demonstrated by comparison with results obtained by GC recommended as an ASTM reference method (D 3606-91). Owing to the miniaturized flow-cell the reagent consumption is kept at a minimum.

This method was later extended successfully to determine toluene (0.01–2.0 vol%) and MTBE (0.035–2.0 vol%).[22,23] Using the same procedure, López-Anreus et al. demonstrated the possibility of determining benzene, toluene, and MTBE in gasoline after vaporization at 90 °C and subsequent determination in an IR multiple pass gas cell in combination with a flow injection system.[24]

3.3 Infrared Filter-based Instrumentation

Filter-based IR instruments utilize only selected bands of the MIR spectral range containing the absorption features of interest. The main advantage compared to FTIR spectroscopy using the information content of the whole spectrum can be seen in the considerably lower costs of the spectroscopic equipment. Owing to full automatization and compactness such instruments are ideally suited to on-site determination of fuel components.

Filter spectrometers use a simple light source emitting MIR radiation, such as a black body filament, which is modulated by a chopper or can be pulsed. The IR beam is guided through a sample cell with fixed optical pathlengh, either in transmission or ATR configuration, to an appropriate IR detector, e.g. MCT (mercury cadmium telluride), DTGS (deuterated triglycine sulfate) or pyroelectric. Fixed frequency narrow bandpass filters usually mounted on a filter wheel are used to select the frequency range of interest.

All approved ASTM methods listed in section 3.6 involving MIR spectroscopy (except GC/FTIR coupling) allow the use of filter instruments as well as FTIR spectrometers.

3.4 Gas Chromatography/Fourier Transform Infrared

The application of the combination GC with FTIR for the determination of oxygenates in gasoline was mainly driven by the fact that conventional GC methods using O-FID or AED gave problems with the reliability or hydrocarbon interference.[25,26] Primary chromatographic separation enables detection and resolution of impurities, for example, from the feed ethers during the blending procedure.

Within GC/FTIR the GC column is directly interfaced to the gas flow cell of an FTIR spectrometer using the so-called light pipe technique with a heated connection as indicated in Figure 2c.

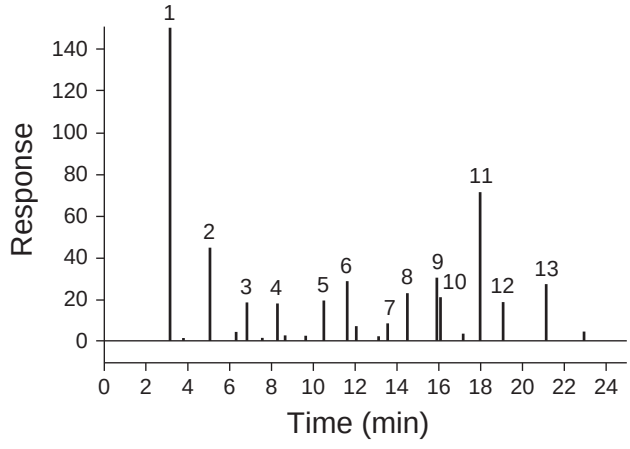

Figure 3 Schematic selective absorbance (wavelength) chromatogram from 1250–1000 cm^{-1} of 1 vol% of typical additives in gasoline. 1,2-Dimethoxyethane serves as internal standard.[11] The unlabeled peaks represent hydrocarbons. 1, methanol; 2, ethanol; 3, 2-propanol; 4, *t*-butanol; 5, 1-propanol; 6, methyl-*t*-butylether; 7, 2-butanol; 8, diisopropylether; 9, *i*-butanol; 10, ethyl-*t*-butylether; 11, 1,2-dimethoxyethane; 12, 1-butanol; 13, *t*-amyl-methylether.

Using selective absorbance reconstruction from the obtained IR spectra, results in a so-called selective absorbance (wavelength) chromatogram as an example are shown in Figure 3, with the most important components indicated as lines.[27] The reconstruction frequency for each compound is determined by the absorbance maximum in the region of interest, $\pm 4\ cm^{-1}$, using the C–O stretching region for alcohols and ethers and the aryl C–H out-of-plane bending region for aromatics, respectively.[28,29] Thus, selectivity over coeluting hydrocarbons is achievable. The separation shows sufficient resolution of all major compounds. ETBE and *i*-butyl alcohol can be separated more clearly by starting at a lower oven temperature and accepting an increased analysis time.

In conclusion, the combination of GC and FTIR yields good selectivity over hydrocarbons for C_1–C_4 alcohols and C_5/C_6 ethers using appropriate absorbance reconstruction frequencies. Diehl et al. have extended the applicable range of GC/FTIR to the quantification of BTEX (benzene, toluene, ethylbenzene, xylene) and 25 individual aromatic hydrocarbons, which presents an alternative method to the ASTM method D 3606 for benzene, in respect of the proposed ASTM approach using GC/MS for the determination of the aromatics.[30,31]

Since 1996 GC/FTIR has been approved as an ASTM standard (D 5986) as detailed in section 3.6.2.

3.5 Chemometrics

The majority of methods presented here take advantage of the exceptionally high information content of IR spectra in the range of 4000–600 cm^{-1}, without including methods combined with separation techniques. Using these, data fuel quality properties such as initial/final boiling point, density, viscosity, aromatic and oxygenates content, and so on can be predicted. In order to gain sufficient accuracy from these measurements, so-called chemometric data evaluation methods have turned out to be a valuable tool supporting data interpretation and automated evaluation of the spectra obtained.

Commonly, these methods require a set of calibration samples to train the system using their IR spectra, which are usually mean centered beforehand.[32] Subsequently, an external set of new test samples is necessary, which has to be uncorrelated with the calibration set, in order to assess the capabilities of the respective multivariate data evaluation method.[33] As a measure of the quality of the prediction provided by the model, the standard error of prediction (SEP) is generally used.[34]

The most commonly applied chemometric methods involve multiple linear regression (MLR), PLS methods or principal components regression (PCR), which are outlined below.

The best results with respect to SEP, precision, repeatability and reproducibility have been found for PLS treatment of IR spectroscopic data, which makes this data evaluation procedure most adequate for routine implementation in combination with IR techniques.[20,35–37]

3.5.1 Multiple Linear Regression

MLR determines the combination of independent prediction variables, which shows the best reproduction of the so-called dependent variables. The criterion used for selection of the best fit is generally the least squares criterion, resulting in a multiple linear correlation coefficient or the percentage of fit. The limitation of this technique can be seen in the problem of ensuring that all prediction variables are independent, because otherwise problems occur when entering the regression step.

A possible solution to this problem with respect to spectroscopic data is to apply methods, which remove a variable from and enter a variable into the set of prediction variables in a stepwise fashion, since both do not necessarily result in the same final set.

3.5.2 Principal Components Regression

PCR avoids redundant information by selecting only the primary dimension of information, the so-called variance, which is fitted by regression to the prediction variables. Irrelevant information, such as the noise in spectral data, remains in the final principal components, because only the primary principal components are usually fed stepwise into the regression model. In contrast to MLR, collinearity is avoided since principal components are respectively orthogonal. However, because the last principal components are not considered, the uncertainty remains that useful information for the prediction might be lost.

3.5.3 Partial Least Squares

PLS is probably the most interesting tool for evaluating spectroscopic data for fuel property prediction.[38,39] With this so-called canonical regression, the degree of relationship between a set of prediction variables and a set of resulting variables is determined. All the information from each source which is relevant to both sets of variables and avoids unrelated information is used, selecting the data by their predictive ability. Hence, frequent cross-validation is necessary to optimize and finally to select the PLS model with the best predictive characteristics.

Owing to achievable SEP, precision, repeatability and reproducibility, PLS treatment of IR spectroscopic data is the most useful chemometric method of automated data interpretation in order to yield standard fuel properties.

3.6 Mid-infrared Spectroscopy-based American Society for Testing and Materials Methods for Fuel Characterization

The acknowledgement of several IR-based methods for the determination of oxygenates and aromatics in fuels by ASTM attests to the fact that spectroscopic methods are gaining increasing importance in this area. The main advantages of these methods can be summarized as being cost effective, simple to run, portable and considerably faster than currently used systems.

3.6.1 Standard Test Method for the Determination of Oxygenates in Gasoline by Infrared Spectroscopy (D 5845-95)

This approved standard method[40] covers the determination of various oxygenates in gasoline (see Table 3) using either a transmission cell or an ATR configuration combined either with a filter-based MIR system, an FTIR spectrometer or a dispersive MIR apparatus.

Generally, the sample is introduced into a liquid sampling cell as shown schematically in Figure 2 and a beam of MIR radiation is imaged through the sample onto an appropriate detector. The spectral region used for data evaluation is either selected mathematically after recording a whole spectrum or in case of filter-based instruments by highly selective bandpass filters. Subsequent multivariate data analysis allows calculation of the concentration of the components of interest.

Table 3 Oxygenates determined by ASTM method D 5845

Substance	Abbreviation	Concentration range (mass%)
Methanol	MeOH	0.1–6
Ethanol	EtOH	0.1–11
t-Butanol	TBA	0.1–14
Methyl *t*-butyl ether	MTBE	0.1–20
Ethyl *t*-butyl ether	ETBE	0.1–20
t-Amyl methyl ether	TAME	0.1–20
Diisopropyl ether	DIPE	0.1–20

The driving force behind the introduction of MIR spectroscopy as standard method is the fact that it is comparably faster, simpler, less expensive and more portable than currently used methods based on GC.[2] Although this test method is not approved for testing for compliance with federal regulations, it applies well to quality control especially in the production of gasoline.

3.6.2 Standard Test Method for the Determination of Oxygenates and Aromatics in Finished Gasoline by Gas Chromatography/Fourier Transform Infrared (D 5986-96)

Using this test method[41] the quantitative determination of oxygenates and aromatics (see Table 4) in finished gasoline is enabled by combining GC with FTIR spectroscopy via a so-called light-pipe system as depicted schematically in Figure 2c.

This methods utilizes a methylsilicone wall-coated open tubular (WCOT) GC column interfaced to an FTIR spectrometer via a light pipe, enabling the compounds of interest to be detected by their characteristic absorption bands after separation in the column.

Table 4 Oxygenates and aromatics determined by ASTM method D 5986

Substance	Abbreviation	Concentration range (vol%)
Methanol	MeOH	0.1–20
Ethanol	EtOH	0.1–20
1-Propanol	1-PrOH	0.1–20
2-Propanol	2-PrOH	0.1–20
t-Butanol	*t*-BuOH	0.1–20
1-Butanol	1-BuOH	0.1–20
2-Butanol	2-BuOH	0.1–20
i-Butanol	*i*-BuOH	0.1–20
Methyl *t*-butyl ether	MTBE	0.1–20
Ethyl *t*-butyl ether	ETBE	0.1–20
t-Amyl methyl ether	TAME	0.1–20
Diisopropyl ether	DIPE	0.1–20
Benzene		0.1–2
Toluene		1–15
Total C_6–C_{12} aromatics		10–40

This method enables testing for oxygenates and aromatics simultaneously, without interference, for quality control and fuel regulation surveys. However, expensive equipment is mandatory owing to the hyphenation of two analytical techniques and the portability is limited.

3.6.3 Standard Test Method for the Determination of Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy (D 4053-98)

With this method[42] the percentage of benzene in full-range gasoline can be assessed in the concentration range from 0.1–5 vol%.

Using a thin film absorption flow cell that enables transmission measurements, benzene can be determined following a correction for interference with heavier aromatic compounds and comparison with calibration blends of known benzene concentration.

3.6.4 Standard Test Method for the Determination of Benzene in Spark-ignition Engine Fuels using Mid-infrared Spectroscopy (D 6277-99)

This ASTM method[43] covers the determination of the percentage of benzene in spark-ignition engine fuels addressing a concentration range of 0.1–5 vol%. Again, a liquid sample cell is applied using either an ATR or transmission cell configuration in combination with an FTIR spectrometer or a filter-based system.

This simple method enables rapid and simple detection of benzene in fuels at low cost and can be applied for quality control in the production of spark-ignition engine fuels.

3.7 Comparison with Other Spectroscopic Techniques

Apart from FTIR spectroscopy in the MIR region, several studies describe the application of NIR spectroscopy (0.9–2.5 μm) to the determination of gasoline and middle distillate fuel properties.[44–51] Furthermore, Fourier transform Raman spectroscopic methods have been used to investigate fuel specifications.[19,52–54]

An NIR spectrum generally results from overtone vibrations and combination tones of the fundamental molecular vibrations occurring in the MIR spectral range and can be considered to be up to a factor of 1000 weaker than the associated absorption band at the fundamental frequency. In contrast to the MIR region, usually broad and overlapping absorption bands have to be evaluated applying chemometric data evaluation almost exclusively. The combination of NIR spectroscopy with optical fibers enables such analysis systems to operate remotely and on-line without sample preparation, as shown for determination of MTBE in gasolines.[55] Owing to developments in the field of MIR fiberoptic sensors it

can be expected that this extension of MIR spectroscopy will be available for gasoline analysis in the near future.

Raman spectroscopy is generally a method well-suited to hydrocarbon analysis. Owing to the strong fluorescence of gasoline samples, dispersive Raman spectroscopy using a visible light source cannot be applied. Alternatively, Fourier transform Raman using an NIR laser, usually Nd:YAG at 1.06 μm, is well suited, because the fluorescence background caused by e.g. aromatics is sufficiently reduced. Fiberoptic Raman spectroscopy has been applied to characterizing petroleum products, determining RON, MON, density and the benzene content following a PLS data evaluation approach.[56] Although appropriate for fuel analysis Fourier transform Raman spectroscopy is still not widely used because of the cost of equipment, which is not yet routinely used or established in petroleum laboratories and controlling agencies.

A special application of NIR as well as Raman spectroscopy is reported by Choquette et al., who show the application of these methods to determine oxygenates in gasoline standard reference material (SRM) as provided, e.g. by NIST (National Institute of Standards and Technology). Since these SRMs are provided in sealed glass ampules, MIR spectroscopy cannot be used since these ampules are not transparent in the MIR spectral range. Both techniques have proven to be feasible for identifying the additives MTBE, ETBE, TAME and EtOH and for quantifying the oxygen concentration to an accuracy within 0.1% oxygen mass fraction in less than 1 min.[57]

Cooper et al. conclude in a comparison of MIR, NIR and Raman spectroscopy that particularly for the detection of MTBE in gasoline, MIR spectroscopy is the method of choice because of the distinct absorption pattern and pronounced absorption bands, especially in the fingerprint region.[19]

4 THE WAY INTO THE MARKET – SELECTED EXAMPLES

Several companies have already released ready to use MIR spectroscopic systems for the determination of gasoline properties, in response to the increasing analysis demand caused by the CAA.

Portable systems applying ASTM-approved MIR technologies (D 5845) have been introduced by Grabner Instruments (IROX™), MIDAC Corporation (FOX™) and PetroSpec (GS360™ and GS1000™) and are mentioned as examples of highly integrated and compact analytical tools. While the first two instruments utilize FTIR techniques, the latter represents a filter-based system. All three instruments provide the opportunity to determine selected fuel additives as listed in Table 2 as well as derived general fuel properties as summarized in Table 1 with sufficient accuracy and according to ASTM demands.

Owing to the low sample consumption (5–7 mL), a measurement time in the range of 1 min and systems that are fully automated, commercialized MIR instruments can be considered to be a serious alternative to laboratory bound conventional fuel analysis based on separation techniques, especially if low cost per sample and portability combined with sufficient accuracy are the requirements.

5 CONCLUSIONS

Owing to the developments in the field of MIR spectroscopy and the increasing demand for fuel properties to be determined by simple, accurate and portable techniques, novel systems have been established for the direct measurement of fuel additives, taking advantage of the specificity of molecular vibrations in the MIR spectral region. Apart from sum parameters such as oxygenates, aromatic and olefinic hydrocarbon content, selected compounds such as ethers, alcohols and benzene can be measured directly using MIR spectroscopic techniques in combination with microflow cells or ATR configurations. With the possible assessment of distinct analytes, general fuel parameters such as the octane number or physical properties can be derived applying chemometric data evaluation techniques.

Considering ASTM approvals, the wide range of applications and the commercial availability of compact and rugged MIR spectroscopic instruments at a competitive price it can be assumed that MIR spectroscopy for the determination of fuel performance characteristics will be established as a routine method in a wide range of applications in the petroleum industry as well as by regulatory agencies.

ABBREVIATIONS AND ACRONYMS

AED	Atomic Emission Detection
ASTM	American Society for Testing and Materials
ATR	Attenuated Total Reflection
BTEX	Benzene, Toluene, Ethylbenzene, Xylene
CAA	Clean Air Act
DIPE	Diisopropyl Ether
DTGS	Deuterated Triglycine Sulfate
ETBE	Ethyl *tert*-Butyl Ether

FIA	Flow Injection Analysis
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/O-FID	Gas Chromatography with Oxygen-selective Flame Ionization Detection
IP	Institute of Petroleum
IR	Infrared
IRS	Internal Reflection Spectroscopy
MCT	Mercury Cadmium Telluride
MIR	Mid-infrared
MLR	Multiple Linear Regression
MS	Mass Spectrometry
MTBE	Methyl *tert*-Butyl Ether
NIR	Near Infrared
NIST	National Institute of Standards and Technology
PCR	Principal Components Regression
PLS	Partial Least Squares
SEP	Standard Error of Prediction
SRM	Standard Reference Material
TAME	*tert*-Amyl Methyl Ether
TBA	Tertiary Butyl Alcohol
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
WCOT	Wall-coated Open Tubular

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Fuels Analysis, Regulatory Specifications for • Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Multivariate Calibration of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Gas Chromatography/Infrared Spectroscopy • Spectral Data, Modern Classification Methods for

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

REFERENCES

1. Clean Air Act Amendments, 42 U.S.C. 1857 *et seq.*, 1990.
2. D 5599, 'Test Method for Determination of Oxygenates in Gasoline by Gas Chromatography and Oxygen Selective Flame Ionization Detector', Annual Book of ASTM Standards, Vol. 05.03, 1995.
3. GC/O-FID EPA Test Method, 'Oxygen and Oxygenate Content Analysis', Code of Federal Regulations, Part 80 of Title 40, Section 80.46(g), Dec. 13 1993.
4. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectroscopy*, Wiley, New York, Vol. 83, 1965.
5. G.E. Fodor, K.B. Kohl, R.L. Mason, 'Analysis of Gasolines by FT-IR Spectroscopy', *Anal. Chem.*, **68**(1), 23–30 (1996).
6. R. Johnson, I. Pankow, D. Bender, C. Price, J. Zogorsky, 'MTBE. To What Extent Will Past Releases Contaminate Community Water Supply Wells?' *Environ. Sci. Technol.*, **9**, 2A–10A (2000).
7. G.R. Verga, A. Sironi, W. Schneider, Ch.J. Frohne, 'Selective Determination of Oxygenates in Complex Samples (Gasolines) with the O-FID (O-specific FID Analyser)', *J. HRC and CC*, **11**(3), 248–252 (1988).
8. E.A. McNamara, S.A. Montzka, R.M. Barkley, R.E. Silvers, 'Superconductor Metal Oxide Catalyst in a Chemiluminiscence Chromatography Detector', *J. Chromatogr.*, **452**, 75–83 (1988).
9. S.R. Goode, C.L. Thomas, 'Gas Chromatography Interfaced to Atomic Emission Spectrometry, Separation, Identification and Quantitation', *J. Anal. At. Spectrom.*, **9**, 73–78 (1994).
10. L.J. Bellamy, *The Infrared Spectra of Complex Molecules*, Chapman and Hall, London, 1980.
11. H.A. Rashid, S.B. Dekran, N.A. Fakhri, H.J. Aziz, 'Determination of Several Physical Properties of Light Petroleum Products Using IR', *Fuel Sci. Technol. Int.*, **7**, 237–250 (1989).
12. J.M. Anrade, S. Muniategui, P. López, D. Prada, 'Costs, Laboratory Safety, Productivity and Faster Research Octane Number and Motor Octane Number Determinations in Industrial Chemistry Laboratories', *Analyst*, **120**, 249 (1995).
13. B.R. Buchanan, D.E. Honigs, 'Calibration (for Pattern Recognition) in the Fiber-optic Region of the Near Infrared', *Appl. Spectrosc.*, **41**(8), 1388–1392 (1987).
14. G.E. Fodor, K.B. Kohl, 'Analysis of Middle Distillate Fuels by Midband Infrared Spectroscopy', *Energy Fuels*, **7**, 598–601 (1993)
15. G.E. Fodor, 'Analysis of Petroleum Products by Midband Infrared Spectroscopy', SAE (Society of Automotive Engineers) Technical Paper Series 941019, 1994.

16. N.J. Harrick, *Internal Reflection Spectroscopy*, Interscience, New York, 1967.
17. M. Morimoto, E. Nishio, 'Determination of Octane Number of Gasoline by Horizontal ATR/FT-IR Method', *Chem. Express*, **7**(6), 505–508 (1992).
18. F.X. Garcia, L.D. Lima, J.C. Medina, 'Detection of Methanol and MTBE in Gasoline by Infrared Using the Circle Cell and Multivariate Calibration', *Appl. Spectrosc.*, **47**(7), 1036–1039 (1993).
19. J.B. Cooper, K.L. Wise, W.T. Welch, R.R. Bledsoe, M.B. Sumner, 'Determination of Weight Percent Oxygen in Commercial Gasoline: A Comparison Between FT-Raman, FT-IR and Dispersive Near-IR Spectroscopies', *Appl. Spectrosc.*, **50**(7), 917–921 (1996).
20. J.M. Andrade, S. Muniategui, D. Prada, 'Prediction of Clean Octane Numbers of Catalytic Reformed Naphthas Using FT-MIR and PLS', *Fuel*, **76**(11), 1035–1042 (1997).
21. M. Gallignani, S. Garrigues, M. de la Guardia, 'Direct Determination of Benzene in Gasoline by Flow-injection Fourier Transform Infrared Spectrometry', *Anal. Chim. Acta*, **274**, 267–274 (1993).
22. M. Gallignani, S. Garrigues, M. de la Guardia, 'Comparative Study of Different Approaches for the Flow Injection–Fourier Transform Infrared Determination of Toluene in Gasolines', *Analyst*, **119**, 653–657 (1994).
23. M. Gallignani, S. Garrigues, M. de la Guardia, J.L. Burguera, M. Burguera, 'Comparative Study of Different Approaches for the Flow Injection–Fourier Transform Infrared Determination of Toluene in Gasolines', *Talanta*, **41**(5), 739–745 (1994).
24. E. López-Anreus, S. Garrigues, M. de la Guardia, 'Vapour Generation Fourier Transform Infrared Spectrometric Determination of Benzene Toluene and *t*-Butyl Methyl Ether in Gasolines', *Anal. Chim. Acta*, **333**, 157–165 (1996).
25. F.P. DiSanzo, 'Determination of Oxygenates in Gasoline Range Hydrocarbons by Capillary Column GC: A User's Experience with an Oxygen Specific Detector', *J. Chromatogr. Sci.*, **28**, 73–74 (1990).
26. K.J. Slatkavitz, P.C. Uden, R.M. Barnes, 'Consideration of an Atmospheric Pressure Microwave Induced Helium Plasma as an Oxygen Selective Gas Chromatographic Detector', *J. Chromatogr.*, **355**, 117 (1986).
27. J.W. Diehl, J.W. Finkbeiner, F.P. DiSanzo, 'Determination of Ethers and Alcohols in Gasolines by Gas Chromatography/Fourier Transform Infrared Spectroscopy', *Anal. Chem.*, **64**, 3202–3205 (1992).
28. J.A. de Haseth, T.L. Isenhour, 'Reconstruction of Gas Chromatograms from Interferometric Gas Chromatography/Infrared Spectrometry Data', *Anal. Chem.*, **49**, 1977–1981 (1977).
29. I.C. Bowater, R.S. Brown, J.R. Cooper, C.L. Wilkins, 'Maximum Absorbance Algorithm for Reconstruction of Gas Chromatograms from Gas Chromatography–Infrared Spectrometry Data', *Anal. Chem.*, **58**, 2195 (1986).
30. J.W. Diehl, J.W. Finkbeiner, F.P. DiSanzo, 'Determination of Benzene, Toluene, Ethyl Benzene and Xylines in Gasoline by Gas Chromatography-Deuterium Isotope–Dilution Fourier Transform Infrared Spectroscopy', *Anal. Chem.*, **65**, 2493–2496 (1993).
31. J.W. Diehl, J.W. Finkbeiner, F.P. DiSanzo, 'Determination of Aromatic Hydrocarbons in Gasolines by Gas Chromatography/Fourier Transform Infrared Spectroscopy', *Anal. Chem.*, **67**, 2015–2019 (1995).
32. M.B. Seasholtz, B.R. Kowalski, 'The Effect of Mean Centering on Prediction in Multivariate Calibration', *J. Chemom.*, **6**, 103–111 (1992).
33. O.M. Kvalheim, T.J. Karstang, 'Interpretation of Latent-variable Regression Models', *Chemom. Intel. Lab. Syst.*, **7**, 39–51 (1989).
34. C.L. Erickson, M.J. Lysaght, J.B. Callis, 'Relationship Between Digital Filtering and Multivariate Regression in Quantitative Analysis', *Anal. Chem.*, **64**, 1155A (1992).
35. K.A. Caswell, T.E. Glass, M. Swann, H.C. Dorn, 'Rapid Prediction of Various Physical Properties for Middle Distillate Fuels Utilizing Directly Coupled Liquid Chromatography/Proton Nuclear Magnetic Resonance', *Anal. Chem.*, **61**, 206–211 (1989).
36. J.J. Kelly, J.B. Callis, 'Non-destructive Analytical Procedure for Simultaneous Estimation of the Major Classes of Hydrocarbon Constituents of Finished Gasolines', *Anal. Chem.*, **62**, 1444–1451 (1990).
37. S. Garrigues, J.M. Andrade, M. de la Guardia, D. Prada, 'Multivariate Calibrations in Fourier Transform Infrared Spectrometry for Prediction of Kerosene Properties', *Anal. Chim. Acta*, **317**, 95–105 (1995).
38. M. Sjöström, S. Wold, W. Lindberg, J.A. Persson, H. Martens, 'Multivariate Calibration Problem in Analytical Chemistry Solved by Partial Least Squares Models in Latent Variables', *Anal. Chim. Acta*, **150**, 61–70 (1983).
39. H. Martens, T. Naes, *Multivariate Calibration*, Wiley, New York, 1989.
40. D 5845, 'Standard Test Method for Determination of MTBE, ETBE, DIPE, MeOH, EtOH and *t*-Butanol in Gasoline by Infrared Spectroscopy', Annual Book of ASTM Standards, 1995.
41. D 5986, 'Standard Test Method for Determination of Oxygenates, Benzene, Toluene, C_8–C_{12} Aromatics and Total Aromatics in Finished Gasoline by gas Chromatography/Fourier Transform Infrared Spectroscopy', Annual Book of ASTM Standards, 1996.
42. D 4053, 'Standard Test Method for Determination of Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy', Annual Book of ASTM Standards, 1998.
43. D 6277, 'Standard Test Method for Determination of Benzene in Spark-ignition Engine Fuels Using Mid Infrared Spectroscopy', Annual Book of ASTM Standards, 1999.

44. P.M. Frederics, J.B. Lee, P.R. Osborn, D.A.J. Swinkles, 'Materials Characterisation Using Factor Analysis of FT-IR Spectra. I Results', *Appl. Spectrosc.*, **39**, 311 (1985).
45. M.J. Lysaght, J.A. van Zee, J.B. Callins, 'Laptop Chemistry: A Fiber-optic, Field-portable, Near-infrared Spectrometer', *Rev. Sci. Instrum.*, **62**, 507–515 (1991).
46. S.J. Swarim, C.A. Drumm, 'Prediction of Gasoline Properties with Near Infrared Spectroscopy and Chemometrics', SAE (Society of Automotive Engineers) Paper No. 912390, 1991.
47. N.R. Crawford, W.W. Hellmuth, D.H. Marcellus, K. Chou, 'Predicting Research Octane Number from Near Infrared Absorbance Data with Neural Networks', *J. Process Control Qual.*, **4**, 13 (1992).
48. S.R. Westbrook, 'Army Use of Near Infrared Spectroscopy to Estimate Selected Properties of Compression Ignition Fuels', SAE (Society of Automotive Engineers) Technical Paper Series 930734, 1993.
49. R. DiFoggio, M. Sadhukhan, M.L. Ranc, 'Near-infrared Offers Benefits and Challenges in Gasoline Analysis', *Oil Gas J.*, **91**, 87–90 (1993).
50. D. Lambert, B. Descales, J.R. Llinas, A. Espinosa, S. Osta, M. Sanchez, A. Martens, 'Online NIR Monitoring and Optimization for Refining and Petrochemical Processes', *Analusis Mag.*, **23**(4), 9–13 (1995).
51. D. Lambert, B. Descales, A. Espinosa, M. Sanchez, S. Osta, J. Gil, A. Martens, M- Valleur, 'NIR Online Advanced Control System for Gasoline Blender', *Analusis Mag.*, **23**(4), 20–25 (1995).
52. M.B. Seasholtz, D.D. Archibard, A. Lorber, B.R. Kowalski, 'Quantitative Analysis of Liquid Fuel Mixtures with the Use of Fourier Transform Near IR Raman Spectroscopy', *Appl. Spectrosc.*, **43**(6), 1067–1072 (1989).
53. J.B. Cooper, P.E. Fletcher, T.M. Vess, W.T. Welch, 'Remote Fiber-optic Raman Analysis of Xylene Isomers in Mock Petroleum Fuels Using a Low-cost Dispersive Instrument and Partial Least-squares Regression Analysis', *Appl. Spectrosc.*, **49**, 586–592 (1995).
54. J.B. Cooper, K.L. Wise, J. Groves, W.T. Welch, 'Determination of Octane Numbers and Ried Vapour Pressure of Commercial Petroleum Fuels Using FT-Raman Spectroscopy and Partial Least Squares Regression Analysis', *Anal. Chem.*, **67**, 4096–4100 (1995).
55. B. De Simas, L. Lane, 'Measurement of MTBE in Gasolines by NIR', in *ISA Proceedings*, Paper No. 90-462, 451–461, International Society for Measurement and Control, 1990.
56. C.J. de Bakker, P.M. Fredericks, 'Determination of Petroleum Properties by Fiber-optic Fourier Transform Raman Spectrometry and Partial Least-squares Analysis', *Appl. Spectrosc.*, **49**(12), 1766–1771 (1995).
57. S.J. Choquette, S.N. Chesier, D.L. Duewer, S. Wang, T.C. O'Haver, 'Identification and Quantitation of Oxygenates in Gasoline Ampules Using Fourier Transform Near-infrared and Fourier Transform Raman Spectroscopy', *Anal. Chem.*, **68**, 3525–3533 (1996).

Fuels Analysis, Regulatory Specifications for

Peter A. Nick
Process Systems Engineering, Yorba Linda, USA

1 Introduction **2**

2 The Regulatory Legacy for Petroleum Fuels **3**
2.1 History of Fuels Formulations 3
2.2 History of Fuels Emissions Reductions 4
2.3 History of Reformulated Fuels 5
2.4 Regulatory Requirements for Diesel Fuels 7
2.5 Future Regulatory Trends 9

3 Chemical Components of Petroleum-derived Fuels **9**
3.1 Crude Oil Chemical Mixture Components 9
3.2 Typical Gasoline Mixture Components 11
3.3 Components in Heavier Fuels 13

4 Petroleum-derived Fuels Performance Criteria **13**
4.1 Spark-ignition Engines (Gasoline) 13
4.2 Aviation Gasoline 16
4.3 Compression–Ignition Engine (Diesel) Fuels 17
4.4 Aviation Turbine/Jet/Marine Fuels 19

5 Sources and Formulation of Transport Fuels **20**
5.1 Refinery Configuration 20
5.2 Fuels Blending 22
5.3 Advanced Control and Monitoring of Gasoline Blending 24

6 Fuels Analysis for Regulatory Requirments **28**
6.1 Introduction 28
6.2 Analytical Procedures – Component Separation Methods 28
6.3 Electromagnetic Field Spectroscopy 31
6.4 Mass Spectrometry 38
6.5 Volatility Measurements 39

7 Applicable Regulatory Measurement Standards by Criteria Component **41**
7.1 Introduction 41
7.2 Choosing the Appropriate Standards 42
7.3 Distillation Methodology 42
7.4 *V*/*L* Methodology 44
7.5 Measurement of Reid Vapor Pressure – Gasoline Products 44
7.6 Detection and Measurement of Aromatics – Gasoline/Avgas Cuts 45
7.7 Detection and Measurement of Aromatics – Turbine and Diesel Fuels 47
7.8 General Ponaffin, Isoponaffin, Aromatic, Naphthene, Olefin Analyses for Aromatics and Other Components 47
7.9 Detection and Measurement of Benzene 48
7.10 Detection and Measurement of Olefins 49
7.11 Detection and Measurement of Sulfur 50
7.12 Detection and Measurement of Oxygenates in Fuel 51
7.13 Detection and Measurement of Nitrogen in Fuels 52
7.14 Detection and Measurement of Metals in Fuels 53
7.15 Measurement of Fuel Densities 53

8 On-line Analyses for Regulatory Control **53**
8.1 General 53
8.2 On-line Analysis for Blender Control – Chemometric Methods 53
8.3 On-line Analysis for Blender Control – Gas Chromatography Methods 56

9 Maintaining Quality Control of On-line Analyses **58**
9.1 Blend Certification Sampling and Testing Procedures 58
9.2 Infrared Analysis Preparation 59
9.3 Calibration and Acceptance Testing of Gas Chromatography/Simulated Distillation Analyzers for D86/Reid Vapor Pressure Analysis 61

10 Reformulated Fuels Analysis Glossary **64**
10.1 Introduction 64

Abbreviations and Acronyms **71**

References **72**

This article provides a general survey of the chemical and physical analyses of light and mid-range hydrocarbon

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

fuels needed to satisfy the environmental and safety requirements of the refining and transport industries. Since the inception of the oil and automobile industries, emissions from vehicles and from fuel refineries have taken a back seat to the performance factors of the fuels themselves. In the middle of the 1900s, new environmental requirements became a commercial consideration as governments, and to a lesser extent, industrial consortia, began to impose requirements on the fuels which had little or nothing to do with performance in the engine. This article will attempt to enumerate the most prevalent of these testing specifications and provide a modest amount of information about what they involve and why they exist. At the time the original draft of this work was written in 1999, the imposition of several new regulatory criteria in the United States had just been imposed on most of the fuels production companies. Few new regulations have been enacted *in the last few years (1995+) prior to this article's printing, but there have been numerous serious proposals for such new legislation and several more geographical areas have come under existing and more stringent legislation. We have tried to assess where future regulations will lead and what specifications and tests might deal with these limits.*

In general, we will not deal here with the engine performance specifications. These are detailed far more adequately elsewhere in the literature. This article will not provide a long treatise on the formulation of transportation fuel manufacturing processes, and neither will it delve deeply into the testing methodology. Enough of these three subjects will be presented to allow the reader to put the regulatory specifications and what they entail into perspective. The emphasis is slanted towards the specification of gasoline range fuels, primarily caused by the great predominance of these commodities, but we do try to cover the heavier transport fuels in detail as well. There is considerable overlap in several of the specifications, at least in terms of methodology, even if the specification document number perhaps differs from light to heavier fuels.

The various testing specifications are presented by a fairly succinct description and are categorized by test type and by species analysis requirements. An attached glossary will enable the reader to become acquainted with the jargon of emissions compliance and refinery technology, since these are both linked to intelligent discussion of fuels production and usage.

1 INTRODUCTION

Within the United States, about 60% of every barrel of crude oil processed in a typical refinery is converted into motor gasoline and another 15% into alternative transportation fuels such as jet fuel, turbine fuel and diesel fuel.[1–3] Performance quality and contractual specification standards are many and exacting for the final product as it is carried to the retail market. These standards have been defined by more than 80 years of testing and experimentation on all kinds of combustion engines in all sorts of worldwide climatological and geographical environs.

This article is *not* about these types of standards. They are described elsewhere in the technical literature in great and accurate detail. They are mentioned here in passing only to remind ourselves that in spite of all other requirements imposed on the transportation fuel markets by regulatory agencies, the major reason for these fuels to exist is their ability to meet the specifications for modern high performance internal combustion (IC) engines. Neither is this article a detailed report about the process of making environmentally friendly fuels. Enough of this topic is presented here to allow the reader to place the types and needs of various analyses and tests within the proper context of a modern refinery but much on this subject will be missing and for a full treatise on the gasoline manufacturing process, the reader would best be served by other references.

What will be attempted here is to provide a general survey of the chemical and physical analyses of light and mid-range hydrocarbon fuels needed to satisfy the environmental and safety requirements of the refining and transport industries. Since the inception of the oil and automobile industries, emissions from cars and from fuel refineries took a back seat to the performance factors of the fuels themselves. In the mid-1960s, however, as many cities worldwide began to choke on combustion-derived air pollutants, new requirements became a part of the commercial scene as governments, and to a lesser extent, industrial consortia, began to impose requirements on the fuels which had little or nothing to do with performance in the engine.

The appropriate analytical procedures for detecting regulated chemical quantities in the laboratory and on-line in the manufacturing process will also be covered. This will not be a comprehensive manual on the analysis procedures themselves. It will not, for example, describe how to run a gas chromatograph or fluorescent indicator absorption (FIA) unit in great detail. This information, again, is well detailed in the literature and readily available to the laboratory and control instrumentation engineers and technicians who design, install, run and maintain this type of equipment. Certain exceptions will be made, to the extent of making important caveats known or of providing a specific emphasis directly related to the analytical need at hand.

The major emphasis will be on the regulatory analysis requirements for gasolines but also discussed will be jet/

turbine and diesel fuels, for completeness of coverage. As a rule, the lighter the molecular weight or gravity of the liquid, the more and easier will be the analytical choices. The heavier the fuel, the fewer simple methods are feasible.

Sections 2 through 5 will cover background and related material. It is important to consider these topics when performing analyses for control of regulatory components in fuels. The fuels must be sold so the performance characteristics must be considered, even when they are not necessarily the controlling factors in the blend as perhaps they once were. Section 6 will discuss the methodology of testing. There are numerous techniques available to the laboratory analyst and the on-line instrument designer. Most give adequate results; some give better results in one circumstance and worse in others. Where applicable such discrepancies and caveats will be mentioned. Section 7 presents the myriad of available standards and specifications that cover the realm of regulatory analysis requirements. The great emphasis will be on the American Society for Testing and Materials (ASTM) standards with several American National Standards Institute (ANSI) and International Standards Organization (ISO) standards mentioned as well. These compose 95% of the world's governing standards at this time and are usually in agreement with one another on most matters. Thus we give them emphasis, rather than unintentionally exclude any other local standards that may apply. This work is completed with a glossary of commonly used terms, that might be encountered in the fuels regulation world with emphasis on governmental acronyms, technical terms, and consortia and trade association abbreviations.

References to these three organizations and their publications of interest will not be listed as the mention of each standard is fairly obvious. The reader should note that revisions to each standard are issued on a fairly regular basis, if not each year. Thus the inclusion of specific yearly issue dates has been omitted and the reader must take care to obtain the latest version from each organization.

2 THE REGULATORY LEGACY FOR PETROLEUM FUELS

2.1 History of Fuels Formulations

Fuel driven engines of any sophistication have been available only since the mid-1880s. Several early prototypes were indirect drive – the Stanley steamer, for example – and could theoretically use any type of combustible fuel such as coal or wood. By 1890, however, the IC engine won the battle for acceptance.

At this time the most suitable fuels for the automobile were either coal tar distillates or the lighter fractions from the distillation of crude oil. During the early 20th century, oil companies were producing gasoline as a simple petroleum distillate (''white oil'') without benefit of stabilization (removal of butanes and lighter components). Automotive engines before long were rapidly being improved and required a more suitable fuel. During the 1910s, laws prohibited the storage of these unstable gasolines on residential properties, so several automotive manufacturers modified IC engines to run on kerosine. However, the kerosine-fueled engine would ''knock'' and crack the cylinder head and pistons. Research at the time demonstrated that the knock was caused by a rapid rise in pressure after ignition, not during preignition as then believed. This lead to the search for several types of antiknock agents, culminating in the common usage of tetraethyl lead for this purpose. Typical mid-1920s gasolines were 40–60 octane.[4–6]

By the 1930s, the petroleum industry had determined that the larger hydrocarbon molecules (kerosine) had major adverse effects on the octane of gasoline, and were developing consistent specifications for desired properties. In the 1940s, catalytic cracking processes were introduced,[7] and gasoline compositions became fairly consistent between brands during the various seasons.

The development of the alkyl leads (tetramethyl lead, tetraethyl lead) and the toxic halogenated scavengers (to keep the lead from plating out on valves and piston seats) meant that petroleum refiners could configure their refineries to produce hydrocarbon streams that would increase octane with small quantities of alkyl lead.[5,8] By adding increasing amounts of alkyl lead compounds, the lead response (positive octane increase) of the gasoline decreases, and so there are economic limits to how much lead should be added. The 1950s saw an increase in the cylinder compression ratios,[9] requiring even higher octane fuels. Lead levels were increased, and some new refining processes (such as hydrocracking), were specifically designed to provide hydrocarbons components with good lead response and octane.[10]

Well into the late 1960s, alkyl leads were added to gasolines for octane boost. The upper limit was $1.14\,g\,Pb\,L^{-1}$, which is well above the ''diminishing returns'' part of the lead response curve for most refinery streams. It is unlikely that much fuel was ever made at that level. Published articles on the subject usually set $1.05\,g\,Pb\,L^{-1}$ as about the economic maximum, and other literature suggests that 1970 100 research octane number (RON) premiums had about $0.7–0.8\,g\,Pb\,L^{-1}$ and 94 RON regulars $0.6–0.7\,g\,Pb\,L^{-1}$, which matches published lead response data.

Table 1 Lead response characteristics of typical refinery cut

Tetraethyl lead ($g\,L^{-1}$)	Catalytic reformate RON	Straight run naphtha RON
0	96	72
0.1	98	79
0.2	99	83
0.3	100	85
0.4	101	87
0.5	101.5	88
0.6	102	89
0.7	102.5	89.5
0.8	102.75	90

Table 1 shows the typical lead response of two basic refinery blend stocks.

Except for the lead technology and the introduction of saturated hydrocracked and alkylated stocks, only minor improvements were made for several years to basic gasoline formulations. The bulk of the automotive research effort was applied to engine design and performance optimization. Most fuel changes were implemented primarily to improve yields and octane, until the 1970s, when unleaded fuels were introduced to protect the exhaust catalysts that were also being introduced for environmental reasons. From 1970 until 1990, gasolines and corresponding refinery processes were slowly changed as lead was phased out. In 1990 the Clean Air Act (CAA) started forcing major and more rapid compositional changes in gasoline. These changes will continue into the 21st century because gasoline combustion is and still will be a major air pollution source.[11]

As the toxicity of the alkyl lead components and the halogenated scavengers became of increasing concern, alternatives were considered. The most famous of these is methylcyclopentadienyl manganese tricarbonyl (MMT), which was used in the US until banned by the United States Environmental Protection Agency (US EPA) in 1978, but is approved for use in Canada and Australia until the year 2000. It is more expensive than alkyl leads and has been reported to increase unburned hydrocarbon emissions and block exhaust catalysts. Other compounds that enhance octane have been suggested, but usually have significant problems such as toxicity, cost, increased engine wear, etc. Other such compounds include dicyclopentadienyl iron and nickel carbonyl.

Kerosine cuts have had far less change than gasolines, at least in terms of postproduction processing. Most of them are still taken from ''straight run'' streams and perhaps stabilized (redistillation) to reduce flash point for storage reasons. Though they briefly had a fling as spark engine fuels, the lighter gasoline stocks quickly took over this role. Kerosine was relegated to lamps, camp stoves, smudge pots, and to a few ''spray type'' furnaces using aerosol-forming burners. With the widespread introduction of turbine-type engines in the late 1940s, this cut became a favorite economic fuel, especially for jet engines. The product specifications for kerosine/jet/turbine fuels became more exacting but still less complex than those of gasoline.[3,8]

Diesel fuel stocks vary somewhat according to their gravity but are still produced from crude unit distillation and, since the late 1960s, from hydrocracker units. Because the compression–ignition (CI) cycle engines that they serve are far less complicated in terms of combustion initiation, less is expected of this refinery product's specification, primarily the removal of light naphtha from the front end of the distillation curve, the cloud point (for cold weather nascent wax solidification) and pour point concerns and its cetane index, diesel's analog to gasoline octane for ignition control. The relative volume usage of diesel in the commercial market is determined by heavy truck and rail needs. Advances in engine size reduction as well as in sheer numbers of these engines in the transport sector stimulated the diesel supply chain to create an expansion in mid-distillate demand. This demand was prompted by use in several diesel-fueled models of passenger cars as a response to the political disruptions in the gasoline fuel supply in the 1970s. This latter market demand factor was quickly attenuated by the 1980s environmental regulatory sector which objected to the aromatics and particulate contents of diesel engine emissions.

2.2 History of Fuels Emissions Reductions[12–17]

With the widespread post-World War II deployment of gasoline-driven automobiles, kerosine-driven jets, turbine-drive generators, and diesel-powered trucks and trains, the types and levels of fuel combustion by-products began to affect the quality of life in several areas of the developed world. Such pollutant emissions were most notable where the combustion engine was most concentrated, in large urban areas.

Cars emit several pollutants as combustion products from the tailpipe (tailpipe or exhaust emissions), and as losses due to evaporation (evaporative emissions, refueling emissions). The volatile organic carbon (VOC) emissions from these sources, along with nitrogen oxides (NO_x) emissions from the tailpipe, will react in the presence of ultraviolet (UV) light (wavelengths of less than 430 nm) to form ground-level (tropospheric) ozone, which is one of the major components of photochemical smog.

Smog has been a major pollution problem ever since coal-fired power stations were developed and domestic trash (rubbish) burns were permitted in urban areas, but these emissions were the first to be addressed and reduced

in the early 1960s. Tailpipe and crankcase emissions were responsible for the majority of ozone-forming pollutants in the late 1960s. Ozone levels in the Los Angeles basin, as a significant example, reached 450–500 ppb in the early 1970s, well above the typical background of 30–50 ppb.[18]

The first of several air pollution control acts was passed by the US Congress in 1962. While its emphasis was on industrial smokestacks and trash incinerators, the principle of government control of emissions was established and the formation of several urban and state coordinated ''smog centers'' were given legal and financial encouragement.

The CAA of 1990, now administered by the US EPA, is the primary statute governing air quality in the United States. It has been amended several times since it was first enacted in 1967. Initially, the CAA focused on vehicle-related controls, such as exhaust catalytic convertors, exhaust gas recirculation (EGR), and crankcase carbon adsorption canisters to reduce ozone-forming (NO_x and unburned hydrocarbons) and carbon monoxide emissions from automobiles.

The automotive manufacturing industry responded with major advances in emission control technology.[19] The total challenges involved were significant. The design of the engine can have very significant effects on the type and quantity of pollutants. Unburned hydrocarbons in the exhaust originate mainly from combustion chamber crevices, such as the gap between the piston and cylinder wall, where the combustion flame cannot completely use the hydrocarbons. The type and amount of unburned hydrocarbons are related to the fuel composition (volatility, olefins, aromatics, final boiling point (FBP)), as well as state of tune, engine condition, and age/condition of the engine lubricating oil.

It soon became apparent that *major* design changes to the typical automobile engine could no longer be avoided. Tuning a carbureted engine, for example, has only a marginal effect on pollutant levels. More was required, so exhaust catalysts were offered as a postengine solution that could ensure pollutants were oxidized. As engine management systems and fuel injection systems have developed and become more sophisticated, the volatility properties of the gasoline have been tuned to minimize evaporative emissions yet maintain low-exhaust emissions. Since the late 1970s, gasoline vehicle emission rates have been cut by about 80%. Overall vehicle emissions will continue to drop each year as older cars are replaced by newer ones.[20–22]

Because leaded antiknock compounds were found to destroy the performance of catalytic converters then under development, US passenger car manufacturers in 1971 began to build engines designed to operate satisfactorily on gasolines of nominal 91 RON. Some of these engines were designed to operate on unleaded fuel, while others required leaded fuel or the occasional use of leaded fuel. The 91 RON was chosen in the belief that enough unleaded gasoline at this level could be made available in the current refinery base.

Beginning with the 1975 model year, most new car models were equipped with catalytic exhaust treatment devices in addition to all the previously mentioned crankcase emissions control technology. The need for gasolines that would not adversely affect such catalytic devices has led to the large-scale availability and growing use of unleaded gasolines.

There was a further reason why alkyl lead compounds were subsequently reduced, and this was the growing recognition of the potentially toxic nature of the emissions from a leaded gasoline-fueled engine. Not only were toxic lead emissions produced, but the added lead scavengers (ethylene dibromide and ethylene dichloride) could react with hydrocarbons to produce toxic organohalogen emissions such as dioxin. Even if catalysts were removed, or lead-tolerant catalysts discovered, alkyl lead compounds would most likely remain banned because of their toxicity and toxic emissions.

2.3 History of Reformulated Fuels[11,12,15,17,20,23–25]

In spite of the emissions control advances, air quality did not improve enough in certain metropolitan areas where the impact of emissions was severe. Lead removal itself did nothing to improve air emissions control, it just kept the catalytic converters working longer. The US Congress turned to fuel formulation changes as another means of reducing emissions. As tailpipe emissions were reduced by the use of improved exhaust emission control systems, the hydrocarbons produced by evaporation of the gasoline during distribution, vehicle refueling, and from the vehicle, become more and more significant. A recent European study found that 40% of artificial VOCs came from vehicles. Many of the problem hydrocarbons are, unfortunately, aromatics and olefins that have relatively high octane values. Removing them in major quantities presented some major changes in the gasoline pool that resulted in potential engine performance problems.[26]

The CAA Amendments of 1991[13,27,28] required the US EPA to adopt regulations for reformulated gasolines (RFGs) in the USAs nine worst ozone nonattainment areas. (''Nonattainment areas'' are areas which do not meet one or more of the air quality standards established by the US EPA.) If they choose, less severe ozone nonattainment areas across the USA have the option to participate in the Federal RFG program. To date, the US EPA has received participation requests from several states, primarily in the northeast USA, and the District

of Columbia. Other less severe ozone nonattainment areas have adopted other approaches, such as limiting gasoline vapor pressure as a more cost-effective means of improving air quality. Almost every state and major metropolitan area in the United States now has its own air quality management district (AQMD).

The program requires numerous changes in fuel constituents to reduce ozone-forming hydrocarbons and toxic constituents in air. Toxic air constituents, as defined in the CAA, are benzene, butadiene, formaldehyde, acetaldehyde, and polycyclic organic matter. The US EPA has established specific limits for some of these changes. For others, the US EPA developed the emissions reduction schedule summarized in Table 2. The goals will be implemented in two phases to mitigate the effect of the necessary processing changes on the petroleum industry. Similar European Union (EU) protocols are given in Table 3.

Phase 1 (begun on January 1, 1995) is required to achieve a reduction of 15–17% of ozone-reforming hydrocarbons during the high-ozone season and an annual reduction of 15–25% of air toxic emissions from vehicles fueled with RFG. For example, with methyl tertiary butyl ether (MTBE)–RFG, benzene and butadiene emissions decrease, whereas formaldehyde and MTBE emissions increase, all within the context of a net decrease in hydrocarbons.

Phase 2, which began January 1, 2000 (except in highly noncompliant areas such as California where it has been law since 1996) is required for most of the US to achieve an annual reduction of 25–29% in ozone-forming hydrocarbon emissions, a 5–7% reduction in ozone-reforming nitrogen oxide emissions, and a 20–22% reduction in air toxic emissions. The US EPA gives refinery planners the option of meeting a ''per

Table 2 Requirements of the Federal CAA and the CARB

	Average 1990	1995 Federal CAA	CARB
Benzene (vol. %)	2	1 maximum	1.0 maximum Average: 0.8 w/1.2 cap
Oxygen (mass %)	0.2	2 minimum	1.8–2.2
Sulfur (ppm)	150, 347[58]	No increase	40 Average: 30 w/80 cap
Aromatics (vol. %)	32.0	25 maximum	25 maximum Average: 22 w/30 cap
Olefins (vol. %)	9.9	5 maximum	6 vol. % maximum Average: 4 w/10 cap
RVP (kPa) (psi)	60 (8.7)	56 (north) (8.2) (north) 50 (south) (7.3) (south)	48 (7.1)
90% Evaporated (ASTM D-86 – T90)	170 °C (338 °F)	No increase	149 °C (330 °F) max Average: 143 °C w/157 °C cap (290 °F)/(315 °F)
50% Evaporated (ASTM D-86 – T50)	103 °C (218 °F)	No increase	99 °C (210 °F) max Average: 93 °C w/104 °C cap (200 °F)/(220 °F)

CARB, California Air Resources Board; RVP, Reid vapor pressure.

Table 3 Proposed requirements for EU market gasolines[a]

	EU Commission limits–min	Max *t*	EU–Parliament Limits – 2000
Benzene (vol. %)	–	2.0	1.0 maximum
Oxygen (mass %)	–	2.3	2.7 max
Sulfur (ppm)	–	200 max	30 max
Aromatics (vol. %)	–	45	35 (30 in 2005)
Olefins (vol. %)	–	18	10% maximum
Summer RVP (kPa)		60	60
% Evaporated at 100 °C (vol. %)	46	–	51 min
% Evaporated at 150 °C (vol. %)	75	–	80 min

[a] Recent revisions to these codes have allowed for seasonality. The EN228 standard now controls RVP.
EU, European Union.

gallon'' standard or an ''averaged'' standard; in the latter case, the refiner measures the average values of their annual production against the limits. US EPA also provided two emissions level calculations (simple and complex models) to determine the levels of criteria pollutants that a particular refiner's sales block will generate.

The simple and complex models[28,29] are collections of equations and specifications which relate the emissions produced by a specific batch of gasoline, when it is used in a typical vehicle, to its composition and properties. As the names suggest, the simple model involves fewer gasoline properties than the complex model. The models are the mechanism which provides formulation flexibility to refiners, while ensuring that they meet the emission goals. The models are based on measured emissions from a variety of vehicles for a range of gasolines with different properties and/or compositions. Since different vehicles emit different amounts of pollutants, the result obtained by applying the model is an average for the total vehicle population. In 1998, most US refiners were required to use the complex model. In the three-year period 1995–1997, refiners were allowed to use either model.

As well as the US EPA requirements mentioned in this section, there are various restrictions introduced by the CAA and state bodies such as the CARB that often have more stringent limits for the above properties, as well as additional limits. The CAA specifies that some regions that exceed air quality standards had to use RFGs all year, starting January 1995. Other regions were required to use oxygenated gasolines for four winter months, beginning November 1992. The RFGs also contain oxygenates. Metropolitan regions with severe ozone air quality problems were required to use RFGs in 1995 that contain at least 2.0 wt% oxygen, reduce 1990 VOC compounds by 15%, and reduce specified toxic emissions, first by 15% (1995), then 25% (2000). Metropolitan regions that exceeded carbon monoxide limits were required to use gasolines with 2.7 wt% oxygen during winter months, starting in 1992.

Because phosphorus adversely affects exhaust catalysts, the 1990 US EPA regulations also limited phosphorus in all gasolines to 0.0013 g Pb L^{-1}, in effect making this a criteria fuels component as well.[22]

Sulfur compliance is currently (1999) around 350 ppm, well above Phase II specifications. In some regions, the average emissions are nearer 500 ppm. Large refiners will have to meet CARB-like specifications by 2004 and smaller refiners between 2006 and 2010 depending upon which piece of legislation is consulted.

California, the birthplace of RFG, is now planning to ban methyl tertiary butyl ether (MTBE)[30] as a danger to groundwater supplies. Unfortunately, no alternative ''official'' specification replaces it, and the matter currently resides in several court venues.

2.4 Regulatory Requirements for Diesel Fuels

As is the case with many environmental requirements relating to vehicular emissions, diesel fuel specifications had their origin in California and are most stringent there. There is also a likelihood that any fuel specifications begun in California may eventually spread to the remainder of the United States and might also come to Europe. It is therefore logical that any discussion regarding trends in diesel fuel specifications should start with those in California.

Particulate emissions, especially the size fraction smaller than 10 μm, are also drawing attention as a serious health concern. The current major source is from (CI = diesel) engines, while the modern spark ignition engine system has no problem meeting regulatory requirements.

2.4.1 California Specifications

The most exhaustive legislation regarding diesel fuel specifications is in place in California,[31] enacted by the California Air Resources Board (CARB). A summary of this legislation is shown in Table 4 below. The legislation affects both the sulfur content and aromatics level of diesel fuels.

The sulfur requirement for diesel fuels is 500 ppm maximum. The sulfur restriction is based on the one to one correlation of fuel sulfur content with sulfate particulate emissions. All refiners in the south coast air quality district (Los Angeles basin) have faced this sulfur requirement since 1985.

The aromatics requirement is 10% maximum for large refiners and 20% maximum for small refiners.

Table 4 CARB diesel fuel reformulation requirements effective October 1993

	1990 Average	Large refiner	Small refiner
Sulfur, wt% max	0.50[a]	0.05	0.05[b]
Aromatics, vol. % max	no spec	10	20
CN,[c] min	40	40	40

[a] The 0.50 value is the ASTM specification for #2 diesel. Most production is typically about 0.30 maximum. Since 1985, sulfur specification in the Los Angeles Basin has been 0.05 weight% max.

[b] Small refiners had until October 1994 to implement sulfur specifications.

[c] CN (cetane number) is not a legislated specification but is based on ASTM standards for #2 diesel fuel.

There has been much controversy over the aromatics requirement. The principle behind the aromatics restriction is to reduce NO_x and particulate emissions. The Coordinating Research Council (USA) (CRC), a joint industry group, has conducted a number of studies attempting to relate diesel emissions to fuel properties. Their summary report concluded that there was a significant correlation of emissions with fuel aromatics content. The correlation held for most types of emissions, including hydrocarbons, NO_x, CO, and particulates. This test program primarily utilized neat (no additive) fuels and no correlations were developed to relate emissions to CN. Subsequent studies that have included fuels with and without cetane improver have shown a much better correlation of emissions with fuel CN. Despite these arguments, the aromatics specification has remained in place. The earlier reported small degree of positive correlation of emissions versus aromatics has been sufficient to maintain the requirements.

Compounding the situation is the method specified by CARB for measuring aromatics. CARB has specified the fluorescent indicator absorption (FIA) test, ASTM D-1319, as the method for determining aromatics. This test is not applicable for many higher boiling diesel fuels. As yet there has been no change in this test specification. There have been a number of papers that have discussed various methods for determining diesel fuel aromatics.[32,33]

There is one option for a refiner to bypass this maximum aromatics requirement. CARB has defined the properties of a reference diesel fuel. This fuel must be derived from straight-run diesel stocks from crude oils that are typically run in California refineries. The reference fuel must then meet the requirements shown in Table 5. The key parts to these requirements are the 10% aromatics requirement and the relatively high natural CN of 48. In order to qualify as an alternative fuel, a specific side-by-side emissions test is performed with the reference fuel and the alternative fuel. If the alternative fuel emissions meet or better those of the reference fuel, the alternative fuel can be qualified. Certain properties of the alternative fuel then become new specifications for that refiner. These properties are also noted on Table 5. While many California refiners are pursuing this option, it still may involve a tightening of specifications. Some estimates indicate that a qualifying alternative fuel may require an aromatics content of 15–25% and a CN in the range of 48 to 52.

One major southern California refiner claims to have a new clean diesel fuel that will surpass Phase II limits, provide higher cetane, reduce NO_x emissions by 5% and reduce particulates by 15%. If true, this may become the future specifications for RFG.

2.4.2 United States Environmental Protection Agency Diesel Fuel Specifications

These specifications include the lower sulfur requirement of 0.05 weight% maximum. Rather than an aromatics specification, the US EPA has proposed a minimum cetane index of 40. This type of specification is thought to be more reasonable with respect to emissions control. Current ASTM specifications for Number 2 Diesel call for a minimum CN of 40. Thus, unless a refiner is using a lot

Table 5 CARB reference fuel specifications

	Reference fuel		Alternative fuel critical property
	Large refiner	Small refiner	
Sulfur content, ppm max	500	500	Yes, 500 ppm max
FIA aromatics, % max	10	20	Yes
Polycyclic aromatics by mass spec, wt% max	1.4	4.0	Yes
Nitrogen, ppm max	10	90	Yes
Natural CN, min	48	47	
Specific gravity (deg API)	0.830–0.860 (33–39)	0.830–0.860 (33–39)	
Viscosity at 40 °C, CST	2.0–4.1	2.0–4.1	
Distillation, D-86, C			
IBP	171–215	171–215	
10% Recovery	204–254	204–254	
50% Recovery	243–293	243–293	
90% Recovery	288–321	288–321	
EP	304–349	304–349	
CN			Yes
Additive package			Report

API, American Petroleum Institute; IBP, initial boiling point; EP, end point; CST, centistokes.

of cetane improver, the cetane index specification will not be a significant change. It could have an impact on refiners who use large quantities of fluid catalytic cracker (FCC) cycle oil in their diesel pool. Refiners with large runs of highly naphthenic crudes could also be affected. This is typically much of California's refining feedstock base.

Depending on the implementation of the California specifications, there will probably be some carry over to the remainder of the country. In the future, this might include some type of aromatics specification or a further increase in the minimum cetane index.

2.4.3 European and Worldwide Diesel Specifications

In much of western Europe, the sulfur specification is 0.30 weight% maximum. Sweden, Germany, and the Netherlands have reduced the sulfur specification to 0.20 weight%. The EU may move toward reducing the sulfur limit, although certain countries may again take the lead. Values that have been discussed range from 0.15 weight% down to the California, US, level of 0.05 weight% (500 ppm wt).

At present, there are no specific aromatics or cetane requirements for EU diesel. The large percentage of the diesel pool is from straight run stocks, so cetane tends to be inherently high. Sweden has tightened aromatics levels in what is called ''city diesel'' at 350 ppm. As more and more heavy oil conversion units are constructed, this picture may change and more universal limitations on aromatics and cetane may be expected.

The rest of the world still does not worry about diesel, with the exception of a few notable Asian countries, and Brazil, which recently has brought nearly all fuel sources under the 300 ppm limit.[34]

2.5 Future Regulatory Trends

The dramatic rush by refiners worldwide to install regulatory controls in the mid 1990s was mercifully tempered by a minimum of new edicts about what specifications had to be met. The cost of meeting even US EPA Phase I requirements was too much for many US refiners, particularly those with a capacity under 50 000 BPD (barrels per day). Dozens of them were merged, closed or downgraded to oil storage terminals, simply from the excessive cost of making regulation gasoline in a competitive market. Refiners in California had a four-year lead on the rest of the country as the CARB standards basically matched US EPA Phase II. As the year 2000 approached, another round (though smaller than the first) of capital spending began in those remaining refineries that had not yet complied with the Phase II specifications.

Based on the fact that the last major legislation was written in 1991 for this regulatory activity, it would be reasonable to assume that within the next three to four years there will be further tightening of some of the standards in Tables 2, 3, 4, and 5. There is currently a rather heated debate on sulfur levels in gasoline and total aromatic content in all fuels. One can only expect the limits on both of these to go down. It is also likely that total reductions of all regulated components will be mandated based on some 1998–2000 timeframe emissions average criteria. There will also be a push to reduce nitrogen and more of the VOCs. One can examine the proposed Tier2 regulations being proposed[35] that now extend to sports utility venicles (SUVs), minivans, and light duty trucks.

The one exception to this trend to tighter specifications may be in the amount and types of oxygenate used. There has recently been a flurry of data and instances detailing contamination of groundwater supplies by MTBE, one of the more prevalent gasoline oxygenates. There are also some credible court challenges to the effectiveness of oxygenated compounds in reducing smog-forming components in air. It may well be a better idea to reduce such emissions by lowering sulfur content drastically, although this idea is also currently based more on circumstantial evidence rather than theoretical considerations.

3 CHEMICAL COMPONENTS OF PETROLEUM-DERIVED FUELS

Gasoline is a volatile mixture of liquid hydrocarbons suitable for use in spark-ignition internal combustion (IC) engines. Gasoline is formulated from the lighter ends of refinery runs and generally contains petroleum molecules having between four and 12 carbon atoms. This material is sometimes called light naphtha in the industry. Gasoline also typically contains small amounts of additives to prevent icing and corrosion and to improve engine performance. Refiners add high-octane oxygenates such as ethanol, methanol or MTBE or aromatics such as benzene, toluene or xylene to enhance octane. Kerosine and diesel stocks have slightly heavier components than gasoline averaging about 2–4 carbon numbers heavier than gasoline respectively on a size distribution scale. This material is often referred to as heavy naphtha.

3.1 Crude Oil Chemical Mixture Components[7,20,21,36–39]

It is generally accepted that the origin of crude oil is from plant life up to 3 billion years ago, but predominantly from 100 to 600 million years ago. The molecular structure of the hydrocarbons and other compounds present in fossil fuels can be linked to the leaf waxes and other plant molecules of marine and terrestrial plants believed to exist during that era. There are various biogenic

marker chemicals such as isoprenoids from terpenes, porphyrins and aromatics from natural pigments, pristane and phytane from the hydrolysis of chlorophyll, and normal alkanes from waxes, whose size and shape cannot be explained by any known geological processes. The presence of optical activity and the carbon isotopic ratios also indicate a biological origin.

Hydrocarbons are any molecules that just contain hydrogen and carbon, both of which are fuel molecules that can be oxidized to form water (H_2O) or carbon dioxide (CO_2). If the combustion is not complete, carbon monoxide (CO) may also be formed.

The way that hydrogen and carbons bond determines which hydrocarbon family they belong to. If they only hold one common bond they are called ''saturated hydrocarbons'' because they cannot absorb additional hydrogen. If the carbons hold two common bonds they are called ''unsaturated hydrocarbons'' because they can be converted into ''saturated hydrocarbons'' by the addition of hydrogen to the double bond. Hydrogens are omitted from the following protocol for convenience, but if one remembers C has between one and four bonds maximum, H has one bond only, and O has two bonds only, the full structures of most hydrocarbons can be quickly drawn.

Gasoline contains over 500 hydrocarbons that may have between 3 and 12 carbon atoms, and gasoline used to have a boiling range from 30–220 °C at atmospheric pressure. The boiling range is narrowing as the initial boiling point (IBP) is increasing, and the FBP is decreasing–both changes are for environmental reasons. Detailed descriptions of structures can be found in any chemical or petroleum text that discusses gasolines.

Saturated hydrocarbons (paraffins, alkanes) tend to burn in air with a clean flame and form the stable, majority compositional segment of gasolines. They are subdivided into normal, iso and cyclic alkanes

- Normal alkanes have a continuous chain of carbon atoms (C_nH_{2n+2}), for example normal heptane **(1)** is

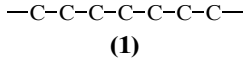

(1)

or C_7H_{16}.

- Iso alkanes have a branched chain of carbon atoms (C_nH_{2n+2}), for example isooctane **(2)** (or 2,2,4-trimethylpentane) is

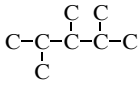

(2)

or C_8H_{18}.

- Cyclic alkanes (naphthenes) have a circle of carbon atoms (C_nH_{2n}), for example cyclohexane **(3)** is

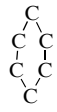

(3)

or C_6H_{12}.

Unsaturated hydrocarbons contain one or more double bonds between carbon atoms.

- Alkenes (olefins) have carbon–carbon double bonds (C_nH_{2n}), for example 2-methyl-2-butene is **(4)** or C_5H_{10}.

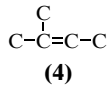

(4)

These components are unstable and reactive with several chemicals, including themselves (polymerization) and are the principal ''gum'' formers in fuels. They also tend to burn in air with a smoky flame and typically have a foul smell. They are very rarely found to any extent in natural ''straight run'' cuts or in hydrotreated feedstocks but are a significant component of thermally cracked stocks such as from the FCC or Coker refinery units.

- Alkynes (acetylenes) have carbon–carbon triple bonds (C_nH_{2n-2}), for example acetylene **(5)**

(5)

or C_2H_2.

These are even more unstable, but are only naturally present in trace amounts, and only to a small extent in some poorly refined gasolines.

- Arenes (aromatic compounds) have one or more ''benzene-type'' rings with highly unsaturated but highly stable intercarbon bonding. For example benzene **(6)**

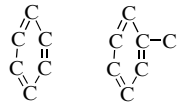

(6) **(7)**

or C_6H_6 and toluene **(7)** or C_7H_8.

These are found to a small but significant extent in most petroleum stocks and are formed in most refinery cracking and reforming processes as well. Typical aromatics levels in gasolines used to be up to 40%, but are gradually being reduced to <20%. From a performance standpoint, however, the benzene, toluene and the xylenes give outstanding gasoline performance results and are difficult to replace in the gasoline supply chain. On the minus side, aromatics attack elastomers and generate smoke.

- Polynuclear aromatics (PNAs or PAHs) contain benzene rings joined together, the simplest of which is naphthalene **(8)**

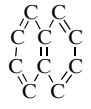

(8)

or $C_{10}H_8$.

These are high boiling, and are only present in small amounts in gasoline. The multi-ringed PNAs are highly toxic, and are not present in gasoline boiling range material but will be found in diesel stocks.

Oxygenates are mostly hydrocarbons which happen to contain oxygen, for example, ethanol C–C–O–H or C_2H_5OH and MTBE (tertiary butyl methyl ether) **(9)** or $C_4H_9OCH_3$. Their structure provides a reasonable anti-knock value, thus they are good substitutes for aromatics, and they may also reduce the smog-forming tendencies of the exhaust gases. They can be produced from fossil fuels such as methanol (MeOH), MTBE, tertiary amyl methyl ether (TAME), or from biomass such as ethanol (EtOH), ethyl tertiary butyl ether (ETBE)). Most oxygenates used in gasolines are either alcohols (C_x–O–H) or ethers (C_x–O–C_y), and contain 1–6 carbon atoms. Oxygenates are added to gasolines to reduce the reactivity of emissions, but are only effective if the hydrocarbon fractions are carefully modified to utilize the octane and volatility properties of the oxygenates. If the hydrocarbon fraction is not correctly modified, oxygenates can increase undesirable smog-forming and toxic emissions.

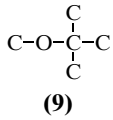

(9)

3.2 Typical Gasoline Mixture Components[39–41]

During the 1990s, various compositional changes to gasolines for environmental and health reasons have resulted in fuels that do not follow historical rules, and the regulations mapped out for 2000–2010 also ensure that the composition will remain in a state of flux. The reformulated gasoline (RFG) specifications, especially the US EPA Jan 1998 Complex model, will introduce major reductions in the distillation range, as is now the case in California (CARB rules), as well as the various limits on composition and emissions.

It is not feasible to list all 500+ hydrocarbons in current gasolines, but those presented in Table 6 are representative of the various classes typically present in a straight run, a conventional (pre-Phase II) and a fully RFG. A traditional fuels cut composition from a normal conversion refinery would include the components given in Table 5. Newer oxygenated fuels (bottom of Table 6) reduce the aromatics and olefins, narrow the boiling range, and add oxygenates up to about 12–15% to provide the octane.

Some of the C_{10+} components will be found in most gasolines, perhaps up to C_{14}, but volatility specifications require that very little of these be residual in the gasoline distillates. Gasoline components will more typically have C_6 to C_{10} components, with an average of around 8.

There are some other properties of oxygenates that have to be considered when they are going to be used as fuels, particularly their ability to form very volatile azeotropes that cause the fuel's vapor pressure to increase, the chemical nature of the emissions, and their tendency to separate into a separate water/oxygenate phase when water is present. The RFGs address these problems more successfully than the original oxygenated gasolines otherwise known as gasohols.

Some other fuel additives are typically formulated into the gasoline by the refiner. A typical gasoline may contain:[8,11,23,42,43]

- Oil-soluble dye, initially added to leaded gasoline at about 10 ppm to prevent misuse as an industrial solvent,
- Antioxidants, typically phenylenediamines or hindered phenols, are added to prevent oxidation of unsaturated hydrocarbons,
- Metal deactivators, typically about 10 ppm of chelating agent such as *N*,*N*′-disalicylidene-1,2-propanediamine is added to inhibit copper, which can rapidly catalyze oxidation of unsaturated hydrocarbons,
- Corrosion inhibitors, about 5 ppm of oil-soluble surfactants are added to prevent corrosion caused either by water condensing from cooling water-saturated gasoline, or from condensation from air onto the walls of almost empty gasoline tanks that drop below the dew point. If gasoline travels along a pipeline, it is possible that the pipeline owner will add additional corrosion inhibitor to the fuel,

Table 6 Fuel property characteristics of common refinery chemical components

	RON	MON	bp	Specific gravity
n-Paraffins (15% typical pre-US EPA)				
n-butane	113	114	−0.5	gas
n-pentane	62	66	35	0.626
n-hexane	19	22	69	0.659
n-heptane (0:0 by definition)	0	0	98	0.684
n-octane	−18	−16	126	0.703
n-decane[a]	−41	−38	174	0.730
n-dodecane[a]	−88	−90	216	0.750
n-tetradecane[a]	−90	−99	253	0.763
iso-Paraffins (30% typical pre-US EPA)				
2-methylpropane	122	120	−12	gas
2-methylbutane	100	104	28	0.620
2-methylpentane	82	78	62	0.653
3-methylpentane	86	80	64	0.664
2-methylhexane	40	42	90	0.679
3-methylhexane	56	57	91	0.687
2,2-dimethylpentane	89	93	79	0.674
2,2,3-trimethylbutane	112	112	81	0.690
2,2,4-trimethylpentane (100:100 by definition)	100	100	98	0.692
Cycloparaffins (12% typical pre-US EPA)				
cyclopentane	141	141	50	0.751
methylcyclopentane	107	99	72	0.749
cyclohexane	110	97	81	0.779
methylcyclohexane	104	84	101	0.770
Aromatics (35% typical pre-US EPA)				
benzene	98	91	80	0.874
toluene	124	112	111	0.867
ethylbenzene	124	107	136	0.867
meta-xylene	162	124	138	0.868
para-xylene	155	126	138	0.866
ortho-xylene	126	102	144	0.870
3-ethyltoluene	162	138	158	0.865
1,3,5-trimethylbenzene	170	136	163	0.864
1,2,4-trimethylbenzene	148	124	168	0.889
Olefins (8% typical pre-US EPA)				
2-pentene	154	138	37	0.649
2-methylbutene-2	176	140	36	0.662
2-methylpentene-2	159	148	67	0.690
cyclopentene	171	126	44	0.774
1-methylcyclopentene[b]	184	146	75	0.780
1,3 cyclopentadiene[b]	218	149	42	0.805
dicyclopentadiene[b]	229	167	170	1.071
Oxygenates				
methanol	133	105	65	0.796
ethanol	129	102	78	0.794
isopropyl alcohol	118	98	82	0.790
MTBE	116	103	55	0.745
ETBE	118	102	72	0.745
TAME	111	98	86	0.776

MON, motor octane number.
[a] Compounds are too heavy for gasoline, but would be fine for kerosine and diesel.
[b] These olefins are not present in significant amounts in gasoline, but have some of the highest blending octanes and cetane values.

- Anti-icing additives, used mainly with carbureted cars, and usually either a surfactant, alcohol or glycol,
- Anti-wear additives, which are used to control wear in the upper cylinder and piston ring area that the gasoline contacts, and are usually very light hydrocarbon oils. Phosphorus additives can also be used for engines without exhaust catalyst systems.
- Deposit-modifying additives, usually surfactants. These include:

1. Carburetor deposits – additives to prevent these were required when crankcase blow-by (PVC, polyvinyl chloride) and exhaust gas recirculation (EGR) controls were introduced. Some fuel components reacted with these gas streams to form deposits on the throat and throttle plate of carburetors.
2. Fuel injector tips operate at about 100 °C, and deposits form in the annulus during hot soak, mainly from the oxidation and polymerization of the larger unsaturated hydrocarbons. The additives that prevent and unclog these tips are usually polybutene succinimides or polyether amines.
3. Intake valve deposits (IVDs) caused major problems in the mid-1980s when some engines had reduced drivability when fully warmed, even though the amount of deposit was below previously acceptable limits. It is believed that the new fuels and engine designs were producing a more absorbent deposit that attracted some of the passing fuel vapor, causing lean hesitation. Intake valves operate about 300 °C, and, if the valve is kept wet, deposits tend not to form, thus intermittent injectors tend to promote deposits. Oil leaking through the valve guides can be either harmful or beneficial, depending on the type and quantity. Gasoline factors implicated in these deposits include unsaturates and alcohols. Additives to prevent these deposits contain a detergent and/or dispersant in a higher molecular weight solvent or light oil whose low volatility keeps the valve surface wetted.
4. Combustion chamber deposits were targeted in the 1990s, as they are responsible for significant increases in emissions. Recent detergent-dispersant additives have the ability to function in both the liquid and vapor phases to remove existing carbon and prevent deposit formation.
5. Octane enhancers are usually formulated blends of alkyl lead or manganese compounds in a solvent such as toluene, and added at the 100–1000-ppm levels. They have been replaced

by hydrocarbons with higher octanes such as aromatics and olefins. These hydrocarbons are now being replaced by a mixture of saturated hydrocarbons and oxygenates.[23,44]

3.3 Components in Heavier Fuels

Heavier petroleum fuels have similar components, only larger on average. Most of the components described above will be present to a smaller degree with significant amounts of higher boiling components, at least for those fuels derived from crude distillation. Many of the compounds with multiple aromatic or naphthenic rings have extended side chains as well. Most of the sulfur, nitrogen, and metals are concentrated in the diesel range components and are more than likely to be embedded inside an aromatic ring (RA). Thiobenzene and dithiobenzene are typical sulfur compounds. Pyridines and pyrroles are common nitrogen contaminant species. Metallic-centered porphyrins are the primary metals species for iron, nickel, and vanadium content.

Most of the same C_6-C_{10} components will be found in most heavier distillates. Kerosine will have significant amounts of up to C_{14} (average about 10) and some diesel components may have a molecular carbon number as high as 18–20, although the average molecular number here is about 11.

4 PETROLEUM-DERIVED FUELS PERFORMANCE CRITERIA

4.1 Spark-ignition Engines (Gasoline)

The critical factors in desirable performance[17,23,41,45–47] of gasoline in a spark-ignition engine[37,38,43,48,49] are volatility, antiknock quality, storage stability, component compatibility, and intake system deposit control. Gasolines are usually determined by government regulation, where properties and test methods are clearly defined. In the USA, several government and state bodies can specify gasoline properties. The US gasoline specifications and test methods are listed in several readily available publications, including the Society of Automotive Engineers (SAE) and the ASTM. These specifications, in fact, change from year to year so they are not listed with the year designation included. To do so would condemn this work to obsolescence very quickly. The latest editions, some having only minor changes in them from previous editions, can be obtained directly through ASTM or found in the nearest technical library.

ASTM D-4814 (*Specification for Automotive Spark-ignition Engine Fuel*) is the major standard for US automobile manufacturers to follow. Some of the considerations addressed therein are discussed in more detail below.

4.1.1 Volatility Considerations

Spark-ignition engines need a volatile[8,50–52] fuel for ease in starting, fast warm-up, and good drivability. In practical terms, this means a fuel that boils in the range 30–215 °C (85–420 °F). Too light a fuel leads to poor fuel economy (volatile volume loss), carburetor icing, and vapor lock in the fuel pump. Too heavy a fuel causes cold weather starting problems, poor drivability, engine deposits, and crankcase oil dilution.

Two primary standards for measuring gasoline volatility are distillation (ASTM D-86) and Reid vapor pressure (RVP) (ASTM D-4953). Distillation measures the boiling range of the hydrocarbons that make up the finished gasoline. RVP measures the pressure that the fuel exerts under conditions approximating those in an automobile carburetor. From these, a third standard vapor/liquid (V/L) ratio, can be calculated using the procedure in ASTM D-2533. The temperature where $V/L = 20$ indicates the tolerance of modern cars to vapor lock. The higher the temperature, the less chance of hot weather stalling. At low temperatures, other factors such as drivability must be considered. These factors are also controlled by ASTM D-86.

Gasoline specifications change seasonally in most areas of the world that have distinguishable seasonality. The reason for these variations is quite simple. In the cylinder only gaseous hydrocarbons burn, so if the air is cold, the fuel has to be very volatile. In summer, a volatile fuel can boil too readily and cause vapor lock, as well as producing high levels of evaporative emissions. The solution for automobile manufacturers was to adjust the volatility of the fuel according to altitude and ambient temperature. This volatility change has been automatically performed for decades by the oil companies without informing the public of the changes. It is one reason why storage of gasoline through the seasons is not a good idea.

Even as gasoline volatility is being reduced, modern engines, with their fuel injection and computerized management systems, can automatically compensate for some of the changes in ambient conditions such as altitude and air temperature, resulting in acceptable drivability using less volatile fuel.

ASTM D-4814 defines six different classes according to location and/or season. As gasoline is distilled, the temperatures at which various fractions are evaporated are calculated. Specifications define the temperatures at which various percentages of the fuel must evaporate. Distillation limits include maximum temperatures when 10% is evaporated (50–70 °C), 50% is evaporated (110–121 °C), 90% is evaporated (185–190 °C), and the

final boiling point (FBP) (225 °C). There is a minimum temperature for 50% evaporated (77 °C), and a maximum amount of residue (2%) after distillation. Vapor pressure limits for each class (54, 62, 69, 79, 93, 103 kPa) are also specified. The US EPA has issued a waiver that does not require gasoline/ethanol blends to meet the required specifications in certain areas of the United States. ASTM D-4814 also defines five classes for vapor lock protection, according to location and/or season. The limit is a maximum V/L ratio of 20 at test temperatures of 41, 47, 51, 56 and 60 °C.

Volatility also affects combustion engine emissions. The higher boiling fractions of the gasoline have significant effects on the emission levels of undesirable hydrocarbons and aldehydes, and a reduction of 40 °C in the FBP will reduce the levels of benzene, butadiene, formaldehyde and acetaldehyde by 25%, and will reduce hydrocarbon emissions by 20%.[11]

4.1.2 Antiknock Quality

Under severe service conditions, spark-ignition engines are knock limited. High temperatures, high compression ratios, lean mixtures, and advanced spark settings lead to knock, which is the explosive combustion of the last part of the fuel/air mixture ''end gas'' drawn into the cylinder. A gasoline's chemical composition determines its ability to resist knock.

The antiknock value[18,21,49] of a gasoline is measured in two standard laboratory engine tests. Both use the same type of single-cylinder variable-compression ratio engine attached to a dynamometer with the operating conditions changed only as needed. ASTM D-2699 describes the standard for determining RON and ASTM D-2700 pertaining to motor octane number (MON).

In both standard procedures, the compression ratio is adjusted to produce a measured level of knock with the test fuel, then the knock intensity with a primary reference fuel is blended to give the same, then reblended for slightly better and finally for slightly worse antiknock performance. Because the reference fuels are known blends of isooctane (octane number of 100) and *n*-heptane (octane number of 0), the octane content of the fuel that exactly matches the knock intensity of the test fuel is its octane number. Of the two procedures, the MON test is more severe, correlating with high-speed high-temperature part-throttle service in an actual vehicle. The MON for most gasolines is lower than the less severe RON, which primarily correlates to a gasoline's ability to resist run-on, after-run, or dieseling. The difference between the two ratings ($R - M$) is called the sensitivity.

Neither test exactly correlates with actual service in a car, where driving conditions vary continuously. Actual ''road'' octane number, otherwise known as antiknock index (AKI) is a better predictor of quality, and various procedures are available to measure this value, either on the road or the chassis dynamometer. Since no two cars rate gasolines exactly alike, RON cannot be used as a control test. As a compromise, the AKI has been developed. This is the arithmetic average of RON and MON, and is sometimes called simply the $(R + M)/2$ value. Where the law requires octane numbers to be posted on gasoline pumps, AKI is posted.

Gasoline specifications are written to reflect the octane number response (ONR) of the typical car population, but they also vary with location and season. Higher ambient temperatures and engine cylinder age increase ONR, while higher absolute humidity and elevation above sea level decrease it.

The most significant determinant of octane, of course, is the chemical structure of the hydrocarbons and their response to the addition of octane enhancing additives. Several other important properties of the fuel itself affect knock. Among them are:

- Front end volatility – Paraffins are the major component in gasoline, and the octane number decreases with increasing chain length or ring size, but increases with chain branching. Overall, the effect is a significant reduction in octane if front end volatility is lost, as can happen with improper or long-term storage. Fuel economy on short trips can be improved by using a more volatile fuel, at the risk of carburetor icing and increased evaporative emissions. The ASTM standards dealing with front-end volatility are D-86, D-2887, D-323, D-4953, D-5191, and D-2533.
- Final boiling point – Decreases in the FBP increase fuel octane. Aviation gasolines have much lower FBPs than automotive gasolines. Note that FBPs are being reduced because the higher boiling fractions are responsible for disproportionate quantities of pollutants and toxins. The ASTM standards dealing with back-end volatility are D-86, D-1160, and D-2887
- Preignition tendency – Both knock and preignition can induce each other.

ASTM 4814 does not set limits for AKI, but changes in engine requirements according to season and location are discussed. Fuels with an AKI of 87, 89, 91 (unleaded), and 88 (leaded) are listed as typical for the USA. Analogous considerations for cetane number (CN) in diesel stocks are governed by ASTM D-613 and ASTM D-976.

4.1.3 Storage Stability

Gasoline deterioration in storage is caused mainly to oxidative processes. Oxidation leads to the formation

of gum, a varnish-like material that can build up in combustion chambers and intake systems, interfering with efficient engine running. In extreme cases, it can even cause ring sticking and engine seizure. To guard against this problem, refiners include antioxidants and metal deactivators, which reduce the catalytic effect of certain metals in promoting oxidation.

Specifications limit the amount of gum in fresh gasoline to 5 mg per 100 mL. These results do not correlate well with actual engine deposits caused by fuel vaporization and probably should be considered as a minimum. Laboratory oxidation tests control the extent to which oxygen will attack gasoline. Compliance with these limits usually ensures that gasolines will remain usable for up to 12 months. Any gasoline known to have been in storage longer than 6 months, however, should be retested before use. Because the oxidation rate doubles for every 10 °C temperature increase, storage conditions are important.

Storage time also increases the likelihood of a water–oxygenate phase separating from the hydrocarbon, so care should be taken with the ''shelf life'' of oxygen-containing fuels. Water tolerance via ASTM D-4814 sets the highest temperature that causes phase separation of oxygenated fuels. The limits vary according to location and month. For Alaska, north of latitude 62, it changes from −41 °C in Dec/Jan to 9 °C in July, but remains 10 °C all year in Hawaii.

Gums are usually the result of copper-catalyzed reactions of the unsaturated hydrocarbons, so antioxidants and metal deactivators are added. Existent gum is used to measure the gum in the fuel at the time tested, whereas the oxidation stability measures the time it takes for the gasoline to break down at 100 °C with 100 psi (690 kPa) of oxygen. A 240-min test period has been found to be sufficient for most storage and distribution systems.

Minimum oxidation stability tests ensure the fuel remains chemically stable, and does not form additional gums during periods in distribution systems, which can be up to 3–6 months. The sample is heated with oxygen inside a pressure vessel, and the delay until significant oxygen uptake is measured.

The ASTM standards dealing with storage stability in gasolines are D-381 and D-525.

4.1.4 Component Materials Compatibility[45,49]

Minor amounts of nonhydrocarbon material found in, or added to, gasolines can have adverse effects on engine life and performance. Phosphorus and lead deactivate oxidation catalysts, and oxygenated materials such as alcohols and ethers can cause seal and fuel-line swelling as well as promoting engine rust.

Sulfur in the fuel creates corrosion and when combusted will form corrosive gases that attack the engine, exhaust and environment. Sulfur also can poison exhaust catalysts. The copper strip corrosion test and the sulfur specification are used to ensure fuel quality. The copper strip test measures active sulfur, whereas the sulfur content reports the total sulfur present.

ASTM D-4814 specifies lead content at the following levels for:

- Leaded brands (where allowed) = 1.1 g Pb L^{-1} maximum,
- Unleaded brands = 0.013 g Pb L^{-1} maximum.

Most local jurisdictions allow for almost no lead, and trace analysis procedures may be needed to determine trace levels.

ASTM D-4814 only mentions copper strip corrosion as the ability to tarnish clean copper, indicating the presence of any corrosive sulfur compounds. Interpretation is left up to the refiner.

Maximum sulfur content via D4814 is specified at the following levels for:

- Leaded brands (where allowed) = 0.15% mass (1500 ppm) maximum
- Unleaded brands = 0.10% mass (1000 ppm) maximum

Sulfur adversely affects exhaust catalysts and is typically emitted as polluting sulfur oxides. Typical US gasoline levels are 0.03% (30 ppm) by mass.

The ASTM standard dealing with materials compatibility for gasolines is D-130.

4.1.5 Intake System Deposit Control

The intake system includes either a carburetor, throttle body injector or port fuel injector, and intake valve stem, tulip, head, and intake manifold runners. Many gasolines contain unstable components that can lead to deposit formation in these high-temperature areas of the induction system. Gasolines for use as fuels in spark-ignition engines must be able to reduce or eliminate deposit formation in critical fuel-delivery system components. Since January 1 1995, the US EPA has mandated the use of detergent additives in gasoline in order to help reduce emissions and improve air quality.

During the 1980s significant problems with deposits accumulating on intake valve surfaces occurred as new fuel injections systems were introduced.[21] These intake valve deposits (IVDs) were different from the injector deposits, in part because the valve can reach 300 °C. Engine design changes that prevent deposits usually ensure that the valve is flushed with liquid gasoline, and provide adequate valve rotation. Gasoline factors that cause deposits are the presence of alcohols or olefins.

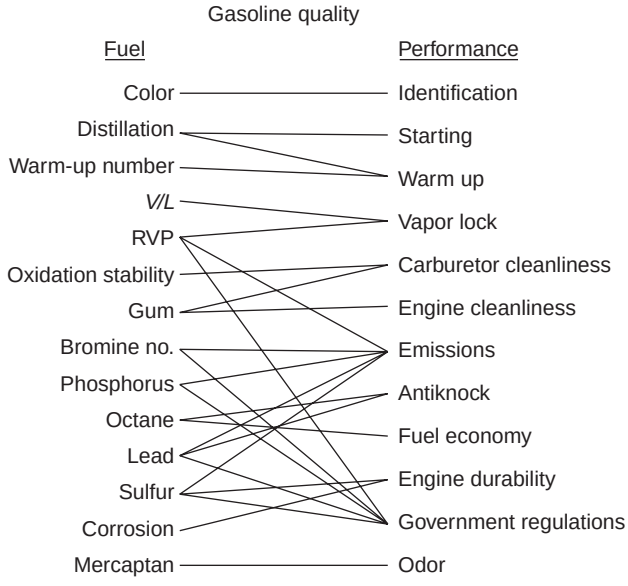

Figure 1 Fuel performance and property relationship for gasoline.

Gasoline manufacturers now routinely use additives that prevent IVD and also maintain the cleanliness of injectors. These usually include a surfactant and light oil to maintain the wetting of important surfaces.

Figure 1 illustrates the fuel property/performance relationships of a modern gasoline.

4.1.6 Brands Differences/Additive Packages

The above specifications and standards are intended to ensure minimal quality standards are maintained, however as well as the fuel hydrocarbons, the manufacturers add their own special ingredients to provide additional benefits. A quality gasoline additive package would include:

- octane-enhancing additives (improve octane ratings)
- antioxidants (inhibit gum formation, improve stability)
- metal deactivators (inhibit gum formation, improve stability)
- deposit modifiers (reduce deposits, spark-plug fouling and preignition)
- surfactants (prevent icing, improve vaporization, inhibit deposits, reduce NO_x emissions)
- freezing point depressants (prevent icing)
- corrosion inhibitors (prevent gasoline corroding storage tanks)
- dyes (product color for safety or regulatory purposes).

4.2 Aviation Gasoline

A special grade of gasoline is aviation gasoline, called avgas for short. It should not be confused with jet fuel which is essentially a kerosine-derived turbine fuel. Avgas is for smaller spark ignition aircraft engines. Most specifications for avgas cater to the safety factors at the expense of the economics. Distillations and volatility are more tightly controlled, as are contaminant levels. The octane requirements are higher to account for both the higher engine cylinder compression ratios and the devastation that could be caused by premature valve and lifter wear if cheap gasolines were routinely available to unsophisticated aircraft owners. Most other specifications follow those of automobile gasoline, including recent US EPA/CARB regulatory requirements. Significant exceptions are mentioned below.

4.2.1 Aviation Gasoline Antiknock Numbers

Aviation gasolines were all highly leaded and graded using two numbers, with common grades being 80/87, 100/130, and 115/145. The first number is the aviation rating (lean mixture rating), and the second number is the supercharge rating (rich mixture rating). In the 1970s a 100 LL grade (low lead = 0.53-mL TEL L^{-1} instead of 1.06-mL TEL L^{-1}, TEL = tetra ethyl lead) was introduced to replace the 80/87 and 100/130. Soon after the introduction, there was a spate of plug fouling, and high cylinder head temperatures resulting in cracked cylinder heads. The old 80/87 grade was reintroduced on a limited scale.

The aviation rating is determined using the automotive motor octane test procedure, and then corrected to an aviation number (AN) using a table in the method – it is usually only 1 – 2 octane units different from the motor value up to 100, but varies significantly above that, e.g. 110 MON = 128 AN.

The second avgas number is the rich mixture method performance number (PN – they are not commonly called octane numbers when they are above 100), and is determined on a supercharged version of the constant fuel ratio (CFR) engine, which has a fixed compression ratio. The method determines the dependence of the highest permissible power (in terms of indicated mean effective pressure) on mixture strength and boost for a specific light knocking setting. The PN indicates the maximum knock-free power obtainable from a fuel compared to isooctane = 100. Thus, a PN = 150 indicates that an engine designed to utilize the fuel can obtain 150% of the knock-limited power of isooctane at the same mixture ratio. This is an arbitrary scale based on isooctane + varying amounts of TEL, derived from a survey of engines performed decades ago. Aviation gasoline PNs are rated using variations of mixture strength to obtain the maximum knock-limited power in a supercharged engine. This can be extended to provide mixture response curves which define the maximum boost (rich, about 11 : 1 stoichiometry) and minimum boost (weak, about 16 : 1 stoichiometry) before knock.[45]

The 115/145 grade has been phased out in North America (though not everywhere else in the world), but the 100 LL is still in use and has more octane than any automotive gasoline. These specifications change much more slowly than automobile fuel specifications because the market is much smaller and safety for use in older aircraft is still considered an important criteria.

The basic ASTM standard dealing with avgas performance is D-910.

4.3 Compression–Ignition Engine (Diesel) Fuels

Compression–ignition (CI) engines run on mid-distillate fuels with boiling ranges higher than that of gasoline. The best diesel fuels are ''straight-run'' stocks, derived from simple distillation of crude oil, but many commercial fuels contain a proportion of catalytically cracked material.

The important properties of diesel fuels are volatility, heating value, ignition quality, CN, viscosity, low-temperature flow, storage stability, component compatibility, and sulfur content. There is no overall defining specification for diesel fuel due to the wide variety of engine types and requirements that the fuel must meet. There are standards, however, for individual diesel fuel qualities, both economic and regulatory.

4.3.1 Volatility

The volatility of a diesel fuel has little influence on its engine performance, except as it affects ignition tendencies. The distillation range of a diesel fuel does not allow much flexibility in this regard because of the interrelationship and interdependence with other specification factors. The controlling factor is the ASTM D-86 distillation 90% evaporated temperature.

Because diesel fuels are classified as nonflammable for freight purposes, minimum flashpoint restrictions are imposed. The ASTM standards dealing with diesel flashpoint are D-93 and D-95.

4.3.2 Heating Value

Fleet operators, rail and shipping companies are concerned about fuel economy. They aim to use the fuel with the greatest heating value, provided that other specification requirements are met. The factors that influence heating value are density and mid-boiling point (bp).

4.3.3 Ignition Quality

This factor influences ease of starting, duration of white smoking after start-up, drivability before warm-up, and intensity of diesel knock at idle. Studies have correlated ignition quality with all regulated emissions. As ignition delay is reduced, the combustion process starts earlier and emissions (primarily CO and hydrocarbons) are reduced.

4.3.4 Cetane Number

Ignition delay is measured by the CN test (ASTM D-613), which uses a single-cylinder, variable compression ratio engine analogous to the octane number engine. The factor measured is ignition delay (rather than knock) at a fixed compression ratio, and it is compared with that of standard reference fuels consisting of blends of *n*-cetane and heptamethylnonane. The cetane content of the blend that matches the ignition delay of the test fuel is the test fuel's CN. It is diesel's analog to gasoline octane.

Few laboratories are equipped with cetane engines, so a number of correlations between CN and physical properties have been developed as substitute methods. Best known is the ASTM D-976 calculated cetane index, which uses the density and mid-bp of the fuel.

Diesel engines vary more widely in their cetane requirements than do automobiles in their octane requirements. In general, the slower an engine's operating speed, the lower the CN of the fuel it can use. Large marine engines can tolerate fuels with CNs as low as 20, while some manufacturers of high-speed passenger car diesel engines specify 55 CN fuel. Railroads purchase fuel to a 37 or 38 CN minimum specification, but most truck manufacturers call for a 40 or 45 minimum. CN can be increased by using cetane improvers. These additives, usually organic nitrates, boost CN by 2 to 7 numbers, depending on the dosage and type of base stock used. These additives directly contribute to NO_x emissions, however, and have come under scrutiny from various regulatory agencies.

4.3.5 Viscosity

Viscosity influences the spray pattern when the fuel is injected into the cylinder. Low-speed marine engines can use higher viscosity fuels than high-speed road transport engines, and still run without excessive smoking. Minimum viscosity limits are imposed to prevent the fuel from causing wear in the fuel injection pump.

4.3.6 Low-temperature Flow

Unlike gasolines, which have freezing points well below even the most severe winter ambient temperatures, diesel fuels have pour points and cloud points well within the range of temperatures at which they might be used. This presents no problem on ships or even in railroad applications where heated storage can be arranged. However, for on- and off-highway use, precautions must be taken to tailor the low-temperature properties of the fuel to the weather.

Seasonal blending to control cloud point (the temperature at which wax separates from the fuel) is the refiner's assurance against field problems. ASTM D-97 is the procedure used to determine a fuel's cloud and pour points. In the winter, there is an increasing tendency to use flow

improvers, polymeric additives that modify the wax structure as it builds up during cooling. These additives keep wax crystals small, so they can pass through the fine pores of fuel filters en route to the injector pump. With suitable base stocks, flow improvers can keep an engine running at temperatures as much as 10 °C below the fuel's cloud point.

In addition, pour point depressants are sometimes used in diesel fuel to ensure that it can be dispensed at low temperatures. Tests such as the cold filter plugging point (CFPP), IP 309, are used to assess the potential of a diesel fuel to have cold weather field problems.

Other ASTM standards dealing with cold environment performance, are D-2500 and D-4539.

4.3.7 Storage Stability

In storage, diesel fuels are attacked by atmospheric oxygen, which can cause deposition of varnish, and for marine fuels containing residual components, asphaltic material. Antioxidants and dispersants are added to prevent such problems, while copper metal deactivators reduce the catalytic effects of screens and other parts. In the presence of water, bacterial action can cause a build-up of slime in the storage system, leading to filter plugging. Biocides are added to inhibit bacterial growth. In cold weather areas, there is the risk of static electric charges building up during high-rate dispensing of distillate fuels. Refiners include antistatic additives in diesel blends to prevent explosions.

The ASTM standards dealing with storage stability in diesel cuts are D-381 and D-5304.

4.3.8 Component Materials Compatibility

Diesel fuels are injected into the engine through precision pumps and fine injector nozzles. Dirt and water contamination must be avoided to protect these critical components. Specifications include tight limits on water and sediment, but some fuel marketers also install final filters at the nozzles of service station pumps to protect against dirt picked up in the distribution system.

Depending on the crude source, diesel fuels contain various amounts of sulfur compounds, which yield corrosive sulfur oxides on combustion. These can cause high rates of engine wear and rapid depletion of engine oil additives. Engine manufacturers often relate oil change intervals to the fuel sulfur content.

Deposit buildup in engines is influenced by fuel quality. Fuels that leave a heavy carbon residue and contain excessive amounts of high-bp materials are prone to cause engine deposits. Therefore, limits are placed on carbon residue and ASTM D-86 90% evaporated temperature.

Again, as for gasoline, the ASTM standard dealing with materials compatibility is D-130.

4.3.9 Sulfur Content

Sulfur content is the first diesel fuel property to be widely controlled by legislation aimed at limiting exhaust emissions. Sulfur is present in all crude oils and refined products. During combustion, sulfur compounds burn to form acidic by-products, SO_2 and SO_3, which form sulfates in the exhaust gas stream. Sulfates are part of a diesel engine's particulate emissions; therefore, controlling fuel sulfur level reduces the level of sulfate pollutants.

4.3.10 Water Content

All diesel fuels contain small amounts of water. The amount that a fuel can hold is controlled by hydrocarbon type and distribution, and bulk temperature. As temperature decreases, the amount of water dissolved in the fuel will also decrease and may lead to a water layer forming on the bottom of the storage vessel. This layer should be minimized by draining the tank regularly to prevent bacterial contamination and the pumping of water into the fuel distribution system. Excessive water in a fuel system can cause corrosion, filter plugging and icing (in the winter).

A tabular summary of diesel fuel performance/property relationships is presented in Table 7.

Table 7 Fuel performance and property relationships for diesel fuel

Distillation	
10%	Starting, warm-up, odor, smoke
50%	Particulates
90%	Heat content, metering
Gravity	
CN; cetane index	Cold starting, warm-up, white smoke, black smoke
Pumpability	
Viscosity	Fuel atomization, penetration, leak back, lubrication
Cold flow	Cold weather operability
Cleanliness	
Appearance	Contamination
Water and sediment	Contamination
Sediment stability	Filter plugging
Demulsibility	Engine damage
Safety	
Flash point	Safety
Miscellaneous	
Sulfur	Engine corrosion, deposits, emissions
Mercaptan	Odor, copper mercaptide gel
Carbon residue	Indicator of contamination
Ash	Engine wear, contamination
Corrosion	Corrosion of copper and brass
Acid number	Indicator of contamination

4.4 Aviation Turbine/Jet/Marine Fuels

The important properties of kerosine-derived turbine fuels are volatility, heating value, low-temperature cloud point, storage stability, component compatibility, and sulfur content. There is no overall defining specification for a generic turbine fuel because of the wide variety of engine types and requirements that the fuel must meet. There are standards, however, for individual turbine fuel types mostly defined by the type and service of the turbine involved. The most important standards are set by military requirements since jet fuels are the biggest users of this fuel stock. In spite of this, there are far fewer performance requirements than for spark or compression ignition engines. The current ASTM standard for turbine fuels performance is D-1655.

There are also some ISO standards for performance and scores of military specifications (nearly one for each country with an airforce), some of them conflicting in minor areas with the other specifications. It would seem that, unlike normal gasoline purchase customers, the military block in the world both sets the refining specifications and buys the product that they have specified and subsequently use. Quantity purchases allow for this customer perquisite. Fortunately, most of these specifications agree in important areas. If they did not, there would be major problems with some jet fuels, of which the only indication might be a crashed plane. Thus, jet fuel misspecification problems are somewhat self-limiting.

The overall performance standard for this fuel is D-1655.

4.4.1 Volatility

The volatility of a turbine fuel has little influence on its engine performance, except as it affects exhaust smoking tendencies and fuel efficiency. The distillation range of jet fuels is tightly defined by military jet engine specifications and have been so for years. As with gasoline, a controlling factor is the ASTM D-86 distillation 90% evaporated temperature.

Because turbine fuels are often handled in large quantities by numerous commercial and military installations, minimum flash point restrictions are imposed. ASTM D-56 is the applicable standard here. This also makes it easier for refiners to store jet cuts than for them to store the more volatile gasoline blend stocks and products.

A consideration for static electricity buildup (non-ionic, nonelectrolyte fluid flowing through a metal conducting pipe tends to do this for gasoline, kerosine, and diesel cuts) should be mentioned. There is no such standard for gasoline or avgas, although the problem can exist with them as well. The diesel flash point and flammability characteristics generally (not always, but typically so) rule out severe static susceptibility. The typical military quartermaster loves to have specifications for nearly everything; therefore the static ignition on jet fuels is maintained. ASTM D-4856 is a rough composite of several historical military specifications combined with some practical materials engineering and grounding practices.

4.4.2 Heating Value

The same considerations apply here as for the diesel fuel heating value commentary.

4.4.3 Low-temperature Flow

Unlike gasolines, which have freezing points well below even the most severe winter ambient temperatures, kerosines would normally have cloud points well within the range of temperatures at which they might be used. This presents no problem on ships or even in railroad applications where heated storage can be arranged. Wax formation in high-altitude aircraft is extremely hazardous to flow of the fuel to the aircraft engine. Since ambient temperatures in such an environment may be $-20\,^{\circ}\mathrm{F}$ ($-29\,^{\circ}\mathrm{C}$) or less, jet fuels have a large number of guidelines for what they may contain. Error in specification formulation for jet fuels is always on the side of safety.

4.4.4 Storage Stability

In storage, jet/turbine fuels like gasoline are also attacked by atmospheric oxygen, which can cause deposition of varnish and gums, but since these fuels typically are used in high volumes, long-term storage problems are not of as much concern. Usually military specifications prescribe refining procedures that minimize the amount of oxygen-sensitive contaminants in the fuel as delivered. Water is always a problem, however, especially in humid climates. In the presence of water, bacterial action can cause a buildup of slime in the storage system, leading to filter plugging. Biocides are added to inhibit bacterial growth. In cold weather areas, there is the risk of static electric charges building up during high-rate dispensing of distillate fuels. Refiners include antistatic additives in jet fuel blends to prevent explosions.

4.4.5 Component Materials Compatibility

Kerosine fuels are injected into the engine through pumps and injector nozzles. Dirt and water contamination must be avoided to protect these components, but the greater need is to protect the turbine engines and their high-speed moving parts. Most jet specifications include tight limits on water and sediment, but some fuel marketers also install final filters at the nozzles of aircraft service

Table 8 Fuel performance and property relationships for jet fuel

Performance characteristic	Fuel property
Engine fuel system deposits and coke	Thermal stability
Combustion properties	Luminometer number; smoke point; aromatics; percent naphthalenes
Fuel metering and range	Gravity; heat of combustion
Fuel atomization	Distillation; viscosity
Compatibility with elastomers and metals in fuel system	Mercaptan sulfur; sulfur
Fuel cleanliness, handling, and storage stability	Flash point; existent gum; potential gum

truck distribution pumps to protect against dirt picked up in the distribution system. Other contaminants are generally specified as they are for gasoline and are tested for in exactly the same fashion.

A tabular summary of turbine fuel performance/property relationships is presented in Table 8.

4.4.6 Benzene and Toluene

The governing performance standard here is ASTM D-3601.

5 SOURCES AND FORMULATION OF TRANSPORT FUELS

5.1 Refinery Configuration[1,3,7,12,32,53]

5.1.1 Introduction

Because the greatest abundance of transportation fuels in the world is derived from petroleum-based hydrocarbons, it makes sense to examine the workings of a petroleum refinery briefly. As mentioned before, naturally occurring petroleum has in its mix light naphtha, heavy naphtha, and light gas oil used respectively in gasoline(automobile and aviation), kerosine, jet/turbine fuels and diesel products. In the earliest days of the industry, these products were simply distilled from the feed oil. It has been said that the modern IC engine was basically built as a means to take advantage of the cheap and abundant ''white gas'' by-products of the refining of petroleum into lamp oil, lube stock and wax products more in demand at the end of the 19th century. Now, of course, the automobile and its demands virtually determine the market for petroleum and its products worldwide.

Today's sophisticated transport machinery has more exacting specifications than can be typically met by using direct distilled, or as commonly known, ''straight-run'' fractions. In order to meet these requirements, refineries are designed to make a number of motor fuel components and streams which vary greatly in octane, volatility and other drivability parameters.

Refineries not only differ in size (BPD of crude processed), they comprise different processes and operations (configurations). The configuration of a refinery depends primarily upon the type of crude(s) available as feedstock and upon the product slate requirement. In general, the products which dictate refinery design are relatively few in number, and the basic refinery processes are based on the high-volume products such as gasoline, jet fuel and diesel fuel. Storage and waste disposal are expensive. It is necessary to sell or use all of the items produced from crude oil even if some, like heavy fuel oil, must be sold at a price less than the crude oil cost. Careful economic balances are required to determine whether certain crude fractions should be sold as is (straight-run) or further processed to increase their value.

5.1.2 Types of Refineries[35]

The simplest and earliest configuration is the skimming (topping) refinery, a schematic of which is shown in Figure 2. Here the major processing step is crude oil distillation, the object being to remove the lighter fractions from the residuum. When only one distillation column is used, it is operated at atmospheric pressure. Typical products will be gases, light naphtha, heavy naphtha, light gas oil, heavy gas oil and a 675 °F+ residuum. A second distillation column operating under vacuum may be added to gain more, and heavier, gas oil. This in turn produces less, but a heavier (975 °F+), residuum.

The liquid sidestream products are further stabilized in individual strippers before leaving as ''straight run'' naphtha, and so on. Contaminants are not removed in straight-run products and are present in proportion to their distribution in the whole crude. Also the quality

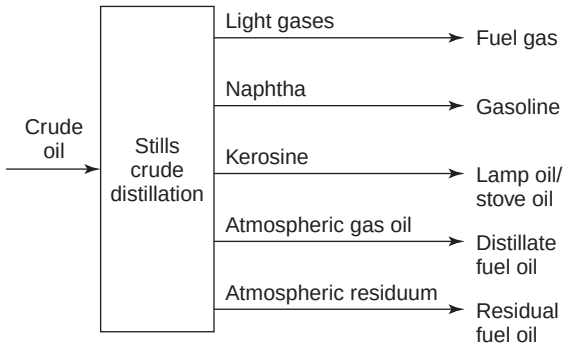

Figure 2 Topping plant scheme.

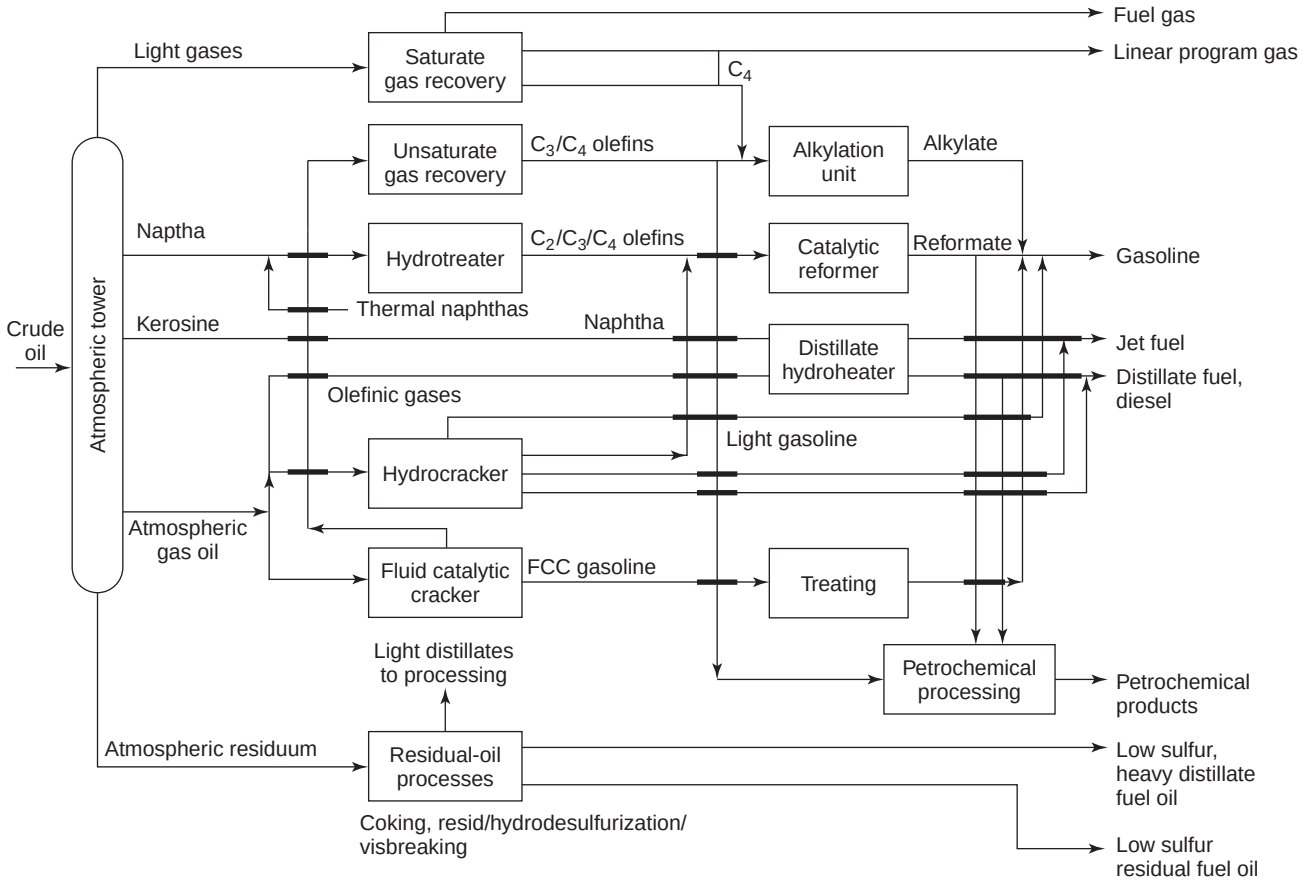

Figure 3 Modern conversion refinery.

(octane, pour point, contaminant levels, etc.) of these straight-run oils is usually at a minimum level.

A more complex configuration results when hydrotreating[54] processes are added to remove contaminants from the straight-run oils (hydroskimming refinery). Figure 3 shows a simplified process block flow diagram of this kind of plant. Olefins and aromatics are also saturated in hydrotreating. To supply the hydrogen consumed in these processes a catalytic reforming operation is often included which itself greatly increases the octane pool potential of the refinery (at the expense of adding a lot of benzene and aromatics to the mix). If reforming does not produce sufficient hydrogen, a hydrogen manufacturing unit must be added, too.

Hydrotreating increases the quality of the liquid products primarily from an environmental cleanliness standpoint by removing the sulfur, nitrogen, and some metals impurities. In some instances, benzenes and other aromatics can also be removed by hydrotreating, though this option is fairly expensive.

Catalytic reforming increases the quality of the naphtha by raising the octane number rating of the clear unleaded gasoline fraction.

When the market for gas oil products is smaller than their yield, they are cracked to lower boiling range sales products such as liquefied petroleum gas (LPG), gasoline and diesel fuel. This is done by catalytic cracking (FCC unit) or hydrocracking units. Because the product slates of these two processes are different, market requirements must be considered in choosing which one to use. In most large refineries, both types of cracking will be used.

Hydrocracking consumes a large quantity of hydrogen (1500 to 2500 standard cubic feet per barrel) which is much greater than that produced by catalytic reforming. A hydrogen manufacturing plant is included in these refineries to supply this excess. The need for a separate hydrogen plant as well as the high operating-pressure requirements, make hydrocracking an expensive operation. This is made up for by the significantly higher quality products which are formed.

If the residuum (heavy fuel oil) produced exceeds the market demands, some type of residuum treatment process must be incorporated. In a conversion refinery a visbreaker or delayed coking or fluid coking is used to crack the residuum to naphtha, distillate and gas oil product. When the coke market is small and the

refinery can use low British Thermal Unit (Btu) fuel gas, flexicoking might be used. Alternatively, a residuum hydrocracker may be used to upgrade the entire stream.

The liquid products from a coking process generally are not very stable. They contain olefins and active aromatics and have high sulfur and nitrogen contents. These liquids are usually hydrotreated to improve their quality.

Figure 3 is a simple block flow diagram of a modern conversion refinery. A comparison of Figures 2 and 3 graphically illustrates some of the increased complexity of modern refinery processing. This particular example has three processes for handling the vacuum residuum: solvent deasphalting, delayed coking and fluid/flexicoking. Many refineries, particularly in cold climates, do find it economical to produce a significant amount of heavy fuel oil in addition to fuel-destined naphtha and coke by-products.

Another option for residuum processing, if the main crude has the proper makeup, is lubricants production. Not all crudes have an abundance of good lube stock components. The main feedstock for lube oil products is usually heavy vacuum gas oil. Lube oil can also be made from hydrotreating/hydrocracking products so there can be a variety of configurations to include this capability. In the refinery of Figure 3 several additional steps would be added. The lube oil feedstock (vacuum gas oil, deasphalted oil, etc.) would be sent to a solvent extraction unit. Here aromatics and other undesirable components are removed to produce a stable oxidation-resistant lube oil product. The lube oil is further treated by solvent dewaxing to lower the pour point. The dewaxed oil is then clay treated to decolorize it to produce the final product.

When a refinery has thermal cracking (visbreaking or coking) or catalytic cracking, significant quantities of light olefins (ethylene, propylene, butadiene, etc.) are produced. Some refineries include processes such as ethylene/propylene separation and butadiene separation in addition to aromatics (benzene, toluene, xylenes) separation. These materials are then sold as petrochemical feedstocks.

A critical process for making reformulated fuels is the alkylation unit. This unit takes isoparaffins and combines them with olefins to produce longer-chain branched naphtha molecules. By using only C_3 to C_5 olefins with isobutane, high octane, low olefin, low RVP, no-aromatics gasoline can be made. While the alkylate product is an extremely valuable commodity, the capital and operating costs of alkylation units are quite high and are catalyzed by either sulfuric or hydrofluoric acid which present refiners with some safety and environmental concerns of their own.

Table 9 shows a textual explanation of the source and distributions of typical refinery streams.

5.2 Fuels Blending

Modern gasolines have such a variety of specifications to be met in each blend[15,18,55–59] run, as many as 30 including those for both performance and environmental criteria. Few if any of the process unit naphtha streams in the refinery (a dozen at most) could singly meet them all. This represents a mathematical imbalance that can be rectified only by using a fairly sophisticated blend procedure to minimize both deviations from specification and economic giveaway. Some of the streams are more expensive to make than others and if one particular blend were to take too much of a more desirable feedstock, say alkylate (low RVP, high octane, no aromatics or olefins), the refinery would be left with no choice but to use the other stocks for its subsequent product demand. This, in turn, gives a greater risk of making off-specification material or will force the refinery to cut production capacity in other units in order to restore blend stock ratios. Neither option should be used very often in today's competitive market. A downgrade of premium fuel to regular grade, or from CARB spec to US EPA Phase I specification may cost as much as $4.00 per barrel of product which is a $280 000 loss for a 70 000 BBL blend run.

The specifications themselves may also vary in type and approach to blending implementation. The normal drivability and regulatory specifications mentioned elsewhere in this article are diverse indeed. Most are expressed as either minima or maxima, some being one or the other, others with widely varying bounds on both sides of a tolerance range. A few such specifications, octane rating being the most prominent, may have tolerance limits so narrow as to be essentially a fixed value. In this case the minimum specification would be determined by commercial or contractual needs and the maximum might be from the standpoint of refinery ''giveaway'' economics.

Figure 4 shows a basic analyzer-control schematic for fuels blending.

Commercial methods for blending motor fuels may be one of two kinds, batch tank blending or the more modern in-line blending. Small sequential batch-to-receiving tank blending (the traditional way) is normally time-consuming and involves trial-and-error testing until a given batch meets specification, making allowances for testing errors. Modern refineries have generally adopted some modification of direct in-line blending or massive blend tank (50 000 BBL + at a time) batching with multiple simultaneous stock feeds. These methods have some outstanding advantages. They are continuous (or mostly so with giant size batches) rather than batch mode, thus saving time. They provide a much closer control in meeting critical specifications and require less tankage. The multiple feedstock protocol generally prescribes

Table 9 Refinery processing source and distribution of streams

Unit	Feeds	From	Products	To
Atmospheric crude	Crude	Tankage	Gas, naphtha	Saturate gas
			Kerosine, distillate	Treating, hydrotreating, product
			Light gas oil, heavy gas oil	Hydrotreating, hydrocracking, product
			Atmospheric residuum	Vacuum, delayed coking, hydrotreating, product
Saturate gas	Gas, naphtha	Crude, hydrotreating, hydrocracking, reforming	Gas	Product
			Propane	Product
			t-Butane	Alkylation, product
			n-Butane, light naphtha	Isomerization, product
			Heavy naphtha	Reforming
Reformer	Naphtha	Hydrotreating, hydrocracking	H_2-rich gas	Hydrotreating, hydrocracking
			Gas	Saturate gas
			Light naphtha	Saturate gas, product
			Reformate	Product
Hydrocracker	Gas oil	Crude, vacuum	Gas	Saturate gas
			Light hydrocrackate	Saturate gas, product
			Heavy hydrocrackate	Reforming
			Distillate	Product
Catalytic cracker	Gas oil	Vacuum, coker (via hydrotreater)	Gas, naphtha	Unsaturate gas
			Light cycle oil, heavy cycle oil	Product
			Tar	Coker, product
Coker	Residuum, tar	Crude, vacuum, catalytic cracker	Gas	Unsaturate gas, product
			Naphtha	Unsaturate gas
			Gas oil	Hydrotreating, catalytic cracking, product
			Coke	Product
Unsaturate gas	Gas, naphtha	Catalytic cracker, coker	Gas	Product
			Olefins	Alkylation
			Naphtha	Product
Vacuum	Atmospheric, residuum	Crude	Gas oil	Hydrotreating, hydrocracking, catalytic cracking, product
			Vacuum residuum	Coker, de-asphalting, visbreaking, product

that the blend header specification must quickly become ''on-spec'' within a certain percentage of the blend having traveled down to the receiving tank and the total blend run, plus the residual ''heel'' in the blend tank must blend together and still meet specifications. True ''in-line'' blending directly to customer carrier (pipeline, railcar, tanker) is even more stringent procedurally. The product must be ''on-spec'' almost immediately and the total blend must meet specs – there obviously is no room at all for reblends in this case.

In-line blending became a practical reality with the advent of sophisticated, automated equipment and instruments, along with distributed (computer) control systems. With this method, the first step is to produce in segregated tankage the required amounts of several appropriate blend components which, according to the blender's calculations, will result in meeting the target blend specifications when the components are blended in the proper proportions. Close control of certain specifications of these components as they are produced, particularly RVP and octane number, is achieved by continuous monitoring, using suitable automated equipment. The final step is to pump these components at controlled rates into a common line that discharges into a vessel which may be a tank, pipeline, barge, or a tanker. This operation can be activated and controlled by a suitable advanced computer control system.

Figure 5 displays the approach to final blend value of RVP and RON for an 87.5 RON regular grade of unleaded fuel. Notice the approach to the target of RVP as opposed to that of the octane. Before RFG, the giveaway strategy would have been reversed.

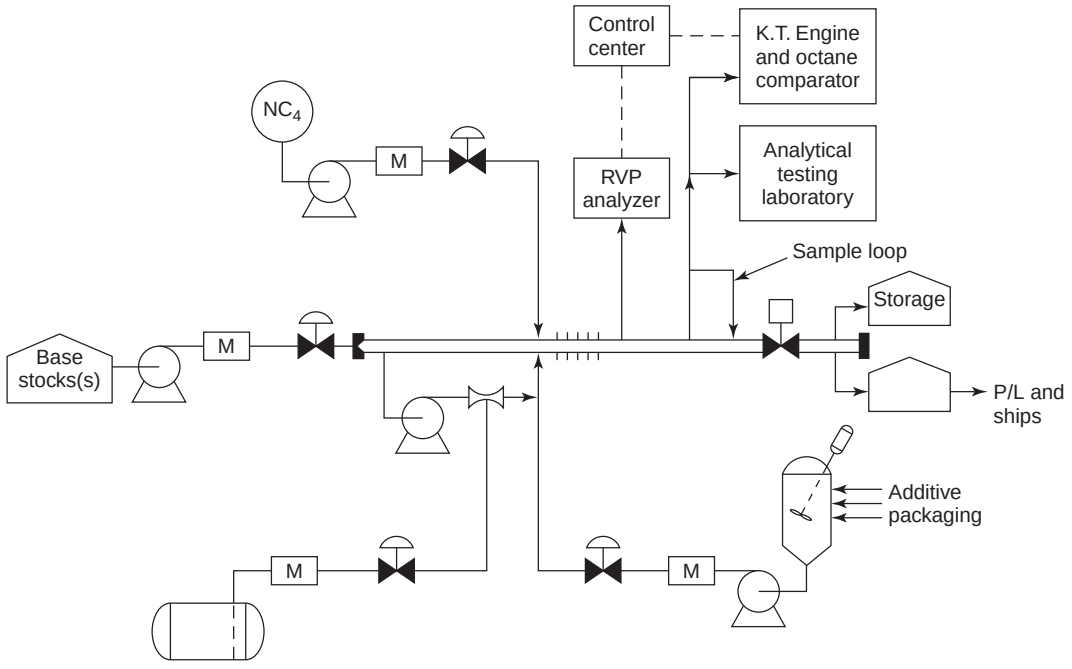

Figure 4 Simple in-line blender process with analyzer control.

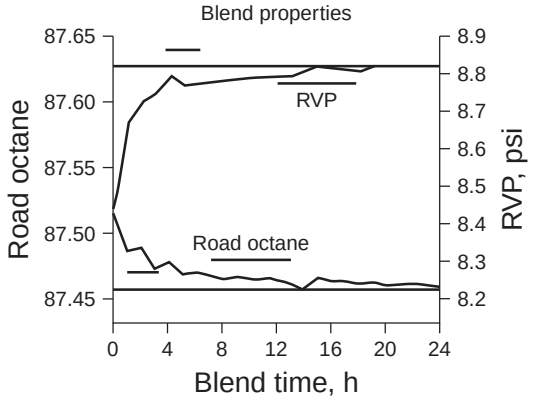

Figure 5 An example of octane and RVP control.

5.3 Advanced Control and Monitoring of Gasoline Blending

5.3.1 Introduction[59]

The following description is a typical procedure that might be followed in the course of setting up and running a gasoline blend in a semi-batch fashion in typical runs of 20 000 to 150 000 barrels sizes depending upon the size and complexity of the refinery.

The key to success is getting all the right information to flow around the system. Blend recipes from the refinery planning group must be received in proper and timely fashion by the blend operators, who must then set up the actual plant flow paths to accomplish the blend. Historically, this function was a simple matter of verifying the availability of stocks and setting the blend ratio by hand or setting up a sample slave-ratio controller to maintain the recipe. Trim control of rates during the run was a relatively simple matter. With the advent of several additional constraints upon the final product specification and availability of almost no new stocks for the planners, the problem of making both customers and regulators happy with the final product took on a new dimension. With several reductions in available degrees of freedom, the mathematical solution to this problem became a matter of a QP (quadratic programming) optimization problem, with the objective function being a residual sum of squares error on quality targets and optimal recipe targets. Without the use of high-quality analysis feedback and a multivariable controller (MVC), it would be nearly impossible for a refinery to blend Phase II fuels and stay in business for long.

5.3.2 Blend Run Set-up for Execution

The first thing that must be transmitted before a blend is configured is the blend summary, which contains approved blend entries downloaded to the blending system from a blend planning engineer. These recipes are based on optimized recipes run through a multiperiod optimizer (typically linear program (LP) with recursion or successive linear program iteration, called SLP) to handle the nonlinearities of predicting stock octane and assay blending. Most commercial systems have what might be

called a blend summary administration screen. The blend summary screen might display such minimum information as blend identification, finished product to be made (type, name, etc.), receiving tank, target volume, plus start and stop times.

Underneath the blend summary screen one should expect a blend entry screen which holds all the information needed functionally to define a blend. This would include one or more subdisplays for blend administration information plus the month/season for quality specifications to be met for this product, target volumes of feeds and product to be blended, receiving tank identifier and initial volume (heel) for finished product tank, percent of blend volume by which the finished product tank must meet specifications.

A more specific type of control screen is for blend recipe control. Information here would cover such items as components or additives to be used in a blend, primary source tank for feed component/additives, secondary component/additive source tanks to use in place of primary tank (if, perhaps, this tank runs low and approaches its minimum safe heel), component recipe percentages defined by the blend engineer, off-line optimizer results that are placed here when an off-line optimizer is run and its results are accepted, minimum and maximum component recipe percents, and other information needed by the blender operator.

Yet another screen typically separates feedstock control from final product control. Such a blend quality control screen allows the blend operator to specify or control the following: the finished product quality, blended product quality values as projected by the blend engineer to start the blend, minimum and maximum product specifications, blend move limits, constraints the quality should have a limit on; maximum (high) constraints, minimum (low) constraints, or should not limit at all (or let the optimizer use these limits), and finally blend header quality values, from the starting recipe to the final recipe projected by the off-line optimizer. Most blender systems also have text entry blocks that allow the operator to make comments to describe any special conditions of this blend, alarm summary statistics, automatic management report generation and routing, and so on.

Once the blend entry is defined and set up by the operator, equipment for flow control of the blend must be selected. The operator chooses the appropriate blend entry (from several that are being planned or that may be available for execution) for run preparation. A flow path must be chosen for each component tank. This path defines the pump(s) and meter(s)/flow controllers(s) to be used for this blend component. A default path is typically suggested by the system. This default path or another path may be used. The path flow limits are calculated for the pump/meter combination of the chosen loop. If the component flow rate required to achieve the blend rate is outside of the equipment flow limits, a message indicates this condition. The operator can choose a different loop or go ahead with this choice on the understanding that the necessary component flow cannot be reached.

A typical blend equipment selection screen displays the following: tank movement status (current flows into and out of the stock tanks), component/additive or finished product tank to be used in the blend, material in the tank, primary and secondary pump and flow meters used for component, product transfer for this particular loop, minimum and maximum limiting pump or meter flow rates, component/additive or finished product flow rate required for specified blend entry.

The basic blender control systems will stop here. The more sophisticated ones will include not only multivariable control functions but economic optimization control as well. This feature can be divided into two phases of operation, off-line and on-line optimization.

5.3.3 Off-line Recipe Blend Optimization[59]

After equipment has been selected for the blend entry, an off-line optimizer can be used to check the feasibility of the intended blend with actual current system values. Current tank qualities are used in the off-line optimization. The off-line optimizer is fundamentally the same as the on-line optimizer in that it uses the current rundown rates, tank inventories, and batch transfers in checking stock availability. This helps to find potential problems with the blend before the blender is actually started. If the product specification qualities cannot be met using the specified component recipe and constraints in the blend entry, the optimizer will give an infeasible result (and hopefully flags it to the operator as such!). This feature provides an important check on the validity of the refinery planning function where the original blend recipe originated. In a typical refinery, the planning group may not have access to all of, or even the same, information that the blend operator has.

The goal of an off-line optimizer is to find a recipe which primarily meets quality specifications for the product at or before the end of the blend, then secondarily reduces quality giveaway. The objective function is a minimization of recipe deviations and quality giveaway. Constraints for this optimization include final product quality constraints, recipe limits, path flow limits, and tank working volume limits.

An infeasible result means that the product quality specifications cannot be met with the components in the current recipe and the current path constraints. The constraints could be made less binding or the recipe could be changed. In any case, the limiting constraints will have

to be modified and the optimizer rerun until a feasible result is achieved.

If the results are feasible, the optimized recipe can be accepted into the blend entry, replacing the previous initial recipe used for the optimization. Off-line optimizers will typically show the blend engineer the following information: type of constraint encountered from beginning of blend to on-spec percentage of total blend, optimized component recipe from on-spec percentage of total blend to end of blend, optimizer result – infeasible or feasible, optimized quality value at both the blend header and receiving tank when on-spec percentage is reached and when end of blend is achieved, path flow limits and infeasibilities, and stock tank conditions at end of blend. Finally, there needs to be a means of accepting the results (basically a yes or no answer.) A feasible result only should be accepted for download into the blender distributed control system (DCS).

5.3.4 Blend Run Initialization, Launch and Monitoring

When the results of the off-line optimizer are satisfactory, the blend entry information can be passed to the header quality control function and downloaded to the DCS to initialize the blender. The recipes, quality targets, and most blend entry information are passed on to the DCS. Recipe and quality targets are passed to the header multivariable controller (MVC) quality controller. The blend entry is then executed from the DCS. After the blend is started and operating near full flow rate, on-line optimization and header quality control can be activated.

Figure 6 shows a schematic of an MVC based CARB blender package.

During the blend, header qualities are monitored by intermittent on-line analyzers, calculated, and manually entered data. The analyzer values are available through the DCS. For every on-line analyzer, a good/bad status is needed to indicate whether the value is useful for control. For each analyzer, a status is required to indicate when a new result is available.

A well-designed blend launch and monitoring screen displays the standard blend identifier information plus header and predicted tank qualities in real time, analyzer status (good/bad) and results along with ongoing predictions of header qualities, real-time tank movement data, updated laboratory results, and alarm messages for problems and constraints encountered during the run. For US EPA Phase II or CARB specification blending, a multivariable advanced controller is essentially a necessity. The operators should always be able to see the status of

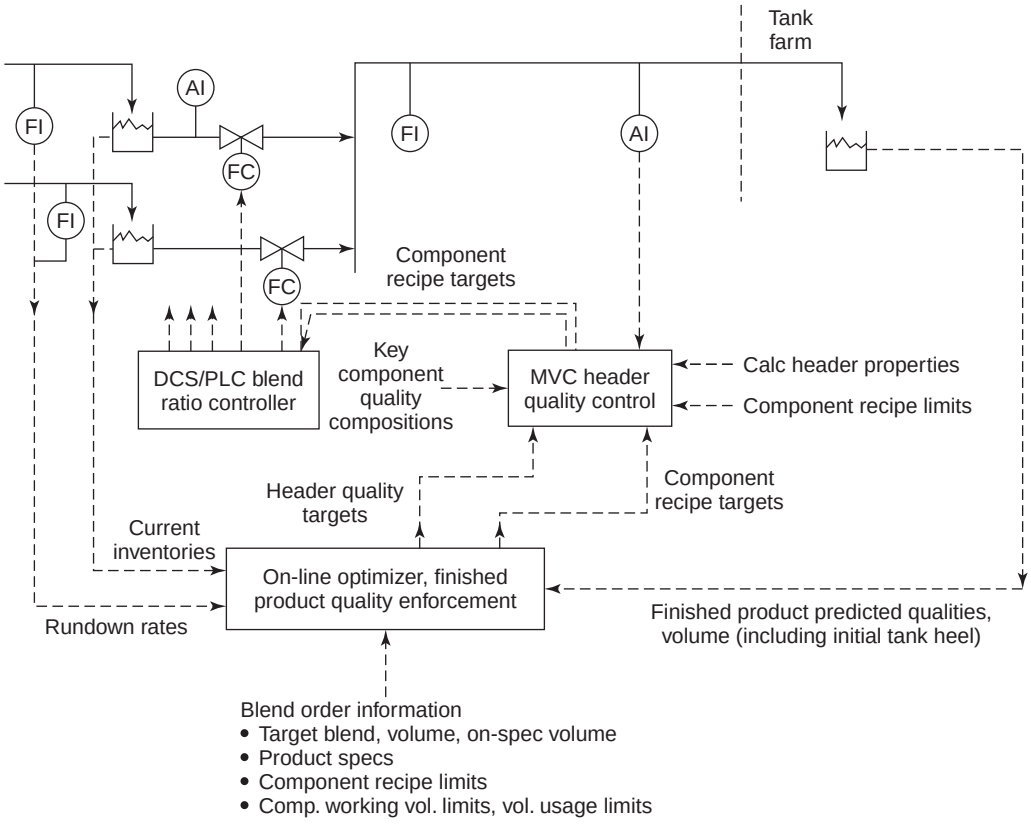

Figure 6 Multivariable blender control (PLC = programmable logic controller).

the MVC unit and the metrics by which its control can be judged.

Figure 7 shows a typical control strategy that typically might be used on a CARB blend recipe. Only three control variables (CVs) are targeted and only a total of seven are available with the stocks available to the blender in this case. Controlling the T10 and the RVP simultaneously, in fact, often poses control problems. If this happens, one or the other criteria must be relaxed somewhat. In typical MVC systems, a two sided control range is an indication that a quality has a regulatory (usually maximum) constraint and a set of economic criteria (if not exactly constraints) or performance quality constraints which are usually the determinants of the lower control constraints. Such conflicting goals are the reason that MVC is essential in making CARB or Phase II gasolines.

Figure 8 shows the essential monitoring function that a quality controller must be able to perform. The final

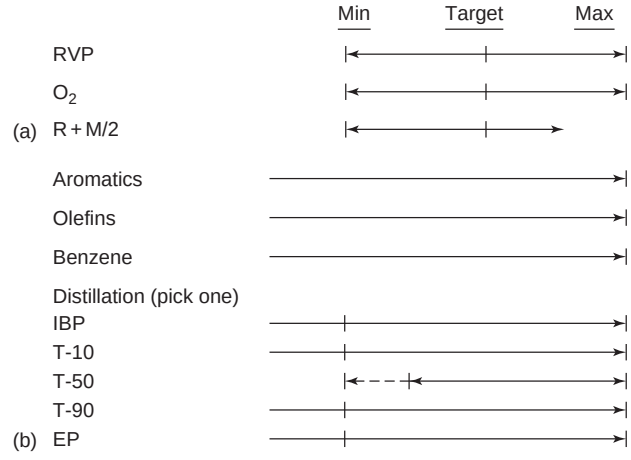

Figure 7 Blend target strategies. Overall design objective is to maximize degrees of freedom to create a blend that meets multiple specification, leading to (a) target and constraint variables versus (b) constraint only (zone) variables.

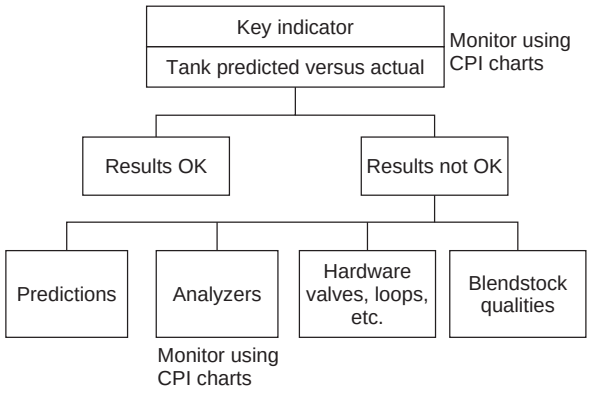

Figure 8 Monitoring the blending system for performance (CPI = chemical process industries).

tank or target pipeline specification must continually be examined for predicted to measured accuracy and corrective action taken if discrepancies occur, including shutdown reoptimization of the recipe and restart of the blend process.

5.3.5 On-line Blend Optimization[59–62]

As the blend run proceeds, the controller varies the signals sent to the individual component flow loop controllers based on the multitude of information received from flow meters, pump motor power meters, analyzers, temperature and pressure indicators and perhaps a parallel predictive or ''virtual'' model of the blend properties as well. The MVC tries basically to minimize the deviation of all flow rates from the specified setpoints. All controller output moves that it makes come from the multivariate nature of the blend problem with competing specifications and quality needs that require constant give and take in order to minimize the deviation from the target of any one quality in the final blend.

Most MVC installations control a combination of predicted qualities modified by regularly updated analyzer values during the course of the run. For whatever reasons, the predicted and measured qualities will diverge to some extent. Occasionally this divergence is significant. In such cases, with the MVC package and being constantly mindful of constraints and competing quality requirements, the blend is still quite feasible but may no longer be as economical as possible. An on-line optimizer provides the blend operator with a chance to redefine the remainder of a blend recipe based on a set of economic and operating criteria.

The result of an on-line optimizer is a new set of setpoint targets. Generally, on-line optimizers will run every 2–10 min, constantly reexamining the ''optimal blend'' problem based upon what has already been passed through the blend header and what is still left in the stock tanks, all the while heeding the same quality specifications and process constraints that govern the MVC. When the new ''optimal'' and current setpoints differ to some significant amount, the on-line optimizer will flag the condition and suggest that the blend operator makes a move. Usually the new setpoints end up being the ideal resting values (IRVs) for the MVC as it wrestles with the multiple specification problem it must satisfy. Making these IRVs the new setpoints will usually make the MVCs job easier and the control of the blend becomes quite stable in operation. Every once in a while, a well-designed on-line optimizer will recognize money-making opportunities such as the opportunity to blend more butane or high RVP straight run naphtha into a blend that is otherwise heading towards a serious RVP ''giveaway'' situation.

6 FUELS ANALYSIS FOR REGULATORY REQUIRMENTS[11,23,31,53,63,64]

6.1 Introduction

The technology available for the chemical and physical characterization of petroleum and chemical feeds and products has improved steadily. Better detectors and better electronics have resulted in higher sensitivity and speed of analysis for many of the more well-established techniques. At the same time, new scientific concepts and the availability of low-cost minicomputers have resulted in the creation of new techniques that provide novel information. The integration of the information from these new, and often extremely expensive, instruments results in a remarkably detailed knowledge of the composition of various process and product streams, particularly when the various separation processes are appropriately combined with the other instrumentation. The following sections provide a brief summary of the more commonly used techniques available today, together with a number of basic references that provide a more detailed discussion of both the scientific principles and areas of application for each system.

6.2 Analytical Procedures – Component Separation Methods[1,3,56,65–68]

Petroleum streams and chemical feedstocks are usually quite complex mixtures and the analysis of these systems can frequently be simplified through some form of prior separation. The separation techniques commonly employed are distillation, to narrow the boiling range, and chromatography, which provides both boiling range and compound-type separations. Much of the recent progress in obtaining more detailed compositional data on petroleum and chemical streams can be attributed to the wide use of chromatography. Such separation techniques are also used to isolate pure compounds or compound types needed as calibration standards for the instrumental methods describe later.

6.2.1 Distillation

Distillation tests are generally performed on products destined to be used under conditions involving conversion to the vapor state. Thus, the distillation characteristics of all fuels are important quality considerations. Since solvent quality depends on the rate of evaporation, distillation characteristics are also important for this type of product. In the case of heavier products such as lubricating oils, distillation is carried out under reduced pressure to avoid cracking. Here distillation characteristics are useful in predicting oil consumption in automotive crankcases in certain types of automotive service.

Distillations are reported in one or more of the following terms:

1. IBP
2. distillation temperatures for increments of volume overhead
3. percent evaporated at specific temperatures
4. EP or maximum temperature
5. dry point
6. recovery or loss
7. residue.

Five standard laboratory procedures for distillation of petroleum-based fuels have been standardized. Three are run at atmospheric pressure and one at reduced pressure. Four of the atmospheric methods, D-86, D-216, D-285, and D-1078 are of a type that results in a very modest degree of fractionation effected by a minimal reflux of the distillate in the neck of the flask. The method at reduced pressure, D-1160, is designed to attain fractionation equivalent to about one theoretical plate. A gas chromatographic (GC) method, D-2887, is now available. It is roughly equivalent to D-86 but requires some conversion of the raw data. D-216 is little used now, as the naphtha fractions from current gas plant liquids recovery almost always goes through a stabilization process designed to fit certain D86 cut requirements.

Table 10 shows the standardized distillation procedures for the various types of petroleum products.

Briefly, the distillation procedure consists of heating a sample in a flask fitted with a thermometer, and then causing the vaporized portions to pass through a tube condenser which extends through an ice–water bath and into a glass cylinder graduated in milliliters. Readings show the amount of sample evaporated at a given temperature. A diagram of a basic laboratory apparatus for D86 analysis is shown in Figure 9.

When the first drop falls from the end of the condenser, the reading of the distillation thermometer is recorded as the IBP. The heating of the sample continues at a uniform rate and the volume of distillate collected in the cylinder is observed and recorded to the nearest 0.5 mL, as the thermometer reaches each multiple of 10 °C, or the temperature is recorded when the distillate level reaches each 10-mL mark on the graduate. The EP or FBP is the highest

Table 10 Distillation procedures

Type of product	ASTM specification
Gasoline, naphtha, kerosine	D-86, D-3710
Gas oil, distillate fuels	D-86, D-1160, D-2887
Natural gasoline	D-216, D-3710
Solvents and diluents	D-1078, D-3710
Heavier products (reduced pressure)	D-1160, D-2887

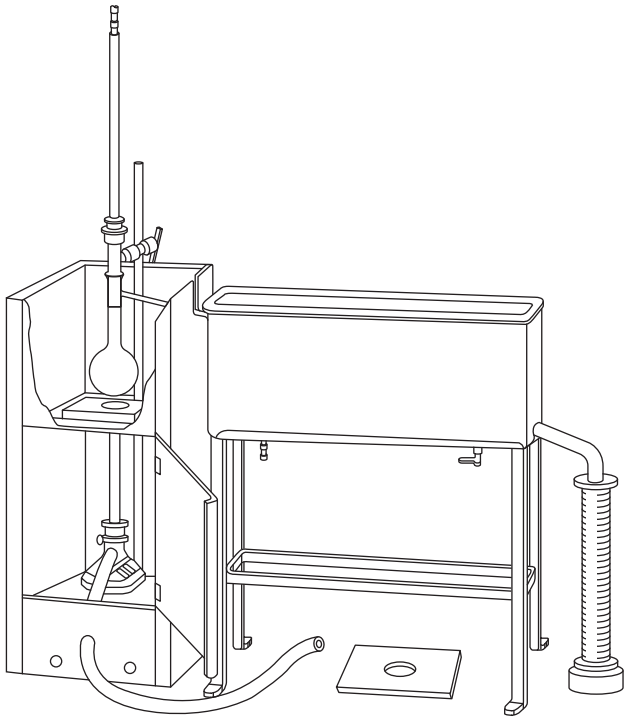

Figure 9 Conventional D86 apparatus.

temperature observed, which is usually reached just as the flask bottom has become dry. Appropriate corrections to the results are made for evaporation loss and barometric pressure. In manual mode, the ASTM distillation procedures require a substantial amount from the laboratory technician. Fortunately, several vendors have made modifications to the apparatus of Figure 9 which automate the heating and volume–temperature accounting process. It should also be noted that the initial and FBPs are somewhat subjective in nature and are, therefore, the least accurate points of the entire assay curve.

All distillation procedures are designed to give results which throw some light on the suitability of the various products for their particular use. All these procedures cannot be described in detail or the significance of their results discussed in this text. However, the quality of motor gasolines as judged from distillation data is presented as being typical of the use of distillation characteristics. Of these procedures, D86 is probably the leading standard. Several methods of measuring volatility via alternative methods, primarily gas chromatography (GC), have come into wide use but they all have the common denominator of requiring calibration standards set by the D86 standard.

6.2.2 Chromatography

As either a separation or as an analytical technique, chromatography[52,66] may be conveniently divided into two categories: GC, and liquid chromatography (LC).

It has been generally found that LC complements GC as a separation tool in the laboratory. LC is ideally suited to samples that are nonvolatile or are sensitive to heat. It generally separates on the basis of compound type. In contrast, GC separates partly on polarity and partly on the basis of volatility, and therefore molecular weight has an influence on the separation. If one employs a special packing in LC, such as a cross-linked polymer with a controlled pore size, separation according to molecular size can be achieved. This is called gel permeation chromatography (GPC). This technique is especially valuable for the determination of molecular weight distribution in polymers and resins.

Here are some additional standards that may be applicable in chromatographic analysis of regulated fuel components.

- ASTM-F1374 – Ionic/organic Extractables of Internal Surfaces-LC/GC/FTIR (Fourier transform infrared) for Gas Distribution System
- ASTM-E1642 – General Techniques of Gas Chromatography Infrared (GC/IR) Analysis, Standard Practice
- ASTM D-3710 – Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography, Standard Practice
- Also ASTM D-3710, D-2427, ASTM F-1374, D3606, and D4815 have pertinent sections that apply to chromatographic methods.

6.2.2.1 Gas Chromatography Since its commercial inception in the 1950s, GC[69] has had astronomical growth in analytical laboratories. The ease of use and versatility of the equipment has allowed gas chromatographs to be used for process as well as laboratory bench instruments.

Basically, GC is a simple technique. A typical GC schematic is shown in Figure 10. A sample is injected onto a column through which a carrier gas is flowing. A typical sample size for a liquid is of the order of a microliter; for a gas, a few milliliters. Liquid samples are vaporized in a heated injection block located ahead of the column. Some solid samples can be handled by using a special solid inlet system. The column is the heart of any chromatographic separation and may take any one of a number of forms. It is either a stainless-steel tube, about 3-mm inner diameter, packed with an inert solid coated uniformly with a liquid of low volatility, or a capillary tube, about 0.3-mm inner diameter, whose interior wall is coated with the liquid. As the sample is carried along by the gas, separation occurs by differential partitioning between the gas and the liquid coating. A large choice of liquid substrates is available, so

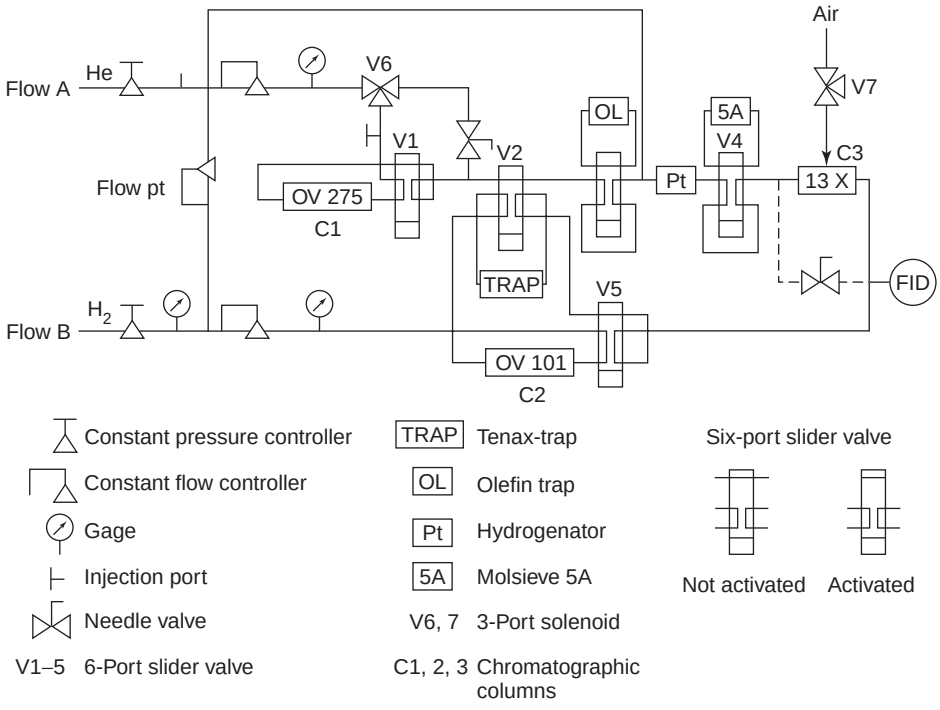

Figure 10 Basic GC schematic.

that the selectivity of the column may be matched to the mixture to be separated. The length of the column may vary from several centimeters for packed columns up to a hundred meters for the capillary. The column is housed in a temperature-controlled oven, the temperature of which can be increased according to a preset program. This had the advantages of either shortening the analysis time or improving the separation, or both.

As sample components are eluted from the column, they are sensed by a detector, the signals from which are recorded. The intensity of the signal is a quantitative representation of the amount of each component. There are a variety of detectors to ensure proper detection of the compounds of interest from the standard hydrocarbon analysis, such as sulfur, nitrogen, halogen and organometallics. Some are general purpose detectors, e.g. thermal conductivity or flame ionization, the latter having essentially equal sensitivity to all hydrocarbons. Others are sensitive to specific elements or compound types. For example, microcoulometers have been used to detect sulfur- and nitrogen-containing compounds in gas oils and other petroleum fractions, while electron capture detectors are highly sensitive to halogens and aromatics. Gas chromatographs have been successfully coupled to the more exotic mass spectrophotometers, Fourier transform infrared (FTIR) and atomic emission spectrum (AE) detectors.

GC is used to characterize a wide variety of hydrocarbon samples. Identification and measurement of the 500 or more components in gasoline are made routinely by GC. For analyses such as these, identification of the components eluting from the column is made by comparing their retention times to the retention times of reference materials, using carefully reproduced chromatographic conditions (column, temperature, flow rate, etc.).

This speciation process is necessarily a slow procedure which usually requires several hours. By contrast, GC distillation data are obtained in much less time, usually one hour or less. If the chromatograph is part of an automatic data handling system, a continuous curve of temperature vs percent overhead can be printed out by a computer-controlled plotter.

Column selection has been expanded to include capillary columns, meaning that the need for knowledge in selecting the right column is important. Capillary columns give the chromatographer the capability successfully to measure compounds that were difficult to measure in the packed column days.

While this new equipment has been developed to enable users to make measurements more accurately and faster at lower detection levels, a whole new set of problems arise with sample and data handling techniques used with new equipment. The trade-off of this new high-tech capability is the need for highly exacting techniques for handling the samples. Since smaller and smaller quantities of material are needed for the analysis, the significance of small levels of contamination becomes magnified. The quantities of material being injected are a

fraction of what they were during the packed column days. Sample sizes were large enough to overcome cross-contamination from a poorly flushed injection valve. Today, one must be concerned about micro levels of cross-contamination, sample cleanliness and sample volumes.

The smaller sample quantities compound the effect of contamination in several ways. From the use of syringes to the use of sampling valves for injection, attention to the sample preparation has become paramount to the analysis, second only to the method/application being used to measure the components of interest. The systems used to introduce the samples have also undergone a radical transition. Many of the new systems are capable of detecting very low quantities of components in a very short time.

Analytical expectations also are different now. In 1980 or even by 1990, refiners were concerned about percent and parts per million levels of detection. Now they are faced with lower levels of detection, parts per million, parts per billion and even parts per trillion levels, with coincident delivery of results in real time.[69]

6.2.2.2 Liquid Chromatography LC has undergone a radical change. The analysis time has been shortened from hours to minutes and the efficiency of the columns has been increased to several thousand theoretical plates per foot. This has been made possible primarily because of the advent of new high-pressure pumps and small-diameter packing materials for columns.

The schematic shown in Figure 10 is equally applicable to LC and GC by replacing the gas elutriant with a liquid one. Between 1–20 μL of a solution of the sample are injected onto a packed column through which a liquid is flowing under pressure. The pressure, which may be up to 40 MPa is needed to achieve adequate flow rates; up to 5 $cm^3\,min^{-1}$. The injection block and column are usually maintained at ambient temperature. The column may be packed with a solid such as silica gel, forming a ''liquid–solid'' chromatograph. This solid may be modified by adding a liquid coating, which may or may not be chemically bonded to the surface, resulting in a ''liquid–liquid'' chromatograph. Separation is achieved by differential partitioning between the moving liquid phase and the stationary solid or liquid phase. The flowing liquid plays an active role in the separation process, and its choice, together with that of the stationary phase, determines the selectivity of the column. To shorten analysis time and to improve resolution, the composition of the moving phase can be changed as the analysis progresses. This is called solvent programming and is analogous to temperature programming in GC.

As the components are eluted from the column, they are sensed by a detector, two types of which are used most commonly. One type, the refractometer, senses a change in the refractive index of the solution and is a general detector, i.e. it will sense any component eluting from the column as long as its refractive index differs from that of the solvent. A second type is a visible range or UV spectrophotometer, and is specific for materials that absorb in a preselected part of the spectrum.

Several LC methods are now used – medium pressure, high-performance liquid chromatography (HPLC), and thin layer methods being the more popular test procedures. They mainly are used on diesel stocks, often to determine aromatics. The calibration techniques to do this are fairly rigorous, and, to date, few standards apply for regulatory purposes.

6.3 Electromagnetic Field Spectroscopy[20,63,67,68,70,71]

6.3.1 General

Electromagnetic field (EMF) spectroscopy is the science of measuring the absorption and emission by atoms and molecules of electromagnetic radiation and of interpreting the results in terms of the composition and structure of such materials. As described in the following sections, spectroscopy can be subdivided into a number of separate fields of endeavor, each related to the various subsections into which the electromagnetic spectrum is commonly divided. The different fields yield different information on the structure and composition of both organic and inorganic materials and between them they are capable of providing a great deal of analytical information of importance to the following and understanding of petroleum and chemical processes. Quantitative spectrophotometric analyses can be accomplished by using Beer's law (Equation 1), where

$$I = \frac{I_0}{I} \quad \text{or} \quad A = abc = \frac{\log I_0}{I} \tag{1}$$

where I is the intensity of the radiation of a given frequency falling upon the sample, I_0 is the intensity transmitted through the thickness b (in cm) of the sample. The zero subscript refers to the transmission of the standard solution at a known calibration value, c is the concentration of the sample (g^{-1}) in the thickness b, and a is the absorptivity or absorbance of the sample, a constant for a material at a given frequency.

The absorbance is more simply related to the concentration and absorptivity than is I and so is therefore commonly employed for quantitative analysis. The ratio I/I_0 is the percentage of radiation transmitted and is commonly known as transmittance, T.

6.3.2 Infrared Spectroscopy

Infrared (IR) absorption spectroscopy,[72,73] covers generally that part of the electromagnetic spectrum between

about 10 000 and 50 Hz (1–200 μm), though in practice the bulk of the work is somewhere between 4000 and 200 Hz (2.5–50 μm) in what is referred to as the near-infrared (NIR). The energy of the radiation in this region of the spectrum corresponds to the energy range characteristic of molecular vibrations, and such radiant energy can be absorbed by resonance interaction. A plot of the amount of energy absorbed vs. the frequency of that energy results in a complex spectrum that is unique to the absorbing molecule. Large catalogs of spectra of pure compounds exist and so it is relatively straightforward to identify an unknown material by comparing its IR spectrum with that of known reference compounds. It is also found that similar chemical structures absorb in a characteristic and relatively small range within the overall spectrum; for example, all carbonyl groups, $>C\cdots O$, absorb in the region 1650–1800 Hz. Therefore, the presence or absence of such functional groups as ketones, alcohols, esters, carbonates, nitriles, and so on, can readily be determined in pure compounds or complex mixtures, both organic and inorganic. Such mixtures may be solids, liquids, or gases.

IR spectroscopy is widely used for the identification of unknown compounds or materials and for the components in many complex mixtures. It can be used to determine the structure of organic and inorganic compounds and to determine both the structure and conformation of polymers. It can also be used quantitatively for the analysis of, for instance, additives in lube oils, copolymer compositions and aromaticity.

6.3.2.1 Fourier Transform Infrared Analysis FTIR technology,[74] one of many flavors of IR spectroscopy, is perhaps the most widely used methodology aside from GC for on-line analysis of hydrocarbons. An FTIR unit typically operates by continuously extracting a sample that is filtered, pressure-regulated, and temperature-controlled in the sample system cabinet. All three of these operations are critical to accuracy. The sampling system must also be responsible for introducing reference samples, wash solvents and sample cell purging.

The analysis is performed by measuring the absorption caused by the compounds as the IR beam passes through the sample cell. The absorption in any one unique frequency is proportional to the concentration of the components in the sample, but with a mixed sample, several different frequency responses may be received additively or may cancel each other via wave physics. The system computer receives the detector signal and simultaneously records an interferogram from the hundreds of spectral patterns it gets. At the end of the test, the computer recalculates the original spectra using fast Fourier transform (FFT) mathematics, a decades old signal processing technology that has become popular with the advent of personal computers and benchtop workstations. The computer then reports the revised component type analysis (from its reference sample database) and component concentrations in standard engineering or scientific units.

Figure 11(a) and (b) show typical GC/FTIR generated spectra for PIANO analyses of catalytic cracker (FCC)

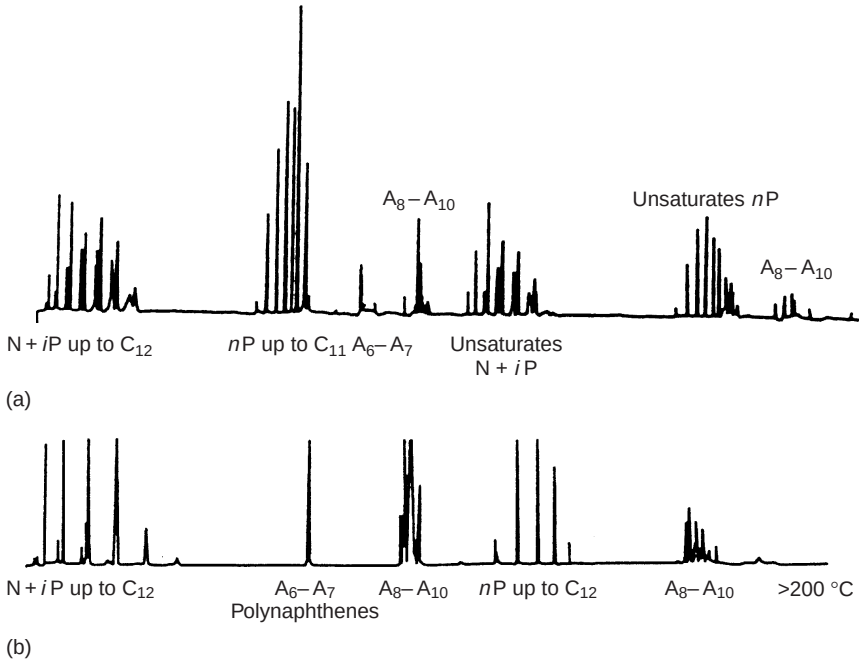

Figure 11 (a) PIANO analysis of catalytic cracker naphtha; (b) PIANO analysis of reformate. P = paraffins, N = naphthenes, A = aromatics, I = isoparaffins, and P = normal paraffins.

Table 11 Data from paraffin, isoparaffin, aromatic, naphthene, olefin (content) (PIANO) analysis of catalytic cracker naphtha (see Figure 11a)

Report in wt%					Report in vol%				
C number	Saturates				C number	Saturates			
	*i*P	*n*P	N	A		*i*P	*n*P	N	A
3	0.00	0.00	0.00	0.00	3	0.00	0.00	0.00	0.00
4	0.60	0.28	0.00	0.00	4	0.77	0.35	0.00	0.00
5	1.89	2.76	0.30	0.00	5	2.19	3.19	0.29	0.00
6	2.48	2.75	1.31	0.09	6	2.73	3.01	1.24	0.07
7	2.19	3.44	2.82	0.54	7	2.31	3.64	2.69	0.45
8	3.75	3.51	4.14	3.57	8	3.82	3.61	3.88	2.98
9	3.99	3.83	4.03	2.88	9	3.96	3.86	3.75	2.39
10	3.20	3.39	3.24	1.52	10	3.11	3.36	2.93	1.27
11	0.49	0.00	0.58	0.00	11	0.47	0.00	0.52	0.00
	18.59	19.96	16.43	8.59		19.35	21.03	15.30	7.16
C number	Unsaturates				C number	Unsaturates			
	*i*P	*n*P	N	Totals		*i*P	*n*P	N	Totals
3	0.00	0.00	0.00	0.00	3	0.00	0.00	0.00	0.00
4	0.50	0.16	0.00	1.55	4	0.63	0.20	0.00	1.96
5	1.92	1.58	0.00	8.45	5	2.23	1.82	0.00	9.72
6	2.81	1.63	0.92	11.98	6	3.09	1.79	0.87	12.80
7	2.35	1.64	2.53	15.51	7	2.47	1.74	2.42	15.72
8	2.78	1.36	3.21	22.32	8	2.84	1.40	3.01	21.55
9	2.79	4.10	1.34	22.96	9	2.77	4.13	1.25	22.12
10	0.31	2.28	0.00	13.94	10	0.30	2.26	0.00	13.22
11	0.00	0.32	0.00	1.39	11	0.00	0.31	0.00	1.30
	13.46	13.06	8.01	98.10		14.33	13.65	7.55	98.38

Table 12 Data from PIANO analysis of reformate (see Figure 11b)

Report in wt%						Report in vol%					
C number	Paraffins					C number	Paraffins				
	n	*i*	N	A	Totals		*n*	*i*	N	A	Totals
3	0.80	0.00	0.00	0.00	0.80	3	1.24	0.00	0.00	0.00	1.24
4	3.23	0.91	0.00	0.00	4.14	4	4.36	1.25	0.00	0.00	5.62
5	3.33	4.58	0.27	0.00	8.18	5	4.17	5.75	0.29	0.00	10.21
6	1.25	8.44	0.32	5.99	15.99	6	1.48	10.06	0.33	5.34	17.21
7	0.31	4.78	0.26	19.39	24.73	7	0.36	5.46	0.27	17.54	23.62
8	0.09	1.47	0.13	24.41	26.11	8	0.10	1.63	0.14	22.07	23.94
9	0.08	0.35	0.07	14.97	15.46	9	0.08	0.38	0.07	13.49	14.02
10	0.00	0.00	0.00	3.22	3.22	10	0.00	0.00	0.00	2.90	2.90
11	0.00	0.00	0.00	0.00	0.00	11	0.00	0.00	0.00	0.00	0.00
	9.09	20.53	1.05	67.98	98.64		11.80	24.53	1.08	61.35	98.76

and reformer naphthas with corresponding data in Tables 11 and 12.

Here are some additional standards that may be applicable for use of IR and UV spectroscopy analysis of regulated fuel components:

- ASTM-E1655 – Infrared, Multivariate, Quantitative Analysis, Standard Practices for
- ASTM-E168 – General Techniques of Infrared Quantitative Analysis, Standard Practices for
- ASTM-E204 – Identification of Material by Infrared Absorption Spectroscopy, Using the ASTM Coded Ba″
- ASTM-E334 – General Techniques of Infrared Micro-analysis, Standard Practices for
- ASTM-E1421 – Describing and Measuring Performance of Fourier Transform Infrared (FTIR) Spectrometers.
- ASTM-F1374 – Ionic/Organic Extractables of Internal Surfaces-GC/FTIR for Gas Distribution System

- ASTM-E1642 – General Techniques of Gas Chromatography Infrared (GC/IR) Analysis, Standard Practice
- ASTM D-4046 – Alkyl Nitrate in Diesel Fuels by Spectrophotometry, Standard Test Method for
- ASTM-E204 – Identification of Material by Infrared Absorption Spectroscopy, Using the ASTM Coded Ba

6.3.3 Ultraviolet/Visible Spectroscopy

UV/visible absorption spectroscopy[71] covers that region of the spectrum between 195 and 1000 nm. Absorption of energy in this region by molecules results in transitions between electronic states and, as with IR spectroscopy, such a spectrum can be used to provide information on the structure and composition of a sample. The most valuable analytical information from a petroleum and chemical processing point of view is obtained on molecules containing highly conjugated systems, particularly aromatic hydrocarbons and conjugated olefins, because such molecules have characteristic spectra and very high absorptivities.

UV/visible spectroscopy is the most powerful method for identifying and measuring trace (ppm–ppb) quantities of PNA hydrocarbons in hydrocarbon systems, particularly in combination with GC and other separation techniques.

6.3.3.1 Fluorescent Indicator Adsorption Fluorescent indicator adsorption (FIA) technology is the most widely used methodology aside from GC for on-line analysis of hydrocarbons with multiple carbon–carbon bonds such as olefins and aromatics. FIA has been an ASTM standard for over 40 years, first appearing as such in 1954, so it is well known and established. It is inexpensive to implement in spite of the fact that it is still a laboratory based manual method. FIA determines the relative amounts of saturates, olefins and aromatics well enough to often substitute for some of the more expensive PIANO testing equipment mentioned in this section. Even so, there has been some controversy about its use for regulatory olefins emissions analysis.

A FIA unit operation is based on the separation of hydrocarbon components according to their adsorption affinities on activated silica gel. A sample is introduced onto the top of a bed of silica gel and packed in a glass column of small, uniform diameter. Isopropyl alcohol is added to desorb the sample and force it down the column. The components with the least affinity for silica gel are the saturates, which move to the front of a traveling band of sample. Those with progressively higher affinities (olefins and aromatics) lag behind.

Fluorescent dyes, visible under UV light mark the zonal boundaries of the saturates, olefins, and aromatics. The amount of each hydrocarbon type, in volume %, is calculated from the length of the zone. Because the fluorescing bands have a measurable length, which can be significant relative to the zone length of a particular hydrocarbon component, differences in ''reading'' the position of the boundary add a source of potential error to the determination.

The presence of significant C_4 and C_5 volatiles in gasoline leads to low-olefin readings compared to other PIANO-type analyses on the same sample composition. A note warning of this possibility appears in the ASTM D-1319 standard which also states that samples containing more than 5% C_4s or 10% C_4s + C_5s should be dependentanized prior to analysis. This extra distillation step does not help the reproducibility of the method.

The author also has had trouble with FIA results for high-olefin stocks such as catalytic gasoline which will routinely contain 25–50% olefins. Any values above 30–35% give FIA readings that are suspect at best.

Here are some additional standards that may be applicable for use of UV spectrophotometry analysis of regulated fuel components:

- ASTM D-1840 – Naphthalene Hydrocarbons in Aviation Turbine Fuels by Ultraviolet Spectrophotometry, Standard Practice.
- ASTM D-1017 – Benzene and Toluene in 250 F and Lighter Petroleum Products by Ultraviolet Spectrophotometry.

6.3.4 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) is a form of radiofrequency spectrometry which uses the weak magnetic properties of atomic nuclei that possess nuclear magnetic moments. The most notable nuclei for petroleum characterization are H and C, the latter of which is only 1% abundant in nature. In the presence of a strong applied magnetic field the nuclear magnetic dipoles have certain preferred orientations, or energy levels, the energy separation of which is directly proportional to the *total* magnetic field sensed by the nuclei. The total magnetic field is the sum of the large applied field and very small molecular magnetic fields which are sensitive to the details of the environment of each nucleus. At a fixed applied magnetic field, chemically distinct nuclei will absorb at slightly different radiofrequencies. These small differences in absorption frequency are referred to as *chemical shifts*. Interaction between nearby magnetically nonequivalent nuclei may result in multiple absorption peaks for the particular nucleus affected. The number of peaks observed and the spacing between them varies in a predictable manner and is referred to as the *spin–spin splitting* pattern.

Both the chemical shift and the spin–spin splitting pattern are useful for determining molecular structures. A third important feature of NMR spectra is that the relative areas of the absorption signals of the various nuclei are proportional to the number of nuclei giving rise to the NMR signals. Thus, quantitative measurements are made by integrating the area under the signals, or peaks, which are due to different groups.

Proton (H) NMR is rapid and nondestructive. It has long been a useful ally of GC and mass spectrometry for characterization of volatile petroleum fractions. Determination of olefin type and the degree of branching in saturated side chains are typical applications. Proton NMR has the additional advantage of being applicable to high-boiling fractions and residues.

Developments in NMR techniques have made C NMR a practical reality. Carbon-13 NMR is difficult because the magnetic isotope (C13) is only 1% of the naturally occurring isotopes of carbon. The great advantage of C13 NMR is its much greater chemical shift range (100 times that of proton NMR). Moreover, C13 NMR probes the carbon skeleton directly, rather than inferring this information via the protons in the molecule. The result is that C13 NMR is many times more powerful than H NMR for characterization of complex hydrocarbons. Carbon-13 NMR provides a direct determination of aromaticity (% aromatic carbon), degree of aromatic substitution, and has the potential to provide a detailed analysis of side chain structure (average chain length, branching, etc.). Exploration of the full potential of C13 NMR for petroleum characterization is an active area of research.

Some ASTM standards for specific component analysis are:

- D-2622, D-4294 for sulfur in petroleum stocks,
- D-4059 for lead in petroleum stocks,
- D-3701, D-4808 for hydrogen content in petroleum stocks.

6.3.5 Emission Spectroscopy[70]

When the atoms of a material are subjected to an electric arc, spark, or hot flame, electrons in the atoms are excited to energy levels higher than the ground state. When these electrons lose their excitation energy, light is emitted in the UV and visible region of the spectra. The electronic transitions involved are discrete so that a set of spectral lines is produced which is characteristic of the emitting atom. Furthermore, the intensity of the spectral lines is proportional to the number of emitting atoms. Thus, the emission spectrum can be used both to identify which elements are present in a material and to measure the concentration of the various elements. Quantitative work does, however, require that suitable standards be available for comparison.

The essential components of an emission spectrograph include (1) a radiation source, (2) an optical analyzer, and (3) a detector.

The radiation source includes the sample to be analyzed and a means of excitation such as the electric arc or spark. With these forms of excitation an electrical current is passed between two electrodes either as a continuous arc or as a spark. If the sample is a metal, the sample itself can serve as one or both electrodes. For other types of sample carbon electrodes are generally used to contain the sample during excitation. The actual design of the electrode will depend on the form of the sample and the elements of primary concern. Both liquid and solid samples can be analyzed.

The optical analyzer serves two functions in the spectrographic system. It provides a sharp image of the radiation emitted from the excited source, and it separates the radiation into its component wavelengths. The focusing required to produce a sharp image of the source is accomplished with lens or focusing mirrors. The spatial separation of the light as a function of wavelength is accomplished with a diffraction grating; a flat piece of glass, metal or plastic on which are ruled a large number of parallel grooves (6000–10 000 grooves per centimeter).

The detection of the dispersed radiation is generally accomplished with a photographic plate or a photoelectric detector. Photographic plates are generally used for qualitative work because a wide range of wavelengths can be recorded simultaneously. The analysis of an unknown spectrum is accomplished by comparing it to reference spectra of known elements. Quantitative analysis is generally performed on instruments which utilize photoelectric detection. In practice, instruments designed for quantitative analysis often have a number of photoelectric detectors (up to 48) which can be set for different wavelengths. It is thus possible to perform simultaneous multielement determinations with this type of equipment. In principle, every element in the periodic table can be detected by emission spectroscopy, but it is routinely useful with approximately 70 elements. Qualitative analysis data can be obtained on as little as 0.1 mg of a solid sample, and under favorable conditions concentrations down to a few parts per billion can be determined. The technique is used extensively in such problems as the analysis of wear metals in lube oils, qualitative analysis of corrosion products, and the determination of catalyst impurities.

6.3.5.1 Flame Ionization Detection This is probably the most common emissions spectroscopic method used for hydrocarbon analysis. Most commercial implementations have excellent sensitivity and stability at parts per

million levels of detection, even in harsh process conditions. Since sample flow, flame and temperatures are constantly monitored, maintenance and error detection are greatly simplified. FIDs (flame ionization detections) are easily automated and pair well with elution-type analyzers such as a GC.

The analyzer operates by burning the sample (vaporized) in a hydrogen-based flame. The hydrocarbons present are ionized in the flame and have the ability to conduct electricity through the flame to a collection grid surrounding the flame. By applying a high-voltage charge to the flame tip, the hydrocarbons in the sample can be measured by monitoring the current reaching the collection grid. This technique is not only sensitive to general hydrocarbons but those with easily oxidized heteroatoms as well – sulfur, chlorine, and so on.

Here are some more standards that may be applicable in emission spectroscopic analysis of regulated fuel components:

- ASTM D-3605 – Trace Metals in Gas Turbine Fuels by Atomic Absorption and Flame Emission Spectroscopy
- ASTM-E115 – Photographic Processing in Optical Emission Spectrographic Analysis, Standard Practice.

6.3.6 Atomic Absorption Spectroscopy

Atomic absorption spectroscopy, a technique related to emission spectroscopy, is used to measure the concentration of a wide range of metallic elements. In this technique, a sample is vaporized in a flame or on a carbon rod and the vapor is irradiated by a lamp which has a cathode made from the element being sought. Electrons in the ground state atoms produced by the flame absorb the resonance lines emitted by the lamp and are thus excited to a higher energy level. The amount of energy absorbed, measured using a photodetector, is related to the concentration of the element present in the sample. The usual operating range of the method is from less than 1 ppm to about 100 ppm.

Some particular applications to the industry include the determination of metals in crude and refined products, in the additives in lube oils, and in aqueous systems such as effluent waters.

Here are some more standards that may be applicable in atomic absorption spectroscopic analysis of regulated fuel components:

- ASTM D-3605 – Trace Metals in Gas Turbine Fuels by Atomic Absorption & Flame Emission Spectroscopy,
- ASTM-E885 – Analyses of Metals in Refuse-derived Fuel by Atomic Absorption Spectroscopy, Standard T
- ASTM-E924 – Quality Assurance of Laboratories Using Molecular Spectroscopy, Standard Guideline for
- ASTM-E863 – Describing Flame Atomic Absorption Spectroscopy Equipment, Standard Practice for
- ASTM D-3237 – Lead in Gasoline by Atomic Absorption Spectroscopy, Standard Test Method for
- ASTM D-4053 – Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy, Standard Test Method for
- ASTM D-3831 – Manganese in Gasoline by Atomic Absorption Spectrometry, Standard Test Method for
- ASTM-E131 – Molecular Spectroscopy, Standard Terminology Relating to

6.3.7 X-ray Analysis Techniques

X-rays are produced when fast-moving electrons, or other X-rays, interact with the atoms of a target material. The primary mechanism of the interaction involves displacement of an inner electron from an atom, followed by the ''falling'' of an outer electron into the resultant hole. An X-ray photon, whose energy is the difference between the energies of the inner and outer electrons, is produced. Electron energies differ from element to element, and so the emitted X-ray energy is characteristic of the atom from which is was generated. Since atoms contain many different electrons, a number of characteristic X-ray energies are associated with each element.

In addition to the characteristic X-ray spectrum of a material, a continuous spectrum is also produced. This is due to a loss of energy when an electron interacts with the strong electric field of an atomic nucleus. This energy loss appears as an X-ray photon. The energies of the X-rays produced by this mechanism are independent of the nature of the atoms being bombarded and appear as a band of X-rays of continuously varying energy.

X-rays are characterized by either their wavelength or their energy. Equation (2) relates the two quantities

$$\text{change in energy} = \frac{hc}{\text{wavelength}} \tag{2}$$

where h is Planck's constant and c is the speed of light.

The X-ray region of the electromagnetic spectrum covers a range from 0.02 Å to slightly over 100 Å (1 Å $= 10^{-10}$ m). The use of X-rays with wavelengths larger than about 2.5 Å, however, requires special vacuum techniques to minimize photon losses caused by air absorption.

There are three different X-ray techniques used for analysis: diffraction, fluorescence (emission) and absorption. X-ray diffraction is used most in analysis of solid crystalline structures and will not be discussed further in this article.

6.3.7.1 X-ray Fluorescence The characteristic radiation produced when a material is bombarded with X-rays or electrons can be used to identify which elements are present in the material and to measure the concentrations of the various elements. The identification of unknown elements in a sample is based on comparing the wavelengths or energies of the observed X-rays to the known values for the elements in the periodic table. The concentrations of the elements are determined by measuring the intensities of the characteristic X-rays and comparing these to the measured X-ray intensities in suitable standards. In practice, the technique is generally limited to elements with a greater atomic number than 11 (sodium). It is often called chemiluminescence.

X-ray fluorescence instrumentation may be classified as either wavelength dispersive or energy dispersive. Wavelength dispersive systems employ a single crystal which separates the various characteristic X-rays by diffraction according to Bragg's law. The resulting diffracted X-rays are detected by a scintillation counter or flow proportional counter. Because X-rays with different wavelengths are spatially separated, measurements in a multielement system are conducted sequentially. Energy dispersive systems utilize a solid-state detector which is capable of both counting the number of incident X-rays and sorting them according to energy. Since information on all elements in a sample is received simultaneously, a device like a computer is needed to store the information. In general, multielement analyses are conducted more rapidly with an energy dispersive system. Furthermore, for most energies,the resolution of a wavelength dispersive system is superior to that of an energy dispersive system.

Elemental analysis on both liquid and solid samples can be carried out by X-ray fluorescence. The optimum concentration range for analysis is generally down to about $1\,\mathrm{mg\,g^{-1}}$. In many cases, however, concentrations as low as a few micrograms per gram can be determined without great difficulty. The major experimental requirement for elemental analysis by X-ray fluorescence is that standards which match the composition of the samples of interest be available for instrument calibration. A variety of sample preparation and measurement techniques have been developed to meet this requirement. The most recent developments in this area include the use of various mathematical techniques to correct measured intensities for changes in the sample matrix. The successful application of these techniques greatly minimizes the amount of sample preparation required.

Some of the major applications of X-ray fluorescence include the determination of metals in catalysts, the determination of trace metals in cracking feedstocks, the determination of lead and sulfur in petroleum fuels, and the determination of additive components in lubricants.

A derivative method involves the bombardment of material with neutrons and measurement of the X-ray spectra emitted. This method is used in some on-line flow measurements.

6.3.7.2 X-ray Absorption The third X-ray technique used for analysis is X-ray absorption. All materials absorb X-rays, some more than others. The relationship which governs the absorption process is the Beer–Lambert law, Equation (3)

$$I = I_o e^{upt} \tag{3}$$

where t is the thickness of the absorber, p is the density of the absorber, u is its mass absorption coefficient and I and I_o are the transmitted and incident X-ray intensities. u is a function of both the elemental composition of the absorber and X-ray wavelength or wavelengths used. At fixed wavelength, the mass absorption coefficient can be expressed in terms of the elemental composition of the absorber in the following manner, Equation (4)

$$u = \sum u_i W_i \tag{4}$$

where u_i is the mass absorption coefficient of the ith element in the sample and W_i is its corresponding weight fraction.

From this it is clear that a measurement of the transmittance (I/I_o) is not sufficient to determine the concentration of all I elements in a sample. Good results can be obtained, however, in specific applications where only one component of a sample is varying, where all other components are comparatively weak absorbers, and where any change in density of the sample is due to the variation in the one component. In such cases the transmittance of the sample at fixed thickness can be calibrated in terms of the weight concentration of the varying component. This technique is well suited, for example, to the continuous monitoring of TEL metals or sulfur in gasoline or in a hydrocarbon matrix.

Here are some additional standards that may be applicable in the use of X-ray analysis of regulated fuel components:

- ASTM D-2599 – Lead in Gasoline by X-ray Spectrometry, Standard Test Method for
- ASTM D-3229 – Low Levels of Lead in Gasoline by Wavelength Dispersive X-ray Spectrometry, Standard Test Method

- ASTM D-2622 – Sulfur in Petroleum Products by X-ray Spectrometry, Standard Test Method for
- ASTM D-4294 – Sulfur in Petroleum Products by Energy-dispersive X-ray Fluorescence Spectroscopy, Standard Practice for
- ASTM-E1015 – Reporting Spectra in X-ray Photoelectron Spectroscopy, Standard Practice for
- ASTM-E1508 – Quantitative Analysis by Energy-dispersive Spectroscopy, Standard Guide for
- ASTM D-5059 – Lead in Gasoline by X-ray Spectroscopy, Standard Test Methods for

6.4 Mass Spectrometry

The mass spectrometer is a most useful analytical tool in petroleum laboratories. This is due to the availability of good commercial instruments and to the capability of mass spectrometry to analyze complex mixtures rapidly and accurately.

The basic principles of mass spectrometry are simple. A sample is allowed to evaporate into a reservoir at about 47×10^{-4} kPa. It then enters the mass tube through a small leak or orifice, the pressure in the mass tube being maintained at between 1.3×10^{-47} kPa (no sample) and 1.3×10^{-46} kPa (with sample). A small proportion (about 0.2%) of the atoms are ionized by electrons emitted from a hot tungsten filament. This ''electron gun'' creates 70-V energy electrons. The ions ''crack'' into charged fragments and these fragments are accelerated through the mass tube in a magnetic field. The ion paths are altered by the magnetic field in direct relation to their mass, the path of a low mass ion being bent more than that of a high mass ion. By varying either the magnetic field or the accelerating voltage in a continuous manner, one mass at a time passes through the exit slit of the mass tube to the collector plate. After amplification, a mass spectrum of mass numbers vs currents is recorded. The fragmentation pattern of a compound serves to identify the components present in the sample.

Mass spectrometry is applied to the analysis of gases, liquids, and solids. This technique requires about a microliter of liquid sample and a few milliliters of gas, at atmospheric pressure, for analysis. Quantitative determination of complex mixtures can be achieved by proper selection of analytical points (mass/charge) for the compounds present and setting up and solving the sets of simultaneous equations obtained.

One application of mass spectrometry in the petroleum field is for the analysis of gaseous mixtures. Mixtures containing 20 or more components can readily be analyzed. Examples include the products from catalytic crackers and hydroformers, which are analyzed for C_1 through C_5 hydrocarbons, fixed gases, and hydrogen. Although a complete isomer breakdown is not made, sufficient information is obtained to satisfy most requirements. This has been a particularly useful measurement in the case of hydroformate make-gases where the sample may contain as much as 95% hydrogen. It is possible to determine some 19 other components in the remaining 5% of the sample.

Mass spectrometry is also used for the analysis of petroleum liquids. Light (C_5 – 121 °C) naphthas can be analyzed for individual paraffins, naphthenes, and aromatics. Samples boiling in the gasoline range can be analyzed for total paraffins, total noncondensed naphthenes, condensed naphthenes (2, 3, 4, etc. rings) and for C_6, C_7, C_8, C_9, C_{10} benzenes and naphthalenes. Heating oils and lube oils can be analyzed for both aromatics and saturates. The aromatics determined are 2, 3, 4, etc. ring-type as well as the same ring-type thiophenes. The saturates determined are the total paraffins, noncondensed naphthenes, and condensed naphthenes (2, 3, 4, etc. rings). Many of these samples, of course, may be solid at room temperature. Other solid samples analyzable by mass spectrometry include refined waxes, which are analyzed for normal paraffins, isoparaffins and naphthenes.

Mass spectrometry is very useful in qualitative analysis. Instrumentation for this includes high-resolution spectrometers and combination units consisting of a gas chromatograph linked to a mass spectrometer.

Mass spectrometers are commercially available with resolutions of one part in 50 000 and greater. With these instruments it is possible to examine high-boiling fractions for specific compound types, and in conjunction with low-voltage mass spectrometry, they have been used very successfully for the identification of oxygen-, sulfur- and nitrogen-containing compound. High-resolution mass spectrometry is also extremely useful for obtaining structural information on relatively pure compounds.

The direct coupling of a gas chromatograph to a mass spectroscopy (GC/MS) has resulted in an extremely powerful tool for the characterization of complex mixtures. Individual peaks in a chromatogram can usually be identified by their mass spectrum. GC/MS has been used in the identification of contaminants in feedstreams at the parts per million level, for the complete characterization of naphthas which can have over 250 peaks in their gas chromatograms, and for determining isomer distribution in products from experimental isomerization or reforming reactors.

A technique called nitric oxide ionization spectrometry evaluation (NOISE, though it really is more of a modification to the standard GC/MS methods) has gained some popularity in the 1990s for petroleum distillates in the range of 250–1000 °F (120–540 °C). In the NOISE method, bp fractions are bombarded with NO^+ ions. A mass spectrometer is used to identify the individual components and reports them by carbon number as well

as by hydrogen deficiency. Knowing these two quantities for all detected components mathematically allows for a PIANO and sulfur analysis of the sample at hand. A drawback of this method is its relatively poor accuracy when applied to C_6 and C_7 hydrocarbons, so that this lighter naphtha has to be removed (easily done by several methods but GC is usually the most readily made automatic) first. It is also helpful to break down a full range cut into several subcuts for separate sequential analysis to improve the accuracy.

Here are some additional standards that may be applicable in the use of mass spectrometry analysis of regulated fuel components:

- ASTM D-1658 – Carbon Number Distribution of Aromatic Compounds in Naphthas by Mass Spectrometry, Standard Practice
- ASTM D-2498 – Isomer Distribution of Straight-chain Detergent Alkylate By Mass Spectrometry, Standard Practice
- ASTM D-2567 – Molecular Distribution Analysis for Monoalkylbenzenes by Mass Spectrometry, Standard Method
- ASTM D-2789 – Hydrocarbon Types in Low Olefinic Gasoline By Mass Spectrometry, Standard Test Method
- ASTM-E899 – Records Management in Mass Spectrometry Laboratories Performing Analysis in Support of
- US EPA-Method-8240B – Volatile Organics by Gas Chromatography/Mass Spectrometry–packed Column Techniques
- US EPA-Method-8270B – Semivolatile Organic Compounds by Gas Chromatography/Mass Spectrometry–capillary
- Patent-5461235 – Mass Spectrometry Apparatus and Method Relating Theretofore
- Patent-5481108 – Method for Ion Detection and Mass Spectrometry and Apparatus Thereof
- ASTM D-2425 – Hydrocarbon Types in Middle Distillates by Mass Spectrometry, Standard Test Method for
- ASTM D-2650 – Chemical Composition of Gases by Mass Spectrometry, Standard Test Method for

6.5 Volatility Measurements

Volatility is a primary factor in liquid fuel quality. The vaporizing tendencies of all types of fuels are important to their combustion characteristics and general utilization. Volatility is also of considerable significance to the quality of other liquid petroleum products, for example, solvents and lubricants.

For some purposes it is necessary to have information on the initial stage of vaporization. To supply this information, flash fire, vapor pressure, and evaporation tests are available. The data from the early stages of the several distillation tests are useful also. In some applications it is important to know the tendency of a product to vaporize partially or completely at ambient temperatures.

6.5.1 Flash and Fire

The *flash point* of a petroleum product is the temperature to which the product must be heated under the prescribed conditions of the test to give sufficient vapor to form a mixture with air that can be ignited by a suitable flame. The *fire point* is the temperature at which the oil's vapor ignites and continues to burn for 5 s. If the flash point is known there is additional value in knowing its fire point. The importance of flash point data has led to the development of standardized procedures, including ASTM D-56, D-92, and D-93.

Petroleum products vary widely in flash and fire points. Gasolines flash well below freezing. Solvents, have flash points ranging from below 21 °C (70 °F) up to 52 °C (125 °F). Kerosines flash at 38 °C (100 °F) to 66 °C (150 °F), while heavier fuel oils flash from 66 °C to 121 °C (250 °F). Lubricating oils have flash points from 275 to 700 °F (135–370 °C). Fire points range from 5.5 °C (10 °F) to 22 °C (70 °F) higher than the flash points, with the greater spread appearing in the heavier products. When the flash point is given, the method used in determining it should be mentioned.

The Cleveland open-cup (ASTM D-92) method, the one almost universally used, employs a brass cup supported by a metal ring, see Figure 12. A flame or electric heater under the cup heats the oil sample, in which the bulb of a thermometer is immersed. A tiny pilot flame is moved over the surface of the heated oil at every 5 °F (2.8 °C) rise in temperature. When the vapor above the oil flashes, the temperature reading indicates the flash point of the oil. Theoretically, it is possible to test the whole range of petroleum products with a single flash test. In actual practice it has been better to use three instruments because flash points of petroleum products vary over a wide range. The tag closed tester (ASTM D-56) is used to test all mobile liquids that flash below 79 °C, but not for products classed as fuel oil. The Pensky–Martens closed tester (ASTM D-93) is used to test diesel and heating fuel oil unless other methods are specified. The open cup tester (ASTM D-92) is used for determining both flash and fire point for all petroleum products except fuel oil and those which have a flash point below 79 °C (175 °F).

Flash points are most generally used to determine fire and explosion hazards. For products with flash points below 33 °C (90 °F), special precautions are necessary for

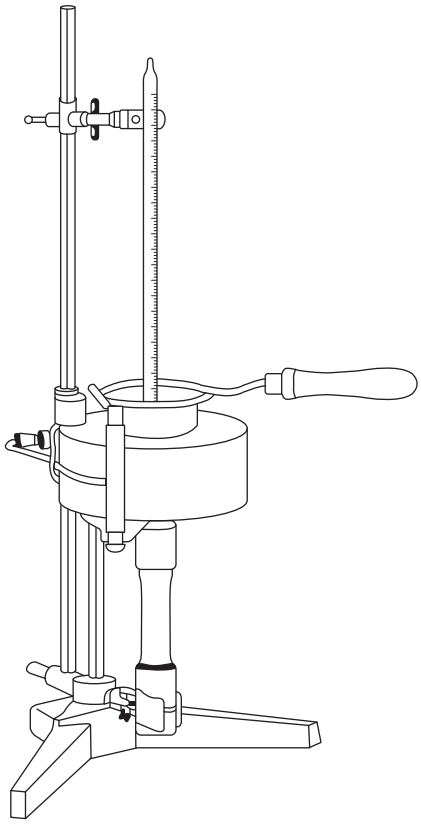

Figure 12 Cleveland open cup tester.

safe handling. Products having flash points in the 33 °C to 66 °C range may also present problems with respect to safe handling and storage. Products with flash points above 66 °C are relatively safe to handle and, in this case, flash points are generally determined as indirect measures of some other quality. For example, flash point is often used by manufacturers and marketers of petroleum products to detect contamination or as an aid in identifying a particular petroleum product.

6.5.2 Reid Vapor Pressure

6.5.2.1 Bomb Methods The RVP[31,55–58,75] apparatus consists essentially of a double-chambered bomb. The upper chamber is fitted with a pressure gage and contains atmospheric air. The lower chamber which has 0.25 the capacity of the upper chamber is filled with a sample of the liquid to be tested. Both the sample and its chamber are first chilled to reduce premature evaporation. The bomb is then sealed, care being take to prevent the introduction of abnormal pressures.

The sealed bomb is immersed in a 38 °C (100 °F) bath. To assure a maximum opportunity for evaporation, the bomb is removed from the path periodically for brief but vigorous shaking. During this process, a small portion of each fraction is vaporized, the amount depending upon its vapor pressure at test temperature and its concentration in the sample. When an equilibrium is reached, that is, when two consecutive gage readings are identical, that gage reading is recorded as the ''uncorrected vapor pressure''.

The bomb pressure consists of the atmospheric pressure plus the sum of the individual vapor pressures of the fractions, each modified according to the ratio of its liquid volume to the total liquid volume. Obviously, this pressure is greater than atmospheric but since the gage registers only pressures above atmospheric, its reading is related only to vapor pressure.

The gage is removed from the bomb, calibrated against a mercury manometer at the value of the uncorrected vapor pressure and any discrepancy noted. The uncorrected value is then corrected for both the gage calibration and the change in the pressure of the water vapor and air in the air chamber between the initial air temperature and 38 °C.

The discussion above is the traditional D-323 method of analysis. Since the water vapor pressure in the vapor over oxygenated fuels is affected by the presence of that oxygenate, D-4953 has its own set of corrections to be made to the ''uncorrected cell vapor pressure''.

6.5.2.2 Minicell Methods Since the bomb methods are slow and time consuming, several alternate methods and equipment have been developed to determine vapor pressure of gasoline samples. These were developed with ease of in-field certification in mind.

Portable Cell Test Apparatus. The type of apparatus suitable for use in this test method employs a small volume test chamber incorporating a transducer for pressure measurements and associated equipment for thermostatically controlling the chamber temperature and for evacuating the test chamber prior to sample introduction. It is often generically called a Grabner cell apparatus after one of the first and most prominent vendors of such portable test equipment. Several other vendors have good equipment on the market as well, but the naming sticks, much as Xerox and Coke, describe a wide class of photostatic copiers or cola drinks.

The test chamber contains between 5 and 50-mL of liquid and is capable of containing a vapor-to-liquid ratio of 1.0 up to 4.0 (minimum by standard). A known volume of chilled air-saturated sample is introduced into the evacuated thermostatically controlled test chamber. The internal volume of the test chamber is five times that of the total test specimen introduced into the chamber. After injection into the test chamber, the test specimen is allowed to reach thermal equilibrium at the test temperature, 37.8 °C (100 °F). The resulting rise in pressure in the chamber is measured using a pressure transducer sensor and indicator. Only total pressure

measurements (sum of the partial pressure of the sample and the partial pressure of the dissolved air) are used in this test method, although some instruments can measure the absolute pressure of the sample as well.

The measured total vapor pressure is converted to a dry vapor pressure equivalent (DVPE) by use of a correlation equation. A correction to this value is necessary to account for the biases between the cell pressure and that which would be measured by standard D-4953 laboratory testing. The 1991 ASTM D-2.02 Round Robin test report mentions several correcting equations.

Automatic Cell Test Method. The chilled sample cup of the automatic vapor pressure instrument is filled with chilled sample and is coupled to the instrument inlet fitting. The sample is then automatically forced from the sample chamber to the expansion chamber where it is held until thermal equilibrium at 37.8 °C (100 °F) is reached. In this process the sample is expanded to five times its volume (4 : 1 vapor-to-liquid ratio). The vapor pressure is measured by a pressure transducer.

The measured vapor pressure is automatically converted to a DVPE value by the instrument. A correction to this value is necessary to account for the observed bias between the test result and that obtained by Test Method D 4953.

The instrument automatically calculates the DVPE from the measured total pressure by means of the following correction equation to correct for the relative bias, found in the 1991 ASTM D-202 Round Robin test program, between the results obtained by this test method and the dry vapor pressure values measured in accordance with Test Method D 4953 as follows, Equation (5):

$$\text{DVPE (kPa)} = (0.941\,X) + 3.22 \quad \text{or}$$
$$\text{DVPE (psi)} = (0.941\,X) + 0.467 \qquad (5)$$

where X is the measured total pressure in kilopascals (pounds-force per square inch).

7 APPLICABLE REGULATORY MEASUREMENT STANDARDS BY CRITERIA COMPONENT[41,47]

7.1 Introduction

This section presents a more detailed listing and discussion of several of the more commonly recognized standards for analysis of fuels constituents for environmental regulatory reasons. It covers the major standards for North America and the EU, as these areas are furthest advanced in environmental concerns. US EPA and EU regulatory standards are the foundation for environmental regulations in much of the rest of the world, even if enforcement in some of these areas lags behind the developed world.

The standards mentioned in this section are not repeated verbatim. The reader can easily obtain these directly and inexpensively, as needed, from either ASTM or ANSI/ISO. In fact, since these standards are routinely updated by several diligent industrial and government committee representatives, any detailed coverage discussed here is likely to be obsolete within a few years.

The content of these standards varies greatly. In some cases, procedures for using specific analytical equipment are provided in great detail, along with illustrations and lots of minutiae. In other cases, they only detail the allowable limits and provide an official ''accuracy'' statement of a given type of analytical procedure for a specific type of measurement on a certain type of fuel product stream. Statistical reliability metrics, calibration standards, and sample preparation methods are three of the most common areas of coverage that these methods provide.

Analytical instrument vendors typically provide ''off the shelf'' systems that are calibrated and verified to one or more of several appropriate standards and analytical environments. ISO 9000 and ISO 14000 requirements for commercial and environmental compliance equipment are fairly strict in removing human error as much as possible from the sampling and measurement process. An IR spectroscopy-based standard may provide a one way of interpreting a spectral chart for a certain composition of sample material and another for a different type of sample. Modern analyzers not only take the spectra, but also calculate the necessary information from digital analysis of the spectral pattern. (The days of doing graphical integration of peak charts by hand – literally counting chart crosshatching scale squares – are over.) To change from one standard methodology to another is often now a matter of choosing a different methodology on a Microsoft Windows™-based computer interface dialog box and hitting the return key on a keyboard. ISO 9000 manufacturing compliance also sets some fairly strong legal requirements for calibration of a piece of test equipment to prevailing standards.

The typical standard values stated in the SI (Système International d'Unitès) units are to be regarded as the standard conversion. Conversion to and from these units to English-language engineering values may or may not be covered. A few of the standards even specify the precision with which such conversions should be reported! In most cases, normal prevailing laboratory practices for accuracy and precision of reported numbers are more than adequate to maintain adherence to the standard. If in doubt, there are even standards for laboratory practice standards that can be consulted.

None of the standards address any safety issues, if any, associated with its use. It is the responsibility of the user of the standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.

7.2 Choosing the Appropriate Standards

With the availability of several perfectly valid alternatives for obtaining chemical analyses, how should one determine which one(s) to use. This may be as simple an answer as referring to the local regulatory agencies' codified requirements. Usually, they present alternatives as well. In some cases, multiple regulatory governing bodies have conflicting specifications.

Table 13 lists the adopted and equivalent test methods for the CARB components whose designated test methods have recently been amended. It should be noted that certain conditions must be met when using the equivalent test methods for aromatic hydrocarbons (aromatics) and low level (1–10 ppm) sulfur.

In fact, each refiner is free to control and do analysis by whatever method they feel comfortable with. Whatever is measured and however the numbers are obtained, they must provide answers compatible with whatever the local governing agencies (all of them) have deemed as the standard of choice.

7.3 Distillation Methodology

7.3.1 D86 Distillation of Petroleum Products

This test method covers the distillation of natural gasolines, motor gasolines, aviation gasolines, aviation turbine fuels, special boiling point spirits, naphthas, white spirit, kerosines, gas oils, distillate fuel oils, and similar petroleum products, utilizing either manual or automated equipment. It is the most widely used distillation standard for naphtha-containing fuels. There is some question about its accuracy with diesel range petroleum fractions but most distillations of this material are reported as D86 or its equivalent from a related test.

7.3.2 D3710 Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography

This test method covers the determination of the boiling range distribution of gasoline and gasoline components. This test method is applicable to petroleum products and fractions with a FBP of 500 °F (260 °C) or lower as measured by this test method. D3710 is designed to measure the entire boiling range of gasoline and gasoline components with either high or low RVP and is commonly referred to as GC distillation. It is essentially a GC-equivalent procedure to obtain D86 assay information. D-3710 has not been validated for gasolines containing oxygenated compounds (for example, alcohols or ethers).

7.3.3 D1160 Distillation of Petroleum Products at Reduced Pressure

This test method covers the determination, at reduced pressures, of the range of boiling points for petroleum products that can be partially or completely vaporized at a maximum liquid temperature of 400 °C. Both a manual method and an automatic method are specified. It is not really suitable for gasoline or kerosine but may be for heavier diesel stocks. It is also not amenable to on-line

Table 13 Criteria fuel component measurement standards

Component	CARB adopted method	CARB equivalent method	US EPA method	ACEA standards
Oxygen	ASTM D-4815	Prop. GC/FTIR[a]		
Olefins	ASTM D-1319a	ASTM D-1159	ASTM D-1319a	ISO 3837
Benzene	ASTM D-5580	Prop. GC/FTIR[a]	ASTM D-3606	EN-238
Aromatics (gasoline)	ASTM D-5580-95	Prop. GC/FTIR w/corr for ASTM D-5580	ASTM D-1319	
RVP	ASTM D-4953-	ASTM D-5191 w/corr. for ASTM D-4953	ASTM D-4953	ISO 3007-1996
T50, T90, FBP[b]	ASTM D-86-95	ASTM D-86	ASTM D-86	ISO 3405-1988
Sulfur				
1–10 ppm	ASTM D-5453	ASTM D-4045 (modified)		
10 ppm+	ASTM D-2622 (modified)	ASTM D-5453		
Phosphorus	ASTM D-3231	ASTM D-3231	ASTM D-3231	
Lead	ASTM D-4059			EN 237

[a] "Proposed ASTM Standard Test Method for Determination of Oxygenates, Benzene, Toluene, C_8–C_{12} Aromatics and Total Aromatics in Finished Gasoline by GC/FTIR."
ACEA, Association des Constructeurs Européens de l'Automobile.
[b] FBP.

analysis as it is clumsy to perform, though it is simpler than TBP (true boiling point). D1160 values can be reasonably converted to TBP and thence to equivalent D86 values via the *API Technical Databook*.[75]

7.3.4 D2887-e1 Boiling Range Distribution of Petroleum Fractions by Gas Chromatography

This test method covers the determination of the boiling range distribution of petroleum products. D2887 is applicable to petroleum products and fractions having a FBP of 538 °C (1000 °F) or lower at atmospheric pressure as measured by this test method. It is limited to samples having a boiling range greater than 55 °C (176 °F), and having a vapor pressure sufficiently low to permit sampling at ambient temperature. It is not to be used for the analysis of gasoline samples or gasoline components, which should be analyzed by Test Method D3710.

7.3.5 D2892 Distillation of Crude Petroleum (15-Theoretical Plate Column)

D2892 describes the procedure for the distillation of stabilized crude petroleum to a final cut temperature of 400 °C AET (atmospheric equivalent temperature). The test method employs a fractionating column having an efficiency of 14 to 18 theoretical plates operated at a reflux ratio of 5 : 1. Performance criteria for the necessary equipment is specified. Some typical examples of acceptable apparatus are presented in schematic form. This test method offers a compromise between efficiency and time in order to facilitate the comparison of distillation data between laboratories. It is commonly referred to as ''standard'' TBP distillation. Although only indirectly related to most gasoline volatility standards, the standard contains a wealth of information. TBP values have the property of being blendable in linear fashion without a severe loss of accuracy. D86 and other ASTM assay methods are not so readily predictable for multiple blended cuts TBP values can be reasonably converted to and from D86, D1160, and D2887 values via the API Technical Databook Method 3A1.1.

The test method gives details of procedures for the production of a liquefied gas, distillate fractions and residuum of standardized quality on which analytical data can be obtained, and the determination of yields of the above fractions by both mass and volume. From the above information a graph of temperature versus mass-percent distilled can be produced. This distillation curve corresponds to a laboratory technique which is defined at 15/5 (15 theoretical plate column, 5 : 1 reflux ratio) or TBP.

This test method can also be applied to any petroleum mixture except LPGs, very light naphthas, and fractions having IBPs above 400 °C.

D2892 contains the following Annexes:

- Annex A1 – Test Method for the Determination of the Efficiency of a Distillation Column.
- Annex A2 – Test Method for the Determination of the Dynamic Holdup of a Distillation Column
- Annex A3 – Test Method for the Determination of the Heat Loss in a Distillation Column (Static Conditions),
- Annex A4 – Test Method for the Verification of Temperature Sensor Location
- Annex A5 – Test Method for Determination of the Temperature Response Time
- Annex A6 – Practice for the Calibration of Sensors
- Annex A7 – Test Method for the Verification of Reflux Dividing Valves
- Annex A8 – Test Method for Dehydration of a Sample of Wet Crude Oil
- Annex A9 – Conversion of Observed Vapor Temperature to atmospheric equivalent temperature (AET)
- Annex A10 – Practice for Performance Check.

7.3.6 D5307 Determination of Boiling Range Distribution of Crude Petroleum by Gas Chromatography

This test method covers the determination of the boiling range distribution of water-free crude petroleum through 538 °C (1000 °F). Material boiling above 538 °C is reported as residue. D-5307 is applicable to whole crude samples that can be solubilized in a solvent to permit sampling by means of a microsyringe.

7.3.7 D1078-95 Distillation Range of Volatile Organic Liquids

This test method covers the determination of the distillation range of liquids boiling between 30 and 350 °C, which are chemically stable during the distillation process. D-1078 is applicable to organic liquids such as hydrocarbons, oxygenated compounds, chemical intermediates, and blends thereof. Typically this has been replaced by the D-86 standard.

7.3.8 Nonstandard Distillation Methods – High-temperature Extended Simulated Distillation

A few simulated distillation GC vendors have equipment to extend the range of the GC/simdist standards above by high-temperature simulated distillation (HTSD). This relatively new method which extends ASTM D-2887 determination of the boiling range distribution of hydrocarbons to a FBP of about 1400 °F (760 °C). The use of capillary GC column and stationary phase technology advances,

together with either programmed temperature vaporization (PTV) or on-column injection techniques, provides adequate separation of from C_5 to C_{20} normal paraffins allowing the characterization of petroleum products from about 97 to 1380 °F (36–750 °C).

7.4 *V/L* Methodology

7.4.1 D2533 Vapor–Liquid Ratio of Spark-ignition Engine Fuels

This test method covers a procedure for measuring the volume of vapor formed at standard atmospheric pressure from a given volume of gasoline. The ratio of these volumes is expressed as the vapor–liquid (V/L) ratio of the gasoline at the temperature of the test.

Dry glycerol can be used as the containing liquid for nonoxygenated fuels. Mercury can be used as the containing liquid with both oxygenated and nonoxygenated fuels. Because oxygenates in fuels may be partially soluble in glycerol, gasoline–oxygenate blends must be tested using mercury as the containing fluid. Test Method D-4815 can be used to determine the presence of oxygenates in fuels first.

7.4.2 D5188 Vapor–Liquid Ratio Temperature Determination of Fuels (Evacuated Chamber Method)

D-5188 covers the determination of the temperature at which the vapor formed from a selected volume of volatile petroleum product saturated with air at 0–1 °C (32–34 °F) produces a pressure of one atmosphere in an evacuated chamber of fixed volume. It is applicable to samples for which the determined temperature is between 36 and 80 °C (97 and 176 °F) and the vapor–liquid ratio is between 8 to 1 and 75 to 1.

When the vapor–liquid ratio is 20:1, the result is intended to be comparable to the results determined by Test Method D-2533. This test method may also be applicable at pressures other than one atmosphere, but the stated precision may not apply. It is, however, applicable to both gasoline and gasoline–oxygenate blends.

Other standards of possible use in distillation and V/L analysis of regulated fuel components are:

- ASTM D-216 – Distillation of Natural Gasoline, Standard Method for. This is now supplanted mostly by D-86 on stabilized naphthas from NGL recovery plants
- ASTM-E1405 – Laboratory Glass Distillation Flasks, Standard Specification for
- ISO-918 – Volatile Organic Liquids for Industrial Use – Determination of Distillation Characteristic
- ASTM D-1093 – Acidity of Hydrocarbon Liquids and Their Distillation Residues, Standard Test Method for
- ISO 3405 Petroleum products – Determination of Distillation Characteristics
- ISO/DIS 3924 Petroleum products – Determination of Boiling Range Distribution – Gas Chromatography Method (Revision of ISO 3924: 1977)
- ISO/DIS 6616 Petroleum products – Determination of Distillation Characteristics at Reduced Pressures.

7.5 Measurement of Reid Vapor Pressure – Gasoline Products

7.5.1 D323 Vapor Pressure of Petroleum Products (Reid Method)

This is the original RVP test standard. This test method covers four procedures for the determination of vapor pressure of gasoline, volatile crude oil, and other volatile petroleum products. Procedure A is applicable to gasoline and other petroleum products with a vapor pressure of less than 180 kPa (26 psi). Procedure B may also be applicable to these other materials, but only gasoline was included in the interlaboratory test program to determine the precision of this test method. Neither procedure is applicable to liquefied petroleum gases or fuels containing oxygenated compounds other than MTBE or alcohols. Procedure C is for materials with a vapor pressure of greater than 180 kPa (26 psi) and Procedure D for aviation gasoline with a vapor pressure of approximately 50 kPa (7 psi).

Because the external atmospheric pressure is counteracted by the atmospheric pressure initially present in the vapor chamber, the RVP is an absolute pressure at 37.8 °C (100 °F) in kilopascals (pounds-force per square inch). The RVP differs from the true vapor pressure of the sample because of some small sample vaporization and the presence of water vapor and air in the confined space. It is *not* suitable for use with oxygenated fuels. For determination of the vapor pressure of gasoline–oxygenate blends, refer to Test Method D4953.

7.5.2 D4953 Vapor Pressure of Gasoline and Gasoline–Oxygenate Blends (Dry Method)

This test method, a modification of test Method D323 (Reid method), provides two procedures to determine the vapor pressure of gasoline and gasoline–oxygenate blends. Again, because the external atmospheric pressure is counteracted by the atmospheric pressure initially present in the air chamber, this vapor pressure is an absolute pressure at 37.8 °C (100 °F) in kilopascals (pounds-force per square inch). This vapor pressure differs from the true vapor pressure of the sample because of some small vaporization of the sample and air in the confined space.

D4953 is applicable to gasolines and gasoline–oxygenate blends with a vapor pressure range from 35 to 100 kPa (5–15 psi). Vapor pressure of gasoline or gasoline–oxygenate blends below 35 kPa (5 psi) or greater than 100 kPa (15 psi) can be determined with D4953, but the precision and bias as described in section 10 of the standard do not apply. For materials with a vapor pressure greater than 100 kPa (15 psi), use the 0 to 200 kPa (0–30 psi) gage as specified in the Annex of Test Method D323.

7.5.3 D5190 Vapor Pressure of Petroleum Products (Automatic Method)

D5190 covers the determination of the total pressure of air-containing volatile petroleum products. It is suitable for testing samples with bps above 0 °C (32 °F) that exert a vapor pressure between 7 and 172 kPa (1 and 25 psi) at 37.8 °C (100 °F) at a vapor-to-liquid ratio of 4 : 1. This test method is suitable for testing gasoline samples that contain oxygenates. No account is taken of dissolved water in the sample.

This test method is suitable for the calculation of a DVPE by means of a correlation equation (see section 6.5.2.2). The calculated DVPE very closely approximates the dry vapor pressure that would be obtained on the same material when tested in accordance with test Method D4953.

7.5.4 D5191 Vapor Pressure of Petroleum Products (Mini Method)

D-5191 covers the use of automated vapor pressure instruments to determine the total vapor pressure exerted in vacuum by air-containing volatile liquid petroleum products. This test method is suitable for testing samples with bps above 0 °C (32 °F) that exert a vapor pressure between 7 and 130 kPa (1.0 and 18.6 psi) at 37.8 °C (100 °F) at a vapor-to-liquid ratio of 4 : 1. Measurements are made on liquid sample sizes in the range of 1–10 mL. No account is taken of dissolved water in the sample. Samples can also be tested at other vapor-to-liquid ratios, temperatures, and pressures, but the precision and bias statements need not apply.

D-5191 is suitable for calculation of the DVPE of gasoline and gasoline–oxygenate blends by means of a linear correlation equation. The calculated DVPE very closely approximates the dry vapor pressure that would be obtained on the same material when tested by test Method D-4953.

7.5.5 D5482 Vapor Pressure of Petroleum Products (Mini Method–Atmospheric)

This test method provides a procedure for the determination of total vapor pressure of petroleum products using automatic vapor pressure instruments. The test method is suitable for testing samples with bps above 0 °C (32 °F) that exert a vapor pressure between 7 and 110 kPa (1.0 and 16 psi) at 37.8 °C (100 °F) at a vapor-to-liquid ratio of 4 : 1. D5482 is applicable to gasolines containing oxygenates. No account is taken of dissolved water in the sample.

Because the external atmospheric pressure does not influence the resultant vapor pressure, this vapor pressure is an absolute pressure at 37.8 °C (100 °F) in kPa (psi). This vapor pressure differs from the true vapor pressure of the sample because of some small vaporization of the sample and dissolved air into the air of the confined space.

The D-5482 test is a modification of Test Method D5191 (mini method) where the test chamber is at atmospheric pressure prior to sample injection. It covers the use of automated vapor pressure instruments that perform measurements on liquid sample sizes in the range from 1 to 10 mL. D5482 is also suitable for the determination of the Test Method D4953 DVPE of gasoline and gasoline–oxygenate blends by means of a correlating equation.

7.5.6 D2889 Calculation of True Vapor Pressures of Petroleum Distillate Fuels

This test method describes the calculation of true vapor pressures of petroleum distillate fuels for which data may be obtained in accordance with Test Method D-86 without reaching a decomposition point prior to obtaining 90 volume% distilled. D-2889 may be used to calculate vapor pressures at temperatures between the 0% equilibrium flash temperature and the critical temperature of the fuel. Provisions are included for obtaining a calculated critical temperature for fuels for which it is not known.

The method is not reliable for distillate fuels having a boiling range of less than 100 °F (38 °C) between the ASTM Test Method D-86 10 and 90 volume% distilled temperatures.

Other applicable standards for RVP analysis are found in:

- ISO 3007 Petroleum Products – Determination of Vapor Pressure – Reid method

7.6 Detection and Measurement of Aromatics – Gasoline/Avgas Cuts

7.6.1 D5580-95 Determination of Benzene, Toluene, Ethylbenzene, p/m-*Xylene,* o-*Xylene, C_9 and Heavier Aromatics, and Total Aromatics in Finished Gasoline by Gas Chromatography*

D5580 covers the determination of benzene, toluene, ethylbenzene, the xylenes, C_9 and heavier aromatics, and total aromatics in finished motor gasoline by

GC. The aromatic hydrocarbons are separated without interferences from other hydrocarbons in finished gasoline. Nonaromatic hydrocarbons that have a bp greater than *n*-dodecane may cause interference with the determination of the C_9 and heavier aromatics. For the C_8 aromatics, *o*-xylene and *m*-xylene co-elute, while ethylbenzene and xylene are separated. The C_9 and heavier aromatics are determined as a single group.

This test method covers the following concentration ranges, in liquid volume%, for the preceding aromatics: benzene, 0.1–5%; toluene, 1–15%; individual C_8 aromatics, 0.5–10%; total C_9 and heavier aromatics, 5–30%, and total aromatics, 10–80%. Results are reported to the nearest 0.01% by either mass or by liquid volume.

Many of the common alcohols and ethers that are added to gasoline to reduce carbon monoxide emissions and increase octane, do not interfere with the analysis. Ethers such as MTBE, ETBE, TAME, and diisopropyl ether (DIPE) have been found to elute from the precolumn with the nonaromatic hydrocarbons to vent. Other oxygenates, including methanol and ethanol, elute before benzene and the aromatic hydrocarbons. 1-Methylcyclopentene has also been found to elute from the precolumn to vent and does not interfere with benzene.

7.6.2 D5769 Determination of Benzene, Toluene, and Total Aromatics in Finished Gasolines by Gas Chromatography/Mass Spectrometry

This test method covers the determination of benzene, toluene, and total aromatics in finished motor gasoline, including RFG containing oxygenated blending components, by GC/MS. It is applicable to the following concentration ranges, in liquid volume%, for the following aromatics: benzene 0.1–3%; toluene 1–15%; and total (C_6–C_{12}) aromatics 10–40%. D-5769 has not been tested by ASTM for gasoline samples containing a concentration of uncalibrated C_{10}–C_{12} aromatic compounds greater than approximately 3 volume%. Also, the test method has not been tested by ASTM for individual hydrocarbon process streams in a refinery, such as reformates, fluid catalytic cracked naphthas, and so on, used in blending of gasolines.

Results are reported to the nearest 0.01% for benzene and 0.1% for the other aromatics by either mass or liquid volume.

7.6.3 D4420-94 Determination of Aromatics in Finished Gasoline by Gas Chromatography

The D4420 test method provides for the determination of benzene, toluene, C_8, C_9 and heavier aromatics, and total aromatics in finished motor gasoline. It is also applicable to gasoline blending components. For analysis of gasolines for benzene and toluene, solely, test Method D-3606 is appropriate.

D-4420 is applicable to the following concentration ranges, in liquid volume percent, for the individual aromatics listed above: benzene, 0.1–5; toluene, 1–20; C, 3–25; C and heavier, 5–30, and total aromatics, 10–80. Results are reported to the nearest 0.1% by liquid volume.

D4420 is slightly more accurate than D-5580, but it has not been determined whether this test method is applicable to gasoline containing oxygenates.

7.6.4 D3257-e1 Aromatics in Mineral Spirits by Gas Chromatography

These test methods cover the determination of ethylbenzene and total eight-carbon (C_8) and heavier aromatics in the concentration range from 0.1 to 30% in mineral spirits having a distillation range from 149 to 210 °C (300–410 °F) as determined by Test Method D86. The procedures permit the identification and calculation of concentrations of aromatic components to 0.1 volume%.

Oxygenated compounds, if present, may interfere and cause erroneous results. Such oxygenated compounds are not normally present in mineral spirits. Two test methods are covered as follows:

- Test Method A, measurement of ethylbenzene content, C plus higher aromatics (except ethylbenzene), and total aromatics by means of a single GC analysis.
- Test Method B, measurement of ethylbenzene content by means of a rapid GC analysis.

7.6.5 D5292 Aromatic Carbon Contents of Hydrocarbon Oils by High Resolution Nuclear Magnetic Resonance Spectroscopy

D-5292 covers the determination of the aromatic hydrogen content (Procedures A and B) and aromatic carbon content (Procedure C) of hydrocarbon oils using high-resolution NMR spectrometers. Applicable samples include kerosines, gas oils, mineral oils, lubricating oils, coal liquids, and other distillates that are completely soluble in chloroform and carbon tetrachloride at ambient temperature. For pulse Fourier transform spectrometers, the detection limit is typically 0.1 mol% aromatic hydrogen atoms and 0.5 mol% aromatic carbon atoms. For continuous wave (CW) spectrometers, which are suitable for measuring aromatic hydrogen contents only, the detection limit is considerably higher and typically 0.5 mol% aromatic hydrogen atoms.

The reported units are mole percent aromatic hydrogen atoms and mole percent aromatic carbon atoms. This test method is not applicable to samples containing more than 1 mass% olefinic or phenolic compounds, and neither does it cover the determination of the percentage mass of aromatic compounds in oils, since NMR signals from both saturated hydrocarbons and aliphatic substituents

on RA compounds appear in the same chemical shift region. For the determination of mass or volume percent aromatics in hydrocarbon oils, chromatographic, or mass spectrometry methods can be used.

7.6.6 D5776 Bromine Index of Aromatic Hydrocarbons by Electrometric Titration

This test method determines the amount of bromine-reactive material in aromatic hydrocarbons and is thus a measure of trace amounts of unsaturates in these materials. It is applicable to materials having bromine indexes below 500.

The D-5776 standard is applicable to aromatic hydrocarbons containing no more than trace amounts of olefins and that are substantially free from material lighter than isobutane and have a distillation EP under 288 °C (550 °F)

7.6.7 D2789 Hydrocarbon Types in Low-olefinic Gasoline by Mass Spectrometry

This test method discusses the determination by mass spectrometry of the total paraffins, monocycloparaffins, dicycloparaffins, alkylbenzenes, indans or tetralins or both, and naphthalenes in gasoline which have an olefin content of less than 3 volume% and a 95% distillation point of less than 210 °C (411 °F) as determined in accordance with Test Method D-86. Olefins are determined by Test Method D-1319, or by Test Method D-875. It has not been determined if D2789 is sufficiently applicable to gasolines containing oxygenated compounds (for example, alcohols and ethers).

7.7 Detection and Measurement of Aromatics – Turbine and Diesel Fuels

7.7.1 D2549 Separation of Representative Aromatics and Nonaromatics Fractions of High-boiling Oils by Elution Chromatography

D-2549 methodology covers the separation and determination of representative aromatics and nonaromatics fractions from hydrocarbon mixtures that boil between 232 and 538 °C (450 and 1000 °F). Alternative procedures are provided for the separation of 2 g or 10 g of hydrocarbon mixture. Some components may not be eluted from the chromatographic column for some types of sample under the conditions used in this method. It is not applicable to gasoline range cuts.

Test Method D-2007 is an alternative method of separating high-boiling oils into polar compounds, aromatics, and saturates fractions.

7.7.2 D2425 Hydrocarbon Types in Middle Distillates by Mass Spectrometry

This test method covers an analytical scheme using the mass spectrometer to determine the hydrocarbon types present in virgin middle distillates 204–343 °C (400–650 °F) boiling range, 5–95 volume% as determined by Method D86. Samples with an average carbon number value of paraffins between C_{12} and C_{16} and containing paraffins from C_{10} and C_{18} can be analyzed. Eleven hydrocarbon types are determined. These include paraffins, noncondensed cycloparaffins, condensed dicycloparaffins, condensed tricycloparaffins, alkylbenzenes, indans or tetralins, or both, C_nH_{n-10} (indenes, etc.), naphthalenes, C_nH_{n-14} (acenaphthenes, etc.), C_nH_{n-16} (acenaphthylenes, etc.), and tricyclic aromatics.

7.8 General Ponaffin, Isoponaffin, Aromatic, Naphthene, Olefin Analyses for Aromatics and Other Components

7.8.1 D3238 Calculation of Carbon Distribution and Structural Group Analysis of Petroleum Oils by the n-d-M *Method*

D3238 discusses the calculation of the carbon distribution and ring content of olefin-free petroleum oils from measurements of refractive index, density and molecular weight (*n-d-M*). This test method should not be applied to oils whose compositions are outside the following ranges: in terms of carbon distribution – up to 75% carbon atoms in ring structure; percentage in RAs not larger than 1.5 times the percentage in naphthenic rings; in terms of ring content – one can have up to four rings per molecule, with not more than half of them aromatic. A correction must be applied for oils containing significant quantities of sulfur.

The composition of complex petroleum fractions is often expressed in terms of the proportions of RA, naphthene rings (RN) and paraffin chains (CP) that would comprise a hypothetical mean molecule. Alternatively, the composition may be expressed in terms of a carbon distribution, that is, the percentage of the total number of carbon atoms that are present in RA structures (% CA), naphthene ring structures (% CN) and in CP (% CP).

7.8.2 D5134 Detailed Analysis of Petroleum Naphthas Through n-*Nonane by Capillary Gas Chromatography*

This test method covers the determination of hydrocarbon components of petroleum naphthas. Components eluting after *n*-nonane (bp 150.8 °C) are determined as a single group. Components that are present at the 0.05 mass % level or greater can be determined.

D5134 applies to olefin-free (<2% olefins by liquid volume) liquid hydrocarbon mixtures including virgin naphthas, reformates, and alkylates. Olefin content can be determined by Test Method D1319. The hydrocarbon mixture must have a 98% distillation point of 250 °C or less as determined by Test Method D3710.

7.8.3 D5443 Paraffin, Naphthene, and Aromatic Hydrocarbon-type Analysis in Petroleum Distillates through 200 °C by Multidimensional Gas Chromatography

This test method provides for the determination of paraffins, naphthenes, and aromatics by carbon number in low-olefinic hydrocarbon streams which have FBPs of 200 °C or less. Hydrocarbons with bps greater than 200 °C and less than 270 °C are reported as a single group. Olefins, if present, are hydrogenated and the resultant saturates are included in the paraffin and naphthene distribution. Aromatics boiling at C_9 and above are reported as a single aromatic group.

D5443 was not intended to determine individual components except for benzene and toluene that are the only C_6 and C_7 aromatics, respectively, and cyclopentane, which is the only C_5 naphthene. The lower limit of detection for a single hydrocarbon component or group is 0.05 mass%. This test method is applicable to hydrocarbon mixtures including virgin, catalytically converted, thermally converted, alkylated and blended naphthas.

Some other standards and reference sources for aromatics analyses in hydrocarbon fuels are:

- API-4439 – Exposure Data on C7 and C8 Aromatics During Handling and Production of Motor Gasolines
- ASTM D-2267 – Aromatics in Light Naphthas and Aviation Gasolines by Gas Chromatography, Standard Test
- ASTM D-875 – Calculation of Olefins and Aromatics in Petroleum Distillates from Bromine Number and Acid a″
- ISO 5661: 1983 Petroleum Products – Hydrocarbon Liquids – Determination of Refractive Index
- ISO 3837: 1993 Liquid Petroleum Products – Determination of Hydrocarbon Types – Fluorescent Indicator Adsorption Method
- ISO 2549 – Separation of Representative Aromatics and Nonaromatics Fractions of High Boiling Oils by Elution Chromatography.

7.9 Detection and Measurement of Benzene

7.9.1 D3606 Determination of Benzene and Toluene in Finished Motor and Aviation Gasoline by Gas Chromatography

This test method provides for a primary standard for the determination of benzene and toluene in finished motor and aviation gasolines by GC. As such, the standard mandates calibrations before routine sample preparation and analysis, all of which adds up to a fairly tedious process. Benzene can be determined between the levels of 0.1 and 5 volume% and toluene can be determined between the levels of 2 and 20 volume%. It has not been determined whether D3606 is applicable to gasolines containing oxygenated compounds (for example, alcohols or ethers). Ethanol is known to interfere with the benzene determination and some modifications of the GC column to correct for this are being considered.

7.9.2 D4053 Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy

This test method covers the determination of the percent benzene in full-range gasoline. It is applicable to concentrations from 0.1 to 5 volume%. D-4053 has not been validated for gasolines containing oxygenates.

7.9.3 D4534-e1 Benzene Content of Cyclic Products by Gas Chromatography

This test method covers the determination of the benzene content of specific cyclic hydrocarbon products. Benzene may be determined over a range from 5–300 mg kg^{-1}. The products in which benzene can be determined include cyclohexane, toluene, individual C_8 aromatics, cumene, and styrene.

Some other standards and reference sources for benzene analysis in hydrocarbon fuels are:

- API-DR-111 – Estimation of Incremental Benzene Exposure Associated with Bulk Gasoline Storage
- API-DR-195 – Naturally Occurring Benzene, Toluene and Xylenes in Soils
- API-DR-221 – Removal of Benzene from Refinery Wastewater
- API-4470-1 – Volume 1: Exhaust Benzene Emissions from Late Model Vehicles
- ASTM D-1017 – Benzene and Toluene in 250F and Lighter Petroleum Products by Ultraviolet Spectrophotometer
- ASTM D-4053 – Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy, Standard Test Method for
- ASTM D-4367 – Benzene in Hydrocarbon Solvents by Gas Chromatography, Standard Test Method for
- ASTM D-4534 – Benzene Content of Cyclic Products by Gas Chromatography, Standard Test Method for
- ASTM D-4600 – Determination of Benzene-soluble Particulate Matter in Workplace Atmospheres, Standard
- ASTM D-5580 – Determination of Benzene, Toluene, Ethylbenzene, *p*/*m*-Xylene, *o*-Xylene, C_9 and Heavier Aromatics
- ASTM D-5769 – Determination of Benzene, Toluene, and Total Aromatics in Finished Gasolines by Gas Chromatography.

See also the above section on aromatics and PIANO analysis standards.

7.10 Detection and Measurement of Olefins

Although no specific recommendations for one method over another is promoted (beyond mentioning the regulatory requirements) in most of section 7, an exception must be made here.

Olefins are a priority control constituent for CARB and US EPA Phase II gasolines. The US EPA has proposed that fuel olefins of ozone-forming potential be measured via the FIA analysis procedures outlined in ASTM D-1319. Several sources, including the author, have noted that FIA olefins for high olefins gasoline blend stocks may be far less accurate than alternative means of measurements.

This applies most to unhydroprocessed thermally cracked streams in a refinery. Chiefly these are FCC and coker light cuts that happen to go to the gasoline pool. The FCC gasoline may have as much as 50% olefinic material, the coker as much as 10%. If it is necessary to control the rates of either of these blend stocks into the blend header with a potential olefins constraint, one should exercise caution in the assumed amount of the stock that will meet all criteria based on linear blend values of properties. A safe option would be to assume 10–15% more olefin in the stock than what is reported by FIA analysis, at least until a detailed biasing study is performed with several months of on-line data available.[76]

7.10.1 D1159 Bromine Numbers of Petroleum Distillates and Commercial Aliphatic Olefins by Electrometric Titration

This test method covers the determination of the bromine number of petroleum distillates that are substantially free of material lighter than isobutane and that have 90% distillation points (by Test Method D-86) under 327 °C (626 °F). This test method is generally applicable to gasoline (including leaded, unleaded and oxygenated fuels), kerosine, and distillates in the gas oil range that fall within the following limits: bromine number at 90% distillation point, under 205 °C (400 °F) = 2 and from 205 to 327 °C (400–626 °F) = 10.

The magnitude of the bromine number is an indication of the quantity of bromine-reactive constituents, not an identification of constituents; therefore, its application as a measure of olefinic unsaturation should not be undertaken without the study given in Annex A1 of the standard.

For petroleum hydrocarbon mixtures of bromine number less than 1.0, a more precise measure for bromine-reactive constituents can be obtained by using Test Method D-2710. If the bromine number is less than 0.5, then Test Method D-2710 or the comparable bromine index methods for industrial aromatic hydrocarbons, Test Method D-1492, must be used in accordance with their respective scopes. The practice of using a factor of 1000 to convert bromine number to bromine index is not applicable to these lower values of bromine number.

7.10.2 D1319-a Hydrocarbon Types in Liquid Petroleum Products by Fluorescent Indicator Adsorption

This test method is for determining hydrocarbon types over the concentration ranges from 5–99 volume% aromatics, 0.3–55 volume% olefins, and 1–95 volume% saturates in petroleum fractions that distill below 315 °C, although some reports sharply limit this effective range. D-1319 may apply to concentrations outside these ranges, but the precision has not been determined. Samples containing dark-colored components that interfere in reading the chromatographic bands cannot be analyzed.

This test method is intended for use with full boiling range products. Comparative data have established that the precision statement does not apply to narrow boiling petroleum fractions near the 315 °C limit. Such samples are not eluted properly, and results are erratic.

The precision statement for this test method has been determined with unleaded fuels that do not contain oxygenated blending components. It may or may not apply to automotive gasolines containing lead antiknock mixtures or oxygenated gasoline blending components, or both.

The following oxygenated blending components: methanol, ethanol, MTBE, TAME and ETBE do not interfere with the determination of hydrocarbon types at concentrations normally found in commercial blends. These oxygenated components are not detected, since they elute with the alcohol desorbent. Other oxygenated compounds must be individually verified. When samples containing oxygenated blending components are analyzed, the results must be corrected to a total-sample basis.

7.10.3 D2710-e1 Bromine Index of Petroleum Hydrocarbons by Electrometric Titration

This test method covers the determination of the amount of bromine-reactive material in petroleum hydrocarbons, and is thus a measure of trace amounts of unsaturates in these materials. It is applicable to materials having bromine indexes below 1000.

The D-2710 test method is applicable only to essentially olefin-free hydrocarbons or mixtures that are substantially free from material lighter than isobutane and have a distillation end point (EP) under 288 °C (550 °F). This procedure has been comparatively tested on materials with bromine indexes in the range from 100 to 1000. These materials include petroleum distillates

such as straight-run and hydrocracked naphtha, reformer feed, kerosine, and aviation turbine fuel. Materials with bromine index greater than 1000 should be tested for bromine number using Test Method D-1159/IP 130. Industrial aromatic hydrocarbons should be tested using Test Method D-1492.

Other possible applicable standards for olefins analysis are:

- ASTM D-875 – Calculation of Olefins and Aromatics in Petroleum Distillates from Bromine Number and Acid
- ISO 3837 Liquid Petroleum Products – Determination of Hydrocarbon Types – Fluorescent Indicator Adsorption (FIA) Method
- ISO 3839 Petroleum Products – Determination of Bromine Number of Distillates and Aliphatic Olefins – Electrometric Method

7.11 Detection and Measurement of Sulfur

7.11.1 D1266 Sulfur in Petroleum Products (Lamp Method)

D-1266 covers the determination of total sulfur in liquid petroleum products in concentrations from 0.01 to 0.4 mass% (Note 1). A special sulfate analysis procedure is described in Annex A1 that permits the determination of sulfur in concentrations as low as 5 mg kg^{-1}.

The direct burning procedure (section 9) is applicable to the analysis of such materials as gasoline, kerosine, naphtha, and other liquids that can be burned completely in a wick lamp. The blending procedure (section 10) is applicable to the analysis of gas oils and distillate fuel oils, naphthenic acids, alkyl phenols, high sulfur-content petroleum products, and many other materials that cannot be burned satisfactorily by the direct burning procedure.

Phosphorus compounds normally present in commercial gasoline do not interfere. A correction is given for the small amount of acid resulting from the combustion of the lead antiknock fluids in gasolines. Appreciable concentrations of acid-forming or base-forming elements from other sources interfere when the titration procedure is employed because no correction is provided in these cases.

7.11.2 D2622 Sulfur in Petroleum Products by X-ray Spectrometry

D2622 covers the determination of total sulfur in liquid petroleum products and in solid petroleum products that can be liquefied with moderate heating or dissolved in a suitable organic solvent. The applicable concentration range will vary to some extent with the instrumentation used and the nature of the sample. Optimum conditions will allow the direct determination of sulfur in essentially paraffinic samples at concentrations exceeding 0.0010 mass%.

Methanol containing fuels M-85 and M-100 may be analyzed with an accompanying loss of sensitivity and precision because of the more absorbing matrix caused by the high oxygen content of these fuels. M-85 is 85% methanol–15% gasoline, and M-100 fuel is 100% methanol. Correction factors are applied to achieve these results.

7.11.3 D3227 Mercaptan Sulfur in Gasoline, Kerosine, Aviation Turbine, and Distillate Fuels (Potentiometric Method)

This test method covers the determination of mercaptan in gasolines, kerosines, aviation turbine fuels, and distillate fuels containing from 0.0003 to 0.01 mass% of mercaptan sulfur. Organic sulfur compounds such as sulfides, disulfides, and thiophene do not interfere. Elemental sulfur in amounts less than 0.0005 mass% does not interfere. Hydrogen sulfide will interfere, if not removed.

7.11.4 D4294 Sulfur in Petroleum Products by Energy-dispersive X-ray Fluorescence Spectroscopy

This standard covers the measurement of sulfur in hydrocarbons such as naphthas, distillates, fuel oils, residues, lubricating base oils and nonleaded gasoline. The concentration range is from 0.05 to 5 mass%.

D-4294 applicable tests may involve hazardous materials, operations, and equipment. The standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices, and then determine the applicability of regulatory limitations prior to use. For specific precautionary statements, see section 7.

7.11.5 D4952 Qualitative Analysis for Active Sulfur Species in Fuels and Solvents (Doctor Test)

This test method is intended primarily for the detection of mercaptans in motor fuel, kerosine, and similar petroleum products. This test is one of the oldest in the refining industry, but still reasonably accurate.

7.11.6 D5623 Sulfur Compounds in Light Petroleum Liquids by Gas Chromatography and Sulfur-selective Detection

This test method covers the determination of volatile sulfur-containing compounds in light petroleum liquids. It is applicable to distillates, gasoline motor fuels (including

those containing oxygenates) and other petroleum liquids with a FBP of approximately 230 °C (450 °F) or lower at atmospheric pressure. The applicable concentration range will vary to some extent depending on the nature of the sample and the instrumentation used; in most cases, however, the test method is applicable to the determination of individual sulfur species at levels of $0.1-100\,mg\,kg^{-1}$.

D-5623 does not purport to identify all individual sulfur components. Detector response to sulfur is linear and essentially equimolar for all sulfur compounds within the standard's scope (1.1). Both unidentified and known individual compounds can be determined. Many sulfur compounds, for example, hydrogen sulfide and mercaptans, are reactive and their concentration in samples may change during sampling and analysis. Consequently, the total sulfur content of samples is estimated from the sum of the individual compounds determined. The D-5623 test method is not the preferred method for determination of total sulfur if some other less roundabout analysis method is available.

Other possible applicable standards for sulfur analysis are:

- ISO 4260 Petroleum Products and Hydrocarbons – Determination of Sulfur Content – Wickbold Combustion Method
- ISO 8754 Petroleum Products – Determination of Sulfur Content – Energy-dispersive X-ray Fluorescence Method
- ISO/DIS 14596 Petroleum Products – Determination of Sulfur Content – Wavelength-dispersive X-ray Fluorescence Spectrometry
- ISO-6326-3 – Part 3: Determination of Hydrogen Sulfide, Mercaptan Sulfur and Carbonyl Sulfide Sulfur B
- ASTM D-1219 – Fuel, Aviation Turbine, Mercaptan Sulfur in (Color Indicator Method)
- ASTM D-129 – Sulfur in Petroleum Products (General Bomb Method), Standard Test Method for
- ASTM D-1323 – Fuels, Aviation Turbine, Mercaptan Sulfur in, Amperometric and Potentiometric Methods
- ASTM D-2747 – Trace Quantities of Total Sulfur in Volatile Organic Liquids (Oxyhydrogen Combustion Method)
- ASTM D-2785 – Trace Quantities of Total Sulfur (Wickbold & Beckman Combustion Apparatus)
- ASTM D-3246 – Microcoulometry, Standard Test Method for Sulfur in Petroleum Gas by Oxidative SCAQMD-307 Speciated Sulfur Content for Reformulated Gasolines.

7.12 Detection and Measurement of Oxygenates in Fuel

Tests applicable to gasoline are not necessarily applicable to its blends with oxygenates. Consequently, the type of fuel under consideration must first be identified in order to select applicable tests. Standard D-4814 outlines the oxygenate requirements and Test Method D-4815 provides a procedure for determining oxygenate concentration in mass percent. These tests have not yet to withstand the test of time, so always exercise appropriate professional judgment in their use.

7.12.1 D4814-c Automotive Spark-ignition Engine Fuel

Though not really an analysis standard, D-4814 sets the guidelines in establishing requirements of automotive fuels for ground vehicles equipped with spark-ignition engines. It is included here as a reminder to analysts that this is the defining specification for the amounts and types of oxygenates that are required to be in RFG, including such items as seasonality and variations in amounts. It changes often and the latest version should always be used for setting control targets in the blender.

D-4814 describes various characteristics of automotive fuels for use over a wide range of operating conditions. It provides for a variation in the volatility and water tolerance of automotive fuel in accordance with seasonal climatic changes at the locality where the fuel is used. For the period May 1 through September 15, the maximum vapor pressure limits issued by the US EPA are specified for each geographical area except Alaska and Hawaii. Variation of AKI with seasonal climatic changes and altitude is discussed in Appendix X1 of D-4814. This specification neither necessarily includes all types of fuel that are satisfactory for automotive vehicles, nor necessarily excludes fuels that can perform unsatisfactorily under certain operating conditions or in certain equipment. The significance of each of the properties of this specification is shown in Appendix X1 of D-4814.

The spark-ignition engine fuels covered in this specification are gasoline and its blends with oxygenates such as alcohols and ethers. This specification does not apply to fuels that contain an oxygenate as the primary component, such as fuel methanol (M85). The concentrations and types of oxygenate are not specifically limited in this specification. However, depending on oxygenate type, as oxygenate content increases above some threshold level, the likelihood of vehicle problems also increases. The composition of both unleaded and leaded fuel is limited by economic, legal, and technical considerations, but their properties, including volatility, are defined by this specification. In addition, the composition of unleaded fuel is subject to the rules, regulations, and CAA waivers of the US EPA. With regard to fuel properties, including

volatility, this specification can be more or less restrictive than the US EPA rules, regulations, and waivers. Refer to Appendix X3 of the Standard for discussions of US EPA rules relating to fuel volatility, lead and phosphorous contents, and use of oxygenates in blends with unleaded gasoline. Contact US EPA for the latest versions of the rules and additional requirements.

D-4814 represents a description of automotive fuel as of the date of publication. The specification is under continuous review, which can result in revisions based on changes in fuel, automotive requirements, or test methods, or a combination thereof. All users of this specification, therefore, should refer to the latest edition.

7.12.2 D4815-a Determination of Methyl Tertiary Butyl Ether, Ethyl Tertiary Butyl Ether, Tertiary Amyl Methyl Ether, Diisopropyl Ether, Tertiary Amyl Alcohol and C_1 to C_5 Alcohols in Gasoline by Gas Chromatography

This test method is designed for the determination of ethers and alcohols in gasolines by GC. Specific compounds determined are: MTBE, ETBE, TAME, DIPE, methanol, ethanol, isopropanol, *n*-propanol, isobutanol, *tert*-butanol, *sec*-butanol, *n*-butanol, and *tert*-pentanol (*tert*-amylalcohol).

Individual ethers are determined from 0.1 to 20.0 mass%. Individual alcohols are determined from 0.1 to 12.0 mass%. Equations used to convert to mass percent oxygen and to volume percent of individual compounds are provided. Alcohol-based fuels such as M-85 and E85, MTBE product, ethanol product and denatured alcohol are specifically excluded from this method. The methanol content of M-85 fuel is considered beyond the operating range of the system.

Benzene, while detected, cannot be quantified using this test method, and must be analyzed by alternate methodology (Test Method D-3606 or D-4420).

Test Method D-4815 also includes procedures for calculating mass oxygen content and oxygenate concentration in volume percent. Appendix X4 of the method provides a procedure for calculating the mass oxygen content of a fuel using measured oxygenate type, oxygenate concentration in volume percent, and measured density or relative density of the fuel.

7.12.3 D5599 Determination of Oxygenates in Gasoline by Gas Chromatography and Oxygen Selective Flame Ionization Detection

This test method covers a GC procedure for the quantitative determination of organic oxygenated compounds in gasoline which have a FBP not greater than 220 °C and oxygenates which have a bp limit of 130 °C. It is applicable when oxygenates are present in the 0.1–20% by mass range.

D-5599 is intended to determine the mass concentration of each oxygenate compound present in a gasoline. This requires knowledge of the identity of each oxygenate being determined (for calibration purposes). However, the oxygen-selective detector used in this test method exhibits a response that is proportional to the mass of oxygen. It is, therefore, possible to determine the mass concentration of oxygen contributed by any oxygenate compound in the sample, whether or not it is identified. Total oxygen content in a gasoline may be determined from the summation of the accurately determined individual oxygenated compounds. The summed area of other, uncalibrated or unknown oxygenated compounds present, may be converted to a mass concentration of oxygen and summed with the oxygen concentration of the known oxygenated compounds.

7.12.4 D5622 Determination of Total Oxygen in Gasoline and Methanol Fuels by Reductive Pyrolysis

The test methods in this standard cover the quantitative determination of total oxygen in gasoline and methanol fuels by means of reductive pyrolysis methods. Precision data are provided for 1.0–5.0 mass% oxygen in gasoline and for 40–50 mass% oxygen in methanol fuels.

Several types of instrument can be satisfactory for these test methods. Instruments can differ in the way that the oxygen-containing species is detected and quantified. However, these test methods are similar in that the fuel is pyrolyzed in a carbon-rich environment.

7.13 Detection and Measurement of Nitrogen in Fuels

7.13.1 D5291 Instrumental Determination of Carbon, Hydrogen, and Nitrogen in Petroleum Products and Lubricants

D-5291 is a catch-all standard to cover the instrumental determination of carbon, hydrogen, and nitrogen in laboratory samples of petroleum products and lubricants. Values obtained represent the total carbon, the total hydrogen, and the total nitrogen.

These test methods are applicable to samples such as crude oils, fuel oils, additives, and residues for carbon, hydrogen and nitrogen analysis. These test methods were tested in the concentration range of at least 75–87 mass% for carbon, at least 9–16 mass% for hydrogen, and 0.1–2.0 mass% for nitrogen.

The nitrogen test method is not applicable to light materials or those containing <0.75 mass% nitrogen, such as hydrotreated gasoline, jet fuel, naphtha, diesel fuel, or chemical solvents. These test methods are not recommended for the analysis of volatile materials such as gasoline, gasoline–oxygenate blends, or gasoline-type aviation turbine fuels.

The results of these tests can be expressed as mass percent carbon, hydrogen or nitrogen.

Other standards of interest for fuels nitrogen analysis are:

- ISO 13379 Petroleum Products – Determination of Alkyl Nitrate in Diesel Fuels – Spectrometric Method
- ASTM D-3228 – Determination of Nitrogen by Kjeldahl Method
- ASTM D-4629 – Determination of Trace Nitrogen by X-ray Fluorescence

7.14 Detection and Measurement of Metals in Fuels

These tests are infrequently used in an era where no-lead is the preferred environment. There are instances, where native or contaminant crude-cut metal porphyrins, for example, need to be determined (they poison conversion catalysts). Appropriate test procedures are:

- ISO 3830 Petroleum Products – Determination of Lead Content of Gasoline–Iodine Monochloride Method
- ISO 6245 Petroleum Products – Determination of Ash
- ISO/FDIS 14597 Petroleum Products – Determination of Vanadium and Nickel Content – Wavelength-dispersive X-ray Fluorescence Spectrometry
- ASTM D-3231 – Determination of Phosphorus Content in Gasolines

7.15 Measurement of Fuel Densities

Although not directly a regulated quantity, the density of a petroleum stock is so important to the proper characterization of the stream that it at least needs to be mentioned. Units reported are degrees API or specific gravity. A relatively simple inspection property is given in:

- ASTM D-287 and D-1298 – Measure Density, Relative Density, or API Gravity of Crude Petroleum and Liquid Hydrocarbon Products

Output signals from certain coriolis-method flowmeters have a direct relationship between signal frequency and liquid density, but these methods are yet to be accepted as standards.

8 ON-LINE ANALYSES FOR REGULATORY CONTROL[63,67,68]

8.1 General

Some of the techniques and standards mentioned in sections 6 and 7 are strictly for laboratory-based analyses only. The techniques are either complicated, potentially hazardous, or the machinery needed to perform the analysis is large and clumsy, not at all adaptable to on-line packaging and monitoring in a modern refinery. Two techniques, however, have shown themselves to be readily adaptable to such installations. These are the spectrophotometric analysis of fuel streams by multiple IR absorption with chemometric analytical techniques and the use of fast sample systems in conjunction with GC.

Other methods besides these have been put in on-line environments, UV absorption and even LC, but these two methods hold by far the greatest interest for gasoline blenders in terms both of accuracy of method and available equipment and the wide variety of the tests they can reliably and economically perform. Therefore, the following discussion will cover these methods in detail from the standpoint of their adoption to real-time on-line monitoring.

8.2 On-line Analysis for Blender Control – Chemometric Methods[71]

This discussion is presented here, as it is critical to obtain accurate IR analytical results. IR is one of the most common of all on-line spectrophotometric methods, so the following is geared for use with on-line varieties of IR analysis, most particularly NIR methods. Other techniques for several different spectrophotometric analysis procedures beyond NIR are described previously in this work. The comments presented below are generally applicable to sample preparation and techniques for other spectrophotometric methods in general.

8.2.1 Near-infrared for On-line Gasoline Analysis – General

Near-infrared (NIR) analysis has a tremendous growth in usage for monitoring several process stream qualities of importance to gasoline blenders, in real-time. Although widespread use of NIR has been delayed by some technical challenges, refineries adapting this technology have had significant provable benefits from the installation. Significant savings derive from allowing optimal process configurations and conditions, using fewer instruments, and cost effectively measuring parameters specified by regulatory agencies.

NIR is a fast spectroscopic technique that can be put readily into industrial grade [i.e. able to withstand dust, splashing, vibration, noise, electromagnetic interference (EMI), temperature extremes – within reason] enclosures for on-line blending. NIR offers low cost per analysis in terms of time, equipment cost, and sample processing. With NIR, all analyses are completed in the time it takes to obtain a spectrum (about 1 min), regardless of how many properties are being analyzed. A single

NIR analyzer can replace multiple traditional analyzers, providing faster results. NIR can be used to monitor physical properties, such as gasoline octane number or diesel CN, or chemical properties such as the percentage of aromatics or saturates.

One of the main advantages of NIR analyzers is that they are compatible with conventional quartz fiber optics. With fiber optics, light can pass from the source to the fiber optic, then through the sample, before it is focused onto the detector. This separates the electronics of the analyzer from the sample, which is desirable, particularly when the sample is a combustible liquid. Despite the advantages of fiber optics, their use can cause difficulty for the analytical measurement. First, there is an economic incentive to place the fiber in the stream, blend header, or pipeline. If this can be done, the need for detailed sample conditioning is avoided and the cost of installing additional sample piping. A primary drawback to this strategy is that gasoline streams contain contaminants that may adversely affect the in-line probe.

Figure 13 shows a schematic plus open cabinet display of a Fourier transform NIR gasoline analyzer typical of those currently on the market.

8.2.2 Infrared Calibration – Model Validation[70,73]

Calibration of an NIR analyzer requires a training set of samples for which the properties of interest have been measured by conventional methods. For many applications, a new calibration must be done for each refinery and, sometimes, for each instrument.

Traditional analytical techniques usually depend on a single variable. NIR, on the other hand, is an example of a chemometric technique. In chemometric techniques, the property to be determined, RVP, as a good example, depends on many variables, often in complex, highly nonlinear relationships that may not always be well known before calibration begins.

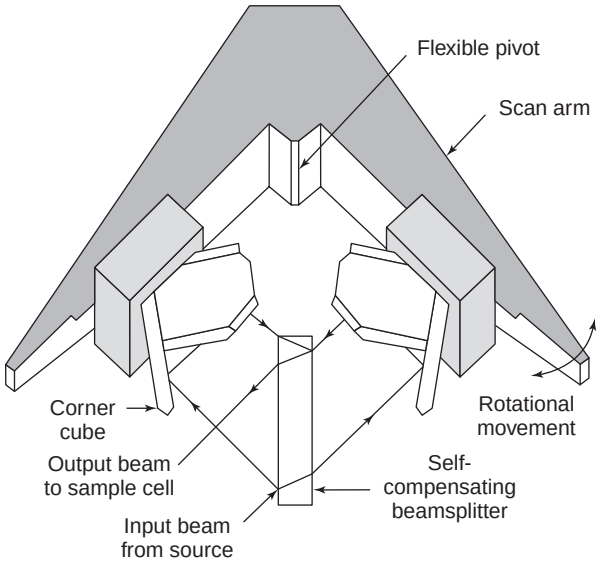

Figure 13 Fourier transform NIR installation.

The process of NIR calibration begins with obtaining the NIR spectra of a training set of samples for which the properties of interest have been measured by traditional means. In the case of a set of gasoline samples, octane numbers would be measured by octane engines, RVP by Grabner cell analysis, and aromatics by laboratory FIA.

Models (equations) that relate the measured properties of the samples to their spectra are generated using a computer and statistical regression analysis or neural net software. These models allow prediction of the properties of unknown samples directly from their spectra. The model predictions are accurate, as long as the unknown samples are sufficiently similar to the training-set samples. Inaccurately predicted samples can be added to the training set to develop a revised model that has greater generality.

Model generality is less of a concern when predicting chemical composition than when predicting physical properties. This is because chemical composition is observed directly as peaks in a spectrum, whereas physical properties are inferred from a complicated correlation to their chemical composition. Generality is also less of a concern when modeling an individual process unit, such as a reformer, because the composition of the product does not vary as widely from day to day as do finished gasolines of several different grades.

Even when unknown samples are too different from the training set to permit accurate (true) predictions, NIR predictions can still be quite precise (repeatable). For those trying to control and optimize a process unit, the immediacy and precision of NIR predictions can be more important than their accuracy (where a bias can always be installed). This is particularly true when trying to determine the changes in process stream properties following a change in operating parameters.

8.2.3 Calibration – Temperature Effects

Temperature affects the analytical measurement because the amount of light absorbed by the sample is a function of the density of the material, and hydrocarbon densities change with temperature. Temperature can also affect the transmission characteristics of the fiber optic. One reason for this is that the fiber, like the sample, absorbs part of the light, and the magnitude of that absorption is a function of temperature. This is a major problem for in-stream placement of NIR sensors in that the temperature becomes as an uncontrolled variable introduced into the calibration model.

Ambient temperature is not an important factor in the transmission of digital signals by fiber optics. The user

should be aware, however, that NIR sources produce continuous light emissions in an analog fashion and that the magnitude of the output is important.

Partial compensation can be made for any of the temperature-induced inaccuracies, but the effects of temperature can greatly exceed the error of the analyzer. The variance of error so introduced by temperature is directly additive to the variance of the total system error. Therefore, like analyzer error, it must be controlled at least to the number of significant figures desired in the measured property.

8.2.4 Calibration – Reference Sampling

A final consideration that deserves attention is the type of reference scan made by the analyzer. An absorbance measurement involves the use of a reference sample to calculate a ratio. Systems that use a stored reference scan, or systems where the reference fiber optic does not experience the same environmental conditions as the sample fiber (length of fiber run, temperature changes, etc.) can add even more error to the absorbance measurement. Several refiners have historically used on-line NIR analysis successfully for conventional and ethanol-blended gasolines (a hot technical topic in the synfuel-oriented 1980s) in feedback closed-loop control. Most find that the use of a ''protofuel'' validation system is invaluable, especially when measuring multiple targets of gasoline blends. There are several ways to do this effectively.

For example, the system can be preset to run before, during, and near the end of the blending operation. If the calibration has changed, as indicated by the statistical analyses, a well-designed NIR system will correct the equations automatically and produce a printed report. This validates the on-line NIR analyzer according to ASTM D-3764.

Protofuels provide a high-quality assurance (QA). Under the best conditions, contaminants may deposit to some extent on the probe, the lamp sources may degrade slightly, and the instrument electronics or optical system may change over time. Validation with protofuels will alert a ''smart'' system and allow it automatically to correct if any of these factors start to have an effect.

Section 9.1 of this article covers sampling techniques from the viewpoint of the regulatory agencies.

8.2.5 Property Measurement Procedures – Sampling

The most common approach to component property measurement is to sample the component tank, measure the properties, and enter the values in a data table. This can take a day or more and occurs weekly or less in most plants. At least one refinery, however, samples and tests its component tanks daily.

A second less-common approach is to use results from the streams as they are released from the process units. These results are then averaged or filtered for use in blending. This process makes double use of unit analyzers or routine samples, but the tests frequently exclude important properties such as octane.

If done less often than daily, the tank-sampling method assumes that component properties are reasonably steady over several days. Both approaches assume tanks are well mixed. Unfortunately, these assumptions frequently are invalid. Component tanks are often active and usually stratified. An active tank is one into which current production goes at the same time as the blend component is drawn off. Tank stratification typically results from a minimal physical size of the tanks and the common practice of putting inlet and outlet nozzles near each other.

Properties of real components vary constantly. This variability comes from feed changes, unit upsets, day-to-night temperature swings affecting overhead condensers, and the thousand natural shocks to which units are subject.

The reason these assumptions are used is that realistic assumptions do not yield a useful calculation method. Most refiners, therefore, have decided that it is better to have useful numbers based on poor assumptions than no numbers.

8.2.6 Property Measurement Procedures – On-line Measurement

The on-line measurement of component properties has long been recognized as desirable, but up to the late 1980s the costs made it impractical for most refiners. The most important property traditionally was octane, the measurement of which with on-line knock engines is slow and capital- and labor-intensive. This provided the original impetus to install on-line control in order to reduce expensive ''octane giveaway''.

The other most critical economic qualities, RVP and distillation, were less expensive to measure on-line, but the payback for these was not as worthwhile either, until the advent of regulatory controls. US EPA Phase I and more certainly Phase II and CARB regulations have completely turned the economics of blending upside down. Now the primary factor is to minimize ''RVP giveaway'' and maximize the amount of heavier naphtha, while still meeting T90 specifications. An octane-barrel is no longer worth what it used to be and is now on last priority after the environmental regulations. While this is the way it should be for a serious ''smog-fighting'' environment, it is also one of the many reasons why reformulated fuels cost more than their petroleum-derived ancestors.

8.2.7 Other Potential Near-infrared Analysis Problems

NIR equations often ''split hairs'' to distinguish between similar-looking spectra. This high sensitivity to subtle differences between spectra means that NIR regression equations can also be sensitive to subtle differences in instrument response. Any instrument instabilities, such as slight drifts or jumps in an NIR instrument's response, can result in unacceptable deviations in the NIR predictions.

Similarly, if there is even a small difference in response between two instruments, the NIR equation developed on the first instrument will probably not work on the second without modification (i.e. calibration transfer), even if both instruments are the same make and model.

Some instrument vendors claim to have resolved the calibration transfer issue in their software or hardware. Others just limit their guarantees to certain systems and ranges of composition. Just as is the case with model generality, the simpler analysis applications are less affected by the issues of instrument stability and calibration transfer.

8.3 On-line Analysis for Blender Control – Gas Chromatography Methods

In order to satisfy a refinery's quality and regulatory control requirements, on-line analysis of several parameters are required. Perhaps the most versatile methodology for this is GC. Much time and effort has gone into this methodology for measurement of several parameters that it is necessary to know with regard to gasoline blends. The results generally show good correlation between the analyzers and the more traditional laboratory analysis methods mentioned previously in this article.

Trends in blending call for specialized gas chromatographs in gasoline and diesel blending for such qualities as V/L (vapor liquid ratio), correlated RVP, and simulated distillation correlation of ASTM D-2887 or D3710 to ASTM D-86. The correlation provided by the GC analyzer's manufacturer is tuned against laboratory results to provide continuous monitoring and evaluation for control and maintenance of ''on-spec'' blends. This improved control reduces component giveaway based on the analyzer's repeatability and accuracy. The readings from the such analyzers are consistently within 0.1–0.2 psi (0.7 to 1.38 kPa) for most gasoline grades, premium, midgrade, and regular unleaded gasoline.

With the new, more effective RVP analyzers in place, and with the addition of the octane engine and simulated distillation analyzers, gasoline properties, including octane measurements, RVP, distillation, and vapor-over-liquid ratio, can also be closed-loop controlled. As a result, gasoline blending now can be implemented with multivariable/optimization controls to capture additional benefits.

Other GC analyzers have been used with varying success in optimization of blend characteristics such as total aromatics, total sulfur, and oxygenates and benzenes to measure octane ratings.

8.3.1 Discrete Component Analyses

The GC/PIANO (paraffins, isoparaffins, aromatics, naphthenes, olefins) technique is the primary test method used for the determination of benzene, aromatics, and olefins contents of the blended gasolines. The method incorporates state-of-the-art, high-resolution GC.

The sample is injected into a capillary column, which is temperature-programmed, then separated into its components (possibly 400 or more compounds.)

The data-acquisition system and software integrate the area under each peak and normalize the data, after adjusting for detector response factors for each compound. Each peak is then grouped into one of the five Piano hydrocarbon types. Finally, a detailed list of the hydrocarbon groups and each component, in weight percent and volume percent, is produced.

8.3.2 Evaluating Distillation Using Gas Chromatography-simulated Distillation Methods

To meet the strict environmental and performance specifications of US EPA/CARB-type gasolines in blend operations, while maintaining or even improving profit margins, accurate determination of the distillation assay is essential. Refiners use D86 T50 and T90 values to set blend run operating parameters and check results. Although it is possible to determine D86 through physical distillation in the laboratory, these methods are time-consuming. One analysis can take from 2 to 6 h. In addition, physical distillation methods require at several milliliters of the product. Although it has been tried in various refining environments, on-line distillation analyzers are clumsy, difficult to maintain and calibrate, and often inaccurate. Simulated distillation provides accurate information on blend components and can be packaged in a manner conducive to continuous real-time analysis. Simulated distillations are GC determinations of the bp range of the components in a sample mixture. ASTM D-86 is the method to which most if not all gasoline is sold. While ASTM D-86 is not the only laboratory distillation method, it is the most common, since it is the easiest to perform.

Two ASTM methods, D-2887 and D-3710, described standardized GC to perform a simulated distillation assay. GC simulated distillation methods offer better distillation results and accurate information on the boiling range distribution. They also offer several operational advantages over physical ASTM methods for D86 and D1160 in the laboratory. Unless the analyst using laboratory-based

physical distillation methods follows the ASTM procedures rigorously and his or her equipment meets the test standards, the test results are too often inconsistent.

It is possible to compare the results of the various methods for determining distillation curves. ASTM method D-2892 obtains TBP results using a laboratory distillation unit with 15 theoretical plates. Simulated distillation results closely match the results obtained from ASTM D-2892. Simulated distillation and ASTM D-86 also give similar results for bps less than 400 °C.

8.3.3 On-line Reid Vapor Pressure Analysis

US EPA Phase I and II regulatory compliance needs, along with the addition of oxygenates and other local regulatory requirements, have redefined gasoline blending rules and left refineries vulnerable to millions of dollars in additional cost. Precise on-line analyzer measurement and closed-loop control of key gasoline specifications, such as RVP, are critical to efficient cost-effective blending under these circumstances.

Accurate and repeatable RVP measurements allow more low-value stocks (such as butane) to be used in the blending, without switching to more expensive blending components. Butanes are readily available, economical, and increase fuel vapor pressure as well as octane. Typically, every 0.1 psi closer to the targeted blend RVP value represents an annual saving of up to $400 000 based on 50 000 BBI/day blended.

Blended gasoline specifications must be measured either in the laboratory or on-line, where process analyzers play a key role. While other refinery processes can be closed-loop controlled using inferential methods, gasoline blending closed-loop control relies totally upon on-line analyzers. Therefore, both analyzer accuracy and reliability are critical in order to control RVP successfully with a closed-loop.

Historically, blenders used continuous mechanical (Hallikainen-type) RVP analyzers. These required intensive maintenance because of their moving parts and, despite modifications to operating procedures, and preventative maintenance, their accuracy and reliability of analytical service was still poor. The modifications included shutting down analyzers when not in use and allowing analyzers from each header to service the others when necessary in order to increase reliability.

There were significant differences between analyzer readings and laboratory results for the gasoline grab samples. There was a bias adjustment for each of the old analyzers, but because of the differences in readings, the technician had to change the bias constantly, making the closed-loop RVP control scheme difficult at best.

The on-line time of the mechanical RVP closed-loop controls was poor, usually below 50%. Because of the uncertainty of the on-line analyzer readings, the rate of RVP giveaway was typically and uneconomically high. Several GC-based analyzers and a few vapor pressure flow-cell analyzers went on the market with the necessary characteristics. A quality on-line RVP analyzer will have the following key qualifications:

1. Its repeatability is 0.005 psi and its reproducibility is 0.15 psi vs an accuracy of 0.3 psi for the old analyzer. Inferential predictive methods are no better than this either.
2. It stays on stream, with little or no maintenance.
3. Its cycle time is less than 20 min.
4. It does not require special calibration beyond standard GC methodology.
5. It incorporates up-to-date microprocessor-based technology.

The RVP controllers will typically take the analyzer signals and adjust butane or light straight run naphtha rates to meet RVP targets or set points. To ensure smooth continuous control, inferential dead-time compensated controllers should be used to treat the discrete RVP analyzer signal. The controllers are carefully tuned based on the analyzer cycle time, the response of the fast sample loop, and the system dynamic.

Because the cycle time of the RVP analyzer is about 8–20 min, the response of the RVP controller is adequate to handle most blending disturbances. Typically, the controller can control RVP within 0.1–0.2 psi (0.7 to 1.38 kPa) of the set point, as illustrated in the above graphs. The newer RVP analyzers also provide a safety check to detect analyzer failure. This is done by hardwiring the analyzer trouble alarm directly to the distributed control system (DCS). It replaces the jump check and the stagnant check to detect bad RVP signals.

Since these GC-based RVP analyzers have been installed in refineries, they have usually performed reliably. Typical problems include such incidences as leaking valves and fittings, loss of air supply to the analyzers, and so on. Required repairs after startup infancy troubles are minimal compared to the previous mechanical systems. Analyzer repeatability is good. In order continually to prove the accuracy of the analyzers, high-purity but readily and economically available reference components such as 2,3-dimethylbutane (RVP 7.8 psi) and 2,2-dimethylbutane (RVP 9.8 psi), and normal pentane (RVP 15.4 psi) should occasionally be used to check analyzer repeatability and reproducibility. Also a Graebner RVP analyzer is used to check against on-line RVP results. This is an important check because most regulatory agencies certify or require certification of results with Graebner-type analyzer measurements on the grab samples.

9 MAINTAINING QUALITY CONTROL OF ON-LINE ANALYSES

9.1 Blend Certification Sampling and Testing Procedures

Samples that accurately represent batch properties are necessary in order to determine if regulatory standards are being met. Several book references and standards describe how to do sampling properly so a general treatise on this topic will not be pursued here. It is useful, however, to consider how the regulatory agencies prefer to specify testing procedures.

CARB regulations state[77] that ''each refiner or importer shall determine the value of each of the [RFG] properties for each batch of RFG it produces or imports prior to the gasoline leaving the refinery or import facility, by collecting and analyzing a representative sample of gasoline taken from the batch.'' ''Batch of reformulated gasoline'' is defined at as ''a quantity of reformulated gasoline which is homogeneous with regard to those properties which are specified for reformulated gasoline certification.'' Therefore, the first concern of batch sampling is to determine whether or not the tank contents are homogeneous.

Gravity analyses of upper, middle, and lower samples are a suggested means of establishing tank homogeneity. US EPA would consider a tank to be homogeneous where the maximum difference in tested gravities between any two samples from different tank strata is no greater than 0.6 API, unless there is reason to believe the tank contents are not mixed in spite of such gravity test results. For example, if samples from a storage tank have noticeably different colors, the gasoline in the tank should not be considered homogeneous even if the samples have gravity tests that are within the 0.6 API range. If a question remains about whether the contents of a storage tank are fully mixed following gravity testing, the party could resolve the homogeneity issue by conducting tests on the upper, middle and lower tank samples for benzene and oxygen. Tank homogeneity could be established using benzene and oxygen tests on upper, middle and lower tank samples without the need for gravity testing. The US EPA would consider a tank to be homogeneous if the maximum difference in benzene tests is 0.10 vol% and the maximum difference in oxygen tests is 0.15 wt%. The benzene and oxygen testing to establish homogeneity (as opposed to certification testing) could use a nonregulatory method such as mid-IR analysis.

In cases where the tank contents are not homogeneous, further in-tank or reblend mixing should be performed before collecting a representative sample for RFG analysis. Product stratification should also be avoided downstream of refiner or importer facilities, because samples must meet the downstream ''per gallon'' standards, and stratification could result in a portion of the gasoline in a tank being out of compliance with ''per gallon'' standards.

The US EPA mandates that storage tanks should be sampled[78] using the method that will best represent the contents of the tank or batch. The refiner, importer, or independent laboratory should use its best professional judgment in determining the procedures that are necessary in order best to represent a given batch within the guidelines of its Appendix D, which contains general instructions and precautions that must be followed when choosing sampling equipment and containers, and when collecting samples.

US EPA preference for sampling storage tanks is a ''running'' or ''all-levels'' sample collected from an unconfined (no gage tube) roof port. A ''running'' or ''all levels'' sample collected from a perforated gage tube is the next best choice. In no case should a sample be collected from a solid gage tube.

US EPA prefers to collect ''running'' samples as opposed to ''all-levels'' samples for two reasons. First, assuming that both ''all-levels'' and ''running'' samples are collected with uniform lowering and retrieval rates, the ''running'' procedure achieves better representation of the tank contents than the ''all-levels'' procedure. This occurs because with the ''running'' procedure, one half of the sample is collected when lowering the apparatus, and the column sampled is undisturbed at that point. The second reason is that ''running'' samples are easier to collect than ''all-levels'' samples because the sample collector is not required to stopper the sample bottle.

If a tank cannot be bottle sampled from the top, then tap sampling is an appropriate substitute. For best representation, a single composite should be collected by proportionally filling the sample container from all available taps. If homogeneity is well documented, the entire sample may be collected from a single tap. If a refinery or importer tank has no roof sampling port or sampling taps, then a pipeline sample is the only other sampling means that is possible.

CARB inspectors obtain gasoline samples using the sampling procedure in section 2296, Motor Fuel Sampling Procedures. Generally, ''running samples'' or ''all-levels samples'' will be obtained from refinery and terminal tanks that have appropriate access from a sample hatch. ''Tap samples'' will be obtained from those tanks that do not have hatches. ''Nozzle samples'' will be obtained from service station nozzles. As to how many samples will be obtained, the CARB inspectors generally obtains two samples, one to be used for RVP analysis and the other for all the other fuel analysis.

In the case of downstream QA sampling from a storage tank which does not have a roof sampling port or taps for

sampling, a sample collected from a truck or barge that has just loaded from that tank is marginally acceptable. The truck or barge should be completely empty before loading, and a ''running'' sample should be collected from the truck or barge compartment.

RVP is the most sensitive RFG property, relative to sampling, and therefore precautions to prevent loss of ''light ends'' must be followed carefully. Also, sampling containers must be clean and rinsed well with the gasoline to be sampled, so that the sample is not contaminated, for example, with trace amounts of heavy metals. When collecting tap samples, the tap and connecting piping must be completely flushed, and the sample container must be bottom-filled strictly according to the procedure outlined in Appendix D or section 2296 depending upon the collection locality. In all cases, the sampler should label the container as soon as possible and note the location of the sampling point and method of collection. This may save several headaches at a later time if there are any questions about the measurement value or methodology.

As far as generally acceptable practices for in-line or ''pipeline'' samplings go, ASTM D-4177, *Standard Practice for Automatic Sampling of Petroleum and Petroleum Products* is probably the most applicable guidance for sampling methodology. This practice covers information for the design, installation, testing, and operation of automated equipment for the extraction of representative samples of petroleum and petroleum products from a flowing stream and storing them in a sample receiver. If sampling is for the precise determination of volatility, use Practice D5842 in conjunction with this practice. For sample mixing, refer to Practice D5854. Petroleum products covered in this practice are considered to be a single phase and exhibit Newtonian characteristics at the point of sampling.

D4177 is applicable to petroleum and petroleum products with vapor pressures at sampling and storage temperatures less than or equal to 101 kPa (14.7 psi). Refer to D5842 when sampling for RVP determination. Petroleum products whose vapor pressure at sampling and sample storage condition is above 101 kPa (14.7 psi), and liquefied gases [that is, liquid natural gas (LNG), liquefied petroleum gas (LPG) etc.], are not covered by this practice.

While the procedures covered by this practice will produce a representative sample of the flowing liquid into the sample receiver, specialized sample handling may be necessary to maintain sample integrity of more volatile materials at high temperatures or extended residence time in the receiver. Such handling requirements and procedures for sampling these fluids are described in Practice D1265, Test Method D1145, and Gas Processors' Association (GPA) 2166.

Annex A2 of 4177 contains theoretical calculations for selecting the sampler location. Annex A3 lists acceptance methodologies for sampling systems and components. Annex A4 gives performance criteria for permanent installations, while Annex A5 has the criteria for portable sampling units. Appendix X1 is a design data sheet for automatic sampling systems; Appendix X2 compares the percent sediment and water to unloading time period.

Other applicable standards for sampling and measurement set-up are:

- API-DR-214 – Sampling and Analytical Methods for Determining Petroleum Hydrocarbons in Gasoline
- API-MPMS-C13-S1 – Chapter 13 – Statistical Aspects of Measuring and Sampling, Section 1
- API-MPMS-C13-S2 – Chapter 13 – Statistical Aspects of Measuring and Sampling, Section 2
- ISO/DTR 9494 Petroleum Liquids – Automatic Pipeline Sampling – Statistical Study of Performance (Future Type 3 Technical Report)
- ISO 3170 Petroleum Liquids – Manual Sampling
- ISO 3171 Petroleum Liquids – Automatic Pipeline Sampling
- ISO 1998-1: 1974 Petroleum Industry – Vocabulary – Part 1
- ISO 1998-2: 1976 Petroleum Industry – Vocabulary – Part 2.

9.2 Infrared Analysis Preparation[73,79]

9.2.1 Chemometric Calibrations

Chemometric calibration techniques require substantial amounts of data for many calibrations. If this calibration data contains unknown or uncontrolled sample and spectrometer variations, the calibration requires more data with all of these variations included. Such a result may not be robust since it becomes specific, not only to the stream being measured, but also to the specific spectrometer and to the other uncontrolled variations in the system. A better calibration method is to reduce or eliminate as many sources of noise as possible. Reducing sources of variation makes the calibration simpler and more robust. The calibration models can benefit from good sample conditioning and the use of stable, properly functioning spectroscopic hardware. This fact is often overlooked at the chemometric calibration stage because techniques such as partial least squares regression (PLS) can ''fit'' noise. This does not mean, however, that the calibration models can predict noise, or deal with ''new'' noise.[50]

Systematic noise such as temperature variation or spectrometer wavelength drift can be partially modeled away by including the temperature variation or wavelength shifting in the calibration set. However, this greatly increases the size and cost of the calibration dataset.

Further, compensating for system noise in this way adds variation into the calibration dataset unrelated to the analyte. This noise in the data can mask and degrade the information of interest in spectrum, and ultimately make the analysis impossible. These issues will be discussed in detail below.

The preferred philosophy is to control as many potential sources of sample-related variations as possible. This means controlling sample, temperature, pressure, flow, particulate, and undesired phases. The sample system is used to facilitate automated true referencing and zeroing of the spectrometer. Also required is the use of a robust and well-designed spectroscopic analyzer that minimizes errors in the data due to wavelength shifts, spectral distortions, stray light, light intensity changes, or nonlinearities.

9.2.2 Sample Filtering

Sample filtering can also be crucial to quality spectroscopy. Particulate in the sample will cause scattering in the spectrum that may lead to curved, nonreproducible spectral backgrounds. A typical means of removing particulates is via centrifugal type filters with a 5–10 μm filter element. An emulsion such as water in a hydrocarbon can cause spectral scattering. This water scattering is particularly prevalent in hydrocarbons in the second and third NIR overtone regions (800–2000 nm). Water can be removed with a centrifugal-type filter. Because water is so insoluble in hydrocarbons, even a few hundred parts per million can cause scattering.

Another use for a sample system is to remove bubbles in the sample. As long as bubbles separate easily from the sample, they can be removed simply by stopping sample flow in the sample cell and allowing the bubbles to float to the top of the cell. Stop flow can also be used to eliminate the possibility of turbulent flow in the sample cell that can cause scattering due to density gradients. Turbulent flow is particularly likely in cells or probes with a path length greater than 1 mm.

It is important in sample system design that the system itself does not create bubbles. In a high volatility sample like gasoline, bubbles can be formed by a pressure drop in the system itself. Therefore flow and low-pressure control valves should be located after the sample cell. Bubbles can also be eliminated by assuring that liquid sample flow is up into the cell. This takes advantage of gravity, so that if a bubble is formed it will flow harmlessly up and out of the cell.

9.2.3 Spectrometer Maintenance

A well-designed fluid switching system can be advantageously used to maintain and monitor optical system and spectrometer performance. This allows reference fluids, protofuels, and wash solvents, as well as dry air or nitrogen to be switched into the sample cell. These functions will be required in a pending ASTM method for validation of NIR spectroscopic analyzers using chemometric-based calibrations.

Most process spectrometers designed for monitoring liquids and gases are ''single beam'' instruments. If a second reference path exists, the reference path is seldom the same as the sample path. The raw single beam spectrum obtained contains information about the sample in the sample cell and the wavelength-dependent response of the spectrometer and sample cell. To remove the effect of the spectrometer response, a spectrometer reference spectrum is required. The spectrum is ratioed against the single beam spectrum to give a spectrometer-independent spectrum. This reference spectrum is ideally a spectrum taken through the empty sample cell. It can be expected that this reference spectrum will change with time. Some changes that can affect the reference include contamination of the sample cell or ageing of the spectrometer light source. On a typical, well-stabilized spectrometer a new reference spectrum is generally recommended once every 8–24 h. This time may vary depending on the sample and the sensitivity of the calibration equations to cell contamination. Obtaining such a new reference is straightforward with a fluid switching system in place and can be an automated scheduled event that requires no manual intervention.

Obtaining a reference spectrum is accompanied by a cell wash cycle to remove sample buildup on the sample cell windows.

A typical wash and reference sequence for an on-line gasoline spectrometer follows:

1. Flush cell with toluene wash solvent – sending the toluene to the process, a flare line, or hydrocarbon waste sewer
2. Flush with pentane – other solvents may be used as long as they are volatile
3. Reverse flow through cell to top-to-bottom
4. Flush with dry nitrogen or air
5. Take the reference spectrum
6. Run diagnostics on the spectrum – if they pass, use the spectrum as the new reference, if not, alarm and keep old reference
7. Return to analysis mode.

If the spectroscopic analyzer does not have a sample system, then the sample cell or probe must be removed to be cleaned or for a reference spectrum to be obtained. Other approaches using a second optical path, such as a short-fiber loop or a dual optical path probe, exist for obtaining a reference spectrum. These approaches are all limited in that they do not take into account all the changes in the system. In particular they miss possible contamination of the sample cell or cell windows.

Depending on the configuration there are usually other assumptions as well, such as assuming that two optically different optical paths have the identical effect on the spectrum.

9.2.4 Spectrometer and Property Calibration Diagnostics

A sample system allows diagnostics of the spectrometer system itself and the property calibration equations. Such diagnostics as cell cleanliness, detector linearity, light source intensity, optical alignment, signal-to-noise, stability, and optical purge quality can all be determined automatically when a reference is taken. These tests assure that the results will be of high quality.

A sample system also allows introduction of a calibration material into the sample cell. A pure compound can be used for wavelength and cell path length checks. A measure of the overall system performance can be obtained by introducing a known sample such as a protofuel and checking the PLS prediction results.

A typical calibration cycle would be as follows:

1. execute reference cycle
2. introduce calibration fluid
3. acquire spectrum
4. run diagnostics, archive results, alarm if indicated
5. return to analysis.

9.2.5 Extractive Sampling

For gases and low-viscosity liquids, an extractive sample system provides a better quality process spectroscopic analysis. An extractive system assures spectral quality and reproducibility and a clean thermostatted sample. An extractive system allows for cost-effective sample multiplexing and automated system diagnostics. Perhaps the feature that is most important to the analysis is that such a system requires minimal manual maintenance.

Extractive sampling allows the use of an optically simple transmission cell. Pure transmission sampling is the preferred method for calibrations because of its high degree of reproducibility from cell to cell. Such a cell typically has windows sufficiently large so as not to shade off the beam. Any sample spectrum distortion due to the cell is minimized. This is crucial to eliminate the adjustment of the PLS calibrations if a cell is replaced or disassembled for cleaning. If transmission sampling is used it is possible for a technician to replace a cell without any adjustments or impact on the calibration.

Extractive sampling is a cost-effective means of sample multiplexing as compared with optical multiplexing. The use of optical multiplexing does not eliminate the need for sample conditioning. Instead, it requires multiple distributed sample conditioning systems. Multiplexing of 1–3 sample streams is common with extractive systems, and many more are possible.

There are cases where extractive sampling is not possible. This may be due to safety, sample viscosity, or very fast cycle time. If so, the preferred approach is a local extractive sample system with a fiber-optic flow-through cell. This will provide all of the sample control of other extractive systems. The use of fiber optics introduces new optical elements into the system. The characteristics of the fiber probe, optical multiplexer, fiber and fiber–spectrometer interface must be carefully controlled to assure spectral reproducibility.

9.3 Calibration and Acceptance Testing of Gas Chromatography/Simulated Distillation Analyzers for D86/Reid Vapor Pressure Analysis

Careful calibration of on-line analyzers is absolutely essential to the success of real-time fuels blending control. Performance of factory and in-field acceptance testing is a critical part of any such analyzer installation.

As part of in-field or factory acceptance tests (FAT) it is not unusual for GC/simulated distillation analyzers to behave inconsistently and show excessive deviation from laboratory data [as much as 100–150 °F (38–66 °C)]. Consequently, such analyzers are often shipped back to the manufacturer, corrected, and their performance verified using additional samples supplied by the refinery laboratory. These samples included all the sample types projected for analysis on these analyzers. Owing to time limitations, not all samples may be run on every analyzer but sufficient data was collected to be confident that the analyzers will be able to perform the required analyses with the expected precision.

Two common problems often arise in gasoline analyzer configuration. A presplitter or stabilizer flash cell may need to be added to each analyzer to prevent overloading the (typically FID) detector with samples containing high C_5 concentrations. Flow sensor placement may also present problems. Typically such sensors should be placed on the outlet side of the sample valve unless pressure relief line or cabinet space requirements dictate otherwise. Inlet flow detectors are often a source of cross-contamination of successive samples.

Acceptance criteria for the correlated D-86 data should be based upon the reproducibility of the D-86 method [15 °F/8.3 °C for initial boiling point (IBP), 19 °F/10.5 °C for FBP and 7–12 °F/3.9–6.7 °C for 10–90% distillation points]. Acceptance for computed RVP numbers is usually 0.2 psi.

Once the factory test has been validated to the proper standard, initial biases of the data can be established. Biases are a set of analyzer offsets that allow the correlation of the analyzer data with laboratory data at each reported distillation and RVP data point. These

Table 14 Typical results for test samples analyzed by GC at acceptance tests

	IBP	10	20	30	50	70	80	90	FBP	RVP
Unleaded super #1										
Lab data	87	136	155	172	207	244	268	308	406	7.44
GC no. 1	107	134	151	168	201	239	269	317	395	7.48
GC no. 2	105	133	151	169	201	239	269	317	394	7.44
GC no. 3	106	134	152	170	202	239	270	317	393	7.51
Unleaded regular #1										
Lab data	96	131	147	164	208	266	305	341	420	7.43
GC no. 1	108	134	150	168	201	277	313	347	424	7.35
GC no. 2	105	133	150	168	201	277	313	348	425	7.37
GC no. 3	107	134	150	169	206	277	312	347	423	7.47
Unleaded super #2										
Lab data	97	140	158	175	211	249	276	315	393	7.4
GC no. 1	104	134	151	169	201	239	270	323	392	7.34
GC no. 2	101	133	151	169	201	240	270	322	392	7.33
GC no. 3	102	134	152	170	203	241	270	321	391	7.36
Unleaded regular #2										
Lab data	97	136	150	166	210	275	309	342	418	7.4
GC no. 1	108	134	150	168	201	277	313	347	424	7.35
GC no. 2	105	133	150	168	201	277	313	348	425	7.37
GC no. 3	107	134	150	169	206	277	312	347	423	7.47
Unicrackate stock										
Lab data	101	121	126	132	146	169	180	192	233	9.39
GC no. 1	124	115	112	116	143	165	183	194	236	10.04
Reformate										
Lab data	113	207	233	251	276	301	316	338	427	2.48
GC no. 2	172	216	235	253	279	310	326	347	421	2.50

are utilized in normal on-line analyzer operation to improve the accuracy of the analyzer data (relative to laboratory values). During normal operations, these factors should occasionally (3–6 months) be revalidated as a part of normal quality assurance (QA) standards work.

Table 14 shows an initial field acceptance test result matrix for three analyzers which failed QA standards. They were returned to the manufacturer for rework and a subsequent retest at the manufacturer's location.[80]

Tables 15 through 17 show the comparison of laboratory results with the GC analyzer distillation and RVP results for the second QA test samples. Also compiled are analyzer computed $V/L20$ data which were included in the analyzer specification as output data but were not measured by the refinery laboratory. These data, as well as the Table 14 data, are shown without any bias offsets of the data, since the statistical analysis must be done on the actual data taken by the instrument. The precision of the GC analyzer data, as measured by the SDs, is better than the precision of the laboratory data for both the distillation points and the RVP results. The pooled standard deviation for the distillation points (averaged over all equipment, samples and distillation points) was 1.1 °F/0.6 °C for the GC analyzers and 3.2 °F/1.8 °C for the laboratory distillation analyzers. For RVP, the pooled SD for the GC analyzers was 0.05 psi compared to 0.22 psi for the laboratory data. Agreement between the GC analyzer data and the laboratory distillation data was acceptable, with a few peculiar exceptions. The 10% and 20% points on the regular gasoline sample were low by 18 °F/10 °C and 13 °F/7.2 °C, respectively. In the front end of the hydrocrackate sample, the IBP was 28 °F/15.6 °C high while the 10–30% points were low. The IBP for the reformate sample was 58 °F/32.2 °C high. Of most concern was the hydrocrackate data, which showed a minimum value in the temperature vs volume% curve around 20% distilled. This appears to be due to the high C_5 content of the sample, which (according to the GC data) comprises about 40% of the sample. For RVP, agreement between the GC data and laboratory data was acceptable, with the exception of the hydrocrackate sample (probably, also due to the high C_5 content) and, marginally, the regular blend sample (0.23 psi difference probably due to the low front-end distillation points).

Fortunately, for the hydrocrackate and reformate samples (those samples with the most serious discrepancy between GC and laboratory), analytical data was

Table 15 Comparison of retest laboratory data and GC analyzer distillation and RVP data

	IBP	10	20	30	50	70	80	90	FBP	RVP	°F	Time (s)
Premium												
Lab D86	98	127	151	170	221	276	300	327	415	7.46		
	91	132	154	173	224	279	301	325	415	7.43		
	104	136	156	175	223	277	301	326	415	7.46		
Lab average	98	132	154	173	223	277	301	326	415	7.45		
Lab SD	6.51	4.5	2.5	2.5	1.5	1.5	0.5	1.0	0.0	0.02		
GC no. 1	107	134	152	177	230	278	310	340	415	7.35	150	112
	106	134	152	177	230	278	310	337	415	7.37	150	113
	106	134	152	176	230	278	309	337	413	7.37	150	115
	106	134	152	176	230	278	309	337	413	7.37	150	120
	107	134	152	177	230	278	309	337	413	7.34	151	122
GC no. 2	107	134	151	176	230	278	310	338	417	7.38	150	102
	107	134	151	176	230	278	310	338	417	7.37	150	103
GC no. 3	107	134	153	177	229	277	308	337	415	7.37	150	152*
GC average	107	134	152	177	230	278	309	338	415	7.37	150	
GC SD	0.52	0.0	0.6	0.5	0.3	0.3	0.7	1.0	1.6	0.01	0.35	
Difference (GC-Lab)	9	2	−2	4	7	1	8	12	0	−0.08		
Regular												
Lab data	103	146	158	175	216	266	303	347	417	7.50		
	96	141	159	174	203	257	292	335	416	7.48		
	106	141	157	173	214	260	293	334	415	7.50		
Lab average	102	143	158	174	211	261	296	339	416	7.49		
Lab SD	5.13	2.8	1.0	1.0	7.0	4.5	6.0	7.2	1.0	0.01		
GC no. 1	104	126	146	164	201	265	298	343	428	7.76	142	184*
	105	125	145	163	201	266	299	343	428	7.75	142	907
	105	124	144	162	201	265	299	344	428	7.76	142	922
	105	125	145	162	201	265	299	344	428	7.75	142	937
GC no. 2	105	124	143	161	202	265	298	344	428	7.67	142	112
	105	124	144	162	202	267	300	344	429	7.66	142	114
	105	124	144	162	203	270	303	344	428	7.61	143	115
	105	124	144	161	202	266	299	344	428	7.66	142	121
	104	123	144	161	202	266	299	344	428	7.66	142	122
GC no. 3	105	130	148	166	200	267	299	342	428	7.82	143	101
	105	130	148	166	201	268	300	343	429	7.78	143	103
GC average	105	125	145	163	201	266	299	344	428	7.72	142	
GC SD	0.40	2.4	1.6	1.8	0.8	1.5	1.3	0.6	0.4	0.07	0.47	
Difference (GC-Lab)	3	−18	−13	−11	−10	5	3	5	12	0.23		

SD, standard deviation.

available for samples collected from those units over a period of time spanning about one month). This allowed a measurement of the applicability of the biases to correct the difference between the laboratory and GC data. For these samples, the bias estimations and the resulting corrected data are shown in Table 18. Here, biases are computed from the first test samples and are applied to the retest samples. These corrections produced acceptable data for all the distillation points and the RVP. It was particularly encouraging that the initial test and retest hydrocrackate samples had different bp curves and vapor pressures (indicating some change in sample composition over that period) but the bias corrections were applicable in the face of the composition change. For each sample, sets of biases will need to be determined in the field and periodically checked.

Prior data for the regular gasoline was not available to test the efficacy of the correction on that sample. However, the same deviations in distillation points and RVP were not observed in the regular initial test sample. The selection of bias corrections will represent a compromise over the range of compositions typical for a particular product or blend stock, and more data will have to be collected to determine the appropriate offsets for the regular gasoline product. Acceptable agreement could be obtained for the regular gasolines by averaging the offsets for the two regular gasolines. Nevertheless, the deviations observed in the uncorrected regular blend data could not be attributed to analyzer error.

With these results, the distillation analyzers were judged to be performing with sufficient accuracy and precision to justify acceptance of the units.

Table 16 Typical comparison of laboratory and GC analyzer distillation and RVP data

Unicrackate (°F)	IBP	10	20	30	50	70	80	90	FBP	RVP	°F	Time (s)
Lab data	89	117	123	128	145	174	188	202	228	9.69		
	95	119	124	130	147	176	190	203	222	10.2		
	90	117	123	128	145	171	185	199	219	10.3		
Lab average	91	118	123	129	146	174	188	201	223	10.1		
Lab SD	3.21	1.1	0.5	1.1	1.1	2.5	2.5	2.0	4.5	0.36		
GC-410	118	110	102	108	141	167	188	202	236	10.9	116	1052
	120	112	103	109	143	167	188	202	237	10.8	117	1107
GC-411	120	112	103	109	142	168	189	202	235	10.9	116	936
	120	112	103	109	142	168	189	202	235	10.9	116	952
	120	112	104	109	142	168	189	202	235	10.9	116	1007
GC-450	117	109	101	107	140	166	187	202	235	11.0	115	1804*
	118	110	103	108	140	167	188	202	234	11.1	115	859
	119	111	103	109	140	167	189	202	235	11.1	115	914
GC average	119	111	103	109	141	167	188	202	235	11.0	116	
GC SD	1.20	1.2	0.8	0.7	1.1	0.7	0.7	0.0	0.8	0.11	0.71	
Difference (GC-lab)	28	−7	−20	−20	−5	−7	0	1	12	0.90		
LT cat naphtha												
Lab data	98	129	141	153	177	208	227	261	310	8.14		
	96	129	140	151	175	204	223	251	315	8.67		
	94	130	141	152	176	205	225	254	316	8.88		
Lab average	96	129	141	152	176	206	225	255	314	8.56		
Lab SD	2.00	0.5	0.5	1.0	1.0	2.0	2.0	5.1	3.2	0.38		
GC no. 1	104	122	138	148	167	205	221	257	324	8.54	132	1022
	104	122	138	148	167	205	221	257	324	8.56	132	1037
GC no. 2	103	121	138	147	164	205	221	254	321	8.44	132	1714*
GC no. 3	104	122	138	147	166	205	221	257	321	8.74	131	1104
	104	122	138	148	166	205	221	257	321	8.67	132	1120
	104	122	138	148	166	205	221	257	321	8.62	132	1136
	104	122	138	147	166	205	221	257	321	8.71	132	1152
	104	122	137	147	166	205	221	257	321	8.69	132	1207
	104	122	138	148	166	205	221	257	322	8.65	132	1223
GC average	104	122	138	148	166	205	221	257	322	8.62	132	
GC SD	0.33	0.3	0.3	0.5	0.8	0.0	0.0	1.0	1.3	0.10	0.33	
Difference (GC-Lab)	8	−7	−3	−4	−10	−1	−4	2	8	0.06		

10 REFORMULATED FUELS ANALYSIS GLOSSARY

10.1 Introduction

This glossary is intended to help interested citizens understand the most commonly used air pollution terms. The glossary is divided into three sections:

1. governmental section: these terms apply to the USA unless stated otherwise
2. technical section
3. standards and associations.

10.1.1 Governmental Section

APCD (Air Pollution Control District): a county agency with authority to regulate stationary, indirect, and area sources of air pollution (e.g. power plants, highway construction, and housing developments) within a given county, and governed by a district air pollution control board composed of the elected county officials (compare AQMD).

AQMD (Air Quality Management District): a group of counties or portions of counties, or an individual county specified in law with authority to regulate stationary, indirect, and area sources of air pollution within the region and governed by a regional air pollution control board comprised mostly of elected officials from within the region (compare APCD).

Attainment area: a geographic area which is in compliance with the national and/or California Ambient Air Quality Standards (see NAAQSs or CAAQSs).

BACT (Best Available Control Technology): the most up-to-date methods, systems, techniques, and production processes available to achieve the greatest feasible emission reductions for given regulated air pollutants and processes. BACT is a requirement of NSR (New Source Review) and PSD (Prevention of Significant Deterioration).

Table 17 Comparison of laboratory and GC (D3710) analyzer distillation and RVP data

	IBP	10	20	30	50	70	80	90	FBP	RVP	°F	Time (s)
Reformate												
Lab data	123	215	237	253	279	305	320	341	431	2.09		
	118	207	231	254	280	306	319	343	440	2.15		
	114	212	235	252	279	305	321	344	429	2.10		
Lab average	118	211	234	253	279	305	320	343	433	2.11		
Lab SD	4.51	4.0	3.0	1.0	0.5	0.5	1.0	1.5	5.8	0.03		
GC no. 2	177	222	237	253	279	310	326	348	423	2.18	198	1801*
	172	217	235	253	279	310	326	348	423	2.26	196	906
	178	222	238	253	279	310	326	348	423	2.18	198	921
GC no. 3	177	222	237	253	279	309	325	348	421	2.17	198	930
	177	222	237	253	279	309	325	347	421	2.19	197	946
	177	222	237	252	279	309	325	348	421	2.16	198	1001
GC average	176	221	237	253	279	310	326	348	422	2.19	198	
GC SD	2.16	2.0	0.9	0.4	0.0	0.5	0.5	0.4	1.1	0.04	0.84	
Difference (GC-Lab)	58	10	3	0	0	5	6	5	−11	0.08		
Hydrotreated straight-run naphtha												
Lab data	218	244	249	256	271	294	309	328	383	1.17		
	219	242	247	253	269	292	306	325	373	1.32		
	208	244	249	254	271	294	308	327	373	1.39		
Lab average	215	243	248	254	270	293	308	327	376	1.29		
Lab SD	6.08	1.1	1.1	1.5	1.1	1.1	1.5	1.5	5.7	0.11		
GC no. 2	231	239	245	250	269	296	311	331	381	1.11	204	1053
	231	239	244	250	269	296	311	331	380	1.11	204	1109
GC no. 3	224	237	243	249	269	295	310	330	380	1.19	203	1700*
GC average	229	238	244	250	269	296	311	331	380	1.14	204	
GC SD	4.04	1.1	1.0	0.5	0.0	0.5	0.5	0.5	0.5	0.05	0.58	
Difference (GC-Lab)	14	−5	−4	−4	−1	3	3	4	4	−0.15		

Table 18 Estimation of biases and effect of biases on distillation and RVP results

	IBP	10	20	30	50	70	80	90	FBP	RVP
Reformate initial test										
Lab test	113	207	233	251	276	301	316	338	427	2.48
GC no. 2	172	216	235	253	279	310	326	347	421	2.5
Computed factor	−59	−9	−2	−2	−3	−9	−10	−9	6	−0.02
Retest										
GC-average	176	221	237	253	279	310	326	348	422	2.19
"Bias"	117	212	235	251	276	301	316	339	428	2.17
Lab average	118	211	234	253	279	305	320	343	433	2.11
Difference	−1	1	1	−2	−3	−4	−4	−4	−5	0.06
Unicrackate Initial test										
Lab data	101	121	126	132	146	169	180	192	233	9.39
GC no. 1	124	115	112	116	143	165	183	194	236	10.04
Computed factor	−23	6	14	16	3	4	−3	−2	−3	−0.65
Retest										
GC-average	119	111	103	109	141	167	188	202	235	11.00
"Bias"	96	117	117	125	144	171	185	200	232	10.35
Refinery lab	91	118	123	129	146	174	188	201	223	10.10
Difference	5	−1	−6	−4	−2	−3	−3	−1	9	0.25

CAAQS (California Ambient Air Quality Standard): a legal limit that specifies the maximum level and time of exposure in the outside air for a given air pollutant and which is protective of human health and public welfare (Health and Safety Code 39606b). CAAQSs are recommended by the California Office of Environmental Health Hazard Assessment and adopted into regulation by the Air Resources Board. CAAQSs are the standards which must be met per the requirements of the California Clean Air Act (CCAA).

CARB (California Air Resources Board): the state's lead air quality agency consisting of a nine-member Governor-appointed board. It is responsible for attainment and maintenance of the state and federal air quality standards, and is fully responsible for motor vehicle pollution control. It oversees county and regional air pollution management programs.

CCAA (California Clean Air Act): a California law passed in 1988 which provides the basis for air quality planning and regulation independent of federal regulations. A major element of the Act is the requirement that local APCD/AQMDs in violation of the CAAQS must prepare attainment plans which identify air quality problems, causes, trends, and actions to be taken to attain and maintain California's air quality standards by the earliest practicable date.

Emission Offset (also known as an emission trade-off): a rule-making concept whereby approval of a new or modified stationary source of air pollution is conditional on the reduction of emissions from other existing stationary sources of air pollution. These reductions are required in addition to reductions required by BACT.

Emission Standard: the maximum amount of a pollutant that is allowed to be discharged from a polluting source such as an automobile or smoke stack.

FCAA (Federal Clean Air Act): a federal law passed in 1970 and amended in 1977 and 1990 which forms the basis for the national air pollution control effort. Basic elements of the act include NAAQSs for major air pollutants, air toxics standards, acid rain control measures, and enforcement provisions.

FIP (Federal Implementation Plan): in the absence of an approved State Implementation Plan (SIP), a plan prepared by the US EPA which provides measures that non-attainment areas must take to meet the requirements of the FCAA.

NAAQS (National Ambient Air Quality Standards): standards set by the federal US EPA for the maximum levels of air pollutants which can exist in the outdoor air without unacceptable effects on human health or the public welfare.

Non-attainment Area: a geographic area identified by the US EPA and/or CARB as not meeting either NAAQS or CAAQS standards for a given pollutant.

PSD (Prevention of Significant Deterioration): a program used in development of permits for new or modified industrial facilities in an area that is already in attainment. The intent is to prevent an attainment area from becoming a nonattainment area. This program, like NSR, can require BACT and, if an CAAQS is projected to be exceeded, emission offsets.

SIP (State Implementation Plan): a document prepared by each state describing existing air quality conditions and measures which will be taken to attain and maintain NAAQSs (see AQMD).

TAC (Toxic Air Contaminant): an air pollutant, identified in regulation by the CARB, which may cause or contribute to an increase in deaths or in serious illness, or which may pose a present or potential hazard to human health. TACs are considered under a different regulatory process (California Health and Safety Code section 39650 et seq.) than pollutants subject to CAAQSs. Health effects to TACs may occur at extremely low levels, and it is typically difficult to identify levels of exposure which do not produce adverse health effects.

US EPA (United States Environmental Protection Agency): the United States agency charged with setting policy and guidelines, and carrying out legal mandates for the protection of national interests in environmental resources.

10.1.2 Technical Section-I, General

AAQS (Ambient Air Quality Standards): health and welfare based standards for clean outdoor air which identify the maximum acceptable average concentrations of air pollutants during a specified period of time (see CAAQS and NAAQS in the section 10.1.1, and Criteria Air Pollutant below).

Acute Health Effect: an adverse health effect which occurs over a relatively short period of time (e.g. minutes or hours).

Aerosol: particles of solid or liquid matter that can remain suspended in air for long periods of time because of extremely small size and lightweight.

Air Pollutants: amounts of foreign and/or natural substances occurring in the atmosphere that may result in adverse effects on humans, animals, vegetation, and/or materials.

Air Quality Simulation Model: a mathematical relationship between emissions and air quality which simulates the transport, dispersion, and transformation of compounds emitted to the air.

Air Toxics: a generic term referring to a harmful chemical or group of chemicals in the air. Typically, substances that are especially harmful to health, such as those considered under US EPA's hazardous air pollutant (HAP) program or California's Assembly Bill 1807 TAC program, are considered to be air toxics. Technically, any compound

that is in the air and has the potential to produce adverse health effects is an air toxic.

Alternative Fuels: fuels such as methanol, ethanol, natural gas, and liquid propane gas that are cleaner burning and help to meet CARB's mobile and stationary emission standards.

Ambient Air: the air occurring at a particular time and place outside of structures. Often used interchangeably with outdoor air.

BACT (Best Available Control Technology): the most up-to-date methods, systems, techniques, and production processes available to achieve the greatest feasible emission reductions for given regulated air pollutants and processes. BACT is a requirement of NSR and PSD.

Chronic Health Effect: an adverse health effect which occurs over a relatively long period of time (e.g. months or years).

Criteria Air Pollutant: an air pollutant for which acceptable levels of exposure can be determined and for which an ambient air quality standard has been set. Examples include ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, and PM-10 (Particulate Matter) (see individual pollutant definitions).

Emission Inventory: an estimate of the amount of pollutants emitted from mobile and stationary sources into the atmosphere over a specific period such as a day or a year.

Hazardous Air Pollutant (HAP): an air pollutant considered by US EPA to be particularly hazardous to health. Emission sources of HAPs are identified by US EPA, and emission standards are set accordingly.

Hydrocarbon: any of a large number of compounds containing various combinations of hydrogen and carbon atoms. They may be emitted into the air as a result of fossil fuel combustion, fuel volatilization, and solvent use, and are a major contributor to smog (also see ROG (Reactive Organic Gas)).

Inversion: a layer of warm air in the atmosphere that lies over a layer of cooler air, trapping pollutants.

Mobile Sources: sources of air pollution such as automobiles, motorcycles, trucks, off-road vehicles, boats, and airplanes (contrast with stationary sources).

Nitrogen Oxides (oxides of nitrogen, NO_x): a general term pertaining to compounds of nitric oxide (NO), nitrogen dioxide (NO), and other oxides of nitrogen. Nitrogen oxides are typically created during combustion processes, and are major contributors to smog formation and acid deposition. NO_x is a criteria air pollutant, and may result in numerous adverse health effects.

Ozone: a strong smelling, pale blue, reactive toxic chemical gas consisting of three oxygen atoms. It is a product of the photochemical process involving the sun's energy. Ozone exists in the upper atmosphere ozone layer as well as at the earth's surface. Ozone at the earth's surface causes numerous adverse health effects and is a criteria air pollutant. It is a major component of smog.

Ozone Precursors: chemicals such as hydrocarbons and oxides of nitrogen, occurring either naturally or as a result of human activities, which contribute to the formation of ozone, a major component of smog.

Photochemical Reaction: a term referring to chemical reactions brought about by the light energy of the sun. The reaction of nitrogen oxides with oxygen in the presence of sunlight to form ozone is an example of a photochemical reaction.

PM-10 (Particulate Matter): a major air pollutant consisting of tiny solid or liquid particles of soot, dust, smoke, fumes, and mists. The size of the particles (10 μm or smaller, about 0.0004 inches or less) allows them to enter the air sacs deep in the lungs easily, where they may be deposited and result in adverse health effects. PM-10 also causes visibility reduction and is a criteria air pollutant.

Pollutant Standards Index (PSI): a numerical index used for reporting severity of air pollution. The higher the index, the higher the level of pollutants and the greater likelihood of health effects.

Pollution Prevention: the use of materials, processes, or practices to reduce, minimize, or eliminate the creation of pollutants or wastes. It includes practices that reduce the use of toxic or hazardous materials, energy, water, and/or other resources.

Risk Assessment: an evaluation of risk which estimates the relationship between exposure to a harmful substance and the likelihood that harm will result from that exposure. Risk assessments are generally expressed as the estimated chance per million that a person, exposed over some period of time (e.g. a 70-year lifetime) and some specified concentration of exposure, will experience a certain effect.

ROG (Reactive Organic Gas): a reactive chemical gas, composed of hydrocarbons, that may contribute to the formation of smog. Also sometimes referred to as nonmethane organic compounds (NMOCs).

Sensitive Groups: identifiable subsets of the general population that are at greater risk than the general population to the toxic effects of a specific air pollutant (e.g. infants, asthmatics, elderly).

Smog: a combination of smoke, ozone, hydrocarbons, nitrogen oxides, and other chemically reactive compounds which, under certain conditions of weather and sunlight, may result in a murky brown haze that causes adverse health effects. The primary source of smog in California is motor vehicles.

SO_2 (Sulfur Dioxide): a strong smelling, colorless gas that is formed by the combustion of fossil fuels. Power plants, which may use coal or oil high in sulfur content, can be major sources of SO_2, SO_3 and other sulfur oxides

contribute to the problem of acid deposition. SO_2 is a criteria pollutant.

Vapor Recovery Systems: mechanical systems that collect and recover chemical vapors resulting from transfer of gasoline from operations such as tank-to-truck systems at refineries, tanker-to-pipeline systems at offshore oil operations, and pump-to-vehicle systems at gasoline stations.

VOCs (Volatile Organic Compounds): hydrocarbon compounds which exist in the ambient air. VOCs contribute to the formation of smog and/or may themselves be toxic. VOCs often have an odor, and some examples include gasoline, alcohol, and the solvents used in paints.

10.1.3 Technical Section – II

Additive: any material added to a base stock to change its properties, characteristics or performance.

Aniline Point: the lowest temperature at which equal volumes of aniline and hydrocarbon fuel or lubricant base stock are completely miscible. A measure of the aromatic content of a hydrocarbon blend, used to predict the solvency of a base stock or the CN of a distillate fuel.

Antistatic Additive: an additive that increases the conductivity of a hydrocarbon fuel to hasten the dissipation of electrostatic charges during high-speed dispensing, thereby reducing the fire/explosion hazard.

Base Stock: the base blending material, usually a straight run or refined naphtha fraction, into which additives are blended to produce finished gasolines.

Carbon Residue: coked material remaining after an oil has been exposed to high temperatures under controlled conditions.

Catalytic Converter: an integral part of vehicle emission control systems since 1975. Oxidizing converters remove hydrocarbons and carbon monoxide (CO) from exhaust gases, while reducing converters control nitrogen oxide (NO_x) emissions. Both use noble metal (platinum, palladium or rhodium) catalysts that can be ''poisoned'' by lead compounds in the fuel or lubricant.

Cetane Index: a value calculated from the physical properties of a diesel fuel to predict its CN.

Cetane Number (CN): a measure of the ignition quality of a diesel fuel, as determined in a standard single cylinder test engine, which measures ignition delay compared to primary reference fuels. The higher the CN, the easier a high-speed, direct-injection engine will start, and the less white smoking and diesel knock after start-up.

Cetane Number Improver: an additive (usually an organic nitrate) that boosts the CN of a fuel.

Cloud Point: the temperature at which a cloud of wax crystals appears when a lubricant or distillate fuel is cooled under standard conditions. Indicates the tendency of the material to plug filters or small orifices under cold weather conditions.

Compression Ratio: in an IC engine, the ratio of the volume of combustion space at bottom dead center to that at top dead center.

Copper Strip Corrosion: a qualitative measure of the tendency of a petroleum product to corrode pure copper.

Density: mass per unit volume.

Detergent: a substance added to a fuel or lubricant to keep engine parts clean. In motor oil formulations, the most commonly used detergents are metallic soaps with a reserve of basicity to neutralize acids formed during combustion.

Detergent/Dispersant: an additive package that combines a detergent with a dispersant.

Detonation: uncontrolled burning of the last portion (end gas) of the air/fuel mixture in the cylinder of a spark-ignition engine. Also known as ''knock'' or ''ping''.

Distillation: the basic test used to characterize the volatility of a gasoline or distillate fuel. For gasoline testing, ASTM D-86 is the defining standard procedure for use.

Emissions (Mobile Sources): the combustion of fuel leads to the emission of exhaust gases that may be regarded as pollutants. Water and CO_2 are not included in this category, but CO, NO_x and hydrocarbons are subject to legislative control. All three are emitted by gasoline engines; diesel engines also emit particulates that are controlled.

End Point: highest vapor temperature recorded during a distillation test of a petroleum stock.

Exhaust Gas Recirculation (EGR): system to reduce automotive emission of nitrogen oxides (NO_x). It routes exhaust gases into the carburetor or intake manifold, where they dilute the air/fuel mixture and reduce peak combustion temperatures, thereby reducing the tendency for NO_x to form.

Flash Point: minimum temperature at which a fluid will support instantaneous combustion (a flash) but before it will burn continuously (fire point). Flash point is an important indicator of the fire and explosion hazards associated with a petroleum product.

Gasohol Blend: a spark-ignition automotive engine fuel containing methanol or denatured fuel ethanol in a base gasoline. It may be leaded or unleaded.

Gasoline: a volatile mixture of liquid hydrocarbons, containing small amounts of additives and suitable for use as a fuel in spark-ignition IC engines.

Gravity: in petroleum products, the mass/volume relationship expressed as Equations (6) and (7):

$$\text{specific gravity} = \frac{\text{mass/unit volume product at } 60\,^{\circ}\text{F } (15\,^{\circ}\text{C})}{\text{mass/unit volume water at } 60\,^{\circ}\text{F } (15\,^{\circ}\text{C})} \tag{6}$$

$$\text{API gravity} = \frac{141.5}{\text{sp. gravity at } 60\,^{\circ}\text{F}/60\,^{\circ}\text{F}\ (30\,^{\circ}\text{C}/30\,^{\circ}\text{F})} - 131.5 \quad (7)$$

Hydrofinishing: a process for treating raw extracted base stocks with hydrogen to saturate them for improved stability.

Induction Period: in an oxidation test, the time period during which oxidation proceeds at a constant and relatively low rate. It ends at the point where oxidation rate increases sharply.

Inhibitor: additive that improves the performance of a petroleum product by controlling undesirable chemical reactions, i.e. oxidation inhibitor, rust inhibitor, and so on.

Insolubles: contaminants found in used oils due to dust, dirt, wear particles or oxidation products. Often measured as pentane or benzene insolubles to reflect insoluble character.

Lead: commonly used name for tetraethyl or tetramethyl lead, an additive used in gasoline to improve octane ratings. Elemental lead is commonly used in sleeve bearing and bushing alloys.

Naphthenic: a type of petroleum fluid derived from naphthenic crude oil, containing a high proportion of closed-ring methylene groups.

Octane Number: a measure of a fuel's ability to prevent detonation in a spark ignition engine. Measured in a standard single-cylinder, variable-compression-ratio engine by comparison with primary reference fuels. Under mild conditions, the engine measures research octane number (RON); under severe conditions (MON). Where the law requires posting of octane numbers on dispensing pumps, the AKI is used. This is the arithmetic average of RON and MON, $(R + M)/2$. It approximates the road octane number, which is a measure of how an average car responds to the fuel.

Octane Requirement (OR): the lowest octane number reference fuel that will allow an engine to run knock-free under standard conditions of service. Octane requirement is a characteristic of each individual vehicle.

Octane Requirement Increase (ORI): as deposits accumulate in the combustion chamber, the octane requirement index of an engine increases, usually reaching an equilibrium value after 10 000–30 000 km. Octane requirement index is a measure of the increase, which may be in the range of three to ten numbers.

Oxidation: occurs when oxygen attacks petroleum fluids. The process is accelerated by heat, light, metal catalysts and the presence of water, acids, or solid contaminants. It leads to increased viscosity and deposit formation.

Oxidation Inhibitor: substance added in small quantities to a petroleum product to increase its oxidation resistance, thereby lengthening its service or storage life; also called antioxidant.

Oxidation Stability: resistance of a petroleum product to oxidation and, therefore, a measure of its potential service or storage life.

Oxygenate: an oxygen-containing ashless organic compound, such as alcohol or ether, that can be used as a fuel or fuel supplement.

Oxygenated Fuels: fuels for IC engines that contain oxygen combined in the molecule, e.g. alcohols, ethers and esters. Term also applies to blends of gasoline with oxygenates, e.g. gasohol, which contains 10% by volume anhydrous ethanol in unleaded gasoline.

Ozone and CO Nonattainment Areas: any area of the continental USA that does not meet the 1990 CAA requirements for carbon monoxide or ground-level ozone pollutants.

Paraffinic: a type of petroleum fluid derived from paraffinic crude oil and containing a high proportion of straight chain saturated hydrocarbons. Often susceptible to cold flow problems.

Positive Crankcase Ventilation (PCV): system for removing blow-by gases from the crankcase and returning them through the carburetor intake manifold to the combustion chamber where the recirculated hydrocarbons are burned. A positive crankcase valve controls the flow of gases from the crankcase to reduce hydrocarbon emissions.

Pour Point: an indicator of the ability of an oil or distillate fuel to flow at cold operating temperatures. It is the lowest temperature at which the fluid will flow when cooled under prescribed conditions.

Pour Point Depressant: additive used to lower the pour point or low-temperature fluidity of a petroleum product.

Preignition: ignition of the fuel/air mixture in a gasoline engine before the spark plug fires. Often caused by incandescent fuel or lubricant deposits in the combustion chamber, it wastes power and may damage the engine.

Refining: series of processes to convert crude oil and its fractions into finished petroleum products, including thermal cracking, catalytic cracking, polymerization, alkylation, reforming, hydrocracking, hydroforming, hydrogenation, hydrogen treating, hydrofining, solvent extraction, dewaxing, de-oiling, acid treating, clay filtration and de-asphalting.

Rerefining: a process of reclaiming used lubricant oils and restoring them to a condition similar to that of virgin stocks by filtration, clay adsorption or more elaborate methods.

Syncrude: unconventional crudes such as those derived from tar sands, oil shale and coal liquefaction.

US EPA Complex Model: implemented on January 1, 1997. The model is more restrictive than the simple model, and contains limits on RVP, oxygen, olefins, benzene,

sulfur, and T-90. In addition, it includes requirements on aromatic content and T-50 temperatures, added in 1998.

US EPA Oxygenated Gasoline: required annually from September 15 to March 15 for use in most CO nonattainment areas. Oxygenated gasoline is defined as a spark-ignition engine fuel meeting ASTM D-4814 specifications and blended to include a minimum of 2.0% mass oxygen and a maximum of 1.0% volume benzene.

US EPA Simple Model: used to define RFG. Effective January 1, 1995. The model includes RVP and oxygen content requirements to reduce volatile organic compound emissions. It caps oxygen, benzene, sulfur, olefins, and T-90 content at levels equal to or lower than a refiner's 1990 baseline.

Vapor Pressure, Reid (RVP): measure of the pressure of vapor accumulated above a sample of gasoline or other volatile fuel in a standard bomb at 100 °F (37.8 °C). Used to predict the vapor locking tendencies of the fuel in a vehicle's fuel system. Controlled by law in some areas to limit air pollution from hydrocarbon evaporation while dispensing.

10.1.4 Associations and Standards Organizations

AAMA	American Automobile Manufacturers' Association (formerly MVMA)
ACEA	Association des Constructeurs Européens de l'Automobile (Association of European Automotive Manufacturers)
ACS	American Chemical Society
AFNOR	Association Français Pétroles de Normalisation
ANFAVEA	Auto Manufacturers' Association (Brazil)
ANSI	American National Standards Institute
APE	Association of Petroleum Engineers (USA)
API	American Petroleum Institute
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing Materials
ATC	Technical Committee of Petroleum Additive Manufacturers (Europe)
BHRA	British Hydromechanics Research Association
BNP	Bureau de Normalisation des Pétroles
BSI	British Standards Institution
BTC	British Technical Council of the Motor and Petroleum Industries (member CEC)
CARB	California Air Resources Board
CEC	Conseil Européen de Coordination pour les Developments des Essais de Performance des Lubrifiants et des Combustibles pour Moteurs (Coordinating European Council)
CEC (Finland)	Finnish Petroleum Federation (member CEC)
CEC/SB	Conseil Europen de Coordination/ Societe Belge (member CEC)
CEFIC	European Chemical Industry Council
CEN	Conseil Européen de Normalisation
CIMAC	International Council on Combustion Engines
CMA	Chemical Manufacturers' Association
CONCAWE	Conservation of Clean Air and Water (Europe)
CRC	Coordinating Research Council (USA)
CUNA	Commissione Tecnica di Unificazione nel l'Autoveicolo (member CEC)
DIN	Deutsche Industrie Norm
DKA	Deutscher Koordinierungsausschuss im Coordinating European Council (member CEC)
ECE	Economic Commission for Europe
EEB	European Environmental Bureau
EFTC	Engine Fuels Technical Committee (of CEC)
EMA	Engine Manufacturers' Association
EU	European Union
GFC	Groupement Français de Coordination (member CEC)
GRPE	Groupe des Rapporteurs pour la Pollution et l'Energie
IFP	Institute Français du Pétrole
IP	Institute of Petroleum (UK)
ISO	International Organization for Standardization
JAMA	Japan Automobile Manufacturers' Association Inc.
JARI	Japan Automobile Research Institute
JASO	Japan Automobile Standards Organization
JIS	Japanese Industrial Standards
JSAE	Society of Automotive Engineers (Japan)
NCM	National Comite Motorproeven (Netherlands) (member CEC)
NPRA	National Petroleum Refiners' Association
PAJ	Petroleum Association of Japan
SAE	Society of Automotive Engineers

SMR	Svenska Mekanisters Riksforenig (member CEC)
SNV	Schweizerische Normenvereinung (member CEC)
US EPA	Environmental Protection Agency

ABBREVIATIONS AND ACRONYMS

ACEA	Association des Constructeurs Européens de l'Automobile
AE	Atomic Emission
AET	Atmospheric Equivalent Temperature
AKI	Antiknock Index
AN	Aviation Number
ANSI	American National Standards Institute
APCD	Air Pollution Control District
API	American Petroleum Institute
AQMD	Air Quality Management District
ASTM	American Society for Testing and Materials
BACT	Best Available Control Technology
bp	Boiling Point
BPD	Barrels Per Day
Btu	British Thermal Unit
CAA	Clean Air Act
CAAQS	California Ambient Air Quality Standard
CARB	California Air Resources Board
CCAA	California Clean Air Act
CFPP	Cold Filter Plugging Point
CFR	Constant Fuel Ratio
CI	Compression–Ignition
CN	Cetane Number
CP	Paraffin Chains
CRC	Coordinating Research Council (USA)
CST	Centistokes
CW	Continuous Wave
DCS	Distributed Control System
DIPE	Diisopropyl Ether
DVPE	Dry Vapor Pressure Equivalent
EGR	Exhaust Gas Recirculation
EMF	Electromagnetic Field
EMI	Electromagnetic Interference
EP	End Point
ETBE	Ethyl Tertiary Butyl Ether
EU	European Union
FAT	Factory Acceptance Tests
FBP	Final Boiling Point
FCAA	Federal Clean Air Act
FCC	Fluid Catalytic Cracker
FFT	Fast Fourier Transform
FIA	Fluorescent Indicator Absorption
FID	Flame Ionization Detection
FIP	Federal Implementation Plan
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatograph/Mass Spectroscopy
GPA	Gas Processors' Association
GPC	Gel Permeation Chromatography
HAP	Hazardous Air Pollutant
HPLC	High-performance Liquid Chromatography
HTSD	High-temperature Simulated Distillation
IBP	Initial Boiling Point
IC	Internal Combustion
IR	Infrared
IRVs	Ideal Resting Values
ISO	International Standards Organization
IVD	Intake Valve Deposit
LC	Liquid Chromatography
LL	Low Lead
LNG	Liquid Natural Gas
LP	Linear Program
LPG	Liquefied Petroleum Gas
MMT	Methylcyclopentadienyl Manganese Tricarbonyl
MON	Motor Octane Number
MTBE	Methyl Tertiary Butyl Ether
MVC	Multivariable Controller
NAAQS	National Ambient Air Quality Standard
NIR	Near-infrared
NMOCs	Nonmethane Organic Compounds
NMR	Nuclear Magnetic Resonance
NOISE	Nitric Oxide Ionization Spectrometry Evaluation
NSR	New Source Review
ONR	Octane Number Response
PIANO	Paraffin, Isoparaffin, Aromatic, Naphthene, Olefin (content)
PLS	Partial Least Squares Regression
PM	Particulate Matter
PN	Performance Number
PNA	Polynuclear Aromatic
PSD	Prevention of Significant Deterioration
PSI	Pollutant Standards Index
PTV	Programmed Temperature Vaporization
PVC	Polyvinyl Chloride
QA	Quality Assurance
QP	Quadratic Programming
RA	Aromatic Ring
RFG	Reformulated Gasoline
RN	Naphthene Rings
ROG	Reactive Organic Gas
RON	Research Octane Number
RVP	Reid Vapor Pressure
SAE	Society of Automotive Engineers
SD	Standard Deviation
SI	Système International d'Unitès

SIP	State Implementation Plan
SLP	Successive Linear Program
TAC	Toxic Air Contaminant
TAME	Tertiary Amyl Methyl Ether
TBP	True Boiling Point
US EPA	United States Environmental Protection Agency
UV	Ultraviolet
VOC	Volatile Organic Carbon

REFERENCES

1. B. Elvers (ed.), *Ullmann's Encyclopedia of Industrial Chemistry*, 5th edition, Resources of Oil and Gas, VCH, Vol. A23, 1993.
2. *BP Statistical Review of World Energy – June 1994*, 'Proved Reserves at End 1993', p. 2, 'Crude Oil Consumption', p. 7, 1994.
3. M. Howe-Grant (ed.), *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edition, Volume 1 Alcohol Fuels; Volume 9 Exhaust Control, Automotive; Volume 12 Gasoline and Other Motor Fuels, Wiley, 1993.
4. T. Midgley, Jr, T.A. Boyd, *Ind. Eng. Chem.*, **14**, 589, 849, 894 (1922).
5. G. Edgar, 'Measurement of the Knock Characteristics of Gasoline in Terms of a Standard Fuel', *Ind. Eng. Chem.*, **19**, 145–146 (1927).
6. 'Texaco to Introduce Clean Burning Gasoline', *Oil Gas J.*, 28 February, 22–23 (1994).
7. G.D. Hobson (ed.), *Modern Petroleum Technology*, 5th edition, Wiley, Chichester, 1984. (Chapter 1. G.D. Hobson, Chapter 9. R.E. Banks, P.J. King, Chapter 19. R.A. Vere, Chapter 20. K. Owen).
8. C.F. Kettering, 'The Effect of the Molecular Structure of Fuels on the Power and Efficiency of Internal Combustion Engines', *Ind. Eng. Chem.*, **36**, 1079–1085 (1944).
9. D.F. Caris, E.E. Nelson, 'A New Look at High Compression Engines', SAE Paper 812A, 1958.
10. I.D.G. Berwick, 'The Rise and Fall of Lead in Petrol', *Phys. Technol.*, **18**, 158–164 (1987).
11. D. Schuetzle, W.O. Siegl, T.E. Jensen, M.A. Dearth, E.W. Kaiser, R. Gorse, W. Kreucher, E. Kulik, 'The Relationship Between Gasoline Composition and Vehicle Hydrocarbon Emissions: A Review of Current Studies and Future Research Needs', *Environ. Health Perspect. Suppl.*, **102**(4), 3–12 (1994)
12. G.F. Unzelman, 'Oxygenate/Hydrocarbon Shift Will Rewrite Gasoline Recipes', *Oil Gas J.*, April 29 (1991).
13. J.M. Wilson, '1992 Will Impact U.S. Refiners', *Hydrocarbon Processing*, February, 71 (1996).
14. G. Yepsin, T. Witoshkin, 'Refiners have Options to Deal with Reformulated Gasoline', *Oil Gas J.*, 8 April, 68–71 (1991).
15. V.E. Pierce, B.B. Bansal, 'Lead Phase-out and Octane Enhancement', *Chem. Eng. Prog.*, March (1986).
16. B.J. Finlayson-Pitts, J.N. Pitts, Jr, 'Volatile Organic Compounds: Ozone Formation, Alternative Fuels and Toxics', *Chem. Ind. (London)*, October, 796–800 (1993).
17. J.G. Calvert, J.B. Heywood, R.F. Sawyer, J.H. Seinfeld, 'Achieving Acceptable Air Quality: Some Reflections on Controlling Vehicle Emissions', *Science*, **261**, 37–45 (1993).
18. G.H. Unzelman, D.W. Michalski, 'Octane Improvement Update – Refinery Processing, Antiknock and Oxygenates', National Petroleum Refiner's Annual Meeting Presentation AM-84-43, March 1984.
19. J.M. Colucci, J.J. Wise, 'Initial Results from the Auto/Oil Quality Improvement Research Program', National Petroleum Refiner's Annual Meeting Presentation AM-91-46, March, 1991.
20. S.K. Hoekman, 'Speciated Measurements and Calculated Reactivities of Vehicle Exhaust Emissions from Conventional and Reformulated Gasolines', *Environ. Sci. Technol.*, **26**, 1206–1216 (1992).
21. E.W. Kaiser, W.O. Siegl, D.F. Cotton, R.W. Anderson, 'Effect of Fuel Structure on Emissions from a Spark-ignited Engine. 2. Naphthene and Aromatic Fuels', *Environ. Sci. Technol.*, **26**, 1581–1586 (1992).
22. K.T. Taylor, 'Catalysts in Cars', *Chemtech*, September, 551–555 (1990).
23. E.M. Goodger, *Alternative Fuels*, Macmillan, 1980. (Chapter 5, also Chapter 7 and Appendix 4.)
24. R.J. Schmidt, P.L. Bogdan, N.L. Gilsdorf, 'Meeting the Challenge of Reformulated Gasoline', *Chemtech*, February, 41–42 (1993).
25. *Knocking Characteristics of Pure Hydrocarbons*, ASTM STP 225, 1958.
26. 'Health Effects of Gasoline', *Environ. Health Perspect. Suppl.*, **101**(6), (1993).
27. Environmental Protection Agency, CFR, Vol. 42(14), 38 582–38 586, Initial Oxygenate Levels.
28. 'Reformulated Gasoline: EPA Simple Model Implementation', *Fuels Reformulation*, 11–21 Sept/October (1994).
29. A. Rhodes, 'U.S. Refiners Make Complex Model RFG as They Prepare for the Next Hurdle', *Oil Gas J.*, January 5 (1999).
30. A. Rhodes et al., 'California Plans to Ban MTBE from Gasoline', *Oil Gas J.*, April 5 (1999).
31. Proposed Amendments to the California Phase 2 Reformulated Gasoline Regulations, California Air Resources Board, April 22, 1994.
32. P. Giammattea, J. Edwards, 'Reliable Refinery Control Using Process NMR', *Control Magazine*, June, 71 (1999).
33. 'New Diesel Fuel Advances Clean Vehicle Emissions'.
34. 'New Low Sulfur Diesel Debuts in Brazil', *Oil Gas J.*, December 12, 22 (1996).

35. S. Romanow, 'How Much Sulfur is Too Much for Clean Fuels', *Hydrocarbon Processing*, June, 15 (1999).
36. I.R. Hills, G.W. Smith, E.V. Whitehead, 'Hydrocarbons from Fossil Fuels and their Relationship with Living Organisms', *J. Inst. Petrol.*, **56**, 127–137 (1970).
37. C.K. Westbrook, W.J. Pitz, 'The Chemical Kinetics of Engine Knock', *Energy Technol. Rev.*, Feb/Mar, 1–13 (1991).
38. C.K. Westbrook, 'The Chemistry Behind Engine Knock', *Chem. Ind. (London)*, 3 August, 562–566 (1992).
39. W.G. Lovell, J.M. Campbell, T.A. Boyd, 'Detonation Characteristics of Some Paraffin Hydrocarbons', *Ind. Eng. Chem.*, **23**, 26–29 (1931).
40. K.A. Boldt, S.T. Griffiths, *Criteria for Quality of Petroleum Products*, ed. J.P. Allinson, Applied Science, McGraw Hill, New York, Chapter 5, 1978.
41. *Annual Compilation of ISO Standards* for years 1986–1998. All ISO standards listed in this work are covered in the annual edition, even if not revised. Year of revision compilation may contain commentary on the standard, but otherwise all ISO standards are available from the American National Standards Institute (ANSI). All standards are potentially subject to annual update.
42. K. Springer, L. Smith, 'Experiments with MTBE-100 as an Automobile Fuel', Tenth International Symposium on Alcohol Fuels, Proceedings, 1, 53, 1993.
43. Automotive Gasolines–Recommended Practice, SAE J312 Jan93, Section 3 & 4, SAE *Handbook, Volume 1*, 1994.
44. D. Simanaitis, D. Kott, 'Octane Enhancers', *Road & Track*, April, 82, 83, 86–88 (1989).
45. *The Automotive Handbook*, 1995 edition, Bosch, 1995.
46. L.M. Sheppard, 'Automotive Sensors Improve Driving Performance', *Ceramic Bulletin*, **71**, 905–913 (1992).
47. *Annual Book of ASTM Standards* for years 1988–1998. All ASTM standards listed in this work are covered in the annual edition, even if not revised. Year of revision compilation may contain commentary on the standard, but otherwise all ASTM standards are available from the American Society for Testing and Materials. All standards are potentially subject to annual update.
48. Automotive Gasolines – Recommended Practice, SAE J312 Jan93, *SAE Handbook, Volume 1*, 1994.
49. Alternative Automotive Fuels – SAE Information Report, SAE J1297 Mar93, SAE Handbook, Volume 1, 1994.
50. J.J. Wise, R.A. Corbett, 'Auto/Oil Progressing Toward Cleaner Fuel Composition', *Fuels Reformulation*, March/April (1992).
51. W.J. Piel, R.X. Thomas, 'Oxygenates for Reformulated Gasolines', *Hydrocarbon Processing*, July, 68–73 1990.
52. A.F. Lau, R.M. Stanley, L. Watkins (eds.), '1991 Vapor Pressure Test Methods Round Robin Program', Research Report RR: D02-1286, ASTM D02.08D.
53. R.H. Perry, C.H. Chilton, *Chemical Engineers' Handbook*, 6th edition, Chapter 3, McGraw-Hill, 1985.
54. K.A. Baron et al., 'Hydrotreating of Light Cycle Oil', National Petroleum Refiner's Annual Meeting Presentation AM-92-20, March 1992.
55. J.R. Felten, K.M. McCarthy, 'Impact of RVP Reduction', *Hydrocarbon Processing*, May, 49 (1987).
56. 'Fuel Blending and Analysis for the Auto/Oil Air Quality Improvement Research Program', SAE paper 902098, 1990.
57. P.A. Nick, 'RVP Estimation for Naphtha Stocks', API Technical Data Committee Internal Memo, September 10, 1993.
58. J.J. Varquez-Esparragoza, 'How to Estimate RVP of Blends', *Hydrocarbon Processing*, August (1992).
59. R. Amadei et al., 'A Proposal for Economic Environmental Gasoline', *Petroleum Quarterly Technology*, Summer 55 (1999).
60. C. Passarella, 'Blend Gasoline Quality Prediction – Online and Offline Implementation of Gasoline Quality Prediction and Blend Index Calculations', National Petroleum Refiners Computer Conference, Paper CC 95-134, Nov. 1995.
61. K.M. Vittachi, J. Jackson, 'Citgo Corpus Christi Refinery Gasoline Blending Upgrade', National Petroleum Refiners Computer Conference, Paper CC 95-133, Nov. 1995.
62. A. Raz, J.A. Barsamian, 'Meet Changing Fuel Requirements with On-line Blend Optimization', *Hydrocarbon Processing*, February, 71 (1996).
63. K.J .Clevett, *Process Analyzer Technology*, John Wiley, New York, 1986.
64. R.J. Quann, S.B. Jaffe, 'Structure Oriented Lumping – Describing the Chemistry of Complex Hydrocarbon Mixtures', *IEC Res.*, **31**(4), 701 (1976).
65. 'National Petroleum Refiners Q&A Forum–1st Summary Article, 1990's Fuel Specifications Will Require Process Changes', *Oil Gas J.*, February 26 (1990).
66. B.L. Karger, L.P. Snyder, C. Howarth, *An Introduction to Separation Sciences*, John Wiley & Sons, NY, 1973.
67. B.G. Liptak, *Process Measurement and Analysis*, Chilton Book Company, Radmor, PE, 1995.
68. H.H. Willard, L.L. Merritt, J.A. Dean, F.A. Settle, D. Van Nostrand, *Instrumental Methods of Analysis*, 6th edition, 1981.
69. M. Johnston et al., 'Virtual Analyzers Help Refiner Meet RFG Specifications', *Hydrocarbon Processing*, June, 45 (1999).
70. *Analytical Emission Spectroscopy*, Marcel Dekker, NY, Vol. 1, 1972.
71. H.H. Jaffe, M. Orchin, *Theory and Application of Ultraviolet Spectroscopy*, John Wiley, NY, 1962.
72. G.A. Bishop, J.R. Starkey, A. Ihlenfeldt, W.J. Williams, D.H. Stedman, 'IR Long-path Photometry: A Remote Sensing Tool for Automobile Emissions', *Anal. Chem.*, **61**, 671A–677A (1989).

73. D.A. Burns, E. Ciurzak, *Handbook of Near-infrared*, Marcel Dekker, New York, 1992.
74. M. Ichikawa et al., 'Estimation of Octane Number of Automobile Gasoline by Fourier Transform Infrared Absorption Spectrometry', *Appl. Spectrosc.*, **46**(6), 966–971 (1992).
75. *API Technical Data Book*, American Petroleum Institute, Washington, D.C., 1992, 1998 editions and revisions (various), Chapters 1, 2, 3, and 5.
76. P.A. Wadsworth, D.C. Villalanti, 'Pinpoint Hydrocarbon Types', *Hydrocarbon Processing*, May, 109 (1992).
77. California Air Resources Board, Transportation Fuel Codes, section 80.65(e)(1), 1990.
78. Environmental Protection Agency, 40CFR Part 80, Appendix D, 1991.
79. S.M. Jacob et al., 'Compositional Modeling Reduces Crude Analysis Time, Predicts Yield', *Oil Gas J.*, July 6 (1998).
80. 'Regional Gasoline Octane Blending – Internal Memo, Author's Private Correspondence, 1988.
81. G.B. Kauffman, 'Midgley: Saint or Serpent?', *Chemtech*, December, 717–725 (1989).
82. Stoichiometric Air/Fuel Ratios of Automotive Fuels – Recommended Practice, SAE J1829 May92, *SAE Handbook, Volume 1*, 1994.
83. 'E.C. Seeks Gasoline Emission Control', *Hydrocarbon Processing*, September, 43 (1990).
84. T.J. Smith, S.K. Hammond, O. Wong, 'Health Effects of Gasoline Exposure. I. Exposure Assessment for U.S. Distribution Workers', *Environ. Health Perspect. Suppl.*, **101**(6), 13 (1993).
85. C. Dahl, W.J. Garcia-Barre, 'Reformulated Gasoline More Profitable in Venezuela than in the U.S.', *Oil Gas J.*, June 21 (1993).
86. J.H. Beynon, *Mass Spectrometry and Its Application to Organic Chemistry*, Elsevier Publishing, NY, 1960.
87. W.J. Price, *Analytical Atomic Absorption Spectrometry*, Heyden & Sons, Toronto, 1972.
88. G.G. Akland, 'Exposure of the General Population to Gasoline', *Environ. Health Perspect. Suppl.*, **101**(6), 27–32 (1993).
89. G.Peaff, 'Court Ruling Spurs Continued Debate Over Gasoline Oxygenates', *Chem. Eng. News*, 26 September, 8–13 (1994).
90. J.M. Campbell, T.A. Boyd, 'Methods of Knock Rating. Section 15. Measurement of the Knocking Characteristics of Automotive Fuels', in *The Science of Petroleum*, Oxford University Press, 3057–3065, 1938.
91. *Preparation of Distillates for Front End Octane Number (RON 100C) of Motor Gasoline IP 325/82 Standard Methods for Analysis and Testing of Petroleum and Related Products*, Wiley, 1994.
92. V. Daniel Hunt, *The Gasohol Handbook*, Industrial Press, 1981.
93. M.H. Naes, *Multivariate Calibration*, John Wiley, New York, 1989.
94. A.J. Lubeck, R.D. Cook, 'Bromine Number Should Replace FIA in Gasoline Olefins Testing', *Oil Gas J.*, Dec. 28 (1992).
95. California Air Resources Board, Transportation Fuel Codes, section 2296, 'Motor Fuel Sampling Procedures', 1991.
96. J. Langston et al., 'Technologies to Achieve 2000 Diesel Specifications', *Petroleum Quarterly Technology*, Summer 5 (1999).
97. 'EPA Proposes Reducing Sulfur Levels in Gasoline', *Chem. Eng. Prog.*, August, 20 (1999).
98. URL: www.epa.gov/orns/consumer/fuels/oxypanel/blueribb.htm.
99. URL: www.epa.gov/orns/consumer/fuels/.
100. A.W. Drew (ed.), *Manual of Hydrocarbon Analysis*, sixth edition, ASTM Publications, 1998.
101. S.W. Kubala et al., 'Need Total Sulfur Content – Use Chemiluminescence', *InTech*, Sept, 71 (1996).
102. *ASOMA* Monthly Newsletter, ASOMA Instruments, Austin TX, August 25, 1998.
103. P.T. Lo, 'On-line RVP Analyses Improve Gasoline Blending', *InTech*, September, 31 (1994).
104. 'Impact of European Standard EN228', *Hydrocarbon Processing*, June, 27 (1999).

Full Range Crudes, Analytical Methodology of

M.A. Ali
King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

1 Introduction **2**
2 History **2**
3 Terminology **2**
4 Composition **3**
4.1 Wellhead Crude Oil 3
4.2 Stabilized Crude Oils 3
5 Characterization Methods **5**
5.1 Physical Methods 6
5.2 Extraction, Filtration and Centrifuging Methods 8
5.3 Fractionation 9
5.4 Distillation 10
5.5 High-performance Liquid Chromatography 10
5.6 Gas Chromatography 11
5.7 Supercritical Fluid Chromatography 12
5.8 Mass Spectrometry 12
5.9 Nuclear Magnetic Resonance Spectroscopy 13
5.10 Infrared Spectroscopy 13
5.11 Thermal Analysis 13
5.12 Elemental Analysis Method 13
5.13 Atomic Absorption Spectrometry 14
5.14 Inductively Coupled Plasma Atomic Emission Spectrometry 14
5.15 Chemiluminescent Detection Method 15
5.16 Titration Methods 15
Acknowledgments **15**
Abbreviations and Acronyms **15**
Related Articles **16**
References **16**

Crude oil at the wellhead is a hot, high pressure fluid that contains gases, oil, water vapor and solid impurities. This mixture is processed in a gas oil separator where it is brought to atmospheric pressure and separated into gases, water, solid and stabilized oil. Characterization of a stabilized crude oil is carried out using a number of analytical methods for both its physical and chemical properties. These methods include use of a densitometer for density measurement and a hydrometer for specific gravity and American Petroleum Institute (API) gravity. The densitometer provides more precise and reproducible measurement of density as well as better control of the temperature of the sample. Pour point is determined using a cooling and pouring method, water content determination by distillation, ash content by a high temperature burning method, while salt content of a crude is determined by an electrometric method. The metal contents are determined using inductively coupled plasma atomic emission spectrometry (ICPAES) and atomic absorption spectrometry (AAS). The ICPAES is faster as compared to AAS and provides simultaneous determination of a wide range of metals. The elemental analysis is carried out by oxidative combustion of the sample in which carbon, hydrogen and sulfur are measured in oxide form, while nitrogen is measured in the form of molecular nitrogen. Sulfur is also determined using atomic emission spectrometry (AES) and chemiluminescent detectors which are very sensitive as compared to the oxidative combustion method. In addition, chemiluminescent detectors also provide nitrogen determination. Molecular weight is determined using thermoelectric measurement of vapor pressure; however the values obtained are influenced by the temperature of the measurement and the type of solvent used for dissolving the sample. Gel permeation chromatography (GPC) is used for average molecular weight determination of crude oil. Physical distillation is used for producing light and middle distillates and residue fractions based on the desired boiling range or volume percent range of a crude oil as well as determination of the boiling range distribution. Gas chromatography (GC) is utilized for determination of boiling range distribution by simulated distillation, volatile compound determination and identification of crude oil spill. The faster analysis time, better reproducibility and equivalent results are some of the advantages of simulated distillation over physical distillation in the determination of the boiling range distribution of a crude oil. Gas chromatography/mass spectrometry (GC/MS) separates and identifies volatiles and low boiling components of a crude oil while mass spectrometry (MS) is utilized for assessing the molecular weights of the compounds. High-performance liquid chromatography (HPLC) is used for separation and quantification of hydrocarbon group types such as saturates, aromatics and polars. Aromatics are further separated into monoaromatics, diaromatics, triaromatics and polyaromatics. Liquid chromatography (LC) is used to fractionate the whole crude oil into desired fractions such as acids, bases and neutral nitrogen compounds. Petroleum resins are separated from whole crude oil by a liquid chromatographic method using ion-exchange resins. Liquid chromatography/mass spectrometry (LC/MS) provides separation and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

positive identification of high boiling compounds of a crude oil. Supercritical fluid chromatography (SFC) is used for separation and identification of hydrocarbon group types. Compared to HPLC, SFC provides better separation of high boiling components. Infrared (IR) spectroscopy is used for determination of type of functional group and hydrogen bonding, thermal analysis for evaluating the thermal stability of the crude oils and classification of hydrocarbons while titrimetry is used for determining acid number, base number and water content.

1 INTRODUCTION

Crude oil or full range crude is a complex mixture of hydrocarbons and heteroatomic hydrocarbons which range from the simple one-carbon atom compound, methane, to highly condensed aromatics to metallic constituents such as porphyrins and asphaltenes. Crude oil at the wellhead is characterized by a hot, high velocity mixture of oil, water, hydrocarbon and nonhydrocarbon gases along with salts and sand. This mixture is then stabilized to remove gases, solids and water from oil which is a viscous brown to black liquid. The main groups of compounds present in the stabilized crude oils are: normal or straight chain alkanes (paraffins); branched alkanes (isoparaffins); cycloalkanes (naphthenes); aromatics; sulfur, nitrogen and oxygen containing compounds; and compounds having metals such as vanadium, nickel and iron. Due to this complexity, various analytical approaches are used to characterize crude oils. The characterization of the crude oil is performed either as a whole or after separating into a number of fractions based on boiling range or compound types. Analytical methods applied to whole crude oil provide information about its physical and chemical properties as well as the presence of relative amounts of different petroleum products.

2 HISTORY

The history of analytical methods starts with the discovery of crude oil. Early characterization techniques were very crude and provide limited information about the properties and composition of crude oil. These techniques and methods were more of manual type and slow and incorporated a wide variety of human errors. These methods were capable of providing only information regarding the light ends of the crude oil. The detection limit of these methods was also quite high and could not determine the trace quantities of metals and heteroatoms such as sulfur, nitrogen and oxygen present in crude oil. With the passage of time, the analytical methods became more automated and the speed and reliability of these methods are now quite high. With the introduction of high temperature methods, it became possible to identify and characterize heavy crude and residue more efficiently. Today, the analytical methods applied to crude oils are highly automated and sophisticated and provide information regarding its physical and chemical properties in a faster and more reliable manner along with very high reproducibility and repeatability.

3 TERMINOLOGY

Alkanes, also known as paraffins, are saturated compounds of carbon and hydrogen having the general formula C_nH_{2n+2}. Alkanes can be straight chain or branched.

Alkenes, also known as olefins, are unsaturated compounds of carbon and hydrogen having the general formula C_nH_{2n}. Alkanes can be straight chain or branched.

API gravity, defined by API, deg = (141.5/specific gravity 60/60 °F) −131.5.

ASTM distillation, an American Society for Testing and Materials standard laboratory batch distillation process used for determining boiling range distribution of naphtha and middle distillate carried out at atmospheric pressure.

Atmospheric residue, residue left in the true boiling point (TBP) distillation of a crude oil carried out according to ASTM D-2892.

Bottom of the barrel, the high boiling residue obtained by distillation of a crude oil.

Carbon residue, the residue formed by evaporation and thermal degradation of a carbon containing material such as crude oil or petroleum residue.

Stabilized crude oil, also known as petroleum, whole crude oil or full range crude or stabilized crude oil, is a naturally occurring liquid hydrocarbon mixture having some gaseous components and may also contain compounds of sulfur, nitrogen, oxygen, metals as well as other elements.

Cut, also called fraction, is a portion of a crude oil having a certain boiling temperature limit.

DTA, differential thermal analysis, is a technique of measuring the amount of heat evolved or absorbed as a function of temperature at which these changes take place within the material. The temperature difference between a test sample and reference material is measured as a function of temperature, while the substance and the reference material are subjected to a controlled temperature heating.

DTG, differential thermogravimetry, is a technique of measuring the amount of material loss as a function of temperature at which these changes take place within

the material. The loss of the material between a test sample and reference material is measured as a function of temperature, while the substance and the reference material are subjected to a controlled temperature heating.

FBP, final boiling point at which 99.5% of the material is evaporated.

Gas oil, a petroleum distillate product heavier than kerosene. When produced from a TBP distillation process it is atmospheric gas oil (AGO), with an approximate boiling range 190–325 °C, and from vacuum distillation it is vacuum gas oil (VGO), with an approximate boiling range 299–500 °C.

IBP, initial boiling point at which 0.5% of the material is evaporated.

Kerosene, distillate fraction which boils in the range 150–300 °C.

Light ends, low boiling hydrocarbons present in a crude oil which are mainly methane to pentane.

MBP, the temperature at which one half of the material is evaporated.

Middle distillate, distillate fraction which boils in the range 150–370 °C.

Naphtha, the distillate which boils in the range 36–200 °C.

Naphthenes, also called cycloalkanes and cycloparaffins, are cyclic compounds of carbon and hydrogen having single C–C bonds.

Sour crude, crude oil containing sulfur greater than 0.5–1.0 wt% percent or which contains a minimum of 0.05 cu ft of hydrogen sulfide per 100 gal of crude.

Vacuum residue, material left undistilled in the vacuum distillation of an atmospheric residue carried out according to ASTM D-1160.

4 COMPOSITION

4.1 Wellhead Crude Oil

Crude oil wellhead streams are usually characterized as turbulent, high temperature, high velocity mixtures of gases, oil, water vapor and solid impurities. As these streams reach the surface, they undergo continuous reduction in temperature and pressure forming a two-phase flow: gas and liquid. These streams need to be processed to separate gas and oil as well as other solid impurities present, because they are very difficult to handle, meter, characterize or transport. These streams are fed to a gas oil separator where the pressure is successively decreased in several steps and brought to atmosphere pressure. At the gas oil separator, the streams are separated into natural gas, crude oil, water and solid. These phases are then separated and analyzed for the assessment of the wellhead stream composition. The composition of the wellhead crude oil depends upon the type of reservoir and the composition of wellhead natural gas. It contains a substantial amount of hydrocarbon gases. As the pressure is reduced, these gases are released from the oil and the crude is stabilized. A typical composition of a wellhead crude oil is given in Table 1 showing the type of components and their percentage.

Table 1 Typical composition of the wellhead crude oil

Component	Mol. %
Nitrogen	0.56
Carbon dioxide	3.55
Hydrogen sulfide	1.00
Methane	45.34
Ethane	5.48
Propane	3.70
i-Butane	0.70
n-Butane	1.65
i-Pentane	0.48
n-Pentane	0.87
C_6-C_{19} compounds	24.35
C_{20+} compounds	12.32

Natural gas at the wellhead is normally associated with the oil. It may be in the form of free gas (gas not in solution) present at the top of the reservoir or dissolved in the oil (gas in solution) and its composition depends upon the type of reservoir. It is a complex mixture of hydrocarbons, mainly methane, ethane, propane and butane gases and some heavy hydrocarbons. Among these, methane and ethane are the most abundant.

4.2 Stabilized Crude Oils

Petroleum or whole crude oil is a viscous brown to black liquid mixture. The main constituents of crude oils are hydrocarbons mixed with varying amounts of heteroatomic hydrocarbon compounds. All hydrocarbon classes except alkenes are practically present in crude oils. Alkanes, cycloalkanes, mono- and polynuclear aromatics are present in crude oils. The ratio of these classes, however, differs appreciably from one crude to another. Light hydrocarbon gases such as methane and ethane may be present in small amounts dissolved in the crude or in large amounts as in associated gas. Associated gas mainly constitutes methane, ethane and propane. This gas is a valuable raw material for many petrochemicals. In addition to the hydrocarbon mixture, crude oils contain variable amounts of heteroatomic hydrocarbon compounds such as sulfur, nitrogen and oxygen compounds. These compounds, sometimes referred to as impurities, affect the ways of processing the crude as well as its market value. A refiner would prefer

Table 2 Average elemental composition (weight percent range) of stabilized crude oils

Element	Amount
Carbon	83.0–87.0
Hydrogen	11.0–14.0
Sulfur	0.05–2.50
Nitrogen	0.01–2.00
Oxygen	0.00–2.00

Table 3 Compounds present in crude oils

Compounds	Representative compounds
Hydrocarbons	
Alkanes (paraffins)	Undecane, dodecane, tricontane, tetratetracontane
Cycloalkanes (naphthenes)	Cyclohexanes, pentacyclics, hexacyclics
Aromatic/polyaromatic compounds	Pyrenes, anthracenes, phenanthrenes
Heteroatomic hydrocarbons	
Sulfur compounds	Mercaptans, thiophenes, benzothiophenes, sulfides
Nitrogen compounds	Pyrroles, pyridines, carbazoles, acridines, indoles
Oxygen compounds	Phenols, carboxylic acids, amides
Metallic compounds	
Vanadium compounds	Vanadium porphyrins
Nickel compounds	Nickel porphyrins

to process a low-sulfur crude since this will reduce the cost of hydrogen required for hydrotreatment of the products. Crude oils also contain trace amounts of heavy metals in the form of organometallic compounds and some inorganic salts, mainly sodium chloride. Some of these heavy metals, such as vanadium and nickel, are poisonous to processing catalysts. Crude oils having a high salt content are desalted before refining to reduce corrosion problems. The averaged elemental composition of stabilized crude oils is shown in Table 2. The compounds found in crude oils are given in Table 3 and generally belong to the following broad classes.

4.2.1 Hydrocarbons

The major constituents of most crude oils are hydrocarbon compounds besides minor concentrations of heteroatomic hydrocarbons. These hydrocarbon compounds are made up of hydrogen and carbon and belong to alkanes, cycloalkanes and aromatics. Alkanes, also known as paraffins, are saturated compounds having the general formula C_nH_{2n+2}. The simplest hydrocarbon compound, methane, may be present in small amounts dissolved in the crude oil. Alkanes are both straight chain as well as branched types. Branched hydrocarbons in the naphtha range are more valuable as gasoline constituents compared to the straight chain isomers since they have higher octane number. Branched hydrocarbons have lower boiling points as compared to the straight chain isomers. Cycloalkanes, also called cycloparaffins or naphthenes, are normally present in crude oils in variable proportions. Cyclohexane, substituted cyclohexanes and substituted cyclopentanes are found in the naphtha fractions. Cycloalkanes are relatively more stable than alkanes Aromatic compounds are normally present in all crude oils regardless of their type. However, C_6-C_8 mononuclear aromatics which are present in naphtha have commercial importance. These include benzene, toluene, xylenes and ethylbenzenes. The simplest aromatic compound, benzene (C_6H_6), is one of the basic raw materials for petrochemical production. Aromatics in this range are not only important petrochemical feedstocks but are also valuable motor fuels. Polynuclear aromatic compounds are present in heavier petroleum fractions and residues. Asphaltenes that are concentrated in heavy fuel oils and asphalt are polynuclear aromatics of a complex structure. It is known that condensed-ring aromatic hydrocarbons and heterocyclic compounds are the major constituents of asphaltenes.

4.2.2 Heteroatomic Hydrocarbons

Many types of heteroatomic hydrocarbon compounds occur in crude oils. The most important are sulfur, nitrogen and oxygen compounds. Metals are also present in trace amounts, mainly in the form of organometallic compounds such as vanadium and nickel porphyrins. Sulfur present in crude oils is mainly in the form of nonacidic organosulfur compounds. Sulfur compounds present in crude oil include mercaptans, sulfides, disulfides, thiophenes, benzothiophenes and dibenzothiophenes. Sour crudes contain a high percentage of hydrogen sulfide which is acidic in nature. However, many of the organosulfur compounds are not thermally stable and hydrogen sulfide is produced during crude processing. High-sulfur crudes are less valuable since they require an additional cost for the removal of sulfur compounds. During hydroprocessing of petroleum products, hydrodesulfurization (HDS) takes place and the sulfur present in the heteroatomic compounds is converted into hydrogen sulfide. Naphtha fed to catalytic reformers is hydrotreated to reduce sulfur compounds to ppm levels to ensure a long life-cycle for the expensive reforming catalysts. Nitrogen compounds present in crude oils are usually low in

concentration and range from simple compounds such as pyridine and pyrrole to more complex structures such as porphyrins. The complex structures of porphyrins are usually found in heavy fuel oils and residues. Nitrogen compounds are basic, neutral and acidic types. These compounds are thermally more stable than the sulfur compounds. During hydroprocessing of petroleum products, hydrodenitrogenation (HDN) takes place and the nitrogen present in the heteroatomic compounds is converted into ammonia. Oxygen compounds in crude oils are more complex than sulfur compounds. Most oxygen compounds are weakly acidic, such as phenol, cresylic acid and naphthenic acids. The total oxygen content of many crudes is very low. Nonacidic compounds such as esters, ketones and amides are less abundant than acidic compounds. Many of these oxygen compounds, however, are concentrated in the heavier portion of the crude. Many metals are found in crude oils; some of the more abundant are sodium, calcium, magnesium, aluminum, iron, vanadium and nickel. These normally occur in the form of inorganic salts such as sodium chloride, or in the form of organometallic compounds such as vanadium and nickel porphyrins, or in the form of salts of carboxylic acids, as in the case of calcium and magnesium. The organometallic compounds are usually concentrated in the heavier fractions and in crude oil residues.

5 CHARACTERIZATION METHODS

The whole crude or full range crude can be analyzed by a number of analytical methods. These analytical methods have been standardized by a number of societies and organizations such as ASTM, Institute of Petroleum (IP) and API. Among them, the ASTM provides a more comprehensive coverage of the analytical methods used for physical and chemical analysis of crude oils. Table 4 shows a summarized list of standard analytical methods used for the characterization of a crude oil. Figure 1 illustrates the variety of analytical techniques which are applied to characterize crude oils. Table 5 shows some of the properties of Arab Heavy and Alamein crude oils.

Table 4 A summary of standard analytical methods applied to crude oils and the type of information obtained

Test	ASTM method	Information obtained
Densitometer method	ASTM D-5002	Density
Hydrometer method	ASTM D-287	Specific gravity and API gravity
Hydrometer method	ASTM D-1298	Density, specific gravity and API gravity
Pour point	ASTM D-97	Determination of low temperature flow property
Electrometric method	ASTM D-3230	Salt content as sodium chloride
Ash content by combustion	ASTM D-482	Metal salts present
Ramsbottom carbon residue	ASTM D-524	Carbonaceous material present
Conradson carbon residue	ASTM D-189	Carbonaceous material present
Molecular weight	ASTM D-2503	Average molecular weight
TBP distillation	ASTM D-2892	Determination of boiling range distribution and collection of fractions for further characterization
Vacuum distillation	ASTM D-1160	Boiling range distribution and collection of fractions for further characterization
Physical distillation	ASTM D-95	Water content
Simulated distillation	ASTM D-5307	Boiling range distribution
Extraction	ASTM D-473	Amount of sediments
Filtration	ASTM D-4807	Amount of sediments
Asphaltene content	IP 143	*n*-Heptane insoluble present
Centrifuging method	ASTM D-96	Water and sediment content
XRF	ASTM D-4294	Sulfur content
Bomb method	ASTM D-129	Sulfur content
Atomic absorption spectroscopy	ASTM D-4628	Metal content
Atomic emission spectroscopy	ASTM D-5623	Sulfur and nitrogen content
Inductively coupled plasma atomic emission spectroscopy	ASTM D-5600	Metal and sulfur content
Potentiometric Karl Fischer titration	ASTM D-4377	Water content
Coulometric Karl Fischer titration	ASTM D-4928	Water content
Potentiometric titration	ASTM D-664	Acid number (acidity)
Potentiometric titration	ASTM D-2896	Base number (basicity)

XRF, X-ray fluorescence spectroscopy.

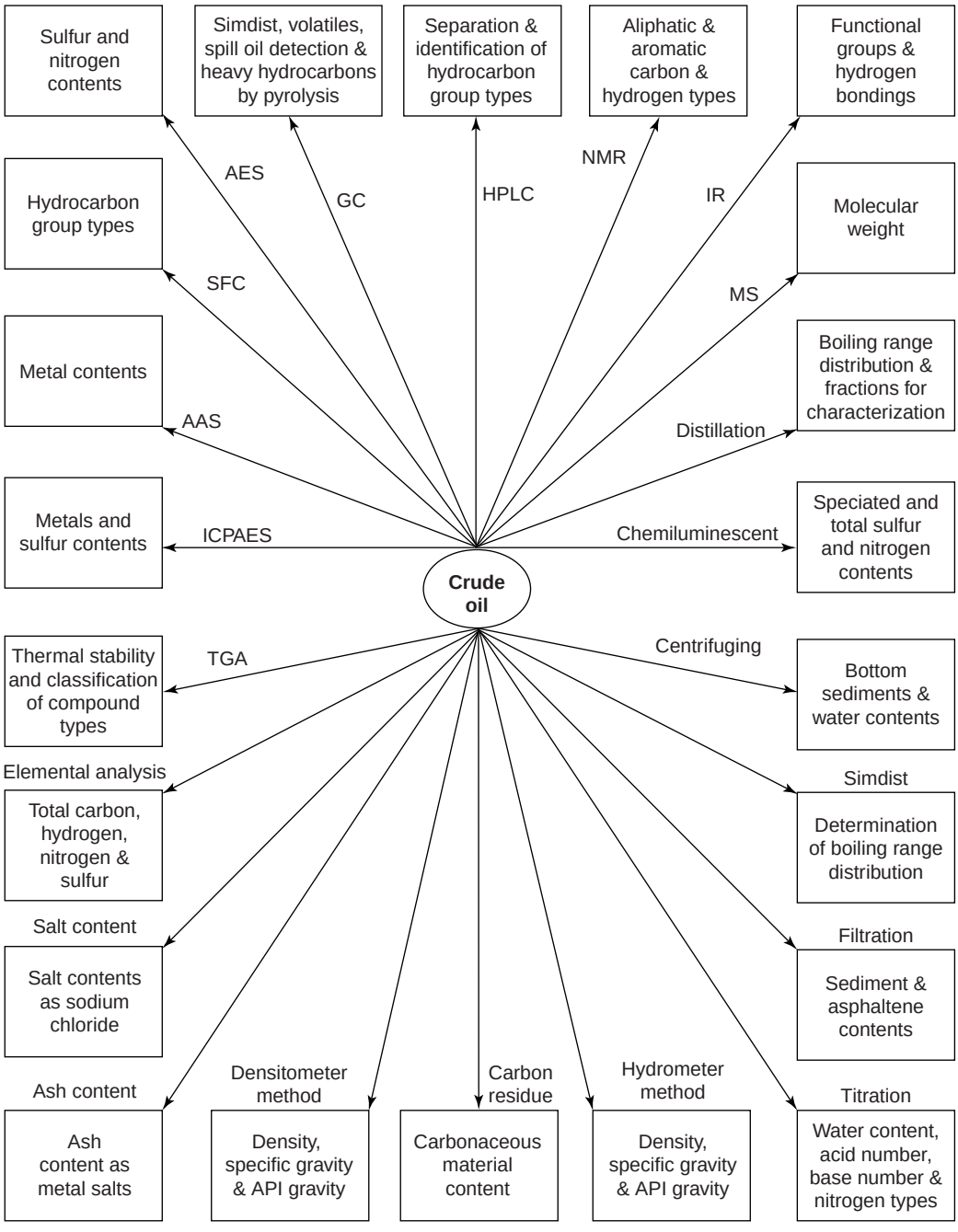

Figure 1 Analytical methods used for the characterization of crude oils. NMR, nuclear magnetic resonance; Simdist, simulated distillation; TGA, thermogravimetric analysis.

5.1 Physical Methods

5.1.1 Density, Specific Gravity and American Petroleum Institute Gravity

The density, specific gravity and API gravity is a rough measure of the quality of a liquid product. A high value of density or specific gravity and a lower value of API gravity would normally mean a heavy crude oil containing a lower percentage of the valuable light and middle fractions and a higher percentage of residue.

The density of a crude oil is the mass to volume ratio and is usually expressed in grams per milliliter ($g\,ml^{-1}$). The density of a crude is measured by a number of methods. These include the hydrometer method[1] in which the crude oil sample is adjusted for its temperature and then poured into a cylinder.

Table 5 Properties of Arab Heavy and Alamein crude oils

Property	Arab Heavy (Saudi Arabia)	Alamein (Egypt)
API gravity at 15.5/15.5 °C (60/60 °F)	28.0	33.4
Carbon residue (wt%)	6.8	4.9
Sulfur content (wt%)	2.8	0.86
Nitrogen content (wt%)	0.15	0.12
Pour point (°C)	−23.8	1.7
Ash content (ppm)	120.0	40.0
Iron (ppm)	1.0	0.0
Nickel (ppm)	9.0	0.0
Vanadium (ppm)	30.0	15.0

The appropriate density range hydrometer is lowered into the sample and allowed to settle until temperature equilibrium is reached. Then, the hydrometer scale is read and the temperature of the sample is recorded. Temperature equilibration is achieved quickly by putting the cylinder in a constant temperature bath. Another method specifies a standard test procedure for density and relative density measurement of crude oils by digital densitometer. In this method, a crude oil sample is introduced into an oscillating sample tube and the change in oscillating frequency with respect to change in mass of the tube is measured. The density of the sample is determined by comparing the measured sample response with the calibration data produced using a standard liquid sample.[2] Specific gravity is measured directly using the ASTM D-1298 method utilizing the hydrometers graduated by specific gravity units. Specific gravity can also be obtained by converting the density values obtained using the above mentioned ASTM D-1298 method. The petroleum measurement tables are used for converting density into specific gravity and API gravity, and vice versa.[3] API gravity is determined using the API hydrometer as the basis of determination.[4] This method is similar to the above mentioned method[1] except that the hydrometers used are graduated by API gravity units. The API gravity is converted to specific gravity and vice versa by employing Equation (1)

$$\text{API gravity, deg} = \left(\frac{141.5}{\text{specific gravity } 60/60\,^{\circ}\text{F}}\right) - 131.5 \quad (1)$$

5.1.2 Pour Point

The pour point of a crude oil is an index of its low temperature mobility. The lowest temperature at which an oil can flow or pour when it is chilled without disturbance under controlled conditions is called the pour point. Pour point is determined by cooling the sample in a glass tube at a specified rate and then examining periodically for its flow characteristics.[5]

5.1.3 Salt Content

There is a great variation in the salt content of crude oils depending mainly on the oilfield source, and possibly on the producing wells or zones within a field. The determination of the salt content of crude oil requires careful sampling. Salt content, expressed as sodium chloride, indicates the amount of salt dissolved in water. Salt content of a crude oil is determined electrometrically and is based on the conductivity of the crude oil solution. In this method, the sample is dissolved in a mixed polar solvent and then subjected to an alternating electrical stress. The flow of current is measured and the salt content is determined by comparing with a calibration curve of current versus salt content for a number of standard mixtures.[6]

5.1.4 Ash Content

The ash content of a crude oil is an indication of metals and salts present as well as dirt and rust. The ash is usually in the form of metal oxides, stable salts and silicon oxides. The ash content is determined by burning a weighed quantity of whole crude oil sample contained in a high temperature resistant container in an atmosphere of air until the residue consisting of ash and carbon remains. The residue left is reduced by heating at 775 °C in a muffle furnace, then cooled and weighed.[7]

5.1.5 Carbon Residue

Carbon residue is a rough indication of the asphaltic compounds and the materials that do not evaporate under conditions of the test, such as metals and silicon oxides. Carbon residue is the residue left after coking and cracking reactions occur during heating of the sample. Conradson carbon residue is determined by subjecting a sample of crude oil to destructive distillation in a crucible for a specified time period under a heating program.[8] Ramsbottom carbon residue is obtained by quick heating of the crude oil sample contained in a glass bulb to 550 °C for a specified time period.[9] Both of these methods provide a variable amount of carbon residue for a given sample. A correlation has been developed by the ASTM which shows that the difference between the two results decreases with increasing the amount of carbon residue in the sample.[9]

5.1.6 Molecular Weight

Molecular weight is determined using a number of methods that include GPC, ultracentrifugation, electron microscope studies and MS. One of the most frequently used methods is the thermoelectric measurement of vapor pressure of a series of solutions of crude oil

in an appropriate solvent. This method is also known as vapor pressure osmometry (VPO). A number of solutions having different concentrations of crude oil are prepared in pure solvent to measure their response. Generally, the solvents used for dissolution should be stable at the temperature of measurement and should not react with the crude oil sample. Some of the solvents frequently used are benzene, toluene, chloroform and 1,1,1-trichloroethane. Prior to proceeding with the crude oil solutions, a calibration curve is developed using a number of standard solutions of benzil in the same solvent in which the sample is dissolved. This calibration curve is used to determine the molecular weight of the crude oil.[10] The value of the molecular weight from VPO must be weighed carefully because the measured value of molecular weight is a function of temperature and the molecular properties of the solvent. Higher dissolution of the crude oil is generally achieved in polar solvents, which leads to dissociation of the molecules and therefore results in low values of molecular weight. On the contrary, nonpolar or less polar solvents allow association of the molecules and therefore high molecular weights are obtained. Another method used for average molecular weight determination is GPC in which the whole crude oil sample is dissolved in an appropriate solvent. Some of the solvents which give best results in terms of solvency and suppression of interfering sorption effects are chloroform and benzene each containing 5% methanol. Other solvents such as trichlorobenzene, pyridine and methylnaphthalene are also used. The solution is filtered and introduced into the column where separation is achieved by the increasing size of the molecules and a plot of elution volume versus detector response is obtained. By comparing the elution volume of a standard sample with an unknown whole crude oil chromatographed under the same conditions, the average molecular weight is determined.[11]

5.1.7 Water Content

Another standard method that utilizes a physical distillation technique is the determination of water present in the crude oil.[12] In this method, a known amount of crude oil sample is mixed with a suitable water-immiscible solvent such as methanol and heated under reflux to boiling. The water is condensed in a graduated trap and measured.

5.2 Extraction, Filtration and Centrifuging Methods

5.2.1 Sediment Content

The sediment present in a crude oil is important for the refining as well as trading point of view. The amount of sediment is determined by extracting with a hot organic solvent in a refractory thimble.[13] Sediment is also determined by the membrane filtration method. In this method, the sample is dissolved in hot toluene and filtered under vacuum through a 0.45-μm porosity membrane filter. The filter with residue is washed, dried and weighed to obtained the amount of sediment present in the sample.[14]

5.2.2 Asphaltene Content

Asphaltenes present in a crude oil are also separated using an *n*-alkane precipitation/extraction method in which crude oil sample is dissolved in an *n*-alkane solvent such as *n*-hexane, and the mixture is stirred for an hour at ambient temperature for dissolution. The *n*-alkane solvent acts as a precipitating agent and the choice of the solvent ranges from *n*-pentane to *n*-decane. In this process, the asphaltenes are precipitated. The resulting mixture is centrifuged at 2000 rpm to separate *n*-alkane-insoluble materials from the soluble portion of the crude oil. Then the solvent containing the soluble portion of the crude oil is decanted. To the insoluble portion, more solvent is added, stirred and the mixture is again centrifuged. This process is repeated until all the insoluble material is separated from the soluble portion of the crude oil. The insoluble portion of the asphaltenes obtained is then soxhlet extracted with the same solvent for 24 h to purify the asphaltenes.[15]

5.2.3 Bottom Sediments and Water

Centrifuging is a process in which the sample contained in a tube is subjected to a force which allows different phases of a mixture to separate from each other. Water and sediments are determined using a method in which a known amount of crude oil is mixed with a water-saturated solvent such as toluene. The mixture is shaken well and then centrifuged in a graduate glass tube at a rate of 1500 to 2000 rpm. The water and sediments are separated from the crude oil, collected at the bottom of the tube and then measured.[16]

5.2.4 Heavy Hydrocarbon Content

Supercritical fluid extraction is a method in which a sample is subjected to extraction by a solvent under supercritical conditions. This method is operated at low temperature using low boiling point solvents such as carbon dioxide as an extracting medium. This method provides the extraction of nondistillable portions of crude oils which are then analyzed using other techniques.[17]

5.2.5 Petroleum Resin Content

Petroleum resins are separated from the whole crude oil or heavy fraction using a liquid chromatographic separation method known as SARA. The abbreviation SARA

is derived from the abbreviations of the main fractions separated from whole crude oil, namely: saturates, aromatics, resins and asphaltenes. In this method, a glass percolating column is packed dry with anionic exchange material Amberlyst 29 at the bottom, and a ferric chloride bound kaolin clay layer in the middle section. The upper part of the ferric chloride bound kaolin clay is packed with anionic exchange material Amberlyst 29 followed by cationic exchange material Amberlyst 15. The column is then purged with nitrogen gas flow followed by *n*-pentane elution and recycling to remove air bubbles and impurities of the packing materials. A weighed quantity of the sample is placed in a cellulose thimble and introduced into the top reservoir of *n*-pentane. The maltenes present in the sample are dissolved in *n*-pentane, leached out of the thimble and carried into the column while *n*-pentane-insoluble asphaltenes remain in the thimble. The resins are retained on the column while the oil remains unretained and collects in the bottom flask. Chloroform elution is then used to remove the resins from the column and then evaporated under nitrogen to recover the resins.[18]

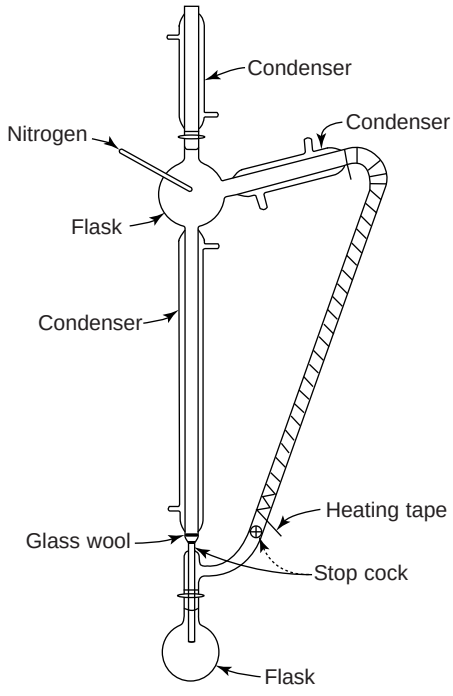

Figure 2 Recycling liquid chromatographic column. (Reprinted from ASTM D-4928,[54] with permission of Marcel Dekker, Inc.)

5.3 Fractionation

Crude oil is a complex mixture of hydrocarbons and nonhydrocarbon-containing sulfur, nitrogen, oxygen and metals such as vanadium, nickel and iron. Crude oil is so complex that it contains basic as well as acidic compounds at the same time. The complexity of crudes increases with increasing boiling point. This is because as the carbon number of molecules increases, the number of isomers increases exponentially. The complexity is also enhanced due to the presence of heteroatomic and metallic compounds. Due to this complexity, crude oils are separated into fractions based upon the type of information required. There are several separation schemes used to produce fractions containing specific compound types. For example, the ion-exchange resins are used to separate acidic and basic compounds whereas ferric chloride–Attapulgus clay complex separates neutral nitrogen compounds.

The United States Bureau of Mines (USBM)–API method[19] provides separation of heavy distillates of a crude oil into saturates, acids, bases, neutral nitrogen compounds and monoaromatic, diaromatic and polyaromatic compounds. This method combines ion-exchange, coordination and adsorption chromatography and the recycling chromatographic column used for this method is illustrated in Figure 2. The crude oil sample is first distilled to produce the distillate which is passed through a glass column having anion resins where the acidic compounds are retained by the resin and other components are washed from the resin by the solvent. The retained acids are removed from the resins using elution with benzene, followed by methanol and finally with methanol saturated with carbon dioxide. Similarly, the acid-free sample is passed through a column of cationic resin to separate basic compounds from the sample. The retained bases are removed from the resin using elution with benzene, followed by methanol and finally with methanol mixed with isopropylamine. The neutral nitrogen compounds are separated by passing the sample through a column having ferric chloride–Attapulgus clay at the top and anion resin at the bottom. The nitrogen compounds are retained by making a complex with the ferric chloride–Attapulgus clay while the aromatics are washed. The retained nitrogen compound–ferric chloride complex is removed by dichloroethane elution and then broken down on contact with anion resin into ferric chloride and nitrogen compounds. The salt is retained by the resin while the nitrogen compounds are recovered. The sample left after acid, base and neutral nitrogen compound removal is chromatographed on a column packed with alumina gel in the bottom and silica gel at the top and eluted with nonpolar to polar solvent. Saturates are eluted using *n*-heptane elution, followed by *n*-heptane containing 5% benzene elution to remove monoaromatics. Diaromatics are separated by increasing the benzene content of the solvent to 15% and finally the polyaromatics are eluted with methanol containing 25% benzene and 25% ether.

5.4 Distillation

The distillation of a crude oil is carried out to determine the boiling range distribution of a crude oil as well as collection of fractions for further characterization using different analytical techniques. This is an important requirement in the preparation of an assay of a crude oil. This information is used to estimate the yield of refinery products and thus the economics of using a crude oil in the refinery. The distillation is carried out by a combination of TBP distillation and vacuum distillation.[20,21] This process is also called physical distillation to distinguish it from simulated distillation. During physical distillation, the crude oil is distilled into various fractions based upon the desired boiling range and it is accomplished in two steps: TBP distillation[20] and distillation under vacuum.[21] The TBP distillation of whole crude oil is achieved using an apparatus having a distillation column equivalent to 16 theoretical plates and operated at a 5 : 1 reflux ratio.[20] The distillation unit is provided with a boiler in which the crude oil is heated, a distillation column, a distilled fraction collector with receivers, a vacuum pump and a refrigeration unit. The distillation unit is equipped optionally with an electronic pressure sensor and a photoelectric level follower to measure the volume of distillate and distillation rate. As a first step, the absorbed gases and moisture present in the crude oil are removed from the sample before distillation and this process is called debutanization. The debutanization is performed under vacuum at 25 °C and the removed gases are trapped under cryogenic conditions. Following the debutanization of crude oil, the distillation is carried out under atmospheric pressure and increased temperature. The temperature of the crude oil in the boiler should not exceed 345 °C. Above this temperature, thermal cracking of the petroleum hydrocarbons is evident and therefore the distillation is carried out under reduced pressure. This is because at this temperature the energy required to distill a fraction exceeds the energy required to break C–C, C–H, C–S and C–N chemical bonds. The TBP distillation is carried out up to 400 °C atmospheric equivalent temperature (AET). The residue from the TBP distillation process is further distilled to produce high boiling distillates and residue.[21] This process, called vacuum distillation, is performed under reduced pressure using equipment consisting of a boiler, short column having a minimum pressure drop, a high vacuum system and a distillate receiver. The vacuum distillation is carried out up to 600 °C AET as a batch process. The collected fractions are measured and the distillation curve is prepared: per cent distilled crude oil as a function of temperature. Another method to carry out distillation under vacuum is called short path distillation. In this method, the preheated residue from atmospheric distillation is pumped at a defined rate onto the surface of a heated vertical cylinder in the evaporation chamber where it is spread by wipers into a thin, uniform and downward-falling film. The lower boiling components of the film are flash evaporated and collected in a condenser, while the higher boiling components move downward and are collected as the residue. As compared to vacuum distillation using a packed column, the short path distillation is carried out under higher vacuum, a continuous process, shorter residence time, and a shorter distance between heated and condensing surface. These factors allow very deep distillation of the residue without decomposition and a distillate fraction up to 700 °C AET is easily prepared.

5.5 High-performance Liquid Chromatography

HPLC of whole crude oil is performed to determine the amount of saturates (paraffins and naphthenes), aromatics and polars. The crude oil is first treated with an *n*-alkane solvent such as *n*-hexane to precipitate the asphaltenes. The resulting mixture is filtered through a 0.5-μm membrane filter to remove asphaltenes (*n*-alkane insolubles) from maltenes (*n*-hexane solubles). The maltenes are further fractionated using HPLC into saturates, aromatics and polars on an aminopropyl packed column. The saturates are eluted first followed by aromatics. When the aromatics are completely eluted the column is backflushed to elute the polar fraction from the column. A complete analytical HPLC run requires 30 min to elute saturates, aromatics and polars. HPLC is operated both in analytical and preparative modes. A typical HPLC system consists of sample injector, solvent delivery system, column, ultraviolet detector and a differential refractometer detector. The analytical HPLC method is used to obtain areas of saturates, aromatics and polars present in the sample solution. The saturates are detected by refractometer detector while the aromatics and polars are detected by ultraviolet detector.[22] In order to collect the fraction from HPLC for further characterization using other techniques, the preparative scale HPLC method is used. The preparative HPLC method employs wider bore columns in which a higher amount of sample can be injected, separated and the fractions collected. The collected fractions are purified by removing solvent under reduced pressure and then weighed. In another method, the whole crude oil is separated into six hydrocarbon types using a normal phase multidimensional HPLC method equipped with two columns and two switching valves. The detection of the fractions is achieved using two detectors: photodiode array detector (PAD) and evaporative light scattering detector (ELSD). These six hydrocarbon types are

1. saturates;
2. monoaromatics (1-ring);
3. diaromatics (2-ring);
4. triaromatics (3-ring);
5. quatroaromatics (4-ring);
6. polars which include 5+ rings and N- and O-heterocyclics.

The elution of components is carried out using a gradient elution program in which three solvents, namely *n*-hexane, dichloromethane and isopropanol, are used. Pure *n*-hexane is used to separate saturates and monoaromatics on a column packed with propylaminocyano (PAC) stationary phase. The aromatics are separated on a dinitroanilinopropyl (DNAP) packed column and the polarity of the solvent is gradually increased by increasing the concentration of dichloromethane in *n*-hexane. When all the aromatics are eluted, a mixture of dichloromethane and isopropanol is used to elute the polars from the column. Both columns are regenerated using a reverse elution program using all three solvents.[23]

5.6 Gas Chromatography

GC of a crude oil is performed to obtain different types of information which include determination of boiling range distribution using simulated distillation, individual component identification and identification of the source of a spilled crude oil using fingerprint chromatograms.

5.6.1 Determination of Boiling Range Distribution by Simulated Distillation Method

Simulated distillation[24] is a method in which the process of physical distillation is simulated using gas chromatographic separation of the components of the sample. This method determines the boiling range distribution of a moisture-free crude oil up to 538 °C and the material boiling above this temperature is termed residue. In this method, the crude sample is dissolved in carbon disulfide and injected into a gas chromatographic column which separates the hydrocarbons according to their boiling points. Both packed columns as well as capillary are used for separating the components. The column temperature is increased in a linear fashion and the area under the whole chromatogram is recorded. Prior to the sample run, a standard mixture of *n*-paraffins (*n*-C_5 to *n*-C_{40}), having a boiling range from 36 to 525 °C, is chromatographed under the same conditions to provide a correlation between the retention time and boiling point.[25] A comparison of the chromatogram of standard with the sample assigned boiling points to the time axis of the sample components. The amount of residue is estimated by making a second run of the crude oil sample having a known amount of internal standard consisting of C_{14}, C_{15}, C_{16} and C_{17} normal paraffins. Using these data, the boiling range distribution of the crude oil sample is calculated and reported as percent distilled versus temperature including IBP and FBP of the sample. IBP and FBP are the temperatures at which 0.5% and 95.5% of the material is eluted from the column, respectively. Figure 3 shows a simulated distillation chromatogram of Arab Berri crude oil. This method is developed further to increase the FBP of the method up to 750 °C and it is called a high temperature simulated distillation. With the introduction of high temperature resistance materials for column preparation, stationary phases and GC oven construction material, the simulated distillation is capable of analyzing crude oils up to C_{120}. In this method, an ultimetal capillary column having an operation temperature stability of 450 °C is used. This column is coated with linear polydimethylsiloxane as stationary phase and temperature programmed up to 450 °C; the hydrocarbons with boiling point up to 750 °C are then separated and identified.[26,27] In order to obtain the boiling range distribution of sulfur compounds in crude oils along with hydrocarbons, a post-column splitter is utilized to achieve detection simultaneously by both a sulfur chemiluminescent detector (SCD) and a flame ionization detector (FID).[28] A comparison of simulated distillation[24] with physical distillation[20,21] to determine the boiling range distribution of a crude oil shows that the simulated distillation incorporates a high level of automation and thus requires very little operator involvement. Besides, a single boiling range distribution analysis using physical distillation requires two to five

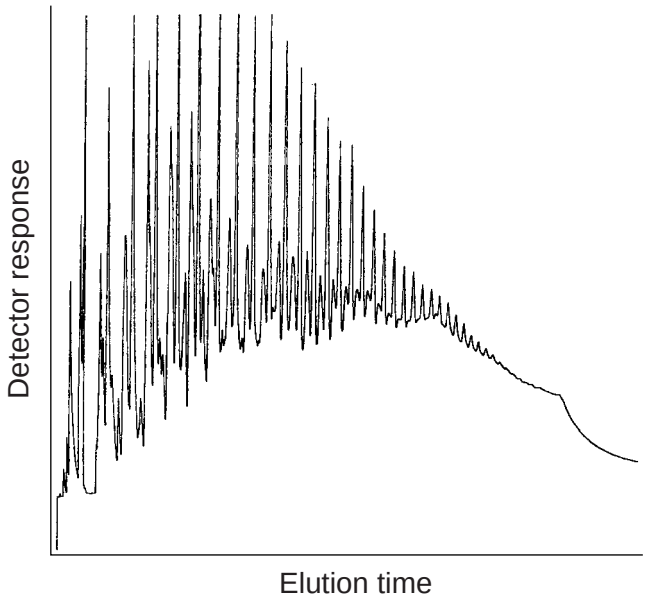

Figure 3 Simulated distillation chromatogram of Saudi Arabian Berri crude oil.

days whereas a complete simulated distillation analysis needed 30 min. The automation in simulated distillation analysis assures a reproducibility of 2–3 °C compared to 10–30 °C in the physical distillation process. The faster analysis time, better reproducibility and equivalent results are some of the advantages of simulated distillation over physical distillation.[29]

5.6.2 Volatile Compound Identification

The individual component identification of a crude oil is performed for volatiles using a backflushing valve system coupled with a GC system. The whole crude oil is introduced into a system with a backflushing valve. When the volatile elution is complete, the sample injected is backflushed to remove the heavy components from the column. Fast analysis of light hydrocarbons present in crude oils is performed using the switching valve, separation at sub-ambient conditions and fast temperature programming. Standard mixtures of hydrocarbons containing volatiles are used to determine the appropriate temperature and switching valve timing of the method. The crude oil is injected and the hydrocarbons up to nonane are eluted while the heavier components are prevented from entering the analytical column and sent to a system vent. This technique permits separation at sub-ambient temperatures in minimum time and the life of the column is extended.[30]

5.6.3 Oil Spill Identification

GC is also used to identify the source of oil spill. In this method, the whole crude oil is injected into a GC column with an FID. This technique provides a chromatogram having a pattern of peaks characteristic of that particular crude provided that the operating conditions remains same. An unknown or suspected spilled crude oil sample is identified by comparing its chromatogram with the standard chromatograms available for potential source crude oils. Using this technique, the source of oil spill is identified. Two-dimensional GC is used to provide better separation and identification of the crude oil components which provide both quantitative and qualitative comparison between the oil spill and the potential source crude.[31] The two-dimensional GC is combined with MS to identify the individual compounds present.[32]

5.6.4 Pyrolysis Gas Chromatography

PGC (pyrolysis gas chromatography) is a technique which provides a rapid assay of crude oil without initial treatment. This technique provides fingerprints of different crude oils for comparison purposes. In PGC, the crude oil sample is heated in a chamber to break the large molecules into smaller volatile molecules which are then separated in a gas chromatograph and characterized easily using an FID. Other detectors such as an NPD (nitrogen phosphorus detector) or an FPD (flame photometric detector) are also used for nitrogen and sulfur compound identification. A comparison of PGC and MS shows a similarity between the two techniques. In PGC, the large molecules are broken down into smaller molecules using thermal energy, whereas in MS this is done using electron impact or chemical ion impact. The molecular fragments are separated in PGC using GC whereas separation of ions is achieved in MS using a strong magnetic field. In PGC, the detection is achieved using an FID and appeared in the form of a chromatogram, whereas in MS the ions are collected and registered in the form of a profile of ions generated which is called a mass spectrum. PGC can be combined with MS to provide a better identification of the fragments obtained during pyrolysis of crude oil.[33]

5.7 Supercritical Fluid Chromatography

SFC is a chromatographic method in which the sample dissolved in carbon disulfide is injected onto a packed silica adsorption column and eluted using a highly pressurized mobile phase such as carbon dioxide. Different compounds of the sample are separated and detected using a suitable detector such as an FID. Compared to GC, SFC has the advantage of possessing a much higher solubilizing power than a gas which allows higher molecular weight and ionic compounds to be chromatographed at lower temperatures. In comparison to LC, the SFC mobile phase exhibits more desirable transport properties resulting in a significant reduction in the resistance to mass transfer in the mobile phase. Other mobile phases used in SFC include nitrous oxide, methanol, *n*-propane, *n*-butene, *n*-pentane, dichlorodifluoromethane and trichlorofluoromethane. SFC is used in the separation of polynuclear aromatic hydrocarbons and vacuum residue as well.[34]

5.8 Mass Spectrometry

In MS, the crude oil sample is placed in a vacuum chamber where a variety of methods are applied to convert the molecules into ions. These methods include fragmenting methods such as electron impact mass spectrometry (EIMS) as well as nonfragmenting techniques of field desorption mass spectrometry (FDMS) and field ionization mass spectrometry (FIMS). The ions are passed through a magnetic field and are separated according to their mass/charge ratios. The ions are collected and registered for the type and amount present in the stream. The pattern obtained after detection of ions, called the mass spectrum, is then compared with the known mass spectrum of pure compounds to find the type of compounds present in crude

oil. The mass spectrum of a crude oil is very complex since thousands of compounds belonging to different classes are present. In practice, mass spectra of the crude oil fractions are obtained which are quite useful and provide valuable information regarding the components present.[35] Low voltage high resolution is another MS method applied to distinguish between molecules of equal masses. Using this method, a nitrogen atom is easily distinguished from a CH_2 group.[36,37] The mass spectrometer coupled with GC as a detector is called GC/MS. In this technique, the compounds eluted from the GC column enter into the mass spectrum and are fragmented into ions. These ions are detected and produce a profile or mass spectrum of the compound which is then compared with mass spectra of known compounds to find its identity. The GC/MS of the whole crude is impractical since the heavy hydrocarbons present in the crude oils are not eluted from GC.

5.9 Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy analysis of a crude oil is performed to determine the type and amount of hydrogen and carbon present. NMR analysis can distinguish whether the hydrogen or carbon is attached to an aliphatic, naphthenic or aromatic type of hydrogen or carbon atom. This leads to the determination of percent paraffinic, naphthenic and aromatic hydrogen and carbon present in the crude oil. This information is quite helpful to differentiate paraffinic, naphthenic and asphaltic types of crude oils. NMR spectroscopy provides structural parameters of crude oils which can produce a detailed analysis when combined with other properties. NMR spectroscopy is performed by dissolving the sample in deuterated chloroform and adding trimethylsilane as a reference. The deuterated chloroform acts as a solvent as well as a locking agent. An NMR spectrum of crude oil is achieved by recording a magnetic resonance signal of a diluted sample in the presence of a very strong magnetic field. The combination of proton and carbon-13 NMR along with elemental analysis of crude oil allows the determination of a number of average structural parameters such as total aromatic and aliphatic carbon, naphthenic and paraffinic carbon, the number and length of paraffinic carbon chains, the degree of chain branching, the number of aromatic methyl substituents, the aromatic ring size and the degree of aromatic ring substitution.[38,39]

5.10 Infrared Spectroscopy

IR spectroscopy is used to characterize crude oils mostly in terms of oxygen and nitrogen containing functional groups. Using IR spectroscopy, acids and bases such as carboxylic acids, phenols, carbazoles and cyclic amides present in the petroleum mixtures are quantified. The type of solvent in which the crude oil is dissolved has a marked effect on the detection of different compounds. For example, tetrahydrofuran (THF) is used as a solvent to minimize association between carboxylic acids and amides and thus help in their detection in the IR spectrum.[40,41] The concentration of hydroxyl compounds is identified using the peak absorption in the region $3540–3600\,cm^{-1}$ of the IR spectrum. IR spectroscopy also determines the hydrogen bondings present between different functional groups such as OH and NH. This is done by recording the IR spectrum of the petroleum mixture at different concentrations in a number of solvents. At low concentration, the hydrogen bonding is less pronounced and the peaks of free OH ($3610\,cm^{-1}$) and NH ($3480\,cm^{-1}$) become visible.[42] IR spectroscopy is also utilized in finding the distribution of H atoms in the aromatic rings and that of CH_2 and CH groups in the aliphatic part of the hydrocarbons.

5.11 Thermal Analysis

Thermal analysis is used to study the characteristics of crude oils. During thermal analysis, a known quantity of crude oil is subjected to programmed heating in the presence of air or inert gas such as nitrogen. The crude oil sample loses moisture and volatile components by evaporation and the process behaves like the physical distillation during the initial heating. During the evaporation process, the heat is required by the molecules and thus endothermic peaks are observed in the thermogram. When the temperature of the crude oil exceeds 340 °C, thermal degradation of the oil is started due to thermal cracking of the hydrocarbons. This is because at higher temperatures the energy provided to the crude oil is enough to break chemical bonds such as C–H, C–C, C–N and C–S present in the petroleum molecules. During thermal analysis, a number of phenomena such as evaporation, decomposition and oxidation take place in the crude oil. Oxidative degradation is characterized by an exothermic peak with corresponding weight loss. Evaporation requires heat and therefore appears as an endothermic peak in the thermogram. Using thermal analysis, the hydrocarbons of a crude oil can be classified into four types; the volatiles and low molecular weight, medium molecular weight and high molecular weight hydrocarbons. The volatiles are characterized by heat absorption leading to evaporation of these hydrocarbons. The other three types of hydrocarbons are identified by the presence of exothermic peaks in the thermogram as a result of oxidative degradation of these hydrocarbons. The volatile contents of crude oils are found more in lighter crude compared to heavier crude.[43]

5.12 Elemental Analysis Method

The analysis of carbon, hydrogen, nitrogen and sulfur is very important in determining the complex nature and

type of the crude oils. Nitrogen and sulfur are very important from the processing and upgrading point of view and are determined to assess the performance of petroleum upgrading processes such as HDS and HDN. In a single instrumental method, carbon, hydrogen, nitrogen and sulfur are determined simultaneously. One of the standard methods[44] describes the determination of carbon, hydrogen and nitrogen in petroleum products. The latest development also provides determination of sulfur along with carbon, hydrogen and nitrogen. In this method, the crude oil sample is introduced into the system where oxidative combustion takes place in an oxygen–helium atmosphere. As a result, the elements C, H, N and S are converted into CO_2, H_2O, NO, NO_2, SO_2 and SO_3 product gases. These gases are passed over a copper reduction tube where oxides of nitrogen are reduced to molecular nitrogen and sulfur oxides are converted to SO_2. After the reduction process, the gas mixture contains CO_2, H_2O, N_2 and SO_2. These components are then separated using an adsorption process in which each component is adsorbed on a separate adsorption column and finally detected individually by a thermal conductivity detector (TCD). The quantification of the elements is achieved through an external standard method.

The total sulfur content of crude oils and heavy products is also determined by burning a sample in a stream of air. The sulfur dioxide produced is further oxidized to sulfuric acid, which is titrated with a standard alkali. Identification of individual sulfur compounds in crude oils is not needed from the refinery point of view since it is time-consuming and all sulfur compounds practically are hydrodesulfurized to hydrogen sulfide and hydrocarbons. Total sulfur in whole crude oil is determined using a variety of methods which include XRF spectroscopy[45] and the bomb method.[46] All of these methods have specific limitations as well as variable detection limits. In XRF spectroscopy, the sample is placed in an X-ray beam and the resulting excited characteristic radiation is measured. By comparing the signal from the sample with multilevel concentration calibration samples, the amount of sulfur present in the sample is determined. In the bomb method, the sample is oxidized by electrical ignition and combustion in a pressurized bomb filled with oxygen. All the sulfur present in the sample is converted to oxides of sulfur which are then converted to barium sulfate and determined gravimetrically.

5.13 Atomic Absorption Spectrometry

AAS is used to determine quantitatively the amount of metals present in crude oils. Generally, two methods are used for sample preparation. The crude oil sample is mixed with a suitable solvent such as mixed xylenes to produce a dilute solution. In the other method, the crude oil is heated to burn and the carbon is removed finally by heating at high temperature in a furnace. The residue is fused with a lithium tetraborate/lithium fluoride mixture. The fused sample is digested in hydrochloric acid solution and diluted with water. The diluted solution containing the metals is then aspirated into an atomic absorption spectrometer having a lamp for a particular metal or group of metals. The absorption intensity of the solution is measured individually for all the elements. Prior to the measurement of each metal, the spectrometer is calibrated using a standard solution containing the particular metal. By comparing absorption intensities of the elements present in the crude oil solution with absorption intensities measured with the standard calibration solutions, the concentrations of elements in the crude oil sample are calculated. Using this technique, almost all of the metals present in the crude oil are determined. These include aluminum, barium, boron, calcium, chromium, copper, vanadium, nickel, iron, magnesium, phosphorus, potassium, sodium, silicon and zinc.[47]

5.14 Inductively Coupled Plasma Atomic Emission Spectrometry

ICPAES is used to determine quantitatively the amount of metals present in crude oils. Generally, two methods are used for sample preparation. The crude oil sample is mixed with a suitable solvent such as mixed xylenes to produce a dilute solution. In another method, the crude oil is heated to burn and the carbon is removed finally by heating at high temperature in a furnace. The residue is fused with a lithium tetraborate/lithium fluoride mixture. The fused sample is digested in hydrochloric acid solution and diluted with water. The diluted solution is then aspirated into the ICPAES system and the emission intensities of all of the spectral lines are measured. Each spectral line represents an element while its intensity indicates its concentration present in the solution. By comparing emission intensities of elements present in the crude oil solution with emission intensities measured with the standard calibration solutions, the concentrations of elements in the crude oil sample are calculated. Using ICP, almost all of the metals present in the crude oil are determined. These include aluminum, barium, boron, calcium, chromium, copper, vanadium, nickel, iron, magnesium, phosphorus, potassium, silicon and zinc.[48,49] AES is also utilized alone for the detection and measurement of sulfur and nitrogen in crude oils. Contrary to FPD, the AES sulfur response is linear, equimolar and exhibits little or no quenching. These characteristics help in determining the concentration of total sulfur as well as individual sulfur compounds. The calibration of the detector is done using a single sulfur compound.[50] Using this detector, sulfur and nitrogen contents at low ppm levels are measured.[51]

5.15 Chemiluminescent Detection Method

A chemiluminescent detector offers rapid quantitation of speciated sulfur containing compounds in crude oils or the total sulfur present. The detector is set up either to quantify compounds separated from the crude oil or for total sulfur measurement. In this method, the whole crude is separated from asphaltenes and injected into the HPLC column where it is separated into different hydrocarbon types. The sulfur compounds present in these separated fractions are then detected by the chemiluminescent detector which is internally or externally calibrated. These compound types are detected for the presence of sulfur on a linear and equimolar basis.[52]

5.16 Titration Methods

A number of titration methods are used to determine different properties of a crude oil. These methods include determination of water content and acid and base compound concentration in a crude oil.

5.16.1 Water Content

The determination of water content is very important in the processing, transportation and purchase of a crude oil. Water content of a whole crude oil is determined using a number of methods depending upon the amount of water present in the sample. Trace quantities of water are determined by using a titration method in which the sample is titrated using Karl Fisher reagent to an electrometric end-point.[53] Ultratrace amounts of water are determined using the coulometric Karl Fisher titration method. In this method, a weight quantity of a properly homogenized crude oil sample is injected into the titration vessel of the Karl Fisher equipment in which iodine is generated coulometrically at the anode. When all the water is titrated, the excess iodine is detected by an electrometric end-point detector and the titration is complete. Since 1 mol of water reacts with 1 mol of iodine stoichiometrically, the quantity of water present in the sample is proportional to the total integrated current.[54]

5.16.2 Acid Number

The acid number of a crude oil is measured to determine the acidic constituents present, which include carboxylic acids and phenols. Acid number is defined as the measure of the amount of acidic substances present in the oil which can be determined by titration with bases under the given test conditions. In this method, a known amount of crude oil sample is dissolved in a mixture of toluene and isopropyl alcohol containing a small amount of water and titrated potentiometrically up to the end-point with alcoholic potassium hydroxide standard solution using a glass indicating electrode and calomel reference electrode.[55]

5.16.3 Base Number

The crude oils contain basic constituents such as amides, pyridines, carbazoles, etc. Base number is defined as the measure of the amount of basic substances present in the oil which can be determined by titration with acids under the given test conditions. In this method, a known amount of sample is dissolved in an anhydrous mixture of chlorobenzene and glacial acetic acid and titrated with a solution of perchloric acid in glacial acetic acid using a potentiometric titrator. The potential readings are measured by means of a glass indicating electrode and a calomel reference electrode and the end-point is detected.[56] The types of nitrogen compounds are determined using a scheme in which the sample is dissolved in a mixture of acetic anhydride and benzene in the ratio 2 : 1 and titrated with 0.05 M perchloric acid prepared in 1,4-dioxane using a glass indicating electrode and a calomel reference electrode system. A number of end-points are detected which indicate the types of nitrogen compounds titrated. Using this method, the nitrogen compounds are classified into strong bases, weak bases and nonbasic types.[57]

ACKNOWLEDGMENTS

The author wish to acknowledge the support of the Center for Refining and Petrochemicals, Research Institute of the King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
AET	Atmospheric Equivalent Temperature
API	American Petroleum Institute
ASTM	American Society for Testing and Materials
DNAP	Dinitroanilinopropyl
DTA	Differential Thermal Analysis
DTG	Differential Thermogravimetry
EIMS	Electron Impact Mass Spectrometry
ELSD	Evaporative Light Scattering Detector
FBP	Final Boiling Point
FDMS	Field Desorption Mass Spectrometry
FID	Flame Ionization Detector
FIMS	Field Ionization Mass Spectrometry
FPD	Flame Photometric Detector

GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
HDN	Hydrodenitrogenation
HDS	Hydrodesulfurization
HPLC	High-performance Liquid Chromatography
IBP	Initial Boiling Point
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
IP	Institute of Petroleum
IR	Infrared
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NPD	Nitrogen Phosphorus Detector
PAC	Propylaminocyano
PAD	Photodiode Array Detector
PGC	Pyrolysis Gas Chromatography
SARA	Saturates, Aromatics, Resins and Asphaltenes
SCD	Sulfur Chemiluminescent Detector
SFC	Supercritical Fluid Chromatography
TBP	True Boiling Point
TCD	Thermal Conductivity Detector
THF	Tetrahydrofuran
USBM	United States Bureau of Mines
VPO	Vapor Pressure Osmometry
XRF	X-ray Fluorescence Spectroscopy

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)** Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Fuel Performance Specifications, Mid-infrared Analysis of • Fuels Analysis, Regulatory Specifications for • High-temperature Simulated Distillation Applications in Petroleum Characterization • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Lube Products. Molecular Characterization of Base Oils • Lubricant Base Oils: Analysis and Characterization of • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of • Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels • Nuclear Magnetic Resonance Characterization of Petroleum • Oil Shale and Shale Oil Analysis • Oxygenate Vapor–Liquid Equilibrium in Gasolines • Petroleum Residues, Characterization of • Refractive Index Technology as a Real Time Viscosity Technique • Use of Inspection Properties to Predict Hydrocarbon Fraction Physical Properties

REFERENCES

1. ASTM D-1298, 'Test Method for Density, Relative Density or API Gravity of Crude Petroleum by Hydrometer Method', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
2. ASTM D-5002, 'Test Method for Density and Relative Density of Crude Oils by Digital Analyzer', Volume 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
3. ASTM D-1250, 'Guide for Petroleum Measurement Tables', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
4. ASTM D-287, 'Test Method for API Gravity of Crude Petroleum and Petroleum Products (Hydrometer Method)', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
5. ASTM D-97, 'Test Method for Pour Point of Petroleum Oils', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
6. ASTM D-3230, 'Test Method for Salts in Crude Oils (Electrometric Method)', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
7. ASTM D-482, 'Test Method for Ash from Petroleum Products', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
8. ASTM D-189, 'Test Method for Conradson Carbon Residue of Petroleum Products', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
9. ASTM D-524, 'Test Method for Ramsbottom Carbon Residue Petroleum Products', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
10. ASTM D-2503, 'Test Method for Relative Molecular Weight of Hydrocarbons by Thermoelectric Measurement of Vapor Pressure', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
11. K.H. Altgelt,'Gel Permeation Chromatography', in *Chromatography in Petroleum Analysis*, eds. K.H. Altgelt, G.H. Gouw, Chromatographic Science Series, Marcel Dekker Inc., New York, USA, Chapter 12, 287–312, Vol. 11, 1979.

12. ASTM D-95, 'Test Method for Water in Petroleum Products and Bituminous Materials by Distillation', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
13. ASTM D-473, 'Test Method for Sediments in Crude Oils and Fuels Oils by Extraction Method', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
14. ASTM D-4807, 'Test Method for Sediments in Crude Oils by Membrane Filtration', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
15. IP 143, 'Test Method for *n*-Heptane Insoluble Asphaltenes Determination', IP Book of Standards, London, United Kingdom, 1999.
16. ASTM D-96, 'Test Method for Water and Sediments in Crude Oils by Centrifuge Method (Field Method)', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
17. I.D. Singh, V. Ramaswami, V. Kothiyal, L. Brouwer, D. Severin, 'Structural Studies of Short Path Distillates and Supercritical Fluid Extract of Petroleum Short Residue by NMR Spectroscopy', *Fuel Sci. Technol. Int.*, **10**(2), 267–280 (1992).
18. K.H. Altgelt, D.M. Jewell, D.R. Latham, M.L. Selucky, 'Separation Schemes', in *Chromatography in Petroleum Analysis*, eds. K.H. Altgelt, G.H. Gouw, Chromatographic Science Series, Marcel Dekker Inc., New York, USA, Chapter 9, 185–214, Vol. 11, 1979.
19. M.F. Ali, M.A. Ali, 'Investigation of Nitrogen Compound Types in High-Boiling Petroleum Distillates from Saudi Arabian Crude Oils', *Fuel Sci. Technol. Int.*, **6**(3), 259–290 (1988).
20. ASTM D-2892, 'Test Method for Distillation of Crude Petroleum', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
21. ASTM D-1160, 'Test Method for Distillation of Crude Petroleum', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
22. M.A. Ali, W.A. Nofal, 'Application of High Performance Liquid Chromatography for the Determination of Hydrocarbon Group Types of Saudi Arabian Crude Oils', *Fuel Sci. Technol. Int.*, **12**(1), 21–33 (1994).
23. W.K. Robbins, 'Quantitative Measurement of Mass and Aromaticity Distribution for Heavy Distillates 1. Capabilities of the HPLC-2 System', *J. Chromatogr. Sci.*, **36**, 457–466 (1998).
24. ASTM D-5307, 'Test Method for Determination of Boiling Range Distribution of Crude Petroleum by Gas Chromatography', section 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
25. N. Kosal, M.A. Ali, A. Bhairi, 'Simulated Distillation of Saudi Arabian Crude Oils by Gas Chromatography', Special Issue of *J. King Abdul Aziz Univ. Eng. Sci.*, 267–273 (1993).
26. D. Villanti, D. Janson, P. Colle, 'Hydrocarbon Characterization by High Temperature Simulated Distillation (HTSD)', Presented at the AIChE Spring National Meeting, Houston, TX, March, 19–23, 1995.
27. J. de Zeeuw, J. Buyten, 'Extended Simulated Distillation up to C_{120} Using Metal Capillary Column with a Temperature Stability up to 450 °C', Paper No. 118, Presented at the Pittcon Conference, New Orleans, Los Angeles, USA, March, 1999.
28. X. Yan, J.-F.A. Borny, 'Simultaneous Nitrogen and Sulfur Simdist, Two Elements with One Detector', Paper No. 1228, Presented at the Pittcon Conference, New Orleans, Los Angeles, USA, March, 1999.
29. D. Villanti, J. Maynard, J. Raia, A. Arias, 'Yield Correlation Between Crude Assay Distillation and High Temperature Simulated Distillation (HTSD)', Presented at the AIChE Spring National Meeting, Houston, TX, March, 9–13, 1997.
30. S.J. Macdonald, K.E. Labneck, D.B. Wheeler, 'Fast Analysis of Light Hydrocarbons in Crude Oil', Paper No. 114, Presented at the Pittcon Conference, New Orleans, Los Angeles, USA, March, 1999.
31. R.B. Gaines, G.S. Frysinger, 'Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography', Paper No. 113, Presented at the Pittcon Conference, New Orleans, Los Angeles, USA, March, 1999.
32. G.S. Frysinger, R.B. Gaines, 'Analysis of Petroleum Fuels by Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry Detection', Paper No. 112, Presented at the Pittcon Conference, New Orleans, Los Angeles, USA, March, 1999.
33. E.J. Gallegos, 'Pyrolysis Gas Chromatography', in *Chromatography in Petroleum Analysis*, eds. K.H. Altgelt, G.H. Gouw, Chromatographic Science Series, Marcel Dekker Inc., New York, USA, 163–184, Vol. 11, 1979.
34. G.H. Gouw, R.E. Jentoff, 'Supercritical Fluid Chromatography', in *Chromatography in Petroleum Analysis*, eds. K.H. Altgelt, G.H. Gouw, Chromatographic Science Series, Marcel Dekker Inc., New York, USA, 313–328, Vol. 11, 1979.
35. M.M. Boduszynski, 'Composition of Heavy Petroleum. 2. Molecular Characterization', *Energy Fuels*, **2**, 697–713 (1988).
36. K.H. Altgelt, M.M. Boduszynski, 'Molecular Characterization of Heavy Petroleum Fractions by Mass Spectrometry', in *Composition and Analysis of Heavy Petroleum Fractions*, Marcel Dekker Inc., Chapter 7, 257–305, 1994.
37. T. Aczel, D.E. Allen, J.R. Harding, E.A. Knipp, 'Computer Techniques for Quantitative High Resolution Mass Spectral Analysis of Complex Hydrocarbon Mixtures', *Anal. Chem.*, **42**, 341–347 (1970).

38. J.K. Brown, W.R. Ladner, 'A Study of the Hydrogen Distribution in Coal like Materials by High Resolution NMR Spectroscopy I – the Measurement and Interpretation of the Spectra', *Fuel*, **36**, 79–86 (1960).
39. J.K. Brown, W.R. Ladner, 'A Study of the Hydrogen Distribution in Coal like Materials by High Resolution NMR Spectroscopy II – a Comparison with Infrared Measurement and the Conversion to Carbon Structure', *Fuel*, **39**, 87–96 (1960).
40. J.F. McKay, T.E. Cogswell, J.H. Werber, D.R. Latham, 'Analysis of Acids in High Boiling Distillates', *Fuel*, **54**, 50–61 (1975).
41. J.F. McKay, J.H. Werber, D.R. Latham, 'Characterization of Nitrogen Bases in High Boiling Distillates', *Anal. Chem.*, **48**, 891–898 (1976).
42. J.C. Petersen, 'An Infrared Study of Hydrogen Bonding in Asphalt', *Fuel*, **46**, 295–305 (1967).
43. M.A. Ali, M.A.B. Siddiqui, S.M.J. Zaidi, 'Thermal Analysis of Crude Oils and Comparison with SIMDIST and TBP Distillation Data', *J. Therm. Anal.*, **51**, 307–319 (1998).
44. ASTM D-5291, 'Test Method for Instrumental Determination of Carbon, Hydrogen and Nitrogen in Petroleum Products and Lubricants', Volume 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
45. ASTM D-4294, 'Test Method for Sulfur in Petroleum Products by Energy-Dispersive X-Ray Fluorescence Spectroscopy', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
46. ASTM D-129, 'Test Method for Sulfur in Petroleum Products (General Bomb Method)', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
47. ASTM D-4628, 'Test Method for Analysis of Barium, Calcium, Magnesium and Zinc in Unused Lubricating Oils by Atomic Absorption Spectrometry', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
48. ASTM D-5600, 'Test Method for Trace Metals in Petroleum Coke by ICP-AES', Volume 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
49. ASTM D-5708, 'Test Method for Determination of Nickel, Vanadium and Iron in Crude Oils and Residual Fuels by Trace Metals in Petroleum Coke by ICP-AES', Volume 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
50. ASTM D-5623, 'Test Method for Sulfur in Light Petroleum Liquids by Gas Chromatography and Sulfur Selective Detection', Volume 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
51. B.D. Quimby, D.A. Grudoski, V. Giarrocco, 'Improved Measurement of Sulfur and Nitrogen Compounds in Refinery Liquids Using Gas Chromatography-Atomic Emission Detection', *J. Chromatogr. Sci.*, **36**, 435–443 (1998).
52. D.-J.-F.A. Borny, M. Homan, J. Werner, 'Speciated and Total Sulfur in Crude Oil and Heavy Oil by HPLC', Paper No. 1097, Presented at the Pittcon Conference, New Orleans, Los Angeles, USA, March, 1999.
53. ASTM D-4377, 'Test Method for Water in Crude Oils by Potentiometric Karl Fischer Titration', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
54. ASTM D-4928, 'Test Method for Water in Crude Oils by Coulometric Titration', Volume 5.03, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
55. ASTM D-664, 'Test Method for Acid Number of Petroleum Products by Potentiometric Titration', Volume 5.01, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
56. ASTM D-2896, 'Test Method for Base Number of Petroleum Products by Potentiometric Perchloric Acid Titration', Volume 5.02, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, USA, 1999.
57. M.F. Ali, M.A. Ali, 'Nonaqueous Potentiometric Titration and Elemental Analysis of High Boiling Distillates of Saudi Arabian Crude Oils', *Fuel Sci. Technol. Int.*, **6**(6), 663–686 (1988).

High-temperature Simulated Distillation Applications in Petroleum Characterization

Dan C. Villalanti and Joseph C. Raia
Triton Analytics Corporation, Houston, Texas, USA

Jim B. Maynard
Shell Development Company, Houston, Texas, USA

1 Introduction 2
2 History 2
3 Comparison of American Society for Testing and Materials Distillation and Simulated Distillation Methods 2
3.1 Physical Distillation Methods 2
3.2 Simulated Distillation Methods 3
4 Methodology 3
4.1 Sample Preparation 4
4.2 High-temperature Simulated Distillation Analysis 4
4.3 Calibration 5
4.4 Determination of Hydrocarbon Yield as a Function of Boiling Point 5
5 Quality Control and Instrument Performance 5
5.1 Statistical Analysis of Quality Control Samples 5
5.2 Instrument Performance and Maintenance 8
5.3 Interlaboratory ''Round-robin'' Test Results 8
6 Petroleum Refining Applications 8
6.1 Atmospheric Crude Distillates 8
6.2 Crude Vacuum Distillates 10
6.3 Resid Processing Characterization 10
7 Crude Oil Characterization 11
7.1 Yield Correlations between Crude Assay Distillation and High-temperature Simulated Distillation 11
7.2 Crude Oil Contaminants 13
8 Petroleum Hydrocarbons in Environmental Samples 13
8.1 Soil Contaminants 13
8.2 Oily Waters 13
Acknowledgments 13
Abbreviations and Acronyms 14
Related Articles 14
References 14

Simulated distillation (SimDist) is a gas chromatography (GC) technique which separates individual hydrocarbon components in the order of their boiling points, and is used to simulate the time-consuming laboratory-scale physical distillation procedure known as true boiling point (TBP) distillation. The separation is accomplished with a nonpolar chromatography column using a gas chromatograph equipped with an oven and injector that can be temperature programmed. A flame ionization detector (FID) is used for detection and measurement of the hydrocarbon analytes. The result of SimDist analysis provides a quantitative percent mass yield as a function of boiling point of the hydrocarbon components of the sample. The chromatographic elution times of the hydrocarbons are calibrated to the atmospheric equivalent boiling point (AEBP) of the paraffins reference material. The SimDist method ASTM (American Society for Testing and Materials) D2887 covers the boiling range 55–538°C (100–1000°F) which covers the n-*alkanes (*n-*paraffins) of chain length about* C_5–C_{44}*. The high-temperature simulated distillation (HTSD) method covers the boiling range 36–750°C (97–1382°F) which covers the* n-*alkane range of about* C_5–C_{120}*. A key difference between ASTM D2887 and HTSD is the ability of the latter technique to handle residue-containing samples (i.e. material boiling > 538°C, 1000°F). SimDist and laboratory-scale physical distillation procedures are routinely used for determining boiling ranges of petroleum crude oils and refined products, which include crude oil bottoms and residue processing characterization. The boiling point with yield profile data of these materials are used in operational decisions made by refinery engineers to improve product yields and product quality. Data from SimDists are valuable for computer modeling of refining processes for improvements in design and process optimization. Precise yield correlations between HTSD and crude assay distillation (methods ASTM D2892 and D5236) have allowed HTSD to be successfully used in place of physical distillation procedures. This has given the refiner the ability to rapidly evaluate crude oils for selection of those with economic advantages and more favorable refining margins. SimDist methods are becoming more widely used in environmental applications. HTSD is useful for characterizing hydrocarbons which can be present as soil and water contaminants; for example, to map and follow hydrocarbon removal processes.*

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

SimDist is a GC technique which separates individual hydrocarbon components in the order of their boiling points, and is used to simulate time-consuming laboratory-scale physical distillation procedures. The separation is accomplished with a chromatography column coated with a nonpolar (hydrocarbon-like), stationary phase, and uses a gas chromatograph equipped with an oven and injector which can be temperature programmed. A FID is used for detection and measurement of the hydrocarbon analytes. The SimDist analysis result provides a quantitative percent mass yield as a function of boiling point of the hydrocarbon components of the sample being analyzed. The chromatographic elution times of the hydrocarbon components are calibrated to the AEBP of the *n*-alkane as described in a method from the ASTM[1] by using *n*-alkane (*n*-paraffin) reference material. In the SimDist method ASTM D2887,[1] the *n*-alkane calibration reference covers the boiling range 55–538 °C (100–1000 °F) which covers the *n*-alkanes with a chain length of about C_5–C_{44}. In the HTSD method, the *n*-alkane calibration reference (a hydrogenated polyolefin wax, polywax 655) covers the boiling range 36–750 °C (97–1382 °F) which covers the *n*-alkanes with a chain length of about C_5–C_{120}. A key difference between ASTM D2887 and HTSD is the ability of HTSD to handle residue-containing samples (i.e. material boiling > 538 °C or 1000 °F).

SimDist and laboratory-scale physical distillation methods are routinely used for determining boiling ranges of petroleum crude oils and refined products. The boiling point with yield profile data of these materials are used in operational decisions made by refinery enginers to improve product yields and product quality. SimDist is valuable for, and can improve results from, computer modeling of refining processes for improvements in design and process optimization. Precise yield correlations between HTSD and crude assay distillation (a procedure which uses methods ASTM D2892[2] and D5236[3]) have allowed HTSD to be successfully used in place of physical distillation procedures. This can result in economic advantages for the refiner in the selection of crude oils, with more favorable refining margins. SimDist methods are also now becoming more widely used in environmental applications; for example, in characterizing hydrocarbons which can be present as soil and water contaminants.

2 HISTORY

SimDist was reported early in the 1960s by Eggertsen et al.[4] and Green et al.[5,6] as a method of simulating the time-consuming laboratory-scale physical distillation "TBP" procedure[7] by using GC. The SimDist technique is based on the fundamental assumption that individual nonpolar hydrocarbon components of a sample elute in the order of their boiling points from a GC column coated with nonpolar (hydrocarbon-like), stationary phase. The elution, or retention, time is dependent upon vapor pressure of the component and its affinity for the stationary phase. This varies with different types of hydrocarbons. For example, aromatic hydrocarbons and cycloalkanes (naphthenes) generally elute earlier than *n*-alkanes having the same boiling points. In SimDist, conditions are selected to give limited column efficiency and resolution unlike other chromatography techniques, which generally aim for highly efficient conditions that achieve high-resolution separation of components. The lower resolution analysis conditions provide distillation data that agree with physical distillation.

SimDist became an ASTM standard method in 1973, with the designation D2887, "Boiling Range Distribution of Petroleum Fractions by GC". The current edition is designated D2887-97. This method covers the determination of the boiling range distribution of petroleum products and fractions having a final boiling point (FBP) of 538 °C (1000 °F) or lower at atmospheric pressure.

HTSD is a relatively recent method which extends ASTM D2887 determination of the boiling range distribution of hydrocarbons to a FBP of about 750 °C (1382 °F). Technological advances in capillary GC columns and stationary phases together with either programmed temperature vaporization (PTV) or on-column injection techniques, provide adequate separation from C_5 to C_{120} normal paraffins and allows the characterization of petroleum products from about 36–750 °C (97–1382 °F). Under the special conditions of HTSD, elution of materials from the GC column occurs at up to 260–316 °C (500–600 °F) below their AEBP. For instance, the elution of C_{110} (AEBP of 735 °C or 1355 °F) occurs at a column temperature of about 427 °C (800 °F). Also under these conditions, little or no evidence of cracking is normally seen in HTSD.

3 COMPARISON OF AMERICAN SOCIETY FOR TESTING AND MATERIALS DISTILLATION AND SIMULATED DISTILLATION METHODS

A summary of ASTM physical distillation and SimDist methods is presented schematically in Figure 1 and discussed in detail below.

3.1 Physical Distillation Methods

Laboratory-scale physical distillation methods are routinely used by petroleum refineries for determining

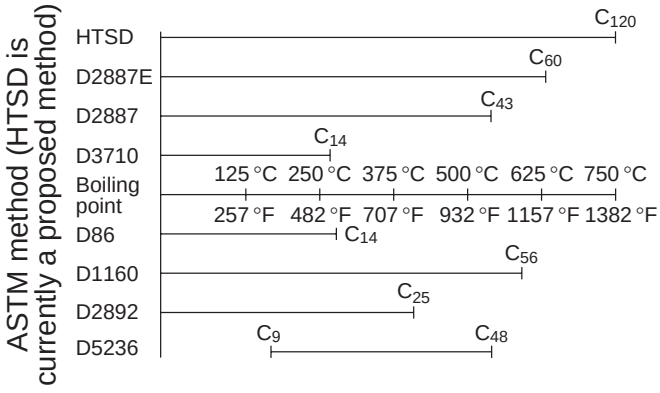

Figure 1 Summary of ASTM physical and simulated distillation.

the boiling ranges of crude oils and their products. ASTM D86[8] has been used as a quality control (QC) test since the 1920s. This method is a one-plate distillation that covers the range of materials boiling up to about 254 °C (490 °F). ASTM D1160[9] is a similar method but is conducted at reduced pressure and covers the determination of the range of boiling points for petroleum products that can be partially or completely vaporized at a maximum liquid temperature of 400 °C (752 °F). This temperature corresponds to an atmospheric equivalent (vapour) temperature of about 600 °C when run under 1 mmHg (the lowest pressure permitted). Petroleum products and fractions in this range may decompose if distilled at atmospheric pressure. ASTM D2892[2] is used for samples with a wide boiling range such as crude petroleum up to a final cut temperature of 400 °C (752 °F) atmospheric equivalent temperature (AET). The method can also be applied to any petroleum mixtures except liquified petroleum gases, very light naphthas, and fractions having initial boiling points (IBPs) above 400 °C. This test method uses a fractionating column having an efficiency of 14–18 theoretical plates operated at a reflux ratio of 5 : 1 for the production of liquefied gas, distillate fractions, and residuum. The method provides for the determination of yields of material as both mass and volume. A graph of temperature versus mass-percent distilled can be produced, and this distillation curve corresponds to a laboratory technique which is defined at 15/5 (15 theoretical plate, 5 : 1 reflux ratio) or TBP. ASTM D5236[3] is for distillation of heavy hydrocarbon mixtures such as heavy crude oils, petroleum distillates, residues, and synthetic mixtures with IBPs greater than 150 °C (300 °F). This method provides for the determination of standard distillation curves to the highest AET possible by conventional distillation. The maximum achievable temperature up to 565 °C (1050 °F) AET is dependent upon the heat tolerance of the charge.

3.2 Simulated Distillation Methods

SimDist methods which have been adopted as ASTM test methods are D2887,[1] D3710,[10] and D5307.[11] Method D2887-97 covers the determination by GC of the boiling range distribution of petroleum products and fractions having a FBP of 538 °C (1000 °F) or lower at atmospheric pressure. Method D3710-95 covers the determination by GC of the boiling range distribution of gasoline and gasoline components with a FBP 260 °C (500 °F) or lower. D5307-97 is for the determination by GC of boiling range distribution of crude petroleum through 538 °C (1000 °F), with material boiling above 538 °C being reported as residue.

HTSD by GC, which is currently under ASTM Committee D-2 proposed development status, is an extension of ASTM method D2887 for application to petroleum products and fractions, and materials containing higher boiling range materials above 538 °C (1000 °F). This method is used for determining the boiling point distribution of petroleum products and fractions in the range up to about 732 °C (1350 °F) at atmospheric pressure, and reports material boiling higher than this as material boiling >732 °C (1350 °F). The method is applicable to a wide range of fully eluting and non-completely eluting materials, which include petroleum crude oil, cat crack feed and product, hydrotreater feed and products, atmospheric residue, vacuum gas oils, deasphalted oils, and vacuum tower bottoms (pitch).

4 METHODOLOGY

By the proper choice of GC conditions and equipment in HTSD, separation in the boiling range from C_5 to C_{120} *n*-alkanes is carried out routinely. To accomplish the goal of eluting heavy materials up to the equivalent of C_{120}, a thin film of column nonpolar stationary phase is used. The film thickness varies from 0.05–0.15 m, which, with a 0.53 mm internal diameter (ID) capillary column results in a phase ratio (volume of the column vs the volume of the stationary phase) high enough to permit the elution of materials from the column at temperatures up to 260–316 °C (500–600 °F) below their AEBP.

GC instrumentation for HTSD is typically equipped with a PTV injector and FID. The analysis is carried out with a relatively short capillary column about 5 m length, 0.53 mm ID, and 0.05–0.15 μm film thickness. The temperature conditions of the column oven and injector are set initially at either near ambient or sub-ambient (−20 °C) and then programmed at a specified linear rate to a final temperature of about 430 °C. Sub-ambient initial conditions are generally required to cover a wide boiling range of samples. Helium carrier gas is used. The detector

signal is recorded as area slices (time intervals) for consecutive increasing retention times. SimDist software is used for data acquisition and results calculations.

4.1 Sample Preparation

A consequence of the conditions necessary for HTSD is a limited concentration capacity of the column due to the small amount of stationary phase. This requires appropriate dilution of the standards and samples (usually in carbon disulfide). Liquid hydrocarbon samples such as petroleum crude oils, gas oils, and solid or semi-solid residue materials, such as vacuum tower bottoms and asphaltic materials, are typically prepared at a concentration of about 2 mass percent in the solvent. Environmental samples such as oil-contaminated soils and waters are generally prepared at higher concentrations on the order of about 50 mass percent solvent to sample to improve sensitivity.

4.2 High-temperature Simulated Distillation Analysis

Because of the highly inert conditions of high-purity fused silica and special treated metal GC columns, the gentle injection techniques, and the short time at maximum temperature, little or no evidence of cracking is normally seen in HTSD. Because of the column breakdown (bleed) during the final portion of an HTSD analysis and the need to dilute the sample to approximately 1–2%, a blank GC run using only the solvent is recorded in the data system. This solvent blank is then subtracted from all subsequent GC runs in the analysis sequence of calibration, QC, and samples (Figure 2). This blank subtraction accomplishes

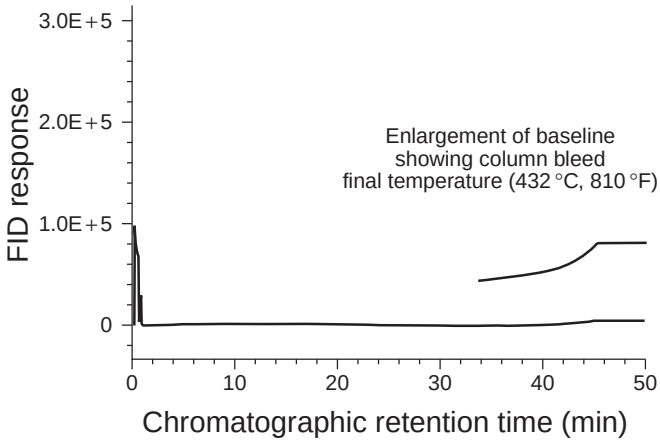

Figure 2 Analysis of carbon disulfide solvent blank by using HTSD.

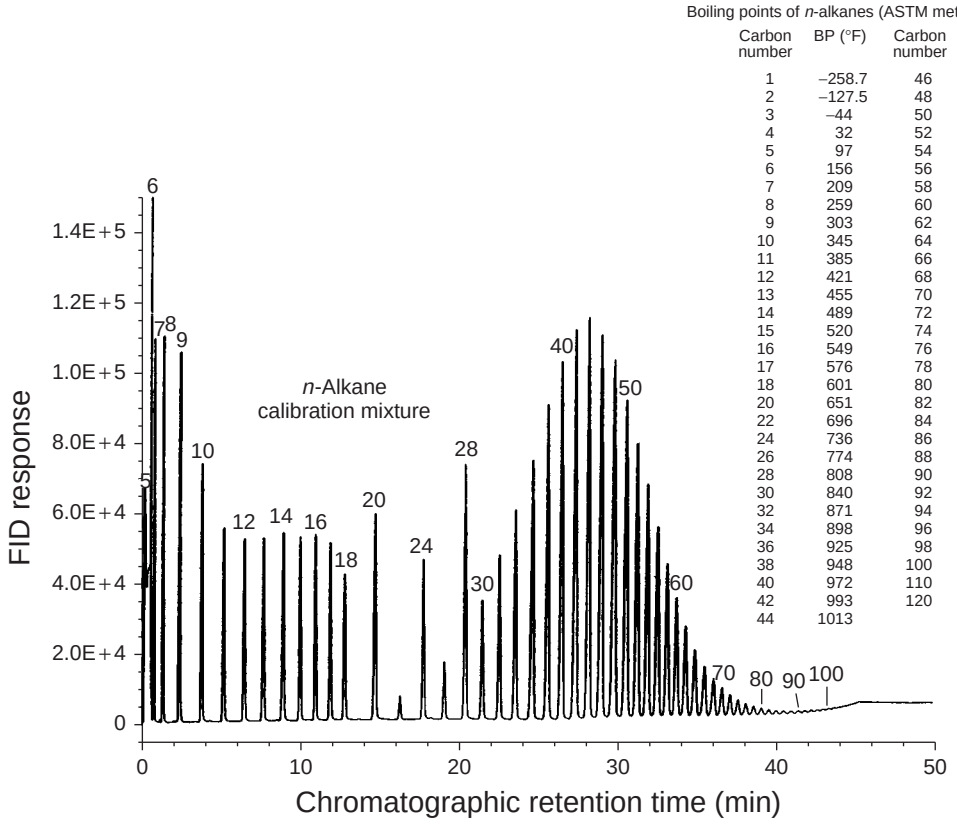

Boiling points of *n*-alkanes (ASTM method D2887-97[1])

Carbon number	BP (°F)	Carbon number	BP (°F)
1	–258.7	46	1033
2	–127.5	48	1051
3	–44	50	1067
4	32	52	1083
5	97	54	1096
6	156	56	1112
7	209	58	1126
8	259	60	1139
9	303	62	1152
10	345	64	1164
11	385	66	1175
12	421	68	1186
13	455	70	1197
14	489	72	1207
15	520	74	1216
16	549	76	1227
17	576	78	1238
18	601	80	1247
20	651	82	1258
22	696	84	1267
24	736	86	1276
26	774	88	1283
28	808	90	1292
30	840	92	1299
32	871	94	1306
34	898	96	1314
36	925	98	1321
38	948	100	1328
40	972	110	1355
42	993	120	1382
44	1013		

Figure 3 Analysis of C_5–C_{110} *n*-alkane calibration reference mixture by using HTSD.

two important goals: (1) it removes the signal present for the solvent, which occurs just following the start of the analysis near the determination of any light end material present; and (2) the column bleed is compensated. The assumption during this process is that the solvent blank and the column bleed profiles are constant during the calibration and sample analysis. It is the HTSD operator's duty to verify this criteria for statistically meaningful results.

4.3 Calibration

The HTSD calibration involves the analysis of (1) *n*-alkane calibration reference (hydrogenated polyolefin wax, polywax 655 which covers the boiling range $C_5 - C_{120}$) as shown in Figure 3, and (2) a reference oil that has been physically distilled by methods ASTM D2892 and D5236, as shown in Figure 4.

The polywax is used to calibrate chromatography elution times of the hydrocarbon components to the AEBP of the paraffin. The physically distilled reference oil is used as an external standard to calibrate the chromatography system for all sample percent recovery calculations. In addition, the HTSD analysis results found for the reference oil are compared to the mass distribution data known from physical distillation of the reference oil, and the statistical error of the difference between HTSD data versus actual boiling point data is reported (Table 1).

4.4 Determination of Hydrocarbon Yield as a Function of Boiling Point

The HTSD analysis result gives a chromatogram of detector signal response as a function of chromatography retention time, and a tabulated report of boiling point distribution of hydrocarbon yield at specific boiling cut-points in cumulative percent mass of sample (Figure 5, and Table 2). The report shows percent mass of material recovered in the analysis, IBP, FBP, and whether material boiling at higher than 732 °C (1350 °F) is present. The IBP is the temperature (corresponding to the retention time) at which a cumulative yield is equal to 0.5% of the total sample. The FBP is the temperature (corresponding to the retention time) at which a cumulative yield is equal to 99.5% of the total sample. The HTSD result can also be expressed in a carbon number distribution format which gives a graphical plot and tabulated report of yield at each *n*-alkane carbon number in incremental and cumulative percent mass (Figure 6).

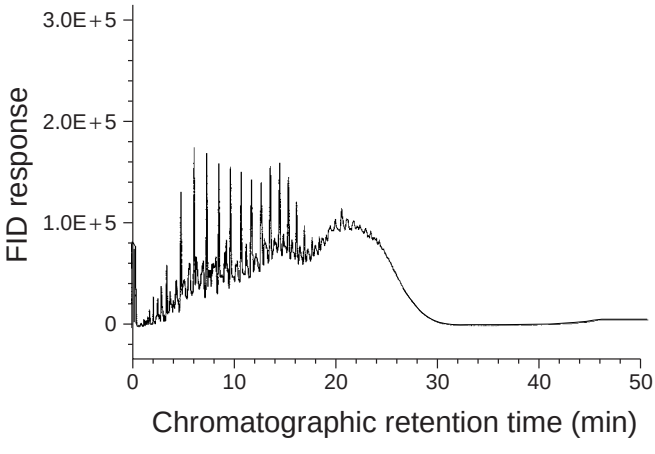

Figure 4 Analysis of ASTM D2892 and D5236 physically distilled reference oil by using HTSD.

Table 1 Results for reference oil obtained using HTSD and two ASTM methods (D2892 and D5236)

Percent area	BP (°F) (ASTM D2892 and D5236)	BP (°F) found (HTSD)	Squared difference (max. acceptance is 139)
10	448	446	4
20	529	526	9
30	603	600	9
40	658	656	4
50	716	715	1
60	777	776	1
70	826	824	4
80	873	871	4
90	923	921	4
		Total:	40

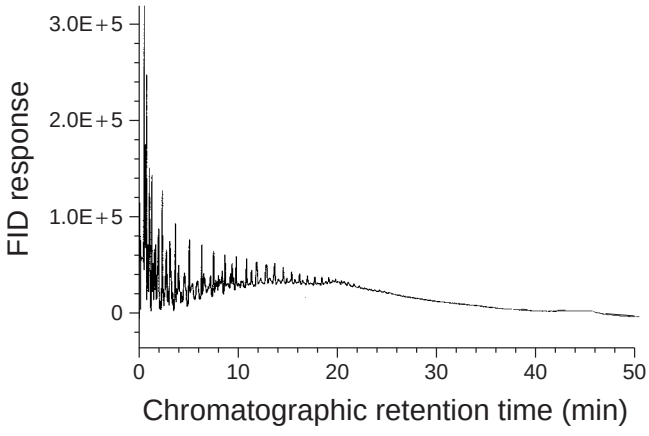

Figure 5 Analysis of Gulf of Mexico crude oil sample by using HTSD.

5 QUALITY CONTROL AND INSTRUMENT PERFORMANCE

5.1 Statistical Analysis of Quality Control Samples

QC and quality assurance (QA) procedures adopted for HTSD are rigorously followed to assure the integrity

Table 2 Tabulated report of results from HTSD analysis of a Gulf of Mexico crude oil

Simdis HT 750 analyzer

Data file name:	C:\HPCHEM\1\DATA\0920A\003F0701.D		
Instrument:	1	Vial:	3
Operator:	Triton Analyt.	Injection:	1
Acquired on:	20 Sep 94, 06:29 PM	Seq. line:	7
Processed on:	21 Sep 94, 11:40 AM		
Sample name:	Gulf of Mexico Crude	Sample amt:	0.0969
Method name:	MHC30.MTH	ISTD amt:	0.0000
Calculation based on:	ESTD	Weight CS2:	4.9479
Sequence name:	0920A Cryogenic sequence		
Blank used:	C:\HPCHEM\1\DATA\0920A\085F0201.D		
Calibration used:	C:\HPCHEM\1\DATA\0920A\090F0301.D		
Reference used:	C:\HPCHEM\1\DATA\0920A\095F0401.D		
Recovery (%):	96.7 found		
Response factor:	1.308E-10		
IBP:	104 (°F)		
FBP:	>1351 (°F)		
End elution after:	44.3 min		

Analysis results: %weight versus boiling point

%	BP (°F)	%	BP (°F)	%	BP (°F)	%	BP (°F)
1	140	25	468	49	687	73	921
2	163	26	478	50	696	74	932
3	191	27	489	51	705	75	943
4	209	28	497	52	715	76	955
5	227	29	507	53	724	77	968
6	240	30	515	54	734	78	981
7	260	31	524	55	743	79	994
8	272	32	535	56	753	80	1007
9	285	33	544	57	762	81	1022
10	298	34	553	58	771	82	1037
11	310	35	563	59	780	83	1052
12	320	36	572	60	789	84	1067
13	331	37	579	61	798	85	1085
14	346	38	588	62	807	86	1099
15	355	39	597	63	816	87	1116
16	370	40	605	64	826	88	1134
17	382	41	614	65	836	89	1152
18	392	42	624	66	846	90	1172
19	405	43	632	67	856	91	1192
20	416	44	641	68	867	92	1214
21	426	45	651	69	877	93	1241
22	436	46	660	70	887	94	1269
23	447	47	669	71	898	95	1298
24	457	48	677	72	909	96	1328

Analysis results: table of cutpoints

BP (°F)	%	BP (°F)	%	BP (°F)	%	BP (°F)	%
150	1.2	375	16.6	650	44.9	1100	86.1
220	4.7	500	28.3	1000	79.4	1200	91.3

of HTSD and statistically meaningful results. The QC includes the initial analysis for calibration of the reference oil that has been physically distilled by methods ASTM D2892 and D5236. As discussed in section 4.3 above, the HTSD yield with boiling point results for the reference oil are compared to the physical distillation mass distribution data, and the statistical error of the difference between HTSD data and the data by physical distillation is calculated and compared with acceptance limits established for the HTSD method.

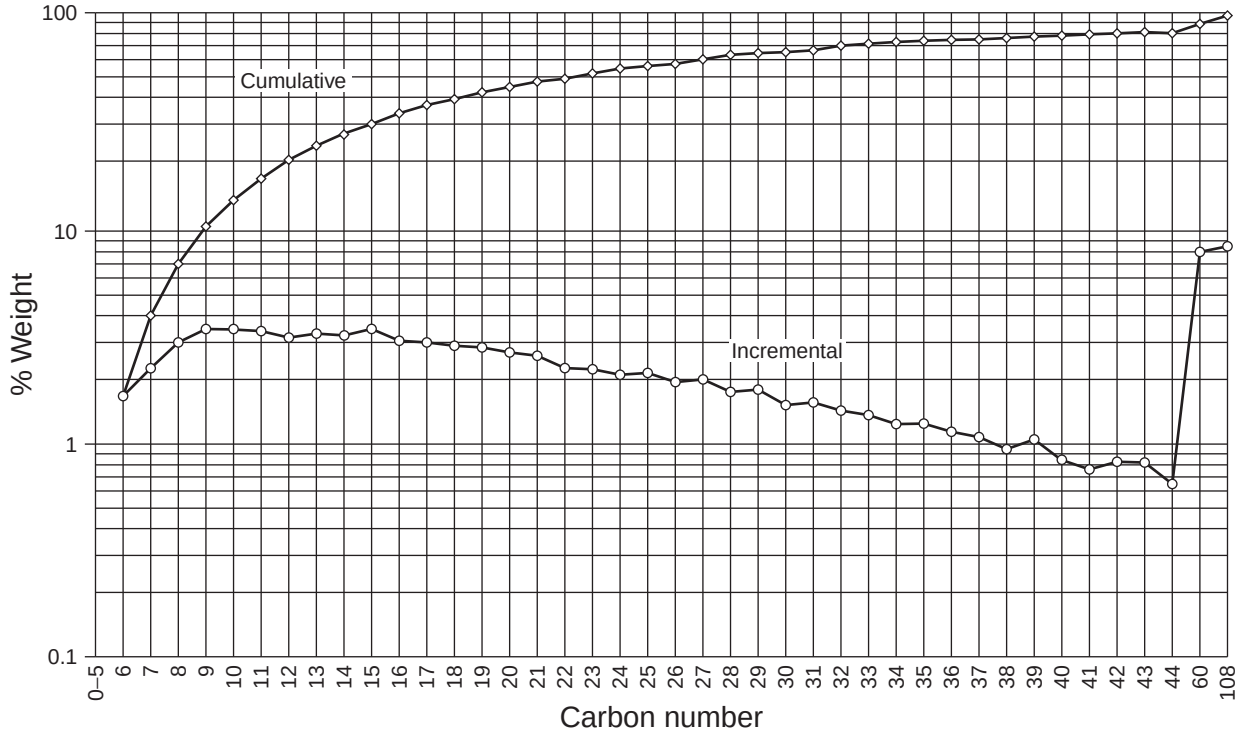

Figure 6 *n*-Alkane carbon number distribution from HTSD analysis of Gulf of Mexico crude oil sample.

Statistical analysis of QC samples is done with materials such as:

- a blend of physically distilled gas oils,
- lube feed stocks,
- hydrotreated residue,
- crude oils of different specific gravity (for example, California crudes and Gulf of Mexico crudes).

For each (QC) sample, the mean and standard deviation (SD) of the temperature vs percent yield is tracked over time. Using an X-type control chart, any excursions beyond ±2 or 3 SD limits will signal an out of control range method (Figures 7 and 8). Appropriate maintenance and corrective action is triggered before a sample analysis is reported.

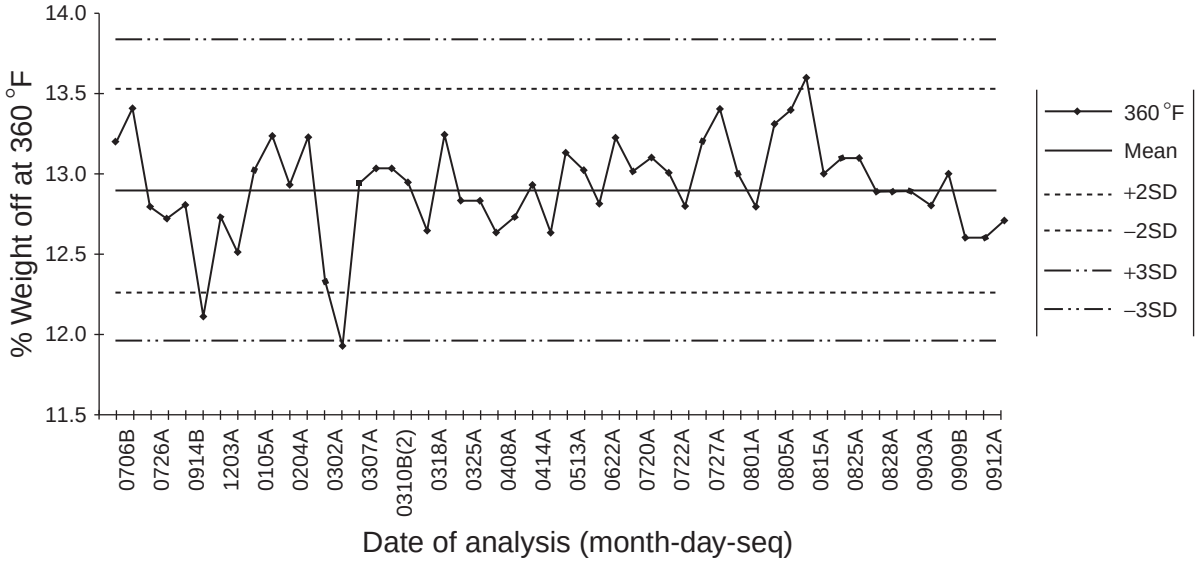

Figure 7 QC X-chart for full range (naphtha and residue-containing) sample (% off at 360 °F) by using HTSD.

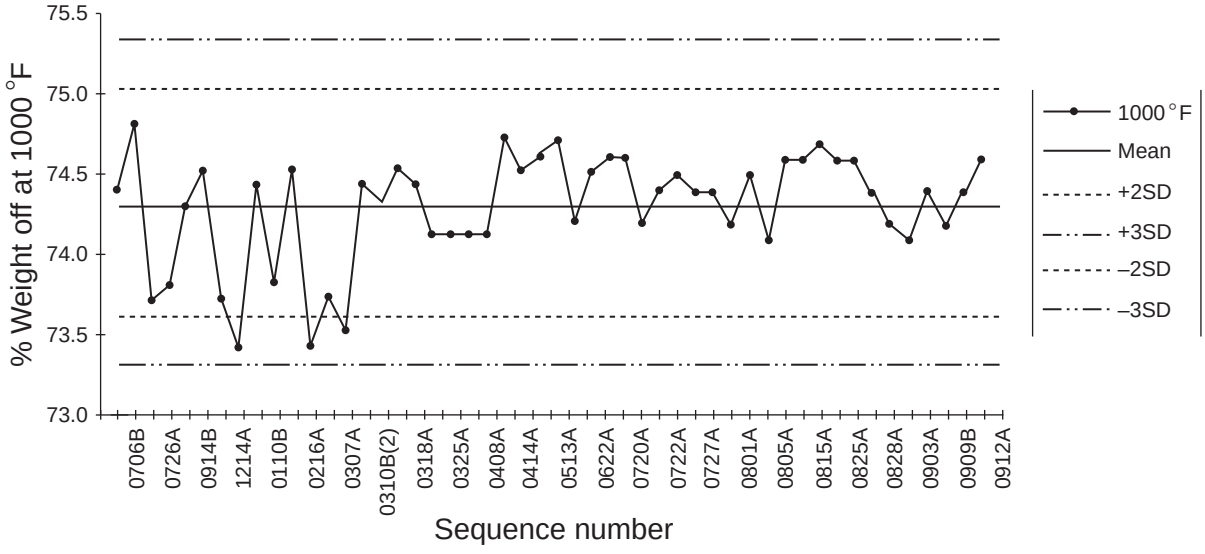

Figure 8 QC X-chart for full range (naphtha and residue-containing) sample (% off at 1000 °F or pitch yield) by using HTSD.

5.2 Instrument Performance and Maintenance

Careful operator evaluation of the system performance is a condition for accurate HTSD data. Visual inspection of the *n*-alkane calibration reference is useful in evaluating system performance. Chromatographic peak shape symmetry of the *n*-alkanes and changes in their retention times can give an indication of the condition of the chromatography column and also the GC injector liner. The solvent blank analysis is also important in monitoring any shifts in baseline during the analysis.

Performance is affected by variables such as loss of the stationary phase at high temperatures (i.e. resulting in loss of film thickness and sample capacity) and the unavoidable build-up from residue containing samples of nonvolatile materials, such as metals, and asphaltenes in the injector and the column. Corrective action often involves replacing the GC injector liner, cutting off a short length of the column at the inlet, or replacing the column with a new one.

5.3 Interlaboratory "Round-robin" Test Results

Since HTSD is currently under ASTM proposed method status, methodologies that are followed in conducting this analysis vary. Users of the technique generally follow a manufacturer's recommended practices for their particular instrumentation. Interlaboratory "round-robin" testing of HTSD draft methods and ruggedness studies for reproducibility are under investigation by ASTM Committee D-2 Study Group and internally by petroleum refining companies that use the technique. The data in Table 3 (provided by Royal Dutch/Shell Laboratories) show interlaboratory round-robin results for five different crude oils by six laboratories. The percent mass of hydrocarbon yield with temperature are given for each crude type and laboratory. Statistical analysis of the data shows the averages, SDs, relative SDs, and the overall average SDs for the HTSD analyses. As shown, precision is best at the cut point yields of 20%, 50%, and 70%, and is lower near the beginning and end cut points of 10% and 90%. The relative SDs of the total percent recoveries of material in the range $<720\,^{\circ}C$ (corresponding to the boiling point of *n*-alkane C_{100}) ranged from about 1% to 5% for the different crude oils tested.

6 PETROLEUM REFINING APPLICATIONS

Distillation is the primary separation process used in crude-oil refining. Although the refinery unit has some nonideality that is not seen in laboratory-scale distillation, laboratory distillation is valuable in simulating the actual plant process. Laboratory analysis data on refinery process streams are often relied on as part of evaluation of unit performance. Process improvement and optimization decisions can be made with the help of laboratory test results that relate to plant occurrences.

6.1 Atmospheric Crude Distillates

Most atmospheric crude units produce atmospheric gas oil (AGO) product for hydrocracker or fluidized catalytic cracker (FCC) feed. A well-designed and proper operating unit will produce AGO with only barely detectable levels of metals present. Generally no volatile metals are found in AGO. Some units, however, produce AGO with relatively high levels of undesirable metals, such as

Table 3 HTSD interlaboratory round-robin results for crude oils

Crude oil and laboratory	Yield at temperature (°C)					%<720 °C (<1328 °F)	% >720 °C (>1328 °F)
	10%	20%	50%	70%	90%		
Arabian Light							
–Lab A	118	179	348	472	653	94.4	5.6
–Lab B	127	185	351	474	660	93.2	6.8
–Lab C	125	182	352	479	680	94.1	5.9
–Lab D	156	174	332	443	592	94.7	5.3
–Lab E	112	167	330	443	589	95.3	4.7
–Lab F	117	175	331	443	589	93.9	6.1
Average	125.8	177.0	340.7	459.0	627.2	94.3	5.7
SD	15.8	6.4	10.7	17.7	41.7	0.7	0.7
% Relative SD	12.5	3.6	3.1	3.9	6.6	0.8	12.5
RAS Burden Medium							
–Lab A	149	236	442	575	ND	87.6	12.4
–Lab B	152	240	446	576	ND	85.7	14.3
–Lab C	149	236	444	582	ND	86.4	13.6
–Lab D	103	195	388	504	647	89	11
–Lab E	138	229	432	558	ND	89	11
–Lab F	153	236	435	559	744	89.2	10.8
Average	140.7	228.7	431.2	559.0	695.5	87.8	12.2
SD	19.2	16.9	21.8	28.6	68.6	1.5	1.5
% Relative SD	13.7	7.4	5.1	5.1	9.9	1.7	12.3
Sahara Blend Extra Light							
–Lab A	87	129	255	357	526	97.2	2.8
–Lab B	99	142	274	387	645	91.6	8.4
–Lab C	103	134	255	354	508	103	0
–Lab D	83	130	253	350	504	102	0
–Lab E	99	133	254	351	499	106	0
–Lab F	101	140	258	353	498	99.2	0.8
Average	95.3	134.7	258.2	358.7	530.0	99.8	2.0
SD	8.2	5.3	7.9	14.1	57.2	5.1	3.3
% Relative SD	8.6	3.9	3.1	3.9	10.8	5.1	165.9
Summatran Light Waxy							
–Lab A	157	235	402	497	686	92.6	7.4
–Lab B	172	248	415	526	ND	88.4	11.6
–Lab C	158	235	398	486	658	95.8	4.2
–Lab D	130	213	389	480	654	95.6	4.4
–Lab E	162	236	402	490	666	95.4	4.6
–Lab F	167	237	399	485	653	96.7	3.3
Average	157.7	234.0	400.8	494.0	663.4	94.1	5.9
SD	14.7	11.4	8.4	16.7	13.6	3.1	3.1
% Relative SD	9.3	4.9	2.1	3.4	2.1	3.3	52.5
Alba Heavy							
–Lab A	255	309	451	559	705	91.8	8.2
–Lab B	258	315	466	581	ND	85.9	14.1
–Lab C	252	305	444	544	687	95.5	4.5
–Lab D	250	305	446	550	697	93.8	6.2
–Lab E	254	308	447	536	687	95	5
–Lab F	248	302	437	532	664	97.4	2.6
Average	252.8	307.3	448.5	550.3	688.0	93.2	6.8
SD	3.6	4.5	9.7	17.9	15.4	4.0	4.0
% Relative SD	1.4	1.5	2.2	3.2	2.2	4.3	59.8
Overall avg SD	12.3	8.9	11.7	19.0	39.3	2.9	2.5

30 ppm mass nickel plus vanadium caused from entrained residue.[12] An example of estimating an entrainment level is to fractionate AGO by ASTM D5236 into two cuts containing the first 95 volume percent and final 5 volume percent. HTSD results, along with results from metals analysis methods, can then be used to help quantify the residue entrainment.

6.2 Crude Vacuum Distillates

Crude vacuum column units produce heavy distillate for downstream processing in an FCC or hydrotreater. Knowing the feedstock distillation curve slope and incremental gas oil contaminants is valuable for optimizing production of heavy vacuum gas oil (HVGO). Understanding contaminant distribution and HVGO product quality is important. For many crudes, high amounts of volatile metals and residue can lead to downstream catalyst poisoning. Laboratory analysis data is important for measuring HVGO yields and quality at different cut points and to determine feedstock contaminant distributions for predictive design work. Distillation curve data for vacuum column feedstocks are obtained in the laboratory by using various distillation techniques such as column distillation, short-path distillation, and SimDist by GC.[12] Column distillation includes spinning-band distillation, batch distillation (ASTM D5236), and continuous-flash vaporization. Physical distillation methods result in overlap between cuts and resultant inaccuracy in the contaminant distribution curve. HTSD analysis of each cut allows the contaminant curve to be corrected for overlap. An example of this application of HTSD is shown in Figure 9. In another example as shown in Figure 10, HTSD was used

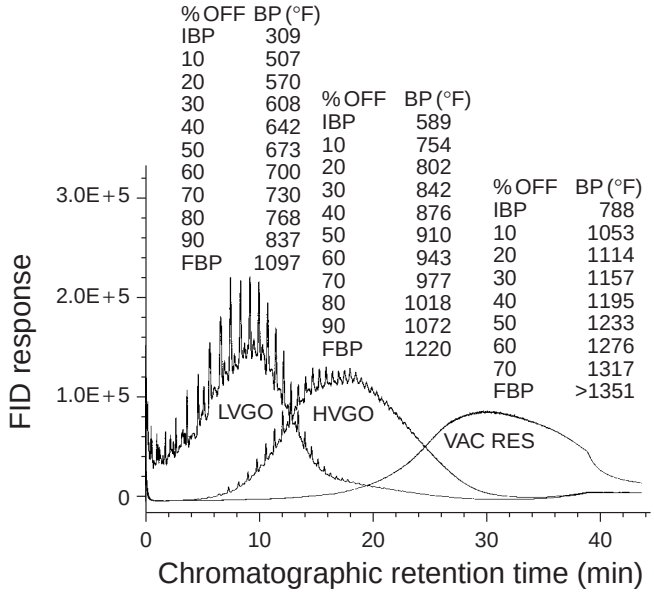

Figure 9 Analysis to characterize refinery vacuum tower products by using HTSD.

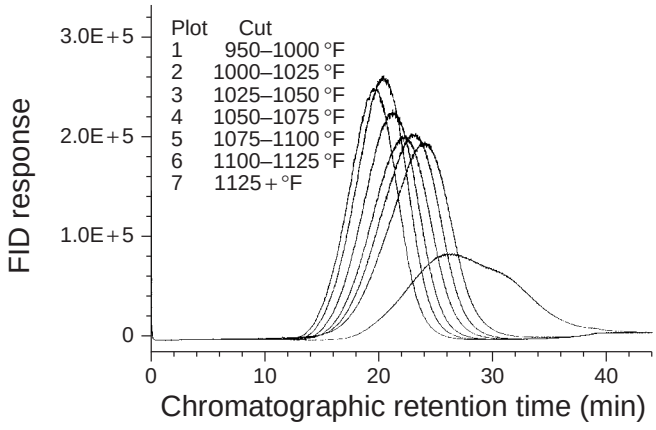

Figure 10 Analysis to characterize preparative physical distillation cuts by using HTSD.

to characterize seven laboratory preparative high vacuum distillation fractions at 25 °F cut points. The HTSD result can determine both the median boiling point of each fraction and also the preparative distillation efficiency or dispersion. With the boiling range of the preparative fraction precisely defined by HTSD and data concerning the properties, obtained by conducting other analyses on the preparative fractions such as density, asphaltenes, Conrad Carbon, and metals, useful vacuum tower optimization models for specific feeds and specific catalyst poisons can be developed.

6.3 Resid Processing Characterization

Crude oil bottoms present challenges to refiners who need to effectively convert them into useful products. Among the many ''bottom of the barrel'' processes available, solvent deasphalting is predominantly used to recover high-value deasphalted gas oil (DAO) from petroleum resid. DAO is excellent feedstock for lube oil, fluid catalytic cracking, and hydrocracking. The residuum oil supercritical extraction (ROSE™) process licensed by Kellogg Brown & Root is a state-of-the-art solvent extraction process widely used in the industry today.[13] HTSD is a useful tool for characterizing hydrocarbons by using boiling point profiles of fractions from the resid solvent extraction process.

The HTSD data presented in Figure 11 and in Table 4 show the chromatograms and percent mass yield as a function of boiling point for sample fractions from a ROSE™ process.[14] In this study, propane was used to extract the deasphalted oil fraction from the petroleum residue.

HTSD has proven valuable for characterizing yield at defined temperatures for heavy feeds and products in the development of technology for upgrading crude and heavy oils. For example, the chromatograms shown

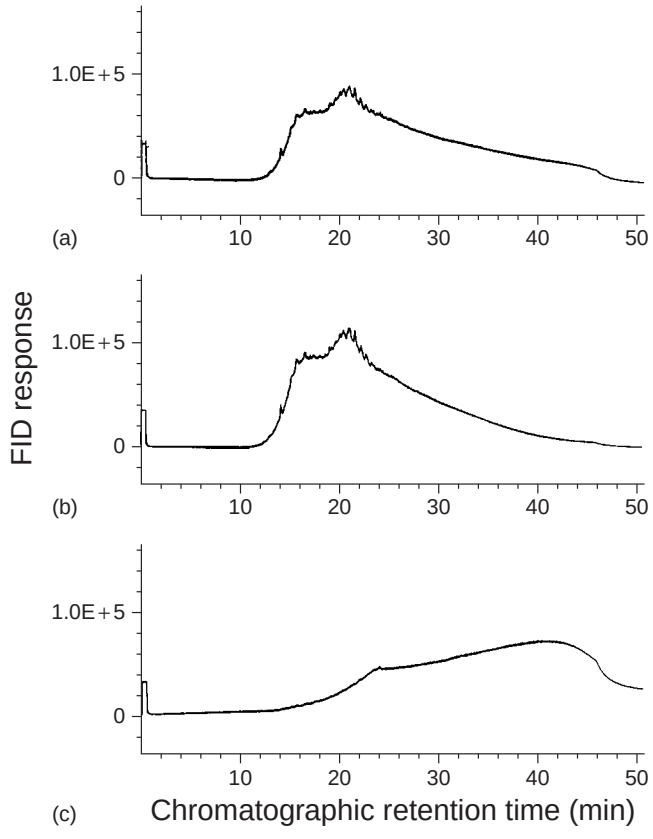

Figure 11 Analysis for resid (a), DAO (b), and asphaltene (c) from the ROSE™ process by using HTSD.

Table 4 HTSD yield (% weight) as a function of boiling point for fractions from the (ROSE™) process

Sample	IBP (°C)	371 °C (700 °F) % wt	482 °C (900 °F) % wt	538 °C (1000 °F) % wt	Recovery at 733 °C (1351 °F) % wt
Resid	317	10.4	46.8	61.6	96.4
Deasphalted oil	314	13.2	57.9	74.0	100
Asphaltene	361	0.7	11.6	22.6	82.6

in Figure 12 along with the data in Table 5 illustrate HTSD results for process feed and product samples from the FLEXCOR T process. This is a proprietary Exxon process to upgrade crude and heavy oils with a high naphthenic acid (i.e. total acid number (TAN)) content. The process reduces TAN (reportedly by about 90%) and as a result reduces the corrosivity of crudes and heavy oils caused by TAN in downstream refineries. The process provides moderate resid conversion under low severity operating conditions. While HTSD for this application is not for detection or measurement of TAN removal, HTSD is valuable for characterizing the reduction in the

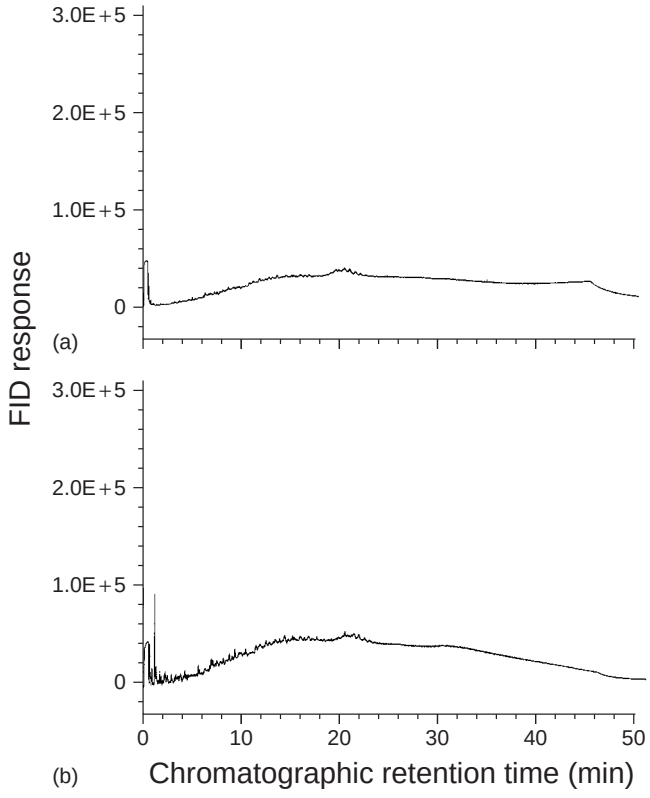

Figure 12 Analysis for feed (a) and product (b) from the Exxon FLEXCOR T process by using HTSD.

Table 5 HTSD yield (% weight) as a function of boiling point for process feed and product samples from the FLEXCOR T process

Sample	IBP (°C)	249 °C (480 °F) % wt	371 °C (700 °F) % wt	538 °C (1000 °F) % wt	Recovery at 733 °C (1351 °F) % wt
Feed	172	4.1	38.4	47.8	80.1
Product	236	7.2	47.0	56.8	85.1

amount of resid and the increase of liquid yield at defined temperatures.

7 CRUDE OIL CHARACTERIZATION

7.1 Yield Correlations between Crude Assay Distillation and High-temperature Simulated Distillation

The crude assay distillation procedure is defined by the methodologies of ASTM Methods D2892 and D5236. The procedure first uses a 15-plate column operating under a reflux ratio of 5 : 1 (D2892) which, as discussed in section 3.1, is known as TBP distillation. The distillation is usually started at ambient pressure (760 mmHg)

and then switched to vacuum conditions (for example, from 100 mmHg and then to 2 mmHg minimum) to extend the method to about 343 °C (650 °F) AEBP. At this point the remaining charge is transferred to a vacuum potstill method (D5236) where the distillation is continued at lower pressure (0.3 and 0.1 mmHg minimum allowed) thus allowing for an AEBP limit of about 538–565 °C (1000–1050 °F). Conversion tables for the vacuum conditions to AEBP are included in the methods.

Although the crude assay distillation (D2892 and D5236) procedure provides only an estimate of the yields of the products of various boiling ranges, the results (when properly and skillfully obtained) are of great importance for the characterization and commercial trading of crudes. However, several areas that can contribute to data variability of the crude assay distillation procedure include:

- the use of two different consecutive distillation methods with widely different characteristics such as efficiency, column hold-up and pressure drop across the column;
- inaccuracies in pressure and temperature readings and the subsequent conversion to AEBP;
- maximum attainable temperatures (AEBP) and cracking limits (maximum liquid temperature tolerance) are different for different types of crudes.

This grouping represented a good range of crudes with widely varying content of pitch, sulfur, nickel, vanadium, oxygen, Conradison Carbon or microcarbon residue, and asphaltene.

The comparisons of the yield curves (expressed in mass percent) of crude assay (ASTM D2892 and D5236) to HTSD were presented for five typical crudes spanning the range of light, intermediate, and heavy API gravities. Each crude was compared at eleven distillation cutpoints (°F): 68, 155, 265, 350, 400, 500, 600, 700, 800, 900, 1000. In another series of comparisons, the percent mass at each of the eleven cutpoints as determined by HTSD was subtracted from the percent mass from crude assay (D2892 and D5236).

The HTSD analysis results of crudes have been compared to crude assay distillation results in a study reported by Villalanti et al.[15] About 100 crudes ranging in API gravity from light to heavy were analyzed. In that study the crudes were grouped into three categories (Table 6) based on API gravity. API gravity is a function of relative density (specific gravity) that is widely used in the petroleum industry. The API value [(141.5/specific gravity 60/60 °F) −131.5 min] requires no statement temperature because it is included in the definition of the measurement; API gravity values decrease as density increases and range from 0 (equivalent to density 1.076) to 100 (equivalent to density 0.6112).

Table 6 Crude type, based on API gravity

Crude type	API gravity	Number of crudes	API range
Light	>30	49	30.1–52.3
Intermediate	20–30	27	20.5–30.0
Heavy	<20	8	9.7–19.5

Table 7 HTSD cutpoint data for a QA sample

HTSD cutpoint (°F)	% Mass	SD
360	12.7	0.3
480	19.6	0.3
650	28.9	0.3
1000	73.0	0.5
1351	98.4	0.9

The QA sample for the HTSD analysis was a full range (naphtha and residue containing) petroleum sample. The HTSD cutpoint statistics determined for the QA sample for more than 500 determinations were as given in Table 7.

The comparison of the distillation yield curves (percent mass) as determined by crude assay and by HTSD showed good agreement. In general, the difference was better than ±2% mass over the cutpoint range of IBP to 538 °C (1000 °F) (Figure 13). The only exception to this is the region around 399 °C (750 °F). The deviation in this region is caused by the change from high efficiency (15 plates) D2892 distillation to low efficiency (one plate) D5236 distillation. The distillation at this point goes from TBP distillation to non-TBP distillation. This results in a larger deviation for HTSD versus physical distillation at the crossover and beyond, and also introduces a bias between D2892 and D5236 (Figures 13 and 14). The fact that HTSD has no crossover effects is one of its main advantages for modelling and monitoring purposes.

The SD of HTSD at cutpoints up to 538 °C (1000 °F) is better than 0.5%. The estimated precision of the

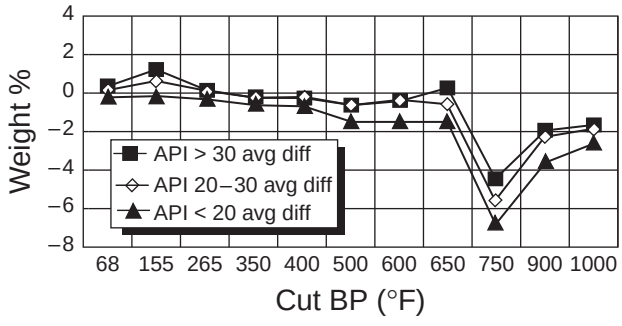

Figure 13 Comparison of physical distillation assay and HTSD as % weight difference in their values at each cut point for crude oils grouped by API gravities.

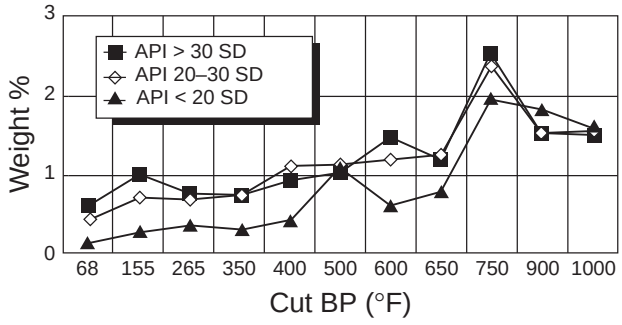

Figure 14 Comparison of physical distillation assay and HTSD as the SD of the difference of their values at each cut point for crude oils grouped by API gravities.

correlation between crude assay distillation and HTSD yield at each cutpoint results in an SD of <2% weight except at the crossover point at 399 °C (750 °F) (Figure 14).

Advantages cited in that study for characterization of crude by HTSD include:

- method precision;
- faster turnaround and less expense than laboratory physical distillation;
- valuable business tool for evaluating new crudes and confirming crude quality before purchase;
- ensuring crude product integrity during transportation and delivery.

7.2 Crude Oil Contaminants

Product integrity must be verified in the course of crude oil trading. Enormous quantities of crude oil are transported and processed daily. Product adulteration can occur whether intentionally or not. A useful application of HTSD is in monitoring product integrity. For example, the blending of pitch (defined as 538 °C+ or 1000 °F+) into crudes has been spotted at concentration levels of about 1.5 percent mass and lower.[16] Advantages of HTSD for this application include speed of analysis, small sample quantity requirement, and extended boiling range of the method.

8 PETROLEUM HYDROCARBONS IN ENVIRONMENTAL SAMPLES

SimDist methods are becoming more widely used in environmental applications. HTSD is useful for characterizing hydrocarbons which can be present as soil and water contaminants; for example, to map and follow hydrocarbon removal processes which may involve biodegradation, thermal, and/or physical separation methods.

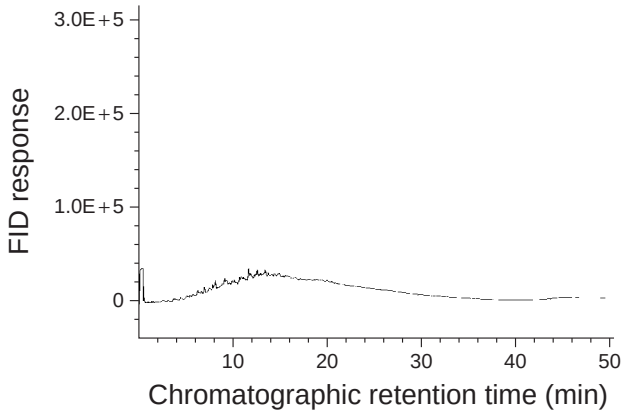

Figure 15 Analysis of environmental soil sample extract by using HTSD.

8.1 Soil Contaminants

An example of a soil extract analysis by HTSD is shown in Figure 15. For analysis the soil is weighed into a small vial and contacted with a weighed amount of solvent such as carbon disulfide, and the extract is analyzed. To improve sensitivity, samples are prepared for analysis at soil to solvent ratios of about 50 percent mass. Other extraction techniques and solvents can also be used. The HTSD result is useful for characterizing the amounts of material over boiling range regions representative of naphtha, gasoline, diesel, and heavier material. The detectable hydrocarbons can be reported on the basis of percent mass of soil as received or on a normalized percent hydrocarbons.

8.2 Oily Waters

In the HTSD analysis of a predominantly aqueous matrix, the sample is weighed into a small vial and shaken with a weighed amount of solvent, such as carbon disulfide, and the extract analyzed. In this case the extract is the bottom layer because carbon disulfide is denser than water. Water to solvent ratios of about 50% mass are used. Other extraction techniques and solvents can also be used. As with soil samples, the HTSD result is useful for characterizing the amounts of petroleum hydrocarbons in the boiling range of naphtha, gasoline, diesel, and heavier material. The detectable hydrocarbons can be reported on basis of percent mass of sample as received or on a normalized percent hydrocarbon basis.

ACKNOWLEDGMENTS

The authors wish to thank the following colleagues for their valuable contributions to this article: Roby Bearden, Jr, Exxon Research and Development Laboratories,

for his discussions concerning HTSD applications in the FLEXCOR T process; Rinus Daane, Royal Dutch/Shell KSLA, for providing data on the interlaboratory round-robin studies of the HTSD analyses of crude oils; Murugesan Subramanian, Kellogg Brown & Root, for information on HTSD applications in refinery ROSE™ solvent deasphalting operations.

ABBREVIATIONS AND ACRONYMS

AEBP	Atmospheric Equivalent Boiling Point
AET	Atmospheric Equivalent Temperature
AGO	Atmospheric Gas Oil
ASTM	American Society for Testing and Materials
DAO	Deasphalted Gas Oil
FBP	Final Boiling Point
FCC	Fluidized Catalytic Cracker
FID	Flame Ionization Detector
GC	Gas Chromatography
HTSD	High-temperature Simulated Distillation
HVGO	Heavy Vacuum Gas Oil
IBP	Initial Boiling Point
ID	Internal Diameter
PTV	Programmed Temperature Vaporization
QA	Quality Assurance
QC	Quality Control
ROSE™	Residuum Oil Supercritical Extraction
SD	Standard Deviation
SimDist	Simulated Distillation
TAN	Total Acid Number
TBP	True Boiling Point

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Soxhlet and Ultrasonic Extraction of Organics in Solids

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Fuels Analysis, Regulatory Specifications for • Full Range Crudes, Analytical Methodology of • Lube Products. Molecular Characterization of Base Oils • Lubricant Base Oils: Analysis and Characterization of • Petroleum Residues, Characterization of

REFERENCES

1. ASTM Method D2887-97, 'Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.02, 1997.
2. ASTM Method D2892-95, 'Test Method for Distillation of Crude Petroleum (15-Theoretical Plate Column)', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.02, 1996.
3. ASTM Method D5236-95, 'Test Method for Distillation of Heavy Hydrocarbon Mixtures (Vacuum Potstill Mixtures)', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.03, 1998.
4. F.T. Eggertsen, S. Groennings, J.J. Holst, 'Analytical Distillation by Gas Chromatography', *Anal. Chem.*, **32**, 904–909 (1960).
5. L.E. Green, J.C. Worman, 'Simulated Distillation of High Boiling Petroleum Fractions', *Anal. Chem.*, **37**, 1620–1621 (1965).
6. L.E. Green, L.J. Schumauch, J.C. Worman, 'Simulated Distillation by Gas Chromatography', *Anal. Chem.*, **36**, 1512–1516 (1964).
7. K.H. Altgelt, T.H. Gouw (eds.), *Chromatography in Petroleum Analysis, Chromatographic Science Series*, Marcel Dekker, Inc., New York, Vol. 11, 1979.
8. ASTM Method D86-96, 'Test Method for Distillation of Petroleum Products', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.01, 1998.
9. ASTM Method D1160-95, 'Test Method for Distillation of Petroleum Products at Reduced Pressure', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.01, 1998.
10. ASTM Method D3710-95, 'Test Method for Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.02, 1998.
11. ASTM Method D5307-97, 'Test Method for Determination of Boiling Range Distribution of Crude Petroleum by Gas Chromatography', in *Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Vol. 05.03, 1998.
12. S.W. Golden, S. Craft, D.C. Villalanti, 'Simulated Distillation by Gas Chromatography', *Hydrocarbon Processing*, **74**(11), 1–8 (1995).
13. M. Subramanian, M.D. Deo, F.V. Hanson, 'Compositional Analysis of Bitumen and Bitumen-derived Products', *J. Chromatogr. Sci.*, **34**, 20–26 (1996).
14. J.C. Raia, D.C. Villalanti, M. Subramanian, B. Williams, 'Application of High-temperature Simulated Distillation

(HTSD) to the Residuum Oil Supercritical Extraction Process in Petroleum Refining', *J. Chromatogr. Sci.*, **38**(1), 1–5 (2000).

15. D.C. Villalanti, J.B. Maynard, J.C. Raia, A.A. Arias, 'Yield Correlations between Crude Assay Distillation and High-temperature Simulated Distillation (HTSD)', Distillation Symposium, Distillation Design and Operation: Troubleshooting, Operation, and Optimization, Paper 51b, AIChE 1997 Spring National Meeting, March 9–13, Institute of Chemical Engineers, New York, 1997.
16. D.C. Villalanti, D. Janson, P. Colle, 'Hydrocarbon Characterization by: High-temperature Simulated Distillation (HTSD)', Distillation Session, Distillation Column Design and Operation-IV: Advances in Distillation Modeling and Simulation, AIChE 1995 Spring National Meeting, March 19–23, Institute of Chemical Engineers, New York, 1995.

Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Christoph Klawun
Applied Automation Inc., Bartlesville, USA

1 Introduction 1
1.1 Hydrocarbons 1
1.2 Gas Chromatography 1
1.3 Detection 2
1.4 Advantages 2
1.5 Limitations 2
1.6 Process Gas Chromatography 3
2 History 3
3 General Methods 4
3.1 Multidimensional Gas Chromatography 4
3.2 Gaseous Hydrocarbons 4
3.3 Liquid Hydrocarbons 4
3.4 Impurity Analysis 4
3.5 Comprehensive Analysis 11
3.6 Heteroatom Analysis 11
3.7 Specific Compounds 12
4 On-line Applications 12
4.1 Natural Gas Analysis 12
4.2 Gasoline Blending 13
4.3 Polymer Production Feedstock 13
5 Off-line Applications 13
5.1 Hyphenated Techniques 13
5.2 Bulk Properties 13
Acknowledgments 14
Abbreviations and Acronyms 14
Related Articles 15
References 15

The analysis of hydrocarbon mixtures originating from mineral oil sources using gas chromatography (GC) is compiled in this article and a general overview of GC and detectors, including hyphenated instrumentation, is given. The advantages and limitations of GC, as well as multidimensional GC are described in more detail from a standpoint of process GC, and the historical development of chromatography is put into perspective. Comprehensive tables of target analytes and methods offer access to all standard methods published by the American Society for Testing and Materials (ASTM) for hydrocarbon analysis with GC. Based on these tables, several subtopics are covered: gaseous and liquid hydrocarbons, heteroatom analysis, comprehensive and impurity analysis, natural gases, gasoline and fuel production, polymer feedstocks, hyphenated techniques, and bulk properties.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The analysis of hydrocarbons with GC is described in this article. The scope of applications is fairly narrow and covers hydrocarbons originating from mineral oil. This includes products supplied by the refining, petroleum, petrochemical, and oil industry, but not biological samples, environmental applications, food, extraterrestrial sources, etc.

1.1 Hydrocarbons

Hydrocarbons are understood as compounds consisting only of hydrogen and carbon with the general molecular formula C_xH_y. In finished petroleum products such as gasoline, it is also of interest to determine fixed gases and hydrocarbons containing sulfur or oxygen. Therefore, a few methods for the analysis of hydrocarbons with these heteroatoms are included as well.

1.2 Gas Chromatography

GC is a technique that separates gaseous mixtures into their components. In many cases, this technique can also be called gas–liquid chromatography: during the separation process, an inert gas (carrier gas, mobile phase) carries the gaseous mixture over a thin layer of an immobilized liquid (stationary phase). As mixture molecules are swept past the liquid film, they briefly permeate the surface and ''dissolve'' in the liquid before being released again into the gas stream. In essence, an equilibrium exists at any given point between the gas and liquid phase for all components; they are partitioned between the gas and liquid phases. This journey of mixture components takes place inside a tube that is either coated on the inside with a stationary phase (open tubular column), or filled with a grainy powder (support phase) that is coated with the stationary phase (packed column). A stronger affinity of a compound for the stationary phase results in a higher ''concentration'' in the liquid phase, effectively slowing the migration of this compound relative to the inert mobile phase. With an appropriately chosen stationary phase, most or all mixture constituents

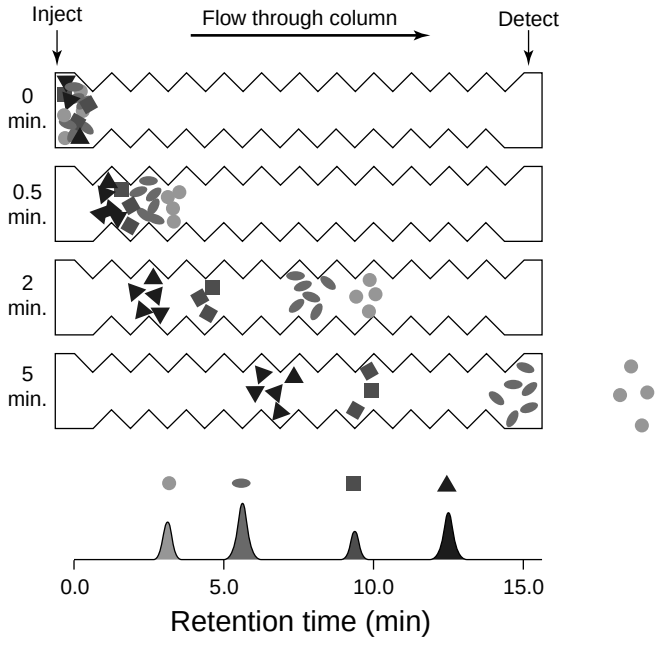

Figure 1 Schematic column separation diagram. The jagged column walls symbolize ''resistance'' against mixture molecules.

exhibit different partition coefficients. Therefore, the time that these components spend inside the column (retention time) will be different, so that they successively emerge (elute) from the column as separately detectable packets (see Figure 1 for a schematic illustration of column chromatography).

GC does not only deal with mixtures that exist as gases at ambient temperatures, but also for liquid mixtures that can be evaporated at operating temperatures of a gas chromatograph. The evaporation is often carried out during the introduction of a sample onto the column (injection). Elevated temperatures inside the column compartment (oven) are used to turn most or all of the injected sample into a gas. Owing to thermally labile samples or limits of the stationary phase, most GC separations do not exceed 200–300 °C. Newer column and oven technology has extended this limit to almost 500 °C. At the other end of the temperature spectrum, subambient separations can start as low as −60 °C.

1.3 Detection

The great majority of hydrocarbon separations with GC are quantitative in nature, although preparative and qualitative GC exists as well. For the determination of separated components, suitable detectors are employed, mostly universal detectors such as thermoconductivity detectors (TCD) or flame ionization detectors (FID). A TCD is a nondestructive detector suitable for quantitating components in a concentration range of approximately 0.01–100%, while FIDs operate best from the mid parts per billion (ppb, 10^{-9}) to high parts per million (ppm, 10^{-6}) range. FIDs exhibit the highest sensitivity for easily combustible compounds such as hydrocarbons. More specialized detectors for sulfur and oxygen as well as multichannel detectors will be introduced in sections 3.6 and 5.1. The continuous monitoring of the column output (effluent) with a detector yields a chromatogram which indicates the separated mixture constituent packets (eluents) as peaks (see bottom of Figure 1).

Sometimes, GC is also used to determine bulk properties such as boiling point distribution or vapor pressure instead of separating individual mixture constituents. In these cases, a chromatogram is evaluated as a whole and the mixture does not have to be separated into individually quantifiable peaks (see section 5.2).

1.4 Advantages

GC instrumentation is robust and a large body of literature and knowledge exists.[1–6] It is well suited for volatile samples whose constituents can be evaporated at temperatures up to 300 °C (sometimes even 450 °C). Compared with other instruments, the capital and maintenance costs are low. Using appropriate columns and methods, GC can be highly selective in detecting and determining a wide range of analytes. When hyphenated with multichannel detectors (see section 5.1), GC is used not only for quantitative, but also for qualitative analyses.

1.5 Limitations

Not many sample matrices are suitable for direct (neat) introduction into a GC. Sample cleanup is often required and aqueous samples or salts are frequently detrimental to a GC's hardware. GC also may not be suitable for the analysis of thermally labile samples. In addition, analytes must be in the gas phase during a split/splitless sample injection and the temperature profile of a separation must be designed to put all analytes into the gas phase. With complex samples, complete separation of all target analytes may require substantial hardware and method development investments, if it can be achieved at all. Potential interferences are difficult to detect if they overlap (co-elute) with the analytes of interest, a problem that can be overcome with multichannel detectors (see section 5.1). Carrier gas must be supplied continuously, which may sometimes be difficult with the most commonly used carrier gases helium (expensive in many parts of the world) and hydrogen (explosion hazards). Duty cycles (times between repeated measurements) for even moderately complex separations (5 min to more than 2 h) may be too long for quick responses based on GC analysis results.

Table 1 Process analysis approaches. (Adapted from McLennan[10])

Type	Advantages	Disadvantages
Off-line	Expert analysts available Flexible operation Controlled environment Sophisticated instrumentation Low unit costs per test	Slow Lack of ownership of data Conflicts of priorities Additional administration costs
At-line	Dedicated instrument Faster sampling process Simpler instrumentation Ownership of data by production Control of priorities	Low equipment utilization Robust equipment necessary
On-line	Fast Automatic feedback possible Dedicated instrumentation	Downtime must be minimized Long or expensive method development 24-h maintenance resource availability Electrical classification required

1.6 Process Gas Chromatography

GC-based hydrocarbon analyses[7–9] are usually carried out in the refining, petroleum, and petrochemical industry, where the majority of GCs are used to provide information about the production process. This can be done in an ''off-line'', ''at-line'', or ''on-line'' fashion. Table 1 outlines the differences between these approaches.[10] The focus on process applications in hydrocarbon analysis with GCs is also reflected in this article, but many methods developed for on-line analysis can also be applied in the laboratory, and vice versa.

Process gas chromatography (PGC) applications[11–13] can usually be classified into either control, mass balance, or monitoring categories.[14] Control applications usually target a few key components from a single sample stream, require short duty cycles, and the analysis results are used to adjust production parameters. Mass balance applications measure most of the individual components, group the remaining ones, and provide a track record for product quality, content, and process efficiency, e.g. in the determination of the heat value of pipeline gas, which is usually expressed in British thermal units (BTUs; 1 BTU = 1.055 kJ). Monitoring applications are normally designed to meet regulatory requirements, and may obtain samples from a large number of sampling points.

2 HISTORY

The earliest chromatography experiments of Runge, Schönbein, and Goppelsröder[15] go back to the nineteenth century. However, the invention of column chromatography[16,17] is attributed to Mikhail S. Tswett, who first demonstrated it in 1903,[18–20] when he successfully separated green leaf pigments to study them in their pure state. In a glass column filled with crushed chalk, he separated the leaf alcohol extract into individually identifiable bands of different color. Hence the name ''chromatography'', which comes from the Greek words $\chi\rho\omega\mu\alpha$ (chroma = color) and $\gamma\rho\alpha\phi o\sigma$ (graphos = writing).

Partition chromatography[21] as we know it today, and GC[22,23] in particular, were invented by Martin and Synge in the 1940s and 1950s, ultimately garnering for them the Nobel Prize in chemistry in 1952.[24] Although all separations were initially done on packed columns, open tubular columns are used in the majority of today's applications.

At the same time GC was developed, mass spectroscopy (MS) began to catch on as a valuable tool for analytical chemists in the early 1950s. The natural fit between these two instruments was discovered soon after that,[25] and the ''hyphenation'' of MS as a versatile multichannel GC detector turned into today's widespread usage of gas chromatography/mass spectroscopy (GC/MS). A GC/MS instrument is often used in qualitative GC work to identify unknown mixture components, and also to quantify known constituents. (For other versatile multichannel GC detectors, see section 5.1.)

Just as in other branches of analytical instrumentation, the increasing use of computers for GC has drastically improved productivity, while also increasing the precision and accuracy of separations through hardware control and chromatogram processing. The quantification of analytes – originally done by cutting out peaks from strip-chart recorder paper and weighing them – has progressed beyond integrators with on-the-fly calculations to computers that process entire chromatograms. Computers also make it possible to calculate bulk properties, such as boiling point distribution, vapor pressure, or BTU content by analyzing complete chromatograms.

3 GENERAL METHODS

The analysis of hydrocarbons with GC covers such a vast body of methods that only a few selected applications can be described. This section deals with ''general purpose'' methods, while more specific problems and their solutions are outlined in sections 4 and 5. There are several organizations that publish standard methods for hydrocarbon analysis, namely the ASTM. The titles and designation numbers of standards published by ASTM pertaining to GC hydrocarbon applications are listed in Table 2. Other organizations for oil, gas, refining, and petroleum, or organizations that publish standards for hydrocarbon analysis with GC are listed in Table 6, and most of the standard titles are accessible and searchable via the Internet.

All ASTM standard methods in Table 2 were examined for target compounds or identified peaks in their reference chromatograms; a complete list of analytes is compiled in Table 3 and common analyte synonyms are show in Table 4. Together, these three tables provide a comprehensive ''roadmap'' to solve common hydrocarbon analysis problems. Much of the text in the following sections serves as entry points into these tables.

3.1 Multidimensional Gas Chromatography

Common chromatography obstacles such as a wide range of target analytes or long separation times are usually overcome with multidimensional gas chromatography (MDGC).[26] Possible variations for MDGC are almost limitless, but they all share common characteristics: during the separation, part of the sample is diverted from the straight *injection* → *column* → *detector* scheme to traps, vents, other columns, and other detectors. Typical applications of MDGC are multiple columns of different selectivity to cover a wider range of components, rearrangement of elution order to speed up separation, and venting of unwanted compounds to protect analytical columns or to complete a separation in less time. Comprehensive two-dimensional separation with gas chromatography (GC × GC) is another tool in the MDGC toolbox to separate mixtures with hundreds of components from a single injection.[27,28] With this technique, the effluent from a longer column is re-injected as sample pulses onto a short column, where separation is accomplished about two orders of magnitude faster than in the longer column, resulting in a three-dimensional chromatogram. As the separation in the primary and certainly in the secondary column becomes faster, narrow injection plugs, rapid sample transfer between columns, and miniaturized hardware become more and more critical.[29,30]

An example of a heartcut and backflush-to-vent MDGC system is shown in Figure 2 and Figure 3.[14] Figure 2 depicts the plumbing arrangement between valves, columns, vents, and detectors, while the flow through the columns is illustrated in Figure 3. This arrangement is used to separate trace amounts of methane and ethane from ethylene, the main constituent. Normally, these components elute in the order CH_4, C_2H_4, and C_2H_6. The separation begins with valves in position 1, stabilizing the flows prior to sample introduction. After the injection with the sample valve (SV), the valves are changed to position 2 (beginning of first heartcut), allowing the methane peak to pass from column 1 through the TCD into column 2. Once the end of the methane peak has been observed with the TCD, the valves are switched back to position 1 (end of first heartcut), which vents the ethylene peak after it emerges from column 1 and the TCD. Once the ethylene elution from column 1 is complete, the valves are switched to position 2 (beginning of second heartcut) to allow ethane (with traces of unseparated ethylene) to flow into column 2 for detection by the FID. Once ethane has completely exited column 1 (end of second heartcut), the valves are switched to position 3, reversing the flow through column 1 and thereby venting all heavier components (backflush to vent), while ethane completes its course through column 2.

3.2 Gaseous Hydrocarbons

Hydrocarbons and related gases that are in their gaseous state at ambient temperature (C_1 through C_4 or C_5) are measured in ASTM standard methods D2504 and D3612 (see also D1945, D1946, D2163, D2426, D2427, D2505, D2593, D2712, and D4424). Measurements in these methods often include fixed gases such as oxygen, nitrogen, and carbon dioxide as well, and are usually applied to natural gas processing or polymer precursor production such as ethylene or propylene.

3.3 Liquid Hydrocarbons

Except for waxes and tars of high molecular weight, most other common hydrocarbons up to about C_{20} exist as liquids at ambient temperature. The melting and boiling points of *n*-alkanes[31] are plotted in Figure 4 as reference points. Most ASTM standard methods for liquid hydrocarbon analysis with GC deal with specific applications or target compounds, but there are also several that are of general nature and do not fit into any specific category: D2306, D2597, D3524, D3525, and D5917.

3.4 Impurity Analysis

The characterization of trace components in the production of high purity compounds or in hydrocarbon waste is important to meet customer specifications, avoid catalyst poisoning in polymer production, or to comply

Table 2 ASTM methods for hydrocarbon analysis with GC and related standards

Number[a]	Update[b]	Title[c]
(D86)	1997	STM for Distillation of Petroleum Products at Atmospheric Pressure
(D323)	1994	STM for Vapor Pressure of Petroleum Products (Reid Method)
D1945	1996	STM for Analysis of Natural Gas by GC
D1946	1990	SP for Analysis of Reformed Gas by GC
D2163	1991	STM for Analysis of Liquefied Petroleum (LP) Gases and Propene Concentrates by GC
D2268	1993	STM for Analysis of High-Purity *n*-Heptane and Isooctane by Capillary GC
D2306	1996	STM for C_8 Aromatic Hydrocarbon Analysis by GC
D2360	1995	STM for Trace Impurities in Monocyclic Aromatic Hydrocarbons by GC
D2426	1993	STM for Butadiene Dimer and Styrene in Butadiene Concentrates by GC
D2427	1992	STM for Determination of C_2 through C_5 Hydrocarbons in Gasolines by GC
D2504	1988	STM for Noncondensable Gases in C_2 and Lighter Hydrocarbon Products by GC
D2505	1988	STM for Ethylene, Other Hydrocarbons, and Carbon Dioxide in High-Purity Ethylene by GC
D2593	1993	STM for Butadiene Purity and Hydrocarbon Impurities by GC
D2597	1994	STM for Analysis of Demethanized Hydrocarbon Liquid Mixtures Containing Nitrogen and Carbon Dioxide by GC
D2712	1991	STM for Hydrocarbon Traces in Propylene Concentrates by GC
D2887	1997	STM for Boiling Range Distribution of Petroleum Fractions by GC
D3054	1995	STMs for Analysis of Cyclohexane by GC
D3257	1997	STMs for Aromatics in Mineral Spirits by GC
D3524	1990	STM for Diesel Fuel Diluent in Used Diesel Engine Oils by GC
D3525	1993	STM for Gasoline Diluent in Used Gasoline Engine Oils by GC
(D3588)	1991	SP for Calculating Heat Value, Compressibility Factor, and Relative Density (Specific Gravity) of Gaseous Fuels
D3606	1996	STM for Determination of Benzene and Toluene in Finished Motor and Aviation Gasoline by GC
D3612	1996	STM for Analysis of Gases Dissolved in Electrical Insulating Oil by GC
D3710	1995	STM for Boiling Range Distribution of Gasoline and Gasoline Fractions by GC
D3760	1998	STM for Analysis of Isopropylbenzene (Cumene) by GC
D3797	1996	STM for Analysis of *o*-Xylene by GC
D3798	1996	STM for Analysis of *p*-Xylene by GC
(D4057)	1997	SP for Manual Sampling of Petroleum and Petroleum Products
(D4177)	1995	SP for Automatic Sampling of Petroleum and Petroleum Products
D4291	1993	STM for Trace Ethylene Glycol in Used Engine Oil
(D4307)	1994	SP for Preparation of Liquid Blends for Use as Analytical Standards
D4367	1994	STM for Benzene in Hydrocarbon Solvents by GC
D4420	1994	STM for Determination of Aromatics in Finished Gasoline by GC
D4424	1990	STM for Butylene Analysis by GC
D4492	1996	STM for Analysis of Benzene by GC
(D4626)	1995	SP for Calculation of Gas Chromatographic Response Factors
D4735	1996	STM for Determination of Trace Thiophene in Refined Benzene by GC
D4806	1998	Standard Specification for Denatured Fuel Ethanol for Blending with Gasolines for Use as Automotive Spark-Ignition Engine Fuel
D4815	1994	STM for Determination of MTBE, ETBE, TAME, DIPE, *t*-Amyl Alcohol and C_1 to C_4 Alcohols in Gasoline by GC
D4864	1990	STM for Determination of Traces of Methanol in Propylene Concentrates by GC
D5008	1997	STM for Ethyl Methyl Pentanol Content and Purity Value of 2-Ethylhexanol by GC
D5060	1995	STM for Determining Impurities in High-Purity Ethylbenzene by GC
D5134	1998	STM for Detailed Analysis of Petroleum Naphthas through *n*-Nonane by Capillary GC
D5135	1995	STM for Analysis of Styrene by Capillary GC
D5303	1992	STM for Trace Carbonyl Sulfide in Propylene by GC
D5307	1997	STM for Determination of Boiling Range Distribution of Crude Petroleum by GC
D5399	1995	STM for Boiling Point Distribution of Hydrocarbon Solvents by GC
D5441	1998	STM for Analysis of MTBE by GC
D5442	1993	STM for Analysis of Petroleum Waxes by GC
D5443	1993	STM for Paraffin, Naphthene, and Aromatic Hydrocarbon Type Analysis in Petroleum Distillates Through 200 °C by Multidimensional GC
D5480	1995	STM for Engine Oil Volatility by GC
D5501	1994	STM for Determination of Ethanol Content of Denatured Fuel Ethanol by GC
D5504	1998	STM for Determination of Sulfur Compounds in Natural Gas and Gaseous Fuels by GC and Chemiluminescence

(*continued overleaf*)

Table 2 (*continued*)

Number[a]	Update[b]	Title[c]
D5541	1994	STM for Analysis of MTBE by GC
D5580	1995	STM for Determination of Benzene, Toluene, Ethylbenzene, *p*/*m*-Xylene, *o*-Xylene, C_9 and Heavier Aromatics, and Total Aromatics in Finished Gasoline by GC
D5599	1995	STM for Determination of Oxygenates in Gasoline by GC and Oxygen Selective Flame Ionization Detection
D5623	1994	STM for Sulfur Compounds in Light Petroleum Liquids by GC and Sulfur Selective Detection
D5713	1996	STM for Analysis of High-purity Benzene for Cyclohexane Feedstock by Capillary GC
D5769	1998	STM for Determination of Benzene, Toluene, and Total Aromatics in Finished Gasolines by GC/MS
D5917	1999	STM for Trace Impurities in Monocyclic Aromatic Hydrocarbons by GC and External Calibration
D5986	1996	STM for Determination of Oxygenates, Benzene, Toluene, C_8-C_{12} Aromatics and Total Aromatics in Finished Gasoline by GC/Fourier Transform IRS
D6228	1998	STM for Determination of Sulfur Compounds in Natural Gas and Gaseous Fuels by GC and Flame Photometric Detection
D6229	1998	STM for Trace Benzene in Hydrocarbon Solvents by Capillary GC
D6296	1998	STM for Total Olefins in Spark-Ignition Engine Fuels by Multidimensional GC
(E260)	1996	SP for Packed Column GC
(E355)	1996	SP for GC Terms and Relationships
(E594)	1995	SP for Testing FIDs Used in GC
(E1510)	1995	SP for Installing Fused Silica Open Tubular Capillary Columns in Gas Chromatographs

[a] Parentheses are for related standards not directly applicable to hydrocarbon analysis with GC.
[b] Year of last revision (sometimes, ASTM methods also indicate years of reapproval or editorial changes).
[c] Abbreviations used for brevity in the standards titles: STM = standard test method; SP = standard practice; ETBE, ethyl *t*-butyl ether; DIPE, diisopropyl ether; MTBE, methyl *t*-butyl ether; TAME, *t*-amyl methyl ether.

Table 3 Target analytes in ASTM methods (see Table 2) for hydrocarbon analysis with GC and related standards. Common synonyms for a number of compounds are listed in Table 4. (Copyright ASTM. Reprinted with permission)

Compound	Formula	CAS no.	Method
Acetone	C_3H_6O	67-64-1	D5441
Acetylene	C_2H_2	74-86-2	D2505, D2712, D3612
t-Amyl methyl ether	$C_6H_{14}O$	994-05-8	D4815, D5441, D5599, D5986
Benzene	C_6H_6	71-43-2	D2268, D3606, D3760, D3797, D4367, D4420, D4492, D4735, D4815, D5060, D5134, D5443, D5580, D5713, D5769, D5986, D6229
Benzothiophene	C_8H_6S	95-15-8	D5504
1,2-Butadiene	C_4H_6	590-19-2	D2593
1,3-Butadiene	C_4H_6	106-99-0	D2426, D2593, D2712, D4424
n-Butane	C_4H_{10}	106-97-8	D1945, D2163, D2427, D2593, D2597, D4424, D5134, D5441
1-Butanethiol	$C_4H_{10}S$	109-79-5	D5504, D5623
1-Butanol	$C_4H_{10}O$	71-36-3	D4815, D5599, D5986
2-Butanol	$C_4H_{10}O$	78-92-2	D4815, D5599, D5986
2-Butanone	C_4H_8O	78-93-3	D5441
1-Butene	C_4H_8	106-98-9	D2427, D2593, D2712, D4424
cis-2-Butene	C_4H_8	590-18-1	D2427, D2593, D2712, D4424, D5441
trans-2-Butene	C_4H_8	624-64-6	D2427, D2593, D2712, D4424, D5441
1-Buten-3-yne	C_4H_4	689-97-4	D2593
Butylbenzene	$C_{10}H_{14}$	104-51-8	D3760, D5769, D5986
s-Butyl methyl ether	$C_5H_{12}O$	6795-87-5	D5441
1-Butyne	C_4H_6	107-00-6	D2593
2-Butyne	C_4H_6	503-17-3	D2593
Carbon dioxide	CO_2	124-38-9	D1945, D1946, D2505, D2597, D3612
Carbon disulfide	CS_2	75-15-0	D5504, D5623
Carbon monoxide	CO	630-08-0	D1946, D2504, D3612
Carbonyl sulfide	COS	463-58-1	D5303, D5504, D5623
Cyclohexane	C_6H_{12}	110-82-7	D2268, D5134, D5443
Cyclopentane	C_5H_{10}	287-92-3	D2268, D2597, D5134, D5443
Cyclopentene	C_5H_8	142-29-0	D5441
Cyclopropane	C_3H_6	75-19-4	D2712

Table 3 (*continued*)

Compound	Formula	CAS no.	Method
trans-Decalin	$C_{10}H_{18}$	493-02-7	D5443
n-Decane	$C_{10}H_{22}$	124-18-5	D5443
1,2-Diethylbenzene	$C_{10}H_{14}$	135-01-3	D5769, D5986
1,4-Diethylbenzene	$C_{10}H_{14}$	105-05-5	D5769, D5986
Diethyl disulfide	$C_4H_{10}S_2$	110-81-6	D5504, D5623
Diethyl sulfide	$C_4H_{10}S$	352-93-2	D5504, D5623
Diisopropyl ether	$C_6H_{14}O$	108-20-3	D4815, D5599, D5986
1,2-Dimethoxyethane	$C_4H_{10}O_2$	110-71-4	D4815, D5599, D5986
2,2-Dimethylbutane	C_6H_{14}	75-83-2	D2268, D2597, D5134
2,3-Dimethylbutane	C_6H_{14}	79-29-8	D2268, D2597, D5134, D5443
2,3-Dimethyl-1-butene	C_6H_{12}	563-78-0	D5441
1,1-Dimethylcyclohexane	C_8H_{16}	590-66-9	D5134
cis-1,2-Dimethylcyclohexane	C_8H_{16}	2207-01-4	D5134, D5443
trans-1,2-Dimethylcyclohexane	C_8H_{16}	6876-23-9	D5134
cis-1,3-Dimethylcyclohexane	C_8H_{16}	638-04-0	D5134
trans-1,3-Dimethylcyclohexane	C_8H_{16}	2207-03-6	D5134
cis-1,4-Dimethylcyclohexane	C_8H_{16}	624-29-3	D5134
trans-1,4-Dimethylcyclohexane	C_8H_{16}	2207-04-7	D5134
1,1-Dimethylcyclopentane	C_7H_{14}	1638-26-2	D2268, D5134
cis-1,2-Dimethylcyclopentane	C_7H_{14}	1192-18-3	D2268, D5134
trans-1,2-Dimethylcyclopentane	C_7H_{14}	822-50-4	D2268, D5134
cis-1,3-Dimethylcyclopentane	C_7H_{14}	2532-58-3	D2268, D5134
trans-1,3-Dimethylcyclopentane	C_7H_{14}	1759-58-6	D2268, D5134
Dimethyl disulfide	$C_2H_6S_2$	624-92-0	D5504, D5623
2,2-Dimethylheptane	C_9H_{20}	1071-26-7	D5134
2,3-Dimethylheptane	C_9H_{20}	3074-71-3	D5134
2,4-Dimethylheptane	C_9H_{20}	2213-23-2	D5134
2,5-Dimethylheptane	C_9H_{20}	2216-30-0	D5134
2,6-Dimethylheptane	C_9H_{20}	1072-05-5	D5134
3,3-Dimethylheptane	C_9H_{20}	4032-86-4	D5134
3,4-Dimethylheptane	C_9H_{20}	922-28-1	D5134
3,5-Dimethylheptane	C_9H_{20}	926-82-9	D5134
4,4-Dimethylheptane	C_9H_{20}	1068-19-5	D5134
2,2-Dimethylhexane	C_8H_{18}	590-73-8	D2268, D5134
2,3-Dimethylhexane	C_8H_{18}	584-94-1	D5134
2,4-Dimethylhexane	C_8H_{18}	589-43-5	D2268, D5134
2,5-Dimethylhexane	C_8H_{18}	592-13-2	D2268, D5134
3,3-Dimethylhexane	C_8H_{18}	563-16-6	D2268, D5134
3,4-Dimethylhexane	C_8H_{18}	583-48-2	D5134
3,5-Dimethyl-1-hexene	C_8H_{16}	7423-69-0	D5441
4,4-Dimethyl-2-neopentyl-1-pentene	$C_{12}H_{24}$	141-70-8	D5441
2,2-Dimethylpentane	C_7H_{16}	590-35-2	D2268, D5134
2,3-Dimethylpentane	C_7H_{16}	565-59-3	D2268, D5134
2,4-Dimethylpentane	C_7H_{16}	108-08-7	D2268, D5134
3,3-Dimethylpentane	C_7H_{16}	562-49-2	D2268, D5134
2,2-Dimethylpropane	C_5H_{12}	463-82-1	D1945, D2597, D2712, D5134
2,2-Dimethyl-1-propanol	$C_5H_{12}O$	75-84-3	D4815, D5599
Dimethyl sulfide	C_2H_6S	75-18-3	D5504, D5623
Diphenyl sulfide	$C_{12}H_{10}S$	139-66-2	D5623
n-Dodecane	$C_{12}H_{26}$	112-40-3	D5443
Ethane	C_2H_6	74-84-0	D1945, D1946, D2163, D2505, D2597, D3612, D5134
Ethanethiol	C_2H_6S	75-08-1	D5504, D5623
Ethanol	C_2H_6O	64-17-5	D4806, D4815, D5501, D5599, D5986
Ethylbenzene	C_8H_{10}	100-41-4	D3257, D3760, D3797, D5060, D5134, D5443, D5580, D5769, D5986
Ethyl *t*-butyl ether	$C_6H_{14}O$	637-92-3	D4815, D5441, D5599, D5986
Ethylcyclohexane	C_8H_{16}	1678-91-7	D5134
Ethylcyclopentane	C_7H_{14}	1640-89-7	D2268, D5134

(*continued overleaf*)

Table 3 (*continued*)

Compound	Formula	CAS no.	Method
Ethylene	C_2H_4	74-85-1	D1946, D2427, D2505, D2712, D3612
Ethylene glycol	$C_2H_6O_2$	107-21-1	D4291
3-Ethylheptane	C_9H_{20}	15869-80-4	D5134
4-Ethylheptane	C_9H_{20}	2216-32-2	D5134
3-Ethylhexane	C_8H_{18}	619-99-8	D5134
2-Ethyl-1-hexanol	$C_8H_{18}O$	104-76-7	D5008
1-Ethyl-1-methylcyclopentane	C_8H_{16}	16747-50-5	D5134
cis-1-Ethyl-2-methylcyclopentane	C_8H_{16}	930-89-2	D5134
trans-1-Ethyl-2-methylcyclopentane	C_8H_{16}	930-90-5	D5134
cis-1-Ethyl-3-methylcyclopentane	C_8H_{16}	2613-66-3	D5134
trans-1-Ethyl-3-methylcyclopentane	C_8H_{16}	2613-65-2	D5134
Ethyl methyl sulfide	C_3H_8S	624-89-5	D5504, D5623
3-Ethylpentane	C_7H_{16}	617-78-7	D2268, D5134
Helium	He	7440-59-7	D1945
n-Heptane	C_7H_{16}	142-82-5	D2268, D5134, D5443
n-Hexane	C_6H_{14}	110-54-3	D2268, D2597, D5134, D5443
2-Hexanone	$C_6H_{12}O$	591-78-6	D5580
1-Hexene	C_6H_{12}	592-41-6	D5443
Hydrogen	H_2	1333-74-0	D1945, D1946, D2504, D3612
Hydrogen sulfide	H_2S	7783-06-4	D1945, D5504, D5623
Indan	C_9H_{10}	496-11-7	D5769, D5986
Isopropylbenzene	C_9H_{12}	98-82-8	D3760, D3797, D5060, D5769, D5986
Isopropylcyclopentane	C_8H_{16}	3875-51-2	D5134
Methane	CH_4	74-82-8	D1945, D1946, D2505, D2597, D3612, D5134
Methanethiol	CH_4S	74-93-1	D5504, D5623
Methanol	CH_4O	67-56-1	D4815, D4864, D5441, D5501, D5599, D5986
2-Methylbutane	C_5H_{12}	78-78-4	D1945, D2163, D2268, D2427, D2597, D4424, D5134, D5441
2-Methyl-1-butene	C_5H_{10}	563-46-2	D2427, D5441
2-Methyl-2-butene	C_5H_{10}	513-35-9	D2427, D5441
3-Methyl-1-butene	C_5H_{10}	563-45-1	D2427, D5441
Methyl *t*-butyl ether	$C_5H_{12}O$	1634-04-4	D4815, D5441, D5541, D5599, D5986
Methylcyclohexane	C_7H_{14}	108-87-2	D2268, D5134, D5443
Methylcyclopentane	C_6H_{12}	96-37-7	D2268, D5134
1-Methyl-2-ethylbenzene	C_9H_{12}	611-14-3	D5769, D5986
1-Methyl-3-ethylbenzene	C_9H_{12}	620-14-4	D5769, D5986
1-Methyl-4-ethylbenzene	C_9H_{12}	622-96-8	D5769, D5986
2-Methyl-4-ethylhexane	C_9H_{20}	3074-75-7	D5134
2-Methyl-3-ethylpentane	C_8H_{18}	609-26-7	D5134
3-Methyl-3-ethylpentane	C_8H_{18}	1067-08-9	D5134
2-Methylheptane	C_8H_{18}	592-27-8	D5134
3-Methylheptane	C_8H_{18}	589-81-1	D5134
4-Methylheptane	C_8H_{18}	589-53-7	D5134
2-Methylhexane	C_7H_{16}	591-76-4	D2268, D5134
3-Methylhexane	C_7H_{16}	589-34-4	D2268, D5134
4-Methyl-1-hexene	C_7H_{14}	3769-23-1	D5443
1-Methylnaphthalene	$C_{11}H_{10}$	90-12-0	D5769, D5986
2-Methylnaphthalene	$C_{11}H_{10}$	91-57-6	D5769, D5986
2-Methyloctane	C_9H_{20}	3221-61-2	D5134
3-Methyloctane	C_9H_{20}	2216-33-3	D5134
4-Methyloctane	C_9H_{20}	2216-34-4	D5134
2-Methylpentane	C_6H_{14}	107-83-5	D2268, D2597, D5134, D5441
3-Methylpentane	C_6H_{14}	96-14-0	D2268, D2597, D5134, D5441
4-Methyl-*cis*-2-pentene	C_6H_{12}	691-38-3	D5441
2-Methylpropane	C_4H_{10}	75-28-5	D1945, D2163, D2427, D2593, D2597, D4424, D5134
1-Methyl-1-propanethiol	$C_4H_{10}S$	513-53-1	D5504, D5623
2-Methyl-1-propanethiol	$C_4H_{10}S$	513-44-0	D5504, D5623
2-Methyl-2-propanethiol	$C_4H_{10}S$	75-66-1	D5504, D5623
2-Methyl-1-propanol	$C_4H_{10}O$	78-83-1	D4815, D5599, D5986

Table 3 (*continued*)

Compound	Formula	CAS no.	Method
2-Methyl-2-propanol	$C_4H_{10}O$	75-65-0	D4815, D5441, D5599, D5986
2-Methylpropene	C_4H_8	115-11-7	D2427, D2593, D2712, D4424, D5441
α-Methylstyrene	C_9H_{10}	98-83-9	D3760
2-Methylthiophene	C_5H_6S	554-14-3	D5623
3-Methylthiophene	C_5H_6S	616-44-4	D5623
Naphthalene	$C_{10}H_8$	91-20-3	D5769, D5986
Nitrogen	N_2	7727-37-9	D1945, D1946, D2504, D2597, D3612
n-Nonane	C_9H_{20}	111-84-2	D5134, D5443
n-Octane	C_8H_{18}	111-65-9	D5134, D5443
Oxygen	O_2	7782-44-7	D1945, D1946, D2504, D3612
1,4-Pentadiene	C_5H_8	591-93-5	D2593
Pentamethylbenzene	$C_{11}H_{16}$	700-12-9	D5443, D5769
2,2,4,6,6-Pentamethyl-3-heptene	$C_{12}H_{24}$	123-48-8	D5441
n-Pentane	C_5H_{12}	109-66-0	D1945, D2268, D2427, D2597, D4424, D5134, D5441, D5443
1-Pentene	C_5H_{10}	109-67-1	D2427, D5441
cis-2-Pentene	C_5H_{10}	627-20-3	D2427, D5441
trans-2-Pentene	C_5H_{10}	646-04-8	D2427, D5441
2-Phenylbutane	$C_{10}H_{14}$	135-98-8	D3760
Propadiene	C_3H_4	463-49-0	D2593, D2712
n-Propane	C_3H_8	74-98-6	D1945, D2163, D2427, D2505, D2593, D2597, D3612, D4424, D5134
1-Propanethiol	C_3H_8S	107-03-9	D5504, D5623
2-Propanethiol	C_3H_8S	75-33-2	D5504, D5623
1-Propanol	C_3H_8O	71-23-8	D4815, D5599, D5986
2-Propanol	C_3H_8O	67-63-0	D4815, D5441, D5599, D5986
Propene	C_3H_6	115-07-1	D2163, D2427, D2505, D2593, D3612, D4424
Propylbenzene	C_9H_{12}	103-65-1	D3760, D5443, D5769, D5986
n-Propylcyclopentane	C_8H_{16}	2040-96-2	D5134
Propyne	C_3H_4	74-99-7	D2593, D2712
Styrene	C_8H_8	100-42-5	D2426, D3797, D5135
Sulfur dioxide	SO_2	7446-09-5	D5504, D5623
n-Tetradecane	$C_{14}H_{30}$	629-59-4	D5443
1,2,3,5-Tetramethylbenzene	$C_{10}H_{14}$	527-53-7	D5769, D5986
1,2,4,5-Tetramethylbenzene	$C_{10}H_{14}$	95-93-2	D5443, D5769, D5986
Thiophene	C_4H_4S	110-02-1	D4735, D5504, D5623
Toluene	C_7H_8	108-88-3	D2268, D3606, D3797, D4420, D5060, D5134, D5443, D5580, D5769, D5986
1,2,3-Trimethylbenzene	C_9H_{12}	526-73-8	D5443, D5769, D5986
1,2,4-Trimethylbenzene	C_9H_{12}	95-63-6	D5443, D5580, D5769, D5986
1,3,5-Trimethylbenzene	C_9H_{12}	108-67-8	D5769, D5986
2,2,3-Trimethylbutane	C_7H_{16}	464-06-2	D2268, D5134
1,1,2-Trimethylcyclohexane	C_9H_{18}	7094-26-0	D5134
1,1,3-Trimethylcyclohexane	C_9H_{18}	3073-66-3	D5134
1,*cis*-2,*cis*-4-Trimethylcyclohexane	C_9H_{18}	1678-80-4	D5443
1,1,2-Trimethylcyclopentane	C_8H_{16}	4259-00-1	D5134
1,1,3-Trimethylcyclopentane	C_8H_{16}	4516-69-2	D2268, D5134
1,*cis*-2,*cis*-3-Trimethylcyclopentane	C_8H_{16}	2613-69-6	D5134
1,*cis*-2,*trans*-3-Trimethylcyclopentane	C_8H_{16}	15890-40-1	D5134
1,*trans*-2,*cis*-3-Trimethylcyclopentane	C_8H_{16}	19374-46-0	D2268, D5134
1,*cis*-2,*cis*-4-Trimethylcyclopentane	C_8H_{16}	2613-72-1	D5134
1,*cis*-2,*trans*-4-Trimethylcyclopentane	C_8H_{16}	4850-28-6	D2268, D5134, D5134
2,4,6-Trimethylheptane	$C_{10}H_{22}$	2613-61-8	D5134
2,2,3-Trimethylhexane	C_9H_{20}	16747-25-4	D5134
2,2,4-Trimethylhexane	C_9H_{20}	16747-26-5	D5134
2,2,5-Trimethylhexane	C_9H_{20}	3522-94-9	D5134
2,3,4-Trimethylhexane	C_9H_{20}	921-47-1	D5134
2,3,5-Trimethylhexane	C_9H_{20}	1069-53-0	D5134
2,4,4-Trimethylhexane	C_9H_{20}	16747-30-1	D5134

(*continued overleaf*)

Table 3 (*continued*)

Compound	Formula	CAS no.	Method
2,2,3-Trimethylpentane	C_8H_{18}	564-02-3	D2268, D5134
2,2,4-Trimethylpentane	C_8H_{18}	540-84-1	D2268, D5134, D5443
2,3,3-Trimethylpentane	C_8H_{18}	560-21-4	D2268, D5134
2,3,4-Trimethylpentane	C_8H_{18}	565-75-3	D2268, D5134
2,4,4-Trimethyl-1-pentene	C_8H_{16}	107-39-1	D5441
2,3,4-Trimethyl-2-pentene	C_8H_{16}	565-77-5	D5441
2,4,4-Trimethyl-2-pentene	C_8H_{16}	107-40-4	D5441
3,4,4-Trimethyl-*trans*-2-pentene	C_8H_{16}	39761-57-4	D5441
n-Undecane	$C_{11}H_{24}$	1120-21-4	D5443
4-Vinyl-1-cyclohexene	C_8H_{12}	100-40-3	D2426
Water	H_2O	7732-18-5	D4815
m-Xylene	C_8H_{10}	108-38-3	D3797, D5060, D5134, D5580, D5769, D5986
o-Xylene	C_8H_{10}	95-47-6	D3797, D5060, D5134, D5443, D5580, D5769, D5986
p-Xylene	C_8H_{10}	106-42-3	D3797, D3798, D5060, D5134, D5580, D5769, D5986

Table 4 Common synonyms for analytes listed in Table 3

Synonym	Name used in Table 3	Synonym	Name used in Table 3
Allene	Propadiene	iC4	2-Methylpropane
Amylene	2-Methyl-2-butene	iC5	2-Methylbutane
Butadiene dimer	4-Vinyl-1-cyclohexene	Isobutane	2-Methylpropane
2-Butanethiol	1-Methyl-1-propanethiol	Isobutanethiol	2-Methyl-1-propanethiol
s-Butanethiol	1-Methyl-1-propanethiol	Isobutanol	2-Methyl-1-propanol
t-Butanethiol	2-Methyl-2-propanethiol	Isobutene	2-Methylpropene
s-Butanol	2-Butanol	Isobutyl alcohol	2-Methyl-1-propanol
t-Butanol	2-Methyl-2-propanol	Isobutylene	2-Methylpropene
Butyl alcohol	1-Butanol	Isobutyl mercaptan	2-Methyl-1-propanethiol
s-Butyl alcohol	2-Butanol	Isodurene	1,2,3,5-Tetramethylbenzene
t-Butyl alcohol	2-Methyl-2-propanol	Isoheptane	2-Methylhexane
s-Butylbenzene	2-Phenylbutane	Isohexane	2-Methylpentane
Butyl mercaptan	1-Butanethiol	Isooctane	2,2,4-Trimethylpentane
s-Butyl mercaptan	1-Methyl-1-propanethiol	Isopentane	2-Methylbutane
t-Butyl mercaptan	2-Methyl-2-propanethiol	Isopropanethiol	2-Propanethiol
Butyl methyl ketone	2-Hexanone	Isopropanol	2-Propanol
C1	Methane	Isopropenylbenzene	α-Methylstyrene
C2	Ethane	Isopropyl alcohol	2-Propanol
C2=	Ethylene	Isopropylether	Diisopropylether
C3	*n*-Propane	Isopropyl mercaptan	2-Propanethiol
C3=	Propene	Mesitylene	1,3,5-Trimethylbenzene
C4	*n*-Butane	Methyl acetylene	Propyne
Carbon oxysulfide	Carbonyl sulfide	Methyl alcohol	Methanol
Crotonylene	2-Butyne	Methylbenzene	Toluene
Cumene	Isopropylbenzene	Methyl disulfide	Dimethyl disulfide
trans-Decahydronaphthalene	*trans*-Decalin	(1-Methylethenyl)-benzene	α-Methylstyrene
Diisobutylene	2,4,4-Trimethyl-1-pentene	(1-Methylethyl)-benzene	Isopropylbenzene
Diisopropyl	2,3-Dimethylbutane	Methylethylketone	2-Butanone
Dimethyl acetylene	2-Butyne	Methyl ethyl sulfide	Ethyl methyl sulfide
1,2-Dimethylbenzene	*o*-Xylene	Methylketone	Acetone
1,3-Dimethylbenzene	*m*-Xylene	Methyl mercaptan	Methanethiol
1,4-Dimethylbenzene	*p*-Xylene	Methyl sulfide	Dimethyl sulfide
Dimethylketone	Acetone	Monoglyme	1,2-Dimethoxyethane
1,1-Dimethylpropyl methyl ether	*t*-Amyl methyl ether	MTBE	Methyl *t*-butyl ether
1,1-Dineopentylethylene	4,4-Dimethyl-2-neopentyl-1-pentene	nC4	*n*-Butane

Table 4 (*continued*)

Synonym	Name used in Table 3	Synonym	Name used in Table 3
DIPE	Diisopropylether	nC5	*n*-Pentane
Durene	1,2,4,5-Tetramethylbenzene	Neohexane	2,2-Dimethylbutane
ETBE	Ethyl *t*-butyl ether	Neopentane	2,2-Dimethylpropane
Ethene	Ethylene	*t*-Pentanol	2,2-Dimethyl-1-propanol
Ethyl acetylene	1-Butyne	Phenyl sulfide	Diphenyl sulfide
Ethyl alcohol	Ethanol	2-Propanone	Acetone
Ethyl disulfide	Diethyl disulfide	Propyl alcohol	1-Propanol
Ethylene glycol dimethyl ether	1,2-Dimethoxyethane	Propylene	Propene
Ethyl mercaptan	Ethanethiol	Propyl mercaptan	1-Propanethiol
Ethyl sulfide	Diethyl sulfide	Pseudocumene	1,2,4-Trimethylbenzene
2-Ethyltoluene	1-Methyl-2-ethylbenzene	TAME	*t*-Amyl methyl ether
3-Ethyltoluene	1-Methyl-3-ethylbenzene	Thionaphthene	Benzothiophene
4-Ethyltoluene	1-Methyl-4-ethylbenzene	*cis*,*cis*,*cis*-1,2,3-Trimethylcyclopentane	1,*cis*-2,*cis*-3-Trimethylcyclopentane
Ethyne	Acetylene	Vinyl acetylene	1-Buten-3-yne

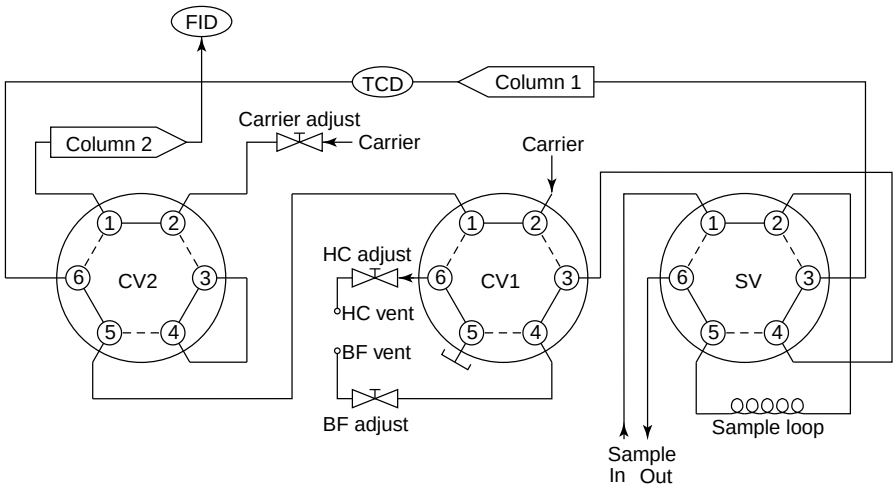

Figure 2 Plumbing diagram for heartcut and backflush-to-vent arrangement. HC = heartcut, BF = backflush, CV = column valve. (Adapted from Combs and Cook, Figure 4.22.[14] Copyright 1995, Chapman & Hall. Reprinted with kind permission from Kluwer Academic Publishers.)

with environmental regulations. ASTM methods D2360, D3054, D3760, D3797, D3798, D4291, D4735, D5060, and D6229 address impurity analyses of a more general scope. (Methods D2268, D2426, D2505, D2593, D2712, D4492, D4864, D5135, D5441, and D5713 are geared more towards specific applications.)

3.5 Comprehensive Analysis

An encompassing analysis of naturally occurring hydrocarbon mixtures and derived products is almost impossible, considering that these mixtures contain hundreds, if not thousands, of closely related compounds. The number of possible isomers for saturated alkanes with the general molecular formula C_xH_{2x+2} expands rapidly as the carbon number increases.[32] Between C_1 and C_6, there are 13 different isomers, but between C_7 and C_{12}, this number increases to 651. This does not even take cycloalkanes, olefins, and aromatics into account. A few ASTM methods specialize in analyzing a section of the hydrocarbon spectrum, identifying many of the components in complex mixtures: D5134, D5442, and D5443.

3.6 Heteroatom Analysis

During the production of petroleum products, it is frequently necessary to remove sulfur-containing components to minimize the sulfur content of the process output. This in turn reduces the generation of SO_2 during the combustion of fuels consumed for propulsion, heating, and energy production in power plants. Sulfur is also undesirable in the production of polymers from

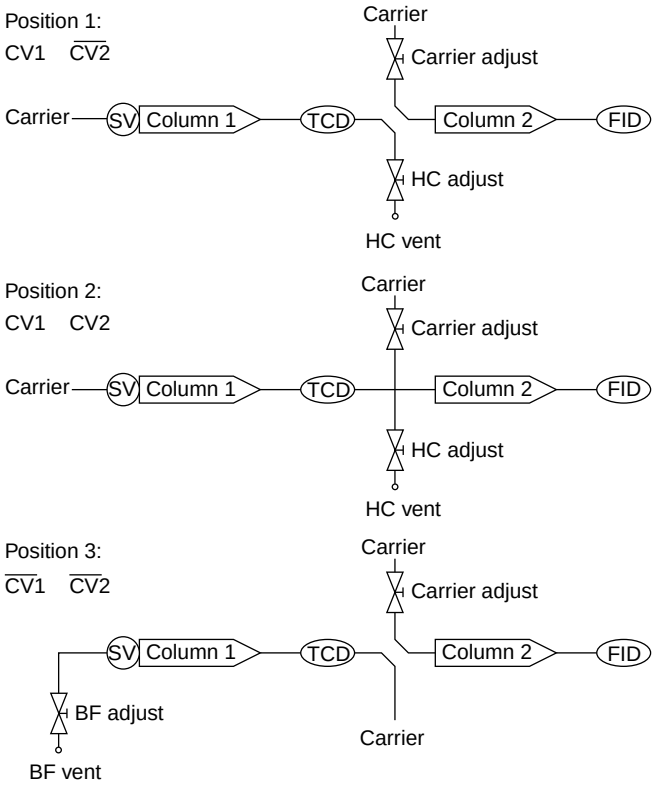

Figure 3 Valve operation diagrams. (Adapted from Combs and Cook[14] Figure 4.23, Copyright 1995, Chapman & Hall. Reprinted with kind permission from Kluwer Academic Publishers.)

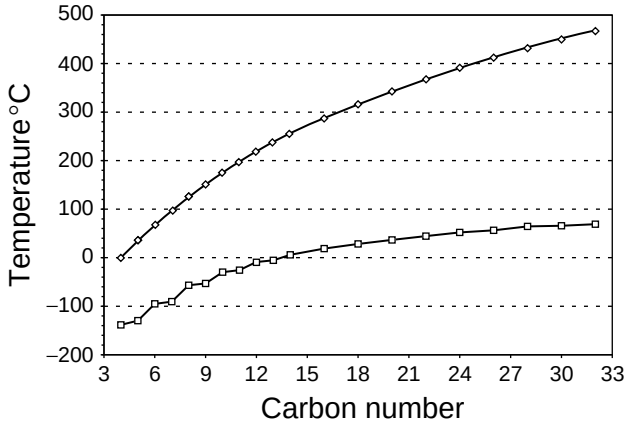

Figure 4 Melting points ( ) and boiling points (◇) for straight chain alkanes from C_4 through C_{32}.

olefins, because it renders many polymerization catalysts useless. In GC methods, sulfur-containing components are frequently detected with a flame photometric detector (FPD) or a pulsed flame photometric detector (PFPD). In these methods, a complete separation from other hydrocarbons is necessary to avoid a loss of the sulfur signal (quenching). This problem and the nonlinear response of FPDs are circumvented with the sulfur chemiluminescence detector (SCD), which is used in ASTM method D5623.

In order to meet governmental emission regulations without compromising driveability, today's gasoline blends usually contain specific amounts of oxygen-containing compounds such as ethers or alcohols. The concentration of these ''oxygenate'' additives does not reach trace levels, so they are normally analyzed with general-purpose detectors (FID, TCD). However, for oxygen-specific detection, an oxygen-selective flame ionization detector (OFID) can be used, and ASTM method D5599 takes advantage of this technology.

Other ASTM methods that target oxygen- or sulfur-containing compounds are D5008, D5303, D5504, D5541, and D6228 (see also D4735, D4806, D4815, D5441, and D5501).

3.7 Specific Compounds

ASTM offers a number of methods that deal with the analysis of specific nontrace target analytes, usually for purity analysis: D2268, D3257, D3606, D4367, D4492, D5135, D5501, D5713, and D5441 (see also D3054, D3760, D3797, D3798, D4291, D4735, D5008, and D5060).

4 ON-LINE APPLICATIONS

A number of hydrocarbon analysis problems are typical applications for PGCs and some of them are outlined in this section. The main differences between laboratory (off-line) and process (on-line) GCs are outlined in Table 1. In addition, PGCs are calibrated almost always externally owing to the difficulty of mixing internal standards into a process stream.

4.1 Natural Gas Analysis

In most cases, the analysis of natural gas with GC is straightforward, including ''fixed gases'' such as H_2, He, N_2, O_2, CO, and CO_2, which can be detected with a TCD using hydrogen or helium carrier gases. If H_2 and He need to be detected, a nitrogen carrier gas is generally used. ASTM methods D1945, D1946, and D2163 (see also D6228) describe the analysis of natural gases. In some cases, CO or CO_2 may have to be measured at trace levels, where FID detectors are usually employed, but neither of these two analytes can be detected directly with an FID. In order to overcome this limitation, these two components are passed through a methanator that converts them to CH_4, which in turn is easily detectable by an FID.

4.2 Gasoline Blending

Gasoline and other propulsion fuels have to bridge several goals, which sometimes seem to conflict with each other: environmental regulations, which vary from country to country and even from state to state, fuel evaporation rates based on seasonal temperatures, driveability (octane number), consumer prices, compatibility with combustion engines manufactured throughout the twentieth century, different fuel grades, and more. Several parameters such as Reid vapor pressure (RVP), total aromatics, total sulfur, total oxygenates, and total olefins are measured to control and adjust the production process,[33] and a couple of ASTM methods assist in characterizing finished fuel products to ensure consistent quality: D2427, D4420, D4806, D4815, D5580, and D6296. There are also a number of other ASTM methods that are not necessarily fuel specific, but may also be needed to analyze additives that are used during the gasoline blending process: D3606, D5441, D5501, D5504, D5541, D5599, D5623, D5769, and D5986.

4.3 Polymer Production Feedstock

A very significant portion of polymer products are manufactured using ethylene, propylene, and other olefin precursors. In most cases, these olefins must meet certain purity standards in order to be usable for polymer production. For example, ''polymer grade'' ethylene must be >99.9% pure, and typical trace impurities are methane and ethane ($CH_4 + C_2H_6 < 0.15\%$), acetylene (<5 ppm), and CO_2 (<10 ppm).[34] ASTM methods that analyze impurities in ethylene, propylene, and butadiene are D2426, D2505, D2593, D2712, D4424, D4864 (see also D2163 and D5303).

5 OFF-LINE APPLICATIONS

Most on-line applications can also be employed in off-line analyses, and vice versa, but the two application areas outlined below are usually done in a laboratory setting.

5.1 Hyphenated Techniques

Many of the previously discussed GC detectors (TCD, FID, SCD) are single-channel detectors, which means that they provide a single signal for each measured point along a chromatogram. Their simplicity can also lead to problems, when target analyte peaks overlap or merge with interferences, or when unknown components need to be identified. Single-channel detector responses cannot discriminate between closely eluting analytes, and identification based on retention time is risky at best. This problem is usually solved by linking or ''hyphenating'' GC instruments with general-purpose instruments such as mass spectrometers, which provide an entire spectrum for each measured point along a chromatogram.[35] Most compounds give rise to different mass spectra, so that a computer-based separation of overlapped components[36] or the identification of unknown components becomes possible.

One of the shortcomings of MS, the inability to distinguish between isomers, was recognized as well. Infrared spectroscopy (IRS) overcomes this problem, but it took much longer to develop a practical interface between IRS and GC.[37] Compared with GC/MS, gas chromatography/infrared spectroscopy (GC/IRS) is much less used today because of higher instrumentation cost, lower reliability, and lower sensitivity.[38] This is particularly true for hydrocarbon analysis with GC/IRS. In rare cases, when the analysis of organometallic compounds is necessary, atomic emission spectroscopy (AES) can also be used as a GC detector for general hydrocarbon analysis.[39,40]

There are many applications for GC/MS[41,42] and most GC laboratories, if not all of them, use MS in at least one of their GC instruments to maximize the versatility of their analytical capabilities. ASTM defines only one GC/MS method, D5769, for the analysis of finished gasoline. For GC/IRS as a much less widely applied technique, ASTM offers a standard method as well. The GC/IRS method D5986 expands the scope of the GC/MS method D5769 in order to distinguish between isomers such as dimethyl- and trimethylbenzenes.

5.2 Bulk Properties

Chromatograms contain a host of information pertaining not only to the concentration of the sample constituents, but also to so-called ''bulk properties'' that are not tied to any particular compound, but to the entire mixture. Such properties include BTU values, specific gravity, vapor-to-liquid ratio (V/L), RVP, carbon numbers, and boiling point distribution. In particular, the determination of boiling point distributions with GC (simulated distillation) is of interest in many areas of the petroleum processing industry,[43] because it avoids costly, time-consuming, and error-prone analyses with actual distillation equipment. Simulated distillation with GC is actually quite simple: the system is calibrated with a mixture of hydrocarbons of known boiling points, typically *n*-alkanes, and several representative aromatics and branched alkanes. During a temperature-programmed separation, the components of the calibration mixture elute at known times and therefore calibrate the time axis as a ''temperature axis''. When an actual sample is analyzed, the total chromatogram area is calculated, and then the first *n*% of this area are correlated to the temperature axis to

Table 5 Vendors of process gas chromatographs

Vendor	Location	Internet
ABB Process Analytics	Lewisburg, WV	www.abb.com/usa
Antek Instruments	Houston, TX	www.antekhou.com
Daniel Industries	Houston, TX	www.danielind.com
Rosemount Analytical (Fisher-Rosemount)	Orrville, OH	www.frco.com/proanalytic
Fluid Data (Onix Process Analytics)	Angleton, TX	www.fluid-data.com
Foxboro	Foxboro, MA	www.foxboro.com
Hewlett-Packard	Wilmington, DE	www.hp.com/go/chem
HNU Systems	Newton, MA	www.hnu.com
Honeywell Industrial Automation & Control	Phoenix, AZ	www.iac.honeywell.com
Houston-Atlas (Onix Process Analytics)	Kingwood, TX	www.houstonatlas.com
Siemens	Munich, Germany	www.ad.siemens.de
Siemens Applied Automation	Bartlesville, OK	www.aai-us.com
Yokogawa Corporation of America	Newnan, GA	www.yca.com

Table 6 Internet resources

Organization	Location	Internet
American Gas Association	Arlington, VA	www.aga.com
American National Standards Institute (ANSI)[a]	New York City, NY	www.ansi.org
American Petroleum Institute	Washington, DC	www.api.org
ASTM[a]	West Conshohocken, PA	www.astm.org
Deutsches Institut für Normung (DIN)[a]	Berlin, Germany	www.din.de
Gas Processors Association (GPA)[a]	Tulsa, OK	www.gasprocessors.com
Institute of Petroleum (IP)[a]	London, United Kingdom	www.petroleum.co.uk
Institut Français du Pétrole (IFP)	Rueil-Malmaison, France	www.ifp.fr
International Organization for Standardization (ISO)[a]	Geneva, Switzerland	www.iso.ch
Varian Associates[b]	Palo Alto, CA	www.varian.com

[a] Publishes standard methods for hydrocarbon analysis with GC, titles accessible via the Internet.
[b] Publishes a number of application notes on the Internet.

obtain ''temperature versus percent off'' value pairs. ASTM defines several methods for simulated distillation (D2887, D3710, D5307, D5399, and D5480), which have to operate within certain temperature and sample constraints.

ACKNOWLEDGMENTS

I wish to acknowledge Ulrich Gökeler (Siemens AG) and Christopher Schalleur (Honeywell Industrial Automation) for providing me with application notes for systems manufactured by their respective companies. I also wish to thank Bob Bade and Jerry Combs for providing valuable input during the writing of this article.

For reference purposes, major vendors of PGCs are listed in Table 5, and most of them provide some information about their gas chromatographs that go beyond mere advertising. Aside from ASTM, many organizations provide standards or application notes for GC hydrocarbon analysis, and many of them are accessible or searchable on the Internet as well (see Table 6).

ABBREVIATIONS AND ACRONYMS

All abbreviations and acronyms used in this article are given below, but standard chemical formulas such as C_2H_6 for ethane or CO for carbon monoxide are not listed.

AES	Atomic Emission Spectroscopy
ASTM	American Society for Testing and Materials
BTU	British Thermal Unit
CV	Column Valve
DIPE	Diisopropyl Ether
ETBE	Ethyl *t*-Butyl Ether
FID	Flame Ionization Detectors
FPD	Flame Photometric Detector
GC	Gas Chromatography
GC × GC	Comprehensive Two-dimensional Separation with Gas Chromatography
GC/IRS	Gas Chromatography/Infrared Spectroscopy

GC/MS	Gas Chromatography/Mass Spectroscopy
IRS	Infrared Spectroscopy
MDGC	Multidimensional Gas Chromatography
MS	Mass Spectroscopy
MTBE	Methyl *t*-Butyl Ether
OFID	Oxygen-selective Flame Ionization Detector
PFPD	Pulsed Flame Photometric Detector
PGC	Process Gas Chromatography
ppb	parts per billion
ppm	parts per million
RVP	Reid Vapor Pressure
SCD	Sulfur Chemiluminescence Detector
SV	Sample Valve
TAME	*t*-Amyl Methyl Ether
TCD	Thermoconductivity Detector
V/L	Vapor-to-Liquid Ratio

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of • High-temperature Simulated Distillation Applications in Petroleum Characterization

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis • Chromatography in Process Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography • Multidimensional Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Literature Searching Methodology • Quality Assurance in Analytical Chemistry

REFERENCES

1. G.A. Eiceman, H.H. Hill, Jr, J. Gardea-Torresdey, 'Gas Chromatography', *Anal. Chem.*, **70**, 321R–339R (1998).
2. R.L. Grob (ed.), *Modern Practice of Gas Chromatography*, 3rd edition, John Wiley & Sons, New York, 1995.
3. G. Schomburg, *Gas Chromatography. A Practical Course*, VCH, Weinheim, 1990.
4. R.P.W. Scott, *Introduction to Analytical Gas Chromatography*, 2nd edition, Marcel Dekker, New York, 1998.
5. H.M. McNair, J.M. Miller, *Basic Gas Chromatography*, John Wiley & Sons, New York, 1998.
6. W. Jennings, E. Mittlefehldt, P. Stremple, *Analytical Gas Chromatography*, 2nd edition, Academic Press, San Diego, CA, 1997.
7. J.P. Durand, 'Le Rôle de la CPG dans l'Industrie Pétrolière et la Pétrochimie. De l'Analyse Détaillée des Hydrocarbures à la Distillation Simulée', *Analusis*, **26**, M17–M21 (1998).
8. J.P. Durand, S. Gautier, E. Robert, M.C. Guilhem, R. Phan-Tan-Luu, 'Experimental Design Optimization of the Analysis of Gasoline by Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **20**, 289–294 (1997).
9. C.T. Mansfield, B.N. Barman, J.V. Thomas, A.K. Mehrotra, R.P. Philp, 'Petroleum and Coal', *Anal. Chem.*, **69**, 59R–93R (1997).
10. F. McLennan, 'Process Analytical Chemistry in Perspective', in *Process Analytical Chemistry*, eds. F. McLennan, B.R. Kowalski, Blackie Academic and Professional, Glasgow, 1–13, 1995.
11. R. Annino, R. Villalobos, *Process Gas Chromatography. Fundamentals and Applications. On-line Analysis for Process Monitoring and Control*, Instrument Society of America, Research Triangle Park, NC, 1992.
12. G. Guiochon, C.L. Guillemin, *Quantitative Gas Chromatography. For Laboratory Analyses and On-line Process Control*, Elsevier, Amsterdam, 1988.
13. R. Villalobos, 'A Window Diagram For Key Component Analysis in On-line Gas Chromatography', *J. High Resolut. Chromatogr.*, **18**, 343–347 (1995).
14. G.L. Combs, R.L. Cook, 'Process Chromatography', in *Process Analytical Chemistry*, eds. F. McLennan, B.R. Kowalski, Blackie Academic and Professional, Glasgow, 106–158, 1995.
15. G. Dascenzo, N. Nicolini, 'Evolution of Paper Chromatography – From Runge Perceptions to Tswett Adsorption', *Fresenius' J. Anal. Chem.*, **337**, 232–234 (1990).
16. J.C. Touchstone, 'History of Chromatography', *J. Liquid Chromatogr.*, **16**, 1647–1665 (1993).
17. Y.I. Yashin, 'The Ninety-year History of Chromatography (1903–1993)', *J. Anal. Chem.*, **49**, 939–949 (1994).
18. M.S. Tswett, 'Physikalisch-chemische Studien über das Chlorophyll. Die Adsorptionen', *Ber. Deutsch. Bot. Ges.*, **24**, 316–332 (1906).
19. L.S. Ettre, K.I. Sakodynskii, 'Tswett, M.S. and the Discovery of Chromatography. 1. Early Work (1899–1903)', *Chromatographia*, **35**, 223–231 (1993).
20. L.S. Ettre, K.I. Sakodynskii, 'Tswett, M.S. and the Discovery of Chromatography. 2. Completion of the

Development of Chromatography (1903–1910)', *Chromatographia*, **35**, 329–338 (1993).
21. A.J.P. Martin, R.L.M. Synge, 'A New Form of Chromatogram Employing Two Liquid Phases', *Biochem. J.*, **35**, 1358–1368 (1941).
22. A.T. James, A.J.P. Martin, 'Gas–Liquid Partition Chromatography. A Technique for the Analysis of Volatile Materials', *Analyst*, **77**, 915–932 (1952).
23. A.T. James, A.J.P. Martin, 'Gas–liquid Partition Chromatography: the Separation and Microestimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid', *Biochem. J.*, **50**, 679–690 (1952).
24. http://www.nobel.se/laureates/chemistry-1952.html
25. J.C. Holmes, F.A. Morrell, 'Oscillographic Mass Spectrometric Monitoring of Gas Chromatography', *Appl. Spectrosc.*, **11**, 86–87 (1957).
26. Hernan J. Cortes (ed.), *Multidimensional Chromatography. Techniques and Applications*, Marcel Dekker, New York, 1990.
27. C.J. Venkatramani, J. Xu, J.B. Phillips, 'Separation Orthogonality in Temperature-programmed Comprehensive Two-dimensional Gas Chromatography', *Anal. Chem.*, **68**, 1486–1492 (1996).
28. C.A. Bruckner, B.J. Prazen, R.E. Synovec, 'Comprehensive Two-dimensional High-speed Gas Chromatography with Chemometric Analysis', *Anal. Chem.*, **70**, 2796–2804 (1998).
29. E.B. Overton, K.R. Carney, 'New Horizons in Gas Chromatography – Field Applications of Microminiaturized Gas Chromatographic Techniques', *TrAC, Trends Anal. Chem.*, **13**, 252–257 (1994).
30. R. Sacks, H. Smith, M. Nowak, 'High-speed Gas Chromatography', *Anal. Chem.*, **70**, 29A–37A (1998).
31. ChemFinder, CambridgeSoft (Cambridge, MA), http://www.chemfinder.com/.
32. L. Bytautas, D.J. Klein, 'Chemical Combinatorics for Alkane-isomer Enumeration and More', *J. Chem. Inf. Comput. Sci.*, **38**, 1063–1078 (1998).
33. D. Thomas, 'Analyzers Reduce Gas Blending ''Give-away'' ', *InTech*, **38**(11), 30–31 (1991).
34. *Composition Measurement in Ethylene Plants*, Industry Applications Bulletin, 7005037-001 Applied Automation, Inc., Bartlesville, OK, 1992.
35. N. Ragunathan, K.A. Krock, C. Klawun, T.A. Sasaki, C.L. Wilkins, 'Multispectral Detection for Gas Chromatography', *J. Chromatogr. A*, **703**, 335–382 (1995).
36. B.N. Colby, 'Spectral Deconvolution for Overlapping GC/MS Components', *J. Am. Soc. Mass Spectrom.*, **3**, 558–562 (1992).
37. L.V. Azarraga, 'Gold Coating of Glass Tubes for Gas Chromatography/Fourier Transform Infrared Spectroscopy 'Light Pipe' Gas Cells', *Appl. Spectrosc.*, **34**, 224–225 (1980).
38. R. White, *Chromatography/Fourier Transform Infrared Spectroscopy and Its Applications*, Marcel Dekker, New York, 1990.
39. J.J. Kosman, R.G. Lukco, 'The Use of GC/AES Multielement Simulated Distillation for Petroleum Product Fingerprinting', *J. Chromatogr. Sci.*, **31**, 88–94 (1993).
40. B.D. Quimby, P.C. Dryden, J.J. Sullivan, 'Selective Detection of Volatile Nickel, Vanadium, and Iron Porphyrins in Crude Oils by Gas-chromatography Atomic-emission Spectroscopy', *J. High Resolut. Chromatogr.*, **14**, 110–116 (1991).
41. F.W. Karasek, R.E. Clement, *Basic Gas Chromatography–Mass Spectrometry. Principles and Techniques*, Elsevier, Amsterdam, 1988.
42. F.G. Kitson, B.S. Larsen, C.N. McEwen, *Gas Chromatography and Mass Spectrometry. A Practical Guide*, Academic Press, San Diego, CA, 1996.
43. J.A. Crandall, R.K. Bade, D.E. Clay, R.E. Warner, S.B. Hunt, 'Simulated Distillation Application of Process Gas Chromatography: The Real World', *Adv. Instrum. Control*, **45**, 605–637 (1990).

Lube Products. Molecular Characterization of Base Oils

Luciano Montanari
EniTecnologie S.p.A., San Donato Milanese, Italy

1 Introduction 1
2 Base Oil Manufacture 2
2.1 Mineral Oils 2
2.2 Hydrocracked Base Oils 2
2.3 Hydroisomerized Base Oils 2
2.4 Polyolefins 3
3 Base Oil Properties 3
4 Chemical Composition 4
5 Conventional Base Oil Testing and Analysis 4
6 Molecular Characterization of Base Oils by Nuclear Magnetic Resonance 4
Acknowledgments 8
Abbreviations and Acronyms 8
Related Articles 8
References 9

Base oils used in the formulation of lubricant products are a complex mixture of paraffinic (prevalently branched), aromatic and naphthenic (cycloparaffinic) molecules ranging in carbon number from 20 to 40+. Most automotive crankcase oils, accessory fluids, and many industrial lubricants contain in excess of 80% base oil in their "finished" formulation.

The compositional analysis of Lube base oils is usually obtained by a two-step procedure. The first step separates the aromatic and saturated fractions by column chromatography (American Society for Testing and Materials, ASTM D2549). The aromatic fraction is then further characterized by mass spectrometry (ASTM D3239) and the saturated subclasses of paraffins and cycloparaffins are measured by a different mass spectrometry method (ASTM D2786).

Recently the Lube base oils have been characterized at the molecular level by means of nuclear magnetic resonance (NMR) spectroscopy. The conventional NMR spectra have the problem of signal overlap due to the presence of a great number of isomers in the base oil composition.

To overcome the problem of signal overlap, some selected multiplet subspectral carbon-13 nuclear magnetic resonance (^{13}C-NMR) analyses were developed and, in particular, gated spin echo (GASPE) and distortionless enhancement by polarization transfer (DEPT) have been applied to obtain quantitative CH_n subspectra in petroleum fractions.

Usually the quantitative data obtained from GASPE have a better accuracy than those from DEPT and it will be shown how GASPE is applied to study the molecular characteristics of the aliphatic moieties of base oils of a different nature.

On the basis of the GASPE data and of the average molecular mass obtained via vapor pressure osmometry (VPO), the average number of branches and aliphatic rings can be calculated. Moreover, on the basis of GASPE, the distribution of side chain lengths and the positions of the methyl groups along the straight chain (methyls are the only groups for which the positions are directly obtained from NMR spectra) were obtained.

All these structural features are related to the finished lubricant's performance in both crankcase and industrial applications.

1 INTRODUCTION

Substances in all states of matter can be used as lubricants provided that they reduce friction and wear between materials in contact.

The majority of lubricants are liquids. They can be classified according to their origin into two main groups: mineral oils derived from petroleum (the majority); and synthetic oils. Most liquid lubricants contain additives to improve their natural properties, or give new properties to them, so they have qualities suitable for their applications and their efficiency may be prolonged.

Mineral lubricating oils are obtained from crude oils: the chemical composition of the crude petroleum has a great influence on obtaining lubricating oils of the required properties. In terms of chemical composition, crude oils are classified into alkanic (or paraffinic) crude oils with more than 50% alkanes, cycloalkanic or cyclic or naphthenic crude oils with more than 50% cycloalkanes, and aromatic crude oils with more than 50% aromatics. Alkanic crude oils are mostly light, wax bearing and of low sulfur content. Cycloalkanic crude oils are often heavy, wax-free or of low wax content and may contain sulfur compounds. In greater detail, the hydrocarbon groups present in the crude oils are straight-chain and branched alkanes, alkylcycloalkanes, alkylaromatics and alkylcycloaromatics.

From a property point of view, comparison of different hydrocarbon classes of the same carbon number shows that *straight-chain alkanes* (*n*-alkanes, *n*-paraffins) have the lowest viscosity, the highest viscosity index

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

(VI > 200), the lowest evaporation loss and the highest flash point. Their pour point is high (40 °C or more) and in high concentration they are undesirable. Generally they are removed from lubricating oils by solvent dewaxing or catalytic cleavage.

Branched alkanes (isoalkanes, isoparaffins) differ from the *n*-alkanes by having alkyl substituents on the main chain. If the alkyl side chains are short and not very numerous, the isoalkanes are similar to the *n*-paraffins. On the contrary, if the alkyls are long and numerous, isoalkanes have a low pour point, a relatively high VI, low volatility and evaporation loss and a high flash point. Such isoalkanes are usually present in small amounts in the conventional mineral oils, but they are predominant in the unconventional oils as those obtained by hydrocracking and hydroisomerization processes (see section 2.3) or manufactured synthetically (polyolefins).

Alkylcycloalkanes (naphthenes) of the same carbon number have higher viscosity, lower VI, lower flash point and higher evaporation loss than the alkanes and the isoalkanes, but their pour points are generally lower.

Alkylaromatics are characterized by one or more aromatic rings: aromatic rings have a greater effect than the cycloalkanic ring on the viscosity of lubricating oil cuts and at the same time decrease the VI.

2 BASE OIL MANUFACTURE

2.1 Mineral Oils

The manufacture of lubricating oils from crude oils includes separation, refining and dewaxing processes. The first step is the separation process, which is the distillation of the crude oil to produce gasoline, kerosene, gas oil and an atmospheric residue. The viscous oil cuts are obtained from the atmospheric residue by vacuum distillation. Usually the temperature does not exceed 410 °C. Vacuum distillation is normally used to produce three or four lubricating oil distillates and vacuum asphalt. Lubricating oil distillates differ from one another by distillation range, average molecular weight, viscosity, and flash and pour points. The main criteria used are the viscosity and the flash point.

The second step is the refining, which substantially changes the chemical composition of lubricating oil distillates and can thus control the properties of the distillates. Its primary effect is to reduce the concentration of the most polar constituents of the distillates. The object of this process is to improve color, the VI, and long-term stability. Nowadays the most diffuse refining process is the solvent refining, which consists of separating wax-free oil cuts into two layers: one rich in less-polar components and poor in solvent, called the raffinate layer; the other rich in polar components and solvent, called the extract layer. The most common selective solvents are furfural, which is suitable for all oil stocks, and phenol, which is used more for waxy and heavy cut oils.

The last step is the dewaxing process, which is the dominant process in the recovery of oils of low pour point and good rheological properties at low temperatures from raffinates. The oil is diluted with the solvent, chilled to low temperature and the separated wax crystals are filtered off. Suitable solvents include propane, ethane and the so-called solvent/antisolvent mixtures, such as methyl ethyl ketone (MEK).

The processes described are used for the production of base oils, which are, nowadays, regarded as conventional ones.

2.2 Hydrocracked Base Oils

Although modern types of mineral oils reinforced by additives are capable of covering a broad range of lubricating requirements, their properties are in some cases significantly surpassed by the unconventional base oils, which include hydrocracked oils, wax hydroisomerized oils and polyolefin oils. Hydrocracking of oils consists of the high-pressure hydrogenation of oils by using suitable catalysts. It enables oils to be obtained with a high VI or even an extremely high VI (up to 150). Hydrogenation changes the composition of all homogeneous oil groups: the viscosity, density, sulfur content and aromatic content are all reduced, whereas the VI, hydrogen content, concentration of cycloalkanic compounds and rings are all increased. Hydrocracking enables high-grade and high-VI oils to be produced economically. Oils of this quality cannot be obtained by solvent refining. The superiority of hydrocracked oils is in their very high VI (up to 130), good response to the additives, lower volatility at the same viscosity and lower biological activity. Less convenient characteristics are poorer stability to light, higher corrosiveness after oxidation, and lower solvent power for polar substances, which are originated by oxidation.

The hydrocracked oils are generally solvent dewaxing and, because of this, their pour points are not very low.

2.3 Hydroisomerized Base Oils

More recently, catalysts containing a hydrogenation component on a molecular sieve have been used for wax isomerization. They have permitted the development of a new process for lowering the pour points of high-wax feeds – a once-through process without requiring a solvent dewaxer – and therefore can be used to produce ultrahigh-VI oils with a low pour point. The oils obtained by this process are called wax hydroisomerized oils.

2.4 Polyolefins

The synthetic oils are the highest performance oils. They are termed synthetic because they are not obtained directly from petroleum. In comparison to mineral oil products, synthetic lubricants have the disadvantage of much higher cost. They can be classified into the following categories:

- synthetic hydrocarbons;
- polyether oils;
- esters of carboxylic acids;
- phosphoric acid esters;
- silicon-containing oils;
- halogenated hydrocarbons and halocarbons.

Among these classes of synthetic oils, we consider only the synthetic hydrocarbons because analysis of the others is not general but is specific for each product.

Synthetic hydrocarbons are produced by polymerization of olefins in the presence of Lewis acids (e.g. Al_2Cl_6, BF_3 and their complexes) or Ziegler catalyst (e.g. $Al(C_2H_5)_{3-x}Cl_x$ and promoters) in the liquid phase. They offer a number of advantages: they have narrow boiling ranges and contain small quantities of unsaturated or polycyclic aromatic compounds and traces of nitrogen-, sulfur-, or oxygen-containing impurities. They have favorable low-temperature characteristics and respond well to additive treatment. Their variable temperature characteristics are good.

In recent years, poly(α-olefins) have received considerable attention as high-performance functional fluids in a variety of applications. They are manufactured in the presence of Lewis acids by oligomerization of 1-alkenes in the range C_8–C_{12}, followed by hydrogenation of the double bond: they are almost pure isoalkanes. These poly(α-olefins) are characterized by high thermal stability, low viscosities even at low temperature, low pour points (down to $-80\,^{\circ}C$), high VI (up to 195), high flash point and low volatility. Almost all of the poly(α-olefins) available today are produced from 1-decene. They can be classified into different grades depending on their $100\,^{\circ}C$ viscosities.

3 BASE OIL PROPERTIES

Lubricants are formulated by blending base oils and additives to meet a series of performance specifications. These specifications relate to the physical properties of the oil when it is new and also ensure that the oil continues to function and protect the engine or machinery in service. Self-evidently, lubricant performance is determined by the base oils and the additives that are used in the formulation.

When selecting the appropriate base oil to use in a formulation, there are a range of properties that can be measured and used to predict performance. Although many of these properties are modified or enhanced by the use of additives, knowledge of the base oil characteristics is vital for effective formulation of any lubricant. The complexity of chemical composition of the base oils requires that most measurements are of bulk physical or chemical properties that indicate the average performance of all the molecular types in the base oil.

The main physical properties are:

- *Viscosity*. This is a measurement of the internal friction within a liquid. It is a vital property of a lubricant because it influences the ability of the oil to form a lubricating film or to minimize friction and reduce wear. The unit of absolute viscosity is the Pascal second ($Pa \cdot s$), although the centipoise (cP) is often used as an alternative unit ($1\,Pa \cdot s = 10^3$ cP).
- *Viscosity/temperature relationship – VI*. The most frequently used method to compare the variation of viscosity with temperature between different oils is by calculation of a dimensionless number, known as the VI. The higher the VI, the less the effect of temperature on the viscosity of the sample.
- *Low-temperature properties*. When a sample of oil is cooled, its viscosity increases in a predictable manner until wax crystals start to form. The matrix of wax crystals becomes sufficiently dense with further cooling to cause an apparent solidification of the oil, but it has not undergone a true phase change. Many lubricating oils have to be capable of flow at low temperatures and a number of properties should be measured. Cloud point is the temperature at which the first signs of wax formation can be detected. Pour point is the lowest temperature at which the sample of oil can be made to flow by gravity alone.
- *High-temperature properties*. The high-temperature properties of oil are governed by the distillation or boiling range characteristics of the oil. Volatility is important because it is an indication of the tendency of the oil to be lost in service by vaporization. Several methods are used to characterize volatility but the most used is the Noack method, where the sample is heated for 1 h at $250\,^{\circ}C$ and the weight loss is measured. The flash point of an oil is important from a safety point of view because it is the lowest temperature at which autoignition of the vapor occurs.

Various other physical properties may be measured, most of them relating to specialized lubricant applications.

The most important chemical property of base oils is their resistance to thermal–oxidative degradation.

Although the formulated lubricant may have many desirable properties when new, oxidation can lead to a dramatic loss of performance in service by oxidative reactions. A variety of different stability tests have been devised to measure resistance to oxidation under different conditions.

4 CHEMICAL COMPOSITION

Even if the properties of lubricating oils are largely due to the additives, the importance of the chemical composition of base oils on the finished lubricant performance is increasing more and more and this is particularly true for improved fuel efficiency, which is being driven by both legislation and product specification. This requires a deep knowledge of the molecular characteristics of the base oils, which is not easy because they consist of large numbers of structurally different compounds. A large number of correlations based on the relative distribution of paraffinic-, aromatic- and naphthenic-type hydrocarbons are reported in the literature[1–7] but these studies do not involve the nature and distribution of the aliphatic moieties, which, vice versa, determine the low-temperature and rheological properties of base stocks. Proton nuclear magnetic resonance (^{1}H-NMR) and ^{13}C-NMR techniques have been used to estimate structural parameters in petroleum cuts[8–14] but they have limitations in resolving the resonance of CH_n ($n = 0–3$) groups in the aliphatic region.

To overcome the problem of signal overlap, some selected multiplet subspectral ^{13}C-NMR analyses were developed and, in particular, GASPE has been applied to obtain quantitative abundances of CH_n groups in petroleum fractions.[15–19] Recently GASPE/NMR[20] was used to compare different base oils: mineral, hydrocracked, wax hydroisomerized, poly(α-olefins) and poly(internal olefins). From GASPE spectra the fractional abundances of CH_n, the number and length of branches, the position of methyl groups along the straight chain and the average number of aliphatic rings were obtained.

5 CONVENTIONAL BASE OIL TESTING AND ANALYSIS

The testing of lubricants is based on chemical and physical properties and on parameters that represent statistical means of the molecular properties. The essential methods have been standardized on an international basis by organizations such as the ASTM; USA, the Institute of Petroleum (IP; UK), the Deutsches Institut fur Normung (DIN; Germany) and the International Organization for Standardization (ISO). Hereafter, the most diffuse tests for analysis of physical properties and hydrocarbon type are summarized:

- *Viscosity*. Kinematic viscosity is measured routinely with ease and great precision in capillary viscometers suspended in constant-temperature baths. Standard methods are ASTM D445 and IP 71, and a number of standard temperatures are used, especially 40 °C and 100 °C.
- *Viscosity/temperature relationship – VI*. VI are calculated by measuring the kinematic viscosities of the sample at two different temperatures (40 °C, 100 °C) and comparing the viscosity change with an empirical reference scale. Full definitions of the methods of calculation are given in ASTM or IP manuals (ASTM D2270, IP 226).
- *Low-temperature properties*. The pour points are measured with ASTM D97 and IP 15, in which the oil is warmed and then cooled at a specified rate until the lowest temperature is reached at which the sample can be made to flow by gravity alone.
- *Volatility*. This is important because it is an indication of the tendency of oil to be lost in service by evaporation. Generally it is measured with the Noack method, where the sample is heated for 1 h at 250 °C and the weight loss is measured (DIN 51581).
- *Hydrocarbon analysis*. The detailed hydrocarbon analysis is obtained by a two-step procedure. The first step separates the aromatic and saturated fractions by column chromatography (ASTM D2549). The aromatic fraction is then characterized further by mass spectrometry (ASTM D3239) and the saturated subclasses of paraffins and cycloparaffins are measured by a mass spectrometry method (ASTM D2786).

6 MOLECULAR CHARACTERIZATION OF BASE OILS BY NUCLEAR MAGNETIC RESONANCE

NMR is the most important spectroscopic technique available for molecular characterization of the hydrocarbon skeleton. On the basis of the positions and intensities of signals, it is possible to identify the types and number of different nuclei, especially H and C in organic compounds. Some important books[21–23] describe the fundamentals of this spectroscopy, which is widely used in different fields such as chemistry, physics, medicine and pharmaceutics.

^{1}H- and ^{13}C-NMR techniques have been utilized also for the estimation of structural parameters of low- and higher boiling petroleum cuts. The ^{1}H-NMR spectra of

base oils consist of two main types of signal (Figure 1): a broad resonance at 6.8–8 ppm due to H on aromatic carbons; and signals at 2–0.5 ppm due to H and aliphatic carbons. In the aliphatic regions it is possible to distinguish the H of methyl groups of paraffinic chains at nearly 0.8 ppm and a sharp line at nearly 1.25 ppm due to CH_2 of paraffinic chains.

The ^{13}C-NMR spectra of base oils (Figure 2) show the signals due to aromatic carbons between 100 and 150 ppm and those of the aliphatic carbons in the region 0–50 ppm.

Particularly in the aliphatic regions there are a great number of peaks due to branched and cycloparaffins: some structural assignments have been proposed[24] but they refer predominantly to the prominent peaks, whereas

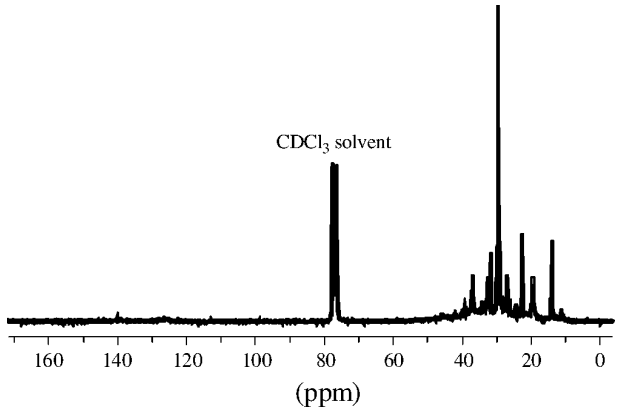

Figure 2 ^{13}C-NMR spectrum of base oil A (Table 1).

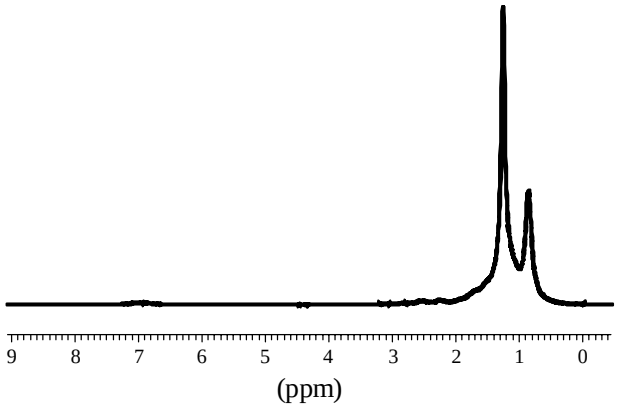

Figure 1 ^{1}H-NMR spectrum of base oil A (Table 1).

a great part of the signal is due to the multiplicity of weak resonances spread over the spectral width. The isomeric forms present originate from these latter; for instance, there are over 360 000 noncyclic alkane isomers for C_{20} alkanes and this number rapidly increases for heavier alkanes. In conventional ^{13}C-NMR spectra it is very difficult to resolve the resonances of CH_n ($n = 0$–3) in the aliphatic region.

To overcome this problem the GASPE technique has been quite useful because it allows the generation of quantitative CH_n subspectra.

An example of the application of the GASPE technique is shown in Figure 3: the ^{13}C-NMR spectrum of the squalane molecule (Figure 3), which is used as a model

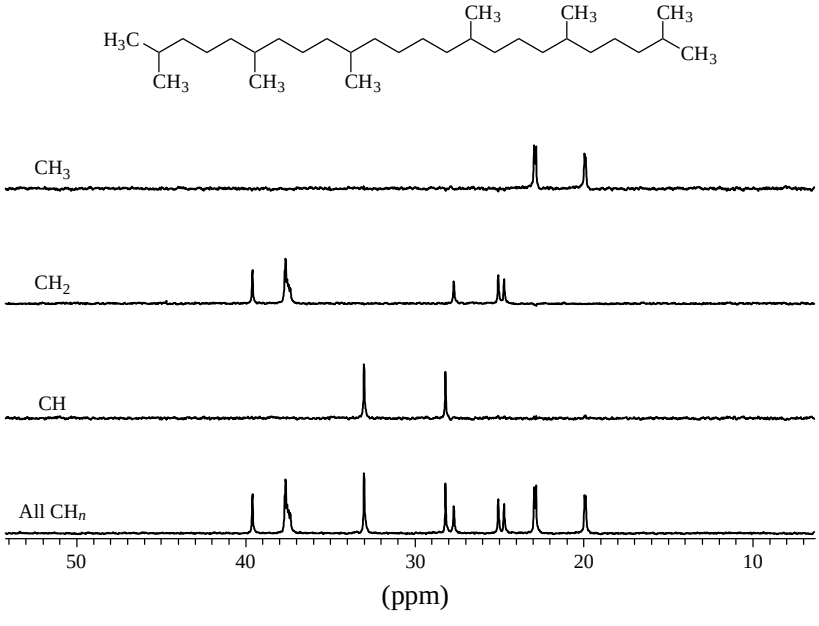

Figure 3 Conventional ^{13}C-NMR (bottom) spectrum of squalane (whose molecular structure is on the top). The GASPE subspectra are labeled according to the component resonance types as CH, CH_2, and CH_3.

compound for base oil, is divided into three subspectra with the C assignments reported in the figure.

The molecule of squalane has the following CH_n groups: $(CH_3)_8(CH_2)_{12}(CH)_6$. The respective CH_n fractions according to Equation (1):

$$F(CH_n) = \frac{I(CH_n)}{\sum_n I(CH_n)} \quad (1)$$

where $I(CH_n)$ is the total intensity of the NMR signals of the CH_n group, are: $F(CH) = 0.20$; $F(CH_2) = 0.53$; $F(CH_3) = 0.27$. The $F(CH_n)$ fractions calculated from GASPE are: $F(CH) = 0.21$; $F(CH_2) = 0.53$; $F(CH_3) = 0.26$. Comparing the data obtained by GASPE with the real values, the GASPE technique results are very accurate.

We have analyzed a few different types of base oils by GASPE/NMR, and in Table 1 they are summarized using letters to define their nature. In Figure 4 the GASPE/NMR spectra of base oil L, as a general example, are shown.

Another advantage of GASPE spectra is to have the methyl signals without the presence of CH_2 signals in the region between 5 and 30 ppm. The methyl signals can be assigned to specific branching sites along the hydrocarbon chain.[25,26] An example of the methyl region of the GASPE spectrum is shown in Figure 5. The main signals of methyl groups are assigned to the hydrocarbon branches shown in Table 2.

Table 1 List of the 17 analyzed lubricating base oils

Base oil	Class[a]
A	SR
B	SR
C	SR
D	SR
E	SR
F	HISO
G	HISO
H	HISO
I	HC + HISO
L	HC
M	HC
N	HC
O	HC
P	HC
Q	PAO
R	PIO
S	PAO

[a] SR = solvent-refined mineral oils; HISO = hydroisomerized; HC = hydrocracked; PAO = poly(α-olefin); PIO = poly(internal olefin).

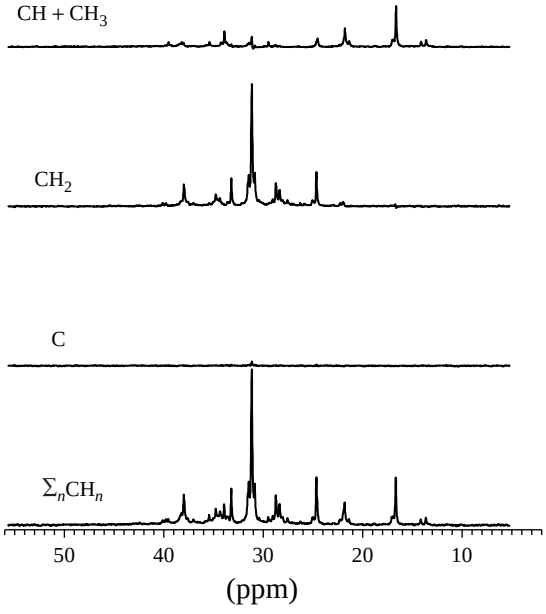

Figure 4 Conventional ^{13}C-NMR (bottom) spectrum of base oil F (aliphatic region). The GASPE subspectra are labeled according to the component resonance types as C, CH_2, and $CH + CH_3$.

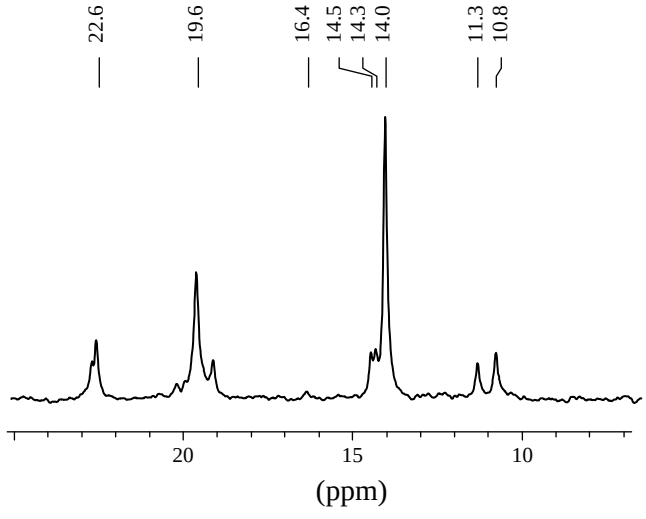

Figure 5 GASPE CH_3 ^{13}C-NMR subspectra of base oil F as a general example.

On the basis of GASPE data, some molecular features have been calculated relating to the number of branches (*NB*) and the number of aliphatic rings (*NAR*) for an average molecule. From GASPE spectra the molar fractions of the CH_n ($n = 0$–3) aliphatic group are obtained using Equation (1). The average number of branches is estimated by multiplying the molar fraction of methyls, $F(CH_3)$, by the average number of carbons (N_C) of the base oil, obtained by VPO, which gives the average

Table 2 Methyl peak assignment of the CH_3 GASPE/^{13}C-NMR subspectrum of hydrocarbon base oil

Type of branch	δ (ppm)	Identification no.
$>CHCH_2\mathbf{CH_3}$	10.8	1
$-(CH_3)CHCH_2\mathbf{CH_3}$	11.3	2
$>CHCH_2CH_2\mathbf{CH_3}$	14.5	3
$-CH(CH_3)CH_2CH_2\mathbf{CH_3}$	14.3	4
$-(CH_2)_nCH_2CH_2CH_2\mathbf{CH_3}$	14.0	5
$-CH_2CH(\mathbf{CH_3})-CH<$	17–15	6
$-CH_2CH(\mathbf{CH_3})-CH_2-$	19.6	7
$-CH_2-CH(\mathbf{CH_3})_2$	22.6	8

number of methyls per molecule, and then subtracting two from the resulting product (the two terminal methyls of the straight chain). The expression is as shown in Equation (2):

$$NB = n\mathrm{CH_3} - 2 = F(\mathrm{CH_3})N_\mathrm{C} - 2 \quad (2)$$

The number of aliphatic rings can be calculated according to Equation (3):

$$NAR = 0.5N_\mathrm{C}[2F(\mathrm{C}) + F(\mathrm{CH}) - F(\mathrm{CH_3})] + 1 \quad (3)$$

Other molecular characteristics are obtained directly from the NMR spectra: for instance, the molar fractions of aromatic hydrogens and carbons are calculated using Equation (4):

$$F(\mathrm{H_{ar}}) = \frac{I(\mathrm{H_{ar}})}{\sum_n I(\mathrm{H}_n)} \quad F(\mathrm{C_{ar}}) = \frac{I(\mathrm{C_{ar}})}{\sum_n I(\mathrm{C}_n)} \quad (4)$$

where the denominators are the sum of all spectral intensities of the relative nuclei. In Table 3 the molecular features of the 17 base oils of Table 1 are shown, with the acronyms used in the text.

In Table 4 the molar fractions of different methyls groups with the identification numbers reported in Table 2, are shown. The hydrocarbon-type analysis of each base oil was carried out with gas chromatography/mass spectrometry (ASTM D2549, ASTM D3239 and ASTM D2786). The volume percentages of isoparaffins, cycloparaffins and aromatic compounds are reported

Table 4 Molar percentages of methyl groups on different branches, numbered as in Table 2

Base oil	1	2	3	4	5	6	7	8
A	3.3	4.7	1.1	4.0	28.9	6.3	39.2	12.5
B	1.1	3.1	1.6	4.6	20.2	9.0	44.7	15.7
C	1.3	4.0	1.8	6.6	24.2	8.6	40.8	12.7
D	1.2	4.0	2.0	8.0	24.2	7.2	39.8	13.6
E	1.5	2.8	3.3	6.2	22.8	8.1	41.7	13.6
F	9.0	5.1	7.1	4.2	23.1	4.9	36.5	10.1
G	8.9	5.3	6.6	3.0	30.1	2.8	33.9	9.4
H	7.1	4.5	7.9	3.8	26.1	6.2	31.8	12.6
I	6.2	5.1	6.4	4.3	22.4	6.7	36.3	12.6
L	7.5	5.2	4.5	4.8	32.4	3.5	30.9	11.2
M	4.5	4.2	5.4	4.7	19.7	7.5	39.2	14.8
N	6.0	4.5	8.0	4.9	27.9	5.3	32.2	11.2
O	5.4	5.6	7.0	4.6	26.1	4.3	33.4	13.6
P	5.1	4.1	5.7	4.1	25.8	8.1	34.9	12.2
Q	1.7	0.8	6.6	0.0	75.7	11.9	3.3	0.0
R	3.3	3.5	2.8	1.8	62.2	8.8	17.6	0.0
S	2.6	0.0	8.1	0.0	72.5	13.3	3.5	0.0

Table 3 Molecular features of the analyzed base oils

Base oil	$F(\mathrm{H_{ar}})$	$F(\mathrm{C_{ar}})$	$F(\mathrm{C})$	$F(\mathrm{CH})$	$F(\mathrm{CH_2})$	$F(\mathrm{CH_3})$	N_C	NB	NAR
A	0.8	0.07	0.02	0.14	0.59	0.18	30	3.4	1.0
B	1.1	0.07	0.02	0.15	0.56	0.20	28	3.6	0.9
C	1.5	0.09	0.02	0.14	0.56	0.19	29	3.5	0.9
D	1.3	0.07	0.02	0.14	0.58	0.19	29	3.5	0.9
E	1.4	0.09	0.02	0.14	0.58	0.17	36	4.1	1.0
F	0.0	0.00	0.00	0.14	0.68	0.18	36	4.5	0.3
G	0.0	0.00	0.00	0.13	0.72	0.15	33	3.0	0.7
H	0.0	0.00	0.00	0.12	0.70	0.18	30	3.4	0.1
I	0.0	0.00	0.02	0.11	0.68	0.19	32	4.1	0.4
L	0.2	0.01	0.00	0.14	0.68	0.17	30	3.1	0.6
M	0.2	0.01	0.01	0.14	0.66	0.18	32	3.8	0.7
N	0.0	0.00	0.03	0.12	0.67	0.18	29	3.2	1.0
O	0.0	0.00	0.01	0.13	0.66	0.20	34	4.8	0.2
P	0.5	0.02	0.03	0.11	0.66	0.18	32	3.8	0.8
Q	0.0	0.00	0.00	0.12	0.73	0.15	33	3.0	0.5
R	0.0	0.00	0.02	0.11	0.70	0.17	35	4.0	0.7
S	0.0	0.00	0.02	0.08	0.75	0.15	44	4.6	0.3

Table 5 Volume percentage (vol. %) distribution of isoalkanes, cycloalkanes and aromatics

Base oil	Isoalkanes (vol. %)	Cycloalkanes (vol. %)	Aromatics (vol. %)
A	23.8	55.1	21.1
B	20.0	61.9	18.1
C	21.1	48.4	30.5
D	23.7	51.7	24.6
E	15.8	47.6	36.6
F	79.1	20.0	0.9
G	80.5	19.2	0.3
H	86.2	13.2	0.6
I	71.1	28.3	0.6
L	65.6	30.4	4.0
M	31.3	66.1	2.6
N	75.8	23.6	0.6
O	53.0	46.4	0.6
P	55.0	39.8	5.2
Q	97.0	3.0	0.0
R	82.9	17.1	0.0
S	97.0	3.0	0.0

Table 6 Physical and rheological properties of base oils

Base oil	Pour point (°C)	Kinematic viscosity at 40 °C (cSt)	Kinematic viscosity at 100 °C (cSt)	VI	Noack (% w/w)
A	−9	29.69	5.23	107	11.6
B	−9	31.61	5.32	100	14.1
C	−9	31.17	5.29	101	11.2
D	−12	29.45	5.22	108	12.0
E	−9	95.38	10.71	95	2.6
F	−21	23.57	5.04	146	12.4
G	−24	29.67	5.79	141	7.5
H	−21	21.07	4.69	146	15.1
I	−18	27.14	5.52	146	11.8
L	−27	17.69	4.02	127	14.8
M	−21	32.65	5.94	128	6.0
N	−27	18.49	4.14	129	10.7
O	−15	32.93	5.92	125	7.6
P	−9	34.67	6.15	126	4.5
Q	−60	29.67	5.90	147	4.8
R	−45	30.36	5.70	131	8.2
S	−63	45.55	7.68	137	4.0

in Table 5, and the physical and rheological properties of the same analyzed base oils are reported in Table 6.

From the analysis of Tables 1–5 it is evident that the molecular features are peculiar to different classes of base oils. For instance, the SR mineral base oils have a greater content of aromatics (vol.% and their H and C fractions), whereas the polyolefins have a greater fraction of methyls on long aliphatic chains (identification number 6 in Table 2). These chemical features certainly influence the physical properties of the base oils: for instance, the higher value of aromatics in SR mineral oils is the cause of their higher pour points, whereas the presence of long chains in the polyolefins should be the cause of their excellent pour points. Some correlation could be tried between performance and molecular feature with statistical models, but at the moment they suffer from the limitations of the low number of significantly different base oils with respect to the large quantity of molecular characteristics shown by these complex compounds. Certainly the possibility of having predictive models of the performance of finished oils based on the molecular characteristics of base oils could represent a very valid tool for lubricant manufacturer choice, especially for the formulation of lubricants with very high performance.

ACKNOWLEDGMENTS

The author is particularly grateful to ENI Group S.p.A and Agip Petroli S.p.A for financial support of his work on lubricants, to Dr Carlo Corno for helpful discussions and suggestions and to Mr Walter Stringo for his technical contribution.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
^{13}C-NMR	Carbon-13 Nuclear Magnetic Resonance
DEPT	Distortionless Enhancement by Polarization Transfer
DIN	Deutsches Institut fur Normung
GASPE	Gated Spin Echo
^{1}H-NMR	Proton Nuclear Magnetic Resonance
IP	Institute of Petroleum
ISO	International Organization for Standardization
MEK	Methyl Ethyl Ketone
NMR	Nuclear Magnetic Resonance
VI	Viscosity Index
VPO	Vapor Pressure Osmometry

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring

Industrial Hygiene **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Lubricant Base Oils: Analysis and Characterization of • Nuclear Magnetic Resonance Characterization of Petroleum • Petroleum Residues, Characterization of

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Nuclear Magnetic Resonance of Geological Materials and Glasses • Parameters, Calculation of Nuclear Magnetic Resonance • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. S. Korcek, R.K. Jensen, 'Relation between Base Oil Composition and Oxidation Stability at Increased Temperatures', *ASPLE Trans.*, **19**, 83–94 (1976).
2. C.J.A. Roelands, J.C. Vlugter, H.I. Waterman, 'The Viscosity-Temperature-Pressure Relationship of Lubricating Oil and its Correlation with Chemical Constitution', *ASPLE Trans.*, **85**, 601–610 (1963).
3. M.P. Smith, A.J. Stipanovic, G.P. Firmstone, W.M. Cates, T.C. Li, 'Comparison of Mineral and Synthetic Base Oils using Correlations for Bench and Engine Tests', *Lubr. Eng.*, **52**, 309–314 (1995).
4. A.J. Stipanovic, M.P. Smith, G.P. Firmstone, J.A. Patel, 'Compositional Analysis of Re-refined and Non-conventional Lubricant Base Oils: Correlations to Sequence VE and Sequence IIIE Gasoline Engine Tests', *SAE Pap.*, **941978**, 1365–1373 (1994).
5. J.A. Supp, R.E. Kornbrekke, S.H. Roby, 'Deposit Formation in Gasoline Engines-Part I: Base Oil Effects in Sequence VE Deposits', *Lubr. Eng.*, **50**, 964–969 (1994).
6. A.J. Stipanovic, J.A. Patel, D.L. Mc.Gregor, V.P. Nero, 'Compositional Analysis of Lubricating Base Oils and Re-refined Products: Correlation to Engine Test Performance', *Prepr.-Am. Chem. Soc., Div.Pet. Chem.*, 1377–1382 (1992).
7. H. Singh, S. Swarup, G.S. Chaudary, 'Physico-Chemical Characterization and Property Correlations in Base Oils and Fractions', *Lubr. Eng.*, **49**, 276–282 (1992).
8. V. Calemma, P. Iwanski, M. Nali, R. Scotti, L. Montanari, 'Structural Characterization of Asphaltenes of Different Origins', *Energy Fuels*, **9**, 225–230 (1995).
9. S.A. Sarpal, G.S. Kapur, A. Chopra, S.K. Jain, S.P. Srivastava, A.K. Bhatnagar, 'Structure–Performance Relationship of Viscosity-index Improvers: I. Microstructural Determination of Olefin Copolymers by NMR Spectroscopy', *Lub. Sci.*, **8**, 49–60 (1995).
10. F.P. Barron, M.R. Bendall, L.G. Armstrong, A.R. Atkins, 'Application of the DEPT Pulse Sequence for the Generation of $^{13}CH_n$ Subspectra of Coal-derived Oils', *Fuel*, **63**, 1276–1280 (1984).
11. M.U. Hasan, A. Bukhari, M.F. Ali, 'Structural Characterization of Saudi Arabian Heavy Crude Oil by NMR Spectroscopy', *Fuel*, **64**, 839–841 (1985).
12. S. Gillet, J.J. Delpuech, P. Valentin, J.C. Escalier, 'Optimum Conditions for Crude Oil and Petroleum Product Analysis of Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **52**, 813–817 (1980).
13. C.E. Snape, 'Comments on the Application of Spin-echo Carbon-13 Methods to Fossil Fuel-derived Materials', *Fuel*, **62**, 621–624 (1983).
14. S.P. Srivastava, A.K. Bhatnagar, G.C. Joshi, 'A Procedure for Quantitative Carbon-13 NMR Application to the Characterization of the Naphta Fraction of Petroleum Crude', *J. Chem. Technol. Biotechnol., Chem. Technol.*, **33A**, 361–366 (1983).
15. D.J. Cookson, B.E. Smith, 'Determination of Carbon C, CH, CH_2 and CH_3 Group Abundances on Liquids Derived from Petroleum and Coal Using Selected Multiplet ^{13}C NMR Spectroscopy', *Fuel*, **62**, 34–38 (1983).
16. D.J. Cookson, B.E. Smith, 'Determination of Structural Characteristics of Saturates from Diesel and Kerosene Fuels by Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **57**, 864–871 (1985).
17. D.J. Cookson, C.L. Rolls, B.E. Smith, 'Structural Characteristics of Branched Plus Cyclic Saturates from Petroleum and Coal Derived Diesel Fuels', *Fuel*, **68**, 788–792 (1989).
18. D. Doddrell, D. Pegg, M. Bendall, 'Distortionless Enhancement of NMR Signals by Polarization Transfer', *J. Magn. Reson.*, **48**, 323–327 (1982).
19. F.A. Dagga, H. Ruegger, 'Average Structural Parameters and Identification of Aromatic Components by 2D NMR Spectroscopy', *Fuel*, **67**, 1255–1262 (1988).

20. L. Montanari, E. Montani, C. Corno, S. Fattori, 'NMR Molecular Characterization of Lubricating Base Oils: Correlation with Their Performance', *Appl. Magn. Reson.*, **14**, 345–356 (1998).
21. R.K. Harris, *Nuclear Magnetic Resonance Spectroscopy*, Longman Scientific & Technical, Essex, UK, 1986.
22. J.K.M. Sanders, B.K. Hunter, *Modern NMR Spectroscopy*, Oxford University Press, New York, 1988.
23. A.E. Derome, *Modern NMR Techniques for Chemistry Research*, Pergamon Press, New York, 1988.
24. A.S. Sarpal, G.S. Kapur, A. Chopra, S.K. Jain, S.P. Srivastava, A.K. Bhatnagar, 'Hydrocarbon Characterization of Hydrocracked Base Stocks by One- and Two-dimensional NMR Spectroscopy', *Fuel*, **75**, 483–490 (1996).
25. L.P. Lindeman, J.Q. Adams, 'Carbon-13 Nuclear Magnetic Resonance Spectrometry-chemical Shifts for the Paraffins through C_9', *Anal. Chem.*, **10**, 1245–1251 (1971).
26. A. Priola, C. Corno, S. Cesca, 'On the Cationic Polymerization of Olefins and the Structure of the Product Polymers. 4. 1-Butene Selectively ^{13}C Enriched in positions 3 and 4', *Macromolecules*, **13**, 1110–1113 (1980).

Lubricant Base Oils: Analysis and Characterization of

Arnold L. Shugarman
76 Lubricants Company, Santa Ana, CA, USA

1 Introduction 1
2 Base Oil Refining Processes 2
3 Base Oil Composition 2
3.1 Hydrocarbon Type Analysis 2
3.2 Elemental Analysis 3
4 Physical and Chemical Properties 3
4.1 Viscosity 3
4.2 Wax-related Properties 4
4.3 Volatility 4
4.4 Other Properties 4
5 Effect of Base oil Composition on Lubricant Performance 5
5.1 Oxidation Tests 5
5.2 Correlation of Base oil Composition and Lubricant Performance 5
6 Toxicology 6
7 Environmental Impact of Base oils 6
Acknowledgments 6
Abbreviations and Acronyms 6
Related Articles 7
References 7

Lubricant base oils, or base stocks, are hydrocarbon lubricant components which are combined with performance additives to make finished industrial and automotive lubricants. A lubricant is a substance that reduces friction and wear between surfaces. Base oils typically comprise 80–99% of the finished lubricant. They are refined from selected crude oils, re-refined from used oil, synthesized from specific organic compounds, or extracted from renewable resources, such as seed oils. Base oils formulated into finished lubricants help to reduce friction and wear between moving surfaces, remove heat, seal, cushion against shock loads, and remove abrasives and contaminants from the lubricated area. Base oil analysis and characterization is important for understanding how the physical and chemical properties of these materials relate to performance in finished lubricants. It is also essential for maintaining quality control (QC) over the manufacturing process, for quality assurance (QA) of the finished product, and for setting specifications between buyer and seller.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Base oils are refined from crude oil, re-refined from used oil, or manufactured to give products with specific chemical and physical properties. Mineral oils are base oils produced from the atmospheric residuum of crude oil. The residua are processed to remove aromatics and nitrogen-, sulfur-, and oxygen-containing compounds which degrade the performance of the finished base oil. Mineral base oils are classified by the molecular type that predominates (paraffinic, naphthenic, or aromatic) and their physical and chemical properties.

Paraffinic base oils tend to have high pour points and must be dewaxed to produce a base oil with acceptable flow properties at low temperatures. Naphthenic base oils have a naturally low wax content and generally have better solvency properties than paraffinic stocks. However, they tend to oxidize more readily than paraffinic oils. Aromatic stocks are valued for their solvency properties rather than their lubricating quality, and are used as process oils in the rubber and printing ink industries. Typical properties of paraffinic, naphthenic, and aromatic oils are shown in Table 1.[1]

High-quality re-refined base oils are processed to be indistinguishable from virgin mineral oils derived from crude oil. They are derived from contaminated used oils from a variety of sources. Analyses of re-refined base oils may include tests for contaminants in used oil that are not found in virgin base oils. (Refer to the appropriate American Society for Testing and Materials (ASTM) standard.[2] Unless stated otherwise, all references to ASTM methods are from the *Annual Book of ASTM Standards*, Vol. 05.)

The most commonly used synthetic base oils fall into one of five classes: olefin oligomers, especially polyalphaolefins (PAOs),[3,4] esters (dibasic and polyol), alkylated aromatics, polyalkylene glycols, and metalloorganic oligomers (phosphate or silicate esters, siloxanes).[5] Chlorine and/or fluorine-substituted hydrocarbons or polyethers are specialty lubricants used where extreme operating conditions are encountered.

Several general references on base oil technology, applications, and characterization have been published and are widely available. Neale has edited a tribology handbook with an excellent review of base oil types and their properties.[6]

Klaus and Tewksbury provide a comprehensive review of base oil and finished lubricant properties.[7] The ASTM has published a monograph on testing petroleum products following ASTM methods[8] and an educational guide on base oil characterization tests which covers lubricant performance and safe handling, composition, performance, and consistency.[2] The

Table 1 Typical Base Oil Properties

Test	ASTM	Paraffinic	Naphthenic	Aromatic
Viscosity, cSt (40 °C)	D 445	40	40	36
Viscosity SUS (100 °F)	D 2161	205	208	192
Viscosity, cSt (100 °C)	D 445	6.2	5.0	4.0
Viscosity, SUS (210 °F)	D 2161	46.9	43.0	40.0
VI	D 2270	100	0	−185
Specific Gravity	D 287	0.8628	0.9194	0.9826
API Gravity	D 287	32.5	22.5	12.5
Viscosity–Gravity const	D 2501	0.807	0.875	0.957
Flash point, COC, °C	D 92	229	174	160
Aniline point, °C	D 611	107	73	17
Pour point, °C	D 97	−15	−30	−24
Color	D 1500	L 0.5	2.0	D 8.0
Molecular weight	D 2503	440	330	246
Refractive index	D 1747	1.4755	1.5068	1.5503
Clay-gel analysis	D 2007			
% Polars		0.2	3	8
% Aromatic		8.5	43	80
% Saturates		91.3	54	12
Carbon atom type	D 3238			
% C_A		2	19	41
% C_N		32	37	36
% C_P		66	44	23

Reprinted by permission of Society of Tribologists and Lubrication Engineers from D.V. Brock, *Lubricant Base Oils, Starting from Scratch – Tribology Basics*, Volume 1, p 16.

journal *Analytical Chemistry* issues a comprehensive, biennial review on the analysis of petroleum and coal which includes lubricants and base oils.[9] For a broader perspective on lubricants, applications, lubricant characterization, and performance, Caines and Haycock[10] have written a reference book on automotive lubricants, and the Association of Iron and Steel Engineers (AISE)[11] has published a manual on industrial lubricants.

2 BASE OIL REFINING PROCESSES

Base oils are traditionally made by solvent refining to physically extract undesirable aromatics and heteroatomic compounds from the feed, solvent-assisted wax crystallization (paraffinic crudes), and clay treatment to decolorize and stabilize the finished product. In more modern processes, some refiners have replaced solvent refining with hydrocracking or severe hydrotreating, solvent dewaxing with catalytic dewaxing or wax isomerization,[12] and clay treating with hydrofinishing. Mineral base oils with properties approaching those of PAOs have been manufactured by isomerizing wax derived from slack wax or Fischer–Tropsch synthesis.[13,14] Base oils produced from vacuum distillates are called neutral oils; bright stocks are base oils refined from vacuum.

3 BASE OIL COMPOSITION

Base oil characterization is important for understanding the relationship between composition and performance of finished products in end-use applications. Rig or engine tests, developed to predict field performance, are often expensive or take a long time to complete. Consequently, a need exists for relatively simple, low-cost tests to predict the performance of a finished lubricant in the customer's equipment.

3.1 Hydrocarbon Type Analysis

Base oils derived from petroleum are complex mixtures of thousands of individual compounds. To better understand how base oil molecular structure, composition, and properties relate to product performance, investigators have examined individual model compounds and characterized the bulk molecular hydrocarbon types in base oils. Over a 26-year period, American Petroleum Institute (API) sponsored a program to measure the properties of 321 high-molecular-weight hydrocarbons.[15] Briant has provided a thorough analysis of these and other data on model compounds.[16]

ASTM has published several test methods to obtain information on the composition of base oils. ASTM D 2140[17] and D 3238 give calculated values for the carbon number distribution among aromatic, naphthenic, and paraffinic compounds based on viscosity (ASTM D 445),

density/specific gravity (ASTM D 1481), and refractive index (ASTM D 1218). ASTM D 3238, also known as the refractive index–density–molecular weight (n–d–M) method, gives similar, though not identical, results to ASTM D 2140.

More detailed base oil compositional information can be obtained by first separating base oil into aromatic and non-aromatic fractions on a chromatographic column and analyzing each fraction separately by mass spectrometry (MS). ASTM D 2549 describes a chromatographic procedure for separating hydrocarbon mixtures boiling between 232 °C and 538 °C on a packed column. The non-aromatics fraction is subjected to analysis by high ionizing voltage MS (ASTM D 2786) to give composition by seven saturated hydrocarbon types and one aromatic type. The aromatics fraction is analyzed by high ionizing voltage, low-resolution MS.

ASTM D 2007 (Clay-gel Analysis) and a technique developed by Hirsch et al.[18] are column chromatographic procedures used to separate high-boiling petroleum fractions into saturated, aromatic, and polar compounds. Barman[19] showed that ASTM D 2007 is subject to cross-contamination of these hydrocarbon classes. Thin-layer chromatography (TLC) with flame ionization detection (FID) is more rapid and accurate. Mohindroo and Preston[20] used supercritical-fluid chromatography (SFC) with flame ionization and ultraviolet (UV) detectors to separate base oils into saturates, polars, monoaromatics, and polynuclear aromatics. Hsu et al.[21] developed a rapid, preparative scale, high-pressure liquid chromatography (LC) procedure to separate base oils into compound groups using two columns and solvents of different polarity.

Thermal diffusion allows separation of hydrocarbons by molecular size under the influence of a thermal gradient. Cyclic and linear molecules diffuse by convection towards opposite ends of the apparatus. The process is time-consuming, but it provides detailed information not available by bulk chromatographic separation.[22]

Chromatography separates compounds into distinct molecular types but provides no quantitative information on the number of aromatic and paraffinic carbon or hydrogen atoms in each fraction. Nuclear magnetic resonance (NMR) and infrared (IR) analyses provide this information directly. ASTM D 5292 describes an NMR method to measure the mole or mass percent of aromatic carbon or aromatic hydrogen atoms in a sample. A. van de Ven et al.,[23] developed an improved correlation between ^{13}C NMR and IR analysis using the aromatic bands at 1610 cm^{-1} and 815 cm^{-1} to quantify the mole percent aromatic carbon in a base oil.

UV absorption (ASTM D 2008) is used to detect total aromatics in a sample, especially in severely hydroprocessed base oils, waxes, and white oils where the aromatics content is low. In ASTM D 2269, aromatics are concentrated before UV analysis by dimethylsulfoxide (DMSO) extraction. Similar procedures have been developed to assess the potential carcinogenicity of base oils and aromatic extract distillates.[24,25]

3.2 Elemental Analysis

Compounds containing nitrogen, oxygen, and sulfur in base oils can have profound effects on the properties and performance of finished lubricants. Elemental analyses are conducted to detect base oil contamination or impurities which survive the process of re-refining used oils. ASTM D 6074 provides a comprehensive list of tests to characterize hydrocarbon base oils (mineral oils) including those for elemental analysis. Sieber and Salmon[26] give an excellent summary of the instrumentation and methodology for spectrometric methods of lube oil analysis.

4 PHYSICAL AND CHEMICAL PROPERTIES

Physical and chemical properties of base oils can be related, ultimately, to their bulk or molecular properties. Physical property tests are used for manufacturing process control, QC, QA, consistency of finished products, and to set specifications between buyer and seller.

4.1 Viscosity

Viscosity is the critical factor in establishing hydrodynamic or full-film lubrication which reduces friction and wear between moving parts. Viscosity is defined as the ratio of shear stress to shear rate and measures resistance to flow. Newtonian fluids, including base oils above the cloud point, obey this equation. Klaus and Tewksbury,[7] Briant et al.,[27] and Alexander[28] provide detailed descriptions of rheology theory and applications to lubricants.

Dynamic, or absolute, viscosities are most often measured in coaxial-cylinder viscometers.[29] Kinematic viscosities are determined by ASTM D 445 which measures the time for a liquid to move through a narrow capillary segment. Absolute viscosity in millipascal sec. or centipoise (cP) equals kinematic viscosity in $mm^2\,s^{-1}$ or centistokes (cSt) times density ($g\,cm^{-3}$). Viscosities are typically reported at two temperatures, 40 °C and 100 °C, and a property defined as the viscosity index (VI) is calculated from the values (ASTM D 2270). High VI oils exhibit a relatively small change in viscosity with temperature. Aliphatic paraffins predominate in high VI oils (Table 1). ASTM D 341 describes a procedure and conditions to calculate the viscosity at one temperature from viscosities measured at two other temperatures.

Table 2 Low temperature – low shear rate tests

Test	ASTM D	Cooling Rate, °C	Shear Rate, s^{-1}
ASTM cloud point	2500	0.6/min	0
ASTM pour point	97/5949/5950 5985	0.6/min	0.1–0.2
Stable pour point	([a])	0 to −40/7 days	0.1–0.2
Mini rotary viscometer	3829	2/h	<100
Mini rotary viscometer	4684	0.33–2.5	<100
Brookfield viscosity	2983	Shock	0.1–12
Scanning Brookfield	5133	1/h	1.7

[a] Federal Stable Pour Test (FED-STD-791, Method 203), Naval Publications and Forms Center, Philadelphia, PA 19120, USA; SAE J-300, Appendix B, Society of Automotive Engineers, Warrendale, PA 15096, USA.

Viscosity also changes with respect to pressure, but the effect is much smaller than the temperature effect. Viscosity–pressure relationships are important in highly loaded contacts in gears or rolling element bearings, where the pressures under elastohydrodynamic lubrication can exceed 345 MPa (50 000 psi). Viscosity as a function of pressure can be measured in falling-weight viscometers, capillary viscometers, vibrating-crystal viscometers, and optical viscometers.[30] The effect of pressure on viscosity decreases in the order naphthenics > paraffinics > esters.

Apparent viscosities of non-Newtonian fluids vary with shear rate and shear stress. These include base oils below the cloud point and fully formulated lubricants containing high-molecular-weight components (VI improvers). For engine oils, high-temperature–high-shear viscosity at 150 °C and $10^6\ s^{-1}$ shear rate represents lubrication conditions in high-speed bearings. Methods to measure viscosity under these conditions include the Tapered-bearing Simulator (ASTM D 4683), Tapered-plug Viscometer (ASTM D 4741), and Multicell Capillary Viscometer (ASTM D 5481).

The Cold-cranking Simulator (CCS) (ASTM D 5293) measures low-temperature base oil viscosities at 10^4–$10^5\ s^{-1}$ shear rates. Wax-related viscosity effects are minimized, and measured viscosities are similar to extrapolated viscosities from data obtained above the cloud point. CCS viscosities of fully formulated engine oils correlate with low-temperature engine startup.

4.2 Wax-related Properties

As paraffinic base oils are cooled, wax crystals grow and form interlocking networks throughout the fluid which eventually stop oil flow at low shear rates. The shear stress required to break down the wax structure and resume lubricant flow is called the yield stress. Several tests have been developed to measure low-temperature, engine oil viscosities at low shear and low shear stress after long cooling cycles. Test conditions mimic, but don't necessarily duplicate, conditions that can cause low-temperature engine oil pumpability failures in the field. Except for the cloud- and pour-point tests, methods listed in Table 2 were developed specifically for fully formulated engine oils.

4.3 Volatility

Base oil volatility is an important factor in crankcase oil consumption, especially at high engine temperatures. Volatility decreases as VI increases for a given viscosity base oil and distillation cut width.

Volatility is measured by an evaporative-loss test (Noack) or gas chromatography (GC) (simulated distillation). In the Noack test, the weight of oil evaporated at 250 °C in one hour in flowing air is measured. Variations of the Noack test exist as Coordinating European Council (CEC) L-40-T-87,[31] Japanese Petroleum Institute (JPI)-5S-41-93, Method B,[32] and ASTM D 5800. An alternative apparatus has been proposed which eliminates use of toxic Woods metal as the heat-transfer medium and allows nearly quantitative collection of the volatiles for further analysis.[33] GC simulates distillation, and is more rapid and precise than Noack methods. ASTM D 2887 measures boiling-point distribution for hydrocarbons boiling between 55 °C and 538 °C. ASTM D 5480 uses an internal standard and was developed specifically to measure volatility of engine oils. Noack volatility is typically 5–15% higher than ASTM D 2887. Choi and Deane,[34] showed that two base oils with the same Noack volatility could have different volatilities by ASTM D 2887, and vice versa. Engine oil volatility at high engine operating temperatures appears to correlate better with Noack volatility.

4.4 Other Properties

ASTM D 6074 summarizes the tests used most often to characterize base oils and finished oils. ASTM D 6158 lists the specific tests and specifications for mineral hydraulic oils, including noncompounded base oils. Klaus and Tewksbury[7] and Godfrey and Herguth[35] describe base oil properties from a more fundamental viewpoint and

provide typical properties for different chemical classes of lubricants. The Annual Book of ASTM Standards, Volume 05, describes over 150 test procedures useful for testing or characterizing base oils.

Base oil characterization involves testing and assessment of appearance, viscometrics, safety, interfacial properties, solvency, oxidation and thermal stability, seal compatibility, and volatility. Appearance is determined by color (ASTM D 1500), cleanliness (International Organization for Standardization (ISO) 4406), and clarity (ASTM D 4860). Water above 80–100 ppm, for example, will impart a haze to base oils at room temperature. Color stability can be estimated by light irradiation, either by UV source or sunlight. Mills and Melchior[36] showed that color instability is due to nitrogen and sulfur heterocyclic compounds in base oils. Safety aspects include flash point (ASTM D 92/93) and toxicity (sections 6 and 7). Flash point is a rough measure of volatility and is used to detect low-boiling contaminants in base oils.

Interfacial properties between water and lube oil determine the demulsibility or the speed and efficiency of separating a mixture of the two (ASTM D 2711). Gas release (ASTM D 3827) and foaming tendency (ASTM D 892/6082) relate to the interfacial properties between base oils and gases. Copper corrosion (ASTM D 130) indicates presence of reactive sulfur compounds in base oils.

Solvency for additives and seal compatibility are governed by the chemical composition of base oils. Highly paraffinic base oils (hydrocracked mineral base oils and PAO) tend to shrink rubber seal materials and have lower solubility for additives than mineral oils containing aromatics and naphthenes, or synthetic base oils containing esters. Aniline point (ASTM D 611) is a useful guide to mineral base oil aromaticity and solvency properties. The effect of base oils and formulated lubricants on seal materials is determined by methods ASTM D 412/471/2240[37] and 5662.

5 EFFECT OF BASE OIL COMPOSITION ON LUBRICANT PERFORMANCE

Bench and rig test studies of base oils and formulated lubricants are used as guides to performance of fully formulated oils in field applications. The goal is to reduce the testing cost and maintain a high confidence level of performance when one base oil is substituted for another in a fully formulated lubricant. Hsu[38] conducted a thorough review of bench tests used to correlate performance of crankcase oils.

5.1 Oxidation Tests

Base oil oxidation tests can be classified as oxygen diffusion-limited (bulk), or reaction rate-controlled (thin film). Tests are run with or without catalyst activator. Catalysts simulate surfaces or contaminants that might accelerate oxidation. Oxidation-test results on uninhibited oils can be misleading. Highly paraffinic oils with low sulfur and nitrogen content oxidize more readily than base oils containing sulfur and aromatics, but are more stable with added oxidation inhibitor.[39]

Bulk oil oxidation tests include the oxidation test for turbine oils (The Institute of Petroleum (IP) 114), turbine oil stability test (ASTM D 943/4310/IP 157), rotary bomb oxidation test (ASTM D 2272), corrosiveness and oxidation stability of highly refined oils (ASTM D 4636/FTM 791-5308), and the universal oxidation/thermal stability test (ASTM D 4871/5846). Except for the rotary bomb test, which is conducted in a pressurized vessel, oxidation stability is evaluated at atmospheric pressure in flowing air.

Hsu and Cummings[40] developed a thin-film, thermogravimetric analysis (TGA) test to determine oxidation, volatility and extent of polymerization of base oils. Yoshida, Stipanovic et al.[41] used high-pressure differential scanning calorimetry (DSC) to measure base oil oxidative stability and correlated results with ASTM D 943. Klaus et al.[42] developed the Penn State Micro-oxidation Test to evaluate oxidative and thermal stability of base oils, and correlate lubricant degradation processes with those observed in automotive and diesel engine tests. Lubricant is oxidized in a thin film. Residues are analyzed by gel permeation chromatography (GPC). Selby[43] recommended the thin-film oxygen uptake test (TFOUT), ASTM D 4742, to determine base oil consistency and quality.

5.2 Correlation of Base oil Composition and Lubricant Performance

As a general rule, base oil oxidation stability increases with increasing saturates content, and sulfur compounds are beneficial at high temperatures in uninhibited oils or when inhibitors have been depleted.[44] Murray et al.[45] concluded that the relationship between VI and oxidation performance of solvent-refined base oils held only for base oils from the same crude source and refining process. Robson[46] found that oxidation of lubricants formulated with hydrocracked base oils correlated inversely with aromatics content. Sulfur did not correlate at the low levels found in the hydroprocessed lubes. Stipanovic et al.[41] used base oil compositional analysis, physical property measurements, and engine test results to develop predictive models based on partial least squares (PLS) and neural networks analyses. Multi-ring naphthene and polyaromatic compounds, thioaromatics, basic nitrogen compounds, and ''resins'' increase oxidation, thermal decomposition, and deposit formation. Supp et al.[47] also found that base oil sulfur increased sludge formation in

the Sequence VE test, but noted that aromatics, especially highly polar base oil components, were beneficial, possibly because they solubilize sludge precursors.

6 TOXICOLOGY

Crude oil atmospheric residua contain polynuclear aromatic, nitrogen-, and sulfur-heterocyclic compounds, some of which may be carcinogenic. While base oil refining removes some or all of these materials, the manufacturer is still responsible for assuring that the final products are non-carcinogenic. Blackburn[48] has reviewed the subject, and background information on test procedures and legal definitions of carcinogenicity are discussed in ASTM D 6074.

Mouse-skin bioassay is the most recognized, direct method for predicting the carcinogenicity of base oils.[49] The time of tumor development and the percentage of animals developing tumors relative to the controls are used to assess carcinogenicity. A bioassay test and DMSO extraction tests have been developed to predict the results of mouse-painting tests to reduce the time and cost to assess carcinogenicity. The Mutagenicity Index (Modified Ames Test), ASTM E1687,[50,51] is a short-term, microbiological, salmonella mutagenesis assay developed to detect mutagens in base oils. The correlation of mutagenicity and carcinogenicity with PNA content is defined for oils with viscosities >18 cSt at 40 °C, distilling between 250–550 °C. A new test to predict base oil carcinogenicity, DNA adduct formation, is now available.[52]

Method IP 346 specifies DMSO extraction of polycyclic aromatics and condensed-ring polar compounds in virgin base oil and measurement of the polycyclic aromatics in the extract.[53] The scope and limitations are described in the CONCAWE review[24] and the European Union Dangerous Substances Directive.[54] Haas et al.[25] developed another DMSO extraction procedure which is similar to ASTM D 2269. The index they developed includes both PNA content and viscosity to predict results from mouse-painting studies.

7 ENVIRONMENTAL IMPACT OF BASE OILS

Environmental impact is assessed by estimating the ecotoxicity and biodegradation of lubricants. Biodegradation is the oxidative decomposition of carbon, hydrogen, and oxygen compounds by microorganisms. Ecotoxicity is the tendency of a material to negatively impact nonhuman organisms.

The general order of biodegradation of base oils is vegetable oils > synthetic esters > mineral oils ~ alkylbenzenes > polyalkyleneglycols > PAOs.[5] The relative order depends on the test conditions and method of reporting. Within a family group, lower-molecular-weight compounds degrade faster and more completely than higher-molecular-weight materials.[4]

Biodegradation processes follow two paths. ''Ultimate'' biodegradation leads to complete biochemical oxidation to carbon dioxide, water, and inorganic salts. ''Bioaccumulation'' or ''primary biodegradation'' is a process in which the original molecule is partially degraded and incorporated into a microorganism as biomass. ''Primary biodegradation'' is followed by monitoring the disappearance of the original substance, while ultimate biodegradation is measured by carbon dioxide production or oxygen consumption.

ASTM D 6046 is a standard classification of hydraulic fluids for their environmental impact. While the document pertains specifically to the biodegradability and acute aquatic toxicity of hydraulic fluids, the methodology and test procedures identified also apply to base oils. ASTM published a companion document, ASTM D 6081, which covers standard practices for aquatic toxicity testing of lubricants and their components. Voltz et al.[55] and Kiovsky et al.[56] reviewed lubricant biodegradation test methods, and provided comparative test data on base oils and fully formulated oils.

ACKNOWLEDGMENTS

The author appreciates the assistance of Dr Dennis Brinkman, Safety Kleen Corporation, Mr Barry Deane, Exxon Research and Development, Mr Tom Kiovsky, BP Oil, Ms Margaret Lemmon, Lubrizol Corporation, and Dr Arthur Stipanovic, Texaco, Inc., with helpful discussions and reference materials.

ABBREVIATIONS AND ACRONYMS

AISE	Association of Iron and Steel Engineers
API	American Petroleum Institute
ASTM	American Society for Testing and Materials
CCS	Cold-cranking Simulator
CEC	Coordinating European Council
DMSO	Dimethylsulfoxide
DSC	Differential Scanning Calorimetry
FID	Flame Ionization Detection
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
IP	The Institute of Petroleum
IR	Infrared

ISO	International Organization for Standardization
JPI	Japanese Petroleum Institute
LC	Liquid Chromatography
MS	Mass Spectrometry
n–d–M	Refractive Index–Density–Molecular Weight
NMR	Nuclear Magnetic Resonance
PAO	Polyalphaolefin
PLS	Partial Least Squares
QA	Quality Assurance
QC	Quality Control
SFC	Supercritical-fluid Chromatography
TFOUT	Thin-film Oxygen Uptake Test
TGA	Thermogravimetric Analysis
TLC	Thin-layer Chromatography
UV	Ultraviolet
VI	Viscosity Index

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges)

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Full Range Crudes, Analytical Methodology of • High-temperature Simulated Distillation Applications in Petroleum Characterization • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Lube Products. Molecular Characterization of Base Oils • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels • Nuclear Magnetic Resonance Characterization of Petroleum • Petroleum Residues, Characterization of • Use of Inspection Properties to Predict Hydrocarbon Fraction Physical Properties

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Infrared Spectroscopy **(Volume 12)**
Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Thermal Analysis **(Volume 15)**
Differential Scanning Calorimetry and Differential Thermal Analysis

REFERENCES

1. D.V. Brock, *Lubricant Base Oils, Starting from Scratch – Tribology Basics*, Society of Tribologists and Lubrication Engineers, Park Ridge, IL, USA, 16, Vol. 1, 1997.
2. ASTM D-6074-97, *Standard Guide for Characterizing Hydrocarbon Lubricant Base Oils*. Annual Book of ASTM Standards, American Society for Testing and Materials, Conshohocken, PA, Vol. 05.03, 1998.
3. R.L. Shubkin, M.E. Kerkemeyer, 'Tailor-making Polyalphaolefins', *J. Synth. Lubrication*, **8**(2), 115–134 (1991).
4. R. Benda et al., 'Polyalphaolefins – Base Fluids for High Performance Lubricants', *J. Synth. Lubrication*, **13**(1), 41–57 (1996).
5. D.H. Culpon, T.C. Mead, 'Synthetic Lubricants', *Lubrication*, **78**(4), 1–16 (1992).
6. M.J. Neale (ed.), *Lubrication – A Tribology Handbook*, Society of Automotive Engineers, Warrendale, PA, 1–13, 1993.
7. E.E. Klaus, E.J. Tewksbury, 'Liquid Lubricants', *Handbook of Lubrication, Theory and Practice of Tribology*, ed. E.R. Booser, CRC Press, Boca Raton, FL, 229–254, Vol. 11, 1984.
8. G.V. Dyroff (ed.), *Manual on Significance of Tests for Petroleum Products*, 6th edition, ASTM Manual Series MNL 1, American Society for Testing and Materials, Conshohocken, PA, 1993.
9. C.T. Mansfield et al., 'Petroleum and Coal', *Anal. Chem.*, **69**(12), 59R–93R (1997).
10. A. Caines, R. Haycock, *Automotive Lubricants Reference Book*, Society of Automotive Engineers, Warrendale, PA, 1996.
11. *The Lubrication Engineers Manual*, 2nd edition, Association of Iron and Steel Engineers, Pittsburgh, PA, 1996.
12. M.W. Wilson et al., 'Isodewaxing Increases VI Capabilities for Lubricant Base Stock Manufacture', in *American Chemical Society Symposium on The Processing, Characterization and Application of Base Oils: Part II*, San Diego, California, 1994.
13. T.R. Lynch et al., 'Basestocks for the Next Millennium', in *American Chemical Society Symposium on Worldwide Perspectives on the Manufacture, Characterization and Application of Lubricant Base Oils*, San Francisco, California, 1997.
14. A. Sequeira, 'Lubricant Base Oil Processing', *Lubrication*, **75**(1), 4 (1989).
15. *Properties of Hydrocarbons of High Molecular Weight*, American Petroleum Institute Research Project 42, American Petroleum Institute, Washington, DC, 1940–1966.

16. J. Briant et al., *Rheological Properties of Lubricants*, Institut Francais du Petrole, Editions Technip, Paris, 123–163, 1989.
17. ASTM D 2140, *Carbon Type Composition of Insulating Oils of Petroleum Origin*, Annual Book of ASTM Standards, Conshohocken, PA, Vol. 10.03, 1998.
18. D.E. Hirsch et al., 'Separation of High-boiling Petroleum Distillates Using Gradient Elution Through Dual-packed (Silica Gel–Alumina Gel) Adsorption Columns', *Anal. Chem.*, **44**(6), 915–919 (1972).
19. B.N. Barman, 'Hydrocarbon-type Analysis of Base Oils and Other Heavy Distillates by Thin-layer Chromatography with Flame Ionization Detection and by the Clay-gel Method', *J. Chromatog. Sci.*, **34**, 219–225 (1996).
20. J.P. Mohindroo, H.G. Preston, 'Hydrocarbon Type Analysis of Base Oils by Supercritical Fluid Chromatography', in *American Chemical Society Symposium on The Processing, Characterization and Application of Base Oils: Part II*, San Diego, California, 1994.
21. S.M. Hsu et al., 'Friction and Wear Characteristics of Molecular Compound Classes from Lubricating Base Oils', Part 1, National Bureau of Standards, Washington, DC, NBSIR 84-2821, 12–28, 1984.
22. J. Briant et al., *Rheological Properties of Lubricants*, Institut Francais du Petrole, Editions Technip, Paris, 134–135, 1989.
23. A. van de Ven et al., 'An Infrared Analysis Method to Determine the Aromatic Carbon Content of Mineral Lubricant Base Oils Based on a Correlation with Carbon-13 NMR Aromaticities', *Lubrication Science*, **7**(3), 197–209 (1995).
24. CONCAWE Report No. 94/51, 'The Use of the Dimethyl Sulfoxide (DMSO) Extract by the IP 346 Method as an Indicator of the Carcinogenicity of Lubricant Base Oils and Distillate Aromatic Extracts', 1994.
25. J.M. Haas et al., 'A Simple Analytical Test and a Formula to Predict the Potential for Dermal Carcinogenicity from Petroleum Oils', *Amer. Indust. Hygiene Assoc. J.*, **48**(11), 935–940 (1987).
26. J.R. Sieber, S.G. Salmon, 'Elemental Analysis of Lubricating Oils and Greases', *Lubrication*, **80**(1), 1–16 (1994).
27. J. Briant et al., *Rheological Properties of Lubricants*, Institut Francais du Petrole, Editions Technip, Paris, 1989.
28. D. Alexander, 'The Viscosity of Lubricants', *Lubrication*, **78**(3), 1–16 (1992).
29. J. Briant et al., 'Rheological Properties of Lubricants', Institut Francais du Petrole, Editions Technip, Paris, 86–87, 1989.
30. E.E. Klaus, E.J. Tewksbury, 'Liquid Lubricants', *Handbook of Lubrication, Theory and Practice of Tribology*, ed. E.R. Booser, CRC Press, Boca Raton, FL, 234–235, Vol. 11, 1984.
31. Coordinating European Council, 'Evaporation Loss of Lubricating Oils', CEC L-40-T-87, Commission of the European Communities, Bruxelles, 1987.
32. 'Determination of Evaporative Loss of Engine Oils (Unified NOACK)', Japanese Petroleum Institute, Keidanren Kaikan, Tokyo, Japan.
33. B.J. Cluff et al., SAE 961227, Society of Automotive Engineers, Warrendale, PA, 1996.
34. E. Choi, B. Deane, 'Basestock Volatility and Oil Consumption', AM-98-47, NPRA Annual Meeting, March 15–17, San Francisco, CA, 1998, National Petroleum Refiners Association, Washington, DC, 1998.
35. D. Godfrey, W.R. Herguth, 'Back to Basics, Physical and Chemical Properties of Industrial Mineral Oils Affecting Lubrication – Parts 1–5', STLE, Park Ridge, IL, 1994.
36. I.W. Mills, J.J. Melchiore, 'Effects of Aromatics and Selected Additives on Oxidation Stability of Transformer Oils', *J. Ind. Eng. Chem. Prod. Res. Develop.*, **6**, 41 (1967).
37. ASTM D 412/471/2240, Annual Book of ASTM Standards, ASTM Conshohocken, PA, Vol. 9.01, 1998.
38. S.M. Hsu, 'Review of Laboratory Bench Tests in Assessing the Performance of Automotive Crankcase Oils', *Lubrication Engineering*, **37**(12), 722–731 (1981).
39. S.M. Hsu, 'Characterization of Lubricating Basestocks for Automotive Crankcase Oils, NBS SP 584, National Bureau of Standards, Washington, DC, 1980.
40. S.M. Hsu, A.L. Cummings, 'Thermogravimetric Analysis of Lubricants', SAE 831682, Society of Automotive Engineers, Warrendale, PA, 1983.
41. T. Yoshida, A.J. Stipanovic et al., 'The Impact of Basic Nitrogen Compounds on the Oxidative and Thermal Stability of Base Oils in Automotive and Industrial Applications'. SAE 981405, Society of Automotive Engineers, Warrendale, PA, 1998.
42. V. Palekar et al., 'Evaluation of High-temperature Liquid Lubricants Using the Penn State Micro-oxidation Test', *Lubrication Engineering*, **52**(4), 327–334 (1996).
43. T.W. Selby, 'Method of Controlling Base Oil Production Quality Using a Thin Film Oxygen Uptake Test Technique', in American Chemical Society Symposium on *The Processing, Characterization and Application of Base Oils: Part II*, San Diego, California, 1994.
44. S.M. Hsu et al., 'Relationship between Lubricating Basestock Composition and the Effects of Additives on Oxidation Stability', SAE 821237, Society of Automotive Engineers, Warrendale, PA, 1982.
45. D.W. Murray et al., 'The Effect of Basestock Composition on Lubricant Performance', *Petroleum Rev.*, **36**, 36–40 (1982).
46. R.J. Robson, 'Base Oil Composition and Oxidation Stability', Symposium on Trends in Lube Oil Base Stocks, American Chemical Society, Division of Petroleum Chemistry Preprints, Washington, DC, 1094–1100, 1984.
47. J.A. Supp et al., 'Deposit Formation in Gasoline Engines. Part 1. Base Oil Effects in Sequence VE Deposits', *Lubrication Engineering*, **50**(12), 964–968 (1994).
48. G.R. Blackburn, 'Assessing the Carcinogenic Potential of Lubricating Base Oils', *Lubrication Engineering*, **54**(8), 17 (1998).

49. K.L. Chasey, R.H. McKee, 'Evaluation of the Dermal Carcinogenicity of Lubricant Base Oils by the Mouse Skin Painting Bioassay and Other Proposed Methods', *J. Appl. Toxicol.*, **13**(1), 57–65 (1993).
50. ASTM E 1687, *Annual Book of ASTM Standards*, ASTM Conshohocken, PA, Vol. 11.03, 1998.
51. G.R. Blackburn et al., 'Predicting Carcinogenicity of Petroleum Distillation Fractions Using a Modified Salmonella Mutagenicity Assay', *Cell Biol. Toxicol.*, **2**(1), 63–84 (1985).
52. M.V. Reddy, K. Randerath, 'Nuclease P1-Enhanced 32P-Postlabeling Assay for Aromatic Carcinogen-DNA Adducts', in *Methods in Toxicology, Part B, In Vitro Toxicology Indicators*, eds. C.A. Tyson, J.M. Frazer, 201–215, Vol. 1, Academic Press, New York, 1994.
53. IP 346, Standard Methods for Analysis and Testing of Petroleum and Related Products, Institute of Petroleum, London, Vol. 2, 1998.
54. Commission Directive 94/69/EC of 19 December 1994 Adapting to Technical Progress for the twenty-first time Council Directive 67/548/ECC on the Approximation of Laws, Regulations and Administrative Provisions Relating to the Classification, Packaging and Labeling of Dangerous Substances. Official Journal of the European Communities No L381, European Commission, Luxembourg, 1994.
55. M. Voltz et al., 'Biodegradability of Lubricant Base Oils and Fully Formulated Products', in *American Chemical Society Symposium on The Processing, Characterization and Application of Base Oils: Part II*, San Diego, California, 1994.
56. T.E. Kiovsky et al., 'Biodegradable Hydraulic Fluids and Related Lubricants', SAE 942287, Society of Automotive Engineers, Warrendale, PA, 1994.

Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams

Stuart E. Scheppele
Research Resources Center, University of Illinois, Chicago, USA

1 Introduction **2**
2 Group (Chemical) Types **2**
2.1 Definition and Significance in Terms of the Concept of Homology 2
2.2 Mass Spectrometric Nomenclature: the Specific *Z* Series Concept 3
2.3 Deduction/Determination of Group Types from High-resolution Mass Spectrometric Homolog Analyses 5
3 Sample Preparation **7**
3.1 Distillation 7
3.2 Solvent Separations – Solvent Extractions 8
3.3 Liquid Chromatography 8
4 Sample Introduction **8**
4.1 All Glass Heated (Batch) Inlet Systems 8
4.2 Direct Introduction Probe 14
5 Ionization Techniques: Applications **21**
5.1 Low-voltage Electron Ionization 21
5.2 Field Ionization 21
5.3 Other Ionization Techniques 21
6 Qualitative Analysis **22**
6.1 Processing of High-resolution Mass Spectra 22
6.2 Resolving Power: Definitions 23
6.3 Dependence of Resolving Power on the Differences in Composition, Mass-to-charge Ratios and Relative Abundances of Successive Ion Beams 24
6.4 Mass Analyzers and the Resolution of Ion Beams 24
7 Quantitative Analysis **27**
7.1 Electron Ionization Cross-sections 28
7.2 Low-voltage Ionization Cross-sections and Ionization Sensitivities 29
7.3 Calculation of Carbon Number Distributions and Chemical Properties 30
7.4 Fundamental Phenomena 32
7.5 Practical Consequences of Fundamental Phenomena 32
7.6 Electron Impact Ionization and Photoionization Cross-sections 34
7.7 Low-voltage Relative Mole Sensitivities for Ionization of Aromatic Compounds 34
7.8 Summary 38
Acknowledgments **39**
Abbreviations and Acronyms **39**
Related Articles **39**
References **39**

The complex chemistry and chemical engineering of fossil fuel materials including petroleums, coal-derived liquids, tar sands, shale oils, refinery streams and refined products can often be adequately simplified by expressing the composition of these materials in terms of group types. Group (chemical) types are series of homologous compounds that exhibit similar chemical properties and a regular gradation of physical properties. A homologous series of compounds, the members of which are termed homologs, contains compounds in which successive members exhibit the same difference in chemical formula with the homologous unit of present interest being CH_2. Mass spectrometry (MS) is an ideal, extensively used technique in analyzing for chemical types in terms either of the total amounts of the various homologous series of compounds present in a sample or of the individual members (homologs) of the various series of homologous compounds present in a sample. The focus of this article is the fundamentals of homolog-type analysis because analyses expressed in terms of the amounts of individual homologs rather then in terms of the total amounts of group types benefits the understanding/prediction of the effect of experimental conditions on feedstock production/transportation, product distributions in feedstock refining, end-product quality and environmental impacts. A homolog-type analysis also permits the determination/prediction of chemical/physical properties of a sample that can be expressed as a function of composition; examples being reaction rates, distillation curves and elemental analyses. The discussion of techniques for obtaining the ''molecular ion mass spectra'' required to produce homolog-type analyses focuses on the use of all glass heated inlet systems (AGHIS) and direct probes for sample introduction, low-voltage electrons for ion formation, and double focusing magnetic sector (DF) and Fourier transform ion cyclotron

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

resonance (FTICR) mass analyzers to resolve isobaric ions.

1 INTRODUCTION

The complex chemistry and chemical engineering of fossil fuel feedstocks including petroleums, coal-derived liquids and tar sands and shale oils, refinery streams and refined products can often be adequately simplified by expressing the composition of these materials in terms of series of compounds that exhibit similar chemical properties and a regular gradation of physical properties. Such series of compounds have historically been termed group types. However, the phrases ''group-type'' and ''chemical-type'' are taken to be synonyms throughout this article. The concept of homology is implicit in this approach to formulating such analyses. A homologous series of compounds, the members of which are termed homologs, contains compounds in which successive members exhibit the same difference in chemical formula.

Mass spectrometric methods for performing chemical-type analyses can be classified according to whether the analytical results are expressed in terms of the total amounts of the various homologous series of compounds present in a sample or the individual members (homologs) of the various series of homologous compounds present in a sample. The extensive literature devoted to the development and applications of mass spectrometric techniques for performing such analyses over the past 50 years testifies to their extensive use in the fossil fuel industry.[1–98] The intent of the literature cited here[1–98] is to provide those who are new to the field of homolog-type analysis with an introduction to this body of literature and to provide both a background and framework for this article which focuses on the fundamentals of homolog-type analysis. The author apologizes to any of his colleagues whose work was not cited.

This article focuses on the fundamentals of homolog-type analysis for a number of reasons. High-resolution mass spectrometric (HRMS) methods are generic in terms of analyzing for chemical types whereas low-resolution methods necessarily assume that the group types comprising the sample are known. Analyses expressed in terms of the amounts of individual homologs rather than in terms of the total numbers of group types clearly benefits the understanding/prediction of the effect of changes in experimental conditions on chemical and physical processes affecting issues such as corrosion, product distributions, end-product quality and environmental impact. Finally, a homolog-type analysis permits prediction of any chemical or physical property of a sample that can be expressed as function of composition. Examples include the calculation of reaction rates, mass balances, average molecular weights for each chemical type and for the sample, the distillation characteristics for each group type and for the sample and elemental analyses for the predicted distillate cuts and the total sample. The discussion of techniques for obtaining ''molecular ion mass spectra'' focuses on the use of AGHIS and direct probes for sample introduction, low-voltage electrons for ion formation, and DF magnetic sector and FTICR mass analyzers to resolve isobaric ions. Topics considered in processing these spectra include the assignment of formulas to ion masses obtained from the conversion of time centroids and frequencies in mass spectra acquired using DF and FTICR instruments, respectively, the generation of quantitative distributions from molecular-ion abundances and the calculation of chemical/physical properties from these distributions. In writing this article the author hopes to accomplish in part the goals of providing the individual who is new to the field with an introduction to the fundamentals of HRMS homolog analysis and with an indication of issues in these methods that need to be addressed and resolved.

2 GROUP (CHEMICAL) TYPES

2.1 Definition and Significance in Terms of the Concept of Homology

As noted above, compounds that differ in composition by CH_2, undergo similar chemical reactions and exhibit a regular gradation of physical properties under specific experimental conditions constitute a group or compound type. A chemical (group)-type is a homologous series of neutral or ionic substances. However, as discussed in section 2.3 a homologous series may contain more then one group type. Isomers differ in structure but not in formula and, hence, are subsets in a set of homologous substances. For example, benzene (C_6H_6), and benzene substituted with alkyl groups containing 1, 2, 3, 4, . . ., carbons differ in composition by CH_2 and, hence, correspond to a homologous series of compounds. The ionization of these compounds in a mass spectrometer produces a homologous series of molecular ions. Furthermore, isomers are subsets of homologs of benzene containing two or more alkyl carbons. For example, the third homolog in the series (C_8H_8) may reflect the presence of one or more of the four possible isomers, ethylbenzene and *o*-, *m*-, and *p*-xylenes. It should be noted that neutral and, hence, ionic homologs consist of compounds and ions containing the various isotopes of their constituent elements.

2.2 Mass Spectrometric Nomenclature: the Specific *Z* Series Concept

2.2.1 Naming Homologous Series

A homolog-type analysis is synonymous with the identification of the homologous compounds which constitute a sample. Unfortunately, the existence of isomers can complicate to varying degrees the nomenclature for these series. For example, benzene and benzene substituted with one or more alkyl groups may unambiguously be referred to as the benzene series with the number of alkyl carbons possessed by each homolog specified numerically. Thus, the name C-2 alkylbenzenes encompasses ethylbenzene and the isomeric dimethylbenzenes. However in many instances the existence of isomers makes this approach to nomenclature quite cumbersome. As an example, consider a homologous series of hydrocarbons beginning with naphthalene ($C_{10}H_8$). The phrase ''naphthalene series'' or simply ''naphthalenes'' would be acceptable nomenclature provided that only naphthalene and its alkylated homologs are present. However, hydroaromatic hydrocarbons such as **(1)** and **(2)**, indenes substituted with one alkenyl group such as **(3)** and naphthenoindenes such as **(4)** and the alkylated homologs of these hydrocarbons are isomers of alkylated naphthalenes containing the same number of carbons. If the series either contains or is believed to contain hydrocarbons such as **(1)** and **(2)** and their alkylated homologs in addition to naphthalenes then the series could be and often is referred to as the naphthalene/trinaphthenobenzene series. The possible presence of hydrocarbons such as **(3)** and **(4)**, isomers of these compounds and their alkylated homologs would require that the series be called the naphthalene/trinaphthenobenzene/alkenylindene/naphthenoindene series. Such difficulties in nomenclature are circumvented by using the *Z*-series concept to classify molecular ions and, hence, their precursor compounds according to the homologous series to which they belong.

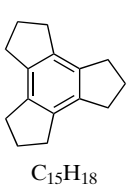

$C_{15}H_{18}$

(1)

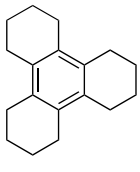

$C_{18}H_{24}$

(2)

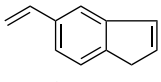

$C_{11}H_{10}$

(3)

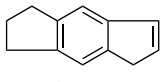

$C_{12}H_{12}$

(4)

2.2.2 Specific Z *Value*

The vast majority of organic compounds in petroleums, synthetic crudes, refinery streams and materials derived from their processing are derived from the combination of two or more of the elements carbon, hydrogen, nitrogen, oxygen and sulfur in various proportions by weight. Petroleums and syncrudes also contain varying amounts of organic compounds substituted with metallic elements such as nickel, vanadium and iron. The general formula of organic and organometallic compounds of present interest is $C_nH_xN_aO_bS_cNi_dV_eFe_f$. Homologs differ in composition by $1, 2, 3, 4, \ldots n$, CH_2 groups. Therefore, we associate $2n$ of the x hydrogens in the general formula with the n carbons leaving $x - 2n$ hydrogens. Thus, the general formula becomes $(CH_2)_nH_{x-2n}N_aO_bS_cNi_dV_eFe_f$. Homologs have the same value of the quantity $x - 2n$. This difference is defined as the specific Z value as shown in Equation (1)

$$Z = (x - 2n)(N_aO_bS_cNi_dV_eFe_f) \tag{1}$$

As seen in Equation (1), a homologous series is designated by the numerical value of Z followed by the symbols for and the numbers of the heteroelements that are present. The symbol H is used for designating hydrocarbons because all elemental coefficients other then those for carbon and hydrogen are zero. Thus, the general formula for compounds and ions of present interest may be written as $(CH_2)_nH_ZN_aO_bS_cNi_dV_eFe_f$.

Compounds and ions of present interest may be viewed as being derived from combinations of basic structural groups. Table 1 lists specific Z values for a number of these structural groups. Thus, the specific Z value for any species considered as being derived from these structural groups species is given by

$$Z = \sum_i \alpha_i Z(X_i) \tag{2}$$

Table 1 Specific Z values for structural groups

Group (X_i)	$Z(X_i)$
C	−2
CH	−1
$(CH_2)_n$	0
CH_3	1
H	1
X=N, O, S	0
XH=NH, OH, SH	1
C=O	−2
C−C/C=C/C≡C	−4
CH=CH	−2
CH=N	−1
C_6H_5	−7

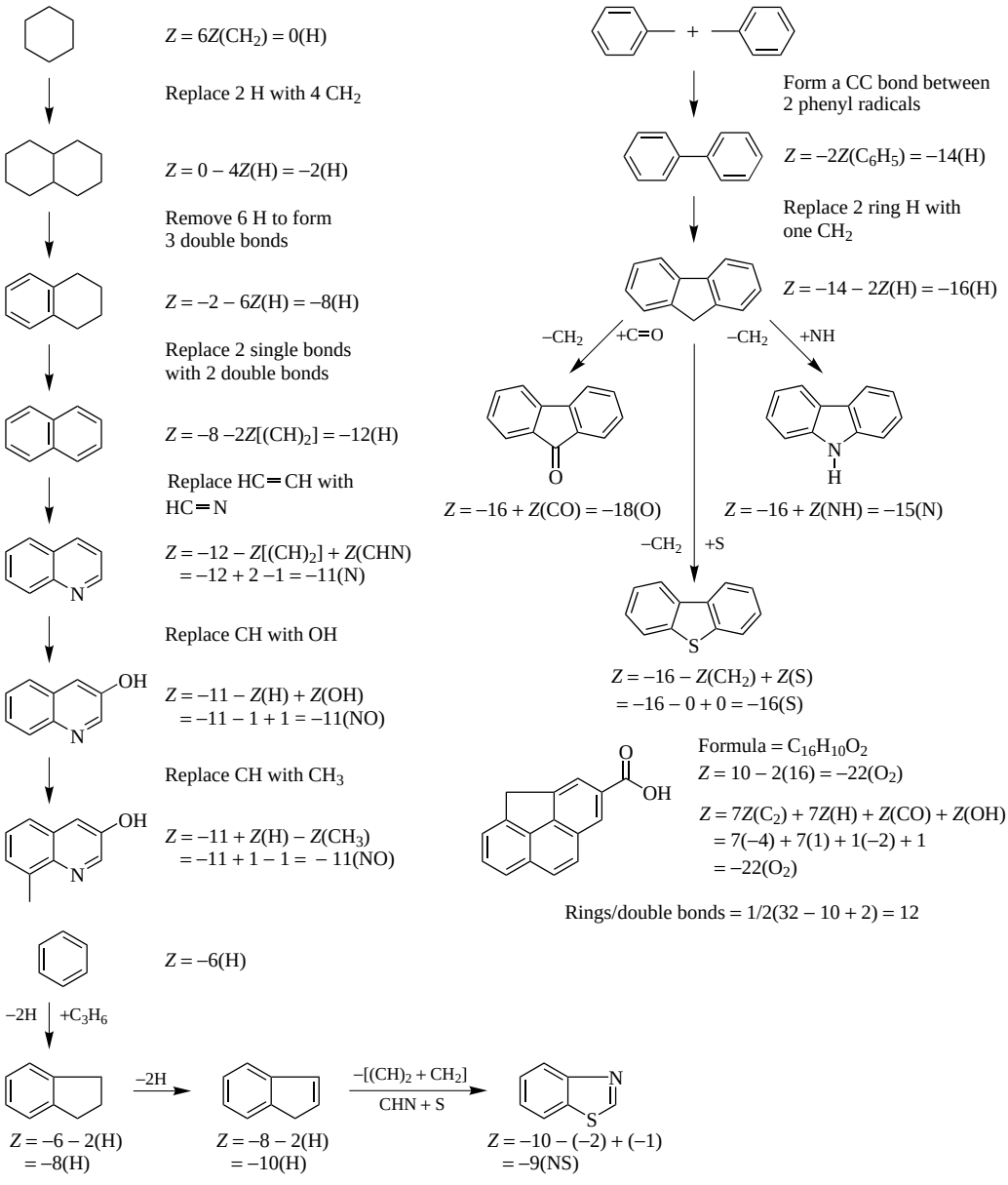

Figure 1 Examples illustrating the specific Z series concept, prediction of Z values for species from Z values of functional groups and relation between Z values and rings plus double/triple bonds.

where α_i is the number of the ith structural group in the structure. From Table 1 and Equation (2) it follows that the number of rings plus double/triple bonds present in a species is

$$\mathcal{R} = \tfrac{1}{2}[2n - x + a + 2] \tag{3}$$

Solving Equation (3) for the quantity $x - 2n$ and substituting the result into Equation (1) results in the Equation (4) which expresses the specific Z value as a function of the number of rings plus double/triple bonds

$$Z(\mathrm{N}_a\mathrm{O}_b\mathrm{S}_c) = a + 2(1 - \mathcal{R}) \tag{4}$$

Figure 1 contains a representative number of examples that illustrate the application of the specific Z series concept, the use of the Z values for functional groups in Table 1 and the relation between the specific Z value and the number of rings plus double/triple bonds.

In the opinion of the author, the practice of referring to the specific Z value as the hydrogen deficiency should be discontinued for several reasons. For example, consider naphthalene ($C_{10}H_8$) which from Equation (1) has a specific Z value of -12(H). As applied to this compound, a logical question that never seems to be answered is what is the chemical significance of the statement

that naphthalene has a hydrogen deficiency of −12? A definition of hydrogen deficiency that has been used is the difference in the number of hydrogens between a compound and the alkane with the same number of carbons. Although this definition gives correct Z values for some species it is not generally valid because it predicts a specific Z value of −14(H) for naphthalene, that is, the difference in hydrogen numbers between $C_{10}H_8$ and $C_{10}H_{22}$ is −14. Addition of 12 hydrogens to naphthalene would presumably result in cyclodecane or a set of isomeric decenes. Another definition that can be found in the literature is the difference in hydrogen numbers between a species and a cycloalkane or an alkene. Clearly, this is a general definition with the caveat that the reference and subject species have the same carbon number. For example, consider pyrrole which has the formula C_4H_5N and hence, $Z = -3$(N). Thus, to get the correct Z value, the formula for pyrrole must be compared to an alkene or cycloalkane having four and not five carbons as the latter would yield $Z = -5$(N) for pyrrole. However this definition with or without the added caveat produces paradoxical results for alkanes. First, there is no alkene or cycloalkane to compare to methane. Second, the difference in hydrogen numbers between all alkanes having carbon numbers greater then 1 and the corresponding alkenes is +2. Thus, alkanes have a hydrogen deficiency of +2. That is, alkanes have an excess not a deficiency of hydrogens.

2.3 Deduction/Determination of Group Types from High-resolution Mass Spectrometric Homolog Analyses

Two concepts underlie mass spectrometric methods for homolog analysis. First, removal of an electron from a homologous series of compounds produces the corresponding series of homologous molecular ions. Thus, use of an ionization technique that generates mass spectra either dominated or characterized by the presence of molecular ions and the accurate measurement of the mass-to charge ratios of these ions suffices to identify the homologs present in a sample. Second, the signal magnitudes of these ions allow the contribution of each of these homologs to the sample to be calculated given quantitative relationships between the isotopically most abundant molecular ion signal magnitudes and the structures of the neutral precursors to these ions. As discussed in subsequent sections, implementation of these concepts is not necessarily straightforward. However, we assume the successful implementation of them in order to consider briefly the equally important topic of determining the number of group types present in a homologous series of compounds. In general, homologs may consist of two or more isomers. The chemistry of isomers depends upon the experimental conditions. Given the compositional complexity of materials of present interest, the possibility that homologous series may contain more then one chemical (group) type should be considered by both the analysts and the chemists/chemical engineers in planning the analytical scheme and/or interpreting the results of a homolog-type analysis.

Figures 2, 3 and 4 contain carbon number distributions (CNDs) from four homolg-type analyses. The phrase ''carbon number distribution'' refers to the format conventionally used in reporting the results of homolog-type analyses. The numbers of the homologs identified as being present in a sample are normally calculated on

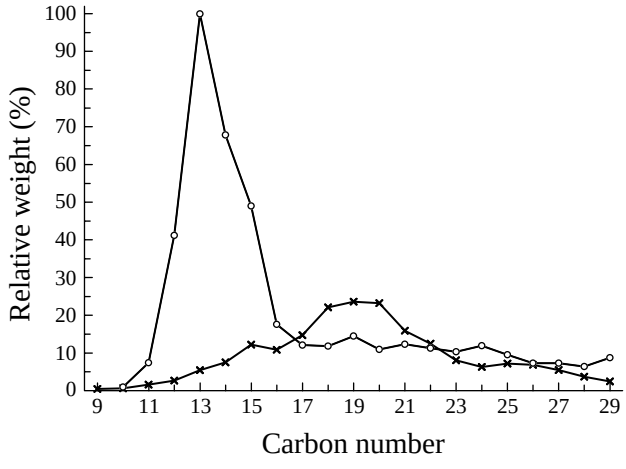

Figure 2 CNDs for the −6(H) (×) and −12(H) (○) Z series obtained from high-resolution FI mass spectra of the aromatic neutrals from separation of coal–liquid-derived oils.

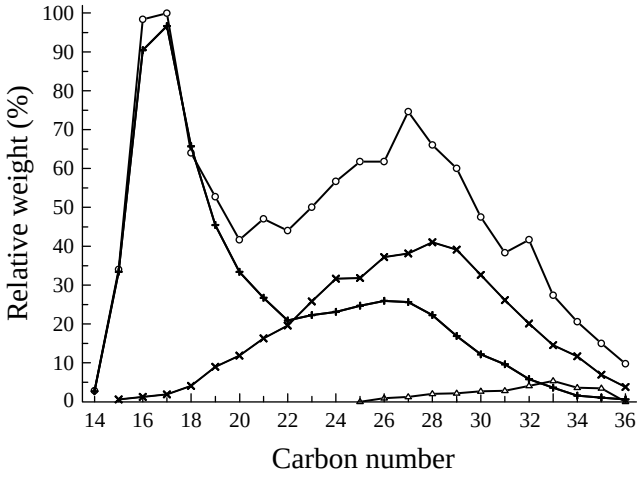

Figure 3 CNDs for the −18(H) Z series in a gas oil and the monoaromatic, diaromatic and triaromatic fractions from its ring number separation. ○, total aromatics; △, ring fraction 1; ×, ring fraction 2; +, ring fraction 3.

a percentage or absolute weight, mole or volume basis. The results are usually reported as a two-dimensional array. The rows or columns are the specific Z series designations for the homologous series and the columns or rows correspond to the number of carbons present in each homolog. Since the numbers of the members of each homologous series identified as being present in the sample are reported as a function of the number of carbons present in each homolog, the tabulation is termed a CND. Figures 2, 3 and 4 contain weight percentages normalized to a value of 100 for the most abundant homolog in the more abundant series, if two or more series are present.

Figure 2 contains CNDs for the -6(H) and -12(H) series obtained from high resolution field ionization (FI) mass spectra of the aromatic neutrals from chromatographic separation of coal–liquid-derived oils.[61,99] The -6(H) aromatic series can only contain benzene, toluene and either *n*-alkylated benzenes or isomeric benzenes substituted with two or more alkyl groups or both. Thus, we conclude that the -6(H) CND in Figure 2 reflects the presence of a single group-type consisting of homologs of benzene which have 3–23 alkyl carbons.

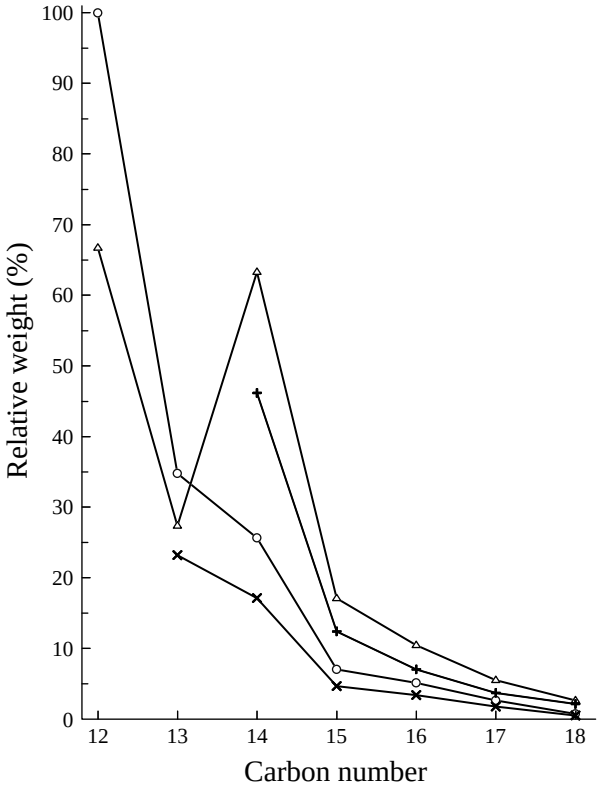

Figure 4 CNDs for the -14(H) Z series in an anthracene oil and in a hydrotreated product. ○, in feed; △, in products; ×, feed group-type in products; +, nonfeed-type(s) in products.

The nature of a CND can often yield significant inferences concerning the presence or absence of more then one group type in a homologous series. For example, in Figure 2 the first member in the -12(H) series has the formula $C_{10}H_8$ and, hence, reflects the presence of naphthalene. This result, the shape of the distribution and the fact that the $C_{10}-C_{16}$ homologs dominate the distribution, leads to the conclusion that the -12(H) series is reasonably attributed to the presence of a single group type, that is, naphthalene and its alkylated homologs. However, definitely to exclude compound types such as **(1–4)** as significant contributors to the -12(H) series would require additional experiments such as performing HRMS analyses on fractions obtained from a ring number separation of the aromatic neutrals.

The bimodal CND for the -18(H) series in Figure 3 clearly indicates the presence of at least two group types. The first homolog in the series has the formula $C_{14}H_{10}$ and is attributed to the presence of anthracene **(5)** and/or phenanthrene **(6)**. Either one or both of these hydrocarbons containing 1–6 alkyl carbons undoubtedly contributes to the $C_{15}-C_{20}$ homologs. However, the shape of the distribution is evidence that at least one additional group type contributes to homologs which have carbon numbers greater than 20. As seen in Figure 3, the CNDs for the -18(H) series from the HRMS analyses of the fractions from the ring number separation of the total aromatic neutrals substantiates this conclusion. The results for the monoaromatic or first ring fraction reveal the presence of a group type consistent with the presence of hydrocarbons containing one aromatic and six naphthene rings such as **(7)** and their alkylated homologs. The results for the second ring fraction support the presence of a diaromatic chemical type reflecting the presence of hydrocarbons such as trinaphthenonaphthalenes and alkylated homologs of these hydrocarbons. The results for the third ring fraction confirm the presence of a three aromatic ring chemical-type containing 14–36 carbons.

CNDs for the -14(H) series in an anthracene oil and one of the products from its hydrotreating are reproduced in Figure 4.[60] The distribution for the feed is consistent with the presence of a single chemical type. This conclusion is supported by gas chromatography/mass spectrometry (GC/MS) analysis which revealed that acenaphthene **(8)** and biphenyl **(9)** accounted for 94% and

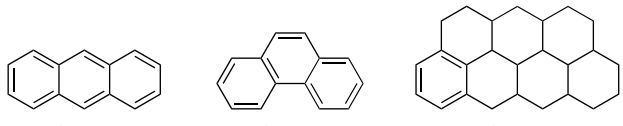

$C_{14}H_{10}$ **(5)**

$C_{14}H_{10}$ **(6)**

$C_{25}H_{32}$ **(7)**

6% of the first homolog, respectively.[60] The CND for the −14(H) series in the products shows that hydrotreating deceased the weight percentages of the C_{12} and C_{13} homologs and resulted in significant increases in the amounts of the C_{14}–C_{18} homologs. This and other results show the occurrence of both hydrogenation reactions introducing hydroaromatics such as **(10)** and **(11)** and hydrogenation/hydrogenolysis reactions introducing aromatics such as **(12)** and **(13)** into the −14(H) series. Estimates of the percentage contributions of **(8)** and **(9)** and homologs of these hydrocarbons and products formed from the hydrogenation of phenanthrenes **(10)** and **(11)** and hydrogenation/hydrogenolysis of phenanthrene **(12)** and pyrene **(13)** to the −14(H) series in the products are shown in Figure 4. These estimates were obtained by assuming that the C_{13}–C_{18} homologs in the feed undergo the same percentage decrease as does the C_{12} homolog.

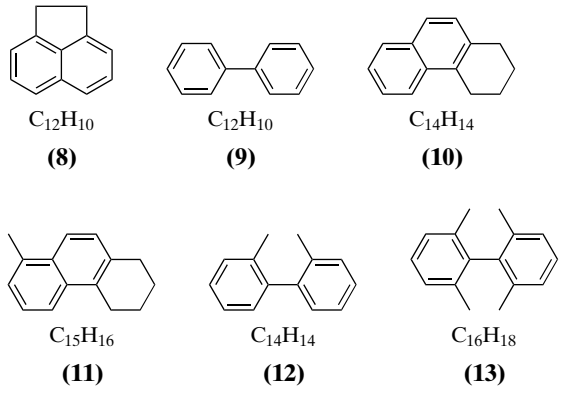

High-resolution low-voltage electron ionization mass spectra (HRLVEIMS) of fossil fuel materials which have atmospheric equivalent boiling points (AEBP) up to about 1050 °F (566 °C) will in general contain series of molecular and fragment ions. The fragmentation processes undergone by molecular ions in a given homologous series and the yields of these ions relative to the molecular ions depends upon a number of factors which are considered in section 7. What is of present interest is the fact that the presence of these fragment ions must be taken into account in processing mass spectra in order to avoid erroneous conclusions regarding the group types present in a sample. The occurrence of this phenomena is illustrated in Figure 5, which contains a plot of the −6(H) homolog molecular ion relative abundances versus carbon number and could be interpreted as reflecting the presence of a −6(H) series in a >650 °F (343 °C) AEBP distillate that contains homologs with carbon numbers ranging from 7–40. However, such a conclusion would be erroneous for two reasons. First, the initial carbon number in the series would be expected

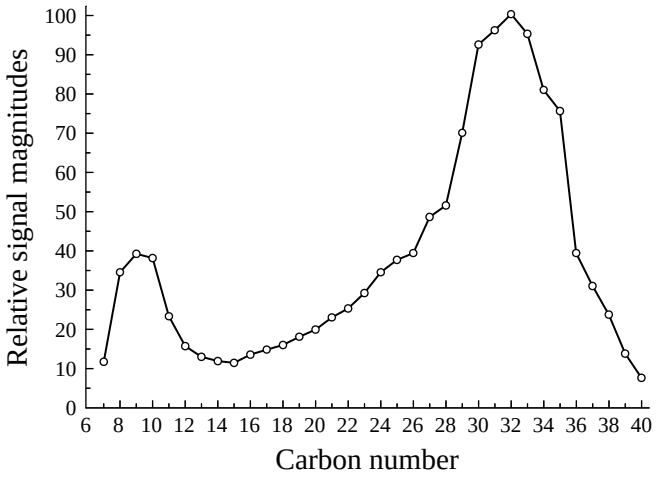

Figure 5 Plot of the relative signal magnitudes versus carbon number for the −6(H) *Z* series in the high-resolution low voltage mass spectrum of gas oil aromatic neutrals.

to be about 15, based upon the boiling range of the sample. Second, one fragmentation reaction of benzene molecular ions substituted with relatively long-chain alkyl groups phenomenologically involves hydrogen transfer from the alkyl group to the ring with loss of an alkene resulting in formation of parent ions.[100,101] Although this process is dominated by α-cleavage at 70 eV, it becomes relatively important at low voltage.[100,101] Thus, the ion abundances from carbon numbers 7 to about 14 are attributed to the occurrence of such rearrangement fragmentation reactions.

3 SAMPLE PREPARATION

It is generally necessary to subject crude oils, syncrudes and materials derived from their processing to physical and/or chemical separations prior to determining the compositions of these materials. An in-depth discussion of the numerous techniques available for performing these separations is clearly beyond the scope of this article. Rather the intent of this section is simply to provide the reader with an overview of a number of approaches to generating fractions that are suitable for analysis. The reader should be aware of the fact that the approach to separating a given material appears often to be determined either by the past history of the analytical chemistry of a given fossil fuel material or by the preferences of a given research group.

3.1 Distillation

Distillation is an important technique in the characterizations of crude oils. It has also been employed as the

first step in the analysis of other fossil fuel materials such as coal liquids and shale oils. Atmospheric pressure distillations resulting in distillate fractions boiling up to approximately 650 °F (343 °C) are conventionally performed using the American Society of Testing and Materials (ASTM) method D2892.[102] Middle distillates contain compounds boiling from about 400 to about 650 °F (205–343 °C). Gas oils or vacuum gas oils which have AEBPs ranging from approximately 650 °F (343 °C) to approximately 1050 °F (566 °C) are obtained by distilling the residue from a D2892 distillation at reduced pressure using ASTM method D1160.[103] Compounds having AEBPs up to about 1300–1400 °F (705–760 °C) can be recovered from a D1160 vacuum residue using the technique of molecular distillation.[104,105] It should be noted that these distillations produce fractions that do not have sharp cut points in terms of composition, with the degree of overlap tending to increase with increasing boiling point. Distillate cuts boiling up to 400 °F (205 °C) are amenable to analysis using techniques such as gas chromatography (GC) and combined GC/MS. Applications of HRMS group-type methods are usually limited to distillate cuts having initial AEBPs of at least 400 °F (205 °C).

3.2 Solvent Separations – Solvent Extractions

The development of coal-liquefaction technology has historically used the technique of solvent separation or solvent extraction to separate coal-derived liquids into fractions termed oils (maltenes), asphaltenes and preasphaltenes.[48,52,61,80,106] The technique has also been applied to other materials such as bitumens.[67] A high-performance liquid chromatography (HPLC) solvent fractionation technique based upon the solubility properties of solvents has been described.[105] Acidic and basic chemical extraction techniques have been used to recover basic and acidic compounds from syncrudes and petroleums.[50,72,79,80,82] It should be noted that the fractions obtained from solvent extractions are operationally defined in terms of the procedure used to generate them.

3.3 Liquid Chromatography

Liquid chromatography is used to separate crude oils, syncrudes and materials derived from their processing into compound classes. The nature of the fractions resulting from these separations are operationally defined in terms of the procedure used to generate them. One approach is to separate a sample into classes of compounds that exhibit for a given procedure either acidic, basic or neutral chemical properties. One method for generating acidic and basic compound classes uses ion-exchange chromatography to remove compounds that exhibit acidic and basic properties with respect to anion and cation resins, respectively.[107,108] Chromatography using $FeCl_3$ impregnated Attapulgus clay as the stationary phase has been used to remove nitrogenous compounds present in neutral fractions. A second technique for isolating acids and bases involves chromatography over acidic and basic alumina to recover compounds that exhibit basic and acidic properties, respectively.[77] A number of chromatographic procedures have been developed for separating neutral fractions into saturated and aromatic compound classes.[50,77,109,110] Solid–liquid chromatography using activated silica gel is required to separate saturates from aromatic neutrals. This separation can be performed using either HPLC or open column (low pressure) chromatography. Gradient elution chromatography through a dual-packed adsorption column containing silica gel and alumina has been used to separate neutral concentrates into saturate and operationally defined monoaromatic, diaromatic and polyaromatic-polar fractions, that is, compounds containing one, two, three or more aromatic rings and/or heteroatoms.[109] Preparative scale HPLC procedures employing stationary phases such as chemically bonded aminopropylsilica,[77] 2,4-dinitroanilinopropylsilica[110] and bonded charge-transfer agents[111] have been developed for fractionating aromatic neutral fractions into operationally defined compound classes such as monoaromatics, diaromatics, triaromatics and tetra and greater aromatics, that is, compounds containing one, two, three and four or more aromatic rings. It should be noted that the resolution achieved in these separations decreases with increasing boiling point.

4 SAMPLE INTRODUCTION

4.1 All Glass Heated (Batch) Inlet Systems

The introduction of fossil-fuel materials having AEBPs up to about 1050 °F (566°C) requires an inlet system capable of being heated and evacuated. The design should be compatible with the introduction of liquids and solids. Quantitative analysis of mass spectral data is necessarily simplified if the composition of the vapor entering the ion source is, at all times, functionally related to the original composition of the sample under analysis. Therefore, a reservoir of known volume is incorporated into the inlet system and a leak of known leak rate is inserted between the reservoir and the ion source. The sample is first expanded into the reservoir, hence, the term expansion volume, and then allowed to flow from the reservoir to the ion source via the leak. The glass line connecting the inlet system to the ion source housing is conventionally termed the leak line. The device which connects to the leak line at the source housing and conducts the vapor to the ion

source is termed the reentrance. All components of the inlet system including the leak line and the reentrance should be capable of achieving and maintaining the same temperature.

4.1.1 Inlet Designs

The first successful reservoir–inlet system meeting these criteria was developed by Lumpkin and Taylor.[112] Gallium was used to maintain the integrity of the vacuum within the inlet system during sample introduction. A capillary pipette was used to introduce volatile liquids via a gallium-covered fritted disk. Solids and relatively nonvolatile liquids were introduced using a removable, heatable holder containing gallium. All subsequently designed high-temperature reservoir inlet systems familiar to the author can be considered to be variants of the Lumpkin–Taylor inlet.[113–117] Recent versions of this inlet are described below.

The Brunfeldt AGHIS was a joint development between R.J. Brunfeldt of the Brunfeldt Co., Bartlesville, OK (presently located in Bixby, OK) and the author and other members of the Instrumental Section at the Bartlesville Energy Technology Center, US Department of Energy, Bartlesville, OK. The basic approach was to design an inlet system incorporating new technology while maintaining the salient concepts of the Lumpkin–Taylor inlet. The methods for sample introduction and for actuating the magnetically operated valves within the oven are shown in Figure 6. Figure 7 shows the construction of the oven and the components of the inlet system within the oven.

As shown in Figure 6, the valve lifters are activated by controls on the bottom of the oven. This design allows for immediate direct access to the interior of the oven by simply lifting off the top cover of the oven. The availability of high-temperature GC septums exhibiting negligible bleed permitted use of a heatable septum inlet for introduction of relatively nonviscous liquids as shown in Figure 6. Figure 6 also shows that a removable glass cup which is heated by an independent oven permits the introduction of solids and liquids having a wide range of volatilities because the initial temperature of the cup can be subambient. Optically flat glass surfaces on the top of the cup and at the end of the glass line exiting the oven assures the integrity of the vacuum.[114]

As seen in Figure 7, the oven incorporates a reservoir which has a volume of 900 mL. Figure 7 also shows the use of high-temperature ceramic insulation to maintain excellent temperature control within the oven. The sections of the septum inlet and the leak line passing through the oven walls are encased in aluminum blocks containing cartridge heaters in order to eliminate cold spots. As is evident in Figure 7, difficulties in performing any required maintenance is minimized by the capability of simply accessing the interior of the oven and by the location of components in the oven. The use of large diameter sample lines provides for rapid and complete residual sample removal. In order to achieve significantly reduced leak rates across the ball valves, each valve

Figure 6 Front of AGHIS oven showing septum and glass-cup inlets and controls for actuating magnetically operated valves.

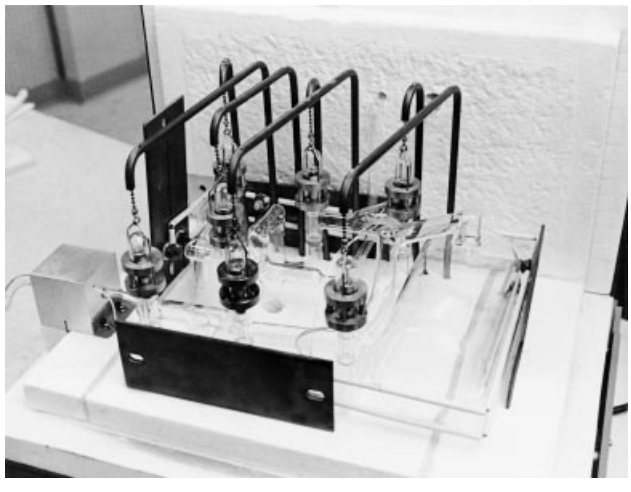

Figure 7 Interior of oven in AGHIS.

is fabricated from glasses having dissimilar hardness. Placement of the valves provides for nonsynergistic variations of pressures within the inlet system.

The design of the reentrance needs to allow for expansion/contraction of the ion source as it is heated/cooled while maintaining good transport of the vaporized compounds into the ion source. With respect to the Kratos MS-50, this requirement was met by installing a Brunfeldt reentrance. This design incorporates threads on the end of the stainless steel housing of the reentrance which is matted via a vacuum flange and a gold gasket to a port in the ion source housing. A bellows and a threaded collar are used to vary the distance between the ion source and the reentrance tube. A thin glass helix inserted in the transfer line in the inlet oven allows for changes in the position of the leak line as the distance between the reentrance glass tube and the ion source is varied. The distance between the glass tube and the ion source is increased/decreased by decreasing/increasing the number of turns of the collar on the reentrance housing threads. For example, with the bellows in a relaxed position, the inlet oven, the leak line, the reentrance glass tube and the ion source are brought to a desired temperature. An aliquot of *p*-xylene is admitted to the inlet and the increase in the signal magnitude of the *m*/*z* 106 ion is monitored as the bellows is compressed to reduce the distance between the reentrance tube and the ion source. The onset of a decrease in the ion signal magnitude is used to establish the occurrence of contact between a port in the ion source and the reentrance tube. The collar is then turned in the opposite direction to relax the bellows first to remaximize the ion signal and then to decrease the ion signal slightly in order to achieve a slight distance between the ion source port and the reentrance glass tube. This procedure has been found to give good transfer of vapor into the ion source while maintaining excellent alignment of the source which is critical for achieving optimum performance of the mass spectrometer at high resolving powers (RPs). In addition, use of this reentrance was found significantly to reduce contamination of both the ion source housing and the glass electrical feedthroughs on the end flange of the ion source.

A turbomolecular pump backed by a mechanical pump is used to reduce the pressure in the system to less than 10^{-3} Torr [1 Torr = (101 325/760) N m^{-2}] as measured by a thermocouple gauge. The design of the power supplies that provide electrical power to the heaters incorporates temperature regulators connected to thermocouples. Thus, the temperatures in the various sections of the inlet including the cup inlet, oven, leak line and reentrance are controlled to within about ±2 °C. Temperatures as high as 325 °C are maintained routinely. The septum inlet has been operated at temperatures as high as 325 °C without noticeable thermal deterioration of the septum, although temperatures up to 300 °C are more commonly used. The time required to raise the temperature of the cup inlet to a given value can be significantly reduced by preheating the door heater shown in Figure 6, for example, the temperature in the cup-inlet oven rose from ambient to 300 °C in about 75 s after preheating the oven for about 4 min.

Gallegos and Pazzi reported the design of a miniature batch inlet system that they termed a dynamic batch inlet system (DBIS).[116] The system incorporated an expansion bulb with a volume of only 2 mL and was used in conjunction with a gas chromatograph. The bulb was placed in the GC oven with the inlet and outlet lines connected to the GC injector and the MS ion source using deactivated uncoated 0.25-mm fused-silica capillary tubing. Hydrogen was used as the carrier gas at a flow rate of about 1 mL min^{-1}. The temperatures of the injector, oven and transfer line were set to 300 °C. Injection volumes of samples such as a 350–550 °C AEBP distillate cut were 0.02–0.05 μL.

Roussis and Cameron evaluated the operational characteristics of a modified version of the Gallegos and Pazzi DBIS fabricated by the Brunfeldt Company.[117] Substitution of helium for hydrogen as the carrier gas and variations in the dimensions of the capillary lines resulted in residence times of at least 15 min. The ability to introduce samples without using a carrier gas was also reported. Excellent agreement was obtained between the results of low-resolution type analyses performed using the DBIS for sample introduction and the corresponding results obtained using a Brunfeldt AGHIS for sample introduction.

The following are several issues that must be addressed in any consideration of substituting a DBIS for an AGHIS in HRMS-type analysis. Although the ability to use small sample sizes would be a clear advantage of a DBIS over an AGHIS, which generally require samples sizes of 5–12 mg, the actual sample sizes required in HRMS applications and the implications for the required volume of the reservoir and the dimensions of the capillary lines connected to the DBIS would have to be determined. In addition, a further increase in the residence time would be required to permit acquisition of HRLVEIMS at a magnetic field scan rate of 1000 s/mass decade, for example, the time for acquiring 40 000–50 000 RP spectra at a magnetic field scan rate of 1000 s/decade is typically greater then 16 min. Also, design changes that would reduce the time-dependent variation in the ion signal magnitude observed by Roussis and Cameron and that would provide the capability for introducing viscous liquids and solids and for evacuating samples would be mandatory.

The problems encountered by the author and coworkers in attempting to use a GC oven-based inlet system developed by The Brunfeldt Company are relevant in regard to attempting to develop a DBIS for use in HRMS

group-type analyses. Vaporized samples were transferred from the reservoir to the ion source of the mass spectrometer using deactivated fused silica capillary tubing via the heated GC reentrance. Satisfactory flow rates and stable ion signal magnitudes were obtained by empirically adjusting the dimensions of the capillary tubing. The inlet system incorporated septum and cup-based inlets for introduction of liquids and solids. Unfortunately, the design placed these inlets inside of the GC oven. Thus, the introduction of samples required that the GC oven be cooled to room temperature. The excessive time required to raise the temperature to say 300 °C using only the GC oven heater was addressed by installing but not without problems an additional heater in the GC oven. A more serious problem that was never satisfactorily resolved was the effect of GC blower fan vibrations on 40 000 RP low-voltage electron ionization (LVEI) mass spectra.

4.1.2 *Effusive Flow and the Time Dependence of Ion Yields*

We now consider the effect of the time dependence of the sample composition in the reservoir on the relative ion abundances (RIAs). Since the rate of effusion of a gaseous species follows molecular flow and is inversely proportional to the square root of its mass,[118] the partial pressures of the components in the reservoir are time dependent. Thus, the composition of the gas flowing into the ion source is time dependent. The effect of neglecting to correct for this phenomenon on the accuracy of an analysis reflects the rate of effusion through the leak in relation to the masses of the gaseous species, the vacuum-seal integrity of the inlet system, mass spectral acquisition time and the reproducibility in the mass spectra. In order to assess the magnitude of the problem, let N_i. be the number of moles of the ith species having molecular weight MW_i in the inlet at time t (s) and at temperature T(K). Neglecting the steady state pressure in the ion source, the rate of decrease in N_i is[118]

$$\frac{-\mathrm{d}N_i}{\mathrm{d}t} = \frac{AN_i}{V}\left(\frac{RT}{2\pi MW_i}\right)^{1/2} \tag{5}$$

where A is the cross-sectional area (cm^2) of the leak, R is the gas constant (8.314×10^7 erg K^{-1} mol^{-1}) and V is the volume of the reservoir plus the glass line between the reservoir and the leak. Integrating Equation (5) between the limits $N(t_0)$, t_0 and $N(t_i)$, t_i, we obtain Equation (6)

$$\ln\frac{N(t_i)}{N(t_0)} = -k_{\mathrm{eff}}(t_i - t_0) \tag{6}$$

where

$$k_{\mathrm{eff}} = \frac{A}{V}\left(\frac{RT}{2\pi MW_i}\right)^{1/2} \tag{7}$$

The rate of effusion of air through a leak was found to be 0.21 cm s^{-1} at 298 K. Thus, Equation (7) yields an area of 1.80×10^{-5} cm^2 for the leak. The area of the leak was independently determined by monitoring the time dependence of the intensity of the molecular ion of acetone. The m/z 58 intensity decreased by a factor of two in 2398 s. Since the inlet temperature was 573 K and since V is 930 cm^3, Equations (6) and (7) yield an area of 2.35×10^{-5} cm^2. The good agreement between the two independent methods for determining the cross-sectional area of the leak demonstrates the vacuum-seal integrity of the inlet system.

The time dependence of the composition of a gaseous mixture in a reservoir inlet is considered assuming a mixture initially composed of equal moles of −6(H) series homologs. Assume that the magnetic field was scanned exponentially from m/z 650 to m/z 70 at 100 s/decade. Acquisition of mass spectra rather than elapsed time is of immediate relevance to the analyst. Therefore, Table 2 presents the effect of effusion of components across the leak on the composition of the mixture in the inlet as a function of the mass spectrum number rather than as a function of elapsed time following introduction of the sample into the ion source. In Table 2, the effusion constants in column 3 were calculated for the masses in column 2 using the values in footnote a. The values in column 4 are the arrival times for the seven molecular ions in a given mass spectrum calculated using the value specified in footnote c. Columns 6 through 15 list the moles of each of the seven components remaining in the inlet relative to the initial moles at the time of detection of each molecular ion in each of the ten mass spectra. As noted in footnote d, these values were calculated using the effusion constants and times in columns 3 and 4, respectively, assuming an elapsed time of 5 min between sample introduction to the ion source and initiation of the first scan and a total time of 6 s between scans.

Comparison of the mole ratios in columns 6 through 15 of Table 2 with the corresponding initial ratios in column 5 reveals that the moles of each component and, hence, their partial pressures, decrease with increasing time. The decrease in these quantities is seen to increase, as expected, with decreasing molecular weight. Therefore, the number of molecules of each component entering the ion source per unit of time diminishes as time increases with the diminution becoming larger as the molecular weight decreases. Consequently, consecutive scans of the mass spectrum of the mixture will exhibit progressively smaller values for both the molecular-ion signal magnitudes and the total ionization. In order to assess the impact of these observations on the accuracy of a quantitative analysis, it is useful to express the time dependence of the composition of the vapor remaining

Table 2 Variation of the relative mole ratios of the −6(H) series homologs comprising a gaseous mixture in a reservoir inlet at 325 °C as a function of mass spectrum number

Components		$k_{eff} \times 10^{-4}$ [a] (s^{-1})	δT [b,c] for M^+ in single scan (s)	Initial	Mole ratios of commponents in inlet[d] at detection of molecular ions in spectrum									
Formula	Mass				1	2	3	4	5	6	7	8	9	10
$C_{42}H_{78}$	582.610	0.82268	4.754	1	0.975	0.967	0.959	0.951	0.943	0.935	0.927	0.919	0.911	0.904
$C_{36}H_{66}$	498.516	0.88936	11.523	1	0.973	0.964	0.955	0.946	0.938	0.929	0.921	0.912	0.904	0.896
$C_{30}H_{54}$	414.423	0.97543	19.547	1	0.969	0.960	0.950	0.941	0.931	0.922	0.913	0.904	0.895	0.886
$C_{24}H_{42}$	330.329	1.09260	29.397	1	0.965	0.954	0.943	0.933	0.922	0.912	0.902	0.892	0.882	0.872
$C_{18}H_{30}$	246.235	1.2654	42.156	1	0.957	0.945	0.933	0.921	0.909	0.897	0.886	0.874	0.863	0.852
$C_{12}H_{18}$	162.141	1.5594	60.302	1	0.945	0.930	0.916	0.901	0.887	0.873	0.859	0.845	0.827	0.818
C_6H_6	78.047	2.2477	92.056	1	0.916	0.895	0.874	0.854	0.835	0.816	0.797	0.779	0.761	0.744

[a] Calculated for values of $A = 2.076 \times 10^{-5}$ cm^2, $V = 930$ cm^3 and $T = 598$ K using masses in column 2.
[b] Time for detection of a molecular ion during acquistion of a given spectrum.
[c] Calculated for exponential scanning from m/z 650 to m/z 70 at 100 s/decade; i.e. $k_s = 2.3026 \times 10^{-2}$ s^{-1}.
[d] Calculated using k_{eff} values in column 3 and the times in column 4 given 300 s as the elapsed time between opening the valve to the leak line and initiation of acquisiton of the first spectrum and a time of 6 s between acquisiton of spectra.

Table 3 Variation of moles of −6(H) series homologs relative to moles of the $C_{42}H_{78}$ −6(H) series homolog comprising a gaseous mixture in a reservoir inlet at 325 °C as a function of mass spectrum

Components		Mole ratios of components in inlet normalized to moles of $C_{42}H_{78}$ at detection of molecular ions in spectrum[a]									
Formula	Mass	1	2	3	4	5	6	7	8	9	10
$C_{42}H_{78}$	582.610	1	1	1	1	1	1	1	1	1	1
$C_{36}H_{66}$	498.516	0.998	0.997	0.996	0.995	0.995	0.994	0.994	0.992	0.992	0.991
$C_{30}H_{54}$	414.423	0.994	0.993	0.991	0.989	0.987	0.986	0.985	0.984	0.982	0.980
$C_{24}H_{42}$	330.329	0.990	0.987	0.983	0.981	0.978	0.975	0.973	0.971	0.968	0.965
$C_{18}H_{30}$	246.235	0.982	0.977	0.973	0.968	0.964	0.959	0.956	0.951	0.947	0.942
$C_{12}H_{18}$	162.141	0.969	0.962	0.955	0.947	0.941	0.934	0.927	0.919	0.908	0.905
C_6H_6	78.047	0.939	0.926	0.911	0.898	0.885	0.873	0.860	0.848	0.835	0.823

[a] Values in columns 3 through 12 calculated from values in columns 6 through 15 in Table 2, respectively.

in the inlet in terms of the mole ratios normalized to the moles of one component because these values are directly proportional to the relative molecular-ion signal magnitudes observed in the mass spectra of a mixture. The values calculated from the relative mole ratios in Table 2 are tabulated in columns 3 through 12 of Table 3. Assuming all other factors to be equal, the relative mole ratios in Table 3 reveal that incorrect molecular ion abundance ratios will be observed in all ten mass spectra. The error in the RIAs increases with decreasing mass for a given spectrum and becomes larger as the spectrum number rises for a given mass. Thus, neglecting to account for changes in the composition of the vapor within the reservoir inlet during acquisition of the mass spectra will introduce errors into a quantitative analysis provided that other experimental errors are less than the changes in composition on a percentage basis. Under this assumption, the variations in the mole ratios shown in Table 3 would, unless accounted for, negate the increase in accuracy obtained by averaging the corresponding molecular ion signal magnitudes in the ten scans of the mass spectrum of the mixture.

We now consider the effect of decreasing the cross-sectional area of the leak on the composition of the vapor in the reservoir during acquisition of a mass spectrum at 1000 s/decade rather than at 100 s/decade. For this scan rate, columns 5 and 6 of Table 4 give the moles of each of the seven components relative to the moles of $C_{42}H_{78}$ remaining in the inlet at the times for detection of the corresponding molecular ions calculated for cross-sectional areas of the leak of 2.076×10^{-5} and 7.378×10^{-6} cm^2, respectively. For the purposes of comparison, the values of the remaining parameters in these calculations were identical to those used in calculating the mole ratios in Tables 2 and 3. Comparison of the mole ratios in columns 5 and 6 of Table 4 reveals that a reduction of ca. 64% in the cross-sectional area of the leak significantly reduces the alteration in the composition of the vapor in the inlet during acquisition of the mass spectrum of the mixture. Such a reduction in

the cross-sectional area of the leak significantly reduces the contribution of the mass dependence of effusion across the leak to the error in analysis especially for the lower mass compounds. Since the times for acquiring one spectrum of the mixture at a magnetic field scan rate of 1000 s/decade and ten spectra of the mixture at a magnetic field scan rate of 100 s/decade are very similar, use of the leak with the smaller cross-sectional area would significantly reduce the variation in the relative mole ratios shown in Tables 2 and 3.

4.1.3 Processing of Mass Spectra

The time dependence of the ion signal magnitudes should obey Equation (6). Thus, the errors introduced into an analysis by changes in the composition of the vapor in the inlet during the time required for acquisition of a mass spectrum can be minimized by extrapolating the logarithm of the ion signal magnitudes to zero time. With respect to mass spectra acquired at 100 s/decade for our hypothetical mixture, the quantitative analysis would employ the seven average molecular ion signal magnitudes calculated from the ten zero-time signal magnitudes for each ion which were obtained by a least-squares fit of the individual signal magnitudes in each of the ten spectra to Equation (6). The times listed in column 3 of Table 4 reveal the impracticality of acquiring multiple scans of a mass spectrum at a magnetic field scan rate of 1000 s/decade if the components in a mixture cover a reasonable mass range. A possible approach to minimizing the error in such situations would be to calculate effusion constants for the various components in the mixture using Equation (7). These values and both the arrival times and signal magnitudes of the corresponding ions would be substituted into Equation (6) to arrive at estimated zero-time signal magnitudes for these ions.

An experimental approach to accounting for the mass dependence of effusive flow would be to acquire mass spectra under conditions of constant timing. The mass spectra of samples would be acquired under conditions such that the signal magnitudes of the various ions are recorded at the same time intervals following introduction of each sample into the reservoir inlet. The sensitivities for those ion m/z values to be used in the analyses are either determined at the same time intervals or calculated at these time intervals from values measured at other known time intervals. Consequently, the signal magnitudes of these ions in the spectra of both the mixtures of unknown composition and the calibrating substance or mixture would then be, within the limits of experimental accuracy, the same fractions of their zero-time signal magnitudes. Since the ion signal magnitudes obtained under these conditions are directly used in calculating quantitative distributions, time has been experimentally eliminated as an independent variable. This approach is especially useful in the analysis of mixtures of noncondensable gases.

The aforementioned approaches to accounting for the compositional dependence of ion signal magnitudes are extremely tedious and time-consuming. Thus, a decision regarding implementation of such data analyses should include a consideration of the uncertainties in the ion signal magnitudes and the availability of ionization sensitivities. The accuracy of the ion abundances reflects both the ion statistics and the variation in instrumental parameters governing the generation, extraction, transmission and detection of the ions. The factors which determine the ion statistics for the various m/z values include the amounts of each component charged, the

Table 4 Effect of the cross-sectional area of the leak on the moles of −6(H) series homologs relative to moles of the $C_{42}H_{78}$ −6(H) series homolog comprising a gaseous mixture in a reservoir inlet at 325 °C during acquisition of a 1000 s/decade mass spectrum

Components		Arrival Time for M^+ (s)[a]	Relative moles in reservoir		
Formula	Mass		Initial	at detection of M^+ for cross-sectional area $\times 10^{-5}$ cm^2 of	
				2.076[b]	0.7378[c]
$C_{42}H_{78}$	582.61	347.535	1	1	1
$C_{36}H_{66}$	498.516	415.234	1	0.992	0.997
$C_{30}H_{54}$	414.423	495.470	1	0.980	0.993
$C_{24}H_{42}$	330.329	593.967	1	0.964	0.987
$C_{18}H_{30}$	246.235	721.564	1	0.939	0.978
$C_{12}H_{18}$	162.141	1220.56	1	0.894	0.961
C_6H_6	78.047	1267.82	1	0.782	0.916

[a] Calculated for exponential scanning from m/z 650 to m/z 70 at 1000 s/decade; i.e. $k_s = 2.3026 \times 10^{-3}$ s^{-1}; 300 s elapsed time between opening leak-line valve and initiation of spectrum acquisition.
[b] Calculated for k_{eff} values in column 3 of Table 2.
[c] Calculate for $k_{eff} \times 10^{-5}$ values of 2.9238, 3.1607, 3.4666, 3.8829, 4.4973, 5.5422 and 7.9882, respectively.

cross-sections for their ionization, and the efficiencies of extraction of ions from the source and of the transmission of ions at a given RP. In this regard, the analysis of multiple scans of the mass spectra obtained from repeated injections of various samples of present interest suggests that a standard deviation of up to about 10% in the molecular-ion relative abundances for components accounting for reasonable weight percentages of the mixtures is not atypical. As discussed later, the situation with regard to the availability of ionization sensitivities and the standard deviations associated with these values is in general unsatisfactory. Based upon these considerations, the mathematical analysis of ion signal magnitudes to account for changes in vapor composition resulting from effusion across a leak is justified in order to obtain highly accurate analysis of mixtures of noncondensable gases. However, such analyses would not in general be warranted for mixtures of condensates given reservoir–inlet system parameters similar to those considered in the present discussion. The mole ratios in Tables 2–4 support this conclusion especially for compounds having masses in excess of ca. 150 amu. Depending upon the accuracy in the ion signal magnitudes, corrections may be justified for lower molecular weight compounds. Finally, the rates of effusion are directly proportional to the cross-sectional area of the leak divided by the volume of the inlet and exhibit a half-power dependence on the absolute temperature divided by the mass of each species. For exponential scanning of the magnetic field, the change in the composition of the vapor in the inlet and, hence, the variation in the relative ion signal magnitudes are exponentially dependent upon the effusion constants and the ion-arrival times. Thus, the effect of compositional changes on accuracy can be minimized by considering the nature of the samples to be analyzed and both the rates for acquiring mass spectra and the functional form of the scan law in developing specifications for a reservoir–inlet system.

4.2 Direct Introduction Probe

The technique of direct-probe introduction of samples into a mass spectrometer ion source has been used in the MS analysis of a variety of fossil fuel materials.[24,28,58,59,62,64,67–69,77,79,89,93,94,105] However, the majority of these studies involved the acquisition of either high-resolution mass spectra for the purpose of determining molecular formulas or low-resolution (unit mass resolved) mass spectra. In only several of these studies were the HRMS acquired both to identify and quantify the homologs present in samples.[62,79,93,94]

4.2.1 Status of Results Obtained Using the Technique

The objectives in using a direct probe rather then an AGHIS for sample introduction are twofold: first, to increase the analyzable molecular weight range because ion-source pressures are most likely to be less than the pressures in reservoir–inlet systems and, second, to reduce any occurrences of thermal decompostions by decreasing the residence time of the gaseous molecules in the system prior to their ionization. In the opinion of the author, the studies reported to date preclude a definite assessment of the extent to which the first objective is attained for various compound classes.

Ions having m/z values ranging from about 100 up to about 1900 were observed in low-resolution FI mass spectra of probe introduced samples obtained from application of molecular distillation and sequential elution fractionation to atmospheric residues.[105] Low resolution FI mass spectra containing ions with m/z values ranging up to about 900–1200 were been obtained from probe distillation of saturated hydrocarbon and aromatic neutral fractions from a petroleum distillate having 50% point AEBPs from about 1050 to 1250 °F (566 to 677 °C).[77] It is unfortunate that the mass spectra in these latter experiments were not also acquired at RPs in the range 10 000–50 000 because ion signal magnitudes, and hence, the analyzable molecular weight range decreases with increasing RP, with the decrease in RP exceeding linearity at RP values greater then 10 000.[83] In this regard, it should be noted that the author has routinely observed ions having low signal magnitudes above m/z 700 in the 40 000 RP LVEI mass spectra of gas–oil–aromatic–neutral fractions introduced into the ion source via a batch inlet at 300–325 °C.

Probe distillation LVEI mass spectra of an aromatic neutral fraction from the 1000–1250 °F (538–677 °C) AEBP distillate of a different crude that were acquired at RPs ranging from 3000–10 000 exhibited ranges of ion m/z values comparable to those that are observed using batch inlet introduction.[62] The same conclusion is indicated to account for probe distillation of the saturates and aromatic neutrals from the maltene fraction of a tar sand bitumen based upon the FI mass spectra and presumably the 70-eV electron ionization (EI) mass spectra obtained at 1000 and 10 000 RP, respectively.[67]

The results from analysis of the 25 000 RP LVEI mass spectra of molten caustic-treated coal extracts indicate that direct probe introduction may increase to some extent the accessible molecular weight range for highly polar compounds containing one or more of the elements N, O and S.[79] However, probe introduction resulted in only about 45% of the pyridine dialysate from one coal being vaporized. Finally, the carbon number ranges of the various hydrocarbon homologous series are consistent with those routinely introduced via a batch inlet.

The technique of probe microdistillation (PMD) was used in analyzing a vanadyl porphyrin fraction isolated from the >700 °C AEBP resid from Venezuelan crude

Cerro Negro. The HRLVEIMS of the volatiles were acquired at 10 000 RP.[94] The mass spectra exhibited ions having m/z values ranging from about 500 to about 850 with the maximum ion signal magnitudes in approximately half of the -28 through -50 $Z(N_4$ VO) series occurring at m/z values less then 700. Approximately two-thirds of the vanadium-containing compounds were found to be distillable.

In the opinion of the author, using the technique of probe introduction to extend the range of analyzable molecular weights of various compound classes significantly requires a detailed study using both a batch inlet and a direct probe to introduce fractions from separation of various fossil fuel materials with the mass spectra acquired over a range of RPs. As discussed below, the issues that attend the incorporation of the technique of probe distillation into HRMS homolog methods would appear to justify such a study.

4.2.2 *Fundamental Considerations in High-resolution Mass Spectrometric Applications*

Since probe distillation of the components of a mixture occurs over a range of temperatures, the pressure of a neutral species in the region of ionization and, hence, the m/z values and signal magnitudes of the corresponding molecular ions observed in mass spectra acquired during a distillation of a sample are temperature dependent. Consequently, quantitative MS employing probe introduction requires the total signal magnitude of the molecular ions detected from the volatilization/ionization of the components as the temperature T_i of the sample in degrees Kelvin is raised from its initial value, T_{int}, to its final value, T_{fin}. The total signal magnitude of the kth molecular ion resulting from volatilization/ionization of the kth component in a mixture is

$$SM(\text{Tot})_k = \int_{T_{\text{int}}}^{T_{\text{fin}}} \text{f}(SM(T)_{k,i})\,\text{d}T \tag{8}$$

One approach to evaluating the integral in Equation (8) is to acquire multiple mass spectra as the probe temperature is raised from its initial to its final value. The signal magnitudes of the corresponding ions in each of the acquired mass spectra are then summed to obtain a total spectrum for distillation of the sample. However, as discussed subsequently, this method for acquiring total molecular ion signal magnitudes is not applicable to the acquisition of HRMS.

When using the technique of probe distillation in performing HRMS-type analyses, Schmidt and Sprecher acquired 25 000 RP LVEI mass spectra of the components of coal extracts that were volatile at four different probe temperatures.[79] Each spectrum at a given temperature was treated as a single analysis and solutions to Equation (8) were obtained by summing the signal magnitudes for the corresponding molecular ions in the spectra acquired at a given probe temperature to obtain a composite spectrum for that temperature. The composite mass spectrum was used to calculate CNDs and other parameters. The fact that this approach entails the acquisition of mass spectra of compounds volatilized at only selected temperatures rather then over the entire temperature range raises a question regarding the extent to which the analytical results represent the actual composition of a sample.

Schronk et al.[62] used the method of PMD developed by Grigsby et al.[119] in the HRMS-type analysis of a monoaromatics fraction from a 1000–1250 °F (538–677 °C) AEBP petroleum distillate. Linear temperature programming at a rate of 7 °C min^{-1} resulted in a distillation time of approximately 28 min. Since 3000 and 10 000 RP mass spectra were acquired using a magnetic field scan rate of 100 s/decade, 25 mass spectra were acquired during a distillation. Variations in the efficiency of extracting ions from an ion source generally occurs during distillations under the condition of linear temperature programming. Since these variations would introduce significant errors into the quantitative results, the ion signal magnitudes are corrected for the occurrence of these variations by acquiring mass spectra containing the signal magnitudes of both the ions from distillation/ionization of a sample and ions $1, 2, 3, \ldots, j$ from ionization of a substance introduced via an expansion volume. The correction factor for the signal magnitudes in the ith mass spectrum acquired at T_i is, Equation (9)

$$CF_i = \frac{\sum_i \sum_j SM_j/N}{\sum_j SM_j} \tag{9}$$

The signal magnitude of the kth ion in the ith mass spectrum corrected for the variation in the ion source extraction efficiency is, Equation (10)

$$SM(\text{cor})_{i,k} = SM(\text{exp})_{i,k} CF_i \tag{10}$$

Equation (11) describes the functional form of $\text{f}(SM(T)_i)$ deduced from an empirical analysis of

$$\text{f}(SM(T)_i) = \frac{A_1 \exp(A_2/T_i)}{1 + \exp[A_3(T_i - A_4)]} \tag{11}$$

the temperature dependence of the ion signal magnitudes resulting from probe distillation of pure compounds.[119] Values of the parameters A_1 through A_4 are obtained from Equation (12) by nonlinear regression analysis

$$SM(\text{cor})_{i,k} = \frac{A_1 \exp(A_2/T_i)}{1 + \exp[A_3(T_i - A_4)]} \tag{12}$$

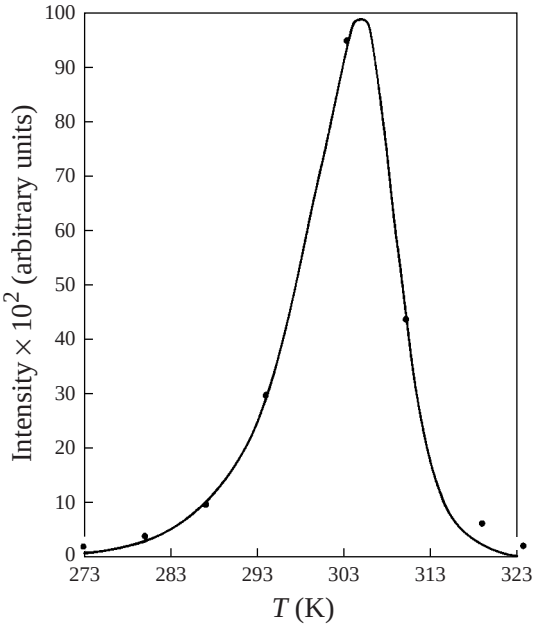

Figure 8 Elimination curve for PMD of 1-methylnaphthalene. Points correspond to experimental FI ion abundances for m/z 142 at eight temperatures. Curve is calculated from least-squares parameters.

Plots of corrected and least squares predicted signal magnitudes versus absolute temperature are termed elimination curves. As an example, the elimination curve obtained from PMD of 1-methylnaphthalene from a 19 component mixture is reproduced in Figure 8.[64] The dots represent the m/z 158 signal magnitude, temperature pairs. The solid curve corresponds to the least-squares fit of these pairs to Equation (12). Equation (12) is seen to describe accurately the dependence of the m/z 158 signal magnitude on temperature.

The degree of accuracy in CNDs obtained for samples introduced via a direct probe is an important matter that has been inadequately addressed. The results obtained from analyzing mixtures containing 19 and 3 components are, to the author's knowledge, the only ones that directly bear on this matter.[64] Table 5 tabulates the results obtained from the low resolution FI mass spectra of the 19-component mixture with samples introduced via a batch inlet and a direct probe. Column 3 of Table 5 lists the weight percentages of the 19 compounds calculated from weights taken. The results of the original FIMS analysis with sample introduction via a batch inlet are tabulated in columns 4 and 5. Columns 6 and 7 of Table 5 list the corresponding results obtained subsequent to storage of the mixture at 0 °C for two years. Comparison of the weight percents in column 6 with the corresponding values in columns 3 and 4, the corresponding percentage deviations in columns 5 and 7 and both the average errors and associated standard deviations and the root

Table 5 Results obtained from PMD/FIMS and FIMS of 19-component mixture

Mixture			Weight percents and percent deviations with sample introduction by Batch inlet				Direct probe	
Compound	MW	Wt. %	6/76		6/78		6/78	
			Wt. %	Dev (%)	Wt. %	Dev (%)	Wt. %	Dev (%)
Indan	118	8.37	8.29	−5.4	7.85	−6.2	7.40	−11.6
1,3,5-Trimethylbenzene	120	7.20	7.10	−1.4	6.58	−8.6	5.78	−19.7
Naphthalene	128	4.14	4.18	1.0	4.30	3.9	5.52	32.1
1,2,3,4-Tetrahydronaphthalene	132	8.16	7.96	−2.4	7.69	−5.8	9.49	16.3
Benzo[*b*]thiophene	134	4.47	4.68	4.7	5.05	13.0	5.96	33.3
1-Methylnaphthalene	142	10.78	11.03	2.3	11.10	3.0	10.61	−1.6
6-Methylbenzo[*b*]thiophene	148	4.82	5.06	5.0	5.50	14.1	5.07	5.2
Acenaphthene	154	2.84	2.86	0.7	2.84	0.0	2.75	−3.2
1,6-Dimethylnaphthalene	156	6.53	6.60	1.1	6.97	6.7	5.89	−9.8
2*a*,3,4,5-Tetrahydroacenaphthene	158	10.34	10.16	−1.7	9.63	−6.9	9.91	−4.2
Fluorene	166	2.68	2.74	2.2	2.81	4.8	2.09	−22.0
Dibenzofuran	168	4.02	4.43	10.2	4.55	13.2	4.44	10.4
Phenanthrene	178	2.98	2.84	−4.7	3.10	4.0	3.57	19.8
9,10-Dihydrophenanthrene	180	9.50	9.44	−0.6	9.55	0.5	9.90	4.2
Dibenzothiophene	184	1.74	1.65	−5.2	1.84	5.8	2.10	20.7
4-Methylphenanthrene	192	1.27	1.23	−3.2	1.14	−10.2	1.38	8.7
Fluoranthene	202	4.59	4.36	−5.0	5.12	11.6	4.54	−1.1
1,2,3,9,10,10*a*-Hexahydropyrene	208	4.22	3.99	−5.5	3.09	−26.8	3.33	−21.1
Chrysene	228	1.35	1.40	3.7	1.35	0.0	0.96	−28.9
Average error ± σ				−0.2 ± 4.3		−0.8 ± 10.1		1.4 ± 18.0
RMS error				±4.2		±9.8		±17.6

FIMS, field ionization mass spectrometric.

mean square (RMS) deviations for the two batch inlet experiments led us to conclude that the mixture was suitable for use in our development of a computer-based method for acquiring and processing probe temperatures and low-resolution mass spectral files.[64] However, the results from the second batch inlet analysis are consistent with the loss of small amounts of the more volatile components indan, 1,3,5-trimethylbenzene and 1,2,3,4-tetrahydronaphthalene during storage and/or handling of the mixture. Also, the variability of the results from this analysis is generally greater then that from the first batch inlet analysis.

Columns 8 and 9 of Table 5 list the weight percentage of each component calculated from the areas under the elimination curves and the percentage deviation in each of these values, respectively. Both the average and RMS errors and the standard deviation in the average error are seen in Table 5 to be larger for the PMD FIMS analysis then for the second batch inlet FIMS analysis. Examination of column 9 reveals that the errors in the PMD FIMS analysis were less then 10% for eight compounds, 10–20% for five compounds and between 20 and 34% for six compounds. The corresponding results from the second batch inlet analysis are 13 compounds, five compounds and one compound, respectively. Thus, at least in this instance, the imprecision in the quantitative results was greater using a direct probe than a batch inlet for sample introduction. The diminished accuracy of the results may be attributed in part to the loss of volatile components upon introduction of the probe into the system and the effect of the rate of temperature increase on the volatilization of these compounds in relation to the time for spectral acquisitions. The values of 33.1%, 22.6% and 44.4% obtained from PMD FIMS analysis of a mixture consisting of 33.2% phenanthrene, 25.1% fluoranthene and 41.7% hexahydropyrene afford a measure of support for this conclusion. These results indicate the need for additional studies using both synthetic blends and fossil-fuel samples with the analytical results for both kinds of samples obtained for both probe and batch inlet introductions in order to establish the ranges in both the accuracy and the precision that are obtainable using direct probe introduction.

Significant problems must be addressed and resolved in order to implement probe introduction HRMS homolog analysis in the molecular weight range were the method(s) would have the most significance, that is, above about mass 400 u for highly polar compounds such as acids and bases and above about 700 u for aromatic neutrals and saturates. In order to begin to define these problems, we consider the probe introduction of the −20(H), −18(H), −16(H) and −32(H) homologs $C_{114}H_{208}$, $C_{114}H_{210}$, $C_{114}H_{212}$ and $C_{115}H_{198}$ under conditions that simulate the presence of these homologs in an aromatic neutral concentrate containing homologs of chemical types ranging in mass from about 750 u to about 1800 u. That is, the amounts of the −20(H), −18(H) and −16(H) homologs to be distilled are equivalent to 0.5 wt% of a 1-mg sample of the aromatic neutral fraction or 3.167×10^{-9}, 3.163×10^{-9} and 3.159×10^{-9} moles; the corresponding values for the −32(H) homolog are 0.05% and 3.163×10^{-10} moles. The LVEI would initially be acquired at 20 000 RP in order to resolve $C_{114}H_{210}$, and $C_{115}H_{198}$, molecular ions having 10 to 1 RIA ratios; see section 6.2. We also assume that: (1) the LVEI mass spectra are characterized by the presence of molecular ions and the absence of fragment ions, (2) the mass spectra are acquired from m/z 1950 to m/z 750 using either a Kratos MS-50 mass spectrometer or an instrument that has a similar dependence of sensitivity on RP with either instrument being fitted with a high-field magnet to permit acquisition of HRLVEIMS at full accelerating voltage, and (3) the magnetic field strength is decayed exponentially at a rate of 300 s/decade with an interscan time of 4 s. The 2000 RP sensitivity for LVEI of phenanthrene is 4.4×10^{17} ions mol^{-1}. The ratio of the sensitivity at 20 000 RP to the sensitivity at 2000 RP is about 0.071. Making the usual assumption that LVEI sensitivities are carbon-number independent results in a 20 000 RP sensitivity for LVEI of $C_{114}H_{210}$ of 3.12×10^{16} ions mol^{-1} and mole sensitivities for −20(H), −32(H) and −16(H) homologs relative to the mole sensitivity for −18(H) homologs of 1.24, 1.84 and 0.67, respectively; see section 7.7. The distributions of $C_{114}H_{208}$, $C_{114}H_{210}$, $C_{114}H_{212}$, and $C_{115}H_{198}$ molecules and, hence, molecular ions containing carbon-13 but not deuterium were calculated as described in section 7.1. Molecules containing from zero through six atoms of carbon-13 were found to account for 99.97% of the −20(H), −32(H), −18(H) and −16(H) neutral and ionic species.

Consider two ions having m/z values such that $(m/z)_2 > (m/z)_1$. For magnetic focusing instruments

$$(m/z)_2 = (m/z)_1(1 + \alpha) \tag{13}$$

The scan law for exponential decay of the magnetic field strength is

$$\ln \frac{(m/z)_2}{(m/z)_1} = -kt \tag{14}$$

where k is the scan law constant in reciprocal seconds and $t_0 = 0$. Combining Equations (13) and (14) and noting that $\alpha = 1/RP$, we obtain Equation (15) for the peak width in seconds as a function of the RP and the rate constant for exponential decay

$$pw_{RP} = \frac{1}{k} \ln \left(\frac{RP}{RP + 1} \right. \tag{15}$$

Equation (14) can be used to obtain a value of k given the time it takes to scan a decade of m/z values, that is, for $t = 300$ s/decade, $k = 7.6753 \times 10^{-3}\ s^{-1}$. Although the actual scan law is not ideal, the author has routinely realized good agreement between theoretically predicted and experimental peak widths at 5% of the peak heights provided that the Kratos MS-50 is set up so that the RP is essentially constant over the range of m/z values being scanned.

Following Schmidt and Sprecher,[79] we assume that probe introduction at temperature T resulted in a constant beam current for 10 min. Since the time for acquiring a spectrum is calculated to be 124.5 s and the interscan time is 4 s, four 20 000-RP LVEI/MS will be acquired during this time interval. For LVEI of each of the $j = 1, 2, \ldots, 7$ isotopically substituted $-20(H)_{i=1}$, $-32(H)_{i=2}$, $-18(H)_{i=3}$, and $-16(H)_{i=4}$ molecules, Equations (16) and (17) were solved to obtain values for the predicted total number of ions and the experimental total number of ions that include estimates of the effects of ion statistics, respectively.

$$N_{i,j} = NOS \times \text{floor}\ \frac{1}{2}\frac{\text{moles}[-Z(H)_i]}{DT}\frac{PA_{i,j}}{100} \times S(\text{mol})_{i=3} RS(\text{mol})_i pw_{RP} \quad (16)$$

$$N\exp_{i,j} = N_{i,j} + \text{floor}(r_{i,j}\sqrt{N_{i,j}}) \quad (17)$$

In Equation (16), NOS is the number of acquired mass spectra, $PA_{i,j}$ are the percentage abundances of the seven isotopically substituted molecules/molecular ions, $S(\text{mol})_{i=3}$ is the mole sensitivity for LVEI of $C_{114}H_{210}$ at 20 000 RP, $RS(\text{mol})_i$ are the relative mole sensitivities for LVEI of $C_{114}H_{208}$, $C_{115}H_{198}$, $C_{114}H_{210}$ and $C_{114}H_{212}$ and DT is the distillation time. In Equation (17), $r_{i,j}$ are sets of random numbers having normal distributions with $\mu = 0$ and $\sigma = 1$. Estimates of the effects of ion statistics on the ion yields were based on the ion sums both for the sake of simplicity and because essentially the same result was obtained from summing the estimates of ion statistics on the ion yields in each of the four spectra for two of the molecules.

The specific Z values, the formulas, and the mass numbers of the various isotopically substituted molecules/molecular ions are listed in columns 2, 3 and 4, respectively, in Table 6. Columns 5 and 6 of Table 6 contain the percentage abundances of $-20(H)$, $-18(H)$ and $-16(H)$ molecules/molecular ions containing 0 through 6 atoms of carbon-13 and $-32(H)$ molecules/molecular ions containing 0 through 5 atoms of carbon-13, respectively. The mass-to-charge ratio and the predicted ion total for each molecular ion are given in columns 7 and 8, respectively. RPs in column 10 were calculated assuming unit RIAs for successive ion beams whereas those in column 11 were calculated for the RIAs in column 9; see section 6.2.

Comparison of corresponding RPs in columns 10 and 11 in Table 6 shows that the relative abundance of successive ion beams can significantly increase the RP required to obtain resolution of successive ion beams. The resolution of a number of successive ions such as 4 and 5, 7 and 8 and 10 and 11 is seen in columns 10 and 11 greatly to exceed the capabilities of state-of-the-art DF mass spectrometers. In this regard, the resolution of $-20(O)$ molecular ions $C_{113}H_{206}O^{+\bullet}$ from $-18(H)$ molecular ions having relative abundances ranging from 1 to 10 would require RPs ranging from about 43 400–50 600. These results together with the discussion in section 6 demonstrate that resolution of many isobaric ions in the "high mass region" will simply not be possible using DF instruments.

Table 7 list the results obtained from processing the composite "experimental" 20 000-RP LVEI/MS from probe distillation of $C_{114}H_{208}$, $C_{115}H_{198}$, $C_{114}H_{210}$ and $C_{114}H_{212}$ at temperature T. The specific Z values and the formulas of the various isotopically substituted molecular ions are listed in columns 2 and 3, respectively. Columns 4 and 5 contain normalized percentage abundances of $-20(H)$, $-18(H)$ and $-16(H)$ molecular ions and $-32(H)$ molecular ions, respectively, calculated from the percentage abundances in Table 6. The m/z values and the predicted ion abundances in Table 6 are shown in columns 6 and 7, respectively. The m/z values and the abundances of the ions in the composite "experimental" spectrum are tabulated in columns 8 and 9, respectively. The following procedure was used to obtain experimental m/z values and abundances for unresolved peaks. Equations (18) and (19) were solved to obtain a m/z value and an experimental ion abundance. Solutions to Equation (19) and the known m/z values for resolved peaks were

$$\text{ions} = \sum_i N_{i,j} \quad (18)$$

$$m/z = \frac{\sum N_{i,j} \times m/z_{i,j}}{\text{ions}} \quad (19)$$

subject to a 5-ppm error multiplied by random numbers having normal distributions with $\mu = 0$ and $\sigma = 1$. The entries in columns 10–13 are the results obtained from deisotoping/formula assignment of the m/z signal magnitude pairs using Equations (18) and (19) as described in Gaun et al.[92]

The mass measurement errors in column 12 are seen to be typical of the values encountered in processing HRLVEIMS in the "conventional" mass range. The

Table 6 Quantities relevant to predicting mass spectra for direct probe introduction of $(CH_2)_{114}H_{-20}$, $(CH_2)_{114}H_{-18}$, $(CH_2)_{114}H_{-16}$ and $(CH_2)_{115}H_{-32}$ at temperature T

Molecules/Molecular ions						m/z of molecular ion[a]	$N_{i,j}^{b}$	RIA[c,d] $r \geq 1$	RP	
No.	$-Z$	Formula	Mass no.	% for $n =$ 114	% for $n =$ 115				$r = 1^{d}$	$r \geq 1^{d}$
1	20	$C_{114}H_{208}$	1576	27.98		1577.6276	621			
								1.28	1572	1604
2	20	$^{13}CC_{113}H_{208}$	1577	35.84		1578.6310	795			
								8.74	1719	1988
3	32	$C_{115}H_{198}$	1578		27.67	1579.5494	91			
								5.54	18 605	20 963
4	20	$^{13}C_2C_{112}H_{208}$	1578	22.75		1579.6343	504			
								1.01	175 515	175 632
5	18	$C_{114}H_{210}$	1578			1579.6433	500			
								4.27	1737	1927
6	32	$^{13}CC_{114}H_{198}$	1579		35.75	1580.5527	117			
								1.80	18 595	19 469
7	20	$^{13}C_3C_{111}H_{208}$	1579	9.54		1580.6377	211			
								3.03	177 600	192 753
8	18	$^{13}CC_{113}H_{210}$	1579			1580.6466	640			
								8.53	1738	2007
9	32	$^{13}C_2C_{113}H_{198}$	1580		22.89	1581.5561	75			
								1.15	18 628	18 848
10	20	$^{13}C_4C_{110}H_{208}$	1580	2.97		1581.6410	65			
								6.25	395 410	448 684
11	18	$^{13}C_2C_{112}H_{210}$	1580			1581.6450	406			
								1.22	113 787	115 612
12	16	$C_{114}H_{212}$	1580			1581.6589	334			
								10.77	1756	2054
13	32	$^{13}C_3C_{112}H_{198}$	1581		9.69	1582.5594	31			
								1.94	18 618	19 595
14	20	$^{13}C_5C_{109}H_{208}$	1581	0.74		1582.6444	16			
								10.63	177 825	207 832
15	18	$^{13}C_3C_{111}H_{210}$	1581			1582.6533	170			
								2.52	175 850	188 492
16	16	$^{13}CC_{113}H_{212}$	1581			1582.6623	428			
								42.8	1758	2198
17	32	$^{13}C_4C_{111}H_{198}$	1582		3.05	1583.5628	10			
								3.33	18 652	20 368
18	20	$^{13}C_6C_{108}H_{208}$	1582	0.15		1583.6477	3			
								17.67	175 961	211 109
19	18	$^{13}C_4C_{110}H_{210}$	1582			1583.6567	53			
								5.13	177 939	199 591
20	16	$^{13}C_2C_{112}H_{212}$	1582			1583.6656	272			
								136.0	1759	2308
21	32	$^{13}C_5C_{110}H_{198}$	1583		0.76	1584.5661	2			
								6.50	16 875	19 193
22	18	$^{13}C_5C_{109}H_{210}$	1583			1584.6600	13			
								8.77	176 073	203 656
23	16	$^{13}C_3C_{111}H_{212}$	1583			1584.6690	114			
								57.0	1594	2018
24	18	$^{13}C_6C_{108}H_{210}$	1584			1585.6634	2			
								17.5	178 164	213 652
25	16	$^{13}C_4C_{110}H_{212}$	1584			1585.6723	35			
								4.38	1580	1756
26	16	$^{13}C_5C_{109}H_{212}$	1585			1586.6757	8			
								8.0	1581	1820
27	16	$^{13}C_6C_{110}H_{212}$	1586			1587.6790	1			

[a] Calculated using 13.003 3548 and 1.007 8250 as the atomic weights for carbon-13 and hydrogen, respectively.
[b] Solutions to Equation (16).
[c] Ratios of successive $N_{i,j}$ values calculated such that the ratio is equal to or greater than 1.
[d] See section 6.2.

Table 7 Processing of mass spectra acquired from direct probe introduction of $(CH_2)_{114}H_{-20}$, $(CH_2)_{114}H_{-18}$, $(CH_2)_{114}H_{-16}$, and $(CH_2)_{115}H_{-32}$ at temperature T

No.	Molecular ions				Mass spectra							
	$-Z$(H)	Formula	Relative % for $n =$		Predicted molecular ion		Experimental molecular ion		Processed molecular ion			
			114	115	m/z	$N_{i,j}$	m/z	$N\text{exp}_{i,j}$	Ions	m/z	$\delta\, m/z$ (ppm)[a]	% δ Ions[b]
1	20	$C_{114}H_{208}$	1.0000		1577.6276	621	1577.6242	604	604	1577.6242	2.2	−2.7
2	20	$^{13}CC_{113}H_{208}$	1.2808		1578.6310	795	1578.6382	764	764	1578.6382	−4.6	−3.9
3	32	$C_{115}H_{198}$		1.0000	1579.5494	91	1579.5571	87	87	1579.5571	−4.9	−4.4
4	20	$^{13}C_2C_{112}H_{208}$	0.8130		1579.6343	504			491	1579.6343		−2.6
							1579.6415	1061				
5	18	$C_{114}H_{210}$			1579.6433	500			570	1579.6478	−2.9	14.0
6	32	$^{13}CC_{114}H_{198}$		1.2920	1580.5527	117	1580.5549	119	119	1580.5549	−1.4	1.7
7	20	$^{13}C_3C_{111}H_{208}$	0.3410		1580.6377	211			205	1580.6377		−2.8
							1580.6507	832				
8	18	$^{13}CC_{113}H_{210}$			1580.6466	640			627	1580.6550	−5.3	−2.0
9	32	$^{13}C_2C_{113}H_{198}$		0.3501	1581.5561	75	1581.5553	68	68	1581.5553	0.51	−9.3
10	20	$^{13}C_4C_{110}H_{208}$	0.1063		1581.6410	65			64	1581.6410		−1.5
11	18	$^{13}C_2C_{112}H_{210}$			1581.6450	406	1581.6526	808	463	1581.6450		14.0
12	16	$C_{114}H_{212}$			1581.6589	334			281	1581.6595	−0.4	−15.9
13	32	$^{13}C_3C_{112}H_{198}$		0.1101	1582.5594	31	1582.5612	28	28	1582.5612	−1.1	−9.7
14	20	$^{13}C_5C_{109}H_{208}$	0.0263		1582.6444	16			15	1582.6444		−6.3
15	18	$^{13}C_3C_{111}H_{210}$			1582.6533	170	1582.6600	631	194	1582.6533		14.1
16	16	$^{13}CC_{113}H_{212}$			1582.6623	428			422	1582.6636	−0.9	−1.4
17	32	$^{13}C_4C_{111}H_{198}$		0.0275	1583.5628	10	1583.5750	3	3	1583.5750	−7.7	−70.0
18	20	$^{13}C_6C_{108}H_{208}$	0.0054		1583.6477	3			3	1583.6477		0.0
19	18	$^{13}C_4C_{110}H_{210}$			1583.6567	53	1583.6568	309	60	1583.6567		13.2
20	16	$^{13}C_2C_{112}H_{212}$			1583.6656	272			246	1583.6570	5.5	−9.6
21	32	$^{13}C_5C_{110}H_{198}$		0.0057	1584.5661	2	1584.5703	1	1	1584.5703	−2.61	−50.0
22	18	$^{13}C_5C_{109}H_{210}$			1584.6600	13			14	1584.6600	<0.1	7.7
							1584.6692	111				
23	16	$^{13}C_3C_{111}H_{212}$			1584.6690	114			97	1584.6706	−1.0	−14.9
24	18	$^{13}C_6C_{108}H_{210}$			1585.6634	2			3	1585.6634	0	50.0
							1585.6826	38				
25	16	$^{13}C_4C_{110}H_{212}$			1585.6723	35			35	1585.6842	−7.5	0.0
26	16	$^{13}C_5C_{109}H_{212}$			1586.6757	8	1586.6830	8	8	1586.6830	−4.6	0.0
27	16	$^{13}C_6C_{110}H_{212}$			1587.6790	1	1587.6823	2	2	1587.6823	−2.1	100.0

[a] Mass-to-charge measurement error in parts per million, i.e. values in column 6 minus values in column 11 divided by values in column 11 multiplied by 10^6.
[b] Percentage differences between predicted numbers of ions in column 7 and the numbers of ions after deisotoping spectra in column 10.

percentage deviations in the ion abundances in column 13 in relation to the ion abundances in column 9 are similar to the repeatability that can be attained in ion abundances calculated from the signal magnitudes in HRLVEIMS in the ''conventional'' mass range. Subject to a number of considerations, Tables 6 and 7 indicate the desirability of performing well-conceived experiments to assess the feasibility of using probe distillation to extend homolog-type analysis to aromatic hydrocarbon fractions having AEBPs up to about 1400 °F (760 °C). The assumed absence of fragment ions is most likely to be invalid for LVEI. The presence of significant numbers of fragment ions relative to molecular ions would undoubedtly complicate spectral interpretations. For example, hydrogen atom loss from $C_{114}H_{210}{}^+$ would result in the formation of $C_{114}H_{209}{}^+$ with a mass that differs from the mass of the −20(H) ion $^{13}CC_{113}H_{208}{}^{+\bullet}$ by only 4.5 mmu (millimass units). Since deisotoping of the experimental ion abundances will be necessary especially with DF instruments, decisions regarding the composition/origin of the ion masses in a sequence will be required. For example, for Table 7 a decision would have to be made regarding the value of a residual abundance for ion 2 that would be attributed to the presence of $C_{114}H_{209}$ fragment ions rather than to ion statistics and/or variations in the abundance of carbon-13/deuterium. In addition, the loss of two hydrogens or H_2 from $C_{114}H_{210}{}^+$ would result in the formation of $C_{114}H_{208}{}^{+\bullet}$. Thus, a decision would be required

about the extent to which the signal magnitude of the mass 1577.6276 ion reflects the presence of a −20(H) homolog in the sample. This extension of homolog methods would clearly benefit from the implementation/development of ionization techniques resulting in mass spectra characterized by the presence of molecular ions, the essential absence of fragment ions and by the development of separation techniques producing well-defined ring number fractions. Also, FTICR MS could be an important technique in method development given the RPs that have been achieved to date in the LVEI mass spectra of lower boiling petroleum distillates.[91,92,97] Finally, the procedure to be used in calculating CNDs must be addressed. The usual practice in the ''conventional'' mass range is to base the calculation of these distributions on the signal magnitudes of the isotopically most-abundant molecular ions. However, as seen in Tables 6 and 7 the dominance of a distribution of molecular ion signal magnitudes by a single isotopic molecular ion decreases with increasing mass-to-charge ratio.

5 IONIZATION TECHNIQUES: APPLICATIONS

5.1 Low-voltage Electron Ionization

For various reasons LVEI is the principal ionization technique used in the HRMS method for determining group types. (See refs. 6, 8, 10, 11, 13, 15–17, 19–24, 26, 31–35, 37–39, 41, 44, 47, 48, 52, 53, 61, 62, 71, 72, 75, 79, 82, 85, 86, 88, 91–94, 97)

A number of commercial ion sources can be operated at low voltage without significant if any modification in a well maintained and optimally tuned DF mass spectrometer such as the Kratos MS-50 to generate ion beams that are intense enough for the routine acquisition of mass spectra of distillates at RPs ranging up to about 50 000. A number analysts including the author have acquired HRLVEIMS at RPs ranging up to 80 000–100 000.[41] The technique is applicable to neutral, acidic and basic aromatic compound types. However, aromatic carboxylic acids are undoubedtly an exception. For example, the author has normally observed significant numbers of molecular ions phenomenologically corresponding to the ionization of homolgs of various aromatic hydrocarbon types in the HRLVEIMS of acid fractions which are known from ^{13}C NMR (nuclear magnetic resonance) and IR (infrared) spectroscopy to contain carboxylic acids. The LVEI technique is not applicable to saturate fractions.

5.2 Field Ionization

The technique of FI has been shown to be applicable to the mass spectrometric analysis of distillates boiling up to 400 °F (204 °C) and saturate, aromatic neutral, acidic and basic fractions from higher boiling distillates. However, FIMS has not enjoyed wide acceptance for a number of reasons. FI sources are not commercially available for all mass spectrometers. Achieving optimum performance of the source on a routine basis is often tedious and difficult. The majority of FI applications have involved the acquisition of low-resolution spectra.[25,29,36,42,43,46,49,51,54,58,59,64,67–69,74,76,77,89,95,105] With this in mind, it should be noted that unit resolution FI mass spectra suffice for the determination of CNDs for saturate fractions. The technique has found few applications in HRMS homolog analysis. Acquisition of mass spectra with adequate ion abundances using a DF mass spectrometer equipped with an FI source is limited under the best of conditions to RPs of less then about 10 000. One approach to using FIMS to determine CNDs involved a convoluted process.[60,61] FIMS were data system acquired at about 2000 RP and oscillographically recorded at 7500–10 000-RP. The heights of resolved and partially resolved peaks in the latter spectra were hand measured and the nominal m/z values of these ions were interpreted in terms of the formulas assigned to ion m/z values observed in the 20 000 or greater RP 70-eV spectra recorded on photoplates. These results were then used to factor the ion abundances in the 2000 RP FIMS into signal magnitudes for the various homologous molecular ions. Using a Kratos MS-50 equipped with a combination FI/EI ion source, Parisi and Scheppele determined CND for low-boiling distillates from FIMS data system acquired at about 7500 RP.[120] The known masses of +2(H) ions were used to convert the time centroids of the peaks in the spectra to masses.

5.3 Other Ionization Techniques

The techniques of chemical ionization (CI), charge exchange (CE) and fast atom bombardment (FAB) have received consideration as alternative ionization techniques in type analysis. The results as described below indicate the possibility of using one or more of these techniques to circumvent a number of existing problems in HRMS homolog-type analysis.

Veloski et al. reported the acquisition of mass spectra of a coal derived distillate and a petroleum-derived residuum obtained at a static RP of 20 000 with ion formation accomplished using low-pressure ammonia CI and LVEI.[93] The ammonia CI results are especially significant because of the problem, as discussed by both the authors and subsequently in sections 6.3 and 6.4, of resolving ions differing in composition by ^{13}CH/N using DF mass spectrometers. As demonstrated by the CNDs obtained from the chemical ionization high resolution mass spectra (CIHRMS) and the HRLVEIMS for the unfractionated

coal-derived distillate, this resolution problem effectively precludes, in general, the determination of low-levels of $Z(\mathrm{N\ O}_b\ \mathrm{S}_c)$ homologs in samples dominated by $Z(\mathrm{H\ O}_b\ \mathrm{S}_c)$ homologs using LVEI as the ionizing agent.

Allgood et al. found that charge-exchange reactions between $C_6H_6^{\bullet+}$ and series of olefins and alkylbenzenes under CI conditions resulted in mass spectra containing essentially only molecular ions.[81] The dependence of the relative molar sensitivities on molecular structure was found to be less then 10–15%. As expected, CE reactions between $C_6H_6^{\bullet+}$ and alkanes and cycloalkanes were not observed allowing homologs of olefinic and aromatic compound types to be differentiated from homologs of these compound types. Hsu and Qian investigated CE reactions between $CS_2^{\bullet+}$ and a series of aromatic hydrocarbons.[90] Electron beam energies that produced m/z 106 to m/z 91 ratios > 50 : 1 gave CS_2 CE conditions that resulted in maximum ion signal magnitudes and minimum formation of fragment ions. However, the molecular structure dependence of sensitivity is greater for CS_2 CE than for C_6H_6 CE because the increase in the alkyl carbon number from 6 to 13 in the −6(H) series was accompanied by an approximately 67% increase in the CS_2 CE sensitivity. There are indications that an increase in the number of condensed aromatic rings results in a substantial increase in the sensitivity for CS_2 CE of aromatic hydrocarbons, for example, the sensitivity for benzo[*a*]pyrene is about 3.3 times greater then the sensitivity for naphthalene.

Dzidic et al. developed and applied a GC low-resolution mass spectrometric method for hydrocarbon-type analysis of 350–850 °F (177–455 °C) distillates using Townsend discharge nitric oxide chemical ionization (TDNOCI).[87] The NO^+ produced under TDNOCI conditions distinguishes between homologs of hydrocarbon types by the occurrence of: (1) CE reactions with aromatics, (2) hydride extraction reactions with saturates and (3) addition reactions as well as both CE and hydride extraction reactions with olefins.

Dzidic investigated fluoride ion CI low-resolution MS with sample introductions using a direct probe as a method for determining naphthenic acids.[78] Unfortunately the use of a quadrupole instrument precludes the resolution of isobaric carboxylate anions and, hence, the determination of CNDs. Fan reported on the characterization of naphthenic acids using negative ion FAB MS.[84] The 2000 RP FAB spectra obtained from Xe bombardment of acid–triethanolamine matrices consisted of only $(M-H)^-$ ions. In contrast to this result, Grigsby et al. found that the positive ion FAB spectra of nitrogenous compounds and a base fraction from anthracene oil exhibited significant $M^{\bullet+}$, $(M+H)^+$, $(M-H)^+$ and fragment ions.[63]

6 QUALITATIVE ANALYSIS

6.1 Processing of High-resolution Mass Spectra

In general, an experimental high-resolution mass spectrum consists of pairs of data points consisting of peak positions and the sum of the signal magnitudes of the detected ions defining the peaks. The phrase "in general" is used because at least two commercial data acquisition systems afford the capability of storing "raw" data points, that is, the digitized signals from the mass spectrometer and the precisely defined times at which the signals were acquired. Peak positions in spectra acquired using a DF mass spectrometer commonly are the times for the peak centroids calculated from the data points defining the peaks. Peak positions in spectra acquired using a FTICR mass spectrometer are the frequencies of the peak centroids calculated from the frequencies across the peak obtained by fast Fourier transformation of the time-dependent signal from the FTICR instrument. The peak positions expressed as either times or frequencies must be converted to masses.

The time centroids of the peaks in spectra acquired exponentially using DF instruments are converted to highly accurate masses using a polynomial of the form,[66] Equation (20)

$$\ln(m/z) = A + Bt + Ct^2 + Dt^3 \tag{20}$$

The coefficients A–D in Equation (20) are determined by least squares analysis of the time centroids of known masses of five successive ions in a spectrum. A polynomial equation such as Equation (20) is used in place of Equation (14) because, as mentioned in section 4.2.2, Equation (14) is an inexact description of the rate of decrease of the magnetic field strength.

The frequency centroids of the peaks in spectra acquired using a FTICR mass spectrometer can be converted to masses using the relationship in Equation (21)

$$m/z = \frac{a}{\nu} + \frac{b}{\nu^2} \tag{21}$$

which is exact in the limit of a pure quadrupolar electrostatic trapping potential.[92,97] Values of the parameters a and b are determined by least squares analysis of the frequency centroids of known masses of three successive ions in a spectrum.

Concurrent acquisition of the spectrum of both know compounds and the sample being analyzed is one approach to establishing values of the masses and time centroids for determining the parameters in Equations (20) and (21). The mass defects of the reference ions must be considerably larger then those for the ions of unknown composition. Unfortunately, the traditional reference standards used with 70-eV electrons,

such as perfluorokerosenes, are not suitable for use with low-voltage electrons. Thus, mixtures consisting primarily of halogenated compounds have been used as mass calibrants.[33,75,80] However, several problems attend the extensive use of chlorinated and/or brominated aromatic compounds. First, these compounds by their nature contain isotopically substituted molecules that result in the formation of isotopically substituted ions that are not germane to the calibration process. The presence of these molecules restricts the amount of sample that can be charged because the pressure range over which an ion source operates efficiently is necessarily restricted. Second, it has not been possible to achieve an optimum mass difference of approximately 14 u between successive molecular ions especially in the mass region above 300 u. Third, the insolubility of a number of these halogenated compounds complicates the preparation of reference blends. The compounds in Table 8 were adopted as calibrants at the Amoco Research Center in order to overcome several of these problems. It should be noted that solutions can be prepared that contain 26 of the 31 compounds. However, the issue of the lack of mass calibration standards above mass 630 remains to be resolved.

The use of ions formed from compounds in a sample either alone or in conjunction with ions formed from reference compounds is another approach to mass calibration. One caveat in the use of this approach is that the time centroids of the peaks used as reference standards must not have derived significant contributions from unresolved ions.

Schmidt and Sprecher used the mass measurement accuracy obtained from assigning formulas to mass converted peak centroids as a criterion for selecting sample peaks to be used in conjunction with reference peaks.[79] The identification of a number of peaks resulting from the detection of $Z(\mathrm{H})$ molecular ions containing carbon-13 rather then $Z(\mathrm{NO_2})$ molecular ions is one of several results indicating that valid results were obtained in this instance. However, the ubiquitous problem of resolving ions that differ in composition by C_3/SH_4 using DF mass spectrometers would in general preclude the use of sample peaks in mass converting time centroids in mass spectra of samples such as aromatic neutral, acidic and basic fractions that contain significant amounts of sulfur compounds. For example, the $-6(\mathrm{H})$ series is the only homologous series of molecular ion peaks in the HRLVEIMS of aromatic neutrals that does not a priori reflect the detection of ions differing in composition by C_3/SH_4. Hsu et al. have approached the problem by performing ring-number separations of aromatic neutral fractions prior to obtaining HRLVEIMS.[86] At least in principle the ring number separation should significantly mitigate the C_3/SH_4 resolution problem. For example, HPLC separation should place the $-10(\mathrm{S})$ series in the diaromatic fraction and the isobaric $-20(\mathrm{H})$ series in the three-ring fraction. It should be noted when using this approach that the resolution in the separation decreases with increasing boiling point and can exhibit a dependency on the origin and history of the aromatic neutral fraction.

Table 8 Compounds for calibrating mass axis in HRLVEIMS

Compound[a]	Formula[b]	Mass
Hexabromobiphenyl[c]	$C_{12}H_4{}^{79}Br_3{}^{79}Br_3$	627.535 177
Tetraiodothiophene[c]	C_4SI_4	587.589 965
Pentyl 2,3,5-triiodobenzoate	$C_{12}H_{13}O_2I_3$	569.804 975
Butyl 2,3,5-triiodobenzoate	$C_{11}H_{11}O_2I_3$	555.789 325
Propyl 2,3,5-triiodobenzoate	$C_{10}H_9O_2I_3$	541.773 675
Ethyl 2,3,5-triiodobenzoate	$C_9H_7O_2I_3$	527.758 025
Methyl 2,3,5-triiodobenzoate	$C_8H_5O_2I_3$	513.742 375
2,3,5-Triiodobenzoic acid[d]	$C_7H_3O_2I_3$	499.726 725
Triiodoanisole[c]	$C_7H_5I_3$	485.747 460
Octafluorodibromobiphenyl	$C_{12}F_8{}^{79}Br^{81}Br$	455.821 851
Di(pentafluorophenyl)phenyl phosphine (Ultramark 443)	$C_{18}H_5F_{10}P$	441.996 919
Diiodofluorene[c]	$C_{13}H_8I_2$	417.871 547
Diiodotetrafluorobenzene	$C_6F_4I_2$	401.802 560
Diiodonaphthalene[c]	$C_{10}H_6I_2$	379.855 897
Decafluorobenzophenone	$C_{13}F_{10}O$	361.978 947
Perfluorobiphenyl	$C_{12}F_{10}$	333.984 032
Dibromotetrafluorobenzene	$C_6F_4{}^{79}Br^{81}Br$	307.828 239
Iodopentafluorobenzene	C_6F_5I	293.896 490
Octafluoronaphthalene	$C_{10}F_8$	271.987 226
Iodonaphthalene	$C_{10}H_7I$	253.959 249
Fluoroiodotoluene[e]	C_7H_6FI	235.949 827
Iodofluorobenzene[e]	C_6H_4FI	221.934 177
Iodobenzene	C_6H_5I	203.943 599
Hexafluorobenzene	C_6F_6	185.990 419
Pentafluorobenzene	C_6HF_5	167.999 841
Tetrafluorobenzene[e]	$C_6H_2F_4$	150.009 263
Trifluorobenzene[e]	$C_6H_3F_3$	132.018 685
Difluorobenzene[e]	$C_6H_4F_2$	114.028 107
Fluorobenzene	C_6H_5F	96.037 528
Benzene-d_6	C_6D_6	84.084 611
Dimethylformamide	C_3H_7NO	73.052 764

[a] Compounds in solution unless otherwise noted.
[b] C = carbon-12.
[c] Compounds in solid mixture.
[d] Fragment ion.
[e] Isomer not critical.

6.2 Resolving Power: Definitions

RP for magnetic deflection mass spectrometers is conventionally defined as the peak separation $(m/z)_{i+1} - (m/z)_i$ corresponding to a 10% valley between two equal-magnitude peaks resulting from detection of successive ion beams having mass-to-charge ratios $(m/z)_i$ and $(m/z)_{i+1}$:

$$RP(\mathrm{DF/MS}) = \frac{(m/z)_i}{(m/z)_{i+1} - (m/z)_i} = \frac{(m/z)_i}{\Delta(m/z)} \quad (22)$$

RP for ions of mass-to-charge ratio $(m/z)_i$ is conventionally defined in FTICR MS as

$$RP(\text{FTICR/MS}) = \frac{(m/z)_i}{\Delta(m/z)} \tag{23}$$

where $\Delta(m/z)$ is the full magnitude-mode mass spectral peak width at half-maximum peak height for an isolated peak.

Both definitions depend on peak shape. However, it is important to note that neither definition considers the dependence of peak separation and, hence, RP on the ratio of the signal magnitudes of adjacent peaks. Unfortunately, equal magnitude peaks are clearly the exception rather then the rule. Thus, the RP required to observe resolution of the majority of successive ion beams in the HRMS spectra of most samples exceeds values obtained from Equations (22) and (23) with the RP required to resolve successive ion beams increasing without limit as the ratio of the larger to smaller signal magnitude continues to increase. Therefore, let r be the ratio of the signal magnitudes SM_i and SM_{i+1} of successive beams calculated such that $r \geq 1$, i.e $r = 1$ if $SM_i = SM_{i+1}$, $r = SM_i/SM_{i+1}$ if $SM_i > SM_{i+1}$ and $r = SM_{i+i}/SM_i$ if $SM_{i+1} > SM_i$. Then for mass spectral peak shapes adequately described as Gaussian the dependence of RP on $(m/z)_i$, $\Delta(m/z)$ and r can be shown to be[121]

$$RP = \frac{(m/z)_i}{2\Delta(m/z)} \quad 1 + \sqrt{\frac{\ln(20r)}{\ln 20}} \tag{24}$$

In Equation (24), it is understood that the peaks resulting from detection of successive ion beams are by definition resolved if the valley height is 10% of either peak height if $r = 1$ and if the valley height is 10% of the lower magnitude peak if $r > 1$. Alternatively, ion beams i and $i + 1$ are said to be resolved if the signal magnitudes at the intersection point of the two peak profiles are 5% of either peak height if $r = 1$ or 5% of the lower magnitude peak if $r > 1$. For Gaussian-shaped peaks it should be noted that the RP defined by Equation (23) is about 2.08 times greater then the RP defined by Equation (22).

6.3 Dependence of Resolving Power on the Differences in Composition, Mass-to-charge Ratios and Relative Abundances of Successive Ion Beams

The RP that is appropriate for an HRMS group-type analysis depends in part on the differences in the m/z values and the relative abundances of the isobaric ions that will be present in the mass spectra of a given sample. The difference in ion m/z values resulting from different ion compositions has been considered by a number of individuals.[14,26,50,52,55,57,65,122,123] Table 9 lists differences in ion m/z values for various compositional differences consistent with the general formula $C_nH_xN_aO_bS_c$ and the observation that the coefficients of the heteroatomic species of present interest will in general have values $0 \leq a \leq 2$, $0 \leq b \leq 4$ and $0 \leq c \leq 2$.[65]

The dependence of the m/z values of resolved ion beams having equal signal magnitudes on both RP and the m/z differences between them is illustrated in Figures 9 and 10 for RPs from 10 000 to 80 000 and $\Delta(m/z)$ values from 2×10^{-3} to 30×10^{-3} u. Figures 9 and 10 show that the range of ion m/z values that can be resolved using RP values ranging from 10 000–80 000 exhibits a strong dependence on the difference in m/z values between successive ions. For example, Figures 9 and 10 show that increasing $\Delta(m/z)$ from 2 to 10 to 30 millimass units (mmu) increases the resolvable range of m/z values from 200 to 800 to greater than 1800. The rate of increase in the m/z value of resolved ion beams at a given RP is seen to increase dramatically with increasing $\Delta(m/z)$. For example, Figures 9 and 10 reveal that the upper limit of the mass range of resolved ions at 40 000 RP increases from 80 to 720 as $\Delta(m/z)$ varies from 2 to 18 mmu. However, Figure 9 shows that increasing the RP to 80 000 has only a marginal effect on the resolvable mass range for ions having $\Delta(m/z)$ differences less then 6 mmu.

The dependence of the RP required to resolve doublets on both the ion compositions and relative abundances are illustrated for a range of $\Delta(m/z)$ values in Table 10. Column 4 lists the $\Delta(m/z)$ values resulting from the compositional differences in column 3 between the doublets having the compositions shown in column 2. Values of the RP required for resolution of the members of each doublet having equal signal magnitudes and the maximum m/z value at which each doublet will be resolved are listed in columns 5 and 6, respectively. Inspection of Equation (24) reveals that the relative abundance of resolved ions increases with decreasing m/z at a given RP. The values in columns 7 and 8 in Table 10 illustrate the dependence of the relative abundance of resolved ions on ion mass for the RPs in column 5. For example, a RP of 6438 is predicted to resolve -26(H) and -12(H) ions having relative abundances ≥ 50 at m/z values ≤ 480 and ≥ 100 for $m/z \leq 466$. Table 10 illustrates the ubiquitous problem of resolving ions that differ in composition by C_3/SH_4 and various isotopically substituted ions.

6.4 Mass Analyzers and the Resolution of Ion Beams

6.4.1 Double Focusing Magnetic Instruments

Large geometry state-of-the-art DF instruments can be used to acquire HRLVEIMS exhibiting adequate ion intensities from about m/z 100 to about m/z 650 at RPs ranging up to about 60 000, depending upon factors such as the magnitude of the total aberrations,

Table 9 Differences in compositions and m/z values between successive ions

No.	Difference in		No.	Difference in		No.	Difference in	
	Composition	$m/z \times 10^{-3,a}$		Composition	$m/z \times 10^{-3,a}$		Composition	$m/z \times 10^{-3,a}$
1	$^{13}CH_3S/C_4$	1.1	23	O_2/S	17.7	45	$C_5N/H_{10}O_4$	54.8
2	$C_7/^{13}CH_7S_2$	2.3	24	CH_4S/O_3	18.6	46	$^{13}CH_5N/O_2$	55.7
3	C_3/SH_4	3.4	25	C_3/H_4O_2	21.1	47	C_2/H_8O	57.5
4	$CN/^{13}C_2$	3.6	26	$^{13}C_2H_{22}/SO$	21.2	48	$^{13}C_2/H_{10}O$	66.4
5	$H_4{}^{34}S/C_2N$	3.9	27	H_2N/O	23.8	49	CH_4O_3/S_2	71.9
6	$^{13}C/CH$	4.5	28	$^{13}C_2C_2/H_2SO$	24.1	50	C_2H_8/O_2	72.8
7	CH_2S/NO_2	5.2	29	$^{13}CHO_2/NS$	25.9	51	$CH_6O/^{34}S$	74.0
8	H_2NS/O_3	6.1	30	$^{13}C_2H_2/CO$	27.4	52	CH_6N/S	78.0
9	CO_3/N_2S	6.5	31	$CH_3N/^{13}CO$	28.3	53	$C_5/H_{12}O_3$	78.6
10	$C_4N/^{13}C_2H_4S$	7.0	32	$CH_3N/^{13}CO$	28.3	54	$C_2/H_{10}N$	81.3
11	$H_4{}^{34}S/^{13}C_2C$	7.5	33	$C_3N/^{13}CH_5O_2$	29.2	55	$^{13}C_2H_6/S$	81.6
12	$^{13}CH/N$	8.1	34	C_4/SO	33.0	56	C_6H_6/NO_2S	82.0
13	$^{13}CH_3{}^{34}S/C_3N$	8.4	35	O_4/S_2	35.5	57	$N_2/H_{12}O$	82.7
14	C_4/H_2NO_2	8.6	36	CH_4/O	36.4	58	C_3H_{10}/NO_2	85.4
15	$^{13}C_2O_2/C_2H_2S$	8.8	37	C_6/H_8O_4	42.2	59	$^{13}CCH_7/S$	86.1
16	C_4/H_2NS	9.2	38	$^{13}CCH_5/NO$	44.5	60	C_5H_4/S_2	87.2
17	CO/N_2	11.2	39	C_3/H_6NO	44.9	61	C_2H_8/S	90.5
18	C_7/H_4SO_3	11.9	40	C_2H_6/NO	49.0	62	C/H_{12}	93.9
19	CH_2/N	12.6	41	$^{13}CH_3O/S$	49.7	63	$CH_8N/^{34}S$	97.8
20	C_4/O_3	15.2	42	C_4O/S_2	50.8	64	$C_2H_8O_2/S_2$	108.3
21	$^{13}C_2H_2/N_2$	16.2	43	C_5H_4/O_4	51.6	65	$C_2H_{10}/^{34}S$	110.4
22	$C_3/^{34}SH_2$	16.5	44	CH_4O/S	54.1			

[a] The mass difference is the absolute value of the difference $(m/z)_2 - (m/z)_1$ for the formula differences F_2/F_1.

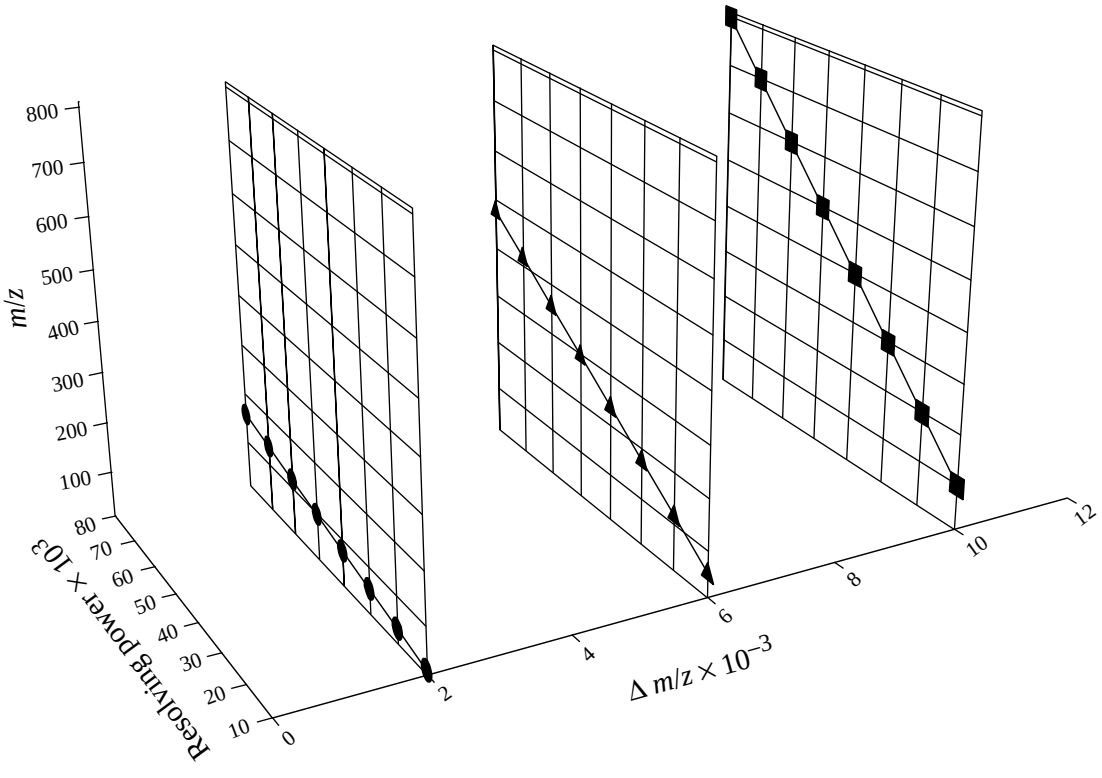

Figure 9 Mass-to-charge ratios of resolved ion beams having equal signal magnitudes as a function of RP and mass-to-charge ratio differences of 2×10^{-3}, 6×10^{-3} and 10×10^{-3} u.

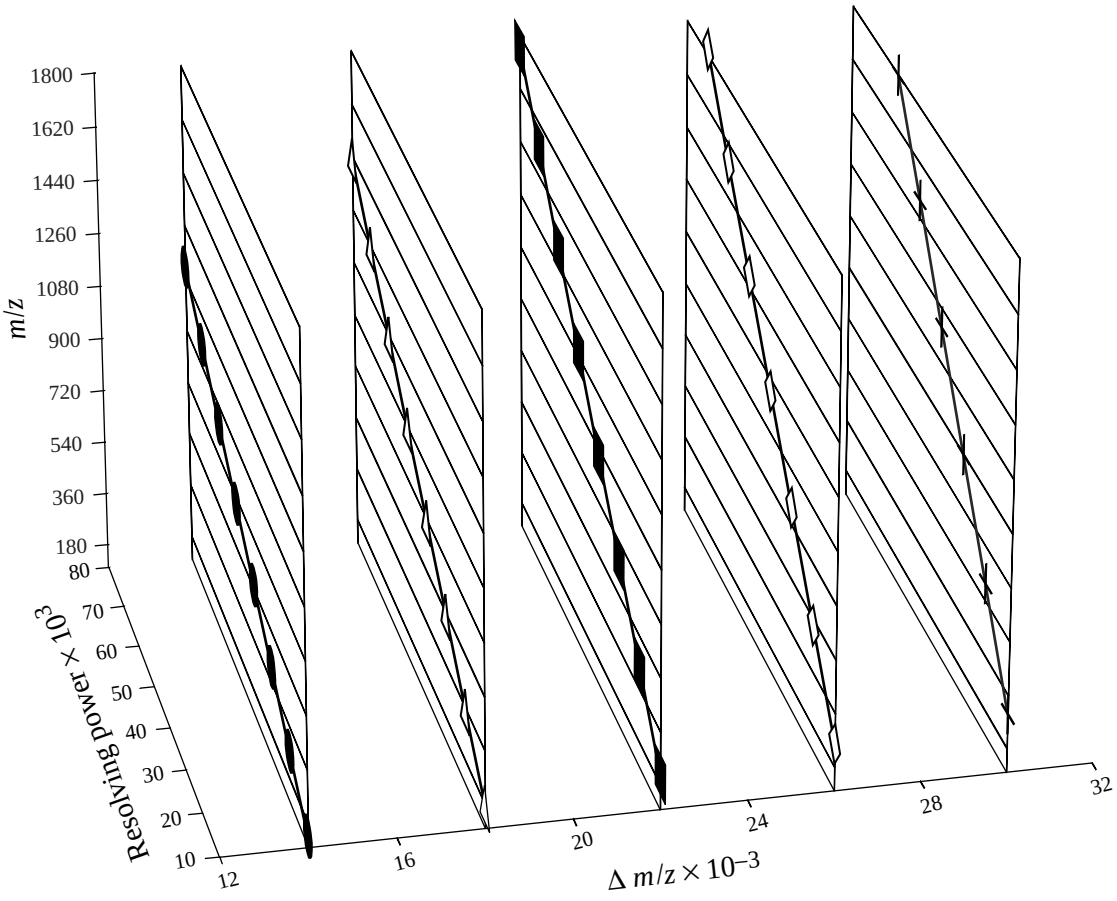

Figure 10 Mass-to-charge ratios of resolved ion beams having equal signal magnitudes as a function of RP and mass-to-charge ratio differences of 14×10^{-3}, 18×10^{-3}, 22×10^{-3}, 26×10^{-3} and 30×10^{-3} u.

which for a given instrument determines the ion source and collector slit widths required,[83] the rates of ion production that can be achieved with commercial electron impact sources operated at low voltage, the ranges in both the masses and the amounts of the compounds in a sample and the amount of sample that can be introduced into a DF mass spectrometer ion source. If we assume that singly charged ions having unit relative abundances are to be resolved at mass 650 then ions having compositional differences resulting in mass differences greater then 65, 32.5. 16.3 and 10.8 mmu will be resolved in spectra acquired at 10 000, 20 000, 40 000 and 60 000 RP, respectively. Doublets with mass differences less then these values will not be resolved over the entire mass range. For example, resolution of doublets 22–65 listed in Table 9 should be realized in LVEI mass spectra acquired at 40 000 RP. However, doublets such as $^{13}C/CH$ and C_4/H_2NO_2 (numbers 6 and 14 in Table 9) will be resolved up to m/z values of only about 180 and 344.

The presence of unresolved doublets must be taken into account in the processing of HRLVEIMS in order to avoid introducing significant errors into the analytical results. For example, the signal magnitudes of odd mass peaks potentially corresponding to unresolved $^{13}CH/N$ doublets need to be corrected for contributions from the ion abundances of the appropriate peak at one mass lower in order to avoid overestimating the contribution of nitrogenous homologs to the sample. Failure to consider that odd mass peaks correspond to unresolved $^{13}C/CH$ doublets could result in the erroneous identification of homologs of types such as $Z(NO_2)$ in a sample rather than homologs of fragment ion types such as $Z(H)$ in a spectrum.

A number of indirect approaches to circumventing this lack of spectral resolution have been developed. One approach involves extrapolations based upon a combination of differences in the initial carbon numbers of unresolved molecular ions and the distillation characteristics of their precursor homologs.[35] Unfortunately a detailed description has apparently never been published. Deconvolution of peaks derived from data discreetly acquired during a scan is another approach to enhancing resolution.[56,57,124] The deconvolution method described

Table 10 Dependence of RP on the compositional difference between successive ion beams and on the relative beam intensities

$-Z$	Formula	Difference in		RP for $r = 1$	m/z at which $r = 1$	m/z at which $r =$	
		Formulas	$m/z \times 10^{-3}$			50	100
26(H)	$C_{45}H_{64}$				604.50080		
		C/H_{12}	93.90	6438		480	466
12(H)	$C_{44}H_{76}$				604.59470		
16(O)	$C_{40}H_{64}O$				560.49572		
		CH_4/O	36.39	15 404		445	432
14(H)	$C_{41}H_{68}$				560.53210		
19(NS)	$C_{34}H_{49}NS$				503.35857		
		S/O_2	17.76	28 346		400	388
19(NO_2)	$C_{34}H_{49}NO_2$				503.37633		
29(N)	$C_{36}H_{43}N$				489.33955		
		$^{13}C/N$	8.11	60 368		389	377
30(H)	$^{13}CC_{36}H_{44}$				489.34766		
35(H)	$^{13}CC_{42}H_{51}$				568.40243		
		$^{13}C/CH$	4.47	127 159		451	438
34(H)	$C_{43}H_{52}$				568.40690		
		C_3/SH_4	3.37	168 667		451	438
24(S)	$C_{40}H_{56}S$				568.41027		
		$^{13}C_2O_2/C_2H_2S$	8.82	64 446		451	438
26(O_2)	$^{13}C_2C_{38}H_{54}O_2$				568.41909		

by Van Katwijk[56,57] is directly applicable to the processing of spectra acquired by exponential variation of the magnetic field strength. A method developed by Scheppele and Nutter analyses a HRMS for the presence of unresolved peaks, corrects the spectrum for contributions of ions containing either carbon-13 or sulfur-34 or both and assigns probable formulas to the mass and intensity corrected peaks. The method has been described in detail by Gaun et al.[92] A related approach to processing unresolved peaks has been described by Roussis.[98]

6.4.2 Fourier Transform Ion Cyclotron Mass Spectrometers

As seen in Figures 9 and 10, in Table 10 and discussed in section 6.3, resolution of a number of important doublets requires RPs that exceed the capabilities of DF instruments. The instrumental resolution of doublets such as C_3/SH_4, $^{13}C/CH$ and $^{13}CH/N$ would significantly benefit HRMS-type analyses by increasing the reliability of the results and significantly reducing the time for spectral processing. It is thus important to note that several authors have demonstrated the resolution of such doublets in the mass spectra of samples of interest using FTICR mass spectrometers.[91,92,97] In addition, the errors in mass measurement were typically 1 ppm or less.[92,97] The mass range over which spectra exhibiting resolutions exceeding 300 000 (based upon the Full Width Half Maximum (FWHM) definition) can be acquired was significantly increased and, hence, the time required to cover the complete mass range of a sample was diminished by replacement of the three Telsa magnet[92] with a 5.6 Telsa magnet.[96,97]

7 QUANTITATIVE ANALYSIS

Cross-sections for ionization are of fundamental significance to quantitative MS because they provide the basis for describing the relationship between the total ionization and the concentration of neutral species in the region of ionization. However, the relationship between ion signal magnitudes and the composition of a mixture uses the concept of sensitivity. The relationship between cross-sections and sensitivities for EI is developed in sections 7.1 and 7.2 in order to provide a precise definition and a clear understanding of the concept of sensitivities. Formulas for calculating CNDs and a number of chemical properties from these distributions are derived in section 7.3. The effect of isomers on the quantitation of homologs is also addressed. Ionization sensitivities are an important factor in determining the accuracy of CNDs. However, the published literature contains only a limited number of studies involving the determination of low-voltage EI sensitivities for ionization of aromatic compounds.[6,8,10,11,13,16,19,22,23,26,34,37,38,47,72,80,97,125–127] Moreover, few attempts at developing correlations that express the functional dependence of LVEI sensitivities on molecular structure quantitatively have been

reported.[6,8,11,16,20,26,72] Development of these correlations is essential in order to predict values for LVEI sensitivities because the number of organic compounds for which LVEI sensitivities can be determined comprises an unacceptably small subset of the compounds either known or thought to be present in distillates of present interest. Qualitative correlations of the dependence of LVEI sensitivities on molecular structure have been published and extensively reviewed.[35,38,80] Unfortunately, these correlations do not allow the prediction of LVEI sensitivities. Consequently, the following approach was adopted in addressing the issue of developing a database of LVEI sensitivities. Fundamental phenomena that affect the shape of the ionization efficiencies in the threshold region are reviewed in section 7.4. The practical consequences of these phenomena on experimental LVEI ion abundances are considered in section 7.5. Section 7.6 addresses the prediction of LVEI sensitivities from EI and photoionization cross-sections. The reproducibility of sensitivities for LVEI of aromatic hydrocarbons are reviewed and discussed in terms of values predicted from relative electron-impact ionization probabilities assuming a linear threshold law for ionization in section 7.7.

7.1 Electron Ionization Cross-sections

Total ionization cross-sections can be calculated from total ion currents determined without mass analysis of the ion beams. Consider compound k with an internal energy defined by temperature T. Values of the total cross-section for ionization of compound k by electrons with an energy V electron volts are given by Equation (25)

$$\sigma_{k,T,V} = \frac{TIC_{k,T,V}}{(n_{\text{tot},k})\text{Vol}^{-1} I_e d} \tag{25}$$

where $TIC_{k,T,V}$ is the total ion current, I_e is the ionizing electron current, d is the ionizing pathlength in cm and $n_{(\text{tot},k)}\text{Vol}^{-1}$ is the number of molecules per cubic centimeter in the ion source. Thus, the cross-section has units of cm^2 molecule^{-1}.

The cross-sections for ionization of the isotopically most abundant molecules of a given compound are of interest because the m/z values and signal magnitudes of the isotopically most abundant molecular ions rather then total ion signal magnitudes are of analytical significance. If $k(1), k(2), k(3), \ldots, k(l)$ designates the order of the variously isotopically substituted molecules of compound k according to decreasing natural abundance, then the number of each of them is $n[k(1)], n[k(2)], n[k(3)], \ldots, n[k(l)]$. If EI of the isotopically most-abundant molecules results in the formation of molecular ($j = 1$) and fragment ($j = 2, 3, 4, \ldots, m$) ions then the cross-section for ionization of the isotopically most-abundant molecules of compound k is given by Equation (26)

$$\sigma_{k(1),T,V} = \sum_{j=1}^{m} \frac{SM(j)_{k(1),T,V}}{PM(j)_{k(1)} GM(j)_{k(1)} n_{k(1)} \text{Vol}^{-1} I_e d R_r SM_v} \tag{26}$$

In Equation (26), the numerator is the signal magnitude of the jth isotopically most-abundant ion formed from ionization of compound $k(1)$ in volts, $PM(j)_{k(1)}$ represents the probability that the jth ion from $k(1)$ will exit the ion source and be transmitted through the mass analyzer and be detected, $GM(j)_{k(1)}$ designates the detector gain for this ion and R_r and SM_v designate the resistor in the ion-current amplifier and the experimental ion signal magnitude per volt of signal from the detector system, respectively.

We now express the number of isotopically most-abundant molecules $n_{k(1)}$ as a function of the total number of moles of compound k, $N_{\text{tot},k}$. The number of isotopically most-abundant molecules is given by Equation (27)

$$n_{k(1)} = FP_{k(1)} N_{\text{tot},k} N_{\text{A}} \tag{27}$$

where $FP_{k(1)}$ is the fractional percentage abundance of the isotopically most-abundant molecules and N_{A} is Avogadro's number. Values of $FP_{k(1)}$ can be obtained to acceptably high levels of accuracy from Equation (28)

$$FP_{k(1)} = \prod_i \frac{a_{(i)}!}{\prod_j b_{(i,j)}!} \prod_j FP(b_{i,j})^{b(i,j)} \tag{28}$$

where a_i represent the number of the ith element and $b_{i,j}$ denotes the number of the jth isotope of the ith element in a formula. The symbols and values for i and j are defined in terms of elements and isotopes in Table 11.

Table 11 Definition and values of parameters on the right hand side (RHS) of Equation (28)

i	Element $a(i)$	i, j	Isotope $b(i, j)$	Fractional % abundance of isotope $b(i, j)$[a]
1	Carbon	1, 1	Carbon-12	0.988 89
		1, 2	Carbon-13	0.011 11
2	Hydrogen	2, 1	Hydrogen	0.999 844
		2, 2	Deuterium	0.000 156
3	Nitrogen	3, 1	Nitrogen-14	0.996 34
		3, 2	Nitrogen-15	0.003 66
4	Oxygen	4, 1	Oxygen-16	0.997 628
		4, 2	Oxygen-18	0.002 000
5	Sulfur	5, 1	Sulfur-32	0.950 18
		5, 2	Sulfur-34	0.042 15

[a] Reproduced by permission of the International Union of Pure & Applied Chemistry (IUPAC) and Pergamon Press from N.E. Holden, R.L. Martin, I.L. Barnes, 'Isotopic Compositions of the Elements 1983', *Pure Appl. Chem.*, **56**(6), 675–694 (1984).

Substituting Equation (27) into Equation (26), we obtain the following expression for the cross-section for ionization of the isotopically most-abundant molecules of compound k expressed as a function of the signal magnitudes of the molecular ion and fragment ions formed from ionization of molecules $k(1)$, the total moles of compound per cubic centimeter of the ion source, the ionizing electron current, the ionizing pathlength, the detector gain and the probability that a molecular ion will exit the ion source and be transmitted through the mass analyzer and be detected, Equations (29) and (30)

$$\sigma_{k(1),T,V} = \frac{1}{\kappa}\sum_{j=1}^{m}\frac{SM(j)_{k(1),T,V}}{PM(j)_{k(1)}GM(j)_{k(1)}FP_{k(1)}N_{\text{tot},k}} \quad (29)$$

where

$$\kappa = R_r SM_v N_A I_e d\text{Vol}^{-1} \quad (30)$$

The total cross-section for ionization of compound k by electrons possessing energy V is a linear combination of the cross-sections for ionization of the various isotopically substituted species weighted by their fractional percentage abundance, that is, Equation (31)

$$\begin{aligned}\sigma_{k,T,V} &= (\sigma_{k(1),T,V})FP_{k(1)} + (\sigma_{k(2),T,V})FP_{k(2)} + (\sigma_{k(3),T,V}) \\ &\quad \times FP_{k(3)} + (\sigma_{k(4),T,V})FP_{k(4)} + \cdots + (\sigma_{k(m)})FP_{k(m)} \\ &= \sum_{l}(\sigma_{k(l),T,V})FP_{k(l)} \end{aligned} \quad (31)$$

Since the Born–Oppenheimer approximation is valid for the range of energies of present interest, the potential energy hypersurfaces for both ionic and neutral species are independent of isotopic substitution. However, ionization cross-sections could be subject to isotope effects because the vibrational and rotational energy levels associated with these electronic states are isotope dependent. In this regard, the ionization cross-sections for small deuterated polyatomic molecules were subject to deuterium isotope effects ranging from about 1.01 to about 1.05.[128,129] Deuterium and heavy atom such as carbon-13 isotope effects would be expected to be considerably smaller for large polyatomic molecules. Thus, Equation (31), reduces to Equation (32)

$$\sigma_{k,T,V} = (\sigma_{k(1),T,V}) \quad (32)$$

for the condition that isotope effects on ionization cross-sections are less than the experimental uncertainties in these values.

7.2 Low-voltage Ionization Cross-sections and Ionization Sensitivities

In order to derive the relationship between LVEI cross-sections and sensitivities we separate the RHS of the equation obtained from combining Equations (29) and (32) into terms for molecular and fragment ions and rearrange the result to obtain

$$\begin{aligned}(\sigma_{k,T,V})FP_{k(1)} &= \frac{SM(1)_{k(1),T,V}}{\kappa PM(1)_{k(1)}GM(1)_{k(1)}N_{\text{tot},k}} + \frac{1}{\kappa} \\ &\quad \times \sum_{j=2}^{m}\frac{SM(j)_{k(1),T,V}}{PM(j)_{k(1)}GM(j)_{k(1)}N_{\text{tot},k}}\end{aligned} \quad (33)$$

If the cross-section for ionization of compounds by low-voltage electrons are either known or assumed to be dominated by molecular ions then the second term in the RHS of Equation (33) can be neglected. For either situation, Equation (33) reduces to

$$(\sigma_{k,T,V})FP_{k(1)} = \frac{SM(1)_{k(1),T,V}}{\kappa PM(1)_{k(1)}GM(1)_{k(1)}N_{\text{tot},k}} \quad (34)$$

The RHS of Equation (34) is the signal magnitude for the isotopically most-abundant molecular ion for a given value of the concentration of the precursor homolog in the ion source, the ionizing electron current, the ionizing pathlength and both the detector gain and the ion-source extraction and analyzer transmission probabilities for these ions. Thus, a solution of Equation (34) is by definition a mass spectrometric response factor, termed an ionization sensitivity, for the kth compound. Thus, the sensitivity for ionization of the kth compound per mole of the kth compound by low-voltage electrons with energy V, termed the mole sensitivity, has the units square centimeters per mole and is by definition

$$S(\text{mol})_{k,T,V} = \frac{SM(1)_{k(1),T,V}}{\kappa PM(1)_{k(1)}GM(1)_{k(1)}N_{\text{tot},k}} \quad (35)$$

Combining Equations (34) and (35) we obtain

$$S(\text{mol})_{k,T,V} = (\sigma_{k,T,V})FP_{k(1)} \quad (36)$$

Equation (36) reveals that the mole sensitivities required for calculating CNDs from molecular ion signal magnitudes could be obtained from total ionization cross-sections provided that the contribution of fragment ions to the total ionization is either minimal or known.

As discussed in sections 7.4 and 7.5 the ionization cross-section is a rapidly changing function in the threshold region. Consideration of Equations (25), (26) and (35) in terms of this fact and other factors such as the performance characteristics of conventional electron ion sources, the focusing of ion beams and the ion statistics accounts for the difficulties that are often experienced in obtaining satisfactorily reproducible ionization efficiency curves and absolute values of cross-sections and mole sensitivities in the threshold region. The influence of these factors can be minimized by determining relative

cross-sections and relative mole sensitivities. These values can be converted to absolute values given the value for the reference compound. It should be noted that the use of relative rather then absolute mole sensitivities expedites the determination of CNDs because the necessity of determining the absolute sensitivity a mass spectrometer routinely is abrogated.

Let $\sigma_{\mathrm{rel},k,T,V}$ and $\mathrm{S(mol)}_{rel,k,T,V}$ designate the cross-section and mole sensitivity, respectively, for ionization of the kth compound relative to the corresponding quantities for ionization of the reference compound ref, where T and V designate the gas temperature and the ionizing energy, respectively. Let ref(1) represent the isotopically most-abundant neutral and ionic reference species. Thus, the relative cross-sections and relative mole sensitivities are, Equations (37), (38) and (39)

$$\sigma_{\mathrm{rel},k,T,V} = \frac{TIC_{k,T,V}}{TIC_{\mathrm{ref},T,V}} \frac{N_{\mathrm{tot,ref}}}{N_{\mathrm{tot},k}} \tag{37}$$

$$\sigma_{\mathrm{rel},k,T,V} = \frac{SM(1)_{k(1),T,V}}{SM(1)_{\mathrm{ref}(1),T,V}} \frac{N_{\mathrm{tot,ref}}}{N_{\mathrm{tot},k}} \frac{PM(1)_{\mathrm{ref}(1)}}{PM(1)_{k(1)}} \times \frac{GM(1)_{\mathrm{ref}(1)}}{GM(1)_{k(1)}} \frac{FP_{\mathrm{ref}(1)}}{FP_{k(1)}} \tag{38}$$

$$S(\mathrm{mol})_{\mathrm{rel},k,T,V} = \frac{SM(1)_{k(1),T,V}}{SM(1)_{\mathrm{ref}(1),T,V}} \frac{N_{\mathrm{tot,ref}}}{N_{\mathrm{tot},k}} \frac{PM(1)_{\mathrm{ref}(1)}}{PM(1)_{k(1)}} \times \frac{GM(1)_{\mathrm{ref}(1)}}{GM(1)_{k(1)}} \tag{39}$$

Equation (40) expresses relative sensitivities on a gram basis as function of molecular ion signal magnitudes and the grams of each compound in the mixture, $g_{\mathrm{tot},k}$, and the grams of the reference compound, $g_{\mathrm{tot,ref}}$

$$S(\mathrm{g})_{\mathrm{rel},k,T,V} = \frac{SM(1)_{k(1),T,V}}{SM(1)_{\mathrm{ref}(1),T,V}} \frac{g_{\mathrm{tot,ref}}}{g_{\mathrm{tot},k}} \frac{PM(1)_{\mathrm{ref}(1)}}{PM(1)_{k(1)}} \times \frac{GM(1)_{\mathrm{ref}(1)}}{GM(1)_{k(1)}} \tag{40}$$

Relative mole sensitivities are converted into relative gram sensitivities using Equation (41)

$$S(\mathrm{g})_{\mathrm{rel},k,T,V} = S(\mathrm{mol})_{\mathrm{rel},k,T,V} \frac{(M_r)_{\mathrm{ref}}}{(M_r)_k} \tag{41}$$

where $(M_r)_{\mathrm{ref}}$ and $(M_r)_k$ are the relative molecular masses of the reference compound and compound k, calculated from the atomic weights of their constituent elements, respectively.

7.3 Calculation of Carbon Number Distributions and Chemical Properties

Consider the ionization of the jth homolog in the ith specific Z series at temperature T by electrons of voltage V. The jth homolog may consist of 0, 1, 2, 3, . . . , l isomers. For the sake of simplicity, the notation for this homolog is taken to be $Z_{i,j,l}$ and is illustrated in Table 12 for the first three members of the −6(H) homologous series.

Consider a sample containing the four compounds shown in Table 12. The first and second homologs consist of only one compound each. However, the third homolog is composed of four isomers. Following conventional practice, we assume that the detector gains and the ion-source extraction/analyzer transmission probabilities are the same for the molecular ions of the compounds listed in Table 12. Thus, using absolute rather then relative sensitivities, the number of moles of the first and second homologs benzene and toluene are, Equations (42) and (43)

$$N_{\mathrm{tot},Z_{-6,1,0}} = \frac{SM(78)_{Z_{-6,1,0},T,V}}{S(\mathrm{mol})_{Z_{-6,1,0},T,V}} \tag{42}$$

and

$$N_{\mathrm{tot},Z_{-6,2,0}} = \frac{SM(92)_{Z_{-6,2,0},T,V}}{S(\mathrm{mol})_{Z_{-6,2,0},T,V}} \tag{43}$$

respectively. Since the third homolog consists of four isomers, the number of moles of this homolog is given by Equation (44)

$$N_{\mathrm{tot},Z_{-6,3,4}} = \frac{SM(106)_{Z_{-6,3,1},T,V}}{S(\mathrm{mol})_{Z_{-6,3,1},T,V}} + \frac{SM(106)_{Z_{-6,3,2},T,V}}{S(\mathrm{mol})_{Z_{-6,3,2},T,V}} + \frac{SM(106)_{Z_{-6,3,3}}}{S(\mathrm{mol})_{Z_{-6,3,3}}} + \frac{SM(106)_{Z_{-6,3,4}}}{S(\mathrm{mol})_{Z_{-6,3,4}}} \tag{44}$$

Molecular ion mass-to-charge ratios and signal magnitudes are the experimentally accessible quantities in a HRMS homolog analysis. Since the first and second −6(H) homologs correspond to the presence of only one compound each, Equations (42) and (43) can be solved to obtain values for the number of moles of benzene and toluene in the mixture. In contrast, solving Equation (44) to obtain the number of moles of the third −6(H) series homolog would require the signal magnitudes of the four

Table 12 Compounds illustrating nomenclature used in Equations (42–45)

Values of			Compound	Formula	Mass number
i	j	l			
−6	1	0	Benzene	C_6H_6	78
−6	2	0	Toluene	C_7H_8	92
−6	3	1	Ethylbenzene	C_8H_{10}	106
−6	3	2	*o*-Xylene	C_8H_{10}	106
−6	3	3	*m*-Xylene	C_8H_{10}	106
−6	3	4	*p*-Xylene	C_8H_{10}	106

C_8H_{10} molecular ions rather then the sum of these signal magnitudes which is the experimentally accessible quantity $SM(106)_{\exp,T,V}$. Therefore, a homolog mole sensitivity which is a value bounded by the least and greatest mole sensitivities for the set of isomers is substituted for the four sensitivities in Equation (44) in order to obtain a value for the number of moles of the third homolog. Hence, Equation (44) reduces to Equation (45)

$$N_{\text{tot},Z_{-6,3}} = \frac{SM(106)_{\exp,T,V}}{S(\text{mol})_{Z_{-6,3},T,V}} \tag{45}$$

The sensitivities used in calculating quantitative distributions are, in general, values for homologs rather then values for specific compounds. The magnitudes of the error so introduced into the quantitative results are a complex function of variables such as the difference between the assumed and correct sensitivity values for the actual isomeric mixture, the signal magnitude of the homolog molecular ion, the effect of mass discrimination on the molecular-ion signal magnitudes and the number of homologs of the compound types present in the sample.

For the general case, consider a distillate sample containing 1, 2, 3, . . . , *i* compound types. Each compound type consists of 1, 2, . . . , *j* homologs. The mole percentage of the *j*th homolog in the *i*th specific *Z* series is

$$MP_{Z_{i,j}} = \frac{N_{\text{tot},Z_{i,j}}}{\sum_i \sum_j N_{\text{tot},Z_{i,j}}} \times 100 \tag{46}$$

where for the sake of simplicity $Z_{i,j}$ has been substituted for notation $Z(N_aO_bS_c)$. Generalizing the result obtained from substituting Equations (42), (43) and (45) into Equation (46), we obtain Equation (47) for calculating mole-percentage CNDs from the signal magnitudes of the isotopically most-abundant molecular ions

$$MP_{Z_{i,j}} = \frac{(SM_{Z_{i,j},T,V})/(S(\text{mol})_{\text{rel},Z_{i,j},T,V})}{\sum_i \sum_j (SM_{Z_{i,j},T,V})/(S(\text{mol})_{\text{rel},Z_{i,j},T,V})} \times 100 \tag{47}$$

Weight-percentage CNDs can be calculated by solving Equation (48)

$$WP_{Z_{i,j}} = \frac{(SM_{Z_{i,j},T,V})/(S(\text{g})_{\text{rel},Z_{i,j},T,V})}{\sum_i \sum_j (SM_{Z_{i,j},T,V})/(S(\text{g})_{\text{rel},Z_{i,j},T,V})} \times 100 \tag{48}$$

or from mole percentage CNDs using Equation (49)

$$WP_{Z_{i,j}} = \frac{MP_{Z_{i,j}}(M_r)_{Z_{i,j}}}{\sum_i \sum_j MP_{Z_{i,j}}(M_r)_{Z_{i,j}}} \times 100 \tag{49}$$

where $(M_r)_{Z_{i,j}}$ is the relative molecular mass of the *j*th homolog in the *i*th specific *Z* series. CNDs on a weight-percentage basis can also be obtained using Equation (50)

$$WP_{Z_{i,j}} = \frac{[(SM_{Z_{i,j},T,V})/(S(\text{mol})_{\text{rel},Z_{i,j},T,V})] \times (M_r)_{Z_{i,j}}}{\sum_i \sum_j [(SM_{Z_{i,j},T,V})/(S(\text{mol})_{\text{rel},Z_{i,j},T,V})] \times (M_r)_{Z_{i,j}}} \times 100 \tag{50}$$

Aczel and Lumpkin have demonstrated the routine calculation of a significant number of chemical properties from CNDs.[35,52,53] The properties calculated include the average carbon number, the average molecular weight and the elemental analysis for each compound type and for the sample. In addition to these properties, Schmidt et al. reported calculating average specific *Z* values and values of f_a, that is, the fraction of the total carbon that is aromatic.[75] Aczel and Lumpkin have also shown that CNDs can be used to calculate distillation characteristics for a sample and to generate CNDs for the compound types expected to be present in the predicted narrow-boiling distillate cuts.[32] Unfortunately many of the essential details of the method have not been described. Chemical properties for narrow boiling distillate cuts should be predictable from the CNDs for these fractions. Scheppele and Greenwood were able to draw conclusions concerning the occurrence of reaction networks in the hydrotreating of anthracene oil from the CNDs for the feed and products.[60]

In regard to such calculations, the average relative molecular mass for a specific *Z* series, $(\overline{M}_r)_{z_i}$, the average relative molecular mass for the sample, $\overline{M}_r$, the average carbon number for a specific *Z* series, $\bar{n}_{Z_i}$, and the average carbon number for the sample, $\bar{n}$, are solutions to Equations (51–54), respectively

$$(\overline{M}_r)_{Z_i} = \frac{\sum_j MP_{Z_{i,j}}(M_r)_{Z_{i,j}}}{TMP_{Z_i}} \tag{51}$$

$$\overline{M}_r = \frac{\sum_i TMP_{Z_i}(\overline{M}_r)_{Z_i}}{100} \tag{52}$$

$$\bar{n}_{Z_i} = \frac{\sum_i MP_{Z_i} n_{Z_i}}{TMP_{Z_i}} \tag{53}$$

$$\bar{n} = \frac{\sum_i TMP_{Z_i}\bar{n}_{Z_i}}{100} \tag{54}$$

where, Equation (55)

$$TMP_{Z_i} = \sum_j MP_{Z_{i,j}} \tag{55}$$

Let Y_u denote the presence of u atoms of element Y in a formula and let $A(Y)$ represent the atomic weight of element Y. The weight percentage of element Y in a sample is thus

$$WP(Y) = A(Y) \sum_i \sum_j \frac{WP_{Z_{i,j}} u_{Z_{i,j}}}{(M_r)_{Z_{i,j}}} \quad (56)$$

where $(M_r)_{Z_{i,j}}$ is the relative molecular mass of a homolog calculated from the atomic weights of its constituent elements.

7.4 Fundamental Phenomena

Consider compound k with an internal energy defined by temperature T. The cross-section for its ionization by electrons having a nominal energy V is given by

$$\sigma_{k,T,E} = \sum_K \int_{-\infty}^{+\infty} K_\kappa Th(V - IP_\kappa)^w m(V - IP_\kappa)\,dV \quad (57)$$

where K_κ are the probabilities for transitions from the ground state of the neutral molecule to the various κ states of the ion having ionization potentials IP_κ, $Th(V - IP_\kappa)^w$ are the threshold laws for the transitions and $m(V - IP_\kappa)$ is the energy distribution in the ionizing beam.[130–132] Equation (57) includes contributions of single-direct ionization, multiple-direct ionization and autoionization to the cross-section. The contribution of multiple-direct ionization to the cross-section for ionization in the threshold region should be negligible for compounds of present interest. Unfortunately, very little definitive evidence exists concerning the contribution of electron-induced auto-ionization to the total ionization in the threshold region for these compounds. Therefore, we necessarily assume that single-direct ionization dominates the cross-section for electron-induced ionization of compounds of present interest. For this case, several theoretical analyses predict a value of one for the exponent w in the threshold law.[130–132] Ionization efficiencies obtained for polyatomic compounds and radicals using energy-resolved electron beams are consistent with a linear or an apparently linear threshold law, that is, if the threshold law is nonlinear, the region of nonlinearity falls within the energy resolution of the electron beam (0.07–0.08 eV).[133–139]

The assumption that the mass spectrum of a polyatomic molecule results from competing and consecutive unimolecular reactions of internally excited molecular ions led to the development of the quasi-equilibrium theory (QET).[140,141] Briefly, The RRKM (Rice–Ramsperger–Kassel–Marcus) rate constants, $k(E)$, for each unimolecular decomposition with a barrier height E_0 are calculated as a function of the internal energy E of the molecular ion. These rate constants and the reaction times are used to compute a breakdown graph that describes the yield of all ions in the reaction network quantitatively as a function of the internal energy of the molecular ion. Convolution of the breakdown graph with both the internal energy distribution function (P(E)) and the electron energy distribution function (EED) yields the mass spectrum. The P(E) corresponds to the convolution of the energy deposition function (P(ED)) and the Boltzmann function describing the thermal energy distribution of the neutrals prior to ionization. The P(ED) describes the probability of producing a molecular ion possessing an internal energy E as a function of the energy of the ionizing electron. Thus, P(ED) depends upon the Franck–Condon factors for electronic transitions from the various vibrational levels of the ground state of a molecule to the different vibrational levels of the various electronic states of the ion and upon the threshold law. The threshold law describes the probability of transferring an energy $E + IP_\kappa$ to the molecular ion as a function of the energy of the incident electron.

7.5 Practical Consequences of Fundamental Phenomena

Both the fundamental factors considered in the previous section and the operating characteristics of conventional electron sources have an impact on the practical application of low-voltage electrons as an ionizing medium. The threshold region for compounds of interest extends from about 6 eV to approximately 13 eV.[142] Equation (57) reveals that the cross-sections for ionization of these compounds by electrons having nominal energies in this range and energy distributions $m(V - IP_\kappa)$ depend upon the magnitudes of both the Franck-Condon factors for accessible transitions and the differences $V - IP_\kappa$. For a given value of the ionizing energy, the function $V - IP_\kappa$ will vary significantly in the threshold region as the ionization potential decreases. This conclusion is substantiated by the values in columns 3 and 4 of Table 13 for the compounds and ionization potentials listed in columns 1 and 2, respectively. Since the ionization potentials for aromatic hydrocarbons are lower than for saturated hydrocarbons, the threshold laws in column 5 are seen to be considerably larger for the former than for the latter. The values in columns 5 and 6 also show that the function $V - IP_\kappa$ increases with increasing aromaticity. A decrease in the electron energy by 3 eV is seen in column 7 to lead to a marked decrease in the magnitude of the threshold law for the three aromatic hydrocarbons.

The previous discussion reveals that the production of large numbers of molecular ions is facilitated by large values of $V - IP_\kappa$. However, the formation of molecular ions with insufficient energy to undergo decomposition at a given nominal value of the electron energy reflects the kinetics of fragmentation, the internal energy of the

Table 13 Effect of the threshold law on the electron impact cross-section for the ground electronic state

Compound	IP[a]	$V - IP_1$ for $V=$		$(V - IP_1)_{rel}$ for $V=$		$(10 - IP_1)/(13 - IP_1)$
		13 eV	10 eV	13 eV	10 eV	
Methane	12.61	0.39	–	1		
Benzene	9.24	3.76	0.76	9.64	1	0.20
Naphthalene	8.14	4.86	1.86	12.46	2.45	0.38
Chrysene	7.68	5.32	2.32	13.64	3.05	0.44

[a] Reprinted with permission from H.M. Rosenstock, K. Draxl, B.W. Steiner, J.T. Herron, 'Energetics of Gaseous Ions', *J. Phys. Chem. Ref. Data*, **6**, (Suppl. 1), Copyright 1977, American Institute of Physics, American Chemical Society, & National Institute of Science and Technology (*J. Phys. Chem. Ref. Data*).

molecular ion and the EED. For a given value of V, the range of internal excitation energies available to the molecular ion is $0 \leq E \leq V - IP_1$. The RRKM rate constants, $k(E)_1$, increase rapidly as E rises above $E_{0,1}$, where $E_{0,1}$ is the barrier height for the lowest energy fragmentation reaction. Thus, production of high yields of stable molecular ions relative to fragment ions requires that only a relatively small fraction of the electrons in the beam have energies in excess of $IP_1 + E_{0,1}$.

The second derivative of the ionization efficiency (SDIE) for formation of an ionized atom in its ground electronic state reduces to the EED reversed with respect to energy provided that ion formation involves single direct ionization obeying a linear threshold law.[132] A SDIE for xenon obtained by numerical differentiation of an experimentally determined first derivative of the ionization efficiency (FDIE) of xenon in the threshold region is reproduced in Figure 11.[143] The FDIE curve

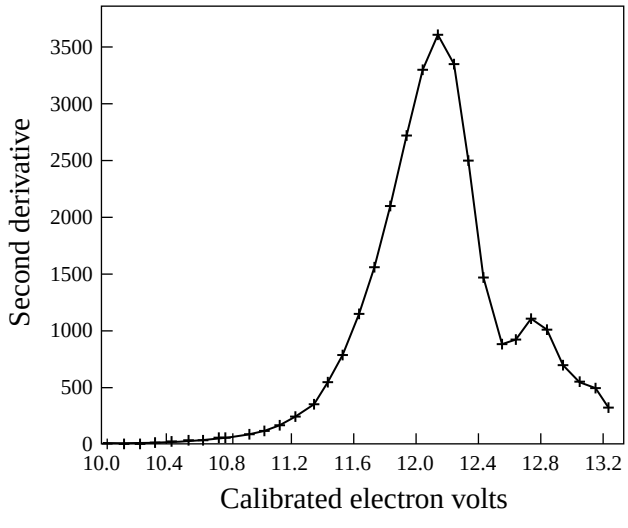

Figure 11 Second derivative of the xenon ionization efficiency curve in the threshold region. The EED is Maxwellian and reversed with respect to the energy axis. The FWHM of the first peak corresponding to the formation of $Xe^{\bullet+}$ in the $^2P_{3/2}$ state is 0.50 eV. The second maximum results from the 6d autoionizing state.

was obtained with a clean ion source and with the electron shield at zero potential with respect to the filament and with the extraction-plate potentials near the block potential to minimize the effect of contact potentials on the EED. The trap current was 20 μA to minimize space charge. The second maximum in the SDIE curve in Figure 11 which occurs at 12.8 ± 0.1 eV is caused by the 6d autoionizing state.[144] In agreement with theory,[132] the SDIE for formation of Xe^+ in the $^2P_{3/2}$ state is seen to correspond to a reversed Maxwellian distribution having a FWHM of 0.50 eV.

As seen in Figure 11, a significant probability exists that the ionizing beam will contain electrons having energies ranging from the nominal value up to about the nominal value plus 1.4 eV. Thus, molecular ions would be observed at nominal electron energies of approximately $IP - 1.4$ eV. Furthermore, the internal energies of the molecular ions can be increased over those present in the neutral precursor by amounts ranging from $V - IP$ up to $V - IP + 1.4$ eV. Thus, for $V = IP$ the EED shown in Figure 11 could produce molecular-ion decomposition reactions having $E_{0,1} \leq 1.4$ eV.

A number of experimental phenomena can increase the spread of electron energies in the ionizing region. For example, the use of 0.030 inch (0.0762 cm) rather than 0.020 inch (0.0508 cm) rhenium ribbon increased the FWHM of the EED from 0.42–0.55 eV to 0.63–0.74 eV.[143] An increase in the contamination of the ion source necessitates an increase (decrease) in the voltage applied to the repeller (extractor) electrodes which results in an increase in the FWHM of the EED. For example, SDIE curves for Xe having FWHM values in excess of 1 eV were obtained for contaminated ion sources. The FWHM of the EED is sensitive to the potential on the electron trap due to field penetration. For example, operation of a CEC 21-110B mass spectrometer at low-ionizing energies required modification of the ion source circuitry because penetration of the field resulting from the high fixed potential applied to the trap into the ionizing region sufficiently broadened the distribution of electron energies so that molecular ions

of benzene were observed at nominal electron energies below 1 eV.

Samples of interest generally require inlet and ion-source temperatures of 250 °C or greater. At these temperatures, the average thermal energy may attain considerable values for large molecules and the distribution of these energies will encompass a considerable range of energies because of the large number of degrees of freedom over which the thermal energy may be distributed. For, example, at 300 °C the average thermal energy of propane is ca. 0.3 eV and the FWHM of the Boltzmann distribution is approximately 0.3 eV.[145] At 327 °C the calculated thermal energy distribution for hexane has an average value of ca. 0.55 eV and extends from slightly greater than 0 eV up to ca. 1.5 eV.[146] The calculated average thermal energy of 2-phenyl-2-methylpropane at this temperature is about 0.8 eV and the calculated distribution ranges from about 0.2 to approximately 1.8 eV.[147] Since the thermal-energy content of the neutral species is fully effective in determining the kinetics of ion decompositions, the yield of fragment ions relative to molecular ions and the degree to which the ionization efficiency curves for these ions asymptotically approaches the energy axis increases with temperature for a given compound and with molecular size for a given temperature. For example, the thermal shifts of the ionization curves for a series of alkanes range from 0.10–0.24 eV as the temperature is raised from 28–142 °C.[148]

Appearance potentials for formation of fragment ions from saturated compounds are only a few tenths of a volt above their ionization potentials.[142] This fact together with the phenomena discussed above precludes the use of low-voltage electrons as the ionizing agent for performing homolog-type analysis of saturated hydrocarbons because of low molecular-ion signal magnitudes. Since the appearance potentials for fragment ions from aromatic compounds are generally greater then about 1–1.5 eV above their ionization potentials,[142] an electron energy can usually be selected that results in acceptably large molecular ion to fragment ion signal magnitude ratios.

7.6 Electron Impact Ionization and Photoionization Cross-sections

Total ionization cross-sections have been determined for a number of organic compounds.[129,149–157] Unfortunately, these cross-sections have only limited applicability with regard to predicting LVEI sensitivities for two reasons. First, except for benzene[129,149,150,153] and a series of its alkylated homologs[129,153] the cross-sections are for the ionization of compounds that are not of interest here. Second, the majority of cross-sections are for ionizations using 70–75 eV electrons. Harrison et al.[129] observed that the 75-eV cross-sections for ionization of benzene and its alkylated homologs increase with increasing carbon number up to a carbon number of nine. Within the limits of data precision, the cross-sections were found to be both carbon number- and isomer-independent for C_9 and C_{10} homologs. In contrast, Alberti et al. observed that the 75-eV cross-sections for benzene and *n*-alkylated homologs containing up to 15 carbons increased with increasing alkyl carbon number by a constant amount per CH_2 group.[153] Munson et al. also observed that the 70-eV cross-sections for -6(H) series homologs containing 2–10 alkyl carbons increased linearly with increasing carbon number but were, within the limits of data precision, isomer independent for a given homolog.[157]

The molecular-structure dependence of the total cross-sections for the ionization of organic compounds by electrons with energies in the region of 70 eV has been considered in terms of atomic and molecular properties. The suggestion[149] that molecular cross-sections correspond to the sum of the atomic cross-sections of the molecules constituent atoms has been shown[129,150,158] to be incorrect for the general case. The additivity approach has been modified to correlate total cross-sections for ionization of several classes of organic compounds by electrons having energies in the region of 70 eV.[129,152,154] EI cross-sections at 70–75 eV and 11.6–11.8 eV photoionization cross-sections[159] for a number of classes of organic compounds have been correlated with molecular properties such as the sum of the atomic cross-sections, polarizability, diamagnetic susceptibility and molecular volumes.[129,149,150,152,154–157,159] These linear correlations are not general but are different for each class of compounds.

It is not, a priori, obvious that low-voltage cross-sections can be reliably predicted from 70–75 eV cross-sections. In this regard, the relative cross-sections for ionization of 41 low-molecular weight alkanes, alkyl chlorides, ketones, aldehydes, ethers and alcohols by 11.6–11.8-eV photons were found[159] to be linearly correlated with those[129,152] for ionization by 70–75-eV electrons. LVEI efficiencies have been reported for a number of relevant aromatic compounds.[160] Since the ionization efficiencies were analyzed in terms of relative ionization probabilities rather then relative cross-sections, the results are considered in the next section.

7.7 Low-voltage Relative Mole Sensitivities for Ionization of Aromatic Compounds

Relative mole sensitivities for LVEI of aromatic hydrocarbons are tabulated in columns 6 through 11 in Table 14.[11,47,72,125–127] Relative gram sensitivities were converted to relative mole sensitivities values using the

molecular masses in column 5 which were calculated from the numbers of carbons and hydrogens in columns 2 and 3, respectively. Specific Z values are listed in column 4. Conditions used in determining these sensitivities are given in the footnotes to Table 14.

The footnotes in Table 14 reveal that the relative sensitivities were not determined using a uniform set of experimental conditions. The effects of differences in experimental conditions must be considered in any evaluation of the reproducibility of relative mole sensitivities. Dempster and Nier-Johnson geometry mass spectrometers were used in a number of the sensitivity determinations. Thus, the agreement might be considered to be quite good given that the mass range in the former instrument is scanned by varying the accelerating voltage rather then the magnetic field.

The inlet and ion-source temperatures used in determining the sensitivities in set 2 were probably at least 250 °C.[16,125] Ion-source temperatures for sets 1 and 5 were not reported; see footnotes a and e. Although the magnitude of the effect is dependent upon the structure of the neutral molecule, the sensitivities would reflect to some degree the use of different inlet-system/ion-source temperatures in these experiments. An increase in the contribution of hot-band transitions to the ionization efficiency caused by an increase in temperature would tend to increase the sensitivity except that sensitivities tend to decrease with increasing temperature due to increases in the thermal-energy content of the neutral which tends to increase molecular-ion fragmentations. For example, we have observed that the m/z 106 to m/z 91 signal magnitude ratio from *p*-xylene decreases as the inlet and ion-source temperatures are increased from 250–325 °C. The effect of different inlet and ion-source temperatures on the internal energies of the neutral prior to ionization and, hence, on sensitivities is unknown.

The footnotes in Table 14 reveal the use of different procedures for establishing the electron energies used in these sensitivity determinations. Crable et al.[11] used a nominal electron energy of 8 eV which was reported to correspond to an actual energy of 10 eV. This difference arises from contact potentials within the ion source. Thus, the actual ionizing energies used by Del Bianco et al.[72] and by Schiller[47] were undoubtedly greater then the reported nominal values of 9.2 and 10.0 eV, respectively.

Table 14 Experimental relative mole sensitivities for low-voltage ionization of aromatic hydrocarbons C_nH_x

Hydrocarbon	Values of			Formula weight	Experimental $S(m)_{rel}$ values from set					
	n	x	$-Z$		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
Benzene	6	6	6	78.112	0.35	0.48		0.62		0.52 ± 0.03
Toluene	7	8	6	92.138		0.89		1.15		0.92 ± 0.04
Ethylbenzene	8	10	6	106.165	1	1	1	1	1	1
o-Xylene	8	10	6	106.165	1.51					1.28 ± 0.03
m-Xylene	8	10	6	106.165	1.69	1.44[g]		1.58[h]		1.41 ± 0.02
p-Xylene	8	10	6	106.165	1.73					1.45 ± 0.02
Propylbenzene	9	12	6	120.192	1.05	1.07	1.00 ± 0.04	0.98		1.02 ± 0.02
1-Methyl-3-ethylbenzene	9	12	6	120.192	1.97	1.57[g]	1.28 ± 0.03			1.50 ± 0.02
1-Methyl-4-ethylbenzene	9	12	6	120.192	2.09		1.33 ± 0.03			1.48 ± 0.02
Butylbenzene	10	14	6	134.218	1.18		0.94 ± 0.03			0.89 ± 0.02
Hexylbenzene	12	18	6	162.271			0.75 ± 0.07			0.71 ± 0.02
Octylbenzene	14	22	6	190.324	1.19					0.85 ± 0.02
Decylbenzene	16	26	6	218.378						0.88 ± 0.02
Hexadecylbenzene	22	38	6	302.537						0.95 ± 0.10
Nonyldecylbenzene	25	44	6	344.617						0.97 ± 0.10
Indan	9	10	8	118.176		1.33	1.24 ± 0.02	1.48		1.31 ± 0.02
Tetralin	10	12	8	132.202	1.95	1.17	1.11 ± 0.06	1.28		1.32 ± 0.05
Indene	9	8	10	116.160	2.76	1.84		1.75		2.05 ± 0.02
Octahydroanthracene	14	18	10	186.293	4.41[h]		1.80 ± 0.14[g]	1.97[g]		
Naphthalene	10	8	12	128.171	4.59	2.09	2.30 ± 0.09	2.55		2.39 ± 0.10
1-Methylnaphthalene	11	10	12	142.197	5.46	2.83	2.34 ± 0.16	3.19[h]		3.01 ± 0.15
2-Ethylnaphthalene	12	12	12	156.224		2.67[g]	2.39 ± 0.07	3.21		3.18 ± 0.18
1,5-Dimethylnaphthalene	12	12	12	156.224		3.13[g]	2.76 ± 0.06	3.68[g,h]		3.63 ± 0.18
2,3,6-Trimethylnaphthalene	13	13	12	170.250		1.74[g]	3.72 ± 0.05			
Biphenyl	12	10	14	154.208		4.03		1.57		2.59 ± 0.03
Acenaphthene	12	10	14	154.208		2.98	2.68 ± 0.03	3.41	3.14	3.71 ± 0.10
Acenaphthalene	12	8	16	152.192						2.62 ± 0.03

(*continued overleaf*)

Table 14 (*continued*)

Hydrocarbon	Values of			Formula weight	Experimental $S(m)_{rel}$ values from set					
	n	*x*	$-Z$		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
Fluorene	13	10	16	166.219		2.49	2.84 ± 0.09	2.07	3.54	3.32 ± 0.16
Anthracene	14	10	18	178.229	10.0			4.01		
						3.96[g]				
Phenanthrene	14	10	18	178.229	9.02		4.19 ± 0.35	3.34		4.65 ± 0.28
9-Methylanthracene	15	12	18	192.256			4.08 ± 0.10	4.04		
						4.28[g]				
1-Methylphenanthrene	15	12	18	192.256			4.22 ± 0.12			
Pyrene	16	10	22	202.251		4.76	3.38 ± 0.11	4.19		5.84 ± 0.41
Fluoranthene	16	10	22	202.255			4.02 ± 0.38	2.55	4.81	4.78 ± 0.05
1,2-Benzofluorene	17	12	22	216.277				3.95		
2,3-Benzofluorene	17	12	22	216.277			4.20 ± 0.06	3.50		6.35 ± 0.45
Chrysene	18	12	24	228.288		4.59		7.76	4.90	6.69 ± 0.65
3,4-Benzfluoranthene	20	12	28	252.309						
Benzo[*a*]pyrene	20	12	28	252.309						9.38 ± 0.94
Perylene	20	12	28	252.309					9.82	
1,12-Benzperylene	22	12	32	276.331					9.37	14.50 ± 2.63
Indeno(1,2,3-*c*,*d*)pyrene	22	12	32	276.331					11.56	

[a] (Reprinted with permission from G.F. Crable, G.L. Kearns, M.S. Norris, 'Low Voltage Mass Spectrometric Sensitivities of Aromatics', *Anal. Chem.*, **32**(1), 13–17 (1960). Copyright 1960, American Chemical Society.) Spectra obtained using CEC 21-103; repeller voltage = 1.5 V; nominal electron energy = 8.0 eV; calibrated electron energy = 10.0 eV; inlet = 175 °C; ion source temperature not specified.

[b] Values calculated from relative gram sensitivities from Aczel;[125] see also Lumpkin and Aczel.[16] Spectra obtained using a nominal electron energy that resulted in a *SM*(106) to *SM*(91) ratio from *p*-xylene of 100 : 1.

[c] Relative mole sensitivities from Scheppele and Sturm.[126] Spectra obtained using Kratos MS-50; resolving power = 11 000; repeller voltage maintained near block potential; electron energy set to produce a *SM*(106) to *SM*(91) ratio from *p*-xylene of 10 : 1; inlet = 300 °C; ion source = 300 °C.

[d] Relative mole sensitivities calculated from relative gram sensitivities. (Reprinted from A. Del Bianco, M. Zaninelli, E. Girardi, 'Determination of Transferable Hydrogen in Coal Liquefaction Solvents by Spectroscopic Methods', *Fuel*, **65**, 1062–1066, Copyright 1986, with permission from Elsevier Science.) Spectra obtained using Kratos MS-25Q; repeller voltage not specified; nominal electron energy = 9.2 eV; inlet = 250 °C; ion source = 240 °C.

[e] Values calculated from relative gram sensitivities. (Reprinted with permission from J.E. Schiller, 'Preparation and Low-voltage Mass Spectrometry Sensitivities of Methylated Polynuclear Aromatic Hydrocarbons', *Anal. Chem.*, **49**(8), 1260–1262 (1977). Copyright 1977, American Chemical Society), using a relative mole sensitivity for naphthalene equal to the average of values from sets 2, 3, 4 and 7. Spectra obtained using Kratos MS-30; repeller voltage maintained at the block potential; nominal electron energy = 10.0 eV; inlet = 325 °C; ion source temperature not specified.

[f] S.E. Scheppele, J.W. Miller.[127] Spectra obtained at 10 000–40 000 RP at 1000 s/decade; electron energy set to produce a *SM*(106) to *SM*(91) ratio of 20; inlet, transfer line, reentrance and ion source temperature were 300 °C.

[g] Isomer not specified.

[h] Average value.

The use of fixed values of the electron energy fails to take into account the effect of field penetrations and changes in the operating conditions of the ion source on the spread of energies in the electron beam in the region of ionization. In an attempt to minimize the effect of this phenomenon, Lumpkin and Aczel,[16,125] Scheppele and Sturm[126] and Scheppele and Miller[127] adjusted the electron energy to maintain a fixed *m*/*z* 106 to *m*/*z* 91 signal magnitude ratio from *p*-xylene. The use of signal magnitude ratios of 100 : 1,[125] 10 : 1[126] and 20 : 1[127] could account for some of the differences in sensitivities between sets 2, 3 and 6. It should be noted that we have routinely found that maintaining a fixed value of this ion abundance ratio requires decreases in the nominal value of the electron energy as the degree of source contamination increases for a given value of the ionizing (trap) current. This result is consistent with a broadening of the EED which in turn could result in variations in ion relative abundances and, hence, variations in relative sensitivities. The extent of acceptable ion-source contamination decreases with increasing trap current for increasing values of *SM*(106)/*SM*(91).

Deverse and King determined relative electron-impact ionization probabilities for the formation of stable molecular ions in the threshold region for a series of aromatic compounds.[160] The dependence of the relative mole sensitivities on the magnitude of the ionizing electron energy can be quantitatively estimated using these relative ionization probabilities. It is thus unfortunate that the work of Deverse and King appears to have been overlooked to date. Using a CEC 21-103C mass spectrometer, these authors determined the parent ion ionization-efficiency curves from near the ionization thresholds to about 12 eV. Equation (57) shows that the ion signal magnitude at each value of the electron energy is the sum of the ion signal magnitudes for formation

of various states of the molecular ion. The ''tails'' of the ionization efficiency curves were reported to be approximately 1.5 eV which from Figure 11 is consistent with a Maxwellian EED having a FWHM of at least 0.6 eV. At higher values of the electron energy, the ionization efficiency curves were observed to increase linearly with increasing ionization energy. The relative ionization probabilities were calculated on a mole basis from the linear portions of the molecular-ion ionization efficiency curves. The electron energy spread precluded detecting the formation of molecular ions in excited states. Consequently, the slopes effectively correspond to the sum of the ionization probabilities for formation of molecular ions in these states. Thus, from Equation (57) we obtain Equation (58) for calculating relative mole sensitivities

$$S(\text{mol})_{\text{rel}} = \frac{RTP_i(V - IP_i)}{RTP_{\text{ref}}(V - IP_{\text{ref}})} \tag{58}$$

where RTP_i and RTP_{ref} are the relative ionization probabilities for the ith and the reference hydrocarbons, respectively. Table 15 presents relative mole sensitivities in columns 7 through 11 calculated in 0.5-eV intervals from 9.5 through 11.5 for the hydrocarbons listed in column 1. These $S(\text{mol})_{\text{rel}}$ values were obtained using Equation (58) and the relative transition probabilities and ionization potentials in columns 5 and 6, respectively, of Table 15.

Table 15 Relative mole sensitivities for low-voltage ionization of aromatic hydrocarbons C_nH_x calculated as a function of electron energy from EI probabilities in the threshold region

Hydrocarbon	Values of			TP[a]	IP[b]	$S(\text{mol})_{\text{rel}}$ calculated at electron energy (eV) Values of				
	n	x	$-Z$			9.5	10.0	10.5	11.0	11.5
Benzene	6	6	6	200	9.25	0.223	0.399	0.474	0.516	0.542
Toluene	7	8	6	287	8.82	0.870	0.901	0.915	0.922	0.926
Ethylbenzene	8	10	6	303	8.76	1	1	1	1	1
o-Xylene	8	10	6	315	8.56	1.321	1.207	1.159	1.132	1.115
m-Xylene	8	10	6	361	8.56	1.513	1.384	1.328	1.298	1.278
p-Xylene	8	10	6	349	8.45	1.634	1.440	1.357	1.311	1.282
Propylbenzene	9	12	6	301	8.72	1.047	1.025	1.016	1.011	1.008
Isopropylbenzene	9	12	6	296	8.69	1.069	1.032	1.016	1.007	1.002
Butylbenzene	10	14	6	305	8.69	1.102	1.063	1.047	1.038	1.032
sec-Butylbenzene	10	14	6	289	8.68	1.057	1.015	0.998	0.988	0.982
tert-Butylbenzene	10	14	6	264	8.68	0.965	0.927	0.911	0.902	0.897
Pentylbenzene	11	16	6	313	8.68[c]	1.145	1.100	1.080	1.070	1.063
Hexylbenzene	12	18	6	311	8.68[c]	1.137	1.093	1.074	1.063	1.056
Indene	9	8	10	410	8.14	2.487	2.030	1.835	1.728	1.659
Indan	9	10	8	282	9.05	0.566	0.713	0.776	0.810	0.832
Tetralin	10	12	8	223	8.73	0.766	0.754	0.749	0.746	0.744
Naphthalene	10	8	12	364	8.14	2.208	1.802	1.629	1.534	1.473
1-Methylnaphthalene	11	10	12	445	7.96	3.056	2.416	2.144	1.993	1.897
2-Methylnaphthalene	11	10	12	453	7.96	3.111	2.460	2.182	2.029	1.932
Biphenyl	12	10	14	463	8.27	2.540	2.132	1.958	1.862	1.801
Acenaphthene	12	10	14	498	7.73	3.931	3.009	2.616	2.399	2.261
Acenaphthalene	12	8	16	473	8.02	3.122	2.493	2.225	2.077	1.983
Fluorene	13	10	16	515	7.93	3.606	2.837	2.510	2.329	2.215
Anthracene	14	10	18	677	7.41	6.310	4.667	3.968	3.581	3.335
Phenanthrene	14	10	18	781	7.91	5.538	4.344	3.837	3.556	3.377
Pyrene	16	10	22	759	7.72	6.025	4.606	4.002	3.668	3.456
Pyrene	16	10	22	759	7.31	7.413	5.434	4.592	4.126	3.831
Fluoranthene	16	10	22	809	7.80	6.134	4.737	4.143	3.814	3.605
Chrysene	18	12	24	811	8.01	5.389	4.295	3.830	3.573	3.409
Chrysene	18	12	24	811	7.72	6.438	4.921	4.276	3.919	3.692
Perylene	20	12	28	855	7.10	9.152	6.599	5.514	4.913	4.531

[a] Reprinted with permission from F.T. Deverse, A.B. King, 'Effect of Molecular Structure on the Ionization Probabilities of Aromatic Molecules', *J. Chem. Phys.*, **41**(12), 3833–3838 (1964). Copyright 1964, American Institute of Physics (*J. Chem. Phys.*). Relative transition probabilities (RTP) were reported on the basis of a value of 200 for benzene.

[b] Reprinted with permission from H.M. Rosenstock, K. Draxl, B.W. Steiner, J.T. Herron, 'Energetics of Gaseous Ions', *J. Phys. Chem. Ref. Data*, **6**, (Suppl. 1.), Copyright 1977, American Institute of Physics, American Chemical Society, & National Institute of Science and Technology (*J. Phys. Chem. Ref. Data*).

[c] Assumed values.

Comparison of the calculated relative mole sensitivities in Table 15 with the corresponding experimental values in Table 14 reveals that the different procedures employed in establishing the electron energies used in determining these latter values could account for a significant fraction of the apparent differences between a significant number of the experimental sensitivities. Inspection of Tables 14 and 15 reveals that the experimental $S(\text{mol})_{\text{rel}}$ values for hydrocarbons in set 1 having specific Z values less than or equal to -12(H), for benzene, chrysene and fluoranthene in set 4 and for pyrene in set 3 are exceptions that must reflect the occurrence of other phenomena.

The increase in senstivities with increasing molecular weight is seen in Table 14 to be generally largest in set 6. However, a potentially plausible explanation for at least a portion of this result is the use of a post-acceleration detector rather than an electron multiplier as used in the other studies given the fact that the gain of electron multipliers decreases with increasing ion m/z value.

The predicted $S(\text{mol})_{\text{rel}}$ values for pyrene and chrysene in Table 15 illustrate the effect of uncertainties in the ionization potentials on the magnitudes of the predicted sensitivities. As seen in column 6 of Table 15, the reported values for the ionization potentials of pyrene and chrysene range from 7.72 eV down to 7.31 eV and from 8.01 eV down to 7.72 eV, respectively. Columns 7 through 11 reveal that the lower ionization potential results in increased predicted $S(\text{mol})_{\text{rel}}$ values for pyrene with the increase decreasing from 23.0% at 9.5 eV to 10.9% at 11.5 eV. The corresponding increases for chrysene are 19.5% at 9.5 eV down to 8.3% at 11.5 eV.

The predicted sensitivities for indan and tetralin in Table 15 are seen to be significantly smaller than any of the experimental values in Table 14. Uncertainties in the ionization potentials can not account for these differences. The differences between the predicted and experimental sensitivities for these compounds is most likely to reflect phenomena such as errors in calculating the relative transition probabilities and/or greater degrees of molecular-ion fragmentation.

The *RTP* and *IP* values in Table 15 account for a number of anomalies that have been observed in attempts to correlate relative mole sensitivities for ionization of unalkylated aromatic hydrocarbons with quantities such as specific Z values and ionization potentials quantitatively for predictive purposes. For example, the *IP* for phenanthrene is about 0.25 eV greater than the *IP* for anthracene. The observation of similar relative mole sensitivities for these two hydrocarbons in Tables 14 and 15 reflects the fact that the larger *RTP* for phenanthrene essentially compensates for this difference in *IP*s. In Table 15, the predicted $S(\text{mol})_{\text{rel}}$ values for pyrene are seen to be larger or smaller than the predicted values for fluoranthene depending upon the *IP* assumed for the former compound. The experimental values in Table 14 indicate but do not prove that pyrene has the larger relative mole sensitivity. If additional experiments either confirm this indication or determine similar $S(\text{mol})_{\text{rel}}$ values for the two hydrocarbons then the result is another example of compensating differences in *RTP* and *IP* values unless the *RTP* values for either pyrene or fluoranthene or both are wrong. Although both biphenyl and acenaphthene are -14(H) series compounds, the experimental and predicted sensitivities for the latter are larger than those for the former. This result is consistent with the fact that the *RTP* and *IP* values for the acenaphthene are larger and smaller, respectively, than the corresponding values for biphenyl.

7.8 Summary

A number of conclusions and recommendations regarding the development of a universal database of low-voltage sensitivities are indicated by the contents of sections 7.1–7.7. The dependence of sensitivities for an agreed range of molecular structures should be ascertained and agreement should be reached regarding the acceptable range in inlet and ion source temperatures. A standard procedure for establishing the electron energy should be adopted. The simplest procedure is the attainment of a specified molecular ion to fragment ion relative abundance using a particular compound, for example, an m/z (106) to m/z (91) ratio of 25 : 1 from *p*-xylene. However, the differences in the electron energy distribution and the nominal value of the electron energy for attainment of a given relative abundance that can arise between mass spectrometers and for a given instrument as a function of the source and analyzer contamination will necessarily be reflected in the values for sensitivities to an extent presently unknown. Only rhenium filaments should be used.[8] The mass dependence of detector gain should be determined for each detector system and sensitivities should be corrected for any differences in these dependencies. It would be desirable to address the consequences of differences in the mass dependence of the analyzer transmission on sensitivities albeit to do so would undoubtedly be difficult. It should be determined to what extent sensitivities such as those in sets 1–6 in Table 14, excluding those discussed above, may reasonably approximate the repeatability and reproducibility that can be realized in the determination of a universal database of low-voltage sensitivities given the experimental variables that affect these values. Finally, publications should be required to specify the basis on which either absolute or relative sensitivities were determined, for example, mole, gram or volume, and the mathematical

operations performed in converting ion abundances to sensitivities.

ACKNOWLEDGMENTS

Development of computer programs for processing high resolution mass spectra by G.L. Nutter is gratefully acknowledged. The Research Resources Center at the University of Illinois at Chicago is thanked for supporting preparation of this article.

ABBREVIATIONS AND ACRONYMS

AEBP	Atmospheric Equivalent Boiling Points
AGHIS	All Glass Heated Inlet System
ASTM	American Society of Testing and Materials
CE	Charge Exchange
CI	Chemical Ionization
CIHRMS	Chemical Ionization High Resolution Mass Spectra
CND	Carbon Number Distribution
DBIS	Dynamic Batch Inlet System
DF	Double Focusing
EED	Electron Energy Distribution Function
EI	Electron Ionization
FAB	Fast Atom Bombardment
FDIE	First Derivative of the Ionization Efficiency
FI	Field Ionization
FIMS	Field Ionization Mass Spectrometric
FTICR	Fourier Transform Ion Cyclotron Resonance
FWHM	Full Width Half Maximum
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HRLVEIMS	High-resolution Low-voltage Electron Ionization Mass Spectra
HRMS	High-resolution Mass Spectrometric
IR	Infrared
LVEI	Low-voltage Electron Ionization
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
P(E)	Internal Energy Distribution Function
P(ED)	Energy Deposition Function
PMD	Probe Microdistillation
QET	Quasi-equilibrium Theory
RHS	Right Hand Side
RIA	Relative Ion Abundance
RMS	Root Mean Square
RP	Resolving Power
RRKM	Rice–Ramsperger–Kassel–Marcus
SDIE	Second Derivative of the Ionization Efficiency
TDNOCI	Townsend Discharge Nitric Oxide Chemical Ionization

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Lube Products. Molecular Characterization of Base Oils • Lubricant Base Oils: Analysis and Characterization of • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Nuclear Magnetic Resonance Characterization of Petroleum • Petroleum Residues, Characterization of

REFERENCES

1. R.A. Brown, 'Compound Types in Gasoline by Mass Spectrometer Analysis', *Anal. Chem.*, **23**(3), 430–437 (1951).
2. R.J. Clerc, A. Hood, M.J. O'Neal, Jr, 'Mass Spectrometric Analysis of High Molecular Weight, Saturated Hydrocarbons', *Anal. Chem.*, **27**(6), 868–875 (1955).
3. S.H. Hastings, B.H. Johnson, H.E. Lumpkin, 'Analysis of the Aromatic Fraction of Virgin Gas Oils by Mass Spectrometer', *Anal. Chem.*, **28**(8), 1243–1247 (1956).
4. H.E. Lumpkin, 'Determination of Saturated Hydrocarbons in Heavy Petroleum Fractions by Mass Spectrometry', *Anal. Chem.*, **28**(6), 1946–1948 (1956).
5. G.F. Crable, N.D. Coggeshall, 'Application of Total Ionization Principles to Mass Spectrometric Analysis', *Anal. Chem.*, **30**(3), 310–313 (1958).
6. H.E. Lumpkin, 'Low Voltage Techniques in High Molecular Weight Mass Spectrometry', *Anal. Chem.*, **30**(3), 321–325 (1958).
7. A. Hood, 'Standardization of Mass Spectra by Means of Total Ion Intensity', *Anal. Chem.*, **30**(7), 1218–1220 (1958).
8. A.G. Sharkey, Jr, G. Wood, J.L. Shultz, I. Wender, R.A. Friedel, 'Analysis of Liquid Products from Coal Hydrogenation by Mass Spectrometry: Use of Rhenium Filaments and Low Ionizing Voltage', *Fuel*, **38**, 315–328 (1959).

9. A. Hood, M.J. O'Neal, 'Status of Application of Mass Spectrometry to Heavy Oil Analysis', *Adv. Mass Spectrom.*, **1**, 175–192 (1959).
10. G.L. Kearns, N.C. Maranowski, G.F. Crable, 'Analysis of Petroleum Products in the C_{12} to C_{20} Range: Application of FIA Separatory and Low Voltage Mass Spectrometric Techniques', *Anal. Chem.*, **31**(10), 1646–1651 (1959).
11. G.F. Crable, G.L. Kearns, M.S. Norris, 'Low Voltage Mass Spectrometric Sensitivities of Aromatics', *Anal. Chem.*, **32**(1), 13–17 (1960).
12. E.G. Carlson, G.T. Paulissen, R.H. Hunt, M.J. O'Neal, Jr, 'High Resolution Mass Spectrometry: Interpretation of Spectra of Petroleum Fractions', *Anal. Chem.*, **32**(11), 1489–1494 (1960).
13. A.G. Sharkey, Jr, J.L. Shultz, R.A. Friedel, 'Application of Low Ionizing Voltage Mass Spectrometry to Oils Derived from Coal', *Fuel*, **41**, 359–371 (1962).
14. E. Kendrick, 'A Mass Scale Based on $CH_2 = 14.000$ for High Resolution Mass Spectrometry of Organic Compounds', *Anal. Chem.*, **35**(13), 2146–2154 (1963).
15. A.G. Sharkey, Jr, J.L. Shultz, R.A. Friedel, 'Advances in Coal Spectrometry. Mass Spectrometry', Washington, US Department of the Interior, Bureau of Mines, RI 6318, 1963.
16. H.E. Lumpkin, T. Aczel, 'Low-voltage Sensitivities of Aromatic Compounds', *Anal. Chem.*, **36**(1), 181–184 (1964).
17. H.E. Lumpkin, 'Analysis of a Trinuclear Aromatic Petroleum Fraction by High Resolution Mass Spectrometry', *Anal. Chem.*, **36**(13), 2399–2401 (1964).
18. D.D. Tunnicliff, P.A. Wadsworth, D.O. Schissler, 'Ordered Exact Mass and Abundance Tables for Mass Spectrometry', *Anal. Chem.*, **37**(4), 543–548 (1965).
19. M. Gomez, W.S. Landers, J.L. Shultz, A.G. Sharkey, Jr, 'Distribution of Phenols in Low-temperature Tar from Low-rank Coals', Washington, US Department of the Interior, Bureau of Mines, RI 6586, 1965.
20. W.K. Reid, W.L. Mead, K.M. Bowen, 'Combination of High Resolution and Low Ionizing Voltages in the Determination of Hydrocarbon and Sulfur Compound Types in Petroleum Fractions Using Mass Spectrometry', *Adv. Mass Spectrom.*, **3**, 731–745 (1966).
21. A.G. Sharkey, Jr, J.L. Shultz, R.A. Friedel, 'Mass Spectra of Pyrolyzates of Several Aromatic Structures Identified in Coal Extracts', *Carbon*, **4**, 365–374 (1966).
22. A.G. Sharkey, Jr, J.L. Shultz, R.A. Friedel, 'Analytical Methods in Mass Spectrometry', Washington, US Department of the Interior, Bureau of Mines, Bulletin 634, 1967.
23. J.L. Shultz, R.A. Friedel, A.G. Sharkey, Jr, 'Mass Spectrometric Analyses of Coal-tar Distillates and Residues', Washington, US Department of the Interior, Bureau of Mines, RI 7000, 1967.
24. A.G. Sharkey, Jr, J.L. Shultz, T. Kessler, R.A. Friedel, 'High Resolution Mass Spectrometry Investigation of Coal-tar Pitch', Proceedings 15th Annual Conference Mass Spectrometry Allied Topics, American Society for Mass Spectrometry Denver, 443–446, 1967.
25. G.G. Wanless, G.A. Glock, Jr, 'Simplified Mass Spectra by Means of a Dual Field Ionization/Electron Impact Source', *Anal. Chem.*, **39**(1), 2–13 (1967).
26. B.H. Johnson, T. Aczel, 'Analysis of Complex Mixtures of Aromatic Compounds by High Resolution Mass Spectrometry at Low Ionizing Voltages', *Anal. Chem.*, **39**(6), 682–685 (1967).
27. E.J. Gallegos, J.W. Green, L.P. Lindeman, R.L. LeTourneau, R.M. Teeter, 'Petroleum Group-type Analysis by High Resolution Mass Spectrometry', *Anal. Chem.*, **39**(14), 1833–1838 (1967).
28. E.J. Gallegos, 'High Resolution Mass Spectrometry of Asphalt Fractions', Proceedings Seventh World Petroleum Congress, **4**, 249–260, 1967.
29. W.L. Mead, 'Field Ionization Mass Spectrometry of Heavy Petroleum Fractions: Waxes', *Anal. Chem.*, **40**(4), 743–747 (1968).
30. C.J. Robinson, G.L. Cook, 'Low-resolution Mass Spectrometric Determination of Aromatic Fractions from Petroleum', *Anal. Chem.*, **41**(12), 1548–1554 (1968).
31. T. Kessler, R. Raymond, A.G. Sharkey, Jr, 'Composition of Pyridine Extracts from Reduced and Untreated Coals as Determined by High-resolution Mass Spectrometry', *Fuel*, **48**, 179–186 (1969).
32. T. Aczel, H.E. Lumpkin, 'Simulated Distillation by High Resolution Low Voltage Mass Spectrometry', Proceedings 18th Annual Conference Mass Spectrometry Allied Topics, American Society for Mass Spectrometry, B37–B39, 1970.
33. T. Aczel, D.E. Allan, J.H. Harding, E.A. Knipp, 'Computer Techniques for Quantitative High Resolution Mass Spectral Analysis of Complex Hydrocarbon Mixtures', *Anal. Chem.*, **42**(3), 341–347 (1970).
34. T. Aczel, H.E. Lumpkin, 'Quantitative Analysis of Nitrogen Compounds', Proceedings 19th Annual Conference Mass Spectrometry Allied Topics, American Society for Mass Spectrometry, Atlanta, 328–330, 1971.
35. T. Aczel, 'Applications of High Resolution Mass Spectrometry to the Analysis of Complex Aromatic Mixtures Derived from Petroleum and Coal', *Rev. Anal. Chem.*, **1**, 226–261 (1971).
36. K.G. Hippe, H.D. Becky, 'Quantitative Group Type Analysis of Benzene-like Hydrocarbon Mixtures with the Field Ion Mass Spectrometer', *Erdol, Kohle, Erdgas, Petrochem.*, **24**(10), 620–626 (1971).
37. J.L. Shultz, A.G. Sharkey, Jr, R.A. Brown, 'Determination of Mass Spectrometric Sensitivity Data for Hydroaromatic Compounds', *Anal. Chem.*, **44**(8), 1486–1487 (1972).

38. H. Pichler, A. Herlan, 'High Resolution Mass Spectrometric Investigation of a High-temperature Coal-tar. I. Introduction, Experiments, Sensitivities', *Erdol, Kohle, Erdgas, Petrochem.*, **26**(6), 338–341 (1973).
39. J.L. Schultz, T. Kessler, R.A. Friedel, A.G. Sharkey, Jr, 'High-resolution Mass Spectrometric Investigation of Heteroatom Species in Coal-carbonization Products', *Fuel*, **52**, 242–246 (1973).
40. J.T. Swansiger, F.E. Dickson, H.T. Best, 'Liquid Coal Compositional Analysis by Mass Spectrometry', *Anal. Chem.*, **46**(6), 730–734 (1974).
41. H.E. Lumpkin, R.M. Elliott, S. Evans, D. Hazelby, W.A. Wolstenholme, 'The Application of Ultra High Resolution Dynamic Scanning to the Analysis of Sulfur-containing Petrochemicals', Proceedings 23rd Conference Mass Spectrometry Allied Topics, American Society for Mass Spectrometry, Dallas, TX, May 25–30, 235–237, 1975.
42. M. Ryska, M. Kuras, J. Mostecky, 'Phenomenology of Adsorption Processes on Emitters in the Field Ionization of Hydrocarbon Mixtures', *Int. J. Mass Spectrom. Ion Processes*, **16**, 257–267 (1975).
43. M. Kuras, M. Ryska, J. Mostecky, 'Analysis of Saturated Hydrocarbons by Field Ionization Mass Spectrometry', *Anal. Chem.*, **48**(1), 196–198 (1976).
44. D. Severin, 'Molecular Ion Mass Spectrometry for the Analysis of High-boiling and Non-boiling Hydrocarbon Mixtures', *Erdol, Kohle, Erdgas, Petrochem.*, **29**(1), 13–18 (1976).
45. A.W. Peters, J.G. Bendoraitis, 'High Resolution Mass Spectrographic Method for the Analysis of Nitrogen- and Oxygen-containing Material Derived from Petroleum', *Anal. Chem.*, **48**(7), 968–973 (1976).
46. S.E. Scheppele, P.L. Grizzle, G.J. Greenwood, T.D. Marriott, N.B. Perreira, 'Determination of Field Ionization Sensitivities for the Analysis of Coal-derived Liquids and Their Correlation with Low-voltage Electron Impact Relative Sensitivities', *Anal. Chem.*, **48**(14), 2105–2113 (1976).
47. J.E. Schiller, 'Preparation and Low-voltage Mass Spectrometry Sensitivities of Methylated Polynuclear Aromatic Hydrocarbons', *Anal. Chem.*, **49**(8), 1260–1262 (1977).
48. T. Aczel, R.B. Williams, R.J. Pancirov, J.H. Karchmer, 'Chemical Properties of Synthoil Products and Feeds', US Department of Energy, MERC-8007, 1977.
49. S.E. Scheppele, G.J. Greenwood, P.A. Benson, 'Field Ionization Sensitivities for Analysis of Coal-derived Liquids Determined as a Function of Ion-source Temperature and Binary Mixture Composition', *Anal. Chem.*, **49**(12), 1847–1850 (1977).
50. J.E. Dooley, C.J. Thompson, S.E. Scheppele, 'Characterizing Syncrudes from Coal', in *Analytical Methods for Coal and Coal Products*, ed. C. Karr, Jr, Academic Press, New York, 467–498, Vol. I, 1978.
51. S.E. Scheppele, C.S. Hsu, T.D. Marriott, P.A. Benson, K.N. Detwiler, N.B. Perreira, 'Field-ionization Relative Sensitivities for the Analysis of Saturated Hydrocarbons from Fossil-energy-related Materials', *Int J. Mass Spectrom. Ion Processes*, **28**, 335–346 (1978).
52. T. Aczel, R.B. Williams, R.A. Brown, R.J. Pancirov, 'Chemical Characterization of Synthoil Feeds and Products', in *Analytical Methods for Coal and Coal Products*, ed. C. Karr, Jr, Academic Press, New York, 499–540, Vol. I, 1978.
53. T. Aczel, H.E. Lumpkin, *Mass Spectrometric Analysis of Heterocompounds in Coal Extracts and Coal-liquefaction Products*, eds. M.L. Gorbaty, B.M. Harney, Advances Chemistry Series 179, American Chemical Society, Washington, DC, 2–24, 1979.
54. M. Ryska, M. Kuras, J. Mostecky, 'Study of Bimolecular Reactions on the Emitter in Field Ionization of Alkane–Alkyne Mixtures', *Int J. Mass Spectrom. Ion Processes*, **29**, 39–46 (1979).
55. R.A. Heppner, J.M. Knox, 'Computer Analysis of High Resolution Mass Spectral Data from Mixtures of Organic Compounds', *Chem. Biomed. Environ. Instrument.*, **11**(5&6), 425–446 (1981).
56. J. Van Katwijk, 'Improved Evaluation of High Resolution Mass Spectra Through the Investigation of Peak Shapes and Isotopic Contributions. I. Simulation of Peak Shapes', *Int. J. Mass Spectrom. Ion Processes*, **39**, 273–286 (1981).
57. J. Van Katwijk, 'Improved Evaluation of High Resolution Mass Spectra Through the Investigation of Peak Shapes and Isotopic Contributions. II. Interpretation of Limited Series of Peaks', *Int. J. Mass Spectrom. Ion Processes*, **39**, 287–310 (1981).
58. S.E. Buttrill, Jr, 'Analysis of Middle Distillate Fuels by Field Ionization Mass Spectrometry', Final Technical Report for Research Performed under Naval Research Laboratory Contract N00173-79-C-0462, 1981.
59. T. Yoshida, Y. Maekawa, T. Higuchi, E. Kubota, Y. Itagaki, S. Yokoyama, 'The Application of Field Ionization and Field Desorption Mass Spectrometry to the Analysis of Coal-derived Oil', *Bull. Chem. Soc. Jpn.*, **54**, 1171–1175 (1981).
60. S.E. Scheppele, G.J. Greenwood, R.J. Pancirov, T.R. Ashe, 'Chemical Composition of Raw and Upgraded Anthracene Oil and the Chemistry of Coal-liquids Upgrading', in *Upgrading Coal Liquids*, ed. R.F. Sullivan, ACS Symposium Series 156, American Chemical Society, Washington, DC, 39–73, Vol. 156, 1981.
61. S.E. Scheppele, P.A. Benson, G.J. Greenwood, Q.G. Grindstaff, T. Aczel, B.F. Beier, 'Consequences of the Mass Spectrometric and Infrared Analysis of Oils and Asphaltenes for the Chemistry of Coal Liquefaction', in *Chemistry of Asphaltenes*, eds. J.W. Bunger, N.C. Li., Advances Chemistry Series

195, American Chemical Society, Washington, DC, 53–82, 1981.
62. L.R. Schronk, R.D. Grigsby, S.E. Scheppele, 'Probe Microdistillation/Mass Spectrometry in the Analysis of High-boiling Petroleum Distillates', *Anal. Chem.*, **54**(4), 748–755 (1982).
63. R.D. Grigsby, S.E. Scheppele, Q.G. Grindstaff, G.P. Sturm, Jr, L.C.E. Taylor, H. Tudge, C. Wakefield, S. Evans, 'Evaluation of Fast Atom Bombardment Mass Spectrometry for Identification of Nitrogen-containing Compounds in Fossil Fuels', *Anal. Chem.*, **54**(7), 1108–1113 (1982).
64. S.E. Scheppele, Q.G. Grindstaff, R.D. Grigsby, K.C. Chung, C.S. Hwang, T.D. Marriott, L.R. Schronk, 'Automated Probe Microdistillation/Mass Spectrometry for the Analysis of High-molecular Weight Compounds in Fossil Fuels', United States Department of Energy, National Technical Information Service, US Department of Commerce, Springfield, Virginia, DOE/BETC/TPR-82/1, 1982.
65. S.E. Scheppele, K.C. Chung, C.S. Hwang, 'Computer Processing of Mass Spectral Data. I. Assignment of Formulas to Experimental Masses; Chemical and Mathematical Principles', *Int J. Mass Spectrom. Ion Processes*, **49**, 143–178 (1983).
66. S.E. Scheppele, Q.G. Grindstaff, R.D. Grigsby, S.R. McDonald, C.S. Hwang, 'Computer Processing of Mass Spectral Data. II. Assignment of Formulas to Experimental Masses Illustrating Program FZM; Effects of Electron Mass and Functional Form of Exponential Scan Law on Accuracy of Mass Measurement', *Int J. Mass Spectrom. Ion Processes*, **49**, 179–209 (1983).
67. J.D. Payzant, A.M. Hogg, D.S. Montgomery, O.P. Strausz, 'A Field Ionization Mass Spectrometric Study of the Maltene Fraction of Athabasca Bitumen. Part I. The Saturates', *AOSTRA J. Res.*, **1**(3), 175–182 (1985).
68. J.D. Payzant, A.M. Hogg, D.S. Montgomery, O.P. Strausz, 'A Field Ionization Mass Spectrometric Study of the Maltene Fraction of Athabasca Bitumen. Part II. The Aromatics', *AOSTRA J. Res.*, **1**(3), 183–202 (1985).
69. J.D. Payzant, A.M. Hogg, D.S. Montgomery, O.P. Strausz, 'A Field Ionization Mass Spectrometric Study of the Maltene Fraction of Athabasca Bitumen. Part III. The Polars', *AOSTRA J. Res.*, **1**(3), 203–210 (1985).
70. R.M. Teeter, 'High-resolution Mass Spectrometry for Type Analysis of Complex Hydrocarbon Mixtures', *Mass Spectrom. Rev.*, **4**, 123–143 (1985).
71. R.H. Harrison, S.E. Scheppele, G.P. Sturm, Jr, P.L. Grizzle, 'Solubility of Hydrogen in Well-defined Coal Liquids', *Chem. Eng. Data*, **30**(2), 183–189 (1985).
72. A. Del Bianco, M. Zaninelli, E. Girardi, 'Determination of Transferable Hydrogen in Coal Liquefaction Solvents by Spectroscopic Methods', *Fuel*, **65**, 1062–1066 (1986).
73. E.J. Gallegos, 'Gas Chromatographic/Mass Spectrometric Analysis of Paraho Shale Oil <260 °C (500 °F)', in *Mass Spectrometric Characterization of Shale Oils*, ed. T. Aczel, ASTM Special Technical Publication 902 Philadelphia, 5–26, 1986.
74. S.E. Scheppele, Q.G. Grindstaff, P.L. Grizzle, 'Determination of Field Ionization Sensitivities for Saturated Hydrocarbons in the Presence of Acetone and Their Application to the Analysis of Shale Oil Saturates', in *Mass Spectrometric Characterization of Shale Oils*, ed. T. Aczel, ASTM Special Technical Publication 902 Philadelphia, 49–80, 1986.
75. C.E. Schmidt, R.F. Sprecher, R.D. Batts, 'Low-voltage, High-resolution Mass Spectrometric Methods for Fuel Analysis: Application To Coal Distillates', *Anal. Chem.*, **59**(17), 2027–2033 (1987).
76. P.H. Wolf, D. Severin, 'Development of a Quantitative FI-MS Technique for the Characterization of High- and Nonboiling Saturated Hydrocarbon Mixtures', *Fresenius' Z. Anal. Chem.*, **329**, 35–38 (1987).
77. M.M. Boduszynski, 'Composition of Heavy Petroleums. 2. Molecular Characterization', *Energy Fuels*, **2**(5), 597–613 (1987).
78. I. Dzidic, A.C. Somerville, J.C. Raia, H.V. Hart, 'Determination of Naphthenic Acids in California Crudes and Refinery Wastewaters by Fluoride Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **60**(13), 1318–1323 (1988).
79. C.E. Schmidt, R.F. Sprecher, 'Analysis of Coal-derived Materials with Low Volatility by Low-voltage, High Resolution Mass Spectrometry in Conjunction with Direct-insertion-probe Techniques', in *Novel Techniques in Fossil Fuel Mass Spectrometry*, eds. T.R. Ashe, K.V. Wood, ASTM Special Technical Publication 1019 Philadelphia, 116–132, 1989.
80. B.D. Batts, J.E. Batts, 'Mass Spectrometry of Coal Liquids', in *Spectroscopic Analysis of Coal Liquids*, ed. J.R. Kershaw, Elsevier, New York, 61–128, 1989.
81. C. Allgood, Y.C. Ma, B. Munson, 'Quantitation Using Benzene in Gas Chromatography/Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **63**(7), 721–725 (1990).
82. R.E. Pauls, M.E. Bambacht, C. Bradley, S.E. Scheppele, D.C. Cronauer, 'Distribution and Characterization of Phenolics in Distillates Derived from Two-stage Coal Liquefaction', *Energy Fuels*, **4**(3), 236–242 (1990).
83. S.E. Scheppele, G.L. Nutter, D.L. Parisi, 'Optimizing Performance of Double Focusing Mass Spectrometers. 1. Determination and Use of Total Aberrations', *Energy Fuels*, **5**(3), 360–370 (1991).
84. T-P. Fan, 'Characterization of Naphthenic Acids in Petroleum by Fast Atom Bombardment Mass Spectrometry', *Energy Fuels*, **5**(3), 371–375 (1991).
85. K.L. Chasey, T. Aczel, 'Polycyclic Aromatic Structure Distributions by High Resolution Mass spectrometry', *Energy Fuels*, **5**(3), 386–394 (1991).
86. C.S. Hsu, M.A. McLean, K. Qian, T. Aczel, S.C. Blum, W.N. Olmstead, L.H. Kaplan, W.K. Robbins, W.W.

Schulz, On-line Liquid Chromatography/Mass Spectrometry for Heavy Hydrocarbon Characterization', *Energy Fuels*, **5**(3), 395–398 (1991).
87. I. Dzidic, H.A. Petersen, P.A. Wadsworth, H.V. Hart, 'Townsend Discharge Nitric Oxide Chemical Ionization Gas Chromatography/Mass Spectrometry for Hydrocarbon Analysis of Middle Distillates, *Anal. Chem.*, **64**(19), 2227–2232 (1992).
88. K.N. Qian, C.S. Hsu, 'Molecular-transformation in Hydrotreating Processes Studied by Online Liquid Chromatography Mass Spectrometry', *Anal. Chem.*, **64**(20), 2327–2333 (1992).
89. Y. Sugimoto, Y. Miki, Y. Niwa, K. Tunnukij, S. Yamadaya, 'Characterization of Compound-class Fractions of Vacuum Gas Oil by Using Field Ionization Mass Spectrometry', *Sekiyu Gakkaishi*, **35**(4), 349–352 (1992).
90. C.S. Hsu, K. Qian, 'CS_2 Charge Exchange as a Low-energy Ionization Technique for Hydrocarbon Characterization', *Anal. Chem.*, **65**(6), 767–771 (1993).
91. C.S. Hsu, Z.M. Liang, J.E. Campana, 'Hydrocarbon Characterization by Ultrahigh-resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **66**(6), 850–855 (1994).
92. S. Gaun, A.G. Marshall, S.E. Scheppele, 'Resolution and Chemical Formula Identification of Aromatic Hydrocarbons and Aromatic Compounds Containing Sulfur, Nitrogen and Oxygen in Petroleum Distillates and Refinery Streams', *Anal. Chem.*, **68**(1), 46–71 (1996).
93. G.A. Veloski, R.J. Lynn, R.F. Sprecher, 'Characterization of Nitrogen-containing Species in Coal- and Petroleum-derived Products by Ammonia Chemical Ionization High Resolution Mass Spectrometry', *Energy Fuels*, **11**(1), 137–143 (1997).
94. R.D. Grigsby, J.B. Green, 'High-resolution Mass Spectrometric Analysis of a Vanadyl Porphyrin Fraction Isolated from the >700 °C Resid of Cerro Negro Heavy Petroleum', *Energy Fuels*, **11**(3), 602–609 (1997).
95. Z. Liang, C.S. Hsu, 'Molecular Speciation of Saturates by On-line Liquid Chromatography Field Ionization Mass Spectrometry', *Energy Fuels*, **12**(3), 637–643 (1998).
96. R.P. Rogers, F.M. White, D.G. McIntosh, A.G. Marshall, '5.6 Telsa Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for Analysis of Volatile Complex Mixtures', *Rev. Sci. Instrum.*, **69**(6), 2278–2284 (1998).
97. R.P. Rogers, F.M. White, C.L. Hendrickson, A.G. Marshall, K.V. Andersen, 'Resolution, Elemental Composition, and Simultaneous Monitoring by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Organosulfur Species Before and After Diesel Fuel Processing', *Anal. Chem.*, **70**(22), 4743–4750 (1998).
98. S.G. Roussis, 'Exhaustive Determination of Hydrocarbon Compound Type Distributions by High Resolution Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **13**(11), 1031–1051 (1999).
99. P.A. Benson, S.E. Scheppele, G.J. Greenwood, T. Aczel, Q.G. Grindstaff, B.F. Beier, 'Consequences of the Mass Spectrometric and Infrared Analysis of Oils and Asphaltenes for the Chemistry of Coal Liquefaction', United States Department of Energy Report FE 2537-10, Dist. Cat. UC-90d, National Technical Information Service, US Department of Commerce, Springfield, Virginia, 1979.
100. H.M. Grubb, S. Meyerson, 'Mass Spectra of Alkylbenzenes', in *Mass Spectrometry of Organic Ions*, ed. F.W. McLafferty, Academic Press, New York, 453–527, 1963.
101. A.B. King, M.R. Basila, 'Mass Spectrometric Determination of the Isomer Distribution of Carbon-16 to Carbon-22 Phenyl-*n*-alkanes', *Anal. Chem.*, **37**(6), 734–738 (1965).
102. Standard Test Method for Distillation of Crude Petroleum (15-Theoretical Plate Column), ASTM Method D 2892, Annual Book of ASTM Standards, American Society for Testing and Materials, Volume 5.02, 1999.
103. Standard Test Method for Distillation of Petroleum at Reduced Pressures, ASTM Method D 1160, Annual Book of ASTM Standards, American Society for Testing and Materials, Volume 5.01, 1999.
104. H.J. Coleman, J.E. Dooley, D.E. Hirsch, C.J. Thompson, 'Compositional Studies of a High-boiling 370–535 °C Distillate from Prudhoe Bay, Alaska, Crude Oil', *Anal. Chem.*, **45**(9), 1724–1734 (1973).
105. M.M. Boduszynski, 'Composition of Heavy Petroleums. 1. Molecular Weight, Hydrogen Deficiency, and Heteroatom Concentration as a Function of Atmospheric Equivalent Boiling Point up to 1400 °F (760 °C)', *Energy Fuels*, **1**(1), 2–11 (1987).
106. M.J. Mima, H. Schultz, W.E. McKinstry, 'Method for the Determination of Benzene Insolubles, Asphaltenes and Oils in Coal-derived Liquids', in *Analytical Methods for Coal and Coal Products*, ed. C. Karr, Jr, Academic Press, New York, 557–568, Vol. I, 1978.
107. D.M. Jewell, J.H. Weber, J.W. Bunger, H. Plancher, D.R. Latham, 'Ion-exchange, Coordination, and Adsorption Chromatographic Separation of Heavy-end Petroleum Distillates', *Anal. Chem.*, **44**(8), 1391–1395 (1972).
108. J.B. Green, R.J. Hoff, P.W. Woodward, L.L. Stevens, 'Separation of Liquid Fossil Fuels into Acid, Base and Neutral Concentrates. 1. An Improved Nonaqueous Ion Exchange Method', *Fuel*, **63**, 1290–1301 (1986).
109. D.E. Hirsch, R.L. Hopkins, H.J. Coleman, F.O. Cotton, C.J. Thompson, 'Separation of High-boiling Petroleum Distillates Using Gradient Elution Through Dual-packed (Silica Gel-alumina Gel) Adsorption Columns', *Anal. Chem.*, **44**(6), 915–919 (1972).

110. P.L. Grizzle, J.S. Thomson, 'Liquid Chromatographic Separation of Aromatic Hydrocarbons with Chemically Bonded (2,4-Dinitroanilinopropyl)silica', *Anal. Chem.*, **54**(7), 1071–1078 (1982).
111. J.S. Thomson, J.W. Reynolds, 'Separation of Aromatic Hydrocarbons Using Bonded-phase Charge-transfer Liquid Chromatography', *Anal. Chem.*, **56**(13), 2434–2441 (1983).
112. H.E. Lumpkin, G.R. Taylor, 'A Solids Inlet System for a Mass Spectrometer', *Anal. Chem.*, **33**(3), 476–477 (1961).
113. L. Peterson, 'Mass Spectrometer All-glass Heated Inlet', *Anal. Chem.*, **34**(13), 1850–1851 (1962).
114. G. Stafford, T.D. Morgan, R.J. Brunfeldt, 'A Mass Spectrometer All-glass Heated Inlet', *Int. J. Mass Spectrom. Ion Processes*, **1**, 87–92 (1968).
115. J. Shen, M.P.T. Bradley, 'Versatile Mass Spectrometric Inlet System for High Boiling Liquid Samples and for Thermoanalysis of Nonvolatile Materials', *Anal. Chem.*, **48**(14), 2291–2292 (1976).
116. E.J. Gallegos, E.C. Pazzi, 'A Dynamic Batch Inlet for Group Type Analysis by Mass Spectrometry', Proceedings 39th ASMS Conference Mass Spectrometry Allied Topics, Nashville, TN, 1091–1092, 1991.
117. S.G. Roussis, A.S. Cameron, 'Simplified Hydrocarbon Compound Type Analysis Using a Dynamic Batch Inlet System Coupled to a Mass Spectrometer, *Energy Fuels*, **11**(4), 879–886 (1997).
118. J.R. Partington, *An Advanced Treatise on Physical Chemistry, Vol. I, Fundamental Principles, the Properties of Gases*, John Wiley, New York, Section III, 1962.
119. R.D. Grigsby, C.O. Hansen, D.G. Mannering, W.G. Fox, R.H. Cole, 'Analysis of Composite Spectra of Mixtures Subjected to Micromolecular Distillation in a Mass Spectrometer', *Anal. Chem.*, **43**, 1135 (1971).
120. D.L. Parisi, S.E. Scheppele, unpublished results.
121. S.E. Scheppele, 'Predicting Optimum Resolving Powers for Homolog-type Analysis of Fossil Fuel Distillates', Proc. 44th ASMS Conference Mass Spectrometry Allied Topics, Portland, OR, 746, 1996.
122. J.H. Lederberg, private communication.
123. B.V. Ioffe, D.A. Vitenberg, I.G. Zenkevich, 'Calculation of Ion Composition in Organic High-resolution Mass Spectrometry', *Org. Mass Spectrom.*, **28**(8), 907–913 (1993).
124. A.G. Ferrige, M.J. Seddon, J. Skilling, N. Ordsmith, 'The Application of ''MaxEnt'' to High Resolution Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 765–770 (1992).
125. T. Aczel, private communication.
126. S.E. Scheppele, G.P. Sturm, Jr, Liquid Fossil Fuel Technology, Bartlesville Energy Technology Center, US Department of Energy, Quarterly Report 83/2, 37–39, 1983.
127. S.E. Scheppele, J.W. Miller, unpublished results.
128. S. Meyerson, H.M. Grubb, R.W. Vander Haar, 'Effects of Isotopic Labeling and Molecular Structure on Probability of Ionization by Electron Impact', *J. Chem. Phys.*, **39**(6), 1445–1446 (1963).
129. A.G. Harrison, E.G. Jones, S.K. Gupta, G.P. Nagy, 'Total Cross-sections for Ionization by Electron Impact', *Can. J. Chem.*, **44**, 1967–1973 (1966).
130. S. Geltman, 'Theory of Ionization Probability Near Threshold', *Phys. Rev.*, **102**(1), 171–179 (1956).
131. I. Kang, I.W.D. Foland, 'Theory of Electron Impact Excitation and Ionization of Atoms and Ions', *Phys. Rev.*, **164**(1), 122–130 (1967).
132. J.D. Morrison, 'Electron Impact Spectroscopy', *Rev. Pure Appl. Chem.*, **5**(1), 22–59 (1955).
133. F.P. Lossing, G.P. Semeluk, 'Threshold Ionization Efficiency Curves for Monoenergetic Electron Impact on H_2, D_2, CH_4 and CD_4', *Int. J. Mass Spectrom. Ion Processes*, **2**, 408–413 (1969).
134. F.P. Lossing, 'Threshold Ionization of Acetylene by Monoenergetic Electron Impact', *Int. J. Mass Spectrom. Ion Processes*, **5**, 190–192 (1970).
135. F.P. Lossing, G.P. Semeluk, 'Free Radicals by Mass Spectrometry. XLII. Ionization Potentials and Ionic Heats of Formation for C_1–C_4 Alkyl Radicals', *Can. J. Chem.*, **48**, 955–965 (1970).
136. F.P. Lossing, 'Free Radicals by Mass Spectrometry. XLIII. Ionization Potentials and Ionic Heats of Formation for Vinyl, Allyl, and Benzyl Radicals', *Can. J. Chem.*, **49**, 357–361 (1971).
137. I.H. Suzuki, K. Maeda, 'Appearance Potential of Methylene-ion from Ethylene', *Int. J. Mass Spectrom. Ion Processes*, **13**, 293–295 (1974).
138. K. Maeda, I.H. Suzuki, Y. Koyama, 'Ionization Efficiency Curves of Ethylene by Electron Impact', *Int. J. Mass Spectrom. Ion Processes*, **14**, 273–283 (1974).
139. I.H. Suzuki, K. Maeda, 'Ionization Efficiency Curves and Fragmentations of Deuterated Ethylenes by Electron Impact', *Int. J. Mass Spectrom. Ion Processes*, **15**, 281–290 (1974).
140. H.M. Rosenstock, M. Krauss, 'Quasi-equilibrium Theory of Mass Spectra', in *Mass Spectrometry of Organic Ions*, ed. F.W. McLafferty, Academic Press, New York, 1–64, 1963.
141. M. Vestal, 'Ionic Fragmentation Processes', in *Fundamental Processes in Radiation Chemistry*, ed. P. Ausloss, John Wiley & Sons, New York, 59–118, 1968.
142. H.M. Rosenstock, K. Draxl, B.W. Steiner, J.T. Herron, 'Energetics of Gaseous Ions', *J. Phys. Chem. Ref. Data*, **6**(Suppl. 1), (1977).
143. R.K. Mitchum, PhD Thesis, Oklahoma State University, 1973.
144. G.G. Meisels, J.Y. Park, B.G. Giessner, 'Ionization and Dissociation of C_4H_8 Isomers', *J. Am. Chem. Soc.*, **92**(2), 254–258 (1970).

145. H. Ehrhardt, O. Osberghaus, 'Temperaturabhangigkeit der Massenspektren von Kohlenwasserstoff molekulen und ihre Bedeutung in Rahnen der statistischer Theorie', *Z. Naturforsch*, **15a**, 575–584 (1960).
146. W. Brand, K. Levsen, 'Thermal Energy Distribution and Energy Deposition During Field Ionization', *Int. J. Mass Spectrom. Ion Processes*, **35**, 1–9 (1980).
147. W. Brand, K. Levsen, 'Approximate Determination of Absolute Unimolecular Rate Constants Using the Field Ionization Kinetic Method', *Int. J. Mass Spectrom. Ion Processes*, **51**, 135–154 (1983).
148. W.A. Chupka, 'Effect of Thermal Energy on Ionization Efficiency Curves of Fragment Ions', *J. Chem. Phys.*, **54**(5), 1936–1947 (1971).
149. J.W. Otvos, D.P. Stevenson, 'Cross-sections of Molecules for Ionization by Electrons', *J. Amer. Chem. Soc.*, **78**, 546–551 (1956).
150. F.W. Lampe, J.L. Franklin, F.H. Field, 'Cross-sections for Ionization by Electrons', *J. Amer. Chem. Soc.*, **79**, 6129–6132 (1957).
151. D. Rapp, P. Englander-Golden, 'Total Cross-sections for Ionization and Attachment in Gases by Electron Impact. I. Positive Ionization', *J. Chem. Phys.*, **43**(5), 1464–1479 (1965).
152. J.A. Beran, L. Kevan, 'Molecular Electron Ionization Cross-sections at 70 eV', *J. Phys. Chem.*, **73**(11), 3866–3876 (1969).
153. R. Alberti, M.M. Genoni, C. Pascual, J. Vogt, 'Electron Impact Ionization/Cross-sections of Organic Molecules. I. Apparatus and Ionization Cross-sections of *n*-Alkanes and *n*-Alkyl Derivatives', *Int. J. Mass Spectrom. Ion Processes*, **14**, 89–98 (1974).
154. W.L. Fitch, A.D. Sauter, 'Calculation of Relative Electron Impact Total Ionization Cross-sections for Organic Molecules', *Anal. Chem.*, **55**(6), 832–835 (1983).
155. A.D. Sauter, J.J. Downs, J.D. Buchner, N.T. Ringo, D.L. Shaw, J.G. Dulak, 'Model for the Estimation of Electron Impact Gas Chromatography/Mass Spectrometry Response Factors for Quadrupole Mass Spectrometers', *Anal. Chem.*, **58**(8), 1665–1670 (1986).
156. F. Turecek, L. Brabec, V. Hanus, V. Zima, O. Pytela, 'Relative Ionization Cross-sections of Oxygenated C(4) Molecules', *Int. J. Mass Spectrom. Ion Processes*, **97**, 117–124 (1990).
157. C. Allgood, R. Orlando, B. Munson, 'Correlations of Relative Sensitivities in Gas Chromatography Electron Ionization Mass Spectrometry with Molecular Parameters', *J. Am. Soc. Mass Spectrom.*, **1**, 397–404 (1990).
158. D.P. Stevenson, D.O. Schissler, *The Chemical and Biological Actions of Radiation*, ed. M. Haissinsky, Academic Books, London, Vol.5, 1961.
159. J.A. Beran, A.K. Dhingra, W.S. Liu, D.F. Mullica, 'Photoionization Cross-sections at 11.6–11.8 eV', *Can. J. Chem.*, **53**, 1616–1621 (1975).
160. F.T. Deverse, A.B. King, 'Effect of Molecular Structure on the Ionization Probabilities of Aromatic Molecules', *J. Chem. Phys.*, **41**(12), 3833–3838

Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Vincent P. Nero
Texas Upstream Technology, Houston, USA

1 Introduction 1
2 General Composition of Commercial Petroleum Products 2
3 General Methodology 2
3.1 Reducing Petroleum Matrix Complexity 2
3.2 Gas Chromatography Combined with Low-resolution Electron Ionization Mass Spectrometry 3
3.3 Direct Insertion Probe Mass Spectrometry/Mass Spectrometry 3
3.4 Specialized Internal Petroleum Surrogates 3
4 Applications 4
4.1 General Strategies for the Mass Spectrometry Analyses of Petroleum Components 4
4.2 General Strategy for Well-characterized Components at Moderate Concentrations: Analysis of the Individual Aromatic Components in Gasoline 4
4.3 General Strategy for Well-characterized Components at Low Concentrations: Analysis of the Trace Components in Refined Oils and Crude Oils by Gas Chromatography/Mass Spectrometry and Liquid Chromatography 8
4.4 General Strategy for the Analysis of Less Defined Components Present at Moderate Concentrations 9
4.5 General Strategy for Well-characterized but Thermally Labile Components at Moderate Concentrations: Direct Insertion Probe Mass Spectrometry/Mass Spectrometry for the Determination of Elemental Sulfur in Crude Oil 11
5 Conclusion 13
Abbreviations and Acronyms 13
Related Articles 14
References 14

Common industrial analytical procedures determine the overall physical or chemical characteristics of a petroleum product, which are necessary to market the product, but these methods provide little or no information concerning specific chemical components. This article focuses on standard methods which can be used to determine the composition of individual components or classes of components within various petroleum matrices. Gas chromatography (GC) coupled with low-resolution mass spectrometry (MS) is the most useful method for quickly identifying and precisely measuring specific components or classes of components in a petroleum matrix. Direct insertion probe mass spectrometry/mass spectrometry (DIP/MS/MS) can be used for less volatile and thermally labile petroleum components.

1 INTRODUCTION

Generally accepted industrial analytical procedures utilized for commercial petroleum products rarely determine any of the numerous actual chemical components in those products. Most industrial procedures only define the overall physical or chemical characteristics of a petroleum product that are necessary to market the product. Physical properties such as boiling range, density, viscosity, flash point, and vapor pressures are measured. Specific chemical characteristics such as carbon, hydrogen, and nitrogen content, total sulfur, average molecular weight, octane number, or cetane number are determined. This article will focus on the composition of individual components or specific classes of components in the common commercially available refined or natural petroleum products. Specific components or classes of components can be identified quickly and precisely quantitated using GC coupled with low-resolution MS, i.e. gas chromatography/mass spectrometry (GC/MS). DIP/MS/MS can sometimes be used for less volatile and thermally labile petroleum components. DIP/MS/MS is also a useful technique for determining the elemental sulfur content of crude oils. The selection of appropriate internal surrogates will greatly enhance identification, precision, and accuracy.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 GENERAL COMPOSITION OF COMMERCIAL PETROLEUM PRODUCTS

Petroleum products are extremely complex. Crude oils contain thousands of hydrocarbon components, many of which have not been fully characterized. Some components are unique to only a few specific crude sources. A second complication is the large and varying range of concentrations of possible components. Individual component concentrations will range from parts per billion levels to percentage levels. This range will far exceed the dynamic limits of most analytical instruments. A third complication is that many components are extremely similar. Although hundreds of these components are structurally unique, still thousands of other components are isomeric. Many of these isomeric petroleum hydrocarbons have very similar chemical and physical properties.

It will be useful to describe the major types of components in common commercial liquid petroleum products. These products are listed below in ascending boiling ranges.

- *Aviation gasoline* contains primarily alkylate, which is a highly branched saturated hydrocarbon mixture ranging from hexanes to nonanes, plus tetraethyllead. It does not contain aromatics or olefinic (unsaturated) hydrocarbons. It has a nominal boiling range of 30–220 °C.
- *Motor gasoline* contains straight run gasoline, which includes all of the natural crude oil components in the gasoline boiling range. Additionally, gasoline contains special refinery streams such as alkylate, reformate (which is mainly toluene), ethylbenzene, xylenes, and C_3-benzenes, some benzene and naphthalenes, plus trace amounts of heavy polycyclic aromatic hydrocarbons (PAHs), olefinic alkylate, and sometimes substantial levels of methyl *tert*-butyl ether (MTBE) or other oxygenates. The nominal boiling range of gasoline is 30–220 °C.
- *Kerosene* has a very significant normal paraffin content ranging from nonane to nonadecane, usually with a maximum near tridecane or tetradecane. Kerosene also contains lower levels of aromatics and some naphthenics (cyclic aliphatics) in the corresponding boiling range. Kerosene does not contain olefinic hydrocarbons. Its nominal boiling range of major components is 140–240 °C.
- *Aviation jet fuel* has a composition and boiling range similar to those of kerosene.
- *Diesel fuel* also has significant normal paraffin content, which ranges from decane to tetracosane, and usually has a maximum near hexadecane. Diesel also contains some aromatics and naphthenics with corresponding paraffinic boiling range. Diesel usually contains ethylhexyl nitrate as a cetane enhancer. Diesel does not contain olefinic hydrocarbons. Its nominal boiling range is 220–340 °C.
- *Home heating fuel* has a composition and boiling range similar to diesel, but does not contain ethylhexyl nitrate.
- *Motor oils* primarily contain octadecane to C_{32} paraffinic, naphthenic, and aromatic hydrocarbons along with various lubricant additives. Motor oils do not contain olefinic hydrocarbons. Modern fresh motor oils do not contain nonalkylated PAHs. Many motor oils contain low levels of distinctive biomarkers. Motor oils may contain or be completely composed of synthetic esters or polyolefins. These synthetic motor oils can be easily differentiated from natural motor oils. The boiling point of motor oils starts around 330 °C and will extend beyond 400 °C.

  Within the above refined petroleum product classifications, the general chemical composition is fairly similar. However, the relative distributions of these chemical components are unique and can generally be useful for distinguishing one fuel from another in the same class.
- *Crude oils*, in contrast to refined products, vary greatly from each other. Still, a few generalities can be made. Crude oils usually contain a distinctive normal paraffin content over a wide boiling range. Crude oils do not contain olefinic hydrocarbons. They have relatively fewer branched-chain alkanes and more cycloparaffins in the gasoline range than motor gasoline. Crude oil will contain low levels of distinctive biomarkers, which are particularly useful fingerprints for identifying a particular crude oil. They generally contain low levels of nonalkylated PAHs. Crude oils may also contain elemental sulfur, S_8.

3 GENERAL METHODOLOGY

3.1 Reducing Petroleum Matrix Complexity

There are two significant physical properties of petroleum products that are generally used to simplify a liquid petroleum matrix: polarity and boiling point. Petroleum products can generally contain two broad polarity classifications, aliphatic and aromatic hydrocarbons. Aliphatic (including saturated, unsaturated and cyclic), hydrocarbons are relatively less polar than aromatic hydrocarbons. Simple chromatographic procedures, such as the use of solid-phase cartridges, usually can be used to separate

aliphatics rapidly from aromatics. Automated high-performance liquid chromatography (HPLC), if available, can also be very useful. Typically, crude oils, gasolines, diesel fuels, and many motor oils contain 15–30% aromatics. Boiling fractions will again simplify the petroleum matrix. Gasoline, kerosene, diesel fuel, and lubricants are, to a first approximation, boiling range fractions of crude oil. GC using nonpolar columns is a quick procedure to separate hydrocarbons by boiling point.

3.2 Gas Chromatography Combined with Low-resolution Electron Ionization Mass Spectrometry

Only GC/MS can resolve and identify petroleum products at the molecular level. The gas chromatograph has the separating power to resolve the thousands of compounds in petroleum products. The mass spectrometer will further separate even unresolved chromatographic peaks based upon their mass spectral characteristics. The mass spectrometer has a sufficient dynamic range and sensitivity to measure both the major and trace components. Most important, the mass spectrometer has the specificity to identify or at least characterize most of these complex hydrocarbons.

3.2.1 Instrumental Conditions for Gas Chromatography/Mass Spectrometry

Most product samples can be separated at least partially by medium-resolution GC using splitless injection on to a nonpolar capillary column (30 m × 0.25 mm ID or 20 m × 0.18 mm ID columns, 100% methylsilicone or 95% methyl–5% phenylsilicone phase). Complex samples require high-resolution GC conditions, using 100 m × 0.25 mm ID or 40 m × 0.10 mm ID capillary columns. The narrow-bore columns are preferable owing to the fast analysis time and low column bleed into the mass spectrometer detector, although some commercial gas chromatographs cannot handle the high head pressure required. The mobile phase velocity is set to 30 cm s^{-1} for helium carrier gas. A slow oven ramp is used to provide maximum chromatographic separation, up to the upper temperature limit of the column (usually 350–400 °C).

The mass spectrometer detector is configured with a direct capillary interface. It is operated in an electron ionization (EI) mode using an ionization energy between 50 and 100 eV. The mass range is set in full scan mode to 29–500 Da at unit resolution. The scan time is set to achieve at least eight scans per GC peak width at half-height. Quantitation is achieved using the most unique mass for each target component, usually the molecular ion, which is measured against the closest eluting internal standard of similar polarity. Isotopic internal surrogates will greatly enhance identification, precision, and accuracy.

3.3 Direct Insertion Probe Mass Spectrometry/Mass Spectrometry

When a petroleum component is not volatile and/or thermally labile, GC will be ineffective. However, DIP/MS/MS can often be an effective substitute for selected compounds. The first of the tandem mass spectrometers replaces the gas chromatograph. It is used to select a characteristic set of masses for a targeted compound, usually its molecular mass. The second mass analyzer is used to discriminate between those compounds with similar masses. An internal surrogate is still required for precision and accuracy. This technique is rapid and requires minimum sample preparation. The high selectivity can be tailored to the analyte even in the most complex of matrices, such as crude oils. DIP/MS/MS can be used to measure elemental sulfur, S_8, in crude oils. This is an excellent example of achieving selectively, high sensitivity, and accuracy with minimal analysis time.

3.3.1 Instrumental Conditions for Direct Insertion Probe Mass Spectrometry/Mass Spectrometry

The surrogate of the compound of interest is spiked at a known concentration into the petroleum sample. The spiked sample is then placed on a direct thermal probe and is slowly sublimed at low temperature directly into the source of the mass spectrometer. Although this slow evaporation or sublimation does provide some separation, the traditional total ion mass scan would still display only a continuous spectrum with hits at virtually every mass. However, when the tandem mass spectrometer is sequentially programmed to scan selective daughter ions that are key masses for the target compound and alternately for the surrogate compound, then quantitation can be achieved.

The mass spectrometry/mass spectrometry (MS/MS) conditions should be optimized for the target analyte. The selected ionization energy should maximize the target component's parent ion. Generally, the molecular ion is selected, but other distinctive ions may be used as the parent ion. The collision offset energy that provides the daughter ion fragments should also be optimized.

3.4 Specialized Internal Petroleum Surrogates

The surrogate compounds should have the same physical and chemical properties as the analyte and should not be naturally present in the petroleum matrix. Isotopically labeled aliphatics and aromatics make ideal surrogates. Isotopic surrogates over the entire petroleum hydrocarbon boiling range are generally available commercially. These surrogates are simple to use and easy to calibrate against target analytes. They both make identifications easier and greatly enhance accuracy of measurements.

Deuterated surrogates should be prepared and added just prior to analysis due to the possibility of isotopic exchange with solvents at high or low pH levels. Carbon-13 surrogates are stable, but are much more expensive. Examples of the various commercially available hydrocarbon surrogates and the corresponding petroleum product are as follows:

- Gasoline-range surrogates – percentage range concentrations of octane-d_{18}, benzene-d_6, toluene-d_8, ethylbenzene-d_{10}, styrene-d_8, butylbenzene-d_{12}, 4-ethyltoluene-d_3, and naphthalene-d_8; parts per million range deuterated PAHs.
- Kerosene- and diesel-range surrogates – parts per thousand range concentrations of hexadecane-d_{34}, methylnaphthalene-d_{10}, butylbenzene-d_{14}; parts per million range deuterated PAHs.
- Motor oil-range surrogates – parts per thousand range concentrations of octacosane-d_{58}, nonylpyrene-d_{28}, butylbenzene-d_{14}; parts per million range deuterated PAHs.
- Heavy crude oil-range hydrocarbons – same as motor oil. Use $^{34}S_8$ to determine elemental sulfur.
- Isotopic derivatives of common fuel additives such as deuterated MTBE, methyl-d_{12} *tert*-butyl ether.

4 APPLICATIONS

4.1 General Strategies for the Mass Spectrometry Analyses of Petroleum Components

Application examples of four general strategies for analyzing components in liquid petroleum matrices will be described. Very similar approaches can be applied for the analysis of just about any individual or group of semivolatile components found in any petroleum product. These general mass spectrometric strategies can be summarized as follows:

- The analysis of known, well-characterized and stable components that are present at moderate concentration levels (approximately 50 ppm or greater). The GC/MS analysis of the specific aromatics and oxygenates in gasoline will be used to illustrate this approach.
- The analysis of known, well-characterized and stable components that are present at low concentration levels (approximately 10 ppm or less). The HPLC cleanup and subsequent GC/MS analysis of the 2-aminonaphthalene in motor oil will be used to illustrate this approach.
- The analysis of less defined components that are present at moderate concentration levels (about 50 ppm or greater). Two approaches will be used as examples. First is the GC/MS analysis of specific alkylated PAHs in crude oil. The comparative analysis of trimethylphenanthrene and trimethyldibenzothiophene will be used to illustrate this approach. Second is the DIP/MS/MS analysis of the specific higher molecular weight alkylated PAHs in crude oil. The analysis of alkylbenzothiophenes and alkyldibenzothiophene in heavy oil fractions will be used to illustrate this approach.
- The analysis of known, well-characterized but thermally labile components that are present at moderate concentration levels (about 100 ppm or greater). The DIP/MS/MS analysis of elemental sulfur in crude oil will be used to illustrate this approach.

4.2 General Strategy for Well-characterized Components at Moderate Concentrations: Analysis of the Individual Aromatic Components in Gasoline

The GC/MS analysis of the specific aromatics and oxygenates in gasoline will be used as an example of the general strategy for the compositional analysis of known components at moderate concentrations. The gasoline matrix is the simplest of the liquid petroleum products, yet it still presents a challenge to analysis. Various GC methods have been used for many years to identify and quantitate gasoline components, but these methods have significant problems. Even when very high-resolution GC methods are used, co-eluting peaks and multiple isomers with different retention times complicate the resulting chromatograms. A stand-alone MS method, American Society for Testing and Materials (ASTM) Method D2789, which does not use a prior chromatographic separation, also has been used to determine the aromatic concentration in gasoline.[1] However, ASTM Method D2789 only provides general gasoline component classifications (i.e. paraffins, monocycloparaffins, dicycloparaffins, alkylbenzenes, indans plus tetralins, and naphthalenes). Furthermore, ASTM Method D2789 lacks precision because it assumes the probable distribution of aromatics in gasoline and requires that the gasoline be relatively free of olefins.

A combination of GC and MS overcomes the lack of analytical target specificity that is characteristic of the individual techniques when those techniques are used independently. GC/MS is an orthogonal hyphenated method that combines multiple complementary separation capacities of both methods. This combination results in an extremely powerful technique for the analysis of gasoline. The general strategy for applying GC/MS to the component analysis of aromatics in gasoline will now be described. A very similar approach can be applied

for the analysis of just about any individual or group of semivolatile components found in any petroleum product.

4.2.1 *Gas Chromatography/Mass Spectrometry of Aromatics in Gasoline Using American Society for Testing and Materials Method D5769*

The United States Environmental Protection Agency (USEPA), specified in the *Federal Register* in 1994 that GC/MS must be used to determine the total aromatics in gasoline.[2] The USEPA established performance guidelines, but only very general procedural guidelines. It permits the use of any GC/MS conditions and internal standards as long as the results are within a relative error of 2.0%. This progressive legislation allowed the petroleum and environment laboratories to develop their own particular methodologies. The Texaco Upstream Technology laboratory worked closely in a joint effort with other laboratories to develop specific details to optimize the performance and minimize the complexity of this new USEPA method. The result of this effort was ASTM Method D5769.[3]

The GC/MS conditions for ASTM Method D5769 are summarized at follows. The gas chromatograph is equipped with a 100% polymethylsiloxane open-tubular GC capillary column. The method suggests using a 60 m × 0.32 mm ID column which is operated at 50 °C (1 min), ramped at 2 °C min^{-1} to 190 °C. The total acquisition time will be 71 min. The mass spectrometer is operated under 70 eV EI conditions. The method required a minimum full-scan sampling rate of five scans across peak widths at half-height.

The ASTM Method D5769 internal standards are perdeuterobenzene, perdeuteroethylbenzene and perdeuteronaphthalene. Perdeuterotoluene is optional. (The hydrogen atoms in a perdeutero compound have been totally substituted for deuterium atoms.) The resulting compound is an internal surrogate with almost the same GC retention time as the corresponding aromatic component. The deuterated surrogates will actually have a slightly lower GC retention time. Of course, the mass spectrum of the deuterated surrogate will be completely different, since deuterium has an atomic mass of 2 Da. However, the deuterated surrogate will have an analogous mass spectrum corresponding to the natural component spectrum. The relative GC/MS response factors for individual pure components (from benzene through all of the major aromatics up to C_{11}) are measured relative to the response of the deuterated standards. A typical calibration, in this case for benzene, is shown in Figure 1. The coordinates are relative response vs relative amount. The slope of the line is the response factor. In our laboratory each of five concentrations across the calibration range was measured at least three times prior to calculating response factors. The fit is essentially linear with a coefficient of correlation of 0.9999. All major aromatic compounds are calibrated separately, generally using molecular ions in the mass spectrum for quantitation for most species.

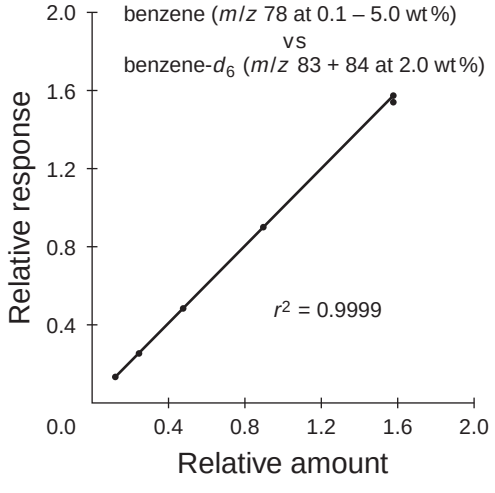

Figure 1 Calibration curve for benzene. The slope is a straight line, which is indicative of a linear response factor.

The deuterated internal standard mix is weighed into a 10-g sample of gasoline to the nearest 100 µg. A 12-component quality control (QC) mix is prepared by weight. The QC sample is then run. ASTM Method D5769 specifies that each individual aromatic species in the QC mix must be measured within 5% relative of the true, weighed concentration. Any additional unspecified experimental conditions used are acceptable, as long as this specification is obtained. The weight percentage of each target aromatic component is then calculated. (The volume percentage of that component is calculated from the density of the pure component relative to the sample density.) A synthetic gasoline consisting of known weighed amounts of pure gasoline components is then run as a QC sample. The measured total aromatics for this QC standard must be within 2.0 wt% of the weighed total aromatics. The ASTM Method D5769 statistical specifications are significantly tighter than those specified by the USEPA, since the USEPA requires that only the total aromatics of the QC mix be within 2% accuracy.

4.2.2 *Gas Chromatography/Mass Spectrometry of Aromatics in Gasoline Using a Very Fast and More Precise Version of American Society for Testing and Materials Method D5769*

This laboratory developed a very rapid and more precise modification for ASTM Method D5769.[4] The analysis

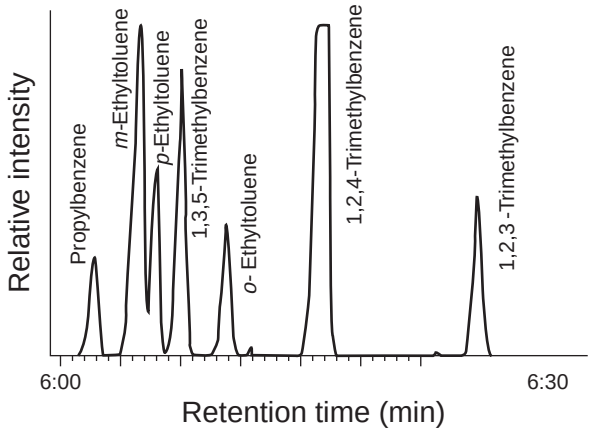

Figure 2 A selected ion chromatogram of *m/z* 120. The narrow-bore column elutes all of the C_3-benzenes within 30 s.

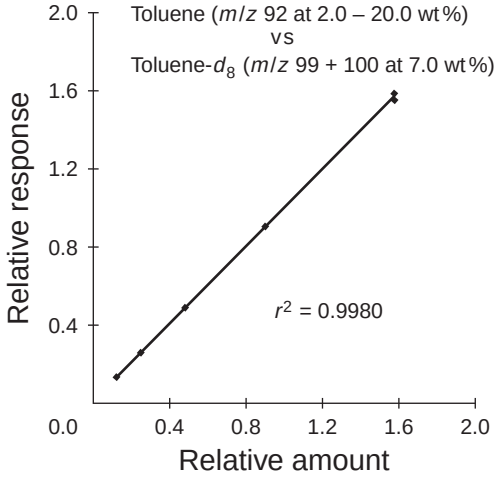

Figure 3 Calibration curve for toluene. The slope is linear, when calibrated against deuterated toluene, using *m/z* 99 and 100.

time was greatly reduced by applying fast GC column technology. Our laboratory specifies using either a 20 m × 0.10 mm ID or 20 m × 0.18 mm ID 100% polymethylsiloxane open-tubular column with a 25 °C min^{-1} ramp rate. These GC conditions reduce the total acquisition time to only 9 min vs 71 min and the chromatography is still well within the chromatographic resolutions specified in ASTM Method D5769. These narrow-bore columns are available from numerous manufacturers. Figure 2 shows a selected ion chromatogram in the range of C_9 alkylbenzenes (trimethylbenzenes, ethylmethylbenzenes, and propylbenzenes). The C_9 benzenes eluted totally within 6 min and with a GC resolution which meets or exceeds that published in ASTM Method D5769 for a 90-min version using the more traditional wider bore and longer columns. The peak widths are approximately 0.5 s. For such a peak width, ASTM Method D5769 requires a full-scan MS sampling rate of 10 scans s^{-1} (five scans across the peak at half-height). The MS sampling rate was actually 13 scans s^{-1} in this case. Most commercial mass spectrometers can easily scan at these rates.

4.2.3 Other American Society for Testing and Materials Method D5769 Modifications

The modified method improves the precision of ASTM Method D5769 while staying within its method parameters. The additional precision was achieved through the following three modifications. First, the addition of perdeuterotoluene to the internal standard mix increases the precision of toluene determination significantly. Perdeuterotoluene is optional for the standard version and has the least precision of any component in the standard ASTM Method D5769. The modified internal standard mix is added at approximately 10% by weight to each sample to give approximately 1 wt% benzene-d_6, 7 wt% toluene-d_8, 1 wt% ethylbenzene-d_{10}, and 1 wt% naphthalene-d_8. The result significantly increases the precision, since toluene generally has one of the highest concentrations of any individual component in gasoline. When toluene is quantitated against perdeuterobenzene, generally a quadratic calibration fit is required. However, when perdeuterotoluene is added at 7% in the sample, the linear range of the calibration curve is extended by almost one order of magnitude. Figure 3 shows linear calibration curves for toluene with an excellent correlation coefficient.

Second, the effective instrumental operational dynamic range of the mass spectrometer is significantly enhanced using microbore GC columns. The sensitivity specification in ASTM Method D5769 requires a detection limit of diethylbenzene at 20 ppm and a calibration curve for toluene that extends to 20%. Significant care must be taken in tuning the mass spectrometer to achieve such a wide dynamic range, but the selection of the proper GC conditions can extend the dynamic range of the GC/MS instrument far beyond the range of the MS detector alone. By using narrow-bore GC columns (0.1–0.18 mm ID) with a fast GC temperature ramp, it is possible to sharpen the small peaks while broadening the large peaks (column saturation), thus limiting the amount of material passing into the ion source at any given time. Although major components, such as toluene, may be saturated in the GC column, the MS detector is not saturated. The precision of the analysis is not affected and even the chromatographic resolution is still acceptable.

Third, the internal standard calculation requirements of ASTM D5769-95 can be another source of error. The method suggests using molecular ions for calibration,

Table 1 Six repetitive results for a gasoline analyzed using the modified version of ASTM Method D5769. Weight/weight percentages

Component	1	2	3	4	5	6	Mean	SD
Benzene	1.91	1.89	1.86	1.92	2.01	1.96	1.93	0.05
Toluene	7.51	7.47	7.30	7.33	7.43	7.50	7.42	0.09
Ethylbenzene	1.53	1.64	1.52	1.55	1.52	1.59	1.56	0.05
m,*p*-Xylene	6.43	6.58	6.56	6.52	6.52	6.44	6.51	0.06
o-Xylene	2.12	2.17	2.16	2.13	2.16	2.16	2.15	0.02
Isopropyl benzene	0.07	0.07	0.08	0.07	0.07	0.07	0.07	0.00
Propylbenzene	0.43	0.43	0.44	0.43	0.45	0.44	0.44	0.01
m-Ethyltoluene	1.85	1.95	1.95	1.92	1.95	1.90	1.92	0.04
p-Ethyltoluene	0.83	0.89	0.89	0.82	0.86	0.83	0.85	0.03
1,3,5-Trimethylbenzene	0.96	0.98	0.99	0.98	0.98	0.97	0.98	0.01
o-Ethyltoluene	0.53	0.51	0.54	0.53	0.52	0.53	0.53	0.01
1,2,4-Trimethylbenzene	2.53	2.57	2.55	2.50	2.56	2.56	2.55	0.03
1,2,3-Trimethylbenzene	0.48	0.50	0.50	0.50	0.50	0.50	0.50	0.01
Indan	1.16	1.15	1.16	1.16	1.15	1.13	1.15	0.01
C_4 benzenes, branched	0.09	0.11	0.09	0.09	0.10	0.08	0.09	0.01
C_4 benzenes, long chain	3.25	3.12	3.23	3.10	3.17	3.19	3.18	0.06
C_4 benzenes, tetramethyl	0.71	0.68	0.69	0.67	0.73	0.68	0.69	0.02
C_6 benzenes	0.18	0.19	0.17	0.20	0.21	0.20	0.19	0.01
Naphthalene	0.28	0.27	0.29	0.27	0.27	0.29	0.28	0.01
C_5 benzenes	0.39	0.37	0.36	0.36	0.37	0.39	0.37	0.01
2-Methylnaphthalene	0.27	0.27	0.28	0.27	0.27	0.27	0.27	0.00
1-Methylnaphthalene	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.00
Total aromatics	33.63	33.93	33.73	33.44	33.92	33.80	33.74	0.19

but many deuterated standards are only commercially available at less than 100% nominal isotopic purity. For example, perdeuteroethylbenzene is available in nominal 98% isotopic purity. Therefore, a problem may occur when this imperfect internal standard is use. The molecular ion (unfragmented ion) will be distributed between *m*/*z* 116 (C_8D_{10}) and *m*/*z* 115 ($C_8D_9H_1$). Since *m*/*z* 116 cannot lose hydrogen, *m*/*z* 115 cannot be a fragment of *m*/*z* 116. A batch of perdeuteroethylbenzene internal standard deuterated to 98.1% has an *m*/*z* 115/116 ratio of 20%, whereas another batch deuterated to 99% has an *m*/*z* 115/116 ratio of 10%. Using just the *m*/*z* 116 ion to quantitate this internal standard, as the method suggests, could cause significant error since the internal standard added to samples could be from a different batch than that used for calibration. In our laboratory, the sum of the intensities observed for both the molecular ions of the zero hydrogen and one hydrogen analogs of each internal standard were used to maximize precision and accuracy. For example, *m*/*z* 99 and 100 were used for toluene-d_8, as seen in Figure 3.

Table 1 shows for a typical sample the precision that can be obtained when incorporating these three easy to implement modifications. Total aromatics of the six repetitive analyses range within 0.48% of each other and within a comfortable 0.30% of the mean. The EPA requires that the calibration mixture, which is less complex to analyze than actual gasoline, to agree within 2.0% of the actual measured preparation value.

4.2.4 American Society for Testing and Materials Method D5769 Bias

ASTM Method D5769 contains some bias. The USEPA mandate does not include the higher aromatics such as C_{10+} alkylnaphthalenes, C_{12+} alkylbenzenes, styrenes, and alkylindenes, which are found in gasoline, but at relatively low concentrations. ASTM Method D5769 tries to correct this problem by selecting pentamethylbenzene to establish the response factors for the uncalibrated heavier alkylbenzenes. Unfortunately, out of the hundreds of C_{11} and C_{12} alkylbenzenes, only pentamethylbenzene and hexamethylbenzene do not contain benzylic carbon–carbon bonds. Alkylbenzenes with a weak benzylic C–C bond have molecular ions that are approximately 30% smaller than those for alkylbenzenes without a benzylic C–C bond. Therefore, estimating these uncalibrated components against the pentamethylbenzene calibration curve will produce results which are 30% lower than the actual value. In our laboratory, uncalibrated components are quantitated using response factors calculated from the aggregate response of the calibrated compounds butylbenzene, 1,4-diethylbenzene, and 1,2-diethylbenzene. These components are reported as 'long chain C_4-benzenes'. When a long-chain alkylbenzene is

used to estimate for uncalibrated compound response, the bias will disappear.

4.2.5 Expanding the Scope of American Society for Testing and Materials Method D5769

Again the method for total aromatics in gasoline can serve as a template for the analysis of any component in gasoline. Additional gasoline components can be added to the list of eluting targets with essentially no additional analysis time or loss in quantitative quality. For example, this laboratory easily modified ASTM Method D5769 to include MTBE and TAME. MTBE and TAME are common oxygenate additives that can be found in gasolines, especially during winter months. The fragment ion at *m*/*z* 73 was used to measure the two ethers, since the molecular ions of these nonaromatic species are weak or absent in their EI spectra. ASTM Method D5599 is a GC method with oxygen flame ionization detection (OFID) that is recommended for measuring gasoline oxygenates. Both methods produce essentially the same answer, but the modified ASTM Method D5769 also simultaneously provides gasoline aromatic content. Figure 4 is a plot of the results for both methods at low and high concentration levels of MTBE and TAME. The response curve is linear throughout the range.

4.3 General Strategy for Well-characterized Components at Low Concentrations: Analysis of the Trace Components in Refined Oils and Crude Oils by Gas Chromatography/Mass Spectrometry and Liquid Chromatography

The analysis of trace-level components in motor oil or heavy crude oils is one of the most challenging petroleum analyses. Heavy hydrocarbon mixtures are more difficult to analyze because the individual components have higher molecular weights. Large molecules contain more possible combinations of atoms. There is a continuous array of molecules with similar boiling points and molecular weights throughout even narrow boiling distributions. When GC/MS alone is applied for the analysis of trace components in a heavy petroleum matrix, the analyte ions can be lost in a continuum of interfering background noise. In these cases, solid-phase chromatography or HPLC can often be used to reduce the complexity of the matrix. The analysis for 2-aminonaphthalene in motor oil using HPLC cleanup prior to GC/MS will be used as example.

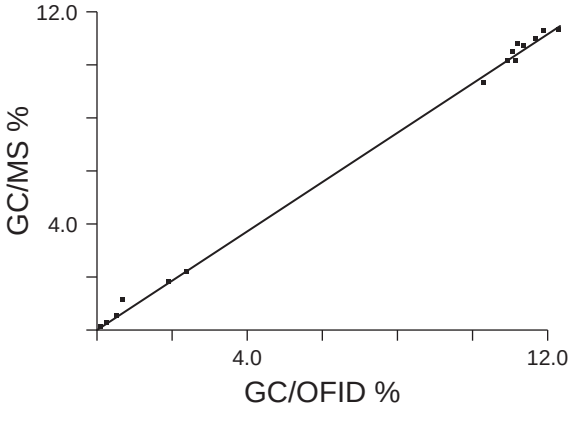

Figure 4 ASTM Method D5769, when modified to include oxygenates, produces the same results as ASTM Method D5599. MTBE and *tert*-amyl methyl ether (TAME) (*m*/*z* 73) calibrated against benzene-d_6 (*m*/*z* 84).

4.3.1 Analysis for 2-Aminonaphthalene in Motor Oil Using High-performance Liquid Chromatography Clean-up Prior to Gas Chromatography/Mass Spectrometry

2-Aminonaphthalene has long been recognized as a human carcinogen. It is crucial to establish that it is not present even at trace levels in petroleum products which may come into contact with potable water or food packaging equipment. The motor oil is spiked at the low parts per million levels with a deuterated surrogate of 2-aminonaphthalene and analyzed by a general strategy similar to the general methodology used for determining the aromatic content of gasoline. The gas chromatograph was operated with column

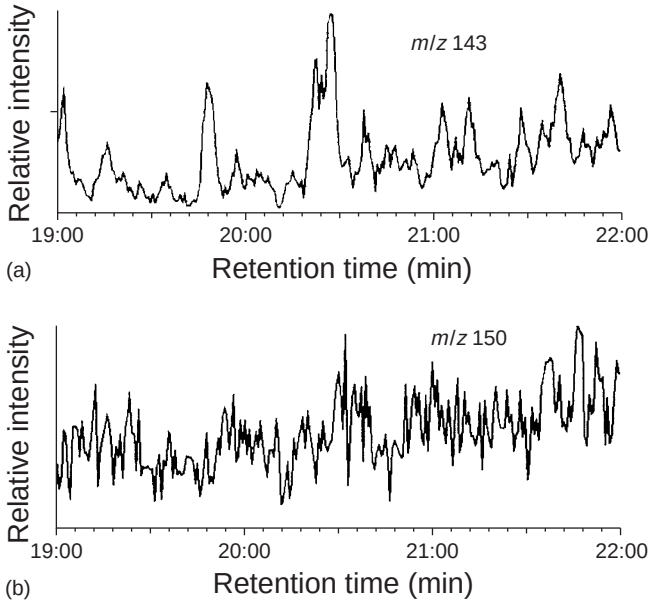

Figure 5 Selected ion traces at (a) *m*/*z* 143 and (b) *m*/*z* 150 for an oil sample spiked at 5 ppm with 1-aminonaphthalene and 2-aminonaphthalene-d_7 prior to HPLC cleanup. The chromatogram fails to indicate their presence above the background noise.

temperature programming from 70 °C at 8 °C min^{-1} to 320 °C. A 1-μL volume of sample was injected on to a 30 m × 0.25 mm ID, 0.25-μm phase thickness, fused-silica capillary 5% phenyl–methylsilica column with a flow rate of 40 cm s^{-1}. The mass spectrometer was operated at nominal mass using EI with an ionization energy of 70 eV. The instrument was scanned over a mass range from 50 to 250 Da every 0.75 s. The analysis failed. There are too many interfering components in the same retention time window and with the same fragment ions. Figure 5(a) and (b) shows GC/MS selected ion chromatograms of motor oil spiked at the 5 ppm level with both 1-aminonaphthalene and 2-aminonaphthalene-d_7. Both compounds will elute between 20 and 21 min under these GC conditions. The molecular ion trace for aminonaphthalenes, m/z 143, is questionable. The molecular ion trace for the deuterated surrogate, m/z 150, shows only noise. Even though the results are ambiguous, they are superior to others. One State and two commercial environmental laboratories actually falsely reported 2-aminonaphthalene present in this oil.

This laboratory developed an HPLC procedure that will quickly and unambiguously determine if 2-aminonaphthalene is not present in refined oils or crude oils at trace levels.[5] First, 25 μL of the original spiked oil was injected on to a 250 × 4.6 mm ID cyano column and eluted with a 5% solution of methylene chloride in hexane, then the fraction containing the aminonaphthalenes was collected and evaporated to about 25 μL.

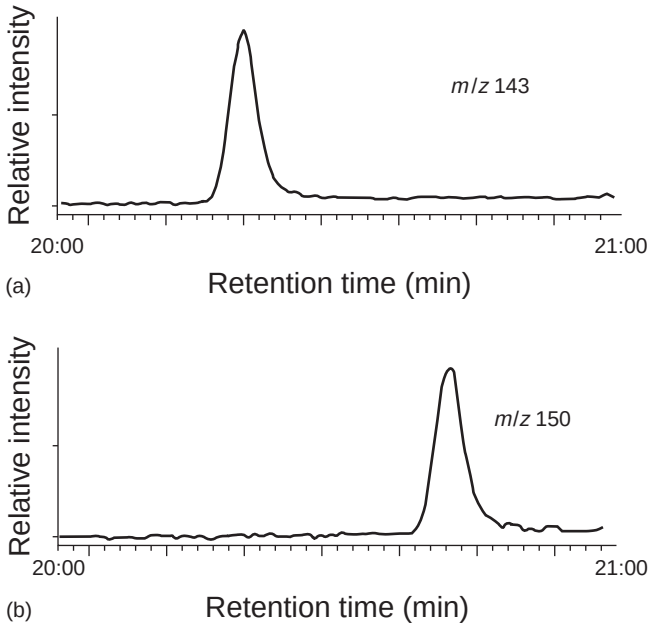

Figure 6 Selected ion traces at (a) m/z 143 and (b) m/z 150 for an oil sample spiked at 5 ppm with 1-aminonaphthalene and 2-aminonaphthalene-d_7 after HPLC cleanup. The chromatogram clearly indicates their presence of both compounds.

Finally, the GC/MS procedure was repeated under the previous conditions. Figure 6(a) and (b) shows selected ion chromatograms of the cleaned-up oil. The deuterated 2-aminonaphthalene surrogate and 1-aminonaphthalene are clearly present at the 5 ppm level. There is no evidence of the natural 2-aminonaphthalene, since that compound would have eluted just after its deuterated surrogate when using these GC conditions. The retention time for the natural (non-deuterated) 2-aminonaphthalene was verified in subsequent analyses.

This procedure has a lower limit of detection of approximately 500 ppb. Still lower detection limits would have been achieved with selected ion detection. However, the mass spectrometer was operated using a full mass range in this case to assure that the HPLC cleanup procedure completely eliminated hydrocarbon interferences from the oil. This procedure was then supplied to the commercial environmental laboratories that initially used straight GC/MS without HPLC cleanup and had initially reported 2-aminonaphthalene to be present in this oil at the low parts per million level. The commercial laboratory then also verified that 2-aminonaphthalene was absent.

Special precautions must be taken when using a deuterated surrogate, since deuterium can exchange with some solvents and matrices, especially under acidic or basic conditions. In one case, one of the labeled aromatic ring positions adjacent to the amino group exchanged with the solvent, methylene chloride, during prolonged storage. A 6-week-old methylene chloride solution showed over 80% 2-aminonaphthalene-d_6 and only 3% of the original 2-aminonaphthalene-d_7. Standards should be prepared shortly before use and stored, if necessary, cold.

4.4 General Strategy for the Analysis of Less Defined Components Present at Moderate Concentrations

The most challenging target analyte is a target that is poorly characterized. Many petroleum components have not yet been isolated and therefore their actual physical properties are unknown. Furthermore, many components may never be isolated because of the complexity of the petroleum matrix and the expense of thorough analyses. In these cases, it is more practical to analyze classes of hydrocarbons, which can be loosely defined by their basic chemical properties and carbon framework. Members of individual class types will consist of isomers and homologs, varying from each other only by position and type of aliphatic groups attached to the particular carbon framework. These carbon frameworks can be generalized as various combinations of aliphatics, olefinics, naphthenics, aromatics, and naphthoaromatics. Traditional methods were developed in the 1950s using batch-inlet electron ionization mass spectrometry (EIMS). These hydrocarbon type analyses are still useful in some cases for

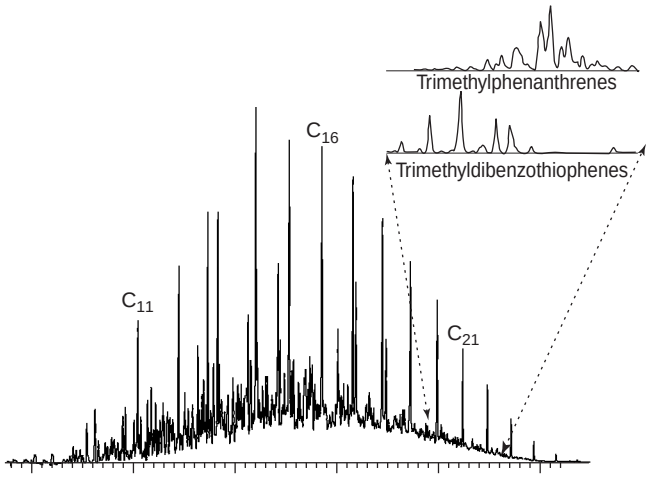

Figure 7 Total ion chromatogram of a diesel fuel with selected ion traces at *m*/*z* 220 and 226, eluting in the 20–22 min time interval.

characterizing middle distillates and vacuum gas oils.[6–8] However, GC/MS and DIP/MS/MS now permit these traditional analyses to be customized more selectively within individual hydrocarbon types. The information is especially useful for predicting general physical and chemical properties of petroleum products.

4.4.1 Gas Chromatography/Mass Spectrometry Analysis of Alkylated Polycyclic Aromatic Hydrocarbons in Petroleum Products

GC/MS response factors for alkylated polyaromatic hydrocarbons are either difficult to generate or unavailable. However, the GC data are internally consistent over short boiling ranges, when comparing within one or within two similar homologous series. For instance, the relative amounts of two different alkylated PAH homologs, which are both alkylated to the same extent, are commonly used to differentiate middle distillate products or crude oils from each other.[9] For example, Figure 7 shows the total ion chromatogram of a diesel fuel and two selective ion traces extracted from it. The selective traces show only trimethylphenanthrenes (*m*/*z* 220) and trimethyldibenzothiophenes (*m*/*z* 226), which eluted in the 20–22 min time interval. The total peak areas of trimethyldibenzothiophenes are then divided by the corresponding sum of the trimethylphenanthrenes. This ratio is called the D3/P3 index and is characteristic of the particular sample. For example, the D3/P3 index of a low-sulfur diesel fuel will range typically from 0.10 to 0.30. An Arabian Light Crude Oil may have an index as high as 4.06. The general pattern within a particular alkyl-PAH sequence is also

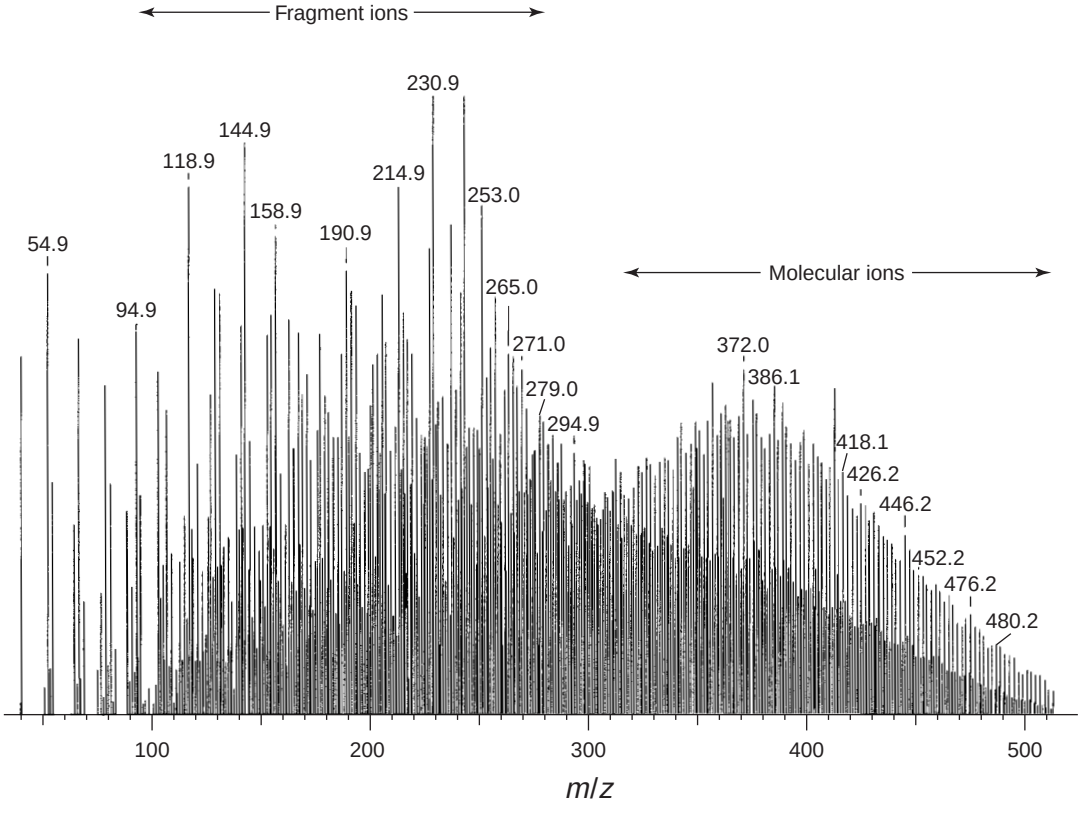

Figure 8 Total-ion, low-voltage EI mass spectrum of a heavy petroleum fraction.

valuable for determining the source or fingerprints for a petroleum product.

4.4.2 Direct Insertion Probe Mass Spectrometry/Mass Spectrometry Analysis of the Specific Alkylated Polycyclic Aromatic Hydrocarbons in Crude Oil

The heavier components of crude cannot be analyzed by GC/MS since they cannot be effectively resolved by GC or are simply not volatile. However, DIP/MS/MS is useful for analyzing specific higher molecular weight alkylated PAHs in crude oil. The general procedure follows: the high-boiling crude oil cut is placed in a quartz direct insertion probe tube and inserted into the source of a mass spectrometer. The sample is heated from 50 to 450 °C at 50 °C min^{-1} and held for 5 min. The mass spectrometer is operated under EI conditions using an ionization energy of 20 eV. The resulting total ion mass, presented in Figure 8, is of low value. It does clearly show a bimodal ion distribution. The lower mass distribution represents fragmentation ions and the higher mass distribution represents a continuum of molecular ions. However, this molecular ion distribution can be nearly resolved using parent mode MS/MS. The parent ions of *m*/*z* 197, 213, 215, 217, 221, 219, 223, and 225 were sequentially scanned. Figure 9 reveals a distribution of dibenzothiophenes that are generated by the parent ions of *m*/*z* 197. Figure 10 reveals a distribution of benzothiophenes that are generated by the parent ions of *m*/*z* 217.

4.5 General Strategy for Well-characterized but Thermally Labile Components at Moderate Concentrations: Direct Insertion Probe Mass Spectrometry/Mass Spectrometry for the Determination of Elemental Sulfur in Crude Oil

Elemental sulfur in crude oil can range from trace amounts to low percentage levels in mixtures, which are sometimes nonhomogeneous. The traditional analyses for measuring sulfur in crude oil have been limited to wet chemical methods. These techniques are time-consuming, are cumbersome, require skilled technicians, and still generally lack accuracy and precision. These analyses are further complicated because of the physical and chemical properties of elemental sulfur (sulfur has a low sublimation temperature, low and differing solubilities in the various crude oils, moderate reactivity with air, and even a modest thermal reactivity with hydrocarbons in the crude oil matrix).

DIP/MS/MS, when used in conjunction with an isotopicaly labeled sulfur-34 surrogate, provides a rapid and accurate method for determination of natural sulfur in crude oils.[10] Elemental sulfur-34 is spiked directly into the crude oil at 0.1 wt% and allowed to dissolve thoroughly overnight. The homogeneous spiked sample is then placed on a direct insertion probe and introduced into the mass spectrometer while being held at 30 °C. The sulfur is non-reactive at this low temperature, but will still sublime easily in the low-pressure environment of the mass spectrometer's ion source. Since the probe was never heated, most of the crude remains on the probe.

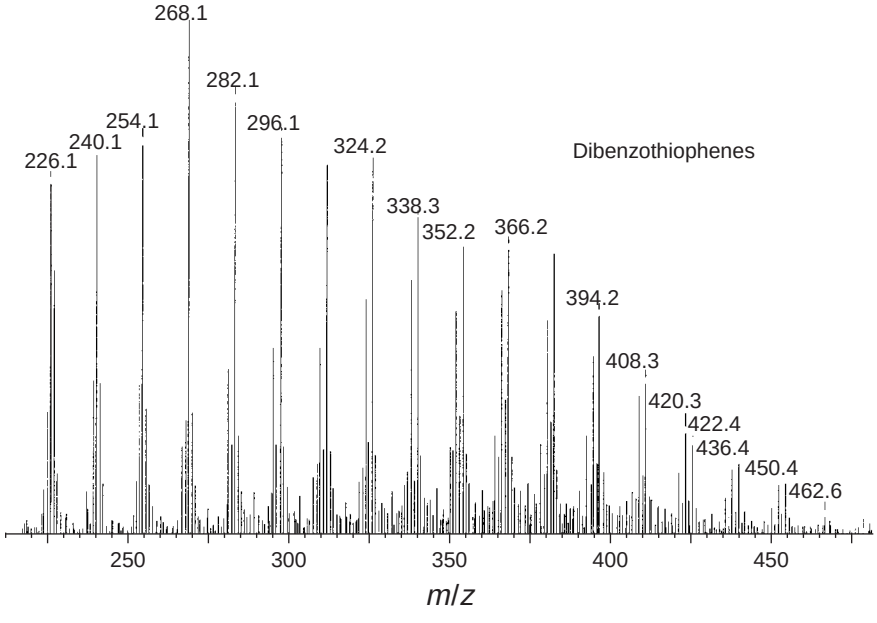

Figure 9 Mass spectrum showing parent ions of *m*/*z* 197.

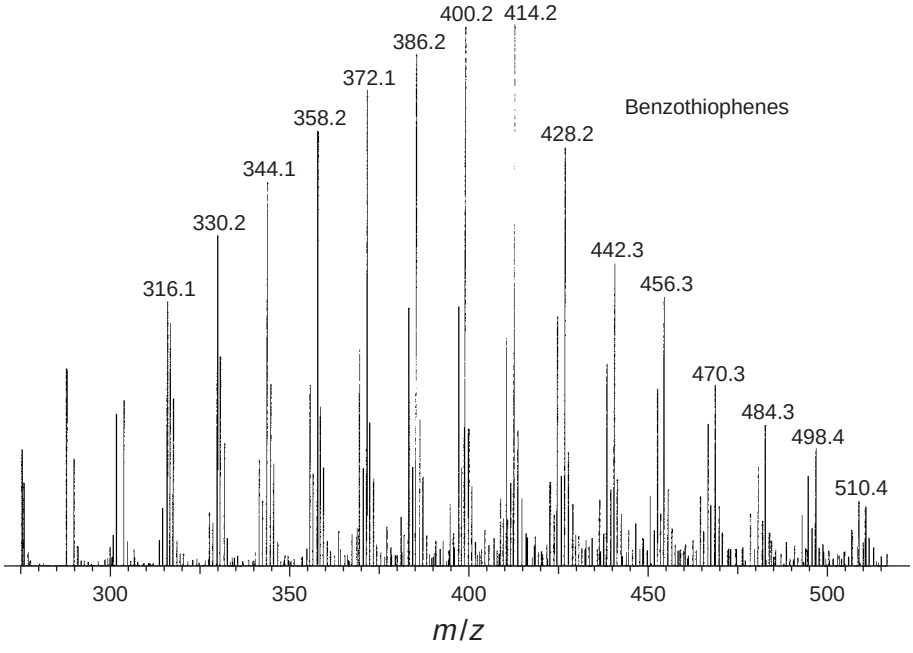

Figure 10 Mass spectrum showing parent ions of *m*/*z* 217.

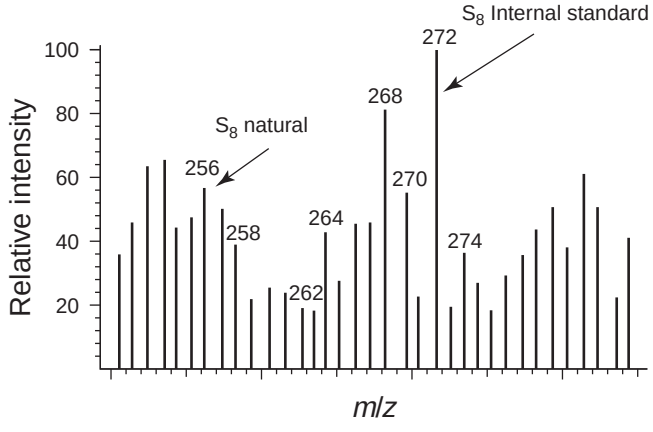

Figure 11 Normal total-ion EI mass spectrum of a direct insertion probe crude oil sample. Sulfur peaks are not prominent.

The mass spectrometer was sequentially programmed for one full-range scan and four daughter mode scans under EI conditions at 70 eV ionization energy. The full-range scan was from 250 to 285 Da. The daughter ions of *m*/*z* 256, 258, 270, and 272 were collected starting at mass 75 to the mass of the daughter ion minus 2 Da. The collision offset was set at 10 eV. Scan time was 1 s for each scan. Figure 11 shows the full-range mass scan. The key molecular ions for elemental sulfur at 256 Da and the surrogate at 272 Da are completely obscured by a continuum of interfering hydrocarbon ions. No successful quantitation could be obtained from the straight full-scan mass spectrum. Figure 12(a) and (b) presents the daughter ion scans of *m*/*z* 256 and 272. The parent ion of *m*/*z* 256 ($^{32}S_8$) produces a strong daughter ion at *m*/*z* 160 ($^{32}S_5$), representing the loss of three atoms of sulfur-32. The parent ion of *m*/*z* 272 ($^{34}S_8$) produces a strong daughter ion at *m*/*z* 170 ($^{34}S_5$), representing the loss of three atoms of sulfur-34. The experimentally measured data for the daughter ions of *m*/*z* 160 and 170 are then normalized for the initial relative abundances of the parent ions of *m*/*z* 256 and 272 in the pure molecular clusters of the natural sulfur and the enriched sulfur. For example, the abundance of $^{32}S_8$ in natural elemental sulfur is 66.5%. Dividing by 0.665 normalizes the experimental measurements. The molecular isotopic abundances for the natural sulfur and the surrogate abundance were easily determined from the mass spectra of the pure components. Figure 13(a) and (b) shows the molecular ion distribution and relative contributions for the parent ions. Equation (1) can be used to calculate weight percentage of sulfur in the oil:

$$\text{sulfur (wt\%)} = {}^{34}\text{S (wt\%)} \frac{\text{normalized response of natural sulfur}}{\text{normalized response of } {}^{34}\text{S}} \quad (1)$$

Two precautions should be considered. First, when working with isotopic internal standards, general response factors will take into account variations of the nominal values in isotopic distributions stated by the supplier. However, when a new batch of standards is

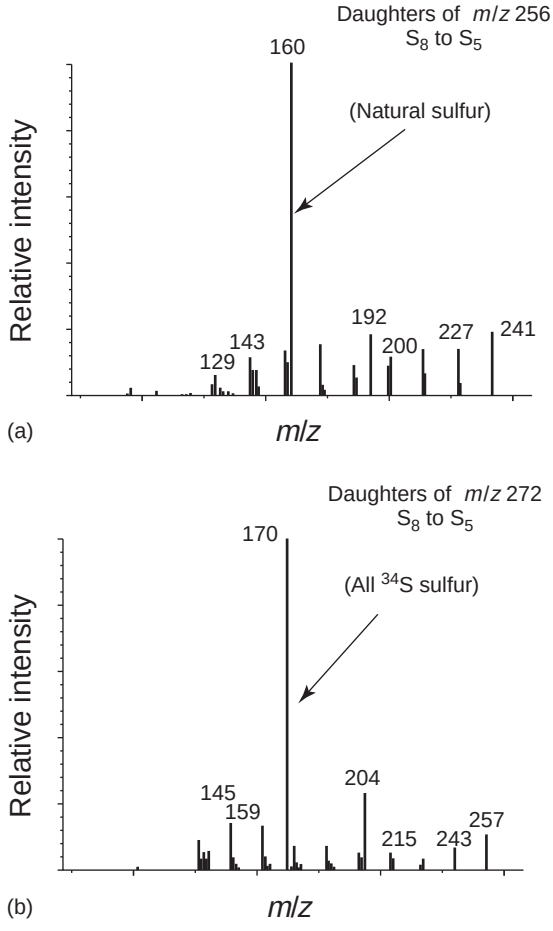

Figure 12 Daughter ion EI mass spectra of a direct insertion probe crude oil sample. S_5 fragment ions are dominant.

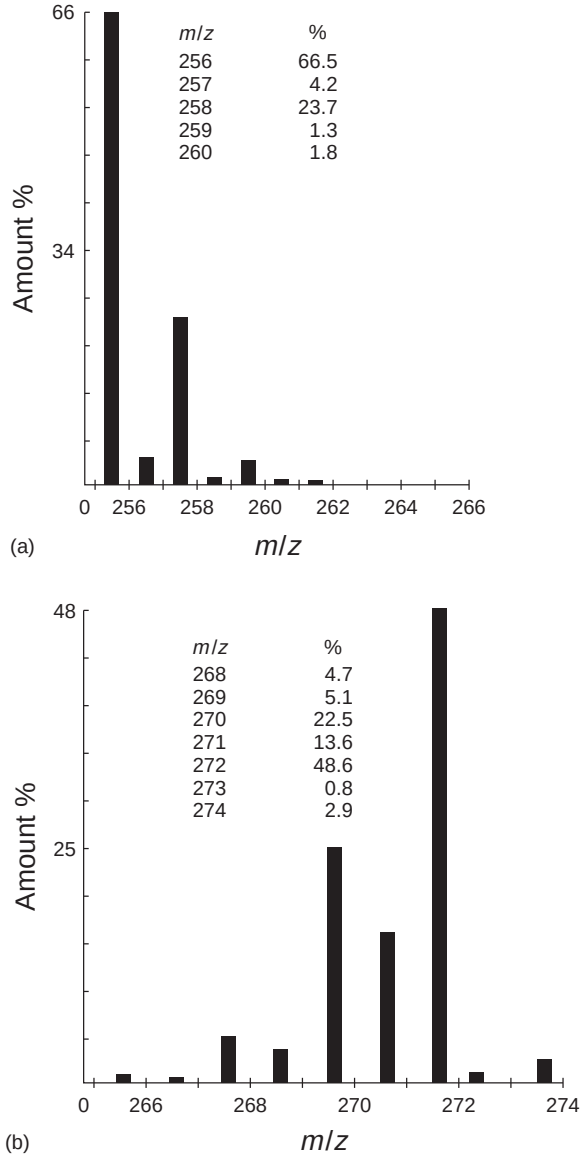

Figure 13 Molecular ion distributions for (a) elemental sulfur (S_8) natural and (b) isotopic sulfur ($^{34}S_8$).

used, new response factors must be generated. Second, elemental sulfur is hygroscopic and the extent of hydrate formation must be determined in the surrogate so that the true amount of spike will be known.

5 CONCLUSION

Specific hydrocarbon components or classes of components can be quickly identified and precisely quantitated using GC coupled with low-resolution MS. DIP/MS/MS can sometimes be used for less volatile and thermally labile petroleum components. The selection of appropriate internal surrogates will greatly enhance identification, precision, and accuracy. The application of four general strategies for analyzing components in liquid petroleum matrices serves as templates only. Very similar approaches can be applied for the analysis of just about any individual or group of semivolatile components found in any petroleum product.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
DIP/MS/MS	Direct Insertion Probe Mass Spectrometry/Mass Spectrometry
EI	Electron Ionization
EIMS	Electron Ionization Mass Spectrometry
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry

HPLC	High-performance Liquid Chromatography
MS	Mass Spectrometry
MS/MS	Mass Spectrometry/Mass Spectrometry
MTBE	Methyl *tert*-butyl Ether
OFID	Oxygen Flame Ionization Detection
PAH	Polycyclic Aromatic Hydrocarbon
QC	Quality Control
TAME	*tert*-amyl Methyl Ether
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)** Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Full Range Crudes, Analytical Methodology of • Lubricant Base Oils: Analysis and Characterization of • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams

REFERENCES

1. ASTM, 'Hydrocarbon Types in Low Olefinic Gasoline by Mass Spectrometry', Method D2789-91, American Society for Testing and Materials, Washington, DC, 1994.
2. EPA 40 CFR Part 80, *Fed. Regist.*, **59**(32), 7826–7828 (1994).
3. ASTM, 'Determination of Benzene, Toluene and Total Aromatics in Finished Gasoline by Gas Chromatography/Mass Spectrometry', Method D5769, American Society for Testing and Materials, Washington, DC, 1995.
4. D.E. Drinkwater, V.P. Nero (eds.), 'The Analysis of Reformulated Gasolines Using Fast Gas Chromatography/Mass Spectrometry', in *Transportation Fuel for a Cleaner Environment*, Taylor and Francis, London, 187–198, 1999.
5. V.P. Nero, J.A. D'Antono, D.E. Drinkwater, 'The Determination of 2-Aminonaphthalene in Petroleum Products by High-performance Liquid Chromatography and Mass Spectrometry', in *Proceedings of the 41st ASMS Conference on Mass Spectrometry Allied Topics*, San Francisco, CA, 206, 1993.
6. ASTM, 'Hydrocarbon Types in Middle Distillates by Mass Spectrometry', Method D2425-93, American Society for Testing and Materials, Washington, DC, 1994.
7. ASTM, 'Hydrocarbon Types in Gas–Oil Saturate Fractions by High Ionization Voltage Mass Spectrometry', Method D2789-91, American Society for Testing and Materials, Washington, DC, 1994.
8. ASTM, 'Hydrocarbon Types in Gas–Oil Aromatic Fractions by High Ionization Voltage Mass Spectrometry', Method D3239-91, American Society for Testing and Materials, Washington, DC, 1994.
9. V.P. Nero, D.E. Drinkwater, 'Environmental Analysis of Petroleum Hydrocarbons', in *Environmental Analysis and Remediation*, ed. R.A. Meyer, John Wiley & Sons, New York, 3668, 1998.
10. V.P. Nero, D.E. Drinkwater, 'The Rapid Analysis of Elemental Sulfur in Complex Petroleum Matrices Using Tandem Mass Spectrometry', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, 1474, 1998.

Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of

W.K. Robbins
ExxonMobil Research & Engineering, Annandale, USA

1 Introduction 1
2 Petroleum Resids 1
2.1 Origin 1
2.2 Range of Composition 2
2.3 Physical Properties 2
3 Determination of Sulfur in Petroleum Resids 3
3.1 X-ray Fluorescence 3
3.2 High-temperature Combustion and SO_2 Measurement 3
3.3 Interferences 4
3.4 S Speciation 4
4 Determination of Nitrogen in Petroleum Residue 4
4.1 Measurement Techniques 4
4.2 Interferences 5
4.3 N Speciation 5
5 Determination of Metals in Petroleum Resid 5
5.1 Measurement Techniques 5
5.2 Sample Preparation 7
5.3 Metal Speciation 8
6 Comparison with Other "Heavy Oil" Matrices 8
Acknowledgments 9
Abbreviations and Acronyms 9
Related Articles 9
References 9

When petroleum (i.e. crude oil) is distilled, a residual fraction containing molecules that do not volatilize is drawn off the bottom. This "resid" has an increased concentration of minor impurities, including sulfur-, nitrogen-, and oxygen-bearing heterocyclics, soluble nickel and vanadium compounds and a variety of trace metals. These impurities both dominate the unique properties of the resid and play a major role in determining the refining processes necessary for its conversion to marketable products. This article introduces the common analytical techniques for the measurement of sulfur, nitrogen, nickel, vanadium, and trace metals in this challenging organic matrix.

1 INTRODUCTION

Crude oils exhibit a broad range of elemental composition because they are formed by a combination of deposition, degradation, and transformation of ancient organics from different biosources.[1,2] In addition to hydrocarbons, crude oil contains a multitude of minor impurities, including sulfur-, nitrogen-, and oxygen-bearing heterocyclics, soluble nickel and vanadium compounds and a variety of trace metals that are removed in refining. When crude oil is distilled, a residual fraction containing molecules that do not volatilize is drawn off the bottom. This "resid" has an increased concentration of the heterocyclics and practically all of the metals. However, resid is valuable because it compares favorably with alternate hydrocarbon resources (bitumen, heavy oil, shale oil, and coal liquids) in terms of these deleterious contaminants and it is accessible in the refinery.[3,4] This article introduces the common techniques available for analysis of sulfur, nitrogen, nickel, vanadium, and trace metals in this challenging organic matrix.

2 PETROLEUM RESIDS

2.1 Origin

Commercially, resids are produced in the distillation unit that is the initial refinery treatment of a crude oil. In this unit, the oil continuously flows in sequence through a desalter, heat exchangers, an atmospheric pipestill (APS) and a vacuum pipestill (VPS). The bottoms from the APS (650 °F$^+$ (343 °C) or atmospheric resid) are reheated and distilled under vacuum (15–100 Torr) in the VPS. The application of vacuum allows the VPS to produce distillation bottoms (vacuum resid) with an atmospheric equivalent boiling point (AEBP) of 1050 °F$^+$ (566 °C).[3,4] In this scheme, both distillation units keep the actual temperature below 650 °F$^+$ (343 °C) to minimize cracking of the oil.

In addition to the continuous distillation in a refinery, laboratories use batch distillation methods to produce experimental "resids" with different AEBP. Distillation to a 650 °F$^+$ resid is done in accordance with ASTM (American Society for Testing and Materials) D 2892.[5] In this 15/5 distillation under reflux, the distillation temperature is gradually increased and a series of distillate fractions removed; the pressure is then reduced to ~10 Torr to reach the AEBP of 650 °F$^+$ without cracking

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the oil. The 650 °F$^+$ bottoms is then charged into a second still where fractions are distilled in accordance with ASTM D 5236 down to pressures down to 0.1 Torr and AEBP of 1050 °F$^+$. Deeper distillations (1050–1300 °F$^+$, AEBP) are carried out at high vacuum with thin film evaporators with short vapor pathlengths.[6,7] Each short-path distillation produces a distillate and a bottoms; so multiple distillations are employed to generate the bottoms from 1050 to 1300 °F$^+$.

2.2 Range of Composition

With the exception of trace metals, the typical elemental composition of crude oil and of the corresponding resids varies only slightly over narrow limits (Table 1).[6] However, the percentage of resid and percentage of sulfur in the resid differ widely among crudes (Table 2).[8] Sulfur and nitrogen concentrations increase with the final boiling point of the resid fraction (Table 3).[7] Although the numbers are not included in the tables, practically all the metals remain in the resid as well. Data for DISTAC distillation of the atmospheric residue from a heavy California crude (Kern River) show that the 1000 °F$^+$ resid is nearly half the 775 °F$^+$ resid and twice the 1300 °F$^+$ resid (Table 4).

Extensive studies have been carried out to characterize the deep vacuum distillates.[7] The resids produced in these studies show steady increases for S, N, and O from 775 to 1300 °F$^+$, with the exception of metals. While no metals are volatilized below 1050 °F$^+$, some metals do, however, distill into cuts between 1050 and 1300 °F$^+$. The results for the heterocyclics in the deep distillates demonstrate that the molecular types found in the lower boiling fractions are found in the resids, as well. Resid can be considered to fit a ball and string model where the balls are fused aromatic (or naphthenoaromatic) rings and the strings are carbon or heteroatom links between the rings. While the number of fused rings in the resid may be larger than in vacuum distillates or may contain metals absent from those fractions, most of the molecular types retain their characteristic reactivity. Therefore, many of the analyses determining the elemental and molecular-type composition of resid are a logical extension of those used for the vacuum distillates.

Table 1 Range of elemental composition for whole crude and residues

Element	Whole crude (range)	Atmospheric resid (range)	Vacuum resid (range)
Carbon (%)	83–87	86–88	83–86
Hydrogen (%)	10–14	10–13	9–13
Sulfur (%)	0.04–6.0	0.05–7	0.1–12
Nitrogen (%)	0.02–1.0	0.05–1.2	0.1–1.5
Oxygen (%)	0.05–1.5	0.2–2.5	0.5–3
Nickel (wppm)	0.1–100	0.5–200	1–300
Vanadium (wppm)	0.1–500	0.2–800	0.5–1000

2.3 Physical Properties

The physical properties of resids make them a challenge for analysis (Table 5). Nothing can be seen or measured optically in resids because they are so dark. Resids may be lighter or heavier than water (API gravity = 10).

Table 2 Atmospheric and vacuum residue distributions for various crudes[8]

Crude	Region	% S in whole crude	Atmospheric resid			Vacuum resid		
			Wt%	% S	% S in 650+	Wt%	% S	% S in 1050+
Tapis	Malaysia	0.023	17.6	0.070	53.6			
Taching	China	0.078				36.9	0.140	66.2
Brass River	West Africa	0.09	26.5	0.14	41.2			
Cabinda	West Africa	0.17	52.9	0.24	74.6			
Brega	North Africa	0.21	36.8	0.41	71.8	9.0	0.69	29.6
Statfjord	North Sea	0.27	36.9	0.6	76.5	9.8	0.74	27.0
Shengli	China	1.00				36.4	1.66	60.4
Alaskan North Slope	Alaska	1.06	52.4	1.63	80.6			
Qatar Land	Middle East	1.28	32.7	2.58	65.8	9.9	3.45	26.6
Midway Sunset	California	1.68	79.6	1.75	82.9	33.7	1.97	39.5
West Texas Sour	Texas	1.70	52.2	2.15	66.0	16.3	2.40	23.0
BCF-24	Venezuela	1.85	61.9	2.58	86.3			
Tia Juana Pesado	Venezuela	2.70	82.2	3.00	91.3			
Arab Heavy(Safaniya)	Middle East	2.85	53.8	4.34	81.9	23.2	6.00	48.8
Cold Lake	Canada	2.91	51.6	4.80	85.1			
Maya	Mexico	3.32	56.4	5.04	85.6	31.2	5.78	54.3
Basrah Heavy	Iraq	3.50	55.9	5.25	83.9	42.4	5.80	70.3
Hondo Monterey	California	5.15	67.2	5.92	77.2	38.6	6.61	49.5

Table 3 Metals and heteroatoms concentrate in bottoms from Tia Juana crude

Boiling range	Yield	S (%)	S (%)	N (%)	V (ppm)	Ni (ppm)	Fe (ppm)
Whole crude	100	1.08	100	ND			
565+	57.4	1.64	87.2	ND			
650+	48.9	1.78	80.6	0.33	185	25	28
700+	44.4	1.84	75.6	0.36			
750+	39.7	1.93	70.9	0.39			
850+	31.2	2.12	61.2	0.45			
950+	23.8	2.35	51.8	0.52			
1050+	17.9	2.59	42.9	0.6	450	64	48

ND = no data or not determined.

Table 4 Residues from Kern River Crude as a function of boiling point (from Speight,[6] p. 436)

Initial boiling point of residue	Whole crude (%)
775	78.4
850	67.2
925	57.2
1000	48.6
1050	41.7
1125	35.7
1200	31.1
1300	25.7

Table 5 Physical properties of atmospheric and vacuum resids[a]

Property	Atmospheric resid	Vacuum resid
Boiling range (°F)	650+	1050+
API gravity	10–35	−5–25
Viscosity (cSt, 210 °C)	6–100	350–55 000
Carbon residue (%)	1–14	5–35
Pour point (F)	20–140	75–200
Asphaltenes	1–15	2–30

[a] Physical property ranges summarized from data in Speight,[6] Altgekt and Boduszynski,[7] and Kinney.[8]
API, American Petroleum Institute.

Measurements by ASTM tests show that resids vary in solubility (asphaltenes), viscosity, and tendency to form coke (carbon residue), all factors that complicate sample preparation.

3 DETERMINATION OF SULFUR IN PETROLEUM RESIDS

There are several standard methods suitable for measuring sulfur in resids. Sulfur can be measured directly in the resid matrix with X-ray fluorescence (XRF) or indirectly by titration or infrared detection (IR) after combustion to SO_2. Standard analytical practice includes a number of variations on these two approaches. Other S-sensitive methods are not as well suited for resid.

3.1 X-ray Fluorescence

For XRF analysis, a sample is placed in a disposable cell with a thin Mylar® or polypropylene window, the cell placed (window down) over a fluorescence detector, and the contents of the cell irradiated at fixed angle by an X-ray tube with Cu, Rh or Cr target or a radioisotope (Fe-55).

There are several options for fluorescence detection. Samples containing 0.1–4.0% can be analyzed with a nondispersive (radioisotope) or energy dispersive detector (ASTM 4294).[5] Instruments are calibrated frequently with sulfur in resid reference samples. While large amounts of heavy metals may interfere, the amount of iron, nickel and vanadium usually found in residual oils does not. The nondispersive and energy dispersive spectrometers are simple enough to be incorporated into portable instruments.

Lower levels can be measured at the sulfur K_α line at 5.373 Å with wavelength dispersive spectrometers (ASTM D 2622).[5] Sulfur levels from 0.0003 to 5.3% can be measured directly after the addition of an internal standard. The background intensity measured at a recommended wavelength of 5.190 Å (5.437 Å for a Rh target) is subtracted from the peak intensity. The resultant net counting rate is then compared to a previously prepared calibration curve or equation to obtain the concentration of sulfur in mass percent.

3.2 High-temperature Combustion and SO_2 Measurement

High-temperature combustion followed by SO_2 detection offers an alternative fast method for determining sulfur in resids without matrix-matched standards. In the combustion techniques, samples are burned under controlled conditions and the mixture of SO_x gases measured by titration or IR detection. ASTM D-1552[5] includes two types of combustion (tube furnace and induction furnaces) and two methods of measurement (iodate titration and IR detection of SO_2).

To achieve complete combustion of the refractory resid (high boiling, viscous and tendency to form coke), the samples must be burned "smoothly" to SO_2, i.e. without splatter, flash, or formation of coke. In the tube furnace combustion, a sample is weighed into a ceramic boat, MgO and V_2O_5 added as combustion accelerator, placed into the cool portion of the furnace, and a flow of oxygen established over the sample. The boat is then pushed

into the combustion zone set at 1371 °C (2500 °F) to produce SO_x where SO_x is ~97% SO_2. In the induction furnace combustion, samples are weighed into a ceramic crucible, covered with an inductor (powdered iron or tungsten) and accelerator (vanadium pentoxide). The crucible is covered and placed into an oxygen stream in the instrument where it is heated for 90 s by an induction coil.

For the titration measurement, the combustion products are passed into an absorber containing an acid solution of potassium iodide and starch indicator that is mounted on an autotitrator equipped with a photodetector. Initially, a faint blue color is established in the absorber solution by the addition of standard potassium iodate solution. As SO_2 is evolved, it bleaches the blue color and the autotitrator adds more iodate. The volume of standard iodate consumed during the combustion is a measure of the sulfur content of the sample.

For the IR measurement, the combustion products are passed through traps that scrub out CO_2, moisture and dust prior to IR absorption measurement. The area response of the IR absorbance is used to calculate the mass percent of sulfur for the sample weight. The IR calibration factor is determined using standards approximating the material to be analyzed. Accuracy is increased if standards bracketing the sample are also analyzed. The combustion techniques have comparable precision but are less convenient than XRF.[9]

3.3 Interferences

For XRF, errors in the sulfur determination can result when the elemental composition (excluding sulfur) of samples differs significantly from the standards. For example, differences in the carbon–hydrogen ratio in sample and calibration standards induce errors in the determination. Compensation is commonly made by the addition of an internal standard or use of a standard addition technique. Because the distance between the sample and detector affects the response, the Mylar® on the cell must be flat, the sample must be free of sediment (or diluted with solvents that do not induce asphaltene formation), and "infinitely thick" (>15 mm) for consistent excitation. For combustion tests, the iodate titration can be affected by >1% chlorine and >0.1% nitrogen, while IR detection will tolerate somewhat higher concentrations.

3.4 S Speciation

Both wet chemical and instrumental techniques show that sulfides and thiophenes are the major sulfur types in resid. Characterization studies of the deep distillates suggest that both types are present within the multiring structures in atmospheric and vacuum resids. Iodate titration and iodine complex tests show that the fraction of sulfur as sulfide decreases as resid boiling point increases.[10,11] Roughly two-thirds of the sulfur in atmospheric resid is estimated to be thiophenic by a variety of techniques: iodate titration,[12] by NMR (nuclear magnetic resonance) measurements after derivatization,[13] excited X-ray absorption fine structure (EXAFS),[14] and X-ray absorption near edge spectrometry (XANES).[15] Selective reduction experiments suggest that the sulfides in resid exist in links between multiring structures as well as in the rings.[16] Sulfoxides detected by XPS (X-ray photoelectron spectrometry) may be limited to the thin layer (<10 molecules) sampled by this technique. Other advanced analytical techniques for elucidating the structure of sulfur species in various fossil resource materials have been reviewed.[17]

4 DETERMINATION OF NITROGEN IN PETROLEUM RESIDUE

4.1 Measurement Techniques

Nitrogen in resid is commonly measured by two techniques. The classic Dumas technique is used for samples containing >0.2% nitrogen; for samples containing less, nitrogen can be measured down to 0.004% by combustion and NO_x chemiluminescence (ASTM D 5762).[5] Other N-measurement techniques are not well suited for the analysis of the forms and concentrations of N found in resids.

The Dumas technique is based on high temperature combustion, gas scrubbing, and reduction of nitrogen oxides for measurement as N_2 (ASTM D-5291 Method C). It is applicable to samples containing N between 0.1 and 50% N, irrespective of the form of nitrogen. In modern instruments, a sample is weighed into a tin capsule, combusted in oxygen (as O_2) at 1700 °C, and then the gases produced are swept through a train of gas-treating tubes with a helium carrier gas into a ballast reservoir. The gas treaters include chromic oxide (oxidizer) and silver-coated cobalt oxide (trap for halogens and sulfur). The scrubbed gases (CO_2, H_2O, NO_x) then are passed over copper at 650 °C to reduce the NO_x gases to molecular N_2. This mixture is then carried through molecular sieves to remove water and through supported magnesium perchlorate to remove CO_2. The N_2 in the He gas is detected with a thermoconductivity detector.

In the chemiluminescence test for resid, a boat inlet system is used. The sample is placed on a boat at room temperature, the boat placed in a combustion tube, and a flow of oxygen–argon established. The boat is advanced into a high-temperature combustion tube where the nitrogen is oxidized to nitric oxide (NO) that

is then allowed to react with ozone creating an excited form of nitrogen dioxide.[5] The light emitted when the excited NO_2^* decays is detected by a photomultiplier tube and the resulting time-integrated signal is a measure of the nitrogen contained in the sample. The detector is calibrated with model nitrogen compounds. Although the scope of the test can cover the 1–1000-wppm range the linearity of the detector is much more limited. Good laboratory practice would suggest using a calibration curve with nitrogen contents close to the range of the samples being analyzed.

4.2 Interferences

The Dumas and chemiluminescence methods are generally free of interference. Assuming that a representative sample has been taken and is within the calibration range of the methods, these analyses can be applied to most resids. As long as the sample is soluble in the mixed solvent used for the titration, the basic nitrogen in resids can also be measured routinely.

4.3 N Speciation

Deep HVGO (heavy vacuum gas oil) distillate data suggest that the most common nitrogen compound types in resid are pyrrole and pyridine benzologs.[7] Dilute solutions of resid in deuterochloroform detect the pyrrolic N–H at 3650 cm^{-1}.[18] NMR studies after derivatization with strong base and labeled methyl iodide confirm the presence of that structure. This ''acidic nitrogen'' accounts for roughly two-thirds of the total.[19] Basic nitrogen titrations suggest similar ratios of these two main nitrogen types.[20] Fourier transform infrared (FTIR) and titration data indicate the presence of amides in some resids.[21] No evidence is found for alkyl or aromatic amines in resid.

5 DETERMINATION OF METALS IN PETROLEUM RESID

5.1 Measurement Techniques

Nickel and vanadium are present as organometallics in crudes at levels characteristic of their geochemical environment.[1,2] Although sodium and iron are also commonly measured in resid, these metals are more likely to be present as entrained brine (Na) or scale and soluble corrosion products (Fe). Other metals have been found in soluble form at trace levels (<0.1 ppm) and at higher levels in the form of entrained brine or corrosion products (scale). The metals in resid can be measured by a number of instrumental techniques, among them XRF, flame atomic absorption (FAA), graphite furnace atomic absorption (GFAA), inductively coupled plasma atomic emission spectrometry (ICPAES), and instrumental neutron activation analysis (INAA).[22,23]

5.1.1 X-ray Fluorescence

The direct XRF techniques used for sulfur can be applied for Fe, Ni, and V if the metal concentrations are >10 ppm. XRF has poor sensitivity for Na. Although XRF has been reported to measure trace metals after acid digestion, precipitation and filtration, the results appear erroneously high.[24]

5.1.2 Flame Atomic Absorption

In FAA, a sample is aspirated into a laminar acetylene–air or acetylene–nitrous oxide flame. The emission lines of the element to be measured (produced by a hollow cathode lamp) are focused through the flame into a spectrometer that measures the absorption of one of the characteristic lines. Corrections are made for background absorption using Zeeman effects or using broad band absorption from a deuterium lamp. Peak heights are measured and results quantified using standards prepared in similar matrix types.[25]

Both direct and indirect FAA methods have been included in standard methods for the determination of Ni, V, Na, and Fe in crude oils and residual fuels by FAA (ASTM D 5863).[5] In the direct method, samples are diluted to 5 to 20% in an organic solvent (xylene or tetralin), depending on the anticipated metal content. Separate analyses are made for sodium (589.0 nm), iron (248.3 nm), and nickel (232.0 nm) as they are aspirated directly into an air–acetylene flame. Concentrations are calculated from calibration curves prepared in the same manner using oil-soluble metals standards. Vanadium is analyzed at 318.4 nm using a nitrous oxide–acetylene flame with 15-ppm Al (oil-soluble standard) added as an ionization suppressant. The dilution, needed to minimize the effect of resid matrix, limits direct FAA to Ni, V, Na and Fe.

In contrast, the matrix is destroyed in the indirect method. Samples of 1 to >20 g can be wet ashed with sulfuric acid and the ash taken up into a dilute nitric acid solution. The solution is then analyzed for each element separately by FAA at the same wavelengths as the direct method. For vanadium, 250-ppm Al is added to suppress ionization and measurements made with a nitrous oxide–acetylene flame. Concentrations are determined from calibration curves in the same aqueous acid matrix.

5.1.3 Graphite Furnace Atomic Absorption

In GFAA, a hollow graphite tube furnace in an inert gas atmosphere is mounted in the optics of an atomic absorption spectrometer. A sample is introduced into

the furnace. The tube is heated (by electric resistance) through a sequence of drying (100–300 °C), ashing (600–1500 °C) and atomization (up to 3000 °C) stages. During the atomization stage, the peak height is measured and corrections made as in FAA. It is not possible to give specific heating cycles because the optimum conditions vary from element to element and between instrument designs. In indirect GFAA, where the samples are ashed prior to injection, the generic (aqueous-based) detection limits for GFAA are generally lower than for FAA, but actual values differ between instruments.[25]

GFAA has also been applied to the direct analysis of metals in crude oils after dilution with toluene or xylene.[26] Nickel measurements are routine; vanadium requires addition of some ashing aid (toluene sulfonic acid, for example) to avoid losses of vanadyl species by volatilization.[27] Some trace metals have been measured by direct GFAA, but each analysis requires optimization of the heat cycles.

5.1.4 Instrumental Neutron Activation Analysis

INAA can be used to measure the metals as a survey technique.[28,29] In INAA, a representative sample is sealed in an inert polyethylene capsule. The capsule is then pneumatically conveyed into a nuclear reactor where it is exposed to a high-intensity neutron flux (for example, 5×10^{12} n cm^{-2} s^{-1}) for a predetermined time (5 min to 8 h). After irradiation, the capsule is moved to a shielded detector until the background emission from short halflife isotopes has decreased sufficiently to allow the elements of interest to be detected. Radiation from the capsule is measured for 3 min to 10 h by a γ-ray spectrometer equipped with a lithium-drifted germanium crystal detector (Table 6).

Table 6 Irradiation conditions for selected nucleotides measured by INAA

Irradiation time (@ 5×10^{12} n cm^{-2} s^{-1})	Measurement time (minutes)	Nucleotide measured	Detection limit (ng g^{-1})
5 min	3 min	^{28}Al	NR
		^{27}Mg	NR
		^{52}V	2
5 min	20 min	^{56}Mn	15
8 h	80 min	^{76}As	5
		^{24}Na	20
		^{42}K	NR
8 h	10 h	^{51}Cr	20
		^{59}Fe	50
		^{58}Co (for Ni)	100
		^{75}Se	10
		^{65}Zn	25

NR = not reported. From Filby and Olsen.[28]

5.1.5 Inductively Coupled Plasma Atomic Emission Spectrometry

In ICPAES, a sample solution is pumped into a 10 000-K argon/oxygen plasma torch. At this temperature, all the organics are destroyed and the metals converted to excited atoms and ions. The emission from the torch is dispersed by a spectrometer onto an array of photodetectors positioned to measure one of characteristic atomic emission lines of the selected elements. Results are determined from calibration curves using external standards. Each ICPAES spectrometer is unique in design, selected wavelengths, and computer capabilities; suggested emission wavelengths are given in standard methods (ASTM D 5708).[5] All measured emission peaks are referenced to a specific element added as internal reference.

In ASTM D5708 (Method A),[5] ICPAES is used in a direct method that measures Ni, V, and Fe. In the method, a sample is diluted to 10% (w/w) with xylene or tetralin. Results are determined from calibration curves using standards in a white oil diluted with the same dilution solvent. The direct ICPAES method is modified slightly for use in measuring trace metals (Ag, Al, B, Ba, Ca, Cd, Cr, Cu, Fe, Mg, Mn, Mo, Na, Ni, P, Pb, Si, Sn, Ti, V, and Zn). A sample is blended with an internal reference (Co) to correct for viscosity effects. Typical working ranges are in the 0.05–10-ppm range.

In the indirect ASTM test for Ni, V, Fe, and Na (D5708-Method B),[5] resid samples are combined with sulfuric acid in a beaker and heated to dryness on a hot plate to char the sample and convert the metals to sulfate salts. The beaker is then placed in a muffle furnace at 525 °C overnight to burn off the char. The inorganic residue is digested with nitric acid and evaporated to incipient dryness. The contents of the beaker are then dissolved in dilute nitric or hydrochloric acid, the solution made to volume, and the metals quantified by ICPAES using aqueous calibration standards. Detection limits vary with the size of sample used and the spectroscopic sensitivity of each metal to be measured. Metals measured by this technique differ slightly from the direct technique (Al, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, Pb, Sb, Tl, V, Y, Zn, and Zr).

5.1.6 Inductively Coupled Plasma Mass Spectrometry

Inductively coupled plasma (ICP) capabilities have been expanded and detection limits lowered by the introduction of inductively coupled plasma mass spectrometry (ICPMS).[30,31] The latter combines ICP excitation with mass spectrometric detection. In this mode, the ICP torch generates a cloud of ions that are accelerated into a mass spectrometer that resolves the excited atoms. Because the torch also destroys the sample matrix, ICPMS can be applied directly to oil samples dissolved

in xylene. In addition to dilution with solvent, resids have been analyzed by ICPMS in the form of aqueous microemulsions.[32] Both internal standard (Bi, Sc, etc.) and standard addition techniques have been used for quantification.[22] If samples are ashed and analyzed in diluent aqueous acid, the ICP methods have even better sensitivity and are less subject to interference.[33] However, care must be taken to avoid Cl (brine salts, methylene chloride or HCl) because the ClO formed in the torch interferes with the measurement of V at mass 51.

5.1.7 Choice of Metals Method

Trace metals can be measured in resid by many of the above techniques. Nickel and vanadium in resids can be determined rapidly by XRF at levels above 10 ppm. Many of the FAA and GFAA techniques developed for individual trace elements in the 1970s are still in use. Where applicable, direct GFAA affords the fastest technique for determining an individual element. For greater accuracy and sensitivity, indirect FAA and GFAA remain the methods of choice for individual elements.

Multielement survey data can be collected on resid by INAA and ICP techniques that use expensive specialized instrumentation. ICP methods have largely supplanted INAA for survey measurements of trace elements because they have lower detection limits for most elements. Direct ICPAES can be used to rapidly screen soluble metals; the indirect version can be more sensitive and can quantify suspended metals. The same restrictions apply to ICPMS which has several orders of magnitude more sensitivity for many elements. INAA and ICP results for Fe, Ni, and V are in excellent agreement with other techniques. However, the advantages of INAA (multielement, simplicity of sample preparation, minimal matrix effects, few interferences) are balanced by long analysis times caused by decay periods and specialized training needed for working with reactor and radioactivity.

5.2 Sample Preparation

5.2.1 Sampling

Crude oils are produced with varying amounts of water (actually salt-rich brines) and sediment; traded crudes always carry some portion of these nonvolatile impurities. Distillation leaves sediment and (now dry) salts in the resid. After desalting, uptake of corrosion scale and addition of process aids, the refinery resid differs from the laboratory resid obtained by distillation of a crude oil that retains the sediment and dry salts.

In comparison with Ni and V, trace metal analyses show wide discrepancies. The variations among reported values appear to be related to sample homogeneity and preparation technique rather than the detection system. This has been reinforced by a systematic study of trace metals in whole crudes that shows natural metals concentrations two orders of magnitude lower than reported elsewhere. A statistical evaluation of the data scatter suggests that many of the trace metals are associated with contamination in transportation or with suspended sediment and water.[34,35]

For accurate trace metals analyses of a submitted sample, relatively large portions of resid should be taken as a heterogeneous material after thorough mixing. Resids can be mixed after heating to a free flowing liquid or as a dry powder after freeze-fracturing with dry ice.

5.2.2 Direct Methods

The physical properties of resid can play a significant role in the direct analysis of metals in resid. Inherently, the direct methods assume that the metals are homogeneously distributed in the resid sample. Because the resids are so dark, it is difficult to determine if sediments are present. If the metal is present as particulate (scale or salts from entrained brine), the direct techniques can give erroneous results. High results may be obtained in XRF if a metal-rich sediment settles onto the Mylar® window in the test cell; low results may occur if particulates escape the atomic absorption flame or ICPAES torch. Variability in direct analyses by ICP methods can arise from factors that influence the sample flow into the ionization region, especially viscosity, surfactants and choice of solvent, and calibration should be done with oils that closely match the sample matrix.

As noted above, metals can be lost in the dry ashing provided as part of the GFAA cycle. Published detection limits for aqueous solutions may not be achievable by direct GFAA analysis of resid samples. The severe ashing conditions required to minimize background interference from resids can volatilize some metals. This is particularly true of the Ni and V porphyrins in resids, but can also apply to the more volatile metals such as mercury, selenium, arsenic, cadmium, antimony and lead. To reduce the severity of ashing conditions and eliminate metals loss, homogenous ashing aids (such as *p*-xylenesulfonic acid) can be added to the diluent.[24] The materials accelerate the destruction of the resid allowing lower ashing temperatures and also convert the metals into a microsulfated ash, i.e. sulfates and oxides that are less volatile than indigenous metal compound types.

5.2.3 Indirect Methods

Most heavy metals are not lost in wet ashing of resid. The exceptions, metals not readily oxidized or inherently

volatile as the oxide (Hg, Se, As, etc.), have been studied separately. The initial addition of sulfuric acid prevents losses of some metals if chlorides are present, for example V as $VOCl_3$. The sample should be well mixed with sulfuric acid and heated gently at first to avoid splatter. Numerous variations have been suggested for ashing samples for trace metals including alternate oxidizing agents, different acids for dissolution, and reagents for fluxing the sulfated ash. Microwave ashing has been substituted for the sample preparation of resid samples for metals analyses. In microwave ashing, the resid is combined with oxidizing acids (sulfuric, nitric, perchloric or mixtures thereof), sealed in Teflon™ vessel, and heated in a microwave oven for a brief period. The resulting solutions are washed from the vessel with aqueous acid, an internal standard added, and made up to volume. Because it uses concentrated heating in a closed vessel, microwave ashing is faster and less subject to contamination than sulfated ashing. On the other hand, their microwave techniques are limited to small samples and care has to be taken to avoid explosions with the sealed vessel. Thus, sulfated ashing, is preferred for routine analyses because it requires less technique and no specialized equipment.

5.3 Metal Speciation

The consistency of the Ni and V results among techniques demonstrates that these elements are homogeneously dispersed in the colloidal resid.[36] A portion of the metals can be measured by the Soret bands characteristic of porphyrins in the ultraviolet–visible spectra.[6,7] Although a larger fraction of the metals are described as ''nonporphyrin'' structures,[37,38] X-ray absorption fine structure (XAFS) analyses show that *all* the Ni and V resid is in molecular environments similar to those in porphyrins.[39,40] Thus, the nonporphyrins appear to be mixtures of closely related tetrapyrrole analogs. This direct spectroscopic evidence is more compelling than the indirect data (reactivity, size exclusion chromatography, ESR (electron spin resonance), etc.) from which alternate structures have been inferred. The lack of a Soret band for these metals may be rationalized by (i) molecular association of tetrapyrrole structures with electron-rich aromatics or heteroaromatics; (ii) a mixture of numerous isomeric porphyrin homologs with overlapping spectra, (iii) molecules with extended conjugation (benzoporphyrins), or a combination of all these.

Among the trace elements, Fe in resid can appear in both soluble (indigenous naphthenates and corrosion products) and insoluble (scale and iron sulfide) forms.[23] For accurate total iron measurements, indirect techniques should be used on well-mixed samples. Other trace metals in resid may also be present in either soluble or suspended form.[1,2] The suspended forms, arising from brine salts, probably contribute more to the trace metal load than soluble metals.[41] The latter are generally naphthenate soaps of basic metals (groups I, II, VIII).[23] Traces of elements that can form organometallics (As, Hg, Se, etc.) are found in crude, but rarely survive distillation.[42]

6 COMPARISON WITH OTHER "HEAVY OIL" MATRICES

Resids are often compared with other viscous black oils that are rich in trace metals and heteroatoms. Although some of these ''oils'' are extracts and others are pyrolyzates, they are lumped together because they represent a major hydrocarbon resource that challenges conventional crude oil processing. In most cases, the methods used for the analysis of resids can be applied to these alternate ''heavy oils''. The following briefly notes the characteristics of each type and the unique difficulties that each presents for the methods.

The elemental composition of oil sand bitumen falls generally into the range of petroleum resid with 1–10% S, 0.4–1.0% N and nickel and vanadium contents hundreds or thousands of parts per million. The higher molecular weight, viscous nature of these oils presents basically the same challenges as found for resids, that is, taking representative samples, avoiding solvent precipitation, and preventing coke formation prior to measurement. The analyses of S, N, Ni, V, and Fe in bitumen, extra heavy oils and tars have been widely reported.[6,7] Classic combustion techniques are routinely applied for S and N. Direct INAA or indirect ICPAES are favored for metals.[28]

Most of the techniques used for petroleum resids can be applied to shale oils with adjustment of sample sizes to keep within the instrumental calibration range.

Table 7 Heteroatom content of some raw shale oils[4] determined by different processes

Process	Paraho DH	Occidental MIS	Dravo	CSR
Shale Source	Colorado	Colorado	Kentucky	Australia
Nitrogen (wt%)	2.02	1.51	1.14	1.3
Oxygen (wt%)	1.38	0.65	1.97	1.3
Sulfur (wt%)	0.70	0.64	3.86	4.9
Iron (wppm)	53	45	110	55
Nickel (wppm)	2.4	6.7	–	3
Vanadium (wppm)	0.17	0.42	ND	153
Arsenic (wppm)	28	27.5	100	26

ND = no data or not determined.

Shale oils are rich in olefins, diolefins and nitrogen. The nitrogen-compound types that best survive the retorting conditions are nitriles, amides and pyrroles.[43] In presence of air, the latter combine readily with the olefinics to form sediments. Thus for quantitative analyses, shale oil samples must be protected from air. The boiling point distribution of heteroatoms of retort oils differs from petroleum resid. As a cracked product, the shale oil exhibits the decrease in sulfur concentration with boiling point while the concentration of nitrogen increases. Oxygen levels are typically higher than resid, but lower than coal liquids. Crude shale oils contain soluble forms of iron, nickel, and arsenic that can be measured by resid techniques.[44] Care must be taken to avoid losses of arsenic in volatile forms, however. Heteroatom contents of shale oils from a variety of sources show variability (Table 7):

ACKNOWLEDGMENTS

The author would like to acknowledge the valuable discussion with Dr Frank McElroy of ExxonMobil Research & Engineering in the preparation of this document.

ABBREVIATIONS AND ACRONYMS

AEBP	Atmospheric Equivalent Boiling Point
API	American Petroleum Institute
APS	Atmospheric Pipestill
ASTM	American Society for Testing and Materials
ESR	Electron Spin Resonance
EXAFS	Excited X-ray Absorption Fine Structure
FAA	Flame Atomic Absorption
FTIR	Fourier Transform Infrared
GFAA	Graphite Furnace Atomic Absorption
HVGO	Heavy Vacuum Gas Oil
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
INAA	Instrumental Neutron Activation Analysis
IR	Infrared
NMR	Nuclear Magnetic Resonance
VPS	Vacuum Pipestill
XAFS	X-ray Absorption Fine Structure
XANES	X-ray Absorption Near Edge Spectrometry
XPS	X-ray Photoelectron Spectrometry
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Neutron Activation in Environmental Analysis

Industrial Hygiene **(Volume 5)**
Dust, Measurement of Trace Elements in

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of • Lube Products. Molecular Characterization of Base Oils • Lubricant Base Oils: Analysis and Characterization of • Oil Shale and Shale Oil Analysis • Petroleum Residues, Characterization of

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Atomic Spectroscopy **(Volume 11)**
Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. B.P. Tissot, D.H. Welte, *Petroleum Formation and Occurrence*, 2nd edition, Springer Verlag, New York, 1984.
2. J.M. Hunt, *Petroleum Geochemistry and Geology*, W.H. Freeman, San Francisco, 1979.
3. J.G. Speight, *The Desulfurization of Heavy Oils and Residua*, Marcel Dekker, New York, 1982.
4. J.G. Speight (ed.), *Fuel Science and Technology Handbook*, Marcel Dekker, New York, 1990.
5. ASTM, *Annual Book of ASTM Standards, American Society for Testing Materials*, West Conshohocken, PA, 1999.

6. J.G. Speight, *The Chemistry and Technology of Petroleum*, 3rd edition, Marcel Dekker, New York, 1998.
7. K.H. Altgekt, M.W. Boduszynski, *Composition and Analysis of Heavy Petroleum Fractions*, Marcel Dekker, New York, 1994.
8. G.T. Kinney, *Oil and Gas Journal Databook*, Pennwell, Tulsa, 23–67, 1987; 235–276, 1992.
9. K. Sakamoto, T. Koyama, Y. Sakamoto, 'Comparison of the Data of Sulfur Content in Heavy Oils by X-ray Fluorescence Method and Combustion Method', *Wakayama-ken Eisei Kogai Kenkyu Senta Nenpo*, **33**, 58–61 (1987).
10. A. Somogyvari, L. Zanzotto, D. Sieben, J. Auger, 'The Aliphatic Sulfide Sulfur Content of Some Western Canadian Crudes, their Distillates, and Residues', *Fuel*, 1269–1272 (1988).
11. O.P. Strausz, E.M. Lown, J.D. Payzant, 'Nature and Geochemistry of Sulfur Containing Compounds in Alberta Petroleums', *Geochemistry of Sulfur in Fossil Fuels*, ACS, Washington, 366–396, 1990.
12. L. Wang, T.K. Green, 'Structural Characterization of Organic Sulfur Compounds in Petroleum by S-Methylation and 2H and ^{13}C NMR', *Shiyou Xeubao, Shiyou Jiagong*, **14**, 80–86 (1998).
13. Z. Wang, X. Li, G. Que, W. Liang, R.H. Xuebao, 'Determination of Sulfidic and Thiophenic Sulfur in Shengli and Gudao Vacuum Residues', Chemical Abstracts Number 124, 121571, 1996.
14. J.M. Ruiz, B.M. Carden, L.J. Lena, E.J. Vincent, J.C. Escaller, 'Determination of Sulfur in Asphalts by Selective Oxidation and Photoelectron Spectroscopy for Chemical Analysis', *Anal. Chem.*, **54**, 688–691 (1982).
15. G.S. Waldo, R.M.K. Carlson, J.M. Moldowan, K.E. Peters, J.E. Penner-Hahn, 'Sulfur Speciation on Heavy Petroleums: Information from X-Ray Absorption Near-edge Structure', *Geochim. Cosmochim. Acta*, **55**, 801–814 (1991).
16. P. Peng, A. Morales-Izquierdo, A. Hogg, O.P. Strausz, 'Molecular Structure of Athabasca Asphaltene: Sulfide, Ether, and Ester Linkages', *Energy Fuels*, **11**, 1171–1187 (1997).
17. C.E. Snape, S.C. Mitchell, K. Ismail, R. Garcia, 'Speciation of Organic Sulfur Forms in Heavy Oils, Petroleum Source Rocks, and Coals', in *Reviews on Analytical Chemistry – Euroanalysis VIII*, eds. D. Littlejohn, D.T. Burns, Royal Society of Chemistry Special Publication, Vol. 154, 103–120, 1994.
18. K. Schuchardt, 'Determination of Nitrogen and Oxygen in Selected Fractions of High-boiling and Non-boiling Petroleum Products', *Dtsch. Ges. Mineraloelwiss. Kohlechem.*, **236**, 162–166 (1983).
19. J.A. Green, J.B. Green, R.D. Grigsby, C.D. Pearson, J.W. Reynolds, J.Y. Shay, G.P. Sturm, J.S. Thomson, J.W. Vogh, R.P. Vrana, S.K.T. Yu, 'Analysis of Heavy Oils, Method Development and Application to Cerro Negro Petroleum', Topical Report NIPER 452, US Department of Commerce, 1989.
20. J.F. McKay, J.H. Weber, D.R. Latham, 'Characterization of Nitrogen Bases in High Boiling Petroleum Distillates', *Anal. Chem.*, **48**, 891–898 (1976).
21. J.B. Green, G.P. Sturm, J.Y. Shay, J.A. Green, J.W. Reynolds, L.L. Young, J.S. Thomas, 'Compound Class Separation of Petroleum Resids: Relating Coke Yields and Heteroatom Content to Chemical Structure', ACS Symposium on Exploration, Characterization and Utilization of California Heavy Fossil Fuel Resources, Preprints Division of Petroleum Chemistry, American Chemical Society, Washington, DC, 37, 956–968, 1992.
22. M. Gonzales, A.W. Lynch, 'Methods Development for Trace Metal Characterization of Crude Oil by Inductively Coupled Plasma Atomic Emission Spectroscopy', ASTM Special Publication 1109 (Instrumental Methods of Elemental Analysis in Petroleum Products and Lubricants), 62–69 (1991).
23. P. Jones, 'Trace Metals and Other Elements in Crude Oil: A literature Review', BP Research Centre, Sunbury, September 1975.
24. A.A. Olajire, R.A. Oderinde, 'Trace Metals in Nigerian Crude Oils and Their Heavy-end Distillates', *Bull. Chem. Soc. Jpn.*, **66**, 630–632 (1993).
25. J.L. Fabec, M.L. Ruschak, 'Determination of Nickel, Vanadium and Sulfur in Crudes and Heavy Crude Fractions by Inductively Coupled Argon Plasma Atomic Emission Spectrometry and Flame Atomic Absorption Spectrometry', *Anal. Chem.*, **57**, 1853–1863 (1985).
26. M.C. Gonzales, A.R. Rodriguez, V. Gonzalez, 'Determination of Vanadium, Nickel, Iron, Copper, and Lead in Petroleum Fractions by Atomic Absorption Spectrophotometry with a Graphite Furnace', *Microchem. J.*, **35**, 94–106 (1987).
27. A.P. Udoh, S.A. Thomas, E.J. Ekanem, 'Application of *p*-Xylenesulfonoic Acid as Ashing Reagent in the Determination of Trace Metals in Crude Oil', *Talanta*, **39**, 1591–1595 (1992).
28. R.H. Filby, S.D. Olsen, 'A Comparison of Instrumental Neutron Activation Analysis and Inductively Coupled Plasma Mass Spectrometry for Trace Element Determination in Petroleum Geochemistry', *J. Radioanal. Nucl. Chem.*, **180**, 285–288 (1994).
29. K.R. Shah, R.H. Filby, W.A. Haller, 'Determination of Trace Elements in Petroleum by Neutron Activation Analysis', *J. Radioanal. Chem.*, **6**, 413–422 (1970).
30. H.M. Al-Swaidan, 'The Determination of Lead, Nickel, and Vanadium in Saudi Arabian Crude Oil by Sequential Injection Analysis/Inductively-coupled Plasma Mass Spectrometry', *Talanta*, **43**, 1313–1319 (1996).
31. J. Li, L. Wu, C. Ji, 'Determination of Ultra-trace Metals in Petroleum by ICP–MS', *Yankuang Ceshi*, **17**, 290–295 (1998).
32. C.J. Lord, 'Determination of Trace Metals in Crude Oil by Inductively Coupled Plasma Mass Spectrometry with

Microemulsion Sample Introduction', *Anal. Chem.*, **63**, 1594–1599 (1991).
33. H.M. Al-Swaidan, 'Trace Metal Determination by Wet Ashing ICP/MS in Saudi Arabian Crude Oils', *Anal. Lett.*, **23**, 1345–1356 (1990).
34. C. Dekker, R. Daane, 'Metals Contents in Crudes Much Lower Than Expected', *Oil Gas J.*, March 1, 44–51 (1999).
35. B. Hitchon, R.H. Filby, K.R. Shah, 'Geochemistry of Trace Elements in Crude Oils, Alberta, Canada', in *The Role of Trace Metals in Petroleum*, ed. T.F. Yen, Ann Arbor Science, Ann Arbor, MI, 111–121, 1975.
36. D.A. Storm, E.Y. Sheu, 'Characterization of the Colloidal Asphaltenic Particles in Heavy Oil', *Fuel*, **74**, 1140–1145 (1995).
37. J.G. Reynolds, W.R. Biggs, S.A. Bezman, 'Reaction Sequence of Metallopetroporhyrins During Heavy Residuum Upgrading', *ACS Monograph 344, Metal Complexes in Fossil Fuels: Geochemistry, Characterization, and Processing*, eds. R.H. Filby, J.F. Branthaver, ACS, Washington, 1987.
38. R.H. Fish, J.G. Reynolds, E.J. Gallegos, 'Molecular Characterization of Nickel and Vanadium Non-porphyrin Compounds Found in Heavy Crude Petroleums and Bitumens', *ACS Monograph 344, Metal Complexes in Fossil Fuels: Geochemistry, Characterization, and Processing*, eds. R.H. Filby, J.F. Branthaver, ACS, Washington, 1987.
39. J.T. Miller, R.B. Miller, A.M.J. van der Eerden, D.C. Konigsberger, 'Structural Determination by XAFS Spectroscopy of Non-porphyrin Nickel and Vanadium in Mayan Residuum, Hydrocracked Residuum, and Toluene-insoluble Solid', *Energy Fuels*, **13**, 719–727 (1999).
40. J.T. Miller, R.B. Fisher, P. Thiyagajan, R.E. Winans, J.E. Hunt, 'Subfractionation and Characterization of Mayan Asphaltene', *Energy Fuels*, **12**, 1290 (1998).
41. R.H. Filby, J.G. VanBerkel, A.E. Bragg, A. Joubet, W.E. Robison, C.A. Grimes, *J. Radioanal. Nucl. Chem.*, **91**, 361–368 (1985).
42. G.N. Aleshin, F.I. Samedov, M.F. Mir-Babaev, V. Kam'-yanow, 'Trace Element Composition of High Molecular Weight Components of Petroleum and Petroleum Residues of Azerbaidzhan', *Neftehhimiya*, **30**, 175–183 (1990).
43. Y. Inoue, K. Gomyo, T. Hisamitsu, H. Ozaki, 'Hydrorefining of Shale Oils (Part 1). Type Analysis of Nitrogen in Shale Oils', *Sekiyu Gakkaishi*, **28**, 374–378 (1985).
44. P.G. Shaw, D. McKown, S.E. Manahan, 'Trace Elements Determinations in Shale Oil Products by Neutron Activation', *Anal. Chim. Acta*, **123**, 65–74 (1981).

Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels

Ernie Baughman
Orbital Sciences, Pomona, USA

1 Introduction 1
2 On-line versus Laboratory Instrumentation 1
3 Near-infrared Spectroscopy – What is it? Overtones? 1
3.1 Warning 2
3.2 Assessment of the Instrument 2
3.3 Examples of Problems with Fiber-optics and How to Address Them 3
3.4 Octane Numbers 5
3.5 Determination of Methyl Tertiary Butyl Ether 5
3.6 Olefins 5
4 Accuracy and Repeatability 5
5 Diesel 7
6 Conclusion 7
Abbreviations and Acronyms 7
Related Articles 7

1 INTRODUCTION

The goal of any manufacturing process is to make acceptable products in an economical fashion. To do this requires timely information about raw materials, material in process and product quality. It is the role of on-line analytical instrumentation in providing some of this information on crude oil, gasoline, and diesel fuel that will be addressed in this section.

2 ON-LINE VERSUS LABORATORY INSTRUMENTATION

Laboratory instrumentation sits in an environmentally controlled atmosphere and is operated by skilled people to provide information on samples collected from the process. The goal of both the on-line instrument and laboratory instruments sounds the same, that is, to provide information, but the conditions they must operate in are very different. The laboratory instrument has a skilled operator, whereas the on-line instrument must operate on its own. The laboratory instrument works on a sample that is extracted from the process line and normally cooled to room temperature and reduced to atmospheric pressure. The on-line instrument takes its own sample and works with it at line pressure and usually at line temperature. If any adjustments are required, the on-line system must make them as no operator is involved. This, of course, must be done by the sample system but the on-line instrument dictates what the sample system must do. One question that arises frequently is the need for temperature control of the sample. Near-infrared (NIR) spectroscopy is a tool which counts molecular bonds within the sample. As the temperature changes the number of molecules within a given path length changes and therefore the spectrum changes. In a system with very little hydrogen bonding, most of the changes can be accounted for with just the change in density. If hydrogen bonding exists, such as with water or alcohol, major changes in the spectra are noted with temperature differences. This need for temperature control can become a significant item for on-line instrumentation and is normally ignored in the laboratory because everything is at the same temperature, the temperature of the room.

3 NEAR-INFRARED SPECTROSCOPY – WHAT IS IT? OVERTONES?

Light can interact with matter in several ways:

1. Simple reflection – we see that every morning in the mirror.
2. Refraction – light as it changes mediums changes direction, for example, going from air into water (our eye glasses are shaped by this principle).
3. Ultraviolet light can be absorbed by certain molecules and has enough energy to move electrons into an excited state.
4. Infrared (IR) light can excite, stretching and bending the bonds that are dipolar.

In the NIR region of the spectrum, that band of wavelengths between visible light and the IR, or wavelengths from 700 nm to 3000 nm, light is able to excite overtones and combinations of some of the molecular bonds that are excited by the IR. The only significant absorbencies in the NIR are due to the hydrogen–X bonds where X is a heavier atom. Therefore, organic compounds where we have bonds between hydrogen and carbon, H–C, show bonds clearly in the NIR region. From the above, one might

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

think that all H–C bonds absorb at the same frequency and with the same intensity. However, this is not the case: just as in the IR range, there are many different bonds, depending on what else is attached to the carbon. For example, methyl bonds will differ from methylene bonds, which will be different again from aromatic hydrogen bonds.

NIR spectroscopy then makes use of the overtones and combination bands to analyze compounds in both quantitative and qualitative terms. Because of the complexity of the spectrum the qualitative aspect of NIR is played down but is used in judging raw materials for suitability in some manufacturing processes. By far the biggest use of NIR is in the quantitative area, not only for chemical make-up of the sample, but also for the physical properties of that sample such as boiling point.

How can the boiling point of a liquid be determined by shining light through it? From elementary physical chemistry we know that the boiling point of a mixture is determined by the mole fraction of each component in the mixture multiplied by the boiling point of each component. A mixture is not a static system. As it starts to boil, certain compounds evaporate first; this changes the composition so now the boiling point is higher. This process continues until all of the liquid has boiled. What could be done is to determine each component in the liquid, calculate its mole fraction and therefore its influence on the boiling point. By summing over all these components the boiling curve of the liquid could be determined. The NIR approach skips the step of identifying each component in the mixture but based on ''experience'' finds a pattern which will predict the boiling points of the liquid. This ''experience'' is called a model – a mathematical relationship between absorbencies at various wavelengths and the boiling points of the liquid. This ''modeling'' is one of the most powerful and frustrating aspects of NIR spectroscopy. It is frustrating in the sense that although the human brain cannot process the NIR spectrum into the underlying components, a computer, ''trained'' using various statistical methods, can. The methods start with a ''training set''. The spectra are collected for a set of samples which are similar to the samples expected to be used by the NIR instrument. Correlations are found between the wavelengths and associated absorbencies with the properties of interest. If these correlations involve very few wavelengths then we can use a method called multiple linear regression (MLR) to develop a ''model'' between the spectrum and the property of interest. More commonly many wavelengths are involved and we use methods such as partial least squares (PLS) or principal component regression (PCR). Both methods can use the entire spectrum and the difference is that PLS uses correlations between the property of interest and the spectrum to find the best fit and PCR uses the biggest spectral differences and tries to find correlations. (I have used PLS because that has worked the best for me where I have tested various methods.) Whatever mathematical tool is used, we call the resulting equation between the spectrum and the property of interest a model. One can apply various models to the same spectrum to determine various properties. In one refinery we are using 14 models to ''predict'' or measure the properties of the gasoline flowing through the NIR.

3.1 Warning

The above tools are extremely powerful and correlations will be found for any property of interest regardless of spectral information. Models should always be tested on a new set of samples, called the verification set, before the model is used. Since this new set of materials has not previously been analyzed by the spectrometer computer system, the results should indicate what will happen with future production samples. Note that the system can only measure samples within the calibration range of the instrument. If samples are outside the calibration range a new ''model'' must be constructed to incorporate the new sample types. Sometimes this involves a totally new model and sometimes the old model can be modified to include the new information. If the process has permanently changed, I recommend building a new model. If the old model did not cover the total range – then extend the model. I have seen different models used for winter and summer grade gasoline, for example, as well as a different model for each grade of gasoline.

3.2 Assessment of the Instrument

Why would one use NIR with the computer algorithms when IR is available? The biggest arguments for the use of NIR for liquid samples are: that sources are more stable; the detectors are easier to work with and more stable; windows are more sturdy (glass vs. salt plates); the path lengths for liquids are much more reasonable (mm and cm in the NIR compared with 0.1 mm or less in the IR); and fiberoptics are readily available in the NIR, at lower cost compared to IR fibers. The IR has a major advantage in the laboratory setting in ease of interpretation of the data; however, in the process environment, where very similar samples are seen minute after minute, for many days, it is the quantitative ability of the NIR that wins over the qualitative advantages of the IR region.

The statement that fiberoptics are readily available in the NIR region needs comment. There are many vendors who supply fibers for the communication industry that operate in the NIR region. In the communication industry the signal is either off or on; by contrast, for the analyzer industry we need to quantitate the size of the signal. In the

communization industry, certain wavelengths are chosen which fit well with the existing fibers, but in the analyzer industry we are using the wavelength range. The point is that fibers designed for the communication industry may not, or more probably will not, work for the analyzer.

3.3 Examples of Problems with Fiberoptics and How to Address Them

Figure 1 is a superimposition of two different pictures. In each case we have 10 m of fiber connected to a power meter. The difference is that at the bottom the fiberoptic coil is covered with black felt. The readings on the power meter tell the story: on the top where light is hitting the fiber, the reading is 389.7 and on the bottom with the fiber covered, the reading is 8.8. What is happening? Light is going through the fiber jacket and coupling into the fiber and that is what the power meter is reading. Once the problem is defined, there is an easy solution: to make the jacket out of material that does not transmit light in this region of the spectrum.

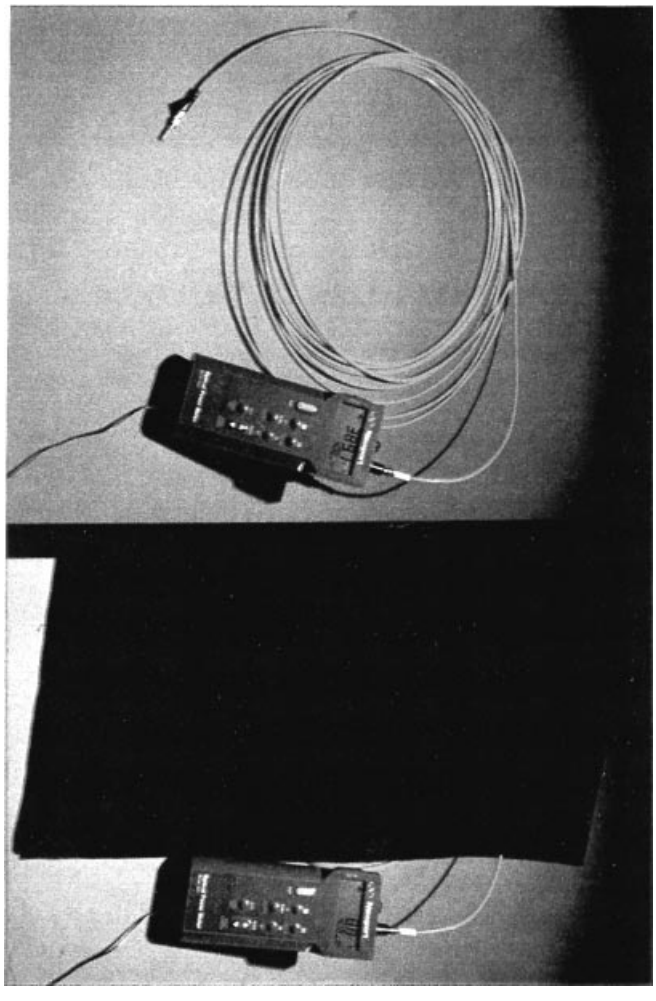

Figure 1 Comparison of light transmission in coated and uncoated fiberoptic coils.

The following is taken from a 1992 Eastern Analytical Symposium talk by G.H. Vickers of Amoco.

Figure 2 shows a standard drawing of an optical fiber. The assumption is that the light turns at the interface between the core and cladding and does not penetrate into the cladding. Figure 3 shows the spectrum of 10 m of plastic-clad silica fiber. The large absorbances are from the plastic cladding CH bands. This fiber is not usable in the NIR region. Figure 4 shows the spectra of a silica-clad, low OH, silica fiber with a Tefzel® jacket as a function of temperature. Note the large changes around 1400 nm. This is due to the change in the hydrogen bonding around the few remaining OH groups in the silica fiber. Also note the bands around 1250 nm and others of similar size. They are due to the CH bonds in the Tefzel® jacket. The cladding jacket is acting like a secondary fiber and becomes a plastic-clad silica fiber. How can this be avoided? Use a jacket material of high index of refraction to extract the cladding modes. The success of this is shown in Figure 5, where we have a silica fiber with a

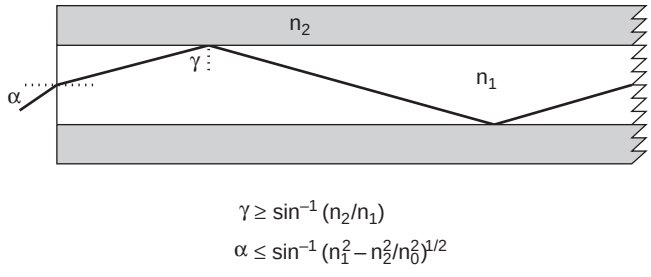

Figure 2 Total internal reflection in an optical fiber.

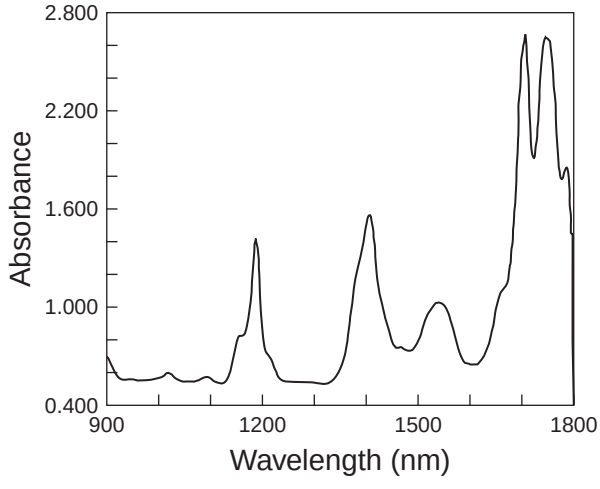

Figure 3 Absorbance spectrum for a plastic-clad silica fiber.

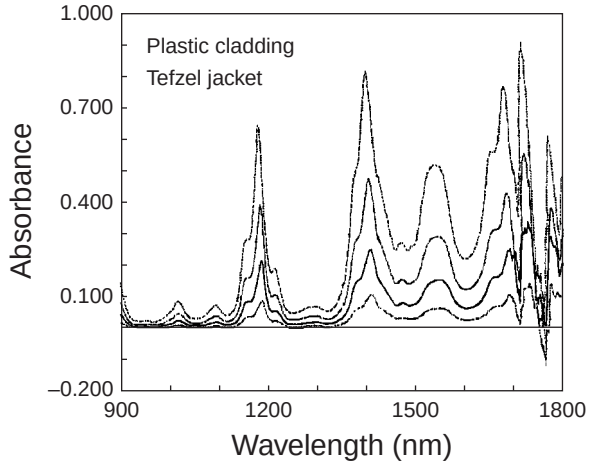

Figure 4 Absorbance spectra of a silica-clad silica fiber with Tefzel® jacket at temperatures from 5 °F to 125 °F.

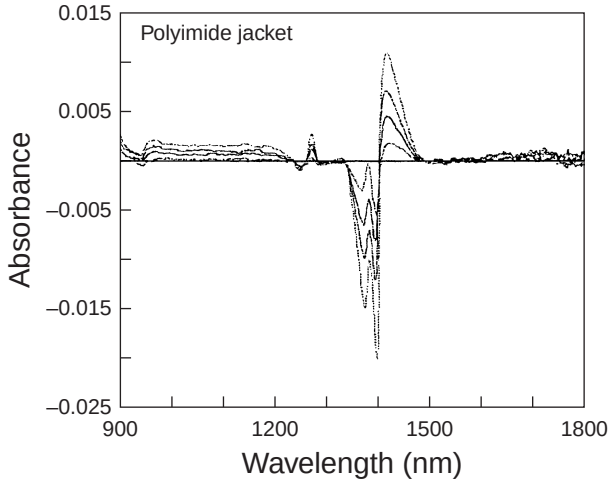

Figure 5 Absorbance spectra of a silica fiber clad with polyimide at temperatures from 5 °F to 125 °F.

silica cladding jacket with polyimide. Note the scale: other than the 1400 water peak, the spectra show changes of less than 0.005 AU. To eliminate this remaining difference we coil the fiber inside the jacket to allow for differences in thermal expansion of the fiber and the jacket as the temperature changes. In other words for fiberoptics to be usable in NIR requires some thought and understanding of the physics of the fiber. Unfortunately, all of the things mentioned to make the fiber useful add to the cost of the fiber.

It was mentioned previously that the optical path length through the cell in the NIR has a range from millimeters to centimeters. The NIR spectral region can be further subdivided according to the C–H vibration overtone. The higher the overtone the shorter the wavelength and the longer the path length for an equivalent absorbance. In general we can use a path length about a factor of 10 longer for each overtone. Therefore, in the first overtone region the path length is about 1 mm, in the second about 1 cm, in the third about 10 cm. With the 10-cm path the sample needs to be free of scattering particles or that effect will dominate the spectrum. There is another problem with the third overtone which must be discussed: interference from the ultraviolet and visible spectrum.

Certainly, most crude oil is black, but not all. The black color is due to the asphaltenes that are present in most crude oils. The peak absorbence of asphaltene is down in the ultraviolet part of the spectrum but the tail of the peak is so large that it extends through the visible and the third overtone of the NIR. However, if one examines the second overtone region, information on the olefin, aromatic and saturate content of the crude oil is available. Some work has been done trying to measure the sulfur and metals content of the crude oil but the results are questionable.

In the mid-1980s, Dr Jim Callis (University of Washington, as part of a project for the Center of Process Analytical Chemistry, CPAC) showed that octane numbers of gasoline (see section 3.4) could be determined from the NIR spectrum. The oil companies that were part of CPAC were skeptical of Dr Callis' results but realized the enormous impact they could have if proved to be real. Several companies undertook their own projects to test Dr Callis' work. The private work confirmed his original data and showed that the octane numbers could be obtained in all three overtones. However, additional effort also showed the sensitivity of the results to wavelength long-term reproducibility. By that I mean, if a model was developed and the spectrometer's wavelength axis drifted, the model's results were meaningless. In order to avoid the inherited drift in the moving grating spectrometers, two avenues were pursued. BP and Bomen worked together to develop the TOP-NIR which is a Fourier transform (FT)–NIR instrument. EXXON and Analect Instruments also developed an FT–NIR approach. Amoco and Perkin-Elmer developed a third overtone diode array system, the PIONIR 1024. Both the FT and the diode array approaches avoid the variation caused by the moving grating but introduce other problems which must be accounted for. With the diode array concept, dimensional stability is vital – hence the spectrograph temperature is controlled within 0.5 °C as the outside temperature varies from −40 °C to 50 °C on the PIONIR 1024. With the FT approach, mirror movement is referenced to a laser wavelength to ensure stability. All three systems have proprietary systems built in to give the required stability on the wavelength and absorbance axis to make the octane measurements.

A report in *Oil & Gas Process International*, 1997, by Kjell Arne Ulrund of Statoil, Norway, reports that

in addition to the octane numbers, they are measuring the MTBE (methyl tertiary butyl ether) concentration, olefins, benzene, initial boiling point, final boiling point and distillation evaporated at 70, 100 and 180 °C, every minute, using the PIONIR 1024. This system allows the operators, either human or computer, to optimize the gasoline blends using materials available in the refinery. The speed of the data allows on-line blending into ships for transport of the fuel.

3.4 Octane Numbers

There are three octane numbers: the Research Octane Number (RON), the Motor Octane Number (MON), and the Pump Octane Number (PON), which is the average of RON and MON. The traditional way to determine the RON and MON is with an engine and compare when the test fuel ''knocks'' in the engine compared to the standard. The standard is a blend, for MON, of iso-octane (2, 2, 4-trimethylpentane, octane number 100) and *n*-heptane (octane number 0). The MON is determined by finding a composition of iso-octane and *n*-heptane that knocks the same as the test fuel under the conditions of the test. It reflects the normal engine requirements at idle conditions. The RON is again determined by an engine test but the conditions simulate an engine under load. More severe conditions require a higher standard and that is toluene, octane number 113. The RON then finds the ratio of *n*-heptane, iso-octane and toluene that has the same knock properties as the test fuel under these new conditions.

Both the RON and MON tests have significant reproducibility problems and require highly trained operators to obtain consistent results. (I have been told that it takes about five years to train someone to correctly run the knock engines.)

Since the octane properties are related to the chemical composition of the fuel, maybe it is not too surprising that the octane number could be determined by the NIR. However, when one considers that a typical gasoline is a mixture of over 250 compounds it is amazing that a 5 to 10 factor PLS will allow determination of the octane number from the NIR spectrum.

3.5 Determination of Methyl Tertiary Butyl Ether

The standard way to measure MTBE in gasoline is with a long, 40-min, gas chromatography (GC) procedure. A 40-min analysis may be acceptable in a laboratory but consider a refinery that is blending 1 million gallons per hour (a single large gasoline refinery). What is that refinery to do? It must put in at least the minimum legal level of MTBE. If the GC analysis is not available until 40 min after analyzing, the refinery must hold that 1 million gallons of gasoline waiting for results or put in enough MTBE to make sure it has the legal minimum. With the NIR report results every minute, the ''excess'' MTBE can be reduced.

How can the NIR measure the ether concentration? Because of the electronegativity of the oxygen, the hydrogens on the carbon adjacent to the oxygen will absorb light at a slightly lower wavelength (higher energy). It is that shift which can be measured.

3.6 Olefins

Olefins are very reactive compounds and at very high concentrations are used to make polymers, e.g. polyethylene. In gasoline they are blamed for gum formation in the fuel system, and as components in the formation of smog if they escape from the system.

The standard method for determination is called fluorescence indication analysis (FIA) (do not confuse with flow injection analysis, also referred to as FIA). The FIA procedure involves flushing the sample down a silica gel column which has been impregnated with indicators; as the gasoline flows through the column, colored bands will show the paraffins, isoparaffins, aromatics, nathphenes, and olefins in the gasoline. Again, this is a very slow test that is nonrepeatable. Factors associated with the lack of repeatability are the column packing, flow through the column, sample size, and the indicator preparation. This is very different from shining light through the sample, as is done with the NIR method.

4 ACCURACY AND REPEATABILITY

Several times the standard tests have been referred to as nonrepeatable, yet these are the standards which are used to calibrate the NIR. How can the NIR do better? There are two different issues to consider: repeatability and accuracy.

The NIR is extremely repeatable in determining the absorbance vs. wavelength, the spectrum values, on an acceptable unit. The model which coverts such data to information, such as octane number, is just a multiplier, a set of numbers. Hence it is totally repeatable. Therefore, the NIR system will give the same answers on the same sample regardless of how many times it is run.

The accuracy of the NIR opens up a totally new argument. One side of the argument says that the NIR cannot be more accurate than the standard, whereas the other side says that if we have enough samples in our calibration set the *random* errors of the standard method will be averaged out and therefore the NIR is more accurate (closer to the truth) than the standard model. Unfortunately, while I side with the second argument,

Table 1 Expected PIONIR performance for diesel/gasoline

Property	Units	Typical range	ASTM method	Repeatability		ASTM reproducibility	SEP		Notes
				ASTM	PIONIR		ASTM	PIONIR	
Cetane number	CN	40 to 56	D613	0.7	0.11	2.9	1.0	1.2	ASTM values at 48 CN
Cetane index (2-term)	CN	40 to 56	D976	N/A[a]	0.12	N/A		0.8	Correlated ±2 CN to engine rating, per method
Cetane index (4-term)	CN	40 to 56	D4737		N.M.[b]			0.8	ASTM values depend on D86 and D1298
Aromatics	vol%	10 to 50	D1319	1.5	0.2	3.3	1.2	2.0	ASTM values at 35% aromatics
Saturates	vol%	48 to 88	D1319	1.7	0.2	5.3	1.9	2.0	ASTM values at 65% saturates
Kinematic viscosity	cSt	1.480 to 3.530	D445	0.1365	0.0240	0.4200	0.1516	0.14	ASTM values at 2.5 cSt
Specific gravity	Gms/me	0.8090 to 0.8710	D1298	0.0005	0.0003	0.0012	0.0004	0.0016	
T10	°F	364 to 478	D86	5.4	1.3	12.6	4.5	9.5	ASTM values for automatic method, slope 2.7
T50	°F	418 to 561	D86	5.4	1.2	15.3	5.5	5.6	ASTM values for automatic method, slope 2.7
T90	°F	476 to 644	D86	5.4	2.3	12.6	4.5	10.3	ASTM values for automatic method, slope 3.6
R250	%	8.5 to 29.5	D86	1	N.M.	3	1.1	1.5	ASTM values for slope 1.5
R350	%	87.0 to 98.0	D86	1	N.M.	2	0.7	1.6	ASTM values for slope 2
Flash point	°C	56 to 86	D93	5	1.0	10	3.6	3.6	ASTM value for automatic method
Cloud point	°C	−8 to +4	D2500	2	0.4	4	1	2	
Pour point	°C	−35 to +10	D97	3	0.4	6	2	5.4	

On-line customer results can typically be more accurate than the general results reported here when calibration models are developed for the customer's specific type of fuel. Calibration model results will depend on the laboratory analysis of the samples using the primary method. Note that the ASTM (American Standard Testing Method) SEP (standard error of perdition, measured on samples not in modeling set) is defined as the reproducibility reported in the method divided by 2.77 and that the PIONIR (Orbital's third overtone NIR instrument) repeatability data are based on 120 repeated scans of a static sample over 48 hours.

[a] Not applicable.

[b] Not measured.

the ''jury is out'' on this entire issue. The problem is that even if the NIR is more accurate on a given sample than the standard method, there is no way to prove it with *one* laboratory test.

5 DIESEL

Another major product of most refineries is diesel fuel. Although there are no literature references to the NIR applications for diesel, the system can be used to determine:

1. Cetane number
2. Cetane index
3. Aromatic concentration
4. Saturate concentration
5. Viscosity
6. Specific gravity
7. T10 (temperature at which 10% of the fuel has been distilled off)
8. T50
9. T90
10. R250 (amount distilled off at 250 °C–note the use of °C vs. °F in T-numbers)
11. R350
12. Flash point (temperature in °C at which fuel supports a flame)
13. Cloud point (temperature at which fuel becomes cloudy, in °C)
14. Pour point (temperature at which fuel will not pour, in °C).

The cetane numbers are measures of fuel quality for a diesel engine like octane numbers are for the gasoline engine. Because of the extreme difference in ignition methods of the two engines, the octane numbers and cetane numbers are almost anti-correlated – straight chain molecules have a low octane number and high cetane number.

The cetane index measures how the fuel will respond to certain additives, cetane improvers – vital information so the refiner can put in the correct amount of additive. The cloud point and pour point are important values but difficult to obtain in the laboratory because of the time and operator attention required to get repeatable results.

The results using one type of NIR instrument available from Orbital Sciences are shown in Table 1.

6 CONCLUSION

What the above has shown is that the NIR spectrum contains valuable information to optimize the refinery operation. One large refinery gave the NIR credit for adding 40 million dollars to the ''bottom line'' in the first year of operation. This saving resulted from fast, precise, and, this refinery believed, accurate results of the NIR.

ABBREVIATIONS AND ACRONYMS

FIA	Fluorescence Indication Analysis
FT	Fourier Transform
GC	Gas Chromatography
IR	Infrared
MLR	Multiple Linear Regression
MON	Motor Octane Number
MTBE	Methyl Tertiary Butyl Ether
NIR	Near-infrared
PCR	Principal Component Regression
PLS	Partial Least Squares
PON	Pump Octane Number
RON	Research Octane Number

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Food **(Volume 5)**
Near-infrared Spectroscopy in Food Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Fuel Performance Specifications, Mid-infrared Analysis of • Fuels Analysis, Regulatory Specifications for • Full Range Crudes, Analytical Methodology of • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Polymers and Rubbers **(Volume 9)**
Near-infrared Spectroscopy of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction • Chemometric Methods in Process Analysis • Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Signal Processing in Analytical Chemistry

Nuclear Magnetic Resonance Characterization of Petroleum

Francis P. Miknis and Daniel A. Netzel
Western Research Institute, Laramie, USA

1 Introduction **1**
1.1 Scope 1
1.2 Historical Development 2
2 Solid-state Nuclear Magnetic Resonance Characterization of Petroleum Source Rocks **2**
2.1 Experimental Procedures and Instrumentation 2
2.2 Nuclear Magnetic Resonance Measurements of Solids 2
2.3 Applications 4
3 Liquid-state Nuclear Magnetic Resonance Characterization of Petroleum Distillate Fractions **7**
3.1 Experimental Procedures and Instrumentation 7
3.2 One-dimensional Nuclear Magnetic Resonance 8
3.3 Two-dimensional Nuclear Magnetic Resonance 11
3.4 Average Molecular Structural Parameters 13
3.5 Applications 14
4 Some Applications of Nuclear Magnetic Resonance to Petroleum Processing **17**
4.1 Average Molecular Structures Using Solid-state Nuclear Magnetic Resonance 17
4.2 Predictive Property Correlations 18
4.3 Determination of Hydrogen Consumption During Processing 20
4.4 Nuclear Magnetic Resonance Studies of Coking 20
4.5 Process Kinetics 21
Abbreviations and Acronyms **22**
Related Articles **22**
Further Reading **22**
Books on Nuclear Magnetic Resonance 22
Journals 23
References **23**

Nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical method for determining the structure of an organic material. Developments in NMR have allowed high-resolution spectra of materials in the solid state to be obtained. Thus, materials that have been intractable, such as kerogens and sedimentary organic matter, can now be characterized for their organic carbon structure. The combination of solid- and liquid-state NMR makes NMR one of the most useful methods for the characterization of petroleum and its products.

Petroleum is formed from the maturation of sedimentary organic matter as a result of burial over geological timescales. Everyone is familiar with the products of petroleum, i.e. gasoline, diesel fuels, chemicals, and so on. All these materials, including the petroleum source rock material, can now be studied by NMR, from the genesis of petroleum to finished products such as transportation fuels. In this article, some applications of solid- and liquid-state NMR to the characterization of petroleum source rocks and petroleum end products are described.

1 INTRODUCTION

1.1 Scope

In an encyclopedia article such as this, which deals with NMR studies of petroleum and related materials, it is not practical to cite and discuss all the applications of NMR to the analysis of petroleum. This would require a great deal more space and the article would be more of a textbook or review-type article than an encyclopedic article. Consequently, the applications are necessarily limited, as are the citations for a given topic. It is hoped to provide sufficient information here for the reader to obtain an appreciation of the uses of NMR in petroleum analyses, or to give sufficient citations to direct the reader to more detailed information about a particular application.

The applications range from the maturation of petroleum source rocks to the end-use of petroleum as distillate products, and include solid- and liquid-state NMR applications. For a more detailed discussion of NMR techniques, the reader is referred to the *Related Articles* section at the end of this article for additional reading on **Solid-state Nuclear Magnetic Resonance** and liquid-state NMR (**Carbon-13 Nuclear Magnetic Resonance Spectroscopy**). Only a rudimentary discussion is included here. The objectives are to introduce the analytical application of NMR to the researcher who wishes to acquire an understanding of some of the methods and terminology that are most often encountered in the literature.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1.2 Historical Development

The possible applications of NMR to the characterization of petroleum and petroleum products were first discussed in 1955 and the first reports of the use of ^{1}H NMR were published in 1957 as discussed in Retcofsky et al.[1] Quantitation of the elemental hydrogen content of petroleum materials using ^{1}H NMR was reported in 1959[2] and the first quantitative ^{13}C NMR spectrum of a petroleum fraction by a time-averaged slow passage technique was reported in 1967.[3] It was not until 1976, shortly after the development of Fourier transform NMR, that quantitative ^{13}C investigation of crude oils and petroleum products started to become routine.[4]

The development of solid-state NMR techniques of cross-polarization (CP) and magic angle spinning (MAS) with high-power decoupling in the mid 1970s provided a great advance in characterizing and understanding the chemical nature of the organic matter in sedimentary rocks. These techniques provide information about the carbon structure of the organic matter in the rock without the need for preparation of kerogen concentrates. Some of the first published CP and CP/MAS ^{13}C NMR spectra were of coals[5,6] and oil shales,[7,8] materials that are largely insoluble and for which few good techniques then existed for probing the carbon structure of these fossil fuels.

2 SOLID-STATE NUCLEAR MAGNETIC RESONANCE CHARACTERIZATION OF PETROLEUM SOURCE ROCKS

2.1 Experimental Procedures and Instrumentation

Sample preparation in solids NMR is relatively straightforward. The sample must be ground to a sufficiently small particle size, so that the NMR spinners can be packed uniformly in order to spin the samples at high speeds without difficulty. A 100-mesh sample size is ideal because this particle size is acceptable for elemental analyses (C, H, O, N, S) as well. Larger particle sizes can be used. The criteria are that the spinner can be balanced to spin easily and that the sample placed in the spinner is representative of the material under investigation. The sample should be free of magnetic materials. In situations where there is not enough sample material to fill a spinner, or the particle size leads to an unbalanced spinner, other materials such as sulfur, or salt (NaCl) can be added to provide ballast to the spinner. Salt works well because the sample material can be recovered easily by washing. This is especially useful for kerogen samples, which are usually prepared in small quantities.

A field of 2.3 Tesla (100 MHz for ^{1}H, 25 MHz for ^{13}C) is about optimum for solid-state NMR measurements on solid fossil-fuel materials including source rocks. At this field strength, spinning rates up to 5 kHz can easily be obtained for spinners of 7.5-mm diameter. In addition, large-volume spinners with diameters of about 14 mm can also be spun at rates of about 4 kHz.[9] These rates are sufficient to minimize interferences caused by spinning sidebands associated with the chemical shift anisotropy. The increased sample size of a large-volume spinner has the advantage of increasing the signal-to-noise ratio (S/N), all other parameters being equal. For example, a typical 7.5-mm sample spinner has a volume of ca. 0.5 mL. A large-volume spinner has a volume of ca. 2.25 mL. Obviously, the greater the sample size, the shorter the time required to obtain a spectrum with a given S/N.

A variety of compounds have been mixed with oil shale and coals as internal standards to determine the percentage of carbons observed in CP and single-pulse NMR measurements.[10–14] In some cases, the spinner material, Delrin®, has been used as an internal standard.[9] An excellent reference material is *tetrakis*(trimethylsilyl) silane (TKS).[11,12] Furthermore, treatment with samarium(II) iodide has been shown to selectively reduce the free radical content, which in turn leads to an increase in the percentage of observable carbon.[12] For example, treatment of Wyodak sub-bituminous coal with samarium(II) iodide increased the percentage of observable carbons by 27%, leading to an increase in the carbon aromaticity of ca. 10% from 0.66 to 0.73. This technique should work on petroleum source rock samples, although it has yet to be tried on these types of sample.

The issue of quantitation in CP has been recognized since the beginning of its application to fossil fuels.[15–21] The main concerns are whether all carbons are observed equally in the CP experiment, and whether the fraction of carbons observed are a faithful enough representation so that the small loss of signal from carbons that are not observed can be ignored. Other concerns relate to such things as sample heterogeneity, spinning sidebands, unpaired electrons, MAS, and the pulse repetition rate. The problems associated with quantitation of ^{13}C NMR measurements in carbonaceous solids have been reviewed.[21]

2.2 Nuclear Magnetic Resonance Measurements of Solids

2.2.1 Cross-polarization with Magic Angle Spinning

CP NMR relies on the presence of an abundant spin system (^{1}H) to enhance the observation of a signal from a dilute spin system (^{13}C). The idea is to transfer polarization (and hence signal intensity) from the abundant ^{1}H spins to the dilute ^{13}C spins.[22] CP is a technique that enhances the sensitivity by a factor of four over the conventional single-pulse method. The

CP experiment consists of four basic timed sequences of radiofrequency (RF) pulses. The four-part procedure consists of (1) polarization of the ^{1}H spin system, (2) spin-locking in the rotating frame, (3) establishment of $^{13}C-^{1}H$ contact, and (4) observation of the ^{13}C free-induction decay (FID). CP/MAS NMR is discussed in greater detail in **Solid-state Nuclear Magnetic Resonance** in this encyclopedia.

2.2.2 Dipolar Dephasing

The technique of dipolar dephasing (DD) is a variation of CP. In this case, the ^{1}H decoupler is switched off for a short time after the CP contact time. During this off-time, signals from different carbon types dephase at different rates, depending on the strength of the $^{1}H-^{13}C$ dipolar interactions. The characteristic time for this dephasing is T_2, the spin–spin relaxation time. Carbons that are directly attached to hydrogen (primary, secondary, and tertiary carbons) experience strong $^{1}H-^{13}C$ interactions and dephase more rapidly (~20 μs) than carbons without attached hydrogen (quaternary carbons) (50–100 μs). After about 50 μs the decoupler is switched on again and the observed signal is due primarily to quaternary carbons. However, because of their rapid rotation in the solid state, the signals from methyl groups (CH_3) do not completely dephase and are observed as part of the aliphatic carbon signal. These methyl carbons can be distinguished from the other quaternary aliphatic carbons on the basis of chemical shifts. Thus DD is a technique that discriminates between protonated and nonprotonated aliphatic and aromatic carbons. This information can be useful in deriving average molecular structures for insoluble solid materials such as kerogen and asphaltenes.

2.2.3 Single-pulse Excitation

The most fundamental and most quantitative NMR measurement is made using single-pulse excitation (SPE), sometimes referred to as Bloch decay. For ^{13}C NMR measurements in solids, SPE measurements are made in conjunction with high-power decoupling to minimize broadening caused by the $^{1}H-^{13}C$ dipole–dipole interaction and with MAS to eliminate broadening caused by the chemical shift anisotropy. SPE can be employed as an independent check of the quantitative reliability of CP/MAS aromaticity measurements.[23,24] The signals from SPE arise from relaxation mechanisms that are different from those in the CP experiment. Therefore, agreement between carbon aromaticities from the two measurements indicates that both measurements are observing the same types and amounts of carbon. However, in solids, ^{13}C spin lattice relaxation times can be on the order of minutes, necessitating long acquisition times to record an NMR spectrum. For example, a typical SPE/DD experiment employing 10 delays, 1000 scans, and a 20-s pulse delay takes 56 h.[24] Thus, the advantage of having quantitative data is partially offset by longer signal accumulation times to obtain a spectrum with a reasonably good S/N. The use of a large-volume sample spinner can negate much of this disadvantage because less time is required to obtain a good S/N.[9] Although NMR spectroscopists realize that there has been an improvement in quantitation

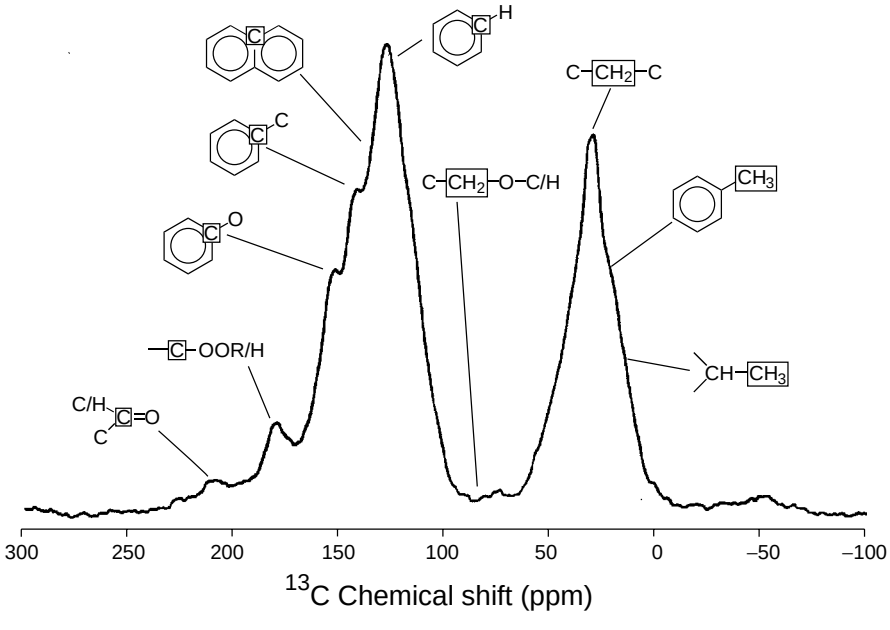

Figure 1 CP/MAS ^{13}C NMR spectrum of sub-bituminous coals, illustrating resolution typical of solid-state NMR.

using SPE methods, CP/MAS still is used more frequently because of the considerable saving in time.

2.3 Applications

2.3.1 Typical Nuclear Magnetic Resonance Spectrum of Carbonaceous Solid

A CP/MAS ^{13}C NMR spectrum of a carbonaceous solid is shown in Figure 1. The spectrum is that of a sub-bituminous coal. It illustrates the resolution typically obtained by CP/MAS NMR for these types of material, provided they have a sufficiently high percentage of total organic carbon (TOC) to give a good S/N in a reasonable amount of time (e.g. 1 h). The positions of different carbon types are shown on the spectrum. A listing of chemical shifts for the most common carbon functional groups is given in Table 1.[25] There is little resolution in the aliphatic carbon region (0–60 ppm), except for shoulders on the methylene band caused by methyl groups on aromatic rings and terminal methyl groups on long-chain paraffins. The aromatic carbon region shows more resolution from alkyl substituents, phenolic carbons, and carbonyl functionality. The main structural parameter obtained by CP/MAS is a measurement of the carbon aromaticity, i.e. the fraction of total carbon that is in aromatic (sp^2) carbon structures. Contributions from carbonyl carbons are often included in aromaticity measurements if the integrated signal intensity for the carbonyl carbons (165–210 ppm) is small. The spectrum was obtained at a spinning rate of 4.5 kHz, so spinning sidebands at 300 and −50 ppm are negligible. A petroleum source rock having a low TOC (<5%) would give a noisier signal and the shoulders, although present on the aliphatic and aromatic bands, would be more difficult to observe.

2.3.2 Natural Maturation with Depth of Burial

Thermal evolution of source rocks during natural maturation as a result of burial, or during artificial maturation as in pyrolysis, changes many of the physical and chemical properties of the organic matter (kerogen) in sedimentary rocks.[26,27] A depiction of thermal degradation of a hypothetical kerogen molecule is shown in Figure 2. In this example, temperature increases from top to bottom. The example shows the types of chemical change that are expected to occur during hydrocarbon generation resulting from burial for periods of geological time, or from elevated temperatures over short periods of time in the laboratory. In either situation, the thermal alteration of kerogen involves hydrogen disproportionation reactions in which the kerogen loses hydrogen to form liquid hydrocarbons, wet gas and dry gas in succession. The remaining hydrogen-deficient kerogen condenses to form larger aromatic clusters which would eventually form graphite.

The main chemical change depicted in Figure 2 is the progressive decrease in H/C (atomic hydrogen to carbon) ratio (increase in the carbon aromaticity) of the kerogen with increasing maturation. This change is brought about by aromatization of aliphatic carbons in hydroaromatic structures (tetralin type) and by cyclization and dehydrogenation of alkyl substituents. In addition, the number of

Table 1 List of carbon functionality used in peak identification of a ^{13}C NMR spectrum

Functionality	Structure	Symbol	ppm
Carbonyl		C=O	210
Carboxyl	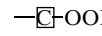	COOR	178
Phenol		ArCO	154
Branched aromatic		ArCC	130-140
Bridgehead aromatic		ArC^C	132
Protonated aromatic		ArCH	128
Oxy-methine		CHO	60-70
Oxy-methylene	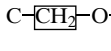	CH_2O	50-60
Oxy-methyl		CH_3O	40
Quaternary aliphatic		C	40
Methine		CH	39
Methylene (C-2)		CH_2(C-2)	23
Methylene		CH_2	30
Aliphatic methyl		CH_3(al)	16
Aromatic methyl		CH_3(arom)	20

Reproduced by permission of Elsevier Science from M.J. Trewhella, I.J.F. Poplet, A. Grint, *Fuel*, **65**, 541–546, copyright 1986.[25]

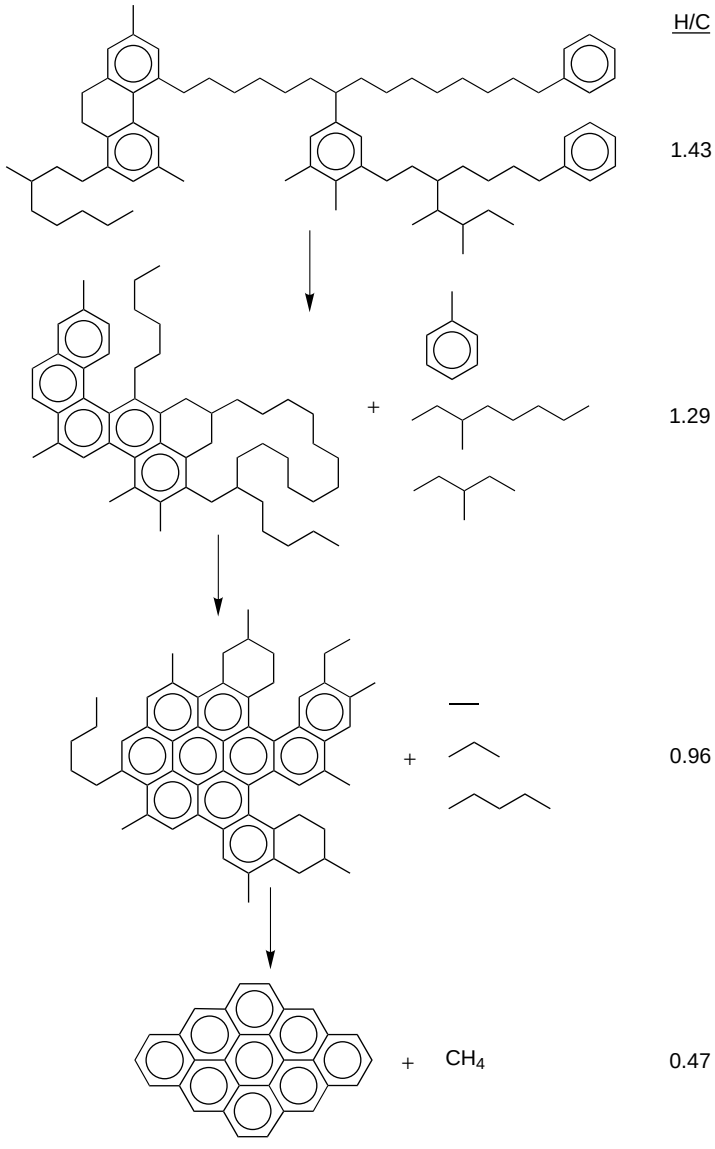

Figure 2 Schematic depicting thermal alteration of kerogen during burial. (Reproduced from *Petroleum Geochemistry and Geology*, by Hunt © 1979, by W.H. Freeman and Co. 1996; used with permission.[26])

side chains on the aromatic ring system decreases with maturation. The products of the reactions progressively become lower in molecular weight, with methane, CH_4, being the ultimate product. For simplicity, only hydrocarbon structures are shown; however, similar changes would be expected to occur with heteroatoms (N, S, O) in the ring system.

2.3.3 Changes in Nuclear Magnetic Resonance Spectra with Depth of Burial

NMR measurements provide important information on the catagenesis and maturation of the organic matter in sedimentary rocks. The chemical changes depicted in Figure 2 are observed as changes in the NMR spectra of sedimentary organic matter as a function of burial depth or temperature. Most NMR measurements show how the distribution of aliphatic and aromatic carbon changes with time and temperature during burial. CP/MAS ^{13}C NMR spectra for a suite of samples from the Wind River Basin in Wyoming representing different depths of burial are shown in Figure 3. The NMR spectra exhibit features that clearly support the types of chemical reaction in the scheme depicted in Figure 2. The spectra show an increase in the aromatic carbon fraction relative to the aliphatic carbon fraction with increasing depth (i.e. temperature). Aliphatic carbon structures are hydrogen rich and are largely responsible for the hydrocarbons that

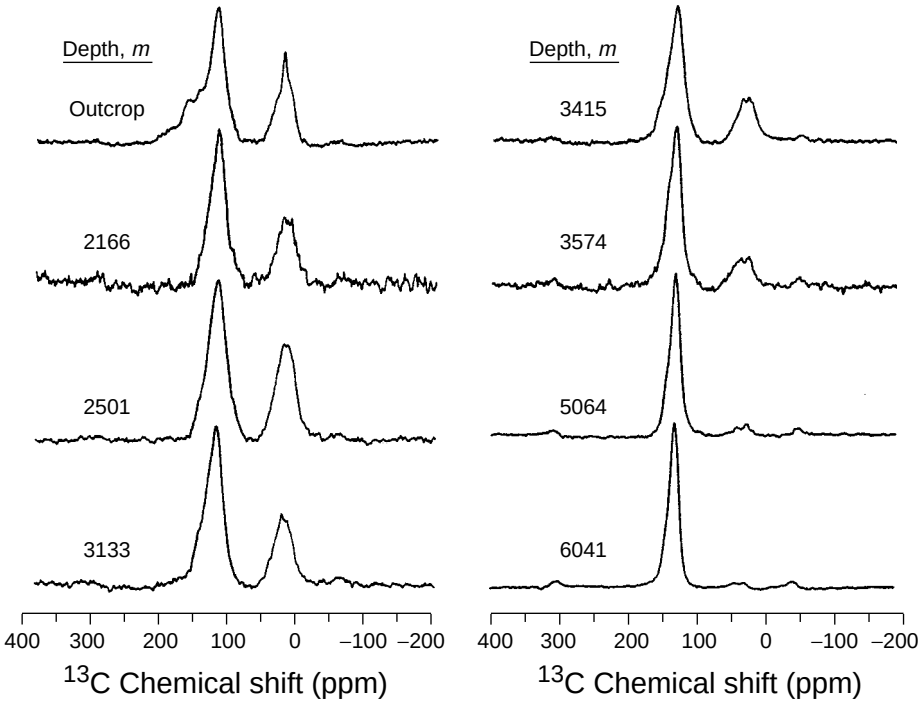

Figure 3 CP/MAS ^{13}C NMR spectra of samples from the Wind River Basin, Wyoming, taken from different burial depths.

are generated during natural and artificial maturation. The increase in carbon aromaticity with depth is due to loss of aliphatic carbons to hydrocarbon generation and to aromatization reactions. However, it is not possible to partition these losses between the two types of reaction without detailed material balances and elemental analyses. Nevertheless, by monitoring the aliphatic and aromatic carbons as a function of burial depth, limits can be placed on the capacity of source rocks to generate hydrocarbons at depth. Such information is difficult, if not impossible, to obtain by other methods. The NMR spectra in this example suggest that in the Wind River Basin, most of the volatile products (oil and gas) are generated down to a depth of about 3570 m. At greater depths there is not a sufficient fraction of aliphatic carbons remaining to generate hydrocarbons.

The aliphatic carbon region (0–90 ppm) shows partial resolution of at least two types of carbon (see Figure 1). These can be assigned to methylene carbons (CH_2) in long chains (30 ppm), and methyl groups (CH_3) attached to aromatic rings (ca. 20 ppm). Methylene carbons in branched and cyclic aliphatic structures are also counted with the methylene carbons at 30 ppm. The resolution is more noticeable in the spectrum of the sample at 3415 m. At greater burial depths, the spectra show a shift in position of the maximum of the aliphatic carbon resonance from 30 ppm for methylene carbons in long chains to that of the methyl groups attached to aromatic rings at ca. 20 ppm. Ultimately, these methyl groups would be the source of methane gas at very high maturity (Figure 2). Resonances at 300 and −50 ppm, most clearly seen in the 6041 m spectrum, are spinning sidebands.

There is a narrowing of the width of the aromatic resonance band (90–200 ppm) with increasing burial depth. This narrowing is indicative of an increase in the number of fused rings in the polynuclear aromatic moiety and of a lesser amount of substitution on the aromatic rings as alkyl side chains are cleaved during oil and gas generation. During maturation, cleavage of substituents bridging aromatic rings also initiate condensation reactions that yield larger aromatic cluster sizes. Aromatization of aliphatic carbons to produce larger more condensed aromatic ring structures would limit the number of substituents per aromatic carbon in the ring. The net effect of all these changes is to narrow the chemical shift dispersion, which results in a narrower aromatic resonance band. The general features of the NMR spectra as a function of burial depth in Figure 3 have been observed for maturation of petroleum source rocks[28] and coals during hydrous pyrolysis.[29]

The CP/MAS ^{13}C NMR spectra also show a resonance band at ca. 180 ppm for the outcrop sample. This band is attributed to carboxylate carbon functionality. During maturation, decarboxylation reactions occur liberating CO_2. The CP/MAS ^{13}C NMR spectra in Figure 3 show that this carbon functionality has disappeared by the time that a depth of 2166 m is reached, illustrating that evolution of CO_2 occurs in the early stages of maturation.

During artificial maturation (pyrolysis), rapid evolution of CO_2 is observed at low temperatures and is thought to be a major contributor to crosslinking and retrograde reactions.[30]

2.3.4 Nuclear Magnetic Resonance Evaluation of Source Rock Potential

The changes that are observed in the NMR spectra of source rocks as a function of burial depth are also observed in the laboratory when the sample is heated to a high temperature for short periods of time.[29,31] As a result, an NMR spectrum can be used to evaluate a rock for its potential to produce liquid hydrocarbons during burial. An example is shown in Figure 4 for the case of an oil shale from Colorado and a sub-bituminous coal from Wyoming. The conversion data were obtained from a Fischer assay of the materials (see **Oil Shale and Shale Oil Analysis**).

The most important factor controlling the generation of oil and gas is the hydrogen content of the organic matter. The Colorado oil shale is from a lacustrine depositional environment and has a highly aliphatic carbon structure; therefore, it has a high hydrogen content. It would be an excellent petroleum source rock, given the property relating to the geological conditions of burial depth and time. The Wyoming sub-bituminous coal is of a terrestrial origin. It has a more aromatic carbon content and therefore is hydrogen deficient relative to the Colorado oil shale. These characteristics are shown in the NMR spectra and they correlate well with the conversion data.

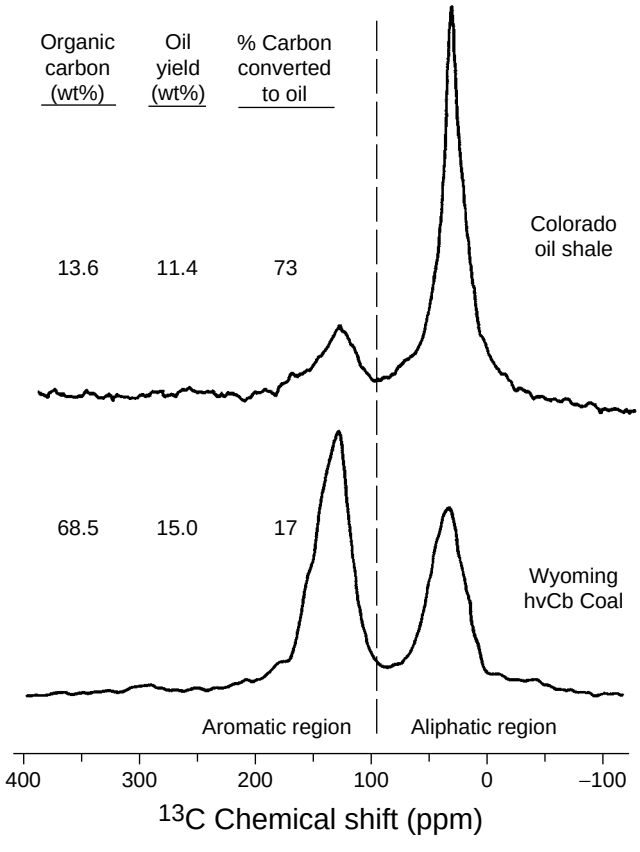

Figure 4 CP/MAS ^{13}C NMR spectra of a Colorado oil shale and a Wyoming sub-bituminous coal, illustrating the effect of carbon structure on conversion to oil.

3 LIQUID-STATE NUCLEAR MAGNETIC RESONANCE CHARACTERIZATION OF PETROLEUM DISTILLATE FRACTIONS

Liquid-state NMR is unsurpassed as a spectroscopic technique for characterizing and quantifying the molecular structure and dynamics of monomolecular materials. In the study of complex materials such as petroleum, the technique is most often used for characterization and quantification of the structure of the various petroleum distillate fractions, including residues. ^{1}H and ^{13}C NMR have been used to study naphtha and gasoline,[32–34] middle distillates,[35,36] jet and diesel fuels,[37–41] light gas oil,[42] heavy oils,[43–45] light crudes,[46] waxes,[47] and residue and heavy ends, including asphaltenes and asphalts.[48–53] Most of the studies involved the use of one-dimensional NMR spectral techniques. However, studies have included the use of two-dimensional (2D) NMR spectroscopic techniques, which facilitate the spectral interpretation of complex mixtures. Compared to gas chromatography and mass spectrometry, a major advantage of NMR spectroscopy is that the technique can also be applied to fractions that have boiling points greater than 550 °C.

3.1 Experimental Procedures and Instrumentation

Petroleum crudes are complex solutions of thousands of different paraffinic, aromatic and heteroaromatic molecules. Neither NMR nor other spectroscopic techniques directly provide enough information fully to delineate the individual components present in the whole crudes. Thus, as with other techniques, fractionation of the crude oil is desirable to reduce the complexity of molecular types in a sample and to obtain meaningful information about the molecular structure from an NMR spectrum. The first fractionation is usually distillation, which yields products of the petroleum crude such as gasoline, kerosene and vacuum gas oils. The distillate products are listed in Table 2. The light distillates can also be easily characterized and quantified by mass spectrometry, other spectroscopic techniques such as infrared, and chromatographic methods. The light as well as heavy distillates can be separated further into saturates, olefins and aromatic structural groups, using solvent extraction and various

Table 2 Common distillation fractions of petroleum

Fraction	Boiling (°C)
Light naphtha	−1–150
Heavy naphtha	150–205
Gasoline	−1–180
Kerosine	205–260
Fuel oil	205–290
Light gas oil	260–315
Heavy gas oil	315–425
Lubricating oil	>400
Vacuum gas oil	425–600
Residuum	>600

chromatographic techniques. Because individual compounds cannot be easily identified in the fractionated samples, recourse must be made to approximate (average) structural characterization methods using information obtained from NMR, elemental and average molecular weight (AMW) data.

Petroleum fractions for NMR studies are usually dissolved in $CDCl_3$ containing a small amount of tetramethylsilane (TMS) as an internal reference for both ^1H and ^{13}C chemical shift measurements. Because of the high sensitivity of present-day NMR spectrometers, only a few milligrams of sample are needed for a ^1H spectrum and approximately 100 mg for a ^{13}C spectrum. To obtain quantitative data from a ^{13}C NMR spectrum, it is often necessary to add a small amount of the relaxation reagent, chromium(III) acetylacetonate, to a $CDCl_3$ solution of a distillate fraction.

Almost all experiments are performed using a Fourier transform NMR spectrometer at frequencies ranging from 100 to 400 MHz for ^1H and from 25 to 100 MHz for ^{13}C. Many of the spectrometers manufactured today are multinuclear, almost fully automated and computer controlled. In addition, many can be programmed to conduct two- and three-dimensional spectral data acquisition.

3.2 One-dimensional Nuclear Magnetic Resonance

3.2.1 ^1H Nuclear Magnetic Resonance

The hydrogen (^1H) resonance signal is the most easily detected signal of any of the elements, and the area of the ^1H resonance quantitatively represents the number of hydrogens of a given type within a molecule. Two disadvantages of the ^1H NMR technique for investigation of complex mixtures are the spin–spin scalar coupling of adjacent hydrogens in a molecule and the small chemical shift range for the different hydrogen types (0–15 ppm). The scalar coupling results in splitting of a given hydrogen type into multiple peaks. The splitting pattern (multiplicity) is useful diagnostically for determining the structural groups adjacent to one another in molecules. However, in complex mixtures of molecules the splitting patterns of the various hydrogens overlap one another, resulting in the loss of resolution and information related to the coupling of adjacent hydrogen groups. This loss of resolution, together with chemical dispersion (many hydrogen types in the different compounds with similar but not equal chemical shifts), results in a broad envelope of resonances for the methyl, methylene, methine, olefinic and aromatic hydrogen types.

Figures 5 and 6 show the 400 MHz ^1H spectra of a regular grade gasoline and of an asphalt, respectively. The high resolution of the resonances peaks observed for the ^1H spectrum of the gasoline relative to asphalt is due to the relatively low molecular weight and lower number of compounds found in gasoline. The peak at 0 ppm is the chemical shift reference, TMS. The peak at 7.24 ppm results from a small residue of $CHCl_3$ in $CDCl_3$. The spectra shown in Figure 5 can be divided into three major hydrogen types: aliphatic, olefinic, and aromatic. The aliphatic hydrogens have chemical shift values ranging from 0.5 to 4.0 ppm. This region can be further divided into resonances associated with the methyl, methylene and methine functional groups. The chemical shift range from 4.0 to 6.0 ppm represents the resonances due to olefinic hydrogens. The resonance profile in this region can be used to determine the types of olefin present.[54] The aromatic hydrogens lie in the chemical shift range from 6.6 to 8.3 ppm, which can be divided further into hydrogens associated with mono-, di-, and tri-aromatic ring systems. The ^1H chemical shift ranges for the various

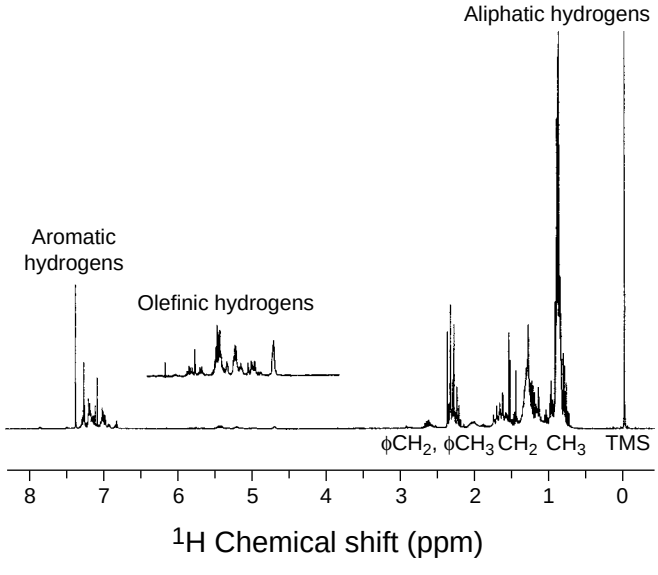

Figure 5 400 MHz ^1H NMR spectrum of a regular grade gasoline.

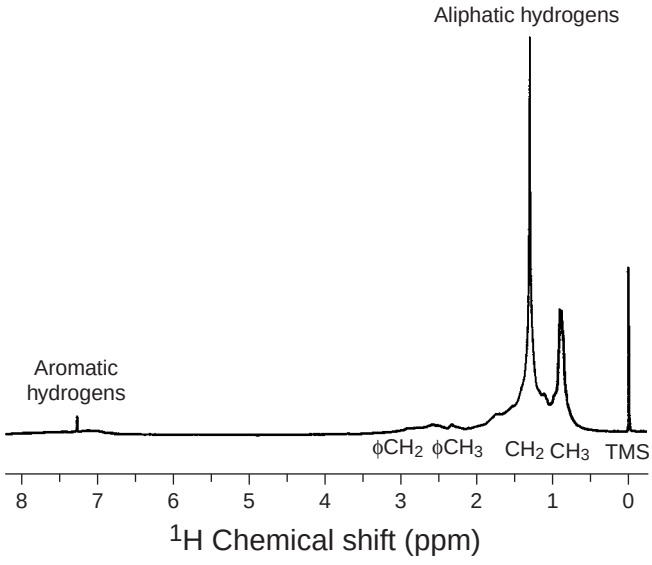

Figure 6 400 MHz ^{1}H NMR spectrum of an asphalt.

hydrogen types in hydrocarbons are given in Table 3. The interpretation of the ^{1}H spectrum is further complicated if there are significant concentrations of aliphatic groups adjacent to heteroatomic groups containing nitrogen, oxygen and sulfur. Thus, before attempting to assign the peak in a ^{1}H spectrum, some knowledge of elemental and functional group composition is desirable. Conclusions can be drawn from ^{1}H NMR data about the average carbon skeleton if one assumes a definite H/C ratio for the aliphatic constituents. However, the combined use of ^{13}C and ^{1}H NMR data is the preferred method.

Table 3 ^{1}H and ^{13}C NMR chemical shift ranges

Definitions	Chemical shift range (ppm from TMS)
^{1}H	
Monoaromatic hydrogens	6.6–7.2
Diaromatic hydrogens	7.2–7.8
Triaromatic hydrogens	7.8–8.3
Olefinic hydrogens	4.0–6.0
Methylene hydrogens alpha to aromatic ring	2.3–4.0
Methyl hydrogens alpha to aromatic ring	1.9–2.3
Naphthenic hydrogens beta to aromatic ring	1.6–1.9
β-CH_2– and β-CH_3 to aromatic ring and straight chain alkane methylene hydrogens	1.0–1.6
γ-CH_3 to aromatic ring and straight chain or branch alkane methyl hydrogens	0.5–1.0
^{13}C	
Total aromatic/olefinic carbons	100–150
Total aliphatic carbons	5–50
Methyl carbons of straight chain alkanes	13.8–14.2
C_2-Methylene carbons of straight chain alkanes	22.7–23.0
C_3-Methylene carbons of straight chain alkanes	31.8–32.0
$C_{4,5,n}$-Methylene carbons of straight chain alkanes	29–30

Several instrumental parameters must be established for reliable quantification of the NMR integration data. Typically, long relaxation delay times between 5 and 20 s, with a relatively small pulse angle (<40°), are used to acquire ^{1}H Fourier transform (FT) spectra. Good integration of the data can easily be obtained, but depends on having correct phasing and a flat baseline. Absolute determination of the hydrogen content by NMR is not often performed, but can be accomplished by adding an accurately weighed pure compound with a single ^{1}H resonance peak to the sample of interest.[55,56] The chemical shift of this reference compound must be in a region which does not interfere with the integration of the hydrogens associated with the sample.

Relative integration values from a ^{1}H spectrum are typically obtained for the different hydrogen types. Because the resonance signals for the total aromatic and aliphatic hydrogen types are widely separated, reasonably good accuracy and precision of the integration can be achieved. However, within the aromatic and aliphatic hydrogen regions, where only partial separation of hydrogen types are observed, the integration of the hydrogen types is less accurate and depends on the chemical shift range assigned to the particular hydrogen type. High field spectra for some samples can increase the resolution between hydrogen types and thus provide better integration values. 2D NMR spectra can also provide good resolution in some cases, but accurate integration can be more difficult.

3.2.2 ^{13}C Nuclear Magnetic Resonance

The natural abundance of the carbon-13 isotope is 1.1%. However, with the advent of Fourier transform NMR, this isotope of carbon is easy to detect and in most cases is the element of choice for investigation of fossil fuel materials. Unlike hydrogen, the scalar coupling between carbon and hydrogen nuclei can be decoupled electronically using a broad-band amplifier. Thus, a single resonance for each different carbon within a molecule is observed, resulting in greater resolution between carbon types. In addition, the chemical shift range (0–220 ppm) is much greater than for hydrogen. The large shift range reduces the extent of overlapping resonances and facilitates carbon-type identification. Unfortunately, chemical dispersion is also a major contributor to the broad resonances observed in carbon spectra of complex mixtures.

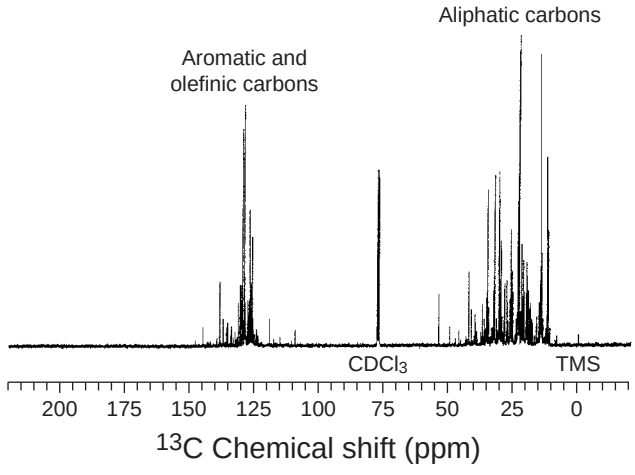

Figure 7 100 MHz ^{13}C NMR spectrum of a regular grade gasoline.

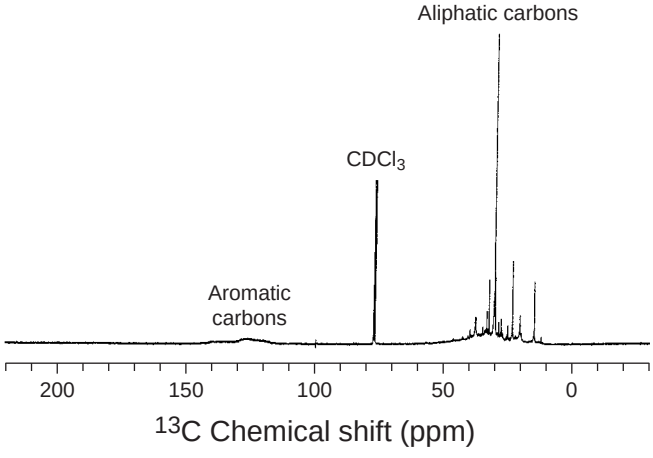

Figure 8 100 MHz ^{13}C spectrum of an asphalt.

^{13}C spectra of a regular grade gasoline and an asphalt are shown in Figures 7 and 8. Because there are fewer compounds, a large ^{13}C chemical shift range and broadband decoupling of the hydrogens, the resonance peaks, which each represent a single carbon type, are well resolved for the gasoline compared to those for the asphalt spectrum. The lower resolution and broader peaks in the asphalt are the result of chemical shift dispersion caused by the many compounds and shorter relaxation times for the carbon atoms. The ^{13}C spectrum of a hydrocarbon has two distinct chemical shift regions. The chemical shift for the aliphatic carbon-types ranges from 5 to 50 ppm and the aromatic carbons range from 100 to 150 ppm. Carbons adjacent to heteroatoms (nitrogen and oxygen) are often resolved in many light petroleum distillates. Phenolic and carbonyl carbons have chemical shifts ranging from 150 to 210 ppm, oxygenated aliphatic carbons have a range of 40 to 60 ppm. The three peaks in each spectrum centered at 78 ppm are due to the carbon atom coupled to the deuterium atom of the solvent $CDCl_3$. The olefinic carbons, if present, would have chemical shift values which overlap the aromatic carbon chemical shift range and are not easily identified in a one-dimensional NMR spectrum. In this case, the ^{1}H spectrum is preferred for identification and quantification of olefins. However, with advanced techniques such as 2D C–H correlation experiments, olefinic carbons can be identified and quantified (see section 3.3). Depending upon the petroleum distillate and subfractions of the distillates, the aromatic and aliphatic region can be further subdivided into different carbon types using various spectral editing and 2D NMR techniques (see sections 3.2.3, 3.3, and 3.7.2). These methods can be used to resolve tertiary and quaternary aromatic carbons and the methyl, methylene, methine, quaternary aliphatic carbons. In the asphalt spectrum (Figure 8) only the methyl (14 ppm) and methylene carbons (23, 30 and 32 ppm) of straight chain alkanes are readily identifiable. Table 3 lists the carbon-13 chemical shift for various carbon types in hydrocarbons.

The nuclear parameters which affect the quantitation of the carbon resonance signals are the ^{13}C spin–lattice relaxation time and the nuclear Overhauser enhancement effect (NOE). Carbon relaxation times are considerably longer than for hydrogens. In addition, aromatic quaternary carbons (bridgehead or from alkyl-substituent) in a molecule have much longer relaxation times relative to other carbons with attached hydrogens. To shorten and equalize the relaxation times among the various carbons, a relaxation reagent is often added to the sample. The most often used relaxation reagent is chromium(III) acetylacetonate at a concentration of about 0.05 M. Even with this reagent, instrumental parameters such as a small pulse angle of $<90^{\circ}$ and a pulse sequence delay of 5–10 s are used. The NOE on the various carbons can differ significantly but can be reduced using an inverse gated decoupling pulse sequence. This pulse sequence minimizes the NOE of all carbon signals, but because the signal enhancement by this effect is lost, time averaging processes must be increased to achieve a good S/N ratio. As with ^{1}H NMR, reliable integration of a carbon spectrum depends on correct phasing and a flat baseline.

3.2.3 Spectral Editing

NMR spectra of petroleum distillates are often recorded with the objective of elucidating the molecular structural characteristics of the sample in an effort to understand better its physicochemical properties and to monitor changes in process conditions. The usual approach involves subdividing the conventional ^{1}H and ^{13}C NMR spectra into chemical shift regions, each of which is

assumed to be associated with a single type of hydrogen or carbon substructure. If suitable experimental precautions are taken, integration yields the relative abundance of the different ^{1}H and ^{13}C aromatic and aliphatic structural types. However, chemical shift–structure relationships are limited in scope because of the ambiguities associated with determining the chemical shift range for many different hydrogen and carbon types in a fossil fuel. It is difficult to know where to subdivide spectra in such a way that structural inferences are both detailed and valid. Too large or too small a range will increase or decrease the relative abundance of the different structural types and increase the uncertainty of structural inferences. In addition, structural information such as the hydrogen connectivity and carbon multiplicity of the structural groups is lost when using conventional ^{1}H and ^{13}C experimental techniques for complex mixtures.

An alternative approach to quantify the NMR spectral data of any petroleum fraction into carbon types (CH_n, $n = 0–3$) is to use spectral editing techniques, i.e. separation of ^{13}C spectra into methyl (CH_3), methylene (CH_2), methine (CH), and nonprotonated or quaternary (C) carbons. The distortionless enhancement by polarization transfer (DEPT),[37,57–60] quaternary only carbon spectrum (QUAT),[61] gated spin-echo ^{13}C NMR (GASPE)[62,63] and the spin-echo broad-band off-resonance decoupling (SEBBORD)[64,65] are a few of the techniques that have been developed to regain the carbon multiplicity information and to define more precisely the chemical shift range for different carbon types in a complex molecule or mixture of molecules. These methods are based upon obtaining a series of spectra by modulating the signal intensities of the CH_n group as a function of either the ^{1}H pulse width, θ (DEPT) or the $^{13}C–^{1}H$ scalar spin-coupling characterized by a spin-coupling constant, J_{CH} (GASPE). A linear combination of these spectra results in edited spectra of the specific carbon types (DEPT CH_n, $n = 1–3$ and GASPE CH_n, $n = 0–3$). Figure 9 shows the conventional ^{13}C spectrum and the ^{13}C edited DEPT spectra for the various carbon types in the saturate fraction obtained from the 200–425 °C distillate cut from the Cerro Negro heavy petroleum. Also shown is the quaternary only (QUAT) ^{13}C spectrum. The spectra shown can be integrated to give the relative fractional amounts of the various carbon types. From a knowledge of the fractional amounts of carbon types, a number of average molecular structural parameters can be determined (see section 3.4).

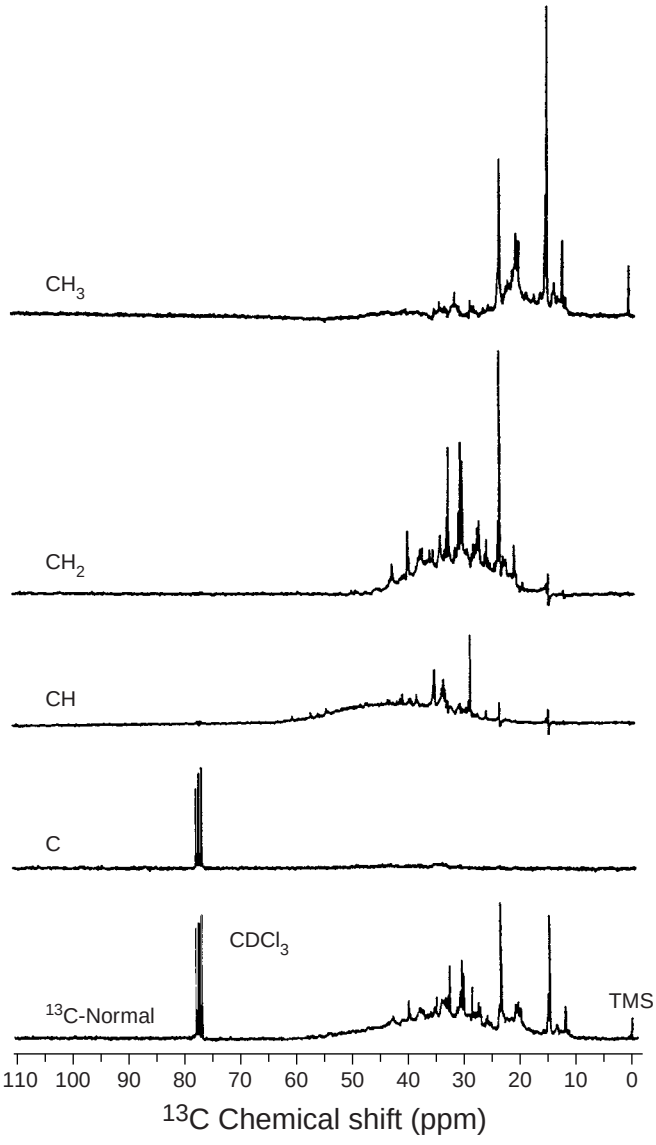

Figure 9 Normal ^{13}C NMR spectrum and the DEPT/QUAT spectral-edited carbon-type spectra for the 200–425 °C saturate fraction from the Cerro Negro heavy petroleum crude. (Reproduced with permission from D.A. Netzel, F.D. Guffey, *Energy Fuels*, **3**, 455–460 (1989).[37] Copyright (1989) American Chemical Society.)

3.3 Two-dimensional Nuclear Magnetic Resonance

In spectra which are inherently crowded, such as those of complex mixtures of different hydrocarbons, overlapping resonances of the various carbon types can often be separated and/or identified by introducing an additional dimension to the normal intensity versus chemical shift (frequency) representation of an NMR spectrum. The introduction of a second frequency dimension can be either a second chemical shift or a scalar-coupling parameter. The application of 2D NMR techniques to fossil fuels is relatively new, but their value is becoming increasingly apparent.[66,67] The application of 2D NMR to the complex mixtures of chemical types encountered in such fossil fuel liquids as shale oils, coal liquids, or petroleum have been reported.[68–72] The application of

2D methods to fossil fuels has been limited to oils and distillates, since heavier fractions, such as asphaltenes, usually do not give sufficient resolution even at high field.

In principle, two classes of 2D experiments exist: *J*-resolved (2D-*J*) and correlated 2D spectra. The first is characterized by one frequency axis containing the coupling (multiplicity) information and by another frequency axis containing the chemical shifts. In a second type, both frequency axes contain chemical shift information. The connection between the two frequency axes is through scalar coupling (homonuclear [^{1}H–^{1}H] as well as heteronuclear [^{1}H–^{13}C]) or through dipolar coupling.

Figure 10 shows a homonuclear (^{1}H–^{1}H) 2D-*J* spectrum[70] of an aromatic petroleum solvent (PETSOL), which mainly consists of a mixture of alkyl benzenes. The *x*-axis represents the ^{1}H chemical shift of the alkyl hydrogens, the *y*-axis the coupling (multiplicity) information, and the *z*-axis the intensity. The multiplicity reflected in the (^{1}H–^{1}H) coupling gives information on neighboring ^{1}H group interactions. Peak 7 shows that the methyl hydrogens are split into a triplet, which indicates the methyl group is adjacent to a group containing two hydrogens (the adjacent methylene group in the alkyl substituent). The three-dimensional presentation of the data can also be done in the form of a contour plot,[70] as shown in Figure 11.

Homonuclear correlated 2D NMR data from a correlation spectroscopy (COSY)-90 experiment[72] are presented in Figure 12 for a monoaromatic diesel fraction. The contour format is employed. Both horizontal and vertical axes pertain to the ^{1}H chemical shift. All ^{1}H groups

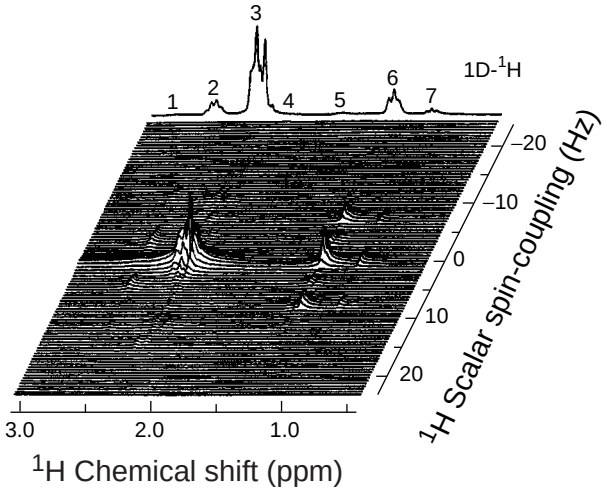

Figure 10 ^{1}H 2D NMR *J*-spectrum of PETSOL. ^{1}H chemical shift (ppm) on the *x*-axis. ^{1}H coupling (Hz) on the *y*-axis. The one-dimensional (1D) ^{1}H NMR spectrum is shown at the top of the diagram. (Reproduced by permission of Elsevier Science from D.J. Cookson, B.E. Smith, *Fuel*, **61**, 1007–1013, Copyright (1982).[70])

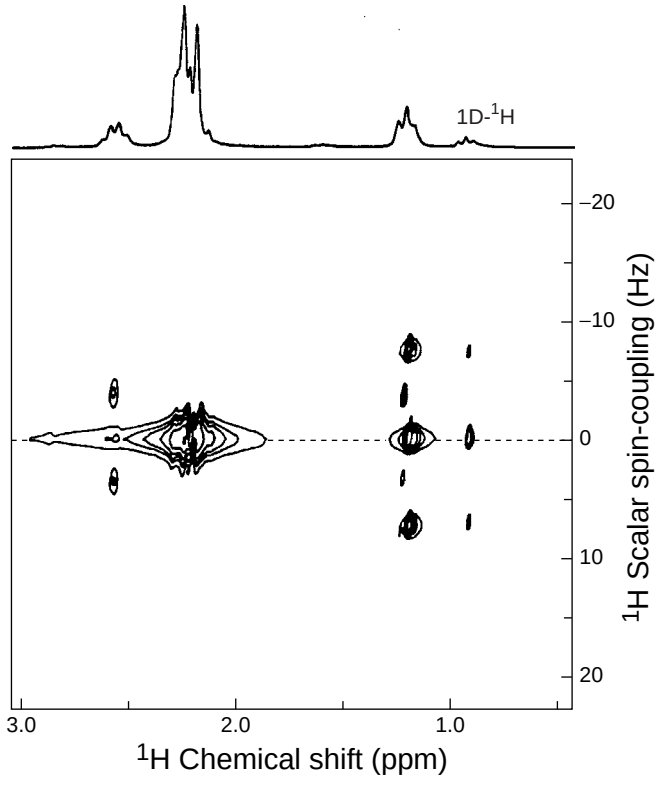

Figure 11 ^{1}H 2D NMR *J* spectrum of PETSOL. This is equivalent to Figure 10 but shown as a contour plot. (Reproduced by permission of Elsevier Science from D.J. Cookson, B.E. Smith, *Fuel*, **61**, 1007–1013, Copyright (1982).[70])

contribute to intensity on the diagonal. Off-diagonal peak intensity arises when two ^{1}H groups of different chemical shifts are coupled. The relationship between regions in the ^{1}H NMR spectrum can be interpreted as follows:[72] peaks 1 and 20 are coupled and belong to the 1-methylindan ring: 2–19, isopropyl; 3–9, indan ring: 4–17, 1-methyltetralin ring: 5–11, tetralin ring; 7-(13/12)-23, propyl; 7-(13/14)-(15/16)-22, butyl; 8,α $-OH_3$ not coupled, and 10–21, 2-methyltetralin ring.

The 2D heteronuclear (^{1}H–^{13}C) correlation (HETCOR) spectrum[71] of an aromatic fraction separated from a high-speed diesel fuel sample is shown in Figure 13. The 2.0–3.5 ppm region of the ^{1}H NMR spectrum is shown along the vertical axis, whereas the horizontal axis shows the ^{13}C DEPT 135° spectrum of the carbons in the aliphatic region (5–50 ppm). In the DEPT spectrum, signals of the CH and CH_3 carbons appear as negative and those from CH_2 appear as positive peaks. The signals between 18 and 20 ppm in the ^{13}C NMR spectrum are due to the methyl substituents of the aromatic carbons. These signals are used to assign the ^{1}H NMR region from 2.0 to 3.5 ppm to the hydrogens of the methyl group attached to an aromatic ring. Upon closer examination of the HETCOR spectrum, the ^{1}H region between 2.4 and

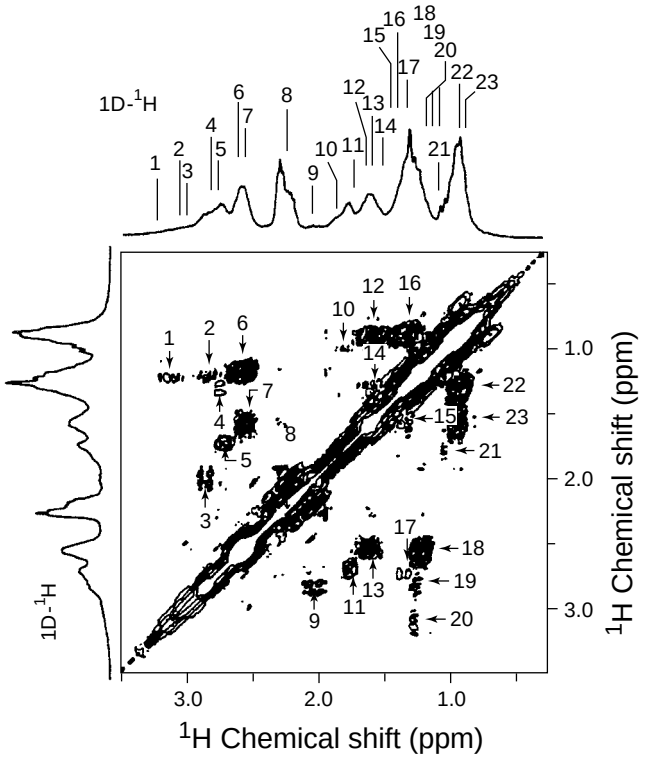

Figure 12 $^{1}H-^{1}H$ Homonuclear COSY 2D NMR spectrum of the aliphatic H_n groups in a fossil fuel diesel (monoaromatic fraction). With the exception of label 8 (α-CH_3 diagonal peak), labels 1–23 relate to cross peaks. (Reproduced with permission from D.J. Cookson, B.E. Smith, *Energy Fuels*, **1**, 111–120 (1987).[72] Copyright (1987) American Chemical Society.)

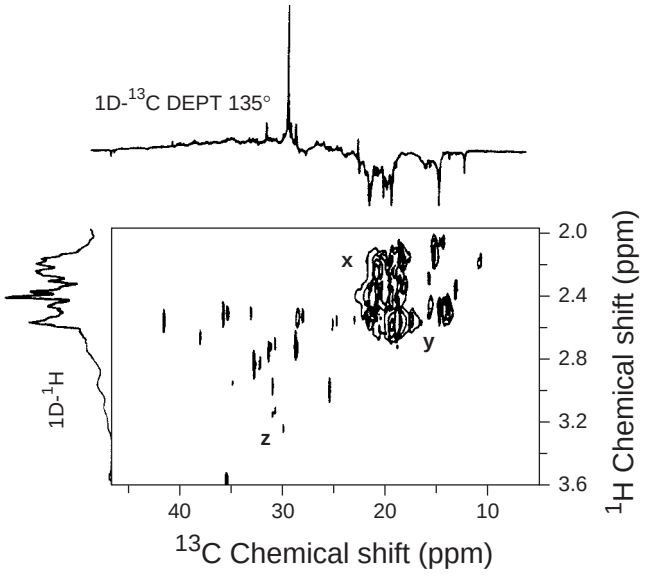

Figure 13 $^{1}H-^{13}C$ HETCOR 2D NMR spectrum of the aromatic fraction of a diesel fuel. (Reproduced with permission from V. Bansal, G.S. Kapur, A.S. Sarpel, V. Kagdiyal, S.K. Jain, S.P. Srivastava, A.K. Bhatnagan, *Energy Fuels*, **12**, 1223–1227 (1998).[71] Copyright (1998) American Chemical Society.)

3.5 ppm contains signals due to both α-CH_2 and α-CH_3 substituents on the aromatics.

A stacked plot of the HETCOR experiment[69] performed on a fossil fuel liquid containing both aromatic and olefinic material is shown in Figure 14. The 1D-^{13}C spectrum (vertical axis) shows complete overlap of the two types of sp^2 carbons. The 1D-^{1}H spectrum (horizontal axis), however, gives complete separation of the aromatic and olefinic hydrogens. Thus, through the use of the 2D/HETCOR spectrum, the olefinic ^{13}C peaks can be assigned based upon the splitting pattern observed in the ^{1}H spectrum. Several of the ^{13}C peaks are identified as specific types of olefin (shown in Figure 14). The unlabeled peaks are unassigned protonated aromatic carbons.

The preceding illustrations of the spectral editing and 2D NMR methods represent a few of the many techniques that are available to the NMR spectroscopist in an effort to identify and quantify carbon types in complex materials.

3.4 Average Molecular Structural Parameters

Detailed compositional characterization of petroleum fractions is extremely important, but has proved to be quite difficult and time-consuming. For these reasons, procedures have been reported which describe a particular fraction in terms of a hypothetical molecule.[73–75] The methods of average molecular structural determination usually incorporate results of elemental analysis, ^{1}H and ^{13}C NMR spectroscopy, and average molecular weight (AMW) determination from vapor pressure osmometry (VPO) to define the structural parameters of an average

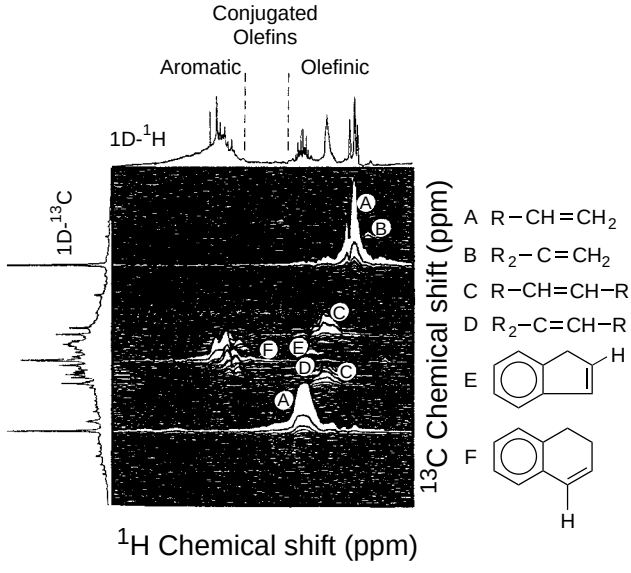

Figure 14 $^{1}H-^{13}C$ HETCOR 2D NMR spectrum of a fossil fuel liquid. (Reproduced by permission of CRC Press LLC from L.W. Dennis, R.E. Pabst, in *NMR of Humic Substances and Coal* (1987).[69] Copyright Lewis Publishers, an imprint of CRC Press, Boca Raton, Florida.)

Table 4 Average parameters generally calculated for liquid fuels

Average number of carbon atoms per alkyl substituent
Average carbon–hydrogen weight ratio of the alkyl groups
Percent substitution of alkyl groups on nonbridge aromatic ring carbons
Average number of aromatic ring carbon atoms per average molecule
Average number of nonbridge aromatic ring carbon atoms per average molecule
Average number of aromatic rings per average molecule
Average number of alkyl substituents per average molecule
Average number of naphthenic rings per average molecules
Molar ratio of aromatic carbon to total carbon in sample
Branchiness index
AMW
Percentage saturate carbon
Percentage naphthenic carbon
Number of naphthalene rings per substituent
Fraction nonbridge aromatic ring carbons

molecule. Table 4 lists some of the average parameters which are generally determined for petroleum fractions. It is important to realize that these average parameters are weighted averages for certain properties of the sample and may not represent actual structures present in the petroleum fraction. There is little evidence to indicate that complex mixtures exhibit the chemical and physical behaviors predicted by a single hypothetical structure. However, it can be reasoned that the narrower the distillation cut, the more the ''average molecule'' represents the chemical and physical properties of the fractions. There are many hidden assumptions in calculating the average molecular structure of a petroleum fraction.[76,77] One important major assumption is that all alkyl groups are present as substituents on aromatic rings. In solid materials, this is probably a good assumption (see section 4.1).

3.5 Applications

3.5.1 Average Molecular Weight

Characterization of oil fractions using average molecular parameters obtained by quantitative NMR analysis is often used to monitor the changes in composition at various stages of conversion processes and for simulating thermophysical properties of crude oil, such as phase equilibria. To simulate the thermophysical properties of crude oil, accurate values for the AMW of the oil distillate fractions are required to establish a relevant equation of state for the fluid. The data obtained through VPO are sometimes questionable.[78]

NMR structural analysis models can be used to obtain AMWs of high-boiling aromatic fractions.[78] A model using a ^{13}C spectral editing technique has been found to be particularly reliable. Tests on 17 narrow high-boiling aromatic fractions of a North Sea crude oil gave very good agreement with mass spectrometry data. In another study, a simple algorithm based on data from NMR spectra gives AMW values consistent with VPO values for different fractions as distillates, residues or asphaltenes.[79]

3.5.2 Characterization of Chemical Class in Petroleum Distillates

Spectral editing techniques, DEPT and QUAT, have been used to characterize and quantify several saturate fractions from a Cerro Negro heavy petroleum crude.[37] Saturate hydrocarbon fractions have been obtained from the 200–425, 425–500 and 550–700 °C distillate cuts and the >700 °C residue. The average molecular structural parameters calculated from the NMR data are given in Table 5. Quantitative ^{13}C NMR subspectral analysis has provided detailed molecular structural information on saturate fractions that have atmospheric equivalent

Table 5 Average molecular structural parameters for the saturate fractions from the Cerro Negro heavy petroleum crude

Structural parameter	Saturate fraction			
	200–425 °C	425–550 °C	550–700 °C	>700 °C
H/C ratio	1.85	1.91	1.84	1.89
Carbon number range	10–25	25–41	41–63	>74
Average number of carbon atoms per molecule (N_c)				
using midpoint of the temperature range	17	32	45	74
from NMR data	19.6	31.1	47.4	49.1
Average molecular weight (MW_A)	271	432	656	682
Average number of branching sites per molecule (BS)	6.6	9.0	14.3	13.3
Average number of branches per molecule (N_B)	7.6	10.1	14.7	13.3
Average number of rings per molecule (N_B) R	2.5	2.4	4.7	3.8
Average number of quaternary carbons per molecule	1.0	1.1	0.5	0.0
Average number of methine carbons per molecule	5.6	7.9	13.8	13.3
Average number of methylene carbons per molecule	8.3	14.8	25.8	28.1
Average number of methyl carbons per molecule	4.7	7.3	7.3	7.7

Reproduced with permission from D.A. Netzel, F.D. Guffey, *Energy Fuels*, **3**, 455–460 (1989).[37] Copyright (1989) American Chemical Society.

boiling points above the temperature for effective use of the GC/MS and ASTM (American Society for Testing and Materials) D-2786 mass spectral techniques. For the 200–425 and 425–550 °C distillate fractions, the NMR data correlate reasonably well with mass spectral data.

Conventional ^{13}C and ^{1}H NMR and spectral editing technique GASPE were used to provide detailed quantitative information regarding the abundance of major and minor alkanes and olefins in jet and diesel fuels from a Fischer–Tropsch syncrude.[80] The jet fuel consisted primarily of *n*-alkanes (87 wt%), while branched alkanes with, on the average, 12 carbon atoms per molecule and one CH_3 branch per molecule, represent another 9 wt%. Olefins represent nearly 4 wt% of the sample. From ^{1}H and ^{13}C NMR spectra of the olefins, the *cis* and *trans* isomers, as well as the position of the double bond, could be distinguished. Similar results were found for the diesel fuel in which the average number of carbon types per molecule was found to be approximately 15.

The branched and cyclic alkanes from the saturate fractions from diesel and kerosene fuels were characterized using the ^{13}C NMR technique, GASPE, to determine the fractional abundance of CH_n groups.[81,82] The *n*-alkane and branched plus cyclic alkanes were separated from each other using silica chromatography and urea clathration. The NMR data were used in calculating several average structure parameters, which included: number of branches per molecule, number of branches per branched molecule, number of branches per cyclic molecule, and number of rings per cyclic molecule.

A direct and fast method for estimating the aromatic carbon content in hydrofinished and hydrotreated types of base oils using ^{1}H NMR spectroscopy was reported by Sarpel et al.[83] In developing the method, the 2D HETCOR spectral analysis technique was used, which facilitated the estimation of quaternary and bridgehead aromatic carbons from the ^{1}H NMR spectrum. Using derived equations, the complete assignment of the ^{1}H NMR spectrum can be made, from which the aromatic content can be determined for the base oils. The methodology is highly useful for routine quality control analysis where a large number of base oils are required to be analyzed in a limited time. In principle, the method should easily be extended to other heavier ends like vacuum gas oils.

Bansal et al.[71] reported the methodology for the estimation of total aromatics and their distribution as mono- and polynuclear aromatics in diesel-range fuel products. The method is also based on the detailed analysis of the ^{1}H NMR spectrum. Advanced NMR techniques such as DEPT and 2D HETCOR were applied for the unambiguous assignment of the α-substitution on the aromatic ring in the ^{1}H NMR spectra (2.0–3.5 ppm region). The proposed ^{1}H NMR-based method correlates very well with the standard HPLC(IP-391/90) and mass spectrometric-based techniques.

1D and 2D NMR methods can be used to elucidate the structural characteristics of the monoaromatic fractions from petroleum and synthetic fuels.[72] Spectral assignments are supported by a range of NMR information, including CH_n types ($n = 0$–3), ^{1}H and ^{13}C chemical shifts, $^{1}H-^{13}C$ and $^{1}H-^{1}H$ resonance connectivities, and ^{1}H multiplicity. From the information provided by the various NMR techniques, the average number of α-CH_3, β-CH_3, γ-CH_3 groups per molecule can be calculated and a range of specific structures can be identified. Overall, the structural data are sufficient to identify similarities and differences for the monoaromatic fractions from different Australian petroleum sources.

3.5.3 Gasoline Octane Number

The octane number of a gasoline is a function of its chemical composition and therefore should correlate with the hydrocarbon types present. Mühl et al.[32] reported a correlation between the reformed gasoline chemical composition and its research octane number (RON) determined by the standard ASTM method (D2699). Using ^{1}H NMR spectral data to quantify the various hydrocarbon hydrogen types in 62 gasoline samples, they obtain the following linear regression relationship, Equation (1)

$$\text{RON (NMR)} = 92.37 + 0.30H_A + 0.31H_B - 0.87H_C - 0.35H_D \quad (1)$$

where H_A is the percentage of aromatic ring hydrogens, H_B is the percentage of α-aromatic CH_3 hydrogens, H_C is the percentage of paraffinic CH hydrogens, H_D is the percentage of paraffinic CH_2 hydrogens.

The method gives a satisfactory level of accuracy and is suitable for determining the octane number of gasoline samples generated in the laboratory when small sample quantity prevents use of the standard method.

3.5.4 Diesel Cetane Number

The cetane number (CN) measures the ignition quality of a diesel fuel. The effect of hydrocarbon type composition of the fuel on the ignition quality is qualitatively well-known. However, analytical methods to identify the amount of each hydrocarbon are time-consuming and thus prevent the establishment of models to represent quantitatively the effects of hydrocarbon type on ignition quality. Gülder and Glavincevski[84] developed a ^{1}H NMR method to predict the CN of diesel fuels from the carbon-type structural composition. Based on the carbon-type composition of 67 fuels, the CN can be predicted

using an expression of the functional form, Equation (2)

$$CN = B_0 + \frac{B_1}{C_A} + \frac{B_2}{C_A^2} + (B_3 C_\alpha + B_4 \ln C_\alpha) + (B_5 C_2 + B_6 C_2^2) + (B_7 C_3 + B_8 C_3^2) \quad (2)$$

where C_A is the number of aromatic carbons, C_α is the number of carbons at the α-position to aromatic rings, C_2 is the number of CH_2 CH carbons including β-CH_2, γ-CH_2 and β-CH_3 to aromatic rings, C_3 is the number of CH_3 carbons including terminal and branched and γ to aromatic rings, and $B_0 = 24.3848$; $B_1 = -286.728$; $B_2 = 587.3567$; $B_3 = 1.5227$; $B_4 = -15.882$; $B_5 = 0.9778$; $B_6 = 0.0047$; $B_7 = -0.2835$; $B_8 = 0.002$.

The constants, B_i, are determined by a multiple regression analysis fit of the data. The accuracy of the correlation is better than the accuracy of the cetane engine measurement on the basis of the spread of CNs determined by a number of standard cetane rating engines.

3.5.5 *Lubricant Base Oils*

Base oils are lubricant components which provide a fluid layer to separate dynamic surfaces. The fluid layer removes heat and wear particles while keeping friction at a minimum. The properties of base oils depend on the chemical composition and structural characteristics of hydrocarbon molecules present in the matrix. NMR techniques have been used to understand the carbon and hydrogen distribution in base oils in terms of average structural parameters (see Adhvaryu et al.[85] and the references they cite). These parameters, in principle, can be correlated with the basic properties required as specification for lube oil formulations. Using ^{13}C NMR, Singh et al.[86] studied industrial base oils refined from light and intermediate viscosity distillates and an aromatic-rich oil. A number of NMR average structural parameter values (Table 6) were correlated with several physicochemical characteristics of the lube base stocks (Table 7). Reasonable good correlations were determined for the percent aromatic carbon with the weight percent of sulfur, aniline point, the UOP characteristic factor, and the viscosity index. The study reveals that NMR could be

Table 6 NMR-derived parameters of lube base stocks

Characteristics	Sample identity			
	A	B	C	D
$\%C_{ar}$	5.0	14.0	25.0	39.2
$\%C_{ar,alk}$	0.6	5.4	9.9	11.2
$\%C_{ar,b}$	0.1	6.6	8.5	13.8
$\%C_{sat}$	95.0	85.2	75.0	60.8
$\%C_N$	22.7	23.4	18.8	21.7
$\%C_{n-p}$	36.5	26.4	27.5	19.7
$\%C_{i-p}$	35.8	35.4	28.7	19.4
%Straight chain alkane	29.5	20.0	20.4	15.5
Average chain length	11.89	10.78	9.08	7.58
f_a	0.05	0.15	0.25	0.39
H_{ar}	1.90	1.49	3.79	8.57
H	2.30	1.79	5.92	14.69
C/H	0.51	0.54	0.57	0.62

Reproduced with permission from H. Singh, A. Adhvaryu, I.D. Singh, G.S. Chaudhary, *preprint paper, ACS Div. Petrol. Chem.*, **42**(1), 255–258 (1997).[86] Copyright (1997) American Chemical Society.

Table 7 Physicochemical characteristics of lube base stocks

Characteristics	Sample identity			
	A	B	C	D
Density ($g\,mL^{-1}$) at 15 °C	0.8772	0.8952	0.9077	1.0005
API gravity (degrees)	29.7	26.5	24.3	9.8
Pour point (°C)	−6	0	−3	19
Kinetic viscosity (cSt) at				
100 °C	5.08	10.46	4.08	24.52
40 °C	28.63	91.26	22.18	849.63
Viscosity index	104	96	67	−35
Aniline point (°C)	94.8	101.6	76.6	–
Conradson carbon residue (%wt)	0.01	0.02	0.05	3.00
Sulfur (%wt)	1.02	1.36	2.28	3.08
Molecular weight (VPO)	370	468	325	390
Characterization factor (UOP)	12.38	12.52	11.79	11.18
Hydrocarbon type composition (ASTM D 2549-91) (%wt)				
Saturates	62.5	60.3	50.2	15.5
Aromatics (including polars)	36.5	39.7	49.8	84.5

API, American Petroleum Institute; UOP, universal oil products.
Reproduced with permission from H. Singh, A. Adhvaryu, I.D. Singh, G.S. Chaudhary, *preprint paper, ACS Div. Petrol. Chem.*, **42**(1), 255–258 (1997).[86] Copyright (1997) American Chemical Society.

a potent technique for predicting various physicochemical properties which are used for base oil specifications (see also section 4.2).

4 SOME APPLICATIONS OF NUCLEAR MAGNETIC RESONANCE TO PETROLEUM PROCESSING

4.1 Average Molecular Structures Using Solid-state Nuclear Magnetic Resonance

The observed NMR signals represent all the carbons in the sample and cannot distinguish whether these carbons are in a single molecular entity or in a mixture of many different molecular types. Sometimes it is advantageous to construct an average molecule from average molecular structural parameters to monitor changes that result from varying process conditions, feedstocks, or other variables. In petroleum fractions, in which aliphatic and aromatic carbons, for example, exist in separate compounds, an average molecule does not represent the average properties of the mixture and is of little value in studying changes in process conditions. In solids such as coals, however, the situation is different because the concentration of paraffins is small. Any aliphatic carbons present are usually assumed to be attached to an aromatic cluster so that an average molecular representation can be constructed and the change in process conditions can be monitored.

Solum et al.[87] have devised an NMR procedure for calculating a number of molecular parameters that are useful in petroleum processing that involves solid or insoluble materials. DD relaxation data and integrated intensities of selected chemical shift ranges in a CP/MAS NMR spectrum are used to calculate 12 basic structural parameters, from which an additional eight structural parameters can be derived. A list of representative chemical shifts and carbon structures is given in Table 1. Definitions of the integrated areas are listed in Table 8. The superscripts on the areas refer to the chemical shift range in parts per million over which the integrals are evaluated. NMR structural parameters and additional lattice parameters derived from these data are given in Table 9. The procedure for calculating the structural parameters in Table 9 is from Solum et al.[87]

Storm et al.[88] used this procedure to develop molecular representations of petroleum asphaltenes and to develop correlations that predict fouling due to sludge formation in visbreaking and hydrocracking processes. Sludge formation occurs when asphaltenes flocculate. Therefore a better understanding of asphaltene molecular properties could lead to a better understanding of the mechanisms of coking and sludge formation in refinery operations. However, it should be stressed that asphaltenes are defined on the basis of solubility and that an asphaltene molecule per se does not exist. Asphaltenes are a mixture of strongly interacting molecules having colloidal properties. Therefore using an average asphaltene structure to represent a mixture is somewhat dubious and should be used with caution.

Nevertheless, Storm et al.[88] derived molecular representations of Ratawi and Alaska North Slope asphaltenes. These differed from other ''average structure'' calculations in that a molecular weight is not independently measured in order to calculate the average molecule. Instead the procedure automatically provides estimates of the AMWs that are consistent with independent determinations. The Ratawi asphaltenes are represented by four condensed rings with four side chains

Table 8 Definition and symbols for the measured and calculated parameters used in DD and CP NMR techniques

Parameter	Measured and calculated
Total intensity of aromatic carbons at dephasing time, t	$M(t) = M_G^0 \exp(-0.5(t/T_{2G})^2) + M_L^0 \exp(-t/T_{2L})$
Intensity of protonated aromatic carbons at $t = 0$	M_G^0
Intensity of nonprotonated aromatic carbons at $t = 0$	M_L^0
Gaussian component of DD relaxation time constant	T_{2G}
Lorentzian component of DD relaxation time constant	T_{2L}
Dephasing time	t
Aliphatic carbon integrated area from 0 to 90 ppm	A^{0-90}
Carbonyl carbon integrated area from 165 to 240 ppm	$A^{165-240}$
Aromatic and carbonyl carbon integrated area from 90 to 240 ppm	A^{90-240}
Phenolic carbon integrated area from 150 to 165 ppm	$A^{150-165}$
Alkyl substituted aromatic carbon integrated area from 135 to 150 ppm	$A^{135-150}$
Aromatic carbons integrated area from 90 to 165 ppm	A^{90-165}
Aliphatic carbons bonded to oxygen integrated area from 50 to 90 ppm	A^{50-90}
Methoxy carbon integrated area from 50 to 60 ppm	A^{50-60}
Methyl carbon integrated area from 0 to 25 ppm	A^{0-25}

Table 9 Calculation of structural and lattice parameters

Definition	Formula
Structural parameters	
1. Fraction of total sp^2 carbon (f_a)	$f_a = A^{90-240}/(A^{0-90} + A^{90-240})$
2. Fraction of total sp^3 carbon (f_{al})	$f_{al} = 1 - f_a$
3. Fraction of carbonyl carbons (f^{CO})	$f^{CO} = f_a(A^{165-240}/A^{90-240})$
4. Fraction of aromatic carbons (f'_a)	$f'_a = f_a - f^{CO}$
5. Fraction of phenolic carbons (f_a^P)	$f_a^P = f'_a(A^{150-165}/A^{90-165})$
6. Fraction of alkyl substituted aromatic carbons (f_a^S)	$f_a^S = f'_a(A^{135-150}/A^{90-165})$
7. Fraction of methyl and methoxy carbons (f_{al}^*)	$f_{al}^* = f_{al}\{(A^{0-25} + A^{50-60})/A^{0-90}\}$
8. Fraction of methine and methylene carbons (f_{al}^H)	$f_{al}^H = f_{al} - f_{al}^*$
9. Fraction of aliphatic carbons bonded to oxygen (f_{al}^O)	$f_{al}^O = f_{al}(A^{50-90}/A^{0-90})$
10. Fraction of quaternary aromatic carbons (f_a^Q)	$f_a^Q = M_L^0/(M_L^0 + M_G^0)$
11. Fraction of tertiary aromatic carbons (f_a^T)	$f_a^T = 1 - f_a^Q$
12. Fraction of protonated aromatic carbons normalized to total carbons (f_a^H)	$f_a^H = f'_a \times f_a^T$
13. Fraction of quaternary aromatic carbons normalized to total carbons (f_a^N)	$f_a^N = f'_a \times f_a^Q$
14. Fraction of aromatic bridgehead carbons (f_a^B)	$f_a^B = f_a^{Q'} - f_a^P - f_a^S$
Lattice parameters	
15. Mole fraction of bridgehead carbons (χ_b)	$\chi_b = f_a^B/f'_a$
16. Mole fraction of bridgehead carbons for linear catenation cluster ($\chi_{b'}$)	$\chi_{b'} = (1/2) - (3/C_{ar})$
17. Mole fraction of bridgehead carbons for circular catenation cluster ($\chi_{b''}$)	$\chi_{b''} = 1 - (6/C_{ar})^{1/2}$
18. Number of aromatic carbons per cluster (C_{ar})	$\chi_b = \{1 + \tanh[(C_{ar} - C_o)/m]\}\chi_{b'}/2 + \{1 + \tanh[(C_{ar} - C_o)/m]\}\chi_{b''}2$ $C_o = 19.57 \quad m = 4.15$
19. Number of attachments per cluster ($\sigma + 1$)	$\sigma + 1 = (f_a^S + f_a^P)C_{ar}/f'_a$
20. Fraction of intact bridges per lattice (P_o)	$P_o = (f_a^S + f_a^P - f_{al}^*)/f_a^S + f_a^P)$
21. Number of bridges and loops per cluster (BL)	$BL = (\sigma + 1)P_o$
22. Number of side chains per cluster (SC)	$SC = (\sigma + 1) - BL$
23. Molecular weight of cluster (MW_c)	$MW_c = (12C_{ar})/f'_a(\%C/100)$
24. Molecular weight per attachment	$M_\delta = [MW - (12f_a^N C_{ar} - 13f'_a)/(f_a^N + f_a^H)]/(\sigma + 1)$

and one connecting bridge, whereas the Alaska North slope asphaltenes are represented by four condensed rings with three side chains and one connecting bridge.

Sludge formation during heavy oil upgrading is a problem in many petroleum refinery processes. Often a soft coke-like substance forms during visbreaking and hydrocracking processes. This limits the severity of the conversion because it accumulates in downstream equipment, causing fouling. Because more petroleum operations use the bottom-of-the barrel for upgrading, methods to predict the production rate of sludge formation from chemical characteristics of the crude oil are of great importance. Storm et al.[89] found that the propensity of a feedstock to form sediment is related to a parameter which characterizes the degree of condensed polynuclear aromaticity of the resid. The polynuclear aromaticity is defined as the fraction of aromatic bridgehead carbons. This value can be obtained from NMR data using the formulas given in Table 9. The propensity to form sludge is higher the more the polynuclear aromatic hydrocarbon cores are highly substituted with alkyl groups. The degree of substitution can also be obtained from NMR data using the formulas in Table 9.

When applying the analytical procedure just described, it should be pointed out that, depending on the material under investigation, not all carbon types are necessarily present in the sample, and, therefore, will not be observed in the normal CP/MAS spectrum. For example, the procedure described was derived for characterization of coals. Phenolic and carboxylate carbon structures are generally part of the organic structure of these materials. In asphaltenes, these carbons are usually absent, or in low abundance and can be ignored.

4.2 Predictive Property Correlations

As the world continues to deplete its supply of high-quality crude oil, heavy crude oils that contain molecules of higher molecular weight and are more aromatic and contain more heteroatoms are being processed to a greater extent than previously. To process these heavy oils, their thermodynamic and transport properties must be known. Methods to estimate these basic parameters generally use normal boiling points and liquid densities at 15.6 °C (60 °F).[90] Many of these correlations are not applicable to fossil fuels that contain a significant number

of heteroatoms. In addition, the liquid densities and boiling points that are needed may be experimentally inaccessible.

NMR can be used effectively to provide molecular structure parameters for use in predictive property calculations.[91–94] Schwarz and Prausnitz[91] used NMR to develop a correlation for thermodynamic properties of heavy fossil-fuel fractions. Elemental analyses provide the relative abundances of C, H, O, N, and S atoms. An NMR spectrum is used to distinguish between: hydrogen bonded to an aromatic carbon, hydrogen bonded one carbon away from an aromatic ring, hydrogen bonded to a nonterminal carbon further than one carbon away from an aromatic ring, and hydrogen bonded to a methyl group that is not bonded to an aromatic ring. Infrared spectroscopy was also used to distinguish between ether and hydroxy oxygen and to distinguish between amines and other nitrogen groups.

Schwarz and Prausnitz[91] give detailed example calculations for converting raw characterization data to specific characterization parameters used in their perturbed-hard-chain equation of state. The procedure is not intended for calculations on conventional fossil fuels where experimental data on boiling points and densities are available. Instead the procedure is intended for heavy fossil fuels, where conventional correlations are unreliable or input data are unavailable.

Cookson and Smith[92] derived simple linear property–composition relationships for a range of 51 jet fuels and 48 diesel fuels using ^{13}C NMR alone. The relationships were of the form, Equation (3)

$$P = b_1 C_n + b_2 C_{ar} + c \tag{3}$$

where P is the property in question, b_1, b_2, and c are coefficients, and C_n is the fraction of total ^{13}C NMR signal intensity attributable to long-chain alkyl-carbon occurring at 32.3, 30.1, 29.8, 23.0 and 14.2 ppm. C_{ar} is the total aromatic carbon signal in the region, 110–160 ppm.

For jet fuels in the boiling range, 190–230 °C, the properties included smoke point, aromatics content, hydrogen content, heat of combustion, freezing point and specific gravity. For diesel fuels in the range 230–320 °C, the properties included pour point, cloud point, hydrogen content, aniline point, specific gravity, diesel index, and cetane index. The results are summarized in Tables 10 and 11. Of particular note is that the relationships all have the same form and use the same type of analytical information.

Table 10 Models for jet fuel properties based on NMR compositional parameters[a]

Property	Method	No. of samples	Range	Coefficients		
				b_1	b_2	c
Smoke point (mm)	IP57	44	15–38	25.7	−66.8	23.6
Aromatics content (vol%)	ASTM D1319	49	1.6–24.2	−1.1	139.4	2.4
Hydrogen content (wt%)	mod. ASTM D3701	36	12.63–14.88	2.47	−5.01	13.62
Gross heat of combustion (MJ kg^{-1})	IP12	33	45.24–46.94	2.02	−4.06	45.87
Net heat of combustion (MJ kg^{-1})	IP12	30	42.54–43.78	1.50	−2.78	42.96
Freezing point (°C)	IP16	25	−50 to −32	49.8	18.1	−61.6
Inverse specific gravity (60/60F)	ASTM D4052	38	1.146–1.292	0.225	−0.131	1.178

[a] Model: property value $= b_1 C_n + b_2 C_{ar} + c$; boiling range = 190–230 °C.
Reproduced with permission from D.J. Cookson, B.E. Smith, *Energy Fuels*, **4**, 152–156 (1990).[92] Copyright (1990) American Chemical Society.

Table 11 Models for diesel fuel properties based on NMR compositional parameters[a]

Property	Method	No. of samples	Range	Coefficients		
				b_1	b_2	c
Pour point (°C)	ASTM D97	32	−21 to 0	75.0	52.8	−47.6
Cloud point (°C)	ASTM D2500	21	−17 to 5	65.3	45.9	−39.5
Hydrogen content (wt%)	mod. ASTM D3701	39	12.39–14.48	2.95	−5.34	13.18
Aniline point (°C)	ASTM D611	36	41.5–82.9	55.8	−69.5	59.5
Inverse specific gravity (60/60F)	STM D4052	33	1.100–1.244	0.251	−0.165	1.126
Diesel index	IP21	42	25.6–80.8	84.3	−77.4	39.9
Cetane index	ref. 12	36	31.2–68.4	53.4[b]	−68.5	44.6

[a] Model as in Table 10; boiling range = 230–320 °C.
Reproduced with permission from D.J. Cookson, B.E. Smith, *Energy Fuels*, **4**, 152–156 (1990).[92] Copyright (1990) American Chemical Society.

4.3 Determination of Hydrogen Consumption During Processing

In the 1990s there has been a significant increase in residuum hydroconversion in refineries throughout the world. To a large extent, this expansion was derived from expectations that the spread in market values between high-sulfur fuel oil and light transportation fuels would increase to support the investments required for these conversion facilities.

Residuum conversion involves increasing the hydrogen-to-carbon (H/C) ratio and decreasing the molecular weight of the distillate products. However, because of the hydrogen imbalances, either carbon is rejected or hydrogen is added to achieve a desired product state. For those processes that involve hydrogen addition, NMR can be used effectively to determine the types of reaction that consume hydrogen under a given set of process conditions.

Finseth et al.[95] proposed a method to monitor the changes in hydrogen utilization during coal liquefaction. The same approach can be used to study the role of hydrogen during residuum upgrading.[96,97] This method incorporates analytical methods that are readily available. In the method of Finseth et al.,[95] the total hydrogen incorporated in the products during conversion is given by Equation (4)

$$\Delta H_T = \Delta H_{hyd} + \Delta H_{NOS} + \Delta H_{C1-C4} + \Delta H_{matrix} \qquad (4)$$

where ΔH_T is the total hydrogen incorporated, ΔH_{hyd} is the hydrogen incorporated by hydrogenation, ΔH_{NOS} is the hydrogen associated with heteroatom removal, ΔH_{C1-C4} is the hydrogen associated with hydrocarbon gas production, ΔH_{matrix} is the hydrogen incorporated in nonvolatile solid and liquid products owing to alkyl bond scission.

Thus, measurement of hydrogen incorporation during processing is reduced to measuring ΔH_T by elemental analysis, measuring ΔH_{hyd} by ^{13}C NMR aromaticity measurements in the solid and liquid reactants and products, and measuring ΔH_{C1-C4} and ΔH_{NOS} by quantitative gas analyses. The term ΔH_{matrix} is then determined by difference. If this term is negative, hydrogen is liberated because of condensation reactions. If ΔH_{matrix} is positive, then thermolysis reactions consume hydrogen during liquefaction.

The method assumes that, on average, one hydrogen atom is incorporated into the organic product for every heteroatom removed from the feed material; hydrogenation consumes one hydrogen atom for every carbon atom reduced (each 0.01 change in aromaticity); and two hydrogen atoms are consumed per mole of light hydrocarbons formed. The rationale for these assumptions is discussed in greater detail by Finseth et al.[95] and by Beret and Reynolds,[96,97] who also used the method to determine hydrogen incorporation in petroleum residuum hydroconversion.

Beret and Reynolds[96] applied this method to study residuum hydroconversion of three California residua processed at different severities using different sets of operating variables such as temperature, pressure, residence time, hydrogen circulation route and catalyst. The approach yielded a detailed description of hydrogen incorporation during residuum hydroconversion. Character of the feedstock, processing conditions, and catalyst all affected how hydrogen was incorporated into the residuum matrix. Hydrogenation and hydrocracking reaction incorporated hydrogen equally at moderate process severities (343 °C or 650 °F). At higher severities (454 °C or 850 °F), the bulk of the hydrogen was used to cap radicals generated from cracking reactions. Thus, by a judicious choice of operating conditions, certain hydrogen incorporation reactions could be favored over others. The feedstock also influences how hydrogen was incorporated. With aromatic feedstocks, hydrogenation reactions compete for hydrogen equally with hydrocracking. With less aromatic feedstocks, hydrocracking dominates over hydrogenation reactions.

In another study, Beret and Reynolds[97] used the method to study the effects of prehydrogenation on hydroconversion of Maya residuum. A Maya 343 °C (650 °F) residuum was prehydrogenated using a standard hydroprocessing catalyst. The 343 °C (650 °F) residuum of this product was then processed separately at selected conditions and the products analyzed to determine how hydrogen was incorporated during processing. The results were compared with direct hydroconversion. The distribution of hydrogen was dependent upon the type and severity of the process. In direct hydroconversion, about 25–30% of the total hydrogen was incorporated from heteroatom removal and hydrocarbon gas formation. The two-step process, which involved mild prehydrogenation, primarily incorporates hydrogen by hydrogenation reactions at the expense of C_4 gas formation and cracking, particularly at low severity. At moderate severity, cracking reactions occur but are suppressed to some extent. The real benefit came from suppression of C_4 gas formation.

4.4 Nuclear Magnetic Resonance Studies of Coking

Delayed coking is the most widely used petroleum residue conversion process in the world.[98,99] Although the process is about 60 years old, there are still many aspects of the chemical reactions taking place that are poorly understood. Of particular importance is the ability to predict product yield and quality from basic feedstock composition. Yield data are important for assessing process economics, and maximum conversion to liquid

products, primarily gasoline and distillate product is generally the main objective. Solid- and liquid-state NMR techniques can provide useful information about delayed coking processes similar to that described in the previous section. Complete NMR characterization (^{1}H and ^{13}C) of a feedstock and delayed coking products, coupled with material balances, can be used to obtain the distribution of carbon and hydrogen atoms in a delayed coking process.

Rodriguez et al.[98] used such an approach to study the delayed coking of a fluid catalytic cracker decanted oil in a pilot plant operation that simulated a conventional commercial operation. They demonstrate how to perform the material balance to reveal changes in the carbon and hydrogen distributions resulting from carbonization reactions during delayed coking. Rodriguez et al.[99] found a significant increase in the number of aromatic carbons, which was due mainly to an increase in the number of quaternary aromatic carbons. This was interpreted as an increase in the extent of condensation reactions.

4.5 Process Kinetics

The reactions that occur during catalytic and thermal upgrading of crude oil are very complex and involve a large number of interacting species. In such complex mixtures, it is not possible to have a knowledge of all the interacting species. Consequently, lumped kinetics methods, wherein similar reactive species are grouped together to develop rate constants, have been developed to analyze reactions in complex mixtures.[100] Lumping can be used to reduce the number of experiments required

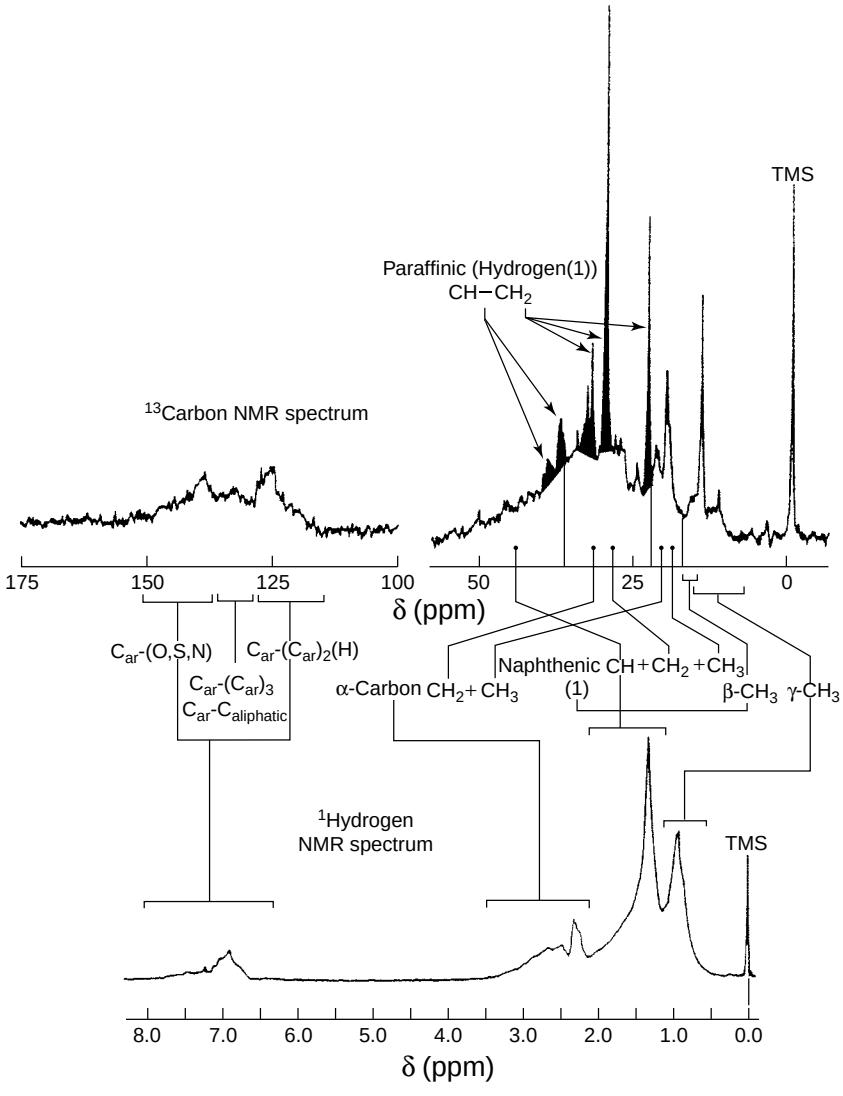

Figure 15 Assignment of carbon and hydrogen types to groups in regions of the ^{13}C and ^{1}H NMR spectra. (Reproduced with permission from M.R. Gray, *Ind. Eng. Chem. Res.*, **29**, 505–512 (1990).[100] Copyright (1990) American Chemical Society.)

to define the kinetic behavior of the reacting system, simplify the network of possible reaction paths, and reduce the overhead of analytical work required to define component-by-component compositions.

NMR spectroscopy provides a method for estimating the concentrations of a selected set of groups to be used in lumped kinetics as long as the sample is soluble in an appropriate solvent. Data from NMR analyses can be combined with elemental analyses to obtain concentrations of structural groups. The main structural groups that can be estimated by combining NMR with elemental analysis are shown in Figure 15. The spectra in Figure 15 are typical of heavy oils and illustrate the regions where signals appear from the different structural groups. The assignment of groups to the ^{13}C NMR spectra involves some approximations that are due to overlapping peaks. The lack of sharp signals (except for paraffinic CH and CH_2) is due to the large number of isomeric compounds that have overlapping resonances. NMR provides a unique way of monitoring changes in the concentrations of these groups independent of molecular weight so long as the samples are soluble. Insoluble material can be analyzed by solid-state NMR, but less structural information can be deduced.

An example of the use of lumped kinetics in petroleum upgrading is that of Gray.[100] Data from thermal and catalytic hydrotreating of a Syncrude coker gas oil in a continuous-stirred reactor were used to study lumping of reaction kinetics based on structural groups. Feed and product compositions were described by a limited set of aliphatic and aromatic carbon groups, based on data from elemental analysis and ^{1}H and ^{13}C NMR spectroscopies. The study demonstrated that stoichiometric relationships or ratios between groups can be deduced from rate data, even in complex mixtures. These ratios could then be used to construct models for hydroprocessing. For example, the rate of heteroatom removal can be used to predict changes in the carbon aromaticity of the oil, which in turn, can be used to estimate the CN.

ABBREVIATIONS AND ACRONYMS

AMW	Average Molecular Weight
ASTM	American Society for Testing and Materials
CN	Cetane Number
COSY	Correlation Spectroscopy
CP	Cross-polarization
DD	Dipolar Dephasing
DEPT	Distortionless Enhancement by Polarization Transfer
FID	Free-induction Decay
GASPE	Gated Spin-echo ^{13}C NMR
H/C	Atomic Hydrogen to Carbon
HETCOR	Heteronuclear Correlation
MAS	Magic Angle Spinning
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Enhancement Effect
PETSOL	Petroleum Solvent
QUAT	Quaternary Only Carbon Spectrum
RF	Radiofrequency
RON	Research Octane Number
SEBBORD	Spin-echo Broad-band Off-resonance Decoupling
S/N	Signal-to-noise Ratio
SPE	Single-pulse Excitation
TKS	*Tetrakis*(trimethylsilyl) silane
TMS	Tetramethylsilane
TOC	Total Organic Carbon
UOP	Universal Oil Product
VPO	Vapor Pressure Osmometry
2D	Two-dimensional

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Full Range Crudes, Analytical Methodology of • Lubricant Base Oils: Analysis and Characterization of • Oil Shale and Shale Oil Analysis • Petroleum Residues, Characterization of

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses

FURTHER READING

Books on Nuclear Magnetic Resonance

F. Kasler, *Quantitative Analysis by NMR Spectroscopy*, Academic Press, New York, 190, 1973.

L. Petrakis, D. Allen, *NMR for Liquid Fossil Fuels*, Elsevier, New York, 242, 1987.

L.D. Field, S. Sternhell (eds.), *Analytical NMR*, John Wiley, New York, 250, 1989.

D.E. Axelson, *Solid State Nuclear Magnetic Resonance of Solid Fossil Fuels, Multiscience Publications*, CANMET and Canadian Government Centre, 1985.

C.A. Fyfe, *Solid State NMR For Chemists*, CFC Press, Guelph, Ontario, Canada, 1983.

M.A. Wilson, *NMR Techniques and Applications in Geochemistry and Soil Chemistry*, Pergamon Press, New York, 1987.

C.L. Petrakis, J.P. Fraissard (eds.), *Magnetic Resonance, Introduction, Advanced Methods, and Applications to Fossil Energy*, NATO ASI Series, D. Reidel Publishing, Dordrecht, The Netherlands, Vol. 124, 1984.

R.L. Wershaw, M.A. Mikita (eds.), *NMR of Humic Substances and Coal*, Lewis Publishers, Chelsea, MI, 1987.

R.E. Botto, Y. Sanada (eds.), *Magnetic Resonance of Carbonaceous Solids*, Advances Chemical Series, American Chemical Society, Washington, DC, Vol. 229, 1993.

Journals

Anal. Chem. (publishes applications reviews every two years, including petroleum).

Energy and Fuels.

Fuel.

J. Petrol Sci. Technol.

REFERENCES

1. H.L. Retcofsky, K.D. Rose, F.P. Miknis, 'Magnetic Resonance of Heavy Fossil Fuels Early History', *ACS Preprint, Div. Petrol. Chem.*, **29**, 232–235 (1985).
2. R.B. Williams, 'Nuclear Magnetic Resonance', *Spectrochim. Acta*, **14**, 24–44 (1959).
3. S.A. Knight, 'Analysis of Aromatic Petroleum Fractions by Means of Absorption Mode Carbon-13 NMR', *Chem. Ind.*, **45**, 1920–1923 (1967).
4. J.N. Shoolery, W.L. Budde, 'Natural Abundance Carbon-13 Nuclear Magnetic Resonance Spectrometry for Crude Oil and Petroleum Product Analyses', *Anal. Chem.*, **48**(11), 1458–1461 (1976).
5. D.L. VanderHart, H.L. Retcofsky, 'Estimation of Coal Aromaticities by Proton-decoupled Carbon-13 Magnetic Resonance Spectra of Whole Coals', *Fuel*, **55**, 202–204 (1976).
6. V.J. Bartuska, G.E. Maciel, J. Schaefer, E.O. Stejskal, 'Prospects for Carbon-13 Nuclear Magnetic Resonance Analysis of Solid Fossil Fuel Materials', *Fuel*, **56**, 354–358 (1977).
7. H.A. Resing, A.N. Garroway, R.N. Hazlett, 'Determination of Aromatic Hydrocarbon Fraction in Oil Shale by ^{13}C NMR with Magic Angle Spinning', *Fuel*, **57**, 450–454 (1978).
8. F.P. Miknis, D.A. Netzel, 'NMR Studies of Oil Shale and Related Materials', in *Magnetic Resonance in Colloid and Interface Science*, Symposium Series, eds. H.A. Resing, C.G. Wade, American Chemical Society, Washington, DC, 182–188, 1976.
9. M. Zhang, G.E. Maciel, 'Large-sample ^{13}C MAS NMR Spectroscopy of Coal: Relaxation and Spin Counting', *Fuel*, **69**, 557–563 (1990).
10. R.E. Botto, R. Wilson, R.E. Winans, 'Evaluation of the Reliability of Solid ^{13}C NMR Spectroscopy for the Quantitative Analysis of Coals: Study of Whole Coals and Maceral Concentrates', *Energy Fuels*, **1**, 173–181 (1987).
11. L.M. Stock, J.V. Muntean, R.E. Botto, 'A New Approach for Quantitative ^{13}C NMR Spectroscopy of Coal', in *New Trends in Coal Science*, NATO ASI Series C, ed. Y. Yurum, Kluwer Academic Publishers, Boston, 159–167, 1988.
12. J.V. Muntean, L.M. Stock, R.E. Botto, '*Tetrakis*(Trimethylsilyl) silane, a Suitable Chemical-shift Standard for Solid-state NMR Spectroscopy', *J. Magn. Reson.*, **76**(3), 540–542 (1988).
13. E.W. Hagaman, F.M. Schell, D.C. Cronauer, 'Oil Shale Analysis by CP/MAS ^{13}C NMR Spectroscopy', *Fuel*, **63**, 915–919 (1984).
14. A. Jurkiewicz, C.E. Bronnimann, G.E. Maciel, 'Spin Dynamics and Spin Counting in the ^{13}C CP/MAS Analysis of Argonne Premium Coals', *Fuel*, **73**, 828–835 (1994).
15. F.P. Miknis, M.J. Sullivan, V.J. Bartuska, G.E. Maciel, 'Cross Polarization Magic Angle Spinning ^{13}C NMR Spectra of Coals of Varying Rank', *Org. Geochem.*, **3**, 19–28 (1981).
16. R.L. Dudley, C.A. Fyfe, 'Evaluation of the Quantitative Reliability of the ^{13}C CP/MAS Technique for the Analysis of Coals and Related Materials', *Fuel*, **61**(7), 651–657 (1982).
17. K.J. Packer, R.K. Harris, A.M. Kenwright, C.E. Snape, 'Quantitative Aspects of Solid-state ^{13}C NMR of Coals and Related Materials', *Fuel*, **62**, 999–1002 (1983).
18. M.A. Wilson, P.J. Collin, R.J. Pugmire, D.M. Grant, 'Relaxation of Carbon-13 and Nitrogen-15 Nuclei by Solvent-refined Coal', *Fuel*, **61**(10), 959–967 (1982).
19. M.A. Wilson, A.M. Vassallo, P.J. Collin, H. Rottendorf, 'High Resolution Carbon-13 Nuclear Magnetic Resonance of Organic Solids for Fossil Fuels', *Anal. Chem.*, **56**(3), 433–436 (1984).
20. C.E. Snape, D.E. Axelson, R.E. Botto, J.J. Delpeuch, P. Tekely, B.C. Gerstein, M. Pruski, G.E. Maciel, M.A. Wilson, 'Quantitative Reliability of Aromaticity and Related Measurements on Coals by ^{13}C NMR. A Debate', *Fuel*, **68**, 547–560 (1989).
21. R.A. Wind, G.E. Maciel, R.E. Botto, 'Quantitation in ^{13}C NMR Spectroscopy of Carbonaceous Solids', in *Magnetic Resonance of Carbonaceous Solids*, eds. R.E. Botto, Y. Sanada, Advances in Chemistry Series, 229, American Chemical Society, Washington, DC, 1993.
22. F.P. Miknis, V.J. Bartuska, G.E. Maciel, *Cross Polarization ^{13}C NMR with Magic Angle Spinning: Some Applications to Fossil Fuels and Polymers*, American Laboratory, Vol. II, 19, Nov. 1979.
23. J.A. Franz, R. Garcia, J.C. Linehan, G.D. Love, C.E. Snape, 'Single-pulse Excitation ^{13}C NMR Measurements on the Argonne Premium Coal Samples', *Energy Fuels*, **6**, 598–602 (1992).
24. G.D. Love, R.V. Law, C.E. Snape, 'Determination of Nonprotonated Aromatic Carbon Concentrations in

Coals by Single-pulse Excitation ^{13}C NMR', *Energy Fuels*, **7**, 639–644 (1993).
25. M.J. Trewhella, I.J.F. Poplet, A. Grint, 'Structure of Green River Oil Shale Kerogen: Determination Using Solid State ^{13}C NMR Spectroscopy', *Fuel*, **65**, 541–546 (1986).
26. J.M. Hunt, *Petroleum Geochemistry and Geology*, 2nd edition, W.H. Freeman and Co., San Francisco, 743, 1996.
27. B.P. Tissot, D.H. Welte, *Petroleum Formation and Occurrence*, 2nd edition, Springer-Verlag, Berlin, 699, 1984.
28. A.G. Requejo, N.R. Gray, H. Freund, H. Thomann, H.M.T. Melchior, L.A. Gebhard, M. Bernardo, C.F. Pictroski, C.S. Hsu, 'Maturation of Petroleum Source Rocks. 1. Changes in Kerogen Structure and Composition Associated with Hydrocarbon Generation', *Energy Fuels*, **6**, 203–214 (1992).
29. F.P. Miknis, D.A. Netzel, R.C. Surdam, 'NMR Determination of Carbon Aromatization During Hydrous Pyrolysis of Coals from the Mesaverde Group, Greater Green River Basin', *Energy Fuels*, **10**, 3–9 (1996).
30. P.R. Solomon, M.A. Serio, G.V. Despande, E. Kroo, 'Cross-linking Reactions During Coal Conversion', *Energy Fuels*, **4**, 42–54 (1990).
31. F.P. Miknis, 'Solid-state ^{13}C NMR in Oil Shale Research: An Introduction with Selected Applications', in *Composition, Geochemistry, and Conversion of Oil Shales*, NATO ASI Series C, ed. C. Snape, Kluwer Academic, Dordrecht, The Netherlands, 69–91, 1995.
32. J. Mühl, V. Srica, M. Jednacak, 'Determination of Reformed Gasoline Octane Number by NMR Spectrometry', *Fuel*, **68**, 201–203 (1989).
33. R.S. Ozubko, D.M. Clugston, E. Furisky, 'Comparison of Mass Spectrometry and Nuclear Magnetic Resonance Spectrometry for Determination of Hydrocarbon Type', *Anal. Chem.*, **53**, 183–187 (1981).
34. S.P. Srivastava, A.K. Bhatnagar, G.C. Joshi, 'A Procedure for Quantitative ^{13}C-NMR. Application to the Characterization of the Naphtha Fraction of Petroleum Crude', *J. Chem. Tech. Biotechnol.*, **33A**, 361–366 (1983).
35. J. Mühl, V. Srlca, B. Mimica, M. Tomaskovic, 'Characterization of Middle Petroleum Fractions by Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **54**, 1871–1874 (1982).
36. V. Srica, J. Mühl, S. Marin-Mudrovic, 'Characterization of Gas Condensate Fractions by Nuclear Magnetic Resonance Spectrometry', *Fuel*, **67**, 1582–1583 (1988).
37. D.A. Netzel, F.D. Guffey, 'NMR and GC/MS Investigation of the Saturate Fractions from the Cerro Negro Heavy Petroleum Crude', *Energy Fuels*, **3**, 455–460 (1989).
38. D.J. Cookson, B.E. Smith, M. Shaw, 'Determination of Hydrocarbon Compound Class Abundances in Kerosene and Diesel Fuels', *Fuel*, **66**, 758–765 (1987).
39. D.J. Cookson, J.L. Latten, I.M. Shaw, B.E. Smith, 'Property–Composition Relationship for Diesel and Kerosene Fuels', *Fuel*, **64**, 509–519 (1985).
40. T.A. Molak, D.W. Aksnes, M. Stocker, 'Spin-lattice Relaxation Time and Chemical Shift for Structural Determination in Crude Oil and Petroleum Products by Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **56**, 725–728 (1984).
41. J.F. Haw, T.E. Glass, D.W. Hauder, E. Motell, H.C. Dorn, 'Direct Coupling of a Liquid Chromatograph to a Continuous Flow Hydrogen Nuclear Magnetic Resonances Detector for Analysis of Petroleum and Synthetic Fuels', *Anal. Chem.*, **52**, 1135–1140 (1980).
42. V. Ramasevamy, S.L.S. Sarowha, I.D. Singh, 'Structural Characterization of Bombay High Crude Oil Fractions by Nuclear Magnetic Resonance Spectroscopy', *Fuel*, **65**, 1281–1283 (1986).
43. H. Shui, B. Shen, J. Gao, 'Study on the Structure of Heavy Oils', *Petrol. Sci. Technol.*, **15**(7–8), 595–610 (1997).
44. M.U. Hasan, M.F. Ali, A. Bukhari, 'Structural Characterization of Saudi Arabian Heavy Crude Oil by NMR Spectroscopy', *Fuel*, **62**, 518–523 (1983).
45. M. Hajek, G. Sebor, I. Lang, Z. Starcuk, O. Weisser, L. Vodicka, 'Characterization of Heavy Crude Oil Fractions by Means of the ^{13}C-NMR Technique', *Collection Czechslov. Chem. Commun.*, **44**, 1160–1166 (1979).
46. M.U. Hasan, M.F. Ali, M. Arab, 'Structural Characterization of Saudi Arabian Extra Light and Light Crudes by ^{1}H and ^{13}C NMR Spectroscopy', *Fuel*, **68**, 801–803 (1989).
47. B.J. Musser, P.K. Kilpatrick, 'Molecular Characterization of Wax Isolated from a Variety of Crude Oils', *Energy Fuels*, **12**, 715–725 (1998).
48. J.C. Ware, C.M. Schramm, 'NMR Analysis of Heavy Crude Oil Resids', *ACS Preprint, Div. Fuel Chem.*, **30**(4), 104–112 (1985).
49. Y. Takegami, Y. Watanabe, T. Suzuki, T-A. Mitsudo, M. Itoh, 'Structural Investigation on Column Chromatographic Vacuum Residues of Various Petroleum Crudes by ^{13}C Nuclear Magnetic Resonance Spectroscopy', *Fuel*, **59**, 253–259 (1980).
50. S-A. Qian, C-F. Li, P-Z. Zang, 'Study of Structural Parameters on Some Petroleum Aromatic Fractions by ^{1}H NMR/IR and ^{13}C, ^{1}H NMR Spectroscopy', *Fuel*, **63**, 268–273 (1984).
51. D.A. Netzel, 'Low Temperature Studies of Amorphous, Interfacial, and Crystalline Phases in Asphalts Using Solid-state ^{13}C Nuclear Magnetic Resonance', *Transport. Res. Rec.*, **1638**, 23–30 (1998).
52. F.P. Miknis, A.T. Pauli, L.C. Michon, D.A. Netzel, 'NMR Imaging Studies of Asphaltenes Precipitation in Asphalts', *Fuel*, **77**, 399–405 (1998).
53. D.A. Netzel, F.P. Miknis, J.C. Wallace, C.H. Butcher, 'Molecular Motions and Rheological Properties of

Asphalts: An NMR Study', in *Asphalt Science and Technology*, ed. A.M. Usmani, Marcel Dekker, New York, 11–58, 1997.
54. F.C. Stehling, K.W. Bartz, 'Determination of Molecular Structure of Hydrocarbon Olefins by High Resolution Nuclear Magnetic Resonance', *Anal. Chem.*, **38**, 1467–1479 (1966).
55. D.A. Netzel, T.F. Turner, F.P. Miknis, J.C. Wallace, Jr, C.H. Butcher, J.M. Mitzel, R.J. Hurtubise, 'Use of 2,2-Dimethoxypropane and ^{1}H NMR to Distinguish and Quantify the External and Internal Sorbed Water in Coals', *Energy Fuels*, **10**, 371–377 (1996).
56. F. Kasler, *Quantitative Analysis by NMR Spectroscopy*, Academic Press, New York, 1973.
57. M.R. Bendall, D.T. Pegg, 'Complete Accurate Editing of Decoupled ^{13}C Spectra Using DEPT and a Quaternary-only Sequence', *J. Magn. Reson.*, **53**, 272–296 (1983).
58. D.A. Netzel, 'Quantitation of Carbon Types Using DEPT/QUAT NMR Pulse Sequences: Application to Fossil-fuel-derived Oils', *Anal. Chem.*, **59**, 1775–1779 (1987).
59. J.M. Dereppe, C. Moreaux, 'Measurement of CH_n Group Abundances in Fossil Fuel Materials Using DEPT ^{13}C NMR', *Fuel*, **64**, 1174–1176 (1985).
60. D.A. Laude, Jr, C.L. Wilkens, 'Identification of Organic Mixtures Components Without Separation: Quantitative and Edited Carbon-13 Nuclear Magnetic Resonance Spectrometry Data for the Analysis of Petroleum Distillates', *Anal. Chem.*, **58**, 2820–2824 (1986).
61. M.R. Bendall, D.T. Pegg, D.M. Doddrell, S.R. Johns, R.I. Willing, 'Pulse Sequence for the Generation of a ^{13}C Subspectrum of Both Aromatic and Aliphatic Quaternary Carbons', *J. Chem. Soc., Chem. Commun.*, 1138–1140 (1982).
62. M. Bauquet, A. Bailleul, 'Routine Method for Quantitative ^{13}C NMR Spectra Editing and Providing Structural Patterns: Application to Every Kind of Petroleum Fractions Including Residues and Asphaltenes', *Fuel*, **65**, 1240–1246 (1986).
63. D.J. Cookson, B.E. Smith, 'Determination of Carbon C, CH, CH_2 and CH_3 Group Abundances in Liquids Derived from Petroleum and Coal Using Selected Multiplet ^{13}C NMR Spectroscopy', *Fuel*, **62**, 34–37 (1983).
64. D.J. Cookson, B.E. Smith, 'Investigation of Aromatic Carbon Sites in Materials Derived from Petroleum and Coal Using ^{13}C NMR Methods', *Fuel*, **62**, 39–43 (1983).
65. D.J. Cookson, B.E. Smith, 'Quantitative Estimations of CH_n Group Abundances in Fossil Fuel Materials Using ^{13}C NMR Methods', *Fuel*, **62**, 986–988 (1983).
66. A. Bax, L. Lerner, 'Two-dimensional Nuclear Magnetic Resonance Spectroscopy', *Science*, **232**, 960–967 (1986).
67. T. Farrar, 'Two-dimensional NMR Spectrometry', *Anal. Chem.*, **59**, 749A–761A (1987).
68. C.E. Snape, 'Application of NMR to Fossil Fuels: An Assessment of Recent Advances', *Fresenius' Z. Anal. Chem.*, **324**, 781–785 (1986).
69. L.W. Dennis, R.E. Pabst, 'Polarization Transfer and 2-D Experiments in NMR of Solutions of Humic Materials and Fossil Fuel Liquids', in *NMR of Humic Substances and Coal*, eds. R.L. Wershaw, M.A. Mikita, Lewis Publishers, Chelsea, MI, 99–128, 1987.
70. D.J. Cookson, B.E. Smith, '^{1}H and ^{13}C NMR Spectroscopic Methods for the Analysis of Fossil Fuel Materials: Some Novel Approaches', *Fuel*, **61**, 1007–1013 (1982).
71. V. Bansal, G.S. Kapur, A.S. Sarpel, V. Kagdiyal, S.K. Jain, S.P. Srivastava, A.K. Bhatnagan, 'Estimation of Total Aromatics and their Distribution as Mono and Global Di-plus Aromatics in Diesel-range Products by NMR Spectroscopy', *Energy Fuels*, **12**, 1223–1227 (1998).
72. D.J. Cookson, B.E. Smith, 'One- and Two-dimensional NMR Methods for Elucidating Structural Characterization of Aromatic Fractions from Petroleum and Synthetic Fuels', *Energy Fuels*, **1**, 111–120 (1987).
73. D.R. Clutter, L. Petrakis, R.L. Stenger, Jr, R.K. Jensen, 'Nuclear Magnetic Resonance Spectrometry of Petroleum Fractions: Carbon-13 and Proton Nuclear Magnetic Resonance Characterization in Terms of Average Molecular Parameters', *Anal. Chem.*, **44**, 1395–1405 (1972).
74. D.M. Carter, 'Nuclear Magnetic Resonance Spectrometric Determination of Average Molecular Structure Parameters for Coal-derived Liquids', *Anal. Chem.*, **50**, 1185–1187 (1978).
75. E.M. Dickinson, 'Structural Comparison of Petroleum Fractions Using Proton and ^{13}C NMR Spectroscopy', *Fuel*, **59**, 290–294 (1980).
76. P.S. Shenkin, 'Hidden Assumptions in the Average Structure Method', *Liq. Fuels Technol.*, **2**(3), 233–256 (1984).
77. M.W. Boduszynski, 'Limitations of Average Structure Determination for Heavy Ends in Fossil Fuels', *Liq. Fuels Technol.*, **2**(3), 211–232 (1984).
78. B. Rousseau, A.H. Fuchs, 'Determination of Average Molecular Weights of High-boiling Aromatic Oil Fractions by ^{13}C and ^{1}H Nuclear Magnetic Resonance', *Fuel*, **68**, 1158–1161 (1989).
79. V. Leon, 'Average Molecular Weight of Oil Fractions by Nuclear Magnetic Resonance', *Fuel*, **66**, 1445–1446 (1987).
80. D.J. Cookson, B.E. Smith, 'Determination of the Structures and Abundances of Alkanes and Olefins in Fischer–Tropsch Products Using ^{13}C and ^{1}H NMR Methods', *Fuel*, **68**, 776–781 (1989).
81. D.J. Cookson, B.E. Smith, 'Determination of Structural Characterization of Saturates from Diesel and Kerosene Fuels by Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **57**, 864–871 (1985).

82. D.J. Cookson, C.L. Rolls, B.E. Smith, 'Structural Characterization of Branched Plus Cyclic Saturates from Petroleum and Coal Derived Diesel Fuels', *Fuel*, **68**, 788–792 (1989).
83. A.S. Sarpel, G.S. Kapur, V. Bansal, S.K. Jain, S.P. Srivastava, A.K. Bhatnager, 'Direct Estimation of Aromatic Carbon (C_a) Content of Base Oils by ^{1}H-NMR Spectroscopy', *Petrol. Sci. Technol.*, **16**(7–8), 851–868 (1998).
84. O.L. Gülder, B. Glavincevski, 'Prediction of Cetane Number of Diesel Fuels from Carbon Type Structure Compositions Determined by Proton NMR Spectroscopy', *Ind. Eng. Chem. Prod. Res. Dev.*, **25**, 153–156 (1986).
85. A. Adhvaryu, Y.K. Sharma, I.D. Singh, 'Studies on the Oxidation Behavior of Base Oils and Their Chromatographic Fractions', *Fuel*, **78**, 1293–1302 (1999).
86. H. Singh, A. Adhvaryu, I.D. Singh, G.S. Chaudhary, 'NMR Based Characterization of Lubricant Base Oils', *ACS Preprint, Div. Petrol. Chem.*, **42**(1), 255–258 (1997).
87. M.S. Solum, R.J. Pugmire, D.M. Grant, '^{13}C Solid-state NMR of Argonne Premium Coals', *Energy Fuels*, **3**, 187–193 (1989).
88. D.A. Storm, J.C. Edwards, S.J. DeCanio, E.Y. Sheu, 'Molecular Representations of Ratawi and Alaska North Slope Asphaltenes Based on Liquid- and Solid-state NMR', *Energy Fuels*, **8**, 561–566 (1994).
89. D.A. Storm, S.J. DeCanio, J.C. Edwards, E.Y. Sheu, 'Sediment Formation During Heavy Oil Upgrading', *Petrol. Sci. Technol.*, **15**, 77–102 (1997).
90. T.E. Daubert, 'Property Predictions', *Hydrocarbon Processing*, 107–112 (1980).
91. B.J. Schwarz, J.M. Prausnitz, 'A Correlation for Thermodynamic properties of Heavy Fossil-fuel Fractions', *Ind. Eng. Chem. Res.*, **27**, 882–892 (1988).
92. D.J. Cookson, B.E. Smith, 'Calculation of Jet and Diesel Fuel Properties Using ^{13}C NMR Spectroscopy', *Energy Fuels*, **4**, 152–156 (1990).
93. T.T. Lee, D.T. Allen, 'Property Estimation Using Structural Characterizations', *Fuel*, **64**, 1754–1759 (1985).
94. D.T. Allen, 'Structural Characterization and Property Estimation for Complex Mixtures', *Fluid Phase Equilibria*, **30**, 353–366 (1986).
95. D. Finseth, D.L. Cillo, R.F. Sprecher, H.L. Retcofsky, R.G. Lett, 'Changes in Hydrogen Utilization with Temperature During Direct Coal Liquefaction', *Fuel*, **64**, 1718–1722 (1985).
96. S. Beret, J.G. Reynolds, 'Hydrogen Incorporation in Residuum Conversion', *Prepr. Pap.-Am. Chem. Soc. Div. Petrol. Chem.*, **30**(4), 664–671 (1985).
97. S. Beret, J.G. Reynolds, 'Effect of Prehydrogenation on Hydroconversion of Maya Residuum, Part II: Hydrogen Incorporation', *Fuel Sci. Technol. Int.*, **8**, 191–219 (1990).
98. J. Rodriguez, J.W. Tierney, I. Wender, 'Evaluation of a Delayed Coking Process by ^{1}H and ^{13}C NMR Spectroscopy: 1. Material Balances', *Fuel*, **73**, 1863–1869 (1994).
99. J. Rodriguez, J.W. Tierney, I. Wender, 'Evaluation of a Delayed Coking Process by ^{1}H and ^{13}C NMR Spectroscopy: 2. Detailed Interpretation of Liquid NMR Spectra', *Fuel*, **73**, 1870–1875 (1994).
100. M.R. Gray, 'Lumped Kinetics of Structural Groups: Hydrotreating of Heavy Distillate', *Ind. Eng. Chem. Res.*, **29**, 505–512 (1990).

Oil Shale and Shale Oil Analysis

Francis P. Miknis and Daniel A. Netzel
Western Research Institute, Laramie, USA

1 Introduction **1**
2 Oil Shale Assay Methods **2**
2.1 Thermal Methods 2
2.2 Nonthermal Methods 4
3 Combined Use of Nuclear Magnetic Resonance and the Material Balance Fischer Assay to Study Oil Shale Process Chemistry **7**
4 Shale Oil Analysis **9**
4.1 Introduction 9
4.2 Separation Methods 9
4.3 Spectroscopic Methods for Shale Oil Analyses 13
5 Analytical Application of High-resolution Nuclear Magnetic Resonance to the Study of Shale Oil Fractions **14**
5.1 Saturates 15
5.2 Olefins 17
5.3 Aromatics 18
6 Conclusion **20**
Acknowledgment **21**
Abbreviations and Acronyms **21**
Related Articles **21**
References **21**

An oil shale can be defined as a compact rock of sedimentary origin with an ash content of more than 33% and containing organic matter that yields oil when destructively distilled, but not appreciably when extracted with ordinary solvents. This is an operational definition. There is no geological or chemical definition of an oil shale. The term is used mostly in an economic sense so that any shallow rock that yields a commercial amount of oil upon pyrolysis may be considered an oil shale. Oil shale deposits occur in at least 50 countries and the estimated potential world supply of oil from shale is 5×10^{12} barrels.

The organic matter in an oil shale is mostly in the form of kerogen, which is defined as that fraction of the organic matter in a sedimentary rock that is insoluble in common petroleum solvents. The insolubility of kerogen is the main reason why an oil shale must be heated to produce liquid products. The amount of shale oil that can be produced by heating an oil shale is normally obtained from a Fischer assay. However, the Fischer assay is somewhat time-consuming and does not provide information about what chemical properties of oil shales are important for producing liquids. Consequently, a number of alternative procedures have been developed to analyze oil shales. These procedures, along with methods of analyzing shale oils, are described in the following sections.

1 INTRODUCTION

An oil shale can be defined as a compact rock of sedimentary origin with an ash content of more than 33% and containing organic matter that yields oil when destructively distilled, but not appreciably when extracted with ordinary solvents.[1] This is an operational definition. There is no geological or chemical definition of an oil shale. The term is used mostly in an economic sense so that any shallow rock that yields a commercial amount of oil upon pyrolysis may be considered an oil shale. The key words are ''commercial amount'' because petroleum source rocks, which often contain only 1% organic matter, can produce commercial amounts of oil over geologic times. On the contrary, oil shales must have a large enough fraction of organic matter to be of economic interest. The organic matter content of an oil shale should be ~2.5 wt%, just to provide the calorific requirements necessary to heat the rock to 500 °C in order to produce shale oils by thermal decomposition of the organic matter.[2] Below this amount of organic matter, the rock cannot be a source of energy because it takes more energy to heat the rock than can be derived from the produced shale oil. A lower limit of 5% organic matter is sometimes used to define a commercial deposit[2] although such a limit is subject to change depending on prevailing economic conditions. Thus, an oil shale can be a petroleum source rock if subjected to the proper burial conditions over geologic time, but a petroleum source rock at shallow depths is not an oil shale because of the requirement of organic matter richness.

Oil shales occur worldwide, span geologic time from Cambrian to present, and were deposited principally in large freshwater lakes (lacustrine environment), shallow seas and continental shelves (marine environment), and in small lakes, bogs and lagoons associated with coal-producing swamps (paludal environment). Oil shale deposits occur in at least 50 countries and the estimated potential world supply of oil from shale is 5×10^{12} barrels. Shale oil industries in Scotland, Australia, France, Russia and China have been active over the years since about

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1860.[3] In Australia, France and Scotland oil shales have been the source of products similar to those obtained from petroleum. However, the discovery of petroleum in the United States in 1859, and elsewhere soon afterwards, caused the demise of the economic production of shale oil, and the situation remains much the same today. Limited but continued use of oil shale as an energy resource has been made since about 1909 in China and 1916 in Russia. In Brazil, a pilot plant has been in operation since 1982 and has produced over three million barrels of shale oil. Oil shale is being used for some power generation in Israel.

The Green River Formation in Colorado, Utah, and Wyoming, USA has been the most extensively studied oil shale formation in the world. A summary of core data and samples collected by the US Geological Survey has been published.[4] Table 1 summarizes the number of Fischer assays performed on Green River Formation oil shales. During the 1970s, there was significant oil shale activity ongoing in the United States. Tracts of oil shale lands in the Green River Formation were leased for commercial development in 1974 and the decade between 1974 and 1984 represented the greatest activity ever in oil shale research in the US. Partly, this was caused by the oil embargo of 1973, the leasing of the oil shale tracts in Colorado and Utah in 1974, and the creation of the US Department of Energy in 1976. This activity spawned a variety of methods for analyzing oil shale. However, in the year 2000 oil shale activity has all but ceased in the US. Understandably, the plentiful supply and low cost of petroleum has suppressed the commercialization of oil shale and other fossil fuel conversion processes (tar sand processing and coal liquefaction).

The organic matter in an oil shale is mostly in the form of kerogen, which is defined as that fraction of the organic matter in a sedimentary rock that is insoluble in common petroleum solvents. Another fraction, bitumen, is defined as that fraction of the organic matter in a sedimentary rock which is soluble in common petroleum solvents. Kerogen constitutes the major portion of the organic matter (~90% or greater), and its insolubility is one reason why an oil shale must be heated to produce liquid products. Because of kerogen's insolubility, it has been very difficult to obtain information about the chemical structure and composition of kerogen. Compositional data can be acquired after first carrying out laborious and time-consuming procedures to remove minerals and to prepare kerogen concentrates.[5] For an introduction to oil shales and kerogen, the reader is referred to books and reviews that are available.[2,3,5–12]

Table 1 Summary of Fischer assay analyses of the Green River Formation oil shale deposits in Colorado, Wyoming, and Utah (modified from Dyni et al.[4])

	Colorado	Wyoming	Utah	Totals
Number of drill holes analyzed	389	63	153	605
Number of Fischer assays	206 000	13 000	32 000	251 000

2 OIL SHALE ASSAY METHODS

To evaluate the potential of an oil shale formation as a future source of liquid hydrocarbons, two important factors must be considered: (1) how much organic matter is in the formation, (2) how much of the organic matter can be converted to oil. The first factor relates to the total quantity of organic matter in the sediment, while the second factor relates to the quality or type of organic matter.

2.1 Thermal Methods

2.1.1 The Fischer Assay

Probably the single most important item of information about an oil shale is its potential to produce oil during heating. The traditional method for evaluating the potential oil yield of an oil shale is the Fischer assay [ASTM (American Society for Testing and Materials) method D-3904-80].[117] The Fischer assay consists of heating a 100 g sample of −8 mesh (2.38 mm) particle size oil shale to 500 °C at 12 °C min^{-1} and maintaining this temperature for 40 min. During this heating cycle, hydrocarbon vapors are distilled from the rock and condense to form a shale oil. The material is collected, and its volume and weight and specific gravity are recorded. From these measurements, the oil potential is reported in gallons of oil per ton of shale (or liters per tonne, depending upon one's preference for units). Obviously the greater this number, the greater the commercial viability of the formation.

The Fischer assay is strictly a specification test. It does not provide any information about the quality of the organic matter in the oil shale. Nevertheless, it is the benchmark for evaluating an oil shale deposit, and also for determining the efficiencies of oil shale retorting processes. For issues involving exchange or leasing agreements of public oil shale lands in the western United States, the Fischer assay is the only accepted method to determine the value of the lands in question.

Because the Fischer assay does not provide information about what chemical properties of oil shales are important for producing liquids, a number of studies have been made to correlate various chemical and physical property measurements with oil yields determined by the Fischer assay. Even the Fischer assay procedure has been

Table 2 Summary of methods of correlating kerogen properties with oil yields

Method	Measured property	Refs.
Elemental analysis	Organic carbon, total hydrogen	16, 17
Laser pyrolysis	Acetylene yield	18, 19
Thermal analysis	Weight loss and evolved gases upon heating	20, 21
Thermophysical	Thermal conductivity, thermal diffusivity, relative dielectric constant	22–25
Densimetric	Oil shale density	26, 27
FTIR spectroscopy	Aliphatic hydrogen	28, 29
ESR	Unpaired electron concentration	30, 31
NMR pulsed	Total hydrogen	32–35
CP/MAS	Aliphatic/aromatic carbon distribution	28, 36–41

FTIR, Fourier transform infrared; ESR, electron spin resonance; NMR, nuclear magnetic resonance; CP/MAS, cross-polarization/magic-angle spinning.

modified to assay different types of oil shale, or to resemble, more closely, a certain type of retorting process.[13–15] Other reasons for the development of alternate assay methods are that the Fischer assay is fairly time-consuming and the desire to have on-line assaying methods at an oil shale retort.

A summary of some methods for correlating oil shale properties with Fischer assay oil yields is given in Table 2. In general, these methods attempt to measure a parameter that is associated with the quantity of organic matter in the sediment and do not take into account the quality of the organic matter, i.e. its chemical structure. As such, most of them are specific to a given oil shale formation. For example, a correlation derived for oil shales from the Green River Formation in Colorado in the United States is probably not valid for a different oil shale formation, such as, for example, oil shales from the Rundle Deposit in Queensland, Australia. A new correlation must be established for every new deposit.

2.1.2 Elemental Analyses

The elemental composition of an oil shale can be correlated with oil yields determined by the Fischer assay.[16,17] The best correlation for Green River Formation shales is between organic carbon and Fischer assay oil yields. This is not surprising since organic carbon constitutes the major element in the composition of an oil shale. Cook[16] obtained Equations (1) and (2) for oil yield based on weight percent (wt%) or gallons per ton

$$\text{Oil (wt\%)} = 0.832\,(\%\text{orgC}) - 0.251 \qquad (1)$$

$$\text{Oil (gal ton}^{-1}\text{)} = 2.22\,(\%\text{orgC}) - 0.771 \qquad (2)$$

where %orgC is the wt% organic carbon. These correlations were obtained prior to the development of solid-state NMR techniques. However, it is of interest to note that the coefficient of 0.832 in Equation (1) is close to the fraction of aliphatic carbon (0.80) in Green River Formation oil shale determined by ^{13}C-NMR measurements.[36] The aliphatic carbon content for any oil shale has been shown to be a good indicator of oil yield.[28,36–41]

For Devonian (Cleveland member) and Mississippian (Sunbury member) black shales from the eastern United States, organic carbon also correlates with Fischer assay oil yield (Figure 1).[17] However, because of differences in maceral compositions (quality) of the organic matter in different deposits,[42] the data are best fit with separate regressions for deposits from different geologic ages.

If the quality of the organic matter does not change significantly in the formation, then correlations based on organic carbon are applicable, and organic carbon content is a good indicator of the quantity of organic matter in the sediment. As a rule of thumb, the Fischer assay oil yields in gal ton^{-1} for Green River Formation oil shales are about 2–2.5 times %orgC. For Devonian oil shales from the eastern United States, the oil yields in gal ton^{-1} are about

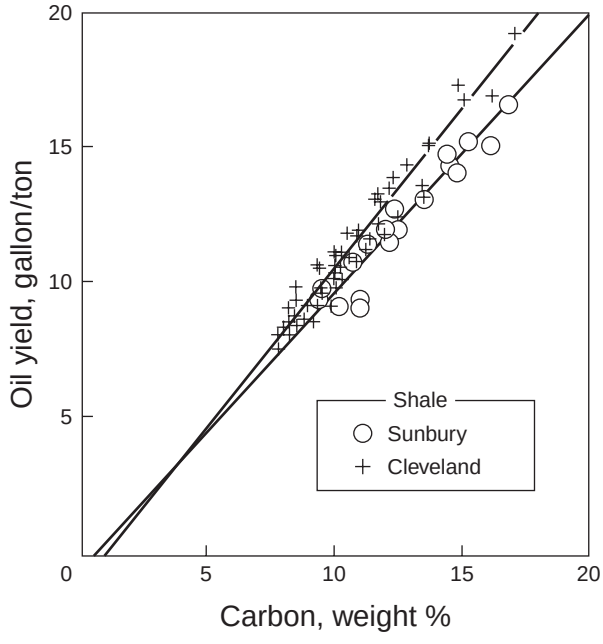

Figure 1 Plot of Fischer assay oil yields and carbon contents for Cleveland and Sunbury oil shales. (Reprinted with permission from T.L. Robl, A.E. Bland, D.W. Koppenaal, L.S. Barron, 'Geochemistry of Oil Shales in Eastern Kentucky', in *Geochemistry and Chemistry of Oil Shales*, eds. F.P. Miknis, J.F. McKay, ACS Symposium Series 230, American Chemical Society, Washington, 159–180, 1983. Copyright 1983 American Chemical Society.)

equivalent to %orgC. For the shales from eastern Kentucky (Figure 1), the Sunbury and Cleveland shales produce 1.03 and 1.18 gal of oil per unit %orgC, respectively.

2.1.3 Laser Pyrolysis

Laser pyrolysis chromatography has also been suggested as a method for analyzing oil shales.[18,19] The method involves using a laser to rapidly pyrolyze the kerogen in an oil shale, followed by detection of the pyrolysis products by gas chromatography (GC). Because of the extremely high temperatures involved during the pyrolysis, acetylene is the major organic species in the product gas mixture. The amount of acetylene produced correlates with the oil yield. With a flame ionization detection (FID) probe only organic species are detected so that the integrated chromatogram is proportional to the amount of organic matter. Some disadvantages of the technique are that the oil shale must be ground, then pelletized before analysis. Because the laser is focused onto a small area, it is important that the pellets be uniform in oil shale composition so that the samples are representative of the deposit. Also, it is important that the amount of laser energy absorbed by the sample be the same for each laser pulse in order to achieve reproducibility.

2.1.4 Thermal Chromatography

A procedure for assaying oil shales based on thermal analytical methods has been suggested.[20,21] This method is a low-temperature method compared with laser pyrolysis. In the thermal chromatographic method, an oil shale is heated ($\sim$30 °C min^{-1}) to a temperature of 600 °C. Hydrocarbon gases that are evolved during heating are measured with a gas chromatographic flame ionization detector. The total amount of evolved gases correlates with the Fischer assay oil yields. A disadvantage of this procedure is that small sample sizes (milligrams) must be used; therefore, sample homogeneity and particle sizes could present problems. An advantage is that the evolved gases can be analyzed by GC to obtain information about the organic matter just as in laser pyrolysis. These methods are similar to the Rock-Eval pyrolysis method used in geochemical analyses for determining kerogen type and maturity.[2] Rock-Eval measurements have been shown to correlate with Fischer assay oil yields.[1]

2.1.5 Microwave and Thermophysical Methods

Microwave radiation has been suggested as an alternative heating method for assaying Colorado oil shales.[22] The dissipation of microwave energy has been shown to be a strong function of the richness of an oil shale. The loss tangent, which is an index of a material's ability to absorb microwaves and which is also related to the dielectric constant of the material, increases nonlinearly as shale richness increases linearly.

Thermal conductivity, thermal diffusivity, dielectric constant, specific heat and sonic velocity measurements have been made as a function of oil shale richness.[23–25] However, none of these measurements appears promising as a method for assaying oil shale. In some cases, the measurements are dominated by the mineralogy. In addition the correlations are nonlinear, as are the microwave correlations, which tends to detract from their usefulness as assay methods.

2.2 Nonthermal Methods

2.2.1 Density Measurements

A simple, but effective, correlation of a physical property measurement with oil yield is the relationship between density and oil yield.[26,27,43] These correlations are based on the simple observation that the oil shales having a high organic content are less dense than oil shales having a low organic content. Thus, by measuring the density of an oil shale, one is essentially measuring the weight fraction of organic matter in an oil shale. The relationship between density and weight percent organic content has been derived by Smith[26] and has the form shown in Equation (3)

$$D_T = \frac{D_A D_B}{A(D_B - D_A) - D_A} \tag{3}$$

where D_T is the absolute density of the oil shale; D_A is the average density of the organic fraction; D_B is the density of the inorganic fraction; and A is the weight fraction of the organic matter. Equation (3) is a general expression and is applicable to any oil shale deposit. For Green River Formation oil shales, D_A is about 1.07, and D_B is about 2.72 g cm^{-3}.[27] An expression relating oil yield in gal ton^{-1} and density for Green River oil shale is shown in Equation (4).

$$\text{Oil (gal ton}^{-1}) = 31.6D_T^2 - 206.0D_T + 326.6 \tag{4}$$

Schultz and Bates[43] have correlated oil yields with oil shale density for oil shales and beneficiated products from oil shales from Alabama, Indiana, Kentucky, Ohio, Michigan and Tennessee. Their correlation between oil yield and densities is shown in Equation (5).

$$\text{Oil (gal ton}^{-1}) = 29.9D_T^2 - 173D_T + 250.3 \tag{5}$$

For New Albany oil shales from Kentucky, oil yields have also been shown to correlate with specific gravity and sound velocity measurements. The latter is somewhat redundant because sound velocity depends on density.[44]

The method relating density to organic content only works well if the mineral and organic components are fairly uniform throughout the deposit, and the percentage conversion of organic matter to oil is uniform. If not, these types of correlations must be developed for each oil shale deposit. These conditions apply to all correlations that measure some chemical or physical property that relates to the quantity, and not to the quality, of the organic matter.

2.2.2 *Spectroscopic Methods*

2.2.2.1 Fourier Transform Infrared FTIR spectroscopic measurements have been applied to obtain information about the aliphatic components in oil shales.[28,29] Spectra are recorded from 4000 to 400 cm^{-1} and corrections are applied for particle scattering and mineral interferences. The CH stretching bands between 3100 and 2750 cm^{-1} are used to obtain a measure of the aliphatic carbon contents, which are then correlated with the Fischer assay oil yields. The FTIR methods have not been widely accepted for studying oil shales. One reason is that, because of the small sample sizes used, obtaining a representative sample is a concern.

2.2.2.2 Electron Spin Resonance Attempts have been made to correlate ESR measurements with oil yields.[30,31] The assumption behind these measurements is that the organic free-radical concentration might correlate with oil yield. Stable organic free radicals are associated with aromatic components, whereas oil yield is more closely associated the hydrogen-rich aliphatic components. Unfortunately unpaired electrons from the minerals also contribute to the ESR signal. Nevertheless, for a set of Colorado oil shales from the same core, the ESR signal gives a fair correlation with oil yields. However, in the case of eastern US oil shales, no correlation between ESR measurements and oil yield was found when samples from different locations were analyzed.[31]

2.2.2.3 Proton (^1H) Nuclear Magnetic Resonance It is reasonable to assume that the amount of hydrogen in an oil shale should be an indicator of the potential oil yield of an oil shale because the amount of available hydrogen is a key factor in producing liquids during pyrolysis. Consequently, correlations have been developed between hydrogen content and Fischer assay oil yields.[32–35] These were based on the use of pulsed NMR methods, in which the signal amplitude measured immediately after a 90° pulse is directly proportional to the amount of hydrogen in the shale. A comparison of a Fischer assay oil yield histogram and a pulsed NMR hydrogen response histogram is shown in Figure 2. The NMR response, which is in arbitrary units, quite faithfully mimics the Fischer assay histogram. Some advantages of pulsed NMR methods are that they are rapid and nondestructive, and can be easily automated. They could be used for rapidly screening an oil shale deposit. Disadvantages are small sample sizes, and interferences due to protons in water and clay minerals. Also, because pulsed NMR provides a measurement of total hydrogen, and not just organic hydrogen, calibrations must be established for each oil shale deposit.

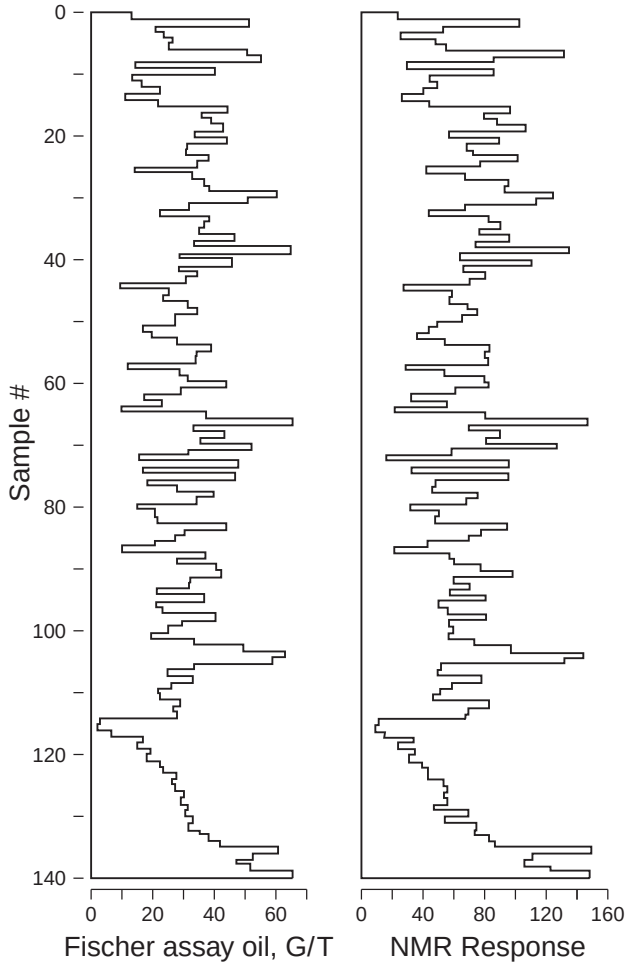

Figure 2 Comparison of depth profile histograms of Fischer assay oil yields and pulsed NMR response for a set of 141 Colorado oil shales. (Reproduced with kind permission from Kluwer Academic Publishers from F.P. Miknis, 'Methods of Oil Shale Analysis', in *Composition, Geochemistry, and Conversion of Oil Shales*, ed. C. Snape, NATO ASI Vol. 455, Kluwer Academic Publishers, The Netherlands, 191–209, Figure 2, 1995.)

2.2.2.4 Carbon (^{13}C) Nuclear Magnetic Resonance Except for the FTIR measurements, the assay methods described in the previous sections have one thing in

common. They all measure, or attempt to measure, some physico-chemical parameter associated with the amount (quantity) of organic matter in an oil shale. While these methods might provide reasonably good correlations with oil yields, they are generally only valid for a specific deposit, and only valid if the type of organic matter, or its composition, does not vary significantly throughout the deposit. None of the methods provides much insight into what chemical structures are important for producing shale liquids during pyrolysis of oil shales.

The importance of kerogen structure in the conversion behavior of oil shales has been known for a long time.[6] However, it was only with the development of solid-state ^{13}C-NMR techniques that direct information about the carbon structure could be obtained. An example of the importance of kerogen chemical structure for producing shale oil is illustrated by the ^{13}C-NMR spectra shown in Figure 3. The total organic carbon is essentially the same for the three oil shales; however, the percentage of conversion of organic carbon to oil during Fischer assay retorting is different for the three shales. The CP/MAS solid-state ^{13}C-NMR spectra show clear differences in the carbon distributions of the oil shales. Most of the organic carbon in the Colorado oil shale is in the form of aliphatic carbon structures (0–50 ppm), whereas a substantial portion of the organic carbon in the Kentucky oil shale is in aromatic carbon structures (100–150 ppm). The oil shale from Morocco has a carbon distribution in between. These data provide a quantitative basis for previous concepts that kerogens with a high H/C ratio (ca. 1.5 or more), so-called Type I kerogens, produce greater quantities of oil during maturation than do Type III kerogens which have an H/C ratio less than unity.[2]

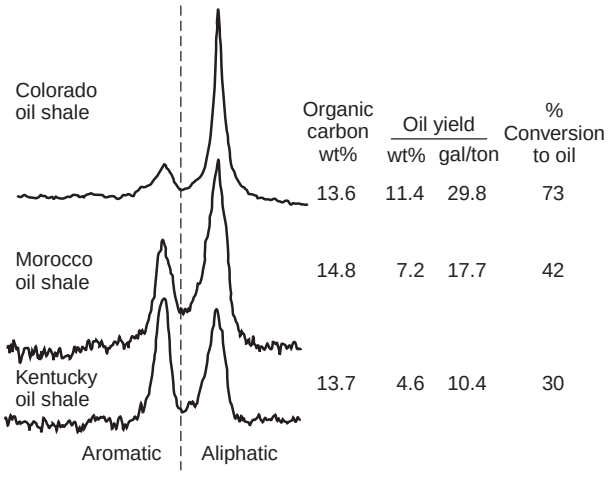

Figure 3 CP/MAS ^{13}C-NMR spectra of oil shales illustrating importance of kerogen structure in conversion to oil.

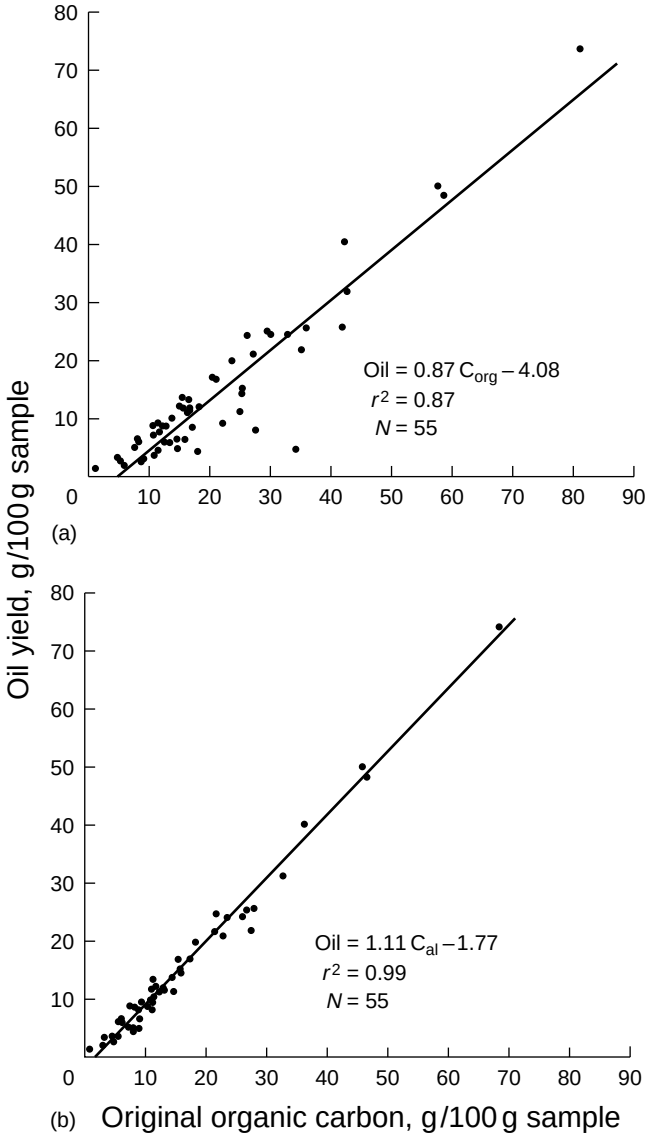

Figure 4 Correlations between aliphatic carbon and total carbon with Fischer assay oil yield. (Reproduced with kind permission from Kluwer Academic Publishers from F.P. Miknis, 'Methods of Oil Shale Analysis', in *Composition, Geochemistry, and Conversion of Oil Shales*, ed. C. Snape, NATO ASI Vol. 455, Kluwer Academic Publishers, The Netherlands, 191–209, Figure 4, 1995.)

By combining the NMR measurements of aliphatic carbon ratios of the kerogen, with organic carbon contents of kerogen, good correlations have been obtained that are independent of depositional environment, geographic location or geologic age.[28,36–41] In Figure 4, correlations are shown between Fischer assay oil yields and total organic carbon and aliphatic carbon for 55 oil shales from different deposits around the world. The data clearly show that chemical structure, i.e. aliphatic carbon, is important for producing shale oils during retorting.

3 COMBINED USE OF NUCLEAR MAGNETIC RESONANCE AND THE MATERIAL BALANCE FISCHER ASSAY TO STUDY OIL SHALE PROCESS CHEMISTRY

The effects of kerogen chemical structure on the conversion behavior of oil shales can be studied by combining solid- and liquid-state NMR measurements with material balance Fischer assay data. In a material balance Fischer assay the elemental composition is determined for the raw shale and the gas, oil and spent shale products. The elemental composition of the gas is calculated from the gas composition which is determined by GC. The assay data can be normalized to the basis of 100 carbon atoms, from which the percentage of carbon in the different products can be calculated. The percentage of organic carbon in the raw shale can be used as an indicator of the quantity of organic matter (kerogen) in raw shale. From these combined measurements, insight into some of the chemistry of kerogen decomposition can be obtained. A knowledge of the chemistry of kerogen decomposition could lead to more efficient retorting processes.

In this section, the effects of carbon structure on the conversion behavior of several oil shales are described for a set of ten oil shales from different deposits around the world.[45] Material balance Fischer assay data are given in Table 3, and normalized NMR conversion data are given in Table 4. During pyrolysis, there is an increase in the amount of aromatic carbons in the products (oil plus residue) over that in the original shale.[46,47] This increase is produced at the expense of aliphatic carbon moieties, which produce aromatic carbon by dehydrogenation of hydroaromatic structures and by ring closure of alkyl groups followed by dehydrogenation.[48] The combined use of solid- and liquid-state ^{13}C-NMR, in conjunction with the material balance Fischer assay conversion data, allows some inferences to be made about the extent of aromatization reactions during heating. The Alaska shale has the greatest loss of aliphatic carbons (22) to aromatization reactions (Table 4). The Australia, Kentucky, Morocco

Table 3 Material balance Fischer assay results for oil shales

Oil shale	Fischer assay (wt%)			Carbon (wt%)				Carbon (conversion %)			Raw shale/ aliphatic carbon (%)
	Oil	Gas	Spent shale	Raw shale	Oil	Gas	Residue	Oil	Gas	Residue	
Alaska	58.00	8.83	26.66	62.7	84.2	55.7	32.5	78.2	7.9	13.9	82
Australia–Stuart	11.65	3.03	75.77	20.3	81.6	41.6	12.1	47.7	6.3	46.0	67
Brazil–Irati	11.14	2.29	83.32	17.5	83.7	36.0	8.6	53.9	4.8	41.3	56
China–Fushun	8.01	2.29	83.98	13.0	83.9	43.5	6.0	52.7	7.7	39.6	56
China–Maoming	7.71	2.70	83.11	15.4	85.4	42.7	8.8	43.8	7.7	48.5	58
Colorado–Anvil Pts	12.19	1.59	84.94	14.4	83.5	46.1	5.3	66.0	4.8	29.2	65
Kentucky–Sunbury	5.31	2.76	88.08	16.2	83.4	36.5	11.8	27.9	6.4	65.7	50
Morocco–Timahdit	4.57	1.88	90.71	8.1	78.7	30.4	4.5	43.8	6.9	49.3	57
Turkey–Göynük	39.16	9.02	41.87	53.4	81.6	38.6	46.7	58.1	6.3	35.6	65
Turkey–Seyitömer	2.41	1.51	88.67	5.6	85.2	40.3	3.0	38.7	11.4	49.9	64

Table 4 Summary of NMR conversion data for 10 selected oil shales

Oil shale	Number of carbons in raw shale		Number of carbons aromatized	Decrease in aliphatic carbons (%)	Increase in aromatic carbons (%)	C_a^{res}/C_a^{raw} [a]	$C_{al}^{res}/C_{al}^{raw}$ [b]
	Aliphatic	Aromatic					
Alaska	82	16	22	26.8	137.5	0.813	0.012
Australia–Stuart	67	30	17	25.4	56.7	1.067	0.209
Brazil–Irati	56	43	8	14.3	18.6	0.744	0.161
China–Fushun	57	40	0	0	0	0.700	0.193
China–Maoming	58	39	12	20.7	30.8	0.974	0.172
Colorado–Anvil Pts	65	34	6	9.2	17.6	0.618	0.123
Kentucky–Sunbury	50	49	21	42.0	42.9	1.143	0.200
Morocco–Timahdit	57	40	18	31.6	45.0	0.975	0.175
Turkey–Göynük	65	32	17	26.2	53.1	0.938	0.092
Turkey–Seyitömer	64	31	18	28.1	58.1	1.161	0.219

[a] Ratio of residue aromatic carbon to raw shale aromatic carbon.
[b] Ratio of residue aliphatic carbon to raw shale aliphatic carbon.

and Turkey shales also show substantial losses of aliphatic carbons to aromatization reactions. When expressed as a percentage of the raw shale aliphatic carbons, a more realistic comparison of the importance of aromatization reactions for the shales is obtained (Table 4). The Kentucky, Morocco and Turkey–Seyitömer oil shales show the greatest percentage decrease in aliphatic carbons. These shales also show the lowest carbon conversions to oil and the highest carbon conversions to residue.

The implication of these observations is that coking (aromatization) reactions are more prevalent in the Kentucky, Morocco and Turkey–Seyitömer shales. However, it cannot always be established to what extent these reactions occur in the solid or liquid state. For the Alaska shale the net effect of the aromatization reactions is to produce oil, because the number of aromatic carbons (16) in the raw shale is not sufficient to account for the number of aromatic carbons (25) in the oil. For the remaining shales, the number of aromatic carbons in the shale oil is substantially less than the number of aromatic carbons in the raw shale. Consequently, the origin of the aromatic carbons in the shale oil is not conclusive. In the case of the Kentucky shale, the number of aromatic carbons (49) in the raw shale is not sufficient to account for the number of aromatic carbons (56) in the residue. Therefore, at least seven of the raw shale aliphatic carbons are converted to aromatic carbons in the residue. This assumes that coking of the shale oil vapor does not occur during Fischer assay.

A plot of the ratio of the number of aromatic carbons in the residue versus the number of aromatic carbons in the raw shale is shown in Figure 5. If this ratio is greater than unity, then the minimum number of aliphatic carbons that have aromatized to form residue can be determined. If the ratio is less than unity, the source of the residue aromatic carbons cannot be determined. Figure 5 shows that the Australia, Kentucky and Turkey–Seyitömer oil shales have ratios greater than unity, suggesting that the aromatization reactions to form coke are most prominent in these shales. The Morocco, China–Maoming and Turkey–Göynük shales have ratios close to unity (>0.93). Except for the Turkey–Göynük shale, all of the shales with ratios greater than or near unity show the lowest aliphatic carbon conversions to oil during Fischer assay. The data in Figure 5 suggest that the shales can be divided into three groups with regard to coking tendency, based on the ratio of residue aromatic carbon to raw shale aromatic carbon. Thus, the coking tendency is Australia, Kentucky, Turkey–Seyitömer > China–Maoming, Morocco, Turkey–Göynük > Alaska, Brazil, China–Fushun, Colorado.

Some of the aliphatic carbons in the kerogen end up as aliphatic carbons in the residue. Although these carbons are not involved in coking reactions, they represent a loss in oil potential. These carbons are most likely short-chain (methyl, ethyl) aliphatic carbons attached to aromatic rings. Figure 6 shows the aliphatic carbon content of the residue, expressed as a percentage of the raw shale aliphatic carbon. The interesting feature of these data is that this percentage is in the 10–20% range for most of the shales, independent of kerogen structure and conversion behavior.

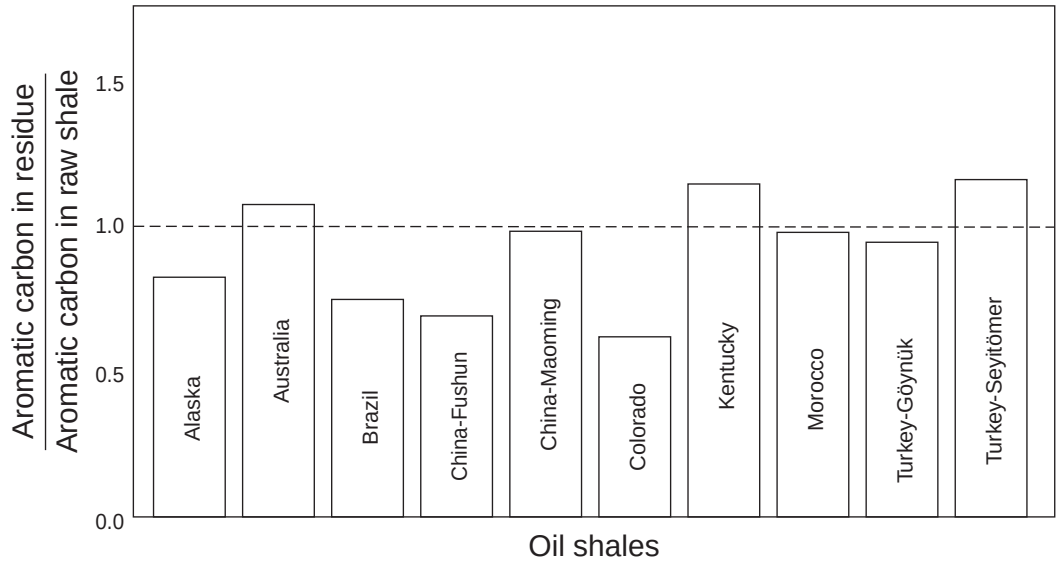

Figure 5 Ratio of aromatic carbon in residue to aromatic carbon in raw shale. (Reproduced with kind permission from Kluwer Academic Publishers from F.P. Miknis, 'Methods of Oil Shale Analysis', in *Composition, Geochemistry, and Conversion of Oil Shales*, ed. C. Snape, NATO ASI Vol. 455, Kluwer Academic Publishers, The Netherlands, 191–209, Figure 5, 1995.)

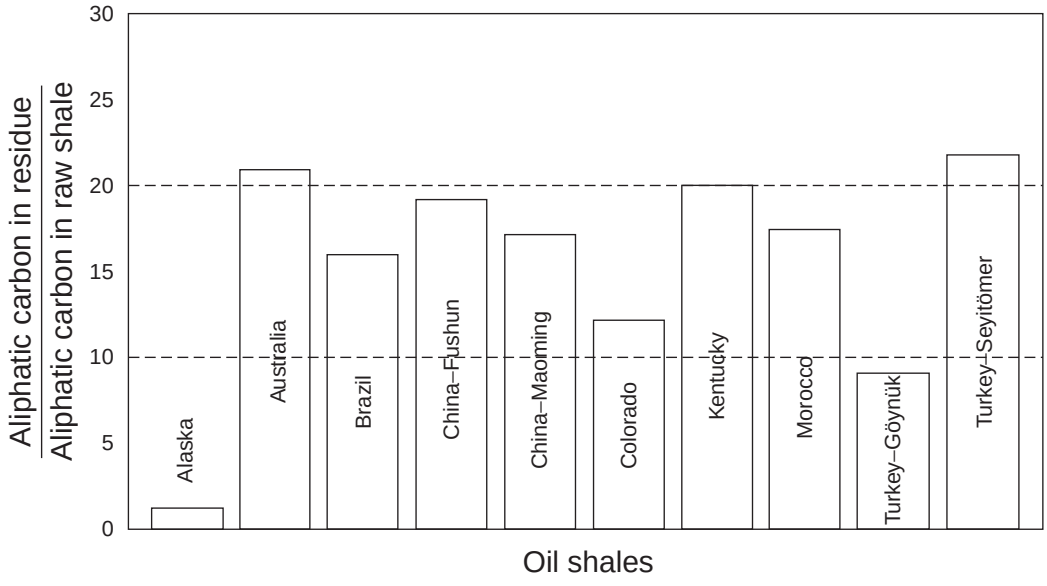

Figure 6 Percentage of raw shale aliphatic carbon in residue. (Reproduced with kind permission from Kluwer Academic Publishers from F.P. Miknis, 'Methods of Oil Shale Analysis', in *Composition, Geochemistry, and Conversion of Oil Shales*, ed. C. Snape, NATO ASI Vol. 455, Kluwer Academic Publishers, The Netherlands, 191–209, Figure 6, 1995.)

4 SHALE OIL ANALYSIS

4.1 Introduction

The physical properties and the chemical compositions of shale oil depend on the source of the raw shale and the type of process used to obtain the oil. Shale oil can be obtained from oil shales by in situ and ex situ (surface) retorting. Laboratory quantities of shale oil for comparative studies and analytical methodology development are often obtained by the Fischer assay method.[49] Chemical analyses of shale oils produced by various retort methods have been used to compare oil shale properties from different locations[50–52] and with depth of the deposit[53,54] and to monitor and/or determine optimum retorting parameters.[55,56] Basically, shale oils differ from petroleum primarily in that they have a higher nitrogen content, and because shale oils are produced by pyrolysis, they contain higher concentrations of olefins, which are generally found in natural crude oils in low concentrations, if at all. Because of these materials, shale oils are relatively thermally unstable,[50] have poor long-term storage stability, and require modifications of the analytical methods used for petroleum characterization.

4.2 Separation Methods

Shale oils as well as other fossil fuels are complex mixtures of hydrocarbons and heteronuclear molecules. Thus, it is necessary to separate them into fractions in order to obtain information on molecular types and to identify individual compounds. The separation techniques most often used are distillation, simulated distillation, and various forms of chromatography – gas, liquid, ion-exchange chromatography (IEC), size-exclusion chromatography (SEC) and high-performance liquid chromatography (HPLC). In addition, shale oils often contain substantial amounts of water that interfere with fractionation of the oil using the various separation techniques and subsequent analyses. Generally, water is removed by distillation using a standard ASTM method.[57]

The elemental composition of the dried oil is determined using commercially available standard combustion instrumentation. Oxygen content is usually obtained by difference. Heating value of the shale oils can be computed from the elemental analysis data and the Boie equation (Equation 6).[58]

$$H = 8400(C) + 27765(H) + 1500(N) + 2500(S) - 2650(O) \quad (6)$$

where H is the gross heating value in cal g^{-1}, and C, H, N, S and O are the normalized weight fractions of these elements in the shale oil. Other properties such as specific gravity, viscosity and pour point are also determined on the raw shale oil using standard ASTM procedures.[59–61] Table 5 lists the elemental composition, carbon and hydrogen aromaticity values, molecular weights and specific gravity for Fischer assay shale oils from different oil shale locations worldwide.

Table 5 Elemental analyses, aromaticities, molecular weights and specific gravities of material balance Fischer assay shale oils

Oil shale	Wt%				Aromaticity		Molecular weight[a]	Specific gravity
	C	H	N	S	Carbon	Hydrogen		
Alaska	84.2	11.9	0.3	0.8	0.319	0.073	310	0.9101
Australia–Stuart	81.6	12.3	1.0	0.7	0.297	0.069	280	0.8902
Brazil–Irati	83.7	11.2	0.9	1.2	0.356	0.080	300	0.9151
China–Fushun	83.9	12.2	2.2	0.7	0.222	0.058	300	0.8873
China–Maoming	85.5	12.2	1.2	0.6	0.301	0.064	300	0.9002
Colorado–Anvil Pts	83.6	12.2	1.9	0.7	0.284	0.049	310	0.9091
Kentucky–Sunbury	83.4	10.0	1.1	2.8	0.495	0.135	280	0.9770
Morocco–Timahdit	78.7	10.0	1.4	7.8	0.422	0.110	270	0.9819
Turkey–Göynük	81.4	11.8	1.3	1.9	0.331	0.081	300	0.9301
Turkey–Seyitömer	85.2	12.4	0.9	0.6	0.346	0.090	280	0.8823

[a] By vapor pressure osmometry in toluene.

4.2.1 Distillation Techniques

4.2.1.1 Hempel Distillation Much of the early work on the characterization of shale oil used the Hempel distillation method[62] to obtain shale oil fractions having boiling point ranges similar to petroleum fractions obtained by distillation. In this method the crude shale oil is distilled into 14 subfractions, each having a boiling range of 25 °C. The first seven subfractions are collected at 760 mmHg (25–200 °C) and the remaining seven subfractions at 40 mmHg (200–350 °C). Specific gravity is measured on all subfractions, including the residue, and the viscosity is measured on fractions 11–14. The residue is analyzed for carbon residue and ash content to give an indication of carbon particulates and mineral matter in the oil.[50,51] Subfractions 1–7 are combined to give the naphtha fraction (25–200 °C), subfractions 8–10 to give the light distillate fractions (106–200 °C at 40 mmHg), and subfractions 11–14 to give the heavy distillate fraction (200–300 °C at 40 mmHg).

Naphtha Fraction. Figure 7 is a schematic diagram of the procedure used to separate further the shale oil naphtha fraction into numerous fractions using an efficient fractionating column for the identification of the compound types and individual compounds.[55,63] Naphtha distillates from crude shale oils produced by several methods of retorting Colorado oil shale were analyzed and found to be remarkably similar. The results indicate that the approximate composition of raw shale oil naphtha is: paraffins and naphthenes 30%; olefins 40%; aromatics 20%; sulfur, nitrogen and oxygen compounds 10%. Paraffins and aliphatic olefins, which comprise about two-thirds of their respective groups, are predominantly straight-chain compounds. The sulfur, nitrogen and oxygen are present principally in the form of thiophenes, pyridines and phenols, respectively.

Light Distillate Fraction. The shale oil distillate fraction that has received the least attention is the light distillate. This is the fraction that boils between 106 and 200 °C at 40 mmHg (equivalent to the 200–300 °C range if it were obtained at 760 mmHg). Analysis of the neutral oil obtained from the light distillate fraction[64] showed the paraffins and naphthenes content to be 10%, olefins 29% and aromatics 61%. The percentages of tar acids and bases were not determined.

Heavy Distillate Fraction. A series of separation techniques, outlined in Figure 8, has been applied to the heavy distillate from Colorado shale oil.[65] The techniques used were adsorption employing several different absorbents, vacuum distillation, thermal diffusion and complex formation. The hydrocarbon content made up approximately half of the heavy distillate (57%). Concentrates obtained from the heavy distillate were those composed of nitrogen compounds (43%) and aromatic, saturated and olefinic hydrocarbons (57%). The concentrates containing the saturates and olefinic materials (35%) were further subdivided into straight- and branched-chain compounds, and the aromatic concentrate (22%) was subdivided into fractions containing different numbers of condensed aromatic rings. For the nitrogen concentrate, separation was made into fractions containing different relative percentages of basic and nonbasic nitrogen.

4.2.1.2 Simulated Distillation Simulated distillation technique is a temperature-programmed gas–liquid chromatographic procedure for determining the distillation profiles of crude oils.[66] The chromatographic data are divided into desired distillate fractions from correlations between GC retention times and normal alkane boiling points. The concentration (wt%) of each distillate fraction is determined using response factors obtained from standard samples. The results

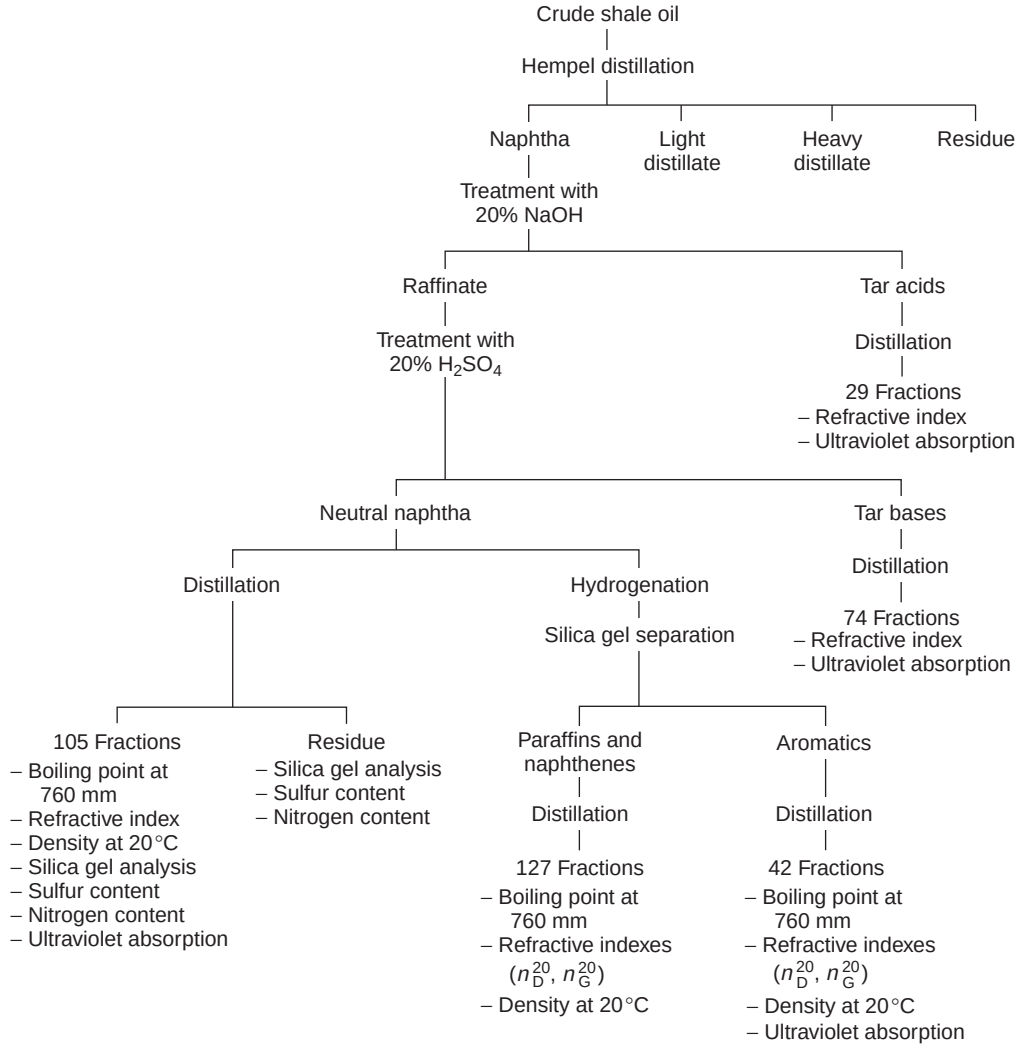

Figure 7 Separation scheme used to analyze the naphtha fraction from shale oil.

are reported as concentration (wt%) of distillable material versus temperature.[67] The boiling ranges of the fractions are chosen to correspond to the true boiling point distillation range of typical crude oil fractions.

The simulated distillation properties of the different Fischer assay shale oils[45] from different oil shales worldwide are given in Table 6. The amount of distillate in each of the fractions does not differ significantly among any of the shale oils, except for possibly the residue. The bulk of the oil (>50%) distills in two boiling ranges, the atmospheric and vacuum gas oil fractions. The temperature at which 50% of the shale oil has distilled off is shown in Figure 9. The average 50% off temperature for all the shale oils is $369 \pm 16\,^{\circ}C$ (95% confidence). This temperature is close to the average temperature of petroleum crude atmospheric gas oil fractions, which is 374 °C.

4.2.2 Chromatographic Techniques

Shale oil and shale oil fractions are extremely complex mixtures and while individual identification of the major components is conceptually and technically possible, the task is exceedingly time-consuming. Thus, characterization of a shale oil or fraction thereof in terms of molecular class (saturates, olefins and aromatics) and average molecular properties of a particular shale oil or fraction is an alternative approach to individual compound identification. There are many chromatographic and extraction techniques which can be used to separate oil into chemical classes of organic compounds.

Adsorption displacement chromatography can be used to separate the shale oil distillates in the 200–325 °C boiling range into saturates, olefins and aromatic fractions.[68] Later papers[65,69] describe the methodology to separate a high-boiling shale oil distillate (gas oil) into

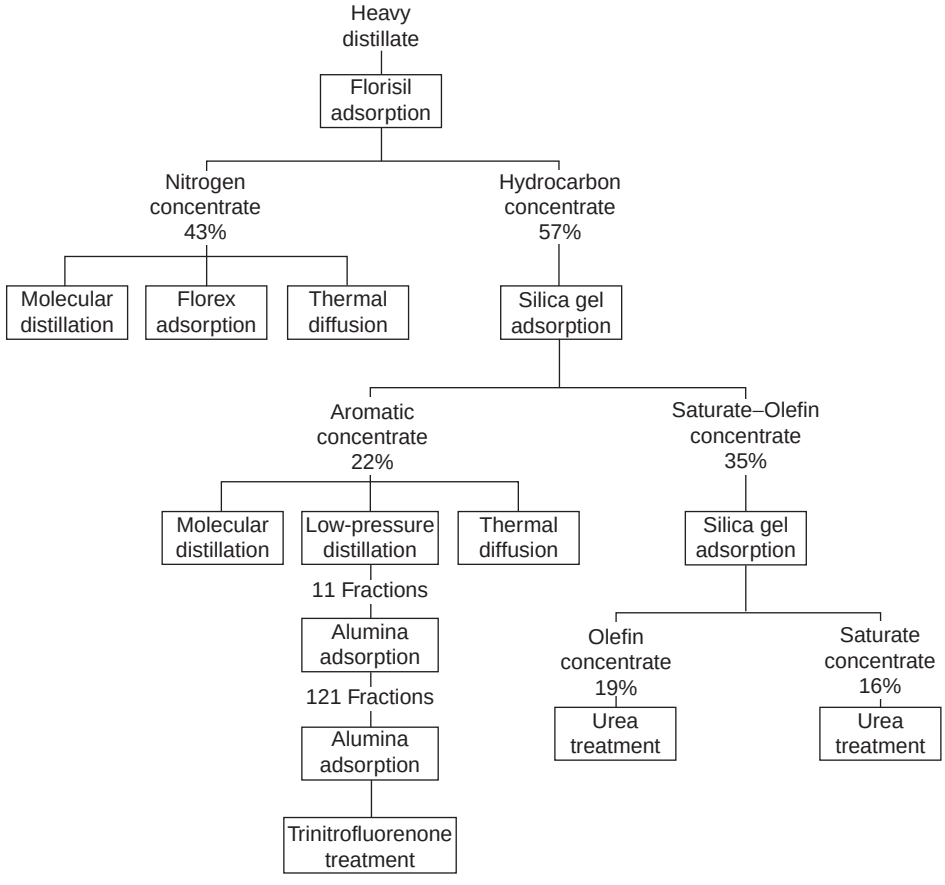

Figure 8 Separation scheme used to analyze the heavy distillate fraction from shale oil.

two concentrates using Florisil as an adsorbent (Figure 8). One concentrate contains most of the nitrogen compounds originally present in the distillate, and the other contains predominantly hydrocarbons. The hydrocarbon concentrate is then further separated into concentrates of aromatics, olefinic and saturated compounds by a two-step process, employing silica gel as the adsorbent. The aromatics are first removed from the saturates and olefins. These latter compounds are then separated from each other, using a much higher adsorbent/sample ratio than is used to remove the aromatics. Further separations of these concentrates are made by vacuum

Table 6 Simulated distillation fractions of material balance Fischer assay shale oils

Fraction	Boiling range (°C)	Wt% of shale oil fraction									
		Alaska	Australia Stuart	Brazil Irati	China Fushun	China Maoming	Colorado Anvil Pts	Kentucky Sunbury	Morocco Timahdit	Turkey Göynük	Turkey Seyitömer
Light straight-run gasoline	32–88	1.1	0.4	0.3	0.5	0.5	0.4	0.8	0.4	0.0	0.5
Naphtha	88–193	9.9	9.3	11.7	7.8	9.7	9.3	12.6	11.4	6.9	14.6
Kerosene	193–271	12.4	15.5	17.1	12.3	15.6	12.7	15.8	15.9	14.2	21.4
Light gas oil	271–321	10.9	12.9	13.1	12.0	14.0	11.2	12.9	13.7	10.8	15.1
Atmospheric gas oil	321–427	27.9	30.0	28.3	29.2	31.6	28.1	28.9	30.6	27.5	29.0
Vacuum gas oil	427–566	28.6	21.6	26.3	25.8	24.2	33.2	21.4	19.9	26.4	18.1
Residue	566+	9.2	10.3	3.2	12.4	4.4	4.1	7.6	8.1	14.2	1.3

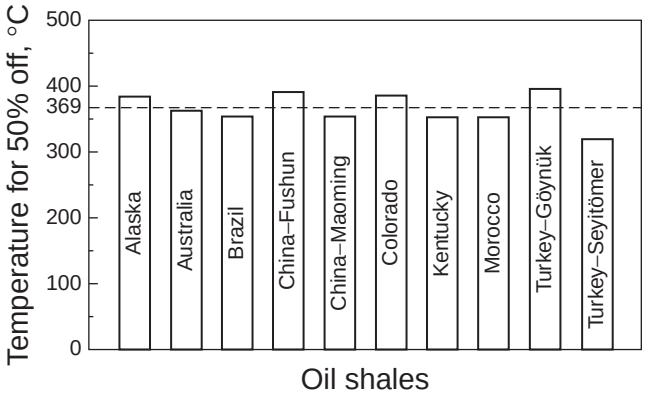

Figure 9 Temperature at which 50 wt% of shale oil distills. (Reproduced with kind permission from Kluwer Academic Publishers from F.P. Miknis, 'Methods of Oil Shale Analysis', in *Composition, Geochemistry, and Conversion of Oil Shales*, ed. C. Snape, NATO ASI Vol. 455, Kluwer Academic Publishers, The Netherlands, 191–209, Figure 7, 1995.)

distillation, thermal diffusion, adduct formation or additional adsorption techniques.

Elution chromatography can be used to separate shale oils into various fractions.[52] Typically, the column is packed with two-thirds silica gel topped with one-third alumina. The linear aliphatic and alicyclic hydrocarbon fraction is first eluted using a light petroleum followed by dichloromethane to elute the aromatic hydrocarbons and thio-aromatics. Methanol is used last to elute the polar nitrogen, sulfur and oxygen fractions. Separation of the aliphatic/alicyclic hydrocarbon fraction into saturates and unsaturates (olefins) is achieved by thin-layer chromatography (TLC) using plates coated with 0.5-mm thick silica gel and containing 10% silver nitrate. Branched and cyclic hydrocarbon fractions are separated from saturated fractions by urea adduction. Oil shales from different locations can be compared using the weight percents calculated for the various fractions.

Hydrogen types and polar materials, including acids and bases in shale oil distillates, can be determined using a combined subtraction and gas chromatographic technique.[70] This method consists of adding an appropriate internal standard to the sample, applying a specific subtraction technique and determining the amount of a compound type from comparison of the before and after gas chromatograms. The precision and detection limits are set by the precision of the gas chromatographic readout at about ±1%. The subtraction technique with GC readout is not useful for trace analyses. However, the data for compound types present in shale oil at the level of a few per cent allow use of the method for more detailed characterization of polar materials.

Brown et al.,[71] using gas–liquid chromatographic separation and spectrometric techniques, have identified 50 individual compounds in the nitrogen base fraction obtained from a hydrocracked shale oil naphtha fraction. They found that alkylpyridines comprised about 64% of the bases and alkylanilines about 33%. The rest of the bases were quinoline and small quantities of cycloalkanopyridines.

Preparative liquid chromatography and glass capillary GC were used to separate shale oil hydrocarbons into fractions of alkanes, alkenes and aromatics.[72–74] In addition, resolution of individual components of branched and cyclic alkanes and alkenes was achieved using a vapor-phase subtraction technique employing 100–200 mesh molecular sieve 5A. A chemically bonded pyrrolidone substrate was used for the HPLC[74] to separate polycyclic aromatic hydrocarbons (PAH) according to the number of aromatic rings and the type of ring condensation.

4.3 Spectroscopic Methods for Shale Oil Analyses

One of the first reported uses of instrumental methods of analysis for crude shale oil fractions was by Ball et al.[55] Using ultraviolet (UV) absorption spectroscopy these authors qualitatively identified numerous aromatic compounds in the naphtha fraction obtained from the distillation of crude shale oil. Dinneen et al.[65] used the technique to investigate high-boiling shale oil distillate. UV spectroscopy was also used in the investigation of the types of sulfur and nitrogen compounds in shale oil.[75–79]

Infrared (IR) spectroscopy was first applied in 1952 to the identification of tar acids[80] isolated from shale oil and later was used to confirm the straight-chain character of the urea–hydrocarbon adduct for saturated and olefinic fractions.[64,65] In 1956 Cook and Church[81,82] used IR spectroscopy to study the pyridine compounds isolated from shale oil tar base fractions. In succeeding years, other nitrogen-type compounds in shale oil also were identified by IR spectroscopy.[56,71,78,79,83] Olefin types found in shale oil were characterized by IR spectroscopy as early as 1960[84] and later by Jackson et al.[85] The polar constituents as well as normal and isoprenoid hydrocarbons isolated from oil shale bitumen and kerogen were studied by Robinson and co-workers.[86–89]

Mass spectrometry (MS) has been used extensively as an analytical tool in the petroleum industry. Thus, it is not surprising that MS has also been applied to shale oils. MS, as an analytical tool, was used by Van Meter et al.[90] to evaluate the Dumas procedure for determining the nitrogen content in shale oils. The first reported use of MS for the identification of hydrocarbons in shale oil was in 1952 by Dinneen et al.[80] A more extensive use of MS for the characterization of saturates and olefins in shale oil was reported by Jensen et al.[91] Normal,[86] isoprenoid[86] and alicyclic[92] hydrocarbons, fatty acids[93] and other polar

constituents,[89,94] steranes[95] and substituted tetralins in oil shale bitumens also have been identified using MS. Using the combined techniques of gas chromatography/mass spectrometry (GC/MS), Morandi and Guffey[96] identified isoprenoids, steranes and terpanes in a Fischer assay shale oil. Ciccioli et al.[97] used the GC/MS technique to study New Albany Shale, and Jones et al.[98] also characterized shale oils and related fuels using the technique.

The application of MS to the analysis of olefins and mixtures of olefins is extremely difficult. However, since shale oil can contain as much as 30% olefinic materials, several studies directed toward the identification of olefins using MS were conducted based upon deuterium labeling[99,100] and derivatization[85] techniques. Earlier MS studies reported direct identification of aliphatic olefins[63] and diolefins.[84]

Morandi and Jensen[76,101] used the mass spectral technique to identify porphyrins in shale oil and oil shale. Other nitrogen compounds were also identified using MS in the raw, refined shale oils,[56,61,77] shale oil naphtha[71] and gas oil fractions.[69] Not only were the nitrogen compounds identified in shale oil by MS, but the sulfur compounds were also identified.[102,103] A compilation of mass spectra of organic sulfur compounds was published by Cook and Dinneen.[104]

Less common analytical techniques were also used to investigate shale oils. Fluorescence spectrometry was used by Hurtubise et al.[105,106] to characterize and identify polynuclear aromatic hydrocarbons in shale oil, and the first systematic study of the free-radical content in raw shale, shale oil and spent shale using electron paramagnetic resonance spectroscopy was reported by Sidwell et al.[30]

^{1}H-NMR techniques have been used extensively to rapidly characterize hydrocarbon types in petroleum. The first reported use of ^{1}H-NMR to study oil shale bitumen was in 1964 by Cummins and Robinson.[86] They used the technique to confirm the methyl branching characteristics of isoprenoid compounds. ^{1}H-NMR was also used to confirm the identity of both isoprenoid compounds and a number of polar constituents isolated from Green River oil shale. The ^{1}H-NMR characterization of saturates and olefins in shale oil was conducted by Jensen et al.[91] and Jackson et al.[85] Several papers were published on the use of ^{1}H-NMR to identify nitrogen compounds in shale oil.[71,77]

As in the case of ^{1}H-NMR, ^{13}C-NMR was first applied to petroleum crude oil investigations and coal liquid fractions. NMR has been used to characterize hydrocarbons in shale oil[107] and to study oil shale in the solid state.[37] Netzel et al.[108] have used both ^{1}H- and ^{13}C-NMR to study the naphtha and light distillate saturated hydrocarbon fraction obtained from in situ shale oil. Both solid-state and high-resolution ^{1}H- and ^{13}C-NMR have been used to compare the hydrocarbon composition of shale oils produced by the pyrolysis and hydropyrolysis of Eastern and Western US oil shales.[109]

5 ANALYTICAL APPLICATION OF HIGH-RESOLUTION NUCLEAR MAGNETIC RESONANCE TO THE STUDY OF SHALE OIL FRACTIONS

NMR spectroscopy is an extremely useful technique for structural characterization of organic compounds and for quantitation of major functional groups in complex mixtures. The advantages of NMR spectroscopy are that it is nondestructive and the spectroscopic data can be reduced relatively easily to give the desired information. The technique is not as sensitive as MS and the application of the NMR to a mixture of similar organic compounds gives only the average structural properties of the molecular composition. As a result, the method is most useful for fractions in a narrow boiling range.

The chemical changes that occur when an oil shale is subjected to various physical and chemical treatments to yield shale oil can be followed using both liquid and solid-state NMR. Figure 10 shows the solid-state CP/MAS ^{13}C-NMR spectrum of a typical Colorado oil shale, and the spectra after the oil shale was (1) treated thermally to yield raw shale oil, (2) extracted with benzene before heating to yield shale oil bitumen, and (3) acid-treated to yield kerogen concentrate. One parameter which can easily be determined from each spectrum is the carbon

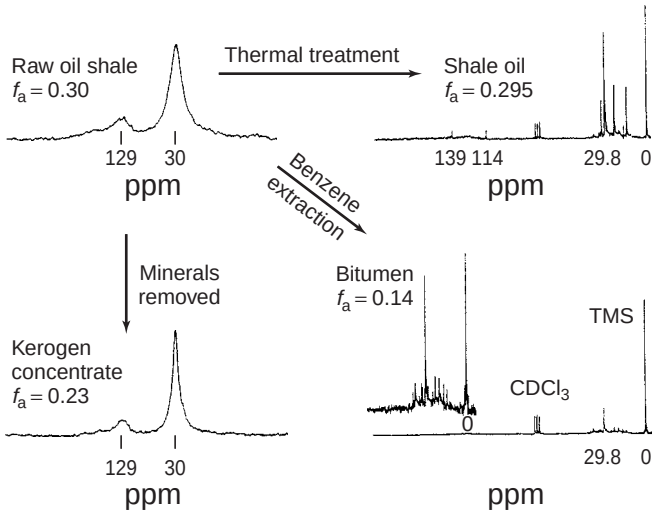

Figure 10 ^{13}C-NMR spectra of Colorado oil shale, kerogen, bitumen and shale oil.

aromaticity, f_a. This parameter is the fraction of aromatic carbons (aromatic integrated area) divided by the total number of carbons (sum of the aromatic and aliphatic integrated areas) measured in the NMR spectrum. The aromaticity values for this oil shale, shale oil, kerogen concentrate and bitumen were found to be 0.30, 0.295, 0.23 and 0.14, respectively.

To obtain additional average molecular structural information on the shale oil by NMR, the Hempel distillation method can be used to fractionate shale oil into naphtha, light distillate, heavy distillate and residue fractions. Figure 11 shows the high-resolution ^{13}C-NMR spectra for the different fractions from a Hempel distillation of an oil shale. Not only can the aromaticity value be measured for each fraction, but also other details about the average aliphatic structure can be obtained. These include the average carbon chain length of the alkane and relative percentages of normal and branched–cyclic alkanes. The naphtha and light distillate fractions can be further subdivided into saturates, olefins and aromatic subfractions by silica gel chromatography. A discussion of the NMR spectra of these subfractions follows.

5.1 Saturates

The ^{1}H-NMR spectra of the saturates from the naphtha and light distillate fractions of shale oil are shown in Figures 12(a) and 12(b), respectively. The resonance signals observed in the aliphatic region of the spectra are due to normal, branched and cyclic alkanes. Spectral assignments of the alkane hydrogens and the quantitative information which can be obtained have been discussed in the literature.[110–112] Table 7 lists the ^{1}H-NMR chemical shift values for the different hydrogen types. The interpretation of the ^{1}H-NMR spectra is straightforward. The resonance signals ranging from 0.5 to 1.0 ppm are due to terminal and branched methyl hydrogens. Methylene hydrogens in straight-chain or branched hydrocarbons have resonances in the range 1.2–1.3 ppm. The signal at 1.27 ppm in each spectrum is thus assigned to the methylene hydrogens. The intensity of this signal is primarily indicative of the alkane chain length. Cyclic methylene hydrogens range from 1.1 to 1.8 ppm. The shoulder at 1.34 ppm on the left of the straight-chain methylene hydrogens is assigned, in part, to these hydrogen types. Because of the overlap of the resonance signals of cyclic

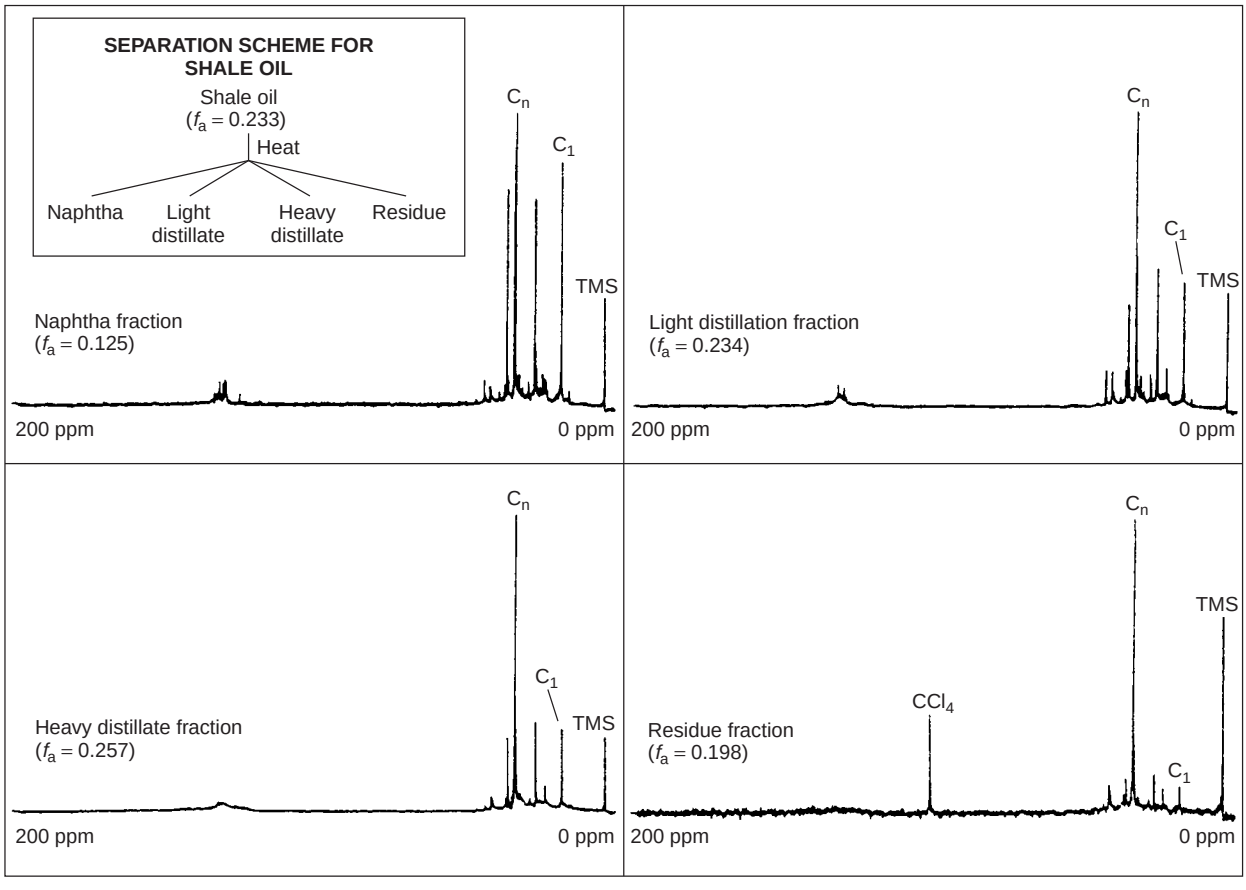

Figure 11 ^{13}C-NMR spectra of fractions obtained from shale oil.

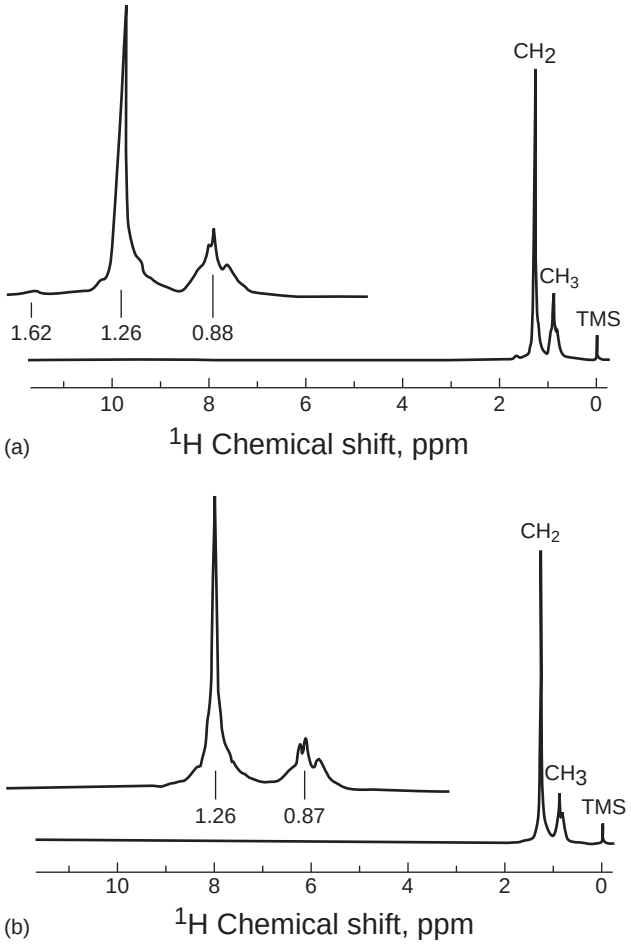

Figure 12 ^{1}H-NMR spectra of the saturate cuts from (a) the naphtha and (b) the light distillate fractions.

Table 7 ^{1}H- and ^{13}C-NMR chemical shift ranges

Definition	Chemical shift range (ppm from TMS[a])
^{1}H	
Monoaromatic hydrogens	6.6–7.2
Diaromatic hydrogens	7.2–7.8
Triaromatic hydrogens	7.8–8.3
Olefinic hydrogens	4.0–6.0
Methylene hydrogens alpha to aromatic ring	2.3–4.0
Methyl hydrogens alpha to aromatic ring	1.9–2.3
Naphthenic hydrogens beta to aromatic ring	1.6–1.9
β-CH_2– and β-CH_3 to aromatic ring and straight-chain alkane methylene hydrogens	1.0–1.6
γ-CH_3 to aromatic ring and straight-chain or branched alkane methyl hydrogens	0.5–1.0
^{13}C	
Total aromatic/olefinic carbons	100–150
Total aliphatic carbons	5–50
Methyl carbons of straight-chain alkanes	13.8–14.2
C_2-methylene carbons of straight-chain alkanes	22.7–23.0
C_3-methylene carbons of straight-chain alkanes	31.8–32.0
$C_{4,5,n}$-methylene carbons of straight-chain alkanes	29–30

[a] TMS, tetramethylsilane.

and straight-chain methylene hydrogens, the amount of quantitative information which can be extracted from these spectra is limited. However, area ratios of the methylene to methyl hydrogens were found to be 2.1 and 2.5 for the naphtha and light distillate saturates, respectively. Thus, semiqualitatively the naphtha saturates have on average shorter alkane chain lengths than the light distillate saturates. No evidence for aromatic (6–9 ppm) or olefinic (4–6 ppm) hydrogens can be found, indicating that the silica gel separation method is quite satisfactory for separating the saturates of the naphtha and light distillate fractions from other hydrogen types.

Figures 13(a) and 13(b) are ^{13}C-NMR spectra of the aliphatic region for the saturate cuts from the naphtha and light distillate fractions, respectively. The relatively few resonances shown in each spectrum suggest that the major constituent carbon types can be identified and quantified. The ^{13}C-NMR chemical shifts for the different carbon types are listed in Table 7. The following discussion refers to the ^{13}C-NMR spectrum of the light distillate saturate fraction (Figure 13b) but is also applicable to the interpretation of the ^{13}C-NMR spectrum of the naphtha saturate fraction. The five intense lines at 14.18, 22.88, 32.16, 29.62 and 29.95 ppm relative to TMS correspond to C-1, C-2, C-3, C-4 and C-5, respectively, of long-chain normal alkanes (see Figure 13b). However, neither the chemical shifts nor intensities correspond to any one *n*-alkane, since the spectrum is a composite of a distribution of several long-chain normal alkanes. The next set of resonances of medium intensity in the ^{13}C-NMR spectrum is due to symmetrically branched linear hydrocarbons having an isoprenoid-like structure. Phytane, pristane and others have all been identified in oil shale.[113] Figure 14(a) shows the identification of the signals with the corresponding carbon atoms in farnesane, a C_{15} isoprenoid.[114] The remaining signals in the ^{13}C-NMR spectrum of the light distillate saturates are due to unsymmetrically branched alkanes and cyclic alkanes. Some representative structures are shown in Figure 14(b).

The relative percentage of the combined branched and cyclic alkanes can be calculated by subtracting the area due to the *n*-alkanes from the total ^{13}C-NMR aliphatic carbon area. The ratio of the area of the branched and

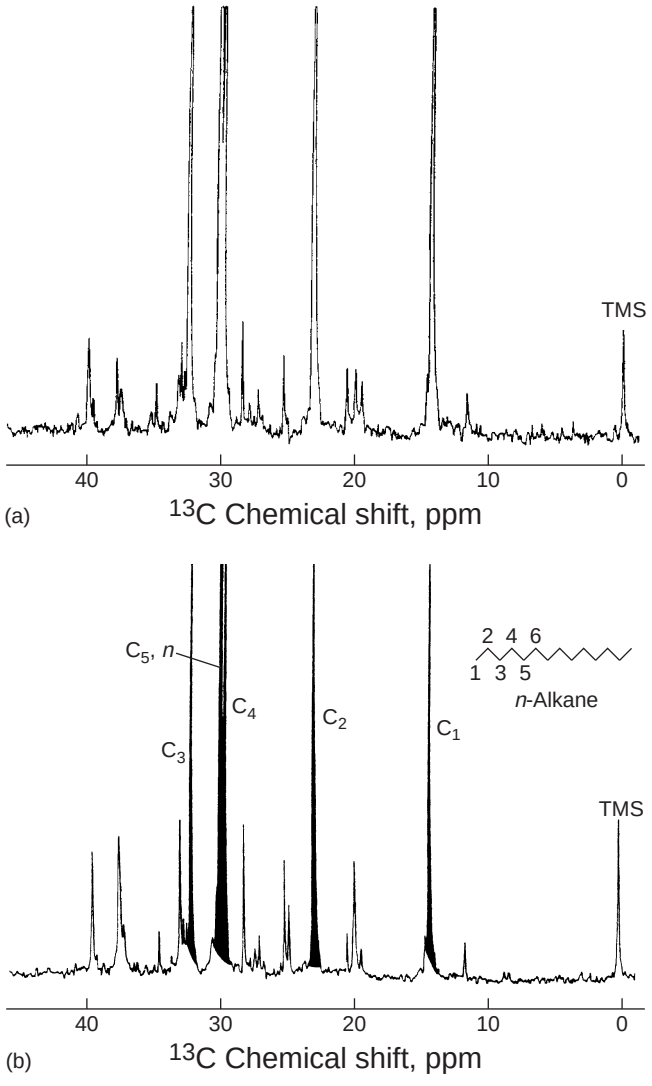

Figure 13 ^{13}C-NMR spectra of the saturate cuts from (a) the naphtha and (b) the light distillate fractions.

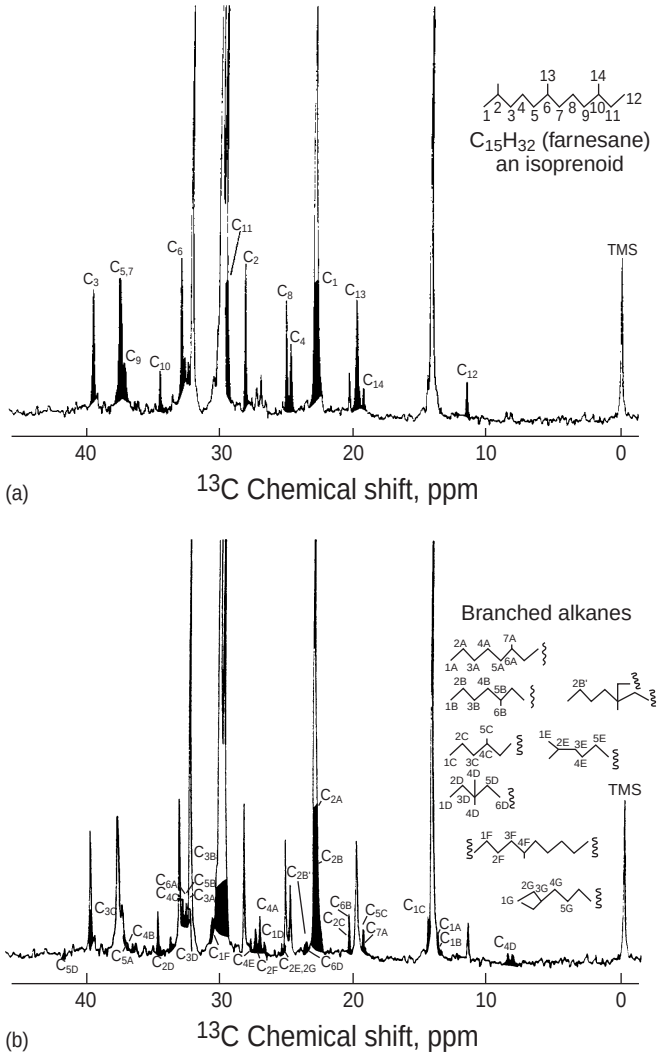

Figure 14 ^{13}C-NMR spectral assignments for the saturate cuts from the light distillate fraction.

cyclic alkane carbons to the total aliphatic carbon area for the light distillate saturate fraction was found to be 0.31. Analytical results from a GC/MS determination on the same sample gave 73% *n*-alkanes, 20% branched alkanes, and 7% unidentified. The sum of the branched alkane and unidentified species content determined by GC/MS (27%) agrees reasonably well with the value found by ^{13}C-NMR (31%). A comparison of the MS and NMR percentage compositions for the saturate cuts from the naphtha and light distillate fractions is listed in Table 8. The data show that the NMR analysis of the saturates can give results comparable to GC/MS analysis.

Table 8 Comparison of the MS and NMR relative percentage compositions of saturate cuts from shale oil

	Naphtha fraction		Light distillate fraction	
	GC/MS	NMR	GC/MS	NMR
n-Alkanes	86	82	73	69
Branched alkanes	5		20	
Cycloalkanes	–	18	–	31
Unidentified	9		7	

5.2 Olefins

^{1}H-NMR can be used to analyze olefins in shale oil fractions. The proton spectra of the olefins obtained from the naphtha and light distillate fractions (Figures 15a and 15b respectively) show the presence of large amounts of methyl and methylene hydrogens and the presence of olefinic hydrogens at about 5.38 ppm relative to TMS. The ^{1}H-NMR olefinic region from 4 to 6 ppm has essentially no interference from hydrogens of other

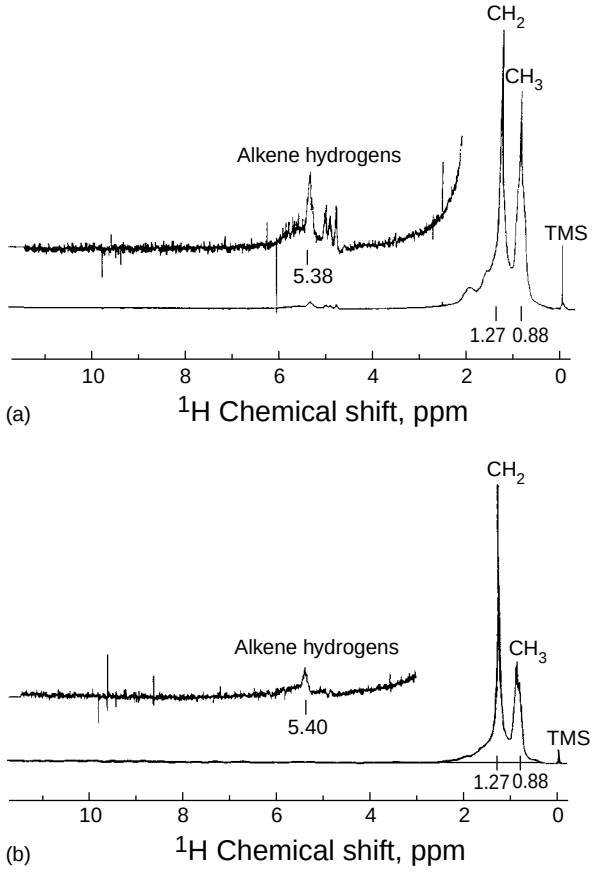

Figure 15 ^{1}H-NMR spectra of the olefin cuts from (a) the naphtha and (b) the light distillate fractions.

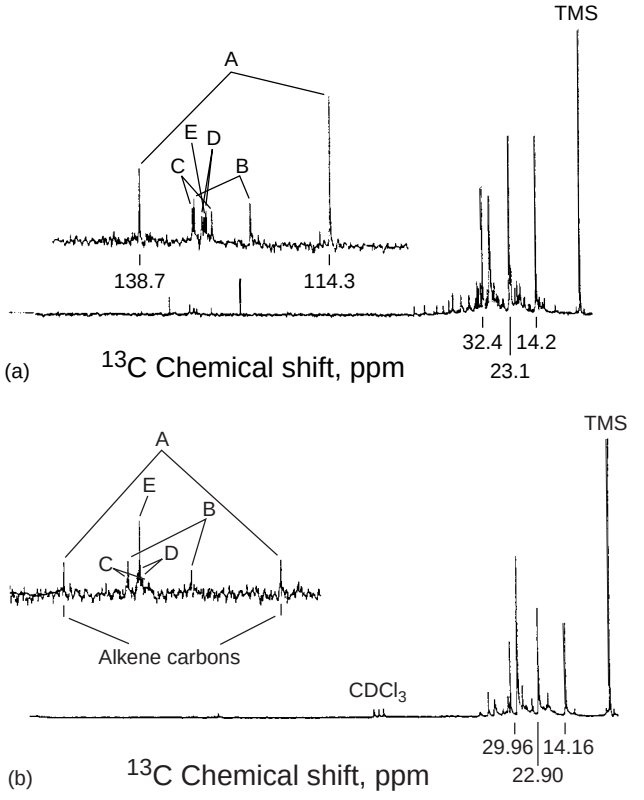

Figure 16 ^{13}C-NMR spectra of the olefin cuts from (a) the naphtha and (b) the light distillate fractions.

types of functional groups. Thus, the olefin content can be measured quantitatively even for fractions with low olefin content. However, the position of the double bond in the alkene molecules cannot always be ascertained from the ^{1}H-NMR spectra but can from ^{13}C-NMR spectra.

The ^{13}C-NMR spectra (Figures 16a and 16b) of the olefin fractions of the shale oil naphtha and light distillation cuts can show clearly the position of the double bond for the different olefins. For the 1-ene olefin the carbons associated with the double bond are at 114 and 138 ppm. As the double bond moves along the carbon chain the differences between the chemical shifts of the two carbon atoms get smaller. For the 5-ene olefin, the carbon atoms associated with the double bond have nearly the same chemical shift. Thus, within the olefinic carbon region (114–138 ppm), it is possible to identify the most dominant double bond positions and determine their corresponding relative percentages. Table 9 lists the positions of the dominant olefinic bonds identified in the olefin subfractions of the naphtha and light distillate fractions. Also included in Table 9 are the observed chemical shifts of the olefinic carbons and the relative percentages. The double bond positions were identified by comparing the observed chemical shifts with those reported by Couperus et al.[115]

5.3 Aromatics

The ^{1}H-NMR spectra of the aromatic naphtha and aromatic light distillate fractions show considerable detail as to the types of protons present. In the naphtha fraction (Figure 17a), the ^{1}H-NMR spectrum shows evidence for the existence of small amounts of di- and triaromatic ring molecules (peaks with ^{1}H-NMR chemical shift values greater than 7.2 ppm). There is no evidence of these molecules in the light distillate fraction (Figure 17b) at the recorded signal-to-noise level. The relative amount of aromatic protons is also less in the light distillate than in the naphtha fraction (8.7% and 23.8%, respectively). In addition, the naphtha fraction has more α-CH_2 (42%) (CH_2 groups attached directly to an aromatic ring) than the light distillate fraction (26.6%). Thus, the ^{1}H-NMR data suggest the naphtha fraction contains mostly substituted monoaromatic compounds with apparently small amounts of di- and triaromatic molecules. The light distillate fraction is, on the other hand, composed mostly of monoaromatics which are less substituted, but the alkyl substituent is of longer carbon chain length.

Table 9 ^{13}C-NMR chemical shifts and relative percentages of normal olefins in naphtha and light distillate fractions obtained from shale oil

Double bond position (*n*-ene)	Naphtha fraction			Light distillate fraction		
	Chemical shift (ppm)		Relative percentage	Chemical shift (ppm)		Relative percentage
	C_n	C_{n+1}		C_n	C_{n+1}	
1-ene	114.34	138.67	36	114.21	138.95	21
2-ene	124.44	131.68	11	124.46	131.73	18
3-ene	131.92	129.44	11	131.91	129.42	–
4-ene	130.16	130.69	42	130.12	130.66	61
5-ene	130.40	130.40		130.41	130.41	

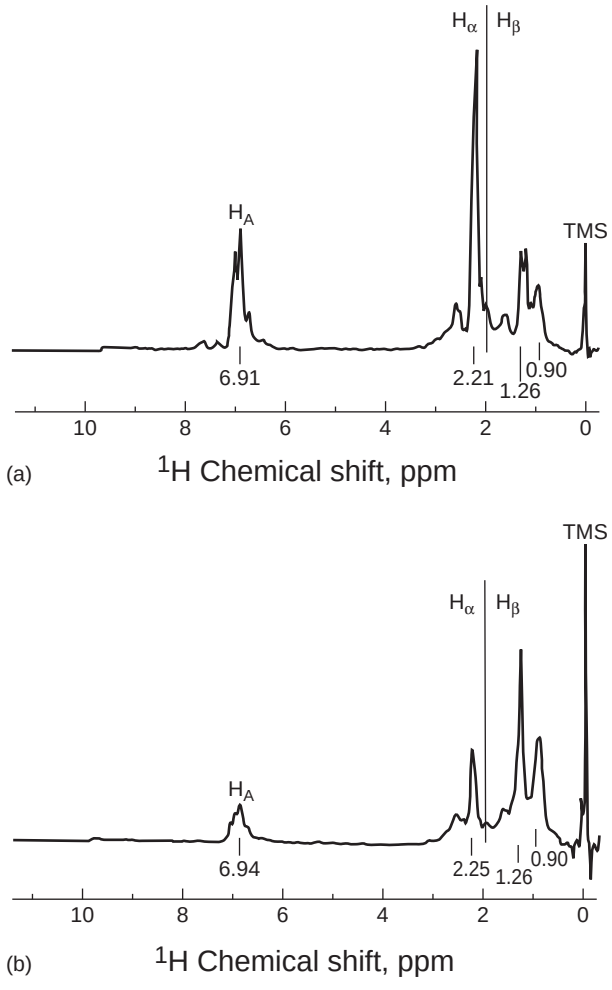

Figure 17 ^{1}H-NMR spectra of the aromatic cuts from (a) the naphtha and (b) the light distillate fractions.

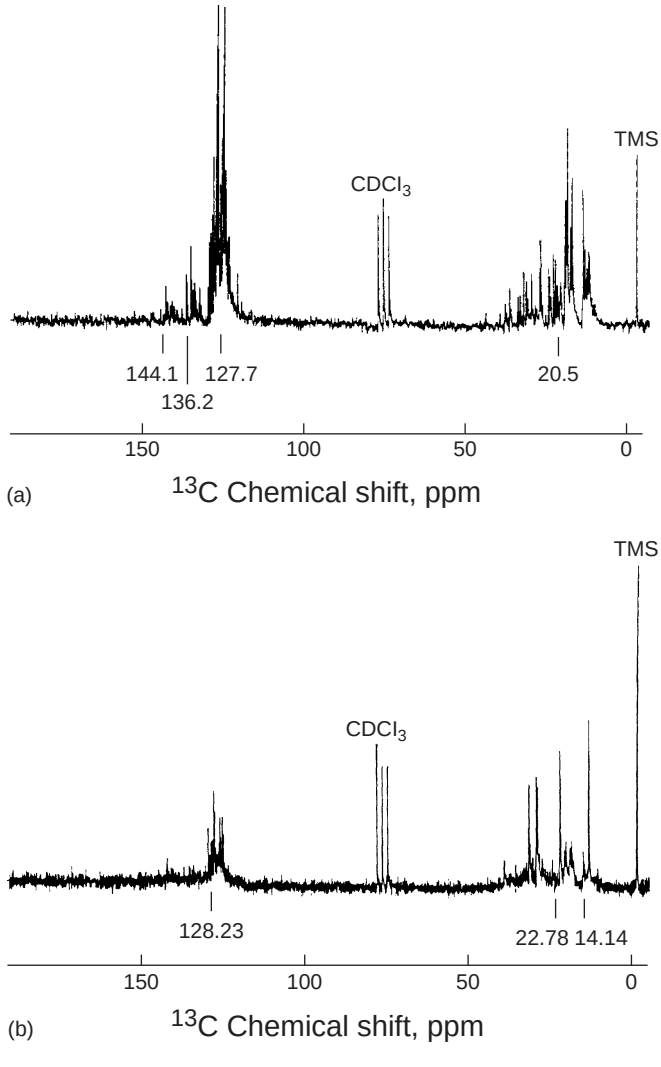

Figure 18 ^{13}C-NMR spectra of the aromatic cuts from (a) the naphtha and (b) the light distillate fractions.

The ^{13}C-NMR spectra of the aromatics for both the naphtha and light distillate fraction (Figures 18a and 18b, respectively) show a large number of narrow and distinct resonances which, in principle, can be interpreted as to specific carbon type present. The naphtha fraction contains more aromatic carbons (greater aromaticity) relative to the light distillate fraction (54% and 32%, respectively). The corresponding relative amounts of

Table 10 Average molecular structural parameters for the aromatic subfraction obtained from the naphtha and light distillate fractions of shale oil

Average molecular structural parameters[116]	Equation[a]	Naphtha fraction	Light distillate fraction
Average number of aromatic carbons per molecule	$\#C_a = C_aMW/12$	8.5	5.3
Average number of saturate carbons per molecule	$\#C_s = A_2CMW/12$	7.2	11.5
Average number of substituents per molecule	$R_s = \%AS\#C_1/100$	4.0	3.3
Average number of carbon atoms per substituent	$n = (H_\alpha + H_\beta)/H_\alpha$	1.8	3.4
Average number of fused aromatic rings per molecule	$R_a = 1 + (\#C_a - \#C_1)/2$	0.6	0.6
Average number of nonbridged aromatic carbons per molecule	$\#C_1 = C_1MW/12$	9.4	6.1
Average number of saturate rings per molecule	$R_N = rR_s$	0.6	−1.0
Average number of saturate rings per substituent	$r = n + 0.5 - 6n/f$	0.2	−0.3
C/H weight ratio in alkyl groups	$f = A_2C/(H_\alpha + H_\beta)H$	5.0	4.9
C/H weight ratio in aromatic groups	$f_1 = A_1C/H_AH$	18.9	23.6
Fraction of carbons which are aromatic	$f_a = A_1/(A_1 + A_2)$	0.5	0.3
Weight fraction of sample which is aromatic carbon	$C_a = f_aC$	0.4	0.3
Weight fraction of sample which is nonbridged aromatic carbons	$C_1 = C_1^s + C_1^u$	0.5	0.3
Proportion of C_1 carbons which are alkyl-substituted	$C_1^s = A_2C/n$	0.2	0.2
Proportion of C_1 carbons which are unsubstituted	$C_1^u = 12H_AH$	0.3	0.1
Percentage substitution of nonbridged aromatics	$\%AS = 100C_1^s/C_1$	42.6	55.3

[a] Input data		Naphtha fraction	Light distillate fraction
A_1, A_2	= Normalized integrals for the aromatic (100–150 ppm) and alkyl (5–50 ppm) regions of the ^{13}C-NMR spectrum	0.540 0.460	0.317 0.684
H_A, H_α, H_β	= Normalized integrals for the aromatic (6.0–9.0 ppm), α-alkyl (2.0–4.0 ppm) and other alkyl (0.3–2.0 ppm) regions of the ^{1}H-NMR spectrum	0.238 0.420 0.341	0.087 0.266 0.647
C	= Weight fraction of carbon	0.792	0.819
H	= Weight fraction of hydrogen	0.091	0.126
MW	= Average molecular weight	238	246

aliphatic carbons were found to be 46% and 68% for the naphtha and light distillate fractions respectively.

A number of average molecular structural parameters[116] can be calculated for the aromatic subfraction of an oil. The underlying assumption is that the aliphatic carbons are attached to the aromatic molecule. Table 10 lists the definitions and equations for the structural parameters as well as the calculated values for the aromatic subfractions obtained from the naphtha and light distillate fractions from the Colorado shale oil. The data indicate that the average number of aromatic carbons per molecule in the naphtha fraction is 8.5 as compared with 5.3 in the light distillate fraction. The average numbers of saturate carbons per molecule are 7.2 and 11.5 and the average numbers of substituents per molecule are found to be 4.0 and 3.3 for the naphtha and light distillate fraction, respectively. The average carbon chain length per substituent for the naphtha and light distillate fractions is 1.8 and 3.4, respectively. From this information and additional data in Table 10, it appears that the naphtha fraction is highly substituted with short alkyl groups. The light distillate aromatic fraction on the other hand appears to have fewer substituents but longer carbon chain lengths per substituent. These conclusions are the same as the conclusions made from the ^{1}H-NMR spectra.

6 CONCLUSION

In 1973 there was a false belief that the world supply of petroleum was being rapidly depleted and by the year 2000 petroleum crude would be completely depleted. At this time a major emphasis was placed on the commercialization of oil shale conversion from which gasoline, diesel and jet fuels can readily be produced. Extensive research was conducted to develop more cost-effective retorting processes and analytical methodology to evaluate oil shale processes and shale oil products. However, since the early 1980s the petroleum supply has increased and the development of an oil shale industry has been limited by the commercial reality that petroleum

can be produced less expensively than shale oil. Until that time when petroleum becomes unavailable or too expensive, there will be little or no incentive to produce synthetic liquid fuels from oil shale on a large scale. How far into the future this will continue is difficult to predict.

Utilization of oil shale to replace petroleum will mean searching for economically efficient and environmentally acceptable recovery processes for the energy-rich organic material locked inside the oil shale's rock matrix. The recovery processes will involve a number of different kinds of technologies, including mining, size reduction, retorting or other means of recovering shale oil from the rock, disposal of the spent shale in an environmentally acceptable manner, and upgrading the shale oil into marketable products. On a commercial scale, to satisfy even a portion of the daily consumption of transportation fuels in the US, the amount of material that has to be mined and disposed of is staggering.

Because oil shales are diverse in their compositions, lithologies and genesis, their behavior during heating will also be different, as has been shown in preceding sections. Ongoing research continues at a low level. However, this has not led to new analytical methods to analyze oil shale and shale oil, so methods described in this article continue to be utilized.

ACKNOWLEDGMENT

The author gratefully acknowledges support of this work by the US Department of Energy under Cooperative Agreement DE-FC21-86 MC11076.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
CP/MAS	Cross-polarization/Magic-angle Spinning
ESR	Electron Spin Resonance
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
IEC	Ion-exchange Chromatography
IR	Infrared
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PAH	Polycyclic Aromatic Hydrocarbons
SEC	Size-exclusion Chromatography
TLC	Thin-layer Chromatography
TMS	Tetramethylsilane
UV	Ultraviolet

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Full Range Crudes, Analytical Methodology of • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Nuclear Magnetic Resonance Characterization of Petroleum

Mass Spectrometry **(Volume 13)**
Electron Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Quadrupole Ion Trap Mass Spectrometer

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses

REFERENCES

1. J.M. Hunt, *Petroleum Geochemistry and Geology*, W.H. Freeman Co., San Francisco, 1979.
2. B.P. Tissot, D.H. Welte, *Petroleum Formation and Occurrence*, 2nd edition, Springer, Berlin, 1984.
3. P.L. Russell, *Oil Shales of the World: Their Occurrence and Exploitation*, Pergamon Press, New York, 1990.
4. J.R. Dyni, F. Gay, T.C. Michalski, 'Oil Shale Data and Drill Cores Collected by the US Geological Survey', ed. J. Gary, Proc. 23rd Oil Shale Symp., Colorado School of Mines, Colorado School of Mines Press, Golden, 10–15, 1990.
5. B. Durand (ed.), *Kerogen*, Editions Technip, Paris, 1980.
6. T.F. Yen, G.V. Chilingarian (eds.), *Oil Shale*, Elsevier, New York, 1976.
7. I.A. Breger, *Organic Geochemistry*, Pergamon Press, London, 1963.
8. G. Eglinton, M.T.J. Murphy, *Organic Geochemistry*, Springer-Verlag, New York, 1969.
9. R. Pelet, 'Oil Shales, What Are They and What Are Their Resources', in *Magnetic Resonance, Introduction, Advanced Methods, and Applications to Fossil Energy*, eds. L. Petrakis, J.P. Fraissard, NATO ASI Series C, D. Reidel, Dordrecht, 219–233, 1984.
10. A.C. Hutton, 'Organic Petrography of Oil Shales', in *Composition, Geochemistry and Conversion of Oil Shales*, ed. C. Snape, NATO ASI Series C, Kluwer Academic Publishers, Dordrecht, 17–33, 1995.

11. F.P. Miknis, J.F. McKay (eds.), *Geochemistry and Chemistry of Oil Shales*, ACS Symposium Series 230, American Chemical Society, Washington, 1983.
12. C. Snape (ed.), *Composition, Geochemistry, and Conversion of Oil Shales*, NATO ASI Kluwer Academic Publishers, Dordrecht, Vol. 455, 1995.
13. C.A. Audeh, 'Improved Oil Shale Retorting', ed. J. Stivers, Proc. Eastern Oil Shale Symp., Lexington, KY, IMMR, Lexington, 263–272, 1990.
14. J.R. Dyni, D.E. Anders, R. Rex, 'Comparison of Hydroretorting, Fischer Assay and Rock-Eval Analyses of Some World Oil Shales', ed. D.J. Lazar, Proc. Eastern Oil Shale Symp., Lexington, KY, IMMR, Lexington, 270–286, 1989.
15. N.C. Watson, 'A Modified Gray–King Assay Method for Small Oil Shale Samples', *Fuel*, **63**, 1455–1458 (1984).
16. E.W. Cook, 'Green River Shale-oil Yields: Correlation with Elemental Analyses', *Fuel*, **53**, 16–20 (1974).
17. T.L. Robl, A.E. Bland, D.W. Koppenaal, L.S. Barron, 'Geochemistry of Oil Shales in Eastern Kentucky', in *Geochemistry and Chemistry of Oil Shales*, eds. F.P. Miknis, J.F. McKay, ACS Symposium Series 230, American Chemical Society, Washington, 159–180, 1983.
18. J.P. Biscar, 'Laser Pyrolysis of Oil Shale', *J. Chromatogr.*, **56**, 348–352 (1971).
19. R.L. Hanson, N.E. Vandenborgh, D.G. Brookins, 'Characterization of Oil Shales by Laser Pyrolysis–Gas Chromatography, *Anal. Chem.*, **47**, 335–338 (1975).
20. P.R. Reed, P.R. Warren, 'Rapid Determination of Recoverable Oil in Oil Shale by Thermal Analysis, *Quart. Colorado Sch. Mines*, **69**, 221–223 (1974).
21. D.R. Johnson, N.B. Young, J.W. Smith, 'Thermal Analysis on Oil Shale: Determination of Potential Oil Yields and Dawsonite, Nahcolite and Nordstrandite Content', Laramie Energy Technology Center Report of Investigations 77/6, Laramie, 1977.
22. A. Judzis, B. Williams, 'Assaying Green River Oil Shale with Microwave Radiation', in *Proc. 10th Oil Shale Symp., Golden*, Colorado Sch. Mines Press, Golden, 207–213, 1977.
23. J. DuBow, R. Nottenburg, K. Rajeshwar, Y. Wang, 'The Effects of Moisture and Organic Content on the Thermophysical Properties of Green River Oil Shale', in *Proc. 11th Oil Shale Symp., Golden*, Colorado Sch. Mines Press, Golden, 350–363, 1978.
24. K. Rajeshwar, J. Du Bow, 'Thermophysical Characterization of Oil Shales–An Overview', in *IGT Symposium on Synthetic Fuels from Oil Shale, Atlanta, GA, December 3–6, 1979*, Institute of Gas Technology Publications, Chicago, 305–321, 1979.
25. K. Rajeshwar, R. Inguva, 'Application of Dielectric Spectroscopy to Chemical Characterization of Oil Shales', *Fuels*, **64**, 931–937 (1985).
26. J.W. Smith, 'Theoretical Relationship Between Density and Oil Yield for Oil Shales', US Bureau of Mines Report Investigations 7248, 1969.
27. J.W. Smith, 'Relationship Between Rock Density and Volume of Organic Matter', Laramie Energy Technology Center Report of Investigations 76/6, Laramie, 1976.
28. E. Evans, B. Batts, N. Cant, 'The Yield of Australian Oil Shales as Determined by Nuclear Magnetic Resonance and Infrared Spectroscopy', *Fuel*, **66**, 326–330 (1987).
29. P.R. Solomon, F.P. Miknis, 'Use of Fourier Transform Infrared Spectroscopy for Determining Oil Shale Properties', *Fuel*, **59**, 893–896 (1980).
30. B.L. Sidwell, L.E. McKinney, M.F. Bozeman, P.C. Egbujor, G.R. Eaton, S.S. Eaton, D.A. Netzel, 'Electron Paramagnetic Resonance Studies of Oil Shale, Shale Oil, and Spent Shale', *Geochim. Cosmochim. Acta*, **43**, 1337–1342 (1979).
31. M. Choudhury, K.F. Rheams, J.W. Harrell, 'Correlation Between Electron Spin Resonance Spectra and Oil Yield in Eastern Oil Shales', *Fuel*, **65**, 1028–1029 (1986).
32. F.P. Miknis, A.W. Decora, G.L. Cook, 'Pulsed Nuclear Magnetic Resonance Studies of Oil Shales–Estimation of Potential Oil Yields', US Bureau of Mines Report of Investigations 7984, US BuMines, Washington, DC, 1974.
33. F.P. Miknis, A.W. Decora, G.L. Cook, 'Pulsed Nuclear Magnetic Resonance Examination of Oil Shales – Estimation of Potential Oil Yields', in *Science and Technology of Oil Shale*, ed. T.F. Yen, Ann Arbor Sci. Publ., Ann Arbor, Chapter 3, 1975.
34. R.D. Sydansk, 'Colorado Green River Formation Oil Shale as Viewed by Pulsed NMR', *Fuel*, **57**, 56–72 (1978).
35. D. Vitorovic, D. Vucelic, M.J. Gasic, N. Juranic, S. Macura, 'Analysis of the Organic Matter of Oil Shales by Nuclear Magnetic Resonance', *Org. Geochem.*, **1**, 89–91 (1978).
36. F.P. Miknis, D.A. Netzel, J.W. Smith, M.A. Mast, G.E. Maciel, '^{13}C NMR Measurements of the Genetic Potentials of Oil Shales', *Geochim. Cosmochim. Acta*, **46**, 977–984 (1982).
37. G.E. Maciel, V.J. Bartuska, F.P. Miknis, 'A Correlation Between Oil Yields of Oil Shales and ^{13}C Nuclear Magnetic Resonance Spectra', *Fuel*, **57**, 505–506 (1978).
38. G.E. Maciel, V.J. Bartuska, F.P. Miknis, 'Improvement in Correlation Between Oil Yields of Oil Shales and ^{13}C NMR Spectra', *Fuel*, **58**, 155–156 (1979).
39. F.P. Miknis, G.E. Maciel, 'Some Examples of ^{13}C NMR Applications to Oil Shales', in *Magnetic Resonance, Introduction, Advanced Methods, and Applications to Fossil Energy*, eds. L. Petrakis, J.P. Fraissard, NATO ASI Series C, D. Reidel, Dordrecht, Netherlands, 545–555, Vol. 124, 1984.
40. F.P. Miknis, P.J. Conn, 'A Common Relation for Correlating Pyrolysis Yields of Coals and Oil Shales', *Fuel*, **65**, 248–250 (1986).

41. E.W. Hagaman, F.M. Schell, D.C. Cronauer, 'Oil Shale Analysis by CP/MAS ^{13}C NMR Spectroscopy', *Fuel*, **63**, 915–919 (1984).
42. T.L. Robl, S.M. Rimmer, L.S. Barron, 'Organic Petrography of Mississippian and Devonian Oil Shales in East-central Kentucky', ed. J. Stivers, Proc. 1990 Eastern Oil Shale Symp., Lexington, KY, IMMR, Lexington, 173–178, 1990.
43. C.W. Schultz, J.B. Bates, 'Specific Gravity-Fisher Assay Relationships Among Eastern Oil Shales', ed. D.J. Lazar, in *Proc. 1989 Eastern Oil Shale Symp., Lexington, KY*, IMMR, Lexington, 533–538, 1989.
44. K.P. Chong, A.I. Leskinen, F.P. Miknis, 'Relationships of Oil Content and Rock Density for New Albany Reference Oil Shale and Sound Velocity Measurements', in *Proc. 1987 Eastern Oil Shale Symp., Lexington, KY*, Kentucky Energy Cabinet Lab. Publisher, Lexington, KY, 119–123, 1987.
45. F.P. Miknis, 'Combined NMR and Fischer Assay Study of Oil Shale Conversion', *Fuel*, **71**, 731–738 (1992).
46. F. Hershkowitz, W.N. Olmstead, R.P. Rhodes, K.D. Rose, 'Molecular Mechanism of Oil Shale Pyrolysis in Nitrogen and Molecular Hydrogen', in *Geochemistry and Chemistry of Oil Shales*, eds. F.P. Miknis, J.F. McKay, ACS Symposium Series 230, American Chemical Society, Washington, DC, 301–316, 1983.
47. A.K. Burnham, J.A. Happe, 'On the Mechanism of Kerogen Pyrolysis', *Fuel*, **63**, 1353–1356 (1984).
48. A.M. Vassallo, M.A. Wilson, J.H. Edwards, ^{13}C NMR Aromaticity Balance on the Products from Flask Pyrolysis of Five Australian Coals', *Fuel*, **66**, 622–626 (1987).
49. A.B. Hubbard, 'Automated Modified Fischer Retorts for Assaying Oil Shale and Bituminous Materials', US Bureau of Mines Report of Investigations 6676, US BuMines, Washington, DC, 1965.
50. R.F. Stevens, G.U. Dinneen, J.S. Ball, 'Analysis of Crude Shale Oil', US Bureau of Mines Report of Investigations 4898, US BuMines, Washington, DC, 1952.
51. C.S. Allbright, R.A. Van Meter, G.U. Dinneen, J.S. Ball, 'Analysis of Crude Shale Oil. 2. Some Brazilian and USA Oils', US Bureau of Mines Report of Investigations 5286, US BuMines, Washington, DC, 1956.
52. P.F.V. Williams, 'The Analysis of Oil Shale and Shale Oil', in *Petroanalysis '81 Advances in Analytical Chemistry in the Petroleum Industry 1975–1982*, ed. G.B. Crump, John Wiley & Sons, New York, Chapter 31, 1982.
53. L.P. Jackson, J.R. Morandi, R.E. Poulson, 'Compositional Variation of Retorted Shale Oils with Stratigraphy: Wyoming Core, Northern Green River Basin', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **22**(3), 66–67 (1977).
54. W.E. Robinson, J.J. Cummings, G.U. Dinneen, 'Changes in Green River Oil-shale Paraffins with Depth', *Geochim. Cosmochim. Acta*, **29**(4), 249–258 (1965).
55. J.S. Ball, G.U. Dinneen, J.R. Smith, C.W. Bailey, R.A. Van Meter, 'Composition of Colorado Shale-oil Naphtha', *Ind. Eng. Chem.*, **41**, 581–587 (1949).
56. R.E. Poulson, 'Nitrogen and Sulfur in Raw and Refined Shale Oils', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **20**(2), 183–197 (1975).
57. ASTM Standards, D95-70, ASTM, Philadelphia, Parts 15, 22, 23, 27, 29, 1982.
58. S. Ringen, J. Lanum, F.P. Miknis, 'Calculating Heating Values from Elemental Compositions of Fossil Fuels', *Fuel*, **58**, 69–71 (1979).
59. ASTM, D70-76, ASTM, Philadelphia, Parts 15, 19, 1982.
60. ASTM, D445-79, ASTM, Philadelphia, Parts 23, 40, 1982.
61. ASTM, D97-66, ASTM, Philadelphia, Parts 23, 40, 1982.
62. N.A.C. Smith, H.M. Smith, O.C. Blade, E.L. Garton, 'The Bureau of Mines Routine Method for the Analysis of Crude Petroleum. I. The Analytical Method', US Bureau of Mines Bulletin 490, US BuMines, Washington, DC, 1951.
63. G.U. Dinneen, R.A. Van Meter, J.R. Smith, C.W. Bailey, G.L. Cook, C.S. Allbright, J.S. Ball, 'Composition of Shale-oil Naphtha', US Bureau of Mines Bulletin 593, US BuMines, Washington, DC, 1961.
64. H.M. Thorne, J.S. Ball, 'The Composition of Shale Oils', in *The Chemistry of Petroleum Hydrocarbons*, eds. B.T. Brooks, S.S. Kurtz, Jr, C.E. Boord, L. Schmerling, Reinhold Publishing Corp., New York, Vol. 1, Chapter 5, 1955.
65. G.U. Dinneen, J.R. Smith, R.A. Van Meter, C.S. Allbright, W.R. Anthoney, 'Application of Separation Techniques to a High-boiling Shale-oil Distillate', *Anal. Chem.*, **27**, 185–190 (1955).
66. H.B. Jensen, R.E. Poulson, G.L. Cook, 'Characterization of a Shale Oil Produced by In-situ Retorting', *Prep. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **15**, 113–121 (1971).
67. J.H. Gary, G.E. Handwerk, *Petroleum Refining*, Marcel Dekker, New York, 32, 1975.
68. G.U. Dinneen, C.J. Thompson, J.R. Smith, J.S. Ball, 'Adsorption Analysis by Displacement Techniques', *Anal. Chem.*, **22**, 871–876 (1950).
69. G.U. Dinneen, G.L. Cook, H.B. Jensen, 'Estimate of Types of Nitrogen Compounds in Shale-oil Gas Oil', *Anal. Chem.*, **30**(12), 2026–2030 (1958).
70. R.E. Poulson, H.B. Jensen, J.J. Duvall, F.L. Harris, J.R. Morandi, 'Determination of Compound Types in Complex Mixtures Using Subtraction and Gas Chromatographic Readout', in *Analysis Instrumentation*, Proc. Instrument Society of America, Pittsburgh, 193, Vol. 10, 1972.
71. D. Brown, D.G. Earnshaw, F.R. McDonald, H.B. Jensen, 'Gas–Liquid Chromatographic Separation and Spectrometric Identification of Nitrogen Bases in

Hydrocracked Shale-oil Naphtha', *Anal. Chem.*, **42**(2), 146–150 (1970).
72. F.P. DiSanzo, P.C. Uden, S. Siggia, 'Shale Oil Hydrocarbon Separation by Preparative Liquid Chromatography and Glass Capillary Gas Chromatography', *Anal. Chem.*, **52**, 906–909 (1980).
73. J.F. McKay, D.R. Latham, 'High Performance Liquid Chromatographic Separation of Olefin, Saturate and Aromatic Hydrocarbons in High-boiling Distillates and Residues of Shale Oil', *Anal. Chem.*, **52**, 1618–1621 (1980).
74. T.H. Maurey, S. Siggia, P.C. Uden, R.J. Crowley, 'High Performance Liquid Chromatographic Separation of Polycyclic Aromatic Hydrocarbons on Microparticulate Pyrrolidone and Application to the Analysis of Shale Oil', *Anal. Chem.*, **52**, 885–891 (1980).
75. G.U. Dinneen, 'Sulfur and Nitrogen Compounds in Shale Oil', *Proc. Am. Petrol. Inst.*, **42**(VIII), 41–44 (1962).
76. J.R. Morandi, H.B. Jensen, 'Isolation of Porphyrins from Shale Oil and Oil Shale', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **8**(1), 293–303 (1964).
77. G.L. Cook, 'Nitrogen Compounds in Colorado Shale Oils', *Prepr. Pap. – Am. Chem. Soc., Div. Petrol. Chem.*, **10**(2), C35–C38 (1965).
78. W.E. Haines, G.L. Cook, G.U. Dinneen, 'Techniques for Separating and Identifying Nitrogen Compounds in Petroleum and Shale Oil', in *Proc. 7th World Petroleum Congress, Petroleum Industry: General Problems*, 83–92, Vol. 9, 1967.
79. R.E. Poulson, H.B. Jensen, G.L. Cook, 'Nitrogen Bases in a Shale-oil Light Distillate', *Prepr. Pap. – Am. Chem. Soc., Div. Petrol. Chem.*, **16**(1), A49–A61 (1971).
80. G.U. Dinneen, J.R. Smith, C.W. Bailey, 'High-temperature Shale Oil, Product Composition', *Ind. Eng. Chem.*, **44**, 2647–2650 (1952).
81. G.L. Cook, F.M. Church, 'Determination of Pyridines by Infrared Spectroscopy', *Anal. Chem.*, **28**, 993–995 (1956).
82. G.L. Cook, F.M. Church, 'Correlations of the Infrared Spectra of Some Pyridines', *J. Phys. Chem.*, **61**, 458–462 (1957).
83. J.R. Morandi, R.E. Poulson, 'Nitrogen Types in Light Distillates from Aboveground and In Situ Combustion Produced Shale Oil', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **20**(2), 162–174 (1975).
84. H.H. Heady, L.G. Adams, G.U. Dinneen, 'Composition of a Shale-oil Naphtha', US Bureau of Mines Report of Investigations 5662, US BuMines, Washington, DC, October 1960.
85. L.P. Jackson, C.S. Allbright, R.E. Poulson, 'Olefin Analysis in Shale Oils', in *Analytical Chemistry of Liquid Fuel Sources*, eds. P.C. Uden, S. Siggia, H.B. Jensen, Advances in Chemistry Series 170, American Chemical Society, Washington, 232–242, 1978.
86. J.J. Cummins, W.E. Robinson, 'Normal and Isoprenoid Hydrocarbons Isolated from Oil-shale Bitumen', *J. Chem. Eng. Data*, **9**, 304–307 (1964).
87. J.I. Fester, W.E. Robinson, 'Oxygen Functional Groups in Green River Oil-shale Kerogen and Trona Acids', in *Coal Science*, ed. R.F. Gould, Advances in Chemistry Series 55, American Chemical Society, Washington, DC, 22–31, 1966.
88. W.E. Robinson, G.U. Dinneen, 'Constitutional Aspects of Oil-shale Kerogen', in *Proc. 7th World Petrol. Congress, Mexico City, Mexico, Drilling and Production*, 669–680, Vol. 3, 1967.
89. D.E. Anders, F.G. Doolittle, W.E. Robinson, 'Polar Constituents Isolated from Green River Oil Shale', *Geochim. Cosmochim. Acta*, **39**, 1423–1430 (1975).
90 R. Van Meter, C.W. Bailey, E.C. Brodie, 'Evaluation of Dumas Method by Mass Spectrometry', *Anal. Chem.*, **23**, 1638–1639 (1951).
91. H.B. Jensen, J.R. Morandi, G.L. Cook, 'Characterization of the Saturates and Olefins in Shale-oil Gas Oil', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **12**(1), F93–F110 (1968).
92. D.E. Anders, W.E. Robinson, 'Cycloalkane Constituents of the Bitumen from Green River Shale', *Geochim. Cosmochim. Acta*, **35**, 661–678 (1971).
93. D.L. Lawlor, W.E. Robinson, 'Fatty Acids in Green River Formation Oil Shale', *Prepr. Pap. – Am. Chem. Soc., Div. Petrol. Chem.*, **10**(1), 5–7 (1965).
94. T. Aczel (ed.), 'Mass Spectrometric Characterization of Shale Oil', ASTM Special Technical Publication No. 902, ASTM, Philadelphia, 1986.
95. P.C. Anderson, P.M. Gardner, E.V. Whitehead, D.E. Anders, W.E. Robinson, 'The Isolation of Steranes from Green River Oil Shale', *Geochim. Cosmochim. Acta*, **33**, 1304–1307 (1969).
96. J.R. Morandi, F. Guffey, 'Identification of Isorprenoids, Steranes, and Terpanes in Fischer Assay Retorted Shale Oils', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **23**(1), 228–239 (1978).
97. P. Ciccioli, J.M. Hayes, G. Rinaldi, K.B. Denson, W.G. Meinschein, 'Graphitized Carbon in Gas–Liquid–Solid Chromatography and Gas Chromatography/Mass Spectrometric Analysis of High Boiling Hydrocarbon Mixtures', *Anal. Chem.*, **51**(3), 400–408 (1979).
98. P.W. Jones, R.J. Jackson, P.E. Strup, A.P. Graffeo, 'Chemical Characterization of Shale Oil and Related Fuels', *Prepr. Pap. – Am. Chem. Soc., Div. Fuel Chem.*, **21**(6), 258–262 (1976).
99. J.R. Morandi, H.B. Jensen, 'Homogeneous Catalytic Deuteration of Olefinic Double Bonds', *J. Org. Chem.*, **34**, 1889–1891 (1969).
100. D.G. Earnshaw, F.G. Doolittle, A.W. Decora, 'Some Mass Spectral and Analytical Aspects of Deuterium Labeling of Straight-chain Olefins', *Org. Mass Spectrom.*, **5**, 801–815 (1971).

101. J.R. Morandi, H.B. Jensen, 'Comparison of Porphyrins from Shale Oil, Oil Shale, and Petroleum by Absorption and Mass Spectroscopy', *J. Chem. Eng. Data*, **11**(1), 81–88 (1966).
102. I.W. Kinney, Jr, G.L. Cook, 'Identification of Thiophene and Benzene Homologs, Mass Spectral Correlations', *Anal. Chem.*, **24**, 1391–1396 (1952); Correction: *Anal. Chem.*, **25**, 1729 (1953).
103. I.W. Kinney, Jr, J.R. Smith, J.S. Ball, 'Thiophenes in Shale-oil Naphtha', *Anal. Chem.*, **24**, 1749–1754 (1952).
104. G.L. Cook, G.U. Dinneen, 'Mass Spectra of Organic Sulfur Compounds', US Bureau of Mines Report of Investigations 6698, US BuMines, Washington, DC, 1965.
105. R.J. Hurtubise, J.F. Schabron, J.D. Feaster, D.H. Therkildsen, R.E. Poulson, 'Fluorescence Characterization and Identification of Polynuclear Aromatic Hydrocarbons in Shale Oil', *Anal. Chim. Acta*, **89**, 377–382 (1977).
106. R.J. Hurtubise, G.T. Skar, R.E. Poulson, 'Determination of Benzo(*a*)pyrene in Shale Oil by Solid-surface Fluorescences', *Anal. Chim. Acta*, **97**, 13–19 (1978).
107. T.T. Coburn, J.H. Campbell, 'Oil Shale Retorting: Part 2 Variation in Product Oil Chemistry During Retorting of an Oil Shale Block', UCRL-52256 Part 2, Lawrence Livermore Laboratory, Livermore, September 1977.
108. D.A. Netzel, D.R. McKay, R.A. Heppner, F.D. Guffey, S.D. Cooke, D.L. Varie, D.E. Linn, '^{1}H and ^{13}C NMR Studies on Naphtha and Light Distillate Saturates Hydrocarbon Fractions Obtained from In-situ Shale Oil', *Fuel*, **60**, 307–320 (1981).
109. D.A. Netzel, F.P. Miknis, 'NMR Study of U.S. Eastern and Western Shale Oils Produced by Pyrolysis and Hydropyrolysis', *Fuel*, **61**, 1101–1109 (1982).
110. R.B. Williams, 'Characterization of Hydrocarbons in Petroleum by Nuclear Magnetic Resonance Spectrometry', ASTM Special Technical Publication No. 224, ASTM, Philadelphia, 168–194, 1958.
111. L.P. Lindeman, 'Characterization of Heavy Saturate Petroleum Fractions by NMR and Mass Spectroscopy', *Prepr. Pap. – Am. Chem. Soc., Div. Petrol. Chem.*, **14**(3), B186–B198 (1969).
112. K.W. Bartz, N.F. Chamberlain, 'Determining Structure of Paraffinic Chains by NMR', *Anal. Chem.*, **36**(11), 2151–2157 (1964).
113. D.E. Anders, W.E. Robinson, 'Geochemical Aspects of the Saturated Hydrocarbon Constituents of Green River Oil Shale – Colorado No. 1 Core', US Bureau of Mines Report of Investigations 7737, US BuMines, Washington, DC, 1973.
114. K.D. Bartle, D.W. Jones, H. Pakdel, 'Isoprenoid Hydrocarbons in Coal and Petroleum: A Preliminary Spectroscopic Study', in *Molecular Spectroscopy*, ed. A.R. West, Heyden and Son, London, Chapter 9, 1976.
115. P.A. Couperus, A.D.H. Calque, J.P.C.M. VanDongen, 'Chemical Shifts of Some Model Olefins', *Org. Magn. Reson.*, **8**, 426–431 (1976).
116. D.M. Cantor, 'Nuclear-magnetic Resonance Spectrometric Determination of Average Molecular Structure Parameters for Coal-derived Liquids', *Anal. Chem.*, **50**, 1185–1187 (1978).
117. ASTM D3904-80, ASTM, Philadelphia, Section 5, 516–528, 1983.

Oxygenate Vapor–Liquid Equilibrium in Gasolines

Stanley I. Sandler
University of Delaware, Newark, USA

Kenneth N. Marsh
University of Canterbury, Christchurch, New Zealand

1 Introduction 1
2 History and Environmental Applications 1
3 Methods Used 2
4 General Comments on the Vapor–Liquid Equilibrium of Oxygenate–Hydrocarbon Systems 5
5 Review of Measurements of the Vapor–Liquid Equilibrium of Oxygenate–Hydrocarbon Systems 7
6 Discussion 7
Acknowledgments 11
Abbreviations and Acronyms 11
Related Articles 11
References 11

Ethers and alcohols are being added to hydrocarbon fuels to produce the current generation of ''clean gasolines'' which are claimed to reduce air pollution. Also, alcohols provide a renewable energy source. However, the addition of oxygenates can change the volatility (as measured by the Reid vapor pressure) of a hydrocarbon mixture. This article reviews the equipment and measurement methods commonly used to study the vapor–liquid equilibrium (VLE) of hydrocarbon–oxygenate and gasoline–oxygenate mixtures. Detailed tables present those oxygenate–hydrocarbon mixtures for which VLE data are available.

1 INTRODUCTION

The development of ''clean gasolines'' in the United States has resulted in considerable interest in the measurement and correlation of VLE data for oxygenate–hydrocarbon mixtures. Of primary interest are ethers and alcohols and the effect of such added oxygenates on the volatility of gasolines. Gasoline volatility is generally reported as the Reid vapor pressure, that is, the pressure that would be produced when the gasoline is placed in an evacuated, closed container and partially vaporized as described by the American Society for Testing and Materials (ASTM) standard ASTM-D 323-94 Vapor Pressure of Petroleum Products (Reid Method). The Reid vapor pressure is an important characterizing parameter for gasoline, and there are United States Environmental Protection Agency (USEPA) limits on its value in commerce.

2 HISTORY AND ENVIRONMENTAL APPLICATIONS

Gasoline is neither a single identifiable component nor a well-defined mixture. Rather, it is a complex mixture of hydrocarbons largely in the boiling range of 40–200 °C. Gasoline is generally blended to produce a product with a desired antiknock or octane rating, little tendency to vapor lock (which depends on the vapor pressure of the fuel), and other specified properties. The composition of gasoline varies as a result of the starting crude oil, the producer, the season, the geographic location of use, the grade, and other factors. For many years tetraethyllead was used to increase the octane number of gasolines. This has been prohibited in many countries for environmental reasons, and also because lead is a poison to the catalytic converters now used on automobiles. Aromatics have high octane numbers, but recent restrictions on the amount allowed in gasoline have resulted in a further decrease in gasoline octane numbers. This has led to the increased use of nonlead octane enhancers. Oxygenates are octane-number enhancers, and several, especially methyl *tert*-butyl ether (MTBE) and ethanol, have been added to gasoline for this reason. Such blends have been referred to as ''clean gasolines''.

As the interest of refiners is to predict the effect of oxygenate addition on gasoline volatility, much of the VLE data are for the addition of an oxygenate to pure hydrocarbons or well-characterized mixtures of hydrocarbons. In particular, VLE data are available for mixtures of an oxygenate with either (1) a single hydrocarbon, (2) a mixture of defined hydrocarbons, or (3) a ''standard gasoline''. Such data are useful for determining parameters in activity coefficient or equation of state models that then allow gasoline refiners to make predictions for the behavior of their oxygenate–gasoline blends.

There are many oxygenates; however, this review with consider only ether- and alcohol-containing systems as only these oxygenates that have been widely used as gasoline additives. The most important data for

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

processing and for end-use in the clean gasolines is VLE, although other data such as liquid–liquid equilibrium (LLE) data are also of interest, especially because of the possibility of liquid–liquid phase splitting in the presence of water, or at high alcohol concentrations.

3 METHODS USED

There are two parts to determining the VLE of oxygenate–gasoline systems: one is producing VLE in the mixture, and the other is determining the composition of the equilibrium phases (although, as described below, in some experimental techniques the composition of one of the phases is gravimetrically prepared).

There are a number of ways that VLE data for mixtures can be obtained. The most complete data set includes the pressure (P), temperature (T), and the mole fraction compositions of the co-existing liquid (x) and vapor (y). Such data are referred to as $P–T–x–y$ data. For the correlation of such VLE data (especially for binary mixtures) it is preferable to have approximately 10 VLE points at each temperature covering the whole liquid composition range, together with the pure component vapor pressures measured in the same equipment. Usually, data for more than one isotherm would be reported. Data are also reported for a variety of compositions at a fixed pressure and therefore varying temperature. Such data are somewhat more difficult to correlate with thermodynamic models because of the temperature dependence of the model parameters, but are useful for distillation column design (as such devices operate close to isobarically), and generally are easier to measure using automated pressure control devices. Other types of VLE measurements include ebulliometry, static cell devices, gas–liquid chromatography (GLC), and headspace chromatography. As will be discussed below, each of these provides useful, but less complete, data in that either only a limited range of compositions are measured, or the compositions of one of the phases are not measured.

Recirculation equipment for VLE measurements has been discussed previously.[1,2] A double-recirculation still, one type of which is shown schematically in Figure 1, is frequently used to obtain $P–T–x–y$ data. [Note: there are many variations of this type of still, as can be seen in the literature.[1,2]] In the still shown, the liquid mixture boils in the still pot, producing a vapor–liquid mixture that is trapped in the inverted funnel region of a helical tube (a Cottrell pump), forcing the mixture into an adiabatic region where VLE is achieved. As the flow continues, the vapor and liquid phases disengage, the vapor is condensed, and the liquid and condensed vapor return

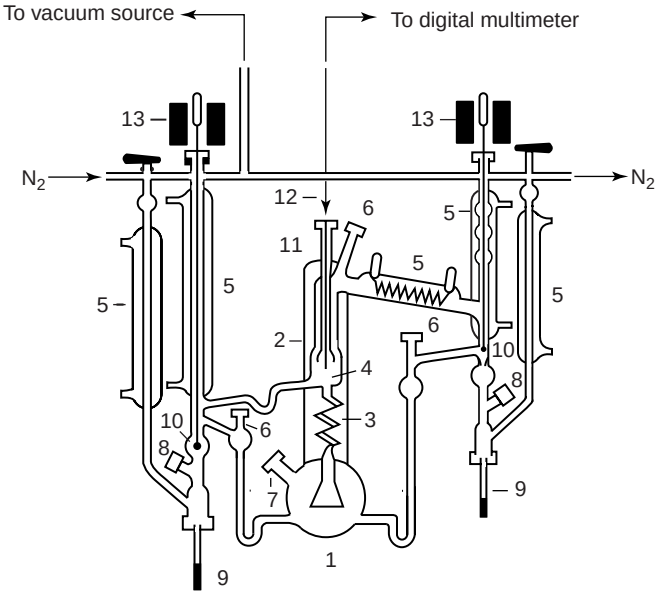

Figure 1 Schematic diagram of a double-recirculation VLE still: (1) still pot; (2) vacuum jacket; (3) Cottrell pump; (4) VLE chamber; (5) condensers; (6) injection ports; (7) filling spout and thermowell; (8) sample bypass valves; (9) sample tubes; (10) glass ball sampling valves; (11) thermowell; (12) platinum resistance thermometer; (13) solenoids to actuate sampling valves.

to the still pot by separate lines. Samples for analysis may be removed from the liquids in each of these lines. Depending on the mixture under study, different analytical methods can be used. For binary mixtures, techniques such as gas chromatography (GC) with an appropriate detector (see **Hyphenated Gas Chromatography**, **Column Technology in Gas Chromatography**, **Trace Organic Analysis by Gas Chromatography with Selective Detectors**), refractometry (for binary mixtures that are liquids under ambient conditions) and densimetry, have been used. For composition measurements involving real and simulated gasolines, which are multicomponent mixtures, GC is most commonly used.

The advantage of complete $P–T–x–y$ data for binary mixtures is that such data can be tested for thermodynamic consistency. However, such tests are not unambiguous. Also, composition measurements, especially of the condensed vapor, are the least accurate of the measurements made in VLE, and for some applications it is not necessary to have such complete data. Ebulliometry is one example of an experimental technique that results in useful, but less complete, data. In this technique, a gravimetrically prepared liquid sample is boiled in a still that recirculates the condensed liquid, and the temperature and pressure are measured. Figure 2 shows one example of such a device, though many variations exist. If the amount of recirculating liquid is not too large

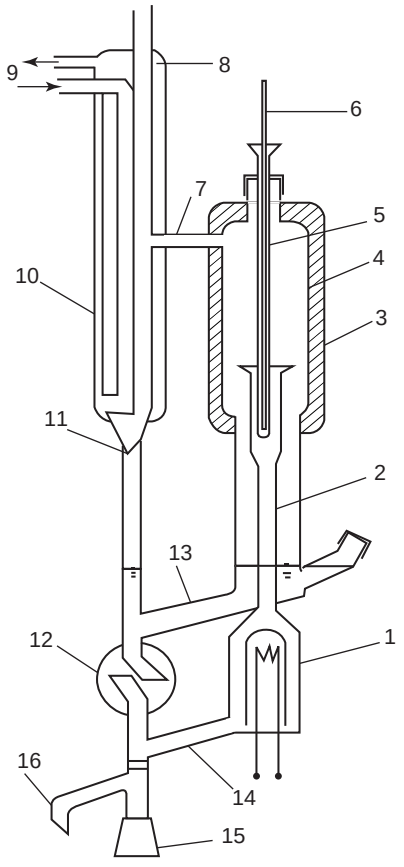

Figure 2 Schematic diagram of an ebulliometer: (1) still pot; (2) Cottrell pump; (3) vacuum jacket and insulation; (4) VLE chamber; (5) thermowell; (6) thermocouple; (7) vapor passover line; (8) connection to pressure controller/vacuum pump; (9) coolant inlet and outlet; (10) condenser; (11) condensed vapor drop; (12) mixing chamber; (13) and (14) liquid recycle line; (15) stopcock; (16) drain port.

(so that the composition of the equilibrium liquid can be considered to be the same as the prepared liquid), one then obtains $P–T–x$ data, but not the composition of the co-existing vapor. Usually in such measurements, the pressure is fixed and the equilibrium temperature measured, since attempting to adjust the pressure might result in vaporization and loss of the more volatile component.

As the relative volatility of the components and/or the amount of condensed vapor holdup in the still increases, there is greater error in the assumption that the prepared and equilibrium compositions are identical. A common procedure is to correct for the liquid holdup using an iterative procedure; however, this requires knowledge of the amount of vapor holdup and the composition of the equilibrium condensed vapor, neither of which is known accurately. Consequently, this measurement technique is generally limited to relative volatilities (Equation 1):

$$\alpha_{ij} = \frac{y_i/x_i}{y_j/x_j} \tag{1}$$

of the two components in a binary mixture of between 0.1 to 10, as beyond this range the error from the unknown vapor phase composition becomes too large.

Ebulliometry is generally used for binary mixtures, and in either one of two ways. The first is by preparing gravimetrically a number of samples and obtaining $P–T–x$ data over the whole composition range. Barker's method[3] or its later variants is then used to calculate the vapor composition and then the activity coefficients. A simplified description of this calculation procedure is as follows. At low pressures, where the vapor phase can be assumed to be ideal, we have for species i at experimental point j that (Equation 2)

$$x_{i,j}\gamma_i(x_{i,j})P_i^{\text{vap}}(T) = y_{i,j}P_j \tag{2}$$

where $\gamma_i(x_{i,j})$ is the activity coefficient of species i at the mole fraction $x_{i,j}$ of that species at the jth experimental point at which the total pressure is P_j and $P_i^{\text{vap}}(T)$ is the vapor pressure of pure component i at the temperature T of the measurement. Summing over both species in a binary mixture at each experimental point gives Equation (3):

$$\sum_i x_{i,j}\gamma_i(x_{i,j})P_i^{\text{vap}}(T) = P_j \tag{3}$$

by assuming an activity coefficient model, for example, the very simple Van Laar model (see, for example, Sandler[4] for a discussion of various activity coefficient models), Equation (4):

$$\ln\gamma_1(x_{1,j}) = \frac{\alpha}{(1+\alpha x_{1,j}/\beta x_{2,j})^2}$$

and

$$\ln\gamma_2(x_{2,j}) = \frac{\beta}{(1+\beta x_{2,j}/\alpha x_{1,j})^2} \tag{4}$$

we have for the calculated pressure P_j^{calc} at each point (Equation 5):

$$x_{1,j}\exp\left[\frac{\alpha}{(1+\alpha x_{1,j}/\beta x_{2,j})^2}\right]P_1^{\text{vap}} + x_{2,j}\exp\left[\frac{\beta}{(1+\beta x_{2,j}/\alpha x_{1,j})^2}\right]P_2^{\text{vap}} = P_j^{\text{calc}} \tag{5}$$

Next, by using all the experimental data, one chooses the values of the adjustable parameters in the activity coefficient model (here α and β) to minimize the value of

an objective function F such as given in Equation (6):

$$F = \sum_{\substack{\text{experimental} \\ \text{points } j}} (P_j^{\text{calc}} - P_j^{\text{exp}})^2 \tag{6}$$

With the parameter values so obtained, the activity coefficient can be calculated at any composition. Then Equation (2) can be used to estimate the unmeasured vapor phase compositions. Note, however, that it is more common to use more sophisticated and somewhat more accurate activity coefficient models than the Van Laar model; also, more sophisticated parameter estimation methods, such as the maximum likelihood method,[5] can be used.

If the experimental data points are closely spaced, it is also possible to estimate the vapor compositions [and from those and Equation (2) the activity coefficients] without making any assumptions about the functional form of the activity coefficient model. Starting from the Gibbs–Duhem equation, and making the mild assumptions of an ideal vapor phase, and that $PV^{\text{ex}}/RT \ll 1$, where V^{ex} is the molar excess volume on mixing at constant temperature and pressure, one has (Equation 7)

$$x_1 \frac{\mathrm{d}\ln(x_1\gamma_1)}{\mathrm{d}x_1} + x_2 \frac{\mathrm{d}\ln(x_2\gamma_2)}{\mathrm{d}x_2} = 0 \tag{7}$$

which written on a point-by-point basis and using Equation (2) can be transformed into Equation (8):

$$\frac{(y_{1,j} - x_{1,j})}{y_{1,j}(1 - y_{1,j})} \frac{\mathrm{d}y_{1,j}}{\mathrm{d}x_{1,j}} = \frac{\mathrm{d}\ln P_j^{\text{exp}}}{\mathrm{d}x_{1,j}} \tag{8}$$

Replacing the derivatives by finite differences (which is why the experimental data must be closely spaced), and using the experimental pressure and liquid mole fractions and an appropriate numerical integration scheme, the value of the vapor-phase mole fraction can be computed for each liquid-phase mole fraction. This is called the synthetic method for obtaining P–T–x–y data using prepared liquid compositions without the direct measurement of the composition of either phase. This method is particularly useful in mixtures such as alcohols–hydrocarbons which exhibit very non-ideal behavior. For such systems most activity coefficient models cannot adequately represent the experimental data, particularly at low concentrations of the alcohol.[6]

The second method of using ebulliometry is to make a few measurements at the high-dilution range of each component, from these data to obtain the slope of the equilibrium temperature with composition at fixed pressure, and then calculate the infinite-dilution activity coefficient of each component in the other from Equation (9):

$$\gamma_i(x_i \to 0) = \gamma_i^{\infty} = \frac{P_j^{\text{vap}} - (\mathrm{d}P_j^{\text{vap}}/\mathrm{d}T)(\partial T/\partial x_i)_{T,x_i \to 0}}{P_i^{\text{vap}}} \tag{9}$$

Once the infinite-dilution activity coefficients have been obtained for each component in the other, their values can be used to obtain the parameters in a two-parameter activity coefficient model, and then to predict the VLE over the complete concentration range. Since ebulliometry is a dynamic boiling method using stills, similar to that used to obtain P–T–x–y data discussed above, it is of use for approximately the same relative volatility range.

To improve the accuracy of ebulliometric measurements, differential ebulliometry is commonly used. In this measurement method, two (or more) ebulliometers are connected to the same manifold, one containing the pure solvent as a reference and the other containing the solvent into which the solute is to be added. Then, by using differential thermometry, the temperature difference as a result of solute addition is measured directly. This is more accurate than using a single ebulliometer and measuring the two temperatures separately.

In static methods, boiling does not occur as it does in the dynamic methods discussed above. Rather, a cell is almost completely filled with a liquid of known (gravimetrically prepared) composition, closed, maintained at constant temperature, and the equilibrium pressure is measured. The advantage of such a cell is that very little of the liquid is vaporized, so except for the case of a very extreme difference in relative volatility, the error introduced by a changing liquid composition as a result of partial vaporization of the liquid is small. The disadvantage of such closed-cell equipment is that any gas dissolved in the liquid results in a significant error in the measured pressure. Therefore, the liquids have to be carefully degassed (by partial vaporization or alternate freezing, evacuation, and thawing) before introduction into the static cell. With this equipment, P–T–x data are generally obtained at constant temperature.

As in ebulliometry, measurements can be over the whole composition range, in which case the results are analyzed in the same manner as described above. Also, measurements can be made in just the high-dilution regions to obtain the slope of the equilibrium pressure with composition at fixed temperature. In this latter case the infinite-dilution activity coefficient of the solute in the solvent in a binary mixture is obtained from Equation (10):

$$\gamma_i(x_i \to 0) = \gamma_i^{\infty} = \frac{P_j^{\text{vap}} + (\partial P/\partial x_i)_{T,x_i \to 0}}{P_i^{\text{vap}}} \tag{10}$$

By obtaining both infinite-dilution activity coefficients, one can fit parameters to an activity coefficient model and estimate the VLE over the whole composition range.

Static cell methods have also been used to determine directly the effect on the Reid vapor pressure of the addition of an oxygenate to a real or simulated gasoline. In this case the oxygenate addition is done gravimetrically, so composition measurements are not required. The main difficulty is initial degassing the gasoline without changing its composition as a result of volatilization of the light components. Degassing of gasoline (or other mixtures) in such cases is generally done by placing the gasoline in the static cell, freezing to liquid nitrogen temperature, evacuating the headspace, thawing, and repeating the process a number of times until the vapor pressure remains constant on repeated cycles. In this way, dissolved gases such as oxygen and nitrogen can be eliminated without evaporation of the light hydrocarbon components.

The accuracy of the static cell measurements can be improved by using a differential method in which there are two (or more) static cells in the same temperature bath. The reference cell contains the pure solvent and the other cell contains solvent to which the oxygenate is to be added. A differential pressure transducer or similar device between the two cells directly measures the pressure difference resulting from addition of the oxygenate.

There also exist several gas–liquid chromatographic methods for determining infinite-dilution activity coefficients in a pure solvent. In each case the chromatographic column is loaded with the solvent and the solute is injected. By measuring the temperature, pressure, column loading, and the carrier gas retention volume, the infinite-dilution activity coefficient of the solute in the solvent can be calculated. Alternatively, by measuring only the temperature and the peak times for both a reference solute and the solute of interest, the ratio of their infinite-dilution activity coefficients in the solvent can be obtained.[7]

Headspace methods provide still another technique for obtaining VLE data. In this method, an equilibrium cell similar to a static cell is used. However, the liquids are not degassed prior to a measurement, and the total pressure generally is not measured. The headspace, that is, the vapor space above the liquid, is sampled and analyzed. Consequently, in this measurement method one obtains the equilibrium vapor composition in equilibrium with the prepared liquid. As with the other methods discussed, this technique can be used over the whole composition range, or only in the high-dilution regions to obtain infinite dilution activity coefficients.

LLE data are usually obtained using relatively simple, thermostated equilibrium cells to which the components are added and stirred to obtain equilibrium as shown in Figure 3. The stirring is stopped, time is allowed to let the phases separate, and then the equilibrium phases are sampled and analyzed. Care must be used on sampling not to contaminate one liquid phase with the other. The same methods of analysis as used for VLE can also be used for LLE. Liquid–liquid phase separation for the systems of interest here occurs when water is present, and in some alcohol–hydrocarbon systems.

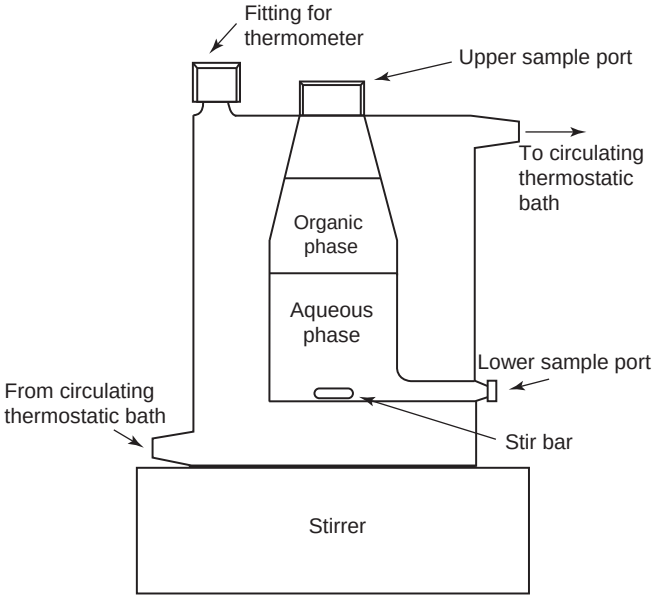

Figure 3 Schematic diagram of an LLE cell.

4 GENERAL COMMENTS ON THE VAPOR–LIQUID EQUILIBRIUM OF OXYGENATE–HYDROCARBON SYSTEMS

As indicated in the tables that follow, there are many data sets for binary mixtures of oxygenates with hydrocarbons, for ternary mixtures of an oxygenate with two hydrocarbons or two oxygenates with a single hydrocarbon. There are relatively few published data for an oxygenate with a multicomponent hydrocarbon mixture or a gasoline. (Given the importance of such information in the fuels industry, it is likely such data have been measured but are proprietary.)

The VLE data show that ethers and hydrocarbons form nearly ideal mixtures. Consequently, the effect of the ether on the volatility (i.e. the Reid vapor pressure) of the gasoline–oxygenate mixture is largely determined by the vapor pressure of the oxygenate. It is partly for this reason that largely C_5 and higher carbon number ethers have been used as gasoline additives. MTBE is the oxygenate most commonly used in the USA, and may

result in a very slight increase in the Reid vapor pressure of a gasoline. Higher ethers such as *tert*-amyl methyl ether (TAME), diisopropyl ether (DIPE), and others, which are less volatile, have little effect on the Reid vapor pressure of the gasoline at the concentrations in which they are used, and may even result in a slight lowering. One method of lowering the Reid vapor pressure of a gasoline is to reduce its content of C_4 hydrocarbons, which has resulted in a refinery stream that can be converted to MTBE. As a result, MTBE is relatively inexpensive and the mostly widely used oxygenated gasoline additive. However, there are concerns about the environmental effects of MTBE, particularly with regard to groundwater contamination.

Alcohols form nonideal mixtures with hydrocarbons. These mixtures show positive deviations from Raoult's law; consequently, if a low-molecular-weight alcohol such as methanol or ethanol is added to gasoline, an increase in the Reid vapor pressure will result. Also, some alcohol–gasoline mixtures, particularly those containing methanol, exhibit liquid–liquid phase splitting. Many binary mixtures of an alcohol and a hydrocarbon have a minimum boiling azeotrope and/or liquid–liquid partial immiscibility. Ethanol is the alcohol most commonly used as a gasoline additive as a result of its availability from renewable or surplus agricultural products, and its low cost. Since ethanol is available domestically, it can be used to reduce a country's dependence on foreign crude oil. Indeed, it has been used in this manner at concentrations exceeding 20 wt% in Brazil. Methanol has been used as a gasoline replacement.

Table 1 MTBE, TAME, DME, DEE, and BME binary systems

Substance	MTBE		TAME		DME		DEE		BME	
	VLE	γ^∞	VLE	γ^∞	VLE	γ^∞	VLE	γ^∞	VLE	γ^∞
Butane	✓	✓			✓					
2-Methylpropane	✓						✓			
Pentane			✓				✓			
2-Methylbutane	✓		✓				✓			
Hexane	✓	✓		✓			✓		✓	
2-Methylpentane	✓	✓	✓				✓			
3-Methylpentane	✓		✓							
Heptane	✓	✓	✓	✓	✓		✓		✓	
Octane	✓	✓	✓				✓		✓	
2,2,4-Trimethylpentane	✓	✓	✓							
Decane							✓		✓	
Dodecane							✓		✓	
Tetradecane		✓								
Hexadecane		✓					✓		✓	
Cyclopentane	✓	✓								
Cyclohexane	✓	✓	✓				✓		✓	
Methylcyclohexane	✓	✓	✓							
Propene	✓									
But-1-ene	✓				✓		✓			
But-2-ene	✓									
2-Methylprop-1-ene	✓						✓			
2-Methylbut-1-ene	✓		✓							
2-Methylbut-2-ene	✓		✓							
Hex-1-ene	✓		✓							
Hept-1-ene	✓									
2,2,4-Trimethylpent-2-ene	✓									
Buta-1,3-diene	✓						✓			
2-Methyl-1,3-butadiene	✓									
Benzene	✓		✓		✓		✓		✓	
Toluene	✓		✓				✓			
1,2-Dimethylbenzene	✓		✓							
1,4-Dimethylbenzene	✓									
DIPE	✓									
DME	✓									
ETBE							✓			
TAME	✓									
MTBE					✓					
Gasoline blend (>3)	✓									

Table 2 DPE, DIPE, ETBE, DBE and DIBE binary systems

Substance	DPE		DIPE		ETBE		DBE		DIBE	
	VLE	γ^∞	VLE	γ^∞	VLE	γ^∞	VLE	γ^∞	VLE	γ^∞
Butane			✓							
Pentane								✓		
2-Methylbutane					✓					
Hexane					✓	✓	✓	✓		
2-Methylpentane					✓					
Heptane	✓		✓	✓	✓	✓	✓	✓		
2,4-Dimethylpentane			✓		✓					
Octane	✓		✓	✓	✓			✓		
2,2,4-Trimethylpentane					✓					
Nonane	✓									
Tetradecane				✓						
Hexadecane		✓		✓						
Octadecane				✓					✓	
Cyclohexane	✓		✓	✓			✓	✓		
Hex-1-ene			✓							
Hept-1-ene					✓					
Oct-1-yne							✓			
Oct-2-yne							✓			
Benzene	✓		✓				✓	✓		
Toluene	✓		✓		✓					
Ethylbenzene	✓		✓	✓						
1,2-Dimethylbenzene					✓					
DIPE	✓									
DEE							✓			

There has been some interest in the VLE of gasoline with mixed oxygenates, especially an ether and an alcohol. Some of the data sets given in the tables reflect that interest.

5 REVIEW OF MEASUREMENTS OF THE VAPOR–LIQUID EQUILIBRIUM OF OXYGENATE–HYDROCARBON SYSTEMS

This review of measurements only covers mixtures of ethers or alcohols with hydrocarbons. Most of the data that have been reported are for binary mixtures, although ternary and multicomponent mixtures are also considered. The properties reviewed are VLE and LLE. The literature search was made using on-line and hard-copy versions of *Chemical Abstracts*, on-line *Current Contents*, various Thermodynamics Research Center electronic databases,[8] the EVLM'97 electronic database,[9] the Dortmund Data Bank,[10,11] and recent volumes of *Fluid Phase Equilibria*, *Journal of Chemical Thermodynamics* and *Journal of Chemical and Engineering Data*.

In the tables, a tick indicates that one or more measurements have been made on a particular system. Table 1 indicates measurements made on MTBE, TAME, diethyl ether (DEE), and butyl methyl ether (BME) with hydrocarbons and with other ethers. Table 2 contains information for measurements made on mixtures of dipropyl ether (DPE), ethyl *tert*-butyl ether (ETBE), dibutyl ether (DBE), and diisobutyl ether (DIBE) with hydrocarbons and other ethers. Table 3 contains information on binary mixtures of methanol and ethanol with hydrocarbons–ethers. Table 4 shows the measurements made on multicomponent mixtures containing ethers–hydrocarbons–other components and Table 5 indicates measurements available on multicomponent mixtures containing alcohols–hydrocarbons–other components. Complete annotated references to the measurements on ethers are given in a recent review.[12]

6 DISCUSSION

The experimental methods used to obtain VLE data in oxygenate–hydrocarbon and oxygenate–gasoline systems are well established, and provide no unusual challenges. Consequently, many data are available. There are a considerable number of VLE data sets for alkanes and cycloalkanes up to C_8 for mixtures with ethers. There are few measurements on alkanes larger than C_8. There are a large number of VLE measurements for mixtures of MTBE with alkenes, but none with alkynes. For aromatic compounds there exists a comprehensive

Table 3 Methanol and ethanol binary systems

Substance	Methanol		Ethanol	
	VLE	LLE	VLE	LLE
Butane	✓	✓	✓	
2-Methylpropane	✓		✓	
Pentane	✓	✓	✓	
2-Methylbutane	✓		✓	
2,2-Dimethylpropane	✓			
Hexane	✓	✓	✓	
2-Methylpentane	✓			
3-Methylpentane	✓	✓		
2,2-Dimethylbutane		✓		
2,3-Dimethylbutane	✓	✓	✓	
Heptane	✓	✓	✓	
Octane	✓	✓	✓	
3-Methylheptane		✓		
2,2,3-Trimethylpentane		✓		
2,2,4-Trimethylpentane	✓		✓	
Nonane	✓	✓	✓	
2,2,5-Trimethylhexane		✓		
Decane	✓	✓	✓	✓
Undecane	✓	✓	✓	
Dodecane	✓		✓	
Tridecane	✓			
Tetradecane	✓	✓	✓	
Hexadecane	✓	✓	✓	✓
Octadecane		✓	✓	
Cyclopentane		✓	✓	
Cyclohexane	✓	✓	✓	
Methylcyclopentane	✓	✓	✓	
Methylcyclohexane	✓	✓	✓	
But-1-ene	✓			
But-2-ene	✓			
2-Methylpropene	✓		✓	
Pent-1-ene	✓			
trans-Pent-2-ene	✓		✓	
2-Methylbut-1-ene	✓			
2-Methylbut-2-ene	✓			
Hex-1-ene	✓		✓	
Hept-1-ene	✓	✓	✓	
2,4-Dimethylpent-2-ene	✓			
Oct-1-ene	✓			
2,2,4-Trimethylpent-2-ene	✓			
Dodec-1-ene		✓		
Buta-1,3-diene	✓		✓	
2-Methylbuta-1,3-diene	✓			
cis-Penta-1,3-diene	✓			
trans-Penta-1,3-diene	✓			
Cyclopenta-1,3-diene	✓			
1-Methylcyclobutene	✓			
Cyclohexene	✓			
But-1-yne	✓			
Hex-1-yne	✓			
Hex-2-yne	✓			
Hex-3-yne	✓			
But-3-en-1-yne	✓			
Benzene	✓	✓	✓	
Toluene	✓		✓	
Ethylbenzene	✓		✓	
1,2-Dimethylbenzene	✓		✓	
1,3-Dimethylbenzene	✓		✓	
1,4-Dimethylbenzene	✓		✓	
1,2,4-Trimethylbenzene		✓		
1,3,5-Trimethylbenzene		✓		
Methyl ethyl ether	✓			
DEE	✓		✓	
MTBE	✓		✓	
Ethyl propyl ether	✓			
BME	✓			
Butyl ethyl ether			✓	
DIPE	✓		✓	
ETBE	✓		✓	
DPE	✓		✓	
TAME	✓		✓	
DBE	✓		✓	

set of VLE measurements, except for ethylbenzene and propylbenzene. There are VLE data for MTBE with various synthetic gasoline blends, and there have been a considerable number of measurements reported for MTBE–alkanes + a third component including alkanes, cycloalkanes, aromatics, ethers, and alcohols. Studies on ternary mixtures with cycloalkanes have been restricted to cyclohexane, while benzene has been the main aromatic compound studied. There is a general lack of LLE data for ternary systems involving alcohols or water.

There are no VLE measurements of TAME with hexane and alkanes larger than C_8 and limited measurements with the branched alkanes. There are sufficient VLE measurements for modeling purposes with the alkenes, though not as extensive as for MTBE, but there are no VLE measurements with the alkynes. Considerably fewer measurements have been reported for ternary mixtures containing TAME than for MTBE. While the measurements cover a cross-section of the classes of mixtures, more measurements are warranted, especially on systems containing TAME–aromatic–water, TAME–aromatic–alcohol, TAME–alkene–water, and TAME–alkene–alcohol.

There exist VLE measurements for binary mixtures with dimethyl ether (DME), DEE, and BME and a good selection of hydrocarbons.There are no measurements for DPE with alkenes, while for some mixtures, e.g. DIPE separately with heptane, cyclohexane, and benzene, many VLE data sets have been reported. There are adequate measurements on binary mixtures with ETBE and DBE, but virtually no measurements with DIBE. There have been few measurements on ternary mixtures containing ethers other than MTBE and TAME. Additional measurement should focus on the LLE for mixtures containing ethers–hydrocarbons–alcohols and ethers–hydrocarbons–water.

For binary mixtures of either methanol and ethanol with mixtures of a wide range of hydrocarbons there are many measurements of both VLE and LLE (where applicable), and there appears to be little need for any additional measurements. There are many measurements

Table 4 Multicomponent systems containing ethers

Ether	Substance 2	Substance 3	VLE or γ^{∞}	LLE
MTBE	Hexane	Heptane	✓	
MTBE	Hexane	Octane	✓	
MTBE	Hexane	Toluene	✓	
MTBE	Heptane	Octane	✓	
MTBE	Heptane	Cyclohexane	✓	
MTBE	Heptane	Benzene	✓	
MTBE	Heptane	Methanol		✓
MTBE	Heptane	Oxolane	✓	
MTBE	Octane	Methanol		✓
MTBE	Octane	Ethanol	✓	
MTBE	Octane	2-Methylpropan-2-ol	✓	
MTBE	2,2,4-Trimethylpentane	Water		✓
MTBE	Nonane	Methanol		✓
MTBE	Decane	Methanol		✓
MTBE	Cyclopentane	Methanol	✓	
MTBE	Cyclohexane	Hex-1-ene	✓	
MTBE	Cyclohexane	Benzene	✓	
MTBE	But-2-ene	Methanol	✓	
MTBE	2-Methylprop-1-ene	Methanol	✓	
MTBE	Benzene	Hex-1-ene	✓	
MTBE	Benzene	Water		✓
MTBE	Toluene	Water		✓
TAME	2-Methylbutane	Methanol	✓	
TAME	Heptane	Cyclohexane	✓	
TAME	Heptane	1-Hexene	✓	
TAME	Heptane	Benzene	✓	
TAME	Heptane	Methanol	✓	
TAME	2,2,4-Trimethylpentane	Water		✓
TAME	Decane	Water		✓
TAME	Hexadecane	Water		✓
TAME	Cyclohexane	1-Hexene	✓	
TAME	Cyclohexane	Benzene	✓	
TAME	Methylcyclohexane	Water		✓
TAME	1-Hexene	Benzene	✓	
TAME	Toluene	Water		✓
TAME	5-component mixture		✓	
DEE	Hexane	Methanol	✓	
DEE	Heptane	Methanol		✓
DEE	Octane	Methanol		✓
DEE	Nonane	Methanol		✓
DEE	Decane	Methanol		✓
DEE	Toluene	Ethanol	✓	
DEE	Cyclohexane	Benzene	✓	
DIPE	Heptane	Cyclohexane	✓	
DIPE	Heptane	1-Hexene	✓	
DIPE	Heptane	Benzene	✓	
DIPE	Cyclohexane	Benzene	✓	
DIPE	Hex-1-ene	Cyclohexane	✓	
DIPE	Hex-1-ene	Benzene	✓	
DBE	Heptane	Methanol		✓
DBE	Octane	Methanol		✓
DBE	Nonane	Methanol		✓
DBE	Decane	Methanol		✓
ETBE	Hexane	Heptane	✓	
ETBE	Heptane	Octane	✓	
ETBE	Octane	Isooctane	✓	

Table 5 Multicomponent systems containing alcohols

Alcohol	Substance 2	Substance 3	VLE	LLE
Methanol	Butane	Cyclohexane	✓	
Methanol	Butane	Benzene	✓	
Methanol	Butane	Water		✓
Methanol	Pentane	Ethanol	✓	
Methanol	Hexane	Cyclohexane	✓	
Methanol	Hexane	Methylcyclohexane	✓	
Methanol	Hexane	DEE	✓	
Methanol	2-Methylpropane	Cyclohexane	✓	
Methanol	2-Methylpropane	Benzene	✓	
Methanol	Heptane	Water		✓
Methanol	Octane	Water		✓
Methanol	Nonane	Water		✓
Methanol	Cyclohexane	Water		✓
Methanol	But-2-ene	MTBE	✓	
Methanol	2-Methylpropene	MTBE	✓	
Methanol	Hex-1-ene	Water		✓
Methanol	Hept-1-ene	Water		✓
Methanol	Cyclohexene	Water		✓
Methanol	2-Methylbuta-1,3-diene	Water		✓
Methanol	2-Methylbuta-1,3-diene	2-Methylbut-2-ene	✓	
Methanol	2-Methylbuta-1,3-diene	2-Methylbutane	✓	
Methanol	Benzene	Ethanol	✓	
Methanol	Benzene	Cyclohexane	✓	
Methanol	Benzene	Toluene	✓	
Methanol	Benzene	1,2-Dimethylbenzene	✓	
Methanol	Benzene	Water		✓
Methanol	Toluene	Methylcyclohexane	✓	
Methanol	Toluene	Heptane	✓	
Methanol	Toluene	Water	✓	✓
Methanol	1,2-Dimethylbenzene	Water		✓
Methanol	1,4-Dimethylbenzene	Water		✓
Methanol	Water	Hydrocarbons	✓	
Ethanol	Butane	Water	✓	
Ethanol	Pentane	Water	✓	
Ethanol	Hexane	Hex-1-ene	✓	
Ethanol	Hexane	Methylcyclopentane	✓	
Ethanol	Hexane	Benzene	✓	
Ethanol	Hexane	Water	✓	✓
Ethanol	Heptane	2,2,4-Trimethylpentane	✓	
Ethanol	Heptane	Benzene	✓	
Ethanol	Heptane	Toluene	✓	
Ethanol	Heptane	Water	✓	✓
Ethanol	Octane	Water	✓	✓
Ethanol	Nonane	Water	✓	✓
Ethanol	Decane	Water	✓	✓
Ethanol	Dodecane	Water	✓	
Ethanol	Tetradecane	Water	✓	
Ethanol	Hexadecane	Water	✓	
Ethanol	Cyclohexane	Benzene	✓	
Ethanol	Cyclohexane	Toluene	✓	
Ethanol	Cyclohexane	Water	✓	✓
Ethanol	Methylcyclopentane	Benzene	✓	
Ethanol	Methylcyclohexane	Toluene	✓	
Ethanol	Benzene	Water	✓	✓
Ethanol	Toluene	Water	✓	
Ethanol	Toluene	DEE	✓	
Ethanol	Ethylbenzene	Water	✓	
Ethanol	1,2-Dimethylbenzene	Water		✓
Ethanol	1,3-Dimethybenzene	Water		✓

Table 5 (*continued*)

Alcohol	Substance 2	Substance 3	VLE	LLE
Ethanol	1,4-Dimethylbenzene	Water		✓
Ethanol	1,3,5-Trimethylbenzene	Water		✓
Ethanol	Benzene + cyclohexane	Water	✓	
Ethanol	Benzene + heptane	Water	✓	
Ethanol	>3-Component gasoline		✓	
Ethanol	>3-Component diesel fuel		✓	

of LLE for mixtures of alcohols–hydrocarbons–water, but there are a limited number of VLE data for such ternary mixtures. As noted, there are few data for alcohols–ethers–hydrocarbons. There are limited data for ternary alcohol–hydrocarbon–hydrocarbon mixtures and more measurements are warranted.

The three IUPAC Workshops on Vapor–Liquid Equilibria and Related Properties in Binary and Ternary Mixtures of Ethers, Alkanes, and Alkanols have stimulated much of the recent experimental work. There is now a very adequate database on VLE and LLE for many mixtures containing ethers.

ACKNOWLEDGMENTS

We wish to acknowledge the extremely valuable help of Professor J. Gmehling of the Universität Oldenburg in providing the information that allowed us to assemble the tables presented here, and to Prin Naimskul of the University of Canterbury for his help. Also, S.I.S. would like to acknowledge the financial support of the US National Science Foundation provided through grant number CTS-9521406

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
BME	Butyl Methyl Ether
DBE	Dibutyl Ether
DEE	Diethyl Ether
DIBE	Diisobutyl Ether
DIPE	Diisopropyl Ether
DME	Dimethyl Ether
DPE	Dipropyl Ether
ETBE	Ethyl *tert*-Butyl Ether
GC	Gas Chromatography
GLC	Gas–Liquid Chromatography
LLE	Liquid–Liquid Equilibrium
MTBE	Methyl *tert*-Butyl Ether
TAME	*tert*-Amyl Methyl Ether
USEPA	United States Environmental Protection Agency
VLE	Vapor–Liquid Equilibrium

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Selective Detectors

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Instrumentation of Gas Chromatography

REFERENCES

1. E. Hala, T. Pick, V. Fried, O. Vilim, *Vapor Liquid Equilibrium*, 2nd edition, Pergamon Press, Oxford, 1967.
2. J.D. Raal, A.L. Mühlbauer, *Phase Equilibria: Measurement and Computation*, Taylor and Francis, London, 1998.
3. J.A. Barker, 'Determination of Activity Coefficients from Total Pressure Measurements', *Aust. J. Chem.*, **6**, 207–210 (1953).
4. S.I. Sandler, *Chemical and Engineering Thermodynamics*, 3rd edition, John Wiley & Sons, New York, 1999.
5. S. Kemeny, J. Manczinger, S. Skold-Jorgensen, K. Toth, 'Reduction of Thermodynamic Data by Means of Multiresponse Maximum Liklihood Principle', *AIChE J.*, **28**, 20–30 (1982).
6. R.H. Stokes, M. Adamson, 'Thermodynamic Functions for System Ethanol + Cyclohexane from Vapor-pressures and Enthalpies of Mixing', *J. Chem. Soc., Faraday Trans. I*, **73**, 1232–1238 (1977).
7. H. Orbey, S.I. Sandler, 'Relative Measurements of Activity Coefficients at Infinite Dilution by Gas Chromatography', *Ind. Eng. Chem. Res.*, **30**, 2006–2011 (1991).
8. *TRC Databases for Chemistry and Engineering*, Thermodynamics Research Center, Texas A&M University, College Station, TX, 77843-3111, USA 1998.

9. I. Wichterle, J. Linek, Z. Wagner, H.V. Kehiaian, 'Vapor–Liquid Equilibrium in Mixtures and Solutions', ELDATA, Montreuil, 1996.
10. *Dortmund Data Bank*, DDBST, Oldenburg, 1998.
11. J. Gmehling, U. Onken, W. Arlt, P. Grenzheuser, J. Rarey, U. Weidlich, B. Kolbe, *Vapor–Liquid Equilibria Data Collection, DECHEMA Chemistry Data Series*, DECHEMA, Frankfurt, Vol. 1, Parts 1–20, 1999.
12. K.N. Marsh, P. Niamskul, J. Gmehling, R. Bölts, 'Review of Thermophysical Property Measurements on Mixtures Containing MTBE, TAME, and Other Ethers with Non-polar Solvents', *Fluid Phase Equilib.*, **159**, 207–227 (1999).

Petroleum Residues, Characterization of

Robert Bacaud
Institute de Recherches sur la Catalyse, Villeurbanne, France

1 Introduction **2**
2 Composition of Petroleum **2**
2.1 Definition of Residues 2
2.2 Constituents 3
3 Determination of Some Physical Parameters of Residues **4**
3.1 Specific Gravity 4
3.2 Viscosity 4
3.3 Carbon Residue 4
4 Chemical Analysis Without Previous Fractionation **5**
4.1 Elemental Analysis 5
4.2 Evaluation Without Previous Fractionation 5
5 Fractionation **7**
5.1 Solvent Extraction 7
5.2 Adsorption Chromatography 7
5.3 Short-path Molecular Distillation 8
5.4 Simulated Distillation by High-temperature Gas Chromatography 9
5.5 Supercritical Fluid Extraction 9
6 Characterization of Fractions **9**
6.1 Determination of Molecular Weight 9
6.2 Determination of Aromatics and Saturates 10
6.3 Hyphenated Techniques 10
6.4 Determination of Individual Compounds 11
7 Rationalization of Analytical Data **11**
7.1 Determination of Structural Parameters from Physical Properties 11
7.2 Determination of Structural Parameters from Spectroscopic Methods 11
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**

Petroleum residues are the remaining fraction left after the distillation of crude oil. According to the conditions of the primary distillation, a distinction is made between atmospheric and vacuum residues (VRs). Since the simplest, lighter constituents of petroleum have been separated by distillation, residues constitute the most complex fraction of crude oil and a detailed inventory of the individual compounds is impossible. Thus characterization methods will aim at giving an adequate description of the pertinent properties for a given purpose. Practically all of the existing analytical techniques have been applied: chromatographic, spectroscopic, solvent extraction, fractionation, etc. The analyst's strategy will consist in choosing the procedure that will be able to provide limited but pertinent information, at an acceptable cost, in a reasonable time. Some easily obtained physical parameters correlate with chemical composition data. This is the case with specific gravity, viscosity, and the content of carbon residue obtained after pyrolysis. Standard test methods for the determination of these parameters have been developed.

The direct determination of compositional parameters such as aromatic carbon and hydrogen content is obtainable through nuclear magnetic resonance (NMR) spectroscopy without previous fractionation. The quantitative distribution of saturated, aromatic, and polar compounds can be measured by thin-layer chromatography (TLC) coupled with flame ionization detection (FID). The parameters obtained through various analytical techniques are generally hardly correlated since they refer to different units (e.g. mole %, weight %, or volume %) that cannot be interconverted.

A more detailed characterization requires previous separation for the preparation of simpler fractions. Although the definition of residues implies they have been produced after a preliminary distillation, short-path molecular distillation, simulated distillation by high-temperature gas chromatography (GC), and supercritical fluid extraction (SFE) apply to residues and provide the distribution of fractions possessing equivalent boiling point up to about 720°C at atmospheric pressure. The fractions resulting from distillation cuts can subsequently be characterized. Simulated distillation by GC can be coupled with other analytical tools such as element-specific detectors or a mass spectrometer.

The separation of components in residues can be performed according to their solubility in solvents possessing increasing polarity or by chromatographic solvent elution after preliminary adsorption of the dissolved sample on a solid adsorbent. Among a wide variety of separation schemes, some have been standardized and the corresponding fractions are designated according to solubility: oils or maltenes, resins, and asphaltenes. The yield of the fractions is indicative of composition and behavior during further processing of residues.

Further characterization of fractions involves either chromatographic or spectroscopic techniques. Chromatographic methods include thin-layer, liquid-phase, and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

size-exclusion modes and provide information related with composition of fractions in terms of hydrocarbon group types (saturated, aromatics, polars) and molecular weight distribution. Spectroscopic methods [ultraviolet (UV), mass, NMR] are suitable for compositional evaluation.

Physical and analytical parameters can be correlated with the distribution of structural groups existing in the hydrocarbon fraction of residues: paraffinic, naphthenic, olefinic, and aromatic (single or fused rings).

1 INTRODUCTION

Fossil fuels, excluding natural gas, represent a challenge for the analytical chemist. The molecules that they contain range from light hydrocarbons to extremely complex macromolecular structures which are the memory or fingerprints of living matter origin. The complexity originates from the process of biological material decomposition and further maturation, which are responsible for the genesis of petroleum. Considered as a global degradation, this process would give rise to the formation of a limited range of products, but, simultaneously, reactive intermediates generated during the course of degradation may recombine or associate, giving rise to the production of new entities which may exhibit a higher degree of complexity than the corresponding fragments. Prior to further processing, crude oil is first subjected to atmospheric or reduced pressure distillation; this separation yields a distillate fraction – commercialized, after adequate treatment, as gasoline, kerosene, fuel oil, etc. – and a residue. Since the simplest, lighter constituents of petroleum have been separated by distillation, residues constitute the most complex fraction of crude oil.

Owing to the associated and micellar nature of residues, an exhaustive inventory of individual components is definitively beyond reach. Hence characterization methods will aim at giving an adequate description of the pertinent properties for a given purpose. Practically all of the existing analytical techniques have been applied, but the main, hardly solvable, problem is the intercorrelation of data obtained from distinct methods. Thus, rather than pursuing a hypothetical universal method, the analyst's strategy will consist in choosing the procedure that will be able to provide limited but pertinent information, at an acceptable cost, in a limited time.

Some easily obtained physical parameters correlate with chemical composition data. This is the case with specific gravity, viscosity, and the content of carbon residues obtained after pyrolysis. Standard test methods for the determination of these parameters have been developed.

The direct determination of compositional parameters such as aromatic carbon and hydrogen content is obtainable through NMR spectroscopy without previous fractionation. The quantitative distribution of saturated, aromatic, and polar compounds can be measured by TLC coupled with FID. The parameters obtained through various analytical techniques are generally hardly correlated since they refer to different units (e.g. mole %, weight %, or volume %) that cannot be interconverted.

A more detailed characterization requires previous separation for the preparation of simpler fractions. Although the definition of residues implies they have been produced after a preliminary distillation, short-path molecular distillation, simulated distillation by high-temperature GC and SFE apply to residues and provide the distribution of fractions possessing equivalent boiling points up to about 720 °C at atmospheric pressure. The fractions resulting from distillation cuts can subsequently be characterized. Simulated distillation by gas chromatography (Simdist GC) can be coupled with other analytical tools such as an element-specific detector or a mass spectrometer.

The separation of components in residues can be performed according to their solubility in solvents possessing increasing polarity or by chromatographic solvent elution after preliminary adsorption of the dissolved sample on a solid adsorbent. Among a wide variety of separation schemes, some have been standardized and the corresponding fractions are designated according to solubility: oils or maltenes, resins, and asphaltenes. The yield of the fractions is indicative of composition and behavior during further processing of residues.

Practically all currently practised analytical techniques are applied to the further characterization of fractions. Chromatographic methods include thin-layer, liquid-phase, and size-exclusion modes and provide information related to the composition of fractions in terms of hydrocarbon group types (saturated, aromatics, polars) and molecular weight distribution. Spectroscopic methods (UV, mass, NMR) are suitable for compositional evaluation.

Physical and analytical parameters can be correlated with the distribution of structural groups existing in the hydrocarbon fraction of residues: paraffinic, naphthenic, olefinic, and aromatic (single or fused rings).

2 COMPOSITION OF PETROLEUM

2.1 Definition of Residues

Petroleum is a mixture of liquid, solid, and gaseous hydrocarbons and other minor components, whose composition is extremely variable according to the origin of the sedimentary deposits from where it is extracted; variability in composition is also observed within a given oilfield. The hydrocarbon components of petroleum can be separated

into distinct products by atmospheric pressure distillation. The nondistillable fraction is named the atmospheric residue. The distillation end-point affects the properties of the resulting residue. The practical limit for atmospheric pressure distillation is about 360 °C, since thermal decomposition occurs at higher temperatures. During petroleum refining, the atmospheric residue is usually submitted to low-pressure distillation, yielding a distillate (vacuum gas oil) and a VR. The atmospheric pressure equivalent temperature of vacuum distillation is about 560 °C.

Residues are black, viscous material, liquid or almost solid at room temperature. The specific gravity varies from 0.95 to 1.02 and the pour point ranges between 10 and 50 °C. The yield of residue obtained from crude oil distillation is extremely variable: light oils produce as low as 10% VR, whereas some heavy oils can yield up to 70% VR. The major part of proven world petroleum reserves is composed of heavy oils and tar sand deposits. The progressive introduction of these new resources in refineries will cause an increase of the portion of residues generated by primary distillation, while the demand for light products tends to grow. As a consequence of this conflictive evolution of offer and demand, processes dealing with conversion of residues for the production of more valuable products such as transportation fuels will acquire increasing importance. In this context, a precise evaluation of the properties of residues in relation to processability will be required.

2.2 Constituents

The complexity of petroleum residues is related to the mechanisms of petroleum genesis. Crude oil is the product of decomposition of plant and animal debris in a marine environment, incorporated in sediments. Some of the constituents are directly correlated with living matter origin:

- paraffinic, naphthenic, terpene, and carotene hydrocarbons are widely distributed in living plants;
- the decarboxylation of fatty acids yields hydrocarbons;
- porphyrins (Ni or V complex of porphine derivatives) are considered as degradation products of chlorophyll and pigments.

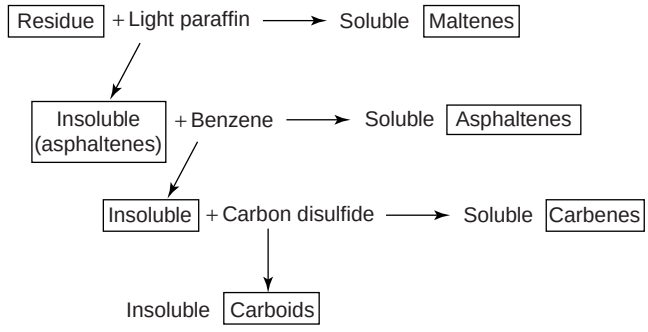

Figure 1 Definition of the major components obtained from residue fractionation.

Since the simplest, lighter constituents of petroleum have been separated by distillation, residues constitute the most complex fraction of crude oil. Attempts to fractionate residues further have been conducted since the early stages of petroleum science. As a general trend, extraction with solvents of increasing polarity provides fractions of increasing molecular weight and/or aromaticity. Solvent-extraction protocols have been developed on an empirical basis and a complex, confusing nomenclature has been derived. Some of the representative fractions obtained by solvent extraction are illustrated in Figure 1. The fraction that is soluble in light paraffins is named maltenes. Further fractionation of maltenes yields oils and resins (extracted with benzene). Asphaltenes, insoluble in paraffins, are soluble in benzene; the remaining insoluble fraction is named carbenes, which are further fractionated into carboids (insoluble in pyridine). The yield of the defined fractions is dependent on the operating conditions: nature of the paraffinic solvent, concentration, temperature. These components are far from representative of a chemical class and their elemental composition is extremely variable (Table 1). Based on the constituents derived from fractionation, petroleum residues may conceptually be visualized as a micellar suspension of asphaltenes and resins within a continuous oily phase. The distribution of the three main components (oil, resins, and asphaltenes) may be related to maturation of petroleum, and some correlations between reservoir characteristics and composition of crude oil have been observed.

Table 1 Yield and range of variation of selected properties of oils, resins and asphaltenes obtained from residues fractionation

Fraction	Yield (%)	Molecular weight	Atomic ratios			
			H/C	N/C	O/C	S/C
Oils	45–60	300–600	1.5–2	0	0	0.001–0.03
Resins	20–40	700–1000	1.3–1.7	0.002–0.008	0.002–0.05	0.003–0.04
Asphaltenes	10–30	2000–10 000	0.9–1.3	0.005–0.03	0.005–0.04	0.001–0.04

3 DETERMINATION OF SOME PHYSICAL PARAMETERS OF RESIDUES

As a raw material destined to yield secondary products, residues are generally subjected to further processing. According to compositional properties such as sulfur, asphaltene, or metal content, some types of residues may possess some economic benefit as a source of fuel, as a feedstock for cracking, or for any given purpose. Others may contain exceedingly low concentrations of oily components for the manufacture of lubricating oil, for instance, that may not be economically acceptable. It is therefore necessary to perform the selection of this raw material on the basis of the subsequent end product.

In view of the extreme complexity of petroleum, in-depth analytical evaluation of residues is an exceedingly difficult task. However, recourse to sophisticated analytical techniques is not necessary for obtaining the relevant properties related to a given use. This is the case with some easily obtained physical parameters that correlate with chemical composition data: specific gravity, viscosity, Conradson carbon, refractive index. Methods for the determination of essential parameters have been standardized by major standards organizations, and test methods have been developed and published by the American Society for Testing and Materials (ASTM), 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959, USA.

3.1 Specific Gravity

It was early recognized that density (relative to water) or specific gravity (sp gr expressed in $g\,cm^{-3}$) is roughly correlated with the expected yield of distillate obtainable from crude oil. Determination can be performed by means of a pycnometer or hydrometer. It is generally expressed as American Petroleum Institute (API) gravity, defined according to Equation (1):

$$\text{degree API} = \frac{141.5}{\text{sp gr at } 15\,^{\circ}\text{C}} - 131.5 \qquad (1)$$

Since residues are highly viscous at 15 °C, the determination may be performed at higher temperature and a correction for thermal expansion applied (ASTM D1217 and D1555).

Automated apparatus (ASTM D5002) for the determination of density are based upon the change in oscillating frequency of a cell containing the liquid to be evaluated, as compared with a reference liquid (generally water). Density is proportional to the square of the variation of oscillation frequency. It is expressed as Equation (2):

$$d_S = d_W + K(T_S^2 - T_W^2) \qquad (2)$$

where d_S is the density of the sample, d_W is the density of water at the temperature of measurement, K is an instrumental constant, and T_S and T_W are the period of vibration for the sample and for water, respectively. Provided that the temperature is controlled at $\pm 0.1\,^{\circ}C$, the observed reproducibility of determination with heavy samples such as residues is better than $0.005\,g\,cm^{-3}$.

This very straightforward, easily obtained property is correlated with sulfur, nitrogen, and asphaltenes content and Conradson carbon residue; as a general trend, higher heteroatoms and asphaltenes content result in increased specific gravity.

3.2 Viscosity

Similarly to density, viscosity is of considerable practical importance for the petroleum industry since it is directly related to the flow characteristics of oil. Viscosity is usually expressed in centipoise ($1\,cP = 10^{-3}\,Pa\,s$) and kinematic viscosity, which relates viscosity to specific gravity, is expressed in centistokes ($1\,cSt = 10^{-6}\,m^2\,s^{-1}$). The viscosity of residues ranges between 10^3 and 10^7 St.

The determination is based either on capillary instruments applying the Poiseuille equation and giving a direct measure of viscosity, or on calibrated orifice devices as the measure of flow time (t). Depending on the orifice geometry, several test methods are used and the corresponding scales of viscosity can be converted to kinematic viscosity (v) by applying the empirical Equation (3):

$$v = At + \frac{B}{t} \qquad (3)$$

Numerical values of A and B are given in Table 2 for commonly used viscosity scales. The variation of viscosity with temperature (viscosity index, ASTM D2270) is an important expression for the use of the oily fraction of residues as a basis for the elaboration of lubricating oils. It is related to compositional characteristics of residues since it has been observed that, at comparable viscosity, naphthenic oils have a higher viscosity index than paraffinic oils.

3.3 Carbon Residue

This parameter is a measure of the amount of carbonaceous residue left after pyrolysis under on inert gas above 500 °C. It is useful to characterize the depositing propensity of residual fuels used in burning equipment

Table 2 Coefficients for conversion of viscosity scales

Scale	A	B
Saybolt	0.0022	1.8
Redwood	0.0026	1.72
Engler	0.00147	3.74

or internal combustion engines. According to the test method, the numerical value is named Conradson carbon (ASTM D189), microcarbon residue (ASTM D4530) or Ramsbottom carbon (ASTM D524). Owing to the empirical character of the determination, data obtained from these methods are not strictly equivalent but can be interconnected.

Interesting relationships have been observed between Conradson carbon residue and compositional characteristics. Residual carbon is proportional to sulfur, nitrogen, and asphaltenes content. An inverse correlation is observed between carbon residue and hydrogen-to-carbon ratio.

4 CHEMICAL ANALYSIS WITHOUT PREVIOUS FRACTIONATION

4.1 Elemental Analysis

Generic methods for elemental analysis are generally applicable for residue evaluation. However, the commercial importance of fuel characterization has justified manufacturers developing dedicated instruments. The specifications and range of application of commercial equipment generally exceed the requirements of current standard test methods.

Carbon content is determined through combustion in oxygen and measurement of the evolved CO_2. Proper control of the combustion is of critical importance for precision. It includes the use of catalytically additive, programmed sequences of previous pyrolysis and combustion, and programmed temperature ramps, which are generally proprietary. Water produced during combustion is absorbed by magnesium perchlorate. According to the instrument design, the CO_2 produced may be determined gravimetrically after absorption in sodium hydroxide deposited on an inert support; alternatively, the content of CO_2 in the collected combustion gases can be determined by infrared (IR) absorption measurement.

Hydrogen content can be obtained simultaneously along with carbon determination by gravimetry of the absorbed water formed during combustion of the sample. Specific low-resolution NMR-based methods (ASTM D4808) have been developed for the measurement of total hydrogen in residues and good correlation is obtained between NMR data and combustion data. Provided that the specific gravity of the sample is previously determined with good accuracy, the routine NMR determination of total hydrogen content does not require weighing of the sample.

Nitrogen determination involves a previous combustion of the sample. According to the type of instrument, elemental nitrogen remaining after combustion is quantified by a thermal conductivity detector, after water and CO_2 have been absorbed and eliminated from the combustion gases. Alternatively, nitrogen may be catalytically oxidized during combustion to nitric oxide (NO); further reaction of the combustion gases with ozone produces excited nitrogen dioxide, which is quantified with a chemiluminescent detector.

Sulfur content determination is generally performed through combustion and further evaluation of the sulfur dioxide produced in the combustion gases. Distinct methods for SO_2 detection are implemented in commercial equipment: simultaneous determination of carbon and sulfur can be performed by IR absorption measurement; alternatively, UV fluorescence of SO_2 is employed in some types of apparatus; another option consists in the absorption of SO_2 in iodide solution followed by coulometric measurement of the iodine produced. This method possesses the distinct advantage that it is an absolute determination that does not require calibration with standards. The range of concentration attainable by both techniques greatly exceeds the sulfur content usually found in residues; the detection limit (DL) may be as low as 100 ppb while the maximum concentration is about 5%. Higher contents can be determined after previous dilution of the sample.

Energy-dispersive X-ray fluorescence (EDXRF) is currently the most popular method for sulfur determination in many petroleum-based matrices since it does not require special skill, and measurement is fully automated. However, caution must be exercised for its application to residues since the samples are generally too viscous to be poured directly into the analytical cells. Special cells, designed for handling pastry materials, may circumvent this difficulty. Dilution of the analyte with a solvent may cause some flocculation which affects the stability of measurement. This drawback may be partly resolved by a sample spinning device incorporated in some instruments.

Metal content (mainly nickel and vanadium) determination is easily performed by EDXRF. The state of the sample is also critical: homogeneity for diluted liquid samples must be preserved, and solid-state samples require a previous preparation of the exposed surface for obtaining reliable results. The elaboration of solid-state standards requires special skill. The determination of metal content can be performed by inductively coupled plasma atomic emission spectroscopy (ICPAES). It requires the previous dissolution of the sample, hence incomplete solubility may cause some errors due to segregation or flocculation of carboid fractions.

4.2 Evaluation Without Previous Fractionation

Analytical techniques to be used for the characterization of residues must be fast, (i.e. applicable without previous fractionation), repeatable, and quantitative. They must

also provide information regarding the distribution of chemical families, alternatively referred to as hydrocarbon group type analysis (HGTA). A rigorous analysis of residues typically involves at least one chromatographic separation step during sample preparation, which is followed by consecutive characterization of the fractions. These fractionation steps are lengthy and tedious, hence many attempts have been made to obtain information related to composition in terms of distribution of saturated, aromatic, or polar compounds, without previous fractionation. Some techniques may provide HGTA data without any previous separation. This is the case with TLC and or ^{13}C-NMR spectroscopy.

Repeatability of analytical techniques is generally considered as a criterion for protocol selection, along with reliable quantitative determination. It must be pointed out that, in the absence of detailed information concerning individual constituents of complex mixtures, quantitation is related to internal consistency. Effectively, the definition of the considered hydrocarbon groups is often ambiguous; for example, if aromatic and saturated hydrocarbons are distinguished by a given analytical method, the distribution of hydroaromatic species between these two classes will depend on the chosen protocol. Hence internal consistency does not imply the exactness of data, and comparisons with additional analytical techniques must be made in order to assess the meaningfulness of one particular technique.

Another kind of difficulty arises from tentative comparisons of data obtained with distinct tools. It comes from the fact that the results provided by different analytical techniques are expressed in distinct units: some are expressed on a weight basis, others on a molar basis. Direct comparison of the results are meaningless, since unit conversion relies upon some external hypothesis concerning the composition of the analyte. Therefore, although analytical data are inter-related, the correlation between two distinct techniques relies upon a third external data or upon some hypothesis concerning the molecular structure.

^{13}C- and ^{1}H-NMR are the methods of choice for aromatic carbon and aromatic hydrogen content determination (ASTM D5292). Samples need previous homogenization before dissolution in deuterated chloroform or in a mixture of carbon disulfide and deuterated chloroform. The aromatic hydrogen content (weight %) is measured by comparing the integrals of the signals in the ^{1}H-NMR spectrum between 5.0 and 10.0 ppm for aromatic hydrogen and between -0.5 and 5.0 ppm for aliphatic hydrogen. Similarly, the aromatic carbon content (weight %) is measured by comparing the integral of the signal in the ^{13}C-NMR spectra between 100 and 170 ppm for aromatic carbon and -10 and 70 ppm for aliphatic carbon, after correction for the aliphatic fraction of reference material. DLs are typically 0.1 mol% aromatic hydrogen atoms and 0.5 mol% aromatic carbon atoms. The olefin and phenol content must be below 1%.

Chromatographic techniques allow the separation and quantitation of hydrocarbon groups, after an adequate calibration procedure. High-performance liquid chromatography (HPLC) is not directly applicable to residues owing to the possible irreversible adsorption of heavy and/or polar compounds on the columns and, hence, incomplete elution.

TLC has long been applied to the separation of complex materials. It requires very simple equipment but cannot provide quantitative information. A substantial improvement of TLC has been accomplished by coupling with quantitative FID. Instead of performing the chromatographic separation on glass plates covered with a layer of porous medium, generally alumina, the sample is deposited on silica rods about 1 mm in diameter and 10 cm long, coated with a separation medium. After elution with solvents, the rods are scanned over the flame of the detector, which provides a quantitative signal. Hence this technique grants the possibility of quantifying the uneluted heavy and polar products, in addition to increasing the speed of the analysis. Hydrocarbon groups are classified according to elution sequences with selected solvent systems, a practice which allows the determination of saturated, alkylaromatic, aromatic, and polar compounds. A typical separation of a residual oil is illustrated in Figure 2.

As with other chromatographic methods, quantitative results are based upon absolute calibration, which requires a previous preparative chromatographic step in order to obtain external standards. However, a rapid

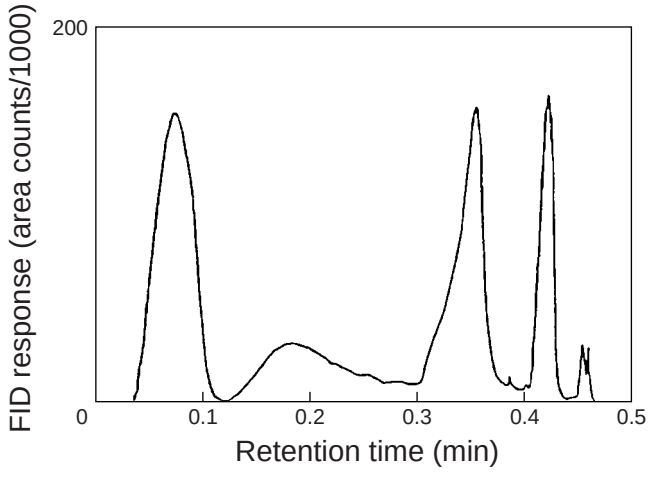

Figure 2 Separation of a residual oil by TLC/FID using a consecutive sequence of elution with *n*-hexane, toluene, and dichloromethane–methanol (95 : 5 v/v). (Reproduced from the *Journal of Chromatographic Science* by permission of Preston Publications, a Division of Preston Industries, Inc.)

internal calibration procedure has been developed.[1] Results are obtained on a weight % basis.

Small-angle X-ray scattering has been applied to the determination of the mean molecular weight of residues, since this parameter is related to the radius of gyration obtained from diffused intensity for diffusion angles smaller than 4° (2θ). This technique provides good agreement with molecular-weight data for asphaltene fractions obtained by gel permeation chromatography (GPC). In the case of raw and hydrotreated residues, polydispersity is evidenced.[2]

5 FRACTIONATION

The purpose of fractionation is the production of a significant amount of different fractions of the residue for further characterization. The composition of the fractions is supposedly simpler than that of the whole sample, thus allowing more specific analytical techniques to be applied. Owing to the extreme complexity of residues, it is clear that an infinite variety of fractionation schemes can be imagined and that the definition and properties of the resulting fractions will depend strongly on the applied procedure. The general goal is to obtain a classification based upon some difference in chemical properties. However, the presence of many multifunctional compounds and the wide distribution of molecular weights are the cause of overlapping of fractions. Hence, owing to the impossibility of defining an absolute scheme of separation, reproducibility must be considered as the only pertinent criterion. As a consequence, the standardization of fractionation protocols is of primary importance, although they are partly based upon empirical practice.

5.1 Solvent Extraction

Numerous separation schemes have been developed for the preparation of fractions from residues. Some proposed fractionations are illustrated in Figure 3(a) and (b). No standard methods have been elaborated. The only exception concerns the precipitation of asphaltenes with *n*-heptane (ASTM D3279). The properties of the generic fractions (asphaltenes, maltenes, resins, wax, and oils) are obviously dependent upon the chosen solvent sequence. For instance, the yield and properties of precipitated asphaltenes vary with the number of carbon atoms in the paraffinic solvent, temperature, relative volume of solvent, and contact time.[3] Consequently, the terminology of the fractions obtained cannot be considered as representative of a class of compounds. The positive aspect of solvent fractionation is the complete recovery of material without loss through irreversible adsorption, as occurs in chromatographic separation techniques. The use of large volumes of solvents and the tedious, lengthy character of the separation procedure advocate the use of alternative methods.

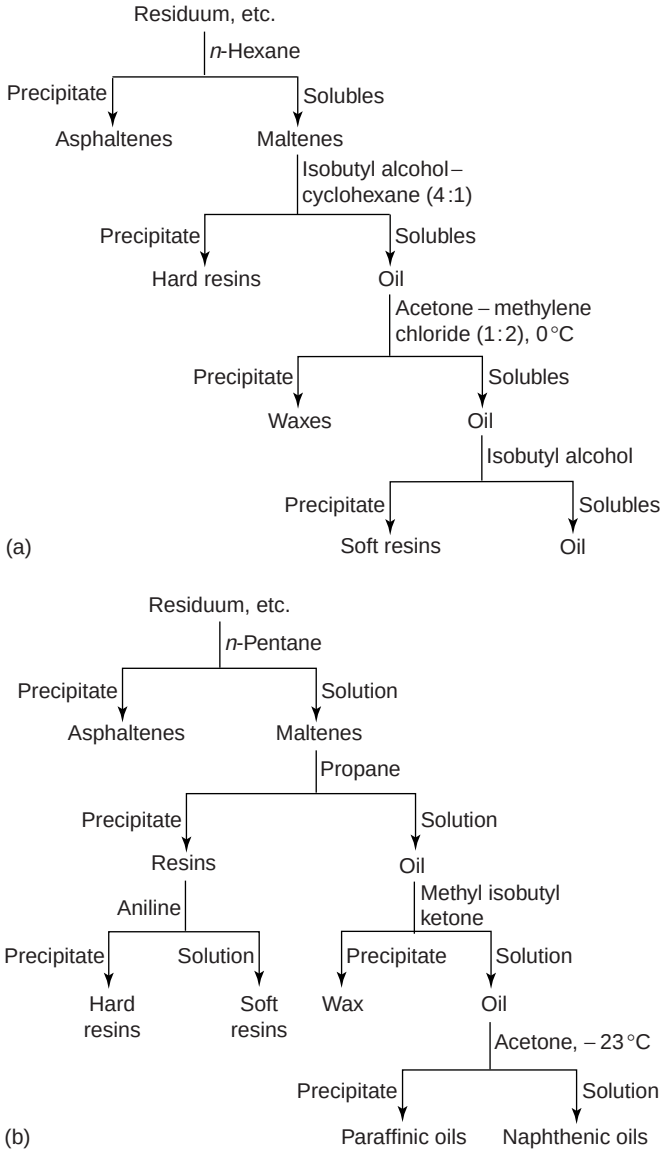

Figure 3 Two solvent fractionation schemes. (Reprinted from J.G. Speight, *The Chemistry and Technology of Petroleum*, Marcel Dekker, New York, 327–328, 1991, by courtesy of Marcel Dekker Inc.)

5.2 Adsorption Chromatography

Chromatographic procedures involve the previous adsorption of a solution of residue on a solid adsorbent. Consecutive elution with solvents yields fractions, characterized by the polarity of the corresponding elution solvent. Since asphaltenes are irreversibly adsorbed, prior separation must be considered, applying a standard

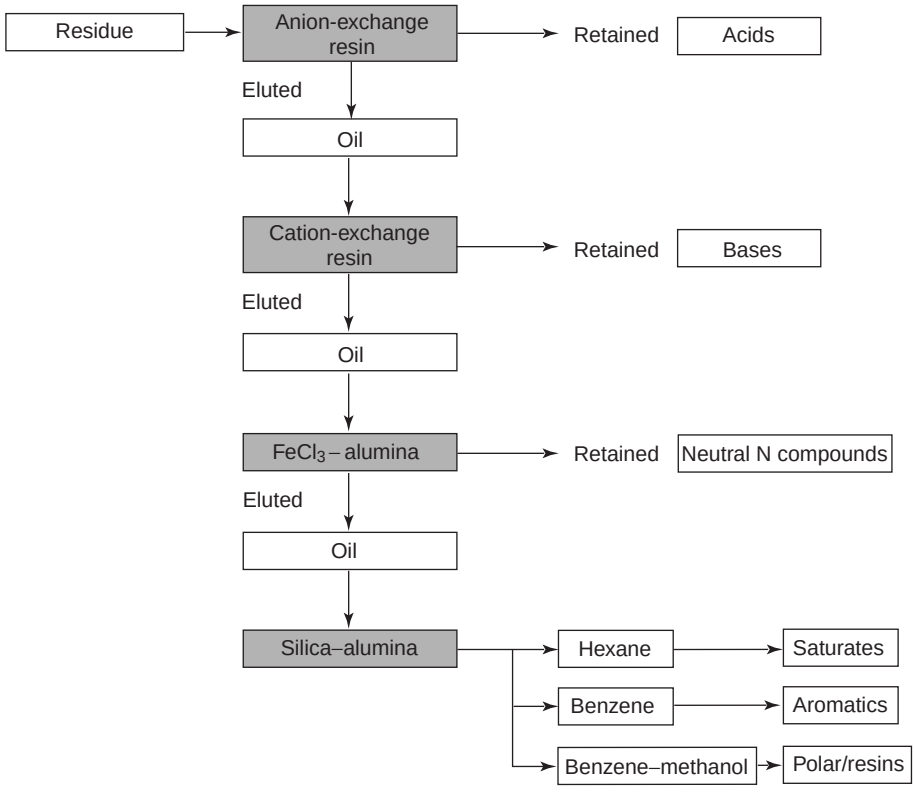

Figure 4 The USBM/API separation procedure.

method. The textural and surface properties of the solid adsorbent directly influence the separation process, hence reproducibility of the solid phase must be guaranteed for the credibility of the analytical method.

The United States Bureau of Mines/American Petroleum Institute (USBM/API) method, involving the separation of acidic and basic compounds on ion-exchange resins, is described in Figure 4. The neutral eluting fraction is further fractionated by complexation with $FeCl_3$, yielding neutral nitrogen compounds and an oil fraction, consecutively separated on a silica–alumina column into saturates and aromatics.

The saturates–aromatics–resins–asphaltenes (SARA) method consists in a simplification of the USBM/API scheme. Instead of using separate columns, the consecutive ion-exchange resins and complexation materials are packaged in a single column. Two methods have been standardized (ASTM D2007 and D4124). Various separation, chemical treatment, and complexation procedures have been described.[3] The main drawback of all of these fractionations is that they are labor intensive and expensive, since they require large quantities of solvents and column packing. Furthermore, they were designed for preparative purpose in view of the subsequent characterization of the extracted fractions by analytical techniques. The recent evolution of analytical instrumentation towards ever more miniaturized systems eliminates the need for the production of large quantities of fractions since microgram quantities are now sufficient for most analytical determinations. As a consequence, chromatographic methods are moving from the preparative to the analytical scale.

Open-column liquid chromatography (LC) is advantageously replaced by medium- or high-pressure HPLC, which gives a much better resolution, requires smaller amounts of sample and is faster. Heavy and polar compounds are irreversibly adsorbed, causing column deterioration and incomplete elution, but this problem can be circumvented by an appropriate choice of stationary and mobile phases and column backflushing. The detection and quantitation of eluted fractions can be performed with a variety of detectors based upon refractive index, UV absorption, evaporative light scattering, etc. The coupling of HPLC with spectroscopic techniques can provide the direct analytical characterization of eluting fractions. Some hyphenated methods are described later.

5.3 Short-path Molecular Distillation

Although the definition of residues implies they have been produced after a preliminary distillation, short-path molecular distillation can yield additional distillate from

VRs.[4] In this low-pressure (10^{-2}–10^{-4} mmHg) distillation apparatus, the feed is introduced continuously and spread out by means of rotating paddles on a cylindrical warm wall. Partial vaporization occurs and, at a close distance, a coaxial condenser allows the vaporized fraction to be collected. The remaining residue is recovered in a separate collector. At the considered pressure, the distance between the cold and hot walls of the device is comparable to the mean molecular path of hydrocarbon molecules. Therefore, the separation process does not correspond to liquid–vapor equilibrium. The control of the hot-wall temperature up to 350 °C allows consecutive fractions to be prepared; it allows the production of 30–50% of distillate fraction from a VR. The mean molecular weight of the heavier distillate fraction may be as high as 800, corresponding to a 700 °C equivalent boiling point at atmospheric pressure.

5.4 Simulated Distillation by High-temperature Gas Chromatography

Simdist GC consists in the injection of a diluted sample into a precolumn (tubing diameter equivalent to column diameter). The column temperature is programmed and the vapors are continuously directed to a nonpolar gas–liquid capillary chromatographic column by an inert carrier gas flow and coupled with FID. The column performance in terms of resolution is not critical since the objective is not to perform a detailed separation but to obtain a global signal. Initially designed for distillate characterization, the range of application has been considerably extended since the introduction of stationary phases withstanding temperatures up to 440 °C, which allow the elution of hydrocarbons containing up to 120 carbon atoms (bp 720 °C). The column dimensions are usually 5 m × 0.5 mm i.d. Aluminum-coated silica columns withstand high temperatures but become breakable; metal columns are preferred. Since the FID response is directly proportional to the mass flow of carbon atoms eluted from the column, the resulting signal characterizes the differential flow of eluting material. The blank signal, resulting from the injection of pure solvent with the same temperature program, is subtracted from the analytical signal, thus producing a distribution of eluted mass % as a function of retention time. This scale can be converted to an equivalent atmospheric boiling point of *n*-alkanes through calibration with standard polyethylene waxes. Uneluted material is trapped in the precolumn section, which may be replaced. Quantitative determination of the eluting fraction requires the use of an internal standard in order to account for the uneluted portion of the sample. The deconvolution of simulated distillation curves can provide information related to the processing of residues.[5]

5.5 Supercritical Fluid Extraction

Supercritical fluids possess unique solvation properties and can be used as a mobile phase in chromatographic processes using either packed HPLC-type columns or capillary GC-type columns. According to the scale and objective of the separation (analytical or preparative), the process is referred to as supercritical fluid chromatography (SFC) or SFE. The most common and attractive fluid is CO_2, which is nontoxic and can be easily separated from the eluting material. Similarly to other chromatographic techniques, SFC can be coupled with standard detectors (FID, UV, etc.) and hyphenated with chromatographic or spectroscopic techniques. It has been applied to hydrocarbon group determination in residues and a comparative study with LC illustrated good quantitation for the evaluation of saturated, aromatic, and polar compounds.[6] Alternatively, simulated distillation can be achieved by SFC at moderate temperatures (150 °C) with capillary columns and gives a similar boiling-point distribution to high-temperature GC.[7]

6 CHARACTERIZATION OF FRACTIONS

6.1 Determination of Molecular Weight

A distinction must be made between colligative methods (based upon the modification of physical properties of solutions) and hydrodynamic methods. The former give access to the number-average molecular weight, W_n, and the latter, concerned with the apparent molecular size, allow the value of the weight-average molecular weight, W_w, to be determined.

In the former group of methods, cryometry and ebullioscopy have now largely been replaced by vapor pressure osmometry, which is easily automated. Instead of a direct determination of osmometric pressure, the flow of vapor from the pure solvent to the solution of sample is estimated through its thermal effect. The range of application is approximately 500–5000 (ASTM D2503).

The latter group of methods includes light scattering and viscosity. When a solution is illuminated by monochromatic light, the diffused intensity, extrapolated to zero concentration of solute, is a function of the radius of gyration of dissolved molecules, which in turn can be correlated with the mean molecular weight, introducing the variation of refractive index with concentration. The practical range of application is limited to molecular weights > 500.

The Mark–Houwink law establishes the relationship between intrinsic viscosity $[\eta]$ and molecular weight, W_w (Equation 4):

$$[\eta] = KW_w^{\alpha} \tag{4}$$

The coefficients K and α are related to molecular geometry, hydrodynamic behavior and solvent properties. Hence this is an indirect determination of molecular weight which can be applied over a wide range. A simplified standard method (ASTM D2502) has been proposed for petroleum-based fractions in a limited molecular-weight range (250–700). The ratio of weight-average to number-average molecular weight (W_w/W_n) is termed the polydispersity index and characterizes the heterogeneity of the sample. For instance, values as high as 4 can be observed for nonfractionated heavy crude samples, whilst that of asphaltene fractions is ca. 1.3. This indicates that the determination of a *mean* value of a given property of a complex, extremely dispersed sample is meaningless and that the information it provides is too vague to be useful.

As a consequence of the complexity of residues and fractions, the distribution of molecular weights is more meaningful than the mean value. It can be obtained by size-exclusion chromatography (SEC). In this technique, molecules are separated according to size variation in solution through permeation into the porous structure of a solid. The column packing generally consists of small particles (5–100 μm) of styrene–divinylbenzene copolymers possessing a pore structure in the range 10–100 nm. Elution is performed with solvents such as toluene, tetrahydrofuran, and chloroform. A large variety of detection methods can be applied, e.g. refractive index, UV absorption, fluorescence, and evaporative light scattering. Calibration of molecular-weight distribution is performed with polystyrene standards or normal paraffins for smaller values of W.

6.2 Determination of Aromatics and Saturates

The distribution and content of aromatic hydrocarbons in oil and oil/aromatic fractions is easily obtainable by UV absorption spectrophotometry, the experimental background of which is well established.[8] The exploitation of UV absorption spectra is based upon multicomponent analysis, which in turn relies upon the additive contribution of each component to the total absorbance at selected wavelengths. Typical attributions of wavelength ranges are as follows: monoaromatics, 195–200; diaromatics, 225–235; triaromatics, 250–258; chrysenes, 260–269; tetraphenes, 280–289; pyrenes, 332–350; and heavier polyaromatics, 385 nm.

Calibration consists in establishing a calculation matrix based upon known samples. Since no clear relationship exists between the molecular structure of a given compound and absorbance, many interferences may affect quantitative analysis. A correction for some identified interfering classes of compounds is possible in the calculation matrix. This is the case for heterocyclic compounds and pyrenes. However, other spectral interferences are more subtle and may be overlooked in commonly practised analytical methods. For instance, the interference caused by benzofluoranthenes is structure dependent: *g*, *h*, and *i* isomers interfere, whereas the *j* isomer does not. As a consequence of variable interferences, UV spectrophotometry may be considered as a fast, comparative method for the continuous monitoring of samples whose composition varies within restricted limits. Results are expressed as mole % aromatic content.

Low-resolution, high-ionization-voltage mass spectrometry (MS) has long been used for the quantitative determination of saturates and aromatics in oil fractions[9] (ASTM D2786). The method relies upon the resolution of a calculation matrix which accounts for the contribution of different classes of aliphatic and aromatic compounds to the production of selected ion fragments. A different approach consists in the use of a low ionization potential (field ionization) in order to maximize the yield of molecular ions. However, since numerous interferences occur, recourse to high resolution is necessary in order to discriminate compounds possessing an equivalent nominal molecular weight but different elemental compositions. The scope of the technique is limited to those products that are volatile at the maximum allowable source temperature (usually 350 °C), corresponding to a 550 °C atmospheric boiling temperature.

^{13}C- and ^{1}H-NMR spectroscopy (section 4.2) obviously can be applied to fractionated products for the determination of aromatic and aliphatic carbon and hydrogen.

6.3 Hyphenated Techniques

As mentioned earlier, an infinite variety of fractionation schemes and corresponding products can be designed. It therefore seems more efficient to proceed to a continuous determination of one specific property of fractions along with the separation process. This approach constitutes the basis for numerous combinations of analytical methods coupled with separation techniques.

Typically, a chromatographic method may be coupled with another chromatographic technique; numerous methods use the coupling of HPLC columns. The combination of normal-phase, reversed-phase, charge-transfer, and size-exclusion modes can be used to separate fractions of different polarity or size, and subsequently proceed to a further separation based upon another process. Similarly, the combination of HPLC and GC can be contemplated for volatile fractions. Chromatography can also be coupled with a specific detector, either an element-specific or multielement detector (e.g. atomic emission detector), or with a spectroscopic technique, e.g. TLC coupled with a UV detector or GC coupled with MS.

6.4 Determination of Individual Compounds

The presence of some target compounds in residue fractions deserves particular attention. This is the case with carcinogenic polycyclic aromatic hydrocarbons (PAHs), which can be concentrated on reversed-phase columns and further quantified by taking advantage of the selectivity of fluorescence detectors.

Metalloporphyrin complexes can be used as potential fingerprints of the origin of crude oils. They are selectively extracted from neutral heavy fractions with acetonitrile and may be further separated into nickel and vanadium groups by HPLC.

7 RATIONALIZATION OF ANALYTICAL DATA

The preceding fractionation and characterization techniques roughly distinguish paraffinic, naphthenic, and aromatic structures in residues. Since establishing an exhaustive inventory of individual molecular components in residues is a hopeless endeavor, structural group analysis (SGA) will attempt to describe the statistical distribution of structural elements, irrespective of the way in which they are combined in molecules, thus providing information intermediate between elemental and molecular analysis. Molecules in heavy fractions may simultaneously contain aromatic and naphthenic rings and paraffinic chains. Hence they cannot be classified according to the previously described separation schemes and they are better described in terms of structural fragment composition.

7.1 Determination of Structural Parameters from Physical Properties

In the early age of the petroleum industry, several simple methods, based upon the analysis of numerous samples, were developed for determining the distribution of carbon atoms (number of rings, percentage of carbon in rings) and hydrogen content from elemental analysis, refractive index, and density determination. For example, the $n–d–M$ method (ASTM D3298) is based upon the measurement of refractive index, density, and molecular weight at two temperatures. It allows the calculation of the percentage of aromatic carbon atoms, carbon atoms in ring structures, mean number of rings, and mean number of aromatic rings.

Numerous variations of these methods have been proposed, e.g. dispersion–refraction, density temperature coefficient, molecular weight–refractive index, and viscosity–density–refractive index. Owing to the empirical character of these methods and the need for some tedious measurements, spectroscopy-based SGA methods are currently preferred.

Table 3 ^{1}H- and ^{13}C-NMR band assignments

Structure type	Chemical shift (ppm from TMS)
Aromatic hydrogen	9.0–5.0
Hydrogen in CH, CH_2, CH_3, α to aromatic ring	5.0–1.9
Hydrogen in CH, CH_2, β or further to aromatic ring	1.9–1.0
Hydrogen in CH_3 γ or further to aromatic ring	1.0–0.5
Aromatic C	160–60
CH	60–37
CH_2 β or further to aromatic ring	37–24
CH_2 next to methyl or naphthenic	24–22.5
CH_3 α to aromatic ring	22.5–20
CH_3 β or further to aromatic ring	20.5–0

TMS, trimethylsilane.

7.2 Determination of Structural Parameters from Spectroscopic Methods

^{1}H-NMR and later ^{13}C-NMR have been extensively used to provide average molecule parameters as a description of petroleum derived fractions. The assignment of ^{1}H- and ^{13}C-NMR signals can be more refined than the simple distinction between aliphatic and aromatic entities presented in section 4.2. A summary of structure assignments is presented in Table 3. Thus, combining ^{1}H- and ^{13}C-NMR data, the following distribution of carbon atoms per mean molecule can be deduced: saturated carbon to aromatic ring; peripheral carbon in condensed ring systems; internal carbon in condensed ring; paraffinic methylene carbon further than α to aromatic ring; and paraffinic methyl carbon.

Additional parameters can be derived by combining NMR data and molecular-weight measurement, such as the degree of substitution of aromatic structures, the number of aromatic rings per molecule, and the number and length of alkyl substituents.[10]

This approach gives average parameters that may not provide information about the actual distribution of components in the mixture. The functional group analysis approach developed by Petrakis et al.[11] is based on the premise that although the molecules in heavy fractions may be extremely complex, they are composed of a limited number of functional groups. The chosen set must account for experimental data, and it must be as concise as possible. An illustrative list of functional groups is described in Table 4. Additional groups may be accounted for, provided that their contribution is relevant and in agreement with analytical data (e.g. oxygen or

Table 4 Definition of functional groups

Benzene		Naphthalene	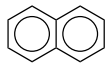
Phenanthrene	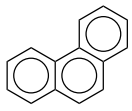	Pyrene	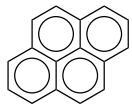
Dibenzothiophene	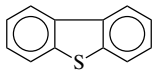	Ether bridge	
Biphenyl bridge	$C_{ar}-C_{ar}$	Aromatic methyl	CH_3-C_{ar}
Aromatic α methyl		Aliphatic methyl	CH_3-C_{al}
Methylene bridge		Hydroaromatic	
Two ring hydroaromatic	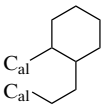	Two ring hydroaromatic bridge	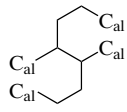

nitrogen atoms). The concentration of the specified set of functional groups is introduced into a set of balance equations concerning the different structures (carbon and hydrogen atoms at different positions) derived from experimental data, with their respective stoichiometric coefficients. The resulting information has a definite superiority over average values of structural groups in the sense that it can be used for the quantitative description of complex samples and as a tool for elucidating reaction pathways. The contribution of estimated functional groups to thermodynamic properties, such as heat capacity,[12] can be evaluated.

ABBREVIATIONS AND ACRONYMS

API	American Petroleum Institute
ASTM	American Society for Testing and Materials
DL	Detection Limit
EDXRF	Energy-dispersive X-ray Fluorescence
FID	Flame Ionization Detection
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HGTA	Hydrocarbon Group Type Analysis
HPLC	High-performance Liquid Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PAH	Polycyclic Aromatic Hydrocarbon
SARA	Saturates–Aromatics–Resins–Asphaltenes
SEC	Size-exclusion Chromatography
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SGA	Structural Group Analysis
Simdist GC	Simulated Distillation by Gas Chromatography
TLC	Thin-layer Chromatography
TMS	Trimethylsilane
USBM/API	United States Bureau of Mines/ American Petroleum Institute
UV	Ultraviolet
VR	Vacuum Residue

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)** High-temperature Simulated Distillation Applications in Petroleum Characterization • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of • Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels • Nuclear Magnetic Resonance Characterization of Petroleum

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Liquid Chromatography **(Volume 13)**
Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis

REFERENCES

1. J. Vela, L. Membrado, V.L. Cebolla, A.C. Ferrando, 'Suitability of Thin-layer Chromatography/Flame Ionization Detection with Regard to Quantitative Characterization of Different Fossil Fuel Products. II. Calibration Methods Concerning Quantitative Hydrocarbon Group Type Analysis', *J. Chromatogr. Sci.*, **36**, 487–495 (1998).
2. D. Espinat, D. Tchoubar, R. Boulet, E. Freund, 'Etude des Produits Lourds par Diffusion Centrale des Rayons X', in *Characterization of Heavy Crude Oils and Petroleum Residues*, eds, Centre National de la Recherche Scientifique, Elf, Institute Français du Pétrole, Total, Technip, Paris, 147–152, 1984.
3. J.G. Speight, *The Chemistry and Technology of Petroleum*', Marcel Dekker, New York, 316–325, 1991.
4. W. Fisher, 'Labor-destillation Hochsiedender Mineralölantelle', *Erdöl Kohle Erdgas Petrochem.*, **36**, 1–3 (1983).
5. R. Bacaud, L. Rouleau, B. Bacaud, 'Modeling Simulated Distillation: a Tool for the Evaluation of Hydroconverted Petroleum Residues', *Energy Fuels*, **10**, 915–920 (1996).
6. E. Lundanes, T. Greibrokk 'Group Separation of Oil Residues by Supercritical Fluid Chromatography', *J. Chromatogr.*, **349**, 439–446 (1985).
7. J.S. Thomson, A.F. Rynaski, 'Simulated Distillation of Wax Samples Using Supercritical Fluid and High Temperature Gas Chromatography', *J. High Resolut. Chromatogr.*, **15**(4), 227–234 (1992).
8. R.A. Burdett, L.W. Taylor, L.C. Jones, Jr, 'Determination of Aromatic Hydrocarbons in Lubricating Oil Fractions by Far Ultraviolet Absorption Spectroscopy', Molecular Spectroscopy, Reprint, Conference of Institute of Petroleum, London, 30–41, 1955.
9. C.J. Robinson, 'Low Resolution Mass Spectrometric Determination of Aromatics and Saturates in Petroleum Fractions', *Anal. Chem.*, **43**(11), 1425–1434 (1971).
10. D.R. Clutter, L. Petrakis, R.L. Stenger, Jr, R.K. Jensen, 'Nuclear Magnetic Resonance Spectrometry of Petroleum Fractions', *Anal. Chem.*, **44**(8), 1395–1405 (1972).
11. L. Petrakis, D.T. Allen, G.R. Gavalas, B.C. Gates, 'Analysis of Synthetic Fuels for Functional Group Determination', *Anal. Chem.*, **55**(9), 1557–1564 (1983).
12. T.T. Le, D.T. Allen, 'Property Estimation Using Structural Characterizations', *Fuel*, **64**, 1754–1759 (1985).

Refractive Index Technology as a Real Time Viscosity Technique[a]

John G. Groetsch, Jr
K-Patents, Inc., Naperville, USA

1 Introduction 1
2 Viscosity of Lubricating Oil 1
3 Basic Lube Oil Processing 1
4 Refractive Index Theory 2
5 In-line Refractometer 3
6 Case Studies 4
7 Conclusion 5
Abbreviations and Acronyms 5
Related Articles 5
References 5

This article discusses productivity enhancement which involves using an on-line digital process refractometer to determine the viscosity index (VI) of lube oil stocks. Since lube oils are Newtonian fluids, refractive index (RI) can be related to VI. A real-time continuous measurement of the RI has eliminated sampling, improved product consistency, reduced waste and has helped to optimize the process.

1 INTRODUCTION

In lube oil production the measurement and control of viscosity is vital. Goals such as uniform product quality and minimum waste can be achieved by continuous monitoring. Unlike periodical sampling, continuous monitoring provides instant feedback about changes in the process. This instant feedback is then used to control the process in real time.

2 VISCOSITY OF LUBRICATING OIL

Viscosity is the internal resistance to a shear force and is a complex property of a fluid. It is, however, a measurable property. The dynamic viscosity of a fluid will change with temperature and the way in which the shear is applied. Thicker fluids have a higher viscosity. Dynamic viscosity is defined by Equation (1):

$$\mu = \frac{\tau}{D} \tag{1}$$

where μ is the dynamic viscosity ($N\,s\,m^{-2}$ or poise), τ is the shear stress ($N\,m^{-2}$) and D is the shear rate (s^{-1}).

Fluids can be divided into two groups: Newtonian and non-Newtonian (see Figure 1). A Newtonian fluid is one in which the viscosity is independent of the shear rate. Examples of Newtonian fluids include water and oils. A non-Newtonian fluid is one in which the viscosity will differ depending upon the shear rate used in making the measurement. There are several types of non-Newtonian fluid. In order to obtain an accurate reading in a non-Newtonian fluid, the shear rate at which the measurement is made must be specified.

The kinematic viscosity of a fluid is the ratio of dynamic viscosity to density. Kinematic viscosity (ν) is expressed in either centistokes (cSt) or millimeters squared per second ($mm^2\,s^{-1}$) and is defined by Equation (2):

$$\nu = \frac{\mu}{\rho} \tag{2}$$

where ρ is the density ($kg\,m^{-3}$).

Lubricating oils with higher kinematic viscosity create a thicker film of the oil that will cling to a metal surface. Depending on the end-use of the lube oil, it may need to be very thin and free flowing or thick with a high resistance to flow. The rate of change of kinematic viscosity with temperature is the VI; it is a unitless number. A high VI indicates a smaller decrease in the kinematic viscosity with increasing temperature (see Table 1).

3 BASIC LUBE OIL PROCESSING

Crude oil is a mixture of various types of hydrocarbon. The crude oil type influences the VI of lube oil that can be produced. Naphthenic crude oils contain cyclic compounds and are used to produce low to medium VI base oils. High VI base oils are commonly made from paraffinic crude oils that contain mostly straight-chained hydrocarbons. The challenge to the refiner is the selection and blending of the crude type to optimize the production of the desired lube oil.

This crude oil mixture is first separated into useful fractions by thermal distillation. The products from thermal distillation are enhanced later by other unit processes such as catalytic cracking, hydrotreating, dewaxing and solvent extraction. Lube oil stocks contain wax and other contaminants that need removal before blending into products such as motor oils. This further processing helps to maximize the desired lube oil production.

[a] Copyright © ISA. All right reserved. Reprinted with permission.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

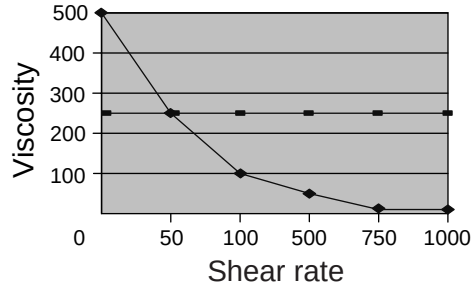

Figure 1 Newtonian and non-Newtonian fluids; non-Newtonian pseudoplastic, ▬ Newtonian.

The propane extraction process removes undesired components. Propane is unique because it extracts both on the basis of molecular weight and by chemical structure. The propane is recovered with the resulting products being asphalt and deasphalted oil feed for the solvent extraction process.

Extractive solvent processing follows propane deasphalting. The most common extractive solvents used are furfural and phenol. The purpose of solvent extraction is to improve the VI, oxidation resistance and color of the lube base stock. After the solvent extraction process is complete, the resulting waxy stream is further processed either chemically or by catalytic methods.

There are several types of dewaxing solvent used in the refining industry. The most common solvent is MEK (methyl ethyl ketone). Most refiners charge the stream produced from the furfural or phenol refining unit to the dewaxing unit.

The types of result achieved in catalytic treatment depend on the type of catalyst used and the severity of the treatment. The pour point of a hydrocarbon liquid is the temperature at which it will no lower flow. Improving pour point is best achieved using a single catalyst. For color improvement and oxygen stability more than one catalyst may be required. Color improvement is directly related to the type of catalyst and operating conditions of the unit.

A summary of the various types of lube oil processes is given in Table 2.

The use of a reliable in-line measurement technique can eliminate sampling and time-consuming laboratory testing for the determination of the VI of oil. The laboratory method for the determination of the VI of lube oil is the American Society for Testing and Materials (ASTM) D 2270-93, the VI is the rate of change of the measured kinematic viscosity. For the calculation of a VI when the kinematic viscosity at 100 °C is less than 70 cSt Equation (3) is used:

$$\mathrm{VI} = \frac{L - U}{L - H} \times 100 \tag{3}$$

where U is the kinematic viscosity at 40 °C ($mm^2\,s^{-1}$ or cSt), L is the kinematic viscosity at 40 °C of an oil of 0 VI (from Table 1 of ASTM D 2270-93) and H is the kinematic viscosity at 40 °C of an oil of 100 VI (from Table 1 of ASTM D 2270-93). For ν values greater than 70 cSt the ASTM method uses another equation.

The relationship between RI and VI for a given crude or crude blend is established by laboratory testing and then confirmed by sampling at a prescribed interval. Based on the results of the laboratory tests, ASTM D 1747-89, the refinery can set target RI values for the process. The final quality or final VI of the lube oil can then be controlled and measured by the in-line refractometer.

Table 2 Processes for finishing lube oils

Process	Function
Solvent (propane) deasphalting	Reduces carbon and sludge formation
Solvent extraction	Improves VI
Solvent dewaxing	Lowers pour point
Catalytic hydrotreating	Improves color and oxygen stability
	Lowers organic acidity

4 REFRACTIVE INDEX THEORY

RI is a measure of the amount of bending or refraction that occurs when light passes from one medium to another. This is caused by the change in the speed of light in each medium. In addition to varying with the medium, the response changes with the concentration of the dissolved solids. The mathematical equation for this principle is known as ''Snell's law'' defined as, Equation (4):

$$\sin i = n \sin r \tag{4}$$

Table 1 Typical values of VI for lubricating oils

Oil type	SAE	°API	Viscosity (cSt) at 40 °C	VI
Gear lubricants	90	28	150–200	95–98
Motor oils	10W–40	30	70–75	120–140
Synthetic oils	5W–40	33	80–90	150–160

SAE, Society of Automotive Engineers; °API, American Petroleum Institute gravity.

where i is the angle of incidence, r is the angle of refraction and n is the RI.

The critical angle (Figure 2) is the point or angle at which light refraction ends and total reflection starts. This is defined by Equation (5):

$$\sin c = \frac{n_1}{n_2} \tag{5}$$

where c is the critical angle, n_1 is the RI of first medium and n_2 is the RI of second medium.

However, it was not until 1874, with the invention of the refractometer by Ernst Abbe, that RI became practical to measure.

5 IN-LINE REFRACTOMETER

The primary measurement of a process refractometer is RI. This is done by directing light through a prism at various angles to the interface between the prism and the liquid being measured. This creates both refracted and reflected light rays that form a light and dark image. The critical angle occurs at the transition or borderline that differentiates the refracted and reflected portion of this optical image. This borderline position changes with RI and/or the dissolved solids concentration.

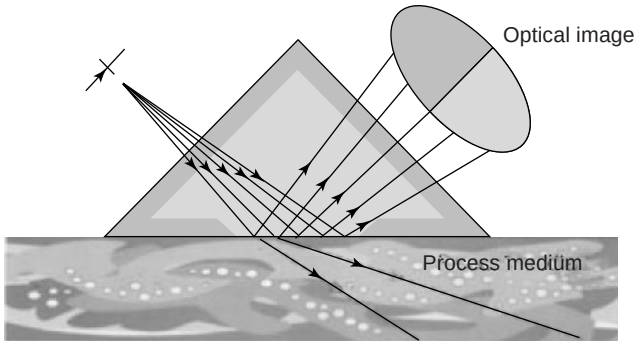

Figure 2 Critical angle.

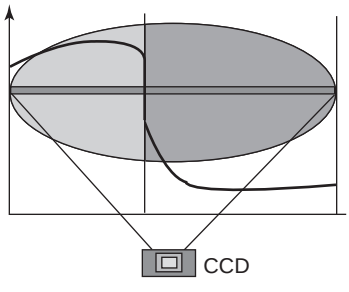

Figure 3 Digital analysis of optical image by CCD.

The in-line process refractometer uses a charge-coupled device (CCD) microchip to determine the position of the borderline and relate its position to RI. The use of the CCD microchip replaces the human eye that serves this function in the case of a hand-held and/or a laboratory refractometer. The CCD then converts this optical image into a digital signal (Figure 3). This conversion eliminates drift and increases the stability of the measurement. The second component to any RI measurement is temperature: RI is temperature dependent and, hence, a correct temperature measurement is important. Only by the correct combination of these two measurements can a stable, accurate and reliable RI value be generated.

The relationship between °API or API gravity of an oil (a density term used in the petroleum industry) and RI is proportional. °API can be defined by Equation (6):

$$^{\circ}\text{API} = \frac{141.5}{\text{specific gravity of the oil}} - 131.5 \tag{6}$$

Specific gravity equals the density of liquid divided by the density of water. Water has a specific gravity of 1.0 and an API gravity of 10°. Materials with high API gravity are lighter and lower in density. The reverse is true for specific gravity.

However, RI can be directly measured whereas °API is calculated. This is why RI has been found to be a more reliable process parameter to monitor. The range of RI at 70 °C for a lube oil varies from 1.4400 to about 1.5000. An accurate in-line determination of RI is required because a change in the RI between 0.0007 to 0.0010 RI units, can result in a change of 1 unit in the final VI of nonprocessed lube oil. The exact magnitude of the change will depend on the oil type and the initial °API of the oil being processed. In general, the change in VI for heavy, high °API, naphthenic-type feeds tends to be greater then that of lighter, lower °API, more paraffinic feeds (Figure 4).

Mounting and location of the prism are important to the measurement. As well as improving the temperature indication, being angled and having an intrusive design promotes self-cleaning. The angle creates turbulent flow and locating away from pipe walls keeps the prism at the same temperature as the process fluid. A minimum velocity is maintained so that self-cleaning can occur and a fresh sample is constantly provided for the measurement. In some lube oil applications the self-cleaning is sometimes supplemented by a steam washing system. The in-line refractometer should be installed in a sample line so that temperature variations are minimized; temperature control is usually done by a heat exchanger. This location assures that any sample collected and later tested by the laboratory will be identical to that measured by the in-line refractometer (Figure 5).

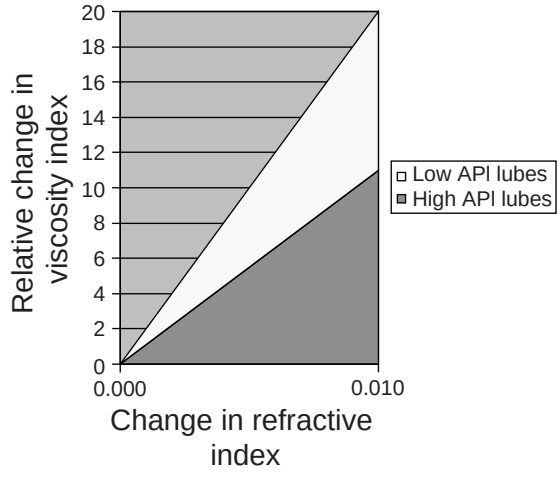

Figure 4 Sensitivity of lube oil VI response.

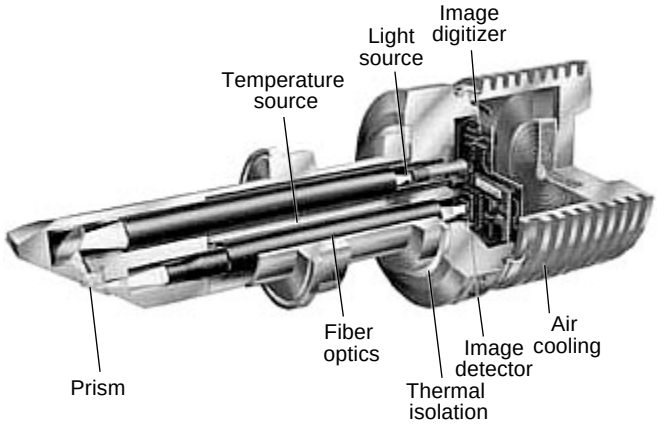

Figure 5 In-line refractometer.

The instrumentation hardware costs for an in-line refractometer are typically below that of other in-line measurement techniques. The final hardware cost is influenced by such factors as cable run, prism washing options and the area classification requirements. This makes refractometers very attractive and competitive as an analytical option.

6 CASE STUDIES

A large lube oil producer has installed several digital process refractometers in their refineries. These units have been in operation for several years. According to one process engineer associated with the project the digital process refractometers operated as expected and provided accurate and repeatable RI numbers. This on-line RI information is then used to predict the final lube oil VI and to adjust the efficiency of the process.

Several types of process change can affect VI, such as, blend stock source and extraction tower temperature. There is a finite residence time for the effect of a given process change to pass through the process unit. However, laboratory samples can be delayed and their results determined well after the unit has been affected. The installation of an in-line device is required to establish an exact correlation between process changes and VI (Figure 6). This delay time can increase production costs and reduce product quality.

A major gulf coast lube oil facility produces a variety of motor oil and lubrication products. Their process is based on Chevron's ISODEWAXING® technology and is designed to operate on a wide variety of crude oils.

Owing to variability of the feed, a reliable and accurate method is required to characterize the process. This refinery makes frequent changes in the crude feed. Changes in the crude feed increase the risk of making off-target product which may contaminate a finished product tank.

Contamination of a finished product tank can adversely affect the refinery in several ways. First, it reduces overall refinery capacity owing to a lack of storage for new production. Second, it tends to reduce net unit feed because the off-test material must be blended in with fresh feed. This is especially costly if the affected unit is currently at its maximum capacity.

This location has two digital process refractometers installed in its new catalytic hydrotreater dewaxing lube oil unit that processes about 10 000 barrels per day or

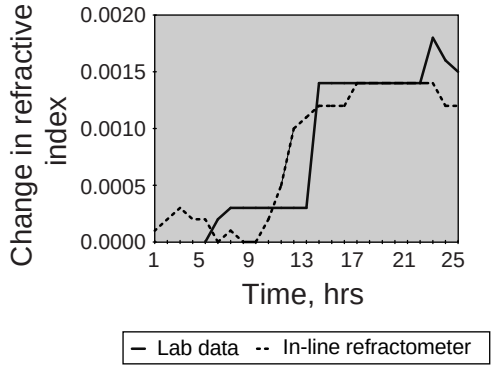

Figure 6 Response of in-line refractometer.

Table 3 Paraffinic base oil prices (April 1998)

Viscosity grade	Price range (cents/gallon)
Light (90–150)	87–110
Medium (200–250)	87–110
Heavy (500–600)	100–111
Bright Stock	123–134

Source: *Hart's Lubricants World.*

420 000 US gallons per day. These refractometers were commissioned early in 1997; the start-up was very simple. According to the process engineer, both units have been stable and very reliable. This plant has found the RI measurement useful during crude blend shifting. The RI reading is used to identify and predict when possibly off-test VI material is going into final product tanks and when best to collect lab samples. If the RI reading is outside the normal level it has been found that the VI will also be outside the required specifications. In this case, the off-test product is diverted from the finished product tank. Operations estimated that because they have avoided producing off-test material, the pay back time for the refractometer was very short, based on the price of lube oil (see Table 3).

7 CONCLUSION

Both viscosity and RI are physical properties of a fluid. They have both been measured by one or more methods for several years. Although laboratory test procedures and protocols are well established in a modern refinery, they can never respond with the speed of an in-line instrument. The profitability of the lube oil unit is related to its ability to obtain quick, accurate and reliable VI information in order to make the correct adjustments to the process.

The following can be stated about the relationship between RI and lube oil VI:

- RI measurement of lube oils has traditionally been done in the laboratory, but now it has moved into the process environment. The process measurement of RI is now reliable, simple and cost effective.
- An in-line process measurement of RI can be used as a real-time predictive tool for the final VI of lube oils in the quickly changing production environment. A quick and accurate response is needed in order to optimize production.
- RI can identify and predict when possibly off-test VI material is going into finished product tanks. Contamination of a finished product tank can adversely affect both the economics and operations of a refinery.

ABBREVIATIONS AND ACRONYMS

°API	American Petroleum Institute Gravity
ASTM	American Society for Testing and Materials
CCD	Charge-coupled Device
MEK	Methyl Ethyl Ketone
RI	Refractive Index
SAE	Society of Automotive Engineers
VI	Viscosity Index

RELATED ARTICLES

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

REFERENCES

1. J.R. Welty, C.E. Wicks, R.E. Wilson, 'Shear Stress in Laminar Flow', in *Fundamentals of Momentum, Heat and Mass Transfer*, 2nd edition, John Wiley & Sons, New York, 98–103, 1976.
2. J. Groetsch, 'Refractive Index Analyzers', in *Analytical Instrumentation, Practical Guides for Measurement and Control*, ed. B. Sherman, Instrument Society of America, 269–282, 1996.
3. J. Tatera, 'Viscosity Analyzers (non ASTM)', in *Analytical Instrumentation, Practical Guides for Measurement and Control*, ed. B. Sherman, Instrument Society of America, 309–323, 1996.
4. D. Greene (ed.), *Hart's Lubricants World*, Phillips Publishing International, Houston, TX, 1998.
5. ASTM Standard D 1747-94, 'Standard Method for Refractive Index of Viscous Materials', in *Annual Book of ASTM Standards*, ASTM, Philadelphia, PA, 587–590, 1994.
6. ASTM Standard D 2270-93, 'Standard Practice for Calculating the Viscosity Index from Kinematic Viscosity at 40 °C and 100 °C', in *Annual Book of ASTM Standards*, ASTM, Philadelphia, PA, 754–759, 1993.

Use of Inspection Properties to Predict Hydrocarbon Fraction Physical Properties

Peter A. Nick
Process Systems Engineering, Yorba Linda, USA

1 Hydrocarbon Characterization and Property Prediction via Inspection Properties **2**
1.1 Introduction 2
1.2 Summary Description of Petroleum Inspection Properties 3
1.3 Chemical Compositions of Petroleum Fractions 6
2 Hydrocarbon Characterization from Distillation Curves **10**
2.1 Introduction 10
2.2 Generation of Distillation Curves 11
2.3 Cut Fraction Characterization – Manual Methods 15
2.4 Establishing the True Boiling Point Curve – Where to Start 16
2.5 Quality of Data 17
2.6 Methods to Use When Lacking "Total Data" 17
2.7 Establishing the Pseudocomponents 18
2.8 Calculation of the Average Cut Parameters 18
2.9 Measures of "True" Boiling Point 19
2.10 Measures of Fractionation Efficiency 23
2.11 Interconversion of Distillation Data 24
3 Prediction of Stream/Stock Properties from Inspection Property Data **31**
3.1 Property Predictions in General 31
3.2 Property Predictions Derived Directly from Observed Data or Other Derived Properties 34
4 Property Prediction for Mixtures of Known Individual Values **47**
4.1 Introduction 47
4.2 Blending Index and Numbers 48
4.3 Other Blend Prediction Methods 50
Glossary of Terms **51**
Abbreviations and Acronyms **54**
References **54**

The chemical complexity of petroleum products has, until recently, precluded calculational techniques based on a definitive analysis of the molecular species involved. Instead, readily observed or easily (quick and inexpensively) measured macroscopic properties (gravity, boiling range, etc.) have been used to correlate information empirically. In some instances, the predictions have been "tweaked" on a local data-centric basis or extended by quasi-theoretical analysis.

Recent computer-aided advancements[134] in multivariate electromagnetic emission, NMR (nuclear magnetic resonance) and/or MS (mass spectroscopy) analyses allow for the reliable and repeatable use of much more detailed experimental data. The ability to speciate cuts by individual chemical components will revolutionize both refining and regulation of refined products in response to the availability of quality data.

So why would one bother with older analysis and prediction methods? It is likely, given the operational sophistication and investment costs needed to deliver this highly speciated information properly, that the old ways will die but slowly. Most refineries and petrochemical plants try to get by with the minimum amount of expensive on-line analyzers. Multivariate data generation is not yet at a state where it can easily be used for real-time process or even product quality control. This barrier will exist for some time to come.

The historical development of crude oil processing technology has generated a great volume of data and "general" correlations, some contradictory. Most of the data upon which these methods are based have been around for 60 or more years. Many have withstood the "test of time", being examined and verified by numerous researchers both public and private. Some of the precision of the older data may not be as good as what is generated today, but new entries to various refinery products' databanks seem to confirm the validity of the best of the older procedures and data.

The classical methods allow one to impute several otherwise unavailable characterizing pieces of data. It is essential that those engaged in the design and analysis of refinery units (which have multiple feed and product streams) become familiar with classical terminology and techniques, and eventually acquire an intuitive sense of what constitutes the "best answer" when confronted with several conflicting alternatives. The subject of this article is the proper methodology of imputing this knowledge from whatever data are at hand.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 HYDROCARBON CHARACTERIZATION AND PROPERTY PREDICTION VIA INSPECTION PROPERTIES

1.1 Introduction

The chemical complexity of petroleum products has, until recently, precluded calculational techniques based on a definitive analysis of the molecular species involved. Instead, readily observed or easily (quick and inexpensively) measured microscopic properties (gravity, boiling range, etc.) have been used to correlate empirical information. In some instances, the data have been ''tweaked'' on a local ''crude-feed'' basis or have been extended by quasi-theoretical analysis. A list of the more common inspection properties for petroleum-derived fuels products is given in Table 1.

Recent computer-aided advancements[134] in multivariate electromagnetic emission, NMR and/or MS spectral analyses allow for the use of such experimental data in real time. The ability to speciate cuts by individual chemical components will revolutionize both refining and regulation of refined products in response to the availability of quality data.

So why would one bother with older analysis and prediction methods? It is likely, given the sophistication and investment costs required to deliver this highly speciated information properly, that the old ways will not die out for a long time; some of them never. Most refineries and petrochemical plants try to get by with the minimum amount of expensive on-line analyzers. Multivariate data generation is not yet at a state where it can easily be used for real-time process or business control. It is likely that this barrier will exist for some time to come.

The classical methods allow one to impute several otherwise unavailable characterizing pieces of data. It is essential that those engaged in the design and analysis of refinery units (which have multiple feed and product streams) become familiar with classical terminology and techniques, and eventually acquire an intuitive sense of what constitutes the ''best answer'' when confronted with several conflicting alternatives. The subject of this article is the proper methodology of imputing this knowledge from whatever data is at hand.

The historical development of crude oil processing technology has generated a great volume of data and ''general'' correlations, many contradictory, and which can be a source of frustration to those accustomed to relatively rigorous true component calculations. Most of the data upon which these methods are based have been around for 60 or more years. They have withstood the ''test of time'', being examined and verified by numerous researchers both public and private. Some of the precision of the older data may not be as good as what is generated today, but new entries to various refinery products' databanks seem to confirm the general efficacy of the older methods and data.

Then, too, a mere listing of components and relative amounts does not always mean that the important processing properties of the stream can be estimated

Table 1 Typical refining inspection properties

- AP
- API gravity
- Bromine(or iodine) number
- CN: Diesel fuel ignition characteristics
- CP
- Distillation data (different types of):
- ASTM D-86 distillation (760 mmHg)
- ASTM D-1160 distillation (usually at 10 mmHg)
- TBP distillation (15/5 distillation at 760 mmHg)
- ASTM D-2887/D-3710(GC Method at 760 mmHg)
- FP
- FRP
- Heat of combustion
- Octane numbers: antiknock characteristics
- PONA analysis (observed/calculated measurement)
- PP
- RC/CC/BMCI
- RI
- RVP
- SP
- Viscosity

The following properties are so quickly and easily derived from basic measured inspection data that they are typically included as a part of the inter-property relationships:

- Watson K factor and other compositional indices
- MW (derived/measured)
- CI
- DI
- Temperature at which $V/L = 20$
- Thermal capacity and transport properties:
  1. Heat of combustion
  2. Heat of vaporization
  3. Thermal heat capacity
  4. Thermal conductivity

The following properties are often measured in a normal sampling in most refineries. They have little bearing on general predictive techniques and will not be mentioned beyond this summary.

- Penetration index for asphalt (observed measurement)
- BS & W content (wt%)(observed measurement)
- Copper number (observed measurement)
- Doctor test (observed measurement)

AP, aniline point; API, American Petroleum Institute; CN, cetane number; CP, cloud point; ASTM, American Society for Testing and Materials; TBP, true boiling point; GC, gas chromatography; FP, flash point; FRP, freeze point; PONA, paraffins, olefins, naphthenes and aromatics; PP, pour point; RC, Ramsbottom carbon number; CC, Conradson carbon number; BMCI, Bureau of Mines correlation index; RI, refractive index; RVP, Reid vapor pressure; SP, smoke point; MW, molecular weight; CI, cetane index; DI, diesel index; BS & W, bottom sediment and water.

accurately. In cases of MWs and critical properties, the multispectral analyses predict well. In the areas of transport properties, such as viscosity and thermal conductivity, predictions from mixtures of pure components, by most published research work, have been patchy at best. Although some efforts have been made public in quantifying the contributory effects of molecular structural interactions, little has been published that is of general value. There will certainly be more opportunities for research in this area, but the current technology will be used to augment and extend pure lab data for the next 10–20 years.

In any event, these correlations have been useful to the author and can be utilized, for lack of any other need, as a check on the values derived from more directly analytical methods.

For editorial purposes, several liberties have been taken in construction of this article. For one, several acronyms are employed in place of explicitly named properties. For example, AP is aniline point, SG is specific gravity, etc. We make so many references to such properties here, both in the text and in equations, that to spell them out time and again would drastically increase the length of the article. We also have relied heavily on two primary compendia of timely information. These are the ASTM standards,[19–49] and the API Technical Data Book.[9–16] Mention of any of these sources in the articles should be considered as an implicit reference. To repeatedly denote them separately would increase text length dramatically. Then too, each of these magnificent documentation sources are constantly being revised and improved. The reader is strongly advised to refer to the *latest* versions of each relevant document, which are readily available from the two organizations, as well as from other many other purveyors of technical information. Finally, several of the correlations and calculation methods are reported here in units other than standard SI, the default for this encyclopedia. In some cases where clarity and brevity allow, we have provided both sets of units. In other cases, we have reported the units as described in the original information source, to avoid transcription errors or in cases where prevailing standards and specifications (or simply potential user bias and familiarity) call for the values in, perhaps, Imperial standard units as opposed to metric. The reader is advised to bear this in mind, especially when dealing with correlations requiring absolute temperatures, such as with the Kelvin or Rankine scales.

1.2 Summary Description of Petroleum Inspection Properties

Among many different types of petroleum properties, the following are the ones that should be considered in order to understand petroleum processing technologies.

The first set of inspection properties are typically measured as a normal daily or even hourly quality control effort. In a few cases, on-line analyzers can be used for real-time process control. The requirement for this capability is the ability to obtain, analyze and report accurately the findings to the process feedback control system on a timely basis (30 s to 20 min, typically).

1.2.1 Aniline Point

The AP is the minimum temperature for complete miscibility of equal volumes of aniline and the sample under test. It is a measure of paraffinicity of petroleum fractions; the higher the AP, the higher is the paraffin content. The standard test for this method is ASTM procedure D-611.

1.2.2 American Petroleum Institute Gravity

The API density standard is by definition (Equation 1):

$$\text{SG} = \frac{141.5}{131.5 + \text{API gravity}} \tag{1}$$

Therefore, by mathematical manipulation (Equation 2):

$$\text{API gravity} = \frac{141.5}{\text{SG}} - 131.5 \tag{2}$$

The standard test for this SG of a petroleum cut is ASTM procedure D-4502.

1.2.3 Bromine (or Iodine) Number

The bromine number is the number of grams of bromine consumed under specified test conditions by 100 g of sample. Bromine number is a measure of olefinicity of a petroleum fraction; the higher the bromine number, the higher is the olefinic content. Iodine is sometimes used for this test instead.

1.2.4 Cetane Index/Cetane Number: Diesel Ignition Characteristics

The cetane number (CN) is the diesel fuel analog to gasoline octane. It is defined (ASTM D-613) as the percentage of normal cetane that must be added to heptamethylnonane in order to match the ignition characteristics of the fuel undergoing tests. Typical refinery diesel stock CNs vary from 35.0 to 50.0.

The CI is a useful calculation for estimating the CN without experiment.

1.2.5 Cloud Point

This is the temperature at which paraffinic waxes (or other high MW components) just start to crystallize

and cause the stock to appear cloudy or hazy. This is a constraint temperature for fuels that must perform in low temperatures (−20 °F and below). Jet fuels and arctic winter gasolines are two such products. The CP is also an important characteristic of lubricants that need to perform in cold temperatures. The formation of wax crystals that can form will inhibit the lubrication effectiveness.

The standard test for this method is ASTM procedure D-2500.

1.2.6 Distillation Data (Different Types of)

1.2.6.1 American Society for Testing and Materials D-86 Distillation (760 mmHg)

- Gasoline
- Distillate

1.2.6.2 American Society for Testing and Materials D-1160 Distillation (Usually at 10 mmHg)

- Heavy oils
- Residuum
- Fractionator bottoms

1.2.6.3 True Boiling Point Distillation (15/5 Distillation at 760 mmHg)

- 15 stages and 5 reflux ratio
- Approximate true boiling temperatures
- Used for blending of fractions
- Pseudo-component generation from the distillation
- This test has several variations in common use but ASTM D-2892 is as close to a standard as exists.

1.2.6.4 American Society for Testing and Materials D-2887/D-3710 (Gas Chromatography Methods at 760 mmHg)

- Programmed GC calibrated to type (boiling range) of stocks (gasoline, kerosine, diesel, etc.). They can be readily correlated to D-86 and TBP and are quick and repeatable up to heavy ends [1000 °F normal boiling point (NBP) maximum] distillation range.

1.2.7 Flash Point

This the lowest temperature to which the substance must be heated in air to form a mixture containing sufficient vapor to be ignited momentarily by a flame under a specific condition. The test methods used are the Cleveland open cup (ASTM D-92), Pensky Martins closed cup (ASTM D-93) and TAG closed cup standard (ASTM D-56). The FP is important in the handling of volatile products such as cleaning solvents and gasolines. For heavier stocks, the FP is used to estimate the oil's ability to vaporize under certain conditions, which affects both the oil's potential flammability hazard and the consumption rate of the oil.

1.2.8 Freeze Point

This parameter is similar to CP but according to a slightly different procedure. It is defined by ASTM D-2386 as the temperature at which the first solid hydrocarbon crystals, which are formed as the sample of the hydrocarbon is slowly cooled with agitation, disappear upon slight heating. Again, cold weather performance of the fuel is the critical property of concern here.

1.2.9 Octane Number: Anti-knock Characteristics

This is defined as the volume percentage of isooctane that must be mixed with *n*-heptane in order to match the knock intensity of the diesel fuel undergoing testing.

Octane number of isooctane = 100.
Octane number of *n*-heptane = 0.0.
Research octane number: engine speed 900 rpm.
Motor octane number: engine speed 600 rpm.

1.2.10 Paraffins, Olefins, Naphthenes and Aromatics Analysis (Observed/Calculated Measurement)

The PONA in a petroleum fraction is often obtained by a combination of GC, NMR and infrared (IR) methods and inspection properties or by multivariate analysis of one or more of these. There are several ASTM standards for a multitude of tests that can be used to gather this information.

1.2.11 Pour Point (American Society for Testing and Materials D-97)

The PP is the lowest temperature at which a high MW hydrocarbon mass starts to pour. A solidified quantity is slowly heated (in 3 °C steps) and tilted to see if the solid mass starts to flow within 5 s. The temperature at which perceptible flow occurs is recorded as the PP. This property is useful in determining some of the processing properties of heavier cuts. Cold weather flow performance of the fuel or lubricant stock is the major concern addressed by this standard.

1.2.12 Ramsbottom Carbon/Conradson Carbon/Bureau of Mines Correlation Index

These tests determine the percentage (wt%) of residue left after pyrolytic carbonization of a petroleum fraction. This is usually done for residuum or fractionator bottoms product. They are a measure of how much coke it will

make in coking processes. All lubricating oils leave behind ''carbonaceous'' deposits at high temperature as well. Deposits of these materials can ultimately affect the engine's performance.

The ASTM D-189 method is used for Conradson carbon and the ASTM D-524 method for Ramsbottom carbon contents. The BMCI is another method of defining the carbon residue of shale oils.[57]

1.2.13 Refractive Index

The RI represents the ability of a liquid to bend and distort the travel path of light across the air–liquid boundary. It is a fairly simple and repeatable test. RI is a basic parameter for determination of aromatics content and, when combined with viscosity, provides a quality estimate of MW. ASTM D-1218 is the controlling standard.

1.2.14 Reid Vapor Pressure

RVP is an approximation of the vapor pressure of naphtha and distillate range petroleum fractions at 100 °F, the estimated automobile carburetor or injector inlet temperature. RVP also gives the refiner an indication of how much light hydrocarbons (especially *n*-butane) are contained in a petroleum fraction. Three ASTM standards, D-323, D-4956 and D-5191, are used to ensure the quality of the data.

1.2.15 Smoke Point (American Society for Testing and Materials D-1322)

The SP is a measure of burning cleanliness of jet fuel or kerosine. It is defined by the ASTM standard as the maximum flame height (millimeters) at which a fuel will burn without smoking. The SP is low for a fuel having a high smoking tendency. In general, removal of aromatic content increases the SP. An SP > 17 is generally acceptable.

1.2.16 Viscosity

Viscosity is an important control property for lube oils and often for some heavy cracked products. There are at least eight scales by which these data can be reported. For most correlations, either centipoise or centistokes is required. This may necessitate some calculations in addition to the primary measurement data one takes in the laboratory. This article will not concern itself with such conversions as they are widely presented elsewhere in the literature.

The following properties are so quickly and easily derived from basic measured inspection data that they are typically included as a part of the inter-property relationships.

1.2.17 Watson K Factor

The most important derived property (sometimes called the UOP *K* factor) is the Watson *K* (K_W) or characterization factor. It is defined by Equation (3):

$$K_W = \frac{[\text{weight-average boiling point } (^\circ\text{R})]^{1/3}}{\text{SG}} \quad (3)$$

It is a measure of the distribution of different types of molecules (paraffins, naphthenes and aromatics) in a petroleum fraction. The higher the K_W factor, the higher is the paraffinicity, and the lower the K_W factor, the higher is the aromatics content. For most refinery stocks K_W varies from 10.0 to 13.0 in general. Note the use of absolute temperature scales in calculating K_W.

Three other compositional characterization factors are used to a limited extent with petroleum fractions. These are the viscosity–gravity constant, the Huang factor and the refracticity index. They will be discussed in limited detail. Most of the methods and data work done to date in this field have involved the K_W factor far more than any other.

1.2.18 Molecular Weight (Derived/Measured)

The molecular weight (MW) can actually be estimated by laboratory procedures,[78,111] but these are slow and labor intensive, so they are very rarely performed in an operations setting.

1.2.19 Cetane Index: Diesel Fuel Ignition Characteristics

The CI is a derived property that is roughly approximated to measured CNs. Direct analysis of CN is a difficult measurement and several correlations have been developed to permit operations to estimate the CI based on other observed properties.

1.2.20 Diesel Index

This is given by Equation (4):

$$\text{DI} = \frac{(\text{AP})\ (\text{API gravity})}{100} \quad (4)$$

This is also an indication of paraffinicity of a petroleum fraction.

1.2.21 Temperature at which V/L = *20*

T at $V/L = 20$ is a fuel specification requirement for most gasolines. It is obtained from the TBP curve with estimates of vapor and liquid volumes where the ratio of the two is 20 : 1. As such, it is easier to derive than to measure.

1.2.22 Thermal Capacity and Transport Properties

This category of properties is most important in the design and prediction of refinery operations. Many of these can be measured directly in the laboratory but few with any simplicity. Derivation of values is then the preferred and a usually accurate alternative in this case.

The properties considered here are outlined below.

1.2.22.1 Heat of Combustion An important property of all fuels is the heat of combustion, since this is a measure of the energy available from the fuel. A knowledge of this value is essential when considering the thermal efficiency of equipment for producing either power or heat. ASTM D-240 is used for the direct determination of heats of combustion of aviation gasolines and of fuel oils. Net heats of combustion of petroleum fuels vary from $45\,925\,kJ\,kg^{-1}$ ($19\,774\,Btu\,lb^{-1}$) for propane to $33\,029\,kJ\,kg^{-1}$ ($14\,200\,Btu\,lb^{-1}$) for petroleum coke. Ordinary calorimetric techniques are employed.

1.2.22.2 Heat of Vaporization Proper and safe separation of hydrocarbons by vapor–liquid equilibria requires an accurate knowledge of the heats of vaporization. This is an important variable for calculating heating duties and thermal processing equipment sizes.

1.2.22.3 Heat Capacity of Liquid–Vapor Prediction of liquid and vapor behavior under heat transfer conditions is of great concern. This is an important variable for calculating details of thermal processing and distillation equipment.

1.2.22.4 Thermal Conductivity of Liquid–Vapor Prediction of this property is often an inaccurate procedure. With hydrocarbons, however, it is surprisingly good. This is also an important variable for calculating details of design and operation of thermal processing equipment.

1.2.22.5 Critical Properties For a pure component, the critical state can be summarized by three interrelated thermodynamic parameters – the critical temperature, critical pressure and critical density. The critical temperature is defined as that temperature above which a gas cannot be liquefied by an increase in pressure. The maximum pressure that will result in formation of a separate liquid phase at the critical temperature is called the critical pressure. The density of a fluid at its critical temperature and pressure is its critical density.

When dealing with mixtures of two or more compounds, this precise definition needs some refining. In theory, there is a state in which all intensive properties of the coexisting liquid and vapor phases become continuously identical. Corresponding states correlations that depend upon a reasonable estimate of the critical state work best with what are known as ''pseudocritical'' quantities. These are traditionally defined as the molar average of the critical temperature and pressure.

The following properties are typical items also measured in a normal sampling in most refineries. With the exception of the asphalt test, they all measure contamination levels. They have little bearing on general predictive techniques and will not be mentioned beyond this summary.

1.2.23 Penetration Index for Asphalt (Observed Measurement)

This is the depth (in millimeters) that a standard needle penetrates vertically into a sample of the asphalt material under specific test conditions.

1.2.24 Bottom Sediment and Water Content (Observed Measurement)

This is a measure of basic sediment and water absorbed and contaminating full-range crude and other hydrocarbon fractions.

1.2.25 Copper Number (Observed Measurement)

This is the amount (milligrams) of mercaptan sulfur in 100 mL of oil sample. It is used to measure mercaptan content in a general petroleum fraction.

1.2.26 Doctor Test (Observed Measurement)

Sodium plumbite solution is used to measure mercaptan content in a gasoline stock.

1.3 Chemical Compositions of Petroleum Fractions

To grasp fully the relationship between performance and fuel properties, is essential to understand the chemical composition of the fuel. This is currently done at great difficulty and expense, but enough data exist that several generalizations can be formulated.

1.3.1 Broad Component Classification of Hydrocarbons[41,56,59,68]

Petroleum and coal-based liquid hydrocarbons contain chemical species from a list of over 1600 pure hydrocarbons and several related compounds. Nearly every crude oil has 300–800 measurable species at some detectable level. Among the categories are the following:

1. Paraffins (alkanes)
   - Straight chain (normal)
   - Branched chain (iso)

Table 2 Typical crude assay ranges[3,7,72,73][a]

Source	Boiling point range of crude (°F)	Volume-% distilled at end boiling point	SG range of crude	MW range of crude
Kuwait crude oil	104–2102	100.0	0.6285–1.10	77–1668
Venezuelan crude oil	195–533	36.0	0.746–0.9593	195–353
Saudi Arabian crude oil	78–963	80.0	0.645–0.9254	68–484
Abu Dhabi crude oil	104–932	87.04	0.6285–0.9465	77–450
Sumatran crude oil	99–934	73.7	0.685–0.892	70–477
Iranian crude oil	93–1022	78.4	0.654–0.9285	71–532
Iraq crude oil	104–2012	100.0	0.628–1.10	77–1532
Qatar crude oil	104–1652	100.0	0.6285–1.054	77–1097

[a] The TBP ranges of the crude are not always complete. Very often, the assay is reported only to the last tangible cut range above the tower resid. This may be diesel, or more likely gas-oil material. Measuring the resid properties often seemed superfluous and unnecessary. Obtaining accurate temperatures and recovery percentages is also very difficult for temperatures much above 800 °F and certainly for those above 1000 °F. Many petroleums have estimated end-points between 1300 and 2000 °F.

Table 3 Typical hydrocarbon assay molecular analysis[94,99,101,102]

Petroleum fraction	MW	Critical temperature (°F)	Critical pressure (psia)	Molecular type analysis (vol%)				
				Paraffin	Naphthene	Olefin	Aromatic	S compounds
1. California naphtha	102.4	536.7	532	36.2	58.2	0.0	5.6	0.0
2. Alaska naphtha	131	626.5	414.7	39.7	39.6	0.0	20.5	0.0
3. Aromatic naphtha	126	626	709	0.0	0.0	16.9	83.1	0.0
4. Jet fuel	144	665	440	18.6	69.8	0.0	10.6	1.0
5. Light naphtha	120	605	463	61.9	30.6	0.0	7.6	0.0
6. Heavy naphtha	142.4	646	371	59.3	30.8	0.0	9.9	0.0
7. Kerosine	162.3	739	361	30.9	64.3	0.0	4.8	0.0
8. Fuel oil	227.5	879	290	29.8	45.6	0.0	22.4	1.3
9. Gas oil	214	868	280	38.8	41.5	0.0	17.2	2.4

2. Cycloparaffins (naphthenes)
   - Single ring
   - Multi-ring
3. Olefins
   - Aliphatic olefins
   - Cycloolefins
4. Aromatics (with/without attached chains)
   - Single ring
   - Multi-ring
5. Sulfur compounds
   - Mercaptans (thiols)
   - Thiophenes (simple, multi-ring)
6. Nitrogen Compounds
   - Amines
   - Pyridine

In general, petroleum fractions are made of combinations of the above components in MW structures ranging from 2 to >600. Tables 2–4 present type and characterization breakdowns of various crude oils and their typical refinery cut fractions.

1.3.2 Crude Oil Analysis by Fractional Product Content[1–18,51,57,72]

Crude oil distillation separates the basic components of petroleum into various streams, mostly on the basis of boiling point. Gasolines, kerosines, diesel, gas-oils, resids, etc. are all standard refinery terminology for some of these cuts. Refer to the glossary if a more detailed explanation is necessary.

A typical crude oil can be described compositionally in several ways:

- A typical crude oil composition speciated by pure boiling range cut fractions is presented in Figure 1.
- Relative amounts of hydrocarbon types in a crude oil are displayed in Figure 2.
- A textual listing of a typical crude fraction assay is presented in Table 4.
- A graphical display of the boiling point curve typical of crude assays is shown in Figure 3.
- The delineation of various products based on TBP value is demonstrated in Figure 4.

Table 4 Typical crude oil assay[3,99,101,102,150,151]

Name, state, and country	Boscan, Zulia, Venezuela	Kern River, CA, USA	Quitman, MS, USA	Tia Juana, Zulia, Venezuela	Gato Ridire, CA, USA	Bachaquero, Zulia, Venezuela
Gravity (API)	9.5	10.7	12.7	13.3	13.8	14.0
Sulfur (wt%)	5.25	1.23	3.85	2.66	5.21	2.64
Viscosity (SSU at 100 °F)	–	6000+	–	452 at 210 °F	6000+	–
Characterization factor:						
at 250 °F	12.20	–	11.9	–	11.62	11.75
at 450 °F	11.60	11.13	12.1	11.40	11.28	11.40
at 550 °F	11.40	11.15	11.85	11.40	11.20	11.15
at 750 °F	11.40	11.15	11.55	11.30	11.20	11.30
Gasoline:						
Octane No. clear	–	–	–	–	–	72.4
% to 400 °F	3.8	1.2	9.4	1.4	11.3	5.0
Octane No. clear	42.0	–	42.5	62.0	67.0	68.0
% to 450 °F	5.2	2.2	11.4	2.9	14.2	7.5
Kerosine distillate:						
% 375–500 °F	4.1	2.7	5.9	4.2	7.5	7.0
API gravity	39.2	32.5	45.0	36.2	33.9	35.0
SP (mm)	19.2	13.0	29.0	15.7	14.5	16.0
Sulfur (wt%)	3.0	0.38	0.75	0.56	2.25	0.50
Distillate or diesel fuel:						
% to 400–700 °F	16.0	19.7	21.0	21.3	20.0	22.5
DI	26.0	30.0	48.0	39.0	32.0	33.0
PP (°F)	−35.0	−30.0	3.0	−40.0	−30.0	−35.0
Sulfur (wt%)	3.3	0.8	1.7	1.3	3.1	1.1
Lube distillate:						
% to 700–900 °F	12.8	22.2	17.6	23.2	14.8	19.0
PP (°F)	55.0	–	–	−5.0	–	−10.0
Viscosity index	25.0	–	37.0	−20.0	−19.0	−17.0
Sulfur (wt%)	4.3	1.5	2.65	2.30	4.5	2.25
Residue (% over 900 °F)	67.4	57.0	52.0	53.5	52.5	53.5

Typically the TBP curve and a few of the important physical properties (viscosity, density, cloud and PPs, etc.) of each cut fraction are reported in a standard abbreviated laboratory analysis of a crude. More recently, a generic speciation by carbon number and type (PONA) is routinely reported if the proper spectrophotometric equipment and analyst knowledge are available.

1.3.3 Chemical Analysis of Refinery Products[47,59,62]

Refinery products are divided into two categories, namely are ''virgin'' or ''straight-run'' streams that are derived from pure physical separation from the crude feed, and converted stocks which have been run through one or more thermal or catalytic reactions. Of the conversion reactions, cracking of large molecules into smaller ones is the most common. This category can be subclassified into thermally cracked stocks, catalytically cracked stocks and hydrocracked stocks. Other reactions common in refineries are reforming, isomerization and alkylation.

The major refinery streams of ultimate value in a refinery are virgin naphtha, thermal crackate, catalytic reformate and hydroprocessed stocks.

Compositional comparison charts are illustrated as follows:

- Figure 2 shows typical components in virgin crude cuts.
- Figure 5 shows typical component compositions for a catalytically reformed stock.
- Figure 6 shows typical component compositions for a catalytically cracked stock.
- Figure 7 shows typical component compositions for a hydrocracked stock.

Catalytic reformate contains the largest amount of aromatics, which increase the octane rating. Thermal crackate made from thermal cracking or coking processes contains significant amounts of olefins which are not stable and thus not suitable as a gasoline component. Virgin naphtha (straight-run naphtha) contains small amounts of aromatics but significant amounts of paraffins.

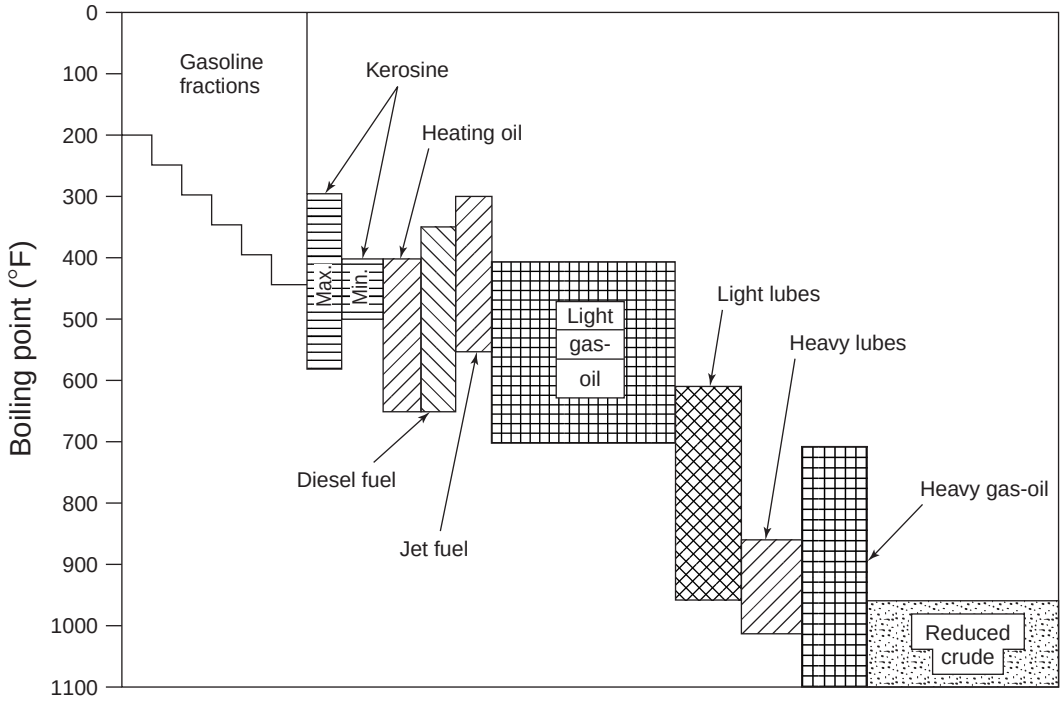

Figure 1 Sample crude oil composition.

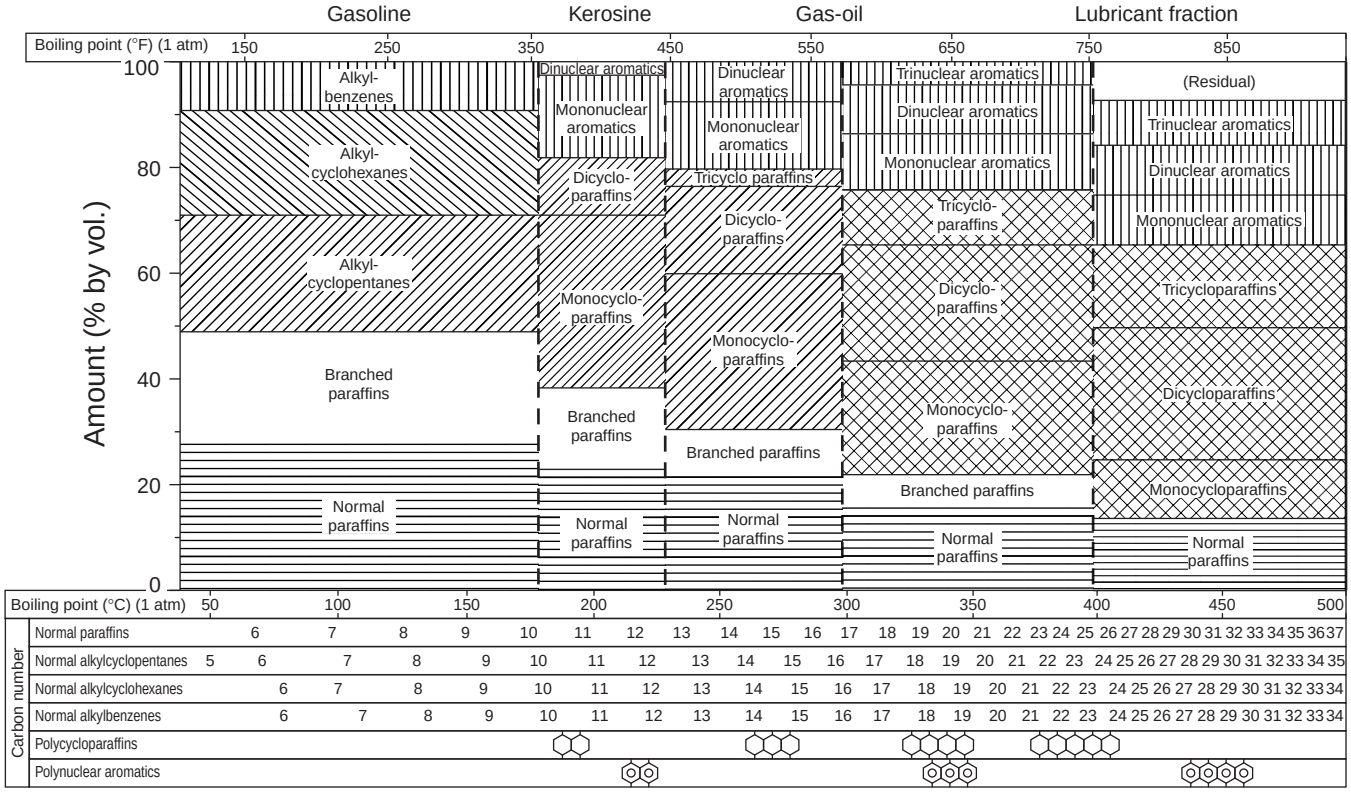

Figure 2 Relative amounts of hydrocarbons in a crude oil.

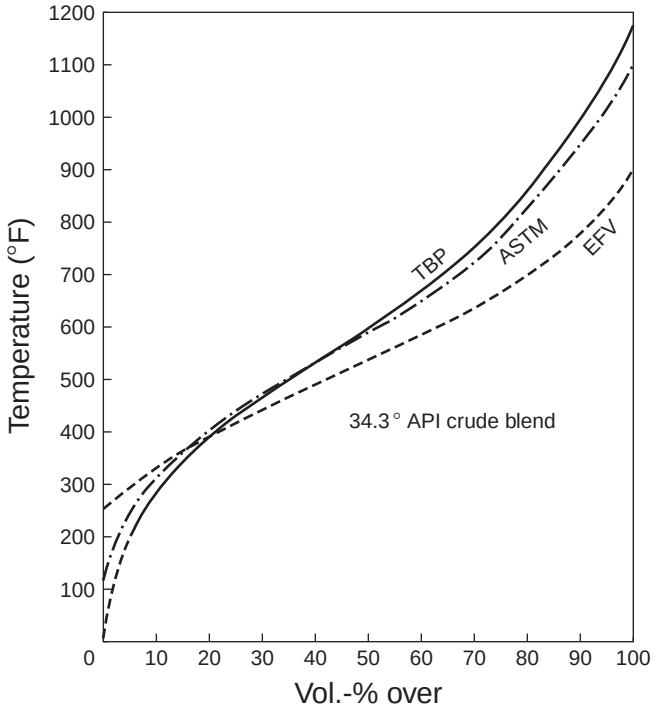

Figure 3 Typical distillation curves. EFV, equilibrium flash vaporization.

1.3.4 Chemical Analysis of Synfuel Products[57,72,73,142]

Most work on hydrocarbons has gone into products derived from crude oil refineries. There is, however, a significant base of stocks derived from shale oil, coal and heavy oil/tar sands recovery processes.

These stocks contain most of the same chemical components as do petroleum-based derivatives but will also contain substantial amounts of several biaromatic multi-ringed components that are not generally major components of the petroleum-based stocks. They tend to have higher nitrogen levels than most crudes and perhaps a slightly lower sulfur content. A partial listing of some of these chemicals is presented in Table 5.

1.3.5 Component Effects on Refinery Products

1.3.5.1 Motor Gasolines Volatility characteristics for motor performance are dependent upon the amount of lighter components such as butane and pentane. Antiknock characteristics are highly dependent upon the components of the stock involved, especially in the C_6-C_9 range.

1. *Paraffins*: octane number increases with increasing MW and branchedness.
2. *Olefins*: octane number is better than that of corresponding (same carbon number) paraffins.
3. *Naphthenes*: have moderate octane numbers.
4. *Aromatics*: have excellent octane numbers but poor volatility and have some toxic characteristics that have resulted in some regulatory constraints on their usage levels in transportation fuels.

1.3.5.2 Other Petroleum Stocks A lowering of aromatic content generally increases (improves) the SP.

The paraffin content correlates with FRP, CP and PPs of several fuel stocks. Lowering the amounts of paraffins and especially isoparaffins generally improves (lowers) these three property values.

The higher the Watson *K* factor, the higher is the paraffinicity; the lower the Watson *K* factor, the higher is the aromatics content.

Highly paraffinic crude oils may deposit waxes in transport lines and processing equipment. On the other hand, the lube oils made from these stocks are generally of superior quality.

Highly aromatic or naphthenic stocks will tend to form more coke in cracking and pyrolysis operations. They may also tend to foul catalyst pellet matrixes in various conversion processing operations.

Table 6 shows a series of gasoline blend components and a select set of their corresponding properties.

2 HYDROCARBON CHARACTERIZATION FROM DISTILLATION CURVES

2.1 Introduction

Hydrocarbon fractions, being a mix of several dozen to hundreds of hydrocarbon compounds, cannot be treated simply as a single pure component. A measure of the various molecular sizes is necessary in order to obtain meaningful thermodynamic information about the cut. The typical measure of molecular size distribution is via a volumetric distillation designed to measure the ''true'' molar boiling point of each incremental volume of overhead batch distillate. These boiling points, combined with corresponding cut point gravities, are used to estimate MWs. Nearly all industrial hydrocarbon characterization procedures require at least two of these quantities.

One more reason why this makes sense is that hydrocarbon cuts are usually separated in the refinery on the basis of distillation column performance. The boiling point range of a cut is highly dependent upon the size of the molecules, as are the subsequent properties of the fraction stream. Since the boiling point of each familiar homolog of components rises in a predictable fashion with size (i.e. MW), the boiling temperature of a cut indicates something about the relative size of the molecules in the

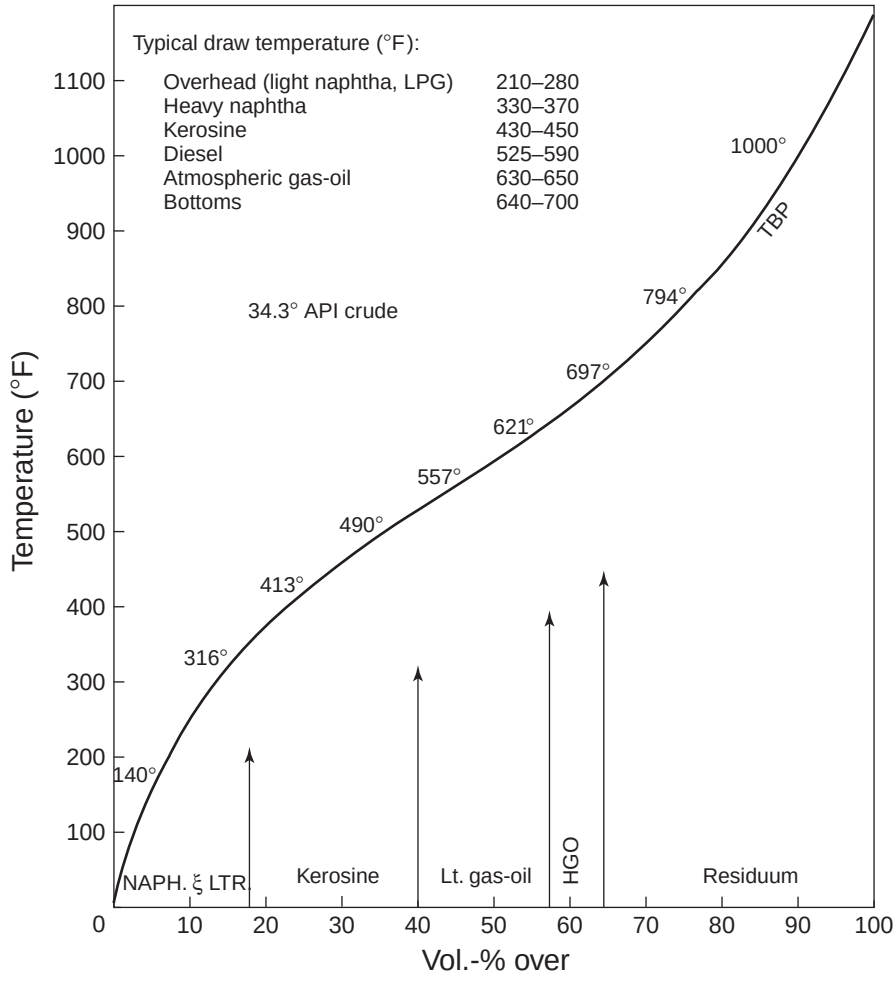

Figure 4 Product yields on TBP curve.

cut and their subsequent expected behavior. Typically, the hundreds of component species in a cut are ''lumped together'' on the basis of comparable boiling points for simplicity. This is why establishing the average ''true'' boiling point of a cut/fraction is so important. One can also then understand that the smaller the cut range, the more likely one is to be able to use these boiling-point lumping models (also called pseudocomponents or hypotheticals) with any reasonable accuracy.

There are numerous ways of using distillation curves to characterize hydrocarbon fractions. No particular purpose would be served by attempting to present all these methods here. Instead, only those considered most reliable and useful will be included.

2.2 Generation of Distillation Curves

There are four primary types of laboratory distillation curves that are used to characterize crude oils and refinery products, and these curves constitute the most important and useful information upon which the process design engineer must rely. Of these, the most useful curve is the true boiling point (TBP) curve. TBP values are rarely a final product specification but these values are most useful in extrapolating properties of various cuts within the measured TBP cut range.

1. The TBP distillation is basically a batch distillation conducted in equipment of high fractionation efficiency, such that each increment of overhead vapor is (theoretically) a pure component at its TBP. The curve is a plot of overhead temperature versus cumulative volume percentage overhead yield. A sample TBP–volume % overhead curve for a full range crude is shown in Figure 3.

   One consistency problem here is the lack of a real standard procedure for conducting a TBP test. ASTM D-2892 is a standard that is often used, but most laboratories modify it for various reasons. Since the various laboratory stills in use employ

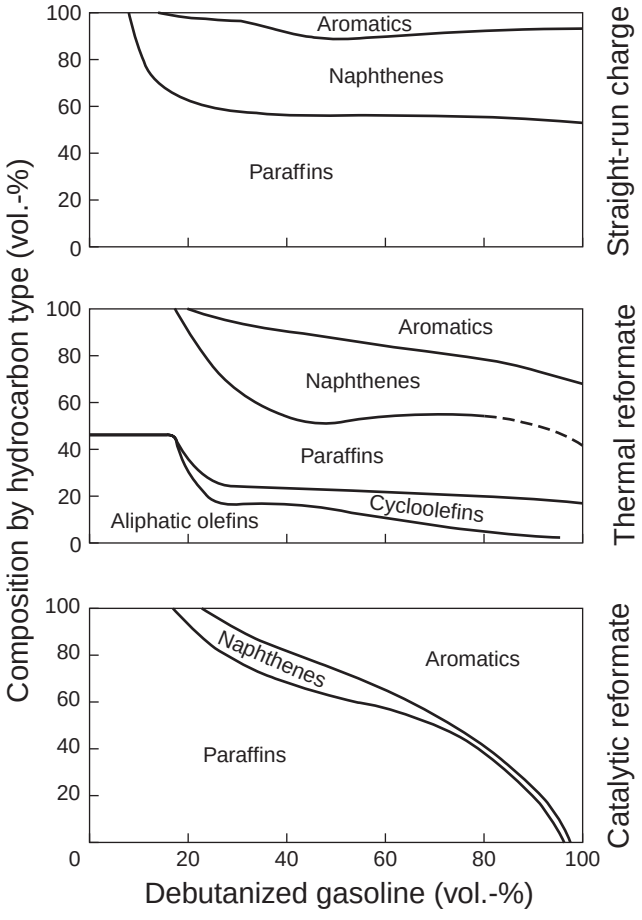

Figure 5 Composition of catalytic reformate vs. thermal reformate and straight-run charge (Pennsylvania crude).

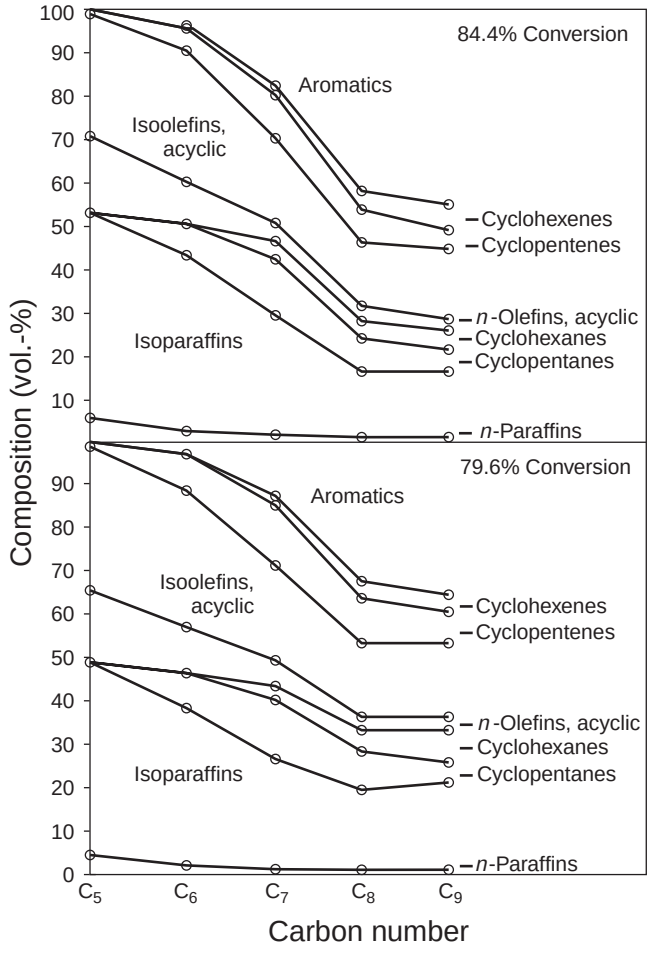

Figure 6 Catalytic cracker naphtha compositions.

different numbers of trays and different reflux ratios, a particular oil sample will yield slightly different TBP curves depending on where the distillation is run.

The terms ''Oldershaw'', ''Podbielniak'', ''Hypercal'' and ''15/5'' (15 trays, 5:1 reflux ratio) refer to distillations that are essentially TBP runs.

The United States Bureau of Mines (USBM) historically used the ''Hempel'' distillation for its crude assay work.[80] The Hempel distillation, and the very similar ASTM D-285, are what may be termed semifractionating distillations. Both of these procedures give some degree of fractionation, but are less efficient than a TBP distillation and should not be regarded as truly equivalent to TBP runs.

2. The newest and most promising methods are based primarily on GC analysis of the crude stream or its fractional cuts. Recent improvements to GC instrumentation allow for direct estimates of TBP based on elution times of component vapor through a chromatographic column. Although limited to 1000 °F boiling points and below, they are quick and reliable. ASTM D-2887 and D-3710 are rapidly becoming laboratory standards.

   Whenever possible, the source of any directly measured TBP distillations should be at least one of the 15/5 or better procedures or the equivalent GC predictive method.

   In some cases, the TBP curve is a derived function from one of two ASTM distillation procedures.

3. The first of these ASTM distillations is covered by ASTM specification D-86 for tests conducted at atmospheric pressure (ATM) for lighter refinery fuel-blend stocks (naphtha, kerosine, light diesel). D-86 values are often a final product specification.

4. The second major ASTM distillation is covered by ASTM specification D-1160 for tests conducted at high vacuum for heavier refinery high-boiling point stocks (heavy diesel, gas-oils, resids, etc.). D-1160 values are often a final product specification.

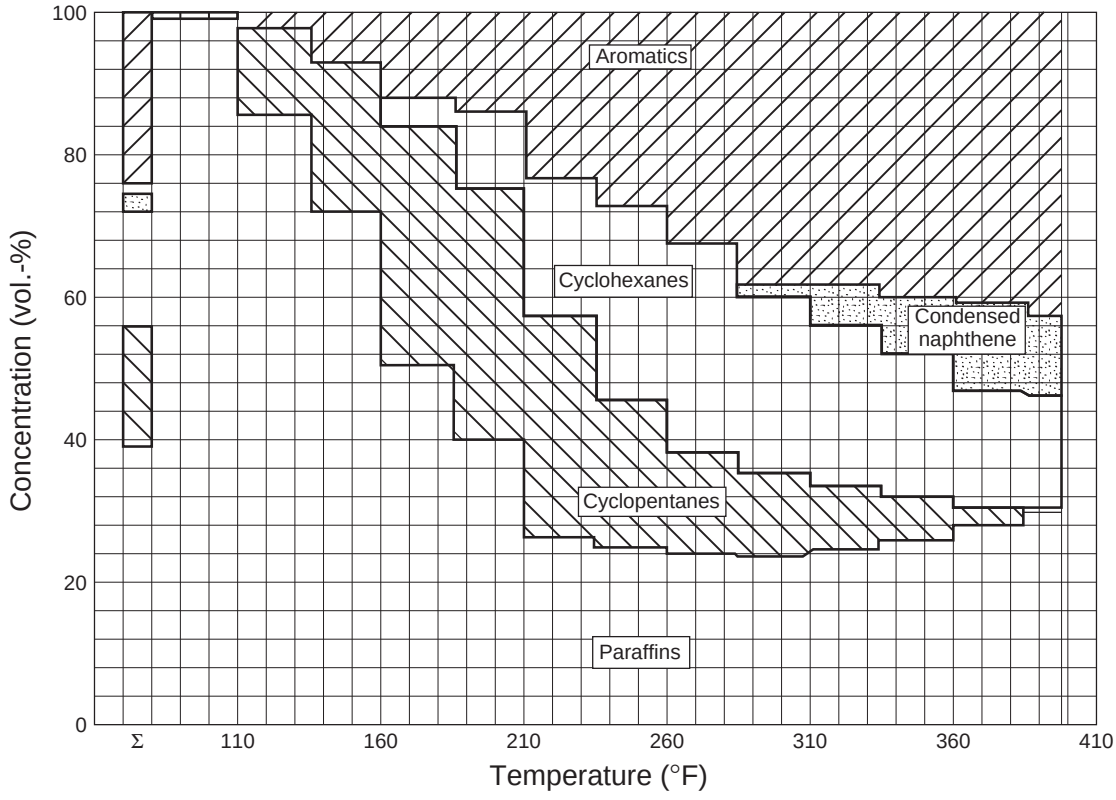

Figure 7 Composition distribution of gasoline from hydrocracking of typical gas-oil.

Table 5 Component data for coal-based hydrocarbons[108]

Component	MW	API	NBP (°F)	Critical temperature (°F)	Critical pressure (psia)
Phenol	94.1	0.13	359.2	791.5	889.1
Tetralin	132.2	13.84	406.1	834.5	510.0
Naphthalene	128.2	−8.46	424.4	887.4	576.1
Biphenyl	154.2	4.04	191.4	960.3	557.0
Fluorene	166.2	−12.89	566.7	1020.5	433.5
Dibenzofuran	168.7	−7.16	548.3	997.6	464.4
Phenanthrene	178.2	−11.79	641.1	1122.2	420.3
Fluoranthene	202.3	−20.61	743.1	1189.7	373.3
Anthracene	178.2	−18.93	647.1	1129.7	411.5
Acenaphthalene	154.2	25.72	517.7	974.7	467.3
Pyrene	202.3	−20.61	743.1	1146.1	377.7
Acenaphthene	154.2	−14.94	531.0	986.6	448.2
Acridine	179.2	9.02	654.6	1142.2	430.6
Carbazole	167.2	−4.71	610.5	1158.3	473.2
Chrysene	228.3	−20.69	825.8	1328.4	346.8
Indole	117.2	−16.65	488.9	963.6	620.2
2-Phenylpyridine	155.2	−3.79	515.9	983.2	486.4
2-Dihydronaphthalene	133.2	10.00	404.3	833.8	508.5
Quinoline	129.2	−2.51	459.7	930.5	551.1
1,2,3,4-Tetrahydroquinoline	133.2	1.99	481.7	928.7	593.7
Naphthacene	228.3	−32.96	829.4	1317.2	348.0
Benz[*a*]anthracene	228.3	−11.18	819.7	1337.0	348.0
Triphenylene	228.3	−11.48	839.1	1364.0	348.0

Table 6 Typical gasoline fuel blending stocks boiling point and molecular analysis

	Blending component										
	Cat. cracked naphtha 1	Cat. cracked naphtha 2	Light cat. cracked naphtha	Light alkylate	Heavy alkylate	Full-range reformate	Light st. run naphtha	C_6 isomerate	Light reformate	Mid-cut reformate	Heavy reformate
Gravity (API)	52.1	51.9	66.8	72.3	55.8	44.2	81.8	83.0	72.0	32.8	29.8
Aromatics (vol.-%)	35.2	35.9	17.6	0.5	1.0	81.1	2.2	1.6	4.6	94.2	93.8
Olefins (vol.-%)	32.6	25.4	44.9	0.2	0.9	1.0	0.9	0.1	1.5	0.6	1.9
Saturates (vol.-%)	32.2	36.6	37.4	99.3	95.1	37.9	96.9	98.3	93.7	5.1	4.2
Benzene (vol.-%)	1.08	1.23	1.24	0.00	0.01	1.17	0.73	0.00	4.01	0.00	0.00
Bromine no.	57.1	41.7	91.4	2.3	0.3	1.2	0.5	3.8	3.1	0.6	0.5
RVP (psi)	4.3	4.6	6.7	4.6	0.3	3.2	10.8	8.0	3.8	1.0	0.3
Temperature D-86 (°F)											
IBP	110	112	95	101	299	117	91	118	138	224	313
*T*05	143	142	117	144	318	158	105	131	169	231	325
*T*10	156	155	124	162	325	192	113	134	174	231	326
*T*20	174	171	130	181	332	224	117	135	179	231	331
*T*30	192	189	130	196	340	244	121	135	182	232	335
*T*40	215	212	149	205	346	258	126	136	185	233	339
*T*50	241	239	164	211	354	270	132	136	188	234	344
*T*60	270	269	161	215	362	280	139	137	190	235	350
*T*70	301	302	200	219	373	291	149	137	192	237	356
*T*80	336	337	224	226	391	304	163	136	194	240	370
*T*90	376	379	257	239	427	322	184	139	196	251	391
EP	431	434	337	315	517	393	258	146	218	316	485
RON	93.2	92.8	93.6	93.2	65.9	97.3	63.7	76.6	57.6	109.3	104.3
MON	61	82.1	79.4	91.2	74.5	86.7	61.2	80.5	56.5	100.4	92.4
(R + M)/2	87.1	87.4	86.5	92.2	70.2	92.0	82.4	79.5	58.0	104.9	95.4
Carbon (wt-%)	56.94	85.86	85.6	8.4	84.39	88.11	83.58	83.44	84.41	90.67	89.52
Hydrogen (wt-%)	13.00	13.56	14.20	16.09	15.54	11.60	16.29	16.49	15.54	9.32	10.34
Nitrogen (ppmw)	46	37	27	0	0	0	0	0	0	0	0
Sulfur (ppmw)	321	522	0	15	15	9	325	10	7	10	8
Heating value (Btu lb^{-1}) (net)	17 300	17 300	18 700	18 400	18 100	18 800	18 400	18 500	18 200	15 500	17 300

2.3 Cut Fraction Characterization – Manual Methods

2.3.1 Establishing the True Boiling Point Curve – Methods

The typical means of measuring molecular size distribution is with a volumetric TBP distillation. This is a laboratory-based batch distillation whereby the overhead recovery of a percentage of the initial charge is recorded versus the average temperature of the bottoms in the still pot. The following methods are used in industry today.

2.3.1.1 Distillation via American Society for Testing and Materials D-86 Procedures (Usually at 760 mmHg Pressure) This procedure is used for lower boiling hydrocarbon cuts such as gasoline, kerosines and even some light diesel oils. This test is a batch distillation in which no vapor fractionation occurs, so the the still vapor temperature is theoretically the dewpoint of a continuously changing (increasing in MW and therefore boiling point) hydrocarbon stream. Although not directly related to composition, this test is still a primary test because of its low cost and high degree of standardization and reproducibility. Nearly all fuel product specifications are expressed in terms of a D-86.

2.3.1.2 Distillation via American Society for Testing and Materials D-1160 Procedures (Usually at 10 mmHg Pressure) This procedure is used primarily for higher boiling hydrocarbon cuts such as heavy oils, resids, fractionator bottoms, etc. The typical vacuum conditions provide a pressure at which a significant recovery of material can be obtained without the need to use high temperatures to vaporize. Hydrocarbons start cracking around 316–343 °C (600–650 °F), so the longer the analyst can carry out the distillation under lower temperatures, the more accurate the analysis of the heavy ends in the initial still charge will be.

This test is a batch distillation in which no vapor fractionation occurs, so the still vapor temperature is theoretically the dewpoint of a continuously changing (increasing in MW and therefore boiling point) hydrocarbon stream. Although not directly related to composition, this test is still a primary test because of its high degree of standardization and reproducibility. Nearly all heavy product specifications are expressed in terms of a D-1160.

2.3.1.3 Distillation via Experimental True Boiling Point Procedures TBP distillation is meant to represent the ''true'' molar boiling point of each incremental amount of recovered overhead still product, i.e. if each small amount of distillate liquid were in fact a pure component, its NBP would correspond to the still pot temperature at the time it went overhead. In practice, this cannot occur, but by increasing the fractionation efficiency of the still, the MW distribution of the overhead liquid can be narrowed. In this way, a fairly good estimate of the ''average'' molar boiling point of a distillate cut at a given temperature can be made. The still efficiency is dependent on the number of rectifying plates and the amount of reflux that is used to maintain liquid traffic in the trayed section. Typical TBP columns (Oldershaw and some of the newer systems) use what is known as a 15/5 specification, that is 15 trays with a 5 : 1 reflux ratio. This standard, however, is more traditional than anything, and most TBP cuts are taken under conditions ranging from 5/3 to 60/20, depending upon the equipment and the care taken by the laboratory analyst.

ASTM D-2892 is as close as anything to a standard for this test. It is prescribed only for a full-range crude analysis but should work well for any wide-boiling cut.

2.3.1.4 Distillation via American Society for Testing and Materials D-2887/D-3710 Gas Chromatography Procedures These are newer procedures, based on a GC reconstruction of the boiling point curve, which are quickly becoming the preferred reporting methods for literature and operations data. They are quick and accurate, although the required equipment is somewhat expensive. ASTM procedure D-3710 covers the analysis of stocks with NBPs <260 °C (500 °F). D-2887 theoretically covers a full-range crude but must typically be limited to NBPs <538 °C (1000 °F). Any molecules that are too large are either cracked in the elution column or end up as coke on the column packing.

D-2887 results are often close enough to ''true'' TBP that some laboratories will substitute them directly. Several correlations exist, however, for converting D-2887 values to ''true'' TBPs and D-86 assays, and should be used preferentially.

Data concerning any hydrocarbon fraction will ultimately be obtained in one of the above forms. Occasionally, the curve for mid-range cuts will be reported as a composite, the lower boiling cuts being measured via the D-86 procedure and the higher boiling cuts being reported as D-1160. This protocol is fine as long as the assumptions are made clear in the reported data.

Several laboratory methods today combine a chromatography step with some form of additional downstream analysis of the eluate. Spectral emission is the most common supplemental analysis method, with gas chromatography/infrared (GC/IR) spectroscopy and gas chromatography/mass spectrometry (GC/MS) being the most advanced in terms of broad applicability. High-performance liquid chromatography (HPLC) with hydrodealkylation (HDA) is promising for determining aromatics in 343+ °C (650+ °F) fractions, but is limited for other components and boiling point ranges. This advanced equipment is also expensive and difficult for

all but trained laboratory operators to use. Therefore, its general usage will grow slowly for some time.

2.4 Establishing the True Boiling Point Curve – Where to Start

The normal method of petroleum fraction characterization requires the generation of a TBP curve. Unless measured directly, a fairly time-consuming and costly procedure, a TBP curve must be derived from another form of distillation.

Figure 3 shows typical relationships between an ASTM distillation curve and a TBP distillation curve for a full-range cut. The ASTM curve has had the 371 °C (700+ °F) portion of the curve corrected to 760 mmHg pressure, because its data were taken at 10 mmHg via the D-1160 procedure, but the relationship of the resulting curve to the more tightly separated TBP curve is still valid. The equilibrium flash vaporization (EFV) curve shown on Figure 3 is an EFV curve of the type that used to be an industry standard method for estimating a single-stage ideal flash. With the availability of inexpensive computer software for stage flashes, this particular curve has fallen into general disuse and will not be covered here.

Several methods have evolved for relating the ASTM curves to a TBP curve. The classical graphical procedure based on both the Edmister–Pollock[70] and Edmister–Okamoto[71] methods is illustrated in Figure 8.

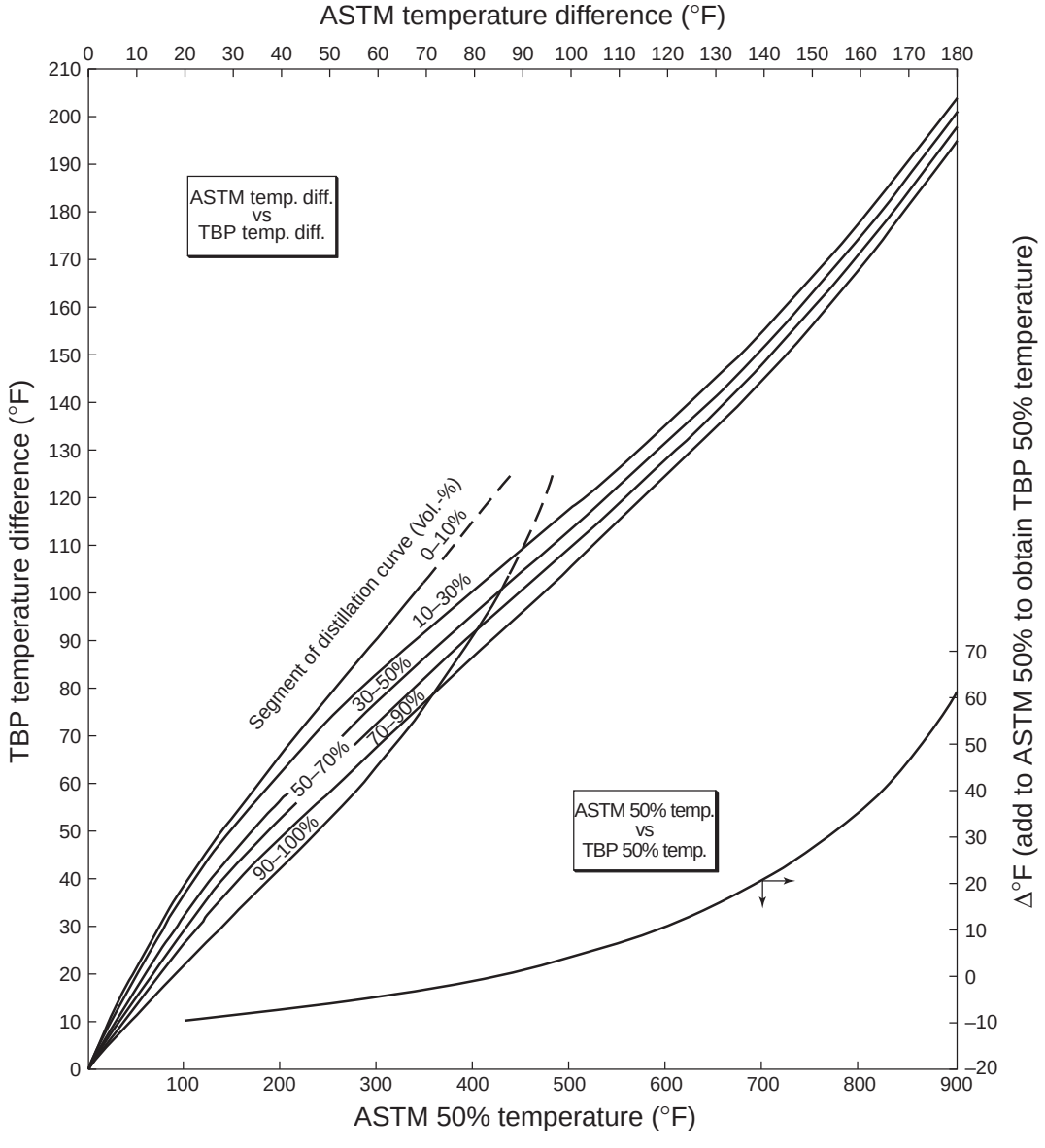

Figure 8 ASTM–TBP relationships.[70,71]

These methods are still recommended by the API[9] and are used by refining engineers. Some use one method exclusively but the blend of the two seems to give the most uniform results. Most companies, and the API itself, have provided computer-compatible equations for distillation conversions based on Figure 8 or refinements thereof. A more detailed discussion of these procedures is covered in subsequent sections of this article.

Figure 9 shows the ''preferred'' methodology of the API Technical Data Committee as the solid lines and the author's suggested deviations as the dashed lines. These procedures provide reasonable and consistent answers.

2.5 Quality of Data

The major problem in defining a TBP curve is not so much the manipulation of the data but rather the quality of the data themselves. ASTM distillations are defined by fairly rigid standards and as long as the equipment meets the standard and is run properly (most new analytical units are digitally controlled), the main source of error is the care with which the analyst reports the data. The quality of directly reported TBP curves, however, has error inherent from both the distillation conditions and the method of analysis, in addition to the biases and errors of the analyst.

Since the generation of ''boiling-lumped'' pseudocomponents requires a TBP curve, the real problem becomes one of defining the curve from the ''best-quality'' data. A good analyst with good equipment, generating a ''true'' TBP curve, is the optimum situation. The word ''true'' is in quotes because sometimes reported TBPs are actually ASTM distillations that have been transformed by either the analyst or the apparatus (the newer ones under microchip control). Use of these latter TBP values leaves one at the mercy of the transform method. Any such transformed data should be thoroughly rechecked before use.

Given the existence of a real or assumed TBP curve, the engineer then starts to divide the curve into volume-specific cuts defined by their corresponding segmental end-point temperature. Figure 12(a) and (b) (see later) shows how this might be done.

In any case, the reader must realize that converting from one distillation curve to another is somewhat of an inexact procedure, and errors as great as 17–22 °C (30–40 °F) may occur with wide-boiling stocks. It is essential, therefore, that the chemist or engineer be realistic about the accuracy of estimated distillation curves. This means checking important results by more than one method, and allowing reasonable safety margins so that the overall workability of a design is not jeopardized by unavoidable inaccuracies.

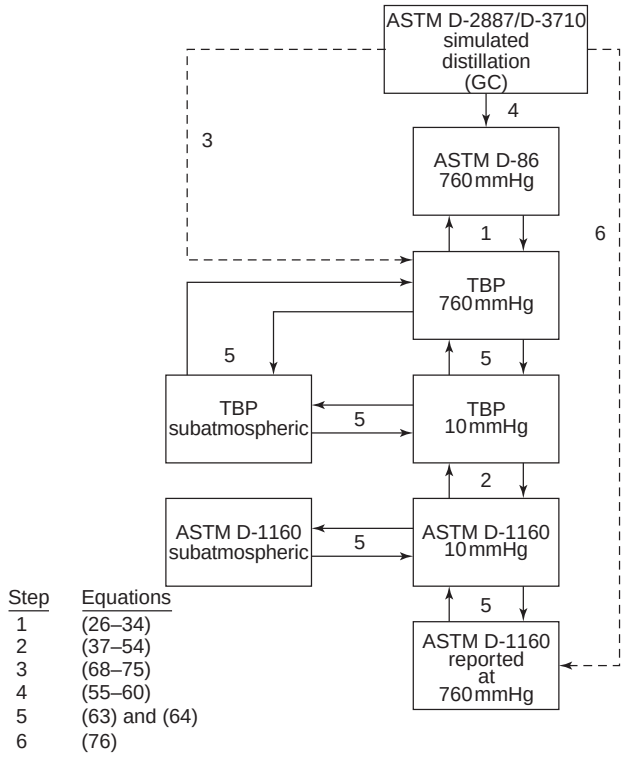

Figure 9 Distillation conversion pathways. Solid lines are API-sanctioned methods. Dashed lines are alternative means of estimation. These may or may not be consistent with conversion results from a different path.

2.6 Methods to Use When Lacking "Total Data"

The above recommendations are dependent upon the existence of some *real* distillation data on the hydrocarbon cuts in question. When such data are not available, the engineer must use judgment and some knowledge about the relative amounts of the various types of components in the stream.

For estimating an ATM tower resid, one may have only a sample of the full-range crude and an idea of the fractional residual. Knowing that most resids (at least those from reasonably efficient fractionation systems) have TBPs starting around 260–288 °C (500–550 °F), one may elect simply to cut off the resid tail of the ''full-range TBP'' at the volume percentage corresponding to the fraction residual and use that as the heavy fraction TBP curve, normalizing the cut fractions to a full 100% basis.

An extension of this line of thinking allows one to specify a full-range cut based on some assumptions about the various parts of that whole stream. One might characterize the effluent of a hydrocracker as consisting of 20 standard vol.-% of light naphtha cuts [32 °C (90 °F) < NBP < 82 °C (180 °F)], 30% heavy naphtha cuts [82 °C (180 °F) < NBP < 196 °C (385 °F)], 40%

of mid-distillate cuts [196 °C (385 °F) < NBP < 287 °C (550 °F)] and the residual.

From these assumptions, a TBP curve for the full cut can be estimated. Although not as accurate as distillation curve analysis, this method will allow the engineer to proceed with work in the absence of ''real'' data.

Typical procedures, assuming the availability of some TBP data, are illustrated in Table 7.

In 1987, the American Petroleum Institute (API)[10] also developed a correlation for interconversion of D-2887 and TBP data. The original intent was to develop the correlation at each ''standard'' volume node (10, 30, 50, 70, 90% points).

In general, the correlation did well except for several diesel–gas-oil range boiling cuts that had a fairly ''flat'' boiling range from the 30–70% range. In such cases, the correlation predicted regressive values of at least one of the 50, 70 or 90% points. This did not go over well in the refining community and the correlation was revised to its current form; see sections 3.12.2.1 and 3.12.2.2 for descriptions. In spite of the problems, it is still an excellent method of predicting TBPs from a very limited amount of data. If one has only 10, 50 and 90% D-86 values (not an uncommon situation), this method may give better results than using interpolations to calculate difference values.

If one must *interpolate* from data not on the ''standard'' reporting volumes nodes, use a quadratic or spline fit procedure. To extrapolate (very carefully – avoid if at all possible!), use a probability plot format. The MWs, and thus the boiling points, of the stream components can be assumed to be normally distributed about a mean value.

Table 7 Construction of full-cut TBP curve

1. If the residual is entirely heavy (>550 °F) material, the TBP will be as follows:

TBP (°F)	Cumulative vol.-% overhead
90	0
180	20
385	50
550	90

2. If the residual is entirely light (<90 °F) material, the TBP curve will be as follows:

TBP (°F)	Cumulative vol.-% overhead
90	10
180	30
385	60
550	100

3. If the residual is partly lights and heavies, the TBP curve, assuming a 2% lights/8% heavies residual ratio as an example, will be as follows:

TBP (°F)	Cumulative vol.-% overhead
90	2
180	22
385	52
550	92

2.7 Establishing the Pseudocomponents

Whereas arrival at a ''final'' TBP curve may be accomplished in several ways, the method of defining pseudocomponents (or commonly called hypotheticals) has less variation.

Petroleum fraction pseudocomponents are traditionally defined in terms of their NBPs. In the simplest method of determination, a certain boiling range of a TBP curve is segregated and the mid-volume boiling temperature is found graphically. In looking at Figure 4, one might define nine pseudocomponents by the heavy cut lines crossing the TBP curve. The components would then be identified as NBP140, NBP316, NBP413, NBP490, NCP557, NBP621, NBP697, NBP794 and finally, NBP1000, as in Figure 4. One could rename them to corresponding celsius, kelvin or any other meaningful scale for tracking purposes.

This last mid-range temperature is estimated from an assumed end-point for the heavy end of the curve. It is most difficult to measure the boiling points of heavy resids, as they tend to be destroyed (cracked) in the process of having their boiling point measured. Unfortunately, this lack of a definable end-point is a major source of error in the characterization of resid fractions. One can use some advanced interpolation methods (details of which are beyond the scope of this article) to aid in end-point estimates, but even these are only of marginal value. It should also be noted that MW correlations based on NBP and gravity become increasingly inaccurate for higher boiling point material.

Using the classical TBP cut method of determining pseudocomponents has its limitations in heavy resid processing.

2.8 Calculation of the Average Cut Parameters

Once the TBP range for the pseudocomponent has been defined, the calculation of MW and related parameters can commence. Nearly all correlations require the NBP and specific gravity (SG) of a cut. Some use these two parameters completely separately and some incorporate the SG in a Watson K (K_W) parameter. Implicit with any TBP curve is a related gravity versus volume fraction curve or a gravity–boiling point curve.

In general, the SG of the overhead product increases as the distillation progresses (i.e. as both volume-% and/or distillation temperature increase). In many instances, the SG curve is reported along with the temperature curve. This is the ideal situation, in that all quantities necessary to define the MW are specified.

Usually just the average whole-stream gravity is reported. In this case, the best procedure is to estimate the individual pseudocut gravities based on some idea of what the SG of the cut does with increasing NBP.

Figure 10 shows a typical representation of a straight-run cut SG–volume-% curve and one of a "cracked" fraction. Both are monotonic, with a constantly increasing dependent function value. The cracked fraction, however, has an inflection point (where the second derivative changes).

Figure 11 shows a typical representation of cut fraction K_W–vol.-% curves for several different hydrocarbon stream types. The characteristic of most interest to refiners is the fact that for thermally cracked and hydrocracker effluents, the K_W of the mid-distillate material tends to be low in comparison with the tails. The explanation for this is fairly simple. K_W is strongly correlated with the degree of paraffinicity and inversely correlated with the degree of aromaticity in a fraction. Most hydrocracking forms paraffinic products, highly saturated although often highly branched. The hydrocracking process and reforming processes also produce a large number of single-ring aromatic compounds with boiling points in the region of kerosine and light diesel with a corresponding reduction in average K_W in this area. In most straight-run fractions, by contrast, the lightest components tend to be paraffinic and the heavy components are more aromatic and naphthenic, giving a continually decreasing K_W vs volume percentage curve.

Defining cut gravity in the absence of real data depends chiefly upon how the variation of the K_W is perceived. For straight-run crudes, the assumption of a constant K_W, based on either the cut-point TBP (preferred by most of the industry) or ASTM-D-2887 mid-point temperature, is the usual method. The engineer, however, knowing something about the true K_W curve, may want to make some educated guesses about the local variations from this assumption.

A few comments should be made with respect to Figure 11. The curves are typical for shape only. Actual plot values are dependent upon the stream and can vary greatly from those shown. Within whole crude cut fractions, constant K_W is not a bad assumption, especially for straight-run fractions. For hydrocracked cuts, it may be invalid since aromatics accumulation from the conversion process tends to build up in the middle distillate range. Crude light straight run (LSR) tend to be more paraffinic on average than the whole crude. This generalization also holds for full-range hydrocracker effluents. Highly paraffinic crudes also have a region of increasing K_W but it tends to start with 650 °F+ NBP material and the change is not as pronounced as with the mid-range hydrocracker effluents. In the case of hydrocracker effluents, the double-inflected curve in Figure 11 may be a more appropriate type to follow.

2.9 Measures of "True" Boiling Point

Once the NBP and SG of a cut have been set, the MW can be obtained in a straightforward fashion. Details of the best current methods of calculating hydrocarbon cut fraction MWs are discussed in section 3 and will not

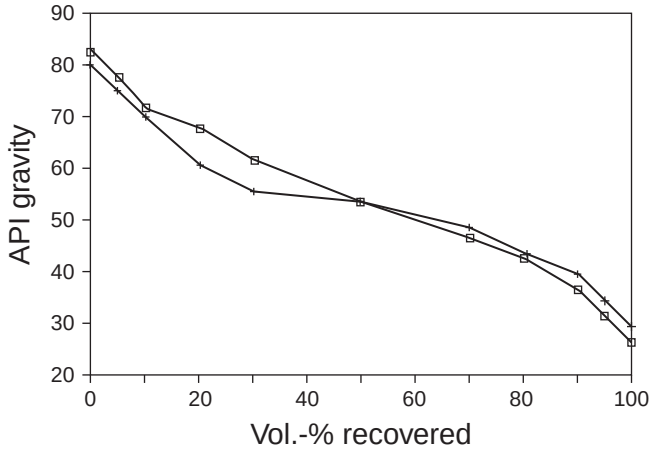

Figure 10 API gravity vs distillation volume-% (ASTM, TBP, etc.). The fractions are of identical average gravity but are not necessarily related by any other means. Curves are for illustrative purposes only. , Typical straight-run cut; +, typical cracked cut.

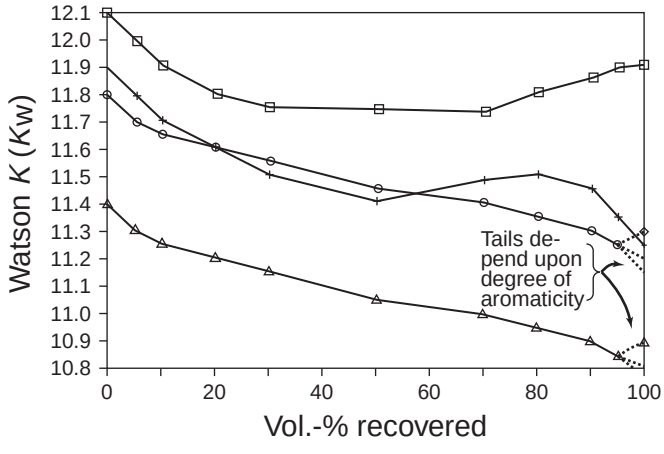

Figure 11 K_W vs distillation volume-% (ASTM, TBP, etc.). , Highly paraffinic crudes; +, typical full-range R-2 effluent; ○, naphthenic crude 1; △, naphthenic crude 2.

be repeated in any great detail here. What is of some controversy is just what the NBP measures. In any cut, there are several measures of a mean boiling temperature.

There is the volumetric average boiling point (VABP), which is *not* identical with the mid-boiling point temperature (see Figure 12a and b), the CABP, which is the theoretical basis for K_W calculations, the MABP (defined in Figure 12a and b) and the MEABP, which the API defines as the average of the CABP and the MABP. These values are outlined graphically in Figure 12(a) and (b) and explained as numerical approximations below.

Several methods typically used by oil companies and process engineering contractors use the VABP, a very simple number to generate or estimate. The Twu[142] correlation used in some process simulation packages such as ''Process/Provision'' uses MABP, and the most recent correlation approved by the API Technical Data Committee uses MEABP, as did earlier versions of their methodology.

Note that for narrowly defined TBP temperature cuts, all these measures of ''average'' boiling point tend to converge. It is only when trying to characterize a wide-boiling fraction that the differences in the NBP values become significant. An exception is in the use of mid-range boiling points for the front and back ends of a cut (i.e. the 0–10 and 90–100% ranges). The mid-range value in these cases can be significantly lower than the VABP (5% vs 7% on a steeply sloped curve). An extra effort should be made to define the VABP better for these tail cuts.

The five primary boiling point vectors can be estimated by algebraic means, given a defined mixture. The temperature units of the first three distillation data functions do not matter as long as they are consistent with each other and with that of the calculated average value. The quantities are defined by Equations (5–8):

$$\text{VABP} = \text{SUM}[V_i \times Tb_i] \tag{5}$$

$i = 1 \rightarrow$ No. of components, where VABP is the volumetric average boiling point, V_I is the volume fraction of component i, and Tb_i is the NBP of the ith component;

$$\text{MABP} = \text{SUM}[X_i \times Tb_i] \tag{6}$$

$i = 1 \rightarrow$ No. of components, where MABP is the molar average boiling point, X_i is the volume fraction of component i and Tb_i is the NBP of the ith component;

$$\text{WABP} = \text{SUM}[W_i \times Tb_i] \tag{7}$$

$i = 1 \rightarrow$ No. of components, where WABP is the weight average boiling point, V_i is the volume fraction of component i and Tb_i is the NBP of the ith component;

$$\text{CABP} = \text{SUM}[V_i \times Tb_i^{1/3}]^3 \tag{8}$$

$i = 1 \rightarrow$ No. of components, where CABP is the cubic average boiling point, V_i is the volume fraction of component i and Tb_i is the NBP of the ith component.

The temperature of the distillation data for Equation (8) must be absolute (in either degrees rankine or kelvin) and must be consistent with each other and with that of the calculated average value.

$$\text{MEABP} = \frac{\text{MABP} + \text{CABP}}{2} \tag{9}$$

where MEABP is the volumetric average boiling point and CABP and MABP are defined as above.

The temperature of the three distillation averages must be consistent with each other. MEABP is typically considered to be the best approximation for the NBP of a cut and, if available, is the temperature of preference. For narrow cuts, however, all of the values should converge within a few degrees of each other.

Given the fact that full-range analyses of crudes and fractions of crudes are time-consuming and costly, the API has developed the following correlations for a broad boiling range hydrocarbon.

From D-86 distillation data, find the VABP with five standard cut values (Equation 10):

$$\text{VABP} = \frac{T_{10} + T_{30} + T_{50} + T_{70} + T_{90}}{5} \tag{10}$$

If data are scarce and a calculation has to be made, one could use (as have some) Equation (11):

$$\text{VABP} = \frac{T_{10} + T_{50} + T_{90}}{3} \tag{11}$$

The author feels, however, that it would be better to intuitively estimate a T_{30} and a T_{70}, then use Equation (10).

The VABP and subsequent ''average'' derivation methods can also be used with TBP curves, provided that they are identified as such, and that a corresponding D-86 vector has been established corresponding to the TBP curve. Often the D-86 and TBP distillation averages are used interchangeably (and improperly) by novices. The primary reason why the D-86 values need be present is for calculating 10–90% slopes (Equation 12):

$$SL_{1090} = \frac{T_{90} - T_{10}}{90 - 10} \tag{12}$$

The other assay average parameters are calculated according to Equations (13–24):

$$\text{WAPB} = \text{VABP} + \text{Temp_diff_wt} \tag{13}$$

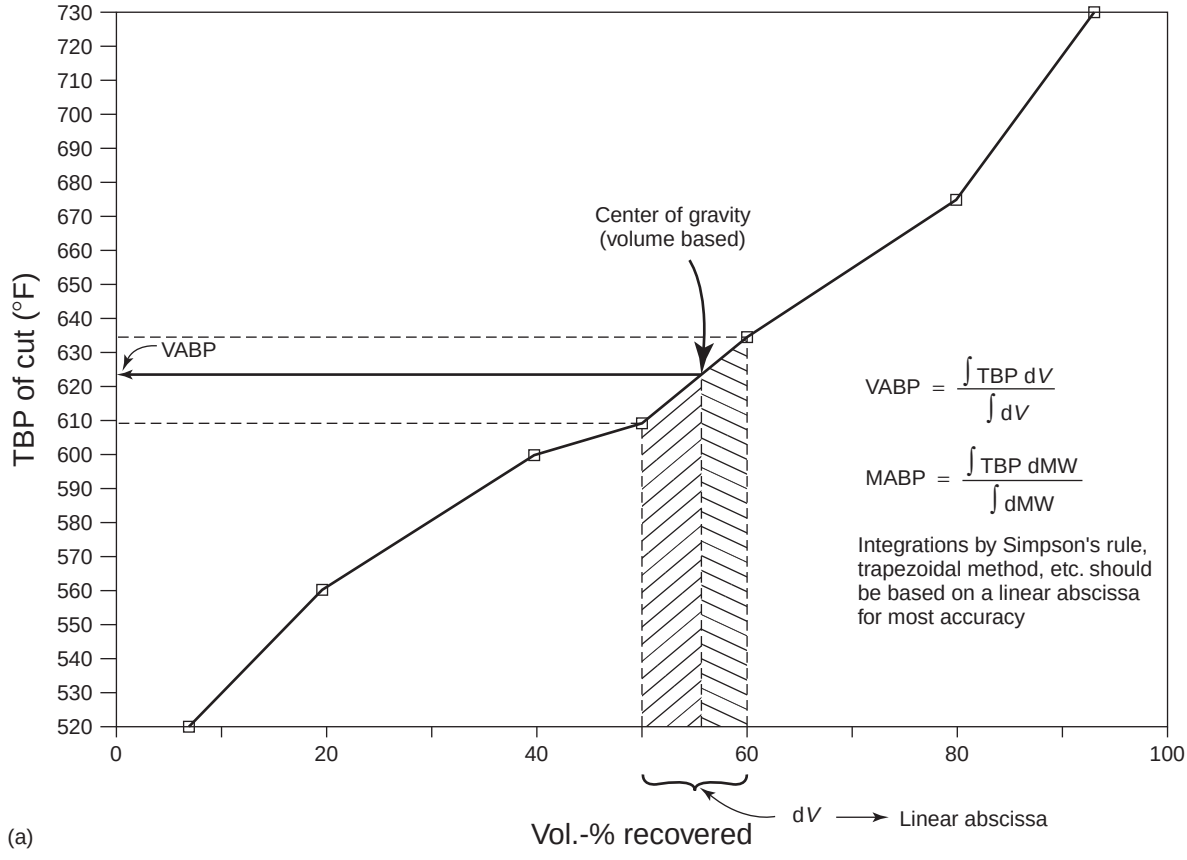

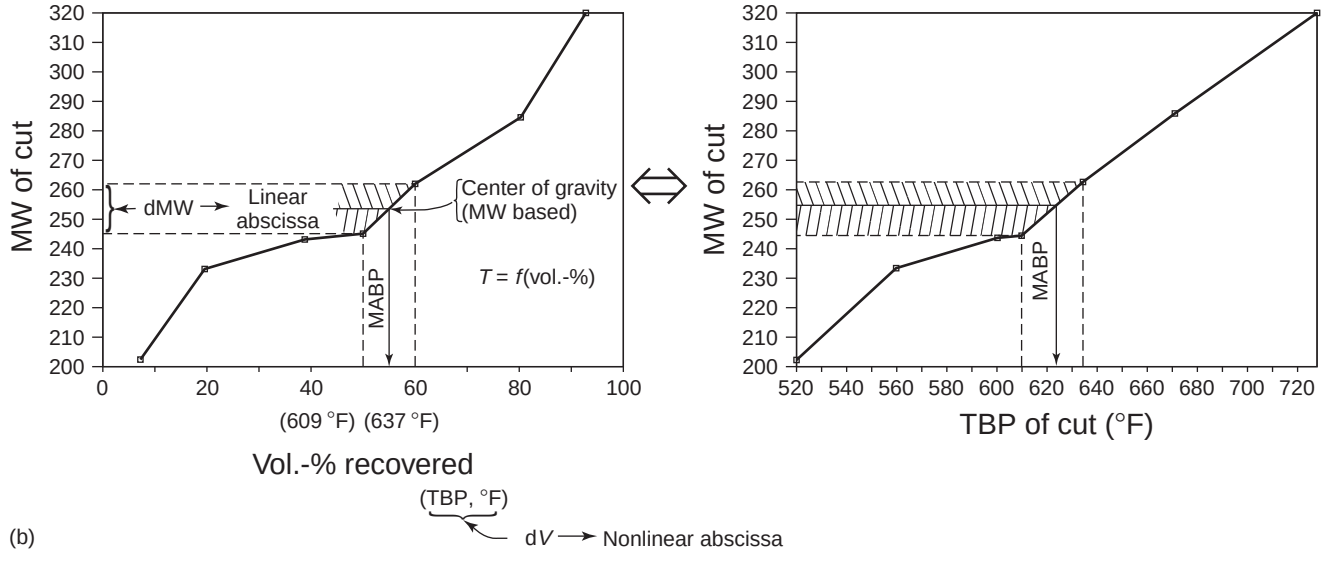

Figure 12 Determination of (a) volume average NBP and (b) molar average NBP. The Watson K uses the cubic average boiling point, $\text{CABP} = \sqrt[3]{\int \text{VABP}^3 \, dV \quad dV}$. Most MW correlations use the molar average boiling point, MABP. 1988/91 API revised correlations use the mean average boiling point, MEABP, which is the average of the cubic and molar average boiling points, $\text{MEABP} = (\text{CABP} + \text{MABP})/2$.

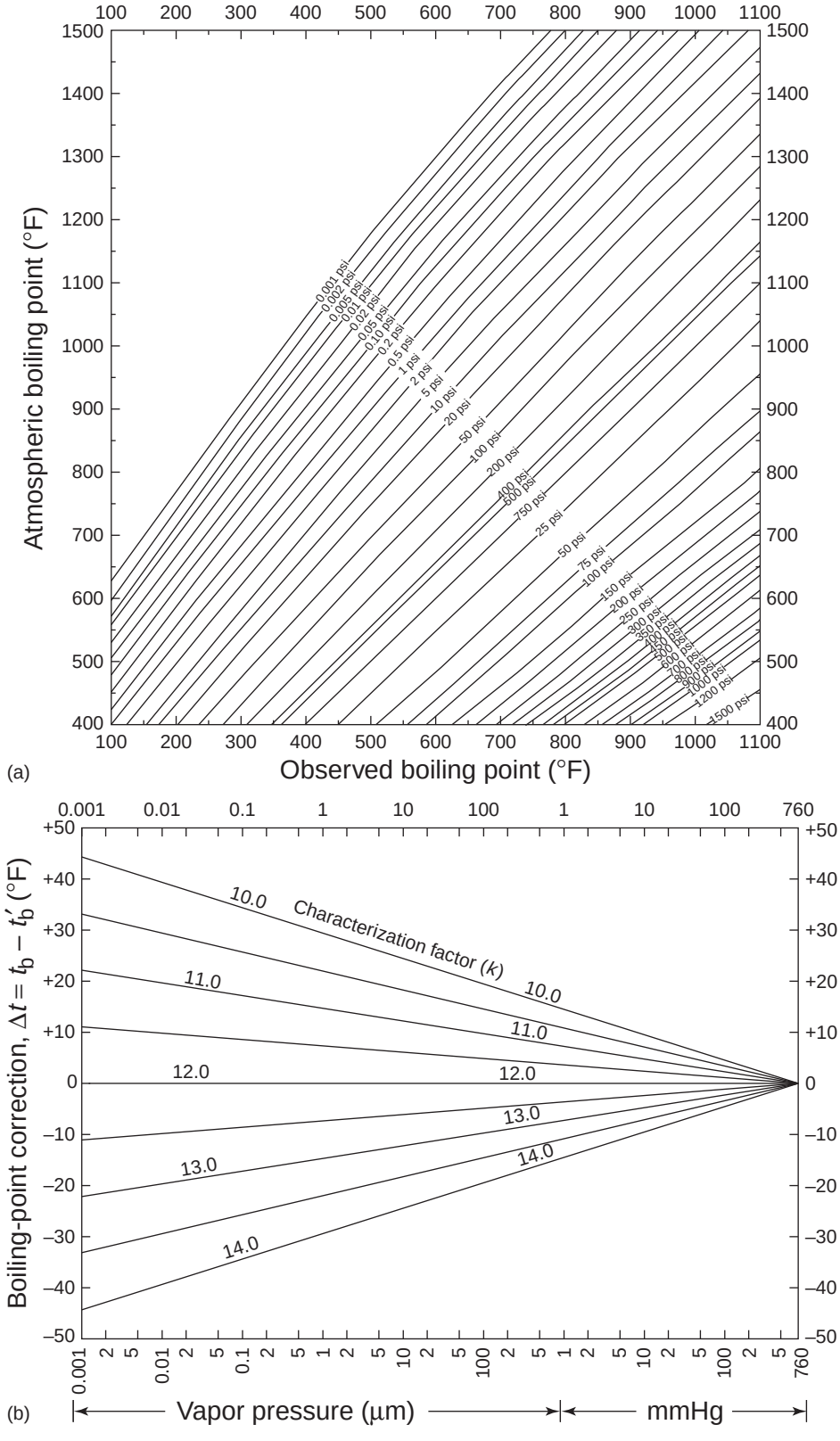

Figure 13 (a) Boiling point conversion chart; (b) Watson K (K_W) correction for (a). Note the logarithmic abscissa in (b).[9,97]

where Temp_diff_wt is determined by Equation (14):

$$\ln(\text{Temp_diff_wt}) = -3.062 - 0.01829(\text{VABP} - 32)^{0.6667} + 4.458SL_{1090}^{0.25} \quad (\text{units of } ^\circ\text{F}) \tag{14}$$

$$\ln(\text{Temp_diff_wt}) = -3.649 - 0.0271(\text{VABP})^{0.6667} + 5.614SL_{1090}^{0.25} \quad (\text{units of } ^\circ\text{C}) \tag{15}$$

$$\text{MABP} = \text{VABP} - \text{Temp_diff_mo} \tag{16}$$

where Temp_diff_mo is determined by Equation (17):

$$\ln(\text{Temp_diff_mo}) = -0.5638 - 0.007981(\text{VABP} - 32)^{0.6667} + 3.047SL_{1090}^{0.3334} \quad (\text{units of } ^\circ\text{F}) \tag{17}$$

$$\ln(\text{Temp_diff_mo}) = -1.152 - 0.01181(\text{VABP})^{0.6667} + 3.707SL_{1090}^{0.3334} \quad (\text{units of } ^\circ\text{C}) \tag{18}$$

$$\text{CABP} = \text{VABP} - \text{Temp_diff_cu} \tag{19}$$

where Temp_diff_cu is determined by Equation (20):

$$\ln(\text{Temp_diff_cu}) = -0.2359 - 0.0691(\text{VABP} - 32)^{0.45} + 1.886SL_{1090}^{0.45} \quad (\text{units of } ^\circ\text{F}) \tag{20}$$

$$\ln(\text{Temp_diff_cu}) = -0.8237 - 0.08997(\text{VABP})^{0.45} + 2.457SL_{1090}^{0.45} \quad (\text{units of } ^\circ\text{C}) \tag{21}$$

$$\text{MEABP} = \text{VABP} - \text{Temp_diff_nbp} \tag{22}$$

Table 8 Comparison of boiling point averages of four different cuts at different VABPs

VABP (°C)	Engler slope (°C %$^{-1}$)	Type of boiling point	Value from chart (°C)	Calculated value (°C)
100	2.0	WABP	106.3	106.7
200	0		200.0	200.0
300	2.8		306.3	306.2
400	5.0		414.6	413.5
100	0.5	MABP	95.1	95.4
200	2.3		172.5	171.9
300	0.8		293.9	294.2
400	3.0		364.9	365.1
100	1.0	CABP	97.4	97.5
200	1.3		197.2	197.4
300	3.8		288.0	288.0
400	2.0		396.8	396.7
100	0.3	MEABP	98.0	98.1
200	3.3		167.7	168.2
300	1.8		290.1	289.7
400	4.0		364	364.7

where Temp_diff_nbp is determined by Equation (23):

$$\ln(\text{Temp_diff_nbp}) = -0.9440 - 0.00865(\text{VABP} - 32)^{0.6667} + 2.998SL_{1090}^{0.3334} (\text{units of } ^\circ\text{F}) \tag{23}$$

$$\ln(\text{Temp_diff_nbp}) = -1.532 - 0.0128(\text{VABP})^{0.6667} + 3.617SL_{1090}^{0.3334} (\text{units of } ^\circ\text{C}) \tag{24}$$

Table 8 shows a comparison of these boiling point averages of four different cuts at different VABPs.[157]

2.10 Measures of Fractionation Efficiency

Distillation assays can have value within themselves. Fractionation performance in crude units and other refinery fractionators is commonly specified as a targeted distillation value to be achieved, such as a D-2887 (90%) value maximum for CARB (California Air Resources Board) gasoline, though this latter method is more often used to set draw rates. A pure measure of column fractionating capacity is often gaged as the *difference* between the ASTM 5% point of a product and the ASTM 95% point of the next lighter product. This ''gap'' is defined according to Equation (25):

$$\text{gap} = \text{heavy curve D86}_{5\%} - \text{lighter curve D86}_{95\%} \tag{25}$$

A high degree of fractionation will result in the 5% point being higher than the 95% point, producing a so-called positive ''gap'' value. Poor fractionation, on the other hand, will result in an overlap where the 95% point is actually higher than the 5% point, or a negative ''gap''. The required fractionation performance is generally specified by the process engineer and is generally higher for the lighter product cuts unless the column has some mechanical problems.

Gaps of 17–28 °C (30–50 °F) are commonly found between naphtha (gasoline) cuts and kerosine or between kerosine and light diesel. Gaps of 0 to −11 °C (0 to −20 °F) are more common between heavier cuts. Figure 14(a–c) illustrates this situation for a full-range crude.

Note that if the fractionation is perfect, the TBP curves of the cuts form a complete unbroken curve that is identical with the TBP curve of the fractionator feed. Perfect fractionation, however, is never achieved in commercial side-draw columns. Some of the heavier product is unavoidably withdrawn with the lighter product and vice versa, thus forming the ''tails'' shown in Figure 14(b). The poorer the degree of fractionation, the greater will be the deviations of the TBP tails from the feed TBP curve, and consequently the greater will

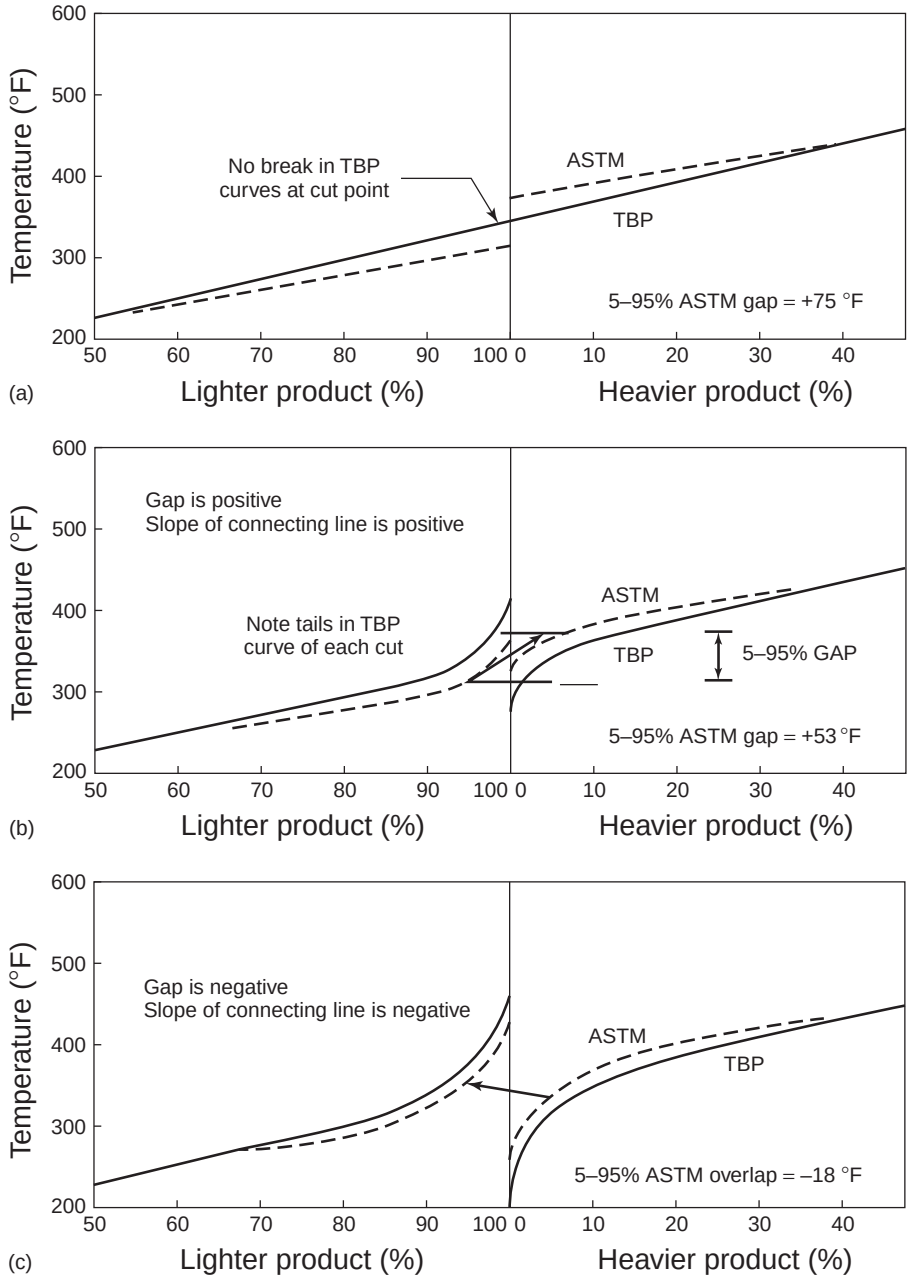

Figure 14 Degrees of fractionation: (a) perfect, (b) good and (c) poor fractionation.[150]

be the deviations of the respective ASTM curves until, eventually, overlap of the adjacent ASTM and TBP 5/95 curve values occurs.

2.11 Interconversion of Distillation Data[9–15]

2.11.1 Introduction

The general use of the TBP (and other) distillation curves as part of the characterization process for cut analyses has been covered to this point. The major focus of this section is to enable the chemist or engineer to arrive at an ATM TBP curve for a stock, particularly from data that were derived from an ASTM standard test. From these data, several other curves and physical properties can be estimated with good accuracy. Very often, these methods can be used (usually with some mathematical manipulation) to impute various inspection data in hydrocarbon streams.

The most important thing to realize is that converting from one distillation curve to another is a rather inexact procedure. Errors as great as 17–22 °C (30–40 °F)

may occur with wide-boiling stocks. It is essential, therefore, that the engineer be realistic about the accuracy of estimated distillation curves. This means checking important results by more than one method, and allowing reasonable safety margins so that the overall workability of a design or analysis dependent upon this data is not jeopardized by unavoidable inaccuracies in the correlations.

Several methods for converting one distillation curve to another are presented in this section. Most are currently recommended by the API Technical Data Committee, the premier nonconfidential information source currently available for hydrocarbon stock thermodynamics. A few are useful correlations that the author has used for this type of work.

In general, for converting an ASTM curve to a TBP curve (or vice versa), Edmister's methods (as shown in Figure 8) and its recent empirical derivatives are generally recommended.

The classical correlation for 50% point ASTM $\rightleftharpoons$ TBP is based on work by Edmister and Pollack[70] in 1948 and is considered by the API to be superior to a later correlation published by Edmister and Okamoto[71] in 1958. In using Figure 8, the 50% point temperature of the required curve is determined from the 50% point of the known curve. A recent API correlation for this graphical method has been published, where the 10, 30, 50, 70 and 90 vol.-% points are known on one or the other side. This method is probably the most accurate available procedure, assuming that one has enough data.

Since most refineries and petrochemical plants try to get by with the minimum amount of expensive on-line analyzers, these methods are well tested.

In a few cases, an analyzer will only report one or two specific volume-% points. Sometimes the residual portions of the distillation curve cannot be obtained. For these situations, one can use an older API correlation based on inter-conversion from one corresponding volume-% directly to that of the target curve, without assuming knowledge of the rest of the source or target value curves. There are some caveats that the user must observe. Trying to construct a full-range curve from the point method can lead to problems, which will be described in detail below. If one is able to generate reliably a full-range ASTM or TBP distillation, the full-range methods should be used.

In any event, this correlation has been useful to the author and can be utilized as a check on the values derived from the other methods.

2.11.2 Physical Distillation Conversion Procedures

The most recent revisions of the API Technical Data Book (1987–98) contain several useful procedures for interconversion of "standard" mixed-hydrocarbon distillation curves.

Figure 9 (solid lines) shows the 1994 API recommended conversion pathways for constructing one curve type from another. An explanation of these procedures, or their acceptable alternatives (dashed lines), follows.

2.11.2.1 Interconversion of American Society for Testing and Materials D-86 and True Boiling Point Distillation Curve at Atmospheric Pressure Full Range Assay[9]

This method should be used only for a fully defined assay of a stream. Trying to derive a full TBP curve from three or more such point applications of this method can create problems.

Equations (26) and (27) are used to interconvert ASTM D-86 and ATM TBP distillation data:

$$\text{TBP}_{50\%} = a(\text{ASTM D-86}_{50\%})^b \tag{26}$$

$$\text{ASTM D-86}_{50\%} = \left(\frac{\text{TBP}}{a}\right)^{1/b} \tag{27}$$

where $a = 0.8718$ and $b = 1.026$ for the interconversion of the 50% points, TBP temperatures at 0, 10, 30, 50, 70, 90 and 95 vol.-% distilled, in degrees Rankine, and ASTM D-86 = observed ASTM D-86 temperatures at the corresponding volume-% distilled, in degrees Rankine.

The procedure works by calculating differential temperatures from the mid-volume boiling point value and translating that value to the other corresponding member of the pairing.

The difference between adjacent cut points can be determined from a series of equations for each interval (Equation 28):

$$\text{TBP_diff(I)} = C_0(\text{D-86_diff})^{C_1} \tag{28}$$

where C_0 and C_1 are coefficients for varying temperature ranges, TBP_diff( ) and D-86_diff( ). Their vectors are shown in Table 9. Temperature values and differences must be in consistent units.

To determine the TBP values at any percentage distilled, add or subtract the proper differentials from

Table 9 Coefficients for ASTM D-86–TBP conversion

Difference vector index value	Cut point range (vol.-%)	C_0	C_1
1	100–90	0.1180	1.661
2	90–70	3.042	0.750
3	70–50	2.528	0.820
4	50–30	3.031	0.8008
5	30–10	4.900	0.7164
6	10–0	7.401	0.6024

the predicted 50% TBP temperature (Equations 29–34):

$$\mathrm{TBP}_{0\%} = \mathrm{TBP}_{50\%} - \mathrm{TBP_diff}(4) - \mathrm{TBP_diff}(5) - \mathrm{TBP_diff}(6) \quad (29)$$

$$\mathrm{TBP}_{10\%} = \mathrm{TBP}_{50\%} - \mathrm{TBP_diff}(4) - \mathrm{TBP_diff}(5) \quad (30)$$

$$\mathrm{TBP}_{30\%} = \mathrm{TBP}_{50\%} - \mathrm{TBP_diff}(4) \quad (31)$$

$$\mathrm{TBP}_{70\%} = \mathrm{TBP}_{50\%} + \mathrm{TBP_diff}(3) \quad (32)$$

$$\mathrm{TBP}_{90\%} = \mathrm{TBP}_{50\%} + \mathrm{TBP_diff}(3) + \mathrm{TBP_diff}(2) \quad (33)$$

$$\mathrm{TBP}_{100\%} = \mathrm{TBP}_{50\%} + \mathrm{TBP_diff}(3) + \mathrm{TBP_diff}(2) + \mathrm{TBP_diff}(1) \quad (34)$$

To determine calculated ASTM D-86 values from measured TBP values, reverse the procedure.

2.11.2.2 Interconversion of American Society for Testing and Materials D-86 and True Boiling Point Distillation Curve at Atmospheric Pressure with Only Three or Fewer Points[10] This method should be used only for point-to-point conversions when no better range of data is available. Trying to derive a full TBP curve from three or more such point applications of this method can create problems. Distillation curves, by definition, are monotonic functions increasing with temperature. In some cases one sees a fairly flat mid-range curve (little temperature increase between the ranges of, for example, 30–40 and 60–70% – a common occurrence with many hydroprocessed stocks and some resid/bottoms streams. These correlations have yielded, on occasion, a distillation curve with the 50 or 70% points equal to or below the corresponding 30% or 50% values. This, of course, is not acceptable. Use this correlation with some care.

Equations (35) and (36) cover point-to-point interconversion of ASTM D-86 and ATM TBP distillation data:

$$\mathrm{TBP} = a(\mathrm{ASTM\ D\text{-}86})^b \quad (35)$$

$$\mathrm{ASTM\ D\text{-}86} = \left(\frac{\mathrm{TBP}}{a}\right)^{1/b} \quad (36)$$

where a and b are constants varying with the percentage of liquid sample distilled as given in Table 10.

2.11.2.3 Interconversion of American Society for Testing and Materials D-1160 Data and True Boiling Point (10 mmHg) Data[9] D-1160 is a common ASTM procedure for distilling heavy stocks whose final boiling temperatures at ATM pressure are well above the cracking initiation temperatures of 316–343 °C (600–650 °F). The presence of a significant amount of cracked products in the outflow of the unit will skew the curve erroneously. A 10 mmHg pressure in the still will safely allow the analysis of stocks that ordinarily could not be measured with ATM equipment. The methodology here is based on API Technical Data Book Procedure 3A2.1, which is the classical Edmister–Okamoto method. Figure 15 is a graphical representation of this method.

Table 10 Coefficients for alternative D-86 TBP conversion (Equations 35 and 36)[a]

Volume-% distilled	*a*	*b*
0	0.9167	1.0019
10	0.5277	1.0900
30	0.7429	1.0425
50	0.8920	1.0176
70	0.8705	1.0226
90	0.9490	1.0110
95	0.8008	1.0355

[a] TBP: TBP temperatures at 0, 10, 30, 50, 70, 90 and 95 vol.-% distilled are in degrees Rankine. ASTM D-86: observed ASTM D-86 temperatures at corresponding volume percentage distilled, in degrees Rankine.

The $T(50)$, $T(70)$, $T(90)$ and $T(\mathrm{EP})$ of the TBP 10 mmHg and D-1160 curves are assumed to be identical by convention. No interconversion is necessary for them.

Warnings/caveats. Many people make the mistake of assuming that the $T(50)$, $T(70)$, $T(90)$ and $T(\mathrm{EP})$ of the TBP 760 mmHg curve and D-1160 10 mmHg curve are identical by convention. Pressure interconversion

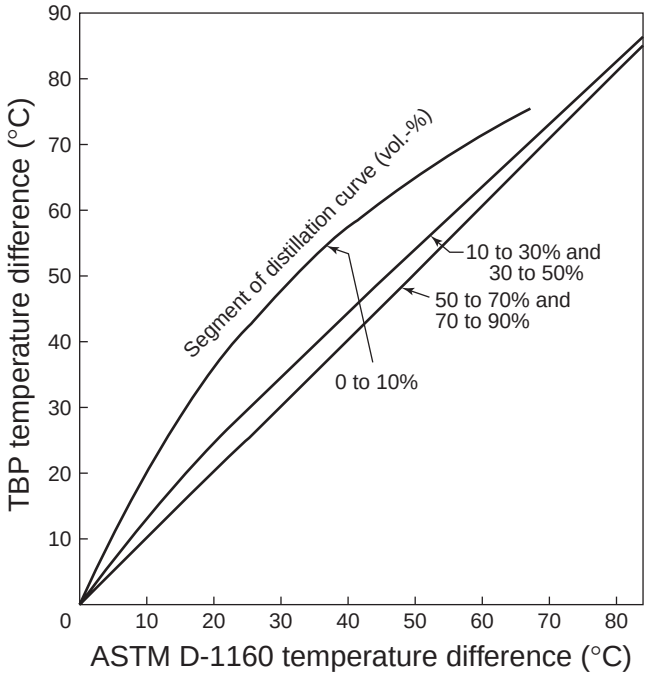

Figure 15 ASTM D-1160–TBP (10 mmHg) data interconversion.

between curve types must be done for corresponding cuts at identical pressures. Pressure corrections can be routinely applied before or after the fact to initial data or derived intermediate values, to yield the final curve. The procedure for this is as follows:

1. Calculate the D-1160 $T(30)$ to $T(50)$ difference (units of °F): Given (Equation 37)

$$\text{TBP_diff}(1) = \text{TBP10}(50) - \text{TBP10}(30) \quad (37)$$

then (Equation 38)

$$\begin{aligned}\text{D-1160_diff}(1) &= \text{D-1160}(50) - \text{D-1160}(30)\\ &= -6.993 \times 10^{-2} + 1.354\text{TBP_diff}(1)\\ &\quad - 2.481 \times 10^{-3}\text{TBP_diff}(1)^2\\ &\quad + 5.894 \times 10^{-6}\text{TBP_diff}(1)^3\\ &\quad - 4.780 \times 10^{-9}\text{TBP_diff}(1)^4 \quad (38)\end{aligned}$$

2. Calculate the D-1160 $T(10)$ to $T(30)$ difference (units of °F): Given (Equation 39)

$$\text{TBP_diff}(2) = \text{TBP10}(30) - \text{TBP10}(10) \quad (39)$$

then (Equation 40)

$$\begin{aligned}\text{D-1160_diff}(2) &= \text{D-1160}(30) - \text{D-1160}(10)\\ &= -6.993 \times 10^{-2} + 1.354\text{TBP_diff}(2)\\ &\quad - 2.481 \times 10^{-3}\text{TBP_diff}(2)^2\\ &\quad + 5.894 \times 10^{-6}\text{TBP_diff}(2)^3\\ &\quad - 4.780 \times 10^{-9}\text{TBP_diff}(2)^4 \quad (40)\end{aligned}$$

3. Calculate the D-1160 $T(\text{IP})$ to $T(10)$ difference (units of °F): Given (Equation 41)

$$\text{TBP_diff}(3) = \text{TBP10}(10) - \text{TBP10}(\text{IP}) \quad (41)$$

then (Equation 42)

$$\begin{aligned}\text{D-1160_diff}(3) &= \text{D-1160}(10) - \text{D-1160}(\text{IP})\\ &= -3.896 \times 10^{-2} + 2.315\text{TBP_diff}(3)\\ &\quad - 1.650 \times 10^{-2}\text{TBP_diff}(3)^2\\ &\quad + 5.977 \times 10^{-5}\text{TBP_diff}(3)^3\\ &\quad - 4.735 \times 10^{-8}\text{TBP_diff}(3)^4 \quad (42)\end{aligned}$$

To determine the D-1160 values at any percentage (besides 50%) distilled, add or subtract the proper differentials from the predicted 50% D-1160 temperature (Equations 43–45) (the units in these equations must be consistent with one another):

$$\begin{aligned}\text{D-1160}_{0\%} &= \text{D-1160}_{50\%} - \text{D-1160_diff}(1)\\ &\quad - \text{D-1160_diff}(2) - \text{D-1160_diff}(3) \quad (43)\end{aligned}$$

$$\begin{aligned}\text{D-1160}_{10\%} &= \text{D-1160}_{50\%} - \text{D-1160_diff}(1)\\ &\quad - \text{D-1160_diff}(2) \quad (44)\end{aligned}$$

$$\text{D-1160}_{30\%} = \text{D-1160}_{50\%} - \text{D-1160_diff}(1) \quad (45)$$

2.11.2.4 Conversion of True Boiling Point Data to/from American Society for Testing and Materials D-1160 Data

1. Calculate the TBP $T(30\%)$ to $T(50\%)$ difference (units of °F): Given (Equation 46)

$$\text{D-1160_diff}(1) = \text{D-1160}(50\%) - \text{D-1160}(30\%) \quad (46)$$

then (Equation 47)

$$\begin{aligned}\text{TBP_diff}(1) &= \text{TBP10}(50\%) - \text{TBP10}(30\%)\\ &= 2.416 \times 10^{-1} + 6.940\text{D-1160_diff}(1)\\ &\quad + 1.813 \times 10^{-3}\text{D-1160_diff}(1)^2\\ &\quad - 3.299 \times 10^{-6}\text{D-1160_diff}(1)^3\\ &\quad + 1.692 \times 10^{-9}\text{D-1160_diff}(1)^4 \quad (47)\end{aligned}$$

2. Calculate the TBP $T(10\%)$ to $T(30\%)$ difference (units of °F): Given (Equation 48)

$$\text{D-1160_diff}(2) = \text{D-1160}(50\%) - \text{D-1160}(30\%) \quad (48)$$

then (Equation 49)

$$\begin{aligned}\text{TBP_diff}(2) &= \text{TBP10}(30\%) - \text{TBP10}(10\%)\\ &= 2.416 \times 10^{-1} + 6.940\text{D-1160_diff}(2)\\ &\quad + 1.813 \times 10^{-3}\text{D-1160_diff}(2)^2\\ &\quad - 3.299 \times 10^{-6}\text{D-1160_diff}(2)^3\\ &\quad + 1.692 \times 10^{-9}\text{D-1160_diff}(2)^4 \quad (49)\end{aligned}$$

3. Calculate the TBP $T(\text{IP})$ to $T(10\%)$ difference (units of °F): Given (Equation 50)

$$\text{D-1160_diff}(3) = \text{D-1160}(10\%) - \text{D-1160}(\text{IP}) \quad (50)$$

then (Equation 51)

$$\begin{aligned}\text{TBP_diff}(3) &= \text{TBP10}(10\%) - \text{TBP10}(\text{IP})\\ &= -7.314 \times 10^{-3} + 0.5368\\ &\quad \times \text{D-1160_diff}(2) - 2.899\\ &\quad \times 10^{-3}\text{D-1160_diff}(2)^2 + 5.452\\ &\quad \times 10^{-5}\text{D-1160_diff}(2)^3 - 9.622\\ &\quad \times 10^{-9}\text{D-1160_diff}(2)^4 \quad (51)\end{aligned}$$

The $T(50\%)$, $T(70\%)$, $T(90\%)$ and $T(\mathrm{EP})$ of the TBP and D-1160 are assumed identical by convention. The units in Equations (52–54) must be consistent with one another.

$$\mathrm{TBP}_{0\%} = \mathrm{TBP}_{50\%} - \mathrm{TBP_diff}(1) - \mathrm{TBP_diff}(2) - \mathrm{TBP_diff}(3) \quad (52)$$

$$\mathrm{TBP}_{10\%} = \mathrm{TBP}_{50\%} - \mathrm{TBP_diff}(1) - \mathrm{TBP_diff}(2) \quad (53)$$

$$\mathrm{TBP}_{30\%} = \mathrm{TBP}_{50\%} - \mathrm{TBP_diff}(1) \quad (54)$$

2.11.2.5 Correcting True Boiling Point to Atmospheric Pressure The following method, based on the Maxwell–Bonnell[9] boiling point conversion, API Procedure 3A4.1/Figure 5A19.1, is probably the most utilized correlation in refinery methods and data analysis. ATM boiling points are the starting point for so many correlations.

This method works well only for actual TBP data at any pressure or for ASTM D-1160 subatmospheric to 10 mmHg data. Some people will try to use it to correct ASTM distillations as well. It is not as accurate this way, although it may give reasonable results. For interconversions of other distillation types, one should use the conversion protocol of Figure 9, which forces the pressure calculation into the TBP realm.

The chart upon which this method is based is provided in equation form below.

Distillation Pressure Interconversion – Numerical Method. For subatmospheric to atmospheric conversion, given a volume-%, a subatmospheric distillation temperature and a K_W (either known or estimated), find the ATM cut point temperature via one of Equations (55–57):

Define

$$\log p^* = \frac{3000T' - 6.762}{43T' - 0.9877} \quad (55)$$

when $T' > 0.0022$ ($p^* < 2$ mmHg).

Define

$$\log p^* = \frac{2663T' - 5.994}{95.76T' - 0.9725} \quad (56)$$

when $0.0013 < T' < 0.0022$ ($2\,\mathrm{mmHg} < p^* < 760\,\mathrm{mmHg}$).

Define

$$\log p^* = \frac{2770T' - 6.413}{36T' - 0.9897} \quad (57)$$

when $T' < 0.0013$ ($p^* = 760$ mmHg).

In these equations, p^* = vapor pressure (mmHg), T' is given by Equation (58):

$$T' = \left(\frac{T_{\mathrm{b;\,atmos}}}{T_{\mathrm{b;\,non\text{-}atmos}}} - 0.0002867(T_{\mathrm{b;\,atmos}}\right) \quad (58)$$

$T_{\mathrm{b;\,atmos}}$ = atmospheric cut-point temperature corrected to $K_W = 12$, in degrees Rankine, and $T_{\mathrm{b;\,non\text{-}atmos}}$ = non-atmospheric cut-point temperature, in degrees Rankine.

The value of $T_{\mathrm{b;\,atmos}}$, and therefore T', must be adjusted until p^* equals the pressure conditions of the subatmospheric distillation being converted. The normalized (to $K_W = 12$) temperature is then converted back to the real K_W basis using Equation (59):

$$DT = T_{\mathrm{b;\,atmos}} - T_{\mathrm{b;\,non\text{-}atmos}} = A(K_W - 12) \times \left(\frac{\log p^*}{760\,\mathrm{mmHg}}\right) \quad (59)$$

For $T_{\mathrm{b;atmos}} > 400\,^\circ\mathrm{F}$, $A = 2.5$.
For $T_{\mathrm{b;atmos}} < 200\,^\circ\mathrm{F}$, $A = 0.0$.
For $200\,^\circ\mathrm{F} < T_{\mathrm{b;atmos}} < 400\,^\circ\mathrm{F}$, $A = 2.5(T_{\mathrm{b;\,atmos}} - 200)/200$

The API has recently proposed a revised correction factor that is similar in form and shown in Equation (60):

$$DT = T_{\mathrm{b;atmos}} - T_{\mathrm{b;non-atmos}} = B(K_W - 15) \times \left(\frac{\log p^*}{760\,\mathrm{mmHg}}\right) \quad (60)$$

For $200\,^\circ\mathrm{F} < T_{\mathrm{b;atmos}} < 600\,^\circ\mathrm{F}$, $B = 3.85$.
For $T_{\mathrm{b;atmos}} < 200\,^\circ\mathrm{F}$, $B = 0.0$.
For $T_{\mathrm{b;atmos}} > 800\,^\circ\mathrm{F}$, $B = 14.9$.
For $600\,^\circ\mathrm{F} < T_{\mathrm{b;atmos}} < 800\,^\circ\mathrm{F}$, $B = 3.85 + 11.02(T_{\mathrm{b;\,atmos}} - 600)/200$.

This latter method using Equation (60) was developed specifically for full-range crude oils. The data from which it was taken were a series of specially assayed common crudes for sale on the world market. For narrower fractions, the Maxwell–Bonnell methods based on Equation (59) are preferable. The user may prefer his/her own plant data to recalibrate the coefficients used.

2.11.2.6 Correcting American Society for Testing and Materials D-86 for Cracking[13,14,75] The following procedure is the classical correction made to ASTM D-86 values in the high-temperature range [>246 °C (475 °F)] to adjust for the cracking of heavier molecules and thus a higher light ends yields than existed in the original sample. The value of this correction procedure has been hotly debated in the last 10 years. An equal number of refining companies can be found on both sides of the controversy – that cracking correction is useless or that it is indispensable. Both sides will claim that proprietary data in their possession prove their respective contentions. For the sake of completeness, it will be included here.

The procedure corresponds to Figure 2.B.1.1 of the API 1964 Technical Data Book[13] (Equation 61). It was

officially removed from the 1987 version 4.[10]

$$\log[\text{D-86_diff}] = -1.587 + 0.00473T_{\text{D-86}} \quad (61)$$

where D-86_diff is the correction that should be made to any D-86 distillation value that exceeds 246 °C (475 °F) and $T_{\text{D-86}}$ is the D-86 value (°F) of concern.

2.11.2.7 Distillation Pressure Interconversion – Gas Chromatography Methods The use of GC equipment is becoming a standard practice for laboratory and on-line analyses of hydrocarbon streams. What must be remembered is that the gas chromatograph will detect the size of the last and presumably largest molecules that elute through the chromatographic column.

The distillation procedures are based on manual observations by an analyst or an automated observation function that mimics an analyst. For this reason, determining the "final boiling point" is a tedious and inaccurate process. It is really difficult to determine the very bottoms of a cut; in fact, it is usually impossible. Absolute purity of the overhead is another issue. GC provides good separation but distillation often does not.

The GC end-point value is artificially skewed to the high end as opposed to what would be observed in a physical distillation. Therefore, the following protocol is proposed (Equation 62):

$$\text{D-2887}_{\text{FBP}} = \text{D-2887}_{98.5\%} \quad (62)$$

Equation (62) is based on the author's experiences with GC analyses and conversions. Using even interpolated values of the 98.5% recovery for the final boiling point is better than using the reported D-2887 end point.

2.11.2.8 Conversion of Simulated Distillation (American Society for Testing and Materials D-2887/D-3710) to American Society for Testing and Materials D-86 Distillation – Method 1 Equation (63) is used to convert what is known as a simulated distillation (SD) in weight-percent to an ASTM D-86 distillation in volume-percent. It is based on API Technical Data Book (1987) procedure 3A3.2.[10] This method provides a one-to-one point correspondence from one distillation type to the other. If one has a full set of distillation data, the next procedure (section 2.12.2.9) may be a preferred method.

$$\text{D-86}(\,) = a(\text{SD})^b F^C \quad (63)$$

where D-86( ) = observed ASTM D-86 temperatures at 0, 10, 30, 50, 70, 90 and 100 vol.-% points, in degrees Rankine, SD( ) = D-2887 temperatures at corresponding weight percentage points, in degrees Rankine. F = a variable parameter given by Equation (64):

$$F = 0.009524(\text{SD}_{10\%})^{0.05437}(\text{SD}_{50\%})^{0.6147} \quad (64)$$

$\text{SD}_{10\%}$ is the simulated distillation temperature at the 10% point, in degrees Rankine, $\text{SD}_{50\%}$ is the simulated distillation temperature at the 50% point, in degrees Rankine, and a, b and c = constants varying with the percentage of liquid sample distilled as given in Table 11.

A standard for SDs, or simdists as they are commonly referred to, is presented in ASTM D-2887 for full-range 0–1000 + °F NBP cuts and in D-3710 which is a procedure for naphtha-based cuts at or below 260 °C (500 °F) NBP. Unfortunately, it is not simple to follow either standard completely to the letter and simdist values may have slightly more variation in their repeatability than is desirable. Use them with care.

2.11.2.9 Conversion of Simulated Distillation (American Society for Testing and Materials D-2887) to American Society for Testing and Materials D-86 Distillation – Method 2 The methodology here is based on API Technical Data Book (1994) Procedure 3A3.1.[9]

This method should be used only for a fully defined assay of a stream. Trying to derive a full TBP curve from three or more such point applications of this method can create problems.

Equations (65–66) are used to interconvert ASTM D-86 and ATM TBP distillation data:

$$\text{D-86}_{50\%} = a(\text{D-2887}_{50\%})^b \quad (65)$$

$$\text{D-86}_{50\%} = \left(\frac{\text{TBP}}{a}\right)^{1/b} \quad (66)$$

where $a = 0.7760$ and $b = 1.040$ for the interconversion of the 50% points, D-2887( ) = vector of simdist boiling point temperatures at 0, 10, 30, 50, 70, 90 and 95 vol.-% distilled, in degrees Fahrenheit, and D-86( ) = vector of observed ASTM D-86 temperatures at corresponding volume percentage distilled, in degrees Fahrenheit.

The procedure works by calculation of differential temperatures from the mid-volume boiling point value and translating that value to the other corresponding member of the pairing.

The difference between adjacent cut points can be determined from a series of equations for each interval

Table 11 GC simulated distillation (SD) (ASTM D-2887) conversion to ASTM D-86 parameters (per Equations 63 and 64)

% Distilled	a	b	c
0	6.0154	0.7445	0.2879
10	4.2262	0.7944	0.2671
30	4.8882	0.7719	0.3450
50	24.1357	0.5425	0.7132
70	1.0835	0.9867	0.0486
90	1.0956	0.9834	0.0354
100	1.9073	0.9007	0.0625

(Equation 67):

$$\text{D-86_diff (I)} = C_0(\text{D-2887_diff})^{C_1} \quad (67)$$

where C_0 and C_1 are coefficients for varying temperature ranges, D-86_diff( ) and D-2887_diff( ). Their vectors are shown in Table 12. Temperature values and differences are in Rankine or Fahrenheit degrees.

2.11.2.10 Conversion of Simulated Distillation (American Society for Testing and Materials D-2887) to American Society for Testing and Materials D-86 Distillation – Method 3[158,159] The methodology here is based on data published by Hewlett-Packard Analyzer Division. It provides a point-to-point translation of data from one assay to the other (Equation 68):

$$\text{D-86}(\ ) = f[\text{D-2887}(\ )] \quad (68)$$

where D-2887( ) = vector of simdist boiling point temperatures at 0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95 and 100 vol.-% distilled, in degrees Fahrenheit, and D-86( ) = vector of observed ASTM D-86 temperatures at the corresponding volume percentage distilled, in degrees Fahrenheit.

To determine estimated ASTM D-2887 values from measured D-86 values, reverse the procedure. In actual practice this would require several iterative calculations, no reverse transform function being readily available. It is, in reality, easier and more reliable simply to take GC data if the information is needed on any sort of regular basis.

The equations for simdist( ) to D-86( ) are Equations (69–77):

$$\text{D-86}_{\text{IP}} = 46.57 + 0.5829\text{D-2887}_{10\%} + 0.3480\text{D-2887}_{\text{IP}} \quad (69)$$

$$\text{D-86}_{10\%} = 33.31 + 0.6156\text{D-2887}_{10\%} + 0.3511\text{D-2887}_{20\%} \quad (70)$$

$$\text{D-86}_{20\%} = 22.41 + 0.4890\text{D-2887}_{30\%} + 0.2753\text{D-2887}_{20\%} + 0.2171\text{D-2887}_{10\%} \quad (71)$$

$$\text{D-86}_{30\%} = 14.43 + 0.4704\text{D-2887}_{30\%} + 0.2837\text{D-2887}_{20\%} + 0.2278\text{D-2887}_{50\%} \quad (72)$$

Table 12 Coefficients for D-86–D-2887 conversion (according to Equations 65–67)

Index value	Cut point range (%)	C_0	C_1
1	100–90	2.603	0.6956
2	90–70	0.3079	1.234
3	70–50	0.1486	1.429
4	50–30	0.0798	1.539
5	30–10	0.06069	1.518
6	10–0	0.3047	1.126

$$\text{D-86}_{50\%} = 4.876 + 0.960\text{D-2887}_{50\%} \quad (73)$$

$$\text{D-86}_{70\%} = 0.911 + 0.5198\text{D-2887}_{80\%} + 0.3326\text{D-2887}_{70\%} + 0.1016\text{D-2887}_{30\%} \quad (74)$$

$$\text{D-86}_{80\%} = 0.279 + 0.7594\text{D-2887}_{80\%} + 0.2833\text{D-2887}_{95\%} - 0.0998\text{D-2887}_{100\%} \quad (75)$$

$$\text{D-86}_{95\%} = -1.973 + 0.6146\text{D-2887}_{90\%} + 0.3191\text{D-2887}_{95\%} \quad (76)$$

$$\text{D-86}_{100\%} = 34.18 + 1.1486\text{D-2887}_{95\%} - 0.5921 \times \text{D-2887}_{90\%} + 0.3154\text{D-2887}_{100\%} \quad (77)$$

To determine calculated ASTM D-2887 values from measured D-86 values, reverse the procedure. The units for the above equations are all degrees Fahrenheit.

2.11.2.11 Conversion of Simulated Distillation (American Society for Testing and Materials D-2887) to True Boiling Point Distillation at 760 mmHg[9] The methodology here is based on API Technical Data Book (1994) Procedure 3A3.3.

Equations (78) and (79) are used to interconvert ASTM D-2887 and ATM TBP distillation data:

$$\text{TBP}_{50\%} = a(\text{D-2887}_{50\%})^b \quad (78)$$

$$\text{D-2887}_{50\%} = \left(\frac{\text{TBP}}{a}\right)^{1/b} \quad (79)$$

where $a = 0.8718$ and $b = 1.026$ for the interconversion of the 50% points, TBP temperatures at 0, 10, 30, 50, 70, 90 and 95 vol.-% distilled, in degrees Rankine, and D-2887 = observed ASTM D-2887 temperatures at corresponding volume percentage distilled, in degrees Rankine.

The procedure works by calculating differential temperatures from the mid-volume boiling point value and translating that value to the other corresponding member of the pairing.

The difference between adjacent cut points can be determined from a series of equations for each interval (Equation 80):

$$\text{TBP_diff (I)} = C_0(\text{D-2887_diff})^{C_1} \quad (80)$$

where C_0 and C_1 are coefficients for varying temperature ranges, TBP_diff( ) and D-2887_diff( ). Their vectors are shown in Table 13. Temperature values and differences are in degrees Rankine or Fahrenheit.

To determine the TBP values at any percentage distilled, add or subtract the proper differentials from the predicted 50% TBP temperature (Equations 81–86):

$$\text{TBP}_{0\%} = \text{TBP}_{50\%} - \text{TBP_diff(4)} - \text{TBP_diff(5)} - \text{TBP_diff(6)} \quad (81)$$

Table 13 Coefficients for D-2887–TBP conversion (according to Equations 78–80)

Index value	Cut point range (%)	C_0	C_1
1	95–90	0.9748	0.8723
2	90–70	0.3153	1.294
3	70–50	0.1986	1.398
4	50–30	0.0543	1.699
5	30–10	0.0119	2.025
6	10–5	0.1578	1.430

$$\text{TBP}_{10\%} = \text{TBP}_{50\%} - \text{TBP_diff}(4) - \text{TBP_diff}(5) \quad (82)$$

$$\text{TBP}_{30\%} = \text{TBP}_{50\%} - \text{TBP_diff}(4) \quad (83)$$

$$\text{TBP}_{70\%} = \text{TBP}_{50\%} + \text{TBP_diff}(3) \quad (84)$$

$$\text{TBP}_{90\%} = \text{TBP}_{50\%} + \text{TBP_diff}(3) + \text{TBP_diff}(2) \quad (85)$$

$$\text{TBP}_{100\%} = \text{TBP}_{50\%} + \text{TBP_diff}(3) - \text{TBP_diff}(2) + \text{TBP_diff}(1) \quad (86)$$

To determine calculated ASTM D-2887 values from measured TBP values, reverse the procedure. The units for the above equations are all degrees Fahrenheit.

2.11.2.12 Conversion of Simulated Distillation (American Society for Testing and Materials D-2887) to American Society for Testing and Materials D-1160 Distillation – Method 3[158] The methodology provides a point-to-point translation of data from one assay to the other (Equation 87):

$$\text{D-1160}(\ ) = f[\text{D-2887}(\ )] \quad (87)$$

where D-2887( ) = vector of simdist boiling point temperatures at 0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95 and 100 vol.-% distilled, in degrees Fahrenheit, and D-1160( ) = vector of observed ASTM D-1160 temperatures at the corresponding volume percentage distilled, in degrees Fahrenheit.

To determine estimated ASTM D-2887 values from measured D-1160 values, reverse the procedure. Unfortunately, this would require several iterative calculations, no reverse transform function being readily available. It is, in reality, easier and more reliable simply to take actual GC data if the information is needed on any sort of regular basis.

The equations for simdist( ) to D-1160( ) are Equations (88–99):

$$\text{D-1160}_{\text{IP}} = 574.7 + 0.3851\text{D-2887}_{10\%} + 0.3660\text{D-2887}_{\text{IP}} - 0.5902\text{D-2887}_{90\%} \quad (88)$$

$$\text{D-1160}_{5\%} = 86.90 - 0.3826\text{D-2887}_{5\%} + 1.931\text{D-2887}_{10\%} - 0.6244\text{D-2887}_{30\%} \quad (89)$$

$$\text{D-1160}_{10\%} = 12.64 + 1.865\text{D-2887}_{10\%} + 0.1127\text{D-2887}_{\text{IP}} - 0.9729\text{D-2887}_{5\%} \quad (90)$$

$$\text{D-1160}_{20\%} = 32.45 + 1.199\text{D-2887}_{20\%} - 0.2398\text{D-2887}_{5\%} \quad (91)$$

$$\text{D-1160}_{30\%} = 10.39 + 1.282\text{D-2887}_{30\%} - 0.3070\text{D-2887}_{10\%} \quad (92)$$

$$\text{D-1160}_{40\%} = 7.249 + 1.117\text{D-2887}_{40\%} - 0.1464\text{D-2887}_{5\%} \quad (93)$$

$$\text{D-1160}_{50\%} = -7.867 + 1.151\text{D-2887}_{50\%} - 0.1702\text{D-2887}_{10\%} \quad (94)$$

$$\text{D-1160}_{60\%} = -20.78 + 2.172\text{D-2887}_{60\%} - 1.187\text{D-2887}_{50\%} \quad (95)$$

$$\text{D-1160}_{70\%} = -28.28 + 0.0398\text{D-2887}_{70\%} + 0.9497\text{D-2887}_{80\%} \quad (96)$$

$$\text{D-1160}_{80\%} = -56.82 + 1.302\text{D-2887}_{80\%} - 0.2840\text{D-2887}_{50\%} \quad (97)$$

$$\text{D-1160}_{90\%} = -54.74 + 1.040\text{D-2887}_{90\%} \quad (98)$$

$$\text{D-1160}_{95\%} = 56.08 + 1.428\text{D-2887}_{95\%} - 0.2514 \times \text{D-2887}_{50\%} + 0.2724\text{D-2887}_{100\%} \quad (99)$$

To determine calculated ASTM D-2887 values from measured D-1160 values, reverse the procedure. The units for the above equations are all degrees Fahrenheit.

3 PREDICTION OF STREAM/STOCK PROPERTIES FROM INSPECTION PROPERTY DATA

3.1 Property Predictions in General[83,89,134,135,143]

The prediction of physical properties is an absolute necessity for any sort of intelligent process and control engineering. It is also a vital tool in the analysis and synthesis of all industrially important organic compounds. As these statements are true of individual volumes of pure components, so too do they apply to mixtures of these components.

Most methods for estimating the properties of hydrocarbon cut fractions were initially derived from relationships that can be ascertained by examining pure components. Boiling point/MW/gravity is an obvious set of positive monotonic correlations where model matches data for nearly all compounds. Refractive index (RI) and

MW for a specific type of compound (aromatics, paraffins, etc.) is another time-tested functional pairing.

Figures 17, 18(a–c), and 19(a) illustrate the variation of various hydrocarbon pure compound physical properties as a function of carbon number and type.

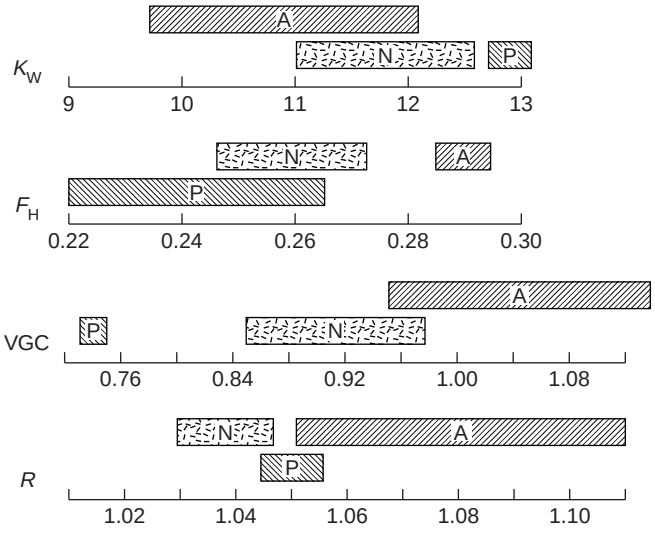

Figure 16 Composition–index relationships: comparison of characterization parameters.

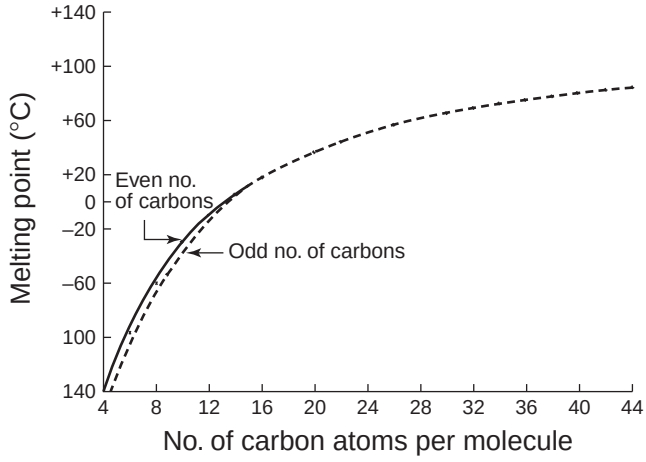

Figure 17 *n*-Alkanes: relationship of melting point and chain length. Similar paraffin analogs exist for other physical properties.

Several of these pure component mathematical models have been modified and extended into the realm of hydrocarbon ''pseudo''-components, a lumping procedure whereby a narrow cut stream is described in terms of one or just a few discrete compounds whose properties are estimated by some of the equations found in this work. MW correlates with gravity and MEABP. Pseudocritical temperature is a strong function of gravity and

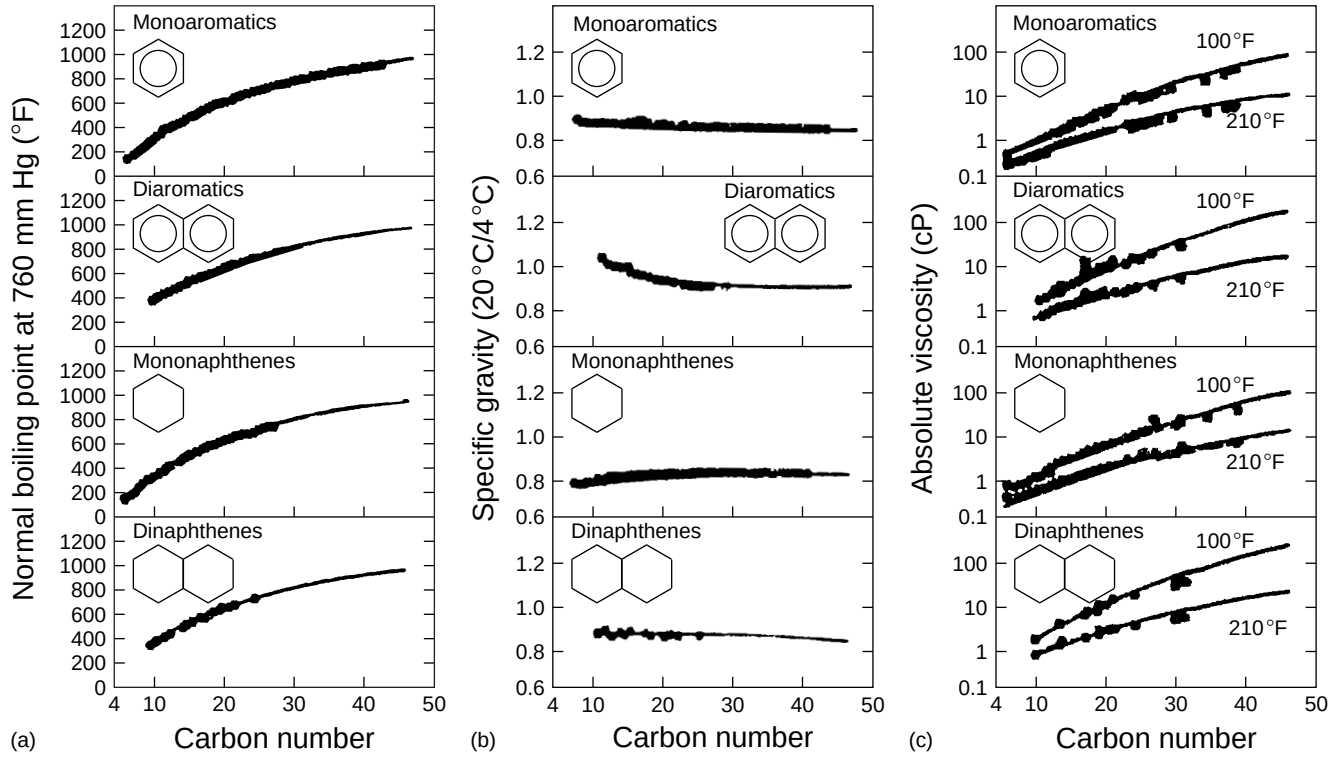

Figure 18 (a) NBP, (b) SG and (c) absolute viscosity versus carbon number for aromatics.

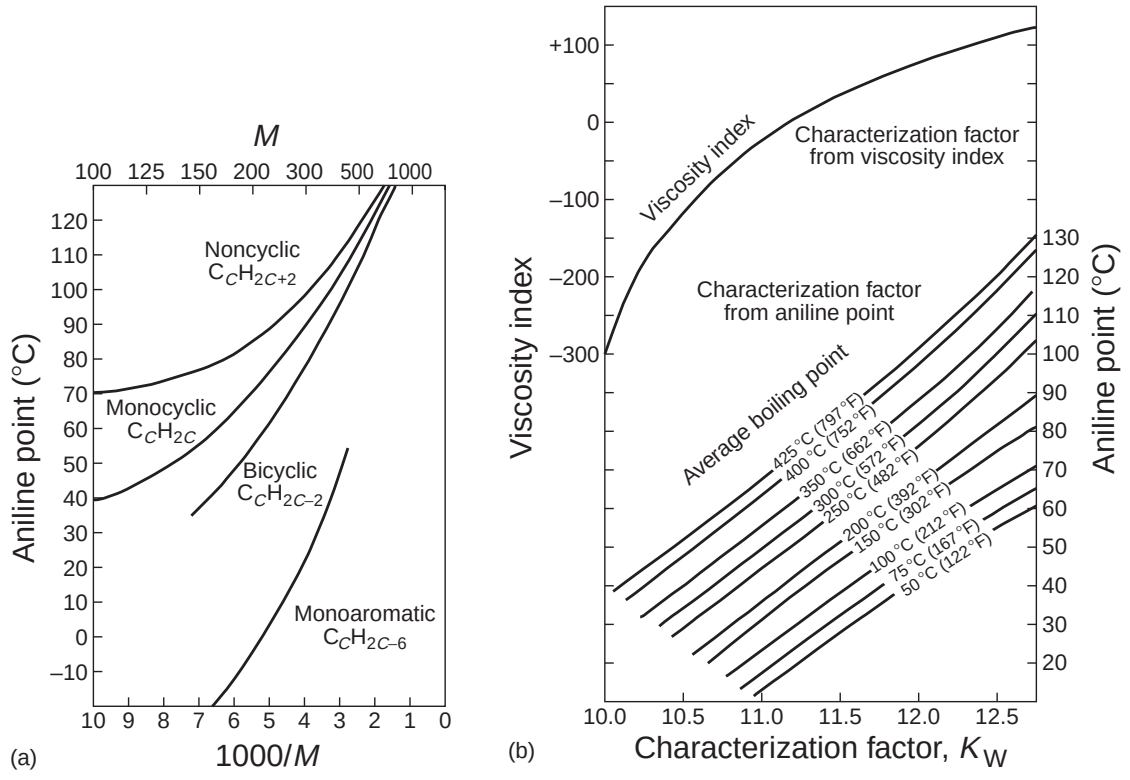

Figure 19 (a) APs of some hydrocarbon families as a function of MW and (b) AP as a function of Watson *K* and average boiling point.

MABP. Any of these properties (MW, gravity, NBP) may be well approximated from a knowledge of any other two of the set.

An historic paper by Watson et al.[152] is a nomograph to predict the AP from K_W, the Watson *K* factor, which Figure 19 summarizes graphically, clearly showing the relationship of AP to MW and molecular type.

In other cases, properties of an entire fraction may be estimated from one or two inspection properties taken for the cut as a whole – without any ''hypothetical'' component lumping being assumed.

None of the methods can be expected to perform perfectly. There is always an error factor involved in the use of any of them. The error derives from two primary sources. The first is in the measurement (bias, average absolute deviations, etc.) of the data upon which the correlation is based. The second is the error of fit for the mathematical derivations from the data. Other error sources include mistranscription from old sources, linear scaling of exponentiation or other transcendental mathematical functions, round off/truncation errors for multiple loop computer processing, change of parameter dimensional units or computational accuracy limitations on significant digits.

Classical statistical analysis tells us that the variance of the correlation is the weighted sum of the variances of the functions and data used to generate the correlation. If all but one of the variances are small, the one that is large will be the determining factor in the overall validity of the fit. Thus, new correlations of any sort are limited by the quality of available data and the amount of work necessary to reduce inaccuracies at all levels of the analysis. Use of established or ''tried and true'' models is often the preferred means of such estimation.

In the following pages, several legacy prediction methods are presented in addition to some more recently examined and published correlations. Use any of them with care. If they yield results that are unexpected, trust one's instincts and review the data and math manipulations for suitability. Two separate correlations that predict a single property give results that may vary significantly from one another, in spite of being derived from the same data. In another laboratory facility down the street, quite the opposite may be true. There is enough variation in data, fit and suitability of these methods that one should always ''calibrate'' real data to the assumed estimation method and use whichever procedure best fits the localized truth the best. For this reason, several similar correlations are presented, based on the gross error of fit for hydrocarbon streams in general. Just because the API Technical Data Committee, for example, fits MW one way does not mean that your data will be so well

served. The reader should possibly try an alternative method presented here or develop one independently if data calibrations fail to show reasonable accuracy and applicability.

3.2 Property Predictions Derived Directly from Observed Data or Other Derived Properties

The following properties can be estimated or predicted directly from simple inspection property measurements. The predicted properties, themselves, can usually (Watson K is an exception) be readily obtained by one or more standard analysis techniques. It may not always be feasible or timely to do so, so these empirical alternatives are presented here. They have been correlated against a significant amount of real historical data.

The correlations presented here either predict derived properties directly from simple inspection property measurements or they use such derived data in hybrid correlations.

3.2.1 Aniline Point

For pure components, several resources[59,63,80,92,96,109] provide tables on the APs of various pure hydrocarbons. This is illustrated by Figure 19(a). Within any hydrocarbon family, the AP increases with MW. Aromatics tend to have the lowest and paraffins the highest values of AP. It is a straightforward exercise to construct a predictive method for the APs for pure compounds, based on MW and compound type.

Several nomographs exist which can be used to estimate APs for petroleum fractions. The Winn nomograph is a traditional part of the API Technical Data Book[10] as Figure 2B6.1. The original Watson nomograph is shown here in Figure 19(b).

The recommended procedure for predicting the CN index of petroleum fractions with lots of information is a recently developed (reregressed) version of the ASTM D-4737-87 procedure. This correlation most accurately predicts the data but requires the liquid density and distillation temperatures at three different specific points.

Other classical AP prediction procedures based on disparate property data are also presented here.

3.2.1.1 Aniline Point from American Petroleum Institute Gravity and Distillation Data – Method 1 The AP can be estimated from known stream gravity and boiling point by this method (Equation 100):

$$\mathrm{AP} = 0.317(\text{D-86}_{50\%}) + 0.40(\text{D-86}_{50\%})(\mathrm{API}) - 298 \tag{100}$$

where $\text{D-86}_{50\%}$ is the cut mid-boiling temperature (degrees Rankine) and API is the API gravity of the cut.

3.2.1.2 Aniline Point from American Petroleum Institute Gravity and Distillation Data – Method 2 An alternative equation to (89) is Equation (101):

$$\begin{aligned}\mathrm{AP}(^\circ\mathrm{F}) = {} & -151.0 + 0.254\mathrm{MEABP} \\ & - 2.87(\mathrm{MEABP})^{0.6667}\mathrm{SG} \\ & + \frac{34.6(\mathrm{MEABP})^{0.3334}}{\mathrm{SG}}\end{aligned} \tag{101}$$

where AP and MEABP are in degrees Rankine.

Equation (101) is more accurate than Equation (100) but it is more complex and requires derived knowledge of the MEABP. It can be estimated as a mid-volume (50%) point or VABP, but the resulting accuracy of the equation will be no better than the previous correlation.

3.2.1.3 Aniline Point from Gravity and Watson K[156] – Method 1 Another derivation of boiling point and gravity is the Watson K characterization factor. It correlates extremely well with AP (Equation 102):

$$\mathrm{AP}(^\circ\mathrm{F}) = \frac{K_\mathrm{W} - 8.3 - 0.0032(\mathrm{API})}{0.0153} \tag{102}$$

where API is the measured or assumed API gravity.

Equation (102) is known as the Woodle method (current API method 2B6.1) and is the API's recommended method, given a lack of complete quality data. One algebraic rearrangement of this correlation is Equation (103):

$$\mathrm{AP}(^\circ\mathrm{F}) = \frac{65.36[\mathrm{MEABP}]^{1/3}}{\mathrm{SG}} - 267.5 - \frac{296.0}{\mathrm{SG}} \tag{103}$$

Another derivative expression of this correlation is Equation (104):

$$\mathrm{AP}(^\circ\mathrm{F}) = -1254 - 0.139\mathrm{MEABP} + 107.8K_\mathrm{W} + 868.7\mathrm{SG} \tag{104}$$

For a narrow range cut, one could conceivably substitute the TBP 50% point or the VABP into the equation. Studies of this idea showed an average error of approximately 8% by doing this. Equation (93) is now API TDB procedure 2B9.1.

The estimated accuracies of AP correlations are given in Table 14.

3.2.1.4 Aniline Point from Gravity and Cetane Index – Method 1 Equation (105) is used:

$$\begin{aligned}\mathrm{AP}(^\circ\mathrm{F}) = {} & 67\,220 + 7.39(\mathrm{API}) - 1.714(\mathrm{CI}) + 2255 \\ & \times \left(\frac{\ln(\mathrm{CI})}{\mathrm{API}}\right)^{0.01} - 70\,600 \left(\frac{\ln((\mathrm{API})^2/\mathrm{CI})}{\mathrm{API}}\right)^{0.01}\end{aligned} \tag{105}$$

where API is the measured or assumed API gravity.

Table 14 Estimated accuracy of AP correlation (AP range 26–250 °F)

Method	Average error (%)	Average bias (%)
Equation (100)	2.87	0.50
Equation (101)	3.88	0.24
Equations (102–104)	3.53	−0.42

Method 2. Equation (106) was formulated as a simple derivation of Equation (105):

$$\mathrm{AP}(^\circ\mathrm{F}) = 60.39 - 0.82(\mathrm{API}) + 2.36(\mathrm{CI}) \quad (106)$$

This procedure, although very simple, is slightly less accurate than using Equation (105). It is reported here, however, as it makes a differential equation relating AP to gravity or CI very simple to evaluate for control or planning optimization systems.

An analysis of the errors for the procedure is given in Table 15.

Both methods tend to predict values higher than experiments. The errors of this method increase greatly when predicted rather than measured CI values are used.

3.2.1.5 Aniline Point from Gravity and Flash Point This method relates the AP to the FP of petroleum fractions. This correlation differs from the previous one because it uses the experimental CI to predict the AP of the fraction from Equation (107):

$$\mathrm{AP}(^\circ\mathrm{F}) = 84 + 0.994(\mathrm{API}) + 0.18\mathrm{FP} - \frac{10\mathrm{API}^2}{\mathrm{FP}^{1.5}} \quad (107)$$

where FP is the flash point (°F) and API is the measured or assumed API gravity.

3.2.1.6 Aniline Point from Smoke Point Equation (108) is used:

$$\mathrm{AP}(^\circ\mathrm{F}) = 12\,140\left\{\frac{\mathrm{SP} + 3500 + 3021\ln(\mathrm{SG})}{3522(\mathrm{SG}) - 1}\right\} \quad (108)$$

where SP is in (mm height).

The validation statistics for these methods are given in Table 16. Equation (97) is now API procedure 2B14.1.

Table 15 Estimated accuracy of AP correlations (AP range 123–184 °F, CI range 31–73 °F)

Method	Average error (%)	Average bias (%)
Equation (105)	2.02	0.08
Equation (106)	3.02	0.35

Table 16 Error analysis for revised prediction methods (FP range 95.0–460.0 °F, AP range 11.0–184.5 °F, SG range 0.76–0.86, SP range 15–42 mm)

	Equation (107)	Equation (108)
Average absolute error (mm)	3.22	1.7
Average error (%)	2.11	7.3
Average bias error (%)	0.18	0.04

3.2.2 Cetane Index[52,53,147]

3.2.2.1 Cetane Index from American Petroleum Institute Gravity and Distillation Data – Method 1 CI can be estimated from known stream gravity and boiling point by one of two methods found in ASTM D-976 (1991–1995 versions) and ASTM D-613 (Equation 109) (temperatures in degrees Fahrenheit).

$$\begin{aligned}\mathrm{CI} = {} & C_7(\text{D-86}_{50\%})^2 + C_6(\text{D-86}_{50\%}) + C_5(\log \mathrm{MBP})^2 \\ & + C_4(\mathrm{API})(\log \mathrm{MBP}) + C_3(\mathrm{API})^2 \\ & + C_2(\log \mathrm{MBP}) + C_1(\mathrm{API}) + C_0 \end{aligned} \quad (109)$$

where log MBP = the logarithm of D-86 mid-boiling point, i.e. (Equations 110 and 111)

$$\mathrm{MBP} = \log(\text{D-86}_{50\%}) \quad (110)$$

$$\mathrm{MBP} = \text{D-976 revised-mean average BP} \quad (111)$$

and C_0–C_7 are coefficients defined in Table 17. The 1995 version is the current API method 2B12.1.

3.2.2.2 Cetane Index from American Petroleum Institute Gravity and Distillation Data – Method 2 A simpler correlation that requires only the liquid density and MEABP is suitable for use. This correlation is a revision of the ASTM D-976-91 procedure. This procedure (the current API method 2B12.1) is represented by Equation (112):

Table 17 CI correlation parameters (Equation 109)

CI parameter	ASTM method		
	D-976 (1991)	(D-976 1995)	D-613
C_0	−420.34	−81.7	419.6
C_1	0.0	0.0	−4.781
C_2	0.0	0.0	−423.5
C_3	0.016	−0.0165	0.01247
C_4	0.192	0.893	2.209
C_5	65.01	8.12	97.83
C_6	0.0	0.0	0.0
C_7	−0.0001809	−0.000032	0.0

$$\mathrm{CI} = 415.26 - 7.673\mathrm{API} + 3.505\mathrm{API}\log(\mathrm{MEABP}) + 0.186\mathrm{MEABP} - 193.8\log(\mathrm{MEABP}) \quad (112)$$

where MEABP is in (°F).

An analysis of the errors for both methods is given in Table 18.

3.2.2.3 Cetane Index from American Petroleum Institute Gravity and Distillation Data – Method 3 A final alternative form of the D-976 method is Equation (113):

$$\mathrm{CI} = 454.7 - 1641S + 774.74S^2 - 0.554(\mathrm{D\text{-}86_{50\%}}) + 97.803(\log \mathrm{D\text{-}86_{50\%}})^2 \quad (113)$$

where S = density (g cm^{-3}) at 15 °C and D-86$_{50\%}$ = distillation temperature (°C) of 50% distillate recovery.

These equations are limited to petroleum liquids with a 90% boiling temperature <343 °C (650 °F). ASTM method D-613 tends to overpredict and is only a slight improvement over ASTM D-976-80. The error of Equation (113) is about the same as that for Equation (112).

3.2.2.4 Cetane Index from Gravity and Assay Data via Expanded Distillation Curve Usage – Alternate Methods CI is also calculated using the methods of ASTM D-4737 (Equation 114):

$$\mathrm{CI} = 45.2 + 0.0892\mathrm{TBP}_{10} + (0.131 + 0.901\mathrm{B})\mathrm{TBP}_{50} + (0.0523 - 0.4208)\mathrm{TBP}_{90} + 0.00049 \times (\mathrm{TBP}_{10}^2 - \mathrm{TBP}_{90}^2) + 107S + 60S^2 \quad (114)$$

where CI = calculated cetane index, SG = specific gravity at 15 °C (60 °F) (g cm^{-3}) and TBP$_{10,50,90\%}$ are true boiling points (°F).

An alternate formulation of Equation (103) using D86 values is Equation (115):

$$\mathrm{CI} = C_0 + C_1A + C_2B + C_3C + C_4BD + C_5CD + C_6A^2 + C_7C^2 + C_8D + C_9D^2 \quad (115)$$

where A = D-86$_{10\%}$ − 488(°F) or D-86$_{10\%}$ − 215(K), B = D-86$_{50\%}$ − 533(°F) or D-86$_{50\%}$ − 260(K), C = D-86$_{90\%}$ − 583(°F) or D-86$_{90\%}$ − 310(K), $D = \exp[-3.5(\mathrm{SG} - 0.85)] - 1$, SG is the specific gravity at 15 °C (g cm^{-3}), D-86$_{10\%}$ = 10% volume boiling temperature (R or K), D-86$_{50\%}$ = mid-boiling temperature (R or K) and D-86$_{90\%}$ = 90% volume boiling temperature (R or K).

Table 18 Estimated accuracy of CI correlation (API range 27–47, MEABP range 360–700 °F)

Method	Average error (%)	Average bias error (%)
Equation (109)	0.6	<0.1
Equation (112)	2.9	<0.2

CI correlation parameters per Equation (115) are given in Table 19 and correspond to API procedure 2B12.2.

The error analysis of CI by Equations (114) and (115) is as follows: CI range 31–73.0, average absolute error 4.31 °F, average error 4.22% and average bias error 0.23%.

3.2.2.5 Cetane Index from Aniline Point Data[110] The method represented by Equation (116) is from Alberta Research Corporation's work. It is fairly good correlation for hydrotreated and hydrocracked heavy naphtha and diesel stocks.

$$\mathrm{CI} = 0.0007685(\mathrm{AP})^2 + 0.471(\mathrm{AP}) + 13.07 \quad (116)$$

where AP is in degrees Fahrenheit.

The reported error analysis of CI by Equation (116) is as follows: AP range 78–166.5 °F, average error 6.43% and average bias error −0.2%.

3.2.2.6 Cetane Index from Aniline Point, Density and Viscosity Data[110]

Canadian General Standards Board Correlation. The Canadian General Standards Board (CGSB) estimates CI via Equation (117):

$$\mathrm{CI} = 77.76 + 0.1765(1.0 + \mathrm{AP}) + 0.003867(\mathrm{AP})^2 - 11.62X + 0.584X^2 - 0.6350V_{40c} \quad (117)$$

where AP is in degrees Celsius, X is given by Equation (118):

$$X = \frac{T_{10} + T_{50} + T_{90} + 820.0}{200D^2} \quad (118)$$

(g cm^{-1}) T_{10}, T_{50}, T_{90} = 10, 50 and 90% D-86 distillation temperatures, (°C), D = density (g cm^{-1}) at 15 °C (kg L^{-1}) and V_{40c} = viscosity at 40 °C (cSt).

Table 19 CI correlation parameters per Equation (115)

CI parameter	ASTM method D-4737-96
C_0	47.7
C_1	−0.0094
C_2	+0.269
C_3	−0.009
C_4	0.317
C_5	−0.2
C_6	0.00004
C_7	−0.000279
C_8	97.9
C_9	1.3

The estimated error analysis of CI by Equation (117) is as follows: AP range 83.8–166.5 °F, average error 7.46% and average bias error −1.7%.

Modified Calculated Carbon Aromaticity Index Method. Equation (119) is used:

$$\mathrm{CI} = -0.3656(\mathrm{CCAI}) + 345.1 \tag{119}$$

where CCAI is given by Equation (120):

$$\mathrm{CCAI} = D - 81 - 141\log[\log(V + 0.85)] \tag{120}$$

D = density ($\mathrm{g\,cm^{-1}}$) at 15 °C ($\mathrm{mg\,L^{-1}}$) and V = kinematic viscosity at 50 °C (cSt). This equation is probably the most promising as a predictor of CI without regard to the source of the diesel stock.

Observed errors of CI by Equation (119) are as follows: AP range 78–166.5 °F, average error 5.61% and average bias error −1.5%.

3.2.3 Carbon Numbers/Index[102,112,117,136,140,146]

A significant portion of any crude stock carbon remains with the heavier residues upon separation, with as little as 0.01 wt-% of the total crude carbon being removed with the gas-oils.

A quick and convenient method was developed for the prediction of the Conradson carbon number (CC) of petroleum fractions. The corresponding data for this correlation were divided into two sections: lower MW gas-oils and higher MW crude oils and residues.

Several equations for carbon number prediction have been reported, as follows:

3.2.3.1 Conradson Carbon Number from American Petroleum Institute Gravity – Method 1 This uses Equation (121):

$$\mathrm{CC\ (wt\text{-}\%)} = -0.2875 + 52\,715(\mathrm{API})^{-3.75} \tag{121}$$

where API = (141.5/SG) − 131.5. This is the better correlation for gas-oils, where CC are in the range 0.01–5.5 wt-% with API gravities between 11 and 40.

3.2.3.2 Conradson Carbon Number from American Petroleum Institute Gravity – Method 2 For full range crudes and crude/vacuum resid fractions, Equation (122) may be a better choice:

$$\mathrm{CC(wt\text{-}\%)} = -33.63 + 85.08(\mathrm{API})^{-0.25} \tag{122}$$

3.2.3.3 Conradson Carbon Number from Watson K and Distillation Data This prediction method (Equation 123), when used for stocks with API < 25, predicts higher CC values for heavy crudes and resids very well. Stocks with lower CC numbers must be used carefully.

$$\mathrm{CC(wt\text{-}\%)} = -507.7 + 2.66\mathrm{API} + 0.024\mathrm{MEABP} + 129.4 \times \ln(\mathrm{MEABP})88.29(\mathrm{API})^{-0.15} - 48.63K_{\mathrm{W}} \tag{123}$$

A summary of the error analysis for the above three methods is given in Table 20.

3.2.3.4 Bureau of Mines Correlation Index from American Petroleum Institute Gravity and Distillation Data This is an alternative method for calculating the CCs of crude oils and residues (Equation 124):

$$\mathrm{BMCI} = \frac{48\,640}{\text{D-86}_{50\%}} + 473.7\mathrm{SG} - 456.8 \tag{124}$$

where D-$86_{50\%}$ is the mid-boiling temperature (kelvin).

3.2.3.5 Interconversion of Ramsbottom Carbon Number (American Society for Testing and Materials D-529) and Conradson Carbon Number (American Society for Testing and Materials D-189) Figure 20 shows the relationship between these two quantities. They are not identical at all. The graph can be represented by Equation (125):

$$\log \mathrm{RC} = a + b[\log(\mathrm{CC})] + c[\log(\mathrm{CC})]^2 \tag{125}$$

where RC is the Ramsbottom carbon (wt-%) and CC is the Conradson carbon (wt-%).

3.2.4 Flash Point

The FP can be estimated from known stream boiling point data by four commonly used and similar methods. On average, the open-cup tests predict a FP 11–17 °C (20–30 °F) higher than that of the closed-cup (Pensky-Martin) estimations. In general, the closed-cup method is used most frequently for light-end fuel oils, whereas the open-cup method is used for the heavier residuum-like fractions.

3.2.4.1 Flash Point from Distillation Data Preferred methods for open-cup FP value predictions are represented by Equations (126–128):

$$\mathrm{FP\ (^{\circ}F)} = 0.64(\text{D-86}_{10\%}) - 100 \tag{126}$$

Table 20 Estimated accuracy of carbon content correlations

Method	Average error (%)	Average bias (%)
Equation (121)	0.17	0.0
Equation (122)	1.19	0.0
Equation (123)	3.02	0.35

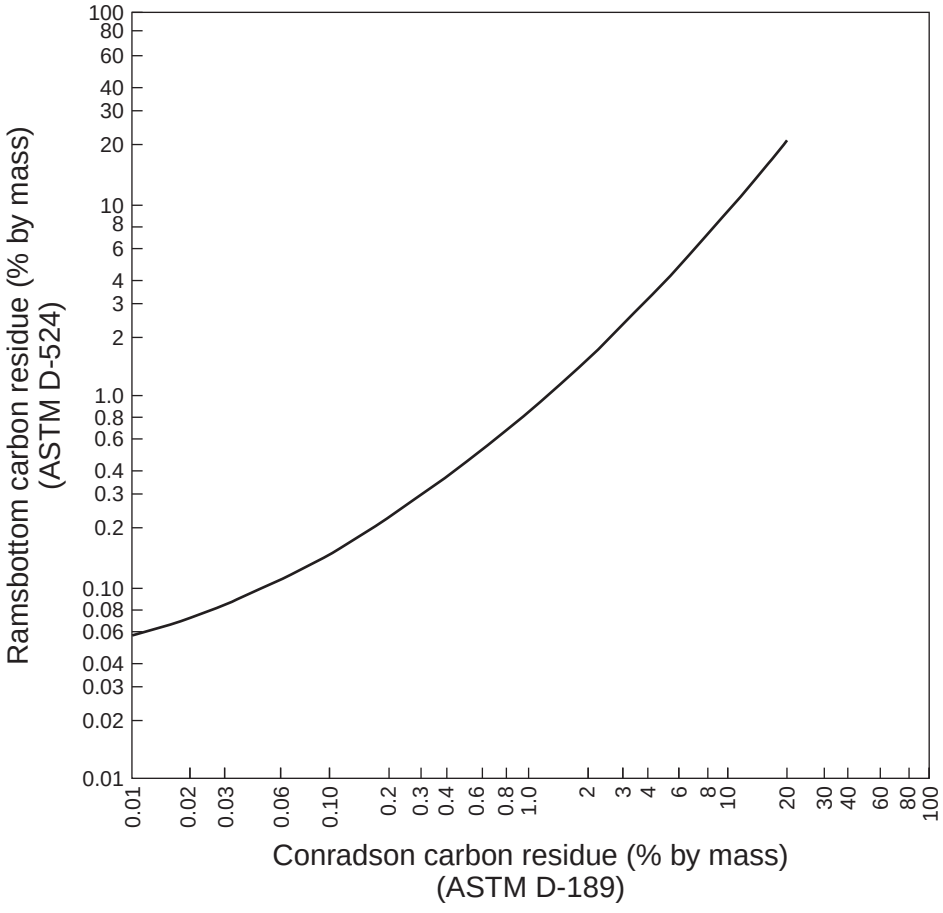

Figure 20 Conversion between Conradson carbon residue and Ramsbottom carbon residue.

where $D\text{-}86_{10\%}$ is the 10% (measured/estimated) value.

$$\text{FP } (^\circ\text{F}) = 0.64(\text{D-}86_{10\%} - \text{D-}86_{\text{IP}}) \tag{127}$$

This equation is a form also used fairly often.

$$\text{FP } (^\circ\text{F}) = 0.68(\text{D-}86_{10\%}) - 109.6 \tag{128}$$

where $D\text{-}86_{10\%} = 10\%$ (measured/estimated) value and $D\text{-}86_{\text{IP}}$ = initial point D-86 value. This equation is the preferred API method for open-cup value prediction and is procedure 2B7.1 in the API Technical Data Book.

Equations (127) and (128) are based on an analysis of the API databank, which showed a possible linear relationship between measured FP and the $D\text{-}86_{(10\%)}$ point. The equation's coefficients were developed in a standard regression fashion. Neither has been adopted as a standard as yet, but they are often useful formulations.

Preferred methods for closed-cup value predictions are represented by Equations (129–131):

$$\text{FP } (^\circ\text{F}) = 0.75(\text{TBP}_{5\%} - 100) \tag{129}$$

where $\text{TBP}_{5\%}$ (°F) is the 5% (measured or estimated) cut TBP value. This method is based on the author's experience at modeling hydrocarbon distillation units. It is presented as an alternative.

$$\text{FP } (^\circ\text{R})^{-1} = \frac{-0.01457 + 2.849}{\text{D86}_{10\%}} + 0.001903 \ln(\text{D-}86_{10\%}) \tag{130}$$

where $D\text{-}86_{10\%}$ is the 10% (measured or estimated) value in degrees Rankine (note: absolute form). Equation (130) is known as the Riazi FP and is a widely used correlation in industry.

This is known as the Nelson equation and is a widely used procedure.

$$\text{FP } (^\circ\text{F}) = 0.69(\text{D-}86_{10\%}) - 118.2 \tag{131}$$

where $D\text{-}86_{10\%}$ is the 10% (measured/estimated) value. This method is based on an analysis of the API databank which showed basically a linear relationship between measured FP and the 10% D-86 point. It is the currently preferred API method for closed-cup predictions.

An analysis of the experimental errors of Equations (125) and (130) is shown in Table 21. The use of other equations may yield adequate results, but these should provide the most accurate estimates, short of fitting one's own proprietary data.

3.2.4.2 Flash Point from Aniline Point and Distillation Data Equation (132) is used:

$$\text{FP } (^\circ\text{R}) = 6989 + 6.0(\text{MEABP}) - 7.67(\text{AP}) - 918.4 \ln\left(\frac{\text{MEABP}}{\text{AP}}\right)^{0.8} - 3683 \times \exp\left(\frac{\text{MEABP}}{\text{AP}}\right)^{0.4} \quad (132)$$

where AP and MEABP are in degrees Rankine.

3.2.4.3 Flash Point from Aniline Point and Smoke Point Equation (133) is used:

$$\text{FP } (^\circ\text{F}) = 116\,814 + 191.0(\text{AP}) + 21.0\exp\left(\frac{\text{AP}}{\text{SP}}\right)^{0.8} - 9419[(\text{AP} + \text{SP})]^{0.5} + 186.0(\text{SP}) \quad (133)$$

where AP is in degrees Fahrenheit and SP is in millimeters.

An analysis of the errors for Equations (132) and (133) is given in Table 22.

Table 21 Analysis of recommended procedures for FP

	Equation (126): Nelson (closed cup)	Equation (130): Riazi (open cup)
FP range (°F)	20–460	95–325
D-$86_{10\%}$ range (°F)	347–855	287–630
Average absolute error (°F)	9.04	6.95
Average bias (°F)	−7.06	−5.38

Table 22 FP, AP and SP error analysis

	Equation (132)	Equation (133)
FP range (°F)	95.0–460.0	105.0–227.3
AP range (°F)	123.0–184.5	123.0–153.0
SP range (mm)	n/a	18.0–27.0
Mean boiling range (°F)	297.0–936.0	n/a
Average absolute error (°F)	8.97	7.62
Average error (%)	4.25	5.24
Average bias error (%)	0.28	0.38

3.2.5 Diesel Index

DI is not an easy measurement to make on a consistent and repeatable basis, yet it is a very useful parameter for the prediction of several other derived property estimates.

3.2.5.1 Diesel Index as a Function of Gravity and Aniline Point DI is defined by a calculation based on the aniline gravity constant (AG), which is the product of AP and API gravity (Equations 134 and 135).

$$\text{AG} = (\text{AP})(\text{API gravity}) \quad (134)$$

$$\text{DI} = \frac{\text{AG}}{100} \quad (135)$$

3.2.5.2 Diesel Index as a Function of Watson K and Mid-boiling Point Equation (136) finds DI as a function of Watson K and D-$86_{50\%}$:

$$\text{DI} = (K_\text{W} - 10.45)[-0.03252(\text{D-}86_{50\%} - 400) + 46.83] \quad (136)$$

where D-$86_{50\%}$ = mid-boiling point − D-$86_{50\%}$ point (°F).

3.2.6 Smoke Point Prediction[84]

3.2.6.1 Smoke Point from Aniline Point and Gravity Data Equation (137) is used:

$$\text{SP} = -255.3 + 2.04(\text{AP}) - 240.8\ln(\text{SG}) + 7727\left(\frac{\text{SG}}{\text{AP}}\right) \quad (137)$$

where SP is in mm height and AP is in degrees Fahrenheit. This is known as the Jenkins equation and is also the API's procedure 2B8.1. The average error of the correlation based on API's databank of experimental measurements is 6.7%, with an average bias of +0.5%. API's use of predicted APs to extend the correlational databank will increase the error and bias of the procedure to an average of 8.8% and 4.0%, respectively.

3.2.6.2 Smoke Point from Aniline Point and Distillation Data Equation (138) is used:

$$\text{SP} = 0.556 - 0.149(\text{MEABP}) + 0.52(\text{AP}) + 0.75\left(\frac{\text{MEABP}}{\text{AP}}\right)^2 + 2084\frac{[\exp(\text{MEABP})]^{0.01}}{(\text{AP})^{2.1}} \quad (138)$$

where SP is in mm height, AP and MEABP are in degrees Fahrenheit.

3.2.6.3 Smoke Point from Flash Point and Distillation Data Equation (139) is used:

$$SP = 1686 - 0.325FP + 5.03(API) - 2252 \times \ln\left(\frac{FP}{API}^{-0.3} - 2.982\frac{(API)^{0.2}}{(FP)^{0.5}}\right) \quad (139)$$

where SP is in mm height and FP is in degrees Fahrenheit.

An analysis of the errors for Equations (138) and (139) is given in Table 23.

3.2.6.4 Smoke Point from Gravity and Distillation Data Equation (140) is used:

$$\ln(SP)(mm) = -1.19 + \frac{0.677(MEABP)^{0.3334}}{SG} - 0.00467(MEABP) \quad (140)$$

where SG is at 60 °F and MEABP is in degrees Fahrenheit.

3.2.6.5 Smoke Point from Watson K and Distillation Data A simplification of Equation (140) gives Equation (141):

$$\ln SP(mm) = -1.028 + 0.474K_W - 0.00168(MEABP) \quad (141)$$

where MEABP is in degrees Rankine.

Equation (140) (now API TDB procedure 2B10.1) turns out to be slightly more accurate than Equation (141), but it is more complex and requires derived knowledge of the MEABP. It can be estimated as a mid-volume (50%) point or VABP, but the resulting accuracy of the equation will be no better than the previous correlation.

3.2.7 Cloud Point, Pour Point and Freeze Point Predictions[89,113,122,155]

These are related measurements and are highly correlated with one another.

3.2.7.1 Cloud Point from Distillation and Gravity Data The API studied a dual-range correlation in this case (Equation 142). The accuracy of either correlation falls off dramatically when it goes out of the temperature ranges shown in Table 24.

Table 23 SP and AP error analysis

	Equation (138)	Equation (139)
Mean boiling range (°F)	297.0–936.0	n/a
SP range (mm)	5.0–42.0	18.5–28.7
AP range (°F)	190.5–644.0	
FP range (°F)	n/a	31.7–187.4
Average absolute error (mm)	216	242
Average error (%)	11.62	9.67
Average bias error (%)	4.10	0.72

$$CP = C_0 + C_1(MEABP)^2 + C_2(MEABP)^{SG} + C_3(SG)^{-1} + C_4(SG)^2 \quad (142)$$

where CP is the cloud point (°F) and MEABP is the mean average boiling point (°F). This is API procedure 2B9.1.

3.2.7.2 Pour Point from Gravity, Molecular Weight and Viscosity Data API Technical Databook 1994 procedure 2B11.1 is represented by Equation (143):

$$PP\ (^\circ F) = 234.95S^A MW^B (V_{100})^C \quad (143)$$

$A = 2.971$, $B = 0.6124 - 0.4376SG$, $C = 0.3103 - 0.8328SG$, SG is the specific gravity, V_{100} = standard 100 °F kinematic viscosity (cSt). See Table 25 for error analysis.

When Equation (132) was run against a database of several hundred waxy crude fractions ($100 < MW < 800$), the average absolute error dropped to ±5.5 °C (10 °F).

3.2.7.3 Freeze Point from Pour Point Data A quick approximation of the PP can be made from either the freeze point (FRP) or the CP of petroleum fractions (Equation 144). The same is true of the other properties, when given one of the three.

$$FRP = 5.252 + 0.9263(PP) \quad (144)$$

where FRP and PP are in degrees Fahrenheit. This is API procedure 2B10.1.

3.2.7.4 Pour Point from Freeze Point Data Equation (145) is just an algebraic manipulation of Equation (144) that works well:

$$PP = 1.079(FRP) - 5.70 \quad (145)$$

where FRP and PP are in degrees Fahrenheit.

Table 24 CI correlation parameters (Equation 142)

CI parameter	Low temp.: $-80\,^\circ F < T < 30\,^\circ F$	Higher temp.: $0\,^\circ F < T < 110\,^\circ F$
C_0	−6583	−14 363
C_1	0.000632	0.000778
C_2	−1.30	−1.832
C_3	3311	−7641
C_4	3862	7676

Table 25 PP error analysis (Equation 143) (temperature range NBP <800 °F, MW range <800)

	For TBP <650 °F	For TBP >650 °F	Combined
Absolute average error(%)	9.6	42.8	28.2
Absolute bias error(%)	−2.6	26.6	15.2

3.2.7.5 Cloud Point from Freeze Point Data The linear Equation (146) gives very good values:

$$\mathrm{CP} = -5.746 + 1.005(\mathrm{FRP}) \tag{146}$$

where FRP and CP are in degrees Fahrenheit.

A second nonlinear equation (Equation 147) will give slightly better results and the extra mathematical complexity is minimal:

$$\mathrm{CP} = 5696 + 3.866(\mathrm{FRP}) - 2329(\mathrm{FRP})^{0.18} \tag{147}$$

where FRP and CP are in degrees Rankine.

An analysis of the errors for Equations (144–147) is given in Table 26. CPs are in the range −80 to 50 °F and FPs are in the range −90 to 50 °F.

3.2.7.6 Freeze Point from Cloud Point Data This is just an algebraic manipulation of Equation (146) above, which works well (Equation 148):

$$\mathrm{FRP} = 0.995(\mathrm{CP}) - 5.715 \tag{148}$$

where FRP and CP are in degrees Fahrenheit.

3.2.7.7 Cloud Point from Pour Point Data This linear function (Equation 149) gives very good values:

$$\mathrm{CP} = 0.9930(\mathrm{PP}) + 4.026 \tag{149}$$

where PP and CP are in degrees Fahrenheit.

This correlation is very accurate at predicting either the CP or PP (see Equation 150), given the other quantity. Correlations on experimental data yielded an absolute average error of approximately 2.25 °C (4.05 °F) and an average bias value of approximately ±0.17 °C (0.3 °F).

3.2.7.8 Pour Point from Cloud Point Data The CP can also be used for the determination of the PP with Equation (150). This equation is just an algebraic manipulation of Equation (149), but is still valid.

$$\mathrm{PP} = 1.007(\mathrm{CP}) - 4.054 \tag{150}$$

where PP and CP are in degrees Fahrenheit.

Table 26 Estimated accuracy of carbon content correlations

Method	Average absolute error (%)	Average absolute bias (%)
Equation (144)	4.00	0.0
Equation (145)	4.39	0.0
Equation (146)	3.97	0.0
Equation (147)	2.91	0.0

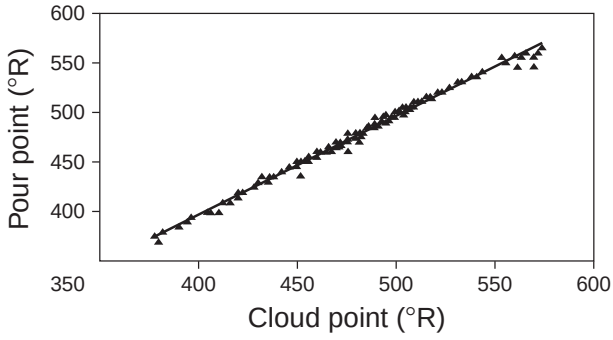

Figure 21 PP vs CP intercorrelation.

Equation (150) is best for CPs between 0 and 110 °F and MEABPs of 275–840 °F.

3.2.7.9 Pour Point from Distillation and Viscosity Data The PP can be determined from a combination of distillation and viscosity data or from distillation alone (Equation 151). This is now API TDB procedure 2B8.1.

$$\mathrm{PP} = 753 + 136[1 - \exp(-0.15V_{100})] - 572(\mathrm{SG}) + 0.0512V_{100} + 0.139(\mathrm{MEABP}) \tag{151}$$

where PP and MEABP are in degrees Rankine, V_{100} is the standard viscosity of the fraction at 100 °F in centistokes. If one lacks viscosity data, Equation (152) can be used:

$$\mathrm{PP} = \frac{3.85 \times 10^{-8}(\mathrm{MEABP})^{5.49}}{10^{z}} + 1.4 \tag{152}$$

where $z = 0.712(\mathrm{MEABP})^{0.315} + 0.133\mathrm{SG}$. The units here are identical with those for Equation (151).

3.2.7.10 Freeze Point from Distillation and Gravity Data The FRP can be determined from a combination of distillation and gravity data from API procedure 2B11.1 (Equation 153):

$$\mathrm{FRP} = -2390 + 1826(\mathrm{SG}) + 122.5\,K_{\mathrm{W}} - 0.135(\mathrm{MEABP}) \tag{153}$$

where FRP and MEABP are in degrees Rankine.

3.2.7.11 Cloud Point from Distillation and Gravity Data The CP can be determined from a combination of distillation and gravity data from API procedure 2B12.1 (Equation 154):

$$\log(\mathrm{CP}) = -7.41 + 5.49\log(\mathrm{MEABP}) - 0.712 \times (\mathrm{MEABP})^{0.315} - 0.133(\mathrm{SG}) \tag{154}$$

where CP is the cloud point in degrees Rankine and MEABP is the mean average boiling point in degrees Rankine.

3.2.8 Refractive Index Use and Prediction[9,133]

The RI of a *pure* compound is easily found in any of several handbooks. It is predictable by an equation such as Equations (155) or (156):

$$\mathrm{RI} = \frac{C_0 + C_1}{\mathrm{MW} + C_2} \tag{155}$$

$$\mathrm{RI} = C_0 + C_1\mathrm{MW} + C_2 \tag{156}$$

where RI is the refractive index (sodium D-line) at 68 °F (20 °C) and C_0, C_1 and C_2 are defined in various handbooks. The values of the coefficients are consistent within compound types such as *n*-alkanes, *n*-alkenes, *n*-alkylcyclopentanes, etc.

3.2.8.1 Refractive Index of Hydrocarbon Mixtures from Gravity and Boiling Point Data API Technical Databook (1994) procedure 2B5.1 is represented by Equation (157):

$$F_\mathrm{H} = C_0 \exp[C_1 T_\mathrm{b} + C_2(\mathrm{SG}) + C_3 T_\mathrm{b}(\mathrm{SG})] T_\mathrm{b}^{C_4} (\mathrm{SG})^{C_5} \tag{157}$$

where C_0–C_5 are defined in Table 27, T_b is the MEABP and F_H is the Huang characterization factor. Then (Equation 158)

$$\mathrm{RI} = \left[\frac{1 + 2F_\mathrm{H}}{1 - F_\mathrm{H}}\right]^{0.5} \tag{158}$$

for the RI at 68 °F (see Table 28 for analysis of RI correlation).

3.2.9 The Characterization Factor[103–106]

By far the most universal characterization parameter used in the prediction of hydrocarbon stocks is the Watson *K* factor (K_W). This characterization factor is an important and special physical property of hydrocarbons defined by Equation (159):

$$K_\mathrm{W} = \frac{[\mathrm{CABP}\ (^\circ\mathrm{R})]^{1/3}}{\mathrm{SG}} \tag{159}$$

Table 27 RI correlation parameters

Coefficient	Light fractions[a]	Heavy fractions[b]
C_0	0.0266	0.02341
C_1	0.0003905	0.0006464
C_2	2.468	5.144
C_3	−0.0005704	−0.0003289
C_4	0.00572	−0.407
C_5	−0.720	−3.333

[a] MW 70–300, boiling point range 90–650 °F.
[b] MW 300–600, boiling point range 650–1000 °F.

where the CABP is in degrees Rankine or Kelvin (see section 2 for discussion of boiling point correlations) and SG is the specific gravity at 60 °F.

The characterization factor is a measure of the distribution of different types of molecules (paraffins, naphthenes and aromatics) in a petroleum fraction. The higher the Watson *K* factor, the higher is the paraffinicity; the lower the Watson *K* factor, the higher is the aromatics content. This is similar to the way in which the AP is often used. The AP varies with boiling point, whereas K_W typically does not. For most refinery stocks, K_W varies from 10.0 to 13.0.

In addition, the Watson *K* factor correlates well with many other physical properties, so that when this is known, it is possible to estimate other properties with good accuracy. The following physical properties correlate well with the characterization factor:

- MW
- pseudocritical temperature
- true critical temperature
- pseudocritical pressure
- viscosity
- aniline number
- viscosity index
- hydrogen content
- heat of combustion
- thermal expansion
- enthalpy corrections

The K_W factors of pure hydrocarbons are given in Table 29. The values for most petroleum oils lie between 11.0 and 12.0. A straight-run (''Virgin'') gas-oil might have a K_W of about 11.7, while a cracked gas-oil of similar boiling range might have a K_W of 11.2. Table 30 lists the Watson *K* ranges of some refinery ''straight-run'' product cuts. Table 31 gives Watson *K* ranges of several refinery ''cracked'' product cuts. The characterization factor of a fraction can be determined to within an accuracy of <0.1 and can be used and interpreted with considerable confidence and reliability.

The most accurate way to determine the characterization factor is from the CABP defined by Equation (8). Also shown are the MW and pseudocritical temperature.

The Watson *K* factor may also be estimated from the AP (Figure 19a) and from viscosity–gravity charts.

Table 28 Analysis of RI correlation

	Average absolute error (%)	Average bias error (%)
MW range (70–600)	0.49	−0.11
Boiling point range (90–1000 °F)	0.27	0.05

Table 29 Characterization factors of selected pure hydrocarbons

Component	K_W
Propane	14.7
Butane	13.5
Pentane	13.0
Hexane	12.8
Heptane	12.7
Octane	12.65
Isobutane	13.85
Isopentane	13.00
2-Methylheptane	12.7
2,2-Dimethylhexane	12.65
3-Ethylhexane	12.5
2,2,4-Trimethylpentane	12.55
Benzene	10.2
Toluene	10.1
m-Xylene	10.4
Ethylbenzene	10.4
Propylbenzene	10.65
Butylbenzene	10.8
α-Methylnaphthalene	9.5
Propylene	14.2
1-Butene	13.05
1-Pentene	12.6
1-Hexene	12.6
2-Hexene	12.5
3-Hexene	12.45
Isobutene	13.1
Cyclobutane	11.35
Cyclopentane	11.2
Cyclohexane	11.05
Methylcyclohexane	11.4
Ethylcyclohexane	11.45
Butylcyclohexane	11.7
Hexylcyclohexane	11.9
Cyclooctane	10.9

Table 30 Watson *K* characterization factors of straight-run fractions from crude oils, listed according to crude base

Crude type	Low-boiling fractions	High-boiling fractions
Paraffinic	12.2+	12.2+
Paraffinic–intermediate	12.2+	11.4–12.0
Paraffinic–naphthenic	12.2+	11.4–
Intermediate–paraffinic	11.5–12.0	12.2+
Intermediate	11.4–12.1	11.4–12.1
Intermediate–naphthenic	11.4–12.1	11.4–
Naphthenic–intermediate	11.5–	11.4–12.1
Naphthenic–paraffin	11.5–	12.2+
Naphthenic	11.4–	14.4–

Other characterization factors of note are the viscosity–gravity constant (VGC), the Huang factor (F_H) and the refractivity intercept (R). Figure 16 illustrates the value range relationship of K_W and these three other factors to composition of the cut.

Table 31 Watson *K* characterization factors of refinery stocks

Source of stock	API gravity	Watson *K*
Raw absorption gasoline	94.6	13.6
Cracked resid from vapor-phase cracking	8.8	10.4
Natural gasoline:		
California	81.9	12.7
Debutanized E.Texas		12.55
Mid-continent	82.2	12.8
Products of thermal cracking:		
Gasolines	58.5–61.2	11.8–12.0
Gasolines		11.5–11.8
Cracking still feeds		10.5–11.5
Recycle stock		10.0–11.0
Light cycle oil	20.0	10.5
Heavy cycle oil	11.0–14.0	10.4–10.5
Gas-oil: paraffin base crude	32.0	11.8
Gas-oil: naphthene base crude	26.0	11.2
Visbreaker resid	8.8	11.1
Cracked resid	2.6–10.3	10.5–10.7
Cracked resid	4.3	9.5
Cracked resid: vapor phase cracking	8.8	10.4
Products of catalytic cracking:		
Hydroformed gasoline	49.0–51.0	11.4–11.5
Cycle oil, once-through		11.3
Hydrocracked light cycle oil: paraffin base		11.4–11.5
Hydrocracked light cycle oil: naphthenic base		10.8–11.0
Hydrocracked heavy cycle oil: paraffin base		11.0
Hydrocracked heavy cycle oil: naphthenic base		10.2–10.8

Several of the equations presented in section 3 are essentially derivatives of these equations. K_W, however, is still the most widely used characterization factor. One major use of these other factors is the speciation of hydrocarbon cuts by compound type.

The refractivity intercept, *R*, is estimated by Equations (160) and (161):

$$R = 1.0482X_p + 1.038X_n + 1.081X_a \quad (160)$$

$$= n - \frac{SG}{2} \quad (161)$$

The viscosity–gravity constant, *VGC*, is estimated by the Equations (162–164):

$$VGC = 0.774X_p + 0.915X_n + 1.040X_a \quad (162)$$

$$= -1.816 + 3.484(SG) - 0.1156 \ln V_{100} \quad (163)$$

$$= -1.948 + 3.535(SG) - 0.1613 \ln V_{210} \quad (164)$$

where X_p, X_n and X_a are the mole fractions of paraffins, naphthenes and aromatics, respectively, n = sodium D-line RI of liquid at 20 °C, V_{100} is the standard viscosity at 100 °F (cSt) and V_{210} is the standard viscosity at 210 °F (cSt).

The Huang factor is defined by rearranging Equation (157):

$$F_H = \frac{n^2 - 1}{n^2 + 2} \tag{165}$$

Coupled with the fact that $X_p + X_n + X_a = 1.00$, Equations (160–165) and (157) can be solved for relative amounts of P, N and A to within an accuracy of approximately 0.05 fraction.

API procedure 2B4.1 presents another estimation procedure. (Equations 166–168):

$$X_P = A_0 + A_1 R + A_2(VGC) \tag{166}$$

$$X_N = B_0 + B_1 R + B_2(VGC) \tag{167}$$

$$X_A = C_0 + C_1 R + C_2(VGC) \tag{168}$$

Coefficients for Equations (166–168) are given in Table 32.

3.2.10 Molecular Weight[9,118,120,125]

MW can be measured directly in the laboratory, and there are several means of doing so. The problem is that they are slow, require methodical attention and are, therefore, expensive to use in everyday operations. Gel permeation chromatography (GPC) and GC/MS techniques are two of the more common of these methods.

There are, fortunately, some correlations that do a good job of MW assessment with a minimum of experimental data. Two such correlations are presented here.

3.2.10.1 Average Molecular Weight as a Function of Gravity and Boiling Point This is expressed by Equation (169):

Table 32 Coefficients for Equations (166–168)

Constant	70 < MW < 200	200 < MW < 600
A_0	−13.36	2.574
A_1	+14.46	1.013
A_2	−1.413	−3.573
B_0	23.98	2.464
B_1	−23.33	−3.670
B_2	0.8152	1.963
C_0	−9.624	−4.088
C_1	+8.874	2.657
C_2	0.5983	1.610

$$\mathrm{MW} = 20.49\left\{\exp[1.165 \times 10^{-4} T_b - 7.787(\mathrm{SG}) + 1.158 \times 10^{-3} - (T_b)(\mathrm{SG})]\right\}(T_b)^{1.260}(\mathrm{SG})^{4.983} \tag{169}$$

where T_b is the MEABP of the cut (°R) and SG is the specific gravity of the cut.

3.2.10.2 Average Molecular Weight as a Function of Gravity and Viscosity This is expressed by Equation (170):

$$\mathrm{MW} = 223.6(V_{100})^{-1.124+1.123(\mathrm{SG})}(V_{210})^{3.476-3.083(\mathrm{SG})} \tag{170}$$

where V_{100} is the cut viscosity at 100 °F (cSt), V_{210} is the cut viscosity at 210 °F (cSt) and SG is the SG of the cut.

3.2.11 Critical Properties[64,82,86,118,119,121,132,151,187]

Traditional mixture estimation of a "pseudocritical" property is merely a linear molar-ratio average blend of all the components. Nobody likes this very much but no-one has come up with a general procedure that is significiantly better at this time.

3.2.11.1 Critical Temperature as a Function of Gravity For a simple hydrocarbon gas we have (Equation 171):

$$T_c = 94.72 + 170.7(\mathrm{SG}) \tag{171}$$

where T_c is the estimated critical temperature (°C) and SG is the gas specific gravity in relation to air at 60 °F, 1 atm.

For a hydrocarbon liquid with Engler (90–10) slope >0.5 °C %$^{-1}$ (Equation 172):

$$\ln T_c = 3.853 + 0.8799 \ln(\mathrm{SG}) + 0.4293 \ln(\mathrm{MABP}) \tag{172}$$

For a hydrocarbon liquid with Engler (90–10) slope <0.5 °C %$^{-1}$ (Equation 173):

$$\ln T_c = 4.052 + 0.4012(\mathrm{SG}) + 0.379 \ln(\mathrm{MABP}) \tag{173}$$

where T_c is the estimated critical temperature (°C), SG is the standard SG of the liquid at 60 °F and MABP is the molar (not mean) average boiling point temperature (°C).

Another method of predicting hydrocarbon liquid critical temperature is the current API procedure 4D1.1 (Equation 174):

$$T_c = 186.2 + 1.667Z - 0.7127 \times 10^{-3} Z^2 \tag{174}$$

where $Z = (\mathrm{SG})(\mathrm{VABP} + 100)$, T_c is the estimated critical temperature (°F), SG is the standard specific gravity of the liquid at 60 °F and VABP is the volumetric average boiling point temperature (°F).

Equation (174) is accurate to within about a 1% average error.

A third valid method of predicting hydrocarbon liquid critical temperature is the current API procedure 4D3.1 (Equation 175):

$$T_c = 10.64\left\{\exp[-5.175^{-3}T_b - 0.5444(SG) + 3.60\times10^{-3} \times (T_b)(SG)]\right\}(T_b)^{0.8107}(SG)^{0.5369} \quad (175)$$

where T_c is the estimated critical temperature (°R), SG is the standard specific gravity of the liquid at 60 °F and T_b is the MEABP temperature (°R) described in section 2. Equation (175) is accurate to within a 0.8% average error.

It is good in the ranges $70 < MW < 295$, $80 < T_b < 650$ °R and $6 < API < 95$.

3.2.11.2 Critical Pressure as a Function of Gravity and Distillation Data For a hydrocarbon gas (Equation 176):

$$P_c = 48.92 - 4.028(SG) \quad (176)$$

where P_c is the estimated critical pressure (10^5 Pa) and SG is the gas specific gravity in relation to air at 60 °F, 1 atm.

For a hydrocarbon liquid with Engler (90–10) slope >0.5 °C %$^{-1}$ (Equation 177):

$$\ln P_c = 8.18 - 1.931\ln(SG) - 0.8727\ln(MABP) \quad (177)$$

and for a hydrocarbon liquid with Englers (90–10) slope <0.5 °C %$^{-1}$ (Equation 178):

$$\ln P_c = 7.742 + 2.802(SG) - 0.7539\ln(MABP) \quad (178)$$

where P_c is the estimated critical pressure (10^5 Pa), SG is the standard specific gravity of the liquid at 60 °F and MABP is the molar (not mean) average boiling point temperature (°C).

A third valid method of predicting hydrocarbon liquid critical pressure is the current API procedure 4D4.1 (Equation 179):

$$P_c = 6.162\times10^6\left\{\exp[-4.725\times10^{-3}T_b - 4.801(SG) + 3.194\times10^{-3}(T_b)(SG)]\right\}(T_b)^{-0.4844}(SG)^{4.085} \quad (179)$$

where P_c is the estimated critical pressure (psia), SG is the standard SG of the liquid at 60 °F and T_b is the MEABP temperature (°R) described in section 2. Equation (179) is accurate to within a 2.6% average error.

It is good in the ranges: $70 < MW < 295$, $80 < T_b < 650$ °R and $6 < API < 95$.

3.2.11.3 Acentric Factor The acentric factor is an important property that it is necessary to have for the calculation of real and pseudocomponent thermodynamic properties. By definition it is given by Equation (180):

$$w = -\log\left(\frac{P^*}{P_c}\right) - 1.00 \quad (180)$$

where P^* is the vapor pressure of the component at $T = 0.7T_c$, temperature is in degrees Rankine and pressure in psia for Equations (177) and (178).

API TDB procedure 2A1.1 allows one to estimate the acentric factor if the critical temperature and pressure are known (Equation 181):

$$w = \frac{(\ln(P^*/P_c) - 5.927 + 6.096/(T_r)) + (1.289\ln(T_r) - 0.1693(T_r)^6)}{(15.25 - 15.69/(T_r)) - (13.47\ln(T_r) - 0.4368(T_r)^6)} \quad (181)$$

where P is the vapor pressure of the component at reduced temperature $T_r = 0.7T_c$. $T_r = T/T_c$.

3.2.12 Thermal Capacities and Transport Properties[9,94,129]

3.2.12.1 Heat of Combustion An important property of all fuels is the heat of combustion, since this is a measure of the energy available from the fuel. ASTM D-240 is the basis used for the direct determination of heats of combustion of aviation gasolines and of fuel oils.

There is an indirect method of calculating or estimating the net heat of combustion of gasolines. The calculation is based on the diesel index (DI) defined according to Equation (182):

$$DI = \frac{(AP)(API\ Gravity)}{100} \quad (182)$$

with (Equation 183)

$$\text{Net Btu lb}^{-1} = 17\,608 + 20.5DI - 7.24\times10^{-2}(DI)^2 \quad (183)$$

where the AP is in degrees Fahrenheit.

3.2.12.2 Liquid Enthalpy (Reduced Temperature and Pressure $T_r < 0.8$, $P_r < 0.98$) API procedure 12B3.1 provides a way of estimating the enthalpy of an unspeciated liquid hydrocarbon mixture (Equation 184):

$$H_L = A_1(T - 259.7) + A_2[T^2 - (259.7)^2] + A_3[T^3 - (259.7)^3] \quad (184)$$

where: H_L is the liquid enthalpy in Btu lb^{-1} T (°F) is the temperature of interest and (Equations 185–187)

$$A_1 = 10^{-3}\left(-1171 + (23.72 + 24.91SG)K_W + \frac{(1150 - 46.54K_W)}{SG}\right) \quad (185)$$

$$A_2 = 10^{-6} \quad (1.0 + 0.8246\,K_W)\left(56.09 - \frac{13.82}{SG}\right. \tag{186}$$

$$A_3 = -10^{-9} \quad (1.0 + 0.8246)\left(9.676 - \frac{2.365}{SG}\right. \tag{187}$$

3.2.12.3 Vapor Enthalpy ($T_r > 0.8$, $P_r > 0.98$) API procedure 12B3.1 provides a way of estimating the enthalpy of an unspeciated hydrocarbon vapor mixture in equilibrium with a saturated liquid (Equation 188):

$$H_L = H_L^* + B_1(T - 0.8T_{pc}) + B_2(T^2 - 0.64T_{pc}^2) + B_3 \times (T^3 - 0.512T_{pc}^3) + RT_{pc}[4.507 + 5.266w] \tag{188}$$

where H_v is the saturated (at dew point) vapor enthalpy in Btu lb^{-1}, H_L is the saturated (at bubble point) liquid enthalpy at temperature T in Btu lb^{-1}, B factors are given by Equations (189–191):

$$B_1 = 10^{-3}\left\{-356.4 + 29.72K_w + B_4 \quad 295.0 - \frac{248.5}{SG} \right\} \tag{189}$$

$$B_2 = 10^{-6}\left\{146.24(77.62 - 2.772K_w) - B_4 \quad 301.4 - \frac{253.9}{SG} \right\} \tag{190}$$

$$B_3 = -10^{-9}\{56.49 - 2.95B_4\} \tag{191}$$

$$B_4 = 10^8\left\{\frac{12.8}{K_w - 1}\left(1.0 - \frac{10.0}{K_w} \quad (SG - 0.885) \times (SG - 0.70)\right\}^2 \right. \tag{191A}$$

For $10.0 < K_w < 12.8$ and $0.70 < SG < 0.885$, else $B_4 = 0$.

T (°F) is the temperature of interest, W is the acentric factor, T_{pc} is the pseudocritical temperature (see section 3.12) in degrees Rankine, $T_r = T/T_{pc}$ is the reduced temperature (dimensionless) and R is the standard gas constant, 1.986 Btu (lb^{-1} mol^{-1}R^{-1}).

3.2.12.4 Isobaric Liquid Heat Capacity of Petroleum Fraction Liquids ($T_r > 0.8$, $P_r > 0.98$) The isobaric (constant pressure) heat capacity of a liquid petroleum fraction is relatively independent of the pressure itself owing to the small compressibility of the liquid. For reduced temperatures below 0.9 (Equation 192):

$$C_p = X_0 + X_1T + X_2T^2 \tag{192}$$

where (Equations 193–195)

$$X_0 = -1.171 + [0.237 + 0.0249(SG)]K_W + \frac{1.150 - 0.04654\,K_W}{SG} \tag{193}$$

$$X_1 = 10^{-4}(1.0 + 0.8246\,K_W)\left(1.122 - \frac{0.2763}{SG}\right. \tag{194}$$

$$X_2 = 10^{-8}(1.0 + 0.8246\,K_W)\left(2.903 - \frac{0.7096}{SG}\right. \tag{195}$$

and T is temperature in degrees Rankine.

3.2.12.5 Isobaric Vapor Heat Capacity of Petroleum Fraction Vapors ($T_r > 0.8$, $P_r > 0.98$) The isobaric (constant-pressure) heat capacity of a vapor-phase petroleum fraction is valid except in the immediate range of criticality (Equation 196):

$$C_p = X_0 + X_1T + X_2T^2 + (\text{pressure correction}) \tag{196}$$

where (Equations 197–200)

$$X_0 = -0.3654 + 0.2972\,K_W + X_3\left(0.2950 - \frac{0.2485}{SG}\right. \tag{197}$$

$$X_1 = -10^{-4} \quad 2.925 - (1.552 - 0.05543\,K_W)K_W + X_4\left(\frac{6.028 - 5.069}{SG}\right. \tag{198}$$

$$X_2 = 10^{-7}[1.695 + 0.0844X_3] \tag{199}$$

$$X_3 = \left(\frac{12.8}{K_W - 1.0} \quad \left(\frac{1.0 - 10}{K_W} \quad (SG - 0.885) \times (SG - 0.70)(10^4)\right.\right.$$
$$\text{for } 10.0 < K_W < 12.8 \text{ and } 0.70 < SG < 0.885$$
$$= 0 \text{ for all other values of } K_W \text{ and SG} \tag{200}$$

and T = temperature in degrees Rankine.

The pressure correction term is a somewhat complex calculation which is beyond the scope of this article. It is zero below about 50 psi and is generally insignificant compared with the error of the correlation below the 150–300 psia range. For higher vapor pressures, any number of good thermodynamic texts will cover the proper use of this factor.

3.2.12.6 Thermal Conductivity – Liquid

Method 1. Liquid Thermal Conductivity as a Function of Molecular Weight. API procedure 12B3.1 provides a way of estimating the enthalpy of an unspeciated hydrocarbon vapor mixture in equilibrium with a saturated liquid (Equation 201):

$$k = \left(C_0 + \frac{C_1}{MW} + \frac{C_2}{MW^2} \quad + T\left(\frac{C_3 + C_4}{MW} + \frac{C_5}{MW^2}\right.\right. \tag{201}$$

where k is the thermal conductivity of the liquid at temperature T, in units of Btu (lb h °F ft^{-1})$^{-1}$ for

temperature in degrees Fahrenheit, $C_0 = 0.01335$, $C_1 = 0.2463$, $C_2 = 1.149$, $C_3 = 3.277 \times 10^{-5}$, $C_4 = 4.188 \times 10^{-5}$ and $C_5 = 0.001843$.

Method 2. Liquid Thermal Conductivity as a Function of Boiling Point. API procedure 12A3.1 provides a way of estimating the enthalpy of an unspeciated hydrocarbon vapor mixture in equilibrium with a saturated liquid (Equation 202):

$$k = (\text{MEABP})^{0.2904}(9.961 \times 10^{-3} - 5.364 \times 10^{-6}T) \quad (202)$$

where k is defined as above and MEABP is the mean average boiling point of the liquid cut.

3.2.12.7 Thermal Conductivity – Vapor API procedure 12A3.2 provides a way of estimating the enthalpy of an unspeciated hydrocarbon vapor mixture in equilibrium with a saturated liquid (Equation 203):

$$k = 0.07577 - 4.1 \times 10^{-5}T \quad (203)$$

where k is defined as above.

3.2.12.8 Viscosity – Liquid Although often a primary inspection property, the viscosity of petroleum cuts is often estimated or is correlated with other variables to quantize derivatives for determining the interrelational effects of change in a property.

Method 1. Liquid Viscosity as a Function of Temperature. API procedure 11A4.1 provides a way of estimating the liquid viscosity of an unspeciated hydrocarbon as a function of any temperature, T (°F) (Equation 204):

$$\text{Log (viscosity)} = B\left(\frac{539.7}{T + 459.7}\right)^2 - 0.8696 \quad (204)$$

where B and S are given by Equations (205) and (206):

$$B = \log V_{100} + 0.8696 \quad (205)$$

$$S = 0.2801 \log V_{100} + 1.862 \quad (206)$$

and V_{100} is the standard liquid viscosity at 100 °F.

Method 2. Liquid Viscosity as a Function of Boiling Temperature. API procedure 11A4.2 is another way of estimating the liquid viscosity of an unspeciated hydrocarbon. One needs slightly more information on the distillation data to use this procedure.

For standard viscosity at 100 °F (Equations 207–209):

$$V_{100} = V_{\text{ref}} + V_{\text{cor}} \quad (207)$$

$$\log V_{\text{ref}} = -1.356 + 8.161 \times 10^{-4}T_{\text{b}} + 8.385 \times 10^{-7}T_{\text{b}}^2 \quad (208)$$

$$\log V_{\text{cor}} = A_1 + A_2K_{\text{W}} \quad (209)$$

where (Equations 210 and 211):

$$A_1 = 34.93 - 0.08844T_{\text{b}} + 6.735 \times 10^{-5}T_{\text{b}}^2 - 1.014 \times 10^{-8}T_{\text{b}}^3 \quad (210)$$

$$A_2 = -2.926 + 0.006984T_{\text{b}} - 5.100 \times 10^{-6}T_{\text{b}}^2 + 7.494 \times 10^{-10}T_{\text{b}}^3 \quad (211)$$

T_{b} is the MEABP of the liquid cut, in degrees Rankine, and V_{100} is the standard liquid viscosity at 100 °F (cSt).

For standard viscosity at 210 °F (Equation 212):

$$\log V_{210} = -1.924 + 2.411 \times 10^{-4}T_{\text{b}} + 0.5113 \log(T_{\text{b}}V_{100}) \quad (212)$$

where T_{b} is the MEABP of the liquid cut, in degrees Rankine, V_{100} is the standard liquid viscosity at 100 °F (cSt) and V_{210} is the standard liquid viscosity at 210 °F (cSt).

3.2.12.9 Density – Liquid Although usually a quick and easy primary inspection property, densities of petroleum cuts are often estimated or correlated with other variables to evaluate derivatives for determining the interrelational effects of change in a property.

Method 1. Liquid Density as a Function of Molecular Weight and Refractive Index. API procedure 2B4.1 provides a way of estimating the liquid density of an unspeciated hydrocarbon as a function of MW and RI (Huang index) (Equation 213).

$$D = 2.831\,\text{MW}^{0.3975}\,F_{\text{H}}^{1.1354} \quad (213)$$

where F_{H} is given by Equation (214):

$$F_{\text{H}} = \frac{n^2 - 1}{n^2 + 2} \quad (214)$$

n is the RI at 68 °F.

Method 2. Liquid Density as a Function of Standard Viscosity. API procedure provides a way of estimating the liquid density of an unspeciated hydrocarbon as a function of standard viscosity measurements (Equation 215):

$$\text{SG} = 0.7717\,V_{100}^{0.3975}\,V_{210}^{1.1354} \quad (215)$$

where SG is the liquid specific gravity, V_{100} is the standard liquid viscosity at 100 °F (cSt) and V_{210} is the standard liquid viscosity at 210 °F (cSt).

4 PROPERTY PREDICTION FOR MIXTURES OF KNOWN INDIVIDUAL VALUES

4.1 Introduction

The characterization of a typical petroleum stock is normally based on easily and routinely obtained inspection

data. How does one handle the case where two stocks, for which a specific inspection property is known for each, are blended? One could estimate average values of gravity and boiling point from some standard mixing rules. For example, gravities blend well linearly with weight percentage or linearly with the inverse of the volume percentage. TBPs blend less well volumetrically but usually give usable results when treated so. With these two quantities, most mixture properties can be estimated. The problem with this technique is that another layer of estimation error has been placed upon the results. Attempts to use simple (i.e. linear) mixing rules to correlate blended values of a hydrocarbon property with the individually estimated or measured blend stock values of the same property have shown marginal results at best.

This difficulty has been overcome by the use of a blend index function and blending numbers, which will be described in this section.

4.2 Blending Index and Numbers

A blend index is a mathematical transformation of a numerical value which can then be blended with other transformed stock values (of the same item) in a reasonably simple algebraic form. The value of the property after transformation is known as the blend number.[50]

Once the blend number of the mixed stream has been calculated, the actual property value of the blended stream can be determined by performing the inverse of the index transform.

Among the values that are best determined for mixtures in this fashion are viscosity, CP, PP, RVP, FP and AP. Suggested blend index transforms (f^{T}), mixing and retransform ($f^{-\mathrm{T}}$) functions for some of these properties are typically of the form in Equations (216–218):

$$\text{index}_I = f^{\mathrm{T}}(\text{property_value})_I \tag{216}$$

$$\text{property_number}_{\text{mix}} = \text{SUM}[(\text{index}_I) \times (\text{vol_pctg_of_stock})_I] \tag{217}$$

$$\text{property_value}_{\text{mix}} = f^{-\mathrm{T}}(\text{property_number}_{\text{mix}}) \tag{218}$$

where I is the index of blend stocks from 1 to N = number of stocks. $f^{\mathrm{T}}(\ )$ and $f^{-\mathrm{T}}(\ )$ are the chosen property value transform and inverse transform functions, and vary greatly in form from one property to another. The forms presented here are those which have been published and are generally used by the petroleum refining community based on commonly available information. The user may want to use custom/proprietary data to modify the blend index function and its numerical parameters, but the forms shown here are generally the best that the author has found for this purpose.

Table 6 shows a series of gasoline blend components and their corresponding properties. The great majority of blend prediction problems in today's refineries involve gasoline stocks. The important values under EPA [Environmental Protection Agency (US)] Phase II/CARB spec fuels are RVP and distillation values, especially D-$86_{90\%}$. The other blend methods described in this section, such as CP and PP, are more applicable to jet fuels and heavier stocks.

4.2.1 Viscosity Blend Number

The viscosity blend number (VBN) is best used with volume averaging, i.e. (Equation 219):

$$\text{VBN}_{\text{average}} = \text{SUM}[(\text{VBN}_I)(\text{volume_pctg_of_stock})_I] \tag{219}$$

where I is the index of blend stocks from 1 to N = number of stocks.

4.2.1.1 Index for Viscosity Blending Number at Set Measurement Temperatures[50] VBN is defined according to Equation (220):

$$\begin{aligned}\text{VBN(at } T_{\text{ref}}) = -[&(-0.04375\text{API} + 0.333\,K_{\text{W}} + 1.467) \\ &\times \log(T_{\text{ref}}) - \log(T_{\text{vis}})] \\ &+ \log[\log(\text{viscosity}_{(\text{CS})} + 0.7)]\end{aligned} \tag{220}$$

where API is the API gravity and $\text{viscosity}_{(\text{CS})}$ and T_{vis} are any laboratory-measured pair of viscosity (centistokes) and temperature (degrees Rankine). Typical viscosity measurement standards are taken at 100 °F or 40 °C, 150 °F or 66 °C and 212 °F or 100 °C. *(Do not forget to convert the temperatures to absolute form before calculating the blend number.)*

The viscosity update/transform function to Equation (220) is Equation (221):

$$\text{viscosity}_{\text{mix}} = [\text{VBN(at } T_{\text{ref}}) + 0.54] \times \frac{K_{\text{W}}}{T_{\text{ref}} - 355} \tag{221}$$

Separate calculations are required for each standard reference temperature. The values for 100, 140 and 212 °F will, for example, have different blend numbers.

4.2.1.2 General Blend Indices at Various Temperatures A general VBN that can be used at any standard T_{vis} is given by Equation (222):

$$\text{VBN} = \frac{\log[\log(\text{viscosity}_{\text{cSt}} + 0.6)] - 0.834}{0.00265} + 154 \tag{222}$$

Table 33 VBNs of a typical gas–oil stream

Cut point mid (vol.-%)	VBN (via Equation 206) at mid (vol.-%)		
	40 °C	66 °C	100 °C
7.93	141.9	30.8	Out of range
30.39	281.5	173.8	18.21
50.9	383.6	282.1	135.3
70.36	465.4	383.7	234.7
76.2	494.6	393.4	281
81.35	Out of range	437.0	319

where viscosity_{cSt} is the measured centistokes at a standard reference temperature (40, 66 or 100 °C).

In the absence of a large amount of data, this method is reasonably good for relating any viscosity measured at any temperature. The VBN method presented here provides only for the prediction of viscosities at the standard temperatures. To determine viscosity at intermediate temperature, one needs to interpolate (extrapolate at your own risk!) from the curve generated at three standard points.

The cut point viscosity data in Table 33 were obtained from a West Texas light vacuum gas-oil (LVGO) cut. Inaccuracies at the extremes of data are common and should be used with caution.

4.2.1.3 Log-linear Blend Function If no explicit blend data are available, a reasonable guess for the viscosity of a blend of streams, all taken at or referenced to the same temperature, can be estimated by blending the logarithm of the viscosities linearly with equivalent volume fractions of the cuts and retransfoming the average by taking the antilogarithm of the value calculated.

4.2.2 Aniline Point Blends[50]

The aniline point (AP) index blends linearly with the relative volume of each stock. The index, mixing and retransformation functions are given by Equations (223–225):

$$\text{AP_index}_I = \exp(0.003652 \times \text{aniline_point}_I) \quad (223)$$

$$\text{AP_index}_{\text{mix}} = \text{SUM}[(\text{AP_index}_I) \times (\text{volume_pctg_of_stock})_I] \quad (224)$$

where I = blend stock index for 1 to N = number of stocks and

$$\text{aniline_point}_{\text{mix}} = \frac{\log(\text{AP_index}_{\text{mix}})}{0.003652} \quad (225)$$

4.2.3 Cloud Point Blends[81]

The cloud point (CP) index blends linearly with the relative volume of each stock. The index, mixing and retransformation functions are given by Equations (226–228):

$$\text{CP_index}_I = \left(\frac{\text{cloud_point}_I}{600}\right)^{20} \quad (226)$$

$$\text{CP_number}_{\text{mix}} = \text{SUM}[(\text{CP_index}_I) \times (\text{volume_pctg_of_stock})_I] \quad (227)$$

where I is the index of blend stocks from 1 to N = number of stocks and

$$\text{cloud_point}_{\text{mix}} = (600 \times \text{CP_number}_{\text{mix}})^{0.05} \quad (228)$$

where cloud_point is measured in absolute form (degrees Rankine for the equations above).

4.2.4 Pour Point Blends[24,80]

4.2.4.1 Method 1 PP index blends linearly with the relative volume of each stock. The index, mixing and retransformation functions are given by Equations (229–231):

$$\text{PP_index}_I = \left(\frac{\text{pour_point}_I}{600}\right)^{12.5} \quad (229)$$

$$\text{PP_index}_{\text{mix}} = \text{SUM}[(\text{PP_index}_I) \times (\text{volume_pctg_of_stock})_I] \quad (230)$$

where I is the index of blend stocks from 1 to N = number of stocks and

$$\text{pour_point}_{\text{mix}} = (600 \times \text{PP_number}_{\text{mix}})^{0.08} \quad (231)$$

where pour_point is measured in degrees Rankine.

4.2.4.2 Method 2 An alternative function that the author has used successfully is given by Equations (232) and (233):

$$\text{PP_index}_I = \left(\frac{\text{pour_point}_I}{280.6}\right)^{9.0909} \quad (232)$$

$$\text{pour_point}_{\text{mix}} = (280.6 \times \text{PP_number}_{\text{mix}})^{0.11} \quad (233)$$

where pour_point is measured in degrees Rankine.

4.2.4.3 Method 3: American Society for Testing and Materials D-97 ASTM D-97 specifies a blend transform function of the following form in Equation (234):

$$(\text{pour_point}_{\text{mix}})^{C_0} = \text{SUM}(V_I^{C_1} T_I^{C_2}) \quad (234)$$

where I is the index of blend stocks from 1 to N = number of stocks and C_0, C_1 and C_2 are coefficients defined in Table 34.

Table 34 Coefficients for ASTM D-97 mixing rule

PP range (°C)		C_0	C_1	C_2
Low	High			
9	51	11.76	0.99	11.76
−12	21	16.67	1.145	16.67
−21	−6	10.87	1.105	10.87

4.2.5 Reid Vapor Pressure Blends[9]

RVP index blends linearly with the relative volume of each stock. The index, mixing and retransformation functions are given by Equations (235–237):

$$\text{RVP_index}_I = (\text{RVP}_I)^{1.25} \tag{235}$$

$$\text{RVP_number}_{\text{mix}} = \text{SUM}[(\text{RVP_index}_I) \times (\text{volume_pctg_of_stock})_I] \tag{236}$$

where I is the index of blend stocks from 1 to N = number of stocks and

$$\text{RVP}_{\text{mix}} = (\text{RVP_number}_{\text{mix}})^{0.8} \tag{237}$$

Of all blended properties, the RVP correlation has the most variation in accepted parameters. Literature values for the index exponent range from 1.07 to 1.32. The user may want to use a customized index function, but a 1.25 index is generally a safe number to start with.

The individual blend values must all correspond to the same blend temperature, 37 °C (100 °F), if following normal standards.

4.2.6 Distillation Temperature Blends

Several methods of blending TBPs can be derived from the procedures in section 2.

ASTM distillation values cannot be accurately blended in any kind of linear fashion if the range of initial to final boiling point of the cut is greater than about 200 °F. This eliminates the heavier cuts altogether and is not a recommended procedure for the lighter cuts either.

A better procedure is to generate the TBP curve for each stock and blend the values based on either volume or weight percentage recovery, depending upon which was used as the independent data acquisition variable. Alternatively, one may take relative volumes of all blend streams at an equivalent boiling point cut and blend these together on a similar pro rata basis. The accuracy of either method is about the same.

The very best TBP blend method is to blend the streams linearly based on the temperature–molar fraction percentage function, but this is a rather complex calculation, and the extra accuracy is often not worth the extra effort. A compromise method comes from the API Technical Data Book,[9] which suggests that TBP values be turned into a molar-based "mean-average" boiling point temperature for a cut fraction. These values are assumed to be linearly blendable at the standard volume fraction recoveries at 10, 30, 50, 70% and 90%.

For wide fraction cuts in general, the API method is recommended. For narrow cuts, simple TBP blending is usually justified.

4.2.7 Flash Point Index Blends[50]

FP index blends linearly with the relative volume of each stock. The two most useful index, mixing, and retransformation functions are as follows.

4.2.7.1 Method 1. Hu–Burns Protocol[81] This is the most commonly used method in industry today (Equations 238–240):

$$\text{FP_index}_I = \left(\frac{735.8}{\text{flash_point}_I}\right)^{19.61} \tag{238}$$

$$\text{FP_number}_{\text{mix}} = \text{SUM}[(\text{FP_index}_I) \times (\text{volume_pctg_of_stock})_I] \tag{239}$$

where I is the index of blend stocks from 1 to N = number of stocks and

$$\text{flash_point}_{\text{mix}} = \left(\frac{\text{FP_number}_{\text{mix}}}{735.8}\right)^{0.051} \tag{240}$$

where flash_point is measured in degrees Rankine.

4.2.7.2 Method 2. Whickey–Chittenden Protocol
This is the current API Technical Data Book procedure 2B15.1 (Equations 241–243):

$$\log(\text{FP_index}_I) = -6.119 + \frac{4345}{\text{flash_point}_I + 383.0} \tag{241}$$

$$\text{FP_number}_{\text{mix}} = \text{SUM}[(\text{FP_index}_I) \times (\text{volume_pctg_of_stock})_I] \tag{242}$$

where I is the index of blend stocks from 1 to N = number of stocks and

$$\text{flash_point}_{\text{mix}} = \frac{4345}{\log(\text{FP_number}_{\text{mix}} + 6.119)} + 383.0 \tag{243}$$

where flash_point is measured in degrees Rankine.

4.3 Other Blend Prediction Methods

4.3.1 Dissolved Gas Effects on Viscosity

One of the most important property estimates that are made incorrectly are liquid viscosity calculations

when dissolved gases are present, even if in known amounts. API Technical Databook procedure 11A7.1 presents a reasonable approach to handling this problem (Equations 244 and 245):

$$\frac{V_m}{V_a} = \frac{(0.294\text{GLR} + 137V_a^{1/3} + 538.4)^3}{[V_a^{1/3}(0.871\text{GLR} + 137 + 538.4)]^3} \quad (244)$$

$$V_t = -1.209 + \frac{1.209 + \log(137V_m)}{T + 139} \quad (245)$$

where V_m is the viscosity (centipoise) of the gas-saturated liquid at 100 °F (37.8 °C), V_a is the viscosity (centipoise) of the gas-free liquid at 100 °F (37.8 °C), V_t is the viscosity (centipoise) of the gas-saturated liquid at any Fahrenheit temperature, GLR is the gas–liquid ratio (std. $ft^3\ Bbl^{-1}$) or liquid and T is the temperature at which the viscosity value is desired (degrees Fahrenheit).

4.3.2 Derived Property Blend Rules

Several blend values of streams with calculated or derived properties can be quickly estimated as blends fractionally prorated to volume, weight or molar bases. Among these are the following:

- Watson K (K_W) – blend is linearly prorated on component weight fraction;
- acentric factor – blend is linearly prorated based on component weight fraction;
- (pseudo)critical temperature – blend is linearly prorated based on component molar fraction;
- (pseudo)critical pressure – blend is linearly prorated based on component molar fraction.

4.3.3 Miscellaneous Property Index Blend Rules of Thumb

It is obvious that prediction of mixture properties is as much art as science. There is no ''right'' way to generate a blending method. There are a few generalizations, however, that can be listed.

- Critically examine any experimental data that are not monotonically inclined in the direction that one would expect. For example, densities and viscosities of liquids tend to decrease with increasing temperature. It should not have any anomalies over the range of data or one could imply that there is a constraint on its usage (above critical point, bad data, etc.).
- Use a transform method that has a mathematical inverse.
- Base the blend mix on either the dependent (first choice) or independent data acquisition variable. Use derived values for the blending vectors with care. The exception is for intrinsic properties that are based on a relative amount of sample. This would include heating value (Btu per weight unit) or density (weight per volume unit). If the dependent data acquisition value does not correspond to the denominator of the property value units, then one might be well advised to make this transform before blending. The heating value, for example, will be very accurate if estimated on a weight percentage proration, even if the data are based on a volume percentage recovery.
- Try to use a linear blend function. This can be done by proper selection of the transform function. Use of higher ordered blending rules often leads to localized anomalies in the generation of the property–index matrix. These anomalies will probably not correspond to the real world and can drastically disrupt the use of such blend functions in planning, real-time control and optimization operations.

GLOSSARY OF TERMS

Alkane	Hydrocarbons conforming to the formula C_nH_{2n+2} with no saturated bonds or heteroatoms in the molecular structure.
Antiknock Index	Synonym for octane rating.
Aromatics	Hydrocarbons with at least one fully saturated ring structure, but with no heteroatoms in the molecular structure
ASTM Distillation	A set of standard procedures governing petroleum distillations. Usually a synonym for ASTM D-86 procedures, sometimes for D-1160.
Atmospheric Resid	The remainder or bottoms of the crude feed to a distillation column that runs at or near atmospheric pressure. Most of the lighter cuts have been taken off in the tower.
Catalytic Reformate	The product of a reaction process that changes naphthene rings into aromatics. Product changes low octane cut to high octane cut.
Coker/Coking	The process/tendency of a hydrocarbon, under thermal stress, to crack and form a mostly carbon solid phase, with simultaneous evolution of some lighter fluid hydrocarbon components

Conversion Reactor	A chemical process in a refinery that cracks and provides subsequent processing such as hydrogenation and distillation for separation of cuts.
Conversion Refinery	A refinery that has one or more conversion reactors. Typically, most sizeable refineries have at least three of the following units: catalytic cracker, coker/ thermal cracker, hydrotreater (usually several), hydrocracker, catalytic reformer.
Cracked	The condition of having long hydrocarbon chains thermally or catalytically broken. In lighter components, the result is often one or more unsaturated bonds between some carbons. In heavier stocks, a mostly carbon (called coke) phase forms with devolution of some lighter hydrocarbons. Thermal cracking is an endothermic process.
Crude Unit	Synonym for the tower/column (or series of same) that first processes crude oil feed to a refinery.
Cut	The name given to a distillation product that is a ''cut'' of the crude feed. Alternatively, it refers to the measured amounts of analyte that come off the column in specified boiling point or TBP ranges.
D-2887 Distillation	The ASTM standard most often associated with GC analysis of petroleum wide-cut boiling ranges.
Diesel	The cut or fraction of a crude that is used as fuel for compression–ignition engines.
FCC	Acronym for fluid catalytic (''cat'') cracker.
Fluid Catalytic Cracking	The process whereby diesel and gas-oil range hydrocarbons are vaporized cracked with the aid of a heterogeneous catalyst at high temperature. The coke formed in the process is subsequently burned to provide the thermal energy necessary to drive the reactions.
Fraction	Synonym for cut.
Gas-oil (atm, vacuum)	The fraction of a crude feed that is too heavy for diesel fuel and still comes off the column at a higher temperature. It generally makes an excellent feed for FCC and hydrocracker units.
Gasoline	The cut or fraction of a crude that is currently used as fuel for internal combustion engines.
GC	Acronym for gas chromatography.
Hydroprocessing	The process whereby diesel and gas-oil range hydrocarbons are vaporized and cracked in the presence of a hydrogen-rich environment with the aid of a heterogeneous catalyst at high temperature. Very little coke is formed in the process and most unsaturated components are saturated. Depending upon the severity, even aromatics can be hydrogenated.
Hypothetical Component	The name given to the lumping of several dozen to hundreds of individual species of hydrocarbons and related organic compounds into a single TBP boiling range. It can vary from using one to several ''hypothetical'' compound to represent the entire collection of species.
Kerosine	The standard name for the cut or fraction of a crude that is currently used as fuel for turbine engines or for campfire fuel. It is slightly heavier than gasoline, but much less volatile and dangerous for handling by humans.
Light Ends	The name for an accumulation of very light (lighter than gasoline) collection of hydrocarbons in a distillation column. Alternatively, this term refers to the lighter (as defined by boiling point) few percent of any distillation curve.

Multivariate	The use of several types of analysis and characterization of petroleum cuts and fractions to arrive at some sort of speciation of the components, at least by type if not by actual accounting. It involves solution of multiple equations with matrix coefficients taken from several different data sources.
Naphtha (Light, Heavy)	Gasoline and kerosine boiling range material. Light naphtha refers to the gasoline components, and heavy naphtha refers to the compounds with boiling points nearer the kerosine range. Usually it is synonymous with the top two or three product cuts from the crude unit.
Naphthenes	Alkanes possessing one or more cyclic but unsaturated rings.
Octane	The tendency of a naphtha fuel to pre-ignite or ''knock'' in an internal combustion engine.
Paraffin	Popular synonym for alkane, also for aliphatic.
Prefrac Column	A sometimes used device that strips a crude of its lighter naphtha cuts before heating the crude to maximum temperature.
Pseudo-component	Synonym for hypothetical component
Real Time	The protocol whereby data about a process are obtained, analyzed and utilized in the control of that process within a short period of time, well under the time constant of the process frequency response function.
Reduced Crude	The term for either atmospheric resid or for crude that has been run through a prefrac column to remove a significant portion of its lighter naphtha content.
Resid	The remainder or bottoms of the crude feed to a distillation column that runs at or near atmospheric pressure. Most of the lighter cuts have been taken off in the tower.
Stabilizer	The common term for a distillation column that is used to strip off very light ends from a hydrocarbon cut in order to make a product specification or safety related reason.
Straight Run	Synonym for virgin cut as applied to physical distillation fractions.
TBP Distillation	The generic term for refluxed laboratory distillation that approximates the ''TBP'' of a petroleum cut or fraction. ASTM D-2892 is ostensibly the standard, but there are so many variations as to make an ''official'' standard superfluous.
Thermal Cracking	The process whereby heat is applied to a hydrocarbon, resulting in its ''cracking''. Delayed cokers work on what is called a ''heat soak'' process whereby heavy hydrocarbon cuts will crack noncatalytically if a high enough temperature is maintained in the reactor for a long enough time. One of the products is coke, hence the name ''coker''.
Topping Refinery	A refinery whose sole purpose is to recover fuel quality material by distillation and hot flashing of crude/vapor feeds. No conversion reactors, except for perhaps a coker unit, are found in this type of refinery.
Tower/Column	The piece of process equipment that is used to separate hydrocarbons on the basis of boiling points by use of a heat source driving all vapors upward and refluxing/condensing the heavier cuts downward by liquid gravity flow. A modern refinery contains dozens of columns. The first unit that a crude feed usually sees is the crude unit, which typically consists of one to three columns operating at different pressure levels.

Transport Properties	The physical properties of a substance that are the primary quantities involved in the characterization of mass, momentum and heat transfer in industrial hydrocarbon processing operations.
Transport Fuels	The naphtha and diesel range cuts that provide the primary bulk of liquid fuels for current domestic transportation methods using some form of a combustion engine, internal combustion, turbine or compression–ignition.
Vacuum Resid	The remainder or bottoms of the crude feed to a distillation column that runs at far below atmospheric pressure. It contains most of the lighter cuts that were not taken off in the atmospheric column. It allows vaporization of components at a lower temperature than at atmospheric pressure, thus reducing the chance of thermal cracking and coking during the processing.
Virgin	Naturally occurring as found in the feed crude oil and not having been processed in a conversion (chemical) reactor. Physical processing such as distillation is acceptable for this nomenclature as long as the chemical structure is not changed.

ABBREVIATIONS AND ACRONYMS

AG	Aniline Gravity Constant
AP	Aniline Point
API	American Petroleum Institute
ASTM	American Society for Testing and Materials
ATM	Atmospheric
BMCI	Bureau of Mines Correlation Index
BS & W	Bottom Sediment and Water
CABP	Cubic Average Boiling Point
CARB	California Air Resources Board
CC	Conradson Carbon Number
CGSB	Canadian General Standards Board
CI	Cetane Index
CN	Cetane Number
CP	Cloud Point
DI	Diesel Index
EFV	Equilibrium Flash Vaporization
EPA	Environmental Protection Agency (US)
FP	Flash Point
FRP	Freeze Point
GC	Gas Chromatography
GC/IR	Gas Chromatography/Infrared
GC/MS	Gas Chromatography/ Mass Spectrometry
GLR	Gas–Liquid Ratio
GPC	Gel Permeation Chromatography
HDA	Hydrodealkylation
HPLC	High-performance Liquid Chromatography
IR	Infrared
LSR	Light Straight Run
LVGO	Light Vacuum Gas-oil
MABP	Molar Average Boiling Point
MEABP	Mean Average Boiling Point
MS	Mass Spectroscopy
MW	Molecular Weight
NBP	Normal Boiling Point
NMR	Nuclear Magnetic Resonance
PONA	Paraffins, Olefins, Naphthenes and Aromatics
PP	Pour Point
RC	Ramsbottom Carbon Number
RI	Refractive Index
RVP	Reid Vapor Pressure
SD or Simdist	Simulated Distillation
SG	Specific Gravity
SP	Smoke Point
TBP	True Boiling Point
USBM	United States Bureau of Mines
VABP	Volumetric Average Boiling Point
VBN	Viscosity Blend Number

REFERENCES

1. L.R. Aalund, 'North Sea Now Offers 14 Export Crudes', *Oil Gas J.*, May 23, 69–74 (1983).
2. L. Aalund, 'North Sea Crudes: Flotta to Thistle', *Oil Gas J.*, June 6, 75–79 (1983).
3. L. Aalund, 'Here are more World-crude Analyses', *Oil Gas J.*, April 12, 72–78 (1976).
4. I.S. Akhmetzhanov, N.K. Nikulina, S.M. Dmitriev, '[An Experimental] Correlation of Flash Point with Boiling Point [is Demonstrated] for Individual Organic Compounds', *Khim. Tekhnol. Topl. Masel*, (8), 45 (1974).
5. J.M. Al-Besharah, S.A. Akashah, 'Binary Mixtures of Petroleum Products and Mixtures of Base Oils Examined', *Oil Gas J.*, March 6, 50–52 (1989).
6. J.M. Al-Besharah, *Ind. Eng. Chem. Res.*, **28**(2), 214 (1989).

7. M.F. Ali, M.V. Hasan, A.M. Bukhari, M. Saleem, 'Arabian Crude Fractions Analyzed', *Hydrocarbon Process.*, **64**(2), 83 (1985).
8. M.F. Ali, M.V. Hasan et al., 'Arabian Crude-oil Residues Evaluated', *Oil Gas J.*, **83**(32), Aug. 12, 71–77 (1985).
9. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 6th edition, American Petroleum Institute, Washington, DC, 1994.
10. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 5th edition, American Petroleum Institute, Washington, DC, 1992.
11. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 4th edition, American Petroleum Institute, Washington, DC, 1987.
12. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 3rd edition, American Petroleum Institute, Washington, DC, 1982.
13. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 2nd edition, American Petroleum Institute, Washington, DC, 1978.
14. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 1st edition, American Petroleum Institute, Washington, DC, 1972.
15. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 6th edition, American Petroleum Institute, Washington, DC, 1996 revision package.
16. American Petroleum Institute (API), *Technical Data Book–Petroleum Refining*, 6th edition, American Petroleum Institute, Washington, DC, 1998 revision package.
17. API Research Project 62, 'Thermodynamics of Hydrocarbons from Petroleum', Bartlesville Energy Research Center, Bartlesville, OK, 1972.
18. API Research Project 42, 'Properties of Hydrocarbons of High Molecular Weight', American Petroleum Institute, Washington, DC, 1966.
19. ASTM, *1998 Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, PA, Section 5, 1998.
20. ASTM D-56-94, 'Standard Test Method for Flash and Fire Points by TAG Closed Cup', American Society for Testing and Materials, Philadelphia, PA.
21. ASTM D-86-95, 'Standard Test Method for Atmospheric Distillation of Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
22. ASTM D-92-90, 'Standard Test Method for Flash and Fire Points by Cleveland Open Cup', American Society for Testing and Materials, Philadelphia, PA.
23. ASTM D-93-90 'Standard Test Methods for Flash Point by Pensky–Martens Closed Tester,' American Society for Testing and Materials, Philadelphia, PA.
24. ASTM D-97-87, 'Standard Test Method for Pour Point of Petroleum Oils', American Society for Testing and Materials, Philadelphia, PA.
25. ASTM D-189-88, 'Standard Guide for Conradson Carbon Residue of Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
26. ASTM D-240-88, 'Standard Guide for Heat of Combustion of Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
27. ASTM D-323-94, 'Standard Guide for Vapor Pressure of Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
28. ASTM D-524-88, 'Standard Test Method for Ramsbottom Carbon Residue of Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
29. ASTM D-611-82, (1987), 'Standard Test Method for Aniline Point and Mixed Aniline Point of Petroleum Products and Hydrocarbon Solvents', American Society for Testing and Materials, Philadelphia, PA.
30. ASTM D-613-88, 'Standard Test Method for Cetane Number', American Society for Testing and Materials, Philadelphia, PA.
31. ASTM D-976-88, 'Standard Test Method for Cetane Number', American Society for Testing and Materials, Philadelphia, PA.
32. ASTM D-976-80, 'Standard Test Method for Cetane Number', American Society for Testing and Materials, Philadelphia, PA.
33. ASTM D-1159-93, 'Standard Test Method for Bromine Numbers of Petroleum Distillates and Commercial Aliphatic Olefins by Electrometric Titration', American Society for Testing and Materials, Philadelphia, PA.
34. ASTM D-1160-90, 'Standard Test Method for Vacuum Distillation of Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
35. ASTM D-1218, 'Standard Test Method for Refractive Index of Organic Liquids', American Society for Testing and Materials, Philadelphia, PA.
36. ASTM D-1332, 'Smoke Point of Aviation Turbine Fuels', American Society for Testing and Materials, Philadelphia, PA.
37. ASTM D-1492-92, 'Standard Test Method for Bromine Index of Aromatic Hydrocarbons by Coulometric Titration', American Society for Testing and Materials, Philadelphia, PA.
38. ASTM D-2500-88, 'Standard Test Method for Cloud Point of Petroleum Oils', American Society for Testing and Materials, Philadelphia, PA.
39. ASTM D-2502-92, 'Standard Test Method for Estimation of Molecular Weight of Petroleum Oils from Viscosity', American Society for Testing and Materials, Philadelphia, PA.
40. ASTM D-2710-92, 'Standard Test Method for Bromine Index of Petroleum Hydrocarbons by Electrometric Titration', American Society for Testing and Materials, Philadelphia, PA.

41. ASTM D-2887-93e1, 'Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography', American Society for Testing and Materials, Philadelphia, PA.
42. ASTM D-2892-85, 'Standard Test Method for Distillation of Crude Petroleum (Reflux Method)', American Society for Testing and Materials, Philadelphia, PA.
43. ASTM D-3710-95, 'Standard Test Method for Boiling Range Distribution of Gasoline Fractions by Gas Chromatography', American Society for Testing and Materials, Philadelphia, PA.
44. ASTM D-4502, 'Standard Test Method for Densities of Petroleum Fractions', American Society for Testing and Materials, Philadelphia, PA.
45. ASTM D-4737-90, 'Standard Test Method for Calculated Cetane Index by Four Variable Equation', American Society for Testing and Materials, Philadelphia, PA.
46. ASTM D-4814-93a, 'Standard Specification for Automotive Spark-ignition Engine Fuel', American Society for Testing and Materials, Philadelphia, PA.
47. ASTM D-4815-93, 'Standard Test Method for Determination of MTBE, ETBE, TAME, DIPE, *tertiary*-Amyl Alcohol and C1 to C4 alcohols in Gasoline by Gas Chromatography', American Society for Testing and Materials, Philadelphia, PA.
48. ASTM D-4953-95, 'Standard Guide for Vapor Pressure of Oxygenated Petroleum Products', American Society for Testing and Materials, Philadelphia, PA.
49. ASTM D-5191-94, 'Standard Guide for Vapor Pressure of Petroleum Products–Pressure Cell Method', American Society for Testing and Materials, Philadelphia, PA. [Equivalent to ISO 3007–1996].
50. C.T. Baird, IV, *Guide to Product Blending*, HPI Consultants, Austin, TX, 1989.
51. Bartlesville Energy Technology Center, 'Analysis of 800 Crude Oils from US Oil Fields', BETC/Rl-78/14, US Department of Energy, Bartlesville, OK, 1978.
52. C.J.S. Bartlett, 'Relationship Between Aromatic Content and Specification Parameters in Diesel Fuels', *Pet. Rev.*, **44**(517), 84 (1990).
53. C.J.S. Bartlett, 'A New Equation [Proposed by ASTM] for Calculation of Cetane Index of Distillate Fuels', *Pet. Rev.*, **41**(485), (1987).
54. A.E. Becker, H.G.M. Fischer, 'A Suggested Index of Diesel Fuel', *SAE J.*, **35**, 376 (Oct. 1934).
55. S.A. Beg et al., 'The Arab Light Crude Study Focuses on Kinematic Viscosity', *Fuel Sci. Technol. Int.*, **7**(2), 187 (1989).
56. R.L. Davidson, 'Evaluation of Crude Oils', in *Petroleum Processing Handbook*, ed. W.F. Bland, McGraw-Hill, New York, 1967.
57. Bureau of Mines Technical Paper No. 610, 'Correlation Index to Aid in Interpreting Crude-oil Analysis', American Petroleum Institute, Washington, DC 1960.
58. R.M. Butler, 'Prediction of Flash Points of Middle Distillates', *Ind. Eng. Chem.*, **48**, 808 (1956).
59. J.A. Carpenter, 'The Composition of Petroleum (Kerosene and Other) Fractions', *J. Inst. Pet. Technol.*, **14**, 446–476 (1928).
60. R.H. Cavett, 'Physical Data for Distillation Calculations–Vapor/Liquid Equilibria', presented at 27th Midyear Meeting of API Division of Refining, San Francisco, May 15, 1962.
61. 'Characterization of Heavy Oils and its Application–5. Characterization of Heavy Oils', *J. Jpn. Pet. Inst.*, **27**(1), 74 (1984).
62. P. Claudy, 'Diesel Fuels: Determination of Onset Crystallization Temperature, Pour Point and Filter Plugging Point by Differential Scanning Calorimetry. Correlation with Standard Test Methods', *Fuel*, **65**(6), 861 (1986).
63. *CRC Netbase*, CRC Press, Cleveland, OH, 1999.
64. T.E. Daubert, 'Characterization Parameters for Hydrocarbons for Physical and Thermodynamic Property Prediction', in *Proceedings of 58th GPA Convention–1985*, Gas Processors Association, Tulsa, OK, 17–25, 1985.
65. T.E. Daubert, N.C. Daubert, 'Documentation for the Basis for Selection of the Contents of Chapter 2 for the API Technical Data Book', Report 3–93, University Microfilms, Ann Arbor, MI, 1993.
66. T.E. Daubert, N.C. Daubert, 'Documentation for the Basis for Selection of the Contents of Chapters 2 and 3 for the API Technical Data Book', Reports 2–86, 3–86, University Microfilms, Ann Arbor, MI, 1986.
67. A. Ecker, 'The Low-temperature Behavior of Petroleum Products. Thermoanalytic Studies', *Erdöl Kohle Erdgas Petrochem Brennst.-Chem.*, **38**(6), 281 (1985).
68. W.C. Edmister, *Applied Hydrocarbon Thermodynamics*, Gulf Publishing, Houston, TX, 1961.
69. W.C. Edmister, D.H. Pollock, 'Phase Relations for Petroleum Fractions', *Chem. Eng. Prog.*, **44**, 905 (1948).
70. W.C. Edmister, K.K. Okamoto, *Pet. Ref.*, **38**(9), 271 (1959).
71. E.O. Egbogah, J.T. Ng, 'An Improved Temperature–Viscosity Correlation for Crude Oil Systems', *J. Pet. Sci. Eng.*, **5**, 197 (1990).
72. M. Ghassemi, 'Physical and Chemical Characteristics of Some Widely Used Petroleum Fuels: A Reference Data Base for Assessing Similarities and Differences Between Synfuel and Petrofuel Products', *Energy Sources*, **7**(4), 4 (1984).
73. M. Ghassemi, A. Panahloo, S. Quinlivan, 'Physical and Chemical Characteristics of Some Widely Used Petroleum Fuels: A Reference Data Base for Assessing Similarities and Differences between Synfuel and Petrofuel Products', *Energy Sources*, **7**(4), 377–401 (1984).
74. B.A. Grigor'ev, A.I. Svidchenko, 'Methods for Calculation Thermal Conductivity of Liquid Petroleum Crudes and Products', *Khim. Technol. Topl. Masel*, **8**(Aug.), 49–50 (1979) (Engl. transl.).
75. S.T. Hadden, *Chem. Eng. Prog.*, **44**, 135 (1948).

76. R.W. Hankinson, L.K. Thomas, K.A. Phillips, *Hydrocarbon Processes*, **48**(4), 106 (1969).
77. O.H. Hariu, R.C. Sage, 'Crude Split Figured by Computer', *Hydrocarbon Processes*, **48**(4), 143–148 (1969).
78. A.E. Hirschler, 'Molecular Weights of Viscous Hydrocarbon Oils: Correlation of Density with Viscosities', *J. Inst. Pet. Technol.*, **32**, 133–161 (1946).
79. O.A. Hougen, K.M. Watson, *Chemical Process Principles*, John Wiley, New York, 1943.
80. D.A. Howes, 'The Chemical Examination of Gasolines and Light Mineral Oils', *J. Inst. Pet. Technol.*, **16**, 54–88 (1930).
81. J. Hu, A.M. Burns, 'New Method Predicts Cloud, Pour, Flash Points of Distillate Blends', *Hydrocarbon Process.*, **49**(11), 216 (1970).
82. P.K. Huang, 'Characterization and Thermodynamic Correlations for Undefined Hydrocarbon Mixtures', *Ind. Eng. Chem. Process. Des. Dev.*, **13**, 359 (1974).
83. S.M. Jacob, R.J. Quann, E. Sanchez, 'Compositional Modeling Reduces Crude-analysis Time, Predicts Yields', *Oil Gas J.*, July 6, 51–56 (1998).
84. G.I. Jenkins, R.P. Walsh, 'Quick Measure of Jet Fuel Properties', *Hydrocarbon Process.*, **47**(5), 161 (1968).
85. G.I. Jenkins, 'The Characterization of Lubricating Oil Basestocks: The Significance and Interdependence of the Inspection Tests, Elemental Analyses, and Carbon Type Analyses', *J. Inst. Pet.*, **52**, 250 (1966).
86. W.B. Kay, *Ind. Eng. Chem.*, **28**, 1408 (1936); *Oil Gas J.*, **91**(44), 74 (1939).
87. M.G. Kesler, B.I. Lee, 'Improve Prediction of Enthalpy of Fractions', *Hydrocarbon Process.*, March, 153–158 (1976).
88. H.U. Khan, 'New Methods Determine Pour Point More Accurately than ASTM D-97', *Oil Gas J.*, **91**(44), 74 (1993).
89. R. Krishna, 'Correlation of Pour Point of Gas Oil and Vacuum Gas Oil Fractions with Compositional Parameters', *Energy Fuels*, **3**(l), 15 (1989).
90. K. Krisnanqkura, 'Simple Method for Estimation of Cetane Index of Vegetable Oil Methyl Esters', *J. Am. Oil Chem. Soc.*, **63**(4), 552 (1986).
91. M. Krutz, H. Krutz, 'A Computerized Information System on Crude Oils', in *1985 API–EPA–US Coast Guard Oil Spill Prevention, Behavior, Control, Cleanup Conference, Los Angeles, 2/25/85–2/28/85*, API Publ. No. 4385, American Petroleum Institute, Washington, DC, 203–207, 1985.
92. J.R. Dean (ed.), *Lange's Handbook of Chemistry*, McGraw-Hill, New York, 1998.
93. F.W. Lane, E.W. Dean, '(San Joaquin Valley Crude)–Viscosity–Temperature Curves of Typical American Crude Oils', *Ind. Eng. Chem.*, September (1921).
94. J.M. Lenoir, H.G. Hipkin, *J. Chem. Eng. Data*, **18**, 195 (1973).
95. H.R. Linden, 'The Relationship of Physical Properties and Ultimate Analysis for Liquid Hydrocarbon', *Oil Gas J.*, **48**(9), 60–65 (1949).
96. B.J. Mair, C.B. Willingham, A.J. Streiff, 'Hydrogenation of the ''Extract'' Portion of the Lubricant Fraction from a Midcontinent Petroleum', *J. Res. Natl. Bur. Stand.*, **21**, 565–580 (1938).
97. J.B. Maxwell, *Data Book on Hydrocarbons–Application to Process Engineering*, Van Nostrand, Princeton, NJ, 19–23, 1950.
98. C.M. McKinney, E.L. Garton, 'Analyses of Crude Oils from 470 Important Oil Fields in the United States', *US Bureau of Mines Report of Investigations*, 1957.
99. H. Mikami, 'Effect of Ambient Pressure on the Smoke Point of Olefines', *Trans. Jpn. Soc. Mech. Eng., Part B*, **59**(563), 211–216 (1993).
100. R.C. Mithoff, 'Method Analyses of Typical Crude Oil from Long Beach Field, California', *Oil Gas J.*, Nov. 6 (1941).
101. R.C. Mithoff, 'Characteristics of California Crude Oils', *Oil Gas J.*, **40**(26), 81 (1941).
102. Y. Nakano, 'Estimation of Flash Point of Emulsified Fuels', *Institution of Chemical Engineers*, Rugby, No. 120, 437–447 (1990).
103. W.L. Nelson, 'CF's of Hydrocarbons', *Oil Gas J.*, Sept. 26, 102 (1966).
104. W.L. Nelson, 'Characterization Factors of Hydrocrackates', *Oil Gas J.*, June 12, 86 (1967).
105. W.L. Nelson, 'Characterization Factor', *Oil Gas J.*, Aug. 8, 63 (1966).
106. Nelson, *Oil Gas J.*, June 24, 325 (1948).
107. S.A. Newman, 'Correlations Evaluated for Coal-tar Liquids', *Hydrocarbon Process.*, December, 133–142 (1981).
108. R. Nokay, 'Estimate Petrochemical Properties', *Chem. Eng.*, **66**(4), 147 (1959).
109. S.G. Pande, 'Cetane Number Predictions of a Trial Index Based on Compositional Analysis', *Energy Fuels*, **3**(3), 308 (1989).
110. S.G. Pande, D.R. Hardy, 'A Practical Evaluation of Published Cetane Indices', *Fuel*, **69**(4), 437 (1990).
111. R. Pasternak, P. Brody, H. Ehrmantraut, 'An Instrument for the Fast and Precise Determination of Molecular Weight', paper presented at ACHEMA-13 Chemical Engineering Conference, Frankfurt-am-Rhein, June 1961.
112. G.S. Patil, 'Estimation of Flash Point', *Fire Mater.*, **12**(3), 127 (1988).
113. K.S. Pedersen, 'Prediction of Cloud Point Temperatures and Amount of Wax Precipitation', *Society of Petroleum Engineers of AIME, (Paper) SPE*, 1–20, 1994.
114. G. Protic-Iovasic, N. Jambrec et al., 'The Determination of Catalytic Reformed Gasoline Octane Number by High Resolution Gas Chromatography', *Fuel*, **69**(4), 525 (1990).

115. A.K. Rhodes, 'Yemen's Light, Sweet Alif Crude Assayed', *Oil Gas J.*, May 23, 78–80 (1994).
116. A.K. Rhodes, 'Two Offshore Australian Crudes Assayed', *Oil Gas J.*, May 9, 76–79 (1994).
117. M.R. Riazi, T.E. Daubert, 'Predicting Flash and Pour Points', *Hydrocarbon Process.*, **66**(9), 81–83 (1987).
118. M.R. Riazi, T.E. Daubert, 'Simplify Property Predictions', *Hydrocarbon Process.*, **59**(3), 115–116 (1980).
119. M.R. Riazi, 'Prediction of Thermophysical Properties of Petroleum Fractions', PhD Thesis, Department of Chemical Engineering, Penn State University, University Park, PA, 1979.
120. M.R. Riazi, T.E. Daubert, 'Molecular Weight of Heavy-oil Fractions from Viscosity', *Oil Gas J.*, Dec. 28, 110–112 (1987).
121. L.C. Roess, 'Determination of Critical Temperature and Pressure of Petroleum Fractions by a Flow Method', *J. Inst. Pet. Technol.*, **22**, 665 (1936).
122. H.P. Ronninsen et al., 'Wax Precipitation from North Sea Crude Oils. 1. Crystallization and Dissolution Temperatures, and Newtonian and Non-newtonian Flow Properties', *Energy Fuels*, **5**, 895 (1991).
123. D.J. Ruzicka, 'Modified Method Measures Bromine Number of Heavy Fuel Oils', *Oil Gas J.*, **85**(31), 48 (1987).
124. T.W. Ryan, J. Erwin et al., 'The Relationships between Fuel Properties and Composition and Diesel Engine Combustion Performance and Emissions', SAE International Congress (Detroit, 28 February–3 March 1994), SAE Meeting Paper No. 941 018, 1994.
125. D.F. Schneider, 'Select the Right Hydrocarbon MW Correlation', *Chem. Eng. Prog. Dec.*, 40–44 (1998).
126. P.B. Semwal, 'New Pour Point Correlations Improve Economics of Distillate Blending', *Oil Gas J.*, **92**(23), 89 (1994).
127. Y.N. Shebeko, A.V. Ivanov, T.M. Dmitreavea, 'Methods of Calculation of Lower Concentration Limits of Combustion of Gases and Vapors in Air', *Sov. Chem. Ind.*, **15**(3), 311 (1983).
128. E.M. Shelton, 'Motor Gasolines, Winter 1983–84', US, Department of Energy Petroleum Products Survey, No. 84/3, June 1984.
129. E.M. Shelton, 'Heating Oils, 1983', US Department of Energy, Bartlesville Energy Technology Center, Petroleum Products Survey, No. 83/4, Aug. 1983.
130. H. Singh, 'Classification of Crude Oils From Refiner's Viewpoint. [A Review of the] Methods and Significance', *Erdöl Kohle Erdgas Petrochem. Brennst.-Chem., Hydrocarbon Technol.*, **41**(II), 458 (1988).
131. N.A.C. Smith, H.M. Smith, C.M. McKinney, 'Refining Properties of New Crudes: Part 4 – Crude Oils from Recent Discoveries in Kansas', *Pet. Process.*, **5**, 960 (1950).
132. Smith, Watson, 'Boiling Points and Critical Properties of Hydrocarbon Mixtures', *Ind. Eng. Chem.*, **29**, 1408 (1937).
133. J. Smittenberg, D. Mulder, 'Relations between Refraction, Density and Structure of Series of Homologous Hydrocarbons. I. Empirical Formulae for Refraction and Density at 20 °C of *n*-alkanes and *n*-alpha-alkanes', *Recl. Trav. Chim. Pays-Bas*, **67**, 813–825 (1948); 'II. Refraction and Density at 20 °C of *n*-alkylcyclo-pentanes, -cyclohexanes and -benzenes', *Recl. Trav. Chim. Pays-Bas*, **67**, 826–838 (1948).
134. M.J. Solomon, R.J. Quann, E. Sanchez, M. Wells, 'Compositional Modeling Reduces Crude-analysis Time, Predicts Yields', *Oil Gas J.*, 6 July, 51 (1998).
135. H. Stenzel, H. Roedicker, 'Results of Substance and Structural Group Analysis of Petroleum Fractions and their Application in Estimating the Correlation for Predicting their Properties', *Chem. Tech.*, **43**(9), 344 (1991).
136. R.M. Stephenson, *Flash Points of Organic and Organometallic Compounds*, Elsevier, Amsterdam, 1987.
137. H.G. Struppe, K. Anders, S. Nowak, 'Analysis of Higher-boiling Hydrocarbon Fractions', *Erdöl Kohle Erdgas Petrochem. Brennst.-Chem., Hydrocarbon Technol.*, **39**(100), 457 (1986).
138. G.J. Su, P.H. Huang, T.M. Chang, *J. Am. Chem. Soc.*, **68**, 1403 (1946).
139. B. Sudbury, 'Analysis of Petroleum/Physical Properties', *Anal. Chem.*, **55**(5), 268 (1983).
140. T. Suzuki, K. Ohtaguchi, K. Koide, 'Correlation and Prediction of Autoignition Temperatures of Hydrocarbons Using Molecular Properties', *J. Chem. Eng. Jpn.*, **25**(5), 606 (1992).
141. W.R. Thomson, *Pet. Rev.*, May 37–40 (1983).
142. Twu, H. Chorng, 'An Internally Consistent Correlation for Predicting Critical Properties and Molecular Weight of Petroleum and Coal Tar Liquid', *Fluid Phase Equil.*, **16**, 137–150 (1984).
143. Twu, H. Chorng, 'Prediction of Thermodynamic Properties of Normal Paraffins Using Only Normal Boiling Point', *Fluid Phase Equil.*, **16**, 65–81 (1984).
144. Twu, H. Chorng, 'Boiling Point as a Third Parameter for Use in a Generalized Equation of State', *Fluid Phase Equil.*, **11**, 189–194 (1983).
145. K. Van Nes, H.A. Van Westen, *Aspects of the Constitution of Mineral Oils*, Elsevier, Amsterdam, 1951.
146. M. Van Winkle, Tag, 'Flash vs. ASTM Distillation', *Pet. Ref.*, **33**(11), 171 (1954).
147. K.S. Virk, 'Diesel Fuel Cetane Number, Aromatic Content and Exhaust Emissions', *J. Am. Chem. Soc.*, **37**(3), 701 (1992).
148. 'Viscosity and Density of Over 40 Lubricating Fluids of Known Composition at Pressures to 150 000 psi and Temperatures to 425 F', ASME Research Committee on Lubrication Report, American Society of Mechanical Engineer, New York, NY, 1953.
149. R.N. Watkins, *Petroleum Refinery Distillation*, Gulf Publishing, Houston, TX, 1979.

150. R.N. Watkins, 'Tia Juana Light, Crude Assay (Appendix 1)', in ed. R.N. Watkins, *Petroleum Refinery Distillation*, 2nd Edition, Gulf Publishing, Houston, TX, 1969.
151. K.M. Watson, W.L. Nelson, 'Improved Methods for Approximating the Critical and Thermal Properties of Petroleum Fractions', *Ind. Eng. Chem.*, **25**, 880 (1933).
152. K.M. Watson, W.L. Nelson, L.A. Murphy, 'Characterization of Petroleum Fractions', *Ind. Eng. Chem.*, **27**, 1460 (1935).
153. K.L. West, R.W. Hankinson, L. Downer, 'Petroleum Measurement Tables of 1980 Reaffirmed', *Oil Gas J.*, **87**(32), 63 (1989).
154. F.W. Winn, 'Physical Properties by Nomogram', *Pet. Ref.*, **36**(2), 157–159 (1957).
155. K.W. Won, 'Thermodynamic Calculation of Cloud Point Temperatures and Wax Phase Compositions of Refined Hydrocarbon Mixtures', *Fluid Phase Equil.*, **53**, 377 (1989).
156. R.A. Woodle, 'New Ways to Estimate the Characterization [Factor (K)] of Lube Cuts', *Hydrocarbon Process.*, **59**(7), 171 (1980).
157. P. Zhou, 'Correlation of the Average Boiling Points of Petroleum Fractions with Pseudocritical Constants', *Int. Chem. Eng.*, **24**(4), 731 (1984).
158. C. Kennard (ed.), 'Correlated ASTM Distillation Based on Simulated Distillation (ASTM D-2887)', Hewlett-Packard Application Note AN145, Hewlett-Packard, Avondale, PA, 1978.
159. D.C. Ford, W.H. Miller, R.C. Thuen, R. Weitzler, 'Correlation of D-2887 Boiling Range Distribution Data', ASTM Special Technical Publication No. 577, 20, American Society for Testing and Materials, Philadelphia, PA, 1974.

Pharmaceuticals and Drugs: Introduction

R.K. Gilpin
Wright State University, Dayton, USA

1 Organization and Information 2
2 Biographical Sketch of the Authors 3
3 Conclusion and Future Trends 5
Abbreviations and Acronyms 5

The field of pharmaceutical analysis includes a wide range of analytes, varying in structure from very simple compounds to complex biomolecules. As such, a host of approaches have been and continue to be used in developing reliable analytical methodology for these analytes. In the broadest sense, there are two types of procedures: those that are designed as simple, reliable means of monitoring formulated products in terms of their identity, strength and quality, and those that are used during the drug discovery and development stages to answer numerous and fundamentally more challenging questions related to safety, therapeutic effectiveness, drug stability and purity, as well as in helping to develop better understanding of the biomechanisms and kinetics.

From the economic standpoint, there is the continuing pressure to reduce production and hence product costs, and from the quality assurance side to develop methods that are more reliable, accurate, and sensitive. Over the last thirty years there has been an increasing reliance on separation-based methods combined with various pre-analysis sample clean-up procedures such as solid phase or supercritical fluid extraction and post-column approaches like derivatization fluorescence detection or electrospray mass spectrometry (MS) to improve their accuracy and sensitivity.

From the regulatory standpoint, an important change that occurred in the United States during the late 1970s was the introduction of Good Laboratory Practices by the Federal Drug Administration. These regulations were the direct result of several perceived problems in the pharmaceutical industry and they resulted in greater accountability in terms of calibration, maintenance and record keeping. With the emergence of an expanding number of multinational companies and increasing global markets, similar trends and demands have followed throughout the world.

From the hardware/equipment standpoint, two of the significant changes that have occurred have been miniaturization and automation. Modern and important techniques that have emerged in the last three decades that have had an important impact on modern pharmaceutical analysis include high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), high-field solution and solids nuclear magnetic resonance (NMR), electrospray MS, and computer automation and robotics. Likewise, in terms of sample handling, solid-phase extraction, supercritical extraction and micro-capillary dialysis sampling have also become important tools.

As the direct result of the above pressures and trends, the increasing competitiveness of the marketplace, and the quest for less expensive, safer, and better drugs with well understood and targeted activities, the pharmaceutical industry has been and continues to be one of the leaders in pioneering new analytical methods. This is easily documented by the thousands of published accounts that appear yearly in the scientific literature, the growth in the job market for those obtaining undergraduate degrees in chemistry and graduate degrees in analytical and allied fields of chemistry, and the growth in vendors marketing analytical hardware and speciality items.

Unfortunately, a ''down-side'' of the remarkable growth in the field of chemical analysis and the automation of modern analytical instrumentation and procedures has been a greater isolation of the analyst from the fundamental principles governing the measurement being made. There are many excellent methods published daily, but, on the negative side, there are numerous accounts that have appeared which are based on what might be considered less than reliable conditions. Generally, most practicing analytical chemists understand how to make their hardware perform effectively, and they have a reasonable empirical/working knowledge of how to systematically change conditions in order to produce a given result, but they sometimes either fail to consider or choose to ignore relatively simple fundamental concepts that influence the reliability and ruggedness of their methods. Many of the problems associated with assay instability result from poorly designed procedures that may appear to provide the correct result but are not optimized to minimize variability. Thus, it is the intent of the current section to address the field of pharmaceutical analysis from both the technique and the topic standpoints. There are articles included that are based on particular techniques, that are topical in nature, and, in a few cases, that survey/review the literature from an accounts perspective. A brief overview of each of these follows.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 ORGANIZATION AND INFORMATION

Alkaloids, Pharmaceutical Analysis of. The alkaloid article provides a historical perspective and topical overview of the analysis of more common and pharmaceutically important, naturally occurring bases. It also includes, in tabular form, a review of representative methodology that has appeared in the literature over the last decade.

Antibiotics, Pharmaceutical Analysis of. The antibiotics article provides a brief introduction to the six major classes of antibiotics, the cephalosporins, penicillins, quinolones, streptomyces, sulfonamides and tetracyclines, as well as several other important miscellaneous compounds such as chloramphenicol, isoniazid and trimethoprim. It also includes extensive tables which contain a survey of the recent literature in terms of the analytical methodology that has been introduced for assaying the six major cases of antibiotics.

Chemical Reagents and Derivatization Procedures in Drug Analysis. Derivatization is often used to improve sensitivity or volatility for analytes measured respectively using liquid or gas chromatography (GC). As such, this article provides a discussion of some of the important derivatization procedures used in combination with either liquid chromatography (LC) or GC.

Chiral Purity in Drug Analysis. With the increasing demand for more effective and targeted drugs, the number of assays for evaluating chiral purity has been expanding. The focus of this article is to provide an overview of approaches currently being used to purify chiral drugs and to measure their enantiomeric purity. Regulatory matters also are discussed.

Combinatorial Chemistry Libraries, Analysis of. The advent of modern combinatorial synthetic methods has made it possible to produce dozens of potential leads in a single experiment. Because of this, there is an expanding demand for rapid and sensitive methods to analyze these compounds. This article discusses important aspects of the topic, including the use of high throughput strategies and approaches for handling small quantities of analytes.

Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures. LC is by far the single most widely used technique for assaying pharmaceuticals, and in a majority of the reported LC methods, the separations are carried out under reversed-phase conditions. This article presents an overview of important considerations in the selection of eluent additives and addresses why and how they are used to optimize the chromatographic separations.

Gas and Liquid Chromatography, Column Selection for, in Drug Analysis. Two of the most important analytical techniques used in pharmaceutical analysis are GC and LC. At the heart of each of these techniques are reliable and selective columns. This article provides useful information related to column selection.

Mass Spectrometry in Pharmaceutical Analysis. MS has been an important analytical tool in pharmaceutical analysis for many years, and it is the premier gas chromatographic detection technique in terms of sensitivity and specificity. Unfortunately, many early attempts at coupling MS with LC were less successful. However, over the last decade, the development of modern electrospray technology combined with multi-dimensional (i.e. MS/MS and MS/MS/MS) mass spectrometric analysis has made it an important and operationally functional tool. This article considers various aspects of the use of modern instrumentation for the analysis of pharmaceutically important peptides.

Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis. With the introduction of modern high-field NMR instrumentation and the ability to carry out multi-dimensional and solids experiments, uses of the technique have expanded beyond its important role as the major structural elucidation technique. This article discusses the application of modern NMR techniques in the pharmaceutical field, including its uses in evaluating enantiomeric purity.

Planar Chromatography in Pharmaceutical Analysis. On a historical basis the use of planar (i.e. thin layer) chromatographic methods has always been an important topic in pharmaceutical analysis. Assays based on this technique have served as quick and convenient means of screening compounds for purity and in evaluating their stability as a function of storage conditions. The current article focuses on important topics related to the analysis of pharmaceutical compounds, using modern planar chromatographic approaches.

Proteins and Peptides Purification in Pharmaceuticals Analysis. Interest in the potential use of proteins, protein fragments, and peptide drugs has increased over the last decade. One of the more important analytical aspects of this work has been to develop reliable and convenient means of preparing sufficient quantities of a candidate compound for testing. This article discusses purification approaches useful for these types of compounds.

Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery. Since the beginning of the modern pharmaceutical industry, the quest for methods of predicting a compound's activity and toxicity from structural features in the molecule have been sought. Although the ability to calculate a many-body problem with the superposition of all of the conformational and motional dynamics is still not possible, much has been done in this area, and the current article discusses important aspects and uses of modern computational approaches in drug discovery.

Robotics and Laboratory Automation in Pharmaceuticals Analysis. This current article addresses important hardware and validation considerations related to the automation of modern pharmaceutical analysis. Likewise, examples of the application of automated and robotic procedures for various dosage forms are given.

Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis. Often the most difficult and time-consuming part of an analysis is rendering the analyte in a form that can be conveniently measured using either a classical or instrumental approach. Sample clean-up is often a topic that is ignored or minimized, but may be the major pitfall in carrying out a successful analysis. The current article addresses important aspects of modern sample preparation, including solid-phase extraction and supercritical fluid extraction.

Steroid Analysis. Steroids represent an important class of compounds that are used to treat a variety of health problems. In many cases, adequate detection is one of the more demanding aspects encountered in developing appropriate analytical methodology. The current article discusses a variety of chromatographic, electrochemical, and spectrometric techniques used to assay steroids, and it presents methodology to evaluate them by therapeutic class.

Vibrational Spectroscopy in Drug Discovery, Development and Production. Modern infrared (IR) spectroscopy has become an increasingly important technique in drug discovery, especially as it relates to product development and production monitoring. This article addresses important instrumental and operational aspects of its use, including different sampling and acquisition techniques.

Vitamins: Fat and Water Soluble, Analysis of. Vitamins fall into two major groups, those that are fat soluble and those that are water soluble. In a number of cases (i.e. the fat-soluble vitamins), very subtle structural differences exist that require relatively demanding separation conditions to obtain the desired assay specificity. The current article considers various analytical procedures that have been developed for each of the two broad vitamin groups.

2 BIOGRAPHICAL SKETCH OF THE AUTHORS

A.P. Argekar: Reader in Chemistry, Institute of Science, University of Mumbai, India. B.S. Karnataka University, M.S. and Ph.D., University of Mumbai, India. Recipient of the Indian Drug Manufacturer's Best Paper Award in 1996. Research interests include pharmaceutical analysis, environmental science, and separation techniques.

L.C. Bailey: Professor of Pharmaceutical Chemistry, Department of Pharmaceutical Chemistry, Rutgers University, Piscataway, NJ. B.S. (1957) Chemistry, Fordham University, NY, M.S. (1965) and Ph.D. (1969) Analytical Chemistry, Rutgers University, NJ. Research interests include analysis of drugs in dosage forms, biochemical analytical methods development, stability of pharmaceutical preparations, standardization of pharmaceutical packaging materials.

J.J. Bao: Senior Scientist, Procter and Gamble Pharmaceuticals, Inc., Cincinnati, OH, and Adjunct Professor of Chemistry at Shandong University, Jinan, China. Ph.D. (1993) Analytical Chemistry, Purdue University, Indiana. Research interests include HPLC, HPCE (high-performance capillary electrophoresis), micro-total analysis systems, high-throughput electrophoretic separations, and various detection techniques, such as MS and laser induced fluorescence.

A.T. Balaban: Professor of Organic Chemistry, Bucharest Polytechnic University, Romania. Ph.D. Chemical Engineering and Dr. Habil. (1974) in Organic Chemistry, Polytechnic University, Bucharest, Dipl. Radiochemist, Bucharest University and Institute of Atomic Physics, Bucharest. Research interests include synthesis of heterocyclics, stable nitrogen free-radicals, nitric oxide donors, catalytic isomerizations of polycyclic aromatic hydrocarbons (PAHs), isotopically labeled compounds, theoretical and computational methods related to drug design, chemical applications of graph theory, topological indices, and molecular modeling.

C.J. Bates: Head of Micronutrient Status Research, MRC Human Nutrition Research Laboratory, Cambridge, UK. B.A. (1961) Biochemistry, Oxford University, M.A. and Ph.D. (1964) Oxford University, Post-Doctoral Fellow (1964–66) Yale University School of Medicine. Research interests include human requirements and status indices for micronutrients, especially vitamins.

P. Campíns-Falcó: Titular Professor of Analytical Chemistry, University of Valencia. B.S. (1981) Chemistry, Ph.D. (1985) Chemistry, University of Valencia, Valencia, Spain and Postdoctoral Fellow (1982–84), University of Valencia. Research interests include new methods of calibration and data interpretation, automated development of analytical chemical methods (multidimensional chromatography and flow injection analysis), global methods of analysis in several matrixes, and testing and correction of systematic bias error.

N.D. Danielson: Professor of Chemistry, Miami University, Oxford, OH. B.S. (1972) Chemistry, Iowa State University, M.S. (1974) Analytical Chemistry, University of Nebraska, and Ph.D. (1978) Analytical Chemistry, University of Georgia. Research interests include CE, HPLC, microflow injection analysis, and chemiluminescence.

B. Dawson: Research Scientist, Health Canada, Ottawa, ON, Canada. B.S. (1972) and M.S. (1974) Chemistry from St. Francis Xavier University, Nova Scotia, Ontario,

Ph.D. Chemistry (1981), Carleton University, Ontario, and Postdoctoral Fellow (1982) University of Western Ontario. Research interests include NMR studies of drugs and related materials, structure elucidation of pharmaceutical impurities, degradation products, and metabolites, illicit and ''designer'' drugs, and natural products, and the applications of NMR spectroscopy for the determination of isomeric content of drugs.

L.P. Dudones: Research Scientist, Ricerca, Inc., Painesville, OH. B.S. (1992) Chemistry, Hiram College, OH, and Ph.D. (1998) Analytical Chemistry, Kent State University, OH. Research interests include chromatographic and NMR studies of chemically modified surfaces, elucidation of chromatographic mechanisms, and HPLC methods development for pharmaceutical analysis.

K. Duff: R&D Supervisor, HPLC Products, Restek Corporation, State College, PA. B.A. (1979) Chemistry, Lycoming College, PA. M.S. (1982) Organic Synthesis/Biochemistry, Rensselaer Polytechnic Institute, NY. Research interests include the development and characterization of novel chromatographic media such as a new line of ''Allure'' LC/MS phases, ''Ultra'' HPLC packings, and numerous other HPLC packing materials and columns.

J.W. Finn: Senior Scientist/Laboratory Head, Genetics Institute. B.S. (1971) University of Minnesota, Minneapolis, MN, Ph.D. (1981) Analytical Chemistry, Kansas State University.

P.A. Gallagher: Research Chemist, Environmental Protection Agency, Cincinnati, OH. B.S. (1993) Chemistry, Thomas More College, KY and Ph.D. (1998) Analytical Chemistry, Miami University. Research interests include HPLC, CE, and inductively coupled plasma/mass spectrometry (ICP/MS).

C.S. Gilpin: Research Librarian, Ohio University Eastern. B.S. (1990) and MLS (1992) Library and Information Science, Kent State University St. Clairsville, OH. Research interests include scientific information retrieval, analytical, pharmaceutical and biomedical information, data management and search systems.

R.K. Gilpin: Professor and Dean of the College of Science and Mathematics, Wright State University, Dayton, OH. B.S. (1969) Chemistry, Indiana State University; Ph.D. (1973) Analytical Chemistry, University of Arizona. Editorial Board Member, *Journal of Chromatographic Science*, and NIH Technology Transfer SSS Member. Research interests include the development of chemically modified surfaces for chromatography, organometallic surface reactions, separation of pharmaceutically and biomedically active compounds, NMR and IR studies of modified surfaces and other ordered systems, and characterization of materials.

D.R. Goodlett: Research Scientist, Department of Molecular Biotechnology, University of Washington, Seattle, WA. B.S. (1982) and M.S. (1988) Auburn University, AL, Ph.D. (1991) Biochemistry, North Carolina State University, and NORCUS Postdoctoral Fellow (1991–1992) Battelle-Memorial Institute, Columbus, OH. Research interests include protein structure–function relationships at the molecular level, development of methods to study signal transduction pathways through mapping of protein phosphorylation sites, and characterization of proteomes both qualitatively and quantitatively.

C.J. Hann: Senior Scientist, Research Division, Solutia, St. Louis, MO. B.S. (1977) Chemistry, Bradley University, IN, Ph.D. (1988) Analytical Chemistry, Kent State University, OH. Research interests include optimization of separation methodology, development of quality assurance procedures for pharmaceutically and biomedically active leads, and process control analysis.

R. Herráez-Hernández: Instructor, Analytical Chemistry, University of Valencia, Valencia, Spain. B.S. (1989) Chemistry, and Ph.D. (1992) Chemistry, University of Valencia. Research interests include the analysis of organic compounds, especially by chromatographic techniques, including multi-dimensional procedures and chiral analysis.

C.E. Kibbey: Group Leader, Separations Group, Analytical Research, Parke-Davis Pharmaceutical Research Ann Arbor, MI. Ph.D. (1993) Analytical Chemistry, University of Michigan. Research interests include the application of HPLC and CE methods development to support drug discovery, characterization of combinatorial libraries for identity and purity, preparative purification of important lead compounds, measurement of physicochemical properties and laboratory automation.

C. Molins-Legua: Instructor of Chemistry, University of Valencia, Valencia, Spain. B.S. (1989) Biochemistry, Ph.D. (1994) Analytical Chemistry, University of Valencia, and FPI Postdoctoral Fellow (1995–1997) R.I.V.M. Research interests include analysis of organic compounds in various matrixes, especially by chromatographic procedures.

Y. Machida: Research Scientist, Analytical Chemistry Department, Product & Technology Development Laboratory, Tanabe Seiyaku Co. Ltd., Osaka, Japan. B.S. (1989), M.S. (1991), and Ph.D. (1998) Pharmacy, Nagoya City University. Research interests include synthesis and characterization of novel chiral stationary phases for drug analysis.

H. Nishi: Research Scientist, Analytical Chemistry Department, Product & Technology Development Laboratory, Tanabe Seiyaku Co. Ltd., Osaka, Japan. B.Tech. (1980), M.Tech. (1982), and Ph.D. (1991) Pharmacy, Kyoto University; Young Pharmaceutical Scientist Award from the Kinki Division of the Pharmaceutical Society of Japan in 1994 and a member of the Editorial

Board of the *Journal of Chromatography A*. Research interests include the separation of enantiomeric drugs by HPLC and CE.

N. North: Team Leader for new automation technology, SmithKline Beecham Pharmaceuticals, UK. Research interests include automation using robotics to develop noninvasive methods for process control.

L.A. Pachla: B.S. (1970) Chemistry, Lawrence Technological University Michigan; Ph.D. (1975) Analytical Chemistry, Purdue University, Indiana. Past member of the Editorial Board of Biomedical Chromatography. Research interests include robotics, electroanalytical chemistry, chromatography, pharmacokinetics, drug metabolism, and bioanalytical chemistry of proteinaceous drugs.

N.G. Pinto: Professor of Chemical Engineering, University of Cincinnati, Cincinnati, OH. B.Tech. (1980) Indian Institute of Technology, M.S. (1982) and Ph.D. (1985) Chemical Engineering, Pennsylvania State University. Research interests include: adsorption and ion exchange chromatography, bioseparations, and the development of novel microsensors.

A. Sevillano-Cabeza: Instructor of Analytical Chemistry, University of Valencia, Valencia, Spain. B.S. Chemistry (1976) and Ph.D. (1983) University of Valencia. Research interests include kinetic analysis and the analysis of organic compounds in various matrixes, especially by chromatographic procedures.

S.A. Tomellini: Associate Professor, Department of Chemistry, University of New Hampshire, Durham, NH. B.S. (1979) University of Rhode Island, Ph.D. (1985) Rutgers University, Visiting Scientist 1991–1992, Laser Biophysics and Spectroscopy Section, Laboratory of Chemical Physics, NIH. Research interests include expert systems in analytical chemistry, FTIR, HPLC, CE, chemical separations, and IR spectroscopy of lipids.

3 CONCLUSION AND FUTURE TRENDS

From the chemical/biological standpoint the more recent introduction of genetic and combinational methods has resulted in and will continue to result in the production of more novel compounds in shorter periods of time. This, in many respects, is like *The Tale of Two Cities*, in that they represent ''the best of times'' for synthetic chemists, since many potential leads can be produced quickly, and ''the worst of times'' for their analytical counterparts, in that large numbers of samples will need to be assayed using relatively small quantities of materials. Separation techniques like microcolumn HPLC and CE, coupled with highly specific and sensitive detection methods like electrospray ionization mass spectrometry (ESIMS), will continue to become more and more important. Likewise, improvements in hardware are making ESIMS more reliable and user friendly, and miniaturization and probe design are making other highly specific techniques like NMR more feasible as on-line detection methods.

ABBREVIATIONS AND ACRONYMS

CE	Capillary Electrophoresis
ESIMS	Electrospray Ionization Mass Spectrometry
GC	Gas Chromatography
HPCE	High-performance Capillary Electrophoresis
HPLC	High-performance Liquid Chromatography
ICP/MS	Inductively Coupled Plasma/Mass Spectrometry
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PAH	Polycyclic Aromatic Hydrocarbon

Alkaloids, Pharmaceutical Analysis of

R.K. Gilpin
Wright State University, Dayton, USA

C.J. Hann
Solutia Inc., St. Louis, USA

1 Introduction **1**
1.1 General Information 1
1.2 Common Properties 2
1.3 Trends in Analytical Methodology 2
2 General Information and Analytical Methods **3**
2.1 Cinchona Alkaloids 3
2.2 Ergot Alkaloids 4
2.3 Opium Alkaloids 5
2.4 Rauwolfia Alkaloids 7
2.5 Tropane Alkaloids 8
2.6 Vinca Alkaloids 9
2.7 Xanthine Alkaloids 11
2.8 Miscellaneous Alkaloids 11
Abbreviations and Acronyms **12**
Related Articles **12**
References **12**

Although alkaloids are naturally occurring bases, additional generalizations are difficult because they include a wide range of structurally dissimilar compounds. They vary greatly in their chemical and physical properties as well as in their distribution in nature. In some cases certain alkaloids are associated with only a single species of plant, whereas others are more widely distributed between biological groupings of plants, and yet others are found in a wide range of unrelated plants. Likewise, the concentration of a particular compound may be highly localized within a given region of one plant and found predominantly in a different region of another plant. In some cases the levels of an alkaloid may be relatively high and its isolation as a natural product may be economically feasible, whereas in other cases the levels may be extremely low and less commercially desirable. The medicinal use of alkaloids in the form of crude plant extracts has been known for several thousand years and today there are hundreds that have been isolated and characterized. However, very few of them are accepted therapeutically and many of these fit into the broad categories of cinchona, ergot, opium, rauwolfia, tropane, vinca and xanthine alkaloids.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 General Information

As a major class of compounds, alkaloids are naturally occurring bases with a wide range of structures, chemical and physical properties, and pharmacological activities. The use of alkaloids as medical agents in the form of crude plant powders and extracts predates the modern pharmaceutical industry by almost four millennia and in some instances prior to modern science, mass poisonings have resulted from their inadvertent usage. This was especially prevalent in Europe during the Middle Ages.

Historically, a primary source of many alkaloids has been flowering plants, although they may be found throughout nature, such as the ergot alkaloids in the grain fungus *Claviceps purpurea*. The first crude alkaloid extract to be studied chemically was opium, which is derived from the latex of the poppy *Papaver somniferum*.[1] Although the initial characterization work on opium that led to the isolation of morphine was carried out in the early 1800s, its use as an analgesic and its narcotic properties had been known for centuries before this.

There are often complex relationships between the alkaloids and their occurrence in nature. In some cases, certain alkaloids such as morphine are associated with only a single species of plant, whereas others like *l*-hyoscyamine are more widely distributed between biological groupings of plants, and yet others (e.g. nicotine) are found in a wide range of unrelated plants.[2] Likewise, the concentration of a given compound may be highly localized within a given area of a particular plant, such as in its leaves, bark, or roots, and this same alkaloid, if present in a different plant, may be found predominantly in another region. In some cases the levels of an alkaloid may be relatively high and its isolation as a natural product may be economically feasible, whereas in other cases the levels may be extremely low and less commercially desirable. A more in-depth discussion of the above items may be found elsewhere.[1–3]

Although today there are several thousand alkaloids that have been isolated and identified structurally, this article will focus only on a relatively small number that are considered to be pharmaceutically more important in terms of their accepted therapeutic value. Some of these compounds are among the earliest alkaloids identified for their medicinal effects, such as morphine, brucine, caffeine, quinine, cinchonine and colchincine.

1.2 Common Properties

Although most alkaloids are crystalline colorless solids, some of the more complex conjugated compounds such as berberine may be colored or, like quinine, fluorescent. A common chemical feature of the alkaloids are that they are bases, which is the basis for many of the commonly used colorimetric methods and thin-layer chromatography (TLC) spray reagents. These are based on the reaction of either organic (e.g. ninhydrin) or inorganic (e.g. Mayer's and Dragendorff's) reagents with the alkaloid to form highly colored products. Beyond this, additional generalizations are more difficult. Structurally, the basic nitrogen or nitrogens (which vary greatly in number and basicity) may be found in a variety of structural environments and hence the equilibrium properties and hydrophobic characteristics vary dramatically between alkaloids.

Illustrated in Figures 1–7 are some of the significant structural differences between a number of the more common pharmaceutically important alkaloids. Nicotine, which is distributed throughout the plant kingdom, is a relatively small molecule with two heterocyclic nitrogens. Similarly caffeine, another widely distributed small molecule, contains four heterocyclic nitrogens but it also contains two additional carbonyl functionalities. Because of their widespread presence and usage, there have been numerous methods developed for assaying both nicotine and caffeine in their natural states and in a host of different formulations and products (Gilpin and Pachla[4–7] and past biannual reviews in this series). Other alkaloids such as berberine, colchicine, and morphine contain only a single nitrogen but are either structurally more complex and/or contain other polar functionalities. For example, the two hydroxyl groups on morphine have a significant influence on its chromatographic properties. Because of these significantly different structural features, the overall ease of analyzing alkaloids and their pharmacological activity vary greatly.

As a result of their basicity, many alkaloids are thermally and photolytically labile, especially in the presence of oxygen. Common breakdown products are the corresponding N-oxides. In the case where the alkaloids contain other reactive groups they may undergo a variety of other reactions and rearrangements. Hydrolysis is often common. As such, during stability testing it is important to evaluate the pharmaceutical products for these likely decomposition candidates. In many cases some of these products, as well as other naturally occurring minor alkaloids and related impurities, may be extremely difficult or impossible to distinguish from the target analyte using simple nonseparation-based analytical procedures. Although details concerning the chemical reactivity of the various classes of alkaloids are presented below, more extensive treatments of this topic may be found elsewhere.[1–3] Commonly, alkaloids, like other organic bases, are stabilized via conversion to their corresponding inorganic or organic salts (i.e. hydrochloride or citrate salts).

1.3 Trends in Analytical Methodology

For over two decades (Gilpin and Pachla[4–7] and past biannual reviews in this series) separation-based procedures have been, and continue to be, the most often used methods for assaying alkaloids and their formulated products. This is consistent with the same trends in other areas of pharmaceutical analysis and often is essential in developing stability-indicating and purity-profiling methods. During this time a variety of techniques have been employed, ranging from simple screening procedures based on the use of an initial thin-layer separation (Table 1[8,9]) in combination with a colorimetric spray reagent (e.g. iodoplatinate, Dragendorff–Munier & Macheboeuf, iodine–potassium iodide reagents[10]) to more elaborate sample pretreatment and work-up procedures in combination with either an isocratic or a gradient elution high-performance liquid chromatography (HPLC) separation. In the latter

Table 1 TLC separation of some common alkaloids according to retention factor (R_f) values

Compound	R_f values for A[a]	R_f values for B[a]	R_f values for C[a]
Ajmaline	0	0.1	0.5
Atropine	0.1	0.2	0.4
Brucine	0.2	0.2	0.4
Chinchonine	0.3	0.2	0.4
Cocaine	0.6	0.6	0.7
Codeine	0.3	0.2	0.4
Colchicine	0	0	0.5
Dihydrocodeine	0.3	0.2	0.4
Dihydromorphinone	0.1	0.1	0.2
Emetine	0.5	0.4	0.7
Ergotamine	0	0	0.2
Homatropine	0.2	0.2	0.4
Morphine	0	0	0.1
Narcotine	0.6	0.5	0.7
Papaverine	0.5	0.4	0.7
Pilocarpine	0.1	0.1	0.4
Quinidine	0.2	0.1	0.3
Quinine	0.2	0.1	0.2
Reserpine	0.5	0.2	0.7
Scopolamine	0.3	0.2	0.6
Serpentine	0	0	0.2
Strychnine	0.4	0.3	0.5
Thebaine	0.5	0.5	0.7
Yohimbine	0.4	0.2	0.6

[a] A = Benzene–ethylacetate–diethylamine (70:20:10); B = chloroform–cyclohexane–diethylamine (40:50:10); C = acetone–chloroform–diethylamine (40:50:10).

instance, assays based on reversed-phase conditions are the most commonly used methods and the most often used eluent additives are simple buffers that are added to control the protonation/deprotonation of the basic nitrogen(s) and hence their retention properties. An in-depth discussion of the influence of eluent pH on solute retention is considered in the article **Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures** in this publication.

For many of the alkaloids, one of the more commonly encountered problems in developing reliable reversed-phase assays is peak tailing. This problem often is exacerbated when more than one nitrogen is present in the alkaloid and/or the alkaloid contains other polar subsitutents. Peak tailing is the result of residual silanol groups that are present on the reversed-phase packing. Because silica is an amorphous material, the number and distribution of these groups change depending on the synthetic route that is employed to produce the silica,[11,12] which in turn dramatically influences the nature and performance of the reversed-phase packings.[13] To the practicing chromatographer, this problem manifests itself as manufacturer-to-manufacturer and batch-to-batch differences in column performance for a given stationary phase. Additionally, this problem is exacerbated for solutes that have polar functional groups that can interact strongly with residual silanols such as amines and heterocyclics, which are common structural features of alkaloids. In order to minimize the residual silanol problems, one of four approaches are generally used: postreaction end-capping; preparation of sterically blocking phases; electronic manipulation of the attached surface groups; and the use of mobile phase additives. The first three of these approaches are controlled by the manufacturer and are important considerations when purchasing a column, especially for strongly interacting solutes like some of the alkaloids. Although the performance of commercially available bonded phases has improved dramatically, there are still many reversed-phase applications where residual silanol activity leads to unacceptable chromatograms in terms of severely tailing peaks, and as columns age the problem of exposed silanol groups increases even for high-performance bonded phases. Many of these unwanted effects can be eliminated or at least minimized through the use of secondary mobile phase additives. Such problems usually can be addressed by the addition of compounds to the eluent that dynamically modify the surface by a competitive sorption mechanism and hence act to suppress undesirable interactions that can arise between basic solutes and residual silanols. The agents used to do this are strongly sorbing compounds (i.e. molecules that contain a polar head group and a nonpolar tail) that do not interfere with detection. The most commonly used compounds to mask silanol activity and hence to improve peak symmetry are alkylamines,[14] however, in a few cases other compounds (e.g. perfluororalkyl surfactants) have been employed.[15] Alkylamines also are used to enhance the performance of normal-phase separations both in terms of HPLC procedures and in terms of routine TLC screening methods (Table 1).

2 GENERAL INFORMATION AND ANALYTICAL METHODS

2.1 Cinchona Alkaloids

The medicinal value of this group of alkaloids has been known since the 17th century, when crude extracts from cinchona bark, a plant species indigenous to the Andes, were first used for the treatment of malaria.[1,2] Although there have been more than two dozen cinchona alkaloids that have been isolated and identified, four of the pharmaceutically more important compounds are quinine, quinidine, cinchonidine and cinchonine. The structures of these are given in Figure 1. They are made up of two parts: a quinoline nucleus and a quinuclidine moiety. Of these, quinine and quinidine are the primary alkaloids of various species of *Cinchona* and *Remijia* and are present at levels of 1–4% and 0.3–3%, respectively.[16]

Upon oxidation, quinine, quinidine, cinchonidine and cinchonine are converted to the corresponding ketones and they undergo acetylation to form O-acetyl derivatives that reconvert to the starting material on hydrolysis. The vinylic group is susceptible to acid attack and rearrangements. In the case of quinine and quinidine, this tendency is greater than it is for the methoxy group. Greater details

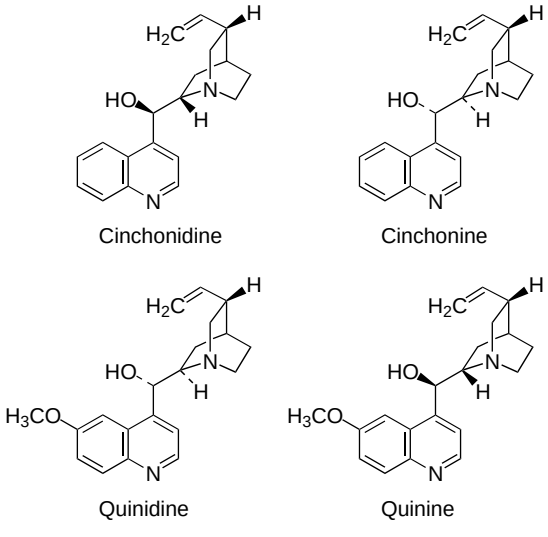

Figure 1 Common cinchona alkaloids.

concerning the chemical reactivity and related properties of these compounds may be found elsewhere.[1,2]

Typically, the cinchona alkaloids are white solids that form sparingly water-soluble mono-salts or highly water-soluble bis-salts. One of the more distinctive spectral features is their fluorescence in acidic media, which has been used in their direct spectrofluorimetric determination as well as by HPLC in combination with fluorimetric detection. Each of these alkaloids has two sites of protonation with respective pK_1 and pK_2 values in the 5.1–5.8 and 9.7–10.0 ranges.[16] When assayed using reversed-phase HPLC conditions, these are the structural features that must be appropriately controlled in order to obtain optimum separation performance.

A representative listing of some of the many methods published for the more common cinchona alkaloids is presented in Table 2.[17–40]

Table 2 Analytical procedures for cinchona alkaloids

Analyte	Technique	Refs.
(+)- and (−)-Cinchonine	LC normal-phase conditions	17
Bisbenzylisoquinoline	LC reversed-phase conditions	18
Cinchoncaine · HCl and 2-hydroxyquinoline-4-carboxylic acid diethylaminoamide	LC and first-derivative spectroscopy	19
Isoquinolines	LC reversed-phase conditions	20, 21
Quinidine and its dihydroxy and dimethoxy derivatives	LC and diode array detection of cinchona bark extracts	22
Quinidine and quinine	FIA using a chemi-luminescence reaction	23
Quinidine and quinine	AA as metal complexes	24, 25
Quinidine	Electrochemical and sensor	26
Quinidine and quinine	Isotachophoresis	27
Quinidine and quinine	Spectrophotometric	28, 29
Quinidine	Spectrofluorimetric	30
Quinine	FIA	31
Quinine	Electrochemical and sensor	32
General	LC reversed-phase conditions	33–36
General	LC reversed-phase conditions/thermospray MS	37, 38
General	Spectrophotometric	39
General	Titrimetry	40

AA, Atomic absorption; FIA, flow injection analysis; LC, liquid chromatography; MS, mass spectrometry.

2.2 Ergot Alkaloids

Historically, the medical value of some of the ergot alkaloids has been known for over 3000 years. They are the oldest known mycotoxins and are found in the *Claviceps purpurea*, a filamentous fungus that grows on rye and other gramineaceous crop plants. During the Middle Ages in Europe, ergot poisoning through their vasconstriction and/or hallucination actions was a common occurrence. In a single epidemic in AD 944 it has been reported that about 20 000 people in France died from ingesting ergot-infested flour.[1]

A common structural feature of many of the ergot alkaloids is their tetracyclic ergoline nucleus, as illustrated

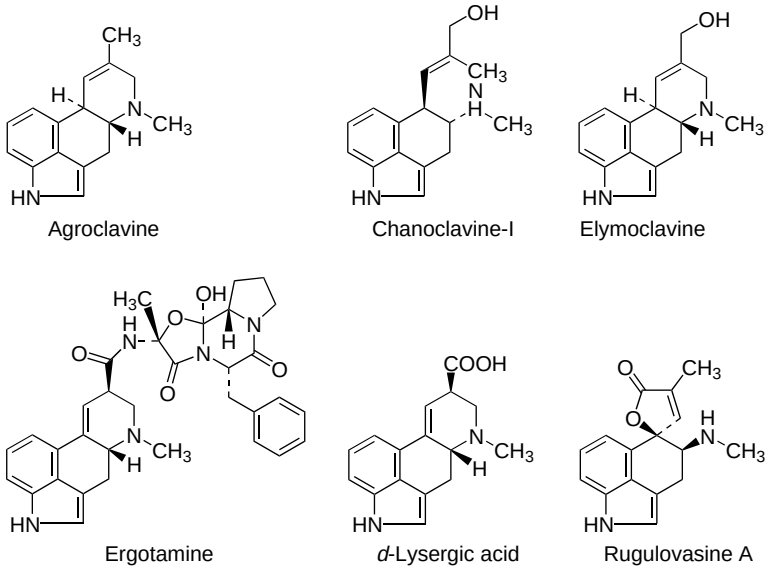

Figure 2 Common ergot alkaloids.

in Figure 2. Although all contain an indole structure, some of the clavine compounds, for example, chanoclavine-I and -II and rugulovasine A and B, may contain fewer rings and some of the more complex compounds, for example, the peptide alkaloid ergotamine, contain additional ring systems. The ergot alkaloids can be divided into four subgroups: the clavine alkaloids; the lysergic acid derivatives; the lysergic acid amides; and the ergot peptide alkaloids. Some of the more common compounds are chanoclavine-I (a precursor of agroclavine and elymoclavine), ergonovine, ergotamine and *d*-lysergic acid.

The most commercially important ergot compounds belong to the peptide subgroup. Typically these compounds must be protected from air oxidation, light, and heat. They hydrolyze to form lysergic acid, proline, a second amino acid, an α-keto acid and ammonia. In the case of ergotamine, a compound used for its antimigraine properties, the hydrolysis products are lysergic acid, proline, L-phenylalanine, pyruvic acid and ammonia.[1] The ergot alkaloids form colored products with sulfuric acid and a characteristic blue product with *p*-dimethylaminobenzaldehyde. In many cases a double bond is present at either the 8,9-position (e.g. agroclavine, chanoclavine-I and -II, and paspalic acid) or the 9,10-position (e.g. ergotamine, penniclavine, setoclavine, and *d*-lysergic acid) (Figure 2). These differences can be distinguished in their respective ultraviolet (UV) spectra where λ_{max} is at 284 nm for the indole structure (i.e. the $\Delta^{8,9}$ compounds) and at 318 nm for the 4-vinyl indole structure (i.e. the $\Delta^{9,10}$ compounds). This difference in UV properties between the two types of structures for the $\Delta^{8,9}$ and $\Delta^{9,10}$ alkaloids can be used in combination with variable or dual-wavelength detection to impart additional specificity to HPLC-based methods.

A representative listing of some of the methods published for the more common ergot alkaloids is presented in Table 3.[41–48]

2.3 Opium Alkaloids

The opium alkaloids have been studied more than any other group.[16] They are derived from the latex of a single species of the poppy *Papaver somniferum* and consist of several closely related compounds, including codeine, morphine, neopine, oripavine, and thebaine. Of these, morphine is the most abundant. A second related group of morphinandienone bases include sinomenine and hasubanonine, metaphenine, and protometaphenine. These latter alkaloids are found in Japanese *Sinomenium* and *Stephania* plants.[1] The structural difference between these two groups of compounds is shown in Figure 3.

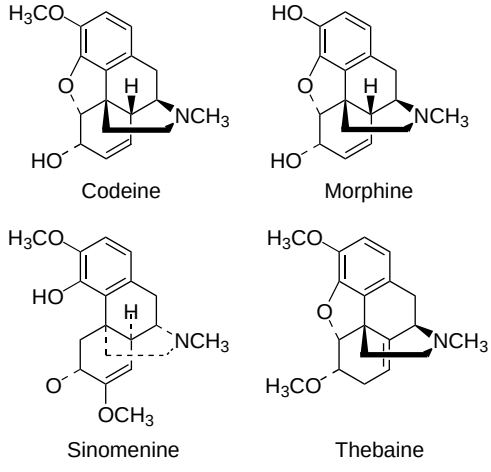

Figure 3 Common opium alkaloids.

Table 3 Analytical procedures for ergot alkaloids

Analyte	Technique	Ref.
Ergonotamine maleate and tartrate	FIA ampometric detection with Kel-F graphite composite electrode	41
Ergonotamine maleate and tartrate	LC fluorescence detection	42
Ergot epimers	CE	43
General	MS	44
General	NMR	45
General	TLC	46
General	CZE enantiomeric separation using cyclodextrins	47
General	LC reversed-phase conditions with eluent additives	48

CE, Capillary electrophoresis; NMR, nuclear magnetic resonance; CZE, capillary zone electrophoresis.

The first compound to be isolated in pure form from crude opium extracts was morphine, by Serturner in 1805, although 2 years prior to this Derosne had reported the separation of a mixture of morphine and noscapine. The presence of these compounds occurs in nature at levels of 4–21% and 4–8%, respectively.[1] Subsequently, codeine was isolated in 1833. This latter compound also can be produced easily via O-methylation of the phenolic group in morphine. Likewise, codeine can be oxidized at this same position to form the corresponding ketone, codeinone, which also results from the acid hydrolysis of thebaine.[3] In general the morphinandienones can undergo two types of acid-catalyzed rearrangements, forming either aporpines or dibenzazonines. The synthesis and various reaction pathways of this group of compounds are well established.[1–3]

Typically the morphinandienones have UV maxima at 235–240 and 275–280 nm. However, for sinomenine, where the double bond at the 4,5-position is missing, the λ_{max} is at 232 and 265 nm. The mass spectrometric,

Table 4 Analytical procedures for opium alkaloids

Analyte	Technique	Refs.
6-Acetylmorphine, diamorphine and morphine	LC to study hydrolysis of dimorphine	49
Apomorphine	LC reversed-phase conditions with C_{18} column	50
Apomorphine	Analytical profile	51
Hydromorphone and morphine	LC	52
Codeine	LC reversed-phase conditions with C_{18} column	53
Codeine	FIA using spectrofluorimetric detection	54
Codeine	LC reversed-phase conditons	55
Codeine	Spectrophotometric	56, 57
Codeine	Isotachophoresis	58
Codeine and byproducts	CE analytes in Kodynal, Ipecarin, Spasmoveralgin, and Alganon formulations	59
Codeine, morphine, papaverine and thebaine	TLC using spectrodensitometry	60
Codeine, morphine, noscarpine, papaverine and thebaine	LC reversed-phase gradient conditons using a base-deactivated C_{18} with 1-heptanesulfonic acid as the eluent modifier	61
Codeine and related alkaloids	LC reversed-phase conditons	62–64
Codeine and related alkaloids	TLC/HPTLC/OPLC	65
Codeine and related alkaloids	Differential pulse polarography	66
Ethylmorphine	LC reversed-phase conditions with C_{18} column	67
Hydrocordone	IR chromatographic isolation/IR identification	68
Morphine	Review of the use of biosensors	69
Morphine	LC reversed-phase gradient conditions using a phenyl column to assay ipecac formulations	70
Morphine	FIA using chemiluminescence detection	71, 72
Morphine	LC reversed-phase conditions	73
Morphine	NIR reflectance	74
Noscapine	TLC/HPTLC/OPLC	75
Noscapine	LC	76
Papaverine	AA by indirect measurement	77
Papaverine	LC reversed-phase conditons with C_{18} column	78
Papaverine	Electrochemical	79–81
Papaverine	TLC using spectrodensitometry detection	82
Papaverine	Colorimetric or UV	83–86
General	CE using guest–host, nonaqueous, or micellar conditions	87–89
General	LC review	90
General	LC reversed-phase conditions	91–97
General	SFC using a packed column	98

HPTLC, High-performance thin-layer chromatography; IR, infrared; NIR, near-infrared; OPLC, overpressured layer chromatography; SFC, supercritical fluid chromatography.

spectroscopic and chromatographic properties of the opium alkaloids have been studied by numerous investigators and they are generally easily analyzed via a variety of techniques. A representative listing of some of the many methods published for the more common opium alkaloids is presented in Table 4.[49–98]

2.4 Rauwolfia Alkaloids

Although approximately 150 species of plants belong to the *Rauwolfia* genus, *R. serpentina*, a plant found in India, is the most important member. However, other species such as *R. vomitoria* and *R. tetraphylla*, which are found respectively in Africa and Central America, have become alternative sources. Likewise, synthetically produced reserpine now competes favorable in price with the natural product.[1] The medicinal use of extracts from this family of plants, like the ergot alkaloids, has been known for about 3000 years but the major active compound reserpine, which occurs at about the 1% level, was not isolated and identified until 1952. Other minor (i.e. at about the 0.1% range) alkaloids that belong to this class are ajmalicine, ajmaline, rescinnamine, reserpiline, and yohimbine (see Figure 4).

Like the ergot alkaloids, a common structural feature of the rauwolfia alkaloids is the indole nucleus, as illustrated in Figure 4. Alkaline hydrolysis of reserpine, the principal alkaloid of this class, produces reserpic acid, 3,4,5-trimethoxybenzoic acid, and methanol. Vigorous oxidation of the resulting reserpic acid leads to loss of the indole structure via production of 4-methoxy-*N*-oxalylanthranilic acid.[2]

Table 5 Analytical procedures for rauwolfia alkaloids

Analyte	Technique	Refs.
Ajmaline	Radioimmunoassay	99
3,4-Dihydroreserpine, isoreserpine, reserpine and 3,4,5,6-tetra-hydroserpine	LC	100
3,4-Dihydroreserpine, isoreserpine, reserpine and 3,4,5,6-tetra-hydroserpine	Electrochemical by differential pulse polarography	101
Rescinnamine	LC normal-phase conditions and fluorescence detection	102
Reserpiline and reserpine	Spectrofluorimetric, differences in fluorescence excitation and emission spectra used	103
Reserpine	Colorimetric	104
Reserpine	LC	105–107
Reserpine	Electrochemical	108
Reserpine	HPLC/TLC	109
Reserpine	Spectrofluorimetric	110
Yohimbine	Review of the interaction of analyte with microcrystalline vs carboxymethylcellulose	111
Yohimbine	Spectrofluorimetric based on the oxidation of the analyte with Ce(IV)	112
General	LC reversed-phase conditions in combination with thermospray MS	113, 114
General	GC/MS	115

GC/MS, Gas chromatography/mass spectrometry.

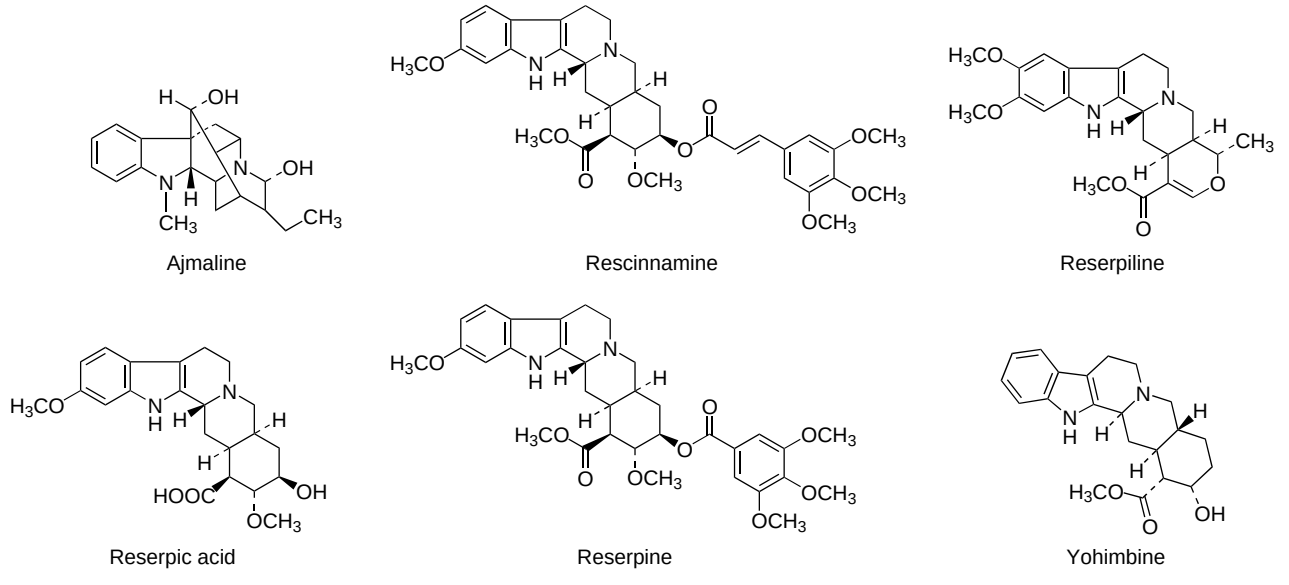

Figure 4 Common rauwolfia alkaloids.

A representative listing of some of the methods published for the more common rauwolfia alkaloids is presented in Table 5.[99–115]

2.5 Tropane Alkaloids

Tropane alkaloids occur in a variety of *Erythroxylaceae*, *Solanaceae* and *Convolvulaceae* plants, which include *Atropa belladona*, *Datura stramonium*, *Erythroxylon coca*, and *Hyosyamus niger*. The most common compounds in this group are cocaine, *l*-hyoscyamine and its racemized form atropine, hyoscine, scopolamine, and meteloidine (Figure 5). The early use of the tropane alkaloids can be traced to the 16th century both in Europe and South America, where crude preparation of *Atropa belladonna* and dried coca leaves were used, respectively, as medical aids. However, their isolation of the active alkaloids was not until the 19th century. Geiger first prepared *l*-hyoscyamine in 1883 and Wohler prepared cocaine in 1862.

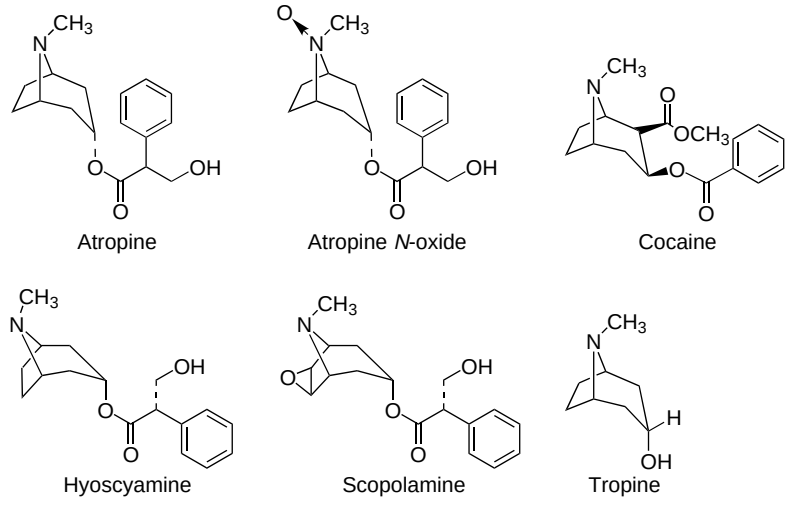

Figure 5 Common tropane alkaloids.

Table 6 Analytical procedures for tropane alkaloids

Analyte	Technique	Refs.
Atropine	LC reversed-phase conditions using cyano column	116
Atropine	Electrochemical and sensor	117
Atropine	LC ion-pairing reagent/column switching made it possible to determine analyte in complex preparations of other gastrointestinal drugs	118
Atropine	LC chiral reagents added to mobile phase to resolve isomers	119, 120
Atropine	LC ion-pairing reagent	121
Atropine	LC reversed-phase conditions	122
Atropine analogs	LC reversed-phase conditions	123
Atropine, cocaine, homatropine and scopolamine	LC enantiometric separation using β-cyclodextrin-bonded phase	124
Atropine, homatropine and scopolamine	CE	125
Coca leaves	Reversed-phase LC/GC comparison: GC better for resolving cocaine and related products; reversed-phase LC faster and more convenient	126
Cocaine	Review of the use of biosensors	69
Cocaine	AA by indirect measurement	77
Cocaine	Reversed-phase LC esterified with optically pure 2-octanol prior to RPLC	127
Scopolamine	Electrochemical and sensor	128
Scopolamine	LC comparison of interaction of analyte with microcrystalline cellulose vs sodium carboxymethylcellulose	129
General	TLC/HPTLC/OPLC	130
General	LC reversed-phase conditions	131–134
General	LC enantiometric separation using β-cyclodextrin-bonded phase	135

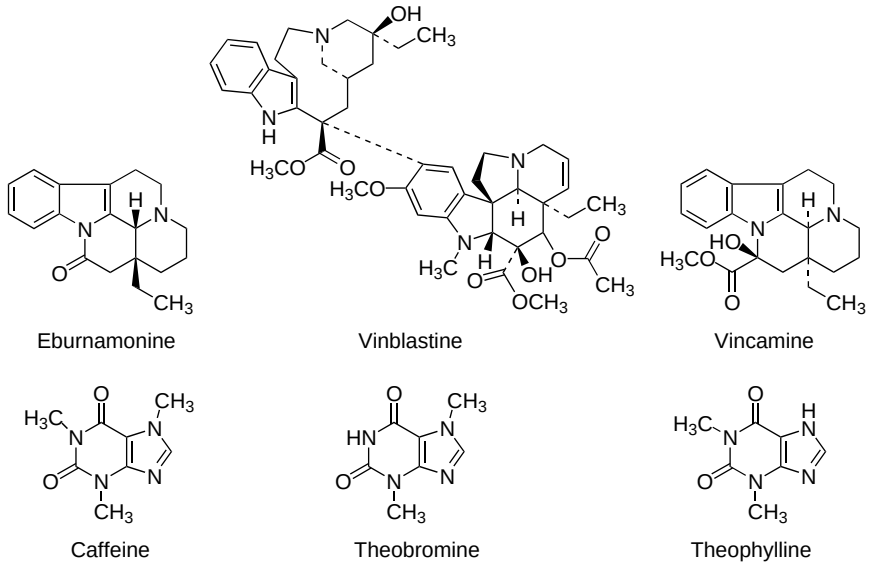

Figure 6 Common vinca and xanthine alkaloids.

A significant structural feature of the tropane alkaloids are that they are esters of an organic acid that is attached at the 3-position in either an α- or β-configuration to the central tropane structure. Very gentle neutral hydrolysis of these compounds produces tropine and the corresponding organic acid. In the case of *l*-hyoscyamine, the most widespread alkaloid of *Solanaceae* plants, the hydrolysis products are tropine and *s*-(−)-tropic acid. In the presence of more vigorous thermal and acidic conditions, tropine can undergo additional reactions to form a variety of products, and oxidation of tropine leads to both equatorial and axial N-oxide isomers. This is illustrated in Figure 5 for atropine.

A representative listing of some of the many methods published for the more common tropane alkaloids is presented in Table 6.[116–135]

2.6 Vinca Alkaloids

There are approximately 100 alkaloids that are present in six species of the genus *Vinca*. This group of plants are found throughout western Asia and the Mediterranean region of Europe. The most important species are *Vinca major* and *Vinca minor*, and the more common compounds include apovincamine, eburnamenine, hervine, reserpinine, sarpagine, (−)-tabersonine, vincadine, vincamajine, and vincamine. Of these, vincamine is the most important alkaloid and may be found at levels up to 2–3%. However, at least half of the vincamine currently used is partially synthesized from tabersonine, and numerous derivatives of vincamine have been prepared.[1]

A typical feature of the *Vinca* alkaloids is the eburna nucleus, which results in a characteristic mass spectrometric pattern. The major fragmentation pathway occurs via Diels–Alder reaction in the C-ring. The radicals produced can undergo two reaction schemes resulting in two major fragments with *m*/*z* differences of 41. For example, in the case of apovincamine, ions are observed at 308 and 267, and for eburnamenine, ions are observed at 249 and 208. However, besides the eburna nucleus, the remaining structural features vary widely between the different alkaloids, as illustrated by eburnamonine, vinblastine and vincamine in Figure 6.

Table 7 Analytical procedures for vinca alkaloids

Analyte	Technique	Ref.
Catharanthine, vinblastine, vincristine and vindoline	LC	136
Vinblastine and degradation products	LC reversed-phase conditions/MS field desorption and chemical ionization	137
Vinblastine, vincristine and indole impurities	LC	138
Vinblastine sulfate	Reviews 131 references and deals with the synthesis, physical properties, stability, and analytical methodology	139
Vinblastine sulfate	LC reversed-phase conditions	140
Vinblastine sulfate	LC reversed-phase conditions and column comparison of α-acid glycoprotein versus human serum albumin	141

Table 8 Analytical procedures for xanthine alkaloids

Analyte	Technique	Refs.
Caffeine	LC reversed-phase conditions with C_{18} column	53
Caffeine	AA or voltametric complex formed with molybdophosphate	142
Caffeine	Oxidimetric titration	143
Caffeine	UV derivative spectrometry	144
Caffeine and Analogs	LC and micellar electrokinetic capillary chromatography	145
Caffeine and Theophylline	Luminescence as function of pH and presence of a heavy atom such as iodine	146
8-Chlorotheophylline	Electrochemical	147–150
Diprophylline	Colorimetric	151
Etoffyline and theophylline	Reversed-phase LC C_{18} column	152
Theophylline	LC reversed-phase conditions with C_{18} column	153–155
Theophylline	LC normal-phase conditions	156
Theophylline	Electrochemical	157
Theophylline	NIR to study dissolution rate, film coating thickness and hardness	158
Theophylline	CE	159
Theophylline	UV and colorimetric	160
Theophylline	Biosensor nafion film containing theophylline oxidase and a ferricytochrome *c* cofactor	161
Theophylline	LC dansyl chloride derivative to enhance detection	162
Theophylline	Colorimetric after treatment with 4-nitroaniline	163
Theophylline	Fluoresence at 615 nm and excitation at 300 nm after treatment with europium(III)	164
Theophylline	Potentiometric	165
Theophylline	Stopped-flow fluorimetry based on measuring kinetics of reaction with Ce(IV)	166
Theophylline	TLC	167
Theophylline	Chromatography	168
Theophylline	Luminescence at room temperature	169, 170
Theophylline analogs	LC reversed-phase conditions	171
General	HPTLC	172, 173
General	LC review	174–178
General	IR	179
General	Luminescence	180

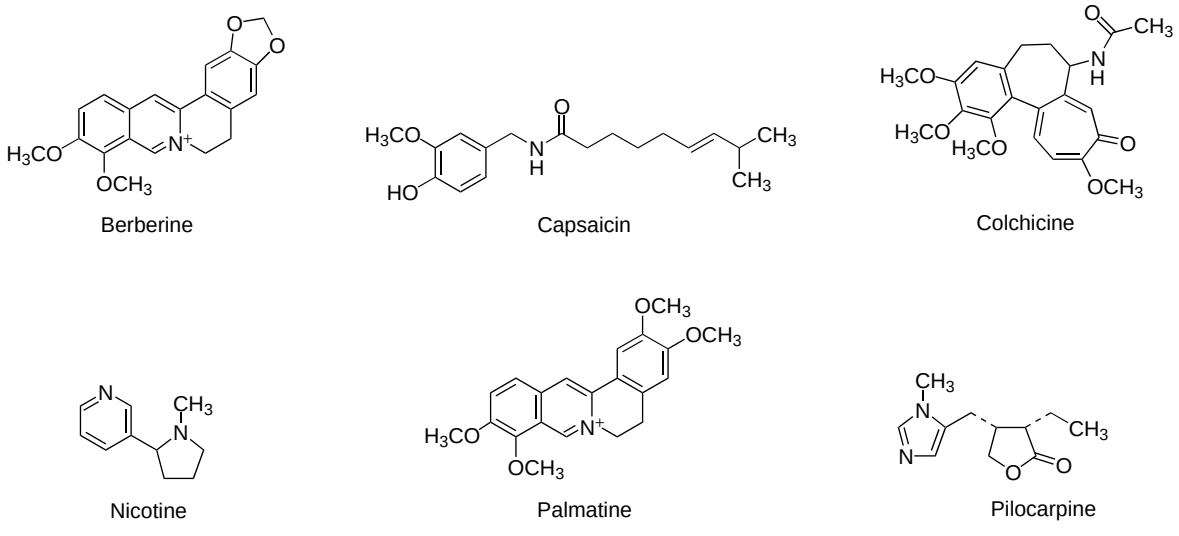

Figure 7 Common miscellaneous alkaloids.

Table 9 Analytical procedures for miscellaneous alkaloids

Analyte	Technique	Refs.
9-Acridone	LC	181
Campothecin	LC to study photodecomposition	182
Berberine	LC	183
Berberine	Electrochemical	184, 185
Berberine	Spectrofluorimetric	186, 187
Berberine	FIA/spectrofluorimetric	188
Berberine	Spectrophotometric	180–191
Capsaicin analogs	LC	192
Catharanthine	Radioimmunoassay	193
Chelidonine protopine	TLC	194
Chelidonium majus alkaloids	LC	195
Colchine	stripping voltametry	196
Indole alkaloids from *Catharanthus roseus*	LC reversed-phase conditions and thermospray MS	197
Indole Alkaloids in suspension culture *Tabernaemontana divaricata*	LC reversed-phase conditions	198
Pilocarpine, degradation products and impurities	LC reversed-phase conditions	199–205
Pilocarpine	LC normal-phase conditions	206
Pilocarpine	LC β-cyclodextrin column	207
Pilocarpine	TLC	208
Pilocarpine	Spectrophotometric	209
Pilocarpine	AA by indirect measurement of mercury complex	210
Protoberberine	TLC/densitometric detection	211
Tetrahydrojatrorrhizing and tetrahydroprotoberberine	LC using a cellulose tris(phenylcarbamate) column	212
Ephedrae herba	LC	213
Senecio vulgaris	LC and NMR to study the interaction of analyte with microcrystalline vs carboxymethylcellulose	214
General	CE influence of structure on electrophoretic mobility	215
General	LC normal-phase conditions using a polyol-derivatized silica column	216
General	LC reversed-phase conditions using cross-linked cyclodextrin columns	217
General	Counter-current chromatography influence of pH and ion-pair formation	218
General	Potentiometric titration	219
General	Colorimetric with 2,6-dichlorophenolindophenol	220

This makes the prediction of chromatographic properties much more unlikely compared to alkaloids with more definable structural changes in terms of their effect on retention (e.g. the common opium alkaloids shown in Figure 3).

A representative listing of some of the methods published for the more common vinca alkaloids is presented in Table 7.[136–141]

2.7 Xanthine Alkaloids

The xanthine alkaloids are found throughout nature and the most common compounds in this group are caffeine, theophylline and theobromine. They share a number of pharmacological properties, including central nervous system, cardiac, respiratory stimulant, and smooth-muscle relaxant.

The central structural feature of the xanthine alkaloids is their purine nucleus (lower row of compounds in Figure 6). These three compounds are easily analyzed by a variety of methods, as summarized in Table 8, including many reversed-phase HPLC assays.[53,142–180]

2.8 Miscellaneous Alkaloids

There are a wide variety of other alkaloids that vary widely in terms of their source of origin, structure, and pharmacological activities. Because of space limitations it is not possible to consider each of these in a more in-depth discussion, but some of the more important are given in Figure 7 and methods for these as well as other alkaloids are given in Table 9.[181–223]

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
CE	Capillary Electrophoresis
CZE	Capillary Zone Electrophoresis
FIA	Flow Injection Analysis
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
OPLC	Overpressured Layer Chromatography
SFC	Supercritical Fluid Chromatography
TLC	Thin-layer Chromatography
UV	Ultraviolet

RELATED ARTICLES

Pharmaceuticals and Drugs **(Volume 8)**
Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

REFERENCES

1. G.A. Cordell, *Introduction to Alkaloids: a Biogenetic Approach*, John Wiley & Sons, New York, 1981.
2. J.S. Glasby, *Encyclopedia of Alkaloids*, Plenum Press, New York, Vols. 1 and 2, 1975; Vol. 3, 1978.
3. K.W. Bentley, *The Chemistry of Natural Products, The Alkaloid*, 2nd edition, Wiley–Interscience, New York, Vol. 1, 1966.
4. R.K. Gilpin, L.A. Pachla, 'Pharmaceuticals and Related Drugs', *Anal Chem.*, **65**(12), 117R–132R (1993).
5. R.K. Gilpin, L.A. Pachla, 'Pharmaceuticals and Related Drugs', *Anal Chem.*, **67**(12), 295R–313R (1995).
6. R.K. Gilpin, L.A. Pachla, 'Pharmaceuticals and Related Drugs', *Anal Chem.*, **69**(12), 145R–163R (1997).
7. R.K. Gilpin, L.A. Pachla, 'Pharmaceuticals and Related Drugs', *Anal Chem.*, **71**(12), 217R–233R (1999).
8. D. Waldi, K. Schnackerz, K.F. Munter, 'Systematic Analysis of Alkaloids on Thin-layer Plates', *J. Chromatogr.*, **6**, 61 (1961).
9. J. Sherma, F. Bernard, *Handbook of Thin-layer Chromatography*, 2nd edition, Marcel Dekker, New York, 1996.
10. J.C. Touchstone, *Practice of Thin Layer Chromatography*, 3rd edition, John Wiley & Sons, New York, 1992.
11. R.K. Iler, *The Chemistry of Silica*, John Wiley & Sons, New York, 1979.
12. K.K. Unger, *Porous Silica*, Elsevier, Amsterdam, 1979.
13. R.K. Gilpin, L. Wu, 'Use of the Reordering/Resolvation of Alkyl-modified Silica to Characterize the Microscopic Heterogeneity of Silica via Liquid Chromatography', *J. Chromatogr.*, **556**, 415–424 (1991).
14. R.K. Gilpin, S.S. Yang, G. Werner, 'A Recent Review of Secondary Mobile Phase Modifiers used to Enhance the Chromatographic Analysis of Pharmaceuticals', *J. Chromatogr. Sci.*, **26**, 388–400 (1988).
15. P. Varughese, M.E. Gangoda, R.K. Gilpin, 'Applications of Fluorinated Compounds as Phases and Additives in Chromatography and their Uses in Pharmaceutical Analysis', *J. Chromatogr. Sci.*, **26**, 401–405 (1988).
16. S. Budavari (ed.), *The Merck Index: an Encyclopedia of Chemicals, Drugs and Biologicals*, 12th edition, Merck & Co., Rathway, NJ, 1996.
17. A.N. Diaz, F.G. Sanchez, A.A. Gallardo, A.G. Pareja, 'HPLC Enantiomeric Resolution of (+)-Cinchonine and (−)-Cinchonidine with Diode-laser Polarimetric Detection', *Instrum. Sci. Technol.*, **24**(1), 47–56 (1996).
18. S.-W. Sun, S.-S. Lee, A.-C. Wu, C.-K. Chen, 'Determination of Bisbenzylisoquinoline Alkaloids by High-performance Liquid Chromatography', *J. Chromatogr. A*, **799**, 337–342 (1998).
19. A. El-Gindy, M.A. Korany, M.F. Bedair, 'First Derivative Spectrophotometric and High-performance Liquid Chromatographic Determination of Cinchocaine Hydrochloride in Presence of its Acid Degradation Product', *J. Pharm. Biomed. Anal.*, **17**, 1357–1370 (1998).
20. P. Pietta, P. Mauri, E. Manera, P. Ceva, 'Determination of Isoquinoline Alkaloids from Peumus Boldus by High-performance Liquid Chromatography', *J. Chromatogr.*, **457**, 442–445 (1988).
21. I. Valka, 'Methods for the Determination of Isoquinoline Alkaloids', *Chem. Usty*, **83**(7), 716–729 (1989).
22. P.C. Majumder, V.S. Giri, N. Banerji, 'Application of HPLC to Determine Quinidine in Presence of its Dihydro and Dimethoxy Derivatives in Cinchona Barks', *J. Inst. Chem. (India)*, **58**(4), 143–145 (1986).
23. I. Koukli, A.C. Calokerinos, 'Determination of Quinine and Quinidine by Continuous-flow Chemiluminescence', *Anal. Chim. Acta*, **236**(2), 463–468 (1990).
24. S. Tsurubou, T. Sakai, S. Kihara, M. Matsui, 'Liquid–Liquid Extraction of Cinchona Alkaloids by Using some Metal Complexes of Optically Pure Usnic Acids', *Anal. Chim. Acta*, **248**(2), 501–506 (1991).
25. M.M. Ayad, S.E. Khayyal, N.M. Farrag, 'Microdetermination of Cinchona Alkaloids by Atomic Absorption Spectroscopy', *Spectrochim. Acta, Part B*, **4013**(9), 1205–1209 (1985).
26. L. Nie, X. Zhang, S. Yao, 'Determination of Quinine in some Pharmaceutical Preparations Using a Ring-coated

Piezoelectric Sensor', *J. Pharm. Biomed. Anal.*, **10**(7), 529–533 (1992).
27. H. Klein, R. Teichmann, 'Determination of Cinchona Alkaloids in Pharmaceutical Preparations by Isotachophoresis (Ionophoresis)', *Pharm. Ztg.*, **132**, 1131–1135 (1987).
28. N.A. El-Sebakhy, A.A. Seif El-Din, M.A. Korany, 'First-derivative Spectrophotometric Determination of Quinine–Quinidine in Cinchona Liquid Extract', *J. Pharm. Belg.*, **41**(4), 222–225 (1986).
29. J. Soucek, E. Halarnek, R. Kysilka, 'Extraction Spectrophotometric Determination of Quinine in Dosage Forms', *Cesk. Farm.*, **35**(9), 388–391 (1986).
30. M. Mariaud, P. Dubois, P. Levillain, 'Determination of Cinchonine and Quinidine in Mixtures by Zero-crossing First-derivative Spectrofluorometry', *Analyst (London)*, **113**(6), 929–932 (1988).
31. M. Polasek, R. Karlicek, P. Solich, 'Quinine Determination in some Mass-produced Pharmaceuticals by Flow Injection Analysis (FIA) with Fluorimetric Detection', *Cesk. Farm.*, **36**(5), 201–206 (1987).
32. H. Cui, 'Atropinium–Scopolaminium Integrated Microconduits in a Potentiometric Analytical System', *Talanta*, **40**(9), 1445–1448 (1993).
33. J.L. Huang, D.J. Morgan, 'Simple Direct Injection High-performance Liquid Chromatographic Method to Determine Quinidine in Plasma', *J. Chromatogr., Biomed. Appl.*, **131**, 278–280 (1993).
34. P. Salvadori, D. Pini, C. Rosini, C. Bertucci, G. Uccello-Barretta, 'Chiral Discriminations with Cinchona Alkaloids', *Chirality*, **4**(1), 43–49 (1992).
35. A. Hermans-Lokkerbol, T. Van der Leer, R. Verpoorte, 'Reversed-phase High-performance Liquid Chromatographic Separation of some Indole and Quinoline Alkaloids from Cinchona', *J. Chromatogr.*, **479**(1), 39–51 (1989).
36. K. Kajiyarna, Y. Ohno, T. Ochiai, Y. Hiraga, K. Takahashi, 'Application of Contour Map Three-dimensional Displays of Ultraviolet Absorbance in High-performance Liquid Chromatography of Natural Drug Materials', *J. Chromatogr.*, **362**(1), 132–137 (1986).
37. S. Abdulrahman, M.E. Harrison, K.J. Welham, M.A. Baldwin, J.D. Phillipson, M.F. Roberts, 'High-performance Liquid Chromatographic–Mass Spectrometric Assay of High-value Compounds for Pharmaceutical Use from Plant Cell Tissue Culture: Cinchona Alkaloids', *J. Chromatogr.*, **562**(1/2), 713–721 (1991).
38. C. Giroud, T. Van der Leer, R. Van der Heijden, R. Verpoorte, C.E.M. Heeremans, W.M.A. Niessen, J. Van der Greef, 'Thermospray Liquid Chromatography/Mass Spectrometry (TSP LC/MS) Analysis of the Alkaloids from Cinchona In Vitro Cultures', *Neth. Planta Med.*, **57**(2), 142–148 (1991).
39. M.S. Karawya, A.M. Diab, N.Z. Szelern, 'A Micro-assay Method for Colchicine in Dosage Forms and in Biological Material', *J. Pharm. Sci.*, **25**(1–4), 161–166 (1984).
40. P.A. Yavich, L.I. Churadze, A.G. Sarabunovich, M.A. Mgebrishvili, 'Quantitative Determination of Colchicine', *Farmatsiya (Moscow)*, **35**(4), 64–66 (1986).
41. F. Belal, J.L. Anderson, 'Flow Injection Determination of Ergonovine Maleate with Amperometric Detection at the Kel-F-graphite Composite Electrode', *Talanta*, **33**(5), 448–450 (1986).
42. U.R. Cieri, 'Determination of Ergotamine Tartrate in Tablets by Liquid Chromatography with Fluorescence Detection', *J. Assoc. Off. Anal. Chem.*, **70**(3), 538–540 (1987).
43. K. Frach, G. Blaschke, 'Separation of Ergot Alkaloids and their Epimers and Determination in Sclerotia by Capillary Electrophoresis', *J. Chromatogr. A*, **808**, 247–252 (1998).
44. A.F. Casy, 'Mass Spectrometry as an Aid to the Identification of Ergots and Dihydroergots: Comparison of Hard and Soft Ionization Techniques', *J. Pharm. Biomed. Anal.*, **12**(1), 41–46 (1994).
45. A.F. Casy, 'Rapid Identification of Ergot Derivatives by ^{1}H-NMR Spectroscopy', *J. Pharm. Biomed. Anal.*, **12**(1), 27–40 (1994).
46. L. Botz, S. Nyiredy, O. Sticher, 'Separation of Ergot Alkaloids by HPTLC, OPLC, and Rotation Planar Chromatographic (RPC) Methods', *J. Planar Chromatogr.-Mod. TLC*, **3**(3/4), 193–195 (1990).
47. S. Fanali, M. Flieger, N. Steinerova, A. Nardi, 'Use of Cyclodextrins for the Enantioselective Separation of Ergot Alkaloids by Capillary Zone Electrophoresis', *Electrophoresis*, **13**(1/2), 39–43 (1992).
48. I.M. Jalal, S.I. Sa'sa, T.A. Yasin, 'Determination of Ergotamine Tartarate and Cyclizine Hydrochloride in Pharmaceutical Tablets by Reverse Phase HPLC', *Anal. Lett.*, **21**(9), 1561–1577 (1988).
49. D.A. Barrett, P.N. Shaw, 'A Stability-indicating HPLC Assay for Diamorphine in Aqueous Solution', *J. Liq. Chromatogr.*, **17**(17), 3727–3733 (1974).
50. M.J. Priston, G. Sewell, 'The Analysis of Apomorphine Formulations of Ambulatory Infusions', *J. Pharm. Sci.*, **1**(2), 91–94 (1995).
51. F.J. Muhtadi, M.S. Hifnawy, 'Analytical Profile of Apomorphine hydrochloride', *Anal. Profiles Drug Subst.*, **20**, 121–171 (1991).
52. T.G. Venkateshwaran, J.T. Stewart, 'HPLC Determination of Morphine–Hydromorphone–Bupivacaine and Morphine–Hydromorphone–Tetracaine Mixtures in 0.9% Sodium Chloride Injection', *J. Liq. Chromatogr.*, **18**(3), 565–578 (1995).
53. A. Huettner, H.G. Eigendorf, 'Simultaneous Determination of Propyphenazone, Caffeine and Codeine in Mixtures by Reverse-phase-HPLC', *Pharmazie*, **41**(1), 59 (1986).
54. N.W. Barnett, T.A. Bowser, R.D. Gerardi, B. Smith, 'Determination of Codeine in Process Streams Using Flow-injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **318**(3), 309–317 (1996).

55. I.N. Papadoyannis, B. Caddy, 'Rapid Analysis for Codeine by Reversed-phase High-performance Liquid Chromatography', *Anal. Lett.*, **19**(9/10), 1065–1081 (1986).
56. G.M. Greenway, A.W. Knight, P.J. Knight, 'Electrogenerated Chemiluminescent Determination of Codeine and Related Alkaloids and Pharmaceuticals with Tris(2,2′-bipyridine)ruthenium(II)', *Analyst (Cambridge, UK)*, **120**(10), 2549–2552 (1995).
57. G.A. Milovanovic, L. Trifkovic, T.J. Janjic, 'Kinetic Determination of Codeine in Pharmaceutical Preparations', *Mikrochim. Acta*, **III**(5/6), 287–293 (1986).
58. J. Pospichalova, V. Jokl, 'Optimization of the Simultaneous Isotachophoretic Determination of Ephedrine and Codeine in Pharmaceuticals', *Pharmazie*, **42**(1), 55–56 (1987).
59. M. Korman, J. Vindevogel, P. Sandra, 'Separation of Codeine and its Byproducts by Capillary Zone Electrophoresis as a Quality Control Tool in the Pharmaceutical Industry', *J. Chromatogr.*, **645**(2), 366–370 (1993).
60. N.R. Ayyangar, S.S. Biswas, A.S. Tambe, 'Separation of Opium Alkaloids by Thin-layer Chromatography Combined with Flame Ionization Detection Using the Peak Pyrolysis Method', *J. Chromatogr.*, **547**(1/2), 538–543 (1991).
61. N.R. Ayyangar, S.R. Bhide, 'Determination of the Five Major Opium Alkaloids by Reversed-phase High-performance Liquid Chromatography on a Base-deactivated Stationary Phase', *J. Chromatogr.*, **366**, 435 (1986).
62. L. Krenn, S. Glantschnig, U. Sorgner, 'Determination of the Five Major Opium Alkaloids by Reversed-phase High-performance Liquid Chromatography on a Base-deactivated Stationary Phase', *Chromatographia*, **47**(1/2), 21–24 (1998).
63. N.R. Ayyangar, S.R. Bhide, 'Separation of Eight Alkaloids and Meconic Acid and Quantitation of Five Principal Alkaloids in Gum Opium by Gradient Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **436**(3), 455–465 (1988).
64. P. Gomez-Serranillos, E. Carrebero, A. Villar, 'Analysis of Poppy Straw and Poppy Straw Concentrate by Reversed-phase High Performance Liquid Chromatography', *Phytochem. Anal.*, **5**(1), 15–18 (1994).
65. N. Bergisadi, Y. Ozsoy, 'Determination of Codeine and Dionin in Codeine–Dionin Combined Tablets', *Acta Pharm. Turc.*, **29**(3), 81–83 (1987).
66. G.E. Baiulescu, S.D. Popescu, 'Determination of Codeine, Dionin and Thebaine by Differential Pulse Polarography', *Anal. Lett.*, **19**(5/6), 587–596 (1986).
67. G. Achilli, G.P. Cellerino, G.V. Melzi d'Eril, F. Tagliaro, 'Determination of Illicit Drugs and Related Substances by High Performance Liquid Chromatography with an Electrochemical Coulometric-array Detector', *J. Chromatogr.*, **729**, 273–277 (1996).
68. D.B. Black, A.W. By, B.A. Lodge, 'Isolation and Identification of Hydrocodone in Narcotic Cough Syrups by High-performance Liquid Chromatography with Infrared Spectrometric Identification', *J. Chromatogr.*, **358**(2), 438–443 (1986).
69. G.G. Guilbault, R.D. Schmid, 'Biosensors for the Determination of Drug Substances', *Biotechnol. Appl. Biochem.*, **14**(2), 133–145 (1991).
70. C.-J. Lai, F.-S. Chen, C.-S. Chien, J.-H. Li, 'A Stability-indicating HPLC Method for the Assay of Morphine Content in Opium–Ipecac Tablets', *Chin. Pharm. J. (Taipei)*, **48**(4), 279–289 (1996).
71. N.W. Barnett, D.G. Rolfe, T.A. Bowser, T.W. Paton, 'Determination of Morphine in Process Streams Using Flow-injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **282**(3), 551–557 (1993).
72. R.W. Abbott, A. Townshend, R. Gill, 'Determination of Morphine by Flow Injection Analysis with Chemiluminescence Detection', *Analyst (London)*, **111**(6), 635–640 (1986).
73. C. Pierron, M.F. Etcheverry, J.M. Panas, G. Ledouble, 'Quantitative Determination of Morphine in Pharmaceutical Preparations Using High-performance Liquid Chromatography', *Ann. Pharm. Fr.*, **45**(6), 475–484 (1987).
74. H. Deslandes, H. DeScheemaeker, D. Barthes, M. Lila, V. Furstoss, 'Determination of Morphine in Poppy Pods by Near-infrared Reflectance Spectrophotometry', *Farmaco, Ed. Prat.*, **41**(12), 388–396 (1986).
75. R.V. Gailonde, S.N. Joshi, 'Analysis of a Drug Preparation Containing Noscapine, Ephedrine Hydrochloride and Chlorpheniramine Maleate by Thin Layer Chromatography', *Indian Drugs*, **23**(10), 575–576 (1986).
76. V. Haikala, 'Rapid High-performance Liquid Chromatography Determination of Noscapine Hydrogen Embonate', *J. Chromatogr.*, **389**(1), 299–305 (1987).
77. M. Eisman, M. Gallego, M. Valcarcel, 'Indirect Flame Atomic Absorption Spectrometric Determination of Papaverine, Strychnine and Cocaine by Continuous Precipitation with Dragendorff's Reagent', *J. Anal. At. Spectrom.*, **8**(8), 1117–1120 (1993).
78. V. Das Gupta, 'Quantitation of Papaverine Hydrochloride in a Discolored Injection', *Drug Stab.*, **1**(1), 132–134 (1996).
79. I.P. Shesterova, E.E. Karibyan, S.S. Turaeva, M.M. Penzina, 'Ionometric Determination of Alkaloids', *Chem. Nat. Comp. Engl. Transl.*, **31**(1), 147–148 (1995).
80. C. Eppelsheim, R. Aubeck, N. Hampp, C. Braeuchle, 'Determination of Ethaverine and Papaverine Using Ion-selective Electrodes', *Analyst (London)*, **116**(10), 1001–1003 (1991).
81. A.F. Shoukry, Y.M. Issa, H. Ibrahim, O.A. El-Rashiedy, 'Papaverine-selective Plastic Membrane Electrode based on Papaverinium Tetraphenylborate', *Anal. Lett.*, **24**(10), 1861–1873 (1991).
82. O.M. Salama, M.I. Walash, 'Densitometric Determination of Papaverine in Opium, Poppy Capsules and

Certain Pharmaceutical Dosage Forms', *Anal. Lett.*, **24**(1), 69–82 (1991).
83. I. Mori, Y. Fujita, H. Kawabe, K. Fujita, 'Spectrophotometric Determination of Papaverine Hydrochloride by Using a Membrane Filter Preconcentration Technique with 2,4,5,7-Tetrachlorofluorescein and Palladium(II)', *Chem. Pharm. Bull.*, **34**(2), 902–905 (1986).
84. B.A. El-Zeany, M. Abd El-Kawy, A.A. Moustafa, M.F. Abdel-Ghany, 'Determination of Papaverine Hydrochloride through its Reaction with N.B.S.', *Bull Fac. Pharm. (Cairo Univ.)*, **31**(2), 135–139 (1993).
85. B.A. El-Zeany, A.A. Moustafa, M. Abdelkawy, M.F. Abdelghany, 'Determination of Papaverine Hydrochloride through its Reaction with Iodine', *Bull Fac. Pharm. (Cairo Univ.)*, **31**(2), 131–134 (1993).
86. B.A. El-Zeany, A.A. Moustafa, M. Abdelkawy, M.F. Abdelghany, 'Simultaneous Spectrophotometric Determination of Papaverine Hydrochloride in Presence of Diphylline in Pharmaceutical Preparations', *Bull. Fac. Pharm. (Cairo Univ.)*, **31**(1), 5–9 (1993).
87. I. Bjornsdottir, S.H. Hansen, 'Determination of Opium Alkaloids in Opium by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **13**(4/5), 687–693 (1995).
88. I. Bjornsdottir, S.H. Hansen, 'Determination of Opium Alkaloids in Crude Opium Using Nonaqueous Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **13**(12), 1473–1481 (1995).
89. V.C. Trenerry, R.J. Wells, J. Robertson, 'Determination of Morphine and Related Alkaloids in Crude Morphine, Poppy Straw and Opium Preparations by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **718**(1), 217–225 (1995).
90. Z. Budvari-Barany, G. Szasz, K. Gyimesi-Forras, 'Optimized and Validated HPLC Methods for Compendial Quality Assessment Opium Alkaloids', *J. Liq. Chromatogr. Relat. Technol.*, **20**(19), 3257–3268 (1997).
91. R. Chiba, Y. Ishii, 'Simultaneous Determination of Yohimbine Hydrochloride, Strychnine Nitrate and Methyltestosterone by Ion-pair High-performance Liquid Chromatography', *J. Chromatogr.*, **588**(1/2), 344–347 (1991).
92. I. S. Ismail, A.C. Galan, 'Determination of the Stability of Morphine Tablets by Ion-pair Reversed-phase Liquid Chromatography', *Anal. Chim. Acta*, **283**(1), 334–337 (1993).
93. S.H. Hansen, 'HPLC Assay of the Opiates in Opium and Cough Mixtures Using Dynamically Modified Silica and UV Absorbance, Fluorescence and Electrochemical Detection', *Int. J. Pharm.*, **32**(1), 7–11 (1986).
94. M. Dolezalova, 'Ion-pair High-performance Liquid Chromatographic Determination of Morphine and Pseudomorphine in Injections', *J. Pharm. Biomed. Anal.*, **10**(7), 507–514 (1992).
95. E.H. Girgis, 'Ion-pair Reversed-phase Liquid Chromatographic Identification and Quantitation of Papaverine Congeners', *J. Pharm. Sci.*, **82**(5), 503–505 (1993).
96. V. Haikala, L. Heimonen, 'Determination of Codeine in Complicated Cough Preparations by Reversed-phase Liquid Chromatography', *Acta Pharm. Fenn.*, **98**(3), 181–188 (1989).
97. A.C. Bello, R.K. Jhangiani, 'Liquid Chromatographic Determination of Morphine Sulfate and Some Contaminants in Injections and Bulk Drug Material: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **71**(5), 1046–1048 (1988).
98. J.L. Janicot, M. Caude, R. Rosset, 'Separation of Opium Alkaloids by Carbon Dioxide Sub- and Supercritical Fluid Chromatography with Packed Columns. Application to the Quantitative Analysis of Poppy Straw Extracts', *J. Chromatogr.*, **437**(2), 351–364 (1988).
99. A. Lodzinska, A. Balter, A. Rozploch, B. Dembinski, 'Spectrofluorimetric Determination of Reserpine', *Chem. Anal. (Warsaw)*, **41**(2), 283–291 (1996).
100. U.R. Cieri, 'Determination of Reserpine, Hydralazine HCl, and Hydrochlorothiazide in Tablets by Liquid Chromatography on a Short, Normal-phase Column', *J. Assoc. Off. Anal. Chem. Int.*, **77**(5), 1104–1108 (1994).
101. E. Dargel, J.B. Mielck, 'HPLC Methods for Separation and Quantitation of Reserpine and its Main Degradation Products', *J. Liq. Chromatogr.*, **13**(20), 3973–3984 (1990).
102. H.L. Rau, A.R. Aroor, P.G. Rao, 'High-performance Liquid Chromatographic Determination of Reserpine in Tablets', *Indian Drugs*, **28**(3), 157–158 (1990).
103. U.R. Cieri, 'Determination of Reserpine and Rescinnamine in *Rauwolfia serpentina* Powders and Tablets: Collaborative Study', *J. Assoc. Off. Anal. Chem. Int.*, **81**(2), 373–380 (1998).
104. B. Dembinski, H. Zawadzki, 'Extraction–Colorimetric Determination of Reserpine as Reserpine Picrate', *Chem. Anal. (Warsaw)*, **31**(3), 437–442 (1986).
105. U.R. Cieri, 'Determination of Reserpine and Rescinnamine in *Rauwolfia serpentina* Preparations by Liquid Chromatography with Fluorescence Detection', *J. Assoc. Off. Anal. Chem.*, **70**(3), 540–546 (1987).
106. U.R. Cieri, 'Determination of Reserpine and Hydrochlorothiazide in Commercial Tablets by Liquid Chromatography with Fluorescence and UV Absorption Detectors in Series', *J. Assoc. Off. Anal. Chem.*, **71**(3), 515–518 (1988).
107. T.N.V. Prasad, E.V. Rao, C.S.P. Sastry, G.R. Rao, 'High-performance Liquid Chromatographic Assay of Chlorothalidone and Reserpine in Single and Combined Dosage Forms', *Indian Drugs*, **24**(8), 398–401 (1987).
108. J. Wang, T. Tapia, M. Bonakdar, 'Sensitive Adsorptive Stripping Voltammetric Measurements of Antihypertensive Drugs', *Analyst (London)*, **111**(11), 1245–1248 (1986).
109. P. Duez, S. Chamart, M. Vanhaelen, R. Vanhaelen-Fastre, M. Hanocq, L. Molle, 'Comparison between High-performance Thin-layer Chromatography–Densitometry and High-performance Liquid Chromatography

for the Determination of Ajmaline, Reserpine and Rescinnamine in *Rauwolfia vomitoria* Root Bark', *J. Chromatogr.*, **356**(2), 334–340 (1986).
110. M. Sanchez, J.J. Sanchez-Aibar, 'Spectrofluorimetric Determination of Reserpine by Oxidation with Cerium-(IV) Sulfate', *Analyst (Cambridge, UK)*, **121**(11), 1581–1582 (1996).
111. A.G. Mekkawi, A.A. Al-Badr, 'Yohimbine', *Anal. Profiles Drug Subst.*, **16**, 731–768 (1987).
112. N.S. El-Shaer, 'Derivative Spectrophotometric and Colorimetric Determination of Yohimbine HCl', *J. Pharm. Sci.*, **11**(1), 5–8 (1997).
113. S. Auriola, T. Naaranlahti, S.P. Lapinjoki, 'Determination of *Catharanthus roseus* Alkaloids by High-performance Liquid Chromatography–Isotope Dilution Thermospray–Mass Spectrometry', *J. Chromatogr.*, **554**(1/2), 227–231 (1991).
114. F. Belal, M. Sano, I. Tomita, 'High-performance Liquid Chromatographic Determination of Yohimbine and Strychnine in Dosage Forms', *Chem. Fharm. Bull.*, **37**(6), 1622–1623 (1989).
115. A.P. Argekar, S.V. Raj, S. Kapadia, 'Quantitative Determination of Reserpine from *Rauwolfia serpentina* Tablets by High Performance Thin Layer Chromatography', *J. Planar Chromatogr.-Mod. TLC*, **9**(2), 148–151 (1996).
116. G.J. Lehr, S.M. Yuen, G.D.J. Lawrence, 'Liquid Chromatographic Determination of Atropine in Nerve Gas', *Assoc. Off. Anal. Chem. Int.*, **78**(2), 339–343 (1995).
117. W. Zhu, W. Wei, L. Nie, S. Yao, 'Piezoelectric Quartz Crystal with Separated Electrode for the Simultaneous Determination of Atropine Sulfate and Sodium Chloride', *Anal. Chim. Acta*, **282**(3), 535–541 (1993).
118. T. Oshima, K. Sagara, F. Hirayama, T. Mizutani, L. He, Y. Tong, Y. Chen, H. Itokawa, 'Combination of Ion-pair Column Switching in High-performance Liquid Chromatography of Tropane Alkaloids', *J. Chromatogr.*, **547**(1/2), 175–183 (1991).
119. E. Heldin, N. Hang Huynh, C. Pettersson, '(2*S*,3*S*)-Dicyclohexyl Tartrate as Mobile Phase Additive for the Determination of the Enantiomeric Purity of (*S*)-Atropine in Tablets', *J. Chromatogr.*, **592**(1/2), 339–343 (1992).
120. E. Arvidsson, S.O. Jansson, G. Schill, 'Chiral Separations of Atropine and Homatropine on Alpha.1-Acid Glycoprotein-bonded Stationary Phase', *J. Chromatogr.*, **506**, 579–591 (1990).
121. G. Santoni, L. Fabbri, G. Renzi, P. Mura, S. Pinzauti, 'Simultaneous Determination of Atropine, Diazepam and Ergotamine by Ion-pair High Performance Liquid Chromatography', *Boll. Chim. Farm.*, **130**(1), 14–16 (1990).
122. O.-W. Lau, C.-S. Mok, 'High-performance Liquid Chromatographic Determination of Atropine and Atropine-like Alkaloids in Pharmaceutical Preparations with Indirect Conductometric Detection', *J. Chromatogr. A*, **766**(1/2), 270–276 (1997).
123. M. Takahashi, M. Nagashima, S. Shigeoka, M. Nishijima, K. Kamata, 'Determination of Atropine in Pharmaceutical Preparations by Liquid Chromatography with Fluorescence Detection', *J. Chromatogr. A*, **775**(1/2), 137–141 (1997).
124. D.W. Armstrong, S.M. Han, Y.I. Han, 'Separation of Optical Isomers of Scopolamine, Cocaine, Homatropine, and Atropine', *Anal. Biochem.*, **167**(2), 261–264 (1987).
125. S. Cherkaoui, L. Mateus, P. Christen, J.-L. Veuthey, 'Development and Validation of a Capillary Zone Electrophoresis Method for the Determination of Atropine, Homatropine and Scopolamine in Ophthalmic Solutions', *J. Chromatogr. B: Biomed. Sci. Appl.*, **696**(2), 283–290 (1997).
126. R.L. Glass, E.L. Johnson, 'Comparison of High-performance Liquid Chromatographic and Gas Chromatographic Analysis of Cocaine in Coca Leaves', *J. Liq. Chromatogr.*, **16**(16), 3543–3555 (1993).
127. R.R. MacGregor, J.S. Fowler, A.P. Wolf, 'Determination of the Enantiomeric Composition of Samples of Cocaine by Normal-phase High-performance Liquid Chromatography with UV Detection', *J. Chromatogr.*, **590**(2), 354–358 (1992).
128. G.E. Pinilla, B.L. Calvo, C.R.M. Garcia-Monco, M.A. Sanchez, 'Polarographic Behavior of 8-Chlorotheophylline and its Determination in Dosage Forms', *Electroanalysis (NY)*, **5**(4), 343–347 (1993).
129. Y. Pramar, V. Das Gupta, 'Drug Development: Quantitation of Scopolamine Hydrobromide when Adsorbed onto Microcrystalline Cellulose and Sodium Carboxymethyl Cellulose in Tablets', *Ind. Pharm.*, **17**(17), 2401–2407 (1991).
130. L. Botz, L.G. Szabo, 'Separation of Tropane Alkaloids by TLC, HPTLC, and OPLC Methods', *J. Planar Chromatogr.-Mod. MC*, **1**(1), 85–87 (1988).
131. G. Satoni, A. Tonsini, P. Gratteri, P. Mura, S. Furlanetto, S. Pinzauti, 'Determination of Aropine Sulfate and Benzalkonium Chloride in Eye Drops by HPLC', *Int. J. Pharm.*, **93**(1/3), 239–243 (1993).
132. I. Papadoyannis, A. Zotou, V. Samanidou, M. Georgarakis, 'Solid-phase Extraction and RP-HPLC Analysis of Atropine Sulfate and Scopolamine-*N*-Butylbromide in Pharmaceutical Preparations and Biological Fluids', *Instrum. Sci. Technol.*, **22**(1), 83–103 (1994).
133. S. Mandal, A.A. Naqvi, R.S. Thakur, 'Analysis of some Tropane Alkaloids in Plants by Mixed-column High-performance Liquid Chromatography', *J. Chromatogr.*, **547**(1/2), 468–471 (1991).
134. L.-Y. He, G. Zhang, Y. Tong, K. Sagara, T. Oshima, T. Yoshida, 'Reversed-phase Ion-pair High-performance Liquid Chromatographic Separation and Determination of Tropane Alkaloids in Chinese Solanaceous Plants', *J. Chromatogr.*, **481**, 428–433 (1989).

135. K.G. Feitsma, K.H. Kooi, B.F.H. Drenth, R.A. DeZeeuw, 'Separation and Determination of Enantiomers of Tropane Alkaloids and Synthetic Derivatives: Possibilities and Limitations', *Pharm. Weekbl.*, **121**(37), 875–880 (1986).
136. S. Mandal, M.L. Maheshwari, 'High-pressure Liquid Chromatographic Determination of Vindoline, Catharanthine, Vincaleukoblastine and Vincristine in Periwinkle Leaf', *Indian J. Pharm. Sci.*, **49**(6), 205–209 (1987).
137. K.N. Thimmaiah, M.J. Thomas, V.S. Sethi, N.M. Made Gowda, 'Desorption Chemical Ionization and Field Desorption Mass Spectrometric Analysis of Antitumor Catharanthus (Vinca) Alkaloids', *Microchem. J.*, **41**(2), 183–190 (1990).
138. D. Drapeau, H.W. Blanch, C.R. Wilke, 'Liquid Chromatographic Isolation of Vincristine and Vinblastine', *J. Chromatogr.*, **390**(2), 297–306 (1987).
139. F.J. Muhtadi, A.F.A.A. Afify, 'Vinblastine Sulfate (Review Supplement)', *Anal. Prof. Drug Subst. Excip.*, **21**, 611–658 (1992).
140. S.K. Volkov, E.I. Grodnitskaya, 'Determination of Relative Vinblastine Quantities in Rose (*Catharanthus Roseus*) by Using HPLC', *Khim.-Farm. Zh.*, **9**, 77–78 (1991).
141. I. Fitos, J. Visy, M. Simonyi, J. Hermanson, 'Chiral High-performance Liquid Chromatographic Separations of Vinca Alkaloid Analogs on α 1-Acid Glycoprotein and Human Serum Albumin Columns', *J. Chromatogr.*, **609**(1/2), 163–171 (1992).
142. A. Bazzi, J. Montgomery, G. Alent, 'Indirect Voltammetric and Atomic Absorption Spectrometric Determination of Caffeine', *Analyst (London)*, **113**(1), 121–124 (1988).
143. S.M. Mayanna, B. Jayaram, 'Oxidimetric Estimation of Caffeine Using Aromatic Sulfonyl Haloamines', *J. Indian Chem. Soc.*, **63**(3), 329–331 (1986).
144. E.N. Vergeichik, A.S. Saushkina, T.T. Likhota, L.S. Raimova, 'Quantitative Analysis of Caffeine–Sodium Benzoate Tablets by Derivative Spectrophotometry', *Farmatsiya (Moscow)*, **35**(5), 43–47 (1986).
145. P. Sun, G.J. Mariano, G. Barker, R.A. Hartwick, 'Comparison of Micellar Electrokinetic Capillary Chromatography and High-performance Liquid Chromatography on the Separation and Determination of Caffeine and its Analogs in Pharmaceutical Tablets', *Anal. Lett.*, **27**(5), 927–937 (1994).
146. M.M. Andino, C.G. De Lima, J.D. Winefordner, 'Luminescence Characteristics of Caffeine and Theophylline', *Spectrochim. Acta, Part A*, **43**(3), 427–437 (1987).
147. M.N.M.P. Alcada, J.L.F.C. Lima, M.C.B.S.M. Montenegro, 'Quinidine Ion-selective Electrode for Potentiometric Determinations in Pharmaceutical Preparations', *Anal. Chim. Acta*, **283**(1), 657–661 (1993).
148. E. Pinilla Gil, L. Calvo Blazquez, R.M. Garcia-Monco Carra, A. Sanchez Misiego, 'Polarographic Behavior of 8-Chlorotheophylline and its Determination in Dosage Forms', *Electroanalysis*, **5**(4), 343–347 (1993).
149. K. Nikolic, M. Medenica, 'Potentiometric Determination of 8-Chlorotheophylline', *Mikrochim. Acta*, **1**(5/6), 325–329 (1986).
150. K.I. Nikolic, K.R. Velasevic, M.F. Ruasse, 'Chlorocoulometric Determination of 8-Chlorotheophylline', *J. Pharm. Belg.*, **42**(1), 44–46 (1987).
151. H.A. Abu-Shady, S.T. Hassib, N.F. Youssef, 'On the Analysis of Diprophylline', *Egypt. J. Pharm. Sci.*, **28**(1/4), 223–234 (1987).
152. E.V. Rao, G.R. Rao, S. Raghuveer, P. Khadgapathi, 'High-performance Liquid Chromatographic Determination of Theophylline and Etophylline in Pharmaceutical Dosage Forms', *Indian J. Pharm. Sci.*, **49**(5), 180–182 (1987).
153. S.U. Alvi, F. Castro, 'A Simultaneous Assay of Theophylline, Ephedrine Hydrochloride and Phenobarbital in Suspensions and Tablets Formulations by High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **9**(10), 2269–2279 (1986).
154. I. Perez-Martinez, S. Sagrado, M.J. Medina-Hernandez, 'Determination of Theophylline in Pharmaceuticals by Micellar Liquid Chromatography and Spectrophotometric Detection', *J. Liq. Chromatogr. Relat. Technol.*, **19**(12), 1957–1966 (1996).
155. V.M. Shinde, P.B. Shetkar, S.V. Vartak, 'Simultaneous Determination of Theophylline and Etofylline in Pharmaceutical Dosages by Reversed Phase HPLC', *Indian Drugs*, **34**(1), 26–31 (1997).
156. A.G. Mwalupindi, I.M. Warner, 'Determination of Theophylline by Liquid Chromatography with Sensitized Lanthanide Luminescence Detection', *Anal. Chim. Acta*, **306**(1), 49–56 (1995).
157. R.M. Shubietah, A.Z. Abu Zuhri, A.G. Fogg, 'Adsorptive Cathodic Stripping Voltammetric Determination of Theophylline at a Hanging Mercury Drop Electrode', *Analyst (London)*, **119**(9), 1967–1970 (1994).
158. J.D. Kirsch, J.K. Drennen, 'Determination of Film-coated Tablet Parameters by Near-infrared Spectroscopy', *J. Pharm. Biomed. Anal.*, **13**(10), 1273–1281 (1995).
159. Q. Dang, L. Yan, Z. Sun, D. Ling, 'Separation and Simultaneous Determination of the Active Ingredients in Theophylline Tablets by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr.*, **630**(1/2), 363–369 (1993).
160. S.R. Marin, E.M. Llobat, M.D. San Martin Ciges, A.R. Mauri Aucejo, 'Spectrophotometric Determination of Theophylline in Pharmaceuticals Employing the Apparent Content Curves to Resolve Spectral Interferences', *Anal. Lett.*, **26**(4), 641–655 (1993).
161. J. Wang, E. Dempsey, M. Ozsoz, M.R. Smyth, 'Amperometric Enzyme Electrode for Theophylline', *Analyst (London)*, **116**(10), 997–999 (1991).

162. C.A. Lau-Cam, R.W. Roos, 'Simultaneous High-performance Liquid Chromatographic Determination of Theophylline and Ethylenediamine in Aminophylline Dosage Forms as their Dansyl Derivatives', *J. Liquid Chromatogr.*, **14**(10), 1939–1956 (1991).
163. S.R. El-Shabourie, S.A. Hussein, S.E. Emara, 'Colorimetric Determination of Theophylline and Aminophylline with Diazotized *p*-Nitroaniline', *Talanta*, **36**(12), 1288–1290 (1989).
164. L.M. Perry, J.D. Winefordner, 'Selective Determination of Theophylline in the Presence of Caffeine by Sensitized Luminescence of Europium(III)', *Talanta*, **37**(10), 965–969 (1990).
165. D. Baylocq, W. Kayata, F. Pellerin, 'Determination of Weak Acids in Concentrated Quaternary Ammonium Media', *Talanta*, **34**(5), 515–517 (1987).
166. M. Carmen Gutierrez, A. Gomez-Hens, D. Perez-Bendito, 'Stopped-flow Fluorometric Determination of Theophylline in Pharmaceutical Preparations', *Analyst (London)*, **113**(4), 559–562 (1988).
167. R.V. Gaitonde, U. Rivankar, 'Analysis of a Drug Preparation Containing Ephedrine Hydrochloride, Theophylline, Chlorpheniramine Maleate and Diazepam by Thin Layer Chromatography', *Indian Drugs*, **24**(10), 486–488 (1987).
168. O.M. Salama, M.I. Walash, 'Quantitative Densitometric Determination of Theophylline', *Anal. Lett.*, **22**(4), 827–839 (1989).
169. L.M. Perry, E.Y. Shao, J.D. Winefordner, 'Room-temperature Phosphorimetry Studies of Caffeine and Theophylline', *Talanta*, **36**(10), 1037–1040 (1989).
170. M.C. Gutierrez, A. Gomez-Hens, D. Perez-Benclito, 'Stopped-flow Fluorometric Determination of Theophylline in Pharmaceutical Preparations', *Analyst (London)*, **113**(4), 559–562 (1988).
171. J.E. Haky, W.M. Foss, B.L. Marks, 'Analysis of Aminophylline in Thigh Cream Formulations by High Performance Liquid Chromatography', *J. Liq. Chromatogr. Relat. Technol.*, **20**(15), 2399–2414 (1997).
172. B. Renger, H. Jehle, M. Fischer, W. Funk, 'Validation of Analytical Procedures in Pharmaceutical Analytical Chemistry: HPTLC Assay of Theophylline in an Effervescent Tablet', *J. Planar Chromatogr.-Mod. TLC*, **8**(4), 269–278 (1995).
173. V.M. Shinde, N.M. Tendolkar, B.S. Desai, 'Simultaneous Determination of Theophylline and Etofylline in Pharmaceutical Dosage [Forms] by HPTLC', *Anal. Lett.*, **28**(1), 45–58 (1995).
174. Z. Budvari-Barany, G. Szasz, K. Gyimesi-Forras, 'Optimized and Validated HPLC Methods for Compendial Quality Assessment. II. Opium Alkaloids', *J. Liq. Chromatogr. Relat. Technol.*, **20**(8), 1233–1242 (1997).
175. A.R. Barnes, 'Determination of Caffeine and Potassium Sorbate in a Neonatal Oral Solution by HPLC', *Int. J. Pharm.*, **80**(2/3), 267–270 (1992).
176. Y.M. El-Sayed, S.I. Islam, 'Comparison of Fluorescence Polarization Immunoassay and HPLC for the Determination of Theophylline in Serum', *Clin. Pharm. Ther.*, **14**, 127–134 (1989).
177. L. Gracza, 'Analysis of Plant Xanthine Derivatives. Part 4. Standardization and Stability of Tinctures', *PZ Wiss.*, **4**(3), 123–126 (1991).
178. J. Thomas, 'Ultra-rapid High-performance Liquid Chromatographic Analysis of Theophylline and Sodium (or Potassium) Anisate in Different Pharmaceutical Preparations', *J. Chromatogr.*, **479**(2), 430–436 (1989).
179. M. Baucells, N. Ferrer, P. Gomez, G. Lacort, M. Roura, 'Determination of Caffeine in Solid Pharmaceutical Samples by FTIR Spectroscopy', *Mikrochim. Acta*, **112**(1/4), 87–98 (1993).
180. L.M. Cabalin, J.J. Laserna, A. Ruperez, J. Fresenius 'Room-temperature Phosphorimetry of Methylxanthine Stimulants on Zeolite-modified Filter Paper Substrates', *Anal. Chem.*, **346**(10/11), 1003–1007 (1993).
181. J. Reisch, W. Probst, D. Groeger, 'The Application of High-performance Liquid Chromatography to Separate and Determine Acridone Alkaloids', *Pharmazie*, **45**(7), 500–501 (1990).
182. B.L. Poehland, N. Troupe, B.K. Carte, J.W. Westley, 'Reversed-phase High-performance Liquid Chromatographic Assay for Camptothecin and Related Alkaloids', *J. Chromatogr.*, **481**, 421–427 (1989).
183. S.-J. Lin, H.-H. Tseng, K.-C. Wen, T.-T. Suen, 'Determination of Gentiopicroside, Mangiferin, Palmatine, Berberine, Baicalin, Wogonin and Glycyrrhizin in the Traditional Chinese Medicinal Preparation Sann-Joong-Kuey-Jian-Tang by High-performance Liquid Chromatography', *J. Chromatogr. A*, **730**(1/2), 17–23 (1996).
184. S. Komorsky-Lovric, 'Electrochemical Studies of Berberine and Jatrorubine by Pulse Polarography', *Mikrochim. Acta*, **1**(5/6), 407–414 (1986).
185. S. Komorsky-Lovric, 'Square-wave Voltammetry of Berberine', *J. Electroanal. Chem. Interfacial Electrochem.*, **219**(1/2), 281–289 (1987).
186. A.L. Ramos Rubio, C. Cruces Blanco, F. Garcia Sanchez, 'Synchronous Scanning First and Second Derivative Spectrofluorometric Determination of the Alkaloid Berberine', *Fresenius' Z. Anal. Chem.*, **323**(2), 153–156 (1986).
187. T. Sakai, Y.S. Chung, N. Ohno, 'Enhancement of Selectivity and Sensitivity for Berberine Determination by Extraction–Fluorophotometry', *Anal. Sci.*, **8**(3), 377–379 (1992).
188. T. Sakai, N. Ohno, Y.S. Chung, H. Nishikawa, 'Spectrofluorometric Determination of Berberine in Oriental Pharmaceutical Preparations by Flow-injection Analysis Coupled with Liquid–Liquid Extraction', *Anal. Chim. Acta*, **308**(1/3), 329–333 (1995).
189. A.I. Zhebentyaev, I.E. Talut, 'Spectrophotometric Determination of Berberine Bisulfate', *Chem. Nat. Compd. (Eng. Transl.)*, **25**(2), 203–204 (1989).

190. T. Sakai, N. Ohno, H. Sasaki, T. Hyuga, 'Extraction–Spectrophotometric Determination of Berberine in Crude Drugs by the Formation of a New Ion Associate', *Anal. Sci.*, **7**(1), 39–43 (1991).
191. T. Sakai, 'Solvent Extraction–Spectrophotometric Determination of Berberine and Benzethonium in Drugs with Tetrabromophenolphthalein Ethyl Ester by Batch-wise and Flow-injection Methods', *Analyst (London)*, **116**(2), 187–190 (1991).
192. H.L. Constant, G.A. Cordell, D.P. West, J.H. Johnson, 'Separation and Quantification of Capsaicinoids Using Complexation Chromatography', *J. Nat. Prod.*, **58**(12), 1925–1928 (1995).
193. B. Deus-Neumann, J. Stoeckigt, M.H. Zenk, 'Radioimmunoassay for the Quantitative Determination of Catharanthine', *Planta Med.*, **53**(2), 184–188 (1987).
194. G. Matysik, L. Jusiak, 'Stepwise Gradient in Thin-layer Chromatography of Chelidonium Alkaloids', *J. Chromatogr.*, **518**(1), 273–276 (1990).
195. C. Bugatti, M.L. Colombo, F. Tome, 'High-performance Liquid Chromatographic Separation of Quaternary Alkaloids of *Chelidonium Majus* L. Roots', *J. Chromatogr.*, **393**(2), 312–316 (1987).
196. J. Wang, M. Ozsoz, 'Trace Measurements of Colchicine by Adsorptive Stripping Voltammetry', *Talanta*, **37**(8), 783–787 (1990).
197. S. Auriola, V.P. Ranta, T. Naaranlahti, S.P. Lapinjoki, 'Thermospray Liquid Chromatographic–Mass Spectrometric Analysis of Catharanthus Alkaloids', *J. Chromatogr.*, **474**(1), 181–185 (1989).
198. R. Van der Heijden, P.J. Lamping, P.P. Out, R. Wijnsma, R. Verpoorte, 'High-performance Liquid Chromatographic Determination of Indole Alkaloids in a Suspension Culture of *Tabernaemontana Divaricata*', *J. Chromatogr.*, **396**, 287–295 (1987).
199. H. Ibrahim, T. Boye, M. Wermeille, R. Gurny, P. Buri, 'Determination of Pilocarpine by HPLC in Presence of Isopilocarpine and a pH-Sensitive Polymer in Ophthalmic Dispersions', *J. Pharm. Biomed. Anal.*, **5**(4), 379–382 (1987).
200. T.Y. Fan, G.M. Wall, K. Sternitzke, L. Bass, A.B. Morton, E. Muegge, 'Improved High-performance Liquid Chromatographic Determination of Pilocarpine and its Degradation Products in Ophthalmic Solutions. Importance of Octadecylsilane Column Choice', *J. Chromatogr. A*, **740**(2), 289–295 (1996).
201. W.N. Charman, A.J. Humberstone, S.A. Charman, 'Comparison of the Behavior of HPLC Stationary Phases for the Determination of Pilocarpine and its Impurities or Degradation Products', *Pharm. Res.*, **9**(9), 1219–1223 (1992).
202. C. Pilatti, M.D. Torre, C. Chiale, M. Spinetto, 'Stability of Pilocarpine Ophthalmic Solutions', *Drug Dev. Ind. Pharm.*, **25**(6), 801–805 (1999).
203. C. Durif, M. Ribes, G. Kister, A. Puech, 'Comparison of the Behavior of HPLC Stationary Phases for the Determination of Pilocarpine and its Impurities or Degradation Products', *Pharm. Acta Helv.*, **63**(11), 294–298 (1988).
204. C. Durif, M. Ribes, G. Kister, A. Puech, 'Validation of an HPLC Technique for Determining Pilocarpine and its Impurities or Degradation Products', *Pharm. Acta Helv.*, **63**(11), 315–320 (1988).
205. S.D. Desai, J.A. Blanchard, 'Simplified and Rapid High-performance Liquid Chromatographic Assay for Pilocarpine Hydrochloride', *J. Chromatogr. Sci.*, **30**(4), 149–152 (1992).
206. G. Balansard, E. Vidal, M. Guigues, B. Ollivier, M. De Meo, 'Pilocarpine and Isopilocarpine Determinations by Normal-phase High-performance Liquid Chromatography. Application to the Study of the Stability of Pilocarpine Based Collyria and Comparison of this Method with a Colorimetric Method Using Hydroxylamine', *Pharm. Acta Helv.*, **61**(2), 47–50 (1986).
207. K.D. Sternitzke, T.Y. Fan, D.L. Dunn, 'High-performance Liquid Chromatographic Determination of Pilocarpine Hydrochloride and its Degradation Products Using a beta-Cyclodextrin Column', *J. Chromatogr.*, **589**(1/2), 159–164 (1992).
208. C. Durif, M. Ribes, G. Kister, A. Puech, 'Rapid, Semi-quantitative, TLC Determination of Pilocarpic and Isopilocarpic Acids in Pilocarpine Eye Drops', *Pharm. Acta Helv.*, **61**(5/6), 135–138 (1986).
209. T. Sakai, N. Ohno, T. Higashi, M. Tanaka, 'Extraction–Spectrophotometric Determination of Pilocarpine in Ophthalmic Solution', *Anal. Sci.*, **1**(3), 275–279 (1985).
210. M.M. Ayad, S.E. Khayyal, N.M. Farag, 'Indirect Determination of Pilocarpine by Atomic Absorption Spectrometry', *Microchem. J.*, **33**(3), 371–375 (1986).
211. P. Xie, Y. Yan, Q. Lin, 'Optimization of the TLC of Protoberberine Alkaloids and Fingerprint Evaluation of the Coptidis Rhizome', *J. Planar Chromatogr.-Mod. TLC*, **5**(5), 302–307 (1992).
212. K. Tagahara, J. Koyama, T. Okatani, Y. Suzuta, 'Chromatographic Resolution of Racemic Tetrahydroberberine Alkaloids by Using Cellulose Tris(phenylcarbamate) Stationary Phase', *Chem. Pharm. Bull.*, **34**(12), 5166–5168 (1986).
213. J. Zhang, Z. Tian, Z. Lou, 'Simultaneous Determination of Six Alkaloids in Ephedrae Herba by High-performance Liquid Chromatography', *Planta Med.*, **54**(1), 69–70 (1988).
214. L.A.C. Pieters, A.J. Vliefinck, 'Comparison of High-performance Liquid Chromatography with Proton Nuclear Magnetic Resonance Spectrometry for the Quantitative Analysis of Pyrrolizidine Alkaloids from Senecio Vulgaris', *J. Liq. Chromatogr.*, **9**(4), 745–755 (1986).
215. M. Unger, 'Capillary Zone Electrophoresis of Alkaloids. Influence of Structure on Electrophoretic Mobility', *J. Chromatogr. A.*, **807**, 81–87 (1998).

216. F. Van Damme, M. Verzele, 'Normal-phase Liquid Chromatography of Alkaloids on Polyol-derivatized Silica Gel', *LC–GC*, **5**(11), 980–985 (1987).
217. B. Zsadon, M. Szilasi, L. Decsei, A. Ujhazy, J. Szeftli, 'Variation of the Selectivity in the Resolution of Alkaloid Enantiomers on Cross-linked Cyclodextrin Polymer Stationary Phases', *J. Chromatogr.*, **356**(3), 428–432 (1986).
218. A. Hermans-Lokkerbol, R. Verpoorte, 'Droplet Counter-current Chromatography of Alkaloids. The Influence of pH-gradients and Ion-pair Formation on the Retention of Alkaloids', *Planta Med.*, **4**, 299–302 (1986).
219. C. Wang, D. Zhang, Y. Guo, H. Zhong, M. Wen, 'Microdetermination of Alkaloids in Organic Solvents by Potentiometric Titration', *Anal. Chim. Acta*, **196**, 299–303 (1987).
220. M. Abdel-Salam, M.S. Mahrous, A.S. Issa, 'Utility of 2,6-Dichlorophenolindophenol for the Spectrophotometric Determination of Certain Alkaloids', *J. Pharm. Belg.*, **41**(4), 226–230 (1986).

Antibiotics, Pharmaceutical Analysis of

R.K. Gilpin
Wright State University, Dayton, USA

C.S. Gilpin
Ohio University–Eastern Campus, St. Clairsville, USA

L.A. Pachla
IBAH, Inc., Ft. Washington, USA

1 Introduction **1**
2 General Information **2**
2.1 Cephalosporins 2
2.2 Penicillins 5
2.3 Quinolones 10
2.4 Streptomycin and Related Analogs 17
2.5 Sulfonamides and Miscellaneous Antibacterial Agents 17
2.6 Tetracyclines 17
Abbreviations and Acronyms **17**
Related Articles **17**
References **18**

An antibiotic is a chemical substance, produced by microorganisms (or recently, synthetic routes) which has the capacity to inhibit the growth of other microorganisms/bacterial agents. The terminology for antibiotics stems from the Latin term of antibiosis *and the noun* antibiotic *was coined by Professor Waksman in 1942. Considered as a single therapeutic class, antibiotics are one of the most diverse groups of medicinal agents. They can be grouped into several subclasses depending on their source, structure and activity. These are the cephalosporins, penicillins, quinolones, streptomyces and tetracycline. In addition to these agents, the sulfonamides and other miscellaneous compounds are also included owing to their antibacterial activity.*

1 INTRODUCTION

Antibiotics include a broad spectrum of compounds that are diverse in structure as well as in their mode of action within an organism resulting from the many different sites where they may block or impede bacterial growth. Common mechanisms of action include the inhibition of cell wall synthesis (e.g. cephalosporin and penicillin), membrane function (e.g. gramicidins, nystatin, and streptomycin), ribonucleic acid metabolism (e.g. griseofulvin and neomycin), deoxyribonucleic acid metabolism (e.g. actinomycin and daunomycin), purine and pyrimidine synthesis (e.g. sparkomycin and tubercidin), protein synthesis (e.g. chloramphenicol, erythromycin, and tetracycline), respiration (e.g. antimycins and oligomycins), and oxidative phosphorylation (e.g. gramicidins and valionomycins). In some instances a given antibiotic may influence multiple sites. For example, streptomycin affects both membrane function and protein synthesis, while griseofulvin affects ribonucleic acid metabolism and deoxyribonucleic acid metabolism. Neomycin affects ribonucleic acid metabolism and protein synthesis. They are therapeutically used as antibacterial, antiinfective, antifungal, antiparisitic, antimicrobial and in some cases anticancer drugs. Certain antibiotics like the tetracyclines may have a broad spectrum of activities and inhibit the growth of a range of microorganisms including Gram-negative and positive bacteria, species of rickettsia and mycoplasma, certain protozoa and large viruses while other antibiotics like erythromycin and troleandomycin are more limited in scope.

Penicillins are the oldest recognized antibiotics originating from the pioneering work of Fleming. They have the 3,3-dimethyl-7-oxo-4-thia-1azabicyclo[3.2.0]heptane-2-carboxylic acid parent structure, I, shown below. The 6 position substitution is respectively a [(aminophenylacetyl)amino], [(phenylacetyl)amino], [[(2-propenylthio) acetyl]amino] or [(phenoxyacetyl)amino] group for the common drugs ampicillin and penicillin G, O, and V. Likewise, a close structurally related group of antibiotics are the cephalosporins, II.

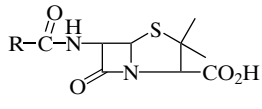

(I)

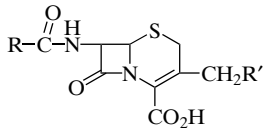

(II)

Some of the more important pathogens where these antibiotics play a significant role in treatment of infections are *Actinomyces israeli*, *Bacillus anthracis*, *Clostridium perfringens*, *Corynebacterium diphtheriae*, *Diplococcus pneumoniae*, *Escherichia coli*, *Hemophilus influenzae*, *Proteus mirabilis*, *Shigella dysenteriae*, *Staphylococcus*

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

aureus, *Streptococcus hemolyticus*, and *Streptococcus viridans*.

Although most antibiotics are produced totally or in part via fermentation processes, compounds like chloramphenicol are manufactured synthetically. Examples of semisynthetic products are ampicillin, methicillin, oxacillin and phenethicillin. These are synthesized by chemically modifying the patient's natural product. In some instances, the initially identified semisynthetic compounds have been later found to be produced biosynthetically. Tetracycline is a good example of this type of compound which was initially synthesized via the catalytic dehalogenation of chlortetracycline but was later produced biosynthetically.

Antibiotics, as a major class of pharmaceuticals, have been assayed by a variety of procedures. In many instances these have involved either chromatographic or spectrophotometric methodology. In the latter, a variety of reagents have been used to form highly colored products. In terms of chromatographic procedures reversed-phase high-performance liquid chromatography (HPLC), and more recently capillary electrophoresis methodology, have become increasingly important. In some instances post-column colorimetric reactions have been used to facilitate detection.

The remainder of this article is divided into six major sections with a representative summary of more recent analytical methodology given in tabular form.

2 GENERAL INFORMATION

2.1 Cephalosporins

The cephalosporins are the second family of β-lactam antibiotics discovered after the penicillins and there are numerous parallelisms between the two families. However, unlike the penicillins, they are generally considered to be broad-spectrum antibiotics for both Gram-positive and Gram-negative pathogens. All of the cephalosporins are semisynthetic derivatives of the natural cephalosporin C antibiotic. Their general mode of action is via inhibiting bacterial wall synthesis similar to the penicillins. There is a marked variation for each cephalosporin against β-lactamase. The most resistant compound against this enzyme is cephalothin.

There are strong historical, biological and chemical similarities between the cephalosporins and the penicillins. The common feature is the β-lactam ring,

Table 1 Methodology for measuring common cephalosporins

Compound	Technique	Method/comments[a]
Cephalosporins (general)		Topics of a general nature that have been considered and are related to the analysis of this class of antibiotics include the separation of various compounds on polymeric ion-exchange resins,[1] comparison of chromatographic methods for purity evaluation of them,[2] measurement of the retention behavior of various cephalosporins,[3] and the use of 2-nitrophoenylhydrazine[4] and second-derivative spectra[5] for determining them.
Cefepime	Spectrometric	Diffuse reflectance mid-IR and X-ray diffraction methods have been developed for characterizing the dihydrate in the monohydrate cefepime · 2HCl drug substance. Excellent correlation was obtained between the two techniques with detection limits better than 1% for both.[6]
Cefaclor	Chromatographic	Two gradient elution HPLC methods useful for identifying impurities, degradation products and formulation excipients have been reported.[7,8] In the latter instance, both photodiode array and MS multidimensional detection approaches were used. HPLC microbore conditions have been used to separate the analyte, defazolin, deftizoxime and cefaloridine.[9] The effects of temperature and eluent composition on the capacity factors of the analyte and the related cephalosporins, cefonicid, cephazolin, cefodizime, cephaloridine, cefamandole, and cephalotin, have been studied.[10]
	Spectrometric	A stability-indicating colorimetric method has been reported where the analyte can be measured in the presence of its degradation products by heating it in an alkaline sorbitol–zinc ion solution for 10–25 min at 90 °C, followed by treatment of the resulting solution with 0.1 N sodium hydroxide to form the piperazine-2,5 derivative, and measurement of the absorbance at 334 nm.[11] Two colorimetric methods have been described that are based on either the oxidation of the target compound with sodium hypochlorite or its reaction with 1-chlorobenzotriazole. Both obey Beer's law and are applicable for dosage forms.[12] Likewise, cefaclor has been measured as its 1 : 1 Ni(II) complex which forms in weakly acidic media but hydrolyzes to a hydroxo complex under basic conditions.[13]

Table 1 (*continued*)

Compound	Technique	Method/comments[a]
Cefadroxil	Chromatographic	An isocratic ion-pairing HPLC method has been developed for assaying three cephalosporins (cefadroxil, cephalexin and cefaclor), and two anti-TB drugs (isoniazid and pyrazinamide). The separations are carried out on a C_{18} column using tetrabutylammonium hydrogensulfate (0.025 M) in combination with pH 3.0 methanol–acetonitrile–triethylamine.[14]
		The analyte has been assayed in capsules by reversed-phase HPLC that employs a C_{18} column, 90 : 10 water–acetonitrile adjusted to pH 3.5 with phosphoric acid as the eluent, *p*-hydroxyphenylacetamide as an internal standard and UV detection at 254 nm.[15] The detection limits for the procedure are 0.8 μg mL^{-1}.
		A comparative study between an HPLC method and a microbiological procedure has been reported. Recoveries were greater than 97% with coefficients of variation below 0.7%.[16] Similarly, multicenter trials have been carried out to evaluate an alternative chromatographic approach.[17]
		Eighteen different C_{18} stationary phases and copolymer columns have been evaluated by European Pharmacopoeia and USP HPLC protocols and a recommendation made that poly(styrene–divinylbenzene) columns be used for measuring the analyte.[18]
	Spectrometric	Tablets and capsules containing the analyte have been assayed by one of three approaches that involve treatment with (1) 3-methyl-2-benzothiazoline hydrazone, HCl and Ce(IV) in sulfuric acid, (2) 4-aminophenazone and potassium hexacyanoferrate or (3) 2,6-dichloro-quinone-4-chlorimide. The absorbance of the resulting colorimetric products are monitored at 410, 510 and 620 nm, respectively.[19] In addition, cefadroxil has been measured colorimetrically as its Cu(II) and V(V) complexes[20] and with ethylene blue[21] as well as using two of the procedures reported above for cefaclor.[11,12]
		In addition to the above colorimetric approaches, cefadroxil, and the related antibiotics cefamandol, cefoxitin, cefapirin and ceftriaxone, have been studied using circular dichroic spectroscopy.[22]
Cefixime	Chromatographic	A robotic procedure has been developed for evaluating the content uniformity of cefixime in oral dosage products.[23]
	Spectrometric	The analyte has been converted to the hydroxamic acid analog with 0.3 mL of methanolic 6.25% hydroxylamine hydrochloride–6.25% NaOH and subsequently to a colored 1 : 3 (metal-to-ligand) complex using methanolic 0.2% Fe(III) perchlorate–1.6% $HClO_4$. The absorbance of the resulting product, which obeys Beer's law from 80 to 320 μg mL^{-1}, is measured at 525 nm.[24]
Cefotaxime	Chromatographic	The analyte has been assayed in the presence of its dimer and major degradation products using CE. Results compare with those obtained under reversed-phase HPLC conditions.[25,26]
	Spectrometric	The analyte has been measured at 228.6 nm using its first-derivative UV spectrum. The procedure can be carried out in the presence of corresponding degradation products.[27] In addition, see the colorimetric procedure for cefixime above.[24]
Cefradine	Chromatographic	An isocratic HPLC method has been described for the analyte that uses a polystyrene–divinylbenzene column operated at 50 °C and an eluent of 14.5 : 10.5 : 75 acetonitrile: 0.02 M Na octanesulfonate: 0.2 M phosphoric acid: water.[28]
Ceftazidime	Chromatographic	A reversed-phase procedure has been developed for studying the storage stability of plastic syringes containing the analyte. The chromatographic separations are carried out on an Ultrasphere-ODS column and detection is at 254 nm.[29] Likewise, ceftazidime has been measured densitometrically following its thin-layer separation.[27]
	Spectrometric	The analyte has been measured at 268.6 nm using its first-derivative UV spectrum. The procedure can be carried out in the presence of corresponding degradation products.[27]
Ceftiofur	Chromatographic	High molecular weight impurities have been determined in ceftiofur using aqueous-based size-exclusion chromatography. Linear responses were observed from 0.005 to 9.25 μg of the impurities with detection limits of 0.03%.[30]

(*continued overleaf*)

Table 1 (*continued*)

Compound	Technique	Method/comments[a]
Ceftriaxone	Chromatographic	A high-performance thin-layer method has been introduced for measuring milligram quantities of the analyte. Recoveries and linearity are acceptable and agree with those obtained by the USP method.[31]
	Electrochemical	Ceftriaxone has been measured electrochemically using cathodic stripping voltammetry[32,33] and differential pulse polarography.[34] In the latter instance it can be determined in the presence of cefobid.
	Spectrometric	See the colorimetric procedure for cefixime above, which is also useful for assaying this analyte.[34]
Cefuroxime	Chromatographic	Cefuroxime has been determined in mixtures that also contain aminophylline and theophylline by reversed-phase LC. The method is capable of assaying as little as 7.5 μg mL^{-1} of the analyte in the presence of 16 and 13 μg mL^{-1} of the co-agents, respectively.[35]
	Spectrometric	The analyte has been measured at 306 nm using its first-derivative UV spectrum. The procedure can be carried out in the presence of corresponding degradation products.[27]
		Other spectrometric methodology that has appeared includes the use of near-IR reflectance[36] and spectrofluorimetric[37] spectroscopy for the quality control of the analyte in tablets. Likewise, a number of other related properties of the analyte may be found elsewhere.[38]
Cephalexin	Chromatographic	Cephalexin has been analyzed in capsules by a reversed-phase HPLC method that uses a C_{18} column, 90 : 10 water–acetonitrile adjusted to pH 3.5 with phosphoric acid as the eluent, *p*-hydroxyphenylacetamide as an internal standard and UV detection at 254 nm. The detection limits are 0.8 μg mL^{-1}.[15]
		The analyte also has been measured in tablets using a C_{18} column, 50 : 50, 0.1% orthophosphoric acid–methanol eluent adjusted to pH 7.5 with triethylamine, methyl paraben as the internal standard and UV detection at 254 nm.[39]
		An alternative HPLC approach has been suggested for measuring the content and stability of the analyte in commercial products.[40] A follow-up nine-laboratory comparative study has demonstrated that this method gives results that compare favorably with a commonly used microbiological method.[41]
	Electrochemical	The analyte has been assayed by first degrading it in alkaline media followed by adsorptive stripping voltammetry.[42] In addition to the quantitation aspects, this work also focuses on developing a better understanding of the mechanism of degradation and the electrode reaction processes.
	Spectrometric	A spectrophotometric procedure has been developed for the analyte and the results obtained using it have been found to compare favorably with those obtained by LC. Additionally, both cephalexin and its acid-induced degradation product are not affected by the presence of excipients.[43]
		Another method has been reported that is based on reacting cephalexin with acetylacetone–formaldehyde, which produces a yellow chromophore that is stable for up to 3 h. This approach is specific for β-lactam antibiotics.[44] Cephalexin also has been determined by alternative colorimetric procedures that involve its reaction with imidazole reagent[45] and with alkaline Co(III).[46]
Cephalothin	Chromatographic	A comparison of LC methods has been reported for the analyte[47] and it has been one of the compounds used as part of a general study that examines postcolumn chemiluminescence as a detection technique for β-lactam ring containing antibiotics in the LC.[48] A number of penicillins also are included.
Cepharadrine	Chromatographic	A multilaboratory comparison of a pharmacopeia bulk drug substance method has been carried out.[49]
	Spectrometric	Cepharadrine can be assayed by the colorimetric method for cefaclor discussed above, using a λ_{max} of 345 nm[11] and by an alternative procedure.[50]

[a] IR, infrared; CE, capillary electrophoresis; C_{18}, octadecyl-modified silica; UV, ultraviolet; USP, United States Pharmacopeia; MS, mass spectrometry; LC, liquid chromatography.

which is hydrolyzed by β-lactamase. These similarities are key to developing analytical methods for this class of drugs. In view of the chemical structure, it is not unusual that β-lactamase enzyme is capable of hydrolyzing the β-lactam ring of both the cephalosporins and penicillins. The carboxylic group is important to both families of compounds. Antibacterial activity and stability (penicillins vs cephalosporins) is defined by the stereospecificity and geometry of the fused side ring. Side chains and stereochemistry play an important role in determining their antibacterial activity.

A listing of the more common cephalosporins and related analytical methodology that has been developed for assaying them is summarized in Table 1.

2.2 Penicillins

The name penicillin was coined in 1929 by Professor Fleming. Penicillins comprise a large family of natural and semisynthetic agents fermented by *Penicillium* and other soil-inhabiting fungi. The penicillins are noted for bactericidal rather than simply bacteriostatic properties and are nontoxic. The unifying chemical structure among them is the central seven-membered thiazolidine–β-lactam ring system. The nucleus common to all penicillins is the fused thiazolidine–amino-β-lactam that consists of L-cysteine and D-valine. The free carboxyl group of the thiazolidine is essential for therapeutic effectiveness. In 1940 several researchers embarked on a scientific endeavor to identify and isolate this class of compounds and as the result of this important pioneering work the penicillins were instrumental during World War II for treating battle casualties.

Crystalline penicillin drug substance and formulated products are sufficiently pure and stable when stored at room temperature and in a dry state. However, all compounds/products deteriorate rapidly at elevated temperatures or when exposed to moisture. The fundamental course of degradation is hydrolysis of the β-lactam ring.

Table 2 Methodology for measuring penicillins

Compound	Technique	Method/comments[a]
Penicillins (general)		Several penicillins have been studied by mercurimetric titration with potentiometric end-point detection. The consecutive formation of reaction products during the titrations were monitored by reversed-phase LC and reaction schemes defined. In the case of benzylpenicillin, at the end of titration reaction, benzylpenicilloaldehyde and a 1 : 1 complex of mercury and penicillamine were present.(51) In addition the titrimetric determination of this group of antibiotics using 2-iodoxybenzoate has been considered.(52)
	Chromatographic	A reversed-phase HPLC method has been developed for the analyte that uses a C_{18} column, 35% methanol in 1 mM potassium hydrogenphosphate buffer as the eluent and postcolumn chemiluminescence for detection.(53) A mixture of seven penicillins have been separated on a C_8 column using gradient elution conditions. During this investigation a variety of flow programs were examined in order to obtain optimal performance in terms of minimal baseline drift and maximum response for the pulsed amperometric detector used.(54) Other general chromatographic reports include a study of the retention properties of monobasic penicillins under ion-pairing reversed-phase conditions,(55) an investigation of their separation on polymeric ion-exchange resins,(56) and an LC/electrochemical method for assaying a variety of compounds.(57)
	Electrochemical	The electrochemical oxidation of penicillins at a gold electrode in a flow injection analysis system also has been reported.(58)
	Spectrometric	General colorimetric procedures have been reported for penicillins that use Ellman's reagent(59) and 2-nitrophenylhydrazine.(60)
Aminopenicillinic acid	Chromatographic	See the general reversed-phase HPLC method above.(53)
Amoxycillin	Chromatographic	Amoxycillin has been assayed by a combined isocratic HPLC approach similar to USP method 23 and a gradient HPLC elution approach. Eleven laboratories took part in evaluating the methods and found that reproducibilities were better than 1.3% and that the isocratic approach was most useful for content control and the gradient approach was more suitable for purity control.(61) The analyte has been simultaneously measured with clavulanic acid using a β-cyclodextrin stationary phase, 65 : 35 pH 4.5 acetate buffer–methanol containing tetraethylammonium acetate as the eluent and detection at 225 nm.(62) A reversed-phase isocratic method has appeared for assaying the analyte in pharmaceutical formulations that also contain cloxacillin and it is capable of simultaneously detecting manufacturing precursors and acid hydrolysis decomposition products of both drugs. The separations are carried out on a C_{18} column using an 85 : 15 20 mM pH 2.0 phosphate buffer–acetonitrile eluent with 100 mM SDS eluent and UV detection at 230 nm.(63) Micellar electrokinetic capillary chromatography(64) and HPLC/MS(65) methods have been developed for evaluating the analyte and its potential impurities. In the first instance the separation is carried out at 15 kV and 25 °C using a 50 μM ID × 44 cm fused-silica capillary and 70 mM pH 6.0 sodium phosphate containing 125 mM SDS and 5% acetonitrile as the run buffer. Detection is at 230 nm.

(*continued overleaf*)

Table 2 (*continued*)

Compound	Technique	Method/comments[a]
	Spectrometric	The analyte has been determined colorimetrically after treating it with dehydroascorbic acid. The approach is specific for penicillins having amino acid side chains and is more accurate and precise than compendia methods.[66] Alternatively, the analyte has been measured at 390 nm following its reaction with acidic formaldehyde at 90 °C for 1 h. The method can be applied to a variety of dosage forms without interferences from excipients, coloring agents and flavor additives.[67] Likewise, acetylacetone–formaldehyde also has been used to assay the analyte in combination products without interference of cloxacillin and excipients.[68] Several other colorimetric methods have been reported for assaying the analyte. These are based on either oxidizing it with sodium hypochlorite or reacting it with 1-chlorobenzotriazole,[69] formation of a chloranil charge-transfer complex[70] and treatment with acenaphthenequinone.[71]
	Miscellaneous	Additional reversed-phase,[72–75] gel permeation,[76] micellar electrokinetic capillary,[77] electrochemical,[78,79] spectrometric,[80,81] and stopped-flow fluorimetric[82] methods may be found elsewhere.
Ampicillin	Chromatographic	An interlaboratory study evaluated several LC assay procedures and found that the USP method was the most selective and that cefradine was a better internal standard than caffeine.[83] However, in a second study involving seven sites, consistency across locations was not good.[84] A LC method has been published that uses luminol chemiluminescence detection to obtain enhanced sensitivity.[85] Additional HPLC conditions for this analyte also are available.[86]
	Electrochemical	Both direct current polarography and differential pulse polarography have been used to assay the analyte in pharmaceutical dosage forms. Prior to carrying out the electrochemical measurements, the analyte is converted to its nitroso derivative using nitrous acid. Linearity of the calibration curves is 8–200 and 2–160 $\mu g\,mL^{-1}$ respectively.[87] Alternative polarographic[88] and conductometric titration[89] methods also have been used to assay the analyte.
	Spectrometric	The analyte has been determined colorimetrically after treating it with dehydroascorbic acid. The approach is specific for penicillins having amino acid side chains and is more accurate and precise than compendia methods.[90] Alternatively, the analyte has been measured at 380 nm following its reaction with acidic formaldehyde at 90 °C for 1 h. The method can be applied to a variety of dosage forms without interferences from excipients, coloring agents and flavor additives.[91] The analyte also has been assayed colorimetrically after formation of a chloranil charge-transfer complex,[70] reaction with acetylacetone–formaldehyde to form a colored product that is measured at 400 nm and obeys Beer's law between 8 and 140 $\mu g\,mL^{-1}$[92] and treatment with other reagents such as dehydroascorbic acid[90] and nitrobenzene.[93] A second-derivative UV procedure has appeared for measuring the analyte and cloxacillin. The formulated product is dissolved in water and spectral readings are made between 260 and 300 nm. The derivative measurements for the two antibiotics are at 269 and 288 nm, respectively.[94] Alternatively, a hydrolysis method has been developed which is specific for ampicillin in the presence of cloxacillin.[95] The suitability of near-IR spectroscopy as an alternative to several compendial tests has been examined and has been found to be able to replace identification, water content and assay for release of ampicillin trihydrate bulk drug substance procedures.[96]
Azlocillin	Chromatographic	A stability-indicating reversed-phase HPLC method has been developed which is capable of resolving degradants formed when the analyte is stored at elevated temperatures.[97]
Benzylpenicillin	Chromatographic	A study has been carried out to examine seven isocratic reversed-phase methods for separating benzylpenicillin and related substances on C_{18} or C_8, phenyl and poly(styrene–divinylbenzene) columns. The manufacturer's method, which uses a C_8 column and 62 : 38 0.05 M pH 3.5 phosphate buffer–methanol as the eluent, gave the best selectivity whereas the selectivity on poly(styrene–divinylbenzene) was poorest.[98] In another study, isocratic reversed-phase HPLC was examined by seven laboratories with the recommendation that it should be combined with gradient elution conditions for purity control.[99] The analyte and related impurities have been assayed by micellar electrokinetic chromatography using a fused-silica capillary operated at 25 °C and 12 kV with pH 6.5 40 mM NaH_2PO_4–150 mM-SDS running buffer and detection at 225 nm.[100]

Table 2 (*continued*)

Compound	Technique	Method/comments[a]
	Electrochemical	Both direct current polarography and differential pulse polarography have been used to assay the analyte in pharmaceutical dosage forms. Prior to carrying out the electrochemical measurements, the analyte is converted to its nitroso derivative using nitrous acid. Linearity of the calibration curves are 8–200 and 2–160 μg mL^{-1}, respectively.(101)
	Spectrometric	A differential colorimetric method has been developed that is based on the initial hydrolysis of the analyte and its subsequent reaction with ferricyanide.(102) The procedure also can be used for ampicillin and benzathine penicillin.
		An indirect spectrophotometric assay has been introduced for formulations containing the analyte and cloxacillin that involves the acid hydrolysis of the analyte followed by the extraction of the products into carbon tetrachloride and measurements made at 520 nm.(103)
		The reaction of the analyte with potassium iodate in acidic media has been investigated and a mechanism proposed.(104)
Carbenicillin	Electrochemical	Both direct current polarography and differential pulse polarography have been used to assay the analyte in pharmaceutical dosage forms. Prior to carrying out the electrochemical measurements, the analyte is converted to its nitroso derivative using nitrous acid. Linearity of the calibration curves are 8–200 and 2–160 μg mL^{-1}, respectively.(101)
Clavulanic acid	Chromatographic	The analyte has been measured by the general reversed-phase(53) and amoxycillin(62) methods presented above.
	Electrochemical	A polarographic method has appeared that can be used to quantify the analyte in the presence of amoxicillin. Clavulanic acid is first hydrolyzed with sulfuric acid and then measured at −0.75 V vs SCE.(105)
	Spectrometric	A stopped-flow fluorimetric kinetic assay has been developed for the simultaneous determination of clavulanic acid and amoxycillin.(106) The method involves the acidic reaction of the analytes with Ce(IV) and detection at 365 nm.
		Near-IR reflectance spectroscopy has been used for the quality control of cream formulations. Samples are extracted with light petroleum after initially dissolving them in methanolic sodium hydroxide.(107)
Cephalothin	Chromatographic	See the general reversed-phase HPLC methodology presented above.(53)
Cloxacillin	Chromatographic	Results have appeared from a collaborative LC study for determining cloxacillin in bulk drug, capsules and injectables that had coefficients of variations of less than 1%.(108) Also see the reversed-phase procedures for amoxycillin above(63) and elsewhere.(109)
	Spectrometric	See the second-derivative(94) and hydrolysis(95) UV methods for ampicillin and spectrometric method for benzylpencillin(102) above.
		Thermospray MS has been used to investigate the solvent degradation of the analyte which has been found to undergo hydrolysis followed by decarboxylation when exposed to aqueous solutions containing either acetonitrile or propanol or esterification when stored in either methanol or ethanol.(110)
Dicloxacillin	Chromatographic	See the general reversed-phase HPLC method above(53) and those appearing elsewhere.(111,112) In the latter instance, the eight laboratories involved in the comparative study reported coefficients of variation for repeatability of less than 0.6% and a reproducibility of 2.6%.
Flucloxacillin	Chromatographic	See the HPLC method for cloxacillin above.(113)
Methicillin	Chromatographic	See the general reversed-phase HPLC method above.(53)
Nafcillin	Spectrometric	The fluorescence intensity of nafcillin at 336 nm has been studied as a function of temperature and it has been concluded that control is essential in making reliable measurements since the intensity fell 1.1% for every 1 °C increase.(113) The analyte has been determined in mixtures that contain methicillin by constant-wavelength synchronous fluorimetry in combination with a partial least-squares multivariate calibration.(114) Besides these methods, the solid state of the analyte in frozen aqueous solutions has been studied by low-temperature X-ray powder diffractometry.(115)
Oxacillin	Chromatographic	See the HPLC method for cloxacillin above.(109)
Penicillamine	Chromatographic	A chiral method has been reported for the analyte that involves its derivatization with *o*-phthalaldehyde and 2-mercaptoethanol followed by separation on a β-cyclodextrin column.(116)
	Electrochemical	A cathodic voltammetric method has been reported that can be used to measure D-penicillamine in the presence of *N*-acetylcysteine and Ni(II).(117)

(*continued overleaf*)

Table 2 (*continued*)

Compound	Technique	Method/comments[a]
	Spectrometric	Penicillamine has been determined after first degrading it to penicillaminic acid using hypobromite and subsequently treating the residual hypobromite with fluorescein and measuring the absorbance at 517 nm.[118]
		Three methods have been suggested for the determination of D-penicillamine in bulk and in capsules which involve oxidation with ferrocyanide and complexation of the iron with either bipyridyl or phenanthroline.[119] These approaches can be used to evaluate the target analyte in the presence of its degradation products and other penicillins.
Penicillin	Chromatographic	See the general reversed-phase HPLC methodology above.[53–57]
Penicillin	Spectrometric	Two flow-injection methods have been developed for the analyte and tiopronin that are based on the oxidation of these drugs by thallium(III) in hydrochloric acid medium and measurement of the thallium(I) formed fluorimetrically at 419 nm.[120] Additional flow-injection methods are available.[121]
		Other spectrometric-based methods that have been reported for the analyte include accounts related to the fabrication of optic chemical sensors[122,123] and derivative spectroscopy.[124,125]

[a] C_8, octyl-modified silica; SDS, sodium dodecyl sulfate; SCE, saturated calomel electrode; ID, inner diameter.

Table 3 Methodology for measuring common quinolones

Compound	Technique	Method/comments[a]
Quinolones (general)	Chromatographic	A study has been carried out to evaluate the capacity factor of 22 fluoroquinolones. Results from the work demonstrate a linear relationship between capacity factor and antibacterial activity.[126] In addition, the influence of pH on the reversed-phase separation of fluoroquinolones[127] has been investigated and HPTLC conditions useful for resolving a variety of fluoroquinolones have been published.[128]
Ciprofloxacin	Chromatographic	LC methods have been described that can be used to quantify ciprofloxacin in the presence of its photodegradation products[129] and also in combination with its ethylenediamine analog or fluoroquinolonic acid[130] and chlorofluoroaniline, dichlorofluoroacetophenone, cyclopropyl acrylate and quinolinic acid.[131] Likewise, a reversed-phase method has been developed for measuring the analyte in tablet formulations that uses a C_{18} column.[132]
		Alternative chromatographic approaches that have been used include a GC method for measuring the analyte and also norfloxacin, sarafloxacin and temafloxacin as their *N*-acyl derivatives[133] and a TLC procedure for evaluating the photodegradation of ciprofloxacin in aqueous solutions.[134]
	Electrochemical	A ciprofloxacin PVC-coated wire electrode has been fabricated that is based on a molecular dispersion of 4-quinolones with dioctyl phthalate as the polymerization solvent mediator. The electrode has a Nernstian response over the range 0.1–100 nM and has been used to measure the analyte in pharmaceuticals via a standard addition technique.[135]
		The polarographic behavior of ciprofloxacin has been studied as a function of pH and a value of 8.5 has been found to be best for carrying out measurements at reduction potentials of −1.44 and −1.64 V.[136]
	Spectrometric	The UV photodecomposition of the analyte in acidic ethanol has been studied by several techniques and the brown crystals that formed were identified to be 7-amino-1-cyclopropyl-6-fluoro-1,4-dihydro-4-oxo-3-quinolonecarboxylic acid. A reversed-phase HPLC method for evaluating the purity of the analyte also is included in the reported work.[137]
		A colorimetric procedure has been described that is based on extracting the yellowish-orange product obtained by the cerium(IV) oxidation of the drug in acidic medium and measuring its absorbance at 345 nm. The color of the organic layer is stable for 4 h.[138] In addition, the dissociation and complexation behavior of the analyte and those of enoxacin, norfloxacin and ofloxacin have been examined.[139]
		Likewise, ciprofloxacin, ofloxacin and norfloxacin have been measured at 370 nm following treatment with Fe(III) nitrate nonahydrate.[140] In addition, several other spectrophotometric methods have appeared for ciprofloxacin that use either methyl orange, bromothymol blue,[141] *p*-benzoquinone,[142] 3-methylbenzothiazolin-2-one with ceric ammonium sulfate[143] or iron[144,145] as colorimetric reagents. In the last instance measurements were carried out by flow injection analysis at a rate of 250 samples h^{-1} with a precision of <0.9%.
Enoxacin	Electrochemical	The voltammetric and polarographic behavior of the analyte has been studied at a mercury electrode and the reduction, which occurs at 1.052 V vs Ag/AgCl, has been found to be irreversible and to form an adsorptive product.[146] An alternative polarographic procedure also has been reported for enoxacin.[147]

Table 3 (*continued*)

Compound	Technique	Method/comments[a]
	Spectrometric	The analytical methodology for ciprofloxacin discussed above[129] can be used to assay enoxacin.
Enrofloxacin	Chromatographic	Enrofloxacin has been quantified in tablets using a C_{18} column, 70 : 30 acetonitrile–water containing 0.1% triethylamine adjusted to pH 4 as the eluent and detection at 282 nm.[148]
	Spectrometric	The analyte also has been measured colorimetrically using either Folin–Ciocalteau reagent or Fe(III).[149]
Flumequine	Electrochemical	Tablets containing the analyte have been assayed polarographically using Britton–Robinson buffer containing methanol as the solubilizer. Well defined cathodic waves are obtained in the pH range 5–10.[150]
	Spectrometric	Flumequine also has been determined spectrophotometrically by direct chelation with iron(III) and treatment of the analyte with iron(III) hydroxamate followed by chelation with iron(III).[151]
Lomefloxacin	Chromatographic	An HPTLC method has been reported for the analyte that uses silica gel plates in combination with 18 : 2 : 1 methanol–water–ammonia as the development solvent. Scanning is carried out at 286 nm using afloxacin as an internal standard.[152]
	Spectrometric	Lomefloxacin has been determined in ground tablets by the formation of colored complexes using acidic solutions of either bromothymol blue, bromophenol blue, bromocresol purple or bromocresol green. The colorimetric measurements of chloroform extracts of the resulting complexes are made respectively at 416, 416, 422 and 412 nm.[152] Likewise, similar methodology also has been reported for measuring lomefloxacin and ofloxacin in another account[153] and it has been assayed with Folin–Ciocalteau reagent.[154]
Norfloxacin	Chromatographic	Both reversed-phase HPLC[155,156] and TLC[157] methods have been reported for measuring the analyte in combination with tinidazole.
	Electrochemical	A variety of electrochemical methods has been reported for norfloxacin that include DC and differential pulse polarography,[158] differential pulse stripping voltammetry,[159] and oscillopolarographic titrimetry.[160]
	Spectrophotometric	Tablets and eye drops containing the analyte have been assayed via initially treating them with 0.1 N H_2SO_4, filtering, mixing with ammonium sulfamate and ceric ammonium sulfate and extracting the yellow chromogen into chloroform. Measurements are made at 350 nm.[161] An alternative colorimetric method has appeared for studying the dissolution of the analyte that is based on its reaction with acidic Fe(III).[162] In addition to these procedures, other spectrometric methods can be found elsewhere;[163–166] also see the spectrometric methodology for ciprofloxacin above.[139,140]
Ofloxacin	Spectrometric	See the colorimetric[140] and related methodology[139] for ciprofloxacin and lomefloxacin above.[153] Other spectrometric methods[167] have been developed for ofloxacin, including an NMR assay.[168]
	Electrochemical	Assays have been reported based on adsorptive stripping,[169] potentiometric,[170] and related electrochemical[171] measurements.
Pefloxacin	Spectrometric	A direct UV method has been described for assaying tablets and composites that involves dilution of an equivalent amount of the powdered material in 0.1 N NaOH and measuring the resulting spectrum from 250 to 290 nm. An orthogonal polynomial function method is used to eliminate irrelevant absorption due to excipients.[172] Alternative UV and spectrofluorimetric procedures have been reported which involve an initial extraction of the tables with methanol and readings made under acidic conditions.[173] The analyte also has been determined colorimetrically following its treatment with iron(III).[174]
Rufloxacin	Chromatographic	The analyte has been examined using a reverse-phase HPLC procedure.[175]
	Spectrometric	A simple and rapid fourth-derivative UV method has been developed for rufloxacin hydrochloride and its main impurity in laboratory mixtures, bulk materials and pharmaceutical forms.[176] Likewise, a fluorimetric assay has been developed for measuring the analyte in pharmaceutical formulations.[177]
Temafloxacin	Chromatographic	HPLC procedures have been reported for determining minor impurities in temafloxacin hydrochloride[178] and for assaying the analyte in bulk drug and dosage forms.[179] In the latter instance, the procedure can be used for difloxacin and sarafloxacin.
	Spectrometric	A fluorimetric assay has been described for the determination of temafloxacin in pharmaceutical formulations. The samples are diluted with 0.1 M sulfuric acid and measured at 460 nm using an excitation wavelength of 276 nm.[180]

[a] HPTLC, high-performance thin-layer chromatography; GC, gas chromatography; TLC, thin-layer chromatography; PVC, poly(vinyl chloride); NMR, nuclear magnetic resonance.

This decomposition is catalyzed by penicillinase and various metals.

A listing of some of the more common penicillins and related analytical methodology that has been developed for assaying them is summarized in Table 2. In many cases HPLC, UV, colorimetric and microbiological approaches are commonly used.

2.3 Quinolones

The quinolones are a newer class of antibiotics and work on them continues to expand. The first compound discovered in the class was nalidixic acid and subsequently other related compounds have been synthesized based upon in vitro results and structure–activity relationships. The quinolones inhibit DNA synthesis in bacterial cells but not in mammalian cells of DNA topoisomerase. Likewise, the quinolones have a diverse ability to extinguish the micro-bioactivity of microbes.

The number of published methods for this class of compounds continues to increase. Although chromatographic and spectroscopic based methods tend to dominate, the compounds are also electrochemically active. A listing of some of the more common quinolones and related

Table 4 Methodology for measuring streptomycins and related analogs

Compound	Technique	Method/comments[a]
Abamectin	Chromatographic	A postcolumn normal-phase HPLC method has appeared for abamectin. Following the separation the analyte is reacted with naphthalene-2-sulfonic acid in an alkaline nonaqueous mixing solvent and the absorbance monitored at 570 nm.[181]
Amikacin	Chromatographic	The analyte has been assayed in the 2–100 $\mu g\,mL^{-1}$ range with relative standard deviations of 0.3–4.2% using a C_{18} column following its derivatization with *o*-phthalaldehyde. The effluent absorbance is monitored at 340 nm.[182]
	Electrochemical	A nonspecific conductometric procedure has been developed to monitor the strength of pediatric injectable dextrose solutions that contain amikacin, netilmicin, and tobramycin prepared by a centralized pharmacy service. The major limitation of the approach is that it cannot be applied to solutions that contain sodium chloride.[183]
		Thirteen aminoglycoside antibiotics have been determined by capillary zone electrophoresis using cetyltrimethylammonium bromide and indirect UV detection.[184]
Bekanamycin	Chromatographic	Bekanamycin and other related aminoglycoside antibiotics, including gentamicin, tobramycin, lincomycin, neomycin, kanamycin, and ribostamycin, have been measured in pharmaceutical preparations by CE in combination with EC. In the study both copper wall-jet electrodes and copper ion-capillary electrodes were evaluated and the latter were found to provide superior stability and reproducibility.[185]
Bleomycin	Chromatographic	A rapid reversed-phase method has been reported for measuring bleomycins in commercial products which uses sodium perchlorate as an ion-pairing reagent.[186]
		The luminescence of bleomycins in combination with Tb(III) has been studied and Tb(III) luminescence detection has been found to be compatible with a wide range of HPLC eluents.[187]
Clarithromycin	Chromatographic	An HPLC method has been developed that can resolve 15 nonpolar related compounds and process impurities that elute after *N*-demethyl-*N*-formyl-6-*O*-methylerythromycin A in clarithromycin bulk drug. All of the nonpolar compounds can be detected at the 0.02% (w/w) level.[188]
		Several other HPLC methods have been reported for clarithromycin.[189–191] In one of these, EC was done using a dual electrode which provided a detection limit of 100 pg[188] and in another case postcolumn reaction with ruthenium(III) and subsequent luminescence detection produced a calibration graph that was linear over four orders of magnitude.[191]
		HPLC with EC also has been used to study clarithromycin residues on various surfaces. After cleaning, any residual clarithromycin remaining can be removed from the surface by a wet swab–dry swab technique.[192]
Clindamycin	Spectrometric	Clindamycin has been determined spectrophotometrically as its ion-pair complex with Rose Bengal in a slightly basic medium. The colored product that forms at pH 7.5 is measured at 555 nm. Likewise, the analyte has been assayed by direct conductometric titration using Rose Bengal in a 2 : 1 ethanol–water mixture.[193]
		Alternative spectrometric and titrimetric methods also have been described for clindamycin. In the first case, the analyte is reacted with potassium iodate in an acidic medium, the liberated iodine extracted with cyclohexane, and the resulting solution measured at 520 nm. In the second method a direct titration of the drug is carried out with *N*-bromosuccinimide.[194]
Dirithromycin	Chromatographic	An HPLC method has been reported for dirithromycin that uses a C_{18} column, an eluent consisting of 44 : 19 : 37 acetonitrile–methanol–50 mM pH 7.5 potassium phosphate buffer, and UV detection at 205 nm.[195]

Table 4 (*continued*)

Compound	Technique	Method/comments[a]
Erythromycin	Chromatographic	The analyte has been separated under isocratic conditions using a C_{18} column and an acetonitrile–pH 6.5 phosphate buffer eluent containing tetrabutylammonium sulfate. The impurities most frequently observed in commercial samples were the propionate ester of erythromycin C and the amide, *N*-propionyl-*N*-demethylerythromycin A.[196]
		A multicenter study involving six laboratories and seven columns has been carried out and a method has been recommended for the analyte that uses a wide-pore poly(styrene–divinylbenzene) stationary phase.[197]
		An alternative study has been carried out that examines several C_{18} columns for their use in separating erythromycin and a gradient elution method has been developed for separating the analyte from seven possible impurities. The method is stated to be simpler, more rugged, faster, and more sensitive than a competing chromatographic procedure that uses a poly(styrene–divinylbenzene) stationary phase and it yields results similar to those obtained by microbiological procedures.[198]
		In addition to the above methodology, a host of other HPLC-based procedures have appeared for erythromycin. Examples of these may be found elsewhere.[199–203]
		The analyte also has been assayed by CE using a 50 mM pH 7.5 phosphate run buffer and an applied voltage of 20 kV. Likewise, 35% ethanol incorporated into a modified 150 mM pH 7.5 phosphate run buffer and an applied voltage of 30 kV have been used to facilitate the separation of erythromycin from erythromycin C, anhydroerythromycin, and *N*-demethylerythromycin A.[204]
	Spectrometric	Erythromycin and related macrolide antibiotics such as oleandomycin, spiromycin, and tylosin have been measured in fermentation broths after treatment with Fe(III). The product that forms in acetic acid–sulfuric acid is measured at 592 nm.[205] Additional colorimetric procedures have been reported that use bromothymol blue, methylthymol blue, and thymol blue as reagents.[206]
		Erythromycin has been determined spectrofluorimetrically via formation of an ion pair with erythrosine B and measuring the chloroform extract at 560 nm using an excitation wavelength of 544 nm. The method can be adapted to an unsegmented flow-injection system and used to assay the analyte in formulations.[207]
Gentamycin	Chromatographic	Injectable solutions containing gentamicin sulfate have been assayed by CE and direct UV detection using a borate buffer.[208] Likewise, the analyte has been analyzed using the CE/EC method for bekanamycin above[185] and by LC thermospray MS.[209] In the latter instance a simplex algorithm was used to optimize the separation.
Gramicidin	Chromatographic	Formulations containing the analyte have been evaluated by a reversed-phase procedure that involves an initial dilution step with aqueous 35% THF, a subsequent separation at 65 °C using an 8 mm PLRP-S column and 38 : 62 THF–water as the eluent and detection at 222 nm.[210]
Kanamycin	Chromatographic	A reversed-phase LC method has appeared for resolving kanamycin sulfate from kanamycins B, C, and D and from other impurities. The procedure involves a two-step gradient separation on a poly(styrene–divinylbenzene) column using aqueous eluents of 20 or 60 g L^{-1} sodium sulfate, 1.3 g L^{-1} sodium octanesulfonate and 50 mL/10.2 M pH 3.0 phosphate buffer and pulsed EC on a gold electrode following the postcolumn addition of sodium hydroxide.[211] The analyte also has been assayed by the CE/EC method described above for bekanamycin[212] as well as by flow injection methodology reported elsewhere.[212] In the latter case, the procedure is linear from 0.1 pM to 10 μM and common excipients do not interfere.
Lincomycin	Chromatographic	Samples containing the analyte have been analyzed after diluting them appropriately with 0.1 M NaOH and separating the resulting solutions on an 80-cm capillary using 0.1 M NaOH as the electrophoretic medium, an applied voltage of 26 kV and wall-jet amperometric detection at +0.675 V vs Ag/AgCl. Using this approach, lincomycin and lincomycin B can be separated within 15 min with a resolution of 2.3.[213] Lincomycins also have been measured by the CE/EC procedure for bekanamycin discussed above[185] as well as procedures appearing elsewhere.[214]
Moenomycin	Chromatographic	An HPTLC procedure has been developed for evaluating the purity of moenomycin A, B and C isolated from the fermentation broth of *Streptomyces bambergiensis.* Separations are carried out on silica gel high-performance plates activated at 110 °C for 30 min and developed using 13 : 7 1-propanol–25% aqueous ammonia. Subsequently, the plates are visualized by treatment with 1 : 2 acetic acid–chlorosulfonic acid, heating them at 100 °C for 5–7 min and scanning in reflectance mode at 550 nm.[215]

(*continued overleaf*)

Table 4 (*continued*)

Compound	Technique	Method/comments[a]
Neomycin	Chromatographic	A reversed-phase procedure has been reported for neomycin sulfate that uses a poly(styrene–divinylbenzene) copolymer column, a 0.2 M pH 3 phosphate buffer containing 70 g L^{-1} sodium sulfate and 1.4 g L^{-1} sodium 1-octanesulfonate as the eluent, and pulsed EC.[216] The relative amounts of the B and C components of neomycin sulfate have been evaluated using silica gel TLC plates, 15 : 85 methanol–20% aqueous sodium chloride, and fluorescence detection after derivatization with 4-chloro-7-nitrobenzo-2-oxa-1,3-diazole.[217] Likewise, an alternative HPTLC method has been described for neomycin A, B, and C that uses a development solvent of 35 : 20 : 20 : 5 methanol–25% aqueous ammonia–acetone–chloroform, and detection with fluorescamine.[218] The analyte also has been assayed by the CE/EC procedure for bekanamycin discussed above.[185]
	Spectrometric	An alternative procedure to the standard agar diffusion method for neomycin has been developed that uses triphenyltetrazolium chloride.[219] Likewise, the analyte has been measured indirectly by atomic absorption spectroscopy following its metal complexation.[220]
Nybomycin	Chromatographic	A combination of TLC, HPLC, and MS methods have been used to study nybomycin isolated from fermentation broths.[221]
Paromomycin	Chromatographic	Similarly, a variety of methods have been used to assay paromomycin sulfate including NMR spectroscopy, elemental analysis, ion chromatography, GC and LC.[222] In the last case, the separation was performed on a C_8 column using 0.2 M Na_2SO_4–1.2 mM sodium 1-heptanesulfonate–0.1% acetic acid as the eluent and fluorescence detection following postcolumn derivatization with *o*-phthalaldehyde and 2-mercaptoethanol. Additionally, LC/MS was performed on a C_{18} column with 2 : 3 methanol–0.7% aqueous heptafluorobutyric acid as the eluent.[222]
Ribostamycin	Chromatographic	Ribostamycin has been assayed by the CE/EC procedure for bekanamycin discussed above.[185]
Rifamycin	Spectrometric	A first-derivative method has been reported for the simultaneous determination of rifamycin SV sodium in injection solutions that also contain lidocaine. Readings are made at 312.0 and 254.6 nm.[223]
Roxithromycin	Chromatographic	Roxithromycin has been measured in tablets by HPLC. The reversed-phase separations are carried out using a C_{18} column, 35 : 65 methanol–0.067 M pH 4.0 phosphate buffer as the eluent, and UV detection at 210 nm.[224]
Streptomycin	Chromatographic	The analysis of streptomycin and 11 related aminoglycoside antibiotics has been carried out by CE utilizing borate complexation and direct UV detection. Quantification of dihydrostreptomycin and streptomycin was accomplished using 160 mM sodium tetraborate decahydrate and the determination of the purity of bulk dihydrostreptomycin was possible by the addition of the cationic surfactant, myristyltrimethylammonium bromide. Other compounds studied included amikacin, bekanamycin, kanamycin A, and tobramycin.[225] In addition, alternative GC[226] and TLC[227] procedures have appeared for the analyte.
	Electrochemical	The analyte and related compounds have been measured at ruthenium dioxide-modified graphite electrodes via an electrocatalytic process.[228]
	Spectrometric	Streptomycin and five other aminoglycoside antibiotics, neomycin, gentamicin, tobramycin, amikacin and kanamycin, have been measured fluorimetrically following their reaction with Eu(III) and monitoring the intensity ratio $R = I_{592\,nm}/I_{616\,nm}$, for free and bound europium ions, which is related to the concentration of the aminoglycoside. The results obtained by this procedure compare favorably with those obtained by a fluorimetric method that uses fluorescamine reagent.[229] In addition, the analyte and relate antibiotics have been assayed colorimetrically using cyanoacetamide.[230]
Spectinomycin	Chromatographic	An isocratic HPLC assay has appeared for spectinomycin dihydrochloride pentahydrate that is based on cation-exchange chromatography on a Dionex IonPac CS3 column at ambient temperature using 150 mM pH 6.0 sodium acetate. The method, which employs pulsed-amperometric detection, is suitable for the analysis of the drug substance.[231]
Tobramycin	Chromatographic	A rapid and sensitive LC method has been developed to quantify the release of tobramycin from polymeric drug delivery systems in vitro. Precolumn derivatization of tobramycin and the internal standard, kanamycin B sulfate, are carried out with 2,4,6-trinitrobenzenesulfonic acid and the products are separated under reversed-phase conditions using a C_8 column.[232] In addition, the CE/EC methodology for bekanamycin discussed above can be used to assay tobramycin.[185]
	Spectrometric	See the spectrometric methodology for streptomycin discussed above[229] and alternative procedures appearing elsewhere.[233]
Tylosin	Chromatographic	See literature.[234–236]
	Spectrometric	See literature.[237]

[a] EC, electrochemical detection.

Table 5 Methodology for measuring common sulfonamides and other miscellaneous antibiotics

Compound	Technique	Method/comments[a]
Sulfonamides (general)	Chromatographic	The reversed-phase behavior of seven common sulfonamides on a C_{18} column using acetonitrile–phosphate buffer as the eluent have been investigated. A number of different conditions, including buffer concentration, type and concentration of positively and negatively charged pairing reagents, column temperature and the ionic state of the solutes have been evaluated and the optimal eluent suggested for each compound.(238)
		Systematic studies have been carried out to characterize the general migration properties of many common sulfonamides under various electrophoretic conditions.(239–241) Similarly, HPTLC has been used to evaluate several sulfonamides.(242)
		A general procedure has been developed for assaying several common sulfonamides in pharmaceutical preparations that involves their initial derivatization to form the corresponding *N*-(1-naphthyl)ethylenediamine dihydrochloride azo dyes and subsequent reversed-phase micellar separation using a pH 7 0.05 M SDS eluent containing 2.4% pentanol.(243)
	Spectrometric	The colorimetric reagent *p*-benzoquinone has been used to measure a variety of sulfonamides in bulk and in dosage forms.(244) Likewise, photochemically induced fluorescence has been found to be a useful approach,(245) as has FTIR spectroscopy.(246)
Salazosulfapyridine	Electrochemical	The voltammetric properties of salazosulfapyridine have been studied by several polarographic procedures at both mercury dropping and static drop electrodes at various pHs. The analyte, which is reduced to a hydrazine derivative, can be determined by scanning differential pulse voltammetry and adsorptive stripping voltammetry in tablet formulations with relative standard deviations of 5% or better and detection limits in the range 0.2–0.5 ng mL^{-1}.(247)
Sulfadimetoxal	Electrochemical	The voltammetric behavior of the analyte has been studied by square-wave techniques and two methods have been developed for measuring it in aqueous samples and formulations.(248)
Sulfadimidine	Electrochemical	The analyte can be measured after reacting it with either salcylaldehyde or cinnamaldehyde. The optimal performance is reached in 30 min at 50 °C.(249)
	Spectrometric	Flow injection analysis methods based on fluorimetric or photochemically induced fluorescence have been developed for the analyte and also for sulfanilimide, sulfaguanidine and sulfacetamide.(250)
Sulfadiazine	Chromatographic	Reversed-phase conditions in combination with EC have been used to assay the analyte and also sulfamerazine and phthalylsulfathiazole. The calibration curves are rectilinear for all three drugs over the range 10–90 μM.(251) Likewise, the analyte has been assayed using an alternative HPLC method.(252)
	Spectrometric	The analyte has been determined colorimetrically by reacting it with dimethylamino-cinnamaldehyde and measuring the absorbance of the red chromogen that forms at 545 nm. The method is linear over the range 0.4–4.8 μg mL^{-1}.(253)
		A continuous-flow diazotization approach has been used for determining the analyte in dosage forms and for measuring its dissolution rates. An advantage is that it eliminates the frequent preparation of in situ generated nitrate solutions.(254)
		Samples containing the analyte are assayed by diluting them with methanol, irradiating the solution with a mercury lamp for 2.5 min, and reading the resulting photochemically induced fluorescence at 353 nm. The method is linear over the concentration range 0.1–4.1 μg mL^{-1}.(255) A similar procedure also can be used to measure sulfamethazine in the presence of sulfamerazine or sulfadiazine.(256)
		The analyte has been measured in the presence of nonsulfonamide antibiotics using a spectrophotometric partial least-squares multivariate method.(257) The same approach can be applied to sulfamerazine, sulfamethazine, sulfaquinoxaline and sulfathiazole.(258) In addition, the analyte has been determined by direct absorbance monitoring.(259)
Sulfaguanidine	Chromatographic	See the HPLC method for sulfadiazine above.(252)
Sulfamerazine	Chromatographic	See the HPLC method for sulfadiazine above.(252)
	Electrochemical	An electrochemical study of the oxidation of sulfamerazine has been carried out at a glassy carbon electrode and based on this an HPLC assay for the analyte, sulfadiazine and phthalylsulfathiazole has been developed.(251)
	Spectrometric	See the spectrometric method for sulfadiazine above.(258)
Sulfamethazine	Spectrometric	The influence of solvents on the photochemical induced fluorescence of the analyte has been examined. Detection limits of 80, 40 and 70 ng mL^{-1} were achievable when methanol, ethanol and 2-propanol were used, respectively.(260) Also see the fluorimetric method for sulfadiazine above.(256)

(*continued overleaf*)

Table 5 (*continued*)

Compound	Technique	Method/comments[a]
Sulfamethizole	Spectrometric	The analyte has been assayed in tablets using its ratio derivative spectrum[261] and after its reaction with *p*-benzoquinone.[244]
Sulfameth-oxazole	Chromatographic	The analyte and trimethoprim have been isolated from a formulation using a supercritical fluid extraction approach.[262] Both direct extraction of the aqueous-based samples and indirect extraction of the formulation from Celite have been evaluated.
	Electrochemical	The voltammetric properties of the analyte have been studied by several polarographic procedures at both mercury dropping and static drop electrodes and its reduction which takes place with difficulty is by a radical mechanism. For additional information, see salazosulfapyridine above.[247]
	Spectrometric	The analyte can be assayed in various formulations by either first-derivative or classical least-squares analysis of the UV methanolic spectra. In the first approach, the absorbance is measured at the zero-crossing wavelength, 288 nm, and in the second the spectrum is recorded from 200 to 350 to 200 nm.[263] The analyte also has been assayed using its ratio derivative spectrum[264] and fluorimetrically as its 1 : 1 β-cyclodextrin inclusion complex.[265]
Sulfametrole	Spectrometric	The analyte has been determined using its first-derivative spectrum.[266]
Sulfanilamide	Spectrometric	The analyte has been determined in binary mixtures that also contain sulfathiazole using the third- and fourth-order derivative spectra.[267] This method can be applied to other sulfanilamides.
Sulfathiazole	Chromatographic	See the HPLC method for sulfadiazine above.[252]
	Spectrometric	Sulfathiazole has been measured using its ratio derivative spectrum[267] and by a first-derivative solid-phase spectrophotometric approach.[268]
Tinidazole	Electrochemical	The analyte can be determined by a polarographic procedure with a well-defined sigmoid-shaped curve with an $E_{1/2}$ of −0.44 V vs SCE using a pH 5.5 Britton–Robinson buffer.[269]
	Spectrometric	The analyte has been assayed in combination with clotrimazole or norfloxacin using a derivative zero crossing point technique. Tablet extracts are prepared in 0.01 M NaOH and readings made between 220 and 300 nm.[270]
Chloram-phenicol	Chromatographic	LC/thermospray MS has been used to measure the analyte and three related compounds.[271]
	Electrochemical	The redox behavior of chloramphenicol under different pH conditions has been studied using a graphite electrode and parallel incident spectroelectrochemistry. Different reaction products have been found to be produced depending on pH.[272]
		The analyte in eye drops containing it were measured in 0.1 M NaOH solution using an Ni/C modified working electrode; a sensitive reductive wave of chloramphenicol was obtained by linear sweep voltammetry. The peak potential was −0.80 V (SCE). The reduction process was quasi-reversible.[273]
		A chloramphenicol sensor has been fabricated that is based on the displacement of an analyte–dye conjugate from a molecularly imprinted polymer. The sensor is immune to interference from analyte-related derivatives and is linear from 3 to 1000 $\mu g\,mL^{-1}$.[274]
Clotrimazole	Spectrometric	The analyte has been determined in combination with tinidazole using a derivative zero crossing point technique. Tablet extracts are prepared in 0.01 M NaOH and readings made between 220 and 300 nm.[270]
Isoniazid	Chromatographic	An isocratic ion-pairing HPLC method has been developed for assaying three cephalosporins (cefadroxil, cephalexin and cefaclor) and two anti-TB drugs (isoniazid and pyrazinamide). The separations are carried on a C_{18} column using tetrabutylammonium hydrogensulfate (0.025 M) in combination with pH 3.0 methanol–acetonitrile–triethylamine.[14]
	Spectrometric	The analyte has been measured colorimetrically after treatment with host of reagents such as azoic diazo component 36,[275] 2,3,5-triphenyltetrazolium chloride,[276] Fe(III) phenanthroline,[277] ethyl-8-quinolinoxy acetate[278] and radiochloramine-B.[279]
Trimethoprim	Chromatographic	Trimethoprim has been measured simultaneously with nitrofuranotion using a micro C_{18} column, a pH 6 water–acetonitrile eluent containing triethylamine and UV detection at 270 nm.[280] Likewise, an alternative HPLC approach has been used to evaluate the analyte in the presence of its degradation products.[281]
	Electrochemical	The voltammetric properties of the analyte have been studied by several polarographic procedures at both mercury dropping and static drop electrodes and its reduction is by either a two- or four-electron process (depending on the acidity) that leads to the di- or tetrahydropyrimidine derivative as a final product. For additional information, see salazosulfapyridine above.[247]

Table 5 (*continued*)

Compound	Technique	Method/comments[a]
	Spectrometric	The analyte has been assayed in various formulations by either first-derivative or classical least-squares analysis of the UV methanolic spectra. In the first approach, the absorbance is measured at the zero-crossing wavelengths, 240 nm, and in the second the spectrum is recorded from 200 to 350 nm.[263] Other first-derivative,[266] ratio-derivative[264] and linear absorbance[282] methods also have been reported. The analyte can be measured in formulations containing sulfamethoxazole via oxidizing it with alkaline persulfate. The stable reaction product is formed after heating for 30 min and it is measured at 355 nm. Beer's law is obeyed in the concentration range $10-60\,\mu g\,mL^{-1}$.[283]

[a] FTIR, Fourier transform infrared.

Table 6 Methodology for measuring common tetracyclines

Compound	Technique	Method/comments
Tetracyclines (general)	Chromatographic	Commercial formulations have been extracted and the analytes separated on a C_{18} column using 84 : 16 0.05 M pH 2.5 phosphate buffer–acetonitrile as the eluent and measured amperometrically at 1.2 V using a glassy carbon electrode. This approach has been demonstrated to be useful for the simultaneous determination of trace amounts of chlortetracycline, demeclocycline, doxycycline, methacycline, minocycline, oxytetracycline and tetracycline and also for the separation of their common impurities (epi, anhydro and epianhydro contaminants).[284] A systematic study of capillary electrophoretic operating conditions has been carried out and optimized methods for the analysis of various tetracyclines have been reported. Different buffer systems are suggested for the separation of tetracycline, oxytetracycline and demeclocycline from their respective major impurities, including the 2-acetyl-2-decarboxamido derivatives. The nonionic surfactant Triton X-100 and methyl-β-cyclodextrin were found to be useful in obtaining improved selectivity in the case of oxytetracycline and demeclocycline.[285] Additional capillary electrophoretic methodology may be found elsewhere[286,287] The micellar electrokinetic migration behavior of six tetracyclines has been investigated in the pH range 5.0–9.0 using ammonium acetate buffers containing SDS. In addition, several other surfactants also were studied. Complete separation of the analytes were achieved in 8 min on a fused-silica capillary using a 15 mM ammonium acetate buffer containing 20 mM SDS at 15 kV. The use of this procedure to obtain information about octanol–water partition coefficients is also discussed.[288]
	Spectrometric	A flow injection assay has been developed that uses aminophenazone and $Fe(CN)_6$ as reagents and it has been used to assay doxycycline, oxytetracycline, rolitetracycline and tetracycline with relative standard deviations <1% for samples in the $10-100\,\mu g\,mL^{-1}$ range.[289]
	Miscellaneous	Other topics of a general nature that have been reported include the analysis of various tetracyclines using simplex optimized flow injection analysis,[290] the use of micellar solutions to enhance the luminescence,[291] the europium-sensitized luminescence of tetracyclines[292] and the loss of tetracyclines via their adsorption on laboratory glassware.[293]
Chlortetracycline	Chromatographic	A study has been carried out involving five laboratories and six different columns to demonstrate that the proposed HPLC method could be used to determine the analyte and also major related substances such as demeclocycline and 4-epichlortetracycline.[294] In two additional cross-laboratory studies, both poly(styrene–divinylbenzene) stationary phases[295] and C_{18} phases[296] were found to be useful for assaying the analyte with variations among the different laboratories between 1 and 2%. Similarly, the usefulness of polymeric stationary phases for separating the analyte and four related substances has been demonstrated in another study.[297] Other analytical methodologies that have been reported for chlortetracycline are a TLC purity method that has been found to be comparable to LC approaches,[298] an HPLC procedure for resolving the keto–enol tautomers of the analyte and 4-epi-chlortetracycline in combination with UV and NMR spectroscopy to elucidate their structures.[299]

(*continued overleaf*)

Table 6 (*continued*)

Compound	Technique	Method/comments
	Spectrometric	The analyte has been measured by solid-surface room temperature phosphorescence[300] and by chemiluminescence.[301,302] In the latter instance, a flow injection approach was employed based on the reaction of the chlortetracycline with $KMnO_4$ in acidic medium in the presence of a micellar agent. It is linear from 1 to 1000 $\mu g\,mL^{-1}$ and has a detection limit of 0.60 $\mu g\,mL^{-1}$ for the analyte and 0.40 $\mu g\,mL^{-1}$ for tetracycline.
Demeclocycline	Chromatographic	This analyte has been examined by HPLC,[303] TLC (see the method for chlortetracycline)[298] and CE.[304] In the last approach doxycycline is detectable at the 0.3% level and measurable at 0.4% in the presence of its degradation products.
	Spectroscopic	The analyte has been determined colorimetrically with sodium molybdate.[305]
Doxycycline	Chromatographic	A CE method has been developed for doxycycline that is capable of measuring it in the presence of possible impurities.[306] Likewise, the TLC fluorimetric methodology discussed above for chlortetracycline can be used to evaluate the purity of the analyte.[298]
	Spectrometric	See the methodology for chlortetracycline above.[300]
Metacycline	Chromatographic	Assays for metacycline have been introduced that use TLC in combination with UV/fluorescence densitometric detection to evaluate its purity.[298,307]
	Spectrometric	See the methodology for chlortetracycline above.[300]
Minocycline	Chromatographic	Both capillary zone electrophoresis[308] and HPLC have been used to measure minocycline.[309,310] In the latter case, acidic eluents in combination with poly(styrene–divinylbenzene) columns are used to resolve minocycline and seven other tetracycline derivatives.
	Electrochemical	The voltammetric behavior of the analyte at a water–nitrobenzene interface has been investigated by normal, semi-differential and semi-integral cyclic voltammetry. Resulting from the work was the development of a semi-differential cyclic voltammetric procedure for measuring the analyte in the range 5–200 $\mu g\,mL^{-1}$ with results comparable to those obtained by the USP method.[311]
Oxytetracycline	Chromatographic	Oxytetracycline has been separated from related impurities and degradation products as its magnesium complex under nonaqueous capillary electrophoretic conditions. The method has been used to assay ointments with high selectivity and good precision.[312] In addition, other CE approaches have been reported for the analyte.[287,313,314] Results from a comparison study of past HPLC methods for the analyte have indicated that the use of polymeric stationary phases gives the most promising results.[315] Likewise, TLC methodology has been reported for oxytetracycline and doxycycline that is useful for evaluating their purity.[316] The separations are carried out on silica gel thin-layer plates using 59 : 35 : 6 dichloromethane–methanol–water with 10% sodium acetate solution adjusted to pH 9.0 as the development solvent.
	Spectrometric	The analyte has been assayed by first extracting it into cyclohexane and then determining it by direct UV measurements in combination with orthogonal polynomial coefficient corrections to minimize interference from excipients.[317] Other spectrometric procedures described for oxytetracycline include one based on its reaction with sodium molybdate,[318] a flow injection approach that uses a carrier stream of iron(III),[319] and its reaction with hydrogen peroxide catalyzed by copper ion to form a chemiluminescent product.[301]
Tetracycline	Chromatographic	The ruggedness of the USP HPLC method for tetracycline has been evaluated. A statistical analysis of column aging and its effect on the separation performance suggests that a C_8 column is to be preferred over a C_{18} column.[320,321] Capillary electrophoretic assays have been developed for measuring the analyte in the presence of its degradation impurities that are based on conventional aqueous[322] and nonaqueous[323] approaches. Also see the electrophoretic general methodology above.[287] Several comparative studies have been carried out for methods related to tetracycline including one where four laboratories evaluated seven different columns for suitability,[324] another where a TLC method for the analyte was compared with an HPLC method[325] and a final study involving capillary electrophoretic methodology.[326] Other related TLC[327] and HPLC[328] methods appear elsewhere.
	Spectrometric	See the methods for chlortetracycline above.[301,302] In the latter method the limit of detection for the analyte is 0.40 $\mu g\,mL^{-1}$. In addition to these continuous flow procedures, other spectrometric methods,[329,330] as well as a postcolumn time-resolved luminescence method for HPLC may be found elsewhere.[331]

analytical methodology that has been reported for them is summarized in Table 3.

2.4 Streptomycin and Related Analogs

The compounds that fall within this category were originally synthesized as organic bases by appropriate strains of actinomycete (*Streptomyces griseus*). The first drug isolated from actinomycetes was streptomycin in 1943. Other compounds have loosely been included in this category because they also were biosynthesized via similar microbial routes.

Streptomycin itself is strongly basic with an empirical formula of $C_{21}H_{39}N_7O_{12}$. It and related compounds are stable when dry and they retain full potency at room temperature. However, the presence of moisture will affect their stability. Solutions are normally stable at room temperature in the pH range 3–7 and optimal solution stability is typically between pH 6 and 7. Exposure to light may darken solutions and impact purity.

Analytical methodology that has been developed for assaying streptomycin and its related analogs is summarized in Table 4.

2.5 Sulfonamides and Miscellaneous Antibacterial Agents

Summarized in Table 5 are representative analytical procedures that are useful for assaying a number of the more common sulfonamides and also miscellaneous compounds that possess antibacterial activity. In addition to the wide variety of chromatographic and spectrometric methodology that has appeared, a number of electrochemical assays have been published for this grouping of compounds and various studies have been carried out to characterize their electrochemical behavior.

2.6 Tetracyclines

The tetracyclines comprise a family of both biosynthetic and semisynthetic derivatives that are characterized by an unusual antimicrobial background. They are therapeutically used against both Gram-positive and Gram-negative bacteria and various viruses. However, these compounds (Table 6) are not active against the common cold or fungi. The first tetracycline (7-chlorotetracycline) was discovered in 1948 and it was the second compound to be developed after chloramphenicol as a broad-based antibiotic. Four other tetracyclines were developed from natural sources (duomycin, oxytetracycline, demeclocyclin and tetracycline). Subsequently, rolitetracycline, methacycline, doxycycline and minocycline were produced as semisynthetic derivatives.

All of the tetracyclines have a common skeleton (i.e. an octahydro analog of naphthacene) but differ in their substituents, which has a significant influence on their individual solubilities. The tetracyclines are amphoteric and can be manufactured as both acidic and basic salts. Differences are observed in their stability at elevated temperatures in both acidic and alkaline media, which can have a significant impact on the bioavailability of an individual compound when administered orally. In the dry state both the free compound and its corresponding salts are stable at room temperature when protected from light and moisture. In solution, the compounds undergo both acid and basic hydrolysis (ring cleavage) but the degree of cleavage is dependent upon the particular drug. Bulk drug substance and drug product should be placed in opaque light-protecting bottles and, if optimal stability is desired, a desiccant added to the container.

Summarized in Table 6 are representative procedures that have been developed to assay the tetracyclines. In general, this methodology parallels trends noted above for the other antibiotics.

ABBREVIATIONS AND ACRONYMS

C_8	Octyl-modified Silica
C_{18}	Octadecyl-modified Silica
CE	Capillary Electrophoresis
EC	Electrochemical Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
ID	Inner Diameter
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PVC	Poly(vinyl chloride)
SCE	Saturated Calomel Electrode
SDS	Sodium Dodecyl Sulfate
TLC	Thin-layer Chromatography
USP	United States Pharmacopeia
UV	Ultraviolet

RELATED ARTICLES

Pharmaceuticals and Drugs **(Volume 8)**
Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis • Nuclear Magnetic Resonance Spectroscopy

in Pharmaceutical Analysis • Planar Chromatography in Pharmaceutical Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis • Vibrational Spectroscopy in Drug Discovery, Development and Production

REFERENCES

1. J.W. Kelly, J.T. Stewart, 'Separation of Selected Beta-lactam Antibiotic Epimers on Gamma-cyclodextrin, Ion-exchange Ethylvinylbenzene–Divinylbenzene Copolymer and Poly(styrene–divinylbenzene) Copolymer Stationary Phases', *J. Liq. Chromatogr.*, **14**(12), 2235–2250 (1991).
2. H. Fabre, A.F. Fell, 'Comparison of Techniques for Peak Purity Testing of Cephalosporins', *J. Liq. Chromatogr.*, **15**(17), 3031–3043 (1992).
3. K. Kovacs-Hadady, J. Szilagyi, *J. Planar Chromatogr. Mod. TLC*, **4**(5–6), 194–198 (1991).
4. M.A. Korany, M.H. Abdel-Hay, M.M. Bedair, A.A. Gazy, 'Colorimetric Determination of Some Penicillins and Cephalosporins with 2-Nitrophenylhydrazine Hydrochloride', *Talanta*, **36**(12), 1253–1257 (1989).
5. J.A. Murillo, J. Rodriguez, J.M. Lemus, A. Alanon, 'Determination of Amoxicillin and Cephalexin in Mixtures by 2nd-derivative Spectrophotometry', *Analyst*, **115**(8), 1117–1119 (1990).
6. D.E. Bugay, A.W. Newman, W.P. Findlay, 'Quantitation of Cefepime-2HCl Dihydrate in Cefepime-2HCl Monohydrate by Diffuse Reflectance IR and Powder X-ray Diffraction Techniques', *J. Pharm. Biomed. Anal.*, **15**(1), 49–61 (1996).
7. L.J. Lorenz, F.N. Bashore, B.A. Olsen, 'Determination of Process-related Impurities and Degradation Products in Cefaclor by High-performance Liquid-chromatography', *J. Chromatogr. Sci.*, **30**(6), 211–216 (1992).
8. B.A. Olsen, S.W. Baertschi, R.M. Riggin, 'Multidimensional Evaluation of Impurity Profiles for Generic Cephalexin and Cefaclor Antibiotics', *J. Chromatogr.*, **648**(1), 165–173 (1993).
9. C.M. Moore, K. Sato, K.Y. Katsumata, 'High-performance Liquid-chromatographic Determination of Cephalosporin Antibiotics Using 0.3 mm/id Columns', *J. Chromatogr.*, **539**(1), 215–220 (1991).
10. J. Martin-Villacorta, R. Mendez, 'Effect of Temperature and Mobile Phase Composition on RP-HPLC Separation of Cephalosporins', *J. Liq. Chromatogr.*, **13**(16), 3269–3288 (1990).
11. L.I. Bebawy, K. Kelani, L. Abdel Fattah, 'Stability-indicating Method for the Determination of Some Cephalosporins in the Presence of Degradation Products', *Spectrosc. Lett.*, **30**(2), 331–343 (1997).
12. M.I. Walash, S. Toubar, S.M. Ahmed, N. Zakhari, 'Spectrophotometric Studies on Determination of Some Cephalosporins and Amoxicillin with Chlorobenzotriazole and Sodium Hypochlorite', *Anal. Lett.*, **27**(13), 2499–2513 (1994).
13. A. Dimitrovska, B. Andonovski, K. Stojanoski, 'A Spectrophotometric Study of the Formation of Nickel(II) Ion Complexes with Cefaclor', *Anal. Lett.*, **29**(6), 937–951 (1996).
14. Y.P. Patel, U.J. Dhorda, M. Sundaresan, A.M. Bhagwat, 'Isocratic, Simultaneous HPLC Determination of Five Antibiotics by Ion-pair Chromatography', *Indian Drugs*, **34**(1), 43–47 (1997).
15. V.M. Shinde, C.V. Shabadi, 'Simultaneous Determination of Cefadroxil and Cephalexin in Capsules by Reversed-phase HPLC', *Indian Drugs*, **34**(7), 399–402 (1997).
16. M.C. Hsu, Y.W. Chang, Y.T. Lee, 'Column Liquid Chromatography and Microbiological Assay Compared for Determination of Cefadroxil Preparations', *J. Chromatogr.*, **609**(1–2), 181–186 (1992).
17. J. Diago, C. Hendrix, J. Hoogmartens, J.H.M. Miller, S. Tedeschi, C. Van der Vlies, *Pharmeuropa*, **4**(2), 144–149 (1992).
18. C. Hendrix, C. Wijsen, L.M. Yunn, E. Roets, J. Hoogmartens, 'Column Liquid Chromatography of Cefadroxil on Poly(styrene–divinylbenzene)', *J. Chromatogr.*, **628**(1), 49–58 (1993).
19. C.S.P. Sastry, K.R. Rao, D.S.F. Prasad, 'Determination of Cefadroxil by Three Simple Spectrophotometric Methods Using Oxidative Coupling Reactions', *Mikrochim. Acta*, **126**(1–2), 167–172 (1997).
20. S.S. Badawy, F.M. Abdelgawad, M.M. Ibrahim, 'Spectrophotometric Studies on Determination of Cefadroxil with Copper(II) and Vanadium(V) in Sulfuric Acid Medium', *Anal. Lett.*, **26**(3), 487–497 (1993).
21. A.A. Alwarthan, F.H. Metwally, S.A. Al-Tamimi, 'Spectrophotometric Assay of Certain Cephalosporins based on Formation of Ethylene Blue', *Anal. Lett.*, **26**(12), 2619–2635 (1993).
22. P. Gortazar, J. Vazquez, 'Discrimination and Direct Determination of Cephalosporins by Circular Dichroism', *J. Pharm. Sci.*, **83**(9), 1204–1208 (1994).
23. J.M. Rosen, M. O'Leary, W. Fotino, R.R. Ryall, L. Gehrlein, *Adv. Lab. Autom. Robot.*, **5**, 363–379 (1989).
24. D. Agbaba, S. Eric, K. Karljikovic-Rajic, S. Vladimirov, D.P. Zivanov-Stakic, 'Spectrophotometric Determination of Certain Cephalosporins Using Ferrihydroxamate Method', *Spectrosc. Lett.*, **30**(2), 309–319 (1997).
25. H. Fabre, G. Castaneda Penalvo, 'Capillary Electrophoresis as an Alternative Method for the Determination of Cefotaxime', *J. Liq. Chromatogr.*, **18**(18–19), 3877–3887 (1995).
26. G. Castaneda Penalvo, E. Julien, H. Fabre, 'Cross Validation of Capillary Electrophoresis and High-performance Liquid Chromatography for Cefotaxime and Related Impurities', *Chromatographia*, **42**(3–4), 159–164 (1996).

27. K. Kelani, L.I. Bebawy, L. Abdel-Fattah, 'Stability-indicating Spectrophotometric and Densitometric Methods for Determination of Some Cephalosporins', *J. AOAC Int.*, **81**(2), 386–393 (1998).
28. C. Hendrix, M. Pijcke, L.M. Yun, E. Roets, J. Hoogmartens, 'Quantitative Analysis of Cefradine by Liquid Chromatography on Poly(styrene–divinylbenzene)', *J. Pharm. Biomed. Anal.*, **11**(2), 145–155 (1993).
29. M.C. Nahata, R.S. Morosco, 'Measurement of Ceftazidime Arginine in Aqueous Solution by HPLC', *J. Liq. Chromatogr.*, **15**(9), 1507–1511 (1992).
30. M.J. Dunn, D.A. Hahn, 'Aqueous High-performance Size-exclusion Chromatographic Assay for High-molecular-weight Impurities in Ceftiofur Sodium', *J. Chromatogr.*, **595**(1–2), 185–192 (1992).
31. S.S. Zarapkar, S.A. Shivalkar, A.A. Dhavante, P.M. Deshpande, S. Kolte, 'High-performance Thin-layer-chromatographic Determination of Ceftriaxone Sodium from Pharmaceutical Preparation', *Indian Drugs*, **32**(5), 232–235 (1995).
32. N.A. El Maali, A.M.M. Ali, M. Khodari, M.A. Ghandour, 'Cathodic Stripping Voltammetric Determination of the Cephalosporin Antibiotic Ceftriaxone at the Mercury Electrode in Aqueous and Biological Media', *Bioelectrochem. Bioenerg.*, **26**(3), 485–492 (1991).
33. S. Altinoz, A. Temizer, 'Differential Pulse Adsorptive Stripping Voltammetric Determination of Ceftriaxone at a Static Mercury Dropping Electrode', *J. Pharm. Sci.*, **79**(4), 351–353 (1990).
34. N.A. El-Maali, A.M.M. Ali, M.A. Ghandour, 'Electrochemical Reduction and Oxidation of 2 Cephalosporin Antibiotics – Ceftriaxone (Rocephin) and Cefoperazone (Cefobid)', *Electroanalysis (NY)*, **5**(7), 599–604 (1993).
35. H. Zhang, J.T. Stewart, 'Determination of a Cefuroxime and Aminophylline–Theophylline Mixture by High-performance Liquid Chromatography', *J. Liq. Chromatogr.*, **15**(17), 3031–3043 (1992).
36. E. Dreassi, G. Ceramelli, L. Savini, P. Corti, P.L. Perruccio, S. Lonardi, 'Application of Near-infrared Reflectance Analysis to the Integrated Control of Antibiotic Table Production', *Analyst*, **120**(2), 319–323 (1995).
37. J.A. Murillo, J.M. Lemus, L. Garcia, 'Spectrofluorometric Analysis of Cefuroxime in Pharmaceutical Dosage Forms', *J. Pharm. Biomed. Anal.*, **12**(7), 875–881 (1994).
38. T.J. Wozniak, J.R. Hicks, 'Cefuroxime', *Anal. Profiles Drug Subst.*, **20**, 209–236 (1991).
39. U.P. Halkar, S.H. Rane, N.P. Bhandari, ' Simultaneous Determination of Cephalexin and Probenecid in Pharmaceutical Preparations by Reversed Phase HPLC', *Indian Drugs*, **34**(9), 539–541 (1997).
40. M.-C. Hsu, Y.-S. Lin, H.-C. Chung, 'High-performance Liquid-chromatographic Method for Potency Determination of Cephalexin in Commercial Preparations and for Stability Studies', *J. Chromatogr. A*, **692**(1–2), 67–72 (1995).
41. M.-C. Hsu, H.-C. Chung, Y.-S. Lin, 'Liquid Chromatographic Determination of Cephalexin Preparations: Interlaboratory Validation', *J. Chromatogr. A*, **727**(2), 239–244 (1996).
42. Q.L. Li, S. Chen, 'Studies on Electrochemical Behavior of Cephalexin', *Anal. Chim. Acta*, **282**(1), 145–152 (1993).
43. P. Campins-Falco, A. Sevillano-Cabeza, L. Gallo-Martinez, F. Bosch-Reig, I. Monzo-Mansanet, 'Comparative Study on the Determination of Cephalexin in its Dosage Forms by Spectrophotometry and HPLC with UV/Vis Detection', *Mikrochim. Acta*, **126**(3/4), 207–215 (1997).
44. I.T. Patel, M.B. Devani, T.M. Patel, 'Spectrophotometric Method for Determination of Cephalexin in its Dosage Forms', *J. AOAC Int.*, **75**(6), 994–998 (1992).
45. A.A. Alwarthan, S. Abdel Fattah, N.M. Zahran, 'Spectrophotometric Determination of Cephalexin in Dosage Forms with Imidazole Reagent', *Talanta*, **39**(6), 703–707 (1992).
46. A.A. Alwarthan, H.A. Al-Lohedan, 'Kinetic Determination of Cephalexin in Drug Formulations', *Talanta*, **41**(2), 225–231 (1994).
47. C. Hendrix, E. Roets, J. Hoogmartens, 'Comparison of Two LC Methods for Analysis of Cefalotin Sodium', *Pharmeuropa*, **5**(2), 150–155 (1993).
48. J.H. Miyawa, S.G. Schulman, J.H. Perrin, 'Postcolumn Chemiluminescence as a Detection Technique in the Liquid Chromatographic Analysis of Beta-lactam Ring Containing Compounds', *Biomed. Chromatogr.*, **11**(4), 224–229 (1997).
49. J. Diego, M. Fischler, C. Hendrix, J. Hoogmartens, J.H. M. Miller, A. Moeller, S. Tedeschi, 'Multi-centre Trial to Assess a Liquid-chromatographic Method for the Determination of Cephradine in Bulk Drug for Pharmacopoeial Use', *Pharmeuropa*, **4**(1), 68–74 (1992).
50. M.B. Devani, I.T. Patel, T.M. Patel, 'Rapid Spectrophotometric Assay of Cepharadine and its Dosage Forms', *Anal. Lett.*, **24**(6), 971–978 (1991).
51. E. Roets, P. Rappe, M. Heeren, J. Hoebus, A. Verbruggen, J. Hoogmartens, 'Investigation of the Reaction Mechanism of the Mercurimetric Determination of Benzylpenicillin', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1141–1149 (1996).
52. F. Belal, A. El-Brashy, F. Ibrahim, 'Titrimetric Determination of Some Penicillins Using 2-Iodoxybenzoate', *J. AOAC*, **73**(6), 896–901 (1990).
53. J.H. Miyawa, S.G. Schulman, J.H. Perrin, 'Postcolumn Chemiluminescence as a Detection Technique in the Liquid Chromatographic Analysis of Beta-lactam Ring Containing Compounds', *Biomed. Chromatogr.*, **11**(4), 224–229 (1997).
54. R.K. Earley, J.S. Miller, L.E. Welch, 'The Application of Flow Programming with Pulsed Amperometric Detection for Liquid Chromatography', *Talanta*, **45**(6), 1255–1266 (1998).

55. D.J. Fletouris, J.E. Psomas, A.J. Mantis, 'Retention Behavior of Some Monobasic Penicillins and their Quaternary Ammonium Ion-pairs in Reversed Phase High Performance Liquid Chromatography', *Chromatographia*, **32**(9–10), 436–440 (1991).
56. J.W. Kelly, J.T. Stewart, 'Separation of Selected Beta-lactam Antibiotic Epimers on Gamma-cyclodextrin, Ion-exchange Ethylvinylbenzene–Divinylbenzene Copolymer and Poly(styrene–divinylbenzene)', *J. Liq. Chromatogr.*, **14**(12), 2235–2250 (1991).
57. E. Kirchmann, L.E. Welch, 'High-performance Liquid Chromatographic Separation and Electrochemical Detection of Penicillins', *J. Chromatogr.*, **633**(1–2), 111–118 (1993).
58. L. Koprowski, E. Kirchmann, L.E. Welch, 'The Electrochemical Oxidation of Penicillins on Gold Electrodes', *Electroanalysis*, **5**(5–6), 473–482 (1993).
59. F.I. Morgil, A. Kestek, 'Analytical Investigation of Penicillins. Part VI. Determination of Some Selected Semisynthetic Penicillins with Ellman's Reagent', *J. Fac. Pharm. Gazi Univ.*, **7**(1), 33–42 (1990).
60. M.A. Korany, M.H. Abdel-Hay, M.M. Bedair, A.A. Gazy, 'Colorimetric Determination of Some Penicillins and Cephalosporins with 2-Nitrophenylhydrazine Hydrochloride', *Talanta*, **36**(12), 1253–1257 (1989).
61. Z. Yongxin, E. Roets, A. Bruzzi, A.J.W. Smelt, C. Van der Vlies, P. Betto, M.L. Moreno, E. Porqueras, H.H. Wendebourg, K. Sinivuo, D. De Kaste, A. Mayrhofer, N. Graby, G. Kister, J.H. Miller, C.J. McB. Nap, J.M. Spieser, J. Hoogmartens, 'Interlaboratory Study of the Analysis of Amoxicillin by Liquid Chromatography', *J. Liq. Chromatogr. Relat. Technol.*, **19**(20), 3305–3314 (1996).
62. T.-L. Tsou, J.-R. Wu, C.-D. Young, T.-M. Wang, 'Simultaneous Determination of Amoxycillin and Clavulanic Acid in Pharmaceutical Products by HPLC with β-Cyclodextrin Stationary Phase', *J. Pharm. Biomed. Anal.*, **15**(8), 1197–1205 (1997).
63. O. Shakoor, R.B. Taylor, 'Analysis of Ampicillin, Cloxacillin and their Related Substances in Capsules, Syrups and Suspensions by High-performance Liquid Chromatography', *Analyst*, **121**(10), 1473–1477 (1996).
64. Y.M. Li, A. Van Schepdael, Y. Zhu, E. Roets, J. Hoogmartens, 'Development and Validation of Amoxicillin Determination by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **812**(1–2), 227–236 (1998).
65. L. Valvo, E. Ciranni, R. Alimenti, S. Alimonti, R. Draisci, L. Giannetti, L. Lucentini, 'Development of a Simple Liquid Chromatographic Method with UV and Mass Spectrometric Detection for the Separation of Substances Related to Amoxicillin Sodium', *J. Chromatogr. A*, **797**(1–2), 311–316 (1998).
66. F.S. El-Shafie, E.A. Gad-Kariem, K.A. Al-Rashood, H.A. Al-Khamees, H.A. El-Obeid, 'Colorimetric Method for the Determination of Ampicillin and Amoxicillin', *Anal. Lett.*, **29**(3), 381–393 (1996).
67. P.P. Shirke, P.C. Balasubramanya, A.R. Samant, V.B. Tirodkar, P. Sethi, 'Simultaneous Spectrophotometric Determination of Ampicillin-amoxicillin and Cloxacillin in Combined Dosage Form', *Indian Drugs*, **31**(8), 370–373 (1994).
68. M.B. Devani, I.T. Patel, T.M. Patel, 'Spectrophotometric Determination of Amoxicillin and its Dosage Forms', *J. Pharm. Biomed. Anal.*, **10**(5), 355–358 (1992).
69. M.I. Walash, S. Toubar, S.M. Ahmed, N. Zakhari, 'Spectrophotometric Studies on Determination of Some Cephalosporins and Amoxicillin with Chlorobenzotriazole and Sodium Hypochlorite', *Anal. Lett.*, **27**(13), 2499–2513 (1994).
70. A.S. Amin, 'Modified Spectrophotometric Method for the Determination of Some Important Antibiotics through Charge-transfer Complexation Reaction with Chloranil', *Spectrosc. Lett.*, **29**(6), 1003–1011 (1996).
71. A.S. Amin, G.O. El Sayed, Y.M. Issa, 'Colorimetric Determination of Ampicillin and 6-Aminopenicillanic Acid Using Acenaphthenequinone as a Chromophoric Reagent', *Anal. Lett.*, **30**(7), 1337–1348 (1997).
72. O. Shakoor, R.B. Taylor, R. Moody, 'Analysis of Amoxicillin in Capsules and Oral Suspensions by High-performance Liquid Chromatography', *Analyst*, **120**(8), 2191–2194 (1995).
73. M.C. Hsu, P.W. Hsu, 'High Performance Liquid Chromatographic Method for Potency Determination of Amoxicillin in Commercial Preparations and for Stability Studies', *Antimicrob. Agents Chemother.*, **36**(6), 1276–1279 (1992).
74. M.A. Abounassif, E.M. Abdel-Moety, M.E. Mohamed, E.R.A. Gad-Kariem, 'Liquid Chromatographic Determination of Amoxicillin and Clavulanic Acid in Pharmaceutical Preparations', *J. Pharm. Biomed. Anal.*, **9**(9), 731–735 (1992).
75. A.F.M. El Walily, F. El-Anwar, M.A. Eid, H. Awaad, 'High Performance Liquid Chromatographic and Derivative Ultraviolet Spectrophotometric Determination of Amoxicillin and Dicloxacillin Mixtures in Capsules', *Analyst*, **117**(6), 981–984 (1991).
76. J. Fernandez, A. Esteban, M. Blanca, A. Ferrer, V. Soriano, 'Ion-pair Reversed Phase and Low Molecular Weight Aqueous Gel Permeation High Performance Liquid Chromatography Methods for the Determination of Amoxicillin Oligomers', *Chromatographia*, **41**(11–12), 651–656 (1995).
77. G.N. Okafo, P. Camilleri, 'Micellar Electrokinetic Capillary Chromatography of Amoxicillin and Related Molecules', *Analyst*, **117**(9), 1421–1424 (1992).
78. S.J. Lyle, S.S. Yassin, 'Polarographic Studies of Metal-ion Complexes of Ampicillin and Amoxicillin', *Anal. Chim. Acta*, **274**(2), 225–230 (1993).

79. E. Kilic, F. Koseoglu, A. Knear, M. Akay, 'Conductimetric Titrimetry for Assay of Selected Antibiotics in Non-aqueous Media', *J. Pharm. Biomed. Anal.*, **13**(12), 1453–1458 (1995).
80. J.A. Murillo, J. Rodriguez, J.M. Lemus, A. Alanon, 'Determination of Amoxicillin and Cephalexin in Mixtures by 2nd-derivative Spectrophotometry', *Analyst*, **115**(8), 1117–1119 (1990).
81. A. Yesilada, M. Ertan, '2nd Derivative Spectrophotometric Determination of Amoxicillin in Oral Suspensions', *Anal. Lett.*, **24**(11), 2033–2041 (1991).
82. P. Izquierdo, A. Gomez-Hens, D. Perez-Bendito, 'Simultaneous Kinetic Flurometric Determination of Amoxicillin and Clavulanic Acid by the Stopped Flow Mixing Technique', *Analyst*, **118**(6), 707–710 (1993).
83. Y. Zhu, E. Roets, Z. Ni, M.L. Moreno, E. Porqueras, J. Hoogmartens, 'Evaluation of Liquid Chromatography Methods for the Separation of Ampicillin and its Related Substances', *J. Pharm. Biomed. Anal.*, **14**(5), 631–639 (1996).
84. Y. Zhu, M.L. Moreno, E. Porqueras, E. Bourke, A. Bruzzi, M. Aletrari, P. Kanari, D. Partasidou, J. Nienhuis, W. Ferigo, J.L. Robert, J.H.M. Miller, J.M. Spieser, E. Roets, J. Hoogmartens, 'Interlaboratory Study of the Analysis of Ampicillin by Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1151–1156 (1996).
85. K. Nakashima, S. Kawaguchi, S. Akiyama, S.G. Schulman, 'High Performance Liquid Chromatography of Penicillins with Penicillin Enhanced Luminol Chemiluminescence Detection', *Biomed. Chromatogr.*, **7**(4), 217–219 (1993).
86. K.A. Al-Rashood, 'Simultaneous Determination of Ampicillin and Dicloxacillin in Pharmaceutical Formulations by High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **18**(12), 2457–2465 (1995).
87. F. Belal, M.S. Rizk, M.J. Eid, 'Polarographic Determination of Some Penicillins through Nitrosation', *J. Pharm. Biomed. Anal.*, **17**(2), 275–282 (1998).
88. S.J. Lyle, S.S. Yassin, 'Polarographic Studies of Metal-ion Complexes of Ampicillin and Amoxicillin', *Anal. Chim. Acta*, **274**(2), 225–230 (1993).
89. E. Kilic, F. Koseoglu, A. Knear, M. Akay, 'Conductimetric Titrimetry for Assay of Selected Antibiotics in Non-aqueous Media', *J. Pharm. Biomed. Anal.*, **13**(12), 1453–1458 (1995).
90. F.S. El-Shafie, E.A. Gad-Kariem, K.A. Al-Rashood, H.A. Al-Khamees, H.A. El-Obeid, 'Colorimetric Method for the Determination of Ampicillin and Amoxicillin', *Anal. Lett.*, **29**(3), 381–393 (1996).
91. P.P. Shirke, P.C. Balasubramanya, A.R. Samant, V.B. Tirodkar, P. Sethi, 'Simultaneous Spectrophotometric Determination of Ampicillin-amoxicillin and Cloxacillin in Combined Dosage Form', *Indian Drugs*, **31**(8), 370–373 (1994).
92. M.B. Devani, I.T. Patel, T.M. Patel, 'Spectrophotometric Determination of Ampicillin and its Dosage Forms', *Talanta*, **39**(10), 1391–1394 (1992).
93. Y.M. Issa, A.S. Amin, 'Spectrophotometric Determination of Ampicillin with Some Nitro Compounds', *Anal. Lett.*, **26**(11), 2397–2407 (1993).
94. A.F. El, F. El-Anwar, M.A. Eid, 'Analysis of Pharmaceutical Formulations Containing Ampicillin and Cloxacillin Using Second-derivative Spectroscopy on a Diode-array Spectrophotometer', *Alexandria J. Pharm. Sci.*, **5**(1), 86–90 (1991).
95. A.O. Akanni, J.S.K. Ayim, 'Determination of Ampicillin in the Presence of Cloxacillin', *J. Pharm. Biomed. Anal.*, **10**(1), 43–47 (1992).
96. W. Plugge, C.J. Van der Vlies, 'Near Infrared Spectroscopy as an Alternative to Assess Compliance of Ampicillin Trihydrate with Compendial Specifications', *J. Pharm. Biomed. Anal.*, **11**(6), 435–442 (1993).
97. M. Pujol, V. Girona, J. Prat, J. Cots, 'HPLC Determination of Azlocillin Sodium for Stability Studies', *Int. J. Pharm.*, **58**(2), 103–106 (1990).
98. Z. Yongxin, A. Verhasselt, E. Roets, A. Perez, E. Porqueras, J. Hoogmartens, 'Evaluation of Liquid Chromatography Methods for the Analysis of Benzylpenicillin and its Related Substances', *J. Chromatogr. A*, **773**(1–2), 147–156 (1997).
99. Z. Yongxin, E. Roets, J. Loidl, P. Inama, A. Perez, E. Porqueras, E. Wachberger, C. Lebbe, A.J.W. Smelt, C. Van Der Vlies, J.H. McB. Miller, J.M. Spieser, J. Hoogmartens, 'Interlaboratory Study of the Analysis of Benzylpenicillin by Liquid Chromatography', *J. Chromatogr. B*, **707**(1–2), 189–193 (1998).
100. Y.X. Zhu, J. Dalle, A. Van Schepdael, E. Roets, J. Hoogmartens, 'Analysis of Benzylpenicillin by Capillary Electrophoresis', *J. Chromatogr. A*, **792**(1–2), 83–88 (1997).
101. F. Belal, M.S. Rizk, M. Eid, 'Polarographic Determination of Some Penicillins through Nitrosation', *J. Pharm. Biomed. Anal.*, **17**(2), 275–282 (1998).
102. M.M. Amer, G. Ruecker, M.G. El-Bardicy, 'Colorimetric Method for the Determination of the Activity-concentration of Some Penicillins', *Egypt. J. Pharm. Sci.*, **30**(1–4), 111–123 (1989).
103. M.I. Helaleh, 'Indirect Spectrophotometric Determination of Benzylpenicillin and Cloxacillin in Pharmaceutical Preparations', *Mikrochim. Acta*, **129**(1–2), 29–32 (1998).
104. A. Saeed, I.A. Khan, 'Extraction Spectrophotometric Method for the Determination of Benzyl Penicillin in Pharmaceutical Preparations', *Indian Drugs*, **33**(2), 71–75 (1996).
105. C. Gonzalez Perez, I. Gonzalez Martin, B. Rodriguez Vazquez de Aldana, 'Polarographic-determination of Clavulanic Acid', *J. Pharm. Biomed. Anal.*, **9**(5), 383–386 (1991).
106. P. Izquierdo, A. Gomez-Hens, D. Perez-Bendito, 'Simultaneous Kinetic Flurometric Determination of

Amoxicillin and Clavulanic Acid by the Stopped Flow Mixing Technique', *Analyst*, **118**(6), 707–710 (1993).
107. P. Corti, E. Dreassi, G. Corbini, L. Montecchi, J. Paggi, 'Application of Near Infrared Reflectance Spectroscopy to Pharmaceutical Quality Control. 2. Assay of Cloxacillin Benzatine in Creams', *Analusis*, **18**(2), 117–121 (1990).
108. M.-C. Hsu, W.F. Huang, 'Collaborative Study of the Determination of Cloxacillin by Column Liquid Chromatography', *J. Chromatogr.*, **600**(2), 333–336 (1992).
109. M. Grover, M. Gulati, S.N. Singh, 'Stability-indicating Analysis of Isoxazolyl Penicillins Using Dual Wavelength High-performance Liquid Chromatography', *J. Chromatogr. B*, **708**(1–2), 153–159 (1998).
110. K.L. Tyczkowska, R.D. Voyksner, A.L. Aronson, 'Solvent Degradation of Cloxacillin and In Vitro Tentative Identification of Degradation Products Using Thermospray Liquid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **594**(1–2), 195–201 (1992).
111. M.C. Hsu, Y.J. Fann, 'Determination of Dicloxacillin Preparations by Liquid Chromatography', *J. AOAC Int.*, **75**(1), 26–29 (1992).
112. M.-C. Hsu, W.F. Huang, 'Liquid Chromatographic Determination of Dicloxacillin Preparations Interlaboratory Study', *J. AOAC Int.*, **76**(5), 1143–1145 (1993).
113. J.A. Murillo, A. Alanon, 'New Rapid Assay for Nafcillin by Spectrofluorometry in Pharmaceutical Dosage', *Anal. Lett.*, **26**(11), 2409–2420 (1993).
114. J.A. Murillo, A. Alanon, P. Fernandez, A. Munoz de la Pena, A. Espinosa-Mansilla, 'Simultaneous Determination of Nafcillin and Methicillin by Different Fluorimetric Techniques Using Partial Least-squares Calibration', *Analyst*, **123**(5), 1073–1077 (1998).
115. R.K. Cavatur, R. Suryanarayanan, 'Characterization of Frozen Aqueous Solutions by Low Temperature X-ray Powder Diffractometry', *Pharm. Res.*, **15**(2), 194–199 (1998).
116. I. Merino, E. Blanco Gonzalez, A. Sanz-Medel, 'Liquid Chromatographic Separation of Penicillamine Enantiomers Derivatized with OPA/2-ME on a Beta Cyclodextrin Bonded Phase', *Mikrochim. Acta*, **107**(1–2), 73–80 (1992).
117. A. Ion, F.G. Banica, A.G. Fogg, H. Kozlowski, 'Cathodic Stripping Voltammetry on D-Penicillamine and *N*-Acetylcysteine in the Presence of Nickel Ion', *Electroanalysis*, **8**(1), 40–43 (1996).
118. M. Abdelkawy, 'Spectrophotometric Determination of Penicillamine by Indirect Method', *Bull Fac. Pharm. (Cairo Univ.)*, **31**(1), 11–14 (1993).
119. G.A. Saleh, H.F. Askal, O.H. Abdelmageed, I.H. Refaat, 'Selective Spectrophotometric Methods for the Analysis of D-penicillamine Bulk Drug and Capsules in the Presence of Some Penicillins and Their Degradation Products', *Bull. Pharm. Sci. Assiut Univ.*, **17**(1), 65–71 (1994).
120. M.M.M.B. Duarte, G. de Oliveira Neto, L.T. Kubota, J.L.L. Filho, M.F. Pimentel, F. Lima, V. Lins, 'Application of Factorial Design to Improve a FIA System for Penicillin Determination', *Anal. Chim. Acta*, **350**(3), 353–357 (1997).
121. Q.S. Li, S.L. Zhang, J. Yu, 'Amperometric Penicillin Biosensor and its Application in Flow Injection Analysis System for Determination of Penicillin in Broth', *Microchem. J.*, **52**(2), 166–173 (1995).
122. B.G. Healey, D. Walt, 'Improved Fiber Optic Chemical Sensor for Penicillin', *Anal. Chem.*, **67**(24), 4471–4476 (1995).
123. J. Polster, G. Prestel, M. Wollenweber, G. Kraus, G. Gauglitz, 'Simultaneous Determination of Penicillin and Ampicillin by Spectral Fibre Optical Enzyme Optodes and Multivariate Data Analysis Based on Transient Signals Obtained by Flow Injection Analysis', *Talanta*, **42**(12), 2065–2072 (1995).
124. B. Morelli, 'Determination of Ternary Mixtures of Antibiotics by Ratio Spectra Zero Crossing First Derivative and 3rd Derivative Spectrophotometry', *J. Pharm. Biomed. Anal.*, **13**(3), 219–227 (1995).
125. B. Morelli, 'Determination of a Ternary Mixture of Penicillin-G Sodium Salt, Penicillin-G Procaine Salt and Dihydrostreptomycin Sulfate by 3rd Derivative Spectrophotometry', *Talanta*, **41**(4), 479–483 (1994).
126. C. Martorell, A.C. Calpena, E. Escribano, J.M. Poblet, J. Freixas, 'Influence of the Chromatographic Capacity Factor (Log k') as an Index of Lipophilicity in the Antibacterial Activity of a Series of 6-Fluoroquinolones – Relationship Between Physicochemical and Structural Properties and their Hydrophobicity', *J. Chromatogr. A.*, **655**(2), 177–184 (1993).
127. Z. Budvari-Barany, G. Szasz, K. Takacs-Novak, I. Hermecz, A. Lore, 'The pH Influence on the HPLC Retention of Chemotherapeutic Fluoroquinolone Derivatives', *J. Liq. Chromatogr.*, **14**(18), 3411–3424 (1991).
128. A.P. Argekar, S.U. Kapadia, S.V. Raj, 'Determination of Some Fluoroquinolones in Pharmaceutical Preparations by a HPTLC Method', *Indian Drugs*, **33**(3), 107–111 (1996).
129. K. Torniainen, E. Maki, 'Development of an Isocratic High-performance Liquid-chromatographic Method for Monitoring of Ciprofloxacin Photodegradation', *J. Chromatogr. A*, **697**(1–2), 397–405 (1995).
130. P.M. Lacroix, N.M. Curran, R.W. Sears, 'High-pressure Liquid Chromatographic Methods for Ciprofloxacin Hydrochloride and Related Compounds in Raw Materials', *J. Pharm. Biomed. Anal.*, **14**(5), 641–654 (1996).
131. S. Husain, S. Khalid, V. Nagaraju, R. Nageswara Rao, 'High-performance Liquid-chromatographic Separation and Determination of Small Amounts of Process Impurities of Ciprofloxacin in Bulk Drugs and Formulations', *J. Chromatogr. A*, **705**(2), 380–384 (1995).

132. R.T. Sane, D.V. Patel, S.N. Dhumal, V.R. Nerurkar, P.S. Mainkar, D.P. Gangal, 'High-performance Liquid-chromatographic Determination of Ciprofloxacin from Pharmaceutical Preparations', *Indian Drugs*, **27**(4), 248–250 (1990).
133. J.A. Morley, L. Elrod, 'Determination of Fluoroquinolone Antibacterials as *N*-Acyl Derivatives', *Chromatographia*, **37**(5–6), 295–299 (1993).
134. S. Tammilehto, H. Salomies, K. Torniainen, 'Qualitative and Quantitative TLC for Monitoring Ciprofloxacin Photodegradation in Aqueous Solutions', *J. Planar Chromatogr. Mod. TLC*, **7**(5), 368–371 (1994).
135. H. Avsec, S. Gomiscek, 'A Study of the Prospects for a Ciprofloxacin PVC Coated Wire Ion-selective Electrode Based on 4-Quinolones', *Anal. Chim. Acta*, **268**(2), 307–309 (1992).
136. P. O'Dea, A. Costa Gracia, A.J. Miranda Ordieres, P. Tunon Blanco, M.R. Smyth, 'Determination of Ciprofloxacin by Differential Pulse Polarography', *Electroanalysis*, **2**(8), 637–641 (1990).
137. K. Torniainen, J. Mattinen, C.-P. Askolin, S. Tammilehto, 'Structure Elucidation of a Photodegradation Product of Ciprofloxacin', *J. Pharm. Biomed. Anal.*, **15**(7), 887–894 (1997).
138. P.V. Bharat, G. Rajani, S. Vanita, 'An Oxidative Spectrophotometric Method for the Determination of Ciprofloxacin in Pharmaceutical Preparations', *Indian Drugs*, **34**(9), 497–500 (1997).
139. D.-S. Lee, H.-J. Han, K. Kim, W.-B. Park, J.-K. Cho, J.-H. Kim, 'Dissociation and Complexation of Fluoroquinolone Analogs', *J. Pharm. Biomed. Anal.*, **12**(2), 157–164 (1994).
140. C.J. Eboka, S.O. Aigba Boa, J.O. Akerele, 'Colorimetric Determination of the Fluoroquinolones', *J. Antimicrob. Chemother.*, **39**(5), 639–641 (1997).
141. Z. Bilgic, S. Tosunoglu, N. Buyuktimkin, 'Two New Spectrophotometric Methods for Ciprofloxacin', *Acta Pharm. Turc.*, **33**(1), 19–22 (1991).
142. S.G. Shanbag, P.P. Thampi, C.S. Thampi, 'Simple Spectrophotometric Method for the Estimation of Ciprofloxacin Hydrochloride and its Dosage Forms', *Indian Drugs*, **28**(6), 279–280 (1991).
143. G. Ramana Rao, A.B. Avadhanulu, D.K. Vatsa, 'Spectrophotometric Determination of Ciprofloxacin in its Dosage Forms', *Indian Drugs*, **27**(10), 532–535 (1990).
144. F.E.O. Suliman, S. Sultan, 'Sequential Injection Technique Employed for Stoichiometric Studies, Optimization and Quantitative Determination of Some Fluoroquinolone Antibiotics Complexed with Iron(III) in Sulfuric Acid Media', *Talanta*, **43**(4), 559–568 (1996).
145. S.M. Sultan, F.E.O. Suliman, 'Flow Injection Spectrophotometric Determination of the Antibiotic Ciprofloxacin in Drug Formulations', *Analyst*, **117**(9), 1523–1526 (1992).
146. G.R. Zhou, H.Z. Fan, J. Pan, 'Polarographic and Voltammetric Behavior of Enoxacin', *Analyst*, **120**(8), 2237–2239 (1994).
147. J.A. Squella, A. Alvarez-Lueje, J.C. Sturm, L.J. Nunez-Vergara, 'Enoxacin – Polarographic Behavior and its Determination in Pharmaceutical Forms', *Anal. Lett.*, **26**(9), 1943–1957 (1993).
148. V.M. Shinde, P.B. Shetkar, 'Selective Determination of Enrofloxacin and Secnidazole from Tablets by Reversed-phase HPLC', *Indian Drugs*, **33**(5), 230–231 (1996).
149. C.P.S. Sastry, K.R. Rao, J.S.V.M.L. Rao, D.S. Prasad, 'Two Simple Spectrophotometric Methods for the Assay of Enrofloxacin in Pharmaceutical Dosage Formulations', *East. Pharm.*, **38**(450), 143–144 (1995).
150. F. Belal, M.S. El-Din, 'Polarographic Behavior and Determination of Flumequine', *Microchem J.*, **42**(3), 300–305 (1990).
151. F.M. Abdel-Gawad, F.M. Abou Atti, 'Spectrophotometric Determination of Flumequine Using Iron(III) Chloride as a Color Developer', *Microchem. J.*, **50**(1), 106–110 (1994).
152. R.A. Sodhi, J.L. Chawla, R.T. Sane, 'Extractive Colorimetric, Microbiological and High-performance Thin-layer-chromatographic Determination of Lomefloxacin', *Indian Drugs*, **34**(9), 512–519 (1997).
153. Y.M. Issa, F.M. Abdel-Gawad, M.A.A. Table, H.M. Hussein, 'Spectrophotometric Determination of Ofloxacin and Lomefloxacin Hydrochloride with Some Sulfonphthalein Dyes', *Anal. Lett.*, **30**(11), 2071–2084 (1997).
154. K.P.R. Chowdary, G.D. Rao, 'Spectrophotometric Method for the Determination of Lomefloxacin Hydrochloride', *Indian Drugs*, **34**(2), 107–108 (1997).
155. A.P. Argekar, S.U. Kapadia, S.V. Raj, 'Simultaneous Determination of Norfloxacin and Tinidazole in Tablets by Reverse Phase High Performance Liquid Chromatography (RP-HPLC)', *Anal. Lett.*, **29**(9), 1539–1549 (1996).
156. C.P. Chen, X.R. Lui, R.J. Wu, 'High-performance Liquid-chromatographic Method for the Determination of Norfloxacin Glutamate and Glucuronate in Solid and Liquid Dosage Forms and its Application to Stability Testing', *J. Pharm. Biomed. Anal.*, **11**(8), 717–721 (1993).
157. A.P. Argekar, S.U. Kapadia, S.V. Raj, 'Simultaneous Determination of Norfloxacin and Tinidazole in Pharmaceutical Preparations by High Performance Thin Layer Chromatography', *J. Planar Chromatogr. Mod. TLC*, **9**(3), 208–211 (1996).
158. A.M.Y. Jaber, A. Lounici, 'Polarographic Behavior and Determination of Norfloxacin in Tablets', *Anal. Chim. Acta*, **291**(1–2), 53–64 (1994).
159. A.M.Y. Jaber, A. Lounici, 'Adsorptive Differential-pulse Stripping Voltammetry of Norfloxacin and its Analytical Application', *Analyst*, **119**(11), 2351–2357 (1994).
160. X.L. Wang, S.M. Zhang, W.W. Zhang, Q. Cai, 'Determination of Norfloxacin by Means of Oscillopolarographic Titration', *Anal. Lett.*, **29**(1), 131–139 (1996).

161. P.V. Bharat, G. Rajani, 'Extractive Spectrophotometric Method for the Determination of Norfloxacin in Pharmaceutical Preparations', *Indian Drugs*, **34**(2), 78–80 (1997).
162. K.P.R. Chowdary, A. Annapurna, 'New Spectrophotometric Method for the Estimation of Norfloxacin in Dosage Forms and in Dissolution Rate Studies', *Indian Drugs*, **29**(13), 612–615 (1992).
163. M. Cordoba-Borrego, M. Cordoba-Diaz, I. Bernabe, D. Corboda-Diaz, 'Determination of Norfloxacin by Fluorescence in the Presence of Different Antacids: Quantification of Analytical Interferences', *J. Pharm. Biomed. Anal.*, **14**(8–10), 977–982 (1996).
164. H. Al-Khamees, 'Colorimetric Determination of Norfloxacin in Tablets by Reaction with *p*-Benzoquinone', *Anal. Lett.*, **28**(1), 109–120 (1995).
165. P.P. Dahibhate, O.D. Chandwani, S.S. Kadam, S.R. Dhaneshwar, 'Simultaneous Spectrophotometric Estimation of Norfloxacin and Tinidazole from Combined Dosage Forms', *Indian Drugs*, **34**(1), 48–49 (1997).
166. S.Z. El Khateeb, S.A. Abdel Razek, M.M. Amer, 'Stability-indicating Methods for the Spectrophotometric Determination of Norfloxacin', *J. Pharm. Biomed. Anal.*, **17**(4–5), 829–840 (1998).
167. F.A. El-Yazbi, 'Spectrophotometric and Spectrofluorometric Determination of Ofloxacin', *Spectrosc. Lett.*, **25**(2), 279–291 (1992).
168. G. Fardella, P. Barbetti, I. Chiappini, G. Grandolini, 'Quantitative Analysis of Fluoroquinolones by H-1-NMR and F-19-NMR Spectroscopy', *Int. J. Pharm.*, **121**(1), 123–127 (1995).
169. A. Tamer, 'Adsorptive Stripping Voltammetric Determination of Ofloxacin', *Anal. Chim. Acta*, **231**(1), 129–131 (1990).
170. M. Tuncel, Z. Atkosar, 'The Determination of Ofloxacin in Tablets by Potentiometry and Conductometry', *Pharmazie*, **47**(8), 642–643 (1992).
171. G.R. Zhou, J. Pan, 'Polarographic and Voltammetric Behavior of Ofloxacin and its Analytical Application', *Anal. Chim. Acta*, **307**(1), 49–53 (1995).
172. D. Mundhe, S.G. Kashedikar, 'Spectrophotometric Estimation of Pefloxacin Mesylate in Pharmaceutical Formulations', *Indian Drugs*, **33**(8), 407–408 (1996).
173. A.K.S. Ahmad, M.A. Kawy, M. Nebsen, 'Spectrophotometric and Spectrofluorimetric Determination of Pefloxacin', *Anal. Lett.*, **30**(4), 809–820 (1997).
174. M. Jelikic-Stankov, D. Veselinovic, D. Malesev, Z. Radovic, 'Spectrophotometric Determination of Pefloxacin in Pharmaceutical Preparations', *J. Pharm. Biomed. Anal.*, **7**(12), 1571–1577 (1989).
175. M.G. Quaglia, A. Farina, P. Melchiorre, E. Bossu, P. Pagella, 'The Analytical Profile of Rufloxacin, a New Fluoroquinolone, by Reversed-phase High-performance Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **8**(8–12), 1075–1078 (1990).
176. M.G. Quaglia, E. Bossu, P. Melchiorre, A. Farina, A. Salvtori, 'Application of Fourth-derivative UV-Spectrophotometry to Pharmaceutical Compounds: Simultaneous Determination of a New Synthetic Fluoroquinolone and its Main Impurity', *Farmaco*, **46**(7–8), 979–985 (1991).
177. A. Farina, 'Fluorimetric Assay of Rufloxacin in Serum and in Pharmaceutical Formulations', *J. Pharm. Biomed. Anal.*, **7**(12), 1579–1583 (1989).
178. L. Elrod, C.L. Linton, B.P. Shelat, C.F. Wong, 'Determination of Minor Impurities in Temafloxacin Hydrochloride by High-performance Liquid Chromatography', *J. Chromatogr.*, **519**(1), 125–136 (1990).
179. J.F. Bauer, L. Elrod, Jr, J.R. Fornnarino, D.E. Heathcote, S.K. Krogh, C.I. Linton, B.J. Norris, J.E. Quick, 'Determination of Temafloxacin, Sarafloxacin, and Difloxacin in Bulk Drug and Dosage Forms by High-performance Liquid Chromatography', *Pharm. Res.*, **7**(11), 1177–1180 (1990).
180. A. Farina, F. Abballe, A. Doldo, M.G. Quaglia, 'Spectrofluorimetric Assay of Temafloxacin in Pharmaceutical Formulations and in Serum', *Spectrosc. Lett.*, **24**(9), 1219–1228 (1991).
181. R.J. Demchak, J.G. MacConnell, 'Normal-phase Chromatography and Postcolumn Colorimetric Detection of Abamectin-8,9-oxide', *J. Chromatogr.*, **511**, 353–358 (1990).
182. M.C. Caturla, E. Cusido, E.D. Westerlund, 'High-performance Liquid-chromatography Method for the Determination of Aminoglycosides Based on Automated Precolumn Derivatization with *ortho*-Phthalaldehyde', *J. Chromatogr.*, **593**(1–2), 69–72 (1992).
183. N. Grunwald, J. Vigneron, N. Blaise, A. Perrin, I. May, M.A. Hoffman, M. Hoffman, 'Determination of Aminoglycoside in Injectable Solutions by a Conductometric Method', *J. Pharm. Clin.*, **16**(3), 189–192 (1997).
184. M.T. Ackermans, F.M. Everaerts, J.L. Beckers, 'Determination of Aminoglycoside Antibiotics in Pharmaceuticals by Capillary Zone Electrophoresis with Indirect UV Detection Coupled with Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr.*, **606**(2), 229–235 (1992).
185. P.D. Voegel, R.P. Baldwin, 'Evaluation of Copper-based Electrodes for the Analysis of Aminoglycoside Antibiotics by CE/EC', *Electroanalysis*, **9**(15), 1145–1215 (1997).
186. H.-P. Fiedler, J. Wachter, 'High-performance Liquid-chromatographic Determination of Bleomycins', *J. Chromatogr.*, **536**(1–2), 343–347 (1991).
187. T.J. Wenzel, K. Zomlefer, S.B. Rapkin, R. Keith, 'Lanthanide Luminescence Detection of Bleomycins and Nalidixic Acid', *J. Liq. Chromatogr.*, **18**(7), 1473–1486 (1995).
188. R.J. Gorski, D.K. Morgan, C. Sarocka, A.C. Plasz, 'Estimation and Identification of Nonpolar Compounds in

Clarithromycin Bulk Drug by High-performance Liquid Chromatography', *J. Chromatogr.*, **540**(1–2), 422–429 (1991).

189. D.K. Morgan, D.M. Brown, T.D. Rotsch, A.C. Plasz, 'A Reversed-phase High-performance Liquid-chromatographic Method for the Determination and Identification of Clarithromycin as the Drug Substance and in Various Dosage Forms', *J. Pharm. Biomed. Anal.*, **9**(3), 261–269 (1991).
190. A. Hornedo-Nunez, T.A. Getek, W.A. Korfmacher, F. Simenthal, 'High-performance Liquid Chromatography of Clindamycin and Clindamycin Phosphate with Electrochemical Detection', *J. Chromatogr.*, **503**(1), 217–225 (1990).
191. M.A. Targove, N.D. Danielson, 'High-performance Liquid Chromatography of Clindamycin Antibiotics Using Tris(bipyridine)–Ruthenium(III) Chemiluminescence Detection', *J. Chromatogr. Sci.*, **28**(10), 505–509 (1990).
192. T.D. Rotsch, M. Spanton, P. Cugier, A.C. Plasz, 'Determination of Clarithromycin as a Contaminant on Surfaces by High-performance Liquid Chromatography Using Electrochemical Detection', *Pharm. Res.*, **8**(8), 989–991 (1991).
193. A.S. Amin, 'Spectrophotometric and Conductometric Determination of Clindamycin Hydrochloride in Pure Form and in Pharmaceutical Preparations', *Analusis*, **23**(8), 415–417 (1995).
194. F.A. El-Yazbi, S.M. Blaih, 'Spectrophotometric and Titrimetric Determination of Clindamycin Hydrochloride in Pharmaceutical Preparations', *Analyst*, **118**(5), 577–579 (1993).
195. B.A. Olsen, J.D. Stafford, D.E. Reed, 'Determination of Dirithromycin Purity and Related Substances by High-performance Liquid Chromatography', *J. Chromatogr.*, **594**(1–2), 203–208 (1992).
196. T. Cachet, M. Delrue, J. Paesen, R. Busson, E. Roets, J. Hoogmartens, 'Analysis of Erythromycin Estolate by Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **10**(10–12), 851–860 (1992).
197. J. Paesen, D.H. Calam, J.H.M.B. Miller, G. Raiola, A. Rozanski, B. Silver, J. Hoogmartens, 'Collaborative Study of the Analysis of Erythromycin by Liquid Chromatography on Wide-pore Poly(styrene–divinylbenzene)', *J. Liq. Chromatogr.*, **16**(7), 1529–1544 (1993).
198. M.M. Nasr, T.J. Tschappler, 'High-performance Liquid Chromatographic Analysis of Erythromycin', *J. Liq. Chromatogr. Relat. Technol.*, **19**(14), 2329–2348 (1996).
199. T. Cachet, P. Lannoo, J. Paesen, G. Janssen, J. Hoogmartens, 'Determination of Erythromycin Ethylsuccinate by Liquid Chromatography', *J. Chromatogr.*, **600**(1), 99–110 (1992).
200. T. Cachet, K. DeTurck, E. Roets, J. Hoogmartens, 'Quantitative Analysis of Erythromycin by Reversed-phase Liquid Chromatography Using Column Switching', *J. Pharm. Biomed. Anal.*, **9**(7), 547–555 (1991).
201. T. Cachet, I. Quintens, J. Paesen, E. Roets, J. Hoogmartens, 'Improved Separation of Erythromycin on Aged Reversed-phase Columns', *J. Liq. Chromatogr.*, **14**(6), 1203–1218 (1991).
202. J. Paesen, E. Roets, J. Hoogmartens, 'Liquid Chromatography of Erythromycin-a and Related Substances on Poly(styrene–divinylbenzene)', *Chromatographia*, **32**(3–4), 162–166 (1991).
203. C. Stubbs, I. Kanfer, 'A Stability-indicating High-performance Liquid-chromatographic Assay of Erythromycin Estolate in Pharmaceutical Dosage Forms', *Int. J. Pharm.*, **63**(2), 113–119 (1990).
204. A.K. Lalloo, I. Kanfer, 'Determination of Erythromycin and Related Substances by Capillary Electrophoresis', *J. Chromatogr. B*, **704**(1–2), 343–350 (1997).
205. P.A. Gallagher, N.D. Danielson, 'Colorimetric Determination of Macrolide Antibiotics Using Ferric Ion', *Talanta*, **42**(10), 1425–1432 (1995).
206. D. Dabrowska, A. Regosz, R. Piekos, M. Mierzwa, B. Paruch, 'A Study of Ion-pair Formation Between Erythromycin and Bromothymol Blue, Methylthymol Blue, and Thymol Blue and their use for Assaying Erythromycin in Dosage Forms', *Microchem. J.*, **41**(2), 210–218 (1990).
207. A. Sanz, V. Tomas, C. Martinez-Lozano, T. Perez-Ruiz, 'Extraction–Spectrofluorometric Method for the Determination of Erythromycin and its Esters in Pharmaceutical Formulations Using Manual and Flow Injection Procedures', *Analyst*, **118**(5), 567–571 (1993).
208. C.L. Flurer, K.A. Wolnik, 'Quantitation of Gentamicin Sulfate in Injectable Solutions by Capillary Electrophoresis', *J. Chromatogr.*, **663**(2), 259–263 (1994).
209. T.A. Getek, M.L. Vestal, T.G. Alexander, 'Analysis of Gentamicin Sulfate by High-performance Liquid Chromatography Combined with Thermospray Mass Spectrometry', *J. Chromatogr.*, **554**(1–2), 191–203 (1991).
210. E. Adams, R. Schepers, L.W. Gathu, R. Kibaya, E. Roets, J. Hoogmartens, 'Liquid-chromatographic Analysis of a Formulation Containing Polymyxin, Gramicidin and Neomycin', *J. Pharm. Biomed. Anal.*, **15**(4), 505–511 (1997).
211. E. Adams, J. Dalle, E. De Bie, I. De Smedt, E. Roets, J. Hoogmartens, 'Analysis of Kanamycin Sulfate by Liquid Chromatography with Pulsed Electrochemical Detection', *J. Chromatogr. A*, **766**(1–2), 133–139 (1997).
212. A.A. Alwarthan, S.A. Al-Tamrah, A. Akel, 'Flow-injection Determination of Kanamycin by Inhibition of the Lucigenin–H_2O_2–Co^{2+} System', *Anal. Chim. Acta*, **292**(1–2), 201–208 (1994).
213. X.M. Fang, X.F. Liu, J. Ye, Y.Z. Fang, 'Determination of Lincomycin and Lincomycin B in Bulk Drug and Pharmaceutical Formulations by Capillary Zone Electrophoresis with Amperometric Detection', *Anal. Lett.*, **29**(11), 1975–1984 (1996).

214. C.L. Flurer, 'Analysis of Macrolide Antibiotics by Capillary Electrophoresis', *Electrophoresis*, **17**(2), 359–366 (1996).
215. B.V. Tyaglov, V.I. Zvenigorodskii, I.A. Sizova, 'Quantitative Chromatographic Determination of Moenomycin Antibiotics', *J. Planar Chromatogr. Mod. TLC*, **10**(3), 200–203 (1997).
216. E. Adams, R. Schepers, E. Roets, J. Hoogmartens, 'Determination of Neomycin Sulfate by Liquid Chromatography with Pulsed Electrochemical Detection', *J. Chromatogr. A*, **741**(2), 233–240 (1996).
217. E. Roets, E. Adams, I.G. Muriithi, J. Hoogmartens, 'Determination of the Relative Amounts of the B-Component and C-Component of Neomycin by Thin-layer Chromatography Using Fluorescence Detection', *J. Chromatogr. A*, **696**(1), 131–138 (1995).
218. W. Funk, T. Kuepper, A. Wirtz, S. Netz, 'Quantitative HTPLC Determination of Neomycin-A, Neomycin-B, and Neomycin-C. 1. Chromatographic Separation and Post-chromatographic Derivatization', *J. Planar Chromatogr. Mod TLC*, **7**(1), 10–13 (1994).
219. C.H. Yamamoto, T.J.A. Pinto, 'Rapid Determination of Neomycin by a Microbiological Agar Diffusion Assay Using Triphenyltetrazolium Chloride', *J. AOAC Int.*, **79**(2), 434–440 (1996).
220. S.M. Abu-El-Wafa, M.A. El-Ries, F.M. Abou-Attai, R.M. Issa, 'Coordination Chemical Studies of Some Polymeric Transition-metal Complexes with Neomycin and their Biological Activity uses – Indirect Determination of Neomycin by Atomic-absorption Spectroscopy (AAS)', *Anal. Lett.*, **22**(13–14), 2703–2716 (1989).
221. J.A. Pope, Jr, R.A. Nelson, C.P. Schaffner, R.T. Rosen, R.C. Pandey, 'Applications of Thin-layer Chromatography, High-performance Liquid-chromatography and Mass Spectrometry in the Fermentation and Isolation of the Antibiotic Nybomycin', *J. Ind. Microbiol.*, **6**(1), 61–69 (1990).
222. L.L. Olson, J. Pick, W.Y. Ellis, P.L. Lim, 'A Chemical Assessment and HPLC Assay Validation of Bulk Paromomycin Sulfate', *J. Pharm. Biomed. Anal.*, **15**(6), 783–793 (1997).
223. F. Onur, S. Gungor, 'Simultaneous Determination of Rifamycin SV Sodium and Lidocaine Hydrochloride in Injection Preparations by First Derivative UV Spectrophotometry', *Sci. Pharm.*, **65**(3), 155–163 (1997).
224. V. de Oliveira, A.M. Bergold, E.E.S. Schapoval, 'High-performance Liquid Chromatographic Determination of Roxithromycin in Tablets', *Anal. Lett.*, **29**(13), 2377–2382 (1996).
225. C.L. Flurer, 'The Analysis of Aminoglycoside Antibiotics by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **13**(7), 809–816 (1995).
226. S. Arrowood, A.M. Hoyt, Jr, M.W. Woods, 'Determination of Streptomycin in Pharmaceutical Preparations by Gas Chromatography', *J. High Resolut. Chromatogr.*, **14**(12), 808–810 (1991).
227. G. Eneva, B. Nikolova-Damyanova, S. Spassov, M. Haimova, *J. Planar Chromatogr. Mod. TLC*, **3**, 232–235 (1990).
228. D. Leech, J. Wang, M.R. Smyth, 'Electrocatalytic Detection of Streptomycin and Related Antibiotics at Ruthenium Dioxide', *Analyst*, **115**(11), 1447–1450 (1990).
229. M. Rizk, Y. El-Shabrawy, N.A. Zakhari, S.S. Toubar, L. Carreira, 'Fluorimetric Determination of Aminoglycoside Antibiotics Using Lanthanide Probe Ion Spectroscopy', *Talanta*, **42**(12), 1849–1856 (1995).
230. N.A. Zakhari, 'Spectrophotometric Assay of Certain Aminoglycosides Using Cyanoacetamide', *Anal. Lett.*, **23**(10), 1843–1856 (1990).
231. J.G. Phillips, C. Simmonds, 'Determination of Spectinomycin Using Cation-exchange Chromatography with Pulsed Amperometric Detection', *J. Chromatogr. A*, **675**(1–2), 123–128 (1994).
232. A.K. Dash, R. Suryanarayanan, 'A Liquid-chromatographic Method for the Determination of Tobramycin', *J. Pharm. Biomed. Anal.*, **9**(3), 237–245 (1991).
233. A. Gaikwad, A. Gomez-Hens, D. Perez-Bendito, 'Kinetic Fluorometric Method for the Determination of Tobramycin by Stopped-flow Mixing Methodology', *Anal. Lett.*, **26**(1), 97–107 (1993).
234. J. Paesen, W. Cypers, R. Busson, E. Roets, J. Hoogmartens, 'Isolation of Decomposition Products of Tylosin Using Liquid Chromatography', *J. Chromatogr. A*, **699**(1–2), 99–106 (1996).
235. J. Paesen, P. Claeys, W. Cypers, E. Roets, J. Hoogmartens, 'Liquid Chromatography of Tylosin-A and Related Substances on Poly(styrene–divinylbenzene)', *J. Chromatogr.*, **699**(1–2), 93–97 (1995).
236. E. Roets, P. Beirinckx, I. Quintens, J. Hoogmartens, 'Quantitative Analysis of Tylosin by Column Liquid Chromatography', *J. Chromatogr.*, **630**(1–2), 159–166 (1993).
237. F.L. Neely, 'Flow Injection Determination of Tylosin in Fermentation Broth', *Anal. Chim. Acta*, **281**(2), 243–247 (1993).
238. N.A. Botsoglou, D.J. Fletouris, E.J. Simeonidou, I.E. Psomas, 'Retention Behavior of Multiple Sulfonamides in Various Lipid Chromatographic Systems', *Chromatographia*, **46**(9–10), 477–482 (1997).
239. M.C. Ricci, R.F. Cross, 'Capillary Electrophoresis Separation of Sulfonamides and Dihydrofolate-reductase Inhibitors', *J. Microcol. Sep.*, **5**(3), 207–215 (1993).
240. C.-E. Lin, C.-C. Chang, W.-C. Lin, 'Migration Behavior and Separation of Sulfonamides in Capillary Zone Electrophoresis. III. Citrate Buffer as a Background Electrolyte', *J. Chromatogr. A*, **768**(1), 105–112 (1997).
241. C.-E. Lin, W.-C. Lin, Y.-C. Chen, S.-W. Wang, 'Migration Behavior and Selectivity of Sulfonamides in Capillary Electrophoresis', *J. Chromatogr. A*, **792**(1–2), 37–47 (1997).

242. H. Salomies, 'Quantitative HPTLC of Sulfonamides in Pharmaceutical Preparations', *J. Planar Chromatogr. Mod. TLC*, **6**(5), 337–340 (1993).
243. M.C. Garcia-Alvarez-Coque, E.F. Simo-Alfonso, G. Ramis-Ramos, J.S.J. Esteve-Romero, 'High-performance Micellar Liquid Chromatography Determination of Sulphonamides in Pharmaceuticals After Azo Dye Precolumn Derivatization', *J. Pharm. Biomed. Anal.*, **13**(3), 237–245 (1995).
244. A.M.I. Mohamed, H.F. Askal, G.A. Saleh, 'Use of *para*-Benzoquinone for the Spectrophotometric Determination of Certain Sulfonamides', *J. Pharm. Biomed. Anal.*, **9**(7), 531–538 (1991).
245. M. Sanchez Pena, F. Salinas, M.C. Mahedero, J.J. Aaron, 'Spectrofluorometric Determination of Sulfonamides in Pharmaceutical Compounds and Foods', *J. Pharm. Biomed. Anal.*, **10**(10–12), 805–808 (1992).
246. K.J. Hartauer, E.S. Miller, J.K. Guillory, 'Diffuse Reflectance Infrared Fourier-transform Spectroscopy for the Quantitative-analysis of Mixtures of Polymorphs', *Int. J. Pharm.*, **85**, 163–174 (1992).
247. M. Kotoucek, J. Skopalova, D. Michalkova, 'Electroanalytical Study of Salazosulfapyridine and Biseptol Components at the Mercury Electrode', *Anal. Chim. Acta*, **353**(1), 61–69 (1997).
248. J.J. Berzas, J. Rodriguez, J.M. Lemus, G. Castaneda, 'Voltammetric Behavior of Sulfadimetoxol Using Square-wave and Adsorptive Stripping Square-wave Techniques', *Electroanalysis*, **9**(6), 474–477 (1997).
249. F. Abdel-Gawad, N. El-Guindi, 'Spectrophotometric Studies on Determination of Sulfadimidine', *Egypt. J. Pharm. Sci.*, **35**, 301–302 (1994).
250. M.C. Mahedero, J.J. Aaron, 'Flow-injection Determination of Sulfonamides with Fluorometric or Photochemical Fluorometric Detection', *Anal. Chim. Acta*, **269**(2), 193–198 (1992).
251. J.M. Pingarron Carrazon, A. Dominguez Recio, L.M. Polo Diez, 'Electroanalytical Study of Sulfamerazine at a Glassy-carbon Electrode and its Determination in Pharmaceutical Preparations by HPLC with Amperometric Detection', *Talanta*, **39**(6), 631–635 (1992).
252. F.M. El Anwar, A.M. El Walily, H.M.H. Abdel, M. El Swify, 'The Analysis of a Triple Sulfonamide in Pharmaceutical Powder Form by HPLC', *Anal. Lett.*, **24**(5), 767–779 (1991).
253. S. Raghuveer, I.R.K. Raju, D.K. Vatsa, C.M.R. Srivastava, 'Spectrophotometric Determination of Sulphadiazine in Pharmaceutical Dosage Forms', *Indian Drugs*, **30**(4), 191–193 (1993).
254. A. Mellado Romero, C. Gomez Benito, J. Martinez Calatayud, 'Continuous-flow Spectrophotometric Determination of Sulfadiazine by Diazotization with In-situ Preparation of Nitrate', *Anal. Chim. Acta*, **308**(1–3), 451–456 (1995).
255. J.J. Aaron, M.I. Acedo Valenzuela, M. Sanchez Pena, F. Salinas, M.C. Mahedero, 'Quantitative Analysis of Sulfadiazine Using Photochemically Induced Fluorescence Detection in Bulk Solution and in a Flow-injection System', *Anal. Chim. Acta*, **314**(1–2), 45–50 (1995).
256. M. Sanchez Pena, M. J. Acedo, F. Salinas, M.C. Mahedero, J. Aaron, 'Analysis of Sulfamethazine in the Presence of Sulfamerazine or Sulfadiazine by First Derivative Photochemically Induced Fluorescence', *J. Pharm. Biomed. Anal.*, **13**(9), 1107–1112 (1995).
257. A. Espinosa-Mansilla, F. Salinas, I. De Orbe Paya, 'Simultaneous Determination of Sulfadiazine, Doxycycline, Furaltadone and Trimethoprim by Partial Least Squares Multivariate Calibration', *Anal. Chim. Acta*, **313**(1–2), 103–112 (1995).
258. A. Espinosa Mansilla, F. Salinas, I. De Orbe Paya, 'Abilities of Partial Least Squares (PLS) Multivariate Calibration in the Analysis of a Quaternary Mixture of Sulfonamides', *Fresenius' J. Anal. Chem.*, **354**(2), 245–249 (1996).
259. C.V.S. Subramanyam, C.L. Chowdary, M.S. Reddy, J.V. Rao, 'Simultaneous Determination of Trimethoprim and Sulfadiazine in Combined Dosage Forms', *Indian Drugs*, **28**(9), 438–440 (1991).
260. M. Sanchez Pena, F. Salinas, M.C. Mahedero, J.J. Aaron, 'Solvent Effect on the Determination of Sulfamethazine by Room-temperature Photochemically Induced Fluorescence', *Talanta*, **41**(2), 233–236 (1994).
261. J.J. Berzas Nevado, J. Rodriguez Flores, M.L. De la Morena Pardo, 'Determination of Sulfamethizole in the Presence of Nitrofurantoin by Derivative Spectrophotometry and Ratio Spectra Derivative', *Talanta*, **38**(11), 1261–1264 (1991).
262. L.J. Mulcahey, L.T. Taylor, 'Supercritical Fluid Extraction of Active Components in a Drug Formulation', *Anal. Chem.*, **64**(9), 981–984 (1992).
263. M.E.M. Hassouna, 'Simultaneous Spectrophotometric Determination of Sulfamethoxazole and Trimethoprim in Drugs', *Anal. Lett.*, **30**(13), 2341–2352 (1997).
264. J.J. Berzas Nevado, J. Lemus Gallego, J.M. Castaneda, G. Penalvo, 'Determination of Sulfamethoxazole and Trimethoprim by Ratio Spectra Derivative Spectrophotometry', *Fresenius' J. Anal. Chem.*, **342**(9), 723–728 (1992).
265. A. Munoz de La Pena, F. Salinas, I. Duran Meras, M. Dolores Moreno, 'Fluorometric Determination of Sulfamethoxazole in Pharmaceutical Preparations in Combination with Trimethoprim by Inclusion in Beta-cyclodextrin Urea', *Anal. Lett.*, **27**(10), 1893–1906 (1994).
266. M.A. Abounassif, M.E.M. Hagga, E.A. Gad-Kariem, M.A. Al-Awady, '1st-Derivative UV Spectrometric Determination of Sulfametrole and Trimethoprim Mixture in Tablets', *Acta Pharm. Jugosl.*, **41**(3), 223–231 (1991).
267. J.J. Berzas Nevado, F. Salinas, I. De Orbe Paya, L.F. Capitan-Vallvey, 'Simultaneous Determination of Sulfathiazole and Sulfanilamide in Pharmaceuticals by

Derivative Spectrophotometry', *J. Pharm. Biomed. Anal.*, **9**(2), 117–122 (1991).
268. L.F. Capitan-Vallvey, I. DeOrbe, M.C. Valencia, J.J. Berzas Nevado, 'Determination of Sulphathiazole in the Presence of Sulphamethazine by 1st-Derivative Solid-phase Spectrophotometry', *Mikrochim. Acta*, **111**(4), 223–229 (1993).
269. D.M. Joshi, A.P. Joshi, Polarographic 'Determination of Tinidazole and Furazolidone', *Indian Drugs*, **33**(7), 338–343 (1996).
270. C.V.N. Prasad, C. Parihar, K. Sunil, P. Parimoo, 'Simultaneous Determination of Tinidazole–Clotrimazole and Tinidazole–Norfloxacin in Combined Tablet Preparations by Derivative Spectroscopy', *Pharm. Sci.*, **3**(7), 337–341 (1997).
271. W.A. Korfmacher, T.A. Getek, E.B. Hansen, C.E. Cerniglia, S. Abou-Khalil, A.A. Yunis, 'Application of Thermospray Liquid Chromatography Mass Spectrometry for the Analysis of Chloramphenicol and 3 Related Compounds', *J. Chromatogr. Sci.*, **28**(5), 236–238 (1990).
272. M. Feng, D.J. Long, Y.Z. Fang, 'Parallel Incident Spectroelectrochemistry Study of Chloramphenicol', *Anal. Chim. Acta*, **363**(1), 67–73 (1998).
273. Z. Min, J. Hu, Y. Lai, Q. Li, Z. Wu, 'Study of the Voltammetric Behavior of Chloramphenicol and its Determination at a Ni/C Modified Electrode', *Anal. Lett.*, **31**(2), 237–249 (1998).
274. S. McNiven, M. Kato, R. Levi, K. Yano, I.R.J. Karube, 'Chloramphenicol Sensor Based on an In Situ Imprinted Polymer', *Anal. Chim. Acta*, **365**(1–3), 69–74 (1998).
275. N.M.A. Mahfouz, K.M. Emara, 'Colorimetric Determination of Isoniazid and its Pharmaceutical Formulations', *Talanta*, **40**(7), 1023–1029 (1993).
276. P.B. Issopoulos, 'Analytical Investigations on Isonicotinic Acid Hydrazide (Isoniazid). 4. Colorimetric Determination of Microquantities of Isoniazid Using 2,3,5-Triphenyltetrazolium Chloride', *Acta Pharm. Jugosl.*, **4**(2), 123–128 (1991).
277. A.M. El-Brashy, F.A. Ibrahim, S.M. El-Ashry, 'Colorimetric Determination of Some Important Hydrazine Derivatives', *Mikrochim. Acta*, **3**(4–6), 105–110 (1991).
278. A.H.N. Ahmed, S.M. El Gizawy, H.I. El Subbagh, 'Spectrophotometric Determination of Isoniazid Using Ethyl-8-quinolinoxy Acetate', *Anal. Lett.*, **25**(1), 73–80 (1992).
279. M. Subhashini, M.S. Subramanian, V.R.S. Rao, 'Determination of Isoniazid in Pharmaceutical Preparations by Reaction with Radiochloramine-B', *J. Pharm. Biomed. Anal.*, **7**(12), 1473–1477 (1989).
280. V.M. Shinde, N.M. Tendolkar, B. Desai, 'Simultaneous Determination of Nitrofurantion and Trimethoprim in Capsules by HPLC', *Indian Drugs*, **32**(6), 283–286 (1995).
281. J.J. Bergh, J.C. Breytenbach, J.L. DuPreez, 'High-performance Liquid-chromatographic Analysis of Trimethoprim in the Presence of its Degradation Products', *J. Chromatogr.*, **513**, 392–396 (1990).
282. M.C. Pascual-Marti, R. Marin Saez, J.M. Iranzo Adrian, 'Determination of Trimethoprim by the Linear Absorbance Method', *Fresenius' J. Anal. Chem.*, **352**(3–4), 396–397 (1995).
283. S.Z. Qureshi, M.I.H. Helaleh, N. Rahman, R.M.A. Jamhour, 'Spectrophotometric Determination of Trimethoprim by Oxidation in Drug Formulations', *Fresenius' J. Anal. Chem.*, **357**(7), 1005–1007 (1997).
284. A.G. Kazemifard, D.E. Moore, 'Evaluation of Amperometric Detection for the Liquid-chromatographic Determination of Tetracycline Antibiotics and their Common Contaminants in Pharmaceutical Formulations', *J. Pharm. Biomed. Anal.*, **16**(4), 689–696 (1997).
285. Y.M. Li, A. Van Schepdael, E. Roets, J. Hoogmartens, 'Optimized Methods for Capillary Electrophoresis of Tetracyclines', *J. Pharm. Biomed. Anal.*, **15**(8), 1063–1069 (1997).
286. M.F.M. Tavares, V. McGuffin, 'Separation and Characterization of Tetracycline Antibiotics by Capillary Electrophoresis', *J. Chromatogr. A*, **686**(1), 129–142 (1994).
287. S. Croubels, W. Baeyens, C. Dewaele, C. van Peteghem, 'Capillary Electrophoresis of Some Tetracycline Antibiotics', *J. Chromatogr. A*, **673**(2), 267–274 (1994).
288. Y.-C. Chen, C.-E. Lin, 'Migration Behavior and Separation of Tetracycline Antibiotics by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **802**(1), 95–105 (1998).
289. R. Karlicek, P. Solich, 'Flow-injection Spectrophotometric Determination of Tetracycline Antibiotics', *Anal. Chim. Acta*, **285**(1–2), 9–12 (1994).
290. S.M. Sultan, F.-E.O. Suliman, S.O. Duffuaa, I.I. Abu-Abdoun, 'Simplex-optimized and Flow-injection Spectrophotometric Assay of Tetracycline Antibiotics in Drug Formulations', *Analyst*, **117**(7), 1179–1183 (1992).
291. Z. Li, M. Feng, J. Lu, Z. Gong, H. Jiang, 'Flow Injection Chemiluminescence Determination of Tetracyclines', *Anal. Lett.*, **30**(4), 797–807 (1997).
292. R.D. Jee, 'Study of Micellar Solutions to Enhance the Europium-sensitized Luminescence of Tetracyclines', *Analyst*, **120**(12), 2867–2872 (1995).
293. A.E. Ciarlone, B.W. Fry, D.M. Ziemer, 'Some Observations on the Adsorption of Tetracyclines to Glass and Plastic Labware', *Microchem. J.*, **42**(2), 250–255 (1990).
294. J. Hoogmartens, 'Liquid Chromatography of Chlortetracycline', *Pharmeuropa*, **5**(2), 149–150 (1993).
295. N.D. Weng, J. De Beer, X. Marcelis, P. Derese, J.H.M. Miller, J. Hoogmartens, 'Collaborative Study of the Analysis of Chlortetracycline Hydrochloride by Liquid Chromatography on Poly(styrene–divinylbenzene) Packing Materials', *J. Pharm. Biomed. Anal.*, **10**(2–3), 199–204 (1992).
296. K. Smets, E. Roets, J. Miller, E. Porqueras, N. Berti, P. Vettori, F. Folliard, J. Hoogmartens, 'Collaborative Study of the Analysis of Chlortetracycline by Liquid

Chromatography on Octylsilyl Silica Gel', *Chromatographia*, **37**(11–12), 640–644 (1993).
297. W. Naidong, E. Roets, J. Hoogmartens, 'High-performance Liquid Chromatography of Chlortetracycline and Related Substances on Poly(styrene–divinylbenzene) Copolymer', *Chromatographia*, **30**(1–2), 105–109 (1990).
298. N.D. Weng, H. Sun, E. Roets, J. Hoogmartens, 'Purity Control of 6 Tetracyclines by Semiquantitative TLC with Fluorescence Detection', *J. Planar Chromatogr. Mod. TLC*, **7**(4), 297–300 (1994).
299. N. Weng, E. Roets, R. Busson, J. Hoogmartens, 'Separation of Keto–Enol Tautomers of Chlortetracycline and 4-Epichlortetracycline by Liquid Chromatography on Poly(styrene–divinylbenzene) Copolymer', *J. Pharm. Biomed. Anal.*, **8**(8–12), 881–889 (1990).
300. T.M. Rossi, R.R. Ryall, 'Analytical Methods for Cleaning Procedures', *Prog. Pharm. Biomed. Anal.*, **13**, 293–302 (1996).
301. X.R. Zhang, W.R.G. Baeyens, A. Van den Borre, G. Van der Weken, A.C. Calokerinos, S. Schulman, 'Chemiluminescence Determination of Tetracyclines Based on their Reaction with Hydrogen Peroxide Catalyzed by the Copper Ion', *Analyst*, **120**(2), 463–466 (1995).
302. Z. Li, M. Feng, J. Lu, Z. Gong, H. Jiang, 'Flow Injection Chemiluminescence Determination of Tetracyclines', *Anal. Lett.*, **30**(4), 797–807 (1997).
303. N. Weng, E. Roets, J. Hoogmartens, 'Quantitative-analysis of Demeclocycline by High-performance Liquid-chromatography', *J. Pharm. Biomed. Anal.*, **7**(12), 1691–1703 (1989).
304. Y.M. Li, A. Van Schepdael, E. Roets, J. Hoogmartens, 'Analysis of Demeclocycline by Capillary Electrophoresis', *J. Chromatogr. A*, **740**(1), 119–123 (1996).
305. M. Jelikic-Stankov, D. Stankov, D. Malesev, Z. Radovic, 'Spectrophotometric Method for the Determination of Demeclocycline Using Sodium Molybdate as Analytical Reagent', *Mikrochim. Acta*, **1**(1–2), 65–70 (1991).
306. A. Van Schepdael, R. Kibaya, E. Roets, J. Hoogmartens, 'Analysis of Doxycycline by Capillary Electrophoresis', *Chromatographia*, **41**(5–6), 367–370 (1995).
307. N.D. Weng, H. Sun, K. Verresen, E. Roets, J. Hoogmartens, 'Assay and Purity Control of Metacycline by Thin-layer Chromatography Combined with UV and Fluorescence Densitometry – A Comparison with Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **9**(9), 717–723 (1991).
308. Y.M. Li, A. Van Schepdael, E. Roets, J. Hoogmartens, 'Capillary Zone Electrophoresis of Minocycline', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1095–1099 (1996).
309. N.D. Weng, J. Thuranira, K. Vermeulen, E. Roets, J. Hoogmartens, 'Quantitative Analysis of Minocycline by Liquid Chromatography on Poly(styrene–divinylbenzene)', *J. Liq. Chromatogr.*, **15**(14), 2529–2545 (1992).
310. P.D. Bryan, J.T. Stewart, 'Separation of Tetracyclines by Liquid Chromatography with Acidic Mobile Phases and Polymeric Columns', *J. Pharm. Biomed. Anal.*, **11**(10), 971–976 (1993).
311. J. Di, X. Xu, J. Luo, 'Determination of Minocycline by Semi-differential Cyclic Voltammetry at a Liquid/Liquid Interface', *Anal. Lett.*, **29**(15), 2691–2700 (1996).
312. J. Tjoernelund, S.H. Hansen, 'Validation of a Simple Method for the Determination of Oxytetracycline in Ointment by Non-aqueous Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **15**(8), 1077–1082 (1997).
313. A. Van Schepdael, I. Van den Bergh, E. Roets, J. Hoogmartens, 'Purity Control of Oxytetracycline by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **730**(1–2), 305–311 (1996).
314. Y.M. Li, A. Van Schepdael, E. Roets, J.F. Hoogmartens, 'Separation of Oxytetracycline and its Related Substances by Capillary Electrophoresis', *J. Liq. Chromatogr. Relat. Technol.*, **20**(2), 273–282 (1997).
315. C.R. White, W.A. Moats, K.L. Kotula, 'Comparative Study of High-performance Liquid-chromatographic Methods for the Determination of Tetracycline Antibiotics', *J. Liq. Chromatogr.*, **16**(13), 2873–2890 (1993).
316. N. Weng, S. Geelen, E. Roets, J. Hoogmartens, 'Assay and Purity Control of Oxytetracycline and Doxycycline by Thin-layer Chromatography – A Comparison with Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **8**(8–12), 891–898 (1990).
317. S. Raghuveer, C.M.R. Srivastava, D.K. Vatsa, 'Spectrophotometric Determination of Oxytetracycline Hydrochloride in its Dosage Forms', *Indian Drugs*, **29**(6), 281–283 (1992).
318. M. Jelikic-Stankov, D. Veselinovic, D. Malesev, Z. Radovic, 'Spectrophotometric Determination of Oxytetracycline in Pharmaceutical Preparations Using Sodium Molybdate as Analytical Reagent', *J. Pharm. Biomed. Anal.*, **7**(12), 1565–1570 (1989).
319. A.A. Alwarthan, S.A. Al-Tamrah, S.M. Sultan, 'Spectrophotometric Determination of Oxytetracycline by Flow Injection', *Analyst*, **116**(2), 183–186 (1991).
320. Y. Vander Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, 'Ruggedness Tests on the High-performance Liquid Chromatography Assay of the United States Pharmacopeia XXII for Tetracycline Hydrochloride – A Comparison of Experimental Designs and Statistical Interpretations', *Anal. Chim. Acta*, **312**(3), 245–262 (1995).
321. Y.V. Heyden, D.L. Massart, Y. Zhu, J. Hoogmartens, J. De Beer, 'Ruggedness Tests on the High Performance Liquid Chromatography Assay of the United States Pharmacopeia 23 for Tetracycline Center Dot HCl: Comparison of Different Columns in an Interlaboratory Approach', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1313–1326 (1996).
322. A. Van Schepdael, J. Saevels, X. Lepoudre, R. Kibaya, N.Z. Gang, E. Roets, J. Hoogmartens, 'Separation of

Tetracycline and its Related Substances by Capillary Zone Electrophoresis', *J. High Resolut. Chromatogr.*, **18**(11), 695–698 (1995).

323. J. Tjornelund, S.H. Hansen, 'Determination of Impurities in Tetracycline Hydrochloride by Non-aqueous Capillary Electrophoresis', *J. Chromatogr. A*, **737**(2), 291–300 (1996).
324. C. Hendrix, E. Roets, J. Crommen, J. De Beer, E. Porqueras, W. Van den Bossche, J. Hoogmartens, 'Collaborative Study of the Analysis of Tetracycline by Liquid Chromatography on Poly(styrene–divinylbenzene)', *J. Liq. Chromatogr.*, **16**(15), 3321–3329 (1993).
325. N. Weng, H. Sun, E. Roets, J. Hoogmartens, 'Assay and Purity Control of Tetracycline by Thin-layer Chromatography. II. Quantitative Aspects: Comparison with Liquid Chromatography', *J. Planar Chromatogr. Mod. TLC*, **5**(3), 152–156 (1992).
326. C.X. Zhang, Z.P. Sun, D.K. Ling, Y.J. Zhang, 'Separation of Tetracycline and its Degradation Products by Capillary Zone Electrophoresis', *J. Chromatogr.*, **627**(1–2), 281–286 (1992).
327. N. Weng, H. Sun, E. Roets, J. Hoogmartens, 'Assay and Purity Control of Tetracycline by Thin-layer Chromatography', *J. Planar Chromatogr. Mod. TLC*, **5**(2), 92–98 (1992).
328. N.H. Khan, P. Wera, E. Roets, J. Hoogmartens, 'Quantitative Analysis of Tetracycline by High Performance Liquid Chromatography on Poly(styrene–divinylbenzene) Packing Materials', *J. Liq. Chromatogr.*, **13**(7), 1351–1374 (1990).
329. U. Saha, A.K. Sen, T.K. Das, S.K. Bhowal, 'Spectrophotometric Determination of Tetracyclines in Pharmaceutical Preparations with Uranyl Acetate', *Talanta*, **37**(12), 1193–1196 (1990).
330. P.K. Hon, W.K. Fung, 'Identification of Tetracyclines by 2nd Derivative Ultraviolet Spectrophotometry', *Analyst*, **116**(7), 751–752 (1991).
331. J.X. Duggan, 'Phosphorimetric Detection in HPLC via Trivalent Lanthanides: High Sensitivity Time Resolved Luminescence Detection of Tetracyclines Using EU^{3+} in a Micellar Post Column Reagent', *J. Liq. Chromatogr.*, **14**(13), 2499–2525 (1991).

Chemical Reagents and Derivatization Procedures in Drug Analysis

Neil D. Danielson
Miami University, Oxford, OH, USA

Patricia A. Gallagher
Environmental Protection Agency, Cincinnati, OH, USA

James J. Bao
Tianjin University, Tianjin, China

1 Introduction **1**
2 Gas Chromatography **2**
2.1 Alkylation 2
2.2 Acylation 4
2.3 Silylation 6
2.4 Sample Handling 9
3 High-performance Liquid Chromatography **10**
3.1 Alkaloids 10
3.2 Amines 12
3.3 Antibiotics 14
3.4 Barbiturates 17
3.5 Carbonyl Compounds and Carboxylic Acids 17
3.6 Catecholamines 18
3.7 Hydroxy Compounds 19
3.8 Steroids 19
3.9 Sulfur Compounds 20
4 Capillary Electrophoresis **21**
4.1 Derivatization for Ultraviolet Detection 22
4.2 Derivatization for Fluorescence Detection 23
4.3 Special Applications 24
4.4 Derivatization Modes 26
5 Conclusion **27**
Abbreviations and Acronyms **27**
Related Articles **28**
References **28**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

This article describes both pre- and postcolumn derivatization chemistry used in conjunction with either chromatography or capillary electrophoresis (CE) to facilitate the determination of drugs. Generally, only prederivatization is used in gas chromatography (GC), principally to enhance the volatility, temperature stability, and/or detectability. The GC section considers derivatization of drugs by reagent class: alkylation, acylation, and silylation. The GC sample-handling section describes an approach that combines the extraction and derivatization steps together. Both pre- and postcolumn derivatization are common approaches for high-performance liquid chromatography (HPLC) and many of these methods have been adapted for CE. Derivatization for HPLC is often directed toward aliphatic amines, carboxylic acids, or alcohols that are difficult to detect at low levels by absorbance, luminescence, or electrochemical means. In addition, small hydrophilic molecules upon prederivatization are often converted into larger more hydrophobic compounds, making reversed-phase HPLC easier or even feasible. The HPLC section considers derivatization of drugs based on the following types: alkaloids, amines, antibiotics, barbiturates, carbonyl compounds and carboxylic acids, catecholamines, hydroxy compounds, steroids, and sulfur compounds. For the GC and HPLC sections, tables are included, giving the structures of the more important derivatizing agents, the analytes, and the corresponding reaction products. A brief rationale for derivatization chemistry with CE concludes this article. For CE, derivatization is often done to improve detectability as the path length for absorbance detection is very short and fluorescence can be effective using laser-based systems.

1 INTRODUCTION

Derivatization in conjunction with chromatography or CE can be done either in an off-line mode, often prior to the separation step (prederivatization), or in an on-line mode (postcolumn derivatization). Generally, only prederivatization is used in GC, principally to enhance the volatility, temperature stability, and/or detectability. Both pre- and postcolumn derivatization are common approaches for HPLC and many of these methods have been adapted for CE. Derivatization for HPLC is often directed toward aliphatic amines, carboxylic acids, or alcohols that are difficult to detect at low levels by absorbance, luminescence, or electrochemical means. In addition, small hydrophilic molecules upon prederivatization are often converted into larger more hydrophobic compounds, making reversed-phase HPLC easier or feasible. For CE, derivatization is often done to improve detectability as the path length for absorbance

detection is very short and fluorescence can be effective using laser-based systems.

The desirable conditions to be met for precolumn derivatization are (i) the reaction stoichiometry and product structure should be known, (ii) the reaction should be reasonably fast and should proceed quantitatively (or at least reproducibly), and (iii) the derivative should be stable, readily separable, and distinguishable from the starting material. One advantage of the prederivatization approach is that simple equipment is commercially available to allow for the reaction chemistry to be done in the batch mode. Subsequent sample analysis using an autoinjector and a standard unmodified chromatograph is straightforward. Alternatively, the use of a robot to carry out the entire prederivatization chemistry and an HPLC instrument equipped with an autosampler means that the sample throughput will usually be limited by the chromatographic step.

Because postcolumn derivatization involves mixing the column effluent with a reagent or passing it through a reactor to form the derivative before it is detected, the automation step is built in. Sample handling before analysis is minimal, reducing the chance of error from sample loss. The conditions desirable to be met for postcolumn derivatization are (i) although the reaction must be rapid, within about 2 min, and be reproducible, the reaction products do not have to be very stable, (ii) the reagent itself must have no, or a very low, detection response, and (iii) the reactor volume must be small to minimize dilution and band broadening.

The remainder of the article is divided into three major sections: GC, HPLC, and CE. The GC section considers derivatization of drugs by reagent class: alkylation, acylation, and silylation. In addition, a short discussion on sample handling is included. The HPLC section considers the derivatization of drugs based on the following types: alkaloids, amines, antibiotics, barbiturates, carbonyl compounds and carboxylic acids, catecholamines, hydroxy compounds, steroids, and sulfur compounds. A brief rationale for derivatization chemistry with CE concludes this article.

2 GAS CHROMATOGRAPHY

Prior to analysis by GC, compounds containing functional groups with active hydrogens such as COOH, OH, NH, and SH need to be protected. Compounds with these functional groups tend to form intermolecular hydrogen bonds, reducing volatility. They are also thermally unstable and can interact with either fused silica or the stationary phase, causing peak broadening. Several good review sources describing derivatization reagents, reactions, and applications for GC are available. At least one commercial company provides a good overview of the commercial reagents and the apparatus available.[1] Two other references are fairly comprehensive in describing many of the published methods for GC derivatization either by reaction class[2] or by the type of organic compound.[3] A recent overview describing gas chromatography/mass spectrometry (GC/MS) methods for drugs of abuse contains prederivatization protocols.[4] Alkylation, acylation, and silylation reaction chemistries are the most common approaches to derivatization for GC and are the primary focus in this article. A specific example and an overview for each of these reaction classes are described below.

2.1 Alkylation

Alkylation represents the replacement of an active hydrogen, found in an organic acid or amine, by an aliphatic or aliphatic–aromatic (e.g. benzyl) group. Alkylation reactions can also be used to prepare ethers from alcohols such as phenol, thioethers from sulfur compounds, and *N*-alkylamines, amides, and sulfonamides from amines. One principal chromatographic use is the conversion of carboxylic acids into esters that are stable and volatile with good chromatographic characteristics. Esterification of carboxylic acids has traditionally been done through reaction with an alcohol in the presence of an acid catalyst. Generally, a 1–2-mg sample of the acid is heated with 100 μL of the alcohol (methanol or ethanol) containing 3-M HCl for 30 min at 70 °C. The alcohol is evaporated, leaving the ester as a residue. A faster reaction in only about 2 min using a boiling water-bath is possible with a BF_3–methanol solution (Table 1, reagent 1). Often the esters are extracted into *n*-heptane followed by evaporation of the solvent.[5] *N*,*N*-Dimethylformamide dimethylacetal (DMFDA) is a rapid methylating agent, which provides quantitative yields upon mixing in a nonaqueous solvent with the sample of interest (Table 1, reagent 2). Carboxylic acids are converted to methyl esters, thiols to thioethers, and phenols to methyl ethers. Primary amines are converted into their *N*,*N*-dimethylaminoethylene derivatives and amino acids are simultaneously modified as this derivative at the amino group and as the methyl ester at the carboxyl group. Aliphatic hydroxyl groups are not methylated. Ethyl, propyl, *n*-butyl, and *t*-butyl acetals may be used analogously.

Alternatively methylation of active functional groups can be done using methyl iodide (CH_3I) or diazomethane ($CH_2{=}N^+{=}N^-$). Sixteen diuretics such as amiloride and hydrochorothiazide were extracted from urine and then reacted with methyl iodide at 70 °C for 2 h to form methyl derivatives. Short columns and high flow rates permitted analysis times by GC/MS of less than 4 min.[6] Carbamazepine, diclofenac, and ibuprofen were methylated

Table 1 Common GC derivatives – alkylation

Number.	Reagent[a]	Compounds	Derivatives
1	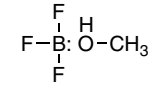 (BF_3–methanol)	R–COOH	$R-COOCH_3$
2	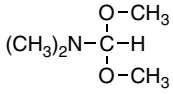 (DMFDA)	R–COOH $R-NH_2$ R–OH R–SH	$R-COOCH_3$ $R-NH-(CH_2)_2-N(CH_3)_2$ $R-O-CH_3$ $R-S-CH_3$
3	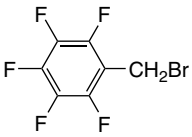 (PFB-Br)	R–COOH R–OH Sulfonamides R–SH	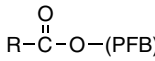 R–O–(PFB) 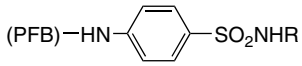 R–S–(PFB)
4	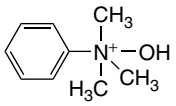 (TMAH)	R–OH $R-NH_2$ R–COOH	$R-O-CH_3$ $R-NH-CH_3$ $R-COOCH_3$

[a] DMFDA, *N*,*N*-dimethylformamide dimethylacetal; PFB-Br, pentafluorobenzyl bromide; TMAH, trimethylanilinium hydroxide.

using diazomethane and determined by GC/MS down to 10 ng L^{-1}, while sulfamethoxazole and its metabolites were determined by liquid chromatography/mass spectrometry (LC/MS).[7]

Pyrolysis of a quaternary ammonium salt (QUAT) in the presence of the organic analyte can be a convenient methylation procedure.[8] Upon injection of the QUAT mixed with the drug having a reactive amino, hydroxyl, or carboxyl group into the hot injection port (250–300 °C), the appropriate methyl derivative is formed (Table 1, reagent 4). The derivative is swept onto the analytical column for separation and detection. Trimethylanilinium hydroxide (TMAH) is generally the commercially recommended reagent, but trimethyl (α,α,α-trifluoro-*m*-tolyl)ammonium hydroxide can be used at a lower injection temperature, which is important for labile unsaturated fatty acids. This method is particularly good for barbiturates, sedatives, xanthine bases, phenolic alkaloids, and dilantin. Pharmaceuticals such as clofibric acid, ibuprofen, carbamazepine, naproxen, ketoprofen, and diclofenac were identified and quantitated by GC/MS via on-line derivatization using tetrabutylammonium salts with large volume (10 μL) sample injection.[9]

Pentafluorobenzyl bromide (PFB-Br) can convert carboxylic acids, phenols, sulfonamides, and thiols into halogenated derivatives that can be easily detected by electron capture (EC) (Table 1, reagent 3). For carboxylic acids, the reaction is run in an organic solvent in the presence of a base such as a tertiary amine for 5 min at 40 °C. The determination of cannabidiol (CBD) and the internal standard tetrahydrocannabidiol in blood plasma involved first derivatization by refluxing a hexane extract with PFB-Br overnight with stirring followed by purification on a Florisil™ column using hexane.[10] A typical EC chromatogram obtained on a packed 5% OV-225 glass column is shown in Figure 1. Although a packed column has limited separation capability, it can handle larger amounts of sample. In addition, because the hexane fraction was evaporated to dryness and reconstituted in a small hexane volume, a detection limit of 50 ng mL^{-1} of CBD in plasma was possible. A second reported example is the solid-phase extraction (SPE) of 19 pharmaceuticals from samples such as sewage, surface, ground, and drinking water. Prederivatization of these pharmaceuticals with PFB-Br before GC/MS was slightly more effective than silylation chemistry (explained in Section 3.3).[11] Detection limits down to 1 ng mL^{-1} were possible. Alternatively, tertiary amines have been derivatized with pentafluorobenzyl chloroformate prior to GC with EC detection.[12]

Alkylchloroformates have been utilized for the derivatization of amphetamines, often before solid-phase microextraction (SPME) and GC/MS. Amphetamine and methamphetamine in urine were immobilized on

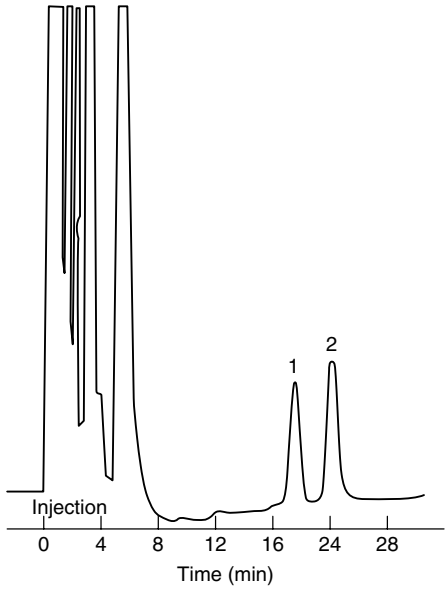

Figure 1 Typical chromatogram from a plasma sample containing 2.0 μg of the internal standard tetrahydrocannabidiol (1) and calculated to contain 1.5 μg of the CBD (2). (Reproduced with permission from Ref. 10.)

an Extrelut extraction column as free bases and then derivatized to *N*-propoxycarbonyl derivatives with propyl chloroformate [$CH_3(CH_2)_2(C{=}O)OCl$].[13] A follow-up study by the same authors used SPME to concentrate the same *N*-propoxycarbonyl derivatives from the vial headspace before GC.[14] Calibration curves were linear from 0.5 to 50 mg L^{-1}. A similar approach was used for amphetamine and ephedrine drugs using 2,2,2-trichloroethyl chloroformate to convert the compounds to the corresponding carbamates before GC/MS determination using electron impact and chemical ionization approaches.[15] Plasma, urine, and hair samples were analyzed and detection limits were from 2 to 5 ng mL^{-1} for the biological liquids and in the 0.1–0.2 ng mg^{-1} range for hair. After conversion of amphetamines to their free base using triethylamine, an SPME fiber coated with an alkylchloroformate was used to sample the headspace and cause derivatization.[16] More conventionally, alkylchloroformate derivatization of cocaine and two derivatives was done before SPME and GC/MS analysis.[17] Detection limits in human hair of 0.1–0.5 ng mg^{-1} were similar to those reported in reference 16.

Ethyl chloroformate activated (*R*)(+)-l-phenylethylamidation of *R* and *S* naproxen and desmethylnaproxen enantiomers before GC separation within 11 min on a 5% phenyl–95% methylsiloxane stationary phase.[18] Using the chiral derivatizing agent (*S*)(−)*N*-trifluoroacetyl-L-prolyl chloride (L-TPC) to modify the amine group, resolution at 2.0 of enantiomeric amphetamines was possible with good linearity by GC/MS to 45 ng mL^{-1}.[19]

2.2 Acylation

Acylation is the conversion of carboxylic acids, thiols, or amines (compounds that have active OH, SH, and/or NH moities) into esters, thioesters, and amides, respectively. Acyl derivatives, even geometric isomers, tend to undergo mass spectrometry (MS) fragmentation that leads to useful structural information and differentiation. Perfluoroacylimidazoles such as trifluoroacetylimidazole (TFAI), pentafluoropropionylimidazole (PFAI), or heptafluorobutylimidazole (HFBI) can react with hydroxyl groups and primary or secondary amines to generate the halogenated derivative and the relatively inert by-product imidazole (Table 2, reagent 1). Generally, the reaction times are 1–3 h at 75 °C. Indoleamines and indole alcohols, which are acid-sensitive compounds, have been modified by this procedure. Metoclopromide, an antispasmodic agent related to procainamide, has been derivatized with HFBI for GC with EC detection.[20] These reagents are often used in bifunctional derivatization schemes involving silylation chemistry. For example, phenolic amines can be silylated with *N*-trimethylsilylimidazole (TMSI) to protect the hydroxyl group and acylated to protect the amine group. Other halogenated imidazoles such as *p*-bromobenzyl- and *p*-chlorobenzylimidazole have been used because the detectability of the derivatives can be enhanced using EC. However, the volatility and stability of the derivatives under GC conditions tend to be lower compared to the fluorinated types.

N-Methylbistrifluoroacetamide (MBTFA), shown in Table 2 as reagent 3, permits trifluoroacetylation of primary or secondary amines and carbohydrates under mild nonacidic conditions.[21] For carbohydrates, a single derivative for mono-, di-, tri-, and tetrasaccharides was obtained. The *N*-methyltrifluoroacetamide by-product is stable and volatile, being compatible with GC. After extraction with *t*-butyl methyl ether and SPE, amphetamines have been derivatized with MBTFA and separated on a 5% phenylmethylsilicone column.[22] On-line flash injection of extracted amphetamines with MBTFA provided GC/MS separation and identification of 12 compounds at limits of detection of 2.5–6.9 ng mL^{-1}.[23]

Perfluoro acid anhydrides [R′(C=O)–O(C=O)R′], such as trifluoro-, pentafluoro-, and heptafluoro-acid anhydride (TFAA, trifluoroacetic anhydride; PFAA, pentafluoropropionic anhydride; HFAA, heptafluorobutyric anhydride) react readily with alcohols, phenols, and urine for the preparation of perfluoroacyl derivatives, which are particularly useful for EC detection (Table 2, reagent 2). Triethylamine is added, as part of the reaction mixture, to neutralize the acidic by-products and drive the reaction to completion. Drugs of abuse are often derivatized in this way before GC/MS confirmation.

Table 2 Common GC derivatives – acylation

Reagent[a]	Compounds	Derivatives
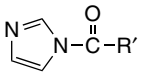	R–OH	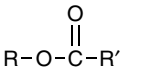
		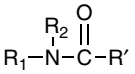
$R' = CF_3$(TFAI), C_2F_5(PFAI), or C_3F_7(HFBI)	$R–NH_2$	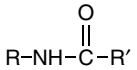
		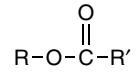
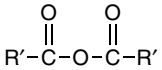	R–OH 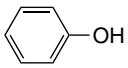	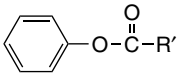
$R' = CF_3$(TFAA), C_2F_5(PFAA), or C_3F_7(HFAA)		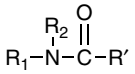
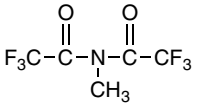 (MBTFA)	$R–NH_2$ R–OH R–SH	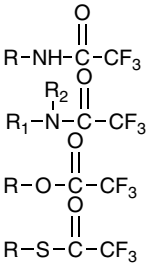

[a] TFAI, trifluoroacetylimidazole; PFAI, pentafluoropropionylimidazole; HFBI, heptafluorobutylimidazole; TFAA, trifluoroacetic anhydride; PFAA, pentafluoropropionic anhydride; HFAA, heptafluorobutyric anhydride; MBTFA, *N*-methylbistrifluoroacetamide.

A GC/MS example of derivatization of amphetamine and methamphetamine by HFAA[24] is shown in Figure 2(a) and (b). The splitting of the peaks is due to the presence of deuterated internal standards. Excess HFAA, which causes a high background level, column contamination, and eventual degradation of the stationary phase, was removed by the procedure outlined below. The drugs of interest were extracted from 2 mL of urine by SPE columns with a 10% 2-propanol in a chloroform elution solvent. Sodium periodate oxidation of interfering drugs was carried out if necessary. The extract was acidified with 10% HCl in methanol before evaporation and reconstitution of the residue in hexane with 0.1-M trimethylamine. HFAA was added and the mixture was heated in a sealed tube for 1 h at 70 °C. After cooling, deionized water was added to the tube, the tube shaken, and the aqueous phase discarded. A 4% ammonia solution was shaken with the remaining hexane volume and then the organic layer removed for injection into the GC/MS system. The limit of quantification and limit of detection were both on the order of 50 ng mL^{-1} and the linearity was observed up to at least 4000 ng mL^{-1}. Newer, but similar, approaches using HFAA for the derivatization of amphetamines and derivatives in biological samples use novel sample preparation methods. Examples of such methods are mixed mode reversed phase-ion exchange SPE[25,26] prior to derivatization, SPME with derivatization using a polydimethylsiloxane fiber,[27] and extraction with toluene containing HFAA.[28]

An important application of PFAA and HFAA derivatization involves the MS differentiation of substituted methamphetamine[29] or phenethylamine[30] isomers from the controlled drug 3,4-methylenedioxymethamphetamine (3,4-MDMA or Ecstasy). Without derivatization, the MS spectra for these GC-separated isomers are virtually identical to that of 3,4-MDMA. Acylation of the NH or NH_2 functional group present in all these compounds with either PFAA or HFAA lowers the basicity of the amine, which allows other fragmentation pathways to occur, resulting in clear distinguishing MS spectra.

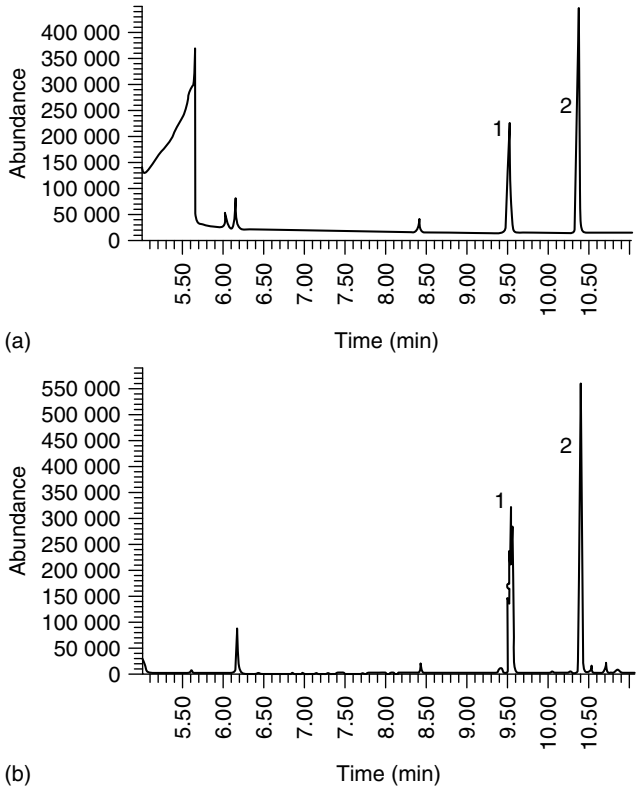

Figure 2 (a) Total ion current chromatogram of HFAA-derivatized amphetamine (1) and methamphetamine (2). The overloaded peak at about 5.6 min is due to excess HFAA. (b) Total ion current chromatogram of HFAA-derivatized amphetamine (1) and methamphetamine (2) after extraction of the excess derivatizing agent. The baseline level in this chromatogram is only about 10% of that in (a). The splitting of the drug peaks in both chromatograms is due to the presence of the deuterated internal standards (amphetamine-d_5 and methamphetamine-d_8). (Reproduced with permission from Ref. 24.)

Comparable approaches using TFAA for derivatization of phenylalkylamines such as methamphetamine[31] and others including phenteramine, cathinone, and mescaline[32] or ketamines[33] and subsequent GC/MS have been reported. Human metabolism of ephedrines including ephedrine and phenylpropanolamine was generally found to be low based on TFAA derivatization and GC/MS results.[34] Piperazine drugs and their metabolites were separated by LC/MS after SPE and forensic confirmation was made by GC/MS of their TFAA derivatives.[35]

The use of nonfluorinated anhydrides is less common. After derivatization with acetic anhydride in the presence of $KHCO_3$, six serotonin uptake inhibitors such as venlafaxine and fluvoxamine were extracted with a polydimethylsiloxane-divinyl benzene microextraction fiber for 30 min before GC/MS determination.[36] Amphetamines, using 2% acetic anhydride, were converted to primary and secondary amides upon injection of the sample solution and furan was used as the reagent for chemical ionization MS to improve selectivity.[37]

2.3 Silylation

Silylation is the most widely used derivatization scheme for GC. Active hydrogens from acid, alcohol, thiol, amine, and amide groups can all be protected, usually with trimethylsilyl (TMS) groups. The ease of silylation of these functional groups can be ordered: alcohol > phenol > carboxylic acid > amine > amide. Steric hindrance is also a factor within a class of organic compounds. For alcohols, the reactivity order is primary > secondary > tertiary, and, for amines, primary > secondary. Trimethylchlorosilane (TMCS) is the simplest reagent that can be used, often to derivatize carboxylic acids (Table 3, reagent 1). TMCS is often used in conjunction with hexamethyldisilazane (HMDS) (Table 3, reagent 2) to improve the silylation of sugars and related compounds. In general, silylation reagents are unstable and must be protected from moisture. Because an excess of the silylating agent is used in the derivatization reaction to minimize the problem of moisture or other acidic components in the sample, it is recommended that the excess silylation reagent be evaporated using a stream of nitrogen before injection of the sample into the GC column. This avoids several problems such as large reagent blank peaks in the chromatogram and fouling of the flame ionization detector by SiO_2 deposits. A polar stationary phase such as poly(ethylene glycol) (Carbowax™ 20 M) or free fatty acid phase is derivatized by excess silylation reagent and cannot be used.

Silylacetamides are represented in the most popular group of silylation agents, owing to their ability to react quickly and quantitatively under mild conditions. *N*,*O*-Bis(trimethylsilyl)acetamide (BSA) is a highly reactive TMS donor (Table 3, reagent 3). *N*,*O*-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) (Table 3, reagent 4) has the advantage of giving a more fluorinated by-product than BSA. This permits derivatization of lower molecular weight analytes without potential overlap of the trifluoroacetamide by-product.

Cannabinol derivatives were volatilized using the BSTFA reaction and then microextracted from the headspace before GC/MS.[38] Limit of detection values ranged from 0.01 to 0.02 ng mg^{-1}. The addition of TMCS to BSTFA promotes the derivatization of amides, secondary amines, and hindered hydroxyl groups. A comprehensive article on drug detection methodology describes the determination of cocaine, heroin, and metabolites in hair, plasma, saliva, and urine samples after isolation by SPE and subsequent derivatization with

Table 3 Common GC derivatives – silylation

Number.	Reagent[a]	Compounds	Derivatives
1	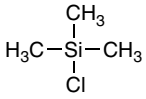 (TMCS)	R–COOH	
2	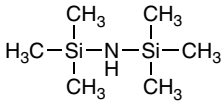 (HMDS)	R–OH	R–O–Si(CH$_3$)$_3$ 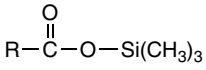
3	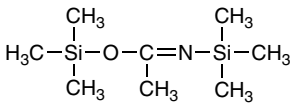 (BSA)	R–OH R–NH$_2$	R–O–Si(CH$_3$)$_3$ 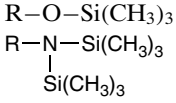
			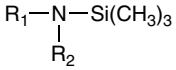
		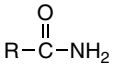	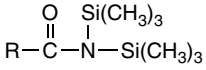
		R–COOH	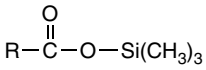
4	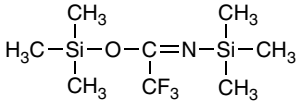 (BSTFA)	Same as BSA, leaving group 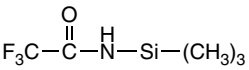 more volatile than that of BSA	
5	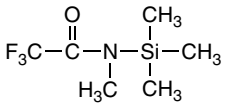 (MSTFA)	Same as BSA, leaving group 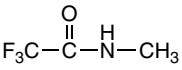 is very volatile	
6	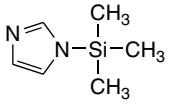 (TMSI)	R–OH (steroids) R–COOH	R–O–Si(CH$_3$)$_3$ 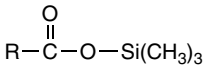
		No reaction with amines	

[a] TMCS, trimethylchlorosilane; HMDS, hexamethyldisilazane; BSA, *N*,*O*-bis(trimethylsilyl)acetamide; BSTFA, *N*,*O*-bis (trimethylsilyl)trifluoroacetamide; MSTFA, *N*-methyl-*N*-trimethylsilyltrifluoroacetamide; TMSI, *N*-trimethylsilylimidazole.

BSTFA with TMCS before analysis by GC/MS.[39] After spiking the plasma, saliva, and hair extracts with internal standards, they were passed through SPE columns and the drugs eluted with a methylene chloride-2 propanol-ammonia solvent. The eluent was evaporated to dryness, reconstituted with acetonitrile, and allowed to react with a BSTFA/TMCS mixture at 60 °C for 30 min before analysis by GC/MS. Representative chromatograms are shown in Figure 3(a–c). Separations were affected on a relatively short 12 m × 0.2 mm i.d. capillary column. The limit of detection for the analyte was about 1 ng mL^{-1} in urine, saliva, and plasma and 0.1 ng mg^{-1} in hair. A BSTFA/TMCS mixture was also shown to be effective for derivatization of the opioid analgesic

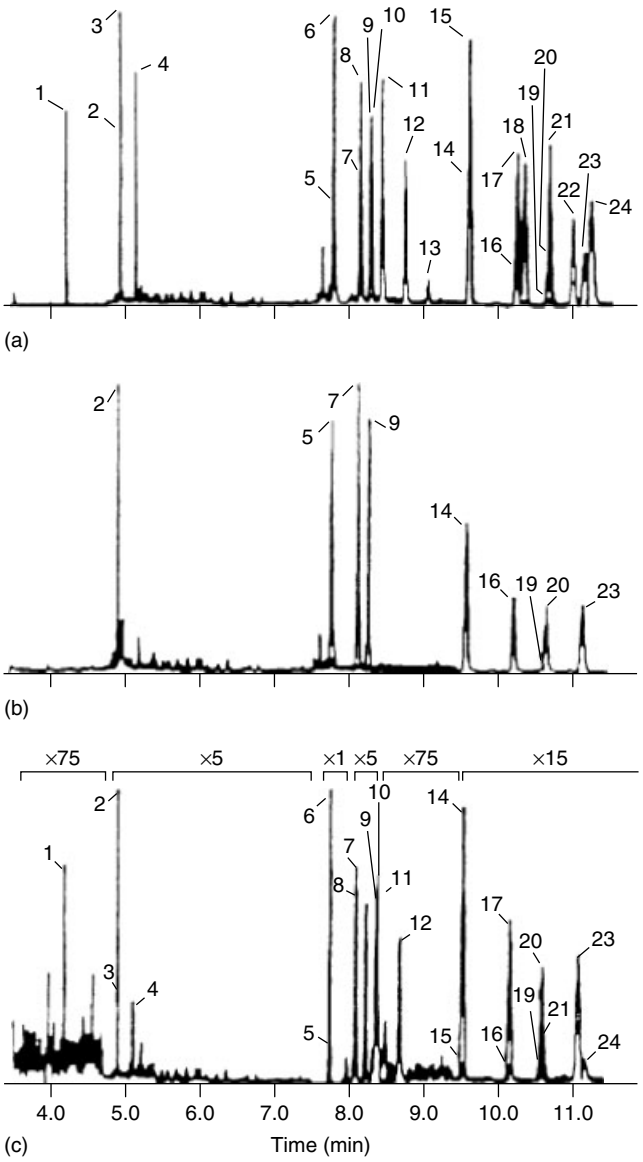

Figure 3 Single ion monitoring recordings of extracts from (a) standard cocaine/opiate hair, (b) drug-free control hair, and (c) a hair sample collected from a heroin user. Analytes are identified as follows: anhydroecgonine methyl ester (1), [^{2}H-3]ecgonine methyl ester (2), ecgonine methyl ester (3), ecgonine ethyl ester (4), [^{2}H-3]cocaine (5), cocaine (6), [^{2}H-3]cocaethylene (7), cocaethylene (8), [^{1}H-3]benzoylecgonine (9), benzoylecgonine (10), norcocaine (11), norcocaethylene (12), benzoylnorecgonine (13), [^{2}H-3]codeine (14), codeine (15), [^{2}H-3]morphine (16), morphine (17), norcodeine (18), [^{2}H-6]-6-acetylmorphine (19), [^{2}H-3]-6-acetylmorphine (20), 6-acetylmorphine (21), normorphine (22), [^{2}H-9]heroin (23), heroin (24). (Reproduced with permission from Ref. 39.)

buprenorphine in association with benzodiazepines,[40] as well as cocaine and derivatives[41,42] before GC/MS. Seven diverse pharmaceuticals including caffeine, estradiol, ibuprofen, ketoprofen, musk ketone, naproxen, and triclosan were also more effectively derivatized using the BSTFA + TMCS combination as compared to TMSI.[43]

N-Methyl-*N*-trimethylsilyl-trifluoroacetamide (MSTFA) has similar donor strength to BSA and BSTFA but generates an even more volatile by-product, *N*-methyltrifluoroacetamide (Table 3, reagent 5). Both, excess MSTFA and the by-product, often elute with the solvent peak, eliminating the presence of extra peaks in the chromatogram. Cocaine and its metabolites, benzoylecgonine and ecgonine methyl ester, were sequentially derivatized first with methyl iodide to obtain the ester derivative and then MSTFA to form the *o*-TMS derivative.[44] The derivatized compounds were separated by GC using a methylphenylsilicone column and nitrogen–phosphorus detection. Again, TMCS can be added to MSTFA to promote the derivatization of amides and hindered amines and hydroxyl groups. After SPE, acidic pharmaceuticals such as salicylate, ibuprofen, naproxen, ketoprofen, and diclofenac, as well as carbamazepin, were derivatized with MSTFA + 10% TMCS and determined by GC/MS at nanogram per milliliter levels in wastewater.[45]

The strongest reagent for silylation of hydroxyl groups is TMSI (Table 3, reagent 6). It does not react with amines or amides and is effective for most steroids, including those with unhindered and highly hindered OH groups. It also reacts quickly and smoothly with carboxyl groups. The carboxylated tetrahydrocannabinol was first isolated from urine using a sulfonated polymer SPE cartridge before derivatization with TMSI to form the TMS derivative that could be detected in the range from 1 to 50 ng.[46]

Silyl imidazole compounds, other than those that derivatize with a TMS group, have also been used in conjunction with GC. Methyltestosterone was derivatized using either the dimethylethylsilyl (DMES)-imidazole or dimethylisopropylsilyl (DMIPS)-imidazole to the corresponding silyl ether derivative.[47] Reaction conditions involve heating at 70 °C for 1 h before removal of the excess reagent by N_2 and reconstitution of the residue in cyclohexane. Because the desired application was the assay of methyltestosterone in bulk powder and tablets, separation on a packed column was adequate in the concentration range 0.1–1.5 mg mL^{-1} (Figure 4a–c).

Alternatively, *N*-methyl-*N*-(*t*-butyldimethylsilyl)-trifluoroacetamide (MTBSTFA) derivatizes hydroxyl, carboxyl, thiol, and primary or secondary amine substituents by adding a *t*-butyldimethylsilyl (TBDMS) group. These TBDMS derivatives are more stable to hydrolysis than TMS compounds and, when analyzed by GC/MS, a strong M-57 fragment is noted, which can be used to determine the molecular weight of the original compound. These advantages have been demonstrated in

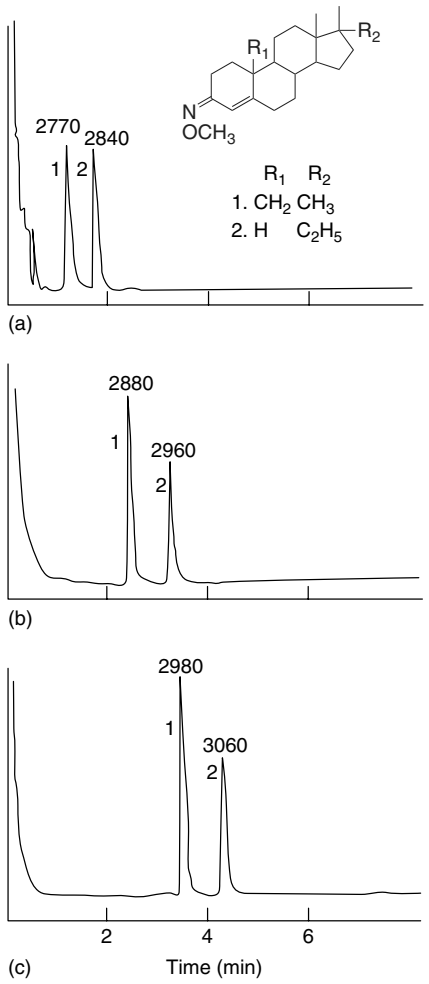

Figure 4 Typical gas chromatograms of (1) methyltestosterone and (2) norethandrolone as (a) TMS, (b) dimethylchlorosilane (DMCS), and (c) DMIPS derivatives. Note the bulkier derivatives are shifted to a cleaner area in the GC trace. (Reproduced with permission from Ref. 47.)

the determination of short- and long-chain carboxylic acids.[48] A factorial experimental design comparison showed MTBSTFA had a better overall performance than BSTFA for selected pharmaceuticals such as salicylate, ibuprofen, naproxen, propranolol, carbamazepine, and others.[49] Thirty drugs of abuse, including amphetamine-type drugs derivatized with HFAA, benzodiazepines and cannabinols with MTBSTFA, and buprenorphine with MSTFA were separated with fast temperature programming and a high flow rate on 30 m × 0.32 mm i.d. capillary columns after SPE from 250 μL of human oral fluid.[50] The same authors focused on the rapid separation of MTBSTFA derivatized benzodiazepines and other drugs (18 in all) in 4.40 min using a high initial column temperature, fast temperature programming, a thin stationary phase film thickness, and hydrogen as the carrier gas.[51] Limits of quantitation were from 1 to 100 ng mL^{-1}.

A wide variety of drugs including benzodiazepines, cannabinoids, opioids, antidepressants, antipsychotics, antiepileptics, and others were derivatized using a two-step MTBSTFA, MSTFA reaction sequence and the subsequent *t*-butyldimethyl silyl and TMS derivatives were determined by GC/MS.[52] A similar two-step derivatization approach was used for β-adrenoceptor ligands and methyl iodide in potassium carbonate and acetone methylated the diuretics before GC/MS of these mixtures in about 5 min using narrow bore capillary columns.[53] To enhance SPME using a polystyrene–divinylbenzene fiber, phenazone drugs were first acetylated on the nitrogen atom using acetic anhydride but then silylated with MTBSTFA, which removes the acetyl group to enhance GC/MS detection.[54]

Pentafluorophenyldimethylsilyl (flophemesyl) derivatives are often generated from steroids by reaction with flophemesylamine at room temperature for 15 min. This is a selective reaction with only primary and secondary hydroxyl groups in steroids reacting in the presence of unprotected ketone groups. GC with EC detection provides detection limits in the nanogram–picogram range. Flophemesyl derivatives also have favorable advantages for GC/MS, producing diagnostic ions that carry more of the current than TMS derivatives.[55]

2.4 Sample Handling

Generally, drugs are found in aqueous matrices such as urine and plasma; extraction with an organic solvent is required to isolate the drug in a nonaqueous solvent that is compatible with most derivatization reagents as well as to remove inorganic salts that are not compatible with GC. An alternative approach, direct derivatization, combines the extraction and derivatization steps together. Using this approach for drugs in untreated biological samples prior to GC analysis has the primary advantages of improved extractability of a derivatized polar compound as well as the avoidance of stability problems. Direct derivatization of drugs in untreated biological samples can mean not only derivatization in the sample matrix followed by extraction but also a two-phase reaction where the derivatization takes place in the organic phase while extraction of the analyte is continuing from the aqueous phase.[56] Extractive alkylation has been applied to valproic acid and ketoprofen, in which the tetrabutylammonium ion acted as the anion-pairing agent to pull the compound into the organic phase, where alkylation to generate the methyl or phenacyl derivative was possible. Phenolic compounds such as clioquinol have been determined in an analogous fashion. Acylation involving perfluorinated anhydrides and chloroformates has also been used for the determination of drugs such

as metanephrine and normetanephrine. Chloroformates can dealkylate tertiary amines to form stable carbamates.

SPME uses a fiber, usually fused silica coated with a stationary phase, to provide extraction, preconcentration, and sample introduction into a single step. The actual syringelike device has an outer septum-piercing needle and a plunger that is the fiber. The fiber can be inserted into a liquid sample matrix or the headspace above a solid or liquid sample. Upon equilibrium, analytes of interest are partitioned between the stationary phase of the fiber and the sample matrix or headspace. After analyte preconcentration into the fiber, the syringe device is inserted into the GC injection port where the analytes are thermally desorbed and swept onto the column for separation. For example, endocrine steroid hormones were extracted from biological and environmental samples using a polyacrylate fiber, which was then introduced into the headspace of a BSTFA solution to cause analyte derivatization before insertion into the GC injection port.[57] It has been reported that SPME, using a carboxene/polydimethylsiloxane fiber, can be effective for adsorption of both the fatty acid analyte and the derivatizing agent phenyltrimethylammonium hydroxide before fiber insertion into the GC injection port where methylation of the carboxylic acid analytes occurs.[58] The optimum sample–fiber interaction time was 20 min and the linearity of response for octyldecylcarboxylic acid was about 2 orders of magnitude from about 3×10^{-6} M. It would seem that this approach could also be applied to some drugs such as naproxen.

3 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

A variety of chemistries can be adopted for either pre- or postcolumn derivatization in conjunction with HPLC because either aqueous or organic solvent compatible reactions are possible. Postcolumn photochemical derivatization has the advantage of instrumental simplicity with no problems of pumping a reagent or detection background from the reagent. Most of the books reviewing fundamental reactions for pre- and postcolumn derivatization chemistry in conjunction with HPLC were published in the 1980s, which appears to have been the most active time for research in this field.[59–65] Many review articles have focused on postcolumn derivatization with HPLC.[66–73] Some specific review articles on derivatization chemistry of pharmaceutical compounds with HPLC have also been published.[74–78] Although the other articles do not try to be comprehensive, Ahuja[77] gave a virtually complete review of the topic up to 1979. Danielson et al.[78] published another fairly complete review covering the period 1979–1987, which is the basis for this part of the article. Again, a specific example and an overview for each major class of pharmaceutical compounds are given.

3.1 Alkaloids

Separation of atropine and ergotamine by normal-phase liquid chromatography (LC) followed by post-column ion-pair extraction with an aqueous solution of 9,10-dimethoxyanthracene-2-sulfonate has been reported.[79] The organic mobile phase was monitored fluorimetrically with detection limits of 40–100 ng. Conversely, methadone, phencyclidine, and their metabolites were separated by reversed-phase HPLC using 9,10-dimethoxyanthracene-2-sulfonic acid in the mobile phase.[80] The resultant ion pairs (Table 4, reagent 3) were extracted with chloroform postcolumn on-line and detected fluorimetrically with detection limits of 1–6 ng mL^{-1} in plasma. A precolumn ion-pair derivatization method for atropine, hyoscyamine, scopolamine, and ergotamine, involving picric acid and normal-phase chromatography, provided ultraviolet (UV) detection limits of about 200 ng.[81,82]

Morphine, which can be oxidized to the dimer pseudomorphine using alkaline hexacyanoferrate (III), has been detected fluorimetrically after this postcolumn reaction.[83] Other opiates, such as normorphine, naborphine, codeine, norcodeine, and others, were also reactive. The excitation and emission wavelengths were 323 and 432 nm respectively. The mobile phase, a 12.5 : 87.5 methanol–0.1 M KBr solution, was propelled through a C_{18} reversed-phase column. The derivatizing agent was 50 mg of $K_3Fe(CN)_6$ in 250 mL of 4 M ammonia solution delivered at 0.4 mL min^{-1}. The selectivity advantage of fluorescence over UV detection for the determination of morphine in urine and serum samples was definitive using the instrumentation shown in Figure 5. Morphine could be determined at the 2–30 μg mL^{-1} level in such biological samples with detection limits of about 10 ng. Fluorescence detection of morphine and related opiates was improved about twofold after postcolumn derivatization with alkaline hexacyanoferrate(III) and micelle formation with the nonionic surfactant Triton X.[84] A detection limit of 0.2 pmol was possible for morphine after precolumn dansylation.[85] The heroin metabolite 6-acetylmorphine has been determined in urine by reversed-phase HPLC with fluorescence detection after automated precolumn oxidation with hexacyanoferrate(III)[86] at a reported detection limit of 1 ppb. Reserpine, an antihypertensive agent, was detected fluorimetrically after postcolumn reaction with nitrous acid and UV irradiation.[87,88] Using

Table 4 Common fluorescent (F) and electrochemical (E) derivatives for HPLC

Number.	Compounds	Reagent(s)	Derivative
1	$R'NH_2$ (F)	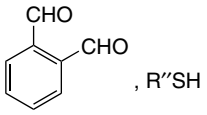	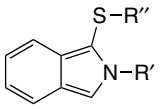
2	RNH_2 (F)	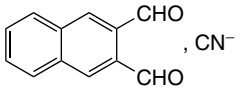	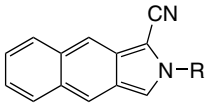
3	R_3NH^+ (F)	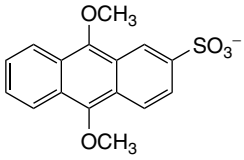	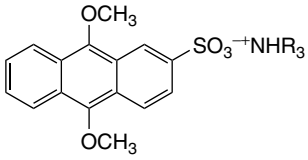
4	RSH (F)	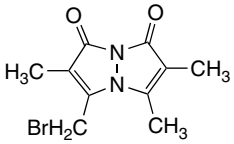	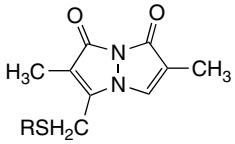
5	RCOOH (E)	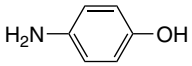	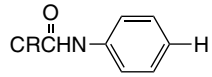
6	Steroid (E)	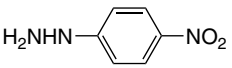	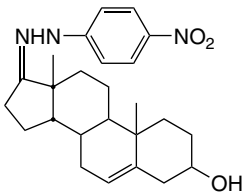

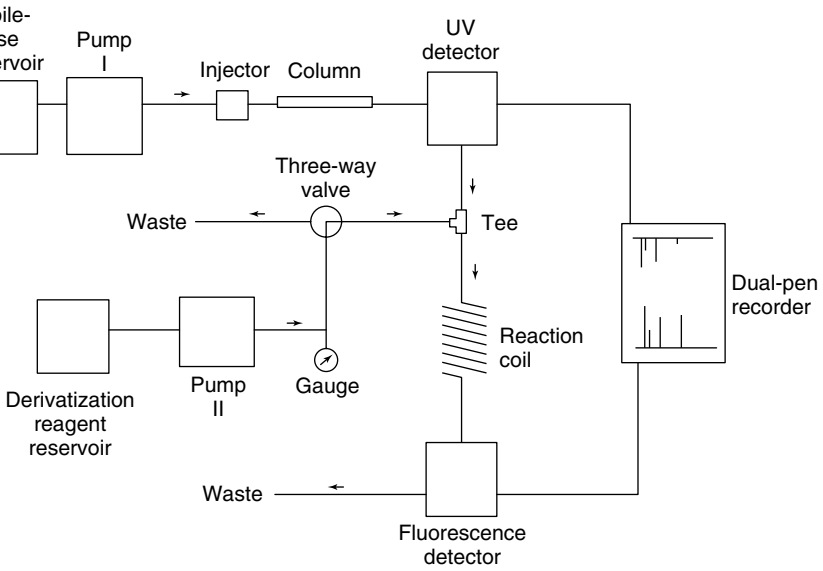

Figure 5 HPLC system for morphine using postcolumn derivatization and fluorescence detection. (Reproduced with permission from Ref. 83.)

this UV photochemical reaction, a 20-fold increase was found in the signal over the native fluorescence of reserpine.

Physostigmine, an acetylcholinesterase inhibitor, was separated from its degradation products and reacted with coulometrically generated bromine.[89] Electrochemical detection of unreacted bromine was inversely proportional to the amount of drug, and a detection limit of 0.5 ng could be attained. An on-line photochemical reaction detector caused decreased fluorescence of ergot alkaloids, permitting the identification of these compounds in complex chromatograms.[90] A similar system has been shown to convert cannabinol into a fluorescent derivative, providing detection limits of less than 1 ng in urine.[91] The use of this approach was extended for the detection of four Cinchona alkaloids and their dihydro derivatives as well as pyridoxine (vitamin B6) after isocratic reversed-phase separation.[92] On-line reaction of cannabinoids with fast blue salt B produced colored derivatives for detection at 490 nm.[93] Pilocarpine was quaternized with *p*-nitrobenzyl bromide before reversed-phase HPLC separation and detection at 254 nm.[94]

3.2 Amines

Primarily, fluorescent methods have been developed for amino acids and peptides. The *o*-phthalaldehyde (OPA), mercaptoethanol (or other thiol) reaction (Table 4, reagent 1) has been studied in both pre- and postcolumn modes; specific articles describing such chemistry with common amino acids are cited in general references.[59–62] Other amino acids such as *s*-carboxymethyl-L-cysteine,[95] baclofen,[96,97] and melphalan[98] have also been derivatized with OPA. Fourteen amino acids were determined fluorimetrically in 35 min by gradient reversed-phase LC after OPA prederivatization and sequential injection.[99] OPA with *N*-acetyl-L-cysteine effectively derivatized 20 amino acid enantiomers found in parental nutrition preparations prior to isocratic LC separation.[100] Carbowax SPME fibers were used to extract OPA-*N*-acetyl-L-cysteine derivatives of amphetamines from urine samples before chiral LC that established the enantiomeric composition at a low milligram per liter level.[101] Phanquinone (4,7-phenanthroline-5,6-dione) was used as a fluorogenic precolumn derivatizing agent for amino acids and was compared with the OPA method.[102] Although a longer reaction time of 60 min at a higher temperature of 60 °C was required for the phanquinone method, the derivative products were more stable. A more recent modification of this method is to use naphthalene-2,3-dicarboxyaldehyde with cyanide ion (Table 4, reagent 2).[103] Detection limits down to 200 fmol and improved stability of the derivatives were the stated advantages.

γ-Aminobutyric acid was modified with dansyl chloride before reversed-phase HPLC and fluorescent detection.[104] Detection at 360 nm was possible for *N*-acetylcysteine after reaction with 2,4-dinitro-1-fluorobenzene.[105] 1,2-Diamino-4,5-dimethoxybenzene was used to form a fluorescent derivative of *p*-hydroxybestatin.[106] A review of precolumn derivatization methods targeting only amino acids is available.[107] The peptide leupeptin was reacted through the guanidino moiety with benzoin to form a fluorescent derivative.[108] Felypressin, a nonapeptide, was derivatized with fluorescamine, providing detection limits of 0.3 ng.[109] A dinitrophthalic anhydride reaction of peptides permitted either electrochemical or absorbance detection.[110] An ion-pair detection technique has been applied to hydrophobic amino acids and peptides.[111]

A variety of other fluorescent methods have also been reported for specific drugs. Tranexamic acid has been detected using OPA.[112] Biogenic amines such as tyramine, tryptamine, and serotonin were reacted with OPA in a post-[113] or precolumn[114] mode. The OPA precolsumn fluorescent reaction has also been applied to L-buthionine-(*S*,*R*)-sulfoximine in plasma[115] and mexiletine.[116] Fluorescent derivatization with fluorescamine has been applied in the postcolumn mode for sulfapyridine[117] as well as in the precolumn mode for tocainide[118] and the antiepileptic drugs vigabatrin and gabapentin.[119] Enantiomers of mexiletine have also been resolved.[120] After conversion of the drug panthenol to aminopropanol, the fluorescamine reaction provided detection limits of 0.4 μg.[121] The determination of debrisoquine and its hydroxy metabolites was possible by reaction through the guanidine moiety to form fluorescent compounds.[122] A postcolumn photochemical reactor caused the cleavage of methotrexate to form the highly fluorescent 2,4-diaminopteridine-6-carboxyaldehyde.[123] Indolethylamines were condensed with an aldehyde or α-keto acid to form fluorescent carboline derivatives.[124] Thiamine (vitamin B_1) was oxidized with hexacyanoferrate(III) in base to give the fluorescent thiochrome in either the precolumn[125] or the postcolumn[126] mode. Riboflavin, already fluorescent, and thiamine, after precolumn oxidation to thiochrome using hexacyanoferrate(III), were determined by HPLC in a variety of food powders such as flour, milk, and beans.[127] Hydrazine compounds, such as isoniazid, were reacted with *m*-fluorobenzyl chloride before reversed-phase separation and detection at 220 nm.[128] A precolumn coumarin reaction permitted fluorescent detection of 5-fluoro-2′-deoxyuridine.[129] Fluorescein isothiocyanate (FITC) modified a bronchodilator before HPLC-fluorescence analysis.[130] Using excitation at 470 nm and emission at 537 nm, the 7-nitrobenzofurazan derivative of gabapentin

was determined by reversed-phase C_{18} silica LC with a limit of quantification of 2 ng mL^{-1} in human serum.[131]

Some UV derivatization methods have been developed for aliphatic and even aromatic amines. Phenylisothiocyanate (PITC) has been used as a precolumn derivatizing agent for amino acids in conjunction with reversed-phase HPLC. Both primary and secondary amines can react with PITC (Table 5, reagent 1(B)) and the aromatic tag provides good retention for the smaller amino acids.[132] Both primary and secondary amines were derivatized with 9-fluorenylmethyl chloroformate (FMOC) before separation by reversed-phase HPLC with UV detection at 254 nm. After SPE of a urine sample, β-phenethylamine was determined with a detection limit of 0.1 ng mL^{-1}.[133] Amphetamine and methamphetamine have been reacted with 4-methoxybenzoyl chloride (Table 5, reagent 1(A)) and other acid chlorides for subsequent UV detection.[134] A comparison of methods for amphetamines has been made.[135] The enantiomeric separation of (−)-1-(9-fluorenyl)ethyl chloroformate chiral derivatives of primary and secondary amphetamines was possible either by UV or fluorescence detection with resolution ranging from 0.9 to about 2.[136] The cardiotonic drug heptaminol has been derivatized with aminoazobenzene-4-isothiocyanate for UV detection at 420 nm[137] or OPA for electrochemical detection.[138] Secondary amines, such as piperazines, were converted into UV derivatives with *m*-toluoylacyl chloride.[139] Enantiomers of metoprolol have been separated after reaction with the chiral reagent (*S*)-(−)-phenylethyl isocyanate.[140] The determination of isophenindamine in the presence of phenindamine

Table 5 Common UV derivatives for HPLC

Number.	Compounds	Reagent(s)	Derivative
1	RNH_2	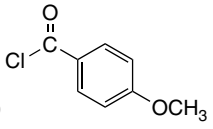	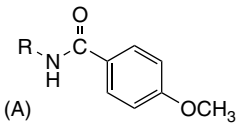
			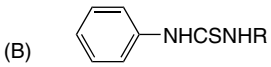
2	R_2NH	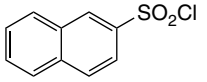	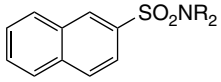
3	R_3N	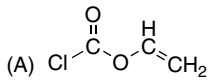	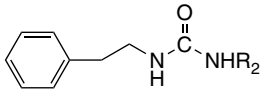
		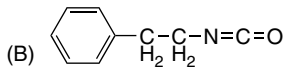	
4	RCOOH	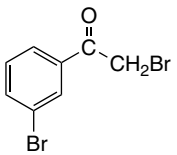	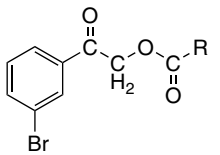
5	RSH	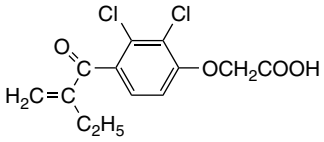	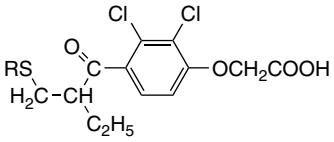
6	Penicillin	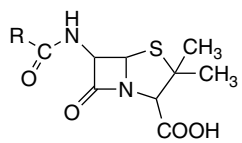	NaOH, $HgCl_2$, ethylenediamine tetraacetic acid (EDTA)
			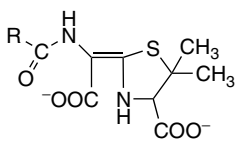

tartrate has been achieved by forming charge transfer complexes using $AgNO_3$ and reversed-phase HPLC.[141]

Antihistamines and other pharmaceuticals with a tertiary amine moiety have been derivatized to provide for luminescence detection. An ion-pair extraction detector using dimethoxyanthracene sulfonate (Table 4, reagent 3) has permitted the fluorescent determination of tertiary amines, such as bromopheniramine and chlorpheniramine,[142,143] ephedrine,[144] and hyoscyamine.[145,146] with detection limits of 200–500 pg. Precolumn reaction of chlorpheniramine with benzyl chloroformate gives a fluorescent derivative with a detection limit of 0.1 ng mL^{-1}.[147] Antihistamines, such as diphenhydramine, were converted through the tertiary amine group into fluorescent derivatives using 2-naphthyl chloroformate.[148] Aliphatic tertiary amines can be determined at the picomole level by postcolumn chemiluminescent detection, using tris(bipyridyl)ruthenium(III).[149] This topic of postcolumn electrochemiluminescence detection has been reviewed for both LC and CE.[150] Antihistamines in urine have also been separated by HPLC and detected with good selectivity using this ruthenium metal complex (Figure 6a–c).[151] Tamoxifen and its metabolites in human serum can be UV photochemically activated to form fluorescent phenanthrenes.[152] Postcolumn UV irradiation of tamoxifen and its derivatives caused rearrangement to a substituted phenanthrene, permitting fluorescent detection with 0.1 ng mL^{-1} detection limits.[153] Several analgesic compounds such as phenazone, propylphenazone, and dimethylaminophenazone, as well as their metabolites, were derivatized in situ with acetic anhydride to make more hydrophobic compounds that could be isolated from the sample matrix using SPE, before their separation by LC/MS/MS.[154]

Derivatives of nicotine, cotinine, and other metabolites in urine after HPLC separation can be detected at 530 nm by using diethylthiobarbituric acid as a color-forming agent.[155] Nicotinic acid and niacinamide were derivatized through the hetereocyclic nitrogen using cyanogen bromide and then an arylamine to give polymethine dyes, which were separated by micellar LC.[156] DL-Carnitine and acetyl-DL-carnitine were reacted with (α-bromo)methylphenyl ketone to form chiral derivatives that were separated on a chiral LC column to determine the enantiomeric purity of the L form.[157] Racemic phenothiazines were *N*-demethylated with vinyl chloroformate to form their secondary amines, which could then be reacted with (*R*)-(+)-1-phenylethyl isocyanate (Table 5, reagent 3), and these compounds were separated by reversed-phase HPLC.[158] Tertiary tetrahydroisoquinolines, such as diclofensine, were oxidized before photochemical conversion to fluorescent isoquinolinium derivatives.[159] After postcolumn ion-pair extraction of

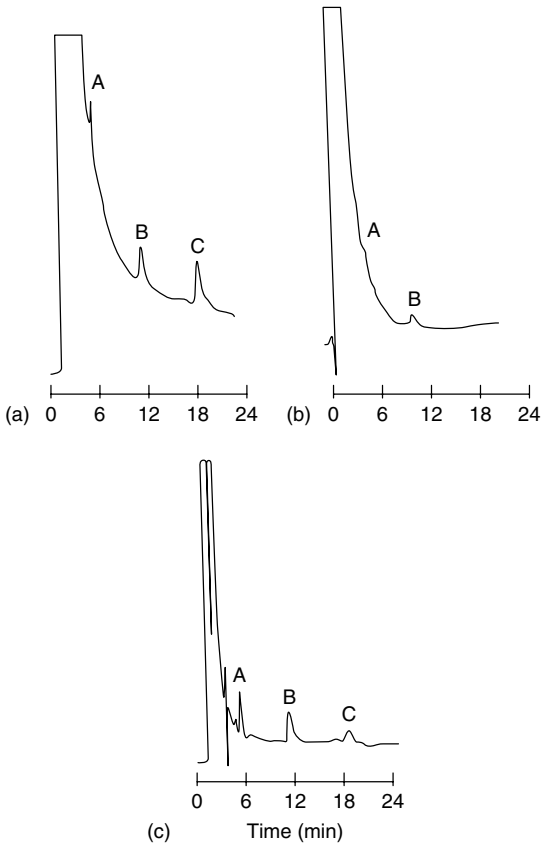

Figure 6 Chromatograms of an undiluted urine sample spiked with (A) 0.15 μg mL^{-1} pheniramine, (B) 0.26 μg mL^{-1} bromopheniramine, and (C) 0.29 μg mL^{-1} diphenhydramine taken with (a) UV (214 nm), (b) UV (254 nm), and (c) Ru $(bpy)_3^{3+}$ chemiluminescence detection. (Reproduced with permission from Ref. 151.)

secoverine, peroxylate chemiluminescence (CL) detection was carried out.[160]

Postcolumn detection of platinum(II) antineoplastic agents such as cisplatin is possible, as shown in Figure 7. The cisplatin-derived species are first reacted with an oxidant such as dichromate to form an activated species, which then combines rapidly with bisulfate to give a UV-absorbing complex. Using knitted open-tubular reactors, delay times of 26 s for the dichromate reaction and 4.7 min for the bisulfate reaction were found to be optimal. Figure 8(b) shows the substantial improvement in absorbance response of the separated derivatives as compared to the detection (Figure 8a) of the unmodified platinum compounds. Detection limits of 5–10 μg mL^{-1} were possible. Cisplatin was also determined in a plasma ultrafiltrate sample at the 5-ng level.[161]

3.3 Antibiotics

This class of pharmaceuticals has received major attention for both pre- and postcolumn derivatization. Fluorescent derivatization of the primary amine group of

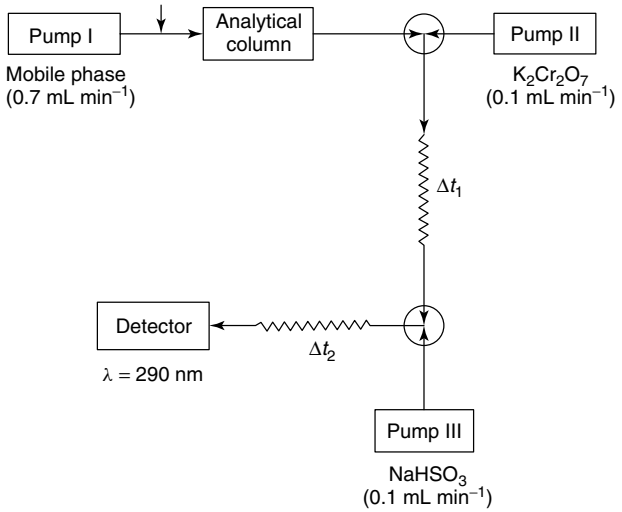

Figure 7 Schematic representation of the Pt–HSO_3^- reaction detector for cisplatin-type compounds. (Reproduced with permission from Ref. 161.)

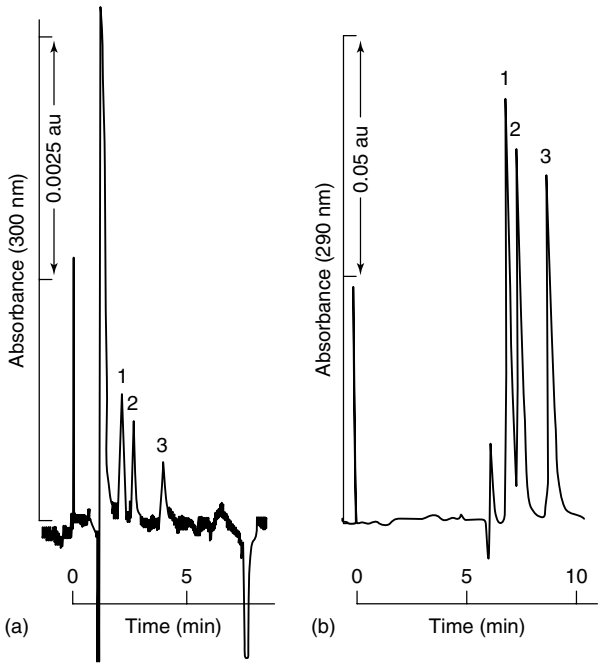

Figure 8 Comparison of UV detection (300 nm) after HPLC separation (a) and reactor response (290 nm) to derivatized platinum (b) for a mixture of *cis*-dichloroplatinum complexes: (1) *cis*-dichloro(1,2-diaminocyclohexane)platinum(II), (2) *cis*-dichloro(ethylenediamine)platinum(II), and (3) *cis*-dichlorodiammineplatinum(II). (Reproduced with permission from Ref. 161.)

many antibiotics with OPA and a sulfhydryl compound, often mercaptoethanol, has been commonly performed (Table 4, reagent 1). Gentamicin,[162,163] penicillin V, after enzymatic conversion to 6-aminopenicillanic acid,[163] penicillin N,[164] cephalosporin C,[165] cycloserine,[166] and fludalanine[167] were assayed by using an OPA postcolumn reactor. Spectinomycin, after postcolumn oxidation with hypochlorite, could also be derivatized with OPA in a second reaction coil.[168] Reversed-phase separation was employed for all these separations because the OPA reaction is carried out in an alkaline buffer solution. Precolumn reactions with OPA for gentamicin,[169] sisomicin,[170] and phosphinothricin and its alanine analog[171,172] before reversed-phase separation have all been reported. Amikacin isomers were first adsorbed on a silica gel column before reaction with OPA; the derivatives were eluted with ethanol before separation by reversed-phase HPLC.[173] A comparison of pre- and postcolumn OPA reaction conditions for gentamicin has been made.[174] For all these OPA fluorescence methods, detection limits are about 1 μg mL^{-1}. For example, β-lactams found in microbial fermentation broths were separated on a C_{18} HPLC column and detected fluorimetrically with excitation at 350 nm and emission at 450 nm after reaction with OPA and mercaptoethanol. Cephamycin, penicillin N, cephalosporin C, and 6-aminopenicillanic acid were separated in about 12 min at a flow rate of 1.5 mL min^{-1} with postcolumn conditions involving the OPA reagent at pH 12, a reaction coil of 12 m, a temperature of 90 °C, and a reagent flow rate of 0.8 mL min^{-1}. Detection limits of less than 0.5 and 1.0 μg mL^{-1} were achieved for penicillin N and cephalosporin C respectively.[175]

Numerous other postcolumn UV, fluorimetric, and electrochemical methods have been reported for tetracycline-type antibiotics and related compounds. Formation of a mercuric mercaptide of penicillins permits UV detection at 310 nm with detection limits of 10 ng.[176] Sulfonamides in egg, milk, and meat samples have been detected at 450 nm after derivatization with *p*-dimethylaminobenzaldehyde.[177] Monensin, narasin, and salinomycin in animal feeds have been reacted with vanillin to give products detectable at 520 nm.[178] Penicillins, such as amoxicillin, ampicillin, and others, can be separated by reversed-phase HPLC with alkaline degradation in the presence of mercuric chloride[179] (Table 5, reagent 6). Methanol promotes this reaction, forming the ring-cleaved product, which absorbs at 274 nm, and is detectable at 50 ng mL^{-1}. It was discovered later that sodium hypochlorite could replace the mercuric chloride reagent while maintaining a 1-min hydrolysis time.[180] Application of this latter method to penicillins in biological samples has been made.[181,182] Fluorescent detection of streptomycin reacted through the guanidino groups with naphthoquinone-4-sulfonate in alkaline solution has been reported in serum samples.[183] Photothermal derivatization of ciprofloxacin and its metabolites permits fluorescent detection throughout a linear range of about 2–1000 ng mL^{-1}.[184]

Photolysis with electrochemical detection of penicillins and cefoperazone has provided detection limits of about 6 ng.[185] Tetracycline, oxytetracycline, chlorotetracycline, demeclocycline, doxycycline, and meclocycline were separated on a C_{18} column and then photoirradiated before chemiluminescent detection with rhodamine and cerium(IV).[186] Detection limits in the 0.1–0.3 ng mL^{-1} range were reported. Electrochemically generated bromine was used as an oxidizing agent for cephalosporins and their decomposition products.[187] The excess bromine was detected at 0.4 V using a glassy carbon electrode. A comparison of this method with UV detection and postcolumn fluorescamine derivatization has recently been summarized.[188] Acidification of a serum sample sometimes improved recovery of cephalosporins when micellar chromatography was used with sodium dodecyl sulfate (SDS) in the mobile phase.[189] Fluorescamine has been used in an automated system for amoxycillin in biological fluids.[190] CL detection of clindamycin phosphate using tris(bipyridyl)ruthenium(III)[191] gave detection limits of 8 ppb compared to 970 ppb using UV detection at 214 nm.

A variety of reagents are also available for precolumn derivatization of antibiotics and UV or fluorimetric detection. Phenacyl esters of some natural penicillins were prepared using dibromoacetophenone before reversed-phase HPLC with UV detection.[192] Nitrobenzene derivatives of neomycin B and C can be separated by normal-phase HPLC and detected at 350 nm.[193] Similar methods using 1-fluoro-2,4-dinitrobenzene to form 2,4-dinitrophenyl derivatives of neomycin B and C[194,195] and amikacin[196,197] have also been published. Neomycin and other aminoglycosides have been converted to benzoyl derivatives before UV detection at 230 nm.[198] Aminoglyosides have also been reacted with 2,4,6-trinitrobenzenesulfonic acid, permitting UV detection at 350 nm.[199] Ion-pair formation of methscopolamine bromide with an aromatic anion during reversed-phase HPLC permitted UV detection.[200] The secondary amine group of spectinomycin was derivatized with 2-naphthalenesulfonyl chloride (Table 5, reagent 2) before normal-phase HPLC and UV detection at 254 nm.[201] Nitrofurantoin drugs furazolidone, nitrofurazone, furaladone, and nitrofurantoin were hydrolyzed and the released side chains derivatized with 2-nitrobenzaldehyde before SPE and determination by LC/MS with atmospheric chemical ionization.[202] Nitrofurantoin antibiotics and metabolites in honey were also determined as 2-nitrobenzaldehyde derivatives by LC/MS/MS and the predominant compound was found to be furazolidone.[203]

The reagent FMOC formed fluorescent derivatives of natural penicillins, cephalosporins, and their precursors in biological samples.[204] β-Lactams were prederivatized with FMOC for 5 min at 20 °C in a borate buffer at pH 7.7 before injection of the derivatives onto a C_{18} reversed-phase column. Excitation and emission wavelengths for the fluorescent detection were 260 and 313 nm respectively. A typical chromatogram is given in Figure 9, showing the modest acetonitrile gradient until after elution of ampicillin in which the acetonitrile is taken to 100%. Detection limits were 0.01 and 0.05 $\mu g\,mL^{-1}$ for 6-aminopenicillanic acid and isopenicillin N, respectively. Application of the method to fermentation broths was made over a range of 0.05–100 $\mu g\,mL^{-1}$. A maleimide reagent gave a fluorescent derivative for penicillamine.[205] An imidazole–mercuric chloride reagent can convert penicillins into mercury stabilized penicillanic acids after reaction at 50 °C for 50 min prior to UV detection at 325 nm.[206] A comparison of this method with the postcolumn OPA fluorescence method has been made.[207] A similar method using 1,2,4-triazole and Hg(II) for ampicillin, amoxicillin, and other antibiotics has been studied.[208–210] Two precolumn derivatization studies with UV detection were checked with real samples. Tobramycin plus the impurities neamine and kanamycin and also the degradation product nebramine were derivatized with 2,4-dinitrofluorobenzene for 20 min at 70 °C in 0.8-mM sulfuric acid. Separation of the derivatives on a C_{18} column with detection at 365 nm was carried out, and the stability of tobramycin in ophthalmic solutions was determined.[211]

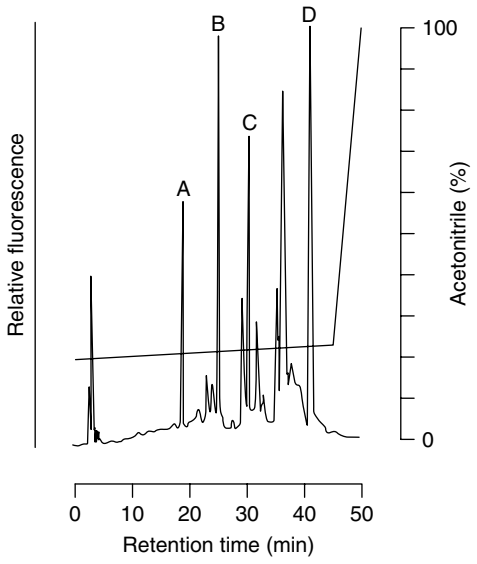

Figure 9 Reversed-phase HPLC of FMOC derivatives of cephalosporins (5 $\mu g\,mL^{-1}$) in a fermentation broth using a borate buffer–acetonitrile gradient mobile phase. (A) Deacetylcephalosporin C; (B) deacetoxycephalosporin C; (C) cephalosporin C; and (D) ampicillin. (Reproduced with permission from Ref. 204.)

3.4 Barbiturates

Precolumn methods primarily include reaction with various alkylating reagents. The highly reactive chlorine of *N*-chloromethylphthalimides permits derivatization of OH and NH functional groups to form UV-absorbing compounds.[212] Detection limits of 5 ng for phenobarbital have been reported.[213]

2-Naphthacyl bromide forms strongly absorbing derivatives of barbiturates at 254 nm with detection in plasma or serum below the therapeutic range.[214] An on-line solid-phase anion-exchange extraction provided enhanced chromatographic selectivity of barbiturates in urine.[215] Postcolumn UV detection of barbiturates can be conveniently enhanced by mixing with a pH 10 borate buffer.[216,217] A wavelength shift for maximum absorbance from 220 to 240 nm provides for detection limits of about 6 μg for butabarbital.

The UV detection of barbiturates can be significantly enhanced through postcolumn photochemical derivatization. Barbiturates have no significant absorbance above 230 nm and detection in the 200–220 nm range can be complicated by the presence of interfering peaks, particularly in biological samples such as serum. Using a 25 m × 0.25 mm i.d. Teflon-knitted open-tubular reactor mounted around a mercury lamp, which provided an irradiation time of about 190 s, excellent signal enhancement at 270 nm was possible for barbiturates (Figure 10a and b). It was determined by HPLC that dealkylation at the 5-position to give ethylbarbituric acid was the mechanism of the photochemical reaction.[218]

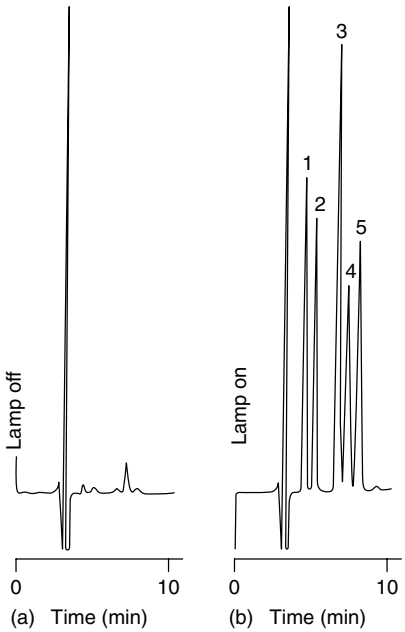

Figure 10 Chromatogram of a standard sample of barbiturates detected at 270 nm, (a) without and (b) with on-line photochemical reaction. (1) Aprobarbital; (2) butethal; (3) pentobarbital; (4) mephobarbital; and (5) secobarbital. (Reproduced with permission from Ref. 218.)

3.5 Carbonyl Compounds and Carboxylic Acids

First, a short summary of the methods involving the reduction of quinone compounds to form fluorescent hydroxy derivatives is given. Vitamin K_1 in human plasma was electrochemically reduced at −0.4 V and the fluorescent derivative detected at levels as low as 25–50 pg.[219] Comparison of this method with UV and electrochemical oxidation detection systems has been made.[220] Danthron (1,8-dihydroxyanthraquinone) has been reduced with dithionite and determined by either flow injection analysis in Modane tablets or HPLC in urine.[221] A comparison of UV and fluorescence chromatograms for danthron spiked in a urine sample showed the selectivity advantage of fluorescence (Figure 11a and b). Dansylhydrazine formed fluorescent derivatives of tetraphenylacetone isomers before separation.[222] Short-chain carboxylic acids were modified using *p*-bromophenacyl bromide (Table 5, reagent 4) to form esters that could be easily detected by UV absorbance.[223]

5-Bromomethylfluorescein was studied for the prederivatization of carboxylic acids for detection by either UV absorbance or fluorescence (standard and laser induced). Model analytes included prostaglandins (unsaturated carboxylic acids) and the drug cefuroxime, and also standard aliphatic and aromatic carboxylic acids. Dicarboxylic acids did not react with bromomethylfluorescein, possibly due to solubility problems in the organic reaction medium.[224] Fatty acids and prostaglandins were converted into *p*-hydroxanilides using *p*-aminophenol (Table 4, reagent 5). These hydroxy derivatives were then oxidized through electrochemical detection after reversed-phase HPLC.[225] A UV-absorbing naphthacyl ester of the prostaglandin carboprost has been formed before normal-phase HPLC.[226] Fluorescent derivatives of prostaglandins using 9-anthryldiazomethane provided detection limits of 100 pg after reversed-phase HPLC.[227] The prostaglandin arbaprostil was derivatized with panacyl bromide before column switching using fluorescent derivatization.[228,229] Prostaglandins have also been derivatized with *p*-(9-anthroyloxyl)phenacyl bromide.[230] Oxidation of prostaglandins to the corresponding 15-oxo derivatives using pyridinium dichromate permitted UV detection at 228 nm and picomole detection limits.[231,232] Prostaglandins have been labeled with the fluorescent reagent 9-anthyldiazomethane through the carboxyl group. This esterification reaction can be carried out under mild conditions such as 40 °C for 30 min. Five prostaglandin derivatives were separated by reversed-phase HPLC using fluorescent detection with

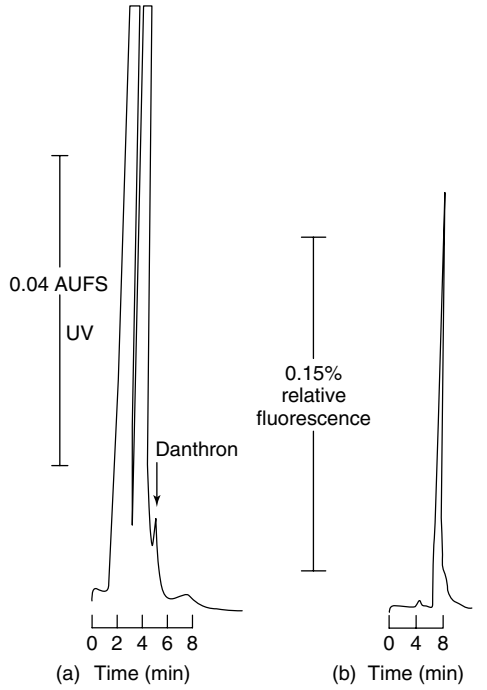

Figure 11 Separation of danthron spiked in urine (2 ng mL^{-1}). (a) UV detection (254 nm). Column, 250 × 4.6 mm C_{18}; mobile phase, 60 : 40 methanol–water at 1.2 mL min^{-1}. (b) Fluorescence detection (388-nm excitation, 510-nm emission). A solution of 90-mM dithionite in 2% borate buffer was pumped into the effluent at 0.4 mL min^{-1}, AUFS–Absorbance units full scale. (Reproduced with permission from Ref. 221.)

excitation at 365 nm and emission at 418 nm at the 8-ng level.[233]

Ascorbic acid from 5 to 800 mg L^{-1} was detected by CL after postcolumn reaction with lucigenin.[234] After derivatization with 1,2-phenylenediamine, both ascorbic acid and dehydroascorbic acid were separated by reversed-phase ion-pair HPLC.[235] The same two compounds were separated by reversed-phase HPLC, and the dehydroascorbic acid was reduced to ascorbic acid with dithiothreitol for UV detection at 267 nm.[236] UV detection of bis(dinitrophenyl)hydrazine derivatives of ascorbic and dehydroascorbic acid was possible at 497 nm.[237] Four forms of ascorbic acid, the previous two plus isoascorbic acid and its dehydro form, were detected postcolumn using benzamidine and fluorescence.[238] Total ascorbic acid was determined fluorimetrically after reaction with diaminodimethoxybenzene.[239]

A number of methods directed toward pain relievers and other compounds with carbonyl groups have been published. Indomethacin formed a fluorescent derivative after postcolumn alkaline hydrolysis, giving a detection limit of 5 pg.[240] Postcolumn alkaline hydrolysis has also been applied to aspirin in plasma in order to form the fluorescent salicylic acid.[241] A wide variety of acidic compounds including acetylsalicylate, ibuprofen, ketoprofen, naproxen, benzafibrate, diclofenac, and tolfenamic acid were determined by HPLC with postcolumn photochemical reaction and fluorescence at the 2–120 ng mL^{-1} level.[242] The enantiomeric composition of ibuprofen in human plasma has been resolved after prederivatization with (*S*)-(−)-1-(naphthenyl)ethylamine[243] or ethyl chloroformate–leucinamide.[244] Flunoxaprofen enantiomers have been separated after reaction with (*S*)-(−)-1-phenylethylamine.[245] Artesunic acid was determined after derivatization with *o*,*p*-nitrobenzyl-*N*,*N*′-diisopropylisourea by HPLC with UV detection.[246] Carnitine, betaine, and related compounds have been reacted to form either 4′-bromophenacyl esters[247] or *p*-nitrobenzyl oxines[248] to permit UV detection at 254 nm. A new acridinium sulfonylamide label permits the chemiluminescent detection of carboxylic acids such as the test compound ibuprofan. This alkylation reaction takes place in dry acetonitrile for 20 min at 50 °C and separation of the derivative is possible on a C_{18} column with an acetonitrile–water–tetrahydrofuran mobile phase with the ion-pairing agent tetrabutylammonium ion. CL detection of the acridinium label is possible by postcolumn addition of potassium hydroxide solution. A detection limit of 60 pg of derivatized ibuprofen was found.[249]

The formation of multiple molecular ions from various alkali metals in LC/MS can reduce reproducibility and make sensitive quantitation difficult. The addition of a primary alkylamine to the mobile phase can cause preferential formation of the alkylammonium adduct ion and suppression of other multiple molecular ions. Octylamine was particularly advantageous in improving the detectability of paclitaxel and docetaxel by about a factor of 4–5 down to about 0.5–1.0 ng mL^{-1}.[250] A similar enhancement was reported for a 3-benzothiepine compound TAK-778 using 1-hexylamine; interaction of the alkyl amine with the phosphoryl group was postulated.[251] Methylammonium acetate was effective in improving the MS sensitivity for simvastin and its hydroxy acid degradation product.[252] Twelve acidic pharmaceuticals and triclosan were determined by ion-pair lC using tri-*n*-butylamine and electrospray ionization-MS.[253] The ion-pairing agent increased the MS signal intensity, permitting detection limits in the 6–200 ng L^{-1} range.

3.6 Catecholamines

Precolumn fluorescent derivatization of catecholamines with OPA and a thiol has been well established (Table 4, reagent 1). Norepinephrine, dopamine, and normetanephrine have been measured at the low picogram level after reversed-phase HPLC.[254,255] An

electrochemical cell placed between the injector and the column was used for oxidation of adrenaline and levodopa to the corresponding quinones for detection in the visible region.[256] Resolution of norephedrine enantiomers after derivatization with either acetylglycosyl isothiocyanates[257] or 4-methyl-5-phenyl-2-oxazolidone[258] has been reported. Fluorescent catecholamine derivatives have been generated using 1,2-diphenylethylenediamine, providing detection limits in the femtomole range.[259]

Postcolumn fluorescent reaction of norepinephrine and epinephrine with trihydroxyindole provided 20–30 pg detection limits.[260] This method was compared with the OPA–thiol reaction in the postcolumn mode.[261,262] The same two catecholamines can be converted into fluorescent products by heating in an alkaline borate buffer after ion-exchange chromatography.[263] Dopamine isomers in serum and urine have been separated before hydrolysis and reaction with *p*-aminobenzoic acid to form fluorescent products.[264] Glycylglycine as an alternative postcolumn reagent to glycinamide increased the rate of formation of fluorescent derivatives after ion-exchange HPLC.[265,266] Catecholamines catalyze the reaction between formaldehyde and *o*-dintrobenzene and permit their detection at 560 nm.[267]

3.7 Hydroxy Compounds

Most precolumn methods focus on enhancing UV detection. Isotachysterol derivatives of vitamin D improved detection at 290 nm after normal-phase HPLC.[268] On-column periodate oxidation of ephedrine sulfate, a nasal decongestant, to form benzyl alcohol has been carried out in a seven-laboratory study.[269] Indirect photometric detection using *n*-heptyl *p*-aminobenzoate in the mobile phase and a C_{18} column has been applied to menthol.[270] Enantiomers of propranolol and 4-hydroxypropranolol were formed upon derivatization with (+)-1-phenylethyl isocyanate[271] or (+)-tetraacetyl β-D-glucopyranosyl isothiocyanate[272,273] and separated by reversed-phase HPLC. Enantiomers of derivatized or underivatized propranolol have been separated.[274] Diastereomeric derivatization of 1-methyl-3-pyrolidinol[275] and proxyphylline[276] for UV detection has been accomplished. Qinghaosu, an antimalaria component in a Chinese herb, can be converted into a UV-absorbing compound using a sodium hydroxide solution before HPLC separation.[277] A derivative of qinghaosu has been esterified with diacetyldihydrofluorescein before HPLC and UV detection.[278] Cholic acid and its derivatives have been labeled with 1-anthroylnitrile before HPLC and fluorescent detection at the femtomole level.[279] Trospium has been converted into the corresponding spiro alcohol before fluorescent derivatization with benoxaprofen chloride and reversed-phase HPLC.[280]

Derivatization can be an effective method to facilitate the chiral separation of pharmaceuticals. Nadolol diastereomers were derivatized with (*R*)-(−)-1-(1-naphthyl)ethyl isothiocyanate to chiral urea derivatives through the secondary amine group by reaction for 5 min at 45 °C. Separation of the RS, SR, RR, and SS diastereomers was straightforward on a C_{18} column using a 60 : 40 water–acetonitrile mobile phase (Figure 12a–c). Fluorescence detection with excitation at 285 nm and emission at 340 nm provided selectivity for the determination of nadolol in plasma samples. The limit of detection was 2.5 ng mL^{-1}, representing 50 pg injected. This chiral derivatization method has been used for the determination of enantiomers of other β-blocker drugs.[281]

A few of the postcolumn approaches are outlined below. Using a photochemical reactor, diethylstilbestrol (DES) has been converted into a fluorescent derivative and determined at the low parts-per-billion level in biological matrixes.[282,283] Anabolic stilbenes, such as DES, have been measured in urine by HPLC and an off-line CL immunochemical assay.[284] The antihypertensive agent fenoldopam was formed using a postcolumn enzyme reactor from the corresponding glucuronide, and detected electrochemically.[285] Cyclodextrins in biological fluids were assayed by negative colorimetric detection after postcolumn complexation with phenolphthalein.[286] Vitamin B_6 (pyridoxine) has been postcolumn derivatized with 2,6-dibromoquinone-4-chlorimide to form a colored product with a maximum absorbance at 650 nm.[287]

3.8 Steroids

Esterified estrogens were converted into their free phenolic forms by acid hydrolysis, and separation could be achieved by reversed-phase HPLC.[288] The chromatographic behavior of estrogen carbonyls, such as equilenin, equilin, and estrone, was improved upon reduction to the 17-α-hydroxy compounds using sodium borohydride.[289] Prederivatization of conjugated estrogens with dansyl chloride permitted fluorescent detection after separation by normal-phase HPLC (Table 4).[290] This method was modified by having the dansylation prederivatization reaction with the estrogens occur on the SPE column before determination by HPLC with MS/MS detection at the parts-per-trillion level.[291] Ketosteroids, such as androsterone, dehydroepiandrosterone, epiandosterone, and etiocholanolone, were derivatized with *p*-nitrophenylhydrazine (Table 4, reaction 6) and detected electrochemically at levels as low as 200 pg.[292]

Isonicotinoyl hydrazine was used to tag steroids, such as corticosterone, to permit fluorescent detection after normal-phase HPLC.[293] Further work using this

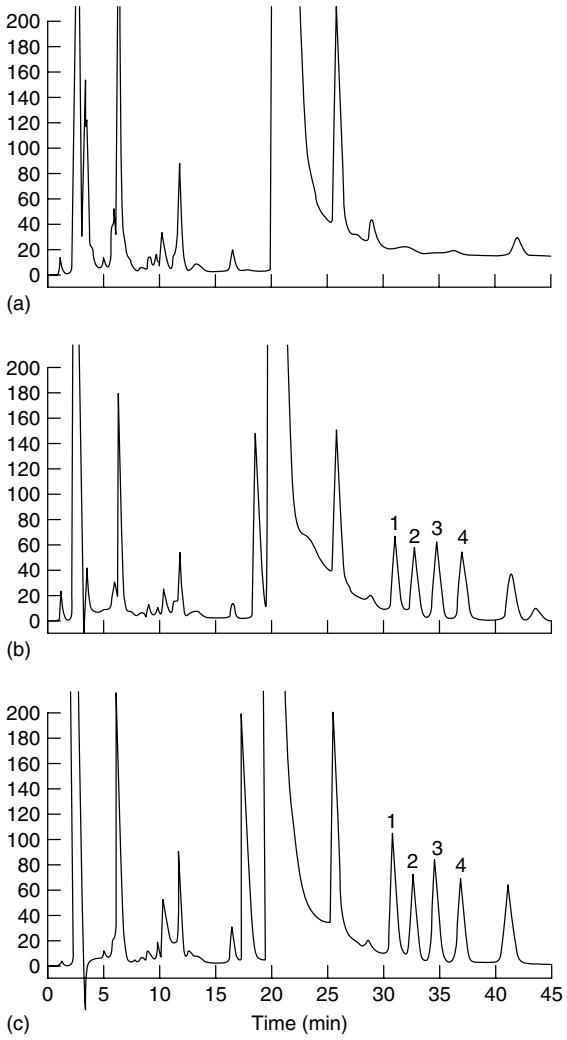

Figure 12 Typical chromatograms of (a) blank control dog plasma, (b) plasma spiked with 50 ng mL^{-1} of each diastereomer, and (c) plasma obtained 2 h after oral administration of 1 mg kg^{-1} of racemic nadolol. Peaks: 1 = (*SR*)-nadolol; 2 = (*RS*)-nadolol; 3 = (*RR*)-nadolol; and 4 = (*S*, *S*)-nadolol. (Reproduced with permission from Ref. 281.)

method with reversed-phase HPLC and optimized reaction conditions permitted detection limits as low as 7–10 ng.[294,295] 17-Oxosteroids were labeled with dansylhydrazine and then chromatographed on a silica column with subsequent fluorescent detection from 60 to 100 pg.[296] A similar method for this class of compounds using 3-chloroformyl-7-methoxycoumarin and reversed-phase HPLC has been published.[297] Hydroxysteroids were derivatized with anthroylnitrile, showing the feasibility of the reaction for fluorescent detection after HPLC.[298]

Six corticosteroids were separated within 25 min and reacted with a lucigenin–KOH solution for CL detection.[299] The reactive site is an α-hydroxycarbonyl group, and detection limits of 2 pg were comparable to those with precolumn fluorescent methods. Corticosteroids were also derivatized postcolumn with glycinamide in the presence of hexacyanoferrate(III) before fluorimetric detection at the 5-ng level.[300] Digoxin, a widely used cardiac glycoside, has been detected fluorimetrically by postcolumn reaction with a solution of ascorbic acid, peroxide, and HCl.[301,302] A detection limit of less than 1 ng was attained.

Digoxin and its metabolites, such as digoxigenin, were derivatized with 1-naphthoyl chloride before separation and fluorescent detection,[303] and 5 ng amounts could be determined in urine or feces. A similar method using 4-nitrobenzoyl chloride has been reported.[304] Dihydrodigoxin, a major metabolite of digoxin, has been separated from digoxin with dual UV and postcolumn fluorescent derivatization detection.[305] The photoreduction of anthroquinone-2,6-disulfonate to the fluorescent 9,10-dihydroxyanthracene-2,6-disulfonate occurs only in the presence of hydrogen atom-donating substrates, such as alcohols, aldehydes, amines, ethers, and saccharides. Using a knitted Teflon reactor, cardiac glycosides, such as digoxin, digoxigenin, and diginatin, have been determined in the range 50–500 ng.[306]

3.9 Sulfur Compounds

Most methods employ a precolumn reaction to form fluorescent or UV derivatives. Dansylaziridine[307] was used to determine cysteine and other thiols. Biological thiols, such as glutathione and ergothioneine, were reacted with monobromobimane (Table 4, reagent 4) before ion-exchange chromatography.[308] This method was also adopted for the determination of dithiols, such as 2,3-dimercaptopropane-1-sulfonic acid, in urine at levels as low as 10 pmol.[309] Fluorescent derivatization of 2-mercaptopropionylglycine using *N*-(7-dimethylamino-4-methyl-3-coumarinyl)maleimide could probably be extended to other thiol compounds.[310] Resolution of the optical isomers of diltiazem was accomplished using UV detection after derivatization with optically pure 2-naphthylsulfonyl-2-pyrollidinecarbonyl chloride.[311] Three N-substituted maleimides were compared for precolumn derivatization of thiols such as penicillamine before electrochemical detection at the picogram level.[312] Ethacrynic acid (Table 5, reagent 5) forms UV-detectable derivatives of thiols, such as captopril, *N*-acetyl-L-cysteine, and mercaptopropionylglycine with detection limits as low as 0.5 μg mL^{-1}.[313]

Postcolumn derivatization of thioethers, such as ampicillin and ranitidine, was accomplished using on-line generated bromine and electrochemical detection of the excess bromine.[314] An analogous method has been reported for phenothiazines, such as thioridazine;

however, the resultant products from the oxidation reaction with bromine are detected fluorimetrically.[315] Photochemical activation of phenothiazines and demoxepam in 2 min and subsequent fluorescence detection provided detection limits a factor of 10 better than UV detection.[316] Applications of photochemical reactors to a variety of pharmaceuticals, such as phenothiazines, have been summarized.[317,318] A precolumn derivatization method for phenothiazine involves desulfurization with Raney nickel to produce diphenylamine, which is electrochemically active.[319] A detection limit of 10 pg was found. The reagent pyrenemaleimide provided derivatization of *N*-acetylcysteine with a detection limit of 10 pmol.[320] Postcolumn complexation of disulfiram and two of its metabolites using Cu^{2+} allowed colorimetric detection at 435 nm with detection limits in the low parts-per-billion range.[321] Quenched peroxalate CL has been employed using immobilized reagents in a postcolumn reactor for thioridazin, sulforidazine, and methimazole.[322]

On-line precolumn derivatization to improve reproducibility has been demonstrated for the determination of busulfan in human serum. A schematic diagram of the HPLC system is shown in Figure 13; all three columns C1, C2, and C3 were C_{18} type with different lengths and/or diameters. After extraction of busulfan from serum, the residue reconstituted in water was injected

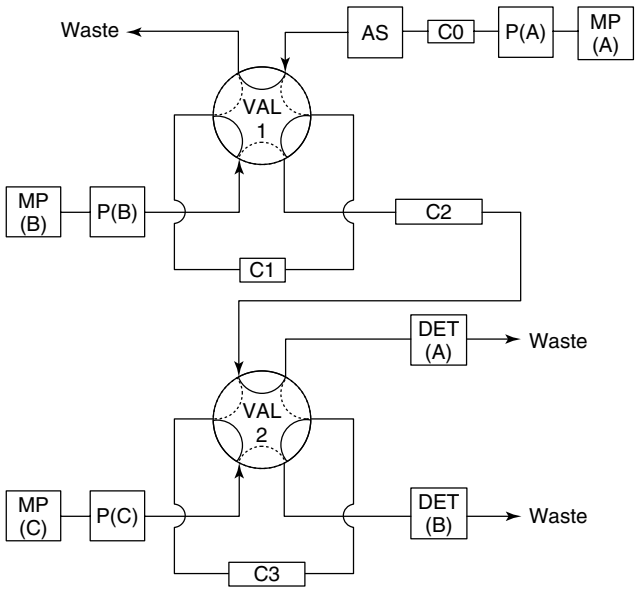

Figure 13 Schematic diagram of the HPLC system. P(A), P(B), and P(C) = pumps; AS = autosampler; VAL 1 and VAL 2 = six–port valves 1 and 2; C0 = clean-up column; Cl, C2, and C3 = columns 1, 2, and 3; DET(A) and DET(B) = UV detectors A and B; MP (A), MP (B), and MP (C) = mobile phases A, B, and C. The solid and dotted lines in the six-port valves indicate valve positions 0 and 1 respectively. (Reproduced with permission from Ref. 323.)

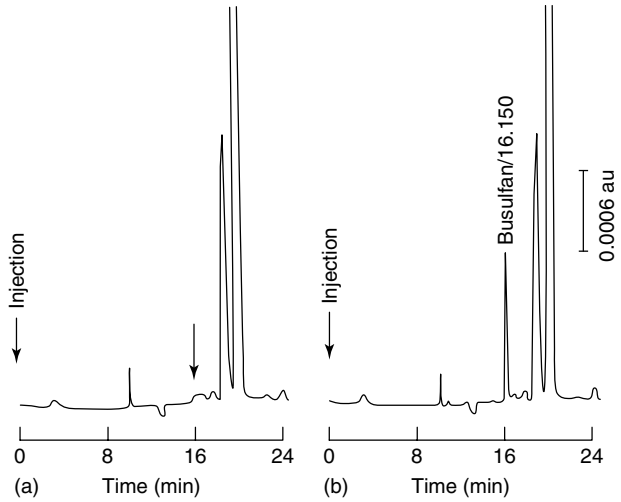

Figure 14 Typical chromatograms of (a) drug-free serum and (b) serum spiked with busulfan (100 ng mL^{-1}) obtained with columns Cl, C2, and C3. The arrow indicates the retention time of the busulfan derivative. (Reproduced with permission from Ref. 323.)

on to C1, where it was derivatized with diethyldithiocarbamate (DCC) in mobile phase A for 5 min. Then the busulfan–DCC derivative from C1 was backflushed on to C2, where it was separated using mobile phase B. Because the background interference from mostly excess DCC overlapped the busulfan–DCC derivative peak completely, valve 2 was used to take a heart cut of this peak of interest and inject it on to C3. The resulting separation using mobile phase C is shown in Figure 14(a) and (b). Detection was at 278 nm and the lower limit of quantitation in serum was 10 ng mL^{-1}. The total time for the derivatization and separation was 33 min.[323] Alternatively, for therapeutic drug monitoring in plasma down to 10 ng mL^{-1}, the determination of busulfan was based on conversion to 1,4-diiodobutane prior to HPLC separation and then UV detection of the iodide formed upon postcolumn photolysis.[324]

4 CAPILLARY ELECTROPHORESIS

CE has developed into a versatile separation technique well suited for the determination of pharmaceutical and biomedical samples. The major advantages of CE over chromatographic separation techniques are its simplicity and efficiency. However, CE suffers from poor concentration sensitivity in detection owing to the extremely small sample volume involved. One way to improve the detection of charged analytes with weak chromophores is to use indirect detection in which an easily detected agent of the same charge is added to the running electrolyte. Replacement of the

visualizing agent by the analytes generates negative peaks as the readout. For example, the antiepiletic drug vigabatrin has poor UV absorptivity but can be determined in dosage forms from 5 to 150 mg L^{-1} using benzyltriethylammonium hydroxide as the visualizing agent.[325] To permit application of a negative voltage, the presence of cetyl trimethylammoniumbromide was required in the running electrolyte but the analyte of interest, valproic acid, could still be visualized as the anion with benzoate at 210 nm over the 1–100 mg L^{-1} range.[326] A similar approach was used to determine gabapentin using 5-sulfosalicylate as the visualizing agent.[327] CE separation and indirect laser-induced fluorescence (LIF) detection of diuretics ethacrynic acid, chlorthalidone, bendroflumethiazide, and bumetanide using fluorescein as the visualizing agent was possible in 3 min with a detection limit of 0.2 mg L^{-1}.[328] A more standard approach to improve the sensitivity of CE detection is to derivatize the analytes with more favorable detection characteristics by adding either an ultraviolet/visible (UV/vis) chromophore or a fluorophore.

Derivatization chemistry previously developed for HPLC is often applicable to CE.[329] At least one comprehensive review specifically focused on the derivatization in CE has appeared.[330] Several other review articles dealt with more specific topics, such as postcolumn luminescence derivatization,[331] CL derivatization[332] diastereomer derivatization,[333] and dyes used to derivatize molecules for CE with LIF detection in drug analysis.[334] The pros and cons of the various modes of derivatization and a detailed comparison of this topic between LC and CE can be found in the literature.[330] Instead of listing the derivatization reagents suited for labeling amino, aldehyde, keto, carboxyl, hydroxyl, and sulfhydryl groups, we intend to focus on the purpose of derivatization in CE from a practical point of view, i.e. enhancing the detectability. Since UV/vis and LIF are the most common detection approaches used in CE for pharmaceutical analysis, the most recent work related to these methods is primarily discussed below. Sections on special applications and derivatization instrumental modes are also included.

4.1 Derivatization for Ultraviolet Detection

Although most organic compounds have UV/vis chromophores and CE detection is possible at 200 nm, derivatization is still important to improve selectivity of compound detection in a complex sample matrix as well as concentration detection limits. Very often, UV/vis detection of the chromophores does not give a satisfactory response owing to the very short light pathlength (50–100 μm) in CE. For example, the anticancer drug prospidin is a piperazinium derivative with chloroxypropyl groups but no UV/vis chromophore. Derivatization with diethyldithioccarbamate at 37 °C for 90 min in a basic solution, with loss of HCl, generates a product that absorbs light at 254 nm. Separation of the desired product from excess derivatizing agent is possible in 10 min by CE with a detection limit of 1 ng L^{-1}.[335]

Rimantadine is a synthetic analog of amantadine. Both are antiviral agents used for prophylaxis and treatment of influenza A. Because rimantadine is almost transparent in the UV/vis range, either indirect detection or derivatization has to be used to identify this compound. The indirect detection method used 5-mM 4-methylbenzylamine in 1 : 4 methanol–water as the absorbing background electrolyte for detection at 210 nm. The derivatization method used rimantadine to react with 1,2-naphthoquinone-4-sulfonic acid in an alkaline medium. CE determination of the derivative at 280 nm was performed in an uncoated capillary (44 cm × 75 − μm i.d.) using a 40-mM tetraborate buffer at pH 9.2. The detection limits were 0.1 and 2 mg L^{-1} for indirect detection and derivatization methods respectively. The methods were used to determine rimantadine in pharmaceutical products and for dissolution testing of Flumadin tablets.[336]

Most of the naturally occurring amino acids do not have proper UV/vis chromophores for CE analysis. Derivatization is a necessity for the determination of amino acids at reasonable concentrations. By using dansyl chloride to derivatize the amino acids, the enantiomeric forms of novel depsipeptide antitumor antibiotic BMY-45012, and its analogs, were determined by CE with UV/vis detection.[337] The compounds were subjected to total hydrolysis in a vacuum hydrolysis tube with 6-M HCl at 110 °C for 24 h. The hydrolyzed residue (about 7.8 mg) was dissolved in a water–acetonitrile solvent. For subsequent derivatization, the solution was further diluted to about 1.7 mg mL^{-1} with water. Dansyl chloride dissolved in acetonitrile (3.0 mg mL^{-1}) was utilized for the derivatization of standard native amino acids and those present in the hydrolyzate. A 50-μL sample solution was mixed with 50 μL of dansyl chloride solution and an aliquot of borate buffer at pH 9.08. The mixture was held at room temperature for 2 h and used directly for injection. The differing interaction of dansyl amino acids with amino-β-cyclodextrin copper complexes in the CE running electrolyte was the basis of the chiral separation of selected dansyl amino acids.[338] The reagent phanquinone (4,7-phenanthroline-5,6-dione) was applied as a prederivatizing agent for amino acids down to the sub micromolar level followed by micellar electrokinetic chromatography (MEKC) with detection at 240 nm.[339]

Examples of other derivatization approaches for amine-containing drugs follow. Ofloxacin acyl chloride was used to modify a wide variety of drugs such as amantadine, transexamic acid, and two amino acids

(glycine and methionine) into derivatives that could be detected at 300 nm.[340] The same derivatizing agent has been applied for the determination of gabapentin and vigabatrin at the 5-μM level.[341] The OPA-3-mercaptopropionic acid derivative of tobramycin could be detected at 230 nm down to 0.1 mg L^{-1}.[342] Chiral separation using β-cyclodextrin of aspartic acid enantiomers derivatized on-line by flow injection with OPA and mercaptoethanol was possible using MEKC.[343] Resolution of five β-blocker drugs, propranolol, oxprenolol, pindolo, metoprolol, and atenolol, after prederivatization with (+)-1-(9-fluorenyl)ethylchloroformate by capillary electrochromatography with a C_{18} column was directed toward determination of their enantiomeric composition.[344]

Relative binding of glycopeptides teicoplanin and ristocetin with alanine terminus peptides before and after reaction with acetic or succinic anhydride was an example of affinity CE.[345] All binding reactions were done on-line in the capillary with the glycopeptides reacting first with the anhydride and then with varying amounts of the alanine terminus peptides. It was ascertained how the modification of drugs can alter their affinities for target molecules.

The aminocychitol antibiotic amikacin can be derivatized with 1-methoxycarbonylindolizine-3,5-dicarbaldehyde at room temperature for 15 min before determination in plasma by MEKC with UV detection at 280 nm and standard fluorescence detection (excitation at 414 nm, emission at 482 nm).[346] However, there was only a twofold gain in sensitivity on switching from UV to fluorescence detection of the amikacin derivative. If laser excitation can be used (as discussed in the next section), the gain in sensitivity should be significantly higher.

4.2 Derivatization for Fluorescence Detection

The principles of fluorescence and the prerequisites for a good fluorophore, including the potential of using diode lasers in combination with a labeling procedure, have been reviewed previously for CE.[330] LIF detection is now commercially available for CE, and more industrial scientists are using LIF detection to detect trace amounts of analytes in complex matrices. LIF detection has several advantages. First, it offers excellent sensitivity;: under optimized conditions, a level of 10^{-10} M is feasible.[347] This is critical to pharmaceutical applications because sensitivity is often the most important parameter to consider when analyzing biological samples. Second, LIF offers a certain selectivity because, unlike UV absorbance, LIF selectively detects compounds that are fluorescent at certain wavelengths, and most organic compounds do not fluoresce. On the other hand, the fact that most organic compounds do not fluoresce is also a major limitation of LIF detection. Since the available excitation wavelengths of various lasers are limited, usually in the visible range, derivatization is often required when using CE with LIF detection. The following examples demonstrate the applicability of derivatization in CE with LIF detection for pharmaceutical analysis.

Sulfur compounds can be converted to fluorescent derivatives through several different reactions. Captopril, an antihypertensive agent, was derivatized through the thiol group with a dicarbocyamine label to give a fluorescent derivative with a wavelength maximum at 675 nm.[348] This wavelength was compatible with a semiconductor laser at 670 nm, which was used for LIF detection in conjunction with CE. Using a methanol-borate running electrolyte, CE also separated other labeled thiols such as cysteine and reduced glutathione from derivatized captopril (Figure 15). The detection limit of 2.5×10^{-8} M for captopril is limited by dilution due to the derivatization reaction. Thiouracil and phenyluracil were derivatized with 5-iodofluorescein before CE separation and LIF detection in the attomole range.[349] Determination of *S*-adenosyl-L-homocysteine and *S*-adenosyl-L-methionine

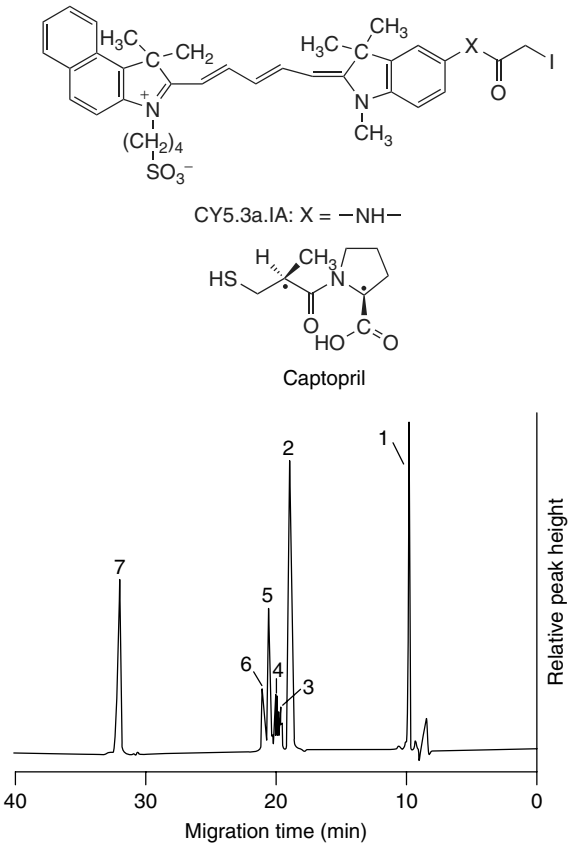

Figure 15 Electropherogram of a number of thiols labeled with CY5.3a.IA: (1) unreacted label; (2) captopril; (3) DL-homocysteine; (4) L-cysteine; (5) 3-mercaptopropionic acid; (6) 2-mercaptoacetic acid; and (7) reduced glutathione. (Reproduced with permission from Ref. 334.)

after chloroacetaldehyde derivatization was possible by capillary electrophoresis/laser induced fluorescence (CE/LIF) in less than 1 min.[350]

Drugs with active oxygen groups such as profens, including pranoprofen, fenoprofen, flurbiprofen, ketoprofen, and ibuprofen, can be derivatized using a water-soluble benzofurazan fluorescent reagent at room temperature using a triphenyl phosphine, diphenyl sulfide catalyst.[351] β-Cyclodextrin was added to the running electrolyte to facilitate the separation in a 50 cm $\times$ 50 μM i.d. capillary. With excitation at 488 nm and emission at 50 nm, CE/LIF detection limits of 0.16–0.3 fmol were attained.

Amines can easily be derivatized with a variety of reagents that generally have optimum excitation wavelengths in the 450–500 nm region, providing compatibility with the commercial 488-nm LIF laser. Huperzine was derivatized using rhodamine B isothiocyanate, and a CE/LIF detection limit of 4 nM was found.[352] Amino acids such as glycine, lysine, and glutamic acid were derivatized on-capillary with 3-(2-furoyl)quinoline-2-caboxaldehyde and detected by LIF at the 20–50-nM level.[353] Using native fluorescence, alkaloids such as normorphine, morphine, 6-acetyl morphine, and codeine could only be detected at the 200 ng mL^{-1} level after CE separation.[353] The same alkaloid mixture after N-demethylation with 1-chloroethyl chloroformate and then derivatization with FITC could be determined by CE/LIF at the 50–100 pg mL^{-1} level. The determination of FITC-derivatized amphetamine, methamphetamine, 3,4-methylenedioxymethamphetamine, and β-phenylethylamine in human urine by CE with or without dodecyl sulfate micelles using chip-based and fused-silica capillary instrumentation with LIF detection has been reported.[354] The results obtained via direct labeling of fortified urine were compared with those generated after FITC labeling of urinary extracts that were prepared by SPE. Using 5 mL of urine with a ''spiked amine'' to FITC ratio of 1 : 250, the SPE extract had a sensitivity of 200 ng mL^{-1} urine. That value is relevant for toxicology drug screening and confirmation. In contrast, with direct labeling of 10 μL of urine that had been alkalinized and diluted for derivatization, the limit of identification was 10 μg mL^{-1}, a value that is too high for practical purposes. Compared to fused-silica capillaries, electrophoresis in microstructures is shown to provide faster separation and higher efficiencies without loss of accuracy and precision.

4.3 Special Applications

4.3.1 Chiral Confirmation

With proper derivatization, CE has been used for the chiral separation of enantiomeric forms of derivatized amino acids. Liu et al. reported the determination of enantiomeric forms of amino acids derived from the novel depsipeptide antitumor antibiotic BMY-45012, and its analogs, the proposed structures shown in Scheme 1.[337] Amino acids were analyzed by complete hydrolysis and the hydrolyzate was derivatized with either dansyl chloride for UV absorbance detection or FITC for LIF detection in CE. For fluorescence detection, the fluorogenic reagent FITC was dissolved in acetone (0.01 M) as a stock solution. Amino acids were derivatized with FITC-derivatizing solution (5.0×10^{-4} M) under basic conditions (borate buffer, pH 9.08). The reaction

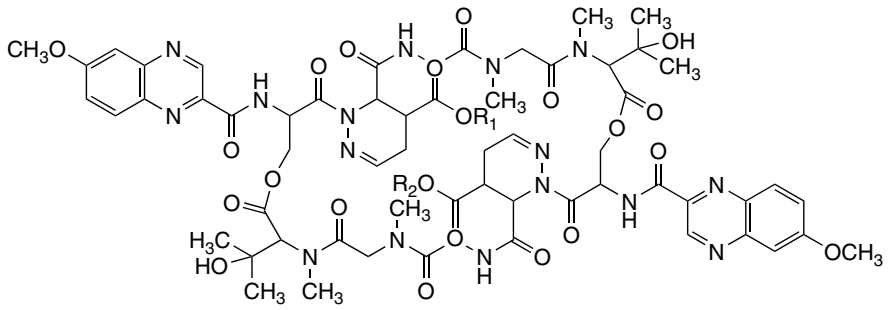

(**1**)

Compound	R_1	R_2
BMY45012	$CH(CH_3)_2$	$CH(CH_3)_2$
Analog-G439B	CH_2CH_3	CH_2CH_3
Analog-G451	$C(CH_3)_3$	$C(CH_3)_3$
Analog-G435	$C(CH_3)_3$	$CH(CH_3)_2$
Analog-G439A	$CH(CH_3)_2$	CH_3

Scheme 1. Structure of antibiotic BMY-45012 and its analogs. (Reproduced with permission from Liu et al. (337).)

was allowed to proceed for 2–4 h in the dark at room temperature and then stored at −20 °C prior to use. Both a metal chelate chiral CE method and a cyclodextrin-mediated host–guest interaction approach in MEKC with LIF detection confirmed the presence of several chiral amino acids, such as D-serine and L-β-hydroxyl-*N*-methylvaline, and the nonchiral amino acid sarcosine in the proposed structure. These methodologies provide a quick and sensitive approach for the determination of amino acid racemization in pharmaceutical natural products and have proved to be useful for structural elucidation refinement.

4.3.2 Oligosaccharide Determination

The derivatization of the pseudo-oligosaccharide acarbose (**2**) and its main metabolite **3** with 7-aminonaphthalene-1,3-disulfonic acid (ANDS) (Scheme 2) in human urine allowed the on-column LIF detection of the pseudo-oligosaccharides in human urine in the nanomolar range.[355] Before derivatization, 0.5-mL samples of urine or spiked urine were evaporated to dryness. A 100 µL volume of 0.08-M ANDS solution in acetic acid–water (3 : 17 v/v) and a 50 µL volume of 0.9-M $NaCNBH_3$ solution in dimethyl sulfoxide (DMSO) were added to each sample residue. The reagent solutions were freshly prepared before derivatization. The reaction tubes were vortex mixed and then incubated overnight at 40 °C. The efficient separation of these derivatives by CE using 100-mM triethylammonium phosphate buffer at pH 1.5 allowed the quantitation of acarbose and **3** in human urine after application of 300 mg of acarbose.

Assessment of the carbohydrate chains in glycoproteins is important for the quality control of such therapeutic antibodies.[356] After release of the four major carbohydrates by glycoamidase, derivatives were formed with 3-aminobenzoic acid and then separated by CE with LIF detection. Good precision in migration times and relative peak areas was found.

Both on-line and off-line capillary electrophoresis/mass spectrometry (CE/MS) trends for the determination of carbohydrates have been reviewed, showing the on-line sheath liquid interface with electrospray ionization (ESI) and ion trap MS detection is the most common instrumental mode.[357] Applications include oligosaccharides from pathogens, carbohydrate-based drugs and metabolites, and characterization of glycans from glycoproteins. It was noted that prior derivatization of neutral carbohydrates to make them charged would likely improve both the CE separation and sensitivity of MS detection.

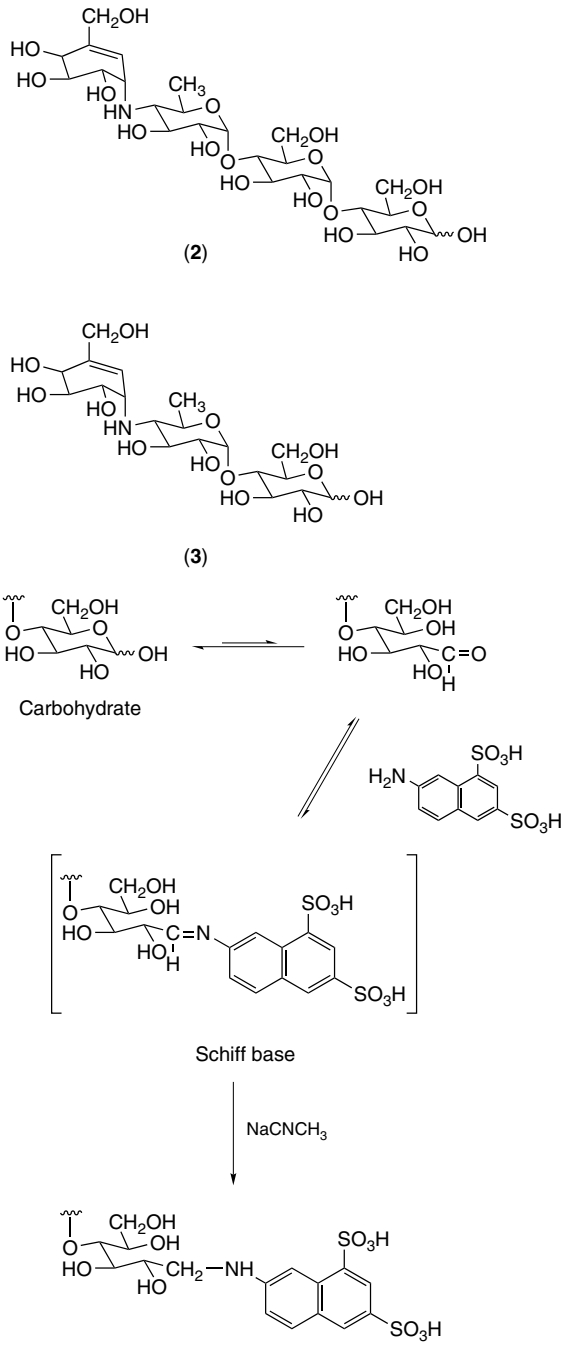

Scheme 2. Reaction of **2** and **3** for derivatization with ANDS by reductive amination. (Reproduced with permission from Rethfeld and Blaschke.(355).)

4.3.3 Derivatization of Protein Samples

With the increased interest in the therapeutic use of the recombinant monoclonal antibody (rMAb) technique, a generic analytical approach for the analysis of size-based rMAb variants is desired. CE/LIF with proper derivatization has been shown to be promising as a

general method. The rMAb can be derivatized with a neutral fluorophore, e.g. 5-carboxytetramethylrhodamine succinimidyl ester (5-TAMRSE) for CE/LIF analysis.[358] Samples containing 2.5 mg of rMAb were buffer exchanged into 800 μL of 0.1-M sodium hydrogencarbonate (pH 8.3) using a NAP-5 column. A 10 μL volume of 5-TAMRSE (1.4 mg mL^{-1}) dissolved in DMSO was then added to 190 μL of rMAb solution and the resultant mixture incubated for 2 h at 30 °C. After incubation, 190 μL of the antibody–dye conjugate was loaded on to a second NAP-5 column and collected in 700 μL of 0.1-M sodium hydrogencarbonate (pH 8.3). The labeled sample, after incubating with SDS, can be separated by CE using a hydrophilic polymer as a sieving matrix. Using these precolumn labeling conditions, the detection of rMAb at a low-nanomolar concentration (9 ng mL^{-1}) is obtained with no apparent decrease in resolution or changes to the distribution of rMAb analyte species in comparison with an unlabeled sample. This assay can be used for monitoring the consistency of the bulk manufacture of a protein pharmaceutical and in providing a size-based separation of product-related variants and also nonproduct impurities. The CE/LIF with derivatization demonstrated comparable resolution and sensitivity to silver-stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) but offered the advantages of enhanced precision and robustness, speed, ease of use, and on-line detection.

CE was used for the study of palmitoyl derivatization of interferon-α_{2b} (p-IFN-α).[359] The derivative was prepared by covalent attachment of the fatty acid to lysine residues in the protein through a reaction with *N*-hydroxysuccinimide palmitate (NHSP) ester (Scheme 3). The first step involved the preparation of an intermediate NHSP, which in turn was reacted with IFN-α. CE was able to study the effect of reaction time and reagent/protein ratio.

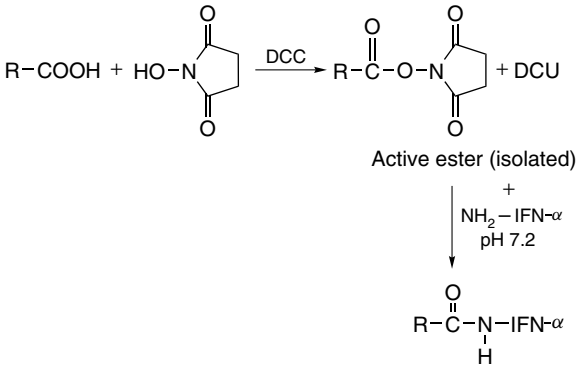

Scheme 3. Synthetic process for the fatty acrylation of IFN-α. R–COOH palmitic acid. DCU-N,N′-dicyclohexylurea DCC-N,N′-dicyclohexylcarbodiimide (Reproduced with permission from Foldvari et al. (359).)

4.4 Derivatization Modes

In general, derivatization in CE can be accomplished by either pre- or postcolumn derivatization. The precolumn derivatization is similar to methods used for derivatization in HPLC. Since CE has a much superior separation power, the interference from the derivatization by-products may be less of a problem than in LC. However, CE is very sensitive to the ionic strength of the sample. Additional sample preparation may be needed if the mixture in the derivatization reaction has a high ionic strength.

Postcolumn detection strategies have also been developed for CE. An overview of the advantages and limitations of postcolumn derivatization for CE can be found.[330] The details about the instrumental developments and applications of postcolumn derivatization in CE can also be found in the literature.[331] Again, much of the same derivatization chemistry developed for HPLC is applicable to CE. One exception is postcolumn photolysis, perhaps due to the difficulty of modifying commercial instruments. Using a Teflon-coated UV transparent capillary and a Pen-Ray Hg UV lamp, an on-column photoreactor has been developed for CE and applied for the determination of phytosiderophores (iron containing species) and the nonprotein amino acid nicotianamine.[360] The relationship of irradiation window length and sensitivity was investigated, and both UV detection and contactless conductivity detection were used.

Various systems to merge the reagent solution with the separation medium have been developed, including coaxial capillary reactors, gap reactors, and free- solution or end-column systems. For all reactor types, the geometry of the system, and the method used to propel the reaction mixture (by pressure or by voltage) appear to be critical to preserve the separation efficiency. To minimize peak broadening, careful designing is necessary in terms of connecting a reagent capillary to form a tee. Because of the small (50–75 μm) i.d. of the capillaries used, normal mixing through a diffusion process is sufficient. The most frequently applied reactors are (i) coaxial, (ii) gap, (iii) free-solution, and (iv) sheath flow. With proper design, plate numbers of over 100 000 could be realized. The strict requirements on the reaction rate in postcolumn derivatization in CE limit the number of different reagents that have been used. For LIF detection, mainly OPA and its naphthalene analog have been used. With careful instrument design, the detection limits of postcolumn derivatization can be comparable to those of precolumn derivatization.

An example of postcolumn CL detection for CE is shown in Figure 16.[361] The CL reagent tris(bipyridyl) ruthenium(II) is added at 10 μL min^{-1} using a syringe pump through a capillary into a reaction tee where it is

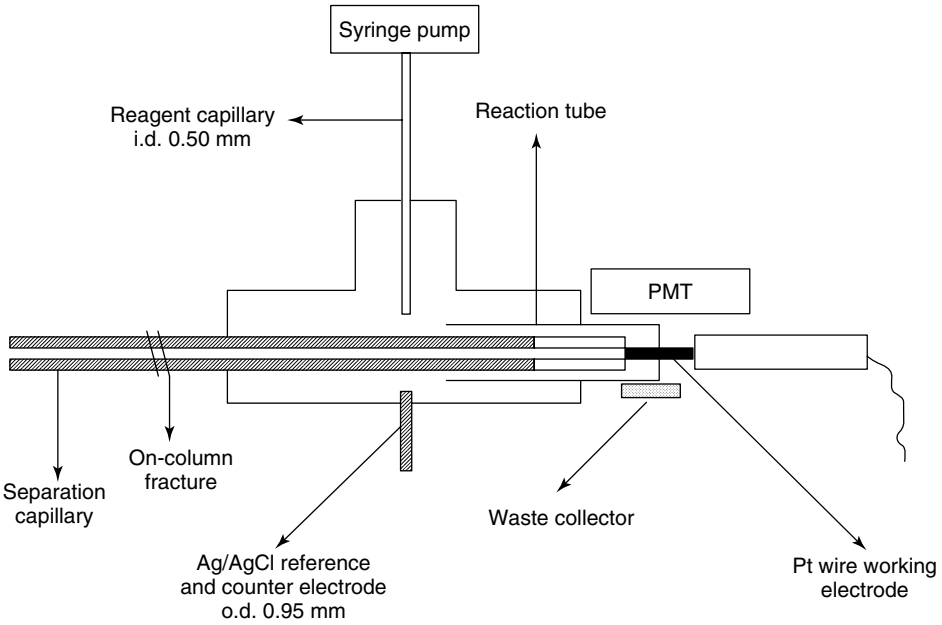

Figure 16 Schematic representation of a cell for CE with postcapillary $Ru(bpy)_3^{3+}$ chemiluminescence detection. (Reproduced with permission from Ref. 361.)

oxidized to the +3 state at a Pt working electrode. The CE electric field has been decoupled from the potential applied to the Pt electrode by a CE capillary fracture that is covered by a Nafion tube. This $Ru(bpy)_3^{3+}$ reactant can undergo CL with CE-separated compounds such as aliphatic amines through detection of the visible light produced at the end of the reaction tube in front of the photomultiplier (PMT) detector. Amino acids proline, valine, and phenylalanine were separated with theoretical plate numbers of 74–84 000 and detection limits in the sub 1-μM range.

5 CONCLUSION

It can be argued that the advances made in MS detection for GC, HPLC, and CE have diminished the importance of chemical derivatization for these techniques. However, chemical derivatization of drugs is critical for GC because these samples, which often contain multiple polar substituents, are simply not volatile or thermally stable. MS detection for HPLC and CE is still expensive and requires considerable operator expertize. Chemical derivatization with HPLC to permit fluorescence detection is certainly one of the more desirable methods for routine use to solve selectivity and detectability problems for drug samples not amenable to GC. As the cost of laser technology comes down, the same statement will eventually be true for CE also. The use of inexpensive light-emitting diodes as fluorescent visible and UV sources for CE also is showing promise.

ABBREVIATIONS AND ACRONYMS

ANDS	7-Aminonaphthalene-1,3-disulfonic Acid
AUFS	Absorbance Units Full Scale
BSA	*N*,*O*-Bis(trimethylsilyl)acetamide
BSTFA	*N*,*O*-Bis(trimethylsilyl)trifluoro-acetamide
CBD	Cannabidiol
CE	Capillary Electrophoresis
CE/LIF	Capillary Electrophoresis/Laser Induced Fluorescence
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CL	Chemiluminescence
DES	Diethylstilbestrol
DMCS	Dimethylchlorosilane
DMES	Dimethylethylsilyl
DMFDA	*N*,*N*-Dimethylformamide Dimethylacetal
DMIPS	Dimethylisopropylsilyl
DMSO	Dimethyl Sulfoxide
EC	Electron Capture
EDTA	Ethylenediamine Tetraacetic Acid
ESI	Electrospray Ionization
FITC	Fluorescein Isothiocyanate

FMOC	9-Fluorenylmethyl Chloroformate
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HFAA	Heptafluorobutyric Anhydride
HFBI	Heptafluorobutylimidazole
HMDS	Hexamethyldisilazane
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LIF	Laser-induced Fluorescence
L-TPC	(*S*)(−)*N*-Trifluoroacetyl-L-prolyl Chloride
MBTFA	*N*-Methylbistrifluoroacetamide
MEKC	Micellar Electrokinetic Chromatography
MS	Mass Spectrometry
MSTFA	*N*-Methyl-*N*-trimethylsilyl-trifluoroacetamide
MTBSTFA	*N*-Methyl-*N*-(*t*-butyldimethylsilyl)-trifluoroacetamide
NHSP	*N*-hydroxysuccinimide palmitate
OPA	*o*-Phthalaldehyde
PFAA	Pentafluoropropionic Anhydride
PFAI	Pentafluoropropionylimidazole
PFB-Br	Pentafluorobenzyl Bromide
p-IFN-α	Palmitoyl Derivatization of Interferon-α_{2b}
PITC	Phenylisothiocyanate
PMT	Photomultiplier
QUAT	Quaternary Ammonium Salt
rMAb	Recombinant Monoclonal Antibody
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
5-TAMRSE	5-Carboxytetramethylrhodamine Succinimidyl Ester
TBDMS	*t*-Butyldimethylsilyl
TFAA	Trifluoroacetic Anhydride
TFAI	Trifluoroacetylimidazole
TMAH	Trimethylanilinium Hydroxide
TMCS	Trimethylchlorosilane
TMS	Trimethylsilyl
TMSI	*N*-Trimethylsilylimidazole
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Pharmaceuticals and Drugs **(Volume 8)**
Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Planar Chromatography in Pharmaceutical Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis • Mass Spectrometry in Pharmaceutical Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Micellar Electrokinetic Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. *GC/HPLC Technical Handbook*, Pierce Biotechnology, Inc., Rockford, IL, 2003 (PDF download).
2. D.R. Knapp, *Handbook of Analytical Derivatization Reactions*, John Wiley & Sons, New York, 1979.
3. K. Blau, J.M. Halket, *Handbook of Derivatives for Chromatography*, 2nd edition, John Wiley & Sons, New York, 1993.
4. T. Kraemer, L.D. Paul, 'Bioanalytical Procedures for Determination of Drugs of Abuse', *Anal. Bioanal. Chem.*, **388**, 1415–1435 (2007).
5. L.D. Metcalf, A.A. Schmitz, 'Rapid Preparation of Fatty Acid Esters from Lipids for Gas Chromatographic Analysis', *Anal. Chem.*, **33**, 514–515 (1966).
6. V. Morra, P. Davit, R. Capra, M. Vincenti, A. Di Stilo, F. Botre, 'Fast Gas Chromatographic/Mass Spectrometric Determination of Diuretics and Masking Agents in Human Urine-Development and Validation of a Productive Screening Protocol for Antidoping Analysis', *J. Chromatogr. A*, **1135**, 219–229 (2006).
7. M. Bataineh, J. Nolte, B. Kuhlmann, N. Zullei-Seibert, M. Borges, M. Grote, 'Degradation Behavior of Selected Pharmaceuticals and their Main Metabolites in Model Systems for Slow Sand Filtration', *Curr. Pharm. Anal.*, **2**, 313–322 (2006).

8. W.C. Krossa, J. MacGee, S. Ramachandran, A.J. Webber, 'Pyrolytic Methylation/Gas Chromatography: A Short Review', *J. Chromatogr. Sci.*, **17**, 177–187 (1979).
9. W.C. Lin, H.C. Chen, W.H. Ding, 'Determination of Pharmaceutical Residues in Waters by Solid Phase Extraction and Large Volume On-Line Derivatization with Gas Chromatography-Mass Spectrometry', *J. Chromatogr. A*, **1055**, 279–285 (2005).
10. A.B. Jones, M.A. Elsohly, J.A. Bedford, C.E. Turner, 'Determination of Cannabidiol in Plasma by Electron Capture Gas Chromatography', *J. Chromatogr.*, **226**, 99–105 (1981).
11. K. Redderson, T. Heberer, 'Multi-Compound Methods for the Detection of Pharmaceutical Residues in Various Waters Applying Solid Phase Extraction (SPE) and Gas Chromatography with Mass Spectrometric Detection (GC/MS)', *J. Sep. Sci.*, **26**, 1443–1450 (2003).
12. P. Hartvig, J. Vessman, A. Kabi, 'Determination of Tertiary Amines as Pentafluorobenzyl Carbamates by Electron Capture Gas Chromatography', *Anal. Lett.*, **7**, 223 (1974).
13. M. Nishida, A. Namera, M. Yashiki, K. Kimura, 'Miniaturized Sample Preparation Method for Determination of Amphetamines in Urine', *Forensic Sci. Int.*, **143**, 163–167 (2004).
14. M. Nishida, M. Yashiki, A. Namera, K. Kimura, 'Single Hair Analysis of Methamphetamine and Amphetamine by Solid Phase Micro-extraction Coupled with In-matrix Derivatization', *J. Chromatogr. B*, **842**, 106–110 (2006).
15. G. Frison, L. Tedeschi, D. Favretto, A. Reheman, S.D. Ferrara, 'Gas Chromatography/Mass Spectrometry Determination of Amphetamine Related Drugs and Ephedrines in Plasma, Urine, and Hair Samples after Derivatization with 2,2,2,-Trichloroethylchloroformate', *Rapid Commun. Mass Spectrom.*, **19**, 919–927 (2005).
16. H. Brown, K.P. Kirkbride, P.E. Pigou, G.S. Walker, 'New Developments in SPME, Part 1: The Use of Vapor-Phase Deprotonation and On-Fiber Derivatization with Alkylchloroformates in the Analysis of Preparations Containing Amphetamines', *J. Forensic Sci.*, **48**, 1231–1238 (2003).
17. F.C.P. deToledo, M. Yonamine, R.L.D. Moreau, O.A. Silva, 'Determination of Cocaine, Benzoylecgonine, and Cocaethylene in Human Hair by Solid Phase Microextraction and Gas Chromatography-Mass Spectrometry', *J. Chromatogr. B*, **798**, 361–365 (2003).
18. M.J. Paik, D.T. Nguyen, I.S. Cho, S. Lee, K.R. Kim, 'Measurement and Identification of R- and (S)-6-O-Desmethylnaproxen Enantiomers as Diastereomeric O-Ethoxycarbonyl/(R)+1-phenylethylamides in Urine by GC and GC-MS', *Chromatographia*, **63**, 623–628 (2006).
19. S.M. Wang, T.C. Wang, Y.S. Giang, 'Simultaneous Determination of Amphetamine and Methamphetamine Enantiomers in Urine by Simultaneous Liquid-Liquid Extraction and Diastereomeric Derivatization Followed by Gas Chromatographic-Isotope Dilution Mass Spectrometry', *J. Chromatogr. B*, **816**, 131–143 (2005).
20. L.M. Ross-Lee, M.J. Eadie, F. Bochner, W.D. Hooper, J.H. Tyrer, 'Electron-capture Gas Chromatographic Assay for Metoclopramide in Plasma', *J. Chromatogr.*, **183**, 175–184 (1983).
21. J.E. Sullivan, L.R. Schewe, 'Preparation and Gas Chromatography of Highly Volatile Trifluoracetylated Carbohydrates Using *N*-Methylbis(trifluoroacetamide)', *J. Chromatogr. Sci.*, **15**, 196–197 (1977).
22. M. Pujadas, S. Pichini, S. Poudavida, E. Menoyo, P. Zucarro, M. Farre, R. de la Torre, 'Development and Validation of a Gas Chromatography-Mass Spectrometry Assay for Hair Analysis of Amphetamine, Methamphetamine, and Methylenedioxy Derivatives', *J. Chromatogr. B*, **798**, 249–255 (2003).
23. E. Hidvegi, P. Fabian, Z. Hideg, G. Somogyi, 'GC-MS Determination of Amphetamines in Serum using On-line Trifluoroacetylation', *Forensic Sci. Int.*, **161**, 119–123 (2006).
24. J.B. Jones, L.D. Mell, 'A Simple Wash Procedure for Improving Chromatography of HFAA Derivatized Amphetamine Extracts for GC/MS Analysis', *J. Anal. Toxicol.*, **17**, 447 (1993).
25. V. Maresova, J. Chadt, L. Prikryl, 'Simultaneous Determination of Amphetamines and Amphetamine-Derived Designer Drugs in Human Urine by GC-MS', *Neuroendocrinol. Lett.*, **27**, 121–124 (2006).
26. V. Habrdova, F.T. Peters, D.S. Theobald, H.H. Maurer, 'Screening for and Validated Quantification of Phenethylamine-type Designer Drugs and Mescaline in Human Blood Plasma by Gas Chromatography/Mass Spectrometry', *J. Mass Spectrom.*, **40**, 785–795 (2005).
27. K.J. Chia, S.D. Huang, 'Simultaneous Derivatization and Extraction of Amphetamine-like Drugs in Urine with Headspace Solid-Phase Microextraction Followed by Gas Chromatography-Mass Spectrometry', *Anal. Chim. Acta*, **539**, 49–54 (2005).
28. A. Kankaanpaa, T. Gunner, K. Ariniemi, P. Lilisunde, S. Mykkanen, T. Seppala, 'Single-Step Procedure for Gas Chromatography-Mass Spectrometry Screening and Quantitative Determination of Amphetamine-type Stimulants and Related Drugs in Blood, Serum, Oral Fluid, and Urine Samples', *J. Chromatogr. B*, **810**, 57–68 (2004).
29. T. Awad, J. DeRuiter, C.R. Clark, 'Chromatographic and Mass Spectral Studies on Methoxy Methyl Methamphetamines Related to 3,4-Methylenedioxymethamphetamine', *J. Chromatogr. Sci.*, **45**, 466–476 (2007).

30. T. Awad, C.R. Clark, J. DeRuiter, 'GC-MS Analysis of Acylated Derivatives of the Side-Chain Regioisomers of 4-Methoxy-3-methyl-phenethylamines Related to Methylenedioxymethamphetamine', *J. Chromatogr. Sci.*, **45**, 477–485 (2007).

31. E.Y. Han, Y. Park, W.Y. Yang, J. Lee, S. Lee, E. Kim, M. Lim, H.S. Chung, 'The Study of Metabolite-to-Parent Drug Ratios of Methamphetamine and Methylenedioxymethamphetamine in Hair', *Forensic Sci. Int.*, **161**, 124–129 (2006).

32. J.Y. Kim, K.S. Jung, M.K. Kim, J. Il Lee, M.K. In, 'Simultaneous Determination of Psychotropic Phenylalkylamines Derivatives in Human Hair by Gas Chromatography/Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **21**, 1705–1720 (2007).

33. H.R. Lin, A.C. Lua, 'Simultaneous Determination of Amphetamines and Ketamines in Urine by Gas Chromatography/Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **20**, 1724–1730 (2006).

34. Y.L. Tseng, M.H. Shieh, F.H. Kuo, 'Metabolites of Ephedrines in Human Urine after Administration of a Single Therapeutic Dose', *Forensic Sci. Int.*, **157**, 149–155 (2006).

35. H. Tsutsumi, M. Katagi, A. Miki, N. Shima, T. Kamata, M. Nishikawa, K. Nakajima, H. Tsuchihashi, 'Development of Simultaneous Gas Chromatography-Mass Spectrometric and Liquid Chromatography-Electrospray Ionization Mass Spectrometric Determination Method for the New Designer Drugs, N-Benzylpiperazine (BZP), 1-(3-Trifluoromethylphenyl)piperazine (TFMPP), and their Main Metabolites in Urine', *J. Chromatogr. B*, **819**, 315–322 (2005).

36. C. Salgado-Petinal, J.P. Lamas, C. Garcia-Jares, M. Llompart, R. Cela, 'Rapid Screening of Selective Serotonin Re-uptake Inhibitors in Urine Samples using Solid Phase Microextraction Gas Chromatography-Mass Spectrometry', *Anal. Bioanal. Chem.*, **382**, 1351–1359 (2005).

37. S.H. Tzing, A. Ghule, J.Y. Liu, Y.C. Ling, 'On-line Derivatization Gas Chromatography with Furan Chemical Ionization Tandem Mass Spectrometry for Screening of Amphetamines in Urine', *J. Chromatogr. A*, **1137**, 76–83 (2006).

38. T. Nadulski, F. Pragst, 'Simple and Sensitive Determination of Delta(9)-tetrahydrocannabinol, Cannabidiol, and Cannabinol in Hair by Combined Silylation, Headspace Solid Phase Microextraction, and Gas Chromatography-Mass Spectrometry', *J. Chromatogr. B*, **846**, 78–85 (2007).

39. W.-L. Wang, W.D. Darwin, E.J. Cone, 'Simultaneous Assay of Cocaine, Heroin, and Metabolites in Hair, Plasma, Saliva, and Urine by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. B*, **660**, 279–290 (1994).

40. S. Pirnay, S. Bouchonnet, F. Herve, D. Libong, N. Milan, P. d'Athis, F. Bud, I. Ricordel, 'Development and Validation of a Gas Chromatography-Mass Spectrometry Method for the Simultaneous Determination of Buprenorphine, Flunitrazepam, and their Metabolites in Rat Plasma: Application to the Pharmacokinetic Study', *J. Chromatogr. B*, **807**, 335–342 (2004).

41. I. Alvarez, F. Palos, A.M. Bermejo, P. Fernandez, M.J. Tabernero, 'Simultaneous Determination of Methadone, Heroin, Cocaine, and their Metabolites in Urine by GC-MS', *Anal. Lett.*, **39**, 1393–1399 (2006).

42. P. Fernandez, L. Bujan, A.M. Bermejo, M.J. Tabernero, 'Gas Chromatographic Determination of Cocaine and its Metabolites in Blood and Urine from Cocaine Users in Northwestern Spain', *J. Appl. Toxicol.*, **24**, 283–287 (2004).

43. S.L. Rice, S. Mitra, 'Microwave-Assisted Solvent Extraction of Solid matrices and Subsequent Detection of Pharmaceuticals and Personal Care Products', *Anal. Chim. Acta*, **589**, 125–132 (2007).

44. J. Ortuno, R. Delatorre, J. Segura, J. Cami, 'Simultaneous Detection in Urine of Cocaine and Its Main Metabolites', *J. Pharm. Biomed. Anal.*, **8**, 911–914 (1990).

45. A. Lajeunesse, C. Gagon, 'Determination of Acidic Pharmaceutical Products and Carbamazepine in Roughly Primary-Treated Wastewater by Solid Phase Extraction and Gas Chromatography-Tandem Mass Spectrometry', *Int. J. Environ. Anal. Chem.*, **87**, 565–578 (2007).

46. S. Huq, A. Dixon, K. Kelly, K.M.R. Kallury, 'Novel Solid-Phase Extraction Protocol for 11-nor-9-carboxy-Delta(9)-tetrahydrocannabinol from Urine Samples Employing a Polymeric Mixed-Mode cation Exchange resin, Strata-X-C, Suitable for Gas Chromatography-Mass Spectrometry or Liquid Chromatography-Mass Spectrometry Analysis', *J. Chromatogr. A*, **1073**, 355–362 (2005).

47. N.A. Zakhari, M.I. Walsh, M.S. Rizk, S.S. Toubar, C.J.W. Brooks, W.J. Cole, 'Gas Chromatographic Assay of Methyltestosterone in Tablets', *J. Pharm. Biomed. Anal.*, **9**, 705–709 (1991).

48. K.R. Kim, M.K. Hahn, A. Zlatkis, E.C. Horning, B.S. Middleditch, 'Simultaneous Gas Chromatography of Volatile and Nonvolatile Carboxylic Acids as *t*-Butyldimethylsilyl Derivatives', *J. Chromatogr.*, **468**, 289–301 (1989).

49. Z.R. Yu, S. Peldszuz, P.M. Huck, 'Optimizing Gas Chromatographic-Mass Spectrometric Analysis of Selected Pharmaceuticals and Endocrine-Disrupting Substances in Water using Factorial Experimental Design', *J. Chromatogr. A*, **1148**, 65–77 (2007).

50. T. Gunnar, K. Ariniemi, P. Lillsunde, 'Validated Toxicological Determination of 30 Drugs of Abuse as Optimized Derivatives in Oral Fluid by Long

Column Fast Gas Chromatography/Electron Impact Mass Spectrometry', *J. Mass Spectrom.*, **40**, 739–753 (2005).

51. T. Gunnar, K. Ariniemi, P. Lillsunde, 'Fast Gas Chromatography Negative-Ion Chemical Ionization Mass Spectrometry with Microscale Volume Sample Preparation for the Determination of Benzodiazepines and Alpha-hydroxy Metabolites, Zalepion, and Zopicione in Whole Blood', *J. Mass Spectrom.*, **41**, 741–754 (2006).
52. T. Gunnar, S. Mykkanen, K. Ariniemi, P. Lillsunde, 'Validated Semiquantitative/Quantitative Screening of 51 Drugs in Whole Blood as Silylated Derivatives by Gas Chromatography-Selected Ion Monitoring Mass Spectrometry and Gas Chromatography Electron Capture Detection', *J. Chromatogr. B*, **806**, 205–209 (2004).
53. C. Brunelli, C. Bichi, A. Di Stilo, A. Salomone, M. Vincenti, 'High-Speed Gas Chromatography in Doping Control: Fast GC and Fast GC/MS Determination of Beta-Andrenoceptor Ligands and Diuretics', *J. Sep. Sci.*, **29**, 2765–2771 (2006).
54. S. Zuhike, U. Dunnbier, T. Heberer, 'Detection and Identification of Phenazone Type Drugs and their Microbial Metabolites in Ground and Drinking Water by Applying Solid Phase Extraction and Gas Chromatography with Mass Spectrometric Detection', *J. Chromatogr. A*, **1050**, 201–209 (2004).
55. C.F. Poole, W.-F. Sye, S. Singhawangcha, F. Hsu, A. Zlatkis, A. Arfwidsson, J. Vessman, 'New Electron Capturing Pentafluorophenyldialkylchlorosilanes as Versatile Derivatizing Reagents for Gas Chromatography', *J. Chromatogr.*, **199**, 123–142 (1980).
56. J. Vessman, K.E. Karlsson, O. Gyllenhaal, 'Direct Derivatization of Drugs in Untreated Biological Samples for Gas Chromatographic Analysis', *J. Pharm. Biomed. Anal.*, **4**, 828–834 (1986).
57. L.H. Yang, C.Y. Lan, H.T. Liu, J. Dong, T.G. Luan, 'Full Automation of Solid Phase Microextraction/on-Fiber Derivatization for Simultaneous Determination of Endocrine-Disrupting Chemicals and Steroid Hormones by Gas Chromatography-Mass Spectrometry', *Anal. Bioanal. Chem.*, **386**, 391–397 (2006).
58. Y.Q. Liu, S.R. Cho, N.D. Danielson, 'Solid Phase Microextraction and On-line Methylation Gas Chromatography for Aliphatic Carboxylic Acids', *Anal. Bioanal. Chem.*, **373**, 64–69 (2002).
59. R.W. Frei, J.F. Lawrence (eds.), *Chemical Derivatization in Analytical Chemistry*, Plenum Press, New York, Chapter 4, Vol. 1, 1981.
60. R.W. Frei, J.F. Lawrence (eds.), *Separation and Continuous Flow Techniques*, Plenum Press, New York, Chapter 5, Vol. 2, 1983.
61. F. Lawrence, *Organic Trace Analysis by Liquid Chromatography*, Academic Press, New York, Chapter 7, 1981.
62. I.S. Krull (ed.), *Reaction Detection in Liquid Chromatography*, Marcel Dekker, New York, 1986.
63. G.L. Hawk, *Biological/Biomedical Applications of Liquid Chromatography*, Marcel Dekker, New York, Vols. 1 and 2, 1986.
64. A., Henschen (ed.), *High Performance Liquid Chromatography in Biochemistry*, VCH, Deerfield Beach, FL, 1985.
65. R. Gupta, I. Sunshine, *CRC Handbook of Chromatography, Drugs*, CRC Press, Boca Raton, FL, Vols. 1 and 2, 1981.
66. R.W. Frei, A.H.M.T. Scholten, 'Reaction Detectors in HPLC', *J. Chromatogr. Sci.*, **17**, 152–160 (1979).
67. R.S. Deelder, M.G.F. Kroll, A.J.B. Beeren, J.H.M. Van Den Berg, 'Post-column Reactor Systems in Liquid Chromatography', *J. Chromatogr.*, **149**, 669–682 (1978).
68. R.W. Frei, 'Reaction Detectors in Modern Liquid Chromatography', *Chromatographia*, **15**, 161–166 (1982).
69. H. Engelhardt, U.D. Neue, 'Reaction Detector with Three Dimensional Coiled Open Tubes in HPLC', *Chromatographia*, **15**, 403–408 (1982).
70. R.W. Frei, H. Jansen, U.A.T. Brinkman, 'Post-column Reaction Detectors for HPLC', *Anal. Chem.*, **57**, 1529A–1539A (1985).
71. H. Jansen, U.A.T. Brinkman, R.W. Frei, 'Post-column Reaction and Extraction Detectors for Narrow-bore High Performance Liquid Chromatography', *J. Chromatogr. Sci.*, **23**, 279–284 (1985).
72. U.A.T. Brinkman, 'A Review of Reaction Detection in HPLC', *Chromatographia*, **24**, 190–200 (1987).
73. C.E. Werkhovern-Goewie, C. De Ruiter, U.A.T. Brinkman, R.W. Frei, G.J. De Jong, C.J. Little, O. Stahel, 'Automated Determination of Drugs in Blood Samples After Enzymatic Hydrolysis Using Precolumn Switching and Post-column Reaction Detection', *J. Chromatogr.*, **255**, 79–90 (1983).
74. S. Gorog, A. Lauko, M. Renyei, B. Hegedus, 'New Derivatization Reactions in Pharmaceutical Analysis', *J. Pharm. Biomed. Anal.*, **1**, 497–506 (1983).
75. L.A. Sternson, 'The Application of Chemical Derivatization to Clinical Drug Analysis', *Xenobiotica*, **17**, 385–396 (1987).
76. R. Tsuji, W. Morozowich (eds.), *GLC and HPLC Determination of Therapeutic Agents*, Marcel Dekker, New York, 1978.
77. S. Ahuja, 'Chemical Derivatization for the Liquid Chromatography of Compounds of Pharmaceutical Interest', *J. Chromatogr. Sci.*, **17**, 168–172 (1979).

78. N.D. Danielson, M.A. Targove, B.E. Miller, 'Pre- and Post-column Derivatization Chemistry in Conjunction with HPLC for Pharmaceutical Analysis', *J. Chromatogr. Sci.*, **26**, 362–371 (1988).

79. J.F. Lawrence, U.A.T. Brinkman, R.W. Frei, 'Continuous Post-column Ion-pair Extraction Detection of Some Basic Organic Compounds in Normal-phase Chromatography', *J. Chromatogr.*, **185**, 473–481 (1979).

80. H. Derendorf, E.R. Garrett, 'High-performance Liquid Chromatographic Assay of Methadone, Phencyclidine, and Metabolites by Postcolumn Ion-pair Extraction and On-line Fluorescent Detection of the Counterion with Applications', *J. Pharm. Sci.*, **72**, 630–635 (1983).

81. J.C. Gfeller, J.M. Huen, J.P. Thevenin, 'Automation of Pre-column Derivatization in High-performance Liquid Chromatography; Application to Ion-pair Partition Chromatography', *J. Chromatogr.*, **166**, 133–140 (1978).

82. J.C. Gfeller, J.M. Huen, J.P. Thevenin, 'Automated Pre-column Derivatisation HPLC', *Chromatographia*, **12**, 368–370 (1979).

83. P.E. Nelson, S.L. Nolan, K.R. Bedford, 'High-performance Liquid Chromatography Detection of Morphine by Fluorescence After Post-column Derivatisation', *J. Chromatogr.*, **234**, 407–414 (1982).

84. P.E. Nelson, 'High-performance Liquid Chromatography Detection of Morphine by Fluorescence After Post-column Derivatisation; II. The Effect of Micelle Formation', *J. Chromatogr.*, **298**, 59–65 (1984).

85. F. Tagliaro, A. Frigerio, R. Dorizzi, G. Lubli, M. Marigo, 'Liquid Chromatography with Pre-column Dansyl Derivatisation and Fluorimetric Detection Applied to the Assay of Morphine in Biological Samples', *J. Chromatogr.*, **330**, 323–331 (1985).

86. H.J.G.M. Derks, K. Van Tuillert, D.P.K.H. Pereboom-De Fauw, G. Zomer, J.G. Loeber, 'Determination of the Heroin Metabolite 6-Acetylmorphine by High-performance Liquid Chromatography Using Automated Pre-column Derivatization and Fluorescence Detection', *J. Chromatogr.*, **370**, 173–178 (1986).

87. J.R. Lang, J.T. Stewart, I.L. Honigberg, 'Post-column Air-segmentation Photochemical Reactor for the Fluorometric Detection of Reserpine by Liquid Chromatography', *J. Chromatogr.*, **264**, 144–148 (1983).

88. J.R. Lang, I.L. Honigberg, J.T. Stewart, 'Post-column Air-segmentation Reactor for the Fluorometric Detection of Reserpine by Liquid Chromatography', *J. Chromatogr.*, **252**, 288–292 (1982).

89. K. Isaksson, 'Evaluation of a Chemical Reaction Detector with Electrochemical Production of Bromine for the Liquid Chromatographic Determination of Physostigmine', *J. Chromatogr.*, **411**, 229–236 (1987).

90. A.H.M.T. Scholten, R.W. Frei, 'Identification of Ergot Alkaloids with a Photochemical Reaction Detector in Liquid Chromatography', *J. Chromatogr.*, **176**, 349–357 (1979).

91. P.J. Twitchett, P.L. Williams, A.C. Moffat, 'Photochemical Detection in High-performance Liquid Chromatography and Its Application to Cannabinoid Analysis', *J. Chromatogr.*, **149**, 683–691 (1978).

92. R. Gatti, M.G. Gioa, V. Cavrini, 'Determination of Cinchona Alkaloids and Vitamin B-6 by High Performance Liquid Chromatography with Fluorescence Detection', *Anal. Chim. Acta*, **512**, 85–91 (2004).

93. H.K. Borys, R. Karler, 'Post-column Derivatization Procedure for the Colorimetric Analysis of Tissue Cannabinoids Separated by High-performance Liquid Chromatography', *J. Chromatogr.*, **205**, 303–323 (1981).

94. M.A. Rendi, P.P. Ellis, J. Gal, 'High-performance Liquid Chromatographic Assay for Pilocarpine in Aqueous Humor', *J. Chromatogr.*, **336**, 337–344 (1984).

95. C. Durruc, Ph. Hermann, A. Haddouche, M.M. Badarani, J.P. Thenot, 'Determination of *S*-Carboxymethyl-L-cysteine in Plasma by High-performance Liquid Chromatography with Column Switching Following Precolumn Derivatization', *J. Chromatogr.*, **417**, 208–215 (1987).

96. E.W. Wuis, E.W.J. Beneken Kolmer, L.E.C. Van Beijsterveldt, R.C.M. Burgers, T.B. Vree, E. Van Der Kleyn, 'Enantioselective High-performance Liquid Chromatographic Determination of Baclofen After Derivatization with a Chiral Adduct of *o*-Phthaldialdehyde', *J. Chromatogr.*, **415**, 419–422 (1987).

97. E.W. Wuis, R.J. Dirks, T.B. Vree, E. Van Der Kleyn, 'High-performance Liquid Chromatographic Analysis of Baclofen in Plasma and Urine of Man After Precolumn Extraction and Derivatization with *o*-Phthaldialdehyde', *J. Chromatogr.*, **337**, 341–350 (1985).

98. D.J. Sweeney, N.H. Greig, S.I. Rapoport, 'High-performance Liquid Chromatographic Analysis of Melphalan in Plasma, Brain and Peripheral Tissue by *o*-Phthalaldehyde Derivatization and Fluorescence Detection', *J. Chromatogr.*, **339**, 434–439 (1985).

99. M.G. Chernobrovkin, I.A. Anan'eve, E.N. Shapovalova, O.A. Shpigun, 'Determination of Amino Acid Enantiomers in Pharmaceuticals by Reversed Phase High Performance Liquid Chromatography', *J. Anal. Chem.*, **59**, 55–63 (2004).

100. C.K. Zacharis, G.A. Theodoridis, A.N. Voulgaropoulos, 'Coupling of Sequential Injection with Liquid Chromatography for the Automated Derivatization and On-line Determination of Amino Acids', *Talanta*, **69**, 841–847 (2006).

101. C. Chafer-Pericas, P. Campins-Falco, R. Herraez-Hernandez, 'Application of Solid-Phase Microextraction Combined with Derivatization to the Enantiomeric

Determination of Amphetamines', *J. Pharm. Biomed. Anal.*, **40**, 1209–1217 (2006).

102. R. Gatti, M.G. Giola, P. Andreatta, G. Pentassugha, 'HPLC-Fluorescence Determination of Amino Acids in Pharmaceuticals after Pre-Column Derivatization with Phanquinone', *J. Pharm. Biomed. Anal.*, **35**, 339–348 (2004).
103. P. de Montigny, J.F. Stobaugh, R.S. Givens, R.G. Carlson, K. Srinivasachar, L.A. Sternson, T. Higuchi, 'Naphthalene-2,3-dicarboxaldehyde/Cyanide Ion; a Rationally Designed Fluorogenic Reagent for Primary Amines', *Anal. Chem.*, **59**, 1096–1101 (1987).
104. G.E. Greismann, W.Y. Chan, O.M. Rennert, 'Determination of γ-Aminobutyric Acid by Reversed-phase High-performance Liquid Chromatography and Pre-column Labeling for Fluorescence Detection', *J. Chromatogr.*, **230**, 121–124 (1982).
105. P.A. Lewis, A.J. Woodward, J. Maddock, 'Improved Method for the Determination of *N*-Acetylcysteine in Human Plasma by High-performance Liquid Chromatography', *J. Chromatogr.*, **327**, 261–267 (1985).
106. J. Ishida, M. Kai, Y. Ohkura, 'Determination of *p*-Hydroxybestatin in Human Serum by High-performance Liquid Chromatography Using Fluorescence Detection', *J. Chromatogr.*, **344**, 267–274 (1985).
107. Q.S. Qu, X.Q. Tang, X.Y. Hu, S.C. Li, 'Applications of Pre-Column Derivatization for Amino Acids Analysis', *Prog. Chem.*, **18**, 789–793 (2006).
108. M. Kai, T. Miura, J. Ishida, Y. Ohkura, 'High-performance Liquid Chromatographic Method for Monitoring Leupeptin in Mouse Serum and Muscle by Pre-column Fluorescence Derivatization with Benzoin', *J. Chromatogr.*, **345**, 259–265 (1985).
109. K. Groningsson, M. Widahl-Nasman, 'Liquid Chromatographic Determination of Felypressin After Pre-column Derivatization with Fluorescamine', *J. Chromatogr.*, **291**, 185–194 (1984).
110. J.L. Meek, 'Derivatizing Reagents for High-performance Liquid Chromatography Detection of Peptides at the Picomole Level', *J. Chromatogr.*, **266**, 401–408 (1983).
111. J. Crommen, 'Ion-pairing Detection Technique in Reversed-phase High-performance Liquid Chromatography of Drugs and Related Compounds', *J. Pharm. Biomed. Anal.*, **1**, 549–555 (1983).
112. P.M. Elworthy, S.A. Tsementzis, D. Westhead, E.R. Hitchcock, 'Determination of Plasma Tranexamic Acid Using Cation-exchange High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **343**, 109–117 (1985).
113. W.J. Hurst, P.B. Toomey, 'High-performance Liquid Chromatographic Determination of Four Biogenic Amines in Chocolate', *Analyst*, **106**, 394–402 (1981).
114. K. Venema, W. Leever, J.O. Bakker, G. Haayer, J. Korf, 'Automated Precolumn Derivatization Device to Determine Neurotransmitter and Other Amino Acids by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **260**, 371–376 (1983).
115. R. Duff, E. Murrill, 'Determination of L-Buthionine-(*S*, *R*)-sulfoximine in Plasma by High-performance Liquid Chromatography with *o*-Phthalaldehyde Derivatization and Fluorometric Detection', *J. Chromatogr.*, **385**, 275–282 (1987).
116. R.N. Gupta, M. Lew, 'Liquid Chromatographic Determination of Mexiletine and Tocainide in Human Plasma with Fluorescence Detection After Reaction with a Modified *o*-Phthalaldehyde Reagent', *J. Chromatogr.*, **344**, 221–230 (1985).
117. H.S. Sista, D.M. Dye, J. Leonard, 'High-performance Liquid Chromatographic Method for Determination of Sulfapyridine in Human Saliva Using Post-column, In-line Derivatization with Fluorescamine', *J. Chromatogr.*, **273**, 464–468 (1983).
118. A.J. Sedman, J. Gal, 'Pre-column Derivatization with Fluorescamine and High-performance Liquid Chromatographic Analysis of Drugs; Application to Tocainide', *J. Chromatogr.*, **232**, 315–326 (1982).
119. A.A. Al-Majed, 'A Derivatization Reagent for Vigabatrin and Gabapentin in HPLC with Fluorescence Detection', *J. Liq. Chromatogr. Relat. Technol.*, **28**, 3119–3129 (2005).
120. O. Grech-Belanger, J. Turgeon, 'High-performance Liquid Chromatographic Assay for Mexiletine Enantiomers in Human Plasma', *J. Chromatogr.*, **337**, 172–177 (1985).
121. H. Umagat, R. Tscherne, 'High Performance Liquid Chromatographic Determination of Panthenol in Bulk, Premix, and Multivitamin Preparations', *Anal. Chem.*, **52**, 1368–1370 (1980).
122. C.O. Meese, P. Thalheimer, M. Eichelbaum, 'High-performance Liquid Chromatographic Method for the Analysis of Debrisoquine and Its (*S*)- (+)- and (*R*)-(−)-Hydroxy Metabolites in Urine', *J. Chromatogr.*, **423**, 344–350 (1987).
123. J. Salamoun, M. Smrz, F. Kiss, A. Salamounova, 'Column Liquid Chromatography of Methotrexate and Its Metabolites Using a Post-column Photochemical Reactor and Fluorescence Detection', *J. Chromatogr.*, **419**, 213–223 (1987).
124. T.R. Bosin, C.A. Jarvis, 'Derivatization in Aqueous Solution, Isolation and Separation of Tetrahydro-β-carbolines and Their Precursors by Liquid Chromatography', *J. Chromatogr.*, **341**, 287–293 (1985).
125. J. Bontemps, L. Bettendorf, J. Lombet, C. Grandfils, G. Dandrifosse, E. Schoffeniels, 'Poly(styrene–divinylbenzene) as Reversed-phase Adsorbent for the

High-performance Liquid Chromatographic Analysis of Thiochrome Derivatives of Thiamine and Phosphorylated Esters', *J. Chromatogr.*, **295**, 486–491 (1984).

126. H. Ohta, T. Baba, Y. Suzuki, E. Okada, 'High-performance Liquid Chromatographic Analysis of Thiamine in Rice Flour with Fluorimetric Post-column Derivatization', *J. Chromatogr.*, **284**, 281–284 (1984).
127. J.K. Fellman, W.E. Artz, P.D. Tassinari, C.L. Cole, J. Augustin, 'Simultaneous Determination of Thiamine and Riboflavin in Selected Foods by HPLC', *J. Food Sci.*, **47**, 2048–2050 (1982).
128. W. Von Sassen, M. Castro-Parra, E. Musch, M. Eichelbaum, 'Determination of Isoniazid, Acetylisoniazid, Acetylhydrazine and Diacetylhydrazine in Biological Fluids by High-performance Liquid Chromatography', *J. Chromatogr.*, **338**, 113–122 (1985).
129. H.C. Michaelis, H. Foth, G.F. Kahl, 'Determination of 5-Fluoro-2′-deoxyuridine in Human Plasma by High-performance Liquid Chromatography with Pre-column Fluorimetric Derivatization', *J. Chromatogr.*, **416**, 176–182 (1987).
130. D.L. Reynolds, L.A. Pachla, 'Analysis of 3-(2-(Ethylamino)propyl)-1,2,3,4-tetrahydro-5*H*(1)-benzopyrano [3,4-*c*]pyridin-5-one in Plasma by Liquid Chromatographic Column Switching After Derivatizing the Secondary Amine with Fluorescein-6-isothiocyanate', *J. Pharm. Sci.*, **74**, 1091–1094 (1975).
131. G. Bahrami, B. Mohammadi, 'Sensitive Microanalysis of Gabapentin by High Performance Liquid Chromatography in Human Serum using Pre-Column Derivatization with 4-Chloro-7-nitrobenzofurazan: Application to a Bioequivalence Study', *J. Chromatogr. B*, **837**, 24–28 (2006).
132. B.A. Bidlingmeyer, S.A. Cohen, T.L. Tarvin, 'Rapid Analysis of Amino Acids Using Pre-column Derivatization', *J. Chromatogr.*, **336**, 93–104 (1984).
133. P. Campins-Falco, R. Herraez-Hernandez, A. Sevillano-Cabeza, I. Trumpler, 'Derivatization of Amines in Solid-phase Extraction Support with 4-Fluorenylmethyl Chloroformate for Liquid Chromatography', *Anal. Chim. Acta*, **344**, 125–136 (1997).
134. C.R. Clark, M.M. Wells, 'Precolumn Derivatization of Amines for Enhanced Detectability in Liquid Chromatography', *J. Chromatogr. Sci.*, **16**, 332–339 (1978).
135. B.M. Farrell, T.M. Jefferies, 'An Investigation of High-performance Liquid Chromatographic Methods for the Analysis of Amphetamines', *J. Chromatogr.*, **272**, 111–128 (1983).
136. P. Campos-Falco, J. Verdu-Andres, R. Herraez-Hernandez, 'Separation of the Enantiomers of Primary and Secondary Amphetamines by Liquid Chromatography after Derivatization with (−)-1-(9-Fluorenyl)ethyl Chloroformate', *Chromatographia*, **57**, 309–316 (2003).
137. M. Bauer, L. Mailhe, L. Nguyen, 'Dosage de l'Heptaminol dans des Preparations Pharmaceutiques par Chromatographie Liquide Apres Derivation par le 4-*N*,*N*-Dimethylaminoazobenzene-4-isothiocyanate', *J. Chromatogr.*, **292**, 468–472 (1984).
138. P. Leroy, A. Nicolas, A. Moreau, 'Electrochemical Detection of Sympatomimetic Drugs, Following Pre-column *o*-Phthalaldehyde Derivatization and Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **282**, 561–569 (1983).
139. S.L. Wellons, M.A. Carey, 'High-performance Liquid Chromatographic Separation and Quantitation of Polyfunctional Amines as Their *m*-Toluoyl Derivatives', *J. Chromatogr.*, **154**, 219–225 (1978).
140. G. Pflugmann, H. Spahn, E. Mutschler, 'Determination of Metoprolol Enantiomers in Plasma and Urine Using (*S*)-(−)-Phenylethyl Isocyanate as a Chiral Reagent', *J. Chromatogr.*, **421**, 161–164 (1987).
141. R.J. Tscherne, H. Umagat, 'Determination of Isophenindamine in Phenindamine Tartrate Using an Argentated High-performance Liquid Chromatographic Mobile Phase', *J. Pharm. Sci.*, **69**, 342–344 (1980).
142. R.W. Frei, J.F. Lawrence, U.A.T. Brinkman, I. Honigberg, 'An HPLC Fluorescence Detection System for Amines', *J. High Resolut. Chromatogr. Commun.*, **2**, 11–14 (1979).
143. J.F. Lawrence, U.A.T. Brinkman, R.W. Frei, 'Extraction Detector for High-performance Liquid Chromatography Using Solvent Segmentation of the Column Effluent', *J. Chromatogr.*, **171**, 73–80 (1979).
144. J.C. Gfeller, G. Frey, J.M. Huen, J.P. Thevenin, 'Determination of Amines Using the Fluorimetric Ion-pair Technique as a Post-column Reactor for High-performance Liquid Chromatography', *J. Chromatogr.*, **172**, 141–151 (1979).
145. C. van Buuren, J.F. Lawrence, U.A.T. Brinkman, I.L. Honigberg, R.W. Frei, 'Reversed-phase Liquid Chromatography of Basic Drugs and Pesticides with a Fluorigenic Ion-pair Extraction Detector', *Anal. Chem.*, **52**, 700–704 (1980).
146. J.C. Gfeller, G. Frey, J.M. Huen, J.P. Thevenin, 'Post-column Derivatization Using the Fluorometric Ion-pair Technique', *J. High Resolut. Chromatogr. Commun.*, **1**, 213–214 (1978).
147. Y. Miyamoto, 'Highly Sensitive Determination of Chlorpheniramine as Fluorescence Derivative of High-performance Liquid Chromatography', *J. Chromatogr.*, **420**, 63–72 (1987).
148. G. Gubitz, R. Wintersteiger, A. Hartinger, 'Fluorescence Derivatization of Tertiary Amines with 2-Naphthyl Chloroformate', *J. Chromatogr.*, **218**, 51–56 (1981).

149. J.B. Noffsinger, N.D. Danielson, 'Liquid Chromatography of Aliphatic Trialkylamines with Post-column Chemiluminescent Detection Using Tris(2, 2′-bipyridine)ruthenium(III)', *J. Chromatogr.*, **387**, 520–524 (1987).

150. N.D. Danielson, 'Analytical Applications: Flow Injection, Liquid Chromatography, and Capillary Electrophoresis', in *Electrogenerated Chemiluminescence*, ed. A.J. Bard, Marcel Dekker, 397–444, 2004.

151. J.A. Holeman, N.D. Danielson, 'Liquid Chromatography of Antihistamines Using Post-column Tris(2, 2′-bipyridine)ruthenium(III) Chemiluminescence Detection', *J. Chromatogr. A*, **679**, 277–284 (1994).

152. R.R. Brown, R. Bain, V.C. Jordan, 'Determination of Tamoxifen and Metabolites in Human Serum by High-performance Liquid Chromatography with Post-column Fluorescence Activation', *J. Chromatogr.*, **272**, 351–358 (1983).

153. M. Nieder, H. Jaeger, 'Quantification of Tamoxifen and *N*-Desmethyltamoxifen in Human Plasma by High-performance Liquid Chromatography, Photochemical Reaction and Fluorescence Detection, and Its Application to Biopharmaceutic Investigations', *J. Chromatogr.*, **413**, 207–217 (1987).

154. S. Zuehlke, U. Duennbler, T. Heberer, 'Determination of Polar Drug Residues in Sewage and Surface Water Applying Liquid Chromatography-Tandem Mass Spectrometry', *Anal. Chem.*, **76**, 6548–6554 (2004).

155. R.D. Barlow, P.A. Thompson, R.B. Stone, 'Simultaneous Determination of Nicotine, Cotinine and Five Additional Nicotine Metabolites in the Urine of Smokers Using Pre-column Derivatisation and High-performance Liquid Chromatography', *J. Chromatogr.*, **419**, 375–380 (1987).

156. M.E. Capella-Peiro, S. Carda-Broch, L. Monferrer-Pons, J. Esteve-Romero, 'Micellar Liquid Chromatographic Determination of Nicotinic Acid and Nicotinamide after Pre-column Konig reaction Derivatization', *Anal. Chim. Acta*, **517**, 81–87 (2004).

157. M.L. Qi, P. Wang, Y.J. Sun, J.J. Yang, 'Determination of Enantiomeric Purity of L-Carnitine and Acetyl-L-carnitine by Chiral Liquid Chromatography', *Chromatograhia*, **59**, 247–250 (2004).

158. J. Maibaum, 'Indirect High-performance Liquid Chromatographic Resolution of Racemic Tertiary Amines as Their Diastereomeric Urea Derivatives After *N*-Dealkylation', *J. Chromatogr.*, **436**, 269–278 (1988).

159. N. Strojny, J.A.F. de Silva, 'Determination of Diclofensine, an Antidepressant Agent, and Its Major Metabolites in Human Plasma by High-performance Liquid Chromatography with Fluorometric Detection', *J. Chromatogr.*, **341**, 313–331 (1985).

160. P.J.M. Kwakman, U.A.T. Brinkman, R.W. Frei, G.J. de Jong, F.J. Spruit, N.G.F.M. Lammers, J.H.M. van den Berg, 'A Post-column Extraction System for the Determination of Tertiary Amines by Liquid Chromatography with Chemiluminescence Detection', *Chromatographia*, **24**, 395 (1987).

161. K.C. Marsh, L.A. Sternson, A.J. Repta, 'Post-column Reaction Detector for Platinum(II) Antineoplastic Agents', *Anal. Chem.*, **56**, 491–497 (1984).

162. J.P. Anhalt, F.D. Sancilio, T. McCorkle, 'Gentamicin C-component Ratio Determination by High-pressure Liquid Chromatography', *J. Chromatogr.*, **153**, 489–493 (1978).

163. G. Seidl, H.P. Nerad, 'Gentamicin C: Separation of C_1, C_{1a}, C_2, C_{2a}, and C_{2b} Components by HPLC Using Isocratic Ion-exchange Chromatography and Post-column Derivatisation', *Chromatographia*, **25**, 169 (1988).

164. W. Buchberger, K. Winsauer, F. Nachtmann, 'Trace Determination of Penicillin V with High-performance Liquid Chromatography', *Fresenius J. Anal. Chem.*, **315**, 525–527 (1983).

165. M.E. Rogers, M.W. Adlard, G. Saunders, G. Holt, 'High-performance Liquid Chromatographic Determination of β-Lactam Antibiotics, Using Fluorescence Detection Following Post-column Derivatization', *J. Chromatogr.*, **257**, 91–100 (1983).

166. D.G. Musson, S.M. Maglietto, S.S. Hwang, D. Gravellese, W.F. Bayne, 'Simultaneous Quantification of Cycloserine and Its Prodrug Acetylacetonylcycloserine in Plasma and Urine by High-performance Liquid Chromatography Using Ultraviolet Absorbance and Fluorescence After Post-column Derivatization', *J. Chromatogr.*, **414**, 121–129 (1987).

167. D.G. Musson, S.M. Maglietto, W.F. Bayne, 'Determination of the Antibiotic Fludalanine in Plasma and Urine by High-performance Liquid Chromatography Using a Packed-bed, Post-column Reactor with *o*-Phthalaldehyde and 2-Mercaptoethanol', *J. Chromatogr.*, **338**, 357–367 (1985).

168. H.N. Meyers, J.V. Rindler, 'Determination of Spectinomycin by High-performance Liquid Chromatography with Fluorometric Detection', *J. Chromatogr.*, **176**, 103–108 (1979).

169. P.J. Claes, R. Busson, H. Vanderhaeghe, 'Determination of the Component Ratio of Commercial Gentamicins by High-performance Liquid Chromatography Using Pre-column Derivatization', *J. Chromatogr.*, **298**, 445–457 (1984).

170. R. Tawa, K. Koshide, S. Hirose, 'Pre-column Derivatization of Sisomicin with *o*-Phthalaldehyde-β-mercaptopropionic Acid and Its Application to Sensitive High-performance Liquid Chromatographic Determination with Fluorimetric Detection', *J. Chromatogr.*, **425**, 143–152 (1988).

171. H.P. Fiedler, A. Plaga, 'Separation of Amino Acids and Antibiotics by Narrow-bore and Normal-bore High-performance Liquid Chromatography with Pre-column Derivatization', *J. Chromatogr.*, **386**, 229–241 (1987).

172. H.P. Fiedler, A. Plaga, R. Schuster, 'Automated On-column Derivatization of the Antibiotics Phosphinothricin and Phosphinothricyl-alanyl-alanine with *o*-Phthalaldehyde and Microbore Column High-performance Liquid Chromatography for Quantitative Determination in Biological Cultures', *J. Chromatogr.*, **353**, 210–216 (1986).

173. S.K. Maitra, T.T. Yoshikawa, C.M. Steyn, L.B. Guze, M.C. Schotz, 'Determination of Amikacin Isomers by High Pressure Liquid Chromatography', *J. Liq. Chromatogr.*, **2**, 823–836 (1979).

174. R. Weigand, R.J. Coombes, 'Gentamicin Determination by High-performance Liquid Chromatography', *J. Chromatogr.*, **281**, 381–385 (1983).

175. M.E. Rogers, M.W. Adlard, G. Saunders, G. Holt, 'High Performance Liquid Chromatographic Determination of β-Lactam Antibiotics, Using Fluorescence Detection Following Post-column Derivatization', *J. Chromatogr.*, **257**, 91–100 (1983).

176. D. Westerlund, J. Carlqvist, A. Theodorsen, 'Analysis of Penicillins in Biological Material by Reversed Phase Liquid Chromatography and Post-column Derivatization', *Acta Pharm. Suec.*, **16**, 187–214 (1979).

177. M.M.L. Aerts, W.M.J. Beek, U.A.T. Brinkman, 'Monitoring of Veterinary Drug Residues by a Combination of Continuous Flow Techniques and Column-switching High-performance Liquid Chromatography; I. Sulfonamides in Egg, Meat and Milk Using Post-column Derivatization with Dimethylaminobenzaldehyde', *J. Chromatogr.*, **435**, 97–112 (1988).

178. W.J. Blanchflower, D.A. Rice, J.T. Hamilton, 'Simultaneous High-performance Liquid Chromatographic Determination of Monensin, Narasin and Salinomycin in Feeds Using Post-column Derivatisation', *Analyst*, **110**, 1283–1287 (1985).

179. J. Haginaka, J. Wakai, 'Liquid Chromatographic Determination of Penicillins by Post-column Alkaline Degradation', *Anal. Chem.*, **57**, 1568–1571 (1985).

180. J. Haginaka, J. Wakai, 'Liquid Chromatographic Determination of Penicillins by Postcolumn Degradation with Sodium Hypochlorite', *Anal. Chem.*, **58**, 1896–1898 (1986).

181. J. Haginaka, J. Wakai, H. Yasuda, T. Uno, K. Takahashi, T. Katagi, 'High-performance Liquid Chromatographic Determination of Ampicillin and Its Metabolites in Rat Plasma, Bile and Urine by Post-column Degradation with Sodium Hypochlorite', *J. Chromatogr.*, **400**, 101–111 (1987).

182. J. Haginaka, J. Wakai, 'Liquid Chromatographic Determination of Amoxicillin and Its Metabolites in Human Urine by Postcolumn Degradation with Sodium Hypochlorite', *J. Chromatogr.*, **413**, 219–226 (1987).

183. H. Kubo, Y. Kobayashi, T. Kinoshita, 'Fluorescence Determination of Streptomycin in Serum by Reversed-phase Ion-pairing Liquid Chromatography', *Anal. Chem.*, **58**, 2653–2655 (1986).

184. H. Scholl, K. Schmidt, B. Weber, 'Sensitive and Selective Determination of Picogram Amounts of Ciprofloxacin and Its Metabolites in Biological Samples Using High-performance Liquid Chromatography and Photothermal Post-column Derivatization', *J. Chromatogr.*, **416**, 321–330 (1987).

185. C.M. Selavka, I.S. Krull, K. Bratin, 'Analysis for Penicillins and Cefoperazone by HPLC–Photolysis–Electrochemical Detection (HPLE–hν-EC)', *J. Pharm. Biomed. Anal.*, **4**, 83–93 (1986).

186. R.S. Valverde, I.S. Perez, F. Franceschelli, M.M. Galera, M.D.G. Garcia, 'Determination of Photoirradiated Tetracyclines in Water by High Performance Liquid Chromatography with Chemiluminescence Detection Based Reaction of Rhodamine B with Cerium(IV)', *J. Chromatogr. A*, **1167**, 85–94 (2007).

187. H. Fabre, W.T. Kok, 'Determination of Cephalosporins and Decomposition Products by Liquid Chromatography with Indirect Electrochemical Detection', *Anal. Chem.*, **60**, 136–141 (1988).

188. J. Haginaka, J. Wakai, H. Yasuda, T. Nakagawa, 'Direct Serum Injection with Micellar Liquid Chromatography; Chromatographic Behavior and Recovery of Cephalosporins', *Anal. Chem.*, **59**, 2732–2734 (1987).

189. M.D. Blanchin, W.T. Kok, H. Fabre, 'New Detection Modes for the Determination of Cephalosporins and Their Decomposition Products', *Chromatographia*, **24**, 625 (1987).

190. J. Carlqvist, D. Westerlund, 'Automated Determination of Amoxycillin in Biological Fluids by Column Switching in Ion-pair Reversed-phase Liquid Chromatographic Systems with Post-column Derivatization', *J. Chromatogr.*, **344**, 285–296 (1985).

191. M.A. Targove, N.D. Danielson, 'High Performance Liquid Chromatography of Clindamycin Antibiotics Using Tris(bipyridyl)ruthenium(III) Chemiluminescence Detection', *J. Chromatogr. Sci.*, **28**, 505–509 (1990).

192. S. Lam, E. Grushka, 'Reversed Phase Liquid Chromatography of Phenacyl Esters of Some Natural Penicillins', *J. Liq. Chromatogr.*, **1**, 33–41 (1978).

193. K. Tsuji, J.F. Goetz, W. VanMeter, K.A. Gusciora, 'Normal-phase High-performance Liquid Chromatographic Determination of Neomycin Sulfate Derivatized with 1-Fluoro-2,4-dinitrobenzene', *J. Chromatogr.*, **175**, 141–152 (1979).

194. P.A. Hartman, 'Liquid Chromatographic Determination of Methscopolamine Bromide in Various Veterinary Formulations', *J. Assoc. Off. Anal. Chem.*, **62**, 1099–1106 (1979).

195. P. Helboe, S. Kryger, 'Improved High-performance Liquid Chromatographic Method for Simultaneous Determination of Neamine, Neomycin B and Neomycin C in Neomycin Sulfate', *J. Chromatogr.*, **235**, 215–220 (1982).

196. L.T. Wong, A.R. Beaubien, A.P. Pakuts, 'Determination of Amikacin in Microlitre Quantities of Biological Fluids by High-performance Liquid Chromatography Using 1-Fluoro-2,4-dinitrobenzene Derivatization', *J. Chromatogr.*, **231**, 145–154 (1982).

197. D.M. Barends, J.S. Blauw, M.H. Smits, A. Hulshoff, 'Determination of Amikacin in Serum by High-performance Liquid Chromatography with Ultra-violet Detection', *J. Chromatogr.*, **276**, 385–394 (1983).

198. T. Harada, M. Iwamori, Y. Nagai, Y. Nomura, 'Analysis of Aminoglycoside Antibiotics as Benzoyl Derivatives by High-performance Liquid Chromatography and Its Application to the Quantitation of Neomycin in the Perilymph', *J. Chromatogr.*, **337**, 187–193 (1985).

199. P. Gambardella, R. Punziano, M. Gionti, C. Guadalupi, G. Mancini, A. Mangia, 'Quantitative Determination and Separation of Analogs of Aminoglycoside Antibiotics by High-performance Liquid Chromatography', *J. Chromatogr.*, **348**, 229–240 (1985).

200. R.B. Binns, K. Tsuji, 'High-performance Liquid Chromatographic Analysis of Neomycin in Petrolatum-based Ointments and in Veterinary Formulations', *J. Pharm. Sci.*, **73**, 69–72 (1984).

201. K. Tsuji, K.M. Jenkins, 'Derivatization of Secondary Amines with 2-Naphthalenesulfonyl Chloride for High-performance Liquid Chromatographic Analysis of Spectinomycin', *J. Chromatogr.*, **333**, 365–380 (1985).

202. P.S. Chu, M.I. Lopez, 'Determination of Nitrofuran Residues in Milk of Dairy Cows using Liquid Chromatography-Tandem Mass Spectrometry', *J. Agric. Food Chem.*, **55**, 2129–2135 (2007).

203. L. Tribalat, O. Paisse, G. Dessalces, M.F. Grenier-Loustalot, 'Advantages of LC-MS-MS Compared to LC-MS for the Determination of Nitrofuran Residues in Honey', *Anal. Bioanal. Chem.*, **386**, 2161–2168 (2006).

204. A.J. Shah, M.W. Adlard, 'Determination of β-Lactams and Their Biosynthetic Intermediates in Fermentation Media by Pre-column Derivatization Followed by Fluorescence Detection', *J. Chromatogr.*, **424**, 325–336 (1988).

205. J.O. Miners, I. Fearnley, K.J. Smith, D.J. Birkett, P.M. Brooks, M.W. Whitehouse, 'Analysis of D-Penicillamine in Plasma by Fluorescence Derivatisation with *N*-[*p*-(2-Benzoxazolyl)phenyl]maleimide and High-performance Liquid Chromatography', *J. Chromatogr.*, **275**, 89–96 (1983).

206. M.E. Rogers, M.W. Adlard, G. Saunders, G. Holt, 'High-performance Liquid Chromatographic Determination of Penicillins Following Derivatization to Mercury-stabilized Penicillenic Acids', *J. Liq. Chromatogr.*, **6**, 2019–2031 (1983).

207. M.E. Rogers, M.W. Adlard, G. Saunders, G. Holt, 'Derivatization Techniques for High-performance Liquid Chromatographic Analysis of β-Lactams', *J. Chromatogr.*, **297**, 385–391 (1984).

208. J. Haginaka, J. Wakai, 'High-performance Liquid Chromatographic Assay of Carbenicillin, Ticarcillin and Sulbenicillin in Serum and Urine Using Pre-column Reaction with 1,2,4-Triazole and Mercury (II) Chloride', *Analyst*, **110**, 1185–1188 (1985).

209. J. Haginaka, J. Wakai, 'High-performance Liquid Chromatographic Assay of Ampicillin, Amoxicillin and Ciclacillin in Serum and Urine Using a Pre-column Reaction with 1,2,4-Triazole and Mercury (II) Chloride', *Analyst*, **110**, 1277–1281 (1985).

210. J. Haginaka, J. Wakai, H. Yasuda, T. Uno, T. Nakagawa, 'High-performance Liquid Chromatographic Assay of Sulbactam Using Pre-column Reaction with 1,2,4-Triazole', *J. Chromatogr.*, **341**, 115–122 (1985).

211. H. Russ, D. McCleary, R. Katimy, J.L. Montana, R.B. Miller, R. Krishnamoorthy, C.W. Davis, 'Development and Validation of a Stability-indicating HPLC Method for the Determination of Tobramycin and Its Related Substances in an Ophthalmic Suspension', *J. Liq. Chromatogr.*, **21**, 2165–2181 (1998).

212. W. Lindner, W. Santi, '*N*-chloromethylphthalimides as Derivatization Reagents for High-performance Liquid Chromatography', *J. Chromatogr.*, **176**, 55–64 (1979).

213. W. Lindner, '*N*-Chloromethyl-4-nitro-phthalimid als Derivatiserungsreagens für die Hochleistungs-Flussigkeits-Chromatographie', *J. Chromatogr.*, **198**, 367–372 (1980).

214. A. Hulshoff, H. Roseboom, J. Renema, 'Improved Detectability of Barbiturates in High-performance Liquid Chromatography by Pre-column Labelling and Ultraviolet Detection', *J. Chromatogr.*, **186**, 535–541 (1979).

215. J. DeJong, M.W.F. Nielen, R.W. Frei, U.A.T. Brinkman, 'Selective On-line Sample Handling for the Determination of Barbiturates in Urine by Liquid Chromatography with Precolumn Technology', *J. Chromatogr.*, **381**, 431–437 (1986).

216. C.R. Clark, J.-L. Chan, 'Improved Detectability of Barbiturates in High Performance Liquid Chromatography by Post-column Ionization', *Anal. Chem.*, **50**, 635–637 (1978).

217. E.P. Scott, 'Application of Postcolumn Ionization in the High-performance Liquid Chromatographic Analysis of Butabarbital Sodium Elixir', *J. Pharm. Sci.*, **72**, 1089–1091 (1983).

218. C.H. Wolf, R.W. Schmid, 'Enhanced UV-detection of Barbiturates in HPLC Analysis by On-line Photochemical Reaction', *J. Liq. Chromatogr.*, **13**, 2207–2216 (1990).

219. J.P. Langenberg, U.R. Tjaden, 'Improved Method for the Determination of Vitamin K_1 Epoxide in Human Plasma with Electrofluorimetric Reaction Detection', *J. Chromatogr.*, **289**, 377–385 (1984).

220. J.P. Langenberg, U.R. Tjaden, 'Determination of (Endogenous) Vitamin K_1 in Human Plasma by Reversed-phase High-performance Liquid Chromatography Using Fluorometric Detection After Post-column Electrochemical Reduction; Comparison with Ultraviolet, Single and Dual Electrochemical Detection', *J. Chromatogr.*, **305**, 61–72 (1984).

221. B.E. Miller, N.D. Danielson, 'Fluorimetric Determination of Danthron in Pharmaceutical Tablets and in Urine', *Anal. Chim. Acta*, **192**, 293–299 (1987).

222. T. Seki, N. Hashida, T. Kanazawa, 'Determination of Tetraphenylacetone in Human Plasma by High-performance Liquid Chromatography with Fluorescence Derivatization Using Dansylhydrazine', *J. Chromatogr.*, **424**, 410–415 (1988).

223. R.L. Patience, J.D. Thomas, 'Rapid Concentration and Analysis of Short Chain Carboxylic Acids: Variation on a Theme', *J. Chromatogr.*, **234**, 225–230 (1982).

224. P.S. Mukherjee, H.T. Karnes, 'Reaction of 5-Bromomethylfluorescein (5-BMF) with Cefuroxime and Other Carboxyl-containing Analytes to Form Derivatives Suitable for Laser-induced Fluorescence Detection', *Analyst*, **121**, 1573–1579 (1996).

225. S. Ikenoya, O. Hiroshima, M. Ohmae, K. Kawabe, 'Electrochemical Detector for High Performance Liquid Chromatography; IV. Analysis of Fatty Acids, Bile Acids and Prosta-glandins by Derivatization to an Electrochemically Active Form', *Chem. Pharm. Bull.*, **28**, 2941–2947 (1980).

226. L.W. Brown, B.E. Carpenter, 'Comparison of Two High-pressure Liquid Chromatographic Assays for Carboprost, a Synthetic Prostaglandin', *J. Pharm. Sci.*, **69**, 1396–1399 (1980).

227. M. Hatsumi, S.-I. Kimata, K. Hirosawa, '9-Anthryldiazomethane Derivatives of Prostaglandins for High-performance Liquid Chromatographic Analysis', *J. Chromatogr.*, **253**, 271–275 (1982).

228. J.W. Cox, R.H. Pullen, M.E. Royer, 'Isolation of Plasma Components by Double Antibody Precipitation and Filtration: Application to the Chromatographic Determination of Arbaprostil [(15*R*)-15-methylprostaglandin E_2]', *Anal. Chem.*, **57**, 2365–2369 (1985).

229. R.H. Pullen, J.W. Cox, 'Determination of (15*R*)- and (15*S*)-15-methylprostaglandin E_2 in Human Plasma with Picogram per Milliliter Sensitivity by Column-switching High-performance Liquid Chromatography', *J. Chromatogr.*, **343**, 271–283 (1985).

230. T.A. Stein, L. Angus, E. Borrero, L.J. Auguste, L. Wise, 'High-performance Liquid-chromatographic Assay for Prostaglandins with the Use of *p*-(9-Anthroyloxy) phenacyl Bromide', *J. Chromatogr.*, **395**, 591–595 (1987).

231. J. Doehl, T. Greibrokk, 'High-performance Liquid Chromatographic Separation and Ultraviolet Detection of Prostaglandins, Oxidized by Pyridinium Dichromate', *J. Chromatogr.*, **282**, 435–442 (1983).

232. J. Doehl, T. Greibrokk, 'High-performance Liquid Chromatographic Separation and Determination of Prostaglandins, Oxidized by Pyridinium Dichromate; Optimization and Applications', *J. Chromatogr.*, **349**, 431–438 (1985).

233. M. Hatsumi, S. Kimata, K. Hirosawa, '9-Anthryldiazomethane Derivatives of Prostaglandins for High-performance Liquid Chromatographic Analysis', *J. Chromatogr.*, **253**, 271–275 (1982).

234. R.L. Veazy, T.A. Nieman, 'Chemiluminescence High-performance Liquid Chromatographic Detector Applied to Ascorbic Acid Determinations', *J. Chromatogr.*, **200**, 153–162 (1980).

235. R.W. Keating, P.R. Haddad, 'Simultaneous Determination of Ascorbic Acid and Dehydroascorbic Acid by Reversed-phase Ion-pair High-performance Liquid Chromatography with Pre-column Derivatisation', *J. Chromatogr.*, **245**, 249–255 (1982).

236. S.J. Ziegler, B. Meier, O. Sticher, 'Rapid and Sensitive Determination of Dehydroascorbic Acid in Addition to Ascorbic Acid by Reversed-phase High-performance Liquid Chromatography Using a Post-column Reduction System', *J. Chromatogr.*, **391**, 419–426 (1987).

237. M.A.J.S. Van Boekel, C.A.J.M. Meeuwissen, 'High-performance Liquid Chromatography of the Bis(dinitrophenyl)hydrazone Derivative of Ascorbic Acid, Applied to Analysis of Ascorbic Acid in Beer', *J. Chromatogr.*, **261**, 176–180 (1983).

238. T. Seki, Y. Yamaguchi, K. Noguchi, Y. Yanagihara, 'Determination of Ascorbic Acid in Human Urine by High-performance Liquid Chromatography Coupled with Fluorimetry After Post-column Derivatization with Benzamidine', *J. Chromatogr.*, **385**, 287–291 (1987).

239. T. Iwata, M. Yamaguchi, S. Hara, M. Nakamura, 'Determination of Total Ascorbic Acid in Human Serum by High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **344**, 351–355 (1985).

240. D. De Zeeuw, J.L. Leinfelder, D.C. Brater, 'Highly Sensitive Measurement of Indomethacin Using a High-performance Liquid Chromatographic Technique

Combined with Post-column In-line Hydrolysis', *J. Chromatogr.*, **380**, 157–162 (1986).

241. D.M. Seibert, F. Bochner, 'Determination of Plasma Aspirin and Salicylic Acid Concentrations After Low Aspirin Doses by High-performance Liquid Chromatography with Post-column Hydrolysis and Fluorescence Detection', *J. Chromatogr.*, **420**, 425–431 (1987).

242. C. Gonzalez-Barreiro, M. Lores, M.C. Casais, R. Cela, 'Simultaneous Determination of Neutral and Acidic Pharmaceuticals in Wastewater by High Performance Liquid Chromatography Post-Column Photocehmically Induced Fluorimetry', *J. Chromatogr. A*, **993**, 29–37 (2003).

243. A. Avgerinos, A.J. Hutt, 'Determination of the Enantiomeric Composition of Ibuprofen in Human Plasma by High-performance Liquid Chromatography', *J. Chromatogr.*, **415**, 75–83 (1987).

244. S. Bjorkman, 'Determination of the Enantiomers of Indoprofen in Blood Plasma by High-performance Liquid Chromatography After Rapid Derivatization by Means of Ethyl Chloroformate', *J. Chromatogr.*, **339**, 339–346 (1985).

245. S. Pedrazzini, W. Zanoboni-Muciaccia, C. Sacchi, A. Forgione, 'Determination of Flunoxaprofen Enantiomers in Biological Fluids by High-performance Liquid Chromatography', *J. Chromatogr.*, **415**, 214–220 (1987).

246. A.-Q. Zou, M. Xie, X.-D. Lou, 'Determination of Artesunic Acid After Chemical Derivatization with *o*,*p*-Nitrobenzyl-*N*,*N*′-diisopropylisourea by High-performance Liquid Chromatography and Ultraviolet Absorption', *J. Chromatogr.*, **410**, 217–221 (1987).

247. P.E. Minkler, S.T. Ingalls, L.S. Kormos, D.E. Weir, C.L. Hoppel, 'Determination of Carnitine, Butyrobetaine, and Betaine as 4′-Bromophenacyl Ester Derivatives by High-performance Liquid Chromatography', *J. Chromatogr.*, **336**, 271–283 (1984).

248. J. Gorham, E. McDonnell, 'High-performance Liquid Chromatographic Method for the Separation and Estimation of Choline, Glycinebetaine Aldehyde and Related Compounds', *J. Chromatogr.*, **350**, 245–254 (1985).

249. O.M. Steijger, D.A. Kamminga, H. Lingeman, U.A.T. Brinkman, 'An Acridinium Sulfonylamide as a New Chemiluminescent Label for the Determination of Carboxylic Acids in Liquid Chromatography', *J. Biolumin. Chemilumin.*, **13**, 31–40 (1998).

250. S. Gao, Z.-P. Zhang, L.E. Edinboro, L.C. Ngoka, H.T. Karnes, 'The Effect of Alkylamine additives on the Sensitivity of Detection for Paclitaxel and Docetaxel and Analysis in Plasma of Paclitaxel by Liquid Chromatography-Tandem mass Spectrometry', *Biomed. Chromatogr.*, **20**, 83–695 (2006).

251. K. Teshima, T. Kondo, C. Maeda, T. Oda, T. Hagimoto, R. Tsukuda, Y. Yoshimura, 'Application of 1-Alkylamines to a Liquid Chromatographic/Turbo Ion Spray Tandem Mass Spectrometric Method for Quantifying Metabolites of a New Bone Anabolic Agent, TAK-778, in Human Serum', *J. Mass Spectrom.*, **37**, 631–638 (2002).

252. J.J. Zhao, A.Y. Yang, J.D. Rogers, 'Effects of Liquid Chromatography Mobile Phase Buffer Contents on the Ionization and Fragmentation of Analytes in Liquid Chromatographic/Ion Spray Tandem Mass Spectrometric Determination', *J. Mass Spectrom.*, **37**, 421–433 (2002).

253. J.B. Quintana, T. Reemtsma, 'Sensitive Determination of Acidic Drugs and Triclosan in Surface and Wastewater by Ion-Pair Reverse-Phase Liquid Chromatography/Tandem Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **18**, 765–774 (2004).

254. L.D. Mell, A.R. Dasler, A.B. Gustafson, 'Pre-column Fluorescent Derivatization for High Pressure Liquid Chromatography with *o*-Phthalaldehyde: Separation of Urinary Catecholamines', *J. Liq. Chromatogr.*, **1**, 261–277 (1978).

255. T.P. Davis, C.W. Gehrke, C.W. Gehrke, T.D. Cunningham, K.C. Kuo, K.O. Gerhardt, H.D. Johnson, C.H. Williams Jr, 'High-performance Liquid-chromatographic Separation and Fluorescence Measurement of Biogenic Amines in Plasma, Urine, and Tissue', *Clin. Chem.*, **24**, 1317–1324 (1978).

256. G.W. Schieffer, 'Pre-column Electrochemical Cell for High-performance Liquid Chromatography', *Anal. Chem.*, **51**, 1573–1575 (1979).

257. N. Nimura, Y. Kasahara, T. Kinoshita, 'Resolution of Enantiomers of Norepinephrine and Epinephrine by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **213**, 327–330 (1981).

258. I.W. Wainer, T.D. Doyle, Z. Hamidzadeh, M. Aldridge, 'Resolution of Norephedrine as its 2-Oxazolidone Derivative: Enantiomeric Separation on a Chiral High-performance Liquid Chromatographic Stationary Phase and Preparative Regeneration of the Resolved Isomers', *J. Chromatogr.*, **268**, 107–111 (1983).

259. A. Mitsui, H. Nohta, Y. Ohkura, 'High-performance Liquid Chromatography of Plasma Catecholamines Using 1,2-Diphenylethylenediamine as Precolumn Fluorescence Derivatization Reagent', *J. Chromatogr.*, **344**, 61–70 (1985).

260. Y. Yui, M. Kimura, Y. Itokawa, C. Kawai, 'Ultramicro Method for the Determination of Picogram Amounts of Norepinephrine and Epinephrine by High-performance Liquid Chromatography', *J. Chromatogr.*, **177**, 376–379 (1979).

261. K.-I. Okamoto, Y. Ishida, K. Asai, 'Separation and Detection of Small Amounts of Catecholamines by High-performance Liquid Chromatography', *J. Chromatogr.*, **167**, 205–217 (1978).

262. Y. Yui, C. Kawai, 'Comparison of the Sensitivity of Various Post-column Methods for Catecholamine Analysis by High-performance Liquid Chromatography', *J. Chromatogr.*, **206**, 586–588 (1981).

263. N. Nimura, K. Ishida, T. Kinoshita, 'Novel Post-column Derivatization Method for the Fluorimetric Determination of Norepinephrine and Epinephrine', *J. Chromatogr.*, **221**, 249–255 (1980).

264. T. Yamamoto, A. Yamatodani, M. Nishimura, H. Wada, 'Determination of Dopamine-3- and -4- *o*-Sulfate in Human Plasma and Urine by Anion-exchange High-performance Liquid Chromatography with Fluorimetric Detection', *J. Chromatogr.*, **342**, 261–267 (1985).

265. T. Seki, Y. Yamaguchi, 'Fluorimetric Determination of Catecholamines Using Glycylglycine as the Reagent for Post-column Derivatization', *J. Chromatogr.*, **287**, 407–412 (1984).

266. T. Seki, Y. Yamaguchi, K. Noguchi, Y. Yanagihara, 'Estimation of Catecholamines by Ion-exchange Chromatography on Asahipak ES-502C, Using Glycylglycine as the Post-derivatizing Agent', *J. Chromatogr.*, **332**, 9–13 (1985).

267. N. Kiba, T. Mizuno, M. Furusawa, 'Determination of Catecholamines by High-performance Liquid Chromatography with Catalytic Photometric Detection', *J. Chromatogr.*, **329**, 147–152 (1985).

268. L.A. Seamark, D.J.H. Trafford, 'High-performance Liquid Chromatography of Vitamin D: Enhanced Ultraviolet Absorbance by Prior Conversion to Isotachysterol Derivatives', *J. Chromatogr.*, **197**, 271–273 (1980).

269. C.C. Clark, 'On-column Periodate Reaction Method for Analysis of Ephedrine Sulfate in Solid Dosage Forms: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **63**, 692–695 (1980).

270. J.E. Parkin, 'High-performance Liquid Chromatographic Assay of Menthol Using Indirect Photometric Detection', *J. Chromatogr.*, **303**, 436–439 (1984).

271. M.J. Wilson, T. Walle, 'Silica Gel High-performance Liquid Chromatography for the Simultaneous Determination of Propranolol and 4-Hydroxypropranolol Enantiomers After Chiral Derivatization', *J. Chromatogr.*, **310**, 424–430 (1984).

272. T. Walle, D.D. Christ, U.K. Walle, M.J. Wilson, 'Separation of the Enantiomers of Intact Sulfate Conjugates of Adrenergic Drugs by High-performance Liquid Chromatography After Chiral Derivatization', *J. Chromatogr.*, **341**, 213–216 (1985).

273. A.J. Sedman, J. Gal, 'Resolution of the Enantiomers of Propranolol and Other Beta-adrenergic Antagonists by High-performance Liquid Chromatography', *J. Chromatogr.*, **278**, 199–203 (1983).

274. M.B. Gupta, J.W. Hubbard, K.K. Midha, 'Separation of Enantiomers of Derivatised or Underivatised Propranolol by Means of High-performance Liquid Chromatography', *J. Chromatogr.*, **424**, 189–194 (1988).

275. I. Demian, D.F. Gripshover, 'High-performance Liquid Chromatographic Determination of Enantiomeric Purity of 1-Methyl-3-pyrolidinol via Derivatization with (*R*,*R*)-*o*,*o*-Dibenzoyltartaric Acid Anhydride', *J. Chromatogr.*, **387**, 532–535 (1987).

276. M. Ruud-Christensen, B. Salvesen, 'Separation of (*R*)- and (*S*)-Proxyphylline as Diastereoisomeric Camphanates by Reversed-phase Liquid Chromatography', *J. Chromatogr.*, **303**, 433–435 (1984).

277. Z. Shishan, Z. Mei-Yi, 'Application of Precolumn Reaction to High-performance Liquid Chromatography of Qinghaosu in Animal Plasma', *Anal. Chem.*, **58**, 289–292 (1986).

278. X.-D. Lou, M. Xie, 'Sub-nanogram Detection of Dihydroartemisinin After Chemical Derivatization with Diacetyldihydrofluorescein Followed by High-performance Liquid Chromatography and UV Absorption', *Chromatographia*, **23**, 112–114 (1987).

279. J. Goto, M. Saito, T. Chikai, N. Goto, T. Nambara, 'Studies in Steroids; CLXXXVII. Determination of Serum Bile Acids by High-performance Liquid Chromatography with Fluorescence Labeling', *J. Chromatogr.*, **276**, 289–300 (1983).

280. G. Schladitz-Keil, H. Spahn, E. Mutschler, 'Fluorimetric Determination of the Quaternary Compound Trospium and Its Metabolite in Biological Material After Derivatization with Benoxaprofen Chloride', *J. Chromatogr.*, **345**, 99–110 (1985).

281. M. Hoshino, K. Yajima, Y. Suzuki, A. Okahira, 'Determination of Nadolol Diastereomers in Dog Plasma Using Chiral Derivatization and Reversed-phase High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr. B*, **661**, 281–289 (1994).

282. A.T.R. Williams, S.A. Winfield, R.C. Belloli, 'Rapid, Specific Method for Diethylstilbestrol Analysis Using an In-line Photochemical Reactor with High-performance Liquid Chromatography and Fluorescence', *J. Chromatogr.*, **235**, 461–470 (1982).

283. R. Verbeke, P. Vanhee, 'Sensitive and Specific Post-column Fluorimetric Determination of Diethylstilbestrol Residues in Extracts of Urine and Animal Tissues at the 1-ppb Level', *J. Chromatogr.*, **265**, 239–255 (1983).

284. R.H. Van Den Berg, E.H.J.M. Jansen, G. Zomer, C. Enkelaar-Willemsen, R. Both-Miedema, R.W. Stephany, 'Identification of Anabolic Stilbene Derivatives in Bovine Urine by High-performance Liquid Chromatography and Off-line Chemiluminescent Immunochemical Detection', *J. Chromatogr.*, **360**, 449–451 (1986).

285. V.K. Boppana, K.-L.L. Fong, J.A. Ziemniak, R.K. Lynn, 'Use of a Post-column Immobilized β-Glucuronidase Enzyme Reactor for the Determination of Diastereomeric Glucuronides of Fenoldopam in Plasma and Urine by High-performance Liquid Chromatography with Electrochemical Detection', *J. Chromatogr.*, **353**, 231–247 (1986).

286. H.W. Frijlink, J. Visser, B.F.H. Drenth, 'Determination of Cyclodextrins in Biological Fluids by High-performance Liquid Chromatography with Negative Colorimetric Detection Using Post-column Complexation with Phenolphthalein', *J. Chromatogr.*, **415**, 325–333 (1987).

287. T. Kawamoto, E. Okada, T. Fujita, 'Post-column Derivatization of Vitamin B_6 Using 2,6-Dibromoquinone-4-chlorimide', *J. Chromatogr.*, **267**, 414–419 (1983).

288. G. Capitano, R. Tscherne, 'Separation and Quantitation of Esterified Estrogens in Bulk Mixtures and Combination Drug Preparations Using High-performance Liquid Chromatography', *J. Pharm. Sci.*, **68**, 311–313 (1979).

289. R.W. Roos, T. Medwick, 'Use of Borohydride Reduction in the Separation of Estrogen Carbonyls', *J. Chromatogr. Sci.*, **17**, 624–627 (1979).

290. R.W. Roos, T. Medwick, 'Application of Dansyl Derivatization to the High Pressure Liquid Chromatographic Identification of Equine Estrogens', *J. Chromatogr. Sci.*, **18**, 626–630 (1980).

291. A. Salvador, C. Moretton, A. Piram, R. Faure, 'On-line Solid Phase Extraction with On-Support Derivatization for High Sensitivity Liquid Chromatography Tandem Mass Spectrometry of Estrogens in Influent/effluent of Wastewater Treatment', *J. Chromatogr. A*, **1145**, 102–109 (2007).

292. K. Shimada, M. Tanaka, T. Nambara, 'Derivatization of Ketosteroids for High-performance Liquid Chromatography with Electrochemical Detection', *Anal. Lett.*, **13**, 1129–1136 (1980).

293. E. Reh, G. Schwedt, 'Fluorometric Reaction Detectors for the HPLC Analysis of Steroids', *Fresenius J. Anal. Chem.*, **303**, 117–120 (1980).

294. G. Schwedt, E. Reh, 'Aufbau und Optimierung Fluorimetrischer Reaktionsdetektoren mit Luftsegmentierten Flussigkeitsstromen fur die Hockdruck-Flussigkeits-Chromatographie; Teil 1. Auswahl Chemischer Reaktionen am Beispiel von Steroiden', *Chromatographia*, **13**, 1779–1784 (1980).

295. G. Schwedt, E. Reh, 'Fluorimetric Reaction Detectors for High-performance Liquid Chromatography; Part 1. Selection of Chemical Reactions', *Chromatographia*, **14**, 249–253 (1981).

296. T. Kawasaki, M. Maeda, A. Tsuji, 'Determination of 17-Oxosteroids in Serum and Urine by Fluorescence High-performance Liquid Chromatography Using Dansyl Hydrazine as a Pre-labeling Reagent', *J. Chromatogr.*, **226**, 1–12 (1981).

297. C. Hamada, M. Iwasaki, N. Kuroda, Y. Ohkura, '3-Chloroformyl-7-methoxycoumarin as a Fluorescent Derivatization Reagent for Alcoholic Compounds in Liquid Chromatography and Its Use for the Assay of 17-Oxosteroids in Urine', *J. Chromatogr.*, **341**, 426–431 (1985).

298. J. Goto, N. Goto, F. Shamsa, M. Saito, S. Komatsu, K. Suzaki, T. Nambara, 'New Sensitive Derivatization of Hydroxysteroids for High-performance Liquid Chromatography with Fluorescence Detection', *Anal. Chim. Acta*, **147**, 397–400 (1983).

299. M. Maeda, A. Tsuji, 'Chemiluminescence with Lucigenin as Post-column Reagent in High-performance Liquid Chromatography of Corticosteroids and *p*-Nitrophenacyl Esters', *J. Chromatogr.*, **352**, 213–220 (1986).

300. T. Seki, Y. Yamaguchi, 'New Fluorimetric Detection Method of Corticosteroids After High-performance Liquid Chromatography Using Post-column Derivatization with Glycinamide', *J. Liq. Chromatogr.*, **6**, 1131–1138 (1983).

301. E. Kwong, K.M. McErlane, 'Development of a High-performance Liquid Chromatographic Assay for Digoxin Using Post-column Fluorogenic Derivatization', *J. Chromatogr.*, **377**, 233–242 (1986).

302. E. Kwong, K.M. McErlane, 'Analysis of Digoxin at Therapeutic Concentrations Using High-performance Liquid Chromatography with Post-column Derivatization', *J. Chromatogr.*, **381**, 357–363 (1986).

303. T.A. Shepard, J. Hui, A. Chandrasekaran, R.A. Sams, R.H. Reuning, L.W. Robertson, J.H. Caldwell, R.L. Donnerberg, 'Digoxin and Metabolites in Urine and Feces: A Fluorescence Derivatization–High-performance Liquid Chromatographic Technique', *J. Chromatogr.*, **380**, 89–98 (1986).

304. P. Jakobsen, S. Waldorff, 'Determination of Digoxin, Digoxigenin and Dihydrodigoxigenin in Urine by Extraction, Derivatization and High-performance Liquid Chromatography', *J. Chromatogr.*, **382**, 349–354 (1986).

305. B. Desta, 'Separation of Digoxin from Dihydrodigoxin and the Other Metabolites by High-performance Liquid Chromatography with Post-column Derivatization', *J. Chromatogr.*, **421**, 381–386 (1987).

306. M.S. Gandelman, J.W. Birks, U.A.T. Brinkman, R.W. Frei, 'Liquid Chromatographic Detection of Cardiac Glycosides and Saccharides Based on the Photoreduction of Anthraquinone-2,6-disulfonate', *J. Chromatogr.*, **282**, 193–209 (1983).

307. E.P. Lankmayr, K.W. Budna, K. Mueller, F. Nachtmann, 'Determination of Thiol Compounds by Selective Fluorescence Derivatization and HPLC', *Fresenius J. Anal. Chem.*, **295**, 371–374 (1979).

308. R.C. Fahey, G.L. Newton, R. Dorian, E.M. Kosower, 'Analysis of Biological Thiols: Quantitative Determination of Thiols at the Picomole Level Based Upon Derivatization with Monobromobimanes and Separation by Cation-exchange Chromatography', *Anal. Biochem.*, **111**, 357–365 (1981).

309. R.M. Maiorino, G.L. Weber, H.V. Aposhian, 'Fluorimetric Determination of 2,3-Dimercaptopropane-1-sulfonic Acid and Other Dithiols by Precolumn Derivatization with Bromobimane and Column Liquid Chromatography', *J. Chromatogr.*, **374**, 297–310 (1986).

310. B. Kagedal, T. Andersson, M. Carlsson, T. Denneberg, A. Hoppe, 'High-performance Liquid Chromatography of 2-Mercaptoproprionylglycine and Its Metabolite 2-Mercaptoproprionic Acid in Plasma and Urine After Treatment with Thiopronine', *J. Chromatogr.*, **417**, 261–267 (1987).

311. R. Shimizu, T. Kakimoto, K. Ishii, Y. Fujimoto, H. Nishi, N. Tsumagari, 'New Derivatization Reagent for the Resolution of Optical Isomers in Diltiazem Hydrochloride by High-performance Liquid Chromatography', *J. Chromatogr.*, **357**, 119–125 (1986).

312. K. Shimoda, M. Tanaka, T. Nambara, 'Sensitive Derivatization Reagents for Thiol Compounds in High-performance Liquid Chromatography with Electrochemical Detection', *Anal. Chim. Acta*, **147**, 375–380 (1983).

313. V. Cavrini, R. Gatti, A.M. DiPietra, M.A. Raggi, 'HPLC Determination of Thiol Drugs in Pharmaceutical Formulations Using Ethacrynic Acid as a Precolumn Ultraviolet Derivatization Reagent', *Chromatographia*, **23**, 680 (1987).

314. W.T. Kok, J.J. Halvax, W.H. Voogt, U.A.T. Brinkman, R.W. Frei, 'Detection of Thioethers of Pharmaceutical Importance by Liquid Chromatography with On-line Generated Bromine', *Anal. Chem.*, **57**, 2580–2583 (1985).

315. W.T. Kok, W.H. Voogt, U.A.T. Brinkman, R.W. Frei, 'On-line Electrochemical Reagent Production for Fluorescence Detection of Phenothiazines in Liquid Chromatography', *J. Chromatogr.*, **354**, 249–257 (1986).

316. U.A.T. Brinkman, P.L.M. Welling, G. DeVries, A.H. M.T. Scholten, R.W. Frei, 'Liquid Chromatography of Demoxepam and Phenothiazines Using a Post-column Photochemical Reactor and Fluorescence Detection', *J. Chromatogr.*, **217**, 463–471 (1981).

317. K. Shimada, T. Mino, M. Nakajima, H. Wakabayashi, S. Yamato, 'Application of the Desulfurization of Phenothiazines for a Sensitive Detection Method by High Performance Liquid Chromatography', *J. Chromatogr. B*, **661**, 85–91 (1994).

318. A.H.M.T. Scholten, U.A.T. Brinkman, R.W. Frei, 'Photochemical Reaction Detectors in Continuous-flow Systems – Applications to Pharmaceuticals', *Anal. Chim. Acta*, **114**, 137–146 (1980).

319. A.H.M.T. Scholten, P.L.M. Welling, U.A.T. Brinkman, R.W. Frei, 'Use of PTFE in Post-column Photochemical Reactors for Liquid Chromatography – Application to Pharmaceuticals', *J. Chromatogr.*, **199**, 239–248 (1980).

320. M. Johansson, D. Westerlund, 'Determination of *N*-Acetylcysteine, Intact and Oxidized, in Plasma by Column Liquid Chromatography and Post-column Derivatization', *J. Chromatogr.*, **385**, 343–356 (1987).

321. H. Irth, G.J. De Jong, U.A.T. Brinkman, R.W. Frei, 'Determination of Disulfiram and Two of Its Metabolites in Urine by Reversed-phase Liquid Chromatography and Spectrophotometric Detection After Post-column Complexation', *J. Chromatogr.*, **424**, 95–102 (1988).

322. P. van Zoonen, D.A. Kamminga, C. Gooijer, N.H. Velthorst, R.W. Frei, 'Quenched Peroxyoxalate Chemiluminescence as a New Detection Principle in Flow Injection Analysis and Liquid Chromatography', *Anal. Chem.*, **58**, 1245–1248 (1986).

323. K. Funakoshi, K. Yamashita, W. Chao, M. Yamaguchi, T. Yashiki, 'High-performance Liquid Chromatographic Determination of Busulfan in Human Serum with On-line Derivatization, Column Switching and Ultraviolet Absorbance Detection', *J. Chromatogr. B*, **660**, 200–204 (1994).

324. A. Jenke, U. Renner, U.S. Schuler, S. Wauer, T. Leopold, E. Schleyer, G. Ehninger, 'Improved Assay for Determination of Busulfan by Liquid Chromatography using Postcolumn photolysis', *J. Chromatogr. B*, **805**, 147–153 (2004).

325. A. Shafaati, C. Lucy, 'Application of Capillary Zone Electrophoresis with Indirect UV Detection to the Determination of a Model Drug, Vigabatrin, in Dosage Forms', *J. Pharm. Biomed. Anal.*, **8**, 190–198 (2005).

326. V. Pucci, R. Mandrolli, M.A. Raggi, 'Determination of Valproic Acid (2-Propylpentanoic acid) in Human Plasma by Capillary Electrophoresis with Indirect UV Detection', *Electrophoresis*, **24**, 2076–2083 (2003).

327. R. Sekar, S. Azhaguvel, 'Indirect Photometric Assay Determination of Gabapentin in Bulk Drug and Capsules by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **36**, 553–667 (2004).

328. X.H. Yang, X.C. Wang, X.M. Zhang, 'Indirect Laser-Induced Fluorescence Detection of Diuretics Separated by Capillary Electrophoresis', *J. Sep. Sci.*, **29**, 677–683 (2006).

329. I.S. Krull, Z. Deyl, H. Lingeman, 'General Strategies and Selection of Derivatization Reactions for Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr. B*, **659**, 1–17 (1994).

330. H.A. Bardelmeijer, H. Lingeman, C. De Ruiter, W.J.M. Underberg, 'Derivatization in Capillary Electrophoresis', *J. Chromatogr. A*, **807**, 3–26 (1998).

331. R. Zhu, W.T. Kok, 'Post-column Derivatization for Fluorescence and Chemiluminescence Detection in Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **17**, 985–999 (1998).

332. A.M. Garcia-Campana, L. Gamiz-Gracia, W.R.G. Baeyens, F.A. Barrero, 'Derivatization of Biomolecules for Chemiluminescence Detection in Capillary Electrophoresis', *J. Chromatogr. B*, **793**, 49–74 (2003).

333. A. Takashi, 'Chiral Separation of Pharmaceuticals Possessing a Carboxy Moiety', *J. Chromatogr. B*, **717**, 295–311 (1998).

334. F. Couderc, E. Causse, C. Bayle, 'Drug Analysis by Capillary Eletrophoresis and Laser Induced Fluorescence', *Electrophoresis*, **19**, 2777–2790 (1998).

335. V.M. Okun, O.V. Aak, V.Y. Kozlov, 'Determination of the Anticancer Drug Prospidin in Human Tissue by High Performance Capillary Electrophoresis Using Derivatization', *J. Chromatogr. B*, **675**, 313–319 (1996).

336. A.L. Revilla, J. Hamacek, P. Lubal, J. Havel, 'Determination of Rimantadine in Pharmaceutical Preparations by Capillary Zone Electrophoresis with Indirect Detection or After Derivatization', *Chromatographia*, **47**, 433–439 (1998).

337. J. Liu, T.T. Dabrah, J.A. Matson, S.E. Klohr, K.J. Volk, E.H. Kerns, M.S. Lee, 'Analysis of Amino Acid Enantiomers Derived from Antitumor Antibiotics Using Chiral Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **16**, 207–214 (1997).

338. B.D. Wu, Q.Q. Wang, L. Guo, R. Shen, J.W. Xie, L.H. Yun, B.H. Zhong, 'Amino Substituted Beta-Cyclodextrin Copper(II) Complexes for the Electrophoretic Enantioseparation of Dansyl Amino Acids: Role of Dual Chelate-Inclusion Interaction and Mechanism', *Anal. Chim. Acta*, **558**, 80–85 (2006).

339. R. Gotti, M.G. Gioia, R. Gatti, V. Cavrini, 'Phanquinone as a Suitable Derivatizing Agent in Micellar Electrokinetic Chromatography and HPLC Analysis of Amino Acids', *J. Sep. Sci.*, **29**, 1259–1267 (2006).

340. F.M. Lin, H.S. Kou, S.M. Wu, S.H. Chen, A.L. Kwan, H.L. Wu, 'An Ionizable Chromophoric Reagent for the Analysis of Primary Amine-Containing Drugs by Capillary Electrophoresis', *Electrophoresis*, **26**, 621–626 (2005).

341. F.M. Lin, H.S. Kou, S.M. Wu, S.H. Chen, H.L. Wu, 'Capillary Electrophoresis Analysis of Gabapentin and Vigabatrin in Pharmaceutical Preparations as Ofloxacin Derivatives', *Anal. Chim. Acta*, **523**, 9–14 (2004).

342. H. Fonge, E. Kaale, C. Govaerts, K. Desmet, A. Van Schepdael, J. Hoogmartens, 'Bioanalysis of Tobramycin for Therapeutic Drug Monitoring by Solid-Phase Extraction and Capillary Zone Electrophoresis', *J. Chromatogr. B*, **810**, 313–318 (2004).

343. Y.Q. Cheng, L.Y. Fan, H.L. Chen, X.G. Chen, Z.D. Hu, 'Method for On-line Derivatization and Separation of Aspartic Acid Enantiomers in Pharmaceutical Applications by the Coupling of Flow Injection with Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **1072**, 259–265 (2005).

344. Z. Aturki, F. Vichi, A. Messina, M. Sinibaldi, 'Indirect Resolution of Beta-Blocker Agents by Reversed Phase Capillary Electrochromatography', *Electrophoresis*, **25**, 607–614 (2004).

345. C.F. Silverado, M. Azad, F.A. Gomez, 'On-Column Derivatization of the Antibiotics Teicoplanin and Ristocetin Coupled to Affinity Capillary Electrophoresis', *Electrophoresis*, **24**, 808–815 (2003).

346. S. Oguri, Y. Miki, 'Determination of Amikacin in Human Plasma by High Performance Capillary Electrophoresis with Fluorescence Detection', *J. Chromatogr. B*, **686**, 205–210 (1996).

347. H. Lingeman, W.J.M. Underberg, A. Takadate, A. Hulshof, 'Fluorescence Detection in High Performance Liquid Chromatography', *J. Liq. Chromatogr.*, **8**, 789–874 (1985).

348. A.J.G. Mank, H. Lingeman, C. Gooijer, 'Semi-conductor Laser Induced Fluorescence Detection in Capillary Electrophoresis Using Pre-column Thiol Derivatization', *J. High Resolut. Chromatogr.*, **17**, 797–798 (1994).

349. T. Perez-Ruiz, C. Martinez-Lozano, A. Sanz, R. Galera, 'An Ultrasensitive Method for the Determination of Thiouracil and Phenylthiouracil using Capillary Zone Electrophoresis and Laser – Induced Fluorescence Detection', *Electrophoresis*, **26**, 2384–2390 (2005).

350. E. Sbrana, E. Bramanti, M.C. Spinetti, G. Raspi, 'S-Adenosyl Methionine/S-Adenosyl-L-Homocysteine Ratio Determination by Capillary Electrophoresis Employed as a Monitoring Tool for the Antiviral Effectiveness of Adenosine Analogs', *Electrophoresis*, **25**, 1518–1521 (2004).

351. C.Z. Huang, T. Santa, K. Okabe, K. Imai, 'Capillary Electrophoresis with Laser Induced Fluorescence Detection of Profens Derivatized with the Water Soluble Fluorogenic Reagent 4-N-(4-N′-Aminoethyl)piperazino-7-nitro-2,1,3-benzoxadiazole', *J. Chromatogr. A*, **1011**, 193–201 (2003).

352. A. Ben Hameda, P. Taborsky, E.M. Pena-Mendez, J. Havel, 'Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI TOF MS) Study of Huperzine A, a Natural Anti-Alzheimer's Disease Product, its Degradation and its Detection by Highly Sensitive Laser Induced Fluorescence (LIF)', *Talanta*, **72**, 780–784 (2007).

353. A. Alnajjar, J.A. Butcher, B. McCord, 'Determination of Multiple Drugs of Abuse in Human Serum using Capillary Electrophoresis with Fluorescence Detection', *Electrophoresis*, **25**, 1592–1600 (2004).

354. A. Ramseier, F. Von Heeren, W. Thormann, 'Analysis of Fluorescein Isothiocyanate Derivatized Amphetamine and Analogs in Human Urine by Capillary Electrophoresis in Chip-based and Fused-silica Capillary Instrumentation', *Electrophoresis*, **19**, 2967–2975 (1998).

355. I. Rethfeld, G. Blaschke, 'Analysis of the Antidiabetic Drug Acarbose by Capillary Electrophoresis', *J. Chromatogr. B*, **700**, 249–253 (1997).

356. S. Kamoda, C. Nomura, M. Kinoshita, S. Nishiura, R. Ishikawa, K. Kakehi, N. Kawasaki, T. Hayakawa, 'Profiling Analysis of Oligosaccharides in Antibody Pharmaceuticals by Capillary Electrophoresis', *J. Chromatogr. A*, **1050**, 211–216 (2004).

357. C. Campa, A. Coslovi, A. Flamigni, M. Rossi, 'Overview on Advances in Capillary Electrophoresis-Mass Spectrometry of Carbohydrates: A Tabulated Review'', *Electrophoresis*, **27**, 2027–2050 (2006).

358. G. Hunt, W. Nashabeh, 'Capillary Electrophoresis Sodium Dodecyl Sulfate Non-gel Sieving Analysis of a Therapeutic Recombinant Monoclonal Antibody: A Biotechnology Perspective', *Anal. Chem.*, **71**, 2390–2397 (1999).

359. M. Foldvari, S. Attah-Poku, J. Hu, Q. Li, H. Hughes, L.A. Babiuk, S. Kruger, 'Palmitoyl Derivatives of Interferon α: Potential for Cutaneous Delivery', *J. Pharm. Sci.*, **87**, 1203–1208 (1998).

360. Y. Xuan, G. Weber, A. Manz, 'Construction and Analytical Application of an On-Column Photo-Reactor for Improved Detection of Iron Species as Plant Metabolites in Capillary Flow Injection and Capillary Electrophoresis', *J. Chromatogr. A*, **1130**, 212–218 (2006).

361. X. Wang, D.R. Bobbitt, 'In Situ Cell for Electrochemically Generated $Ru(bpy)_3^{3+}$ Based Chemiluminescence Detection in Capillary Electrophoresis', *Anal. Chim. Acta*, **383**, 213–220 (1999).

Chiral Purity in Drug Analysis

Yoshio Machida and Hiroyuki Nishi
Tanabe Seiyaku, Co. Ltd, Osaka, Japan

1 Introduction 1
2 Production of Chiral Drugs 2
2.1 History 2
2.2 Simulated Moving Beds Technology 3
2.3 Application of Simulated Moving Beds Technology to Chiral Drug Substances 4
3 Regulation of Chiral Drugs 4
3.1 International Conference on Harmonization 4
3.2 Impurities in New Drug Substances and Products 5
3.3 Guidance for Chiral Active Substances 5
4 Evaluation Methods for Chiral Drugs 5
4.1 Gas Chromatography 7
4.2 High-performance Liquid Chromatography 7
4.3 Capillary Electrophoresis 12
5 Conclusion 18
Abbreviations and Acronyms 18
Related Articles 18
References 19

The preparation and evaluation of optically active compounds have become very important for the development of new drugs. The determination of chiral purity of drugs is essential in order to guarantee safety and effectiveness. Isomeric impurities may have unwanted toxicological, pharmacological, or other side-effects. Therefore we decided to prepare an article entitled ''Chiral purity in drug analysis''. This article is divided into five sections (introduction, production, regulation, evaluation methods and conclusion). In the introduction, the importance of assessing the isomeric impurities in drug analysis will be discussed. In the second section, production methods for optically active compounds, except the stereoselective synthesis method, will be summarized briefly. In particular, new trends in the production of chiral drugs, such as the simulated moving beds (SMB) method, will be described. In the third section, regulatory perspectives in the USA (Food and Drug Administration (FDA)), EC, and Japan will be described and some regulatory guidelines and guidances by, for example, the FDA, the Committee for Veterinary Medicinal Products (CVMP), and the International Conference on Harmonization (ICH) will be mentioned. In the fourth section, methods for the determination of stereoisomeric drugs will be summarized. Methods such as gas chromatography (GC), high-performance liquid chromatography (HPLC), and capillary electrophoresis (CE) have been summarized. The focus will be on direct evaluation methods, i.e. those using chiral stationary phases (CSPs). Practical details for testing a drug's optical purity are also given.

1 INTRODUCTION

The stereoisomers of a molecule have the same constitution in both molecular formula and atomic bonding, but differ in the spatial orientation of the atoms or groups of atoms within the molecule. Enantiomers are stereoisomers with the same molecular formula, but which differ in the spatial arrangement of atoms within the molecule. They are nonsuperimposable mirror images, and quite often differ in their pharmacological activity, either qualitatively or quantitatively. Usually, only one of the isomers fully contributes to the therapeutic action, and the other is often classified as ''isomeric ballast''.[1] In the racemic mixtures that are used the usefulness of kinetics is generally questionable. Exposure of patients to the ''isomeric ballast'' present in about 50% of the most commonly used drugs will probably remain for many decades. However, approximately 50% of all drugs in therapeutic use have a chiral center, and hence about 25% of chiral drugs have been marketed as the racemate,[2,3] because the choice between single stereoisomers (homochiral drugs) and composite chiral drugs (mixtures of stereoisomers) depends upon therapeutic advantages (such as a reduction in xenobiotic load), possible adverse side-effects, and development costs. Therefore, the preparation of enantiopure compounds has become very important in many branches of chemistry, and the evaluation of enantiopure compounds is required during the development of new drugs. Moreover, determination of the chiral purity of drugs is essential in order to guarantee safety and effectiveness. Isomeric impurities may have unwanted toxicological, pharmacological, or other side-effects. One enantiomer sometimes shows high toxicity while another is effective. Some examples of activity differences between enantiomers are noted in Table 1. As can be seen from the table, the production of enantiopure compounds, regulation of enantiopure compounds and evaluation methods for chiral drugs are very important in the development of new chirally active substances and products.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Activity differences between enantiomers

Compounds	Activities of each enantiomer	Ref.
Acebutolol	Acebutolol, a β-blocking drug, is used in the treatment of hypertension and is marketed as a racemate. (*S*)-(−)-acebutolol is the eutomer.	4
Amphetamine	Amphetamine, an inhibitor of dopamine, is used clinically as a racemate, but (*S*)-(+)-amphetamine is four times as effective as its antipode.	5
Flurbiprofen	Both enantiomers possess analgesic activity and the drug is marketed as the racemate. (*S*)-(+)-flurbiprofen shows greater antiinflammatory activity.	6
Ketamine	The anesthetic potency of (*S*)-(+)-ketamine is found to be three times higher than that of (*R*)-(−)-ketamine.	7
Ketoprofen	Ketoprofen, a nonsteroidal anti-inflammatory drug, is marketed as the racemate, but only the *S*-enantiomer is pharmacologically active.	8
Levamisole	Levamisole is the (*S*)-(−)-isomer of tetramisole, and has effective anthelmintic activity; whereas the (*R*)-(+)-isomer, dexamisol, causes vomiting.	9
Naproxen	Except for naproxen, chiral nonsteroidal antiinflammatory drugs are marketed for clinical use as the racemate. However, in vitro studies have shown that the antiinflammatory activity exists almost solely in the (*S*)-enantiomer.	10
Propranolol	Propranolol, a β-adrenoceptor blocking drug, is administered as a racemate, but only (*S*)-(−)-propranolol possesses effective activity.	11
Propoxyphene	α-Dextropropoxyphene is marketed as an analgesic drug, whereas α-levopropoxyphene has no analgesic activity but is marketed as an antitussive agent. β-Dextro and β-levopropoxyphene are substantially inactive.	12
Salbutamol	Salbutamol, a β_2-adrenoceptor stimulant, is used clinically as the racemate. The activity resides predominantly in the (*R*)-(−)-enantiomer with little or no activity attributed to the (*S*)-(+)-enantiomer.	13
Sotalol	D-(+)-Sotalol possesses antiarrhythmic efficacy, whereas L-(−)-sotalol possesses β-adrenoceptor blocking activity.	14
Timolol	(*S*)-(−)-Timolol is one of the few β-adrenoceptor blocking drugs marketed as the pure enantiomer, whereas (*R*)-(+)-timolol is considerably less potent a β-adrenoceptor antagonist than (*S*)-(−)-timolol.	15
Verapamil	Verapamil is a calcium channel blocker with antiarrhythmic, antianginal, and antihypertensive properties as a racemate, but (−)-verapamil has more negative dromotropic activity than (+)-verapamil.	16
Warfarin	Warfarin, is a blood anticoagulant used clinically as a racemate. (*S*)-(−)-warfarin is five times as active as (*R*)-(+)-warfarin.	17

In this paper, production methods for optically active compounds, especially new trends, such as the SMB method are described. Regulatory perspectives which exist in the USA, EU and Japan are outlined, and some regulatory guidelines and guidances by, for example, the FDA and ICH are mentioned. Methods for the determination of stereoisomeric drugs, such as GC, HPLC, and CE, are summarized. The focus will be on the direct evaluation methods, i.e. those using CSPs. Practical details for testing the purity of a drug's enantiomers are also demonstrated.

2 PRODUCTION OF CHIRAL DRUGS

2.1 History

The need for separating enantiomers for pharmaceutical applications is increasing, as regulatory aspects become more stringent. It is well known that isomers can have different therapeutic values and therefore enantiomer separation is an important issue, particularly in health-related fields. During the past few decades there have been three principal methods for producing enantiopure compounds. The first method involves numerous synthetic strategies to prepare enantiopure building blocks, or chiral auxiliary reagents.[18,19] This approach is based on the development of optically active catalysts, which transfer their chiral information during the creation step of new centers of chirality. The second method involves the separation of enantiomers from racemic mixtures. One classical separation method for the preparation of enantiopure compounds is the recrystallization of diastereomeric salts.[20] However, this is not suitable for industrial scale-up and automation.[21] A third method uses conventional chromatographic techniques for the separation of enantiomers. This method is also difficult for scale-up because separation factors are low. Recently, much attention has been paid to continuous chromatography, particularly SMB technology because of its improved efficiency. Continuous chromatography in SMB eliminates the drawbacks of batch chromatography; namely, dilution of species and low utilization of

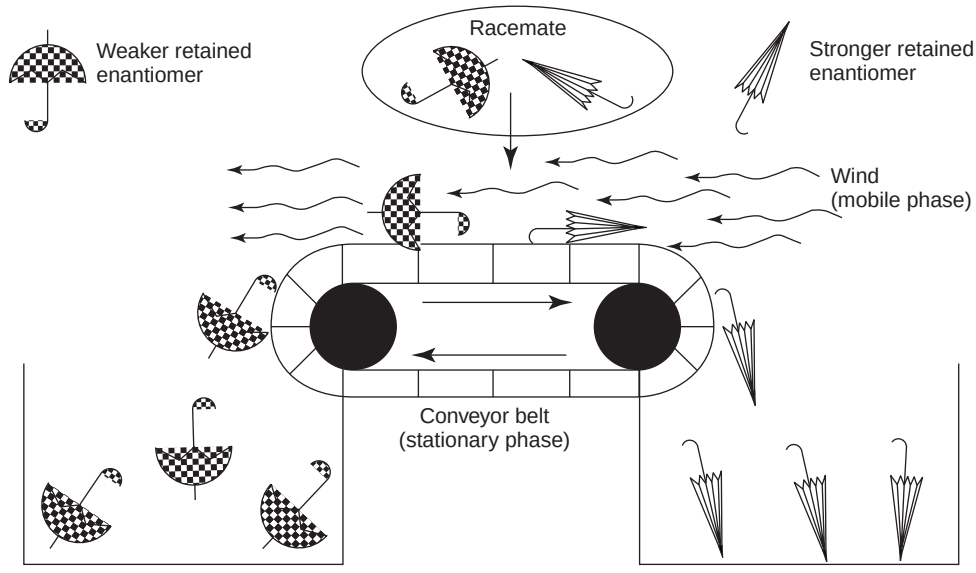

Figure 1 Concept of the simulated moving-bed adsorber system for optical resolution. (Reproduced from Negawa and Shoji,[25] with permission from Elsevier Science.)

adsorbents. As a production method for enantiopure compounds, SMB is described in more detail in section 2.2.

2.2 Simulated Moving Beds Technology

The concept of SMB technology has been known since 1961, when the first patent by Broughton from Universal Oil Products (UOP, Des Plaines, IL, USA) described the separation of structurally related hydrocarbons.[22–24] The principle of SMB technology is a rotary valve, which periodically changes the position of feed, eluent, extract and raffinate lines along the bed. Negawa & Shoji[25] reported the concept of the SMB system, as shown in Figure 1; that is, a situation where the stationary phase is moving, but in the real equipment the feed line, desorbent feed line, raffinate line and extract lines move intermittently.

It has been shown[26–29] that this can be accomplished in units consisting of a set of fixed bed chromatography columns arranged in a similar manner to that illustrated in Figure 2 (boxes 1–12), where the periodic movement of inlet and outlet ports is in the same direction as the fluid flow. In Figure 2 the binary mixture (A + B) to be separated constitutes the feed stream to the unit. The more adsorbable component, A, is collected in the extract stream, whereas the less adsorbable one, B, is collected in the raffinate stream. Each section of the unit plays a specific role in the operation. Section 1 (boxes 1–5) lies between the desorbent inlet port and the extract port; section 2 (box 6) lies between the extract port and the feed inlet port; section 3 (boxes 7–9) lies between the feed inlet port and the raffinate outlet port; section 4

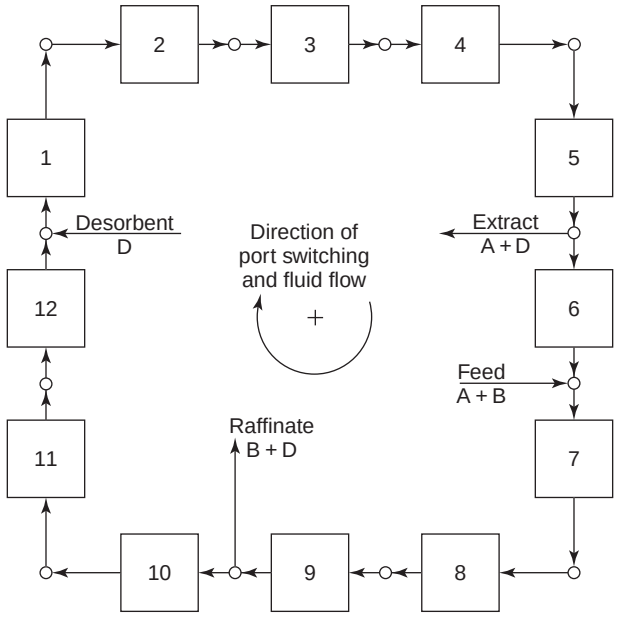

Figure 2 Scheme of a four-section simulated moving bed unit for continuous adsorptive separations with port distribution 5-1-3-3. (Reproduced from Mazzotti et al.,[27] with permission from Elsevier Science.)

(boxes 10–12) lies between the raffinate outlet port and the desorbent inlet port. Separation is performed in the two central sections. Component B is carried by the mobile phase, while component A is retained by the stationary phase. The eluent (D), is used to desorb component A from the first section, in order to regenerate the adsorbent. Finally, component B is adsorbed in the

Table 2 Application of SMB technology to the separation of drug enantiomers

Compound	Pharmaceutical	CSPs used	Extract (e.e.)	Raffinate (e.e.)	Refs.
Agonist at muscarinic receptor	Agonist at muscarinic receptor	Chiralpak AD	97.8	99.5	31
EMD 53986	The precursor of EMD53998	Chiralpak AD	>99	N.R.	32
EMD 53998	Ca-sensitizing drug	ChiraSpher	98.5	80.4	33
Enflurance	Inhalation anesthetic	γ-CD derivative	90 18	12 96	28
Ester of quinoline mevalonic acid (DOLE)	Pharmaceutical intermediate of a cholesterol reducing agent for hyperlipidemia	Chiracel OF	92.0	98.0	34
Foradil (formoterol)	β_2-Adrenoceptor agonist	Chiralcel OJ	98.8	99.2	29, 35
Guaifenesin	Antitussive agent	Chiralcel OD	97.8	100	29, 35
Chiral epoxide	Chiral epoxide used in pharmaceutical development	Chiralcel OD	98 85	91 97	36
Chiral epoxide (intermediate)	Intermediate in enantioselective synthesis	Chiralcel OD	98.2	96.5	37
Morphanthridine	α_2-Adrenoceptor antagonist	Chiralpak AD	N.R.	N.R.	37
Mouse IgG_1	Monoclonal antibody against penicillin amidase	Sepharose 4 FF	99	N.R.	39
Orimeten (aminoglutethimide)	Aromatase inhibitor	Chiralcel OJ	99.6	99.8	29, 35
Praziquantel	Used in the treatment of parasitic disease schistosomiasis	Cellulose triacetate	90.0	93.7	40
Propranolol	β-Adrenoceptor blocking drug	Chiralcel OD	100	100	41
Tramadol	A centrally acting analgesic	Chiralpak AD	99	>99.8	42

e.e., enantiomeric excess; N.R., not reported; CD, cyclodextrin; Sepharose 4 FF, cyanogen bromide-activated gel filtration type (Pharmacia, Uppsala, Sweden).

fourth section of the unit, to regenerate the desorbent itself.

2.3 Application of Simulated Moving Beds Technology to Chiral Drug Substances

SMB technology is useful for separating large amounts of compounds as it is a continuous separation process. It has found new applications in the areas of biotechnology, pharmaceuticals and fine chemistry. Pilot and industrial SMB for such applications have been developed by UOP and Separex.[30] Scaling down of the Sorbex flowsheet becomes less economical than using a system of individual beds segmented by valves and feed and product lines. Therefore, there is a growing demand in the pharmaceutical industry for efficient and cost effective methods for purifying optical isomers. More recent applications are related to chiral technology.[31–42] More than 98.0% enantiomeric excess has been obtained by using polysaccharide-type CSPs. One of these compounds, NK-104 (Nissan Chemical Industries Ltd., Tokyo, Japan), which is a synthetic hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase inhibitor, will be the first enantiopure drug produced by SMB technology in the near future.[34] Examples of enantiopure drugs produced by using SMB are summarized in Table 2. Recently, new applications of SMB chiral separation by using the GC[43] and supercritical fluid chromatography[44] have been published. In elution chromatography, high purity and high recovery cannot be achieved with low resolution. However, SMB chromatography, as a manufacturing process step, is a cost-effective method for producing chiral drugs.

3 REGULATION OF CHIRAL DRUGS

3.1 International Conference on Harmonization

In recent years, the regulatory authorities and industry associations set-up to promote international harmonization of regulatory requirements have undertaken many important initiatives. Regulatory guidance on the nature and extent of investigations into the properties of chirally active substances is required, in order that the production and marketing of compounds with adverse effects can be avoided. One of the goals of harmonization is to identify and reduce differences in the technical requirements for drug development among regulatory agencies.[45]

Six groups (Table 3), which represent the regulatory bodies and research-based industries in the EU, Japan and the USA, organize the ICH, and discuss harmonization of the technical requirements for registration of pharmaceutical products in these regions. Most new drugs and medicines are developed in the EU, Japan and the USA, and therefore when the technical requirements are agreed, new drugs can be developed and marketed without delay in these three regions. Moreover, the ICH steering committee includes representatives from each ICH sponsor and the IFPMA, as well as observers

Table 3 The founder members of the ICH

Region	Regulatory body	Research-based industry
Europe	EC (European Commission)	EFPIA (European Federation of Pharmaceutical Industries' Associations)
USA	FDA	PhRMA (Pharmaceutical Research and Manufacturers of America)
Japan	MHW (Ministry of Health and Welfare, Japan)	JPMA (Japan Pharmaceutical Manufacturers Association)
Observers	WHO Representatives of EFTA Canada Health Protection Branch	
ICH steering committee	IFPMA	

WHO, World Health Organization; EFTA, European Free Trade Association; IFPMA, International Federation of Pharmaceutical Manufacturers Association.

from the WHO, the Canada Health Protection Branch, and the EFTA. The ICH is a joint initiative, involving regulators and industry as equal partners in the scientific and technical discussions of the testing procedures required to ensure and assess the safety, quality and efficacy of medicines. Regulatory perspectives of the ICH are explained in detail in some regulatory guidelines and guidance.

3.2 Impurities in New Drug Substances and Products

The nature and concentration of impurities in new drug substances produced by chemical synthesis are regulated by the ICH guidelines "*Impurities in New Drug Substances*" (Topics Q3A)[46] and "*Impurities in New Drug Products*" (Topics Q3B),[47] which deal with chemical and safety aspects. Classification and identification of impurities are included in chemical aspects. Threshold limits for drug impurities should depend on each individual drug and be based on scientific rationale. The structures of impurities that occur at concentrations of 0.1%, or greater have to be determined. New drugs used for preclinical and/or clinical studies can have impurities. If the proposed concentration of an impurity is not specified, the safety of that impurity has to be studied according to the threshold limits given in Table 4. Although impurities in new drug substances and products are regulated, the guidelines do not apply to enantiomeric impurities in new chiral drugs, even though enantiomeric impurities may be regarded as conventional impurities.

Table 4 The qualification threshold

Maximum daily dose	Qualification threshold
Not more than 2 g	0.1%, or 1 mg per day, intake (whichever is lower)
More than 2 g	0.05%

3.3 Guidance for Chiral Active Substances

The FDA's policy statement for the development of new stereoisomeric drugs was issued in 1992.[48] Later, a *Note for Guidance on the Investigation of Chiral Active Substances* was approved in June 1997 by the CVMP. This "Note" is valid from January 1, 1998.[49] According to the "Note", the identity and stereoisomeric purity of the active substance, its optical rotation (OR), melting point, and characteristics when undergoing liquid chromatography using a CSP have to be studied.

However, chiral impurities of new optically active drug substances are excluded from the ICH Guidelines[46,47] because of practical difficulties in quantifying them and identifying the given thresholds. Chiral impurities in chiral new drug substances and new drug products should be treated according to a decision tree, as outlined in Figure 3, and the ICH Guidelines "*Test for new drug substances that are optically active*" (Table 5) (Topics Q6A).[50] At present, the concentration of the antipode, that is, the chiral impurity of chiral drugs, is not regulated in Topics Q6A, but, ideally, the antipode should be regulated as an impurity. As described in Table 4, if the toxicity of the antipode is unclear, methods have to be available which will detect 0.05% chiral impurity.

4 EVALUATION METHODS FOR CHIRAL DRUGS

Over the last two decades, chiral analytical technology and chiral purity in drug analysis have acquired increasing prominence, because various aspects of pharmaceutical activity, such as pharmacology, drug quality control, toxicology, require these analysis. For these purposes, powerful and widely applicable analytical tools are needed to determine the enantiomeric purity of chiral drugs. Chromatographic methods using GC, HPLC, and electrophoretic methods such as CE have been developed.

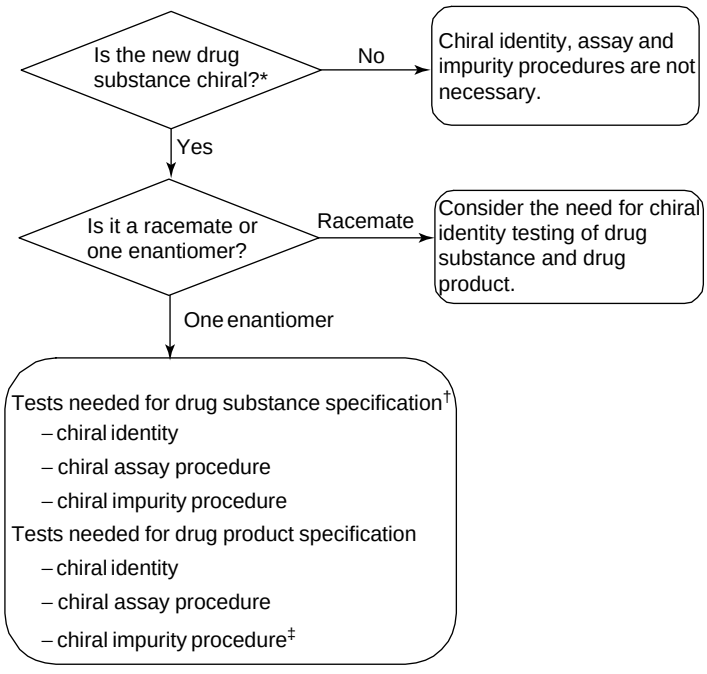

* Chiral substances of natural origin are not addressed in this Guidance.

† As with other impurities arising in and from raw materials used in the synthesis of drug substances control of chiral quality could be established alternatively by applying limits to appropriate starting material or intermediates when justified from development studies. This will be essentially the case when there are multiple chiral centers (e.g. three or more), or when measurement in drug substance at an appropriate sensitivity is technically impossible.

‡ A chiral assay may be necessary if chiral impurity testing is done via a method which is equivalent to an assay procedure.

Figure 3 Decision tree to establish the identity, assay and chiral impurities procedures in a new drug product containing a chiral drug substance.

Table 5 Tests for new drug substances and products that are optically active

	Drug substance	Drug product
Impurity or degradation product	Impurities: for chiral drug substances that are developed as a single enantiomer, control of the other enantiomer should be considered in the same manner as for other impurities. However, technical limitations may preclude the same limits of determination or qualification being applied. If it is technically difficult to effect control in the drug substance itself, assurance of control could be given by appropriate testing of a starting material or intermediate, with suitable justification.	Degradation products: control of the other enantiomer in a drug product is necessary if that enantiomer has been shown to be a degradation product.
Assay	An enantioselective determination of the drug substance should be part of the specification. It is considered acceptable for this to be achieved either through use of a chiral assay procedure or by the combination of an achiral assay together with appropriate methods of controlling the enantiomeric impurity.	Where development studies have demonstrated that the enantiomer is not a degradation product, an achiral assay may be sufficient. However, a chiral assay is preferred or, alternatively, the combination of an achiral assay plus a procedure to control the presence of the opposite enantiomer.
Identity	The identity test(s) should be capable of distinguishing a single enantiomer from its opposite enantiomer. Where a drug substance is a racemate, the identity method should be capable of verifying the racemic nature and distinguishing it from either enantiomer.	An identity test should be established that is capable of verifying the presence of the correct enantiomer or the racemate, as appropriate.

In principle, there are three methods of separating enantiomers: (1) by using enantiopure derivatizing reagents to convert enantiomers to diastereomers, which can then be separated on an achiral column; (2) by using chiral mobile phases (except GC); and (3) by using CSPs.

4.1 Gas Chromatography

The first CSP for separating enantiomers by GC was synthesized in 1977 by coupling polysiloxane and L-valine-*tert*-butylamide, which is now available commercially as Chirasil-Val (Chrompack, Middelburg, The Netherlands). Separation of 11 mixtures of alcohol enantiomers, as isopropyl urethane derivatives, by Chirasil-Val is shown in Figure 4. The separation was completed within 40 min.[51] Separation of the enantiomers of sulfur compounds[52] was also successful within 30 min. Saeed et al.[53] and König et al.[54] synthesized chiral polysiloxane attached to either L-valine-*tert*-butylamide or L-valine-(*S* or *R*)-phenylethylamide as a chiral moiety. Fused-silica capillary columns coated with various CDs linked to the polymer have also been developed. Chirasil-Dex, where the β-CD was linked to the polymer through only one 6-*O*-octamethylene spacing group, was introduced by Jung & Schurig,[55] and is now available commercially. A summary of recently developed, commercially available CSPs is given in Table 6. The advantages of GC for the separation of enantiomers are the high efficiency, sensitivity, and speed of chiral separation. The disadvantages are that the volatility, thermal stability and resolvability of the analyte limit the usefulness of the technique.[56] However, GC is useful for separating enantiomers of nonaromatic ring compounds, which cannot be detected by the systems used with HPLC.

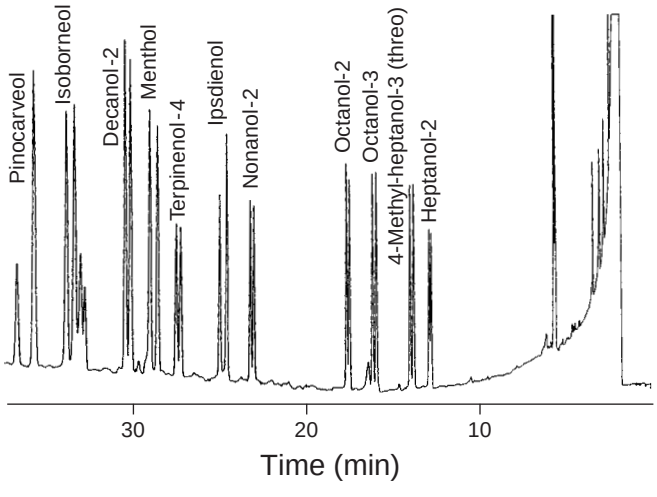

Figure 4 Enantiomer separation of a mixture of alcohols as isopropyl urethane derivatives on Chirasil-Val. (Reproduced from König,[51] with permission from John Wiley & Sons Limited.)

4.2 High-performance Liquid Chromatography

4.2.1 Chiral Derivatization

Approximately 30% of chiral compounds have been analyzed by indirect methods, because they could not be separated by using CSPs. The indirect method is based on the reaction of a racemic mixture with a chiral reagent to form a pair of diastereomers, which possess different physicochemical properties and hence can be separated by an achiral column. The acid chloride of (*S*)-(−)-*N*-trifluoroacetylproline is a widely used derivatizing agent[57] and has been used for resolving racemic amines and alcohols by normal-phase chromatography. (*S*)-(−)-*N*-1-(2-naphthylsulfonyl)-2-pyrrolidine carbonyl chloride (NSP-Cl),[58] a derivatizing reagent for primary or secondary amines and alcohols, has been successfully used for determining the enantiomeric purity of diltiazem hydrochloride (a Ca^{2+} channel blocker),[59] and enprostil (an anti-ulcer drug). The derivatized enantiomers were separated by normal-phase chromatography. 2,3,4,6-Tetra-*O*-acetyl-β-D-glucopyranosyl isothiocyanate (GITC), which was developed by Kinoshita et al.[60] and Nimura et al.[61] as a derivatizing reagent for an amino group, has also been successfully used in reversed-phase HPLC for determining the enantiomeric purity of denopamine,[62] etilefrine, phenylephrine and trimetoquinol hydrochloride. 1-(9-Fluorenyl)ethyl chloroformate (FLEC) has been widely used for the diastereomeric derivatization of primary or secondary amines and alcohols.[63] For carboxylic acids, or their derivatives, of biological interest, separation of enantiomers can be achieved by using reversed-phase HPLC with an ODS column, after derivatization with alanine-β-naphthylamide. This method was used by Kagawa et al.[64] to determine the enantiomeric purity of acetyl-L-carnitine, which may be important as it can stimulate actions of learning behavior and can be used as a drug. Typical chromatograms are shown in Figure 5.

4.2.2 Chiral Stationary Phases

Direct methods using CSPs have been used extensively because CSPs based on amino acid derivatives,[65–68] cellulose,[69–71] CDs,[72–74] ovomucoid,[75,76] bovine serum albumin,[77] crown ethers,[78,79] and antibiotics (Figure 6)[80–85] are now available commercially (Table 7).

Recently, in the development of new chiral drugs, most determinations of chiral purity in drug analysis have been performed by using CSPs. For example, in our laboratory, direct separation of the enantiomers of diltiazem hydrochloride, which was previously achieved by derivatization with NSP-Cl,[58] was achieved by using

Table 6 Commercially available CSPs used in GC for chiral separation

Type of CSP	Name	Chiral moiety	Supplier
CD	CHIRALDEX A-DA	α-CD, dialkyl-	Astec
	CHIRALDEX A-PH	α-CD, hydroxypropyl-	Astec
	CHIRALDEX A-TA	α-CD, trifluoroacetyl-	Astec
	CHIRALDEX B-DA	β-CD, dialkyl-	Astec
	CHIRALDEX B-DM	β-CD, dimethyl-	Astec
	CHIRALDEX B-PH	β-CD, hydroxypropyl-	Astec
	CHIRALDEX B-PM	β-CD, permethylated	Astec
	CHIRALDEX B-TA	β-CD, trifluoroacetyl-	Astec
	CHIRALDEX G-DA	γ-CD, dialkyl-	Astec
	CHIRALDEX G-BP	γ-CD, butyryl-	Astec
	CHIRALDEX G-PH	γ-CD, hydroxypropyl-	Astec
	CHIRALDEX G-PN	γ-CD, propionyl-	Astec
	CHIRALDEX G-TA	γ-CD, trifluoroacetyl-	Astec
	CP-Chirasil-Dex-CB	β-CD, permethylated	Chrompack
	CP-CD-β-2,3,6-M-19	β-CD, 2,3,6-tri-*O*-methylated	Chrompack
	Cyclodex-B	β-CD, permethylated	J&W
	CycloSil-B	30% 2,3-Di-*O*-methyl-6-*O*-TBDMS-β-CD embedded in DB-1701	J&W
	HP-Chiral-10B	10% Permethylated β-CD	HP
	HP-Chiral-20B	20% Permethylated β-CD	HP
	Rt-βDEXcst	Proprietary CD material doped into CPDP-14	Restek
	Rt-βDEXm	Permethylated β-CD doped into CPDP-14	Restek
	Rt-βDEXsa	2,3-Di-acetoxy-6-*O*-TBDMS-β-CD doped into CPDP-14	Restek
	Rt-βDEXse	2,3-Di-ethyl-6-*O*-TBDMS-β-CD doped into CPDP-14	Restek
	Rt-βDEXsm	2,3-Di-methyl-6-*O*-TBDMS-β-CD doped into CPDP-14	Restek
	Rt-βDEXsp	2,3-Di-propyl-6-*O*-TBDMS-β-CD doped into CPDP-14	Restek
	Rt-γ DEXsa	2,3-Di-acetoxy-6-*O*-TBDMS-γ-CD doped into CPDP-14	Restek
	α-DEX 120	20% Permethylated α-CD in SPB-35	SUPELCO
	α-DEX 225	25% 2,3-Di-*O*-acetyl-6-*O*-TBDMS-α-CD embedded in SPB-20	SUPELCO
	α-DEX 325	25% 2,3-Di-*O*-methyl-6-*O*-TBDMS-α-CD embedded in SPB-20	SUPELCO
	β-DEX 110	10% Permethylated β-CD in SPB-35	SUPELCO
	β-DEX 120	20% Permethylated β-CD in SPB-35	SUPELCO
	β-DEX 225	25% 2,3-Di-*O*-acetyl-6-*O*-TBDMS-β-CD embedded in SPB-20	SUPELCO
	β-DEX 325	25% 2,3-Di-*O*-methyl-6-*O*-TBDMS-β-CD embedded in SPB-20	SUPELCO
	γ-DEX 120	20% Permethylated γ-CD in SPB-35	SUPELCO
	γ-DEX 225	25% 2,3-Di-*O*-acetyl-6-*O*-TBDMS-γ-CD embedded in SPB-20	SUPELCO
	γ-DEX 325	25% 2,3-Di-*O*-methyl-6-*O*-TBDMS-γ-CD embedded in SPB-20	SUPELCO
π-Acid and π-base phases	CP-Chirasil-D-Val	D-Valine-*tert*-butylamide modified methyl silicone	Chrompack
	CP-Chirasil-L-Val	L-Valine-*tert*-butylamide modified methyl silicone	Chrompack
	G-column G-800	L-Valine-*tert*-butylamide modified methyl silicone	CITI

Astec, Advanced Separation Technologies, Whippany, NJ, USA; Chrompack, Chrompack International, Middelburg, The Netherlands; CITI, Chemicals Inspection & Testing Institute, Tokyo, Japan; HP, HEWLETT PACKARD, Palo Alto, CA, USA; J&W, J&W Scientific, Folsom, CA, USA; Restek, RESTEK CORP, Bellefonte, PA, USA; SUPELCO, SUPELCO Inc., Bellefonte, PA, USA.
SPB-20, 20% phenyl–80% dimethylsiloxane; SPB-35, 35% diphenyl–65% dimethylsiloxane; CPDP-14, 14% cyanopropyl/phenyl–86% dimethyl polysiloxane.

ovomucoid CSP, where protein is embedded on a silica-gel matrix.[86,87] With this method, good linearity between the peak area and the amount of sample used permitted determination of the enantiomeric purity of the antipode at levels down to about 0.1%. In 1994, direct separation of the enantiomers of imidapril hydrochloride, which previously required chiral derivatization, was achieved by using the CSP Chiralpak WH, a ligand-exchange type column where Cu^{2+} is used as a mobile phase additive.[88] With this method, good linearity between the peak area and the amount of sample used, combined with a signal-to-noise ratio of 3, gave a detection limit of about 0.2% for the antipode. In our laboratory approximately 70% of chiral separations of drugs have been achieved by using CSPs.

Dolezalova & Tkaczykova[89] reported the enantiomeric purity of levodopa and carbidopa (L-DOPA and L-CDOPA, respectively, both of which are used for the treatment of Parkinson's disease), and methyldopa (L-MDOPA, used as an antihypertensive) by using Chirobiotic T, in which teicoplanin is used as a chiral moiety. The detection limits of the antipode on L-DOPA, L-MDOPA and L-CDOPA were 0.03%, 0.3% and 0.4%, respectively. Torrens et al.[90] reported the enantiomeric

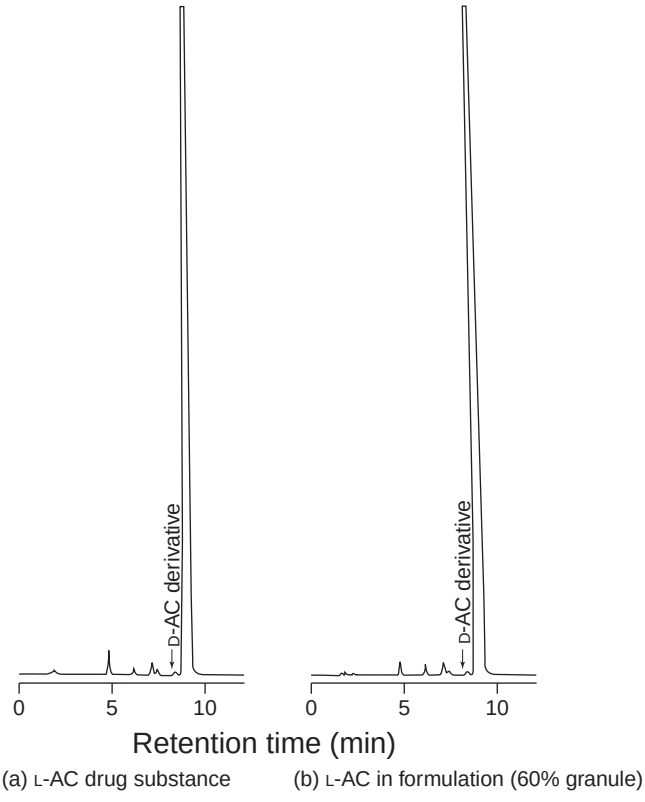

Figure 5 Enantiomeric purity testing of acetyl-L-carnitine. (Reproduced from Kagawa et al.,[64] with permission from Elsevier Science.)

purity of cizolirtine, a potent analgesic that is currently in phase II clinical trials, by three independent methods, namely, ^{1}H-NMR (nuclear magnetic resonance), CE (using β-CDs as chiral resolving agents) and HPLC (using a Chiral AGP column). Typical chromatograms are shown in Figure 7. The detection limits of the antipode obtained by the three methods were: 1% with ^{1}H-NMR, and 0.5% with either HPLC or CE.

Despite the numerous CSPs now available commercially, much research is still carried out to develop novel CSPs that have wider or more specialized enantioselectivity. Pirkle & Terfloth[91] developed a new CSP derived from *S*-naproxen as the chiral moiety. Haginaka et al.[92] investigated the separation of enantiomers on a CSP based on ovoglycoprotein and the effect of aggregation of ovoglycoprotein. 18-Crown-6-tetracarboxylic acid, which has a high enantioselectivity for primary amino compounds, has been immobilized to silica gel as the chiral selector.[93–95]

In addition to the development of CSPs, some applications for mobile phase additives have also been published. Separation of the enantiomers of hydrophobic amino compounds by using Crownpak CR(+), dynamically coated chiral crown ether as chiral moiety, has been studied.[96] Here, the enantiomeric hydrophobic amino compounds were easily eluted and successfully separated by the addition of β-CD (an inclusion reagent) and/or various salts. Typical chromatograms are shown in Figure 8.

Novel detectors for the stereospecific analysis of chiral drugs in plasma have been developed and investigated. In 1999, Kanazawa et al.[97] reported the separation of enantiomers of lorazepam, ibuprofen and flubiprofen on Chiralcel OD, using UV, circular dichroism and OR detection systems. Typical chromatograms of ibuprofen are shown in Figure 9. Chiroptical detection systems, such as circular dichroism and OR, are useful for pharmacokinetic studies of chiral drugs in biological samples, because the front peak and other peaks having no chirality are not detected by these methods.

For future CSPs, molecular imprinting (MI) technology is likely to be introduced. The technique of MI is shown schematically in Figure 10.[98,99] The selected analyte or

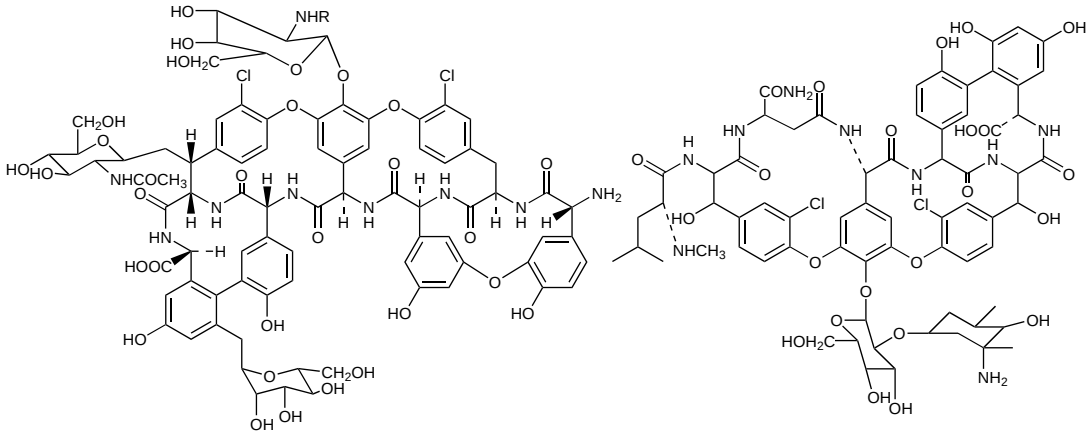

Figure 6 The structures of antibiotics used as chiral selectors.

Table 7 Commercially available CSPs used in liquid chromatography for chiral separation

Type of CSP	Name	Chiral moiety	Supplier
Amylose-based phases	CHIRALPAK AD/AD-RH	Amylose tris(3,5-dimethylphenyl carbamate)	Daicel
	CHIRALPAK AS	Amylose tris[(*S*)-phenylethyl carbamate]	Daicel
Crown ether	CROWNPAK CR (+ or −)	2,3 : 4,5-Bis[1,2-(3-phenyl)]-1,6,9,12,15,18-hexaoxacycloeicosa-2,4-diene	Daicel
CD	ChiraDex	β-CD	Merck
	ChiraDex GAMMA	γ-CD	Merck
	Chiral CD-Ph	β-CD phenylcarbamate	Shiseido
	Cyclobond I 2000	β-CD	Astec
	Cyclobond I 2000 Ac	β-CD acetylated	Astec
	Cyclobond I DMP	β-CD-3,5-dimethylphenyl carbamate	Astec
	Cyclobond I RN/SN	β-CD (*R* or *S*)-naphthylethyl carbamate	Astec
	Cyclobond I RSP	β-CD (*R*,*S*)-hydroxypropyl	Astec
	Cyclobond I SP	β-CD (*S*)-hydroxypropyl	Astec
	Cyclobond II	γ-CD	Astec
	Cyclobond II Ac	γ-CD acetylated	Astec
	Cyclobond III	α-CD	Astec
	Cyclobond III Ac	α-CD acetylated	Astec
	CYCLOSE 2-β-OH	β-CD	Chiralsep
	CYCLOSE 6-β-OH	β-CD	Chiralsep
	NUCLEODEX β-OH	β-CD	Nagel
	NUCLEODEX α-PM	α-CD permethylated	Nagel
	NUCLEODEX β-PM	β-CD permethylated	Nagel
	NUCLEODEX γ-PM	γ-CD permethylated	Nagel
	Orpak CDA-453 HQ	α-CD	Shodex
	Orpak CDB-453 HQ	β-CD	Shodex
	Orpak CDC-453 HQ	γ-CD	Shodex
	Orpak CDBS-453	β-CD	Shodex
	SUMICHIRAL OA-7000	β-CD	SCAS
	SUMICHIRAL OA-7100	β-CD	SCAS
	SUMICHIRAL OA-7500	β-CD permethylated	SCAS
	ULTRON ES-CD	β-CD	Shinwa
	ULTRON ES-PhCD	β-CD phenylcarbamate	Shinwa
	YMC CHIRAL α-CD BR	α-CD	YMC
	YMC CHIRAL β-CD BR	β-CD	YMC
	YMC CHIRAL γ-CD BR	γ-CD	YMC
Cellulose-based phases	Chiralcel OA	Cellulose triacetate (CTA II)	Daicel
	Chiralcel OB/OB-H	Cellulose tribenzoate	Daicel
	Chiralcel OC	Cellulose tris(phenylcarbamate)	Daicel
	Chiralcel OD/OD-H/OD-R/OD-RH	Cellulose tris(3,5-dimethylphenyl carbamate)	Daicel
	Chiralcel OF	Cellulose tris(4-chlorophenyl carbamate)	Daicel
	Chiralcel OG	Cellulose tris(4-methylphenyl carbamate)	Daicel
	Chiralcel OJ/OJ-R	Cellulose tris(4-methylbenzoate)	Daicel
	Chiralcel OK	Cellulose tricinnamate	Daicel
	Chiralcel CA-1	Cellulose triacetate	Daicel
	CHIRAL TRIACEL	Cellulose triacetate	Nagel
	CHIRAL TRIBENCEL	Cellulose tribenzoate	Nagel
	Cellulose triacetate	Cellulose triacetate	Merck
	CONBRIO-TAV	Cellulose triacetate	Perstorp Biolytica
Metal coordination	Ceramospher Chiral RU-1/RU-2	Δ-Ru(1,10-phenanthroline)$_3^{2+}$ in magnesium silicate sodium	Shiseido
Polyacrylamide phase	Chiraspher	Poly-[(*S*)-*N*-acryloylphenylalanine ethyl ester]	Merck
Protein	AF pak ABA-894	Bovine serum albumin	Shodex
	Bioptic AV-1/AV-2	Avidin	GLS
	CHIRAL AGP	α_1-Acid glycoprotein	Daicel
	CHIRAL AGP	α_1-Acid glycoprotein	Regis
	Chiral CBH	Cellobiohydrolase	Regis

Table 7 (*continued*)

Type of CSP	Name	Chiral moiety	Supplier
	Chiral HSA	Human serum albumin	Regis
	Chirobiotic T	Teicoplanin	Astec
	Chirobiotic V	Vancomycin	Astec
	RESOLVOSIL BSA-7/BSA-7PX	Bovine serum albumin	Nagel
	SUMICHIRAL AGP	α_1-Acid glycoprotein	SCAS
	SUMICHIRAL CBH	Cellobiohydrolase	SCAS
	SUMICHIRAL HSA	Human serum albumin	SCAS
	ULTRON ES-BSA	Bovine serum albumin	Shinwa
	ULTRON ES-OVM	Ovomucoid	Shinwa
	ULTRON ES-PEPSIN	Pepsin	Shinwa
π-Acid and π-base phases	Bakerbond DNBPG	*N*-(3,5-Dinitrobenzoyl)-phenylglycine	Baker
	Bakerbond DNBLeu	*N*-(3,5-Dinitrobenzoyl)-leucine	Baker
	(R)-α-Burke 2	Dimethyl *N*-3,5-dinitro-benzoyl-*a*-amino-2,2-dimethyl-4-pentenyl phosphonate	Regis
	CHIRALPAK OT(+)	Poly(triphenylmethylmethacrylate)	Daicel
	CHIRALPAK OP(+)	Poly(diphenyl-2-pyridylmethylmethacrylate)	Daicel
	ChiraSpher	Poly(*N*-acryloyl-L-phenylalanine ethyl ester)	Merck
	β-Gem 1 (R,R)- or (S,S)-	*N*-3,5-dinitrobenzoyl-3-amino-3-phenyl-2-(1,1-dimethylethyl)-propanoate	Regis
	Kromasil CHI-DMB	*O*,*O*′-Bis (3,5-dimethylbenzoyl)-*N*,*N*′-diallyl-L-tartar diamide	Eka
	Kromasil CHI-TBB	*O*,*O*′-Bis (4-*tert*-butylbenzoyl)-*N*,*N*′-diallyl-L-tartar diamide	Eka
	Leucine (L- or D-)	*N*-(3,5-Dinitrobenzoyl)-leucine (L- or D-)	Regis
	Naphthyl alanine (L- or D-)	*N*-(1-Naphthyl) alanine (L- or D-)	Regis
	L-Naphthylleucine	*N*-(1-Naphthyl) leucine	Regis
	NUCLEOSIL CHIRAL-2	*N*-(3,5-Dinitrobenzoyl)-D-phenylglycine	Nagel
	NUCLEOSIL CHIRAL-3	*N*-(3,5-Dinitrobenzoyl)-L-phenylglycine	Nagel
	Phenylglycine (L- or D-)	*N*-(3,5-Dinitrobenzoyl)-phenylglycine (L- or D-)	Regis
	Pirkle 1-J (3R,4S)- or (3S,4R)-	Unusual β-lactam	Regis
	Si100-DNB-Leu (L- or D-)	*N*-(3,5-Dinitrobenzoyl)-leucine (L- or D-)	Serva
	Si100-DNB-Phgly (L- or D-)	*N*-(3,5-Dinitrobenzoyl)-phenylglycine (L- or D-)	Serva
	SUMICHIRAL OA-1000	(*S*)-Naphthylethylamine	SCAS
	SUMICHIRAL OA-2000/2000I	(*R*)-Phenylglycine 3,5-dinitrobenzoylamide	SCAS
	SUMICHIRAL OA-2100	(*R*)-Phenylglycine, (*S*)-chlorophenylisovaleric acid	SCAS
	SUMICHIRAL OA-2200	(*R*)-Phenylglycine, (1*R*,3*R*)-chrysanthemic acid	SCAS
	SUMICHIRAL OA-2500/2500I	(*R* or *S*)-Naphthylglycine 3,5-dinitrobenzoylamide	SCAS
	SUMICHIRAL OA-3000	(*S*)-Valine *tert*-butylurea	SCAS
	SUMICHIRAL OA-3100	(*S*)-Valine 3,5-dinitrophenylurea	SCAS
	SUMICHIRAL OA-3200	(*S*)-*tert*-Leucine 3,5-dinitrophenylurea	SCAS
	SUMICHIRAL OA-3300	(*R*)-Phenylglycine 3,5-dinitrophenylurea	SCAS
	SUMICHIRAL OA-4000/4100	(*S*)-Valine, (*S* or *R*)-1-naphthylethylamine	SCAS
	SUMICHIRAL OA-4400/4500	(*S*)-Proline, (*S* or *R*)-1-naphthylethylamine	SCAS
	SUMICHIRAL OA-4600/4700	(*S*)-tert-Leucine, (*S* or *R*)-1-naphthylethylamine	SCAS
	SUMICHIRAL OA-4800/4900	(*S*)-Indoline-2-carboxylic acid, (*S* or *R*)-1-naphthylethylamine	SCAS
	SUPELCOSIL LC-(*R* or *S*)-Naphthylurea	(*R* or *S*)-Phenylurea	SUPELCO
	SUPELCOSIL LC-(*R* or *S*)-Phenylurea	(*R* or *S*)-Phenylurea	SUPELCO
	SUPELCOSIL LC-(*R* or *S*)-Dinitrobenzoylphenylglycine	(*R* or *S*)-Dinitrobenzoylphenylglycine	SUPELCO
	Spherisorb Chiral-1	(*R*)-Phenylethylamine urea	Phase Sep
	Spherisorb Chiral-2	(*S*)-Phenylethylamine urea	Phase Sep
	ULMO (*R*,*R*)- or (*S*,*S*)-	Pirkle type CSPs were developed by Austrian Researchers, Uray, Linder, and Maier	Regis
	Whelk-O1/O2 (*R*,*R*)- or (*S*,*S*)-	Polysaccharide derived	Regis
	YMC CHIRAL NEA (*R* or *S*)	(*R* or *S*)-Naphthylethylamine	YMC

(*continued overleaf*)

Table 7 (*continued*)

Type of CSP	Name	Chiral moiety	Supplier
Phases for ligand-exchange chromatography	CHIRALPAK WE	(1*R*, 2*S*)-1,2-Diphenyl-2-amino ethanol	Daicel
	CHIRALPAK WH	L-Proline	Daicel
	CHIRALPAK WM	L-*tert*-Leucine	Daicel
	CHIRALPAK MA(+)	Alanine derivative	Daicel
	Chirosolve L- or *O*-pip	L-Pipecolic acid	JPS Chimie
	Chirosolve L- or *O*-phe	L-Phenylalanine	JPS Chimie
	Chirosolve L- or *O*-val	L-Valine	JPS Chimie
	Chirosolve L- or *O*-hypro	L-Hydroxyproline	JPS Chimie
	Chirosolve L- or *O*-pro	L-Proline	JPS Chimie
	Chirosolve L- or *O*-porretine	L-Porretine	JPS Chimie
	Davankov Column	This CSP was developed by Professor V. Davankov	Regis
	NUCLEOSIL CHIRAL-1	L-Hydroxyproline	Nagel
	Orpak CDX-853	L-Amino acid derivative	Shodex
	Si100 Hypro-Cu	Hydroxyproline	Serva
	Si100 Pro-Cu	Proline	Serva
	Si100 Val-Cu	Valine	Serva
	SUMICHIRAL OA-5000	*N*-(*S*)-Dioctyl-D-penicillamine	SCAS
	SUMICHIRAL OA-5500	(*R*)-2-Amino-1,1-bis(2-butoxy-5-*tert*-butylphenyl)-phenylpropanol	SCAS
	SUMICHIRAL OA-6000	(*R*,*R*)-Tartaric acid mono-(*R*)-1-(α-naphthyl)ethylamide	SCAS
	SUMICHIRAL OA-6100	(*R*,*R*)-Tartaric acid mono-L-valine-(*S*)-1-(α-naphthyl)ethylamide	SCAS
	MCI GEL CRS10W (DLAA)	*N*,*N*-Dioctyl-L-alanine	MCI
	MCI GEL CRS15W (LDAA)	*N*,*N*-Dioctyl-D-alanine	MCI

Astec, Advanced Separation Technologies, Whippany, NJ, USA; Baker, J. T. Baker, Phillipsburg, NJ, USA; Chiralsep, Chiralsep, Parc de la Boissiere, La Frenaya, France; Daicel, Daicel Chemical Industries, Tokyo, Japan; Eka, Eka Chemicals, Bohus, Sweden; GLS, GL Sciences Inc., Tokyo, Japan; JPS Chimie, Bevaix, Switzerland; MCI, Mitsubishi Chemical, Tokyo, Japan; Merck, E. Merck, Darmstadt, Germany; Nagel, Macherey-Nagel GmbH & Co. KG, TUV, Rheinland Germany; Perstorp Biolytica, Perstorp Biolytica, Lund, Sweden; Phase Sep, Phase Separation, Clwyd, UK; Regis, Regis Chemical, Morton Grove, IL, USA; SCAS, Sumika Chemical Analysis Service, Ltd, Osaka, Japan; Serva, Serva Feinbiochemica, Heidelberg, Germany; Shinwa, Shinwa Kako, Kyoto, Japan; Shiseido, Shiseido, Tokyo, Japan; Shodex, Showa denko, Tokyo, Japan; SUPELCO, SUPELCO Inc., Bellefonte, PA, USA; YMC, YMC, Kyoto, Japan.

template molecules is first allowed to establish binding interactions with polymerizable chemical functionalities, that is, hydrogen bonds, ion pairing, $\pi-\pi$ interaction or hydrophobic-effect-driven reactions, and the resulting complexes or adducts are subsequently copolymerized with crosslinkers into a rigid polymer. After the template molecules are extracted, specific recognition sites are left in the polymer in which the spatial arrangement of the functional groups in the polymer network, together with the shape, are complementary to the imprinted molecule. Therefore, an MI polymer appears to have the potential for chiral recognition, and an analytical method using an MI polymer will be a powerful tool for chiral separation of drugs.

4.3 Capillary Electrophoresis

CE is another powerful method for separating enantiomers.[100–103] Relatively fast separation, with high resolution, can be achieved with one capillary tube, without a special chiral column or an organic solvent. Furthermore, CE has many attractive features, such as the speed of method development and analysis, robustness, simplicity and cost. The separation solution can easily be altered to find the optimum separation media, and an expensive chiral selector can be used because only small amounts of media are required. Among various different CE modes, both capillary zone electrophoresis (CZE) and electrokinetic chromatography (EKC), in which only a chiral selector is added to the usual running buffer solution, are most widely used for enantiomer separations.[100,101] EKC, where an ionic pseudo-stationary phase is employed, is one branch of the CE techniques and has the capability to separate electrically neutral analytes. Various ionic additives or aggregates, such as micelles, microemulsions, and vesicles have been used as a pseudo-stationary phase in EKC. Among them, EKC using micelles, i.e. micellar EKC (MEKC) has become most popular.[104,105] Normally, an ionic surfactant (e.g. sodium dodecyl sulfate, SDS; or bile salts) solution is used as a running buffer solution in MEKC.

Typically, for the enantiomer separation of ionic analytes, neutral chiral selectors are employed in the CZE mode. CDs and their various derivatives have been

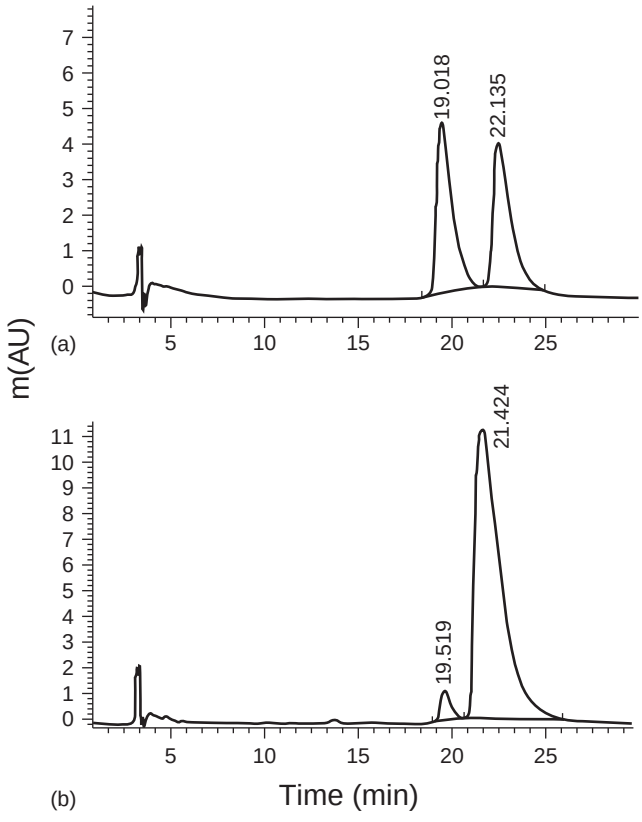

Figure 7 Enantiomer separation of cizolirtine using a Chiral AGP. (a) HPLC enantioseparation of racemate. (b) Determination of ~4% of S-form. (Reproduced from Torrens et al.,[90] with permission from John Wiley & Sons Limited.)

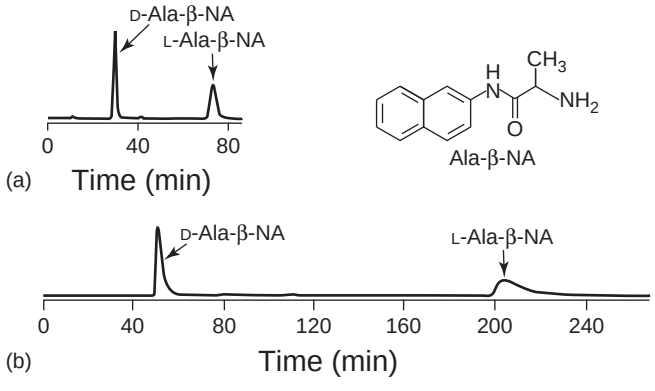

Figure 8 Enantiomer separation of alanine-β-naphthylamide (chiral derivatizing reagent) by the modified mobile phase. (Reproduced from Machida et al.,[96] with permission from Elsevier Science.) (a) Mobile phase: methanol–500 mM potassium chloride (15 : 85). (b) Mobile phase: methanol–10 mM perchloric acid (15 : 85).

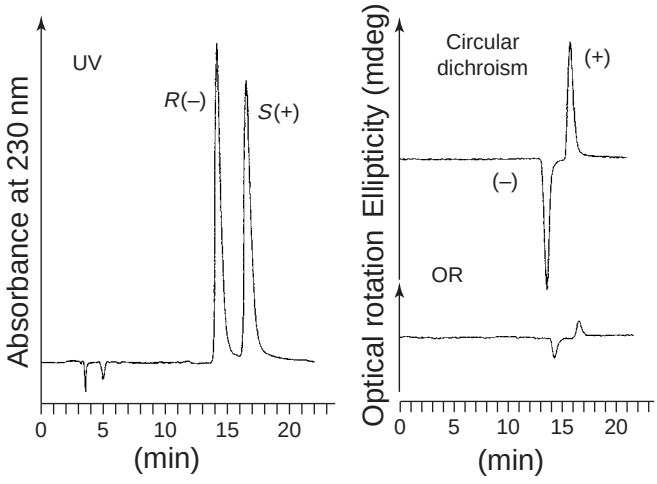

Figure 9 UV, circular dichroism and OR chromatograms of ibuprofen. (Reproduced from Kanazawa et al.,[97] with permission from the Society of Chromatographic Sciences.)

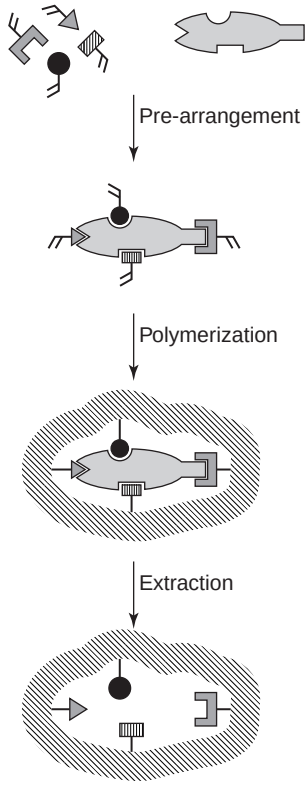

Figure 10 Schematic representation of imprint formation following the self-assembly approach. (Reproduced from Ramström and Ansell,[98] with permission from John Wiley & Sons Limited.)

found to be most effective in CZE for a wide range of drugs.[106,107] On the other hand, charged pseudostationary phases having chirality are added to the running buffer solution in EKC. Therefore, EKC enantiomer separation can be applied to both ionic and nonionic analytes. CDs are also effective chiral selectors in the MEKC mode.[101,104] This mode is called CD modified MEKC (CD/MEKC). CDs, polysaccharides, proteins,

crown ethers, and chiral surfactants, which are also effective chiral moieties in HPLC have been found to be useful in CE for the separation of enantiomers. Typical chiral selectors that have been successfully employed for the CE enantiomer separations are shown in Figure 11.

4.3.1 Use of Cyclodextrins and Crown Ethers

Among the various chiral selectors used in CE for separation of enantiomers, CDs have attained the greatest popularity, because they are capable of recognizing a wide range of enantiomers. When neutral CDs are used, the separation mode is called CD/CZE, and enantiomers of basic and acidic drugs can be separated. One example is shown in Figure 12, where ropivacaine and its enantiomer (R-form), and other related derivatives (enantiomers) have been successfully separated by CD/CZE. The running buffer was 0.1 M phosphoric acid, pH 3.0 (adjusted by triethanolamine), containing 10 mM 2,6-dimethyl (DM)-β-CD.[108] Detection was at 206 nm. The method also detected the required 0.1% of enantiomeric impurity, and proved to be robust. Some of these CE enantiomer purity testings have been accepted as a part of regulatory submissions or submitted to the FDA or the pharmacopoeias. Enantiomers of various ionic drugs have been successfully separated by this CD/CZE mode, using one of these CD derivatives.[106,107]

Recently, CE enantiomer separations by the combination of neutral CDs and other additives, such as surfactants (CD/MEKC mode),[109] charged CDs (dual CD system),[110,111] and crown ethers,[112] have been demonstrated. By combining CDs with another additive that can interact with the analytes, enantioselectivity has been much improved. For example, the enantiomers of 1-aminoindan, could not be separated by CZE with either 18-crown-6 or DM-β-CD, but were successfully separated by CZE using 5 mM 18-crown-6 and 5 mM DM-β-CD. Enantiomers of some other aminocompounds, such as 1,2,3,4-tetrahydro-1-naphthylamine, and 1-methyl-3-phenylpropylamine, have also been successfully separated by a dual system using CDs and crown ethers.[112]

Charged CDs have also been successfully used alone (CD/EKC) or with other additives, such as neutral CDs (dual CD system). Various anionic CD derivatives, such as sulfobutyl ether-, phosphated-, sulfated-, or carboxymethylated-, are now available commercially.[113] Positively charged CD derivatives, such as 2-hydroxypropyltrimethylammonium β-CD, heptamethylamino- and 6^A-methylamino-β-CD, have been synthesized and used in CE for separation of enantiomers.[114] Enantiomers of acidic pharmaceuticals that have a carboxylic group, such as nonsteroidal anti-inflammatory agents (e.g. ibuprofen, naproxen, ketoprofen) have been successfully separated by using these cationic CD derivatives, that is, by the CD/EKC mode.

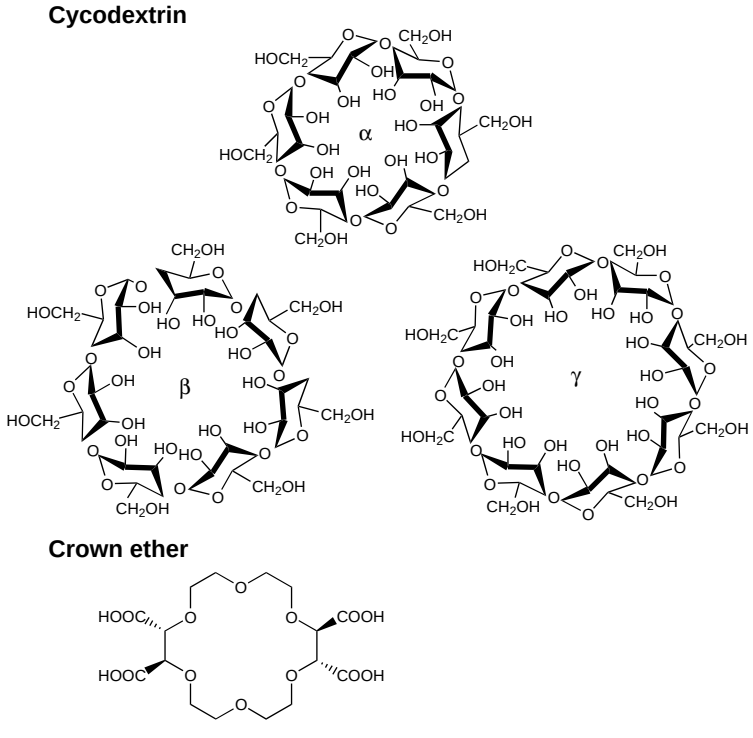

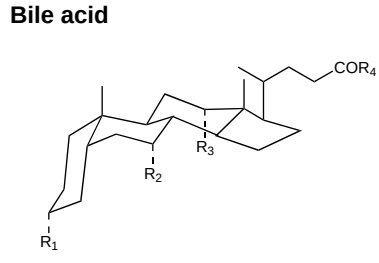

Bile salt	R_1	R_2	R_3	R_4
SC	OH	OH	OH	ONa
STC	OH	OH	OH	$NHCH_2CH_2SO_3Na$
SDC	OH	H	OH	ONa
STDC	OH	H	OH	$NHCH_2CH_2SO_3Na$

Figure 11 Typical chiral selectors used for CE enantiomer separations.

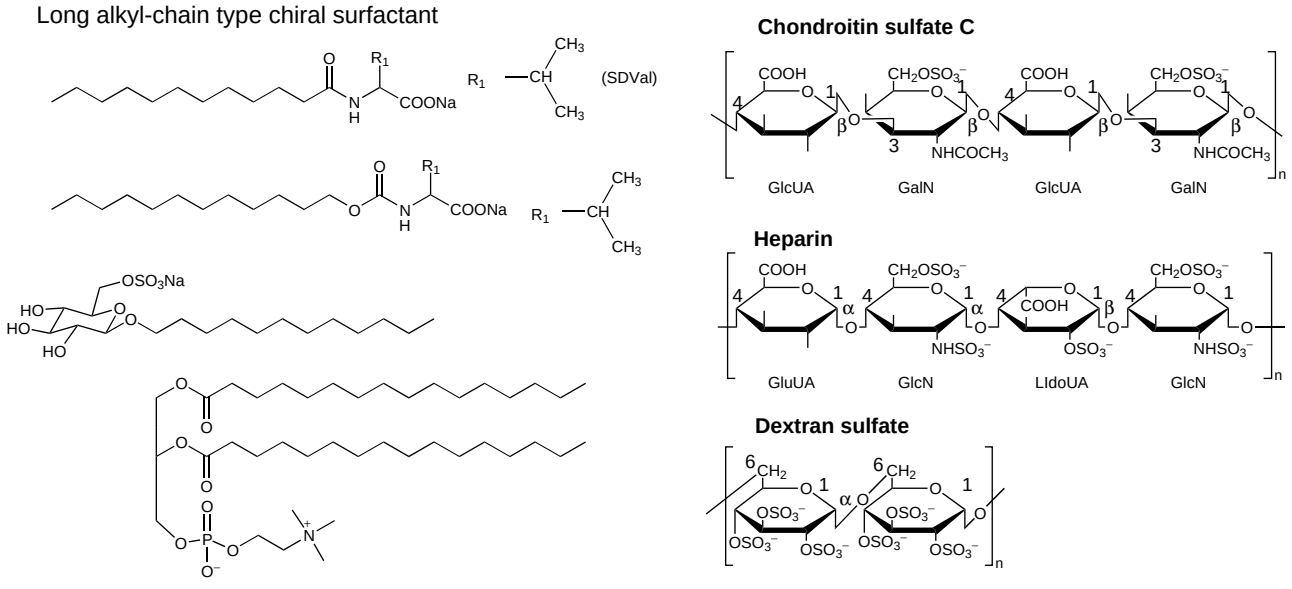

Proteins used for separation of enantiomers

Protein	Molecular mass	pI	Carbohydrate (%)
BSA	67 000	4.7	–
HSA	68 000	4.7	–
Fungal cellulase	60 000–70 000	3.9	6
Cellobiohydrolase I	60 000–70 000	3.9	6
α_1-AGP	44 000	2.7	45
Avidin	70 000	10	20.5
OVM	28 000	4.5	30
Conalbumin	77 000–87 000	6.0–6.8	–
Casein(α-casein)	26 200	4.7	–
Ovoglycoprotein	25 000	3.9	20

Figure 11 (*Continued*)

It is often difficult to control the position and the number of the substituents in these synthesized CD derivatives. The substitution distribution significantly influences the enantioselectivity. One example is the separation of enantiomers of denopamine and trimetoquinol by CZE with DM-β-CD.[115] Resolution values obtained from six different commercial DM-β-CD were different and showed that the degree of methyl substitution probably differed. This was confirmed by HPLC purity testing and matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI/TOFMS) analysis. To overcome these problems, single-isomer sulfated CDs have been developed recently.[116] Various enantiomers were successfully separated, as in the randomly derivatized charged CDs. Reproducibility in the enantioselectivity will be controlled by using these single-isomer derivatives.

The chiral crown ether, 18-crown-6-tetracarboxylic acid, has been shown to be effective for the separation of chiral amino drugs.[117] Compared with the results obtained by HPLC with a commercially available chiral crown ether column (Crownpak CR(+)), CE with 18-crown-6-tetracarboxylic acid was effective for the separation of enantiomers where the amine functional group and the stereogenic center is long. In a nonaqueous CE mode with formamide, enantiomer separation of 1-phenylethylamine, which has not been separated by any other mode of CE, was achieved by adding the tetra-*n*-butylammonium salt to the running buffer containing 18-crown-6-tetracarboxylic acid.[118] It is expected that various chiral crown ethers that can be applied in the CE system will be developed, because, at present, only 18-crown-6-tetracarboxylic acid is available commercially.

4.3.2 Use of Polysaccharides, Antibiotics and Proteins

Polysaccharides, antibiotics, and proteins have been successfully used in CE for separation of enantiomers. Polysaccharides, such as chondroitin sulfates, dextran

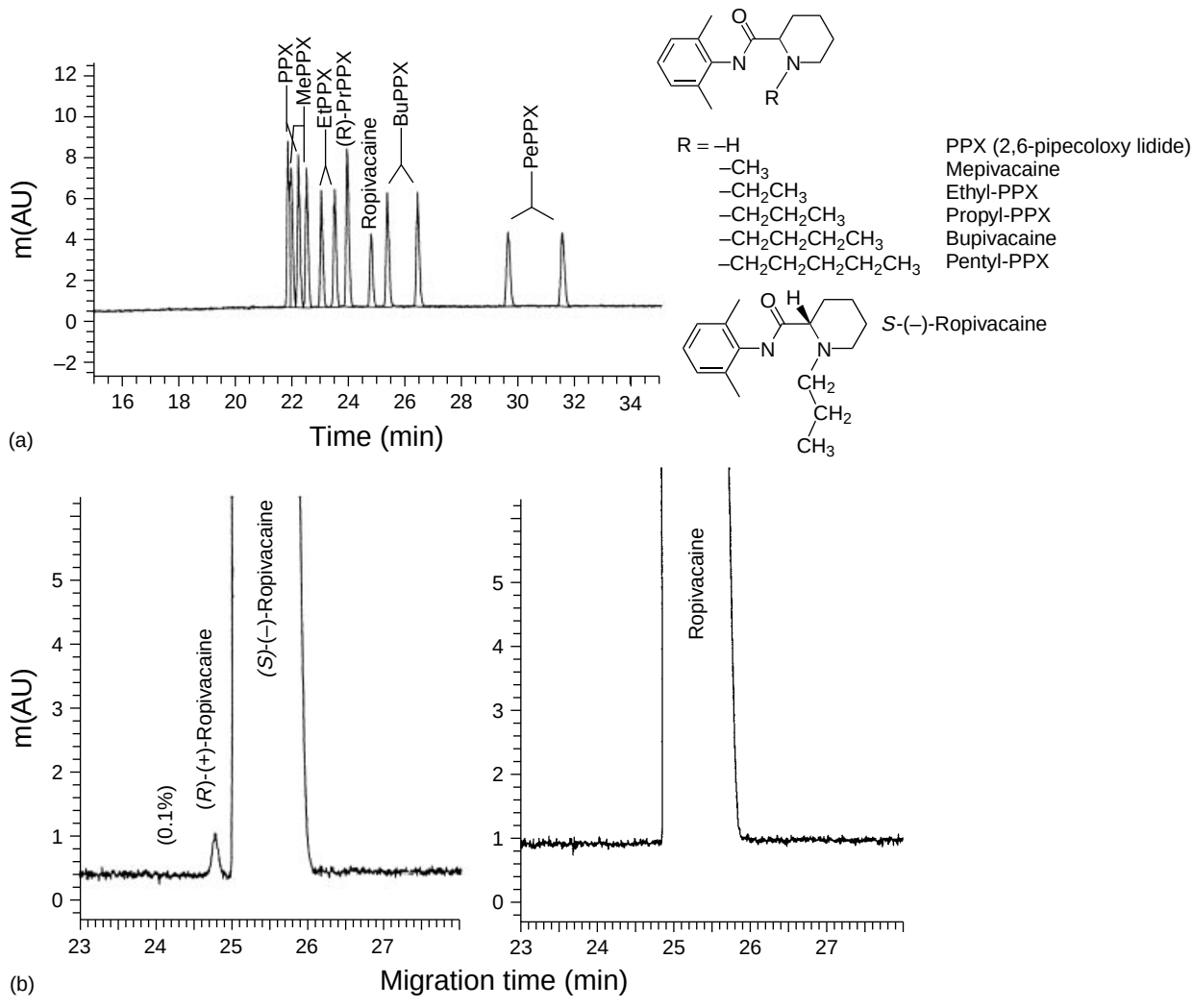

Figure 12 Enantiomer purity testing of ropivacaine (4*S*-form) by CZE using DM-β-CD. (a) Separation of related enantiomers including the ropivacaine minor enantiomer (4*R*-form); and (b) detection limit (0.1%) of the minor enantiomer. Applied voltage, 30 kV. Separation tube, 50 μm i.d. × 80.5 cm (effective length 72.0 cm). Detection, 206 nm. Temperature, 30 °C. (Reproduced from van der Griend and Groningsson,[108] with permission from Elsevier Science.)

sulfates, heparin, dextrans, dextrins, pentosan polysulfate, and dermatan sulfate, have been used as chiral selectors.[119–123] Dextran sulfate and pentosan polysulfate are semisynthetic sulfated polyanions. When charged polysaccharides are used as chiral selectors, the ionic character is important. In these polysaccharides, pentosan polysulfate has more ionizable groups per monomer unit (about two) than do chondroitin sulfates, dermatan sulfate (about one) and heparin, (one to two), and fewer than dextran sulfate (two to three). Enantiomers of tryptophan analogs[121] and basic pharmaceuticals[119–122] have been successfully separated by CE with ionic polysaccharides. One example is shown in Figure 13. Enantiomers of diltiazem hydrochloride (a Ca^{2+} channel blocker, in which the 2S,3S-form is active) and sulconazole (an antifungal, and a racemic drug) have been successfully separated by CZE using simply dextrin addition to a pH 2.5 phosphate buffer.[123]

Glycopeptide macrocyclic antibiotics such as vancomycin, teicoplanin, ristocetin A, and avoparcin, and some other antibiotics, such as rifamycins (B and SV), and aminoglycosides (kanamycin, fradiomycin, and streptomycin) appear to be among the most useful chiral selectors available for CE.[124,125] Recently, two different vancomycin analogs, namely A82846B and LY307599, and the teicoplanin family antibiotic MDL 63246 have been used to analyze negatively charged enantiomers.[124] High enantioselectivity has been shown by using these antibiotics, for example in HPLC with CSP moiety. However, some attention should be paid to the use of these

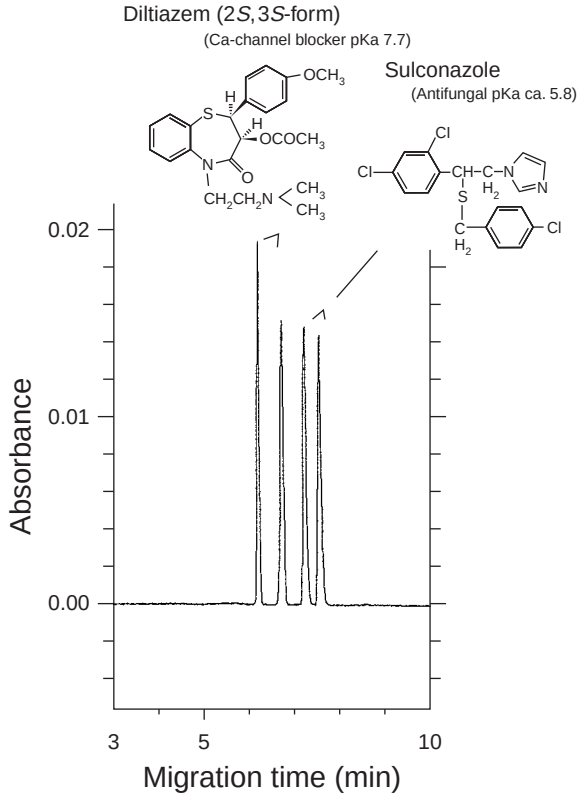

Figure 13 Enantiomer separations of diltiazem hydrochloride and sulconazole by CZE using 6% dextrin (JP grade) in 20 mM phosphate buffer, pH 2.5. Applied voltage, 30 kV. Separation tube, 75 µm i.d. × 47 cm (effective length 40 cm). Detection, 220 nm. Temperature, 20 °C. (Reproduced from Nishi et al.,[123] with permission from Vieweg Publishing.)

antibiotics in CE because their strong UV absorption affects the sensitivity of the technique.

Proteins such as bovine serum albumin, human serum albumin, α_1-acid glycoprotein, avidin, ovomucoid, casein, cellobiohydrolase, cellulase (fungal), transferrin (human serum), and pepsin, have also been found useful in CE for enantiomer separation.[126–128] Proteins have strong UV absorption in the typical UV detection wavelength range of 200–280 nm, therefore partial filling techniques have been used to avoid severe disturbances in UV detection.[127,128] This technique, where only a part of the capillary is filled with the selector, is also very useful in CE for enantiomer separations with antibiotics.

4.3.3 Micellar Electrokinetic Chromatography and Capillary Electrochromatography for Separation of Enantiomers

Chiral surfactants have been used for enantiomer separation in MEKC,[129,130] either when used on their own or in conjunction with other buffer additives such as CDs. Natural surfactants such as bile salts, digitonin, and saponins, and semi-synthetic surfactants derived from naturally occurring sugars[131,132] and amino acids[133,134] have been used. Recently polymerized surfactants have been used as chiral selectors in MEKC.[135,136] Compared with conventional micelles, which consist of a monomer surfactant, there are various advantages in polymer type micelles. The elimination of the dynamic equilibrium between the monomer and micelle may enhance chiral recognition. The critical micelle concentration of these polymer type micelles is essentially zero. This means that the polymer can be used over a wider range of concentrations. Organic modifiers can also be used without disrupting the formation of the micelle.

Separation by capillary electrochromatography (CEC) has recently attracted considerable interest due to the improved efficiency, as in conventional micro-HPLC. In addition, a wide variety of different packing materials developed for CEC are already available commercially.[137] The application of CEC has mainly focused on separations of electrically neutral drugs, because the separation mechanism of CEC is chromatography itself. Separation of enantiomers is another area where CEC holds tremendous potential. Capillaries packed with CDs as the immobilized stationary phases[138,139] have been used. A wide variety of CSPs such as polysaccharide derived,[140] naproxen derived, and Whelk-O,[141] have been prepared. MI has also been used in CEC to obtain capillaries with predetermined chiral selectivities.[142,143]

4.3.4 Regulatory Aspects

The acceptability of CE methods in drug analysis (including New Drug Applications (NDAs)) by regulatory authorities is one of the topics for pharmaceutical companies. Validations according to the ICH[144,145] or pharmacopoeia guidelines[146] are required in order that CE can be used as routine quality control methods by the pharmaceutical companies or when presenting NDAs. In modern, automated CE instruments, where a capillary is thermostated by an oven or liquid coolant, and sample injection and capillary washing are performed by computer, usually around 1% relative standard deviation (RSD) are obtained for the peak area ratios in the repeated injection (five or six injections). Accordingly, in the testing method of a drug, typically 2% or 3% values (corresponding to 2 RSD or 3 RSD values) are adopted as a criterion for the system suitability test (see below).

Appropriate CE data, meeting the requirements of ICH guidelines, have been included in a regulatory submission.[147,148] An assay that indicates the stability and uniformity of content of the cholesterol-lowering agent BMS-188494 by the MEKC method has been accepted as part of a regulatory submission to the

FDA.[147] A running buffer containing 0.1 M borate and 0.025 M SDS was used, and detection was carried out at 200 nm.

Pharmacopeia have also recognized the advancing application of CE within pharmaceutical companies. The savings in solvent purchase and disposal have been highlighted, particularly from an environmental viewpoint. Furthermore, a draft general chapter on CE has been published for the US Pharmacopeia (USP).[146] The first USP monographs describing CE were published in 1997.[149,150] These are chiral separation methods for epinephrine borate ophthalmic solution,[149] and a separate procedure for the analysis of ethambutol content in tablets.[150] In epinephrine borate injection, DM-β-CD has been used for separating the enantiomers (−)-epinephrine and (+)-epinephrine. Relative migration times, tailing factors of peaks, resolution and RSD values for replicate injections are specified in system suitability testing as follows. Both resolution values between the enantiomers, and the (−)-form and internal standards are not less than 1.5. Tailing factors for all peaks are not more than 2.5. An RSD of the ratio of (−)-epinephrine (active component) and (−)-pseudoepinephrine (internal standards) is not more than 3.0%. The relative migration times, resolution and RSD have been also specified in the assay of ethambutol. Resolution between the active component and internal standards is not less than 2. The RSD for the peak area ratio is not more than 2.0%. As a result of the ICH, groups representing the USP, European Pharmacopoeia and Japanese Pharmacopoeia have discussed the harmonization of general methods and monographs for drug substances. This has led to the publication of harmonized information concerning CE, as official inquiry stage draft.[151]

5 CONCLUSION

Recently, regulatory authorities have discussed the level of antipode and stereoselective analytical methods with industry associations. Some drafts such as ICH guidelines (Q3A, Q3B and Q6A) have been proposed. For chiral analysis and chiral purity determination in drug analysis, various CSPs have been developed and most are now available commercially. Further CE technology has also been used for chiral analysis because of its advantages. By using these technologies, a concentration of 0.1% of the minor enantiomer, can be easily determined, as required by regulatory authorities. In the near future, it is expected that more than 90% of enantiomers will be separated by using novel CSPs or chiral separation technologies. On the other hand, high sensitivity, nonseparated chiral purity determination methods using mass spectrometry (MS) or NMR technology may be introduced for drug analysis.

ABBREVIATIONS AND ACRONYMS

CD	Cyclodextrin
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CSP	Chiral Stationary Phase
CVMP	Committee for Veterinary Medicinal Products
CZE	Capillary Zone Electrophoresis
EFTA	European Free Trade Association
EKC	Electrokinetic Chromatography
FDA	Food and Drug Administration
FLEC	1-(9-Fluorenyl)ethyl Chloroformate
GC	Gas Chromatography
GITC	2,3,4,6-Tetra-*O*-acetyl-β-D-glucopyranosyl Isothiocyanate
HMG-CoA	Hydroxymethylglutaryl Coenzyme A
HPLC	High-performance Liquid Chromatography
ICH	International Conference on Harmonization
IFPMA	International Federation of Pharmaceutical Manufacturers Association
MALDI/TOFMS	Matrix-assisted Laser Desorption Ionization/Time-of-flight Mass Spectrometry
MEKC	Micellar Electrokinetic Chromatography
MI	Molecular Imprinting
MS	Mass Spectrometry
NDA	New Drug Application
NMR	Nuclear Magnetic Resonance
NSP-Cl	(*S*)-(−)-*N*-1-(2-Naphthylsulfonyl)-2-pyrrolidine Carbonyl Chloride
OR	Optical Rotation
RSD	Relative Standard Deviation
SMB	Simulated Moving Beds
USP	US Pharmacopeia
WHO	World Health Organization

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry • Drugs of Abuse, Analysis of

Food **(Volume 5)**
Electrophoresis and Isoelectric Focusing in Food Analysis • Liquid Chromatography in Food Analysis

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography

REFERENCES

1. E.J. Ariëns, 'Stereochemistry, a Basis for Sophisticated Nonsense in Pharmacokinetics and Clinical Pharmacology', *Eur. J. Clin. Pharmacol.*, **26**, 663–668 (1984).
2. E.J. Ariëns, E.W. Wuis, 'Bias in Pharmacokinetics and Clinical Pharmacology', *Clin. Pharmacol. Ther.*, **42**, 361–363 (1987).
3. E.J. Ariëns, E.W. Wuis, E.J. Veringa, 'Stereoselectivity of Bioactive Xenobiotics: A Pre-Pasteur Attitude in Medicinal Chemistry, Pharmacokinetics and Clinical Pharmacology', *Biochem. Pharmacol.*, **37**, 9–18 (1988).
4. M.M. Piquette, R.T. Foster, C.T. Kappagoda, F. Jamali, 'Pharmacokinetics of Acebutolol Enantiomers in Humans', *J. Pharm. Sci.*, **80**, 313–316 (1991).
5. M. Simonyi, I. Fitos, J. Visy, 'Chirality of Bioactive Agents in Protein Binding Storage and Transport Processes', *Trends Pharmacol. Sci.*, **7**, 112–116 (1986).
6. N.M. Davies, 'Clinical Pharmacokinetics of Flurbiprofen and its Enantiomers', *Clin. Pharmacokinet.*, **28**, 100–114 (1995).
7. H.A. Adams, A. Thiel, A. Jung, G. Fengler, G. Hempelmann, 'Studies Using *S*-(+)-ketamine on Probands. Endocrine and Circulatory Reactions, Recovery and Dream Experiences', *Anaesthesist*, **41**, 588–596 (1992).
8. F. Jamali, D.R. Brocks, 'Clinical Pharmacokinetics of Ketoprofen and its Enantiomers', *Clin. Pharmacokinet.*, **19**, 197–217 (1990).
9. D. Thienpont, J. Brugmans, K. Abadi, S. Tanamal, 'Tetramisole in the Treatment of Nematode Infections in Man', *Am. J. Trop. Med. Hyg.*, **18**, 520–525 (1969).
10. N. Dubois, N. Muller, F. Lapicque, P. Gillet, P. Netter, R.J. Royer, 'Stereoselective Protein Binding of Non-steroidal Anti-inflammatory Agents. Pharmacological Implications', *Therapie*, **48**, 335–339 (1993).
11. K. Stoschitzky, W. Lindner, W. Kiowski, 'Stereoselective Vascular Effects of the (*R*)- and (*S*)-Enantiomers of Propranolol and Atenolol', *J. Cardiovasc. Pharmacol.*, **25**, 268–272 (1995).
12. R. Kato, 'The Study of the Racemic Drugs', *Gekkan Yakuji*, **29**, 2039–2042 (1987).
13. D.W. Boulton, J.P. Fawcett, 'Enantioselective Disposition of Salbutamol in Man Following Oral and Intravenous Administration', *Br. J. Clin. Pharmacol.*, **41**, 35–40 (1996).
14. J. Brachmann, T. Beyer, C. Schmitt, W. Schols, M. Montero, T. Hilbel, M. Schweizer, W. Kubler, 'Electrophysiologic and Antiarrhythmic Effects of D-Sotalol', *J. Cardiovasc. Pharmacol.*, **20**(Suppl 2), S91–95 (1992).
15. R. Richards, A.E. Tattersfield, 'Bronchial Beta-adrenoceptor Blockade Following Eyedrops of Timolol and its Isomer L-714,465 in Normal Subjects', *Br. J. Clin. Pharmacol.*, **20**, 459–462 (1985).
16. H. Echizen, T. Brecht, S. Niedergesass, B. Vogelgesang, M. Eichelbaum, 'The Effect of Dextro-, Levo-, and Racemic Verapamil on Atrioventricular Condition in Humans', *Am. Heart J.*, **109**, 210–217 (1985).
17. Y.W. Lam, 'Stereoselectivity: an Issue of Significant Importance in Clinical Pharmacology', *Pharmacotherapy*, **8**, 145–157 (1988).
18. LHASA (Logic and Heuristics Applied to Synthetic Analysis) is a program for synthesis planning, an expert system to assist chemists in designing efficient routes to target molecules for organic synthesis. LHASA Ltd, School of Chemistry, University of Leeds (Leeds, UK).
19. CAMEO (Computer-Assisted Mechanistic Evaluation of Organic Reactions) is an interactive computer program that predicts the products of organic reactions given the starting materials and reaction conditions. W. L. Jorgensen Group Dept. of Chemistry, Yale University (New Haven, CT, USA).
20. A. Collet, 'Optical Resolution by Crystallization Methods', in *Chiral Separation by HPLC*, ed. A.M. Krstulovic, Ellis Horwood Ltd, UK, 81–104, 1989.
21. B. Pynnonen, 'Simulated Moving Bed Processing: Escape from the High-cost Box', *J. Chromatogr., A*, **827**, 143–160 (1998).
22. D.B. Broughton, D.B. Carson, 'The Molex Process', *Petroleum Refiner*, **38**, 130–134 (1959).
23. D.B. Broughton, C.G. Gerhold, U.S. Patent No. 2,985,589, May 23, 1961.
24. D.B. Broughton, 'Molex: Case History of a Process', *Chem. Eng. Prog.*, **64**, 60–65 (1968).
25. M. Negawa, F. Shoji, 'Optical Resolution by Simulated Moving-bed Adsorption Technology', *J. Chromatogr.*, **590**, 113–117 (1992).
26. C. Migliorini, M. Mazzotti, M. Morbidelli, 'Continuous Chromatographic Separation Through Simulated Moving Beds under Linear and Nonlinear Conditions', *J. Chromatogr., A*, **827**, 161–173 (1998).
27. M. Mazzotti, G. Storti, M. Morbidelli, 'Optimal Operation of Simulated Moving Bed Units for Nonlinear Chromatographic Separations', *J. Chromatogr., A*, **769**, 3–24 (1997).
28. M. Juza, O.D. Giovanni, G. Biressi, V. Schurig, M. Mazzotti, M. Morbidelli, 'Continuous Enantiomer Separation of the Volatile Inhalation Anesthetic Enflurane

with a Gas Chromatographic Simulated Moving Bed Unit', *J. Chromatogr., A*, **813**, 333–347 (1998).
29. E.R. Francotte, P. Richert, M. Mazzotti, M. Morbidelli, 'Simulated Moving Bed Chromatographic Resolution of Chiral Antitussive', *J. Chromatogr., A*, **796**, 239–248 (1998).
30. R.M. Nicoud, F. Charton, 'Complete Design of a Simulated Moving Bed', *J. Chromatogr., A*, **702**, 97–112 (1995).
31. D.W. Guest, 'Evaluation of Simulated Moving Bed Chromatography for Pharmaceutical Process Development', *J. Chromatogr., A*, **760**, 159–162 (1997).
32. M. Schulte, R. Ditz, R.M. Devant, J.N. Kinkel, F. Charton, 'Comparison of the Specific Productivity of Different Chiral Stationary Phases Used for Simulated Moving-bed Chromatography', *J. Chromatogr., A*, **769**, 93–100 (1998).
33. J. Strube, U. Altenhoner, M. Schmidt-Traub, M. Schulte, 'Dynamic Simulation of Simulated Moving-bed Chromatographic Processes for the Optimization of Chiral Separations', *J. Chromatogr., A*, **769**, 81–92 (1997).
34. S. Nagamatsu, K. Murazumi, S. Makino, 'Chiral Separation of a Pharmaceutical Intermediate by a Simulated Moving Bed Process', *J. Chromatogr., A*, **832**, 55–65 (1999).
35. E.R. Francotte, P. Richert, 'Applications of Simulated Moving-bed Chromatography to the Separation of the Enantiomers of Chiral Drugs', *J. Chromatogr., A*, **769**, 101–107 (1997).
36. T. Proll, E. Kusters, 'Optimization Strategy for Simulated Moving Bed Systems', *J. Chromatogr., A*, **800**, 135–150 (1998).
37. E. Küsters, G. Gerber, F.D. Antia, 'Enantioseparation of a Chiral Epoxide by Simulated Moving Bed Chromatography Using Chiralsel-OD', *Chromatographia*, **40**, 387–393 (1995).
38. E. Küsters, J. Nozulak, 'The Preparative Chromatographic Enantioseparation of a Racemic Morphanthridine Analogue on a Chiral Stationary Phase', *Chromatographia*, **47**, 440–442 (1998).
39. N. Gottschlich, V. Kasche, 'Purification of Monoclonal Antibodies by Simulated Moving-bed Chromatography', *J. Chromatogr., A*, **765**, 201–206 (1997).
40. C.B. Ching, B.G. Lim, E.J.D. Lee, S.C. Ng, 'Preparative Resolution of Praziquantel Enantiomers by Simulated Counter-current Chromatography', *J. Chromatogr., A*, **634**, 215–219 (1993).
41. S. Nagamatsu, S. Makino, 'Chiral Separation by SMB System in a Pilot Scale', PP2-67, poster at 9th International Symposium on Chiral Discrimination 1997, Nagoya, 1997.
42. E. Cavoy, M.F. Deltent, S. Lehoucq, D. Miggiano, 'Laboratory-developed Simulated Moving Bed for Chiral Drug Separations. Design of the System and Separation of Tramadol Enantiomers', *J. Chromatogr., A*, **769**, 49–57 (1997).
43. M. Juza, O. Di Giovanni, G. Biressi, V. Schurig, M. Mazzotti, M. Morbidelli, 'Continuous Enantiomer Separation of the Volatile Inhalation Anaesthetic Enflurane with a Gas Chromatographic Simulated Moving Bed Unit', *J. Chromatogr., A*, **813**, 333–347 (1998).
44. M. Perrut, 'Advances in Supercritical-fluid-chromatographic Processes', *J. Chromatogr., A*, **658**, 293–313 (1994).
45. http://www.ifpma.org/ich2.html
46. ICH guideline (Topics Q3A), Impurities in New Drug Substances. http://www.ipfma.org/ich1.html (1999).
47. ICH guideline (Topics Q3B), Impurities in New Drug Products. http://www.ipfma.org/ich1.html (1999).
48. 'FDA's Policy Statement for the Development of New Stereoisomeric Drugs' (Announcement), *Chirality*, **4**, 338–340 (1992).
49. CVMP, Note for Guidance. Investigation of Chiral Active Substances. Committee for Veterinary Medicinal Products, 1998. http://www.endra.org/enhome.html
50. ICH guideline (Topics Q6A), Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products (1999). These have been discussed in a state of Step 3 (where a regulatory rapporteur is designated from the EU, MHW or FDA, and comments are collected by the regulatory agencies in the three regions and exchanged with the other regulatory bodies) by the ICH members and its technical Expert Working Groups, at its meeting in Brussels from 8 to 11 March, 1999. When the full recommendations are incorporated into domestic regulations or other appropriate administrative measures, according to national/regional internal procedures, the draft guidance entitled 'Q6A Specifications' will be finished as a single set of global specifications (a state of Step 5). http://www.ipfma.org/ich1.html
51. W.A. König, 'Separation of Enantiomers by Capillary Gas Chromatography with Chiral Stationary Phases', *J. High Resolut. Chromatogr. Commun.*, **5**, 588–595 (1982).
52. E. Bayer, E. Kusters, G.J. Nicholson, 'Gas Chromatographic Separation of Enantiomeric Sulphur Compounds on Chirasil-Val', *J. Chromatogr.*, **320**, 393–396 (1985).
53. T. Saeed, R. Sandra, M. Verzele, 'Synthesis and Properties of a Novel Chiral Stationary Phase for Resolution of Amino-acid Enantiomers', *J. Chromatogr.*, **186**, 611–618 (1979).
54. W.A. König, I. Benecke, S. Sievers, 'New Results in the Gas Chromatographic Separation of Enantiomers of Hydroxy Acids and Carbohydrates', *J. Chromatogr.*, **217**, 71–79 (1981).
55. M. Jung, V. Schurig, 'Enantiomeric Separation by GC on Chirasil-dex: Systematic Study of Cyclodextrin Concentration, Polarity, Immobilization, and Column Stability', *J. Microcol. Sep.*, **5**, 11–22 (1993).

56. V. Schurig, 'Enantiomer Separation by Gas Chromatography on Chiral Stationary Phases', *J. Chromatogr., A*, **666**, 111–129 (1994).
57. A.B. Attygalle, E.D. Morgan, R.P. Evershed, S.J. Rowland, 'Comparison of Three Derivatives for Enantiomeric Separation of Chiral Alcohols and the Absolute Configuration of Myrmica and Octan-3-ol', *J. Chromatogr.*, **260**, 411–417 (1983).
58. R. Shimizu, T. Kakimoto, K. Ishii, Y. Fujimoto, H. Nishi, N. Tsumagari, 'New Derivatization Reagent for the Resolution of Optical Isomers in Diltiazem Hydrochloride by High-performance Liquid Chromatography', *J. Chromatogr.*, **357**, 119–125 (1986).
59. M. Chaffman, R.N. Brogden, 'Diltiazem. A Review of its Pharmacological Properties and Therapeutic Efficacy', *Drug*, **29**, 387–454 (1986).
60. T. Kinoshita, Y. Kasahara, N. Nimura, 'Reversed-phase High-performance Liquid-chromatographic Resolution of Non-esterified Enantiomeric Amino-acids by Derivatization with 2,3,4,6-tetra-*O*-acetyl-beta-D-glucopyranosyl isothiocyanate and 2,3,4-tri-*O*-acetyl-alpha-D-arabinopyranosyl Isothiocyanate', *J. Chromatogr.*, **210**, 77–81 (1981).
61. N. Nimura, A. Toyama, T. Kinoshita, 'Optical Resolution of Amino-acid Enantiomers by High-performance Liquid Chromatography', *J. Chromatogr.*, **316**, 547–552 (1984).
62. H. Nishi, N. Fujimura, H. Yamaguchi, W. Jyomori, T. Fukuyama, 'Reversed-phase HPLC Separation of the Enantiomers of Denopamine after Derivatization with GITC Chiral Reagent', *Chromatographia*, **30**, 186–190 (1990).
63. S. Einarsson, B. Josefsson, P. Moller, D. Sanchez, 'Separation of Amino Acid Enantiomers and Chiral Amines Using Precolumn Derivatization with (+)-1-(9-Fluorenyl)ethyl Chloroformate and Reversed-phase Liquid Chromatography', *Anal. Chem.*, **59**, 1191–1195 (1987).
64. M. Kagawa, Y. Machida, H. Nishi, 'The Enantiomeric Excess Determination of Acetyl-L-carnitine by Reversed-phase High-performance Liquid Chromatography Using Chiral Derivatization', *J. Chromatogr., A*, **857**, 127–135 (1999).
65. W. Lindner, 'Recent Developments in HPLC Enantioseparation: a Selected Review', *Chromatographia*, **24**, 97–107 (1987).
66. M. Schleimer, W.H. Pirkle, V. Schurig, 'Enantiomer Separation by High-performance Liquid Chromatography on Polysiloxane-based Chiral Stationary Phases', *J. Chromatogr.*, **679**, 23–34 (1994).
67. S. Oi, H. Ono, H. Tanaka, M. Shijo, S. Miyano, 'Axially Dissymmetric Bianthracene-based Chiral Stationary Phase for the High-performance Liquid Chromatographic Separation of Enantiomers', *J. Chromatogr.*, **679**, 35–46 (1994).
68. Y. Machida, H. Nishi, K. Nakamura, H. Nakai, T. Sato, 'Enantiomeric Separation of Diols and β-Amino Alcohols by Chiral Stationary Phase Derived from (*R*,*R*)-tartramide', *J. Chromatogr., A*, **757**, 73–79 (1997).
69. Y. Okamoto, M. Kawashima, R. Aburatani, K. Hatada, T. Nishiyama, M. Masuda, 'Chromatographic Resolution. XII. Optical Resolution of Beta-blockers by HPLC on Cellulose Triphenylcarbamate Derivatives', *Chem. Lett.*, 1237–1240 (1986).
70. Y. Okamoto, Y. Kaida, H. Hayashida, K. Hatada, 'Tris(1-phenylethylcarbamate)s of Cellulose and Amylose as Useful Chiral Stationary Phases for Chromatographic Optical Resolution', *Chem. Lett.*, 909–912 (1990).
71. Y. Okamoto, E. Yashima, 'Polysaccharide Derivatives for Chromatographic Separation of Enantiomers', *Angew. Chem. Int. Ed.*, **37**, 1020–1043 (1998).
72. D.W. Armstrong, W. DeMond, B.P. Czech, 'Separation of Metallocene Enantiomers by Liquid Chromatography: Chiral Recognition via Cyclodextrin Bonded Phases', *Anal. Chem.*, **57**, 481–484 (1985).
73. S.C. Chang, G.L. Reid, III, S. Chen, C.d. Chang, D.W. Armstrong, 'Evaluation of a New Polar-organic High-performance Liquid-chromatographic Mobile Phase for Cyclodextrin-bonded Chiral Stationary Phases', *Trends in Analytical Chemistry*, **12**, 144–153 (1993).
74. K. Nakamura, H. Fujima, H. Kitagawa, H. Wada, K. Makino, 'Preparation and Chromatographic Characteristics of a Chiral-recognizing Perphenylated Cyclodextrin Column', *J. Chromatogr., A*, **694**, 111–118 (1995).
75. T. Miwa, M. Ichikawa, M. Tsuno, T. Hattori, T. Miyakawa, M. Kayano, Y. Miyake, 'Direct Liquid-chromatographic Resolution of Racemic Compounds. Use of Ovomucoid as a Column Ligand', *Chem. Pharm. Bull.*, **35**, 682–686 (1987).
76. J. Haginaka, C. Seyama, H. Yasuda, K. Takahashi, 'Investigation of Enantioselectivity and Enantiomeric Elution Order of Propranolol and its Ester Derivatives on an Ovomucoid-bonded Column', *J. Chromatogr.*, **598**, 67–72 (1992).
77. S. Allenmark, B. Bomgren, H. Boren, 'Direct Liquid-chromatographic Separation of Enantiomers on Immobilized Protein Stationary Phases. III. Optical Resolution of a Series of *N*-Aroyl-D,L-amino-acids by High-performance Liquid Chromatography on Bovine Serum Albumin Covalently Bound to Silica', *J. Chromatogr.*, **264**, 63–68 (1983).
78. T. Shinbo, T. Yamaguchi, K. Nishimura, M. Sugiura, 'Chromatographic Separation of Racemic Amino-acids by use of Chiral Crown Ether-coated Reversed-phase Packings', *J. Chromatogr.*, **405**, 145–153 (1987).
79. T. Shinbo, T. Yamaguchi, H. Yanagishita, D. Kitamoto, M. Sugiura, 'Improved Crown Ether-based Chiral Stationary Phase', *J. Chromatogr.*, **625**, 101–108 (1992).
80. D.W. Armstrong, Y. Tang, S. Chen, Y. Zhou, C. Bagwill, L.R. Chen, 'Macrocyclic Antibiotics as a New Class

of Chiral Selectors for Liquid Chromatography', *Anal. Chem.*, **66**, 1473–1484 (1994).
81. S. Chen, Y. Liu, D.W. Armstrong, J.I. Borrell, T.B. Martinez, J.L. Matallana, 'Enantioresolution of Substituted 2-Methoxy-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitriles on Macrocyclic Antibiotic and Cyclodextrin Stationary Phases', *J. Liq. Chromatogr.*, **18**, 1495–1507 (1995).
82. D.W. Armstrong, Y. Liu, K.H. Ekborgott, 'A Covalently Bonded Teicoplanin Chiral Stationary Phase for HPLC Enantioseparations', *Chirality*, **7**, 474–497 (1995).
83. E.H.Y. Aboul-Enein, V. Serignese, 'Enantiomeric Separation of Several Cyclic Imides on a Macrocyclic Antibiotic (Vancomycin) Chiral Stationary Phase under Normal- and Reversed-phase Conditions', *Chirality*, **10**(4), 358–361 (1998).
84. O.K.H. Ekborg, Y. Liu, D.W. Armstrong, 'Highly Enantioselective HPLC Separations Using the Covalently Bonded Macrocyclic Antibiotic, Ristocetin A, Chiral Stationary Phase', *Chirality*, **10**(5), 434–484 (1998).
85. A. Berthod, Y. Liu, C. Bagwill, D.W. Armstrong, 'Facile Liquid-chromatographic Enantioresolution of Native Amino-acids and Peptides Using a Teicoplanin Chiral Stationary Phase', *J. Chromatogr., A*, **731**, 123–137 (1996).
86. H. Nishi, N. Fujimura, H. Yamaguchi, T. Fukuyama, 'Direct High-performance Liquid Chromatographic Separation of the Enantiomers of Diltiazem Hydrochloride and its 8-Chloro Derivative on a Chiral Ovomucoid Column', *J. Chromatogr.*, **633**, 89–96 (1993).
87. H. Nishi, N. Fujimura, H. Yamaguchi, T. Fukuyama, 'Direct Enantiomeric Separation of Racemic Precursors to Diltiazem Hydrochloride and Clentiazem Maleate by HPLC on a Chiral Ovomucoid Column', *Chromatographia*, **35**, 305–310 (1993).
88. H. Nishi, K. Yamasaki, Y. Kokusenya, T. Sato, 'Optical Resolution of Imidapril Hydrochloride by High-performance Liquid Chromatography and Application to the Optical Purity Testing of Drugs', *J. Chromatogr., A*, **672**, 125–133 (1994).
89. M. Dolezalova, M. Tkaczykova, 'HPLC Enantioselective Separation of Aromatic Amino and Hydrazino Acids on a Teicoplanin Stationary Phase and the Enantiomeric Purity Determination of L-Isomers Used as Drugs', *Chirality*, **11**, 394–403 (1999).
90. A. Torrens, J.A. Castrillo, J. Frigola, L. Salgado, J. Redondo, 'Optical Resolution and Enantiomeric Purity Determination of the Analgesic Cizolirtine', *Chirality*, **11**, 63–69 (1999).
91. W.H. Pirkle, G.J. Terfloth, 'Naproxen-derived Segmented and Sidechain-modified Polysiloxanes as Chiral Stationary Phases', *J. Chromatogr., A*, **704**, 269–277 (1995).
92. J. Haginaka, H. Matsunaga, T. Tsukamoto, 'Separation of Enantiomers on a Chiral Stationary Phase Based on Ovoglycoprotein. III. Effect of Aggregation of Ovoglycoprotein on Chiral Resolution', *J. Chromatogr., A*, **830**, 81–89 (1999).
93. Y. Machida, H. Nishi, K. Nakamura, H. Nakai, T. Sato, 'Enantiomer Separation of Amino Compounds by a Novel Chiral Stationary Phase Derived from Crown Ether', *J. Chromatogr., A*, **805**, 85–92 (1998).
94. M.H. Hyun, J.S. Jin, W. Lee, 'Liquid Chromatographic Resolution of Racemic Amino Acids and their Derivatives on a New Chiral Stationary Phase Based on Crown Ether', *J. Chromatogr., A*, **822**, 155–161 (1998).
95. Y. Machida, H. Nishi, K. Nakamura, 'Separation of the Enantiomers Amino and Amide Compounds on Novel Chiral Stationary Phases Derived from Crown Ether', *Chromatographia*, **49**, 621–627 (1999).
96. Y. Machida, H. Nishi, K. Nakamura, 'Enantiomer Separation of Hydrophobic Amino Compounds by High-performance Liquid Chromatography Using Crown Ether Dynamically Coated Chiral Stationary Phase', *J. Chromatogr., A*, **830**, 311–320 (1999).
97. H. Kanazawa, Y. Kunito, Y. Matsushima, S. Okubo, F. Mashige, 'Stereospecific Analysis of Chiral Drugs in Plasma by Chromatography-chiroptical Detector', *Chromatography*, **20**, 81–87 (1999).
98. O. Ramström, R.J. Ansell, 'Molecular Imprinting Technology: Challenge and Prospects for the Future', *Chirality*, **10**, 195–209 (1998).
99. K. Ensing, Theo de Boer, 'Tailor-made Materials for Tailor-made Applications: Application of Molecular Imprints in Chemical Analysis', *Trends in Analytical Chemistry*, **18**, 138–145 (1999).
100. R. Vespalec, P. Bocek, 'Chiral Separation by Capillary Zone Electrophoresis: Present Status of the Art', *Electrophoresis*, **18**, 843–852 (1997).
101. K. Verleysen, P. Sandra, 'Separation of Chiral Compounds by Capillary Electrophoresis', *Electrophoresis*, **19**, 2798–2833 (1998).
102. G. Gubitz, M.G. Schmid, 'Chiral Separation Principles in Capillary Electrophoresis', *J. Chromatogr., A*, **792**, 179–225 (1997).
103. B. Chankvetadze, *Capillary Electrophoresis in Chiral Analysis*, John Wiley & Sons, New York, 1997.
104. H. Nishi, S. Terabe, 'Micellar Electrokinetic Chromatography. Perspective in Drug Analysis', *J. Chromatogr., A*, **735**, 3–27 (1996).
105. J. Vindevogel, P. Sandra, *Introduction to Micellar Electrokinetic Chromatography*, Huthig, Heidelberg, 1992.
106. S. Fanali, 'Identification of Chiral Isomers by Capillary Electrophoresis', *J. Chromatogr., A*, **735**, 77–121 (1996).
107. S. Fanali, 'Controlling Enantioselectivity in Chiral Capillary Electrophoresis with Inclusion-complexation', *J. Chromatogr., A*, **792**, 227–267 (1997).
108. C.E.S. van de Griend, K. Groningsson, 'Validation of a Capillary Electrophoresis Method for the Enantiomeric Purity Testing of Ropivacaine, a New Local Anaesthetic Compound', *J. Pharm. Biomed. Anal.*, **14**, 295–304 (1996).

109. S. Terabe, Y. Miyashita, Y. Ishihama, O. Shibata, 'Cyclodextrin-modified Micellar Electrokinetic Chromatography: Separation of Hydrophobic and Enantiomer Compounds', *J. Chromatogr.*, **636**, 47–55 (1993).
110. M. Fillet, L. Fotsing, J. Crommen, 'Enantioseparation of Uncharged Compounds by Capillary Electrophoresis Using Mixtures of Anionic and Neutral β-Cyclodextrin Derivatives', *J. Chromatogr., A*, **817**, 113–119 (1998).
111. I.S. Lurie, 'Separation Selectivity in Chiral and Achiral Capillary Electrophoresis with Mixed Cyclodextrins', *J. Chromatogr., A*, **792**, 297–307 (1997).
112. W.X. Huang, S.D. Fazio, R.V. Vivilecchia, 'Achievement of Enantioselectivity of Non-polar Primary Amines by a Non-chiral Crown Ether', *J. Chromatogr., A*, **781**, 129–137 (1997).
113. Y. Tanaka, M. Yanagawa, S. Terabe, 'Separation of Neutral and Basic Enantiomers by Cyclodextrin Electrokinetic Chromatography Using Anionic Cyclodextrin Derivatives as Chiral Pseudo-stationary Phases', *J. High Resolu. Chromatogr.*, **19**, 421–433 (1996).
114. G. Schlte, B. Chankvetadze, G. Blaschke, 'Enantioseparation in Capillary Electrophoresis Using 2-Hydroxypropyltrimethylammonium Salt of β-Cyclodextrin as a Chiral Selector', *J. Chromatogr., A*, **771**, 259–266 (1997).
115. Y. Kuwahara, H. Nishi, 'Fast Drug Analysis by Capillary Electrophoresis. II. Effects of Purity of Dimethyl β-Cyclodextrin as Chiral Selector on Enantiomer Separations', *Yakugaku Zasshi*, **119**, 288–298 (1999). (In Japanese.)
116. J.B. Vincent, G. Vigh, 'Systematic Approach to Methods Development for the Capillary Electrophoretic Analysis of a Minor Enantiomer Using a Single-isomer Sulfated Cyclodextrin. A Case Study of L-Carbidopa Analysis', *J. Chromatogr., A*, **817**, 105–111 (1998).
117. H. Nishi, K. Nakamura, H. Nakai, T. Sato, 'Separation of Enantiomers and Isomers of Amino Compounds by Capillary Electrophoresis and High-performance Liquid Chromatography Utilizing Crown Ethers', *J. Chromatogr., A*, **757**, 225–235 (1997).
118. Y. Mori, K. Ueno, T. Umeda, 'Enantiomeric Separations of Primary Amino Compounds by Non-aqueous Capillary Zone Electrophoresis with a Chiral Crown Ether', *J. Chromatogr., A*, **757**, 328–332 (1997).
119. H. Nishi, 'Enantioselectivity in Capillary Electrophoresis with Polysaccharides', *J. Chromatogr., A*, **792**, 327–347 (1997).
120. R.M.C. Sutton, K.L. Sutton, A.M. Stalcup, 'Chiral Capillary Electrophoresis with Noncyclic Oligo- and Polysaccharide Chiral Selectors', *Electrophoresis*, **18**, 2297–2304 (1997).
121. X. Wang, J.-T. Lee, D.W. Armstrong, 'Separation of Enantiomers by Capillary Electrophoresis Using Pentosan Polysulfate', *Electrophoresis*, **20**, 162–170 (1999).
122. R. Gotti, V. Cavrini, A. Andrisano, G. Mascellani, 'Dermatan Sulfate as Useful Chiral Selector in Capillary Electrophoresis', *J. Chromatogr., A*, **814**, 205–211 (1998).
123. H. Nishi, S. Izumoto, K. Nakamura, H. Nakai, T. Sato, 'Dextran and Dextrin as Chiral Selectors in Capillary Zone Electrophoresis', *Chromatographia*, **42**, 617–630 (1996).
124. C. Desiderio, S. Fanali, 'Chiral Analysis by Capillary Electrophoresis Using Antibiotics as Chiral Selector', *J. Chromatogr., A*, **807**, 37–56 (1998).
125. D.W. Armstrong, U.B. Nair, 'Capillary Electrophoretic Enantioseparations Using Macrocyclic Antibiotics as Chiral Selectors', *Electrophoresis*, **18**, 2331–2342 (1997).
126. D.S. Hage, 'Chiral Separations in Capillary Electrophoresis Using Proteins as Stereoselective Binding Agents', *Electrophoresis*, **18**, 2311–2321 (1997).
127. S. Fanali, G. Caponecchi, Z. Aturki, 'Enantiomeric Resolution by Capillary Zone Electrophoresis: Use of Pepsin for Separation of Chiral Compounds of Pharmaceutical Interest', *J. Micro. Sep.*, **9**, 9–14 (1997).
128. Y. Tanaka, S. Terabe, 'Separation of the Enantiomers of Basic Drugs by Affinity Capillary Electrophoresis Using a Partial Filling Technique and α_1-Acid Glycoprotein as Chiral Selector', *Chromatographia*, **44**, 119–128 (1997).
129. H. Nishi, 'Enantiomer Separation of Drugs by Electrokinetic Chromatography', *J. Chromatogr., A*, **735**, 57–76 (1996).
130. P. Cammilleri, 'Chiral Surfactants in Micellar Electrokinetic Capillary Chromatography', *Electrophoresis*, **18**, 2322–2330 (1997).
131. Y. Mechref, Z.E. Rassi, 'Capillary Electrophoresis of Herbicides. II. Evaluation of Alkylglucoside Chiral Surfactants in the Enantiomeric Separation of Phenoxy Acid Herbicides', *J. Chromatogr., A*, **757**, 263–273 (1997).
132. Y. Mechref, Z.E. Rassi, 'Capillary Electrophoresis of Herbicides. III. Evaluation of Octylmaltopyranoside Chiral Surfactant in the Enantiomeric Separation of Phenoxy Acid Herbicides', *Chirality*, **8**, 518–524 (1996).
133. A.G. Peterson, E.S. Ahuja, J.P. Foley, 'Enantiomeric Separation of Basic Pharmaceutical Drugs by Micellar Electrokinetic Chromatography Using a Chiral Surfactant, *N*-Dodecoxycarbonyvaline', *J. Chromatogr., B*, **683**, 15–28 (1996).
134. W. Ding, J.S. Fritz, 'Carbamate Chiral Surfactants for Capillary Electrophoresis', *J. Chromatogr., A*, **831**, 311–320 (1999).
135. E. Billiot, J. Macossay, S. Thibodeaux, S.A. Shamsi, I.M. Warner, 'Chiral Separations Using Dipeptide Polymerized Surfactants: Effect of Amino Acid Order', *Anal. Chem.*, **70**, 1375–1381 (1998).
136. J. Wang, I.M. Warner, 'Chiral Separations Using Micellar Electrokinetic Chromatography and a Polymerized Chiral Micelle', *Anal. Chem.*, **66**, 3773–3776 (1994).
137. K.D. Altria, N.W. Smith, C.H. Turnbull, 'A Review of the Current Status of Capillary Electrochromatography

Technology and Applications', *Chromatographia*, **46**, 664–673 (1997).
138. D. Wistuba, H. Czesla, M. Roeder, V. Schurig, 'Enantiomer Separation by Pressure-supported Electrochromatography Using Capillaries Packed with a Permethyl-β-cyclodextrin Stationary Phase', *J. Chromatogr., A*, **815**, 185–188 (1998).
139. F. Lelievre, C. Yan, R.N. Zare, P. Garel, 'Capillary Electrochromatography: Operating Characteristics and Enantiomer Separations', *J. Chromatogr., A*, **723**, 145–156 (1996).
140. K. Krause, M. Girod, B. Chankvetadze, G. Blaschke, 'Enantioseparations in Normal- and Reversed-phase Nano-high-performance Liquid Chromatography and Capillary Electrochromatography Using Polyacrylamide and Polysaccharide Derivatives as Chiral Stationary Phases', *J. Chromatogr., A*, **837**, 51–63 (1999).
141. C. Wolf, P.L. Spence, W.H. Pirkle, E.M. Derrico, D.M. Cavender, G.P. Rozing, 'Enantioseparations by Electrochromatography with Packed Capillaries', *J. Chromatogr., A*, **782**, 175–179 (1997).
142. L. Schweitz, L.I. Andersson, S. Nisson, 'Capillary Electrochromatography with Predetermined Selectivity Obtained Through Molecular Imprinting', *Anal. Chem.*, **69**, 1179–1183 (1997).
143. L. Schweitz, L.I. Andersson, S. Nisson, 'Molecular Imprint-based Stationary Phases for Capillary Electrochromatography', *J. Chromatogr., A*, **817**, 5–13 (1998).
144. ICH guideline (Topics Q2A), Text on Validation of Analytical Procedures (1995).
145. ICH guideline (Topics Q2B), Validation of Analytical Procedures: Methodology (1997).
146. 'Capillary Electrophoresis', *Pharm. Forum*, **22**, 1727–1735 (1996).
147. A.E. Bretnall, M.M. Hodgkison, G.S. Clarke, 'Micellar Electrokinetic Chromatography Stability Indicating Assay and Content Uniformity Determination for a Cholesterol-lowering Drug Product', *J. Pharm. Biomed. Anal.*, **15**, 1071–1075 (1997).
148. K. Persson, O. Astrom, 'Separation of Ibuprofen, Codeine Phosphate, their Degradation Products and Impurities by Capillary Electrophoresis. II. Validation', *J. Chromatogr., A*, **826**, 95–102 (1998).
149. 'Epinephryl Borate Ophthalmic Solution', *Pharm. Forum*, **23**, 3991–3992 (1997).
150. 'Ethambutol Hydrochloride Tablets', *Pharm. Forum*, **23**, 3992–3993 (1997).
151. 'Capillary electrophoresis (Stage 4)', *Jap. Pharm. Forum*, **8**, 178–201 (1999).

Combinatorial Chemistry Libraries, Analysis of

Christopher E. Kibbey
Parke-Davis Pharmaceutical Research, Division of Warner–Lambert Company, Ann Arbor, USA

1 Introduction 1
2 Combinatorial Chemistry 2
2.1 Solid-phase Synthesis 2
2.2 Mixture Synthesis 2
2.3 Split Synthesis 2
2.4 Spatially Addressed Synthesis 3
2.5 Encoding Strategies 3
2.6 Solution-phase Synthesis 7
3 Library Characterization by Fourier Transform Infrared Spectroscopy 7
3.1 On-resin Analysis by Fourier Transform Infrared Spectroscopy 7
3.2 Off-resin Analysis by Fourier Transform Infrared Spectroscopy 10
4 Library Characterization by Mass Spectrometry 13
4.1 On-resin Analysis by Mass Spectrometry 14
4.2 Off-resin Analysis by Mass Spectrometry 14
5 Library Characterization by Nuclear Magnetic Resonance Spectroscopy 19
5.1 On-resin Analysis by Nuclear Magnetic Resonance Spectroscopy 19
5.2 Off-resin Analysis by Nuclear Magnetic Resonance Spectroscopy 24
6 Library Screening by Affinity Selection and Mass Spectrometry 27
6.1 Ligand Selection by Affinity Chromatography 28
6.2 Ligand Selection by Gel-filtration 28
6.3 Ligand Selection by Ultrafiltration 28
6.4 Bio-affinity Characterization Mass Spectrometry 31
7 Combinatorial Library Quantitation 32
7.1 Quantitation using ^{1}H-Nuclear Magnetic Resonance Spectroscopy 32
7.2 Quantitation using Chemiluminescent Nitrogen Detection 34
7.3 Quantitation using Evaporative Light Scattering Detection 34
Abbreviations and Acronyms 35
Related Articles 36
References 37

Many of the analytical techniques applied to the analysis of traditional pharmaceutical compounds are used to characterize combinatorial chemistry libraries. In both cases, the medicinal chemist is primarily concerned with verifying the identity and purity of the compounds synthesized. The identity of a compound usually is confirmed spectroscopically by Fourier transform infrared (FTIR) spectroscopy, mass spectrometry (MS), and/or nuclear magnetic resonance (NMR) spectroscopy, while purity is determined using a chromatographic technique, such as high-performance liquid chromatography (HPLC). Combinatorial chemistry libraries may contain 100–100 000 compounds present as single species or as mixtures. The need to efficiently characterize such large numbers of compounds has been the primary motivation behind the development of many of the chromatographic–spectroscopic techniques applied to the analysis of combinatorial chemistry libraries. Many of these techniques have utility outside of the realm of normal library characterization. For example, the combination of affinity chromatography and MS has proven useful for screening combinatorial chemistry libraries against specific biological targets. The major impediments to the analysis of combinatorial chemistry libraries have been the challenges of designing analytical methods compatible with the wide chemical diversity of compounds encountered in these libraries, and the unique sample preparation and handling requirements imposed by combinatorial synthesis.

1 INTRODUCTION

The drive to improve the efficiency of the drug discovery process has created the need for rapid compound synthesis and efficient methods for screening compounds for biological activity. The fields of combinatorial chemistry and high-throughput screening matured during the early 1990s to meet the challenges of modern pharmaceutical research. Solid-phase synthesis (SPS) and parallel solution-phase synthesis techniques have the potential to deliver hundreds of thousands of compounds in a relatively short period of time. However, the emergence of these combinatorial chemistry techniques has created the additional demand for new analytical techniques to follow the course of chemical reactions and characterize final products. This article begins with a brief overview of the basic concepts of combinatorial

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

chemistry, and follows with a description of many of the proven analytical techniques employed in combinatorial chemistry laboratories.

2 COMBINATORIAL CHEMISTRY

The field of combinatorial chemistry originally focused on the automated synthesis of peptide and small molecule libraries for drug discovery. Invariably, these compounds were assembled on chemical scaffolds immobilized on a solid support. However, with the rapid development of new technologies to synthesize compounds en masse, and the application of these techniques to fields outside of pharmaceutical research, the field of combinatorial chemistry has been in need of a more comprehensive definition. While a formal definition has yet to emerge, combinatorial chemistry may be described as a process of synthesizing sets of structurally related compounds by reacting all possible combinations of a limited number of chemical reagents according to a general reaction scheme in a manner compatible with the subsequent screening of the products for specific properties. The following sections provide an overview of the techniques that have emerged in the field of combinatorial chemistry.

2.1 Solid-phase Synthesis

The beginnings of modern SPS date back to Merrifield's work on solid-phase peptide synthesis (SPPS) in the early 1960s.[1] By the early 1970s a number of researchers, including Rapoport and Crowley,[2] Patchornik and Kraus,[3] and Leznoff and Wong,[4] had reported the preparation of nonpeptide heterocyclic compounds by solid-phase organic synthesis (SPOS). Automated peptide synthesizers became commercially available in the early 1980s. Further interest in SPS was sparked during this time by breakthroughs in phosphoramidite chemistry[5] which brought forth DNA SPS. A quantum leap in SPS during the mid-1980s grew out of the efforts of Geysen et al.,[6] Furka et al.,[7] Houghten,[8] Lam et al.,[9] and Fodor et al.[10] to increase the efficiency of SPPS through parallel synthesis. By the mid-1990s, many of the major pharmaceutical companies in the United States had established research groups to exploit the benefits of SPOS.

The primary advantages of SPOS over traditional synthesis are twofold. First, reactions on solid supports can be driven to completion through the use of excess reagents. Second, products can be isolated with reasonable purity by simply washing the excess reagents away with organic solvent. Using modern SPS techniques, it is possible to synthesize hundreds of thousands to millions of compounds simultaneously. Despite its universal acceptance in the pharmaceutical industry, SPOS is still in its infancy. The repertoire of synthetic organic reactions that have been carried out on solid-phase supports[11,12] represents only a small fraction of the chemistries available to the field of organic chemistry. A factor that has limited the use of SPOS in pharmaceutical research is the three to six month lead-time required to optimize new chemistries on solid supports.

The majority of SPS work is carried out on spherical polymeric supports (beads) composed of either polystyrene cross-linked 1–2% divinylbenzene, or polyethylene glycol–polystyrene block copolymer. A notable exception is the functionalized polyethylene–poly(*N*,*N*-dimethylacrylamide/methacrylic acid) graft copolymer ''pins'' developed by Geysen et al. for peptide epitope mapping.[6,13] The surface of the resin beads used in SPOS contains active functional groups (linkers), which serve as sites for attachment during synthesis. The structures of a number of the more common linkers used in SPOS are shown in Figure 1.

2.2 Mixture Synthesis

The most straightforward approach to synthesizing multiple compounds in parallel on a solid support is to treat an appropriate resin-bound intermediate with a solution containing equimolar quantities of various reagents. Cleavage of the reaction products from the solid support will yield, in theory, an equimolar mixture of the desired products. This is precisely the approach taken by researchers, such as Geysen et al.,[14] involved in the synthesis of peptide libraries in the mid-1980s. The potential of this technique, however, is less promising when one considers that the reactivity of a reagent is highly compound specific. Rather than providing an equimolar distribution of products, the method usually results in a wide distribution of product yields. When the technique is applied to peptide synthesis, small variations in the reactivity of amino acids can be overcome through variation of the concentration of each reagent in the reaction mixture.[15] If the reactions are very efficient, it is possible to obtain equimolar product mixtures by using only one equivalent of each reagent in the coupling step.[16] Mixture synthesis is generally impractical for the synthesis of nonpeptide libraries.

2.3 Split Synthesis

Furka et al.[17,18] developed the ''split synthesis'' strategy for parallel synthesis to overcome the difficulties involved in using reagents with different chemical reactivities. In the split synthesis technique, a common pool of resin-bound intermediates is divided into equal pools, each pool is coupled with a separate reagent, and the resin

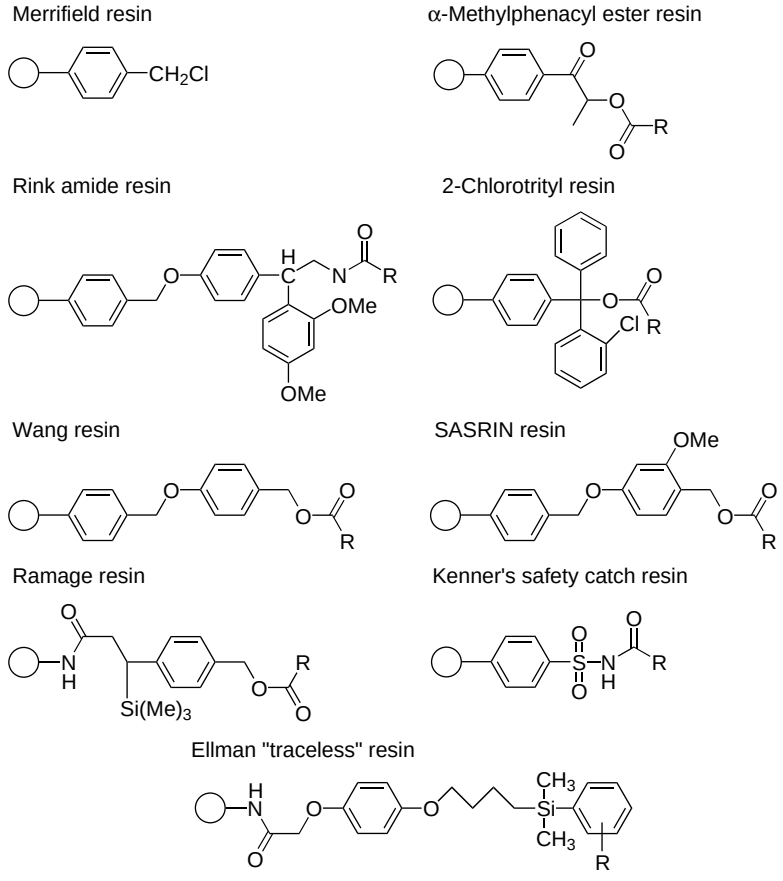

Figure 1 Structures of common resin linkers used in SPOS.

pools are then recombined into one. Multiple iterations of dividing, coupling, and pooling resin beads result in a mixture of resin-bound products, where each single bead consists of only one set of product molecules. An example of the split synthesis of a library of nine hypothetical compounds is illustrated in Figure 2. The advantage of the split synthesis technique is that the number of products grows geometrically with each iteration of the divide–couple–pool cycle. For oligomers the number of possible products equals X^n, where X is the number of starting monomers and n is the number of reaction cycles.

2.4 Spatially Addressed Synthesis

Spatially addressed synthesis is a special case of parallel synthesis developed by Fodor et al.[10] at Affymax Research. The technique uses photolithography to synthesize arrays of up to 100 000 discrete compounds on silica wafers approximately 1 cm^2 in area. Active functional groups are attached to the surface of the silica substrate and their termini blocked with photolabile protecting groups (see Figure 3). Specially designed masks are laid over the surface of the silica wafer, and laser light is used to cleave exposed protecting groups from the surface. The wafer is then immersed in a reaction solution consisting of a reagent derivatized with a photolabile protecting group. The deprotected active sites on the silica wafer react with the protected reagents in the solution to form a protected oligomer. Peptide libraries may be synthesized through repeated application of the laser deprotection and coupling steps. The structure of any compound in the spatially addressable array may be inferred from its location on the silica surface and knowledge of the masking scheme and order of reagent additions.

2.5 Encoding Strategies

Combinatorial libraries prepared by split synthesis always result in mixtures, rather than discrete products. When a ''hit'' is obtained in a biological assay, it becomes necessary to identify the active member in the test mixture. Deconvolution of an active well usually involves iterative resynthesis and testing. An example of deconvolution of a 1000 component library of oligomers is shown in Figure 4. In this example, the combinatorial library is prepared by split synthesis involving ten sets

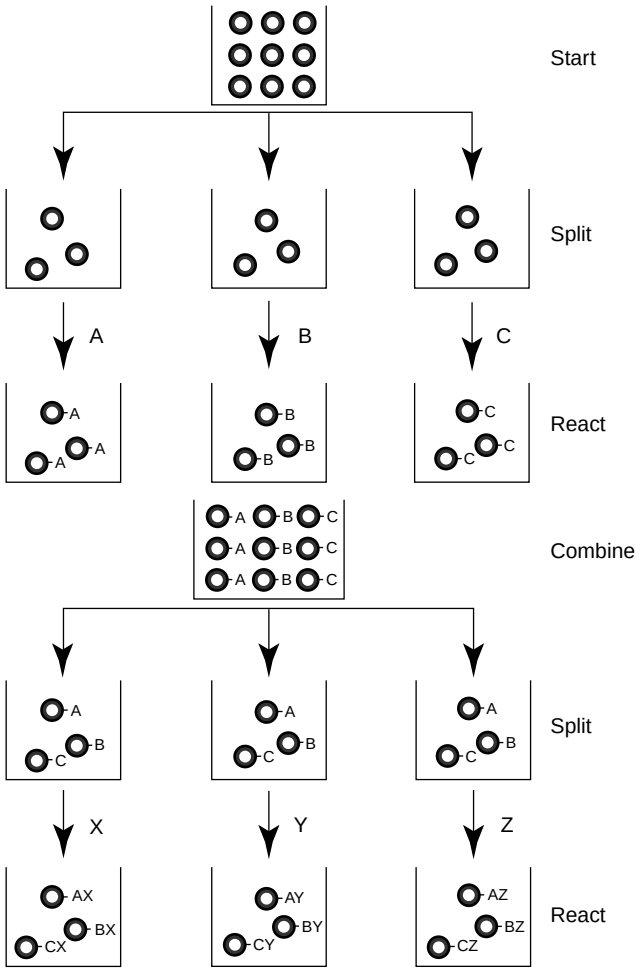

Figure 2 Split synthesis of a nine-member combinatorial library.

of monomers and three sequential reaction steps (i.e. $10 \times 10 \times 10 = 1000$) to give ten sets of oligomers. If the third well is biologically active, then the identity of the third monomer of the active compound is known from the last step of the synthesis. The identity of the second monomer of the active compound may be determined through resynthesis of a 100 component library by randomizing the first monomer, terminating the split at the end of the second reaction step, and fixing the position of the third monomer. In turn, the identity of the first

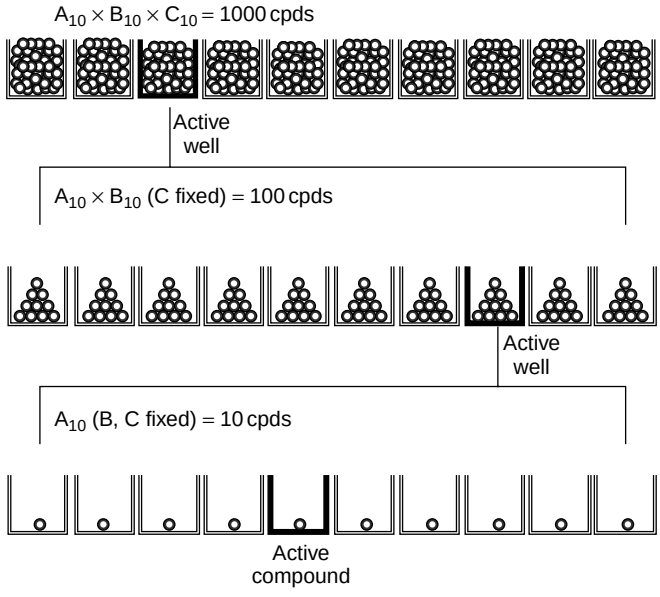

Figure 4 Deconvolution of a 1000-component library by iterative screening and resynthesis.

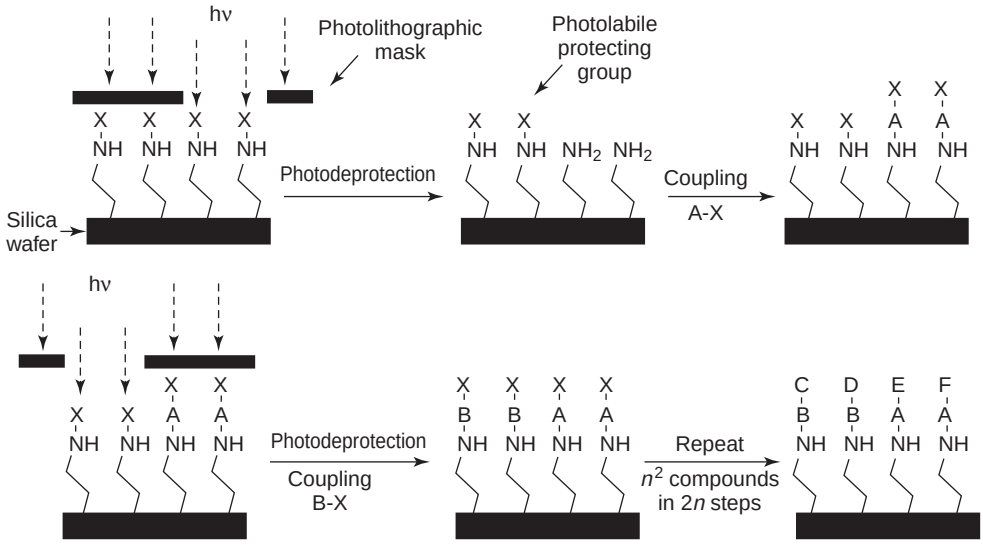

Figure 3 Light-directed, spatially addressable chemical synthesis. (Reprinted with permission from L.A. Thompson, J.A. Ellman, *Chem. Rev.*, **96**, 555–600 (1996). Copyright 1996 American Chemical Society.)

monomer of the active oligomer may be determined by fixing the positions of the second and third monomers.

The added burden of performing multiple resynthesis and the complications involved in deconvoluting libraries containing multiple active compounds have led researchers to devise other strategies for identifying the active components in mixture libraries. A number of encoding strategies have been developed by various researchers, and those based on DNA strands, peptide strands, molecular tags and radiofrequency (RF) transponders are described. Simply stated, an encoding strategy involves ''tagging'' resin beads at each step of a split synthesis, such that the identity of a compound on any bead can be determined by decoding the tag sequence. A generic example of library encoding is shown in Figure 5. The advantages of the encoding strategy are threefold. First, tagging schemes permit identification of all of the biologically active compounds in a mixture library, whereas deconvolution usually gives the most active library member. Second, rigorous analytical characterization of the library products is avoided, because compound identity is inferred from the tag sequence. A compound's identity can be verified faster by decoding a tag sequence, than by structural characterization. Third, encoding requires fewer synthetic steps than the iterative synthesis involved in library deconvolution. It is important to bear in mind, however, that the encoding process is independent of the reaction steps in split synthesis. Hence, failure of a reaction step will lead to an incorrect product assignment.

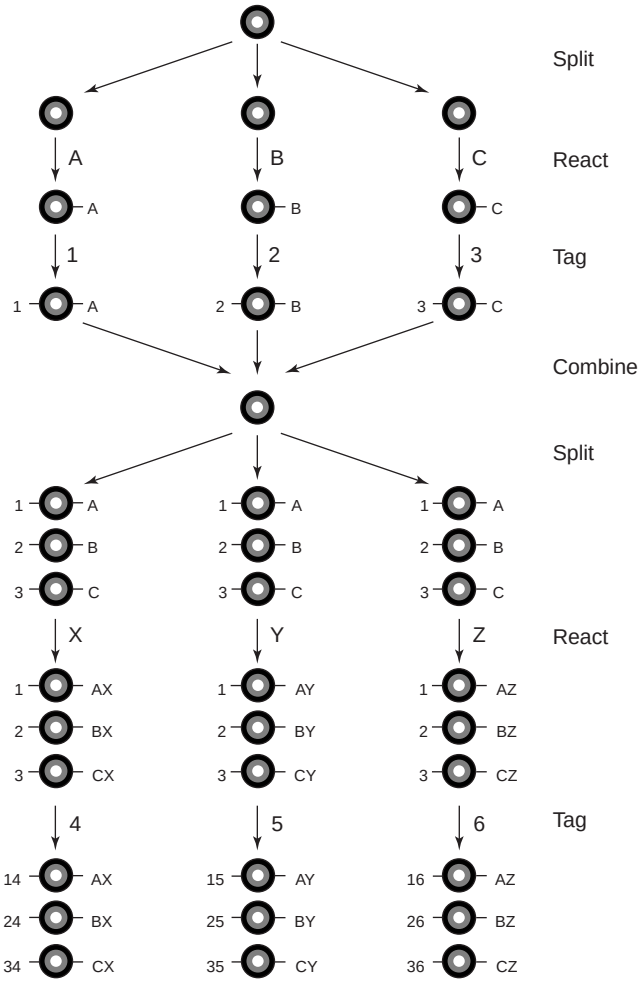

Figure 5 Encoding a solid-phase combinatorial library using molecular tags.

2.5.1 Oligonucleotide Tags

The first tagging strategy for combinatorial chemistry was reported by Brenner et al.[19,20] and involved encoding peptide libraries using oligonucleotide sequences. A pair of alternating combinatorial syntheses was used to attach the DNA strands and peptide monomers to separate sites on the bead surface. This encoding technique requires at least 15 nucleotides in the coding region for effective hybridization. Thus, a combinatorial synthesis involving five coupling steps would be encoded using combinations of triplets.

The library is decoded by amplifying the encoded oligonucleotide strands by the polymerase chain reaction (PCR), followed by DNA sequencing. The PCR amplification step provides enough sensitivity to allow single bead decoding. If multiple beads are decoded, separate PCR amplifications must be performed on each bead. A disadvantage of this encoding strategy is that the peptide products must remain immobilized on the bead surface during biological assay. Interference from the oligonucleotide tags could bias binding assays. Further, DNA strands are sensitive to temperature, redox reagents, and strong acids and bases, and hence they are of limited use in the synthesis of nonpeptide combinatorial libraries.

2.5.2 Peptide Tags

In a manner analogous to the use of oligonucleotide tags, peptides have been used to encode combinatorial libraries prepared by split synthesis. Two distinct applications of peptide encoding have been reported.[21,22] In the first approach, a bifunctional linker containing two orthogonal protecting groups is attached to a polystyrene resin. One of the protecting groups is acid labile and the other is removed by bases. The two sites on the linker serve as points of attachment for the peptide tag and the peptide ligand. A pair of alternating combinatorial syntheses is used to construct the ligand and tag on the solid support. The linker itself is acid labile, and is cleaved from the solid support prior to screening to yield a solution containing the coupled ligand/tag pairs. The active ligands are

isolated by affinity selection following biological assay. The identity of these active peptides is determined from Edman sequencing of the peptide tags.

Salmon et al.[22] devised a novel orthogonal release approach to screening and identifying active peptides in combinatorial libraries. Instead of relying on separate ligands and tags, the method uses a single peptide for both functions. Peptides are synthesized on an orthogonal linker possessing three pH sensitive cleavage sites, such that each linker molecule incorporates three copies of the peptide. In the assay procedure, approximately 500 resin beads are placed into each well of a 96-well microliter plate. A pH 4.5 buffer is applied to the beads to release the first third of the immobilized peptide molecules from the linker. The filtrate from each well is screened for biological activity, and the beads in the active wells are redistributed into clean assay plates as one bead per well. The single beads are treated with dilute base and the second peptide molecule from the linker is screened for activity. The identity of the active peptide is determined through Edman sequencing of the remaining peptide attached to the linker. Peptide sequencing by electrospray ionization mass spectrometry (ESIMS) and tandem mass spectrometry (MS/MS) also has been reported.[23]

2.5.3 Molecular Tags

To overcome the synthetic constraints imposed by the use of oligonucleotide and peptide tags, Still et al. developed a method for encoding combinatorial libraries using electrophoric tags.[24,25] Electrophoric tags are halogen-substituted phenoxyalkyl alcohols and are analyzed during library decoding by gas chromatography (GC) with electron capture detection (ECD). Prior to attachment on a polystyrene resin, the tag molecule is derivatized with an oxidatively labile linker (3-methoxy-4-hydroxybenzoic acid), and the linker's carboxylic acid group is converted to a diazoketone group. The derivatized tag is subsequently immobilized on the surface of the polystyrene resin through a metal-catalyzed carbene insertion reaction. The immobilized tags are extremely stable to the harsh conditions often encountered in organic synthesis, which has made this tagging technique ideally suited to the synthesis of nonpeptide combinatorial libraries.[26,27]

A novel feature of the encoding strategy developed by Still et al. is that the molecular tags are coupled through oxidatively labile linkers, while ligands are attached using photolabile groups. Molecular tags and ligands are released from the polystyrene support by independent mechanisms, thus eliminating interference during assay and decoding. In addition, the high sensitivity of ECD requires only picomoles of tag molecules for analysis. Hence, only about 1% of the loading capacity of the resin is required for tagging, leaving the bulk of the resin surface available for ligand immobilization.

As with other chemical tagging strategies, the preparation of combinatorial libraries encoded with molecular tags involves alternating cycles of ligand synthesis and encoding reactions. The choice of tag molecules and the sequence of their immobilization are based on their order of elution by GC. Molecular tags based on dialkylamines also have been reported.[28]

2.5.4 Radiofrequency Transponders

RF-encoded combinatorial chemistry makes use of miniature solid-state transponders to record encoding information during a split synthesis. These devices are encased in glass, which renders them inert to most organic reactions. Two types of RF transponder have been reported.[29,30] One type of transponder is designed to receive, store, and emit binary encoded RF signals. The specific reagents used in each step of a split synthesis are assigned unique bar codes by a computer, and these are transmitted to the transponder during synthesis. The entire synthetic pathway for a compound can be read from the information stored on the device. The second type of RF transponder does not store encoded data, but rather emits a unique numeric identifier when passed over a scanning device. A computer is used to track the reaction history for this second type of transponder in a split synthesis. The main advantages of RF transponders are that the encoding process does not require additional chemical reactions, and library decoding does not involve analytical characterization of molecular tags.

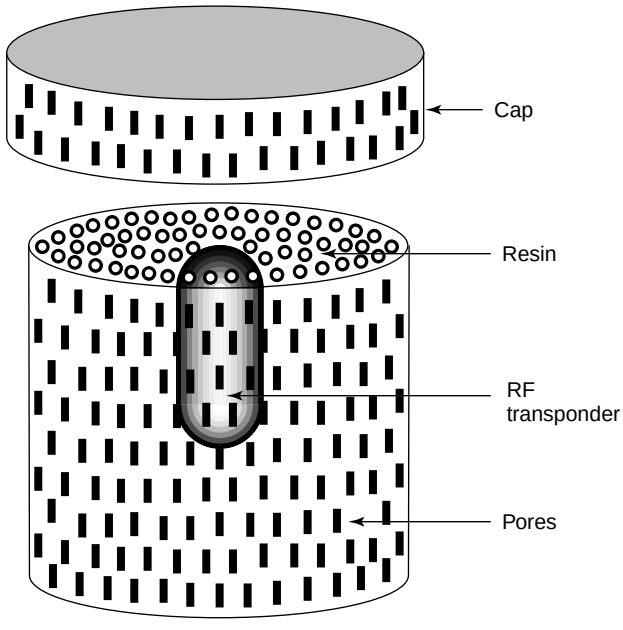

Figure 6 RF encoding of a single combinatorial library member. Multiple resin beads contained within the reactor are treated as a single unit during synthesis.

Both types of transponder are physically handled in the same manner. A set of porous capsules is filled with an appropriate resin, and a microchip transponder is placed inside each one (see Figure 6). The capsules are sealed and treated as individual reactors in a split synthesis. Organic reactions may be performed using conventional laboratory glassware,[31] and the capsules are easily manipulated manually, or with the aid of laboratory robots.

2.6 Solution-phase Synthesis

The application of liquid-handling robots to organic synthesis and the simultaneous design of apparatus for temperature control and agitation of multiple independent reaction vessels during the early to mid-1990s provided chemists with an attractive alternative to solid-phase combinatorial synthesis. Solution-phase synthesis is performed by reacting all possible combinations of a set of reagents in parallel. A primary advantage of solution-phase synthesis is that there is no need to develop strategies for the attachment of substituents to a solid support, or their subsequent cleavage.[11] Hence, lead times for reaction optimization are much shorter for solution-phase synthesis compared with SPS. Furthermore, solution-phase syntheses can be scaled to multi-milligram yields. Disadvantages of solution-phase synthesis are that the technique is limited to short reaction sequences, and the products can be difficult to purify. Most laboratories apply solution-phase synthesis to the generation of combinatorial libraries of discrete compounds, rather than mixtures. The population of solution-phase libraries is usually limited to only a few hundred compounds.

3 LIBRARY CHARACTERIZATION BY FOURIER TRANSFORM INFRARED SPECTROSCOPY

Infrared (IR) spectroscopy is a powerful sample characterization technique with broad application in organic and analytical chemistry. The mid-IR region provides both important functional group information ($4000-1300\,cm^{-1}$) and structural information ($1300-650\,cm^{-1}$). FTIR instruments are equipped with interferometers, which allow multiple spectral measurements of a sample to be made and averaged, and the resulting IR spectrum to be displayed in the frequency domain. The improvement in signal-to-noise that results from signal averaging affords FTIR instruments greater sensitivity than is possible with grating-based IR instruments. The high sensitivity of FTIR, its nondestructive nature, low cost, and ease of use make it ideally suited to reaction monitoring and final product characterization of combinatorial libraries.

While IR spectroscopy may be used quantitatively, its application to the characterization of organic molecules is usually qualitative. The widths of the bands in IR spectroscopy are quite narrow and lead to deviations from Beer's law over wide ranges of absorption. If quantitation is desired, the IR bands monitored should be reasonably wide and well resolved from nearby absorptions so that an accurate integration can be performed. The identity of a compound usually is confirmed by verifying the presence or absence of characteristic absorption bands in the product's IR spectrum. Changes in the intensity of specific bands in the IR spectra of reagents or product molecules are followed during reaction monitoring.

3.1 On-resin Analysis by Fourier Transform Infrared Spectroscopy

Rapid growth in the field of combinatorial chemistry during the mid-1990s lead to an increased interest in SPOS. The lengthy process of optimizing reaction strategies for use on solid supports has created the need for rapid and sensitive analytical techniques for monitoring solid-phase reactions. Ideally, a suitable technique for reaction monitoring would consume an insignificant amount of sample; involve little sample preparation; provide rapid analysis; possess high sensitivity; and have sufficient specificity to distinguish reagents from product. In traditional organic synthesis thin-layer chromatography (TLC) is the analytical technique of choice for monitoring reactions. FTIR spectroscopy has proven to be an equally capable technique for SPOS. Application of FTIR spectroscopy to reaction monitoring and final product characterization on solid supports is described in the following sections.

3.1.1 Fourier Transform Infrared of Pressed Potassium Bromide Pellets

Many of the early pioneers of SPS relied extensively upon IR spectroscopy to characterize resin-bound compounds. Crowley and Rapoport[32] quantitated the extent of chloromethylation of polystyrene resin based upon the H−C−Cl bending vibration at $1250\,cm^{-1}$ in the preparation of resin linkers for the Dieckmann cyclization of mixed esters. Frechet and Schuerch[33] characterized resin-bound intermediates by IR spectroscopy during a six-step synthesis of polystyrene linkers used in the SPS of oligosaccharides. Goldwasser and Leznoff[34] employed IR spectroscopy to characterize products from the SPS of monoester monoamides and monoester monoalcohols. Larsen et al.[35] followed the stepwise SPS

of a peptide using Fourier transform Raman spectroscopy (FTRS). Narita et al.[36,37] employed FTIR spectroscopy to characterize the conformation of peptides immobilized on polystyrene beads.

Preparation of samples for direct analysis by FTIR involves mixing approximately 10 mg of the resin beads with 190 mg of KBr, and compressing the mixture to form a pellet.[38] The quality of the resulting pellet is highly dependent on the pressing technique used, and the best results are obtained if the pressing die is warmed to 50 °C. The IR spectra of resin-bound products are usually obscured by absorptions from the polymer matrix, and background subtraction of the resin matrix is required to obtain interpretable sample spectra.

There are a number of disadvantages to treating resin-bound compounds like conventional solid samples during their preparation for analysis by FTIR. First, the KBr pellet technique consumes too much material. A 10-mg sample of 50-μm diameter resin contains approximately 100 000 beads, and most SPSs are carried out on less than 100 mg of resin. Second, this sample preparation technique is incompatible with mixture libraries prepared by split synthesis. The individual resin beads present in mixture libraries represent a number of unique products, and these must be characterized separately.

3.1.2 Single-bead Fourier Transform Infrared Microspectroscopy

Yan and Kumaravel[39] developed the technique of single-bead Fourier transform infrared microspectroscopy (SBFTIRM) to overcome the limitations of conventional FTIR analysis of bead-bound combinatorial libraries. The instrumentation required for SBFTIRM consists of a conventional FTIR instrument equipped with an IR microscope accessory. The IR microscope includes an $X-Y$ sample stage, a magnification objective, and a liquid-nitrogen-cooled mercury–cadmium–telluride (MCT) detector. Spectra may be obtained by conventional transmission mode, or by attenuated total reflectance (ATR). The latter technique detects only compounds on a resin bead's outer surface, whereas transmission mode IR spectra represent compounds on the surface and interior porous regions of the resin.

Sample preparation in SBFTIRM is straightforward. A drop of resin suspension is withdrawn from the reaction solution and the resin beads are washed with organic solvent and briefly dried under vacuum. The beads are placed on a NaCl window mounted on the $X-Y$ stage of the IR microscope. Spectra are recorded in transmission or ATR mode by averaging 64 scans from

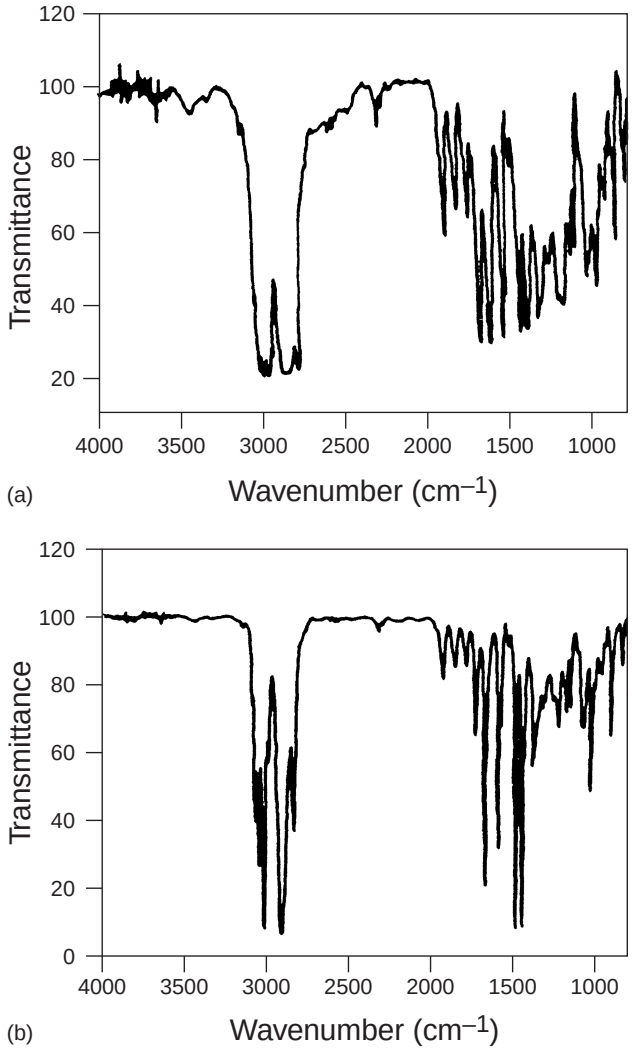

Figure 7 SBFTIRM spectrum of (a) a single Merrifield bead and (b) a flattened bead. (Reprinted from B. Yan, G. Kumaravel, *Tetrahedron*, **52**(3), 843–848 (1996), with permission from Elsevier Science Ltd.)

a single resin bead. The FTIR spectra of compounds obtained in this manner are complicated by absorbances from the resin matrix. The quality of single-bead IR spectra can be dramatically improved by flattening the beads to a thickness of 10–15 μm between a pair of NaCl windows.[39] Comparison of the FTIR spectra obtained on a single Merrifield resin bead (Figure 7a) and a flattened bead (Figure 7b) illustrates the improvements in band shape and spectral resolution that result from the transmission of IR energy through a short, uniform pathlength.

SBFTIRM is superior to NMR and MS analyses in that the technique is extremely sensitive and does not require cleavage of the compound from the solid support. SBFTIRM is compatible with resin beads as small as

50 μm in diameter, and the technique has a detection limit of approximately 100 pmol of bound compound.[40] In addition, single-bead analysis by SBFTIRM produces spectra that are representative of the larger population of resin beads in SPS.

SBFTIRM has been used to measure the rates of a variety of organic reactions on solid-phase supports.[41–44] These experiments have helped dispel a number of myths surrounding SPOS.[45] Among the results are the following:

- the rates of organic reactions on solid supports are comparable to their counterparts measured in solution
- positioning the reactive site away from the bead surface through attachment of poly(ethylene glycol) spacers has no significant impact on the rates of organic reactions relative to rates measured on conventional resins
- comparison of the rates of solid-phase reactions monitored by transmission mode and ATR-mode SBFTIRM indicate that reaction kinetics on the bead surface and the bead interior are similar.

Overall, the technique of SBFTIRM is rapid, information rich, and has provided a greater understanding of SPOS.

3.1.3 Diffuse Reflectance Infrared Fourier Transform

Diffuse reflectance infrared Fourier transform (DRIFT) employing micro-sample-cups was developed by Deben et al.[46] as an alternative to conventional methods of FTIR analysis of resin-bound compounds. The micro-DRIFT technique requires only one tenth of the amount of resin used to prepare KBr pellets, and is simpler to perform than SBFTIRM. Sample preparation involves grinding 0.5–1.0 mg of resin with 30 mg of KBr, transferring the mixture into a DRIFT microcup, and placing the microcup into a nitrogen-purged DRIFT cell in the sample compartment of the FTIR. The micro-DRIFT technique yields FTIR spectra that compare favorably with spectra obtained by SBFTIRM. This technique, however, is not compatible with the characterization of mixture libraries prepared by split synthesis.

3.1.4 Photoacoustic Fourier Transform Infrared

Gosselin et al.[47] reported a simple, rapid, nondestructive technique to monitor solid-phase organic reactions based on photoacoustic Fourier transform infrared (PAFTIR) spectroscopy. PAFTIR spectroscopy measures the pressure variations of an inert gas above a sample that result from the sample's absorption of modulated IR radiation. This absorbed radiation heats the boundary layer of gas above the sample and sets up a pressure wave that is detected by a sensitive microphone. Multiple scans are averaged, and the resulting spectrum is Fourier transformed to the frequency domain. In contrast to transmission, ATR, and DRIFT techniques, PAFTIR is not susceptible to interference from light scattering and reflection by the sample. Approximately 10 mg of resin is required to obtain a reasonable spectrum, and no additional sample preparation is required. The resin may be used in subsequent reactions following PAFTIR. PAFTIR is not compatible with split synthesis SPOS.

3.1.5 Fourier Transform Infrared of Pins

Characterization of compounds attached to polyethylene pins by ATR/FTIR has been described by Gremlich and Berets.[48] In this technique, an internal reflectance probe is placed on a pin and pressed down to ensure good contact with the pin surface. Spectra are acquired in ATR mode and signal averaging is employed to improve the signal-to-noise ratio. The technique requires no sample preparation, is fast and nondestructive. Additional organic reactions may be performed on the pin surface following FTIR analysis.

3.1.6 Continuous Reaction Monitoring

Pivonka et al. developed a flow-through cell that allows solid-phase organic reactions to be monitored in situ by FTIR.[49] The flow cell consists of a 50-μm thick Teflon® gasket sandwiched between two IR transparent windows as shown in Figure 8. Inlet and outlet ports allow organic solvents to be pumped through the cell. The entire cell is mounted on the $X-Y$ stage of an IR microscope and spectra of individual resin beads may be recorded in transmission or ATR mode. The apparatus allows the FTIR analysis of multistep syntheses to be performed in real time.

Sample preparation involves placing a few resin beads near the center of the Teflon® gasket with the aid of a small drop of fluorolube gel. The flow cell is then assembled and positioned on the IR microscope. A series of solvents and reagents are passed through the flow cell, and FTIR spectra of the resin bead are collected at regular intervals during the synthesis. This approach to reaction monitoring has three advantages over SBFTIRM. First, continuous monitoring of reactions in situ does not require additional sample preparation. Second, this technique is compatible with monitoring organic reactions with short half-lives. Lastly, analysis of the same bead throughout an experiment makes it possible to enhance the FTIR spectra of resin-bound

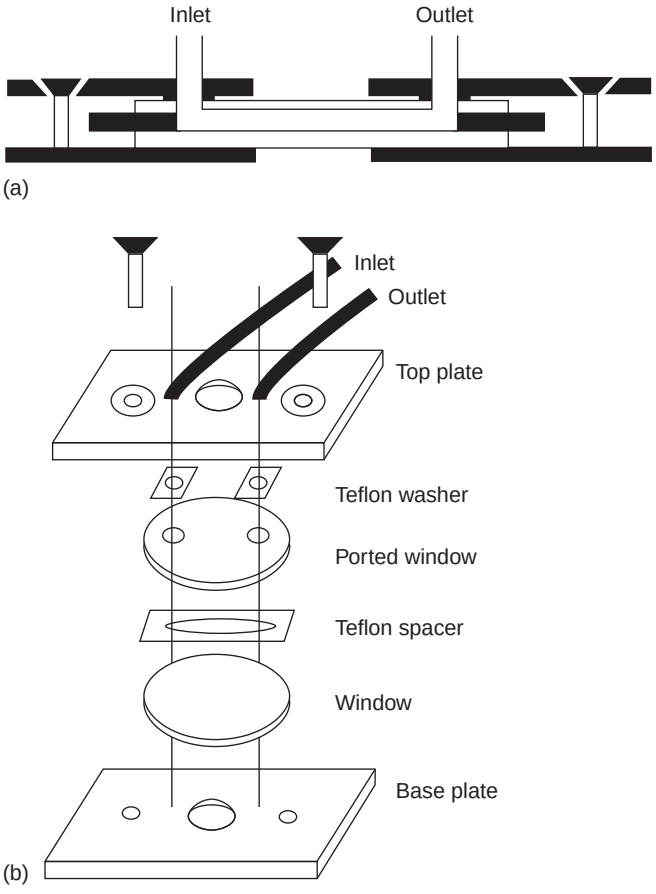

Figure 8 Flow cell designed for in situ reaction monitoring of SPOS by FTIR. Side view of flow cell (a); exploded view of cell assembly (b). (Reproduced by permission of the Society for Applied Spectroscopy from D.E. Pivonka, K. Russell, T. Gero, *Appl. Spectrosc.*, **50**(12), 1471–1478 (1996).)

products through signal averaging followed by subtraction of the blank resin spectrum.

3.2 Off-resin Analysis by Fourier Transform Infrared Spectroscopy

Development of MS and NMR-based techniques for characterizing solution-phase combinatorial libraries has overshadowed analysis of these types of samples by FTIR. Indeed, widespread application of FTIR in combinatorial chemistry has been limited to the analysis of solid-phase libraries. Considerable research during the 1980s and early 1990s led to the development of suitable interfaces for coupling HPLC with FTIR spectroscopy. Many of the approaches to high-performance liquid chromatography/Fourier transform infrared (HPLC/FTIR) could be automated to perform flow-injection FTIR, or HPLC/FTIR analysis of solution-phase libraries. A description of several flow-through and solvent-elimination HPLC/FTIR interfaces available is given next.

3.2.1 Flow-through Fourier Transform Infrared Interfaces

A flow-through HPLC/FTIR interface provides a direct link between the liquid chromatograph and the IR spectrometer. Effluent from an HPLC column passes through a flow cell mounted within the sample compartment of an IR spectrometer. This arrangement allows IR spectra of eluting components to be recorded in real time. No sample preparation is required and solutes may be recovered after analysis. The major disadvantages of flow-through interfaces are that they suffer from low sensitivity, and are prone to losses of solute information due to regions of solvent opacity in the mid-IR.

3.2.2 High-performance Liquid Chromatography/ Fourier Transform Infrared Flow Cells

The first flow cells designed for HPLC/FTIR consisted of a spacer sandwiched between a pair of NaCl or KBr windows through which the column effluent passed.[50–52] These early designs, however, were of little practical value for routine compound analysis. Many of the solvents used in liquid chromatography (LC) absorb across regions of the mid-IR and interfere with solute detection. In addition, the use of strong IR-absorbing solvents severely limits the cell pathlengths that can be tolerated, which further reduces solute sensitivity. The use of IR-transparent solvents, such as $CHCl_3$ and $CDCl_3$, as eluants in normal-phase HPLC improves solute detection by FTIR,[53] but limits the types of solutes that may be characterized by the technique. Microbore (1 mm i.d.) HPLC columns operate at lower flow rates than analytical (4.6 mm i.d.) columns, thereby increasing solute concentrations in the mobile phase 20-fold. The use of microbore columns can help overcome the sensitivity limitations associated with the use of short pathlength flow cells in HPLC/FTIR.[54] A schematic diagram of a zero dead volume flow cell designed for microbore HPLC/FTIR is shown in Figure 9.[55] This flow cell has a pathlength of 0.45 mm, an internal volume of 0.33 μL, and can detect less than 50 ng of 2,6-di-*tert*-butylphenol at 3641 cm^{-1}.

Flow cells designed for reversed-phase HPLC/FTIR are based on ATR technology and constructed with ZnSe cell windows. In contrast to the HPLC/FTIR flow cells, IR radiation passing through an ATR cell penetrates only a few wavelengths into the exterior sample region per internal reflection. Thus, the effective pathlength of ATR flow cells is short enough to keep interference from water absorption to an acceptable level. McKittrick et al.[56]

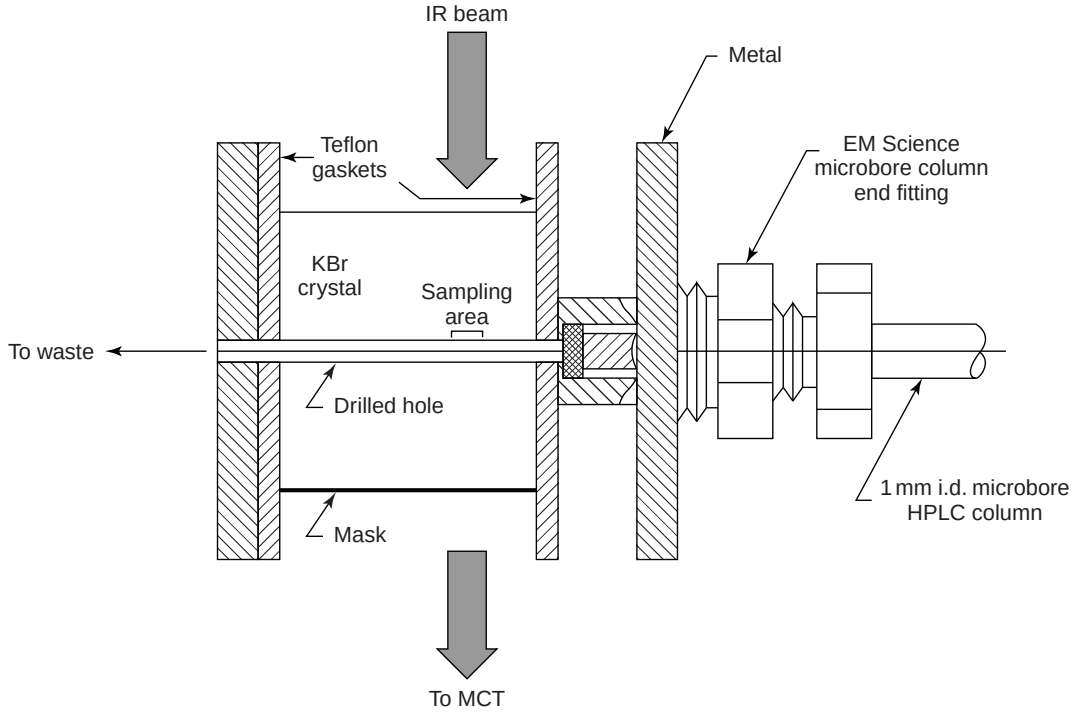

Figure 9 Schematic diagram of a zero dead volume flow cell for microbore HPLC/FTIR. (Reproduced with permission from C.C. Johnson, L.T. Taylor, *Anal. Chem.*, **56**, 2642–2647 (1984). Copyright 1984 American Chemical Society.)

compared the performance of micro and ultramicro cylindrical internal reflectance (CIRCLE) cells for the detection of caffeine and theophylline by reversed-phase HPLC/FTIR. The micro CIRCLE cell had an internal volume of 24 μL and an effective pathlength of 4.65 μm, while the ultramicro CIRCLE cell had an internal volume of 1.75 μL and an effective pathlength of 3.18 μm. These types of flow cell, however, suffer from poor sensitivity. The detection limits for caffeine with the micro and ultramicro flow cells were 0.1 mg and 0.5 mg, respectively.

3.2.3 Solvent-elimination Interfaces for High-performance Liquid Chromatography/ Fourier Transform Infrared

Solvent-elimination interfaces for HPLC/FTIR overcome the disadvantages of solvent opacity and poor solute sensitivity characteristic of flow-through interfaces by removing the liquid solvent from solutes of interest prior to IR spectroscopy. These interfaces are compatible with many modes of chromatography, including normal-phase, reversed-phase, and gradient HPLC. The solvent elimination process concentrates solutes prior to analysis, and hence achieves high sensitivities with FTIR spectroscopy. The major disadvantages of solvent-elimination interfaces stem from their complexity and incompatibility with real-time analysis. Furthermore, volatile solutes may be lost during desolvation, and thermally labile compounds are susceptible to degradation. When used with aqueous solvents, these interfaces cannot tolerate flow rates greater than 0.5 mL min^{-1}. Three solvent-elimination interfaces for HPLC/FTIR are described below.

3.2.4 The High-performance Liquid Chromatography/ Diffuse Reflectance Infrared Fourier Transform Interface

The first high-performance liquid chromatography/diffuse reflectance infrared Fourier transform (HPLC/DRIFT) interface was developed by Kuehl and Griffiths[57] in the late 1970s. In this technique, the mobile phase exiting the column is deposited into a series of diffuse-reflectance cups filled with KBr. The chromatographic solvent is allowed to evaporate, and the DRIFT spectra of the solute residue on the KBr substrate are recorded. The HPLC/DRIFT technique provides greater solute sensitivity than flow-through interfaces,[58] but does not allow continuous analysis of the chromatographic eluant. Further, the quality of DRIFT spectra is dependent on the particle size of the KBr substrate, which can change during solvent evaporation. Conventional HPLC/DRIFT interfaces are not compatible with reversed-phase HPLC, because the KBr substrate is water soluble. However, Kalasinsky et al.[59] demonstrated that water could be eliminated from the mobile phase through post-column treatment with 2,2-dimethoxypropane. Water reacts with

2,2-dimethoxypropane stoichiometrically to yield two moles of methanol and one mole of acetone. Alternatives to chemical removal of water from reversed-phase eluants include: the use of industrial-grade diamond powder as a diffuse reflectance substrate;[60] solute extraction in methylene chloride prior to solute deposition on KBr;[61] and aqueous solvent removal with a heated nebulizer.[62]

3.2.5 Mono-disperse Aerosol Generation Interface for Combining High-performance Liquid Chromatography with Fourier Transform Infrared

The mono-disperse aerosol generation interface for combining high-performance liquid chromatography with Fourier transform infrared (MAGIC/HPLC/FTIR) was adapted from a similar device for coupling HPLC with MS.[63] A schematic of an HPLC system equipped with this interface is shown in Figure 10.[64] During operation, the HPLC mobile phase exiting the column is converted to a uniform stream of droplets in the aerosol generator. Dispersion of the droplets is achieved with the aid of a stream of helium gas. Solvent is removed by evaporation at ambient temperature, and solvent vapor and helium gas are removed from the stream of solute particles under vacuum. Solute particles are collected on the surface of an IR-transparent collection plate, which is moved continuously during analysis. IR spectra of the deposited solutes are collected off-line. MAGIC/HPLC/FTIR can remove solvents containing up to 100% water at flow rates of up to 1.0 mL min^{-1}.[65,66] The interface is compatible with the analysis of thermally labile compounds, and as little as 10–100 ng of deposited compound can be detected by FTIR.

3.2.6 The Reflectance–Absorbance Interface

Gagel and Biemann developed the continuously recording interface for HPLC/FTIR in the mid-1980s.[67,68] An illustration of the interface is shown in Figure 11. The effluent from an HPLC column is continuously sprayed onto the surface of a slowly rotating aluminum mirror. Solvent is evaporated by means of a stream of heated nitrogen. Solutes are deposited onto the surface of the aluminum mirror in a spiral pattern. The rotation speed of the reflective surface and the efficiency of solvent elimination govern the resolution between deposited solute bands. Reflectance–absorbance FTIR spectra of the deposited solute bands are collected off-line using a separate reading device mounted in the sample compartment of the IR spectrometer (see Figure 11c). Spectra obtained in this manner are susceptible to Christiansen scattering (i.e. a sudden rise on one side of an absorption band and a tailing on the other side) depending on the particle size of the deposited solute and refractive index changes that may occur in the vicinity of an absorption band. A representative FTIR spectrum of anthracene is shown in Figure 12.

A continuously recording interface for HPLC/FTIR is available commercially. The commercial instrument replaces the heated gas nebulizer in Gagel and Biemann's original design with an ultrasonic nebulizer and vacuum desolvation chamber, and solutes are collected on an aluminum-backed germanium disc instead

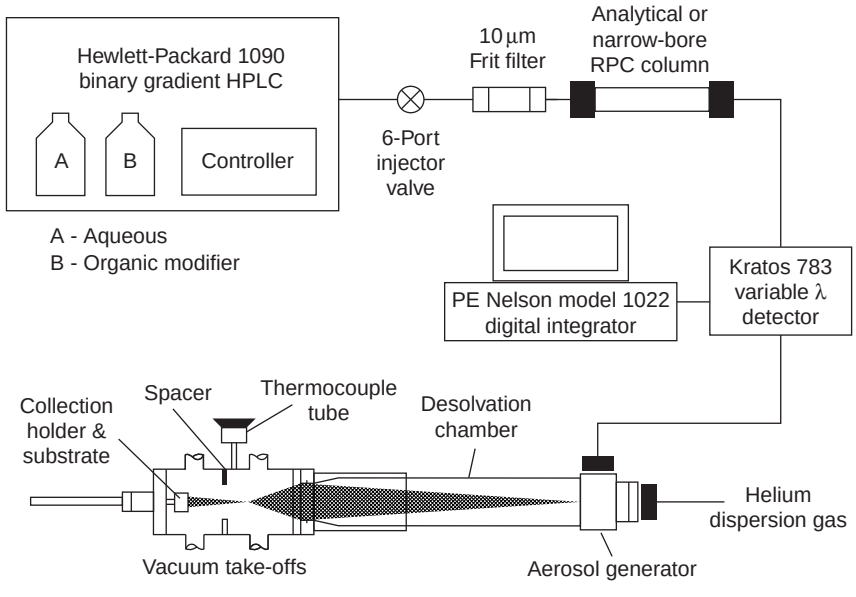

Figure 10 MAGIC/HPLC/FTIR. (Reproduced with permission from V.E. Turula, J.A. de Haseth, *Anal. Chem.*, **68**, 629–638 (1996). Copyright 1996 American Chemical Society.)

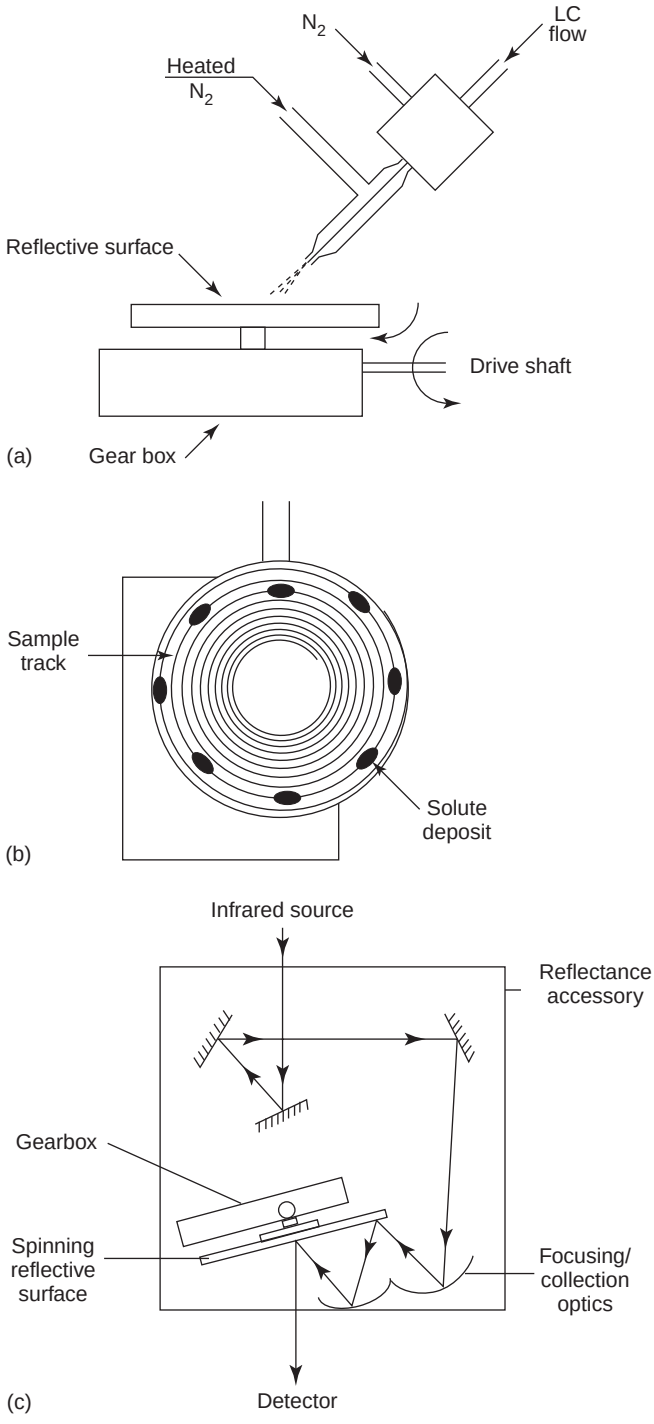

Figure 11 Schematic diagram of a continuously recording interface for HPLC/FTIR illustrating (a) the recording device, (b) deposited solutes, and (c) the reading device. (Reproduced with permission from J.J. Gagel, K. Biemann, *Anal. Chem.*, **58**, 2184–2189 (1986) and J.J. Gagel, K. Biemann, *Anal. Chem.*, **59**, 1266–1272 (1987). Copyright 1986, 1987 American Chemical Society.)

of an aluminum mirror. Detection limits are in the 10–100 ng range. This interface is compatible with both

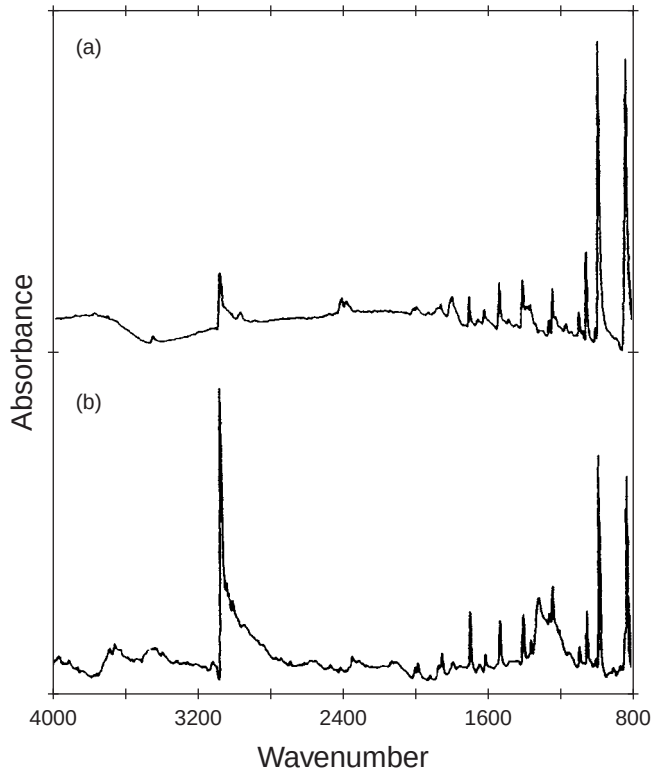

Figure 12 IR spectra of anthracene: (a) transmission spectrum taken in a KBr pellet, and (b) that obtained using a continuously recording interface for HPLC/FTIR. (Reproduced with permission from J.J. Gagel, K. Biemann, *Anal. Chem.*, **58**, 2184–2189 (1986). Copyright 1986 American Chemical Society.)

normal- and reversed-phase HPLC, but only volatile buffers may be used in the mobile phase. Somsen et al.[69] incorporated a liquid–liquid extraction step upstream of solute deposition to make the interface compatible with the use of nonvolatile buffers (e.g. 10 mM potassium phosphate) in reversed-phase eluants.

4 LIBRARY CHARACTERIZATION BY MASS SPECTROMETRY

MS is one of the most sensitive analytical techniques available for the characterization of organic compounds. Modern mass spectrometers are capable of detecting low-molecular-weight organic compounds in the picomole to femtomole range. In addition to its high sensitivity, MS provides important molecular-weight and structural information that can be used to monitor reactions, or confirm the identity of final products. Specific compounds may be readily confirmed by noting the presence of characteristic molecular ions in a sample's mass spectrum. The inherent simplicity of MS data makes it possible to develop sophisticated algorithms to interpret

the mass spectra of combinatorial libraries in an efficient manner. Indeed, a number of commercial software packages are available for processing MS data and reporting sample results in condensed form. The development of similar tools for FTIR and NMR data reduction has languished due to the complexities involved in interpreting these types of spectra. Together these attributes have contributed to the widespread acceptance of MS among combinatorial chemists.

4.1 On-resin Analysis by Mass Spectrometry

MS can only detect ions in the gas phase. Compounds originating from SPS must be cleaved from the surface of the solid support prior to MS analysis, and the cleavage step can take up to 24 h to perform. From this perspective, MS is not as convenient a technique for reaction monitoring in SPS as SBFTIRM and magic-angle spinning (MAS) NMR spectroscopy. Nonetheless, the high sensitivities that can be achieved with MS make the technique ideally suited to the analysis of single resin beads. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOFMS) has emerged as the MS technique of choice for characterizing peptides and low-molecular-weight organic compounds covalently bound to solid-phase supports.

4.1.1 Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry

Preparation of resin-bound samples for MALDI/TOFMS is tedious and difficult to fully automate. Before compounds obtained from SPS can be analyzed, they must be cleaved from the solid support and combined with a suitable ultraviolet (UV) absorbing matrix. The matrix is required for the matrix-assisted laser desorption/ionization (MALDI) process, and sample ionization efficiencies can be highly matrix dependent. Typical matrices employed in MALDI/TOFMS include 2,5-dihydroxybenzoic acid, α-cyano-4-hydroxycinnamic acid, and 3,5-dimethoxy-4-hydroxycinnamic acid. Procedures for cleaving a resin-bound compound depend on the nature of the linker employed in SPS. Compounds immobilized through acid-labile linkers can be removed by treating the resin with a solution of 1–2% trifluoroacetic acid (TFA) in dichloromethane. A more convenient method is to expose the resin beads to the acidic vapor produced from a solution of 15% TFA in dichloromethane for 3–5 min. Base-labile linkers can be cleaved using 1–2% ammonia in tetrahydrofuran (THF). Photo-labile linkers require treatment with UV light to cleave resin-bound compounds.

The application of MALDI/TOFMS to the characterization of peptide libraries has been reported by a number of research groups.[70–74] High-quality MALDI/TOFMS can be obtained from as little as 2–5% of the total peptide isolated from a single resin bead. In addition, MALDI/TOFMS may be used to assess the quality of a library by revealing the presence of deletion peptides, the products of side reactions and incomplete-deprotection products. Analysis of peptide libraries by imaging time-of-flight secondary ion mass spectrometry (TOF/SIMS) has also been reported.[75]

The characterization of nonpeptide bead-bound combinatorial libraries by MALDI/TOFMS requires the same sample preparation involved in the analysis of peptide libraries. Zambias et al.[76] characterized a peptoid attached via a Rink amide linker to a polymeric bead by MALDI/TOFMS employing 2,5-dihydroxybenzoic acid as a matrix. This same matrix was employed by Haskins et al.[77] to characterize the individual components of a five-member combinatorial library. Brummel et al.[78] reported that the MALDI/TOF mass spectrum of a angiotensin II antagonist (see Figure 13) synthesized on a polystyrene resin compared favorably with the electrospray ionization (ESI) mass spectrum of the same compound. Quantitative MALDI/TOFMS has been applied to study the incorporation of combinatorial libraries of porphyrins and peptide–DNA hybrids into liposome membranes.[79] An automated system for the analysis of combinatorial samples by matrix-assisted laser desorption/ionization Fourier transform mass spectrometry (MALDI/FTMS) was developed by Tutko et al.[80] The automated MALDI/FTMS instrument allows the characterization of 20 samples per hour and can be used for accurate mass determinations and MS/MS.

4.2 Off-resin Analysis by Mass Spectrometry

MS is typically applied to the characterization of libraries of compounds that have been cleaved from solid-phase resins, or that were synthesized in parallel by solution-phase techniques. Demands for high mass resolution, sensitivity, and speed of analysis place a number of constraints on instrument design. Automation requires that the samples be swept into the mass spectrometer in a liquid stream, as exemplified by flow injection analysis mass spectrometry (FIA/MS) and liquid chromatography/mass spectrometry (LC/MS) techniques. The design of interfaces that can both accommodate the introduction of analytes in solution and generate gas-phase ions suitable for mass analysis and detection is central to the successful application of MS in combinatorial chemistry.

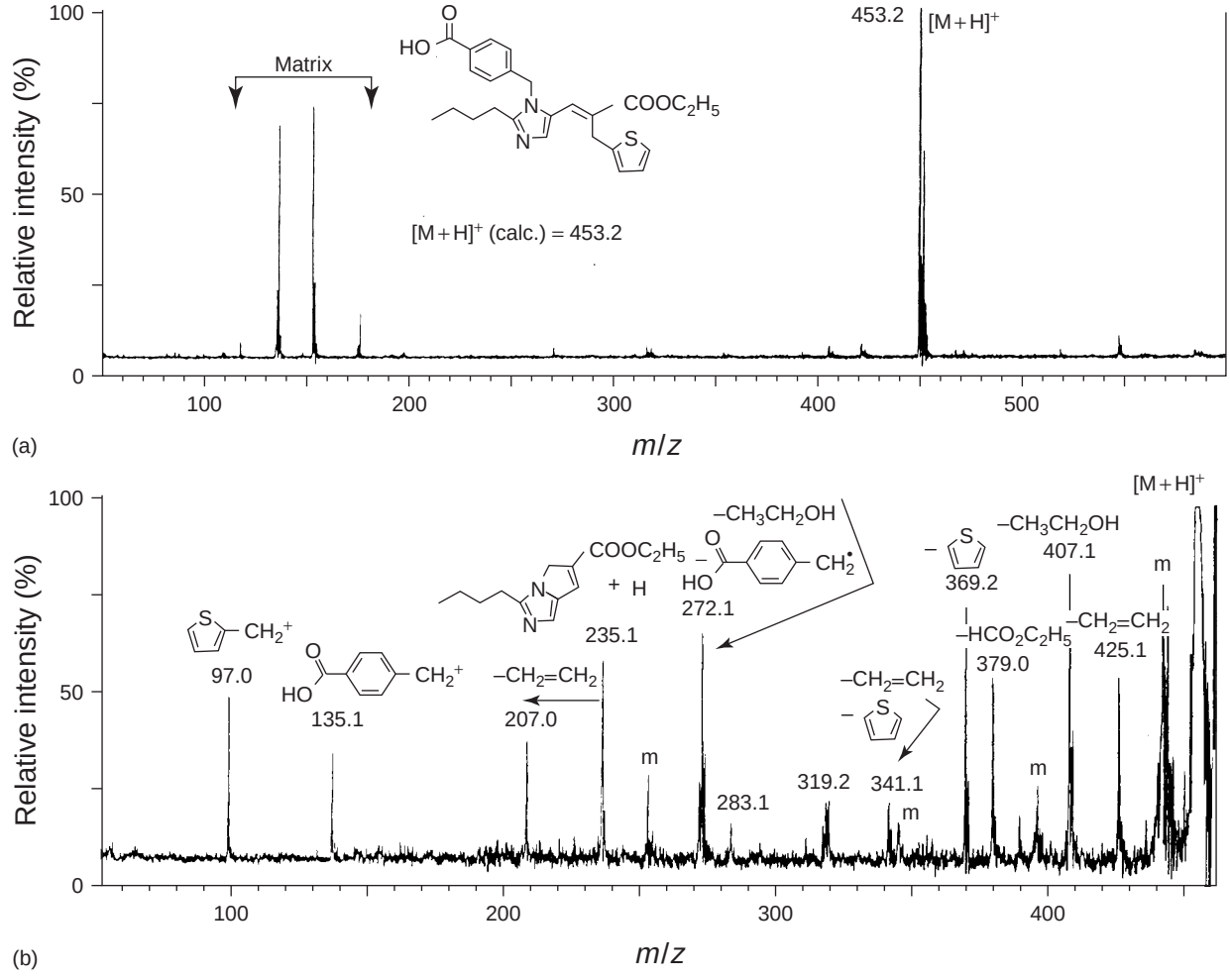

Figure 13 (a) MALDI mass spectrum and (b) metastable ion MALDI mass spectrum of an angotensin II antagonist. (Reproduced with permission from C.L. Brummel, J.C. Vickerman, S.A. Carr, M.E. Hemling, G.D. Roberts, W. Johnson, J. Weinstock, D. Gaitanopoulos, S.J. Benkovic, N. Winograd, *Anal. Chem.*, **68**, 237–242 (1996). Copyright 1996 American Chemical Society.)

4.2.1 Liquid Chromatography/Mass Spectrometry Interfaces

All LC/MS interfaces are designed to perform two basic tasks. First, they provide a mechanism for stripping solvent molecules from solutes present in the sample. These solute molecules are then converted into gas-phase ions prior to entering the mass analyzer. Depending on the design of the interface, the desolvation and ionization processes may occur under conditions of low vacuum or atmospheric pressure. The two most common sample introduction techniques used in the analysis of libraries of compounds present in solution are ESI and atmospheric pressure chemical ionization (APCI).

In the ESI process, the liquid sample is pumped through a narrow-bore capillary whose outlet is held at a potential of a few kilovolts relative to an opposing counter-electrode. A volatile electrolyte generally is introduced into the sample to increase the sample's ionic strength. The liquid becomes charged as it emerges from the electrospray capillary, and coulombic repulsion causes the stream to spread out as a plume of charged droplets. The mechanism leading to ion formation in electrospray is not well understood, but may be regarded as a process wherein charged droplets decrease in size through both evaporation and coulombic repulsive forces. Eventually these highly charged droplets reach a radius at which the electrical field gradient is large enough for ionized solute molecules to desorb from the droplet surface. The probable mechanisms involved in ESI have been described by Fenn[81,82] and Kebarle and Tang.[83]

A schematic diagram of a liquid chromatography/electrospray ionization (LC/ESI) interface is shown in Figure 14(a). ESI takes place at atmospheric pressure, and the sample ions formed are directed through a small orifice and a series of skimmers into the vacuum region of the mass spectrometer. An annular sheath gas surrounding the spray needle may be applied to enhance

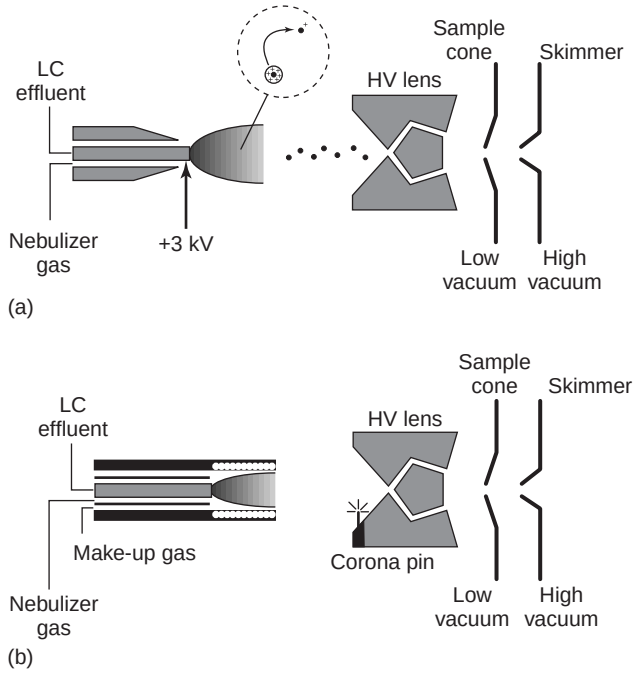

Figure 14 Schematic diagram of (a) a LC/ESI interface for MS, and (b) an LC/APCI interface.

the desolvation process. The electrospray process is a soft ionization technique, yielding primarily $[M + H]^+$ or $[M - H]^-$ ions even with thermally labile and nonvolatile molecules. In addition, ESI can generate multiply charged ions, which makes the technique well suited to the analysis of biomolecules. Electrospray interfaces are generally operated at flow rates of 1–200 μL min^{-1}, with well-designed interfaces capable of flows as high as 1 mL min^{-1}. Liquid chromatography coupled with electrospray ionization mass spectrometry (LC/ESIMS) is generally performed using narrow-bore (2 mm i.d.) or microbore (1 mm i.d.) columns. If separations are carried out on analytical-scale (4.6 mm i.d.) columns, the column effluent usually is split to reduce the flow rate entering the electrospray interface. Coupling ESIMS with LC has been described in a number of review articles.[84,85]

In APCI, a heated nebulizer is used to remove solvent from the solute prior to the ionization step. As its name implies, sample ionization in APCI takes place at atmospheric pressure. A schematic diagram of a representative atmospheric pressure chemical ionization mass spectrometry (APCIMS) interface is shown in Figure 14(b). The liquid sample enters the APCI interface through a narrow capillary around which a concentric flow of nitrogen gas is provided to pneumatically generate a fine aerosol. An additional flow of nitrogen makeup gas helps direct the sample aerosol towards the heated region of the APCI probe. The temperature of the probe is maintained such that effective desolvation and solute volatalization are achieved. The vaporized aerosol then passes by a corona discharge needle held at a potential of 2 kV, where chemical ionization of the solute is achieved in the surrounding solvent vapor at atmospheric pressure. The solvent vapor serves as a reagent gas for the formation of the reactant ions necessary in the chemical ionization process. The solute ions produced through ion–molecule reactions with the reagent gas then enter the intermediate pressure region of the mass spectrometer through a sampling cone. The sampling cone is maintained at 10–100 V to help dissociate analyte–solvent cluster ions before the analyte ions pass through a skimmer on their way to the mass analyzer.

The transport regions used in APCI and ESI interfaces are nearly identical in design, and most modern systems allow rapid conversion between the two interface types.[86] APCI interfaces are compatible with liquid flow rates as high as 1–2 mL min^{-1}. As with ESI, APCI generally produces little fragmentation, although sample fragmentation can be induced where desired by increasing the voltage of the sampling cone. The two ionization techniques differ in the greater abundance of multiply charged ions that are observed in ESI mass spectra.

4.2.2 Electrospray Ionization Mass Spectrometry

ESIMS is well suited to the analysis of low-molecular-weight (e.g. 10–1000 Da) organic compounds containing ionizable functional groups and large molecules (e.g. $>10^3$ Da) with moderate to high polarity. It is not surprising, therefore, that ESIMS is the ionization technique most often applied to the MS characterization of combinatorial libraries. Analysis of nonpeptide libraries by ESIMS has focused mainly on the characterization of single compounds cleaved from individual resin beads,[87–89] or single compounds obtained from parallel synthesis.[90,91] The ESIMS spectra of low-molecular-weight organic compounds generally are characterized by singly charged molecular ions, and multiply charged ions are only rarely observed. Compounds containing basic functional groups yield intense positively charged molecular ions with little fragmentation. Hydrophobic molecules are more difficult to ionize by addition of a proton in ESIMS, and hence have a higher tendency to fragment.

Dunayevskiy et al.[92] employed ESIMS to study the composition of six sub-libraries synthesized to mimic the diversity present in a 65 341 component mixture library. Analysis of the 36–55 member sub-libraries provided qualitative information on the yields of synthetic reactions used to generate the mixture library, and hence insight into the library's actual composition. This work illustrated a number of difficulties that may be encountered when

characterizing mixtures of small organic molecules by ESIMS. First, peak intensities do not directly correlate with sample concentration due to differences in ionization efficiency among organic compounds present in mixtures. Furthermore, polar solvents such as dimethyl sulfoxide (DMSO), used to prepare combinatorial libraries for biological testing, have a tendency to form intense adduct ions with library compounds. Complete suppression of product molecular ions in a combinatorial library by DMSO adduct ions has been reported by Kyranos and Hogan.[93] Separation of the solute from the solvent by reversed-phase chromatography prior to MS analysis is an effective means of eliminating DMSO adduct formation.

The likelihood of a small-molecule mixture library containing isobaric compounds increases as the population of the library grows, hence direct application of ESIMS with nominal mass resolution to large mixture libraries is not practical. The instrumentation and data precision requirements for the MS determination of structures from combinatorial mixtures have been described by Blom.[94] In addition, Hughes[95] devised a rules-based strategy for designing combinatorial syntheses such that libraries of low-molecular-weight organic molecules always contain unique nominal product masses.

ESIMS is an extremely useful technique for the characterization of peptide libraries containing mixtures. Metzger et al.[96,97] were the first to apply ion-spray and ESIMS to the analysis of synthetic peptide libraries. Side-reactions during SPPS can lead to peptide libraries that deviate significantly from expected molar composition. MS is ideally suited to peptide analysis, because it can readily distinguish between peptides differing by as little as one amino acid. While a number of MS ionization techniques are available for peptide analysis, many suffer from ion suppression effects. Ionization of some components present in a mixture may be partially or completely suppressed, even though the same compounds yield intense ions when analyzed separately. Ionization yields for individual peptides present in mixtures with ESIMS, however, have been reported to be nearly identical.[98] Lambert et al.[99] reported that a single peptide possessing a unique mass may be identified from a library containing up to 100 000 peptides using a combination of LC/ESIMS with selected ion monitoring. The feasibility of characterizing peptide mixtures by MS has been explored mathematically using computer simulations of mass spectra for peptides containing 2–7 amino acids.[100] ESIMS also has been used to automate the purity assessment of synthetic peptides.[101] The total ionizable material present in a sample is compared with the intensity of the target peptide and each of the identifiable contaminants present in order to arrive at an estimation of sample purity.

4.2.3 Atmospheric Pressure Chemical Ionization Mass Spectrometry

Moderately polar compounds within a mass range of 100–1000 Da ionize readily by APCI. APCI generally produces greater fragmentation and fewer multiply charged ions than ESI does. While fragment ions in a mass spectrum can complicate spectrum interpretation, their presence often helps confirm structure identity. These characteristics make APCI well suited to the identification of nonpeptide organic compounds. The ionization process in APCI can produce both positively and negatively charged molecular ions, and predicting the relative intensities of the two for a given compound can be difficult. Iwabuchi et al.[102] examined the intensities of the molecular ions observed in positive-ion and negative-ion mode APCI for a series of hydroxymethylglutaryl–Coenzyme A reductase inhibitors. Lipophilic compounds in the series gave more intense molecular ions in positive-ion mode, while hydrophilic compounds gave more intense negatively charged molecular ions. Further, it is not often easy to predict whether APCI or ESI will be better suited to a given combinatorial library. Hence the designs for automated systems for high-throughput library characterization often make provision for both types of ionization technique in conjunction with positive and negative ion detection.[103] Positive-ion mode APCIMS has been applied to the characterization of a variety of organic compound classes, including quinolones,[104] pyrazoles,[105] and dihydropyridones.[106]

4.2.4 Tandem Mass Spectrometry

MS/MS is based on a simple fragmentation experiment in which a parent molecular ion dissociates into a corresponding product ion and neutral fragment molecule. The parent molecular ion may be positively or negatively charged, and the product ions produced have the same charge as the parent ion. MS/MS historically has been performed using triple quadrupole instruments, wherein the first and third quadrupoles are used to scan parent and product ion masses, and the intermediate quadrupole is employed as a collision chamber. Fragmentation of parent ions emitted from the first quadrupole is achieved through a process known as collision-induced dissociation (CID). Argon, nitrogen, methane, or ammonia gas are introduced individually or as mixtures into the second quadrupole. Collisions between parent ions and target gas molecules increase the internal energy of the parent ions. This collisional activation is followed by fragmentation into product ions and neutral fragment molecules. MS/MS may also be performed on ion-trap mass spectrometers.[107] When carried out on ion-trap instruments, these experiments involve tandem

operations over time in the same space, whereas MS/MS experiments on triple quadrupole instruments involve tandem operations in different regions of space.

There are three basic types of MS/MS experiments: the product-ion scan, the neutral-loss scan, and selected reaction monitoring. A description of each of these experiments carried out on a triple-quadrupole instrument is given below. The product-ion scan provides a fragmentation mass spectrum of a single parent ion. The m/z of the parent ion is fixed by the first quadrupole, and the masses of all of the product ions formed from CID of the parent in the second quadrupole are scanned by the third quadrupole. A product-ion scan can be used to distinguish between multiple components introduced directly into the mass spectrometer as a mixture, or as a single band eluted during a chromatographic separation. A neutral-loss scan involves setting the first and third quadrupoles such that they scan for a constant difference in mass between the parent and product ions. The first quadrupole passes a select range of incoming parent ions into the collision region defined by the second quadrupole. Of the product ions generated through CID, only those with a specific mass difference from their respective parent ions are scanned by the third quadrupole. The mass difference between scans of the first and third quadrupoles equals the mass of the neutral fragment from CID of the parent ions. The neutral-loss experiment is used to confirm the presence of specific functional groups that are known to yield characteristic neutral loss molecules, such as CO_2 from carboxylic acids, or NO radicals from nitroaromatic compounds. A selected reaction monitoring experiment is performed by fixing the first quadrupole to pass a single parent ion to the collision chamber, and setting the third quadrupole to pass only one product ion fragment. While selected reaction monitoring provides less structural information than the other two MS/MS experiments, the technique delivers greatly enhanced sensitivity. The sensitivity enhancement results from a decrease in instrument noise, and the selected reaction monitoring experiment is best suited to the analysis of trace amounts of compound, such as single-bead characterization.

While MS/MS experiments can be performed on ion-trap instruments, the processes involved differ from those executed on triple-quadrupole mass spectrometers. For example, a product-ion scan carried out on an ion-trap mass spectrometer involves scanning the RF voltage to eject all ions in the trap, up to but not including the ion of interest. These trapped ions then undergo collision with a background gas inside the ion-trap chamber. The product-ion spectrum is recorded by ramping the RF voltage a second time to eject product ions from the trap. A distinct advantage of ion-trap instruments is that a product ion from the initial CID can be trapped and subjected to an additional CID cycle, and this process may be repeated through many generations of product ion. The ability to perform such MS^n experiments on ion-trap mass spectrometers provides the combinatorial chemist with a powerful tool for structure elucidation.

MS/MS has a number of possible applications to combinatorial chemistry, including structure characterization, identification of side products, deconvolution of isobaric mixtures, single-bead analysis, and library decoding. Keough et al.[108] employed a hybrid magnetic sector/orthogonal acceleration single-pass time-of-flight (TOF) mass spectrometer to measure the exact masses of product ions formed during MS/MS studies of compounds

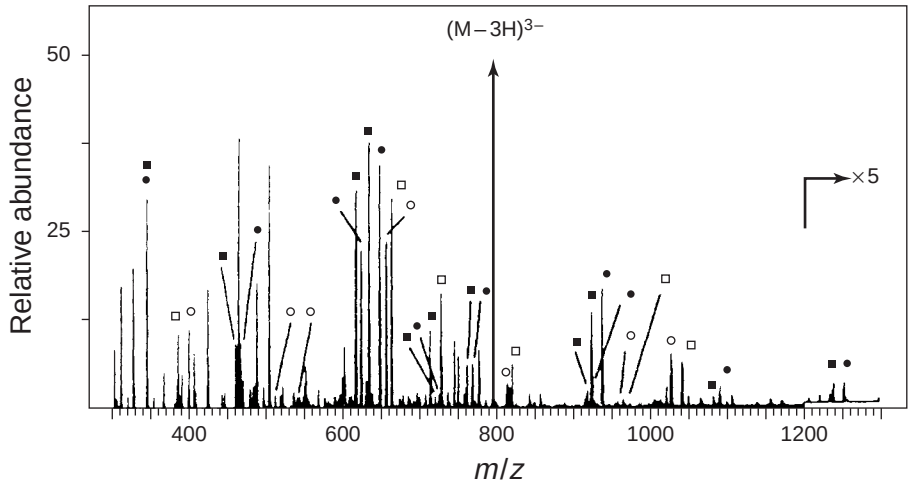

Figure 15 Product-ion mass spectrum of two isomeric 8-mer sequences, UGACACTG ●, w series ions; ○, a-base series ions and TGACACUG (■, w series ions; , α-base series ions). (Reproduced with permission from S.C. Pomerantz, J.A. McCloskey, T.M. Tarasow, B.E. Eaton, *J. Am. Chem. Soc.*, **119**, 3861–3867 (1997). Copyright 1997 American Chemical Society.)

from combinatorial libraries. In this work, the combination of exact mass measurement with MS/MS proved crucial to characterizing a major side-product present in a combinatorial sample, and differentiating between isobaric residues in peptides. Pomerantz et al.[109] developed a MS/MS procedure that allows sequence analysis of mixtures of closely related oligonucleotides. Oligonucleotide sequencing involves performing a product-ion scan on each of the parent ions observed in the sample mixture. Mixtures containing oligonucleotides with the same mass yield product-ion mass spectra with partially overlapping signals. Proper interpretation of these spectra requires application of special algorithms, which are based on fragment-ion abundance patterns observed in sequencing of single oligonucleotides. The product-ion spectrum of two isomeric oligonucleotides is shown in Figure 15.[109] MS/MS has also been used to decode secondary amine tags cleaved from resin beads following SPOS.[110] In this example, a separation step was required to resolve isobaric masses prior to MS/MS in parent-ion scanning mode. Dansylated secondary amine tags could be detected unambiguously at the femtomole level, making the technique compatible with decoding of single resin beads.

5 LIBRARY CHARACTERIZATION BY NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

NMR spectroscopy is an information-rich technique for probing the structure of organic compounds. NMR spectroscopy can be used to differentiate between most structural, conformational, and optical isomers. The chemical shift information and coupling constant parameters provided by ^{1}H NMR allow specific protons within an organic molecule to be assigned unambiguously. By performing heteronuclear ^{1}H/^{13}C shift and homonuclear shift correlated experiments, the exact structure of an organic molecule may be determined. In contrast to MS, NMR spectroscopy is a noninvasive and nondestructive technique. A sample may be recovered after a NMR experiment and used in subsequent analyses.

5.1 On-resin Analysis by Nuclear Magnetic Resonance Spectroscopy

Because it is a powerful tool for structural characterization of small organic molecules, there has been significant interest in the application of NMR spectroscopy to combinatorial chemistry. However, conventional NMR techniques are only compatible with homogeneous samples. Hence, characterization of compounds originating from SPS by standard NMR requires that the products be cleaved from the solid support prior to analysis. This presents a significant barrier to the characterization of resin-bound intermediates and reaction monitoring by NMR. Two sample-handling techniques that are amenable to the NMR analysis of resin-bound compounds are standard NMR using gel-phase samples, and MAS NMR. These gained wide acceptance within the combinatorial chemistry community during the 1990s and are discussed below.

NMR spectroscopy is most effective when the spectra of organic compounds possess good spectral resolution (i.e. narrow linewidths). There are two major factors that work to broaden the linewidths observed in NMR. The first of these is magnetic susceptibility inhomogeneity. All compounds possess an inherent magnetic susceptibility, which is a measure of how the compound affects a surrounding magnetic field. In an NMR experiment, a homogeneous sample will produce a uniform change in the applied magnetic field, and this can be corrected by a process called ''shimming''. A resin-bound compound, however, is heterogeneous, and possesses regions of differing magnetic susceptibility that cannot be corrected by shimming. These magnetic susceptibility mismatches result in broadening of the normal NMR linewidths. Second, significant rotational freedom is required to overcome the dipolar splittings and chemical shift anisotropy present in rigid molecules.

The quality of NMR spectra obtained with solid-phase resins depends primarily on the resin structure, with choice of solvent playing a secondary role.[111,112] Many of the resins used in SPS are based on cross-linked polystyrene that is functionalized with linker molecules attached to the polymer surface through tethers of varying length. The length of the tether influences the mobility of compounds attached to the resin. Tentagel™ resins have the longest tethers, and hence provide the narrowest observed NMR line widths. Wang type resins have short tethers and restrict the mobility of attached compounds. Compounds attached directly to polystyrene with no tether produce NMR spectra of poor quality. The shorter the tether on the polymer support, the more important proper choice of solvent becomes to produce quality NMR spectra. Resins swell to different degrees in various solvents, and the greater the extent of swelling the narrower the NMR line widths will be. Chloroform-*d*, dichloromethane-d_2, benzene-d_6, dimethylformamide-d_7, and THF-d_8 all provide significant swelling of polystyrene-based resins. Dichloromethane-d_2 is generally preferred as a solvent because it provides the highest degree of swelling and its high volatility facilitates recovery of the resin after analysis. Methanol-d_4, D_2O, acetonitrile-d_3, and dimethylsulfoxide-d_6 do not appreciably swell polystyrene-based resins and are poor solvent choices for monitoring SPS by NMR.

5.1.1 Gel-phase Nuclear Magnetic Resonance

While ^{1}H NMR is sufficiently sensitive to permit rapid analysis of solid-phase reactions, the signals obtained from resin samples by the gel-phase approach are typically too broad to be useful for structural determinations. Standard NMR of gel-phase samples is generally limited to the detection of heteronuclei such as ^{13}C, ^{19}F, ^{15}N, and ^{31}P, where there is significantly greater chemical shift dispersion than in ^{1}H NMR. The linewidths observed in ^{1}H NMR of gel-phase samples are typically greater than 100–300 Hz.

Gel-phase NMR spectra are obtained using conventional NMR sample tubes and high-resolution NMR spectrometers. Sample preparation involves placing between 50 and 100 mg of resin in a 5-mm NMR tube and adding sufficient deuterated solvent (0.2–0.3 mL) to cover the solid sample. Sonication or vortexing of the tube's contents may be necessary to achieve a uniform gel. Solvent molecules entrained in a different environment within the solvent-swollen polymer matrix can lead to a second set of resonances in the NMR spectra of the gel samples. This can be corrected by using a significant excess of solvent, or by preswelling the resin with the solvent prior to placing it into the NMR tube.

Gel-phase ^{13}C-NMR was employed extensively during the 1970s and early 1980s as an analytical technique to follow SPPS. Sternlicht et al.[113] pioneered gel-phase ^{13}C-NMR to characterize amino acids bound to cationic exchange resins. Leibfritz et al.[114] employed gel-phase ^{13}C-NMR in conformational studies on polypeptides bound to poly(oxyethylene) supports. The first application of gel-phase NMR to follow resin-bound intermediates in SPS was reported by Epton et al. in 1980.[115] In this work, the step-wise synthesis of polypeptides on a phenolic cross-linked poly(acryloylmorpholine)-based matrix was monitored by recording the ^{13}C-NMR spectra of the resin following purification by recrystallization. The choice of the resin matrix was based on its superior swelling characteristics over poly(styrene) matrices in polar organic solvents, such as DMSO, that solubilize peptides well. Giralt et al.[116] followed the step-wise synthesis of the peptide H–Asn–(N–Me)–Ala–Thr–NH_2 on a benzhydrylamine resin by gel-phase ^{13}C-NMR. Both Epton et al. and Giralt et al. reported that narrow signals are obtained for carbons that are further away from the polymer backbone, while the NMR signals for carbons close to the backbone broaden and diminish in intensity as the peptide chain grows. The latter is attributed to decreased mobility of nuclei nearer the support.

Application of gel-phase NMR in SPOS was first reported by Manatt et al.[117] in 1980. In 1982, Jones et al.[118] employed gel-phase ^{13}C-NMR to characterize resin-bound intermediates from the SPOS of insect pheromones. As with gel phase ^{13}C-NMR of peptides, the

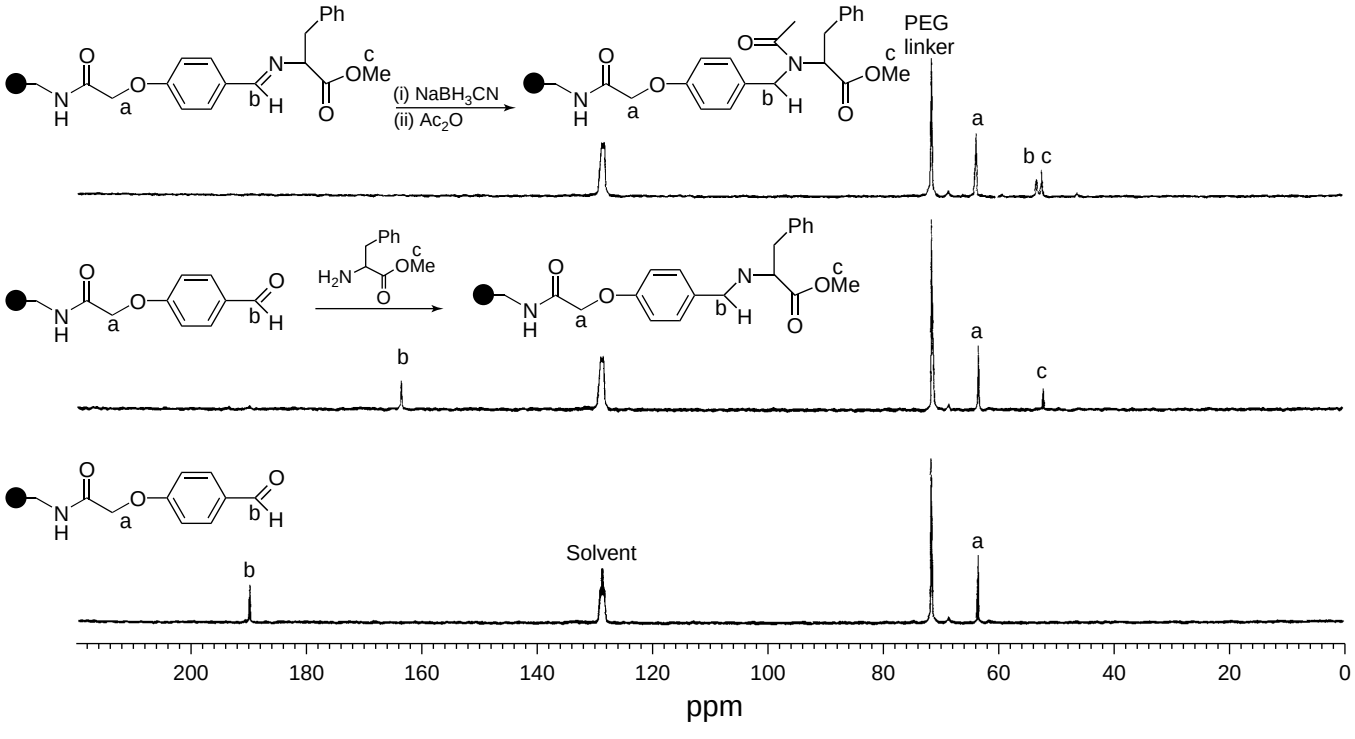

Figure 16 Fast ^{13}C-NMR monitoring of a stepwise reductive amination sequence. (Reproduced with permission from G.C. Look, C.P. Holmes, J.P. Chin, M.A. Gallop, *J. Org. Chem.*, **59**, 7588–7590 (1994). Copyright 1994 American Chemical Society.)

resonances of carbon atoms close to the polymer backbone were observed to be broadened relative to those associated with side chains of the resin-bound intermediates. Blossey et al.[119] reported the gel-phase ^{13}C-NMR characterization of resin-bound steroids obtained by SPOS on dehydrocholic and cholic acids immobilized on chloromethylated poly(styrene–2% divinylbenzene). Correlation with solution-phase ^{13}C-NMR spectra permitted complete assignment of each of the resonances observed in the gel-phase ^{13}C-NMR spectra of the resin-bound steroids.

Gel-phase ^{13}C-NMR involves averaging thousands of transients to improve signal-to-noise, making the technique impractical for monitoring reactions in real time. The long analysis times required in ^{13}C-NMR are primarily due to the low natural abundance of the ^{13}C isotope. By using ^{13}C enriched reagents, Look et al.[120,121] have reported obtaining high-quality NMR spectra from as few as 64 transients. Fast ^{13}C-NMR analysis requires as little as 20 mg of resin, and can be performed within 15–30 min with most of this spent drying the resin sample. Representative ^{13}C-NMR spectra from a stepwise reductive amination sequence are shown in Figure 16. The limited availability of ^{13}C-enriched intermediates may make this characterization technique impractical for some SPOS applications.

Manatt et al.[122] incorporated ^{19}F nuclei into amino-protecting groups and used these to monitor SPPS by gel-phase ^{19}F-NMR. These intermediates were immobilized on 1% cross-linked, chloromethylated polystyrene, and provided reasonably narrow NMR linewidths. Loadings as low as 0.005 mequiv g^{-1} could be detected by gel-phase ^{19}F-NMR.

Swayze[123] incorporated isotopically enriched [^{15}N]alanine in the synthesis of a combinatorial library of 3,4,8-trisubstituted 1,4-diazabicyclo[3,4,0]nonan-2-one compounds. The labeled amino acid nitrogen was involved in the formation of the key heterocyclic ring via an intramolecular Mitsunobu reaction, the formation of which was monitored by gel-phase ^{15}N-NMR.

Gel-phase ^{31}P-NMR has two advantages over ^{13}C-NMR for characterizing phosphorus-containing compounds immobilized on solid supports. First, the 100% natural abundance of the ^{31}P isotope makes the technique 383 times more sensitive than ^{13}C-NMR. Spectra can be obtained in as little as 10 min using the same sample preparation techniques described for gel-phase ^{13}C-NMR. Second, most polymer resins used in SPS do not contain phosphorus, and hence do not interfere with the observed resonances in ^{31}P-NMR. Bardella et al.[124] were the first to apply ^{31}P-NMR to the characterization of oligonucleotides prepared by SPS. The technique has also been applied to monitor the SPPS of phosphotyrosine-containing peptides.[125] Johnson and Zhang[126] used gel-phase ^{31}P-NMR to monitor the Horner–Wadsworth–Emmons synthesis of a combinatorial library of alkenes. Completion of this reaction is indicated by the disappearance of the multiplet at δ 22 due to the resin-bound diethylphosphonoacetamide starting material and the simultaneous appearance of a broad resonance for diethylphosphonate at δ 0. Quarrell et al.[127] studied the ^{31}P-NMR spectra of α-triphenylphosphonium-*p*-toluic acid bromide in solution and immobilized on Tentagel™ resin in order to characterize the mobility of compounds immobilized on this widely used solid-phase support.

5.1.2 Magic-angle Spinning Nuclear Magnetic Resonance

In solution, rapid isotropic motion averages the dipolar splittings and chemical shift anisotropy to zero, which yields sharp lines in the corresponding NMR spectrum. In solids, both the dipolar coupling and the chemical shift anisotropy contain an angular dependence of the form $(3\cos^2\theta - 1)$. This angular dependence becomes zero when θ equals roughly 54.73°, hence this angle is referred to as the ''magic angle''.

Characterization of cross-linked polystyrene gels by ^{1}H-MAS NMR was first reported in the mid-1980s.[128] A lack of suitable techniques for determining the structure of resin-bound compounds led to the investigation of ^{1}H MAS NMR as an analytical tool for SPOS beginning in the mid-1990s. The acquisition of high-resolution ^{1}H-MAS NMR spectra, however, would not have been possible were it not for the development of NMR probes that combined the use of both MAS and ''magnetic-susceptibility-matched'' materials. Indeed, the first reported 500 MHz ^{1}H-NMR spectrum of a resin-bound compound was recorded using the Nano-NMR™ probe commercialized by Varian Instruments (Palo Alto, California).[129] A comparable high-resolution MAS probe is available from Bruker (Rheinstetten, Germany). A comparison of a 16-scan 500 MHz ^{1}H-NMR spectrum of a resin-bound compound obtained using a conventional 5-mm liquids probe with that obtained by ^{1}H-MAS NMR using a Nano-NMR™ probe is shown in Figure 17. In addition to providing a significant improvement in resolution through removal of magnetic susceptibility line broadening, the MAS probe required only one tenth of the sample used in the conventional liquids probe.

Modern MAS probes designed for combinatorial analysis are sensitive enough to allow ^{1}H-MAS NMR spectra to be obtained on single resin beads. The ^{1}H-MAS NMR spectra from the stepwise SPS of a hydantoin precursor are shown in Figure 18.[130] Sample preparation involved swelling a single macro bead in DMSO-d_6 and placing it in a Bruker MAS probe. The spectra were generated within 6–8 min by averaging 128–168 scans. The observed

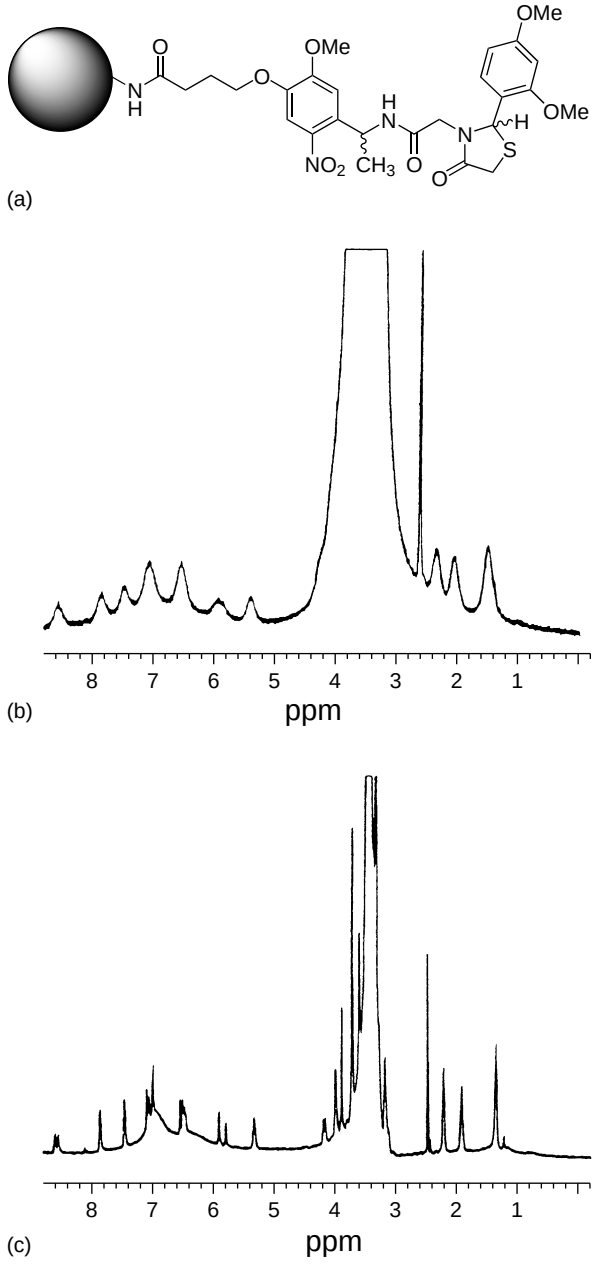

Figure 17 A 500 MHz ^{1}H-NMR spectrum of a resin-bound compound (a) obtained with a conventional 5-mm liquids probe (b) and a 500 MHz ^{1}H-MAS NMR spectrum of the same compound using a Varian Nano-NMR™ probe (c). (Reproduced with permission from W.L. Fitch, G. Detre, C.P. Holmes, J.N. Shoolery, P.A. Keifer, *J. Org. Chem.*, **59**, 7955–7956 (1994). Copyright 1994 American Chemical Society.)

line widths were in the order of 5–8 Hz. Sarkar et al.[131] have reported similar success in obtaining ^{1}H-MAS NMR spectra from single resin beads using the Varian Nano-NMR™ probe. Chin et al. applied MAS NMR to reaction monitoring and structure determination of compounds attached to multipin crowns.[132]

The quality of the spectra obtained by ^{1}H-MAS NMR can be enhanced by applying a small decoupling pulse to the signal prior to data collection (presaturation). In some cases, ^{1}H-MAS NMR spectra are complicated by the presence of broad peaks related to the polystyrene matrix of the resin. Garigipati et al.[133] have reported that this type of interference from the resin matrix can be eliminated by application of spin-echo pulse sequences in the ^{1}H-MAS NMR experiments.

Stöver and Fréchet[134] demonstrated the superiority of ^{13}C-MAS NMR over gel-phase ^{13}C-NMR in the characterization of solvent-swollen polymers. The narrow line widths observed in the MAS spectra were attributed to the cancellation of residual motional anisotropy within the polymer network. The first application of ^{13}C-MAS NMR to combinatorial chemistry involved the structural characterization of norborane-2-carboxylic acid epimers attached to a polystyrene resin.[135] The ^{13}C-MAS NMR spectrum of the resin yielded sharp signals for all of the carbons in the norbornyl system, with an average resolution of 13 Hz. Remarkably, the ratio of peak intensities in the ^{13}C-MAS NMR spectrum suggested a 60 : 40 *exo/endo* ratio of epimers on the resin, whereas the ratio of epimers in nonbound norborane-2-carboxylic acid is 70 : 30.

Macdonald et al.[136] applied ^{31}P MAS NMR to the characterization of oligonucleotides covalently bound to controlled porous glass beads. Changes in the oxidation state of the phosphorus atom and the addition or removal of protecting groups were readily apparent from the shifts in the resonances observed in the ^{31}P MAS NMR spectra. However, the spectra of many of the oligonucleotides consisted of isotropic ^{31}P resonances. Thus, ^{31}P MAS NMR cannot distinguish between di- and trinucleotides, or between linear and branched trinucleotides immobilized on controlled porous glass beads.

The high quality of MAS NMR spectra allows the same types of two-dimensional (2-D) NMR experiments that are performed to assess the structure of compounds in solution to be carried out on solid-phase supports. The first total structural assignment of a resin-bound compound by NMR was reported in 1995.[137] Heteronuclear multiple quantum coherence (HMQC) spectroscopy and total correlation spectroscopy (TOCSY) MAS NMR spectra for Wang-bound Fmoc-lysine-*t*-butoxycarbonyl (Boc) (see Figure 19) enabled complete assignment of the aliphatic resonances in the compound.

While the line widths in ^{1}H-MAS NMR spectra are sufficiently narrow to distinguish between protons based on chemical shift information, they are too broad to reveal couplings. Shapiro et al.[138] have reported that coupling information can be obtained from resin-bound compounds by performing 2-D *J*-resolved NMR experiments. Spin-echo correlated spectroscopy (SECSY) can

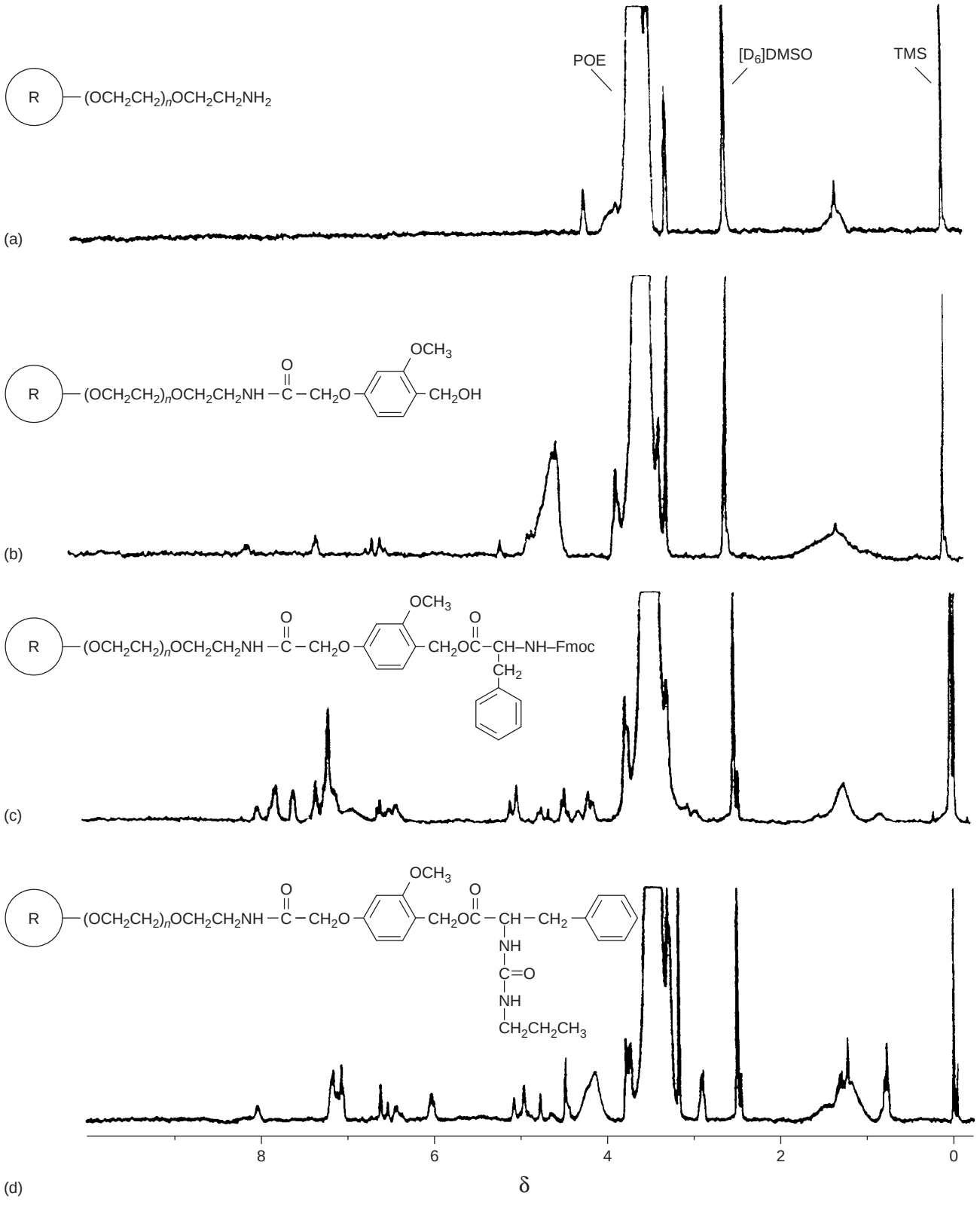

Figure 18 ^{1}H-MAS NMR spectra of a suspension of individual resin particles in the hydantoin reaction sequence: spectra of (a) the support-bound polyoxyethylene spacer, (b) with acid-labile linker, (c) after treatment with 9-fluorenylmethyloxycarbonyl (Fmoc) phenylalanine, and (d) after scission of the Fmoc group and treatment with isocyanate. (Reproduced by permission of Wiley-VCH from M. Pursch, G. Schlotterbeck, L.H. Tseng, K. Albert, W. Rapp, *Angew. Chem. Int. Ed. Engl.*, **35**, 2867–2869 (1996).)

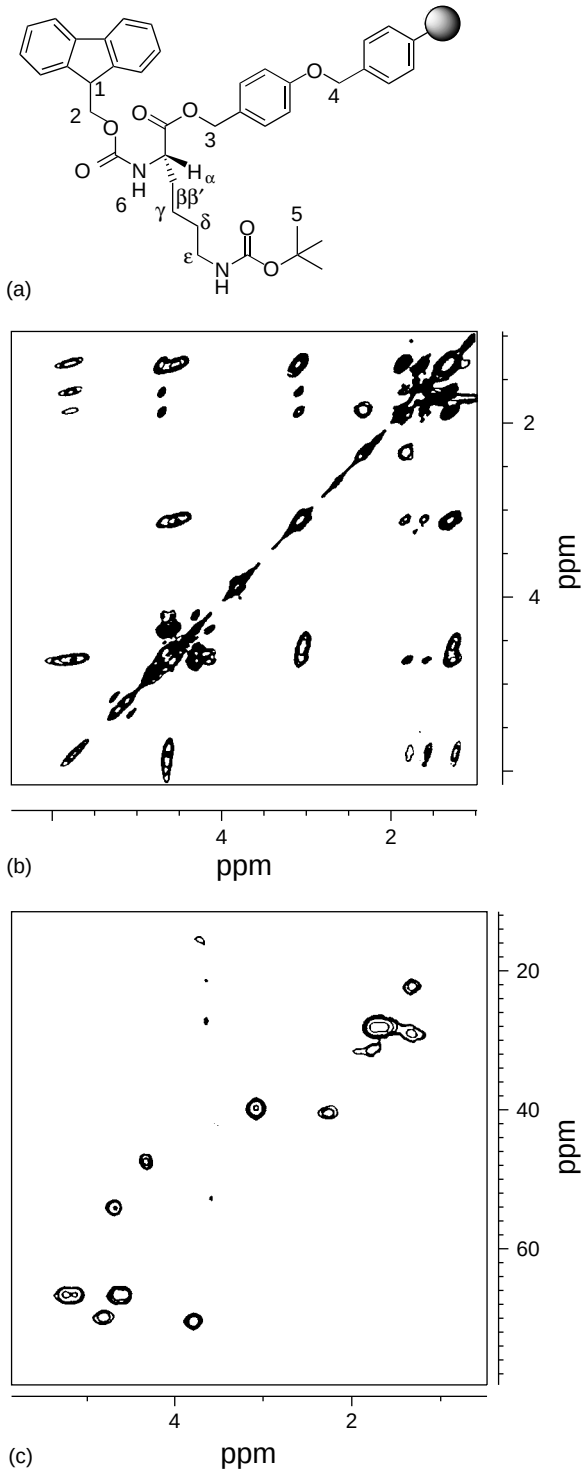

Figure 19 TOCSY (b) and HMQC (c) MAS NMR spectra of Wang-bound Fmoc-lysine-Boc (a). (Reprinted with permission from R.C. Anderson, J.P. Stokes, M.J. Shapiro, 'Structure Determination in Combinatorial Chemistry: Utilization of Magic Angle Spinning HMQC and TOCSY NMR Spectra in the Structure Determination of Wang-bound Lysine', *Tetrahedron Lett.*, **36**(30), 5311–5314. Copyright 1995, with permission from Elsevier Science.)

simultaneously provide spin connectivity and coupling information[139] for protons in resin-bound molecules. SECSY data contains the same connectivity information provided by homonuclear shift correlation spectroscopy (COSY), but retains the enhanced resolution of 2-D *J*-resolved experiments.

Nuclear Overhauser effect (NOE) spectroscopy is routinely applied to characterize the three-dimensional (3-D) structure of peptides and proteins in solution. Several research groups have used NOE analysis to determine the secondary structure of resin-bound peptides.[140,141]

5.2 Off-resin Analysis by Nuclear Magnetic Resonance Spectroscopy

Despite the strengths of NMR as a powerful spectroscopic method, it is difficult when using most NMR techniques to distinguish between a compound of interest and any impurities or contaminants that may be present in the sample. Structure elucidation of mixtures of unknown organic compounds with overlapping ^{1}H-NMR signals is difficult, especially in spectral regions that are crowded with many resonances. What is needed for the NMR analysis of combinatorial libraries is a means of distinguishing between the individual sample components during spectral acquisition. Flow-through interfaces have been developed to directly couple HPLC and NMR, and this allows individual components present in a sample to be spatially resolved prior to their characterization by NMR. A second approach distinguishes between individual components of a sample solution based on differences in their diffusion rates, and does not require a separation step to be performed prior to NMR analysis.

5.2.1 Flow-through Nuclear Magnetic Resonance Interfaces

The first flow-through interfaces for liquid chromatography/nuclear magnetic resonance (LC/NMR) were developed in the late 1970s.[142,143] However, the NMR spectrometers used in these systems employed iron magnets with low field strengths, so sensitivity was severely limited (e.g. 100-μg detection limit). Developments during the early 1980s[144–146] led to the first applications of on-line LC/NMR employing superconducting magnets. Flow-through interfaces for LC/NMR consist of a hollow glass tube, through which the LC eluant flows, and an RF coil wrapped around the outside of the detection window. The design of the RF coil and related circuitry, and the internal volume and geometry of the flow cell differ by manufacturer. Diagrams of a Bruker (Karlsruhe, Germany) continuous-flow NMR probe and typical LC/NMR system configuration are shown in Figure 20.[147] The Bruker flow cell is a

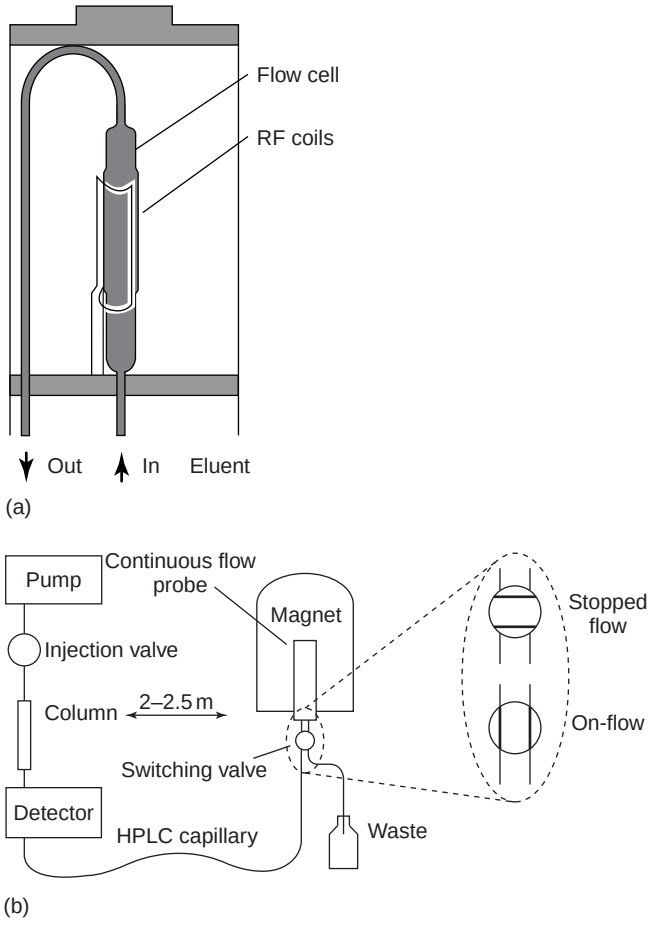

Figure 20 (a) Schematic diagram of the geometry of a continuous-flow NMR probe and (b) experimental arrangement for coupling an HPLC system with an NMR spectrometer. (Reprinted with permission from K. Albert, 'On-line Use of NMR Detection in Separation Chemistry', *J. Chromatogr. A*, **703**, 123–147. Copyright 1995, with permission from Elsevier Science.)

U-shaped glass tube with a slight bulge at the detection window. An RF coil is attached to the outside of the glass tube at the bulge. The LC/NMR probe is preceded by a switching valve (see Figure 20b) that allows spectra to be collected in either continuous-flow or stopped-flow mode. The latter allows NMR experiments involving the collection of hundreds to thousands of transients to be performed without interrupting the LC separation.

In contrast to conventional NMR liquids probes, which typically spin the sample at 20 Hz to expose the sample to a uniform magnetic field, the continuous-flow probes used in LC/NMR are static. As a consequence, the NMR line widths observed in LC/NMR are 3–4 times greater than those in conventional NMR spectra. Continuous-flow LC/NMR probes produce additional line broadening when used in on-flow rather than stopped-flow mode, because the observed nuclei have limited residence times within the detection region of the probe compared with their spin–lattice and spin–spin relaxation times. Dorn reported that a sample should be resident in the detection region of a LC/NMR flow cell for a minimum of 1–6 s to minimize line broadening.[144]

The volume of sample within the region of the RF coil of a continuous-flow LC/NMR probe is only slightly smaller than the total volume of the glass cylinder containing the detection area. This high filling factor helps offset the negative effects of failing to spin the LC/NMR cell within the magnetic field. The detection limits that can be achieved with ^{1}H-NMR continuous-flow probes are dependent on the number of equivalent nuclei within the detection coil, and the spin–lattice relaxation (T_1) times of the observed protons. Albert[147] reported a detection limit of 500 ng for the benzyl CH_2 protons of *n*-butylbenzylphthalate obtained with a flow cell with a volume of 120 µL on a 600-MHz instrument. Detection under stopped-flow conditions is approximately an order of magnitude more sensitive than in on-flow mode, because of the improvements in the signal-to-noise ratio that can be achieved with long acquisition times.

Miniaturization of the basic flow-through cell design has been described for microcolumn and nanoscale separation techniques (e.g. capillary zone electrophoresis, capillary electrochromatography, and capillary HPLC).[148–150]

5.2.2 *Liquid Chromatography/Nuclear Magnetic Resonance*

LC/NMR has been applied extensively in drug metabolism studies in the pharmaceutical industry.[151–155] The relative ease with which information-rich spectra can be obtained using this technique has enabled the rapid identification of important drug metabolites in a variety of complex matrices, including blood, urine, tissue, and bile. There have been comparatively fewer applications of LC/NMR to combinatorial chemistry, presumably because commercial systems have only recently been developed for combinatorial chemistry and software for data reduction of libraries of compounds is not widely available. Chin et al.[156] reported the analysis of a mixture of four isomeric (dimethoxybenzoyl)glycines by high-performance liquid chromatography/nuclear magnetic resonance (HPLC/NMR) in 1998. The isomers were separated by reversed-phase HPLC, and NMR spectra were generated using both on-flow and stopped-flow conditions on a 500-MHz instrument. A UV chromatogram of the sample mixture is shown in Figure 21(a), and the ^{1}H-NMR spectra of the four components present are shown in Figure 21(b).

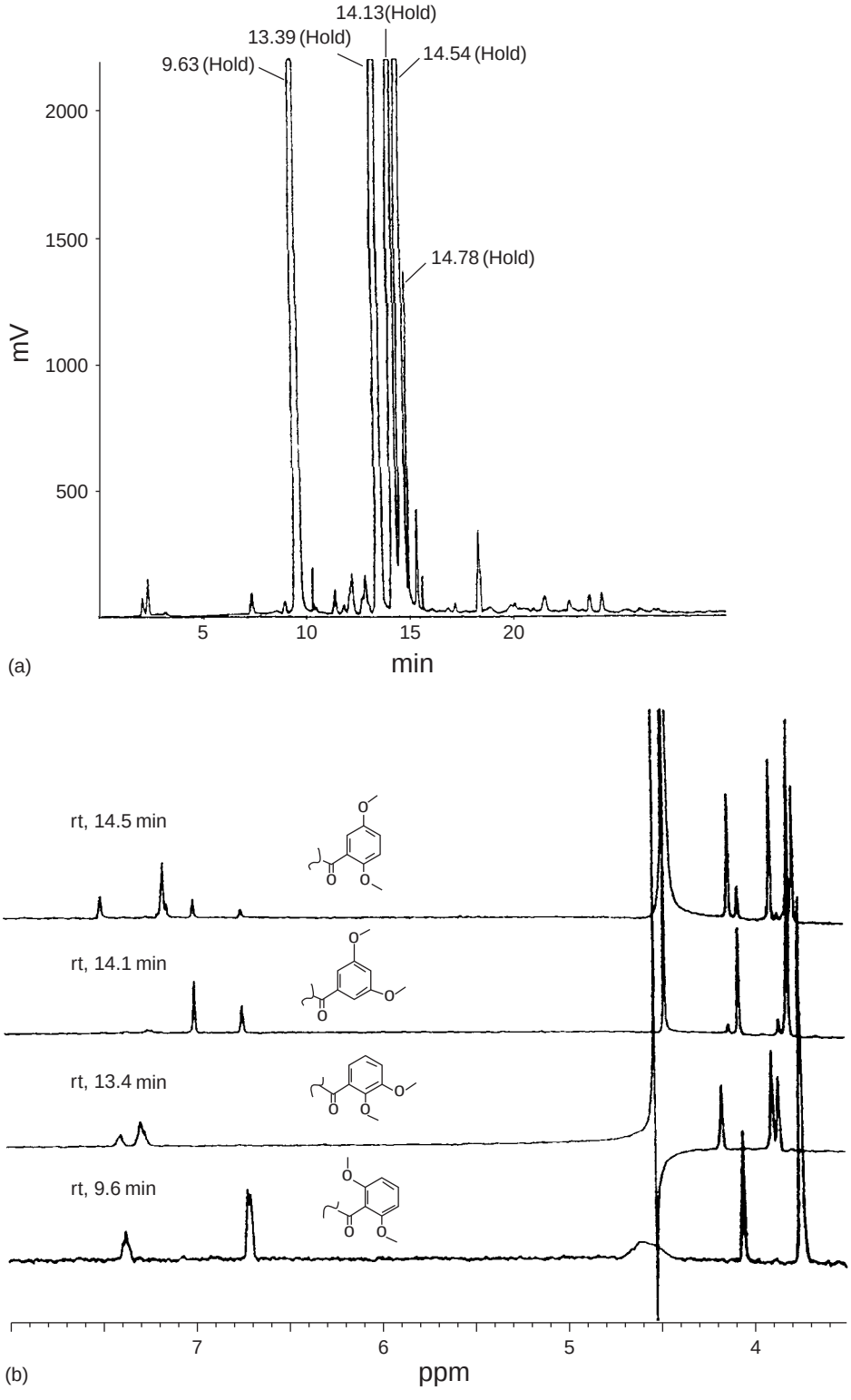

Figure 21 (a) UV trace of the chromatographic separation for a mixture of four isomeric (dimethoxybenzoyl)glycines and (b) their corresponding 500 MHz ^{1}H-NMR spectra. (Reproduced with permission from J. Chin, J.B. Fel, M. Jarosinski, M.J. Shapiro, J.R. Wareing, *J. Org. Chem.*, **63**, 386–390 (1998). Copyright 1998 American Chemical Society.)

Suppression of signals from the chromatographic eluent is a critical challenge to LC/NMR. There are a number of techniques available for eliminating interference from proton-containing solvents in LC/NMR systems. The most direct approach to eluent signal suppression involves using deuterated or halogenated solvents in the LC mobile phase. However, the high cost of deuterated solvents makes this approach impractical for conventional reversed-phase applications. Electronic pulse techniques provide the best approach to solvent signal suppression in LC/NMR. Presaturation to suppress interfering solvent signals requires lengthy saturation periods during which sample spectra cannot be acquired, which limits application of this technique to stopped-flow analysis.[157] Binomial suppression methods, such as the 1–1 hard pulse and 1331 pulse sequences[158,159] are compatible with on-flow analysis and can be used to suppress multiple solvent resonances simultaneously. Pulsed-field gradients also provide a useful means of suppressing background water signals in one-dimensional (1-D) and 2-D NMR experiments.[160–162] Developments in solvent suppression techniques for LC/NMR have been reviewed.[163]

5.2.3 Diffusion-resolved Nuclear Magnetic Resonance

Morris and Johnson[164,165] developed a 2-D NMR technique that allows ^{1}H-NMR spectra of discrete chemical species present in mixtures to be distinguished based on differences in their diffusion coefficients. 2-D NMR diffusion-ordered spectroscopy (DOSY) is performed in two steps. The first step of the DOSY experiment involves generating the 2-D NMR data using pulsed-field-gradient NMR techniques. The data are subsequently transformed and displayed with chemical shift information plotted in one dimension and diffusion coefficients in the second dimension. Transformation of NMR signal intensities versus gradient pulse areas into diffusion spectra is performed using standard computer algorithms. DOSY is potentially a powerful technique for characterizing mixtures of low-molecular-weight compounds of similar size, such as those present in combinatorial mixture libraries. A significant advantage of DOSY over LC/NMR is that spectra on individual compounds present in a mixture can be obtained directly, without requiring an initial chromatographic separation.

Extension of the standard DOSY experiment to include ^{13}C correlation information as a third dimension has been reported by Barjat et al.[166] The 3-D diffusion-ordered spectroscopy/heteronuclear multiple quantum coherence (DOSY/HMQC) spectrum of a mixture of quinine, camphene, and geraniol in methanol-d_4 is shown in Figure 22.[166] Lin et al.[167,168] combined pulsed field gradients with TOCSY in a technique they termed diffusion-encoded spectroscopy (DECODES). A limitation of DOSY is that overlapped resonances in the region between 0.5 and 2.0 ppm cannot always be resolved by diffusion alone. TOCSY, however, can resolve overlapped resonances by separating coupled spin systems in the second dimension. Hence, the DECODES technique is more effective in resolving individual components present in complex mixtures.

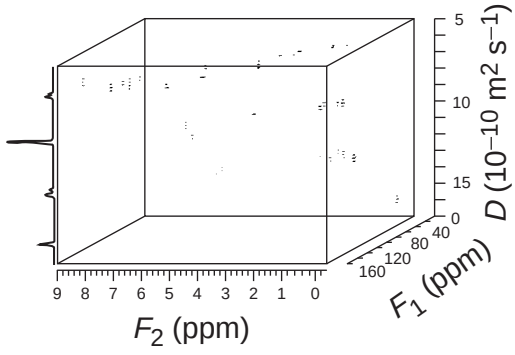

Figure 22 3-D DOSY/HMQC spectrum of a mixture of quinine, camphene, and geraniol in methanol-d_4. (Reproduced by permission of Academic Press Inc. from H. Barjat, G.A. Morris, A.G. Swanson, *J. Magn. Reson.*, **131**, 131–138 (1998).)

6 LIBRARY SCREENING BY AFFINITY SELECTION AND MASS SPECTROMETRY

Classical approaches to drug discovery involve synthesizing, purifying, characterizing and biologically screening compounds one at a time. While there have been significant advances in methods for synthesizing large numbers of compounds in combinatorial libraries, there are relatively few techniques available for screening mixtures of compounds in ways that allow detection of specific receptor–ligand interactions and preserve the native binding of these interactions. Most screening approaches use classical bioassays to test the activity of single compounds, but perform these assays in parallel to increase sample throughput (i.e. high throughput screening). Affinity-based assays, while compatible with mixtures of ligands, employ immobilized receptors or immobilized ligands and yield binding constants that may not mimic true solution-phase binding.

During the mid-1990s, a number of affinity-based screening techniques were developed to overcome the disadvantages of conventional assays. These improved strategies combine the high specificity of mass spectrometric detection with novel approaches for distinguishing receptor-bound ligands from unbound species present in combinatorial mixture libraries.

6.1 Ligand Selection by Affinity Chromatography

Affinity chromatography is a technique wherein a mixture of ligands is chromatographically separated based on their interaction with an immobilized receptor. The target receptor usually is immobilized on a solid support, and packed into a column. A solution containing the mixture library is passed through the column, and those compounds that interact strongly with the receptor are fully retained. Compounds that interact weakly with the immobilized receptor, or not at all, pass through the column unretained. The strongly retained compounds then are eluted from the affinity column using a high-ionic-strength buffer. The active mixture is passed through a desalting column to remove buffer salts, and individual components in the mixture are separated by reversed-phase HPLC prior to their characterization by ESIMS. A disadvantage of affinity chromatography as a compound selection technique is that the immobilization of receptors to the solid support may alter their binding with ligand molecules.

A variety of approaches to library screening by affinity selection and MS have been described, and they all utilize affinity separation for compound selection and ESIMS for compound identification. Nedved et al.[169] developed an automated system for screening benzodiazepine combinatorial libraries. This system consists of an immunoaffinity column containing antibodies to benzodiazepines and two reversed-phase HPLC columns coupled to a triple-quadrupole mass spectrometer through a series of electronic switching valves (see Figure 23).[169] A mixture of benzodiazepines is passed through the immunoaffinity column, where active compounds bind to the immobilized antibodies. A pH step-gradient is used to release the benzodiazepine–antibody complexes from the immunoaffinity column onto a restricted access media stationary phase that separates the active benzodiazepines from the antibody. A C-8 reversed-phase column is then used to separate the mixture of active benzodiazepines before they enter the mass spectrometer. Kelly et al.[170] adopted a more conventional approach to identify peptides that bind strongly to the Src homology 2 (SH2) domain of phosphatidylinositol 3-kinase. Active peptides eluted from a SH2 affinity column were desalted on a small molecule trap column prior to their characterization by ESIMS.

Karger et al.[171–173] developed an approach to library screening based on affinity capillary electrophoresis and mass spectrometry (ACE/MS). In affinity capillary electrophoresis (ACE), the electrophoretic mobility of a ligand in solution is altered upon complexation with a receptor molecule. Complexation with the receptor shifts the mobility of the ligand to that of the complex, and this shift in mobility corresponds to the ligand's binding affinity to the receptor. In the ACE/MS experiment, the capillary is filled with a running buffer containing the receptor. A mixture of ligands is injected into the inlet side of the capillary prior to the start of electrophoresis. Ligand molecules (both bound and free) are introduced into the electrospray source of the mass spectrometer as they migrate past the capillary exit. An advantage of ACE/MS is that the compound selection and compound identification steps of the library screening process are combined into a single step. The ACE/MS technique has been applied successfully to the screening of peptide libraries, but the technique may not be applicable to libraries of small-molecule organic compounds due to the complexities involved in developing electrophoretic separation conditions appropriate for these types of compounds.

6.2 Ligand Selection by Gel-filtration

Gel-filtration columns are packed with hydrophilic polymer beads of controlled porosity. These columns separate compounds primarily on the basis of molecular size and shape. Large molecules are excluded from the interior porous region of the beads and elute at or near the void volume of the column. Small molecules travel deep within the interior of the beads and take longer to elute from the column. The use of gel-filtration columns in library screening has focused primarily on the separation of receptor-bound ligands from combinatorial mixture libraries. Typically, a solution containing compounds from the mixture library is incubated with a solution containing the target receptor. The mixture then is passed through a gel-filtration column, where the bound ligands elute with the receptor molecules, while the unbound ligands are retained for longer. The receptor-bound ligands are liberated from the target and characterized by LC/ESIMS. A variant of this procedure has been incorporated into an automated screening methodology termed Selectronics™ and is illustrated in Figure 24.[174] In addition, this general screening procedure has been used in competitive binding experiments to identify library components with optimum binding affinity.[175]

Gel-filtration separations are best suited to ligands that have high affinity for the receptor molecule ($K_d < 1\,\mu M$).[176] Loss of ligand during the gel-filtration step is possible if the rate of dissociation of the receptor–ligand complex is fast compared with the separation time. In addition, irreversible retention of the receptor on the gel-filtration support can lead to poor ligand recoveries.

6.3 Ligand Selection by Ultrafiltration

Ultrafiltration membranes are a convenient and more robust alternative to gel-filtration columns for separating

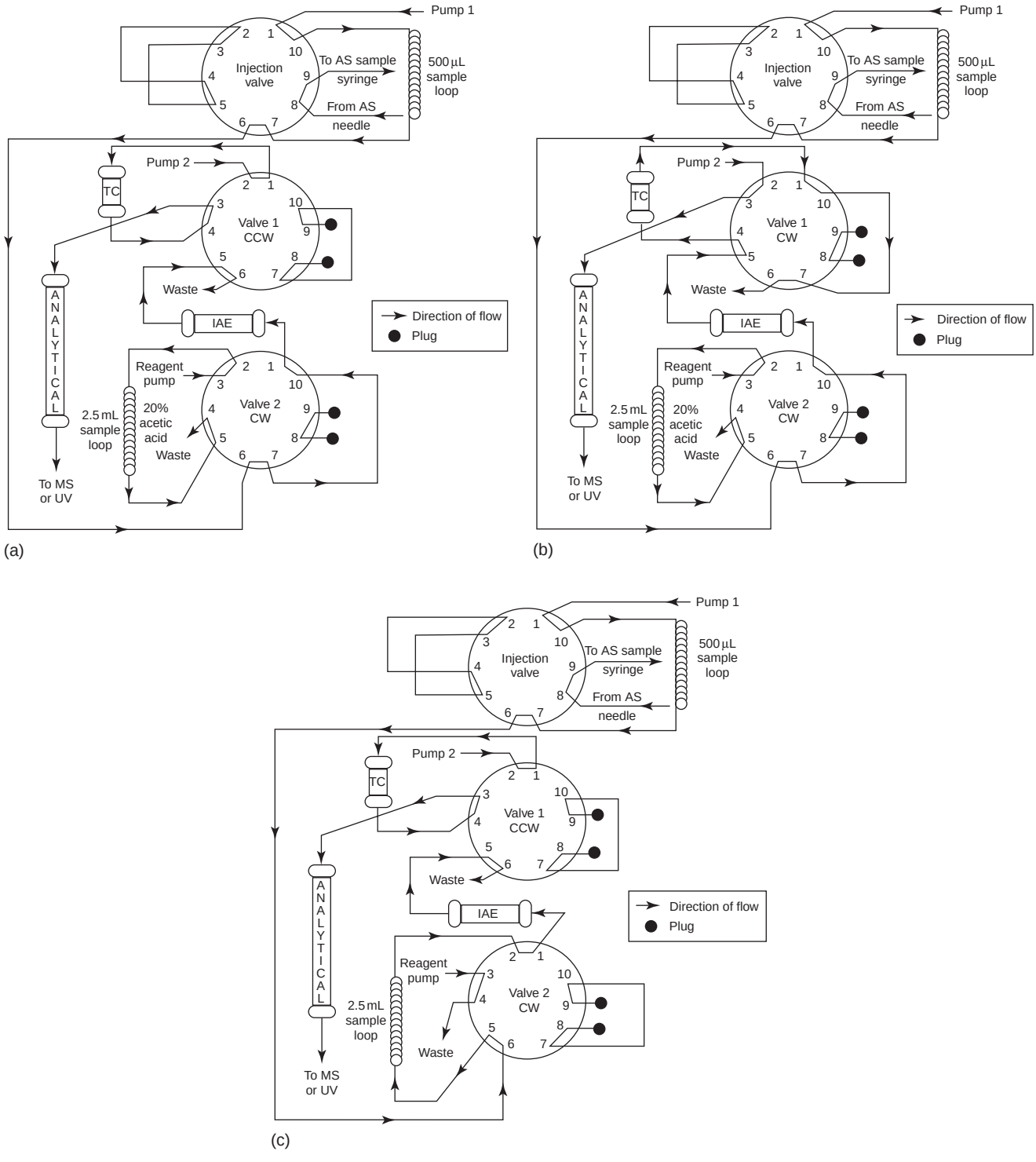

Figure 23 Characterization of benzodiazepine combinatorial libraries by on-line immunoaffinity extraction coupled with HPLC–ion spray MS. Valve positions for (a) loading antibody and chemical library onto the immunoaffinity extraction (IAE) column, (b) desorbing benzodiazepine–antibody complexes onto the trap column (TC), and (c) backflushing the trap column on to the analytical reversed-phase column. (Reproduced with permission from M.L. Nedved, S. Habibi-Goudarzi, B. Ganem, J.D. Henion, *Anal. Chem.*, **68**(23), 4228–4236 (1996). Copyright 1996 American Chemical Society.)

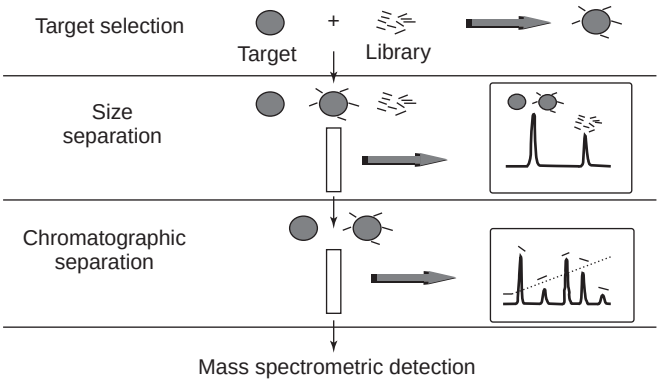

Figure 24 The Selectronics™ screening strategy of diverse libraries with solution-phase selection. (Reproduced with kind permission of Kluwer Academic Publishers from Y.F. Hsieh, N. Gordon, F. Regnier, N. Afeyan, S.A. Martin, G.J. Vella, *Mol. Diversity*, **2**, 189–196 (1996).)

molecules on the basis of size. Ultrafiltration membranes are manufactured of microporous cellulose triacetate or polysulfone in a variety of molecular weight cutoffs. The nominal molecular weight cutoff of an ultrafiltration membrane refers to the molecular weight of a globular protein that is 90% rejected by the membrane. Compounds with molecular weights significantly below the molecular weight cutoff pass freely through the membrane. Due to their relatively smaller surface areas, sample losses related to irreversible adsorption are rarely observed with these devices.

Van Breemen et al.[177,178] developed the technique of pulsed ultrafiltration MS by combining on-line ultrafiltration with electrospray MS. Pulsed ultrafiltration MS allows the rapid screening of combinatorial libraries for ligands that bind to receptors in solution. An outline of this technique is shown in Figure 25. The technique employs a flow-through chamber consisting of two compartments separated by an ultrafiltration membrane. During pulsed ultrafiltration, receptor-bound ligands remain in one half of the ultrafiltration chamber, while unbound compounds pass through the membrane and are washed away. Release of bound ligands is achieved by adding a small aliquot of an organic solvent to the sample compartment. The released compounds are then continuously monitored by ESIMS. The pulsed ultrafiltration chamber design allows the receptor to be reused for subsequent binding experiments, and is compatible with competitive binding studies and detection of strong-affinity ligands present at dilute concentrations in the presence of higher concentrations of weaker binders. For example, the adenosine deaminase inhibitor, erythro-9-(2-hydroxy-3-nonyl)adenine, was readily detected from a mixture of 20 adenosine analogs despite being present at a 10-fold lower concentration than the other ligands.[177]

Wieboldt et al.[179] employed ultrafiltration membranes to screen mixtures of benzodiazepines for binding affinity to antibodies raised to nitrazepam, temazepam, and oxazepam. Benzodiazepine mixtures were incubated with a specific antibody in 10 mM ammonium acetate buffer at pH 5.0, and the unbound drug ligands were separated by centrifugal filtration through a 50 000 molecular weight cutoff ultrafiltration membrane. The bound ligands were

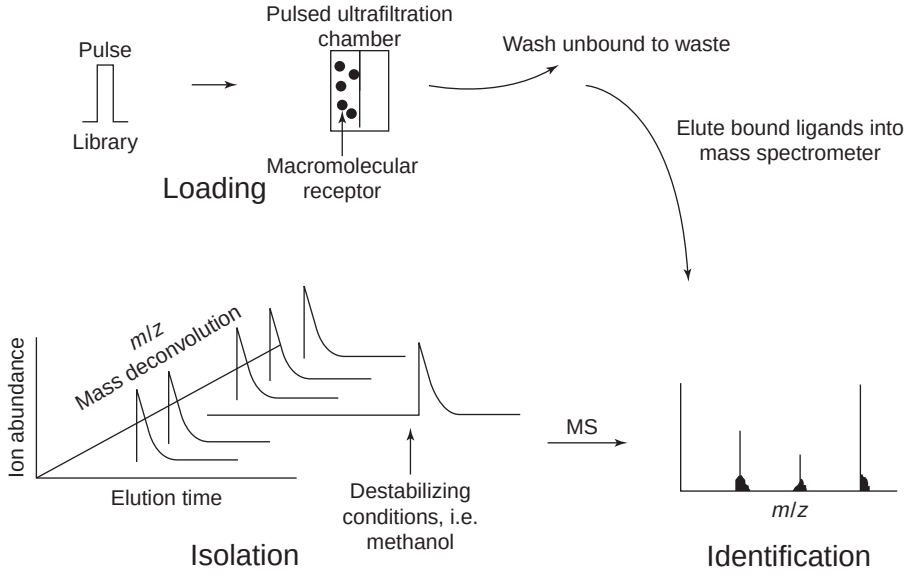

Figure 25 Scheme showing the use of pulsed ultrafiltration MS for screening a combinatorial library for compounds that bind to a macromolecular receptor. (Reproduced with permission from R.B. van Breemen, C.R. Huang, D. Nikolic, C.P. Woodbury, Y.Z. Zhao, D.L. Venton, *Anal. Chem.*, **69**, 2159–2164 (1997). Copyright 1997 American Chemical Society.)

characterized by HPLC and ESIMS, following their release from the antibody–drug complex by treatment with 1% TFA in water.

6.4 Bio-affinity Characterization Mass Spectrometry

All of the approaches to library screening described so far treat affinity selection and identification of active compounds as separate steps. As a result, these techniques have limited sensitivity due to the loss of ligand introduced by the additional sample handling required. An alternative approach developed by Smith et al.[180,181] eliminates the need for distinct separation and purification steps prior to ESIMS characterization. Bio-affinity characterization mass spectrometry (BACMS) combines ESI with Fourier transform ion cyclotron resonance (FTICR) MS in a technique that enables the direct characterization of noncovalent receptor–ligand complexes formed in solution. The basic BACMS experiment is depicted in Figure 26.[181] A solution containing the affinity receptor and ligand library is infused into the electrospray source, where the individual species and receptor–ligand complexes are transferred into the gas phase. These gas phase ions are trapped in the FTICR chamber, and the receptor–ligand complexes of interest are isolated by selected-ion accumulation. The selected-ion accumulation process is sufficiently gentle to prevent dissociation of the noncovalent complexes and allows these complexes to be observed with a good signal-to-noise ratio. This accumulation step is immediately followed by low-energy dissociation, which isolates the selected ligand molecules in the FTICR chamber. Collisionally activated dissociation then may be performed to obtain structural information on the high binding affinity ligands.

BACMS has been used to screen 256–289 member tripeptide libraries for tight binding inhibitors to carbonic anhydrase II obtained from bovine erythrocytes.[182] In this work, the relative intensities of the ions dissociated from the carbonic anhydrase II–inhibitor complexes correlate well with the binding constants of these inhibitors in solution (see Figure 27). This correlation between solution-phase and vapor-phase ordering of binding constants suggests that the data from BACMS experiments may be used to predict the order of binding affinities in solution.

Higger et al.[183] monitored the reaction between 1-chloro-2,4-dinitrobenzene (CDNB) and components of a H–γ–Glu–Cys–Xxx–OH (where Xxx is any natural amino acid except Cys, Val, Thr, and Pro) tripeptide library catalyzed by the enzyme glutathione-S-transferase (GST). Assessment of GST substrate specificity relied on detection of the products formed by reaction of the tripeptide library with CDNB, rather than on identification of tight binding ligands to the enzyme. In addition, the presence of two substrate-binding pockets on the

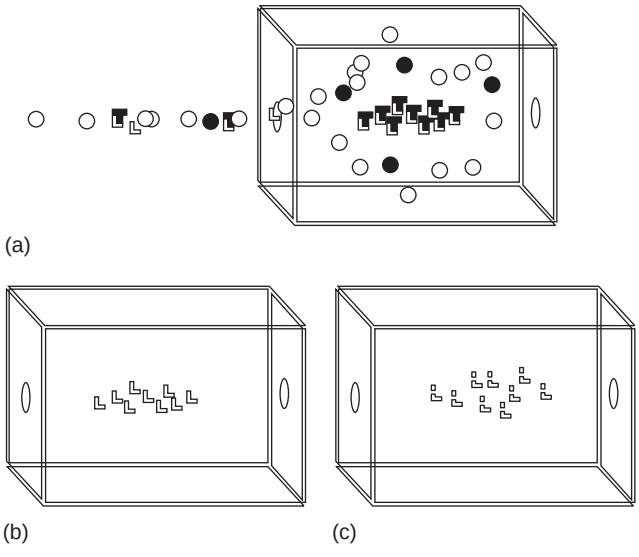

Figure 26 BACMS: (a) a mixture solution is ionized by ESIMS and the complex of interest is accumulated in the FTICR trap; (b) the noncovalent complex is perturbed to liberate the ligand species; then (c) the ligand ions are subjected to collision-activated dissociation for structural information. (Reproduced by permission of John Wiley & Sons, Ltd. from J.E. Bruce, G.A. Anderson, R. Chen, X. Cheng, D.C. Gale, S.A. Hofstadler, B.L. Schwartz, R.D. Smith, *Rapid Commun. Mass Spectrom.*, **9**, 644–650 (1995).)

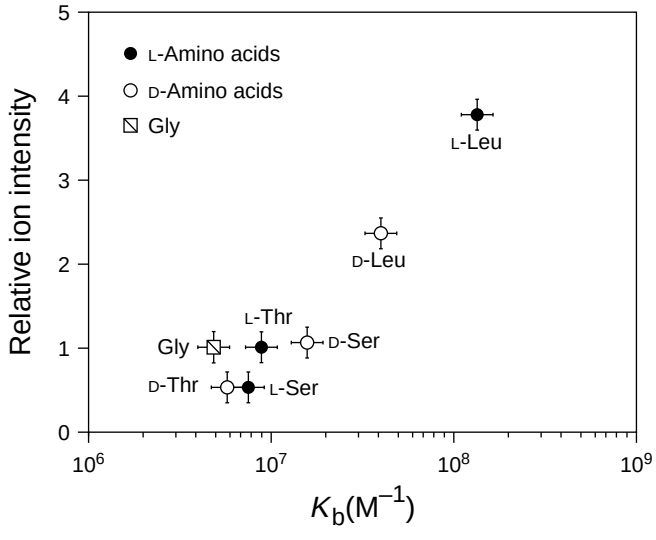

Figure 27 Correlation of relative ion intensities versus K_b in solution for seven peptide inhibitors to carbonic anhydrase II. (Reproduced with permission from J. Gao, X. Cheng, R. Chen, G.B. Sigal, J.E. Bruce, B.L. Schwartz, S.A. Hofstadler, G.A. Anderson, R.D. Smith, G.M. Whitesides, *J. Med. Chem.*, **39**, 1949–1955 (1996). Copyright 1996 American Chemical Society.)

GST enzyme allows optimization of substrate specificity to be extended to two dimensions. Two substrate libraries each targeting one of the two binding sites on the GST enzyme would result in the formation of products representing the best binding substrates from each library.

7 COMBINATORIAL LIBRARY QUANTITATION

Classical methods for handling compounds in a traditional drug discovery environment require large quantities (0.1–1.0 g) of highly pure material. Product yields from organic synthesis usually are calculated based on the mass of purified product obtained. Accurate measurement of pharmacological activity requires that known concentrations of test compounds be prepared by dissolving an accurately weighed amount of compound in an accurately measured volume of solvent. Product yields from combinatorial library synthesis often are too low (1–10 mg) to permit accurate weighing. In addition, split-synthesis techniques produce mixtures that are difficult to quantitate on a per-compound basis. Consequently, there has been considerable interest in developing reliable methods for determining the reaction yields and solution concentrations of combinatorial samples that do not require individual compounds to be accurately weighed.

The combination of HPLC and UV detection provides a rapid, reproducible, and sensitive means of quantifying compounds in solution. Accurate quantitation by UV, however, requires the use of well-characterized reference standards due to the wide variability in the molar absorptivity of organic compounds, and this has presented a serious impediment to the quantitation of combinatorial samples. Accurate quantitation of combinatorial samples requires a detector whose response depends only on the amount (concentration) of compound present in solution and provides a signal that is independent of molecular structure. Three detectors have emerged that are compatible with the requirements of HPLC and provide the degree of ''universal'' response required for accurate quantitation of combinatorial samples.

7.1 Quantitation using ^{1}H-Nuclear Magnetic Resonance Spectroscopy

Sample quantitation by ^{1}H-NMR is superior to quantitation by UV detection, because the integral intensity of a given resonance band in a ^{1}H-NMR spectrum is directly proportional to the number of proton nuclei contributing to that signal. Optimizing the instrument parameters to minimize quantitation errors, however, requires an understanding of the processes involved in acquiring a spectrum. During an NMR experiment, the bulk magnetization of protons in the analyte molecule is perturbed through application of an RF pulse. After a short delay, the exponential free induction decay (FID) signal produced as the system returns to equilibrium is digitized, averaged, and transformed to yield a frequency domain spectrum. The exponential decay of the FID signal is a function of the spin–spin (T_2) relaxation time, while return of the perturbed spin system to equilibrium is governed by the spin–lattice (T_1) relaxation time. In other words, the acquisition time in an NMR experiment depends on T_2, while the pulse repetition rate depends on T_1. The value of T_1 usually is greater than T_2 for organic molecules with a molecular weight less than 1000 Dal.

A significant problem in obtaining accurate quantitation in ^{1}H-NMR is related to variable spin–lattice relaxation times among protons in the sample molecule. Quantitation errors arise when the RF pulse rate is faster than T_1 for any of the sample nuclei. The situation becomes more complex when sample quantitation is performed using flow-through NMR systems. The effective T_1 and T_2 values of a proton in a flow-through system are related to their static values and the residence time of the molecule in the NMR detector coil. As a consequence, quantitative NMR experiments should be designed such that the RF pulse rate is five times the value of T_1 for the proton with the longest spin–lattice relaxation time, and the sample is resident within the region of the NMR detector coil for 1–6 s during measurement.[184] A number of additional experimental factors that can lead to spectral distortions in NMR and interfere with sample quantitation have been described.[185,186]

Quantitation by ^{1}H-NMR spectroscopy involves measuring the integral intensity of a specific sample resonance band relative to that of a resonance band corresponding to an internal standard. The major requirements placed on the choice of internal standard are: the proton resonances of the internal standard should not overlap the resonances from the sample; the internal standard must be chemically compatible with the sample; and the nuclei in the internal standard should have short T_1 relaxation times. Hexamethyldisiloxane is a useful internal standard for ^{1}H-NMR, because of its small chemical shift (0.07 ppm) relative to tetramethylsilane. In a conventional quantitation experiment, the sample and internal standard are combined in a standard NMR sample tube, and a ^{1}H-NMR spectrum recorded in the usual manner. The concentration of sample in solution is calculated as shown in Equation (1):

$$M_a = \frac{I_a}{n_a}\frac{n_s M_s}{I_s} \quad (1)$$

where n represents the number of protons assigned to the resonance signal, M is the molar concentration, and I is the integrated intensity of the internal standard (s) and analyte (a). This method will lead to erroneous quantification results if the sample contains impurities with ^{1}H-NMR resonances that overlap the signals corresponding to the compound of interest. Accurate quantitation of mixtures is achieved by performing a chromatographic separation prior to ^{1}H-NMR analysis.

The requirements for sample quantitation in continuous-flow HPLC/NMR experiments have been elaborated by Godejohann et al.[187] There are two methods for incorporating an internal standard in an HPLC/NMR experiment: direct addition of the internal standard to the HPLC mobile phase; and injection of a fixed amount of internal standard following sample injection. The former technique is employed in gradient HPLC to compensate for the influence of varying eluent composition on the intensity of resonance bands in the NMR spectrum. Injection of a fixed volume of the internal standard following sample injection is compatible with isocratic HPLC separations.

The technique of sample quantitation by ^{1}H-NMR with internal standard has been applied successfully to the determination of the potency of active ingredients in pharmaceutical dosage forms[188] and the yield of products originating from the SPS of combinatorial libraries.[189] A representative ^{1}H-NMR spectrum of a product from a combinatorial library in the presence of internal standard is shown in Figure 28.

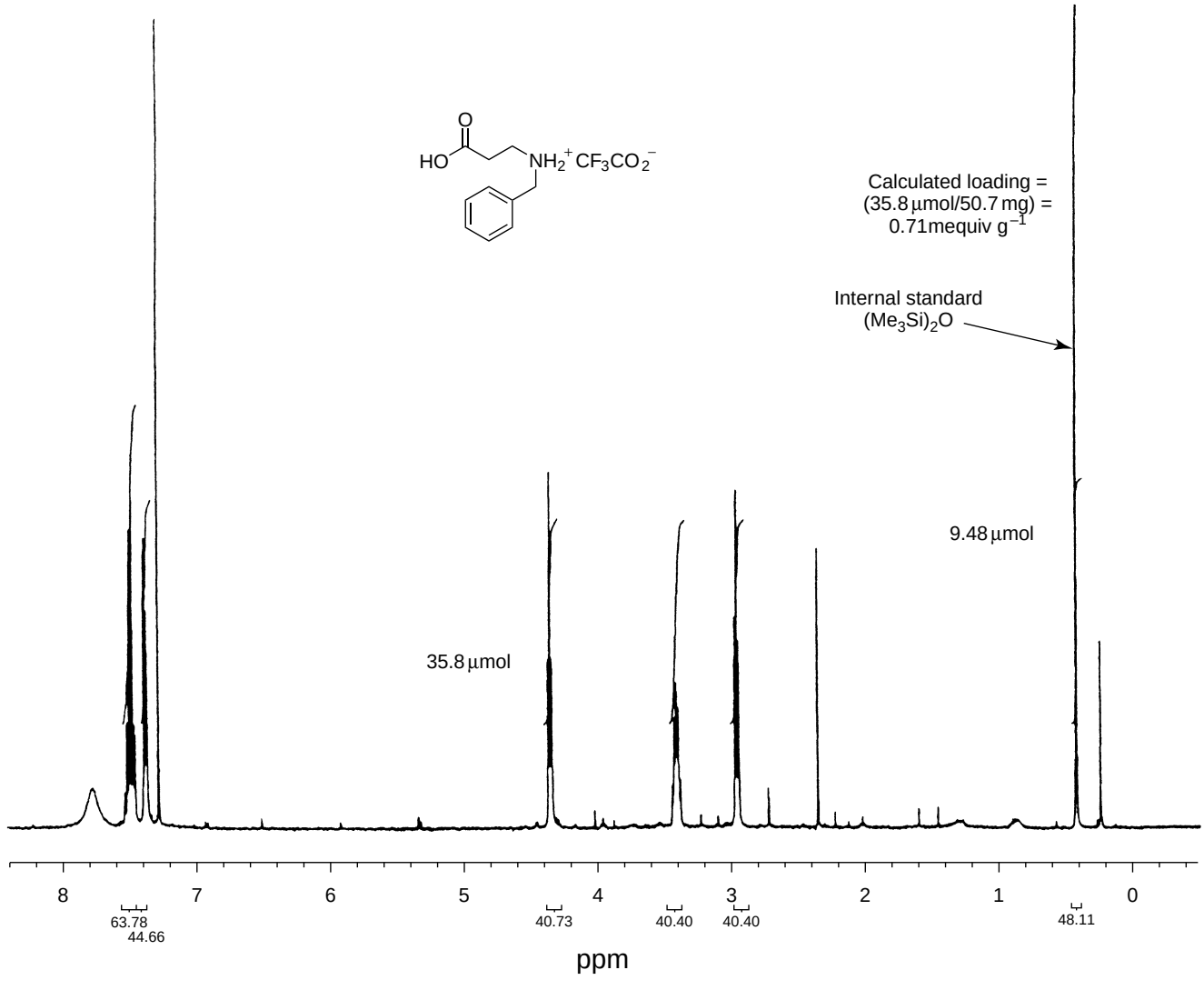

Figure 28 Quantitation of the direct cleavage product from a SPOS combinatorial library by ^{1}H-NMR with internal standard. (Reproduced with permission from B.C. Hamper, S.A. Kolodziej, A.M. Scates, R.G. Smith, E. Cortez, *J. Org. Chem.*, **63**, 708–718 (1998). Copyright 1998 American Chemical Society.)

7.2 Quantitation using Chemiluminescent Nitrogen Detection

Chemiluminescent nitrogen detection (CLND) instruments respond only to nitrogen present in organic compounds. Sample detection by CLND involves oxidizing the sample at approximately 1000 °C to convert nitrogen in the compound to nitric oxide. The nitric oxide is reacted with ozone to produce nitrogen dioxide in an excited state, which releases a photon as it returns to the ground electronic state. The emitted photons are detected using a photomultiplier tube, whose output is amplified to yield an analog signal. In contrast to UV detectors, the response of a chemiluminescent nitrogen detector is linearly related to the total mass of compound present in the sample and is independent of concentration.

The first practical application of CLND coupled with HPLC was reported by Fujinari et al.[190] for the analysis of ammonium nitrogen in wastewater. The combination of reversed-phase HPLC and CLND has been applied successfully to the purity assessment of synthetic peptides obtained by SPPS.[191] This combination of a high resolution separation technique and ''universal'' detection also has been utilized to assess product yields and purities of nonpeptide combinatorial libraries.[192,193] A series of chromatograms comparing the response of an equimolar mixture of nitrogen-containing organic compounds separated by gradient reversed-phase HPLC with UV detection at 214 nm and 270 nm and CLND is shown in Figure 29.

CLND has been reported to yield similar response to nitrogen-containing, small-molecule organic compounds over a linear range of 25 to 6400 pmol of nitrogen.[193] Furthermore, quantitation by CLND results in an error of approximately 10% over a wide range of compound classes. The use of CLND as a detector in reversed-phase HPLC imposes practical limitations on the choice of mobile phase eluents and buffers. Mobile phase components must be nitrogen-free, hence organic alcohols (e.g. methanol) are frequently used in place of acetonitrile as mobile phase modifiers.

7.3 Quantitation using Evaporative Light Scattering Detection

Like CLND, evaporative light scattering (ELS) detectors respond to the total mass of compound injected, rather than yielding a concentration-dependent response. The principle of operation of an ELS detector is based on measurement of the intensity of light scattered by the analyte following evaporation of the carrier solvent (e.g. the HPLC mobile phase). The HPLC column effluent enters the detector through a narrow bore tube where it is mixed with a high-velocity stream of nitrogen gas and passed through a nebulizer. The nebulized sample stream enters a heated tube, maintained at a sufficient temperature to cause rapid evaporation of the carrier solvent. The stream of solid analyte particles that is produced passes a detection system consisting of a source and photodiode detector placed 90° apart and perpendicular to the central axis of the drift tube. As the analyte passes through the detection region, the solid particles scatter light through a combination of Mie-scattering, reflection and refraction.[194,195] The intensity of the scattered light is measured by the photodiode and amplified to produce an analog signal proportional to the

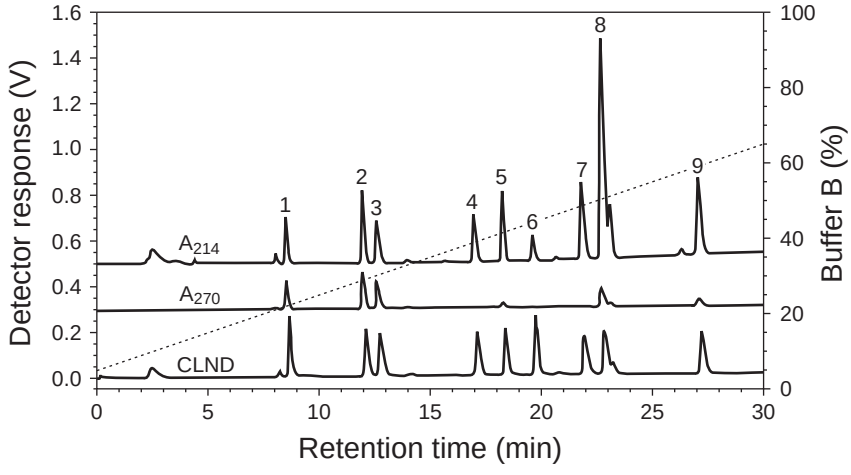

Figure 29 Chromatograms of a mixture of 'drug-like' compounds separated by reversed-phase HPLC with gradient elution and detected by UV at 214 nm and 270 nm and by CLND. The injected sample contained 2 nmol each of (1) caffeine, (2) triprolidine, (3) chlorpheniramine, (4) diphenylalanine, (5) 6-nor-6-allylsergic acid diethylamide, (6) angiotensin II, (7) diphenhydramine, (8) doxepin, and (9) dibucaine. (Reproduced with permission from E.W. Taylor, M.G. Qian, G.D. Dollinger, *Anal. Chem.*, **70**, 3339–3347 (1998). Copyright 1998 American Chemical Society.)

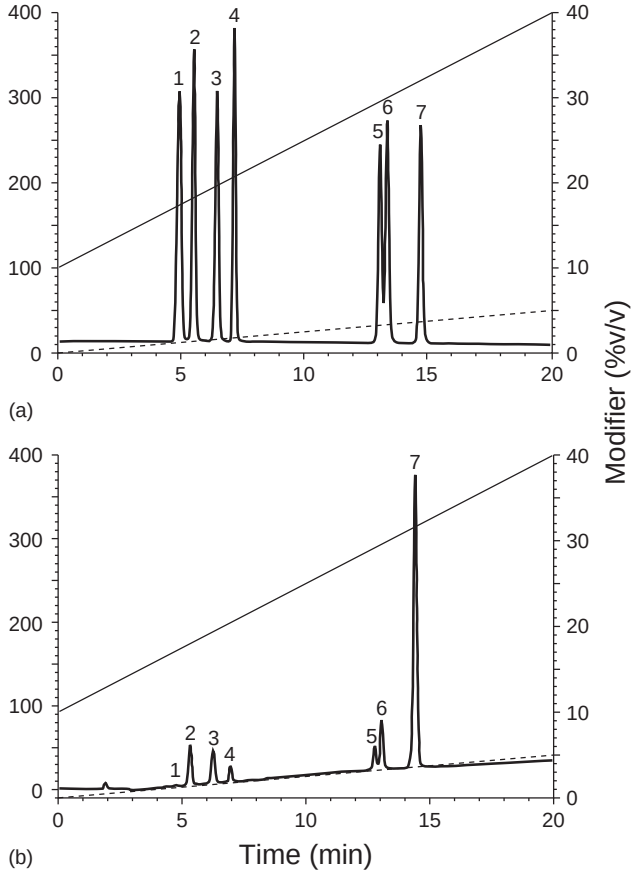

Figure 30 Chromatograms of a steroid mixture containing approximately 5 μg each of (1) pregnenolone, (2) estrone, (3) 4-androstene-3,17-dione, (4) β-estradiol, (5) cortisone, (6) hydrocortisone, and (7) prednisone injected in the column and separated by gradient normal-phase HPLC with (a) ELS and (b) UV detection at 260 nm. (Reproduced with kind permission of Kluwer Academic Publishers from C.E. Kibbey, *Mol. Diversity*, **1**, 247–258 (1995).)

amount of analyte present in the injected sample. ELS detectors exhibit a sigmoidal response to compounds, but, the detector's response can be made linear over a range of 2–3 orders of magnitude by plotting intensity versus mass on a log/log scale. The ELS detector's nonlinear response is attributed to the physics of Mie-scattering. The Mie-scattering efficiency of particles is maximal over a narrow range of particle diameters. Small radius particles generated at low analyte concentrations scatter less efficiently, while at high analyte concentrations changes in the surface-to-volume ratio of larger radius particles leads to less efficient light-scattering.

When combined with HPLC instruments, ELS detectors are highly effective at providing on-line quantitation of small organic compounds derived from combinatorial libraries.[196] Reliable quantitation requires that a well-characterized standard belonging to the same class of compounds present in the combinatorial library be used to calibrate the detector. The pair of chromatograms shown in Figure 30 illustrate the more "universal" response of ELS detection compared with conventional UV detection towards compounds of similar structural class (e.g. steroids). The response of the ELS detector depends on the ability of the nebulizer to produce analyte particles of the optimal size (0.8 μm) for effective Mie-scattering. This process is dependent on a variety of factors including solvent vaporization enthalpy, nebulizer gas flow rate, solvent composition, analyte molecular weight, and analyte polarity. The best results are obtained with HPLC separations performed under normal-phase conditions with solutes of a mass greater than 150–200 Dal.

ABBREVIATIONS AND ACRONYMS

ACE	Affinity Capillary Electrophoresis
ACE/MS	Affinity Capillary Electrophoresis and Mass Spectrometry
APCI	Atmospheric Pressure Chemical Ionization
APCIMS	Atmospheric Pressure Chemical Ionization Mass Spectrometry
ATR	Attenuated Total Reflectance
BACMS	Bio-affinity Characterization Mass Spectrometry
Boc	*t*-Butoxycarbonyl
CDNB	1-Chloro-2,4-Dinitrobenzene
CID	Collision-induced Dissociation
CIRCLE	Cylindrical Internal Reflectance
CLND	Chemiluminescent Nitrogen Detection
COSY	Correlation Spectroscopy
DECODES	Diffusion-encoded Spectroscopy
DMSO	Dimethyl Sulfoxide
DOSY	Diffusion-ordered Spectroscopy
DOSY/HMQC	Diffusion-ordered Spectroscopy/ Heteronuclear Multiple Quantum Coherence
DRIFT	Diffuse Reflectance Infrared Fourier Transform
ECD	Electron Capture Detection
ELS	Evaporative Light Scattering
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization Mass Spectrometry
FIA/MS	Flow Injection Analysis Mass Spectrometry
FID	Free Induction Decay
Fmoc	9-Fluorenylmethyloxycarbonyl
FTICR	Fourier Transform Ion Cyclotron Resonance
FTIR	Fourier Transform Infrared

FTRS	Fourier Transform Raman Spectroscopy
GC	Gas Chromatography
GST	Glutathione-S-transferase
HMQC	Heteronuclear Multiple Quantum Coherence
HPLC	High-performance Liquid Chromatography
HPLC/DRIFT	High-performance Liquid Chromatography/Diffuse Reflectance Infrared Fourier Transform
HPLC/FTIR	High-performance Liquid Chromatography/Fourier Transform Infrared
HPLC/NMR	High-performance Liquid Chromatography/Nuclear Magnetic Resonance
IR	Infrared
LC	Liquid Chromatography
LC/ESI	Liquid Chromatography/ Electrospray Ionization
LC/ESIMS	Liquid Chromatography Coupled with Electrospray Ionization Mass Spectrometry
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/NMR	Liquid Chromatography/Nuclear Magnetic Resonance
MAGIC/HPLC/ FTIR	Mono-disperse Aerosol Generation Interface for Combining High-performance Liquid Chromatography with Fourier Transform Infrared
MALDI	Matrix-assisted Laser Desorption/ Ionization
MALDI/FTMS	Matrix-assisted Laser Desorption/ Ionization Fourier Transform Mass Spectrometry
MALDI/TOFMS	Matrix-assisted Laser Desorption/ Ionization Time-of-flight Mass Spectrometry
MAS	Magic-angle Spinning
MCT	Mercury–Cadmium–Telluride
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PAFTIR	Photoacoustic Fourier Transform Infrared
PCR	Polymerase Chain Reaction
RF	Radiofrequency
SBFTIRM	Single-bead Fourier Transform Infrared Microspectroscopy
SECSY	Spin-echo Correlated Spectroscopy
SH2	Src homology 2
SPOS	Solid-phase Organic Synthesis
SPPS	Solid-phase Peptide Synthesis
SPS	Solid-phase Synthesis
TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TLC	Thin-layer Chromatography
TOCSY	Total Correlation Spectroscopy
TOF	Time-of-flight
TOF/SIMS	Time-of-flight Secondary Ion Mass Spectrometry
UV	Ultraviolet
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis • Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Affinity Chromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Liquid Chromatography/Mass Spectrometry • Tandem Mass Spectrometry: Fundamentals and Instrumentation

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • High-performance Liquid Chromatography Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton

REFERENCES

1. R.B. Merrifield, 'Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide', *J. Am. Chem. Soc.*, **85**, 2149–2154 (1963).
2. H. Rapoport, J.I. Crowley, 'Cyclization via Solid Phase Synthesis. Unidirectional Dieckmann Products from Solid Phase and Benzyl Triethylcarbinyl Pimelates', *J. Am. Chem. Soc.*, **92**(21), 6363–6365 (1970).
3. A. Patchornik, M. Kraus, 'Reactive Species Mutually Isolated on Insoluble Polymeric Carriers. I. Directed Monoacylation of Esters', *J. Am. Chem. Soc.*, **92**(26), 7587–7589 (1970).
4. C.C. Leznoff, J.Y. Wong, 'Use of Polymer Supports in Organic Synthesis. Synthesis of Monotrityl Ethers of Symmetrical Diols', *Can. J. Chem.*, **50**(17), 2892–2893 (1972).
5. M.H. Caruthers, 'Gene Synthesis Machines: DNA Chemistry and its Uses', *Science*, **230**(4723), 281–285 (1985).
6. H.M. Geysen, R.H. Meloen, S.J. Barteling, 'Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid', *Proc. Natl. Acad. Sci. USA*, **81**(13), 3998–4002 (1984).
7. A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, 'General Method for Rapid Synthesis of Multicomponent Peptide Mixtures', *Int. J. Pept. Protein Res.*, **37**(6), 487–493 (1991).
8. R.A. Houghten, 'General Method for the Rapid Solid-phase Synthesis of Large Numbers of Peptides: Specificity of Antigen–Antibody Interaction at the Level of Individual Amino Acids', *Proc. Natl. Acad. Sci. USA*, **82**(15), 5131–5135 (1985).
9. K.S. Lam, S.E. Salmon, E.M. Hersh, V.J. Hruby, W.M. Kazmierski, R.J. Knapp, 'A New Type of Synthetic Peptide Library for Identifying Ligand-binding Activity', *Nature*, **354**(6348), 84–86 (1991).
10. S.P.A. Fodor, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, D. Solas, 'Light-directed, Spatially Addressable Parallel Chemical Synthesis', *Science*, **251**(4995), 767–773 (1991).
11. L.A. Thompson, J.A. Ellman, 'Synthesis and Applications of Small Molecule Libraries', *Chem. Rev.*, **96**, 555–600 (1996).
12. N.K. Terrett, M. Gardner, D.W. Gordon, R.J. Kobylecki, J. Steele, 'Combinatorial Synthesis – The Design of Compound Libraries and their Application to Drug Discovery', *Tetrahedron*, **51**(30), 8135–8173 (1995).
13. H.M. Geysen, S.J. Rodda, T.J. Mason, G. Tribbick, P.G. Schoofs, 'Strategies for Epitope Analysis Using Peptide Synthesis', *J. Immunol. Methods*, **102**(2), 259–274 (1987).
14. H.M. Geysen, S.J. Rodda, T.J. Mason, 'A Priori Delineation of a Peptide which Mimics a Discontinuous Antigenic Determinant', *Mol. Immunol.*, **23**(7), 709–715 (1986).
15. J.M. Ostresh, J.H. Winkle, V.T. Hamashin, R.A. Houghten, 'Peptide Libraries: Determination of Relative Reaction Rates of Protected Amino Acids in Competitive Couplings', *Biopolymers*, **34**(12), 1681–1689 (1994).
16. T. Carell, E.A. Wintner, A. Bashir-Hashemi, J. Rebek, Jr, 'Screening Method for Isolation in Solution of Biologically Active Compounds from a Molecular Library', *Angew. Chem., Int. Ed. Engl.*, **33**(20), 2059–2061 (1994).
17. F. Sebestyen, G. Dibo, A. Kovacs, A. Furka, 'Chemical Synthesis of Peptide Libraries', *Bioorg. Med. Chem. Lett.*, **3**(3), 413–418 (1993).
18. A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, 'General Method for Rapid Synthesis of Multicomponent Peptide Mixtures', *Int. J. Pept. Protein Res.*, **37**(6), 487–493 (1991).
19. S. Brenner, R.A. Lerner, 'Encoded Combinatorial Chemistry', *Proc. Natl. Acad. Sci. USA*, **89**, 5381–5383 (1992).
20. J. Nielsen, S. Brenner, K.D. Janda, 'Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry', *J. Am. Chem. Soc.*, **115**, 9812–9813 (1993).
21. J.M. Kerr, S.C. Banville, R.N. Zuckermann, 'Encoded Combinatorial Peptide Libraries Containing Non-natural Amino Acids', *J. Am. Chem. Soc.*, **115**, 2529–2531 (1993).
22. S.E. Salmon, K.S. Lam, M. Lebl, A. Kandola, P.S. Khattri, S. Wade, M. Patek, P. Kocis, V. Krchnak, D. Thorpe, S. Felder, 'Discovery of Biologically Active Peptides in Random Libraries: Solution-phase Testing After Staged Orthogonal Release from Resin Beads', *Proc. Natl. Acad. Sci. USA*, **90**, 11 708–11 712 (1993).
23. B.B. Brown, D.S. Wagner, H.M. Geysen, 'A Single-bead Decode Strategy using Electrospray Ionization Mass Spectrometry and a New Photolabile Linker: 3-Amino-3-(2-nitrophenyl)propionic acid', *Molecular Diversity*, **1**, 4–12 (1995).
24. M.H.J. Ohlmeyer, R.N. Swanson, L.W. Dillard, J.C. Reader, G. Asouline, R. Kobayashi, M. Wigler, C.W. Still, 'Complex Synthetic Chemical Libraries Indexed with Molecular Tags', *Proc. Natl. Acad. Sci. USA*, **90**, 10 922–10 926 (1993).
25. H.P. Nestler, P.A. Bartlett, W.C. Still, 'A General Method for Molecular Tagging of Encoded Combinatorial Chemistry Libraries', *J. Org. Chem.*, **59**, 4723–4724 (1994).
26. J.C. Chabala, J.J. Baldwin, J.J. Burbaum, D. Chelsky, I.H. Dillard, G. Li, M.H.J. Ohlmeyer, T.L. Randle, J.C. Reader, L. Rokosz, N.H. Sigal, 'Binary Encoded Small-molecule Libraries in Drug Discovery and Optimization', *Perspect. Drug Disc. Des.*, **2**, 305–318 (1994).
27. J.J. Burbaum, M.H.J. Ohlmeyer, J.C. Reader, I. Henderson, L.W. Dillard, G. Li, T.L. Randle, N.H. Sigal, D. Chelsky, J.J. Baldwin, 'A Paradigm for Drug Discovery Employing Encoded Combinatorial Libraries', *Proc. Natl. Acad. Sci. USA*, **92**, 6027–6031 (1995).

28. Z.J. Ni, D. Maclean, C.P. Holmes, M.M. Murphy, B. Ruhland, J.W. Jacobs, E.M. Gordon, M.A. Gallop, 'Versatile Approach to Encoding Combinatorial Organic Synthesis using Chemically Robust Secondary Amine Tags', *J. Med. Chem.*, **39**, 1601–1608 (1996).
29. K.C. Nicolaou, X.Y. Xiao, Z. Parandoosh, A. Senyei, M.P. Nova, 'Radiofrequency Encoded Combinatorial Chemistry', *Angew. Chem., Int. Ed. Engl.*, **34**(20), 2289–2291 (1995).
30. E.J. Moran, S. Sarshar, J.F. Cargill, M.M. Shahbaz, A. Lio, A.M.M. Mjalli, R.W. Armstrong, 'Radiofrequency Tag Encoded Combinatorial Library Method for the Discovery of Tripeptide-substituted Cinnamic Acid Inhibitors of the Protein Tyrosine Phosphatase PTP1B', *J. Am. Chem. Soc.*, **117**, 10 787–10 788 (1995).
31. S. Shi, X.Y. Xiao, A. Czarnik, 'A Combinatorial Synthesis of Tyrphostins via the 'Directed Sorting' Method', *Biotechnol. Bioeng.*, **61**(1), 7–12 (1998).
32. J.I. Crowley, H. Rapoport, 'Cyclization via Solid Phase Synthesis. Unidirectional Diekmann Products from Solid Phase and Benzyl Triethylcarbinyl Pimelates', *J. Am. Chem. Soc.*, **92**, 6363–6365 (1970).
33. J.M. Frechet, C. Schuerch, 'Solid-phase Synthesis of Oligosaccharides. I. Preparation of the Solid Support. Poly[*p*-(propen-3-ol-1-yl)styrene]', *J. Am. Chem. Soc.*, **93**, 492–496 (1971).
34. J.M. Goldwasser, C.C. Leznoff, 'The Solid-phase Synthesis of Monoester Monoamides and Monoester Monoalcohols from Symmetrical Diacid Chlorides', *Can. J. Chem.*, **56**, 1562–1568 (1978).
35. B.D. Larsen, D.H. Christensen, A. Holm, R. Zillmer, O.F. Nielsen, 'The Merrifield Peptide Synthesis Studied by Near-infrared Fourier Transform Raman Spectroscopy', *J. Am. Chem. Soc.*, **115**, 6247–6253 (1993).
36. M. Narita, S. Honda, H. Umeyama, T. Ogura, 'Infrared Spectroscopic Conformational Analysis of Polystyrene Resin-bound Human Proinsulin C-peptide Fragments. β-Sheet Aggregation of Peptide Chains during Solid-phase Peptide Synthesis', *Bull. Chem. Soc. Jpn.*, **61**, 1201–1206 (1988).
37. M. Narita, S. Isokawa, S. Honda, H. Umeyama, H. Kakei, S. Obana, 'Individuality of Amino Acid Residues in Protected Peptides. Conformational and β-Sheet Structure-disrupted Behaviors of Resin-bound Peptides', *Bull. Chem. Soc. Jpn.*, **62**, 773–779 (1989).
38. J.I. Crowley, H. Rapoport, 'Unidirectional Diekmann Cyclizations on a Solid Phase and in Solution', *J. Org. Chem.*, **45**, 3215–3227 (1980).
39. B. Yan, G. Kumaravel, 'Probing Solid-phase Reactions by Monitoring the IR Bands of Compounds on a Single 'Flattened' Resin Bead', *Tetrahedron*, **52**(3), 843–848 (1996).
40. B. Yan, G. Kumaravel, H. Anjaria, A. Wu, R.C. Petter, C.F. Jewel, J.R. Wareing, 'Infrared Spectrum of a Single Resin Bead for Real-time Monitoring of Solid-phase Reactions', *J. Org. Chem.*, **60**, 5736–5738 (1995).
41. B. Yan, Q. Sun, J.R. Wareing, C.F. Jewell, 'Real-time Monitoring of the Catalytic Oxidation of Alcohols to Aldehydes and Ketones on Resin Support by Single-bead Fourier Transform Infrared Microspectroscopy', *J. Org. Chem.*, **61**, 8765–8770 (1996).
42. B. Yan, J.B. Fell, G. Kumaravel, 'Progression of Organic Reactions on Resin Supports Monitored by Single Bead FTIR Microspectroscopy', *J. Org. Chem.*, **61**, 7467–7472 (1996).
43. B. Yan, H. Gstach, 'An Imadazole Synthesis on Solid Support Monitored by Single Bead FTIR Microspectroscopy', *Tetrahedron Lett.*, **37**, 8325–8328 (1996).
44. R.E. Marti, B. Yan, M.A. Jarosinski, 'Solid-phase Synthesis via 5-Oxazolidinones. Ring-opening Reactions with Amines and Reaction Monitoring by Single-bead FTIR Microspectroscopy', *J. Org. Chem.*, **62**, 5615–5618 (1997).
45. R.V. Hoffman, 'Progression of Organic Reactions on Resin Supports Monitored by Single-bead FTIR Microspectroscopy', *Chemtracts*, **10**, 822–827 (1997).
46. I. Deben, J. Goorden, E. van Doornum, H. Ovaa, E. Kellenbach, 'MicroDRIFT: Rapid and Efficient IR Analysis of Solid Phase Organic Chemistry Reactions', *Eur. J. Org. Chem.*, **4**, 697–700 (1998).
47. F. Gosselin, M. DiRenzo, T.H. Ellis, W.D. Lubell, 'Photoacoustic FTIR Spectroscopy, a Nondestructive Method for Sensitive Analysis of Solid-phase Organic Chemistry', *J. Org. Chem.*, **61**, 7980–7981 (1996).
48. H.U. Gremlich, S.L. Berets, 'Use of FTIR Internal Reflection Spectroscopy in Combinatorial Chemistry', *Appl. Spectrosc.*, **50**(4), 532–536 (1996).
49. D.E. Pivonka, K. Russell, T. Gero, 'Tools for Combinatorial Chemistry: In Situ Infrared Analysis of Solid-phase Organic Reactions', *Applied Spectrosc.*, **50**, 1471–1478 (1996).
50. D.W. Vidrine, D.R. Mattson, 'A Practical Real-time Fourier Transform Infrared Detector for Liquid Chromatography', *Appl. Spectrosc.*, **32**(5), 502–506 (1978).
51. D.W. Vidrine, 'Use of Subtractive Techniques in Interpreting On-line FTIR Spectra of HPLC Column Eluates', *J. Chromatogr. Sci.*, **17**(8), 477–482 (1979).
52. K.H. Shafer, S.V. Lucas, R.J. Jakobsen, 'Application of the Combined Analytical Techniques of HPLC/FTIR, GC/FTIR, and GC-MS to the Analysis of Real Samples', *J. Chromatogr. Sci.*, **17**(8), 464–470 (1979).
53. R.S. Brown, L.T. Taylor, 'Microbore Liquid Chromatography with Flow Cell Fourier Transform Infrared Spectrometric Detection', *Anal. Chem.*, **55**, 1492–1497 (1983).
54. C.C. Johnson, L.T. Taylor, 'Normal-phase Liquid Chromatography/Fourier Transform Infrared Spectrometry for Analysis of Nonpolar Material with Semipreparative, Analytical, and Microbore Columns', *Anal. Chem.*, **55**, 436–441 (1983).
55. C.C. Johnson, L.T. Taylor, 'Zero Dead Volume Flow Cell for Microbore Liquid Chromatography with Fourier

Transform Infrared Spectrometric Detection', *Anal. Chem.*, **56**, 2642–2647 (1984).
56. P.T. McKittrick, N.D. Danielson, J.E. Katon, 'A Comparison between a Micro and an Ultramicro Circle® Cell for On-line FTIR Detection in a Reverse Phase HPLC System', *J. Liq. Chromatogr.*, **14**(2), 377–393 (1991).
57. D. Kuehl, P.R. Griffiths, 'Novel Approaches to Interfacing a High Performance Liquid Chromatograph with a Fourier Transform Infrared Spectrometer', *J. Chromatogr. Sci.*, **17**(8), 471–476 (1979).
58. C.M. Conroy, P.R. Griffiths, K. Jinno, 'Interface of a Microbore High-performance Liquid Chromatograph with a Diffuse Reflectance Fourier Transform Infrared Spectrometer', *Anal. Chem.*, **57**(4), 822–825 (1985).
59. V.F. Kalasinsky, K.G. Whitehead, R.C. Kenton, J.A.S. Smith, K.S. Kalasinsky, 'HPLC/FTIR Interface for Normal- and Reversed-phase Analytical Columns', *J. Chromatogr. Sci.*, **25**, 273–280 (1987).
60. M.A. Castles, L.V. Azarraga, L.A. Carreira, 'Continuous, On-line Interface for Reversed-phase Microbore High-performance Liquid Chromatography/Diffuse Reflectance Infrared Fourier Transform Analysis', *Appl. Spectrosc.*, **40**(5), 673–680 (1986).
61. C.M. Conroy, P.R. Griffiths, P.J. Duff, L.V. Azarraga, 'Interface of a Reverse-phase High-performance Liquid Chromatograph with a Diffuse Reflectance Fourier Transform Infrared Spectrometer', *Anal. Chem.*, **56**(14), 2636–2642 (1984).
62. A.M. Robertson, D. Littlejohn, M. Brown, C.J. Dowle, 'Assessment of a Thermospray Interface for the Coupling of High-performance Liquid Chromatography and Fourier Transform Infrared Spectrometry', *J. Chromatogr.*, **588**, 15–24 (1991).
63. R.C. Willoughby, R.F. Browner, 'Monodisperse Aerosol Generation Interface for Combining Liquid Chromatography with Mass Spectroscopy', *Anal. Chem.*, **56**(14), 2625–2631 (1984).
64. V.E. Turula, J.A. de Haseth, 'Particle Beam LC/FTIR Spectrometry Studies of Biopolymer Conformations in Reversed-phase HPLC Separations: Native Globular Proteins', *Anal. Chem.*, **68**, 629–638 (1996).
65. R.M. Robertson, J.A. de Haseth, J.D. Kirk, R.F. Browner, 'MAGIC-LC/FTIR Spectrometry: Preliminary Studies', *Appl. Spectrosc.*, **42**(8), 1365–1368 (1988).
66. R.M. Robertson, J.A. de Haseth, R.F. Browner, 'MAGIC-LC/FT-IR Spectrometry', *Mikrochim. Acta*, **2**(1–6), 199–202 (1988).
67. J.J. Gagel, K. Biemann, 'Continuous Recording of Reflection–Absorbance Fourier Transform Infrared Spectra of the Effluent of a Microbore Liquid Chromatograph', *Anal. Chem.*, **58**, 2184–2189 (1986).
68. J.J. Gagel, K. Biemann, 'Continuous Infrared Spectroscopic Analysis of Isocratic and Gradient Elution Reversed-phase Liquid Chromatography Separations', *Anal. Chem.*, **59**, 1266–1272 (1987).
69. G.W. Somsen, E.W.J. Hooijschuur, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, T. Visser, 'Coupling of Reversed-phase Liquid Column Chromatography and Fourier Transform Infrared Spectrometry Using Post-column On-line Extraction and Solvent Elimination', *Anal. Chem.*, **68**, 746–752 (1996).
70. R.S. Youngquist, G.R. Fuentes, M.P. Lacey, T. Keough, 'Matrix-assisted Laser Desorption Ionization for Rapid Determination of the Sequences of Biologically Active Peptides Isolated from Support-bound Combinatorial Peptide Libraries', *Rapid Commun. Mass Spectrom.*, **8**, 77–81 (1994).
71. B.J. Egner, G.J. Langley, M. Bradley, 'Solid Phase Chemistry: Direct Monitoring by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry. A Tool for Combinatorial Chemistry', *J. Org. Chem.*, **60**(9), 2652–2653 (1995).
72. B.J. Egner, M. Cardno, M. Bradley, 'Linkers for Combinatorial Chemistry and Reactions Analysis using Solid Phase In Situ Mass Spectrometry', *J. Chem. Soc., Chem. Commun.*, **21**, 2163–2164 (1995).
73. G. Talbo, J.D. Wade, N. Dawson, M. Manoussios, G.W. Tregear, 'Rapid Semi-on-line Monitoring of Fmoc Solid-phase Peptide Synthesis by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Lett. Pept. Sci.*, **4**, 121–127 (1997).
74. Z. Yu, X.C. Yu, Y.H. Chu, 'MALDI-MS Determination of Cyclic Peptidomimetic Sequences on Single Beads Directed Toward the Generation of Libraries', *Tetrahedron Lett.*, **39**, 1–4 (1998).
75. C.L. Brummel, I.N.W. Lee, Y. Zhou, S.J. Benkovic, N. Winograd, 'A Mass Spectrometric Solution to Address Problem of Combinatorial Libraries', *Science*, **264**, 399–402 (1994).
76. R.A. Zambias, D.A. Boulton, P.R. Griffin, 'Microchemical Structural Determination of a Peptoid Covalently Bound to a Polymeric Bead by Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry', *Tetrahedron Lett.*, **35**, 4283–4286 (1994).
77. N.J. Haskins, D.J. Hunter, A.J. Organ, S.S. Rahman, C. Thom, 'Combinatorial Chemistry: Direct Analysis of Bead Surface Associated Materials', *Rapid Commun. Mass Spectrom.*, **9**, 1437–1440 (1995).
78. C.L. Brummel, J.C. Vickerman, S.A. Carr, M.E. Hemling, G.D. Roberts, W. Johnson, J. Weinstock, D. Gaitanopoulos, S.J. Benkovic, N. Winograd, 'Evaluation of Mass Spectrometric Methods Applicable to the Direct Analysis of Non-peptide Bead-bound Combinatorial Libraries', *Anal. Chem.*, **68**, 237–242 (1996).
79. K. Berlin, R.K. Jain, C. Tetzlaff, C. Steinbeck, C. Richert, 'Spectrometrically Monitored Selection Experiments: Quantitative Laser Desorption Mass Spectrometry of Small Chemical Libraries', *Chem. Biol.*, **4**(1), 63–77 (1997).
80. D.C. Tutko, K.D. Henry, B.E. Winger, H. Stout, M. Hemling, 'Sequential Mass Spectrometry and

MS^n Analysis of Combinatorial Libraries by Using Automated Matrix-assisted Laser Desorption/Ionization Fourier Transform Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**, 335–338 (1998).
81. C.K. Meng, J.B. Fenn, 'Formation of Charged Clusters During Electrospray Ionization of Organic Solute Species', *Org. Mass Spectrom.*, **26**(6), 542–549 (1991).
82. J.B. Fenn, 'Ion Formation from Charged Droplets: Roles of Geometry, Energy and Time', *J. Am. Soc. Mass Spectrom.*, **4**(7), 524–535 (1993).
83. P. Kebarle, L. Tang, 'From Ions in Solution to Ions in the Gas Phase–the Mechanism of Electrospray Mass Spectrometry', *Anal. Chem.*, **65**(22), 972A–986A (1993).
84. J.B. Fenn, M. Mann, K. Chin, S.F. Wong, C.M. Whitehouse, 'Electrospray Ionization – Principles and Practice', *Mass Spectrom. Rev.*, **9**(1), 37–70 (1990).
85. A.P. Bruins, 'Electrospray, Technique and Applications', *J. Chim. Phys. Phys.-Chim. Biol.*, **90**(6), 1335–1344 (1993).
86. B.A. Andrien, J.G. Boyle, 'Electrospray and APCI Interface for Mass Spectrometry', *Spectroscopy*, **10**(2), 42–44 (1995).
87. M. Staòková, O. Issakova, N.F. Sepetov, V. Krchòák, K.S. Lam, M. Lebl, 'Application of One-bead One-structure Approach to Identification of Nonpeptidic Ligands', *Drug Dev. Res.*, **33**, 146–156 (1994).
88. A.M. Bray, D.S. Chiefari, R.M. Valerio, N.J. Maeji, 'Rapid Optimization of Organic Reactions on Solid Phase using the Multipin Approach: Synthesis of 4-Aminoproline Analogous by Reductive Amination', *Tetrahedron Lett.*, **36**(28), 5081–5084 (1995).
89. W.J. Haap, J.W. Metzger, C. Kempter, G. Jung, 'Composition and Purity of Combinatorial Aryl Ether Collections Analyzed by Electrospray Mass Spectrometry', *Molecular Diversity*, **3**, 29–41 (1997).
90. D.G. Powers, D.S. Casebier, D. Fokas, W.J. Ryan, J.R. Troth, D.L. Coffen, 'Automated Parallel Synthesis of Chalcone-based Screening Libraries', *Tetrahedron*, **54**, 4085–4096 (1998).
91. H. An, B.D. Haly, A.S. Fraser, C.J. Guinosso, D.P. Cook, 'Solution Phase Combinatorial Chemistry. Synthesis of Novel Linear Pyridinopolyamine Libraries with Potent Antibacterial Activity', *J. Org. Chem.*, **62**, 5156–5164 (1997).
92. Y. Dunayevskiy, P. Vouros, T. Carell, E.A. Winter, J. Rebek, Jr, 'Characterization of the Complexity of Small-molecule Libraries by Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **67**, 2906–2915 (1995).
93. J.N. Kyranos, J.C. Hogan, Jr, 'High-throughput Characterization of Combinatorial Libraries Generated by Parallel Synthesis', *Anal. Chem.*, **70**, 389A–395A (1998).
94. K.F. Blom, 'Strategies and Data Precision Requirements for the Mass Spectrometric Determination of Structures from Combinatorial Mixtures', *Anal. Chem.*, **69**, 4354–4362 (1997).
95. I. Hughes, 'Design of Self-coded Combinatorial Libraries to Facilitate Direct Analysis of Ligands by Mass Spectrometry', *J. Med. Chem.*, **41**(20), 3804–3811 (1998).
96. J.W. Metzger, K.H. Wiesmüller, V. Gnau, J. Brünjes, G. Jung, 'Ion-spray Mass Spectrometry and High-performance Liquid Chromatography–Mass Spectrometry of Synthetic Peptide Libraries', *Angew. Chem., Int. Ed. Engl.*, **32**, 894–896 (1993).
97. J.W. Metzger, S. Stevanoviæ, J. Brünjes, K.H. Wiesmüller, G. Jung, 'Electrospray Mass Spectrometry and Multiple Sequence Analysis of Synthetic Peptide Libraries', *Methods (San Diego)*, **6**, 425–431 (1994).
98. J.W. Metzger, C. Kempter, K.H. Wiesmüller, G. Jung, 'Electrospray Mass Spectrometry and Tandem Mass Spectrometry of Synthetic Multicomponent Peptide Mixtures: Determination of Composition and Purity', *Anal. Biochem.*, **219**(2), 261–277 (1994).
99. P.H. Lambert, J.A. Boutin, S. Bertin, J.L. Fauchère, J.P. Volland, 'Evaluation of High Performance Liquid Chromatography/Electrospray Mass Spectrometry with Selected Ion Monitoring for the Analysis of Large Synthetic Combinatorial Peptide Libraries', *Rapid Commun. Mass Spectrom.*, **11**, 1971–1976 (1997).
100. P.A. Demirev, R.A. Zubarev, 'Probing Combinatorial Library Diversity by Mass Spectrometry', *Anal. Chem.*, **69**, 2893–2900 (1997).
101. S.S. Smart, T.J. Mason, P.S. Bennell, N.J. Maeij, H.M. Geysen, 'High-throughput Purity Estimation and Characterisation of Synthetic Peptides by Electrospray Mass Spectrometry', *Int. J. Peptide Protein Res.*, **47**, 47–55 (1996).
102. H. Iwabuchi, E. Kitazawa, N. Kobayashi, H. Watanabe, M. Kanai, K. Nakamura, 'Studies on Drug Metabolism Using Liquid Chromatography/Mass Spectrometry: Comparison of Three Liquid Chromatographic/Mass Spectrometric Interfaces', *Biol. Mass Spectrom.*, **23**, 540–546 (1994).
103. G. Hegy, E. Görlach, R. Richmond, F. Bitsch, 'High Throughput Electrospray Mass Spectrometry of Combinatorial Chemistry Racks with Automated Contamination Surveillance and Results Reporting', *Rapid Commun. Mass Spectrom.*, **10**, 1894–1900 (1996).
104. A.A. MacDonald, S.H. DeWitt, E.M. Hogan, R. Ramage, 'A Solid Phase Approach to Quinolones using the Diversomer® Technology', *Tetrahedron Lett.*, **37**, 4815–4818 (1996).
105. J.R. Booth, J.C. Hodges, 'Polymer-supported Quenching Reagents for Parallel Purification', *J. Am. Chem. Soc.*, **119**, 4882–4886 (1997).
106. M.W. Creswell, G.L. Bolton, J.C. Hodges, M. Meppen, 'Combinatorial Synthesis of Dihydropyridone Libraries and their Derivatives', *Tetrahedron*, **54**, 3983–3998 (1998).
107. R.E. March, R.J. Strife, C.S. Creaser, in *Practical Aspects of Ion-trap Mass Spectrometry*, eds. R.E. March, J.F. Todd, CRC Press, Boca Raton, Vol. 3, 27–88, 1995.

108. T. Keough, M.P. Lacey, M.M. Ketcha, R.H. Bateman, M.R. Green, 'Orthogonal Acceleration Single-pass Time-of-flight Mass Spectrometry for Determination of the Exact Masses of Product Ions Formed in Tandem Mass Spectrometry Experiments', *Rapid Commun. Mass Spectrom.*, **11**, 1702–1708 (1997).
109. S.C. Pomerantz, J.A. McCloskey, T.M. Tarasow, B.E. Eaton, 'Deconvolution of Combinatorial Oligonucleotide Libraries by Electrospray Ionization Tandem Mass Spectrometry', *J. Am. Chem. Soc.*, **119**, 3861–3867 (1997).
110. S.J. Lane, A. Pipe, 'Unambiguous Bead Decoding by Microelectrospray Capillary Electrochromatography Tandem Mass Spectrometry of Dansylated Secondary Amine Tags from Encoded Combinatorial Organic Synthesis', *Rapid Commun. Mass Spectrom.*, **12**, 667–674 (1998).
111. P.A. Keifer, 'Influence of Resin Structure, Tether Length, and Solvent upon the High-resolution ^{1}H NMR Spectra of Solid-phase-synthesis Resins', *J. Org. Chem.*, **61**, 1558–1559 (1996).
112. P.A. Keifer, L. Baltusis, D.A. Rice, A.A. Tymiak, J.N. Shoolery, 'A Comparison of NMR Spectra Obtained for Solid-phase-synthesis Resins Using Conventional High-resolution, Magic-angle-spinning, and High-resolution Magic-angle-spinning Probes', *J. Magn. Reson., Ser. A*, **119**, 65–75 (1996).
113. H. Sternlicht, G.L. Kenyon, E.L. Packer, J. Sinclair, 'Carbon-13 Nuclear Magnetic Resonance Studies of Heterogeneous Systems. Amino Acids Bound to Cationic Exchange Resins', *J. Am. Chem. Soc.*, **93**, 199–208 (1971).
114. D. Leibfritz, W. Mayr, R. Oekonomopulos, G. Jung, '^{13}C NMR Spectroscopic Studies on the Conformation During Stepwise Synthesis of Peptides Bound to Solubilizing Polymer Supports', *Tetrahedron*, **34**, 2045–2050 (1978).
115. R. Epton, P. Goddard, K.J. Ivin, 'Gel Phase ^{13}C NMR Spectroscopy as an Analytical Method in Solid (Gel) Phase Peptide Synthesis', *Polymer*, **21**, 1367–1371 (1980).
116. E. Giralt, J. Rizo, E. Pedroso, 'Application of Gel-phase ^{13}C-NMR to Monitor Solid Phase Peptide Synthesis', *Tetrahedron*, **40**, 4141–4152 (1984).
117. S.L. Manatt, D. Horowitz, R. Horowitz, R.P. Pinnell, 'Solvent Swelling for Enhancement of Carbon-13 Nuclear Magnetic Resonance Spectral Information from Insoluble Polymers: Chloromethylation Levels in Crosslinked Polystyrenes', *Anal. Chem.*, **52**, 1529–1532 (1980).
118. A.J. Jones, C.C. Leznoff, P.I. Svirskaya, 'Characterization of Organic Substrates Bound to Cross-linked Polystyrenes by ^{13}C-NMR Spectroscopy', *Org. Mag. Reson.*, **18**, 236–240 (1982).
119. E.C. Blossey, R.G. Cannon, W.T. Ford, M. Periyasamy, S. Mohanraj, 'Synthesis, Reactions, and ^{13}C FT NMR Spectroscopy of Polymer-bound Steroids', *J. Org. Chem.*, **55**, 4664–4668 (1990).
120. G.C. Look, C.P. Holmes, J.P. Chinn, M.A. Gallop, 'Methods for Combinatorial Organic Synthesis: The Use of Fast ^{13}C NMR Analysis for Gel Phase Reaction Monitoring', *J. Org. Chem.*, **59**, 7588–7590 (1994).
121. G.C. Look, M.M. Murphy, D.A. Campbell, M.A. Gallop, 'Trimethylorthoformate: A Mild and Effective Dehydrating Reagent for Solution and Solid Phase Imine Formation', *Tetrahedron Lett.*, **36**(17), 2937–2940 (1995).
122. S.L. Manatt, C.F. Amsden, C.A. Bettison, W.T. Frazer, J.T. Gudman, B.E. Lenk, J.F. Lubetich, E.A. McNelly, S.C. Smith, D.J. Templeton, R.P. Pinnell, 'A Fluorine-19 NMR Approach for Studying Merrifield Solid-phase Peptide Synthesis', *Tetrahedron Lett.*, **21**, 1397–1400 (1980).
123. E.E. Swayze, 'The Solid Phase Synthesis of Trisubstituted 1,4-Diazabicyclo[4.3.0]nonan-2-one Scaffolds: On Bead Monitoring of Heterocycle Forming Reactions Using ^{15}N NMR', *Tetrahedron Lett.*, **38**(50), 8643–8646 (1997).
124. F. Bardella, R. Eritja, E. Pedroso, E. Giralt, 'Gel-Phase ^{31}P-NMR. A New Analytical Tool to Evaluate Solid Phase Oligonucleotide Synthesis', *Bioorg. Med. Chem. Lett.*, **3**, 2793–2796 (1993).
125. Z. Tian, C. Gu, R.W. Roeske, M. Zhou, R.L. van Etten, 'Synthesis of Phosphotyrosine-containing Peptides by the Solid-phase Method', *Int. J. Peptide Protein Res.*, **42**, 155–158 (1993).
126. C.R. Johnson, B. Zhang, 'Solid-phase synthesis of Alkenes Using the Horner–Wadsworth–Emmons Reaction and Monitoring by Gel-phase ^{31}P NMR', *Tetrahedron Lett.*, **36**(51), 9253–9256 (1995).
127. R. Quarrell, T.D.W. Claridge, G.W. Weaver, G. Lowe, 'Structure and Properties of TentaGel™ Resin Beads: Implications for Combinatorial Library Chemistry', *Molecular Diversity*, **1**, 223–232 (1995).
128. B. Schneider, D. Doskocilova, J. Dubal, 'Motional Restrictions and Chain Conformation in Various Swollen Crosslinked Polystyrene Gels from ^{1}H-NMR Line-shape Analysis', *Polymer*, **26**, 253–259 (1985).
129. W.L. Fitch, G. Detre, C.P. Holmes, J.N. Shoolery, P.A. Keifer, 'High-resolution ^{1}H NMR in Solid-phase Organic Synthesis', *J. Org. Chem.*, **59**, 7955–7956 (1994).
130. M. Pursch, G. Schlotterbeck, L.H. Tseng, K. Albert, W. Rapp, 'Monitoring the Reaction Progress in Combinatorial Chemistry: ^{1}H MAS NMR Investigations on Single Macro Beads in the Suspended State', *Angew. Chem., Int. Ed. Eng.*, **35**, 2867–2869 (1996).
131. S.K. Sarkar, R.S. Garigipati, J.L. Adams, P.A. Keifer, 'An NMR Method to Identify Nondestructively Chemical Compounds Bound to a Single Solid-phase-synthesis Bead for Combinatorial Chemistry Applications', *J. Am. Chem. Soc.*, **118**, 2305–2306 (1996).
132. J. Chin, B. Fell, M.J. Shapiro, J. Tomesch, J.R. Wareing, A.M. Bray, *J. Org. Chem.*, **62**, 538–539 (1997).

133. R.S. Garigipati, B. Adams, J.L. Adams, S.K. Sarkar, 'Use of Spin Echo Magic Angle Spinning ^{1}H NMR in Reaction Monitoring in Combinatorial Organic Synthesis', *J. Org. Chem.*, **61**, 2911–2914 (1996).
134. H.D.H. Stöver, J.M.J. Fréchet, 'Direct Polarization ^{13}C and ^{1}H Magic Angle Spinning NMR in the Characterization of Solvent-swollen Gels', *Macromolecules*, **22**, 1574–1576 (1989).
135. R.C. Anderson, M.A. Jarema, M.J. Shapiro, J.P. Stokes, M. Ziliox, 'Analytical Techniques in Combinatorial Chemistry: MAS CH Correlation in Solvent-swollen Resins', *J. Org. Chem.*, **60**, 2650–2651 (1995).
136. P.M. Macdonald, M.J. Damha, K. Ganeshan, R. Braich, S.V. Zabarylo, 'Phosphorus 31 Solid State NMR Characterization of Oligonucleotides Covalently Bound to a Solid Support', *Nucleic Acids Res.*, **24**, 2868–2876 (1996).
137. R.C. Anderson, J.P. Stokes, M.J. Shapiro, 'Structure Determination in Combinatorial Chemistry: Utilization of Magic Angle Spinning HMQC and TOCSY NMR Spectra in the Structure Determination of Wang-bound Lysine', *Tetrahedron Lett.*, **36**(30), 5311–5314 (1995).
138. M.J. Shapiro, J. Chin, R.E. Marti, M.A. Jarosinski, 'Enhanced Resolution in MAS NMR for Combinatorial Chemistry', *Tetrahedron Lett.*, **38**(8), 1333–1336 (1997).
139. J. Chin, B. Fell, S. Pochapsky, M.J. Shapiro, J.R. Wareing, '2D SECSY NMR for Combinatorial Chemistry. High-resolution MAS Spectra for Resin-bound Molecules', *J. Org. Chem.*, **63**, 1309–1311 (1998).
140. C. Dhalluin, C. Boutillon, A. Tartar, G. Lippens, 'Magic Angle Spinning Nuclear Magnetic Resonance in Solid-phase Peptide Synthesis', *J. Am. Chem. Soc.*, **119**, 10 494–10 500 (1997).
141. R. Jelinek, A.P. Valente, K.G. Valentine, S.J. Opella, 'Two-dimensional NMR Spectroscopy of Peptides on Beads', *J. Mag. Res.*, **125**, 185–187 (1997).
142. N. Watanabe, E. Niki, 'Direct-coupling of FT-NMR to High Performance Liquid Chromatography', *Proc. Jpn. Acad., Ser. B*, **54**(4), 194–199 (1978).
143. E. Bayer, K. Albert, M. Nieder, E. Grom, T. Keller, 'On-line Coupling of High-performance Liquid Chromatography and Nuclear Magnetic Resonance', *J. Chromatogr.*, **186**, 497–507 (1979).
144. H.C. Dorn, '^{1}H-NMR: A New Detector for Liquid Chromatography', *Anal. Chem.*, **56**, 747A–758A (1984).
145. E. Bayer, K. Albert, M. Nieder, E. Gromm, G. Wolff, M. Rindlisbacher, 'On-line Coupling of Liquid Chromatography and High-field Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **54**(11), 1747–1750 (1982).
146. D.A. Laude, C.L. Wilkins, 'Nuclear Magnetic Resonance Detection for the On-line Identification of Liquid Chromatography Eluents', *Trends Anal. Chem.*, **5**, 230–235 (1986).
147. K. Albert, 'On-line Use of NMR Detection in Separation Chemistry', *J. Chromatogr. A*, **703**, 123–147 (1995).
148. N. Wu, T.L. Peck, A.G. Webb, R.L. Magin, J.V. Sweedler, '^{1}H-NMR Spectroscopy on the Nanoliter Scale for Static and On-line Measurements', *Anal. Chem.*, **66**, 3849–3857 (1994).
149. B. Behnke, G. Schlotterbeck, U. Tallarek, S. Strohschein, L.H. Tseng, T. Keller, K. Albert, E. Bayer, 'Capillary HPLC-NMR Coupling: High-resolution ^{1}H NMR Spectroscopy in the Nanoliter Scale', *Anal. Chem.*, **68**(7), 1110–1115 (1996).
150. R. Subramanian, A.G. Webb, 'Design of Solenoidal Microcoils for High-resolution ^{13}C NMR Spectroscopy', *Anal. Chem.*, **70**, 2454–2458 (1998).
151. U.G. Sidelmann, C. Gavaghan, H.A.J. Carless, M. Spraul, M. Hofmann, J.C. Lindon, I.D. Wilson, J.K. Nicholson, '750-MHz Directly Coupled HPLC-NMR: Application to the Sequential Characterization of the Positional Isomers and Anomers of 2-, 3-, and 4-Fluorobenzoic Acid Glucuronides in Equilibrium Mixtures', *Anal. Chem.*, **67**, 4441–4445 (1995).
152. K. Albert, M. Kunst, E. Bayer, H.J. DeJong, P. Genissel, M. Spraul, W. Bermel, 'Investigation of Cyclopropyl-containing Drugs by On-line High Performance Liquid Chromatography/Nuclear Magnetic Resonance', *Anal. Chem.*, **61**, 772–775 (1989).
153. M. Spraul, M. Hofmann, P. Dvortsak, J.K. Nicholson, I.D. Wilson, 'High-performance Liquid Chromatography Coupled to High-field Proton Nuclear Magnetic Resonance Spectroscopy: Application to the Urinary Metabolites of Ibuprofen', *Anal. Chem.*, **65**, 327–330 (1993).
154. M. Spraul, M. Hofmann, J.C. Lindon, R.D. Farrant, M.J. Seddon, J.K. Nicholson, I.D. Wilson, 'Evaluation of Liquid Chromatography Coupled with High-field ^{1}H NMR Spectroscopy for Drug Metabolite Detection and Characterization: The Identification of Paracetamol Metabolites in Urine and Bile', *NMR Biomed.*, **7**, 295–303 (1994).
155. M. Spraul, M. Hofmann, J.C. Lindon, J.K. Nicholson, I.D. Wilson, 'Metabolite Investigation by HPLC Coupled With High-field NMR: State-of-the-art', *Methodol. Surv. Bioanal. Drugs*, **23**, 21–32 (1994).
156. J. Chin, J.B. Fell, M. Jarosinski, M.J. Shapiro, J.R. Wareing, 'HPLC/NMR in Combinatorial Chemistry', *J. Org. Chem.*, **63**, 386–390 (1998).
157. E. Bayer, K. Albert, M. Niedler, E. Gromm, G. Wolff, M. Rindlisbacher, 'On-line Coupling of Liquid Chromatography and High-field Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **54**, 1747–1750 (1982).
158. G.M. Clore, B.J. Kimber, A.M. Gronenborn, 'The 1–1 Hard Pulse: A Simple and Effective Method of Water Resonance Suppression in FT ^{1}H NMR', *J. Magn. Reson.*, **54**, 170–173 (1983).

159. D.L. Turner, 'Binomial Solvent Suppression', *J. Magn. Reson.*, **54**(1), 146–148 (1983).
160. R.E. Hurd, 'Gradient-enhanced Spectroscopy', *J. Magn. Reson.*, **87**(2), 422–428 (1990).
161. G. Wider, K. Wuethrich, 'A Simple Experimental Scheme Using Pulsed Field Gradients for Coherence-pathway Rejection and Solvent Suppression in Phase-sensitive Heteronuclear Correlation Spectra', *J. Magn. Reson., Ser. B*, **102**(2), 239–241 (1993).
162. T.L. Hwang, A.J. Shaka, 'Water Suppression that Works. Excitation Sculpting using Arbitrary Waveforms and Pulsed Field Gradients', *J. Magn. Reson., Ser. A*, **112**(2), 275–279 (1995).
163. S.H. Smallcombe, S.L. Patt, P.A. Keifer, 'WET Solvent Suppression and its Application to LC NMR and High-resolution NMR Spectroscopy', *J. Magn. Reson., Ser. A*, **117**(2), 295–303 (1995).
164. K.F. Morris, C.S. Johnson, Jr, 'Diffusion-ordered Two-dimensional Nuclear Magnetic Resonance Spectroscopy', *J. Am. Chem. Soc.*, **114**, 3139–3141 (1992).
165. K.F. Morris, C.S. Johnson, Jr, 'Resolution of Discrete and Continuous Molecular Size Distributions by Means of Diffusion-ordered 2D NMR Spectroscopy', *J. Am. Chem. Soc.*, **115**, 4291–4299 (1993).
166. H. Barjat, G.A. Morris, A.G. Swanson, 'A Three-dimensional DOSY/HMQC Experiment for the High-resolution Analysis of Complex Mixtures', *J. Magn. Reson.*, **131**, 131–138 (1998).
167. M. Lin, M.J. Shapiro, 'Mixture Analysis in Combinatorial Chemistry: Application of Diffusion-resolved NMR Spectroscopy', *J. Org. Chem.*, **61**, 7617–7619 (1996).
168. M. Lin, M.J. Shapiro, J.R. Wareing, 'Diffusion-edited NMR-affinity NMR for Direct Observation of Molecular Interactions', *J. Am. Chem. Soc.*, **119**, 5249–5250 (1997).
169. M.L. Nedved, S. Habibi-Goudarzi, B. Ganem, J.D. Henion, 'Characterization of Benzodiazepine ''Combinatorial'' Chemical Libraries by Online Immunoaffinity Extraction, Coupled Column HPLC – Ion Spray Mass Spectrometry – Tandem Mass Spectrometry', *Anal. Chem.*, **68**(23), 4228–4236 (1996).
170. M.A. Kelly, H. Liang, I.I. Sytwu, I. Vlattas, N.L. Lyons, B.R. Bowen, L.P. Wennogle, 'Characterization of SH2–ligand Interactions via Affinity Selection with Mass Spectrometric Detection', *Biochemistry*, **35**(36), 11 747–11 755 (1996).
171. Y.H. Chu, D.P. Kirby, B.L. Karger, 'Free Solution Identification of Candidate Peptides from Combinatorial Libraries by Affinity Capillary Electrophoresis/Mass Spectrometry', *J. Am. Chem. Soc.*, **117**(19), 5419–5420 (1995).
172. Y.H. Chu, Y.M. Dunayevskiy, D.P. Kirby, P. Vouros, B.L. Karger, 'Affinity Capillary Electrophoresis–Mass Spectrometry for Screening Combinatorial Libraries', *J. Am. Chem. Soc.*, **118**(33), 7827–7835 (1996).
173. Y.M. Dunayevskiy, Y.V. Lyubarskaya, Y.H. Chu, P. Vouros, B.L. Karger, 'Simultaneous Measurement of Nineteen Binding Constants of Peptides to Vancomycin Using Affinity Capillary Electrophoresis–Mass Spectrometry', *J. Med. Chem.*, **41**(7), 1201–1204 (1998).
174. Y.F. Hsieh, N. Gordon, F. Regnier, N. Afeyan, S.A. Martin, G.J. Vella, 'Multidimensional Chromatography Coupled with Mass Spectrometry for Target-based Screening', *Molecular Diversity*, **2**, 189–196 (1996).
175. S. Kaur, L. McGuire, D. Tang, G. Dollinger, V. Huebner, 'Affinity Selection and Mass Spectrometry-based Strategies to Identify Lead Compounds in Combinatorial Libraries', *J. Protein Chem.*, **16**, 505–511 (1997).
176. Y.M. Dunayevskiy, J.L. Lai, C. Quinn, F. Talley, P. Vorous, 'Mass Spectrometric Identification of Ligands Selected from Combinatorial Libraries using Gel Filtration', *Rapid Commun. Mass Spectrom.*, **11**(11), 1178–1184 (1997).
177. R.B. van Breemen, C.R. Huang, D. Nikolic, C.P. Woodbury, Y.Z. Zhao, D.L. Venton, 'Pulsed Ultrafiltration Mass Spectrometry: A New Method for Screening Combinatorial Libraries', *Anal. Chem.*, **69**, 2159–2164 (1997).
178. Y.Z. Zhao, R.B. van Breemen, D. Nikolic, C.R. Huang, C.P. Woodbury, A. Schilling, D.L. Venton, 'Screening Solution-phase Combinatorial Libraries Using Pulsed Ultrafiltration/Electrospray Mass Spectrometry', *J. Med. Chem.*, **40**(25), 4006–4012 (1997).
179. R. Wieboldt, J. Zweigenbaum, J. Henion, 'Immunoaffinity Ultrafiltration with Ion Spray HPLC/MS for Screening Small-molecule Libraries', *Anal. Chem.*, **69**(9), 1683–1691 (1997).
180. X. Cheng, R. Chen, J.E. Bruce, B.L. Schwartz, G.A. Anderson, S.A. Hofstadler, D.C. Gale, R.D. Smith, J. Gao, G.B. Sigal, 'Using Electrospray Ionization FT-ICR Mass Spectrometry to Study Competitive Binding of Inhibitors to Carbonic Anhydrase', *J. Am. Chem. Soc.*, **117**(34), 8859–8860 (1995).
181. J.E. Bruce, G.A. Anderson, R. Chen, X. Cheng, D.C. Gale, S.A. Hofstadler, B.L. Schwartz, R.D. Smith, 'Bioaffinity Characterization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(8), 644–650 (1995).
182. J. Gao, X. Cheng, R. Chen, G.B. Sigal, J.E. Bruce, B.L. Schwartz, S.A. Hofstadler, G.A. Anderson, R.A. Smith, G.M. Whitesides, 'Screening Derivatized Peptide Libraries for Tight Binding Inhibitors to Carbonic Anhydrase II by Electrospray Ionization–Mass Spectrometry', *J. Med. Chem.*, **39**(10), 1949–1955 (1996).
183. M. Wigger, J.P. Nawrocki, C.H. Watson, J.R. Eyler, S.A. Benner, 'Assessing Enzyme Substrate Specificity Using Combinatorial Libraries and Electrospray Ionization–Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 1749–1752 (1997).

184. J.F. Haw, T.E. Glass, H.C. Dorn, 'Conditions for Quantitative Flow FT-^{1}H NMR Measurements Under Repetitive Pulse Conditions', *J. Mag. Reson.*, **49**, 22–31 (1982).
185. C. Szantay, Jr, 'High-field NMR Spectroscopy as an Analytical Tool for Quantitative Determinations: Pitfalls, Limitations, and Possible Solutions', *Trends Anal. Chem.*, **11**(9), 332–344 (1992).
186. L. Griffiths, A.M. Irving, 'Assay by Nuclear Magnetic Resonance Spectroscopy: Quantification Limits', *Analyst*, **123**, 1061–1068 (1998).
187. M. Godejohann, A. Preiss, C. Mugge, 'Quantitative Measurements in Continuous-flow HPLC/NMR', *Anal. Chem.*, **70**, 590–595 (1998).
188. S.T. Eberhart, A. Hatzis, R. Rothchild, 'Quantitative NMR Assay for Aspirin, Phenacetin, and Caffeine Mixtures with 1,3,5-Trioxane as Internal Standard', *J. Pharm. Biomed. Anal.*, **4**(2), 147–154 (1986).
189. B.C. Hamper, S.A. Kolodziej, A.M. Scates, R.G. Smith, E. Cortez, 'Solid Phase Synthesis of β-Peptoids: *N*-Substituted β-Aminopropionic Acid Oligomers', *J. Org. Chem.*, **63**, 708–718 (1998).
190. E.M. Fujinari, L.O. Courthaudon, 'Nitrogen-specific Liquid Chromatography Detector Based on Chemiluminescence. Application to the Analysis of Ammonium Nitrogen in Wastewater', *J. Chromatogr.*, **592**, 209–214 (1992).
191. R. Bizanek, J. Damon, M. Fujinari, E.M. Fujinari, 'Chemiluminescent Nitrogen Detection as a New Technique for Purity Assessment of Synthetic Peptides Separated by Reversed-phase HPLC', *Peptide Res.*, **9**(1), 40–44 (1996).
192. W.L. Fitch, A.K. Szardenings, E.M. Fujinari, 'Chemiluminescent Nitrogen Detection for HPLC: An Important New Tool in Organic Analytical Chemistry', *Tetrahedron Lett.*, **38**(10), 1689–1692 (1997).
193. E.W. Taylor, M.G. Qian, G.D. Dollinger, 'Simultaneous On-line Characterization of Small Organic Molecules Derived from Combinatorial Libraries for Identity, Quantity, and Purity by Reversed-phase HPLC with Chemiluminescent Nitrogen, UV, and Mass Spectrometric Detection', *Anal. Chem.*, **70**, 3339–3347 (1998).
194. T.H. Mourey, L.E. Oppenheimer, 'Principles of Operation of an Evaporative Light-scattering Detector for Liquid Chromatography', *Anal. Chem.*, **56**, 2427–2434 (1984).
195. J.M. Charlesworth, 'Evaporative Analyzer as a Mass Detector for Liquid Chromatography', *Anal. Chem.*, **50**, 1414–1420 (1978).
196. C.E. Kibbey, 'Quantitation of Combinatorial Libraries of Small Organic Molecules by Normal-phase HPLC with Evaporative Light Scattering Detection', *Molecular Diversity*, **1**, 247–258 (1995).

Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures

R.K. Gilpin and L.P. Dudones
Wright State University, Dayton, USA

1 Introduction **1**
1.1 Understanding the Problem or Sample 1
1.2 Important Physical and Chemical Aspects of the Analyte 2
1.3 Selection of the Separation Mode 5
1.4 Establishing the Starting Eluent Conditions 7
1.5 Isocratic versus Gradient Methods 8
2 Sample Preparation **9**
2.1 Analyte and Matrix Considerations 9
2.2 Sample Work-up 9
3 Methods Optimization **10**
3.1 General Trends 10
3.2 Controlling the Chemical Equilibria 10
3.3 Nonideal Column Effects 12
3.4 Use of Secondary Additives to Enhance Performance 12
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

High-performance liquid chromatography (HPLC) is one of several separation techniques that are based on the differential migration of an analyte (solute) through a fixed medium as the result of a driving force, the eluent. In the case of HPLC, the eluent is typically either a binary or ternary mixture of solvents and the stationary phase is most often a porous adsorbent or a chemically bonded phase attached to a porous substrate. The rate of migration of an individual solute and the separation between groups of solutes are governed by their interactions with the stationary phase. Thus, the optimization of an HPLC method involves the selection/adjustment of the physical and chemical parameters which control the rate of migration of the solute in such a fashion that baseline resolution is obtained between it and other co-analytes or interfering compounds within the mixture. In doing this, given a series of closely related compounds, it is important to identify structural differences between solutes and to maximize the interactions with the stationary phase arising from them. This article considers the important physical and chemical aspects of the solute, eluent, and stationary phase in terms of chemical equilibria, solute structure, eluent conditions, operational parameters, and nonideal effects. Likewise, as several approaches may lead to acceptable analytical separations, other performance criteria such as simplicity, reliability, speed, and cost also are discussed in terms of arriving at the optimal method.

1 INTRODUCTION

Since the late 1960s HPLC-based methods have become increasingly popular in the analysis of pharmaceuticals and currently are the most widely used procedures.[1–3] Their growing acceptance has resulted from the inherent versatility and reliability of the methodology, as well as from improvements in hardware performance and ease of automation of the procedures. Likewise, fundamental sample considerations such as thermal instability, low volatility, and matrix complexity have been important contributing factors in their increasing use.

1.1 Understanding the Problem or Sample

In developing appropriate cost-effective HPLC methods, as with analytical procedures in general, it is important to consider the total problem to be solved. This not only includes the development and optimization of a method in terms of analyte and matrix considerations, separation and detection design, and validation, but it also should take into account the end-usage and user skills. Often, these latter aspects may be the overriding factors in determining the final assay design. If the intended application of a new procedure is as a routine high-throughput assay for assuring that a single active ingredient is within acceptable levels during manufacture of the formulated product, then simplicity, reliability, speed, and cost become important factors in selecting the overall methodology.

What is eloquent in the research and development laboratory may be too complex, expensive, and unreliable for the quality assurance setting. Often simple isocratic approaches are more effective than gradient elution procedures and methods that are based on them are easier to carry out by the less trained practitioner. Similarly, ease of use and reliability are two important considerations in choosing a given approach such as the use of reversed phase (RP) conditions over normal phase

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

(NP) conditions, even though in many cases the latter may be more selective in resolving a series of closely related polar solutes or structural isomers. In cases where compounds are easily separated, there may be distinct advantages in more effective utilization of hardware and decreased costs in supplies and labor by using less retentive and shorter columns operated at higher flow rates. Although rapid analysis (RA) is a concept that emerged two decades ago[4] and has been applied to many types of pharmaceutical separations, it works best for assays involving single components or simple combination products because they do not require the higher resolving power of longer columns. In many of these types of procedures only 1000–2000 plates may be needed to produce an effective separation.[5] Columns for RA, which are available from a number of manufactures, are typically 30–50 mm in length and are packed with 2–5 μm porous materials. However, columns packed with nonporous particles have been developed recently as alternative media for carrying out highly efficient fast separations.[6] Rapid analysis liquid chromatography (RALC) has been used to assay a variety of compounds ranging from simple tablet formulations of acetaminophen[7] to combined products and more problematic compounds such as alkaloids,[8] antibiotics,[9] barbiturates,[8] and hormones.[10]

Alternatively, if the problem to be solved involves more complex samples or less routine usage such as that encountered during the discovery and development stages of a drug, then specificity, versatility, and flexibility are the more important factors. Some typical examples of these are encountered in: (1) the development of a reference standard material, (2) evaluating the purity of the drug substance, (3) stability testing of the active ingredient and its various formulations, and (4) measurements of bioavailability and pharmokinetics. In many cases gradient and multilevel elution approaches and combination column assays may not only be more convenient but necessary, in order to resolve a mixture of compounds produced via a complex synthetic route or if the target compound is a natural product or is produced via a fermentation process. Additionally, gradient elution approaches may provide an important advantage in the detection of trace impurities, especially if they vary significantly in their structures and polarity.

Generally, most practising chromatographers understand how to make their hardware work effectively, and they have a reasonable empirical/working knowledge of how to systematically change conditions in order to effect a given separation, but they sometimes either fail to consider or choose to ignore relatively simple fundamental concepts that influence the reliability and ruggedness of their method. Many of the problems associated with assay instability, such as erratic or slow changes in solute retention and poorly formed or varying peak shapes, result from poorly designed procedures that may provide the desired resolution but are not optimized to minimize variability. Thus, this article addresses the topic of optimization, pointing out whenever possible strategies that improve performance, especially when the assay is to be transferred from a research setting to the quality assurance laboratory and used routinely to monitor production runs.

1.2 Important Physical and Chemical Aspects of the Analyte

Besides the general sample/problem considerations noted above, a fundamental understanding of the physical and chemical aspects of the analyte are important in the systematic design of a reliable method. The first question that should always be asked in adopting an HPLC strategy should be ''is a liquid chromatographic approach the best solution?'' If the sample is thermally stable and volatile, then the answer for most analytes is no. For these types of analytes, methods based on gas chromatography are likely better approaches in terms of cost, simplicity, and reliability. However, if volatility and thermal stability are concerns, then liquid chromatography has been and continues to be the preferred approach in a majority of cases.[1–3] A working knowledge of the analyte's solubility in different solvents and its molecular properties in terms of size, structure, and equilibrium characteristics is important in selecting the overall separation and detection approaches, in choosing the column and eluent, and if necessary in fine-tuning the method in terms of the secondary modifiers needed. Likewise, an understanding of the possible contaminants and by-products that may arise during the chemical synthesis or isolation steps and of degradation products that may be formed as the result of drug instability or drug excipient and package interactions are important in validating assay specificity.

In the general design of a separation, given a series of closely related analytes, it is important to identify that portion of the solute which contains the structural differences and to enhance their interactions with the stationary phase. In some cases these are easily recognized; however, with increasing molecular complexity or with structurally dissimilar compounds, the problem becomes more difficult. As the approaches most often used in pharmaceutical analysis generally involve some form of RP separation, an important item of additional information is the protonation and deprotonation characteristics of the analyte. Although there may be a number of reversed-phase liquid chromatography (RPLC) conditions that will produce a satisfactory separation, some are preferable based on assay ruggedness, cost, and ease of use. It cannot be over emphasized that many of the commonly encountered

problems are associated with the solute and the stationary phase in terms of simple equilibrium considerations. Examples of these types of problems are fluctuations in the retention factor, k', and poorly formed peaks when RP separations are carried out at or near a solute's pK_a, and significant peak tailing for solutes which interact strongly with residual silanols. The logical selection of eluent additives, such as buffers, ion-pairing compounds, and masking reagents, requires a general understanding of the equilibria that govern the solute and its interactions with the stationary phase as well as knowledge about the nonideal nature of the stationary phase and residual silanol activity.

Shown in Figure 1 is the RP separation of a series of important pharmaceutical additives carried out using a conventional short (4.6 mm internal diameter × 30 mm) octadecyl column.[8] Figure 2(a) is the same separation using identical eluent conditions on a higher-efficiency wide-bore (7.0 mm internal diameter × 30 mm) RA column. In performing this latter separation, the flow rate of the eluent was adjusted to give an equivalent linear flow velocity to that used in Figure 1. The accompanying chromatograms (Figures 2b and c) illustrate how analysis speed can be enhanced by simple increases in flow rate.

Under the RP conditions used in Figures 1 and 2, the differences in retention of the ester homologs of *p*-hydroxybenzoic acid are governed by the interactions of the alkyl portion of the compound and the immobilized alkyl chains of the bonded octadecyl phase. The relationship between retention and size (i.e. carbon number) of the aliphatic portion of the paraben homologs is logarithmic, and a plot of ln k' against the carbon number is linear with a slope related to the incremental methylene selectivity.[11] The substitution of a series of $-CF_2-$ units in place of methylene units also results in similar incremental group additivity.[12] The predictable nature of retention for homologs is illustrated in Figure 3

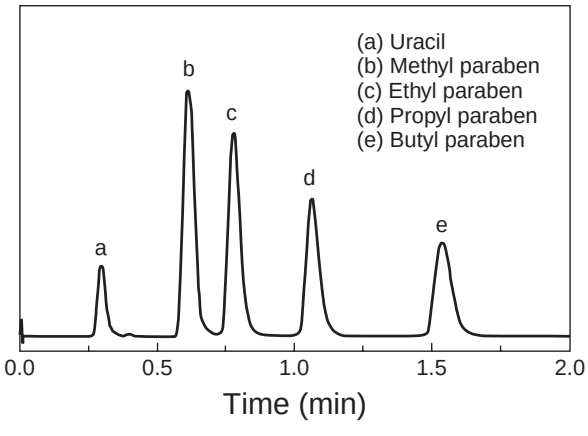

Figure 1 RP separation of paraben homologs under RA conditions; column 4.6 mm × 30 mm octadecyl. (Redrawn from ref. 8 with permission.)

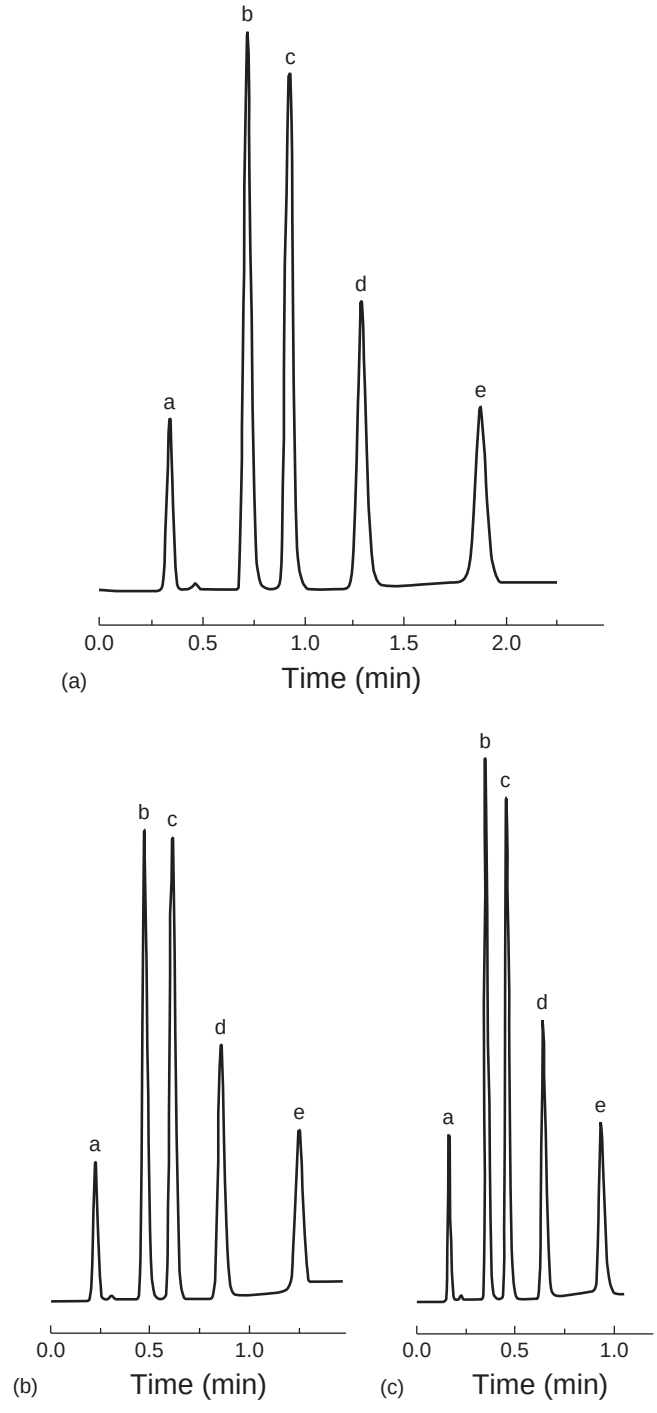

Figure 2 RP separation of paraben homologs under RA conditions; column 7.0 mm × 30 mm octadecyl. Linear flow velocities compared to that used in Figure 1: (a) same; (b) ×1.5; (c) ×2. (Redrawn from ref. 8 with permission.)

within a given column and between columns packed with similar stationary phases. Nonionizable homologs, like the parabens, are often useful as internal standards as they cover a range of predictable retention times under various eluent pH conditions, and they have similar detector

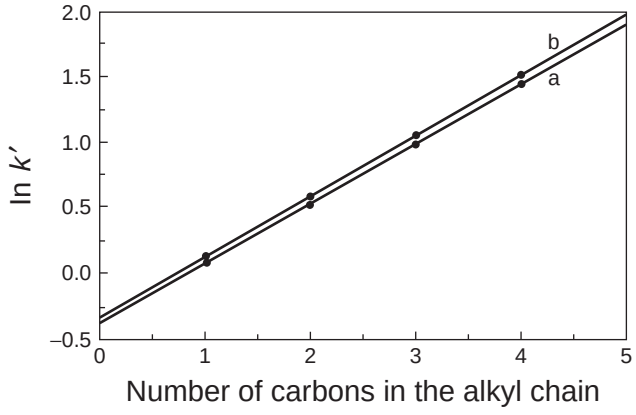

Figure 3 Plot of ln k' versus carbon number: (a) for the chromatogram shown in Figure 1; (b) for the chromatogram shown in Figure 2.

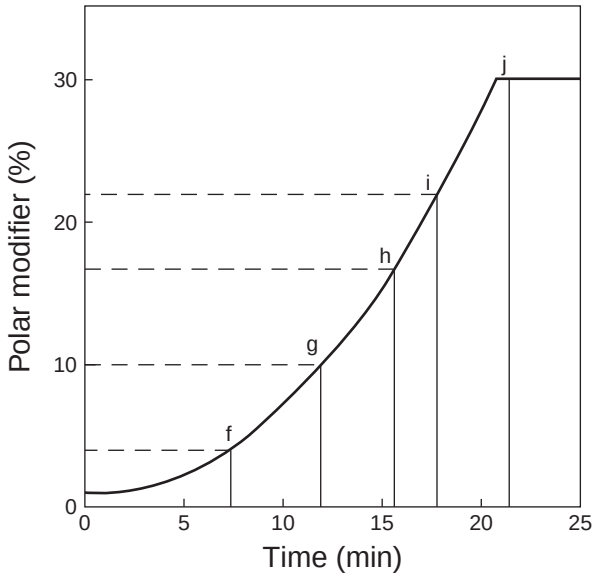

Figure 4 Gradient profile for the separation of the compounds given in Table 1: carrier solvent, *n*-hexane with 0.25% acetic acid; gradient solvent, 2-propanol with 0.25% acetic acid; solid vertical lines, retention times for later-eluting compounds; dashed horizontal lines, amount of polar modifier in eluent when the compounds elute.

responses. Similar linear relationships are obtained for other classes of compounds including ionizable solutes by controlling their deprotonation. This topic is discussed in greater detail in section 3.2.

Structural features, such as the addition of a polar group (e.g. a hydroxyl group), decrease the overall hydrophobicity of the solute and hence decrease retention under RP conditions. These same features result in increased retention under NP conditions. This is illustrated by the data shown in Table 1 for the anti-inflammatory drug tolmetin and related compounds.[13] Comparing the retention data (i.e. elution order) of compounds b, c, and f, which differ only by the substituent at the R_1 position, the greater interaction of the carboxyl group requires a stronger eluent (8.6 min into the gradient profile shown in Figure 4) compared to the corresponding methyl and ethyl esters that elute at the initial mobile phase conditions (1.2 and 1.6 min into the gradient). Although not shown, these same three compounds should be easily resolvable under RP conditions but in the reverse order with the

Table 1 Influence of structure on the gradient elution separation of a nonsteroidal anti-inflammatory drug and related compounds under NP conditions

Compound[a]	R_1	R_2	R_3	Retention time (min)
b	$-CH_2COOC_2H_5$	—H	$CH_3-C_6H_4-CO-$	1.2
c	$-CH_2COOCH_3$	—H	$CH_3-C_6H_4-CO-$	1.4
d	$-CH_2COOH$	—H	—H	3.4
e	—COOH	$CH_3-C_6H_4-CO-$	$-CH_3$	3.4
f (tolmetin)	$-CH_2COOH$	—H	$CH_3-C_6H_4-CO-$	8.6
g	$-CONH_2$	$CH_3-C_6H_4-CO-$	$-CH_3$	11.9
h	$-CH_2COOH$	$CH_3-C_6H_4-CO-$	—H	15.6
i	$-CH_2CONH_2$	—H	$CH_3-C_6H_4-CO-$	17.8
j	$-CH_2CONH_2$	$CH_3-C_6H_4-CO-$	—H	21.8

[a] 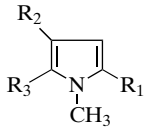

same predictable incremental methylene additivity, as discussed above, once the carboxyl group's deprotonation is suppressed in the case of the parent acid, tolmetin.

The separation summarized in Table 1 is an example of the resolving power of NP methods for structurally similar compounds containing polar substituents. Although NP methods have become less popular because they are more susceptible to retention variability and column fouling, these types of problems can often be minimized. Approaches for doing this are discussed in greater detail in following sections. Under certain conditions, such as the separation of *cis* and *trans* isomers, NP methods may be the best choice.

1.3 Selection of the Separation Mode

The selection of the appropriate separation mode, whether it is based on an RP, NP, ion exchange chromatography (IEC), or size exclusion chromatography (SEC) mechanism, is dependent on the solubility and size of the analyte. In some cases, such as the analysis of common inorganic ions, the choice may be relatively straightforward, the methodology almost turnkey (such as the analysis of cations or anions via ion chromatography (IC)) and the retention predictable.[14] By far the most common application of IC is for measuring inorganic ions and it has been applied to a variety of products including injectable solutions, nutritive liquids, and vitamin formulations.[1–3] Similarly, the application of size exclusion methods have, to a large extent, been used for molecular weight characterizations and determinations of additivities (e.g. polyethylene glycols[15] and microcrystalline cellulose[16]) or to study the properties of a few individual compounds (e.g. the chemical stability of insulin[17]). More recently SEC usage has been increasing with the emerging interest in the development of novel type pharmaceuticals (e.g. polymeric drugs[18]) and biotechnology products from peptides, proteins, and DNA. Some recent examples where it is used include determination of the purity and homogeneity of recombinant hepatitis B virus antigens,[19] the analysis of acidic fibroblast growth factor in viscous formulations,[20] and the purification of plasmid DNA for cancer gene therapy.[21]

Another increasingly important problem is the stereochemical purity of pharmaceuticals, as numerous pharmacologically active agents are chiral and their two enantiomeric forms usually exhibit different physiological properties. Typically, three chromatographic approaches are used to resolve chiral pharmaceuticals:

- the preanalysis conversion of the enantiomers to diastereomers followed by their separation using achiral stationary and mobile phases;[22]
- the direct resolution of the isomeric pairs using achiral stationary phase with a chiral mobile phase additive;[23]
- the direct resolution of the enantiomers on a chiral stationary phase (CSP).[24]

Of these approaches, the latter has become the most popular because of ease and convenience, as well as availability of a host of CSPs. These can be classified into five major groups:[25,26] (I) Pirkle-type (donor–acceptor) phases, (II) inclusion (chiral cavity) materials, (III) helical polymers, (IV) ligand exchangers, and (V) immobilized peptides and proteins. In the latter instance, chiral recognition is a natural property of biological molecules, and when they are properly immobilized, either by physical adsorption or by covalent attachment, some if not most of their indigenous binding affinities are preserved. Of the reported separations being carried out, many employ columns with type I and type II packing. Although all of the above approaches and examples are important, further consideration of these topics is beyond the general scope and emphasis of the current article.

Although most small multifunctional-containing analytes can be separated by the use of either RP or NP conditions, RPLC is typically the route of choice, as noted earlier. This tendency has arisen based on practical concerns such as greater column ruggedness and ease of clean-up of RP packings, and operational considerations such as often longer equilibration times and instability of NP packings, even though in many cases the actual ability of NP approaches to resolve target compounds may be better. On a thermodynamic basis the problems associated with NP methods are the result of larger heats of adsorption (ΔH) that are two to three times greater than the corresponding ΔH values for solutes interacting with bonded alkyl phases. In extreme cases, for highly polar and larger molecules that contain many strongly interacting groups such as amine, amide and hydroxyl functionalities, irreversible or nearly irreversible sorption many occur.

Under NP conditions, two of the most commonly encountered problems are erratic fluctuations in retention and slowly drifting retention. In many instances these undesirable effects are due to short-term variations in laboratory temperature and slow column equilibration rates, respectively. In the first instance erratic changes in k' can be minimized through appropriate temperature control of the column and eluent. In general for most RP applications a $\pm 1\,^{\circ}C$ variation in laboratory temperature (i.e. change in the column temperature) will result in only a small variation in k' (<3%). However, under NP conditions it is not unusual to obtain two to three times this change and under certain conditions even an order of magnitude difference may be observed through significant

polarity changes in the sorbed layer of solvent present on the surface.[27]

The need for temperature control is especially important when NP separations are carried out on polar surfaces such as silica or alumina when using nonpolar carrier eluents (e.g. hexane and small amounts of a highly polar modifier such as an alcohol). These same types of conditions lead to slowly drifting retention, as the packing material may require hundreds of column volumes of the eluent to pass through it to reach equilibrium. This effect is illustrated in Figure 5(a) for a standard 4.6 mm i.d. × 250 mm silica column using *n*-hexane as the carrier eluent and modifying it with varying amounts of three polar alcohols. The curves were generated assuming a simple single-site titration-type adsorption model, an eluent flow rate of 1.0 mL min^{-1}, and a column containing 4 g of silica with surface area of 500 m^2 g^{-1} and five silanol groups per 100 A^2 available as adsorption sites. As it is not uncommon to use small quantities (<1%) of an alcohol modifier to obtain useful k' values, several hours may be needed for the column to reach equilibrium. The horizontal line in Figure 5(a) shows the time needed for the column to reach equilibrium using 0.5 vol% methanol, ethanol, and propanol (2.3, 3.2, and 4.2 h, respectively). Figure 5(b) illustrates this same effect as a function of three different column sizes. An even worse scenario in terms of long equilibration times occurs when hexane or another nonpolar solvent is used alone, due to the presence of trace amounts of dissolved water. This arises because a completely dry silica surface is significantly more retentive of polar compounds than a surface that is fully hydrated. The equilibration time problem often can be minimized if an adequate separation can be developed using an eluent containing larger quantities of the polar modifier and by more careful control of temperature and surface hydration.

As many pharmaceutically active compounds are water soluble and are relatively small to intermediate in size, in a majority of cases they can be resolved more conveniently using RP methodology. By far, the more common approaches employ alkyl-bonded phases in combination with binary hydro-organic eluents. Further, as a majority of the pharmaceutically active compounds contain ionizable groups, such as carboxyl, amine, and amide functionalities, it is necessary to control the pH of the eluent. In the simplest case, where only a single group or at least similar groups are involved, it is advisable to carry out the separation where the solute exists predominately in its uncharged state. This topic is considered in more detail in section 3.2. Generally, this simple strategy works better for acidic compounds than for bases because of equilibrium considerations and pH limitations of the stationary phase being used to carry out the separation. A working eluent pH range for most silica-based materials is 3–8.5. For short periods of time or when using some of the sterically stabilized bonded phases, this range may be exceeded. However, in doing this, it should be recognized that once phase-loss begins, it proceeds at an increasing

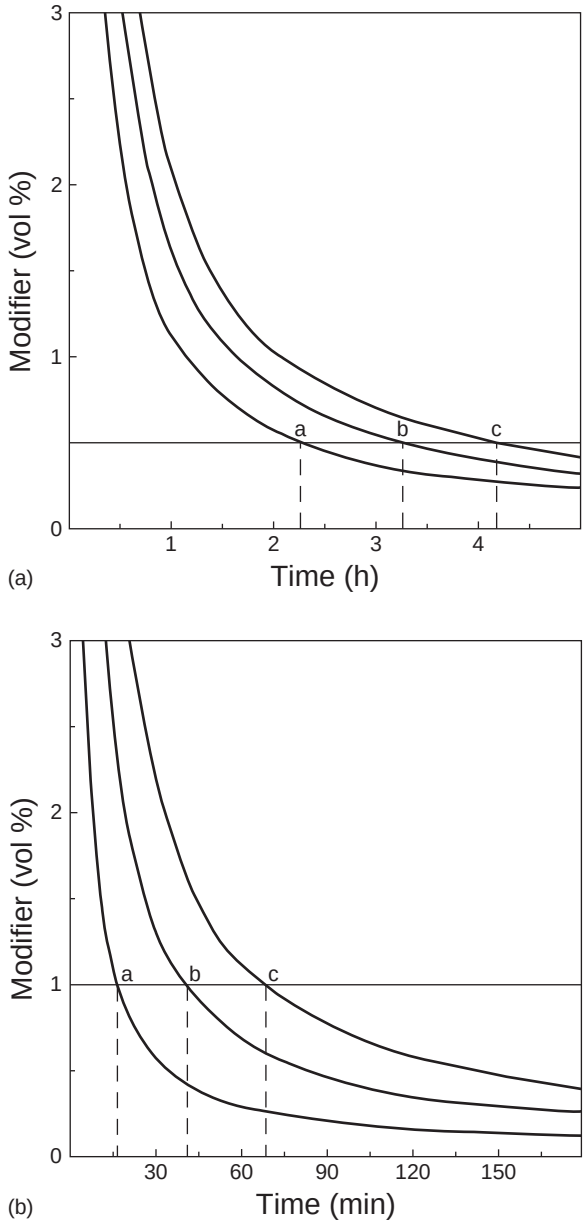

Figure 5 (a) Predicted equilibration rates under NP conditions for a 4.6 mm inside diameter (i.d.) × 250 mm column (i.e. containing 4 g silica with a surface area of 500 m^2 g^{-1} and five silanol groups per square nanometer). The eluent is *n*-hexane with varying amounts of (curve a) methanol, (curve b) ethanol, and (curve c) propanol as the polar modifier. (b) Predicted equilibration rates under NP conditions using *n*-hexane with 1% methanol as the modifier on columns packed with the same material described in Figure 5(a) but packed into columns with dimensions of (curve a) 2.3 mm i.d. × 250 mm, (curve b) 4.6 mm i.d. × 150 mm, and (curve c) 4.6 mm i.d. × 250 mm.

rate as the underlying substrate undergoes hydrolysis. Initially phase-loss results in slowly decreasing retention and increasing peak asymmetry and, subsequently, it can lead to complete breakdown of the underlying silica matrix. The manifestation of this is column voiding and erratic peak shapes. Refilling the column with additional packing material may provide a short-term solution to some of the more severe peak-shape problems, but it will never return the column to its original level of performance as column breakdown proceeds by phase-loss and chemical and physical degradation of the porous matrix.

In cases where the analyte or analytes are very polar, their interactions with simple alkyl phases under RP conditions may not lead to sufficient retention to resolve the analytes. Under these conditions, alternate more polar phases, such as cyanoalkyl and aminoalkyl materials, may be useful. A good example of this strategy is the separation of simple carbohydrates using amine-type phases. Alternatively, very good separations of polar solutes sometimes can be obtained on unmodified materials, such as silica, using NP chromatography.

1.4 Establishing the Starting Eluent Conditions

There are two basic strategies for establishing the starting eluent conditions. One of these involves an isocratic approach and the other a gradient-elution approach. The isocratic approach starts with a solvent with high eluting power, which is systematically decreased by increments in subsequent separations. In the case of RP conditions, the eluent contains a high level of the organic cosolvent. In order of usage and increasing elute strength, the three principle organic cosolvents are methanol, acetonitrile, and tetrahydrofuran. It is not uncommon to start with nearly 100% of the organic component for highly hydrophobic compounds. For NP methodology, the scouting experiment begins with a high level of the polar modifier (say 20–30% 2-propanol) added to the nonpolar carrier eluent (e.g. hexane). In carrying out the isocratic approach, it is important that the eluting strength of the initial mobile phase is high enough so that all compounds to be separated will elute quickly from the column, and that changes in the eluent's composition are made systematically. For RP methods, a 10 or 15% decrease in the organic eluent component provides a rapid means of arriving at an appropriate range of starting conditions and, for NP eluents, halving or approximately halving the polar modifier's concentration provides an adequate change.

A useful quantitative scale (i.e. the eluotropic series) for selecting solvents is based on the work of Snyder and colleagues[28,29] who examined how strongly solvent molecules adsorb to polar substrates. Although this scale was developed initially for alumina, it has been modified for silica. Experimentally it is often helpful, once three or four experiments are carried out, to make a simple plot of $\ln k'$ versus the percentage of organic component in the eluent. An extrapolation of this plot, which in many cases is linear or nearly linear, provides a convenient way of predicting the retention profile of a compound over a broader range of eluent compositions or in fine-tuning the selection of a eluent between compositions tried. This is illustrated in Figure 6 for the separation shown in Figure 7.

When employing the gradient-elution approach as a scouting technique, one starts with a mobile phase of low eluting power and programs its strength to increase with time. Often a 10–15-min profile is a good scouting time.

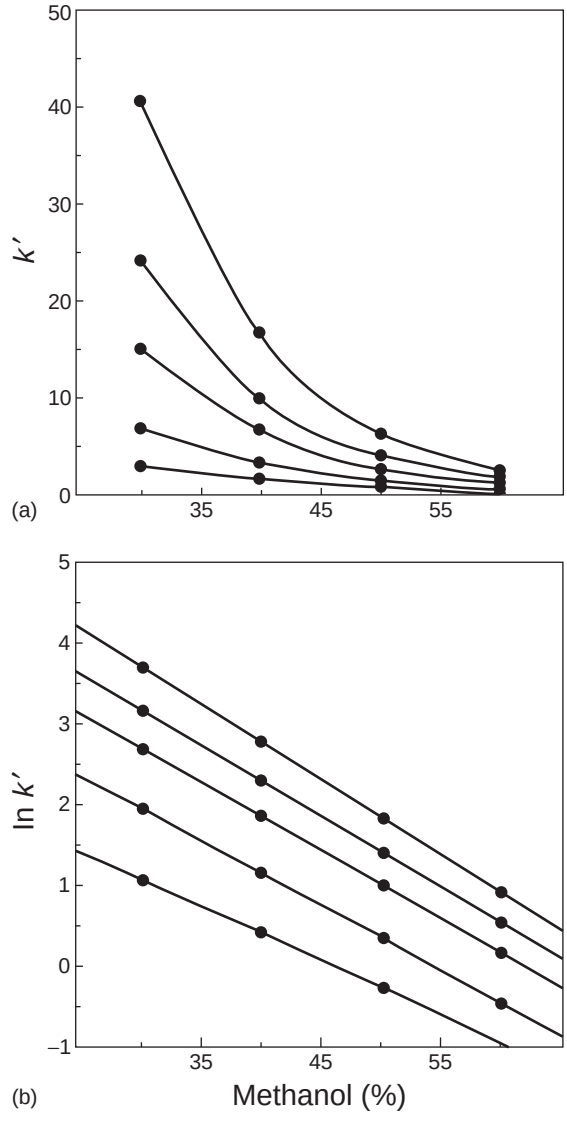

Figure 6 Retention data for the RP separation shown in Figure 7: (a) shows changes in k' versus percentage methanol in the eluent; (b) shows changes in $\ln k'$ versus percentage methanol in the eluent.

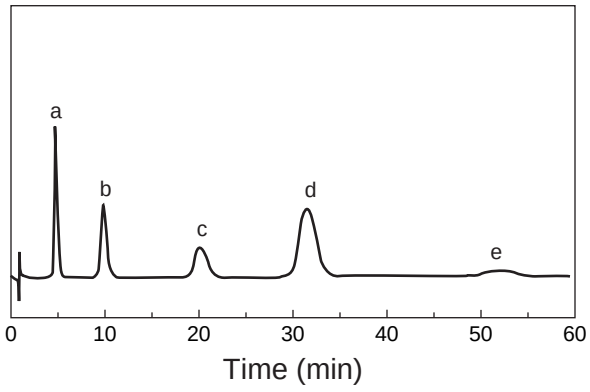

Figure 7 RP separation using methanol–water as the eluent. The chromatogram was obtained with a 30 : 70 mixture. Changes in retention (i.e. k' and $\ln k'$) versus eluent composition are given in Figure 6.

From the gradient profile it is possible to predict starting conditions. This is illustrated by the gradient profile shown in Figure 4 and the respective data summarized in Tables 1 and 2 for the NP separation of tolmetin.[13] The predicted starting conditions from an extrapolation of the gradient at the time a particular solute elutes from the column is listed in the last column of Table 2. The published isocratic conditions are summarized in column two. The predicted and actual conditions are in good agreement. Although not illustrated, this same strategy also is applicable for the selection of starting RP isocratic conditions from a gradient profile that is run from high aqueous conditions to high organic conditions.

In using either the isocratic or gradient approach to predict an initial set of mobile phase conditions, the practicing chromatographer is faced with the following question: ''How does one know if all compounds elute from the column?'' Given a known set of analytes to be resolved, this is easily evaluated. However, given a truly unknown sample or an analyte in a complex matrix, this question is more difficult to answer. Unknown/ghost peaks or slowly drifting baselines often reflect the elution of highly retained compounds. It is often advisable to use thin-layer chromatography (TLC) as a quick screening tool. Under most conditions, if a compound shows some level of migration (even if it is only a very small amount) under a given set of TLC development conditions, then it will be eluted from an equivalent HPLC stationary phase using a similar mobile phase.

1.5 Isocratic versus Gradient Methods

Although it is not always possible to develop isocratic methods for every assay, they tend to be more reliable and less expensive and, as exemplified by the number appearing in the literature yearly compared to the number

Table 2 Reported and predicted level of the polar modifier (isopropanol) needed for the isocratic separation of the compounds in Table 1 under NP conditions

Compound[a]	Reported amount of polar modifier (%)	Retention time (min)	Predicted amount of polar modifier (%)
b	1	1.15	1
c	1	1.40	1
d	3	2.10	2
e	3	1.80	2
f (tolmetin)	10	1.60	4
g	10	2.70	10
h	20	2.60	17
i	20	3.40	22
j	30	5.00	30

[a]

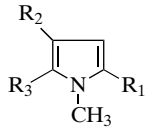

of gradient procedures, they are the first choice in simple-component assays. Additionally, in many cases they are carried out using relatively simple hardware, requiring the operator to blend the eluent off-line. Because of the popularity of the RP methodology, this operation most often involves the preparation of binary mixtures of either methanol and water or acetonitrile and water. For the trained chemist this is a very simple procedure of measuring a known amount of each component and carefully combining them to arrive at the correct eluent composition. Likewise, to the trained chemist, volume changes due to mixing and preferential enrichment of one of the two components are well-known concepts if one component is added directly to the other in the same volumetric vessel. However, it is not uncommon for individuals who are less familiar with these concepts, first to add a given amount of one of the components to a graduated cylinder followed by the direct addition of the second component. This latter practice should be avoided as it can lead to mixing errors and measurable variability in the chromatographic separation. This is illustrated in Figure 8 for the separation shown in Figure 7 and the elution relationships in Figure 6. A 1-mL mixing error, shown by the dashed lines intersecting at points b, results in a very significant change in k'.

In cases where gradient methodology is necessary or at least more convenient, an added consideration in the optimization is selection of an appropriate profile in terms of linear, concave, or convex addition of the stronger

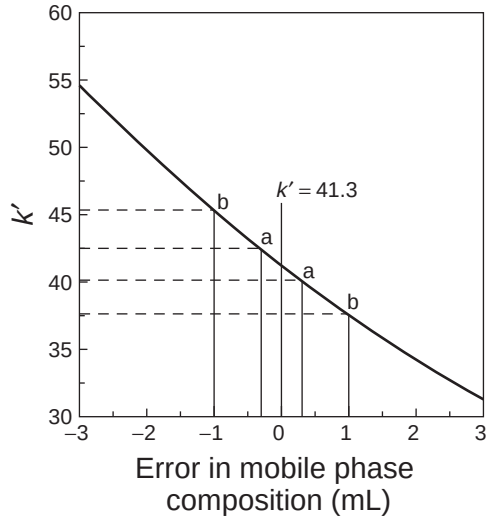

Figure 8 Change in k' for peak e in Figure 7 as a function of mixing error in the preparation of a 30 : 70 methanol–water eluent.

eluent component. Carrying out the initial separation with a linear gradient profile is often a good starting point. Based on this the gradient can be optimized to enhance the resolution of compounds in the initial part of the chromatogram by employing concave addition of stronger eluent components. Conversely, if the later peaks need additional resolution and the early peaks are well separated a convex gradient is used.

2 SAMPLE PREPARATION

2.1 Analyte and Matrix Considerations

One of the most challenging steps in any analysis, including those developed for assaying pharmaceutical products, is often conversion of the target compound into a form suitable for carrying out the chromatographic separation. In many pharmaceutical products, it is not unusual for the analyte to be present in combination with a variety of coadditives and, in the case of biological samples, at differing concentrations depending on source and disposition. This, coupled with physical constraints such as low volatility and thermal instability, has made HPLC methods the most widely used approaches for the quantization of pharmaceuticals in complex formulations and biologicals such as plasma, urine, and bile. Likewise for these same reasons, RP chromatography is especially well suited for these types of samples due to its ability to separate the analyte of interest from a wide range of contaminants. Nevertheless, sample pretreatment steps are usually necessary in order to convert the analyte into a form where the active ingredient is present at suitable concentrations and/or is free of contaminants and particulate matter that can interfere with the separation or damage the column and hardware. Even in cases where the separation can be carried out almost directly, (i.e. for formulations that can be chromatographed after a simple dissolution or dissolution/centrifugation steps) more extensive sample clean-up may be advisable in order to extend column lifetime. Ideally, the pretreatment steps should be rapid and inexpensive. Alternatively, in certain instances it may be more cost-effective to use a precolumn or to sacrifice column lifetime for speed and convenience.

2.2 Sample Work-up

In selecting the best set of pretreatment conditions, the analyst must first consider the physical state of the sample and its complexity. Clear liquids and capsules (i.e. analytes contained in a matrix that is free of particulate matter, dyes, and other coadditives) are generally the simplest types to analyze as either an aliquot of the liquid or a known amount of the powder can be diluted with the mobile phase or a compatible solvent containing an internal standard. Single-dosage forms such as capsules and tablets are analyzed either directly (i.e. to evaluate tablet-to-tablet uniformity) or after they are ground into a fine powder in order to obtain a homogeneous and representative sampling (i.e. composite assays). Many times the resulting formulations are dissolved in a suitable solvent (i.e. a solvent where the analyte is readily soluble and which is miscible with the mobile phase) and any suspended materials removed by centrifugation or filtration. Subsequently, a known aliquot of the supernatant or filtrate is diluted using a solvent containing the internal standard. Selection of an appropriate internal standard generally involves finding a compound that is similar in structure to the analyte and one that elutes near it in the chromatogram. As noted above, for RP methods a methyl homolog is often a good choice, if available, because it will have similar detection properties to the analyte and its retention will be predictable. If the homolog selected contains a larger number of carbon atoms than the analyte, it will elute later in the chromatogram. Conversely, if it has fewer carbon atoms, it will elute earlier. Likewise, a structurally identical or similar compound containing a hydroxyl group is also a good candidate for an internal standard and should elute more quickly than the analyte under RP conditions.

More complex samples typically require greater preparation, such as the removal of particulate matter by centrifugation or filtration or matrix proteins from biological media via precipitation with acetonitrile, ethanol, or trichloroacetic acid. These pretreatment steps are usually followed by some type of extraction in order to further isolate the analyte from soluble formulation coadditives

or biologicals that will either interfere with the assay or foul the column. Traditionally, liquid–liquid extraction has been the method employed most often, and for many applications this continues to be a useful approach, especially for more concentrated samples and when there are distinct differences in the solubility and equilibrium properties of the analyte from those of the contaminants. The disadvantages of liquid–liquid extraction procedures are that they are time-consuming and utilize relatively large volumes of solvents which can be expensive to purchase. Similarly, disposal of the solvent can add considerable cost to the assay and, for more dilute analytes, additional concentration steps (e.g. evaporation) often are required prior to injection of the analyte.

Alternatives to liquid–liquid extraction include adsorption and partitioning methods to concentrate the analyte from the matrix. Partitioning methods, particularly modern solid-phase extraction (SPE), have become popular and increasing in use because they remove particulate matter with concomitant sample concentration, can be carried out on small sample volumes, and are applicable to a wide variety of samples in terms of their physical and equilibrium properties. These are attractive features, especially for biological fluids that are limited to small volumes. A further advantage of SPE is that it typically can be carried out rapidly using minimal volumes of the elution solvents. Further discussions on this topic may be found elsewhere.[30,31]

The SPE technique is performed using small, disposable cartridges filled with 100–1000 mg of inexpensive sorbents that are available in a wide range of polarities and separation modes, analogous to HPLC phases. NP (silica and alumina), bonded NP (diol, cyano, and amino), RP (phenyl, octyl, octadecyl), cation (carboxylic acid, sulfonic acid, aromatic sulfonic acid), and anion (primary and secondary amino, quaternary amine) exchange, and size exclusion cartridges are commercially available from a number of vendors who also provide technical support.

Choosing an SPE sorbent is analogous to selection of an appropriate HPLC packing and depends on the solubility and equilibrium characteristics of the analyte as well as those of the sample matrix. Materials for SPE are chosen such that the analyte(s) of interest is either retained while the impurities pass through the cartridge or, conversely, the analyte(s) passes through the cartridge and the impurities are retained. The first approach is often preferred, because it can be used to purify and concentrate the analyte in one step. In the sample-retained mode, when the analyte is more hydrophilic than the sample matrix, NP-type sorbents are logical choices. Under this set of conditions, highly polar analytes are retained while less polar impurities can be made to pass through the cartridge. However, when the analyte is more hydrophobic than the sample matrix RP sorbents will selectively retain it with high efficiency and unwanted polar interferences can be made to elute. The reverse strategy is applicable when operating in the analyte-nonretained mode.

3 METHODS OPTIMIZATION

3.1 General Trends

Since the late 1970s an important trend in the area of pharmaceutical development has been the ever increasing use of HPLC, especially RPLC.[1–3] Thousands of compounds are analyzed daily by RPLC and new agents are continually under development which will use RPLC because of its ruggedness compared to other separation modes such as normal-phase liquid chromatography (NPLC), the solubility of many drugs in hydro-organic eluents, and the selection and optimization of conditions are easier. A close inspection of the recent pharmaceutical literature[1–3] reveals that by far the most typical set of RP conditions used employ either 150- or 250-mm columns packed with octadecyl or octyl bonded phases. These are used in combination with an isocratic mobile phase prepared from either water–methanol or water–acetonitrile and with the addition of secondary compounds to control solute ionization and to correct for nonideal surface effects.

3.2 Controlling the Chemical Equilibria

The most often used eluent additives are simple buffers, added to control the protonation and deprotonation of ionizable compounds and hence their retention properties. This is easily illustrated by the separation of a series of simple fatty acids with the general structure RCOOH. As these homologs contain the same functional group but again vary only in terms of the length of the alkyl chain, one would want to enhance (maximize) the interaction of the alkyl chain with an alkyl bonded phase by controlling the solute's dissociation. Experimentally, this is accomplished by adjusting the pH of the eluent to about 1–1.2 pH units below the pK_a of the carboxyl group which is approximately 4.7. When the eluent is buffered to 3.5–3.7, the solute exists predominately in its protonated, uncharged, form and its retention is governed by strong interactions between the nonpolar alkyl portion of the molecule and the immobilized alkyl chains of the bonded phase. There are many examples of pharmaceutically important agents which can be conveniently separated via relatively simple RP methodology by controlling the eluent's pH. These range from common products such as aspirin and ibuprofen to related analgesics or anti-inflammatories such as fenbufen, flurbiprofen, ibufeac, ketoprofen, and the anticonvulsant valproic acid.

Figure 9(a) illustrates the relationship between retention and pH for simple monoprotic organic acids. In the deprotonated form ($RCOO^-$), the solute elutes quickly from the column, whereas in the protonated form (RCOOH), it is retained to an increasing degree based on the aliphatic portion of the molecule. The mirror image of this behavior is observed for basic solutes.[32] In the fully protonated form, the relationship between retention and size (i.e. the carbon number) of the aliphatic portion of the homolog is logarithmic. A plot of ln k' versus carbon number is linear with a slope related to incremental methylene selectivity (Figure 9b).[11] Similar relationships also are obtained for other classes of solutes and the retention for a given series is often easily predictable from a simple plot of ln k' against the carbon number. The substitution of a series of nonpolar groups in place of methylene units also results in similar incremental additivity.[12] Likewise, an increase in the hydrophobic character of a series of homologs, such as through the addition of an aromatic ring, results in increased retention as illustrated by the parallel relationship that is offset to higher retention values for the alkyl phenols (Figure 9b, line a) compared to the fatty acids (Figure 9b, line b).

A simple rule is that by decreasing the charge (i.e. by dissociation) of the solute under common RP conditions one increases the interaction of the solute with the stationary phase, hence increasing retention and the possibility of resolving like compounds. As such, the preferred set of conditions for the separation of the compounds shown in Figure 9, as well as with other simple cases where only a single monoprotic group is present (such as the acid drugs shown in Scheme 1) or at least similar groups are involved, is in the plateau region of the curves. However, once the eluent pH is adjusted to this region, further adjustments to lower pH will have little effect on the quality of the separation, but will lead to increased chances of phase instability and decreases in performance of the column. Unfortunately in some cases, because of column stability considerations, the pH of the eluent cannot be adjusted low

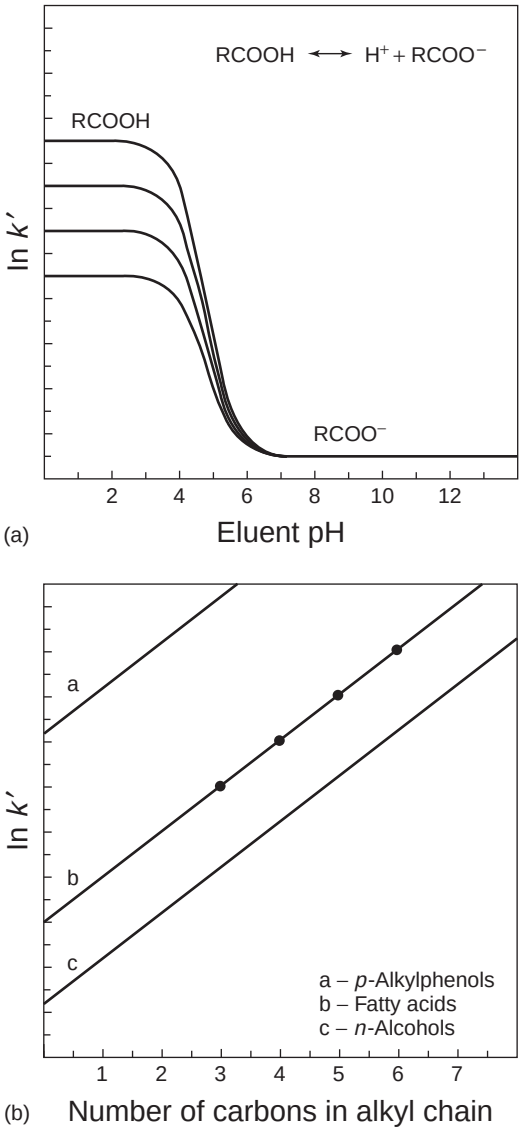

Figure 9 Retention properties of simple monoprotic organic acids: (a) influence of eluent pH; (b) chain-length effect for different homologs.

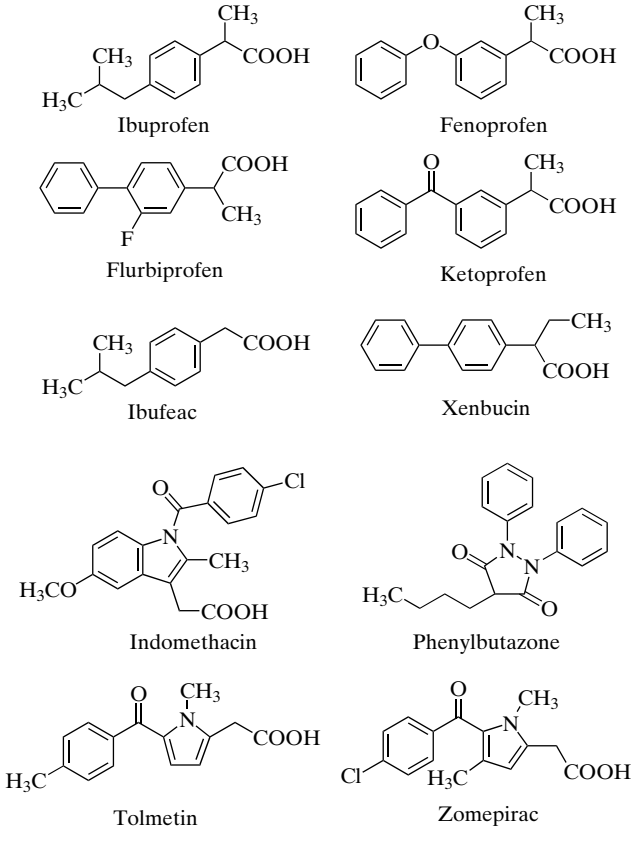

Scheme 1

enough or high enough to fully protonate or deprotonate the solute. A working range for most silica-based bonded phases, is 3.0–8.5. In cases where the solute's equilibrium cannot be adequately controlled within this range, an ion-pairing reagent can be added to the eluent to control the solute's ionization. The most commonly used of these are alkyl (often C_5 to C_8) sulfates and sulfonates, and tetraalkyl (up to C_4) ammonium salts. Although the mechanisms involved in ion-pair separations are complex, its use on an empirical basis is relatively straightforward. The retention of the ionized solutes increases with increasing size of the alkyl portion of the ion-pair salt and conversely decreases with decreasing size.

As the sample becomes more complex in terms of the presence of multiple functionality and other structural features, it is often not possible to predict in advance many of the subtleties controlling retention. Considering some of the related nonsteroidal analgesic/anti-inflammatories appearing in Scheme 1, it is possible to predict that ibufeac will elute before ibuproben, as the molecules are structurally very similar except that ibuprofen contains an additional methyl carbon. However, it is much more difficult to predict how long indomethacin will be retained compared to tolmetin or the elution order of ketoprofen, fenoprofen, and flurbiprofen. Nevertheless, for the latter set of compounds, because they are structurally similar, it is reasonable to suggest that an assay developed for ketoprofen might be a useful starting point for fenoprofen. Additionally, it is reasonable to suggest that the pH of the eluent for separating the first six compounds in Scheme 1, as well as the remaining more complex structures should be about 3.5, as the only important structural feature requiring eluent buffering is the carboxyl group.

3.3 Nonideal Column Effects

Irrespective of the synthetic routes used to produce silica-based packings, an important controlling parameter inherent to all of these types of phases is the nature of the substrate surface (i.e. its chemical and structural heterogeneity).[33] Porous silica is an amorphous material with a three-dimensional structure that contains several types of polar silanol groups on its surface.[34,35] The distribution of these groups and morphology determine the extent and heterogeneity of the resulting bonded phases. The number and distribution of silanols change depending on the synthetic route that is employed to produce silica, as well as post-synthesis processing (i.e. thermal treatment and rehydration) to mechanically stabilize the material. Additional information on the physical and chemical properties of silica may be found elsewhere.[36,37]

When silicas are produced, it is not unusual to obtain materials that are equivalent based on an examination of their macroscopic properties, such as total number of silanol groups, surface area, porosity, etc. yet dramatically different in terms of microscopic heterogeneity which influences the nature of bonded phases produced from such materials.[38] To the practicing chromatographer, this problem manifests itself as manufacturer to manufacturer and batch to batch differences in column performance for a given stationary phase. Additionally, this problem is exacerbated for solutes that have polar functional groups that can interact strongly with residual silanols such as amines and heterocyclics. These are not uncommon types of compounds encountered on a daily basis in the pharmaceutical industry.

In order to minimize the residual silanol problems, one of four approaches is generally used, these being postreaction end-capping, preparation of sterically blocking phases, electronic manipulation of the attached surface groups, and the use of mobile phase additives. End-capping involves a resilanization of the column with a small reactive silane such as trimethylchlorosilane following the initial bonding process. The second approach utilizes surface ligands that can sterically block silanol accessability.[39–41] The third approach, electronically manipulated phases, involves imbedding a functional group within the bonded chain's interior that can interact with neighboring residual silanols (e.g. alkylamide phases).[42–46] Each of these three approaches has allowed manufacturers, as well as researchers, to prepare columns with enhanced chromatographic performance. Unfortunately, not all peak-tailing problems have been eliminated by these and, as columns age, the problem of exposed silanol groups increases even for high-performance bonded phases. Many of these unwanted effects can be eliminated or at least minimized through the use of secondary mobile phase additives.

3.4 Use of Secondary Additives to Enhance Performance

Although the performance of commercially available bonded phases has improved dramatically in recent years, there are still many RP applications where residual silanol activity leads to unacceptable chromatograms in terms of severely tailing peaks. Such problems can usually be addressed by the addition of compounds to the eluent that dynamically modify the surface by a competitive sorption mechanism, and hence act to suppress undesirable interactions that can arise between basic solutes and residual silanols. This process is sometimes referred to as silanol masking and the agents used to do this are strongly sorbing compounds (i.e. molecules that contain a polar head group and a nonpolar tail) that do not interfere with detection. The most commonly used compounds to

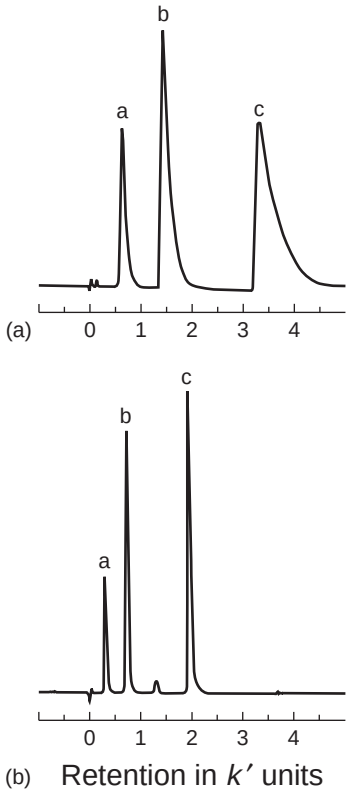

Figure 10 The RP separation of a simple test mixture of (peaks a) aniline, (peaks b) *N*-methylaniline, and (peaks c) *N*,*N*-dimethylaniline with 4.6 mm × 250 mm column packed with octyl-silica, with 60 : 40 v/v% methanol–water as eluent: (a) without masking agent; (b) 5 mM PFPA (*N*-(2-hydroxyethoxyethyl)-perfluoropropylamide) added as a masking agent. (Redrawn from ref. 48 with permission.)

mask silanol activity, and hence improve peak symmetry, are alkylamines;[47] however, in a few cases other compounds, such as perfluororalkyl surfactants, have been employed.[48,49]

The chromatograms appearing in Figure 10 illustrate the dramatic effect a masking agent can have on the separation performance of solutes (e.g. simple aromatic amines) that interact strongly with residual silanol groups. Both separations were carried out on a standard commercial octyl bonded phase using 60 : 40 methanol–water. However, chromatogram (a) was obtained without the addition of a masking agent and chromatogram (b) was obtained with a masking agent present in the eluent.[47]

ABBREVIATIONS AND ACRONYMS

CSP	Chiral Stationary Phase
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
i.d.	Inside Diameter
IEC	Ion Exchange Chromatography
NP	Normal Phase
NPLC	Normal-phase Liquid Chromatography
PFPA	*N*-(2-Hydroxyethoxyethyl)-perfluoropropylamide
RA	Rapid Analysis
RALC	Rapid Analysis Liquid Chromatography
RP	Reversed Phase
RPLC	Reversed-phase Liquid Chromatography
SEC	Size Exclusion Chromatography
SPE	Solid-phase Extraction
TLC	Thin-layer Chromatography

RELATED ARTICLES

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Ion Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography

REFERENCES

1. R.K. Gilpin, L.A. Pachla, 'Pharmaceutical and Related Drugs', *Anal. Chem.*, **67**, 295R–313R (1995).
2. R.K. Gilpin, L.A. Pachla, 'Pharmaceutical and Related Drugs', *Anal. Chem.*, **69**, 145R–163R (1997).
3. R.K. Gilpin, L.A. Pachla, 'Pharmaceutical and Related Drugs', *Anal. Chem.*, **71**, 217R–233R (1999).
4. J.H. Knox, 'Practical Aspects of LC Theory', *J. Chromatogr. Sci.*, **15**, 352–364 (1977).
5. S. Lindsay, *High Performance Liquid Chromatography*, 2nd edition, John Wiley and Sons, London, England, 1992.
6. D.C. Lommem, L.R. Snyder, 'Fast HPLC Separations with Small Nonporous Particles', *LC GC Magazine of Separation Science*, **11**, 222–228 (1993).
7. R.K. Gilpin, M.H. Gaudet, 'Rapid High-performance Liquid Chromatographic Determination of Acetaminophen in Dosage Form Using a Totally Aqueous Mobile Phase', *J. Chromatogr.*, **248**, 160–164 (1982).
8. *Alletch Bulletin 413*, Alltech Associates, Deerfield, IL, 1998.
9. K.M. Kirkland, D.A. McCombs, J.J. Kirkland, 'Rapid High-resolution High-performance Liquid Chromatographic Analysis of Antibiotics', *J. Chromatogr.*, **660**, 327–337 (1994).

10. M.B. Hurley, J.S. Rowarth, 'Comparison of a Fast Liquid Chromatography Column with a Standard Column for the Rapid Isocratic Quantitation of ABA', *J. Chromatogr. Sci.*, **36**, 345–348 (1998).
11. R.K. Gilpin, M. Jaroniec, S. Lin, 'Dependence of Methylene Selectivity on the Composition of Hydro-organic Eluents for Reversed-phase Liquid Chromatographic Systems with Alkyl Bonded Phases', *Chromatographia*, **30**, 393–399 (1990).
12. M. Jaroniec, S. Lin, R.K. Gilpin, 'Characterization of the Surface Composition of Alkyl Bonded Phases Under Reversed-phase Liquid Chromatographic Conditions Using Homologs of Alkanoate and Perfluoroalkanonate Esters as Solute Probes', *Anal. Chem.*, **62**, 2092–2098 (1990).
13. R.K. Gilpin, C.A. Janicki, 'A High-pressure Liquid Chromatographic Method for Surveying the Purity of Tolmetin Sodium Dihydrate', *J. Chromatogr.*, **147**, 501–506 (1978).
14. H. Small, 'Ion Chromatography', in *Modern Analytical Chemistry Series*, ed. D. Hercules, Plenum Press, New York, 1989.
15. J. Vandam, P. Daenens, 'Molecular-weight Identification of Polyethylene Glycols in Pharmaceutical Preparations by Gel-permeation Chromatography', *J. Pharm. Sci.*, **82**, 938–941 (1993).
16. S.S. Cutie, C.G. Smith, 'Determination of Methocel A15-LV Cellulose Ether in Blends with Microcrystalline Cellulose', *J. Chromatogr. A*, **693**, 371–375 (1995).
17. J. Brange, S. Havelund, P. Hougaard, 'Chemical-stability of Insulin: 2) Formation of Higher Molecular-weight Transformation Products During Storage of Pharmaceutical Preparations', *Pharm. Research*, **9**, 727–734 (1992).
18. K. Juliano, W.L. Champion, M.A. Schreiber, J.A. Blackwell, 'Quantitation and Determination of Molecular Weight Distribution of Residual Water Soluble Extractable Polyamines in DMP 504, Polyalkylamine Bile Acid Sequestrant, by Aqueous High Performance Size Exclusion Chromatography', *J. Pharm. Biomed. Anal.*, **16**, 499–506 (1997).
19. L. Costa, G. Moya, F.M. Gonzalez, A. Caballero, V. Falcon, M. Vega, 'Determination of Purity and Molecular Homogeneity of Recombinant Hepatitis B Virus Surface Antigen by Gel Filtration HPLC', *Biotechnol. Apl*, **11**, 86–90 (1994).
20. M.W. Bruner, J. Goldstein, C.R. Middaugh, M.A. Brooks, D.B. Volkin, 'Size Exclusion HPLC Method for the Determination of Acidic Fibroblast Growth Factor in Viscous Formulations', *J. Pharm. Biomed. Anal.*, **15**, 1929–1935 (1997).
21. N.A. Horn, J.A. Meek, G. Budahazi, M. Marquet, 'Cancer Gene-therapy Using Plasmid DNA-purification of DNA for Human Clinical-trials', *Human Gene Therapy*, **6**, 565–573 (1995).
22. W. Linder, 'Chromatographic Chiral Separations', in *Chromatographic Science Series*, eds. M. Zief, L.J. Crane, Marcel Dekker, New York, 1988.
23. G. Szepesi, M. Gazdag, 'Enantiomer Separations and Their Application in Pharmaceutical Analysis Using Chiral Eluents', *J. Pharm. Biomed. Anal.*, **6**, 623–639 (1988).
24. W.H. Pirkle, T.C. Pochapsky, 'Considerations of Chiral Recognition Relevant to the Liquid Chromatography Separation of Enantiomers', *Chem. Rev.*, **89**, 347–362 (1989).
25. I.W. Wainer, R.M. Stiffin, Y.Q. Chu, 'Chiral Separations', in *Proceeding of the Chromatogr. Soc. International Symp. on Chiral Separations*, eds. D. Stevenson, I.D. Wilson, Plenum Press, New York, 1988.
26. I.W. Wainer, 'HPLC Chiral Stationary Phases for the Stereochemical Resolution of Enantiomeric Compounds: The Current State of the Art', in *Drug Stereochemistry Analytical Methods and Pharmacology*, 2nd edition, ed. I.W. Wainer, Marcel Dekker, New York, 1993.
27. R.K. Gilpin, W.R. Sisco, 'Effect of Temperature on Precision of Retention Measurements in Liquid Chromatography', *J. Chromatogr.*, **194**, 285 (1980).
28. L.R. Snyder, 'Principles of Adsorption Chromatography; the Separation of Nonionic Organic Compounds', in *Chromatographic Science Series*, Marcel Dekker, New York, 1968.
29. B.L. Karger, L.R. Snyder, C. Horvath, *An Introduction to Separation Science*, John Wiley and Sons, New York, 1973.
30. M. Zief, S.V. Kakodkar, 'Solid Phase Extraction for Sample Preparation', in *Pharmaceutical and Biomedical Applications of Liquid Chromatography*, eds. C.M. Riley, W.J. Lough, I.W. Wainer, Pergamon, 127–143, 1994.
31. M. Zief, 'Sample Preparation for Pharmaceuticals in a Solid-Phase Extraction', in *Liquid Chromatography in Pharmaceutical Development*, ed. I.W. Wainer, Aster Publishing, Springfield, OR, 133–146, 1985.
32. C. Horvath, W. Melander, 'Liquid Chromatography with Hydrocarbonaceous Bonded Phases; Theory and Practice of Reversed-phase Chromatography', *J. Chromatogr. Sci.*, **15**, 393 (1977).
33. R.K. Gilpin, 'Conformational Changes and Molecular Dynamics of Simple Silica Immobilized Systems', *J. Chromatogr. A*, **656**, 217–229 (1993).
34. L.R. Snyder, J.W. Ward, 'The Surface Structure of Porous Silica', *J. Phys. Chem.*, **70**, 3941–3952 (1966).
35. R.P.W. Scott, 'The Silica Gel Surface and its Interactions with Solvent and Solute in Liquid Chromatography', *Adv. Chromatogr.*, **20**, 167–215 (1982).
36. R.K. Iler, *The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry*, Wiley, New York, 1979.
37. K.K. Unger, *Porous Silica, its Properties and Use as Support in Column Liquid Chromatography*, Elsevier Scientific Publishing, Amsterdam, 1979.

38. R.K. Gilpin, L. Wu, 'Use of the Reordering/Resolvation of Alkyl-modified Silica to Characterize the Microscopic Heterogeneity of Silica Via Liquid Chromatography', *J. Chromatogr.*, **556**, 415–424 (1991).
39. S.S. Yang, R.K. Gilpin, 'Admantyl-modified Silica as a Reversed-phase Liquid Chromatographic Packing', *Anal. Chem.*, **59**, 2750–2752 (1987).
40. S.S. Yang, R.K. Gilpin, 'Reversed-phase Separation of a Standard Mixture of Polycyclic Aromatic Hydrocarbons on an Admantyl Surface', *J. Chromatogr.*, **439**, 414 (1988).
41. S.S. Yang, R.K. Gilpin, 'Analysis of Cough/Cold Products Using an Admantyl Column', *J. Chromatogr. Sci.*, **26**, 416 (1988).
42. P. Kasturi, B. Buszewski, M. Jaroniec, R.K. Gilpin, 'Reordering/Resolvation Studies of Alkylamide Phases', *J. Chromatogr. A*, **659**, 261–265 (1994).
43. B. Buszewski, M. Jaroniec, R.K. Gilpin, 'Influence of Eluent Composition on Retention and Selectivity of Alkylamide Phases Under Reversed-phase Conditions', *J. Chromatogr. A*, **668**, 293–299 (1994).
44. B. Buszewski, P. Kasturi, R.K. Gilpin, M.E. Gangoda, M. Jaroniec, 'Chromatographic and Related Studies of Alkylamide Phases', *Chromatographia*, **39**, 155–161 (1994).
45. B. Buszewski, M. Jaroniec, R.K. Gilpin, 'Studies of Physicochemical and Chromatographic Properties of Mixed Amino-alkylamide Bonded Phases', *J. Chromatogr. A*, **673**, 11–19 (1994).
46. T. Czajkowska, I. Hrabovsky, B. Buszewski, R.K. Gilpin, M. Jaroniec, 'Comparison of the Retention of Organic-Acids on Alkyl and Alkylamide Chemically Bonded Phases', *J. Chromatogr. A*, **691**, 217–224 (1995).
47. R.K. Gilpin, S.S. Yang, G. Werner, 'A Recent Review of Various Secondary Mobile Phase Modifiers Used to Enhance Chromatographic Analysis of Pharmaceuticals', *J. Chromatogr. Sci.*, **26**, 388–400 (1988).
48. P. Varughese, M.E. Gangoda, R.K. Gilpin, 'Applications of Fluorinated Compounds as Phases and Additives in Chromatography and Their Uses in Pharmaceutical Analysis', *J. Chromatogr. Sci.*, **26**, 401–405 (1988).
49. M.E. Gangoda, R.K. Gilpin, 'Comparative Study of Alkyl and Fluoroalkyl *N*-(2-Hydroxy-ethoxyethyl)-Amines as Reversed-phase Liquid Chromatographic Modifiers', *J. Chromatogr.*, **448**, 365–370 (1988).

Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Keith J. Duff
Restek Corporation, Bellefonte, USA

1 Choosing the Separation Technique – Gas Chromatography Versus High-performance Liquid Chromatography 2
1.1 Introduction 2
1.2 Sample and Other Considerations 2
1.3 Literature Comparisons 3
1.4 Ideal Methods Development 4
2 Column Selection 4
2.1 Tubing Choices 4
2.2 Dimensions 4
2.3 Coverage and Film Thickness 6
2.4 Stability and Reproducibility 7
2.5 Phase Type 8
3 Chemistry of Bonded Phases 11
3.1 Gas Chromatography Coatings 11
3.2 Modified Silica Phases 11
3.3 Nonsilica High-performance Liquid Chromatography Phases 12
3.4 Direct Inject Columns 14
4 Modes of Chromatographic Separation 14
5 Phase Selection by Analyte Functional Group Type 16
5.1 Acids 16
5.2 Bases 17
5.3 Acid/Base Mixtures 17
5.4 Zwitterions 17
5.5 Polar Compounds 17
5.6 Halogenated Compounds 17
5.7 Chiral Compounds 18
6 Phase Selection by Analyte Class 21
6.1 Adrenergic 21
6.2 Alkaloids 22
6.3 Amino Acids 23
6.4 Antibacterial/Antifungal 23
6.5 Anti-inflammatory 23
6.6 Antipsychotic 23
6.7 Anthracene Derivatives 23
6.8 Barbiturates 23
6.9 Benzazepines/Benzodiazepines 24
6.10 β-Blockers 24
6.11 Blood Volatiles 24
6.12 Cardiovascular 24
6.13 Catecholamines/Phenols 24
6.14 Cephalosporins/Penicillins 25
6.15 Cold Medicines 25
6.16 Cyclosporins 25
6.17 Drugs of Abuse 25
6.18 Enkephalins/Endorphins 25
6.19 Glycosides 26
6.20 Hormones 26
6.21 Mycotoxins 26
6.22 Nitrosamines 26
6.23 Nucleosides/Nucleotides/Bases 26
6.24 Parabens 26
6.25 Peptides 26
6.26 Pharmaceutical Processing Solvent Residues 27
6.27 Phenylalkylamines 27
6.28 Porphyrins 27
6.29 Procainamides 27
6.30 Profens 28
6.31 Proteins 28
6.32 Steroids 28
6.33 Sulfur Drugs 28
6.34 Taxol and Related Compounds 29
6.35 Terpines 29
6.36 Tetracyclines 29
6.37 Tricyclic Antidepressants 29
7 Association of Official Analytical Chemists and United States Pharmacopeia Methods 29
Acknowledgments 31
Abbreviations and Acronyms 31
Related Articles 31
References 32

Within the realm of pharmaceutical drug analyses, two techniques currently dominate the field. High-performance liquid chromatography (HPLC) is at the forefront, followed somewhat distantly by gas chromatography (GC). HPLC is more prevalent than GC owing mainly to the heat-labile nature of the majority of drug substances. Still GC continues to play a very active role. In comparison to HPLC and GC, combined, the frequency of use of other techniques is quite small. Perhaps the major reason for this is that chromatography not only involves qualitative and quantitative detection of targeted species, but also provides a medium for isolating these analytes from potential interfering impurities prior to measurement. The proficiency of this combination is unparalleled by any other blending of means to clean up a sample and detect

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the target compound. These more prominent techniques are often complementary, although each has its own strengths and limitations. General guidelines are given to assist the chromatographer to elect HPLC and/or GC for a particular separation. Considerations for choosing hardware type and sizes, solid support and bonded phase are presented. Particular emphasis is placed on the latest, which is of primary importance for obtaining rugged methods. The reader is directed towards optimum phase selection based on the compound of interest's functional group type, or by its drug activity classification (e.g. barbiturate, cardiovascular, etc.). Assistance in developing new methods frequently arises when starting with an existing published method for a compound with similar structure. This article provides 514 references for more in-depth separation information. In addition, background knowledge on bonded-phase chemistry, chromatographic modes of separation, approaches to chromatographic chiral drug discrimination and processes for developing rugged methods is contributed.

1 CHOOSING THE SEPARATION TECHNIQUE – GAS CHROMATOGRAPHY VERSUS HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

1.1 Introduction

GC and HPLC are generally considered to be complementary techniques. Choosing one over the other is often dictated more by equipment availability[1] and familiarity than by which technique is better suited for a particular application. However, guidelines exist for the chromatographer to make educated choices for the optimum system, when both means are available. Traditionally, GC chromatograms have provided an order of magnitude more theoretical plates than HPLC, although newer HPLC columns packed with smaller particles approach GC performance in that regard. Higher theoretical plates enable easier separations of more complex samples that contain numerous constituents. A chromatogram containing about 15 peaks would be considered complex for most HPLC separations, but is more commonplace for GC chromatograms. HPLC is capable of *separating* a wider range of analytes than GC. Earlier, GC was capable of *detecting* a wider range of analytes with better sensitivity than HPLC. However, this advantage has diminished with the advent of light scattering detectors, and ''hyphenated technologies'' such as liquid chromatography/mass spectrometry (LC/MS)[2–4] and liquid chromatography/nuclear magnetic resonance (LC/NMR).[5] In fact, Gelpi[6] postulated that LC/MS is replacing gas chromatography/mass spectrometry (GC/MS) analyses for drug metabolism, therapeutic drug monitoring (TDM) and pharmacology, since the former is applicable to highly polar and larger molecular weight compounds without complicated derivatization and handling procedures.

There is evidence that HPLC is currently the more popular technique and is growing faster than GC. In a 1995 survey,[6] total HPLC instrument sales ($840.00 million) were higher than GC instrument sales ($575.00 million). A 1997 survey of analytical instrument end-users[7] reported that HPLC (1600 units) was the main technique used, followed by GC (1300 units). Other techniques, such as UV/VIS (ultraviolet/visible), electrophoresis, FTIR (Fourier transform infrared) and AA (atomic absorption) were less utilized than HPLC or GC. Around 50% of HPLC owners surveyed expected increased HPLC usage over the next five years.[8] In contrast, about half of the GC users considered their technique to be mature with few new applications emerging. Within the pharmaceutical industry, HPLC was the most commonly used instrumentation in the laboratory (35%) followed by GC (15%).[9] Pharmaceutical QA/QC (quality assurance/quality control) laboratories employed HPLC (49%) in higher percentage than GC (19%), and HPLC was expected to demonstrate a significantly higher growth rate than GC over the next five years.[10]

In reflection of the above surveys, this article focuses proportionally more on HPLC than on GC pharmaceutical separations. Nevertheless, it appears that GC will remain an important tool in pharmaceutical analyses for some time.

It is important to note that the references provided in this work are by no means an exhaustive coverage of the available literature for the given topics. Rather, they serve to exemplify distinguished points and provide starting conditions for the analyst in methods development of unique compounds.

1.2 Sample and Other Considerations

Perhaps the most important points in choosing between GC and HPLC methodology are the volatility and thermal stability of the sample. Nonvolatile samples are not suitable for GC analyses unless they are derivatized to a more volatile form. Gouw and Jentoft[11] estimated that only 15% of known compounds can be volatilized without thermal degradation. There are no such restrictions with HPLC, which is at least part of the reason for the increased popularity of HPLC over GC in pharmaceutical separations. Sample volatility is dependent on the composition and structure of the compound and its molecular weight. As a general rule of thumb, pharmaceutical chemicals above about 200–300 molecular weight and those containing polar functional groups are better suited for HPLC analyses.

Derivatization for GC analysis is accomplished by reducing a nonvolatile sample's polarity. Electrophilic and other derivatization reagents are commercially available from a variety of chromatography companies and are usually shipped in sealed ampules to prevent atmospheric moisture from deactivating the reagent. Some of the more popular reagents include silanes, acylation reagents and alkylation reagents. They react with compounds that contain exchangeable or active protons (protic samples), such as alcohols, thiols, carboxylic acids and amines, to form the corresponding ethers, thioethers, esters and amides, or with aldehydes and ketones to form oximes. In addition to increasing sample volatility, many derivatized samples also exhibit higher detection sensitivity, selectivity and chemical stability than the parent sample. Disadvantages in derivatized sample methodology include the added preparative time and expense, and the need to determine the percent yield of the derivatized product for quantitative measurements. Reproducible derivatizations may be particularly difficult with samples that contain more than one reactive site. By-products and side products may also co-elute and interfere with chromatographic peaks of interest.

Although not as prevalent as in GC, HPLC derivatizations are also occasionally performed. For strategies and synthetic methods in pharmaceutical HPLC sample derivatizations, see Ahwja[12] and Danielson et al.[13]

Although the analyte itself may be volatile, the sample matrix may preclude use of GC analyses. Nonvolatile components can build up within the GC column, affecting subsequent chromatographic performance, or even plug the column. Prepurification by solid-phase extraction (SPE), membrane filtration, flash chromatography and other procedures are viable options, but add to analysis time and cost.

Determining residual solvents in drug substances is an application particularly suited for GC analysis. Twenty-two common process solvents were determined by Kersten[14] and the results were not affected by weakly acidic, weakly basic or neutral drug matrices.

Analyte stability may play a role in the choice of technique. Heat-labile compounds may be more suited for HPLC than GC. GC may be more appropriate than HPLC for compounds that can potentially react with silica surface silanols. As discussed below (section 3.3), nonsilica HPLC phases are available to extend the useful range of HPLC packings. Notwithstanding, sample stability of analytes is often overlooked in methods development, but is often an important consideration.

1.3 Literature Comparisons

The complementary nature of HPLC and GC techniques is exemplified by the separation of atropine [51-55-8] and obidoxime [7683-36-5] parenteral solution, an emergency treatment for toxic organophosphates poisoning. Pohjola and Harpf[15] found that simultaneous GC analysis required tedious and time-consuming sample preparation owing to their different extraction properties, and obidoxime is thermally labile. Retention characteristics differed widely for the two components, making simultaneous HPLC analysis unsuccessful, even under gradient conditions. The authors settled for a GC method for atropine and an HPLC method for obidoxime.

Bile acids analysis generally gave identical results by GC and HPLC/UV (high-performance liquid chromatography/ultraviolet). However, Scalia et al.[16] chose HPLC as the more convenient system for routine analyses since it bypassed derivatization reactions required for GC analysis. GC was then used for verification of HPLC results.

Similarly, Guinebault et al.[17] picked HPLC as the faster and simpler technique for the determination of cicloprolol [63659-12-1]. The GC/EC (gas chromatography/electron capture) method gave somewhat lower levels of detection at $1\,ng\,mL^{-1}$ (compared to $5\,ng\,mL^{-1}$ for the HPLC fluorescence method), and was used for lower level samples.

Regarding TDM, Mehta[18] discussed several advantages of HPLC over GC. The former technique was considered to be quicker and more widely applicable. Sample cleanup was simpler, which is important for TDM. GC methods are often limited to 1-μL injections, whereas up to 100 μL can be injected on an HPLC column. Furthermore, HPLC is more adaptable to automation than GC.

Qualitative and quantitative results by HPLC analyses of nitrosamines were found to be in agreement with GC analyses as reported by Röper.[19] HPLC/UV detection of diethylnitrosamine at 254 nm matched the sensitivity of GC/FID (gas chromatography/flame ionization detection). The significant advantage in this study was that the HPLC method could be run in half the time that it took to run the GC method.

A review of HPLC and GC analyses of basic neurotransmitters and their metabolites was recorded by Yi and Brown.[20] Although the report summarized work published before LC/MS reached a level of commercial importance, HPLC was already seen as gaining on GC/MS, in use for routine analysis of diseases.

Suitability studies between HPLC and GC techniques, such as those described above, are not directly comparable. Contrarily, the conclusions obtained from such an endeavor rely heavily on the particular equipment, conditions and methodology employed. They serve only to introduce the types of problems that might be encountered in investigating a particular HPLC or GC technique. In a practical sense, the actual choice of one technique over the other might best be accomplished by

finding published work on how similar mixtures were analyzed, keeping in mind the considerations discussed above (see section 1.2).

1.4 Ideal Methods Development

The ideal chromatogram presents impurity peaks eluting at or near the solvent front, followed by peaks of interest emerging with baseline resolution (defined by ASTM (American Society for Testing and Materials) as being $R_s = 1.5$) and no more. This scenario provides the best approach in obtaining simple, relatively inexpensive, rugged, reproducible and fast methods. Although ''real'' sample mixtures rarely can be separated so perfectly, the professional chromatographer should strive to procure quality in developing as close to the ideal method as possible. Many pharmaceutical methods are Food and Drug Administration (FDA)-regulated, placing severe restrictions on analytical modifications. As a result, analysts are often locked into running poor methods for years to come. It is very time-consuming and expensive to revalidate or develop a new method. The prudent choice is to develop good methods up front, although the necessary time often is not afforded to the analyst. As a result, analytical laboratories specializing in cGMP (continuous good management practices) methods development have sprung into existence, particularly in the 1990s, to provide this needed service. The most versatile laboratories have expertise in both HPLC and GC for acquiring sound methods. For strategies on methods development, see section 5.3.

2 COLUMN SELECTION

2.1 Tubing Choices

Stainless steel (SS) tubing is the most widely used tubing for HPLC columns. They are rigid and can easily withstand packing pressures that typically go up to 10 000 psi (6.9×10^7 N m^{-2}). The material is relatively inert to a wide variety of solvents and chemicals. A smooth internal diameter (ID) of the SS tube can be honed, which is desirable for packing high-efficiency columns. Root mean square (RMS) units are used in measurement of smoothness of the tubing ID, by a profilometer instrument. A RMS of about 16 or better is standard for 4.6 mm or larger SS HPLC columns, 2-mm ID columns are generally 32 RMS, and 1-mm ID columns are usually 40 RMS or worse.

Reduced bore SS tubes of about 2 mm or less are more difficult to hone and quality control than larger bores. Some are coated with a glass lining, which may provide a smooth internal surface enabling higher efficiency in the packing process. Antithetically, the glass-lined tubes are somewhat susceptible to cracking, which critically inhibits packing good columns. Other commercial SS tubes are micropolished with an abrasive mixture that is forced through the column.

Particularly in some protein and peptide separations, polyetheretherketone (PEEK®) tubing is thought to provide a more inert surface than SS. PEEK® column frits (used to hold the HPLC packing material in the tube) may be employed in conjunction with PEEK® columns. Although PEEK® is a ''hard'' plastic, it is less rigid than SS and it is often more challenging to pack PEEK® columns with the same efficiency as SS columns.

A newer approach to produce smooth, inert HPLC tubes entails coating a proprietary silicone-containing layer of material known as Silcosteel® onto the ID of the SS tube. The strength and ruggedness of the SS tube is retained and excellent improvement in deactivation is realized. The coating treatment was originally developed for deactivating fused silica and flexible SS GC columns and also for air monitoring SS canisters, particularly for use with invasive sulfur-containing samples. Its use has been extended to deactivate bonded silica HPLC particles.

Early GC capillary columns, since the 1950s, were largely made of SS. They were rugged and flexible, but lacked inertness. Latter developments in packed glass columns improved inertness at the expense of being more fragile. In 1979, the more rugged and flexible fused silica tubing was discovered, which was considered to be a breakthrough in GC technology. Capillary fused silica columns are even more inert than glass columns. One limitation they do have is that they become brittle at temperatures exceeding 350 °C. Some improvements in thermal stability have been attempted by devising aluminum cladding or high temperature polyimides as an outer column coating. Another approach was the aforementioned invention of silica-lined, thin-walled, SS capillary columns.

2.2 Dimensions

Majors[21] reported in a 1994 survey that the 250 × 4.6-mm column (53.2%) is the most popular dimension used, followed by the 150 × 4.6-mm (40.9%) size. The 4.6-mm ID columns are popular partially because system peak bandspreading is less of a factor for large ID columns compared to small ID columns according to the relationship in Equation (1):

$$\sigma^2_{\text{total}} = \sigma^2_{\text{sys}} + \sigma^2_{\text{col}} \quad (1)$$

where, σ^2_{total} is the total bandspreading, σ^2_{sys} is the bandspreading due to HPLC system effects and σ^2_{col} is the bandspreading due to HPLC column effects.

Thus, as column ID decreases, the σ^2_{col} term becomes smaller and σ^2_{sys} contributes proportionally more to the total peak bandspreading. Columns down to about 3-mm ID typically do not require significant changes in an existing HPLC system to obtain peak efficiencies similar to a 4.6-mm ID column. However, column IDs of 2 mm or less may require changing the HPLC hardware to minimize the now dominant σ^2_{sys} term. This may include switching to narrower ID and/or length connective tubing, replacing the flow cell with one of smaller volume, and/or using a pump designed to deliver accurate slower flow rates.

Approximate optimum flow rates for the corresponding column ID are shown in Table 1. Note that regardless of the column ID, the linear velocity remains constant for optimum efficiency. The optimum linear velocity is about 0.15 cm min^{-1} in HPLC separations. This value can change slightly depending on the column packing material. The advantages of using smaller ID HPLC columns include solvent savings (Table 1) and increased sensitivity for samples of limited quantity.

Larger ID columns are desirable for semi-preparative (about 10-mm ID) and preparative (about 20-mm ID or larger) separations. Mobile phase and other chromatographic conditions are first explored on a smaller ID analytical column to save time, solvent and costs. After desirable conditions are found, increasing sample loads are run to determine the ''overload'' point. Overload is not strictly defined, but is relative to the needs of the analyst. It is the highest load that can be run before the desired separation is lost. For chemists pursuing 98^+% purity for drug candidate testing, near baseline conditions may be required. Alternatively, a marginal separation may provide the required purity level by the shave/recycling process (Figure 1).

Once the analytical column scouting work has been completed, the following calculations in Equation (2) and Equation (3) may be used to ascertain the corresponding flow rate and load, respectively, for the semi-preparative/preparative column chromatography runs:

Table 1 Optimum flow rates for HPLC columns of varying ID

Column ID (mm)	Optimum flow rate (mL min^{-1}) or normalized solvent usage (mL)
0.3	0.0043
0.5	0.012
1.0	0.047
2.0	0.19
3.0	0.43
4.0	0.76
4.6	1.0
7.0	2.3
10	4.7
20	9.5

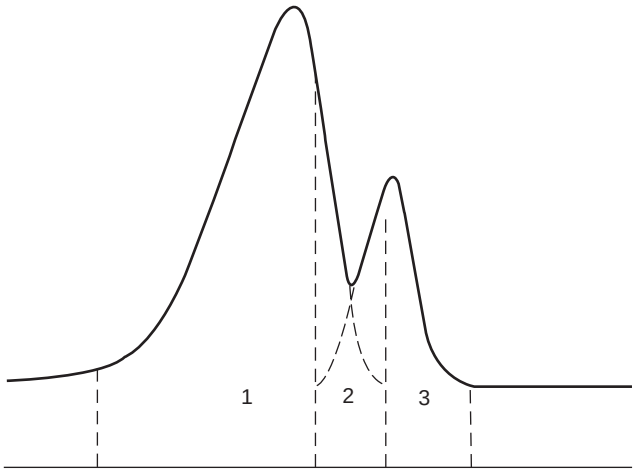

Figure 1 Shave/recycling technique for a nonbaseline resolved preparative HPLC separation. Fraction 1, collect for >98% purity of first compound. Fraction 2, collect and recycle in subsequent injection. Fraction 3, collect for >98% purity of second compound.

$$F_p = F_a \frac{D_p^2}{D_a^2} \quad (2)$$

where F_p is the flow rate of preparative column, F_a is the flow rate of analytical column, D_p is the diameter of preparative column and D_a is the diameter of analytical column.

(Note that this calculation was used to determine the optimum flow rates in Table 1).

$$S_p = S_a \frac{D_p^2 L_p}{D_a^2 L_a} \quad (3)$$

where S_p is the sample load to inject on preparative column, S_a is the sample load injected on analytical column, L_p is the length of preparative column and L_a is the length of analytical column.

It is important to note that Equation (2) and Equation (3) do not necessarily hold true unless the packing material in the preparative column is identical to that in the analytical column. Some chromatographers prefer to use a larger particle diameter (for example, 12-μm preparative particles instead of 5-μm analytical particles) in preparative columns to reduce backpressure, and because of their lower initial cost. However, preparative separations using small particle sizes are not only feasible, but also can be advantageous. Substantially higher efficiency preparative columns can be obtained, and faster flow rates may be used without significant sacrifice of efficiency.[22]

The larger sample amounts injected on preparative columns may cause the UV (ultraviolet) signal to go off scale, even at maximum attenuation. If a nonlinear setting is not available on the detector, the signal may be reduced by setting the detector wavelength away from the UV_{max} (ultraviolet maximum) of the analyte.

Theoretically, column ID does not affect column efficiency. In practice, however, column IDs of about 2 mm or less are more difficult to pack and generally generate about 10 000–20 000 plates m^{-1} lower efficiencies than the corresponding larger ID columns. In contrast, theoretical plates increase in direct proportion to column length. Thus, a 200-mm length column should provide twice the efficiency of a 100-mm column, all other factors being equal. For routine analytical work, lengths of 150 mm or 250 mm are commonplace. Lengths of 50 mm or less are popular for fast separations and may be packed with smaller particle sizes to boost the efficiency lost by using shorter column lengths. Sizes longer than 250 mm may be utilized to gain any needed efficiency to attain a preferred separation. Size exclusion chromatography (SEC) prevalently uses longer columns or two coupled columns to attain desired separations.

Regarding GC columns, the ID choice is primarily based on the concentration range of the targeted analytical compounds. Overall sample concentrations may need to be adjusted to fall within the linear range of the detector being used. The column ID is then chosen that can function within that range. Table 2 includes suggested column IDs based on sample capacity. Resolving power (measured by plates m^{-1} or Trennzahl values) increases with decreasing ID (see Table 2). With very complex samples, the smallest ID column within the sample range capacity should be chosen. If samples are dirty, a wider ID column may be beneficial in reducing column plugging. Column IDs of 0.32 mm or 0.53 mm are recommended for splitless, direct or on-column injections to promote better sample transfer efficiency and increase quantitation. Instrument types and inlet systems may require certain flow rates, dictating the ID selection. Packed column GC inlet instruments are limited to 0.32 or 0.53-mm ID columns because their higher flow rates are essential to improve sample introduction, unless system modifications are introduced. Capillary injection systems accommodate any column ID Finally, instruments with sample valves and purge and trap inletting systems operate best with higher flow rates, which may be afforded by 0.53-mm ID columns.

Resolution increases with the square root of, and analysis time increases directly proportional to, the column length for isothermal separations. More than half of the columns used in laboratories are 30 m in length. Longer GC column lengths are standard for low-boiling complex mixtures, while shorter thicker filmed columns are popular in preventing peak tailing from highly active compounds. Shorter columns also tend to display lower bleed levels and have utility for fast separations of simple mixtures, or for extremely high-molecular-weight compounds.

2.3 Coverage and Film Thickness

Percent carbon values for HPLC bonded phases are frequently reported by commercial manufacturers as an indication of the hydrophobic retentivity of the packing. However, this can be deceiving. A better measurement is the coverage calculation, which in addition to percent carbon, also takes the particle's surface area into consideration. The coverage value furnishes a gauge of phase ligand density, hydrophobicity, surface silanol availability, mixed mode silanol activity and possibly phase stability. These points are addressed in more detail below. Of the coverage formulas used by chromatographers, perhaps the most precise calculation is that introduced by Berendsen and de Galan[23] as follows in Equation (4)

$$\text{Coverage } (\mu\text{mol m}^{-2}) = \%C \frac{10^6}{100 \times n \times 12.01 - (\%C \times (\text{MW} - 1))} \text{SA} \quad (4)$$

where $\%C$ is the percent carbon of bonded phase (measured by elemental analysis), n is the number of carbon atoms in one bonded ligand, MW is the molecular weight of bonded ligand and SA is the surface area of silica.

Table 2 Typical GC column characteristics[a]

Column ID	0.18 mm	0.25 mm	0.32 mm	0.53 mm
Helium (20 cm s^{-1})	0.3 cm^3 min^{-1}	0.7 cm^3 min^{-1}	1.2 cm^3 min^{-1}	2.6 cm^3 min^{-1}
Hydrogen (40 cm s^{-1})	0.6 cm^3 min^{-1}	1.4 cm^3 min^{-1}	2.4 cm^3 min^{-1}	5.2 cm^3 min^{-1}
Sample capacity	<50 ng	50–100 ng	400–500 ng	1000–2000 ng
Trennzahl values	40	30	25	15
Theoretical plates/m	5300	3300	2700	1600
Effective plates/m	3900	2500	2100	1200

[a] From Restek Corporation with permission.

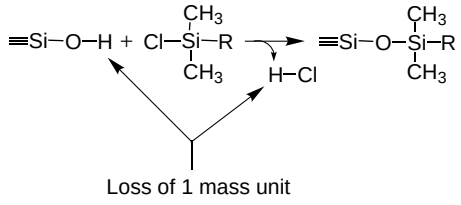

Figure 2 Silanol proton loss consideration in formulating a bonding coverage calculation.

This formula addresses the fact that a proton was removed from the bonded silanol as depicted in Figure 2.

Coverage can also be an indicator of phase ratio (amount of solute in the stationary phase per amount of solute in the mobile phase). Increasing the phase ratio can, in turn, increase resolution by increasing the capacity factor term in the resolution equation (Equation 5). Higher coverage packings generally exhibit increased phase ratios, excluding situations where mixed mode separations play significant roles.

$$R_s = \frac{1}{4} \times \frac{\alpha - 1}{\alpha} \times \frac{k'_1}{k'_1 + 1} \times N^{1/2} \qquad (5)$$

where α is the selectivity, defined as k'_2/k'_1, k' is the capacity factor of the peak and $N^{1/2}$ is the square root of the number of theoretical plates for a particular packed column.

In GC separations, film thickness can be regarded as a reflection of the phase ratio. Thicker films are used for low-boiling compounds to increase retention and, thereby, resolution. Moderately thick films (1 or 1.5 μm) are utilized for samples that elute between 100 and 200 °C. Samples that elute up to 300 °C can be run on standard films (0.25 or 0.5 μm).

2.4 Stability and Reproducibility

Silica-based HPLC packings are still the most widely used phases. They possess excellent mechanical strength, are relatively inexpensive, can be easily modified to render desired chromatographic performance, and can be manufactured reproducibly with control of both particle and pore size distributions. One shortcoming is that they are limited to a pH range of about 2 to 7.5. Kirkland[24] reported that higher pH conditions hydrolyze the siloxane backbone, while lower pH conditions hydrolyze off the bonded-phase moiety. Acid hydrolysis typically results in a loss of chromatographic retention with repeated injections, as bonded phase is gradually liberated from the column. Base hydrolysis maintains retention since the bonded phase attachment remains intact, but abruptly loses efficiency, when the silica backbone finally collapses. Nevertheless, analysts continually push the pH limitations of silica in developing methods for difficult mixtures. Commercial column manufacturers have responded by developing technology to extend the useful pH range of silica packings. One strategy was to manufacture bulky side chains on the bonded phase in an attempt to hinder hydrolysis of the bonded phase sterically. The commercial product, Zorbax® Stablebond®, reportedly increased acid stability.[25] Another study applied larger 150-Å pore, pure silica particles ($1\,\text{Å} = 10^{-10}\,\text{m}$), and high coverage bonding chemistry to obtain essentially the same level of acid stability.[26] Advances in base stability of silica packings have also been reported.[27] In addition, nonsilica packings are used for methods run at pH extremes (see section 3.3).

An investigation into commercially available modified silica HPLC packings will reveal that, almost without exception, bonded phases are prepared with no less than three carbons between the reagent silicon atom and any heteroatom present in the reagent. Thus, there are 3-cyanopropylsilyl phases, 3-aminopropylsilyl phases, 3-glycidoxypropylsilyl phases, etc. The reason for this is that spacers of less than three carbons result in an inherently less stable compound. A review of the stability of bonded phases was published by Unger.[28]

There have been some claims that cyanopropyl bonded phases hydrolyze under acidic HPLC conditions to the corresponding propylcarboxylic acid as shown in Scheme 1. It is true that synthetic chemists use cyano functionality to mask the eventual preparation of carboxylic acid moieties. Still, the cyano hydrolysis in synthesis deprotection procedures[29] is persistently much more invasive than regularly practiced HPLC conditions. This author found no sign of cyano group hydrolysis of a cyanopropyldimethylhydroxysilane model under conditions ranging from pH = 2 to pH = 7, by proton NMR (nuclear magnetic resonance) or HPLC using chemical probes.

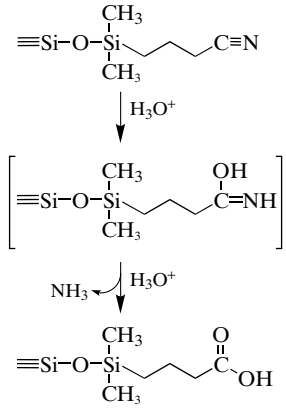

Scheme 1 Hypothetical acid hydrolysis of an HPLC cyano bonded phase.

Ester groups are susceptible to hydrolysis by acidic mobile phases, which should be avoided with both ester bonded phases and/or ester analytes. Whenever practical, ester solutes should be run under neutral pH conditions.

There are commercial phases that are not covalently bonded to solid supports, but rather are "ionically" coated onto the surface of particles. These phases, such as some chiral cellulose packings, are strictly limited to certain types of solvent exposure. Even slight amounts of solubilizing solvent can ruin a column.

Reproducibility of HPLC bonded phases is largely dependent on the replication of the bare silica. Bonding chemistry is also important. Monofunctional bonded phases are inherently easier to reproduce than di- and trifunctional bonded phases (see section 3.2). Some commercial difunctional and trifunctional bonded phases (see section 3.2) contain residual alkyloxysilane or chlorosilane functionality following bonding. Under aqueous mobile-phase conditions, these residual reactive groups might slowly hydrolyze, resulting in an unstable baseline (bleed) over a long period of time, with constantly changing chromatographic performance. Even if a stable baseline can eventually be obtained, reproducibility of this product under similar bonding tactics is customarily poor. Commercial polymeric aminopropyl phases, used for sugar/carbohydrate separations, are prime examples of this type of hydrolysis problem.

HPLC and GC bonding reagents should be purified of any catalysts used in their formation before attachment to silica. It is very difficult to remove them afterwards. Their presence may cause unstable baselines, peak tailing or possible catalytic decomposition of some analytes (e.g. proteins).

Column bleed in GC phases ordinarily is larger for thick films than thin films of the same coating. The phases have varying degrees of inherent thermal stability depending on their chemical nature. Manufacturers routinely list proper operating temperatures, which should be followed to prevent phase degradation. Tolerable residual bleed normally occurs even at low temperatures owing to trace amounts of oxygen, metals, catalysts and other impurities that slowly degrade the polymer film.

The reproducibility of commercial brand stationary-phase performance over several years, whether HPLC or GC, should not be taken for granted. Welinder et al.[30] describe some of the real-life enigmas associated with unannounced manufacturer's phase alterations over time. The authors stress the importance of validation, GMP (good management practice), SOP (safe operating procedures) and QA/QC for the column manufacturer, who should be treated as a contract laboratory.

2.5 Phase Type

Pharmaceutical methods development, whenever possible, should be attempted on a reversed-phase (such as a C_{18}, C_8, phenyl, etc.) column. This is due to the simple hydrophobic mode of separation that is most familiar to the chromatographer. Reversed-phase separations are often rugged, and columns are relatively inexpensive and readily available.

The C_{18} phase (also called octadecylsilyl, or ODS phase) was one of the first modified silica types prepared, owing to the availability of the starting material rather than for any specific scientific motive. Its use rapidly gained in notoriety during the early years of HPLC and it remains the most popular phase today.[21]

It is crucial to note that all commercial C_{18} phases are not alike. In fact, it can be favorably argued that no two C_{18} phases are alike. Figure 3 illustrates the vast possible differences in C_{18} brand column performance.

Shorter hydrocarbon chain reversed-phase columns may be used to reduce retention of highly adsorbed compounds. Phenyl phases add possible dipole:dipole or dipole:induced dipole interactions for another dimension in solute/stationary phase recognition. In particular, phenyl phases appear to have proficient selectivity of carbamates. Cyano phases are one of the best choices for basic pharmaceuticals, showing excellent peak shape and complementary selectivity to C_{18} phases. Cyano columns are also a good choice for steroids, and can be used under both reversed-phase and normal-phase conditions. De Smet and Massart[31] found the cyano phase to be so versatile for normal- and reversed-phase separations of basic drugs, that they used it exclusively.

A modified C_{18} phase engineered for use under high aqueous mobile-phase conditions is of particular interest in reversed-phase retention of polar solutes. Traditional C_{18} phases contain hydrophobic chains that may fold under such conditions to avoid the hydrophilic mobile phase surroundings.[32] Modified commercial C_{18} phases based on this concept adopt a similar fundamental strategy in introducing polar functionality close to the silica backbone. In this manner, the C_{18} chain, hypothetically, remains extended since there is a hydrophilic environment both in the mobile-phase volume and towards the solid support surface.

Newer reversed-phase columns include the intrinsically base deactivated (IBD) phases. They are categorized as those packings that contain a polar functional group within, or intrinsic to, the hydrocarbon chain (see Figure 4). The intrinsic heteroatom(s) apparently is involved in deactivating active surface silanols that would otherwise cause undesirable peak tailing of basic compounds.[32] The heteroatom also appears to play a major role in selectivity, as reversals in elution order are

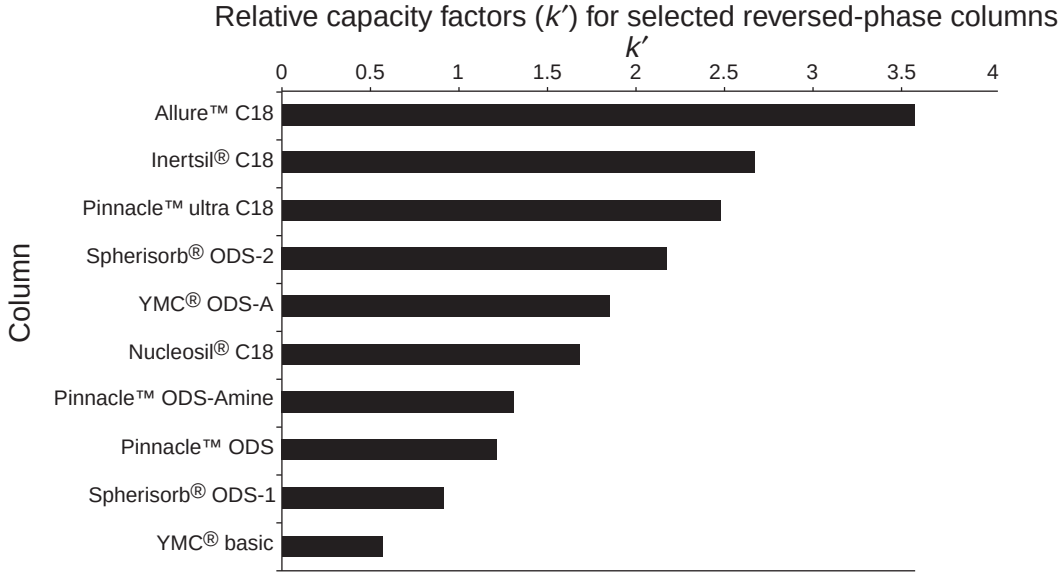

Figure 3 Hydrophobic retention differences of some commercial C_{18} phases. (From Restek Corporation with permission.)

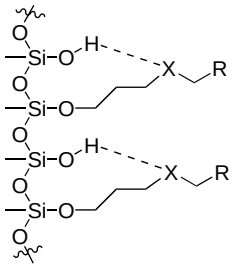

Figure 4 IBD packings. X is a polar group.

commonplace compared to C_{18} phases. Solute mixtures of acids, bases and zwitterions can often simultaneously be separated with good peak shape for all solute types using simple mobile-phase conditions. Separations on this phase are somewhat more affected by small differences in mobile-phase composition than C_{18} phases are, which may make developed methods slightly less rugged. However, some of these phase types show immunity to chain folding, similar to the modified C_{18} phases described above. A few IBD phases are commercially available.[32]

Another variety of reversed-phase column is the halogenated stationary phases discussed in section 5.6.

The most prevalent normal-phase columns employed in the pharmaceutical industry are the amino phase columns for sugar-containing drugs, saccharides, and so on, and the diol phase for polar compounds. The latter phase is regularly used in SEC, especially for proteins and large peptides (see section 4).

Ion-exchange columns are used for highly polar, ionizable drugs that cannot be retained by reversed-phase columns. Weak anion-exchange (WAX) columns contain an immobilized primary or secondary amine functionality that is positively charged at mobile phase pHs below the pK_a of the amine (pH about 7.5 or less). Strong anion-exchange (SAX) columns contain a tertiary amine group that remains positively charged throughout the pH range. The WAX and SAX phases, therefore, are designed to retain anionic drugs (see section 4). Weak cation-exchange (WCX) phases carry a negatively charged carboxylic acid residue at mobile phase pHs around 6 or more. Strong cation-exchange (SCX) phases bear a negatively charged sulfonate group essentially throughout the pH range. Consequently, WCX and SCX phases are capable of retaining basic drugs (see section 4).

In summary, phase selection is primarily empirical. Quantitative structure–retention relationships (QSRR) studies have progressed in identifying at least one factor or descriptor responsible for a given reversed-phase HPLC separation, such as partition coefficient,[33] electronic characteristic,[34] solute bulkiness[35–37] and shape.[38] Yet many more descriptors may be involved in the final separation result. In one QSRR study, Kaliszan and Osmialowski[39] identified 16 structural descriptors in addition to some polar descriptors. An excellent review of QSRR application to reversed-phase HPLC is that authored by Kaliszan.[36]

Retention indices are used as indicators for choosing GC phases for classes of mixtures. For example, benzene serves as an indicator where aromatics or saturated hydrocarbons may elute using a particular GC phase. Similarly, butanol can be used as an indicator for where polar alcohols and aldehydes may elute. Coated GC phases may impart strong dipole and hydrogen bonding characteristics, such as the polyethyleneglycol

100% dimethylpolysiloxane

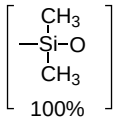

Polarity: nonpolar
Uses: solvents, petroleum products, pharmaceutical samples, waxes

5% diphenyl – 95% dimethylpolysiloxane

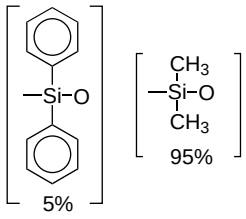

Polarity: nonpolar
Uses: flavors, environmental samples, aromatic hydrocarbons

6% cyanopropylphenyl
94% dimethylpolysiloxane

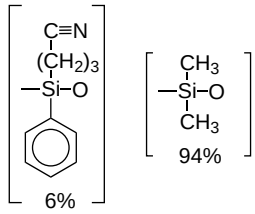

Polarity: slightly polar
Uses: volatile compounds, insecticides, residue solvents in pharmaceutical products

20% diphenyl – 80% dimethylpolysiloxane

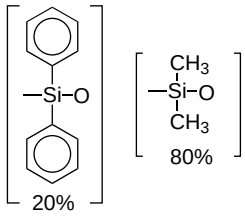

Polarity: slightly polar
Uses: volatile compounds, alcohols

35% diphenyl – 65% dimethylpolysiloxane

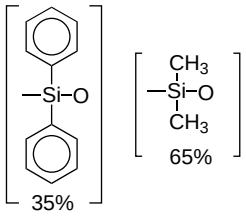

Polarity: intermediately polar
Uses: pesticides, Aroclors, amines, nitrogen-containing herbicides

14% cyanopropylphenyl
86% dimethylpolysiloxane

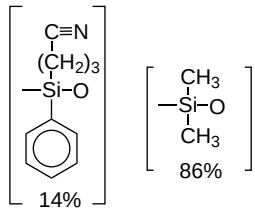

Polarity: intermediately polar
Uses: pesticides, Aroclors, alcohols oxygenates

50% phenyl – 50% methylpolysiloxane

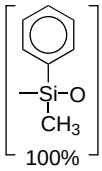

Polarity: intermediately polar
Uses: triglycerides, phthalate esters, steroids, phenols

65% diphenyl – 35% dimethylpolysiloxane

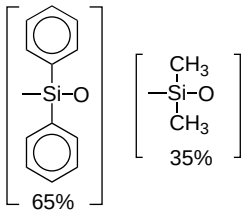

Polarity: intermediately polar
Uses: triglycerides, rosin acids, free fatty acids

trifluoropropylmethylpolysiloxane

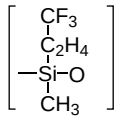

Polarity: selective for lone pair electrons
Uses: environmental samples, solvents, Freons, drugs, ketones, alcohols

50% cyanopropylphenyl
50% phenylmethylpolysiloxane

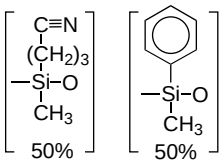

Polarity: polar
Uses: FAMEs, carbohydrates

Carbowax@ PEG

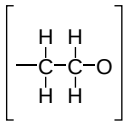

Polarity: polar
Uses: FAMEs, flavors, acids, amines, solvents, xylene isomers

90% biscyanopropyl
10% cyanopropylphenylpolysiloxane

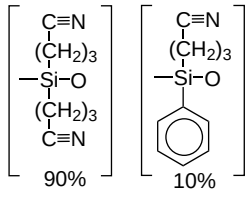

Polarity: very polar
Uses: FAMEs, *cis*/*trans* & dioxin isomers, rosin acids

Figure 5 Some commercially available GC phases. FAME, fatty acid methyl ester; PEG, polyethylene glycol. (From Restek Corporation with permission.)

phase attraction for low-molecular-weight alcohols. This adds a powerful resolving mode for mixtures that would otherwise elute based solely on their relative boiling points, which may be too similar to resolve. In general, polar phases retain polar solutes longer than do nonpolar phases.

3 CHEMISTRY OF BONDED PHASES

3.1 Gas Chromatography Coatings

Most GC phases are constructed by coating a static or dynamic layer of polymerizable silica-containing reagent(s) on the inside wall of a tube and then polymerizing it in place by heat or other processes. Film thickness can be controlled by the original reagent concentration. The reagent may contain a variety of functional groups, each contributing a unique quality to the separation process. Figure 5 shows a few marketed phases.

3.2 Modified Silica Phases

The silica particle surface may contain chemical diversity including siloxanes, geminal silanols, vicinal silanols, isolated silanols, and metal contaminants (Figure 6). A surface chemistry variation between brands is one of the primary reasons why no two C_{18} packings chromatographically behave the same. Geminal and isolated silanols are thought to be the most active silanols in causing undesirable peak tailing. Metal contaminants, such as sodium, magnesium and iron, also play a role in activating silanols. Vicinal silanols, hypothetically, are indicative of a more homogeneous surface, and are less active.

Active silanols may actually improve peak shape for acidic solutes compared to a nonactive packing.[40] These older ''type A'', active silicas are not suitable for basic or zwitterionic compounds. Newer, ''type B'', pure silica packings give superior peak shape for basic and zwitterionic solutes, and perform well for acids as well. Reviews covering the effect of silanols on basic compounds were composed by Nawrocki[41] and Cox.[42]

Traditional bonding chemistry involves covalently bonding a monofunctional, difunctional or trifunctional chlorosilane reagent to a surface silanol group in the presence of a tertiary amine proton scavenger, as shown in Scheme 2. Alkoxysilanes, such as trimethoxysilyl-, triethoxysilyl-, dimethylmethoxysilyl-, etc. may also be used and do not require a proton scavenger. The order of reactivity of silyl electrophiles is: SiCl > SiOMe > SiOEt. Presence of water in the bonding of di- or trifunctional reagents results in polymerization

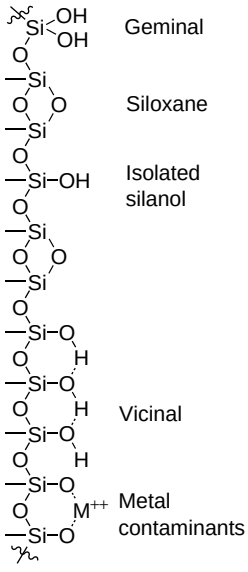

Figure 6 Surface chemistry of silica particles.

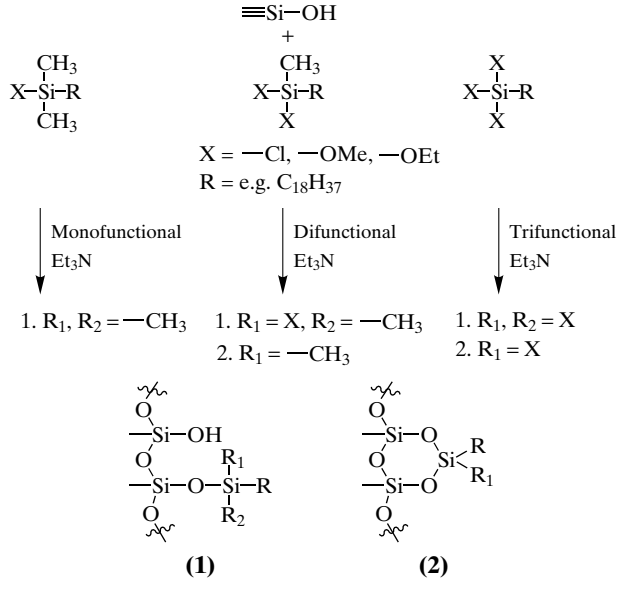

Scheme 2 Traditional bonding chemistry on the silica surface.

of the reagent that may or may not covalently bond to silica silanols. Unbonded polymerized reagent may be difficult to wash out and could appear as bleed under certain mobile-phase conditions. The degree of polymer character of phases may be measured by a National Institute of Standards and Technology (NIST) test mix developed by Sander and Wise.[43] The mix is designated as standard reference material (SRM) 869. It contains three polycyclic aromatic hydrocarbons (PAHs): benzo[*a*]pyrene **(3)** (BaP) [50-32-8], 1,2-3,4-5,6-7,8-tetrabenzonaphthalene **(4)** (TBN) [34478-91-6]

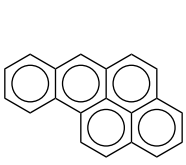
(3)

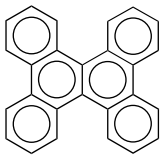
(4)

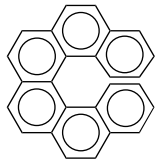
(5)

and phenanthro[3,4-*c*]phenanthrene **(5)** (PhPh) [87-83-7]. Level of polymeric C_{18} bonded phase is measured based on the resulting selectivity and elution order of the PAH components in the SRM 869 mix run on the phase in question.

An end-capping step most often is performed following bonding, to reduce the number of active surface silanols further. The end-capping reagent, being smaller than the bonding reagent, bonds to silanols not accessible to the bonding reagent. Examples of end-capping reagents include trimethylchlorosilane [75-77-4] and hexamethyldisilazane [999-97-3].

Ludwig et al.[26] concluded that bonded-phase ligand density might play a role in hydrolytic stability of the bonded phase. Maximum ligand density or coverage is obtained by using excess bonding reagent and good synthetic technique. Moisture can hydrolyze the reactive bonding reagent and/or polymerize the precursor. Higher coverage also assists in deactivating active silanols. Larger pore size particles (e.g. 200–500 Å) may be more available to bonding reagents than 60–100-Å pore size particles, and may display superior base deactivation due to higher coverage. Larger pores are also preferred for separation of larger molecules such as proteins. Bonded phases with smaller pore sizes have greater surface area, higher percent carbon and longer retention.

3.3 Nonsilica High-performance Liquid Chromatography Phases

Polymeric and other nonsilica HPLC packings avoid active silanols altogether. State-of-the-art polymer particle and pore sizes are controlled with desirable narrow distribution. The pH stability is improved over silica-based packings as shown in Table 3. Polymeric columns still suffer in that they have lower efficiencies than silica-based columns. Polymers also tend to shrink and swell in certain solvents and are not as rigid as silica. For years, researchers have tried to overcome these shortcomings. Progress has been made, but the advantage remains with silica. Silica phases contain at least 10 000 plates m^{-1} more efficiency than the best polymeric phases in well-packed columns. Polymer phases still shrink and swell to some degree, even with high cross-linking. The major utilities of polymeric phases have taken foundation in applications requiring extreme pH conditions, and in ion exchange, size exclusion and biomolecule separations.

Porous glassy carbon (PGC) phases, such as the commercial product called Hypercarb®, can be run at any pH. The solid support is sp^2-hybridized, 100% carbon, formed at high temperatures, with a morphology described as being interwoven ribbon-like strands.[44] Figure 7 depicts how the material is believed to exist as a chromatography medium.

The flat surface of the ribbons acts as the location for certain adsorption processes of the solute. Positional isomers of derivatized phenyl groups, for example, are well discriminated by this adsorption mode of separation.[44] Molecules capable of more contact points with the PGC surface are retained longer than compounds with few contact points as pictured in Figure 8. The *o*-xylene isomer has four contact points on the PGC surface and is retained the longest. The second longest retained isomer is the *p*-xylene isomer which has four contact points temporarily on one side (a, b) of the molecule, but that can flip to the other side (c, d). Therefore, there may be times when it has just two points of contact as opposed to constantly having four contact points, as in the case of the ortho isomer. Finally, the meta isomer has just three points of contact with the PGC surface, and elutes first. This elution order is opposite to that seen on conventional C_{18} phases.

Stereoisomers are also good candidates for PGC phases as outlined by Karlsson and Pettersson,[45] where the flat surface aids in the ''three point attachment'' model first introduced by Dalgleisch.[46] The discriminating PGC phase was used by Azzoui et al.[33] to analyze

Table 3 Characteristics of some commercially available polymeric HPLC phases

Packing range	Manufacturer	Type	Pore size (Å)	Particle size (μm)	pH
PLP-1	Hamilton®	PS/DVB	75	5, 10	1–13
TSK Octadecyl-NPR	Toso-Haas®	PMM	Nonporous	2.5	2–12
Jordi® Gel DVB	Jordi®	DVB	100, 500, 10^3	5	1–14
ODP-50	Asahipak®	PVA	200	5, 9	2–13

PS/DVB, polystyrene/divinylbenzene copolymer; PMM, polymethylmethacrylate polymer; DVB, divinylbenzene polymer; PVA, polyvinylalcohol polymer.

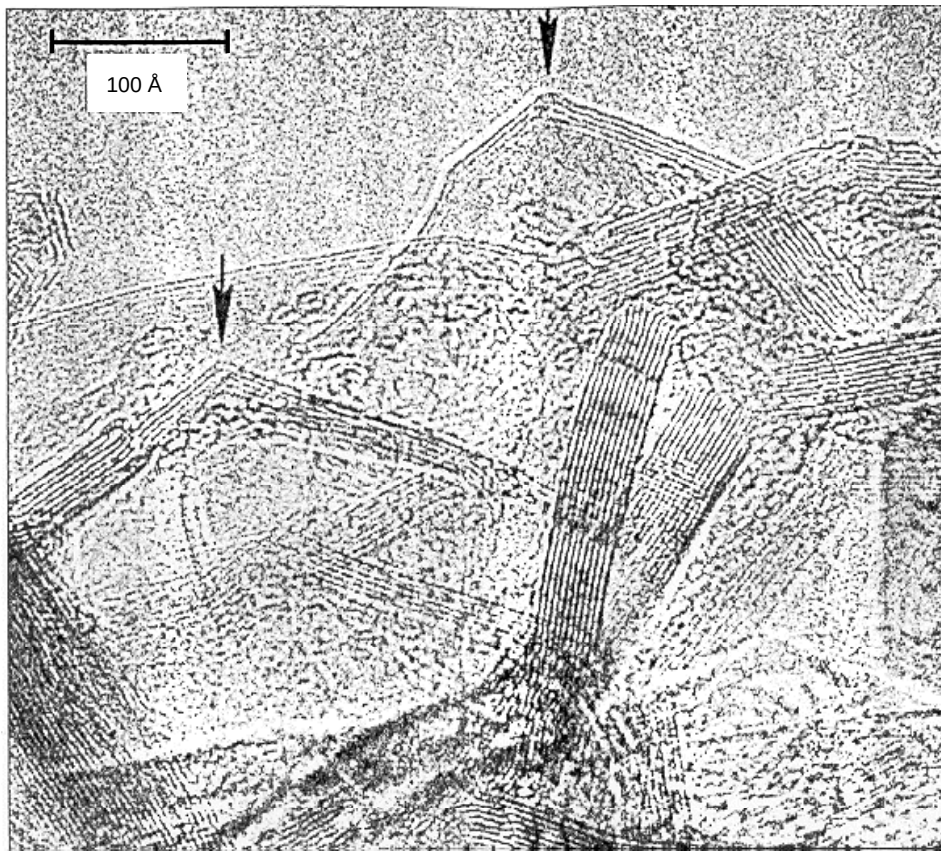

Figure 7 PGC stationary phase. (Reprinted from J. Knox et al.[44] *J. Chromatogr.*, **352**, 3–25. Copyright (1986), with permission from Elsevier Science.)

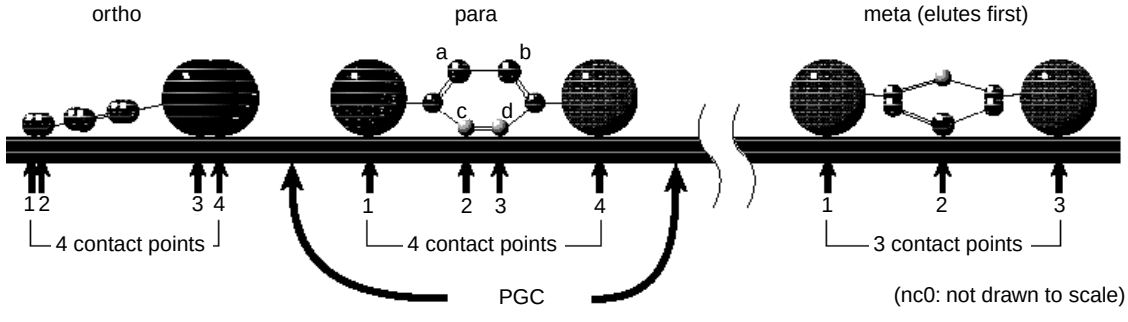

Figure 8 Xylene isomer separation on a PGC phase.

benzodioxinic isomers that could not be separated by normal or reversed-phase HPLC or by GC.

Perhaps the greatest potential for PGC phases is in retaining very polar compounds. Solutes that do not retain on even the most retentive C_{18} phases should be tried on the PGC phase, which also operates on a reversed-phase mechanism. Solvent strength is very different for PGC than for C_{18} phases. Acetonitrile and methanol are strong solvents for C_{18} phases, but are relatively weak for PGC phases. The strength of solvents for PGC phases is solute dependent and not readily definable.[47]

PGC phases have also been prepared at low temperatures, where the degree of glassy carbon formation is controlled.[48] The degree of PGC formation dictates solute selectivity. The higher the process temperature used, the more the phase resembles Hypercarb®. Therefore, desired chromatographic activity can be ''tuned'' by controlling the temperature of PGC formation.

Low-temperature PGC phases have not been commercialized at the time of this writing.

Zirconia-based HPLC phases have recently entered the market under the tradename, Zirchrom®. The 3-μm zirconia solid support is extremely stable and is operable from pH 0 to 14.[49] Zirconia monomeric bonded phases appear to be significantly less stable than the corresponding silica monomeric phases. However, polymer, PGC and ionic coatings on the zirconia particles are available. Mass transfer appears to be slower on zirconia phases than on silica bonded phases, which is compensated for by running separations as high as 200 °C. The higher temperature conditions also reduce higher backpressures caused by the 3-μm particle by decreasing mobile phase viscosity.

Modified alumina phases have also been reported by Holland et al.[50] Although they are not robust throughout the entire pH range as are PGC and zirconia phases, they are stable from pH 2 to 13, which is an advantage over silica phases. Lingeman and Underberg[51] reviewed underivatized alumina which can be used both as an anion exchanger or a cation exchanger for the analysis of basic and acidic compounds, dependent on the pH of the mobile phase. Selectivity for acidic and neutral compounds was particularly good, although the retention mechanism was rather complex on alumina phases.

3.4 Direct Inject Columns

Considerable cost and sample preparation time savings can be realized using phases designed to handle the direct injection of serum and urine samples for pharmaceutical analysis. Early attempts involved automated column-switching techniques between two columns, which are still used today. The sample plasma or urine sample reaches the first column, where proteins and undesirable molecules are unretained and pass directly to waste. Meanwhile drugs to be analyzed are strongly retained and trapped on the first column. A valve is switched directing a stronger mobile phase to elute (usually by backflushing) the trapped drugs from the first column onto an analytical column and then to a detector. The first column (precolumn) is usually bonded with a diol-containing phase of some sort. This provides a ''protein friendly'' environment to which proteins will not stick. The second, analytical column must be sufficiently retentive to perform the drug separation under the strong mobile phase conditions needed to elute them off the precolumn. Yu and Westerlund[52] described such a system using an alkyl-diol precolumn and a C_{18} analytical column for direct plasma analysis of methotrexate [59-05-2]. Clarke and Robinson[53] reviewed several HPLC clean-up procedures, among which was the column-switching technique that gave high drug recovery rates of from 95% to 100%, and with low coefficients of variation. Other acclamations of the practice of this technique for the analysis of drugs in biological samples have been published.[54–56]

Phases were then developed with the concept of using one column that excludes proteins, but retains small molecules, especially drugs. Such a phase eliminates the need for the more complicated column-switching techniques. Pinkerton and Hagestam[57] were the first to introduce this notion, whereby small molecules were retained in the pores, which were bonded with a glycerylpropyl/glycine-L-phenylalanine-L-phenylalanine (GFF) phase. The outside surface was treated by carboxypeptidase A, which is too large to enter the pores. The external surface, after enzymatic treatment, contained glycerolpropyl/glycine chemistry, which occluded protein molecules from entering packing pores. Small solutes were retained in the pores by the GFF phase that survived enzyme treatment. The designation, internal surface reversed phase (ISRP) was given to describe the chemistry. Mohammad and Morrison[58] separated methylene violet biological stain in a DNA-containing sample matrix, on an ISRP GFF column. Perry[59] published an in-depth review of this packing. Shihabi[60] discussed some of the earlier pros and cons associated with ISRPs compared to other direct injection techniques of serums, which are still applicable today.

Other ISRP-like packings have been developed which are modifications of the original to varying degrees. Haginaka[61] summarized several of these phases including semipermeable surface (SPS), dual zone (DZ), shielded hydrophobic phase (SHP) and mixed functional phase (MFP). Collectively, they are known as restricted access phases.

4 MODES OF CHROMATOGRAPHIC SEPARATION

The most widely used modes of pharmaceutical analyses include reversed-phase, normal-phase and ion-exchange chromatography. Mechanisms describe, on the molecular level, how the separations occur.

Reversed-phase separations may take place by partitioning or adsorption mechanisms. Although there is still some debate, it is generally believed that C_{18} and similar hydrophobic phases retain solutes by their partitioning from the mobile phase into a solvated organic coating on the hydrocarbon stationary phase. Retention of solutes, therefore, is involved in a liquid:liquid partitioning process, as shown in Figure 9.

In contrast, PGC phases are thought to operate by an adsorption mode, which may explain the exceptional

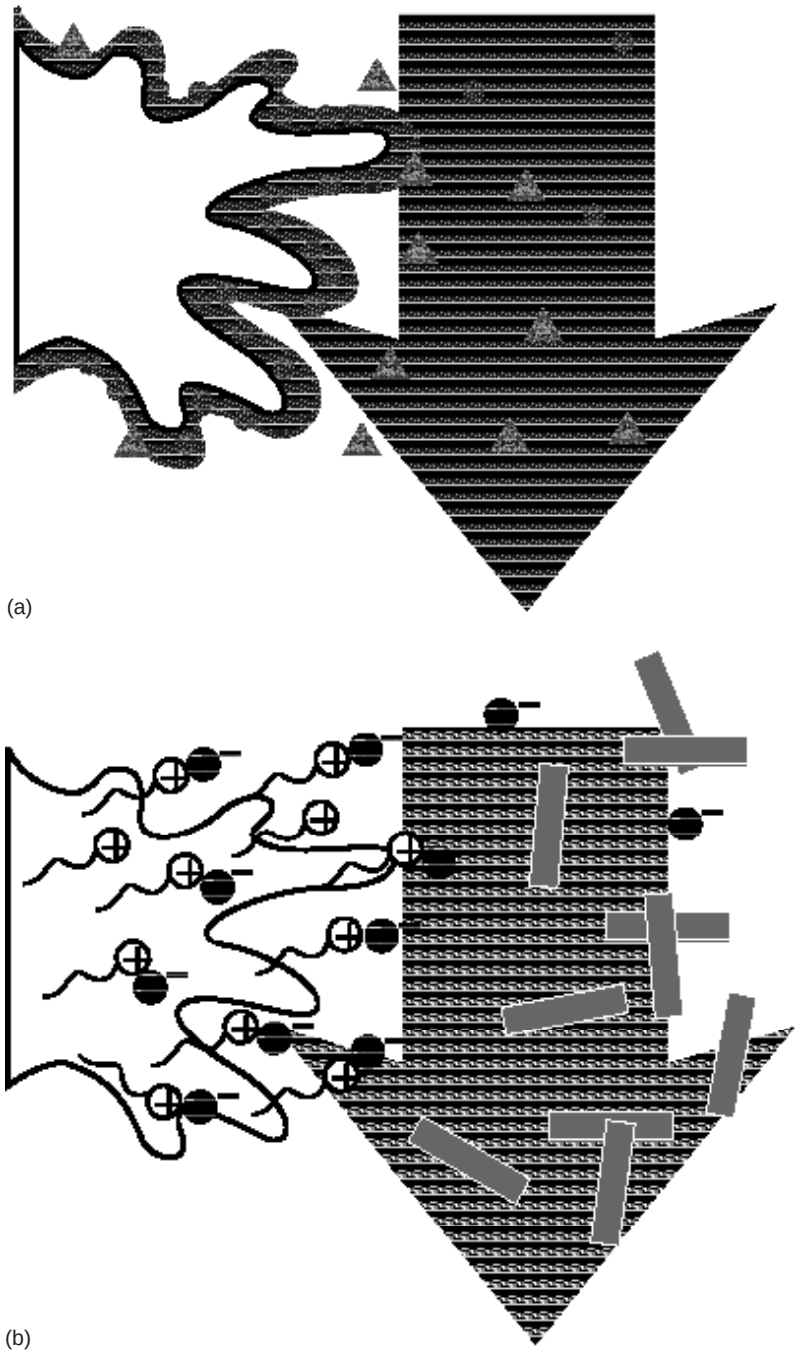

Figure 9 (a) Partitioning, and (b) ion-exchange mechanisms. (From Restek Corporation with permission.)

selectivity these phases show for positional and stereoisomers. Figure 8 depicts a PGC hydrophobic surface that molecules interact with directly.

Normal-phase separations also occur either by partitioning or by adsorption. Prevalent thinking associates diol, amino and cyano bonded phases with the partitioning mechanism, whereas bare silica is postulated to act by adsorption.

Ion-exchange mechanisms include WAX, SAX, WCX and SCX, as discussed in section 2.5 and pictured in Figure 9. For ion exchange of acidic compounds, the mobile-phase pH should be at least one unit above the pK_a of the solute (pK_as of carboxylic acids are typically about 5) to deprotonate and ionize the acid. Basic solutes (pK_as of secondary and primary amines are ordinarily about 8 or 9, respectively) require a

mobile-phase pH of at least one unit lower than the pK_a of the analyte to protonate the species. A strong mobile-phase solvent is favored to reduce secondary modes of separation from transpiring. Elution of target solutes from the stationary phase ensues by mobile-phase additive ions displacing them from the ionic stationary phase. Gradients of increasing buffer concentration are typical, although isocratic ion-exchange separations are not uncommon. In practice, few separations are purely ion exchange, but also contain, for example, some secondary partitioning.

5 PHASE SELECTION BY ANALYTE FUNCTIONAL GROUP TYPE

5.1 Acids

Low-molecular-weight or polar acids are difficult to retain by reversed-phase columns. Retention of acids increases, while bases decrease, with lower pH mobile phases. Correspondingly, acids are less soluble and bases are more soluble in aqueous mobile phases at lower pH. Obviously, there are other factors that also contribute to reversed-phase retention. Most polar acid separations, therefore, are conducted at low pH to gain retention. However, peak broadening occurs when the analyte solubility in the acidic mobile phase becomes a significant issue.

To retain the most polar acid samples, earlier HPLC methods resorted to complex mobile phases with ion pairing agents on active, type A, silica C_{18} packings. For example, Honnegger et al.[62] separated ascorbic acid [50-81-7] and uric acid [69-93-2] from tissue extract using an active packing and a pH 3.5 mobile phase consisting of phosphate buffer, EDTA (ethylenediaminetetraacetic acid), and sodium octylsulfate. Methods of such enhanced complexity are difficult to reproduce.

Chang et al.[63] employed normal-phase amine and β-cyclodextrin (CD) columns to separate substituted benzoic acids under simpler mobile-phase conditions using acetic acid, isopropyl alcohol and methanol. Similarly, Mallet et al.[64] analyzed dihydroxyeicosatetraenoic acids by normal phase on bare silica using hexane, isopropanol, methanol and acetic acid as mobile phase. Separation by GC required derivatization to the corresponding *t*-butyldimethylsilyl or *t*-butylmethoxyphenylsilyl esters. However, normal-phase methods are generally not as robust as reversed-phase methods.

Acids separations may be performed on SAX or WAX columns. Under ion-exchange conditions, the mobile phase pH is adjusted to at least one unit higher than the pK_a of the solute to ionize the acid. As noted in section 2.5, WAX columns must also be used with a mobile-phase pH that is at least one unit below the pK_a of the bonded amine moiety. These two necessary conditions customarily place the working range of WAX columns at about pH 6–7.5. Silica-based SAX columns are also limited to approximately the same pH range, since higher pH mobile phases may hydrolyze the silica backbone. Xia and Gilmer[65] successfully retained the highly polar sialic acid [131-48-6] using a SAX column. Polymer-based SAX columns may be operated at higher pHs. Schmucker,[66] and Bulusu et al.[67] include reviews of some earlier ion-exchange separations of acids.

More recent developments in HPLC column technology seek to perform separations of polar acids by novel reversed-phase packings that provide more robust methods than nonreversed-phase modes of separation. Among these phases are the aforementioned IBD and high aqueous reversed-phase packings (see section 2.5). In another approach, Tischler and Cardellina[68] used a polybutyldiene-coated alumina phase to separate triterpene carboxylic acids without the need for any ion pairing, ion suppressing or buffering additives.

Acidic solutes of moderate to low polarity are relatively easy to separate on either Type A or Type B hydrocarbon phases. Abscisic acid [14375-45-2] and indol-3-ylacetic acid [87-51-4] contain sufficient hydrophobicity to be retained by conventional reversed-phase C_{18} packings.[69] Kwong[70] surveyed a variety of ways that the moderately polar acetylsalicylic acid (aspirin) [50-78-2] and its metabolites can be separated, including the use of C_{18} phases. More hydrophobic solutes, such as the fatty acid γ-linolenic acid [463-40-1], can be converted to esters for reversed-phase C_{18},[71,72] or for polyethyleneglycol phase GC[72] separation. However, derivatization often is not necessary for HPLC fatty acid separations.[73–76] The type of sample preparation for GC or LC (liquid chromatography) separations of pharmaceutically important organic acids depends largely on analytical objectives, as outlined by Liebich.[77] Mukherjee and Karnes[78] reviewed carboxylic acid derivatization reagents for HPLC, if the situation so warrants the additional effort required for this sample preparation.

Phenols should be recognized as weak acids owing to the acidic nature of the phenolic proton. Thomassin et al.[79] reported successful separations of several *p*-hydroxybenzoic esters and *p*-hydroxybenzoic acids on a standard C_{18} packing (also, see parabens in section 6.24). Cabaleiro and Cela[80] employed an active C_{18} column and protonated both stationary-phase silanols and the polyphenolic solutes with 1% acetic acid in the methanol and acetonitrile mobile-phase gradients. Polyphenolic acid retention on a PGC is dependent on the aromatic ring substitution, as documented by Vazquez et al.[81]

HPLC and GC separations of bile acids are summarized by Scalia.[82]

5.2 Bases

Tailing of basic solutes on silica HPLC packings is commonly rationalized by analyte interaction with surface silanols (see section 3.2). Tetraalkyl ammonium ions, sometimes referred to as ion suppressing agents, mask active surface silanols, greatly improving peak shape for basic solutes. Such additives were particularly important in earlier separation techniques before type B silica packings became available.[83] Even with type B silica phases, improvements may be seen with ion suppressing agents.[84] Peak shape can also be improved by using acidic mobile phases that protonate silica siloxides. This may be undesirable for polar basic compounds since lower pH generally elutes bases faster than when the pH is neutral. Good base deactivated phases offer a clear advantage in such situations.

Normal-phase separations of basic compounds on bare silica and alumina were reviewed by Lingeman and Underberg.[85] Additional base solute separations are listed in section 6.

5.3 Acid/Base Mixtures

Mobile-phase conditions optimized for acids may not be favorable for bases and vice versa. Conventional packings do not handle a mixture of both kinds very successfully. Problems with retention and peak shape of one or both solute types are commonplace, even with all but the most base deactivated, type B phases. On C_{18} phases, bases decrease in retention and sharpen in peak shape, while acids decrease in retention and broaden in peak shape with decreasing pH (for example, see results published by Das-Gupa and Mathew[86]). One of the advantages of IBD phases is that they show excellent peak shape for acid/base mixtures for a wide pH range and have complementary selectivity to C_{18} phases.[32] However, highly base deactivated C_{18} phases are more predictive of retention characteristics and separate most mixtures encountered by analysts. Mutton[87] claimed that one phase type successfully separated 95% of 1250 acid and base pharmaceutical candidates synthesized in the preceding year.

For methods development of acids and bases on reversed-phase columns, Zhu et al.[88] recommend starting by varying the gradient steepness and beginning at low pH. If the variation of these two parameters does not achieve the desired selectivity, the effect of temperature should then be explored. Snyder[89] estimated relative contributions of temperature, solvent strength,[90] column type and solvent type towards selectivity and weighed these factors against the ease of operation in changing them. Varying percent organic in the mobile phase, temperature and solvent type (methanol or acetonitrile) was advocated first. If ineffectual, solvent type (methanol, acetonitrile or tetrahydrofuran), column type, pH and/or ion pair reagent concentration deviations should be tried next.

5.4 Zwitterions

Because they contain both acid and base functionality on the same molecule, zwitterionic solutes can sometimes pose challenging problems in obtaining good peak shape and retention. As might be expected, IBD phases have excellent peak shape for zwitterionic compounds, in addition to acids and bases.[32]

Stationary phases of multimode separation ability have been prepared by immobilizing a zwitterionic molecule onto silica particles. Yu et al.[91,92] bonded glutamic acid to silica and were able to separate acids and bases simultaneously, including nucleosides. Chou and Yang[93] simultaneously ran acids, bases and amino acids on an immobilized zwitterionic stationary phase. Yang et al.[94] invoked a similar ''like-retains-like'' strategy in preparing two other zwitterionic stationary phases.

5.5 Polar Compounds

Retention of polar ionic solutes is discussed in sections 5.1 through 5.4. As with those solutes, methods development of nonionic polar analytes is best performed under reversed-phase conditions, if possible. Newer high aqueous reversed-phase columns and IBD phases (section 2.5) are ideal for extra retention accrual.

5.6 Halogenated Compounds

HPLC stationary phases have been developed that provide exceptional selectivity for pharmaceutical compounds containing halogen atoms. Perhaps conceived by a ''like-retains-like'' strategy, these stationary phases contain halogenated groups themselves. The most popular halogen for this purpose is the fluorine atom, possibly because of its strong electronegative nature. One of the more common commercially available fluorinated packings is known as Fluofix®, a partially fluorinated branched hydrocarbon stationary phase (Figure 10).

Maisch[95] depicted several advantages of this phase. The fluorinated packing separated fluorinated adrenocortical hormones and the epimers 3-β-panaxadiol [19666-76-3] and 3-α-panaxadiol, which could not be resolved on traditional C_{18} phases. It appears to show good selectivity for positional and *cis/trans* isomers as well. Jinno and

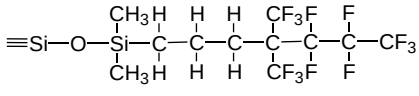

Figure 10 Fluorinated Fluofix® stationary phase.

Nakamura[96] attributed the special selectivity of polar solutes to the larger polarity of the fluorinated phase compared to C_{18} phases. Aboul-Enein and Serignese[97] accomplished the separation of taxol A [33069-62-4] and five of its derivatives using a Fluofix® column.

Taxol and related compounds have also been separated on a pentafluorophenyl (PFP) phase.[98] Another PFP phase, a heptadecafluorodecyl phase and a 3-(heptafluoroisopropoxy)propyldimethylsilane phase exhibited marked retention differences compared to C_{18} phases, affected by different solute functional groups.[99]

Krafft et al.[100] credited the greater transfer energies of the CF_2 over the CH_2 groups with achieving longer retention and better separation of a series of perfluoroalkylated compounds on two perfluoroalkylamino stationary phases compared to a C_{18} phase. Rectilinear plots of capacity factor versus number of carbon atoms were obtained. Also, utility was found for fluorinated stationary phases in the analysis of fluorinated surfactants,[100] flavonoids and fluorinated phenols,[101] and for high recovery rates of proteins.[102] Similar to silica-based fluorinated bonded phases, Hirayama et al.[103] discovered that a perfluoroalkyl acrylate phase provided more efficient separation of fluorine-containing compounds than a C_{18} packing. It is noteworthy that one nonfluorinated phase, poly(vinyl *p*-*t*-butylbenzoate), also showed preferential retention of halogenated solutes compared to C_{18} silica, polystyrene/divinylbenzene and poly(methyl methacrylate–ethyl dimethacrylate) phases.[104]

Dhanesar and Poole[105] compared GC fluorinated phases to nonfluorinated phases in separating some halogenated solutes. Organic compound retention was determined to be about 2–6 times less on the fluorinated versus nonfluorinated stationary phases of similar polarity.

Fluorinated phases are not the only phases suitable for halogenated compound separations. Conventional C_{18} packings often are sufficient. For example, 5-fluoro-2′-deoxycytidine [10356-76-0] and 5-trifluoromethyl-2′-deoxycytidine [66384-66-5] were detected on a C_{18} phase by Briggle et al.[106] Leone and Collins[107] employed simple isocratic mobile phases in separating nine simple perhalogenated compounds on C_{18} phases. Interpolation curves of log mass distribution ratios versus number of solute chlorine atoms accurately predicted observed retention times.[108] Yamaguchi and Hanai[109] compared high versus low carbon load and end-capped versus nonend-capped C_{18} phases for solutes including some halogenated phenols. They concluded that the low carbon load, nonend-capped C_{18} phases gave superior selectivity, but suffered from short column lifetime. De Vries and Rekker[110] separated halogen-substituted aromatic compounds on bare silica and alumina packings under normal-phase conditions. Retention was significantly affected by the substitution pattern of the aromatic solutes and by the bulk and mobility of the mobile-phase molecules.

Documented GC analysis of halogenated pharmaceuticals, as with nonhalogenated pharmaceuticals, is far less prevalent in recent literature than is HPLC analysis publications. GC analysis finds usefulness with more volatile compounds in pharmaceutical analysis, such as drug headspace analysis,[111] urinary halogenated hydrocarbon degradation products in drugs of abuse screening,[112] and analysis of the highly volatile anesthetic, halothane.[113] Some organohalogen drugs in legal medicine analysis by GC/MS were summarized by Suzuki et al.[114]

5.7 Chiral Compounds

Drug candidates containing a chiral center(s) present a unique challenge to the analytical chemist. Racemic compounds, differing only by the arrangement of atoms in space, represent the most challenging of all mixtures to separate. Isolation of the individual enantiomers is very important for evaluation of their respective biological activity and toxicity, which may be quite different. For example, Silber et al.[115] reported that the (*S*)-isomer of propranolol is 100 times more potent as a β-blocker than the (*R*)-isomer. In extreme cases, racemic drugs may be considered to be only 50% pure owing to the presence of the inactive enantiomer. Krstulovic[116] summarized some complex activity, toxicity and pharmacokinetic scenarios possible with racemic drugs. Although the US FDA does not currently require it, potential biological activity differences associated with stereoisomers make it prudent for pharmaceutical companies to determine each enantiomer's individual activity and toxicity. Chromatographic resolution of enantiomers has rapidly expanded as a result. This work mainly involves three processes for chromatographic separation of racemic mixtures, in using chiral stationary phases (CSPs), chiral mobile phase additives (CMPAs) and a CSP and CMPA simultaneously.

Choice of CSPs largely remains empirical at the time of this writing. The best approach is to choose a column that is known to separate racemates of similar structure, and/or meets the general requirements outlined in the literature for successful chiral recognition of the CSP for the solute. Trace analysis of enantiomers may benefit from having the option to choose between a (*R*)-CSP and the corresponding (*S*)-CSP. It is generally desirable to have the minor enantiomer elute before the major enantiomer, so that the latter does not tail into the former, in closely eluting peaks. If, for example, a (*R*)-CSP elutes the minor enantiomer last, then switching to the opposite (*S*)-CSP will reverse the elution order, as desired, without changing any other chromatographic conditions.

In chiral separations where mixtures contain overlapping metabolic peaks, two or more columns with different selectivities may be used to effect separation. For example, Fried and Wainer[117] used a diol silica precolumn to perform an achiral separation of verapamil [52-53-9]/verapamil metabolites and norverapamil [67812-42-4]/norverapamil metabolites, which was then coupled to a chiral column for enantiomer separation. Without the initial achiral separation, overlap of nonenantiomeric peaks was problematic. Ducharme et al.[118] commented that chiral methods validation ideally should be conducted on each enantiomer and the racemate, in order to document the amount of interference between the enantiomers, if any. Figure 11 exemplifies one possible interference, where Fujima et al.[119] recorded racemization of lorazepam [846-49-1] at 25 °C. The baseline rise between eluting enantiomer peaks is indicative of racemization and the method should be modified to avoid this problem.

Chromatographic separations of enantiomers was initiated in 1952 by Dalgleish[46] in his ''three point interaction'' model depicting the stereogeometric interaction of racemic solutes with a single-enantiomer stationary phase. (Biochemists may readily associate the analogy

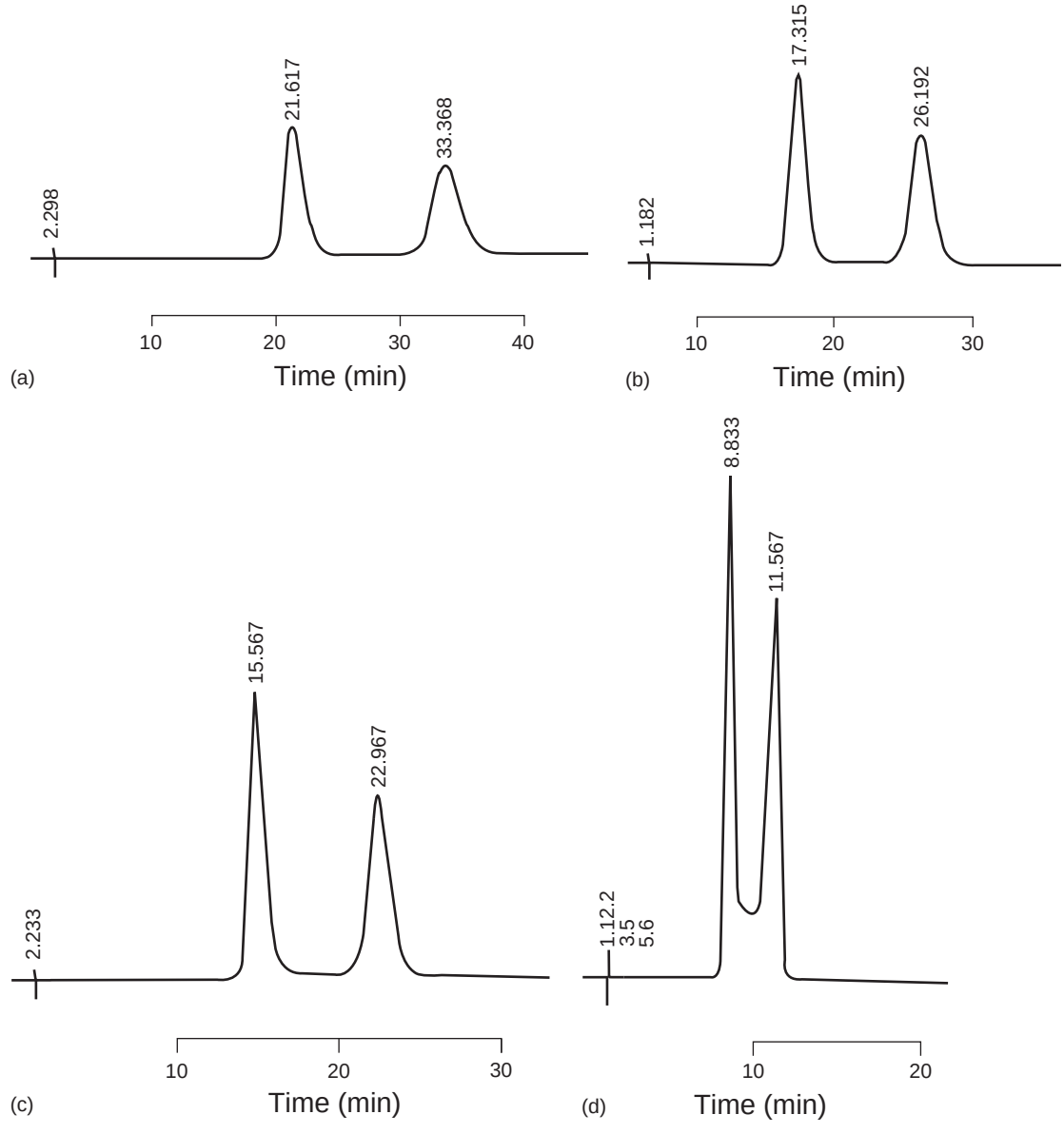

Figure 11 Temperature-dependent racemization of lorazepam enantiomers on an ovomucoid (OVM) chiral column. Column temperature: 7 °C (a), 10 °C (b), 15 °C (c) and 25 °C (d). (Reprinted from H. Fujima et al.[119] *J. Liq. Chromatogr.*, **16**(4), 879–891, 1993, by courtesy of Marcel Dekker, Inc.)

of this model to the Ogsten, ''three point landing'' model[120] diagramming stereochemical interactions of active biomolecules with enzymes/proteins). The model involves interactions between three groups bonded to the chiral center on the analyte with three groups on the CSP, as displayed in Figure 12. Solute enantiomer a has a more favorable ''fit'' to the stationary phase regarding attractive interactions than does solute enantiomer b. Therefore, a is retained longer than b. It is important to note that not all three interactions need to be attractive interactions. Steric hindrance and charge:charge repulsions may also contribute.

Some of the earliest commercially available chiral HPLC phases were the so-called ''brush type'' or ''pirkle'' phases, the latter of which was named after the inventor. One of the three points of interaction which is a characteristic of pirkle phases is $\pi : \pi$ attraction. The CSP dinitrobenzoylphenylglycine (DNBPG) molecule in Figure 13 has electron-withdrawing groups on the aromatic ring that produce an electron-deficient or π-acidic phenyl ring. This CSP is designed for solutes containing an electron-rich or π-basic aromatic ring, which acts as one of the three interactions for chiral discrimination. Solutes that are successfully separated on this phase often, but not always, contain an amide group on the ring, similar to the CSP. This structural ''requirement'' can be met for primary and secondary amine analytes by derivatization with, for example, benzoyl chloride or naphthoyl chloride.

For π-acid solutes, CSPs containing π-base phases are commercially available, such as the naphthylethylurea CSP displayed in Figure 14. With this phase, a suitable

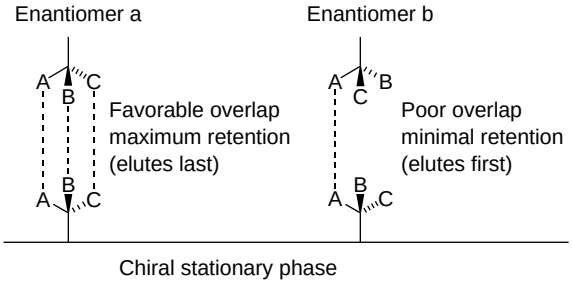

Figure 12 Dalgleish three point interaction model.

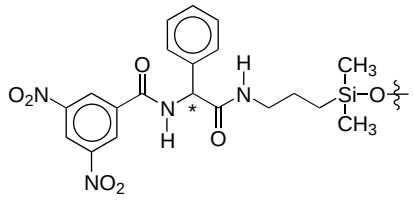

Figure 13 π-Acidic CSP. ⋆ CSP chiral center.

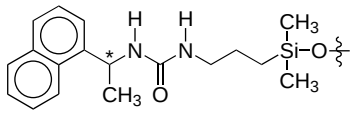

Figure 14 π-Basic CSP. ⋆ CSP chiral center.

derivatization reagent for primary and secondary amines might be 3,5-dinitrobenzoylchloride.

Many commercial brush-type phases have been developed over the years, with varying degrees of success. In general, they do not demonstrate a broad range of applicability without solute derivatization.

The spectrum of underivatized compounds capable of being separated on one particular phase increased with the advent of CD phases. Commercial CD phases are currently available for both HPLC and GC separations. The three most prominent phases, α-, β-, and γ-CD, contain 6, 7, and 8 glucopyranose rings, respectively, hooked by 1,4-linkages as shown in Figure 15. The respective cavity sizes are 4.5 Å, 6 Å and 8 Å in diameter.

Retention may occur by either inclusion of a hydrophobic portion of the analyte into the cavity or by ''capping'' the top of the CD cavity.[121] Hydroxy groups at the mouth of the CD molecule are available for hydrogen bonding as part of the discrimination process, which is especially important for enantiomers, and also to some extent in separations of structural and geometric isomers. Commercial permethylated CD columns are also available, and appear to separate an even broader range of racemates than unmodified CD phases. They should be tried first, unless the analyte contains hydroxy functionality, which underivatized CD phases may better recognize through hydrogen bonding.

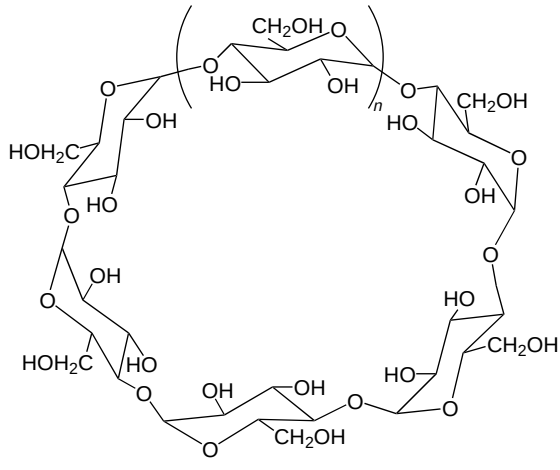

Figure 15 CD structures. α-CD ($n = 1$), β-CD ($n = 2$) and γ-CD ($n = 3$).

Immobilized proteins (affinity phases) are fairly versatile in separating a wide range of underivatized chiral solutes. They take advantage of the natural chiral recognition processes of biological systems. Commercial columns include bovine serum albumin (BSA), human serum albumin (HSA), OVM, α_1-glycoprotein (AGP) and other phases. The proteins are usually covalently bonded to a 300-Å silica support, which contains a large enough pore size to incorporate the protein. Among the serum albumin phases, it is interesting to note that, although there is a high homology between the various sources (bovine, human, sheep, pig, etc.), Tittelbach and Gilpin[122] found that small differences in the amino acid sequence can have a pronounced effect on solute selectivity. Lower column temperatures often increase resolution, presumably by freezing protein conformation, thereby improving enantiomer recognition. Some drawbacks of protein phases include slower mass transfer characteristics causing peak broadening and mobile-phase restrictions that are recommended by column manufacturers to avoid denaturing the protein.

Derivatized cellulose CSPs provide a broad range of separations as summarized by Aboul-Enein and Islam.[123] With the exception of triacetyl and tribenzoylcellulose, derivatized cellulose phases are largely covered by patents owned by Daicel Chemical Industries Ltd., Tokyo, Japan.

Chiral ligand exchange chromatography (CLEC) was introduced by Davankov[124,125] in 1970. Amino acid racemates may be separated without derivatization, using the CLEC phase, which incorporates an immobilized single enantiomer amino acid. The amine and carboxylate functional groups complex with a metal ion, such as copper, which in turn complexes with the amino acid solutes. The solute enantiomer that forms the most stable complex is retained longer, giving rise to the separation process. Disadvantages of this technology include long equilibration times associated with the metal-containing complex mobile phase, and that the technology is not compatible with MS (mass spectrometry) detection.

Chiral crown ether phases embody cavities involved in the separation of racemic mixtures containing primary amines. Sousa et al.[126] first introduced these phases, which were latter modified for free amino acid racemate separations.[127] Secondary and tertiary amines do not separate on these phases, which limits their usefulness. However, Hilton and Armstrong[128] have successfully separated dipeptides.

Chiral imprinted gels are very specific for a targeted racemic mixture. A solute enantiomer is used as a template, bound in a cross-linked polymer. The template molecule is then removed by hydrolysis, leaving a cavity that retains the identical enantiomer or one with very similar structure.

Another application-specific chiral phase was developed for racemic barbiturates, hydantoins and glutarimides by Feibush et al.[129] The phase contains functionality strategically located to form triple hydrogen bonds with the solutes, similar to biochemical base pair formation.

A more recent development of CSP technology is the macrocyclic chiral phases published by Szczepaniak and Szymanski,[130] and then expanded upon by Armstrong et al.[131] Molecular cavities, reminiscent of crown ether and CD cavities, but larger and of different chemistry, are thought to be involved in the separation process. The number and types of racemate that are separable is reported to be greater than any other single chiral phase, although the reproducibility of the commercial phase may prove to be difficult to control. Among the most popular covalently bonded macrocyclic chiral phases are vancomycin [123409-00-7] for N-protected amino acids and drugs[131] including warfarin [81-81-2],[132] teicoplanin [61036-62-2] for protected amino acids[132] and underivatized amino acids, peptides, α-hydroxycarboxylic acids, cyclic amides and amines,[133] and ristocetin A [1404-55-3] for derivatized or underivatized primary and secondary amino acids and di- and tripeptides.[134]

CMPAs can be used in conjunction with a standard C_{18} or other reversed-phase column. Here, the recognition occurs by the complexing of the CMPA with each of the enantiomers. The complex formation results in two pseudodiastereomers, which nonchiral phases are capable of discriminating between.

Chiral separations can be predictably improved by using both a CSP and a structurally homologous CMPA in concert. Somewhat surprisingly, the beneficial CMPA (of opposite stereochemistry to the CSP) does not elute the less retained solute enantiomer from the column relatively faster than the other solute enantiomer, but rather performs a proposed stereoselective ''salting out'' process. The longer retained solute isomer, hypothetically, is forced to reside longer on the bonded phase by being salted out of the mobile phase. Therefore, enhanced CSP/CMPA enantiomer separations occur solely due to the more retained enantiomer, while the less retained enantiomer is not affected in a chiral recognition sense. Duff et al.[135] proposed a ''push/pull'' model to describe this process, as shown in Figure 16.

6 PHASE SELECTION BY ANALYTE CLASS

6.1 Adrenergic

Pharmaceutical separations of adrenergic β-blockers can be accomplished on ordinary C_{18} phases. Maguregui et al.[136] analyzed urinary alprenolol [13655-52-2],

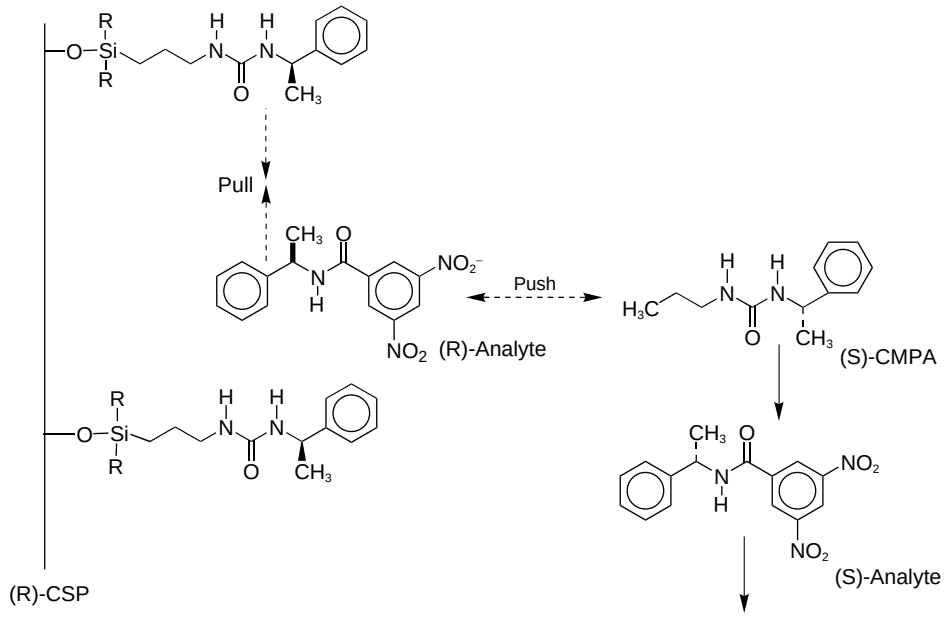

Figure 16 Push/pull model for CSP/CMPA synergism. (Reproduced from K. Duff et al.[135] *Chiralty*, **5**, 201–206. Copyright 1993, with permission from Wiley-Liss, a subsidiary of John Wiley & Sons.)

atenolol [29122-68-7], metoprolol [37350-58-6], nadolol [42200-33-9], oxprenolol [6452-71-7] and timolol [26839-75-8] at the parts-per-billion level using a common C_{18} column and electrochemical detection. Pindolol [13528-86-9] in urine and plasma was determined by Ohta et al.[137] using a column-switching technique (see section 3.4), with the second column (the analytical column) being a C_{18} phase. It is important to note that lower wavelengths of about 230 nm should be used for many β-blockers in order to detect them by UV absorbance.

Chiral separations for racemic atenolol, bevantolol [59170-23-9], celiprolol [56980-93-9], metipranolol [22664-55-7], metoprolol, oxprenolol, pindolol, propranolol and sotalol [3930-20-9] were accomplished by Ceccato et al.[138] on an AGP phase (see section 5.7). Attempted chiral resolutions of 18 β-blockers were conducted on a newer generation pirkle phase and compared to separations on a cellulose phase (see section 5.7).[139] Toyo'oka et al.[140] obtained low levels of detection of ten racemic β-blockers by derivatization with a chiral fluorescent reagent, followed by HPLC analysis of the resulting diastereomers on a C_{18} phase.

Although they are not nearly as well documented in the literature, chiral β-blocker separations can also be directed by GC CD phases, as demonstrated in Figure 17.[141]

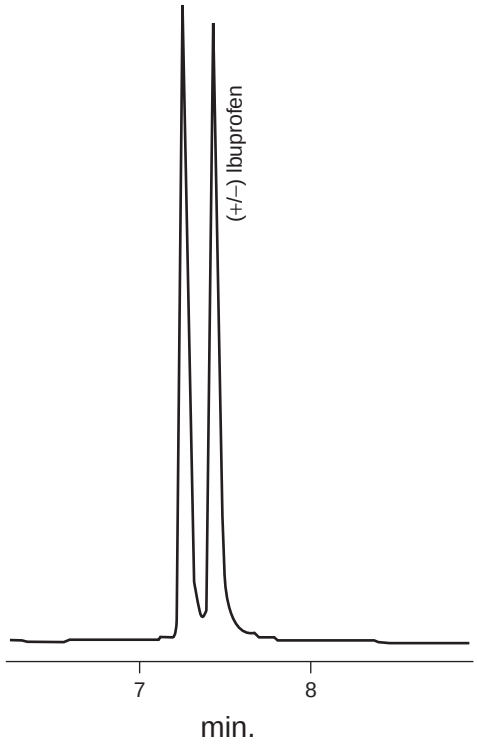

Figure 17 GC Separation of racemic ibuprofen. 30 m, 0.32-mm ID, 0.25 m Rt-βDEXsm, on column injection, concentration approximately 125 ng each enantiomer; oven temperature 175–200 °C at 2 °C min^{-1}, injection/det. temperature 200 °C/230 °C; carrier gas, helium at 60 cm s^{-1}; detector type, GC/FID; split ratio 13 : 1 using cup splitter sleeve. (From Restek Corporation with permission.)

6.2 Alkaloids

Isoquinoline [119-65-3] and other alkaloids are best separated by HPLC. The choice of separation mode

varies with the sample matrix type. For example, alkaloids in poppy extracts are preferably separated by ion pair chromatography on an aminopropyl phase, while those in biological liquids may primarily be separated by ion-exchange or reversed-phase chromatography.[142] Metabolites of alkaloids are separable from their parent compounds by HPLC.[143] Review articles on alkaloid analyses have been published by Kabulov et al.,[144] Papadoyannis[145] and Valka.[146]

6.3 Amino Acids

Earlier amino acid separations (also see section 5.4) were run by methods using dedicated cation-exchange columns such as outlined by Walker and Mills.[147] Post-column derivatization with, for instance, ninhydrin[148] was common to improve detection. A large variety of precolumn derivatization reagents gradually emerged for the analysis of amino acids. Review articles can be found for *o*-phthaldialdehyde,[149] dabsyl chloride,[150] phenylisothiocyanate[151] and fluoren-9-ylmethyl chloroformate.[152] Selective derivatization reagents, for example benzoin for arginine [74-79-3] and glyoxal for tryptophan [54-12-6], are reviewed by Cui et al.[153] The derivatized amino acids can conveniently be run on reversed-phase columns.

6.4 Antibacterial/Antifungal

Watanabe et al.[154] analyzed methylated oryzalides and oryzalic acid rice extracts containing antibacterial activity by GC/MS. The derivatized terpenes were separated on a methyl silicon GC column and detected by selective ion monitoring. Fluconazole [86386-73-4] in human plasma was analyzed by Beijnen et al.[155] on a HP-1 GC phase. Other GC methods require modification of the analyte, such as derivatization of plasma canellal [66550-09-2] with bis(trimethylsilyl)trifluoroacetamide and subsequent GC/MS analysis on a DB-1 column.[156] Saponified steroidal lanosterol [79-63-0] isomers were derivatized by Howell and Mallet[157] to the corresponding trimethylsilyl or *t*-butyldimethylsilyl ethers before isolation on a SE-30 column.

Antifungal agents are more prevalently separated on reversed-phase HPLC columns. A cyclic hexapeptide antifungal agent was separated from plasma or urine on a conventional C_8 column by Schwartz et al.[158] Zornes and Stratford[159] used the same type of C_8 phase to assay a lipopeptide antifungal agent in plasma. Another reversed-phase column, this time a C_{18} phase, was employed by Chung and Paik[160] in discovering the presence of the antifungal active alkaloid ingredient, berberine [2086-83-1], in Japanese quince. Other examples of antifungal separations on conventional reversed-phase HPLC columns include fluconazole,[161,162] nystatin [1400-61-9],[163] heptaenes,[164] the nonapeptides, herbicolin A [74188-23-1] and B [74188-24-2],[165] and bromosalicylchloranilide [3679-64-9].[166] Both reversed-phase and normal-phase columns were used by Bergeron et al.[167] in isolating antifungal polyphenols from milkwort plants. Albet et al.[168] used a cyano phase to analyze the imidazole antifungal, sertaconazole nitrate [99592-39-9], which provided excellent separation of the target from impurities. Stopher and Gage[169] used a SEC column to elute directly injected plasma proteins, while trapping voriconazole on a C_{18} guard column. HPLC separation was then conducted on a 250 × 4.6-mm C_{18} phase. A less conventional β-CD column was successfully employed by Kim and Lin[170] in analyzing a triazole drug in human plasma. Reviews of antifungal analyses have been published by Woestenborghs and Heykants,[171] Ristuccia[172] and Warnock et al.[173]

6.5 Anti-inflammatory

Anti-inflammatory (also see section 6.30) drug analyses are largely accomplished by reversed-phase chromatography. Review articles covering methods through the 1980s include those written by Van Overbeke et al.,[174] Bojarski,[175] McElnay[176] and Wilson.[177] The former two referrals specifically cover enantiomer separations. Ho and Chen[178] have summarized some analytical methods.

6.6 Antipsychotic

A novel basic antipsychotic agent (also see section 6.9) and its N-oxide metabolites in plasma were analyzed using a C_{18} phase and an ion pairing reagent.[179] Reversed-phase HPLC has also been suggested as a possible diagnostic tool for postpartum psychosis.[180] Psychotic subjects showed a different pattern of compounds in their cerebral spinal fluid in comparison to normal subjects.

6.7 Anthracene Derivatives

Van den Berg and Labadie[181] separated several structurally similar cathartic (purgative) agents on an active C_{18} phase. Excellent separation was obtained for the anthracene-like compounds chrysophanol [481-74-3], physcion [521-61-9], emodin [518-82-1], aloe emodin [481-72-1] and rhein [478-43-3]. A custom copper(II)phthalocyanine sulfonylaminopropyl silica phase was found to separate electron-rich anthracene and other compounds by Saito et al.[182] Anthracene moieties are also used as fluorophore derivatization agents, such as that described by Goto et al.[183] for 3-oxo-bile acids.

6.8 Barbiturates

(See section 6.17.)

6.9 Benzazepines/Benzodiazepines

Benzazepines and benzodiazepines are sufficiently lipophilic that they are often analyzed on a C_8[184–189] or phenyl[190] phase instead of a C_{18} phase. Still, C_{18} phases dominate in the literature.[191–210] Bare silica, cyano and amino normal-phase columns have also been used for several benzodiazepines containing hydrogen bond donor or acceptor groups.[211] Radwanska et al.[212] studied dibenzazepines' interaction with melanin by analyzing the drugs on a melanin-bound aminopropyl silica phase and hydrophobicities were determined on an immobilized artificial membrane (IAM) phase. The chemistry of the IAM phase is designed to mimic cell membranes, as outlined by Alvarez et al.[213]

Benzazepines and benzodiazepines may exist in two twist boat conformational forms, a cup-shaped *anti* form and an extended *syn* form.[214] These ''frozen'' conformations may be visualized by proton NMR, and conformer adsorption coalescence can be obtained by heating the NMR sample. It is interesting to note that Fellegvari et al.[215] documented the ability to study benzodiazepine conformation interconversions, also by HPLC.

Benzodiazepine positional isomers were separated by Jin et al.[216] on a C_{18} phase. Franzelius and Besserer[217] separated diastereomeric isomers using a β-CD column (see section 5.7). Enantiomeric separations were accomplished on cellulose CSPs (see section 5.7) without the need for derivatization.[218] Some benzodiazepine enantiomers, specifically related to the 3-chiral position, were separated by Nishikawa et al.[219] and Fitos et al.[220] Separations related to the 5-chiral position are also reported.[220,221]

This class of analytes is somewhat unique among the pharmaceuticals in that, in addition to the many HPLC methods, there are also numerous GC separations appearing in the literature. Most methods are performed by GC/MS and use a variety of columns including SPB-5,[222] DB-1,[223] DB-5[224] and a 5% phenylmethylsilicone phase.[225] Methods for urinary samples,[223,224,226–231] blood/plasma samples,[222,225,232–238] and tissue samples are reported.[239] Noninvasive analyses have been gaining emphasis and forensic science has made advancements in technology, reflected by the development of analyses of benzodiazepine in human hair[240–242] and sweat.[243]

6.10 β-Blockers

(See section 6.1.)

6.11 Blood Volatiles

Typically, determination of blood volatiles is conducted by GC/MS using headspace analysis. Blood level ethanol [64-17-5] and other volatiles in drunk drivers were measured by Schuberth[244] on a DB-WAX phase coated with polyethyleneglycol. Isovaleraldehyde [590-86-3] in plasma from hepatic encephalopathy subjects was separated on a polyoxyethylene glycol phase by Al Mardini et al.[245] Dunn et al.[246] examined *N*-nitrosamines in blood using a carbowax column. Other volatiles analyzed in blood include organic pollutants,[247–254] halocarbons,[255–257] anesthetics,[258–261] multiple drugs taken by a suicide subject,[262] mercury [7439-97-6] (following ethylation derivatization),[263] blood cell membrane peroxidation residues,[264] solvent thinner[265] and EDTA[60-00-4] (following esterification).[266]

HPLC analysis of blood volatiles generally requires derivatization. For example, Vollmer et al.[267] derivatized ethylene glycol [107-21-1] with benzoyl chloride before HPLC separation on a C_{18} column. However, literature applications are largely of GC, not HPLC, methodology.

6.12 Cardiovascular

Cardiovascular agents containing basic moieties are best separated on type B silica reversed-phase packings. Nevertheless, many literature methods include type A reversed-phase separations. Using the more active type A silica requires ion-pairing and/or ion-suppressing reagents to be added to the mobile phase. Bonazzi et al.[268] used an active C_{18} phase to determine the angiotensin-converting enzyme (ACE) inhibitors, ramipril [87333-19-5], quinapril [85441-61-8], benazepril [86541-75-5], lisinopril [83915-83-7] and enalapril [75847-73-3]. Even with sodium heptanesulfonate mobile-phase additive, a low pH (2.5) was required to avoid peak splitting and band broadening. The more complex mobile phases were also needed for similar cardiovascular separations on type A packings.[269–272] Normal-phase bare silica columns have been used for cardiovascular separations, both with water[273] and without water[274] in the mobile phase. Review articles on the HPLC separation of eicosanoids[275] and on GC analysis of lipoprotein (a) [159577-00-1] and homocysteine [454-29-5] for risk factor studies of cardiovascular disorders[276] have been published.

Chiral separations of cardiovascular drugs include pindolol on an AGP affinity column,[277,278] and several brush-type phases were compared for a series of basic solutes.[279]

6.13 Catecholamines/Phenols

The HPLC analysis of catecholamines (also see section 6.24) has grown since the 1970s and is presently the major method of analysis. Cation-exchange columns are used under acidic conditions, where the amine functionality of the catecholamine is essentially completely

protonated. Newer type B reversed-phase columns often give excellent peak shape and selectivity and are a good first choice phase for methods development. The reader is referred to several review articles for HPLC separations.[20,280–283] Analysis by GC/MS is also popular and is reviewed in the literature.[284,285]

Phenolic compounds are weak acids and may be separated by anion-exchange columns as previously discussed (see section 5.1). Simple reversed-phase packings are a preferred alternative. Bosch et al.[286] documented a review of phenolic compound separations.

6.14 Cephalosporins/Penicillins

Cephalosporins/penicillins are relatively labile compounds, subject to acid and, especially, base hydrolysis. Care should be exercised in choosing a pH at least one unit away from the pK_a of the solute. Penicillins are generally strong organic acids, with pK_a values less than 3. Therefore, HPLC separations most often are run with a mobile phase pH in the range 4.5–7.5.[287] Active reversed-phase packings often show unacceptable peak tailing, chiefly at neutral pH mobile phases. Before type B packings became available, methods using active C_{18} and other reversed-phases resorted to ion-pairing or ion-suppressing mobile-phase additives.[288–291] A method for metronidazole [443-48-1] analysis by Venkateshwaran and Stewart,[292] demonstrates that a type B reversed phase can be used with simpler mobile phases (phosphate buffer and methanol in this case). The IBD phases (see section 2.5) are particularly valuable for analyzing this class of compound yielding excellent peak shape and selectivity even under neutral pH conditions.

Elevated temperatures accelerate decomposition. Because of the lack of thermal stability, GC methods for cephalosporins/penicillins are few.[293,294]

6.15 Cold Medicines

Cold medicine ingredients are relatively easily separated with good peak shape on type B silica reversed-phase packings. Caffeine [58-08-2] is often used in reversed-phase test mixes as a control for testing the level of base deactivation of packings. It maintains fairly sharp peak shape even with the more active phases, whereas other basic compounds may interact with the silanols so strongly that they do not elute.

6.16 Cyclosporins

The undecapeptide cyclosporin A [59865-13-3] is an immunosuppressant agent administered to organ transplant patients, which is slightly better than standard azathioprine and prednisone treatment. Whole-blood monitoring of cyclosporin A (as opposed to untrustworthy plasma analysis) is important in monitoring graft rejection which occurs below the 100 ng mL^{-1} level.[295] HPLC monitoring has been shown to be more accurate than radioimmunoassay which gave spuriously high results in multiple doses.[295] Bowers and Mathews[296] suggested that a broad unsymmetrical HPLC peak shape associated with the molecule is due to the various tertiary isomers in which the compound exists. Moyer et al.[297] used the minimum retention of a C_1 column and elevated temperature to reduce the apparent bandspreading for a more aesthetic chromatogram. Nishikawa et al.[298] latter found that cyclosporins A, C [59787-61-0] and D [63775-96-2] undergo interconversions between two molecular forms, whose chromatographic peak shapes are affected by both temperature and flow rate. Other researchers have used more retentive phases to discern structural cyclosporin[299,300] and cyclosporin metabolite[301,302] differences, particularly with LC/MS analyses.[303] Nazir et al.[304] used a highly selective PGC (see section 3.3) to separate two cyclosporins that differ in structure by just one methylene group.

6.17 Drugs of Abuse

Reviews of methods of analysis of illicit drugs (also see sections 6.9, 6.20, 6.27 and 6.32) include drug screening in law enforcement,[305] forensic and biomedical sciences,[306,307] Olympic[308] and Pan American Games[309] sports and the workplace.[310,311] GC/MS methods are the most prevalent in the literature and are required in the *Mandatory Guidelines for Federal Workplace Drug Testing Programs*.[311] A dated, yet thorough review of GC/MS methods (which include the GC phases employed) covering publications from 1981 to 1991 was documented by Maurer.[312] Published HPLC methods are also numerous and include review articles by Shihabi and McCormick[313] and Turcant and Kohn.[314]

6.18 Enkephalins/Endorphins

Enkephalins are either of two pentapeptides having the formula, H_2N-Tyr-Gly-Gly-Phe-X, where X is leucine or methionine. Endorphins are any of three peptides: the 30 amino acid C-terminal of β-lipotropin (β-endorphin) [19666-76-3], the 16 residues of the N-terminal of β-endorphin (α-endorphin)[59004-96-5], or the 17 residues of the N-terminal of β-endorphin (γ-endorphin) [61512-77-4]. Both enkephalins and endorphins bind to opioid receptors and are neurotransmitters that play a role in pain perception, among other biochemical functions.

HPLC is clearly the method of choice for determination of these peptides. In particular, several LC/MS methods have been developed that provide powerful analytical data.[315–318] Common C_{18} columns are often used in

these analyses. Several reversed phases were compared by Hearn and Aguilar[319] for retention and band broadening of the peptides. Likewise, Purcell et al.,[320] compared a C_4 and C_{18} phase for retention and bandwidth and studied conformational changes of the solutes during chromatographic migration. Conformational changes of β-endorphin on a C_{18} phase are believed to occur depending on mobile-phase organic content, flow rate and temperature.[321,322]

6.19 Glycosides

Glycosides (also see section 6.20) are compounds that contain a carbohydrate molecule, particularly from natural products in plants. Cardiac glycosides occur in certain plants (digitalis, strophanthus, urginea and others) and are biologically active as cardiac muscle contractants. In general, glycosides can be separated on conventional C_{18} HPLC phases.[323,324] GC separations require modification, such as trimethylsilyl derivatization, prior to analysis.[324–326]

6.20 Hormones

Mueller et al.[327] reviewed methods for analyzing performance-enhancing glycoprotein hormones by GC/MS. Some early LC/MS methods for analyzing hormones (also see sections 6.17 and 6.32) were summarized by Vouros and Karger.[328] Other reviews of both LC/MS and GC/MS were published by Shackleton[329] and Bowers[330] for steroids, and anabolic steroids and peptide hormones, respectively.

6.21 Mycotoxins

Separations of mycotoxins such as coumarins are probably most often performed by HPLC.[331] However, Betina[332] listed many mycotoxin methods (462 references) employing both HPLC and GC. Aflatoxins, ochratoxin A [303-47-9], trichothecenes, zearalenone [17924-92-4] and patulin [149-29-1] were among the mycotoxin analyses reviewed by Scott.[333]

6.22 Nitrosamines

Although they are sometimes used as antitumor agent pharmacophores, some nitrosamines are alkylating agents that invoke carcinogenic activity. The deleterious effects of using tobacco goods may partially be due to the presence of nitrosamines in these products.[334,335] The general public became alarmed when analytical chemists began to find potentially dangerous nitrosamines in many unexpected places such as baby bottle rubber nipples and pacifiers,[336,337] fried bacon[338] and protective gloves.[339] Nitrosamine carcinogenocity may occur by methylation of DNA residues, which can be analyzed by HPLC following hydrolysis. Kawasaki et al.[340] used this technique to measure cytosine derivatives on a SCX HPLC phase.

6.23 Nucleosides/Nucleotides/Bases

Under the hydrophobic mechanism of HPLC reversed-phase column retention, the elution order (and also the order of polarity) of the titled constituents of nucleic acids is: nucleotides > nucleosides > bases. Bases and nucleosides[341,342] are generally sufficiently retained and relatively easily separated on C_{18} phases. Some C_{18} phases may also give adequate retention for simple mixtures of nucleotides. The more lipophilic nucleotides may require special polar interactions for retention, such as those available on IBD reversed phases, or standard normal phases (see section 2.5). Alternatively, the chromatographer might take advantage of the negative charge of the phosphate moiety of the nucleotide molecule in choosing an anion-exchange packing to attain sufficient retention.[343] Reviews have recently been published on HPLC separations directly related to these compounds.[344–349]

Nucleotide and nucleoside samples analyzed by GC/MS are often first hydrolyzed to smaller units. Purine nucleosides are easily acid hydrolyzed to the corresponding free base and pentose.[350,351] However, pyrimidine nucleosides are more resistant to acid hydrolysis. Nucleosidases may be considered for pretreatment, since they hydrolyze both types of nucleoside. The base is then derivatized to a more volatile form, usually with a silane reagent. Columns typically contain DB-5 or SPB-5 phases. The final MS analysis step may provide useful information as biological markers for cancer, AIDS (acquired immunodeficiency syndrome), and other clinical applications.[352–354]

6.24 Parabens

Parabens are most often analyzed by reversed-phase HPLC using a C_{18} or C_8 phase. These phenolic esters have two major uses in pharmaceuticals. They are frequently used as internal standards for drug analysis[355–358] and they are present in a variety of formulations as preservatives/excipients. Pharmaceutical preservatives may include methyl,[359,360] ethyl,[359,361] propyl,[359,360] isopropyl,[360] *n*-butyl,[359,360] and/or isobutyl[360] parabens. It is noteworthy that paraben ''preservatives'' might actually participate in certain pharmaceutical degradation pathways, such as interactions with sorbitol [50-70-4] and/or glycerol [56-81-5].[361]

6.25 Peptides

A wide variety of peptide types (also see sections 6.4 and 6.18) and polarities exist and so, too, there is a diversity

of phases that separate them. Review articles summarize choices of reversed-phase, SEC, ion-exchange, hydrophobic interaction, and other phases for peptide separations.[362,363] Reversed-phase packings, specifically, are covered by Kirkland et al.[364] Evidence provided by Thunecke et al.[365] suggests that a C_{18} phase catalyzes isomerization of *cis-trans* peptidyl proline dipeptides. Because of this and other limitations of silica-based packings, nonsilica peptide columns have also been investigated. Certain polymer phases are stable throughout the pH range showing increased selectivity for peptides compared to conventional silica-based reversed-phase packings, but suffer in that they are less pressure stable and may shrink and swell in particular solvents.[366] Polymer phases that contain polar groups are more favorable for peptide and proteins.[367] Alumina (a Lewis acid) modified with C_{18} chemistry forms phases that are too active and are not suitable for peptide separations.[368]

Chiral separations on simple peptides have been accomplished on molecular imprint,[369,370] affinity[371] and brush-type[372,373] columns. CD chiral phases of various sizes (α-, β- or γ-CD) may be chosen depending on the size of the peptide (see section 5.7), as exemplified by Florance and Konteatis.[374] The primary amine residue of peptides can also be taken advantage of (see section 5.7) by performing chiral separations on crown ether columns.[375]

Peptide and protein separations prevalently use trifluoroacetic acid in the mobile phase. Acid exposure to some reversed-phase columns may catalyze hydrolysis of the bonded phase.[376] Therefore, acid-stable columns are particularly important for these classes of compound (see section 2.4). Another packing consideration is the pore size. The packing material should be chosen so that the solute peptide length is smaller than the pore size of the particle. For medium to larger peptides, a 300-Å pore size is optimum.[377] Smaller peptides may be used on the more universal 100-Å or 120-Å pore size packings. Packings of 200-Å pore size are also commercially available. Selection of too large a pore size packing is not necessarily desirable, since larger pore particles contain less surface area and, hence, display less retention. When performing calculations for optimum pore size, it is important not to overlook that the manufacturer's listing of the pore size is for the bare particle before bonding. A catalogued 120-Å C_{18} phase actually has an effective pore size somewhat less than that, since the bonded C_{18} molecule occupies space within the pore.

GC/MS peptide analysis involves hydrolyzing the peptide and then derivatizing the individual amino acids or fragments before injection. Columns include DB-1,[378] BP-5,[379] and PS255.[380] One particularly interesting publication by Cano[381] used this general procedure to characterize 2- to 135-million-year-old bacteria. Study of ancient bacteria might lead to production of novel antibiotic, fungicidal, etc., proteins and peptides.

6.26 Pharmaceutical Processing Solvent Residues

Granules, tablets, film coatings and plastic sealants used for pharmaceutical drugs may contain harmful solvent residues. Analytical monitoring of these solvents is accomplished by headspace GC analysis using phases such as methyl silicone-coated columns. Foust and Bergren[382] developed a separation method for 27 solvents which required just 20 min.

6.27 Phenylalkylamines

Phenylalkylamines (also see section 6.17) are important structures in pharmaceutical drug design and analysis. The solute's basic residue is subject to active silanol interaction resulting in peak tailing. Active C_{18} phases typically show tailing, even when low-pH mobile phases are used in the attempt to protonate the active silanols.[383,384] Similarly, Murayama et al.[385] found that a titania C_{18} phase is very active for phenylalkylamines. Base deactivated reversed-phase HPLC packings are the best choice for these solutes (see section 5.2).

Maurer[386] performed GC/MS analysis on some derivatized amphetamine designer drugs using a methylsilicone-coated column. For chiral separations of primary amine phenylalkylamines, chiral crown ether columns (see section 5.7) may be utilized.[387]

6.28 Porphyrins

Naturally occurring cytochromes contain iron–porphyrin groups involved in electron-transfer systems or hydroxylation reactions. Hematoporphyrin compounds, which are tumor localizers, have been separated by HPLC on C_{18}[388] or C_8[389] phases with UV detection at 360–400 nm. A bare silica normal-phase column has also been used by Dellinger and Brault.[390] Conventional C_{18} phases were utilized in separating other porphyrin compounds including porphyrin c [68045-92-1] and *N*,*N*′-diacetylporphyrin c [31444-63-0],[391] and porphyrin ether glycerides.[392] Tin protoporphyrin [14325-05-4] and other metalloporphyrins were separated on a DVB phase.[393]

6.29 Procainamides

Antiarrhythmic procainamides are separated on a variety of reversed-phase columns including C_{18},[394,395] hexyl,[396] cyano,[397,398] and phenyl[399,400] packings. Gisch et al.[401] found a *t*-butyl phase gave better peak shapes than an *n*-butyl phase, presumably owing to improved shielding of undesirable silanol interactions by the former bulky substituent. Although procainamides

are not as problematic regarding tailing as are tricyclic antidepressants, they still are sensitive enough to be useful as a first test for packing silanol activity. Several column manufacturers have adopted these solutes in quality control analysis of their bonded phases. Procainamides are also used as internal standards in many HPLC analyses including ranitidine [66357-35-5][402] and propranolol [318-98-9].[403]

6.30 Profens

The profens (also see section 6.5) are readily separated on common C_{18} phases.[404–406] Profens may display good peak shape on active type A packings since they are acidic in nature, but type B phases also perform well (see section 5.1). Enantiomer separations of profens is well studied, including isolations on brush-type,[407] cellulose,[408,409] L-leucinamide,[410] OVM,[411] AGP,[412] ergot alkaloid,[413] avidin[414] and CD[415] phases.

6.31 Proteins

The same generalities associated with the chromatography of large peptides (see section 6.25) also hold true for proteins. Reviews have been published on the choice of phase and mode of separation for protein separations, including reversed-phase, hydrophobic interaction, ion-exchange, gel permeation, bioaffinity, and other types.[416–419] More polar stationary phases are desirable in cases where retention of protein integrity and function is needed. Reversed phases denature proteins, thereby destroying their biological activity. A hydroxyethyl methacrylate phase by Coupek and Vins[420] exemplifies a ''protein-friendly'' phase.

Pore sizes become an important factor in selecting phases for larger protein analytes. The pore should be adequately large enough to accommodate the solute. Packings of 300 Å are of sufficient size for most applications. For proteins approaching a molecular weight of about 500 000 Da or more, a 500-Å phase may be the best choice.

Proteins are somewhat unique in that there is no typical adsorption–desorption mechanism normally occurring in reversed-phase separations. Instead, the proteins apparently stick to the packing material until a ''magical'' ratio of % aqueous : % organic mobile phase composition is reached under gradient conditions, which releases the proteins from the phase. Another particular trait of some peptides and proteins is that they may adhere to SS column and frit surfaces. For this reason, some chromatographers prefer to run protein HPLC separations using all PEEK® connective tubing and column hardware. More extensive studies in this area are needed to quantitate protein interaction with various material types.

6.32 Steroids

Both reversed-phase and normal-phase columns may be used for analyzing steroids (also see sections 6.4, 6.5, 6.17 and 6.20). The cyano phase works by either mechanism, depending on the mobile phase employed. It gives excellent separations, but is often overshadowed by the use of the more familiar C_{18} phase. Reversed-phase methods exist for anabolic steroids,[421,422] corticosteroids,[423–427] bile-acid steroids,[428] cholesterol steroids,[429] phytoecdysteroids,[430,431] brassinosteroids[432,433] and oxysterols.[434] An interesting property of steroids in chromatographic separations is that they tend to show improved separations at lower temperatures, akin to chiral separations.[435]

Low-volatility steroids can be analyzed by GC/MS,[436] including the ecdysteroids.[437] Steroid profiling is another good application for GC/MS analysis as reviewed by Honour.[438]

6.33 Sulfur Drugs

Reversed-phase C_{18} HPLC packings are most often used for analysis of sulfur compounds related to pharmaceutical applications. The antibacterial sulfa drug, sulfadoxine [2447-57-6], was measured in plasma by Astier et al.[439] and Kohl et al.[440] determined disulfiram [97-77-8] and sulfamethoxazole [723-46-6] in urine. Sulfatrim [723-46-6] dissolution[441] and elemental sulfur [7704-34-9] in pharmaceutical preparations[442] testings were also conducted on C_{18} phases. Dangi et al.[443] measured sulfur mustard [505-60-2] in blood samples on a C_{18} phase with UV detection at 200 nm.

Consumer concern over the practice of routinely adding antibiotics to animal feeds has led to the development of methods to analyze these additives in food products. Houglum et al.[444] determined sulfamethazine [57-68-1] in animal feeds with a C_{18} phase. The additives may become incorporated in the meat products, which also is desirable to analyze for health reasons. Thus, Milner et al.[445] resolved sulfadimethoxine [122-11-2] in fish fillets. Sulphadiazine [68-35-9], sulphadimidine [57-68-1], sulphadimethoxine [122-11-2], sulfadoxine [2447-57-6], sulfamerazine [127-79-7], sulfamethoxazole, sulfamonomethoxine [1220-83-3], sulfaquinoxaline [59-40-5], sulfathiazole [72-14-0] and sulfasomidine [515-64-0] were measured in beef, chicken and pork meats.[446] Patthy[447] analyzed sulfaquinoxaline in chicken meat and liver.

Chiral separations for *N*-aryl thiazolinones and their derivatives were run on a brush-type phase.[448] A dimethylsilicone phase was used for GC/MS analysis of anticarcinogen components in garlic extract.[449]

6.34 Taxol and Related Compounds

Taxol [33069-62-4] and related compounds are under intense study for their antitumor and antileukemia activities. These natural products are extracted from yew trees for experimentation and drug development. Chromatographic analyses have been developed for isolating the products from the yew bark, leaves and/or stems, using C_{18} or C_8,[450–453] cyano,[454–456] phenyl,[456–458] porous graphitic carbon,[459–461] PFP,[98,462] and special phases designed especially for taxol separations.[463,464] Limited supplies of yew trees prompted research into growing cell cultures to be harvested for their taxol production. Analytical and preparative methods have been developed for this purpose.[465–468] Methods have also been developed for monitoring taxol levels in experimental mouse plasma models[469–471] and human biological fluids and tissues.[472–482]

6.35 Terpines

In general, GC/MS is currently the leading technique in analyzing terpines (also see section 6.4). For example, the naturally occurring α-terpineol [98-55-5] is an antiseptic drug found in petitgrain oil, long leaf pine oil, cajeput oil, gentian roots and other sources. It has been analyzed on PEG phases,[483,484] and a cholesteryl acetate phase.[485] Separation of the enantiomers has been investigated on substituted CD phases.[483,486,487]

6.36 Tetracyclines

Comparisons were run between several C_{18} and C_8 phases for the assay of tetracycline hydrochloride. Under various conditions the C_8 phases performed as well[488] or better[489] than the C_{18} phases. Polymer phases appear to offer advantages over even base deactivated C_{18} phases in peak shape, separations and recoveries of oxytetracycline [6153-64-6], tetracycline [60-54-8] and chlortetracycline [64-72-2].[490] Although tetracycline is not a particularly large molecule, 1000-Å pore polymer particles seem to be optimum for this solute.[491,492]

In addition to urine[493] and serum[494] analyses, tetracycline is measured in meat, milk and cheese owing to health concerns.[495] An HPLC method was also developed for this drug in discolored teeth, which was found to be caused by incorporation of minocycline [10118-90-8] into the dentine and enamel.[496]

6.37 Tricyclic Antidepressants

Drugs within the class of tricyclic antidepressants represent some of the most challenging compounds to obtain good symmetry in HPLC separations. They are relatively strong bases that can have undesirable potent interactions with active silanols. Type B reversed-phase silica packings are highly recommended for these compounds. Most literature separations using C_{18} phases, whether they are deactivated or not, use ion-pairing[497,498] or ion-suppression,[499,500] additives, and/or acidic mobile phases.[501,502] Acidic mobile phases with cyano phases often provide better peak shape than hydrocarbon phases. This is reflected by the appearance in the literature of several publications using cyano columns.[503–508] Mixed phases have also been used to improve peak shapes of tricyclic antidepressants.[509] An approach to choosing the best stationary phase is described by Kirkland and Henderson.[510] Some direct chiral separations of tricyclic compounds were performed on an AGP (see section 5.7) phase.[511]

Separations by GC are not as prevalent as HPLC methods. Columns used include DB-1,[512] OV-1 and OV-25[513] and SPB-1.[514]

7 ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS AND UNITED STATES PHARMACOPEIA METHODS

Appointed volunteers from all over the world perform collaborative validation work in providing the AOAC (Association of Official Analytical Chemists) Official Methods. The studies are designed to establish accuracy, precision, sensitivity, range, specificity, ruggedness, limit of measurement, practicality and other parameters of methods under typical laboratory conditions. The methods are used in government regulations, various commercial product analyses, legal proceedings, and in requirements for some national and international trades. Collaborative study reports for first action methods are published in the *Journal of AOAC International*, while both first and final action methods are published in the compendium, *Official Methods of Analysis*.

The United States Pharmacopeia (USP) provide similar services in developing methods. Regarding bonded phases, a USP method may specify a particular packing type. A list of some USP column classifications is shown in Table 4.

Other assistance available to the chromatographer includes a large selection of applications available from the various commercial column manufacturers. Although most applications are run with standards rather than ''real-life'' samples, the general separations and retention characteristics furnish a good starting point for column selection and chromatographic conditions. Many chromatography column companies offer technical support to assist in the decision making, albeit somewhat biased, naturally, towards the manufacturer's products. The largest convention, by far, for HPLC and GC column manufacturers is the yearly Pittsburgh Conference normally

Table 4 USP column coding system

USP column designation	Description	USP column designation	Description
L1	C_{18} chemically bonded to porous 3–10-μm silica or ceramic microparticles	L18	5–10-μm porous silica bonded with cyano and amino groups
L2	C_{18} chemically bonded to silica gel with a controlled surface porosity bonded onto 30–50-μm solid spherical core	L19	9-μm SCX resin composed of sulfonated cross-linked styrene–divinylbenzene copolymer in the calcium form
L3	5–10-μm porous silica	L20	5–10-μm porous silica bonded with dihydroxypropane groups
L4	Silica gel with a controlled surface porosity bonded onto 30–50-μm solid spherical core	L21	5–10-μm rigid, spherical styrene divinylbenzene copolymer
L5	Alumina with a controlled surface porosity bonded onto 30–50-μm solid spherical core	L22	A 10-μm cation-exchange resin made of porous polystyrene gel with sulfonic acid groups
L6	Sulfonated fluorocarbon polymer which acts as a strong cation exchanger coated on 30–50-μm solid spherical core	L23	A 10-μm ion-exchange resin made of porous polymethacrylate or polyacrylate gel with quaternary ammonium groups
L7	C_8 chemically bonded to porous 5–10-μm silica or ceramic microparticles	L24	A semi-rigid hydrophilic gel consisting of 32–63-μm vinyl polymers with numerous hydroxy groups on the matrix surface
L8	Aminopropylsilane chemically bonded to porous 10-μm silica or ceramic microparticles	L25	A packing with the ability to separate compounds with a molecular weight range from 100 to 5000, especially neutral, anionic and cationic water-soluble polymers
L9	Irregular, 10-μm porous silica bonded with a strongly acidic cation exchanger	L26	5–10-μm porous silica bonded with butyl groups
L10	5–10-μm porous silica bonded with nitrile groups	L27	30–50-μm porous silica
L11	5–10-μm phenyl bonded porous silica	L28	High-purity, 100-Å spherical silica bonded with an anionic function (amine) as well as a conventional reversed-phase C_8 functionality
L13	5–10-μm porous silica bonded with trimethylsilane	L29	Spherical, 5 mm, 80-Å γ-alumina coated with polybutadiene used for reversed-phase chromatography
L14	10-μm silica bonded with a strong basic quaternary ammonium anion exchanger	L30	3–10-μm porous silica bonded with ethyl groups
L15	3–10-μm hexyl bonded porous silica	L31	Latex particles bonded with a quaternary amine attached to a core of 8.5-μm macroporous particles with 2000-Å pores and composed of ethylvinylbenzene cross-linked with 55% divinyl benzene which functions as a strong anion exchanger
L16	5–10-μm porous silica bonded with dimethylsilane	L32	Irregular, 5–10-μm silica bonded with an L-proline copper complex which acts as a chiral ligand exchange packing
L17	7–11-μm SCX resin composed of sulfonated cross-linked styrene–divinylbenzene copolymer in the hydrogen form	L33	A specially processed, pH stable, spherical silica-based packing capable of separating proteins with a molecular weight from 4000 to 400 000
		L34	9-μm SCX resin composed of sulfonated cross-linked styrene–divinylbenzene copolymer in the lead form

held in March, where chromatographers may scout and learn about the newest developments in column phase technology.

ACKNOWLEDGMENTS

The author would like to thank Stephanie Risbon and C. Vernon Bartlett for assistance in preparing this manuscript, and Kent Rauch for his aid in preparing figures. This work is dedicated to my Father.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
ACE	Angiotensin-converting Enzyme
AGP	α_1-Glycoprotein
AIDS	Acquired Immunodeficiency Syndrome
AOAC	Association of Official Analytical Chemists
ASTM	American Society for Testing and Materials
BaP	Benzo[*a*]pyrene
BSA	Bovine Serum Albumin
CD	Cyclodextrin
cGMP	Continuous Good Management Practices
CLEC	Chiral Ligand Exchange Chromatography
CMPA	Chiral Mobile Phase Additive
CSP	Chiral Stationary Phase
DNBPG	Dinitrobenzoylphenylglycine
DVB	Divinylbenzene Polymer
DZ	Dual Zone
EDTA	Ethylenediaminetetraacetic Acid
FAME	Fatty Acid Methyl Ester
FDA	Food and Drug Administration
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/EC	Gas Chromatography/Electron Capture
GC/FID	Gas Chromatography/Flame Ionization Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GFF	Glycerylpropyl/glycine-L-phenylalanine-L-phenylalanine
GMP	Good Management Practice
HPLC	High-performance Liquid Chromatography
HPLC/UV	High-performance Liquid Chromatography/Ultraviolet
HSA	Human Serum Albumin
IAM	Immobilized Artificial Membrane
IBD	Intrinsically Base Deactivated
ID	Internal Diameter
ISRP	Internal Surface Reversed Phase
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/NMR	Liquid Chromatography/Nuclear Magnetic Resonance
MFP	Mixed Functional Phase
MS	Mass Spectrometry
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
ODS	Octadecylsilyl
OVM	Ovomucoid
PAH	Polycyclic Aromatic Hydrocarbon
PEEK®	Polyetheretherketone
PEG	Polyethylene Glycol
PFP	Pentafluorophenyl
PGC	Porous Glassy Carbon
PhPh	Phenanthro[3,4-*c*]phenanthrene
PMM	Polymethylmethacrylate Polymer
PS/DVB	Polystyrene/Divinylbenzene Copolymer
PVA	Polyvinylalcohol Polymer
QA/QC	Quality Assurance/Quality Control
QSRR	Quantitative Structure–Retention Relationships
RMS	Root Mean Square
SAX	Strong Anion-exchange
SCX	Strong Cation-exchange
SEC	Size Exclusion Chromatography
SHP	Shielded Hydrophobic Phase
SOP	Safe Operating Procedures
SPE	Solid-phase Extraction
SPS	Semipermeable Surface
SRM	Standard Reference Material
SS	Stainless Steel
TBN	1,2-3,4-5,6-7,8-Tetrabenzonaphthalene
TDM	Therapeutic Drug Monitoring
USP	United States Pharmacopeia
UV	Ultraviolet
UV_{max}	Ultraviolet Maximum
UV/VIS	Ultraviolet/Visible
WAX	Weak Anion-exchange
WCX	Weak Cation-exchange

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Proteins and Peptides Purification in Pharmaceuticals Analysis • Steroid Analysis

Gas Chromatography **(Volume 12)**
Column Technology in Gas Chromatography

Liquid Chromatography **(Volume 13)**
Affinity Chromatography • Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography

REFERENCES

1. W.F. Smyth, E.D. Chabala, 'The Determination of Selected Organic Substances of Significance in a Developing Country', *Fresenius' J. Anal. Chem.*, **345**(11), 701–704 (1993).
2. K. Vekey, D. Edwards, L.F. Zerilli, 'Liquid Chromatography–Mass Spectrometry Coupling and its Application in Pharmaceutical Research', *J. Chromatogr.*, **488**, 73–85 (1989).
3. L.D. Bowers, 'High-performance Liquid Chromatography/Mass Spectroscopy: State of the Art for the Drug Analysis Laboratory', *Clin. Chem.*, **35**(7), 1282–1287 (1989).
4. E. Brewer, J. Henion, 'Atmospheric Pressure Ionization LC/MS/MS Techniques for Drug Disposition Studies', *J. Pharm. Sci.*, **87**(4), 395–402 (1998).
5. S.A. Korhammer, A. Bernreuther, 'Hyphenation of High-performance Liquid Chromatography (HPLC) and Other Chromatographic Techniques (SFC, GPC, GC, CE) with Nuclear Magnetic Resonance (NMR): A Review', *Fresenius' J. Anal. Chem.*, **354**(2), 131–135 (1996).
6. E. Gelpi, 'Biomedical and Biochemical Applications of Liquid Chromatography–Mass Spectrometry', *J. Chromatogr., A*, **703**, 1–2, 59–80 (1995).
7. G. Wilkinson, 'World-wide Analytical Instrument Markets', *Anal. Instrument Industry Reporter*, **13**(15), (1997).
8. G. Wilkinson, 'Analytical Instrument Markets – Challenges and Opportunities in ''Post-adolescence'' ', *Anal. Instrument Industry Reporter*, **14**(1), (1997).
9. G. Wilkinson, 'Analytical Instrument End-user Survey', *Anal. Instrument Industry Reporter*, **14**(14), (1997).
10. 'News & Notes', *Pharmaceutical Cosmetic Quality*, 7–8, Nov/Dec 1997.
11. T.H. Gouw, R.E. Jentoft, 'Practical Aspects in Supercritical Fluid Chromatography', *Adv. Chromatogr.*, **13**, 1–40 (1975).
12. S. Ahwja, 'Chemical Derivatization for the Liquid Chromatography of Compounds of Pharmaceutical Interest', *J. Chromatogr. Sci.*, **17**, 168–172 (1979).
13. N.D. Danielson, M.A. Targove, B.E. Miller, 'Pre- and Postcolumn Derivatization Chemistry in Conjunction with HPLC for Pharmaceutical Analysis', *J. Chromatogr. Sci.*, **26**, 362–371 (1988).
14. B.S. Kersten, 'Drug Matrix Effect on the Determination of Residual Solvents in Bulk Pharmaceuticals by Wide-bore Capillary Gas Chromatography', *J. Chromatogr. Sci.*, **30**(4), 115–119 (1992).
15. J. Pohjola, M. Harpf, 'Determination of Atropine and Obidoxime in Automatic Injection Devices Used as Antidotes Against Nerve Agent Intoxication', *J. Chromatogr., A*, **686**(2), 350–354 (1994).
16. S. Scalia, U. Cova, M. Foganolo, S. Landi, A. Medici, 'Determination of Free Bile Acids in Raw Materials and Bulk Products by HPLC and GC', *Anal. Lett.*, **27**(9), 1789–1804 (1994).
17. P. Guinebault, A. Haddouche, G. Bianchetti, 'High-performance Liquid and Gas Chromatographic Methods for the Determination of Ciclprolol', *J. Chromatogr.*, **525**, 359–367 (1990).
18. A.C. Mehta, 'The How and Why of Therapeutic Drug Monitoring', *Anal. Proc.*, **32**(8), 347–351 (1995).
19. H. Röper, 'Direkte Qualitative Bestimmung von Nitrosaminen aus Wässeriger Lösung durch Gaschromatographie oder Hochdruckflüssigkeitschromatographie', *J. Chromatogr.*, **166**, 305–309 (1978).
20. Z. Yi, P.R. Brown, 'Chromatographic Methods for the Analysis of Basic Neurotransmitters and Their Acidic Metabolites', *Biomed. Chromatogr.*, **5**(3), 101–107 (1991).
21. R.E. Majors, 'Twenty-five Years of HPLC Column Development: A Commercial Perspective', *LC-GC*, **12**(7), 508–518 (1994).
22. K.J. Duff, R.C. Ludwig, 'The Packing and Evaluation of Small-particle Preparative Columns', *Am. Lab.*, **26**(5), 32KK–32MM (1994).
23. G.E. Berendsen, L. de Galan, 'Preparation and Chromatographic Properties of Some Chemically Bonded Phases for Reversed-phase Liquid Chromatography', *J. Liq. Chromatogr.*, **1**(5), 561–586 (1978).
24. J.J. Kirkland, J.W. Henderson, J.J. Destefano, M.A. Van Straten, H.A. Claessens, 'Stability of Silica-based, End-capped Columns with pH 7 and 11 Mobile Phases for Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr., A*, **762**(1–2), 97–112 (1997).
25. J.L. Glajch, J.J. Kirkland, 'Stable, Sterically Protected, Monofunctional-silane Bonded-phase Columns for High-performance Liquid Chromatography', *LC-GC*, **8**(2), 140–150 (1990).
26. R.C. Ludwig, T.S. Reid, R.A. Henry, K.J. Duff, 'Stability of High-ligand-density HPLC Bonded-phase Silica Media at Low pH', *LC-GC*, **16**(2), 158–168 (1998).

27. J.J. Kirkland, M.A. Van Straten, H.A. Claessens, 'Reversed-phase High-performance Liquid Chromatography of Basic Compounds at pH 11 with Silica-based Column Packings', *J. Chromatogr.*, **797**(1–2), 111–120 (1998).
28. K.K. Unger, *Porous Silica, Its Properties and Use As Support in Column Liquid Chromatography*, Elsevier Scientific, New York, 83–96, 1979.
29. H.O. House, *Modern Synthetic Reactions*, 2nd edition, Georgia Institute of Technology, Atlanta, GA, 552–553, 1964.
30. B.S. Welinder, T.Kornfelt, H.H. Sorensen, 'Stationary Phases, the Achilles Heel in HPLC of Proteins', *Anal. Chem.*, **67**(1), 39A–43A (1995).
31. M. De Smet, D.L. Massart, 'High-performance Liquid Chromatography of Basic Drugs on Cyanopropyl-bonded Phases: Practical Aspects', *Trends Anal. Chem.*, **8**(3), 96–101 (1989).
32. K.J. Duff, 'Improved HPLC Separations Using Intrinsically Base-deactivated Columns', *Am. Lab.*, **30**(4), 32AA–32JJ (1998).
33. K. Azzoui, M. Lafosse, S. Lazar, V. Thiery, L. Morin-Allory, 'Separation of Benzodioxinic Isomers in LC: A Molecular Modelling Approach for the Choice of the Stationary Phase', *J. Liq. Chromatogr.*, **18**(15), 3021–3034 (1995).
34. R. Kaliszan, *Quantitative Structure – Chromatographic Retention Relationships*, J. Wiley & Sons, New York, 1987.
35. R. Kaliszan, 'Quantitative Structure–Retention Relationships', *Anal. Chem.*, **64**(11), 619A–631A (1992).
36. R. Kaliszan, 'Quantitative Relationships Between Molecular-structure and Chromatographic Retention – Implications in Physical, Analytical and Medicinal Chemistry', *CRC Crit. Rev. Anal. Chem.*, **16**(4), 323–383 (1986).
37. R. Kaliszan, in *High-performance Liquid Chromatography*, eds. P.R. Brown, R.A. Hartwick, J. Wiley & Sons, New York, Chapter 14, 1989.
38. L.C. Sander, S.A. Wise, 'Influence of Stationary Phase Chemistry on Shape Recognition in Liquid Chromatography', *Anal. Chem.*, **67**(18), 3284–3292 (1995).
39. R. Kaliszan, K. Osmialowski, 'Correlation Between Chemical Structure of Non-cogeneric Solutes and Their Retention on Polybutadiene-coated Alumina', *J. Chromatogr.*, **506**, 3–16 (1990).
40. K.J. Duff, J.R. Miller, 'A Novel Column for Difficult HPLC Separations of a Broad Spectrum of Analytes', Poster, The Pittsburgh Conference, Chicago, IL, 1996.
41. J. Nawrocki, 'The Silanol Group and its Role in Liquid Chromatography', *J. Chromatogr., A*, **779**(1–2), 29–71 (1997).
42. G.B. Cox, 'The Influence of Silica Structure on Reversed-phase Retention', *J. Chromatogr., A*, **656**(1–2), 353–367 (1993).
43. L.C. Sander, S.A. Wise, 'Synthesis and Characterization of Polymeric C_{18} Stationary Phases for Liquid Chromatography', *Anal. Chem.*, **56**, 504–510 (1984).
44. J.H. Knox, B. Kaur, G.R. Millward, 'Structure and Performance of Porous Graphitic Carbon in Liquid Chromatography', *J. Chromatogr.*, **352**, 3–25 (1986).
45. A. Karlsson, C. Pettersson, 'Enantiomeric Separation of Amines Using *N*-Benzoxycarbonylglycyl-L-proline as Chiral Additive and Porous Graphitic Carbon as Solid Phase', *J. Chromatogr.*, **543**(2), 287–298 (1991).
46. C.E. Dalgleish, 'The Optical Resolution of Aromatic Amino Acids on Paper Chromatograms', *J. Chem. Soc.*, 3940–3942 (1952).
47. M. Balcan, T. Cserhati, E. Forgacs, 'Influence of Organic Modifiers on the Retention Characteristics of Graphitized Carbon Columns', *Anal. Lett.*, **30**(4), 883–893 (1997).
48. T.M. Engel, S.V. Olesik, M.R. Callstrom, M. Diener, 'Liquid Chromatography Based on a Low-temperature Method of Producing Glassy-carbon', *Anal. Chem.*, **65**(24), 3691–3700 (1993).
49. M.P. Rigney, E.F. Funkenbusch, P.W. Carr, 'Physical and Chemical Characterization of Microporous Zirconia', *J. Chromatogr.*, **499**, 291–304 (1990).
50. K.B. Holland, J.M. Washington, D.C. Moe, C.M. Conroy, 'Alumina-based Stationary Phases for HPLC Applications', *Am. Lab.*, 51–62 February (1992).
51. H. Lingeman, W.J.M. Underberg, 'Non-modified Stationary Phases for the Analysis of Basic Compounds', *Trends Anal. Chem.*, **7**(9), 346–351 (1988).
52. Z. Yu, D. Westerlund, 'Ion-pair Chromatography of Methotrexate in a Column-switching System Using an Alkyl-diol Silica Precolumn for Direct Injection of Plasma', *J. Chromatogr., A*, **742**, 113–120 (1996).
53. G.S. Clarke, M.L. Robinson, 'Automated Determination of Drugs in Biological Samples Using HPLC', *Anal. Proc.*, **22**, 137–138 (1985).
54. P. Campins-Falco, R. Herraez-Hernandez, A. Sevillano-Cabeza, 'Column-switching Techniques for High-performance Liquid Chromatography of Drugs in Biological Samples', *J. Chromatogr.*, **619**, 177–190 (1993).
55. E. Doyle, R.D. McDowall, G.S. Murkitt, V.S. Picot, S.J. Rogers, 'Automated Sample Handling for Drug Analysis from Biological Fluids Using HPLC', *Anal. Proc.*, **26**, 294–295 August (1989).
56. J.W. Veals, C.C. Lin, 'Column-switching HPLC Methods for Drug Analysis in Biological Fluids', *Am. Lab.*, 42–47 April (1988).
57. T.C. Pinkerton, I.H. Hagestam, 'Chromatographic Method and Means', US Patent 4544485, 1985.
58. T. Mohammad, H. Morrison, 'Simultaneous Determination of Methylene Violet, Halogenated Methylene Violet and Their Photoproducts in the Presence of DNA by High-performance Liquid Chromatography Using an Internal Surface Reversed-phase Column', *J. Chromatogr. B, Biomed. Appl.*, **704**(1–2), 265–275 (1997).
59. J.A. Perry, 'The Internal Surface Reversed Phase. Concepts and Applications', *J. Liq. Chromatogr.*, **13**(6), 1047–1074 (1990).

60. Z.K. Shihabi, 'Review of Drug Analysis with Direct Serum Injection on the HPLC Column', *J. Liq. Chromatogr.*, **11**(8), 1579–1593 (1988).
61. J. Haginaka, 'Drug Determination in Serum by Liquid Chromatography with Restricted Access Stationary Phases', *Trends Anal. Chem.*, **10**(1), 17–22 (1991).
62. C.G. Honegger, W. Krenger, H. Langemann, A. Kempf, 'Automated High-performance Liquid Chromatographic Method with Column Switching for the Determination of Neurotransmitters and Related Compounds, Ascorbic Acid and Uric Acid in Tissue Extracts', *J. Chromatogr., Biomed. Appl.*, **54**, 2 (1986).
63. C.A. Chang, Q. Wu, L. Tan, 'Normal Phase High-performance Liquid Chromatographic Separations of Positional Isomers of Substituted Benzoic Acids with Amine and β-Cyclodextrin Bonded-phase Columns', *J. Chromatogr.*, **361**, 199–207 (1986).
64. A.I. Mallet, R.M. Barr, J.A. Newton, 'Analysis of Dihydroxyeicosatetraenoic Acids by Gas Chromatography–Mass Spectroscopy', *J. Chromatogr., Biomed. Appl.*, **51**, 1 (1986).
65. J. Xia, P.J. Gilmer, 'Organic Modifiers in the Anion-exchange Chromatographic Separation of Sialic Acids', *J. Chromatogr., A*, **676**(1), 203–208 (1994).
66. G. Schmucker, 'High-performance Liquid Ion-exchange Chromatography', *J. Liq. Chromatogr.*, **10**(8–9), 1887–1901 (1987).
67. S. Bulusu, G.A. Mills, V. Walker, 'Analysis of Organic Acids in Physiological Fluids by High-performance Liquid Chromatography', *J. Liq. Chromatogr.*, **14**(9), 1757–1777 (1991).
68. M. Tischler, J.H. Cardellina II, 'Simple Method for the Separation of Triterpene Carboxylic Acids', *J. Liq. Chromatogr.*, **16**(1), 15–19 (1993).
69. G. Guinn, D.L. Brummett, R.C. Beier, 'Purification and Measurement of Abscisic Acid and Indol-3-ylacetic Acid by High-performance Liquid Chromatography', *Plant Physiol.*, **81**(4), 997–1002 (1986).
70. T.C. Kwong, 'Analysis of Acetylsalicylic Acid (Aspirin) and its Metabolites by Liquid Chromatography', *J. Liquid Chromatogr.*, **10**(2–3), 305–321 (1987).
71. A.G. Netting, 'Number of Double Bonds in Fatty Acids From Fats and Oils by HPLC Using Pentafluorophenyl Esters', *J. Am. Oil Chem. Soc.*, **63**(9), 1197–1199 (1986).
72. J.H. Kim, K.S. Choi, S.J. Lee, 'Studies on the Separation and Quantitation Analysis of γ-Linolenic ((*Z*,*Z*,*Z*)-Octadeca-6,9,12-trienoic) Acid From *Oenothera Jacquin* (Evening Primrose)', *Taehan Hwahakhoe Chi*, **30**(4), 403–405 (1986).
73. J.-T. Lin, T.A. McKeon, A.E. Stafford, 'Gradient Reversed-phase High-performance Liquid Chromatography of Saturated, Unsaturated and Oxygenated Free Fatty Acids and Their Methyl Esters', *J. Chromatogr., A*, **699**(1–2), 85–91 (1995).
74. A.B. Kroumova, G.J. Wagner, 'Methods for Separation of Free, Short, Medium, and Long Chain Fatty Acids and for Their Decarboxylation', *Anal. Biochem.*, **225**(2), 270–276 (1995).
75. M.S. Rao, K. Hidajat, C.B. Ching, 'Reversed-phase HPLC: The Separation Method for the Characterization and Purification of Long Chain Fatty Acids – A Review', *J. Chromatogr. Sci.*, **33**(1), 9–21 (1995).
76. S. Antonopoulou, N.K. Andrikopoulos, C.A. Demopoulas, 'Separation of the Main Neutral Lipids into Classes and Species by PR-HPLC and UV Detection', *J. Liq. Chromatogr.*, **17**(3), 633–648 (1994).
77. H.M. Liebich, 'Sample Preparation for Organic Acids in Biological Fluids', *Anal. Chim. Acta*, **236**(1), 121–130 (1990).
78. P.S. Mukherjee, H.T. Karnes, 'Ultra-violet and Fluorescence Derivatization Reagents for Carboxylic Acids Suitable for High-performance Liquid Chromatography: A Review', *Biomed. Chromatogr.*, **10**(5), 193–204 (1996).
79. M. Thomassin, E. Cavalli, Y. Guillaume, C. Guinchard, 'Comparison of Quantitative High-performance Thin-layer Chromatography and the High-performance Liquid Chromatography of Parabens', *J. Pharm. Biomed. Anal.*, **15**(6), 831–838 (1997).
80. O. Cabaleiro, R. Cela, 'Computer Aided Optimization of the Gradient Separation of Complex Mixtures of Polyphenolic Acids by Microcolumn Liquid Chromatography', *J. Microcolumn Sep.*, **8**(4), 231–247 (1996).
81. A. Vazquez, F. Perez, R. Cela, 'HPLC Retention Behavior of 24 Polyphenolic Acids on Porous Graphitized Columns', *Quim. Anal. (Barcelona)*, **14**(2), 77–83 (1995).
82. S. Scalia, 'Bile Acid Separation', *J. Chromatogr., B, Biomed. Appl.*, **671**(1–2), 299–317 (1995).
83. S.H. Hansen, P. Helboe, M. Thomsen, 'Separation of Basic Drugs by High-performance Liquid Chromatography Using Dynamically Modified Silica', *Trends Anal. Chem. (Pers. Ed.)*, **7**(10), 389–393 (1988).
84. D.W. Hill, A.J. Kind, 'The Effects of Type B Silica and Triethylamine on the Retention of Drugs in Silica Based Reverse Phase High-performance Chromatography', *J. Liq. Chromatogr.*, **16**(18), 3941–3964 (1993).
85. H. Lingeman, W.J.M. Underberg, 'Non-modified Stationary Phases for the Analysis of Basic Compounds', *Trends Anal. Chem. (Pers. Ed.)*, **7**(9), 346–351 (1988).
86. V. Das Gupa, M. Mathew, 'Effect of Mobile Phase pH on the Separation of Drugs Using High-performance Liquid Chromatography', *Drug Dev. Ind. Pharm.*, **21**(7), 833–837 (1995).
87. I.M. Mutton, 'Application of a Comparative Evaluation of Several Reversed-phase Columns to the Automated Analysis of Candidate Pharmaceuticals', *J. Chromatogr., A*, **697**(1–2), 191–201 (1995).
88. P.L. Zhu, J.W. Dolan, L.R. Snyder, D.W. Hill, L. Van Heukelem, T.J. Waeghe, 'Combined Use of Temperature and Solvent Strength in Reversed-phase Gradient Elution. III. Selectivity for Ionizable Samples as a

Function of Sample Type and pH', *J. Chromatogr., A*, **756**(1–2), 51–62 (1996).
89. L.R. Snyder, 'Changing Reversed-phase High-performance Liquid Chromatography Selectivity. Which Variables Should Be Tried First?', *J. Chromatogr., B*, **689**, 105–115 (1997).
90. L.R. Snyder, M.A. Quarry, J.L. Glajch, 'Solvent-strength Selectivity in Reversed-phase HPLC', *Chromatographia*, **24**, 33–44 (1987).
91. L.W. Yu, T.R. Floyd, R.A. Hartwick, 'Synthesis and Characterization of Chemically Bonded Silica-based Zwitterion Exchangers for HPLC', *J. Chromatogr. Sci.*, **24**(5), 177–182 (1986).
92. L.W. Yu, R.A. Hartwick, 'Zwitterionic Stationary Phases in HPLC', *J. Chromatogr. Sci.*, **27**(4), 176–185 (1989).
93. T.-Y. Chou, M.-H. Yang, 'Multifunctional Zwitterion-exchange Stationary Phase for HPLC', *J. Liq. Chromatogr. Relat. Technol.*, **19**(17–18), 2985–2996 (1996).
94. M.-H. Yang, K.-C. Chang, J.-Y. Lin, 'Multifunctional Ion-exchange Stationary Phases for High-performance Liquid Chromatography', *J. Chromatogr., A*, **722**(1–2), 87–96 (1996).
95. A. Maisch, 'HPLC Phase with Fluorocarbon Coating', *LaborPraxis*, **20**(4), 66–73 (1996).
96. K. Jinno, H. Nakamura, 'Retention Characteristics of Fluorinated Bonded Silica Phase in Reversed-phase Liquid Chromatography', *Chromatographia*, **39**(5–6), 285–293 (1994).
97. H.Y. Aboul-Enein, V. Serignese, 'Liquid Chromatographic Determination of Taxol and Related Derivatives Using a New Polyfluorinated Reversed-phase Column', *Anal. Chim. Acta*, **319**(1–2), 187–190 (1996).
98. S.L. Richheimer, D.M. Tinnermeier, D.W. Timmons, 'High-performance Liquid Chromatographic Assay of Taxol', *Anal. Chem.*, **64**(20), 2323–2326 (1992).
99. P.C. Sadek, P.W. Carr, 'Study of Solute Retention in Reversed-phase High-performance Liquid Chromatography on Hydrocarbonaceous and Three Fluorinated Bonded Phases', *J. Chromatogr.*, **288**(1), 25–41 (1984).
100. M.-P. Krafft, M. Le Blanc, J.G. Riess, A. Berthod, 'Highly Fluorinated Stationary Phases for Analysis of Polyfluorinated Solutes by Reversed-phase High-performance Liquid Chromatography', *Anal. Chem.*, **60**(18), 1969–1972 (1988).
101. T. Monde, T. Kamiusuki, T. Kuroda, K. Mikumo, T. Ohkawa, H. Fukube, 'High-performance Liquid Chromatographic Separation of Phenols on a Fluorocarbon-bonded Silica Gel Column', *J. Chromatogr., A*, **722**(1–2), 273–280 (1996).
102. X. Geng, P.W. Carr, 'Use of Fluorinated Bonded Phases in Reversed-phase High-performance Liquid Chromatography of Proteins', *J. Chromatogr.*, **269**(2), 96–102 (1983).
103. C. Hirayama, H. Ihara, S. Nagaoka, K. Hamada, 'Totally Organic Stationary Phase from Perfluoroalkyl Acrylate for Liquid Chromatography', *J. Chromatogr.*, **465**(2), 241–248 (1989).
104. K. Hosoya, E. Sawada, K. Kimata, T. Araki, N. Tanaka, 'Preparation and Chromatographic Properties of Uniform Size Cross-linked Macroporous Poly(vinyl *p*-*t*-butylbenzoate) Beads. Evaluation of Preferential Retention Toward Organohalides', *J. Chromatogr., A*, **662**(1), 37–47 (1994).
105. S. Dhanesar, C.F. Poole, 'Support-deactivated Fluorocarbon Stationary Phase for Gas Chromatography', *Anal. Chem.*, **55**(13), 2148–2151 (1983).
106. T.V. Briggle, D.A. Boothman, C.D. Pfaffenberger, S. Greer, 'Analysis of 5-Fluoro-2′-deoxycytidine and 5′-Trifluoromethyl-2′-deoxycytidine and Their Related Antimetabolites by High-performance Liquid Chromatography', *J. Chromatogr. Biomed. Appl.*, **54**, 2 (1986).
107. P. Leone, C.H. Collins, 'Isocratic Reversed-phase High-performance Liquid Chromatographic Separation of Simple Perhalogenated Compounds', *J. Chromatogr.*, **553**(1), 399–405 (1991).
108. P. de A. Leone, C.A. Bertran, C.H. Collins, 'Interpolation Curves for Identification of Perhalogenated Compounds Separated by Reversed-phase HPLC', *J. High Resolut. Chromatogr.*, **12**(7), 493–495 (1989).
109. J. Yamaguchi, T. Hanai, 'Selectivity Related to Carbon Loading and End-capping of Octadecyl-bonded Silica Gels in Reversed-phase Liquid Chromatography of Phenolic Compounds', *Chromatogr. Sci.*, **27**(12), 710–715 (1989).
110. G. DeVries, R.F. Rekker, 'Retention Studies of Alkyl- and Halogen-substituted Aromatics on Normal-phase Silica and Alumina Columns. IV. Effect of the Mobile Phase Structure on Retention Data', *J. Liq. Chromatogr.*, **16**(2), 383–402 (1993).
111. Y. Hirose, R. Matsumoto, S. Yamada, T. Nozaka, M. Ishino, A. Tanaka, 'The Determination of Volatile Halogenated Organic Compounds in Drugs', *Jpn. J. Toxicol. Environ. Health*, **40**(3), 298–301 (1994).
112. P. Lillsunde, T. Korte, 'Comprehensive Drug Screening in Urine Using Solid-phase Extraction and Combined TLC and GC/MS Identification', *J. Anal. Toxicol.*, **15**(2), 71–81 (1991).
113. T. Gabrio, G. Galetzka, 'Gas Chromatographic Testing of Halothane Purity', *Pharmazie*, **42**(5), 306–307 (1987).
114. O. Suzuki, H. Seno, A. Ishii, 'Analysis of Drugs and Poisons in Legal Medicine', *J. Mass Spectrom. Soc. Jpn.*, **44**(3), 321–333 (1996).
115. B. Silber, N.H.G. Holford, S. Riegelman, 'Stereoselective Disposition and Glutcuronidation of Propranolol in Humans', *J. Pharm. Sci.*, **71**(6), 699–704 (1982).
116. A.M. Krstulovic, 'Racemates Versus Enantiomerically Pure Drugs: Putting High-performance Liquid Chromatography to Work in the Selection Process', *J. Chromatogr.*, **488**, 53–72 (1989).
117. K. Fried, I.W. Wainer, 'Column-switching Techniques in the Biomedical Analysis of Stereoisomeric Drugs:

Why, How and When', *J. Chromatogr., B*, **689**, 91–104 (1997).
118. J. Ducharme, C. Fernandez, F. Gimenez, R. Farinotti, 'Critical Issues in Chiral Drug Analysis in Biological Fluids by High-performance Liquid Chromatography', *J. Chromatogr., B*, **686**, 65–75 (1996).
119. H. Fujima, H. Wada, T. Miwa, J. Haginaka, 'Chiral Separation of Lorazepam on Ovomucoid-bonded Columns: Peak Coalescence Due to Racemization', *J. Liq. Chromatogr.*, **16**(4), 879–891 (1993).
120. A.L. Lehninger, 'Biochemistry', 2nd edition, Worth Publishers, New York, 462–463, 1975.
121. D.W. Armstrong, X. Wang, L.W. Chang, H. Ibrahim, G.R. Reid III, T.E. Beesley, 'Comparison of the Selectivity and Retention of β-Cyclodextrin vs. Heptakis-2,3-*O*-dimethyl-β-cyclodextrin LC Stationary Phases for Structural and Geometric Isomers', *J. Liq. Chromatogr. Relat. Technol.*, **20**(20), 3297–3308 (1997).
122. V. Tittelbach, R.K. Gilpin, 'Species Dependency of the Liquid Chromatographic Properties of Silica-immobilized Serum Albumins', *Anal. Chem.*, **67**(1), 44–47 (1995).
123. H.Y. Aboul-Enein, M.R. Islam, 'Structural Factors Affecting Chiral Recognition and Separation on Cellulose Based Chiral Stationary Phases', *J. Liq. Chromatogr.*, **13**(3), 485–492 (1990).
124. S.V. Rogozhin, V.A. Davankov, 'Chromatography of Ligands as a New Method for Studying Mixed Complexes. Stereoselective Effects in α-Amino Acid Complexes of Copper(II)', *Dokl. Akad. Nauk. S.S.S.R.*, **192**, 1288 (1970).
125. V.A. Davankov, A.S. Bochkov, 'Resolution of Racemates by High-performance Liquid Chromatography', *Adv. Chromatogr.*, **22**, 71–116 (1983).
126. L.R. Sousa, G.D.Y. Sogah, D.H. Hoffmann, D.J. Cram, *J. Am. Chem. Soc.*, **100**, 4569 (1978).
127. T. Shinbo, T. Yamaguchi, K. Nishimura, M. Sugiura, 'Chromatographic Separation of Racemic Amino Acids by Use of Chiral Crown Ether-coated Reversed-phase Packings', *J. Chromatogr.*, **405**, 145 (1987).
128. M. Hilton, D.M. Armstrong, 'Evaluation of the Enantiomeric Separation of Dipeptides Using a Chiral Crown Ether LC Column', *J. Liq. Chromatogr.*, **14**(20), 3673–3683 (1991).
129. B. Feibush, A. Figueroa, R. Charles, K.D. Onan, P. Feibush, B.L. Karger, 'Chiral Separation of Heterocyclic Drugs by HPLC: Solute–Stationary Phase Base-pair Interactions', *J. Am. Chem. Soc.*, **108**, 3310 (1986).
130. W. Szczepaniak, A. Szymanski, 'Characterization of Macrocyclic Bonded Phases in High-pressure Liquid Chromatography Using Phenol Derivatives', *Chem. Anal.*, **32**(3), 273–284 (1987).
131. D.M. Armstrong, Y. Tang, S. Chen, Y. Zhou, C. Bagwill, J.-R. Chen, 'Macrocyclic Antibiotics as a New Class of Chiral Selectors for Liquid Chromatography', *Anal. Chem.*, **66**(9), 1473–1484 (1994).
132. S. Chen, Y. Liu, D.W. Armstrong, J.I. Borrell, B. Martinez-Teipel, J.L. Matallana, 'Enantioresolution of Substituted 2-Methoxy-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitriles on Macrocyclic Antibiotics and Cyclodextrin Phases', *J. Liq. Chromatogr.*, **18**(8), 1495–1507 (1995).
133. D.W. Armstrong, Y.B. Liu, K.H. Ekborg-Ott, 'A Covalently Bonded Teicoplanin Chiral Stationary Phase for HPLC Enantioseparations', *Chirality*, **7**(6), 474–497 (1995).
134. K.H. Ekborg-Ott, Y. Liu, D.W. Armstrong, 'Highly Enantioselective HPLC Separations Using the Covalently Bonded Macrocyclic Antibiotic, Ristocetin A, Chiral Stationary Phase', *Chirality*, **10**(5), 434–484 (1998).
135. K.J. Duff, H.L. Gray, R.J. Gray, C.C. Bahler, 'Chiral Stationary Phases in Concert with Homologous Chiral Mobile Phase Additives: Push/Pull Model', *Chirality*, **5**, 201–206 (1993).
136. M.I. Maguregui, R.M. Alonso, R.M. Jimenez, 'High-performance Liquid Chromatography with Amperometric Detection Applied to the Screening of Beta-blockers in Human Urine', *J. Chromatogr., B, Biomed. Appl.*, **674**(1), 85–91 (1995).
137. T. Ohta, S. Niida, H. Nakamua, 'Selective Attraction of Beta-blockers from Biological Fluids by Column-switching High-performance Liquid Chromatography Using an Internal-surface Phenylboronic Acid Precolumn', *J. Chromatogr., B, Biomed. Appl.*, **675**(1), 168–173 (1996).
138. A. Ceccato, P. Hubert, J. Crommen, 'Direct Liquid Chromatographic Enantioseparation of Sotalol and Other Beta-blockers Using an α-1-Acid Glycoprotein-based Chiral Stationary Phase', *J. Chromatogr., A*, **760**(2), 193–203 (1997).
139. P.V. Petersen, J. Ekelund, L. Olsen, S.V. Ovesen, 'Chiral Separations of Beta-blocking Drug Substances Using Pirkle-type α-Burke 1 Chiral Stationary Phase', *J. Chromatogr., A*, **757**(1–2), 65–71 (1997).
140. T. Toyo'oka, M. Toriumi, Y. Ishii, 'Enantioseparation of Beta-blockers Labelled with a Chiral Fluorescent Reagent, *R*-(−)-DBD-PyNCS, by Reversed-phase Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **15**(9–10), 1467–1476 (1997).
141. 'The Restek Advantage', **3**, Spring/Summer, 1998.
142. J.-L. Janicot, M. Caude, R. Rosset, 'High-performance Liquid Chromatography of Opium Alkaloids', *Analusis*, **14**(9), 441–455 (1986).
143. D.E.M.M. Vendrig, J.J.M. Holthuis, 'Trace Level Analysis of Vinca Alkaloids in Biological Fluids', *Trends Anal. Chem. (Pers. Ed.)*, **8**(4), 141–145 (1989).
144. B.D. Kabulov, A.L. D'yakonov, S.V. Zalyalieva, 'High-performance Liquid Chromatography of Alkaloids', *Chem. Nat. Compd. (Engl. Transl.)*, **27**(5), 521–540 (1992).

145. I.N. Papadoyannis, 'Instrumental Analytical Techniques Used for Datura Alkaloid Analyses', *Instrum. Sci. Technol.*, **22**(3), 241–258 (1994).
146. I. Valka, 'Determination of Isoquinoline Alkaloids', *Chem. Listy*, **83**(7), 716–729 (1989).
147. V. Walker, G.A. Mills, 'Quantitative Methods for Amino Acid Analysis in Biological Fluids', *Ann. Clin. Biochem.*, **32**(1), 28–57 (1995).
148. G. Rogerson, 'Amino Acid Analysis by HPLC with Precolumn Derivatization', *Int. Anal.*, **1**(4), 19–22 (1987).
149. X.S. Zhang, Z.Y. Chu, L. Huang, M. Lin, J. Qu, C.S. Lin, 'Analysis of Compound Amino Acid Injection by Precolumn Derivatization with *o*-Phthaldialdehyde and RP-HPLC', *Yaowu Fenxi Zazhi*, **16**(1), 6–9 (1996).
150. M. Nieder, G.J. Hughs, S. Frutiger, E. Grom, 'Operationally Safe Amino Acid Analysis with the Dabsyl Chloride System', *LaborPraxis*, **14**(1–2), 26–31 (1990).
151. S.A. Cohen, D.J. Strydom, 'Amino Acid Analysis Utilizing Phenyl Isothiocyanate Derivatives', *Anal. Biochem.*, **174**(1), 1–16 (1988).
152. H.-M. Zhang, F.-X. Zhou, I.S. Krull, 'Comparison of 9-Fluorenylmethoxycarbonyl and 9-Fluoreneacetyl-tagged Silica-based Derivatization Reagents in High-performance Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **10**(8), 577–586 (1992).
153. H. Cui, J. Leon, E. Reusaet, A. Bult, 'Selective Determination of Peptides Containing Specific Amino Acid Residues by High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr., A*, **704**(1), 27–36 (1995).
154. M. Watanabe, Y. Kono, Y. Esumi, T. Teraoka, D. Hosokawa, Y. Suzuki, A. Sakurai, M. Watanabe, 'Studies on a Quantitative Analysis of Oryzalides and Oryzalic Acids in Rice Plants by GC/SIM', *Biosci. Biotechnol. Biochem.*, **60**(9), 1440–1443 (1996).
155. J.H. Beijnen, P.L. Meenhorst, R. Van Gijn, W.A. Hazelager, C.H.W. Koks, W.J.M. Underberg, 'Gas Chromatographic Analysis, with Electron Capture Detection, of the Antifungal Drug Fluconazole in Microvolumes of Human Plasma', *J. Pharm. Biomed. Anal.*, **9**(10–12), 1173–1175 (1991).
156. M.M. El-Sherei, F.S. El-Feraly, M.A. Elsohly, D.F. Stanford, 'Quantitation of Canellal, an Antifungal Sesquiterpene, in Canella Bark and in Biological Fluids', *Fitoterapia*, **58**(4), 272–275 (1987).
157. S.A. Howell, A.I. Mallet, 'Novel Lanosterol Isomer Produced in Response to Azole Antifungals', *J. Steroid Biochem.*, **36**(5), 505–506 (1990).
158. M. Schwartz, W. Kline, B. Matuszewski, 'Determination of a Cyclic Hexapeptide (L743872), a Novel Pneumocandin Antifungal Agent in Human Plasma and Urine by High-performance Liquid Chromatography with Fluorescence Detection', *Anal. Chim. Acta*, **352**(1–3), 299–307 (1997).
159. L.L. Zornes, R.E. Stratford, 'Development of a Plasma High-performance Liquid Chromatographic Assay for LY303366, a Lipopeptide Antifungal Agent, and its Application in a Dog Pharmokinetic Study', *J. Chromatogr., B, Biomed. Appl.*, **695**(2), 381–387 (1997).
160. I.-M. Chung, S.-B. Paik, 'Separation and Activity Test of Antifungal Substance from *C. Japonica* Extract', *Anal. Sci. Technol.*, **10**(2), 153–159 (1997).
161. C.H.W. Koks, H. Rosing, P.L. Meenhorst, A. Bult, J.H. Beijnen, 'High-performance Liquid Chromatographic Determination of the Antifungal Drug Fluconazole in Plasma and Saliva of Human Immunodeficiency Virus-infected Patients', *J. Chromatogr., B., Biomed. Appl.*, **663**(2), 345–351 (1995).
162. K.K. Hosotsubo, H. Hosotsubo, M.K. Nishijima, T. Okada, N. Taenaka, I. Yoshiya, 'Rapid Determination of Serum Levels of a New Antifungal Agent, Fluconazole, by High-performance Liquid Chromatography', *J. Chromatogr., Biomed. Appl.*, **94**(1), 223–228 (1990).
163. B. Sauer, R. Matusch, 'High-performance Liquid Chromatographic Separations of Nystatin and Their Influence on the Antifungal Activity', *J. Chromatogr., A*, **672**(1–2), 247–253 (1994).
164. O. Raatikainen, S. Auriola, J. Tuomisto, 'Identification of Aromatic Moieties and Mycosamine in Antifungal Heptaenes with High-performance Liquid Chromatography, High-performance Liquid Chromatography–Mass Spectrometry and Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **585**(2), 247–254 (1991).
165. M. Greiner, G. Winkelmann, 'Rapid Identification and Determination of Herbicolin A and B by High-performance Liquid Chromatography', *J. Chromatogr.*, **502**(2), 437–442 (1990).
166. S.S. Zarapkar, B.B. Salunkhe, U.B. Salunkhe, U.J. Joshi, S.V. Sawant, R.V. Rele, High-performance Liquid Chromatographic Determination of Bromosalicylchloranilide from Pharmaceutical Preparations', *Indian Drugs*, **26**(8), 437–439 (1989).
167. C. Bergeron, A. Marston, J.-L. Wolfender, S. Mavi, C. Rogers, K. Hostettmann, 'Isolation of Polyphenols from *Polygala Gazensis* and Liquid Chromatography–Mass Spectrometry of Related African *Polygala* Species', *Phytochem. Anal.*, **8**(1), 32–36 (1997).
168. C. Albet, J.M. Fernandez, E. Rozman, J.A. Perez, A. Sacristan, J.A. Ortiz, 'Determination of Sertaconazole Nitrate, a New Imidazole Antifungal, by High-performance Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **10**(2–3), 205–211 (1992).
169. D.A. Stopher, R. Gage, 'Determination of a New Antifungal Agent, Voriconazole, by Multidimensional High-performance Liquid Chromatography with Direct Plasma Injection onto a Size-exclusion Column', *J. Chromatogr., B, Biomed. Appl.*, **691**(2), 441–448 (1997).
170. H. Kim, C.-C. Lin, 'A Chiral Liquid Chromatographic Method for the Determination of the Enantiomers of the Racemic Triazole Antifungal Drug (Sch 39304)

in Human Plasma', *J. Pharm. Biomed. Anal.*, **13**(11), 1415–1419 (1995).
171. R. Woestenborghs, J. Heykants, 'Assay Methods for Imidazole- and Triazole-like Antifungals and Some Anthelmintics', *Methodol. Surv. Biochem. Anal.*, **20**, *Anal. Drugs Metab., Incl. Anti-Infect. Agents*, 153–156 (1990).
172. P.A. Ristuccia, 'Liquid Chromatographic Assays of Antimicrobial Agents', *J. Liq. Chromatogr.*, **10**(2–3), 241–275 (1987).
173. D.W. Warnock, M.D. Richardson, A. Turner, 'High-performance Liquid Chromatographic (HPLC) and Other Non-biological Methods for Quantitation of Antifungal Drugs', *J. Antimicrob. Chemother.*, **10**(6), 467–478 (1982).
174. A. Van Overbeke, P. Sandra, A. Medvedovici, W. Baeyens, H.Y. Aboul-Enei, 'Application of Chiralcel OJ in Supercritical Fluid Chromatography for the Resolution of Different Groups of Frequently Used Drug Racemates', *Chirality*, **9**, 126–132 (1992).
175. J. Bojarski, 'Chromatography of Enantiomers of 2-Arylpropionic Acids', *J. Liq. Chromatogr.*, **12**(14), 2685–2706 (1989).
176. J.C. McElnay, 'HPLC Analysis of Anti-inflammatory Agents (Non-steroidal and Steroidal) (in Biological Fluids)', *Methodol. Surv. Biochem. Anal.*, **18**, *Bioanal. Drugs Metab.*, 87–95 (1988).
177. T.D. Wilson, 'Recent Advances in HPLC Analysis of Analgesics', *J. Liq. Chromatogr.*, **9**(11), 2309–2409 (1986).
178. C. Ho, G.L. Chen, 'Stability-indicating High-performance Liquid Chromatographic Assay Methods for Drugs in Pharmaceutical Dosage Forms. Part I.', *Yaowu Shipin Fenxi*, **4**(4), 271–292 (1996).
179. M. Matsuda, M. Sakashita, T. Yamaguchi, T. Fujii, 'Determination of a Novel Anti-psychotic Agent AD 5423 and its Metabolites in Plasma by High-performance Liquid Chromatography with Fluorescence Detection', *J. Pharm. Biomed. Anal.*, **15**(9–10), 1449–1456 (1997).
180. G. Urden, M. Thoernwall, S. Lyrenaes, L. Lindstroem, F. Nyberg, 'Classification of CSF Samples from Normal and Post-partum Psychotic Women Using Chromatographic Profiles with Bilinear Projections: A Multivariate Approach', *Biomed. Chromatogr.*, **10**(4), 149–154 (1996).
181. A.J.J. Van den Berg, R.P. Labadie, 'High-performance Liquid Chromatographic Separation and Quantitative Determination of 1,8-Dihroxyanthraquinones in Plant Cell Cultures', *J. Chromatogr.*, **329**(2), 311–314 (1985).
182. Y. Saito, M. Mifune, J. Odo, Y. Otsuki, M. Mitsuhashi, Y. Mori, A.E.H. Gassim, J. Haginaka, 'Copper(II) Phthalocyanine Sulphonyl–Aminopropyl Silica Gel for Separation of pi-Electron Rich Compounds by High-performance Liquid Chromatography', *Anal. Sci.*, **7**(5), 805–806 (1991).
183. J. Goto, Y. Saisho, T. Nambara, 'Sensitive Fluorescence Labelling Reagents for High-performance Liquid Chromatography of Carbonyl Compounds', *Anal. Sci.*, **5**(4), 399–402 (1989).
184. M.L. Nedved, S. Habibi-Goudarzi, B. Ganem, J.D. Henion, 'Characterization of Benzodiazepine ''Combinatorial'' Chemical Libraries by Online Immunoaffinity Extraction, Coupled-column HPLC–Ionspray Mass Spectrometry–Tandem Mass Spectrometry', *Anal. Chem.*, **68**(23), 4228–4236 (1996).
185. L.A. Larew, B.A. Olsen, J.D. Stafford, M.V. Wilhelm, 'Comparison of Theory-based and Empirical Modelling for the Prediction of Chromatographic Behavior in the Ion-pairing Separation of Benzodiazepine-derived Pharmaceutical Compounds', *J. Chromatogr., A*, **692**(1–2), 183–193 (1995).
186. W.E. Lambert, E. Meyer, Y. Xue-Ping, A.P. De Leenheer, 'Screening, Identification, and Quantitation of Benzodiazepines in Post-mortem Samples by HPLC with Photodiode Array Detection', *J. Anal. Toxicol.*, **19**(1), 35–40 (1995).
187. H.-L. Tai, R.J. Osiewicz, D.P. Bofinger, 'Salt-free Isocratic High-performance Liquid Chromatographic Method for the Quantitation of Benzodiazepines in Serum', *Clin. Biochem.*, **26**(3), 165–171 (1993).
188. M. Mazhar, S.R. Binder, 'Analysis of Benzodiazepines and Tricyclic Antidepressants in Serum Using a Common Solid-phase Cleanup and a Common Mobile Phase', *J. Chromatogr.*, **497**, 201–212 (1989).
189. H.J.J. Willems, A. Van der Horst, P.N.F.C. De Goede, G.J. Haakmeester, 'Determination of Some Anticonvulsants, Antiarrhythmics, Benzodiazepines, Xanthines, Paracetamol and Chloramphenicol by Reversed-phase HPLC', *Pharm. Weekbl., Sci. Ed.*, **7**(4), 150–157 (1985).
190. I.M. McIntyre, M.L. Syrjanen, K. Crump, S. Horomidis, A.W. Peace, O.H. Drummer, 'Simultaneous HPLC Gradient Analysis of Fifteen Benzodiazepines and Selected Metabolites in Postmortem Blood', *J. Anal. Toxicol.*, **17**(4), 202–207 (1993).
191. K.K. Akerman, J. Jolkkonen, M. Parviainen, I. Penttila, 'Analysis of Low Dose Benzodiazepines by HPLC with Automated Solid-phase Extraction', *Clin. Chem.* (Washington, DC), **42**(9), 1412–1416 (1996).
192. E. Tanaka, M. Terada, S. Misawa, C. Wakasugi, 'Simultaneous Determination of Twelve Benzodiazepines in Human Serum Using a New Reversed-phase Chromatographic Column on a 2-μm Porous Microspherical Silica Gel', *J. Chromatogr., B, Biomed. Appl.*, **682**(1), 173–178 (1996).
193. S.M. Sultan, A.H. El-Mubarak, 'High-performance Liquid Chromatographic Method for the Separation and Quantification of Some Psychotherapeutic Benzodiazepines Optimized by the Modified Simplex Procedure', *Talanta*, **43**(4), 569–576 (1996).

194. Y. Guillaume, C. Guinchard, 'Thermodynamic Behavior of Mixed Benzodiazepines by a New Liquid Chromatographic Method', *Chromatographia*, **40**(3–4), 193–196 (1995).
195. Y. Guillaume, C. Guinchard, 'Study and Optimization of Column Efficiency in HPLC: Comparison of Two Methods for Separating Ten Benzodiazepines', *J. Liq. Chromatogr.*, **17**(7), 1443–1459 (1994).
196. C. Lacroix, F. Wojciechowski, P. Danger, 'Monitoring of Benzodiazepines (Clobazam, Diazepam, and Their Main Active Metabolites) in Human Plasma by Column-switching High-performance Liquid Chromatography', *J. Chromatogr., Biomed. Appl.*, **128**(2), 285–290 (1993).
197. A.M.A. Verweij, P.J.L. Lipman, P.G.M. Zweipfenning, 'Quantitative Liquid Chromatography, Thermospray–Tandem Mass Spectrometry (LC/TSP/MS/MS) Analysis of Some Thermolabile Benzodiazepines in Whole Blood', *Forensic Sci. Int.*, **54**(1), 67–74 (1992).
198. M.A. Perillo, D.A. Garcia, 'Non-labelled Benzodiazepines Partitioned into Synaptosomal Membranes: Their Extraction and Quantification by High-performance Liquid Chromatography', *Biomed. Chromatogr.*, **6**(4), 183–190 (1992).
199. I.S. Lurie, D.A. Cooper, R.F.X. Klein, 'High-performance Liquid Chromatographic Analysis of Benzodiazepines Using Diode Array, Electrochemical and Thermospray Mass Spectrometric Detection', *J. Chromatogr.*, **598**(1), 59–66 (1992).
200. C.M. Moore, K. Sato, Y. Katsumata, 'Rapid Monitoring of Benzodiazepines in Clinical Samples by Using Online Column-switching HPLC', *Clin. Chem.* (Winston-Salem, NC), **37**(6), 804–808 (1991).
201. A. Boukhabza, A.A.J. Lugnier, P. Kintz, P. Mangin, 'Simultaneous HPLC Analysis of the Hypnotic Benzodiazepines Nitrazepam, Estazolam, Flunitrazepam and Triazolam in Plasma', *J. Anal. Toxicol.*, **15**(6), 319–322 (1991).
202. P.R. Puopolo, M.E. Pothier, S.A. Volpicelli, J.G. Flood, 'Single Procedure for Detection, Confirmation and Quantification of Benzodiazepines in Serum by Liquid Chromatography with Photodiode Array Detection', *Clin. Chem.* (Winston-Salem, NY), **37**(5), 701–706 (1991).
203. F.T. Noggle, Jr, C.R. Clark, J. De Ruiter, 'Liquid Chromatographic Separation of Some Common Benzodiazepines and Their Metabolites', *J. Liq. Chromatogr.*, **13**(20), 4005–4021 (1990).
204. D. Ladewig, A. Scholer, 'Urine Analysis for Benzodiazepines Among Opiate Consumers', *Drug Alcohol Depend.*, **25**(2), 145–148 (1990).
205. H. Bun, V. Duplan, P. Crevat-Pisano, M. Llurens, A. Durand, 'Rapid Determination of the Benzodiazepine Antagonist Flumazenil, Ro 15–1788, by High-performance Liquid Chromatography', *Biomed. Chromatogr.*, **3**(6), 269–271 (1989).
206. D. Wilhelm, A. Kemper, 'High-performance Liquid Chromatographic Procedure for the Determination of Clozapine, Haloperidol, Droperidol and Several Benzodiazepines in Plasma', *J.Chromatogr.*, **525**(1), 218–224 (1990).
207. V.M. Haver, W.H. Porter, L.D. Dorie, J.R. Lea, 'Simplified High-performance Liquid Chromatographic Method for the Determination of Clonazepam and Other Benzodiazepines in Serum', *Ther. Drug Monit.*, **8**(3), 352–357 (1986).
208. M. Chiarotti, N. De Giovanni, A. Fiori, 'Analysis of Benzodiazepines. I. Chromatographic Identification', *J. Chromatogr.*, **358**(1), 169–178 (1986).
209. R. Gill, B. Law, J.P. Gibbs, 'High-performance Liquid Chromatography Systems for the Separation of Benzodiazepines and Their Metabolites', *J. Chromatogr.*, **356**(1), 37–46 (1986).
210. N. De Giovanni, M. Chiarotti, 'Analysis of Benzodiazepines. [II.] High-performance Liquid Chromatography–Fluorescence Detection After Molecular Rearrangement to Acridanones', *J. Chromatogr.*, **428**(2), 321–329 (1988).
211. M.C. Pietrogrande, F. Dondi, G. Blo, P.A. Borea, C. Bighi, 'Retention Behavior of Benzodiazepines in Normal-phase HPLC. Silica, Cyano and Amino Phases Comparison', *J. Liq. Chromatogr.*, **11**(6), 1313–1333 (1988).
212. A. Radwanska, T. Frackowiak, H. Ibrahim, A.-F. Aubry, R. Kaliszan, 'Chromatographic Modelling of Interactions Between Melanin and Phenothiazine and Dibenzazepine Drugs', *Biomed. Chromatogr.*, **9**(5), 233–237 (1995).
213. F.M. Alvarez, C.B. Bottom, P. Chikhale, C. Pidgeon, 'Immobilized Artificial Membrane Chromatography. Prediction of Drug Transport Across Biological Barriers', in *Molecular Interactions in Bioseparations*, ed. T.T. Ngo, Plenum Press, New York, 151–167, 1993.
214. S.D. Kimball, D.M. Floyd, J.C. Barrish, J. Das, J.T. Hunt, J. Krapcho, G. Rovnyak, K.J. Duff, M. Lago, V.G. Lee, R.V. Moquin, S. Spergel, J.Z. Gougoutas, M. Malley, A. Pudzianowski, S.F. Flaim, S. Moreland, A. Hedberg, R.E. White, 'Structure Activity Studies and Receptor Binding Model of 1-Benzazapine-2-one Calcium Antagonists', *Trends Med. Chem.*, **90**, 271–276 (1992).
215. I. Fellegvari, J. Visy, K. Valko, T. Lang, M. Simonyi, 'Investigation of Conformational Diastereomers of 2,3-Benzodiazepines by High-performance Liquid Chromatography', *J. Liq. Chromatogr.*, **12**(14), 2719–2732 (1989).
216. H. Jin, Y. Zhang, Y. Guo, S. Jin, 'HPLC Separations of Isomers of Tetrahydro-1,5-benzothiazepines and Tetrahydro-1,5-benzodiazepines', *Chromatographia*, **22**(1–6), 153–156 (1986).
217. C. Franzelius, K. Besserer, 'Identification and Quantitation of Intact Diastereomeric Benzodiazepine

Glucuronides in Biological Samples by High-performance Liquid Chromatography', *J. Chromatogr. Biomed. Appl.*, **124**(1), 162–167 (1993).
218. A. Van Overbeke, W. Baeyens, H. Oda, H.Y. Aboul-Enein, 'Direct Enantiomeric HPLC Separation of Several 2-Arylpropionic Acids, Barbituric Acids and Benzodiazepines on Chiralcel OJ-R Chiral Stationary Phase', *Chromatographia*, **43**(11–12), 599–606 (1996).
219. T. Nishikawa, Y. Hayashi, S. Suzuki, H. Kubo, H. Ohtani, 'On-column Enantiomerization of 3-Hydroxybenzodiazepines During Chiral Liquid Chromatography with Optical Rotation Detection', *J. Chromatogr., A*, **767**(1–2), 93–100 (1997).
220. I. Fitos, J. Visy, M. Simonyi, J. Hermansson, 'Separation of Enantiomers of Benzodiazepines on the Chiral-AGP Column', *J. Chromatogr., A*, **709**(2), 265–273 (1995).
221. C. Bertucci, A. Demuro, G. Giannaccini, 'HPLC Resolution of C5 Chiral 4,5-Dihydro-1,4-benzodiazepines: Stereochemical Characterization and Enantioselective GABAA Receptor Binding', *J. Pharm. Biomed. Anal.*, **15**(6), 803–809 (1997).
222. A.J.H. Louter, E. Bosma, J.C.A. Schipperen, J.J. Vreuls, U.A.T. Brinkman, 'Automated Online Solid-phase Extraction–Gas Chromatography with Nitrogen–Phosphorous Detection: Determination of Benzodiazepines in Human Plasma', *J. Chromatogr., B, Biomed. Appl.*, **689**(1), 35–43 (1997).
223. R. Meatherall, 'GC/MS Confirmation of Urinary Benzodiazepine Metabolites', *J. Anal. Toxicol.*, **18**(7), 369–381 (1994).
224. W.A. Joern, 'Confirmation of Low Concentrations of Urinary Benzodiazepines, Including Alprazolam and Triazolam, by GC/MS: An Extractive Alkylation Procedure', *J. Anal. Toxicol.*, **16**(6), 363–367 (1992).
225. J.E. Paanakker, H.J.M. Van Hal, 'Capillary Gas Chromatographic Assay for the Routine Monitoring of the Antidepressant Mepirzepine in Human Plasma', *J. Chromatogr., Biomed. Appl.*, **61**, 1 (1987).
226. O. Beck, Z. Lin, K. Brodin, S. Borg, P. Hjemdahl, 'The Online Screening Technique for Urinary Benzodiazepines: Comparison with EMIT, FPIA, and GC/MS', *J. Anal. Toxicol.*, **21**(7), 554–557 (1997).
227. J.L. Valentine, R. Middleton, C. Sparks, 'Identification of Urinary Benzodiazepines and Their Metabolites: Comparison of Automated HPLC and GC/MS After Immunoassay Screening of Clinical Specimens', *J. Anal. Toxicol.*, **20**(6), 416–424 (1996).
228. S.B. Needleman, M. Porvaznik, 'Identification of Parent Benzodiazepines by Gas Chromatography–Mass Spectroscopy (GC/MS) from Urinary Extracts Treated with β-Glucuronidase', *Forensic Sci. Int.*, **73**(1), 49–60 (1995).
229. P.D. Camara, L. Audette, K. Velletri, P. Breitenbecher, M. Rosner, W.C. Griffiths, 'False-positive Immunoassay Results for Urine Benzodiazepine in Patients Receiving Oxaproxin (Daypro)', *Clin. Chem.* (Winston-Salem, NC), **41**(1), 115–116 (1994).
230. C.E. Jones, F.H. Wians, Jr, L.A. Martinez, G.J. Merritt, 'Benzodiazepines Identified by Capillary Gas Chromatography–Mass Spectrometry, with Specific Ion Screening Used to Detect Benzophenone Derivatives', *Clin. Chem.* (Winston-Salem, NC), **35**(7), 1394–1398 (1989).
231. S.J. Mule, G.A. Casella, 'Quantitation and Confirmation of the Diazolo- and Triazolo- Benzodiazepines in Human Urine by Gas Chromatography–Mass Spectrometry', *J. Anal. Toxicol.*, **13**(3), 179–184 (1989).
232. R. Herraez, A.J.H. Louter, N.C. van de Merbel, U.A.T. Brinkman, 'Automated Online Dialysis for Sample Preparation for Gas Chromatography: Determination of Benzodiazepines in Human Plasma', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1077–1087 (1996).
233. P. Lillsunde, L. Michelson, T. Forsstrom, T. Korte, E. Schultz, K. Ariniemi, M. Portman, M.-L. Sihvonen, T. Seppala, 'Comprehensive Drug Screening in Blood for Detecting Abused Drugs or Drugs Potentially Hazardous for Traffic Safety', *Forensic Sci. Int.*, **77**(3), 191–210 (1996).
234. H. Gjerde, E. Dahlin, A.S. Christophersen, 'Simultaneous Determination of Common Benzodiazepines in Blood Using Capillary Gas Chromatography', *J. Pharm. Biomed. Anal.*, **10**(5), 317–322 (1992).
235. C. Drouet-Coassolo, C. Aubert, P. Coassolo, J.-P. Cano, 'Capillary Gas Chromatographic–Mass Spectrometric Method for the Identification and Quantification of Some Benzodiazepines and Their Unconjugated Metabolites in Plasma', *J. Chromatogr., Biomed. Appl.*, **79**, 2 (1989).
236. H. Bun, C. Dechamp, P. Coassolo, J.P. Cano, 'Determination by Means of Capillary Gas Chromatography of the Benzodiazepines Clobazam, Clonazepam, Diazepam, Nitrazepam, and Their Principal Metabolites, Currently Used in the Treatment of Epilepsy', *Anal. Lett.*, **22**(4), 897–913 (1989).
237. M. Kinter, S. Singh, R.A. Felder, 'Quantitation of Selective Dopaminergic Drugs in Plasma by Gas Chromatography–Mass Spectrometry Following Solid-phase Extraction', *J. Chromatogr.*, **496**(1), 201–208 (1989).
238. P. Levandoski, K. Straub, D. Shah, R. DeMarinis, C. De Mey, 'High-sensitivity Assay for the α-2 Antagonist 6-Chloro-3-methyl-2,3,4,5-tetrahydro-1*H*-3-benzazepine and its Desmethyl Metabolite in Plasma Using Gas Chromatography–Mass Spectrometry with Ammonia–Carbon Tetrachloride Chemical Ionization', *Biomed. Environ. Mass Spectrom.*, **13**(10), 523–530 (1986).
239. D. Hoehne, G. Bohn, H.-P. Their, 'Determination of Benzodiazepine Residues in Biological Tissues and Fluids by Gas–Liquid Chromatography', *Fresenius' Z. Anal. Chem.*, **328**(1–2), 99–104 (1987).
240. V. Cirimele, P. Kintz, B. Ludes, 'Screening for Forensically Relevant Benzodiazepines in Human Hair by Gas Chromatography–Negative Ion Chemical Ionization Mass Spectrometry', *J. Chromatogr., B, Biomed. Appl.*, **700**(1–2), 119–129 (1997).

241. Y. Gaillard, G. Pepin, 'Screening and Identification of Drugs in Human Hair by High-performance Liquid Chromatography–Photodiode Array UV Detection and Gas Chromatography–Mass Spectrometry After Solid-phase Extraction. A Powerful Tool in Forensic Medicine', *J. Chromatogr., A*, **762**(1–2), 251–267 (1997).
242. M. Yegles, F. Mersch, R. Wennig, 'Detection of Benzodiazepines and Other Psychotropic Drugs in Human Hair by GC/MS', *Forensic Sci. Int.*, **84**(1–3), 211–218 (1997).
243. P. Kintz, A. Tracqui, P. Mangin, 'Sweat Testing for Benzodiazepines', *J. Forensic Sci.*, **41**(5), 851–854 (1996).
244. J. Schuberth, 'Volatile Compounds Detected in Blood of Drunk Drivers by Headspace Capillary Gas Chromatography–Ion-trap Mass Spectrometry', *Biol. Mass Spectrom.*, **20**(11), 699–702 (1991).
245. H. Al Mardini, C.J. O'Brien, K. Bartlett, R. Williams, C.O. Record, 'Thermal Desorption–Gas Chromatography of Plasma Isovaleraldehyde (3-Methylbutyraldehyde) in Hepatic Encephalopathy in Man', *Clin. Chim. Acta*, **165**(1), 61–71 (1987).
246. S.R. Dunn, J.W. Pensabene, M.L. Simenhoff, 'Analysis of Human Blood for Volatile *N*-Nitrosamines by Gas Chromatography–Chemiluminescence Detection', *J. Chromatogr.*, **377**, 35–47 (1986).
247. S.R. Antoine, I.R. DeLeon, R.M. O'Dell-Smith, 'Environmentally Significant Volatile Organic Pollutants in Human Blood', *Bull. Environ. Contam. Toxicol.*, **36**(3), 364–371 (1986).
248. J. Angerer, G. Scherer, K.H. Schaller, J. Mueller, 'Determination of Benzene in Human Blood as an Indicator of Environmental Exposure to Volatile Aromatic Compounds', *Fresenius' J. Anal. Chem.*, **339**(10), 740–742 (1991).
249. P.H. Cramer, K.E. Boggess, J.M. Hosenfeld, J.C. Remmers, J.J. Breen, P.E. Robinson, C. Stroup, 'Determination of (Exogenous) Organic Chemicals in Human Whole Blood: Preliminary Method Development for Volatile Organics', *Bull. Environ. Contam. Toxicol.*, **40**(4), 612–618 (1988).
250. R. Uehori, T. Nagata, K. Kimura, K. Kudo, M. Noda, 'Screening of Volatile Compounds Present in Human Blood Using Retention Indices in Gas Chromatography', *J. Chromatogr.*, **411**, 251–257 (1987).
251. I. Ojanpera, R. Hyppola, E. Vuori, 'Identification of Volatile Organic Compounds in Blood by Purge and Trap PLOT–Capillary Gas Chromatography Coupled with Fourier-transform Infra-red Spectroscopy', *Forensic Sci. Int.*, **33**(1), 201–209 (1996).
252. J. Schuberth, 'A Full Evaporation Headspace Technique with Capillary GC and ITD: A Means for Quantitating Volatile Organic Compounds in Biological Samples', *J. Chromatogr. Sci.*, **34**(7), 314–319 (1996).
253. F. Mangani, R. Cenciarini, 'Solid-phase Microextraction Using Fused-silica Fibers Coated with Graphitized Carbon Black', *Chromatographia*, **41**(11–12), 678–684 (1995).
254. K.-M. Chan, E.T. Wong, W.S. Matthews, 'Severe Isopropanolaemia without Acetoneaemia or Clinical Manifestations of Isopropanol Intoxication', *Clin. Chem.* (Winston-Salem, NC), **39**(9), 1922–1925 (1993).
255. R. Kroneld, 'Recovery and Reproducibility in Determination of Volatile Halocarbons in Water and Blood', *Bull. Environ. Contam. Toxicol.*, **34**(4), 486–496 (1985).
256. P. Gerhards, J. Horst, 'Quantitative Analysis of LHKW (Volatile Halogenated Hydrocarbons) in Blood', *Labor-Praxis*, **16**(4), 340–344 (1992).
257. R.L. Dills, S.D. Kent, H. Checkoway, D.A. Kalman, 'Quantification of Volatile Solvents in Blood by Static Headspace Analysis', *Talanta*, **38**(4), 365–374 (1991).
258. T. Daldrup, H. Haddad, C. Heller, 'Gas Chromatographic Determination of Volatile Anaesthetics (Enflurane and Isoflurane) with Calibration by Proton NMR Spectroscopy', *Fresenius' Z. Anal. Chem.*, **330**(4–5), 457–458 (1988).
259. J. Flynn, J.D. O'Keeffe, W.S. Wren, I.M. Shanahan, 'Gas Chromatographic Determination of Volatile Anaesthetic Agents in Blood. I. Preparation of Standard Gas Mixtures of Volatile Anaesthetic Agents', *Analyst (London)*, **114**(10), 1207–1210 (1989).
260. J. Flynn, S. Masud, J.D. O'Keefe, W.S. Wren, I.M. Shanahan, 'Gas Chromatographic Determination of Volatile Anaesthetic Agents in Blood. II. Clinical Studies', *Analyst (London)*, **114**(10), 1211–1213 (1989).
261. K. Saito, T. Takayasu, J. Nishigami, T. Kondo, M. Ohtsuji, Z. Lin, T. Ohshima, 'Determination of the Volatile Anaesthetics Halothane, Enflurane, Isoflurane, and Sevoflurane in Biological Specimens by Pulse-heating GC/MS', *J. Anal. Toxicol.*, **19**(2), 115–119 (1995).
262. B. Levine, R. Jones, K. Klette, M.L. Smith, E. Kilbane, 'Intoxication Involving BRON and Verapamil', *J. Anal. Toxicol.*, **17**(6), 381–383 (1993).
263. L. Liang, N.S. Bloom, M. Horvat, 'Simultaneous Determination of Mercury Speciation in Biological Materials by Gas Chromatography–Cold Vapor Atomic Fluorescence Spectrometry After Ethylation and Room Temperature Pre-collection', *Clin. Chem.* (Winston-Salem, NC), **40**(4), 602–607 (1994).
264. E.N. Frankel, A.L. Tappel, 'Headspace Gas Chromatography of Volatile Lipid Peroxidation Products from Human Red Blood Cell Membranes', *Lipids*, **26**(6), 479–484 (1991).
265. X.P. Lee, T. Kumazawa, K. Sato, K. Watanabe, H. Seno, O. Suzuki, 'Determination of Solvent Thinner Components in Human Body Fluids by Capillary Gas Chromatography with Trapping at Low Oven Temperature for Headspace Samples', *Analyst (Cambridge)*, **123**(1), 147–150 (1998).
266. R.L. Sheppard, J. Henion, 'Determining EDTA in Blood', *Anal. Chem.*, **69**(15), 477A–480A (1997).

267. P.A. Vollmer, D.C. Harty, N.B. Erickson, A.C. Balhon, R.A. Dean, 'Serum Ethylene Glycol by High-performance Liquid Chromatography', *J. Chromatogr., B, Biomed. Appl.*, **685**(2), 370–374 (1996).
268. D. Bonazzi, R. Gotti, V. Andrisano, V. Cavrini, 'Analysis of ACE Inhibitors in Pharmaceutical Dosage Forms by Derivative UV Spectroscopy and Liquid Chromatography (HPLC)', *J. Pharm. Biomed. Anal.*, **16**(3), 431–438 (1997).
269. C.-K. Lai, T. Lee, K.-M. Au, A.Y.-W. Chan, 'Uniform Solid-phase Extraction Procedure for Toxicological Drug Screening in Serum and Urine by HPLC with Photodiode-array Detection', *Clin. Chem.* (Washington, DC), **43**(2), 312–325 (1997).
270. B.K. Kraemer, K.M. Ress, F. Mayer, V. Kuehlkamp, H.M. Liebich, T. Risler, L. Seipel, 'Rapid High-performance Liquid Chromatographic Method for the Quantification of Mexiletine and its Metabolites in Serum', *J. Chromatogr., Biomed. Appl.*, **85**(2), (*J. Chromatogr.*, **493**), 414–420 (1989).
271. J.A. Menius, Jr, D.J. Schumacher, E.A. Hull-Ryde, C.Y. Leung, R.G. Cummings, J.E. Lowe, 'Quantitation of Amiodarone and De-ethylamiodarone from Blood Serum and Myocardium Using Reversed-phase HPLC', *J. Liq. Chromatogr.*, **10**(12), 2625–2637 (1987).
272. P.T. Pollak, S.G. Carruthers, D.J. Freeman, 'Simplified Liquid Chromatographic Assay of Amiodarone and Desethylamiodarone After Solid-phase Extraction', *Clin. Chem.* (Winston-Salem, NC), **32**(5), 890–893 (1986).
273. B.R. Simmons, J.T. Stewart, 'HPLC Separation of Selected Cardiovascular Agents on Underivatized Silica Using an Aqueous Organic Mobile Phase', *J. Liq. Chromatogr.*, **17**(12), 2675–2690 (1994).
274. R.J. Flanagan, R.K. Bhamra, S. Walker, S.C. Monkman, D.W. Holt, 'Measurement of Cardioactive Drugs in Biological Samples by HPLC', *J. Liq. Chromatogr.*, **11**(5), 1015–1040 (1988).
275. S. Fischer, 'Analysis of Cardiovascular Eicosanoids in Man with Special Reference to HPLC', *Chromatographia*, **22**(7–12), 416–420 (1986).
276. D.W. Jacobsen, 'Cardiovascular Disorders (Risk Assessment)', *Anal. Chem.*, **65**(12), 367R–373R (1993).
277. F. Mangani, 'Cardiovascular Drugs in Serum with Online Sample Clean Up Using Zone-cutting Technique', *Gilson Chromatogr. Appl.*, **7**(1), 5–6 (1996).
278. F. Mangani, G. Luck, C. Fraudeau, E. Verette, 'Online Column-switching High-performance Liquid Chromatography Analysis of Cardiovascular Drugs in Serum with Automated Sample Clean Up and Zone-cutting Technique to Perform Chiral Separation', *J. Chromatogr., A*, **762**(1–2), 235–241 (1997).
279. T. Cleveland, 'Pirkle-concept Chiral Stationary Phases for the HPLC Separation of Pharmaceutical Racemates', *J. Liq. Chromatogr.*, **18**(4), 649–671 (1995).
280. N.R. Musso, C. Vergassola, A. Pende, G. Lotti, 'HPLC with Electrochemical Detection of Catecholamines in Human Plasma. A Mini Review', *J. Liq. Chromatogr.*, **13**(6), 1075–1090 (1990).
281. K. Kojima, H. Parev, S. Parev, T. Nagatsu, 'Microbore HPLC for Biological Samples: Catecholamines, Peptides and Proteins', *Prog. HPLC*, **4**, 211–227 (1989).
282. A. Yoshida, M. Yoshioka, H. Parvez, 'High-performance Liquid Chromatography of Metabolites of Catecholamines and Serotonin in Urine, Plasma, Cerebrospinal Fluid and Brain Tissue. I. Analytical Methodology', *Prog. HPLC, Supercrit. Fluid Chromatogr. Micro-HPLC*, **4**, 229–271 (1989).
283. G.A. Gerhardt, C.J. Drebing, R. Freedman, 'Measurement of Plasma Free Catecholamines and Their Metabolites by High-pressure Liquid Chromatography with Electrochemical Detection', *Neurol. Neurobiol.*, **42C**, *Progr. Catechola.*, 189–196 (1988).
284. J. Zamecnik, 'Quantitation of Epinephrine, Norepinephrine, Dopamine, Metanephrine and Normetanephrine in Human Plasma Using Negative Ion Chemical Ionization GC/MS', *Can. J. Anal. Sci. Spectrosc.*, **42**(4), 106–112 (1997).
285. K.F. Faull, 'Application of Combined Gas Chromatography/Mass Spectrometry to Neurotransmitter Research', *Life Sci.*, **41**(7), 889–892 (1987).
286. F. Bosch Reig, P. Campins Falco, J. Verdu Andres, 'Preconcentration, Sample Clean Up, and HPLC Determination of Phenol and its Chloro-, Methyl-, and Nitro-derivatives in Biological Samples. A Review', *J. Liq. Chromatogr.*, **18**(11), 2229–2241 (1995).
287. J.O. Miners, 'The Analysis of Penicillins in Biological Fluids and Pharmaceutical Preparations by High-performance Liquid Chromatography: A Review', *J. Liq. Chromatogr.*, **8**(15), 2827–2843 (1985).
288. C.M. Moore, K. Sato, H. Hattori, Y. Katsumata, 'Improved HPLC Method for the Determination of Cephalosporins in Human Plasma and a New Solid-phase Extraction Procedure for Cephazolin and Ceftizoxime', *Clin. Chim. Acta*, **190**(1–2), 121–123 (1990).
289. R.D. Toothaker, D.S. Wright, L.A. Pachla, 'Recent Analytical Methods for Cephalosporins in Biological Fluids', *Antimicrob. Agents Chemother.*, **31**(8), 1157–1163 (1987).
290. C.Y. Chan, K. Chan, G.L. French, 'Rapid High-performance Liquid Chromatographic Assay of Cephalosporins in Biological Fluids', *J. Antimicrob. Chemother.*, **18**(4), 537–545 (1986).
291. M. Margosis, 'HPLC of Penicillin Antibiotics', *Adv. Chromatogr. (New York)*, **28**, 333–362 (1989).
292. T.G. Venkateshwaran, J.T. Stewart, 'Determination of Metronidazole in Vaginal Tissue by High-performance Liquid Chromatography Using Solid-phase Extraction', *J. Chromatogr., B, Biomed. Appl.*, **672**(2), 300–304 (1995).

293. C.E. Parker, J.R. Perkin, K.B. Tomer, Y. Shida, K. O'Hara, 'Nanoscale Packed Capillary Liquid Chromatography–Electrospray Ionization Mass Spectrometry: Analysis of Penicillins and Cephems', *J. Chromatogr., Biomed. Appl.*, **127**(1), 45–57 (1993).
294. M.I. Al Ashmawi, M. El Sadek, 'Gas Chromatographic and Gas Chromatographic–Mass Spectrometric Analysis of Penicillin V Benzathine (Benzathine Phenoxymethylpenicillin) Oral Suspension', *Egypt. J. Pharm. Sci.*, **28**(1–4), 361–369 (1987).
295. T.P. Moyer, P. Johnson, S.M. Faynor, S. Sterioff, 'Cyclosporin: A Review of Drug Monitoring Problems and Presentation of a Simple, Accurate Liquid Chromatographic Procedure That Solves These Problems', *Clin. Biochem.*, **19**(4), 83–89 (1986).
296. L.D. Bowers, S.E. Mathews, 'Investigation of the Mechanism of Peak Broadening Observed in the High-performance Liquid Chromatographic Analysis of Cyclosporine', *J. Chromatogr.*, **333**, 231–238 (1985).
297. T.P. Moyer, J.R. Charlson, L.E. Ebnet, 'Improved Chromatography of Cyclosporine', *Ther. Drug Monit.*, **8**(4), 466–468 (1986).
298. T. Nishikawa, H. Hasumi, S. Suzuki, H. Kubo, H. Ohtani, 'Interconversion of Cyclosporin Molecular Form Inducing Peak Broadening, Tailing and Splitting During Reversed-phase Liquid Chromatography', *Chromatographia*, **38**(5–6), 359–364 (1994).
299. N. George, M. Kuppusamy, K. Balaraman, 'Optimization of High-performance Liquid Chromatographic Conditions for the Determination of Cyclosporins A, B and C in Fermentation Samples', *J. Chromatogr.*, **604**(2), 285–289 (1992).
300. D. Fiore, 'Simultaneous HPLC Quantitation of Cyclosporin Homologs with Automated Sample Clean Up and Analysis', *LC-GC*, **4**(7), 656–658 (1986).
301. G. Khoschsorur, H.J. Semmelrock, S. Roedl, T. Auer, W. Petek, F. Iberer, K.H. Tscheliessnigg, 'Rapid, Sensitive High-performance Liquid Chromatographic Method for the Determination of Cyclosporin A and its Metabolites M1, M17 and M21', *J. Chromatogr., B, Biomed. Appl.*, **690**(1–2), 367–373 (1997).
302. W.T. Liu, G.A. Levy, P.Y. Wong, 'Measurement of AM19 and Other Cyclosporine Metabolites in the Blood of Liver Transplant Patients with Stable Liver Function', *Ther. Drug Monit.*, **17**(5), 479–486 (1995).
303. J. Abian, A. Stone, M.G. Morrow, M.H. Creer, L.M. Fink, J.O. Lay Jr, 'Thermospray High-performance Liquid Chromatography–Mass Spectrometric Determination of Cyclosporins', *Rapid Commun. Mass Spectrom.*, **6**(11), 684–689 (1992).
304. T. Nazir, L.A. Gould, C. Marriott, G.P. Martin, M.B. Brown, 'High-performance Liquid Chromatography of a Cyclosporin A Formulation on a Porous Graphitic Carbon Column', *Chromatographia*, **46**(11–12), 628–636 (1997).
305. F. Sadeghipour, E. Varesio, C. Giroud, L. Rivier, J.L. Veuthey, 'Analysis of Amphetamines by Capillary Electrophoresis and Liquid Chromatography: Application to Drug Seizures and Cross Validation', *Forensic Sci. Int.*, **86**(1–2), 1–13 (1997).
306. F. Tagliaro, F.P. Smith, Z. De Battisti, G. Manetto, M. Marigo, 'Hair Analysis, a Novel Tool in Forensic and Biomedical Sciences: New Chromatographic and Electrophoretic/Electrokinetic Analytical Strategies', *J. Chromatogr., B, Biomed. Appl.*, **689**(1), 261–271 (1997).
307. C.M. Selavka, I.S. Krull, 'Forensic Determination of Drugs of Abuse Using Liquid Chromatography with Electrochemical Detection: A Review', *J. Liq. Chromatogr.*, **10**(2–3), 345–375 (1987).
308. R.K. Mueller, J. Grosse, R. Lang, D. Thieme, 'Chromatographic Techniques – the Basis of Doping Control', *J. Chromatogr., B, Biomed. Appl.*, **674**(1), 1–11 (1995).
309. J. Segura, J.A. Pascual, R. Ventura, J.I. Ustaran, A. Cuevas, R. Gonzalez, 'International Co-operation in Analytical Chemistry: Experience of Anti-doping Control at the XI Pan American Games', *Clin. Chem.* (Winston-Salem, NC), **39**(5), 836–845 (1993).
310. D. Catlin, D. Cowan, M. Donike, D. Fraisse, H. Oftebro, S. Rendic, 'Testing Urine for Drugs', *J. Autom. Chem.*, **14**(3), 85–92 (1992).
311. B.A. Goldberger, E.J. Cone, 'Confirmatory Test for Drugs in the Workplace by Gas Chromatography–Mass Spectrometry', *J. Chromatogr., A*, **674**(1–2), 73–86 (1994).
312. H.H. Maurer, 'Systematic Toxicological Analysis of Drugs and Their Metabolites by Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **580**, 3–41 (1992).
313. Z.K. Shihabi, C.P. McCormick, 'Role of HPLC in Therapeutic Drug Monitoring and Drug-of-abuse Screening', *BioChromatography*, **5**(3), 121–126 (1990).
314. A. Turcant, A. Kohn, 'Toxicological Screening by Automated HPLC/UV. Identifying Drugs in Poisonings – Fast', *Hewlett-Packard Peak*, **3**, 2–4 (1994).
315. D.M. Desiderio, J.J. Kusmierz, X. Zhu, C. Dass, D. Hilton, J.T. Robertson, H.S. Sacks, 'Mass Spectrometric Analysis of Opioid and Tachykinin Neuropeptides in Non-secreting and ACTH-secreting Human Pituitary Adenomas', *Biol. Mass Spectrom.*, **22**(1), 89–97 (1993).
316. R.D. Voyksner, D.C. Chen, H.E. Swaisgood, 'Optimization of Immobilized Enzyme Hydrolysis Combined with High-performance Liquid Chromatography Thermospray Mass Spectrometry for the Determination of Neuropeptides', *Anal. Biochem.*, **188**(1), 72–81 (1990).
317. R.D. Voyksner, T.W. Pack, 'Semi-quantitative Method for High-molecular Weight Neuropeptides by High-performance Liquid Chromatography–Thermospray Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **18**(10), 897–903 (1989).
318. W.M. Muck, J.D. Henion, 'Determination of Leucine Enkephalin and Methionine Enkephalin in Equine

Cerebrospinal Fluid by Micro-bore High-performance Liquid Chromatography and Capillary Zone Electrophoresis Coupled to Tandem Mass Spectrometry', *J. Chromatogr.*, **495**, 41–59 (1989).
319. M.T.W. Hearn, M.I. Aguilar, 'High-performance Liquid Chromatography of Amino Acids, Peptides, and Proteins. LXXIII. Investigations on the Relationships Between Molecular Structure, Retention and Band-broadening Properties of Polypeptides Separated by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **397**, 47–70 (1987).
320. A.W. Purcell, M.-I. Aguilar, M.T.W. Hearn, 'High-performance Liquid Chromatography of Amino Acids, Peptides and Proteins. CXXIII. Dynamics of Peptides in Reversed-phase High-performance Liquid Chromatography', *Anal. Chem.*, **65**(21), 3038–3047 (1993).
321. L.S. Monger, C.J. Olliff, 'Evidence of HPLC-induced Conformational Changes in β-Endorphin', *J. Pharm. Pharmacol.*, **42**, Suppl., 17P (1990).
322. L.S. Monger, C.J. Olliff, 'Reversed-phase Chromatographic Behaviour of β-Endorphin: Evidence of Conformational Change', *J. Chromatogr.*, **595**, 125–135 (1992).
323. M. Berchter, J. Meister, C. Hammes, 'MALDI TOF MS: A New Analytical Technique for Characterizing Products Derived from Renewable Resources', *Fett/Lipid*, **99**(11), 384–391 (1997).
324. H. Schulenberg-Schell, B. Neuss, H. Sahm, 'Quantitative Determination of Various Hapanoids (Pentacyclic Triterpenoids) in Micro-organisms', *Anal. Biochem.*, **181**(1), 120–124 (1989).
325. R. Julkunen-Tiitto, 'Chemotaxonomical Screening of Phenolic Glycosides in Northern Willow Twigs by Capillary Gas Chromatography', *J. Chromatogr.*, **324**(1), 129–139 (1985).
326. P.C. Elwood, W.K. Reid, P.D. Marcell, R.H. Allen, J.F. Kolhouse, 'Determination of the Carbohydrate Composition of Mammalian Glycoproteins by Capillary Gas Chromatography–Mass Spectrometry', *Anal. Biochem.*, **175**(1), 202–211 (1988).
327. R.K. Mueller, J. Grosse, R. Lang, D. Thieme, 'Chromatographic Techniques – The Basis of Doping Control', *J. Chromatogr., B, Biomed. Appl.*, **674**(1), 1–11 (1995).
328. P. Vouros, B.L. Karger, 'Combined HPLC/Mass Spectrometry: Its Challenges, Present and Future', *New Methods Drug Res.*, **1**, 45–68 (1985).
329. C.H.L. Shackleton, 'Profiling Steroid Hormones and Urinary Steroids', *J. Chromatogr.*, **379**, 91–156 (1986).
330. L. Bowers, 'Analytical Advances in Detection of Performance Enhancing Compounds', *Clin. Chem.* (Washington, DC), **43**(7), 1299–1304 (1997).
331. V.V. Shkarenda, P.V. Kuznetsov, 'Current State of the Liquid Column Chromatography of Coumarins. II. High-performance Liquid Chromatography of Coumarin Derivatives', *Chem. Nat. Compd. (Engl. Transl.)*, **29**(2), 137–150 (1993).
332. V. Betina, 'Chromatographic Methods as Tools in the Field of Mycotoxins', *J. Chromatogr.*, **477**(2), 187–233 (1989).
333. P.M. Scott, 'Recent Developments in Methods of Analysis for Mycotoxins in Foodstuffs', *Trends Anal. Chem. (Pers. Ed.)*, **12**(9), 373–381 (1993).
334. C. Kutzer, B. Branner, W. Zwickenpflug, E. Richter, 'Simultaneous Solid-phase Extraction and Gas Chromatographic–Mass Spectrometric Determination of Haemoglobin Adducts from Tobacco-specific Nitrosamines and Aromatic Amines', *J. Chromatogr. Sci.*, **35**(1), 1–6 (1997).
335. S.G. Carmella, S.S. Hecht, 'Study of Chemical Carcinogenesis. LXXII. High-performance Liquid Chromatographic Analysis of Metabolites of the Nicotine-derived Nitrosamines N′-Nitrosonornicotine and 4-(Methylnitrosamino)-1-(3-pyridyl)butan-1-one', *Anal. Biochem.*, **145**(2), 239–244 (1985).
336. N.P. Sen, S.W. Seaman, S.C. Kushwaha, 'Improved Method for Determination of Volatile Nitrosamines in Baby Bottle Rubber Nipples and Pacifiers', *J. Assoc. Off. Anal. Chem.*, **70**(3), 434–438 (1987).
337. H.C. Thomas, Jr, S.M. Billedeau, B.J. Miller, 'Gas Chromatographic – Thermal Energy Analysis Method for *N*-Nitrosodibutylamine in Latex Infant Pacifiers: Collaborative Study', *J. Assoc. Off. Anal. Chem.*, **69**(3), 504–507 (1986).
338. R.A. Gates, J.W. Pensabene, W. Fiddler, 'Comparison of Three Methods for Determination of *N*-Nitrosopyrrolidine in Fried Dry-cured Bacon', *J. Assoc. Off. Anal. Chem.*, **67**(2), 236–239 (1984).
339. W. Fiddler, J.W. Pensabene, W.I. Kimoto, 'Investigation of Volatile Nitrosamines in Disposable Protective Gloves', *Am. Ind. Hyg. Assoc. J.*, **46**(8), 463–465 (1985).
340. H. Kawasaki, S. Ninomiya, H. Yuki, 'High-performance Liquid Chromatographic Determination of 3-Methylcytosine in Deoxyribonucleic Acid Treated with Carcinogenic Methylating Agents in vitro and in vivo', *Chem. Pharm. Bull.*, **33**(3), 1170–1174 (1985).
341. C.W. Gerhke, K.C. Kuo, 'Ribonucleoside Analysis by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **471**, 3–36 (1989).
342. K. Nakano, T. Yasaka, K.H. Schram, M.L.J. Reimer, T.D. McClure, T. Nakao, H. Yamamoto, 'Isolation and Identification of Urinary Nucleosides. Applications of High-performance Liquid Chromatographic Methods to the Synthesis of 5′-Deoxyxanthosine and the Simultaneous Determination of 5,6-Dihydrouridine and Pseudouridine', *J. Chromatogr.*, **515**, 537–546 (1990).
343. M. Reti, E. Kutejova, 'HPLC and FP (Fast Protein) LC of Oligonucleotides, DNA Fragments, Peptides and Proteins', *Biologia (Bratislava)*, **45**(4), 341–357 (1990).
344. E.L. Esmans, D. Broes, I. Hoes, F. Lemiere, K. Vanhoutte, 'Liquid Chromatography–Mass Spectrometry in Nucleotide, Nucleotide and Modified Nucleotide Characterization', *J. Chromatogr., A*, **794**, 1–2 (1998).

345. H.K. Tewary, P.L. Iversen, 'Qualitative and Quantitative Measurements of Oligonucleotides in Gene Therapy: Part I. In Vitro Models', *J. Pharm. Biomed. Anal.*, **15**(7), 857–873 (1997).
346. Y. Baba, 'Prediction of the Behaviour of Oligonucleotides in High-performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr., Biomed. Appl.*, **129**(1–2), 41–55 (1993).
347. T. Grune, W.G. Siems, 'Reversed-phase High-performance Liquid Chromatography of Purine Compounds for Investigation of Biomedical Problems: Application to Different Tissues and Body Fluids', *J. Chromatogr., Biomed. Appl.*, **129**(1–2), 15–40 (1993).
348. R.C. Simpson, P.R. Brown, 'High-performance Liquid Chromatographic Profiling of Nucleic Acid Components in Physiological Samples', *J. Chromatogr.*, **379**, 269–311 (1986).
349. D. Perrett, 'Some Applications of High-performance Liquid Chromatography in the Field of Nucleotides, Nucleosides and Bases', *Biochem. Soc. Trans.*, **13**(6), 1067–1070 (1985).
350. W.G. Stillwell, H.-X. Xu, J.A. Adkins, J.S. Wishnok, S.R. Tannanbaum, 'Analysis of Methylated and Oxidized Purines in Urine by Capillary Gas Chromatography/Mass Spectrometry', *Chem. Res. Toxicol.*, **2**(2), 94–99 (1989).
351. C.W. Keuzenkamp-Jansen, J.M. van Baal, R.A. De Abreu, J.G.N. de Jong, R. Zuiderent, J.M.F. Trijbels, 'Detection and Identification of 6-Methylmercapto-8-hydroxypurine, a Major Metabolite of 6-Mercaptopurine, in Plasma During Intravenous Administration', *Clin. Chem.* (Washington, DC), **42**(3), 380–386 (1996).
352. K. Nakano, T. Nakao, K.H. Schram, W.M. Hammargren, T.D. McClure, M. Katz, E. Petersen, 'Urinary Excretion of Modified Nucleosides as Biological Marker of RNA Turnover in Patients with Cancer and AIDS', *Clin. Chim. Acta*, **218**(2), 169–183 (1993).
353. W.M. Hammargren, K.H. Schram, K. Nakano, T. Yasaka, 'Identification of a Novel Nucleotide, $1,N_6$-Dimethyladenosine, in Human Cancer Urine', *Anal. Chim. Acta*, **247**(2), 201–209 (1991).
354. R.A. Sherwood, 'Measurement of Nucleotide Deaminases by High-performance Liquid Chromatography and Their Use in Clinical Chemistry', *Biomed. Chromatogr.*, **5**(6), 235–239 (1991).
355. R.O. Williams III, V. Mahaguna, M. Sriwongjanya, 'Determination of Diazolidinyl Urea in a Topical Cream by High-performance Liquid Chromatography', *J. Chromatogr., B, Biomed. Appl.*, **696**(2), 303–306 (1997).
356. U.P. Halkar, S.H. Rane, N.P. Bhandari, 'Simultaneous Determination of Cephalexin and Probenecid in Pharmaceutical Preparations by Reversed-phase HPLC', *Indian Drugs*, **34**(9), 539–541 (1997).
357. C.W. Chang, S.L. Hsiu, P.P. Wu, S.C. Kuo, P.D.L. Chao, 'HPLC Assays of Naringin and Hesperidin in Chinese Herbs and Serum', *Yaowu Shipin Fenxi*, **5**(2), 111–120 (1997).
358. O. Al-Deeb, B.M. Al-Hadiya, N.H. Foda, 'Quantitative Analysis of Mebeverine in Dosage Forms by HPLC', *Chromatographia*, **44**(7–8), 427–430 (1997).
359. M.A. Abuirjeie, E. Abdel-Hamid, A.A. Abdel-Aziz, E.A. Ibrahim, 'Comparative Study on HPLC and Microbiological Determination of *p*-Hydroxybenzoate Esters (Parabens) in Pharmaceutical Formulations', *Anal. Lett.*, **23**(1), 67–82 (1990).
360. M.J. Akhtar, S. Khan, I.M. Roy, I.A. Jafri, 'High-performance Liquid Chromatographic Determination of Phenoxetol, Methyl Paraben, Ethyl Paraben, *n*-Propyl Paraben, Isobutyl Paraben, *n*-Butyl Paraben and Conconazole Hydrochloride', *J. Pharm. Biomed. Anal.*, **14**(11), 1609–1613 (1996).
361. M.J. Thompson, B.J. Clark, A.F. Fell, M.L. Robinson, 'Liquid Chromatographic Studies on the Potential Degradation of Preservatives in Formulated Drug Products', *Anal. Proc.* (London), **30**(2), 106–109 (1993).
362. K. Stulik, V. Pacakova, J. Suchankova, H.A. Claessens, 'Stationary Phases for Peptide Analysis by High-performance Liquid Chromatography: A Review', *Anal. Chim. Acta*, **352**(1–3), 1–19 (1997).
363. J. Cummings, 'Problems Associated with HPLC Analysis of Natural Product Anticancer Drugs', *Chromatogr. Anal.*, **33**, 5–7 (1994).
364. J.J. Kirkland, B.E. Boyes, J.J. DeStefano, 'Changing Band Spacing in Reversed-phase HPLC: A Technique for Varying Column Stationary-phase Selectivity', *Am. Lab.*, **26**(14), 36–37, 39–42 (1994).
365. F. Thunecke, A. Kalman, F. Kalman, S. Ma, A.S. Rathore, C. Horvath, 'Kinetic Study on the *cis-trans* Isomerization of Peptidyl-proline Dipeptides', *J. Chromatogr., A*, **744**(1–2), 259–272 (1996).
366. E. Krause, B. Bienert, 'Separation of Peptides by Liquid Chromatography on Polyalkylene Columns', *J. Liq. Chromatogr.*, **15**(10),1773–1784 (1992).
367. C. Hirayama, H. Ihara, T. Yoshinaga, H. Hirayama, Y. Motozato, 'Novel Packing for High-pressure Liquid Chromatography. Partially Alkylated and Cross-linked PMLG (Poly(4-Methyl-L-Glutamate)) Spherical Particles', *J. Liq. Chromatogr.*, **9**(5), 945–954 (1986).
368. J.E. Haky, A.R. Raghani, B.M. Dunn, L.F. Wieserman, 'Evaluation of Octadecyl-bonded Alumina for Separations of Proteins and Peptides by Reversed-phase High-performance Liquid Chromatography', *Chromatographia*, **32**(1–2), 49–55 (1991).
369. I.A. Nicholls, O. Ramstrom, K. Mosbach, 'Insights into the Role of the Hydrogen Bond and Hydrophobic Effect on Recognition in Molecularly Imprinted Polymer Synthetic Peptide Receptor Mimics', *J. Chromatogr., A*, **691**(1–2), 349–353 (1995).
370. M. Kempe, K. Mosbach, 'Separation of Amino Acids, Peptides and Proteins on Molecularly Imprinted

Stationary Phases', *J. Chromatogr., A*, **691**(1–2), 317–323 (1995).
371. P. Jadaud, I.W. Wainer, 'Stereochemical Recognition of Enantiomeric and Diastereomeric Dipeptides by High-performance Liquid Chromatography on a Chiral Stationary Phase Based Upon Immobilized α-Chymotrypsin', *J. Chromatogr.*, **476**, 165–174 (1989).
372. C. Hirayama, H. Ihara, K. Tanaka, 'Chromatographic Resolution of Dipeptide Enantiomers and Diastereomers on Chiral Stationary Phases from Poly(L-Leucine) or Poly(L-Phenylalanine)', *J. Chromatogr.*, **450**(2), 271–276 (1988).
373. W.H. Pirkle, R. Daeppen, D.S. Reno, 'Applications for Various Versions of Chiral Stationary Phases', *J. Chromatogr.*, **407**, 211–216 (1987).
374. J. Florance, and Z. Konteatis, 'Chiral High-performance Liquid Chromatography of Aromatic Cyclic Dipeptides Using Cyclodextrin Stationary Phases', *J. Chromatogr.*, **543**(2), 299–305 (1991).
375. B. Esquivel, L. Nicholson, L. Peerey, M. Fazio, 'Enantiomeric Resolution of Underivatized Small Peptides by HPLC with a Chiral Crown Ether Stationary Phase', *J. High Resolut. Chromatogr.*, **14**(12), 816–823 (1991).
376. J.L. Glajch, J.J. Kirkland, J. Koehler, 'Effect of Column Degradation on the Reversed-phase High-performance Liquid Chromatographic Separation of Peptides and Proteins', *J. Chromatogr.*, **384**, 81–90 (1987).
377. M. Kawakatsu, H. Kotaniguchi, H.H. Freiser, K.M. Gooding, 'The Effects of Pore Diameter and Ligand Chain Length on Fast Liquid Chromatography of Proteins and Peptides', *J. Liq. Chromatogr.*, **18**(4), 633–647 (1995).
378. C.D. Marquez, M.-L. Lee, S.T. Weintraub, P.C. Smith, 'Quantitative Analysis of Exogenous Peptides in Plasma Using Immobilized Enzyme Cleavage and Gas Chromatography–Mass Spectrometry with Negative Ion Chemical Ionization', *J. Chromatogr., B, Biomed. Appl.*, **700**(1–2), 9–21 (1997).
379. P. Cao, M. Moini, 'Rapid Derivatization and Gas Chromatography–Mass Spectrometry Analysis of Dipeptides in Aqueous Solutions', *Rapid Commun. Mass Spectrom.*, **11**(4), 349–352 (1997).
380. M. Zollinger, U. Amsler, J. Brauchli, 'Quantification of *N*-Acetylaspartylglutamate, an N-Terminal Blocked Dipeptide Neurotransmitter Candidate, in Brain Slice Superfusates by Gas Chromatography–Mass Spectrometry', *J. Chromatogr.*, **532**(1), 27–36 (1990).
381. R. Cano, 'Characterizing Ancient Bacteria', *Anal. Chem.*, **68**(19), 609A–611A (1996).
382. D.W. Foust, M.S. Bergren, 'Analysis of Solvent Residues in Pharmaceutical Bulk Drugs by Wall-coated Open-tubular Gas Chromatography', *J. Chromatogr.*, **469**, 161–173 (1989).
383. K. Mathys, R. Brenneisen, 'HPLC and TLC Profiles of Phenylalkylamines of Khat (*Catha Edulis* Forsk.) Confiscated in Switzerland', *Pharm. Acta. Helv.*, **68**(2), 121–128 (1993).
384. H.-J. Helmlin, R. Brenneisen, 'Determination of Psychotropic Phenylalkylamine Derivatives in Biological Matrices by High-performance Liquid Chromatography with Photodiode Array Detection', *J. Chromatogr.*, **593**(1–2), 87–94 (1992).
385. K. Murayama, H. Nakamura, K. Takahashi, 'Reversed-phase Separation of Basic Solutes with Alkaline Eluents on an Octadecyl Titania Column', *Anal. Sci.*, **10**(5), 815–816 (1994).
386. H.H. Maurer, 'On the Metabolism and the Toxicological Analysis of Methylenedioxyphenylalkylamine Designer Drugs by Gas Chromatography–Mass Spectrometry', *Ther. Drug Monit.*, **18**(4), 465–470 (1996).
387. X.-C. Zhou, H. Yan, Y.Y. Chen, C.Y. Wu, X.R. Lu, 'Chiral Crown Ether-anchored Polysiloxanes as Capillary Gas Chromatography Stationary Phases', *J. Chromatogr., A*, **753**(2), 269–277 (1996).
388. J.W. Owens, M. Robins, R. Robinson, R. Smith, 'Chromatographic Analysis of Photodynamically Significant Porphyrin Dimers and Trimers', *J. Chromatogr., B, Biomed. Appl.*, **682**(2), 327–336 (1996).
389. J.C.M. Meijers, C.K. Lim, A.M. Lawson, T.J. Peters, 'Analysis of Tumor-localizing Haematoporphyrin Derivative by High-performance Liquid Chromatography and Fast Atom Bombardment Mass Spectrometry', *J. Chromatogr.*, **352**, 231–239 (1986).
390. M. Dellinger, D. Brault, 'Normal-phase High-performance Liquid Chromatography of Free-acid Dicarboxylic Porphyrins and Haematoporphyrin Derivative on Silica', *J. Chromatogr., Biomed. Appl.*, 66 (1987).
391. P.A. Scourides, C. Morstyn, R.W. Henderson, 'Analytical and Preparative Reversed-phase HPLC of Porphyrin C and *NN′*-Diacetylporphyrin C', *J. Liq. Chromatogr.*, **9**(13), 2879–2895 (1986).
392. J. Sliwiok, P. Kus, A. Kozera, 'Chromatographic Characteristics of Porphyrine Ether Glycerides', *Chromatographia*, **34**(5–8), 431–432 (1992).
393. J. Bauer, C. Linton, B. Norris, 'Stability-indicating High-performance Liquid Chromatographic Analysis of Tin Protoporphyrin and Other Free-acid Metalloporphyrins', *J. Chromatogr.*, **445**(2), 429–432 (1988).
394. R.N. Gupta, R.R. Hansebout, 'Optimization of the Determination of 4-Aminopyridine in Human Serum and Urine by Column Liquid Chromatography', *J. Chromatogr., B, Biomed. Appl.*, **677**(1), 183–189 (1996).
395. D. Raphanaud, M. Borensztejn, J.P. Dupeyron, F. Guyon, 'High-performance Liquid Chromatography of Procainamide and *N*-Acetylprocainamide in Human Blood Plasma', *Ther. Drug Monit.*, **8**(3), 365–367 (1986).
396. R. Verbesselt, T.B. Tjandramaga, P.J. De Schepper, 'High-performance Liquid Chromatography Determination of Twelve Antiarrhythmic Drugs in Plasma Using Solid-phase Column Extraction', *Ther. Drug Monit.*, **13**(2), 157–165 (1991).

397. E. VasBinder, T. Annesley, 'Liquid Chromatographic Analysis of Mexiletine in Serum, with Alternate Application to Tocainide, Procainamide and *N*-Acetylprocainamide', *Biomed. Chromatogr.*, **5**(1), 19–22 (1991).
398. H.F. Proelss, T.B. Townsend, 'Simultaneous Liquid Chromatographic Determination of Five Antiarrhythmic Drugs and Their Major Active Metabolites in Serum', *Clin. Chem.* (Winston-Salem, NC), **32**(7), 1311–1317 (1986).
399. J.D. Coyle, J.J. MacKichan, H. Boudoulas, J.J. Lima, 'Reversed-phase Liquid Chromatography Method for Measurement of Procainamide and Three Metabolites in Serum and Urine: Percent of Dose Excreted as De-ethyl Metabolites', *J. Pharm. Sci.*, **76**(5), 402–405 (1987).
400. T.M. Welscher, R.L. McQuinn, S.F. Chang, 'High-performance Liquid Chromatographic Procedure with Fluorescence Detection for the *m-O*-Dealkylated Lactam Metabolite of Flecainide Acetate in Human Plasma', *J. Chromatogr., Biomed. Appl.*, **75**, 2 (1988).
401. D.J. Gisch, R. Ludwig, L.A. Witting, 'Comparison and Characterization of Normal- and Tertiary-butyl Alkyl Bonded Phases for Reversed-phase Liquid Chromatography', *J. High Resolut. Chromatogr.*, **12**(6), 409–410 (1989).
402. T.L. Loys, T.B. Perschy, A.E. Gooding, J.J. Tomlinson, 'Robotic Solid-phase Extraction and High-performance Liquid Chromatographic Analysis of Ranitidine in Serum or Plasma', *Biomed. Chromatogr.*, **6**, 311–316 (1992).
403. W. Lin, P.D. Frade, 'Simultaneous Quantitation of Eight Tricyclic Antidepressants in Serum by High-performance Liquid Chromatography', *Ther. Drug Monit.*, **9**(4), 448–455 (1987).
404. T. Hirai, S. Matsumoto, I. Kishi, 'Simultaneous Analysis of Several Non-steroidal Anti-inflammatory Drugs in Human Urine by High-performance Liquid Chromatography with Normal Solid-phase Extraction', *J. Chromatogr., B, Biomed. Appl.*, **692**(2), 375–388 (1997).
405. S. Wanwimolruk, S.Z. Wanwimolruk, A.R. Zoest, 'Sensitive HPLC Assay for Ketoprofen in Human Plasma and its Application to Pharmacokinetic Study', *J. Liq. Chromatogr.*, **14**(20), 3685–3694 (1991).
406. R.T. Sane, V.J. Banavalikar, M.D. Joshi, V. Nayak, 'High-performance Liquid Chromatographic Determination of Ketoprofen in Pharmaceuticals', *Indian Drugs*, **27**(1), 54–57 (1989).
407. W.H. Pirkle, Y. Liu, 'Incremental Development of Chiral Selectors for Underivatized Profens', *J. Chromatogr., A*, **736**(1–2), 31–38 (1996).
408. R. Lovlin, M. Vakily, F. Jamali, 'Rapid, Sensitive and Direct Chiral High-performance Liquid Chromatographic Method for Ketoprofen Enantiomers', *J. Chromatogr., B, Biomed. Appl.*, **679**(1–2), 196–198 (1996).
409. R.A. Carr, G. Caille, A.H. Ngoc, R.T. Foster, 'Stereospecific High-performance Liquid Chromatographic Assay of Ketoprofen in Human Plasma and Urine', *J. Chromatogr., B, Biomed. Appl.*, **668**(1), 175–181 (1995).
410. F. Pehourcq, F. Lagrange, L. Labat, B. Bannwarth, 'Simultaneous Measurement of Flurbiprofen, Ibuprofen and Ketoprofen Enantiomer Concentrations in Plasma Using L-Leucinamide as the Chiral Coupling Component', *J. Liq. Chromatogr.*, **18**(20), 3969–3979 (1995).
411. N. Mano, Y. Oda, N. Asakawa, Y. Yoshida, T. Sato, T. Miwa, 'Studies of Ovomucoid-, Avidin-, Conalbumin-, and Flavoprotein-conjugated Chiral Stationary Phases for Separation of Enantiomers by High-performance Liquid Chromatography', *J. Chromatogr., A*, **687**(2), 223–232 (1994).
412. J. Hermansson, I. Hermansson, 'Dynamic Modification of the Chiral Bonding Properties of a CHIRAL-AGP Column by Organic and Inorganic Additives. Separation of Enantiomers of Anti-inflammatory Drugs', *J. Chromatogr., A*, **666**(1–2), 181–191 (1994).
413. M. Sinibaldi, M. Flieger, L. Cvak, A. Messina, A. Pichini, 'Direct Resolution of Optically Active Isomers on Chiral Packings Containing Ergoline Skeletons. II. Enantioseparation of Carboxylic Acids', *J. Chromatogr., A*, **666**(1–2), 471–478 (1994).
414. Y. Oda, N. Asakawa, S. Abe, Y. Yoshida, T. Sato, 'Avidin Protein-conjugated Column for Direct Injection Analysis of Drug Enantiomers in Plasma by High-performance Liquid Chromatography', *J. Chromatogr., Biomed. Appl.*, **110**(1–2), 133–141 (1991).
415. M.D. Beeson, G. Vigh, 'Examination of the Retention Behaviour of Underivatized Profen Enantiomers on Cyclodextrin Silica Stationary Phases', *J. Chromatogr.*, **634**(2), 197–204 (1993).
416. K.K. Unger, R. Janzen, G. Jilge, 'Packings and Stationary Phases for Biopolymer Separations by HPLC', *Chromatographia*, **24**, 144–154 (1987).
417. R. Hynek, 'High-performance Liquid Chromatography of Proteins', *Chem. Listy*, **89**(2), 93–99 (1995).
418. N. Cooke, P. Shieh, N. Miller, 'High-performance Hydrophobic-interaction Chromatography of Proteins', *LC-GC*, **7**(12), 954–958 (1989).
419. G.W. Welling, R. Van der Zee, S. Welling-Wester, 'Column Liquid Chromatography of Integral Membrane Proteins', *J. Chromatogr., Biomed. Appl.*, 62 (1987).
420. J. Coupek, I. Vins, 'Hydroxyethyl Methacrylate-based Sorbents for High-performance Liquid Chromatography of Proteins', *J. Chromatogr., A*, **658**(2), 391–398 (1994).
421. E.H.J.M. Jansen, L.A. Ginkel, R.H. Van den Berg, R.W. Stephany, 'High-performance Liquid Chromatographic Separation and Detection Methods for Anabolic Compounds', *J. Chromatogr., Biomed. Appl.*, **118**(1–2), 111–124 (1992).
422. L.A. Van Ginkel, 'Immunoaffinity Chromatography, its Applicability and Limitations in Multi-residue Analysis of Anabolizing and Doping Agents. Review', *J. Chromatogr., Biomed. Appl.*, **102**(2), 363–384 (1991).

423. J.E. Kountourellis, C.K. Markopoulou, K.O. Ebete, J.A. Stratis, 'Separation and Determination of Some Corticosteroids Combined with Bamipine in Pharmaceutical Formulations by High-performance Liquid Chromatography', *J. Liq. Chromatogr.*, **18**(17), 3507–3517 (1995).
424. S. Kurosawa, T. Yoshimura, T. Kurosawa, H. Chiba, K. Kobayashi, T. Koike, M. Tohma, 'Simultaneous Determination of 18-Oxygenated Corticosteroids by High-performance Liquid Chromatography with Fluorescence Detection', *J. Liq. Chromatogr.*, **18**(12), 2383–2396 (1995).
425. C. Valenta, H. Janout, 'Corticosteroid Analysis by HPLC with Increased Sensitivity by Use of Pre-column Concentration', *J. Liq. Chromatogr.*, **17**(5), 1141–1146 (1994).
426. T. Yoshitake, J. Ishida, S. Sonezaki, M. Yamaguchi, 'High-performance Liquid Chromatographic Determination of 3-α,5-β-Tetrahydroaldosterone and Cortisol (Hydrocortisone) in Human Urine with Fluorescence Detection', *Biomed. Chromatogr.*, **6**(5), 217–221 (1992).
427. P. Volin, 'High-performance Liquid-chromatographic Analysis of Corticosteroids', *J. Chromatogr., B, Biomed. Appl.*, **671**(1–2), 319–340 (1995).
428. M.C. Vescina, A. Mamianetti, N.M. Vizioli, S.E. Lucangioli, V. Rodriguez, A. Orden, D. Garrido, C.N. Carducci, 'Evaluation of Faecal Bile-acid Profiles by HPLC After Using Disposable Solid-phase Columns', *J. Pharm. Biomed. Anal.*, **11**(11–12), 1331–1335 (1993).
429. B. Ruan, N. Gerst, G.T. Emmons, J. Shey, G.J. Schroepfer, Jr, 'Sterol Synthesis. A Timely Look at the Capabilities of Conventional and Silver Ion High-performance Liquid Chromatography for the Separation of C27 Sterols Related to Cholesterol Biosynthesis', *J. Lipid Res.*, **38**(12), 2615–2626 (1997).
430. R. Lafont, E.D. Morgan, I.D. Wilson, 'Chromatographic Procedures for Phytoecdysteroids', *J. Chromatogr., A*, **658**(1), 31–53 (1994).
431. E.D. Morgan, M. Pilar Marco, 'Advances in Techniques for Ecdysteroid Analysis', *Invertebr. Reprod. Dev.*, **18**(1–2), 55–56 (1990).
432. K. Gamoh, S. Takatsuto, 'Liquid Chromatographic Assay of Brassinosteroids in Plants', *J. Chromatogr., A*, **658**(1), 17–25 (1994).
433. N. Ikekawa, 'Analysis of Brassinosteroids – Plant Growth-promoting Substances', *Trends Anal. Chem. (Pers. Ed.)*, **9**(10), 337–342 (1990).
434. L.L. Smith, 'Analysis of Oxysterols by Liquid Chromatography', *J. Liq. Chromatogr.*, **16**(8), 1731–1747 (1993).
435. S.U. Sheikh, J.C. Touchstone, 'High-performance Liquid Chromatography of Steroids at Sub-ambient Temperatures', *Chim. Oggi*, **10**, 25–28 (1987).
436. K. Robards, P. Towers, 'Chromatography as a Reference Technique for the Determination of Clinically Important Steroids', *Biomed. Chromatogr.*, **4**(1), 1–19 (1990).
437. R.P. Evershed, M.C. Prescott, L.J. Goad, H.H. Rees, 'High-temperature Gas Chromatography–Mass Spectrometry of Steroid Derivatives', *Biochem. Soc. Trans.*, **15**(1), 175–179 (1987).
438. J.W. Honour, 'Steroid Profiling', *Ann. Clin. Biochem.*, **34**(1), 32–44 (1997).
439. H. Astier, C. Renard, V. Cheminel, O. Soares, C. Mounier, F. Peyron, 'Simultaneous Determination of Pyrimethamine and Sulfadoxine in Human Plasma by High-performance Liquid Chromatography After Automated Liquid–Solid Extraction', *J. Chromatogr., B, Biomed. Appl.*, **698**(1–2), 217–223 (1997).
440. C. Kohl, H. Oelschlaeger, D. Rothley, 'Identification of Drug Glucuronides in Human Urine by (Reverse-phase)-HPLC After Derivatization', *J. Pharm. Biomed. Anal.*, **12**(2), 249–254 (1994).
441. K.P. Shah, M. Chang, C.M. Riley, 'Automated Analytical Systems for Drug Development Studies. II-a System for Dissolution Testing', *J. Pharm. Biomed. Anal.*, **12**(12), 1519–1527 (1994).
442. K. Kurosaka, A. Kuchiki, H. Nakagawa, 'Determination of Sulfur in Pharmaceutical Preparations Using Reversed-phase High-performance Liquid Chromatography', *Chem. Pharm. Bull.*, **39**(8), 2138–2139 (1991).
443. R.S. Dangi, K. Jeevaratnum, K. Sugendran, R.C. Malhotra, C.D. Raghuveeran, 'Solid-phase Extraction and Reversed-phase High-performance Liquid Chromatographic Determination of Sulfur Mustard in Blood', *J. Chromatogr., B, Biomed. Appl.*, **661**(2), 341–345 (1994).
444. J.E. Houglum, R.D. Larson, R.M. Neal, 'Liquid Chromatographic Determination of Sulfamethazine (Sulphadimidine) in Feeds', *J. Assoc. Off. Anal. Chem.*, **71**(5), 1054–1056 (1988).
445. N.P. Milner, M.R. Johnson, K.J. Perry, 'Determination of Sulfadimethoxine and Ormetoprim Residues in Channel Catfish Fillets After Treatment with Romet and Evaluation of a Commercially Available Rapid Diagnostic Test for Drug Residues in Fish Fillets', *J. AOAC Int.*, **77**(4), 875–881 (1994).
446. M. Horie, K. Saito, Y. Hoshino, N. Nose, N. Hamada, H. Nakazawa, 'Identification and Quantitation of Residual Sulfa Drugs in Meat by High-performance Liquid Chromatography with Photodiode Array Detection', *Shokuhin Eiseigaku Zasshi*, **31**(2), 171–176 (1990).
447. M. Patthy, 'Trace Analysis of Sulfaquinoxaline in Animal Tissues by High-performance Liquid Chromatography', *J. Chromatogr., Biomed. Appl.*, **26**, 1 (1983).
448. W.H. Pirkle, M.E. Koscho, Z.P. Wu, 'High-performance Liquid Chromatographic Separation of the Enantiomers of *N*-Aryloxazolinones, *N*-Aryl Thiazolinones and Their Sulfur Derivatives on a Synthetic Chiral Stationary Phase', *J. Chromatogr., A*, **726**(1–2), 91–97 (1996).
449. D.S. Weinberg, M.L. Manier, M.D. Richardson, F.G. Haibach, T.S. Rogers, 'Identification and Quantification of Anticarcinogens in Garlic Extract and Liquorice Root

Extract Powder', *J. High Resolut. Chromatogr.*, **15**(10), 641–654 (1992).
450. K. Glowniak, G. Zgorka, A. Jozefczyk, M. Furmanowa, 'Sample Preparation for Taxol and Cephalomannine Determination in Various Organs of *Taxus* Spp.', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1215–1220 (1996).
451. D.R. Lauren, D.J. Jensen, J.A. Douglas, 'Analysis of Taxol, 10-Deacetylbaccatin III and Related Compounds in *Taxus Baccata*', *J. Chromatogr., A*, **712**(2), 303–309 (1995).
452. S.D. Harvey, J.A. Campbell, R.G. Kelsey, N.C. Vance, 'Separation of Taxol from Related *Taxus Brevifolia* Extracts by Isocratic Elution Reversed-phase Microcolumn High-performance Liquid Chromatography', *J. Chromatogr.*, **587**(2), 300–305 (1991).
453. L. Xu, A. Liu, 'Determination of Taxol in the Extract of *Taxus Chinensis* by Reversed-phase HPLC', *Zhongguo Yaoxue Zazhi*, **24**(7), 552–555 (1989).
454. S.O. Auriola, A.-M. Lepisto, T. Naaranlahti, S.P. Lapinjoki, 'Determination of Taxol by High-performance Liquid Chromatography–Thermospray Mass Spectrometry', *J. Chromatogr.*, **594**(1–2), 153–158 (1992).
455. J.H. Cardellina II, 'HPLC Separation of Taxol and Cephalomannine', *J. Liq. Chromatogr.*, **14**(4), 659–665 (1991).
456. K.M. Witherup, S.A. Look, M.W. Stasko, T.G. McCloud, H.J. Issaq, G.M. Muschik, 'High-performance Liquid Chromatographic Separation of Taxol (A) and Related Compounds from *Taxus Brevifolia*', *J. Liq. Chromatogr.*, **12**(11),2117–2132 (1989).
457. K.C. Chan, A.B. Alvarado, M.T. McGuire, G.M. Muschik, H.J. Issaq, K.M. Snader, 'High-performance Liquid Chromatography and Micellar Electrokinetic Chromatography of Taxol and Related Taxanes from Bark and Needle Extracts of *Taxus* Species', *J. Chromatogr., B, Biomed. Appl.*, **657**(2), 301–306 (1994).
458. T.P. Castor, T.A. Tyler, 'Determination of Taxol in *Taxus Media* Needles in the Presence of Interfering Components', *J. Liq. Chromatogr.*, **16**(3), 723–731 (1993).
459. V. Nemeth-Kiss, E. Forgacs, T. Cserhati, G. Schmidt, 'Determination of Taxol Species Grown in Hungary by High-performance Liquid Chromatography–Diode Array Detection: Effect of Vegetative Period', *J. Chromatogr., A*, **750**(1–2), 253–256 (1996).
460. V. Nemeth-Kiss, E. Forgacs, T. Cserhati, G. Schmidt, 'Taxol Content of Various *Taxus* Species in Hungary', *J. Pharm. Biomed. Anal.*, **14**(8–10), 997–1001 (1996).
461. E. Forgacs, 'Use of Porous Graphitized Carbon Column for Determining Taxol in *Taxus Baccata*', *Chromatographia*, **39**(11–12), 740–742 (1994).
462. W.J. Wong, K.J. Pardue, R.C. Ludwig, D.F. Conklin, K.J. Duff, 'New HPLC Column Offers Longer Lifetime and Improved Resolution of Taxanes', *Supelco Rep.*, **13**(3), 2–3 (1994).
463. R.E.B. Ketchum, D.M. Gibson, 'Rapid Isocratic Reversed-phase HPLC of Taxanes on New Columns Developed Specifically for Taxol Analysis', *J. Liq. Chromatogr.*, **16**(12), 1519–1530 (1993).
464. D.-R. Wu, K. Lohse, H.C. Greenblatt, 'Preparative Separation of Taxol in Normal- and Reversed-phase Operations', *J. Chromatogr., A*, **702**(1–2), 233–241 (1995).
465. Y.G. Wu, W.H. Zhu, 'High-performance Liquid Chromatographic Determination of Taxol and Related Taxanes from *Taxus Callus* Cultures', *J. Liq. Chromatogr. Relat. Technol.*, **20**(19), 3147–3154 (1997).
466. R.E.B. Ketchum, D.M. Gibson, 'A Novel Method of Isolating Taxanes from Cell Suspension Cultures of Yew (*Taxus*)', *J. Liq. Chromatogr.*, **18**(6), 1093–1111 (1995).
467. E.R.M. Wickremesinhe, R.N. Arteca, 'Methodology for the Identification and Purification of Taxol and Cephalomannine (Taxols A and B) from *Taxus Callus* Cultures', *J. Liq. Chromatogr.*, **16**(15), 3263–3274 (1993).
468. F. Bitsch, W. Ma, F. Macdonald, M. Nieder, C.H.L. Shackleton, 'Analysis of Taxol and Related Diterpenoids from Cell Cultures by Liquid Chromatography–Electrospray Mass Spectrometry', *J. Chromatogr. Biomed. Appl.*, **126**(2), 273–280 (1993).
469. A. Sparreboom, O. van Tellingen, W.J. Nooijen, J.H. Beijnen, 'Determination of Paclitaxel and Metabolites in Mouse Plasma, Tissues, Urine and Faeces by Semi-automated Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr., B, Biomed. Appl.*, **664**(2), 383–391 (1995).
470. A. Sharma, W.D. Conway, R.M. Straubinger, 'Reversed-phase High-performance Liquid Chromatographic Determination of Taxol in Mouse Plasma', *J. Chromatogr., B, Biomed. Appl.*, **655**(2), 315–319 (1994).
471. J. Leslie, J.M. Kujawa, N. Eddington, M. Egorin, J. Eiseman, 'Stability Problems with Taxol (A) in Mouse Plasma During Analysis by Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **11**(11–12), 1349–1352 (1993).
472. M. Andreeva, P.D. Niedmann, L. Binder, V.W. Armstrong, H. Meden, M. Binder, M. Oellerich, 'A Simple and Reliable Reversed-phase High-performance Liquid Chromatographic Procedure for Determination of Paclitaxel (Taxol) in Human Serum', *Ther. Drug Monit.*, **19**(3), 327–332 (1997).
473. G. Theodoridis, R. Verpoorte, 'Taxol Analysis by High-performance Liquid Chromatography: A Review', *Phytochem. Anal.*, **7**(4), 169–184 (1996).
474. A. El-Yazigi, A. Yusuf, 'Expedient Liquid Chromatographic Assay for Paclitaxel in Plasma After its Administration to Cancer Patients', *Ther. Drug Monit.*, **17**(5), 511–515 (1995).
475. I. Royer, P. Alvinerie, J.P. Armand, L.K. Ho, M. Wright, B. Monsarrat, 'Paclitaxel Metabolites in Human Plasma and Urine: Identification of 6-α-Hydroxytaxol, 7-Epitaxol and Taxol Hydrolysis Products Using Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass

Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(6), 495–502 (1995).
476. M.T. Huizing, H. Rosing, F. Koopman, A.C.F. Keung, H.M. Pinedo, J.H. Beijnen, 'High-performance Liquid Chromatographic Procedures for the Quantitative Determination of Paclitaxel (Taxol) in Human Urine', *J. Chromatogr., B, Biomed. Appl.*, **664**(2), 373–382 (1995).
477. D. Song, J.L.-S. Au, 'Isocratic High-performance Liquid Chromatographic Assay of Taxol in Biological Fluids and Tissues Using Automated Column Switching', *J. Chromatogr., B, Biomed. Appl.*, **663**(2), 337–344 (1995).
478. R.A. Newman, A. Fuentes, T.Y. Minor, K.T. McManus, D.A. Garteiz, 'Determination of Echinomycin (NSC-526417) in Human Plasma: Comparison of Conventional HPLC to a Capillary HPLC, Electrospray Ionization and Mass Spectrometry System', *J. Liq. Chromatogr.*, **17**(2), 403–417 (1994).
479. T.A. Willey, E.J. Bekos, R.C. Gaver, G.F. Duncan, L.K. Tay, J.H. Beijnen, R.H. Farmen, 'High-performance Liquid Chromatographic Procedure for the Quantitative Determination of Paclitaxel (Taxol) in Human Plasma', *J. Chromatogr., Biomed. Appl.*, **132**(2), 231–238 (1993).
480. J.C. Vergniol, R. Bruno, G. Montay, A. Frydman, 'Determination of Taxotere in Human Plasma by a Semi-automated High-performance Liquid Chromatographic Method', *J. Chromatogr., Biomed. Appl.*, **120**(1–2), 273–278 (1992).
481. J. Rizzo, C. Riley, D. Von Hoff, J. Kuhn, J. Phillips, T. Brown, 'Analysis of Anticancer Drugs in Biological Fluids: Determination of Taxol with Application to Clinical Pharmacokinetics', *J. Pharm. Biomed. Anal.*, **8, 2**, 159–164 (1990).
482. S.M. Longnecker, R.C. Donehower, A.E. Cates, T.-L. Chen, R.B. Brundrett, L.B. Grochow, D.S. Ettinger, M. Colvin, 'High-performance Liquid Chromatographic Assay for Taxol in Human Plasma and Urine and Pharmacokinetics in a Phase I Trial', *Cancer Treat. Rep.*, **71**(1), 53–59 (1987).
483. S. Moellenbeck, T. Koenig, P. Schreier, W. Schwab, J. Rajaonarivony, L. Ranarivelo, 'Chemical Composition and Analyses of Enantiomers of Essential Oils from Madagascar', *Flavour Fragrance J.*, **12**(2), 63–69 (1997).
484. F. Chialva, C. Frattini, A. Martelli, 'Unusual Essential Oils with Aromatic Properties. III. Volatile Components of Gentian Roots', *Z. Lebensm.-Unters.-Forsch.*, **182**(3), 212–214 (1986).
485. T.J. Betts, 'Cholesteryl Acetate as a Stationary Phase for the Gas Chromatography of Some Volatile Oil Constituents', *J. Chromatogr.*, **600**(2), 337–340 (1992).
486. C. Bicchi, A. D'Amato, V. Manzin, A. Galli, M. Galli, 'Cyclodextrin Derivatives in the Gas Chromatographic Separation of Racemic Mixtures of Volatile Compounds. VII. The Use of 2,6-Di-*O*-methyl-3-*O*-pentyl-β-cyclodextrin Diluted in Phases with Different Polarity in the Separation of Racemates in Complex Mixtures', *J. Chromatogr., A*, **666**(1–2), 137–146 (1994).
487. T.J. Betts, 'Use of Three Molecularly Toroid Phases for the Gas Chromatography of Some Volatile Oil Constituents, and Comparison with Liquid Crystal Phases', *J. Chromatogr.*, **606**(2), 281–284 (1992).
488. Y. Van der Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, 'Ruggedness Tests on the High-performance Liquid Chromatography Assay of the United States Pharmacopeia XXII for Tetracycline Hydrochloride. A Comparison of Experimental Designs and Statistical Designs Interpretations', *Anal. Chim. Acta*, **312**(3), 245–262 (1995).
489. Y. Van der Heyden, D.L. Massart, Y. Zhu, J. Hoogmartens, J. De Beer, 'Ruggedness Tests on the High-performance Liquid Chromatography Assay of the United States Pharmacopeia 23 for Tetracycline Hydrochloride: Comparison of Different Columns in an Inter-laboratory Approach', *J. Pharm. Biomed. Anal.*, **14**(8–10), 1313–1326 (1996).
490. C.R. White, W.A. Moats, K.L. Kotula, 'Comparative Study of High-performance Liquid Chromatographic Methods for the Determination of Tetracycline Antibiotics', *J. Liq. Chromatogr.*, **16**(13), 2873–2890 (1993).
491. W. Naidong, J. De Beer, X. Marcelis, P. Derese, J.H.M. Miller, J. Hoogmartens, 'Collaborative Study on the Analysis of Chlortetracycline Hydrochloride by Liquid Chromatography on Polystyrene–Divinylbenzene Packing Materials', *J. Pharm. Biomed. Anal.*, **10**(2–3), 199–204 (1992).
492. W. Naidong, E. Roets, J. Hoogmartens, 'High-performance Liquid Chromatography of Chlortetracycline and Related Substances on Poly(Styrene–Divinylbenzene) Copolymer', *Chromatographia*, **30**(1–2), 105–109 (1990).
493. B.W. LeDuc, D.J. Greenblatt, H. Friedman, 'Automated High-performance Liquid Chromatographic Analysis of Tetracycline in Urine', *J. Chromatogr.*, **490**(2), 474–477 (1989).
494. K. Iwaki, N. Okumura, M. Yamazaki, 'Rapid Determination of Tetracycline Antibiotics in Serum by Reversed-phase High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr.*, **619**, 319–323 (1993).
495. E. Brandsteterova, P. Kubalec, L. Bovanova, P. Simko, A. Bednarikova, L. Machackova, 'SPE and MSPD as Pre-separation Techniques for HPLC of Tetracyclines in Meat, Milk and Cheese', *Z. Lebensm.-Unters.-Forsch. A*, **205**(4), 311–315 (1997).
496. S. Tanase, H. Tsuchiya, J. Yao, S. Ohmoto, N. Takagi, S. Yoshida, 'Reversed-phase Ion-pair Chromatographic Analysis of Tetracycline Antibiotics. Application to Discoloured Teeth', *J. Chromatogr., B, Biomed. Appl.*, **706**(2), 279–285 (1998).
497. F.S. Messiha, 'Determination of Carbamazepine by HPLC Electrochemical Detection and Application for Estimation of Imipramine, Desipramine, Doxepin and Nordoxepin', *Alcohol*, **3**(2), 135–138 (1986).

498. R. Gill, S.O. Wanogho, 'High-performance Liquid Chromatographic System for the Separation of Tricyclic Antidepressants and Related Drugs Using ODS-Hypersil', *J. Chromatogr.*, **391**(2), 461–464 (1987).
499. W.A. Dunn, L.A. Lockrey, M.W. McCain, T.J. Siek, 'Report of a Suicide Involving Digoxin and Doxepin', *J. Anal. Toxicol.*, **18**(2), 122–123 (1994).
500. M.P. Segatti, G. Nisi, F. Grossi, M. Mangiarotti, C. Lucarelli, 'Rapid and Simple High-performance Liquid Chromatographic Determination of Tricyclic Antidepressants for Routine and Emergency Serum Analysis', *J. Chomatogr.*, **536**(1–2), 319–325 (1991).
501. R.J. Hughs, M.D. Osselton, 'Comparison of Methods for the Analysis of Tricyclic Antidepressants in Small Whole Blood Samples', *J. Anal. Toxicol.*, **13**(2), 77–83 (1989).
502. A.L. Fisher, W.F. Bayne, D.G. Musson, 'Sensitive Method for Assay of a Novel Tricyclic Compound Using Coulometric Electrochemical Detection', *J. Pharm. Sci.*, **78**(11), 971–974 (1989).
503. G.L. Lensmeyer, D.A. Wiebe, B.A. Darcey, 'Application of a Novel Form of Solid-phase Sorbent (Empore Membrane) to the Isolation of Tricyclic Antidepressant Drugs from Blood', *J. Chromatogr. Sci.*, **29**(10), 444–449 (1991).
504. G. Carfagnini, A. Di Corcia, M. Marchetti, R. Samperi, 'Antidepressants in Serum Determined by Isolation of Two Online Sorbent Cartridges and Liquid Chromatography', *J. Chromatogr.*, **530**(2), 359–366 (1990).
505. M.A. Evenson, D.A. Engstrand, 'Sep-Pak HPLC Method for Tricyclic Antidepressant Drugs in Human Vitreous Humour', *J. Anal. Toxicol.*, **13**(6), 322–325 (1989).
506. W. Lin, P.D. Frade, 'Simultaneous Quantitation of Eight Tricyclic Antidepressants in Serum by High-performance Liquid Chromatography', *Ther. Drug Monitor.*, **9**(4), 448–455 (1987).
507. R.C. Parry, G.D. Wachob, 'Weak Cation-exchange Solid-phase Extraction of Tricyclic Antidepressant Drugs from Serum', *LC-GC*, **5**(6), 498–500 (1987).
508. E.G. Lovering, N. Beaulieu, R.C. Lawrence, R.W. Sears, 'Liquid Chromatographic Method for Identity, Assay and Content Uniformity of Five Tricyclic Drugs', *J. Assoc. Off. Anal. Chem.*, **68**(2), 168–171 (1985).
509. AH.J. Issaq, J. Gutierrez, 'Mixed Packings in High-performance Liquid Chromatography. II. Mixed Packings vs. Mixed Ligands', *J. Liq. Chromatogr.*, **11**(14), 2851–2861 (1988).
510. J.J. Kirkland, J.W. Henderson, 'Reversed-phase HPLC Selectivity and Retention Characteristics of Conformationally Different Bonded Alkyl Stationary Phases', *J. Chromatogr. Sci.*, **32**(11), 473–480 (1994).
511. D.T. Witte, J. Bosman, T. De Boer, B.F.H. Drenth, K. Ensing, R.A. De Zeeuw, 'Influence of Chemical Structure of Tricyclic Tertiary Dimethylamines on Chiral Separation by Reversed-phase High-performance Liquid Chromatography After Derivatization with (−)-Menthyl Chloroformate', *J. Chromatogr.*, **553**(1), 365–372 (1991).
512. X.-P. Lee, T. Kumazawa, K. Sato, O. Suzuki, 'Detection of Tricyclic Antidepressants in Whole Blood by Headspace Solid-phase Microextraction and Capillary Gas Chromatography', *J. Chromatogr. Sci.*, **35**(7), 302–308 (1997).
513. O. Papp, G. Szasz, J. Kokosi, I. Hermecz, 'Correlation Between Structure and Gas Chromatographic Behaviour of Nitrogen-containing Heterocyclic Compounds. III. Variation of Ring Size', *J. Chromatogr.*, **537**(1–2), 377–383 (1991).
514. H. Hattori, E. Takashima, T. Yamada, O. Suzuki, 'Detection of Tricyclic Antidepressants in Body Fluids by Gas Chromatography with a Surface Ionization Detector', *J. Chromatogr.*, **529**(1), 189–193 (1990).

Mass Spectrometry in Pharmaceutical Analysis

David R. Goodlett
Goodlett Scientific, Seattle, USA

David C. Gale
Roche Bioscience, Palo Alto, USA

Stacey Guiles
Hershey Medical School, Hershey, USA

Jonathan B. Crowther
Janssen Research, Beerse, Belgium

1 Introduction **2**
2 Biological Mass Spectrometry **2**
2.1 Mass Spectrometers 2
2.2 Electrospray Ionization 3
2.3 Electrospray Ionization/Quadrupole Instrumentation 3
2.4 Electrospray Ionization/Ion Trap Instrumentation 5
3 Sample Handling Considerations **5**
3.1 Sample Handling for High Sensitivity 5
3.2 Chromatographic Considerations 6
3.3 Electrospray Ionization Advantages 6
4 General Laboratory **6**
4.1 Following Good Laboratory Practice 6
4.2 Automation 7
4.3 Managing Instrument Time 8
5 Quantitation of Peptides by Electrospray Ionization/Mass Spectrometry **8**
5.1 Analysis of Peptides in Physiological Fluids 8
5.2 Tandem Liquid Chromatography/Mass Spectrometry Quantitation of Peptides in Urine 9
5.3 Discussion 11
6 Chiral Analysis **11**
6.1 Chiral Amino Acid Analysis 11
6.2 Liquid Chromatography/Mass Spectrometry Chiral Purity 12
6.3 Discussion 13
7 Comparative Tandem Mass Spectrometry **13**
7.1 Background on Collision-induced Dissociation of Peptides 13
7.2 Chiral Analysis of Peptides by Tandem Mass Spectrometry 14
7.3 Discussion 14
8 Nonconventional Mass Spectrometry **16**
8.1 Noncovalent Complex Analysis 16
8.2 Thermal Denaturation of Noncovalent Complexes 17
8.3 Discussion 17
9 Conclusions **18**
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Mass spectrometry (MS) applied to the biological sciences in the pharmaceutical industry covers a broad range of application topics, such as quantitative analysis of drugs in physiological fluids, qualitative analysis of noncovalent complexes, structural analysis to determine chirality and protein sequence analysis. Regardless of the application, electrospray ionization (ESI) is the primary ionization method for transfer of nonvolatile biologics like proteins and peptides into the gas phase for analysis. It is most often used as an interface between high-performance liquid chromatography (HPLC) separation methods and MS. The ESI process converts analytes in solution to ions in the gas phase via electrostatic nebulization at atmospheric pressure. It is characteristic of ESI to produce multiply charged analytes in a concentration-dependent manner that is sensitive to changes in solvent, pH and ionic strength. A high voltage is applied to a liquid containing analytes as they traverse a capillary, and when the liquid droplets exit this capillary at atmospheric pressure they are directed into the source region of a mass spectrometer where desolvation is completed. Ions are then focused by an electronic gradient across a series of lenses and separated by mass/charge (m/z) using various mass spectrometer designs. Currently, the two most popular mass separation devices are quadrupoles (either single or triple quadrupole (TQ)) and ion traps (ITs). Both types of mass spectrometers are used with ESI to achieve sensitivities as low as attomoles (10^{-18} mol) of analyte injected onto a microcapillary HPLC column. The type of analyte, as long as it is capable of accepting a charge (addition of H^+ or loss of charge in negative ion mode) is relatively transparent to ESI and so a diverse range of analyte classes with disparate molecular weights can be ionized, from peptides (500–5000 amu) to proteins (10 000–100 000 amu). Depending on the construction of the source, ESI can be carried out with reversed-phase HPLC at flow rates from 200 nL min^{-1} to 2 mL min^{-1} and also with capillary electrophoresis. The ease of use of ESI compared to prior techniques like fast atom bombardment

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

has made it very popular with biologists, chromatographers and chemists who might not otherwise have been interested in MS. In fact, for many scientists ESI has made MS almost as ''user-friendly'' and thus accessible as an ultraviolet (UV)–visible detector for HPLC.

1 INTRODUCTION

The modern pharmaceutical laboratory utilizes MS for many purposes, including basic biological research, quantitative analysis, qualitative analysis and routine monitoring for quality control of products. Before the mid-1980s and fast atom bombardment[1,2] ionization, MS was mainly used for the quantitation of small volatile organic molecules and structural analysis via chemical and electron impact ionization. The incorporation of MS into all aspects of pharmaceutical research was made possible by two developments in the late 1980s, specifically the development of two ionization methods, which revolutionized the way in which biological analytical sciences were conducted. These two ionization methods or techniques – ESI[3] and matrix-assisted laser desorption ionization (MALDI)[4] – gained immediate popularity because of their ability to ionize, without significant fragmentation or prior derivatization of analytes, nonvolatile biological molecules such as proteins and peptides. For many laboratories ESI quickly replaced fast atom bombardment as the interface between HPLC and MS,[5] owing to the ease with which it could be adapted to scanning instruments such as quadrupoles. Furthermore, the ESI process allowed simple quadrupole instruments with a limited m/z range to detect analytes outside the m/z range of the instrument, because of the addition of multiple charges to analytes during the ESI process.

Qualitative MS is often used to help biologists gain insight into biological processes or systems. Understanding noncovalent, intermolecular interactions between two macromolecules, between a small molecule and a macromolecule or simply mapping disulfide bonds within a protein can be performed much faster with MS than traditional methods.[6] Studies similar to these can provide details about how a receptor molecule may bind a ligand, providing important information prior to crystal growth for crystallographic studies. For instance, cross-linking a ligand to a receptor, followed by enzymatic digestion and MS analysis, provides information about sites juxtaposed under native conditions that is valuable for the interpretation of crystallographic data. If the receptor–ligand complex cannot be crystallized, then the MS data may be the only route to understanding the interaction.

Quantitative studies have been, and continue to be, an important application of MS. With the advent of ESI, the variety of molecules accessible for analysis became much wider than when electron impact, chemical ionization and fast atom bombardment were the only good choices for ionization. Generally, MS has been used to quantitate small organic molecules (<500 Da) in plasma and urine. However, even large molecules such as proteins can, with an appropriate standard, be quantitated using ESI or MALDI without enzymatic digestion to peptides. Quantitation of peptides in plasma and urine is remarkably similar to the analysis of small molecules from these matrices. Peptides, however, generally suffer from poor pharmacokinetics after oral administration, which is generally due to the short half-life observed for many peptides.[7] This is often controlled through modification of the structure to prevent rapid digestion by proteolytic enzymes present in the gut and circulatory system.[8]

MS now is also used for routine sequencing of proteins for the drug discovery process. With the advent of DNA and protein databases, MS has replaced Edman chemistry as the method of choice for sequence analysis for all but complete unknowns. The MS sequence methods rely on correlative analysis between tandem mass spectrometry (MS/MS) data that is analyzed by computer algorithms such as SEQUEST[9] and sequences that exist in databases. Edman chemistry[10] will, however, continue to be important for sequence analysis of organisms for which no genome sequence is available and for post-translational modification such as phosphorylation of certain peptide sequences that are intractable to MS methods.

In addition to the use of MS in basic research, the application of this technology for quality control is important. Methods that allow for rapid assessment of drug substance batch-to-batch quality are routinely devised, as are methods for the analysis of peptide or protein purity. After a review of the basic MS instrumentation currently in use and some general aspects of working in an analytical laboratory, examples will be provided for MS applications in the quantitative analysis of peptides in urine, chiral analysis of peptides, comparative MS/MS analysis and thermal denaturation of DNA complexes. All of these topics are of widespread interest and constitute ongoing areas of analytical development in the pharmaceutical science. For many of these applications, the type of analyte used is more or less transparent to the analytical process and thus is useful for small molecules as well as peptides and proteins.

2 BIOLOGICAL MASS SPECTROMETRY

2.1 Mass Spectrometers

At the simplest level, MS consists of three processes: ionization (e.g. MALDI or ESI); mass (m/z) separation

(e.g. time-of-flight (TOF), quadrupole, IT); and ion detection. Ion detection, being mostly MS instrument independent, will not be covered here. Analytes in solution or as crystals on a surface are converted to ions by the ionization process where protons/electrons are added or subtracted to impart some detectable m/z value. Ions are then separated from each other in vacuo based on a process that is dependent on the type of mass spectrometer being used. For protein and peptide applications, the four most common types of mass spectrometers utilized are MALDI coupled to a TOF mass spectrometer and ESI coupled to a TQ, a single quadrupole (SQ) or an IT mass spectrometer. The applications presented in this chapter will focus on ESI and so MALDI coupled with a TOF mass spectrometer will not be reviewed directly.

2.2 Electrospray Ionization

In ESI, the analyte solution passes through a needle to which a high potential (~1000–5000 V) is applied at atmospheric pressure.[11,12] This results in electrostatic nebulization of the liquid, producing charged liquid droplets that are directed to the mass spectrometer inlet, which is at some lower potential than the ESI voltage. As the charged droplets traverse the source region they are desolvated by various processes, depending on the instrument design. The process by which analytes in solution are converted to ions in the gas phase continues to be a topic of debate among research laboratories and will not be discussed further. After desolvation, solvent-free ions are focused through a series of lenses into the mass analyzer. It is characteristic of the ESI process to produce peptide ions bearing multiple charges. The number of observed charge states is a function of the number of ionizable groups contained in the analyte. The observed distribution of charge states can be affected by changes in pH, solvent and tuning conditions.[13,14] For instance, tryptic peptides analyzed by positive ion mode electrospray ionization/mass spectrometry (ESI/MS) at a pH of 3.5 will normally carry at least two charges.[15] This is because the hydrolysis of a peptide bond by trypsin leaves a lysine or arginine amino acid at the carboxyl-terminus of the peptide. In general, peptides and proteins will bear a number of charges in positive ion mode roughly equivalent to the number of basic residues, such as the amino acids lysine, arginine, histidine and proline, present in the sequence. If the protein or peptide is of unknown molecular weight, adjacent charge states can be used to determine the molecular weight. Most instruments come with software to do this deconvolution but it is based on simple algebra. To determine molecular weight from the mass spectrum of an unknown analyte we can use the mass spectrum in Figure 1 as an example. Choose two adjacent charge states suspected of being related and fill in the following equation: $(m_2)/(m_2 - m_1) = (649)/(649 - 433) = 3$. If this algebra produces an integer, as in this case, then the integer represents the number of charges (e.g. in positive ion mode protons or sodium ions) on m_1; i.e. 433 is an $[M + 3H]^{3+}$ ion. If $(m_2)/(m_2 - m_1)$ produces a noninteger, then the adjacent m/z values are not related (i.e. they arise from different analytes or the same analyte but have different charge-bearing adducts). Molecular weight is then obtained from the following equation: $(m_1)(m_2)/(m_2 - m_1) - (m_2)/(m_2 - m_1) = (433)(3) - (3) = 1296$. Note that for the negative ion mode of peptides, protons are removed from the peptide.

2.3 Electrospray Ionization/Quadrupole Instrumentation

Quadrupoles consist of four parallel rods with opposite pairs of rods connected electrically carrying DC and AC voltages of opposite polarity. Voltage is varied across the rods such that at a given moment in time only ions of a certain m/z ratio oscillate in a synchronous and linear path through the filter. All other ions follow a nonlinear trajectory out of the quadrupole and are lost. To produce a mass spectrum, the quadrupole is ''scanned'' across a defined m/z range. A typical scan with a quadrupole instrument takes several seconds, because the AC and DC voltages are varied for a scan and then reset to start again. In contrast TOF and IT mass analyzers acquire an entire mass spectrum in milliseconds. A TOF mass analyzer has a theoretically unlimited m/z range and so is popular for analyzing proteins, whereas quadrupoles and ITs are typically limited in m/z range, thus relying on the multiple charging observed with ESI sources to analyze high-molecular-weight compounds.

A TQ mass spectrometer consists of three sets of quadrupoles that can either transmit all ions (i.e. radiofrequency (RF)-only mode of operation) or function as a mass filter allowing ions of a specific m/z to pass.[16] Generally the first (Q1) and third (Q3) quadrupoles function as mass filters whereas the second quadrupole (Q2) always operates in RF-only mode. When filled with argon (~3–5 mTorr) and with an accelerating voltage applied, Q2 serves as a collision cell, passing all fragment ions from a precursor ion selected by Q1. There are five basic modes of operation for a TQ:

1. In **MS mode**, Q1 or Q3 is scanned to produce a ''main beam'' or standard mass spectrum whereas the non-scanned quadrupole functions in RF-only mode. This scanning mode would provide data that are exactly analogous to the data that one would observe using an SQ instrument.

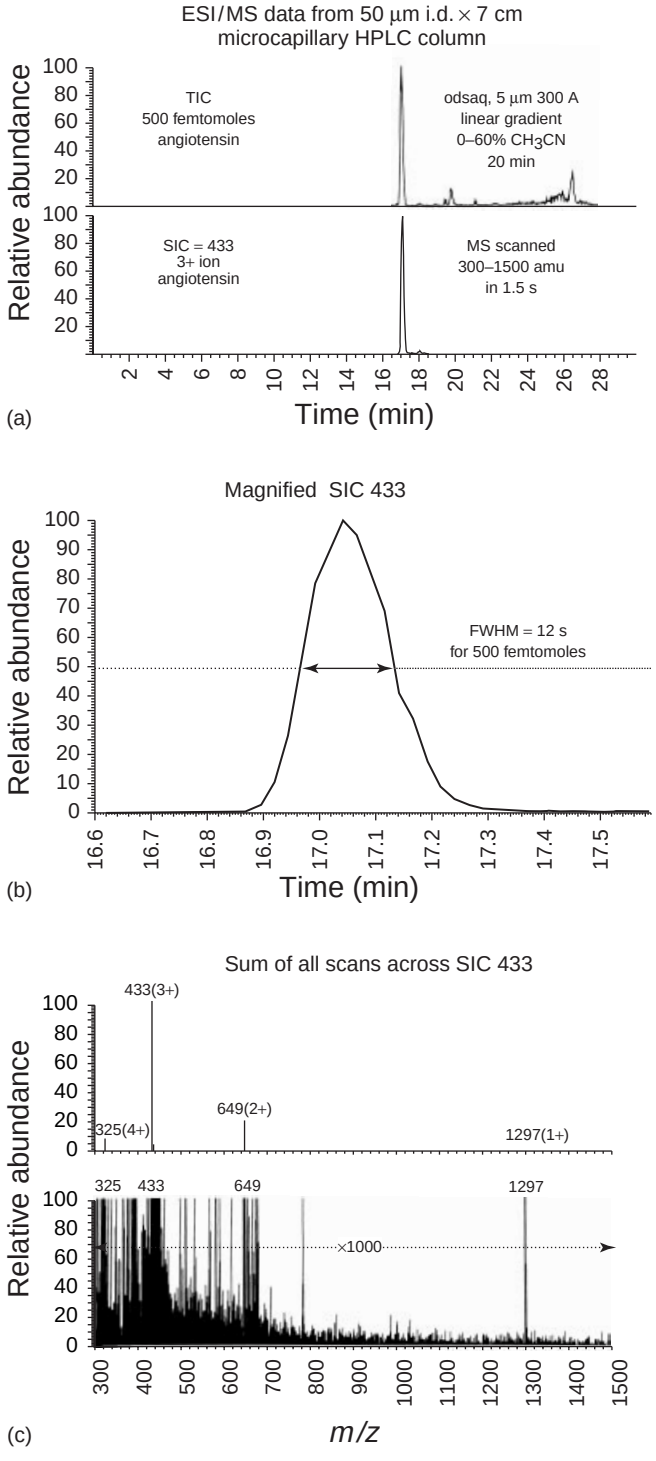

Figure 1 Results from Liquid chromatography/mass spectrometry (LC/MS) system suitability standard test. (a) Total ion chromatogram (TIC) for 500 fmol of angiotensin and single ion chromatogram (SIC) for the 3+ ion. (b) Abscissa magnified around the SIC of the 3+ ion to show peak shape. (c) Sum of mass spectra across the full width at half-maximum (FWHM) shown in (b) normalized to the 3+ ion and the same mass spectrum with the ordinate magnified 1000×. Data acquired on an SQ scanning 300–1500 amu in 1.5 s.

2. In **MS/MS mode**, Q1 transmits ions of a selected m/z. These ions are accelerated into Q2, which is filled with argon gas to generate fragment ions. These fragment ions are then transmitted by Q3, which is scanned to produce a mass spectrum consisting of fragment ions originating from the Q1-selected m/z ion.
3. In **neutral loss mode**, Q1 and Q3 are scanned simultaneously but with Q3 at an m/z offset, corresponding to a neutral loss of interest (e.g. doubly-charged phosphorylated peptides can be detected by scanning for a neutral loss of 49 m/z (loss of H_3PO_4) that undergoes a facile β-elimination). The resultant mass spectrum consists only of ions that lose the specified neutral m/z fragment.[17]
4. To generate a **precursor ion scan**, Q1 is scanned while Q3 transmits ions of only one select m/z value. Thus, all ions transmitted through Q1 are sequentially fragmented in Q2 and any that produce a fragment ion equal to the m/z value set for Q3 produce the so-called precursor ion mass spectrum.[18]
5. **In-source fragmentation** provides an opportunity for triple MS on a TQ instrument or MS/MS on an SQ instrument, or in negative ion mode as a diagnostic ion scan for loss of phosphate from a peptide.[19] Fragmentation occurs in a high-pressure region of the differentially pumped ESI source, as a result of collisions with atmospheric gases. It is an indiscriminate process where charge states with $n > 1$ are preferentially fragmented by imparting a greater amount of kinetic energy than is necessary to simply focus the ions into Q1. All of the above scan types, 1–4, can be carried out on ions generated by in-source collision-induced dissociation (CID).

The two most commonly used MS modes for quantitation are variations on the MS mode and MS/MS mode described above: specifically, selected ion monitoring (SIM), which is sometimes referred to as multiple ion monitoring; and selected reaction monitoring (SRM), which is sometimes referred to as multiple reaction monitoring. These methods of quadrupole operation are important for quantitative analysis. An example of SIM quantitation is presented in section 5. Generally in quantitative studies, a stable isotope label of the drug substance or an analog similar in chemical properties to the drug substance is added as an internal standard (see section 5, where this is discussed in depth). The internal standard is added at a constant amount to the biological sample prior to any sample preparation. Once the sample has been prepared, it can be injected onto the mass spectrometer. In SIM mode, the mass spectrometer scans a small m/z window (typically <0.4 amu) at the m/z of both analyte and internal

standard. This is repeated continuously over the entire chromatographic run. The areas under the curve for each standard concentration are then plotted, so that a linear regression can be used to calculate the quantity of analyte in each unknown sample. It is important to prepare the standard curve in the same matrix that will be used in the study and that the analysis of standards and samples be reproducible. The second type of method commonly used for quantitation is SRM. This technique has significant advantages over SIM in both selectivity and specificity, resulting in a potential speed advantage. In this mode, the TQ remains in MS/MS mode and continuously switches between two different MS/MS reactions: one for the analyte and one for the internal standard. Specific fragment ions, instead of precursor ions, as with SIM, are used for quantitation. Many laboratories are utilizing SRM quantitation with TQs, to achieve the lowest possible detection limits for potent drug substances. The selectivity and specificity of SRM reduces the amount of chromatography required, thus time-consuming (i.e. costly) chromatography can be avoided. Chromatography is still utilized, but it is done rapidly, where an entire run may only last a few minutes. Generally the analyte is desalted, retained on the column for a short period of time and eluted. With an isocratic method, limited column equilibration is required. With a rapid gradient method, as short as possible column re-equilibration is performed. Samples in very complex matrices[20] can be analyzed with little clean-up prior to analysis owing to the specificity of the MS/MS reaction. The sample clean-up is still required with SRM quantitative studies, owing to the ion suppression effects of salts, proteins, and other endogenous compounds. It is also important to demonstrate, through the analysis of blanks, that no endogenous compounds interfere with the analyte-to-fragment ion transition selected for quantitation because it is the unique combination of analyte mass and fragment ion mass that identifies one analyte from another.

2.4 Electrospray Ionization/Ion Trap Instrumentation

A major difference between a TQ and an IT mass spectrometer is that in an IT instrument the mass analyzer also serves as the collision cell.[21] Quadrupole ITs use RF-only quadrupoles to guide ions from the ESI source into the IT, similar to quadrupole instruments. The mass analyzer is significantly different, however. Two different types of electrodes are used to create the IT: a ring electrode and two end-cap electrodes. Ions enter and leave the trap through holes in the end-cap electrodes and are trapped in a field created by applying an RF voltage to the ring electrode. The magnitude of the voltage determines the frequency and motion of trapped ions. Ions above a certain m/z remain in the trap and are collisionally cooled to stabilize their trajectories with helium gas. This results in a collapse of the sinusoidal cycling of ions from a larger to a smaller trajectory over a short period of time. A mass spectrum is acquired by ramping the RF voltage linearly while applying a small voltage across the end-cap electrodes, which causes ''resonance ejection'' of ions of successive m/z. Resonance ejection refers to ions becoming unstable in the trap and thus being ejected axially through the end-cap electrodes, where they are detected by an off-axis conversion dynode with an electron multiplier detector.

Through this ion trapping/ejecting process, tandem MS experiments (MS/MS) and MS^n experiments are conducted. The process of fragmenting ions in an IT begins by ejecting all ions at larger and smaller m/z than the ion of interest. Once the IT contains only the m/z of interest, the selected m/z value is energized by altering voltages on the electrodes, so that it collides with the dampening gas (helium). Fragment ions are then detected by scanning out all of the ions from the trap, as described above. If sufficient signal persists, further rounds of MS/MS (MS^n) can be conducted by leaving a fragment ion of interest in the trap for additional fragmentation events. The amount of energy imparted to the trapped ion can be varied with a parameter referred to on some instruments as the relative collision energy. Other parameters that can be varied on an IT include the number of ions allowed to enter the trap in any one event and the number of such events performed in replicate for averaging purposes (i.e. increasing signal-to-noise). Commercial IT mass spectrometers have data-dependent fragmentation capabilities and are therefore ideally suited to liquid chromatography/tandem mass spectrometry (LC/MS/MS).

3 SAMPLE HANDLING CONSIDERATIONS

3.1 Sample Handling for High Sensitivity

This field is better known as microsample handling and is of significant importance to those working with any sample that can be described by the phrase ''the world's supply''. This usually means that the biologist is nervous about handing over six months' worth of work to someone with no direct appreciation for the precious sample. Or it may mean that the samples are worth thousands of dollars and the study (e.g. clinical trial) cannot be repeated easily. There are some simple rules worth mentioning that can be followed for working with peptides and other small molecules. Generally, almost any transfer or purification step applied to a sample will reduce the quantity by 50%. This is a generalization of course, but the point is to plan carefully the route from

tube to MS so as to avoid potential sample loss. Complete lyophilization to dryness should be avoided because some peptides (and other analytes as well), once dried in a polypropylene tube, will not go back into solution. It is of course impossible to predict a priori which will behave this way. Often new columns will irreversibly adsorb a portion of the sample, thus new HPLC columns should be preconditioned prior to injecting a valuable sample. Finally, the mass spectrometer should be tuned using an analyte that is as near identical to the sample and under the same conditions that the sample will be introduced as possible. Failure to tune the MS under appropriate conditions can easily negate any advantage provided by careful sample handling.

3.2 Chromatographic Considerations

In the field of Proteomics,[22] working with the smallest i.d. capillary HPLC columns possible is a requirement in order to avoid sample loss and to optimize sensitivity.[23,24] Samples, often present at the limit of detection, are extracted from polyacrylamide gels and any manipulation of sample must be minimized. LC/MS sensitivity is optimized by using 50-μm i.d. capillary liquid chromatography (LC) columns packed using stainless-steel chambers (Mass Evolution, Houston, TX) pressurized to 1000 psi. Samples are loaded directly using the same pressure chamber into the capillary column without any chance for loss present in a normal injector where transfer to a syringe is required. The fused silica capillaries (50 μm i.d.) are slurry packed to 10 cm with C18 silica beads in 70% ethanol.[25–27] Porous frits are placed at the capillary terminus to hold the packing in place by sintering underivatized silica beads (LichroSorb Si60) in a flame. Peptides are eluted quickly with a fast linear gradient of 0–60% solvent B in 15 min at 150 nL min^{-1}. Standard HPLC pumps operated at 50 μL min^{-1} can be used for the microcapillary column described above using a precolumn restrictive flow splitter that is 50 μm i.d. and 200 cm in length. This fast gradient method, together with a trace amount of ion pairing reagent, maximizes the chromatographic resolution and produces a sharp elution profile. Typical solvents are aqueous with 0.2% acetic acid and 0.005% heptafluorobutyric acid in one pump and acetonitrile alone in the second pump. Acid is not needed in solvent B for good chromatography and its addition only increases the chemical noise observed in the mass spectrometer. For such small columns, microspray sources are easily and economically constructed from commercially available baseplates and parts (Mass Evolution, Houston, TX), such as a Valco stainless-steel union to which the ESI voltage is applied. Capillary columns terminate inside the Valco union and a fused silica capillary (10 cm × 50 μm × 180 μm) manually tapered to ~5 μm serves as an emitter. No sheath gas or liquid is necessary for such small flow rates. This same device can also be used in negative ion ESI/MS. These techniques have been used to great advantage for sequencing and identifying peptides present in extracts at the very limit of detection by ESI/MS in the fields of apoptosis[28] and immunology.[28–30] Additional examples of the use of microcapillary columns to solve structural problems in biology can be found at the World Wide Web site: http://staff.washington.edu/goodlett.

3.3 Electrospray Ionization Advantages

An additional advantage can be made of the concentration-dependent nature of ESI, which allows the flow of liquid through an ESI source to be decreased or modulated relative to some normally constant flow rate during a separation without sacrificing sensitivity.[31] The advantage provided by a reduction in flow through the ESI source can be thought of as follows. If at a given moment in time (i.e. one scan) 100 analyte molecules pass through the ESI source and there is current available to convert only ten analytes to ions in the gas phase, then 90 will be wasted. By reducing the flow through the ESI source, the available ESI current can be utilized more efficiently. Consider the following simplified example where flow is reduced and the chromatographic ''peak'' broadened such that 25 analytes instead of 100 pass through the source during one scan. Now, with the only experimental change being reduced flow through the ESI source, such that there is still enough current available to convert ten analytes per scan to ions, only 15 analyte molecules are wasted per scan. The next three MS scans across this broadened ''peak'' will detect, in turn, 30 analytes as ions. So, instead of 10% of the sample being analyzed, 40% is analyzed via this reduced elution speed process. This may seem insignificant, but it has proven to be important for both electrophoretic and capillary liquid chromatography/electrospray ionization/mass spectrometry (LC/ESI/MS) applications.[5,32] Several names have been used to describe this process, including ''peak parking''[33] and ''voltage drop'';[34] regardless of the name, it is the concentration-dependent nature of ESI that allows more time for analysis across a given chromatographic area without sacrificing sample.

4 GENERAL LABORATORY

4.1 Following Good Laboratory Practice

Good laboratory practice (GLP) regulations are a management-based quality system enforced by the Food and Drug Administration (FDA) for nonclinical studies.

It is intended for nonclinical laboratory studies that support applications for research or marketing permits for products regulated by the FDA. Laboratory analysis done for clinical trial support is also expected to be done under GLP. In general, following GLP is mainly about using a commonsense approach to laboratory management. Scientists are asked to record all work in a logical and reproducible manner such that it can be easily repeated or reconstructed. It involves signing and dating all notebook entries so that a trail is left for decision-making in future experimental design and for auditors. Standard operating procedures should be developed and adhered to for instrumentation used in the various studies. Instrument maintenance logs are kept for each piece of equipment so that trouble-shooting can be carried out logically and orderly. Instruments should be tested routinely to ensure that they are in good working order. When a system that is routinely tested using the same conditions fails, then logical systematic procedures should be in place to correct the problem. Each set of experiments should include, at the very least, a set of system suitability checks (i.e. standards and blanks run under conditions identical to the unknowns) placed at the beginning and end of each data set to ensure that the instrument or system is, and has, functioned properly over the course of the analysis.

For LC/MS systems there are four tests that the system suitability standard can assess:

1. Is the HPLC system functioning properly?
2. Are the solvents clean?
3. Is the mass spectrometer tuned and calibrated properly?
4. Is the HPLC column and associated plumbing in good condition?

All of these questions can be answered by one LC/MS analysis performed prior to analyzing real samples. To answer these questions a system suitability standard is designated for the laboratory LC/MS work. It should be an analyte in the same class as that being analyzed or a well-characterized sample. For example, a system suitability test for microcapillary high-performance liquid chromatography/mass spectrometry (HPLC/MS) using a 50-μm i.d. × 10-cm C18 column could be the analysis of 500 fmol of angiotensin. The peptide is injected onto the capillary column and then eluted under standard linear gradient conditions (Figure 1a). The results obtained are analyzed in light of the above questions. If the HPLC system has functioned properly to deliver solvents as programmed, then the elution time for the peptide will be reproducible. If elution time is outside of the expected range, then either the HPLC system is not functioning as programmed or perhaps the column is beginning to fail. The peak shape should be symmetrical and elute over a short time period of 10–15 s (Figure 1b). If the peak is too broad or not symmetrical, then there may be a problem with the chromatography or perhaps additional dead volume has been introduced somewhere in the HPLC system. If the HPLC/MS system is working properly, then the signal-to-noise ratio of the analyte should be above a preset value. If the signal-to-noise ratio of the analyte is below an acceptable value, then perhaps the solvents should be changed or the mass spectrometer cleaned. For a 500-fmol angiotensin injection on the above-mentioned capillary column dimensions, a signal-to-noise ratio of at least 1000 : 1 should be attained at the apex of the chromatographic peak (Figure 1c). Finally, the calculated and observed molecular weight for the analyte (angiotensin) should match closely. Of course many of the potential symptoms and problems in an HPLC/MS system are related and some skill is needed to ascertain the exact cause of decreased system performance. However, a simple, well-designed, well-characterized, robust system suitability test carried out prior to the analysis of samples would provide confidence in the observed results and should enable the novice operator to begin trouble-shooting a complex scientific instrument if something were to go wrong.

4.2 Automation

Almost every step in the sample analysis process can be and has been automated. Some automated processes are more efficient than others, but all automation is driven by the high costs of drug discovery and production. This cost includes employee costs and the years of research prior to discovery that are followed by methodical FDA testing requirements once a new drug candidate has been identified. The desire for higher throughput in pharmaceutical companies can be observed easily from the combinatorial chemistry initiatives being employed at almost all companies. The need for larger compound libraries, in the hopes of finding better lead candidates, has relied heavily on the automation of numerous synthetic procedures. MS has been utilized heavily in the combinatorial chemistry field, performing rapid molecular weight measurements, high-throughput purification, and decoding libraries from hundreds of combinatorial chemistry samples per day. Almost all of the MS instruments utilized in the combinatorial field are part of an automated system, where multiple probe injectors, high-capacity autosamplers, fraction collectors, analog detectors, and liquid delivery systems are used together. With the increased size in focused libraries, additional needs have been placed on the discovery research groups responsible for lead identification, development, and selection. These groups have begun to utilize robotic liquid handlers, 96-well solid-phase extraction (SPE), fast generic gradient methods, and as many mass spectrometers as necessary to analyze

samples as rapidly as possible from a variety of in vitro and in vivo screens. From sample work-up to data analysis, devices have been built that free chemists for nonroutine tasks that require problem-solving skills. This drive to increase throughput and efficiency can be beneficial for scientists operating in this environment. The increased use of automation may enable employees to focus on the creative processes of managing projects, understanding processes, and interacting with other scientists to understand better the important scientific problems facing society today.

In the field of protein discovery science, data-dependent analysis is commonly used in a first-pass analysis of samples.[35] The entire HPLC/MS analysis is automated once the scientist loads a sample into the autosampler tray. The number of injections per vial and the injection volume can be controlled by various instrument programs. The time-consuming data analysis process can also be automated to some degree; indeed the scientist can load a sample in the evening and data printouts are waiting in the printer bin when the scientist returns the next day. With modems and networking, progress can be monitored and analyses interrupted remotely if necessary to avoid waste of valuable samples.

As an example, a routine for sequencing peptides would be designed to select an ion for MS/MS, based on ion current intensity. As carried out on a TQ instrument, the mass spectrometer would be set-up (i.e. tuned) to operate with Q2 at a constant argon pressure of 3 mTorr. The offset on Q3 is increased to compensate for the dampening effect of this gas. No fragmentation is observed from such low-energy collision with the gas because the ions during a normal MS scan are not accelerated into the collision gas. However, once the instrument control language (ICL) procedure selects an ion for MS/MS, then the Q2 offset is increased as a function of the m/z value selected for MS/MS. More energy for fragmentation is required for higher m/z values than smaller, but less for higher charge states than lower ones, because the energy barrier to fragmentation is lowered by Coulombic repulsion when more than one charge is present. For this automated MS/MS process the instrument is not optimized for a specific ion of interest and so Q2 offset energy is increased slightly during three consecutive MS/MS scans such that the final summed MS/MS spectrum is a composite of MS/MS scans carried out all on the same ion but at various energies. This rotation of energy compensates for the lack of operator input during the process and produces readily interpretable spectra. Once the three MS/MS scans for the selected ion are acquired, the ICL procedure returns the mass spectrometer to full scan or simple MS mode and the process of ion selection begins again. The ion just selected for MS/MS goes into a list of ions to prevent it from being immediately selected again should it prove to be still the most intense ion in the first scan after switching back to MS mode. In this manner, a 100 MS/MS spectra can be acquired in a single LC run. On an IT where scans are of the order of milliseconds rather than seconds, many more MS/MS spectra can be acquired per LC run. This advantage can be of value when seeking to get as complete a coverage as possible on the peptides present in a tryptic digest. Additional information and ICL routines can be found at the World Wide Web site: http://depts.washington.edu/~ruedilab/aebersold.html.

4.3 Managing Instrument Time

Mass spectrometers, being so-called universal-type detectors, rarely sit idle once installed. If the instrument is easy to use and delivers the expected results then chemists and biologists will keep the instrument busy. Mass spectrometers tend to be set-up in three broad categories of user: the single-operator instrument, which is usually a more complicated design or usage; the open-access instrument, for multiple users who are often the least experienced; and the multiple-operator instrument, which is used by many experienced users. In the last 10 years mass spectrometers have become more user-friendly to the point where instruments are more often sold today to chemists, chromatographers or biologists with no prior experience in setting them up. This ease of use has made the open-access laboratory and the multiple-user set-up possible where before specialists were needed just to maintain the instrument. Many open-access instruments are usually maintained by an analytical department, or other departments that are heavy users of MS instrumentation simply hire at least one person trained in MS. Manufacturers have made interfacing to a mass spectrometer easy for the novice by using computer software such as Windows NT, with which most people are familiar. For many MS instruments, operation may be as easy as any other laboratory instrument, such as a gas chromatography, HPLC or UV–visible spectrophotometer.

5 QUANTITATION OF PEPTIDES BY ELECTROSPRAY IONIZATION/MASS SPECTROMETRY

5.1 Analysis of Peptides in Physiological Fluids

Analysis of trace levels of drugs in urine by automated procedures that avoid SPE is of particular interest in the pharmaceutical sciences. Benefits to direct sample analysis over SPE include a decrease in sample preparation

time and a reduction in nonspecific losses that can occur during SPE. Although SPE produces adequate sensitivity for most studies, it is time-consuming and costly. Methods that allow peptides to be analyzed directly from plasma or urine are also of interest because of the low half-life of most peptides in physiological fluids. For a small peptide such as thymopentin (Arg-Lys-Asp-Val-Tyr), it can be as low as 30 s. The limited bioavailability of orally administered peptides is due in large part to enzymatic and acid hydrolysis in the gut and stomach.[7,8] Resistance to hydrolysis may be improved by modification of the amino- and carboxyl-terminii. Even so, methods that provide rapid processing of peptides in physiological fluids for analysis are needed for pharmacokinetic studies where the concentration of peptides is low at the start of extraction.

''Column-switching'' protocols have been employed where analytes in crude mixtures need to be analyzed directly by MS. Such a protocol avoids SPE preparation by relying on interaction of the analyte with the stationary phase while most endogenous components are washed to waste. Selective chromatographic conditions retain the analyte until sufficiently clean and then pneumatically switched valves redirect flow from waste to MS for detection. Such techniques have been demonstrated for the analysis LSD analogs and metabolites in urine.[36,37] They allow clean-up with no precolumn sample preparation and direct quantitation by MS methods.

When coupled with HPLC separations, the ionization technique of choice for charged analytes such as peptides is ESI. However, the presence of alkali salts and other nonvolatile compounds in a physiological sample degrades sensitivity by competing efficiently with the analyte for available ESI current. Thus, it is important to remove as many of the nonanalyte components as possible prior to ESI/MS analysis. For MS detection with ESI, clean-up is necessary because the salts and urea present in urine decrease MS performance by physically occluding the atmospheric inlet and coating the lenses used to complete desolvation and then focus ions. An additional problem with salt in a sample is that alkali metals such as sodium are efficiently bound to carboxylic acid moieties, decreasing the relative amount of $[\text{peptide} + \text{H}]^+$ available for detection by single or multiple ion monitoring of a narrow m/z range around the $[\text{peptide} + \text{H}]^+$ value. With such orthogonal techniques as ion exchange and HPLC coupled on-line to MS it is important that the ion exchange cycle be followed by thorough rinsing to avoid such problems as ion adduction to analytes. The following study demonstrates the use of a ''column-switching'' chromatography system for automated, on-line detection of peptides in urine by ESI/MS.

5.2 Tandem Liquid Chromatography/Mass Spectrometry Quantitation of Peptides in Urine

Prior to use, refrigerated urine is treated with trifluoroacetic acid (TFA) to give a final concentration of 1% and a pH of approximately 3.0. Excess endogenous materials precipitated by TFA are removed by centrifugation at 3000 rpm for 30 min. The supernatant from acidified urine is filtered through 0.2-μm Anodisc. For ion-exchange HPLC, a 1 M potassium phosphate stock solution filtered through a 0.2-μm Anodisc filter is used to prepare mobile phases A (75 mM) and B (700 mM) by dilution. Before final dilution with HPLC-grade water, acetonitrile is added to A and B so that the final concentration is 7.5% and the pH is adjusted to 3.5 with 85% phosphoric acid. For reversed-phase HPLC, mobile phases C and D contain 0.05% TFA with 2% and 60% methanol, respectively. Solutions of IRI-695 (*N*-acetyl-L-arginyl-L-prolyl-L-aspartyl-*NH*-isobutyl) are prepared at 0, 2, 10, 100, 500, and 1000 ng mL^{-1} in water or urine and each contains 400 ng mL^{-1} of IRI-959 (*N*-acetyl-L-arginyl-L-prolyl-L-aspartyl-*NH*-isopentyl).

A diagram of the column-switching apparatus is shown in Figure 2. Pump one (P1) is a Waters Model 616 (Division of Millipore Corp., Milford, MA, USA) capable of delivering four different mobile phases; pump two (P2) consists of two Shimadzu Model LC-10AD units linked to perform a linear gradient; and pump three (P3) is a single Shimadzu Model LC-10AD. Additional instruments include a Model 717 autosampler, a Model 600S controller, and two Model 486 tunable absorbance UV detectors. Both UV detectors are connected to Hewlett-Packard 3395 recording integrators (Palo Alto, CA, USA). Two Rheodyne (Cotati, CA, USA) high-pressure automated six-port switching valves are used with pneumatic actuators engaged at 80 psi. The HPLC columns are as follows: for cation exchange, a Spherisorb Phase Sep SCX, 250 mm × 4.6 mm i.d. column with 5-μm particles (Norwalk, CT); for trapping prior to reversed-phase HPLC, a YMC S3 ODS, C18, 50 mm × 4 mm i.d. column with 3-μm particles (Wilmington, NC, USA); and for analytical reversed-phase separation, a YMC Basic S5, 250 mm × 2 mm i.d. column with 5-μm particles. All columns are kept at 35 °C inside a Shimadzu CTO-10A column heater.

The mass spectrometer is operated at unit resolution in multiple ion monitoring mode, scanning continuously during tandem liquid chromatography/mass spectrometry (LC/LC/MS). The range scanned for analyte (IRI-695; $[\text{M} + \text{H}]^+ = 485$) and internal standard (IRI-959; $[\text{M} + \text{H}]^+ = 499$) is set at $[\text{M} + \text{H}]^+ \pm 1.0$ amu. One complete scan of the $[\text{M} + \text{H}]^+$ value for both peptides took 2 s. Chromatographic separations are monitored by UV absorption at 210 nm after the ion-exchange column

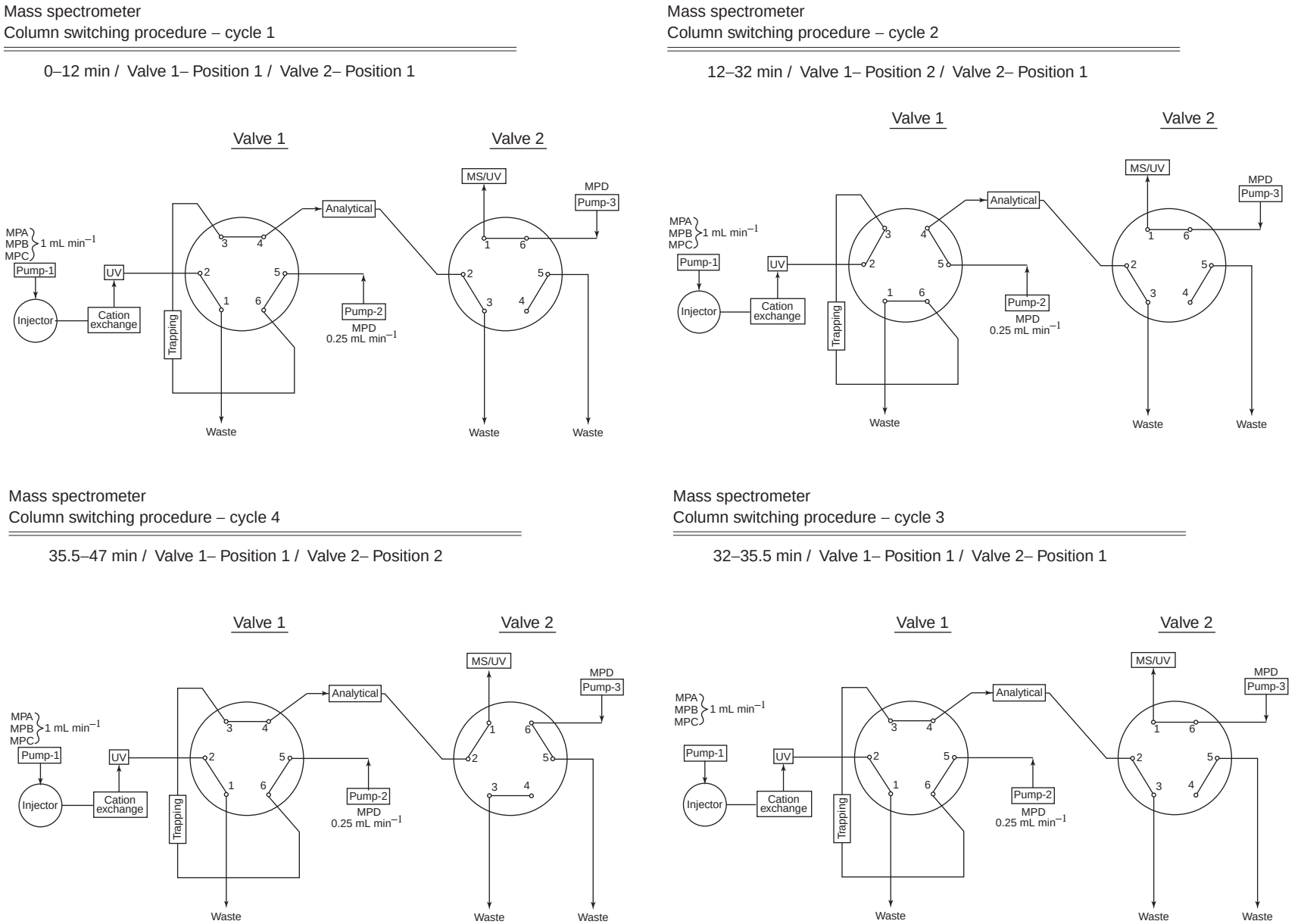

Figure 2 Solvent flow diagram for the four cycles of orthogonal ion-exchange chromatography/reversed-phase HPLC (LC/LC/MS) used to clean up peptides in raw urine prior to on-line analysis by ESI/MS. MPA–MPD refer to mobile phases A–D.

and after the analytical column. Figure 2 shows the four stages of one LC/LC/MS cycle. Pump P1 maintains flow through the ion-exchange column at 1 mL min^{-1} through all four stages and delivers mobile phases A, B and C. Pumps P2 and P3 deliver mobile phases C and D at 0.25 mL min^{-1}. Pump P3 provides a constant flow of mobile phase D to the mass spectrometer with a post-column splitter sending one-third of that to the ESI source and two-thirds to the second UV detector.

5.3 Discussion

This LC/LC/MS method consists of four cycles (Figure 2). The only sample preparation prior to injection onto the cation-exchange column involved acidification of raw urine followed by centrifugation. This ensures that peptides are at an acidic pH sufficient for cation exchange to work efficiently and removes endogenous components that flocculate during refrigeration. During cycle 1 (Figure 2), peptides in urine are loaded on a cation-exchange column and washed just as during SPE preparation. After washing the column to remove as much endogenous material as possible, the peptides are transferred directly by a gradient of increasing potassium phosphate concentration to a reversed-phase trapping column. The trapping column and analytical column avoid contamination by most of the endogenous urinary waste because the valve between the cation-exchange and trapping column does not flow directly toward the trapping column until immediately prior to gradient elution of the peptides.

Valve switching times were determined by monitoring UV absorbance of peptides after injecting a 0.1 mg mL^{-1} standard solution. The first UV detector, located after the cation-exchange column (Figure 2), allows valve switching times to be optimized so that most UV-absorbing material goes to waste rather than to the trapping column. Based on the area under the curve, 90% of the UV-absorbing material goes to waste prior to elution of the peptides. The isopentyl group of IRI-959 ensures chromatographic resolution from IRI-695, which contains an isobutyl group instead of the isopentyl group. Peptides IRI-695 and IRI-959 were resolved chromatographically on both the cation-exchange and reversed-phase columns.

An initial concern with ESI/MS detection was the high potassium phosphate concentration needed to carry out ion exchange. Any peptide adducts would reduce the limit of detection of the method by reducing the $[M + H]^+$ species available for detection. In solution, peptides readily form adducts with potassium and/or phosphate ions that are easily observed by ESI/MS. However, neither potassium nor phosphate adducts to the peptides were observed by ESI/MS. One other concern prior to the study was the high content of urea in urine. It was expected

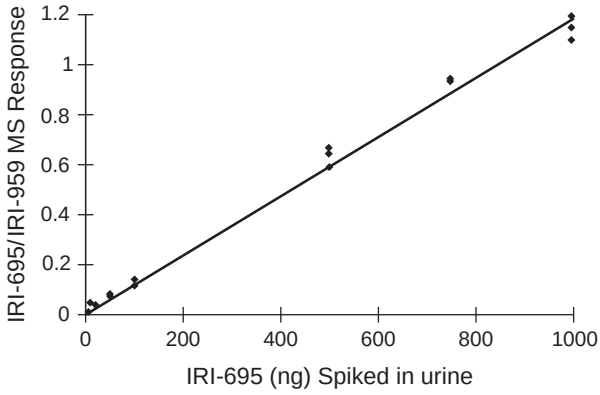

Figure 3 Linear regression used to quantitate peptides in urine. To generate the curve an internal standard (IRI-959) was spiked at a constant amount and the concentration of the analyte (IRI-695) was varied from 2 to 1000 ng.

that the chromatography would be degraded and the atmospheric pressure inlet clogged because the average concentration of urea in urine is 330 mM. To test whether or not urea would have an effect on peptide recovery and cause the mentioned physical problems, IRI-695 and IRI-959 were spiked into a 330 mM urea solution and analyzed by LC/LC/MS. This high concentration of urea alone did not change the recovery of peptides, nor did it cause any other obvious physical problems with equipment, which suggests that most of the urea was removed during the cation-exchange cycle (data not shown).

Ruggedness of the columns was determined by multiple urine injections. Approximately 100 urine injections were done without any change of HPLC columns. Inclusion of the internal standard in each sample provided a measure of column performance. Any change in chromatographic resolution could be noted and changes made. However, during the course of developing this method the reproducibility of retention times proved the method to be rugged. The limit of detection for IRI-695 in water was 2 ng mL^{-1} and in urine 10 ng mL^{-1}. Performance of the ESI-heated capillary used in desolvation of analyte did not degrade over the course of 100 analyses. Detection of IRI-695 in urine by LC/LC/MS was linear from 10 to 1000 ng mL^{-1} (Figure 3) and sensitive, with a limit of detection of 10 ng mL^{-1}. The LC/ESI/MS process is completely automated and chromatographic fidelity is maintained after over 100 separations without a change of column.

6 CHIRAL ANALYSIS

6.1 Chiral Amino Acid Analysis

Characterization of peptides for an investigational new drug or new drug application submission to the FDA

requires that all impurities above 0.1% level be identified. It is expected that most impurities will be related to the product chemically. For peptide drugs this means that the related substances may be oxidized forms, truncated forms resulting from degradation of the parent peptide, the parent peptide with protecting groups intact from the synthesis or simply a chiral impurity. Using reversed-phase HPLC, most related substances can be separated from the parent peptide and purified for characterization. In an automated setting, impurities might be characterized on-line with some sort of data-dependent MS/MS ICL procedure. If this is not practical because of limited sample, then scale-up for purification is necessary.

Related substances that are isobaric with the peptide drug are likely to be stereoisomers where one or more amino acid is in the D-amino acid configuration rather than the bioactive L-amino acid form. Substitution of a D-amino acid for an L-amino acid may result in a biologically inactive drug or, worse, impart some sort of toxic property not present in the all-L-amino acid peptide. Thus, chiral analysis of a peptide drug and any related substances is an important part of the overall characterization process before submission to the FDA. Methods that provide rapid and accurate assessment of chiral purity have been developed and will be reviewed in sections 6 and 7.[38,39]

6.2 Liquid Chromatography/Mass Spectrometry Chiral Purity

An LC/MS method that allows fast determination of the chiral purity of a peptide is described that involves hydrolysis in deuterium-enriched HCl (DCl)–deuterium-enriched acetic acid (DOAc) followed by derivatization with a chiral reagent (Scheme 1) to create diastereomers that can be separated and quantitated by LC/MS. During peptide hydrolysis the ratio of D-/L-amino acids present in the original peptide can change significantly by acid racemization through an enol intermediate.[38] Peptide hydrolysis in deuterated solvents labels amino acids that racemize during hydrolysis with one deuterium at the chiral center. Racemized amino acids, labeled with deuterium, can be distinguished from non-racemized amino acids by MS because of the increase in molecular weight by one mass unit. Labeling with 1-fluoro-2,4-dinitrophenyl-5-L-alanine (FDAA) allows the five L-amino acids found in a control peptide thymopentin and all possible D-amino acid pairs to be separated in one chromatographic experiment on a conventional reversed-phase column.[38] All peptides were hydrolyzed for 18 h at 130 °C in 1:1 HCl–HOAc or DCl–DOAc. Hydrolyzed peptides were lyophilized and then dissolved in 400 μL of water. To this was added 100 μL of 1 N $NaHCO_3$ and then 200 μL of 38.7 mM FDAA in acetone. The solution was vortexed and incubated at 40 °C for 1 h. Reactions were quenched by 50 μL of 2 N HCl. For LC/MS analysis, samples were diluted 1:10 with solvent A described below. Amino acids labeled with FDAA are referred to as 1-fluoro-2,4-dinitrophenyl-5-L-alanine-labeled or DNPA-amino acids in the text.

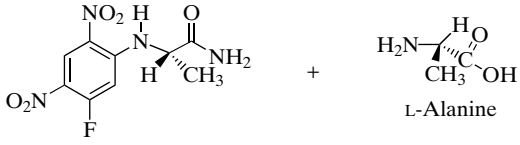

1-Fluoro-2,4-dinitrophenyl-5-L-alanine amide

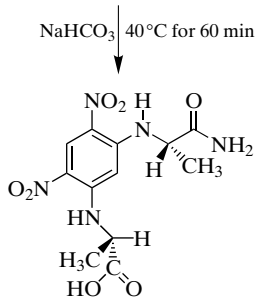

1-L-Alanine-2,4-dinitrophenyl-5-L-alanine amide

Scheme 1 Derivatization of an amino acid with FDAA amide after peptide/protein acid hydrolysis. The derivatization reaction produces a diastereomer, as in this example with alanine, that becomes 1-L-alanine-2,4-dinitrophenyl-5-L-alanine amide or DNPA-alanine.

HPLC separation was by linear gradient using a YMC basic 2.0 × 250-mm column with post-column flow splitting. Two-thirds of the flow was split for UV detection at 340 nm and one-third was split for MS detection. Solvent A was 0.05% TFA, 1.0% methanol and 5.0% acetonitrile and solvent B was 0.05% TFA, 1.0% methanol and 60.0% acetonitrile. Alternatively, a 10 mM ammonium formate (pH 5.2) solvent system was used that was prepared with the same methanol and acetonitrile concentrations as per solvents A and B above but without the TFA. This was done by bringing solvents A and B to volume (1.0 L total each) using 10 mM ammonium formate instead of water. The flow rate at the pump was 0.25 mL min^{-1} and in 20 min the gradient went from initial conditions of 80%A–20%B to 20%A–80%B. For LC/MS, flow from the column was 0.083 mL min^{-1}, such that with the coaxial sheath (methanol) the total flow was 0.183 mL min^{-1} and +3.0 kV was applied for ESI. When a single amino acid was suspected of being in the D-amino acid form then the mass spectrometer was operated at unit resolution and scanned ±5.0 units around the ion of interest. To increase sensitivity at unit resolution, only one ion of interest was scanned per LC/MS run. Alternatively, single ion monitoring (e.g. for $[\text{DNPA-lysine} + H]^+ = 393$)

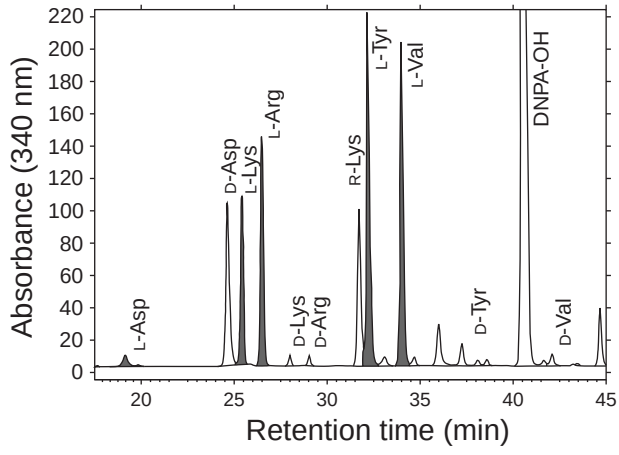

Figure 4 Chromatographic separation of the 10 D- and L-amino acid diastereomers possible from hydrolysis of the peptide Arg-Lys-Asp-Val-Tyr.

was also carried out to observe simultaneously each of the five amino acids in thymopentin. Controls using D- and L-amino acids found in thymopentin were used to establish order of elution. The relative chiral purity of each D-/L-amino acid pair was calculated from the integrated area of extracted ion chromatograms (Figure 4).

6.3 Discussion

The method has been used to establish the chiral purity of a thymopentin-related substance. The unknown was originally present at less than 0.1% level relative to thymopentin. From proton nuclear magnetic resonance data it was suspected that the impurity contained D-Asp or D-Val and possessed a second lysine residue. Three standards were synthesized for comparison: Arg-Lys-Lys-Asp-Val-Tyr; Arg-Lys-Lys-D-Asp-Val-Tyr; and Arg-Lys-Lys-Asp-D-Val-Tyr. After processing through the above protocol, it was concluded that the impurity was composed of only L-amino acids. The impurity then had to be isolated for sequence analysis by MS and Edman chemistry. It was found to have the sequence Arg-Lys-Lys-Asp-Val-Tyr. Note that Figure 4 shows the separation of all D- and L-amino acids involved in this study. A more comprehensive study showed that all L-amino acids and their D-amino acid pairs could be separated in a single LC/MS run.[38]

7 COMPARATIVE TANDEM MASS SPECTROMETRY

7.1 Background on Collision-induced Dissociation of Peptides

Interpretative CID routines like SEQUEST compare observed CID spectra of peptides to artificial spectra of peptides found in sequence databases and generate a list of most likely matches. Such routines work very well, but lack the ability to incorporate chemical information found in the peptide sequence; the artificial CID spectrum created by a computer routine will be a series of m/z values of equal intensity. An objective comparison of the tendency of a series of related peptides to generate specific fragments can be used to yield information useful in screening peptide drugs for specific impurities. This section will provide an example of such a comparative analysis.

Before preceding it is necessary to review briefly the fragment ion nomenclature used by mass spectrometrists to describe fragment ions used for sequence analysis. For sequence analysis, proteins are processed by proteolysis to peptides, which are then analyzed easily by electrospray ionization/tandem mass spectrometry (ESI/MS/MS). Trypsin is often used because proteins tend to have numerous cleavage sites for trypsin (i.e. Lys and Arg), but also because peptides produced by tryptic digestion of a protein have either lysine or arginine at the carboxyl-terminus and thus with a charge fixed at the carboxyl-terminus will produce a nearly complete series of y-ion fragments. Peptide fragments that retain charge on the carboxyl portion are termed y-ions and on the amino-terminus b-ions (Figure 5a).[40] An example of how fragment ions are calculated for the pentapeptide Arg-Lys-Asp-Val-Tyr is shown in Figure 5(b).

Data accumulated over the years allow certain predictions to be made about the CID spectrum of a peptide based on the amino acid sequence (i.e. the

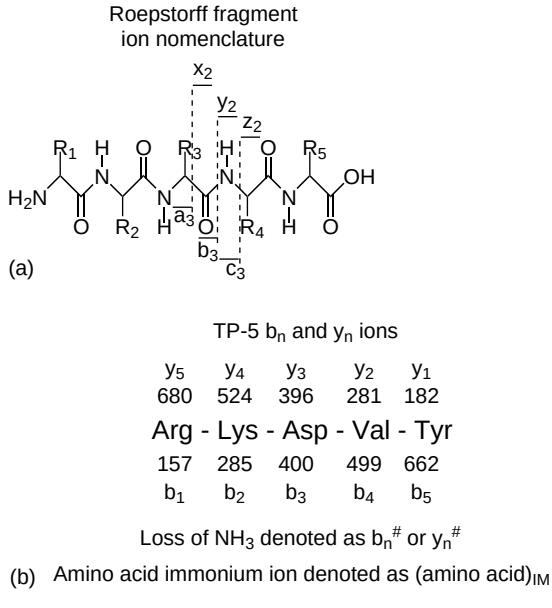

Figure 5 (a) Peptide fragment ion nomenclature as suggested by Roepstorff and Fohlman.[40] (b) Example of calculating b- and y-ion fragments for a peptide (Tp-5 or thymopentin) from nominal residue masses.

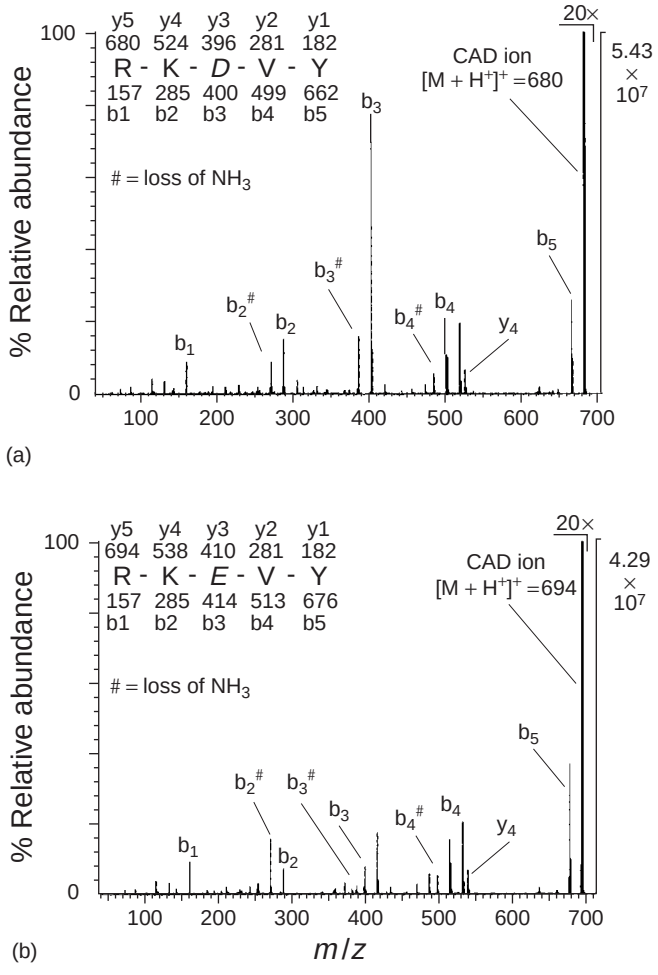

Figure 6 Low-energy MS/MS spectra of the $[M + H]^+$ ions for the peptides RKDVY (Arg-Lys-Asp-Val-Tyr) (a) and RKEVY (Arg-Lys-Glu-Val-Tyr) (b) acquired by ESI on a TQ mass spectrometer. Note the intense b3 ion formation in (a) relative to (b).

peptide chemistry). For instance, a peptide with the amino acid proline in the sequence will, because of the basicity of the proline imide bond being greater than that for any of the amide bonds, fragment preferentially at the amino-terminal side of proline (i.e. at the imide bond).[41] Peptides that contain aspartic acid tend to fragment on the carboxyl side of aspartic acid.[42] It is thought that this happens because of the ability of the aspartic acid side chain to form a stable five-membered ring with the adjacent amide bond and thus influence departure of the carboxyl fragment. Glutamic acid, which would form a less-favored six-membered ring, does not show the same tendency (Figure 6).[42] Although the interpretation of such chemical information remains largely a difficult process limited to skilled analysts, comparative information can be used to analyze related peptides for slight chemical differences such as presented by changes at a chiral center.[43]

An alternative to the LC/MS chiral method presented above was devised using comparative MS/MS of synthetic standards.[39] Just as in section 6 above, it was of interest to know when a D-amino acid isomer of the peptide Arg-Lys-Asp-Val-Tyr (RKDVY) appeared during stability studies or as a result of production procedures. The study used the series of peptides based on the peptide RKDVY, where each L-amino acid was replaced with a D-amino acid such that there were six peptides for comparison. The study compared the CID of each of six peptides using electrospray ionization/triple quadrupole tandem mass spectrometry (ESI/TQ/MS/MS) and electrospray ionization/ion trap tandem mass spectrometry (ESI/IT/MS/MS) and looked at both the 1+ and 2+ ions for each of the peptides.[39] It was shown that peptides in this isomeric series produced fragments that could be correlated with unique structural features of individual peptides.

7.2 Chiral Analysis of Peptides by Tandem Mass Spectrometry

CID spectra of the following peptide 1+ and 2+ ions were compared: Arg-Lys-Asp-Val-Tyr, D-Arg-Lys-Asp-Val-Tyr, Arg-D-Lys-Asp-Val-Tyr, Arg-Lys-D-Asp-Val-Tyr, Arg-Lys-Asp-D-Val-Tyr and Arg-Lys-Asp-Val-D-Tyr. All peptides ($1.5\,\text{pmol}\,\mu\text{L}^{-1}$) were introduced into an ESI source by direct infusion at $3\,\mu\text{L}\,\text{min}^{-1}$ in an acidic solution. Low-energy CID spectra were obtained on a Finnigan tsq 7000 TQ mass spectrometer and a Finnigan lcq IT mass spectrometer (San Jose, CA). For comparison, CID spectra were averaged for 1 min and collision energies were optimized for one of four sets of experiments: triple quadrupole collision-induced dissociation (TQ/CID) of 1+ ion ($680\,m/z$), TQ/CID of 2+ ion ($340\,m/z$), ion trap collision-induced dissociation (IT/CID) of 1+ ion and IT/CID of 2+ ion. Specifically the fragment ions b3 (400 amu) and b4 (499 amu) were compared. For graphical comparison, the ratio of the observed intensity of b3/b4 ions in each isomer was calculated and all values normalized to the same ratio for the all-L-amino acid peptide. This type of data analysis allows an unbiased assessment of the tendency of a given isomer to form the fragment ions b3 and b4. As seen in the following discussion, when bar graphs of such data are used for comparison an objective analysis can be done even by a novice.

7.3 Discussion

Some general trends are apparent when comparing the IT and TQ/CID of peptides. The most obvious is the reduction in complexity in the IT data. Fewer fragment ions are observed with the IT/CID of a 1+ ion (Figure 7a, c) than with the TQ (Figure 7b, d). This is likely due

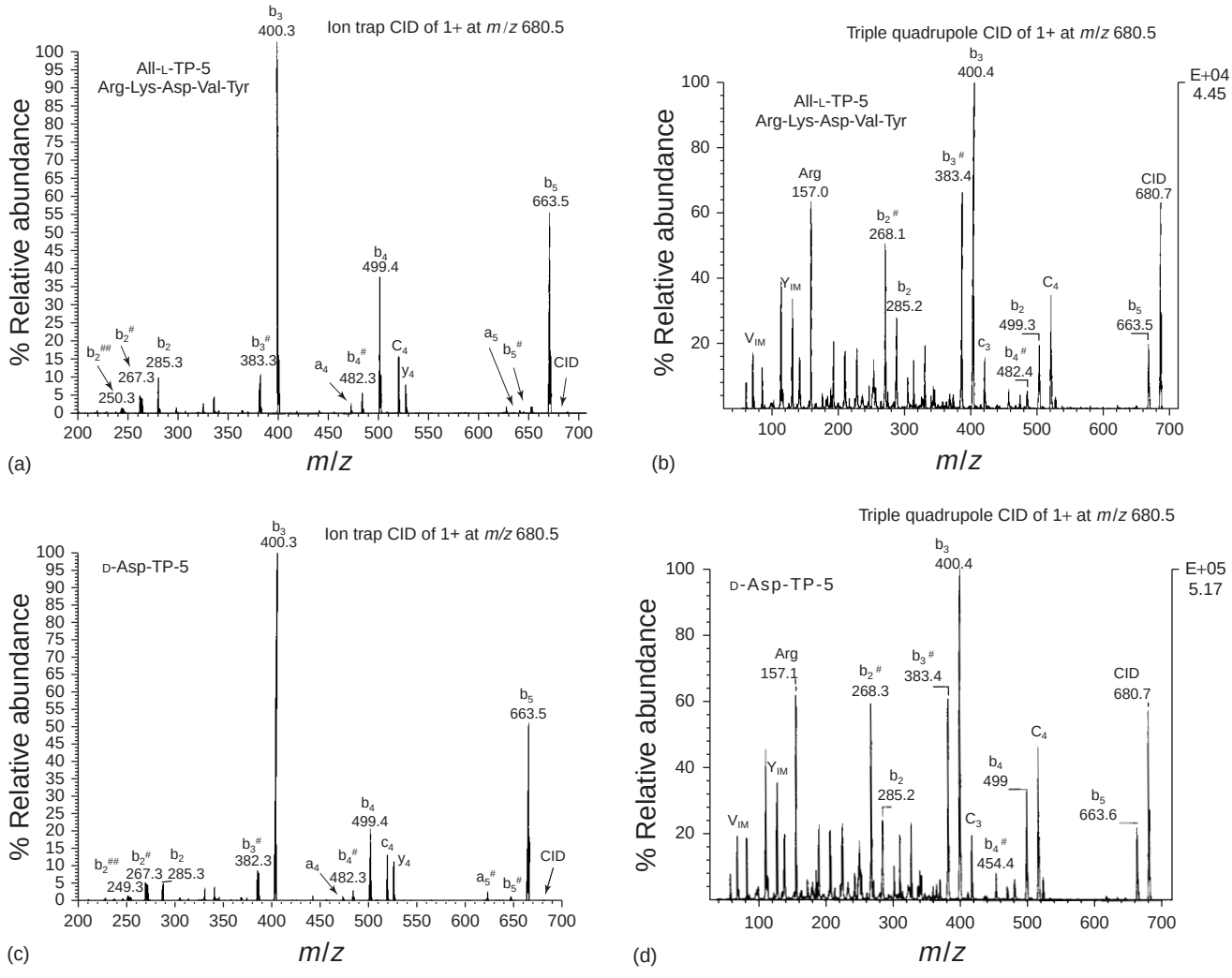

Figure 7 Low-energy MS/MS spectra of the $[M + H]^+$ ions for the peptide Arg-Lys-Asp-Val-Tyr (all-L-thymopentin or all-L-TP-5) in an IT (a) and TQ (b) and for the peptide Arg-Lys-D-Asp-Val-Tyr in an IT (c) and TQ (d).

in part to temporal differences between CID in an IT and a TQ. The IT acquires spectra about 10 times faster than the TQ. The TQ, being a relatively slow (1–2 s per scan) scanning instrument, allows more time for secondary fragmentation of b- and y-ions than the IT. It is often the case that many of the fragment ions present in the MS/MS spectrum of a peptide from a TQ instrument will be difficult to interpret, whereas most of the fragment ions for the same peptide can be assigned in an IT. A disadvantage of the IT data is that there are no immonium ions present.[44] Peptide ions will fragment during CID to constituent amino acids that consist of a basic unit where an immonium group bears the charge and the amino acid side chain is retained, providing a unique mass for each of the 21 possible amino acids except leucine/isoleucine, which are isobaric. Immonium ions are diagnostic of the amino acids present in a peptide.

The CID of the 1+ ions, but not the 2+ ions, showed marked differences between isomers. The trend for the 1+ ions is shown in Figure 8. The lack of a trend in the fragments from the 2+ ions is likely to be due to Coulombic repulsion. A peptide with two charges requires less energy for fragmentation than a peptide with only one charge due to Coulombic repulsion. The repulsion between two charges on one peptide molecule lowers the energy barrier to fragment ion formation. This effect likely prevents observation of any subtle differences between isomers that are apparent with the 1+ ions. Such subtle differences between isomers are observed in the 1+ ion series on both instruments.

Both instruments showed differences in fragmentation when D-amino acids were substituted for L-amino acids near positions 3 and 4 of RKDVY, but the IT 1+ ion data showed a more prevalent trend than the TQ 1+ ion data. Substitution of D-amino acids near the amino-

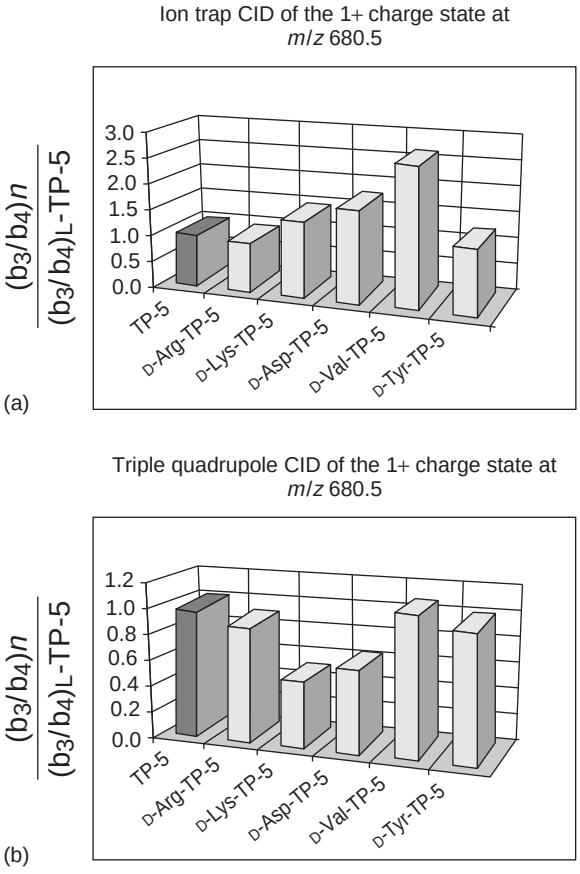

Figure 8 Comparative analysis of the tendency to form the b3 ion versus the b4 ion in the isomeric series for peptide Arg-Lys-Asp-Val-Tyr (TP-5 or thymopentin) in an IT (a) and TQ (b). The b3/b4 ratio of ion intensity for each D-amino acid isomer of TP-5 is normalized to that observed for TP-5.

and carboxyl-terminii did not affect fragmentation of the b3/b4 ratio. This is expected, based on the tendency of this given peptide sequence (i.e. chemistry) to form intense b3 and b4 ions relative to all the other possibilities. The most obvious trend is in the IT 1+ ion data (Figure 8a) where the D-Val-thymopentin isomer was the most likely to form the b3 ion relative to the b4. The D-Val-thymopentin isomer likely decreases the energy barrier to the formation of this b3 fragmentation pathway via a specific steric constraint that promotes the formation of the putative five-membered ring between the aspartic acid side chain and the ''b3'' amide bond (Figure 8a). Recall that in an all-L-amino acid peptide each amino acid in the sequence points away from the backbone in an alternating fashion. As soon as one D-amino acid is introduced, then the adjacent L- and D-amino acids will have side chains on the same side of the amide bond backbone rather than the normal where side chains project to one side or the other in an alternating fashion down the backbone. Interestingly, the 1+ ion CID data from the TQ does not show the same strong trend toward the D-Val-isomer (Figure 8b). The D-Val isomer is still the most likely to produce the strongest b3/b4 ion ratio, but now the series dips at the D-Lys isomer. In this case the D-Lys isomer must interfere with the formation of the five-membered ring by the L-Asp side chain and the ''b3'' amide bond. Given the longer time frame of the CID experiment in the TQ the D-Lys isomer must allow stabilization of a structure that prevents facile cleavage of the b3 amide bond. Whatever the physicochemical explanation, such graphical displays of data allow even the novice to find trends. In this case comparative MS/MS analysis on either instrument shows a definite influence from the substitution of an L-amino acid for a D-amino acid. As such, these comparative analyses could be used as a less labor-intensive method to screen impurities in libraries than the chiral analysis presented in section 6. Unless there is a compelling reason to use TQ/MS, then IT/MS analysis would offer the advantages of faster scans, lower instrument start-up costs and less complex CID data to interpret.

8 NONCONVENTIONAL MASS SPECTROMETRY

8.1 Noncovalent Complex Analysis

One area of MS that is nonconventional but appropriate for biological studies is the study of noncovalent complexes. In the past 10 years a number of articles have been published describing this research.[6] Basic to these studies is the use of soft ionization techniques like ESI to transfer noncovalent complexes from solution to the gas phase. Argument continues about whether or not the ionization process significantly perturbs the system under study. However, a number of studies would indicate that when proper controls using techniques such as nuclear magnetic resonance and circular dichroism are conducted, then the results can be trusted. In fact, one of the earliest studies showed that the dissociation constant K_d for a series of enzyme inhibitors had the same relative affinities in the gas phase as in solution.[45] Thermal denaturation of RNAse S in the gas phase also correlated with the change in enthalpy (ΔH) of dissociation measured by traditional calorimetry.[46] As a last application an example will be presented showing that the thermal denaturation of DNA measured in a manner analogous to that for RNAse S exhibits a melting temperature T_m close to that shown by circular dichroism.

Guanine-rich DNA can be found in the telomeric region of certain chromosomes.[47] When removed from the chromosome and analyzed as small oligonucleotides these guanine-rich sequences form quadruple-stranded complexes in the presence of cations such as K^+, Ca^{2+}

and Na^+, but dissociate into single-stranded DNA in the absence of cations. Proton nuclear magnetic resonance, circular dichroism and electrophoresis have been used to study the cation-dependent thermal stability of quadruplex DNA. Also, it has been shown previously that these quadruplex–cation structures and more common duplex structures can be transferred intact into the gas phase.[48,49] The following example shows that on-line thermal denaturation of the quadruplex–cation structures can be used to estimate the T_m of such non-covalent structures by ESI/MS.

8.2 Thermal Denaturation of Noncovalent Complexes

Liquid-phase thermal denaturation was conducted in 60 cm of an 80-cm length of 50-μm i.d. × 180-μm o.d. fused-silica capillary wrapped in heating tape, with segments of about 8 cm at the ESI sources and 12 cm at the syringe remaining unheated. Temperature was measured with a thermal couple placed inside the heating tape and next to the fused-silica capillary. For thermal denaturation studies the capillary was filled with sample and incubated at a given temperature for 10 min prior to acquiring ESI data. After the thermal equilibration period the syringe pump was turned on and the sample infused at 0.5 μL min^{-1}. The thermally equilibrated sample was pushed out of the syringe using distilled, deionized water. The synthetic oligomer for the study had the sequence d(CGCGGGGGCG). The single-stranded oligomer was prepared for ESI/MS by HPLC purification to remove low-molecular-weight impurities and residual salt. Quadruplex-DNA was prepared in the presence of 10 mM sodium phosphate (pH 7.6) containing 0.1 mM ethylenediaminetetraacetic acid (EDTA) from single-stranded d(CGCGGGGGCG) according to established protocols.[50]

For MS a home-built SQ instrument was used that was assembled from Extrel components and consisted of a nozzle-skimmer atmospheric pressure inlet followed by an RF-only quadrupole surrounded by a cryopumping surface.[44] The front plate was held at ambient temperature and the lens was at 1.0 kV. A counter-current stream of nitrogen gas at ambient room temperature was used to assist desolvation and a co-axial stream of SF_6 within the ESI source prevented discharge from the aqueous solutions. ESI was carried out in negative ion mode.

8.3 Discussion

At room temperature in 10 mM sodium phosphate (pH 7.6) and 0.1 mM EDTA, ions corresponding to both single-stranded d(CGCGGGGGCG) (i.e. G4) and quadruplex G4 (i.e. $(G4)_4$) were observed at a low nozzle-skimmer bias of −150 V (Figure 9a): so-called

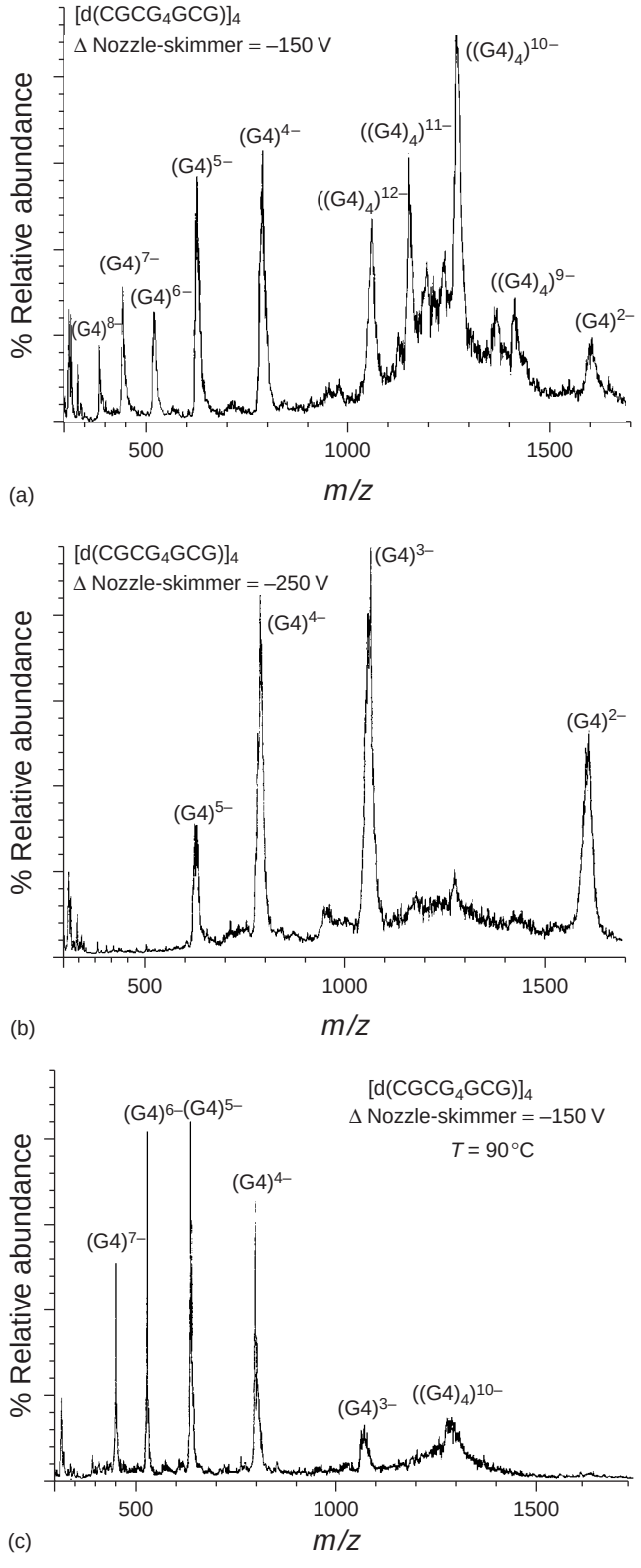

Figure 9 Thermal denaturation of quadruplex DNA ($(G4)_4$) in solution at room temperature and low nozzle-skimmer bias (a) analyzed by negative ion ESI/MS. The $(G4)_4$ DNA dissociates at high temperature (b) or high nozzle-skimmer bias (c) into single-stranded DNA (G4).

noncovalent friendly conditions on this particular instrument. The $(G4)_4$ species could be converted largely to the single-stranded G4 by raising the nozzle-skimmer bias to −250 V (Figure 9b) or by raising the solution temperature to 90 °C (Figure 9c). This simple result agrees with circular dichroism studies carried out under identical conditions of ionic strength and ionic species present, which showed that the quadruplex DNA had a T_m of 90 °C. A study of a related quadruplex $[d(CGCGGGGCG)]_4$ showed that it had a T_m of approximately 80 °C. This measurement also agreed with circular dichroism studies and the chemical intuition that one less guanine would lower thermal stability.[51] Although such studies are decidedly difficult to carry out, a number of pharmaceutical companies have made a significant investment in using ESI/MS analysis of noncovalent complexes to screen ligand libraries as a guide prior to further drug development.

9 CONCLUSIONS

As can be seen by the diversity of applications provided, the use of MS in the field of pharmaceutical sciences is vast and a single chapter cannot adequately address all the issues at hand. It covers fields as diverse as protein chemistry, combinatorial chemistry, and clinical chemistry. This article sought to address some of the more common uses of MS in pharmaceutical sciences by presenting novel approaches to quantitation of peptides in urine, chiral analysis of peptides, comparative MS/MS and thermal stability of noncovalent complexes. Each of these single areas constitutes a field of research in itself. So it is hoped that the reader gains enough insight from the discussions to have a basic understanding of each field of study presented here.

The future of MS in pharmaceutical sciences promises to include more often the use of Fourier transform ion cyclotron resonance/mass spectrometry (FTICR/MS). An excellent review of this technique was published recently and so the fundamentals of FTICR/MS operation will be left for the reader's initiative.[52] The reason why FTICR/MS is likely to become so important are fourfold: high resolution (e.g. 1 000 000 @ 1000 amu as $m/\Delta m$ at 50% peak height); high mass accuracy (at least 1 ppm @ 1000 amu; i.e. for 1 ppm @ 1000 amu the mass can be measured accurately to the thousandth place as 1000.001 amu); high sensitivity (zepto- to attomoles for small molecules); and broad dynamic range; all these advantages are combined in one instrument. Additionally, all of these benefits can be carried out on-line with HPLC or capillary electrophoresis and with either ESI or MALDI ionization. The possibility to make accurate m/z measurements to 0.1 ppm (@ 1000 amu) should make it feasible to identify proteins in sequence databases based on the accurate mass tag provided by a single peptide molecular weight.[53] This advantage will allow the identification of analytes in defined libraries without time-consuming MS/MS procedures and promises to increase significantly the rate at which samples can be screened.[54]

ACKNOWLEDGMENTS

DRG acknowledges the following people for scientific counsel and inspiration over the past 10 years, during which much of this work was done: Richard B. van Breemen, Richard D. Smith, Charles G. Edmonds, Rachel Ogorzalek Loo, Harold R. Udseth, John W. Tolan, Jeffery Shabanowitz, John R. Yates, III, Ruedolf H. Aebersold and Donna M. Goodlett. For contribution of samples, sample preparation and/or instrument time, James Blake, Charles C. Hardin, Perlette Abuaf, Tarit Mukherjee, Paul Salomons, Johnson & Johnson, Inc., Bristol-Myers Squibb, Inc. and Finnigan, Inc. are recognized. Further information regarding MS-related topics can be found at the World Wide Web site for the American Society for MS: http://www.asms.org.

ABBREVIATIONS AND ACRONYMS

CID	Collision-induced Dissociation
DNPA	1-Fluoro-2,4-dinitrophenyl-5-L-alanine-labeled
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray Ionization
ESI/IT/MS/MS	Electrospray Ionization/Ion Trap Tandem Mass Spectrometry
ESI/MS	Electrospray Ionization/Mass Spectrometry
ESI/MS/MS	Electrospray Ionization/Tandem Mass Spectrometry
ESI/TQ/MS/MS	Electrospray Ionization/Triple Quadrupole Tandem Mass Spectrometry
FDA	Food and Drug Administration
FDAA	1-Fluoro-2,4-dinitrophenyl-5-L-alanine
FTICR/MS	Fourier Transform Ion Cyclotron Resonance/Mass Spectrometry
FWHM	Full Width at Half-maximum
GLP	Good Laboratory Practice
HPLC	High-performance Liquid Chromatography
HPLC/MS	High-performance Liquid Chromatography/Mass Spectrometry

ICL	Instrument Control Language
IT	Ion Trap
IT/CID	Ion Trap Collision-induced Dissociation
LC	Liquid Chromatography
LC/ESI/MS	Liquid Chromatography/ Electrospray Ionization/ Mass Spectrometry
LC/LC/MS	Tandem Liquid Chromatography/ Mass Spectrometry
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption Ionization
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
RF	Radiofrequency
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SQ	Single Quadrupole
SRM	Selected Reaction Monitoring
TFA	Trifluoroacetic Acid
TOF	Time-of-flight
TQ	Triple Quadrupole
TQ/CID	Triple Quadrupole Collision-induced Dissociation
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Food **(Volume 5)**
Proteins, Peptides, and Amino Acids Analysis in Food

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Peptide Diastereomers, Separation of

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. M. Barber, R.S. Bordoli, R.D. Sedgwick, A.N.J. Tyler, 'Fast Atom Bombardment of Solids: A New Ion Source for Mass Spectrometry', *Chem. Soc. Commun.*, 325–327 (1981).
2. D.R. Goodlett, F.B. Armstrong, R.J. Creech, R.B. van Breemen, 'Formylated Peptides from Cyanogen Bromide Digests Identified by Fast Atom Bombardment Mass Spectrometry', *Anal. Biochem.*, **186**, 116–120 (1990).
3. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, 'Electrospray Ionization for Mass Spectrometry of Large Biomolecules', *Science*, **246**, 64–71 (1989).
4. F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**, 1193A–1293A (1991).
5. W.M.A. Niessen, A.P. Tinke, 'Liquid Chromatography/ Mass Spectrometry: General Principles and Instrumentation', *J. Chromatogr. A*, **703**, 37–57 (1995).
6. J.A. Loo, 'Bioanalytical Mass Spectrometry: Many Flavors to Choose', *Bioconj. Chem.*, **6**, 644–665 (1995).
7. J. Crowther, V. Adusumalli, T. Mukherjee, K. Jordan, P. Abuaf, N. Corkum, G. Goldstein, J. Tolan, 'Determination of Nanogram Levels of Peptide Drug in Rabbit and Human Plasma using High-performance Liquid Chromatography Coupled with Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **66**, 3256–2361 (1994).
8. A. Vidyasagar, N. Corkum, A. Jacala, T. Mukherjee, D.R. Goodlett, J.B. Crowther, I.W. McConnell, G. Goldstein, 'Pharmacokinetics and Toxicokinetics of an Orally Active Tripeptide, IRI-695, in Animals', *Biopharm. Drug Dispos.*, **17**, 25–41 (1996).
9. J. Eng, A.L. McCormack, J.R. Yates, III, 'An Approach to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in a Protein Database', *J. Am. Soc. Mass Spectrom.*, **5**, 976–989 (1994).
10. R.H. Aebersold, D.B. Teplow, L.E. Hood, S.B. Kent, 'Electroblotting onto Activated Glass. High Efficiency

Preparation of Proteins from Analytical Sodium Dodecyl Sulfate–Polyacrylamide Gels for Direct Sequence Analysis', *J. Biol. Chem.*, **261**, 4229–4238 (1986).
11. R.B. Cole, *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications*, New York, Wiley, 1997.
12. R.D. Smith, H.R. Udseth, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, 'Capillary Electrophoresis Mass Spectrometry', High Resolution Separation and Analysis of Biological Macromolecules, Part B, *Methods Enzymol.*, **271**, 448–486 (1996).
13. R.R. Ogorzalek Loo, D.R. Goodlett, R.D. Smith, J.A. Loo, 'Observation of Noncovalent Ribonuclease S-Protein/S-Peptide Complex by Electrospray Ionization Mass Spectrometry', *J. Am. Chem. Soc.*, **115**, 4391–4392 (1993).
14. J.A. Loo, R.R. Ogorzalek Loo, D.R. Goodlett, R.D. Smith, A.F. Fuciarelli, D.L. Springer, B.D. Thrall, C.G. Edmonds, 'Elucidation of Covalent and Noncovalent Associations in Proteins by Electrospray Ionization Mass Spectrometry', in *Techniques in Protein Chemistry IV*, ed. R.H. Angeletti, Academic Press, New York, 23–32, 1993.
15. R. Aebersold, D. Figeys, S. Gygi, G. Corthals, P. Haynes, B. Rist, Y. Zhang, D.R. Goodlett, 'Towards an Integrated Analytical Technology for the Generation of Multidimensional Protein Expression Maps', *J. Prot. Chem.*, **17**, 533–535 (1998).
16. R.A. Yost, C.G. Enke, 'Triple Quadrupole Mass Spectrometry for Direct Mixture analysis and Structure Elucidation', *Anal. Chem.*, **51**, 1251A–1264A (1979).
17. T.R. Covey, E.C. Huang, J.D. Henion, 'Structural Characterization of Protein Tryptic Peptides via Liquid Chromatography/Mass Spectrometry and Collision-induced Dissociation of their Doubly Charged Molecular Ions', *Anal. Chem.*, **63**, 1193–1200 (1991).
18. M. Wilm, G. Neubauer, M. Mann, 'Parent Ion Scans of Unseparated Peptide Mixtures', *Anal. Chem.*, **68**, 527–533 (1996).
19. M.J. Huddleston, R.S. Annan, M.F. Bean, S.A. Carr, 'Selective Detection of Phosphopeptides in Complex Mixtures Electrospray Liquid Chromatography/Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 710–717 (1993).
20. J. Henion, E. Brewer, G. Rule, 'Sample Preparation for LC/MS/MS: Analyzing Biological and Environmental Samples', *Anal. Chem.*, **70**, 650A–656A (1998).
21. K.R. Jonscher, J.R. Yates, III, 'The Quadrupole Ion Trap Mass Spectrometer – A Small Solution to a Big Challenge', *Anal. Biochem.*, **244**, 1–15 (1997).
22. D.R. Goodlett, A. Timperman, S.P. Gygi, J. Watts, G. Corthals, D. Figeys, R. Aebersold, 'Strategies and Methods for Proteome Analysis', in *Proteome and Protein Analysis*, ed. R.-M. Kamp, Springer-Verlag, Heidelberg, 1999.
23. J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, 'Examination of Small-diameter Capillaries for Increasing Peptide and Protein Sensitivity in CE/MS', *Electrophoresis*, **14**, 448–457 (1993).
24. J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, 'Attomole Level Capillary Electrophoresis/Mass Spectrometry', *Anal. Chem.*, **64**, 3194–3196 (1992).
25. M.A. Moseley, L.J. Deterding, K.B. Tomer, J.W. Jorgenson, 'Nanoscale Packed-capillary Liquid Chromatography Coupled with Mass Spectrometry using a Coaxial Continuous Flow Fast Atom Bombardment Interface', *Anal. Chem.*, **63**, 1467–1473 (1991).
26. D.F. Hunt, J.E. Alexander, A.L. McCormack, P.A. Martino, H. Michel, J. Shabanowitz, N. Sherman, M.A. Moseley, J.W. Jorgenson, K.B. Tomer, 'Mass Spectrometric Methods for Protein and Peptide Sequence Analysis', in *Techniques in Protein Chemistry II*, ed. J.J. Villafranca, Academic Press, New York, 441–454, 1991.
27. R. Kennedy, J.W. Jorgenson, 'Preparation and Evaluation of Packed Capillary Liquid Chromatography Columns with Inner Diameters from 20 μm to 50 μm', *Anal. Chem.*, **63**, 1467–1473 (1991).
28. S.A. Susin, H.K. Lorenzo, N. Zamzami, I. Marzo, G. Brothers, B. Snow, E. Jacotot, P. Costantini, N. Larochette, D.R. Goodlett, R. Aebersold, G. Pietu, M.-C. Prevost, D. Siderovski, J. Penninger, G. Kroemer, 'Molecular Characterization of Mitochondrial Apoptosis-inducing Factor', *Nature*, **397**, 441–446 (1999).
29. A.L. Cox, J. Skipper, Y. Chen, R.A. Henderson, T.L. Darrow, J. Shabanowitz, V.H. Engelhard, D.F. Hunt, C.L. Slingluff, 'Identification of a Peptide Recognized by Five Melanoma-specific Human Cytotoxic T Cell Lines', *Science*, **264**, 716–719 (1994).
30. N. Lee, D.R. Goodlett, H. Marquardt, D.E. Geraghty, 'HLA-E Surface Expression Depends on Binding of TAP-dependent Peptides Derived from Certain HLA Class I Signal Sequences', *J. Immunol.*, **160**, 4951–4960 (1998).
31. D.R. Goodlett, J.H. Wahl, H.R. Udseth, R.D. Smith, 'Reduced Elution Speed Detection for Capillary Electrophoresis/Mass Spectrometry', *J. Microcolumn Sep.*, **5**, 57–62 (1993).
32. R.D. Smith, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, 'Capillary Electrophoresis/Mass Spectrometry', *Anal. Chem.*, **65**, 574A–584A (1993).
33. M.T. Davis, T.D. Lee, 'Rapid Protein Identification using a Microscale Electrospray LC/MS System on an Ion Trap Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **9**, 194–201 (1997).
34. D. Figeys, G.L. Corthals, B. Gallis, D.R. Goodlett, A. Ducret, M.A. Corson, R. Aebersold, 'Data-dependent Modulation of Solid-phase Extraction Capillary Electrophoresis for the Analysis of Complex Peptide and Phosphopeptide Mixtures by Tandem Mass Spectrometry: Application to Endothelial Nitric Oxide Synthase', *Anal. Chem.*, **71**, 2279–2287 (1999).
35. A. Ducret, I. van Oostveen, J.K. Eng, J.R. Yates, III, R. Aebersold, 'High Throughput Protein Characterization by Automated Reversed-phase Chromatography/Elec-

trospray Tandem Mass Spectrometry', *Protein Sci.*, **7**, 706–719 (1998).
36. J. Cai, J.D. Henion, 'On-line Immunoaffinity Extraction-coupled Column Capillary Liquid Chromatography/Tandem Mass Spectrometry: Trace Analysis of LSD Analogs and Metabolites in Human Urine', *Anal. Chem.*, **68**, 72–78 (1996).
37. G.S. Rule, J.D. Henion, 'Determination of Drugs from Urine by On-line Immunoaffinity Chromatography/High-performance Liquid Chromatography/Mass Spectrometry', *J. Chromatogr.*, **582**, 103–112 (1992).
38. D.R. Goodlett, P. Abuaf, T. Mukherjee, J.B. Crowther, N. Corkum, G. Goldstein, 'Determination of Peptide Chiral Purity by Liquid Chromatography/Electrospray Ionization/Mass Spectrometry', *J. Chromatogr. A*, **707**, 233–244 (1995).
39. D.R. Goodlett, D.C. Gale, J.C. Crowther, T. Mukherjee, 'Interpretation of Peptide Fragment Ion Intensities for a Series of D- and L-Amino Acid Isomers', *Proceedings of 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, 692, 1996.
40. P. Roepstorff, J. Fohlman, 'Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides', *J. Biomed. Mass Spectrom.*, **11**, 601 (1984).
41. B.L. Schwartz, M.M. Bursey, 'Some Proline Substituent Effects in the Tandem Mass Spectrum of Protonated Pentaalanine', *Biol. Mass Spectrom.*, **21**, 92–96 (1992).
42. D.R. Goodlett, 'Aspartic Acid Effect in the CID of Peptides by ESI/MS/MS', *Proceedings of 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, 1333, 1997.
43. B.L. Schwartz, R.D. McClain, B.W. Erickson, M.M. Bursey, 'Differentiation Between Selected Pairs of Tripeptide Diastereomers by Tandem Mass Spectrometry on a Hybrid Tandem Mass Spectrometer', *Rapid. Commun. Mass Spectrom.*, **7**, 339–342 (1993).
44. D.F. Hunt, J.R. Yates, III, J. Shabanowitz, S. Winston, C.R. Hauer, 'Protein Sequencing by Tandem Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **83**, 6233–6236 (1986).
45. B. Ganem, Y.-T. Li, J.D. Henion, 'Observation of Noncovalent Enzyme Substrate and Enzyme–Product Complexes by Ion-spray Mass Spectrometry', *J. Am. Chem. Soc.*, **113**, 7818–7819 (1991).
46. D.R. Goodlett, R.R. Ogorzalek Loo, J.A. Loo, J.H. Wahl, H.R. Udseth, R.D. Smith, 'A Study of the Thermal Denaturation of Ribonuclease S by Electrospray Ionization Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**, 614–622 (1994).
47. W.I. Sundquist, A. Klug, 'Telomeric DNA Dimerizes by Formation of Guanine Tetrads Between Hairpin Loops', *Nature*, **342**, 825–829 (1989).
48. D.R. Goodlett, D.G. Camp, C.C. Hardin, M. Corregan, R.D. Smith, 'Direct Observation of a DNA Quadruplex by Electrospray Ionization/Mass Spectrometry', *Biol. Mass Spectrom.*, **22**, 181–183 (1993).
49. D.C. Gale, D.R. Goodlett, K.J. Light-Wahl, R.D. Smith, 'Observation of Duplex-DNA Noncovalent Complexes by Electrospray Ionization/Mass Spectrometry', *J. Am. Chem. Soc.*, **116**, 6027–6028 (1994).
50. C.C. Hardin, T. Watson, M. Corregan, C. Bailey, 'Cation-dependent Transition between the Quadruplex and Watson–Crick Hairpin Forms of d(CGCG3GCG)', *Biochemistry*, **31**, 833–841 (1992).
51. D.R. Goodlett, C.C. Hardin, M. Corregan, R.D. Smith, 'Estimation of Quadruplex DNA T_m by Electrospray Ionization Mass Spectrometry', *Proceedings of 41st ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, CA, 258a, 1993.
52. A.G. Marshall, C.L. Hendrickson, G.S. Jackson, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer', *Mass Spectrom. Rev.*, **17**, 1–35 (1998).
53. J.E. Bruce, G.A. Anderson, J. Wenn, R. Harkowitz, R.D. Smith, 'High Mass Measurement Accuracy for 100% Sequence Coverage of Enzymatically Digested Bovine Serum Albumin from an ESI/FTICR Mass Spectrum', *Anal. Chem.*, **71**, 2595–2599 (1999).
54. B. Rist, D.R. Goodlett, S.P. Gygi, J.E. Bruce, G.A. Anderson, R.D. Smith, R. Aebersold, 'Evaluation of Mass Resolution on a Quantitative Method of Protein Profiling', *Proceedings of 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 283–284, 1999.

Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Brian Dawson
Bureau of Biologics and Radiopharmaceuticals, Health Canada, Ottawa, Canada

1 Introduction **1**
2 Analysis of Drugs **2**
2.1 Sample Types 2
2.2 Sample Preparation 2
2.3 Nuclear Magnetic Resonance Experiments 3
3 Applications **9**
3.1 Confirmation of Known Compounds 9
3.2 Identification of Unknown Compounds 9
3.3 Quantification 10
3.4 Impurity Profiling 11
3.5 Optical Purity 11
3.6 Analysis of Body Fluids 11
4 Comparison with Other Techniques **11**
4.1 Advantages 11
4.2 Disadvantages 12
5 Conclusions **12**
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

Nuclear magnetic resonance (NMR) spectroscopy is an extremely powerful tool for the detection, identification and quantification of drugs and related substances. The whole range of one-dimensional (1-D) and two-dimensional (2-D) NMR techniques is available for performing the required analyses. These NMR methods may be used for routine purposes, such as the confirmation of the identity of the drug or quantification of the amount of drug substance present in a formulation. However, the area where NMR stands out as an analytical tool is in the identification of unknown compounds such as metabolites or drug degradation products. NMR is also used for impurity profiling or determination of the drug's optical purity.

Although NMR has been used for many years to analyze drugs, even the most modern spectrometers lack the sensitivity obtainable by other techniques such as mass spectrometry (MS) or high-performance liquid chromatography (HPLC). However, NMR is a nondestructive technique which provides essential structural information that cannot easily be obtained from these other methods. Recent advances in technology have allowed the interfacing of NMR instruments with other equipment such as HPLC devices to provide even more robust analytical methodology.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The nuclei of atoms with an odd number of protons, neutrons, or both, have an intrinsic nuclear *spin*. The spinning of charged particles – the circulation of charges – produces a magnetic moment along the axis of spin. The simplest nucleus possessing a spin is the proton – the nucleus of the hydrogen atom. If hydrogen atoms are placed in an external magnetic field, their magnetic moments, according to quantum mechanics, will become aligned in one of two ways – with or against the external field. For hydrogen, the more stable alignment is with the field and energy must be absorbed to ''flip'' the proton over to the less stable alignment against the field. The amount of energy needed to flip the proton depends on the strength of the external field: the stronger the field, the greater the energy required to produce the flip. The energy required corresponds to electromagnetic radiation in the radiofrequency (RF) range. The observation frequency is given by Equation (1)

$$\nu = \frac{\gamma H_0}{2\pi} \tag{1}$$

where ν = frequency, in cycles per second; γ = a nuclear constant (the gyromagnetic ratio) and H_0 = strength of the static magnetic field, in tesla. When the nucleus is irradiated at the proper frequency, absorption occurs, followed by the nuclei re-emitting the RF radiation to return to the lower energy state. This process gives rise to a signal and the spectrum arising from this signal is called an NMR spectrum. The frequency at which a nucleus absorbs depends on the magnetic field that it *feels*, and this *effective* field strength is not exactly the same as the *applied* field strength. The effective field strength at each nucleus depends on its environment – the electron density, closeness of other nuclei, etc. – and hence in each case a slightly different applied frequency is required in order for absorption to take place. Thus each different nucleus will give rise to its own signal.

There are two basic NMR spectrometer designs – continuous wave (CW), and pulsed or Fourier transform

(FT). CW spectrometers have low-field water-cooled electromagnets which consume large amounts of electricity. The newer Fourier transform nuclear magnetic resonance (FTNMR) spectrometers use liquid-helium-cooled superconducting magnets which consume very little electrical power, but require relatively expensive cryogens (liquid nitrogen and liquid helium) to keep them working. Although CW spectrometers are perfectly suitable for many routine applications, they have largely been replaced by the higher-field pulsed FTNMR instruments.

Because the energy difference between the nuclei aligned with and against the field is small, the number of nuclei in each spin state are almost equal. There is, however, a slight excess of spins in the more stable alignment (of the order of 1 in 10^5 for protons at 200 MHz). It is the signal from this excess which gives rise to the NMR signal, thus explaining the relative insensitivity of this technique. However, over the last 20 years, the field strength (and thus sensitivity) has grown by leaps and bounds. Even the more routine modern instruments (300–500 MHz) are capable of producing good quality proton NMR spectra with microgram amounts of small drugs and related compounds. Lower field spectrometers are useful for acquiring routine spectra from smaller molecules where sample quantity is not limited. Higher-field magnets (currently up to 18.8 T, with a proton frequency of 800 MHz) are required for more demanding research applications.

When regulatory laboratories are performing analyses of pharmaceutical products, the analytical techniques used must meet a number of strict criteria. Besides being sufficiently sensitive, they must be highly selective, reproducible and suitable to give qualitative and, in many cases, quantitative analyses. Other desirable features for more routine pharmaceutical methods include that they require the minimum amount of sample pretreatment and that they be able to handle mixtures of substances without prior separation. As will be seen from the following discussion, NMR spectroscopy meets all these requirements and is becoming more and more important for identification of unknown substances, as well as for routine analyses, in many pharmaceutical laboratories throughout the world.

Organic compounds, including pharmaceuticals, are composed of relatively few chemical elements (hydrogen, carbon, oxygen, nitrogen, and phosphorus, as well as the halogens and sometimes metals). Each of these elements has one or more isotopes that may be detected by NMR experiments. However, due to low natural abundance and/or low relative sensitivity of the nuclei, many of these elements are not routinely studied. Those that are commonly studied include hydrogen (^{1}H), carbon (^{13}C), fluorine (^{19}F) and phosphorus (^{31}P). In order to study the structure of larger macromolecules (proteins, for example), it is usually necessary to enrich (i.e. label) the samples with ^{13}C, ^{15}N and sometimes ^{2}H (see section 2.3.2.5).

2 ANALYSIS OF DRUGS

2.1 Sample Types

The most important role NMR plays in pharmaceutical analysis is its use in elucidation and/or confirmation of structures for drug-related substances. This would include the drugs themselves, either as raw materials or in formulations. NMR has also been used extensively to study drug impurities including solvents, synthetic precursors, synthetic intermediates, and decomposition products. In the case of natural products NMR may be used to determine the identity of co-extractives. Another key area where NMR has proved to be invaluable is in the study of drug metabolism where it has been used for identification and quantification of many metabolites.

In addition to the high-resolution solution NMR techniques, several other magnetic resonance methodologies are sometimes used in pharmaceutical studies. Solid-state NMR has been used to study polymorphism of drug powders. Micro-imaging NMR may be used to study dissolution of drug tablets. The use of whole-body imaging NMR is a very powerful tool for clinical diagnosis. Extensive discussion of these areas is beyond the scope of this article.

2.2 Sample Preparation

The methods used to prepare the pharmaceutical samples for NMR analyses depend on the source of the sample and the information required. Procedures range from simple dissolution to complex extraction and derivatization reactions which must be done under very stringent conditions.

For the simplest case of ''pure'' drug raw material from a small pharmaceutical, a suitable quantity (usually a few milligrams) of the sample is dissolved in a deuterated NMR solvent. The choice of solvent depends on the solubility and stability of the drug. Commonly used deuterated solvents include chloroform ($CDCl_3$), water (D_2O), dimethylsulfoxide (DMSO-d_6), methanol (CD_3OD) or a mixture of chloroform and methanol ($CDCl_3/CD_3OD$). The amount of sample used depends on the quantity of sample available, the field strength of the spectrometer and the type of NMR experiments to be performed (see section 2.3).

In cases where the pharmaceutical product is a mixture, such as in tablets or elixirs where the drug is mixed with excipients such as sugars or starch, sample preparation

methods depend on the nature of the components. It is often possible to choose a solvent which will dissolve only the drug, leaving the other excipients behind, or to choose two immiscible solvents where the drug will go into one solvent while the excipients go into the other. For example, deuterochloroform will extract many drugs from formulations where the excipients are sugars, since sugars will not dissolve in this solvent but may be taken up in water. In many situations, however, this simple procedure will not work and one must use the chemical characteristics of the drug to separate it from its excipients. Drugs may conveniently be divided into four general categories – acidic, basic, amphoteric and neutral. Each category has different partition coefficients between aqueous and immiscible organic solvents. The hydrogen ion concentration in the aqueous phase, which varies with pH, affects the ionic state or polarity of the drug, depending on the polar substituents present. It is often possible to separate drugs from the different classes from each other or from other organic compounds by proper adjustment of the pH. For example, if the aqueous solution is made acidic, organic solvents may be used to remove acidic and neutral components. On the other hand, if the aqueous phase is made basic, basic and neutral components may be extracted with organic solvents. For amphoteric drugs such as morphine, the pH must be adjusted to its isoelectric point before extraction is possible. At this pH, the drug exists in a nonpolar or nonionized form, making it more soluble in appropriate organic solvents than in water. It is possible to isolate or remove neutral drugs or other organic components from acidic and basic drugs by extracting a basic aqueous solution with an organic solvent, and then washing this solvent with aqueous acid. Only the neutral materials should remain in the organic solvent.

Other procedures such as extraction and derivatization may be required for drugs derived from plant materials. Special precautions may be necessary when the drugs in these materials are not stable under certain conditions – in the presence of acid or base for example. However, because of the complexity of the extracts from most plants, NMR is usually not the method of choice for analyzing these materials.

For drugs where only one optical isomer is allowed in the product, or where there is a limit set on the isomeric content, procedures involving the addition of chiral shift reagents, chiral solvating agents (CSAs), or reaction with chiral derivatizing agents (CDAs) are required (see section 2.3.1.4).

In cases where the study will include the modern ''hyphenated techniques'' (NMR/HPLC, NMR/MS, HPLC/NMR/MS, etc.) the sample preparation may have to be changed to accommodate these other methodologies (see section 2.3.3).

When the pharmaceutical product under study is a protein or polypeptide, considerable effort may be necessary to find optimal conditions (pH, temperature, buffer strength, etc.) at which the sample is soluble, stable and nonaggregating at relatively high concentrations (usually >1 mM).

2.3 Nuclear Magnetic Resonance Experiments

Since the mid-1980s, NMR techniques have advanced dramatically. As the number of experiments increases, many spectroscopists find it difficult to keep up with all the newest procedures available to help solve problems. Most develop a set of experimental techniques which work well for them in obtaining the kind of information they require. What follows is a discussion of the general types of experiments which are most often used in the analyses of pharmaceuticals.

2.3.1 One-dimensional Techniques

2.3.1.1 High-resolution Spectra When most scientists with a passing knowledge of NMR think about this technique, they recall the traditional 1-D high-resolution proton (^{1}H) and carbon (^{13}C) spectra. These spectra are highly informative in the hands of a trained NMR spectroscopist. For those who may not be experts in the field, these spectra may still be used as ''fingerprints'' for comparison with spectra of authentic materials. For this type of work, the use of ^{13}C-NMR spectra is most suitable. ^{1}H-NMR spectra are not as convenient for a number of reasons, the main one being that the appearance of the proton spectra differs significantly when obtained at different field strengths – the higher the field, the more first-order (and thus simpler) the spectra become. At lower field strengths, the ^{1}H-NMR spectra for many drugs consist of broad and uninterpretable envelopes of signals, whereas at higher fields the corresponding spectra contain large numbers of sharp resonances which may be readily assigned using some of the modern techniques discussed below. This may be clearly seen by comparing the 80 MHz and 400 MHz spectra of phencyclidine hydrochloride (Figure 1). On the other hand, ^{13}C spectra look very similar, regardless of the field strength used. The only differences would be for those signals which may only be resolved from each other at higher field strengths.

^{1}H-NMR spectra can give detailed information on the nature of unknown substances. Each different proton in the molecule gives rise to a peak or group of peaks. When there is no overlap in the ^{1}H-NMR spectrum, integration may be used to obtain the relative number of each type of proton. The chemical shift of the protons indicates the environment in which they exist. The majority of protons fall in a range of about

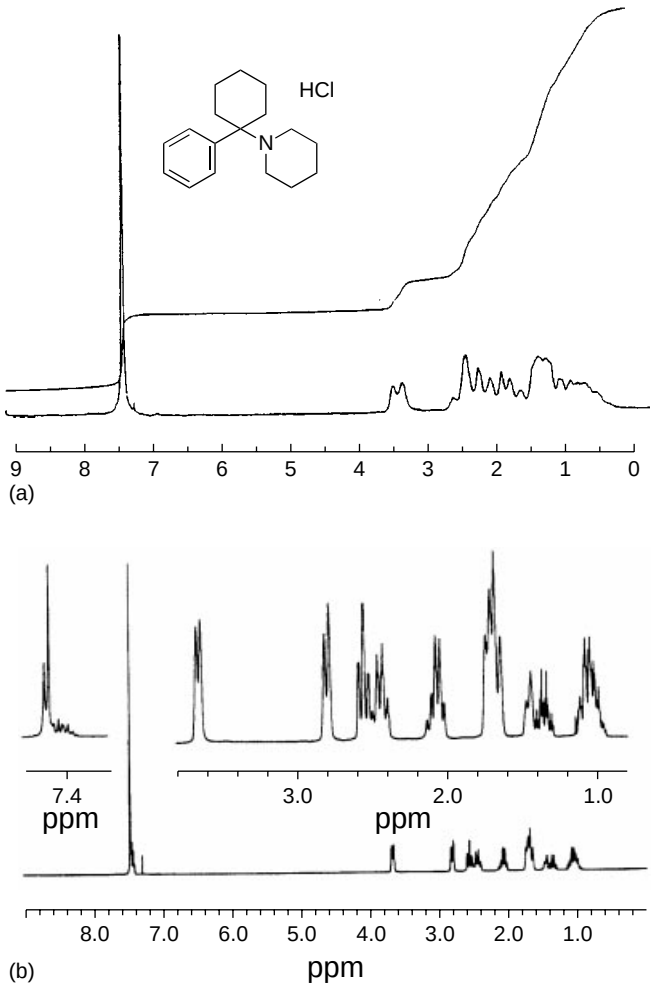

Figure 1 Structure and ^{1}H-NMR spectra for phencyclidine hydrochloride: (a) 80 MHz spectrum; (b) 400 MHz spectrum (blow-ups of the two chemical shift regions illustrate the sharpness of the peaks compared with those of the 80 MHz spectrum in (a)).

10 ppm. The chemical shifts in organic solvents are usually referenced to tetramethylsilane (TMS), which is assigned a value of 0.00 ppm. In aqueous solutions, chemical shifts may be referenced to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), also at 0.00 ppm. Chemical shifts provide evidence as to whether the protons are aliphatic (0.8–3.0 ppm), olefinic (4.5–6.5 ppm), or aromatic (6.5–8.5 ppm). Protons attached to carbons that are bonded to nitrogen resonate at about 2.0–3.5 ppm, while if the carbon is bound to an oxygen, the protons resonate at about 3.3–4.5 ppm. Aldehyde protons are found around 8.5–10.0 ppm. Complete tables of chemical shifts may be found in most introductory NMR text books. These would include older books such as those by Bible,[1] Abraham,[2] and Emsley et al.,[3] as well as some more recent ones such as that by Günther.[4] A collection of ^{1}H-NMR spectra for a large number of compounds may be

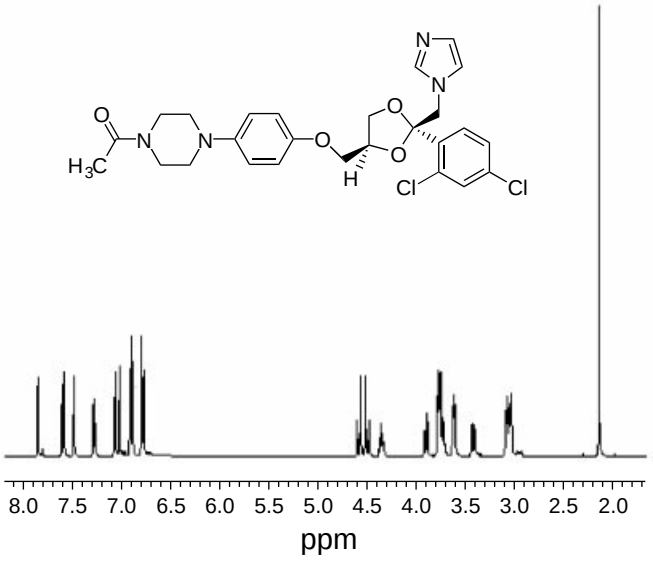

Figure 2 Structure and ^{1}H-NMR spectrum of ketoconazole.

found in the *Aldrich Library*.[5] Examination of the proton spectrum of the drug ketoconazole (Figure 2), shows resonances for many types of protons in the chemical shift ranges indicated.

High-resolution ^{13}C-NMR spectra also provide a great deal of information. The chemical shift range for carbon is more than 200 ppm. As with protons, the chemical shift ranges are indicative of the kind of carbon. Carbon chemical shifts follow basically the same order as protons, with aliphatic carbons being at highest fields, followed by olefinic and then aromatic carbons. Acid and ester carbonyls are at lower fields than aromatic carbons, followed by aldehydes and isolated (nonconjugated) carbonyls at lowest fields. As with protons, tables of typical chemical shifts may be found in most introductory NMR text books which discuss ^{13}C-NMR. These would include older books such as those by Stothers,[6] Wehrli and Wirthlin[7] and Levy and Nelson[8] as well as the newer books mentioned previously.[4,5] Figure 3(a) shows the ^{13}C-NMR spectrum of ketoconazole, with resonances in the ranges indicated.

The number of carbon signals (excluding those from the solvent) usually gives a direct count of the number of chemically distinct carbons in the molecule being studied. Exceptions occur when two or more different carbons have chemical shifts that are too close to be resolved under the conditions of the experiment, when one or more carbons are coupled to another magnetic nucleus that is not decoupled (^{19}F or ^{31}P, for example), or in special circumstances when equilibria between two or more forms of the drug exist in solution (due to hindered rotation, ring inversion, keto–enol equilibria, etc.). In a routine ^{13}C-NMR spectrum, the signals obtained for

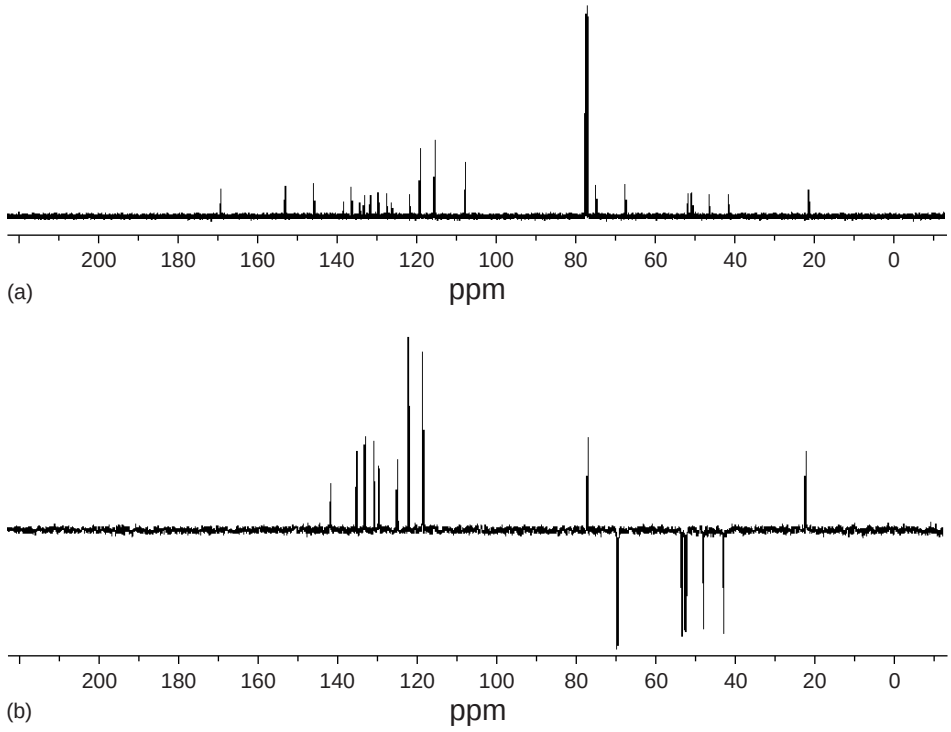

Figure 3 ^{13}C-NMR spectra for ketoconazole: (a) normal carbon spectrum (large peaks at 77 ppm are due to the solvent $CDCl_3$); (b) distortionless enhancement by polarization transfer (DEPT) spectrum (CH and CH_3 resonances give positive peaks, CH_2 gives negative peaks, quaternary carbons – including the $CDCl_3$ solvent – do not give any peaks).

quaternary carbons are smaller than those for protonated carbons. It is possible to obtain carbon spectra such that each carbon gives rise to the same size peak by using long relaxation times between scans and taking precautions to eliminate the nuclear Overhauser effect (NOE) (see section 2.3.1.2). When this type of quantitative experiment is performed, one can easily determine if there are unresolved carbon shifts or if any carbons are coupled to any other nuclei, thus enabling the analyst to get an accurate count of the number of carbons in the molecule.

One very important class of 1-D experiment is that which gives a count of the number of protons attached to each carbon. Perhaps the best known of these is DEPT. With a suitable choice of experimental parameters, the ^{13}C spectra obtained will have positive peaks for CH and CH_3 groups, negative peaks for CH_2 groups and no signal for quaternary carbons. Varying the parameters allows one to obtain spectra with only CH or CH_2 or CH_3 carbons appearing. Thus one can readily determine the number of protons attached to every carbon in the compound being studied. The DEPT spectrum for ketoconazole is shown in Figure 3(b). Comparison with the regular carbon spectrum (Figure 3a) illustrates how easy it often is to determine the number of protons attached to each carbon.

Although it is more common to obtain high-resolution spectra for protons and carbons, spectra may be obtained for many other nuclei. Those which are most often studied include ^{15}N, ^{19}F and ^{31}P. Many drugs contain nitrogen, but the low natural abundance and relative insensitivity of the ^{15}N nucleus make it a very difficult atom to study without isotopic labeling. Comparatively large amounts of material are required in order to obtain spectra in a reasonable period of time. Despite this, a number of reports have been published containing ^{15}N chemical shift data for pharmaceuticals. The other two nuclei (^{19}F and ^{31}P) are essentially 100% abundant and are NMR sensitive, but unfortunately they occur in relatively few drugs.

2.3.1.2 Nuclear Overhauser Effect When magnetic nuclei are spatially close to each other, there is a dipolar interaction. If the chemical shift of one of these nuclei is irradiated in order to saturate its spin, any nuclei that are close to it may experience a small change in the size of their signal, referred to as the NOE. For small molecules, such as traditional pharmaceuticals, this effect is usually a signal enhancement. (For larger proteins, the effect may be close to zero or even negative.) The closer the nuclei are to each other, the greater the effect on the signal. The maximum amount of enhancement for small molecules is only a few percent of original size of the signal. Because the strength of the dipolar interaction falls off at a rate which is proportional to

the inverse of the distance raised to the sixth power, nuclei must normally be within about 5 Å of each other to see any observable enhancement. To observe the small signal enhancements, one must obtain a difference spectrum by subtracting a reference spectrum obtained under ''identical'' conditions. (The only difference is that the reference spectrum is obtained while irradiating at a blank part of the spectrum instead of at the chemical shift frequency for one of the nuclei.) When the difference spectrum is plotted, one obtains a large negative peak for the irradiated nucleus and small positive peaks for nuclei experiencing an observable NOE. Thus this procedure often indicates which protons are close to each other in space, thus allowing the spectroscopist to determine the relative stereochemistry in unknown molecules.

2.3.1.3 Lanthanide Shift Reagent Studies The magnetic moments of the unpaired electrons of paramagnetic reagents are capable of producing relatively large chemical shift changes in molecules which are in close proximity. Most paramagnetic reagents are detrimental to the appearance of NMR spectra because they cause severe line-broadening of all the signals. However, a few complexes of the rare earth elements (lanthanides) do cause substantial chemical shift changes without substantial line-broadening on lower-field-strength instruments. The most common of these, referred to as lanthanide shift reagents (LSRs), are composed of complexes of europium, praesodymium or ytterbium with a number of different diketones. The most popular LSRs are complexes with 6,6,7,7,8,8,8-heptafluoro-2,2-dimethyloctane-3,5-dione [$Eu(fod)_3$, $Pr(fod)_3$ and $Yt(fod)_3$] or 2,2,6,6-tetramethylheptane-3,5-dione [$Eu(dmp)_3$, $Pr(dmp)_3$ and $Yt(dmp)_3$].

When a shift reagent is added to a solution of the pharmaceutical, it forms a weak bond to the most basic site in the drug – usually a nitrogen or oxygen. The amount of change in the chemical shift is inversely proportional to the distance between the lanthanide atom and the proton (or carbon). The experimental procedure followed in this type of study is to obtain a spectrum of the drug alone and in the presence of varying amounts of the shift reagent. By following the shifting of the different nuclei, one can determine their relative distances from the reagent, thus providing help in assigning the shifts to the appropriate nuclei. In the case of proton studies, the spreading out of the resonance signals may cause the spectrum to be more first-order and thus easier to interpret.

The amount of line-broadening experienced by nuclei with a given concentration of LSR is directly proportional to the square of the magnetic field strength. Thus, these reagents are not suitable for use on the more modern high-field spectrometers. However, on lower-field instruments they have been used for a wide variety of drugs over the years.

2.3.1.4 Optical Purity Methods Another important class of 1-D experiments is that used to determine the optical isomer content of drugs. For a number of drugs only one isomer should be present in the commercial product. There are three basic types of NMR experiments which may be used to determine the optical purity of pharmaceuticals. The first method to be developed was the use of chiral lanthanide shift reagents (CLSRs). A related method developed later was the use of chiral solvents or CSAs. The final method is the use of CDAs.

The use of LSRs was discussed in section 2.3.1.3. A number of shift reagents have optically active ketone groups. Some of the common CLSRs include complexes with the ligands 3-(trifluoroacetyl)-*d*-camphor [$Eu(tfc)_3$], 3-(heptafluorobutyryl)-*d*-camphor [$Eu(hfc)_3$] and *d*,*d*-dicampholylmethane [$Eu(dcm)_3$]. These materials work in basically the same manner as LSRs, but when these compounds are used they may induce different chemical shift changes for nuclei of optical antipodes. The CLSRs have been used to obtain optical isomer contents for many drugs.

An alternative to CLSRs is the use of one of a series of compounds referred to as chiral solvents or CSAs. These are optically active compounds possessing an aromatic group and a polar functional group such as an amine, alcohol or acid. When a suitable CSA is added to a solution of a drug which also has a polar group, the CSA forms a weak hydrogen bond with the drug. Because the CSA is optically pure, it forms different complexes with each optical antipode of the drug. The aromatic ring current of the CSA is located at different distances from the corresponding nuclei in the drug isomers. This causes the nuclei of the drug to give rise to separate signals for each antipode, thus enabling the optical purity of the drug to be determined. Unlike the CLSRs, no line-broadening occurs when CSAs are used. Higher-field spectrometers provide better separation of signals than lower-field instruments. The structures of some of the more common CSAs are shown below **(1–8)**.

A third method for determining the optical isomer content of drugs is the use of CDAs. As with CSAs, the CDAs usually contain an optical center and an aromatic group. They also contain a functional group (such as an acid chloride) which is capable of reacting and forming a covalent bond with the drug. As with the CSAs, the aromatic group is in different proximity to the nuclei of the drug's antipodes, causing differential shifts in some of the corresponding resonances. In many cases, this difference is sufficient to allow accurate quantification of each isomer. Once again, there is no line-broadening as with CLSRs and the higher-field

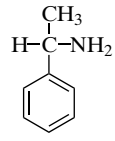

(*S*)-(−)-α-Methylbenzylamine

(1)

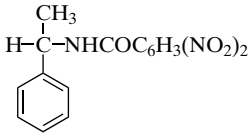

(*R*)-(−)-*N*-(3,5-Dinitrobenzoyl)-α-methylbenzyamine

(2)

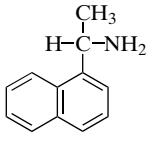

(*S*)-(−)-1-(1-Naphthyl)ethylamine

(3)

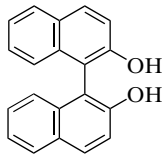

(*R*)-(+)-1,1′-Bi-2-naphthol

(4)

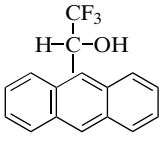

(*R*)-(−)-2,2,2-Trifluoro-1-(9-anthryl)ethanol

(5)

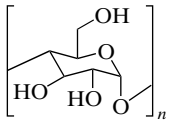

α-Cyclodextrin (*n*=6)
β-Cyclodextrin (*n*=7)
γ-Cyclodextrin (*n*=8)

(6-8)

spectrometers give better separation than the lower-field instruments.

2.3.2 *Two-dimensional Techniques*

2-D techniques in NMR spectroscopy all involve multiple-pulse experiments containing three basic stages – preparation, evolution and detection. The spectrum that is finally obtained depends on the nature of the preparation and the length of the evolution time. The effects of evolution are not observed directly, but are detected by carrying out a series of experiments in which the evolution time is systematically varied. By following this procedure, two distinct sets of spectroscopic features are obtained – those that influence the magnetization during the evolution time and those that influence the magnetization during the acquisition time. In 1-D NMR Fourier transformation is used to convert the time dependence of the magnetization evolution during the acquisition time into frequency data, while 2-D NMR uses a second Fourier transformation to convert the time dependence of the evolution time into a second frequency.

2.3.2.1 *$^{1}H-^{1}H$ Correlation Spectroscopy Experiments*

One class of 2-D experiments, which may be performed on any modern high-resolution NMR spectrometer, is referred to as correlation spectroscopy (COSY). One of the most common 2-D experiments is $^{1}H-^{1}H$-COSY. The results for this type of experiment are usually displayed as a 2-D contour plot with the 1-D proton spectrum plotted along both the *x*- and *y*-axes. A typical COSY spectrum is shown in Figure 4 for the drug ketoconazole. The contours in this plot indicate the chemical shifts of the protons that are coupled to each other. Thus, to determine which protons are coupled to a particular proton resonance, one simply scans down a column beneath the peak plotted on the *x*-axis. Contours will be found on rows corresponding to peaks plotted on the *y*-axis for protons to which there is detectable coupling. By varying the experimental parameters, one can look for either larger short-range or smaller long-range coupling constants. This experiment allows the spectroscopist to determine which protons are separated from each other by two, three or even four bonds.

2.3.2.2 *$^{13}C-^{13}C$ Correlation Spectroscopy Experiments*

The other homonuclear shift correlation experiment which may be used by spectroscopists is $^{13}C-^{13}C$-COSY. This technique is often referred to in the literature by the acronym INADEQUATE (incredible natural abundance double quantum transfer experiment). This is a very powerful technique which shows all the

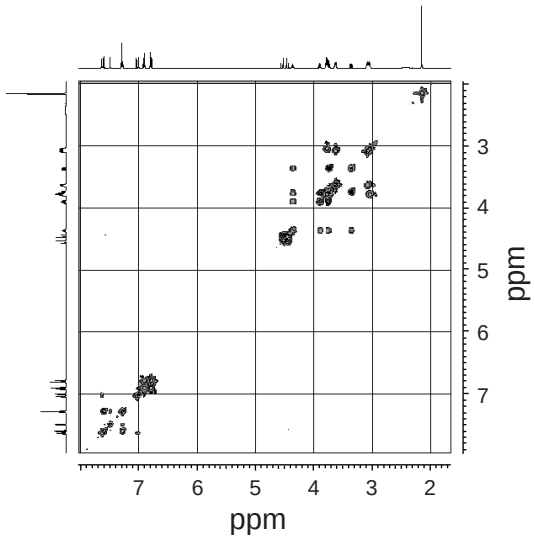

Figure 4 $^{1}H-^{1}H$-COSY spectrum of ketoconazole. Contours can be used to determine which protons are coupled to each other by examining the proton spectra plotted along the top and side.

one-bond carbon–carbon connections in the molecule, enabling the complete skeleton of the compound to be determined, except for those parts joined by heteroatoms. However, because of the extremely low sensitivity of this experiment, and the length of time required to accumulate the data, it is seldom used. Instead, spectroscopists use a combination of other techniques to get the same information much more quickly.

2.3.2.3 $^{13}C-^{1}H$ Correlation Spectroscopy Experiments Because of the importance of determining the connections between hydrogens and carbons, a large number of $^{13}C-^{1}H$-COSY experimental techniques have been developed. In the 2-D contour plots for this type of experiment, the carbon spectrum appears along one axis and the proton spectrum along the other. Parameters may be chosen to emphasize large coupling constants to observe one-bond couplings, or to emphasize smaller couplings so that two- and three-bond couplings may be observed as well. A $^{13}C-^{1}H$-COSY spectrum showing one-bond couplings for ketoconazole is presented in Figure 5. In a related experiment, called relayed coherence transfer (RCT), parameters are used which give contours for each carbon atom corresponding to the directly attached protons as well as all the protons coupled to them (i.e. protons on nearby carbons).

2.3.2.4 Nuclear Overhauser Effect Spectroscopy In section 2.3.1.2 the 1-D NOE experiment was discussed. There is a corresponding 2-D version of this experiment which goes by the acronym NOESY (nuclear Overhauser effect spectroscopy). Although the $^{1}H-^{1}H$ version of this experiment is more popular with the pharmaceutical analyst, the $^{13}C-^{1}H$ version is also available. In the 2-D plot, contours are found for each pair of nuclei which would have produced a medium to strong NOE in the 1-D counterpart. In the 2-D experiment, some weak interactions that are detectable in the 1-D method are sometimes missed. However, in a single 2-D experiment one can get most of the results that would be obtained in many 1-D NOE experiments.

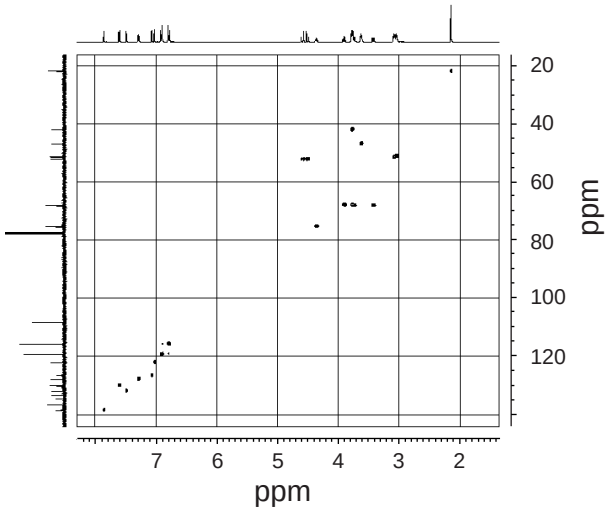

Figure 5 $^{13}C-^{1}H$-COSY spectrum of ketoconazole. Contours can be used to determine which carbon resonance (plotted along the side) is associated with each proton resonance (plotted along the top). This experiment was set up to determine which protons are directly attached to which carbon. By varying the acquisition conditions, it is possible to obtain a plot showing which carbons are 2 or 3 bonds away from each proton (data not shown).

During recent years, advances in NMR technology have led to techniques which provide increased sensitivity over these 2-D methods just discussed. These advances include the introduction of ''inverse'' or ''reverse'' detection probes, gradient probes, and decouplers which are capable of producing shaped pulses. These exciting advances mean that the pharmaceutical analyst can determine the structures of unknown samples with ever decreasing amounts of material.

2.3.2.5 Protein Techniques The proton spectra of individual amino acids are relatively easy to obtain and interpret. However, in proteins the large number of signals creates extensive overlap, making it impossible to interpret the normal 1-D spectrum. For larger proteins, even the 2-D experiments discussed previously are often not sufficient to allow chemical shift and structural assignments. Use of higher order (three-dimensional (3-D) and four-dimensional (4-D)) experiments has allowed the spreading of signals into more dimensions, making them more tractable. To perform many of these multidimensional experiments, it is necessary to have protein that is enriched (i.e. isotopically labeled) with ^{13}C and ^{15}N (and sometimes ^{2}H).

There are several problems involved in performing higher dimensional experiments. Each increase in dimension greatly increases the experimental time required to obtain the data. The multidimensional experiments require much higher concentrations of protein to reduce the amount of signal averaging for each acquisition. The 3-D and 4-D experiments require more complex spectrometer hardware and probes than the 1-D and 2-D techniques. The resulting data in the higher dimensional experiments also have reduced digital resolutions in some dimensions.

One class of 3-D experiments involves combining two 2-D ^{1}H experiments. This serves to reduce the overlap of signals by spreading them into the third dimension. The advantage of this type of experiment is that it does not require any isotopic labeling. A second class of 3-D experiments combines a 2-D ^{1}H experiment with a 2-D $^{1}H-X$ nucleus experiment (where X is usually ^{13}C or ^{15}N). This type of experiment takes advantage of the

wider shift range of the *X* nucleus and also allows the chemical shifts for the *X* nuclei to be assigned. This class of experiment requires that the protein be labeled with the *X* nucleus. Usually, data from several of these experiments are required to make full assignments for larger proteins. A third class of experiments is used to trace connectivities along the protein peptide backbone or along the side-chains. These experiments, which correlate three different nuclei (^{1}H, ^{13}C and ^{15}N), require doubly labeled protein. The NMR probes must have at least three frequency channels and the spectrometer must have as many as five transmitters. These experiments allow determination of correlations between sequential residues. For example, some experiments allow the correlation of the amide proton on one residue with the α-carbon of the preceding residue. Additional new experiments and refinements of older experiments have appeared over the last several years, and no doubt will continue to do so.

2.3.3 Hyphenated Techniques

One of the drawbacks of NMR spectroscopy has traditionally been its inability to easily positively identify small amounts of compounds in the presence of large amounts of other material. If the compound had resonances that were not overlapped with those of the main material, one could get some evidence of its presence and of its level, but because most of its resonances could not be observed, the identification usually had to be confirmed by another method. With techniques such as MS, gas chromatography (GC) or HPLC is used to separate each component before analysis. Over the last several years considerable work has been done on coupling NMR with HPLC in an analogous manner (see Keifer[9] and references therein).

A number of problems had to be solved in order to make HPLC/NMR a feasible technique for pharmaceutical research. One of the main concerns was with the sensitivity of NMR spectrometers. Since only small amounts of sample are injected onto analytical columns, older NMR instruments could not be used to detect the components coming off the column. Also, the mobile phase used on the HPLC systems gave rise to large peaks in the NMR spectrum, making it very difficult to observe the resonances for the drug-related material of interest. These problems have largely been addressed and HPLC/NMR spectroscopy is now a more routine, commercially available technique that is being used in many pharmaceutical laboratories around the world. There have recently been reports of combining this technique with MS (HPLC/NMR/MS) to produce a technique which is extremely powerful for pharmaceutical analysis.

Over the last few years there have been many reports of using NMR spectrometers as the detectors for a variety of other separation techniques. These include capillary HPLC, capillary electrophoresis (CE), capillary zone electrophoresis (CZE), and capillary electrochromatography (CEC). Although not yet commercially available, these techniques may soon provide pharmaceutical researchers with very powerful structural elucidation methods in their studies of drug-related materials.

3 APPLICATIONS

3.1 Confirmation of Known Compounds

While the majority of drugs may be just as easily identified by other methods, such as infrared (IR) spectroscopy and MS, NMR is used in many pharmaceutical laboratories. Some drugs that possess hygroscopic or polymorphic properties are difficult to analyze by IR spectroscopy. Also, many drugs fragment in such a manner that their mass spectra are ambiguous and difficult to interpret, especially in cases where there may be a number of closely related isomers. NMR provides an attractive alternative approach for these types of drugs since it is not affected by these kinds of problems and provides a very powerful method for distinguishing between isomeric drugs.

3.2 Identification of Unknown Compounds

NMR spectroscopy provides the analyst with one of the most powerful techniques for the identification of unknown compounds. For laboratories with modern spectrometers, the full range of experiments described in section 2.3 is available. While IR spectroscopy and MS are useful for identification of known compounds, especially by means of their library search capabilities, NMR is usually a key spectroscopic technique for identifying true unknowns.

The exact series of NMR experiments used to solve unknown structures may vary somewhat, depending on the preferences of the particular spectroscopist. However, the general procedure will be the same. The analyst will carry out the following series of experiments until enough information has been obtained to identify the unknown, or until enough information is obtained so that the investigator will be able to determine which specialized experiment is required to complete the study.

First, the drug-related sample is prepared for study as described in section 2.2. Then the spectroscopist obtains a high-resolution ^{1}H-NMR spectrum using an appropriate deuterated NMR solvent. Careful inspection of this spectrum gives a great deal of structural information. Chemical shifts and integration reveal the relative number of protons in each area (aliphatic, olefinic, aromatic) and the coupling patterns may reveal information on the proximity of many protons. Next, ^{13}C-NMR and

DEPT spectra are obtained to determine the number of different carbons and the numbers of protons attached to each. Chemical shifts of the carbons give important information about each carbon's environment. Often, the combination of proton and carbon data is enough to solve the structure of the unknown. If it is not, 2-D spectra and NOE experiments are used to establish connectivity and proximity of the nuclei in the molecule. The information obtained is often sufficient to establish the skeleton of the molecule. Sometimes, especially when there are NMR "silent" nuclei (such as chlorine, bromine, oxygen, nitrogen, etc.) present, it may be necessary to obtain MS and/or IR data in order to complete the structure. While it is true that spectra can be obtained for most of the "silent" nuclei, it is usually much quicker and easier to use these other spectroscopic techniques to establish their presence.

In studies to determine the identity of unknown compounds, it is often helpful to compare the NMR chemical shifts with those of related known substances. For example, if one were trying to identify drug metabolites or impurities, one could compare the spectra with those of the drug itself. If the chemical shift assignments are not already known, one can use the above sequence of experiments to establish them. Then, by determining which resonances are present in both and which have changed or disappeared, one can determine which part of the drug has changed. This is of particular advantage in cases where one has a relatively large amount of the drug and the amount of unknown available is small, thus limiting the types of experiments which may be performed on the unknown.

3.3 Quantification

When quantification is required, sample preparation procedures are more demanding than those for routine samples (discussed in section 2.2). If absolute quantification is required, the amount of sample used must be accurately known and an accurately known amount of a suitable standard must be added to the sample before the analysis. If only relative quantities of substances present are required, there is no need for an internal standard. If an extraction procedure is necessary, one must be sure that the drug-related material is quantitatively removed.

An appropriate internal standard must be chosen to ensure accurate quantification. This standard must meet a number of strict criteria. It must be pure (or at least be of known purity) and be soluble and stable in the solvent used for analysis. It must not react with any compounds present (including the solvent). For convenience, it is usually easier to work with a solid standard, but if a liquid is chosen, it should be nonvolatile in order to avoid losses during the sample preparation and analysis. Lastly, the standard must have an appropriate NMR spectrum such that both the drug-related material and standard have well separated resonances whose integrals or peak heights may be used in the quantification. It is usually best to choose a standard which has either a suitable singlet or narrow multiplet to minimize the chances of overlap of signal resonances.

The choice of nuclei for NMR quantification is important. For relatively simple mixtures, ^{1}H-NMR, because of the shorter relaxation times and higher sensitivity when compared with ^{13}C-NMR, is often the method of choice. However, for more complex mixtures, ^{13}C-NMR may prove to be the better option. Although NMR is well suited for quantification of drugs, relatively few publications have appeared in this area. In part this is due to the fact that this method is considered to be so routine as to not warrant publishing these kinds of results. More recently, however, a number of publications have appeared (e.g. Lacroix et al.[10,11]) which show that NMR gives analytical results comparable to those obtained by other methods, such as HPLC (see sections 3.4 and 3.5). These papers point out the advantages of NMR over these alternative methods, indicating that NMR should be used more routinely for quantification purposes.

There are basically two approaches to obtaining the accurate quantification of drugs or impurities by NMR techniques. In the first, one must wait a relatively long time ($>5\,T_1$ for the longest relaxation time) to ensure equal responses for each of the nuclei being studied. The second method requires that spectra be acquired and processed under identical conditions and that a calibration curve, with appropriate amounts of the standard and analyte covering the range of interest, be obtained. The first method has the advantage of not requiring pure compounds (used to create the calibration curve in the second) because the relative response of the nuclei is known. Its main disadvantage is the amount of time required to acquire the spectra due to the long relaxation delay between pulses. Another disadvantage appears in the processing of the data. Because of differences in peak widths, areas from integrations must be used and care must be taken to avoid problems due to baseline distortions which may adversely affect the analyses. The second method has the advantage of shorter acquisition times for the spectra, but requires that appropriate standards be available and calibration curves be obtained. Another advantage of the second method is that because a calibration curve is being used, it is possible to use peak heights instead of areas (although areas may still be used). Using peak heights with calibration curves has allowed the simultaneous quantification of a relatively large number of impurities in the drug fenofibrate[10] at relatively low levels (<0.1%). This is considerably better than results reported for other drugs using integration, where the lower limits are reported to be about 1%.

3.4 Impurity Profiling

When a ^{1}H-NMR spectrum is taken for a drug, signals are obtained for the protons of the impurities as well as for the pharmaceutical itself. Careful examination of these ''extra'' peaks will often allow the identification of one or more impurities. Integration, or comparison of relative peak heights, will usually give a very good estimate of the percentage of each impurity present. As long as there are unique resonance peaks for each impurity, large numbers of impurities may be quantified from a single spectrum. This was amply illustrated by Lacroix et al.[10] for the drug fenofibrate. In this study, 12 known impurities, as well as a number of unknowns, were quantified at levels of <0.1% in drug raw materials from several manufacturers.

In cases where the impurity(ies) cannot be identified in the presence of the drug, it is often possible to separate it (them) from the drug using the extraction methods discussed in section 2.2. Once the drug resonances no longer mask the impurity peaks, it is much easier to perform the required NMR techniques to identify the material. Alternatively, if an HPLC method exists for separating the impurity, the eluted peak may be collected and analyzed by NMR. If the HPLC and NMR are interfaced, the impurity may be studied without having to go through the collection step.

3.5 Optical Purity

For a number of drugs only certain optical isomers are supposed to be present in the commercial product. Using the methods described in section 2.3.1.4, NMR methods for determining the optical purity for many of these drugs have been reported. Most of the earlier studies were done on lower-field-strength instruments, using CLSRs. The limit of quantification for these studies was typically 1–5%. The later studies, using higher-field-strength instruments and CSAs or CDAs provide much greater sensitivity, with detection limits in some cases of much lower than 0.1%. For example, Lacroix et al.[11] used (*R*)-(−)-2,2,2-trifluoro-1-(9-anthryl)ethanol with drug raw material for timolol maleate, enabling them to easily detect the optical impurity at levels of 0.1% at a relatively low field strength (400 MHz). These kinds of results make NMR a very attractive method for determining if pharmaceutical products meet the requirements set by pharmacopoeias of many countries.

3.6 Analysis of Body Fluids

In many cases analysts studying drug metabolism will analyze samples of body fluids in attempts to measure levels of the drugs and/or their metabolites. The concentration of drugs in the blood varies over a wide range – from tens of milligrams per milliliter for ethanol to nanograms per milliliter for most drugs and their metabolites. However, in some extreme cases of overdoses, drug levels of 100 mg mL^{-1} are observed. The lower ranges of these drug and metabolite levels are beyond the sensitivity of NMR, but the higher end values are not. Detection limits for ^{1}H-NMR are below the microgram per milliliter range on newer high-field spectrometers using the hyphenated techniques discussed in section 2.3.3. Without these methods, it has proved difficult to obtain these lower limits due to the background signals from endogenous compounds in the blood and urine samples. The fluorine signals of fluorinated steroids are detectable at submicrogram-per-milliliter levels because there is no background. For nonfluorinated drugs, analysts who do not have access to the hyphenated techniques have the option of performing sample work-up procedures such as solid-phase extraction (SPE) or preparative HPLC clean-up before analysis. This also allows the analyst the opportunity to concentrate the drug or metabolite from several milliliters down to the volume (<0.5 mL) required for NMR analysis, thus bringing many more samples into the range of levels detectable by NMR.

4 COMPARISON WITH OTHER TECHNIQUES

4.1 Advantages

NMR will detect all organic compounds present in the sample. Other techniques, such as HPLC or MS, will often not detect impurities. For example, these techniques will not observe any compounds that either elute in the solvent front or do not come off the column at all under the conditions used. Careful examination of the impurity peaks in the NMR spectrum will often give significant clues as to their identity and allow quantitative estimates as to the amounts present. HPLC provides little information about the identity of unknown impurities and is not reliable for quantifying them since it does not give equimolar responses for different materials. In fact, compounds with low or no ultraviolet (UV) absorbance may not be detected, even when present at significant levels.

NMR is a nondestructive technique. As long as the sample material is stable in the solvent chosen for the study, the entire sample may be recovered and then used for other studies. Thus, since sample is used up in other techniques such as HPLC and MS, the NMR studies should be performed first whenever there is a limited amount of sample.

NMR spectroscopy provides the most powerful structure elucidation technique available to the pharmaceutical analyst for unknown materials. One area where

NMR outshines other spectroscopic techniques is in the determination of isomers. Its true value comes from its ability to establish the proximity of nuclei in a molecule using coupling constants (through one, two or three bonds) or with NOE measurements (for nuclei within approximately 5 Å). The NMR chemical shifts and coupling constants easily allow the discrimination of geometric or positional isomers, even when there are no authentic materials for comparison purposes. It would be much more difficult, if not impossible, to get this same information using other spectroscopic techniques such as IR spectroscopy or MS (without reference standards).

The most common method for an analyst to determine the optical purity of drugs has been by either specific optical rotation or HPLC, using a chiral column or a chiral derivatization procedure. However, NMR has a number of advantages over optical rotation and HPLC in studies aimed at determining the enantiomeric composition of any organic compounds, including pharmaceuticals. This is especially true if methods are required that are capable of analyzing for optical antipodes at levels of less than a few percent. This may be important if one wishes to confirm that a pharmaceutical meets the requirements of a country's pharmacopoeia. Using modem spectrometers and the techniques described previously, it is usually possible to develop methods that can quantify the drug's enantiomer at or below levels of 0.1%. Some of the advantages of NMR are given here.

1. NMR results are accurate and precise, even at low percentages. Optical rotation cannot be used to determine low levels of impurity, especially in drugs with a low specific optical rotation.
2. The time required for analysis is usually less for NMR than for HPLC. This is especially true if there is no existing method, since it is usually quicker to develop an NMR method than an HPLC method. If a method already exists, it is still much faster to get the results by NMR. For NMR, the procedure would normally involve simply dissolving the drug in an appropriate solvent, adding the calculated amount of CSA, acquiring the proton spectrum and measuring the proper peaks in the spectrum. This may normally be done within half an hour. For HPLC, one must set up the liquid chromatography system (put in the proper column, prepare the mobile phase, etc.), check the performance of the column by running a standard and then run the sample. All this requires much more than half an hour.
3. NMR methods are more ''robust'' than HPLC methods. Once a method has been developed for NMR, it should work on any spectrometer of equal field strength (or higher field strength for CSAs and CDAs). Because of the variability of columns, HPLC methods have to be checked and sometimes modified when changing columns.
4. NMR provides ''bonus'' information. Since NMR detects all signals for the nucleus under study, it will detect not only the drug (and its enantiomer), but also any impurities present.

4.2 Disadvantages

One of the main criticisms of NMR as an analytical technique is its relative insensitivity when compared with other methods such as MS or HPLC. While these other instruments may require only nanograms or less of the sample, NMR usually needs micrograms or more. While this limitation is of little concern for drug raw materials or dosage forms, it does limit its use in analyses of body fluids for drugs and their metabolites.

Another common complaint about NMR is the cost of the instrumentation. While it is true that highest-field-strength spectrometers are very expensive (millions of dollars each), the models that have been around for several years (currently field strengths of 200–600 MHz) tend to be comparable in cost to mass spectrometers. In general, NMR spectrometers have proven to be very reliable instruments, with relatively low repair and maintenance costs. Thus, if one considers the cost per year, NMR spectrometers may not be as expensive as is generally thought.

Another concern often brought up when discussing the use of NMR is the need for highly specialized expertise in the area. While it is true that considerable expertise is required in order to take advantage of the full capabilities of modern spectrometers, very little knowledge is needed to use NMR results for more routine purposes. Thus, for positive identification of known compounds, simple comparisons of spectra with those of authentic materials is all that is required. This is very similar to the use of IR spectroscopy ''fingerprinting'' to identify compounds. It is also possible to automate the identification by use of library searches of NMR databases, comparable to those for IR spectroscopy or MS.

5 CONCLUSIONS

NMR has played a key role in pharmaceutical analysis in the past and promises to play an even more crucial role in the future. The ever increasing sensitivity and capabilities of the most modern spectrometers provide pharmaceutical analysts and researchers with more powerful tools to tackle their analytical problems. There is no doubt of the vital role to be played by NMR in studies

of the newer biotech pharmaceuticals in the areas of structure and drug interactions. In studies of traditional synthetic drugs, NMR, especially in conjunction with other analytical techniques such as MS and HPLC, will continue to have a critical function.

ABBREVIATIONS AND ACRONYMS

CDA	Chiral Derivatizing Agent
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CLSR	Chiral Lanthanide Shift Reagent
COSY	Correlation Spectroscopy
CSA	Chiral Solvating Agent
CW	Continuous Wave
CZE	Capillary Zone Electrophoresis
DEPT	Distortionless Enhancement by Polarization Transfer
DSS	2,2-Dimethyl-2-silapentane-5-sulfonate
FT	Fourier Transform
FTNMR	Fourier Transform Nuclear Magnetic Resonance
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
IR	Infrared
LSR	Lanthanide Shift Reagent
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
RCT	Relayed Coherence Transfer
RF	Radiofrequency
SPE	Solid-phase Extraction
TMS	Tetramethylsilane
UV	Ultraviolet
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
4-D	Four-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. R.H. Bible, *Interpretation of NMR Spectra*, Plenum Press, New York, 1965.
2. R.J. Abraham, *Analysis of High Resolution NMR Spectra*, Elsevier, New York, 1971.
3. J.W. Emsley, J. Feeney, L.H. Sutcliffe, *High Resolution Nuclear Magnetic Resonance Spectroscopy*, Pergamon Press, New York, Vol. 1 & 2, 1965.
4. H. Günther, *NMR Spectroscopy, Basic Principles, Concepts and Applications in Chemistry*, 2nd edition, John Wiley & Sons, New York, 1994.
5. C.J. Pouchert, J. Behnke, *The Aldrich Library of* ^{13}C *and* ^{1}H *FTNMR Spectra*, 1st edition, Aldrich Chemical Company Inc., Milwaukee, Vol. 1–3, 1993.
6. J.B. Stothers, *Carbon-13 NMR Spectroscopy*, Academic Press, New York, 1972.
7. F.W. Wehrli, T. Wirthlin, *Interpretation of Carbon-13 NMR Spectra*, Heyden, New York, 1976.
8. G.C. Levy, G.L. Nelson, *Carbon-13 Nuclear Magnetic Resonance for Organic Chemists*, Wiley-Interscience, New York, 1972.
9. P.A. Keifer, 'NMR Tools for Biotechnology', *Curr. Opin. Biotechnol.*, **10**, 34–41 (1999).
10. P.M. Lacroix, B.A. Dawson, R.W. Sears, D.B. Black, T.D. Cyr, J.-C. Ethier, 'Fenofibrate Raw Materials: HPLC Methods for Assay and Purity and an NMR Method for Purity', *J. Pharm. Biomed. Anal.*, **18**, 383–402 (1998).
11. P.M. Lacroix, B.A. Dawson, R.W. Sears, D.B. Black, 'HPLC and NMR Methods for the Quantitation of the (*R*)-Enantiomer in (−)-(*S*)-Timolol Maleate Drug Raw Materials', *Chirality*, **6**, 484–491 (1994).

Planar Chromatography in Pharmaceutical Analysis

A.P. Argekar
The Institute of Science, Mumbai, India

1 Introduction 1
2 History 2
3 Sample Preparation 2
4 Pharmaceutical Applications 2
4.1 Pharmaceutical Products 4
4.2 Purity Profiles 4
4.3 Stability Studies 5
4.4 Process Monitoring 6
4.5 Herbal Drugs 6
5 Pharmaceutical Regulations and Methods 8
6 Validation and Quality Assurance 10
6.1 Analytical Method Validation 10
6.2 Quality Assurance of Experimental Data and Documentation 11
7 Method Development 12
7.1 Selection of the Stationary Phase 12
7.2 Selection of the Vapor Phase 13
7.3 Selection of Suitable Solvents for the Mobile Phase 13
7.4 Optimization of the Mobile Phase 13
7.5 Selection of the Development Mode 14
8 Comparison with Other Methods 14
Acknowledgment 15
CAS Registry Numbers 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Planar chromatography is a liquid chromatography in which the stationary phase is arranged in the form of a planar or flat bed and the mobile phase moves by capillary action. Thin-layer chromatography (TLC) is the most widely used planar chromatography. Modern TLC is an instrumental version of conventional TLC known as high-performance thin-layer chromatography (HPTLC). Rotation planar chromatography (RPL), high-speed thin-layer chromatography (HSTLC), and over-pressured liquid chromatography (OPLC) are some other forms of planar chromatographic methods.

The application of HPTLC techniques for pharmaceutical analysis, such as the determination of active pharmaceutical ingredients (API) and the content uniformity test (CUT) of dosages, purity and stability tests of drugs and drug products, process monitoring of formulations, and herbal drug analysis, is described in this article. The method development of the high-performance liquid chromatography (HPLC) technique, various pharmaceutical regulations and methods, validation and quality assurance (QA) of HPTLC experimental data and documentation are also described. The comparison of HPTLC with other chromatographic techniques is also discussed.

1 INTRODUCTION

Planar chromatography is a liquid chromatographic technique in which the stationary phase is arranged in the form of a planar or flat bed and the mobile phase moves by capillary action. The most widely used planar chromatography is TLC. Modern TLC is a fully instrumented version of conventional TLC, also known as HPTLC. In HPTLC perfectly uniform surfaces, 150–200 μm in thickness, are prepared from particles with a narrow size distribution, typically with diameters between 5 μm and 15 μm.

HPTLC is an off-line technique where every stage of the analysis protocol can be visualized. It is capable of providing faster separations, better separation efficiency, reduced zone diffusion, higher sensitivity, and more accurate quantification than other techniques for a wide variety of applications. The optimized instrumentation coupled with high levels of automation offers precise control over sample application, chromatographic development, derivatization, detection, and determination. The special feature of HPTLC is its ability to simultaneously spot, develop, and automatically scan a number of separations on a single plate. Typically, 18–36 samples can be run on a single plate with development times of 3–20 min over migration distances of about 2–7 cm using 2–5 μL of loading sample. HPTLC has a higher detection limit than HPLC, but this can be overcome by sample postchromatographic derivatization or by bioautography. The greatly increased development time on HPTLC chromatoplates has made it necessary to employ forced flow of the mobile phase instead of capillary flow. This feature has been exploited in RPL. In HSTLC electroosmosis is employed to force the eluent. OPLC is the true planar version of HPLC, where pressurized flow is developed using a pump system for admission of the mobile phase.

Pharmaceutical analysis is traditionally defined as the branch of analytical chemistry dealing with drugs, both

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

as bulk drug substances and pharmaceutical products (formulations). This article is concerned mainly with planar chromatography together with capillary flow, namely HPTLC for pharmaceutical analysis.

2 HISTORY

The history of TLC has been reviewed by several authors.[1–7] Ismailov and Shraiber separated certain medicinal compounds on an alumina spread on a glass plate, in 1938. Meinhard and Holl in 1949 used a binder to adhere alumina to microscope slides to separate certain inorganic ions. In the early 1950s, Kirchner and colleagues used sorbents held on glass plates with the aid of a binder, and the plates were developed with the conventional ascending procedure used in paper chromatography. Stahl introduced the term "TLC" in the late 1950s. Quantitative TLC was introduced in 1954 by Kirchner et al.[8] Densitometry in TLC was initially reported in the mid-1960s by Dallas et al.,[9] Genest,[10] and Thomas et al.[11] HPTLC plates having fine particle layers were produced commercially in the mid-1970s, and in the late 1970s forced-flow planar chromatography was introduced.

TLC will continue to grow as a highly selective, sensitive, quantitative, rapid, and automated technique for the analysis of all types of samples and analytes, and for preparative separations.

3 SAMPLE PREPARATION

Sample preparation for TLC is not as demanding as for other chromatographic techniques as the plates are disposable products. However, several steps for sample pretreatment may be necessary, which include mechanical crushing, extraction steps, filtration and sometimes enrichment of interesting desired components, or removal of undesirable impurities. Proper sample preparation is an important prerequisite for the success of TLC separation. The sample preparation procedure is to dissolve the drug with complete recovery of intact compounds of interest and minimum of matrix with a suitable concentration of analytes for direct application on the TLC or HPTLC plate. Besides maximizing the yield of analytes in the selected solvent, the stability of analytes during extraction and analysis must be ensured. Therefore, the choice of a suitable solvent is very important. For normal-phase chromatography using silica gel-precoated plates, the solvent for dissolving the sample should be nonpolar and volatile as far as possible. This is because polar solvents are likely to induce circular chromatography when the sample is employed incrementally, leading to spreading of the spot and thus loss of separation efficiency. For reversed-phase chromatography polar solvents are usually used for dissolving the sample; however, such polar solvents must wet the layer so that the sample penetrates the layer uniformly. It is preferable to use the minimum possible amount of solvent, the quantity employed being limited to ensure complete extraction of analytes but the minimum of extraneous components. Sample and reference substances should be dissolved in the same solvent to ensure comparable distribution at the starting zone. Most sample preparation procedures require concentration or evaporation of the sample extracts. It is important that evaporations are carried out without loss or degradation of the analyte. Common methods for concentration are rotary evaporation and nitrogen blow down. Common practice is to evaporate solutions just to dryness and then to dissolve the residue in an exact volume of the same or different solvent from which a known aliquot or the total sample is applied to chromatoplates.

4 PHARMACEUTICAL APPLICATIONS

There is a comprehensive literature dealing with TLC systems for pharmaceutical analysis.[12–26] Manufacturers of HPTLC instrumentation provide instrument validation and support for all users performing analysis in the fields of herbal, bulk drugs, and pharmaceutical formulations. Pharmaceutical applications of HPTLC analysis are summarized in Table 1. The only limitation of HPTLC

Table 1 Uses of HPTLC in the pharmaceutical industry

Industry sector	HPTLC analytical task
Manufacturing	Process monitoring, fermentation broth analysis, cleaning, and residue analysis
QC	Raw material assay, multicomponent formulations, uniformity of content, individual container analysis, impurity profiles
Formulation (research and development)	Formulation analysis, stability studies, sustained release, bioequivalance studies
Synthesis (research and development)	Process optimization, preparative chromatography
Crude drug	Fingerprinting, active ingredient quantitation, label claim check, preparative analysis, adulteration check

QC, quality control.

Table 2 HPTLC of some drug formulations

Drug(s)	Stationary phase	Mobile phase	UV detector wavelength (nm)	Ref.
Pseudoephidrine sulfate, azatidine maleate and dextromethophan HBr in tablets, simultaneously	Silica gel 60 F^{a}_{254}	CH_2Cl_2–acetone–methanol–triethylamine (70:40:5:2)	257	27
Sulfamethoxazole and trimethoprim in tablets, simultaneously	Silica gel 60 F_{254}	Chloroform–methanol (9:1)	254	28
Heloperidol and benzhexol HCl, simultaneously	Silica gel 60 F_{254}	Toluene–methanol–triethylamine (45:5:1)	258	29
Pseudoephidrine HCl (I) and guaiphenesin (II) in decongestant tablet	Silica gel 60 F_{254}	Ethyl acetate–methanol–concentrated ammonia (17:2:1)	210 for (I) 276 for (II)	30
Ibuprofen and paracetamol in tablets, simultaneously	Silica gel 60 F_{254}	Hexane–ethyl acetate–acetic acid (18:5:2)	265	31
Azatadine maleate, pseudoephidrine HCl, and paracetamol (phenylhydramine as internal standard)	Silica gel 60 F_{254}	Dichloromethane–methanol–triethylamine (74:5:2)	257	32
Naproxen (I) and ibuprofen (II) in tablet	Silica gel 60 F_{254}	Ethyl acetate–acetic acid (19:1)	254 for (I) 268 for (II)	33
Paracetamol, ibuprofen, and chloroxazone, simultaneously	Silica gel 60 F_{254}	Chloroform–toluene–dioxan (7:2:2)	210	34
Ketoprofen (I), fusidic acid (II), molsidomine (III), dipyrone (IV), pitofenone HCl (V), and fenpiverinium HBr (VI) in tablet and injections	Silica gel 60 F_{254}	Chloroform–ethyl acetate–methanol (6:1:4) for (I), chloroform–cyclohexane–acetic acid–methanol (40:6:4:1) for (II), chloroform–acetone (4:1) for (III), tetrahydrofuran–methanol–acetic acid (5:3:2) for (IV) and (V)	290	35
Tinidazole and furazolidone in suspension	Silica gel	Chloroform–methanol–ammonia (90:10:1)	335	36
Theophylline and etofylline in tablets	Silica gel	Chloroform–methanol (9:1)	278	37
Nitrinedipine (I) and Nimodipine (II)	Silica gel	Toluene–ethyl acetate (1:1)	236 for (I) 240 for (II)	38
Paracetamol (I), ascorbic acid (II), caffeine (III), and phenylephrine (IV)	Silica gel	Dichloromethane–ethyl acetate–ethanol–formic acid (7:4:8:1) and dichloromethane–ethyl acetate–ethanol (5:5:1)	274 for I and IV 264 for II 254 for I	39
Amiloride HCl and frusemide	Silica gel	Ethyl acetate–methanol–18% ammonia (75:10:8)	275	40
Aspirin and dipyridamole	Silica gel	Ethyl acetate–ethanol–13.5 M ammonia (15:3:3)	290	41
Reserpine and other alkaloids	Silica gel	Ethyl acetate–cyclohexane–DEA (210:90:1)	366	42
Norfloxacin and tinidazole	Silica gel	*n*-Butanol–ethanol–12.5% ammonia (20:5:11)	293	43
Metronidazole and nalidixic acid	Silica gel	Ethyl acetate–chloroform–methanol–ammonia (5:5:3:1)	320	44
Rifampicin (I), isoniazid (II), and pyrazinamide (III)	Silica gel	Ethyl acetate–ammonia–ethyl alcohol–cyclohexane (20:9:4.5:5)	440 for (I) 275 for (II) and (III)	45
Lansoprazole	Silica gel	Ethyl acetate–methanol–13.5 M ammonia (10:3:4)	288	46
Isoniazid and pyridoxine HCl, simultaneously	Silica gel	Acetone–carbon tetrachloride–6.5 M ammonia (21:7:2)	275	47
Amikecin	Silica gel	Methanol–25% ammonia–chloroform (10:10:2)	498	48
Ciprofloxacin, enrofloxacin, lomefloxacin, norfloxacin, ofloxacin, and pefloxacin	Silica gel	Butanol–ethanol–ammonia (20:5:11)	280	49

(*continued overleaf*)

Table 2 (*continued*)

Drug(s)	Stationary phase	Mobile phase	UV detector wavelength (nm)	Ref.
Salbutamol sulfate	Silica gel	Methanol–chloroform–triethylamine (5.5 : 4.5 : 0.05)	276	50
Atenol and nitrendipine, simultaneously	Silica gel	Chloroform–methanol–toluene–25% ammonia (2 : 2.5 : 5.5 : 0.1)	233	51
Pyridoxin HCl and doxylamine succinate in tablets, simultaneously	Silica gel	Acetone–chloroform–methanol–25% ammonia (7 : 1.5 : 0.3 : 1.2)	269	52

[a] Contains a fluorescent substance.
DEA, diethylamine; UV, ultraviolet.

so far as pharmaceutical analysis is concerned is that it cannot be employed for dissolution testing studies.

4.1 Pharmaceutical Products

Drugs are defined as substances or products that are intended for modification or exploration of physiological states for the recipient. The drugs are prescribed in a variety of forms such as tablets, capsules, injections, ointments, intravenous fluids, syrups, etc. termed as pharmaceutical products, formulations, or dosage forms. As HPTLC requires minimal sample cleanup, interference from the excipients in tablets, the high quantity of fatty matter in creams, and the sugar base of lozenge formulations pose no problems in HPTLC analysis. Therefore, the active drug components either in single or mixed dosage forms are estimated economically and speedily. Table 2, summarizes HPTLC applications for the determination of the API from formulations.

As specified in *United States Pharmacopeia* XXIII, dosage forms containing 50 mg or less of active ingredient per unit must comply with a CUT. The criteria specified is that out of 10 units individually tested the contents of each unit must be within 85–115% of the labeled claim with a relative standard deviation (RSD) not greater than 6.0%. However, if one unit is outside the 85–115% range of the labeled claim and no unit is outside 75–125%, or of RSD not greater than 6.0% or both, 20 additional units must be tested. Not more than one unit out of 30 so tested is to be outside the range 85–115% and no unit is to be outside 75–125% with RSD not above 7.8%. These tests are laborious and time-consuming. However, HPTLC has been found to be the most suitable and economically viable method for testing the dosage form for compliance in respect of content uniformity. Using the built-in CUT feature in CATS software, three complete CUTs can be performed in 2 h. (CATS software is a Windows-based software used in quantitative TLC and is supplied by CAMAG, Switzerland.)

4.2 Purity Profiles

The purity of a drug and the determination of the nature and quantity of any impurities present are key questions for the modern pharmaceutical analyst. The nature and the quantity of impurities in a drug product depend upon the route of synthesis used to produce the drug substance (bulk drugs), purification procedures, manufacturing of pharmaceutical products or formulation, and the conditions under which they have been stored. Purity tests of bulk drugs are generally focused on related substances such as homologs, analogs, by-products from the synthesis or degradation, and the enantiomeric purity. HPTLC has revolutionized the ability to determine substance purity and is currently the most important check, giving essentially the fingerprint of a synthesis.

Low-level substances with relatively high retention times are poorly resolved and detected owing to peak broadening in HPLC. However, in reversed-phase high-performance thin-layer chromatography (RPHPTLC) substances with low R_f (retention factor) values form strong and sharp zones, and hence have better detection sensitivity. Those substances with low retention times may appear as unresolved in HPLC, but in RPHPTLC they will usually appear between R_f values of 0.5 and 0.9 and are easily detected. In RPHPTLC, the impurities can be detected either at the starting line or near the solvent front provided the plates have been properly alleviated by prewashing prior to development. Large quantities of slowly eluting substances would completely spoil the HPLC column after collection and may show memory effects, whereas slowly migrating substances in RPHPTLC will be retained near the origin.

When studying the purity profile of drugs using HPTLC, the following precautions should be considered:

- After the TLC plates have been developed, they should not be dried at an elevated temperature. This avoids any decomposition, oxidation, and hydrolysis during drying.

- The solvents to be used for sample preparation or as components of mobile phase should be checked by running a blank TLC plate in the solvent. The plate should not show any spot visible under UV light or with the reagent likely to be used for the detection of impurities. This is absolutely essential because any impurity present in the solvent can give erroneous results, in that the additional spots can be mistaken to have been contributed by the substance under examination.

The application of HPTLC to the purity profile study of some drugs is illustrated by the densitograms in Figures 1–3.

Commercially available chiral HPTLC plates have made it easy to purify and quantify enantiomeric drugs. Optical purity control of L-dopa and D-pencillamine (Figures 4 and 5) are two excellent examples showing the application of HPTLC for enantiomeric drug separations.

4.3 Stability Studies

The purpose of stability testing is to provide evidence on how the quality of a drug substance or pharmaceutical product varies with time under the influence of a variety of environmental factors such as temperature, humidity, and light. Stability testing permits the establishment of recommended storage conditions, retest periods, and shelf lives.

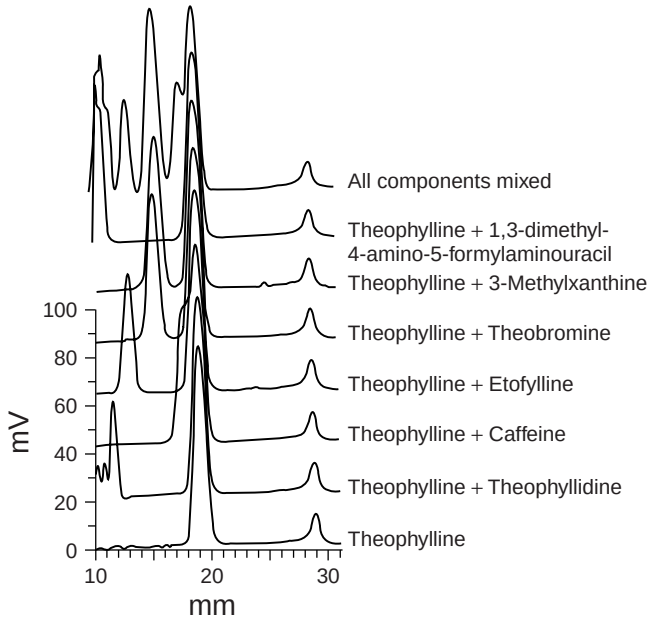

Figure 1 Densitogram of the purity profile of theophylline scanned at 274 nm using Camag scanner II with CATs software; silica gel 60 F_{254} HPTLC plates; toluene–2-propanol–acetic acid (16 : 2 : 1 v/v) as mobile phase; unsaturated twin-trough chamber. (Reprinted by permission of publisher Research Institute for Medicinal Plants, Hungary from *J. Planar Chromatogr.*, **8**, 271 (1995).)

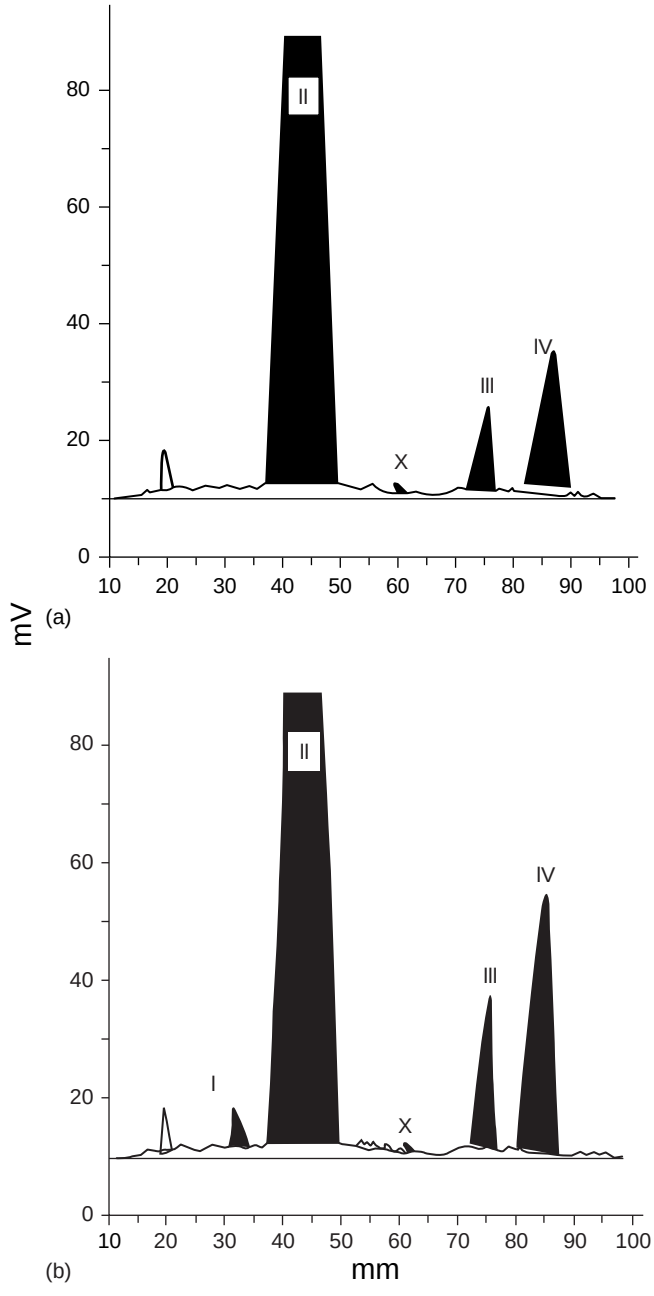

Figure 2 Densitogram of the purity profile of estradiol benzoate scanned at 229 nm with TLC Camag scanner II with CATs software, silica gel 60 F_{254} HPTLC plates; hexane–acetone–2-propanol (90 : 5 : 5 v/v) as mobile phase; unsaturated twin-trough chamber; I estradiol; II estradiol benzoate; III estradiol-3-benzoate-17-acetate; IV estradiol dibenzoate. (a) A typical batch (100 µg of estradiol). (b) batch spiked with impurities. X – is the impurity of unknown identity. (Reprinted by permission of publisher Research Institute for Medicinal Plants, Hungary from *J. Planar Chromatogr.*, **8**, 350 (1995).)

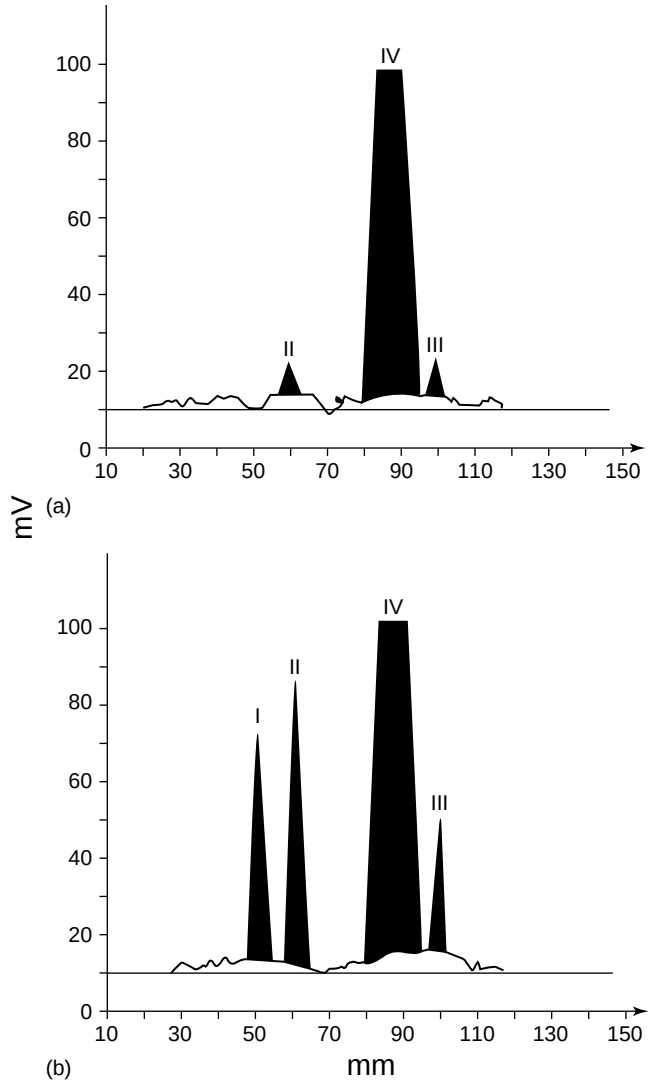

Figure 3 Densitogram of the purity profile of danazol scanned at 252 nm with TLC Camag scanner II with CATs software, silica gel 60 F_{254} HPTLC plates; chloroform–ethyl acetate (88 : 12 v/v) containing 0.02% ethylamine as mobile phase; unsaturated twin-trough chamber; I, II, III are impurities; IV is danazol. (a) A typical batch of danazol (100 μg). (b) batch spiked with impurities. (Reprinted by permission of publisher Research Institute for Medicinal Plants, Hungary from *J. Planar Chromatogr.*, **6**, 199 (1993).)

Information on the stability of a drug substance or dosage from under defined storage conditions is an integral part of a systematic approach to stability evaluation. Stress testing helps to determine the intrinsic stability characteristics of a molecule by establishing degradation pathways to identify the probable degradation products and to evaluate the stability by indicating the power of the analytical procedure used.

HPTLC has been employed as a stability-indicating method for stability studies of several drugs and formulations. Three such examples are illustrated in Figures 6–8.

4.4 Process Monitoring

Profitability is directly linked to production efficiency, which in turn can be established by numerous research and development experiments. Each such experiment must be monitored closely. Based on the results of the impurity profile of the drug, the process is modified. Thus an assurance of product quality and consistency is very much in the hands of analysts. In process analysis the assay is performed either by the production department or the QC laboratory. Production departments are therefore equipped to conduct those types of in-process controls necessary to keep the process running, whereas controls that might require special equipment are conducted by the QC laboratory. The methods should be very easy to perform, fast, accurate and reproducible. They should help in the processing of drug substances during intermediate stages. Chances of contamination from previous processes is to be monitored and indicated.

HPLC and HPTLC are the two most widely used techniques in the pharmaceutical industry for process control. Process optimization needs hundreds of quick analyses, which are well executed by an HPTLC system at a low cost. Moreover one HPTLC system can support several laboratories simultaneously.

The drug norfloxacin is produced by condensing 3-chloro-4-fluoro aniline (CFA) and diethyl ethoxy methylene malonate first to yield the acrylate which, on further cyclization, yields the *N*-ester. On ethylation this ester gives the ethyl-*N*-ester which, on hydrolysis, gives acid and norfloxacin. This production of norfloxacin can be monitored by HPTLC technique, as all the compounds have absorption in the UV region at 288 nm. The densitogram of a typical production batch of norfloxacin is shown in Figure 9. The production process optimization and monitoring of oxfendazole can also be performed using HPTLC (Figure 10).

4.5 Herbal Drugs

Plant substances have been used in medicines since ancient times. The composition of the drug varies widely depending on the species, geographical location, origin, climate, growth period, part of the plant, storage, and preparation procedure. Herbal formulations used in the preparation of therapeutic medicines derive their potency from constructive and symbiotic effects due to the presence of two or more crude drugs. This makes the analysis of herbal formulations more complex.

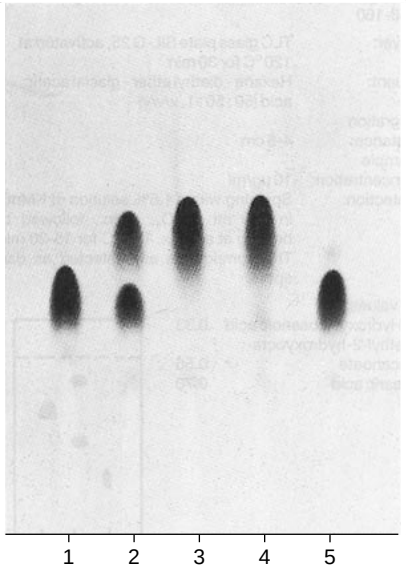

Figure 4 Chromatoplate showing the optical purity of L-dopa. Layer, Chiralplate™; eluent, methanol–water–acetonitrile (50:50:30 v/v/v); developing time, 45–60 min; detection, 0.1% ninhydrin reagent. 1 = L-dopa; 2 = D,L-dopa; 3 = D-dopa; 4 = 3% L-dopa in D-dopa; 5 = 3% D-dopa in L-dopa. (Reprinted by permission of K. Gunther, from *Z. Anal. Chem.*, **322**, 513–514 (1985).)

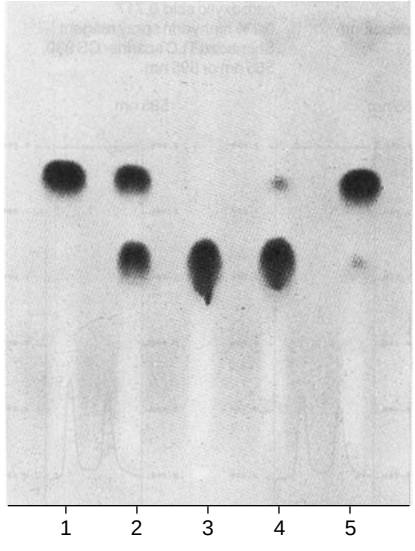

Figure 5 Chromatoplate of optical purity of D-pencillamine. Layer, Chiralplate™; eluent, methanol–water–acetonitrile (50:50:200 v/v/v); developing time, 30 min; detection, 0.1% ninhydrin spray reagent. 1 = L-5,5-dimethyl-4-thiozolidine carboxylic acid (L-3); 2 = D,L-5,5-dimethyl-4-thiozolidine carboxylic acid (D,L-3); 3 = D-5,5-dimethyl-4-thiozolidine carboxylic acid (D-3); 4 = 3% L-3 in D-3; 5 = 3% D-3 in L-3. (Reprinted by permission of K. Gunther, from *Arch. Pharm. (Weinheim)*, **319**, 461–465 (1986).)

HPTLC has been found to be the most accurate and facile technique for the analysis of herbs and their extracts over other techniques for the following reasons:

- Very little sample clean-up is needed as the layer is disposable.
- Preparative work can be very easily done.
- Fingerprinting of a large number of components of the sample matrix is possible.
- Active ingredients can be quantified.
- Short run times.
- Large numbers of analyses are possible, for reliable statistics.
- Many components in herbal medicines are fluorescent.

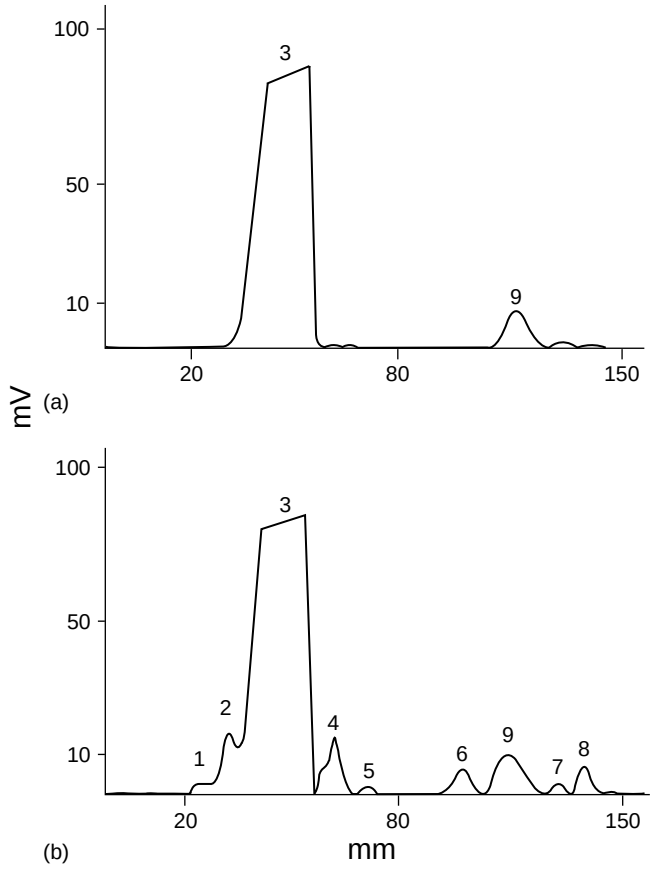

Figure 6 Densitogram of the stability profile of oxytetracycline eye drops scanned at 254 nm. Layer, silica gel TLC plates impregnated with aqueous solution of disodium ethylenediaminetetraacetic acid and activated at 130 °C; eluent, chloroform–methanol–water (65 : 25 : 5 v/v/v); saturated TLC tank. (a) On the day of the preparation, (b) On the 10th day of storage. Peaks: 1 = α-APC-OTC (α-apoxytetracycline); 2 = EOTC (4-ethoxytetracycline); 3 = OTC (oxytetracycline hydrochloride); 5 = β-APC-OTC (β-apoxytetracycline); 4,6,7 and 8 = unidentified; 9 = solvent front caused by demixing. (Reproduced by permission of J. Torole from *Acta. Pharmaceutica Hungarica*, **64**, 65 (1994).)

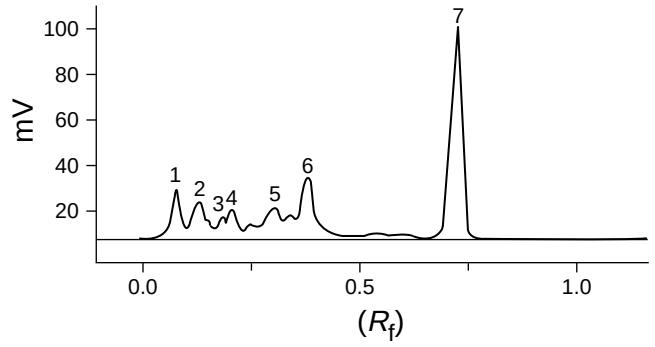

Figure 7 Densitogram of stability profile of lansoprazole scanned at 288 nm with TLC Camag scanner II with CATs software. Layer, silica gel 60 F_{254} HPTLC plates (Merk); mobile phase, ethyl acetate–methanol–13.5 M ammonia (10 : 3 : 4 v/v/v); unsaturated twin-trough chamber. Peaks: 1–6 are degraded substances; 7 = lansoprazole. (Reprinted by permission of publisher, Research Institute for Medicinal Plants, Hungary, from *J. Planar Chromatogr.*, **9**, 298 (1996).)

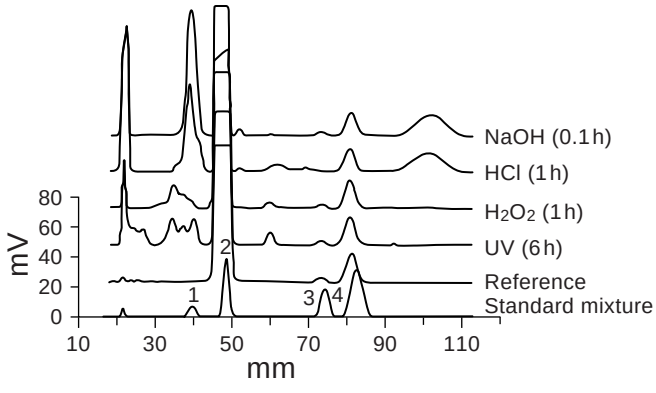

Figure 8 Densitogram of stability profile of estradiol benzoate in solution; chromatographic conditions as described in Figure 2. Peaks: 1 = estradiol; 2 = estradiol benzoate; 3 = estradiol-3-benzoate-17-acetate; 4 = estradiol dibenzoate. (Reprinted by permission of publisher, Research Institute for Medicinal Plants, Hungary, from *J. Planar Chromatogr.*, **8**, 351 (1995).)

A large number of papers on the application of planar chromatography to the characterization of herbal medicines have been published in Chinese journals. Table 3 shows HPTLC applications for herbal formulations.

5 PHARMACEUTICAL REGULATIONS AND METHODS

There are several regulations that have to be followed in the development of pharmaceuticals as well as in their production. Regulatory approval is required prior to the investigational new drug (IND) application, before initial human testing, and also the new drug application (NDA) which summarizes the data obtained from all the studies needed for marketing approval of the drug as a medicine.

An important part of the development process is safety evaluation, primarily the toxicology tests, which run from 6 to 24 months in different species. By this time bioanalytical studies have been performed, as well as control of the formulations used in the tests. After approval for marketing the authorities exercise control of products on the market and require past production stability data. Public interest in the quality of drugs is also reflected in the compilation of substance monographs in compendia, known as pharmacopoeias, which contain

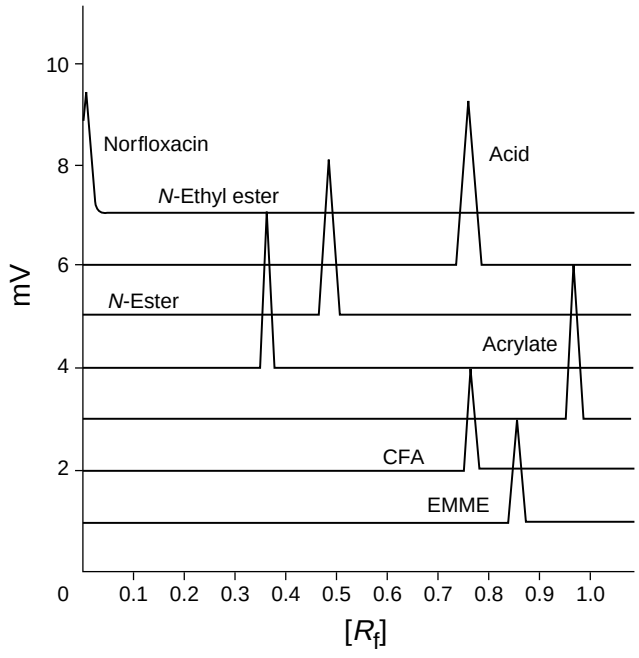

Figure 9 Densitogram of process control for the synthesis of norfloxacin. Layers, 60 F_{254} silica gel on aluminum sheet HPTLC plates (Merk); mobile phase, toluene–ethyl acetate–glacial acetic acid (10 : 3 : 4 v/v/v); twin-trough chamber; Camag TLC scanner with CATs software; detector wavelength 254 nm. EMME, ethoxy methyl malonate ester. (Reproduced by permission of Dr. S.U. Kapadia from PhD Thesis, University of Mumbai, 1996.)

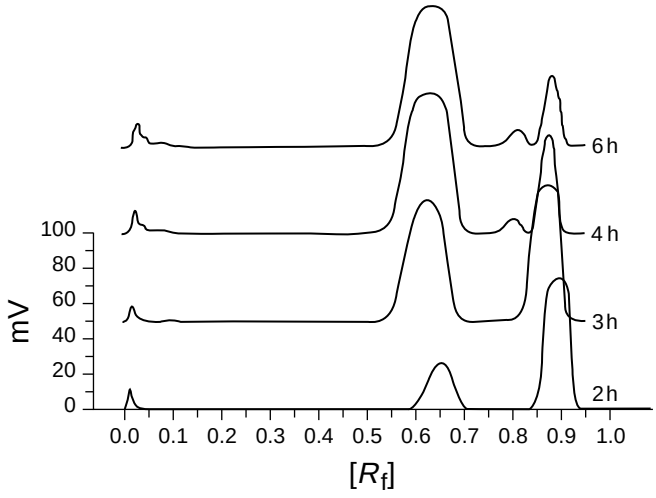

Figure 10 Densitogram showing the process optimization/monitoring of the synthesis of oxfendazole. Precoated HPTLC (Merk) plates; twin-trough chamber (Cmag); mobile phase, *n*-hexane–ethyl acetate–acetic acid (50 : 45 : 5 v/v/v); detector wavelength 254 nm. (Reproduced by permission of Dr. P.D. Sethi from *High-performance Thin-layer Chromatography. Quantitative Analysis of Pharmaceutical Formulations*, CBS Publishers and Distributors, New Delhi, 1996.)

general analytical methods and also requirements on the formulation of the substances.

Good Manufacturing Practice (GMP) regulations have been a cornerstone of pharmaceutical production and control since the 1970s. These regulations state clearly what has to be done to safeguard quality from the beginning of the production process to the end, namely documentation, staff qualifications, standard of facilities, technical standards, handling of material, labeling, etc. Control guidelines are mentioned but not in the same detail as in current good laboratory practice (GLP). These rules state that all documentation of analytical methods should fulfil certain performance criteria, that

Table 3 HPTLC of some herbal formulations

Mobile phase	λ (nm)	Compounds	Remarks	Ref.
Hexane–ethyl acetate (2 : 1)	268	Phyllanthin and hypophyllanthin	In phyllanthus amurus and polyherbal preparations	53
Hexane–ethyl acetate (1 : 1)	260 and 409	Phyllanthus	Fingerprinting evaluation for shelf preparation	54
Hexane–ethyl acetate (2 : 1)	254	Phyllanthus	For rational search for antihepatotoxic agent	55
Chloroform–toluene (2 : 1)	366	Ginger	In tablets	56
Ethyl acetate–methanol–acetic acid different proportions by gradient elution	254	Convallatoxin	In *Herba convallariae*	57
Methanol–water–ethyl acetate (1 : 1 : 5)	254	Vasicine	In *Adatoda vasica*	58
(1) Hexane–methanol–acetone (15 : 3 : 2); (2) benzene–acetone–ethanol (8 : 1 : 1); (3) chloroform–acetone (4 : 1)	254	Diazepam	In arogyakadha – an ayurvedic preparation	59
Ethyl acetate–methanol–13.5% ammonia	264	18-β-Glycyrrhetinic acid	In polyherbal preparations of liquorice syrup, cream, tablets	60
Chloroform–ether–methanol (5 : 2 : 1)	560	Madecassic acid	Tablets and capsules containing *Centella asiatica*	61

Stationary phase is silica gel.

instruments should have maintenance records, and that their performance and those of the method should be documented. Every analyst should in principle document an analytical method and its aims.

Since the late 1970s various leading drug regulatory bodies such as United States Food and Drug Administration (USFDA), World Health Organization (WHO) and Commission of Pharmaceutical Manufacturing Product (CPMP) of the European communities, in consultation and collaboration with various research organizations and industries, have been monitoring the efficacy of various drugs. The Centre for Drug Evaluation and Research (CDER) in the USA is one of the leading agencies which issues timely guidance as regard to various current GMP norms, formulation designs, and biostudies.

Owing to the stringent regulatory and ethical requirements, it becomes equally important to monitor the quality of drug substances with respect to their identity, potency, purity (including chiral purity), and dosage at the time of release and at various stages of shelf life. These are achieved through analytical techniques. The analytical results form the basis of the evaluation of drugs and pharmaceutical products, demonstrating their quality, efficacy and safety during drug development and product registration, and verifying their quality in QC process. Rapid and cost-effective provision of suitable results are, therefore, a fundamental challenge in pharmaceutical analytical chemistry.

Analytical methodology is in a period of transition due to technological advances and modified policies of the governmental regulatory agencies. In many countries the regulations and the practices of drug manufacturing have changed because of the recently developed ability of analysts equipped with modern instrumentation to determine low concentrations of impurities. Although HPLC is the method of choice, HPTLC is becoming increasingly popular in the field of pharmaceutical analysis.

6 VALIDATION AND QUALITY ASSURANCE

6.1 Analytical Method Validation

Analytical method validation is a summary of several validation steps involving exact proof and evaluation of the suitability, correctness, and precision of each chromatographic and nonchromatographic step and of the instrumental components. Regulatory recommendations have been developed to promote correct analytical method validation through ICH (International Conference for Harmonization) guidelines, and several articles have been published covering TLC.[15,16,19,24,62,63]

Table 4 summarizes the validation parameters specific to planar chromatography for pharmaceutical analysis. Errors originating from sample preparation, application, calibration, etc. are evaluated and the selectivity of separation and peak purity are checked. If pure standards of each analyte are available, the information may be obtained by carrying out chromatographic analysis of each standard separately under the predetermined experimental conditions. Assessment of peak purity is best performed by subjecting samples to appropriate stress conditions in order to produce decomposition products; peak overlapping may then be more easily recognized. The chromatograms of treated and untreated samples either unspiked and spiked with known concentrations of impurity standards are recorded at two or three preselected wavelengths. Any change in peak shape resolution, or peak symmetry, determined by comparing the chromatographs obtained from treated and untreated samples,

Table 4 HPTLC validation parameters for pharmaceutical analysis[a]

Validation characteristics	Types of analytical procedure		Limit	Assay content/potency
	Identification	Quantitative testing for impurities		
Accuracy	−	+	−	+
Precision:				
– Repeatability	−	+	−	+
– Intermediate precision	−	+[b]	−	+[b]
Specificity[c]	+	+	+	+
Detection limit	−	−[d]	+	−
Quantitation limit	−	+	−	−
Linearity	−	+	−	+
Range	−	+	−	+

[a] + signifies that this characteristic is normally evaluated, − signifies that it is not.
[b] In cases where repeatability has been measured, intermediate precision is not needed.
[c] Lack of specificity in one analytical procedure can be compensated by other supporting analytical procedure(s).
[d] May sometimes be needed.

can indicate nonhomogeneity of a chromatographic peak. Peak homogeneity can also be investigated by two-dimensional chromatography using a different mobile phase in the second direction to the first direction of mobile-phase flow. After trying various mobile-phase compositions during the second run and if single peak chromatograms are obtained the peak can be considered as a pure one.

The ruggedness testing of HPTLC is performed by three different analysts trained in HPTLC, analyzing the same sample under the same experimental conditions to determine problems with reproducibility of the method. Various experiments to be performed for testing HPTLC ruggedness are tabulated in Table 5. The results of evaluation are reported as a percentage of the RSD. A TLC method is judged to be rugged when the system still yields acceptable RSD results in all variations and it passes the system suitability test (Tables 6 and 7). Any parameters favorably influencing the result must be reported and the degree of control must be specified.

Table 7 System suitability data specific to planar chromatography

Test	Data to be checked[a]		
	A	B	C
Orientary R_f value	+	+	+
Minimum value of peak symmetry	+	+	+
Maximum value of peak symmetry	+	+	−
Desired peak weight	+	+	+
Detection limit	+	−	−
Limit for quantification	+	−	−
Precision (for IQ)	+	+	+
Precision (for QSP)	+	(+)	(+)

[a] A = full test – the method is used after a long period of time or it is an attempt to use the method for a different analytical purpose; B = basic test – the method is in continuous use on the same analysis but something, such as the chromatoplate, has been replaced; C = abbreviated check – the system is in continuous use under the same experimental conditions; − = check is not required; + = to be checked; (+) = to be checked if necessary.

6.2 Quality Assurance of Experimental Data and Documentation

QA of the experimental data is an important topic in method validation. If the results of the method validation fails, the QA function rejects the analytical results giving instructions to the analyst to repeat the method developed. The second function of QA includes the documentation activities of the method validation, such as

Table 5 Ruggedness testing of different HPTLC operations

Operation	Experiments performed for ruggedness checking
Sample preparation	Vary shaking time, pH, temperature, and the composition of the extracting solvents phase ratio and the number of extractions
Sample application	Vary effects of spot shape, size, and stability
Separation	Vary chamber system, mobile phase composition, temperature, and development distance for at least three different plates of the same or different samples
Spot visualization (postchromatographic derivatization)	Vary spraying or dipping conditions, temperature, and reaction time
Quantitative evaluation	Vary conditions for drying plates and detection wavelength

Table 6 System quantification and analytical tasks specific to planar chromatography

Data to be checked	Identification			Purity testing			Assay			Stability test		
	IQ	SSQ	QSP	IQ	SSQ	QSP	IQ	SSQ	QSP	IQ	SSQ	QSP
R_f or R_m values from five replicate applications of the standard solution	−	+	−	−	+	−	−	+	−	−	+	−
Precision calculated from five replicate applications of the standard solution	−	−	−	+	−	−	+	−	−	+	−	−
Precision calculated from three parallel runs of spiked placebo or blank solutions with different concentrations of analytes. Each concentration of sample is applied in triplicate	−	−	−	−	−	(+)	−	−	(+)	−	−	(+)
Peak symmetry	−	+	−	−	+	−	−	+	−	−	+	−
Peak height for one or more peaks	−	−	−	+	−	−	+	−	−	+	−	−

$R_m = \log[(1 - R_f)/R_f]$.
IQ, instrument qualification; SSQ, separation system qualification; QSP, qualification of sample preparation; − = check not required; + = to be checked; (+) = to be checked if necessary.

- compilation of the validation package;
- creation of standard operating procedures (SOP) for satisfactorily validated analytical methods according to GLP requirements;
- statistical assessment of the methodology.

To demonstrate adequate performance of the TLC system, all results must be documented and criteria formulated that should be controlled prior to the use of the TLC system (system suitability test). The necessary items for documentation are as follows:

1. Listings of:
   (a) all samples used for the experiments including sources, lot numbers, analytical results, and storage conditions;
   (b) all required instrumentation used, together with their manufacturing companies and model numbers;
   (c) all necessary reagents used for sample preparation, their required dilutions or mixtures, and storage conditions indicating shelf life;
   (d) all materials and reagents used for mobile-phase preparation including their required dilutions and mixings in detail;
   (e) type, particle size, dimensions, and thickness of chromatoplates used;
   (f) type and size of chamber used for the experiments, experimental conditions such as sample applications, temperature, saturation, developing distance, etc.;
   (g) visualization techniques including detailed description of preparation of spraying reagents, spraying conditions;
   (h) detection parameters.
2. Detailed description of the entire analytical process including sample preparation, standard preparation and step-by-step procedure (including any presaturation, preequilibration, weights, volumes, extraction times, solvent range adjustments, centrifugation times), method for evaluation and calculation.
3. Summary of developmental data for formulated products.
4. Degradation scheme for the active ingredients and for formulations.
5. All validation data, including definitions, determinations and calculations for the analytical performance parameters, together with densitograms.
6. Proposed criteria for system suitability test.
7. Methods reported as properly written SOPs in detail.
8. Detailed evaluation of the test method validation procedures containing all of the recommendations and precautions to be taken into consideration when the TLC method is being routinely used; some recommendations should be given as to when and why revalidation might be necessary.

7 METHOD DEVELOPMENT

Method development, which results in sufficient separation, is one of the most critical steps of qualitative and quantitative planar chromatography. The main steps involved in method development of HPTLC for pharmaceutical analysis are:

1. selection of the stationary phase
2. selection of the vapor phase
3. selection of suitable solvents
4. optimization of the mobile phase
5. selection of the development mode.

7.1 Selection of the Stationary Phase

Silica gel is the most widely used stationary phase for pharmaceutical analysis. Modified and chemically bonded phases are also used. Practically all stationary phases used in normal-phase and reversed-phase HPLC are now available for TLC.

Precoated TLC or HPTLC plates 20×20 cm in size with aluminum or polyester supports are usually employed, mainly for economic reasons. These plates can be cut to size and shape to spot a particular analyte using scissors. Prewashing of chromatographic plates is an essential step. After washing, the plates must be dried for a sufficient time to ensure complete removal of washing liquids in a drying cupboard, by keeping them between two blank glass plates. The washed plates should always be stored in a dust-free atmosphere under ambient conditions, preferably in desiccators without the use of grease for sealing and any drying agents. Prewashing of plates must be done to at least 1–2 cm longer than the subsequent chromatographic development so that any dirt accumulated at the front does not interfere with the densitometric scanning. Methanol is the most commonly used prewashing solvent. Mixtures of chloroform in methanol (1 : 1), ethyl acetate in methanol (1 : 1), and chloroform–methanol–ammonia (90 : 10 : 1) have also been used as solvents for prewashing. Methylene chloride–methanol (1 : 1) is best suited for removing impurities picked up during storage in the laboratory. One may also use 1% ammonia or 1% acetic acid in methanol for prewashing.

7.2 Selection of the Vapor Phase

Selection of the vapor phase is a variable offered in planar chromatography. The type of chamber used decides the vapor phase. Basically there are two types of chamber. In normal (N) chambers there is a distance of more than 3 mm between the layer and the wall or between the layer and the lid of the chamber in the horizontal development of the chromatographic tank. If this distance is smaller the chamber is said to have the sandwich (S) configuration. For saturated and unsaturated systems both these chambers can be used. Characterization of the chamber saturation is done by the marker dye test, as proposed by Nyiredy et al.[64] The time required for chamber saturation depends upon the nature and composition of the mobile phase and layer thickness. When dealing with solvents of low polarity, such as aliphatic hydrocarbons, toluene, and their mixtures, preloading of the dry layer with solvent vapors should be avoided. However, preequilibrium is often recommended for high-polarity solvents such as methanol. Development in a nonsaturated or partially saturated atmosphere is recommended with solvents used in a composition leading to phase separation, such as mixtures of *n*-butanol, water, and glacial acetic acid. However, in the case of reversed-phase thin-layer chromatography (RPTLC) it is always preferable to saturate the chamber with methanol. If chromatographic separation is to be carried out at a controlled relative humidity of the chamber, then a suitable liquid is placed in a separate device prior to inserting the plate in the tank. The relative humidity of the room is usually controlled by a dehumidifier. However, if experiments are to be carried out at specific relative humidity, then a solution of sulfuric acid or salt solutions may be employed. Twin-trough chambers are suitable for all developing techniques and comparatively smaller volumes of the mobile phase are required.

7.3 Selection of Suitable Solvents for the Mobile Phase

The use of poor-grade solvents in preparing the mobile phases has been found to decrease the resolution, spot detection, and R_f reproducibility. The mobile phase composition is generally selected by a controlled process of trial and error, based on experience in the field. The mobile phase should be chosen taking into consideration the chemical properties of analytes and the sorbent layer. The basis of the strategy for solvent selection is the solvent classification by Snyder[65] who classified more than 80 solvents into eight groups for normal-phase chromatography according to their properties as proton acceptors, proton donors, and their dipole characteristics. From these eight groups 29 solvents were chosen which are called neat solvents and used in planar chromatography. An efficient method of screening a large number of neat solvents has been described by Birkinshaw and Waters.[66] With automatic CAMAG TLC Sampler III the same sample solution is applied spot-wise in a pattern on an HPTLC plate. Thereafter, 45 µL of different solvents are applied to the same positions at a high dosage rate (500 nL s^{-1}) so that circular chromatograms develop. From the position and sharpness of the fraction rings formed, conclusions can be drawn as to the suitability of the solvents. From the solvents showing good separation, their homologs or other solvents of the same group may be tested. Consequently the structure and the properties of the compounds to be separated do not have to be known for these experiments. After these experiments the solvents giving the most appropriate separations are chosen for further optimization of the mobile phase.

7.4 Optimization of the Mobile Phase

Mobile-phase composition is highly important in planar chromatography. Since the early days of chromatography considerable effort has been devoted to optimization because of the specific problems encountered in liquid chromatography. The widespread availability of computers has been of assistance for both liquid chromatography and TLC.[67,68] Optimizing the mobile-phase composition for HPTLC consists of four levels of operation.

7.4.1 First Level

Neat solvents from different selectivity areas are tested as described above for solvent selection. If acceptable resolution in the medium R_f range is achieved here, the operator can proceed directly to level three with these solvents, i.e. the investigation of solvent mixtures. If this is not the case, level two must be followed.

7.4.2 Second Level

Solvents that leave the main fractions near the start or move them close to the front need to be adjusted in strength. If the R_f values are too high, the solvent strength is decreased by adding *n*-hexane or toluene or, if they are too low, the solvent strength is increased by the addition of water, methanol, or ethanol, usually in the ratio of 9 : 1 or 8 : 2.

7.4.3 Third Level

At this level mixtures of solvents from different selectivity groups are investigated by adjusting strength, if required. These solvent mixtures can be binary, ternary, or even quaternary. It is best to start with the center and outside selectivities, i.e. using binary mixtures in the ratios 1 : 1, 9 : 1 and 1 : 9, or ternary mixtures in the ratios

1:1:1, 8:1:1, 1:8:1, and 1:1:8. At this (third) level addition of small amounts of acidic (acetic acid) or basic (triethylamine) modifier can be considered and tested whenever this appears promising. Also at this level, if not sooner, the operator should work with the developing device to be used later in routine analysis, such as twin-trough, horizontal, or automatic developing chamber.

7.4.4 Fourth Level

At this level final optimization of the mobile phase to be used for a particular separation is made. To obtain the best separation, small variations in the proportions of different solvents may be necessary.

Solvent selection based on experience and chromatographic intuition can be very time-consuming when applied to complex mixtures. For these, more systematic strategies have been elaborated:

- simplex method[69,70]
- prisma model method[71,72]
- window diagram method[73]
- overlapping resolution map method.[74,75]

7.5 Selection of the Development Mode

Ascending linear development is most frequently used for TLC. However, horizontal development has become more common in recent years. Circular development, where the solvent system migrates readily from the centre of the plate to the periphery, is well established for compounds in the lower R_f range.[76,77] The separation power of the circular development mode is greater than for the horizontal development mode. In the anticircular development mode the solvent system enters the layer at a circular line and flows towards the centre. As the solvent flow velocity decreases with the square of the distance, but the area wetted also decreases as the square of the distance travelled, the speed of the solvent system migration is practically constant. Therefore, this mode is the fastest with respect to separation distance. Anticircular development is a widely accepted approach in analytical TLC if the resolution must be increased within the higher R_f range.[21]

Of the multiple development techniques, unidimensional multiple development is effective to improve separation in lower R_f ranges; incremental multiple development improves the zone centre separation; gradient multiple development increases the separation capacity of the chromatographic system; bivariate multiple development is effective for samples of differing polarity. For mixtures of components spanning a wide retention range some form of gradient development is required to obtain even spacing of the developments in the chromatogram. The arrival of the automated multiple development chamber has done much to popularize gradient development in TLC. In a typical automated separation the chromatogram is developed repeatedly in the same direction with successive developments being over increasing migration distances; increments are generally between 1 and 5 mm and a complete program usually comprises 10–30 cycles. Between developments the solvent is drained from the chamber and the mobile phase contained in the layer is evaporated under vacuum. New solvent of any selected composition is then mixed and fed into the developing chamber and proceeds to ascend the layer. All steps are fully automated under capillary flow-controlled conditions – the useful plate length from a separation is limited to above 80 mm, whereas in an optimized separation typical zones have widths of 2–3 mm. Thus the separation capacity for baseline-resolved zones is 25–40, a considerable increase over the separation capacity obtained using conventional development modes.

8 COMPARISON WITH OTHER METHODS

Comparisons of TLC to other chromatographic methods have been reviewed.[78–84] TLC involves the concurrent processing of multiple samples and standards on an open layer developed by a mobile phase. Development is performed usually without pressure, in a variety of modes, including simple one-dimensional, multiple, circular, and multidimensional. The detection of the zones is done statistically with an assortment of diverse possibilities. Paper chromatography is fundamentally very similar to TLC, differing mainly in the nature of the stationary phase. TLC is superior to paper chromatography because it is faster, more efficient, and allows more versatility in the choice of stationary and mobile phases. HPTLC layers are smaller, contain known sorbents with smaller, more uniform particle size, are thinner, and are developed for shorter distances compared to TLC layers. These factors lead to faster separations, reduced zone diffusion, better separation efficiency, lower detection limits, less solvent consumption, and the ability to spot more samples per plate.

HPLC involves the elution under pressure of sequential samples in a closed on-line system with dynamic detection of solutes, usually by absorption. The predominant mode of HPLC is reversed phase on bonded silica columns, whereas in normal-phase TLC silica gel is most widely used. The reasons for this difference include the change in properties of silica gel columns caused by continuous mobile-phase flow (which is unimportant in silica gel layers because they are used only once), and the impracticability of using solvents with high percentages of water to develop many brands of bonded reversed-phase layers.

TLC is the most versatile and flexible chromatographic method for the separation of all types of organic and inorganic molecules that can be dissolved and are not volatile. It is rapid because precoated layers are usually used as received without preparation. Even though it is not fully automated like HPLC, TLC has the highest sample throughput because up to 30 individual samples and standards can be applied to a single plate and separated at the same time. The ability to separate samples simultaneously in parallel lanes in TLC is important in applications requiring high sample throughput such as screening of biological fluids for drugs, determining drug residues in animal products, study of drug impurity profiles, etc. Modern computer-controlled scanning instruments and automated sample application and developing instruments allow accuracy and precision in quantification that are in many cases equivalent to HPLC and GC. There is a wide choice of layers and developing solvents (acidic, basic, completely aqueous, aqueous–organic). Solvents that can cause interference with HPLC UV detection can be used in TLC because the mobile phase is removed from the plate prior to detection. Every sample is prepared on a fresh layer, so that problems involved with carryover and cross-examination of samples and sorbent regeneration procedures are avoided. Mobile-phase consumption is low, minimizing the costs of solvents and waste disposal. Because layers are not normally reused, sample preparation methods are less demanding and complex. Also, impure samples can be applied without concern for additional (ghost) peaks and for noneluting compounds that shorten the life of HPLC columns. The wide choice of development methods and pre- or postchromatographic detection reagents leads to unsurpassed specificity in TLC, and all components in every sample, including irreversibly sorbed substances, can be detected.

There is no need to rely on peaks drawn by a recorder or to worry about sample components possibly remaining uneluted on the layer. Being an off-line method, the various steps of the procedure are carried out independently. As an example of the advantage of this approach, zones can be scanned repeatedly with a densitometer using different parameters that are optimum for individual sample components. TLC and HPTLC are cost effective and environment friendly, and the chromatograms can be preserved as well as being photo- or videodocumented.

ACKNOWLEDGMENT

I am indebted to my research students who have helped me during the literature search and preparation of the manuscript.

CAS REGISTRY NUMBERS

Amikecin	37517-28-5
Amiloride HCl	2016-88-8
L-Ascorbic acid	50-81-7
Aspirin	50-78-2
Atenolol	29122-68-7
Azatadine maleate	3978-86-7
Benzhexol HCl	52-49-3
Caffeine	58-08-2
Chloroxazone	95-25-0
Ciprofloxacin	85721-33-1
Convallatoxin	508-75-8
Danazol	17230-88-5
Dextromethophan HBr	125-69-9
Diazapam	439-14-5
Dipyradamole	58-32-2
Dipyrone	68-89-3
L-Dopa	59-92-7
Doxylamine succinate	469-21-6
Enrofloxacin	93106-60-6
Estradiol benzoate	50-28-2
Etofylline	519-37-9
Fenpiverinium HBr	125-60-0
Frusemide	54-31-9
Furazolidone	67-45-8
Fusidic acid	6990-06-3
18-β-Glycyrrhetinic acid	1449-05-4
Guaiphenesin	93-14-1
Heloperidol	52-86-8
Ibuprofen	15687-27-1
Isoniazide	54-85-3
Ketoprofen	22071-15-4
Lansoprazole	103577-45-3
Lomefloxacin	9079-51-7
Madecassic acid	18449-41-7
Metronidazole	443-48-1
Molsidomine	25717-80-0
Nalidixic acid	389-08-2
Naproxen	22204-53-1
Nimodipine	66085-59-4
Nitrinedipine	39562-70-4
Norfloxacin	70458-96-7
Ofloxacin	82419-36-1
Oxfendazole	53716-50-0
Oxytetracycline	79-57-2
Paracetamol	103-90-2
Pefloxacin	70458-92-3
D-Pencillamine	52-67-5
Phenylephrine	59-42-7
Phenylhydramine	100-63-0
Pitofenone	54063-52-4
Pseudoephidrine HCl	345-78-8

Pseudoephidrine sulfate	7460-12-0
Pyrazinamide	98-96-4
Pyridoxine HCl	65-23-6
Reserpine	50-55-5
Rifampicin	13292-46-1
Salbutamol sulfate	51022-70-9
Sulfamethaxazole	723-46-6
Theophylline	58-55-9
Tinidazole	19387-91-8
Trimethoprim	738-70-5
Vasicine	6159-55-3

ABBREVIATIONS AND ACRONYMS

API	Active Pharmaceutical Ingredients
CDER	Centre for Drug Evaluation and Research
CFA	3-Chloro-4-fluoro Aniline
CPMP	Commission of Pharmaceutical Manufacturing Product
CUT	Content Uniformity Test
DEA	Diethylamine
EMME	Ethoxy Methyl Malonate Ester
EOTC	4-Ethoxytetracycline
GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
HSTLC	High-speed Thin-layer Chromatography
ICH	International Conference for Harmonization
IND	Investigational New Drug
IQ	Instrument Qualification
NDA	New Drug Application
OPLC	Over-pressured Liquid Chromatography
OTC	Oxytetracycline Hydrochloride
QA	Quality Assurance
QC	Quality Control
QSP	Qualification of Sample Preparation
RPHPTLC	Reversed-phase High-performance Thin-layer Chromatography
RPL	Rotation Planar Chromatography
RPTLC	Reversed-phase Thin-layer Chromatography
RSD	Relative Standard Deviation
SOP	Standard Operating Procedures
SSQ	Separation System Qualification
TLC	Thin-layer Chromatography
USFDA	United States Food and Drug Administration
UV	Ultraviolet
WHO	World Health Organization
α-APC-OTC	α-Apoxytetracycline
β-APC-OTC	β-Apoxytetracycline

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Planar Chromatography in Clinical Chemistry

REFERENCES

1. N. Pelick, H.R. Bolliger, H.K. Mangold, in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, 85, Vol. 3, 1966.
2. E. Stahl, in *Thin-layer Chromatography – A Laboratory Handbook*, 2nd edition, ed. E. Stahl, Springer Verlag, Berlin, 1, 1969.
3. J. Sherma, in *A History of Analytical Chemistry*, eds. H.A. Laitenen, G.W. Ewing, ACS Division of Analytical Chemistry, Washington, DC, 306, 1977.
4. E. Stahl, *J. Chromatography*, **165**, 59 (1979).
5. J.G. Kirchner, *Thin-layer Chromatography*, 2nd edition, Wiley, New York, 1978.
6. J.G. Kirchner, in *Thin-layer Chromatography Quantitative Environmental and Clinical Applications*, eds. J.C. Touchstone, D. Rogers, Wiley-Interscience, New York, 1980.
7. U. Wintermeyer, *The Root of Chromatography – Historical Outline of the Beginning to Thin-layer Chromatography*, GIT Verlag, Darmstadt, 1989.
8. J.G. Kirchner, J.M. Miller, R.G. Rice, *J. Agr. Food Chem.*, **2**, 1031 (1954).
9. M.S.J. Dallas, C.B. Barret, F.P. Padley, *Quantitative Analysis of Triglyceride Mixtures by TLC on SilicaGel Impregnated with SilverNitrate*, Joyce Loebl Publication, 3, 1964.
10. C.E. Genest, *J. Chromatogr.*, **18**, 531 (1965).
11. A.E. Thomas, J.E. Scharoun, H. Ralston, *J. Am. Oil Chem. Soc.*, **42**, 790 (1965).
12. A.H. Stead, R. Gill, T. Wright, J.P. Gibbs, A.C. Moffat, *Analyst*, **107**, 1106 (1982).
13. B. Fried, J. Sherma, *Pharmaceuticals in Thin-layer Chromatography – Techniques and Applications*, Marcel Dekker, New York, 1986.
14. E.C.G. Clerke, *Isolation and Identification of Drugs*, Pharmaceutical Press, London, Vol. 1, 1969 and Vol. 2, 1975.
15. G. Szepesi, *Chromatography in Steroid Analysis in the Pharmaceutical Industry*, ed. S. Gorog, Ellis Horwood, Chichester, 103, 1989.

16. G. Szepesi, 'Steroids', in *Quantitative Thin-layer Chromatography and Its Industrial Applications*, ed. L. Treiber, Marcel Dekker, New York, 1986.
17. T.C. Touchstone, *Quantitative Thin-layer Chromatography*, John Wiley & Sons, New York, 1973.
18. L.L. Ng, 'Pharmaceuticals and Drugs', in *Handbook of Thin-layer Chromatography*, eds. J. Sherma, B. Fried, Marcel Dekker, New York, 717, 1991.
19. G. Szepesi, 'Quantitation', in *Modern Thin-layer Chromatography*, ed. N. Grinberg, Marcel Dekker, New York, 249, 1990.
20. G. Szepesi, 'Instrumentation', in *Modern Thin-layer Chromatography*, ed. N. Grinberg, Marcel Dekker, New York, 285, 1990.
21. J.C. Touchstone, J. Sherma, 'Densitometry', in *Thin-layer Chromatography*, John Wiley & Sons, New York, 1979.
22. C.F. Poole, S.K. Poole, T.A. Dean, in *Recent Advances in Thin-layer Chromatography*, eds. F.A.A. Dallas, H. Read, R.J. Ruane, D. Wilson, Plenum Press, New York, 11, 1988.
23. D.E. Jansen, in *Hand Book of Thin-layer Chromatography*, eds. J. Sherma, B. Fried, Marcel Dekker, New York, 113, 1991.
24. G. Szepesi, Sz. Nyiredy, *J. Pharm. Biomed. Anal.*, **10**, 1007 and 1017 (1992).
25. P.D. Sethi, *High Performance Thin-layer Chromatography. Quantitative Analysis of Pharmaceutical Formulations*, CBS Publishers and Distributors, New Delhi, 1996.
26. G. Szepsi, Sz. Nyiredy, in *Handbook of Thin-layer Chromatography*, 2nd edition, eds. J. Sherma, B. Fried, Marcel Dekker, New York, 819, 1996.
27. R.A. Sodhi, J.L. Chawla, R.T. Sane, *Indian Drugs*, **34**(8), 433–436 (1997).
28. J.K. Lalla, Siu. Bhat, N.R. Sandu, M.U. Shah, *Indian Drugs*, **34**(5), 275–282 (1997).
29. R.A. Sodhi, J.L. Chawla, R.T. Sane, *Indian Drugs*, **33**(12), 601–603 (1996).
30. M.B. Lippstone, E.K. Grath, J. Sherma, *J. Planar Chromatogr. – Mod. TLC*, **9**(6), 456 (1996).
31. P.P. Shirke, M.K. Patel, V.A. Tamhane, V.B. Tirodkar, P.D. Sethi, *Indian Drugs*, **30**(12), 653–654 (1993).
32. J.L. Chawla, R.A. Sodhi, R.T. Sane, *Indian Drugs*, **33**(5), 208–212 (1996).
33. M.B. Lippstone, J. Sherma, *J. Planar Chromatogr. – Mod. TLC*, **8**(6), 427–429 (1995).
34. R.A. Sodhi, J.L. Chawla, R.T. Sane, *Indian Drugs*, **33**(6), 280–285 (1996).
35. D. Agbaba, O. Grozdanovic, L. Popovic, S. Vladimirov, D. Zivanov-Stakic, *J. Planar Chromatogr. – Mod. TLC*, **9**(2), 116–119 (1996).
36. N.M. Tendolkar, B.S. Desai, J.S. Gaudh, V.M. Shinde, *Anal. Lett.*, **28**(9), 1641–1653 (1995).
37. V.M. Shinde, N.M. Tendolkar, B.S. Desai, *Anal. Lett.*, **2**(1), 45–58 (1995).
38. V.M. Shinde, N.M. Tendolkar, B.S. Desai, *Indian Drugs*, **31**(3), 119–122 (1994).
39. M. El-Sadak, A. El-Shanawany, A. Aboul-Khier, G. Ruecker, *Analyst (London)*, **115**(9), 1181–1184 (1990).
40. A.P. Argekar, S.V. Raj, S.U. Kapadia, *Indian Drugs*, **32**(4), 116–171 (1995).
41. A.P. Argekar, S.S. Kunjir, *J. Planar Chromatogr. – Mod. TLC*, **9**(1), 65–67 (1996).
42. A.P. Argekar, S.V. Raj, S.U. Kapadia, *J. Planar Chromatogr. – Mod. TLC*, **9**(2), 148–151 (1996).
43. A.P. Argekar, S.V. Raj, S.U. Kapadia, *J. Planar Chromatogr. – Mod. TLC*, **9**(3), 208–211 (1996).
44. A.P. Argekar, S.V. Raj, S.U. Kapadia, *Indian Drugs*, **33**(4), 167–170 (1996).
45. A.P. Argekar, S.S. Kunjir, K.S. Purandare, *J.Pharm. Biomed. Anal.*, **14**(11), 1645–1650 (1996).
46. A.P. Argekar, S.S. Kunjir, *J. Planar Chromatogr. – Mod. TLC*, **9**(4), 296–299 (1996).
47. A.P. Argekar, S.S. Kunjir, *J. Planar Chromatogr. – Mod. TLC*, **9**(5), 390–394 (1996).
48. A.P. Argekar, S.V. Raj, S.U. Kapadia, *J. Planar Chromatogr. – Mod. TLC*, **9**(6), 459–461 (1996).
49. A.P. Argekar, S.V. Raj, S.U. Kapadia, *Indian Drugs*, **33**(3), 107–111 (1996).
50. A.P. Argekar, S.G. Powar, *J. Planar Chromatogr. – Mod. TLC*, **11**(4), 254–257 (1998).
51. A.P. Argekar, J.G. Sawant, *J. Chromatogr.*, **22**(10), 1561–1569 (1999).
52. J.G. Sawant, PhD Thesis, University of Mumbai, 1999.
53. S. Ded, S.K. Mandal, *Indian Drugs*, **33**(8), 415–416 (1996).
54. R.T. Sane, M.M. Phadke, P.S. Hijli, M.M. Shah, *Indian Drugs*, **34**(4), 89–94 (1997).
55. V.K. Kuber, J.L. Chawla, R.T. Sane, *Indian Drugs*, **34**(1), 36–42 (1997).
56. R.T. Sane, M.M. Phadke, P.S. Hijli, M.C. Shah, P.P. Deshpande, *Indian Drugs*, **33**(9), 462–464 (1996).
57. G. Matysik, J. Kowalski, M. Staszewska, *Chromatographia*, **43**(9–10), 559–562 (1996).
58. D.B.A. Narayana, S. Agrawal, S.K. Luthra, N.S. Shrinivas, *Indian Drugs*, **32**(12), 583–586 (1995).
59. V.W. Kamble, M.V. Garad, S.V. Padalkar, *J. Planar Chromatogr. – Mod. TLC*, **8**(2), 143–144 (1995).
60. A.P. Argekar, S.S. Kunjir, *J. Planar Chromatogr. – Mod. TLC*, **9**, 220–223 (1996).
61. S.S. Kunjir, PhD Thesis, University of Mumbai, 1996.
62. G. Szepesi, *J. Planar Chromatogr. – Mod. TLC*, **6**, 187 (1993).
63. G. Szepesi, *J. Planar Chromatogr. – Mod. TLC*, **6**, 259 (1993).
64. Sz. Nyiredy, Zs. Fater, L. Boltz, O. Sticher, *J. Planar Chromatogr. – Mod. TLC*, **5**, 308 (1992).
65. L.R. Snyder, *J. Chromatogr. Sci.*, **16**, 223 (1978).
66. F.L. Birkinshaw, D.G. Waters, *J. Planar Chromatogr. – Mod. TLC*, **8**, 319–323 (1995).
67. J.C. Barriidge, *Techniques for the Automated Optimization of HPLC Separation*, John Wiley & Sons, New York, 1985.

68. L.R. Snyder, J.L. Glajch, J.J. Kirkland, *Practical HPLC Method Development*, John Wiley & Sons, New York, 1988.
69. L.G. Sabate, X. Thomas, *J. High. Resolut. Chromatogr.*, **7**, 104–106 (1984).
70. B.M.J. De Spiegeleer, P.H.M. De Moerloose, G.A.S. Slegers, *Anal. Chem.*, **59**, 62–64 (1987).
71. Sz. Nyiredy, B. Meier, C.A.J. Erdelmeier, O. Sticher, *J. High Resolut. Chromatogr.*, **8**, 186–192 (1985).
72. K. Dallenbach-Toelke, Sz. Nyiredy, S.Y. Meszaros, O. Sticher, *J. High Resolut. Chromatogr.*, **10**, 362–371 (1987).
73. D. Nurok, *Chem. Rev.*, **89**, 363–375 (1989).
74. J.L. Glajch, J.J. Kirkland, K.M. Squire, J.M. Minor, *J. Chromatogr.*, **199**, 57–61 (1980).
75. S.F.Y. Li, H.K. Lee, C.P. Ong, *J. Chromatogr.*, **506**, 245–252 (1990).
76. F. Geiss, *Fundamentals of Thin-layer Chromatography (Planar Chromatography)*, Huthig, Heidelberg, 1987.
77. Sz. Nyiredy, in *Chromatography*, 5th edition, ed. E. Heftman, Elsevier, Amsterdam, 109–150, 1993.
78. J. Sharma, *J. Assoc. Off. Anal. Chem.*, **74**, 435 (1991).
79. L.R. Treiber, *J. Chromatogr. Sci.*, **24**, 220 (1986).
80. U.A.Th. Brinkman, *J. Planar Chromatogr. – Mod. TLC*, **1**, 150 (1988).
81. H.E. Hauck, *Fresenius' J. Anal. Chem.*, **327**, 24 (1987).
82. D.C. Fenimore, C.M. Devis, *Anal. Chem.*, **53**, 252A (1981).
83. B. Fried, J. Sherma, *Thin-layer Chromatography – Techniques and Applications*, 3rd edition, Marcel Dekker, New York, 1994.
84. J. Sherma, B. Fried (eds.), *Handbook of Thin-layer Chromatography*, 2nd edition, Marcel Dekker, New York, 1996.

Proteins and Peptides Purification in Pharmaceuticals Analysis

Marvin E. Thrash, Jr and Neville G. Pinto
Department of Chemical Engineering, University of Cincinnati, USA

1 Introduction 1
2 Background 2
3 Applications 6
3.1 Reversed-phase Chromatography 6
3.2 Hydrophobic Interaction Chromatography 8
3.3 Ion-exchange Chromatography 12
3.4 Affinity Chromatography 14
3.5 Size-exclusion Chromatography 16
3.6 Electrophoresis 19
4 Detection of Column Response 20
5 Scale-up 23
6 Conclusions 25
Abbreviations and Acronyms 26
Related Articles 26
References 26

The biopharmaceutical market has exploded in the last 10 years, with annual sales easily exceeding a billion dollars per year. Analysis and purification are essential components in the production of therapeutic proteins and peptides, and chromatography is at the core of all pharmaceutical separations. The more popular chromatographic techniques used in the purification of proteins and peptides are reversed-phase chromatography (RPC), hydrophobic interaction chromatography (HIC), ion-exchange chromatography (IEC), affinity chromatography (AC) and size-exclusion chromatography (SEC). In order to meet the challenges of the future it is essential to understand the fundamentals of each technique. In this article an overview of the more popular chromatographic techniques employed to purify biomolecules is presented. Relevant examples are also given and a discussion of detection and scale-up is provided.

1 INTRODUCTION

Downstream purification is an essential component of any chemical process. Products from industrial chemical, biochemical and biopharmaceutical processes all require a certain level of purity before they are commercially sold. Pharmaceutical products in particular must meet stringent purity requirements. Pharmaceutical production plants often rely on chromatography as an integral part of their purification procedures. With the advent of biopharmaceuticals, chromatography is being utilized as a means of purifying very complex biological mixtures containing therapeutic proteins and peptides. The intent of this article is to discuss the role of chromatography in biopharmaceutical processes. Before the discussion on chromatography begins, a brief introduction concerning the potential uses of proteins and peptides as therapeutic agents is presented.

Proteins and peptides manufactured by pharmaceutical companies are part of the ever-evolving biopharmaceutical arena. Simply put, biopharmaceuticals are drugs produced by genetically engineered microorganisms, transgenic animals or transgenic plants. These drugs are usually proteins or peptides. Proteins and peptides are large macromolecules that are responsible for performing very specific tasks in living organisms. The specific role of each molecule is dictated by the amino acid sequence. One example of this specificity can be seen in the digestive enzyme pepsin.[1] This protein is found in the intestinal mucosa of most animals. Pepsin has only one purpose and that is to fragment the proteins and peptides consumed with each meal. Pepsin cleaves each protein at the same location regardless of type. In principle if biopharmaceutical drugs can be developed and designed specifically to target complicated diseases such as AIDS (acquired immunodeficiency syndrome) or cancer, thousands if not millions of people will benefit.

The specificity of protein molecules has intrigued medical researchers for years. This is because the challenge in all pharmaceutical research is to develop medications that are specific enough to disable the infectious agent without harming the human host. In some cases the medical community has been very successful. An excellent example of this can be seen in the development of the smallpox vaccine. When the human body has been exposed to the vaccine the immune system produces very specific antibodies that are capable of protecting the host from a smallpox infection. Antibodies are simply protein molecules whose sole purpose is to protect the host from microbial or viral invaders. The introduction of this vaccine has greatly minimized the risk of this virus once again becoming rampant among the general

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

population. This approach cannot however be universally applied to other types of viral infection. An example of this is AIDS. Scientists have not yet found a nonvirulent form of the AIDS virus that can be used for vaccination purposes. A possible alternative route that can be used in fighting viruses like AIDS may potentially be found in a specific class of biopharmaceuticals called monoclonal antibodies (Mabs). Mabs are antibodies that have been specifically engineered to target specific types of foreign invaders.[2] For example if a certain species is immune to the AIDS virus, this means that this species must possess the ability to produce antibodies that are capable of neutralizing the virus. If this antibody can be identified, then perhaps genetically engineered microorganisms or transgenic animals can produce it. In principle, when Mabs are introduced to the human host they should protect the human in the same manner that they protect their natural host.

Researchers are also testing the effectiveness of biopharmaceuticals in the treatment of nonviral diseases such as cancer. Surgery, radiation and chemotherapy are three common approaches used to fight this disease. These techniques have helped many patients to extend their lifespan significantly; however, for some cancer patients radiation and chemotherapy produce extremely undesirable side effects. In some cases the chemotherapy and/or radiation do not stop or reverse the progress of the cancer. This has lead biopharmaceutical researchers to look for a better approach. One example of this can be seen in the ongoing research with endosatin and angiostatin.[2] These proteins are specifically designed to constrict the blood flow to cancerous tumors. These and other anticancer biopharmaceuticals are still under investigation. If they do prove to be effective in clinical trials, they should, in principle, produce little or no side effects because of the ''task specificity'' of protein molecules.

Still another example of the potential of biopharmaceuticals is gene therapy.[3] Gene therapy can be divided into two categories. The first category is somatic cell gene therapy, where the genetic components of certain cells are altered to assist the host in fighting off certain diseases. This alteration is not inheritable by the offspring of the host. The second category is germ line gene therapy, where the genetic material of a fertilized egg is altered. The offspring inheriting these genetic traits are permanently altered. The overall goal of gene therapy is to protect the host from monogenic diseases such as adenosine deaminase (ADA) deficiency, cystic fibrosis, Duchenne muscular dystrophy, polycystic kidney disease, melanoma, neuroblastoma, brain tumors, and many more.

The first biopharmaceutical drug approved by the FDA (Food and Drug Administration) was Humulin, which was developed by The Eli Lilly Company. FDA approval of this drug came in the mid-1970s, 15 years after the discovery of DNA. By 1987 four biopharmaceutical drugs were approved by the FDA, and by mid-1998 approximately 79 biopharmaceuticals were approved for use. Currently, biotech drugs generate over one billion dollars of revenue annually, and the figure is expected to rise sharply.[2]

The production of biopharmaceutical products is linked to the identification of specific genes or receptor sites and genetic engineering technology. Once the desired protein drug has been identified, it is up to the engineering and manufacturing community to produce this drug on a large scale. In a typical biotech process the desired protein is created by genetically engineered microorganisms. If the desired protein resides in the cytoplasm of the microorganism, the cells are ruptured to release the protein into solution. If the protein is soluble, the fermentation broth is filtered or centrifuged to remove excess cellular material. This is usually followed by an enrichment step that could typically involve membrane filtration or extraction. If the protein cannot be adequately purified by these techniques then chromatography is employed as the next step.

Chromatography is at the core of most analytical separations. It has been extensively used in the biotech industry for the purification of proteins and peptides. The popularity of chromatography can be attributed to the versatility of this technique. There are a number of stationary phases that can be commercially purchased. Each stationary phase has the ability to separate analytes by exploiting the differing affinities each analyte has for the immobilized ligands on the stationary phase surface. Additionally, the mobile phase composition and/or pH are key variables that can also be exploited to achieve the successful purification of the desired product.

In this article we will attempt to identify the chromatographic techniques employed in biopharmaceutical processes. Additionally, the use of detectors will be covered and issues relating to scale-up will also be discussed.

2 BACKGROUND

Proteins and peptides are composed of amino acids. The sequence of amino acids dictates the role of a given protein or peptide. In general a peptide has one specific amino acid sequence. A protein on the other hand may have multiple amino sequences and proteins may also possess inorganic compounds, such as iron in the case of hemoglobin. The physical configuration of proteins and peptides also depends on the amino acid sequence. In an aqueous environment each molecule is folded in a manner that forces a preponderance of hydrophobic amino acid

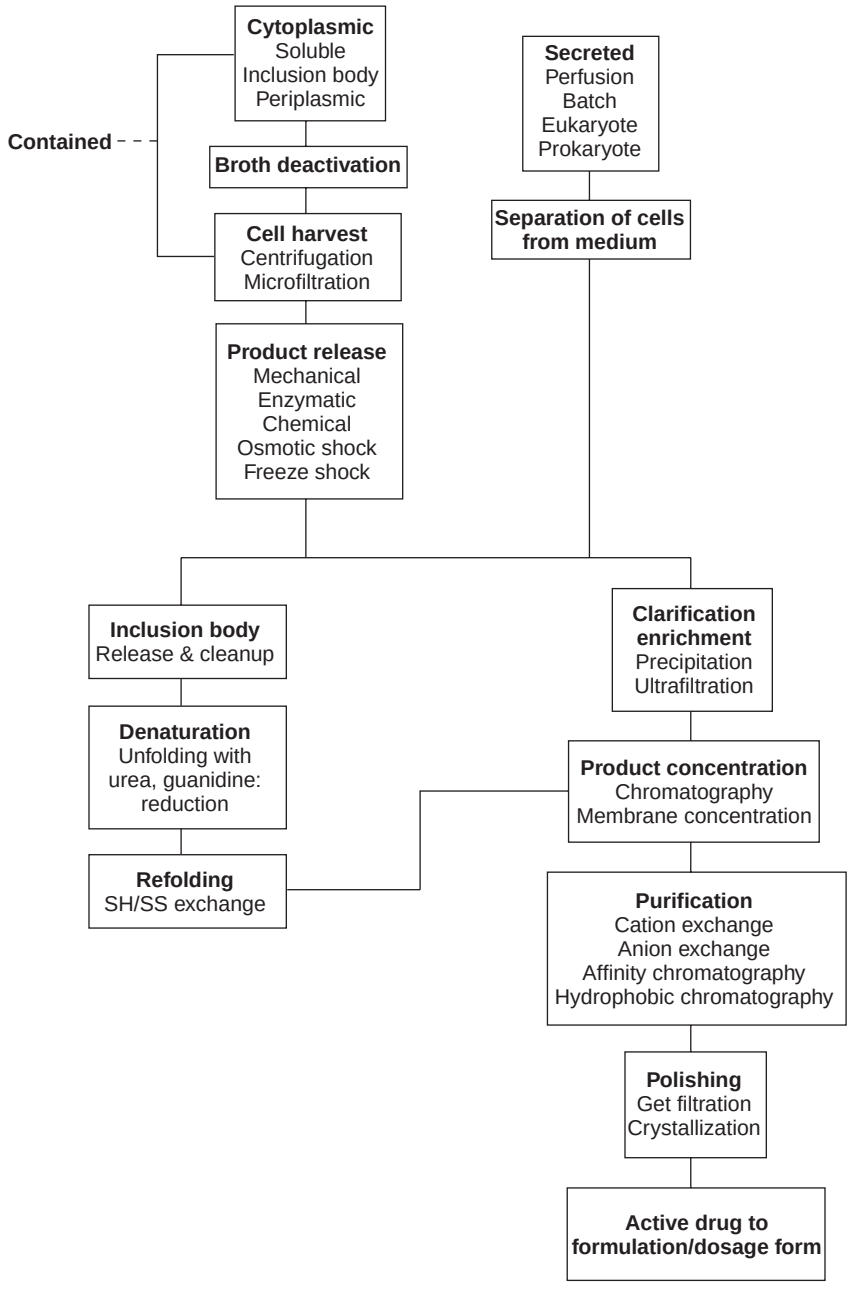

Figure 1 Generalized block diagram of recombinant DNA process. (Reproduced with permission from A. Sadana, in *Biopharm*, 26–33, March 1994,[3] Copyright by Advantstar Communications Inc. Advanstar Communications Inc. retains all rights to this article.)

groups to the interior of the molecule and hydrophilic amino acid groups to the exterior of the molecule. The activity of these macromolecules is highly dependent on the physical structure. When a protein or peptide loses its unique structure, it is considered denatured. From an engineering standpoint high temperature and excessive agitation can cause denaturation. Additionally, exposure to certain chemicals and harsh environments can trigger unfolding. An example of a harsh environment would be an environment in which the pH diminishes the activity of the protein. Thus it is essential that the production and purification process conditions be specified in such a way as to keep the structure of these molecules intact.

A general flowchart showing the production and purification of recombinant DNA proteins is shown in Figure 1.[4] The first step in this process is fermentation.

The desired protein manufactured by the cells is either contained internally within the cytoplasm or excreted through the cell wall. If the protein is excreted through the cell wall, the cell lysis (disruption) step is eliminated. Also the recovery of the protein becomes easier because the number of impurities in the bioreactor effluent is far fewer. Many bioprocesses do however require disruption. During cell disruption the desired protein is released into solution along with numerous other molecules that comprise the cell wall and internal structure. Ultrafiltration membranes remove many of the impurities released into solution by disruption. Since proteins are known to adsorb to hydrophobic materials, the membranes used in the clarification process should be hydrophilic in nature to minimize adsorption and permanent denaturation. The filtration process must remove as many impurities as possible. At this point in the process the composition of the solution must be known because chromatography is generally the next step. Accurate knowledge of the feed composition is essential in identifying the type of stationary phase and the operating conditions that are needed to achieve further product purity.

Figure 1 reveals that IEC, AC, HIC and gel filtration chromatography (GFC) are the popular chromatographic options for protein purification. IEC is an obvious choice because proteins are charged molecules and IEC exploits differences in charge to purify biomolecules. HIC exploits differences in the hydrophobicity of proteins. The speed and resolution of this technique is comparable to IEC and SEC. HIC also has additional benefits in that it can remove endotoxins, nucleic acids and viruses from the sample matrix.[5] AC utilizes highly specific immobilized ligands on the stationary phase. These ligands attract molecules possessing similar or complementary moieties. Finally GFC separates proteins and peptides by their size. In GFC the smallest molecules exit the column last and the largest molecules exit the column first.

An example of a plasmid DNA process[6] is shown in Figure 2. This figure again shows chromatography as an integral part of the purification process. In this case the stationary phase will either be an anion-exchange resin or a size-exclusion gel. The selection of the stationary phase generally depends on the nature of the impurities. If some form of adsorptive chromatography is used, it is generally desirable to have the product as the most strongly adsorbed component.[3] With this strategy the impurities will exit the column first. If SEC is used, it is essential that the size differences between the product and the impurities be large enough to achieve a good separation. As previously stated, it is preferable for the desired product to be the most strongly adsorbed. Under this scenario the desired product can be captured with minimal risk of contamination. If the desired product is not the strongest adsorbing analyte then the process designer must ensure that there is adequate resolution between the product and the leading and trailing impurities. A schematic flow sheet of a chromatographic unit operation[7] is shown in Figure 3. It is evident from this diagram that the operator and/or the control system must be aware of everything that is flowing through the system. It is also apparent that good resolution is essential to minimize contamination and maximize yield. The actual time the product tank inlet valve is opened and the amount of time this valve stays open is highly dependent upon the resolution that is achieved by the column.

The heart of the system shown in Figure 3 is the column. The proper selection of the column stationary phase is vital. The selection of the mobile phase can also influence the separation, but the stationary phase undoubtedly has a large effect on the separation. The chromatographic designer must look for physical parameters that distinguish between the product and the impurities. As previously stated, molecular size difference is often used as a means of purifying proteins and peptides. In many other cases molecular charge differences can be exploited, while in other applications differences in solubility and/or hydrophobicity are used to achieve a separation. Affinity can also provide an effective means of achieving purity. One method by which chromatographic designers can explore these physical differences is by collecting capacity factor data for the product and each impurity in the sample matrix. The stationary phase that provides the greatest differential between the product capacity factor and the capacity factors for the impurities is normally chosen for plant operations. Another way of obtaining the same data is to produce pure component isotherms. The quantity of material required for these experiments may be an issue; however, it is important to note that isotherm data are critical because the capacity factor approach is not adequate when the chromatographic system is overloaded, as is the case in many pharmaceutical applications. Components with the greatest attraction for the stationary phase will also have the highest loading. The stationary phase that demonstrates the greatest differential between the product loading and impurity loading will normally be chosen for plant operations. The former method is currently the most popular, because correlations have been developed that relate the capacity factor to peak resolution. These correlations also allow designers to estimate column length for scale-up.

An overview of the various stages of protein fractionation is shown in Table 1.[8] In the initial fractionation step the main goal is to capture and concentrate similar

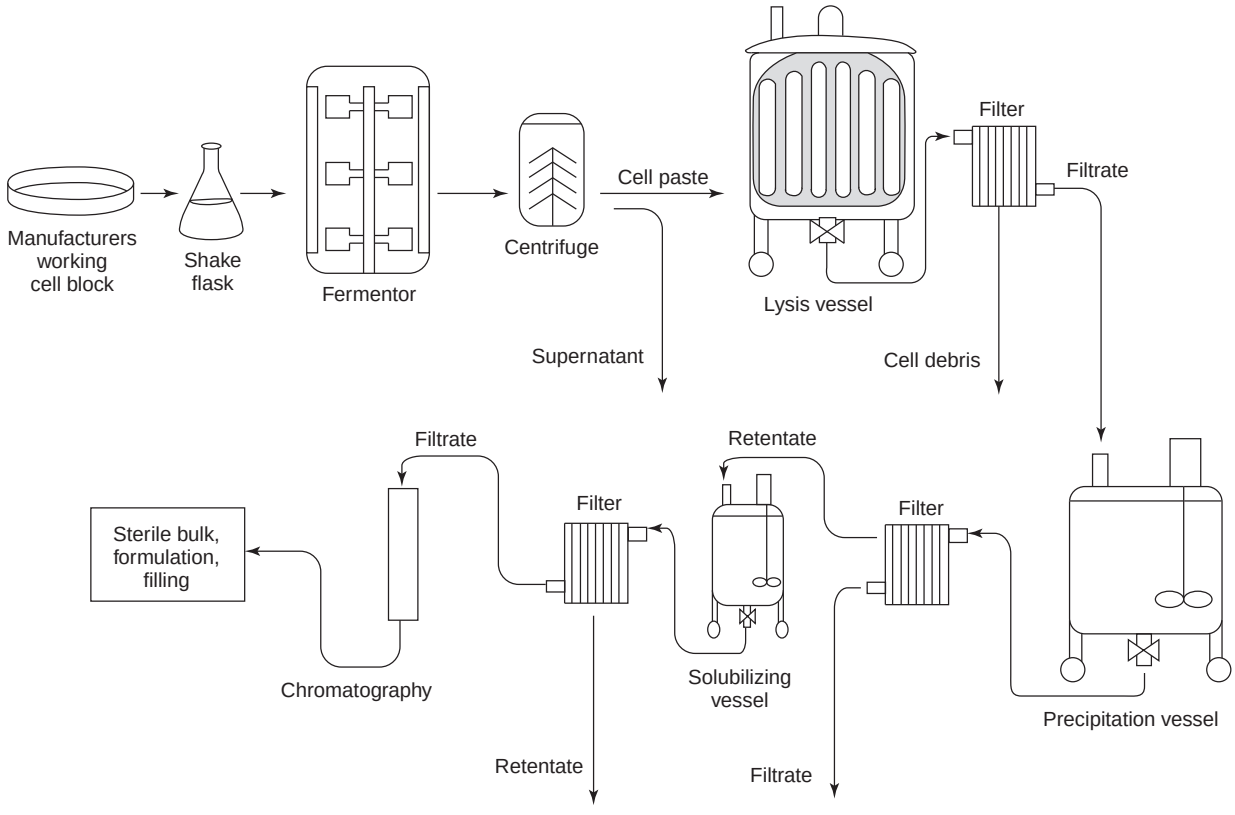

Figure 2 Flow diagram for plasmid DNA process. (Reproduced by permission from M. Marquet, N.A. Horn, J.A. Meek, in *Biopharm*, 27–37, September 1995,[6] Copyright by Advantstar Communications Inc. Advanstar Communications Inc. retains all rights to this article.)

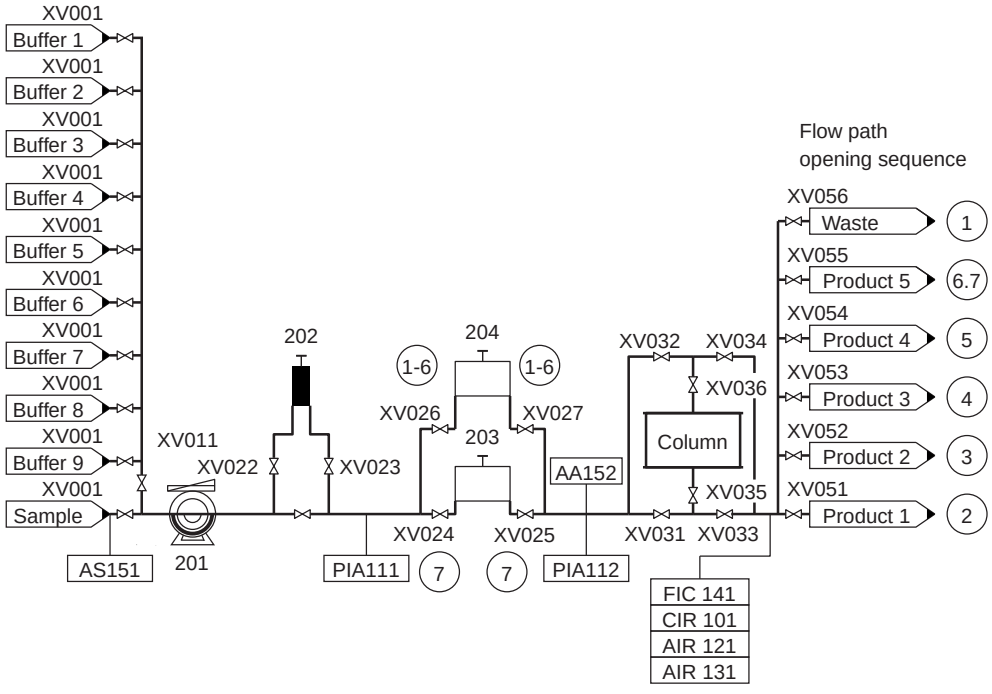

Figure 3 Detailed flow diagram of production scale chromatographic unit. (Reproduced by permission from P.A. Gariopy, R.S. Moore, G. Sofer, in *Biopharm*, 32–36, September 1998,[7] Copyright by Advantstar Communications Inc. Advanstar Communications Inc. retains all rights to this article.)

Table 1 Purification goals and methods[a]

	Initial fractionation	Intermediate purification	Final purification
Aim	Concentration Group separation Conditioning	High-resolution separation	Final polishing Polymer removal Transfer to formulation buffer
Chromatographic separation technique	Ion exchange Hydrophic interaction Affinity Gel filtration	Ion exchange Hydrophobic interaction Affinity	Gel filtration Reversed phase

[a] Reproduced from G. Ganetsos and P.E. Barker, in *Preparative and Production Scale Chromatography*, 1993,[8] by courtesy of Marcel Dekker Inc.

species into groups. The goal of intermediate fractionation is to separate the components that were concentrated into groups in the initial stage. The final stage is polishing. The goal here is to attain the required purity as specified by the FDA. As seen in Table 1, the various modes of chromatography that are commonly employed in protein and peptide purification are reversed phase, hydrophobic interaction, ion exchange, affinity and size exclusion. These techniques that are used to purify peptides and proteins in pharmaceutical applications will be discussed in the next sections. Because of the competitive nature of the pharmaceutical business, pharmaceutical companies rarely publish articles describing their process capabilities. Therefore the examples cited in this article come from research articles that have been published in key scientific journals.

3 APPLICATIONS

3.1 Reversed-phase Chromatography

RPC is the most popular form of chromatography used in the purification of pharmaceuticals. The popularity of this technique is due to its flexibility. In RPC the column is packed with a nonpolar stationary phase. The solute is brought into contact with the stationary phase by a polar mobile phase and is eventually displaced from the stationary phase with a suitable solvent. The main role of RPC in the biopharmaceutical industry is in the purification of polypeptide mixtures. RPC is well suited to handle very complex mixtures of polypeptides. This is because a number of parameters such as pH, temperature, mobile phase composition and solvent gradient can all be adjusted to obtain the desired product in a highly purified state. This type of flexibility underscores the popularity of this technique.

The separation of components in RPC is primarily achieved by differences in solute hydrophobicity and solubility. The mobile phase is usually an aqueous buffer that contains ion-pairing agents such as trifluoroacetic acid (TFA). In most applications the ion-pairing agent is added to buffer in low concentrations to assist in solubilizing hydrophobic macromolecules such as large polypeptides. The stationary phase in many RPC applications usually contains *n*-octadecyl, *n*-octyl or *n*-butyl moieties bonded to the support. These stationary phases are all hydrophobic in nature with *n*-octadecyl being the most hydrophobic and *n*-butyl being the least hydrophobic. The particles comprising the stationary phase are highly porous. The average particle diameter ranges from approximately 3 μm for analytical columns to 100 μm for preparative columns. A summary of common RPC stationary phases[9] is presented in Figure 4.

The elution of analytes in RPC is accomplished by adding an organic modifier to the mobile phase. Typical organic modifiers that are commonly used are acetonitrile, isopropanol, ethanol and methanol. The solubility of each analyte in the ''eluting'' mobile phase is also a key factor because the elution order of each analyte is a function of the mobile phase solvent strength. The solvent strength of the mobile phase depends on the type and concentration of the organic modifier used. In a typical application, analytes are desorbed from the stationary phase by a linear gradient in which the concentration of the organic modifier is linearly increased. When the concentration of the organic modifier reaches a certain level the analyte's affinity for the mobile phase becomes greater than its affinity for the stationary phase. In other words when the solvent strength reaches a certain level it is thermodynamically more favorable for the analyte to be in the mobile phase than adsorbed to the surface phase. Therefore subtle differences in solubility can be exploited to provide adequate separation and purification of sample mixtures.

A number of mathematical models describing RPC have been published. Among the most common are the solvaphobic theory[10] and the stoichiometric displacement model (SDM).[11] A detailed description of each theory is beyond the scope of this chapter and the reader is

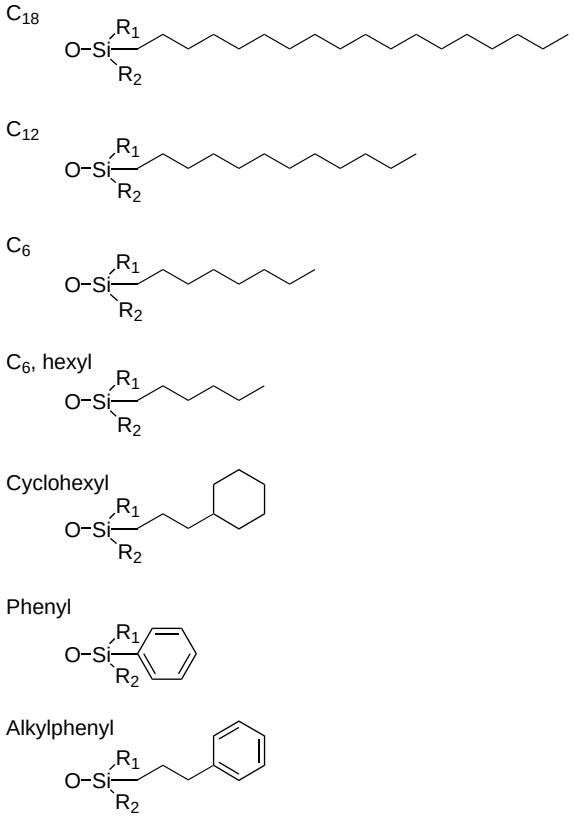

Figure 4 Examples of RPC stationary phases. (Reproduced by permission of Wiley-VCH, a subsidiary of John Wiley & Sons, Inc., from Uwe D. Neue, *HPLC Columns*, 1997,[9] Copyright Wiley-VCH, 1997.)

referred to the references cited for a detailed description. It is however useful to outline the simplest of these models briefly, the SDM. This model is based on the premise that when a solvated molecule is adsorbed by the adsorbent, a stoichiometric number (Z) of solvent molecules are released at the interface between the adsorbent and the solute. This SDM equation for the equilibrium constant (K) in RPC is shown in Equation (1):

$$K = \frac{P_b[D_o]^z}{P_m[L_d]^n} \tag{1}$$

where P_b and P_m are the adsorbed protein concentration and unbound protein concentration, respectively, D_o is the solvent concentration and L_d is the ligand density. The reported expression for the capacity factor is (Equation 2):

$$\log k' = z \log \frac{1}{D_o} + \log[I] \tag{2}$$

where $I = ([L_d]^n \phi K)/[D_o]^z$ and ϕ is the phase ratio. If the ligand density and phase ratio are known, the equilibrium constant can be calculated directly. Once K is known, the linear region of the isotherm can be approximated from Equation (2). In order to approximate the nonlinear portion of the isotherm, activity coefficient data are needed.

Although RPC is widely used in the pharmaceutical industry, it is only used in limited applications for the purification of therapeutic proteins. Typical biopharmaceutical applications that are amenable to RPC are the purification of polypeptides and small proteins that tend to be mildly hydrophobic in nature. The limitations are primarily due to the problem of denaturation. Proteins and certain polypeptides have unique three-dimensional structures. When proteins with strong hydrophobic interiors are brought into contact with a strongly hydrophobic stationary phase, the protein will rearrange its structure to allow the hydrophobic interior to have maximum exposure to the surface. If the hydrophobic patches have a strong affinity for the stationary phase, irreversible adsorption can sometimes occur. If irreversible adsorption does not occur some proteins can still be permanently denatured. In either case, these effects lead to low recoveries or insufficient purification (mixture of active and denatured protein). Insufficient purification is usually denoted by the presence of very broad peaks in the chromatogram.

An example of the effect of protein hydrophobicity in RPC[12] is seen in Figure 5. In this figure the chromatograms of three different proteins of increasing hydrophobicity from (a) to (c) are presented. These results were obtained using 5-μm C_4-silica particles in a 250×4.6-mm internal diameter (ID) column. The least hydrophobic protein, ribonuclease A, is shown in Figure 5(a). This chromatogram is described as being ''good''. This is because of the symmetrical shape of the peak and a recovery that exceeds 80%. Figure 5(b), which is labeled ''bad'', is for alcohol dehydrogenase. This peak may indicate conformational changes and is labeled ''bad'' because in this application wide protein peaks in RPC are usually the result of conformational changes. Figure 5(c) is β-galactosidase. The peak shape for this protein is labeled ''ugly'' because it is extremely broad with multiple bands present.

The hydrophobicity of 33 example proteins and peptides[12] is presented in Table 2. This table indicates the appearance of a relationship between the hydrophobicity of a protein and the quality of the RPC chromatogram. Some exceptions to this general relationship are also apparent. For example carbonic anhydrase (hydrophobicity = 129) yields a good chromatogram while β-lactoglobulin (hydrophobicity = 127) yields a bad chromatogram. Perhaps an interesting experiment would be to rank the peak shapes against their heats of adsorption. This type of experiment could possibly explain these inconsistencies.

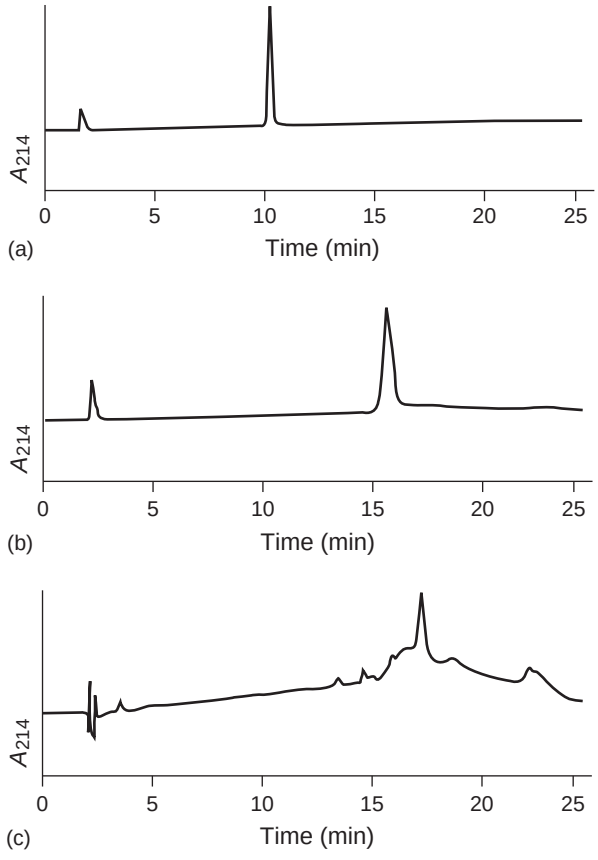

Figure 5 Peak shape as a function of protein hydrophobicity. (Reprinted from W.G. Burton K.D. Nugent, T.K. Slattery, B.R. Summers, 'Separation of Proteins by Reversed-phase High-performance Liquid Chromatography: 1. Optimizing the Column', *J. Chromatogr.*, **443**, 363–379, 1988,[12] Copyright 1988, with permission from Elsevier Science.)

A good example of the ability RPC has in resolving polypeptide mixtures is shown in Figure 6.[13] One of the objectives of this work was to determine if reversed-phase high-performance liquid chromatography (RPHPLC) could be used in identifying polypeptide variants in hemoglobin. Hemoglobin has six polypeptide chains. The chromatograph of normal polypeptide chains in hemoglobin is shown in Figure 6. Figure 7 shows the chromatograms of individuals with variant β-chains. This example plus others reported by Huisman[13] indicate the utility and reliability of RPHPLC. This method cannot however be used to predict which mutant is present although it can nonetheless indicate the presence of an abnormality.

Another example of the ability of RPHPLC to resolve polypeptide mixtures is shown in Figure 8,[14] for the separation of human growth hormone variants. The results here demonstrate the previously mentioned effect of solubility on resolution. Each variant (a, b and c) had slightly different solubility in the mobile phase solvent. The difference, although very slight, was significant enough to be amplified by RPC.

The separation of five peptides ranging in size from 2 to 14 amino acid residues is shown in Figure 9.[15] Unlike the previous two examples these peptides were eluted under isocratic conditions using a phosphoric acid buffer at pH 3. Isocratic conditions indicate that the mobile phase composition is constant. Under isocratic conditions the separation is dependent upon the difference in the distribution coefficients for each analyte. Additional examples of protein and peptide purification with RPC can be found in the work of several authors.[16–25] Additional information and references discussing the use of RPC in purifying proteins and peptides can be found in Aguilar and Hearn.[26]

3.2 Hydrophobic Interaction Chromatography

HIC has become a popular technique used in the purification of proteins. HIC separates sample components by exploiting differences in hydrophobicity. This is the same principle employed by RPC; however, there are some significant differences. The first difference is that HIC usually takes place in the presence of salt. Solvents such as organic modifiers are not commonly used. The second difference is that the stationary phases in HIC are not as hydrophobic as in RPC. The ligand density is lower[27,28] and the molecular structures of HIC ligands are slightly different.[5,9] Since the hydrophobicity of the stationary phase is lower, the risk of protein denaturation is also lower. This reduction in hydrophobicity also reduces (not eliminates) the affinity of the protein for the stationary phase. This phenomenon has been in evidence in studies focusing on the energetics of adsorption.[29,30] In these experiments it was shown that the ΔH value associated with the adsorption of proteins on reversed-phase supports is exothermic. This is indicative of the fact that proteins, in particular the more hydrophobic proteins, will have stronger affinities for the stationary phase. This also explains why certain proteins irreversibly adsorb to reversed-phase supports. It is simply more energetically favorable for the protein to remain on the surface than to become solubilized in the mobile phase solvent. Similar studies were also conducted on hydrophobic stationary phases.[31–33] The results revealed that in this case of HIC protein adsorption was entropically driven. The reported ΔH values associated with adsorption were endothermic. From the standpoint of energetics, HIC appears to be a milder process. This milder nature may be the underlying reason why HIC is generally better suited to protein purification than RPC.

In HIC, protein elutions are usually accomplished through linear salt concentration gradients. The presence of salt has a great influence on the equilibrium behavior

Table 2 Hydrophobicity ranking of 33 proteins: characteristics and classification of protein standards[a]

Protein	Abbreviation	Subunit mol. wt.	No. of subunits	Hydrophobic[c]	pI	Initial class[d]
Insulin[b]	INS	5700	1	95	5.3	G
Aprotinin[b]	APR	6500	1	66		B
Cytochrome c	CYC	12 500	1	92	9.2	G
Ribonuclease A[b]	RNS	13 500	1	75	8.8	G
α-Lactalbumin	ALA	14 200	1	114		G
Lysozyme[b]	LYS	14 300	1	100	11.0	G
Hemoglobin	HEM	17 100	4	123		B
Myoglobin	MYG	17 400	1	122	7.1	B
β-Lactoglobin	BLA	18 400	1	127	5.8	B
Trypsin inhibitor	TIN	20 500	1	119		G
Ferritin	FER	21 000	22	144	4.3	U[e]
Human growth hormone[b]	HGH	21 500	1	160	4.8	B
α-Chymotrypsin	CTN	21 600	1	120	8.4	B
Papain	PAP	22 000	1	117		B
α-Chymotrypsinogen	CTA	25 000	1	123	8.8	B
Carbonic anhydrase[b]	CAH	29 000	1	129		G
Lactic dehydrogenase	LDH	35 000	4	145	4.9	U
Alcohol dehydrogenase	ADH	37 000	4	131		B
Glycerol-3P-dehydrogenase	GPD	39 000	4	121	6.3	B
Ovalbumin[b]	OVA	45 000	1	150	4.7	B
β-Amylase	BAM	50 000	4	139		B
Human glycoprotein[b]	HGP	50 000	3	109	4.7	B
β-Glucosidase	BGU	65 000	2	130		U
Bovine serum albumin	BSA	68 000	1	136	5.0	U
Lactoperoxidase[b]	LPO	85 000	1	139	9.2	U
Jack bean urease	JBU	92 000	6	134		U[e]
Amyloglucosidase[b]	AGS	97 000	1	155		U
Phosphorylase B	PHB	97 000	1	146	6.3	U
β-Galactosidase	BGL	115 000	5	144	5.1	U
Collagen	COL	120 000	1	78		B
Immuno γ-globulins	IGG	160 000	1	116		U[e]
Thyroglobulin	THY	335 000	2	153		U[e]
Fibrinogen	FIB	340 000	1	120		B

[a] Reprinted from W.G. Burton et al., 'Separation of Proteins by Reversed-phase High-performance Liquid Chromatography: 1. Optimizing the Column', *J. Chromatogr.*, **443**, 363–379, 1988,[(12)] Copyright 1988, with permission from Elsevier Science.
[b] Standard used in indicator test mixture.
[c] Scale of protein hydrophobicity is based on retention times relative to lysozyme (100).
[d] G = "good", B = "bad" and U = "ugly".
[e] Protein which did not exhibit ideal behavior under final conditions (but now only B rather than U). Slight modifications of these conditions make them behave in an ideal fashion.

in HIC. It has been postulated that the equilibrium distribution in HIC is a function of the preferential interactions between the protein, the salt and the water.[(34)] The free energy of a dissolved protein will increase when a salt such as NaCl is added to the solution. As a result of the presence of salt the protein will become preferentially hydrated; hydration excludes salt from the immediate vicinity of the protein. As the salt concentration is increased, the interactions between the protein and salt will also increase. This is because at higher salt concentrations the salt molecules displace some of the water molecules surrounding the protein. The protein, in an effort to reduce its free energy, will adsorb to the surface. The free energy of the protein is effectively reduced because the amount of protein surface area exposed to the salt has been minimized.

The selection of salt is an important decision in HIC separations. This is because the salt will alter the surface tension of water. If the surface tension of water is too high the protein will precipitate out. Once again the underlying reason for this is energetics. The free energy of the protein in the solid phase is lower because the protein–salt interactions are minimized when the protein has precipitated out of solution. A list of the common salts used in HIC[(35)] with their effects on the surface tension of water is presented in Table 3.

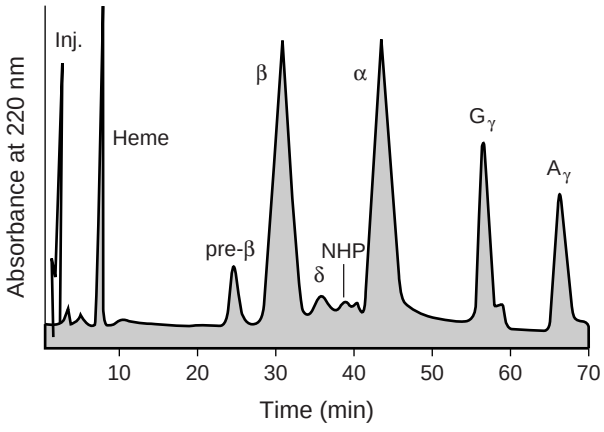

Figure 6 Chromatogram of normal polypeptide chains in hemoglobin. Column dimensions: 300 × 4.6-mm ID. Column stationary phase: Vydac C4 RP (Separations Group, Hesperia, CA, USA); Developer A: 80% water, 20% acetonitrile (CAN) and 0.1% TFA. Developer B: 40% water, 60% CAN and 0.1% TFA. Elution was accomplished by reducing developer A from 51% to 40% over 50 min. This step was followed by a further reduction in developer A from 40% to 20% over 40 min. The column was then purged with developer B for 10 min and equilibrated with 51% developer A. The flow rate was 1.0 mL min^{-1}. (Reprinted from T.H.J. Huisman, 'High Performance Liquid Chromatographic Analysis of Human Hemoglobins and Their Polypeptide Chains: Its Use in the Identification of Variants', *Anal. Chim. Acta*, **352**, 187–200, 1997,[13] Copyright 1997, with permission from Elsevier Science.)

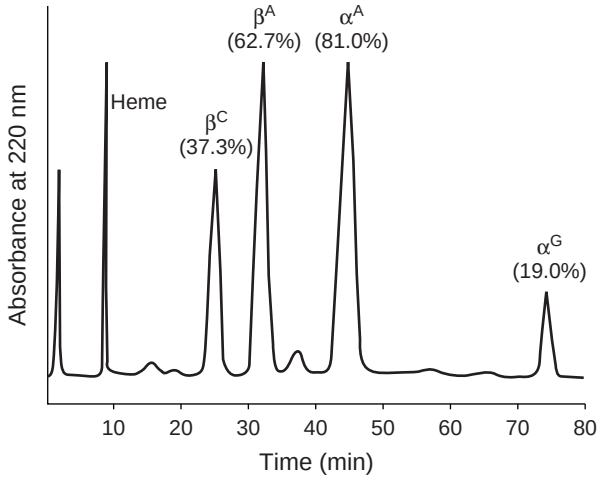

Figure 7 Chromatogram of variant polypeptide chains in hemoglobin. Column dimensions: 300 × 4.6-mm ID. Column stationary phase: Vydac C4 RP (Separations Group, Hesperia, CA, USA); Developer A: 80% water, 20% acetonitrile (CAN) and 0.1% TFA. Developer B: 40% water, 60% CAN and 0.1% TFA. Elution was accomplished by reducing developer A from 51% to 40% over 50 min. This step was followed by a further reduction in developer A from 40% to 20% over 40 min. The column was then purged with developer B for 10 min and equilibrated with 51% developer A. The flow rate was 1.0 mL min^{-1}. (Reprinted from T.H.J. Huisman, 'High Performance Liquid Chromatographic Analysis of Human Hemoglobins and Their Polypeptide Chains: Its Use in the Identification of Variants', *Anal. Chim. Acta*, **352**, 187–200, 1997,[13] Copyright 1997, with permission from Elsevier Science.)

The selection of the stationary phase is also another very important consideration. Although HIC is milder than RPC, protein conformation changes are still possible.[36] With this in mind, the system designer must carefully analyze the peak shapes of the purified analyte in question. If split peaks are present then a conformational change has occurred. If broad peaks are present then the possibility of a conformational change exists. The word ''possibility'' is used because broad peaks can also arise from excessive protein interactions with the stationary phase. If the desired protein has undergone a conformational change then the resulting effect will be a lower recovery. Examples of commercially available HIC stationary phases[5] are presented in Figure 10. Additional examples can be found in Neue.[9]

As previously stated, the speed and resolution of HIC are comparable with IEC. HIC is also advantageous because of its ability to remove endotoxins, nucleic acids and viruses from certain samples.[5] Also, as previously mentioned, the risk of denaturation is much smaller than RPC, and the use of harsh solvents for the purposes of elution are generally not required for most HIC applications.

A number of models describing HIC have also been published.[37,38] One such model is the preferential

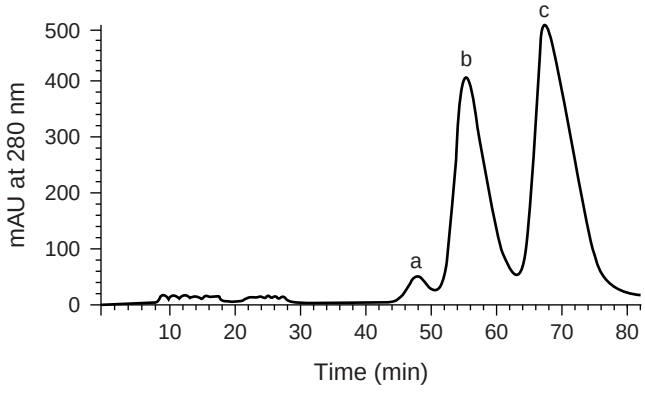

Figure 8 Chromatogram of human growth hormone variants. Column dimensions: 300 × 7.5-mm ID; column stationary phase: Polymer Labs RP-S 8-μm particles with a pore size of 300 Å; column temperature: 40 °C; mobile phase: 25 mM ammonium acetate, pH 7.5; elution: linear gradient containing 34–39% 1-propanol over 100 min; flow rate: 1.0 mL min^{-1}. (Reprinted from G. Teshima, E. Canova-Davis, 'Separation of Oxidized Human Growth Hormone Variants by Reversed-phase High-performance Liquid Chromatography: Effect of Mobile Phase pH and Organic Modifier', *J. Chromatogr.*, **625**, 207–215, 1992,[14] Copyright 1992 with permission from Elsevier Science.)

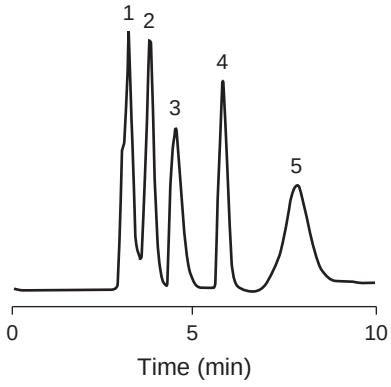

Figure 9 Separation of five polypeptides. The separation of five peptides ranging in size from 2 to 14 amino acid residues is shown. Column dimensions: 250 × 4.6-mm ID; column stationary phase: Unisphere-PBD (polybutadiene) 8-µm particles with a pore size of 250 Å; mobile phase: 0.02-M K_2PO_4 at pH 3.0 and acetonitrile (55 : 45); flow rate was 1.0 mL min^{-1}. Elution order: 1, 1-tyrosil-1-phenylalanine; 2, methionine enkephalin; 3, DL-alanyl-DL-phenylalanine; 4, somatostatin; 5, bombesin. (Reprinted from P.T. Jedrzejewski, L.T. Taylor, 'Comparison of Silica-, Alumina-, and Polymer-based Stationary Phases for Reversed-phase Liquid Chromatography', *J. Chromatogr. Sci.*, **33**, 438–445, August, 1995,[15] Copyright 1995, by permission of Preston Publications, A Division of Preston Industries, Inc.)

Table 3 Effects of various salts on water surface tension[a]

Salt	Molal surface tension increment $\sigma \times 10^3$(dyn g cm^{-1}mol^{-1})
K_3-citrate	3.12
Na_2SO_4	2.73
K_2SO_4	2.58
$(NH_4)_2SO_4$	2.16
Na_2HPO_4	2.02
K_2-tartrate	1.96
NaCl	1.64
$KClO_4$	1.40
NH_4Cl	1.39
NaBr	1.32
$NaNO_3$	1.06
$NaClO_3$	0.55

[a] Reproduced by courtesy of Marcel Dekker Inc., from R.E. Shansky, S.-L. Wu, A. Figueroa and B.L. Karger, 'Hydrophobic-interaction Chromatography of Biopolymers' in *HPLC of Biomolecules: Methods and Applications*, 1990.[35]

interaction analysis model.[38] A summary of this model is presented in this section; however, a more thorough discussion can be found in Perkins et al.[38] This model focuses on the influence of salt on the equilibrium constant. Preferential interaction analysis has been used to study protein stabilization and precipitation. If the adsorption process in HIC is described by the stoichiometric Equation (3):

$$p\mathrm{P} + s\mathrm{S} \Longleftrightarrow c\mathrm{C} \tag{3}$$

where P is the protein, S is the surface site and C is the protein–ligand complex, and the variables p, s and c represent the number of water molecules associated with each entity in the equation, then the equilibrium constant for this process can be formulated as Equation (4):

$$\ln K_{\mathrm{obs}} = c\ln(m_{\mathrm{C}}) - p\ln(m_{\mathrm{P}}) - s\ln(m_{\mathrm{S}}) \tag{4}$$

where m represents the molal quantities of P, S and C. From the preferential interaction model, the equilibrium can be defined by Equation (5):

$$\left(\frac{\partial \ln(k')}{\partial \ln(m_3)}\right)_{T,P} = \frac{\Delta v_+ + \Delta v_-}{g} - \frac{n\Delta v_1}{m_1 g} m_3 \tag{5}$$

where Δv_1 is the net displacement of water molecules, $(\Delta v_+ + \Delta v_-)$ is the net change in modulator ions, m_3 is the modulator concentration, m_1 is the water concentration, n is the total number of cations and anions per formula unit, and g is the change in the $\ln(m_3)$ per change in the $\ln(a_3)$, where a is the activity of the modulator. Integration of Equation (5) leads to Equation (6):

$$\ln(k') = c + \left(\frac{\Delta v_+ + \Delta v_-}{g} - \frac{n\Delta v_1}{m_1 g}\right) m_3 \tag{6}$$

where c is an integration constant. From this model the number of displaced water molecules can be calculated. From these data an estimate of the required area for one protein molecule to adsorb on a given surface at a given modulator concentration can be made. This may provide a potential insight into determining the

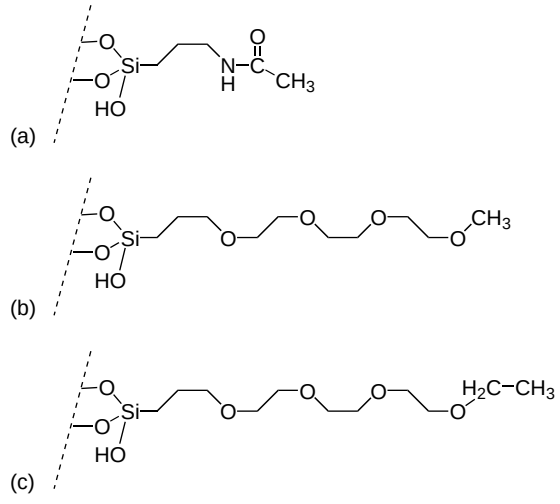

Figure 10 Examples of HIC stationary phases. (a) Acetamide bonded phase, (b) methylether bonded phase, (c) ethylether bonded phase. (Reproduced by permission of Wiley-VCH, a subsidiary of John Wiley & Sons, Inc., from Uwe D. Neue, *HPLC Columns*, 1997,[9] Copyright Wiley-VCH, 1997.)

degree to which conformation changes are occurring. Once these data have been collected under certain conditions, they can then be used to estimate retention and changes in retention with respect to changes in modulator concentration.

An example of the resolving power of HIC is presented in Figure 11,[39] which shows the separation of two RNase variants. These variants were produced by deamination at position 67 of RNase. One derivative contained ASP (aspartyl) and the other iso-ASP. As seen in Figure 11, these molecules were eluted with a linear gradient that effectively lowered the salt concentration. As previously stated, lowering the salt concentration will lower the surface tension of the water. The surface tension is eventually lowered to a point where it becomes energetically favorable for the protein to interact with the water molecules rather than the surface.

Another example of protein purification using HIC is presented in Figure 12.[40] These results were collected on an HIC monolith column. The advantage of a monolith is that the contact surface area of the stationary phase is significantly increased. Figure 12 shows that all of the components were resolved in less than 4 min. This separation was carried out in a 50 × 8-mm ID column. The effect of salt on the system is the same as described in the previous example.

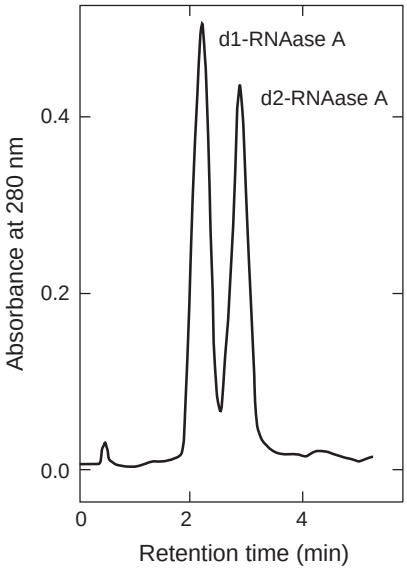

Figure 11 Separation of two RNase variants. Column: Spherogel HIC-CAA column (Beckman) equilibrated in 70% buffer A (3 M ammonium sulfate, 0.5 M ammonium acetate, pH 6.0) and 30% buffer B (0.5 M ammonium acetate, pH 6.0). Elution was carried out by increasing the concentration of buffer B from 30 to 70% in 10 min at a flow rate of 1.0 mL min^{-1}. (Reprinted with permission from A. Di Donato, M.A. Ciardiello, M. de Nigris, R. Piccoli, L. Mazzarella, G. D'Alessio, *J. Biol. Chem.*, **268**(7), 4745–4751, 1993,[39] Copyright the American Society for Biochemistry and Molecular Biology.)

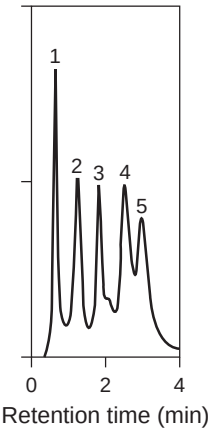

Figure 12 Protein purification using an HIC monolith column. Column dimensions: 50 × 8-mm ID; column stationary phase: rigid porous polyacrylamide monolith; mobile phase: 10% butyl methacrylate; elution: linear gradient that steadily reduced the salt concentration in the mobile phase from 1.5 to 0.1 M ammonium sulfate in 0.01 M sodium phosphate buffer (pH 7) in 3.3 min; flow rate: 3.3 mL min^{-1}. The proteins shown in this chromatogram are cytochrome c (1), ribonuclease (2), carbonic anhydrase (3), lysozyme (4), and chymotrypsinogen (5). (Reprinted from Xie et al., 'Rigid Porous Polyacrylamide-based Monolithic Columns Containing Butyl Methacrylate as a Separation Medium for the Rapid Hydrophobic Interaction Chromatography of Proteins', *J. Chromatogr., A*, **775**, 65–72, 1997,[40] Copyright 1997, with permission from Elsevier Science.)

A third example showing the usefulness of HIC is shown in Figure 13(b). One of the three large peaks shown in Figure 13(b) is purified Mabs.[41] In this example the HIC column was preceded by an ion-exchange column. The chromatogram from the ion-exchange column is shown in Figure 13(a). The shaded area in Figure 13(a) gives the approximate location of the Mabs. Ion-exchange column fractions from 18–24 min were collected and injected into the HIC column. This produced the chromatogram shown in Figure 13(b); the shaded portion indicates the location of the purified Mabs. At the time this experiment was performed, the accepted way of purifying Mabs was by precipitation induced by the presence of ammonium sulfate. The disadvantage of precipitation is that a portion of the protein is normally lost. This experiment demonstrated the use of chromatography as a viable means of purification with minimal losses in protein activity. Additional examples of the resolving power of HIC can be found in the work of several authors.[42–46] Additional references and information on the use of HIC in the purification of proteins can be found in Cooke et al.[28] and Shansky et al.[35]

3.3 Ion-exchange Chromatography

IEC is the most common technique employed for protein purification. IEC exploits differences in protein charge

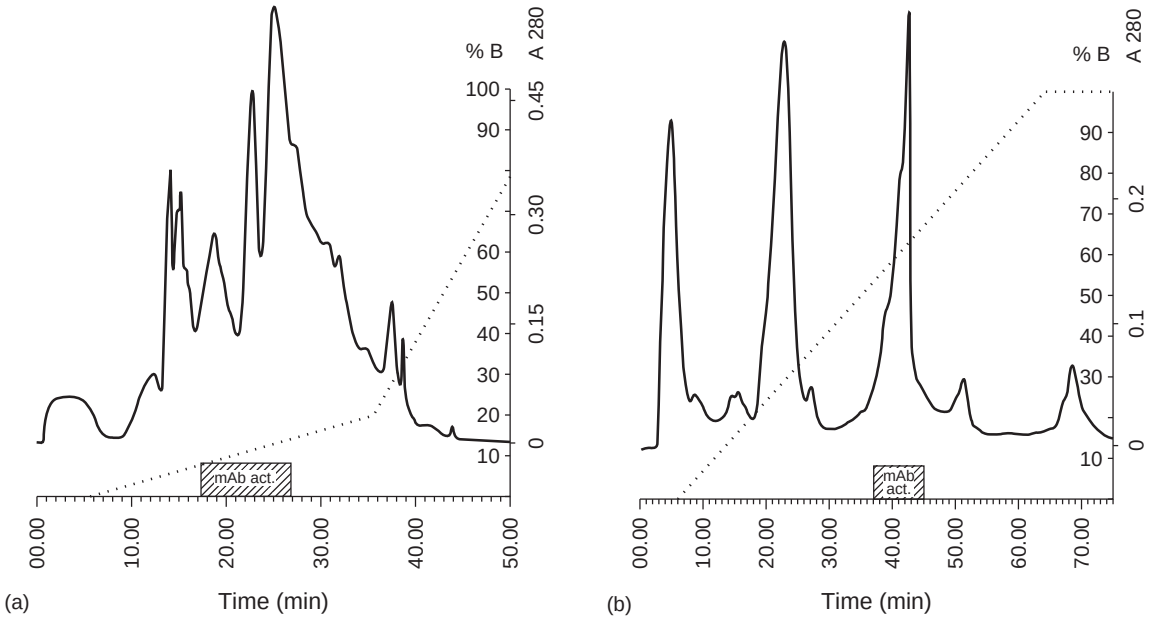

Figure 13 Separation of Mabs using IEC and HIC. The chromatogram of the ion-exchange column is shown in (a). The shaded area in (a) represents the approximate location of the Mabs. An aliquot from the ion-exchange column fractions from 18–24 min was injected into the HIC column producing the chromatogram shown in (b). The shaded area in (b) represents the location of the Mabs. HIC column dimensions: 75 × 7.5-mm ID; HIC stationary phase: TSK gel Phenyl 5 PW. HIC column equilibration: buffered solution containing 0.030 mol L^{-1} Tris-HCl and 1.0 mol L^{-1} sodium sulfate (pH 7.5). Elution was accomplished by reducing the ionic strength and the polarity of the mobile phase with a linear gradient for 80 min. The final salt concentration was 0 mol L^{-1} and the mobile phase contained 5% 2-propanol. The flow rate was 1 mL min^{-1}. The sample load was approximately 0.3 mg of protein. (Reprinted from B. Pavlu, U. Johansson, C. Nyulen, A. Wichman, 'Rapid Purification of Monoclonal Antibodies by High Performance Liquid Chromatography', *J. Chromatogr.*, **359**, 449–460, 1986,[41] with permission from Elsevier Science.)

to purify mixtures. Proteins are essentially very large electrolytes. Adjusting the solution pH to a value that is different from the protein's pI (the pH at which the protein is neutral) can alter the charge of a protein molecule. If the solution pH is above the pI of the protein, the protein will possess a net negative charge. Likewise the protein will possess a net positive charge when the solution pH is less than the pI. In IEC the stationary phase will have the opposite polarity of the desired protein. In general a weakly charged protein sample will require a strongly charged stationary phase to resolve the mixture. Strongly charged protein samples will usually require a weakly charged stationary phase to resolve the mixture. If a strongly charged molecule is brought into contact with a strongly charged stationary phase, the molecule could be irreversibly adsorbed. If the proteins and the stationary phase are too weakly charged, sample resolution will not occur. It should be noted that protein adsorption in IEC can be enthalpically or entropically driven. The ΔH of adsorption depends on the pH, the salt concentration and the stationary phase.[47–50] Thus it is possible to adsorb a protein with a net positive charge on an anion-exchange column, if the conditions selected result in an entropically driven process.

Like HIC, ion-exchange separations take place in the presence of aqueous buffers and salt. Buffers are important because their presence stabilizes the mobile phase pH. The choice of a buffer usually depends on the operating pH. Ideally, the pK_a of the buffer should be as close to the operating pH as possible. Some typical IEC buffers are found in the *Pharmacia Biodirectory*.[51] The presence of salt does however serve a different purpose in IEC. The salt acts as a displacing agent in IEC. For example in the case of an anion-exchange column the salt anions exchange with negatively charged proteins on the surface. In IEC the stationary phase, solution pH, column temperature and salt concentration are key design parameters. Some typical ion-exchange stationary phases are shown in Table 4.[52]

The stoichiometric equation describing the ion-exchange process is shown in Equation (7):

$$Z_n C_p + Z_p N_n \Longleftrightarrow Z_n N_p + Z_p C_n \qquad (7)$$

where C_p and N_p are the protein concentrations in solution and on the surface respectively. C_n and N_n are the respective concentrations of the salt counterion in solution and salt counterion on the surface. Z_p is the effective protein charge, and Z_n is the charge of the salt counterion.

Table 4 Examples of IEC stationary phases

Group	pH range	Structure
Anion exchangers		
Q (Quaternary ammonium)	2–12	$-CH_2-N^+(CH_3)_2-CH_3$
DEAE (Diethylaminoethyl)	2–9	$-O-CH_2-CH_2-NH^+(C_2H_5)_2$
QAE (Quaternary aminoethyl)	2–12	$-O-(CH_2)_2-N^+(C_2H_5)_2-CH_2-CHOH-CH_3$
Cation exchangers		
SP (Sulfopropyl)	2–12	$-(CH_2)_2-CH_2-SO_3^-$
S (Methyl sulfonate)	2–12	$-CH_2-SO_3^-$
CM (Carboxymethyl)	6–11	$-O-CH_2-COO^-$

[a] Reproduced by permission of Humana Press Inc., from D. Sheehan, R. FitzGerald, in *Protein Purification Protocols*, 1996.[52]

A thermodynamically consistent model that has been developed to describe the equilibrium relationship in Equation (7) is the nonideal surface solution (NISS) model.[53] The NISS equilibrium constant is defined by Equation (8):

$$K = \frac{a_p^s}{a_p} \left(\frac{a_n}{a_n^s}\right)^{\frac{z_p}{z_n}} \tag{8}$$

where a_p^s and a_p are the respective activities of the protein on the surface and in solution. a_n^s and a_n are the respective activities of the modulator on the surface and in solution. The activity of each species on the surface is defined as the product of $\gamma_i x_i$, where i represents either the protein or counterion on the surface, γ is the activity coefficient and x is the mole fraction. The activity of the adsorbed species has been calculated using a model developed by Talu and Zweibel.[54] From Equation (8) an expression for the capacity factor was derived and Equation (9) results.

$$\ln k' = \ln[K\lambda\phi] - \frac{z_p}{z_n}[\gamma c_n] \tag{9}$$

where K is the equilibration constant, λ is ion-exchange capacity of the resin, ϕ is the phase ratio, z_p is the protein charge, z_n is charge of the counterion, c_n is modulator concentration, and γ is the salt activity coefficient. If the resin capacity and phase ratio are known then Equation (9) can be used the calculate the equilibrium constant K and the protein charge z_p.

An example of IEC is presented[55] in Figure 14. In this work 11 egg-white proteins are separated by three methods. The known proteins in egg white are shown in Table 5.[55] The conclusions reached from this experiment were that IEC offered a more complete analysis when compared to RPC or gel permeation chromatography.

Another example of the resolving power of IEC is seen in Figure 15,[56] which shows the separation of six isoforms of ovalbumin using displacement chromatography. In displacement chromatography, the solutes are allowed to adsorb to the stationary phase. The solutes are then displaced from the adsorbent by a displacer, which has a higher affinity for the adsorbent than any of the solutes. This separation exemplifies the enhanced resolution potential that displacement chromatography possesses. An elution separation of the same mixture gave only three isoforms, but six were identified with displacement chromatography. Figure 15 illustrates an additional feature of ion-exchange displacement chromatography. All the isoforms are produced at concentrations substantially higher than the feed. The displacement process leads to concentration as well as separation of the sample components. Additional examples of IEC can be found.[57–66]

3.4 Affinity Chromatography

An excellent review of AC is presented by Subramanian[67] and highlights are presented in this section. AC is by far the most specific method that can be employed for the isolation of protein molecules. In RPC, HIC and IEC the adsorption process is consistent with the principles of physisorption. The interactions are not necessarily site specific, the heats of adsorption

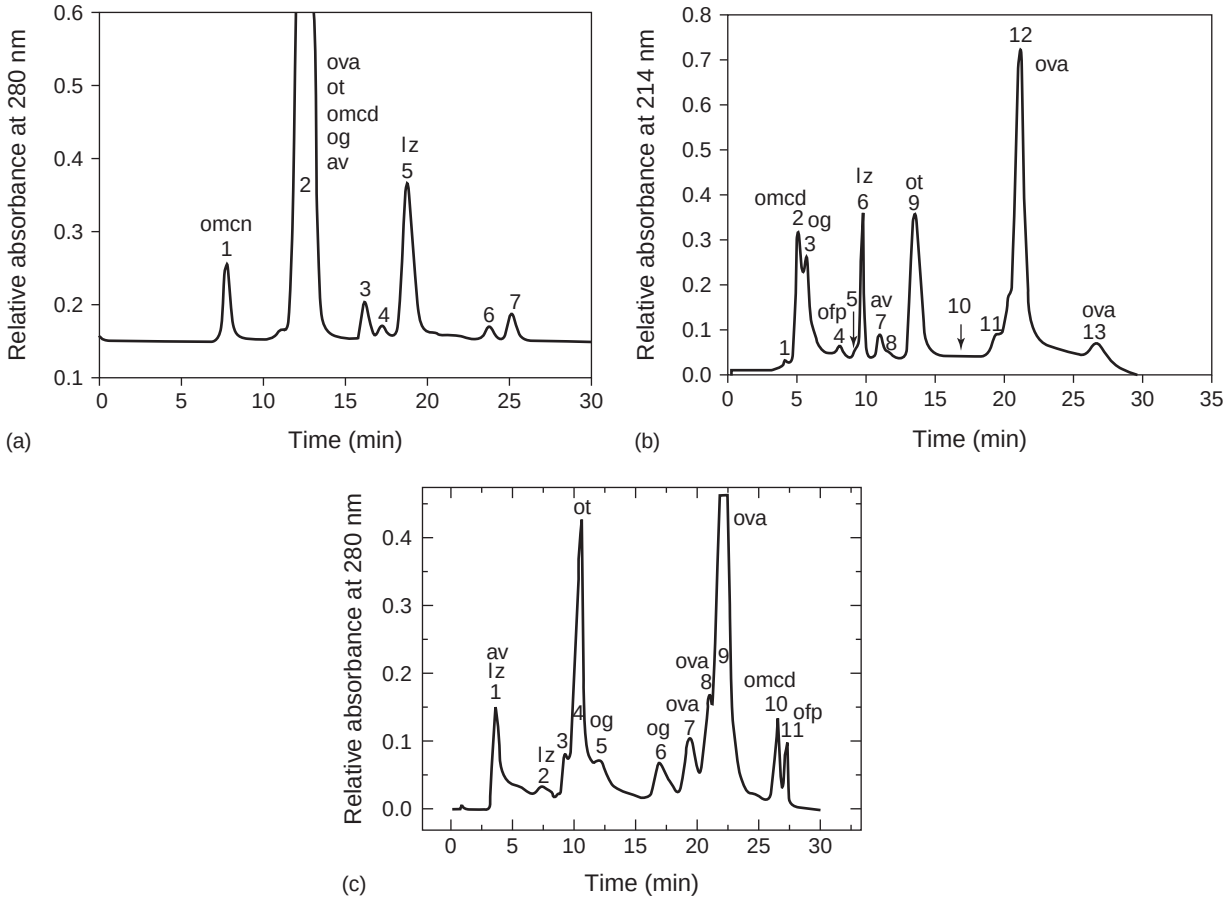

Figure 14 Separation of hen egg-white proteins by (a) gel permeation chromatography, (b) reversed-phase chromatography and (c) ion-exchange chromatography. In this application the IEC column was reported to produce the best resolution. IEC column dimensions: 5 × 0.5-cm ID; IEC column stationary phase: Mono Q HR 5/5. The IEC column was equilibrated with 0.02 M Tris-HCl (pH 9.0). Elution of the protein from the IEC column was accomplished by a stepwise gradient that increased the concentration of NaCl. The flow rate was 1.0 mL min^{-1}. (Reprinted from A.C. Awade, T. Efstathiou, 'Comparison of Three Liquid Chromatographic Methods for Egg-white Protein Analysis', *J. Chromatogr., B*, **723**, 69–74, 1999,[55] with permission from Elsevier Science.)

are relatively low, and the process is usually rapid and reversible. In AC the interactions between the analytes and the surface ligands are extremely specific. In a more strict sense the interactions are site specific. This is because the stationary phase contains biospecific ligands that exclusively elicit very specific and limited types of interaction. The stationary phase ligands will only interact with specific groups on the protein molecule. This type of specificity is normally seen in chemisorption processes. Note that chemisorption is usually a slow process that takes place at elevated temperatures. In AC the process is usually rapid and typically occurs at ambient temperatures. This technique is used for the purification of highly valuable proteins. It is of particular value when the protein concentration is very dilute.

The principle underlying AC is shown in Figure 16. This emphasizes the biospecificity of the protein–ligand binding. Dilute protein solutions can be concentrated on the order of a 1000-fold. Interferon has been purified on an immobilized antibody column with a 5000-fold increase in specific activity. A list of the overall advantages and disadvantages of AC is shown in Table 6. Of the disadvantages listed, the slow adsorbent development time can potentially be the most difficult obstacle to overcome. This is due in part to the need for biospecific ligands. The search for biospecific ligands can be time-consuming and expensive. Once a ligand has been identified it must be attached to the support without losing its activity. Since proteins are large macromolecules the ligands must be spaced far enough apart to allow the protein–ligand complex to form. Finally, the stationary phase must be validated in a pharmaceutical plant.

Some examples of AC ligands and their target proteins are presented in Table 7. A critical parameter to consider in the selection of an AC ligand is the dissociation constant

Table 5 List of egg-white proteins[a]

Protein	Amount of total protein (%)	Mol. weight	pI
Ovalbumin	54	45 000	4.5
Ovotransferrin	12–13	77 700	6.0
Ovomucoid	11	28 000	4.1
Lysozyme	3.4–3.5	14 300	10.7
Ovomucin	1.5–3.5	$0.22–270 \times 10^6$	4.5–5.0
G2 ovoglobulin	1.0	47 000	4.9–5.3
G3 ovoglobulin	1.0	50 000	4.8
Ovoflavoprotein	0.8	32 000	4.0
Ovastatin	0.5	$7.6–9.0 \times 10^5$	4.5–4.7
Cystatin	0.05	12 000	5.1
Avidin	0.05	68 300	10.0
Thiamine-binding protein	Not determined	38 000	Not determined
Glutamyl aminopeptidase	Not determined	320 000	4.2
Minor Glycoprotein 1	Not determined	52 000	5.7
Minor Glycoprotein 2	Not determined	52 000	5.7

[a] Reproduced from A.C. Awade and T. Efstathiou, 'Comparison of Three Liquid Chromatographic Methods for Egg-white Protein Analysis', *J. Chromatogr., B*, **723**, 69–74, 1999,[55] Copyright 1999, with permission from Elsevier Science.

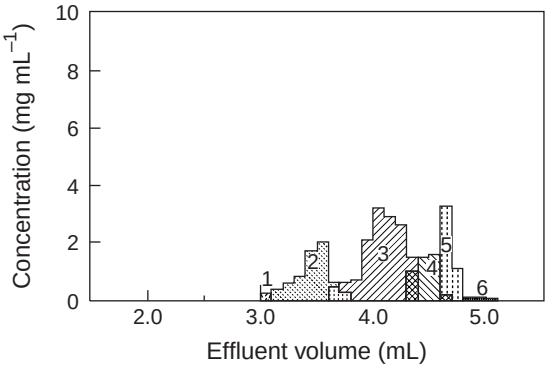

Figure 15 Separation of ovalbumin isoforms. Column dimensions: 250 × 4.6-mm ID. Column stationary phase: PAE 300 (Amicon). Mobile phase: histidine (pH 6.0); flow rate: 0.5 mL min^{-1}; regenerate: 0.1 M sodium citrate in 0.3 M sodium bicarbonate (pH 9); regenerant flow rate: 0.5 mL min^{-1}. (Reprinted from S.C.D. Jen, N.G. Pinto, 'Use of the Sodium Salt of Poly(vinylsulfonic acid) as a Low-molecular-weight Displacer for Protein Separations by Ion-exchange Displacement Chromatography', *J. Chromatogr.*, **519**, 87–98, 1990,[56] with permission from Elsevier Science.)

K_d, Equation (10)

$$K_d = \frac{[PL]}{[P][L]} \tag{10}$$

where PL is the protein–ligand complex concentration, P is the protein concentration and L is the ligand concentration.

The risks of irreversible protein adsorption and denaturation are minimized when the value of K_d is between 10^4 and 10^6. At higher K_d values the risk of protein damage increases. At lower K_d values the interaction between the protein and surface ligands may not be strong enough to retain the protein.

Once a protein is sufficiently bound to the stationary phase the column is then washed to remove excess debris. After washing, the protein is eluted from the column. Protein elution can be accomplished by varying pH, modulator concentration or possibly the concentration of a specific displacing cofactor or substrate. The latter two options can be cost prohibitive. If these options fail, then chaotropic agents such as urea or ethylene glycol can be used to desorb and elute the protein. It must be noted, however, that these chaotropic agents can be harsh and as a result some proteins can become denaturated.

An example of AC is shown in Figure 17,[68] which shows the purification of different types of blood cells. The stationary phase used was PEG 20M-sepharose 6B. The essential chemical structure of the stationary phase is shown in Figure 18. In all of the cases the adsorption of blood cells onto the stationary phase was induced by the presence of salt. This suggests that hydrophobic interactions may have played a key role in adsorption process. Additional information on AC can be found in Narayanan.[69] Additional examples of the resolving power of AC are given by various authors.[70–73]

3.5 Size-exclusion Chromatography

SEC, sometimes referred to as gel permeation, isolates proteins by using molecular characteristics such as size and/or shape. Molecules with similar molecular weight can be separated by shape or by significant conformational differences. In general the smaller molecules with high diffusivities permeate the matrix of the stationary phase

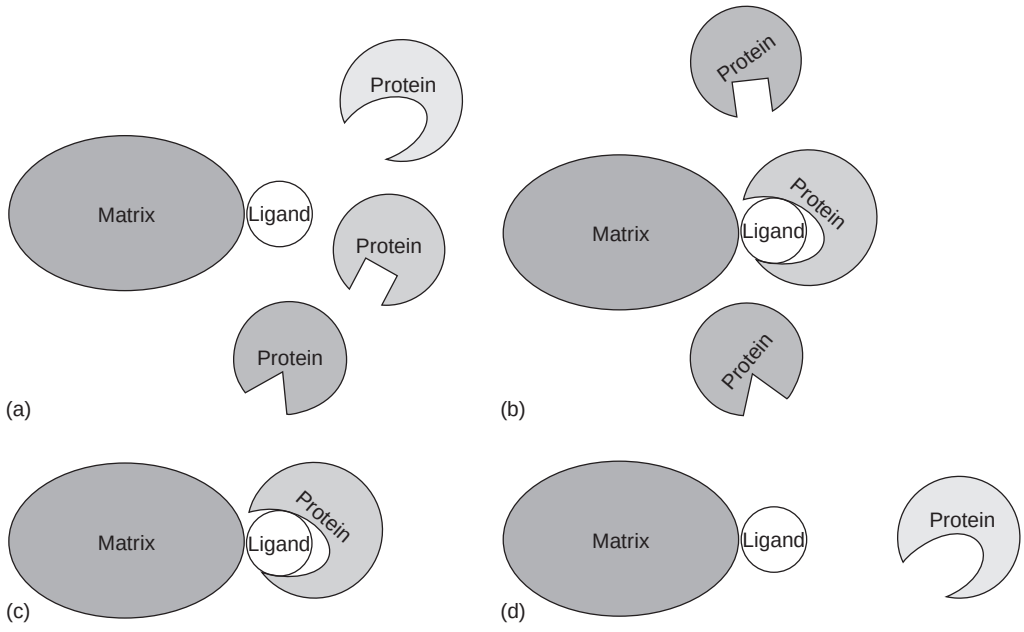

Figure 16 Overview of AC. (a) Sample is applied to the adsorbent, (b) target protein is bound to the ligand, (c) other proteins are washed away, (d) target protein is eluted from the adsorbent. (Reproduced by permission of VCH Publishers from G. Subramanian, *Process Scale Liquid Chromatography*, 1995.[67])

Table 6 Advantages and disadvantages of AC[a]

Advantages	Disadvantages
Overall process time may be rapid	Ligand may be expensive
Overall number of steps in a process are often reduced	Development of an adsorbent may be very slow
Capacity of the adsorbent can be high or low	Regeneration of an adsorbent can be difficult
Process is not limited be volume of sample	
Resolution can be very high	

[a] Reproduced by permission of Wiley-VCH Publishers from G. Subramanian, *Process Scale Liquid Chromatography*, 1995.[67]

Table 7 Typical AC ligands[a]

Ligand	Target protein
Protein G	Immunoglobulins
Protein A	Immunoglobulins
Protein L	Immunoblobulins
Antigen	Antibodies
Antibody	Antigen
Hormone	Receptor
Nucleotides	Nucleic acid binding proteins
Protease inhibitors	Proteases
Cofactor	Enzyme
Substrate	Enzyme
Triazine dyes	Dehydrogenases, kinases, polymerases
Streptavidin/avidin	Biotin-labeled proteins
Phenyl boronate	Glycoproteins
Lectins	Glycoproteins
Heparin	Fibronectin

[a] Reproduced by permission of Wiley-VCH Publishers from G. Subramanian, *Process Scale Liquid Chromatography*, 1995.[67]

while the larger molecules with lower diffusivities tend to be carried through the column by the mobile phase. An overview of this mechanism[74] is shown in Figure 19. This technique is primarily used for desalting, buffer exchange and purification; however, it has also been used to isolate protein conformational changes that have been induced by a harsh solvent or by an extremely hydrophobic stationary phase. GFC can also be used to determine the molecular weight distribution of synthesized molecules.

SEC stationary phases are usually gels.[67] The gel matrix can be composed of individual polymer chains, aggregated polymer chains or combinations of both. Examples of the individual polymer gels would be the dextran and polyacrylamide gels. These gels have a protein exclusion limit of approximately 1×10^6 for globular proteins. If these gels are loosely cross-linked they tend to be mechanically weak. High-pressure applications could destroy the gel. For preparative applications the more tightly cross-linked gels provide the mechanical strength required for process applications. Agarose gels are an example of an aggregated polymer. Agarose has a high degree of cross-linking which gives it rigidity and mechanical strength. The composite gels are denoted by

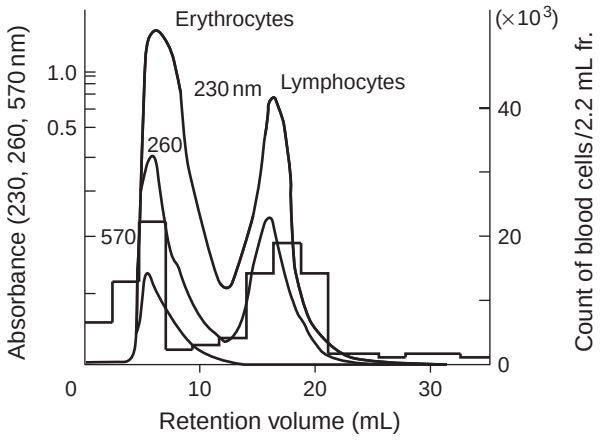

Figure 17 Separation of blood cells. Column dimensions: 25 × 0.9-cm ID; column stationary phase: polyethylene glycol (PEG) 20M-sepharose 6B; eluent: 8% (w/w) dextran T40; mobile phase: 0.09 M sodium phosphate buffer at pH 7.4; flow rate: 4.4 mL h^{-1}. (Reprinted from Y. Shibusawa, 'Surface Affinity Chromatography of Human Peripheral Blood Cells', *J. Chromatogr., B*, **722**, 71–88, 1999,[68] with permission of Elsevier Science.)

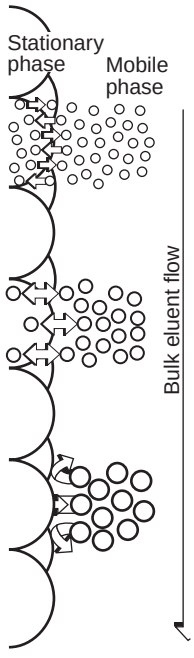

Figure 19 Overview of SEC. In SEC smaller molecules with high diffusivities diffuse through the stationary phase while larger molecules with lower diffusivities are essentially carried through the column by the mobile phase. (Reproduced by permission of Academic Press, from G.K. Sofer, L.-E. Nystrom, *Process Chromatography: a Practical Guide*, 1989.[74])

their higher pressure limits across a wide fractionation range.

The choice of a stationary phase in SEC is usually driven by two factors. This first of these is the fractionation range of the proteins in question and the process conditions (pressure, sample load, flow rate, ... etc.). An example of a selection chart is shown in Figure 20.[74] Each solid line corresponds to a different stationary phase. Once the molecular weight of each component is known the distribution coefficient can be determined from the plot. The optimal stationary phase will be the one in which the distribution coefficients are significantly different.

An example of SEC is presented in Figure 21(a) and (b),[75] which show the effect of mobile phase flow rates and sample volume on resolution. As previously stated, one of the main uses of SEC is desalting. In all cases the salt was effectively removed from the sample components. This is because the salt molecules have a much higher diffusivity than that of the proteins. Figure 21 shows that desalting was less sensitive to the operating conditions, however, protein resolution was sensitive to increased flow rates and increased sample volumes. Additional information on SEC can be found in review articles cited in Irvine[76] and Barth et al.[77] Examples demonstrating the ability of SEC to purify biomolecules are given by several authors.[77–81]

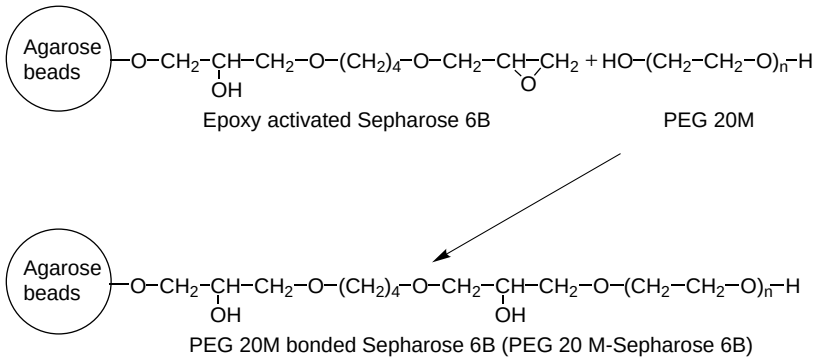

Figure 18 PEG 20M-sepharose 6B. (Reprinted from Y. Shibusawa, 'Surface Affinity Chromatography of Human Peripheral Blood Cells', *J. Chromatogr., B*, **722**, 71–88, 1999,[68] with permission from Elsevier Science.)

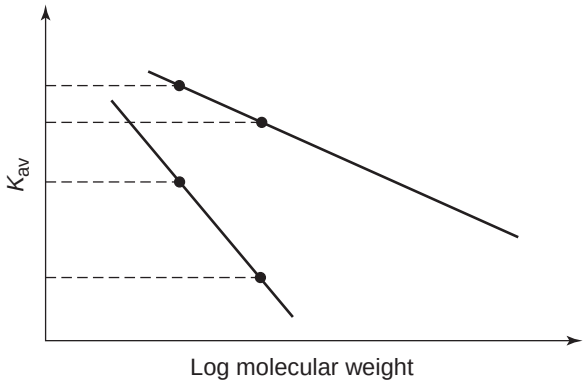

Figure 20 Example of a selection chart for SEC. The ideal stationary phase will be the one in which the distribution coefficients of each of the sample components are significantly different. (Reproduced by permission of Academic Press, from G.K. Sofer, L.-E. Nystrom, *Process Chromatography: a Practical Guide*, 1989.[74])

3.6 Electrophoresis

Electrophoresis was originally developed for the separation of proteins. This technique is used extensively for protein purification on a small scale. Since electrophoresis is primarily limited to small-scale purifications, a brief overview is presented in this section. References are also presented for readers requiring additional information about this technique.

Protein separation takes place in a column with a diameter of 0.5 mm or less and in the presence of an electric field. In a typical application the electric field placed on the column is generated by a negatively charged anode at the inlet and a positively charged cathode at the exit. The degree of separation depends on the charge-to-mass ratio for each protein. The protein molecules with the greatest negative charge will have the highest affinity for the cathode provided that the relative difference in protein molecular weights is small. The positively charged anions will also move toward the cathode because the electroosmotic flow of the buffer is essentially carrying or pushing these molecules through the column. In capillary electrophoresis the mobility of each analyte is a key factor in determining the selectivity. The mobility (μ) of an analyte is related the electric field strength and the velocity by Equation (11):

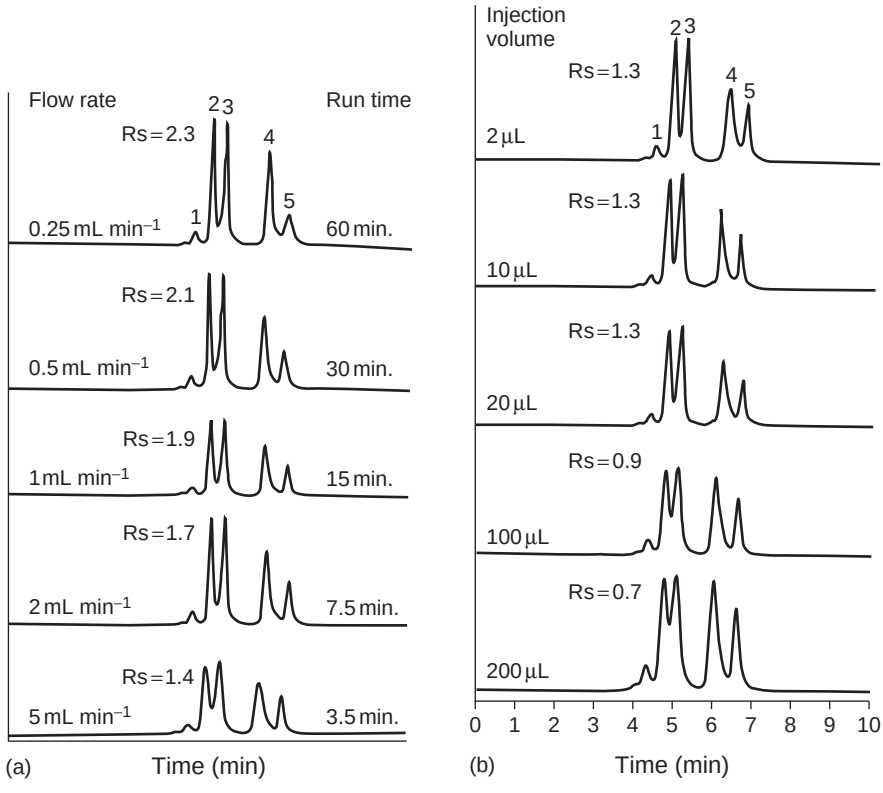

Figure 21 SEC protein fractionation studies. Column dimensions: 250 × 9.4-mm ID; mobile phase of 200 mM sodium phosphate, pH 7.0. The flow rate varied from 0.25 to 5 mL min^{-1} in (a). The separation conditions and column were the same in (b) except that the flow rate was held constant at 2 mL min^{-1} and the injection volume was varied from 2 to 200 μL; 1, bovine serum albumin dimer; 2, bovine serum albumin; 3, ovalbumin; 4, lysozyme; 5, sodium azide. Results obtained by changing (a) flow rate and (b) sample size. (Reprinted from R.D. Ricker, L.A. Sandoval, 'Fast Reproducible Size-exclusion Chromatography of Biological Macromolecules', *J. Chromatogr., A*, **743**, 43–50, 1996,[75] with permission from Elsevier Science.)

$$v = \mu E \tag{11}$$

where v is the molecule velocity ($cm\,s^{-1}$), and E is the electric field strength ($V\,cm^{-1}$), and μ has units of ($cm^2\,s^{-1}\,V^{-1}$). The electric force F_e experienced by a given protein molecule is given by Equation (12):

$$F_e = qE \tag{12}$$

where q is the ion charge. The frictional force F_f for a spherical ion is given by Equation (13):

$$F_f = -6\pi\eta r v \tag{13}$$

where η is the solution viscosity and r is molecular radius. When these forces are balanced, Equation (14):

$$qE = 6\pi\eta r v \tag{14}$$

Solving for v and substituting the result into Equation (13) provides a theoretical prediction for the mobility in measurable terms,[82] Equation (15):

$$\mu = \frac{q}{6\pi\eta r v} \tag{15}$$

In Equation (15) the mobility is directly proportional to the molecular charge and inversely proportional to the molecular size and mobile phase viscosity. Since the net charge is a function of pH, the mobility of an ion is dependent upon the pH of the mobile phase.

A number of alternative versions[83] of electrophoresis are employed when the difference in protein mobilities is small. Isoelectric focusing is a form of electrophoresis that separates proteins by their pI values. A pH gradient is established in the column with a low pH at the inlet and higher pH at the outlet. Capillary gel electrophoresis separates proteins by size. A highly cross-linked polymer is present in the column. The separation is essentially the same as that employed by SEC. Micellar electrokinetic capillary chromatography is another technique commonly used. The method is good for separating proteins with similar charge. A micelle-forming surfactant such as sodium dodecyl sulfate is added to the mobile phase. The more hydrophobic proteins diffuse to the interior while the less hydrophobic proteins stay in the mobile phase. The proteins that diffuse into the micelle have effectively lowered their charge-to-mass ratio, thus reducing their mobilities. The less hydrophobic proteins emerge from the column first while those inside the micelles emerge last.

A thorough review of electrophoresis is presented by the work of several authors.[83–86] Examples of the resolving power of electrophoresis can be found in Jorgensen et al.[87] Rabel and Stobaugh,[88] and Dolink.[89]

4 DETECTION OF COLUMN RESPONSE

Detectors are an essential component of any chromatographic process. The output of a chromatographic detector displays a series of peaks that represent either a product or an impurity. Detectors are used in three different applications. These are manufacturing, design and research. In manufacturing, the detector is used to verify if the desired separation or purification of components was successful. In this application the chromatographer is looking for peak resolution. Good peak resolution means the product peaks should ideally be distinct with little or no overlap with other peaks representing impurities or other products.

The second application is in design. Before a column is specified for production a number of experiments are conducted to determine the proper stationary phase and mobile phase composition. The system designer uses data directly from the detector output to calculate peak resolution and the required number of theoretical plates. The resolution between any two peaks can be calculated by the empirical relationship shown by Equation (16):

$$R_s = \frac{1}{4}\frac{\alpha_1 - 1}{\alpha_1}\frac{k_1'}{1 + k_1'}N_1^{\frac{1}{2}} \tag{16}$$

where R_s is the resolution between component peak 1 and component peak 2, N_1 is the number of theoretical plates required for component 1 and k_1 is the capacity factor for component 1. α_1 is the ratio of k_1/k_2, where k_2 is the capacity factor for component 2. α is commonly referred to as the selectivity. N_1 can be further defined by Equation (17)

$$N_1 = \frac{L}{H} = 5.54\left(\frac{t_R}{W_{\frac{1}{2}}}\right)^2 \tag{17}$$

where L is the column length, H is the height equivalent of one theoretical plate, t_R is the retention time, and $W_{1/2}$ is the peak width in time units at half peak height. These two equations show that resolution is a function of peak shape and retention time. All of these data are taken from the detector output and used to specify the column and the method of achieving the desired separation or purification, thus making the detector a key part of the design phase.

In the third application, the detector output can be used as an investigative tool in analytical applications to explore the interaction between a given analyte and a given stationary phase. The peak shape reflects the nature of the interaction between an analyte and the surface. Broad peaks or peaks with long tails are indicative of strong affinities between the analyte and the stationary phase. They can also be indicative of mass

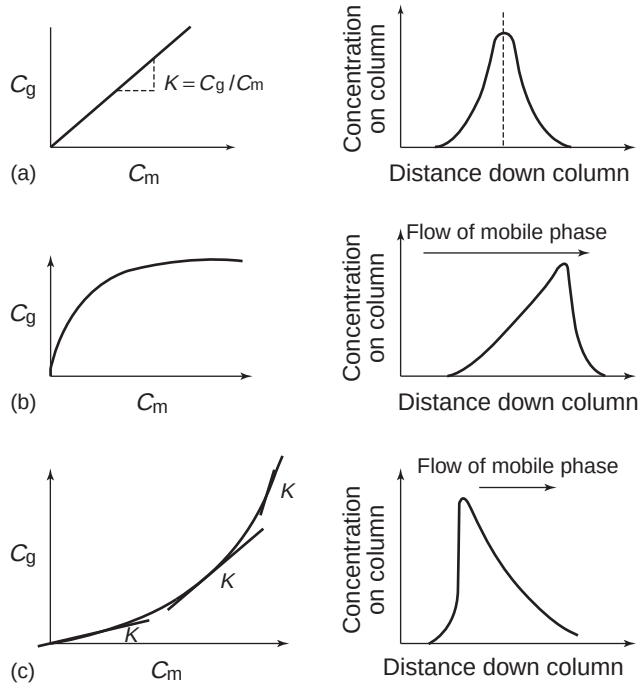

Figure 22 This demonstrates the effect of the isotherm on the component peak measured by the detector. (a) Linear isotherm, (b) Langmuir isotherm, (c) anti-Langmuir isotherm. (Reproduced with permission from P.A. Sewell, D. Kelly, *Chromatographic Separations*, 1987,[90] Copyright, John Wiley & Sons Ltd, 1987.)

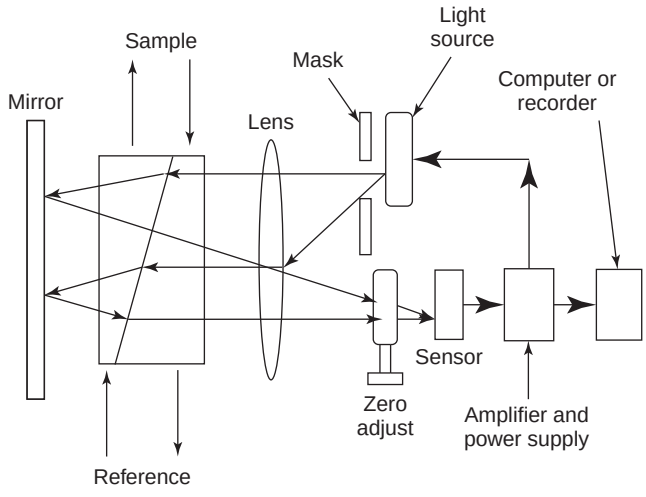

Figure 23 Functional overview of refractive index detector. (Reproduced with permission from E.D. Katz, *High Performance Chromatography: Principles and Methods in Biotechnology*, 1996,[91] Copyright, John Wiley & Sons Ltd, 1996.)

transfer resistance. As previously stated, in RPC broad peaks can also indicate that part of the protein sample has undergone a conformational change. As shown in Figure 22(a) Gaussian-shaped peaks are indicative of linear behavior.[90] Here the distribution coefficient for a given component is independent of concentration. Figure 22(b) demonstrates that peaks with sharp fronts and diffuse tails are indicative of Langmuirian behavior. In this scenario the distribution coefficient of a given component decreases when the mobile phase concentration of that component is increased. These peak shapes are typically seen in overloaded preparative applications. Figure 22(c) shows that peaks with diffuse fronts and sharp backs are indicative of anti-Langmuirian behavior. Here the distribution coefficient will increase when the component concentration in the mobile phase is increased. In general detectors are divided into two categories: bulk property detectors and solute property detectors.[91] A thorough review of these detectors is covered by Katz[91] and highlights are presented in this section.

The bulk property detectors sense bulk property changes in the mobile phase solution that contains the solute. These data are compared to a reference that contains pure mobile phase solution. An example of this type of detector is the refractive index detector. This type of detector essentially measures the refraction of light that is passed through a cell containing the solute and the mobile phase. A schematic of this detector is shown in Figure 23. Light is passed through a lens and then through the sample cells. The photocell sensor measures the refraction of light in both cells and records the difference. This value is linearized and then amplified for the output. Refractive index detectors are widely used, inexpensive and simple. The sensitivity of this detector is of the order of $1 \times 10^{-6}\,\mathrm{g\,mL^{-1}}$. They are, however, highly sensitive to temperature and pressure changes. Also the use of a reference cell prohibits the use of gradients. These restrictions severely limit the use of this type of detector in protein purification methods.

Another type of bulk detector is the electrical conductivity detector. This detector is both sensitive and linear; however, it is limited to detecting the presence of ionic materials such as organic acids, bases, salts, amino acids and peptides. The detector cell has two electrodes that are energized with an AC voltage ranging from 1 to 10 kHz. When ionic analytes pass through the cell the impedance of the system changes. This change in impedance is then linearized to produce an output that is linearly related to the solute concentration. The sensitivity of this detector is of the order of $5 \times 10^{-9}\,\mathrm{g\,mL^{-1}}$.

The second category of detector is the solute property detectors. These detectors sense some chemical or physical property change in the solute. An example of this is the UV (ultraviolet) detector. These detectors pass UV light through the sample cell. The solute absorbs a portion of the light and the remainder of light passes through the cell into a photocell where the signal is measured,

linearized and sent to the output. There are three types of UV detector. The first type is a fixed wavelength detector (Figure 24a). The light source in these detectors emits light at fixed wavelengths. These types of detectors have only limited use in analytical applications because of the presence of a reference cell. As with the refractive index detector, the presence of a reference cell eliminates the use of gradients as a means of eluting products from the column.

The variable wavelength detector is the second type of UV detector (Figure 24b). The light source in these detectors is a deuterium lamp. Deuterium lamps are broad band UV sources that emit UV light at a number of different wavelengths. Light is deflected from two concave mirrors onto a grating. The light is then reflected from the grating onto two concave mirrors in series and then through a plane mirror followed by a lens. The light continues through the flow cell where a portion is absorbed and the remaining unabsorbed light is measured by a photocell. The operating principle is similar to the fixed wavelength detector with two exceptions. First, there is no reference cell, and second the wavelength of the light-emitting source can be changed by rotating the grating. Only one wavelength is passed through the sample, however the ability to select wavelengths allows users to make adjustments for the detection of different products during different runs. Note the major limitation of this detector is that it can only detect sample components that absorb at the wavelength it is configured for. Sample components that absorb light at a wavelength different from the one selected will not be detected.

A third type of UV detector is the diode array detector (Figure 24c). The main difference between this detector and the variable wavelength detector is that multiple wavelengths are passed through the sample and the absorption of multiple wavelengths can be measured simultaneously by the photodiode array. If the sample matrix contains components that absorb light at different wavelengths, detection of each component is possible with a photodiode array. The sensitivity of this detector is of the order of $5 \times 10^{-8}\,\mathrm{g\,mL^{-1}}$.

Another type of solute property detector is the fluorescence detector. Fluorescence detectors are extremely sensitive instruments. They can detect solutes at dilute concentrations of the order of $10^{-9}\,\mathrm{g\,mL^{-1}}$. A diagram of a fluorescence detector is shown in Figure 25. Light from a deuterium source is passed through a series of mirrors and ultimately through the sample. A portion of the light is deflected 90° through the sample cell wall containing rhodamine which acts as a fluorescence substance. The fluorescent light is detected by a photocell. Only one wavelength can be detected. This type of detector is primarily suited to analytical applications where trace analysis is routinely performed. Since samples are rarely fluorescent, they must be derivatized with a fluorescing agent. The higher cost of fluorescence detectors plus

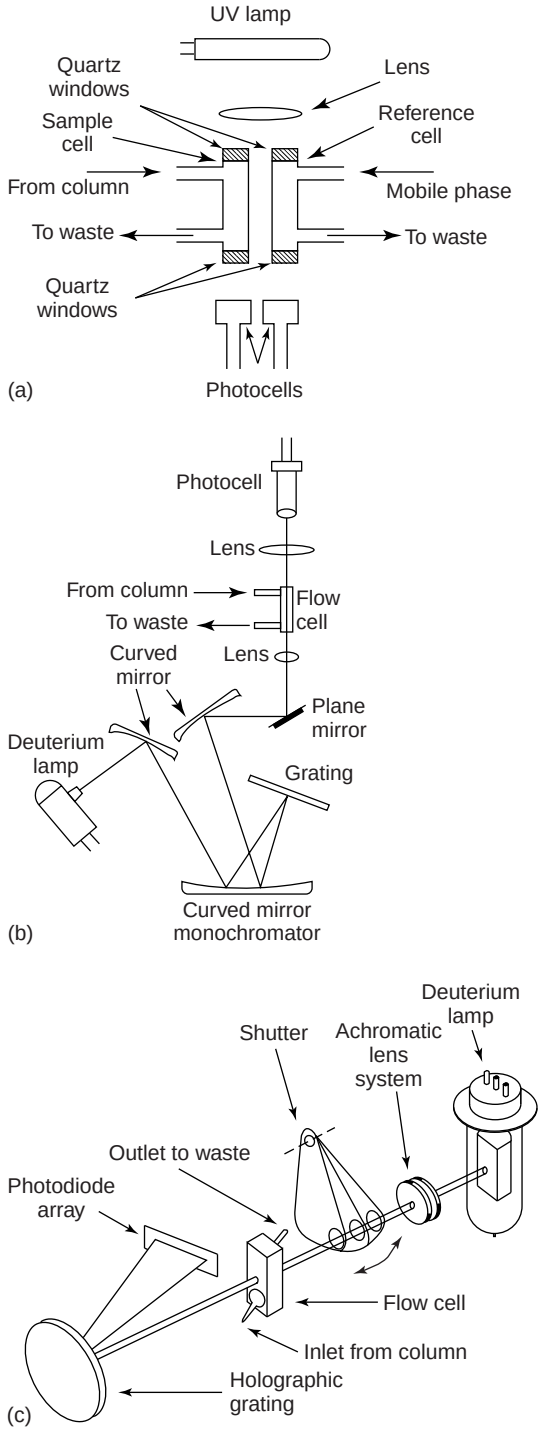

Figure 24 Functional overview of UV detectors. (a) Fixed wavelength detector, (b) variable wavelength detector, (c) diode array detector. (Reproduced with permission from E.D. Katz, *High Performance Chromatography: Principles and Methods in Biotechnology*, 1996,[91] Copyright, John Wiley & Sons Ltd, 1996.)

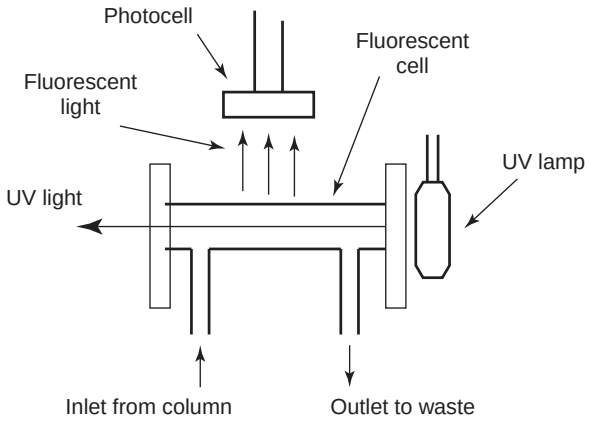

Figure 25 Functional overview of fluorescence detector. (Reproduced with permission from E.D. Katz, *High Performance Chromatography: Principles and Methods in Biotechnology*, 1996,[91] Copyright, John Wiley & Sons Ltd, 1996.)

Table 8 Key parameters used for detector specification

Parameter	Definition
Dynamic range	The concentration range over which a detector will provide a response
Response index or linearity	An empirical factor with a nominal range of 0.98–1.02
Linear dynamic range	The concentration range in which the response index remains between 0.98 and 1.02
Detector response	Detector voltage output per unit change in solute concentration
Detector noise level	Any perturbation on the detector output that is not related to an eluted solute
Detector sensitivity	Minimum solute concentration that provides a signal equivalent to twice the noise level
Total system dispersion	Variance contribution to a peak from the various components of the detector
Pressure sensitivity	Change in detector output per unit change in operating pressure
Flow sensitivity	Change in detector output per unit change in flow rate
Operating temperature range and temperature sensitivity	Acceptable temperature range of surroundings Detector output change per unit temperature change

the additional complications associated with derivatization severely limit the use of these types of detector in preparative protein purification applications. They are, however, used in analytical applications for the analysis of analytes at low concentration or analytes that do not possess a UV chromaphore.

The selection of a detector is as important as the selection of the stationary phase and the mobile phase composition. There are many factors to consider when purchasing a detector.[91] A list of key factors is presented in Table 8 along with a brief description. Other types of detectors such as mass spectroscopy and light scattering are also used in analytical experiments. A review of these types of detectors can be found in Rabel and Stobaugh[88] and Dolink.[89]

5 SCALE-UP

The goal of preparative chromatography is to isolate the desired protein at a purity level that is appropriate for dosing. The objective of scale-up is to reproduce the resolution, analysis time and yield achieved in an analytical scale system. Each of these objectives is influenced by certain parameters in the chromatographic process and therefore a discussion of each is presented in this section.

As previously stated, the resolution is defined by Equation (16). This equation shows the resolution to be a function of the number of theoretical plates, the retention time and the selectivity of the desired component. As seen in Equation (17) the number of theoretical plates is a function of the column length. These equations plainly show that the resolution will increase when the column length is increased. This phenomenon is presented in Figure 26. Since the goal of preparative applications is purification, a longer column appears to be more desirable. Although the longer column in this example produces superior results it is important to note that the analysis time also increases. In analytical applications this may be acceptable because the goal of many analytical separations is to gain knowledge about the composition. On the other hand in preparative applications the analysis time directly affects the operating cost, which in turn affects the product cost. Since the preparative biopharmaceutical market is highly competitive, it is critical that the stationary phase, flow rate, column length, column width and particle size be specified with the underlying goal of producing acceptable resolution while maintaining product throughput.

Particle size and porosity are also important parameters. Smaller porous particles are advantageous because the sample is exposed to more active surface area. Additionally, the diffusion distance is shorter in smaller particles. The enhanced surface area coupled with a shorter diffusion path lays the foundation for a more efficient separation. High efficiency provides the framework

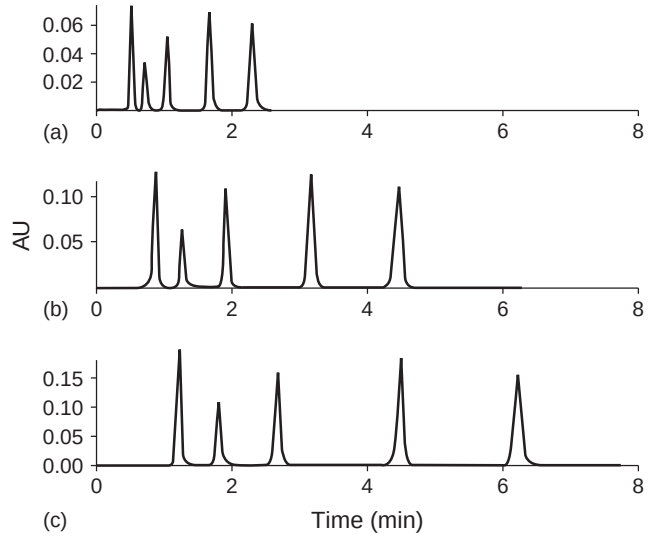

Figure 26 Effect of column length on resolution. This figure shows the comparison of the performance of 5-μm columns of 5, 10 and 15-cm lengths. (a) 5-cm column length, (b) 10-cm column length, (c) 15-cm column length. The data from each run were collected at the same flow rate. (Reproduced by permission of Wiley-VCH, a subsidiary of John Wiley & Sons, Inc., from Uwe D. Neue, *HPLC Columns*, 1997,[9] Copyright Wiley-VCH, 1997.)

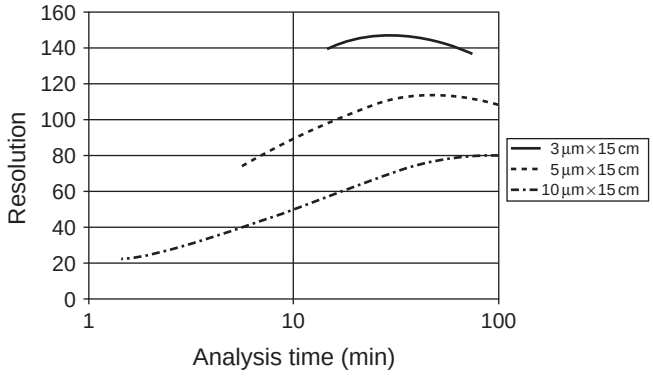

Figure 27 Effect of particle size on resolution. Comparison of the performance of columns of constant length but variable particle size. ———, 3 μm × 15 cm; - - - - -, 5 μm × 15 cm; — · — · —, 10 μm × 15 cm. (Reproduced by permission of Wiley-VCH, a subsidiary of John Wiley & Sons, Inc., from Uwe D. Neue, *HPLC Columns*, 1997,[9] Copyright Wiley-VCH, 1997.)

for acceptable resolution. An example of this is presented in Figure 27.[9] Particle type (IEC, RPC, HIC, etc.) and particle size have the largest influence on yield and resolution. There is however a trade-off with smaller particles. As the particle size decreases the pressure drop increases, which results in higher pumping cost. In addition to a higher pressure drop, the friction generated by flow through a tightly packed bed increases the temperature within the column. Increases in temperature alter the equilibrium in such a way that the overall adsorption of the analyte is reduced. Also smaller particles are more expensive.

The combined effect of column length and particle size is shown in Figure 28.[9] Here we see that the maximum resolution is equivalent in all cases. Note that as the column length is increased the maximum occurs at a later time. This figure also shows that the resolution is diminished as the analysis time is increased. This is the result of band spreading within the column, primarily caused by axial diffusion at low flow rates.

The column diameter is another design parameter that must be altered during scale-up to accommodate the increased volume. The effects of column diameter are shown in Figure 29,[9] where increasing the column diameter has a minimal effect on the separation. Because of this it is generally preferred to scale up a system by increasing the column diameter while leaving the length constant. Note that the largest diameter used in this illustration is much smaller than the diameter of generally used preparative columns. It is not uncommon to see preparative columns with diameters of 100 mm. Wider diameter columns can, however, introduce the potential for flow distribution problems. If the sample is not evenly distributed when it enters the column, broad peaks and poor resolution will the result.

In SEC scale-up is achieved by maintaining dynamic similarity,[92] when the analytical column and the preparative column have the same L/D and $DV\rho/\mu$ values, where

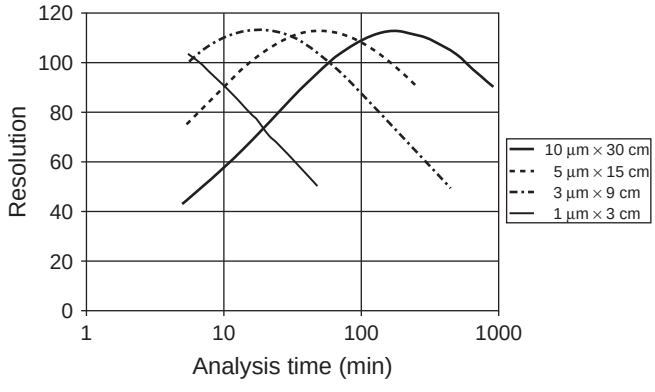

Figure 28 Combined effect of particle size and column length on resolution. Comparison of four columns with constant ratio of column length to particle size. The maximum resolution achievable is the same for all columns, but the maximum occurs at different analysis times. With the 1-μm column, the maximum cannot be reached since it is outside the pressure limit assumed for this graph. ———, 10 μm × 30 cm; - - - - -, 5 μm × 30 cm; — · — · —, 3 μm × 9 cm; ———, 1 μm × 30 cm. (Reproduced by permission of Wiley-VCH, a subsidiary of John Wiley & Sons, Inc., from Uwe D. Neue, *HPLC Columns*, 1997,[9] Copyright Wiley-VCH, 1997.)

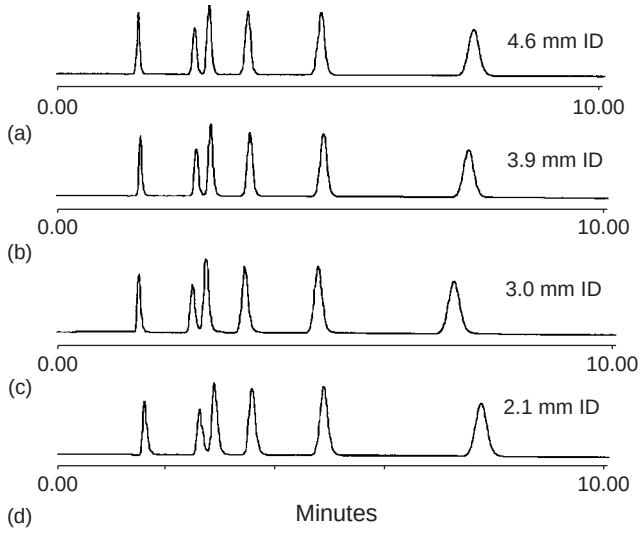

Figure 29 Effect of column diameter on resolution. Comparison of four columns of different IDs but constant length and particle size. The flow rates and injection volumes were scaled in proportion to the column cross-section. (a) 4.6 mm ID, (b) 3.9 mm ID, (c) 3.0 mm ID, (d) 2.1 mm ID. (Reproduced by permission of Wiley-VCH, a subsidiary of John Wiley & Sons, Inc., from Uwe D. Neue, *HPLC Columns*, 1997,[9] Copyright Wiley-VCH, 1997.)

L is the length, D is the diameter, V is the velocity, ρ is the density and μ is the viscosity. The advantage of dynamic similarity is that the elution profile remains unchanged after scale-up. The dynamic similarity approach may also be applicable to HIC, particularly if the scale-up of an HIC column results in a longer length. This is because a longer residence time in the column could lead to conformational changes in the protein structure. If the preparative column and the analytical column are kept dynamically similar (i.e. identical residence times for the sample), conformational changes can be minimized.

The scale-up process requires a tremendous amount of analytical experimentation. To circumvent this, some researchers have focused on the use of rate models to predict elution profiles. A key requirement of a rate model is that the model must accurately describe the equilibrium. Two prominent models that have been reported in the literature for ion exchange are the NISS model[53] developed by Li and Pinto and the steric mass action (SMA) model[93] developed by Brooks and Cramer. Both of these models attempt to describe protein equilibrium on ion-exchange supports. The advantage of modeling is that the amount of experimentation required for scale-up can be reduced. A number of scenarios can be simulated to determine the optimum process parameters. In principle, once these parameters have been obtained, the number of scale-up experiments can be limited to verifying the computer generated results.

Mathematical modeling of protein equilibrium is still in its infancy and more work needs to be done. As modeling efforts continue to develop and improve, our fundamental knowledge of protein–adsorbent interactions will also improve. This increased knowledge will be key in developing purification protocols for future biopharmaceutical products.

6 CONCLUSIONS

The arena of biopharmaceutical proteins and peptides has grown into a billion-dollar-a-year business. This market is expected to grow at a faster pace in the next decade, and with growth will come the need to perform thorough and rapid purifications.

Chromatography is at the core of all biopharmaceutical separations. An overview of the more commonly used chromatographic techniques has been presented. From this overview it was shown that RPC is primarily used for the separation of polypeptides and small proteins that are not very hydrophobic. Protein purification is limited in RPC because of the possibility of denaturation. Proteins are large three-dimensional macromolecules that have a number of hydrophobic patches within their structure. If the stationary phase is strongly hydrophobic then irreversible adsorption or permanent denaturation can occur. HIC is better suited to protein purification than RPC because the stationary phase is less hydrophobic. The environment in HIC is milder and protein elution can be accomplished with salt gradients that reduce the ionic strength of the buffer. Because of this the use of harsh solvents is usually not needed. IEC is also well suited to protein purification because of the amphophilic nature of proteins. Proteins are electrolytic molecules that possess a net charge when the mobile phase pH is either above or below the pI of the protein. IEC exploits small differences in protein net charge to purify mixtures of proteins. AC is used to concentrate valuable proteins that are initially very dilute in solution. The stationary phase in AC elicits specific responses from an analyte. The binding of an analyte with an affinity stationary phase can be thought of in an analogous fashion as a ''hand in a glove'' or a ''key in a lock''. This technique can concentrate the desired analyte over 1000-fold. The stationary phase in AC can be more difficult to synthesize and thus more expensive. SEC is a very popular technique in protein purification. SEC separates proteins by size and/or shape and is frequently used for desalting.

Protein detection can be achieved by a number of methods. The most popular detector is the UV spectrometer. Fluorescence, refractive index, conductivity and light scattering are also used but to a lesser extent in preparative

applications. These techniques are more common in a laboratory setting.

Scale-up of chromatographic columns is usually achieved by increasing the column diameter to allow for a higher throughput. The dynamic similarity approach is used in the scale-up of SEC columns. This technique minimizes the analysis time and maintains the peak order and elution times of each product.

As previously stated, the demand for rapid and thorough protein purification will increase as the biopharmaceutical market continues to grow. To meet this need it is imperative that our knowledge of protein adsorption must also grow. It is vital to focus on research that improves our fundamental understanding of the mechanisms governing protein adsorption. As our knowledge in this phenomenon improves, it will be possible to refine existing techniques and create new protocols to meet better the needs of the ever growing biopharmaceutical market.

ABBREVIATIONS AND ACRONYMS

AC	Affinity Chromatography
ADA	Adenosine Deaminase
AIDS	Acquired Immunodeficiency Syndrome
ASP	Aspartyl
FDA	Food and Drug Administration
GFC	Gel Filtration Chromatography
HIC	Hydrophobic Interaction Chromatography
ID	Internal Diameter
IEC	Ion-exchange Chromatography
Mab	Monoclonal Antibody
NISS	Nonideal Surface Solution
PBD	Polybutadiene
PEG	Polyethylene Glycol
RPC	Reversed-phase Chromatography
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SDM	Stoichiometric Displacement Model
SEC	Size-exclusion Chromatography
SMA	Steric Mass Action
TFA	Trifluoroacetic Acid
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Food **(Volume 5)**
Proteins, Peptides, and Amino Acids Analysis in Food

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Chromatography of Membrane Proteins and Lipoproteins • Gel Electrophoresis in Protein and Peptide Analysis • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis • Planar Chromatography in Pharmaceutical Analysis

Liquid Chromatography **(Volume 13)**
Affinity Chromatography • Capillary Electrophoresis • Reversed Phase Liquid Chromatography

REFERENCES

1. D.M. Matthews, 'Protein Absorption', Wiley-Liss, New York, 1991.
2. A.M. Thayer, 'Great Expectations', *Chem. Eng. News*, August 10, 19–31 (1998).
3. A. Sadana, 'Bioseparation Steps for Proteins and Other Biological Products, Part 1: Cell Disruption', *Biopharm*, March, 26–33 (1994).
4. D.A. Parasrampuria, 'Therapeutic Delivery Issues in Gene Therapy, Part 1: Vectors', *Biopharm*, March, 38–45 (1998).
5. P. Gagnon, E. Grund, T. Lindback, 'Large-scale Process Development for Hydrophobic Interaction Chromatography, Part 1: Gel Selection and Development of Binding Conditions', *Biopharm*, April, 21–27 (1995).
6. M. Marquet, N.A. Horn, J.A. Meek, 'Process Development for the Manufacture of Plasmid DNA Vectors for Use in Gene Therapy', *Biopharm*, September, 27–37 (1995).
7. P.A. Gariopy, R.S. Moore, G. Sofer, 'Multiuse Chromatography System: a Clean-in-place Case Study', *Biopharm*, September, 32–36 (1998).
8. G. Ganetsos, P.E. Barker, *Preparative and Production Scale Chromatography*, Marcel Dekker, New York, 1993.

9. U.D. Neue, *HPLC Columns*, Wiley-VCH, New York, 1997.
10. W.R. Melander, C. Horvath, 'Theory of Reversed-phase Chromatography', in *High Performance Liquid Chromatography*, ed. C. Horvath, Academic Press, New York, 201, 1980.
11. X. Geng, F.E. Regnier, 'Retention Model for Proteins in Reversed-phase Liquid Chromatography', *J. Chromatogr.*, **296**, 15–30 (1984).
12. W.G. Burton, K.D. Nugent, T.K. Slattery, B.R. Summers, 'Separation of Proteins by Reversed-phase High-performance Liquid Chromatography: 1. Optimizing the Column', *J. Chromatogr.*, **443**, 363–379 (1988).
13. T.H.J. Huisman, 'High Performance Liquid Chromatographic Analysis of Human Hemoglobins and Their Polypeptide Chains: its Use in the Identification of Variants', *Anal. Chim. Acta*, **352**, 187–200 (1997).
14. G. Teshima, E. Canova-Davis, 'Separation of Oxidized Human Growth Hormone Variants by Reversed-phase High-performance Liquid Chromatography: Effect of Mobile Phase pH and Organic Modifier', *J. Chromatogr.*, **625**, 207–215 (1992).
15. P.T. Jedrzejewski, L.T. Taylor, 'Comparison of Silica-, Alumina-, and Polymer-based Stationary Phases for Reversed-phase Liquid Chromatography', *J. Chromatogr. Sci.*, **33**, 438–445 (1995).
16. F. Iinuma, K. Hamase, S. Matsubayashi, M. Takahashi, M. Watanabe, K. Zaitsu, 'Sensitive Determination of Melatonin by Precolumn Derivatization and Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr., A*, **835**, 67–72 (1999).
17. M.A. Chlenov, E.I. Kandyba, L.V. Nagornaya, I.L. Orlova, Y.V. Volgin, 'High-performance Liquid-chromatography of Human Glycoprotein Hormones', *J. Chromatogr.*, **631**, 261–267 (1993).
18. R. Rosenfeld, K. Benedek, 'Conformational Changes of Brain-derived Neurotrophic Factor During Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **632**, 29–36 (1993).
19. E. Watson, W.C. Kenney, 'Multiple Peak Formation From Reversed-phase Liquid Chromatography of Recombinant Human Platelet-derived Growth Factor', *J. Chromatogr.*, **606**(2), 165–170 (1992).
20. D. Rapaport, G.R. Hague, Y. Pounty, Y. Shai, 'pH-dependent and Ionic Strength-dependent Fusion of Phospholipid Vesicles Induced By Pardaxin Analogs or by Mixtures of Charge-reversed Peptides', *Biochemistry*, **32**(13), 3291–3297 (1993).
21. J.J. Lacapere, J. Garin, B. Trinnaman, N.M. Green, 'Identification of Amino Acid Residues Photolabeled with 8-Azidoadenosine 5′-Diphosphate in the Catalytic Site of Sarcoplasmic Reticulum CA-ATPase', *Biochemistry*, **32**(13), 3414–3421 (1993).
22. K. Mock, M. Hail, I. Mylchrest, J. Zhou, K. Johnson, I. Sardine, 'Rapid High-sensitivity Peptide-mapping by Liquid Chromatography Mass Spectrometry', *J. Chromatogr.*, **646**(1), 169–174 (1993).
23. T.-M. Chang, B. Erway, W.Y. Chey, 'Rapid Small-scale Preparation of Gastrointestinal Hormones by High-performance Liquid Chromatography on a C_{18} Column', *J. Chromatogr.*, **326**, 121–127 (1985).
24. A.T. Jones, J.N. Keen, N.B. Roberts, 'Human Pepsin 3B Peptide Map Sequence-analysis, Genotype and Hydrophobic Nature', *J. Chromatogr.*, **646**(1), 207–212 (1993).
25. S. Linde, M.E. Roder, S.G. Hartling, C. Binder, B.S. Welinder, 'Separation and Quantitation of Serum Proinsulin and Proinsulin Intermediates in Humans', *J. Chromatogr.*, **548**(1–2), 371–380 (1986).
26. M.-I. Aguilar, M.T.W. Hearn, 'High-resolution Reversed-phase High-performance Liquid Chromatography of Peptides and Proteins', *Methods Enzymol.*, **270**, 3–26 (1996).
27. L. Szepesy, G. Rippel, 'High Performance Hydrophobic Interaction Chromatography of Proteins', *LC/GC Int.*, **5**(11), 24–29 (1992).
28. N. Cooke, P. Stein, N. Miller, 'High Performance Hydrophobic Interaction Chromatography or Proteins', *LC/GC Int.*, **3**(1), 8–13 (1990).
29. D.W. Lee, B.Y. Cho, 'Investigation of the Retention Behavior and Structural-change of Proteins in Reversed-phase and Hydrophobic Interaction Chromatography', *J. Liq. Chromatogr.*, **17**(12), 2541–2558 (1994).
30. D.W. Lee, B.Y. Cho, 'A Study of the Retention Behavior of Proteins in High-performance Liquid Chromatography(I): the Effect of Solvent and Temperature on Retention Behavior of Proteins in Reversed-phase Chromatography', *Bull. Korean Chem. Soc.*, **14**(4), 510–514 (1993).
31. D.W. Lee, B.Y. Cho, 'A Study of the Retention Behavior of Proteins in High-performance Liquid Chromatography(II): the Effect of Salt and Temperature on Retention Behavior of Proteins in Hydrophobic Interaction Chromatography', *Bull. Korean Chem. Soc.*, **14**(4), 515–519 (1993).
32. M.A. Esquibel-King, A.C. Dias-Cabral, J.A. Queiroz, N.G. Pinto, 'A Study of Hydrophobic Interaction Adsorption of Bovine Serum Albumin Under Overloaded Conditions Using Flow Microcalorimetry', *J. Chromatogr., A*, **865**, 111–122 (1999).
33. M.A. Esquibel-King, 'An Investigation of Hydrophobic Interaction Chromatography Using Flow Microcalorimetry', Masters Thesis, University of Cincinnati, July, 1999.
34. T. Arakawa, 'Thermodynamic Analysis of the Effect of Concentrated Salts on Protein Interaction with Hydrophobic and Polysaccharide Columns', *Arch. Biochem. Biophys.*, **248**(1), 101–105 (1986).
35. R.E. Shansky, S.-L. Wu, A. Figueroa, B.L. Karger, 'Hydrophobic-interaction Chromatography of Biopolymers', in *HPLC of Biological Macromolecules: Methods*

and Applications, eds. K.M. Gooding, F.E. Regnier, Marcel Dekker, New York, 1990.
36. S.-L. Wu, A. Figueroa, B.L. Karger, 'Protein Conformational Effects in Hydorphobic Interaction Chromatography: Retention Characterization and The Role of Mobile Phase Additives and Stationary Phase Hydrophobicity', *J. Chromatogr.*, **371**, 3–27 (1986).
37. X.D. Geng, L. Guo, J. Chang, 'Study of the Retention Mechanism of Proteins in Hydrophobic Interaction Chromatography', *J. Chromatogr.*, **507**, 1–23 (1990).
38. T.W. Perkins, D.S. Mak, T.W. Root, E.N. Lightfoot, 'Protein Retention in Hydrophobic Interaction Chromatography: Modeling Variation with Buffer Ionic Strength and Column Hydrophobicity', *J. Chromatogr., A*, **766**, 1–14 (1997).
39. A. Donato, M.A. Ciardiello, M. de Nigris, R. Piccoli, L. Mazzarella, G. D'Alessio, 'Selective Deamidation of Ribonuclease A, Isolation and Characterization of the Resulting Isoaspartyl and Aspartyl Derivatives', *J. Biol. Chem.*, **268**(7), 4745–4751 (1993).
40. S. Xie, F. Svec, J.M.J. Frechet, 'Rigid Porous Polyacrylamide-based Monolithic Columns Containing Butyl Methacrylate as a Separation Medium for the Rapid Hydrophobic Interaction Chromatography of Proteins', *J. Chromatogr., A*, **775**, 65–72 (1997).
41. R. Pavlu, U. Johansson, C. Nyulen, A. Wichman, 'Rapid Purification of Monoclonal Antibodies by High-performance Liquid Chromatography', *J. Chromatogr.*, **359**, 449–460 (1986).
42. P. Gagnon, P.G. Cartier, J.J. Maikner, R. Eksteen, Michael Kraus, 'A Systematic Approach to the Purification of Monoclonal Antibodies', *LC/GC Int.*, **11**(1), 26–34 (1993).
43. V. Stocchi, P. Cardoni, P. Ceccaroli, G. Piccoli, L. Cucchiarini, R.D. Bellisand, M. Dacha, 'Hydrophobic Interaction Chromatography – a Fast Technique for High-resolution of Enzymes Present in Multiple Forms', *FASEB J.*, **7**(7), A1131 (1993).
44. J. Hiyama, A. Surus, A.G.C. Renwick, 'Purification of Human Pituitary LH and Thyrotropin by Hydrophobic Chromatography', *J. Endocrinol.*, **125**(3), 493–500 (1990).
45. P.J. Padfield, R.M. Case, 'Separation of the Proteins Present in Pancreatic Juice Using Hydrophobic Interaction Chromatography', *Anal. Biochem.*, **171**, 294–299 (1988).
46. S.C. Goheen, R.S. Matson, 'Purification of Human Serum Gamma Globulins by Hydrophobic Interaction Chromatography', *J. Chromatogr.*, **326**, 235 (1985).
47. W.R. Bowen, D.T. Hughes, 'Ion-exchange of Proteins: a Microcalorimetric Study of the Adsorption of Bovine Serum Albumin on Anion-exchange Materials', *J. Colloid Interface Sci.*, **158**, 395–402 (1993).
48. P. Raje, N.G. Pinto, 'Importance of Heat of Adsorption in Modeling Protein Equilibria for Overloaded Chromatography', *J. Chromatogr., A*, **796**, 141–156 (1998).
49. D.S. Gill, D.J. Roush, R.C. Willson, 'Adsorption Heterogeneity and Thermodynamic Driving Forces in Anion Exchange Equilibria of Cytochrome b5', *J. Colloid Interface Sci.*, **167**, 1–7 (1994).
50. A. Chandavarkar, N.G. Pinto, 'Modeling Adsorption Isotherms of Proteins for Overloaded Ion-exchange Chromatography', in *Fundamentals of Adsorption*, ed. F. Meunier, Elsevier Science, New York, 413–418, 1998.
51. *Pharmacia Biodirectory '98*, Amersham Pharmacia Biotech, Piscatoway, New Jersey, 1998.
52. D. Sheehan, R. FitzGerald, 'Ion-exchange Chromatography', in *Protein Purification Protocols*, ed. S. Doonan, Humana Press, Totowa, NJ, 1996.
53. Y. Li, N.G. Pinto, 'Model for Ion-exchange Equilibria of Macromolecules in Preparative Chromatography', *J. Chromatogr., A*, **702**, 113–123 (1995).
54. O. Talu, I. Zweibel, 'Multicomponent Adsorption Equilbria of Nonideal Mixtures', *AIChE J.*, **32**, 1263 (1986).
55. A.C. Awade, T. Efstathiou, 'Comparison of Three Liquid Chromatographic Methods for Egg-white Protein Analysis', *J. Chromatogr., B*, **723**, 69–74 (1999).
56. S.-C.D. Jen, N.G. Pinto, 'Use of the Sodium Salt of Poly(vinylsulfonic acid) as a Low-molecular Weight Displacer for Protein Separations by Ion-exchange Displacemanent Chromatography', *J. Chromatogr.*, **519**, 87–98 (1990).
57. S.-C.D. Jen, N.G. Pinto, 'Nonlinear Chromatography of β-Lactoglobulins A and B: Non-Langmuirian Behavior', *IEC Res.*, **34**(8), 2685–2691 (1995).
58. A. Singh, N.G. Pinto, 'Polymeric Short-fiber Chromatographic Supports for Downstream Processing of Biomolecules', *React. Polym.*, **24**, 229–242 (1995).
59. J.K. King, N.G. Pinto, 'Short Fibrous Supports for Preparative Chromatographic Separations of Biomolecules', *J. Chromatogr.*, **609**, 61–68 (1992).
60. R. Necina, K. Amatschek, A. Jungbauer, 'Capture of Human Monoclonal Antibodies from Cell Culture Supernatant by Ion Exchange Media Exhibiting High Charge Density', *Biotechnol. Bioeng.*, **60**(6), 689–698 (1998).
61. H. Yamakawa, K. Higashino, O. Ohara, 'Sequence-dependent DNA Separation by Anion-exchange High-performance Liquid Chromatography', *Anal. Biochem.*, **240**, 242–250 (1996).
62. Y.-B. Yang, K. Harrison, 'Influence of Column Type and Chromatographic Conditions on the Ion-exchange Chromatography of Immunoglobulins', *J. Chromatogr., A*, **743**, 171–180 (1996).
63. J. Underwood, G.J. Underwood, G. Cartwright, M.T.W. Hearn, 'High-performance Liquid Chromatography of Amino Acids, Peptides and Proteins: XC III Comparison of Methods for the Purification of Mouse Monoclonal Immunoglobulin M Autoantibodies', *J. Chromatogr.*, **476**, 269–290 (1989).
64. K.N.B. Roberts, W.H. Taylor, 'Improved Separation of Human Pepsins from Gastric Juice by High-performance

Ion-exchange Chromatography', *J. Chromatogr.*, **476**, 291–297 (1989).
65. D.J. Burke, J.K. Duncan, C. Siebert, G.S. Ott, 'Rapid Cation-exchange Chromatography of Hemoglobulins and Other Proteins', *J. Chromatogr.*, **359**, 533–540 (1986).
66. M. Carlquist, A. Rokaeus, 'Isolation of a ProForm of Porcine Secretion By Ion-exchange and Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **296**, 143–151 (1984).
67. G. Subramanian, *Process Scale Liquid Chromatography*, VCH Publishers, New York, 1995.
68. Y. Shibusawa, 'Surface Affinity Chromatography of Human Peripheral Blood Cells', *J. Chromatogr., B*, **722**, 71–88 (1999).
69. S.R. Narayanan, 'Preparative Affinity Chromatography of Proteins', *J. Chromatogr., A*, **658**, 237–258 (1994).
70. F.B. Anspach, O. Hilbeck, 'Removal of Endotoxins of Affinty Sorbents', *J. Chromatogr., A*, **711**, 81–92 (1995).
71. H. Lakhiari, D. Muller, J. Jozefonvicz, 'Interaction Mechanisms Between Insulin and *N*-Acetylneuraminic Acid in Affinity Chromatography', *J. Chromatogr., A*, **711**, 93–103 (1995).
72. S.B. Tensen, L. Thim, T. Christensen, H. Woeldike, E. Boel, K. Hjortshoej, M. Tansen, 'Affinity Chromatography of Recombinant Rhizomucormikoki Aspartic Proteinase on Si-300 Bacitracin', *J. Chromatogr.*, **476**, 227–233 (1989).
73. K. Shimura, M. Kazama, K.-I. Kasai, 'High Performance Affinity Chromatography of Plasmin and Plasminogen on a Hydrophobic Vinyl-polymer Gel Coupled with *p*-Aminobenamidine', *J. Chromatogr.*, **292**, 369–382 (1984).
74. G.K. Sofer, L.-E. Nystrom, *Process Chromatography: a Practical Guide*, Academic Press, San Diego, 1989.
75. R.D. Ricker, L.A. Sandoval, 'Fast Reproducible Size-exclusion Chromatography of Biological Macromolecules', *J. Chromatogr., A*, **743**, 43–50 (1996).
76. G.B. Irvine, 'Size-exclusion High-performance Liquid Chromatography of Peptides: a Review', *Anal. Chim. Acta*, **352**, 387–397 (1997).
77. H.G. Barth, B.E. Boyes, C. Jackson, 'Size Exclusion Chromatography', *Anal. Chem.*, **68**, 445R–466R (1996).
78. F. Ahmed, B. Modrek, 'Biosep-SEC-S High-performance Size-exclusion Chromatographic Columns for Proteins and Peptides', *J. Chromatogr.*, **599**, 25–33 (1992).
79. L. Tarditi, M. Camagna, A. Parisi, C. Vassarotto, L.B. Demonte, M. Letarte, F. Malavasi, M. Mariani, 'Selective High-performance Liquid Chromatographic Purification of Bispecific Monoclonal Antibodies', *J. Chromatogr.*, **599**, 13–20 (1992).
80. A.K. Taneja, S.Y.M. Lan, R.S. Hodges, 'Separation of Hydrophobic Peptide Polymers by Size Exclusion and Reversed Phase High-performance Liquid Chromatography', *J. Chromatogr.*, **317**, 1–10 (1984).
81. L.J. Delucas, D.D. Muccio, 'Purification of Bovine Rhodopsin by High-performance Size Exclusion Chromatography', *J. Chromatogr.*, **296**, 121–128 (1984).
82. J.P. Landers, *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, Florida, 1994.
83. C. Schoneich, S.K. Kwok, G.S. Wilson, S.R. Rabel, J.F. Stobaugh, T.D. Williams, D.G. Vandervelde, 'Separation and Analysis of Peptides and Proteins', *Anal. Chem.*, **65**, 67R–84R (1993).
84. W.G. Kuhr, C.A. Monnig, 'Capillary Electrophoresis', *Anal. Chem.*, **64**, 389R–407R (1992).
85. N.J. Adamson, E.C. Reynolds, 'Rules Relating Electrophoretic Mobility, Charge and Molecular Size of Peptides and Proteins', *J. Chromatogr., B*, **699**, 133–147 (1997).
86. I. Messana, D.V. Rossetti, L. Cassiano, F. Misti, B. Giardina, M. Castagnola, 'Peptide Analysis by Capillary (Zone) Electrophoresis', *J. Chromatogr., B*, **699**, 149–171 (1997).
87. T.K. Jorgensen, L.H. Bagger, J. Christiansen, Gitte H. Johnsen, J.R. Faarbaek, L. Jorgensen, B.S. Welinder, 'Quantifying Biosynthetic Human Growth Hormone on *Escherichia coli* with Capillary Electrophoresis Under Hydrophobic Conditions', *J. Chromatogr., A*, **817**, 203–214 (1998).
88. S.R. Rabel, J.F. Stobaugh, 'Applications of Capillary Electrophoresis in Pharmaceutical Analysis', *Pharm. Res.*, **10**(2), 171–186 (1993).
89. V. Dolink, 'Capillary Zone Electrophoresis of Proteins', *Electrophoresis*, **18**, 2353–2361 (1997).
90. P. Sewell, D. Kelly, *Chromatographic Separations*, John Wiley & Sons, Chichester, 1987.
91. E.D. Katz, *High Performance Liquid Chromatography: Principles and Methods in Biotechnology*, John Wiley & Sons, Chichester, 1996.
92. K. Vorauer, M. Skias, A. Trkola, P. Schulz, A. Jungbauer, 'Scale-up of Recombinant Protein Purification by Hydrophobic Interaction Chromatography', *J. Chromatogr.*, **625**, 33–39 (1992).
93. C.A. Brooks, S.M. Cramer, 'Steric Mass-action Ion Exchange: Displacement Profiles and Induced Salt Gradients', *AIChE J.*, **38**(12), 1969–1978 (1992).

Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery

Alexandru T. Balaban
Texas A&M University, Galveston, TX, USA

1 Introduction 2
2 Classical and Present Day Quantitative Structure–Activity Relationship Paradigms 3
3 Chemical Structure: Constitutional and Steric Isomerism 4
3.1 Atomic Bonds 4
3.2 Molecular Bonds 4
3.3 Isomers 4
4 Chemical Coding of Structures 5
5 Hydrophobicity or Lipophilicity Parameters 5
6 Descriptors for Electronic, Polar, and Steric Factors 6
7 The Free–Wilson (de novo) Method 7
8 Topological Indices and their Uses in Quantitative Structure–Activity Relationships 8
9 Three-dimensional Shape Descriptors 10
10 Molecular Similarity and Dissimilarity 11
11 Case Studies of Quantitative Structure–Activity Relationships 12
12 Quantitative Structure–Activity Relationships in Environmental Chemistry and Toxicology 13
13 Statistical Data and Computational Approaches 14
14 Conclusions 14
Abbreviations and Acronyms 16
Related Articles 17
Further Reading 17
References 17

Quantitative structure–activity relationships (QSARs) are mathematical equations or other types of functions (such as the weights of connections in artificial neural networks) relating chemical structures to their biological activity. The purpose of QSAR studies is to predict novel structures with either beneficial activity as drugs for human or veterinary medicine (bactericides, antiviral or anticancer drugs, and metabolic regulators, such as hypoglycemics, hypotensives, etc.), or selective toxicity for various unwanted higher organisms (pesticides such as fungicides, insecticides, acaricides, weed killers, etc.) Quantitative structure–property relationships (QSPRs) are similar, but the property may be physical or chemical. The main problems involve

1. *finding a mathematical representation of chemical structures (usually organic molecules), represented either as molecular graphs by their constitution or connectivity without considering the three-dimensional (3-D) (stereochemical) factors, or as 3-D objects including stereochemical information;*
2. *measuring the biological activities of a series of molecules;*
3. *finding the QSAR between each type of biological activity and the most convenient molecular descriptors.*

Physicochemical parameters that are closely related to drug transport and ligand binding include lipophilicity, polarity, polarizability, and electronic influence on hydrogen-donor and hydrogen-acceptor binding. The next task is to use the QSAR to predict which novel structures to prepare in order to obtain molecules with the desired biological activity. Then the cycle is usually repeated, because a single pass seldom affords the optimal solution.

The main mathematical descriptors and the techniques for obtaining QSARs are reviewed. An important molecular parameter, which may be measured experimentally or computed from the chemical structure, is lipophilicity or hydrophobicity; Hansch introduced the n-octanol–water partition coefficient as a measure of this property, which, to a large extent, determines the ability of molecules to penetrate the lipophilic, bilayer, extra- or intracellular membranes. Lipophilicity levels that are too high lead to insolubility in water, i.e. it is difficult to administer the drug orally via the digestive tract. The main constitutional and 3-D molecular descriptors are described and examples are given. To be statistically valid, correlations must involve orthogonal or orthogonalized descriptors. Linear, multilinear, and nonlinear types of correlations are reviewed. Screening of virtual combinatorial libraries together with high-throughput combinatorial synthesis and testing provides modern tools for more efficient drug design.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

1 INTRODUCTION

A fitting introduction for the importance of analytical chemistry for QSAR in drug design[1] is provided in the following quotation, reproduced with permission[2]:

> Over the last two decades the art of drug discovery has changed dramatically with the introduction of new analytical tools. Analytical chemistry revolutionized both the analysis of chemical compounds and the study of biological processes. Today crystallography and NMR (nuclear magnetic resonance) contribute significantly to biostructural research and have led to the unraveling of many details about the structure and functions of macromolecules, such as nucleic acids and proteins. The second revolution, developed in parallel and which is now indispensable, concerns the use of computers in molecular design and in the lead discovery process.

These phrases, written in 1994 by van de Waterbeemd,[2] must now be supplemented by the contribution brought about by high-throughput analytical techniques, which detect biological activity in minute amounts of substances synthesized simultaneously in combinatorial libraries numbering thousands of compounds. Only a very small number of these substances are then prepared in sufficient amounts for chemical and biological characterization and testing, namely, those that show a positive response to fluorescent, radiochemical, immunochemical, or other analytical techniques. Thus, although the total number of substances registered by *Chemical Abstracts* is now about 30 million, with an average yearly increase of 0.6–1.0 million, this number will not jump spectacularly with the advent of combinatorial libraries, because only characterized compounds are registered.

It is estimated that before the advent of combinatorial chemistry, for each drug that reached the market, about 5000 compounds had to be synthesized and tested over about 15 years, and thus the cost was US $\$10^{10}$–$10^{11}$. It is hoped that both the cost and the duration of new drug launches will be decreased by combinatorial libraries (virtual and experimental) and high-throughput screening, and that QSARs will help to reduce the number of substances that have to be tested.

The vast majorities of novel compounds characterized each year are organic substances with presumed beneficial biological activity, and are synthesized by the major drug companies and by research universities. The necessity of continuously looking for new medicinal drugs originates in a multitude of factors, among which the most important are as follows:

1. the fact that pathogens develop resistance (nowadays few of the ''wonder antibiotics'' of the 1950s are still useful);
2. the discovery of genetically derived metabolic disorders (the Human Genome Project having completed its task, the complete mapping of the human genome) provides clues for curing such disorders, mainly through computer-aided design of small molecules that can act as receptor blockers or activators for crucial proteins;
3. the need for improving the efficiency of existing drugs learning from the experience of the past, and giving rise to new generations of drugs;
4. developments in the knowledge of biochemical pathways and of the interconnection between the complex networks involved, as well as the opening of new areas such as the mechanisms of memory, neurology disorders, and various forms of cancer.

The huge costs associated with drug discovery, including in vitro and in vivo tests (the latter finalized in lengthy clinical trials), have to be recovered by the drug manufacturing companies during a relatively short time interval under the present patent laws. Insurance companies prefer generic drugs that are less expensive after the patent expires, but very few manufacturers of generic drugs contribute to the research effort.

Computational techniques have now made possible a third type of screening, in addition to the in vitro and in vivo tests, often called *in silico tests*. Virtual combinatorial libraries, including far larger numbers of structures than can presently be explored experimentally even by the most high-performance combinatorial approaches, can be screened by computers, provided that a sufficient number of reliable experimental data exist to obtain meaningful QSAR studies. An eloquent example is provided by a recent report published by Grassy et al.[3] The activity of the lead immunosuppressive peptide for skin and heart allograft survival in mice was increased about 100 times by a computer-assisted design using 13 molecular descriptors. With 35 natural and nonnatural amino acids that could occupy the seven positions of the relevant portion of the peptidic chain, the virtual combinatorial library would have numbered $35^7 \cong 64$ billion structures. A first filter based on the lipophilicity reduced the number of amino acids to six (one nonnatural and five natural amino acids). All the resulting $6^7 \cong 280\,000$ structures were calculated to ascertain whether their 13 descriptors fitted into the favorable domains previously determined from the learning set of about 20 active and inactive compounds. These descriptors, independent of 3-D factors and therefore very fast to compute, are discussed later. Only 26 peptides passed this test, i.e. one in about 10 000, and they were analyzed by molecular dynamics simulations to compare the similarity between their conformational space and that of the active peptides.

The five peptides that passed this last filter were synthesized and tested in vivo; one of these was about 100 times more active than the best peptide known until then.

Among the molecular descriptors, this review describes topological indices (TIs) in more detail than is usually encountered in the overviews of QSARs. The reasons for doing this are that

1. such molecular descriptors deserve to be better known because they have a bright future for screening of virtual libraries;
2. chemometric methods may also profit from this approach;
3. TIs have been intensively developed for QSARs and QSPR purposes by the author.

2 CLASSICAL AND PRESENT DAY QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP PARADIGMS

From the rich literature on QSARs, several post-1978 references have been selected.[4–73] The literature on QSPRs is almost as rich, and again only a few leading references have been provided.[66,74–80] The classical approach to drug design involved the following steps[5]:

1. A known active structure (the lead compound) may be discovered by chance in random screening, by testing many natural products, or in the search for another type of bioactivity (as examples, the erectile activity of Viagra-like compounds was discovered in clinical trials during the search for a hypotensive drug, and oral antidiabetics such as tolbutamide were developed after observing the hypoglycemic activity of new sulfa drugs). Alternatively, enzyme inhibitors may be designed by analogy with enzyme substrates, by the receptor-mapping method, or by Lindquist's transition state analogs.
2. An exploratory series patterned after the lead compound is generated either by combinatorial chemistry (the preferred contemporary method, when a high-throughput screening is available and when the number of congeneric structures is considerable) or by normal stepwise syntheses. At this stage, rational design methods can come into play, as is shown further (Free–Wilson method, extra thermodynamic approach, etc.).
3. On the basis of the bioactivity results of this test series, it is possible to formulate a QSAR using various approaches (molecular descriptors and computer programs, i.e. software) at different levels of sophistication.
4. Finally, on the basis of the QSAR study, it should be possible to predict structures that optimize the beneficial activity of the lead compound, and at the same time minimize the unwanted side effects, most importantly the toxicity.

The ratio between the beneficial and toxic doses is called *the therapeutic ratio*, and drug designers strive to arrive at structures with optimal therapeutic ratios. (Other unwanted side effects are also important, as shown by the example of thalidomide, which had passed clinical trials and was put on the market in Europe for one year, because at that time no test had discovered that one of its two stereoisomers had severe teratogenic effects.)

Present day paradigms may change the weight and the order of experimental versus computational approaches. The facility of virtual library screening when there is enough experimental material with established bioactivity data enhances the weight of the computer-assisted drug design (CADD) phase. Actually, neither this acronym nor computer-aided molecular design (CAMD) are the best descriptors, and should be replaced by computer-aided ligand design (CALD), as argued by Boyd.[67] CALD is more accurate because computers are not used to design drugs directly, but instead are used to design small molecules (ligands) that may bind specifically to a larger biomolecule, and thus may have a beneficial effect; however, many such ligands have to be tested until a drug is obtained. Combinatorial chemistry techniques also greatly increase the potential for experimental exploration of congeners having the same molecular skeleton (scaffold) and differing in their substitution patterns. In both cases, however, one has to have a lead compound and a test series for an initial QSAR. As a result, one can either explore huge virtual combinatorial libraries and experimentally test only a few compounds,[3] or one can rely on the power of combinatorial synthesis and high-throughput testing to experimentally explore thousands of compounds, and then develop a powerful and very reliable QSAR. In both approaches, several such cycles may be necessary.

Many QSAR data on biological activities have not been obtained with whole organisms or living cell cultures, but with purified enzymes. Often their molecular structure has been determined by X-ray diffractometry in the crystal, or more recently by NMR in solution (with geometry closer to that which is responsible for the catalytic activity of the enzyme). In some cases, the mechanism of this catalytic activity is known in detail. The accuracy and reproducibility of QSAR data for enzyme inhibition are much higher than those of in vivo studies.

3 CHEMICAL STRUCTURE: CONSTITUTIONAL AND STERIC ISOMERISM

3.1 Atomic Bonds

Covalent bonds are the strongest chemical bonds and they hold atoms together in molecules that are unchanged in all three states of matter (crystalline, liquid, and gaseous). Neither phase transitions nor solvation has any effect on such bonds. Ionic and metallic bonds are usually weaker; the former are compensated for in aqueous solutions by ion – dipole interactions so that ions move independently. For ionic or metallic solids, only empirical formulae (e.g., NaCl, $CaCl_2$), which provide information on the relative numbers of atoms, are meaningful. However, for compounds with covalent bonds, molecular formulae are more meaningful than empirical formulae, because they also provide information on the absolute number of atoms in the molecule. Thus, acetylene and benzene both have the same empirical formula, CH, but their molecular formulae are different: C_2H_2 and C_6H_6, respectively.

3.2 Molecular Bonds

Intermolecular forces (hydrogen bonds, dipole interactions, dispersion or London forces, hydrophobic interactions) are not chemical bonds, i.e. do not lead to stable chemical compounds, but are responsible for the existence of condensed phases (liquid and crystalline), and play an important role in biochemical processes. These forces are much weaker than chemical bonds, and hence they lead to associations that are reversible and depend strongly on the surroundings, as is necessary in enzymatic reactions. The van der Waals' intermolecular interactions are both weakly attractive at larger distances r and highly repulsive at lower distances. The simplest expression for van der Waals' interaction energies is the Lennard–Jones potential energy: $E_{vdW} = (a/r^{12}) - (b/r^6)$, with constants a and b to be determined.

3.3 Isomers

Most of the chemical bonds in organic compounds are covalent bonds. Isomers are different compounds that share the same molecular formula. There is just one stable compound, acetylene, with the molecular formula C_2H_2; however, acetylene, which has a triple bond between two CH groups, may rearrange to vinylidene, $M{=}C{=}CH_2$, when adsorbed on a metallic catalyst (M). There are many structures isomeric with benzene, but only one of these has a planar regular hexagonal structure with six CH groups and ''delocalized'' or ''aromatic'' C–C bonds. Four of these isomers are called *valence isomers* because like benzene they have six CH groups, but the bonding is different.[81]

It is customary to classify isomers into *constitutional* and *steric* isomers. Constitutional isomers differ in their connectivity patterns, e.g. *n*-butane and isobutane (which have the common molecular formula C_4H_{10}) have a linear and a branched chain, respectively, of four carbon atoms with single bonds between them. Their constitutional formulae may be written in Smiles notation,[32] which neglects single bonds, as CCCC and CC(C)C, respectively (when also neglecting the hydrogen atoms, as is usual in molecular graphs for hydrocarbons). Other examples of constitutional isomers are ethanol and dimethyl ether (C_2H_6O), which have the hydrogen-depleted constitutional formulae CCO and COC, respectively; 1-fluoropropane and 2-fluoropropane (C_3H_7F), with constitutional formulae FCCC and CC(F)C, respectively; and 1-butene and 2-butene (C_4H_8) with constitutional formulae C=CCC and CC=CC, respectively. All such constitutional isomers may be represented mathematically by molecular graphs, where vertices (points) symbolize atoms, and edges (lines) correspond to covalent bonds. In other types of graphs used in chemistry, such as reaction graphs, vertices symbolize chemical species and edges represent elementary reaction steps.

Stereoisomers have the same constitutional formulae, but differ in the geometrical 3-D disposition of bonds. Steric isomers are of two kinds:

- *diastereoisomers* or *diastereomers*, whose properties differ as much as those of constitutional isomers; they are characterized either by differences in distances between nonbonded atoms (e.g. cis/trans or Z/E isomers of polysubstituted compounds with rings such as 1,2-disubstituted cycloalkanes or with double bonds such as 2-butene) or by the presence of several chiral centers and an absence of mirror-image symmetry;
- *enantiomers*, which are nonsuperimposable mirror images of one another, and whose properties are identical except for those involving interaction with chiral forms of energy (such as plane-polarized photons) or matter (such as chiral solvents and biomolecules or bioreceptors).

Natural compounds synthesized enzymatically by living cells are produced as single enantiomers, whereas chemical reactions carried out in the absence of chiral catalysts yield mixtures of enantiomers (mixtures of equal amounts of two enantiomers with opposite chirality are called *racemic mixtures*, after the historical discovery by Louis Pasteur in the case of tartaric acids). A tetrahedral carbon atom with sp^3 hybridization and with four different substituents can have two nonsuperimposable (chiral) configurations leading to

a pair of enantiomers; a pentacovalent atom with trigonal – bipyramidal geometry gives rise to 10 pairs of enantiomers when it has five different substituents; a hexacovalent atom with octahedral geometry affords 15 pairs of enantiomers when all its six substituents are different.

In addition to these types of isomerism, which are seen at normal temperatures because of differences in molecular properties and the possibility of separating isomers, there is also the possibility of conformational isomerism, which does not usually result in isolatable compounds at normal temperatures because of low energy barriers toward rotation around single bonds. However, at lower temperatures, or when rotation barriers are high owing to electronic or steric factors, ''rotamers'' may become stable enough to be separated.

4 CHEMICAL CODING OF STRUCTURES

For a long time, the Chemical Abstracts Service (CAS) used an extended Morgan algorithm for coding molecular constitutions (connectivity), and more recently a 3-D coding algorithm, CONCORD, has been used. So far, however, parts of the 3-D structures have not been correct in their absolute configuration. The CAS registry number is uniquely associated with a substance or an industrially produced mixture of substances, but bears no relationship to the chemical structure. Linear notation systems (e.g. the Wiswesser line notation) rely on elaborate conventions and are outmoded.

By contrast, the algorithmic structure display of CAS[82] is able to convert a chemical graph into a connection table and vice versa, allowing rapid structure or substructure searches of all known compounds. This fact makes chemistry the best documented science because it does not rely on words and semantics for accessing information on any chemical structure or substructure. Independent of their own language, chemists can communicate the constitution of chemical compounds via graphs. However, when words are needed, as when searching for properties or phenomena, documentation in chemistry is no longer any better than in other sciences such as mathematics, physics, biology, or medicine.

For molecules with N nonhydrogen atoms represented as constitutional (molecular) hydrogen-depleted graphs, chemical constitution is easily encoded as an adjacency matrix **A**, whose entries a_{ij} are 1 for two adjacent atoms i and j (i.e. atoms which are connected by a covalent bond) and 0 otherwise. The result is a symmetrical matrix with N rows and columns. The sum v_i over row or column i is called the *degree of vertex i*; for instance, a quaternary carbon atom (connected to four other C atoms) has degree $v_i = 4$, and a secondary carbon atom has degree $v_i = 2$.

More detailed topological information is provided by the distance matrix **D**, whose entries d_{ij} represent topological distances between vertices i and j, i.e. the number of bonds along the shortest path between these vertices. Thus, both matrices **A** and **D** have zeros along the main diagonal and the same entries 1, but all off-diagonal zeros in **A** are replaced in **D** by integers that are larger than 1. Sums of entries along row or column i in **D** are called *distance sums* (distasums) d_i.

Both matrices **A** and **D** contain all the necessary information to calculate the molecular constitution. TIs have been devised using these matrices as a basis. It had been conjectured that the eigenvalues (EVAs) or characteristic polynomials of **A** or **D** could uniquely characterize a chemical graph, but this conjecture was proved to be unfounded.[83]

5 HYDROPHOBICITY OR LIPOPHILICITY PARAMETERS

Historically, the QSAR has its origins in the first attempts (Crum-Brown and Frazer, 1868) to quantitatively connect the biological effects of alkaloids with their chemical structure. In 1893, Richet discovered an inverse relationship between differences in the toxicity of organic compounds and their water solubility. Meyer (1899) and Overton (1901) expressed a correlation between narcotic activity and lipophilicity using oil–water partition coefficients, forerunners of log P values. Fühner and Neubauer (1906) showed that narcotic activities increase regularly in homologous series, suggesting the additivity of group contributions, a theory which was developed further by Bruice, Kharasch, and Winzler in 1956. Ferguson's principle, formulated in 1939 at the ICI laboratories, linked the narcotic activity with thermodynamic equilibria governed by the partition coefficient, creating the basis for the development of halothane anesthetics. Further investigations by Albert, initiated in the 1940s, linked the antiseptic activity of acridine derivatives with their ionization equilibria. Zahradnik and Hansen are also considered to be pioneers in this area because in 1960–1962 they applied Hammett's approach to biological data. Details of references for these early efforts may be found in some of the cited monographs.[5,31–39,49]

The official birthdate for the QSAR is considered to be 1962, when Hansch et al.[84] made the connection between biological activities (expressed logarithmically in terms of the molar concentration C of compounds producing a standard response in a constant time interval) and the octanol–water partition coefficient P for a substituted

derivative $X–R$ of the parent aromatic compound H–R, expressed by analogy with Hammett's parameters as Equation (1).

$$\log\left(\frac{P_{X-R}}{P_{H-R}}\right) = \pi_X \tag{1}$$

Hansch et al.[85,86] then combined two parameters for explaining biological activities, one parameter being the Hammett substituent constant σ, and the other the lipophilicity, π or log P, as shown in Equation (2).

$$\log\left(\frac{1}{C}\right) = k_1\pi + k_2\sigma + k_3 \tag{2}$$

On the basis of the observation that there is often an optimal lipophilicity, a parabolic equation in terms of log P or π (Equation 3) was introduced by Hansch and Clayton in the early 1960s and reviewed in 1969.[87,88] They also combined lipophilicity not only with an electronic parameter (Hammett's σ) but also with a steric parameter (Taft's E_s constant, Equation 4), and devised nonlinear relationships in terms of log P (Equations 5 and 6):

$$\log\left(\frac{1}{C}\right) = k_1\pi - k_2\pi^2 + k_3\sigma + k_4 \tag{3}$$

$$\log\left(\frac{1}{C}\right) = k_1\pi + k_2E_s + k_3\sigma + k_4 \tag{4}$$

$$\log\left(\frac{1}{C}\right) = k_1\log P - k_2\log(k_3P + 1) + k_4\sigma + \cdots + k_5 \tag{5}$$

$$\log\left(\frac{1}{C}\right) = k_1\pi - k_2\log(k_3 10^{\pi} + 1) + k_4\sigma + \cdots + k_5 \tag{6}$$

Equations (2)–(6) are all extrathermodynamic.[5,37] The choice of *n*-octanol proved to be an ideal one because, being similar to biological membranes (a polar head plus a nonpolar tail), it dissolves most druglike substances (unlike alkanes, lower alcohols, or water). In addition, although it is practically insoluble in water, octanol dissolves enough in water (2.3 M) for solutes to pass into the organic phase with their hydration shell so that they do not associate with each other. Also, octanol has a low volatility and does not absorb in the ultraviolet (UV) region. Differences between partition coefficients for aqueous solutes using *n*-hexane and octanol as organic solvents may serve as quantitative measures for the hydration/dehydration properties of molecules or substituents. Such hydrophobic interactions are responsible for micelle formation and are important for biomolecules such as proteins, or for receptor–agonist (pharmacophore) interactions.

Partition coefficients (log P) may be obtained experimentally either by the "shake flask" method, or more conveniently by high-performance liquid chromatography (HPLC) measurements on reversed-phase HPLC columns and subsequent adjustment of the retention times with correction factors using standard substances.[89] However, theoretical algorithms for computing log P values for any known or unknown molecule are available nowadays. Starting with Hopfinger and continuing with Reynolds, Klopman, Bodor, Reddy, Katagi, and their coworkers, various quantum-chemical methods were devised.[5,31–39,49] The algorithms most often used at present are those using either

1. a fragmental approach initiated by Rekker et al.[90–94] who recognized that to be able to include araliphatic and aliphatic compounds, a different method had to be employed see also[95];
2. atomic contributions, which form the basis of a computational method proposed by Ghose and Crippen (110 types of covalently bonded atom types).[96]

Rekker's idea was later developed further.[19,97–101] Chou and Jurs launched a computer program called *CLOGP*[102,103]; other programs are also available.[104]

6 DESCRIPTORS FOR ELECTRONIC, POLAR, AND STERIC FACTORS

In 1937, Hammett took the initiative in expressing quantitatively the electron-donating or electron-withdrawing effect of substituents X.[103,105,106] He investigated meta- or para-X-substituted benzyl derivatives, and found that rates of reactions involving the bond-breaking of the benzyl group (k_X) were correlated with ionization constants $K_{a,X}$ of the benzoic acids with the same substitution pattern. Electronic effects can be expressed by the substituent constants σ_{meta} or σ_{para} defined as in Equation (7).

$$\sigma_X = \log K_{a,H}(\text{benzoic acid}) - \log K_{a,X}(\text{X} = \text{substituted benzoic acid}) \tag{7}$$

The rate correlation can be expressed by Equation (8):

$$\log\left(\frac{k_X}{k_H}\right) = \rho\sigma_X \tag{8}$$

where ρ denotes the sensitivity of the reaction to electronic effects.[33,102,103] Since Equation (8) relates

thermodynamic to kinetic data (activation energies), it is a linear free-energy relationship (LFER) like the Brønsted relationship. Many physical properties of substituted benzene derivatives can also be correlated with Hammett's σ parameters.

In addition to electronic effects transmitted through space or via bonds, polar interactions result either from polarization or polarizability of molecules. *Polarization* is a permanent effect due to electrical charges or dipole moments. As an example, the greatest difference in normal boiling points for two homologous molecules differing only by one methylene (CH_2 group) is observed for oxalonitrile (cyanogen), which boils at $-22\,^{\circ}C$, and malononitrile, which boils at $219\,^{\circ}C$ under normal pressure; the explanation lies in the fact that the former dinitrile is linear and symmetric and has no dipole moment, whereas the latter dinitrile is V-shaped and has a high dipole moment (3.58 D).[107] *Polarizability* is due to the dynamic response of electrons in molecules and gives rise to dispersion forces. These forces can best be described by molar refractivities (MRs), which have an electronic component proportional to the polarizability, and are obtained as shown in Equation (9) from the refractive indices n and the molar volumes V (hence they are also related to steric parameters).

$$MR = \frac{V(n^2 - 1)}{n^2 + 2} \tag{9}$$

ortho-Substituted benzoic acids and benzylic derivatives do not behave normally in Hammett-type correlations because of steric effects. Taft, therefore, later introduced his steric parameter E_s.[108] Subsequent developments included the separation of inductive effects from resonance effects, the extension of the Hammett equation to polycyclic, polysubstituted, or heterocyclic aromatics, the definition of various scales for electrically charged substituents or reagents, and the use of nonaromatic scaffolds linking the substituent with the reaction center (such as four-substituted bicyclo[2.2.2]octane-carboxylic acids).[109,110]

Other steric parameters have been introduced by Charton[110] and by Verloop et al.[111] Verloop et al. introduced five parameters for substituents, called *STERIMOL*; the first (L) indicates the length of the substituent in the direction of its attachment to the molecular scaffold; the remaining four ($B_1 - B_4$) are dimensions along four directions perpendicular to the first, so that substituent flexibility can also be taken into account. The ratios L/B_1 and B_4/B_1 (the latter being the ratio of maximum to minimum width) measure the dimensions of the substituent when approximated as an ellipsoid. Usually, QSAR equations employing the STERIMOL method involve only these last three parameters. All previously mentioned steric parameters, together or separately, may enter QSAR equations in linear or quadratic polynomials.

Beginning in 1973, Simon and others introduced the concept of *minimal topological difference* (*MTD*) (an earlier version was called *minimal steric difference* (*MSD*)).[6–8] The postulates of this approach are as follows:

1. the binding site of the biological receptor is assumed to be a cavity into which the agonist (drug or ligand molecule), in its graph-theoretical representation neglecting hydrogen atoms fits relatively tightly;
2. atoms or groups of atoms within the cavity are favored sterically, and contribute by lowering the potential energy ($\varepsilon_j = -1$);
3. atoms or groups of atoms interfering with the cavity wall inhibit binding sterically, and increase the total energy ($\varepsilon_j = +1$);
4. atoms or groups of atoms left outside the cavity are of no steric significance, and have no contribution to binding ($\varepsilon_j = 0$).

Mapping several bioactive molecules over one another in their most favorable topology or geometry gives rise to an imaginary hypermolecule. From overlapping vertices with each nonhydrogen atom j of the hypermolecule and agonist i (x_{ij} is 1 when overlapping and 0 otherwise) one can reckon the MTD_i for molecule i as shown in Equation (10) (S is the number of cavity vertices having $\varepsilon_j = -1$ in the hypermolecule):

$$MTD_i = -S + \sum_j \varepsilon_j x_{ij} \tag{10}$$

Simon's approach was the basis for later studies by Hopfinger[112] and Amoore.[113]

7 THE FREE–WILSON (de novo) METHOD

Assuming that biological effects of substituents attached to a molecular scaffold are additive, Free and Wilson[114] started with a series of compounds with the same skeleton, but with various substituents in more than one position, and with each substituent occurring more than once at each position in which it is found. Ban and Fujita reformulated Free and Wilson's method using logarithms of biological activities, and the unsubstituted scaffold as standard.[115] The mathematical background of this QSAR involves indicator variables (0 for the absence and 1 for the presence of a substituent at a particular position) in a matrix, plus the vector of the corresponding potency

values, allowing a multiple regression calculation. This method is subject to the following two restrictions.

1. The minimum number of compounds in the series must be larger than the number $\tilde{N}$ of parameters in the regression equation, i.e. (Equation 11)

$$\tilde{N} = 1 + \sum_j (n_i - 1) \tag{11}$$

   where j is the number of available positions on the scaffold for substitution and n_i is the number of various substituents at position i. Usually, an excess of about five compounds over $\tilde{N}$ yields a minimum number of degrees of freedom for accurate statistical analyses.
2. No two substituents should always be present in combination; if they are, their sum can be considered as a single pseudosubstituent.

An example of antiadrenergic activities of *N*,*N*-dimethyl-α-bromophenethylamines with $j = 2$ positions for para (Y) and/or meta (X) substituents follows.[116,117] The series consisted of 22 compounds, and the $n_i = 5$ substituent positions were occupied by X/Y = H, F, Cl, Br, I, or Me groups. The resulting QSAR equation is shown in Equation (12):

$$\begin{aligned}\log\left(\frac{1}{C}\right) = {} & -0.30(\pm 0.50)[m\text{-F}] + 0.21(\pm 0.29)[m\text{-Cl}] \\ & + 0.43(\pm 0.27)[m\text{-Br}] + 0.58(\pm 0.50)[m\text{-I}] \\ & + 0.45(\pm 0.27)[m\text{-Me}] + 0.34(\pm 0.30)[p\text{-F}] \\ & + 0.77(\pm 0.30)[p\text{-Cl}] + 1.02(\pm 0.30)[p\text{-Br}] \\ & + 1.43(\pm 0.50)[p\text{-I}] + 1.26(\pm 0.33)[p\text{-Me}] \\ & + 7.82(\pm 0.27)\end{aligned} \tag{12}$$

and the statistical data were $n = 22$, $r = 0.969$, $s = 0.194$, $F = 17$.

One shortcoming of the Free–Wilson model is the necessity for series that are large enough to have structural variation in at least two different positions of substitution. A large number of parameters is thus needed to describe relatively few compounds; in the above example, $\tilde{N} = 10$ parameters with 22 compounds leaves only $22 - (10 + 1) = 11$ degrees of freedom. If a substituent occurs only once in the data set, this single-point determination will contribute with its full error to this one biological value. Most significantly, predictions for combinations of substituents that were not present in the series are limited to n_i^j minus the number of degrees of freedom, i.e. $5^2 - 11 = 14$ compounds in this case. No extrapolation to other types of substituents or other substituent positions is allowed.

If the numbers of substituents in positions $a, b, c, \ldots$ are denoted by $n_a, n_b, n_c, \ldots$, the above number of predictions should be replaced by $(n_a + 1)(n_b + 1)(n_c + 1)\cdots - V$, where V is the total number of observed values.

A mixed approach, combining the Free–Wilson and the Hansch methods, solves some of these drawbacks.[118,119] Although log P lipophilicities do not differentiate among stereoisomeric substituents, the Free–Wilson method can take such factors into account. The latter method was also successfully applied to peptide design.[120]

8 TOPOLOGICAL INDICES AND THEIR USES IN QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIPS

The central problem in QSARs is to convert chemical structures into mathematical molecular descriptors that are relevant to the mechanism of the biological activity. Katritzky and Gordeeva.[79] showed that for QSPR studies, TIs are among the best descriptors, but for QSAR studies they have to be complemented with other parameters (such as log P, which is involved in molecular transport across biological membranes).

Recalling what was mentioned in Section 4, one may start from the simplest representation of agonist molecules (potential drugs in CALD) as molecular graphs. Their constitution is encoded in matrix form (adjacency matrix **A**, or distance matrix **D**). Irrespective of the arbitrary labeling of vertices, the sums over row i or column i of these two matrices, v_i and d_i, respectively, are local vertex invariants (LOVIs). The degree v_i of vertex i in organic molecules can be an integer from 1 to 4, but there is no limit for the distasum d_i.

From these or related LOVIs, one can obtain global molecular descriptors, called *TIs*, via mathematical operations such as summation, summation of squares, summation of products over edge endpoints, etc. The simplest *(first-generation)* TIs are integer numbers obtained from these integer LOVIs.[121,122] Thus, one of the first and still useful TIs was proposed by Wiener as the half-sum of distasums, and the result shown in Equation (13), is known as *Wiener's index*[123]:

$$W = \frac{\left(\sum_i d_i\right)}{2} \tag{13}$$

Another first-generation TI is the sum of squares of vertex degrees,[124] shown in Equation (14).

$$M_1 = \sum_i v_i^2 \tag{14}$$

Centric indices for acyclic graphs were obtained by Balaban in 1979 by summing squares of numbers of vertices pruned stepwise from the periphery toward the graph center, which is either one vertex or a pair of adjacent vertices.[125] This idea was later generalized for cyclic graphs,[126] and for substituents.[127]

Second-generation TIs are real numbers based on integer LOVIs. One of the most successful TIs, the molecular connectivity (χ), was proposed by Randic in 1975 as the sum (over all edges $i-j$ in the graph) of the reciprocal square root of products of vertex degrees for edge endpoints (Equation 15), where the LOVIs are vertex degrees v_i and v_j[128]:

$$\chi = \sum_{ij} (v_i v_j)^{-1/2} \tag{15}$$

This idea was then generalized for longer paths than edges (which can be considered as paths of length 1 in the graph). The three carbon atoms in a propane graph form a path of length 2, an *n*-butane graph is a path of length 3, and so on. An isobutane graph is a cluster (star graph) on four vertices. Two Kier–Hall monographs[129,130] presented many successful QSAR data using their extended connectivities; a third book describes their electrotopological state index.[131] All these indices have recently been reviewed.[132–134]

To take into account the nature of atoms symbolized by vertices, Kier and Hall advocated the use of ''valence connectivity indices''.[135] These are calculated using formulae similar to Randic's (Equation 15), but products of edge endpoint (or path–vertex) invariants are no longer vertex degrees but weights (valence delta values δ_i) given by Equation (16):

$$\delta_i = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1} \tag{16}$$

where Z_i^v stands for the number of valence electrons in atom i, Z_i is its atomic number, and H_i is the number of hydrogen atoms attached to atom i.

Kappa indexes developed by Kier are calculated relative to structures that correspond to the least-branched (linear chain) and most-branched (star graph) molecule with the same number A of nonhydrogen atoms. For paths ${}^L P_i$ of length L (i.e. the count of contiguous two-bond fragments), the result is shown in Equation (17).

$${}^L\kappa = \frac{2\left({}^L P_{\max} \times {}^L P_{\min}\right)}{\left({}^L P_i\right)^2} \tag{17}$$

As mentioned earlier, an edge is a path of length $L = 1$. The nature of atoms is taken into account via a corrective factor α, which is derived from the covalent radius of an atom relative to that of an sp^3-hybridized carbon. For ${}^2\kappa_\alpha$ as shown in Equation (18) we obtain

$${}^2\kappa_\alpha = \frac{(A+\alpha-1)(A+\alpha-2)^2}{({}^2P_i+\alpha)^2} \tag{18}$$

Later, Kier introduced a flexibility index,[21,133,136] defined as $\phi = ({}^1\kappa_\alpha \times {}^2\kappa_\alpha)/A$. For a linear alkane, $\phi = A - 1$, i.e. the count of $C(sp^3)-C(sp^3)$ bonds.

No TI proposed so far is free of degeneracy (more than one structure sharing the same value of the index), and all TIs depend both on molecular branching and molecular size. For some TIs it is possible to reconstruct the molecular constitution from the value of its TI if its degeneracy is not too high; this solution of the ''inverse problem'' allows one to make predictions in QSARs.[137,138]

In 1982, Balaban developed the average distance connectivity index (J) based on distasums,[139] to reduce the degeneracy of previous indices (Equation 19).

$$J = \left[\frac{q}{\mu+1}\right] \sum_{ij} (d_i d_j)^{-1/2} \tag{19}$$

The numbers of graph edges, q, and of cycles in the graph, μ, are introduced into Equation (19) to avoid the automatic increase of J with graph size and cyclicity. Indeed, for an infinite linear carbon chain, it was demonstrated that $J = \pi = 3.14159$.[140] The nature of atoms can be taken into account by means of parameters based on their relative atomic numbers, electronegativities, or covalent radii, with respect to those of carbon atoms.[141]

A different approach to lowering the high degeneracy of some TIs such as the Wiener index was to make use of Shannon's formula in information theory, which uses base-two logarithms. The first such ''information theoretic indices'' were proposed by Balaban and Filip,[142] Bonchev,[143] Trinajstic,[144] and Basak et al.[145–147] The former types of indices consider separately the information on each class of summands. The latter indices include all atoms (hydrogens also) in the constitutional formula, and consider the information content provided by various classes of atoms.

Third-generation TIs are based on new, real-number LOVIs and, of course, must also be real numbers. Since the set of integers, although infinite, is smaller than the set of real numbers, such TIs have less chance to degenerate than previous ones. A matrix such as **A** or **D** can be converted into a system of linear equations whose unknowns are real-number LOVIs and whose coefficients are the matrix entries supplemented by two column vectors, which can convey chemical or topological

information; one vector replaces the zeros on the main diagonal, and the other vector provides the free terms of the equations. The result is a ''triplet TI'' denoted by the components of the triplet, namely, the matrix, the main diagonal vector, and the free term vector. Then the LOVIs may be assembled into a TI by various operations. Thus, AZV triplet LOVIs summed over all vertices yield the ''AZV topological index'', which correlates well with some physical properties of alkane hydrocarbons. **A** stands for the adjacency matrix, Z stands for the atomic number of each nonhydrogen atom, and **V** is the free term vector for the vertex degree. The combinatorial nature of the triplet combinations and assembly operations affords a large variety of TIs that may be adapted to different databases.[148]

Another type of LOVI results from applying information theory at the vertex level, providing an informational LOVI; then a formula similar to Randic's index (Equation 15) assembles such LOVIs into four types of TIs denoted by the letters U, V, X, and Y.[149]

The adjacency matrix **A** considers only nearest-neighbor interactions and is identical to the Hückel matrix in quantum chemistry; however, the LOVIs derived from the distance matrix **D** place the largest weight on the more distant vertices. In practice, chemical interactions are strongest between adjacent atoms, but more remote atoms also exert an influence that gradually decreases with the distance, as shown by the through-bond inductive effect. Matrices modeling this feature have been devised. One is the reciprocal distance matrix whose entries are reciprocal values of topological distances; the TIs based on this matrix are called *Harary indices*.[150,151] A different idea is to devise ''regressive distances'' or ''regressive vertex degrees'' by means of nonsymmetrical matrices whose entries in row i represent sums of topological distances or degrees, respectively, for vertices in shells around vertex i; an exponentially decreasing factor multiplies each successive shell. The results are indices that slightly change the distasum or the vertex degree, respectively, according to the molecular topology.[152,153]

Newer ideas for TIs involve the adaptation of TIs to the corresponding database; one can explore the optimal molecular descriptor for that database by allowing the exponent in the formulae for χ or J (Equations 15 and 19) to differ from $-1/2$,[154] or by computing, for each database, a new type of Randic's χ index in which the adjacency matrix has, on its main diagonal, variable numbers for each type of atom instead of zeros.[155]

Fourth-generation TIs are those that include information on stereochemical features of molecules. This type of TI is now being developed, and so far there has been only limited success for diastereomers and enantiomers.[74,156–159]

One should be aware that there are intercorrelations among TIs[160,161]; therefore, one should either eliminate collinearly correlated TIs using the statistical packages available for principal component analysis (PCA),[162] or one should use Randic's orthogonalization procedure for intercorrelated parameters.[163,164]

The great advantage of TIs is their ease of calculation for any imagined structure at low ''CPU (central processing unit) cost''. Many QSPR studies with TIs have been published, and they can be found in reviews[165–167] and in treatises on chemical graph theory.[144,168,169] To facilitate the computation of TIs, commercial programs such as CODESSA, POLLY, or MOLCONN are available.[170–172]

Recalling the lead optimization with 13 molecular descriptors[3] described in Section 1, it is now appropriate to provide details on these descriptors for the amino acids. Four of these descriptors were TIs: three Kier–Hall indices (ϕ, ${}^2\kappa_\alpha$, and χ^v) and Balaban's J index. Other parameters were the lipophilicity (log P) calculated according to Ghose and Crippen,[96] the ellipsoidal and molar volumes, the MR, and the dipole moment. The remaining four descriptors were simple numerical data for the numbers of atoms (oxygen, nitrogen) or groups (ethyl, hydroxyl). The conclusion that one may draw from this approach confirms Katritzky's favorable assessment of the value of TIs used together with other descriptors for QSAR studies,[79] and contradicts criticisms about using TIs in QSARs.[49] If TIs prove their utility for peptides whose stereochemistry is so important, then it is even more likely that they will work (together with other descriptors) for peptidomimetics or any other smaller molecules in drug design.

9 THREE-DIMENSIONAL SHAPE DESCRIPTORS

Surpassing the fourth-generation TIs, a series of shape descriptors has been reported in the literature. The rapidly increasing performance of computers now makes it possible to perform elaborate calculations for quantum-chemical parameters.[168,173] The approach initiated by Mezey has allowed shape simulations to be made for various biologically important molecules such as amino acids or even macromolecular compounds.[79,174–176]

The most promising of all the 3-D/QSAR methods is comparative molecular field analysis (CoMFA).[22,177–180] A 3-D grid is defined so as either to (i) include the most active or the most rigid pharmacophore (ligand), or (ii) model the receptor cavity defined by its molecular docking with the potential bonding sites. Calculations are carried out at each intersection point of the

imaginary grid using force-field (molecular mechanics) methods or Gaussian approximations, Lennard-Jones potentials for noncovalent ligand–receptor interactions, and Coulomb electrostatic attractive/repulsive forces. Statistical evaluations are carried out by using partial least squares (PLS) and PCA.[162] Cross validation is achieved by the leave-one-out procedure from the studied series, and the error is measured by a parameter named *PRESS* (*predicted residual sum squares*). The most difficult problems in CoMFA studies are connected with conformational aspects of flexible molecules and with molecular alignment.

The growing literature on CoMFA has been reviewed, and is too elaborate to be presented here in more detail.[50,53–60,67] For just the brief period between 1993 and 1997, the bibliography on CoMFA has 383 references.[59]

Other descriptors that have been developed include the 3-D/WHIM (three-dimensional/weighted holistic invariant molecular) and G-WHIM (grid-weighted holistic invariant molecular) descriptors,[51] which encompass size, shape, symmetry, and atom distribution, as well as the EVA descriptor,[52] derived from the vibrational frequencies; all of these are independent of structural alignment. Further methods related to CoMFA are HASL (hypothetical active site lattice),[181] and CoMMA (comparative molecular moment analysis), which is independent of the dilemmas raised by molecular superposition.[182] As discussed in section 10, similarities and dissimilarities between aligned molecules can be detected by means of CoMSIA (comparative molecular similarity indices analysis).[60]

10 MOLECULAR SIMILARITY AND DISSIMILARITY

Molecular similarity is important when examining potential candidates for enzyme inhibitors, because inhibitors usually fit into the receptor cavity of the enzyme in the same way as the natural substrate, but their structural difference leads either to irreversible bonding or an intermediate that is no longer able to participate in subsequent reactions. A classical example is that of sulfonamides that mimic *para*-aminobenzoic acid (PAB) and thus act as bactericides. Among the numerous other examples are the anticancer agent 5-fluorouracil, and acyclovir or zidovudine (azidothymidine (AZT)) as anti-AIDS (acquired immune deficiency syndrome) viral agents; in these examples, the drugs mimic pyrimidine or purine bases essential for nucleic acid replication. On the other hand, molecular dissimilarity (diversity) is important when looking for lead compounds in a large database such as the Cambridge Structural Database (X-ray crystallographic data) or the National Cancer Institute Databases, because one wishes to have as much structural variation as possible.

Several books have been dedicated to this topic. TI can also be used for this purpose.[183] Here, we discuss only the more recent developments.

For substructure searches, Burden[184] suggested using molecular identification numbers obtained from the two lowest EVAs (and occasionally eigenvectors) of a connectivity matrix **B** obtained from the hydrogen-depleted graph; the diagonal element b_{ii} is the atomic number of atom i; entries b_{ij} are 0.1, 0.15, 0.2, and 0.3 for single, aromatic, double, and triple bonds, respectively, between connected atoms i and j, and 0.01 otherwise; endpoints (vertices with degree 1) are distinguished by augmenting their b_{ij} entry by 0.01. Over 20 000 structures have been successfully tested and discriminated, with the two lowest EVAs taken to 16 decimal places. A further step was undertaken by Pearlman[185] who proposed the BCUT (Burden–Chemical Abstracts–University of Texas) approach, which consists of taking the lowest and highest EVAs, but using modified matrices: instead of atomic numbers, more relevant data for intermolecular interactions are inserted on the main diagonal, namely, atomic charges, polarizabilities, and H-bond donor and acceptor abilities (corresponding to electrostatic, dispersion forces, and hydrogen bonding, respectively). Using the program CONCORD[186] to generate 3-D structures from 2-D (two-dimensional) structures, it is possible to introduce properly scaled information on the 3-D structure as the off-diagonal entries. With their ''autochoose'' algorithm, Pearlman et al.[186] obtained a robust program for comparing the diversities of large databases at a relatively low CPU cost.

A seductive and fairly fruitful approach to drug design was made possible by rapid advances in hardware. It involves interactive computer graphics,[10–14] and benefits from the fact that color on a computer screen is free, unlike color in print. This is a powerful argument in favor of electronic publishing; the future of information processing has been changed by the computer as much as by Gutenberg several centuries ago; and the ''information superhighway paradox''[187] will probably be solved in the very near future by electronic publishing.

Like 3-D physical models of molecules, computer simulations may be seen, rotated, and moved around, but, in addition, they offer many advantages:

- the geometry (interatomic distances, bond angles, torsion angles) determined by analytical methods can be accurately reflected by the model;
- the molecular energy can be calculated for any real or imagined conformation either by molecular

mechanics (force field) calculations or by semiempirical quantum-chemical methods (ab initio methods are still too CPU-costly for all but small molecules);
- these calculations indicate which is the ground-state conformation and how much energy would be needed to change it so that it would fit the receptor cavity;
- several computer models can be superimposed to afford a 3-D image of the supermolecule and of the excluded or included volume;
- molecular docking may be simulated and dynamic approaches can be programmed.

Among the usual modes of presenting stereodrawings in color, one of the easiest to visualize is to depict the molecular contours with dots (short vectors) because one may see what lies behind the surface and one may include the molecular skeleton inside this surface; alternatively, wire, ball-and-stick, or space-filling molecular models may be imitated. Some of these models are slow to draw and therefore are not appropriate for interactive real-time graphics. Expensive computers already exist, which allow the molecular designer to view molecular models stereoscopically with polarized eyeglasses (as in 3-D movie theaters) and to manipulate them with a trackball. Chemists hope that programmers will soon be able to produce holograms, which could also be handled in print.

11 CASE STUDIES OF QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIPS

Dihydrofolate reductase (DHFR) is an important enzyme, present in all organisms, from bacteria and protozoa to plants, animals, and humans. Folic acid **(1)** is a vitamin for humans, and when it is present in insufficient amounts during pregnancy, severe malformations of the central nervous system occur. It consists of three ''building blocks'': pteridine, PAB, and L-glutamic acid. Bacteria (but not humans) synthesize DHFR using PAB and therefore when sulfonamides (sulfa drugs) are present, these mimic PAB and block the synthesis of **(1)** by acting as selective bactericides. Methotrexate **(2)** is a powerful inhibitor of DHFR, and acts as anticancer agent, preferentially attacking rapidly proliferating cells. The dihydro derivative of folic acid **(3)** is reduced by DHFR, affording tetrahydrofolic acid **(4)**, which is able to react with formaldehyde in vitro or with serine in vivo yielding the five-membered ring intermediate **(5)**. This compound reacts in many single carbon transfer reactions, such as the methylation of deoxyuridylate affording thymidylate (thymine is one of the four bases in DNA), and in amino acid syntheses.

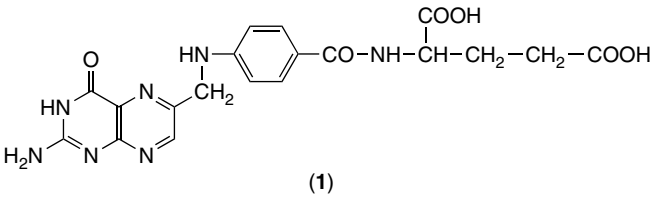

(**1**)

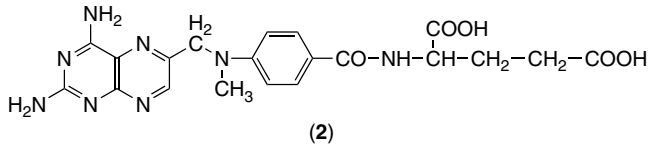

(**2**)

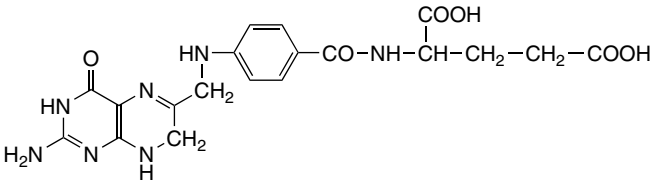

(**3**)

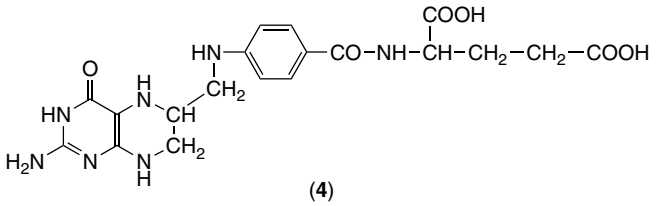

(**4**)

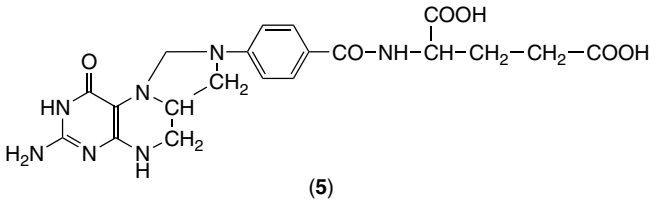

(**5**)

Other medically useful inhibitors of DHFR, in addition to methotrexate, are the antibacterial trimethoprim (**(6)**, $R = \mathrm{OMe}$), the anticoccidial diaveridine (**(6)**, $R = \mathrm{H}$), employed in veterinary practice for ''egg and poultry factories'', and the antimalarial pyrimethamine **(7)**.[14]

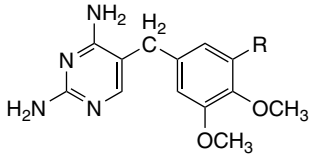

(**6**)

The amino acid homology in the primary structure of DHFR isolated from various species is high when comparing vertebrates among themselves, but low when

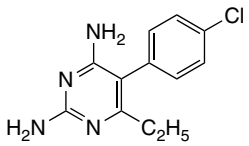

(7)

comparing vertebrates to bacteria or bacteria among themselves. Binding of the substrate occurs through a combination of hydrogen bonding, electrostatic, and hydrophobic forces. The selectivity originates in subtle differences among these forces, which determine the geometry of the association between enzyme and ligand.

Equation (20) is one of the QSAR equations for the inhibition of bovine liver DHFR by trimethoprim analogs[188,189]:

$$\log\left(\frac{1}{C}\right) = 0.62(\pm 0.13)\pi_{\text{meta}} + 0.33(\pm 0.18)\sum\sigma + 4.99(\pm 0.07) \quad (20)$$

where $n = 23$, $r = 0.931$, and $s = 0.146$.

The hydrophobicity term of one meta-substituent of the benzyl group in analogs of **(6)** has the largest coefficient; the $\sum\sigma$ term includes the electronic contributions of substituents in the benzylic ring. For other vertebrates, including humans, the QSAR equations are nonlinear, but not much different. On the other hand, the inhibition of *Escherichia coli* DHFR (a bacterial enzyme) is expressed by Equation (21)[190,191]:

$$\log\left(\frac{1}{K_i}\right) = 1.38(\pm 0.30)\text{MR}_{3,5} + 0.82(\pm 0.35)\text{MR}_4 + 5.77(\pm 0.25) \quad (21)$$

where $n = 23$, $r = 0.918$, and $s = 0.250$.

Reliable and accurate apparent inhibition equilibrium constants K_i express the biological activity; the $\text{MR}_{3,5}$ and MR_4 terms are truncated MR values of the substituents in the meta and para positions of the benzyl ring, respectively.

Equation (22) represents another equation for the same bacterial enzyme[62]:

$$\log\left(\frac{1}{\text{IC}_{50}}\right) = 0.72(\pm 0.27)\sigma_{\text{para}} + 8.85(\pm 0.16) \quad (22)$$

where $n = 11$, $r = 0.892$, and $s = 0.185$

Here the biological activity is expressed by the less accurate IC_{50} (50% inhibitory concentration) values. The differences between Equations (20), (21), and (22) reflect the fact that hydrophobic interactions are more important in vertebrate DHFRs, whereas polar interactions are more important in bacterial DHFRs.[14,62]

12 QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIPS IN ENVIRONMENTAL CHEMISTRY AND TOXICOLOGY

The toxicity of new chemicals is an integral part of drug design, since the therapeutic ratio contains toxicity in the denominator. In addition, the toxicity of chemicals released into the environment has become an important problem for present day civilizations. One may recall the case of organomercurials concentrated by marine organisms that led to casualties in Japan, the ''Silent Spring'' caused by DDT (1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)-ethane) and emotionally described by Rachel Carson in her well-known book, or the chemical accidents in Seveso (Italy) and Bhopal (India). The polychloro-dibenzodioxins released in the Seveso accident, which were also present in the defoliant (Agent Orange) used in the Vietnam war, as well as methylisocyanate, which caused many fatalities in Bhopal, are extremely and acutely toxic compounds. The less acutely toxic carcinogenic polycyclic aromatic hydrocarbons (PAHs) such as benzo[*b*]pyrene and benz[*a*]anthracene are, however, as dangerous when inhaled for a long time because they undergo oxidation in the liver to the corresponding epoxy-dihydroxy-tetrahydro derivatives at the ''bay-region'' benzenoid ring, and this ultimate carcinogen causes permanent damage to cellular DNA. Cigarette smoke as well as car exhaust gases have a significant content of such carcinogens. The Ames test is a quick means of determining mutagenic properties of chemicals, and it is inferred that there is a good correlation between mutagenic and carcinogenic activity.

The theoretical prediction of toxicities by means of QSARs is no longer wishful thinking, as stressed by Karcher and Devillers,[26] as well as by Basak.[29] Similarities between QSARs in CALD and in environmental sciences exist (Hansch approach, multivariate techniques) but the differences are more marked. In the latter sciences, the analyzed response is usually for the whole organism (although the target enzyme or organ is also relevant), the receptor and mechanism are seldom known, and the mechanism of action may be either specific or nonspecific. The aims of QSAR studies in these two fields are, of course, quite different: to find new drugs for human or veterinary use on the one hand and, on the other hand, to identify (and find ways of coping with) toxic wastes or to design pesticides.

Randic's TI $^1\chi$ was able to model soil sorption of PAHs by means of the ratio (K_{om}) between organic matter sorbed by the soil and dissolved in soil water, leading to

the QSAR shown in Equation (23)[28]

$$\log K_{om} = 0.53^1\chi + 0.43 \quad (23)$$

where $n = 72$, $r = 0.977$, $s = 0.282$, $F = 1478$, and $EV = 95.4\%$.

An early QSAR study of the toxicity (50% lethal dose, LD_{50}) of phenols for mice is shown in Equation (24)[192]:

$$\log\left(\frac{1}{C}\right) = 0.44(\pm 0.06)C\log P + 1.76(\pm 0.17) \quad (24)$$

where $n = 25$ and $r^2 = 0.894$.

Though the mechanism by which mice are killed is not clear, Equation (24) confirms Ferguson's principle mentioned in Section 5.

For the toxicity of aniline derivatives on mouse embryo fibroblast cells, the QSAR shown in Equation (25) was found[193]:

$$\log\left(\frac{1}{C}\right) = -0.55(\pm 0.19)\sigma^+ - 0.22(\pm 0.15)\log P + 4.07(\pm 0.222) \quad (25)$$

where $n = 20$ and $r^2 = 0.849$.

In the last two cases, comparison with QSAR equations for organic reactions provides indirect support that free radicals may be involved in the toxic effect.

A much-studied topic deals with aquatic toxicity for vertebrates such as fish (many studies involved fathead minnows, *Pimephales promelas*), single-cell eukaryotes such as *Tetrahymena pyriformis*, or gram-negative prokaryotes such as *Vibrio fischeri* (formerly *Photobacterium phosphoreum*). One can conduct either static tests in solutions of various concentrations, or dynamic (flow) tests, for diverse time intervals. Narcosis or respiratory uncoupling caused by phenols or anilines, and acute toxicity caused by various congeneric chemicals, have been effectively modeled by QSAR equations. In most cases, a large part of the variance is explained by log P values, indicating that the transport across biomembranes must be an important factor.

An example for acute toxicities (50% lethal concentration (LC_{50})) with rainbow trout for 96-h exposure to various pesticides such as orthophosphates, carbamates, aromatic halides is shown in Equation (26)[26]:

$$\log\left(\frac{1}{C}\right) = 0.288\log P + 1.56 \quad (26)$$

where $n = 14$, $r = 0.902$, $s = 0.195$, $F = 52.5$, and $P < 0.01\%$.

Mekenyan et al. considered a range of conformations for flexible molecules and devised QSAR equations for polychlorinated biphenyls, dibenzofurans, and dibenzo-*p*-dioxins in terms of quantum-chemical, topological, geometrical, and hydrophobic parameters.[76,194] For multiple conformations, the COREPA (common reactivity pattern) approach can be applied not only to toxicophores but also to pharmacophores.[195,196]

Experimental data for the basic toxicity of about 1500 high-production-volume chemicals is being carried out by about 200 US chemical and oil companies at high cost; the US Environmental Protection Agency has emphasized the need for observing new protocols that spare test animals. Evidently, any theoretical method that would provide reasonable guesses for the toxicity of the many thousands of chemicals produced in smaller amounts would spare considerable funds and effort.

13 STATISTICAL DATA AND COMPUTATIONAL APPROACHES

A few relevant literature data on chemometrics and statistics are provided briefly.[197] As mentioned earlier, PCA multiple regression and PLS are the methods of choice. Cluster analysis and pattern recognition techniques help in detecting similarity or diversity in databases of chemical structures. Two such techniques are the *k*-nearest neighbor (*k*-NN) method, and a class modeling technique called *SIMCA* (*similarity, chemistry, and analogy*), which places objects from a high-dimensional space into lower dimension boxes.[39] The DARC-PELCO method[168,169,198] originated from the idea that information is structured in successive layers around a focal point, and developed corresponding algorithms. The autocorrelation method was applied successfully by Devillers in QSAR for molecular modeling and in environmental sciences.[198]

For a QSAR that varies nonlinearly with structural parameters, newer and very efficient methods are artificial neural networks[199–203] and genetic algorithms,[204] which can be applied either separately, or more efficiently together.

14 CONCLUSIONS

Computational chemistry has reached a level of trustworthiness by producing valuable results in medicinal chemistry. Success stories for drug design were few until 20 years ago[205–207]; an early discovery based on QSARs was pralidoxime (2-pyridine-aldoxime methiodide, **(8)**, an acetylcholinesterase inhibitor, which removes phosphate from the serine residue of acetylcholinesterase in cases of

organophosphate or nerve gas poisoning). QSARs also contributed to the discovery of cimetidine **(9)**, as well as its furanic analog ranitidine **(10)** both of which are used for treating gastric ulcers as histamine agonists for H_2 receptors. Nowadays, the medicinal drugs obtained by means of QSAR, CALD, and molecular modeling studies are so numerous that even a nonexhaustive enumeration would be too long to include here. Only three newer examples of commercial drugs found with the contribution of QSAR studies will be included: the antibiotic norfloxacin **(11)**, the antihypertensive agent losartan **(12)**, and the antiviral drug indinavir **(13)** used for treating autoimmune deficiency syndrome, all benefited from the results of molecular mechanics and molecular simulation simulations based on X-ray diffractometry data.

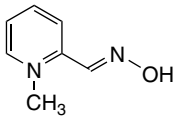

(**8**)

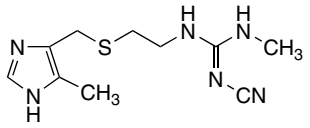

(**9**)

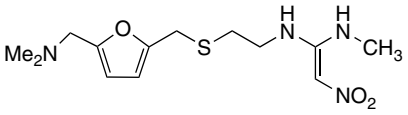

(**10**)

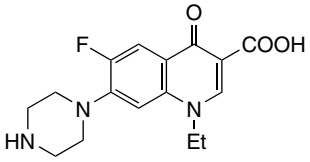

(**11**)

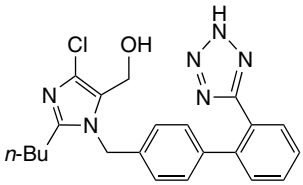

(**12**)

Nowadays, every major pharmaceutical company has CALD (molecular modeling and computer-assisted ligand design) and combinatorial chemistry sections,

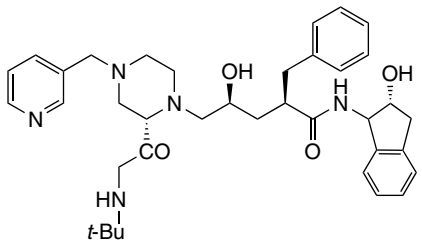

(**13**)

as well as proprietary databases. The tools of the trade are

1. molecular graphics and data visualization;
2. quantum chemistry at the ab initio and semiempirical levels MNDO (modified neglect of differential overlap), AM1 (Austin Method 1), and PM3 (parameterized method 3) for obtaining atomic charges, frontier orbital energies, and super-delocalizabilities;
3. molecular mechanics based on force fields and molecular dynamics;
4. molecular modeling and conformational searching on the potential energy hypersurface;
5. molecular structure databases for 2-D and 3-D structures or substructures, and criteria for similarity or dissimilarity;
6. QSARs and QSPRs using all the available panoply of molecular descriptors, from TIs to de novo (Free–Wilson) and Hansch's lipophilicity approaches, alternating computational and experimental stages for the determination of bioactivities in order to gradually improve the QSARs;
7. pharmacophore and receptor analysis using advanced statistical methods, such as artificial neural networks and genetic algorithms.

Hansch et al. recently wrote a fascinating book chapter, which should become a citation classic, since it opens many ''doors of perception'' and has many paragraphs and data worthy of being quoted.[63] In brief, the authors describe how their database of about 10 000 QSAR data (of which one-third are for biological/enzymatic systems and two-thirds for organic reactions) can be used to understand the chemistry of living cells, and thus to devise better drugs or pesticides. This chapter could become a gold mine for medicinal chemists and pharmacologists.

A few quotes from this last chapter (reproduced with permission of Taylor & Francis) are in order:

> The advent of ''sexy'' 3-D pictures of ligands bound to enzymes of established structure captured researchers' attention the way that the Lorelei of old entranced the sailors on the Rhein. Mechanism based on physical organic chemistry was forgotten .... Can SAR

> (structure–activity relationship) problems be solved de novo with 3-D pictures? Our belief is that mechanistic physical organic chemistry is on the verge of an exciting new venture in helping to elucidate chemico-biological interactions that will do vastly more than simply justify its existence to make only incremental advances in our study of the chemistry of life.

The authors of this chapter indicate how to mine their rich database for relevant, meaningful, and manageable information. Then, they show by example what one can learn from such data. By grouping together QSAR equations in terms of electronic and/or lipophilicity parameters, it becomes possible to gain insight into mechanisms of enzymatic reactions. Thus, the *signs and values* of the *slopes and intercepts* in such equations are powerful indicators of the similarities and differences between complicated biochemical reactions and simple reactions that are well understood mechanistically. As biological QSARs are more difficult and expensive (and frequently less precise and with smaller numbers of data) than physical–organic QSARs, ''lateral'' support by the latter can add confidence in the results of the former studies. Bridges must be built between these two strictly compartmentalized areas of research. Numerous examples are presented and discussed, and several suggestions for plausible mechanisms and for promising classes of new drugs are made in this chapter.

It seems appropriate to conclude with a few other quotations from this chapter:

> The rather abstract, ''boring'', multivariate equations have left many who stopped to think of the enormous complexity of the problem of formulating QSAR from a set of chemicals perturbing a cell or a mouse with the feeling that they were watching a con game ....
> The 3-D graphics distracted many from trying to relate biological QSAR to the much better understood mechanistic studies from the fields of physical, organic, and biochemistry. But the enormous driving force to understand how chemicals react with the various forms of life (and its constituent parts) and the economic prospects in terms of designing the billion-dollar drug as well as understanding toxicology, have been relentless encouragement for more and more researchers to attempt what some still regard as impossible .... Because there are virtually an unlimited number of organic chemicals and a huge number of biological systems (DNA, enzymes, organelles, cells, and whole organisms) with which they can interact, we must develop generalizations, limited as they may be, to be used in the planning stage of synthesis of the myriad chemicals we seem to need to facilitate our existence .... Our present system is a small start on the problem of the design and construction of a computerized means of keeping account of what has been done and how it can be used, but we are confident that it will grow rapidly.

ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immune Deficiency Syndrome
AM1	Austin Method 1
AZT	Azidothymidine
CADD	Computer-assisted Drug Design
CALD	Computer-aided Ligand Design
CAMD	Computer-aided Molecular Design
CAS	Chemical Abstracts Service
CoMFA	Comparative Molecular Field Analysis
CoMMA	Comparative Molecular Moment Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
COREPA	Common Reactivity Pattern
CPU	Central Processing Unit
DDT	1,1,1-Trichloro-2,2-bis(*p*-chloro-phenyl)-ethane
DHFR	Dihydrofolate Reductase
EVA	Eigenvalue
G-WHIM	Grid-weighted Holistic Invariant Molecular
HASL	Hypothetical Active Site Lattice
HPLC	High-performance Liquid Chromatography
IC_{50}	50% Inhibitory Concentration
k-NN	*k*-Nearest Neighbor
LC_{50}	50% Lethal Concentration
LD_{50}	50% Lethal Dose
LFER	Linear Free-energy Relationship
LOVI	Local Vertex Invariant
MNDO	Modified Neglect of Differential Overlap
MR	Molar Refractivity
MSD	Minimal Steric Difference
MTD	Minimal Topological Difference
NMR	Nuclear Magnetic Resonance
PAB	*para*-Aminobenzoic Acid
PAH	Polycyclic Aromatic Hydrocarbon
PCA	Principal Component Analysis
PLS	Partial Least Squares
PM3	Parameterized Method 3
PRESS	Predicted Residual Sum Squares
QSAR	Quantitative Structure–Activity Relationship
QSPR	Quantitative Structure–Property Relationship
SAR	Structure–Activity Relationship
SIMCA	Similarity, Chemistry, and Analogy
TI	Topological Index
UV	Ultraviolet
2-D	Two-dimensional
3-D	Three-dimensional

3-D/WHIM Three-dimensional/Weighted Holistic Invariant Molecular

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Immunochemistry

Peptides and Proteins **(Volume 7)**
Molecular Modeling in Peptide and Protein Analysis • Protein–Drug Interactions

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Combinatorial Chemistry Libraries, Analysis of

Chemometrics **(Volume 11)**
Chemometrics

FURTHER READING

M.A. Johnson, G.M. Maggiora, eds, *Concepts and Applications of Molecular Similarity*, John Wiley & Sons, New York, 1990.

G.M. Maggiora, M.A. Johnson, M.S. Lajiness, A.B. Miller, T.R. Hagadone, 'Looking for Buried Treasures: The Search for New Drug Leads in Large Chemical Databases', *Math. Comput. Model.*, **11**, 626–629 (1988).

A.C. Wood, '3D Similarity Indices and Their Application in QSAR Studies', in *Molecular Similarity in Drug Design*, ed. P. Dean, Chapman & Hall, New York, 24–56, 1995.

P. Willett, *Similarity and Clustering in Chemical Information Systems*, Research Studies Press, Letchworth, 1987.

T.I. Oprea, ed. *Chemoinformatics in Drug Discovery*, Wiley-VCH, Weinheim, 2005.

A.R. Leach, V.J. Gillet, *An Introduction to Chemoinformatics*, Kluwer, Dordrecht, 2005.

W. Jahnke, D.A. Erlanson, eds. *Fragment-based Approaches to Drug Discovery*, Wiley-VCH, Weinheim, 2006.

S. Ekins, ed. *Computational Toxicology. Risk Assessment for Pharmaceutical and Environmental Chemicals*, Wiley-Interscience, Hoboken, New Jersey, 2007.

REFERENCES

1. J.R. Chrétien, 'The State of the Art in Chemometrics and Analytical Chemistry', *Anal. Bioanal. Chem.*, **372**, 511–512 (2002).
2. H. van de Waterbeemd, ed, 'Advanced Computer-assisted Techniques in Drug Discovery', in *Methods and Principles in Medicinal Chemistry*, VCH Publishers, New York, Vol. 3, 1995.
3. G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floc'h, R. Buelow, 'Computer-assisted Rational Design of Immunosuppressive Compounds', *Nat. Biotechnol.*, **16**, 748–752 (1998).
4. R. Franke, *Theoretical Drug Design Methods*, Elsevier, Amsterdam, 1984.
5. Y.C. Martin, *Quantitative Drug Design: A Critical Introduction*, Marcel Dekker, New York, 1978.
6. A.T. Balaban, A. Chiriac, I. Motoc, Z. Simon, *Steric Fit in QSAR*, *Lecture Notes in Chemistry*, Springer, Berlin, Vol. 15, 1980.
7. Z. Simon, A. Chiriac, S. Holban, D. Ciubotariu, G. I. Mihalas, *Minimum Steric Difference: The MTD Method for QSAR Studies*, Research Studies Press, Letchworth, John Wiley & Sons, New York, 1984.
8. H.D. Holtje, G. Folkers, eds, 'Molecular Modeling. Basic Principles and Applications', in *Methods and Principles in Medicinal Chemistry*, VCH Publishers, New York, Vol. 5, 1996.
9. T. Nogrady, *Medicinal Chemistry: A Biochemical Approach*, Oxford University Press, New York, 375–394, Chapter 8, 1985.
10. K. Prout, 'Introduction: Molecular Graphics and Molecular Design', in *Molecular Graphics and Drug Design*, *Topics in Molecular Pharmacology*, eds. A.S.V. Burgen, G.C.K. Roberts, M.S. Tute, Elsevier, Amsterdam, 1–14, Vol. 3, 1986.
11. J.G. Vinter, 'Molecular Modelling Systems', in *Molecular Graphics and Drug Design*, *Topics in Molecular Pharmacology*, eds. A.S.V. Burgen, G.C.K. Roberts, M.S. Tute, Elsevier, Amsterdam, 15–56, Vol. 3, 1986.
12. G.R. Marshall, I. Motoc, 'Approaches to the Conformation of the Drug Bound to the Receptor', in *Molecular Graphics and Drug Design*, *Topics in Molecular Pharmacology*, eds. A.S.V. Burgen, G.C.K. Roberts, M.S. Tute, Elsevier, Amsterdam, 115–156, Vol. 3, 1986.
13. M. Mabilia, R.A. Pearlsetin, A.J. Hopfinger, 'Computer Graphics in Molecular Shape Analysis', in *Molecular Graphics and Drug Design*, *Topics in Molecular Pharmacology*, eds. A.S.V. Burgen, G.C.K. Roberts, M.S. Tute, Elsevier, Amsterdam, 157–182, Vol. 3, 1986.
14. J.N. Champness, L.F. Kuyper, C.R. Beddell, 'Interaction between Dihydrofolate Reductase and Certain Inhibitors', in *Molecular Graphics and Drug Design*, *Topics in Molecular Pharmacology*, eds. A.S.V. Burgen, G.C.K. Roberts, M.S. Tute, Elsevier, Amsterdam, 335–362, Vol. 3, 1986.
15. E.J. Ariëns, 'QSAR Methods and Stereoselectivity in Action', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 3–6, 1989.
16. J.C. Dearden, 'QSAR–A European Viewpoint', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 7–14, 1989.

17. S.H. Unger, 'Consequences of the Hansch Paradigm', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 15–22, 1989.

18. C. Hansch, 'Comparative Structure–Activity Relationships', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 23–30, 1989.

19. A.J. Leo, 'Increasing the Accuracy of Computer Calculated log P (Octanol/Water) Values', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 53–58, 1989.

20. R. Hyde, 'QSAR Parameters à la Carte from Computer Chemistry', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 91–96, 1989.

21. L.B. Kier, 'An Index of Flexibility from Molecular Shape Descriptors', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 105–110, 1989.

22. R.D. Cramer III, D.E. Patterson, J.D. Bunce, 'Recent Advances in Comparative Molecular Field Analysis (CoMFA)', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 161–166, 1989.

23. Y.C. Martin, E.B. Dahne, C.S. May, D. Weininger, J.H. van Drie, 'Strategies in Drug Design Based on 3D-Structures of Ligands', in *QSAR: Quantitative Structure–Activity Relationships in Drug Design*, ed. J.L. Fauchère, Alan R. Liss, New York, 177–182, 1989.

24. D. Hadzi, B. Jerman-Blazic, eds, *QSAR in Drug Design and Toxicology*, *Pharmacochemistry Library*, Elsevier, Amsterdam, Vol. 10, 1987.

25. J.L.M. Hermens, A. Opperhuizen, eds, *QSAR in Environmental Toxicology–IV*, Elsevier, Amsterdam, 1991.

26. W. Karcher, J. Devillers, 'SAR and QSAR in Environmental Chemistry and Toxicology: Scientific Tool or Wishful Thinking?' in *Practical Applications of QSAR in Environmental Chemistry and Toxicology*, eds. W. Karcher, J. Devillers, Kluwer Academic Publishers, Dordrecht, 1–12, 1990.

27. J.C. Dearden, 'Physico-chemical Descriptors', in *Practical Applications of QSAR in Environmental Chemistry and Toxicology*, eds. W. Karcher, J. Devillers, Kluwer Academic Publishers, Dordrecht, 25–60, 1990.

28. A. Sabljic, 'Topological Indices and Environmental Chemistry', in *Practical Applications of QSAR in Environmental Chemistry and Toxicology*, eds. W. Karcher, J. Devillers, Kluwer Academic Publishers, Dordrecht, 61–82, 1990.

29. S.C. Basak, 'A Nonempirical Approach to Predicting Molecular Properties Using Graph-theoretic Invariants', in *Practical Applications of QSAR in Environmental Chemistry and Toxicology*, eds. W. Karcher, J. Devillers, Kluwer Academic Publishers, Dordrecht, 83–104, 1990.

30. J. Devillers, R.L. Lipnick, 'Practical Applications of Regression Analysis in Environmental QSAR Studies', in *Practical Applications of QSAR in Environmental Chemistry and Toxicology*, eds. W. Karcher, J. Devillers, Kluwer Academic Publishers, Dordrecht, 129–144, 1990.

31. M.S. Tute, 'History and Objectives of Quantitative Drug Design', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 1–32, Vol. 4, 1990.

32. D. Weininger, J.L. Weininger, 'Chemical Structures and Computers', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 59–83, Vol. 4, 1990.

33. K. Bowden, 'Electronic Effects in Drugs', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 205–240, Vol. 4, 1990.

34. P.J. Taylor, 'Hydrophobic Properties of Drugs', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 241–290, Vol. 4, 1990.

35. A.J. Leo, 'Methods of Calculating Partition Coefficients', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 295–320, Vol. 4, 1990.

36. Y.C. Martin, E.N. Bush, J.J. Kyncl, 'Quantitative Description of Biological Activity', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 349–374, Vol. 4, 1990.

37. T. Fujita, 'The Extra Thermodynamic Approach to Drug Design', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 497–560, Vol. 4, 1990.

38. H. Kubinyi, 'The Free–Wilson Method and Its Relationship to the Extrathermodynamic Approach', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 589–644, Vol. 4, 1990.

39. W.J. Dunn III, S. Wold, 'Pattern Recognition Techniques in Drug Design', in *Comprehensive Medicinal Chemistry*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 674–891, Vol. 4, 1990.

40. C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, *ACS Professional Reference Book*, American Chemical Society, Washington, DC, 1995.

41. C. Hansch, A. Leo, D. Hoekman, *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*, *ACS*

Professional Reference Book, American Chemical Society, Washington, DC, 1995.

42. C. Hansch, T. Fujita, eds, *Classical and Three-dimensional QSAR in Agrochemistry*, American Chemical Society, Washington, DC, 1995.
43. P.C. Jurs, S.L. Dixon, L.M. Egolf, 'Representations of Molecules', in *Chemometric Methods in Molecular Design*, *Methods and Principles in Medicinal Chemistry*, ed. H. van de Waterbeemd, VCH Publishers, New York, 15–38, Vol. 2, 1995.
44. L.B. Kier, 'Atom-level Descriptors for QSAR Analyses', in *Chemometric Methods in Molecular Design*, *Methods and Principles in Medicinal Chemistry*, ed. H. van de Waterbeemd, VCH Publishers, New York, 39–48, Vol. 2, 1995.
45. L.B. Hall, 'Total Response Surface Optimization', in *Chemometric Methods in Molecular Design*, *Methods and Principles in Medicinal Chemistry*, ed. H. van de Waterbeemd, VCH Publishers, New York, 91–103, Vol. 2, 1995.
46. R. Franke, A. Gruska, 'Principal Component and Factor Analysis', in *Chemometric Methods in Molecular Design*, *Methods and Principles in Medicinal Chemistry*, ed. H. van de Waterbeemd, VCH Publishers, New York, 113–164, Vol. 2, 1995.
47. S. Wold, 'PLS for Multivariate Molecular Modeling', in *Chemometric Methods in Molecular Design*, *Methods and Principles in Medicinal Chemistry*, ed. H. van de Waterbeemd, VCH Publishers, New York, 195–218, Vol. 2, 1995.
48. S. Clementi, S. Wold, 'How to Choose the Proper Statistical Method', in *Chemometric Methods in Molecular Design*, *Methods and Principles in Medicinal Chemistry*, ed. H. van de Waterbeemd, VCH Publishers, New York, 319–338, Vol. 2, 1995.
49. H. Kubinyi, *QSAR: Hansch Analysis and Related Approaches*, *Methods and Principles in Medicinal Chemistry*, VCH Publishers, New York, Vol. 1, 1993.
50. R.D. Clark, A.M. Ferguson, R.D. Cramer, 'Bioisosterism and Molecular Diversity', in *3D QSAR in Drug Design: Ligand–Protein Interactions and Molecular Similarity*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, Vols. 9–11, 1998.
51. R. Todeschini, P. Gramatica, 'The WHIM Theory and QSAR Applications', in *3D QSAR in Drug Design: Ligand–Protein Interactions and Molecular Similarity*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 355–380, Vols. 9–11, 1998.
52. T.W. Heritage, A.M. Ferguson, D.B. Turner, P. Willett, 'EVA: A Novel Theoretical Descriptor for QSAR Studies', in *3D QSAR in Drug Design: Ligand–Protein Interactions and Molecular Similarity*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 381–398, Vols. 9–11, 1998.
53. Y.C. Martin, '3D QSAR: Current State, Scope, and Limitations', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 3–23, Vols. 12–14, 1998.
54. U. Norinder, 'Recent Progress in CoMFA Methodology and Related Techniques', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 25–39, Vols. 12–14, 1998.
55. R.T. Kroemer, P. Hecht, S. Guessregen, K.R. Liedl, 'Improving the Predictive Quality of CoMFA Models', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 41–56, Vols. 12–14, 1998.
56. W.J. Dunn, A.J. Hopfinger, '3D QSAR of Flexible Molecules Using Tensor Representation', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 167–182, Vols. 12–14, 1998.
57. E.A. Coats, 'The CoMFA Steroids as Benchmark Dataset for Development of 3D QSAR Methods', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 199–213, Vols. 12–14, 1998.
58. K.H. Kim, G. Greco, E. Novellino, 'Critical Review of Recent CoMFA Applications', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 257–315, Vols. 12–14, 1998.
59. K.H. Kim, 'List of CoMFA References 1993–1996; 1997', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 317–338, Vols. 12–14, 1998.
60. G. Klebe, 'Comparative Molecular Similarity Indices Analysis: CoMSIA', in *3D QSAR in Drug Design: Recent Advances*, *Perspectives in Drug Discovery and Design*, eds. H. Kubinyi, G. Folkers, Y.C. Martin, Kluwer, ESCOM, Dordrecht, 87–104, Vols. 12–14, 1998.
61. J. Devillers, D. Domine, S. Bintein, W. Karcher, 'Comparison of Fish Bioconcentration Models', in *Comparative QSAR*, ed. J. Devillers, Taylor & Francis, Washington, DC, 1–50, 1998.

62. C.D. Selassie, T.E. Klein, 'Comparative QSAR of the Inhibition of Dihydrofolate Reductase', in *Comparative QSAR*, ed. J. Devillers, Taylor & Francis, Washington, DC, 235–280, 1998.

63. C. Hansch, H. Gao, D. Hoekman, 'A Generalized Approach to Comparative QSAR', in *Comparative QSAR*, ed. J. Devillers, Taylor & Francis, Washington, DC, 285–368, 1998.

64. O. Ivanciuc, A.T. Balaban, 'Graph Theory in Chemistry', in *Encyclopedia of Computational Chemistry*, eds. P.V.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 1169–1190, 1998.

65. H. Kubinyi, 'Quantitative Structure–Activity Relationships in Drug Design', in *Encyclopedia of Computational Chemistry*, eds. P.V.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 2309–2320, 1998.

66. P.C. Jurs, 'Quantitative Structure–Property Relationships (QSPR)', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 2320–2330, 1998.

67. H. Kubinyi, 'Comparative Molecular Field Analysis (CoMFA)', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 448–460, 1998.

68. D.B. Boyd, 'Drug Design', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 795–804, 1998.

69. A.J. Leo, 'Octanol/Water Partition Coefficients', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 1960–1968, 1998.

70. J. Devillers, 'Environmental Chemistry: QSAR', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 930–941, 1998.

71. P.G. Mezey, 'Shape Analysis', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 2582–2589, 1998.

72. P. Willett, 'Structural Similarity Measures for Similarity Searching', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 2748–2756, 1998.

73. M. Randic, 'Topological Indices', in *Encyclopedia of Computational Chemistry*, eds. Pv.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, John Wiley & Sons, Chichester, 3018–3032, 1998.

74. A.T. Balaban, 'From Chemical Graphs to 3D Molecular Modeling', in *From Chemical Topology to Three-dimensional Geometry*, ed. A.T. Balaban, Plenum Publishing Corporation, New York, 1–24, Chapter 1, 1997.

75. P.G. Mezey, 'Descriptors of Molecular Shape in 3D', in *From Chemical Topology to Three-dimensional Geometry*, ed. A.T. Balaban, Plenum Publishing Corporation, New York, 25–42, Chapter 2, 1997.

76. O. Mekenyan, G.D. Veith, '3D Molecular Design: Search for Active Conformers in QSAR', in *From Chemical Topology to Three-dimensional Geometry*, ed. A.T. Balaban, Plenum Publishing Corporation, New York, 43–72, Chapter 3, 1997.

77. S.C. Basak, G.D. Grunwald, G.J. Niemi, 'Use of Graph-theoretic and Geometric Molecular Descriptors in Structure–Activity Relationships', in *From Chemical Topology to Three-Dimensional Geometry*, ed. A.T. Balaban, Plenum Publishing Corporation, New York, 73–116, Chapter 4, 1997.

78. M. Randic, M. Razinger, 'On Characterization of 3D Molecular Structure', in *From Chemical Topology to Three-Dimensional Geometry*, ed. A.T. Balaban, Plenum Publishing Corporation, New York, 159–236, Chapter 6, 1997.

79. A.R. Katritzky, E.V. Gordeeva, 'Traditional Topological Indices vs. Electronic, Geometrical and Combined Molecular Descriptors in QSAR/QSPR Research', *J. Chem. Inf. Comput. Sci.*, **33**, 835–857 (1993).

80. H. van de Waterbeemd, ed, *Structure–Property Correlations in Drug Research*, Academic Press, R.G. Landes, Austin, 1996.

81. A.T. Balaban, M. Banciu, V. Ciorba, *Annulenes, Benzo-, Hetero-, Homo-Derivatives and Their Valence Isomers*, CRC Press, Boca Raton, Vols. 1–3, 1986.

82. H.E. Helson, 'Structure Diagram Generation', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, Wiley-VCH, New York, 313–398, Vol. 13, 1999.

83. A.T. Balaban, F. Harary, 'The Characteristic Polynomial Does Not Uniquely Determine the Topology of a Molecule', *J. Chem. Doc.*, **11**, 258–259 (1971).

84. C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, 'Correlation of Biological Activity of Phenoxyacetic Acids with

Hammett Substituent Constants and Partition Coefficients', *Nature*, **194**, 178–180 (1962).

85. C. Hansch, T. Fujita, '$\rho - \sigma - \pi$ Analysis; Method for the Correlation of Activity and Chemical Structure', *J. Am. Chem. Soc.*, **86**, 1616–1626 (1964).
86. T. Fujita, J. Iwasa, C. Hansch, 'A New Substituent Constant, π, Derived from Partition Coefficients', *J. Am. Chem. Soc.*, **86**, 5175–5179 (1964).
87. C. Hansch, 'A Quantitative Approach to Biochemical Structure–Activity Correlations', *Acc. Chem. Res.*, **2**, 232–239 (1969).
88. C. Hansch, J.M. Clayton, 'Lipophilic Character and Biological Activity of Drugs. II. The Parabolic Case', *J. Pharm. Sci.*, **62**, 1–21 (1973).
89. J. McCall, 'Liquid–Liquid Partition Coefficients by High-pressure Liquid Chromatography', *J. Med. Chem.*, **18**, 549–552 (1975).
90. G.G. Nys, R.F. Rekker, 'Statistical Analysis of a Series of Partition Coefficients with Special Reference to the Predictability of Folding of Drug Molecules. The Introduction of Hydrophobic Fragmental Constants (f Values)', *Chim. Ther.*, **9**, 521–535 (1973).
91. G.G. Nys, R.F. Rekker, 'The Concept of Hydrophobic Fragmental Constants (f Values). II. Extension of Its Applicability to the Calculation of Lipophilicities of Aromatic and Heteroaromatic Structures', *Eur. J. Med. Chem.*, **9**, 361–375 (1974).
92. R.F. Rekker, *The Hydrophobic Fragmental Constant*, Elsevier, Amsterdam, 1977.
93. R.F. Rekker, R. Mannhold, *'Calculation of Drug Lipophilicity'*, VCH Publishers, Weinheim, 1992.
94. M. Rekker, R.F. Rekker, K. Dross, G. Bijoo, G. de Vries, 'The Lipophilic Behaviour of Organic Compounds. 1. An Updating of the Hydrophobic Fragmental Approach', *Quant. Struct.-Act. Relat.*, **17**, 517–536 (1998).
95. V. Pliska, B. Testa, H. van de Waterbeemd, eds, *Lipophilicity in Drug Action and Toxicology*, *Methods and Principles in Medicinal Chemistry*, VCH Publishers, New York, Vol. 4, 1996.
96. A.K. Ghose, G.M. Crippen, 'Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure–Activity Relationships. I. Partition Coefficients as a Measure of Hydrophobicity', *J. Comput. Chem.*, **7**, 565–577 (1986).
97. A. Leo, C. Hansch, D. Elkins, 'Partition Coefficients and Their Uses', *Chem. Rev.*, **71**, 525–616 (1971).
98. A. Leo, 'Calculating log P_{oct} from Structure', *Chem. Rev.*, **93**, 1281–1306 (1993).
99. A. Leo, P.Y.C. Jow, C. Silipo, C. Hansch, 'Calculation of Hydrophobic Constant (log P) from π and f Constants', *J. Med. Chem.*, **18**, 865–868 (1975).
100. C. Hansch, A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, New York, 1979.
101. V. Pliska, B. Testa, H. van de Waterbeemd, eds, *Lipophilicity in Drug Action and Toxicology*, VCH Publishers, Weinheim, 1996.
102. J.T. Chou, P.C. Jurs, 'Computer-assisted Computation of Partition Coefficients from Molecular Structures Using Fragment Constants', *J. Chem. Inf. Comput. Sci.*, **19**, 172–178 (1979).
103. MEDCHEM Software, *Release 3.54 Daylight Chemical Information Systems*, Irvine, (VMS version, 1989; UNIX version, 1992).
104. Advanced Chemistry Development, *ACD/LogP Version 0.9*, Toronto, 1995.
105. L.P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, 1940, 1970.
106. C.D. Johnson, *The Hammett Equation*, Cambridge University Press, London, 1973.
107. A.T. Balaban, D. Mills, S.C. Basak, 'Normal Boiling Points of 1,ω-Alkanedinitriles: the Highest Increment in a Homologous Series', *J. Chem. Inf. Comput. Sci.*, **39**, 769–774 (1999).
108. R.W. Taft, 'Separation of Polar, Steric, and Resonance Effects in Reactivity', in *Steric Effects in Organic Chemistry*, ed. M.S. Newman, John Wiley & Sons, New York, 556–675, 1956.
109. O. Exner, 'The Hammett Equation–The Present Position', in *Advances in Linear Free Energy Relationships*, eds. N.B. Chapman, J. Shorter, Plenum Press, New York, 1–70, 1972.
110. M. Charton, 'The Quantitative Treatment of the *ortho* Effect', *Prog. Phys. Org. Chem.*, **8**, 235–318 (1971).
111. A. Verloop, W. Hoogenstraaten, J. Tipker, 'Development and Application of New Steric Substituent Parameters in Drug Design', in *Drug Design*, ed. E.J. Ariëns, Academic Press, New York, 165–207, Vol. 7, 1976.
112. A.J. Hopfinger, 'A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based on Molecular Shape Analysis', *J. Am. Chem. Soc.*, **102**, 7196–7206 (1980).
113. J.E. Amoore, 'Current Status of the Steric Theory of Odor', *Ann. N. Y. Acad. Sci.*, **116**, 456–476 (1964).
114. S.M. Free Jr, J.W. Wilson, 'A Mathematical Contribution to Structure–Activity Studies', *J. Med. Chem.*, **7**, 395–399 (1964).
115. T. Ban, T. Fujita, 'Mathematical Approach to Structure–Activity Study of Sympathomimetic Amines. Norepinephrine Uptake Inhibition', *J. Med. Chem.*, **12**, 353–356 (1969).

116. H. Kubinyi, O.-H. Kehrhahn, 'Quantitative Structure–Activity Relationships. 1. The Modified Free–Wilson Approach', *J. Med. Chem.*, **19**, 578–586 (1976).

117. H. Kubinyi, 'Quantitative Structure–Activity Relationships. 2. A Mixed Approach, Based on Hansch and Free–Wilson Analysis', *J. Med. Chem.*, **19**, 587–600 (1976).

118. A. Cammarata, 'Interrelationship of the Regression Models Used for Structure–Activity Analyses', *J. Med. Chem.*, **15**, 573–577 (1972).

119. J.M. Clayton, W.P. Purcell, 'Hansch and Free–Wilson Analyses of Inhibitory Potencies of Some 1-Decyl-3-carbamyl-piperidines Against Butyrylcholinesterase and Comparison of the Two Methods', *J. Med. Chem.*, **16**, 661–667 (1973).

120. V. Pliska, 'Semiempirical Structure–Activity Relationships in Peptide Pharmacology', in *Perspectives in Peptide Chemistry*, eds. A.N. Eberle, R. Geiger, T. Wieland, Karger Publishers, Basel, 221–235, 1981.

121. A.T. Balaban, O. Ivanciuc, 'Historical Development of Topological Indices', in *Topological Indices and Related Descriptors in QSAR and QSPR*, eds. J. Devillers, A.T. Balaban, Gordon & Breach, The Netherlands, 21–57, 1990.

122. A.T. Balaban, 'Using Real Numbers as Vertex Invariants for Third-generation Topological Indices', *J. Chem. Inf. Comput. Sci.*, **32**, 23–28 (1992).

123. H. Wiener, 'Structural Determination of Paraffin Boiling Points', *J. Am. Chem. Soc.*, **69**, 17–20 (1947).

124. I. Gutman, B. Ruscic, N. Trinajstic, C.F. Wilcox Jr, 'Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes', *J. Chem. Phys.*, **62**, 3399–3405 (1975).

125. A.T. Balaban, 'Chemical Graphs. Part 35. Five New Topological Indices for the Branching of Tree-like Graphs', *Theor. Chim. Acta*, **5**, 239–261 (1979).

126. D. Bonchev, A.T. Balaban, O. Mekenyan, 'Generalization of the Graph Centre Concept and Derived Topological Indices', *J. Chem. Inf. Comput. Sci.*, **20**, 196–213 (1980).

127. A.T. Balaban, S. Bertelsen, S.C. Basak, 'New Centric Topological Indexes for Acyclic Molecules (Trees) and Substituents (Rooted Trees), and Coding of Rooted Trees', *Math. Chem. (MATCH)*, **30**, 55–72 (1994).

128. M. Randic, 'On Characterization of Molecular Branching', *J. Am. Chem. Soc.*, **97**, 6609–6615 (1975).

129. L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.

130. L.B. Kier, L.H. Hall, *Molecular Connectivity in Structure–Activity Studies*, Research Studies Press, Letchworth, 1986.

131. L.B. Kier, L.H. Hall, *The Electrotopological State*, Academic Press, New York, 1999.

132. L.H. Kier, L.B. Hall, 'Molecular Connectivity Chi Indices for Database Analysis and Structure–Property Modeling', in *Topological Indices and Related Descriptors in QSAR and QSPR*, eds. J. Devillers, A.T. Balaban, Gordon & Breach, The Netherlands, 307–360, 1999.

133. L.H. Kier, L.B. Hall, 'The Kappa Indices for Molecular Modeling of Molecular Shape and Flexibility', in *Topological Indices and Related Descriptors in QSAR and QSPR*, eds. J. Devillers, A.T. Balaban, Gordon & Breach, The Netherlands, 455–489, 1999.

134. L.H. Kier, L.B. Hall, 'The Electrotopological State: Structure Modeling for QSAR and Database Analysis', in *Topological Indices and Related Descriptors in QSAR and QSPR*, eds. J. Devillers, A.T. Balaban, Gordon & Breach, The Netherlands, 491–562, 1999.

135. L.B. Kier, L.H. Hall, 'Molecular Connectivity. Part 7. Specific Treatment of Heteroatoms', *J. Pharm. Sci.*, **65**, 1806–1809 (1976).

136. L.B. Kier, 'An Index of Molecular Flexibility from Kappa Shape Attributes', *Quant. Struct.-Act. Relat.*, **8**, 221–224 (1989).

137. N.S. Zefirov, V.A. Palyulin, Yu.V. Radchenko, 'Problem of Generation of Structures with Given Properties–Solution of the Inverse Problem for the Case of the Centric Balaban Index', *Dokl. Akad. Nauk SSSR*, **316**, 921–924 (1991).

138. M.I. Skvortseva, I.I. Baskin, O.L. Slovokhovotova, V.A. Palyulin, N.S. Zefirov, 'Inverse Problem in QSAR/QSPR Indices for the Case of Topological Indices Characterizing Molecular Shape (Kier Indices)', *J. Chem. Inf. Comput. Sci.*, **33**, 630–634 (1993).

139. A.T. Balaban, 'Highly Discriminating Distance-based Topological Index', *Chem. Phys. Lett.*, **80**, 399–404 (1982).

140. A.T. Balaban, 'Topological Indices Based on Topological Distances in Molecular Graphs', *Pure Appl. Chem.*, **55**, 199–206 (1983).

141. A.T. Balaban, 'Chemical Graphs. 48. Topological Index J for Heteroatom-containing Molecules Taking into Account Periodicities of Element Properties', *Math. Chem. (MATCH)*, **21**, 115–122 (1986).

142. A.T. Balaban, P. Filip, 'Computer Program for Topological Index J (Average Distance Sum Connectivity)', *Math. Chem. (MATCH)*, **16**, 163–190 (1984).

143. D. Bonchev, *Information-theoretic Indices for Characterization of Chemical Structure*, Research Studies Press, Letchworth, 1993.

144. N. Trinajstic, *Chemical Graph Theory*, 2nd edition, CRC Press, Boca Raton, 225–273, 1992.

145. S.C. Basak, 'Information-Theoretic Indices of Neighborhood Complexity and Their Application', in *Topological Indices and Related Descriptors in QSAR and QSPR*,

eds. J. Devillers, A.T. Balaban, Gordon & Breach, The Netherlands, 563–593, 1999.

146. S.C. Basak, B.D. Gute, G.D. Grunwald, 'A Hierarchical Approach to the Development of QSAR Models Using Topological, Geometrical and Quantum Chemical Parameters', in *Topological Indices and Related Descriptors in QSAR and QSPR*, eds. J. Devillers, A.T. Balaban, Gordon & Breach, The Netherlands, 675–696, 1999.

147. S.C. Basak, V.R. Magnuson, 'Determining Structural Similarity of Chemicals Using Graph-theoretic Indices', *Discrete Appl. Math.*, **19**, 17–44 (1988).

148. P.A. Filip, T.S. Balaban, A.T. Balaban, 'A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlational Ability', *J. Math. Chem.*, **1**, 61–83 (1987).

149. A.T. Balaban, T.S. Balaban, 'New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances', *J. Math. Chem.*, **8**, 383–397 (1991).

150. O. Ivanciuc, T.S. Balaban, A.T. Balaban, 'Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices', *J. Math. Chem.*, **12**, 309–318 (1993).

151. D. Plavsic, S. Nikolic, N. Trinajstic, Z. Mihalic, 'On the Harary Index for Characterization of Chemical Graphs', *J. Math. Chem.*, **12**, 235–250 (1993).

152. M.V. Diudea, O. Minailiuc, A.T. Balaban, 'Molecular Topology. Part 4. Regressive Vertex Degrees (New Graph Invariants) and Derived Topological Indices', *J. Comput. Chem.*, **12**, 527–535 (1991).

153. A.T. Balaban, M.V. Diudea, 'Real Number Vertex Invariants: Regressive Distance Sums and Related Topological Indices', *J. Chem. Inf. Comput. Sci.*, **33**, 421–428 (1993).

154. O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, and A.T. Balaban, 'Investigation of alkane branching (and resulting partial ordering) by topological indices', *MATCH, Commun. Math. Comput. Chem*. **42**, 155–180 (2000).

155. M. Randic, S.C. Basak, 'Optimal Molecular Descriptors Based on Weighted Path Numbers', *J. Chem. Inf. Comput. Sci.*, **39**, 261–266 (1999).

156. H.P. Schultz, E.B. Schultz, T.P. Schultz, 'Topological Organic Chemistry. Part 9. Graph Theory and Molecular Topological Indices of Stereoisomeric Organic Compounds', *J. Chem. Inf. Comput. Sci.*, **35**, 864–870 (1995).

157. M. Randic, M. Razinger, 'Molecular Topographic Indices', *J. Chem. Inf. Comput. Sci.*, **35**, 130–147 (1995).

158. L. Pogliani, 'On a Graph-theoretical Characterization of Cis/Trans Isomers', *J. Chem. Inf. Comput. Sci.*, **34**, 801–804 (1994).

159. J.V. Julian-Ortiz, C.G. Alapent, I. Rios-Santaorina, R. Garcia-Domenech, J. Galvez, 'Prediction of Properties of Chiral Compounds by Molecular Topology', *J. Mol. Graphics Modell.*, **16**, 14–18 (1998).

160. I. Motoc, A.T. Balaban, 'Topological Indices: Intercorrelations, Physical Meaning, Correlational Ability', *Rev. Roum. Chim.*, **26**, 593–600 (1981).

161. I. Motoc, A.T. Balaban, O. Mekenyan, D. Bonchev, 'Topological Indices: Inter-relations and Composition', *Math. Chem. (MATCH)*, **13**, 369–404 (1982).

162. SAS Institute, *SAS/STAT User's Guide*, Release 6.03 Edition, SAS Institute, Cary, 1988.

163. M. Randic, 'Orthogonal Molecular Descriptors', *New J. Chem.*, **15**, 517–525 (1991).

164. H. Randic, 'Resolution of Ambiguities in Structure–Property Studies by the Use of Orthogonal Molecular Descriptors', *J. Chem. Inf. Comput. Sci.*, **31**, 311–320 (1991).

165. A.T. Balaban, I. Niculescu-Duvaz, Z. Simon, 'Topological Aspects in QSAR for Biologically-active Molecules', *Acta Pharm. Jugosl.*, **37**, 7–36 (1987).

166. N. Voiculetz, A.T. Balaban, I. Niculescu-Duvaz, Z. Simon, *Modeling of Cancer Genesis and Prevention*, CRC Press, Boca Raton, 1990.

167. A.T. Balaban, I. Motoc, D. Bonchev, O. Mekenyan, 'Topological Indices for Structure–Activity Correlations', in *Steric Effects in Drug Design*, *Topics in Current Chemistry*, eds. M. Charton, I. Motoc, Springer-Verlag, Berlin 21–55, Vol. 114, 1983.

168. D.H. Rouvray, 'The Topological Matrix in Quantum Chemistry', in *Chemical Applications of Graph Theory*, ed. A.T. Balaban, Academic Press, London, 175–221, 1976.

169. J.E. Dubois, 'Ordered Chromatic Graphs and Limited Environmental Concepts', in *Chemical Applications of Graph Theory*, ed. A.T. Balaban, Academic Press, London, 330–370, 1976.

170. A.R. Katritzky, V.S. Lobanov, M. Karelson, *CODESSA Manual*, University of Florida, Gainesville, 1995.

171. S.C. Basak, D.K. Harris, V.R. Magnuson, *Program 'POLLY 2.3'*, Copyright of the University of Minnesota, 1988.

172. L.H. Hall, *Programmes MOLCONN and MOLCONN-Z*, Hall Associates Consulting, Quincy, MA, 2001.

173. M. Karelson, V.S. Lobanov, A.R. Katritzky, 'Quantum-chemical Descriptors in QSAR/QSPR Study', *Chem. Rev.*, **96**, 1027–1043 (1996).

174. P. Mezey, *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH Publishers, New York, 1993.

175. P.D. Walker, G.A. Arteca, P.G. Mezey, *'Program GSHAPE 90. Database'*, University of Saskatchewan, Saskatoon, 1990.

176. P.D. Walker, P.G. Mezey, *'Program MEDLA 93. Database'*, University of Saskatchewan, Saskatoon, 1993.

177. R.D. Cramer III, D.E. Patterson, J.D. Bunce 'Comparative Molecular Field Analysis (CoMFA). Part 1. Effect of Shape on Binding of Steroids to Carrier Proteins', *J. Am. Chem. Soc.*, **110**, 5959–5967 (1988).

178. D.E. Patterson, R.D. Cramer, A.M. Ferguson, R.D. Clark, L.E. Weinberger, 'Neighborhood Behavior: A Useful Concept for Validation of Molecular Diversity Descriptors', *J. Med. Chem.*, **39**, 3049–3059 (1996).

179. R.D. Clark, R.D. Cramer, 'Taming the Combinatorial Centipede', *Chemtech*, **27**, 24–30 (1997).

180. Y.C. Martin, K.-H. Kim, C.T. Lin, 'Comparative Molecular Field Analysis: CoMFA', in *Advances in Quantitative Structure–Property Relationships*, ed. M. Charton, JAI Press, Greenwich, 1–52, 1996.

181. A.M. Doweiko, 'Three-Dimensional Pharmacophores from Binding Data', *J. Med. Chem.*, **37**, 1769–1778 (1994).

182. B.D. Silverman, D.E. Platt, 'Comparative Molecular Moment Analysis (CoMMA): 3D QSAR Without Molecular Superposition', *J. Med. Chem.*, **39**, 2129–2140 (1996).

183. S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, 'Determining Structural Similarity of Chemicals Using Graph-theoretic Indices', *Discrete Appl. Math.*, **19**, 17–44 (1988).

184. F.R. Burden, 'Molecular Identification Number for Substructure Searches', *J. Chem. Inf. Comput. Sci.*, **29**, 225–227 (1989).

185. R.S. Pearlman, *'Diverse Solutions User's Manual'*, University of Texas, Austin, 1996.

186. R.S. Pearlman, A. Rusinko, J.M. Skell, R. Balducci, *'CONCORD'* distributed by Tripos International, St. Louis, MO 63144, University of Texas, Austin.

187. A.T. Balaban, 'Mozart, Ford, and the Information Superhighway Paradox', *Am. Sci.*, **85**, 207–209 (1997).

188. R.L. Li, C. Hansch, B.T. Kaufman, 'A Comparison of the Inhibitory Action of 5-(Substituted Benzyl)-2,4-diaminopyrimidines on Dihydrofolate Reductase from Chicken Liver with that from Bovine Liver', *J. Med. Chem.*, **25**, 435–440 (1982).

189. J.M. Blaney, S.W. Dietrich, M.A. Reynolds, C. Hansch, 'Quantitative Structure–Activity Relationship of 5-(X-benzyl)-2,4-diamino Pyrimidines Inhibiting Bovine Liver Dihydrofolate Reductase', *J. Med. Chem.*, **22**, 614–617 (1979).

190. C. Hansch, R.L. Li, J.M. Blaney, R. Langridge, 'Comparison of the Inhibition of *Escherichia coli* and *Lactobacillus casei* Dihydrofolate Reductase by 2,4-Diamino-5-(Substituted Benzyl)-pyrimidines. Quantitative Structure–Activity Relationships. X-ray Crystallography and Computer Graphics in Structure–Activity Analysis', *J. Med. Chem.*, **25**, 777–784 (1982).

191. S.W.Dietrich, J.M.Blaney, M.A.Reynolds, C.Hansch, 'Quantitative Structure–Activity Relationships. Comparison of the Inhibition of *Escherichia coli* and Bovine Liver Dihydrofolate Reductase by 5-(Substituted Benzyl)-2,4-Diaminopyrimidines', *J. Med. Chem.*, **23**, 1205–1212 (1980).

192. G.L. Biagi, O. Gandolfi, M.C. Guerra, A.M. Barbaro, G. Cantelli-Forti, 'R_m Values of Phenols. Their Relationship with log P Values and Activity', *J. Med. Chem.*, **18**, 868–872 (1975).

193. A. Harada, M. Hanazawa, J. Saito, K. Hashimoto, 'Quantitative Analysis of Structure–Toxicity Relationships and Substituted Anilines by Use of Balb./3T3 Cells', *Environ. Toxicol. Chem.*, **11**, 973–980 (1992).

194. O.G. Mekenyan, G.D. Veith, D.J. Call, G.T. Ankley, 'A QSAR Evaluation of *Ah* Receptor Binding of Halogenated Aromatic Xenobiotics', *Environ. Health Perspect.*, **104**, 1302–1310 (1997).

195. O. Mekenyan, J. Ivanov, S. Karabunarliev, S.P. Bradbury, G.T. Ankley, W. Karcher, 'A Computationally-based Hazard Identification Algorithm that Incorporates Ligand Flexibility. 1. Identification of Potential Androgen Receptor Ligands', *Environ. Sci. Technol.*, **31**, 3702–3711 (1997).

196. O. Mekenyan, N. Nikolova, S. Karabunarliev, S.P. Bradbury, G.T. Ankley, B. Hansen, 'New Developments in a Hazard Identification Algorithm for Hormone Receptor Ligands', *Quant. Struct.-Act. Relat.*, **18**, 1–15 (1999).

197. P.C. Jurs, *'Computer Software Applications in Chemistry'*, John Wiley & Sons, New York, 1986.

198. J.E. Dubois, J.P. Doucet, A. Panaye, B.T. Han, *DARC Site Topological Correlations: Ordered Structural Descriptors and Property Evaluation*, 613–673, 1999.

199. J. Devillers, ed, *Neural Networks in QSAR and Drug Design*, Academic Press, London, 1996.

200. J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, New York, 1990.

201. R.C., Eberhart, R.W., Dobbins, eds, *Neural Network PC Tools. A Practical Guide*, Academic Press, San Diego, 1990.

202. P.D.Wasserman, *Neural Computing. Theory and Practice*, Van Nostrand Reinhold, New York, 1989.

203. J. Zupan, J. Gasteiger, *Neural Networks for Chemists. An Introduction*, VCH Publishers, Weinheim, 1993.

204. J.Devillers, ed, *Genetic Algorithms in Molecular Modeling*, Academic Press, London, 1996.

205. D.B. Boyd, 'Successes in Computer-assisted Drug Design', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH Publishers, New York, 355–371, Vol. 1, 1990.

206. D.B. Boyd, 'Compendium of Software for Molecular Modeling', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH Publishers, New York, 303–380, Vol. 7, 1995.

207. D.B. Boyd, 'Molecular Modeling Software in Use: Publication Trends', in *Reviews in Computational Chemistry*, eds. K.B. Lipkowitz, D.B. Boyd, VCH Publishers, New York, 317–354, Vol. 6, 1995.

Robotics and Laboratory Automation in Pharmaceuticals Analysis

Nigel North
SmithKline Beecham Pharmaceuticals, Harlow, UK

1 Introduction 2
2 History 2
3 Pharmaceutical Applications 3
4 System Development and Instrument Qualification 5
5 Method Development 5
6 Method Validation 6
7 Technology Transfer 7
8 Business Benefits 7
Acknowledgments 8
Abbreviations and Acronyms 8
Related Articles 8
References 8

Laboratory robotics is being increasingly applied in pharmaceutical development to help meet the needs of increasing productivity, decreasing drug development time and reducing costs. A robot can be defined as a device that can perform a variable but programmed series of physical manipulations including moving objects, sample identification, weighing, extraction, filtration and dilution. Robots can be classified according to the type of movement the robotic arm can perform. Three of the most common geometries for laboratory robots are: Cartesian (three mutually perpendicular axes); cylindrical (parallel action arm pivoted about a central point); and anthropomorphic (multijointed, human-like configuration). Each has been used in the pharmaceutical industry with varying degrees of success. Cylindrical and anthropomorphic robots generally provide more flexible ''human-like'' automation that includes transfer, weighing, extraction and filtration of samples.

In a typical analysis, four major steps are performed: sampling; sample preparation; sample measurement; and data collection/analysis/reduction to provide a report for the customer. Most laboratory robotic systems are employed to automate the sample preparation stage which is normally the most labor-intensive step of the analysis process. Automation has been applied in pharmaceutical analysis in a number of areas: raw materials release; in-process testing; release of finished products; and stability testing of the dosage form. In pharmaceutical analysis the assay, content uniformity and dissolution tests are the most common methods that use automation. Developing an automated assay or content uniformity method for a dosage form requires several aspects to be investigated before validation of the procedure can proceed, including: extraction; filter; carry-over studies; and stability of solution.

Validation can be defined as establishing documented evidence which provides a high degree of assurance that a specific process will generate a product, in our case an automated method, that will meet predetermined specifications and quality attributes. Qualification is part of validation and has been defined for installation, operation and running of the system under workload for a specific application. In the pharmaceutical industry the ''4Qs'' model is generally used to qualify the unit. Validation of manual analytical methods is well understood but documented in a number of guidelines in slightly different ways. It has been only recently that single documents (International Conference on Harmonisation) have emerged that describe the terminology and methodology for method validation. For robotic methods, equivalency between the manual and automated procedure is an important extra validation test that is often performed since it is unusual not to have a manual already in use. This test may not be relevant if accuracy has been also demonstrated for the automated method. For successful transfer of automated methods, the robotic systems used in research and development (R&D) should mirror those used at the commercial production site. The automated equipment must be rugged and easy to use for the production quality control (QC) laboratory to invest in the technology. Complex robotic systems that need specialist, dedicated staff, have not been generally successful in the QC laboratory.

Robotic systems are a high capital cost and this is often justified by increasing throughput since automatic systems can operate 24 h a day and over weekends unattended, thereby also reducing the cost per analysis. During the 1990s, drug development times have significantly reduced, which has required more responsive ''just-in-time analysis'' and this has been an integral part of cost reduction for the release of raw materials and commercial products. Automated systems also enable scientists to use their time more productively by spending time on evaluating data and on more innovative tasks rather than performing routine repetitive operations.

Although the application of laboratory robotics is increasing, limitations including high capital cost, relative complexity of operation and poor connectivity between the robot and laboratory information management systems, are hindering its implementation.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Laboratory robotics is being increasingly applied in pharmaceutical development to help meet the needs of increasing productivity, decreasing drug development time and reducing costs. A robot can be defined as a device that can perform a variable but programmed series of physical manipulations,[1] including moving objects, sample identification, weighing, extraction, filtration and dilution. Robots can be classified according to the type of movement the robotic arm can perform.[2] Three of the most common geometry's for laboratory robots are: Cartesian (three mutually perpendicular axes); cylindrical (parallel action arm pivoted about a central point); and anthropomorphic (multijointed, human-like configuration). Each has been used in the pharmaceutical industry with varying degrees of success. The PAL system manufactured by CTC Analytics and shown in Figure 1 is an example of a Cartesian robot that performs dedicated relatively simple tasks such as sample dilution, mixing and filling a loop of an autoinjector with a fixed volume of solution for sample analysis. Cylindrical and anthropomorphic robots generally provide more flexible ''human-like'' automation that includes transfer, weighing, extraction and filtration of samples.

In a typical analysis four major steps are performed: sampling; sample preparation; sample measurement; and data collection/analysis/reduction to provide a report for the customer. Most laboratory robotic systems are employed to automate the sample preparation stage which is normally the most labor-intensive step of the analysis process. The results of one survey[3] (Figure 2) indicated that about 60% of the analysis time is spent on sample preparation.

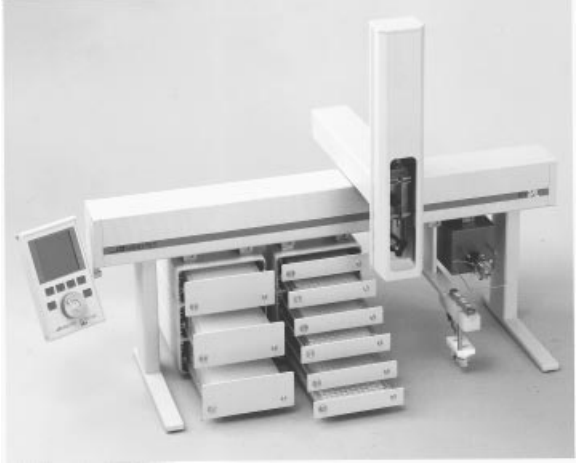

Figure 1 Photograph of a CTC Analytics PAL system. (Reproduced by permission of CTC Analytics AG.)

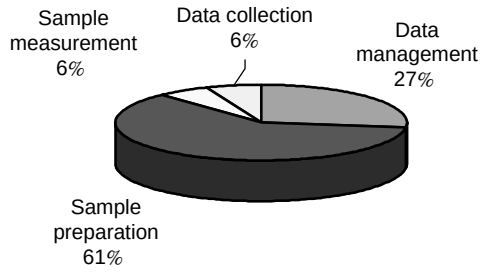

Figure 2 Breakdown of time spent on key steps in sample analysis process.

Automatic analyses often involve similar steps to the manual procedures but robots have an important advantage of performing these repetitive tasks accurately and precisely for long periods of time without the need for breaks. A key benefit is higher productivity,[4] improved quality in terms of increased accuracy and precision of the data generated. Furthermore, automation also frees up analysts' time for important activities such as data interpretation and more innovative tasks. Unattended analysis also reduces the exposure of the scientists to hazardous chemicals.

Although the application of laboratory robotics is increasing, limitations including high capital cost, relative complexity of operation and poor connectivity between the robot and laboratory information management systems, are hindering its implementation.

2 HISTORY

Considerable effort has been devoted over the past decades to increasing the analysis speed and as well as to improving data handling and report generating software. In contrast, automation of sample preparation has lagged behind in its development. Early examples of laboratory automation include continuous flow analyzers,[5] flow injection analysis,[6] and automatic titrators. With the development of inexpensive microelectronics and microcomputers, it is possible to build flexible and more

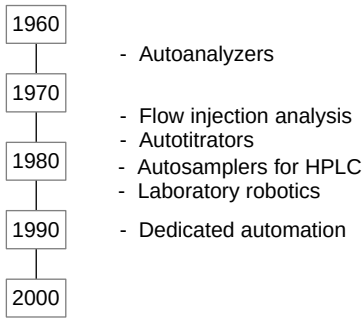

Figure 3 Diagram of the evolution of laboratory automation.

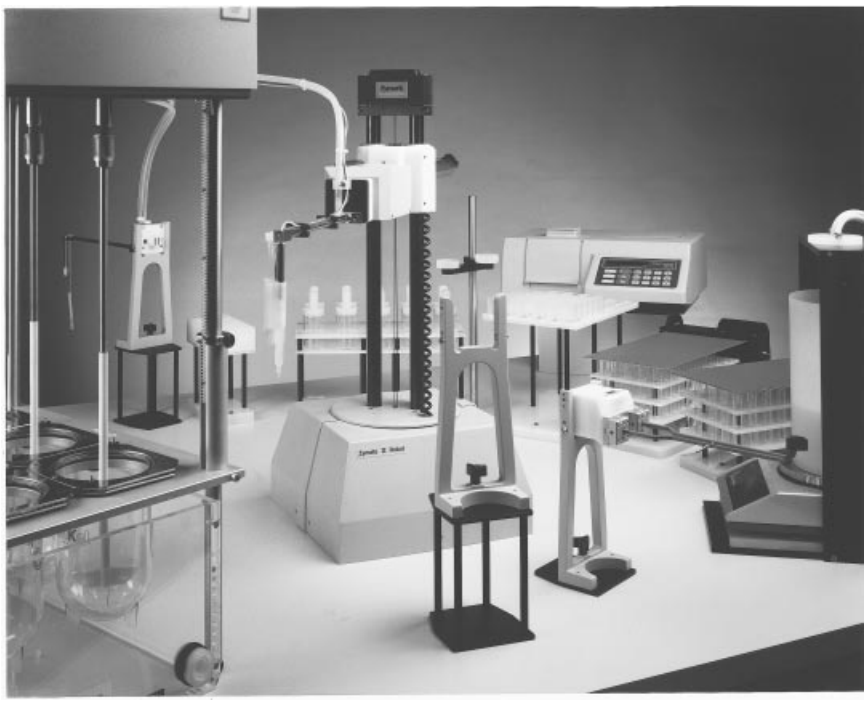

Figure 4 Photograph of Zymate laboratory robot. (Reproduced by permission of Zymark Corporation.)

sophisticated automated systems. These developments have led to programmable laboratory robots that perform a series of physical manipulations using some form of mechanical arm. The deployment of these devices has had a major impact on how routine laboratory tasks are performed. A diagram illustrating the evolution of automation is shown in Figure 3.

The first robotic instruments were introduced into the pharmaceutical laboratory in the early 1980s. The Zymate cylindrical robot (Figure 4) typified these systems. This robot is a modular system that uses its own programming language, scheduler and a range of accessories to automate a wide range of pharmaceutical analytical procedures including dissolution testing and tablet processing. At that time the main competitor was Perkin-Elmer's MasterLab which employed a Mitsubishi robot, an electronics module and associated communications interfaces that enabled the system to be controlled by a high-level menu-driven programming language. This robot used anthropomorphic arm geometry with five axes of motion. During the last 10 years many other manufacturers including CRS Robots, Sands Technology, Source for Automation, Novi and Hewlett-Packard ORCA (Optimized Robot for Chemical Analysis), to name but a few, have developed laboratory robotic systems. Although these robots are laboratory based they are relatively large systems that require their own specially designed bench area and dedicated scientists to develop and maintain the equipment. To overcome these disadvantages and to enhance reliability, smaller, more dedicated systems have now been developed that are simpler robotically and are controlled by easy-to-use menu-driven programs. An example of this smaller, bench-sized robot is Zymark's Benchmate Tablet Processing Workstation. There is now also a trend towards automation that does not use a robotic arm but instead employs dedicated accessories to perform tasks like dissolution testing. A good example of this type of automation is Source for Automation's Automated Dissolution Module. This equipment has the added advantage of being scalable such that single or multiple dissolution units can be linked and controlled by one computer. These smaller automation systems are also more amenable to operating at-line in the production area and this, combined with the recent development of rapid non-invasive methods of analysis such as near infrared, provides the capability of generating analytical results within the manufacturing process.

3 PHARMACEUTICAL APPLICATIONS

Laboratory automation has been applied in pharmaceutical analysis in a number of areas. A summary of the basic steps in the production of a dosage form is given in Figure 5. The individual components that are used to formulate the dosage form need to be tested against a specification before the material can be released for the manufacture of a dosage form. The excipients

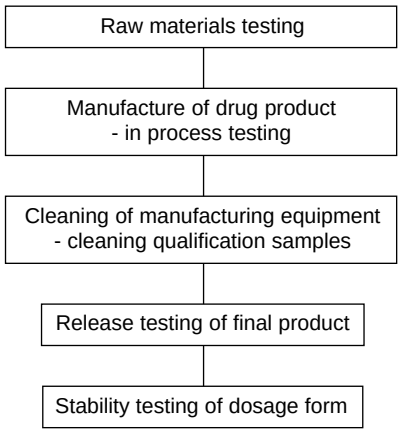

Figure 5 Diagram of the pharmaceutical analysis steps supporting the dosage form production process.

(nonactive ingredients) that give the formulation its fundamental properties, and the drug, must be tested against predefined specifications. Most excipients used have monographs in the United States or European pharmacopeias and involved traditional wet chemical methods. A comprehensive review of the application of flow injection analysis to automating wet chemical methods has published by Betteridge.[7] The heavy metals test, which is a fundamental test for many materials, has been automated using flow injection analysis.[8]

A key element of building quality into the manufacturing of a product is being able to obtain information on the state of the process while it is running. During the 1990s this has started to be addressed in the pharmaceutical industry. Rapid, rugged, reliable and ideally noninvasive analytical techniques are required to provide this information. Conventional laboratory robots have been developed to analyze powders,[9] using high-performance liquid chromatography (HPLC), from blending operations in solid dosage form manufacture.[10] Automation using noninvasive methods such as near infrared has been successfully developed by Sekulic et al.[11] to provide in-process control information. A comprehensive review of the application of near infrared to pharmaceutical analysis has been performed by Blanco et al.[12]

Following the manufacture of a drug product there is a need to ensure the equipment used is clean before the next manufacture can commence. Scypinski et al.[13] automated the analysis of cleaning samples employing the Zymark Benchmate workstation.

Drug products are released by checking the batch against a predefined specification. The most common specification tests include determination of the assay, degradation products, content uniformity and dissolution of the drug product. Content uniformity is a high-volume repetitive test in which the sample preparation has been automated using the Zymate Tablet Processing Workstation by Barrett,[14] Lagade et al.[15] and Sadlowski et al.[16] More recently, Stanley and Park[17] compared the Zymark Tablet Processing Workstation II and the Waters Tablet Processing System robots using content uniformity determinations. Drug content and degradation product determinations have also been automated by Stanley.[18] Metered dose inhaler (MDI) product analysis is highly labor intensive. Robotic systems have been developed by Monk[19] to provide a repeatable method of handling and firing the MDI product into a collection apparatus, measuring weight loss and dosage content determination. The Anderson impactor is well established for measuring particle size in MDIs and a robot to automate this process is shown in Figure 6.

There has been an increased demand for dissolution testing for solid oral dosage forms. One of the first totally automated dissolution methods was developed by Kostek et al.[20] in which the robotic dissolution system was coupled with an ultraviolet/visible (UV/VIS) spectrophotometer equipped with an on-line flow cell for real-time collecting of samples and the immediate acquisition of data. With complex matrices of drug products, some automated methods needed to have samples collected off-line for analysis;[21] Rios et al.[22] used flow injection analysis and Kostek et al.[23] used real-time HPLC analysis. With the development of fiber optics Kostek et al.[24] have applied the technology to enable the UV/VIS analysis to be performed in situ. Doddapaneni et al.[25] proposed an alternative approach for dissolution testing of nonroutine formulation development samples by employing a segmented strategy using an automated dissolution system

Figure 6 Diagram of a Novi robotic system for automating the Anderson impactor. (Reproduced by permission of M.P. Smith.)

linked to several fraction collectors/dilutor dispensers with an off-line UV/VIS spectrophotometer. Stability testing of finished product forms a major part of the workload in a pharmaceutical development laboratory and this is likely to increase with globalization in the pharmaceutical industry.[26]

4 SYSTEM DEVELOPMENT AND INSTRUMENT QUALIFICATION

Robotic systems produced in the early 1990s were flexible and could be configured to perform a variety of sample preparation tasks. These instruments were developed in a time when validation requirements for computerized equipment were less stringent. As a consequence the effort and time to implement these systems was significantly less than that required now as we move into the next millennium. Manufacturers have responded by producing smaller, more dedicated automated systems that are easier to validate and use. An example is the evolution from the Zymate robot to Benchmate-type systems by Zymark Corporation.

Equipment vendors and users have worked together to share the task of equipment qualification and validation. Validation can be defined as establishing documented evidence which provides a high degree of assurance that a specific process will generate a product, in our case an automated method, that will meet predetermined specifications and quality attributes. Qualification is part of validation and has been defined by the US Pharmaceutical Manufacturers' Association Computer System Validation Committee for installation, operation and running of the system under workload for a specific application. In the pharmaceutical industry the ''4Qs'' model is generally used to qualify the unit and this is outlined in Figure 7.

The Pharmaceutical Analytical Sciences Group in the UK has developed guidelines[27] on the requirements for the qualification of analytical equipment. Design qualification is meant to ensure that the equipment meets the purchaser's specification and conforms to a recognized quality standard (e.g. ISO Standard 9001). The vendor should also allow the purchaser access to the software source code and validation documentation. Under design qualification a system of change control for system upgrades also needs to be in place. Protocols should also be provided by the manufacturer for the installation, operational and performance qualification stages. Installation qualification is designed to ensure the automated system that you receive is what you ordered. The environmental conditions for the location of the robot should also be checked to make sure they are appropriate. The equipment is assembled, generally by the vendor, and the system switched on. Individual units of the robot, for example balances, homogenizers, diluters and filtration devices, are checked against the manufacturer's specifications at the operational qualification stage. Calibration checks (e.g. for balances) should be made using traceable standards. Finally, the performance qualification involves checking that all the units that make up the robot function correctly as a complete system. A typical method is run at this point to facilitate this performance check.

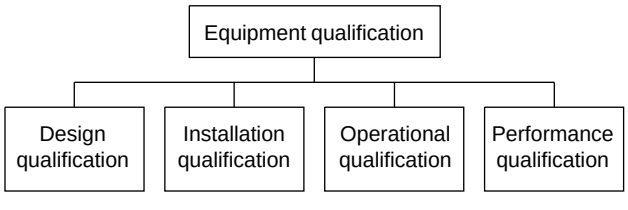

Figure 7 Diagram of ''4Qs'' model.

Software provided with the automated system should be developed using a life cycle management approach and according to a quality standard (e.g. ISO Standard 9003). The vendor should also provide certified evidence that the software has been validated (e.g. TickIT[28]).

Although a lot of effort is required to validate an automated system in the highly regulated environment of the pharmaceutical industry, for the most part, this makes good business sense and helps ensure complex automated equipment operates in the way in which it was intended.

5 METHOD DEVELOPMENT

In pharmaceutical analysis the assay, content uniformity and dissolution tests are the most common methods that use automation. Developing an automated assay or content uniformity method for a dosage form requires several aspects to be investigated before validation of the procedure can proceed.[29] The efficiency of the extraction step with respect to composition and pH of the extraction solvent together with, if used, the homogenization conditions of speed and time need to be studied. To optimize these factors efficiently, an experimental design (Figure 8) can be used to map the response surface of the extraction profile.

Suitable studies should also be conducted on the filter membrane to ensure significant quantities of drug are not retained and demonstrate that extractables from the filter do not interfere with the assay. The capacity of the filter to retain the filtrate without blocking is also important to enable the robotic system to work reliably. Automated sample preparation systems typically dispense solvents and perform dilutions gravimetrically.

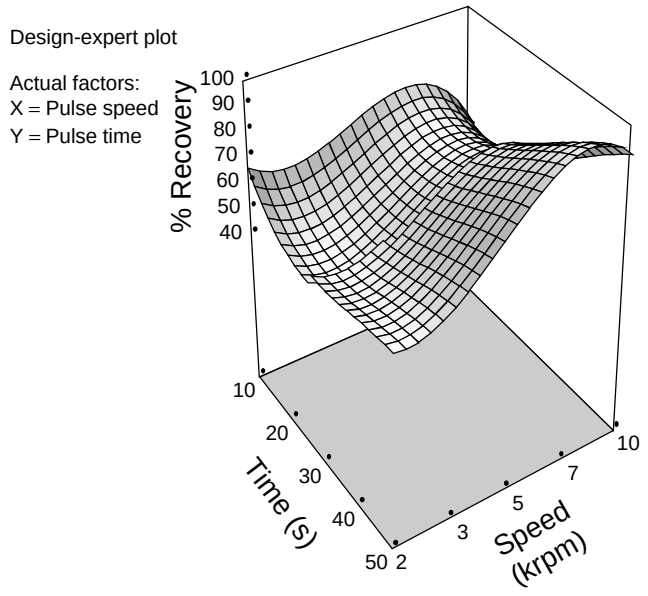

Figure 8 Experimental design used for extraction studies.

Consequently, the density of each sample solution that is manipulated gravimetrically is needed to ensure volumetric requirements of the manual method are met. A robotic method usually performs analysis serially using common components such as extraction vessels and tubing. Carry-over of drug from one sample to the next must, therefore, be investigated typically by employing a sample containing no drug (blank). If a homogenizer is used to grind and extract a dosage form, the volume of solvent retained by the homogenizer probe should be determined to see if this is significant. A robotic system comprises many components that come into contact with the drug and extraction solvent, consequently checks need to be carried out to ensure extractables are not leached from the equipment and significant amounts of drug are not absorbed by the system (e.g. tubing). Stability of solution must also be investigated to ensure that not only is there satisfactory chemical stability but also there is acceptable physical stability in terms of evaporation of the sample and standard solution.

6 METHOD VALIDATION

Method validation should only occur after the method has been developed and optimized. It provides an assurance of the reliability during normal use and it can be described as the means of providing documented evidence that the method performs in the manner in which it was intended. Validation of manual analytical methods in the pharmaceutical industry is well understood but documented in a number of guidelines in slightly different ways. It has been only recently that single documents have emerged that describe the terminology and methodology for method validation.[30,31] The parameters investigated, together with their definitions, are summarized below.

1. Accuracy – closeness of agreement between the true or reference value and the value found.
2. Precision – closeness of agreement between a series of measurements obtained from multiple samplings of the same homogeneous sample.
   (a) Repeatability – precision under the same operating conditions over a short interval of time.
   (b) Intermediate precision – precision within laboratories: different days, analysts and equipment.
   (c) Reproducibility – precision between laboratories
3. Linearity – ability (within a given range) to obtain tests results which are directly proportional to the concentration (amount) of analyte in the sample.
4. Range – interval between the upper and lower concentration (amounts) of analyte for which the method has suitable precision, accuracy and linearity.
5. Specificity – ability to assess unequivocally the analyte in the presence of components of the matrix.
6. Quantitation limit – lowest amount of analyte in a sample that can be quantitatively determined with suitable precision and accuracy.
7. Detection limit – lowest amount of analyte in a sample that can be detected but not necessarily quantitated as an exact value.
8. Robustness – capacity of a method to remain unaffected by small, but deliberate, changes in method parameters. Provides an indication of the method's reliability during normal usage.

It has been recognized that not all the parameters described above are tested for each type of analytical method. The type of method and its intended use dictate which tests should be investigated. The scheme recommended by the *United States Pharmacopeia*[32] for assay validation is shown in Table 1.

For robotic methods equivalency between the manual and automated procedure is an important extra validation test that is often performed since it is unusual not to have a manual procedure already in use. This test may not be relevant if accuracy has also been demonstrated for the automated method. In addition, the test for intermediate precision for an automated method may not be appropriate to investigate since a robot should not be affected by analyst-to-analyst or day-to-day variability.

The validation parameters above, combined with the data generated during method development, provide

Table 1 *United States Pharmacopeia* data elements required for assay validation

Validation parameter	Assay for major component	Assay minor component (e.g. impurity)		Assay[b]
		Quantitative	Limit test	
Accuracy	Yes	Yes	[a]	[a]
Precision	Yes	Yes	No	Yes
Specificity	Yes	Yes	Yes	[a]
Detection limit	No	No	Yes	[a]
Quantitation limit	No	Yes	No	[a]
Linearity	Yes	Yes	No	[a]
Range	Yes	Yes	[a]	[a]
Ruggedness[c]	Yes	Yes	Yes	Yes

[a] May be required, depending on the nature of the test.
[b] Performance characteristic (e.g. dissolution).
[c] *United States Pharmacopeia* definition – degree of reproducibility of results under a variety of conditions expressed as percentage relative standard deviation.

the necessary assurance that the method will perform satisfactorily when used by trained analysts.

7 TECHNOLOGY TRANSFER

Technology transfer from R&D to a commercial site is a key step in the pharmaceutical development process. The technology involved is not only the manufacturing process for the dosage form but also the associated analytical methods to control the quality of the product.[33] These methods include assay, content uniformity of individual dosage units, degradation profile for the drug and dissolution methods. In the pharmaceutical industry validated methods must be transferred. The transfer process for analytical methods can range from repeating some of the validation tests, for example, linearity, accuracy and precision, to just repeating the analysis of several lots of well-characterized batches of drug product. If an automated method is available then there is usually an associated manual method. In this case there are a number of options for technology transfer and these include transfer of both the manual and automated method or just the latter.

For successful implementation of automated methods, the robotic systems used in R&D should mirror those used at the commercial production site. The automated equipment must be rugged and easy to use for the production QC laboratory to invest in the technology. Complex robotic systems that need specialist, dedicated staff have not generally been successful in the QC laboratory. For manual methods, the transfer tests not only the methodology but also the analysts that use the method. However, when transferring an automated method between robots of the same type, less validation may be required, provided that the equipment has been qualified.[34] If there is confidence that the two robots are validated then just an electronic transfer of the method may be possible, which represents a considerable reduction of effort. Typically, repeat analysis of at least one batch of each strength of drug product is carried out on both automated systems. The data can be compared statistically; however, this can result in false conclusions particularly when the two data sets have high precision. This scenario can lead to statistical differences that are not analytically significant. The acceptance criteria for comparison of methods should, therefore, include an absolute difference (%) between the means of the two sets of results.[34] Successful technology transfer hinges on good communication between the analyst(s) that developed the method and the laboratory that receives the procedure.

8 BUSINESS BENEFITS

A key goal for pharmaceutical businesses during the 1990s has been reducing the time to develop a product and get it to market. This benefit also means that the time it takes to find the compounds that will not be commercialized is decreased. From a business viewpoint this is equally important in that the time and money can be re-allocated to developing more new chemical entities.

The increasing cost of drug development and loss of sales that can occur if the drug is delayed, by as little as a day, is very significant for a billion-dollar-a-year product. Automation is playing an important role in helping to reduce the time to market and can be justified in many ways.[35] Robotic systems usually are a high capital cost and this is often justified by increasing throughput, since automatic systems can operate 24 h a day and over weekends unattended, thereby also reducing the cost per analysis.[36] During the 1990s drug development times have significantly reduced, which has required more responsive ''just-in-time'' analysis and this has been an integral part of cost reduction for the release of raw materials and commercial products. Robots perform operations highly reproducibly without getting tired, which should improve the quality of the data. Automated systems also enable scientists to use their time more productively by spending time on evaluating data and on more innovative tasks, rather than performing routine repetitive operations. Robotic systems also have the capability of storing an audit record of manipulations performed by the system, so if an unexpected result occurs when the instrument is working unattended, it

should be possible to find out when and what happened. This provides a high level of regulatory compliance. When analyzing hazardous materials, automated systems reduce the exposure of the operator to these substances. Robotic systems can also effectively smooth out peaks and troughs in work loads that without automation could result in staff working longer hours, introducing shift work or contracting out the extra samples. These benefits more than offset the initial capital cost of the robot, and return on investment is typically less than 2 years for most applications.[35]

ACKNOWLEDGMENTS

I would like to thank Chris Banton, Les Brockhurst, John Gostick and John Stanley, who have worked with me over several years, for their enthusiasm and support in implementing automated methods.

ABBREVIATIONS AND ACRONYMS

HPLC	High-performance Liquid Chromatography
MDI	Metered Dose Inhaler
QC	Quality Control
R&D	Research and Development
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory

Steel and Related Materials **(Volume 10)**
Automation of Analytical Control in the Steel and Metals Industry

REFERENCES

1. R.D. Braun, 'Introduction to Instrumental Analysis', McGraw-Hill, New York, 1987.
2. M. Crook, 'Robotics: Past, Present and Future', *Anal. Proc.*, **30**, 165–167 (1993).
3. R.E. Majors, 'An Overview of Sample Preparation', *LC–GC*, **9**(1), 16–20 (1991).
4. G.L. Hawk, J.N. Little, F.H. Zenie, 'A Robotic Approach to Automated Sample Preparation', *Am. Lab.*, June, 96–104 (1982).
5. L. Snyder, J. Levine, R. Stoy, A. Conetta, 'Automated Chemical Analysis: An Update on Continuous-flow Approach', *Anal. Chem.*, **48**, 942A–956A (1976).
6. J. Ruzicka, E.H. Hanson, 'Flow Injection Analysis Part X. Theory, Techniques and Trends', *Anal. Chim. Acta*, **99**, 37–76 (1978).
7. D. Betteridge, 'Flow Injection Analysis', *Anal. Chem.*, **50**(9), 832A–846A (1978).
8. E. Bylund, R. Anderson, J. Carlson, in Proceedings from the Workshop on Flow Injection Analysis, No. 1, Swedish Chemical Society, 1977.
9. S. Smith, 'The Versatile TPWII – From Powders to Liquids', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 512–523, 1996.
10. T.D. Groeshner, 'The Implementation of the TPWII for On-line Analysis of Process Evaluation Samples', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 552–559, 1996.
11. S.S. Sekulic, H.W. Ward, D.R. Brannegan, E.D. Stanley, C.L. Evans, T.S. Sciavolino, P.A. Hailey, P.K. Aldridge, 'On-line Monitoring of Powder Blend Homogeneity by Near Infrared Spectroscopy', *Anal. Chem.*, **68**(3), 509–513 (1996).
12. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela, 'Near Infrared Spectroscopy in the Pharmaceutical Industry', *Analyst*, **123**, 135R–150R (1998).
13. S. Scypinski, T. Sadlowski, J. Baiano, 'Analysis of Cleaning Validation Samples with the Zymark Benchmate Workstation', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 368–378, 1996.
14. P.J. Barrett, 'Robotic Content Uniformity Testing of Effervescent Tablets', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 299–313, 1992.
15. L.G. Lagade, C.A. Cupo, T.K. Sawhney, D.E. Martynuk, K.J. Halloran, H.J. Mortko, 'The Use of Zymark Tablet Workstation for the Analysis of Pharmaceutical Tablets', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 314–331, 1992.
16. T. Sadlowski, P.F. George, S. Scypinski, 'A Modified Pytechnology Robotic System for Analysis of Tablet Dosage Uniformity', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 332–348, 1992.
17. J.R. Stanley, A.S. Park, 'A Comparison of the Zymark TPWII and Waters TPS Tablet Processing Robots', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 586, 1996.
18. J.R. Stanley, 'Development and Validation of an Automated Drug Content and Degradation Profile Analysis Method for Pharmaceutical Tablets', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 442–457, 1995.

19. A. Monk, 'The Robotic Automation of the Determination of Doses from Metered Dose Inhalers Under a Wide Range of Simulated Patient Conditions', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 610–622, 1995.
20. L.J. Kostek, B.A. Brown, J.E. Curley, 'Fully Automated Dissolution Testing', in *Advances in Laboratory Automation – Robotics*, eds. G.L. Hawk, J. Stimaitis, Zymark Corp., Hopkinton, MA, 701–720, 1985.
21. S.C. Lo, S.M. Donahue, C.W. Brown, 'Automated Drug Dissolution that Uses a UV-visible Diode Array Spectrophotometer', *J. Pharm. Sci.*, **82**(4), 350–354 (1993).
22. A. Rios, M.D. Luque Decastro, M. Valcarel, 'Flow Injection Analysis: New Approach to Pharmaceutical Determination', *J. Biomed. Anal.*, **3**, 105–121 (1985).
23. L.J. Kostek, B.A. Brown, J.E. Curley, *Advances in Laboratory Automation – Robotics*, eds. G.L. Hawk, J. Stimaitis, Zymark Corp., Hopkinton, MA, 311–328, Vol. 5, 1989.
24. L.J. Kostek, P.K. Aldridge, D.W. Melvin, B.A. Williams, K. Bratin, S. Sekulic, 'Automated Dissolution Testing Utilizing On-line Fiber Optic Probe UV Analysis', in Proceedings of the International Symposium on Laboratory Automation and Robotics, 355–367, 1995.
25. H. Doddapaneni, N. Palepu, J. Jushchyshyn, 'Dissolution Testing: Segmented Automation Techniques in a Development Environment', in Proceedings of the International Symposium on Laboratory Automation and Robotics, 524–532, 1996.
26. N. North, S. Smith, 'Laboratory Automation – Some Perspectives on the Challenges in the Implementation of the Technology in Pharmaceutical Development', in Proceedings of the International Symposium on Laboratory Automation and Robotics, 474–483, 1997.
27. M. Freeman, M. Leng, D. Morrison, R.P. Munden, 'Position Paper on the Qualification of Analytical Equipment', *Pharmaceut. Technol. Eur.*, November, 40–46 (1995).
28. 'The TickIT Guide to Software Quality Management System Construction and Certification using EN 29001', Issue 2, February, DISC TickIT Office, London, 1992.
29. J. Gostick, J. Stanley, A. Walsh, 'Development and Validation for Automated Methods for Finished Product', *Pharmaceut. Technol.*, March 2000, in press.
30. 'Validation of Analytical Methods: Definitions and Terminology', International Conference on Harmonisation, Topic Q2A, The European Agency for the Evaluation of Medicinal Products, London, 1995.
31. 'Validation of Analytical Methods: Methodology', International Conference on Harmonisation, Topic Q2B, The European Agency for the Evaluation of Medicinal Products, London, 1996.
32. *United States Pharmacopeia XXII*, 'Validation of Compendial Methods', United States Pharmacopeial Convention Inc., Rockville, 1710–1712, 1990.
33. J. Sadowitz, S. Li, 'Keys to Successful Method Validation and Method Transfer in the R&D Environment', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 1998.
34. T.A. Diehl, 'Automated Method Transfer: A Tale of Two Laboratories', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 1998.
35. F.H. Zenie, 'Superior Productivity – in the Laboratory and Beyond', *J. Automat. Chem.*, **15**(2), 23–33 (1993).
36. S. Smith, 'The Benchmate II TPW – From Installation to Financial Success', in Proceedings of the International Symposium on Laboratory Automation and Robotics, Zymark Corporation, Hopkinton, USA, 299–313, 1995.

Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

P. Campíns-Falcó, A. Sevillano-Cabeza, R. Herráez-Hernández, and C. Molins-Legua
Universitat de Valencia, Valencia, Spain

1 Introduction **1**
2 Basic Theory **2**
2.1 Sorbents and Interaction Mechanisms 2
2.2 Method Development 3
3 Off-line Versus On-line Procedures **4**
3.1 Off-line Automated Solid-phase Extraction 5
3.2 On-line Automated Solid-phase Extraction 5
4 Pharmaceutical Analysis **6**
4.1 Quality Control 6
4.2 Drug Monitoring 8
5 Comparison with Liquid–Liquid Extraction **10**
5.1 Instrumentation 10
5.2 Recovery and Precision 10
6 Other Applications of Solid-phase Extraction **11**
6.1 Concentration 11
6.2 Derivatization 14
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

Solid-phase extraction (SPE) using small cartridges filled with sorbents of a small particle size has rapidly established itself as an important sample clean-up technique. It has prospered at the expense of liquid–liquid extraction (LLE) which is considered labor intensive and frequently plagued by problems, such as emulsion formation and use of large volumes of hazardous solvents. A remarkable characteristic of SPE is its easy adaptation to on-line mode by column switching techniques; switching can be effected manually or by automated controllers.

The same analyte/sorbent interactions that are exploited in liquid chromatography (LC) are of use in SPE, but particle sizes employed are greater. Chemically bonded silica (SI), usually with C_{18} and C_8 organic groups, is by far the most commonly used material. At present, cross-linked polystyrene and other polymeric resins and mixed-mode SPE sorbents are finding increasing use. A guide to improve sample clean-up procedure can be established.

There are different approaches to automation for both off-line and on-line SPE, which involve different levels of apparatus cost. SPE is generally easier and cheaper than LLE. The precision obtained is similar for both techniques, while the recoveries obtained by SPE are generally better than those obtained by LLE.

Enrichment and derivatization can also be carried out by use of sorbents as supports.

1 INTRODUCTION

In order to understand the place of SPE or other sample clean-up procedures in the pharmaceutical field, the drug evolution cycle should be considered. This can be divided into different steps: discovery; chemical process development; metabolism, bioavailability and pharmacokinetics; formulation studies; stability studies; toxicology and safety testing; commercial production and marketing. This cycle gives different kinds of samples that may need a pretreatment prior to their measurement. These pharmaceutical samples are bulk-drug, preformulations, formulations and biological samples. The last samples are also connected with topics such as therapeutic drug monitoring and toxicological analysis developed in many cases in the pharmaceutical field.

Nonvolatile species can be removed from either solid or liquid matrices by a number of techniques which range from distillation to dialysis, LLE, and SPE, including both ion exchange and sorption. The traditional LLE method as a sample clean-up procedure for *pharmaceutical samples*, has been superseded by SPE. This approach with porous solid particles (Rohm and Haas Amberlite XAD™ resins) goes back to the early 1970s.[1] Practical use of this method was undoubtedly hindered by the necessity of grinding, sizing and purifying the early XAD™ resins. Now, with the ready availability of pure bonded-phase SI particles and, more recently, of pure polymeric resins of appropriate particle size, SPE is becoming the preferred method for isolation of organic components from predominantly aqueous samples.

Disposable cartridges filled with sorbents of a small particle size are employed, and the sample and the different mobile phases are flushed with a syringe as a low-pressure pump. The clean-up principles are similar to other open-column techniques. This method is quick and simple to operate, most sample handling losses are avoided and it is easily automated and coupled to LC.[2,3] LC has probably emerged as the method of choice in the pharmaceutical

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

field. Gas chromatography (GC) has gained more prominence for drug screening and coupled with mass spectrometry (MS) for confirmation analysis too.

In recent years, an increasing number of LC methods incorporating on-line sample clean-up by SPE on disposable precolumns by use of column switching have been developed. Switching devices permit the off-line multistep methods to be transformed into single-step procedures by the on-line purification of the sample. The on-line coupling of SPE and capillary GC still poses a number of problems.[4]

A topic close to pharmaceutical analysis is *residue analysis*, as the determination of antibiotic residues in food products of animal origin. SPE on disposable cartridges or precolumns has also been proposed as a substitute for or combined with LLE[5] in this field.

In systematic *toxicological analysis* the substances present are not known at the start of the analysis. In such an undirected search the extraction procedure must be a general procedure where a compromise must be reached in that the substances of interest are isolated at a yield as high as possible and the interfering substances from the biological material are removed.[6] For a long time the traditional sample work-up technique in analytical toxicology was LLE, often combined with sample pretreatment procedures such as conjugate hydrolysis, digestion and protein removal. Although LLE proved to be suitable in a substantial number of cases, the disadvantages of this technique, for example matrix interferences, emulsion formation, use of large volumes of hazardous solvents, have troubled the analyst. The SPE approach can partially overcome the drawbacks of LLE.

In recent years the development of suitable materials for SPE and an increasing demand for on-line sample treatment procedures have provided a new impetus to extraction approaches.

2 BASIC THEORY

This section briefly summarizes the most used sorbents and methods in SPE. The same analyte/sorbent interactions that are exploited in high-performance liquid chromatography (HPLC) are used in SPE. Ion exchangers have been also included due to their use in sample pretreatment processes and because the process involved in the separation is a physico-chemical process, as in the sorption.

2.1 Sorbents and Interaction Mechanisms

The first sorbent used for SPE was XAD™ resin as mentioned above. This resin is a low-polarity styrene–divinylbenzene copolymer which possesses the macroreticular characteristics essential for high sorptive capacity. In addition to polymeric resins, the solid-phase sorbents available include common inorganic sorbents (SI, alumina, Florisil™), siloxane-bonded SI materials (octadecyl, octyl, ethyl, cyclohexyl (CH), phenyl (PH), 3-cyanopropyl (CN), diol (2OH), 3-aminopropyl (NH_2), etc.), ion-exchange macroreticular resins, and specialty products such as resin-bonded phenylboronic acid (PBA) and activated carbon. Chemically bonded SI, usually with a C_{18} or C_8 organic group, is by far the most commonly used material for SPE. At present, cross-linked polystyrene and other porous polymeric resins[7] and mixed-mode SPE sorbents[8] are finding increasing use.

The packing material with particle sizes in the range of 30–60 μm is sandwiched between two porous frits with a pore diameter of about 20 μm in a column. The bottom end of most cartridges is terminated in a leur lok fitting. The cartridges are available in several sizes containing from 50 mg to 10 g of sorbent, with the 100–500 mg sorbent cartridges being the most used in the pharmaceutical field. Approximate values for the capacity and bed volume of the columns are 5% sorbent mass or 0.4–0.6 mequiv mL^{-1} for ion-exchange sorbents and 120 μL per 100 mg of sorbent, respectively.

A more recent approach to SPE is the use of a membrane (0.5 mm thick, 25 or 47 mm in diameter) loaded with an appropriate solid material immobilized within a stable, inert matrix of fibrillated poly(tetrafluoroethylene) (PTFE). These membranes are 90% by weight particle-loaded and 10% PTFE with very dense packing of the sorbent of a smaller average particle size (8 μm) than used in the conventional SPE cartridges. The capacity of the membranes is similar to that of the cartridges, but the small bed volume allows for reduced elution volumes, resulting in more concentrated eluates. Particle-loaded membranes are supplied in disk format for large-volume use and are contained in disposable cartridges for the small volumes appropriate for pharmaceutical analysis.

For the on-line mode, the majority of the applications use particle sizes in the 10–40 μm range. The internal diameters of both the sample clean-up column and analytical column should be the same, in order to minimize extracolumn band broadening. The vast majority of studies employ an internal diameter in the 2–4.6 mm range. Short precolumns (between 5 and 10 mm) are preferable because the duration of the flushing needed to remove undesirable matrix compounds is minimized. Packed precolumns are commercially available from several manufacturers, but manual slurry or dry packing of a precolumn does not present any real problems and, in addition to rapid exchange, allows the easy screening of new packing materials.

Table 1 Sorbents and interaction mechanisms in SPE

Interaction	Sorbent	Chromatographic mode	Analyte functional groups	Matrix
	SI-based bonded phase			
Hydrophobic	C_{18}	Reversed-phase	Aromatic rings	Aqueous
Nonpolar–nonpolar	C_8		Alkyl chains	
Van der Waals forces	C_2			
	CH			
	PH			
Hydrophilic				
Polar–polar	CN	Normal-phase	Hydroxyls	Nonpolar
Hydrogen bridges	2OH		Amines	
$\pi-\pi$ bond	NH_2		Heteroatoms (S, O, N)	
Dipole–dipole				
Induced dipole–dipole				
Ion-exchange	SXC	Ion-exchange	Amines	Aqueous
	PRS		Pyrimidines	
	CBA			
	SAX		CBAs	
	PSA		Sulfonic acids	
	NH_2		Phosphates	
	DEA			
Multiple				
Hydrophobic/ion-exchange	Mixed bonded SI	Reversed-phase and Ion-exchange	Acidic, neutral, and basic compounds	Aqueous
	No bonded phase			
Hydrophilic/hydrophobic	SI	Normal-phase	Generally used to provide class separation by dividing the sample into fractions containing a similar number and type of functional groups	Nonpolar
	Alumina			
	Florisil™			
	Diatomaceous earth			
	Activated carbon			
Hydrophobic/hydrophilic	Copolymer resins	Reversed-phase	Aromatic rings Alkyl chains	Aqueous
Covalent	PBA	Reversed-phase	Nucleotides, nucleosides Carbohydrates Catecholamines	Aqueous

C_{18}, octadecyl; C_8, octyl; C_2, ethyl; SXC, benzenesulfonic acid; PRS, propylsulfonic acid; CBA, carboxylic acid; SAX, quaternary amine; PSA, primary/secondary amine; DEA, diethylaminopropyl.

There are three general extraction mechanisms used in SPE: polar, nonpolar and ion-exchange. The main retention mechanisms for the most used sorbents are shown in Table 1.

Although the same type of SPE material can be obtained from different manufacturers, the results using SPE material from different manufacturers, and even results obtained from different batches from the same manufacturer, may show significant differences in behavior, i.e. in particle size distribution and flow velocities.

2.2 Method Development

It is recommended to find the sorbent that will provide the best results. For that, the potential interactive properties of the isolate(s) and matrix, taking note of functional groups, should be considered (Table 1). Also other known influences that may affect the extraction negatively are important, such as protein binding, properties of the isolate(s) like stability, solubility and pK_a.

If the analyte(s) has nonpolar (alkyl chains, aromatic rings), polar (hydroxyls, amines, keto groups), or ionic (amines, CBAs) functional groups then these properties can be selectively employed to facilitate retention. The undesired constituents of the matrix may interfere with retention of the analytes, requiring consideration of a different mechanism.

The same operational modes as in HPLC are used in SPE. Table 1 shows the option generally chosen for each sorbent.

The sample processing involves four distinct steps. Pretreatment of the sorbent with an activating solvent (such as methanol for chemically bonded SI and porous polystyrene resins) must be carried out to obtain better surface contact and to remove impurities from the sorbent. It is returned to the ready condition by washing

with three or four bed volumes of a solvent as similar to the sample solvent in polarity, ionic strength, and pH as possible. The sample is then sorbed onto the cartridge. The maximum volume that can be sampled depends on the breakthrough volume for the analyte(s). The cartridge is then rinsed with a weak solvent to remove undesirable matrix components. In the final step the analytes are eluted by a solvent of sufficient strength to desorb the analytes in a small volume without displacing more strongly sorbed matrix components.

The typical solvents used for nonpolar extraction are methanol, acetonitrile, ethyl acetate, chloroform, acidic methanol and hexane. Methanol, isopropanol, and acetone are used for polar extraction. Alkaline and acidic buffers and high ionic strength solutions are employed for ion-exchange extraction. Acidic buffer and acidic methanol are the solvents selected for covalent extraction.

The sample and solvents are flushed into the cartridge manually by use of a syringe, by centrifugation or in a vacuum manifold that permits the processing of between 10 and 24 columns at one time generally. SPE membranes require a filtration manifold or special apparatus for multiple extraction of several disks. For the on-line option, switching valves and pumping systems are necessary.

Different strategies exist to improve the sample clean-up by SPE, taking into consideration the sorbent used. Table 2 summarizes these strategies.

Table 2 Strategies for improving sample clean-up by SPE

Problem	Sorbent	Strategies
Inadequate analyte retention	Polar	Change to more nonpolar solvent
	Nonpolar	Change to more polar solvent
		Ionic strength ≤ 0.1
		Bad activation
		Ionic analyte
	Ion-exchanger	Dilution
		Change the pH
		Decrease the ionic strength
Inadequate elution	Polar	Study the secondary mechanism
	Nonpolar	Modify the ionic strength
		Increase the nonpolar solvent
	Ion-exchanger	Increase the elution volume
Poor sample clean-up	Polar, nonpolar, ion-exchanger	Increase the strength of the matrix washing solvent
Poor reproducibility	Polar, nonpolar, ion-exchanger	Protein binding
		Analyte absorption on the sorbent pH

3 OFF-LINE VERSUS ON-LINE PROCEDURES

Off-line SPE may be converted into an on-line procedure using column switching techniques. Its use in HPLC was first reported in 1973.[9] The term ''column-switching'' or ''coupled column chromatography'' includes, in the widest sense, all techniques by which the direction of the flow of the mobile phase is changed by valves, so the effluent from a primary column is passed to a secondary column for a defined period of time. Switching can be effected manually or by automated controllers.

Coupling LC/LC is a well-established technique for sample clean-up in the analysis of pharmaceutical samples. Its success depends on the type of sample, the precolumn, the eluent and the column-switching configuration.[2] The only significant difference with the off-line SPE is the switching network.

In the simplest configuration, only a switching valve is required in addition to a basic chromatograph. It should be indicated that during the sampling and clean-up steps, the eluent in the analytical column is stagnant. Therefore, this configuration may lead to considerable baseline fluctuations depending on the elution conditions, which limits its applicability. More powerful systems can be achieved at a reasonable cost by using an additional pumping system. In such a way, the analytical column is equilibrated during the clean-up stage, and only the fraction of eluent containing the compounds of interest is transferred to the analytical column (Figure 1).

In spite of the level of instrumentation, different transfer modes can be used. In forward flush (also called straight flush) configurations the flow direction in the precolumn is not changed during the transfer stage, which prevents solid particles retained at the head of the precolumn from being sent to the second column. However, for samples containing compounds which are strongly retained in the precolumn, backflush configurations (in which the flow direction of the mobile phase through the precolumn is reversed for the transfer onto the secondary column) are preferable, because the

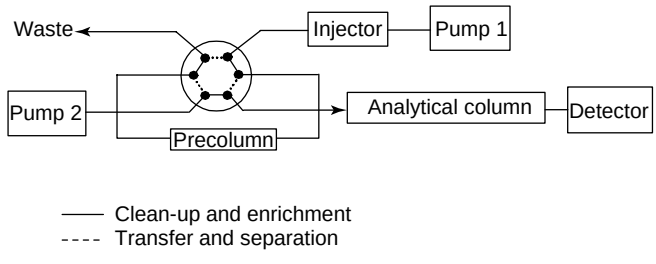

Figure 1 Typical scheme of a column switching system used for on-line SPE.

desorption of the most retained components after every injection is easier.

In order to improve sample clean-up, several columns can be coupled easily in the on-line mode by using several switching valves. In linear coupled-column systems, the peak capacities of the individual columns are summed. A review of commonly used switching networks is given by Ramsteiner.[10]

One of the most obvious trends in analytical laboratories is the growing need for automation. SPE has been traditionally performed in non- (or semi)automated mode. However, an important reason for growth is that SPE is amenable to automation, and today there are several examples of commercially available systems for automated off-line and on-line SPE.

Compared with manual SPE, automated SPE also improves the reproducibility because variations in experimental parameters are reduced or even eliminated. Another important advantage of automated SPE in pharmaceutical analysis is safety, because workers are protected from toxic substances and from samples obtained from patients with a range of diseases (HIV or hepatitis B, for instance).

There are different approaches to automation of SPE, which involve different levels of apparatus cost. These approaches can be categorized into off-line and on-line systems.[11]

3.1 Off-line Automated Solid-phase Extraction

In off-line methods the SPE cartridges are not directly connected to instrument, normally a liquid or a gas chromatograph. Unattended sample work-up and analysis can be performed as a result.

The simplest approach to automation is based on the employment of liquid-handling devices which include SPE in their functions. These workstations usually incorporate precision pumps (which deliver measured volumes of liquid to test tubes) with multisolvent capabilities, replaceable pipettes, and random access to vessels. The system also includes a SPE rack, with the cartridges or disks, and a liquid-handling probe. The probe runs in the trays (usually x–y–z), so it can reach any of the positions: samples, tubes or SPE cartridges, manipulating the sample and providing air or nitrogen (to move liquids through the cartridges) and the extraction or washing solvents. Specially designed probes or column caps create a tight seal in the cartridges. The syringe pump delivers conditioning solvents, which are then pushed through the cartridges by means of the pressurization line. Next, the sample is added and the pressurization step is repeated. If required, one or more washing solvents are also added, and pressure is again applied. Finally, the cartridge (which is over the waste container) is moved, so the sample is eluted into a clear container for further processing. Optional equipment such as a bar-code reader is also available in many commercial systems. All operations are controlled by a keypad controller or by a computer. In semiautomated systems, the extracts collected are manually injected onto the chromatograph. However, total automation is possible by connecting the workstation to the analytical column, for instance by means of a loop injector. Examples of this approach to automation are the Gilson ASPEC™ or the Hamilton Microlab SPE systems.

A more sophisticated form of SPE automation is that using flexible or semiflexible robotic manipulation. Semiflexible systems incorporate a robot-like device which performs only a few of the steps involved in sample preparation. Some modern chromatography autosamplers fit into this category. These devices extend the possibilities of unattended sample work-up because operations typically involved in sample conditioning (internal standard addition, dilution, digestion or derivatization) can be also performed in a fully automated way. The Prep-Station (Hewlett Packard) system is a typical example of this kind of instrument.

In flexible systems, the robot can be programmed to perform all operations involved in the sample conditioning procedure: weighting, dilution, and of course, those required in SPE. Nevertheless, robots are too expensive to be used in SPE operations. For this reason, in normal applications the robot is interfaced to a SPE work-station. The robot may move cartridges from a dispenser to racks holders, remove spent cartridges or transfer collection tubes, but rarely does the robot wait for liquids to be dispensed or for air to pass through the cartridges; instead, the SPE workstation performs these operations. Robots compatible with SPE have been commercialized by different companies such as Zymark Corporation or Hewlett Packard, among others.

3.2 On-line Automated Solid-phase Extraction

In on-line systems, the SPE device is inserted into the liquid or gas stream of the chromatograph, which means that it becomes part of the chromatograph. Automated sample preparation with on-line SPE is accomplished by using column-switching. When the precolumn is loaded with the sample (normally a biofluid), the analytes are selectively retained. The precolumn is then washed to remove endogenous interferents, so that they are sent to waste. The analytes are finally desorbed and transferred by means of an electronically controlled switching valve, which passes the mobile phase through the precolumn, and the drug is eluted out to the analytical LC or GC column. In such a way, off-line sample manipulation is reduced to a minimum and the entire analysis can be

easily automated. Hence, SPE/LC or SPE/GC can be considered particular forms of multidimensional LC/LC and LC/GC chromatography respectively.

The SPE/LC approach is particularly well suited for pharmaceutical analysis, as a vast majority of the samples of interest are compatible with liquid chromatographic systems (especially those operating under reversed-phase conditions). Besides the inherent advantages of automation, this methodology offers some advantages over off-line automated systems, for example, protection of light-sensitive or oxidation-sensitive analytes, because during sample processing they are kept away from light and air respectively. Moreover, implementation is relatively inexpensive compared with off-line SPE devices. In many applications, the cost is little more than a switching valve and a precolumn unit over the chromatograph needed for a manual method. However, an additional pump may sometimes be necessary for washing the precolumn. Although good stability is normally achieved for the repetitive injection of 10–100 μL of untreated samples, a clear disadvantage is that periodic regeneration or replacement of the precolumn may be needed to prevent clogging, especially when processing biofluids with a large amount of proteins (for example blood or plasma). Some commercial systems incorporate cartridge holders that can be loaded with several cartridges. The individual cartridges (which act as a precolumn) are connected to the analytical column by means of a switching arrangement. Each cartridge is periodically replaced by a fresh one, so clogging or memory effects are minimized. Alternatively, SPE in the individual cartridges is performed off-line, and then the cartridges are automatically inserted into the flow stream. Examples of this type of device are the AASP (Varian) and Prospect (Spark Holland) systems.

Whereas on-line SPE/GC is quite popular in environmental analysis, its use is not widespread in the pharmaceutical field. This can be explained by the successful performance of SPE/LC methods, and also by the inherent difficulties of coupling SPE to GC, with either solvent or thermal desorption. Moreover, in pharmaceutical analysis care must be taken to prevent water-containing eluents from being transferred to the GC part of the system. This problem may be overcome by drying the cartridge with an inner gas stream before the transfer step. Another possibility is to place a drying cartridge between the SPE precolumn and the GC system. Nevertheless, with modern equipment the introduction of up to 100 μL of solvent can be achieved.

A precolumn can also be used to couple a SPE workstation to a GC chromatograph, as illustrated in Figure 1. Larger sample volumes (up to several milliliters) can be processed in the SPE cartridges. The analytes are subsequently reconcentrated in the precolumn, which facilitates the complete transfer of the relevant fraction of solvent to the GC system. A proper selection of the precolumn packing would also improve the selectivity, for example by using a packing material with retention capabilities different to those of the SPE cartridges. If required, the precolumn can be dried before connection to the GC unit. As a result, excellent sensitivity and selectivity can be achieved. For example, the system described in Figure 2 enables the detection of some benzodiazepines at low parts per billion levels from only 1 mL of plasma.[12] Similarly, a precolumn can be used to couple a dialyzer module to a GC chromatograph (also to a LC chromatograph). The dialysis step adds selectivity as sample macromolecules do not reach the SPE unit.

A very interesting alternative for on-line automated systems with GC and HPLC is solid-phase microextraction (SPME), because difficulties derived from the injection of large volumes of solvents can be overcome. Automated SPME has been achieved by using polymer-coated fused-SI fibers. The fiber previously fixed to a microliter syringe is automatically inserted into the sample vial. The analytes diffuse and partition in the polymeric phase. Next, the fiber is removed from the vial, and the analytes are thermally desorbed, on-column, in the injector of the gas chromatograph. For HPLC an interface consisting of a six-port switching valve and a desorption camera are needed. This interface replaces the injection circuit of the liquid chromatograph.

4 PHARMACEUTICAL ANALYSIS

4.1 Quality Control

The control of pharmaceutical quality is one of the primary objectives of drug manufacturers. Quality control (QC) has been defined as actions such as testing, monitoring and inspecting taken to detect and control defects. The control of drug substances and drug product quality involves developing sets of tests and specifications for raw materials, intermediates of interest and final products, performing the tests on the subject material, compiling the results and submitting them to regulatory authorities on a regular basis. From the synthesis of a drug onwards, its identity, strength, purity and quality are continuously assessed and monitored.

The British and USA pharmacopoeias recommend microbial assay, ultraviolet (UV) spectrophotometric and HPLC methods for the determination of pharmaceutical preparations. HPLC is currently probably the most popular for impurity and dissolution testing and formulation determination. However, very few publications exist on QC applications of LC other than several examples that

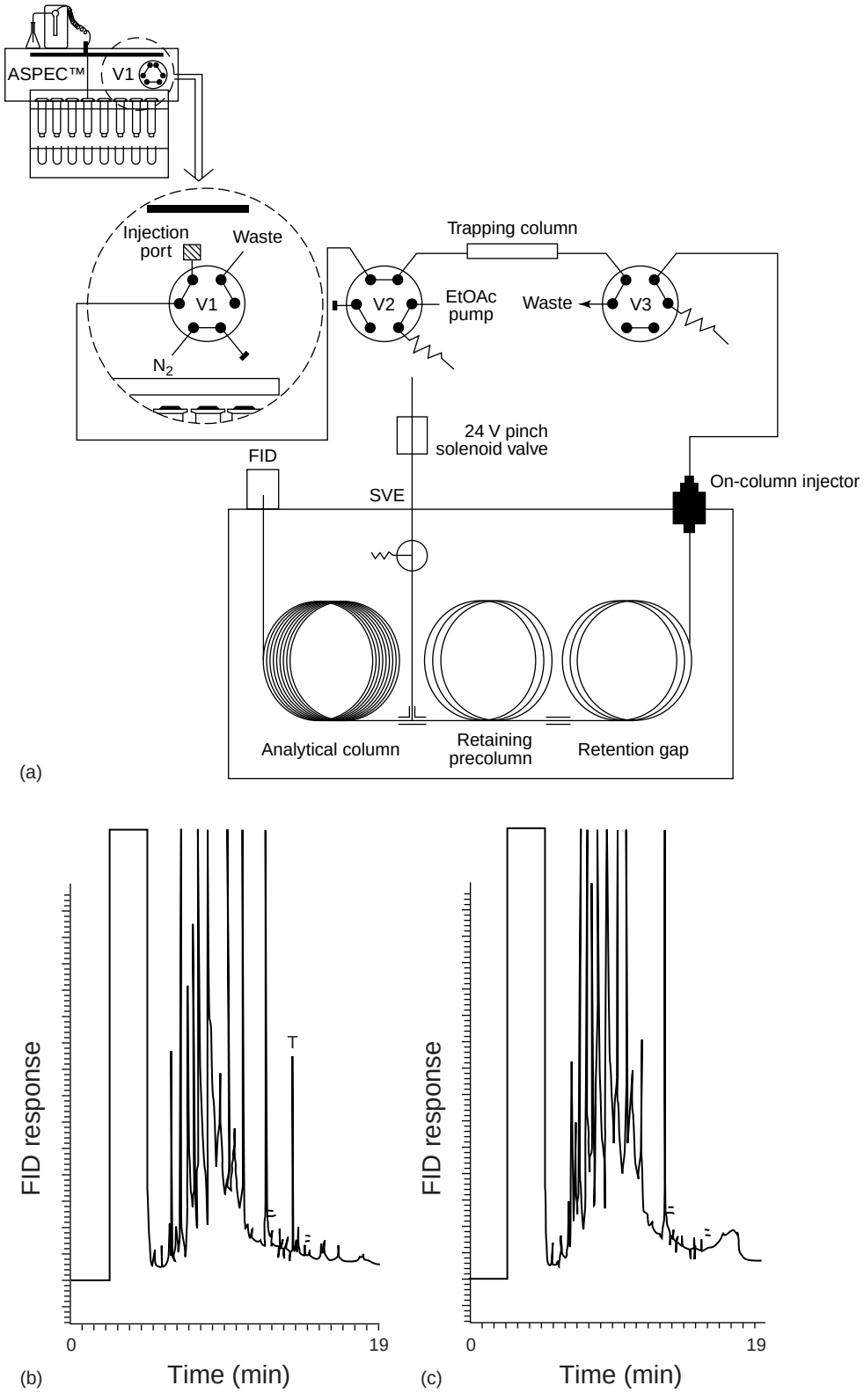

Figure 2 Set-up for on-line SPC/GC (a), and the chromatograms obtained from 1 mL of plasma spiked with 100 ng mL^{-1} of trazodone (T) (b) and blank plasma (c). (Reproduced by permission of Vieweg Publishing from A.J.H. Louter, R.A.C.A. van der Wagt, U.A.Th. Brinkman, *Chromatographia*, **40**, 500–506 (1995).) V1, V2 and V3 are switching valves; SVE, solvent vapor exit; FID, flame ionization detector.

list QC or quality assurance in their titles as part of a method development study.

In the development of new drugs the determination of the purity is important in order to establish the acceptability of batches for safety assessment and clinical trials. In general, process impurities include trace substances resulting from the drug substance synthesis, and trace side-reaction products or contaminants not removed by the isolation and purification scheme. Regulatory authorities set various levels for individual and total impurities in both drug substance and drug product that must be reported and investigated. Ideally, the total amount of impurity, as well as the quantity of each individual impurity, should be monitored in the bulk drug. Prabhu et al.[13] described the isolation and concentration of impurities in 3-[2-(2-benzoxazolyl)ethyl]-5-ethyl-6-methyl-2(1H)-pyridone (L-696229) using sequential preparative HPLC, SPE, and LLE. Three SPE cartridges connected in series were used.

An example of the use of SFE in studies related to dissolution testing can be found in Kenney et al.[14] The antiarrhythmic drug L-768673 in pharmaceutical preparations was retained on a C_{18} SPE column and was analysed by HPLC and UV detection at 245 nm.

We have proposed spectrophotometric procedures that integrate sample clean-up and derivatization in the same process by using C_{18} solid supports. The potential of this technique was evaluated determining the content of DL-amphetamine sulfate in centramine tablets using 1,2-naphthoquinone-4-sulfonic acid (NQS).[15] This methodology has been also used for the determination of cephalexin in capsules and oral suspension (both singly and in combination with bromhexine).[16]

While development of racemic drugs is still permitted, examples are known in which enantiomeric pairs both have the same or different biological activity or in which one member is active and the opposite isomer is completely inactive. Enantiomeric purity is controlled in drug manufacture, storage and shipment for both bulk drug and product for molecular entities with chiral centers. The specific chiral LC methods have involved three general techniques, namely precolumn derivatization to yield diastereomers, mobile phases with constituents to afford separation of enantiomer pairs, and chiral stationary phases. Lots of references can be found on these techniques and several reviews have been written.[3,17,18] We have used the first option for chiral analysis of amphetamine in pharmaceutical samples using a mixture of *o*-phthaldialdehyde and *N*-acetyl-L-cysteine as derivatizing reagent.[19] The assay uses a 20 mm × 2.1 mm inner diameter precolumn packed with a Hypersil™ C_{18} (30 μm) material for purification and derivatization. The resulting derivatives were subsequently separated in a C_{18} achiral column using a mobile phase of acetonitrile–methanol–acetate buffer and detected by fluorescence.

4.2 Drug Monitoring

The analysis of drugs in a biofluid matrix by HPLC or any other separation–detection scheme is usually difficult owing to the large number of substances present in the sample and because the compounds of interest are often present at very low concentration. The most problematic fluids are those that contain a large fraction of proteins: blood, plasma, and serum. Cerebrospinal and interstitial fluids, as well as urine, are generally more compatible with LC, owing to their low protein content. Solid samples can be processed after solubilization and homogenization.

Many methods have been published for the determination of drugs in biological matrices using SFE cartridges for sample clean-up. Table 3 shows representative examples using different sorbents for screening analysis and determination of several drugs. For each analysis Table 3 also gives the separation and detection techniques employed. Several kinds of drugs have been included, such as anticonvulsants, anti-asthmatics, antiarrhythmics, antibiotics, diuretics and illicits. Although the most used detector in HPLC is the UV/visible detector, examples of other detectors are shown in Table 3. Some reports with GC are also recommended. As an example of an analysis that needs a derivatization step, amphetamine and methamphetamine determination has been included in Table 3. SPE has been proposed for both reversed-phase HPLC and normal-phase HPLC. An overview of screening procedures using diatomaceous earth, polystyrene–divinylbenzene copolymers and mixed-mode bonded SI sorbents can be found in Franke and de Zeeuw.[6]

As mentioned earlier, the use of column-switching techniques has become an area of major interest. Its application to drugs in biological samples by HPLC methods incorporating on-line sample clean-up by SPE using column-switching has been reviewed.[2] The time required for the sample clean-up process using this technique is drastically reduced, as minimum off-line sample manipulation is involved. Therefore, this technique can be very useful where a large series of samples must be processed. As a result, the precision is generally improved. Some examples of that improvement in the analytical results are given in papers about on-line sample clean-up with an octadecylsilica-C_{18} SPE precolumn and using column-switching for screening of diuretics and probenecid[38,39] and the determinations of acetazolamide[40] and triamterene[41] in urine samples using an octadecylsilica-C_{18} SPE precolumn.

Antibiotics are used in food-producing animals not only for treatment of disease, but also to maintain health and

Table 3 Representative examples of the use of SPE cartridges for sample clean-up in drug monitoring

Analyte	Studied sorbent	Matrix	Separation and detection techniques	Reference
Screening of amiloride, acetazolamide, hydrochlorothiazide, triamterene, chlorthalidone, furosemide, cyclothiazide, bendroflumethiazide, bumetanide, ethacrynic acid, probenecid and spironolactone	C_{18}, C_8, C_2, CH, PH and cyano	Urine	HPLC with photodiode array	Campíns-Falcó et al.[20,21]
Screening of stimulants, narcotics, sedative hypnotics, antidepressants, antipsychotics, antihistamines, cardiovascular drugs, analgesics and some derivatives	Bond Elut Certify (mixed-mode)	Urine or serum	HPLC with photodiode array	Lai et al.[22]
Screening of cephalexin, cefotaxime, cefazolin, cefuroxime and cefoxitin	C_{18}, C_8, C_2, CH, PH, cyano cartridges and 3M Empore disk C_{18}	Urine	HPLC with diode array	Gallo-Martinez et al.[23]
Screening of benzodiazepines and imidazopyridines, antidepressants, antipsychotics, central analgesics, anticonvulsants, antimalarials, β-blocking and β-against agents, drugs to treat cardiac diseases, alkaloids, anticoagulants, analgesics and anti-inflammatory drugs, barbiturates and carbamates	C_{18}	Powdered hair	HPLC with photodiode array and GC/MS	Gaillard and Pepin[24]
Determination of teophiline, paraxanthine, caffeine and cefoxitin	C_{18}	Urine	HPLC with diode array	Campíns-Falcó et al.[25,26]
Determination of ethacrynic acid	C_{18}	Urine	HPLC with diode array	Campíns-Falcó et al.[27]
Determination of probenecid	C_{18}	Urine	HPLC with diode array	Campíns-Falcó et al.[28]
Determination of cefuroxime	3M Empore disk C_{18}	Urine	HPLC with diode array	Gallo-Martinez et al.[23]
Determination of cocaine, benzoylecgonine and cocaethylene	Copolymeric phase combining C_8 and strong cation-exchange	Urine	HPLC with diode array	Clauwaert et al.[29]
Amphetamine and methamphetamine with NQS as derivatizing reagent	C_{18}, C_8, C_2, CH, PH and cyano cartridges	Urine	HPLC with diode array	Campíns-Falcó et al.[30] Molins-Legua et al.[31]
Methadone	Hydrophobic and cation-exchange disk (mixed-mode)	Serum	HPLC with UV	Rudaz and Veuthey[32]
Determination of antimalarial drug proguanil and its metabolites, cycloguanil and 4-chlorophenyl biguanide	CN-endcapped	Plasma, whole blood and urine	HPLC with UV	Bergqvist et al.[33]
Phenytoin	C_{18}	Human breast milk, maternal plasma	HPLC with UV	Shimoyama et al.[34]
Oxycodone	C_{18}	Plasma	HPLC and electrochemical	Wright et al.[35]
Tricyclic antidepressants	Bond Elut TCA	Plasma	Capillary GC with nitrogen–phosphorus	de la Torre et al.[36]
Determination of ecstasy (methylenedioxymethamphetamine), methamphetamine and methylenedioxyethamphetamine	C_8	Urine	GC/MS	Szignan et al.[37]

promote growth. The use of unauthorized antibiotics or the failure to follow label directions for approved antibiotics could result in unsafe antibiotic residues in food products. Therefore, monitoring antibiotic residues in food forms part of a general policy to prevent unapproved uses of antibiotics. A review of the developments in LC methods of analysis for antibiotics (including aminoglycosides, chloramphenicol, sulfonamides, tetracyclines, macrolides and β-lactams) has been published.[5] This paper also covers clean-up procedures such as SPE.

5 COMPARISON WITH LIQUID–LIQUID EXTRACTION

LLE has been the most used technique for sample clean-up, but SPE is now becoming the preferred method for this step of the analysis. This section focuses mainly on a comparison of the two methods.

5.1 Instrumentation

The equipment involved in using a sorption column for off-line processing can be very simple for manual use as described in earlier sections. For automated off-line SPE more sophisticated apparatus is required. On-line assemblies require only a switching valve for the simplest configuration, and additional pumps and valves for specific applications. In summary, SPE can work with or without expensive apparatus and the partial or total automation of the on-line mode is relatively easy.

In the simplest form of LLE a separatory funnel is needed. When the distribution constant of the solutes are very small, continuous extractors or countercurrent distribution apparatus is required. A continuous liquid–liquid extractor can be coupled to a gas or liquid chromatograph in three different ways: indirectly, by collecting the flow of the phase of interest in vials after the sample has been aspirated or injected into a flow injection manifold and subsequently injecting their contents manually into the chromatograph; on-line, via a simple turntable acting as an interface between the two continuous separation systems; and on-line, but directly, with no interface or discontinuity between the extractor and the chromatograph. Simplicity and the degree of automation of these three alternatives increase in the same order as they have been mentioned. In all cases SFE is generally easier and cheaper than LLE.

5.2 Recovery and Precision

Disadvantages of LLE are matrix interferences, emulsion formation, and the use of large volumes of hazardous solvents. SPE can overcome these drawbacks. The methods used for breaking an emulsion in LLE are generally time-consuming. The most common method is the addition of salt (e.g. NaCl) either as a solid or a saturated aqueous solution. Settling of the lower phase can be accelerated in a separatory funnel by the application of heat, as with an infrared lamp. More difficult emulsions may require high-speed centrifugation to separate two phases. Another technique is the use of low-melting-point organic solvents, such as ether, for the extraction of aqueous samples. Emulsions in this case are first separated by centrifugation, then the mixture is frozen in dry ice or liquid nitrogen. Finally, the organic phase is simply poured off leaving the frozen aqueous phase behind. Another option is to use filter papers.

Where extraction gives low recovery of the desired simple components despite optimum adjustment of solvent polarity and pH, recovery can be increased by using large volumes of solvent for the same sample quantity, and by using successive or continuous extraction. Following LLE of the sample, the solvent phase is normally filtered and may be further dried over anhydrous sodium sulfate, concentrated by evaporation and subjected to further clean-up procedures. The rate and extent of extraction may be different for a solute in a test system and in the presence of a sample matrix. Partial association of drug substances with proteins in plasma samples has been recognized.

Diuretics can provide a good example for testing the recoveries obtained for acidic, basic and neutral drugs in urine samples processed by LLE and off-line and on-line SI-based bonded-phase SPE. Table 4 shows results obtained in our research group.[20,21,38,39] As can be seen in Table 4, the acidic diuretics present better recoveries by LLE in acidic conditions, while basic diuretics increase their recoveries in basic conditions. Neutral diuretics present similar values for both acidic and basic extractions. The urine blanks obtained by SPE in different SI-based bonded-phase cartridges are similar to those obtained by LLE in acidic conditions (Figure 3). The standard deviations obtained are similar to those obtained by LLE, but the recoveries achieved are better, although only a unique cartridge has been used for sample clean-up. C_{18} sorbent provides higher recoveries for all the analytes. The washing solvent used for clean-up of the matrix was water. The results obtained with 2 mL and 0.5 mL of water are shown in Table 4. The smaller amount of water (0.5 mL) should be only used if acetazolamine is the compound being analyzed.

The efficiency of the on-line method is similar to that found with the C_{18} cartridges (Figure 3 and Table 4). The recoveries obtained for standards and spiked urine are the same, indicating that the sample matrix does not affect the recoveries of diuretics and probenecid. The precision is excellent with coefficients of variation below 6% in spiked urine samples. Also, it is greatly improved

Table 4 Comparative study of LLE and off-line and on-line SPE for diuretics

Diuretic	pK_a	% Recovery													
		LLE		SPE											
				Off-line										On-line	
				In urine (2.0 mL washing step)						Urine (0.5 mL washing step)					
		Acidic	Basic	C_{18}	C_8	C_2	CH	PH	CN	C_{18}	C_8	C_2	CH	PH	In water	In urine
Amiloride	8.7[a]	0	24	103	104	59	85	83	5	110	72	54	109	89	88.2	93
				7	10	8	9	9		3	3	13		5	0.6	4
Acetazolamide	7.4, 9.1[a]	80	0	<16	0	0	0	0	0	63	71	26	61	68	47	46
										6	4	6	2		5	3
Hydrochloro-thiazide	7.0, 9.2[a]	71	63	27	29	8	19	21	0	45.5	57	27	53.5	49	63	61
	9.5, 11.3[b]	9	8	5	6	3	6	3		0.7	2	7	0.7		2	2
Triamterene	6.2[a]	26	74	98	94	94	68	66	20	120	139	104	128	96	98	101
		1	10			17				3		20			2	4
Chlorthalidone	9.35[b]	85	75	114	115	126	102	105	<1	146	138	138	145	116	95	92
		4	8	8	2	4	15	3		6	1		5	5	2	3
Furosemide	3.8, 7.5[b]	41.7	25	85	103	76	81	84	3	92	92	70	87	78	102	99
		0.5	4	12	10	6	14	5		2	6		4	8	2	3
Cyclothiazide	9.1, 10.5[b]	53	62	102	100	108	80	83	10	81	80	88	76	72	99	–
		2	6	12	7	6	10	8		3	4		2		3	
Bendroflu-methiazide	8.5,[a] 9.0[b]	56	60	98	98	109	86	89	19	102	92	103	89	79	99	–
		8	9	5	12	3	7	2		10	5		2	8	2	
Bumetanide	5.2, 10[a]	74	39	93	96	91	100	92	–	101	79	64	74	59	–	–
	3.6, 7.7[b]	9	3													
Ethacrynic acid	3.5[a]	72	49	83	102	114	74	84	–	89	100	113	96	93	–	–
		3														
Probenecid	3.4[b]	73	45	87	95	134	75	103	0	127	119	99	120	97	–	–
				5	12	7	6	4			9		6			
Spironolactone	–	69	62	92	94	106	81	88	52	89	83	98	81	71	–	–
		6	5	8	8	12	9	4		6	3			9		

[a] From Hansch.[42]
[b] From Ventura and Segura.[43]

in comparison to LLE or SPE on disposable cartridges, as no on-line operation is involved.

Styrene-divinylbenzene (SDB) resins are especially interesting for analysing urine samples, since sulfate and glucoronide conjugates can be isolated. Extraction yields are in the order of 80%, with a precision of 7–27%, which are on average rather high.[6] These types of compounds are not amenable for classic solvent extraction. The SDB resin is also used for isolating substances from other biological matrices such as blood, serum, bile, gastric content, and tissues.

The extraction yields of substances isolated from different biological samples may vary considerably, however. The resin has to be cleaned very carefully. Recently, new SDB-based SPE columns (e.g. Bond Elut ENV, Varian) have become available, which overcome the above drawbacks. Moreover, SDB material in extraction disks can minimize volumes of elution solvents while still allowing relatively large sample volumes; this is a main characteristic of the SDB sorbents. Screening methods for acidic, basic and neutral compounds using this type of material are not yet available.

6 OTHER APPLICATIONS OF SOLID-PHASE EXTRACTION

6.1 Concentration

In pharmaceutical testing (stability or dissolution tests, for example) high concentrations of analytes are usually involved. Consequently, dilution of the samples is often necessary prior to the actual measurement. In contrast, many other applications usually require some form of analyte enrichment: identification and/or quantification of drug impurities in pharmaceutical preparations (by-products from the synthesis or degradates); pharmacokinetic studies; therapeutic drug monitoring; and in those situations where the sample volume is restricted, particularly if the sample has to be divided between several screening tests. It should be noted that, in a majority of such cases, the analytes also require purification owing to the large number of substances present in the samples.

Analyte enrichment with SPE column or disk cartridges is based on collecting the analytes in a volume of solvent

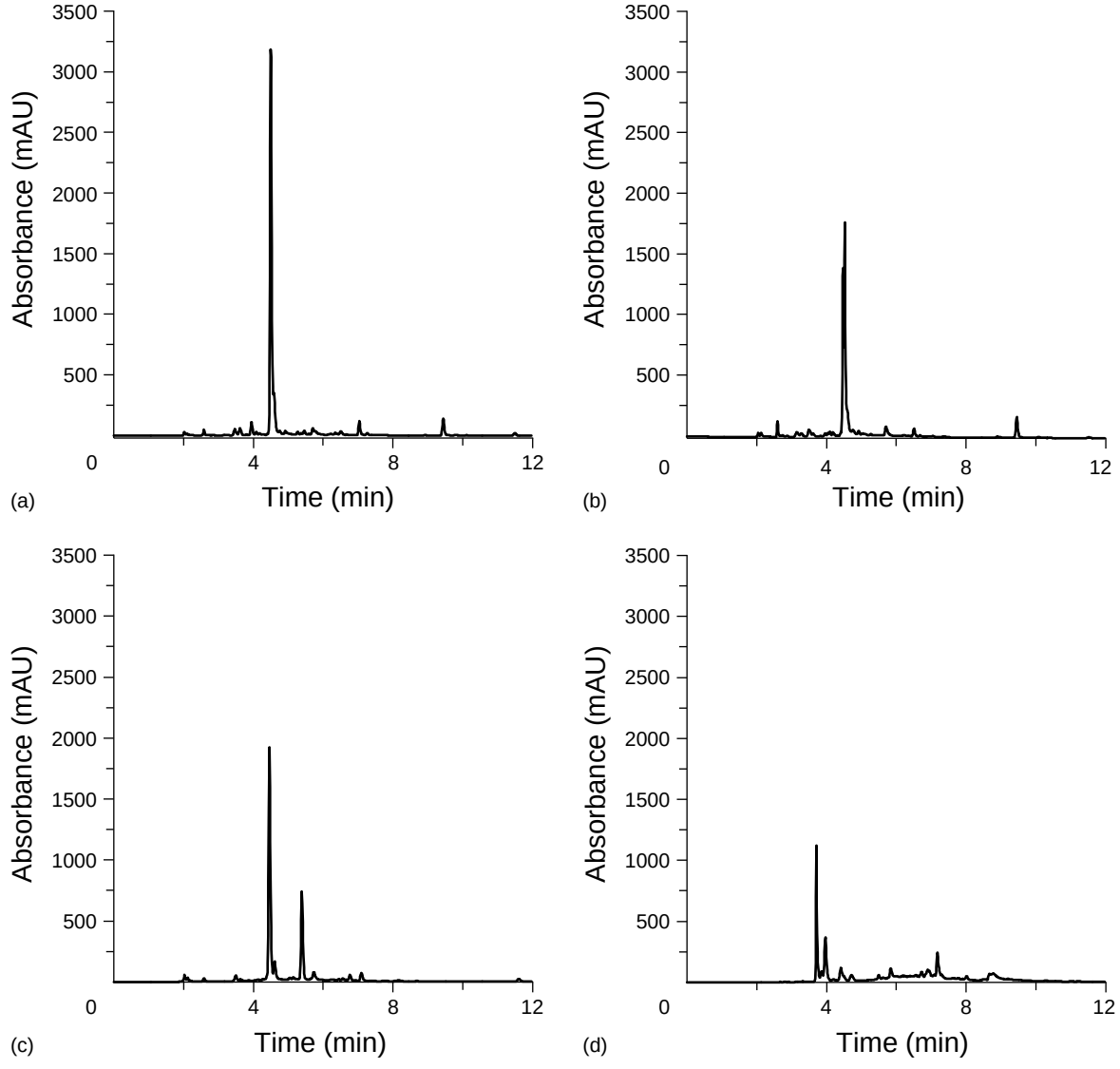

Figure 3 Typical chromatograms obtained at 230 nm from a urine sample by (a) acidic LLE, (b) basic LLE, (c) off-line SPE into a C_{18} column cartridge and (d) on-line SPE with column switching using a precolumn packed with a C_{18} packing.

lower than that of the sample. Unfortunately, in this process, undesirable matrix components may also be concentrated, especially if low selective packings (such as common reversed-phase materials) are used for retention of the compounds of interest. For this reason, washing of the cartridges with an appropriate solvent or mixture of solvents is often needed to improve the selectivity.

In principle, the enrichment factor is only limited by the losses of the analytes from the cartridges due to breakthrough during both the sampling and the washing steps. Therefore, the success of the enrichment procedure depends on a variety of factors: sample volume and type; characteristics of the cartridge; nature and volume of the washing eluent; and volume of solvent used to collect the analytes.

In order to achieve maximum enrichment, a careful selection of the SPE cartridges should be made according to the basic principles indicated in previous sections. The main parameters to be considered are type of packing material, particle size and loading capacity. Cartridges packed with 100 mg of packing material are generally suitable for most applications involving the analysis of up to 1 mL of the most common biofluids. In many instances, apolar packings (C_{18} or C_8) provide nearly quantitative recoveries of the analytes and the required selectivity, as salts and other very polar matrix constituents are wasted out with a relatively small volume of washing solvent (normally water or a buffer). However, more selective packings (protein-coated phases or immobilized antibodies) may be required for some applications.

For a given SPE cartridge, the volume of washing solvent should be as low as possible in order to prevent breakthrough. Therefore, the use of large volumes of samples may not compensate the losses of interesting compounds during the washing stage. With respect to the sample type, the most problematic fluids are those containing a high content of particulate material. In those cases, the employment of large sample volumes may be limited by the low flow rates attainable not only during the sample loading but also during the washing and desorption steps. This effect can be minimized if the samples are previously filtered or centrifuged to prevent blocking of the cartridge frits.

However, in order to achieve the maximum enrichment factor, the volume of solvent used to collect the analytes should be as low as possible. This volume is mainly determined by the amount of packing material in the cartridge, provided that a strongly eluotropic solvent is used for desorption. Cartridges containing 100–200 mg of packing permit the collection of the analyte in relatively small volumes, typically a few hundred microliters. Lower volumes lead to unsatisfactory reproducibility because, after the desorption step, small (and variable) fractions of the eluent may remain in the cartridges. As in LLE, the enrichment factor can be improved by evaporating the collected extracts to dryness, followed by redissolution. Evaporation, however, is time-consuming and may be inadequate if volatile compounds are going to be analyzed. It should be noted that, if chromatography is the final step of the analysis, the sensitivity attainable with the SPE column or with disk cartridges is limited by the fact that only a fraction of the collected extracts is finally injected into the chromatograph.

On-line SPE is an attractive alternative for enrichment of the analytes in chromatography (especially in LC).[2] In this approach enrichment is based on retention of the analytes in a precolumn when a large volume of sample is flushed through it; the precolumn is then flushed with an appropriate solvent, and the enriched analytes are finally transferred to the analytical column.

The main problems associated with the direct injection of biofluids in chromatographic systems are precolumn clogging and irreversible adsorption of matrix constituents to the stationary phases. For these reasons, the volume of samples that can be processed with on-line SPE are significantly lower than those that can be handled with SPE cartridges. Otherwise, rapid pressure development occurs, and the performance of both the precolumn and the analytical column may seriously deteriorate after a few injections. Although special packings (such as restricted access materials) have been designed for the on-line SPE in the biomedical field,[42] sample volumes of 50–100 μL can be adequately processed with conventional packings. Nevertheless, the sensitivity attainable with on-line SPE is similar to that of the off-line methods, as in the latter approach only a fraction of analyte is introduced in the chromatographic system. For this reason, on-line SPE is the best alternative for trace enrichment when the sample volume is restricted.

Moreover, careful selection of working conditions allows the injection of large sample volumes with suitable stability for most applications. For example, a straight-flush configuration (in which the flow direction of the mobile phase through the precolumn is not changed for the transfer onto the analytical column during the transfer step) protects the analytical column from solid particles or matrix components retained at the head of the precolumn, thus making possible the injection of relatively large volumes of biofluids. In addition, the precolumn lifetime can be extended by including a regeneration step in the final procedure. For regeneration, the precolumn is disconnected from the analytical column, and then it is flushed with a strongly eluotropic solvent. Since regeneration can be carried out during the chromatographic separation, the total analysis time is not increased. In many instances, occasional cleaning (e.g. when daily work is finished) allows the analysis of several samples with suitable stability and performance of the system for routine work. This is illustrated in Figure 4, which shows the effect of the successive injection of 250 μL of untreated plasma on back-pressure. Although back-pressure in the precolumn (packed with a 40-μm

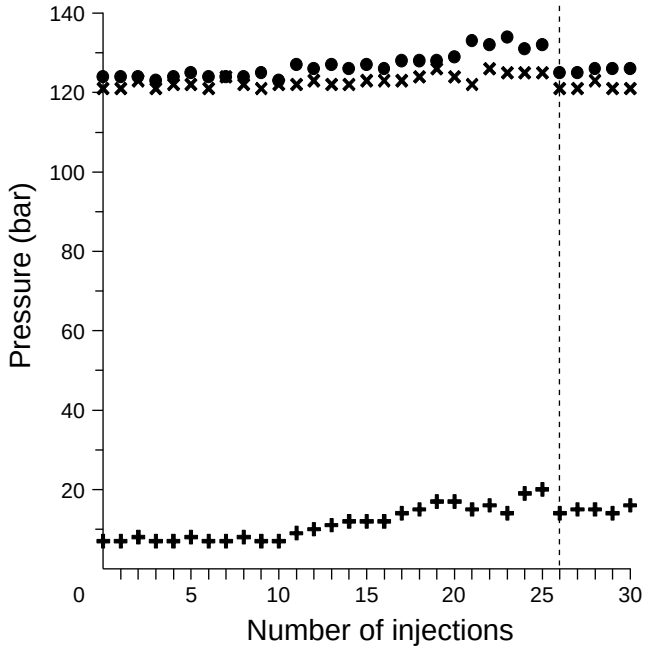

Figure 4 Effect of the number of injections on the back-pressure of the precolumn (+), the analytical column (×), and on the two columns linked (●). The dashed line indicates cleaning both columns with acetonitrile (approximately 15 mL).

C_{18} stationary phase) is approximately duplicated every 25 injections, the pressure at the top of the analytical column remains constant. Cleaning of the precolumn with an organic solvent (e.g. acetonitrile) every 25 injections reduces pressure and permits the system to be used for the analysis of many other samples. Under these conditions, quantification at sub parts per million levels is possible for a large number of compounds in plasma or urine.

Enrichment via a precolumn can also be used to overcome dilution, if dialysis is used to remove macromolecular constituents in biofluids (see section 3.2).

6.2 Derivatization

In pharmaceutical analysis, chemical derivatization is often needed to enhance the sensitivity and/or the selectivity of the overall analytical process, particularly when using LC. In this respect, most of the described procedures are based on precolumn solution derivatization because these reactions do not suffer from kinetic limitations and can be conducted under very flexible conditions (pH, temperature). In the analysis of biological samples, a preliminary step is analyte purification because the reagent can be reactive toward matrix constituents. In addition, derivatization generally involves extra sample handling aimed at removing unreacted reagent or to preconcentrate the derivatives. As a result, the final procedure may be very tedious and time-consuming.

Precolumn derivatization may be simplified in some instances with SPE materials. Krull et al.[43] proposed a very interesting approach based on using solid-phase reagents for simultaneous clean-up and derivatization. The solid-phase reagent contains ionic or covalent labile tags that possess specific detector enhancement properties. The choice of the pore size of the solid support (or the percentage of cross-linking when using a polymer network) permits size exclusion, thus enabling retention of small molecules but hindering the large ones. After purification, the adsorbed analytes are made to react with the reactive part of the solid-phase material. Finally, the derivatives formed are desorbed for further analysis. This methodology has been successfully used for the derivatization of several amines with a variety of reagents in both off-line and on-line modes.

Another approach to derivatization with SPE supports is based on the successive addition of the SPE material (normally a resin) and the derivatization reagent to the samples. The derivatives are retained on the SPE material, and isolated from the matrix and from the excess of reagent by filtration. Next, the derivatives are desorbed from the resin, which is then separated by filtration. This methodology has been applied, for example, to simplify multiple derivatization required for the analysis of prostaglandins in plasma by GC.[46]

Alternatively, derivatization can be carried out in conventional SPE cartridges or precolumns (in on-line SPE) as has been developed in our group. In the former approach, the SPE cartridge is used to selectively retain and/or preconcentrate the analytes, and then to retain the derivatives formed when an aliquot of the reagent is flushed though it. The analyte and the reagent are made to react for a given period of time. Then, the excess of reagent can be removed (if

Table 5 Methods for determination of drugs in urine, which integrate clean-up and derivatization in the same process

Analyte	Studied sorbent	Derivatizing reagent	Separation and detection techniques	Reference
Amphetamine and methamphetamine	C_{18} cartridge and precolumn	NQS	HPLC with diode array	Campíns-Falcó et al.[45] Herráez-Hernández et al.[46]
Norephedrine, ephedrine, pseudoephedrine, amphetamine, 3-phenylpropylamine and methamphetamine	C_{18}, C_8, C_2, CH, PH, CN cartridges and 3M Empore C_{18} disk	3,5-Dinitrobenzoyl chloride	HPLC with diode array	Herráez-Hernández et al.[47]
Norephedrine, ephedrine, pseudoephedrine amphetamine, methamphetamine and 3-phenylpropylamine	C_{18} cartridge and precolumn	9-Fluorenylmethyl chloroformate	HPLC with fluorescence	Herráez-Hernández et al.[48]
Amphetamine and methamphetamine	C_{18} cartridge and precolumn	Dansyl chloride	HPLC with fluorescence and chemiluminescence	Molins-Legua et al.[49]
Putrescine, cadaverine, spermidine and spermine	C_{18} cartridge	Dansyl chloride	HPLC with fluorescence	Molins-Legua et al.[50]

required) by flushing the cartridge with an appropriate solvent. Finally, the derivatives are collected for further processing. The on-line methodology is based on similar principles, but the time of reaction is controlled by stopping the flow rate in the precolumn after injecting the reagent. It is important to note that, under this approach, sample clean-up, enrichment and derivatization can be performed without manual intervention. In addition, minimum instrumentation is needed over conventional equipments: an isocratic pump, a precolumn and a switching valve. This methodology has been applied to the LC determination of different types of compounds such as amino acids, amphetamines, and polyamines in plasma and urine, in combination with reagents typically used to improve UV or fluorescence detection, as can be seen in Table 5.

ABBREVIATIONS AND ACRONYMS

CBA	Carboxylic Acid
CH	Cyclohexyl
CN	Cyanopropyl
DEA	Diethylaminopropyl
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
MS	Mass Spectrometry
NH_2	Aminopropyl
NQS	1,2-Naphthoquinone-4-sulfonic Acid
PBA	Phenylboronic acid
PH	Phenyl
PRS	Propylsulfonic Acid
PSA	Primary/Secondary Amine
PTFE	Poly(tetrafluoroethylene)
QC	Quality Control
SAX	Quaternary Amine
SDB	Styrene-divinylbenzene
SI	Silica
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
SXC	Benzenesulfonic Acid
UV	Ultraviolet
2OH	Diol

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Automation in the Clinical Laboratory • Capillary Electrophoresis in Clinical Chemistry • Drugs of Abuse, Analysis of • Laboratory Instruments in Clinical Chemistry, Principles of • Product Development for the Clinical Laboratory • Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis

Food **(Volume 5)**
Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Robotics and Laboratory Automation in Pharmaceuticals Analysis

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography • Gradient Elution Chromatography • Reversed Phase Liquid Chromatography

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. G.A. Junk, J.J. Richard, M.D. Grieser, D. Witiak, J.L. Witiak, M.D. Arguello, R. Vick, H.J. Svec, J.S. Fritz, G.V. Calder, 'Use of Macroreticular Resins in the Analysis of Water for Trace Organic Contaminants', *J. Chromatogr.*, **99**, 745–762 (1974).
2. P. Campíns-Falcó, R. Herráez Hernández, A. Sevillano Cabeza, 'Column-switching Techniques for High-performance Liquid Chromatography of Drugs in Biological Samples', *J. Chromatogr.*, **619**, 177–190 (1993).
3. K. Fried, I.W. Wainer, 'Column-switching Techniques in the Biomedical Analysis of Stereoisomeric Drugs: Why, How and When', *J. Chromatogr. B*, **689**, 91–104 (1997).
4. K. Grob, 'Development of the Transfer Techniques for On-line High-performance Liquid Chromatography – Capillary Gas Chromatography', *J. Chromatogr.*, **703**, 265–276 (1995).

5. B. Shaikh, W.A. Moats, 'Liquid Chromatographic Analysis of Antibacterial Drug Residues in Food Products of Animal Origin', *J. Chromatogr.*, **643**, 369–378 (1993).
6. J.P. Franke, R.A. de Zeeuw, 'Solid-phase Extraction Procedures in Systematic Toxicological Analysis', *J. Chromatogr. A*, **713**, 51–59 (1998).
7. J.S. Fritz, P.J. Dumont, L.W. Schmidt, 'Methods and Materials for Solid-phase Extraction', *J. Chromatogr. A*, **691**, 133–140 (1995).
8. M.W. Powell, 'Development of a Binary Solid-phase Extraction Cartridge for Use in Screening Water Samples for Organic Pollutants', *J. Chromatogr. A*, **697**, 101–105 (1995).
9. J.F.K. Huber, R. Van der Linden, E. Ecker, M. Oreans, 'Column Switching High-pressure Liquid Chromatography', *J. Chromatogr.*, **83**, 267–277 (1973).
10. K.A. Ramsteiner, 'Systematic Approach to Column Switching', *J. Chromatogr.*, **456**, 3–20 (1988).
11. R.E. Majors, 'Automation of Solid-phase Extraction', *LC/GC*, **6**, 346–350 (1993).
12. A.J.H. Louter, R.A.C.A. van der Wagt, U.A. Th. Brinkman, 'On-line Automated Sample Preparation – Capillary Gas Chromatography for the Analysis of Plasma Samples', *Chromatographia*, **40**, 500–506 (1995).
13. S.V. Prabhu, J.M. Ballard, R.A. Reamer, D.K. Ellison, 'Isolation and Identification of Impurities in L-696,229 Drug Substance', *Talanta*, **40**, 989–994 (1993).
14. R.R. Kenney, R.J. Forsyth, H. Jahansouz, 'Solid-phase Extraction and Liquid Chromatographic Quantitation of Anti-arrhythmic Drug L-768673 in a Microemulsion Formulation', *J. Pharm. Biomed. Anal.*, **17**(4–5), 679–687 (1998).
15. P. Campíns-Falcó, C. Molíns-Legua, A. Sevillano-Cabeza, R. Porras-Martinez, 'Derivatization of Amphetamines and Methamphetamine with 1,2-Naphthoquinone 4-Sulfonic Acid into Solid-phase Extraction Cartridges. Determination of Amphetamine in Pharmaceutical and Urine Samples', *Analyst*, **122**, 673–677 (1997).
16. L. Gallo-Martinez, A. Sevillano-Cabeza, P. Campíns-Falcó, F. Bosch-Reig, 'A New Derivatization Procedure for the Determination of Cephalexin with 1,2-Naphthoquinone-4-sulphonate in Pharmaceutical and Urine Samples Using Solid-phase Extraction Cartridges and UV–Visible Detection', *Anal. Chim. Acta*, **370**, 115–123 (1998).
17. A.C. Mehja, 'Direct Separation of Drug Enantiomers by High Performance Liquid Chromatography with Chiral Stationary Phases', *J. Chromatogr.*, **426**, 1–13 (1988).
18. S. Görög, M. Gazdag, 'Enantiomeric Derivatization for Biomedical Chromatography', *J. Chromatogr. B*, **659**, 51–84 (1994).
19. M.D. Pastor-Navarro, R. Porras-Martinez, R. Herráez-Hernández, P. Campíns-Falcó, 'Automated Determination of Amphetamine Enantiomers Using a Two-dimensional Column-switching Chromatographic System for Derivatization and Separation', *Analyst*, **123**, 319–324 (1998).
20. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Solid-phase Extraction Techniques for Assay of Diuretics in Human Urine Samples', *J. Liq. Chromatogr.*, **14**(9), 3575–3590 (1991).
21. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Improved Detection Limits for Screening of Diuretics by Coupled Liquid Chromatography and UV–V Spectrophotometry', *J. Chromatogr. B*, **612**, 245–251 (1993).
22. C.K. Lai, T. Lee, K.M. Au, A.Y.W. Chan, 'Uniform Solid-phase Extraction Procedure for Toxicological Drug Screening in Serum and Urine by HPLC Photodiode-array Detection', *Clin. Chem.*, **43**(2), 181–184 (1997).
23. L. Gallo-Martinez, P. Campíns-Falcó, A. Sevillano-Cabeza, R. Herráez-Hernández, 'Improved Solid-phase Extraction Procedure for Assay of Cephalosporins in Human Urine Samples', *J. Liq. Chromatogr.*, **21**, 2191–2203 (1998).
24. Y. Gaillard, G. Pepin, 'Screening and Identification of Drugs in Human Hair by High-performance Liquid Chromatography–Photodiode-array Detection and Gas Chromatography–Mass Spectrometry after Solid-phase Extraction. A Powerful Tool in Forensic Medicine', *J. Chromatogr. A*, **762**(1–2), 251–267 (1997).
25. P. Campíns-Falcó, F. Bosch-Reig, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Determination of Teophiline and Paraxanthine in Urine Samples by HPLC Using the H-point Standard Additions Method', *Anal. Chim. Acta*, **268**, 73–80 (1992).
26. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Determination of Caffeine in Human Urine Samples Free of the Interference of its Metabolites by Reversed-phase Liquid Chromatography Using Solid-phase Extraction from Sample Clean-up', *J. Liq. Chromatogr.*, **16**, 1297–1314 (1993).
27. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Sensitive Determination of Ethacrinic Acid in Urine Samples by Reversed-phase Liquid Chromatography with UV–Visible Detection Using Solid-phase Extraction Techniques for Sample Clean-up', *Anal. Chim. Acta*, **270**, 39–44 (1992).
28. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Sensitive Determination of Probenecid in Urine Samples by Reversed-phase Liquid Chromatography and UV–Visible Detection Using Solid-phase Extraction Techniques for Sample Clean-up', *Chromatographia*, **35**(5/6), 317–320 (1993).
29. K.M. Clauwaert, J.F. Van Bocxlaer, W.L. Lambert, A.P. De Leenheer, 'Analysis of Cocaine, Benzoylecgonine, and Cocaethylene in Urine by HPLC with Diode Array Detection', *Anal. Chem.*, **68**, 3021–3028 (1996).
30. P. Campíns-Falcó, C. Molins-Legua, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Improved Amphetamine and

Methamphetamine Determination in Urine by Normal-phase High-performance Liquid Chromatography with Sodium 1,2-Naphthoquinone-4-sulphonate as Derivatizing Agent and Solid-phase Extraction for Sample Clean-up', *J. Chromatogr. B.*, **663**, 235–245 (1995).

31. C. Molíns-Legua, P. Campíns-Falcó, A. Sevillano-Cabeza, 'Amphetamine and Methamphetamine Determination in Urine by Reversed-phase High-performance Liquid Chromatography with Sodium 1,2-Naphthoquinone 4-sulphonate as Derivatizing Agent and Solid-phase Extraction for Sample Clean-up', *J. Chromatogr. B*, **672**, 81–88 (1995).
32. S. Rudaz, J.L. Veuthey, 'Stereoselective Determination of Methadone in Serum by HPLC Following Solid-phase Extraction on Disc', *J. Pharm. Biomed. Anal.*, **4**(8–10), 1271–1279 (1996).
33. Y. Bergqvist, L. Funding, B. Krysen, T. Leek, K. Yvell, 'Improved Validated Assay for the Determination of Proguanil and its Metabolites in Plasma, Whole Blood, and Urine Using Solid-phase Extraction and High-performance Liquid Chromatography', *Ther. Drug Monit.*, **20**(3), 325–330 (1998).
34. R. Shimoyama, T. Ohkubo, K. Sugawara, T. Ogasawara, T. Ozaki, A. Kagiya, Y. Saito, 'Monitoring of Phenytoin in Human Breast Milk, Maternal Plasma and Cord Blood Plasma by C_{18} Solid-phase Extraction and Liquid Chromatography', *J. Pharm. Biomed. Anal.*, **17**(4–5), 863–869 (1998).
35. A.W.E. Wright, J.A. Lawrence, M. Iu, T. Cramond, M.T. Smith, 'Solid-phase Extraction Method with High-performance Liquid Chromatography and Electrochemical Detection for Quantitative Analysis of Oxycodone in Human Plasma', *J. Chromatogr.*, **712**, 169–175 (1998).
36. R. de la Torre, J. Ortuño, J.A. Pascual, S. Gonzalez, J. Ballesta, 'Quantitative Determination of Tricyclic Antidepressants and Their Metabolites in Plasma by Solid-phase Extraction (Bond-Elut TCA) and Separation by Capillary Gas Chromatography with Nitrogen–Phosphorus Detection', *Ther. Drug Monit.*, **20**(3), 340–346 (1998).
37. J. Szignan, D. Lange, P. Gerhards, 'Determination of Designer Drugs and Ecstasy', *Lab. News*, **581**, 12–14 (1998).
38. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Column-switching Techniques for Screening of Diuretics and Probenecid in Urine Samples', *Anal. Chem.*, **66**, 244–248 (1994).
39. P. Campíns-Falcó, R. Herráez-Hernández, M.D. Pastor-Navarro, 'Analysis of Diuretics in Urine by Column-switching Chromatography and Fluorescence Detection', *J. Liq. Chromatogr.*, **20**(12), 1867–1885 (1997).
40. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Application of Column-switching Techniques to the Determination of Medium Polarity Drugs: Determination of Acetazolamine in Urine', *J. Chromatogr. B*, **654**, 85–90 (1994).
41. P. Campíns-Falcó, R. Herráez-Hernández, A. Sevillano-Cabeza, 'Determination of Triamterene in Urine by HPLC Using Fluorescence Detection and Column-switching', *Chromatographia*, **38**(1/2), 29–34 (1994).
42. D. Wasterlund, 'Direct Injection of Plasma into Column Liquid Chromatographic Systems', *Chromatographia*, **24**, 155–164 (1987).
43. I.S. Krull, M.E. Szulc, A.J. Bourque, F.-X. Zhou, J. Yu, R. Strong, 'Solid-phase Derivatization Reactions for Biomedical Liquid Chromatography', *J. Chromatogr.*, **659**, 19–50 (1994).
44. J.M. Rosenfeld, M. Mureika-Russell, M. Love, 'Solid-phase Sample Preparation Method for Prostaglandins: Integration of Procedures for Isolation and Derivatization for Gas Chromatographic Determination', *J. Chromatogr.*, **489**, 263–272 (1989).
45. P. Campíns-Falcó, A. Sevillano-Cabeza, C. Molins-Legua, M. Kohlmann, 'Amphetamine and Methamphetamine Determination in Urine by Reversed-phase High Performance Liquid Chromatography with both Sample Clean-up and Derivatization with NQS on Solid-phase Cartridges', *J. Chromatogr. B*, **687**, 239–246 (1996).
46. R. Herráez-Hernández, P. Campíns-Falcó, A. Sevillano-Cabeza, 'On-line Derivatization into Precolumns for the Determination of Drugs by Liquid Chromatography and Column-switching: Determination of Amphetamines in Urine', *Anal. Chem.*, **68**(5), 734–739 (1996).
47. R. Herráez-Hernández, P. Campíns-Falcó, Adela Sevillano-Cabeza, 'Liquid Chromatographic Analysis of Amphetamine and Related Compounds in Urine Using Solid-phase Extraction and 3,5-Dinitrobenzoyl Chloride for Derivatization', *J. Chromatogr. Sci*, **35**, 169–175 (1997).
48. R. Herráez-Hernández, P. Campíns-Falcó, A. Sevillano-Cabeza, 'Determination of Amphetamine and Related Compounds in Urine Using On-line Derivatization into Octadecyl Silica Columns with 9-Fluorenylmethyl Chloroformate and Liquid Chromatography', *J. Chromatogr. B*, **679**, 69–78 (1996).
49. C. Molins-Legua, P. Campíns-Falcó, A. Sevillano-Cabeza, 'Off-line Dansylation of Amines Using C_{18} Solid-phase Packings: Study of the Fluorescence and Chemiluminescence Detection by Post-column Derivatization with Oxalic Acid Bis(2,4,6-trichloro phenyl ester)/H_2O_2 in Liquid Chromatography. Determination of Amphetamines in Urine Samples', *Anal. Chim. Acta*, **378**, 83–93 (1999).
50. C. Molins-Legua, P. Campíns-Falcó, A. Sevillano-Cabeza, M. Pedrón-Pons, 'Urine Polyamines Determination Using Dansyl Chloride Derivatization in Solid-phase Extraction Cartridges and HPLC', *Analyst*, **124**, 477–482 (1999).
51. C.C. Hansch, in *Comprehensive Medicinal Chemistry*, eds. R.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, Vol. 6, 1990.
52. R. Ventura, J. Segura, 'Detection of Diuretic Agents in Doping Control', *J. Chromatogr. Biomed.*, **687**, 127–144 (1996).

Steroid Analysis

Leonard C. Bailey, Thomas Medwick, and Yijie Dong
Rutgers University, Piscataway, USA

1 Introduction **2**
2 Discovery and History **2**
3 Medicinal Categories **2**
4 Methods used in Steroid Analysis **3**
4.1 Ultraviolet/Visible Spectroscopy 3
4.2 Infrared Spectroscopy 3
4.3 Nuclear Magnetic Resonance Spectroscopy 4
4.4 Chromatographic Methods 4
4.5 Mass Spectrometry 6
4.6 Immunoassay Methods 7
5 Sterols **8**
5.1 Gas Chromatography 8
5.2 High-performance Liquid Chromatography 8
6 Estrogens **9**
6.1 Identification Tests for Estrogens 9
6.2 Assays of Raw Materials and Medicinal Products 9
6.3 Analysis of Estrogens in Biological Media 9
6.4 Estrogen Analysis by Gas Chromatography 11
6.5 Estrogen Analysis by High-performance Liquid Chromatography 12
6.6 Bioassays Using High-performance Liquid Chromatography 12
7 Androgens **12**
7.1 Identification Tests for Androgens 12
7.2 Assays of Raw Materials and Medicinal Products 13
7.3 Analysis of Androgens in Biological Media 13
7.4 High-performance Liquid Chromatography 14
7.5 Mass Spectrometry 14
7.6 Immunoassay of Androgens 14
8 Progestins **14**
8.1 Identification Tests for Progestins 15
8.2 Assays of Raw Materials and Medicinal Products 15
8.3 Analysis of Progestins in Biological Media 15
8.4 Measurement of Progestins 16
8.5 Progestin Analysis by Gas Chromatography 16
8.6 Progestin Analysis by High-performance Liquid Chromatography 17
9 Contraceptives **18**
10 Corticosteroids **18**
10.1 High-performance Liquid Chromatography 18
10.2 High-performance Liquid Chromatography/Mass Spectrometry 19
10.3 Gas Chromatography/Mass Spectrometry 19
10.4 Immunoassay Combined with High-performance Liquid Chromatography and/or Gas Chromatography 19
10.5 Micellar Electrokinetic Capillary Chromatography 20
11 Cardiac Steroids **20**
11.1 Separation and Purification 20
11.2 Immunoassays 20
11.3 High-performance Liquid Chromatography 20
11.4 Thin-layer Chromatography 21
11.5 Micellar Electrokinetic Chromatography 21
12 Vitamin D **21**
12.1 United States Pharmacopeia Analysis 21
12.2 Gas Chromatography 21
12.3 High-performance Liquid Chromatography 21
12.4 Protein Binding Analysis 22
12.5 Mass Spectrometry 22
13 Bile Acids **22**
13.1 Analysis of Bile Acids in Biological Media 22
13.2 General Methods 22
13.3 Enzymatic Analysis 22
13.4 Gas Chromatography 22
13.5 High-performance Liquid Chromatography 22
13.6 Mass Spectrometry 23
13.7 Capillary Zone Electrophoresis 23
Abbreviations and Acronyms **23**
Related Articles **24**
References **24**

The successful determination of steroids in dosage forms and biological samples presents particular problems to the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

analyst in view of the great number of members of this class and their varied properties. This article discusses, according to their medicinal categories, the various classes of steroids and steroid-like compounds which can be encountered by the pharmaceutical analyst, the general methods used in their determination, and the application of these methods to nine classes of steroidal substances.

1 INTRODUCTION

The successful determination of steroids in dosage forms and biological samples presents particular problems to the analyst in view of the great number of members of this class and their varied properties. The addition of any of a number of various functional groups to the generally non-polar cyclopentanophenanthrene nucleus yields compounds of widely different polarities, making extraction from the various matrices a uniquely difficult task. Steroids exhibit uncertain stability, especially to heat, as experienced in gas chromatography (GC), thereby frequently making derivatization necessary. The generally nonconjugated nature of the final structures has few chromophores, making detection and quantitation by ultraviolet (UV) analysis problematic. The extremely low levels found in biological samples make necessary extensive extraction and concentration before analysis. This article discusses, according to their medicinal categories, the various classes of steroids and steroid-like compounds which can be encountered by the pharmaceutical analyst, the general methods used in their determination, and the application of these methods to nine classes of steroidal substances.

2 DISCOVERY AND HISTORY

Steroids possess the skeleton of cyclopentanophenanthrene or a similar skeleton derived by one or more bond scissions, or ring expansions or contractions. An alkyl side chain may be present at C-17, and methyl groups are often present at C-10 and C-13. The ring lettering and the numbering systems for steroids are illustrated in Figure 1,

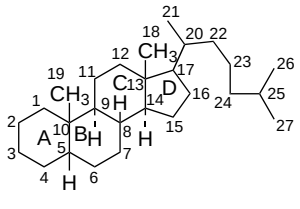

Figure 1 The ring lettering and the numbering system for steroids, using 5α-cholestane.

using 5α-cholestane. Steroid nomenclature derives from this compound for sterols and from three other simple steroids. These are 5α-androstane, the precursor of the androgens, which has no side chain at C-17, 5α-pregnane, the typical compound for the progestins and corticoids, which has a $-C_2H_5$ at C-17, and 5α-estrane, which has no C-17 side chain and no C-19 methyl group.

Because of the important biological activities and wide medicinal use of steroids, the analysis of these compounds is important in both clinical biochemistry practice and industrial applications. The great diversity of steroid structures and the wide range of their physical properties present special problems for the analysis of the different classes of steroids.

3 MEDICINAL CATEGORIES

The major therapeutic classes of steroids, defined by their physiological function, are the sterols, the androgens, the estrogens, the progesterones, the corticoids (which include the glucocorticoids and the mineralocorticoids), and vitamin D. Other therapeutically important substances with steroid-like structures include bile acids, and the cardiac glycosides.

Sterols are steroids carrying a hydroxyl group at C-3 and most of the skeleton of cholestane. Additional carbon atoms may also be present in the side chain. Among the sterols, cholesterol is unique because it is the biochemical precursor of virtually all other steroid hormones. The determination of free cholesterol in human serum is of great clinical importance in the diagnosis of hepatic disease and other studies of clinical pathologies.

Androgens and anabolic agents include natural hormones and semi-synthetic analogs. Naturally occurring androgenic hormones include testosterone and its potent metabolite dihydrotestosterone. Testosterone has two important functions: androgenic and anabolic. The primary use of androgens and anabolic agents is as androgen replacement therapy in men.

The estrogenic steroids exhibit similar chemical characteristics, which include an aromatic A ring and a hydroxyl group at the C-3 position. Conjugation appears very important in estrogen transport and metabolism and primarily involves the combination of estrogen with glucuronic acid and, to a lesser degree, with sulfate. The conjugated estrogens also have physiological activity.

Estrogens are of clinical and analytical interest for many reasons. They have been widely used in birth control, the prevention and treatment of osteoporosis, the treatment of estrogen deficiency, and the treatment of advanced breast and prostate cancer.

Equine estrogens (conjugated estrogens, esterified estrogens) are largely mixtures of estrone (E1) sodium

sulfate and equilin sodium sulfate with traces of other estrogenic substances. Originally they were obtained from horses, in particular pregnant mares, but now they also are produced semisynthetically from diosgenin and other natural precursors.

The progesterones are involved in the preparation for and maintenance of pregnancy. Progestins are compounds that have biological activities similar to progesterone. This group consists of progesterone (4-pregnene-3,20-dione) and its derivatives, and 19-nortestosterones. The most common use of the progestins, as for the estrogens, is for inhibition of ovulation.

Corticoids are steroid hormones produced by the adrenal cortex. The adrenal cortex hormones are classified by their biological activities into three major groups: mineralocorticoid, glucocorticoids with moderate to low salt retention, and the glucocorticoids with high salt retention. The most important steroids produced by the adrenal cortex are aldosterone which is the primary mineralocorticoid in humans, and hydrocortisone, the primary glucocorticoid in humans which has become very useful for its anti-inflammatory effects. The glucocorticoids have a profound effect on carbohydrate metabolism. Mineralocorticoids affect ions involved in the regulation of calcium transport. Apart from mammalian hormones, the steroid structure is found in many naturally occurring compounds of biological interest such as cardiac glycosides, bile acids, sterols, ecdysterols, and steroid sapogenins and certain antibiotics.

Cardiac steroids are a group of compounds that increase the contractility of the heart and can be very useful in the treatment of congestive heart failure. The most familiar of these is digoxin. Since its therapeutic index is relatively narrow and toxicity is a serious problem in therapeutic use, accurate analytical methods are essential. The cardiac steroids are found in a variety of plant species. *Digitalis purpurea* and *D. lanata* have been the most important overall sources of cardiac steroids.

The bile acids are anionic detergents biosynthesized from cholesterol and found in the intestinal tract and in bile. Their action is to emulsify fats, fat-soluble vitamins, and other lipids so that they may be absorbed. Almost all of the cholesterol not used for the formation of steroid hormones is converted to bile acids, which then react with glycine or taurine to form conjugates, which are secreted in the bile. In the intestine, the conjugates are converted to cholic acid, deoxycholic acid, and several other bile acids, many of which are then reabsorbed. Human bile contains five major bile acids, viz., cholic, chenodeoxycholic, deoxycholic, ursodeoxycholic, and lithocholic acid.

The vitamins D are fat-soluble seco-steroids, so called because they have undergone a fission of the B ring between carbons 9 and 10. Ergocalciferol (vit D_2) is formed in plants and cholecalciferol (D_3) in animals and both have identical but low activity in humans until metabolized to active forms.

4 METHODS USED IN STEROID ANALYSIS

4.1 Ultraviolet/Visible Spectroscopy

Ultraviolet/visible (UV/VIS) spectroscopy, which became available as a routine tool in the 1940s, has an important role in both quantitative and qualitative steroid analysis because of its good sensitivity and selectivity, low cost, and ease of automation. Using a scanning diode-array UV spectrophotometer as a detector following high-performance liquid chromatography (HPLC) separation broadens the applicability of both techniques.

UV/VIS spectrophotometry permits the easy characterization and quantification of those steroids containing conjugated unsaturated systems, such as '4-en-3-one' type, estrogens, and trienes of the vitamin D series. Unconjugated steroids with isolated double bonds can also be analyzed using lower wavelengths. A variety of chemical derivatizations have been employed to enhance the UV method in the analysis of UV non-absorptive steroids or to increase the sensitivity and selectivity. UV analysis of steroids depends usually on the identification of the following chromophoric groups.

A set of simple rules, developed through examination of UV spectroscopic data for a wide variety of compounds, including steroids, allows the prediction of λ_{max}, (often within 2–3 nm of the observed value) for conjugated dienes and polyenes.[1] The parent diene is assigned a base absorption at 214 nm and increments are added for substituents. For example, each additional double bond results in a shift of +30 nm and a heteroatom substituent adds 5 nm. For α,β-unsaturated ketones, the UV absorption associated with the $\pi \rightarrow \pi^*$ transition is the most important. The rules for calculating λ_{max} for such systems are similar with those rules for conjugated dienes and polyenes.[1]

The phenolic hydroxyl group at C-3 in the estrogens is a substituent of common interest in steroid UV analysis, owing to its effect on band maxima. For example, the UV absorption spectrum of estradiol (E2) shows a maximum at 280 nm, with strong absorption below 200 nm, which represents the main maximum of benzene. For steroids that are aromatic in both the A ring and the B ring, such as certain estrogens, the spectra are more complicated, with characteristic absorption at 280 and 342 nm, and also strong end absorption below 200 nm.

4.2 Infrared Spectroscopy

In the 1950s, infrared spectroscopy (IRS) was the only effective method for the identification and structure

elucidation of steroids. Although its importance decreased after the developments of nuclear magnetic resonance (NMR) and mass spectrometry (MS), it is still an essential tool for the rapid identification of steroids. It has rarely been used for quantitative analysis because of the low sensitivity and difficulty of sample preparation. The potassium bromide pellet method is mainly used and is generally specified in the United States Pharmacopeia (USP) 23 for the identification procedure.

Infrared (IR) spectra can provide evidence of functional groups in the sample, but can only give very limited information about the skeleton structure. In the analysis of steroids by IRS, most of the useful bands are above $1500\,cm^{-1}$. The pattern of peaks below $1500\,cm^{-1}$ cannot be assigned to specific vibrations, but occasionally it is characteristic of the steroid and might be useful for identification by comparing the spectrum with that of an authentic sample.

4.3 Nuclear Magnetic Resonance Spectroscopy

The major utility of NMR is to provide information about steroid structure, since it yields information about the molecular framework and functional groups, whereas IR only provides information on functional groups and UV only on conjugated systems. Another significant advantage of NMR is that it is non-destructive and samples can be recovered for further testing. More detailed information about applications of NMR in steroid analysis are available from various sources.[2,3,4,5]

4.4 Chromatographic Methods

Although steroids can be determined by using many methods, the identification and determination of the individual steroids require an efficient separation that can be performed only by extraction and chromatography. Chromatographic methods are important tools for the separation and the determination of steroids because of their high sensitivity, selectivity, efficiency, and separation power.

4.4.1 Thin-layer Chromatography

Thin-layer chromatography (TLC) is still one of the most important tools for the rapid separation and identification of steroids because of its ease of operation, its low cost of operation, and its ability to analyze many samples. The data obtained from TLC are used for the estimation of parameters in other analysis, such as capacity factor in HPLC.

Silica gel, a polar acidic material, is the adsorptive material most usually used and it can be modified for special purposes with other substances, such as bases, buffers, and salts. Aluminum oxide, a basic material, has been used for the separation of C-19 androgens. High-performance thin-layer chromatography (HPTLC) uses microparticulate silica and can achieve increased resolution owing to the higher surface area. Reversed-phase thin-layer chromatography (RPTLC), in which the surface is made non-polar by bonding C_{18}, C_8, or C_1 groups to the silica surface, is also available.

Aside from its relatively low resolution power compared to other chromatographic methods, one of the disadvantages of TLC is difficulty in visualization on the plate of the steroids of interest. For adsorptive steroids, UV spectroscopy can be employed. For nonadsorptive steroids, a wide variety of methods, which usually involve spraying and/or heating with some reagents, is available for visualizing the steroids of interest. Two other optical methods have been used, namely direct fluorescence detection following UV excitation, and fluorescence quenching. The latter technique involves incorporation of a fluorescent compound into the silica gel; the fluorescence of this compound is quenched in the presence of a UV absorber, and renders the compound visible. All of these visualization techniques can be divided into non-destructive and destructive methods. Since inspection of a TLC plate by eye has only a sensitivity of 1 to 10 μg and a reproducibility of about 20%, instrumental detection is considered important and sometimes essential for TLC.

4.4.2 Gas Chromatography

Since its introduction, GC has rapidly become a standard method in the biological and clinical analysis of steroids owing to its high sensitivity, selectivity, and separation power. In modern practice, capillary columns, usually wall-coated open tubular (WCOT) columns in which the stationary phase is coated on the inside of a column with an inside diameter of about 0.2 mm, have replaced the once widely used packed column. These capillary columns may have lengths up to 30 m, and now they provide one of the most powerful means for the separation of steroids. Megabore columns with internal diameters of 0.5–0.75 mm and greater capacity but decreased resolution are also available.

In GC, more than in HPLC and TLC, efficient separation depends mainly on the selection of an appropriate stationary phase since there is a limited choice of mobile phases. The stationary phase is a thermostable compound, which is coated on an inert support, usually diatomaceous earth, or directly on the walls of the column. Chemical bonding of the stationary phase to the support or the column wall reduces ''bleed'' or the slow elution of the stationary phase from the column. Chemical modification can be used to produce a variety of stationary phases with different selectivities.

In the GC of steroids, silicone oils are the most generally employed mobile phases, since they are suitable for the

separation of steroids based on their molecular weight or shapes. The carrier gas is usually nitrogen because it is cheaper, although helium is less dense than nitrogen and gives better resolution. Capillary GC separation of steroids and derivatives are best achieved by using non-polar stationary phases. Only estrogens, sterols, and bile acids require slightly more polar phases.

In order to improve the separation and/or to increase the sensitivity of detection in some cases, chemical derivatizations are frequently necessary for the GC analysis of steroids. The choice of derivatives depends on both the detection method and the chromatographic requirements. Post- and precolumn reactions may be employed. For hydroxyl groups, ethers are usually formed, most commonly trimethylsilyl (TMS) ether derivatives. Other ethers also have been used, such as dimethylethylsilyl ethers for the separation of bile acids and halogenated ethers for use with electron capture detection (ECD). The use of tertiary butyldiethylsilyl ethers is very common for quantitative gas chromatography/mass spectrometry (GC/MS), because these derivatives give mass spectra without extensive fragmentation. Other derivatives such as formates and acetates have also been used. Oxo groups can be derivatized as oximes or enol-TMS ethers. For the carboxyl group, methyl esters are usually formed. The carboxyl group is frequently derivatized precolumn, usually by forming methyl ester derivatives with diazomethane.

Three kinds of detector have been used in steroid analysis. The flame ionization detector is the most commonly used, since it can give out a signal that varies linearly in response factor to most of the steroids and has a large linear dynamic range. Nitrogen–phosphorus detection (NPD), also known as thermionic ionization detection (TID), can detect nitrogen- and/or phosphorus-containing steroids or derivatives selectively. ECD is extremely useful for halogen containing compounds. However, since steroids do not usually contain nitrogen or halogen, derivatization is necessary before the use of gas chromatography/nitrogen–phosphorus detection (GC/NPD) and gas chromatography/electron capture detection (GC/ECD).

GC/ECD has been replaced by GC/MS because of the difficulty in quantification by using ECD, since it has a narrow range of linearity. When operated properly, GC/MS can be as sensitive and more selective then GC/ECD. A major advantage of GC/MS is that some of the steroids can be detected without derivatization, but its high cost limits its use in routine determinations.

4.4.3 High-performance Liquid Chromatography

HPLC has become an important tool for the separation and the determination of steroids because of its high sensitivity, selectivity, efficiency, and better chromatographic properties compared to the other methods.

One of the main advantages of HPLC over other chromatographic methods is that no decomposition or rearrangement occurs during the chromatography as in GC and TLC. In addition, HPLC is a powerful tool for the separation of conjugated compounds and isomers, which are relatively difficult to separate by GC. Although derivatization procedures have been described for almost all of the steroids in HPLC, almost all of them are used for increasing detectability and have no important effect on the separation.

Usually, HPLC columns used in steroid analysis are 10–50 cm in length, 4–5 mm in internal diameter, and with 60 000–80 000 theoretical plates per meter. Microbore HPLC using columns 20–50 cm long and with internal diameters of 0.5–2 mm can provide higher sensitivity and efficiency, although it requires specialized pumping equipment and sample injectors as well as microvolume detector cells.

Both normal and reversed-phase high-performance liquid chromatography (RPHPLC) can be used for steroid separation. In normal HPLC, high-performance silica and alumina stationary phases have been successfully used. In RPHPLC, silica packing bonded on the surface with octadecyl or diol groups are most used. The separation depends largely on partition so that the selectivity varies with the length of the carbon chain on the column, the form of the support material, and the properties of the mobile phase.

Mobile phases consisting of only one or two solvents are those mainly used in steroid separations; however, multisolvent mobile phases may be especially useful for particular separations. A preliminary analysis by TLC can be used to test the suitability of the mobile phase chosen. Although most steroids are well separated by HPLC, the best separation of a variety of steroids has been achieved by gradient elution, which includes linear, non-linear, and stepped modes. Beside solvent constitution, the pH, as well as added ion-pair reagents and modifiers are also influential factors on the selectivity of the mobile phase. Ion-pair chromatography is preferred for the separation of steroid conjugates and bile acids. Micellar chromatography, in which some compounds, such as sodium dodecyl sulfate (SDS) are added, has been claimed to improve the separation of HPLC, especially the separation of optical, geometrical, and structural isomers. Nishi has reviewed pharmaceutical applications of micellar liquid chromatography (MLC).[6]

UV absorptive steroids or derivatives can be detected directly by UV detection, which is the most widely used detection technique in the conventional HPLC analysis of steroids. Post- and precolumn derivatization have been used to form UV-absorbing derivatives, which may also increase the sensitivity. Since the sensitivity of HPLC with UV detection is around 1 ng per injection, UV

detectors cannot provide enough sensitivity and selectivity for many steroids, in particular the polar steroids, such as bile acids. As a consequence, laser-induced fluorescence (LIF) detection may be used[7] to increase the limits of detection and quantitation. Fluorescence and electrochemical detectors have achieved determination of steroids at the picogram level. The sensitivity of detection has been extended by developments in biospecific detection principles, e.g. radioimmunoassay (RIA) coupled to HPLC.[8–12] This area has been reviewed.[13,14]

4.5 Mass Spectrometry

Modern MS methods can analyze almost all steroids and steroid conjugates. MS has been used as a qualitative analysis method by comparing the observed spectrum with the spectra of known compounds, and/or by deducing the structure from the presence of characteristic fragments. MS also has been used in steroid quantification since the intensity of signals is proportional to the quantity of the compound present.

The main two roles of MS in steroid analysis are structure elucidation and quantitative analysis. The MS measurements for structure elucidation include the determination of molecular weights, the determination of elemental and isotopic compositions, and the evaluation of the fragmentation characteristics of the analytes. For molecular weight determination, it is often desirable to suppress fragmentation through the use of soft ionization, but for other structural studies the fragmentation products of the analytes provide most information. Modern MS can offer high sensitivity for the qualitative and quantitative analysis of both isolated analytes and of specific analytes in a mixture.

MS analysis involves three essential steps: ionization of the sample, resolution of the resulting ions, and detection of the ions. Sample introduction and data processing are also integral and important parts of MS. The principal techniques of ionization and sample introduction employed in MS analysis of steroids are summarized in Table 1.

Steroids can be analyzed directly by MS using a heated sample probe, which allows the insertion of the analyte directly into the vacuum and in the vicinity of the ionization chambers. However, because of the nature of most samples presented for analysis, it is usually more effective to integrate MS on-line with other techniques, which include GC, HPLC, supercritical fluid chromatography (SFC), and capillary electrophoresis (CE). In these cases, sample introduction can be performed through these instruments but usually a special interface is necessary to separate the analyte from the sample matrix.

MS methods are categorized mainly by the ionization techniques used, because the ionization step has the most important impact on the appearance of the mass spectrum and the application of MS. The methods using low internal energy, such as CI and field ionization are called 'soft' ionizations and give a high yield of protonated and cation-coupled molecular ions with little or no fragmentation. Other methods, in which high internal energy is transferred, such as electron impact (EI) ionization, usually cause extensive fragmentation of the sample.

Depending on the type of instrument used, the ions have kinetic energy from a few eV to as high as 10 KeV. Resolution of the ions expelled from the ion source according to their m/z ratios is accomplished by studying their flight through one or more analyzers, including momentum (magnetic sector), resonance frequency (quadrupole, ion cyclotron resonance and ion trap), kinetic energy (electric sector) and velocity (time-of-flight). The most commonly used MS ion-resolution methods for steroid analysis are the magnetic sector and quadrupole instruments. They are compatible with almost all of the ionization methods except plasma desorption ionization.

Data processing systems assure the efficient use of the enormous amount of data generated by MS and the precise control of complex instrumental parameters for various applications. The main functions of the data systems include the digitization and rapid storage of the data, mass and intensity assignments for every ion, and representation of the data in an easily comprehensible graphic or tabulated form.

Determination of the isotope composition of steroids is another important application of MS in steroid analysis,

Table 1 Principal techniques of ionization and sample introduction employed in steroid analysis by MS. (From Makin and Gower[1])

Ionization techniques	Sample introduction	Analyte type
Electron ionization	Direct insertion probe, GC	Unconjugated steroids (after derivatizations)
CI	Direct insertion probe, GC	Unconjugated steroids (after derivatizations)
Electron capture	GC	Unconjugated steroids (after derivatizations)
Thermospray	Flow injection analysis, LC	Unconjugated steroid glucuronides
FAB/SIMS	Direct insertion probe, GC Flow injection analysis, LC	Conjugated
Electrospray	Flow injection analysis, LC	Conjugated

CI, chemical ionization; FAB/SIMS, fast atom bombardment/secondary ion mass spectrometry; LC, liquid chromatography.

and can be easily accomplished with the use of low-resolution measurements. The ion clusters observed in every steroid spectrum are due to the naturally occurring minor isotopes such as ^{2}H, ^{13}C, and ^{18}O, etc. in which especially informative are the isotope pairs of ^{35}Cl and ^{37}Cl, and ^{79}Br and ^{81}Br, and the complex isotope pattern of certain metals. Because an isotope can be relatively easily measured by MS, stable-isotope labeled substances offer a convenient way to study the metabolism of steroids in humans. The ^{14}C radioisotope is also commonly used in drug metabolism studies at enrichment levels. For the same reason, ^{2}H, ^{13}C, and ^{18}O labeled steroids are extensively used in reaction and fragmentation mechanism studies and as internal standards in quantitative measurement.

4.5.1 Quantitative Determinations

Since its introduction, MS, especially when combined with GC or LC, has been one of the most useful forms of quantification available since it excels in sensitivity, accuracy, dynamic range, and specificity at trace level. As a result, MS has become the method of choice for the quantitative determination of many biologically important steroids. The various concepts of quantitative MS have been reviewed by Millard.[15]

4.5.2 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) provides a method for determining the precursor ions of the fragmentation ions. MS/MS can be performed by using a triple quadrupole combination or a double-focusing sector instrument. Hybrid instruments, which couple a double focusing instrument with a quadrupole collision cell and quadrupole mass analyzer, have also been used. Since MS/MS is so highly specific, low-level constituents in biological samples or chemical mixtures can frequently be determined with little or no sample purification.

Derivatizations are occasionally employed in MS for one or more of the following purposes: to meet the requirement of sample volatility, to improve sample stability, to improve chromatographic properties, to enable functional group analysis, and to modify the fragmentation pattern.

4.5.3 Gas Chromatography/Mass Spectrometry and High-performance Liquid Chromatography/Mass Spectrometry

GC/MS is one of the most important techniques in pharmaceutical studies and in steroid analysis because the resolving power of capillary GC is combined with the high-specificity and sensitivity of MS. For instance, since isomeric steroids may be difficult to distinguish by MS but can be easily separated by GC, they can be identified by GC/MS. Similarly, the combination of HPLC with MS/MS offers special importance in the analysis of large numbers of samples in complex matrices. When coupled with GC or HPLC, the MS can be used in a selected/multiple ion monitoring mode and/or total monitoring mode. The most common and informative mode of operation is scanning the mass spectrum at frequent intervals across the full mass range. The individual spectra from each peak can be subsequently examined and used for identifying the analytes.

4.6 Immunoassay Methods

Competitive binding assays, in which the binding proteins included antibodies and specific receptors, were introduced by the early 1970s. Immunoassays have been proven to be more appropriate analytical tools for steroids in biological materials, such as serum or plasma, and alternatively urine and saliva. The current immunoassays for steroids have much improved sensitivity and precision as well as reduced cost and simplified procedures.

Although immunoassays can be highly specific and sensitive, errors may occur because both the characteristics of the antibody employed and the sample matrix can affect the assay. Cross-reactivity is always a major concern in immunoassays. For steroid analysis, cross-reactivity may come from structurally related steroids, derivatives of the analyte, or metabolites of the analyte.

Both monoclonal and polyclonal antibodies have been used in immunoassays of steroids The main problem in the antibody production is that steroids and relative compounds are too small to be immunogenic. They have to be coupled to large carrier compounds, proteins such as bovine serum albumin (BSA), to elicit the production of antibodies. Many factors of the steroid–protein conjugates may affect the antibodies produced, including the method used, the position and number of the linkage, and the steric structure of the conjugate.

Several kinds of label have been used for detection, including radioactive, enzymatic, fluorescent and chemiluminescent compounds. Radioactive labels are most commonly used, but there is an increase in the use of non-radioactive labels, mainly because these procedures are safer, cheaper, simpler, and sometimes more precise and sensitive.

Tritium has been the most widely used radiolabel in the immunoassays of steroids, although iodinated labels are also frequently used. Multiple labeling of the steroid is usually necessary in order to produce high specificity. The main advantages of iodine labels over tritium labels include higher specificity, lower cost and lower quenching.

Enzymes have been proven to be sensitive and versatile labels in the immunoassays of steroids because the end-point signal is amplified by the enzymatic reaction.

Versatility results from the variety of enzymes and end-point reactions that can be used. The most widely used enzymes in steroid analysis are horseradish peroxidase (HRP) and alkaline phosphatase.

Fluorescence detection is generally a very sensitive method, but practically its sensitivity is restricted to a concentration range of 10^{-9}–10^{-12} mol L^{-1} owing to the background signal produced by other fluorescent materials, particularly in biological samples. The fluorescent molecules that have been most commonly used in steroid analysis including fluorescein, rhodamine, umbelliferone, and rare earth chelates. The latter are especially useful because of the long half-life time and the differing fluorescent spectra of various rare earth chelates, which makes possible the simultaneous measurement of several analytes by using different chelates, such as those of europium, samarium, and terbium.

Chemiluminescence is effected by a chemical reaction, usually an irreversible oxidation reaction. This reaction may be mediated by enzymes, temperature, or electric charge applied, and they are termed, respectively, bioluminescence, thermochemiluminesence and electrochemiluminescence. Many organic compounds are luminescent upon oxidation and have been used in immunoassays as labels, including the cyclic hydrazides such as isoluminol (6-amino-2,3-dihydrophthalazine-1,4-dione) and its derivatives, aminonaphthyl hydrazides, and acridinium esters and their derivatives. Chemiluminescence has also been used in enzyme immunoassays (EIAs) as an end-point.

5 STEROLS

Because of the implications of free cholesterol in atherosclerosis, heart disease, and hepatic diseases, accurate methods for its analysis in human serum are necessary for precise clinical diagnosis. Accurate determination of cholesterol is also important in the food industry. Historically, the methods used for such analyses have been relatively simple to allow for rapid estimation, and have been based mainly on colorimetric or enzymatic reactions. Because these methods lack sensitivity and suffer from interference, they are not reliable in the analysis for cholesterol in physiological fluids other than plasma. Hence, several chromatographic methods[16] have been developed, including GC[17,18] and HPLC.[19]

5.1 Gas Chromatography

GC has great importance in the analysis of free sterols but, because of their polarity, it is often necessary to analyze them as their derivatized forms. The shape of the sterol molecule is critical in determining the separation for sterols in GC. Planar sterols generally have longer retention times than non-planar ones on a column of given polarity. Long capillary columns have been used for separation of isomeric C_{24} sterols. The great advantage of GC over other chromatographic techniques is that it is ideal for coupling with MS. The major disadvantage of GC is that many sterols are thermally unstable.

Rodriguez-Palmero[20] reported an accurate method for the determination of sterol content by capillary gas chromatography (CGC). The procedure includes: dichloromethane–methanol (2 : 1, v/v) extraction, saponification at 80 °C, derivatization to form TMS ethers and GC. The method shows good accuracy, precision, and sensitivity and is suitable for the determination of sterols in food.

Garcia Regueiro[21] developed a reproducible procedure for the determination of eight cholesterol oxides using on-column injection and solvent venting capillary gas chromatograph. The method consists of the combined use of solid-phase fractionation and TLC. The determination of cholesterol oxide TMS derivatives was performed by using CGC. On-column CGC gave absolute area relative standard deviation (RSD) values of 3% to 6%.

5.2 High-performance Liquid Chromatography

As the technology of HPLC has improved over the last decade, it has been more widely applied to the analysis of sterols, with both reversed-phase and normal-phase HPLC having been utilized. Specially synthesized stationary phases with chemically bonded β-cyclodextrin and tomatine have been used in order to attain better separation. The HPLC method has also been applied to the determination of cholesterol in liposome drug products and raw materials.

Since most sterols lack a strong chromophore, UV detection is generally unsatisfactory unless the detector is operated below 250 nm and even as low as 220 nm. Special purpose detectors, such as flame ionization, differential refractometric, and radiochemical have also been employed in order to improve detection.

Okazaki[22] compared the precipitation and direct methods for high-density lipoprotein cholestrol (HDLC) assay with HPLC. The HPLC method, which provided quantitative and qualitative information with high precision, was regarded as being a reliable approach for HDLC assay. The HPLC procedure can be also used to evaluate alternative methods for cholesterol assay.

Rezanka[23] reported the analysis of sterol esters using tandem high-performance liquid chromatography (HPLC/HPLC) and capillary gas chromatography/mass spectrometry (CGC/MS) with CI. Non-polar lipids were

separated by normal-phase HPLC into individual classes and the sterol esters were subsequently separated by RPHPLC. Further separation and identification of the fractions after HPLC were effected by CGC/MS with positive and/or negative CI.

6 ESTROGENS

6.1 Identification Tests for Estrogens

As specified by the USP 23, IR and UV at 200–400 nm are the identification tests prescribed for pharmaceutically important estrogens. The TLC method is used as the identification test for sterile E1 suspension and E2 tablets. For sterile E1 suspension, a 0.25-mm layer of chromatographic silica gel is used and the solvent system consists of a mixture of benzene and acetone (4 : 1). After developing, the plate is sprayed with dehydrated alcohol and sulfuric acid (3 : 1), and is then heated at 105 °C for ten minutes.

Their aromatic ring A and the phenolic group at position 3 can characterize estrogens. They show characteristically complicated absorption, with principal maxima at 280 and 342 nm, as well as very strong end absorption down to 200 nm. The pH of the solution is important in determining these phenolic substances. The spectrum of an estrogen with a free phenolic hydroxyl group has an absorption band of medium intensity at around 278 nm. However, at a pH above 9–10, the band shifts about 20 nm higher and the intensity increases owing to the ionization of the phenolic group.

In the IR technique, the potassium bromide pellet method is used. The characteristic group wavenumbers of estrogens are 1613–1610, 1590–1589, and 1503–1490 cm^{-1} which are due to the C=C or C≡C stretching. The diagnostic value of the absorptions due to the phenolic ring is not comparable with that of the data obtained by UV spectra.

6.2 Assays of Raw Materials and Medicinal Products

Because of their broad therapeutic spectrum, there are numerous pharmaceutical products containing estrogens. As shown in Table 2, HPLC is the most common method used in the analysis of raw materials and medicinal products of estrogens in the USP 23.

6.3 Analysis of Estrogens in Biological Media

Generally, analysis of estrogens in biological material requires extraction of the analyte from the sample, separation into individual components, and their determination. Sometimes it is possible to extract the estrogens from the sample directly by simple solvent extraction but, more often, it is necessary to hydrolyze conjugates

Table 2 Methods used in the assay of estrogens in USP 23

Compound	Method	Detector	Packing	Mobile phase or solvent
E1	HPLC	280 nm	ODS	ACN–0.05 M KH_2PO_4, 1 : 1
Injection and sterile suspension	Gravimetric			
E2 and tablets	HPLC	205 nm	ODS	ACN–H_2O, 55 : 45
Vaginal cream	HPLC	280 nm	ODS	ACN–H_2O, 1 : 1
Pellets and sterile suspension	Visible	520 nm		Iron-phenol reagent
Valerate and valerate injection and cypionate and cypionate injection	HPLC	280 nm	ODS	ACN–NH_4NO_3 0.033 M, 7 : 3
Estriol	UV 281 nm	–	–	EtOH
Ethinyl E2	HPLC	280 nm	ODS	ACN–H_2O, 1 : 1
Tablets	Visible	538 nm	–	MeOH–H_2SO_4, 3 : 7
Equilin	HPLC	280 nm	ODS	ACN–H_2O, 35 : 65
Conjugated estrogens and tablets and esterified estrogens and tablets	GC	FID	Phenylcyano-propyl methyl silicone	He
Estropipate and tablets	HPLC	213 nm	ODS	ACN–0.025 M KH_2PO_4, 65 : 35
Vaginal cream	HPLC	280 nm	ODS	ACN–MeOH–0.05 M KH_2PO_4, 18 : 18 : 64
Mestranol	Visible	545 nm	–	MeOH–H_2SO_4, 3 : 7
Dienestrol and cream	HPLC	254 nm	ODS	Degassed methanol

ODS, octadecyl silane; ACN, acetonitrile.

before the extraction. Several methods are available, including acid hydrolysis and enzymatic hydrolysis using β-glucuronidase and/or sulfatases.

There are a number of reports of the use of gel filtration for extracting conjugates. A review of isolation techniques using gel filtration has been published.[24] Separation of individual estrogen glucuronides and sulfates was achieved by using diethylaminoethyl Sephadex.[25] Simple extraction of estrogens from plasma or hydrolyzed urine has been achieved by using G-15 and G-25 Sephadex.[26,27]

A number of methods for the extraction of estrogens are summarized in Table 3. Martinetti[42] described a novel extraction method for the evaluation of plasma levels of E2, E1, and E1-S. The method consisted of solid-phase extraction (SPE) followed by a highly specific RIA. The sensitivity of the assay was 0.6 pg mL^{-1}, 2.0 pg mL^{-1} and 7.0 pg mL^{-1} for E2, E1, and E1-S respectively.

Hirako[41] reported the establishment of a reliable extraction method of bovine plasma E1-S followed by RIA. E1-S in bovine plasma was extracted with a small reversed-phase cartridge. The steroid conjugate retained in the cartridge was eluted with 40% (v/v) methanol. E1-S was separately recovered from other steroids by the stepwise increase in methanol concentration in the elution solvent. The recoveries of E1-S eluted with 40% methanol were more than 90%. The concentration measured by RIA with the eluent of 40% methanol was consistent for plasma extraction volumes of 0.5–2.0 mL.

Immunoaffinity extraction has been used for liquid chromatographic determination of estrogens. Ikegawa[43] reported an immunoaffinity extraction for the HPLC determination of equilin and its metabolites in plasma. The antibody raised against an equilin 3-*O*-carboxymethyl-ether-BSA conjugate was characterized as having a high affinity for equilin and equilenin. The adsorbates were recovered by elution with 90% aqueous (v/v) methanol without any interfering peaks on the chromatogram.

Quantification of the estrogen analytes has been accomplished by a variety of methods including colorimetry, fluorimetry, double-isotope derivatization, immunoassay, GC, GC/MS, and HPLC. The early methods may be outdated, but those with low specificity may be combined with modern separation methods like HPLC to offer useful assays. As an example, the quantitative structure–chromatographic retention relationship of six underivatized equine estrogens has been reported.[44]

The immunoassays in use are of the competitive type. Some efforts have been made towards non-competitive immunoassay of estrogens.[45] A wide variety of labels are used, including ^{125}I, ^{3}H, enzymes, or fluorescent compounds. Separation of the free and bound analyte can be done using dextran-coated charcoal (DCC), second antibody precipitation, or solid-phase assay. Some form of manipulation can avoid the separation step, e.g. delayed fluorescence. Alternative end-points have enabled some estrogen assays to be simplified and made more sensitive, such as EIA and fluorescence immunoassay (FIA) methods. The detection limits (DLs) achieved are nanograms per litre by RIA and EIA, and micrograms per litre by FIA.

A novel noncompetitive assay for the measurement of estrone-3-glucuronide (EG) in diluted urine was reported by Barnard.[45] The method is based on the use of two

Table 3 Extraction of estrogens (partly adapted from Makin and Gower[1])

Estrogen	Solvent (or method)	Sample dilution	Refs.
E1 and E2	Diethyl ether	1 : 1, Plasma at pH 9.0	28
Estriol	Diethyl ether	1 : 3, Plasma–water	29
Estriol	Ethyl acetate–hexane, 1 : 8	Plasma–buffer, 1 : 1	30
2-Hydroxyestrone	Dichloromethane	Plasma	31
6-Oxo-E2-17β	Diethyl ether	Plasma	32
E1 and E2	Diethyl ether	Diluted urine	33
E1 glucuronide and sulfate	Ethanol	Amniotic fluid	34
Estriol 16α-glucuronide	Precipitated with $(NH_4)_2SO_4$, redissolved in methanol–acetone	Urine	35
E1-S	THF–ethyl acetate	Plasma	36
Estrogen glucuronide and sulfates	Methylene blue	Plasma	37
Estrogen glucuronide and sulfates	Chloroform-*n*-butanol	Urine	38
E1 and E1 conjugates	Columns or cartridges of octadecylsilane-coated silica	Urine, serum, amniotic fluid	39, 40
E1-S	Reverse phase cartridge	Bovine plasma	41
E2, E1 and E1-S	Selective solid phase extraction	Plasma	42
Equilin, equilenin, their metabolites and glucuronic acid conjugates	Immunoaffinity extraction	–	43

THF, tetrahydrofuran; E1-S, estrone sulfate.

types of antibodies, the β-type and α-type that recognize different epitopes within the variable region of the primary anti-EG antibody (Ab1). The β-type antibody (Ab1) is analyte sensitive and competes with the analyte for an epitope of the primary antibody. On the other hand, the α-type is analyte insensitive, but does not bind the Ab1 in the presence of the β-type due to epitope proximity.

RIA still is the most widely used competitive immunoassay. Combined with chromatographic methods, especially HPLC, RIA is becoming a very powerful tool in the analysis of estrogens. Meanwhile, EIA and FIA have enabled some estrogen assays to be simplified and made more sensitive and have been used as bases for automated immuno-analyzer systems.

An EIA was used by De Lauzon[46] to predict ovulation in women. Henderson[47] reported that EG concentrations in urine were measured by an EIA. In this work, a monoclonal antibody of EG was generated and incorporated into competitive EIAs, with HRP as the enzyme label. The sensitivity achieved was 1.1 ± 0.1 pg per well.

Time-resolved fluorescence with rare earth chelate labels has been used[48] to facilitate the simultaneous measurement of more than one ligand. Barnard[49] reported the development of a time-resolved FIA for the simultaneous measurement of EG and pregnanediol-3-α-glucuronide (PG) in samples of early morning urine using appropriate rare earth chelates with differing emission spectra. The procedure used different lanthanide chelates as labels and measured europium and samarium fluorescence. The method demonstrated sensitivity and precision (all coefficients of variation, 5–8%) across the relevant working ranges for each analyte.

Martinetti[42] reported the development of a rapid and ultrasensitive RIA method for E2, E1, and E1-S determination with selective SPE followed by a highly specific RIA. The sensitivity of the assay was 0.6 pg mL^{-1}, 2.0 pg mL^{-1}, and 7.0 pg mL^{-1} for E2, E1, and E1-S, respectively.

Ranadive[50] described a rapid, convenient RIA of E1-S. The assay had a dynamic range of 0.05–90 μg L^{-1} with a DL of 0.009 μg L^{-1}. Cross-reactivities with structurally related estrogens were <5%.

Lonning[51] described a sensitive assay for measurement of plasma E1-S in patients on treatment with aromatase inhibitors. E1-S was separated, hydrolyzed, and purified as unconjugated E1. E1 was subsequently reduced to E2, purified, and measured by a highly sensitive RIA using E2-6-(*O*-carboxymethyl) oximino-(2(-)[^{125}I]) iodohistamine as a ligand. The sensitivity limit of the method was 2.7 pM.

Table 4 lists some typical immunoassay methods for the analysis of estrogenic steroids.

6.4 Estrogen Analysis by Gas Chromatography

GC has proven extremely useful for the measurement of estrogens in biological fluids, such as plasma and urine. Before the development of the immunoassays, GC with flame ionization detection or, after suitable derivatization, ECD provided highly sensitive methods for estrogen analysis. However, extensive purification is needed before GC analysis. Owing to the presence of hydroxyl groups in their structure, most estrogens are highly polar and require derivatization prior to analysis for complete separation. Derivatization reagents used included TMS ethers, methyloxime ethers, and ethyloxime ethers.

CGC, using stationary phases more polar than those used for neutral steroids, provides improved resolution. Among the polar phases used are OV-210 (McReynolds number 1520), OV-225 (McReynolds number 1813),

Table 4 Immunoassays for estrogen estimation

Method	Compounds	Antibody or ligand	Label	Sensitivity	Ref.
FIA	EG	2nd antibody against the anti-EG antibody (Ab1)	Fluorescence of Eu	50 nmol L^{-1}, cross-reaction < 0.1%	45
EIA	EG in urine	Monoclonal antibody of EG	HRP	1.1 ± 0.1 pg per well.	47
	E2-17β in saliva	Monoclonal antibody	HRP	7.3 pM, cross-reaction < 0.2%	52
FIA	EG, PG (simultaneous)	Monoclonal antibody of EG/EG/BSA	Fluorescence of Eu and Sm	40 nmol L^{-1}	48
RIA	E2, E1, and E1-S	–	–	E1 – 0.6 pg mL^{-1}, E2 – 2.0 pg mL^{-1} E1-S – 7.0 pg mL^{-1}	42
	E1-S	E1-6-*O*-carboxy-methyl-oxime : BSA conjugate	^{125}I	DL of 0.009 μg L^{-1}, cross-reaction < 5%	50
	Bovine saliva E1-S	–	–	29.7 pg per tube	53
	Plasma E1-S	E2-6-*O*-carboxy-methyl-oximino-2-[^{125}I] iodohistamine	^{125}I	2.7 pM	51

and Silar 10c (McReynolds number 3682). These have been successful in separating both human and equine estrogens. Estrogens are identified by retention time and preferably, in addition, by on-line MS. GC/MS has the potential to provide higher sensitivity for analysis as well as confirmation of structure although it is technically the most challenging method for analysis and is not the method of choice for routine analysis. The limit of detection is at the $\mu g\,L^{-1}$ level, comparable to FIA.

Bagnati[54] reported the analysis of diethylstilbestrol, dienestrol and hexestrol in biological samples by immunoaffinity extraction and GC-negative-ion CI MS. Novak[55] reported pregnenolone and E2 identification in some biological samples by means of GC/MS analysis.

Casademont[56] described simultaneous determination, in calf urine, of twelve anabolic agents (including diethylstilbestrol, hexestrol, dienestrol, 17β-E2, E1, and 17α-ethinylestradiol) as heptafluorobutyryl derivatives by capillary GC/MS. Urine samples were extracted with octadecylsilica columns and cleanup was performed in two steps with basic alumina and silica SPE cartridges. The extracts obtained were derivatized with heptafluorobutyric anhydride and analyzed by GC/MS. The coefficient of variation obtained was lower than 13%. The limit of detection was $0.5\,ng\,mL^{-1}$ for all the compounds studied.

6.5 Estrogen Analysis by High-performance Liquid Chromatography

HPLC with a suitable choice of column and solvent system can perform estrogen analysis with high specificity. Preparative HPLC has been used to separate steroids prior to analysis by chemical or immunological methods. Reversed-phase LC has been preferred for the separation of the intact conjugated estrogens and is widely used for quality control. The DLs of HPLC using UV detection ranged from 0.5 to $5\,mg\,L^{-1}$ and were at $\mu g\,L^{-1}$ levels when using a fluorescent detector.

Combining the separation power of HPLC and the selectivity and sensitivity of biochemical assays leads to powerful tools for the assays of compounds in complex matrices, e.g. HPLC combined with RIA in the analysis of estrogens. A novel approach is the use of biochemical interactions, such as a receptor affinity detection system (RAD), as an on-line detector linked to a LC separation column.[57]

β-Cyclodextrins have been investigated as a mobile phase additive to improve the separation of 17-α- and 17-β-E2 and E1,[58] and equilin and three E1 derivatives.[59] In these studies, evidence is presented for inclusion complexation between the model estrogens and β-cyclodextrins leading to improved resolution and shorter separation time.

Katayama described a sensitive method for the estimation of seven estrogens in plasma.[60] In that method, estrogens were derivatized to esters and the limit of detection was $1–2\,ng\,L^{-1}$ of plasma. Similar work by Su[61] reported the HPLC analysis of six conjugated and unconjugated estrogens in serum, in which the mobile phase was optimized with ACN and an ion-pairing agent (tetrabutylammonium hydroxide).

An HPLC method for the direct and simultaneous determination of estriol 3- and 16-glucuronides in pregnancy urine was described by Iwata.[62] The method is based on direct derivatization of the glucuronic acid moiety in estriol glucuronides in urine with 6,7-dimethoxy-1-methyl-2 (^{1}H)-quinoxalinone-3-propionylcarboxylic acid hydrazide. The resulting fluorescent derivatives were separated by column-switching chromatography using a first column for cleanup of the derivatives and a second column for the complete separation of the derivatives. The derivatives were detected spectrofluorimetrically at 445 nm with excitation at 367 nm. The DLs (signal-to-noise ratio = 3) for estriol 3- and 16-glucuronides were 150 and 180 fmol in a $5\,\mu L$ sample of urine (14 and $17\,ng\,mL^{-1}$ urine), respectively. The method is highly sensitive and simple without any cleanup such as conventional SPE.

6.6 Bioassays Using High-performance Liquid Chromatography

On-line coupling of a RPHPLC with a RAD system based on the human estrogen receptor is described by Oosterkamp.[57] The RAD is performed using a postcolumn reaction detection system with open-tubular reaction coils. Estrogens are detected by using a recombinant steroid binding domain of the human estrogen receptor as affinity protein, and a fluorescent estrogen, coumestrol, as reporter molecule. A DL of $5\,nmol\,L^{-1}$ was achieved for compounds with high affinity for the estrogen receptor, such as 17-β-E2 and diethylstilbestrol, and $20\,nmol\,L^{-1}$ for compounds with low affinity. The system can also be operated without requiring a separation step.

HPLC coupled with RIA has been used in the assessment of serum E2 and E1 levels during hormone replacement therapy.[63] It was also reported as a sensitive method in the determination of intratumor E1 and E2 in primary breast cancer tissues.[64]

7 ANDROGENS

7.1 Identification Tests for Androgens

As specified by the USP, IR and UV at 200–400 nm are the identification tests available and common for pharmaceutically important androgens. TLC also has been used as an identification test for some androgens.

Table 5 Methods used in the assay of androgen-containing medicinals in USP 23

Compound	Method	Detector	Packing	Mobile phase or solvent
Testosterone and sterile suspension	UV	241 nm	–	MeOH
Cypionate and cypionate injection	GC	FID	1% Trifluoropropyl-methyl polysil-oxane on siliceous earth	He
Enanthate and enanthate injection and propionate and propionate injection	UV	380 nm	–	Isoniazid Reagent
Nandrolone decanoate	UV	239 nm	–	EtOH–$CHCl_3$
injection	UV	380 nm	–	Isoniazid Reagent
Phenpropionate	UV	239 nm	–	EtOH–$CHCl_3$
injection	UV	380 nm	–	Isoniazid Reagent
Oxandrolone	Back titration	–	–	HCl
Tablets	GC	FID	Methylsilicone oil on siliceous earth	He
Oxymetholone and tablets	UV	315 nm	–	MeOH
Stanozolol	Nonaqueous titration	–	–	$HClO_4$
tablets	UV	235 nm	–	EtOH
Methyltestosterone	HPLC	241 nm	ODS	ACN–H_2O, 55 : 45
Capsules and tablets	UV	241 nm	–	EtOH
Fluoxymesterone and tablets	HPLC	254 nm	Porous silica	Special[a]

[a] Butyl chloride–H_2O saturated butyl chloride–THF–MeOH–HOAc, 475 : 475 : 70 : 35 : 30.

7.2 Assays of Raw Materials and Medicinal Products

As shown in Table 5, various methods are used in the analysis of raw materials and medicinal products containing androgens in the USP.

7.3 Analysis of Androgens in Biological Media

Androgens, such as testosterone, occur largely in conjugated form in urine, and require the liberation of the steroids before analysis. The methods used include solvolysis at an acid pH for cleavage of steroid sulfates,[65] as well as enzymatic methods, involving the use of β-glucuronidase and sulfatase for cleavage of glucuronides and sulfates.[66,67]

Solvents that are partially miscible or immiscible with water, such as ethyl acetate, ether, chloroform, and dichloromethane, have been used to extract ''free'' (unconjugated) steroids from biological matrices. In addition, ion-exchange resins, such as Amberlite XAD-2, have been utilized extensively for extraction and subsequent separation of steroid hormones, both ''free'' and conjugated.[68] Gel filtration has also proven to be extremely valuable.[69]

Since 1980, SPE cartridges have been used widely as adsorptive extractants. After application of the urine sample, most of the interfering substances are washed away with water while the analytes remain adsorbed. Elution with a stronger solvent, such as methanol, removes the analytes. As little as 2 mL of methanol is needed to provide some 98% recovery of steroids with SPE cartridges. However, some workers have reported that the eluates from these contain impurities that interfere with subsequent GC/MS quantification of steroids. Some researchers prefer to use HPLC for final purification.[40,70]

Once the steroids and/or steroid conjugates have been extracted from the biological matrix (whether by liquid or solid extraction), an initial separation into steroid classes (if only unconjugated compounds are to be analyzed) or into ''free'' steroids, glucuronides, mono- and di-sulfates may be required. HPLC, capillary GC, and TLC have been widely used in the separation and quantification of androgens as described below.

Using high-performance TLC plates, a complete separation of androgens and their metabolites with methylene chloride–cyclohexane–acetone (70 : 25 : 5, v/v/v) has been reported.[71] Following derivatization of the steroids while still on the plates by cinnamic aldehyde, 4-dimethylaminobenzaldehyde, and sulfuric acid, quantitative determination was achieved by fluorescence.

Since some steroids are difficult to resolve by TLC, special modifications, such as multiple running (using the same system twice or different systems) or two-dimensional TLC, were tried in order to improve separations.[72–75] Appreciable increases in resolution were achieved, but at the sacrifice of speed.

In GC, the use of capillary columns has resulted in improved resolving power over packed columns and has proved invaluable in analyses of extracts of various biological fluids.

In order to increase volatility and improve heat stability in GC, many substances have been used as derivatizing agents. Among these have been the *O*-methyl oximes for derivatization of ketonic functions and the TMS ethers for hydroxyl functions.[76,77] Others have used the oxime instead of the *O*-methyl oxime, coupled with TMS ethers.[78] For hydroxyl group derivatization, alternatives to TMS ethers are the *t*-butyldimethylsilyl (TBDMS) ethers.[69] Other derivatives available for steroid hydroxyl groups are acetates[79] or trifluoroacetates.[80]

7.4 High-performance Liquid Chromatography

The most important uses of HPLC have generally been the separation and measurement of steroids in tissue preparations and perfusates, or synthesized in chemical reactions, and also for analysis of extracts of steroids endogenous to tissues such as the adrenals and testes. Moreover, HPLC has been valuable in the purification of steroid-containing extracts prior to immunoassay, being particularly useful in situations where high-specificity antisera are not available.

Makin and Heftmann[81] have reported on the application of HPLC to the analysis of C_{21}, C_{19}, and C_{18} steroids. The technique has not been used extensively for the quantification of C_{19} steroids in human body fluids, because the majority of detectors normally used with HPLC cannot measure the low concentrations often present. To overcome the problem of the relative insensitivity of UV detection at 240 nm, Payne[82] used specific bacterial hydroxysteroid dehydrogenases for detection and quantification after HPLC separation of serum steroids.

In attempting to improve resolution of steroids or to reduce the overall elution time for a chromatographic run, while still retaining acceptable resolving power, reports of various modifications to HPLC technology, such as using multistep gradient elution and using microbore columns, have been published.[83,84]

7.5 Mass Spectrometry

Coupled with capillary GC or HPLC, MS has been extensively employed in androgen analysis. The techniques used included MS with selected ion monitoring (SIM), metastable peak monitoring, and MS/MS. Fast atom bombardment/mass spectrometry (FAB/MS), after separation by HPLC, has been used in the quantification of intact conjugates without the need for prior hydrolysis.[85]

GC/MS has been useful in studies of metabolism and biosynthesis of androgens.[86,87] Moreover, GC/MS is proven to be a simple analytical screening procedure for the estimation of total androgen conjugates in post-competition urine inspection.[88–89] Hartmann[90] reported that DLs of 0.02–0.1 g kg^{-1} were reached by GC/MS (EI) in the SIM mode. Cairns[91] described the analysis of testosterone esters by MS/MS.

7.6 Immunoassay of Androgens

In addition to its utility for routine analysis of androgens in clinical laboratories, RIA has also been used in analytical and biosynthetic studies in vitro or in vivo.

Separation of the various forms of testosterone has been accomplished by various methods. Older procedures such as equilibrium dialysis (EQD),[92] centrifugal ultrafiltration,[93] and ammonium sulfate precipitation have been modified to increase reproducibility and permit analysis of large numbers of samples.[94]

Since RIA is by nature a heterogeneous technique, the separation of bound and free analyte is necessary for accurate counting. In addition to protein precipitation, DCC has been used extensively, as well as hydroxylapatite.[95] However, these methods prolong the assays and reduce the possibility of automation. Fiet[96] described a technique called scintillation proximity assay (SPA), which eliminates the separation of bound and free steroid in RIA.

Improvements in EIAs, such as enzyme-linked immunosorbent assay (ELISA), for testosterone and other C_{19} steroids, which give greater sensitivity,[97] improved validation, and reduction of non-specific binding, have been published. These improvements have been achieved by the use of various reporter enzymes, such as alkaline phosphatase, HRP,[98] β-galactosidase, and penicillinase.[99]

For FIAs, the most common labels are coumarin derivatives and fluorescein. The disadvantage of these and many other fluorescent compounds is that they give high background signals. In order to resolve this difficulty, time-resolved fluorometry, using the chelates of the lanthanides europium, samarium, and terbium has been developed.[100,101] This method is capable of providing FIAs with sensitivities of 10^{-15}–10^{-16} mol L^{-1}.

8 PROGESTINS

The main clinical use of assays for progestins is determining whether a woman has probably ovulated in a given cycle and if the luteal phase is adequate for pregnancy to be possible. The determination of synthetic progesterones has became important owing to their increasing uses as contraceptive agents. The quantification of neurosteroids (steroids in brain), of which pregnenolone is one of the most important, has evoked intense interest and has been done by RIA, GC/MS, and HPLC.[102,103]

8.1 Identification Tests for Progestins

For pharmaceutically important progestins, IR and UV at 200–400 nm are the most frequently used identification tests specified by the USP 23. In the IR techniques, the potassium bromide pellet method is used.

In the UV technique, the Δ^4-3-keto group is the characteristic chromophoric group in progestins. The conjugation band is fairly intense (λ_{max} 240 nm, $\varepsilon \sim 1.7 \times 10^4$ L mol^{-1} cm^{-1}). The effect of the second double bond at the Δ^6 position is significant, causing about a 40 nm bathochromic and a considerable hyperchromic shift, e.g. megestrol acetate.

In IR spectra, progestins show bands of medium intensity (~1635–1605 cm^{-1}) due to unsaturation as α,β-unsaturated ketones. Together with the carbonyl absorption (1680 cm^{-1}), this is one of the best ways to identify progestins. More detailed correlations of progestin IR bands are available in the literature.[1]

8.2 Assays of Raw Materials and Medicinal Products

Assays of progestins as raw materials and in medicinal products are summarized in Table 6. UV is not only an important detection mode in HPLC, but also is a useful analytical method for identification purposes although it may require suitable derivatization of the analyte.

8.3 Analysis of Progestins in Biological Media

Generally, analysis of progestins in biological material requires extraction of the analyte from the sample, separation into individual components, and their subsequent determination. Although extraction is time-consuming and expensive, it is often necessary because it can increase the specificity and sensitivity of the final assay system by reducing the matrix effect. In some cases, good correlation and agreement between direct and extraction methods have been demonstrated.[104]

The extraction of the progestins from the biological material is dependent on the matrix involved. In some cases, it is necessary to hydrolyze conjugates, e.g. glucuronides and sulfates, with enzymes before extraction. There is usually a compromise on the polarity of the extraction solvent; the less polar extraction reagents will give the lower extraction efficiencies for progesterone but will also reduce interferences from unwanted polar cross-reactants.

Meng[105] reported a method for combined analysis of conjugated progesterone metabolites and bile acids in serum and urine of pregnant women, which put emphasis on extraction and separation. Total neutral steroids were extracted with ODS-bonded silica. Groups of conjugates were separated on the lipophilic ion-exchange triethylaminohydroxypropyl Sephadex LH-20 (TEAP-LH-20). Sequences of hydrolysis/solvolysis and separations on TEAP-LH-20 permitted separate analyses of steroids, glucuronides, monosulfates, and disulfates. Extraction and solvolysis of steroid conjugates were monitored by FAB/MS.

Clouet[106] described the identification of endogenous 19-nortestosterone in pregnant ewes by GC/MS. SPE and liquid–liquid extraction were used in that method. In addition, G60 purification and HPLC purification were used to increase the sensitivity of the final assay.

The quantitative determination of pregnanolone in rat brain was achieved by RPHPLC with fluorescence

Table 6 Methods used in the assay of progestins and oral contraceptives in USP 23

Compound	Method	Detector	Packing	Mobile phase
Progesterone and injection and sterile suspension	HPLC	254 nm	ODS	H_2O–*i*PrOH, 72 : 28
Hydroxyprogesterone caproate	UV	240 nm	–	EtOH
Caproate injection	UV	380 nm	–	Isoniazid Reagent
Megestrol acetate and tablets	HPLC	280 nm	ODS	ACN–H_2O, 55 : 45
Medroxyprogesterone acetate and acetate tablets	HPLC	254 nm	ODS	ACN–H_2O, 40 : 60
Sterile suspension	HPLC	254 nm	Porous silica microparticles	Butyl chloride, hexane and ACN 700 : 300 : 80
Norethindrone	UV	240 nm	–	EtOH
Tablets	UV	380 nm	–	Isoniazid Reagent
Norethindrone and ethinyl E2 tablets	HPLC	200 nm	ODS	ACN–H_2O, 60 : 40
Norethindrone and mestranol tablets	HPLC	200 nm	C_8	ACN–H_2O, %0 : 50
Norgestrel	UV	241 nm	–	EtOH
Tablets	UV	380 nm	–	Isoniazid Reagent
Levonorgestrel	UV	241 nm	–	EtOH

Table 7 Extraction of progestins

Progestin	Solvent (or method)	Sample	Ref.
Progesterone	Diethyl ether	Human plasma	107
Progesterone	Addition of ethanol (10%, v/v), followed by extraction with light petroleum	Human plasma	108
NETA	Basic pH into diethyl ether–dichloromethane (3 : 2, v/v).	Human plasma	109
19-Nortestosterone	SPE, C_{18} column	Urine of ewe	106
Pregnenolone	SPE, Bond Elut C_8 cartridge	Rat brain	103
NET and six metabolites	Extrelut column, pH 5	Human plasma	110
Conjugated progesterone metabolites	ODS-bonded silica	Serum and urine of pregnant women	105

NET, norethisterone; NETA, norethisterone acetate.

detection using 1-anthroyl cyanide as the derivatization reagent and the internal standard method.[103] The desired fraction was obtained from a rat brain with the combined use of a Bond Elut C_8 cartridge for the SPE and lipophilic gel (piperidino-hydroxypropyl Sephadex LH-20) for the ion-exchange chromatography. The achieved DL was 0.2 ppb (initial volume: 20 mL).

These and other common methods for the extraction of progestins are summarized in Table 7.

8.4 Measurement of Progestins

Presently, determination of progestins is mainly performed by immunoassays and less frequently by other methods such as HPLC, GC, and GC/MS. Competitive immunoassays are the principal methods for progestin measurement. A wide variety of labels are used, including ^{125}I, ^{3}H, enzymes, or fluorescent compounds. Separation of the free and bound analyte use methods based on affinity and methods based on coupling of the antibody to some form of solid phase, such as DCC absorption, as well as ammonium sulfate precipitation, and second antibody precipitation.

Use of nonisotope labels has enabled some progestin assays, such as EIA and FIA, to be simplified and made more sensitive. Currently available automated procedures for plasma progesterones utilize a similar principle with a progesterone-peroxidase conjugate.[111] Table 8 shows some examples of immunoassays reported for the analysis of progestins.

8.5 Progestin Analysis by Gas Chromatography

GC was established as the most sensitive method for the measurement of progestin in the 1960s. At that time, the maximum sensitivity was 0.1–0.2 ng mL^{-1}, although specificity was dependent on the efficient separation. Improvements were subsequently made by

Table 8 Immunoassays for progestin

Method	Compounds	Label	Refs.
EIA	NET in plasma and urine	HRP	112
	Plasma progesterone	HRP	113
	Plasma progesterone	Alkaline phosphatase	114
	PG in urine	Lysozyme	115
FIA	PG in urine	Fluorescence	116
RIA	Progesterone in serum	^{3}H	117
	Plasma progesterone	^{3}H, ^{125}I	118, 119
	NET and norgestrel	^{125}I-TME	120
	Cyproterone	^{3}H or ^{125}I	121
	Medroxyprogesterone acetate (MPA)	^{3}H	122
	NET (NE)	^{3}H	123
	NET (NE)	^{125}I	124
	Norgestrel in plasma	^{125}I	125
	PG in urine	^{3}H	126
	Plasma progesterone	^{125}I-TME	127
	Plasma progesterone	^{125}I-TME	128
	Salivary progesterone	–	129

TME, tyrosine methyl ester.

employing more sensitive detection systems, including argon ionization,[130] flame ionization,[131] and electron capture,[132] which used various derivatization reagents including chloroacetate, chlorodifluoroacetate, and hepta-fluorobutyrate in order to increase the sensitivity of electron capture. Lucis[133] has reviewed several GC procedures for the determination of urinary pregnanediol. The methods involved enzymatic hydrolysis, column chromatography on alumina, and separation and quantitation by GC as free steroids or by formation of derivatives.

Pommier developed an analytical method for the determination of NETA in human plasma by capillary GC/MS with testosterone acetate as the internal standard.[109] The compounds were converted into their pentafluoropropionyl derivatives before determination. The GC/MS used a mass selective detector at m/z 486 for NETA. The detection range was 0.10–10 ng mL^{-1}.

Mathur[102] described a GC/MS method for identification and quantification of brain steroids and their conjugates, which includes pregnenolone, 3β-hydroxy-5α-pregnan-20-one, and 3α-hydroxy-5α-pregnan-20-one. The steroids exist as unconjugated compounds and as sulfates, lipoidal esters, and sulfolipids.

Simultaneous determination of NET and six metabolites in human plasma was achieved by capillary GC/MS.[110] The compounds are determined after enzymatic hydrolysis and conversion to bistrimethylsilyl derivatives. The reproducibility and accuracy of the method were found suitable, over the range 0.5–8 ng mL^{-1} for NET and five of the metabolites.

Table 9 summarizes the progress made in the analysis of progestins by GC.

8.6 Progestin Analysis by High-performance Liquid Chromatography

HPLC is an excellent analytical method, but derivatization may be necessary to increase the response of progestins because they are not very responsive to the commonly used detectors. Compounds with poor detector response can be determined using indirect UV or fluorescence detection. In this technique, a UV- or fluorescence-absorbing species is added to the mobile phase and the detection results from the reduction in response of the probe on elution of the sample. RPHPLC is the main mode used in the indirect detection of both charged and uncharged species. In HPLC, β-cyclodextrins have been used either chemically bonded as a stationary phase or added to the mobile phase. It has been postulated that inclusion complexation between the progestin and β-cyclodextrin leads to improved resolution and shorter separation time. Lamparczyk[58] investigated the influence of mobile-phase composition, concentration of β-cyclodextrin, and temperature in the HPLC separation of norgestrel. Enantiomers were detected by using UV detection at 240 nm. The results showed that chiral recognition is maintained at 20 °C, baseline separation is observed at 0 °C, and the best chiral separation was achieved in the range from −5 °C to 0 °C.

The quantitative determination of pregnenolone in rat brain was achieved by RPHPLC with fluorescence detection using 1-anthroyl cyanide as the derivatization reagent and the internal standard method.[103]

Agnus[136] reported the indirect photodetection of pregnanolone by HPLC on a Cyclobond column using UV detection at 280 nm and testosterone as the internal standard. A UV-absorbing species (testosterone) was added to the mobile phase and detection resulted from the perturbation of the distribution equilibrium of the probe on injection of the sample. This method can be used as an alternative for the detection of poorly UV-absorbing species, but both the sample and the marker must interact with β-cyclodextrin.

Gorog[137] described the role of HPLC/diode-array UV spectroscopy in the identification of minor components

Table 9 GC and GC/MS methods for progestin estimation

Progestin	Method	Ref.
Progesterone in serum	GC/Argon ionization	130
Progesterone in serum	GC/Flame ionization	131
Progesterone in serum	GC/Electron capture	132
Pregnanediol in urine	GC	133
Progesterone in saliva	GC/MS, SIM	134
Medroxyprogesterone acetate	GC/MS	135
NETA in human plasma	GC/MS, SIM	109
Endogenous 19-nortestosterone in pregnant ewes	GC/MS, Total ion monitoring and selective ion monitoring	106
Profiles of conjugated progesterone metabolites and bile acids in serum and urine of pregnant women	GC/MS, Total ion monitoring and selective ion monitoring	105
Pregnenolone, 3β-hydroxy-5α-pregnan-20-one, 3α-hydroxy-5α-pregnan-20-one, and their conjugates	GC/MS, selective ion monitoring	102

Table 10 HPLC methods for progestin analysis

Compound	Column	Solvent (v : v)	Detector	Ref.
Hydroxylated impurities of norgestrel	S1 – 10 silica	Chloroform – acetone	Diode-array UV	137
Norgestrel enantiomers	Knauer ODS-1, Supelco ODS	CH_3CN-H_2O, 25 : 75	UV	58
Pregnenolone	J'sphere ODS-18	CH_3CN-H_2O, 15 : 2	Fluorescence	103
Pregnanolone	Cyclobond column	CH_3OH-H_2O, 65 : 35	Indirect UV	136

(impurities, degradation products, metabolites) in various matrices of progestin drug products. Hydroxylated impurities (degradation products) of norgestrel (6α and β, 10β-hydroxyl derivatives) were identified on the basis of their UV spectra and retention time matching with synthesized impurities.

Table 10 summarizes some representative HPLC methods.

9 CONTRACEPTIVES

Since contraceptives contain only estrogens and progestins, the methods for the combinations are similar to those described under previous topics.

Dorantes[138] reported a modification of the USP dissolution method for the analysis of norethindrone and ethinyl E2 tablets. Jiang[139] reported the HPLC analysis of injectable contraceptive preparations containing NET enanthate and E2 valerate.

The separation of seven ethynyl steroids registered as oral contraceptives by micellar electrokinetic chromatography (MEKC) has been described by Poole.[140]

10 CORTICOSTEROIDS

Because of the serious side effects associated with the medicinal use of corticosteroids as anti-inflammatories, many analogs are available and are official in the USP 23. Assays of corticoid raw materials specified in USP are shown in Table 11.

In the search for corticosteroids with improved therapeutic profiles, a great many compounds have been developed that are closely related structurally and have very similar physical properties. These similarities give rise to unique difficulties in accurately analyzing for the corticosteroids in their medicinal products and, more importantly, in biological matrices where their levels are diagnostic for many pathological processes. For these reasons, corticosteroid assays require effective separation techniques, as well as specific and selective detection methods. Since the concentrations in biological matrices are small, sensitivity is also an important requirement in order to lessen estimation error.

Acceptable separation efficiency can be achieved in many cases by simple extraction or back-extraction with suitable solvents. If this is not sufficient, a chromatographic step may be necessary using TLC, paper chromatography, column chromatography, or HPLC. CE has also proven to be a valuable separation modality.[6,141] The last two methods show the most promise for clinical analyses where speed and the potential for automation are important.

Since most of these steroids have acceptable absorptivities, detection can be done efficiently by UV spectroscopy. In more difficult situations, immunoassays, such as RIA, can improve sensitivity. The use of a mass spectral detector after separation by GC or HPLC yields not only sensitivity but also specificity through the use of SIM.

10.1 High-performance Liquid Chromatography

Normal- and reversed-phase HPLC operating in a wide polarity range have been developed for the rapid screening of natural and synthetic corticosteroids The developments in HPLC analysis of corticosteroids for the determination of clinically important steroids in biological specimens have been reviewed.[142,143] Various sample preparation techniques were also described.

Takeda has developed a chemiluminescence high-performance liquid chromatographic method for the determination of corticosteroids and tetrahydrocorticosteroids.[144] Garg[145] described the simultaneous analysis of prednisone, prednisolone, and their major hydroxylated metabolites in urine by HPLC. A rapid HPLC analysis and stability study of hydrocortisone 17-butyrate in cream preparations is described by Wanwimolruk.[146] Chen[147] reported the stereochemical analysis of betamethasone and dexamethasone by derivatization and HPLC.

Table 11 Methods used in the assay of corticoid raw materials in USP

Compound	Method	Detector (nm)	Packing	Mobile phase or solvent
Cortisone acetate and tablets and sterile suspension	HPLC	254	Porous silica	[a]
Desoxycorticosterone acetate and injection and pellets	Visible	525	–	Blue tetrazolium reagent
Hydrocortisone and acetate and hemisuccinate	HPLC	254	Porous silica	[a]
Sodium phosphate	UV	239	–	Alkaline phosphatase and methylene blue
Sodium succinate	Visible	525	–	Blue tetrazolium reagent
Valerate	HPLC	254	ODS, 30–50 μm	ACN–H_2O, 45 : 55
Betamethasone	HPLC	254 or 240	ODS	ACN–H_2O, 37 : 63
Acetate	HPLC	254	ODS	ACN–H_2O–HOAc, 700 : 800 : 63
Benzoate	HPLC	254	ODS	ACN–H_2O, 60 : 40
Dipropionate	HPLC	254	ODS	ACN–H_2O, 50 : 50
Sodium phosphate	HPLC	254	ODS	MeOH–0.05 M K_2HPO_4, 1 : 1
Valerate	HPLC	254	ODS	ACN–H_2O, 60 : 40
Dexamethasone	HPLC	254	ODS	ACN–H_2O, 7 : 3
acetate and sodium phosphate	HPLC	254	Phenyl silica	0.07 M TEA(pH 5.4)–MeOH, 74 : 26
Fluocinolone acetonide	HPLC	254	ODS	H_2O–ACN–THF, 77 : 13 : 10
Fluocinonide	HPLC	254	ODS	ACN–H_2O, 50 : 50
Flumethasone pivalate	Visible	520	–	Blue tetrazolium reagent
Flunisolide	HPLC	254	Octyl silane	H_2O–ACN, 3 : 2
Prednisolone and acetate	HPLC	254	Porous silica	[a]
Hemisuccinate	UV	243	–	EtOH–$CHCl_3$
Sodium phosphate	UV	241	–	Alkaline phosphatase and methylene blue
Tebutate	HPLC	254	Porous silica	Isooctane–THF–EtOH, 89 : 10 : 1
Prednisone	HPLC	254	ODS	H_2O–THF–MeOH, 688 : 250 : 62
Triamcinolone	HPLC	254	ODS	MeOH–H_2O, 60 : 40
Acetonide	HPLC	254	ODS	ACN–H_2O, 30 : 70
Diacetate	HPLC	254	ODS	5 mM K_2HPO_4–ACN–THF, 62 : 37 : 1
Hexacetonide	HPLC	235	Porous silica	Isopropanol–CH_2Cl_2, 3 : 97
Fludrocortisone acetate	Visible	525	–	Blue tetrazolium reagent

[a] Butyl chloride–H_2O saturated butyl chloride–THF–MeOH–HOAc, 95 : 95 : 14 : 7 : 6.

10.2 High-performance Liquid Chromatography/Mass Spectrometry

Render[148] reported the electrospray mass spectrometric characterization of a nonpolar 18-hydroxy-11-deoxycorticosterone dimer. The dimer was purified by HPLC and further analyzed by electrospray MS.

Fiori[149] reported the identification of corticosteroids as illegal feed additives in milk replacers by LC/MS using CI. SPE (C_{18}), C_8 column separation, and data acquisition (positive ions) in the scan range m/z 200–550 were used to differentiate and identify compounds at levels in the range of 20–100 ppb.

10.3 Gas Chromatography/Mass Spectrometry

GC/MS provides excellent sensitivity as well as improved specificity and a number of reports of its use have been published. Several techniques have been described for cortisol using both radiolabeled and stable-isotope dilution methods.[150,151] The technique can also be used for the determination of cortisone in human plasma.[152,153]

10.4 Immunoassay Combined with High-performance Liquid Chromatography and/or Gas Chromatography

Delahaut[154] reported the quantitative determination of several synthetic corticosteroids by GC/MS after purification by immunoaffinity chromatography. Primary extracts from different matrices (liver, milk, urine, and feces) were first purified on C_{18} cartridges. The CI negative ionization mode was used. The estimated detection and quantification limits were, respectively, 0.25 and 0.5 ppb or lower.

In order to increase selectivity of the assay of corticosteroids and simplify the sample pretreatment procedure, Lovgren[155] evaluated MLC using Tween 20 as surfactant preceding immunoassay. It was shown that the separation was dependent on several parameters, such as temperature, the concentration of Tween 20, pH and ionic strength of the mobile phase, and nature of the stationary phase.

10.5 Micellar Electrokinetic Capillary Chromatography

Wiedmer[156] has discussed the successful combination of MEKC with electrospray ionization mass spectrometry (ESIMS) for the analysis of some corticosteroids. Optimization of the surfactant concentration and the length of the injected micellar plug were necessary for efficient analysis.

Bumgarner[157] enhanced separation of corticosteroids by using mixed micelles of short-chain alkyl surfactants and bile salts in electrokinetic chromatography. It was found that separation patterns are mostly influenced by the number of hydroxyl functional groups on the steroidal backbone of the bile salts, while the type of ionic head group has little effect on the steroid separation. And the addition of the short-chain alkylsulfonate sodium butanesulfonate to the mixture of taurocholate and glycodeoxycholate greatly improved the separation of the 17 corticosteroids and provided a baseline separation of all solutes.

Nishi[141] achieved the separation and determination of lipophilic corticosteroids and benzothiazepin analogs by MEKC using bile salts.

11 CARDIAC STEROIDS

Since the medicinal properties of the *Digitalis* species have been known for centuries, there has always been considerable interest in analytical methods. Formerly, the assays in plants and pharmaceutical preparations utilized colorimetric and fluorometric assays. More modern methods are used for the medicinal agents containing the major alkaloids of the *Digitalis* species, digoxin and digitoxin, which are official in the USP. Analyses of these products are all done in a similar manner, by RPHPLC using ODS columns with 3–5 μm packing and UV detection at 218 nm.

Since there is a small margin between therapeutic and toxic concentrations in body fluids, more sensitive methods such as RIA, which was originally developed for digoxin analysis, and fluorescence polarization assay were required. However, HPLC is superior for the separation of digitalis metabolites, which often cannot be differentiated by immunological methods because of considerable cross-reactivity with the antibody. In order to improve sensitivity, the HPLC method is often combined with RIA and fluorescence polarization immunoassay (FPIA) detection. Embree[158] compared digoxin analyses by HPLC/postcolumn derivatization and FPIA.

11.1 Separation and Purification

Ikeda[159] developed an analytical method for the determination of cardiac glycosides in *Digitalis lanata* leaves by RPTLC. The procedure consisted of extraction of dry leaf powder with 50% methanol and cleanup by Sep-Pak cartridges prior to RPTLC analysis on an octadecylsilyl bonded silica gel plate, using a developing solvent of ACN–methanol–0.5 M NaCl (1 : 1 : 1). They also reported a quantitative HPLC method for analysis of cardiac glycosides in *Digitalis purpurea* leaves.[160]

11.2 Immunoassays

Miles[161] described the determination of salivary digoxin with a dry strip immunoassay. Analysis of salivary digoxin using a rapid dry chemistry, enzyme-labeled immunometric assay (ELIA) was compared with FPIA. There were no significant differences found between ELIA and FPIA determinations of salivary digoxin, serum digoxin, or saliva/serum concentration ratios.

Lovgren[162] described the application of new non-competitive flow injection EIAs for determination of digoxigenin. Antibody fragments reacting with digoxigenin and labeled with alkaline phosphatase, were used to convert 4-methyl umbelliferyl phosphate to a fluorescent product measured downstream.

Oosterkamp[12] reported analysis of digoxin and its metabolites by using FIA as an on-line postcolumn detector coupled to HPLC. In a later study, he further discussed different strategies for using this technique.[9]

11.3 High-performance Liquid Chromatography

Belsner[163] developed a postcolumn derivatization method for the fluorescence detection of cardenolides in RPHPLC by fluorescence derivatization of cardiac glycosides with concentrated acids. The column effluent is blended with concentrated acids in a knitted tube reactor, accelerated by Cu(II) acetate or Co(II) nitrate in H_2SO_4. Cardiac glycoside levels of less than 100 pg glycoside in 1 mL of plasma are detectable.

Wolfender[164] studied cardenolide glycosides in crude plant extracts by LC combined with thermospray and continuous-flow FAB/MS.

11.4 Thin-layer Chromatography

Ponder[165] developed a method for the determination of digoxin and related compounds in digoxin drug substance and tablets by using HPTLC. Separation of the three compounds was accomplished on a C_{18} reversed-phase plate using water–methanol–ethyl acetate (50:48:2) as the mobile phase. The analytes were determined by densitometry using absorbance for digoxin and fluorescence for the related compounds.

11.5 Micellar Electrokinetic Chromatography

Debusschere[166] describes the influence of some parameters, such as, concentration of surfactant, pH, and addition of organic modifiers and urea on the resolution of four cardiac glycosides by MEKC. Efficiency was improved with a shorter analysis time.

12 VITAMIN D

Vitamin D exists in two forms, vitamin D_2 (ergocalciferol), which is found in yeast and formed from ergosterol by irradiation, and vitamin D_3 (cholecalciferol), which is found in fish oils and is also formed in animal tissue by irradiation of 7-dehydrocholesterol. These compounds are very similar structurally, each exhibiting an open B ring due to scission of the bond between the 9 and 10 carbons. They have minimal biological activity and are metabolized to 25-hydroxyvitamin D (25-OHD_3), which is more active, and 1,25-dihydroxyvitamin D (1,25-$(OH)_2D_3$), which is still more active. The concentrations of 25-OHD_3 and 1,25-$(OH)_2D_3$ reflect the nutritional state of the individual and, therefore, their quantification becomes important in the diagnosis of patients with calcium and phosphate deficiency.

12.1 United States Pharmacopeia Analysis

The USP 23 contains two alternative assays for the vitamins D, one chemical and one chromatographic (HPLC). Because of the complicated formulations of most vitamin preparations and the instability of vitamin D to air and light, both require extensive cleanup and separation procedures. In the chemical method, the sample is first saponified in alkaline solution, extracted with ether, and passed successively through two glass chromatographic columns, one containing siliceous earth, and the other Fuller's Earth. A color is developed in the purified extract using antimony trichloride–acetyl chloride reagent. Absorbance is measured at 500 and 550 nm.

In the HPLC method, the sample is saponified and extracted, then passed at high pressure first through a reversed-phase cleanup column containing an octyl stationary phase. Then the fraction containing vitamin D is passed through a normal-phase column containing porous silica microparticles. Detection is by UV absorption at 254 nm.

12.2 Gas Chromatography

GC has been used successfully to analyze for vitamins D_2 and D_3. However, the procedure is not without problems. The first of these concerns the formation of the pyro and isopyro forms in which the B ring is closed between the 9 and 10 positions as a result of the high temperature environment in the chromatograph. The isomers can be separated using packed columns of medium polarity. The second problem concerns the polarity of the compounds and their metabolites, which are increasingly hydroxylated and usually require derivatization in order to assure separation and detectability. Silylation of hydroxyl groups and oximation of keto functions are the common methods. Use of GC/MS with these derivatives for the measurement of plasma concentrations has been described for all the major vitamin D metabolites.[167]

12.3 High-performance Liquid Chromatography

HPLC separations are usually monitored by the use of UV detectors, since all vitamin D metabolites have an absorption maximum of 264 nm with molar absorptivities on the order of 18 000. In order to enhance the sensitivity of detection of peaks emerging from an HPLC column, other forms of detection have been investigated and one, using fluorescence detection, provides significant improvement in sensitivity. Vitamin D and its metabolites are not themselves fluorescent and in order to utilize this methodology some precolumn derivatization is required. A number of methods for 25-OHD_3 using the formation of fluorescent triazoline adducts have been described.[168,169] The use of substituted 1,2,4-triazoline-3,5-dione (TAD) as a reagent for the protection of the vitamin D *cis*-triene structure during chemical synthesis or modification has been known for many years.[170]

Masuda[171] reported a highly sensitive HPLC assay method using an electrochemical detector for assays of vitamin D metabolites. The assay involves extracting lipids from plasma with methylene chloride and methanol, purification on Zorbax SIL column with 5.5% (v/v) isopropanol in hexane, and quantification by HPLC coupled to ECD. A coulometric system, composed of the dual electrode analytical cell and a guard cell, was used for electrochemical detection of the eluting compounds. The DL was approximately 50 pg mL^{-1} for 25$(OH)D_3$ and 24,25$(OH)_2D_3$ in plasma.

12.4 Protein Binding Analysis

The majority of binding assays for vitamin D and its metabolites rely on some form of saturation analysis using plasma binding proteins, intracellular receptor proteins, or antibodies as the binding proteins. Two sensitive RIAs have been described.[172,173] A comprehensive list of these methods is given in a number of reviews.[174,175]

12.5 Mass Spectrometry

HPLC can be directly linked to a mass spectrometer (HPLC/MS), and there have been descriptions of the use of thermospray LC with MS for vitamin D metabolites.[176–178] A method for investigating the conjugates of vitamin D_3 using FAB/MS has been described.[179] Litwiller[180] used FAB/MS to prove the existence of a monoglucuronide of 1,25-$(OH)_2D_3$.

13 BILE ACIDS

Bile acids, e.g. cholic acid, are formed in the liver from cholesterol and, therefore, have structural similarities, although they are more highly hydroxylated. They are conjugated in the liver with nitrogenous bases, e.g. glycine and taurine, to yield the bile salts glycocholate and taurocholate. These may also be sulfated and/or glucuronidated. They are excreted into the intestine to emulsify fats for digestion and then reabsorbed and returned to the liver. This process is called enterohepatic recirculation and occurs ten to twelve times a day. Since these processes are so intimately associated with liver and intestinal function, the clinical analysis of bile acids is important in the diagnosis of certain liver or intestinal disorders. Also their assays have been used for probing the complex pathways of cholesterol metabolism.

13.1 Analysis of Bile Acids in Biological Media

Because any biological sample contains a complicated mixture of acids, salts, and conjugated metabolites, efficient extraction from the sample matrix is essential. Since all forms are ionic, the use of an ion-exchange resin such as XAD-2 has been reported to be suitable.[85] SPE cartridges containing octadecylsilyl silica have been used.[40]

Once extracted from their biological matrix, bile acid mixtures may be fractionated into unconjugated, conjugated, sulfated, and glucuronidated moieties. Alme[181] described a method using gel chromatography with Sephadex LH-20, but Setchell[182] found it was first necessary to subject the extract to a cation exchanger to improve separation. Further separation may be achieved using GC or HPLC. Chaplin[183] reported that bile acids and their conjugated forms also may be separated by anion-exchange chromatography in alkaline media (0.9 M sodium acetate, 0.1 M sodium hydroxide, and 15% v/v ACN) on a CarboPac PA-100 column.

13.2 General Methods

The most important techniques currently used to measure bile-acid concentrations include enzymatic, GC, GC/MS, HPLC, HPLC/MS, and capillary zone electrophoresis (CZE). Scalia has reviewed the major chromatographic techniques and combined detection systems for the determination of bile acids.[184]

13.3 Enzymatic Analysis

An enzymatic TLC method used the enzyme 3α-hydroxysteroid dehydrogenase to determine total bile acids with a free 3α-hydroxy group.[185] Since the enzyme does not work with bile acids which are sulfated, glucuronidated, or esterified at the C-3 position, it may be necessary to include a preliminary hydrolysis procedure. It is also a convenient way to distinguish between free and conjugated analytes. Further work indicated that, in order to enhance resolution in the analysis of human bile, it was essential to extract with an anion-exchange resin prior to HPLC. In a similar manner, Kato described an enzymatic determination of serum 3α-sulfated bile acids concentration with bile acid 3α-sulfate sulfohydrolase.[186]

13.4 Gas Chromatography

Even since the development of efficient HPLC procedures for bile acids, CGC has provided an excellent method for bile acid analysis, especially when combined with MS. However, since bile acids found in biological matrices are hydroxylated and conjugated, preliminary treatment is necessary to hydrolyze them and increase their volatility. Hydrolysis can be accomplished in acid solution or by using sulfatase or glucuronidase enzymes. Derivatization involves silylation, methylation, or reduction with $LiAlH_4$ for hydroxyl groups and oximation for keto groups. Street[187] has reported on capillary column GC methods for the determination of major bile acids at very low concentrations in human serum and urine. However, since urinary bile acids are conjugated to a greater extent than serum bile acids, preliminary treatment of the urine is necessary to hydrolyze and derivatize them to increase their volatility. Owing to its ability to achieve greater specificity, GC/MS is a superior method for urinary bile acids.[105,188]

13.5 High-performance Liquid Chromatography

Despite the excellent resolving power of HPLC, bile acids have been difficult to separate because of their structural

similarity and extensive hydroxylation. In addition, bile acids do not have strong UV chromophores detectable at wavelengths higher than 220 nm and they also do not have strong fluorophores. In order to increase resolution and sensitivity, pre- and postcolumn derivatization methods have been developed. Precolumn derivatization procedures included the use of 4-bromoethyl-7-methoxy coumarin,[189] 1-bromoacetylpyrene,[190] 1-anthroyl nitrite,[191] and dansylhydrazones.[192] Cavrini[193] used 2-bromoacetyl-6-methoxynaphthalene as a prechromatographic fluorescent-labeling reagent for the HPLC analysis of bile acids. Guldutuna[194] accomplished a sensitive and precise separation and measurement of commonly occurring free and conjugated bile acids in serum, liver biopsies, bile, gastric juice, and feces using HPLC with fluorescence labeling. Sakakura[195] developed a method for the simultaneous determination of bile acids in rat bile and serum by RPHPLC with post-column enzymatic reaction and fluorescence detection. Without prior fractionation and alkaline hydrolysis, 26 unconjugated or conjugated bile acids were determined. The limits of detection were 1–5 pmol.

Kurosawa[196] reported an analysis of stereoisomeric C_{27}-bile acids by HPLC with fluorescence detection. The method involved the derivatization of the C_{27}-bile acids into fluorescent esters with 3-(4-bromomethylphenyl)-7-diethylaminocoumarin. The separation of the 16 kinds of bile acids containing stereoisomers was carried out using a reversed-phase Inertsil C_8-column by gradient elution. The DLs were about 15 fmol.

Bile salts are all surface-active agents and, as such, their ability to solubilize fats in the intestine depends on their hydrophilic–hydrophobic balance, more commonly called hydrophilic–lipophilic balance (HLB). Scholmerich[197] attempted to correlate HLB with the octanol–water partition coefficient. This can be done either by direct octanol–water partition experiments or by RPHPLC using an octyl (C_8) column. Partition coefficients appear to be influenced strongly by the number of hydroxyl groups present on the sterol nucleus and only weakly by the orientation of the groups. In contrast, capacity factors are very sensitive to differences in orientation of ring hydroxyls.[198]

13.6 Mass Spectrometry

Fast atom bombardment spectrometry has been coupled with collision-induced fragmentation for the analysis of taurine-coupled bile acids.[199] Unconjugated, glycine-conjugated, sulfated, and glucuronidated bile acids are first coupled with taurine by using a carbodiimide reagent. The taurine-conjugated bile acids give more intense spectra than the original parent compounds. MS coupled to HPLC for the analysis of bile acids in biological fluids also has been reported.[200] Warrack[201] described ion spray liquid chromatographic/mass spectrometric characterization of bile acids.

13.7 Capillary Zone Electrophoresis

Yarabe[202] demonstrated that CZE with indirect photometric detection is a viable approach to the separation and identification of free bile acids along with their taurine and glycine derivatives. Various parameters such as pH, organic solvent concentrations, column temperature, and type of chromophore electrolyte were investigated to optimize the electrophoretic separation and to maximize the peak capacity. The quality of separation of bile acids can be dramatically improved by incorporating γ-cyclodextrin (γ-CD) in the running electrolyte. As a result, a CZE separation of all 15 bile acids was possible in approximately 30 min using 5 mM-adenosine 5′-monophosphate, 7 mM γ-CD in 75% (v/v) methanol at pH 7.0 (apparent).

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
BSA	Bovine Serum Albumin
CE	Capillary Electrophoresis
CGC	Capillary Gas Chromatography
CGC/MS	Capillary Gas Chromatography/ Mass Spectrometry
CI	Chemical Ionization
CZE	Capillary Zone Electrophoresis
DCC	Dextran-coated Charcoal
DL	Detection Limit
E1	Estrone
E1-S	Estrone Sulfate
E2	Estradiol
ECD	Electron Capture Detection
EG	Estrone-3-glucuronide
EI	Electron Impact
EIA	Enzyme Immunoassay
ELIA	Enzyme-labeled Immunometric Assay
ELISA	Enzyme-linked Immunosorbent Assay
EQD	Equilibrium Dialysis
ESIMS	Electrospray Ionization Mass Spectrometry
FAB/MS	Fast Atom Bombardment/ Mass Spectrometry
FAB/SIMS	Fast Atom Bombardment/Secondary Ion Mass Spectrometry
FIA	Fluorescence Immunoassay

FPIA	Fluorescence Polarization Immunoassay
GC	Gas Chromatography
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GC/NPD	Gas Chromatography/Nitrogen–Phosphorus Detection
HDLC	High-density Lipoprotein Cholestrol
HLB	Hydrophilic–Lipophilic Balance
HPLC	High-performance Liquid Chromatography
HPLC/HPLC	Tandem High-performance Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
HRP	Horseradish Peroxidase
IR	Infrared
IRS	Infrared Spectroscopy
LC	Liquid Chromatography
LIF	Laser-induced Fluorescence
MEKC	Micellar Electrokinetic Chromatography
MLC	Micellar Liquid Chromatography
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NET	Norethisterone
NETA	Norethisterone Acetate
NMR	Nuclear Magnetic Resonance
NPD	Nitrogen–Phosphorus Detection
ODS	Octadecyl Silane
PG	Pregnanediol-3-α-glucuronide
RAD	Receptor Affinity Detection
RIA	Radioimmunoassay
RPHPLC	Reversed-phase High-performance Liquid Chromatography
RPTLC	Reversed-phase Thin-layer Chromatography
RSD	Relative Standard Deviation
SDS	Sodium Dodecyl Sulfate
SFC	Supercritical Fluid Chromatography
SIM	Selected Ion Monitoring
SPA	Scintillation Proximity Assay
SPE	Solid-phase Extraction
TAD	1,2,4-Triazoline-3,5-dione
TBDMS	*t*-Butyldimethylsilyl
THF	Tetrahydrofuran
TID	Thermionic Ionization Detection
TLC	Thin-layer Chromatography
TME	Tyrosine Methyl Ester
TMS	Trimethylsilyl
USP	United States Pharmacopeia
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
WCOT	Wall-coated Open Tubular
1,25-$(OH)_2D_3$	1,25-Dihydroxyvitamin D
25-OHD_3	25-Hydroxyvitamin D
γ-CD	γ-Cyclodextrin

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry

Pharmaceuticals and Drugs **(Volume 8)**
Chemical Reagents and Derivatization Procedures in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis • Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Capillary Electrophoresis • Gradient Elution Chromatography • Micellar Electrokinetic Chromatography • Reversed Phase Liquid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. H.L.J. Makin, D.B. Gower, D.N. Kirk, *Steroid Analysis*, 1st edition, Chapman and Hall, London, 1995.
2. S. Gorog, *Steroid Analysis in the Pharmaceutical Industry*, Ellis Horwood, Chichester, 1989.
3. W. Kemp, *Organic Spectroscopy*, 3rd edition, Macmillan, London, 1991.
4. J.K.M. Sanders, B.K. Hunter, *Modern NMR Spectroscopy*, Oxford University Press, Oxford, 1987.

5. D.H. Williams, I. Fleming, *Spectroscopic Methods in Organic Chemistry*, 4th edition, McGraw-Hill, London, 1987.
6. H. Nishi, 'Pharmaceutical Applications of Micelles in Chromatography and Electrophoresis', *J. Chromatogr. A*, **780**, 243–264 (1997).
7. M. Amin, K. Harrington, R. von Wandruszka, 'Determination of Steroids in Urine by Micellar HPLC with Detection by Sensitized Terbium Fluorescence', *Anal. Chem.*, **65**, 2346–2351 (1993).
8. S. Boschi, R. Deiasio, P. Mesini, G.F. Bolelli, R. Sciajno, R. Pasquali, M. Capelli, 'Measurement of Steroid Hormones in Plasma by Isocratic High Performance Liquid Chromatography Coupled to Radioimmunoassay', *Clin. Chim. Acta.*, **231**(1), 107–113 (1994).
9. H. Irth, A.J. Oosterkamp, U.R. Tjaden, J. van der Greef, 'Strategies for On-line Coupling of Immunoassays to High-performance Liquid Chromatography', *Trends Anal. Chem.*, **14**(7), 355–361 (1995).
10. A.J. Oosterkamp, B. Hock, H. Irth, 'Novel Monitoring Strategies for Xenoestrogens', *Trends Anal. Chem.*, **16**(10), 544–553 (1997).
11. A.J. Oosterkamp, H. Irth, M.T. Villaverde Herraiz, U.R. Tjaden, J. van der Greef, 'Theoretical Concepts of On-line Chromatographic – Biochemical Detection Systems', *J. Chromatogr. A.*, **787**, 27 (1997).
12. H. Irth, A.J. Oosterkamp, W. van der Welle, U.R. Tjaden, J. van der Greef, 'On-line Immunochemical Detection in Liquid Chromatography Using Fluorescein-labeled Antibodies', *J. Chromatogr.*, **633**, 65–72 (1993).
13. J. Emneus, G. Marko-Varga, 'Biospecific Detection in Liquid Chromatography', *J. Chromatogr.*, **703**, 191–243 (1995).
14. M. De Frutos, F.E. Regnier, 'Tandem Chromatographic/Immunological Analysis', *Anal. Chem.*, **65**(1), 17A–24A (1993).
15. B.J. Millard, *Quantitative Mass Spectrometry*, Heyden, London, 1978.
16. E.B. Hoving, 'Chromatographic Methods in the Analysis of Cholesterol and Related Lipids', *J. Chromatogr. B: Biomed. Appl.*, **671**(1–2), 341–362 (1995).
17. J.J. Klansek, P. Yancey, R.W. St. Clair, R.T. Fischer, W.J. Johnson, J.M. Glick, 'Cholesterol Quantitation by GLC: Artifactual Formation of Short-chain Steryl Esters', *J. Lipid Res.*, **36**(10), 2261–2266 (1995).
18. M. Tanaka, S. Kanamaru, 'Capillary Gas Chromatography Quantification of Cholesterol in Copper-oxidized Low-density Lipoprotein', *Biol. Pharm. Bull.*, **16**(6), 538–543 (1993).
19. M. Amelio, R. Rizzo, F. Varazini, 'Determination of Sterols, Erythrodiol, Uvaol and Alkanols in Olive Oils Using Combined Solid-phase Extraction, High-performance Liquid Chromatographic and High-resolution Gas Chromatographic Techniques', *J. Chromatogr.*, **606**(2), 179–185 (1992).
20. M. Rodriguez-Palmero, S. de la Presa-Owens, A.I. Castellote-Bargallo, M.C. Lopez Sabater, M. Rivero-Urgell, M.C. de la Torre-Boronat, 'Determination of Sterol Content in Different Food Samples by Capillary Gas Chromatography', *J. Chromatogr. A.*, **672**(1–2), 267–272 (1994).
21. J.A. Garcia Regueiro, C. Maraschiello, 'Procedure for the Determination of Eight Cholesterol Oxides in Poultry Meat Using On-column and Solvent Venting Capillary Gas Chromatography', *J. Chromatogr. A.*, **764**(2), 279–293 (1997).
22. M. Okazaki, K. Sasamoto, T. Muramatsu, S. Hosaki, 'Evaluation of Precipitation and Direct Methods for HDL-cholesterol Assay by HPLC', *Clin. Chem.*, **43**(10), 1885–1890 (1997).
23. T. Rezanka, 'Analysis of Sterol Esters from Alga and Yeast by High-performance Liquid Chromatography and Capillary Gas Chromatography–Mass Spectrometry with Chemical Ionization', *J. Chromatogr.*, **598**(2), 219–226 (1992).
24. R.P. Ager, R.W.A. Oliver, 'Separation of Oestrogen Conjugates in Biological Fluids and Synthetic Mixtures on Sephadex G-type Gels', *J. Chromatogr.*, **271**, 325–340 (1983).
25. P.I. Musey, D.C. Collins, J.R.K. Preedy, 'Isocratic Separation of Estrogen Conjugates on DEAE-Sephadex', *Steroids*, **29**, 657–688 (1977).
26. W. Eechaute, G. Demeeste, 'Fluorimetric Determination of Total Estrogens after Gel-filtration of Enzymically Hydrolysed Urine', *J. Clin. Endocrinol. Metab.*, **25**, 480–487 (1965).
27. N.S. Jiang, R.J. Ryan, A. Albert, 'Radioimmunoassay of Serum Estrogen', *Clin. Chem.*, **19**, 740–747 (1973).
28. R.A. Hawkins, 'Studies on the Oestrogens in Peripheral Plasma of the Non-pregnant Human', PhD. thesis, University of Leeds, 1974.
29. P. Moutsatsou, 'Non-protein Bound Oestriol in Human Pregnancy Plasma', PhD. thesis, Columbia University, 1986.
30. H. Katagari, F.Z. Stanczyk, U. Goebelsmann, 'Estriol in Pregnancy', *Steroids*, **24**, 225–238 (1974).
31. I. Yoshizama, J. Fishman, 'Radioimmunoassay of 2-Hydroxy Estrone in Human Plasma', *J. Clin. Endocrinol. Metabol.*, **32**, 3–6 (1971).
32. A.E. Bolton, F.J. Rutherford, 'Evidence for the Presence of 6-Keto Oestrodiol-17 in Human Plasma – implications for Oestrodiol-17 Radioimmunoassays', *J. Steroid Biochem.*, **7**, 71 (1976).
33. A.C. Speight, 'Biologically Available Oestrogens in Plasma and Urine', PhD. thesis, University of Leeds, 1978.
34. B.K. Young, H. Jirku, M. Levits, 'Estriol Conjugates in Amniotic Fluid at Mid-pregnancy and Term', *J. Clin. Endocrinol. Metabol.*, **35**, 208–218 (1972).
35. S.L. Cohen, E. Oran, 'The Conjugated Estrogens', *Can. J. Biochem.*, **50**, 1245–1248 (1972).

36. C.P.W. Tsang, 'Changes in Plasma Levels of Estrone Sulfate and Estrone in the Pregnant Ewe around Parturition', *Steroids*, **23**, 855–868 (1974).
37. A. Remy-Martin, O. Prost, M. Nicollier, J. Purnod, G.L. Adessi, 'Estrone Sulfate Concentrations in Plasma of Normal Individuals, Postmenopausal Women with Breast Cancer and Men with Cirrhosis', *Clin. Chem.*, **29**, 86–89 (1984).
38. C.S. Kammer, J.W. Goldzieher, 'Extraction of Steroid Conjugates from Aqueous Solutions', *Anal. Biochem.*, **28**, 492–502 (1969).
39. F. Andreolini, C. Borra, F. Caccamo, A. Di Corcia, I. Nicoletti, R. Samperi, F. Improta, 'Estriol and its Conjugates in Late Pregnancy Determined by Extraction with Carbopack B and Liquid Chromatography with Fluorometric Detection', *Clin. Chem.*, **31**, 1698–1702 (1985).
40. C.H.L. Shackleton, J.O. Whitney, 'Use of Sep-Pak* Cartridges for Urinary Steroid Extraction: Evaluation of the Method for use Prior to Gas Chromatographic Analysis', *Clin. Chim. Acta*, **107**, 231–243 (1980).
41. M. Hirako, H. Takahashi, I. Domeki, 'Establishment of a Reliable Extraction Method of Estrone Sulfate from Bovine Plasma for Detection of the Peripheral Level During the Regular Estrous Cycle by Radioimmunoassay', *Domestic Animal Endocrinology*, **13**(3), 187–195 (1996).
42. A. Martinetti, E. Seregni, E. Bajetta, G.F. Bolelli, L. Ferrari, S. Massaron, C. Botti, E. Bombardieri, 'Development of a Rapid and Ultrasensitive RIA Method for Estrogen (E2, E1, E1-S) Determination with Selective Solid-phase Extraction', *Intern. J. Biol. Markers*, **12**(3), 102–105 (1997).
43. S. Ikegawa, M. Itoh, N. Murao, H. Kijima, M. Suzuki, T. Fujiyama, J. Goto, 'Immunoaffinity Extraction for Liquid Chromatographic Determination of Equilin and its Metabolites in Plasma', *Biomed. Chromatogr.*, **10**(2), 73–77 (1996).
44. J. Novakovic, V. Pacakova, J. Sevcik, T. Cserhati, 'Quantitative Structure–Chromatographic Retention Relationship Study of Six Underivatized Equine Estrogens', *J. Chromatogr. B: Biomed. Appl.*, **681**(1), 115–123 (1996).
45. G. Barnard, Y. Amir-Zaltsman, S. Lichter, B. Gayer, F. Kohen, 'The Measurement of Oestrone-3-glucuronide in Urine by Non-competitive Idiometric Assay', *J. Steroid Biochem. Mol. Biol.*, **55**(1), 107–114 (1995).
46. S. De Lauzon, B. Desfosses, N. Christeff, C. Hanquez, N.T. Cittanova, 'A Direct Dot-enzyme Immunoassay to Detect Human Ovulation', *J. Steroid Biochem. Mol. Biol.*, **42**, 223–228 (1992).
47. K.M. Henderson, M. Camberis, A.H. Hardie, 'Evaluation of Antibody- and Antigen-coated Enzyme Immunoassays for Measuring Oestrone-3-glucuronide Concentrations in Urine', *Clin. Chim. Acta*, **243**(2), 191–203 (1995).
48. G. Barnard, F. Kohen, 'Monitoring Ovarian Function by the Simultaneous Time-resolved Fluorescence Immunoassay of Two Urinary Steroid Metabolites', *Clin. Chem.*, **44**(7), 1520–1528 (1998).
49. G. Barnard, F. Kohen, H. Mikola, T. Lovgren, 'Measurement of Estrone-3-glucuronide in Urine by Rapid Homogeneous Time-resolved Fluorimmunoassay', *Clin. Chem.*, **35**, 555–559 (1989).
50. G.N. Ranadive, J.S. Mistry, K. Damodaran, M.J. Khosravi, A. Diamandi, T. Gimpel, V.D. Castracane, S. Patel, F.Z. Stanczyk, 'Rapid, Convenient Radioimmunoassay of Estrone Sulfate', *Clin. Chem.*, **44**(2), 244–249 (1998).
51. P.E. Lonning, D. Ekse, 'A Sensitive Assay for Measurement of Plasma Estrone Sulfate in Patients on Treatment with Aromatase Inhibitors', *J. Steroid Biochem. Mol. Biol.*, **55**(3–4), 409–412 (1995).
52. K. Tamate, M. Charleton, J.P. Gosling, D. Egan, M. Ishikawa, P.F. Fottrell, M.M. Kane, 'Direct Colorimetric Monoclonal Antibody Enzyme Immunoassay for Estradiol-17 beta in Saliva', *Clin. Chem.*, **43**(7), 1159–1164 (1997).
53. T. Ohtaki, M. Moriyoshi, K. Nakada, T. Nakao, K. Kawata, 'Radioimmunoassay of Saliva Estrone Sulfate in Pregnant Sows', *J. Veterinary Med. Sci.*, **59**(9), 759–763 (1997).
54. R. Bagnati, M. Grazia Castilli, L. Airoldi, 'Analysis of Diethylstilbestrol, Dienestrol and Hexestrol in Biological Samples by Immunoaffinity Extraction and Gas Chromatography–Negative-ion Chemical Ionization Mass Spectrometry', *J. Chromatogr.*, **527**(2), 267 (1990).
55. F. Novak, E. Van Beek, J. Lambert, 'Pregnenolone and Estradiol Identification in the Brine Shrimp, *Artemia* sp., by means of Gas Chromatographic–Mass Spectrometric Analysis', *Comp. Biochem. Physiol. B*, **95**(3), 565 (1990).
56. G. Casademont, B. Perez, J.A. Garcia Regueiro, 'Simultaneous Determination, in Calf Urine, of Twelve Anabolic Agents as Heptafluorobutyryl Derivatives by Capillary Gas Chromatography–Mass Spectrometry', *J. Chromatogr. B: Biomed Appl.*, **686**(2), 189–198 (1996).
57. A.J. Oosterkamp, W. van der Welle, H. Irth, U.R. Tjaden, J. van der Greef, 'Reversed-phase Liquid Chromatography Coupled On-line to Receptor Affinity Detection Based on the Human Estrogen Receptor', *Anal. Chem.*, **68**, 1201–1206 (1996).
58. H. Lamparczyk, P.K. Zarzycki, J. Nowakowska, 'Effect of Temperature on Separation of Norgestrel Enantiomers by High-performance Liquid Chromatography', *J. Chromatogr. A*, **668**, 413–417 (1994).
59. B.J. Spencer, W.C. Purdy, 'HPLC Separation of Equilin, Estrone, and Estrone Derivatives with Cyclodextrins as Mobile Phase Additives', *J. Liquid Chromatogr.*, **18**(20), 4063–4080 (1995).
60. M. Katayama, H. Taniguchi, 'Determination of Estrogens in Plasma by High-performance Liquid Chromatography after Pre-column Derivatization with

2-(4-carboxyl-phenyl)-5,6-dimethylbenzimidazole', *J. Chromatogr.*, **616**, 317–322 (1993).
61. S.Y. Su, G.K. Shiu, 'High-performance Liquid Chromatographic Analysis of Six Conjugated and Unconjugated Estrogens in Serum', *Biomed. Chromatogr.: BMC*, **6**(6), 265 (1992).
62. T. Iwata, T. Hirose, M. Yamaguchi, 'Direct Determination of Estriol 3- and 16-glucuronides in Pregnancy Urine by Column-switching High-performance Liquid Chromatography with Fluorescence Detection', *J. Chromatogr. B: Biomed. Sci. Appl.*, **695**(2), 201–207 (1997).
63. T. Yasui, H. Uemura, N. Yoneda, M. Irahara, T. Aono, S. Sunahara, K. Hata, 'Assessment of Serum Estrogen levels by RIA with HPLC during Hormone Replacement Therapy', *Nippon Sanka Fujinka Gakkai Zasshi – Acta Obstetrica et Gynaecologica Japonica*, **47**(7), 655–661 (1995).
64. N. Nagaya, M. Toi, S. Sunahara, F. Kurimoto, T. Tominaga, 'Determination of Intratumoral Estrone (E1) and Estradiol (E2) in Primary Breast Cancer Tissues by Sensitive HPLC-RIA', *Ganto Kagaku Ryoho [Japanese Journal of Cancer and Chemotherapy]*, **24**(3), 329–336 (1997).
65. S. Burstein, S. Lieberman, 'Hydrolysis of Ketosteroid Hydrogen Sulfates by Solvolysis Procedures', *J. Biol. Chem.*, **233**, 331–335 (1958).
66. L.D. Dikkeschei, B.G. Wothers, P.H.B. Willemse, 'The Determination of Delta-5-androstenediol and its Sulfate in Serum and Urine by Gas Chromatography–Mass Spectrometry', *Clin. Endocrinol.*, **39**(4), 469–475 (1993).
67. L.D. Dikkeschei, P.H.B. Willemse, B.G. Wothers, 'Delta-5-androstenediol and its Sulfate in Serum and Urine of Normal Adults and Patients with Endocrine Diseases', *Clin. Endocrinol.*, **39**(4), 475 (1993).
68. R. Heikenen, T. Fotsis, H. Adlercreutz, 'Analytical Methods: Physicochemical Assay', *J. Steroid Biochem.*, **19**, 175–180 (1983).
69. S.H. Andersson, J. Sjovall, 'Analysis of Profiles of Unconjugated Steroids in Rat Testicular Tissue by Gas Chromatography–Mass Spectrometry', *J. Steroid Biochem.*, **23**, 469–475 (1985).
70. N.A. Schmidt, H.J. Borburgh, T.J. Penders, 'Steroid Profiling – An Update', *Clin. Chem.*, **31**, 637–639 (1985).
71. S.W. Golf, V. Graef, J.T. Schiller, 'Thin-layer Chromatography – the Forgotten Alternative for the Quantitation Determination of Steroids', *Biochem. Chromatogr.*, **2**, 189–192 (1987).
72. I. Louveau, M. Bonneau, D.B. Gower, 'Biosynthesis of 16-Androstene Steroids and Testosterone by Porcine Testis Tissue In Vitro: Effect of Age and Relationships with Fat 5-Alpha-androstenone Levels In Vivo', *Acta Endocrinologica*, **125**(5), 526 (1991).
73. G.M. Cooke, 'Phospholipases Modulate Pig Testicular Androgen and 16-Androstene Biosynthetic Pathways In Vitro', *J. Steroid Biochem. Mol. Biol.*, **41**, 99–107 (1992).
74. G.M. Cooke, 'Identification of Phospholipids Capable of Modulating the Activities of some Enzymes Involved in Androgen and 16-Androstene Biosynthesis in the Immature Pig Testis', *J. Steroid Biochem. Mol. Biol.*, **42**, 151–159 (1992).
75. C.H.L. Shackleton, *Analysis of Steroids*, Marcel Dekker, New York and Basel, 153–243, 1985.
76. T.K. Kwan, C.H. Poh, R. Perumal, 'Intracellular Distribution of Pregnenolone Metabolites in Testis of Macaque', *Biochem. Internat.*, **7**(5), 885–894 (1988).
77. T.K. Kwan, D.J. Trafford, H.L. Makin, A.I. Mallet, D.B. Gower, 'GC/MS studies of 16-Androstenes and other C_{19} Steroids in Human Semen', *J. Steroid Biochem. Mol. Biol.*, **43**(6), 549–556 (1992).
78. W.G.E.J. Schoonen, J.G.D. Lambert, 'Gas Chromatographic–Mass Spectrometric Analysis of Steroids and Steroid Glucuronides in the Seminal Vesicle Fluid of the African Catfish, *Clarias gariepinus*', *General Comparative Endocrinol.*, **68**, 375–386 (1987).
79. P.R. Dow, J.C. Javis, G.J. Gow, 'Metabolism of Testosterone by Human Semen', *J. Steroid Biochem.*, **22**, 371–376 (1985).
80. T. Yamaguchi, K. Konnon, K. Kawabe, 'Simultaneous Determination of Five Androgens in Early Neonatal Urine by Selected Ion Monitoring GC/MS Using $19d_3$ Steroids as Internal Standards', *Endocrinologica Japonica*, **32**, 279–285 (1985).
81. H.L.J. Makin, E. Heftmann, *High-performance Liquid Chromatography of Steroid Hormones*, Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, 183–224, 1988.
82. D.W. Payne, D. Holtzclaw, E.Y. Adashi, 'A Convenient Unified Scheme for the Differential Extraction of Conjugated and Unconjugated Serum C_{19} Steroids on Sep-Pak C_{18}-cartridges', *J. Steroid Biochem.*, **33**, 289–295 (1985).
83. M.W. Capp, M.H. Simonian, 'Separation of the Major Adrenal Steroids by Reversed-phase High-performance Liquid Chromatography', *Anal. Biochem.*, **147**, 374–381 (1985).
84. R.B. Taylor, K.E. Kendel, R.G. Reid, 'Chromatography of Progesterone and its Major Metabolites in Rat Plasma Using Microbore HPLC Columns with Conventional Injection and Detection Systems,' *J. Chromatogr.*, **385**, 383–392 (1987).
85. S.J. Gaskell, 'Quantification of Steroid Conjugates Using Fast Atom Bombardment Mass Spectrometry', *Steroids*, **55**, 458–462 (1990).
86. T.K. Kwan, M.A. Kraevskaya, H.L. Makin, D.J. Trafford, D.B. Gower, 'Use of Gas Chromatographic–Mass Spectrometric Techniques in Studies of Androst-16-ene and Androgen Biosynthesis in Human Testis; Cytosolic Specific Binding of 5-Alpha-androst-16-en-3-one', *J. Steroid Biochem. Mol. Biol.*, **60**(1–2), 137–146 (1997).

87. A.I. Mallet, K.T. Holland, P.J. Rennie, W.J. Watkins, D.B. Gower, 'Applications of Gas Chromatography–Mass Spectrometry in the Study of Androgen and Odorous 16-Androstene Metabolism by Human Axillary Bacteria', *J. Chromatogr.*, **562**(1–2), 647–658 (1991).
88. L. Dehennin, P. Lafarge, P. Dailly, D. Bailloux, J.P. Lafarge, 'Combined Profile of Androgen Glucurono- and Sulfo- Conjugates in Post-competition Urine of Sportsmen: A Simple Screening Procedure Using Gas Chromatography–Mass Spectrometry', *J. Chromatogr. B: Biomed. Appl.*, **687**(1), 85–91 (1996).
89. M. Becchi, R. Aguilera, P. James, 'Gas Chromatography/Combustion/Isotope-ratio Mass Spectrometry Analysis of Urinary Steroids to Detect Misuse of Testosterone in Sport', *Rapid Commun. Mass Spectrom.*, **8**(4), 304 (1994).
90. S. Hartmann, H. Steinhart, 'Simultaneous Determination of Anabolic and Catabolic Steroid Hormones in Meat by Gas Chromatography–Mass Spectrometry', *J. Chromatogr. B, Biomed. Sci. Appl.*, **704**(1–2), 105–117 (1997).
91. T. Cairns, E.G. Siegmund, B. Rader, 'Analysis of Testosterone Esters by Tandem Mass Spectrometry', *J. AOAC Int.*, **76**(2), 306 (1993).
92. M.J. Wheeler, 'The Determination of Bio-available Testosterone', *Ann. Clin. Biochem.*, **32**(4), 345–357 (1995).
93. M.G. Forest, M.A. Rivarola, C.J. Migeon, 'Percentage Bindings of Testosterone, Androstenedione and Dehydroepiandrosterone in Human Plasma', *Steroids*, **12**, 323–343 (1968).
94. A. Barrini, I. Liberale, E. Menini, 'Simultaneous Determination of Free Testosterone and Testosterone Bound to Non-sex Hormone Binding Globulin by Equilibrium Dialysis', *Clin. Chem.*, **39**, 936–941 (1993).
95. D.J.H. Trafford, H.L.J. Making, 'Use of Hydroxylapatite in Immunoassay', *Methods Enzymol.*, **70**, 291–298 (1980).
96. J. Fiet, Ph. Boudou, J.M. Barthier, 'Application of Scintillation Proximity Assay to Homogeneous Radioimmunoassay of Androstenedione, Dehydroepiandrosterone and 11-Hydroxyandrostenedione', *Clin. Chem.*, **37**, 293 (1991).
97. F.H. Pruslin, R. Winston, T.C. Rodman, 'Caveats and Suggestions for the ELISA', *J. Immunolog. Methods*, **137**(1), 27 (1991).
98. E. Ali, J. Sengupta, T.K. Dhar, 'Sandwich Immunoassay of Small Molecules. I. Investigation with Testosterone as Model Hapten', *J. Immunolog. Methods*, **147**(2), 173 (1992).
99. P.N. Rao, I.B. Taraporewala, 'A Sensitive Enzyme-linked Immunosorbent Assay for Testosterone: Use of a Novel Heterologous Hapten Conjugated to Penicillinase', *Steroids*, **57**, 154–161 (1992).
100. T.M. Jackson, R.P. Ekins, 'Theoretical Limitations on Immunoassay Sensitivity', *J. Immunolog. Methods*, **87**, 13–20 (1987).
101. E. Soini, H. Kojola, 'Time-resolved Fluorescence of Lanthanide Probes and Applications in Biotechnology', *CRC Crit. Rev. Anal. Chem.*, **18**, 105–154 (1987).
102. C. Mathur, V.V.K. Prasad, V.S. Raju, M. Welch, S. Lieberman, 'Steroids and their Conjugates in the Mammalian Brain", *Proc. Natl. Acad. Sci. USA*, **90**, 85–88 (1993).
103. K. Shimada, T. Nakagi, 'Studies on Neurosteroids. IV. Quantitative Determination of Pregnenolone in Rat Brain using HPLC', *J. Liquid Chromatogr. Rel. Technol.*, **19**(16), 2593–2603 (1996).
104. S.P. Haynes, J.M. Corcoran, C.J. Eastan, F.A. Doy, 'Radioimmunoassay of Progesterone in Unextracted Serum', *Clin. Chem.*, **26**, 1607–1609 (1992).
105. L.J. Meng, J. Sjovall, 'Method of Combined Analysis of Profiles of Conjugated Progesterone Metabolites and Bile Acids in Serum and Urine of Pregnant Women', *J. Chromatogr. B, Biomed. Appl.*, **688**(1), 11 (1997).
106. A. Clouet, B. Lebizec, M.P. Montrade, F. Monteau, F. Andre, 'Identification of Endogenous 19-Nortestosterone in Pregnant Ewes by GC-MS', *Analyst.*, **122**, 471–474 (1997).
107. I.M. Thorneycroft, S.C. Stone, 'Radioimmunoassay of Serum Progesterone in Women Receiving Oral Contraceptive Steroids', *Contraception*, **5**, 129–146 (1972).
108. K.K. Dighe, W.M. Hunter, 'A Solid Phase Radioimmunoassay for Plasma Progesterone', *Biochem. J.*, **143**, 219–231 (1974).
109. F. Pommier, A. Sioufi, J. Godbillon, 'Quantitative Determination of Norethisterone Acetate in Human Plasma by Capillary Gas Chromatography with Mass-selective Detection', *J. Chromatogr. A*, **750**, 75–81 (1996).
110. F. Pommier, A. Sioufi, J. Godbillon, 'Simultaneous Determination of Norethisterone and Six Metabolites in Human Plasma by Capillary Gas Chromatography with Mass-selective Detection', *J. Chromatogr. B*, **674**, 155–165 (1995).
111. M.A. Pesce, S.H. Boudarian, M. Sheehan, C.F. Henkel, 'An Automated Biotin-Streptavidin Procedure for Progesterone Assay', *Clin. Biochem.*, **25**, 451–455 (1992).
112. A. Turkes, G.F. Read, D. Riad-Fahmy, 'A Simple High-throughput Enzyme Immunoassay for Norethisterone (Norethindrone)', *Contraception*, **25**, 505–514 (1982).
113. H.X. Wang, R.A. Scott, G.H.G. Thorpe, 'An Enhanced Chemiluminescence Enzyme Immunoassay for Serum Progesterone', *Steroids*, **44**, 317–328 (1985).
114. G.H.G. Thorpe, I. Bronstein, L.J. Kricka, 'Chemiluminescent Enzyme Immunoassay of Alpha-fetoprotein Based on an Adamantyl Dioxetane Phenyl Phosphate Substrate', *Clin. Chem.*, **35**, 2319–2321 (1989).
115. S.J. Thornton, R.J. Pepperell, J.B. Brown, 'Home Monitoring of Gonadotropin Ovulation Induction Using

the Ovarian Monitor', *Fertility Sterility*, **54**, 1076–1082 (1990).
116. J.S. Kesner, E.A. Knecht, E.F. Krieg, 'Time-resolved Immunofluorometric Assays for Urinary Luteinizing Hormone and Follicle Stimulating Hormone', *Anal. Chim. Acta*, **285**, 13 (1994).
117. G.E. Abraham, R. Swerdloff, D. Tulchinsky, W.D. Odell, 'Radioimmunoassay of Plasma Progesterone', *J. Clin. Endocrinol. Metabolism*, **32**, 619–624 (1971).
118. E.H.D. Cameron, J.J. Scarisbrick, 'Radioimmunoassay of Plasma Progesterone', *Clin. Chem.*, **19**, 1403–1408 (1973).
119. E.H.D. Cameron, J.J. Scarisbrick, S.E. Morris, G. Read, '^{125}I-Iodohistamine Derivatives as Tracers for the Radioimmunoassay for Progestagens', in *Fifth Tenovus Workshop; Steroid Immunoassay*, Alpha Omega Publishing Co., Cardiff, 153–164, 1975.
120. R.J. Warren, K. Fotherby, 'Radioimmunoassay of Synthetic Progestagens, Norethisterone and Norgestrel', *J Endocrinol.*, **62**, 605–618 (1974).
121. B. Nieuweboer, K. Lubke, 'Radioimmunological Determination of Cyproterone Acetate', *Hormone Res.*, **8**, 210–218 (1977).
122. K.T. Kriton, J.C. Cornette, 'Return of Ovulatory Cyclicity Following an Intramuscular Injection of Medroxyprogesterone Acetate', *Contraception*, **10**, 39–45 (1974).
123. C. Walls, C.W. Vose, C.E. Horth, R.F. Palmer, 'Radioimmunoassay of Norethisterone after Ethynodiol Diacetate Administration', *J. Steroid Biochem.*, **8**, 167–171 (1977).
124. S.G. Hillier, G.F. Read, 'Radioimmunoassay for Plasma Norethisterone and Testosterone Using Antisera raised against C-11 Conjugated Haptens and Radio-iodinated Ligands', *J. Endocrinol.*, **67**, 5 (1975).
125. P.F. Brenner, D.R. Mishell, F.Z. Stanczyk, U. Goebelsmann, 'Serum Levels of D-Norgestrel, Luteinizing Hormone, Follicle-stimulating Hormone, Estradiol and Progesterone in Women during and following Injection of Combinational Oral Contraceptives Containing D, L-Norgestrel', *Am. J. Obstet. Gynecol.*, **129**, 133–140 (1977).
126. W.P. Collins, P.O. Collins, M.J. Kilpatrick, et al., 'The Concentrations of Oestrone-3-glucuronide, LH and Pregnanediol-3-glucuronide as Indices of Ovarian Function', *Acta Endocrinol.*, **93**, 336–348 (1979).
127. R. Stupnicki, T. Kowalczyk-Preewloka, E. Kula, L.Y. Kasongo, 'HME – a Novel Derivative for Direct Iodination in Steroid Radioimmunoassay', *J. Steroid Biochem.*, **28**, 663–667 (1987).
128. J.E.T. Corrie, W.A. Ratcliffe, J.S. Macpherson, 'The Provision of Labeled Traces for Radioimmunoassay of Haptens: A General Approach?', *J. Immunol. Methods*, **51**, 159–166 (1982).
129. J.E.T. Corrie, 'A Direct Radioimmunoassay for Salivary Progesterone Using a Radiolabeled Tracer', in *Ninth Tenovus Workshop; Immunoassay of Steroids in Saliva*, Alpha Omega Publishing Co., Cardiff, 127–133, 1984.
130. W.P. Collins, I.F. Sommerville, 'Quantitative Determination of Progesterone in Human Plasma by Thin-layer and Gas–Liquid Chromatography', *Nature*, **203**, 836–839 (1964).
131. H.J. Van der Molen, 'Gas Phase Chromatography of Progesterone and Related Steroids', in *Gas Phase Chromatography of Steroids*, Springer Verlag, Heidelberg, 150–235, 1967.
132. K.H. Raisinghani, R.I. Dorfman, E. Forchielli, 'A Gas Chromatographic Method for the Estimation of Nanogram Quantities of Progesterone Using Electron Capture Detection', Abstracts of Meeting on Gas Chromatographic Determination of Steroid Hormones, Rome, 1966.
133. O.J. Lucis, 'Progesterone and its Metabolites', in *Biochemistry of Women: Methods of Clinical Investigation*, CRC Press, Cleveland, OH, 45–78, 1974.
134. H.M. Leith, P.L. Truran, S.J. Gaskell, 'Quantification of Progesterone in Human Saliva', *Biomed. Environmen. Mass Spec.*, **13**, 257–261 (1986).
135. G. Phillipou, R.G. Firth, 'Specific Quantification of Plasma Medroxyprogesterone Acetate by GC/MS', *Clin. Chem. Acta*, **103**, 129–133 (1980).
136. B. Agnus, N.-M. Gosselet, B. Sebille, 'Indirect Photodetection of Pregnanolone on a Cyclobond Column by High-performance Liquid Chromatography', *J. Chromatogr. A*, **663**, 27–33 (1994).
137. S. Gorog, M. Bihari, E. Csizer, F. Dravetz, M. Gazdag, B. Herenyi, 'Estimation of Impurity Profiles of Drugs and Related Materials. Part 14: The Role of HPLC/diode-array UV Spectroscopy in the Identification of Minor Components (Impurities, Degradation Products, Metabolites) in Various Matrices', *J. Pharm. Biomed. Anal.*, **14**(1–2), 85–92 (1995).
138. A. Dorantes, S. Stavachansky, 'Modification of the USP Dissolution Method for the Analysis of Norethindrone and Ethinyl Tablets', *J. Pharm. Sci.*, **83**(3), 379 (1994).
139. L.X. Jiang, Z.J. Wang, S.A. Matlin, 'HPLC Analysis of Injectable Contraceptive Preparation Containing Norethisterone Enanthate and Estradiol Valerate', *J. Liquid Chromatogr.*, **13**(17), 3473 (1990).
140. S.K. Poole, C.F. Poole, 'Separation of Pharmaceutically Important Estrogens by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **749**, 247–255 (1996).
141. H. Nishi, T. Fukuyama, M. Matsuo, S. Terabe, 'Separation and Determination of Lipophilic Corticosteroids and Benzothiazepin Analogs by Micellar Electrokinetic Chromatography Using Bile Salts', *J. Chromatogr.*, **513**, 279–295 (1990).
142. L. Valvo, A. Paris, A.L. Savella, B. Gallinella, S.E. Ciranni, 'General High-performance Liquid Chromatographic Procedures for the Rapid Screening of Natural

and Synthetic Corticosteroids', *J. Pharm. Biomed. Anal.*, **12**(6), 805–810 (1994).
143. P. Volin, 'High-performance Liquid Chromatographic Analysis of Corticosteroids', *J. Chromatogr. B: Biomed. Appl.*, **671**(1–2), 319–340 (1995).
144. M. Takeda, M. Maeda, A. Tsuji, 'Chemiluminescence High-performance Liquid Chromatography of Corticosteroids Using Lucigenin as Post-column Reagent', *Biomed. Chromatogr.*, **4**(3), 119–122 (1990).
145. V. Garg, W.J. Jusko, 'Simultaneous Analysis of Prednisone, Prednisolone and their Major Hydroxylated Metabolites in Urine by High-performance Liquid Chromatography', *J. Chromatogr.*, **561**(1), 39 (1991).
146. S. Wanwimolruk, 'Rapid HPLC Analysis and Stability Study of Hydrocortisone 17-Butyrate in Cream Preparations', *Pharm. Res.*, **8**(4), 547 (1991).
147. S.H. Chen, S.M. Wu, H.L. Wu, 'Stereochemical Analysis of Betamethasone and Dexamethasone by Derivatization and High-performance Liquid Chromatography', *J. Chromatogr.*, **595**(1), 203 (1992).
148. N.B. Render, G.W. Taylor, G.P. Vinson, 'Electrospray Mass Spectrometric Characterization of a Corticosteroid Dimer', *Rapid Commun. Mass Spectrom.*, **11**(2), 219–223 (1997).
149. M. Fiori, E. Pierdominici, F. Longo, G. Brambilla, 'Identification of Main Corticosteroids as Illegal Feed Additives in Milk Replacers by Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **807**(2), 219–227 (1998).
150. N.V. Esteben, A.L. Vergey, D.J. Liberato, 'Stable Isotope Dilution Method Using Thermospray Liquid Chromatography/Mass Spectrometry for Quantification of Daily Cortisol Production in Human', *Biochem. Environment. Mass Spectrom.*, **15**, 603–608 (1988).
151. D.G. Patterson, M.B. Patterson, P.H. Culbreth, 'Determination of Steroid Hormones in a Human-serum Reference Material by Isotope Dilution–Mass Spectrometry: A Candidate Method for Cortisol', *Clin. Chem.*, **30**, 619–626 (1984).
152. H. Shibasaki, T. Furuta, Y. Kasuya, 'Quantification of Corticosteroids in Human Plasma by Liquid Chromatography–Thermospray Mass Spectrometry Using Stable Isotope', *J. Chromatogr. B*, **692**, 7 (1997).
153. H. Shibasaki, T. Furuta, Y. Kasuya, 'Stable Isotope Dilution Mass Spectrometry for the Simultaneous Determination of Cortisol, Cortisone, Prednisolone and Prednisone in Plasma', *J. Chromatogr.*, **579**(2), 193 (1992).
154. P. Delahaut, P. Jacquemin, Y. Colemonts, M. Dubois, J. De Graeve, H. Deluyker, 'Quantitative Determination of Several Synthetic Corticosteroids by Gas Chromatography–Mass Spectrometry after Purification by Immunoaffinity Chromatography', *J. Chromatogr. B: Biomed. Sci. Appl.*, **696**(2), 203–215 (1997).
155. U. Lovgren, M. Johansson, K. Kronkvist, L.E. Edholm, 'Biocompatible Sample Pretreatment for Immunochemical Techniques Using Micellar Liquid Chromatography for Separation of Corticosteroids', *J. Chromatogr. B: Biomed. Appl.*, **672**(1), 33–44 (1995).
156. S.K. Wiedmer, M. Jussila, M.L. Riekkola, 'On-line Partial Filling Micellar Electrokinetic Capillary Chromatography–Electrospray Ionization/Mass Spectrometry of Corticosteroids', *Electrophoresis*, **19**(10), 1711–1718 (1998).
157. J.G. Bumgarner, M.G. Khaledi, 'Mixed Micelles of Short Chain Alkyl Surfactants and Bile Salts in Electrokinetic Chromatography: Enhanced Separation of Corticosteroids', *J. Chromatogr. A.*, **738**(2), 275–283 (1996).
158. L. Embree, K.M. McErlane, 'Comparison of Digoxin Analysis by High-performance Liquid Chromatography/Post-column Derivatization and Fluorescence Polarization Immunoassay', *Xenobiotica*, **20**(6), 635 (1990).
159. Y. Ikeda, Y. Fuji, M. Umemura, T. Hatakeyama, M. Morita, M. Yamazaki, 'Quantitative Determination of Cardiac Glycosides in *Digitalis lanata* Leaves by Reversed-phase Thin-layer Chromatography', *J. Chromatogr. A*, **746**(2), 255–260 (1996).
160. Y. Ikeda, Y. Fuji, I. Nakaya, M. Yamazaki, 'Quantitative HPLC Analysis of Cardiac Glycosides in *Digitalis purpurea* Leaves', *J. Nat. Prod.*, **58**(6), 897–901 (1995).
161. M.V. Miles, J.R. Miranda-Massari, R.E Dupuis, M.R. Mill, A.L. Zaritsky, M. Nocera, S.T. Lawless, 'Determination of Salivary Digoxin with a Dry Strip Immunometric Assay', *Therapeutic Drug Monitoring*, **14**(3), 249–254 (1992).
162. U. Lovgren, K. Kronkvist, B. Backstrom, L.E. Edholm, G. Johansson, 'Design of Non-competitive Flow Injection Enzyme Immunoassays for Determination of Haptens: Application to Digoxigenin', *J. Immunolog. Meth.*, **208**(2), 159–168 (1997).
163. K. Belsner, B. Buchele, 'Fluorescence Detection of Cardenolides in Reversed-phase High-performance Liquid Chromatography after Post-column Derivatization', *J. Chromatogr. B: Biomed. Appl.*, **682**(1), 95–107 (1996).
164. J.L. Wolfender, K. Hostettmann, F. Abe, T. Nagao, H. Okabe, T. Yamauchi, 'Liquid Chromatography Combined with Thermospray and Continuous-flow Fast Atom Bombardment Mass Spectrometry of Glycosides in Crude Plant Extracts', *J. Chromatogr. A*, **712**(1), 155–168 (1995).
165. G.W. Ponder, J.T. Stewart, 'High-performance Thin-layer Chromatographic Determination of Digoxin and Related Compounds, Digoxigenin Bisdigitoxoside and Gitoxin, in Digoxin Drug Substance and Tablets', *J. Chromatogr. A*, **659**(1), 177–183 (1994).
166. L. Debusschere, C. Demesmay, J.L. Rocca, G. Lachatre, H. Lofti, 'Separation of Cardiac Glycosides by Micellar Electrokinetic Chromatography and Microemulsion Electrokinetic Chromatography', *J. Chromatogr. A*, **779**(1–2), 227–233 (1997).

167. R.D. Coldwell, D.J.H. Trafford, M.J. Varley, 'Measurement of 25-Hydroxyvitamin D2, 25-Hydroxyvitamin D3, 24,25-Dihydroxyvitamin D2 and 25,26-Dihydroxyvitamin D2 in a Single Plasma Sample by Mass Fragmentography', *Clin. Chim. Acta*, **180**(2), 157 (1989).
168. M. Shimizu, T. Yamazaki, S. Yamada, 'Fluorescence-labeling Reaction of Vitamin D Metabolites and Analogs with Fluorescent 1,2,4,-Tirazoline-3, 5,-dione', *Bioorganic Medicinal Chem. Lett.*, **3**, 1808–1814 (1993).
169. S.R. Wilson, M.L. Tulchinsky, Y.H. Wu, 'Electrospray Ionization Mass Spectrometry of Vitamin D Derivatives', *Bioorganic Medicinal Chem. Lett.*, **3**, 1805–1808 (1993).
170. D.J. Aberhart, A.C.-T. Hsu, 'Studies on the Adduct of 4-Phenyl-1,2,4,-triazoline-3,5-dione with Vitamin D3', *J. Org. Chem.*, **41**, 2098–2102 (1976).
171. S. Masuda, T. Okano, M. Kamao, Y. Kanedai, T. Kobayashi, 'A Novel High-performance Liquid Chromatographic Assay for Vitamin D Metabolites Using a Coulometric Electrochemical Detector', *J. Pharm. Biomed. Anal.*, **15**(9–10), 1497–1502 (1997).
172. R. Bouillon, E. van Herck, I. Jans, et al., 'Two Direct Assays for 25-Hydroxyvitamin D', *Clin. Chem.*, **30**, 1731–1736 (1984).
173. L. Hummer, C. Christiansen, 'A Sensitive and Selective Radioimmunoassay for Serum 24,25-Dihydroxycholecalciferol in Man', *Clin. Endocrinol.*, **21**, 71–79 (1984).
174. C. Porteous, R.D. Coldwell, D.J.H. Trafford, 'Recent Development in the Measurement of Vitamin D and its Metabolites in Human Body Fluids', *J. Steroid Biochem.*, **28**, 785–801 (1987).
175. G. Jones, D.J.H. Trafford, H.L.J. Makin, 'Vitamin D: Cholecalciferol, Ergocalciferol and Hydroxylated Metabolites', in *Modern Chromatographic Analysis of Vitamins*, 2nd edition, Marcel Dekker, New York, 73–151, 1992.
176. K. Takamura, H. Hoshino, N. Harima, 'Identification of Vitamin D2 by Thermospray-interface Mass Spectrometry', *J. Chromatogr.*, **543**(1), 241–243 (1991).
177. D. Watson, K.D. Setchell, R. Ross, 'Analysis of Vitamin D and its Metabolites Using Thermospray Liquid Chromatography/Mass Spectrometry', *Biomed. Chromatogr.*, **5**(4), 153–160 (1991).
178. D. Vicchio, A. Yergey, K. O'Brien, 'Quantification and Kinetics of 25-Hydroxyvitamin D3 by Isotope Dilution Liquid Chromatography/Mass Spectrometry', *Biol. Mass Spectrom.*, **22**(1), 53–58 (1993).
179. I. Jardine, G.F. Scanlan, V.R. Mattox, et al., 'Analysis of Steroid and Vitamin D Glucuronides and Sulfates by Fast Atom Bombardment Mass Spectrometry', *Biomed. Mass Spectrom.*, **11**, 4–9 (1984).
180. R.D. Litwiller, V.R. Mattox, I. Jardine, et al., 'Evidence of a Monoglucuronide of 1,25-Dihydroxyvitamin D_3 in Rat Bile', *J. Biol. Chem.*, **257**, 7491–7494 (1982).
181. B. Alme, A. Bremmelgaard, J. Sjovall, et al., 'Analysis of Metabolic Profiles of Bile Acids in Urine Using a Lipophilic Anion Exchanger and Computerized Gas–Liquid Chromatography/Mass Spectrometry', *J. Lipid Res.*, **18**, 339–362 (1977).
182. K.D.R. Setchell, A.M. Lawson, N. Tanida, et al., 'General Methods for the Analysis of Metabolic Profiles of Bile Acids and Related Compounds in Feces', *J. Lipid Res.*, **24**, 1085–1100 (1983).
183. M.F. Chaplin, 'Analysis of Bile Acids and their Conjugates Using High-pH Anion-exchange Chromatography with Pulsed Amperometric Detection', *J. Chromatogr. B: Biomed. Appl.*, **664**(2), 431–434 (1995).
184. S. Scalia, 'Bile Acid Separation', *J. Chromatogr. B: Biomed. Appl.*, **671**(1–2), 299–317 (1995).
185. T. Iwata, K. Yamasaki, 'Enzymatic Determination and Thin-layer Chromatography of Bile Acids in Blood', *J. Biochem.*, **56**, 394–424 (1964).
186. T. Kato, M. Yoneda, K. Nakamura, I. Makino, 'Enzymatic Determination of Serum 3-Alpha-sulfated Bile Acids Concentration with Bile Acid 3-Alpha-sulfate Sulfohydrolase', *Digestive Dis. Sci.*, **41**(8), 1564–1570 (1996).
187. J.M. Street, D.J.H. Trafford, H.L.J. Makin, 'Capillary Gas-chromatography of Glycine-conjugated Bile Acids without Prior Hydrolysis', *J. Lipid Res.*, **27**, 208–224 (1986).
188. T. Murai, R. Mahara, T. Kurosawa, A. Kimura, M. Tohma, 'Determination of Fetal Bile Acids in Biological Fluids from Neonates by Gas Chromatography–Negative Ion Chemical Ionization Mass Spectrometry', *J. Chromatogr. B: Biomed. Sci. Appl.*, **691**(1), 13–22 (1997).
189. S. Okuyama, D. Uemura, J.H. Hirata, 'The Improved Method of High Performance Liquid Chromatographic Separation of Individual Bile Acids', *Chem. Lett.*, 461–462 (1979).
190. S. Kamada, M. Marda, A. Tsuji, et al., 'Separation and Determination of Bile Acids by HPLC Using Immobilized 3-Alpha-hydroxysteroid Dehydrogenase and an Electrochemical Detector', *J. Chromatogr.*, **239**, 773–783 (1983).
191. J. Goto, M. Saito, T. Chikai, et al., 'Determination of Serum Bile Acids by HPLC with Fluorescent Labeling', *J. Chromatogr.*, **276**, 289–300 (1983).
192. A.D. Reid, P.R. Baker, 'HPLC with Reversed-phase Radial Compression Column of Serum Bile Acid Hydrazone Derivatives', *Biochem. Soc. Trans.*, **13**, 1235–1236 (1985).
193. V. Cavrini, R. Gatti, A. Roda, C. Cerre, P. Roveri, 'HPLC-fluorescence Determination of Bile Acids in Pharmaceuticals and Bile After Derivatization with 2-Bromoacetyl-6-methoxynaphthalene', *J. Pharm. Biomed. Anal.*, **11**(8), 761–770 (1993).
194. S. Guldutuna, T. You, W. Kurts, U. Leuschner, 'High-performance Liquid Chromatographic Determination of Free and Conjugated Bile Acids in Serum, Liver Biopsies, Bile, Gastric Juice and Feces by Fluorescence Labeling', *Clin. Chim. Acta*, **214**(2), 195–207 (1993).

195. H. Sakakura, M. Suzuki, N. Kimura, H. Takeda, S. Nagata, M. Maeda, 'Simultaneous Determination of Bile Acids in Rat Bile and Serum by High-performance Liquid Chromatography', *J. Chromatogr.*, **621**(2), 123–131 (1993).
196. T. Kurosawa, H. Sato, M. Sato, H. Takechi, M. Machida, M. Tohma, 'Analysis of Stereoisomeric C_{27}-Bile Acids by HPLC with Fluorescence Detection', *J. Pharm. Biomed. Anal.*, **15**(9–10), 1375–1382 (1997).
197. J. Scholmerich, M.S. Becher, K. Schmidt, 'Influence of Hydroxylation and Conjugation of Bile Salts on their Membrane-damaging Properties – Studies on Isolated Hepatocytes and Lipid Membrane Vesicles', *Hepatology*, **4**, 661 (1984).
198. S. Lam, A. Karmen, 'Bile Acids,' in *HPLC of Small Molecules. A Practical Approach*, ed. C.K. Lim, IRI Press, Oxford, 103–116, 1982.
199. J. Zhang, W.J. Griffiths, T. Bergman, 'Derivatization of Bile Acids with Taurine for Analysis by Fast Atom Bombardment Mass Spectrometry with Collision-induced Fragmentation', *J. Lipid Res.*, **34**(11), 1895 (1993).
200. A. Roda, A.M. Gioacchini, C. Cerre, M. Baraldini, 'High-performance Liquid Chromatographic–Electrospray Mass Spectrometric Analysis of Bile Acids in Biological Fluids', *J. Chromatogr. B: Biomed. Appl.*, **665**(2), 281–294 (1995).
201. B.M. Warrack, G.C. DiDonato, 'Ion Spray Liquid Chromatographic/Mass Spectrometric Characterization of Bile Acids.', *Biol Mass Spectrom.*, **22**(2), 101–111 (1993).
202. H.H. Yarabe, S.A. Shamsi, I.M. Warner, 'Capillary Zone Electrophoresis of Bile Acids with Indirect Photometric Detection', *Anal. Chem.*, **70**(7), 1412–1418 (1998).

Vibrational Spectroscopy in Drug Discovery, Development and Production

Sterling A. Tomellini
University of New Hampshire, Durham, USA
James W. Finn
Genetics Institute, Andover, USA

1 Introduction 1
2 Instrumental Considerations 3
3 Sampling 4
4 Microsampling and Microscopy 7
5 Applications 8
5.1 Mid-infrared 8
5.2 Near-infrared Spectroscopy 14
5.3 Fourier Transform Raman Spectroscopy 18
6 Conclusion 19
Acknowledgments 20
Abbreviations and Acronyms 20
Related Articles 20
References 21

The area of vibrational spectroscopy encompasses a number of versatile analytical techniques, many of which have been applied successfully to a wide variety of pharmaceutical analyses. While there are many vibrational spectroscopic techniques, some more esoteric or generally applicable than others, this chapter will concentrate on pharmaceutical applications limited to three specific areas: near-infrared (NIR) spectroscopy, mid-infrared (MIR) spectroscopy and Fourier transform Raman spectroscopy (FTRS). The goal of this work is to provide the interested reader with the background necessary to better understand these vibrational spectroscopic techniques and provide an overview of the breadth of applications for these techniques in analyses of pharmaceutical importance. Recently reported applications will be emphasized. The wide variety of sampling methods associated with these techniques will be highlighted throughout. Recent developments which are expected to have an impact on future pharmaceutical applications will also be discussed. Literature and other resources for advanced study are also provided.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The many variations and permutations of infrared (IR) and Raman spectroscopies are collectively known as ''vibrational spectroscopy''.[1,2] The unifying feature underlying all of the vibrational spectroscopies is that the effects being measured are related to molecular motion (i.e. stretching or bending of atoms in a molecule, lattice vibrations, etc.). Vibrational spectroscopy, in one form or another, has been exploited to provide both quantitative and qualitative information about pharmaceutically important samples for decades, despite the sometimes complex nature of the spectral data obtained. The purpose of this chapter is to provide the interested reader with a broad overview of the subject area, the background necessary to understand the application of vibrational spectroscopy to samples of pharmaceutical importance, to review representative recent applications and provide a broad array of corresponding references to facilitate more in-depth inquiry.

The first experience which most students in one of the chemically related sciences have with vibrational spectroscopy occurs in an introductory organic chemistry laboratory. The student's first IR spectrum is generally acquired for a solid or liquid sample either synthesized in the laboratory or provided as an ''unknown'' by the instructor. The amount of light absorbed (or transmitted) by the sample in the MIR region of the spectrum is recorded using a suitable instrument after generally limited sample preparation. The student's knowledge of IR spectroscopy is often limited to having learned in class that the wavelengths at which the compound absorbs IR radiation are indicative of the chemical functionalities in the sample. Once a spectrum is acquired the student is directed to one of the many IR wavelength versus functionality correlation charts found in textbooks. Depending on the number and type of functionalities in the sample, the complex spectral pattern obtained can prove to be overwhelming to the student. Using both positive information (the sample absorbs light in a region indicative of a given functionality) and negative information (the sample does not absorb in the region indicative of another functionality), the student completes his/her first experience in vibrational spectroscopy by listing those functionalities which are probably present in the sample and those functionalities which are probably absent. Another common approach taken by students to determine the likely identity of the sample, or at least to narrow down the possibilities, is to manually compare the spectral data obtained with the spectra contained in a collection of spectra, either on paper or computer readable format. Thus, the concepts of spectral interpretation and spectral matching are also

usually part of the student's introduction to vibrational spectroscopy. The training for most nonchemistry and many chemistry students at both the undergraduate and graduate levels will not include in-depth discussions of other types of vibrational spectroscopy (i.e. NIR, Raman), more complex sample and data-handling techniques or general discussions of other than routine applications of vibrational spectroscopy to structure identification.

The unifying characteristic of all of the types of spectroscopy that come under the general heading of vibrational spectroscopy is that they measure the effect that electromagnetic radiation has on the molecular vibrations of the sample. The reader is cautioned, however, that one of the confusing aspects of vibrational spectroscopy is that different units, i.e. either wavenumbers (cm^{-1}) or wavelength (μm, nm), are used by practitioners of the various methods. The most common units in current use are presented here for comparison.

The energies required to induce transitions between vibrational levels range from the low-energy photons of the far infrared (FIR) (10–200 cm^{-1} or 1000–50 μm), to the moderate-energy photons of the MIR region (200–4000 cm^{-1} or 50–2.5 μm), to the higher-energy photons of the NIR region (14 300–4000 cm^{-1} or 0.7–2.5 μm or 700–2500 nm). It is worth noting at this point that most MIR instruments have an upper limit which extends beyond 4000 cm^{-1} but have a lower limit of 400 cm^{-1}. Low-energy FIR photons are used to probe the low-energy transitions of the sample, for example the lattice vibrations of a polymer. MIR photons are sufficiently energetic to promote changes from the ground vibrational energy level to the first excited state energy levels for the fundamental vibrations of the molecule, and it is for this reason that MIR spectroscopy is useful for functional group analysis. The more energetic NIR photons promote transitions from the ground state to higher than the first vibrational excited state (i.e. overtones). These photons are also capable of interacting with higher vibrational energy states of the molecule, which are combinations of the fundamental frequencies.

While the difference in energy between vibrational states is important for determining which wavelengths and in which region of the spectrum an absorption will occur, the magnitude of the light absorbed is directly related to the probability of a photon being absorbed by the molecule. If the probability is high that a given photon will be absorbed, then the absorbance measured will be high. For fundamental reasons, the probability of an NIR photon being absorbed by a molecule is far less than the probability of an MIR photon being absorbed. For this reason, to obtain a spectrum of a sample in the MIR region requires fewer molecules to interact with the light than in the NIR region. The molar absorptivity (ε) of a compound is a measure of the probability of a photon being absorbed. Molar absorptivities are generally lower in the NIR than in MIR region of the spectrum. This physical observation results in both advantages and disadvantages for these particular forms of spectroscopy which will be discussed in more detail using appropriate applications.

MIR and NIR spectroscopic techniques expose the sample to photons having a range of energies, some of which correspond exactly to the energy level transitions in the molecules of the sample. When the energy of the photon matches an energy level difference of the molecule, the photon may be absorbed, while photons having energies which do not match exactly an energy level of the molecule will not be absorbed. Thus, the energy levels of the molecule can be viewed as being probed directly when performing MIR and NIR absorption spectroscopy. The many forms of Raman spectroscopy are unique among the vibrational spectroscopic techniques in that they are not based on the absorption of a photon, but instead on the inelastic scattering of a photon by the molecules of the sample. The excitation photons used in Raman methods are not absorbed by the molecules of the sample and do not exactly match an energy level difference of the sample molecules. Instead Raman spectra are acquired by measuring the differences in energy between that of the monochromatic excitation photons and the photons scattered by the sample, this difference termed the "Raman shift" and is usually plotted in cm^{-1}.

The simplest way to get a feeling for Raman scattering is to invoke the presence of a "virtual" state. Unlike the energy levels of a molecule which are real, a virtual state is not a real energy state of the molecule, which means that the molecule will not exist in that state for any appreciable time. It is easiest to think of the virtual state as the energy through which the molecule passes in its transition between its initial interaction with the excitation photon and the energy that it will have at the point at which a photon is scattered. Figure 1 presents a simplified vibrational energy level diagram for a molecule having a

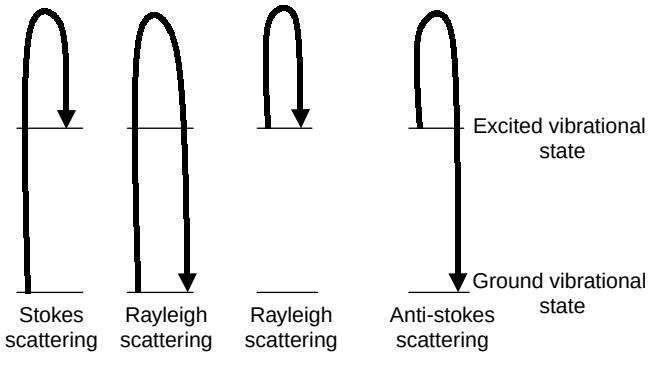

Figure 1 Schematic diagrams indicating energy level transitions for Rayleigh scattering, Stokes Raman scattering and anti-Stokes Raman scattering.

ground state and an excited state as well as a schematic description of the scattering processes. For this simplified example, at the time when the molecule is exposed to the incident photon, it can be in either the ground or excited vibrational energy level. If the molecule interacts with the incident photon and returns to the same energy level as it started, then the scattered photon will be of the same energy as the incident photon. This phenomenon is an ''elastic'' scattering process known as Rayleigh scattering. Raman scattering occurs when the initial and final energy states of the molecule are different after interacting with the incident photon. Most of the molecules of the sample will initially be in the ground vibrational energy level. Interaction with the incident photon can result in the molecule being promoted to a vibrationally excited state. The Raman scattered photon which results from this interaction will have less energy than the excitation photon, with the difference being the energy between the ground and excited states for the molecule. If the molecule in our example is initially vibrationally excited and goes to the ground vibrational state after interaction with the incident photon, then the Raman scattered photon will be of higher energy than the incident photon and thus be Raman shifted to a shorter wavelength. Since the energy difference between the excited and ground states determines the difference in energy between the incident and scattered photon, the Raman shift observed is the same for both the Stokes (Raman shifted to longer wavelength) and anti-Stokes (Raman shifted to shorter wavelength) bands. Stokes bands will, however, be more intense since there is a higher probability of the sample molecules being in the ground state when they interact with the incident photon. It should be emphasized that the overall likelihood of an incident photon being Raman scattered is very small. The vast majority of scattered photons will be Rayleigh not Raman scattered. Thus, it is critical in developing Raman instrumentation to separate the Raman scattered photons from the Rayleigh scattered photons and, in the design of the experiment, to prevent other sources of sample emission, such as fluorescence, from obscuring the weak Raman signal.

Although Raman and IR spectral data are similar in many aspects, there are many theoretical and practical differences between Raman spectroscopy and MIR (or NIR) spectroscopy. For example, the so-called selection rules, which indicate if a given transition will be active, and therefore is likely to be observed, or inactive, and therefore not likely to be observed, are different for the Raman scattering and IR absorption processes. Similarly, water, which strongly absorbs in both the MIR and NIR, is a weak Raman scatterer. Thus, water is not considered a good solvent for obtaining routine MIR spectra of dissolved analytes but is a very good solvent for obtaining Raman spectra of dissolved species. Glass, which has strong bands in the MIR region and therefore, not used for IR sampling, is a weak Raman scatterer, which is one of the reasons glass capillary tubes are often used for Raman sample holders. It is also experimentally easier to obtain Raman spectroscopic data for low-energy transitions (less than 400 cm^{-1}) than by IR spectroscopic techniques.

One of the experimental variables in Raman spectroscopy is the frequency of the incident photons. The intensity of Raman scattering increases in proportion to ν^4, which is the reason that conventional Raman spectra are acquired using visible (VIS) and ultraviolet (UV) laser sources.[3] Unfortunately, UV and VIS photons also are capable of promoting fluorescence either by the analyte or by impurities in the sample. Fluorescence is especially problematic even for the analysis of nonfluorescent analytes in real-world samples since the presence of small amounts of fluorescent impurities can provide sufficient background to obscure the photons due to the weaker Raman scattering. One approach to overcoming fluorescence induced by the incident radiation is to use lower frequency, NIR photons, which are not sufficiently energetic to promote fluorescence for most commonly encountered elements and compounds. The all important decrease in Raman scattering when using NIR incident photons is complicated further by the need to use different detectors than are used for detecting VIS and UV photons. The ability to largely overcome inherent problems with sample fluorescence experienced by conventional Raman spectroscopic techniques have justified the design and development, and has resulted in the commercial availability, of Fourier transform (FT) Raman instrumentation incorporating NIR neodymium : yttrium aluminum garnet (Nd : YAG) lasers. Owing to the higher laser powers generally utilized in FTRS and the reduced intensity of Raman scattering, it should be noted that significant advantages may be gained by using conventional, dispersive Raman spectroscopy to analyze pharmaceutical samples which do not fluoresce.

2 INSTRUMENTAL CONSIDERATIONS

Two distinct types of instruments are used to acquire spectral data in the MIR and NIR regions of the spectrum.[3–5] To understand the difference, it is important to remember that the detectors for these regions of the spectrum are not capable of simultaneously distinguishing between different wavelengths of light. Thus one approach to acquire spectra is to separate the light into discrete wavelengths and allow only photons with a single, known wavelength to reach the detector at any time. This type of instrument often contains a grating to disperse the light, which is why they are classified as ''dispersive'' instruments. It

is important to note that for this type of instrument, at any given moment, most of the light from the source is not reaching the detector, is not providing information, and is, therefore, effectively wasted. Many instruments, especially those used in the NIR for process measurements, utilize optical filters, instead of gratings, to select the wavelength reaching the detector at any given time, but the same principles apply.

An alternative approach can be used for obtaining intensity versus wavelength data using a detector, which is not capable of distinguishing individual wavelengths. In this approach, all wavelengths are allowed to reach the detector at the same time; however, each wavelength in the spectral range is uniquely encoded. Thus, although the detector measures the combined intensity for all of the encoded wavelengths at a given time, suitable decoding provides information on the intensity at any given wavelength. This encoding–measuring–decoding approach is employed in the FT instruments used to obtain MIR, NIR and FT Raman spectra. The most common approach to encoding the IR light is the use of a Michaelson interferometer containing a moving mirror. The light exiting from the interferometer is encoded uniquely at any given point in the mirror's travel. An entire IR spectrum is acquired with each complete sweep of the mirror.

Several advantages which FT instruments demonstrate over dispersive instruments are often listed.[3,6] The most common are the stray light advantage, the throughput advantage, and the multiplex advantage. The stray light advantage arises since only light encoded or modulated by the interferometer provides information. Stray light which is not modulated but reaches the detector does not interfere with the measurement. The throughput advantage, which practically may be more significant for an FT Raman experiment than a Fourier transform infrared (FTIR) experiment, simply stated is that more light from the source impinges on the detector at any given time compared with using a dispersive spectrometer. Clearly, at any given time in a dispersive spectrometer, only a small amount of the source radiation is allowed to reach the detector with most of the source's radiation providing no information. The multiplex advantage is simply that all wavelengths of light are measured simultaneously when using either an FTIR or FT Raman instrument. Another advantage gained when using these types of instruments is that there is an internal laser standard which ensures the wavelength accuracy of the spectral data obtained. It should be noted that the actual importance of these and other so-called advantages for the user's particular application must be individually assessed.

Many of the advances in vibrational spectroscopic techniques and the types of applications that are now possible are due to the vast computational capabilities, which are provided with all FT instruments and available for virtually all other vibrational spectroscopic instrumentation. Advances in computational resources (speed, memory, data storage, and display) have influenced greatly the development and expanded the areas of potential application for vibrational spectroscopic techniques over the past 20 years. In many cases, advanced spectral data-processing techniques have been used to overcome inherent instrumental limitations and difficulties in sampling, compensate for experimental variability and reduce the complexity of the spectral data obtained.

As probably expected, there is generally a high degree of similarity between most vibrational spectroscopic instruments of a given type, even among vendors. Examples of instrument layouts and components along with general descriptions of laboratory and process instrumentation can be found in most undergraduate instrumental methods of analysis texts,[3,4] vibrational spectroscopy monographs,[2,6] and other texts on analytical instrumentation.[5,7] Rapid developments, especially advances in specific components such as optical detectors, can be incorporated very quickly in commercially available instrumentation. The value of information provided on instrument vendor websites, especially concerning current instrument design and specifications should not be underestimated.[8]

Advances in instrumentation and computational capabilities suggest re-evaluations may be prudent for analyses deemed too difficult to perform only a few years ago.

3 SAMPLING

One of the advantages of vibrational spectroscopy is the wide variety of sampling options available.[2–4,6] Spectroscopic data can be acquired for almost any type of sample, often with little or no sample preparation. Minimizing sample preparation is often of paramount importance, not only to reduce analysis time but also to reduce the possibility of changing the sample or introducing impurities during the process. Generally, gaseous, liquid and solid samples can be analyzed using appropriate sampling techniques. With experience, samples that are otherwise thought not to be appropriate for analysis by vibrational spectroscopy, can be analyzed with the appropriate choice of sampling technique.

When choosing a sampling technique where the amount of light absorbed is to be measured (Figure 2), it is important to remember the relationship known as Beer's law shown in Equation (1)

$$A = -\log\left(\frac{I}{I_o}\right. = \varepsilon bc \qquad (1)$$

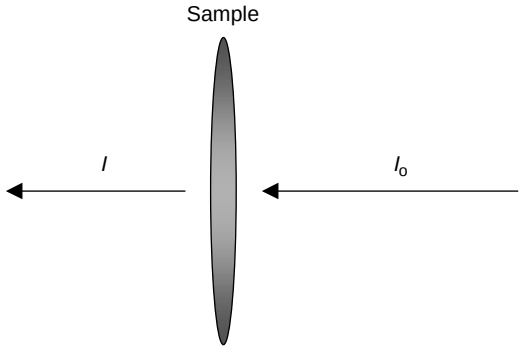

Figure 2 Schematic diagram for a typical transmission measurement. I_o is the intensity of light impinging on the sample at a given wavelength. I is the intensity of light at that wavelength after passing through the sample.

where A is the absorbance, I_o is the intensity of the incident light at a given wavelength, I is the intensity of the light at that wavelength after passing through the sample, ε is the molar absorptivity in units of liter-moles per centimeter, b is the pathlength or thickness of the sample in centimeters and c is the concentration of the analayte in moles per liter. The molar absorptivity for a given analyte varies with wavelength and spectroscopic region. It is generally observed that molar absorptivity decrease significantly in going from the UV/VIS to the MIR and NIR regions for analytes which absorb in these regions. Thus, to acquire NIR spectra, higher analyte concentrations and longer pathlengths are generally necessary than when obtaining MIR spectra. Although this may at first been seen as a disadvantage for NIR spectroscopy, it will be shown that the ability to use a longer pathlength is often considered one of the advantages of NIR for industrial applications.

The physical state of the sample often suggests the sampling technique to use. MIR spectroscopic data can be acquired for solid samples in many ways. One of the most common sampling techniques is the alkali halide or, as generally practiced, potassium bromide (KBr) pellet technique. KBr is used since it does not absorb in the MIR region and also because it cold flows when pressure is applied. Simply, the solid sample is ground to a powder, which is then mixed with spectral grade KBr powder. The mixture is placed in a special die set and sufficient pressure is applied to cause the KBr to cold flow, forming a KBr pellet, often contained in a metal collar which is part of the die set. The collar containing the pellet is placed in the instrument and an absorbance spectrum is acquired. The ratio of sample to KBr used and the thickness of the KBr pellet formed can be adjusted to obtain the desired level of analyte absorbance. Performing a quantitative analysis using the alkali halide technique can be challenging due to difficulties in preparing reproducible and representative samples.

A similar technique, often used by synthetic chemists to obtain qualitative information, is to prepare what is known as a ''mull''. The solid sample is ground to a powder and dispersed in a few drops of mineral oil. This mixture is applied to a KBr or NaCl optical window and the absorbance spectrum for the mixture is obtained. Since mineral oil contains carbon–hydrogen and carbon–carbon bonds, which absorb in the MIR, regions of the spectrum containing such absorption are not utilized. To obtain spectral information for regions of the spectrum obscured by the bands of the mineral oil, a different mulling agent, generally a perfluorinated oil, is used. The perfluorinated oil absorbs in the region for carbon–carbon and carbon–fluorine bonds but provides the spectral regions not available with a mineral oil mull. Whenever preparing a mull or an alkali halide pellet, it is prudent to remember that potential chemical and/or physical changes may result from grinding the sample.

Another common way to obtain spectra of solids is attenuated total reflectance (ATR).[9,10] Sampling by ATR is demonstrated diagrammatically in Figure 3. ATR sampling utilizes an optical element, which is made from a material having a high refractive index, for example zinc selenide, germanium, or diamond. Sampling is based on the principle that as light passes through a material with a high refractive index, an evanescent wave propagates at the surface of the material in contact with a medium having a lower refractive index. The evanescent wave penetrates into this lower refractive index medium for some distance. The intensity of the evanescent wave decreases exponentially with distance away from the element. If the contacting medium, which in this case is the sample, absorbs light at a particular wavelength, then a decrease in the intensity of light at this wavelength emerging from the ATR element will be observed. The depth of penetration of the evanescent wave depends on the refractive indices of the ATR element, the contacting medium (generally air or the sample) and the angle of

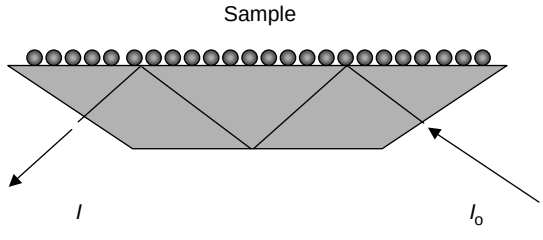

Figure 3 Schematic diagram of sampling by ATR indicating the lightpath and sample arrangement. I_o is the intensity of light impinging on the ATR element at a given wavelength. I is the intensity of light at that wavelength emerging from the ATR element.

incidence of the light on the ATR element. The depth of penetration is wavelength dependent and generally is found to be something between a fraction of a wavelength to several wavelengths in magnitude. The number of reflections in the ATR element can be adjusted by changing the thickness of the ATR element, the length of the ATR element or the angle of incidence of the light entering the element.

Since the depth of penetration of the evanescent wave into the adjacent medium is generally small, the sample must be in intimate contact with the ATR element. Samples which are soft can be pressed on to the ATR element to achieve such contact usually without damaging the element. Care must be taken not to damage ATR elements made from soft materials when pressing hard solid samples or powders into such elements. Likewise, care must be taken not to crack and break ATR elements made from brittle materials when pressing the sample on to the surface. This is especially true when obtaining data for hard samples which are not flat. Thus, in addition to spectroscopic concerns (depth of penetration, wavelength cutoffs and regions of absorption) the type of sample also determines the material from which the ATR element is made. Since the sample is in intimate contact with the ATR element, the chemical reactivity of the sample and the ATR element must also be taken into account.

Another sampling method for obtaining spectra of solids is diffuse reflectance (Figure 4). Diffuse reflectance spectra are obtained by directing light on to a sample and collecting the light that is reflected, diffusely, not specularly, from the sample. The amount of light at the wavelengths which the sample absorbs will be reduced in the diffusely reflected light. Thus, to obtain a spectrum, similar to an absorption spectrum, the diffusely reflected light collected from a sample is ratioed to the light diffusely reflected from a standard. Solid samples are often diluted with potassium chloride and ground to a powder, especially when obtaining a MIR spectrum. The diffuse reflectance technique is widely use for obtaining NIR spectra, often with no sample preparation required.

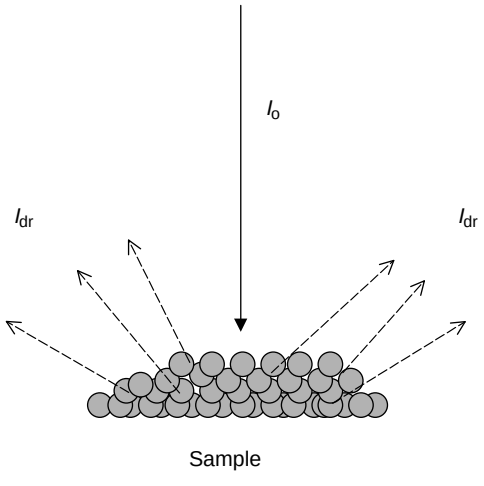

Figure 4 Schematic diagram for typical diffuse reflection measurement. I_o is the intensity of light impinging on the sample at a given wavelength. I_{dr} is the intensity of light at that wavelength after being diffusely reflected by the sample.

Spectra for liquid samples can be acquired using several sampling techniques. Transmission spectra can be acquired with the use of windows, which are transparent in the IR region of interest. The choice of the window material is made based, not only on the spectral region required for the analysis, but also on the chemical reactivity and solubility of the window material. The most common MIR transmitting materials used for this purpose are potassium bromide (KBr) and sodium chloride (NaCl). Both are water soluble, thus eliminating their use with aqueous solutions or even nonaqueous solutions containing water as an impurity. The lower wavenumber cutoff is $675\,cm^{-1}$ and $400\,cm^{-1}$ for NaCl and KBr respectively. Nonwater-soluble MIR windows, for example calcium fluoride and barium fluoride, are also readily available.

For nonvolatile solvents where maintaining the pathlength is not critical, spectra can be acquired by simply placing a drop of the liquid between two IR transmitting windows. Liquid cells generally use spacers to maintain a given pathlength and are often used if quantitative data are to be obtained. One concern in obtaining spectra for analytes in solution is the absorbance profile of the solvent. The primary concern is that the solvent may absorb strongly in the region of the spectrum where the analyte absorbs, thereby making it difficult if not impossible to obtain spectral information for the analyte. Water is a particularly difficult solvent for IR spectroscopy since it absorbs strongly in the MIR region of the spectrum, especially in the $1600\,cm^{-1}$ and $3000\,cm^{-1}$ regions of the spectrum. To obtain IR spectra of aqueous solutions, very short pathlengths (i.e. $6\,\mu m$) must be used. The choice of solvent and concentration of analyte in the sample dictates the pathlength used for the analysis. Flow cells are also available.

Liquid samples can also be analyzed using ATR techniques. Many types of ATR liquid probes are commercially available. Additionally, there are many different designs for liquid ATR cells which are available commercially, ranging from a simple-to-use trough design, where the liquid in contact with the ATR element is contained in a trough, to more elaborate ATR liquid flow cells. As mentioned above, the depth of penetration in the ATR experiment is determined by several factors, including the refractive index of the ATR element, the refractive index of the sample, which is a liquid in this case, and the angle of incidence of the light. Liquid ATR cells are normally

designed such that the actual thickness of the sample in contact with the ATR element is much greater than the depth of penetration, with the only limit being the amount of sample available. Thus, ATR sampling may prove to be experimentally easier when obtaining spectra of viscous samples or samples which absorb strongly and would require, therefore, very short pathlengths to obtain useful transmission spectra.

Spectra can also be obtained for gas phase samples using appropriate sampling devices. The lower analyte concentrations of gaseous samples requires the use of much longer pathlengths than are used with liquid samples, often several centimeters to several meters. Transmission cells can be constructed from inert materials, such as glass or stainless steel with appropriately chosen IR transmitting windows positioned in the light path to allow the light to pass through the sample.

Similar sampling devices are used for acquiring MIR and NIR spectra, although the lower molar absorptivities for analytes in the NIR region lead to some differences. The lower molar absorptivities allow longer pathlengths to be used for NIR transmission experiments and, as mentioned above, diffuse reflectance spectra can usually be acquired without diluting or grinding the sample. The virtual elimination of sample preparation is a major advantage of NIR spectroscopy. Fiber optic probes are more widely used in NIR than MIR sampling.[11]

Sampling is usually more straightforward for acquiring FT Raman spectra. The goal is to present the sample in such a way that light from the excitation laser impinges on the sample and the Raman scattered radiation is directed to the collection optics. Sampling is further simplified by the fact that glass is a weak Raman scatterer and a reasonable window material in the NIR region of the spectrum. Thus, glass nuclear magnetic resonance (NMR) tubes and melting point capillary tubes make economical sample containers for obtaining FT Raman spectra of either liquids or powdered solid samples. Caution must be exercised however in the choice of glass tube in that some glasses contain impurities or constituents that fluoresce when exposed to the NIR excitation from the laser source. An intense fluorescence background can obscure the weaker Raman bands. Another sampling concern when obtaining Raman spectra is that the sample can be degraded by the intense laser radiation. Likewise, temperature-dependent changes in the spectra may be observed due to heating of the sample by the laser. Rotating the sample so that the same spot is not constantly in the laser focus and externally cooling the sample are options for overcoming these problems. FT Raman spectra can also be acquired for solid samples with design of an appropriate sample holder. It should be noted obtaining representative spectral data for large samples may be a concern because of the small area irradiated by the laser.

4 MICROSAMPLING AND MICROSCOPY

Microsampling and microscopic imaging, both IR and Raman, are areas of vibrational spectroscopy which have undergone major advances in the past 10 years. Although these techniques are often classified as being distinct from vibrational spectroscopic techniques using macroscopic sampling, it could be rationalized that these imaging techniques are simply extensions of microsampling with unique and powerful advantages. Interfacing a microscope to an IR or Raman instrument provides the opportunity to obtain spectral information for microsamples, often called microspectroscopy, and to obtain vibrational spectroscopic images.[12–16]

The power of the microimaging techniques lies in their ability to provide spatial images for specific frequencies within the MIR wavenumber range. Since the frequencies of absorption in this range are correlated with known chemical functionalities, functionality mapping can easily be provided within the spatial resolution of the instruments. The limiting spatial resolution for IR imaging is typically in the range of 5–10 μm, which, while not high by optical imaging standards, is a limitation often overcome by the quality and quantity of spectral data obtained. The spectral resolution reported is typically $4\,cm^{-1}$. Spectra can be acquired for samples either by transmission or reflection.

Many approaches have been taken to develop microsampling/microimaging instruments. One approach is to interface a microscope suitably modified to contain optics for the spectral region of interest with a conventional scanning FTIR instrument. This type of instrument uses a single IR detector element. Spectral data are acquired for one position of the sample at any given time. Clearly this instrumental arrangement can be utilized to obtain IR spectra of small samples, which is really better termed microspectroscopy rather than imaging. To obtain an IR image of a sample, the sample stage is moved a precise distance between data acquisition for adjacent points.[13] The scientific literature contains many elegant studies performed using this rather mechanical approach to IR imaging, which although time-consuming has been practiced for many years and is representative of most of the IR microscopic imaging systems in use today. Recently, IR array detectors have become available. The availability of these detectors is likely to revolutionize the use of IR imaging in the laboratory, assuming the cost for these instruments eventually reaches a level which makes them available to reasonably well equipped laboratories.

One recent development in IR microscopic imaging systems is the combination of a step scan FTIR instrument with a suitably modified microscope containing IR optics and an IR array detector, typically a 64×64 array of mercury cadmium telluride (MCT) elements.[15,16] To understand how this instrument obtains IR spectra, it is necessary to review how a conventional scanning FTIR works. When using a continuously scanning interferometer and a single detector, as is common in laboratory instruments, with or without a microscope attachment, a complete IR spectrum is obtained each time the moving mirror of the interferometer makes a complete scan. The detector simply measures the intensity of the IR radiation falling on it at discrete times (i.e. distances moved by the mirror) as the mirror makes a continuous scan. This approach works well for a single detector element, but does not work if the data acquired from multiple elements (i.e. 4096 elements in a 64×64 MCT array) must be acquired and saved before moving on to the next data acquisition position of the mirror. To overcome this problem, the interferometer is changed from one that acquires data by making a continuous scan to one that has a mirror which stops and acquires data for each appropriate position in the mirror travel. Thus, although each full scan produces a complete interferogram, each point of the interferogram is acquired by stopping the mirror, collecting the IR intensity impinging on each element of the array, and saving these data before proceeding to the next mirror position where the process is repeated. Clearly advances in this instrumentation are likely to be made as larger, faster detector arrays and associated electronics become available in the future.

Applications and issues in microspectroscopy, with special emphasis placed on determining the types of particles of particular interest to people working with pharmaceuticals have been covered extensively by Aldrich and Smith.[12] They provide an appendix listing many sources of particulates commonly found in pharmaceuticals. They also describe general approaches, methods and problems encountered in identifying particulate matter and show how IR microspectroscopy can provide information which complements or confirms the information obtained using other techniques. Applications of IR microimaging in several fields, including minerals, biological materials and semiconductor materials have been reviewed by Krishnan et al.[13] and others.[16] While not applications of pharmaceuticals directly, the examples provided demonstrate the broad applicability of the technique. Owing to the recent developments in array technology and the limited availability and high cost of the latest versions of the FTIR imaging instruments, references to their use for solving pharmaceutical problems are limited. Still it is not hard to imagine, based on the applications already in the literature, that the impact of this new technology and likely improvements in hardware, software, and pricing will result in the use of such systems for solving pharmaceutical problems. An interesting hybrid sampling technique which has recently become available combines the features of ATR and microscopy. Using this technique the sample being analyzed is visually imaged through the ATR element, thus allowing the scientist to see the region being sampled by the element. This micro-ATR sampling technique, using a diamond-tipped ATR element, has been reviewed by Coates and Reffner.[17] As pointed out by these authors, the development of FTIR instruments has greatly increased the opportunity for using sampling techniques where energy throughput considerations were limiting for the older dispersive IR instruments, such as ATR and microsampling applications. Another development has been the introduction of diamond surface ATR elements.[17] Clearly, the use of diamond in addition to providing mechanical advantages also provides more chemical compatibility than many of the other high-refractive-index materials commonly used as ATR elements. (MIR diamond ATR devices are available and have been used for process analysis applications.[18]) Owing to the high cost of diamond, diamond-tipped ATR elements have been used for microsampling. For these devices, the area sampled is approximately 1–2 mm in diameter. The light absorbed by the diamond could interfere with an analysis, but this effect is limited by the small pathlength of diamond through which the IR light must travel owing to the limited thickness of the diamond coating and the limited number of ATR reflections (i.e. 1) in the diamond material.

The ability to visually observe the area being sampled by viewing through the ATR element clearly has merit. A sampling device employing a diamond, micro-ATR element and a charge-coupled device camera which allows such sample viewing has been developed and has recently became commercially available. Coates and Reffner[17] present several applications as examples of sampling using this device, including the analysis of ink printed on a business card, the analysis of several layers of a fiber-optic cable, and the analysis of hair. This approach to IR sampling may prove to be a practical and possibly lower-cost alternative to IR microscopy for routine analysis of at least some types of samples.

5 APPLICATIONS

5.1 Mid-infrared

Pharmaceutical applications of MIR spectroscopy have been published for decades, but the commercial availability of lower-cost FTIR instruments has greatly increased

the range of applications and data processing possible. One area of application which has received a great deal of attention in the literature is the analysis of polymorphs. Several selected examples of polymorph analysis are presented here. For additional references and background the reader is directed to one of the several excellent reviews published.[19,20]

Hartauer et al.[21] described the quantitative analysis of polymorphs of the antibiotic sulfamethoxazole with diffuse reflectance FTIR utilizing the partial least squares (PLS) method for data analysis. The use of diffuse reflectance instead of the KBr pellet or mull techniques is highlighted because of advances in the signal-to-noise ratio obtained with commercially available diffuse reflectance infrared Fourier transform (DRIFT) accessories. Data are collected for light, which is diffusely, not specularly reflected, by the material of interest, which is usually dispersed in a nonabsorbing matrix such as potassium chloride. The authors point out that to obtain quantitative information from the reflectance data obtained, the Kubelka–Munk relationship is employed. This relationship produces a linear relationship between the Kubelka–Munk intensity and the concentration of scattering species, although there are several assumptions inherent in the use of this relationship. The scattering coefficient for the material of interest must be constant, and the light reaching the detector must be diffusely, not specularly, reflected from the sample. As emphasized by the authors, there are several experimental variables which need to be monitored to control the amount of specularly reflected light and the scattering coefficient of the material. Among these variables are some which can be controlled by sample preparation, including the particle size of the scattering material, the packing density, and the homogeneity of the sample, as well as the optical design of the diffuse reflectance accessory used. Fitting the overlapping peaks of the spectra obtained for appropriate calibration standards using the PLS regression approach allowed the concentration of each polymorphic form in an unknown sample to be calculated. For calibration, pure samples of the two sulfamethoxazole polymorphs were prepared and characterized. Sample preparation included grinding the samples to obtain a more homogeneous particle size, although sieving would have been employed if grinding was found to affect the polymorphs. The ground samples were then vigorously mixed with micronized KCl powder, which had a particle size of 1–10 μm. The mixed powder was then placed in a sample cup and compacted using a 2-kg weight. The importance of using a reproducible sample method was stressed. It was noted that significantly sharper spectral bands were obtained for the polymorph which had been ground. Experiments were performed to check the linearity of the Kubelka–Munk intensity with concentration. At lower concentrations of the polymorphs (i.e. less than 2% in KCl) good linearity was observed for several of the less intense bands of each of the polymorphs, but the authors caution that intense bands in the spectra do not demonstrate such linearity, possibly because of increases in specular reflection with increased absorptivity at a given wavelength. Using PLS methods to analyze the data for several calibration standards provided average absolute errors between the actual and predicted values of 0.006% and 0.004% for the two polymorphs of sulfamethoxazole, given as a percentage of the each of the polymorphs in KCl.

Previously, Deeley et al.[22] compared FTRS with NIR excitation to FTIR spectroscopy for analyzing binary mixtures containing two of the five polymorphs of cortisone acetate. The polymorphs of this compound are known to be insensitive to grinding, thus grinding can be used to prepare homogeneous mixtures of the polymorph standards. Both FT Raman and FTIR spectra were recorded for this study using the same optical bench. One potential advantage of FTRS is that no sample preparation is required for powdered samples, although often minimal sample preparation is required for diffuse reflection FTIR measurements. One of the concerns in Raman spectroscopy is the amount of excitation laser power to use for the measurement. Increasing the excitation power of the laser increases the Raman scattering of the sample, but also may thermally damage the sample or increase the number of photons due to fluorescence by the sample or impurities. For this study, excitation powers ranging from 125 to 250 mW mm^{-2} were used. It should be noted that since the intensity of Raman scattered radiation from a sample is dependent on excitation power and the Raman experiment is generally performed as a single-beam experiment (i.e. no ratio performed to correct for source fluctuations), the laser power must be constant for quantitative studies. The authors also state that the degree of spectral difference for the polymorphs of most compounds is similar when comparing FT Raman and FTIR spectra. The exception is that the IR spectra of polymorphs involving hydrogen bonding tend to show larger differences than those observed in their Raman spectra. Spectra were analyzed using a commercially available modified principle components regression analysis package. Experimental characterization of two of the polymorphs of cortisone acetate by FTRS was reported with standard errors of prediction of less than 2.5% in some cases, showing that in some cases this technique is a reasonable alternative to FTIR for characterizing polymorphs.

Salari and Young[23] used FTIR/ATR to identify and quantify the three polymorphs of a model compound, ganciclovir, in mixtures. They point out that the use of ATR for analyzing polymorphs in pharmaceuticals was suggested by Higuchi et al.[24] in 1969. ATR spectra can be acquired quickly, with little or no sample

preparation required, and this sampling technique is nondestructive. Samples are generally not subjected to excessive pressures, in contrast to the pressures used to prepare KBr pellets to obtain transmission spectra. Experiments were performed to determine the concentrations of the three known polymorphs of the antiviral, ganciclovir. The polymorphs have distinctive spectra in the region between 1800 cm^{-1} and 600 cm^{-1}. The calibration set included the pure polymorphs along with binary and ternary mixtures prepared from the pure polymorphs.

ATR spectra were acquired with the use of a commercially available, 9-bounce diamond horizontal ATR accessory. Limited pressure was applied to maintain contact between the sample and the diamond ATR element. The PLS algorithm was used for quantitation. The mean absolute errors between actual and predicted concentrations of the polymorphs in the mixtures for one of the validation sets ranged from approximately 3 to 6 w% and were polymorph dependent.

Crystals of famotidine, an H_2-receptor antagonist, have been studied by several researchers using several spectroscopic methods, including FTIR.[25,26] The solubility and therefore the bioavailability of the different polymorphs of a drug may be different, which is one of the reasons necessitating the efforts to identify and characterize them. Hagedüs et al.[25] studied two polymorphs of the drug crystallized from water. To prepare one form, the water was allowed to cool very slowly while the second form was prepared by pouring the aqueous solution over ice. FTIR spectra were recorded at 1 cm^{-1} resolution for the crystals dispersed in KBr and pelletized. Spectra for the two polymorphs obtained are clearly different. Using the absorption band at 3506 cm^{-1}, it was estimated that the presence of between 1% and 2% of the crystal form prepared by the rapid cooling method could be determined in a mixture being primarily the crystal form obtained by slowly cooling the aqueous solution.

Hassan et al.[26] prepared and characterized three polymorphic forms of famotidine which were determined to be present in crystals prepared from several solvents and mixtures of solvents. The ''A'' form was crystallized slowly from acetonitrile. The ''B'' form was crystallized from methanol and mixtures of water–methanol, methanol–diethyl ether, acetonitrile–diethyl ether, tetrahydrofuran–diethyl ether, tetrahydrofuran–n-hexane, methanol–benzene and methanol–toluene. The ''C'' form was prepared by dissolving the famotidine in boiling tetrahydrofuran followed by filtering and adding diethyl ether so that the crystals formed in a tetrahydrofuran–diethyl ether solution. Transmission IR spectra were acquired for the crystals prepared as KBr pellets. It was noted that grinding and the pressure required to prepare the KBr pellets did not affect the polymorphic form of the compound. Significant differences were observed in the spectra of the A crystal form, which was prepared from acetonitrile and crystals obtained from tetrahydrofuran–ether or methanol. A distinct peak at 1672 cm^{-1} was observed for the crystal form obtained from acetonitrile. A similar absorption was observed by Hegedüs et al.[25] for one of the polymorphs that they had characterized. Likewise, both groups observed a peak at 3506 cm^{-1} in one of the crystal forms. Other spectra observed for the two other crystal forms were also highlighted.

An idea of the range of applications of MIR spectroscopy can be seen in the following examples. Wasacz et al.[27] described the use of MIR diffuse reflectance data to determine the hydration state of pharmaceuticals, specifically a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitor. They described the spectral differences that are observed when spectra are taken for the sample incorporated into a KBr pellet or taken by diffuse reflectance without use of a halide diluent. Diffuse reflectance data were acquired at several temperatures, from 25 to 185 °C for samples which had been exposed to different atmospheric humidity conditions. They monitored the spectral features for adsorbed water and water of hydration which is incorporated into the crystal. Also monitored were changes in the OH stretching region for magnesium sterate, a compound commonly used in pharmaceutical formulations, which exists in anhydrous and several hydrated forms. Differences observed in the OH stretching region for spectra acquired by the variable temperature DRIFT technique allowed the characterization of both the di- and trihydrated forms of this compound. Their conclusion was that using variable temperature DRIFT spectroscopy and exposing the compound to different humidity conditions provided ''unambiguous characterization of the pharmaceutical hydrates''.

Several papers have reported using FTIR spectroscopy to monitor interactions between pharmaceutical compounds and carriers. Hirasawa et al.[28] described using FTIR spectroscopy to characterize interactions between naproxen and lactose, which was used as the carrier for the solid dispersion. Samples were prepared by mixing various proportions of the drug, heating to 220 °C, which is the melting point of lactose, stirring the resulting melt to provide a homogeneous mixture, and then rapidly cooling to solidify. FTIR spectra were acquired for the solid samples as KBr pellets. Monitoring the bands in the carbonyl region for crystalline naproxen, and both physical mixtures and solid dispersions of naproxen and lactose provided information on the type of interactions which occur for the drug in the solid dispersion. The spectra of the carbonyl region for these samples provided information on the presence of the free carboxylic

group, a dimeric structure and a hydrogen bonding interaction between the lactose and naproxen. In the solid dispersions, the degree of hydrogen bonding between the lactose and the drug was found to be dependent on the carrier/drug ratio.

FTIR spectroscopy has also been used to monitor changes in the matrix of polymer-based, controlled-release delivery systems.[29] It has been used as part of a study to characterize degradable poly(lactide-*co*-glycolide) (PLGA) films employed to deliver tetracycline. Of particular interest for this study was the effect of incorporating a soluble salt, sodium chloride, at different loadings into the polymeric system. Drug release studies were performed by placing the tetracycline containing film samples in phosphate-buffered saline and held at 37 °C for up to 42 days. Samples of the films were taken at predetermined times and lyophilized. FTIR spectra of cast films of these samples were acquired up to the 21-day time point. Peaks which are characteristic for lactic acid, glycolic acid, and PLGA observed in the 1400–1500 cm^{-1} region were monitored over time. While no change in the peak intensities was observed during the first day of immersion, for samples removed in subsequent days the intensity of the glycolic acid and PLGA peaks decreased relative to the lactic acid peak. These changes in relative peak intensity with time were interpreted to indicate a difference in the rate of degradation between the glycolic acid components and the lactic acid components of the polymer. Spectral differences were also noted in the 1000–1300 cm^{-1} region for samples with and without sodium chloride incorporated into the polymer. These differences were interpreted to indicate that the salt containing samples were more crystalline.

FTIR, X-ray diffraction (XRD) and thermal analysis were used to characterize formulations of the calcium-channel blocker, nifedipine, in binary systems with the phospholipids, dipalmitoylphosphatidylcholine (DPPC) or dimyristoylphosphatidylcholine (DMPC).[30] It is been shown that the dissolution and absorption of drugs which are very sparingly soluble in water can be improved by formulating with small amounts of a phospholipid, such as DPPC or DMPC. The study included physical mixtures, solid dispersions and ground mixtures of nifedipine and phospholipid at approximately 9 : 1 w/w ratios. FTIR data were acquired for these samples as KBr pellets. The authors concluded, based on the FTIR spectra obtained, that preparing the nifedipine–phospholipid mixtures by either grinding or coprecipitation resulted in no molecular or crystal structure changes.

The ability of specific compounds to enhance the transdermal delivery of drugs has been investigated on humans, in vivo, using ATR coupled with FTIR.[31] Oleic acid is known to be a skin permeation enhancer. Such compounds reduce the skin barrier function, which has been correlated with increased transepidermal water loss (TEWL). Studies were performed exposing the patches of skin on the arms of the volunteers with test solutions of mixtures of oleic acid and propylene glycol. Both TEWL and ATR/FTIR measurements were performed. The FTIR measurements were made by placing a ZnSe ATR element directly in contact with the skin. The area of skin sampled was 7 cm^2.

The IR absorption bands in the carbon–hydrogen stretching region (~3000 cm^{-1}) and the carbonyl region (~1720 cm^{-1}) were monitored to ascertain the location and effect of oleic acid on the skin. The position and bandwidth of the methylene asymmetric stretches of the stratum corneum lipids provided an indication of the perturbing effect that the oleic acid has on the stratum corneum lipid barrier. The greater the perturbation, the larger the shift of the methylene asymmetric stretching frequency to higher wavenumber. Carbonyl bands at 1740 cm^{-1} and 1710 cm^{-1} were used as indicators of the esterified carbonyl group of the esterified lipids and the free carbonyl group of the oleic acid, respectively. Monitoring the absorbance of the free carboxylic acid group provided an indication of the movement of the oleic acid from the surface either by diffusing deeper in the skin or spreading laterally through the skin. Changes in the 1710 cm^{-1}/1740 cm^{-1} absorbance ratio over time could also possibly indicate esterification of the oleic acid functional group.

An interesting application of FTIR/ATR of interest for pharmaceuticals is the determination of diffusion coefficients for liquids in anhydrous lanolin and poly(ethylene glycol) (PEG) ointment.[32] The diffusion coefficients were determined for several test liquids, including pyridine, acetophenone, methyl salicylate, and 4-chloro-4′-fluorobutyrophenone. A horizontal ATR cell with a 45° angle of incidence ZnSe ATR element was used for these measurements. The experiments were performed by placing a layer of lanolin or PEG ointment on the ATR element. The layer was contained in a template prepared by punching a hole in a 0.157 mm-thick sheet of Teflon®. The thickness of the layer was limited to the thickness of the template by scraping off any excess with a razor blade. A bead of the same material was placed around the hole and served as a liquid barrier. The diffusion coefficients were determined by placing 20-μL aliquots of the test solution in the well formed, and acquiring IR spectra over time as the compounds diffused into the layer. The depth of penetration of the IR beam into the layer was estimated to be from 0.6 to 0.9 μm. As the concentration of the liquid within this depth of penetration increased with time due to diffusion, the magnitude of the absorbance of the characteristic peaks for the liquid increased. The increase in absorbance was monitored over time with these data used to calculate the diffusion

coefficients for the test liquids. The diffusion coefficients measured ranged from 0.56×10^{-7} to 7.2×10^{-7} $cm^2\,s^{-1}$.

Many studies have shown that particle size is an important variable that must be controlled to obtain quantitative results from diffuse reflectance FTIR measurements.[33,34] Otsuka and Matsuda[33] exploited the relationship between intensity of light absorbed and particle size, as derived by Duyckaerts,[35] to measure the particle size of phenytoin in pharmaceutical preparations. Diffuse reflectance spectra were obtained for the phenytoin powders dispersed in micronized KBr powder at 0.5%. The reflectance spectra obtained were corrected using the Kubelka–Munk relationship. The intensity of the phenytoin absorption at 1724 cm^{-1} was measured after correcting the baseline in that region. A linear relationship was found between the absorbance at 1724 cm^{-1} and the reciprocal of the mean particle size, both for pure phenytoin and a binary mixture of phenytoin and lactose.

One of the strengths of MIR and NIR spectroscopy is the ability to analyze solid samples directly. Solid dosage forms of benzodiazepines were studied by reflectance spectroscopy by Ficarra et al.[36] These studies were performed using a dispersive IR spectrometer operating at 8.5 cm^{-1} resolution at 2000 cm^{-1}. Samples were mixed with KBr powder in proportions of 1 : 10 w/w. Plots of the Kubulka–Munk transformed reflectance data versus wavenumber were produced for several pure benzodiazepines. Diazepam in both laboratory-prepared excipient mixtures used to simulate the commercial drug product and commercially available capsules were analyzed using the same sample preparation procedures. The results presented indicate that using this method the concentration of diazepam in drug product can be determined with a high degree of confidence.

The use of NIR and MIR diffuse reflectance spectroscopy to verify the identity and content of drug formulations in both capsules and tablets was investigated.[37] The goal of the study was to develop a method for routine identity testing for clinical packaging. Requirements included that the method be rapid and could be applied by minimally skilled technicians. Formulations containing simvastatin, enalapril maleate, lovastatin and finastride were employed in this study. Distinguishing between lovastatin and simvastatin, which differ only by an α methyl group of an ester carbonyl, was taken as a difficult identification problem. The MIR region was found to be capable of providing such discrimination, while NIR spectra were not able to provide as clear discrimination at low concentrations. Sample treatment included milling the capsules and tablets with a grinding mill. A grinding time in excess of 5 min was found to affect the quantitation but not drug identification. A resolution of 8 cm^{-1} was determined to be sufficient for this analysis. For tablets and capsules containing 0.2–40 mg of drug, the method was capable of determining 1% (w/w) of drug in the specific formulations tested.

Proteins of pharmaceutical interest have also been the subject of MIR studies. ATR/FTIR was used to study silk fibrogen gel, which was being investigated as a new oral dosage form for elderly patients.[38] Silk fibrogen is a protein which is harmless to humans, has unique properties, and provides a dosage form with a jelly-like consistency, which makes it easier to swallow. Spectra were acquired for the silk fibrogen in solution and as a gel by ATR using a germanium ATR element. Changes in the structure of the protein from random coil to β-structure as the protein gels were determined by monitoring the amide I and amide II bands of the protein between about 1650 cm^{-1} and 1520 cm^{-1}.

Carpenter et al.[39] reviewed the use of MIR spectroscopy for monitoring the secondary structure of proteins to establish formulations which provide stable lyophilized products. It is known that some excipients can prevent proteins from denaturing during the freeze-drying process. Using IR spectroscopy, the secondary structure of the protein can be monitored in aqueous solution, the frozen sample and as a freeze-dried solid. Comparison of the secondary structure, especially the degree of unfolding of the protein, is indicative of the stability of the freeze-dried product. The positions of the amide I bands of the protein, which occur in the 1700–1600 cm^{-1} region, are sensitive to the secondary structure of the protein. Monitoring changes in the band intensity and position provides information on changes in the native state of the protein in going from the protein in aqueous solution to the lyophilized solid. The effect of excipients, such as sucrose, on the stability of the protein can thus be evaluated. Obtaining the IR spectra for aqueous solutions containing several milligrams per milliliter is reasonably straightforward with the use of appropriate, short-pathlength sample cells and modern FTIR instruments. Spectral data processing, however, can be somewhat involved since the amide I bands of the protein overlap the strong water absorption band in this region. Spectral subtraction is performed to remove the water contribution from the spectrum of the protein in solution. The amide I bands of the protein are broad and overlapping, but taking the second derivative of the resulting spectrum assists in determining the positions of the individual amide I bands of the protein. The positions of these bands are indicative of the presence of the various secondary structures (i.e. α-helix, β-sheet, unordered) present in the protein.

Often FTIR instruments are coupled with other instrumentation resulting in a powerful analytical tool for a specific type of analysis. Rodriguez and Bugay[40] combined a thermogravimetric instrument with an FTIR instrument using a commercially available interface to

determine the identity of the compounds volatilized from a sample upon heating. The evolved gases were carried using helium into a multipass gas cell which was held at 160 °C. FTIR spectra were acquired at a rate of 1 min^{-1} at a spectral resolution of 4 cm^{-1} as the samples were heated from ambient temperature to 300 °C at a rate of 10 °C min^{-1}. The availability of spectral libraries of vapor phase species simplifies compound identification. Experiments can be performed on a little as 1 mg of sample and analysis times are short. Basically, the experiment follows the weight loss of the material as the temperature increases and correlates changes observed in the spectrum of the gas in the spectroscopic cell with changes in the sample's mass.

Studies were conducted to determine the form of several solvates and provide information on the presence of residual solvents for several pharmaceutical compounds. One example presented shows how the authors used the combined thermogravimetric analysis (TGA)/FTIR technique to determine that a butyl acetate solvate was formed in a 1 : 1 mole ratio for a compound under investigation for the treatment of hypercholesterolemia. The authors show that TGA/FTIR, when used in combination with other techniques such as differential scanning calorimetry (DSC) and XRD, provides a powerful tool for studying solid pharmaceuticals.

Photoacoustic (PA) spectroscopy is another hybrid spectroscopic technique which, although not as widely utilized as the other sampling techniques discussed, may provide unique advantages for some pharmaceutically important sample types.[6] The FTIR/PA spectroscopy experiment is based on the principle that when a sample absorbs light, a portion of the energy absorbed is dissipated as heat. If light is absorbed close to the surface of the sample, the temperature of the surface and the atmosphere immediately surrounding the sample will increase. Modulating the source of the light results in a modulation of the surface temperature, and likewise the temperature of the surrounding atmosphere, resulting in the formation of an acoustic wave which can be detected using a microphone. PA spectra have been acquired using both dispersive IR and FTIR instruments. The use of FTIR instruments employing step-scan, rather than continuously scanned, interferometers for PA applications has been of recent interest.[41,42]

Neubert et al.[43] used FTIR/PA spectroscopy to determine the concentration of the drugs dithranol and brivudin in a petroleum jelly formulation. To determine the concentration of active ingredient in such formulations often requires the use of a separation technique such as high-performance liquid chromatography (HPLC) or capillary zone electrophoresis (CZE) to physically separate the active ingredients from the matrix prior to quantitation. Spectroscopic techniques can eliminate the need for employing a physical method of separation, with the minimal sample preparation required for the PA technique being a major advantage. Using bands between 1545 cm^{-1} and 1770 cm^{-1} for brivudin and between 1560 cm^{-1} and 1670 cm^{-1} for dithranol a good correlation was found between the concentrations determined by HPLC and by PA spectroscopy for samples ranging in concentration between 1.0 and 10.0 wt% dithranol in petroleum jelly and for CZE and PA spectroscopy between 0.5 and 10.0 wt% brivudin in petroleum jelly. The integrated intensities for the bands of dithranol and brivudin were ratioed to the intensity of a band due to the petroleum jelly. The data were acquired using a step-scan interferometer with a total scan time of approximately 21 min. The PA spectra were acquired at a resolution of 12 cm^{-1} and were normalized by ratioing to a spectrum of carbon black.

One area which has potential for being utilized more for pharmaceutical applications is vibrational optical activity (VOA). Though VOA includes both Raman optical activity and the corresponding IR version, vibrational circular dichroism (VCD), recent developments are the reason for concentrating on VCD here.[2,44] VCD can be used to obtain information on the optical purity of a mixture of enantiomers, help determine the absolute configuration of a molecule or provide conformational information.

The basis for VCD is that optically active molecules interact differently with right circularly polarized light than left circularly polarized light. One of the enantiomers of a molecule may, for example, absorb more right circularly polarized light, than left circularly polarized light at the same wavelength. A VCD instrument is designed to measure the difference between the amount of light absorbed for the two polarizations at a given wavenumber, or ΔA. The VCD spectrum is a plot of ΔA vs wavenumber.

The magnitude of ΔA is very small, approximately 5×10^{-4} times the magnitude of the absorbance.[2] Measuring such a small difference between the amount of light absorbed for the different polarizations makes the measuring of a VCD spectrum instrumentally challenging. Both dispersive and FTIR-based VCD instruments have been assembled. The modification of either instrument to acquire VCD spectra involves passing alternately left and right polarized light through the sample and measuring the difference in the amount of light absorbed. To produce the left and right circularly polarized light, two polarizers are placed in the optical train prior to the sample. First, unpolarized light produced by the IR source is linearly polarized by passing through a wire grid polarizer. This linearly polarized IR light then is converted to either right or left circularly polarized

light using a photoelastic modulator (PEM). The PEM is operated so as to alternately produce right and left circularly polarized light.

VCD, in addition to providing information on absolute configuration of an optically active compound is also capable of providing information on the conformation of such a compound in different environments. For example, Nafie and Freedman[44] discuss the use of VCD for studying the conformation of the antiarrhythmic drug, (*S*)-propranolol, in solution. The recent introduction of a commercial, dedicated, FTIR-based, VCD instrument has the potential to increase the use and importance of VCD for pharmaceutical applications in the future.[44]

5.2 Near-infrared Spectroscopy

One of the best places to get started for anyone interested in NIR spectroscopy is the many reviews on NIR applications in the pharmaceutical industry.[8,19,45–48] NIR spectroscopy has been employed for the analysis of pharmaceuticals for over 30 years. The history of the application of NIR to pharmaceuticals has been reviewed by Kirsch and Drennen.[46] The review also discusses briefly the many chemometric methods which have been applied to NIR spectra and reviews the application of NIR of solid dosage formulations and intact dosage forms. In order to obtain NIR spectra for solid dosage forms, the authors indicate that sample positioning is one of the most important variables in the experiment, being the ''single largest source of error''. One approach to reducing this error is to acquire several spectra of a given tablet while rotating the tablet between spectral data acquisition. The average spectrum for the tablet is then computed thereby reducing the errors associated with positioning.

Corti et al.[47] indicate that several of the advantages of NIR spectroscopy include being nondestructive, noninvasive and applicable to many sample types. Another advantage is that the lack of response of glass in the NIR region allows samples in glass containers to be analyzed. The authors also comment that the shorter wavelengths in the NIR result in more linear and reproducible calibrations for data acquired using diffuse reflectance as compared to the use of this sampling technique for obtaining MIR spectra. One of the limitations of NIR analysis mentioned is that the calibration standards must be analyzed using an independent reference method. Two of the most important aspects of the NIR analysis are the choice of the chemometric model for quantitation and determining which standards to use for calibration. Corti et al.[47] discuss briefly the differences between the multiple linear regression (MLR) method and the full spectrum methods of principal component analysis (PCA), principal component regression (PCR) and PLS. Pretreatment of the data prior to the application of these models is also a consideration. Choosing the correct number and type of standards is also critical. Clearly the goal is to maximize the information provided by standards that are representative of the samples to be analyzed while minimizing the number of standards and time required to perform the calibration. Approaches to determining the correct standards for calibration are discussed. Several applications of NIR analyses are also discussed.

Another excellent, detailed source of information on NIR analysis of pharmaceuticals is the recent review by Blanco et al.[8] which sites over 200 references. Many examples are provided on both quantitative and qualitative NIR analyses. Sampling and data treatment methods are also discussed. A timely section listing the websites of many Internet sources of NIR information is also provided.

Higgins[48] provides an insightful evaluation of the use of NIR spectroscopy for industrial pharmaceutical applications. Among the points raised are that NIR spectroscopy is very useful for confirming the identity of materials and that the technique is ideal for pharmaceutical applications since the materials are generally organic and of high purity. Clearly this application is based on having high-quality raw material standard reference spectra available. The nondestructive nature of this type of analysis along with its ability to sample through many types of containers, including plastic bags, blister packages, disposable glass vials, and other packaging materials, also adds to the versatility of the technique. Liquids and powders can also be conveniently and routinely sampled. The application of NIR spectroscopy to monitoring raw materials, in-process monitoring, and postproduction monitoring, are highlighted. An example presented is the use of NIR spectroscopy to evaluate the degree of powder blending in production. In this example, the spectral response of both the active ingredient and excipients are indicative of the degree of homogeneity of the product and, thus, an indicator of the extent of blending which has taken place. Other applications mentioned include the use of NIR spectroscopy for the quantitative determination of water in freeze dried injection products, direct assays of bulk tablets and monitoring fermentation processes. One important point raised by the author, concerning the pharmaceutical applications of NIR spectroscopy, is that validated methods using NIR technology are now accepted by pharmaceutical regulators. Higgins also points out that NIR spectroscopy must eventually be accepted as a replacement for other, more traditional, methods of analysis.

Catalano[49] discussed the implementation of NIR in the process environment. Included in the discussion were the different worlds in which the people who are involved in dealing with the process and the laboratory people

exist as well as differences in the environments in which process and laboratory instruments must operate. He also states the importance of working with a vendor in the development of process monitoring applications since the vendor should have more experience in performing such analyses under process conditions. Interestingly, Catalano stresses the importance and roles of the people involved in developing and implementing the NIR application in the process environment, especially for environments where there is no history of using such instrumentation for process applications.

Ciurczak[50] explains the development of NIR spectroscopy from the time of the discovery of this region of the electromagnetic spectrum by Hershel, through the development of the early instruments and applications of NIR to food products. Comparisons are made between MIR spectroscopy and NIR spectroscopy. What is interesting is that those factors which are often thought of as being limitations in MIR spectroscopy are exploited in NIR spectroscopy, especially for process applications. For example, MIR spectroscopy is often thought of as being an insensitive technique by analytical chemists due to the molar absorptivities being several orders of magnitude lower than those found in the UV/VIS region. The molar absorptivities in the NIR region are approximately an order of magnitude lower than those in the MIR region, which leads directly to the use of longer pathlengths for NIR measurements. The need to use longer pathlengths and/or higher concentrations, while generally thought of as a limitation in the analytical laboratory, is a benefit in process analysis. The limited sensitivity in the NIR region allows the use of larger pathlength flow-through sample ''cells'' which do not clog and reduce the need to dilute samples or otherwise handle the samples, as is often required when using other forms of spectroscopy. Ciurczak also briefly describes the need for mathematical data manipulation to obtain quantitative results from the NIR data acquired. Several methods and applications are mentioned. The importance of having samples and standards analyzed by a primary method available for calibration of the NIR system is also stressed.

Determining the moisture content of final drug products and components is often important. NIR spectroscopy provides a means for nondestructive analysis which requires little if any sample preparation. Zhou et al.[51] evaluated the effect of several experimental variables on the determination of the moisture content of hygroscopic drug substance at moisture levels ranging from 0.5 to 11.4% (w/w). These analyses make use of the strong water absorption bands in the NIR region (1420 and 1920 nm). Diffuse reflectance spectra were acquired from 1100 to 2500 nm for the samples in sealed glass bottles, with 16 scans being averaged for each spectrum. The reflectance spectrum for each sample was ratioed to the spectrum of a white ceramic reference. The authors chose to analyze the spectral data obtained using PLS although simple linear regression and MLR are also often employed for such analyses due to the strong water absorptions and lack of spectral interferences in these regions. The ability of PLS to remove spectral noise and model nonlinearity were given as potential reasons for using the method for this application. The results indicate that a standard error of prediction of 0.11% could be obtained in the moisture range studied for these hygroscopic samples in sealed glass containers. The use of first derivative spectra does provide a better calibration model for the 1850–1936 nm region.

Berntsson et al.[52] used NIR diffuse reflectance spectroscopy to determine the water content of hard gelatin capsules with analysis times of approximately 2 min. As in the previous example the spectra containing the strong water absorption bands were acquired in the 1100–2500 nm region. The moisture content of the capsules is important, since if it is above approximately 18% the capsules become soft, whereas below a water content of approximately 10% the capsules become brittle. Thus, the moisture content of the capsules needs to be checked before filling, which is why a fast, at-line method for moisture content evaluation is of interest. These researchers compared the results obtained using diffuse reflectance NIR spectroscopy with data obtained from loss on drying experiments. Calibration standards of the gelatin capsules were prepared by placing the gelatin capsules in chambers having known constant relative humidity for up to 20 days. A low moisture standard was also prepared by placing gelatin capsules in a chamber with desiccant for 5 days.

MLR and PLS were utilized to process the spectral data acquired. The effects of several spectral manipulations performed prior to fitting the data were also evaluated. They found that MLR using three wavelengths chosen by an iterative procedure produced the best calibration model for the data presented. Interestingly, none of the wavelengths chosen were the most intense for water absorption. All 70 wavelengths were used for calibration with PLS, which is a more robust technique than MLR and can detect outliers. The PLS weighting vectors, unlike the selected wavelengths in MLR, were found to be heavily weighted towards the regions where water absorbs strongly. Though reportedly not observed in this study, it was noted that other researchers have observed wavelength-independent changes in the spectral baseline with water absorption on silica.[53] It was also noted that while empty, hard gelatin capsules were used for this study, the method should also be applicable to filled capsules, assuming appropriate calibrations standards are utilized.

Borer et al.[54] evaluated the important sources of variability for analyses using NIR reflectance spectroscopy. The study concentrated on determining dose levels for tableted drug product samples. Several potential contributors to variability in the NIR measurement were evaluated, including the iris or aperture setting, the segment, which is a calculation variable indicating the number of data points over which the slope is calculated when converting the spectral data to a second derivative spectrum, the orientation of the sample platform, the number of scans which are averaged to obtain a spectrum, the number of samples of each tablet analyzed, the time in days over which the experiment is performed and the frequency at which a reference spectrum is acquired. The magnitude of the contribution of each of the sources of variability was evaluated. For example, the authors concluded that the segment value was a significant factor in the variability observed in the study indicating the slope should be calculated over 10 points. The orientation was important for embossed samples, while the number of scans was not an important source of variability. The effects of other potential sources of variability, including lot-to-lot variations in the tablets, inhomogeneity within a lot of tablets, and temperature were not evaluated in this study. The study shows the importance of evaluating the magnitude of the contribution of each of the many potential sources of variability in NIR and other spectroscopic analyses.

Comparisons between different data acquisition methods, types of NIR instruments and data manipulations for quantitative analyses of solid pharmaceutical samples was evaluated by Dreassi et al.[55] Experiments were performed using two NIR instruments, one an FT NIR instrument and the other a dispersive NIR instrument. The methods of sampling compared included direct diffuse reflectance measurements of the samples and sampling using a fiber optic probe. The PLS algorithm was employed for quantitation. Samples analyzed included powders containing benzydamine hydrochloride and tricetol, as well as ibuprofen and paracetamol as pills or tablets. Quantitation based on the second derivative spectra were found to provide the best calibration. The authors comment that using higher derivative spectra reduces the effects of particle size on the reflectance data obtained, thereby improving calibrations for samples having a distribution of particle sizes.

Wargo and Drennen[56] used diffuse reflectance NIR spectroscopy to characterize the degree of homogeneity of powder blends. Blend homogeneity is an important consideration in preparing drug product and is typically measured prior to tableting or encapsulation. The homogeneity of blends of hydrochlorothiazide, fast-flow lactose, croscarmellose sodium and magnesium stearate were investigated in this study. Data were acquired using a dispersive NIR instrument by collecting spectra from 1200 to 2400 nm sampling directly through the bottom of the borosilicate sample vials. The homogeneity of a blend is often determined by measuring the amount of the active ingredient in several samples from the same batch. If the active ingredient is found to be homogeneously distributed, the other components of the blend are also considered to be evenly distributed. Data treatment for the NIR spectra included linearization by performing a $\log(1/R)$ transformation, where R is the reflectance. The homogeneity of samples with the same component concentrations mixed for different time intervals in a twin-shell blender were assessed. The degree of homogeneity for a separate blend sampled over time was also assessed. The reflectance data obtained were evaluated to determine sample homogeneity using three methods: chi-squared analysis, the bootstrap error-adjusted single-sample technique (BEST) and the modified BEST technique. Second derivative spectra were calculated prior to performing a PCA. Chi-squared analysis is performed on individual wavelengths while the two BEST techniques utilize the entire spectrum. The results obtained indicate that NIR spectroscopy has great potential for assessing blend homogeneity as well as being a useful tool in studies designed to determine optimal mixing time and conditions.

The sampling and sample handling techniques used for acquiring spectral data is only one of the issues that must be addressed when determining an appropriate method for an analysis. The approach taken to process the spectral data is often a major concern in developing the analytical method. Often an analytical method is developed to confirm the classification of samples in a given group. One example of this type of analysis is the verification of drug lots used for clinical studies. Candolfi et al.[57] evaluated several approaches to processing the NIR reflectance data obtained for capsules and tablets containing several concentrations of active ingredient, several excipients and for several lots. The goal was to determine which data-processing method was best for classifying the spectral data obtained. Discriminating methods were the focus of the study. A comparison was made using both parametric methods (linear discriminate analysis and quadratic discriminate analysis) and nonparametric methods (K-nearest neighbor). NIR reflectance data were acquired for the spectral region from 10 000 to 4000 cm^{-1} for the powdered tablets and capsules using a FT/NIR instrument. The authors conclude that for clinical study lots it might be best to use a two-step approach for final classification. The first step is to discriminate between the different classes present and the second step is to use another data treatment method to positively identify the sample.

Blanco et al.[58] studied the effect which the choice of calibration standards has on the quantitative results obtained for NIR diffuse reflectance analysis of both the active ingredient and major excipients of a solid pharmaceutical formulation. An optical fiber probe was used to acquire the spectroscopic data. The active ingredient for this study was otilonium bromide with microcrystalline cellulose, maize starch and sodium starch glycolate excipients. The importance of using standards for calibration which adequately represent the variability expected to be observed in the samples was discussed. Sources of variability between laboratory-prepared standards and standards obtained from production runs were discussed. The need for the use of a multivariate calibration method, such as PLS regression analysis, is demonstrated. The quantitative results obtained using both laboratory-prepared mixtures of the ingredients with and without additional production samples as standards are presented and discussed. The effects of calibration standards choice, spectral region used for analysis and several mathematical treatments of the spectral data, including first and second derivative spectra, were evaluated. This study clearly points out the importance of systematically evaluating how representative the standards used for quantitation are when developing a NIR diffuse reflectance method and throughout the life of the application of the method developed.

Sources of variability in the NIR data were investigated systematically for tablets, and hard and soft gelatin capsules by Candolfi et al.[59] Changes in both chemical and physical variables are possible sources of the spectral variability observed in such analyses. The tablets and capsules sampled contained both active ingredients, ranging from 4 to 10% in concentration, and between 4 and 10 excipients. The active ingredients were different for the various sample types. The major excipient for the tablets was lactose while the major excipient for the hard gel capsules was cellulose. The sources of error evaluated included positioning of the sample, batch-to-batch variability, interday variability and within-batch variability. The effects of spectral preprocessing methods, such as second derivatization and standard normal variate transformation on the results obtained was also evaluated. It was pointed out that while data processing and manipulation can overcome some of the variability observed in the spectral data obtained, improving the quality of the data is also important. One variable which must be controlled is the moisture content of the samples. Water absorbs strongly in the NIR region. Even if the regions of the strongest water absorption bands are avoided in interpreting the spectral data, water also affects the baseline throughout this region. Instrumental drift was also noted as problematic. These authors suggest that to improve the quality of data obtained, the NIR instrument should be located in an area where both the temperature and humidity are controlled.

The major sources of error were found by Candolfi et al. to differ for the different sample types evaluated. For example, sample positioning was found to be an important variable for both the soft and hard gelatin capsules examined. Time was also found to be a significant cause of the variation observed for these samples. One problem with analyzing gelatin capsules which could add to the variability observed is that the gelatin is sensitive to changes in humidity. For tablets, significant batch-to-batch and within-batch variability was observed which indicates a large number of representative samples which must be included in the training sets for this type of drug delivery system.

An interesting application of NIR spectroscopy assesses the degree of cross-linking of the gelatin-comprising hard gelatin capsules, soft gelatin capsules, and gelatin coatings, which is important due to their wide use in the pharmaceutical, food supplement and cosmetic industries. The degree of cross-linking affects the dissolution properties of these materials. Gold et al.[60,61] have used NIR to study the effect of formaldehyde on both hard and soft gelatin capsules. In one of these studies,[60] NIR spectroscopy was coupled with PCA to monitor the extent of formaldehyde-induced cross-linking in soft elastic gelatin capsules. Capsules were filled with PEG solutions (PEG 400) containing up to 0.40% formaldehyde by volume and allowed to interact for 48 h. The filling solution was then removed and the inside surfaces were thoroughly rinsed with PEG and diethyl ether. Reflectance NIR spectra were obtained for these capsules. Following a series of spectral corrections and mathematical manipulations, the resulting spectra were analyzed using PCA. Baseline shifts were observed which were attributed to changes in the water content of the capsules exposed to the formaldehyde-containing solutions since water is such a strong absorber in the NIR. Changes were, however, also observed in the absorbances at 1780 and 2200 nm which were considered to be the result of changes in gelatin cross-linking caused by the formaldehyde. These spectral changes are likely due to the effects in the NIR region of the spectrum of bonds being formed and broken during the crosslinking process.

Gold et al.[61] also used NIR reflectance spectroscopy to study the effects of exposing empty, unopened, hard gelatin capsules to an atmosphere containing 150 ppb of formaldehyde for up to 24 h. After formaldehyde exposure, the capsules were filled with amoxicillin and NIR reflectance spectra obtained. The NIR data were correlated with data obtained for the dissolution of amoxicillin from the formaldehyde-treated capsules.

5.3 Fourier Transform Raman Spectroscopy

FTRS has many properties which make it attractive for pharmaceutical applications.[62] One important property which has been exploited for biological samples is that water is a weak Raman scatterer, which means that Raman spectra can be obtained for aqueous solutions and samples which contain a high water content. Another of the complications of IR spectroscopy, the requirement that IR-transmitting materials be used for sample containers, is also diminished. Raman spectra can be acquired for samples in glass containers directly, although some care must be taken since the ingredients and impurities found in some glasses may fluoresce when exposed to the intense NIR radiation of the laser source. Probably the most important advantage which FT Raman using NIR excitation has over conventional Raman spectroscopy using UV or VIS wavelengths for excitation is that fluorescence from the sample is generally reduced. Sample fluorescence, which is generally far more intense than the Raman scattering, can result from either a fluorescent analyte or impurities in the sample. The energy of the NIR photons used for excitation in FTRS is not sufficient to result in the electronic transition required for the fluorescence of most species. One other aspect of Raman spectroscopy which is often mentioned is its complementary nature with respect to IR spectroscopy. Absorption bands which are weak in the IR spectrum of the analyte are often strong in the Raman spectrum of the material. One disadvantage of FTRS using NIR excitation, as pointed out by Spragg,[62] is that the high-power, invisible NIR light used for sample illumination gives rise to important safety issues. For this reason, the laser, sample, optics and detector are enclosed in commercial instruments and interlocked to prevent exposure to laboratory personnel.

Petty et al.[63] reviewed the advantages of using Raman spectroscopy, FTRS in particular, for pharmaceutical applications. Many advantages of using FTRS for analysis are noted and several examples are presented. Citing the fact that many packaging materials, again glass in particular, are poor Raman scatterers, which allows spectra to be acquired without removing the sample from its container, Spragg[62] points out that most materials are weak absorbers of the NIR wavelength normally used for excitation in FT Raman. Obtaining Raman spectra for samples in their containers is a rather extreme demonstration that the technique requires little to no sample preparation. An example of an analysis where this property is important is a study involving the compound tegafur, which is cytotoxic.[63] Special handling procedures are normally required when analyzing this material, but can be avoided since Raman spectra can be acquired without removing the sample from its container.[63]

Another important advantage of Raman spectroscopy highlighted by Petty et al.[63] is that it provides an easy means to obtain information on the low-energy vibrations occurring between 400 and 50 cm^{-1}. Most FTIR instruments have a lower wavenumber limit of 400 cm^{-1}, a constraint imposed by the use of KBr optical components (i.e. windows, interferometer, beamsplitter, etc.). To obtain IR spectra in the region below 400 cm^{-1}, the far IR region of the spectrum, requires the use of optical components fabricated from other materials, for example CsI, which is very hygroscopic. Another complication is that water vapor absorbs strongly in the FIR region. Lattice vibrations, which may help distinguish between different polymorphs of a compound, are an example of such low-energy transitions. An example is presented for two crystal forms of an angiotensin II receptor antagonist. Distinct and characteristic differences are observed in the FT Raman spectra in the region between 50 and 425 cm^{-1} for the two crystal forms of the compound.

Hendra[64] explained why FTRS is very often applicable to mixtures of pharmaceutical importance. Raman scattering for noncrystalline, saturated compounds composed of first-row elements is often weak compared with drugs, which are often crystalline, aromatic compounds containing elements not in the first row. Thus, although the drug may be present at lower concentrations than some of the other materials of the drug product, for example a polymer such as cellulose, the Raman scattering as a result of the drug substance may dominate the spectrum obtained. Hendra[64] uses the example of lactose as an excipient which may, however, cause problems since it is a crystalline material and is a strong Raman scatterer. Further, if the lactose in the formulation absorbs water and becomes hydrated, the Raman spectrum of the material will vary depending on the degree of hydration. Hendra[64] also points out that as with IR spectra, the Raman spectra obtained are dependent on the temperature of the sample. It is pointed out that both the intensity and widths of the Raman bands can change with sample temperature.

An early application of FTRS was in the study of drugs which had been dispersed in polymeric materials used in drug delivery systems.[65,66] FTRS using a NIR laser for excitation is more applicable to these analyses than conventional Raman spectroscopy using VIS lasers due to the decreased interference from fluorescence when using NIR excitation and the decreased concern about photodegradation. Poly(sebacic anhydride) is given as an example of a biomedical polymer. The authors point out that by studying changes in the spectral properties of these polymers as the material is exposed to a degradation process provides for the identification of

the end products and the degree of polymer modification. For example, the end products of hydrolysis or exposure to light can be determined simply, often in situ, via FTRS. Such studies may provide information concerning the release of a drug incorporated in the polymer matrix. Analysis of FT Raman spectra of pure drug substance and drug substances dispersed in a polymer matrix provides information on the extent and type of drug – polymer interaction. For example, comparison of the spectra of the drug diclofenac, pure and as dispersed at a concentration of 20% (w/w) in sodium alginate, indicate that there is no molecular interaction between the drug and the polymer. Analysis of concentrations up to 60% (w/w) diclofenac in sodium alginate did, however, show a nonlinear, concentration-dependent intensity change for bands because of the aromatic structure of the molecule. The authors point out that using NIR excitation for FTRS may not eliminate all sample fluorescence and may present some problems for aqueous samples. In another study, Tudor et al.[66] highlighted the use of FTRS using NIR excitation at 1064 nm provided by a Nd : YAG laser and showed spectra for several sympathomimetic amines and poly(anhydrides).

Langkilde et al.[67] employed FTRS to characterize two polymorphs of an unnamed pharmaceutical compound. These two polymorphs are known to be stable at different temperatures and have distinct Raman spectra in the carbonyl region (bands at 1716 or 1724 cm^{-1}) due to the presence of a conjugated ester group which packs differently in the crystal structure of the two forms of the compound. According to these authors the bands are better resolved for the polymorphic forms in the FT Raman spectra than in FTIR spectra, although this may be due to sample treatment. The authors also point out that other researchers have found that polymorphs having distinct IR spectra do not have distinct Raman spectra.[22] A spinning-ring-type sample device was developed and used to increase the area of material sampled; thus the spectra obtained represent more closely a true average for the material present in an inhomogeneous sample. Data treatment included subtracting spectra of each of the pure polymorphs from the spectra of the mixture to determine the contribution of the other polymorph to the observed Raman intensity. For this study, it was determined that the presence of one polymorphic form in the other could be determined from about 1.8 to 15.4% of the minor form. The authors caution that one potential problem is that the absolute intensity of the Raman signal is dependent on particle size, with larger particles resulting in greater Raman intensities. Thus, particle size needs to be considered when performing such analyses, and efforts should be made to control particle size between samples and to ensure that different components in the mixture have the same particle size.

Taylor and Zografi[68] demonstrated the use of FTRS for monitoring the crystallinity of indomethacin powders. The Raman spectra of the γ crystalline and amorphous indomethacin differ in the carbonyl region. Raman-scattering bands for the crystalline and amorphous forms of the compound occur at 1698 and 1680 cm^{-1}, respectively, although there is some overlap due to broadening of the band for the amorphous state. To overcome errors caused by the overlap, spectra were acquired for known standards containing known ratios of the crystalline and amorphous indomethacin. Peak intensity ratios were calculated and used for calibration. Raman spectra were acquired at 4 cm^{-1} resolution using the 1064-nm line of a Nd : YAG laser for excitation. Several different experiments were performed to investigate potential sources of error. The use of intensity ratios overcomes much of the potential error associated with fluctuations in the excitation laser intensity. Although the absolute intensity of Raman scattering increases with increasing laser power, the intensity ratios measured were not found to be very sensitive to such fluctuations. One of the largest sources of error was found to be the homogeneity of mixing of the amorphous and crystalline components of the sample. The effect of particle size on Raman signal intensity also needs to be investigated. The authors, however, indicate that for indomethacin under the conditions studied, FT Raman spectroscopic analysis is capable of detecting the presence of crystalline or amorphous indomethacin to as low as the 1% level with linear calibration being obtained over the entire crystallinity range.

6 CONCLUSION

The range of applications presented show clearly that vibrational spectroscopy, especially MIR, NIR and FTRS, are well suited to analyzing samples of importance to the pharmaceutical industry. The many sampling methods available, the limited sample preparation generally required, and the nondestructive nature of the vibrational spectroscopy, are well suited to the analysis of pharmaceutical products, raw materials, and samples. Discussions with representatives of several instrument manufacturers, however, have indicated that many pharmaceutical applications of their technology, especially in the area of process control and analysis, are proprietary and not reported in the peer review literature. Thus, for some types of applications, important information should be obtained directly from the instrument designers, manufacturers, and suppliers. Instrument manufacturers' Internet websites often prove to be excellent sources of the latest details of instrument design, at least

for that particular manufacturer, and also often provide limited information on proprietary uses for that instrumentation. As with other types of analytical instrumentation, vibrational spectroscopic instrumentation is constantly being developed and improved. Some of the most exciting developments in vibrational spectroscopy are the advances currently being made in vibrational imaging. As the cost of MIR imaging instruments reaches a point where they become more widely available, their use in the analysis of pharmaceutically important samples will be more readily apparent. Likewise, continued developments in FT Raman instrumentation and increases in general availability in industrial laboratories will ensure that the unique capabilities of this technique will be used for pharmaceutically related applications.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Sue Avery, Grace Blinn, and Robert Kowalski of the Genetics Institute Information Center for their assistance in obtaining literature sources. Lisa Housiantis, Stephan Koza, and Hubert Scoble, Biopharmaceutical Characterization and Analysis Department, Genetics Institute, are acknowledged for helpful discussions, suggestions and encouragement throughout. The authors also acknowledge the assistance of Joseph Danahy of Computer Information Services at the University of New Hampshire. Also gratefully acknowledged is the assistance provided by the many experts and representatives of vibrational spectroscopic instrument manufacturers who were contacted. Their willingness to provide information, references, etc., is greatly appreciated.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflectance
BEST	Bootstrap Error-adjusted Single-sample Technique
CZE	Capillary Zone Electrophoresis
DMPC	Dimyristoylphosphatidylcholine
DPPC	Dipalmitoylphosphatidylcholine
DRIFT	Diffuse Reflectance Infrared Fourier Transform
DSC	Differential Scanning Calorimetry
FIR	Far Infrared
FT	Fourier Transform
FTIR	Fourier Transform Infrared
FTRS	Fourier Transform Raman Spectroscopy
HMG-CoA	3-Hydroxy-3-methylglutaryl Coenzyme A
HPLC	High-performance Liquid Chromatography
IR	Infrared
MCT	Mercury Cadmium Telluride
MIR	Mid-infrared
MLR	Multiple Linear Regression
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
PA	Photoacoustic
PCA	Principal Component Analysis
PCR	Principal Component Regression
PEG	Poly(ethylene glycol)
PEM	Photoelastic Modulator
PLGA	Poly(lactide-*co*-glycolide)
PLS	Partial Least Squares
TEWL	Transepidermal Water Loss
TGA	Thermogravimetric Analysis
UV	Ultraviolet
VCD	Vibrational Circular Dichroism
VIS	Visible
VOA	Vibrational Optical Activity
XRD	X-ray Diffraction

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Clinical and Diagnostic Analysis • Infrared Spectroscopy in Microbiology • Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Near-infrared Spectroscopy in Food Analysis

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis

Pulp and Paper **(Volume 9)**
Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Liquid Chromatography/Infrared Spectroscopy • Microspectroscopy • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy • Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation • Raman Microscopy and Imaging • Raman Scattering, Fundamentals

REFERENCES

1. N.B. Colthup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, 3rd edition, Academic Press, Inc., Boston, 1990.
2. M. Diem, *Introduction to Modern Vibrational Spectroscopy*, Wiley & Sons, New York, 1993.
3. D.A. Skoog, F.J. Holler, T.A. Nieman, *Principles of Instrumental Analysis*, 5th edition, Harcourt Brace & Company, Philadelphia, 1998.
4. H.A. Strobel, W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, 3rd edition , Wiley & Sons, New York, 1989.
5. J.D. Ingel, Jr, S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall College Division, New Jersey, 1988.
6. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley & Sons, New York, 1986.
7. J.E. Brown, A.C. Gilby, B.G. Lipták, T.M. Cardis, 'Infrared Analyzers', in *Analytical Instrumentation*, ed. B.G. Lipták, Chilton Book Company, Radnor, PA, 171–185, 1994.
8. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela, 'Near-infrared Spectroscopy in the Pharmaceutical Industry', *Analyst*, **123**, 135R–150R (1998).
9. N.J. Harrick, *Internal Reflection Spectroscopy*, Wiley & Sons, New York, 1967.
10. F.M. Mirabella, Jr, N.J. Harrick, *Internal Reflection Spectroscopy: Review and Supplement*, Harrick Scientific Corp., Ossining, 1985.
11. D.C. Hassell, E.R. Bowman, 'Process Analytical Chemistry for Spectroscopists', *Appl. Spectrosc.*, **52**, 18A–29A (1998).
12. D.S. Aldrich, M.A. Smith, 'Pharmaceutical Applications of Infrared Microspectroscopy', in *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, Inc., New York, 323–375, 1995.
13. K. Krishnan, J.R. Powell, S.L. Hill, 'Infrared Microimaging', in *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, Inc., New York, 85–110, 1995.
14. J.L. Koenig, C.M. Snively, 'Fast FTIR Imaging: Theory and Applications', *Spectroscopy*, **13**, 22–28 (1998).
15. M. Diem, S. Boydston-White, L. Chiriboga, 'Infrared Spectroscopy of Cells and Tissues: Shining Light onto a Novel Subject', *Appl. Spectrosc.*, **53**, 148A–161A (1999).
16. P. Colarusso, L.H. Kidder, I.W. Levin, J.C. Fraser, J.F. Arens, E.N. Lewis, 'Infrared Spectroscopic Imaging: From Planetary to Cellular Systems', *Appl. Spectrosc.*, **52**, 106A–120A (1998).
17. J. Coates, J. Reffner, 'Visualization of Micro-ATR Infrared Spectroscopy', *Spectroscopy*, **14**, 34–45 (1999).
18. Applications Literature, ASI Applied Systems, Inc. – A Mettler-Toledo Company, Millersville, MD.
19. D.E. Bugay, A.C. Williams, 'Vibrational Spectroscopy', in *Physical Characterization of Pharmaceutical Solids*, ed. H.G. Brittain, Marcel Dekker, Inc., New York, 1995.
20. H.G. Brittain, 'Spectral Methods for the Characterization of Polymorphs and Solvates', *J. Pharm. Sci.*, **86**, 405–412 (1997).
21. K.J. Hartauer, E.S. Miller, J.K. Guillory, 'Diffuse-reflectance Infrared Fourier-transform Spectroscopy for the Quantitative Analysis of Mixtures of Polymorphs', *Int. J. Pharm.*, **85**, 163–174 (1992).
22. C.M. Deeley, R.A. Spragg, T.L. Threlfall, 'A Comparison of Fourier Transform Infrared and Near-infrared Fourier Transform Raman Spectroscopy for Quantitative Measurements: An Application in Polymorphism', *Spectrochim. Acta*, **47A**, 1217–1223 (1991).
23. A. Salari, R.E. Young, 'Application of Attenuated Total Reflectance FTIR Spectroscopy to the Analysis of

Mixtures of Pharmaceutical Polymorphs', *Int. J. Pharm.*, **163**, 157–166 (1998).
24. W.I. Higuchi, W.E. Hamlin, S.C. Mehta, 'Infrared Attenuated Total Reflectance (ATR) Method for Observing the Water-mediated Surface Phase Reversion of Methylprednisolone II to I during Dissolution', *J. Pharm. Sci.*, **58**, 1145–1146 (1969).
25. B. Hagedüs, P. Bod, K. Harsányi, I. Péter, A. Kálmán, L. Párkányi, 'Comparison of the Polymorphic Modifications of Famotidine', *J. Pharm. Biomed. Anal.*, **7**, 563–569 (1989).
26. M.A. Hassan, M.S. Salem, M.S. Sueliman, N.M. Najib, 'Characterization of Famotidine Polymorphic Forms', *Int. J. Pharm.*, **149**, 227–232 (1997).
27. F.M. Wasacz, S.W. Strand, D.E. Bugay, K.R. Morris, 'The Characterization of the Hydration State of Pharmaceuticals by Variable Temperature Diffuse Reflectance Infrared Spectroscopy', *Proc. SPIE-Int. Soc. Opt. Eng.*, **1575**, 305–306 (1992).
28. N. Hirasawa, K. Danjo, M. Haruna, A. Otsuka, 'Physiochemical Characterization and Drug Release Studies of Naproxen Solid Dispersions Using Lactose as a Carrier', *Chem. Pharm. Bull.*, **46**, 1027–1030 (1998).
29. W.L. Webber, F. Lago, C. Thanos, E. Mathiowitz, 'Characterization of Soluble Salt-loaded, Degradable PLGA Films and their Release of Tetracycline', *J. Biomed. Mater. Res.*, **41**, 18–29 (1998).
30. S. Yamamura, J.A. Rogers, 'Characterization and Dissolution Behavior of Nifedipine and Phosphatidylcholine Binary Systems', *Int. J. Pharm.*, **130**, 65–73 (1996).
31. H. Tanojo, H.E. Junginger, H. Boddé, 'In vivo Human Skin Permeability Enhancement by Oleic Acid: Transepidermal Water Loss and Fourier-transform Infrared Spectroscopy studies', *J. Controlled Release*, **47**, 31–39 (1997).
32. D.E. Wurster, V. Buraphacheep, J.M. Patel, 'The Determination of Diffusion Coefficients in Semisolids by Fourier Transform Infrared (FTIR) Spectroscopy', *Pharm. Res.*, **10**, 616–620 (1993).
33. M. Otsuka, Y. Matsuda, 'A Novel Method for Particle-size Evaluation of Bulk Powders in Pharmaceutical Preparations Using Fourier-transformed Powder Diffuse Reflectance Infrared Spectroscopy', *Pharm. Sci.*, **1**, 167–168 (1995).
34. M. Otsuka, Y. Matsuda, 'Comparative Evaluation of Mean Particle Size of Bulk Drug Powder in Pharmaceutical Preparations by Fourier-transformed Powder Diffuse Reflectance Infrared Spectroscopy and Dissolution Kinetics', *J. Pharm. Sci.*, **85**, 112–116 (1996).
35. G. Duyckaerts, 'Contribution á l'analyse Quantitative par Les Spectres d'absorption Infrarouge des poudres I. Examen théorique de la question', *Spectrochim. Acta*, **7**, 25–31 (1955).
36. P. Ficarra, A. Villari, R. Ficarra, G. Mondio, 'Analysis of Pharmaceutical Solid Forms by Diffuse Infrared Spectroscopy. Note II', *Farmaco (Italy)*, **42**, 241–252 (1987).
37. J.A. Ryan, S.V. Compton, M.A. Brooks, D.A.C. Compton, 'Rapid Verification of Identity and Content of Drug Formulations Using Mid-infrared Spectroscopy', *J. Pharm. Biomed. Anal.*, **9**, 303–310 (1991).
38. T. Hanawa, A. Watanabe, T. Tsuchiya, R. Ikoma, M. Hidaka, M. Sugihara, 'New Oral Dosage Form for Elderly Patients: Preparation and Characterization of Silk Fibroin Gel', *Chem. Pharm. Bull.*, **43**, 284–288 (1995).
39. J.F. Carpenter, S.J. Prestrelski, A. Dong, 'Application of Infrared Spectroscopy to Development of Stable Lyophilized Protein Formulations', *Eur. J. Pharm. Biopharm.*, **45**, 231–238 (1998).
40. C. Rodriguez, D.E. Bugay, 'Characterization of Pharmaceutical Solvates by Combined Thermogravimetric and Infrared Analysis', *J. Pharm. Sci.*, **86**, 263–266 (1997).
41. R.A. Palmer, 'Step-scan FTIR: A Versatile Tool for Time- and Phase-resolved Vibrational Spectroscopy', *Spectroscopy*, **8**, 26–36 (1998).
42. E.Y. Jian, W.J. McCarthy, D.L. Drapcho, 'FTIR PAS: A Versatile Tool for Spectral Depth Profiling Chemical Analysis', *Spectroscopy*, **13**, 21–32 (1999).
43. R. Neubert, B. Collin, S. Wartewig, 'Direct Determination of Drug Content in Semisolid Formulations Using Step-scan FTIR Photoacoustic Spectroscopy', *Pharm. Res.*, **14**, 946–948 (1997).
44. L.A. Nafie, T.B. Freedman, 'Vibrational Circular Dichroism: An Incisive Tool for Stereochemical Applications', *Enantiomer*, **3**, 283–297 (1998).
45. K.M. Morisseau, C.T. Rhodes, 'Pharmaceutical Uses of Near-infrared Spectroscopy', *Drug. Dev. Ind. Pharm.*, **21**, 1071–1090 (1995).
46. J.D. Kirsch, J.K. Drennen, 'Near-infrared Spectroscopy: Applications in the Analysis of Tablets and Solid Pharmaceutical Dosage Forms', *Appl. Spectrosc. Rev.*, **30**, 139–174 (1995).
47. P. Corti, E. Dreassi, S. Lonardi, 'Near Infrared Reflectance Analysis: Features and Applications in Pharmaceutical and Biomedical Analysis', *Il Farmaco*, **48**, 3–20 (1993).
48. M. Higgins, 'Pinpointing Production Problems with NIR Analysis', *Manufacturing Chemist*, **68**, 38–39 (1997).
49. V. Catalano, 'In Process Near-IR Spectroscopy, People Count as much as Technology', *Today's Chemist at Work*, **4**, 22–27 (1995).
50. E. Ciurczak, 'Letting you in on the Well-kept Secret of Near-IR Spectroscopy', *Today's Chemist at Work*, **4**, 31–36 (1995).
51. X. Zhou, P. Hines, M.W. Borer, 'Moisture Determination in Hygroscopic Drug Substances by Near Infrared Spectroscopy', *J. Pharm. Biomed. Anal.*, **17**, 219–225 (1998).
52. O. Berntsson, G. Zackrisson, G. Östling, 'Determination of Moisture in Hard Gelatin Capsules Using Near-infrared Spectroscopy: Applications to At-line Process

Control of Pharmaceutics', *J. Pharm. Biomed. Anal.*, **15**, 895–900 (1997).
53. A. Fong, G.M. Hieftje, 'Near-infrared Measurement of Relative and Absolute Humidity through Detection of Water Adsorbed on a Silica Gel Layer', *Anal. Chem.*, **67**, 1139–1146 (1995).
54. M.W. Borer, X. Zhou, D.M. Hays, J.D. Hofer, K.C. White, 'Evaluation of Key Sources of Variability in the Measurement of Pharmaceutical Drug Products by Near Infrared Reflectance Spectroscopy', *J. Pharm. Biomed. Anal.*, **17**, 641–650 (1998).
55. E. Dreassi, G. Ceramelli, P. Corti, M. Massacessi, P.L. Perruccio, 'Quantitative Fourier Transform Near-infrared Spectroscopy in the Quality Control of Solid Pharmaceutical Formulations', *Analyst*, **120**, 2361–2365 (1995).
56. D.J. Wargo, J.K. Drennen, 'Near-infrared Spectroscopic Characterization of Pharmaceutical Powder Blends', *J. Pharm. Biomed. Anal.*, **14**, 1415–1423 (1996).
57. A. Candolfi, W. Wu, D.L. Massart, S. Heuerding, 'Comparison of Classification Approaches Applied to NIR-spectra of Clinical Study Lots', *J. Pharm. Biomed. Anal.*, **16**, 1329–1347 (1998).
58. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela, 'Quantitation of the Active Compound and Major Excipients in a Pharmaceutical Formulation by Near Infrared Diffuse Reflectance Spectroscopy with Fiber Optical Probe', *Anal Chem. Acta*, **333**, 147–156 (1996).
59. A. Candolfi, D.L. Massart, S. Heuerding, 'Investigation of Sources of Variance which Contribute to NIR-spectroscopic Measurement of Pharmaceutical Formulations', *Anal. Chem. Acta*, **345**, 185–196 (1997).
60. T.B. Gold, R.G. Buice, Jr, R.A. Lodder, G.A. Digenis, 'Detection of Formaldehyde-induced Crosslinking in Soft Elastic Gelatin Capsules Using Near-infrared Spectrophotometry', *Pharm. Dev. Technol.*, **3**, 209–214 (1998).
61. T.B. Gold, R.G. Buice, Jr, R.A. Lodder, G.A. Digenis, 'Determination of Extent of Formaldehyde-induced Crosslinking in Hard Gelatin Capsules by Near-infrared Spectrophotometry', *Pharm. Res.*, **14**, 1046–1050 (1997).
62. R.A. Spragg, 'Pharmaceutical Analysis Using Fourier Transform (FT) Raman Spectroscopy', *Am. Lab.*, **27**, 35–37 (1995).
63. C.J. Petty, D.E. Bugay, W.P. Findlay, C. Rodriguez, 'Applications of FT-Raman Spectroscopy in the Pharmaceutical Industry', *Spectroscopy*, **11**, 41–45 (1996).
64. P.J. Hendra, 'Fourier Transform-Raman Spectroscopy in Pharmaceutical Analysis and Research', *Am. Lab.*, **28**, 17–24 (1996).
65. M.C. Davies, J.S. Binns, C.D. Melia, 'Fourier Transform Raman Spectroscopy of Polymeric Biomaterials and Drug Delivery Systems', *Spectrochim. Acta*, **46A**, 277–282 (1990).
66. A.M. Tudor, C.D. Melia, J.S. Binns, P.J. Hendra, S. Church, M.C. Davies, 'The Application of Fourier-transform Raman Spectroscopy to the Analysis of Pharmaceuticals and Biomaterials', *J. Pharm. Biomed. Anal.*, **8**, 717–720 (1990).
67. F.W. Langkilde, J. Sjöblom, L. Tekenbergs-Hjelte, J. Mrak, 'Quantitative FT-Raman Analysis of Two Crystal Forms of a Pharmaceutical Compound', *J. Pharm. Biomed. Anal.*, **15**, 687–696 (1997).
68. L.S. Taylor, G. Zografi, 'The Quantitative Analysis of Crystallinity Using FT-Raman Spectroscopy', *Pharm. Res.*, **15**, 755–761 (1998).

Vitamins: Fat and Water Soluble, Analysis of

Christopher J. Bates
MRC Human Nutrition Research, (formerly MRC Dunn Nutritional Laboratory), Cambridge, UK

1 Introduction 2
2 History 8
3 Fat-soluble Vitamins 8
3.1 Vitamins A, E and Carotenoids in Serum, Plasma and Breast Milk 8
3.2 Vitamins A, E and Carotenoids in Foods and Pharmaceuticals 11
3.3 25-Hydroxy and 1,25-Hydroxyvitamin D in Serum, Plasma and Breast Milk 11
3.4 Vitamin D and its Derivatives in Foods and Pharmaceuticals 12
3.5 Vitamin K in Serum, Plasma and Breast Milk 12
3.6 Vitamin K in Foods and Pharmaceuticals 13
4 Water-soluble Vitamins 13
4.1 Vitamin C in Serum, Plasma, White Cells, Urine and Breast Milk 13
4.2 Vitamin C in Foods and Pharmaceuticals 14
4.3 Thiamin (Vitamin B_1) Indices in Blood, Urine and Breast Milk 14
4.4 Thiamin (Vitamin B_1) in Foods and Pharmaceuticals 15
4.5 Riboflavin (Vitamin B_2) Indices in Blood, Urine and Breast Milk 15
4.6 Riboflavin (Vitamin B_2) in Foods and Pharmaceuticals 16
4.7 Vitamin B_6 Indices in Blood, Urine and Breast Milk 16
4.8 Vitamin B_6 in Foods and Pharmaceuticals 16
4.9 Folate Indices in Blood, Urine and Breast Milk 17
4.10 Folate in Foods and Pharmaceuticals 18
4.11 Vitamin B_{12} Indices in Serum, Plasma and Breast Milk 18
4.12 Vitamin B_{12} in Foods and Pharmaceuticals 18
4.13 Niacin Indices in Blood, Urine and Breast Milk 19
4.14 Niacin in Foods and Pharmaceuticals 19
4.15 Biotin Indices in Blood, Urine and Breast Milk 19
4.16 Biotin in Foods and Pharmaceuticals 19
4.17 Pantothenate Assays (Status, Foods, Pharmaceuticals) 20
4.18 Other Substances with Vitamin-like Properties 20
5 Problems Encountered in Interpreting Vitamin Status Indices 20
6 Summary 21
Abbreviations and Acronyms 21
Related Articles 21
References 22

The vitamins are a diverse group of compounds, both chemically and analytically, because they comprise a range of biomolecules whose common properties reside solely in the fact that they are essential dietary components, which are needed in relatively small amounts to sustain life and good health. In addition, the major analytical challenges that they present are derived from the facts that (1) there is a need to quantitate them in a wide range of biological matrices, which include both foods and body fluids (for status indices); (2) the concentrations that are present are usually very low, and the ratio to other components, which can be chemically very similar, is small; (3) they may be present in several or many chemically diverse, but biologically interconvertible, forms; (4) some of them are labile to heat, extremes of pH, degradative enzymes and so on and (5) there is no single analytical approach that can quantitate all of them together, within a biological matrix.

They have been divided, broadly, into fat- and water-soluble groups, and each of these has been further subdivided, mainly on a functional basis. Historically, the B-group vitamins were measured by various microbiological assays, and the others were measured (if at all) by semispecific colorimetric assays. Nowadays, there are several more-powerful techniques available, which include high-pressure liquid chromatography, with a range of detectors operating over a continuous range of selective modalities; radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA)-based techniques, which use specific protein-binding selectivities; enzyme reactivation techniques for certain B-vitamin status assays, and so on. The use of mass spectrometry (MS) is gaining

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

ground and can be combined with several other separation techniques, such as gas–liquid chromatography (GLC), capillary electrophoresis (CE) and so on. Measurement of vitamins in simple pharmaceutical mixtures (vitamin supplements etc.) is now reasonably straightforward.

The technical advances that these methodologies represent have greatly broadened the horizons of the vitamin analyst, and have permitted even some of the more difficult biological matrices to be addressed. However, the unsolved challenges of accurate vitamin analysis show little sign of receding, and there are very few, if any, laboratories worldwide that are currently able to measure all of the known vitamins in all known biological matrices. This remains a major challenge and a goal for the twenty-first century.

1 INTRODUCTION

This article describes the analytical procedures available and problems likely to be encountered for the four fat-soluble vitamins and the nine water-soluble vitamins that are that are generally recognized as being essential organic components in the human diet (Table 1). There are also other dietary organic substances which may be beneficial, or even essential under certain circumstances, but which are not currently recognized or classified as vitamins: these include several carotenoid pigments, a range of bioflavonoid polyphenols, choline, inositol, amino acids such as taurine and so on. These and other organic substances may also be essential vitamins for animals other than man, but this aspect of the subject is not included in the following account. Nevertheless, powerful separation techniques have been developed that can detect and quantitate a large number of organic components in a single analytical run. This is especially true for the carotenoids, which are frequently combined with fat-soluble vitamin A and E analyses.

Each of the vitamins comprises a small group of compounds with interrelated or interconvertible structures and biological functions. Thus chemically related compounds may perform similar functions; each vitamin is converted to several different functional forms within many living tissues and certain degradation products of these vitamins may provide useful evidence about vitamin status even though they are no longer functional. This implies that the analytical methods that are used to measure vitamin contents or status must either measure and summate all of the forms that are present, or measure a representative single form or sample of the forms that are present and then extrapolate to the sum-total of all the forms. Provided that the different forms have similar activities and that they occur in a constant ratio to each other, this is an acceptable compromise, but if this is not the case, then the problem of accurate estimation becomes more complicated. Thus the newer techniques of chemical separation and quantitation may have some disadvantages compared with the older procedures based on bioassay, but since the latter are generally difficult to automate, are usually very time-consuming and lack precision, they are, with some exceptions, steadily losing ground. Table 1 shows the synthetic form for all 13 vitamins, which would normally be used in pharmaceuticals (vitamin supplements, fortified foods etc.), and a small selection of the naturally occurring forms found in foods, blood and tissues. The latter, however, are too numerous to list exhaustively in the present review: for the folates alone, for example, the possible variations in polyglutamate chain lengths, in carbon substituents and in redox state imply several dozen biological variants. In Table 1 an indication of the relative stability of the different vitamins is included. This, too, is a complex subject because each of the different forms of the vitamins has a different stability and because the biological stability may be affected by the matrix, e.g. by protein-binding. However, it is important when planning sample collection and pretreatment procedures for subsequent analysis, to take appropriate precautions against degradation. For most of the vitamins, simple freezing at low temperature is generally adequate, but vitamin C is particularly prone to rapid oxidation, requiring extraction into a metal-chelating acid, such as metaphosphoric acid. The biological forms of folate are also prone to oxidation and will benefit from the addition of a reducing agent.

Vitamins are, by definition, micronutrients, i.e. nutrients that are present and are required in very small amounts compared with the major macronutrient components of the diet, namely the proteins, fats, polysaccharides and sugars. Indeed, nearly all of the vitamins have catalytic functions in living tissues, being recycled and reused in the biochemical pathways in which they participate, whereas the macrodietary components are continuously used up and are not recycled. This implies that the analytical challenge for the vitamins is that of a series of chemically diverse minor components within a complex matrix of major components (in foods, tissues or body fluids) and that the analytical procedures must therefore be capable of extracting them efficiently and quantitating them in the presence of large amounts of potentially interfering substances. The analytical procedures must possess an adequate combination of separation efficiency (without major losses) and analytical specificity to ensure accuracy of estimation and freedom from positive or negative interferences. In the older analytical procedures the extraction step was usually relatively crude and simple and specificity was achieved mainly by biological selectivity, e.g. by microbiological or

Table 1 Natural and synthetic forms of the vitamins and their relative stabilities in aqueous solution

Vitamin	Synthetic form	Some naturally occurring forms	Relative stability in aqueous solutions[a]	Formula
Fat-soluble vitamins				
A	Retinol acetate	Retinol Retinyl esters Retinaldehyde (Retinal) Retinoic Acid	Reasonably stable in non-haemolyzed plasma; less stable once extracted. Store plasma below –40 °C	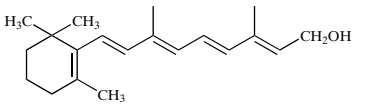 *Retinol (Vitamin A)* **(1)**
E	α-Tocopherol acetate	α-, β-, γ- and δ-tocopherols and tocotrienols	Less stable than vitamin A, being fairly easily oxidized. Store below –70 °C	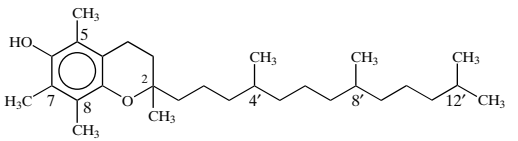 α-*Tocopherol (vitamin E)* **(2)**
D	Vitamin D_3 (cholecalciferol)	Vitamins D_2 and D_3 and hydroxylated forms, especially 25(OH)– and $1,25(OH)_2$– vitamin D	Reasonably stable in plasma; store below –20 °C	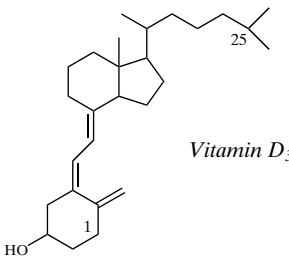 *Vitamin D_3* **(3)**
K	Menadione	Vitamin K_1 (phylloquinone) and bacterial menaquinones	Reasonably stable in plasma; store below –20 °C	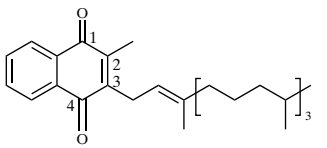 *Phylloquinone (vitamin K_1)* **(4)**

(*continued overleaf*)

Table 1 (*continued*)

Vitamin	Synthetic form	Some naturally occurring forms	Relative stability in aqueous solutions[a]	Formula
Water-soluble vitamins				
C	L-ascorbic acid	L-ascorbic and dehydroascorbic acids	Extremely unstable in neutral or alkaline solution in oxygen; must be acidified, e.g. with metaphosphoric acid. Store below –70 °C	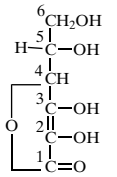 *Vitamin C (L-Ascorbic acid)* **(5)**
B_1 (thiamin)	Thiamin hydrochloride	Mainly thiamine diphosphate (cocarboxylase)	Destroyed by heating in alkaline solution. Store samples below –20 °C	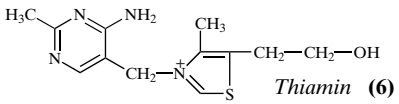 *Thiamin* **(6)**
B_2 (riboflavin)	Riboflavin	Riboflavin phosphate (FMN) and FAD	Destroyed by exposure to light; otherwise stable. Store below –20 °C	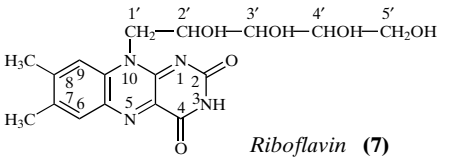 *Riboflavin* **(7)**
B_6 (pyridoxine)	Pyridoxine hydrochloride	Pyridoxal 5¢-phosphate, pyridoxal, pyridoxamine 5¢-phosphate	Very easily destroyed by light exposure. Store below –70 °C	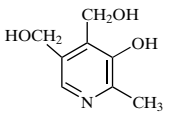 *Pyridoxine* **(8)**
Folate	Pteroyl glutamic acid	5-methyl and 5-formyl tetrahydrofolates (polyglutamate forms inside cells)	Easily destroyed by oxidation (natural forms only). Store below –70 °C; red cell extracts require a reducing agent, e.g. ascorbic acid as a preservative	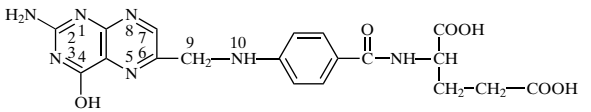 *Pteroyl glutamic acid* **(9)**

B_{12}	Cyanocobalamin	Hydroxo and methyl cobalamins	Fairly stable, but can be destroyed in the presence of ascorbate. Store below −70 °C
Niacin	Nicotinamide, nicotinic acid	Pyridine nucleotide coenzymes: NAD, NADP	Niacin in food is generally stable; the nucleotide cofactors are much less so.
Biotin	*D*-Biotin	Biotin, biocytin, biotin-containing enzymes	Relatively stable
Pantothenic acid	Pantothenic acid	CoA	Pantothenic acid is very stable; CoA is not.

Cyanocobalamin **(10)**

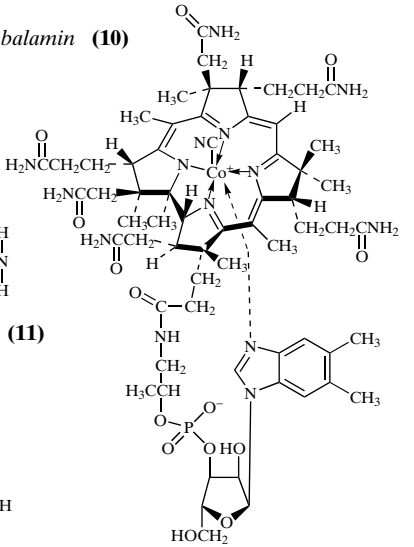

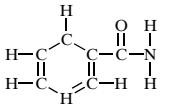

Nicotinamide **(11)**

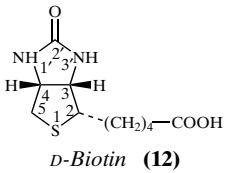

D-Biotin **(12)**

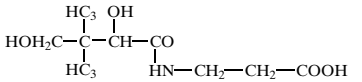

Pantothenic acid **(13)**

[a] Refers to the naturally occurring forms. For several of the vitamins in tissues the partial degradation of the native forms which occurs during storage of food may not affect the biopotency, even though the analytical patterns change. However, the biopotency of some vitamins is easily destroyed by prolonged storage or cooking (e.g. vitamin C) or light exposure (e.g. riboflavin, vitamin B_6). CoA, Coenzyme A; FAD, Flavin adenine dinucleotide; FMN, Flavin mononucleotide; NAD, Nicotinamide adenine dinucleotide; NADP, Nicotinamide adenine dinucleotide phosphate.

Table 2 Orders of magnitude of concentration of the vitamins and some other major components in adult human blood serum and in two vitamin-rich food sources

Vitamin or other component	Molecular weight[a]	Concentration in human serum ($g\,L^{-1}$)	Concentration in food	
			Soft fruit ($g\,kg^{-1}$)	Liver ($g\,kg^{-1}$)
Protein	–	60–80	10	200
Fat	–	1–10	1	70–100
Glucose	180	1	50[b]	0[b]
Cholesterol	387	1–3	–	2–4
Fat-soluble vitamins				
A	286	$0.3–0.8 \times 10^{-3}$	(0.1×10^{-3})[c]	$100–300 \times 10^{-3}$
E	431	$5–20 \times 10^{-3}$	$0–10 \times 10^{-3}$	$2–4 \times 10^{-3}$
D	385	$(10–50 \times 10^{-6})$[d]	–	$2–10 \times 10^{-6}$
K	451	$0.1–1 \times 10^{-6}$	$20–100 \times 10^{-6}$	$(100–200 \times 10^{-6})$[e]
Water-soluble vitamins				
C	176	$2–20 \times 10^{-3}$	$50–1000 \times 10^{-3}$	$100–200 \times 10^{-3}$
B_1 (thiamin)	301	$1–5 \times 10^{-6}$	$0.1–0.5 \times 10^{-3}$	$2–4 \times 10^{-3}$
B_2 (riboflavin)	376	$50–250 \times 10^{-6}$	$0.1–0.5 \times 10^{-3}$	$30–40 \times 10^{-3}$
B_6 (pyridoxine)	247	$5–30 \times 10^{-6}$	$0.1–1 \times 10^{-3}$	$4–7 \times 10^{-3}$
Folate	459	$3–20 \times 10^{-6}$	$0–300 \times 10^{-6}$	$1–6 \times 10^{-3}$
B_{12} (cobalamin)	1355	$0.2–1 \times 10^{-6}$	–	$0.2–1 \times 10^{-3}$
Niacin	122	–	$1–6 \times 10^{-3}$	$100–150 \times 10^{-3}$
Biotin	244	$0.3–1 \times 10^{-6}$	$2–20 \times 10^{-6}$	$0.3–2 \times 10^{-3}$
Pantothenic acid	219	$1–2 \times 10^{-6}$	$1–4 \times 10^{-3}$	$60–90 \times 10^{-3}$

[a] The molecular weight varies to some extent with the form that is present; the value given is that of the most common form of the native vitamin. For vitamin B_6, this is pyridoxal 5′-phosphate and for folate, 5-methyl tetrahydrofolate monoglutamate.
[b] Total sugars.
[c] As precursor carotenoids; no preformed vitamin A.
[d] Mainly as 25-hydroxyvitamin D.
[e] Mainly as menaquinones.

animal growth assays, using growth media or feeds lacking only the vitamin being tested. More modern procedures generally include more complex extraction steps and/or cleanup and separation procedures to achieve specificity, so that the final quantification procedure can then be less specific but, as a result, is more controllable and precise. Table 2 gives an order-of-magnitude estimate of the vitamin concentrations in human blood plasma and in two types of food, and it indicates the considerable magnitude of difference in concentration between the vitamins and the major components of these typical matrices. The lower the concentration of the vitamins, the greater the analytical challenge; thus vitamins D, K and B_{12} present an even greater challenge than the other vitamins. However, this is only part of the analytical problem, because a multiplicity of different forms (folate), poor stability (vitamin C) or a tight binding to macromolecules (several B-vitamins), all increase the analytical difficulties in addition to their low concentrations.

In the ensuing review it has not been possible to describe all of the variants of all of the analytical procedures for vitamins which have been published: this would be neither feasible within the space available, nor particularly useful, as the reader who requires such detailed information on such a wide range of alternative approaches will need to consult a range of reviews, electronic searches and/or information sources. Therefore I have focused primarily on the principles of the currently used analytical methods with some recent examples. I have also included some examples of newly developed techniques that may become better established in the future, such as CE. Analysis of pharmaceutical preparations is generally more straightforward from the viewpoint of analytical complexity than that of biological matrices; therefore this aspect has received less emphasis than the analysis of blood and urine samples (for status measurements) and of food samples (for food-monitoring and dietary analysis). Some of the older assay methods remain appropriate for laboratories in developing countries, which lack sophisticated modern analytical equipment and the means to maintain, repair and provide consumables for it. Also, dried blood spot methods are being developed for blood collection at sites where no special storage facilities are available, e.g. in developing countries.

Several volumes of the multivolume series *Methods in Enzymology* include a series of specific chapters on selected vitamin analysis methods, with emphasis on practical details and Ball[(1)] has extensively summarized and reviewed the methods that are available for the

analysis of vitamins in foods. There are a number of other useful reviews.[2–16]

All vitamin assays require careful validation and quality control (QC). There are external quality assurance schemes (EQAS) now available for some of the vitamins, mostly for blood (status) assays, and there are also a limited number of biological materials available with assigned values for vitamins to assist with method validation. However, all analytical laboratories will also need to maintain quality assurance (QA) procedures to ensure the long-term stability of their methodologies. This is usually achieved by storing pools of relevant biological sample materials, if possible with three or four different concentrations of each analyte, covering the entire analytical range usually encountered, subdivided into small identical aliquots and stored at very low temperature to ensure long-term stability. By including these QA samples at frequent intervals throughout the course of an assay series, it is possible to assess the extent of drift and of instability in the assay procedures, including the extraction stages (which may be at least as critical as the final analytical run, since the latter is usually calibrated by pure standards). Another important precaution is to add, at as early a stage as possible, a known amount of an internal standard or standards that are similar to the endogenous vitamin(s) but are entirely absent from the sample being analyzed, and which can be quantitated in the same analytical run as the required vitamin(s). These internal standards can then be used to correct for losses during purification, provided these are not excessive. Measurement of the percentage recovery of pure added vitamin gives a further indication of the overall efficiency of the extraction procedures and controls for the presence of interfering effects of matrix components in the assay. An even more powerful technique is to add an isotope (radioactive or stable) of the same vitamin(s) that are being measured and to use the resulting isotope-recovery information to correct for extraction- and purification-related losses. With the increasing availability of MS as an analytical tool, the use of stable isotope-labelled material as an internal recovery standard is becoming increasingly attractive. The formation of characteristic fragments by the breakdown of organic molecules during electron or chemical ionization bombardment in the mass spectrometer, promises to simplify the analytical procedure still further, by reducing the need for complete physical isolation of each organic species.

One important aspect of the analytical program which can only be touched upon in the present article is the principle by which the most appropriate sampling procedures are defined and developed. For foods, some important considerations are the choice of a representative subsample or the sampling of different portions of the entire sample so as to allow for sample heterogeneity. For status assays a choice must often be made between whole blood, serum, plasma, washed red cells or some other cellular component of blood, or a urine sample, which may (or may not) be a timed sample, e.g. 24 h. The blood sample may (or may not) be a fasting sample. Serum, plasma or urine concentrations tend to reflect recent vitamin intakes; blood cellular concentrations usually reflect longer-term intakes and long-term status. For some vitamins, secondary or functional indices such as enzyme activities, or the extent of reactivation of partially cofactor-depleted enzymes by their specific cofactors, may be used to measure and define status, and there is a range of functional status assays that cannot be included in this article (but for which see Bates).[16] Another important question, and a possible pitfall for the unwary, is the variety and choice of different units for reporting concentration. In the case of blood concentrations most journals now insist on SI (Système International d'Unités) (molar) units, but the older literature may also include values expressed as weight/volume, or a variety of non-standard units such as international units (IUs), which require separate definition for each vitamin. The interconversion factors for some of these are given in Table 3. In the case of food analysis the vitamins are generally expressed by weight per 100 g or weight per kilogram of fresh or cooked food. For food analysis it is also essential to define whether the food is cooked or raw, and whether or not it includes the inedible parts, such as the stones of fruit or the bones of meat and fish and so on. It may also be necessary to specify whether the concentrations are expressed on a wet or a dry weight basis. These and

Table 3 Conversion factors for IUs of fat-soluble vitamins[a]

Vitamin	Form	IUs per microgram (IU μg^{-1})
A	Retinol	3.33
	Retinyl acetate	2.91
	Retinyl palmitate	1.82
	β-Carotene	0.56
E[b]	D L-α-Tocopherol acetate (RRR)	1000
	D-α-Tocopherol acetate (RRR)	1360
	D-α-Tocopherol (RRR)	1490
	D-β-Tocopherol (RRR)	750
	D-γ-Tocopherol (RRR)	150
	D-δ-Tocopherol (RRR)	50
	D-α-Tocotrienol	450
	D-β-Tocotrienol	80
D[c]	Cholecalciferol (vitamin D_3)	40
	Ergocalciferol (vitamin D_2)	40
	25-Hydroxyvitamin D	ca. 200

[a] See Hall[1] and Bender[38] for further details of the interconversion factors.
[b] There are a large number of isomeric forms of vitamin E; only the more common ones are listed here.
[c] Vitamins D_3 and D_2 are equipotent for humans, but not for other mammals.

other potential problems in interpreting and comparing vitamin status indices are further discussed in section 5.

2 HISTORY

During the latter half of the nineteenth and first half of the twentieth century, most of the vitamins that are now recognized as being essential dietary components were isolated, crystallized, identified and then synthesized, and the methods used were the standard organic chemistry techniques that were then available or being developed, such as solvent partitioning, fractional crystallization, paper and thin-layer chromatography (TLC), GLC and so on. For some of the vitamins, physicochemical techniques (such as fluorimetry) and the formation of chemical derivatives with characteristic absorption or fluorescent spectra, were used for the identification and analysis. With the opportune discovery that many of the vitamins (especially in the B-group) were essential for the growth of particular microorganisms in purified growth media, it became possible to select particular species and strains of microorganisms which, with defined media which lacked a particular vitamin, could be used to quantitate that vitamin in liquid samples or in tissue extracts of biological origin. Gradually a wide range of assay techniques became available, which were tailored to the analysts' particular requirements, so that the food vitamin assay procedures then gradually diverged from those used for status assays or for pharmaceutical preparations. A brief selection of some historically important methods for the individual vitamins is listed below.

For the fat-soluble vitamins (A, E, D, K) and for vitamin C, the bioassay methods that were developed were in animals, generally based on growth stimulation in rodents such as rats or guinea pigs, or in chickens. Bioassay procedures for vitamin A,[17] E,[18] D,[19] and K[20] were of this type; that for vitamin D depended on X-ray pictures of bone growth (line test) and that for vitamin K depended on tests of the efficiency of the blood clotting process. Early procedures for the spectrophotometric assay of vitamin A in tissue extracts depended on the use of Lewis acids, such as antimony trichloride or trifluoroacetic acid,[21] which gave a transient colored product. Vitamin E could be assayed by virtue of its ability to protect against oxidative haemolysis of red blood cells.[22] Vitamin K status can be investigated by a variety of tests of the integrity of the blood clotting cascade which constitute specific functional tests of vitamin K status.[23] Vitamin C was originally measured by animal assays,[24] but a number of semispecific chemical assays were also developed at an early stage, based on the powerful reductive potency of vitamin C, or the formation of fluorescent derivatives or of colored osazones, after oxidation.[25]

Most of the B vitamins can be measured by specific microbiological assays,[26–33] and many of these are still in use, especially for food analysis, because they can usually cope with a multiplicity of different forms of each vitamin. However, the extraction and assay conditions require careful attention to detail[1] and there remain some serious problems of disagreement between different laboratories about the procedures.[30] For some of the B-vitamins, fluorimetric assays have been developed.[34,35] and there are protein-binding RIAs for folate and vitamin B_{12}.[36,37] Some of these have been developed for status assays, others are more suitable for foods and feeds. As clinical chemistry laboratories have developed a need for vitamin status assays, commercial kits have been developed, whose emphasis has been on simplicity, speed of use and high throughput, in contrast to the older assays, which were slow and time-consuming (but which in some cases were nevertheless more reliable than their modern counterparts).

3 FAT-SOLUBLE VITAMINS

3.1 Vitamins A, E and Carotenoids in Serum, Plasma and Breast Milk

Most clinical and epidemiological investigations of vitamin A status rely on the direct estimation of vitamin A in serum or plasma. Because many of the vitamin A assays that have been developed recently can also measure vitamins E and carotenoids, this group of analytes is generally measured in a combined assay, using high-performance liquid chromatography (HPLC) to separate the individual components and a variety of detection methods (optical density, fluorescence or electrochemical reaction: Table 4). Serum or plasma vitamin A concentrations reflect body stores only when the latter are severely depleted (liver concentration $< 70\,\mu mol\,kg^{-1}$, plasma concentration $< 0.7\,\mu mol\,L^{-1}$). When the liver stores are greater than $70\,\mu mol\,kg^{-1}$, plasma concentrations are largely independent of them. Serum or plasma concentrations that are considered to be either low ($<0.7\,\mu mol\,L^{-1}$) or deficient ($<0.35\,\mu mol\,L^{-1}$) occur more commonly in young children than in adults, and most commonly in poorer countries, where good dietary sources of preformed vitamin A from animal foods, or dietary supplements and of rich sources of precursor carotenoids from fruit and vegetables are unavailable. Poor maternal status can arise during pregnancy and lactation in those countries where dietary sources are inadequate, and the concentration of vitamin A in breast milk (expressed as a ratio to the fat content) may

Table 4 Currently used vitamin status assays

Vitamin	Concentration assay	Sample type	Functional assay	Sample type	Comments
A	Retinol by HPLC (optical density or fluorescence detection)	Serum or plasma	Relative dose response test; Conjunctival impression cytology	Serum or plasma. Goblet cells from conjunctiva	Most clinical and epidemiological studies measure plasma concentration
Carotenoids (including several vitamin A precursors)	Individual carotenoids by HPLC (optical density, fluorescence or electrochemical detection)	Serum or plasma	N/A	–	Usually now included with vitamin A assay
E	Tocopherols by HPLC (optical density, fluorescence or electrochemical detection)	Serum or plasma	Oxidative lysis of erythrocytes	Erythrocytes	Concentration assay is usually combined with vitamin A and carotenoids
D	25-Hydroxy vitamin D by HPLC or RIA	Serum or plasma	(Seldom used)	–	RIA available as a commercial kit assay
K	Phylloquinone and menaquinones by HPLC (fluorescence or electrochemical detection)	Serum or plasma	Blood clotting assays; PIVKA	Whole blood; plasma	Concentration assay confined to specialist laboratories. Clotting assays are less complex but also less specific
C	Ascorbic and dehydroascorbic acids by HPLC (various detectors) or by direct assays w/o separation	Serum or plasma; buffy coat	N/A	–	Buffy coat assay is a better index of body stores, but is more complex and seldom performed now
B_1 (thiamin)	Total thiamin by direct fluorimetric assay, or HPLC with fluorescence detection.	Urine	Transketolase (activation coefficient)	Erythrocyte lysates	Transketolase is most commonly used but is difficult to standardize
	Thiamin diphosphate by HPLC.	Red cells			
B_2 (riboflavin)	Total riboflavin by direct assay or HPLC; fluorimetric detection	Urine	Glutathione reductase (activation coefficient)	Erythrocyte lysates	Glutathione reductase is most commonly used
B_6 (pyridoxine)	Pyridoxal phosphate or total B_6 by HPLC	Plasma	Aspartate amino-transferase (activation coefficient)	Erythrocyte lysates	Wide variety of assays used; no general consensus on preference
	Pyridoxic acid: HPLC	Urine			
Folate	Total folate	Serum or plasma	Various (e.g. polymorph lobe counts, formimino glutamate, homocysteine)	Various	Serum folate is most commonly used; red cell folate is a better measure of stores
	Total folate	Red cells			

(*continued overleaf*)

Table 4 (*continued*)

Vitamin	Concentration assay	Sample type	Functional assay	Sample type	Comments
B_{12} (cobalamin)	Vitamin B_{12} by protein-binding or microbiol assay	Serum	MMA	Serum or urine	Combination of serum B_{12} and MMA is recommended
Niacin	Niacin metabolites (N^1-methyl nicotinamide; pyridones) by fluorimetry/HPLC.	Urine	N/A	–	Not often required, therefore not widely available
	Pyridine nucleotides in whole blood	Whole blood			
Biotin	Total biotin by protein-binding or microbiological assay	Serum	Lymphocyte propionyl CoA carboxylase (activation coefficient)	Lymphocytes	Not often required, therefore not widely available
Pantothenate	Total pantothenate by microbiological assay	Serum	N/A	–	Not often required, therefore not widely available

MMA, Methyl malonic acid; PIVKA, protein induced by vitamin K absence.

be a better index of maternal intakes and status than blood levels are.[39] Unfortunately the breast milk content varies considerably with stage of lactation, so its interpretation can be complex. Because vitamin A can be toxic at very high intakes (e.g. more than ten times the recommended intake taken regularly, or more than a hundred times the recommended intake taken intermittently), there may also be a need to investigate plasma or serum levels for evidence of overload. In this situation the concentration of retinyl esters, especially in the chylomicrons, is greatly increased so that the ratio of retinyl esters to retinol (which is normally very low), rises dramatically.

Vitamin E occurs in blood plasma or serum in a variety of different forms (reflecting the varying composition of different dietary sources), of which all-rac D-α-tocopherol is the most biologically potent (Table 3), and is usually the most prevalent form. Although modern HPLC-based (and similar) assay procedures can easily separate and quantitate each of the individual forms of the vitamin, this degree of detail is seldom necessary for the definition of normal ranges. However, it has been found to be preferable to judge status on the basis of vitamin E: lipid ratios rather than of vitamin E concentrations per unit volume, because of the strong covariance of vitamin E with lipid levels. Severe deficiency is associated with vitamin E concentrations below 1.1 μmol g^{-1} total plasma lipid, or below 2.2 μmol mol^{-1} cholesterol, but it has not yet proved possible to define an optimum concentration for long-term protection against peroxidative damage. Unlike vitamin A, vitamin E concentrations are not constrained by a specific binding protein and plasma levels are approximately proportional to intakes over a wide range. Overt toxicity has not been reported, even at very high intakes.

Like vitamin E, plasma carotenoid concentrations are approximately proportional to dietary intakes over a wide range; they are concentrated in the lipidic components of plasma or serum and most of them are generally considered to be nontoxic over a wide range of intakes. α- and β-Carotenes and the cryptoxanthins are partially converted to vitamin A within the intestinal wall and in other tissues, and are therefore useful contributors to vitamin A stores, especially for people whose diets do not contain animal products or dietary supplements, and therefore lack preformed vitamin A. Other dietary plant carotenoids, such as lycopene and the xanthophylls (lutein, zeaxanthin) are not precursors of vitamin A in humans, but most of the carotenoids are thought to possess useful antioxidant properties and are therefore of potential health significance with respect to the limitation of free radical-generated tissue damage. As noted earlier, they are not classified as vitamins per se, but because some of them are vitamin A precursors, because they are potentially protective, because they can provide an index of fruit and vegetable consumption and because they can be measured in conjunction with vitamin A and E, they are generally included in the assay procedures and protocols for these vitamins, especially in the context of epidemiological studies of population vitamin status.

Examples of modern assay procedures for vitamin A in serum, plasma, milk and tissues can be found in the literature.[40–73] Some of these also include simultaneous procedures for vitamin E and/or the carotenoids. Other aspects of the analytical methods and status determination issues for vitamin E have been

described.[74–90] Reviews that concentrate particularly on these vitamins are by Davidek and Velisek,[4] De Leenheer et al.,[6] and Furr et al.[52] In the author's laboratory, a minor modification of method given by Thurnham et al.[47] has been in constant use for several years, and has proven to be robust, precise and accurate for the analysis of vitamins A and E and seven carotenoids in human plasma samples for survey work.

Some interesting new developments, which have not yet been widely adopted, include the use of retinoic acid as an alternative index of vitamin A status,[41] the use of oxidation product(s) of vitamin E as an alternative index of vitamin E states,[74] the use of CE for the separation of fat-soluble vitamins before detection[50,53,91] and the use of MS to assist in the specific recognition of these vitamins or the use of stable isotopes for recovery assessment.[44,60,62,68,69,73,82,92] Vitamin E and its oxidation products have been assayed in the cellular components (membranes) of the blood.[77,80]

In the USA, there is an EQAS for these analytes, run by the National Institute of Standards and Technology (NIST) (Micronutrients Measurement QA Program, NIST, Gaithersburg, MD 20899, USA; fax: +1 301 977 0685; e-mail: jbthomas@nist.gov). In the UK, an equivalent scheme is run by Ms Maureen Goodall, tel: +44 981 296 2804; fax: +44 181 641 2633.[a]

3.2 Vitamins A, E and Carotenoids in Foods and Pharmaceuticals

The assay procedures used for these vitamins in foods overlap considerably with those used for status assays, the main differences residing in the extraction procedures, which are necessarily more complex than those needed for blood analysis, partly because the food matrix is much more variable and complex, and partly because the liberation of the free vitamins from their fatty acid esters by saponification is usually an essential part of the food analysis procedure, whereas in blood serum or plasma, most of them are present in unesterified forms.

Bioassays for these nutrients were the precursors, and in a sense the gold standards, for the later physicochemical-type assays. They bridged the gap between the functions of the vitamins, as essential components of the diet (of humans and animals) for their survival and health, and the identification of their unique chemical structures and detailed biochemical functions, which permitted the development of more sophisticated and more rapid assays, based on their physicochemical properties. For vitamin A, and for its precursor carotenoids, the carotenes and the cryptoxanthins, the animal bioassays depended on growth rate (in rats), a vaginal smear test (correction of keratinization in vaginal smears in ovariectomized rats) and a liver vitamin A accumulation or storage assay, which measured the amount of vitamin A accumulated after a predetermined dietary amount of the test food.[1] These bioassays were able to test the relative potencies of different vitamin A precursors, but only with respect to their potencies in rats, which may differ from those in humans. Analogous bioassays have also been used for vitamin E. Three approaches that have been used successfully are the foetus resorption test in vitamin E-deficient rats, an erythrocyte peroxidative haemolysis test in deficient rats and a liver-storage test in rats or chicks.[1]

Saponification of food homogenates before extraction is classically performed by heating with potassium hydroxide, frequently with the additional of methanol or ethanol (which helps to denature the proteins) and frequently with antioxidants such as pyrogallol, butylated hydroxytoluene (BHT) and so on. Extraction uses nonpolar solvents such as diethyl ether, hexane, tetrahydrofuran, acetonitrile and so on.

HPLC separation of vitamin A, E (multiple forms) and carotenoids in food extracts is similar, in principle, to that described in section 3.1, except that the range of components is likely to be more complex. Gradient elution with complex mixtures of solvents may be needed. Several examples[49,71,75,93–106] may be added to those listed elsewhere.[1,4,5,11]

The assay of vitamins A, E and carotenoids in pharmaceutical preparations is similar in principle to that in foods, but because the concentrations are higher and the complexity of the material is less, the analytical task is less demanding. HPLC is the preferred technique, but a simpler assay, based on optical density or fluorescence, may be feasible for some types of preparation.[67,106]

3.3 25-Hydroxy and 1,25-Hydroxyvitamin D in Serum, Plasma and Breast Milk

Vitamin D shares with several other vitamins the property that a dietary source of the preformed vitamin is not the only biological source of the vitamin. It is, however, the only vitamin that is formed by the action of sunlight on a precursor within the skin, and it is the extent of exposure to ultraviolet (UV) light (of a limited range of wavelengths), rather than the amount of the precursor present, which is critical for the amount of de novo formation. Because of this, wide natural variations in the amount of exposure to sunlight (latitude, season and clothing) are critical in determining whether sufficient vitamin D is produced de novo, or whether food sources are essential, in order to achieve adequate status and freedom from the classical vitamin D deficiency diseases: rickets and osteomalacia. Status is measured

[a] All contact details correct May 1999.

biochemically by blood analysis and there are two principal and important circulating forms of the vitamin: 25-hydroxyvitamin D and 1,25-hydroxyvitamin D. The 25-hydroxy form reflects the sum of the contributions to the vitamin from food and from the action of UV light on the skin, and this is the analyte that is normally used to define biochemical status. However, the 1,25-dihydroxy form is the active functional (hormone-like) metabolite, whose concentration indicates biological efficacy, so that measurement of this metabolite may also be necessary if there are biological or pathological reasons why the interconversion pathway(s) might be disrupted. The principles behind these metabolic processes and the practical choices for the status analyst, are discussed in greater detail in several reviews.[107–111] Although the measurement of vitamin D per se is important for food analysis and for the study of tissues other than blood, its value as a clinical status index is limited.

There are many published methods for 25-hydroxy-vitamin D based on HPLC. A selection of publications is listed.[112–117] Somewhat simpler procedures, with a more rapid throughput rate, have been developed on the principle of specific binding proteins or specific antibodies to synthetic analogues of the naturally occurring vitamin, which generally require a radioactively labeled form of the vitamin as a marker of the extent of binding and hence of the concentration of the unlabeled analyte in the biological sample extract.[118–129] Although the outcome of independent assessment of RIA-based commercial kit assays has not always been favorable,[130] in the author's laboratory this choice of assay has proved adequate for survey work requiring long-term reproducibility and procedural simplicity.[129] There have been several studies using MS[131,132] and some studies of hydroxylated derivatives of vitamin D other than the well-known 25-hydroxy and 1,25-dihydroxy vitamin D.[115,133] An EQAS in the UK for serum or plasma is run by Mr Graham Carter, DEQAS, QA Working Party, Charing Cross Hospital, London W6 8RF, tel: +44 181 846 7074; fax: +44 181 846 7009; email: g.carter@cxwms.ac.uk.

For lactating mothers, the concentration of vitamin and its hydroxylated metabolites varies with the status of the mother. The measurement of breast milk concentrations[39,116,134] can therefore provide additional status information.

3.4 Vitamin D and its Derivatives in Foods and Pharmaceuticals

There are two main forms of vitamin D in foods: vitamin D_3 (cholecalciferol), which is derived (in vertebrates) from the action of sunlight on 7-dehydrocholesterol, and vitamin D_2 (ergocalciferol), which differs only in the presence of a double bond at C-22–C-23. Vitamin D_2 is derived from ergosterol, which is found in plants, fungi and invertebrates. There are smaller amounts of the 25-hydroxy derivatives of those compounds, present in some animal tissues, and there may be significant amounts of previtamin D_3, which is the intermediate between 7-dehydrocholesterol and vitamin D_3 and possesses about half of its antirachitic activity in experimental animals.

The older, but still potentially useful, bioassay for vitamin D activity of foods is the ''rat line test'', which measures the potency of the test substance to induce new calcification in the tibia, radius or ulna of a vitamin D-deficient rat, and hence to form a deposit of silver after staining with silver nitrate (the silver nitrate is converted to silver phosphate by reaction with the deposited calcium phosphate, and then to metallic silver by photochemical reduction).

Modern assay procedures depend on HPLC after saponification (if necessary, to remove triglycerides), extraction into a nonpolar solvent, evaporation and re-solution in the mobile phase. Preliminary cleanup procedure(s), such as solid-phase extraction, are likely to be necessary before the analytical separation and quantitation. Both normal-phase and reversed-phase HPLC systems have been used; representative selections of these have been reviewed.[1,5,11] As with the other fat-soluble vitamins, pharmaceutical preparations can be assayed by similar procedures, but simpler alternatives may be appropriate.[67,106]

3.5 Vitamin K in Serum, Plasma and Breast Milk

Until quite recently, the technical difficulties associated with the measurement of vitamin K (phylloquinone and menaquinone) contents of body fluids were such that status measurements had to depend on indirect measurements, or ''functional'' indices, such as the efficiency of blood clotting in vitro, or the accumulation of undercarboxylated blood proteins which were specifically caused by a lack of vitamin K (e.g. PIVKA), and most recently by undercarboxylated osteocalcin.[135] These are extremely sensitive and valuable status indices, but for some purposes, such as the comparison of dietary vitamin K intakes with blood vitamin levels, it has become necessary to develop the analysis of vitamin K per se in serum, plasma and breast milk. (The last was especially important because breast-fed babies are at relatively high risk of vitamin K deficiency, and a variety of interventions have been devised to decrease this risk, including the provision of vitamin K supplements for nursing mothers to raise their breast milk levels.)

Because of the low concentration of vitamin K in body fluids and the lack of a specific and sensitive physical property, chemical derivatization process or RIA, its quantitation required the development of highly efficient HPLC assays, usually with at least two sequential

separation stages, to achieve an adequate purity and a high enough concentration for accurate quantitation by fluorimetry or electrochemical detection.[135–146] Some of these procedures have focused specifically on vitamin K_1 (alias phylloquinone: the principal form of vitamin K that is present in plant foods); others have included the measurement of the series of menaquinones from bacterial sources that share vitamin K activity, although the relative biopotencies are not known precisely.[137,139,147–149] Phylloquinone has also been quantitated by a stable isotope-dilution gas chromatography/mass spectrometry (GC/MS) assay.[150] Tissue assays have been developed for experimental metabolic studies in humans and animals.[136,147,149,151,152] After many years in which its relevance for human nutrition was largely ignored, the importance of adequate intake and status has recently received attention, which is reflected by the greatly increased numbers of publications.

An EQAS for vitamin K assays has recently been made available in the UK: Dr Alison Cox or Dominic Harrington, Haemophilia Centre, St Thomas' Hospital, London SE1 7EH, tel: +44 171 928 9292; fax: +44 171 401 3125.

3.6 Vitamin K in Foods and Pharmaceuticals

Assays of vitamin K in foods should include both phylloquinone and the menaquinones of bacterial origin. Since good physicochemical assays for this vitamin have only recently become available (and most food table compilations do not yet have values for vitamin K because its content in many foods is not known accurately), the availability of a bioassay has been critically important. The most commonly used procedure depends on the increase in blood clotting time that is observed in young chicks after feeding a vitamin K-deficient diet, and the fact that test substances containing vitamin K will reduce this clotting (prothrombin) time, towards the normal value.

Because vitamin K is destroyed by strong alkali, the removal of triglycerides before extraction requires enzymatic hydrolysis. Vitamin K can be extracted into nonpolar solvents or into supercritical (liquid) carbon dioxide. HPLC can be of the normal or reversed-phase type and the preferred detectors are of the fluorescent or electrochemical type, since these are both more sensitive and potentially more specific than the standard optical density type for substances like vitamin K. Examples of procedures developed between 1988 and 1998 have been described.[1,5,11]

4 WATER-SOLUBLE VITAMINS

4.1 Vitamin C in Serum, Plasma, White Cells, Urine and Breast Milk

In principle, the measurement of vitamin C concentrations and status should be easier and more straightforward then that of some other vitamins, because it is present at higher concentrations and has a characteristic reducing activity, which can be harnessed on the basis of colorimetric assays linked to redox dyes. In practice, however, the task of devising accurate and reliable vitamin C status assays has proved difficult, partly because of its rapid oxidative degradation in solution and partly because of interference by other components of body fluids, tissues, foods and so on. Since the early studies on vitamin C and scurvy which relied on animal bioassays, there have appeared dye-reduction methods, methods linked to transition-metal-ion reduction, methods linked to the formation of colored osazones with dinitrophenylhydrazine, methods linked to the formation of fluorescent adducts with orthophenylene diamine, HPLC methods with optical density, fluorescence or electrochemical detection, GC of volatile derivatives and GC/MS methods.[153,154] With the advent of user-friendly HPLC apparatus, this approach now tends to be favored[155–174] because it is reasonably specific and is less subject to interference, especially if combined with a relatively specific detector, such as the newer electrochemical or coulometric devices.[156,161,162,164,166,168,170,173,175]

Because of the problem of rapid degradation, it is essential that vitamin C be preserved efficiently in the samples that are collected, either by acidification with a chelating acid or by adding a reducing agent, which must not interfere with the assay-detection system. A variety of alternative procedures have been described.[160,167,176–179] With care, it is possible to preserve samples for more than a year at low temperatures and with suitable stabilizing agents.

In the author's laboratory, a fluorescence-based assay in a centrifugal analyzer[180] has proved satisfactory for plasma vitamin C analysis in survey samples, and this technique compares consistently well with more time-consuming HPLC-based assays. Vitamin C assays have begun to utilize CE[181–183] and GC/MS is also proving to be a powerful alternative for certain specific tasks and for method validation.[154,184,185] The NIST QA scheme, described for vitamins A, E and carotenoids, also runs a plasma vitamin C QA scheme.

Today, vitamin C status is usually measured by serum or plasma concentrations, which vary between undetectable levels in depleted subjects, up to concentrations around 100 μM in people with a generous intake. For several decades, however, there was a preference for the use of buffy coat (total white cell) vitamin C, which was shown, in animal models, to be a more accurate reflection of whole body status. This has now fallen into disfavor, partly because the fieldwork procedures are complex and time-consuming and partly because the result is very sensitive to variations in the proportions of the different types of white cells in the buffy coat, which in turn is affected by infections and so on. Serum or plasma

concentrations of vitamin C (and of other water-soluble vitamins) are sensitive to recent dietary intakes of rich food sources or supplements, but this potential problem can be minimized by collecting the blood sample after an overnight fast. The concentration of vitamin C in urine is less useful as a status indicator, partly because of instability (which varies with pH and other urinary constituents), and partly because it varies with variations in kidney function. Breast milk vitamin C concentrations do vary with variations in status,[39] but they are rarely used to investigate it, partly because the range is narrower than that in plasma.

4.2 Vitamin C in Foods and Pharmaceuticals

There are several potential problems that can complicate the assay of vitamin C in foods. First, some foods in some countries may contain the unnatural isomer of L-ascorbic acid: D-erythorbic acid or D-isoascorbic acid, which may be added, especially to cooked meats, as a color preservative and to reduce the formation of carcinogenic nitrosamines from the nitrite (which is added as an antimicrobial agent). The D-isomer of vitamin C has very similar chemical properties to those of ascorbic acid, but it has little of its biological potency because it is poorly absorbed and is rapidly excreted. Therefore, a separation of the two isomers (generally by HPLC) is necessary if the D-isomer is present. Second, as noted earlier, ascorbic acid is easily oxidized, especially in the presence of transition-metal ions; therefore its extraction into a preservative medium and subsequent analytical steps must be performed with careful attention to antioxidant precautions. Most extracts of foods contain a significant proportion of the oxidized but biologically active form of the vitamin, dehydroascorbic acid, which may be present in the native food per se, or which may be formed by oxidation during the extraction procedure. This must be included in the final assay. There are many variants of food vitamin C assays described in the literature. Rich sources of the vitamin without interfering substances can be assayed by several relatively simple physicochemical assays (e.g. the measurement of dye reduction), but less-rich sources generally require more complex assay procedures.

Bioassay procedures for vitamin C are cumbersome and time-consuming: they are generally based on the rate of maturation of the odontoblast cells in the incisor teeth of guinea pigs who have been receiving a diet which is deficient in vitamin C. It is necessary to use guinea pigs, because most other species, (other than humans and the higher primates), can synthesize vitamin C de novo from sugars and do not require a dietary source of the vitamin.

For physicochemical assays of the vitamin, extraction into a chelating acid is required and this should typically be metaphosphoric acid, or a mixture of metaphosphoric and acetic acids. For those extracts that can then be assayed without further separation, the choice lies between the classical reaction with 2,6-dichlorophenol indophenol, or a related dye, which is stoichiometrically decolorised by the reductive action of vitamin C, provided that all of it is present in the reduced form. A wide variety of reducing agents cause positive interference in this assay, however. Another relatively crude assay, which is only adequate for some fruit drinks or pharmaceutical preparations, is the measurement of the optical density at 267 nm (pH 6.0), before and after the copper-catalyzed oxidation of the vitamin. A more specific fluorometric assay, which has the advantage that it includes both ascorbic and dehydroascorbic acids because the former is deliberately oxidized to the latter, depends on the reaction of dehydroascorbate with *o*-phenylene diamine to give a fluorescent quinoxaline (see section 4.1). Several methods have been used to convert ascorbic acid to dehydroascorbic acid in this assay; the most specific is ascorbate oxidase enzyme, isolated from certain plants. A semiautomated version of this assay was adopted by the Association of Official Analytical Chemists (AOAC) for the determination of vitamin C in foods in 1990.[186]

HPLC has undoubtedly become the most reliable procedure for vitamin C analysis, especially as it is the only technique (of those that are widely available) that can distinguish between the D- and L-isomers of the vitamin. In addition, it is possible to measure dehydroascorbate by running the assay with and without a reducing agent, used to convert the dehydro form to the reduced form. Detection can be achieved by optical density, or preferably by electrochemical (or coulochemical) detection, the latter being considerably more sensitive and specific. Column types can be weak anion exchange, or ion exclusion, or reversed phase, preferably with an ion-pairing agent, usually an organic amine, with suitable precautions to avoid the precipitation of metaphosphate salts on the column. A large number of alternative procedures have been described.[1,11,175] In addition, CE methods have been developed,[1,181,183] and GC/MS has also been described.[187] The use of guaicol peroxidase as an alternative to ascorbate oxidase for the specific oxidation of ascorbic acid in a crude spectrophotometric assay has been described and a variety of flow injection analyses, some of which employ immobilized enzymes, have been reported.[1] These and other simplified procedures.[67,104] are likely to be suitable for many types of pharmaceutical preparation with relatively high concentrations of the vitamin.

4.3 Thiamin (Vitamin B_1) Indices in Blood, Urine and Breast Milk

For many years, the assessment of thiamin status was principally by urine analysis,[2] and was based on the

principle that thiamine could be oxidized completely to the fluorescent derivative, thiochrome, by alkaline ferricyanide treatment. This provided a reliable and sensitive measure of status in individuals and populations with relatively generous intakes of the vitamin, because the excess of intake over requirements was almost quantitatively excreted in the urine. In severely depleted or partially depleted subjects, however, the test was less discriminatory. For this and other reasons, including the difficulty of collecting an accurately timed urine sample, the enzyme reactivation test based on red cell transketolase activity was developed by Brin, described by Sauberlich et al.[2] and Bates.[16] This depends on the principle that with progressive tissue thiamin depletion, certain thiamin cofactor-dependent enzymes such as red cell transketolase become progressively depleted of their essential cofactor and therefore lose enzyme activity. Addition of the missing cofactor in vitro restores at least some of the lost activity. The extent of depletion and hence tissue status can be estimated in terms of the ratio of reactivated to nonreactivated enzyme activity, in vitro. The advantages of this approach are that it is highly sensitive to status variations in the critical range of moderate-to-severe depletion, and because it depends on a ratio of two enzyme activities, any procedural variations that may affect the absolute activities (pH, temperature, assay components and so on) are less important than they would be for activity measurements per se.

Although the measurement of thiamin status by the red cell transketolase activation coefficient has been widely used especially in population surveys,[188] it does have some disadvantages, notably that the cofactor-depleted apoenzyme is slowly destroyed in vivo, that the red hemolysates needed for the assay need careful preservation before the assay and that there are no EQAS, or well-established validation procedures, for the assay. Direct thiamin assay procedures have been improved and made gradually more sophisticated since the late 1970s.[189–205] Many of these have made use of recent advances in HPLC[191,194–205] and the measurement of blood or red cell concentrations of thiamine phosphate esters has thus become an attractive alternative to the existing status assays.[191,193,194,197–205]

The thiamin content of breast milk does not reflect maternal status to a degree that can make it useful as a status assay,[39] and other functional assays of thiamin status have generally proved too cumbersome and not sufficiently reliable for widespread adoption.

4.4 Thiamin (Vitamin B_1) in Foods and Pharmaceuticals

Thiamin is present mainly as the cofactor, thiamin diphosphate or thiamin pyrophosphate (TPP) which is noncovalently bound to apoenzyme protein. It can be extracted without loss of vitamin activity at acid or neutral pH, but in alkali it is easily destroyed by oxidation. Indeed, one of the most commonly used assay procedures relies on its oxidation to the inactive product, thiochrome, by alkaline ferricyanide. It can also be inactivated by UV irradiation or exposure to sulfite, which cleaves apart the two halves of the molecule. Extraction from foods is achieved either by heating with dilute (0.1 N) hydrochloric acid, typically at ca. 108 °C for plant foods such as cereals, or at 121 °C for animal foods, or by treatment with wide-spectrum proteolytic enzymes, e.g. those of fungal (diastase) and plant (papain) origin.

The soluble extracts that are obtained by these extraction procedures can be assayed either by microbiological growth assays (*Lactobacillus fermentum* or *L. viridiscens*), or by fluorimetric assays, which may or may not require HPLC separation to achieve an initial purification of the vitamin. AOAC fluorimetric assays of thiamin in food extracts rely on open column chromatography followed by a simple extraction of the oxidation product, thiochrome, into butan-2-ol, before the measurement of its fluorescence. HPLC-based methods have been devised, which either separate thiamin and then convert it to thiochrome in a postcolumn reaction, or which oxidize the thiamin in a precolumn reaction and then separate the thiochrome before fluorimetric quantitation. Examples of these two alternative approaches are listed.[1,11,191,192,206–212] GC separation has also been employed.[4,11,213] Assay of thiamin in pharmaceutical preparations.[67,106] is relatively straightforward.

4.5 Riboflavin (Vitamin B_2) Indices in Blood, Urine and Breast Milk

As with thiamin, the earlier studies of riboflavin status and requirements depended on the measurement of urinary excretion, preferably in a timed urine sample.[2] This procedure has similar limitations to those encountered for thiamin, and as a test of status it has now largely been replaced by the red cell glutathione reductase test.[16,214] Like transketolase for thiamin status, glutathione reductase in red cells loses its riboflavin-derived cofactor, FAD, progressively during tissue riboflavin depletion, and the extent of this depletion is accurately reflected in the ratio of reactivated to nonreactivated enzyme activity in vitro. The activation coefficient, which is the ratio of reactivated to nonreactivated enzyme activities, increases from 1.0 (no difference) in a fully replete subject, to values well above 2.0 in severe deficiency. The normal range is usually deemed to be from 1.0 to 1.3 or 1.4, depending on the details of the assay procedure. This test has proven to be rugged and reliable; it is adequately discriminatory for population studies and it has been widely used and adopted. Nevertheless, the problems that exist for the

transketolase test are also true to some extent for the glutathione reductase test and there remains a need for concentration-type assays.

Riboflavin and its esters can be assayed by a wide range of techniques. Advances in urinary riboflavin assays have been described.[215–217] Other developments, such as luminometric assay[218] and competitive-binding assay, based on specific binding proteins have also been described.[219–222] HPLC assays[216,223–227] and enzyme-based assays have been described.[228,229] Clearly there is now a wide range of choices and the particular selection will depend on the particular application.

Unlike thiamin, the riboflavin content of breast milk could be a moderately useful indicator of maternal status,[39] although it has not often been used for this purpose.

4.6 Riboflavin (Vitamin B_2) in Foods and Pharmaceuticals

Riboflavin is relatively stable to heat and to variations in pH, except when exposed to light, when it is rapidly and irreversibly degraded. The product of light-catalyzed degradation in alkaline solution is the highly fluorescent compound, lumiflavin or 7,8-trimethylisoalloxazine, and this forms the basis of the majority of nonbiological assay procedures for riboflavin. In foods, the vitamin occurs as free riboflavin, as riboflavin phosphate (FMN) and as the more complex enzyme cofactor FAD. There are also forms of riboflavin that are covalently bound to certain enzymes, but these are generally unavailable as food sources of the vitamin.

Like thiamin, riboflavin is usually extracted from foods by hot dilute mineral acids such as 0.1 N hydrochloric acid, typically by autoclaving at 121 °C, which converts FAD to FMN. Enzyme preparations such as fungal diastase are used, especially if complete conversion of FMN to riboflavin is required.

Microbiological growth assays have been widely used for the quantitation of riboflavin in food extracts, typically *Lactobacillus casei* (rhamnosus) or *Enterococcus faecalis*. Removal of lipids may be necessary to avoid interference in the assay. The principal physicochemical approach to riboflavin assay relies on fluorimetry, with or without HPLC separation. For rich sources of riboflavin, simple cleanup procedures and a suitable assay blank will permit quantitation without chromatographic separation, but HPLC-based assays have been gaining acceptance.[1,11,206,212,224,230–238,239] Extracts are likely to contain both free riboflavin and FMN and these have different fluorescence response factors which must be taken into account. Riboflavin can easily be assayed by its fluorescence in extracts of pharmaceutical preparations.[67,106]

4.7 Vitamin B_6 Indices in Blood, Urine and Breast Milk

Vitamin B_6 occupies a somewhat paradoxical and atypical position, in that severe and overt deficiency is hardly ever encountered (it is not a characteristic of poor quality third world diets), yet it remains a potential problem for optimal nutrition, because evidence from several sources indicates that the ratio of this vitamin to other dietary components, particularly to protein, may be suboptimal for many human populations. A variety of assays for the vitamin itself or its degradation products (notably pyridoxic acid), and a variety of functional status assays have been proposed and developed.[16] Among the latter are the pyridoxal phosphate-dependent glutamate and aspartate aminotransferases that occur in red blood cells. Enzyme reactivation assays with pyridoxal phosphate in vitro have been developed which are analogous to the transketolase assay for thiamin status and the glutathione reductase assay for riboflavin status. In competition with this approach is the direct measurement of several forms of vitamin B_6, or of pyridoxal phosphate, in serum, plasma or whole blood samples.[239–252] Although HPLC separation and fluorimetric detection have been widely used for the measurement of vitamin B_6 in blood,[239–247,249,254] the use of a pyridoxal phosphate-requiring enzyme, stripped of its cofactor, such as tyrosine decarboxylase apoenzyme, has been used,[253] and offered as a commercial kit assay for the measurement of pyridoxal phosphate and hence of vitamin B_6 status. The use of microbiological assays for vitamin B_6, which were used for status measurements in early studies, and which remain a mainstay of food vitamin B_6 analysis, are not commonly used for status measurements now, being too cumbersome and incompatible with modern clinical laboratories.

Unfortunately there is poor agreement between the different indices of vitamin B_6 status, each of which is prone to a different set of confounding factors which can complicate the interpretation of vitamin B_6 status. There is some ongoing research that may resolve these problems. Meanwhile it is probably advisable to use a combination of several status indices if possible. Breast milk vitamin B_6 concentrations vary with variations in maternal vitamin B_6 intake,[39] but are seldom used to determine status.

There are a number of functional indices of vitamin B_6 status which focus on specific intermediates of amino acid turnover, but these are of research interest, rather than being in routine use, e.g. for surveys.

4.8 Vitamin B_6 in Foods and Pharmaceuticals

Vitamin B_6 occurs in several different forms in foods. In unfortified foods the principal forms are pyridoxal,

pyridoxal phosphate, pyridoxamine phosphate and pyridoxine glucosides; fortified foods may also contain pyridoxine. Some of these forms are tightly bound to protein, and the sensitivity of vitamin B_6 to light-catalyzed degradation requires the use of rigorous precautions to exclude short-wavelength light during the analytical procedures. A major degradation product, formed from pyridoxal phosphate in vitro is the fluorescent pyridoxic acid 5′-phosphate, which is biologically inactive. The pyridoxine glucosides that occur in considerable amounts in plant tissues have a lower biological potency than the nonglucosidic forms of the vitamin, further complicating the relationship between the analytical values and the vitamin potencies of native foods.

Generally, foods are first treated with mineral acids such as dilute hydrochloric acid, under autoclave conditions, to release vitamin B_6 from the food matrices. If more detailed information is required about the forms that are present in the native foods, then an extraction at low temperatures with protein denaturants such as trichloracetic, metaphosphoric acids and so on, may be preferable. Treatment with β-glucosidase can be used to release pyridoxine from the glucosides and to measure the percentage of glucosylated vitamin present in the food. Other enzymes, such as proteases, have also been used to achieve gentle release of the vitamin from the food matrix.

Microbiological growth assays of vitamin B_6 in food extracts have used yeasts, *Saccharomyces carlsbergensis* (uvarum) or *Kloeckera apiculata*, and these have been used alone, or in combination with partial cleanup procedures, or chromatography. Fluorimetry, with or without HPLC separation has also been used and a range of procedures have been described.[1,11,104,241,242,244,254] Procedures for pharmaceutical preparations are also available.[67,106]

4.9 Folate Indices in Blood, Urine and Breast Milk

The measurement of folate status, together with that of vitamin B_{12}, has for many years been a standard function of those routine hematological laboratories which include ''hematinic nutrient'' assays in their armamentarium. Both nutrients are related to megaloblastic (macrocytic) anemia, and are therefore relevant to the diagnosis of this and related hematological abnormalities. The importance of subclinical folate deficiency, and of folate status in general, has become an important public health issue, because of the discovery that it is linked to the risk of neural tube defects (spina bifida, anencephaly) during very early pregnancy, and that it may be linked to vascular disease risk, via its effect on circulating homocysteine concentrations.

Both serum (or plasma) and red cell folate concentrations can give status information. The serum (plasma) assay is the more widely used and is simpler to perform, but it is also more affected by day-to-day fluctuations. The concentration in the red cells, which is much higher than that in plasma, is fixed at the time of production of the cells in the bone marrow and is therefore an integrated average index, representing a four-month period. However, it needs to be measured in a hemolyzed whole blood sample (the conjugate enzyme in the plasma being needed to release the monoglutamate form from the intracellular polyglutamate complex); it requires a hematocrit measurement and the sample needs to be stored with a reducing agent such as ascorbic acid. Storage conditions for the samples require careful attention.[255]

Traditionally, folate has been measured by a microbiological (growth) assay, using *L. casei*. The older rather cumbersome nephelometric assays using test tubes have now been modified to make use of the technology of ELISA plates and their automated readers.[256–258] With the discovery of a specific folate-binding protein in milk, alternative assays have been developed on the competitive protein-binding principle and marketed as a range of commercial kit assays, which are widely used by laboratories performing routine clinical or population survey work.[259–264] The problems of interlaboratory harmonization of these and other routine assays have not yet been fully solved[261–263] and there is a need for a ''gold standard'' assay, perhaps based on MS.[265,266] However, the existence of EQAS for folate and B_{12} in blood samples gives some degree of QC for these assays, even though the different versions of the assay do not give identical results. Contact details for these QA schemes are: (in the UK) Dr M.J. Lewis, National EQAS (UK), tel: +44 161 720 2483; fax +44 161 720 2886; (in the USA) College of American Pathologists Survey (Illinois), tel: +1 847 832 7000; fax: +1 847 446 8807.

Blood and tissue folate concentrations and, in particular, the study of the many different forms of folate that occur in tissues, require efficient separation and sensitive detection techniques, such as those of HPLC. A range of alternatives is available[267–274] and research in this field is intense. Although urine is not generally a useful source of information about intact folate, folate degradation products[275] and in particular their use for stable isotope folate labeling (turnover) studies.[276,277] are being studied in urine.

Folate in breast milk is not a useful source of information about maternal status.[39] There are several functional indices of folate status, of varying degrees of specificity and usefulness. These include plasma homocysteine (which is strongly dependent on folate status, but is also affected by other B vitamins), the excretion of formiminoglutamic acid (FIGLU) after a standard histidine dose, the deoxyuridine suppression test, which measures the efficiency of the de novo

pathway of thymidine synthesis, and the nuclear lobe count of circulating polymorphonuclear leukocytes. All of these are research, as distinct from routine clinical or epidemiological, tools.

4.10 Folate in Foods and Pharmaceuticals

The assay of folate in foods poses a major challenge because of the multiplicity of forms that are present, their limited stability and the complexities posed by variable bioavailability, especially of the polyglutamate forms. The tetrahydro forms of folate present in foods are easily oxidized, and some of the oxidized forms are biologically inactive. For this reason, it is advisable to include a reducing agent, usually ascorbic acid, in the extraction medium, and to minimize exposure to high temperatures or pro-oxidant conditions. Because of the great complexity of the folate polyglutamates that occur in foods and tissues, it is common practice to simplify the structure by an enzymatic deconjugation to the monoglutamate forms during, or immediately after, the initial extraction process.[1,278–282] Some procedures use thermal denaturation of proteins, others rely entirely on combination of enzymes, e.g. proteases to liberate protein-bound folate, α-amylase to degrade starch and folate conjugates from various sources to deconjugate the folate polyglutamates.

The three main approaches to the final assay and quantification steps have been microbiological growth assays, generally with *L. casei*, radioassays, using a competitive protein-binding principle and HPLC. Microbiological assay has the potential advantage that, with the correct choice of conditions, it should be possible to achieve equal potency for all biologically active forms of folate in the extracts. In practice this has proved difficult[1] and there remains a serious degree of uncertainty about the reliability of the *L. casei*-based assays for food folate analysis. Radioassay-based techniques also suffer from similar problems, because it is very difficult to achieve and maintain assay conditions in which all (and only) the biologically active forms of folate have equal affinity for the chosen folate binder. Although some laboratories have used this approach,[1,282] it appears to have serious drawbacks.

A considerable number of HPLC-based assays have been described[1,11,267,268,281–283] and there is a useful review.[284] The use of GC/MS is finding important applications in this area.[266,267] The problem of monitoring pharmaceuticals is much less difficult than that of food analysis, because the only form that is usually present is pteroylglutamic acid, which is stable and easily measured.

4.11 Vitamin B_{12} Indices in Serum, Plasma and Breast Milk

Not only are folate and vitamin B_{12} closely linked metabolically by their involvement in one-carbon transfer reactions, they are also similar with respect to the choices of status assays that are available: namely microbiological, competitive protein binding and HPLC. Because undiagnosed vitamin B_{12} deficiency due to pernicious anemia or related conditions can cause irreversible neurological damage, it is imperative that early diagnosis is made; hence the development of several commercial kit assays for clinical use, generally for use in hospital hematology laboratories. Initially, these were based on the semispecific ''R-binders'', which failed to distinguish between vitamin B_{12} and some structurally related, but biologically inactive corrinoids. These were replaced by intrinsic factor-based binding assays, which possess greater specificity and have proved to be more satisfactory.[285–293] The most recent development in this field is the use of fluorescent or chemiluminescent end-points, rather than radioisotope-labeled ligands, for protein binding assays for both folate and vitamin B_{12}.[264,290,292,293]

More detailed investigations of vitamin B_{12} status and its metabolism have required HPLC[294] or GC/MS techniques.[295,296] Because serum vitamin B_{12} (the usual clinical status assay) may fail to detect a deficiency in some individuals and some metabolic states, additional measurements based on the circulating vitamin B_{12}-binding proteins[297] or a functional assay based on serum or urinary concentrations of MMA[298] have been developed. Raised levels of MMA provide a specific indicator of tissue vitamin B_{12} deficiency, which is becoming of increased usefulness now that the initial problems of accurate measurement by GC/MS or CE, have been overcome. Other functional assays for vitamin B_{12} are also shared by folate, e.g. raised plasma homocysteine, the deoxyuridine suppression test and increased polymorphonuclear leukocyte lobe counts. The Schilling test specifically tests for impaired absorption of vitamin B_{12} using radioactive cobalt-labeled vitamin B_{12} in vivo. Breast milk vitamin B_{12} levels do vary with maternal status[39] and this is of potential concern for those nursing mothers whose dietary intake is extremely low (e.g. vegans) or those with B_{12} absorption defects.

4.12 Vitamin B_{12} in Foods and Pharmaceuticals

Vitamin B_{12} in foods is, like folate, relatively susceptible to degradation during extraction; it may be destroyed by strongly acidic or alkaline conditions and it is also susceptible to strong oxidizing or reducing agents. Extraction is designed to liberate protein-bound forms of the vitamin and to convert them to the more

stable forms, cyanocobalamin or sulfitocobalamin. This is usually achieved by heating in buffers under mildly acid conditions: metabisulfite can be used to achieve conversion to the sulfito form.

Microbiological assay is achieved by growth assays using *L. leichmannii* (now *L. delbruecki* (*lactis*)), or by the protozoan *Ochromonas malhamensis*.[1,7,9,285] Radioassay procedures, usually based on intrinsic factor as the specific protein binder and ^{57}Co-labeled cyanocobalamin as the tracer, have been used successfully. HPLC-based methods and GC/MS-based methods are potentially available[294,295,299] but have not yet been widely used. The form that is present in pharmaceuticals is cyanocobalamin, which is stable and fairly easily quantitated.[67,106]

4.13 Niacin Indices in Blood, Urine and Breast Milk

Whereas the measurement of niacin status was a priority during the outbreak(s) of pellagra in the first half of this century, especially in the southern states of the USA, it has had a lower profile in recent years and the assay procedures have therefore not progressed. The most widely used status assays are those based on urinary excretion of its degradation products, and whereas the early assays were relatively crude and based on the formation of a fluorescent derivative of N^1-methylnicotinamide alone, more recent versions have separated and quantitated both the N^1-methyl derivative and the pyridone breakdown products, thus achieving a more reliable picture.[300–304] It has been suggested that an N^1-methyl nicotinamide: pyridone ratio less than 1.0 in urine may indicate a deficiency of niacin.[16] Breast milk concentrations are moderately responsive to maternal status.[39] Studies[303,305–307] have suggested that the concentration of niacin-derived compounds formed from niacin in blood fractions may give useful information about niacin status, and although this has not yet been developed into a routine status assay, it appears to be of significant research importance.

4.14 Niacin in Foods and Pharmaceuticals

Studies of the niacin potency of foods need to take account of the fact that there are two main diet-derived sources of body niacin stores: preformed niacin (nicotinamide and its coenzymes and nicotinic acid) and the precursor amino acid, tryptophan. Tryptophan contents can be measured by amino acid analysis after alkaline hydrolysis (tryptophan is destroyed by acid hydrolysis) and it is usually assumed as an approximation that one sixtieth of the dietary tryptophan is available for niacin synthesis.[38,308]

Nicotinic acid and its amide are chemically stable, but some of the niacin in cereals may be locked up in the unavailable complex, trigonelline, which can only be liberated by alkali treatment (as occurs in the cooking of Mexican tortillas). Food extraction for analysis generally uses mineral acid extraction under autoclave conditions (e.g. 1 N sulfuric acid), although gentler conditions may be preferable for some foods. Final assay quantitation is either by microbiological assay using *L. plantarum* or *K. apiculata*, by colorimetry involving a reaction with cyanogen bromide and sulfanilic acid, or by HPLC with optical density detection.[1,7–9,11,99,309] GLC-based methods have also been used.[3,11] Niacin in pharmaceutical preparations is generally in the form of either nicotinamide or nicotinic acid, and both forms have been used in large daily doses for specific therapeutic purposes.

4.15 Biotin Indices in Blood, Urine and Breast Milk

Although biotin is an essential vitamin, overt biotin deficiency in humans is rare and is usually associated with inborn (genetic) errors of metabolism; therefore few analytical laboratories are called upon to measure biotin status. Biotin concentrations in plasma or urine can be measured by a microbiological assay,[16,310] and this was the traditional procedure for many years. Recently, specific protein-binding assays, based on the specific biotin-binder avidin or other specific binders, have been developed.[311–317] Recent studies of urinary excretion products[310,318–320] promise to provide more specific information. As with most vitamins, HPLC has proved useful for the separation and quantitation of biotin and its catabolites.[310,321–323] A status assay based on the in vitro reactivation of the biotin-requiring enzyme, lymphocyte propionyl CoA carboxylase, appears promising.[16,324] The concentrations of biotin in plasma and in red blood cells are similar,[16] and there is little information to suggest whether either one or the other is preferable as a source of status information. There is little information about the variability of biotin in breast milk.[39] Thus studies of biotin status indicators remain at an early stage of development.

4.16 Biotin in Foods and Pharmaceuticals

Biotin is relatively stable, but can be destroyed by extremes of pH and temperature and by powerful oxidizing agents. There are bound forms, notably biocytin, in foods and tissues and for foods of animal origin it is necessary to extract with high concentrations of strong mineral acids at high temperature (e.g. 6 N H_2SO_4 in an autoclave). Different tissues require different extraction conditions for maximum yield. Proteolysis, for example with papain, may be effective and preferable.

Microbiological growth assay of extracts is achieved using *L. plantarum* or by radioassay using the specific

biotin-binding protein, avidin, as the binder and ^{14}C or ^{3}H-biotin as the tracer.[1,7–9,316] Several new assay procedures, based on HPLC or specific binders have been described in a series of reports.[310–315] These will need to be assessed for their potential usefulness in food biotin determinations. Pharmaceutical preparations containing biotin can be assayed by electrophoresis or chromatography-based assays.[67,106]

4.17 Pantothenate Assays (Status, Foods, Pharmaceuticals)

Pantothenate deficiency is practically unknown in humans and is only likely to arise if pantothenate antagonists are ingested. Older studies of the pantothenic acid content of blood and tissues were based on microbiological assays,[325–327] but enzymatic, mass spectrometric and other approaches have now been reported.[327–329] There is little information about its variation in breast milk.[39] Measurements of the pantothenic acid contents of foods are needed, particularly in the context of the development of food tables, and both microbiological and chemical (e.g. specific binding) procedures have been developed for this purpose.[325–333]

4.18 Other Substances with Vitamin-like Properties

There are, of course, innumerable substances present in foods that may have beneficial and protective properties, even though specific deficiency syndromes for them have not been described except in special circumstances, such as inborn errors, or particular diseases, or at extremes of the age range. The analysis of dietary and blood carotenoids has already been referred to, in sections 3.1 and 3.2; some carotenoids are vitamin A precursors and most are antioxidants, although their relevance to human health remains poorly understood. Another, numerous group of dietary antioxidants is the polyphenols, including the bioflavonoids. HPLC-based assays are being developed to quantitate these, both in foods and in body fluids such as blood and urine. Several of the essential and quasi-essential components of our diet are lipidic: these include the essential fatty acids, plus the major phospholipid components choline and inositol and the important lipid-carrier L-carnitine. These can be assayed by the classical techniques of lipid analysis (GLC and so on) or by enzyme-based methods. Other quasi-essential dietary components, such as taurine, are amino acids and can be assayed by the classical procedures of amino acid analysis.

During most of the twentieth century, analytical technology has been unable to measure large numbers of diverse chemical substances in a single run, but it is likely that this limitation will diminish in the future, with the perfection of new and powerful recognition procedures, such as controlled fragmentation MS, combined with more efficient preliminary separation methods. It should, therefore, become possible to quantitate a much greater range of nutrients than has been possible hitherto and to start to investigate the long-term effects of subtle dietary modulation, focusing on lesser-known components, which may have been ignored in the past. These may prove even more significant for long-term health than the 13 substances which have already, largely for historical reasons, been given the status of vitamins.

5 PROBLEMS ENCOUNTERED IN INTERPRETING VITAMIN STATUS INDICES

For most vitamins, the amount present in the deeper tissues of the body and the amount obtained from the diet over a period of time bear some relationship to the concentrations that are found in serum or plasma, blood cells, ratio of urinary excretion and so on, so that one or more of these indicators can be used as status and/or intake indicators for each of the vitamins. However, there are many confounding factors that can disturb and complicate the relationship between the status indices and the intakes or body stores of the vitamins which they aim to reflect. First, the amounts in serum or plasma may be modulated by specific binding proteins: this is particularly true for vitamin A (retinol) in plasma, but to a lesser extent also for several other vitamins. Both vitamins E and K are concentrated in circulating lipoproteins or lipids and their serum or plasma levels are thus proportional to the lipid levels as well as to body stores. Another important factor that can affect serum or plasma concentrations of most nutrients is the acute phase reaction: some nutrients are transferred into the plasma by infections or inflammatory processes, others are transferred out, so that the ratio of circulating levels to body stores changes for the duration of the acute phase reaction. This is especially true for retinol, but it is also observed for several water-soluble vitamins. Some vitamin concentrations are also sensitive to variations in kidney function, which can alter the overall composition of the circulating fluids.

Vitamin concentrations within the cellular components of the blood should be less seriously affected by the disturbances listed above, and indeed it is true that cellular concentrations can be better long-term status indicators for some of the vitamins, such as folate, thiamin and riboflavin. However, there are no red cell status indicators available for the fat-soluble vitamins or for vitamin B_{12}. Although red cell indicators can be useful, they may

also be complicated, e.g. by anemia, or by certain genetic diseases. White cell-based indices should also be valuable, but the complexity and variability over time of the white cell fraction and the relative difficulty of isolating pure cell types, has made this approach difficult to develop.

Urinary concentrations of B-vitamins were favored as indicators in the early years of vitamin studies, but they have proven less attractive lately, because of the difficulty of collecting accurate 24-h (or other timed) urine samples, and because the rate of urinary excretion is not very sensitive to status variations, when status is relatively poor. This arises from the fact that the vitamins are normally only excreted into the urine when the amount in the blood rises above that of tissue requirements. Degradation products may be formed at a rate that is proportional both to intake and to status, and this has been used for studies on niacin and vitamin B_6.

6 SUMMARY

The vitamins are a diverse group of compounds, both chemically and analytically, because they comprise a range of biomolecules whose common properties reside solely in the fact that they are essential dietary components, which are needed in relatively small amounts to sustain life and good health. In addition, the major analytical challenges that they present are derived from the facts that (1) there is a need to quantitate them in a wide range of biological matrices, which include both foods and body fluids (for status indices); (2) the concentrations that are present are usually very low and the ratio to other components, which can be chemically very similar, is small; (3) they may be present in several or many chemically diverse but biologically interconvertible, forms; (4) some of them are labile to heat, extremes of pH, degradative enzymes and so on and (5) there is no single analytical approach that can quantitate all of them together within a biological matrix.

Vitamins have been divided broadly into fat- and water-soluble groups and each of these has been further subdivided, mainly on a functional basis. Historically, the B-group vitamins were measured by various microbiological assays and others were measured (if at all) by semispecific colorimetric assays. Nowadays, there are several more powerful techniques available, which include high-pressure liquid chromatography, with a range of detectors operating over a continuous range of selective modalities; RIA and ELISA-based techniques, which use specific protein-binding selectivities, enzyme reactivation techniques for certain B-vitamin status assays, and so on. The use of MS is gaining ground and can be combined with several other separation techniques, such as GLC, CE and so on. Measurement of vitamins in simple pharmaceutical mixtures (vitamin supplements etc.) is now reasonably straightforward.

The technical advances that these methodologies represent have greatly broadened the horizons of the vitamin analyst, and have permitted even some of the more difficult biological matrices to be addressed. However, the unsolved challenges of accurate vitamin analysis show little sign of receding, and there are very few, if any, laboratories worldwide that are currently able to measure all of the known vitamins in all known biological matrices. This remains a major challenge and a goal for the twenty-first century.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
BHT	Butylated Hydroxytoluene
CE	Capillary Electrophoresis
CoA	Coenzyme A
ELISA	Enzyme-linked Immunosorbent Assay
EQAS	External Quality Assurance Schemes
FAD	Flavin Adenine Dinucleotide
FIGLU	Formiminoglutamic Acid
FMN	Flavin Mononucleotide
GC/MS	Gas Chromatography/Mass Spectrometry
GLC	Gas–Liquid Chromatography
HPLC	High-performance Liquid Chromatography
IU	International Unit
MMA	Methyl Malonic Acid
MS	Mass Spectrometry
NAD	Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NIST	National Institute of Standards and Technology
PIVKA	Protein Induced by Vitamin K Absence
QA	Quality Assurance
QC	Quality Control
RIA	Radioimmunoassay
SI	Système International d'Unités
TLC	Thin-layer Chromatography
TPP	Thiamin Pyrophosphate
UV	Ultraviolet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory • Capillary Electrophoresis in Clinical Chemistry • Gas Chromatography and Mass Spectrometry in Clinical Chemistry • Immunochemistry • Micro Total Analytical Systems in Clinical

Chemistry • Serum Proteins • Urinalysis and Other Bodily Fluids

Food **(Volume 5)**
Enzyme Analysis and Bioassays in Food Analysis • Fluorescence Spectroscopy in Food Analysis • Lipid Analyses in Food • Liquid Chromatography in Food Analysis • Vitamins Analysis in Food

Pharmaceuticals and Drugs **(Volume 8)**
Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis

Liquid Chromatography **(Volume 13)**
Affinity Chromatography • Biopolymer Chromatography • Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Ion Chromatography • Micellar Electrokinetic Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
Artificial Intelligence and Expert Systems in Mass Spectrometry • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. G.F.M. Ball, *Bioavailability and Analysis of Vitamins in Foods*, Chapman and Hall, London, 1998.
2. H.E. Sauberlich, J.H. Skala, R.P. Dowdy, *Laboratory Tests for the Assessment of Nutritional Status*, CRC Press, Cleveland, OH, 1974.
3. J. Velisek, J. Davidek, 'Gas–Liquid Chromatography of Vitamins in Foods: the Water-soluble Vitamins', *J. Micronutr. Anal.*, **2**(1), 25–42 (1986).
4. J. Davidek, J. Velisek, 'Gas–Liquid Chromatography of Vitamins in Foods: the Fat-soluble Vitamins', *J. Micronutr. Anal.*, **2**(2), 81–96 (1986).
5. G.F.M. Ball, 'Applications of HPLC to the Determination of Fat-soluble Vitamins in Foods and Animal Feeds (A Review 1977–87)', *J. Micronutr. Anal.*, **4**, 255–283 (1988).
6. A.P. De Leenheer, H.J. Nelis, W.E. Lambert, R.M. Bauwens, 'Chromatography of Fat-soluble Vitamins in Clinical Chemistry', *J. Chromatogr.*, **429**, 3–58 (1988).
7. D.J. Favell, 'Vitamin Analysis – Microbiological Methods', *J. Micronutr. Anal.*, **7**(3–4), 237–246 (1990).
8. H.A. Lee, N.C. Mills, P.M. Finglas, M.R.A. Morgan, 'Rapid Biospecific Methods of Vitamin Analysis', *J. Micronutr. Anal.*, **7**(3–4), 261–270 (1990).
9. I.D. Lumley, P.R. Lawrance, 'Advances in the Determination of Vitamins in Foodstuffs – Methods Used at the Laboratory of the Government Chemist', *J. Micronutr. Anal.*, **7**(3–4), 301–314 (1990).
10. F. Fidanza (ed.), 'Vitamin Nutriture Methodology', in *Nutritional Status Assessment: A Manual for Population Studies*, Chapman and Hall, London, 185–330, 1991.
11. A. Rizzolo, S. Polesello, 'Chromatographic Determination of Vitamins in Foods', *J. Chromatogr.*, **624**, 103–152 (1992).
12. FLAIR Concerted Action No. 10 Status Papers, 'Vitamins', *Int. J. Vitamin Nutr. Res.*, **63**, 247–295 (1993).
13. G.F.M. Ball, 'Determination of the Fat-soluble Vitamins in Foods by High-performance Liquid Chromatography', in *Handbook of Food Analysis*, ed. L.M.L. Nollet, Marcel Dekker, New York, 601–647, Vol. 1, 1996.
14. L.F. Russell, 'Water-soluble Vitamins', in *Handbook of Food Analysis*, ed. L.M.L. Nollet, Marcel Dekker, New York, 649–713, Vol. 1, 1996.
15. E.W. Gunter, B.G. Lewis, S.M. Koncikowski, *Laboratory Procedures Used for the Third National Health and Nutrition Examination Survey* (NHANES III), 1988–1994, Atlanta, GA. US Dept of Health and Human Services PHS, CDC, NCEH, Hyattsville, MD, NCHS, 1996.
16. C.J. Bates, 'Vitamin Analysis', *Ann. Clin. Biochem.*, **34**, 599–626 (1997).
17. P.L. Harris, 'Bioassay of Vitamin A Compounds', *Vitamins Hormones*, **18**, 347–370 (1960).
18. S.R. Ames, 'Biopotencies in Rats of Several Forms of alpha-Tocopherol', *J. Nutr.*, **109**, 2198–2204 (1979).
19. AOAC, 'Vitamin D in Milk, Vitamin Preparations and Feed Concentrates. Rat Bioassay. Final Action', in *AOAC Official Methods of Analysis*, 15th edition, ed. K. Helrich, Association of Official Analytical Chemists, 936.14, Arlington, VA, 1990.
20. E. Robaidek, 'Bioassay Methods for Nutrients in Processed Foods', *Food Technol.*, **37**(1), 81–83 (1983).

21. J.B. Neeld, W.N. Pearson, 'Macro- and Micro-methods for the Determination of Serum Vitamin A using Trifluoroacetic Acid', *J. Nutr.*, **79**, 454–462 (1963).
22. L. Friedman, W. Weiss, F. Wherry, O. Klein, 'Bioassay of Vitamin E by the Dialuric Acid Haemolysis Method', *J. Nutr.*, **65**, 143–160 (1958).
23. P.M. Allison, L.L. Mummah-Schendel, C.G. Kindberg, C.S. Harms, N.U. Bang, J.W. Suttie, 'Effects of Vitamin K Deficient Diet and Antibiotics in Normal Human Volunteers', *J. Lab. Clin. Med.*, **110**, 180–188 (1987).
24. C.I. Bliss, P. Gyorgy, 'Ascorbic Acid. III. Animal assay', in *The Vitamins, Chemistry, Physiology, Pathology, Methods*, 2nd edition, eds. P. Gyorgy, W.N. Pearson, Academic Press, New York, 49–51, Vol. VII, 1967.
25. J.H. Roe, C.A. Kuether, 'The Determination of Ascorbic Acid in Whole Blood and Urine through the 2,4-Dinitrophenylhydrazine Derivative of Dehydro-ascorbic Acid', *J. Biol. Chem.*, **147**, 399–407 (1943).
26. P.M. Scholes, 'The Microbiological Assay of Thiamine. Two Methods Modified for Use with Small Amounts of Test Material', *Analyst*, (London), **85**, 883–889 (1960).
27. H.A. Kornberg, R.S. Langdon, V.H. Cheldelin, 'Microbiological Assay for Riboflavin', *Anal. Chem.*, **20**, 81–83 (1948).
28. E.W. Toepfer, M.M. Polansky, 'Recent Developments in the Analysis for Vitamin B_6 in Foods', *Vitamins Hormones*, **22**, 825–832 (1964).
29. E.E. Snell, L.D. Wright, 'A Microbiological Method for the Determination of Nicotinic Acid', *J. Biol. Chem.*, **139**, 675–686 (1941).
30. D.R. Phillips, A.J.A. Wright, 'Studies on the Response of *Lactobacillus casei* to Different Folate Monoglutamates', *Br. J. Nutr.*, **47**, 183–189 (1982).
31. H.R. Skeggs, '*Lactobacillus leichmannii* Assay for Vitamin B_{12}', in *Analytical Microbiology*, ed. F. Kavanagh, Academic Press, New York, 551–565, 1963.
32. V.R. Williams, 'Growth Stimulants in the *Lactobacillus arabinosis* Biotin Assay', *J. Biol. Chem.*, **159**, 237–238 (1945).
33. H.R. Skeggs, L.D. Wright, 'The Use of *Lactobacillus arabinosis* in the Microbiological Determination of Pantothenic Acid', *J. Biol. Chem.*, **156**, 21–26 (1944).
34. D.C. Egberg, R.H. Potter, 'An Improved Automated Determination of Riboflavin in Food Products', *J. Agric. Food Chem.*, **23**, 815–820 (1975).
35. K.J. Carpenter, E. Kodicek, 'The Fluorimetric Estimation of N^1-methyl-nicotinamide and Its Differentiation from Coenzyme I', *Biochem. J.*, **46**, 421–426 (1950).
36. S. Waxman, C. Schreiber, V. Herbert, 'Radioisotopic Assay for Measurement of Serum Folate Levels', *Blood*, **38**, 219–228 (1971).
37. D.S.C. Lee, B.W. Griffiths, 'Human Serum Vitamin B_{12} Assay Methods – A Review', *Clin. Biochem.*, **18**, 261–266 (1985).
38. D.A. Bender, *Nutritional Biochemistry of the Vitamins*, Cambridge University Press, Cambridge, 1992.
39. C.J. Bates, A. Prentice, 'Breast Milk as a Source of Vitamins, Essential Minerals and Trace Elements', *Pharm. Ther.*, **62**, 193–220 (1994).
40. A.C. Ross, 'Separation and Quantitation of Retinyl Esters and Retinol by High Performance Liquid Chromatography', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 68–74, Vol. 123, 1986.
41. J.L. Napoli, 'Quantification of Physiological Levels of Retinoic Acid', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 112–124, Vol. 123, 1986.
42. G.L. Catignani, 'An HPLC Method for the Simultaneous Determination of Retinol and α-Tocopherol in Plasma or Serum', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 215–219, Vol. 123, 1986.
43. L.A. Kaplan, J.A. Miller, E.A. Stein, M.J. Stampfer, 'Simultaneous, High-performance Liquid Chromatography Analysis of Retinol, Tocopherols, Lycopene and Alpha and Beta Carotene in Serum and Plasma', in *Methods in Enzymology*, ed. L. Packer, Academic Press, London, 155–167, Vol. 189, 1990.
44. M.E. Cullum, M.H. Zile, S.W. Veysey, 'Analysis of Retinol and Dideuterated Retinol in Rat Plasma by Gas Chromatography Combined Mass Spectrometry', *Int. J. Vitamin Nutr. Res.*, **54**(1), 11–16 (1984).
45. W.A. MacCrehan, E. Schonberger, 'Determination of Retinol, Alpha-Tocopherol and β-Carotene in Serum by Liquid Chromatography with Absorbance and Electrochemical Detection', *Clin. Chem.*, **33**, 1585–1592 (1987).
46. Y.M. Peng, M.J. Xu, D.S. Alberts, 'Analysis and Stability of Retinol in Plasma', *J. Natl. Cancer Inst.*, **78**(1), 95–99 (1987).
47. D.I. Thurnham, E. Smith, P.S. Flora, 'Concurrent Liquid-Chromatographic Assay of Retinol, α-Tocopherol, β-Carotene, α-Carotene, Lycopene, and β-Cryptoxanthin in Plasma, with Tocopherol Acetate as Internal Standard', *Clin. Chem.*, **34**(2), 377–381 (1988).
48. N.E. Craft, E.D. Brown, J.C. Smith, 'Effects of Storage and Handling Conditions on Concentrations of Individual Carotenoids, Retinol and Tocopherol in Plasma', *Clin. Chem.*, **34**(1), 44–48 (1988).
49. K.S. Epler, R.G. Ziegler, N.E. Craft, 'Liquid Chromatographic Method for the Determination of Carotenoids, Retinoids and Tocopherols in Human Serum and in Food', *J. Chromatogr.*, **619**, 37–48 (1993).
50. Y. Ma, Z. Wu, H.C. Furr, C. Lammi-Keefe, N.E. Craft, 'Fast Minimicroassay of Serum Retinol (Vitamin A) by Capillary Zone Electrophoresis with Laser-excited Fluorescence Detection', *J. Chromatogr.*, **616**, 31–37 (1993).
51. A.L. Sowell, D.L. Huff, P.R. Yeager, S.P. Caudill, E.W. Gunter, 'Retinol, α-Tocopherol, Lutein/Zeaxanthin, β-Cryptoxanthin, Lycopene, α-Carotene, *trans*-β-Carotene, and Four Retinyl Esters in Serum Determined

Simultaneously by Reversed-phase HPLC with Multiwave Detection', *Clin. Chem.*, **40**(3), 411–416 (1994).
52. H.C. Furr, A.B. Barua, J.A. Olson, 'Analytical Methods', in *The Retinoids: Biology, Chemistry and Medicine*, 2nd edition, eds. M.B. Sporn, A.B. Roberts, D.S. Goodman, Raven Press, New York, 179–209, 1994.
53. H. Shi, Y. Ma, J.H. Humphrey, N.E. Craft, 'Determination of Vitamin A in Dried Human Blood Spots by High Performance Capillary Electrophoresis with Laser-excited Fluorescence Detection', *J. Chromatogr., B*, **665**, 89–96 (1995).
54. P. Jordan, D. Brubacher, U. Moser, H.B. Stahelin, K.F. Gey, 'Vitamin E and Vitamin A Concentrations in Plasma Adjusted for Cholesterol and Triglycerides by Multiple Regression', *Clin. Chem.*, **41**, 924–927 (1995).
55. B. Finckh, A. Kontush, J. Commentz, C. Hubner, M. Burdelski, A. Kohlschutter, 'Monitoring of Ubiquinol-10, Ubiquinone-10, Carotenoids, and Tocopherols in Neonatal Plasma Microsamples Using High-performance Liquid Chromatography with Coulometric Electrochemical Detection', *Anal. Biochem.*, **232**, 210–216 (1995).
56. A. Ben-Amotz, 'Simultaneous Profiling and Identification of Carotenoids, Retinols and Tocopherols by High-performance Liquid Chromatography Equipped with Three-dimensional Photodiode Array Detection', *J. Liquid Chromatogr.*, **18**(14), 2813–2825 (1995).
57. A.B. Burua, D. Kostic, M. Barua, J.A. Olson, 'Determination of Retinol and Retinoic Acid in Capillary Blood by High-performance Liquid Chromatography', *J. Liq. Chromatogr.*, **18**(7), 1459–1471 (1995).
58. L. Got, T. Gousson, E. Delacoux, 'Simultaneous Determination of Retinyl Esters and Retinol in Human Livers by Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr., B*, **668**(2), 233–239 (1995).
59. G.N. Noll, 'High-performance Liquid Chromatographic Analysis of Retinal and Retinol Isomers', *J. Chromatogr., A*, **721**(2), 247–259 (1996).
60. R.B. van Breeman, C-R Huang, 'High-performance Liquid Chromatography–Electrospray Mass Spectrometry of Retinoids', *FASEB J.*, **10**, 1098–1101 (1996).
61. J.-P. Steghens, A.L. van Kappel, E. Riboli, C. Collombel, 'Simultaneous Measurement of Seven Carotenoids, Retinol and α-Tocopherol in Serum by High-performance Liquid Chromatography', *J. Chromatogr., B*, **694**, 71–81 (1997).
62. R. Andreoli, M. Careri, P. Manini, G. Mori, M. Musci, 'HPLC Analysis of Fat-soluble Vitamins on Standard and Narrow Bore Columns with UV Electrochemical and Particle Beam MS Detection', *Chromatographia*, **44**, 605–612 (1997).
63. J.R. Lane, L.W. Webb, R.V. Acuff, 'Concurrent Liquid Chromatographic Separation and Photodiode Array Detection of Retinol, Tocopherols, all *trans*-α-Carotene, all *trans*-β-Carotene and the mono-*cis* Isomers of β-Carotene in Extracts of Human Plasma', *J. Chromatogr., A*, **787**, 111–118 (1997).
64. B.J. Burri, T.R. Neidlinger, A.O. Lo, C. Kwan, M.R. Wong, 'Supercritical Fluid Extraction and Reversed-phase Liquid Chromatography Methods for Vitamin A and Beta-Carotene. Heterogeneous Distribution of Vitamin A in the Liver', *J. Chromatogr., A*, **762**(1–2), 201–206 (1997).
65. T. Teerlink, M.P. Copper, I. Klaassen, B.J. Braakhuis, 'Simultaneous Analysis of Retinol, All-*trans*- and 13-*cis*-Retinoic Acid and 13-*cis*-4-Oxoretinoic Acid in Plasma by Liquid Chromatography Using On-column Concentration After Single-phase Fluid Extraction', *J. Chromatogr., B*, **694**(1), 83–92 (1997).
66. D.L. Duewer, J. Brown-Thomas, M.C. Kline, W.A. MacCrehan, R. Schaffer, K.E. Sharpless, W.E. May, 'NIST/NCI Micronutrients Measurement Quality Assurance Program: Measurement Repeatabilities and Reproducibilities for Fat-soluble Vitamin-related Compounds in Human Sera', *Anal. Chem.*, **69**(7), 1406–1413 (1997).
67. I.N. Papadoyannis, G.K. Tsioni, V.F. Samanidou, 'Simultaneous Determination of Nine Water and Fat Soluble Vitamins After SPE Separation and RP-HPLC Analysis in Pharmaceutical Preparations and Biological Fluids', *J. Liq. Chromatogr. Rel. Technol.*, **20**(19), 3203–3231 (1997).
68. G. Tang, J. Qin, G.G. Dolnikowski, 'Deuterium Enrichment of Retinol in Humans Determined by Gas Chromatography Electron Capture Negative Chemical Ionization Mass Spectrometry', *J. Nutr. Biochem.*, **9**(7), 408–414 (1998).
69. R.B. van Breeman, D. Nikolic, X. Yu, Y. Xiong, M. van Lieshout, C.E. West, A.B. Schilling, 'Development of a Method for Quantitation of Retinol and Retinyl Palmitate in Human Serum using High-performance Liquid Chromatography–Atmospheric Pressure Chemical Ionization–Mass Spectrometry', *J. Chromatogr., A*, **794**, 245–251 (1998).
70. M.A. Abahusain, J. Wright, J.W.T. Dickerson, M.A. El-Hazmi, H.Y. Aboul Enein, 'Determination of Retinol, α-Tocopherol, α- and β-Carotene by Direct Extraction of Human Serum using High-performance Liquid Chromatography', *Biomed. Chromatogr.*, **12**, 89–93 (1998).
71. A.B. Barua, J.A. Olson, 'Reversed-phase Gradient High-performance Liquid Chromatographic Procedure for Simultaneous Analysis of very Polar to Non-polar Retinoids, Carotenoids and Tocopherols in Animal and Plant Samples', *J. Chromatogr., B*, **707**, 69–79 (1998).
72. Y. Liu, M.J. Xu, L.M. Canfield, 'Enzymatic Hydrolysis, Extraction and Quantitation of Retinol and Major Carotenoids in Mature Human Milk', *J. Nutr. Biochem.*, **9**, 178–183 (1998).
73. R.B. van Breeman, D. Nikolic, X. Xu, Y. Xiong, M. van Lieshout, C.E. West, A.B. Schilling, 'Development of a Method for Quantitation of Retinol and Retinyl

Palmitate in Human Serum using High-performance Liquid Chromatography–Atmospheric Pressure Chemical Ionization–Mass Spectrometry', *J. Chromatogr., A*, **794**(1–2), 245–251 (1998).

74. M. Schultz, M. Leist, A. Elsner, R. Brigelius-Flohe, 'α-Carboxyethyl-6-hydroxychroman as Urinary Metabolite of Vitamin E', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 297–310, Vol. 282, 1997.
75. J.K. Kramer, L. Blais, R.C. Fouchard, R.A. Melnyk, K.M. Kallury, 'A Rapid Method for the Determination of Vitamin E Forms in Tissues and Diet by High-performance Liquid Chromatography Using a Normal-phase Diol Column', *Lipids*, **32**(3), 323–330 (1997).
76. D.I. Thurnham, J.A. Davies, B.J. Crump, R.D. Situnayake, M. Davis, 'The Use of Different Lipids to Express Serum Tocopherol: Lipid Ratios for the Measurement of Vitamin E Status', *Ann. Clin. Biochem.*, **23**, 514–520 (1986).
77. G.T. Vatassery, W.E. Smith, 'Determination of α-Tocopherylquinone (Vitamin E Quinone) in Human Serum, Platelets and Red Cell Membrane Samples', *Anal. Biochem.*, **167**, 411–417 (1987).
78. M.E. Murphy, J.P. Kehrer, 'Simultaneous Measurement of Tocopherols and Tocopheryl Quinones in Tissue Fractions Using High-performance Liquid Chromatography with Redox Cycling Electrochemical Detection', *J. Chromatogr.*, **421**, 71–82 (1987).
79. E.W. Gunter, W.J. Driskell, P.R. Yeager, 'Stability of Vitamin E in Long-term Stored Serum', *Clin. Chem.*, **175**, 329–336 (1988).
80. G.T. Vatassery, W.E. Smith, H.T. Quach, 'A Liquid Chromatographic Method for the Simultaneous Determination of α-Tocopherol and Tocopherolquinone in Human Red Blood Cells and Other Biological Samples where Tocopherol is Easily Oxidized During Sample Treatment', *Anal. Biochem.*, **214**, 426–430 (1993).
81. M. Maes, S. Weeckx, A. Wauters, H. Neels, S. Scharpe, R. Verkerk, P. Demedts, R. Desnyder, 'Biological Variability in Serum Vitamin E Concentrations: Relation to Serum Lipids', *Clin. Chem.*, **42**(11), 1824–1831 (1996).
82. D.C. Liebler, J.A. Burr, L. Philips, A.J.L. Ham, 'Gas Chromatography–Mass Spectrometry Analysis of Vitamin E and its Oxidation Products', *Anal. Biochem.*, **236**, 27–34 (1996).
83. J.F. Koprivnjak, K.R. Lum, M.M. Sisak, R. Saborowski, 'Determination of Alpha-, Gamma(+Beta)- and Delta-Tocopherols in a Variety of Liver Tissues by Reverse-phase High Pressure Liquid Chromatography', *Comp. Biochem. Physiol., B*, **113**(1), 143–148 (1996).
84. E.J. Konings, H.H. Roomans, P.R. Beljaars, 'Liquid Chromatographic Determination of Tocopherols and Tocotrienols in Margarine, Infant Foods and Vegetables', *J. AOAC Int.*, **79**(4), 902–906 (1996).
85. H. Takeda, T. Shibuya, K. Yanagawa, H. Kanoh, M. Takasaki, 'Simultaneous Determination of Alpha-Tocopherol and Alpha-Tocopherolquinone by High-performance Liquid Chromatography and Coulometric Detection in the Redox Mode', *J. Chromatogr., A*, **722**(1–2), 287–294 (1996).
86. M. Podda, C. Weber, M.G. Traber, L. Packer, 'Simultaneous Determination of Tissue Tocopherols, Tocotrienols, Ubiquinols and Ubiquinones', *J. Lipid Res.*, **37**(4), 893–901 (1996).
87. C. Leray, M. Andriamampandry, G. Gutbier, J. Cavadenti, C. Klein-Soyer, C. Gaichet, J.-P. Cazenave, 'Quantitative Analysis of Vitamin E, Cholesterol and Phospholipid Fatty Acids in a Single Aliquot of Human Platelets and Cultured Endothelial Cells', *J. Chromatogr., B*, **696**, 33–42 (1997).
88. O. Sommerburg, L.Y. Zang, F.J.G.M. van Kuijk, 'Simultaneous Detection of Carotenoids and Vitamin E in Human Plasma', *J. Chromatogr., B*, **695**(2), 209–215 (1997).
89. G.W. Burton, M.G. Traber, R.V. Acuff, D.N. Walters, H. Kayden, L. Hughes, K.U. Ingold, 'Human Plasma and Tissue α-Tocopherol Concentrations in Response to Supplementation with Deuterated Natural and Synthetic Vitamin E', *Am. J. Clin. Nutr.*, **67**, 669–684 (1998).
90. M.A. Tirmenstein, B.W. Watson, N.C. Haar, M.W. Fariss, 'Sensitive Method for Measuring Tissue Alpha-Tocopherol and Alpha-Tocopheryloxybutyric Acid by High-performance Liquid Chromatography with Fluorimetric Detection', *J. Chromatogr., B*, **707**(1–2), 308–311 (1998).
91. R. Neubert, J. Schiewe, 'Vitamin Analysis using Capillary Zone Electrophoresis', *Am. Biotechnol. Lab.*, **14**(4), 12, 14 (1996).
92. R.J. Caimi, J.T. Brenna, 'High Sensitivity Liquid Chromatography–Combustion Isotope Ratio Mass Spectrometry of Fat-soluble Vitamins', *J. Mass Spectrom.*, **30**, 466–472 (1995).
93. M.G. Feruzzi, L.C. Sander, C.L. Rock, S.J. Schwartz, 'Carotenoid Determination in Biological Microsamples Using Liquid Chromatography with a Coulometric Electrochemical Array Detector', *Anal. Biochem.*, **256**, 74–81 (1998).
94. A.Z. Mercadente, D.B. Rodriguez-Amaya, G. Britton, 'HPLC and Mass Spectrometric Analysis of Carotenoids from Mango', *J. Agric. Food Chem.*, **45**(1), 120–123 (1997).
95. R.B. van Breeman, 'Innovations in Carotenoid Analysis Using LC/MS', *Anal. Chem.*, **68**, 299A–304A (1996).
96. G.W. Chase, A.R. Long, 'Liquid Chromatographic Method for Analysis for All-rac-alpha-tocopherol Acetate and Retinyl Palmitate in Milk-based Infant Formula Using Matrix Solid-phase Dispersion', *J. AOAC Int.*, **81**(3), 582–586 (1998).
97. A. Pfalzgraf, H. Steinhart, M. Frigg, 'Rapid Determination of Alpha-Tocopherol in Muscle and Adipose Tissues

of Pork', *Z. Lebensm.-Unters.-Forsch.*, **200**(3), 190–193 (1995).
98. R. Rettenmaier, W. Schüep, 'Determination of Vitamins A and E in Liver Tissue', *Int. J. Vitamin Nutr. Res.*, **62**(4), 312–317 (1992).
99. S. Albala-Hurtado, M.T. Veciana-Nogues, M. Izquierdo-Pulido, A. Marine-Font, 'Determination of Water-soluble Vitamins in Infant Milk by High-performance Liquid Chromatography', *J. Chromatogr., A*, **778**(1–2), 247–253 (1997).
100. S. Albala-Hurtado, S. Novella-Rodriguez, M.T. Veciana-Nogues, A. Marine-Font, 'Determination of Vitamins A and E in Infant Milk Formulae by High-performance Liquid Chromatography', *J. Chromatogr., A*, **778**(1–2), 243–246 (1997).
101. M.M. Delgado-Zamarreno, A. Sanchez-Perez, M.C. Gomez-Perez, M.A. Fernando-Moro, J. Hernandez-Mendez, 'Determination of Vitamins A, E and K_1 in Milk by High-performance Liquid Chromatography with Dual Amperometric Detection', *Analyst*, **120**(10), 2489–2492 (1995).
102. R.B. McGeachin, C.A. Bailey, 'Determination of Carotenoid Pigments, Retinol and Alpha-Tocopherol in Feeds, Tissues and Blood Serum by Normal Phase High-performance Liquid Chromatography', *Poult. Sci.*, **74**(2), 407–411 (1995).
103. M.M. Delgado-Zamarreno, A. Sanchez-Perez, M.C. Gomez-Perez, J. Hernandez-Mendez, 'Directly Coupled Sample Treatment – High-performance Liquid Chromatography for On-line Automatic Determination of Liposoluble Vitamins in Milk', *J. Chromatogr., A*, **694**(2), 399–406 (1995).
104. S. Buskov, P. Moller, H. Sorensen, J.C. Sorensen, S. Sorensen, 'Determination of Vitamins in Food Based on Supercritical Fluid Extraction prior to Micellar Electrokinetic Capillary Chromatographic Analyses of Individual Vitamins', *J. Chromatogr., A*, **802**(1), 233–241 (1998).
105. B. Finckh, A. Kontush, J. Commentz, C. Hubner, M. Burdelski, A. Kohlshutter, 'Monitoring of Ubiquinol-10, Ubiquinone-10, Carotenoids and Tocopherols in Neonatal Plasma Microsamples using High-performance Liquid Chromatography with Coulometric Electrochemical Detection', *Anal. Biochem.*, **232**, 210–216 (1995).
106. L. Fotsing, M. Fillet, I. Bechet, Ph. Hubert, J. Crommen, 'Determination of Six Water-soluble Vitamins in a Pharmaceutical Formulation by Capillary Electrophoresis', *J. Pharm. Biomed. Anal.*, **15**, 1113–1123 (1997).
107. C.E. Porteous, R.D. Coldwell, D.J.H. Trafford, H.L.J. Makin, 'Recent Developments in the Measurement of Vitamin D and its Metabolites in Human Body Fluids', *J. Steroid Biochem.*, **28**(6), 785–801 (1987).
108. S.J. Iqbal, 'Vitamin D Metabolism and the Clinical Aspects of Measuring Metabolites', *Ann. Clin. Biochem.*, **31**, 109–124 (1994).
109. B.W. Hollis, 'Assessment of Vitamin D Nutritional and Hormonal Status: What to Measure and How to Do It', *Calcif. Tissue Int.*, **58**, 4–5 (1996).
110. B.W. Hollis, 'Detection of Vitamin D and its Major Metabolites', in *Vitamin D*, eds. D. Feldman, F. Glorieux, W. Pike, Academic Press, London, 587–606, 1997.
111. H. Schmidt-Gayk, R. Bouillon, H.J. Roth, 'Measurement of Vitamin D and its Metabolites (Calcidol and Calcitriol) and their Clinical Significance', *Scand. J. Clin. Lab. Invest.*, **57**, Suppl. 227, 35–45 (1997).
112. W.H. Bradbury, R.D. Coldwell, D.J.H. Trafford, H.L.J. Makin, 'A Semi-automated High-performance Liquid Chromatographic System for the Determination of 25-Hydroxyvitamin D in Human Plasma: Elimination of Interference by Barbiturates and Use of Photodiode Array Detection', *J. Pharm. Biomed. Anal.*, **6**(6–8), 919–925 (1988).
113. T.C. Chen, A.K. Turner, M.F. Holick, 'A Method for the Determination of Circulating Concentration of 1,25-dihydroxyvitamin D', *J. Nutr. Biochem.*, **1**, 320–327 (1990).
114. T.C. Chen, A.K. Turner, M.F. Holick, 'Methods for the Determination of the Circulating Concentration of 25-Hydroxyvitamin D', *J. Nutr. Biochem.*, **1**, 315–319 (1990).
115. M. Shimizu, Y. Gao, T. Aso, K. Nakatsu, S. Yamada, 'Fluorometric Assay of 25-Hydroxyvitamin D_3 and 24*R*, 25-Dihydroxyvitamin D_3 in Plasma', *Anal. Biochem.*, **204**, 258–264 (1992).
116. A.F. Hagar, L. Madson, L. Wales, H.B. Bradford, 'Reversed-phase Liquid Chromatographic Determination of Vitamin D in Milk', *J. AOAC Int.*, **77**(4), 1047–1051 (1994).
117. S. Masuda, T. Okano, M. Kamao, Y. Kanedai, T. Kobayashi, 'A Novel High-performance Liquid Chromatographic Assay for Vitamin D Metabolites Using a Coulometric Electrochemical Detector', *J. Pharm. Biomed. Anal.*, **15**, 1497–1502 (1997).
118. T.A. Reinhardt, B.W. Hollis, '1,25-Hydroxyvitamin D Microassay Employing Radioreceptor Techniques', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 176–185, Vol. 123, 1986.
119. W.E. Duncan, T.C. Aw, J.G. Haddad, 'Assay for 1,25-dihydroxyvitamin D Using Rabbit Intestinal Cytosol-binding Protein', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 185–190, Vol. 123, 1986.
120. S.C. Manolagas, 'Cytoreceptor Assay for 1,25-dihydroxyvitamin D', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 190–198, Vol. 123, 1986.
121. M.J. Beckman, H.F. DeLuca, 'Assay of 1,25-Dihydroxyvitamin D_3 from Serum Samples: Use of Receptor-binding or Enzyme-coupled Reporter Analysis', in *Methods in Enzymology*, eds. D.B. McCormick, J.W.

Suttie, C. Wagner, Academic Press, London, 164–174, Vol. 282, 1997.

122. B.W. Hollis, 'Quantitation of 25-Hydroxyvitamin D and 1,25-Dihydroxyvitamin D by Radioimmunoassay Using Radio-iodinated Tracers', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 174–186, Vol. 282, 1997.
123. B.W. Hollis, J.Q. Kamerud, S.R. Selvaag, J.D. Lorenz, J.L. Napoli, 'Determination of Vitamin D Status by Radioimmunoassay with an ^{125}I-labelled Tracer', *Clin. Chem.*, **39**, 529–533 (1993).
124. S. Wei, H. Tanaka, T. Kubo, M. Ichikawa, Y. Seino, 'A Multiple Assay for Vitamin D Metabolites Without High-performance Liquid Chromatography', *Anal. Biochem.*, **222**, 359–365 (1994).
125. B.W. Hollis, J.Q. Kamerud, A. Kurkowski, J. Beaulieu, J.L. Napoli, 'Quantification of Circulating 1,25-dihydroxyvitamin D by Radioimmunoassay with an ^{125}I-labelled Tracer', *Clin. Chem.*, **42**(4), 586–592 (1996).
126. W.D. Fraser, B.H. Durham, J.L. Berry, E.B. Mawer, 'Measurement of Plasma 1,25-Dihydroxyvitamin D Using Novel Immunoextraction Technique and Immunoassay with Iodine-labelled Vitamin D Tracer', *Ann. Clin. Biochem.*, **34**, 632–637 (1997).
127. C. Lind, J. Chen, I. Byrjalsen, 'Enzyme Immunoassay for Measuring 25-Hydroxyvitamin D_3 in Serum', *Clin. Chem.*, **43**(6), 943–949 (1997).
128. H. Koyama, J.M. Prahl, A. Uhland, M. Nanjo, M. Inaba, Y. Nisizawa, H. Morii, Y. Nishii, H.F. DeLuca, 'A New, Highly Sensitive Assay for 1,25-Dihydroxyvitamin D not requiring High-performance Liquid Chromatography: Application of Monoclonal Antibody against Vitamin D Receptor to Radio-receptor Assay', *Anal. Biochem.*, **205**(2), 213–219 (1992).
129. J. Abbotts, P.S.W. Davies, A. Prentice, D.M. Stirling, L. Yan, C. Bates, '25(OH)-Vitamin D Assay in Plasma: Experience in Using a Commercial Kit Assay for Survey Work', *Ann. Clin. Biochem.*, **32**, 591–592 (1995).
130. R. Vieth, A. Chan, A. Pollard, '^{125}I-RIA Kit Cannot Distinguish Vitamin D Deficiency as Well as a More Specific Assay for 25-hydroxyvitamin D', *Clin. Biochem.*, **28**, 175–179 (1995).
131. R.D. Coldwell, D.J.H. Trafford, M.J. Varley, D.N. Kirk, H.L.J. Makin, 'Stable Isotope-labelled Vitamin D Metabolites and Chemical Analogs: Synthesis and Use in Mass Spectrometric Studies', *Steroids*, **55**, 418–432 (1990).
132. D. Watson, K.D.R. Setchell, R. Ross, 'Analysis of Vitamin D and Its Metabolites Using Thermospray Liquid Chromatography/Mass Spectrometry', *Biomed. Chromatogr.*, **5**, 153–160 (1991).
133. E.B. Mawer, G. Jones, M. Davies, P.E. Still, V. Byford, N.J. Schroeder, H.L.J. Makin, C.W. Bishop, J.C. Knutson, 'Unique 24-hydroxylated Metabolites Represent a Significant Pathway of Metabolism of Vitamin D_2 in Humans: 24-hydroxyvitamin D_2 and 1,24-dihydroxy vitamin D_2 Detectable in Human Serum', *J. Clin. Endocrinol. Metab.*, **83**, 2156–2166 (1998).
134. B.W. Hollis, N.E. Frank, 'Quantitation of Vitamin D_2, Vitamin D_3, 25-Hydroxyvitamin D_2 and 25-Hydroxyvitamin D_3 in Human Milk', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 167–176, Vol. 123, 1986.
135. L.J. Sokoll, J.A. Sadowski, 'Comparison of Biochemical Indexes for Assessing Vitamin K Nutritional Status in a Healthy Adult Population', *Am. J. Clin. Nutr.*, **63**, 566–573 (1996).
136. M.J. Shearer, 'Assay of K Vitamins in Tissues by High-performance Liquid Chromatography with Special Reference to Ultraviolet Detection', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 235–251, Vol. 123, 1986.
137. J. Tamaoka, 'Analysis of Bacterial Menaquinone Mixtures by Reverse-phase High-performance Liquid Chromatography', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 251–256, Vol. 123, 1986.
138. K.W. Davidson, J.A. Sadowski, 'Determination of Vitamin K Compounds in Plasma or Serum by High-performance Liquid Chromatography Using Post-column Chemical Reduction and Fluorimetric Detection', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 408–421, Vol. 282, 1997.
139. S.J. Hodges, 'Assay of Menaquinones in Plasma Utilizing Dual-electrode Electrochemical Detection', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 434–438, Vol. 282, 1997.
140. P.T. McCarthy, D.J. Harrington, M.J. Shearer, 'Assay of Phylloquinone in Plasma by High-performance Liquid Chromatography with Electrochemical Detection', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 421–433, Vol. 282, 1997.
141. J.P. Langenberg, U.R. Tjoden, 'Determination of (endogenous) Vitamin K_1 in Human Plasma by Reversed-phase High-performance Liquid Chromatography Using Fluorometric Detection After Post-column Electrochemical Reduction. Comparison with Ultraviolet, Single and Dual Electrochemical Detection', *J. Chromatogr.*, **305**, 61–72 (1984).
142. W.E. Lambert, A.P. de Leenheer, M.F. Lefevere, 'Determination of Vitamin K in Serum Using HPLC with Post-column Reaction and Fluorescence Detection', *J. Chromatogr. Sci.*, **24**, 76–79 (1986).
143. Y. Haroon, D.S. Bacon, J.A. Sadowski, 'Liquid-chromatographic Determination of Vitamin K_1 in Plasma, with Fluorimetric Detection', *Clin. Chem.*, **32**, 1925–1929 (1986).
144. M. Guillaumont, M. Leclercq, H. Gosselet, K. Makala, B. Vignal, 'HPLC Determination of Serum Vitamin K_1

by Fluorometric Detection After Post-column Electrochemical Reduction', *J. Micronutr. Anal.*, **4**, 285–294 (1988).

145. E. Jakob, I. Elmadfa, 'Rapid HPLC Assay for the Assessment of Vitamin K_1, A, E and Beta-Carotene Status in Children (7–19 Years)', *Int. J. Vitamin Nutr. Res.*, **65**, 31–35 (1995).
146. W.A. MacCrehan, E. Schönberger, 'Determination of Vitamin K_1 in Serum Using Catalytic-reduction Liquid Chromatography with Fluorescence Detection', *J. Chromatogr., B.*, **670**, 209–217 (1995).
147. Y. Usui, 'Assay of Phylloquinone and Menaquinones in Human Liver', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 438–446, Vol. 282, 1997.
148. J.M. Conly, 'Assay of Menaquinones in Bacterial Cultures, Stool Samples and Intestinal Contents' in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 458–466, Vol. 282, 1997.
149. H.H.W. Thijssen, M.J. Drittij-Reijnders, 'Vitamin K Status in Human Tissues: Tissue-specific Accumulation of Phylloquinone and Menaquinone-4', *Br. J. Nutr.*, **75**, 121–127 (1996).
150. G. Fauler, H.J. Leis, J. Schalamon, W. Muntean, H. Gleispach, 'Method for the Determination of Vitamin $K_{1(20)}$ in Human Plasma by Stable Isotope Dilution/Gas Chromatography/Mass Spectrometry', *J. Mass Spectrom.*, **31**, 655–660 (1996).
151. K. Hirauchi, T. Sakano, S. Notsumoto, T. Nagaoka, A. Morimoto, K. Fujimoto, S. Masuda, Y. Suzuki, 'Measurement of K Vitamins in Animal Tissues by High-performance Liquid Chromatography with Fluorometric Detection', *J. Chromatogr.*, **497**, 131–137 (1989).
152. M.J. Shearer, A. Bach, M. Kohlmeier, 'Chemistry, Nutritional Sources, Tissue Distribution and Metabolism of Vitamin K with Special Reference to Bone Health', *J. Nutr.*, **126**, 1181S–1186S (1996).
153. S.T. Omaye, J.D. Turnbull, H.E. Sauberlich, 'Selected Methods for the Determination of Ascorbic Acid in Animal Cells, Tissues and Fluids', in *Methods in Enzymology*, eds. D.B. McCormick, L.D. Wright, Academic Press, London, 3–14, Vol. 62, 1979.
154. J.C. Deutsch, 'Gas Chromatographic/Mass Spectrometric Measurement of Ascorbic Acid and Analysis of Ascorbic Acid Degradation in Solution', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 13–24, Vol. 279, 1997.
155. L.W. Doner, K.B. Hicks, 'High-performance Liquid Chromatographic Separation of Ascorbic Acid, Erythorbic Acid, Dehydroascorbic Acid, Dehydro-erythorbic Acid, Diketogulonic Acid and Diketogluconic Acid', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 3–10, Vol. 122, 1986.
156. E. Kimoto, S. Terada, T. Yamaguchi, 'Analysis of Ascorbic Acid, Dehydroascorbic Acid, and Transformation Products by Ion-pairing High-performance Liquid Chromatography with Multiwavelength Ultraviolet and Electrochemical Detection', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 3–12, Vol. 279, 1997.
157. A.J. Speek, J. Schrijver, W.H.P. Schreurs, 'Fluorometric Determination of Total Vitamin C in Whole Blood by High-performance Liquid Chromatography with Precolumn Derivatization', *J. Chromatogr.*, **305**, 53–60 (1984).
158. M.J. Sanderson, C.J. Schorah, 'Measurement of Ascorbic Acid and Dehydroascorbic Acid in Gastric Juice by HPLC', *Biomed. Chromatogr.*, **2**, 197–202 (1987).
159. M.A. Kutnink, W.C. Hawkes, E.E. Schaus, S.T. Omaye, 'An Internal Standard Method for the Unattended High-performance Liquid Chromatographic Analysis of Ascorbic Acid in Blood Components', *Anal. Biochem.*, **166**(2), 424–430 (1987).
160. S.A. Margolis, T.P. Davies, 'Stabilization of Ascorbic Acid in Human Plasma and Its Liquid Chromatographic Measurement', *Clin. Chem.*, **34**, 2217–2223 (1988).
161. P.W. Washko, W.O. Hartzell, M. Levine, 'Ascorbic Acid Analysis Using High-performance Liquid Chromatography with Coulometric Electrochemical Detection', *Anal. Biochem.*, **181**, 276–282 (1989).
162. A.M. Bode, R.C. Rose, 'Quantitation of Ascorbic Acid by Liquid Chromatography with Coulometric Electrochemical Detection', *J. Micronutr. Anal.*, **8**(1), 55–62 (1990).
163. D. Schofield, P.M. Guyan, J.M. Braganza, 'Problems with the Quantitative Analysis of Dehydroascorbic Acid and Ascorbic Acid in Plasma by HPLC', *Biochem. Soc. Trans.*, **18**(6), 1179–1180 (1990).
164. K.R. Dhariwal, P.W. Washko, M. Levine, 'Determination of Dehydroascorbic Acid Using High-performance Liquid Chromatography with Coulometric Electrochemical Detection', *Anal. Biochem.*, **189**, 18–23 (1990).
165. H.P. Wagner, M.J. McGarrity, 'The Use of Pulsed Amperometry Combined with Ion-exclusion Chromatography for the Simultaneous Analysis of Ascorbic Acid and Sulfite', *J. Chromatogr.*, **546**(1–2), 119–124 (1991).
166. P.W. Washko, R.W. Welch, K.R. Dhariwal, Y. Wang, M. Levine, 'Ascorbic Acid and Dehydroascorbic Acid Analysis in Biological Samples', *Anal. Biochem.*, **204**, 1–14 (1992).
167. C.J. Bates, 'Use of Homocysteine to Stabilise Ascorbic Acid, or to Reduce Dehydroascorbic Acid, During HPLC Separation of Large Volumes of Tissue Extracts', *Clin. Chim. Acta*, **205**, 249–252 (1992).
168. D.A. Schell, A.M. Bode, 'Measurement of Ascorbic Acid and Dehydroascorbic Acid in Mammalian Tissues Utilizing HPLC and Electrochemical Detection', *Biomed. Chromatogr.*, **7**, 267–272 (1993).

169. C.J. Bates and Members of EC FLAIR Concerted Action No.10, 'Micronutrient Measurement, Absorption and Status: Plasma Vitamin C Assays: A European Experience', *Int. J. Vitamin Nutr. Res.*, **64**, 283–287 (1994).
170. K. Umegaki, K. Inoue, N. Takeuchi, M. Higuchi, 'Improved Method for the Analysis of Ascorbic Acid in Plasma by High-performance Liquid Chromatography with Electrochemical Detection', *J. Nutr. Sci. Vitaminol. Tokyo*, **40**(1), 73–79 (1994).
171. M.A. Ross, 'Determination of Ascorbic Acid and Uric Acid in Plasma by High-performance Liquid Chromatography', *J. Chromatogr.*, **657**, 197–200 (1994).
172. S. Wang, I.M. Schram, R.B. Sund, 'Determination of Plasma Ascorbic Acid by HPLC: Method and Stability Studies', *Eur. J. Pharm. Sci.*, **3**, 231–239 (1995).
173. J. Lykkesfeldt, S. Loft, H.E. Poulsen, 'Determination of Ascorbic Acid and Dehydroascorbic Acid in Plasma by High-performance Liquid Chromatography with Coulometric Detection – Are They Reliable Biomarkers of Oxidative Stress', *Anal. Biochem.*, **229**, 329–335 (1995).
174. S.A. Margolis, R.M. Schapira, 'Liquid Chromatographic Measurement of L-Ascorbic Acid and D-Ascorbic Acid in Biological Samples', *J. Chromatogr., B*, **690**, 25–33 (1997).
175. H. Iwase, I. Ono, 'Determination of Ascorbic Acid in Food by Column Liquid Chromatography with Electrochemical Detection Using Eluent for Pre-run Sample Stabilization', *J. Chromatogr., A*, **806**, 361–364 (1998).
176. S.A. Margolis, R.C. Paule, R.G. Ziegler, 'Ascorbic and Dehydroascorbic Acids Measured in Plasma Preserved with Dithioreitol or Meta-phosphoric Acid', *Clin. Chem.*, **36**, 1750–1755 (1990).
177. L.M. Edmond, D.I. Thurnham, S.A. Bingham, 'The Stability of Plasma Vitamin C During Storage in Cryo-bio Straws in Liquid Nitrogen and in Metaphosphoric Acid at −70 °C Over a 16-month Period', *Proc. Nutr. Soc.*, **54**, 187A (1996).
178. S.A. Margolis, D.L. Duewer, 'Measurement of Ascorbic Acid in Human Plasma and Serum: Stability, Intralaboratory Repeatability and Interlaboratory Reproducibility', *Clin. Chem.*, **42**(8), 1257–1262 (1996).
179. M.R. Lowik, J. Schrijver, M. Wedel, 'Vitamin C Analysis in Whole Blood, Plasma and Cells Using Reduced Glutathione as Preservative (Stabilizer): Losses and Redistribution', *Int. J. Vitamin Nutr. Res.*, **61**(1), 43–45 (1991).
180. J.P. Vuilleumier, E. Keck, 'Fluorometric Assay of Vitamin in Biological Materials Using a Centrifugal Analyser with Fluorescence Attachment', *J. Micronutr. Anal.*, **5**, 25–34 (1989).
181. K. Fukushi, S. Takeda, S. Wakida, M. Yamane, K. Higashi, K. Hiiro, 'Determination of Ascorbic Acid in Vegetables by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **772**, 313–320 (1997).
182. A. Prochazkova, L. Krivankova, P. Bocek, 'Quantitative Trace Analysis of L-Ascorbic Acid in Human Body Fluids by On-line Combination of Capillary Isotachophoresis and Zone Electrophoresis', *Electrophoresis*, **19**(2), 300–304 (1998).
183. M.W. Davey, G. Bauw, M. van Montagu, 'Analysis of Ascorbate in Plant Tissues by High Performance Capillary Zone Electrophoresis'', *Anal. Biochem.*, **239**(1), 8–19 (1996).
184. J.C. Deutsch, J.F. Kolhouse, 'Ascorbate and Dehydroascorbate Measurements in Aqueous Solutions and Plasma Determined by Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **65**, 321–326 (1993).
185. L.J. Bluck, A.P. Izzard, C.J. Bates, 'Measurement of Ascorbic Acid Kinetics in Man Using Stable Isotopes and Gas Chromatography/Mass Spectrometry', *J. Mass Spectrom.*, **31**(7), 741–748 (1996).
186. AOAC, 'Vitamin C (total) in Food. Semi-automated Fluorometric Method. Final Action 1985', in *AOAC Official Methods of Analysis*, 15th edition, ed. K. Helrich, Association of Official Analytical Chemists, Arlington, VA, 984.26, 1990.
187. G. Wingsle, T. Moritz, 'Analysis of Ascorbate and Dehydroascorbate in Plant Extracts by High-resolution Selected Ion Monitoring Gas Chromatography–Mass Spectrometry', *J. Chromatogr., A*, **782**, 95–103 (1997).
188. S. Finch, W. Doyle, C. Lowe, C.J. Bates, A. Prentice, G. Smithers, P.C. Clarke, *National Diet and Nutrition Survey: People Aged 65 Years and Over. Vol. 1, Report of the Diet and Nutrition Survey*, The Stationery Office, London, 1998.
189. E.E. Edwin, 'An Improved Procedure for the Determination of Thiamine', in *Methods in Enzymology*, eds. D.B. McCormick, L.D. Wright, Academic Press, London, 51–59, Vol. 62, 1979.
190. D.T. Wyatt, R.E. Hillman, 'Cyanogen Bromide-based Assay of Thiamin', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 91–97, Vol. 279, 1997.
191. T. Kawasaki, 'Determination of Thiamin and Its Phosphate Esters by High-performance Liquid Chromatography', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 15–19, Vol. 122, 1986.
192. V. Fayol, 'High-performance Liquid Chromatography Determination of Total Thiamin in Biological and Food Products', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 57–66, Vol. 279, 1997.
193. C.M.E. Tallaksen, T. Bohmer, J. Karlsen, H. Bell, 'Determination of Thiamin and Its Phosphate Esters in Human Blood, Plasma and Urine', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 67–74, Vol. 279, 1997.
194. J. Gerrits, H. Eidhof, J.W.I. Brunnekreeft, J. Hessels, 'Determination of Thiamin and Thiamin Phosphates

in Whole Blood by Reversed-phase Liquid Chromatography with Precolumn Derivatization', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 74–82, Vol. 279, 1997.

195. H.J. Mascher, C. Kikuta, 'High-performance Liquid Chromatography Determination of Total Thiamin in Human Plasma', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 83–90, Vol. 279, 1997.
196. J. Schrijver, A.J. Speek, J.A. Klosse, H.J.M. van Rijn, W.H.P. Schreurs, 'A Reliable Semiautomated Method for the Determination of Total Thiamine in Whole Blood by the Thiochrome Method with High-performance Liquid Chromatography', *Ann. Clin. Biochem.*, **19**, 52–56 (1982).
197. M. Kimura, Y. Itokawa, 'Determination of Thiamin and Thiamin Phosphate Esters in Blood by Liquid Chromatography with Post-column Derivatization', *Clin. Chem.*, **29**, 2073–2075 (1983).
198. T. Yamaguchi, K. Uchimura, H. Takenaka, K. Fukumoto, Y. Kurokawa, M. Hamada, S. Sugimoto, S. Matsukura, K. Inoue, N. Shibuya, 'Quantitative Determination of Thiamin and Its Phosphate Esters in Human Whole Blood by High-performance Liquid Chromatography and Clinical Application', *J. Clin. Biochem. Nutr.*, **2**, 203–216 (1987).
199. M. Baines, 'Improved High-performance Liquid Chromatographic Determination of Thiamin Diphosphate in Erythrocytes', *Clin. Chim. Acta.*, **153**, 43–48 (1985).
200. M. Baines, G. Davies, 'The Evaluation of Erythrocyte Thiamin Diphosphate as an Indicator of Thiamin Status in Man, and Its Comparison with Erythrocyte Transketolase Activity Measurements', *Ann. Clin. Biochem.*, **25**, 698–705 (1988).
201. H. Iwata, T. Matsuda, H. Tonomura, 'Improved High-performance Liquid Chromatographic Determination of Thiamin and Its Phosphate Esters in Animal Tissues', *J. Chromatogr.*, **450**, 317–323 (1988).
202. R.C. Rose, L.A. France, 'Determination of Thiamin and Its Biological Metabolites by High-performance Liquid Chromatography', *J. Micronutr. Anal.*, **5**, 227–233 (1989).
203. A.L. Bailey, P.M. Finglas, 'A Normal Phase High-performance Liquid Chromatographic Method for the Determination of Thiamin in Blood and Tissue Samples', *J. Micronutr. Anal.*, **7**, 147–157 (1990).
204. A.L. Bailey, P.M. Finglas, A.J.A. Wright, S. Southon, 'Thiamin Intake, Erythrocyte Transketolase (EC 2.2.1.1) Activity and Total Erythrocyte Thiamin in Adolescents', *Br. J. Nutr.*, **72**, 111–125 (1994).
205. P.L.M. Lynch, E.R. Trimble, I.S. Young, 'High-performance Liquid Chromatographic Determination of Thiamine Diphosphate in Erythrocytes Using Internal Standard Methodology', *J. Chromatogr., B*, **701**, 120–123 (1997).
206. P.M. Finglas, R.M. Faulks, 'The HPLC Analysis of Thiamin and Riboflavin in Potatoes', *Food. Chem.*, **15**, 37–44 (1984).
207. R.B.H. Wills, P. Wimalasiri, H. Greenfield, 'Comparative Determination of Thiamin and Riboflavin in Foods by High-performance Liquid Chromatography and Fluorometric Methods', *J. Micronutr. Anal.*, **1**, 23–29 (1985).
208. P. Wimalasiri, R.B. Wills, 'Simultaneous Analysis of Thiamin and Riboflavin in Foods by High-performance Liquid Chromatography', *J. Chromatogr.*, **318**(2), 412–416 (1985).
209. C. Hasselmann, D. Franck, P. Grimm, P.A. Diop, C. Soules, 'High-performance Liquid Chromatographic Analysis of Thiamin and Riboflavin in Dietetic Foods', *J. Micronutr. Anal.*, **5**(2), 269–280 (1989).
210. C. Vidal-Valverde, A. Reche, 'An Improved High-performance Liquid Chromatographic Method for Thiamin Analysis in Foods', *Z. Lebensm.-Unters.-Forsch.*, **191**(4–5), 313–318 (1990).
211. A. Laschi-Loquerie, S. Vallas, J. Viollet, M. Leclerq, V. Fayol, 'High-performance Liquid Chromatographic Determination of Thiamine in Biological and Food Products', *Int. J. Vitamin Nutr. Res.*, **62**(3), 248–251 (1992).
212. A. Sims, D. Shoemaker, 'Simultaneous Liquid Chromatographic Determination of Thiamine and Riboflavin in Selected Foods', *J. AOAC Int.*, **76**(5), 1156–1160 (1993).
213. R.E. Echols, R.H. Miller, W. Foster, 'Analysis of Thiamine in Milk by Gas Chromatography and the Nitrogen–Phosphorus Detector', *J. Dairy Sci.*, **69**(5), 1246–1249 (1986).
214. C.J. Bates, 'Human Riboflavin Requirements and Metabolic Consequences of Deficiency in Man and Animals', *World Rev. Nutr. Diet.*, **50**, 215–265 (1987).
215. J.A. Tillotson, M.M. Bashor, 'Fluorometric Titration of Urinary Riboflavin with an Apoflavoprotein', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 234–237, Vol. 122, 1986.
216. S. Huang, M.J. Caldwell, K.L. Simpson, 'Reverse Phase Open-column Chromatography Determination of Urinary Riboflavin', *Int. J. Vitamin Nutr. Res.*, **63**, 217–222 (1993).
217. S.-I. Huang, M.J. Caldwell, K.L. Simpson, 'Urinary Riboflavin Determination by C_{18} Reversed-phase Open-column Chromatography', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 343–351, Vol. 280, 1997.
218. K. Decker, A. Hinkkanen, 'Luminometric Determination of Flavin Adenine Dinucleotide', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 185–192, Vol. 122, 1986.
219. H.B. White III, 'Competitive Binding Assays for Riboflavin and Riboflavin-binding Protein', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 221–226, Vol. 122, 1986.

220. G.S. Cha, M.E. Meyerhoff, 'Solid Phase Enzyme-linked Competitive Binding Assay for Riboflavin', *Anal. Biochem.*, **168**, 216–227 (1988).
221. V.M. Kodentsova, O.A. Vrzhesinskaya, V.B. Spirichev, 'Fluorometric Riboflavin Titration in Plasma by Riboflavin-binding Apoprotein as a Method for Vitamin B_2 Status Assessment', *Ann. Nutr. Metab.*, **39**, 355–360 (1995).
222. A. Kozik, 'Microtitre-plate Enzyme-linked Ligandsorbent Assay of Riboflavin (Vitamin B_2) in Human Plasma and Urine', *Analyst*, **121**(3), 333–337 (1996).
223. A.J. Speek, F. van Schaik, J. Schrijver, W.H.P. Schreurs, 'Determination of the B_2 Vitamin Flavin Adenine Dinucleotide in Whole Blood by High-performance Liquid Chromatography with Fluorometric Detection', *J. Chromatogr.*, **228**, 311–316 (1982).
224. M.A. Marletta, D.R. Light, 'Flavins', in *Modern Chromatographic Analysis of the Vitamins*, eds. A.P. de Leenheer, W.E. Lambert, Dekker, New York, 413–433, 1985.
225. W.E. Lambert, P.M. Cammaert, A.P. de Leenheer, 'Liquid Chromatographic Measurement of Riboflavin in Serum and Urine with Isoriboflavin as Internal Standard', *Clin. Chem.*, **31**, 1371–1373 (1985).
226. A. Lopez-Anaya, M. Mayersohn, 'Quantification of Riboflavin, Riboflavin 5-phosphate and Flavin Adenine Dinucleotide in Plasma and Urine by High-performance Liquid Chromatography', *J. Chromatogr.*, **423**, 105–113 (1987).
227. L.F. Russell, J.T. Vanderslice, 'A Comprehensive Review of Vitamin B_2 Analytical Methodology', *J. Micronutr. Anal.*, **8**, 257–310 (1990).
228. S. Harbron, H.J. Eggelte, B.R. Rabin, 'Amplified Colorimetric Assay of Alkaline Phosphatase Using Riboflavin 4-phosphate: A Simple Method for Measuring Riboflavin and Riboflavin 5′ Phosphate', *Anal. Biochem.*, **198**, 47–51 (1991).
229. M.J. Kim, H.J. Kim, J.M. Kim, B. Kim, S.H. Han, G.S. Cha, 'Homogeneous Assays for Riboflavin Mediated by the Interaction between Enzyme-biotin and Avidin-riboflavin Conjugates', *Anal. Biochem.*, **231**, 400–406 (1995).
230. S.H. Ashoor, M.J. Knox, J.R. Olsen, D.A. Deger, 'Improved Liquid Chromatographic Determination of Riboflavin in Milk and Dairy Products', *J. AOAC*, **68**, 693–696 (1985).
231. B.K. Ayi, D.A. Yuhas, N.J. DeAngelis, 'Simultaneous Determination of Vitamins B_2 (Riboflavin) and B_6 (Pyridoxine) in Infant Formula Products by Reverse Phase Liquid Chromatography', *J. AOAC*, **69**, 56–59 (1986).
232. H. Johnsson, C. Branzell, 'High Performance Liquid Chromatographic Determination of Riboflavin in Food – a Comparison with a Microbiological Method', *Int. J. Vitamin Nutr. Res.*, **57**, 53–58 (1987).
233. P.M. Finglas, R.M. Faulks, 'Critical Review of HPLC Methods for the Determination of Thiamin, Riboflavin and Niacin in Food', *J. Micronutr. Anal.*, **3**(4), 251–284 (1987).
234. E.S.P. Reyes, K.M. Norris, C. Taylor, D. Potts, 'Comparison of Paired-ion Liquid Chromatographic Method with AOAC Fluorometric and Microbiological Methods for Riboflavin Determination in Selected Foods', *J. AOAC*, **71**, 16–19 (1988).
235. V. Ollilainen, P. Mattila, P. Varo, P. Koivistoinen, J. Huttunen, 'The HPLC Determination of Total Riboflavin in Foods', *J. Micronutr. Anal.*, **8**, 199–207 (1990).
236. A.K. Hewavitharana, 'Method for the Extraction of Riboflavin for High-performance Liquid Chromatography and Application to Casein', *Analyst*, **121**, 1671–1676 (1996).
237. L.F. Russell, J.T. Vanderslice, 'Non-degradative Extraction and Simultaneous Quantitation of Riboflavin, Flavin Mononucleotide and Flavin Adenine Dinucleotide in Foods by HPLC', *Food Chem.*, **43**, 151–162 (1992).
238. G.M. Greenway, N. Kometa, 'On-line Sample Preparation for the Determination of Riboflavin and Flavin Mononucleotides in Foodstuffs', *Analyst*, **119**, 929–935 (1994).
239. H. Millart, D. Lamiable, 'Determination of Pyridoxal 5′-Phosphate in Human Serum by Reversed-phase High-performance Liquid Chromatography Combined with Spectrofluorimetric Detection of 4-Pyridoxic Acid 5′-Phosphate as a Derivative', *Analyst*, **114**, 1225–1228 (1989).
240. M.S. Chauhan, K. Dakshinamurti, 'Fluorometric Assay of Pyridoxal' in *Methods in Enzymology*, eds. D.B. McCormick, L.D. Wright, Academic Press, London, 405–410, Vol. 62, 1979.
241. S.P. Coburn, J.D. Mahuren, 'Cation-exchange High-performance Liquid Chromatographic Analysis of Vitamin B_6', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 102–111, Vol. 122, 1986.
242. H. Tsuge, 'Determination of Vitamin B_6 Vitamers and Metabolites in a Biological Sample', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 3–12, Vol. 280, 1997.
243. H.J. Mascher, 'High-performance Liquid Chromatography Determination of Total Pyridoxal in Human Plasma', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 12–21, Vol. 280, 1997.
244. J.D. Mahuren, S.P. Coburn, 'Determination of 5-Pyridoxic Acid, 5-Pyridoxic Acid Lactone, and Other Vitamin B_6 Compounds by Cation-exchange High-performance Liquid Chromatography'', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 22–30, Vol. 280, 1997.
245. J.F. Gregory, 'Determination of Pyridoxal 5′-Phosphate as the Semi-carbazone Derivative Using High-performance Liquid Chromatography', *Anal. Biochem.*, **102**, 374–379 (1980).

246. J.B. Ubbink, W.J. Serfontein, L.S. de Villiers, 'Stability of Pyridoxal 5-Phosphate Semicarbazone: Applications in Plasma Vitamin B_6 Analysis and Population Surveys of Vitamin B_6 Nutritional Status', *J. Chromatogr.*, **342**(2), 277–284 (1985).
247. P. Linares, M.D. Luque-de Castro, M. Valcarcel, 'Simultaneous Determination of Pyridoxal and Pyridoxal 5-Phosphate in Human Serum by Flow Injection Analysis', *Anal. Chem.*, **57**(11), 2101–2106 (1985).
248. J.B. Ubbink, W.J. Serfontein, L.S. de Villiers, 'Analytical Recovery of Protein-bound Pyridoxal-5′-Phosphate in Plasma Analysis', *J. Chromatogr.*, **375**(2), 399–404 (1986).
249. P. Edwards, P.K.S. Liu, G.A. Rose, 'A Simple Liquid Chromatographic Method for Measuring Vitamin B_6 Compounds in Plasma', *Clin. Chem.*, **35**, 241–245 (1989).
250. D. Hess, J.P. Vuilleumier, 'Assay of Pyridoxal-5′-Phosphate, Pyridoxal and Pyridoxic Acid in Biological Material', *Int. J. Vitamin Nutr. Res.*, **59**(4), 338–343 (1989).
251. J.E. Leklem, 'Vitamin B_6: A Status Report', *J. Nutr.*, **120**, 1503–1507 (1990).
252. S. Kurioka, N. Ishioka, J. Sato, J. Nakamura, T. Ohkubo, M. Matsuda, 'Assay of Vitamin B_6 in Human Plasma with Graphitic Carbon Column', *Biomed. Chromatogr.*, **7**, 162–165 (1993).
253. J.P. Vuilleumier, H.E. Keller, F. Fidanza, J. Schrijver, A.J. Speek, H. van den Berg, 'Pyridoxal 5′-Phosphate in Plasma by Radioenzymatic Assay Using Tyrosine Decarboxylase Apoenzyme', in *Nutritional Status Assessment – A Manual for Population Studies*, ed. F. Fidanza, Chapman and Hall, London, 276–280, 1991.
254. G.W. Chase, W.O. Landen, R.R. Eitenmiller, A.-G.M. Soliman, 'Liquid Chromatographic Determination of Thiamine, Riboflavin and Pyridoxine in Infant Formula', *J. AOAC Int.*, **75**, 561–565 (1992).
255. G. Komaromy-Hiller, K.L. Nuttall, E.R. Ashwood, 'Effect of Storage on Serum Vitamin B_{12} and Folate Stability', *Ann. Clin. Biochem.*, **27**, 249–253 (1997).
256. E.W. Newnham, J.F. Tsai, 'Microbiological Analysis of 5-Formyltetrahydro Folic Acid and Other Folates Using an Automatic 96-well Plate Reader', *Anal. Biochem.*, **154**(2), 509–515 (1986).
257. S. O'Broin, B. Kelleher, 'Microbiological Assay on Microtitre Plates of Folate in Serum and Red Cells', *J. Clin. Pathol.*, **45**, 344–347 (1992).
258. D.W. Horne, 'Microbiological Assay of Folates in 96-Well Microtiter Plates', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wayne, Academic Press, London, 38–53, Vol. 281, 1997.
259. S. Waxman, C. Schreiber, 'Determination of Folate by Use of Radioactive Folate and Binding Proteins', in *Methods in Enzymology*, eds. D.B. McCormick, L.D. Wright, Academic Press, London, 468–483, Vol. 66, 1980.
260. R.D. Brown, R. Jun, W. Hughes, R. Watman, B. Arnold, H. Kronenberg, 'Red Cell Folate Assay: Some Answers to Current Problems with Radioassay Variability', *Pathology*, **22**, 82–87 (1990).
261. H. van den Berg, P.M. Finglas, C. Bates, 'FLAIR Intercomparisons on Serum and Red Cell Folate', *Int. J. Vitamin Nutr. Res.*, **64**, 288–293 (1994).
262. D.J. Raiten, K.D. Fisher (eds.), 'Assessment of Folate Methodology Used in the Third National Health and Nutritional Examination Survey (NHANES III, 1988–1994)', *J. Nutr.*, **125**, 1375S–1398S (1995).
263. E. Gunter, B.A. Bowman, S.P. Caudill, D.B. Twite, M.J. Adams, E.J. Sampson, 'Results of an International Round Robin for Serum and Red Cell Folate', *Clin. Chem.*, **42**, 1689–1694 (1996).
264. D.H. Wilson, R. Hermann, S. Hsu, T. Biegalski, L. Sohn, C. Forsythe, M. Norotny, M. Beggs, G. Mandarino, 'Ion Capture Assay for Folate with the Abbott IMx Analyzer', *Clin. Chem.*, **41**, 1780–1781 (1995).
265. C.R. Santhosh-Kumar, J.C. Deutsch, K.L. Hassell, N.M. Kolhouse, J.F. Kolhouse, 'Quantitation of Red Blood Cell Folates by Stable Isotope Dilution Gas Chromatography/Mass Spectrometry Utilizing a Folate Internal Standard', *Anal. Biochem.*, **225**, 1–9 (1995).
266. C.R. Santhosh-Kumar, J.F. Kolhouse, 'Molar Quantitation of Folates by Gas Chromatography/Mass Spectrometry', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 26–38, Vol. 281, 1997.
267. S.D. Wilson, D.W. Horne, 'High-performance Liquid Chromatographic Separation of the Naturally Occurring Folic Acid Derivatives', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 269–273, Vol. 122, 1986.
268. P.J. Bagley, J. Selhub, 'Analysis of Folates Using Combined Affinity and Ion-pair Chromatography', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 16–25, Vol. 281, 1997.
269. T. Rebello, 'Trace Enrichment of Biological Folates on Solid Phase Adsorption Cartridges and Analysis by High-pressure Liquid Chromatography', *Anal. Biochem.*, **166**(1), 55–64 (1987).
270. J. Selhub, 'Determination of Tissue Folate Composition by Affinity Chromatography Followed by High-pressure Ion Pair Liquid Chromatography', *Anal. Biochem.*, **182**, 84–93 (1989).
271. J.C. Gounelle, H. Ladjimi, P. Prognon, 'A Rapid and Specific Extraction Procedure for Folates Determination in Rat Liver and Analysis by High-performance Liquid Chromatography with Fluorimetric Detection', *Anal. Biochem.*, **176**(2), 406–411 (1989).
272. M.-C. Etienne, N. Speciale, G. Milano, 'HPLC of Folinic Acid Diastereoisomers and *S*-Methyltetrahydrofolate in Plasma', *Clin. Chem.*, **39**, 82–86 (1993).
273. M.D. Lucock, I. Daskalakis, C.J. Schorah, M.I. Levine, R. Hartley, 'Analysis and Biochemistry of Blood Folate', *Biochem. Mol. Med.*, **58**, 93–112 (1996).

274. C.M. Pfeiffer, J.F. Gregory III, 'Enzymatic Deconjugation of Erythrocyte Polyglutamyl Folates During Preparation for Folate Assay: Investigation with Reversed-phase Liquid Chromatography', *Clin. Chem.*, **42**(11), 1847–1854 (1996).
275. J. McPartlin, J. Scott, 'Identification and Assay of Folate Catabolites in Human Urine', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wayne, Academic Press, London, 70–77, Vol. 281, 1997.
276. J.P. Toth, J.F. Gregory, 'Analysis of Folacin by GC/MS: Structures and Mass Spectra of Fluorinated Derivatives of *Para*-amino Benzoyl Glutamic Acid', *Biomed. Environ. Mass Spectrom.*, **17**, 73–79 (1988).
277. L.M. Rogers, C.M. Pfeiffer, L.B. Bailey, J.F. Gregory III, 'A Dual-label Stable-isotopic Protocol is Suitable for Determination of Folate Bioavailability in Humans: Evaluation of Urinary Excretion and Plasma Folate Kinetics of Intravenous and Oral Doses of ($^{13}C_5$) and ($^{2}H_2$) Folic Acid', *J. Nutr.*, **127**, 2321–2327 (1997).
278. J.I. Martin, W.O. Landen, A.M. Soliman, R.R. Eitenmiller, 'Application of a Tri-enzyme Extraction for Total Folate Determination in Foods', *J. AOAC*, **73**, 805–808 (1990).
279. S. de Souza, R. Eitenmiller, 'Effects of Different Enzyme Treatments on Extraction of Total Folate from Various Foods Prior to Microbiological Assay and Radioassay', *J. Micronutr. Anal.*, **7**(1), 37–58 (1990).
280. T. Tamura, Y. Mizuno, K.E. Johnston, R.A. Jacob, 'Food Folate Assay with Protease, α-Amylase, and Folate Conjugase Treatments', *J. Agric. Food Chem.*, **45**, 135–139 (1997).
281. F.M. Pfeiffer, L.M. Rogers, J.F. Gregory III, 'Determination of Folate in Cereal-grain Food Products Using Trienzyme Extraction and Combined Affinity and Revised-phase Liquid Chromatography', *J. Agric. Food Chem.*, **45**, 407–413 (1997).
282. T. Tamura, 'Determination of Food Folate', *J. Nutr. Biochem.*, **9**, 285–293 (1998).
283. L.T. Vahteristo, V. Ollilainen, P.E. Koivistoinen, P. Varo, 'Improvements in the Analysis of Reduced Folate Monoglutamates and Folic Acid in Food by High-performance Liquid Chromatography', *J. Agric. Food Chem.*, **44**(2), 477–482 (1996).
284. M.D. Lucock, M. Green, M. Priestnall et al., 'Optimisation of Chromatographic Conditions for the Determination of Folates in Foods and Biological Tissues for Nutritional and Clinical Work', *Food Chem.*, **53**, 329–338 (1995).
285. P. Gimsing, W.S. Beck, 'Determination of Cobalamins in Biological Material', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 3–14, Vol. 123, 1986.
286. D.B. Matchar, D.C. McCrory, D.S. Millington, J.R. Feussner, 'Performance of the Serum Cobalamin Assay for Diagnosis of Cobalamin Deficiency', *Am. J. Med. Sci.*, **308**(5), 276–283 (1994).
287. V. Herbert, N. Colman, D. Palat, C. Manusselis, G. Drivas, E. Block et al., 'Is there a ''Gold Standard'' for Human Serum Vitamin B_{12} Assay?', *J. Lab. Clin. Med.*, **104**, 829–841 (1984).
288. D.S.C. Lee, B.W. Griffiths, 'Human Serum Vitamin B_{12} Assay Methods – A Review', *Clin. Biochem.*, **18**, 261–266 (1985).
289. J.J. O'Sullivan, R.J. Leeming, S.S. Lynch, A. Pollock, 'Radioimmunoassay that Measures Serum Vitamin B_{12}', *J. Clin. Pathol.*, **45**, 328–331 (1992).
290. S.C. Kuemmerle, G.L. Boltinghouse, S.M. Delby, T.L. Lane, R.P. Simondsen, 'Automated Assay of Vitamin B_{12} by the Abbott IMx Analyzer', *Clin. Chem.*, **38**(10), 2073–2077 (1992).
291. J. Arnaud, A. Cotisson, G. Meffre, M. Bourgeay-Causse, C. Augert, A. Favier, J.P. Vuillez, G. Ville, 'Comparison of Three Commercial Kits and a Microbiological Assay for the Determination of Vitamin B_{12} in Serum', *Scand. J. Clin. Lab. Invest.*, **54**(3), 235–240 (1994).
292. S. Wentworth, J.A. McBride, W.H. Walker, 'Chemiluminescence Receptor Assay for Measuring Vitamin B_{12} Serum Evaluated', *Clin. Chem.*, **40**(4), 537–540 (1994).
293. L.S.W. Steijns, J.K. Braams-Wiatrowska, H.J. Luiting, J. van der Weide, 'Evaluation of Nonisotopic Binding Assays for Measuring Vitamin B_{12} and Folate in Serum', *Clin. Chim. Acta.*, **248**, 135–141 (1996).
294. D.W. Jacobsen, R. Green, K.L. Brown, 'Analysis of Cobalamin Coenzymes and Other Corrinoids by High-performance Liquid Chromatography', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 14–22, Vol. 123, 1986.
295. D.P. Sundin, R.H. Allen, 'Analysis of Cobalamin and Cobalamin Analogs by Gas Chromatography/Mass Spectrometry', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 247–255, Vol. 281, 1997.
296. H. Chassaigne, R. Lobinski, 'Direct Species-selective Determination of Cobalamins by Ionspray Mass Spectrometry and Ionspray Tandem Mass Spectrometry', *Analyst*, **123**, 131–137 (1998).
297. K.C. Das, C. Manusselis, V. Herbert, 'Determination of Vitamin B_{12} (Cobalamin) in Serum and Erythrocytes by Radioassay and of Holo-*trans* cobalamin II (Holo-TC II) and Holo-haptocorrin (Holo-TC I and III) in Serum by Adsorbing Holo-TC II on Microfine Silica', *J. Nutr. Biochem.*, **2**, 455–464 (1991).
298. E.J. Norman, J.A. Morrison, 'Screening Elderly Populations for Cobalamin (Vitamin B_{12}) Deficiency Using the Urinary Methylmalonic Acid Assay by Gas Chromatography Mass Spectrometry', *Am. J. Med.*, **94**(6), 589–594 (1993).
299. H. Iwase, I. Ono, 'Determination of Cyanocobalamin in Foods by High-performance Liquid Chromatography with Visible Detection after Solid-phase Extraction and

Membrane Filtration for the Precolumn Separation of Lipophilic Species', *J. Chromatogr., A*, **771**, 127–134 (1997).
300. J.S. Sandhu, D.R. Fraser, 'Measurement of Niacin Metabolites in Urine by High Pressure Liquid Chromatography. A Simple Sensitive Assay of Niacin Nutritional Status', *Int. J. Vitamin Nutr. Res.*, **51**, 139–144 (1981).
301. K. Shibata, T. Kawada, K. Iwai, 'Simultaneous Microdetermination of Nicotinamide and Its Major Metabolites, N^1 Methyl-2-pyridone-5-carboxamide and N^1-methyl-4-pyridone-3-carboxamide, by High-performance Liquid Chromatography', *J. Chromatogr.*, **424**, 23–28 (1988).
302. K. Shibata, H. Matsuo, 'Correlation Between Niacin Equivalent Intake and Urinary Excretion of Its Metabolites, N^1-Methylnicotinamide, N^1 Methyl-2-pyridone-5-carboxamide and N^1-Methyl-4-pyridone-3-carboxamide in Humans Consuming a Self-selected Food', *Am. J. Clin. Nutr.*, **50**, 114–119 (1989).
303. A. Somogyi, D. Siebert, F. Bochner, 'Determination of Endogenous Concentrations of N^1-Methylnicotinamide in Human Plasma and Urine by High-performance Liquid Chromatography', *Anal. Biochem.*, **187**, 160–165 (1990).
304. R.B. Sashidher, Abhaykumar, H.V.V. Murthy, 'Analysis of Urinary Niacin and Quinolinic Acid: Comparison of Chemical and Microbiological Methods', *J. Micronutr. Anal.*, **7**(1), 59–66 (1990).
305. B.R. Clark, 'Fluorometric Quantitation of Picomole Amounts of 1-Methylnicotinamide and Nicotinamide in Serum', in *Methods in Enzymology*, eds. D.B. McCormick, L.D. Wright, Academic Press, London, 3–8, Vol. 66, 1980.
306. Y. Miyauchi, N. Sano, T. Nakamura, 'Simultaneous Determination of Nicotinic Acid and Its Two Metabolites in Human Plasma Using Solid-phase Extraction in Combination with High-performance Liquid Chromatography', *Int. J. Vitamin Nutr. Res.*, **63**(2), 145–149 (1993).
307. E.L. Jacobson, M.K. Jacobson, 'Tissue NAD as a Biochemical Measure of Niacin Status in Humans', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 221–230, Vol. 280, 1997.
308. L.J. Machlin (ed.), *Handbook of Vitamins*, 2nd edition, Marcel Dekker, New York, 1991.
309. G.W. Chase, W.O. Landon, A.-G.M. Soliman, R.R. Eitenmiller, 'Liquid Chromatographic Analysis of Niacin in Fortified Food Products', *J. AOAC Int.*, **76**, 390–393 (1993).
310. D.M. Mock, 'Determinations of Biotin in Biological Fluids', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 265–275, Vol. 279, 1997.
311. R.D. Nargesi, D.S. Smith, 'Fluorimetric Assays for Avidin and Biotin', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 67–72, Vol. 122, 1986.
312. S. Lizano, S. Ramanathan, A. Feltus, A. Witkowski, S. Daunert, 'Bioluminescence Competitive Binding Assays for Biotin Based on Photoprotein Aequorin', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 296–303, Vol. 279, 1997.
313. E.Z. Huang, Y.H. Rogers, 'Competitive Enzymatic Assay of Biotin', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 304–308, Vol. 279, 1997.
314. E.Z. Huang, 'Competitive Agglutination Assay of Biotin', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 308–320, Vol. 279, 1997.
315. D. Shiuan, C.H. Wu, Y.-S. Chang, R.-J. Chang, 'Competitive Enzyme-linked Immunosorbent Assay for Biotin', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 321–326, Vol. 279, 1997.
316. R. Bitsch, I. Salz, D. Hotzel, 'Biotin Assessment in Foods and Body Fluids by a Protein Binding Assay (PBA)', *Int. J. Vitamin Nutr. Res.*, **59**(1), 59–64 (1989).
317. G.J. Wellenberg, J.N. Banks 'Enzyme-linked Sorbent Assay to Quantify D-Biotin in Blood', *J. Sci. Food Agric.*, **63**, 1–5 (1993).
318. D.M. Mock, G.L. Lankford, J. Cazin, 'Biotin and Biotin Analogs in Human Urine: Biotin Accounts for Only Half of the Total', *J. Nutr.*, **123**, 1844–1851 (1993).
319. N.I. Mock, M.I. Malik, P.J. Stumbo, W.P. Bishop, D.M. Mock, 'Increased Urinary Excretion of 3-Hydroxyisovaleric Acid and Decreased Urinary Excretion of Biotin are Sensitive Early Indicators of Decreased Biotin Status in Experimental Biotin Deficiency', *Am. J. Clin. Nutr.*, **65**, 951–958 (1997).
320. J. Zempleni, D.B. McCormick, D.M. Mock, 'Identification of Biotin Sulfone, Bisnorbiotin Methyl Ketone and Tetranorbiotin-l-sulfoxide in Human Urine', *Am. J. Clin. Nutr.*, **65**, 508–511 (1997).
321. D.M. Bowers-Komro, J.L. Chastain, D.B. McCormick, 'Separation of Biotin and Analogs by High-performance Liquid Chromatography', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 63–67, Vol. 122, 1986.
322. N.G. Hentz, L.G. Bachas, 'Fluorophore-linked Assays for High-performance Liquid Chromatography Postcolumn Reaction Detection of Biotin and Biocytin', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 275–286, Vol. 279, 1997.
323. G.I. Rehner, J. Stein, 'High Performance Liquid Chromatographic Determination of Biotin in Biological Materials After Crown Ether-catalyzed Fluorescence Derivatization with Panacyl Bromide', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie,

C. Wagner, Academic Press, London, 286–295, Vol. 279, 1997.
324. A. Velazquez, S. Zamudio, A. Baez, R. Murguia-Corral, B. Rangel-Peniche, A. Carrosco, 'Indicators of Biotin Status: A Study of Patients on Prolonged Total Parenteral Nutrition', *Eur. J. Clin. Nutr.*, **44**, 11–16 (1990).
325. D. Solberg, I.K. Hegna, 'Microbiological Assay of Pantothenic acid', in *Methods in Enzymology*, eds. D.B. McCormick, L.D. Wright, Academic Press, London, 201–204, Vol. 62, 1979.
326. B.W. Wyse, W.O. Song, J.H. Walsh, R.G. Hansen, 'Pantothenic Acid', in *Methods of Vitamin Assay*, 4th edition, eds. J. Augustin, B.P. Klein, D. Becker, P.B. Venugopal, John Wiley & Sons, New York, 339–416, 1985.
327. C.M. Smith, W.O. Song, 'Comparative Nutrition of Pantothenic Acid', *J. Nutr. Biochem.*, **7**, 312–321 (1996).
328. R.K. Airas, 'Pantothenase-based Assay of Pantothenic Acid', in *Methods in Enzymology*, eds. F. Chytil, D.B. McCormick, Academic Press, London, 33–35, Vol. 122, 1986.
329. K. Banno, 'Measurement of Pantothenic Acid and Hopantenic Acid by Gas Chromatography/Mass Spectroscopy', in *Methods in Enzymology*, eds. D.B. McCormick, J.W. Suttie, C. Wagner, Academic Press, London, 213–219, Vol. 279, 1997.
330. J.H. Walsh, B.W. Wyse, R.G. Hansen, 'A Comparison of Microbiological and Radioimmunoassay Methods for the Determination of Pantothenic Acid in Foods', *J. Food Biochem.*, **3**, 175–189 (1979).
331. H.C. Morris, P.M. Finglas, R.M. Faulks, M.R.A. Morgan, 'The Development of an Enzyme-linked Immunosorbent Assay (ELISA) for the Analysis of Pantothenic Acid and Analogs, Part 1 – Production of Antibodies and Establishment of ELISA Systems', *J. Micronutr. Anal.*, **4**(1), 33–45 (1988).
332. P.M. Finglas, R.M. Faulks, H.C. Morris, K.J. Scott, M.R.A. Morgan, 'The Development of an Enzyme-linked Immunosorbent Assay (ELISA) for the Analysis of Pantothenic Acid and Analogues, Part II – Determination of Pantothenic Acid in Foods', *J. Micronutr. Anal.*, **4**(1), 47–59 (1988).
333. A. Gonthier, P. Boullanger, V. Fayol, D.J. Hartmann, 'Development of an ELISA for Pantothenic Acid (Vitamin B_5) for Application in the Nutrition and Biological Fields', *J. Immunoassay*, **19**(2&3), 167–194 (1998).

Polymers and Rubbers: Introduction

T. Provder
Olmsted Falls, USA

1 Background 1
2 About this Section 2
2.1 Chromatography 2
2.2 Spectroscopy 3
2.3 Morphology 3
2.4 Thermal Analysis/Rheology 4

1 BACKGROUND

The technological directions of polymer-related industries at the beginning of the 21st century have been determined by the operative business and societal driving forces of the last decade of the 20th century. The resultant technological directions affect the product development cycle and shape characterization needs of the materials required. The role of polymer characterization, analysis and testing in the product development cycle is depicted in Figure 1. Product development is no longer a direct straight-line process from product design to product performance, bridged by polymer and product characterization, analysis and testing. The product development cycle is subject to the many constraints produced by the operative temporal business and societal driving forces. These constraints have grown during the last decade and include the following:

- product development costs
- raw material supply
- shorter product development and market introduction cycles
- energy conservation
- waste reduction
- governmental regulations
- safety, health, and environmental considerations
- product quality and consistency
- emphasis on customer needs and satisfaction
- public consumerism
- improved product–process–customer economics
- global competition
- improvement of shareholder value

As can be seen from Figure 1, the role of polymer characterization, analysis and testing is to facilitate product development in a highly constrained environment. For example, polymer product development has been strongly influenced by environmental considerations and government regulations. The ''Clean Air Act'' of 1990 is the driving force behind the development of coatings with significantly lower volatile organic content. Plastic packaging is being developed with built-in environmental degradability. For plastic containers, it has become very desirable to use plastics that can be readily recycled. The Environmental Protection Agency in the United States is increasingly taking a ''cradle to grave'' philosophy with regard to the minimization of organic volatiles generation, minimization of product waste generation and energy usage associated with the chemistries used in the development of a commercial product.

The global operative business and societal driving forces of the last decade have caused polymer-related industries to focus strongly on their core businesses and technological competencies. This focusing has produced a more targeted approach to product development and a significant change in corporate Research and Development culture. The product development process and its associated commercialization is no longer a sequential process from Research and Development to product introduction into the marketplace. The process has become highly nonlinear, nonsequential, and iterative, with many concurrent parallel paths. Often, product Research and Development, process scale-up, plant manufacturing, and quality assurance overlap, occur in parallel and feed back to each other in order to speed up product innovation and shorten the product development and market introduction cycle.

On a more fundamental level, very few new commodity building blocks (monomers) are expected to be developed because of economic, environmental and regulatory considerations. Increased strategic use of low levels of specialty building blocks, coupled with innovations in polymer reaction processes, is expected in order to add value to a polymer product within acceptable economic constraints. In general, structure–morphology–property considerations are becoming of paramount importance in the development of new products. Examples include high-performance engineering plastics and composites that require strategically designed polymers, polymer alloys, and blends; and strategically designed polymers for electronic and biopolymer applications. It is not a question of what building blocks are put together, but how to put them together to make unique polymer products. This approach requires polymer structure–morphology control down to the molecular level in order to enhance and improve properties in a cost-effective manner. A recent example of this approach is the development of new polyethylene materials with unique end-use properties through the application of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

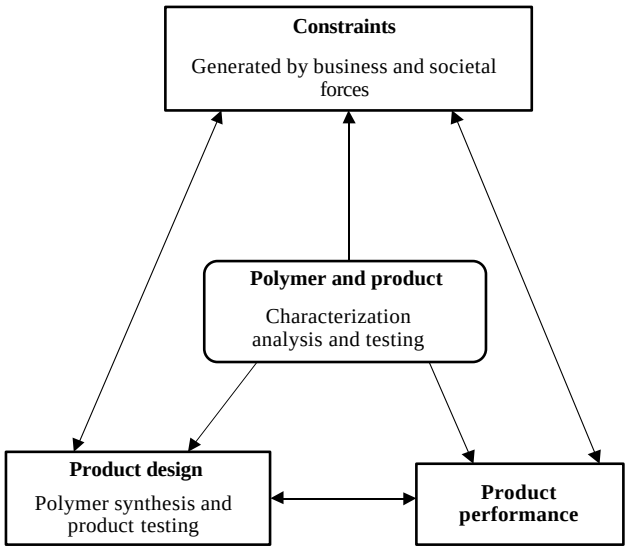

Figure 1 Role of polymer characterization in the product development cycle.

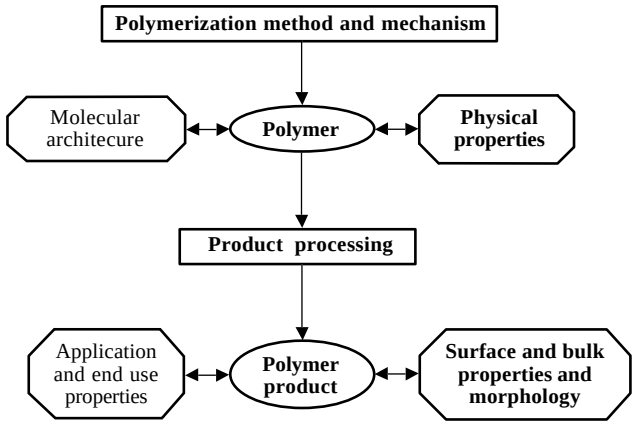

Figure 2 Product development cycle.

metallocene chemistry in conjunction with innovative reaction processes.

The polymer product development process depicted in Figure 2 shows that polymer characterization methodology is required in each step of the process to achieve two goals:

1. characterize the molecular architecture and physical properties produced by a particular polymerization method, mechanism and process;
2. characterize the polymer product resulting from product processing in order to relate surface and bulk properties and morphologies to application and end-use properties.

This process also is highly iterative, with changes sometimes being required in the polymer design or product processing, or both, during the product development and market introduction cycle.

Due to the ever-increasing complex nature of polymeric products, measurement of average properties is no longer adequate to characterize and elucidate the nature of such complex materials. A combination of polymer characterization and analysis methods, the use of ''hyphenated'' characterization techniques and/or multidimensional analytical approaches is required to provide a synergism of analytical and characterization information to establish structure, property and morphology relationships that can form a knowledge bridge between polymerization mechanism and end-use performance.

On the basis of an ever-expanding list of constraints to product development produced by the operative global business and societal driving forces, polymer characterization analysis and testing can be expected to take on an even greater role in the polymer product development cycle during the 21st century. This chapter on the application of polymer characterization and analysis methods to polymers and rubbers is a reflection of that expectation.

2 ABOUT THIS SECTION

This section is organized into four main areas of instrumental applications to polymers and rubbers: chromatography; spectroscopy; morphology; and thermal analysis/rheology. Each of these four instrumental application areas include the key polymer characterization methodologies and techniques that are expected to be highly relevant for analyzing complex polymer systems now and in the future. Each of these four main areas is by no means all-inclusive. However, reference will be made to other sections and articles in this encyclopedia that complement, overlap, reinforce and fill any voids or omissions in the four instrumental applications areas cited above.

2.1 Chromatography

There are seven articles in the chromatography part of this section, which include the following topics: (1) **Gas Chromatography in Analysis of Polymers and Rubbers**, (2) **Inverse Gas Chromatography in Analysis of Polymers**, (3) **Supercritical Fluid Chromatography of Polymers**, (4) **Size-exclusion Chromatography of Polymers**, (5) **Coupled Liquid Chromatographic Techniques in Molecular Characterization**, (6) **Field Flow Fractionation in Analysis of Polymers and Rubbers**, and (7) **Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation**. This is a rather comprehensive section on the application of chromatography methods to the analysis of polymers and rubbers.

Contained in these articles are many examples of the use of hyphenated techniques, manifested by the use of multiple detectors with a given chromatographic method, as well as the use of multidimensional chromatographic methods. The reader also is referred to the *Coatings* section article on **Gas Chromatography in Coatings Analysis** for very closely allied polymer applications. In addition, the reader is referred to some of the theory and instrumentation sections and articles. In particular, the articles in the *Mass Spectrometry* section on **Gas Chromatography/Mass Spectrometry** and **Liquid Chromatography/Mass Spectrometry** are worth noting, as well as many of the articles in the *Liquid Chromatography* section, with particular emphasis on the articles on **Gradient Elution Chromatography** and **Supercritical Fluid Chromatography**. For the reader interested in the process applications of chromatographic methods and their potential applications to polymer reactors, the reader is referred to the section on *Process Instrumental Methods* and the articles on **Chromatography in Process Analysis**.

2.2 Spectroscopy

The spectroscopy part of this section consists of six articles on infrared spectroscopy and two articles on nuclear magnetic resonance spectroscopy. The infrared spectroscopy articles cover the following applications: (1) **Infrared Spectroscopy in Analysis of Polymers and Rubbers**, (2) **Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships**, (3) **Infrared Spectroscopy in Analysis of Polymer Crystallinity**, (4) **Infrared Spectroscopy in Analysis of Polymer Degradation**, (5) **Infrared Spectroscopy in Analysis of Plastics Recycling**, and (6) **Near-infrared Spectroscopy of Polymers and Rubbers**. The infrared spectroscopy articles cover a wide range of applications and do include examples of infrared spectroscopy-hyphenated techniques. Although this section does not contain an article on applications of Raman spectroscopy, the reader is referred to the *Coatings* section for an article called **Infrared and Raman Spectroscopy and Imaging in Coatings Analysis** and to **Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis** in the section on *Electronic Absorption and Luminescence*. The reader also is referred to the theory and instrument section on infrared spectroscopy for details on instrumentation, infrared theory, data analysis and interpretation.

The articles on nuclear magnetic resonance in the section on *Polymers and Rubbers* cover the following applications: (1) **Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers**, and (2) **Nuclear Magnetic Resonance, Imaging of Polymers**. For applications involving the liquid state and polymers and coatings, the reader should see the article on **Nuclear Magnetic Resonance of Coating and Adhesive Systems** in the *Coatings* section. Also, the reader is referred to the theory and instrumentation section on nuclear magnetic resonance and electron spin resonance spectroscopy for details on instrumentation, nuclear magnetic resonance theory, data analysis and interpretation. The article on high-performance liquid chromatography with nuclear magnetic resonance in this section is of particular relevance for polymer-related applications.

Although there are no articles on mass spectrometry and fluorescence spectroscopy in the *Polymers and Rubbers* section, the reader is referred to the articles in the section on *Mass Spectrometry* and the section on *Electronic Absorption and Luminescence* with emphasis on the articles **Fluorescence Lifetime Measurements, Applications of**, and **Fluorescence Imaging Microscopy**. The reader also is referred to the article on **Atomic Spectroscopy in Coatings Analysis** in the *Coatings* section, for its obvious relevance to other polymer systems. For the reader interested in process applications involving spectroscopy, with potential applications to polymer reactors, the reader is referred to the section on *Process Instrumental Methods*, with specific attention to articles on **Near-infrared Spectroscopy in Process Analysis**, **Ultraviolet/Visible Spectroscopy in Process Analyses**, **Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis**, and **Raman Spectroscopy in Process Analysis**.

2.3 Morphology

The morphology part of this section consists of five chapters covering a range of applications with articles as follows: (1) **Atomic Force Microscopy in Analysis of Polymers**, (2) **Neutron Scattering in Analysis of Polymers and Rubbers**, (3) **X-ray Scattering in Analysis of Polymers**, (4) **Positron Annihilation Spectroscopy of Polymers and Rubbers**, and (5) **Pyrolysis Techniques in the Analysis of Polymers and Rubbers**. The reader is referred to the *Coatings* section for the articles on **Microscopy of Coatings** to evaluate how these techniques are used to assess the morphology of polymeric based coatings systems. Additionally, the theory and instrumentation section on X-ray photoelectron spectroscopy and Auger electron spectroscopy should be perused. The reader should also access the application section on surfaces, which has several articles on a range of microscopy techniques which have relevance for polymer surface morphology characterization. One area of morphology analysis not included in this section is the determination of the particle size of polymeric particles. There is an entire application section devoted to particle size analysis. The vast majority of

these articles are highly relevant to the measurement and characterization of polymeric particles.

2.4 Thermal Analysis/Rheology

The thermal analysis/rheology part of this section consists of five articles covering a range of applications that are highly complementary to one another in terms of information that can be generated on polymeric systems. These articles are as follows: (1) **Thermogravimetry of Polymers**, (2) **Dynamic Mechanical Analysis of Polymers and Rubbers**, (3) **Mechanical Properties of Polymers and Rubbers**, (4) **Dielectric Spectroscopy in Analysis of Polymers**, and (5) **Surface Energetics of Polymers and Rubbers, Characterization of**. The *Coatings* section is once again recommended to the reader for its polymer-related articles on **Mechanical Properties of Solid Coatings**, **Thermal Analysis of Coatings** and **Rheology in Coatings, Principles and Methods**. The reader is additionally referred to the theory and instrumentation section on thermal analysis for details on instrumentation, thermal analysis methods theory, data analysis and interpretation.

It is hoped that the articles in this section on applications of polymer characterization analysis and testing will provide the reader with an information and knowledge base relevant to the technical challenges of 21st century product analysis, problem solving and product development.

Atomic Force Microscopy in Analysis of Polymers

S.N. Magonov
Digital Instruments Veeco Metrology Group, Santa Barbara, USA

1 Introduction **1**
2 Scanning Probe Microscopy and Its Development **2**
2.1 Basic Principles of Scanning Probe Microscopy 2
2.2 Atomic Force Microscopy Instrumentation 5
2.3 Atomic Force Microscopy Operational Modes 8
3 Examination of Surface Topography and Local Properties **18**
3.1 Surface Topography and Roughness 18
3.2 Surface Topography in Heterogeneous Polymer Samples 23
3.3 Nanoindentation and Related Techniques 24
4 High-resolution Imaging of Polymer Structures **25**
4.1 Visualization of Single Polymer Chains 25
4.2 Morphology and Nanostructure of Crystalline Polymers 28
4.3 Nanostructure of Oriented Samples 31
5 Compositional Imaging of Heterogeneous Polymer Systems **38**
5.1 Visualization of Amorphous and Crystalline Components 38
5.2 Imaging Polymer Blends 40
5.3 Studies of Microphase Separation in Block Copolymers 44
6 Examination of Polymer Thermal Transitions **46**
6.1 Imaging Polymers at Elevated Temperatures 47
6.2 Visualization of Melting and Crystallization of Polyethylene 48
6.3 Monitoring of Phase Transitions in Mesomorphic Polysiloxanes 52
7 Expectations **55**
Acknowledgments **58**
Abbreviations and Acronyms **58**
References **58**

Atomic force microscopy (AFM), in which a sharp probe is employed for profiling surfaces with unique resolution, has developed into an invaluable multidisciplinary technique for advanced characterization of polymer materials. In its basic function, AFM provides high-resolution imaging of surface structures between the scales of a few nanometers to hundreds of micrometers. This capability is useful for quantitative analysis of surface microroughness of technological surfaces with high sensitivity and accuracy. The direct visualization of polymer structures, from single macromolecules to granular and lamellar nanostructures, to microphase separation and other morphologic patterns, makes AFM an important complementary tool for many of the techniques (light scattering, chromatography, optical and electron microscopy, X-ray diffraction analysis, etc.) employed in polymer analysis.

The advances in AFM of polymers are related to a better understanding of tip–sample force interactions and their use for examination of local sample properties. Different responses of the probing tip to surface locations with different mechanical and adhesive properties are the key for compositional imaging of heterogeneous polymer systems. Compositional mapping with AFM, which does not require any special sample treatment, substantially enhances its microscopic value. Visualization of individual components in block copolymers, polymer blends and semicrystalline materials can be performed with nanometer resolution. In addition to lateral compositional imaging, AFM probing of subsurface structures can be achieved for polymer samples with rubber-like top layers.

A further expansion of AFM applications is expected with the imaging of polymers at different temperatures. Monitoring of polymer structural changes at thermal phase transitions could help in the understanding of polymer melting, crystallization, recrystallization, glass transition and physical aging. In such applications, AFM will complement various thermal techniques.

The review describes the basics of AFM, practical aspects of its applications to polymers and presents a broad range of examples illustrating the main trends in AFM of polymer materials.

1 INTRODUCTION

Polymer macromolecules and polymer materials are characterized by a number of analytical methods that help us learn about their chemical structure, chain organization and mobility, polymer morphology and nanostructure, mechanical and thermal behavior and many

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

other properties. For example, diffraction and microscopy techniques are widely used for examination of structural hierarchy in polymer samples. Electron microscopy is a direct imaging tool that presents structures in real space. Diffraction techniques (X-ray, neutron, light, electron, etc.) are reciprocal space methods and a conversion of their data into a real space demands the use of structural models which might be subjective. Analysis of polymer thermal transitions and mechanical properties of polymers at different temperatures is performed using a number of calorimetric and mechanical methods. These methods and some others (NMR (nuclear magnetic resonance), vibrational spectroscopy, dielectric spectroscopy, optical birefringence and dichroism) indirectly provide valuable information about structural organization in polymer materials. Most of the methods mentioned are well established and routinely applied in research and industrial laboratories. Advantages of these analytical techniques as well as their shortcomings are known, and the latter limit our knowledge about the structure–property relationship in polymer materials.

The increasing importance of these materials in modern technology demands a more detailed characterization of polymers. Progress in polymer synthesis allows precise control of polymer chain structure and, therefore also allows fine tuning of architecture and properties of novel macromolecular structures and polymer materials. Studies of new catalysts, chain structure and its branching, and monitoring the self-assembly of polymer molecules into building blocks of supramolecular architecture, are only a few of the tasks to be accomplished. Also, the rapid development of nanotechnology and the important role of polymer materials in this field requires better understanding of polymer behavior and properties on micrometer and submicrometer scales. This knowledge is invaluable for the engineering of polymer nanomaterials, which will be top players in a variety of technological applications.

Many expectations for advanced polymer analysis are related to the use of AFM. This method was introduced in 1986[1] for high-resolution imaging of surfaces and was first applied to polymer samples in 1988.[2] The original idea was to reach atomic scale resolution in the profiling of surface structures and it was successfully demonstrated on a number of different materials. With the development of AFM applications, it appeared that it is a multidisciplinary technique with applications far beyond those of traditional microscopy. The sharp tip used in AFM, which has an apex tens of nanometers in diameter, can serve not only as a profilometry stylus but also as a tester for probing mechanical, adhesive and even more specific (magnetic, electric, optical, thermal, etc.) sample properties. One of the first important lessons learned from the AFM studies of polymer and biological samples was the need for low tip–sample forces to prevent damage to these soft materials. Since the introduction of an AFM tapping mode, which is most suitable for studies of soft materials, progress in its applications to polymer materials became substantial. Nowadays, this field is of interest to an increasing number of researchers in academic and industrial laboratories.

In this chapter, AFM applications to polymer materials will be overviewed. This goal is challenging because of the rapid developments in this field. Therefore, the main idea is to collect examples with longevity. A short introduction describes the history of scanning probe microscopy (SPM), the family of methods in which AFM is the most important player. Different AFM modes are illustrated by images of polymer samples. AFM applications to polymers will be described by numerous examples, which present direct visualization of polymer chains, nanometer-scale architecture of crystalline polymers, block copolymers, three-dimensional (3D) compositional maps of heterogeneous polymer systems, and polymer morphology together with its changes at thermal phase transitions. From the presented results, the reader will get an idea how AFM complements a large number of other analytical techniques and why for some polymer samples it can provide data which are not accessible by other means. Future expectations in AFM of polymers will be summarized at the end of the article.

2 SCANNING PROBE MICROSCOPY AND ITS DEVELOPMENT

2.1 Basic Principles of Scanning Probe Microscopy

The invention of scanning tunneling microscopy (STM) in 1982,[3] which allows a real-space visualization of surface structures with atomic resolution, has revolutionized microscopy. By placing a sharp metallic probe in the immediate vicinity of a conducting or semiconducting sample and applying an electric potential between these two electrodes, it is possible to detect a tunneling current and to use it for atomic-scale imaging, Figure 1(a). The latter has been achieved using a piezoceramic actuator, which pivots the probe over equicurrent profiles with subansgtrom precision.

A most remarkable example of the first STM applications is the image, which reveals a surface Si (7×7) reconstruction pattern, Figure 1(b). This image unambiguously demonstrated that atoms and atomic-scale defects could be visualized with the new technique and the observed packing motif confirmed one of the structural models of reconstruction on the Si surface. Initially, STM measurements were conducted in UHV (ultrahigh vacuum) but similar studies could also be performed under ambient conditions. Such possibilities have

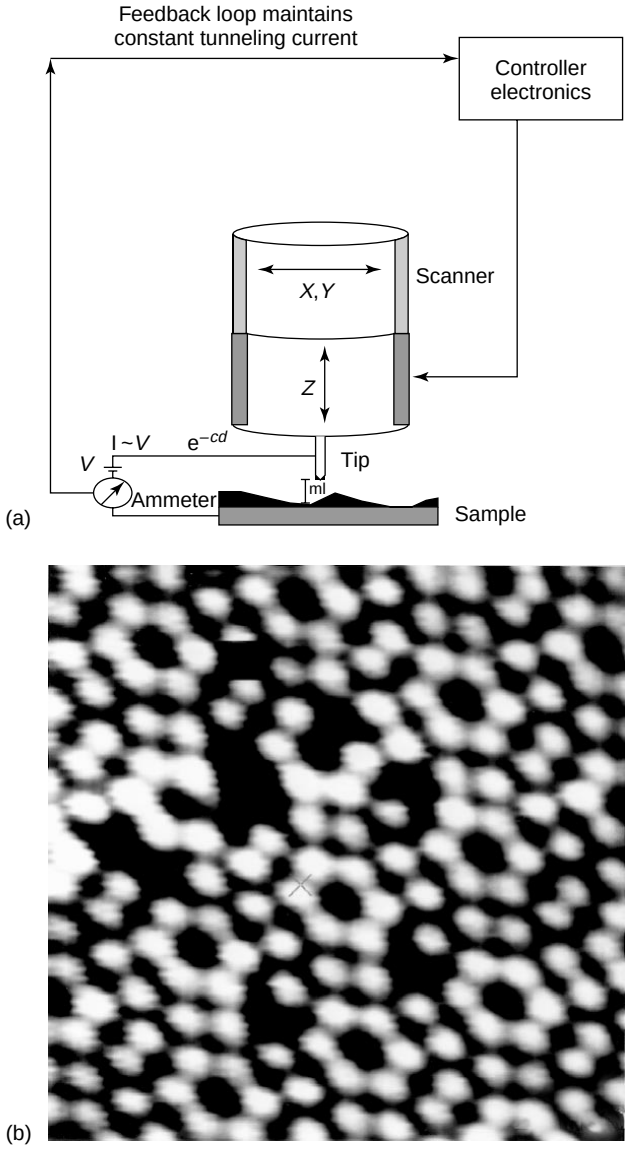

Figure 1 (a) Set-up for STM, in which a metallic tip and a conducting sample are the two electrodes. Bias voltage is applied to the electrodes and the detected tunneling current is employed for surface imaging. (b) Atomic-scale STM image of a Si (7 × 7) surface, showing surface reconstruction. Scan size 10 nm × 10 nm. (STM image courtesy of Prof D. Sarid.)

attracted a large number of researchers to this field, especially after the first commercial scanning tunneling microscopes became available in 1988. It has been established that STM images actually present maps of partial electron density (electron density at the Fermi level) measured at the tip–sample separation.[4] Therefore, the rational interpretation of STM images demands a theoretical analysis,[5] and a direct correlation between the image and surface topographic structure might be misleading. Another difficulty of STM is related to the possibility of strong tip–sample force interactions, which are often not taken into account. Because the tip force could induce surface deformation, on some samples the STM images have been actually obtained on the deformed surfaces.[6] Despite these difficulties, STM is used for surface studies of conducting and semiconducting materials in a large number of research laboratories worldwide. STM studies of conducting polymers and carbon materials (carbon fibers, carbon black, etc.) have been performed to visualize surface topography, morphology and molecular-scale order.

The invention of AFM[1] was initiated by a natural desire to extend STM applications by making possible atomic-scale imaging of nonconducting materials. Instead of tunneling current, a more universal tip–sample force interaction was employed for surface imaging in AFM. A simple optical lever method[7] was successfully used for detection of small forces between the apex of a microfabricated tip, which is positioned at one end of a tiny cantilever, and a sample surface, Figure 2(a–c). In this method, the bending of the cantilever due to small tip–sample attractive and repulsive forces is magnified on a positional photodetector and a related signal is used for profiling the sample surface at a constant force level. Because the tip–sample force is proportional to the total electron density of the top surface atoms, AFM images more directly relate to surface topography than STM images. In contrast to STM applications, the first AFM studies were conducted under ambient conditions, and only later the possibility of AFM measurements in UHV was demonstrated. Imaging samples under liquid and in electrochemical environments are also possible with both techniques.

With developments in AFM applications the validity of this technique became evident. High-resolution profiling of surfaces has facilitated observations of atomic-scale surface steps, atomic and molecular lattices in a large number of inorganic and organic crystals.[5] Observations of single DNA molecules by AFM have promoted its application to different biological samples. It is also understood that the tip–sample force interactions in AFM should be considered while imaging soft materials (polymer and biological materials). Normal and lateral forces between the tip and the sample in contact AFM mode are of tens and hundreds of nanonewtons. Minimization of these forces is needed for nondestructive imaging of soft samples. This has been achieved in imaging under liquid.[8] With the introduction of a tapping mode,[9] in which lateral forces are practically eliminated, AFM studies of soft materials have further broadened.

The success of STM and AFM applications led to the development of a number of related SPM techniques, which are used for mapping surface magnetic, electric, thermal, and optical properties. These methods are also known as magnetic force microscopy (MFM),

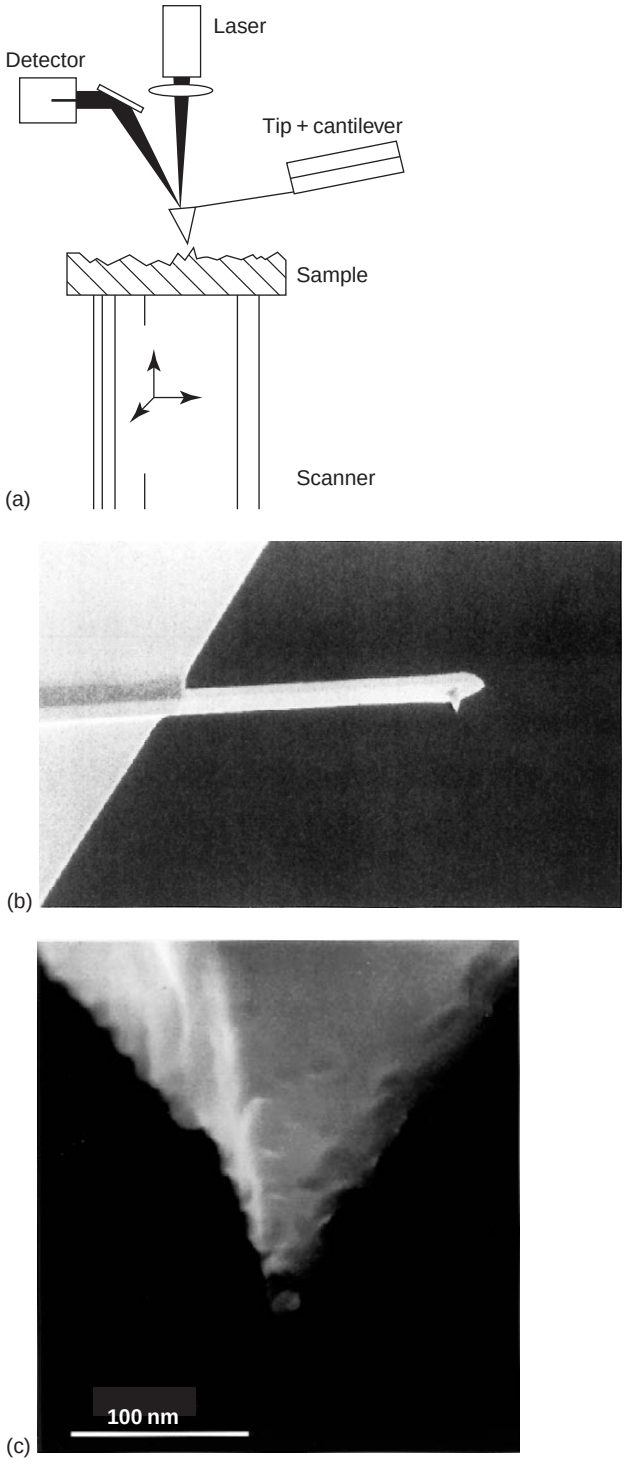

Figure 2 (a) AFM set-up showing optical lever detection employed in AFM for detection of the cantilever deflection caused by tip–sample force interactions. (b) Scanning electron micrograph of etched Si probe, which consists of a 125-μm long cantilever and a sharp tip. (c) Apex of a Si tip.

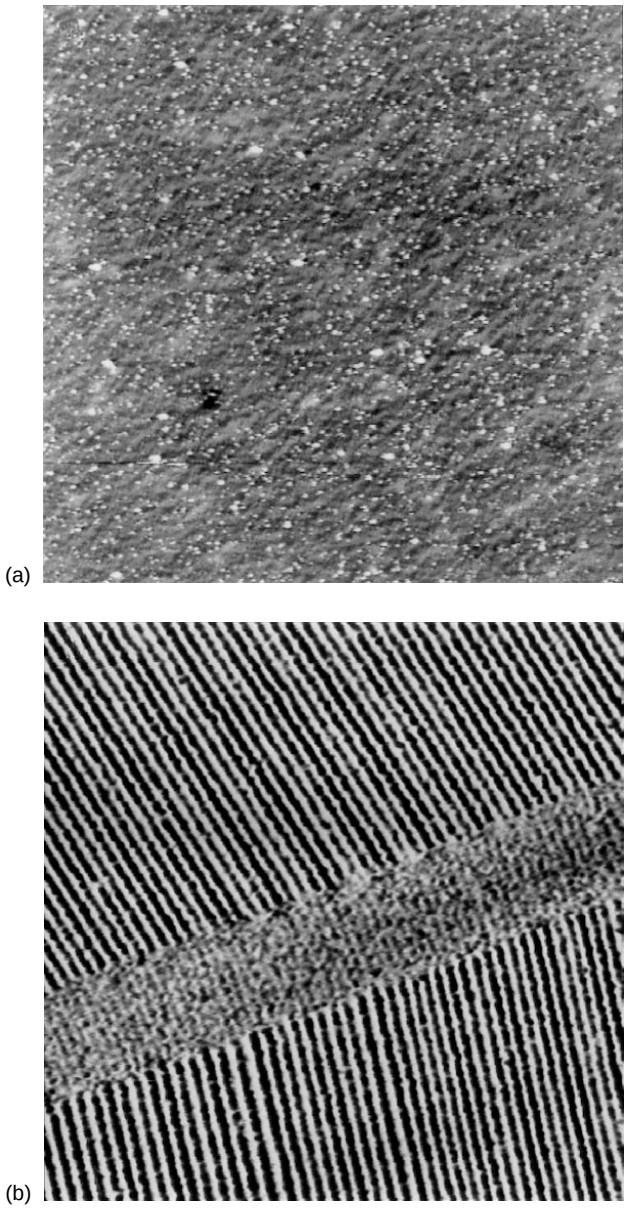

Figure 3 Height image (a) and phase image (b) of a magnetic video tape obtained by MFM. The images were simultaneously recorded on 25 μm × 25 μm area. The height image shows surface topography and the frequency image reveals magnetic pattern.

electric force microscopy (EFM), scanning capacitance microscopy (SCM), scanning near-field optical microscopy (SNOM), and scanning thermal microscopy (SThM), and so on. A large number of applications have been developed with these techniques, however, a separation of the effects of these interactions and surface topography is a key problem in such applications. In MFM and EFM this problem is solved using a two-pass lift technique.[10] In this technique, the topography is measured in the first scan and magnetic or electric properties (using appropriate probes) are measured in the second scan over the same area. In the second scan, the

probe motion reproduces the measured surface profile (feedback is off) but with a vertical lift of 20–50 nm. Changes in frequency or phase of the oscillating ferromagnetic probe during the second scan reveal magnetic surface structures or surface variations in dielectric constant when a conducting probe is used. In this way, the effect of long distance electromagnetic forces is separated from those of mechanical tip–sample forces. The height and frequency images in Figure 3(a) and (b) give a practical example of how surface topography and magnetic pattern of video magnetic tape are visualized in MFM. The magnetic structures of the film are seen in the frequency image. This was an example showing that the measurement of the surface topography is a practical requirement of all scanning probe methods.

There is no doubt that AFM is the leading scanning probe technique. It allows examination of practically any kind of sample and is applied in increasing numbers of industries, while other SPM techniques are rarely used for industrial applications. So far, the semiconductor and data storage industries have benefited most of all. Automated AFM systems have been introduced for on-line control of roughness of Si wafers[11] and of pole tip recession in magnetic heads.[12] The characterization of polymer materials with AFM, which substantially complements more traditional microscopic and diffraction methods, is developing rapidly in research and development laboratories in the plastics, rubber, paint, biomaterials and paper industries.

2.2 Atomic Force Microscopy Instrumentation

A generic scheme of the atomic force microscope is shown in Figure 4. In this configuration, a piezoceramic scanner carries the probe over the sample. In the alternative configuration, Figure 2(a), the scanner moves the sample whereas the probe is fixed. Scanners include two cylinders, one assigned for vertical z-travel, another for lateral x, y motion. Because of nonlinearity, creep and other features inherent in polycrystalline piezoceramic materials, a linearization of the 3D motion of the scanner is performed with software correction. Different lateral and vertical standards are used for scanner calibration. After that, measurements of lateral dimensions can be performed with angstrom resolution and measurements of vertical dimensions can even attain subangstrom resolution. The maximal area, which is commonly covered by piezoceramic scanners, is ~150 μm on the side. Surface corrugations up to 5 μm in height can be profiled. Additional $X-Y$ stages are employed for positioning surface regions of interest under the probe.

Another key feature of AFM, which defines its unique capabilities for surface imaging, is the high force sensitivity of the optical lever detection. Surface imaging is realized with the electronic feedback system, which utilizes a signal from the four-segment optical positional detector. (Vertical segments are used for measurement of the cantilever deflections caused by normal tip–sample force. Horizontal segments are used for measurement of the cantilever torsion that is due to lateral tip–sample forces.) This signal, often called the error signal, is the difference between the actual value of the normal tip–sample force and its set-point value chosen by the operator. The differential signal is used to adjust the vertical probe–sample separation and keep the tip–sample force interaction constant during scanning. All these functions are realized using an electronic controller, which also communicates with an image acquisition computer. The variations in the cantilever–sample separation are converted into a height image with bright (dark) contrast assigned to high (low) surface locations.

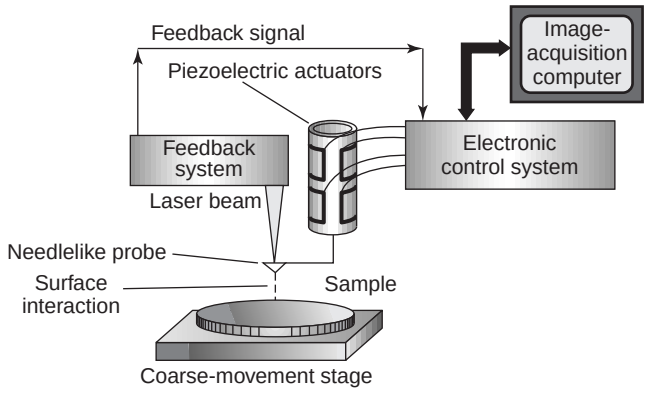

Figure 4 Generic scheme of atomic force microscope with its main components.

Commercial scanning probe microscopes are manufactured by a number of companies in the USA, Japan, the United Kingdom and Russia. The most popular microscopes are the Nanoscope IIIa™, MultiMode™ and the Dimension™, which are shown in Figure 5. Both instruments can be equipped with tensile stages and heating accessories. The MultiMode™ is a compact microscope, which provides a whole spectrum of AFM modes and requires only minimal vibration and acoustic isolation. The limited size of the sample ($\leq$ 2 cm in diameter) is a limitation, which is overcome by the Dimension™ microscope. A close-up of the Dimension™ AFM head and a large sample (8 inch, or 20 cm, wafer) is shown in Figure 5(b). However, with the increase in instrument size more protection from acoustic and vibrational noise is necessary. This is accomplished by placing the microscope into a special enclosure, Figure 5(c). With the further use of AFM instruments in the semiconductor industry, the size of microscopes used

is increasing making possible examination of 300-mm wafers and their automatic on-line handling by robotic systems.

2.2.1 *Atomic Force Microscopy Probes*

Microscopic probes are invaluable components of all SPM instruments, and their quality substantially defines the practical value of these techniques. There are a large variety of AFM probes (the integrated cantilever and tip) that are commercially manufactured. Most of them are microfabricated from Si_3N_4 or Si. Various coatings are applied to these probes in order (1) to increase a reflectivity of their backside for improved detection of the bouncing laser beam, (2) to provide electric or ferromagnetic properties of the probe, (3) to implement a thermosensitive element at the apex, (4) to improve the tip apex wear. The cantilevers have triangular or rectangular shapes and their length varies in the 100–400-nm range. The tips are pyramidal or conical and several micrometers in height. The opening angle of these tips varies in a range between 30–90° and the apex radius can be as small as 5 nm. Specially prepared tips are required for profiling deep trenches (hundreds of nanometers wide and up to several micrometers deep). Triangular-shaped Si_3N_4 probes are most often applied to contact mode measurements and their spring constants are in the 0.1–0.01-N m^{-1} range. There are a larger variety of etched Si probes. Their rectangular-shaped cantilevers have stiffness in the broad range from 0.1 to 500 N m^{-1} and their resonant frequencies are in the range from 10 to 500 kHz. Soft probes are applied to contact mode measurements, whereas stiff probes are required for the tapping mode in order to retract the tip from a sample surface in each oscillation cycle. (Adhesion and capillary forces caused by liquid contamination layers attract the tip to the surface.) However, stiff probes might cause a surface deformation of polymer samples even in the tapping mode. Therefore,

(a)

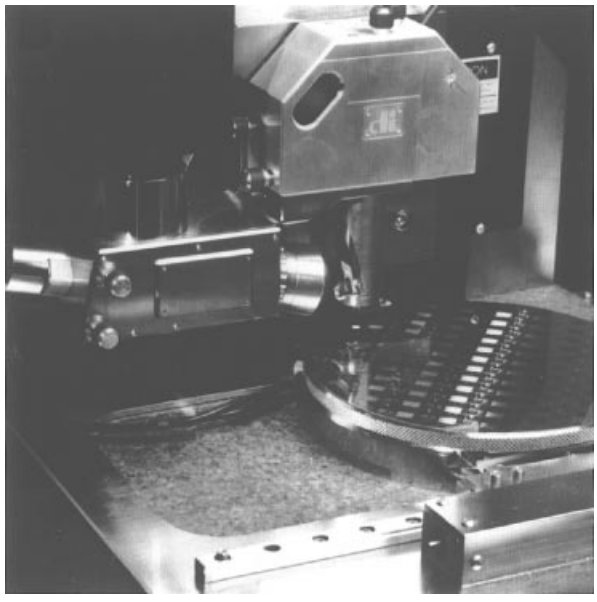

(b)

(c)

Figure 5 (a) MultiMode™ atomic force microscope. (b) Close-up of the head and sample in the Dimension™ 3100 AFM microscope. (c) The Dimension™ 3100 AFM with an acoustic hood, electronic blocks, computer and monitors.

Figure 5 (*Continued*)

an optimal choice of the probe is needed for imaging of a particular polymer sample. The quality factor (Q) of the probe cantilever is another parameter to be considered in tapping mode applications. High-Q cantilevers are more sensitive to small force variations, whereas low-Q cantilevers are more suitable for precise height measurements in the tapping mode and also for fast scanning.

Besides the probes described, which are routinely used in AFM studies of polymers, several types of modified tip are known. For example, chemically modified probes can be prepared by coating the Si tips with gold and by chemisorption of thiolo-organics on gold surfaces. In these ways, tips with different chemical functionality and different physical and chemical properties can be prepared. Applying such probes one might expect detection of specific force interactions between the modified probes and surface locations with a different chemical nature. There are a number of results, which indicate such effects.[13] However, it is difficult to assign these results exclusively to the different chemical natures of these surface regions without considering differences in their mechanical and adhesive properties. The latter makes the interpretation of these effects much more complicated. Another problem of chemically modified Si or Si_3N_4 probes is the integrity of the coating at the apex during scanning that cannot be checked directly.

2.2.2 Sample Preparation

The attractive features of AFM applications to polymers and other materials include the ability to perform measurements under ambient conditions or under liquid with minimal sample preparation. Many polymer samples (thin layers, films, fibers) can be examined without any preparation. Spin-cast films and those prepared by dipping the substrate into polymer solutions are proper samples for AFM. In studies of the morphology and nanostructure of crystalline polymers, the flat sample surfaces can be prepared by melting a piece of polymer on flat substrate (mica graphite, Si wafer) and cooling. Afterwards a crystallized sample can be separated from the substrate, and the flat polymer surface which was in contact with the substrate, is most suitable for AFM studies. For polymer materials, which are prepared by extrusion, compression or injection molding, flat sample surfaces for AFM study can be made by ultramicrotome. In contrast to TEM (transmission electron microscopy), which requires preparation of ultrathin sections, polymer blocks with a flat surface are proper samples for AFM imaging. However, preparation of extremely smooth polymer surfaces is a not trivial procedure and it requires an optimal choice of knife material, and optimal temperature of the knife and the sample. AFM images reveal the imperfections of this sample preparation (scratches, chattering patterns, contamination traces, etc.) Freeze fracture of some polymer samples also can result in flat surface areas.

Another important advantage of AFM studies compared with electron microscopy is that there is no need for sample staining or coating with a metal layer. Differences in mechanical and adhesive properties of components in heterogeneous polymer systems, such as semicrystalline polymers, block copolymers, polymer blends and composites are employed for compositional AFM imaging. Compositional mapping of polymer blends and composites is successfully performed on flat polymer surfaces prepared using an ultramicrotome. However, caution should be exercised in observations of submicrometer structures and, particularly, features in the tens of nanometers size range because these features can be at least partially due to material fracture that occurs during cutting. This area has not been explored yet.

2.2.3 Imaging Artifacts

AFM, as well as other microscopic methods, is not free of various artifacts. Imperfections could be caused by the inexperience of the operator and related to choice of nonoptimal imaging parameters, including set-point values, scanning rate, and gains. Samples with varying roughness require adjustment of these parameters and tracking of the surface steps might be difficult. The ability to measure surface profiles while scanning in different directions (from $-X$ to $+X$ or from $+X$ to $-X$) helps to clarify these effects and to avoid them by optimizing scanning parameters. Also specific features of piezoceramic scanners such as creep, nonlinearity, and cross effects between different axes should be taken into consideration. These effects are enhanced when the sample is tilted with respect to the microscope plane.

Most obvious artifacts are clearly seen in the images. Among these are double tip effects, when additional ghost patterns appear together with the real ones. A bad tip causes these artifacts. This might happen because of a manufacturing defect, traces of sample material which have stuck to the tip, or because of tip damage. The tip apex can be damaged and flattened while scanning hard samples. A sharp Si probe becomes dull after scanning on a Si wafer at elevated tip forces. Then imaging of surface features, which are sharper than the dull probe, leads to the appearance of the tip profiles in the images. This is inverse imaging: the sample images the tip. Periodical linear patterns with a repeat distance of 1.5–2.5 μm, which are often seen in the images of flat samples such as mica are caused by optical interference, which occurs on the surfaces of the sample and the cantilever. Adjustment of

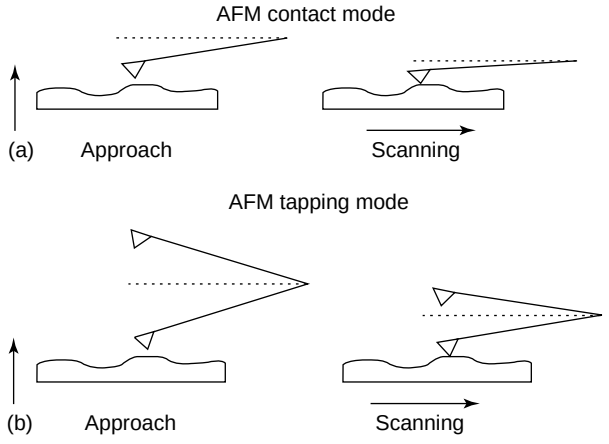

Figure 6 Main AFM operation (a) tapping, (b) contact modes. Note that in the tapping mode the cantilever base keeps the same deflection before and after engagement/scanning.

the position of the detecting laser beam on the cantilever minimizes this effect.

2.3 Atomic Force Microscopy Operational Modes

Several operation modes have been developed in AFM. The main imaging modes, contact mode and oscillatory or tapping mode differ in the way the tip interacts with the sample, Figure 6(a) and (b). In the contact mode, the tip scans a sample surface while being in permanent contact with that surface, Figure 6(a). The cantilever deflection is a measure of the force which is exercised by the sample. The tip–sample force is maintained constant during scanning by an electronic feedback. The correct choice of set-point force is an important decision to be made by the microscopist. If a polymer sample is heterogeneous with components which have different mechanical properties, then with an increase in tip-force the height image reflects a larger tip-induced surface deformation of the weaker component. This is demonstrated in the height images, which were obtained in the contact mode at different tip forces on the sample of multilayer polyethylene (PE), Figure 7(a–c). In this sample, layers of PE with different number of branches and different densities (0.92 g cm^{-3}, high-density material; 0.86 g cm^{-3}, ultralow-density) alternate with each other. Narrow strips of high-density polyethylene (HDPE) are elevated in height images recorded at high forces. This is a result of a strong tip-induced deformation of weak layers of ultralow-density polyethylene (ULDPE). In varying the tip force, one should check if such deformation is elastic or it might cause sample damage. Therefore, in applying contact AFM mode to soft materials, one should be aware of how accurately the height image presents a surface topography.

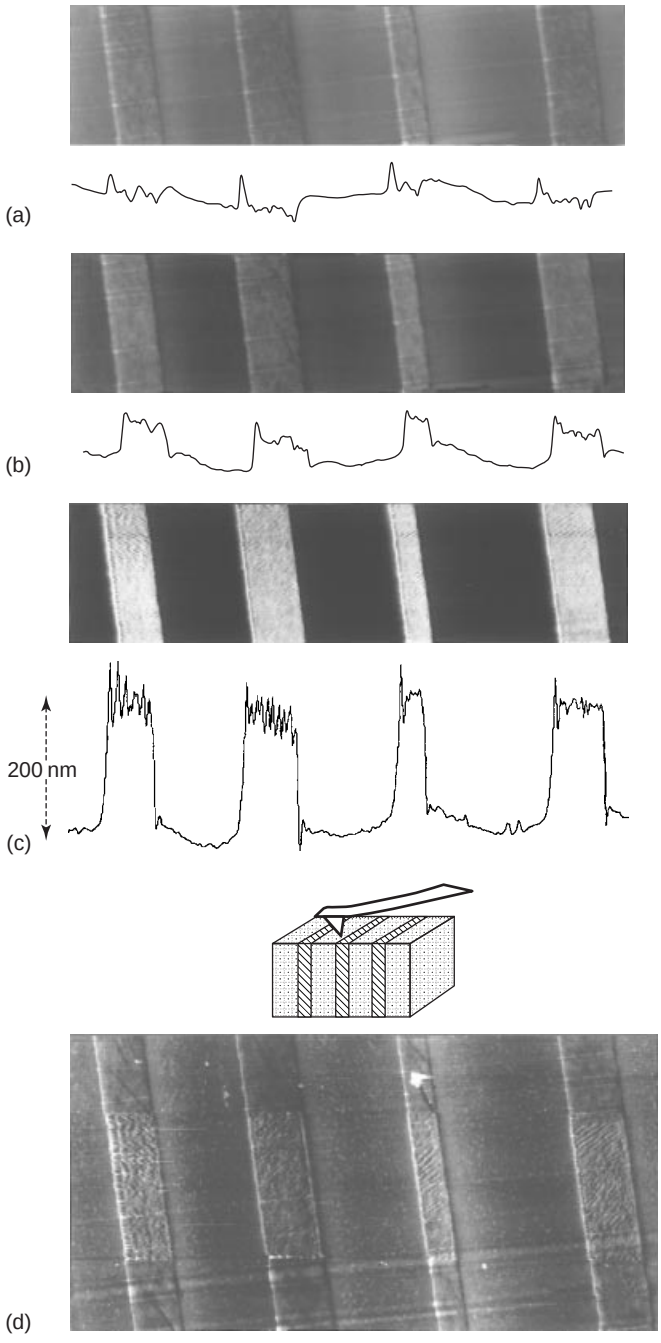

Figure 7 (a–c) Height images of multilayer PE obtained at different tip-forces (ca. 10, 30 and 80 nN) over the same area (60 μm × 15 μm). Below the height images are the cross-section profiles taken across the layers in the middle of the area shown in the images. (d) Height image of a larger area (60 μm × 30 μm) of the same sample obtained in the tapping mode. The surface region, which was previously scanned in contact mode, is in the center of the image.

In the tapping mode (Figure 6b), the tip comes into contact with the sample in every cycle of its oscillatory motion at (or near) the resonant frequency of the

cantilever. A drop in the cantilever amplitude when the sample is brought close to the probe is used to measure tip–sample forces. During scanning in the tapping mode, the probe comes only into intermittent contact with the sample surface. As a result, more soft samples can be examined in the tapping mode than in the contact mode. This is demonstrated by the height image (Figure 7d), which was obtained on a multilayer PE sample with the same probe that was used for contact mode images in Figure 7(a–c). In Figure 7(d), the area previously scanned and damaged in the contact mode is seen. Variations in the tip–sample force interactions in the tapping mode did not lead to damage of this sample. The advantages and disadvantages of both modes will be described below.

Probing the mechanical or adhesive sample properties with the AFM tip can be performed locally, at one surface point, without scanning. Force curves are recorded for this purpose. Force curves describe the probe deflection (contact mode) or variations in the cantilever amplitude (tapping mode) versus the vertical travel of the sample or the probe base towards and away from each other. Depending on the level of the tip force the surface deformation can be elastic or plastic. In the latter case, the tip makes an indent on the sample. Special probes with diamond tips, which have an apex of several tens of nanometers, are employed for such studies. After indenting, the surrounding area is scanned and the indent dimensions are determined from the AFM image obtained. Such a procedure is the basis of a nanoindentation technique that is widely used for measurements of local mechanical properties of different materials and coatings. Examples of local mechanical measurements of polymers will be discussed in section 3.3.

2.3.1 Contact Mode

In this mode, lateral scanning of the probe over the sample surface is initiated when the vertical bending of the cantilever, which is due to the repulsive tip–sample force, reaches the set-point level chosen by the operator. After this happens, lateral motion of the probe is initiated by the feedback control. Continuous motion of the probe from the $-X$ to $+X$ direction (fast direction) proceeds with small advances in the Y direction. Scanning rates are typically in the 0–3-Hz range and can be higher when scanning is performed on atomically flat surfaces or in submicrometer scan areas. Imaging is performed continuously and is monitored on a computer screen. Single images (typically 512×512 data arrays) are collected as data files. Various filtering and data analysis procedures can be applied to AFM images. A height image of banded spherulite of poly(vinylidene fluoride) (PVDF) is shown in Figure 8(a) which reflects the surface topography of the sample. Some details of topography are better seen in the deflection (or error signal) image, Figure 8(b), which resembles a map of derivatives of topography scan lines. For many practical applications, it is also useful to record the variations of the lateral tip–sample forces. To emphasize friction effects, imaging should be performed with the slow scanning direction perpendicular to the cantilever length. Lateral force images, which are collected simultaneously with the height images, reveal surface regions that have different mechanical and adhesive properties. This is demonstrated by the images of a rubber sample, shown in

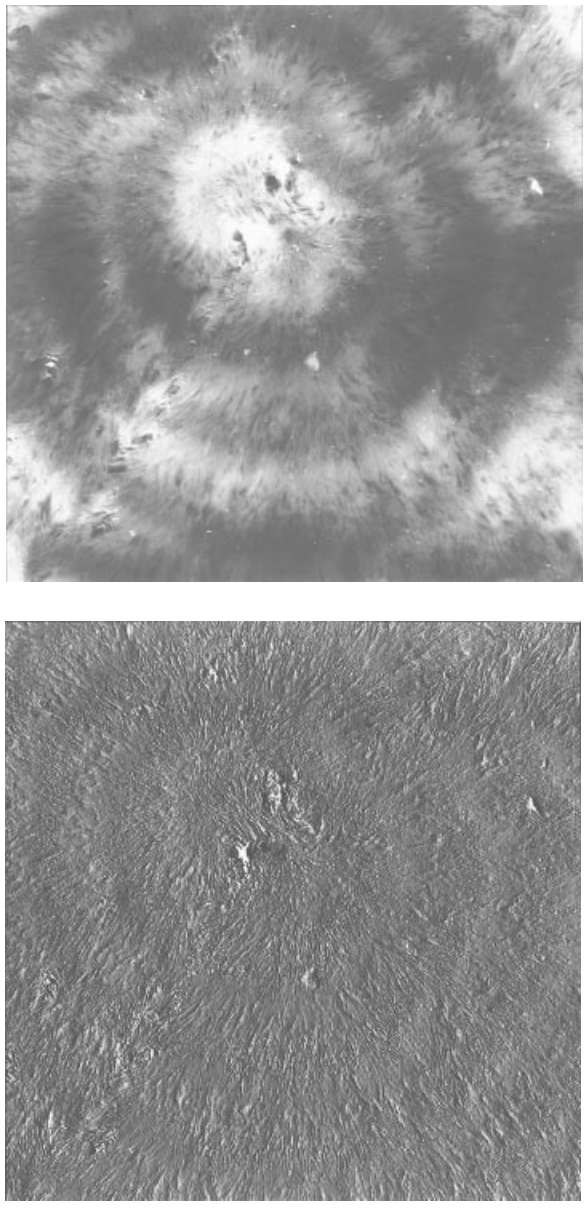

Figure 8 Height (a) and deflection (b) images of banded spherulites of PVDF obtained in contact mode. The images were recorded simultaneously on surface area 50 μm × 50 μm.

Figure 9(a) and (b). Areas of different contrast, which are seen in the lateral force image, correspond to different components of this blend. Such differences might not be distinguishable in the height image.

The contact mode has proven itself capable of imaging atomic- and molecular-scale periodical lattices. Such images have been obtained for a large number of organic and inorganic crystals and also for polymers. Molecular-scale images of polymers were obtained on oriented samples of PE,[14] polytetrafluoroethylene (PTFE),[15] polypropylene[16] and polybutene-I.[17] The periodical patterns can be obtained in the height, deflection and lateral force images. The latter might provide a better contrast of the atomic-scale features. One such image is shown in Figure 10(a), where a periodical pattern of oriented PTFE is seen. Filtering of the image using a fast Fourier transform (FFT) procedure, Figure 10(b), gives a much clearer view of the chain arrangement. The interchain distance (0.56 nm) and the pitch (0.82 nm) along the polymer chains are consistent with X-ray data, which reveals a slightly twisted conformation of PTFE chains. Another AFM image in Figure 10(c) shows oriented chains of PE, which were revealed on the surface of highly oriented sample.

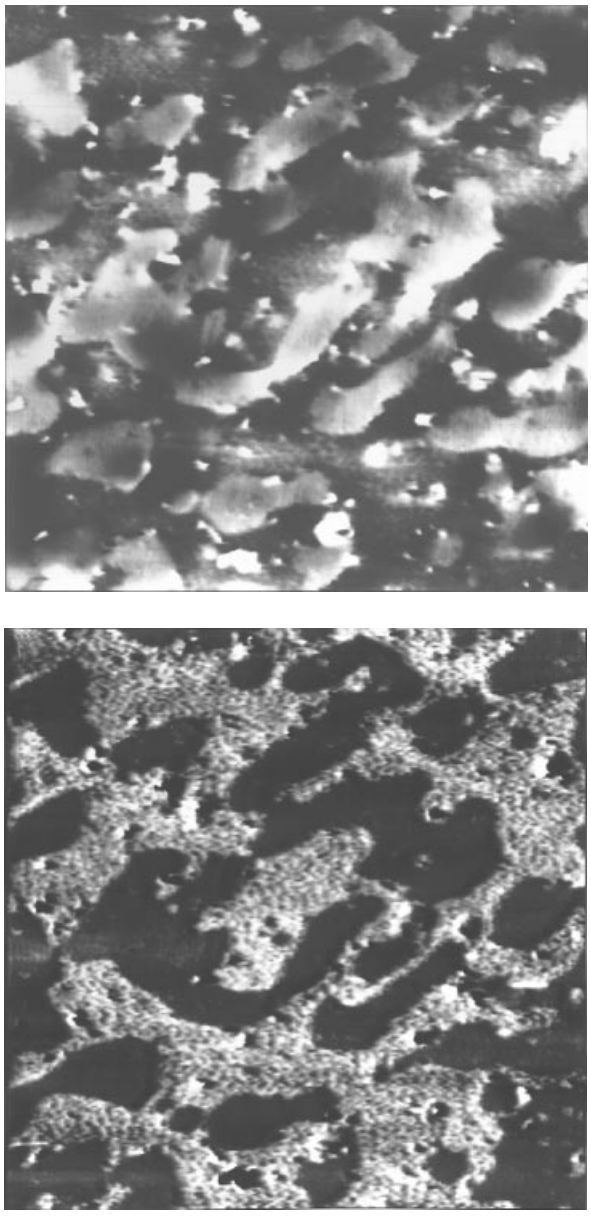

Figure 9 Height image (a) and lateral force image (b) of a blend of synthetic and natural rubber. The images were recorded simultaneously on surface area 12 μm × 12 μm.

Several AFM studies aimed to check molecular packing in polymer single crystals[18] and in samples prepared by epitaxial crystallization.[19] Images of normal and cyclic alkanes, which are short chain analogs of PE, reveal molecular-scale packing of their crystals in full agreement with the crystallographic data. Crystallization of long chain molecules proceeds via chain folding and single crystals of a number of polymers prepared from dilute solutions are actually thin lamellae with a well-defined crystallographic shape. An essential question, which has not yet been answered is, how perfect is the molecular arrangement of the lamellar surface? Different structural defects, such as chain ends, loops, and ''switchboard'' chain segments, could cause deterioration in the perfect adjacent-reentry arrangement expected for perfect crystals. A number of atomic-scale images, which show periodical lattices of single crystals of polymers (PE,[18] poly(ethylene oxide)[20]) are most likely to have been obtained with the probe penetrating through the topmost surface layer (or depressing it) and thus revealing chain packing inside these crystals. Similar data have been obtained with X-ray diffraction techniques. Imaging of polymer crystal surfaces at low tip–sample forces shows a grain nanostructure of lamellar surface,[21] which will be discussed in section 5.1.

It is worthwhile to consider critically the capabilities of atomic-scale images in contact AFM mode. When applied to thin and ultrathin layers (thickness above and below 100 nm, respectively), small and imperfect crystals, atomic-scale AFM images give structural information that might be difficult and even impossible to obtain by diffraction techniques. However, there is a definite limitation to such studies that is related to the fact that such imaging is not true atomic-scale imaging. It is most likely that several apex atoms stay in contact with the surface when the tip moves and the tip–force variations reflect the corresponding surface lattice. Therefore, in contact AFM mode observation of atomic-scale defects (e.g. vacancies) is not achieved. Later, the atomic-force defects were observed in UHV studies of several semiconductors (including Si (7 × 7) surface) when

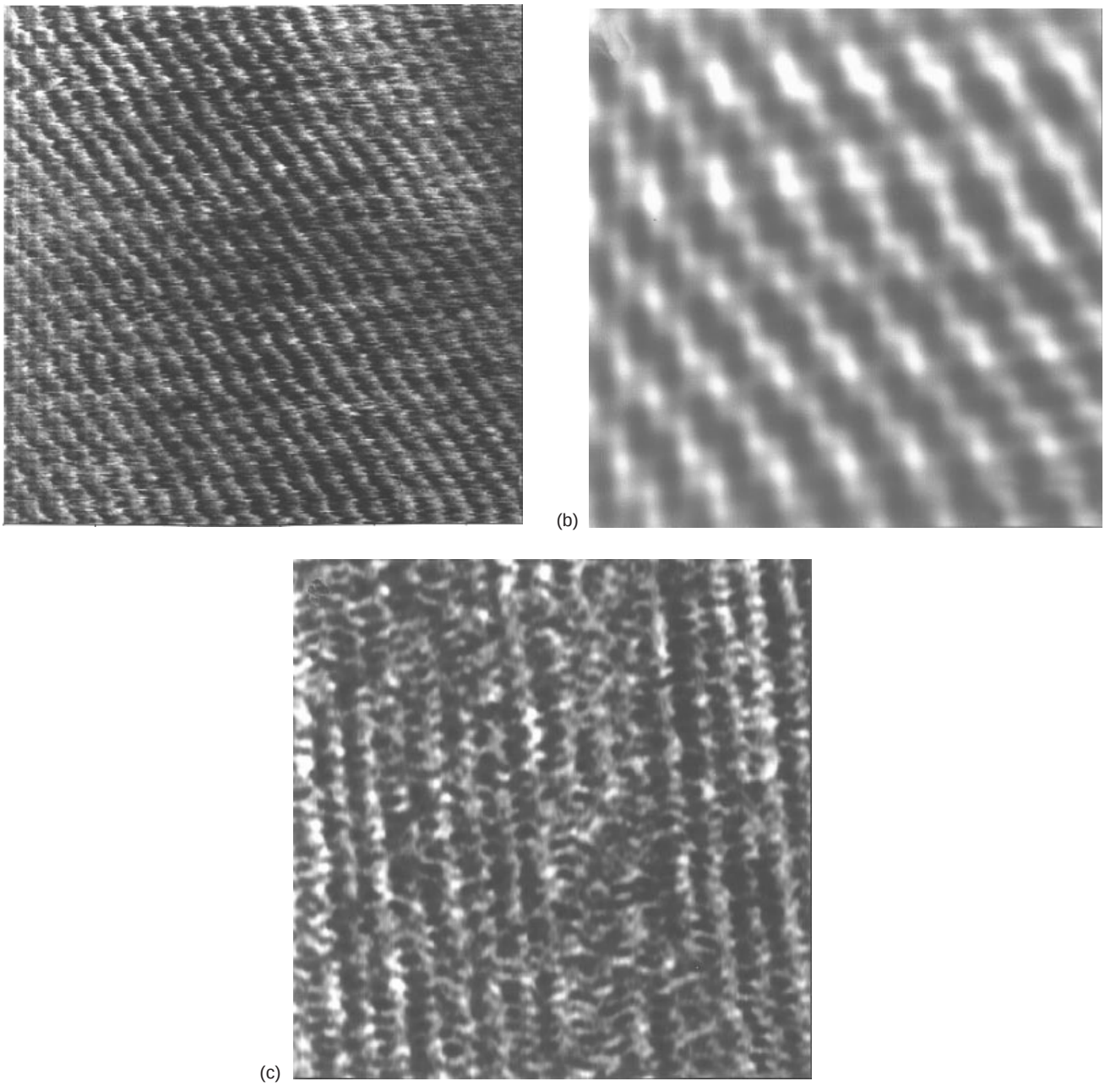

Figure 10 (a) Lateral force image of PTFE chains in oriented layer deposited by rubbing. Image size 14 nm × 14 nm. (b) A part of the image in (a) after FFT filtering. Image size 5 nm × 5 nm. (c) Deflection image obtained on a surface of highly oriented PE prepared by cold extrusion. Image size 8 nm × 8 nm.

the oscillatory AFM mode was applied in so-called noncontact regime.[22] An UHV environment, which is essential for detection of ultrasmall tip–sample forces, is a serious obstacle to the broadening of such applications.

2.3.2 *Force Curves*

Following the first applications of contact AFM mode to polymers and, particularly, to biological objects it has become evident that tip–sample forces lead to surface modification and sample damage. Therefore, one should consider the tip–sample forces in more detail using force curves. The force curve describes the cantilever deflection as a function of the scanner travel, Figure 11(a). (In practice, force curve measurements are performed at frequencies in the 1–100-Hz range.) Upon approach of the sample, a probe initially experiences a small attractive force, which causes the cantilever to bend

downwards. Then the tip comes into contact with the sample and the cantilever deflects upward, reflecting repulsive tip–sample forces. When the sample is retracted from the probe, the tip is stuck to the sample by the capillary force. A layer of liquid contamination that presents on samples in air causes this capillary force. Therefore, a pull-off force is needed for tip disengagement and this causes hysteresis in the force curves. In some cases, the pull-off force might be used as a measure of adhesion between the probe and the surface.

The contributions of different forces acting on the probe and the samples are shown schematically in Figure 11(b). From this picture it is clear that under ambient conditions the tip force includes the capillary force of the liquid contamination layer. By adjusting the set-point deflection to the attractive force level near pull-off point, the imaging can be achieved with the lowest force possible for a given probe. However, this condition is unstable owing to possible tip disengagement and is not realized practically. Under these conditions, the capillary effect limits the minimal force for imaging in contact mode to ~5 nN. By placing the probe and the sample under liquid the capillary force could be avoided. This is

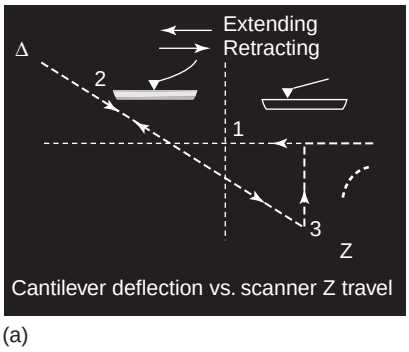

(a)

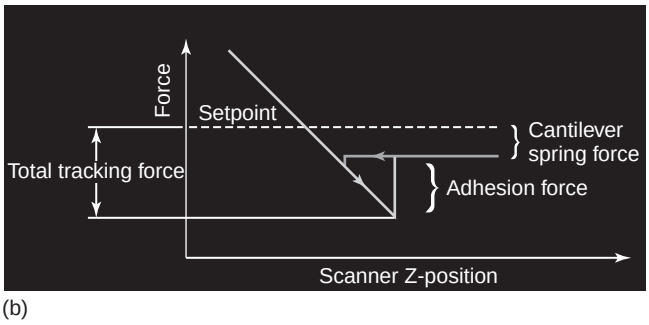

(b)

Figure 11 (a) When the sample approaches the tip (←, right-to-left) attractive force (point 1) slightly bends the probe down (Δ, cantilever deflection). Further sample travel will induce upward bending of the cantilever caused by repulsive forces (point 2). On reverse travel (→), the tip is stuck to the sample until the probe spring stiffness overcomes the adhesion/capillary forces (point 3). (b) Scheme of different forces acting on a probe.

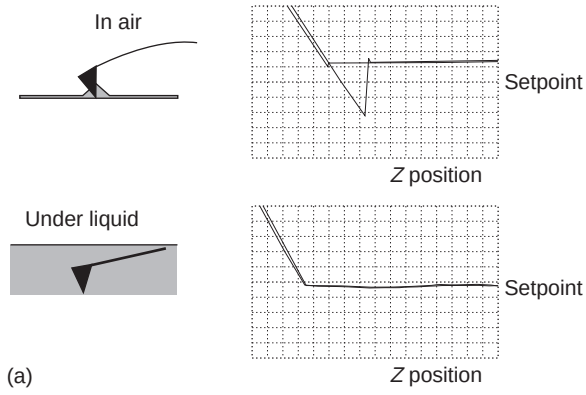

(a)

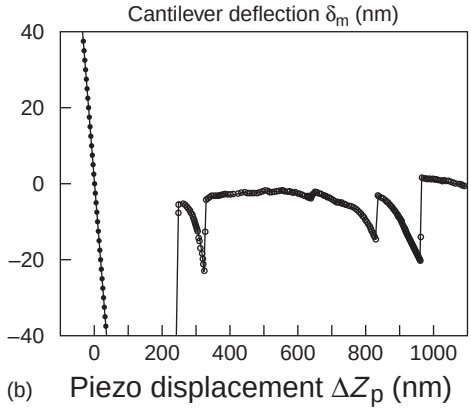

(b)

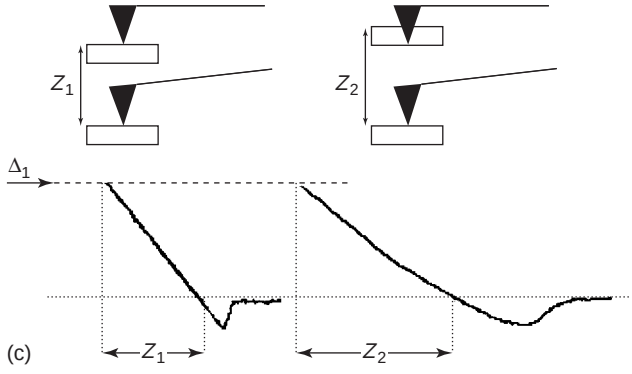

(c)

Figure 12 (a) Force curves recorded in the contact mode in air and in liquid. (b) Force curve obtained on grafted poly (dimethylsiloxane). (Reproduced with permission from Aime and Gauthier.[23]) (c) Force curves obtained on glass and ULDPE samples. Strong deformation of the ULDPE sample ($z_1 - z_2$) at a cantilever deflection Δ_1 is evident from these curves.

seen from the comparison of force curves in Figure 12(a). In general, the tip forces under liquid can be one to two orders of magnitude less than in air. This finding had encouraged AFM applications to biological systems because measurements in the aqueous environment, which is more suitable for such samples, can be performed. However, imaging under liquid has its own limitations especially when knowledge of the material structure and

behavior under ambient conditions is required and when interaction with liquid is not desirable.

It has been found that surface material or even a single macromolecule, when stuck to the tip, is stretched during the disengagement. In such cases, the jumps in the force curve can be associated with a break in the stretched material or bonds in single macromolecules. This situation is described in Figure 12(b) which shows the force curve, which was measured on polydimethylsiloxane surface grafted to silica substrate.[23] First, a pull-out force is recorded when the tip become unstuck from the surface. On further withdrawal of the probe, the cantilever bends again because the polymer material stuck to the tip is stretched and this stretching force exceeds the cantilever's stiffness. The next disconnection happens when the ''bridge'' between the tip and the surface is broken. The events described happen in the tens of piconewtons force range. These highly sensitive force measurements allow mechanical properties of single macromolecules, including DNA to be evaluated, opening up another interesting research area.[24]

Analysis of force curves in the repulsive regions shows how mechanical properties of polymer samples can be examined. When the sample is rigid, such as a piece of Si wafer or glass, the cantilever bends upwards in a linear fashion defined by the cantilever's stiffness. For many polymer samples, their stiffness is close to that of the cantilever and the force curve could reflect the tip-induced surface deformation or a partial tip penetration into the sample. The slope of the force curve and the difference in the way the scanner travels to reach the same cantilever deflection define the tip-induced deformation of a soft sample, Figure 12(c). A more detailed analysis of the force curves in needed to clarify the effects of the sample viscoelasticity.[25]

The important conclusion from force curve studies is that they can be applied to local probing of mechanical and adhesive surface properties. An extension of this approach is realized in another mode known as *force volume*. In this mode, force curves are collected at a large number of points (64×64, 128×128 or 256×256) in the area of interest, and the data at any particular force level is presented as a map. In this way, surface maps of mechanical properties (taken as force data at strong repulsive force levels) or adhesive properties (taken as force data in the disengagement region) can be obtained.[26] A similar idea was implemented in the *pulsed force mode*, in which maps of mechanical and adhesive properties are collected at a broad range of frequencies.[27]

Mapping of the mechanical surface properties had also been achieved in *force modulation* (FM) mode. In FM, the probe, which scans the sample in the contact mode, is driven into a vertical oscillation with an additional piezo-actuator at its resonant frequency (5–8 kHz). This actuator is positioned at the fixed cantilever end. With the tip staying in contact with the sample, the cantilever with its base being the whole assembly periodically moves up and down (Figure 13a) and the cantilever amplitude in the deflection direction is measured by the laser signal reflected from the back of the cantilever. In other words, the force on the sample is modulated around the set-point level, which is chosen to scan in the contact mode. FM data (fast changing, amplitude in deflection) are collected simultaneously with topographic information (slow changing, averaged deflection). The amplitude response depends on the mechanical properties of the sample: on stiff surfaces the tip bounces from the sample and the amplitude is large, whereas on soft samples the tip deforms the sample or penetrates into it and the deflection amplitude is small. This is demonstrated by the images of the multilayer PE sample, Figure 13(b) and (c). The height image shows the sample topography with elevated and narrow strips of high-density PE layers. In the amplitude image, the darker contrast (larger amplitudes are coded by a darker contrast) of these narrow layers confirms their high modulus in comparison with that of the wide layers of ULDPE. Initially FM was introduced to differentiate between the responses of surface regions or components (e.g. carbon fibers in epoxy matrix) with a different elastic modulus, however, its capabilities are much broader though not well explored.

To extend the application of FM to viscoelastic materials, it was suggested that the phase shift between two periodical signals be measured, one for the driver and the other for the photodetector. In this approach, some analogy can be found between this approach and dynamic mechanical analysis (DMA), in which the phase lag between an oscillating load and the induced sample deformation are used for evaluation of the viscoelastic properties of polymer samples. Indeed, there is some correlation between the driving amplitude in FM and sample deformation in DMA, and also between the cantilever deflection in FM and the load applied to the sample in DMA. However, in AFM of polymers the sharp tip deforms a sample or penetrates into it. A theoretical description of the tip deformation of the viscoelastic media is rather complicated, and nonlinear tip–sample force interactions substantially limit transferability of AFM data to the macroscopic mechanical behavior of polymers determined in DMA experiments. An additional challenge is an appropriate description of the cantilever mechanics in FM. Also it is worthwhile to note that the combination of the contact and oscillatory modes employed in this mode restricts its applications to soft materials.

2.3.3 *Tapping Mode*

The need for small tip–sample force interactions has been recognized since the introduction of AFM. An

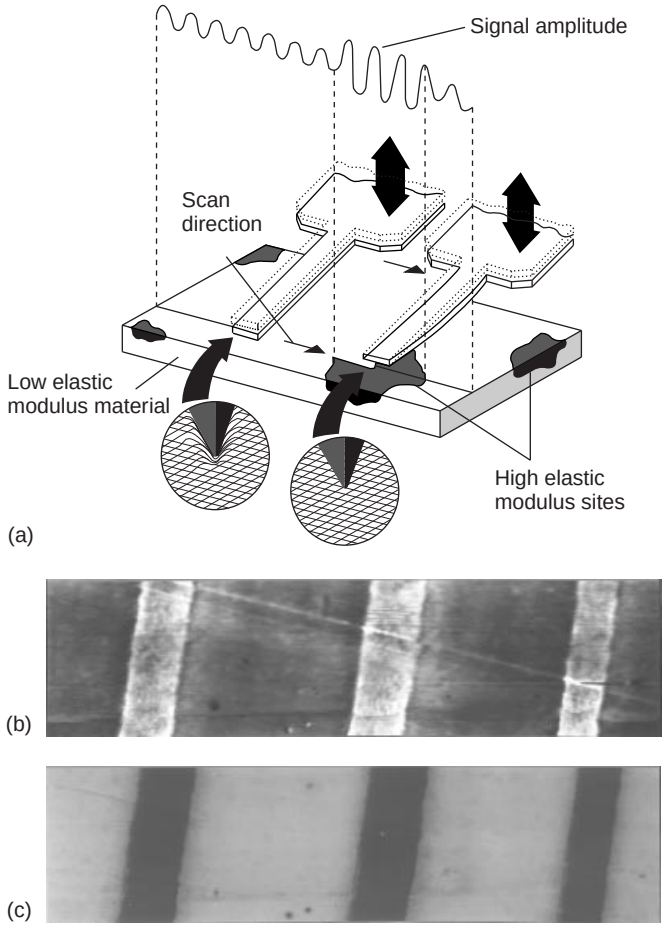

Figure 13 (a) Schematic view of FM mode. Height image (b) and amplitude image (c) obtained on multilayer PE in the FM mode. The images were obtained simultaneously on surface area 60 μm × 15 μm. The dark contrast in the amplitude image corresponds to larger deflection amplitudes.

initial suggestion was to make use of noncontact imaging in a pure attractive force mode.[28] Sensitive probes and oscillatory cantilever motion were needed for the detection of long range weak van der Waals interactions. Although several applications of noncontact AFM imaging have been developed (true atomic scale imaging in UHV, profiling of semiconductor structures with maximum resolution of few nm, etc.) progress in AFM of polymers is entirely determined by use of the tapping mode. In this mode, the tip intermittently contacts the sample for a short time thus avoiding development of the inelastic sample deformations.[9] In the tapping mode, lateral tip–sample forces are practically eliminated, allowing this mode to be applied to a broad range of soft materials. It appears that optimization of the tip–sample forces is the main issue for most AFM applications to polymers. In one respect, minimization of the tip force is required for a gentle profiling of polymer samples. Low-force imaging, which means a small tip–sample contact area, will also be beneficial for high-resolution imaging of samples such as polymers that are not well ordered. In another respect, tapping at elevated forces is extremely useful for the compositional imaging of heterogeneous polymer systems. Therefore, it is important to demonstrate how to optimize the tip–sample force interactions in the tapping mode.

In the tapping mode, reduction of the cantilever amplitude, which is determined at (or near) the resonant frequency of the cantilever, owing to tip–sample forces is used to obtain feedback during scanning. The amplitude of the free-oscillating cantilever, A_0, is typically in the 1–200-nm range. Whether to use a set-point amplitude A_{sp} or a set-point ratio, r_{sp} ($= A_{sp}/A_0$) is the choice of the operator. In contrast to the contact mode, in which the cantilever deflection is directly related to the tip force, the amplitude changes do not fully determine the tip–sample forces in this oscillatory mode. For imaging the surfaces of homogeneous materials, this does not cause problems, although they might occur in imaging of heterogeneous samples with surface corrugations of a few nanometers, see section 3.1.

The driving frequency in the tapping mode is typically chosen at the resonant frequency of the probe. For systematic AFM studies at different A_0 and A_{sp} it is important to find the resonant frequency of the probe in the immediate vicinity of the surface. (Far away from the surface, the resonant frequency of the cantilever will be higher because of reduced damping of the cantilever motion in a near-surface confined environment.[29]) A typical amplitude resonance curve of the AFM cantilever resembles that of a harmonic oscillator, Figure 14(a). The frequency dependence of the cantilever phase is also shown. The phase as well as the resonant frequency of the oscillator is a sensitive measure of the forces acting on the probe. Attractive forces acting on the AFM probe cause a negative shift of its resonant frequency; repulsive forces cause a positive shift. If the phase of the oscillation is measured at the initial resonance frequency, w_0 (as happens in tapping mode), the phase of the oscillation will be negative (with respect to its level at w_0) in the attractive force regime and positive in the repulsive force regime. The phase changes are instructive for clarification of the tip–sample force interactions in the tapping mode. The amplitude versus distance and phase versus distance curves (Figure 14b) show how amplitude and phase of the oscillating cantilever change when the separation between the sample and the cantilever's base is reduced. When an amplitude drop is small, a negative phase indicates that the overall tip–sample interactions are attractive. With a further amplitude drop, repulsive interactions become dominant as seen by the switch in phase from negative to positive. These interactions also depend on the magnitude

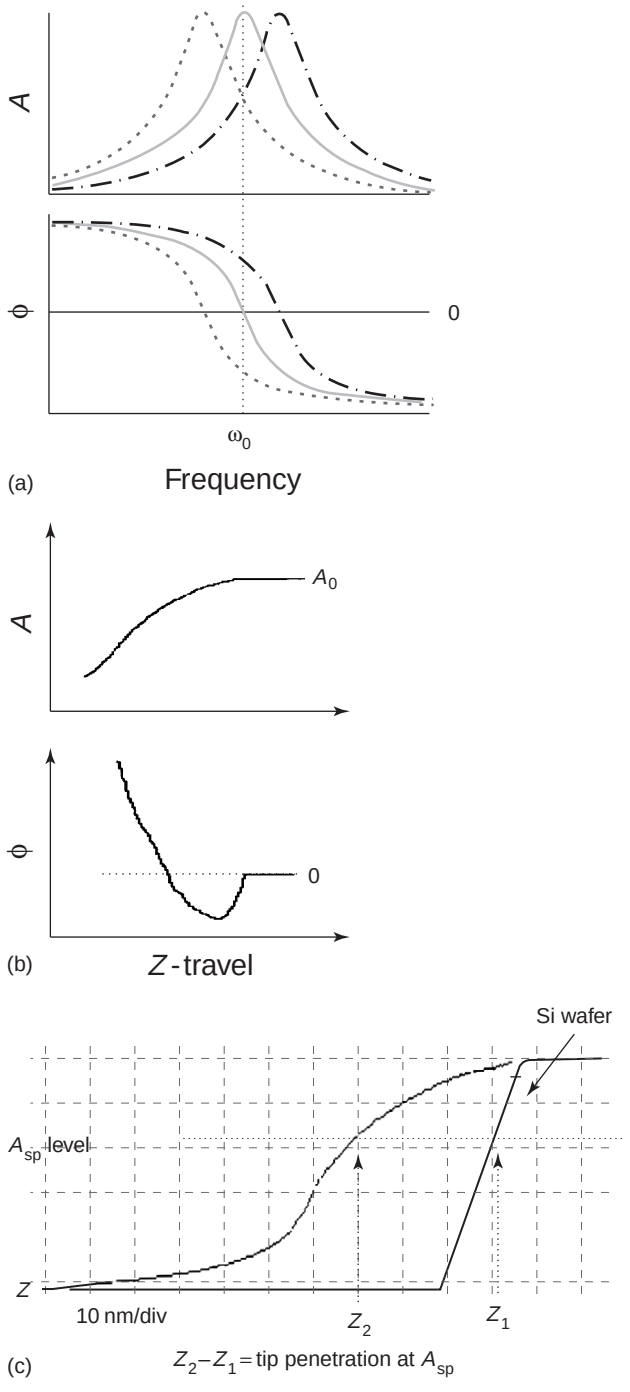

Figure 14 (a) Amplitude versus frequency and phase versus frequency relationships for harmonic oscillator with resonant frequency w_0. Attractive forces cause a low-frequency shift of these curves (- - - -), repulsive forces cause a high-frequency shift (– · – · –). (b) Amplitude versus travel and phase versus travel curves recorded in tapping mode on high-density PE. (c) Amplitude versus travel curves obtained on ULDPE and a Si wafer.

of the initial amplitude A_0. When A_0 is small (1–5 nm), phase changes are negative at all amplitudes. Under these conditions, tip–sample interactions are always attractive. At large amplitudes (above 60 nm), the phase changes are always positive owing to the dominant repulsive force interactions. These results are typically obtained on surfaces of PE samples.

Amplitude curves recorded on polymer samples deviate from the straight line which describes the amplitude drop when the sample is hard such as Si wafer. In the latter case, the amplitude drops linearly with the sample travel and vanishes totally when the sample surface coincides with the cantilever baseline, see also Figure 6(b). (On further vertical travel the cantilever bends up and this leads to the appearance of a deflection signal similar to that in the contact mode.) On the soft samples, owing to this tip-induced deformation or even penetration, the amplitude curve is very different from that on the hard sample. As an example, the amplitude curve which was obtained on ULDPE is shown in Figure 14(c). At each A_{sp} value, the difference between this curve and a straight line, which simulates the curve on the rigid sample, defines an amount of tip-induced sample deformation or of tip penetration. Several examples of imaging at different tip penetrations will be given in sections 5.1 and 5.2.

Although the force interactions in the tapping mode are described by the amplitude and phase curves, it is important to understand how A_0 and A_{sp} influence the AFM images of polymer samples. This question was discussed earlier, and it was found that with an increase of A_0 and a decrease of A_{sp}, tip forces become stronger. An evaluation of the tip–sample forces in the tapping mode is not straightforward because the force changes in each cycle. Therefore, it will be instructive to judge these interactions by considering force-induced sample changes. The model system, a binary blend of HDPE and ULDPE prepared in a melt, can be used as a convenient model sample for examining tip–force effects. The elastic moduli of the components are quite different, therefore it becomes possible to observe tip–force-induced height changes. Height and phase images of the different sample imaging conditions are collected in Figure 15(a). The topography of the sample shows a number of bumps, which appear to be high-density PE regions. With the increase in A_0 and decrease in A_{sp}, these structures become more pronounced related to the strong deformation in their surroundings, which is ULDPE. The contrast in the phase images reveals different components of this blend, however, an assignment of ''dark'' or ''bright'' regions to individual components is rather ambiguous.

To understand better the relationship between the phase image contrast and surface regions with different densities it is instructive to analyze cross-section height and phase profiles taken along the same surface regions, Figure 15(b). These data were obtained at $A_0 = 100$ nm

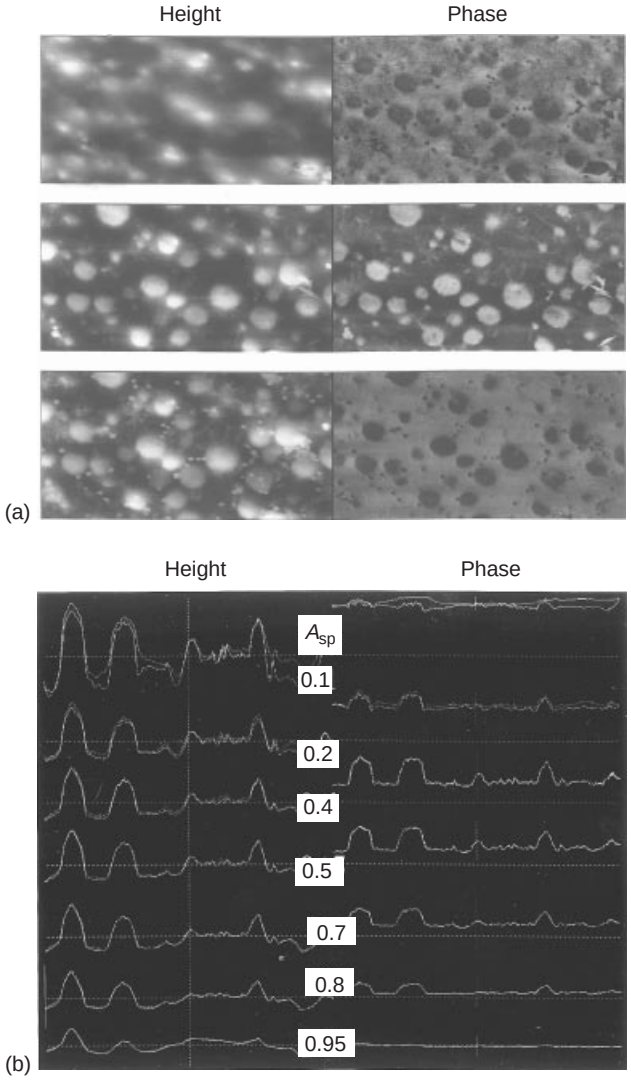

Figure 15 (a) Height and phase images of the HDPE/ULDPE blend, prepared in a melt. The images were obtained on surface area 3 μm × 3 μm The images at the top were obtained at $A_0 = 20$ nm and $r_{sp} = 0.7$; the images in the middle were obtained at $A_0 = 100$ nm, $r_{sp} = 0.5$, and the images at the bottom were obtained at $A_0 = 80$ nm and $r_{sp} = 0.15$. (b) Set of height and phase cross-sections along the 3-μm long regions of the HDPE/ULDPE blend. Measurements were performed at $A_0 = 100$ nm and different r_{sp}.

and r_{sp} in the 0.95–0.1 range. When r_{sp} is close to 1, the profile in the height image presents a true surface topography. Note that the phase curve is located almost at the zero level indicating an absence of noticeable tip–sample force interactions. This imaging regime is known as *light tapping*.[30] With the decrease in r_{sp}, the ''bumps'' become higher owing to stronger tip deformation of the soft regions. This deformation can be described as elastic because the height and phase changes at r_{sp} in the 1–0.1 range are reversible. The phase changes also become evident with high-density spots exhibiting more positive phase changes. The maximum phase difference of the stiff and soft regions is observed at $r_{sp} = 0.4$–0.5, and it gradually disappears at small r_{sp}. Under the same conditions, the effective height of the ''bumps'' continues to increase.

The height changes in the low-density and high-density regions versus r_{sp}, which were recorded at different A_0 and A_{sp}, are summarized in Figure 16(a). The height versus r_{sp} curves show that at small A_0 the decrease of r_{sp} from 1 to 0.4 does not lead to any substantial height changes, and only at lower r_{sp} does the tip-induced surface deformation cause an increase of the height of the high-density regions. With an increase of A_0, the height versus r_{sp} dependence becomes different. At high r_{sp}, the height has already increased and it exhibits a local maximum at $r_{sp} = 0.5$. With a further decrease in r_{sp}, the height slightly decreases, and at small r_{sp} it increases again. The similarity in the height changes for all amplitudes at very small r_{sp} (0.2–0.1) is most likely to be due to a loss of the ''tapping'' character of the probe. Under these conditions the tip spends most of the time in contact with the sample (quasicontact) and the tip-induced surface deformation is determined mostly by the cantilever spring constant and it does not depend on A_0.

The phase changes which were measured on the regions of different density at $A_0 = 100$ nm and $A_0 = 20$ nm are shown in Figure 16(b) and (c). In imaging with large A_0, the positive difference between the phases in the high-density and low-density regions is observed for almost all r_{sp}. The maximal phase contrast between components is at $r_{sp} = 0.4$–0.5, i.e. under the same conditions as when the difference in the mechanical properties is visualized in the height images. This correlation allows the suggestion to be made that with *hard tapping*, the phase contrast is governed mainly by the different mechanical properties of the sample regions or components. More complicated phase behavior with contrast reversal is seen for imaging at smaller amplitudes when a more complex interplay between capillary effects, adhesive interactions and mechanical properties defines the phase changes. Therefore, imaging with larger A_0 and $r_{sp} \sim 0.4$ (''*hard tapping*'') is of more practical importance because it facilitates the assignment of phase contrast to the components with different mechanical properties. In the case of different PE, acrylic blends and elastomeric systems, stiffer components can be identified by their more positive (brighter) phase contrast in the phase images obtained by *hard tapping*. The great sensitivity of the phase changes to the sample heterogeneity is a key feature of the tapping mode in its application to polymers. In many heterogeneous polymer samples the differences in the elastic moduli of the components are not so drastic as in the case of the PE blends discussed above.

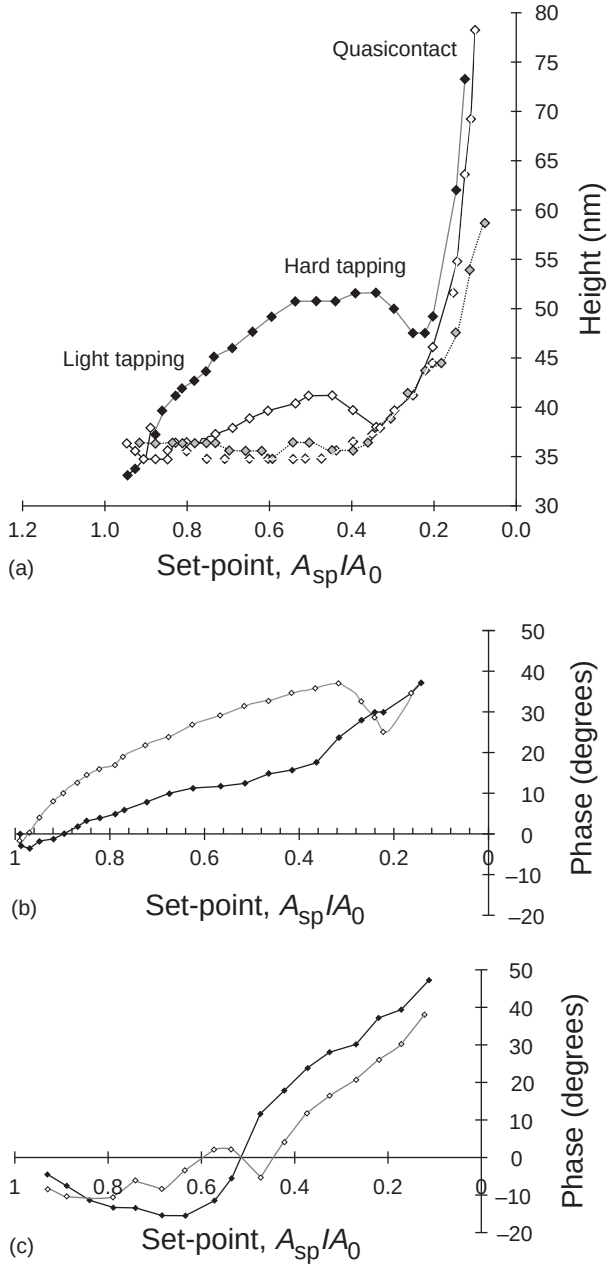

Figure 16 (a) ''Height'' difference between regions of HDPE and ULDPE obtained at different A_0 and different r_{sp}. Unlinked empty diamonds correspond to $A_0 = 5$ nm; cross-hatched diamonds to $A_0 = 20$ nm; linked empty diamonds to $A_0 = 40$ nm; black diamonds to $A_0 = 80$ nm. (b) and (c) Phase differences between the regions of HDPE (open points) and ULDPE (filled points) obtained at different $r_{sp} = A_{sp}/A_0$. $A_0 = 20$ nm (c) and $A_0 = 100$ nm (b).

Therefore, even with *hard tapping*, the height variations in these materials are hardly distinguishable, whereas the phase contrast provides component recognition.

Phase imaging has become a useful addition to the tapping mode because the phase contrast allows visualization of different components of heterogeneous polymers. However, assignment of phase contrast is not straightforward and a number of theoretical approaches are applied to address this problem. From the experimental data, phase contrast can be assigned to variations in mechanical or adhesive properties of components or surface regions. In a more general case, phase changes in the tapping mode strongly depend on energy dissipation.[31] Correlation between the energy dissipation at the tip–sample junction and the sample position is shown in Figure 17(a–c) together with the amplitude and the phase curves. All these dependencies were obtained on the surface of Si wafer.

In the absence of dissipation, when the amplitude of the oscillating cantilever is not restricted by the set-point value, the relation between the phase and amplitude of the cantilever is defined by symmetric branches $\phi = \pm \arcsin(A/A_0)$. These branches are relevant to the operation either in the attractive force regime (negative phase) or in the repulsive force regime (positive phase). In such cases (for example in the lift scan in MFM or EFM), phase and amplitude are strictly bound. Because the amplitude is kept constant (feedback is off) in the tapping mode, this relationship breaks down and phase changes are caused by the energy dissipation in the tip–sample junction, Figure 17(a–c). The experimental phase curves lie within the ''arcsin'' branches and the transition from the attractive to repulsive force regions means a ''jump'' from the lower to the upper branch, Figure 17(d). Although this transition is not specific in terms of energy dissipation, it can complicate the interpretation of phase contrast, which is defined with respect to the 90-degree level. (For simplicity this level so far has been defined as 0°). Figure 17(d) also demonstrates the correspondence between experimental data and the calculations, which were performed considering the interacting cantilever as a weak nonlinear system.[32] A number of phase curves derived from the calculations, which accommodate elastic deformation, adhesion and dissipation terms (K, w and α_s, respectively), all fall within the ''arcsin'' branches. It was also indicated that at low r_{sp} the tapping mode probe becomes a strong nonlinear vibration system. Therefore despite the fact that tapping mode results obtained on some elastomeric systems can be treated in terms of a simple harmonic oscillator[33] this approach is not general. To summarize, progress in the theoretical description of the tapping mode has brought a better understanding of the probe behavior and emphasized the energy dissipation in the tip–sample junction as a primary event. However, separation of the contributions of the different interactions to the phase behavior remains a challenging problem in AFM analysis of multicomponent polymer samples.

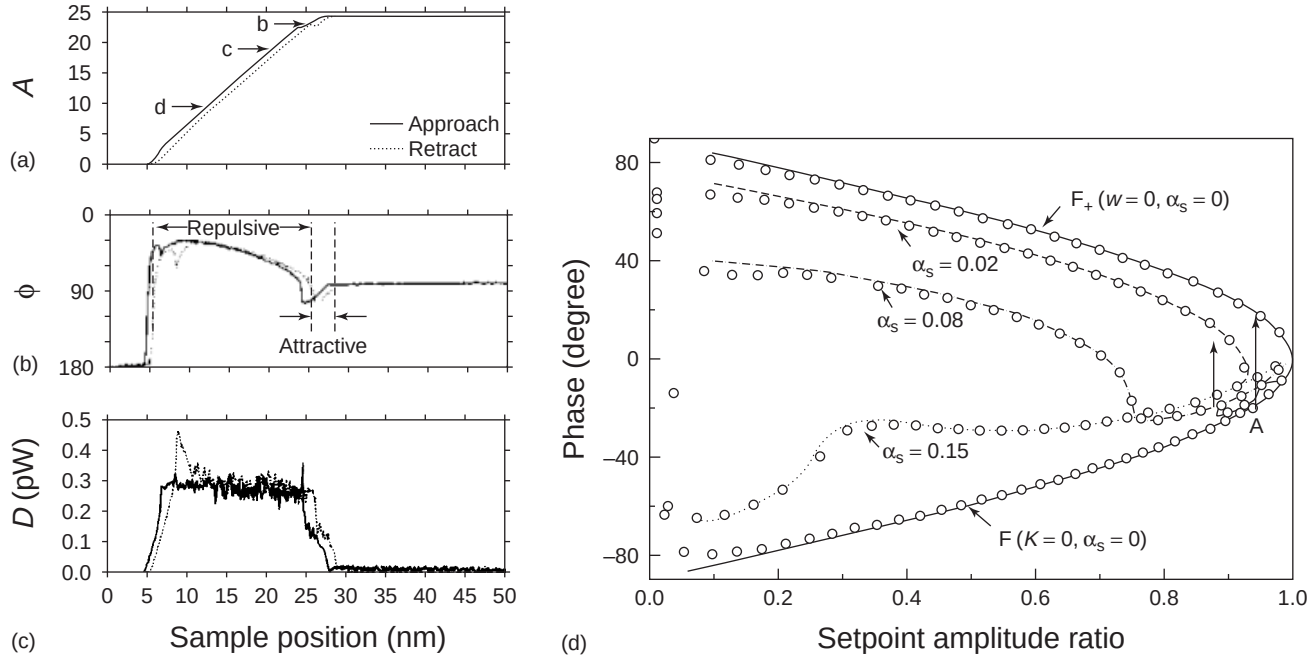

Figure 17 (a) Amplitude, (b) phase and (c) dissipation of the cantilever as a function of sample travel. (Reproduced from Cleveland et al.[31]) (d) Phase versus A_{sp} under different interaction conditions. (Reproduced from Wang.[32]) F_+ and F_- correspond to repulsive dominated and attractive dominated branches, respectively. Mechanical and adhesive interactions ($K = 2$ GPa and $w = 0.4\,\mathrm{N\,m^{-1}}$) are the same for all curves except F_+ and F_-.

$$P_{tip} = \frac{kA^2 w_0}{2Q}\ \frac{A_0}{A}\sin(\varphi) - 1$$

where P_{tip} is power dissipated at the tip–sample junction. A is the set-point amplitude, A_0 is the amplitude of a free-oscillating cantilever and Q is the quality factor.

3 EXAMINATION OF SURFACE TOPOGRAPHY AND LOCAL PROPERTIES

A number of practical examples, which illustrate the unique capabilities of AFM in studies of polymers, are presented in sections 3–5. Many examples show how AFM complements other characterization techniques. Most results were obtained in the tapping mode, recording height and phase images. Height images reveal surface topography and characteristic morphologic features of crystalline polymers, of oriented samples and of microphase separation. In studies of homogeneous samples, phase contrast often helps to visualize fine features of polymer morphology and nanostructure. Phase imaging is invaluable for compositional mapping of heterogeneous polymer systems. In tapping mode imaging, the driving frequency is chosen at the resonant frequency of the cantilever in the immediate vicinity of the sample surface. This does not exclude the possibility that other frequencies can also be used for imaging. For recording height images, which most correctly present true surface topography, small A_0 and high A_{sp} are applied. Such conditions are most favorable for high-resolution imaging owing to the minimal tip–sample contact area. Imaging at an A_{sp} of 0.5–0.4 A_0 and large A_0 (*hard tapping*) was used for compositional mapping of heterogeneous polymer samples. These conditions should be considered only as a preliminary idea of those needed for imaging of polymer samples. In practice, one should use a whole spectrum of parameters to optimize AFM imaging towards the problem being studied.

3.1 Surface Topography and Roughness

Initially, high-resolution surface profilometry was a primary goal of AFM, and this feature is useful in imaging research and industrial polymer samples. Surface corrugations from several nanometers to several microns can be examined quantitatively. Two examples of surface roughness measurements with AFM are presented in Figure 18(a) and (b). Surface corrugations in the height image of poly(methyl methacrylate) (PMMA) photoresist, which was deposited on a Si wafer by spin coating, (Figure 18a) are described by the following roughness parameters: $R_{ms}(R_q) = 0.394$ nm and $R_a = 0.314$ nm. After chemical mechanical polishing bare

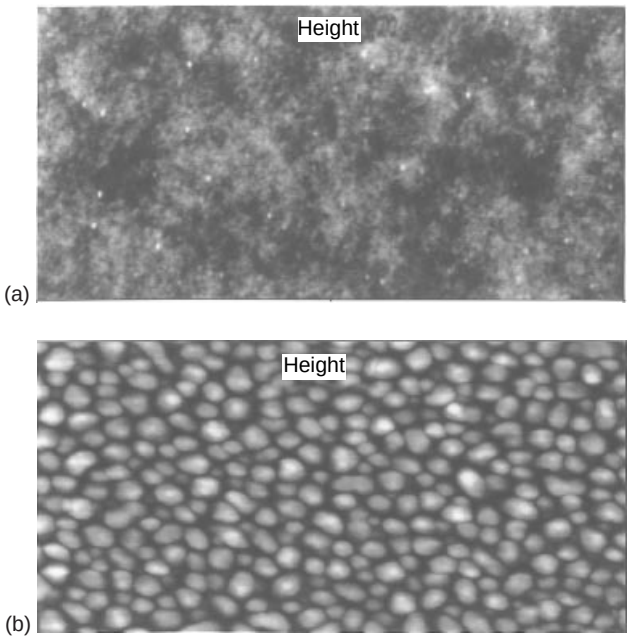

Figure 18 AFM height images of a photoresist (PMMA) layer on Si wafer (a) and polycrystalline Si sample (b). Image size is $2\,\mu m \times 2\,\mu m$ in (a) and $3\,\mu m \times 3\,\mu m$ in (b).

Si wafers are characterized by smaller roughness with R_{ms} (R_q) and R_a below 0.2 nm. AFM evaluation of the surface roughness of Si wafers is employed for quality control in the semiconductor industry.[11] Considerable roughness is desirable for polycrystalline Si surfaces, which are used in the manufacture of stack capacitors, and the AFM image of the polycrystalline surface in Figure 18(b) reveals its surface roughness with $R_{ms}(R_q) = 8.113\,nm$; $R_a = 6.553\,nm$.[34] Imaging with *light tapping* facilitates accurate roughness measurements because it provides not only a small contact area but also prevents wear of the probe apex.

The technological surfaces described above are relatively flat, and they do not require special probes, in contrast to those applied to more complicated measurements of surface structures with critical dimensions, such as steps and trenches. In order to measure angles of steep surface steps or roughness at the bottom of narrow trenches (a few microns in depth and hundreds of nanometers in width) AFM probes with special geometry are needed.

The surfaces of most polymer samples are relatively flat and can be examined with conventional Si probes. Measurements of surface roughness with AFM are useful for studies of paint and polymer coatings. The height image of acrylic paint (Figure 19a) shows a surface with averaged corrugations $R_{ms}(R_q) = 36.771\,nm$ and $R_a = 28.057\,nm$. Graphs, which show a linear correlation of AFM data with gloss measurements at two different angles of incidence, are presented in Figure 19(b).

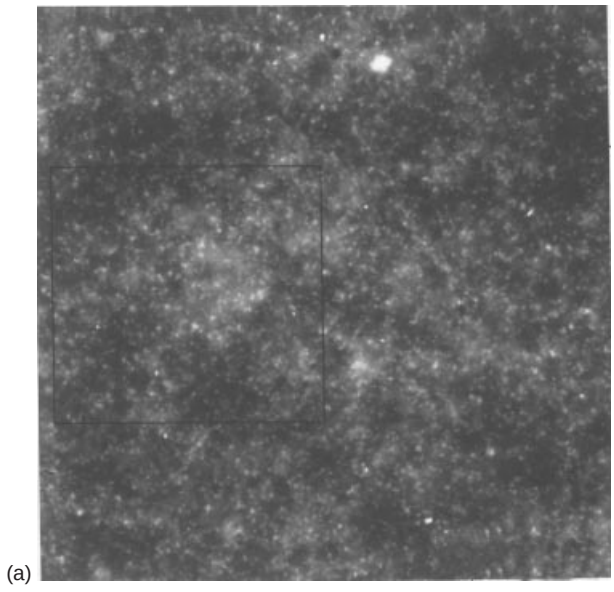

Image statistics	
Z range	539.18 nM
RMS (Rq)	36.771 nM
Ra	28.057 nM
Srf. area	10068 μM²
Srf. area diff	0.689%

Box statistics	
RMS (Rq)	35.947 nM
Mean roughness (Ra)	28.038 nM
Surface area diff	0.687%

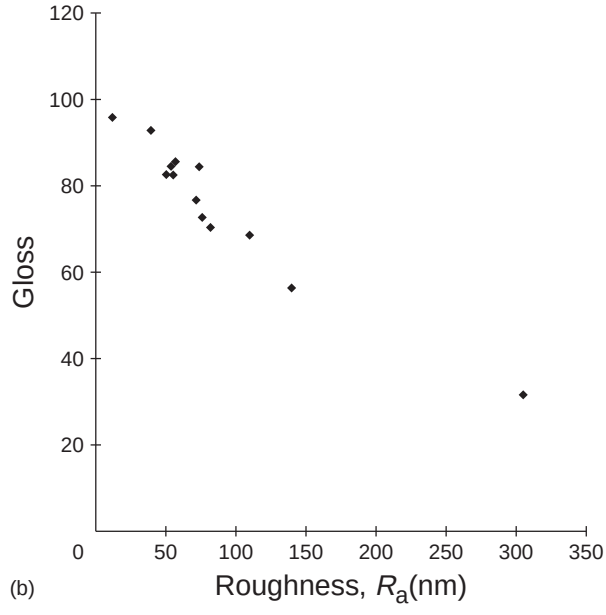

Figure 19 (a) Height image of one of 18 PVC paints. Contrast is in the 0–500-nm range. Image size is $100\,\mu m \times 100\,\mu m$. Statistics for the image and the box, which confines a part of the surface area (black lines) are shown beside the image. (b) The relationship between surface roughness (R_a) and gloss, measured at 60 (filled dots) and 20 (open dots) degrees, for 13 different paints.

Quantitative estimates of the surface roughness of various coatings provided by AFM are useful for correlation with their optical properties and appearance. AFM evaluation of surface roughness is important for samples of thin and transparent coatings on substrates which are difficult to examine by optical means.

More extensive AFM studies were performed on model systems of the paints such as acrylic latexes.[35,36] AFM images reveal the morphology of latex surfaces and their packing motifs (Figure 20a). At higher magnification, the submicron structure of individual latex particles can be examined. Height and phase images of the polystyrene (PS) latex layer show only small surface corrugations on individual particles, which are noticeable only in the phase image, Figure 20(b). A ''berry-like'' particle structure of PMMA latex is seen in the AFM images in Figure 20(c). Again, these structural features are more pronounced in the phase image.

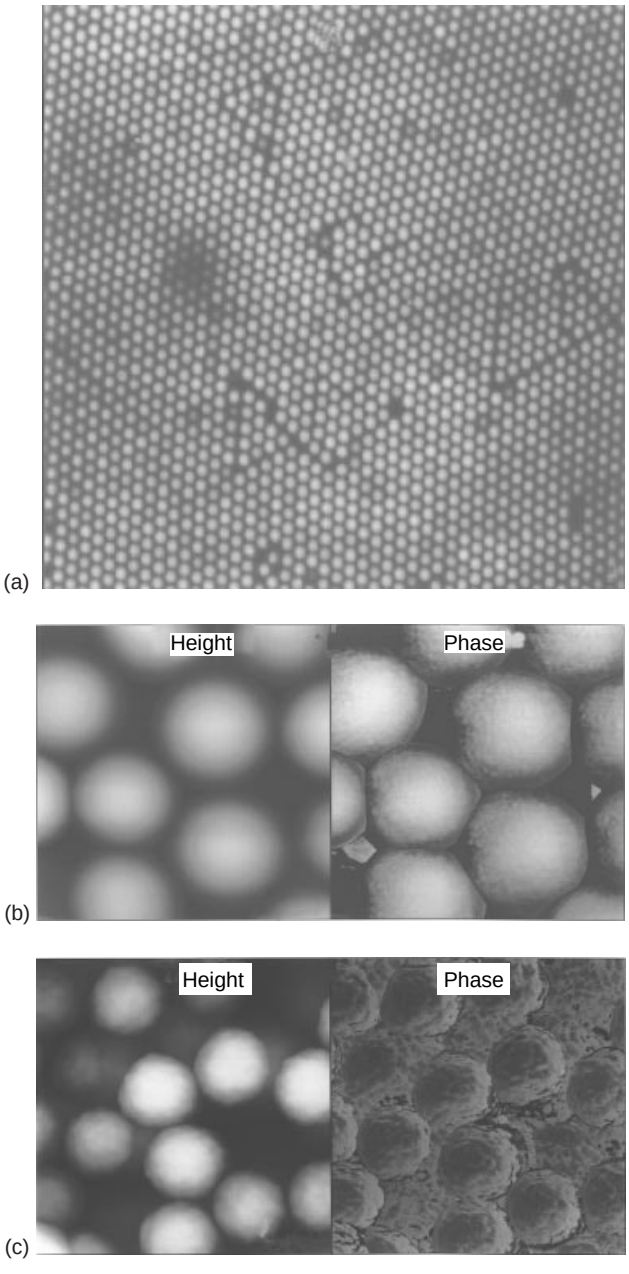

Figure 20 (a) Height image of PS latex layer showing close-packed layer morphology. Scan size 20 μm × 20 μm. (b) High-resolution images of PS latex. Scan size 1 μm × 1 μm. (c) High-resolution images of PMMA latex. Scan size 1.5 μm × 1.5 μm.

It is worth noting that because of the perfect shape of latex particles, which can be prepared in different sizes, they are suitable for testing the quality of AFM probes. The surface area difference (a difference between surface area of corrugated features and surface area of the flat plane), which is calculated from height images of a single latex layer, would be greater if the tip correctly tracks surface locations between neighboring particles. In this way probes that are most appropriate for reliable measurement of surface corrugations can be chosen. The latter are often employed for monitoring film formation in polymer latex systems.

Many polymer surfaces show morphological features with various degrees of order and regularity. The surface topography of liquid crystalline polymers reflects their regular architecture, which correlates with their optical behavior. An optical micrograph and AFM images of a cholesteric liquid crystalline copolymer, which both reveal a focal-conical texture, are presented in Figure 21(a) and (b).[37] Thermal annealing of this copolymer at 153 °C induces its transition from the focal-conical to the fingerprint texture, Figure 21(c). Selective optical reflectivity of the cholesteric copolymer is directly related to its texture and its spectrum correlates with the width of rings of the fingerprint texture seen in the AFM images.

Another example of a polymer sample with well-defined morphology is an industrial film of microporous isotactic polypropylene (iPP) (Celgard™ 2400). The morphology of this film, which was examined with SEM and AFM (Figure 22a–c), is characterized by alternating fibrillar and lamellar regions.[38] The lamellar regions are slightly elevated compared to the fibrils. The latter are not closely packed and, therefore, precise profiling of these fibrils with an AFM tip might be difficult owing to the tip convolution. These structural features (fibrils and holes between neighboring fibrils) are more correctly reproduced in the SEM micrograph (Figure 22a) than in the AFM height image (Figure 22b). However, visualization of the polymer nanostructure is superior with AFM, which is evident from the phase image in Figure 22(b). Therefore AFM is useful for detection of minor structural changes in polymers caused by various technological treatments. Comparison of the images of the Celgard™ 2400 membrane prior to and

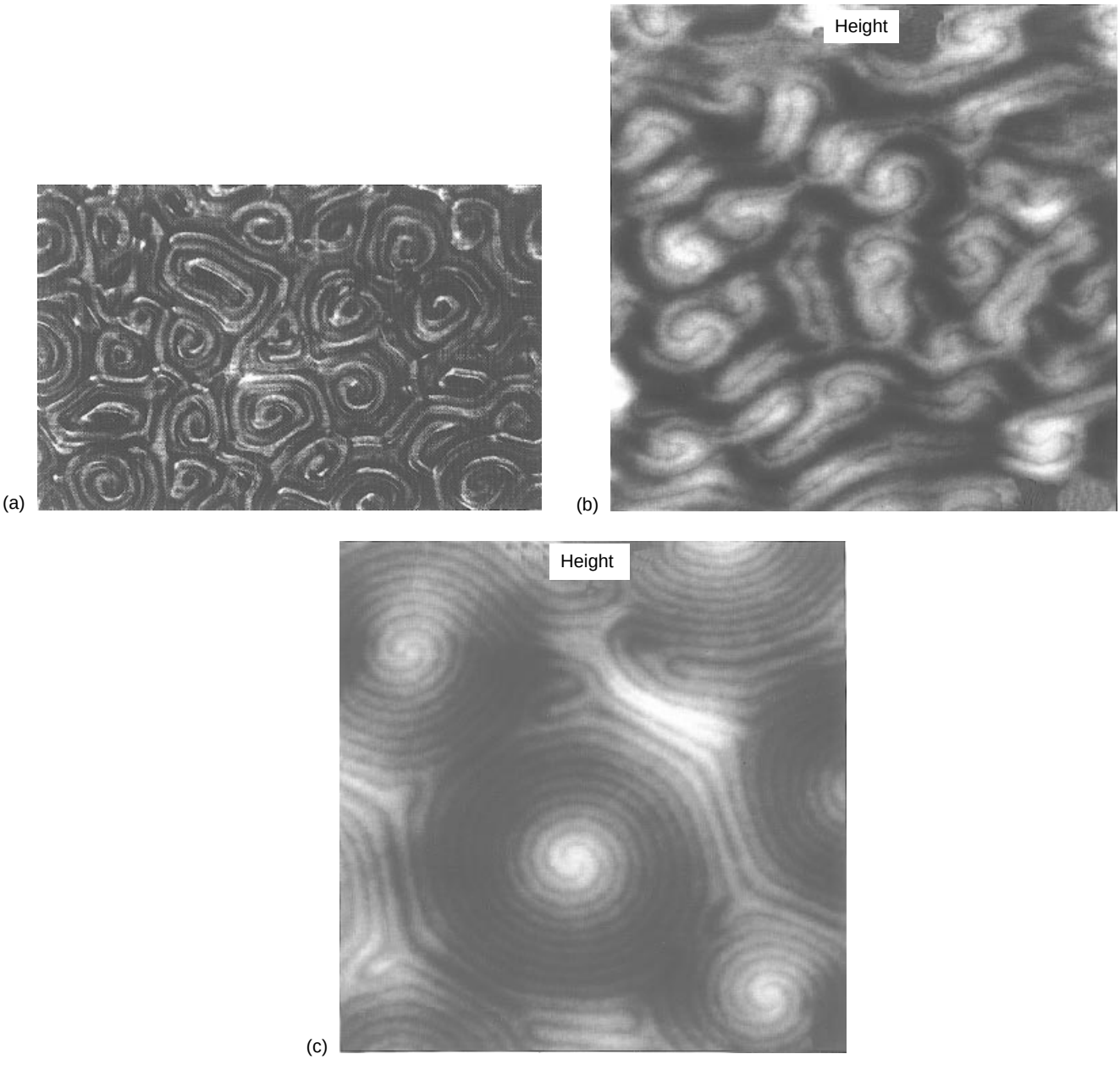

Figure 21 (a) Optical micrograph of focal-conical texture of cholesteric liquid crystalline polymer. (b) and (c) AFM images revealing focal-conical and fingerprint textures of cholesteric liquid crystalline copolymer, respectively. Scan size 10 μm × 10 μm.

after argon plasma treatment (Figure 23a and b) shows drastic changes in its nanostructure, whereas overall morphological features remain the same. These changes are caused by condensation of small water droplets from air. It happens because the plasma treatment makes the iPP surface hydrophilic.

Visualization of the morphology and nanostructure of ultrathin polymer films is a capability offered only by AFM. The height image of poly(ε-caprolactone) (ε-PCL) grafted to a Si surface (Figure 24) demonstrates this capability. The image shows the crystalline morphology of a 10-nm thick ε-PCL film, which was prepared by spin casting.[39] Numerous fibrils are spread over the surface along with flat lying lamellar sheets. These parts of growing two-dimensional (2D)-spherulites are the main structural features of the crystalline layer. A deficit of material prevented complete coverage of the substrate surface, and the cross-section profile taken along the *x*-axis, which is shown above the image, can be used for thickness measurements. The same measurements of thin polymer or organic layers, which cover the substrate completely, can be facilitated by scratching away part of

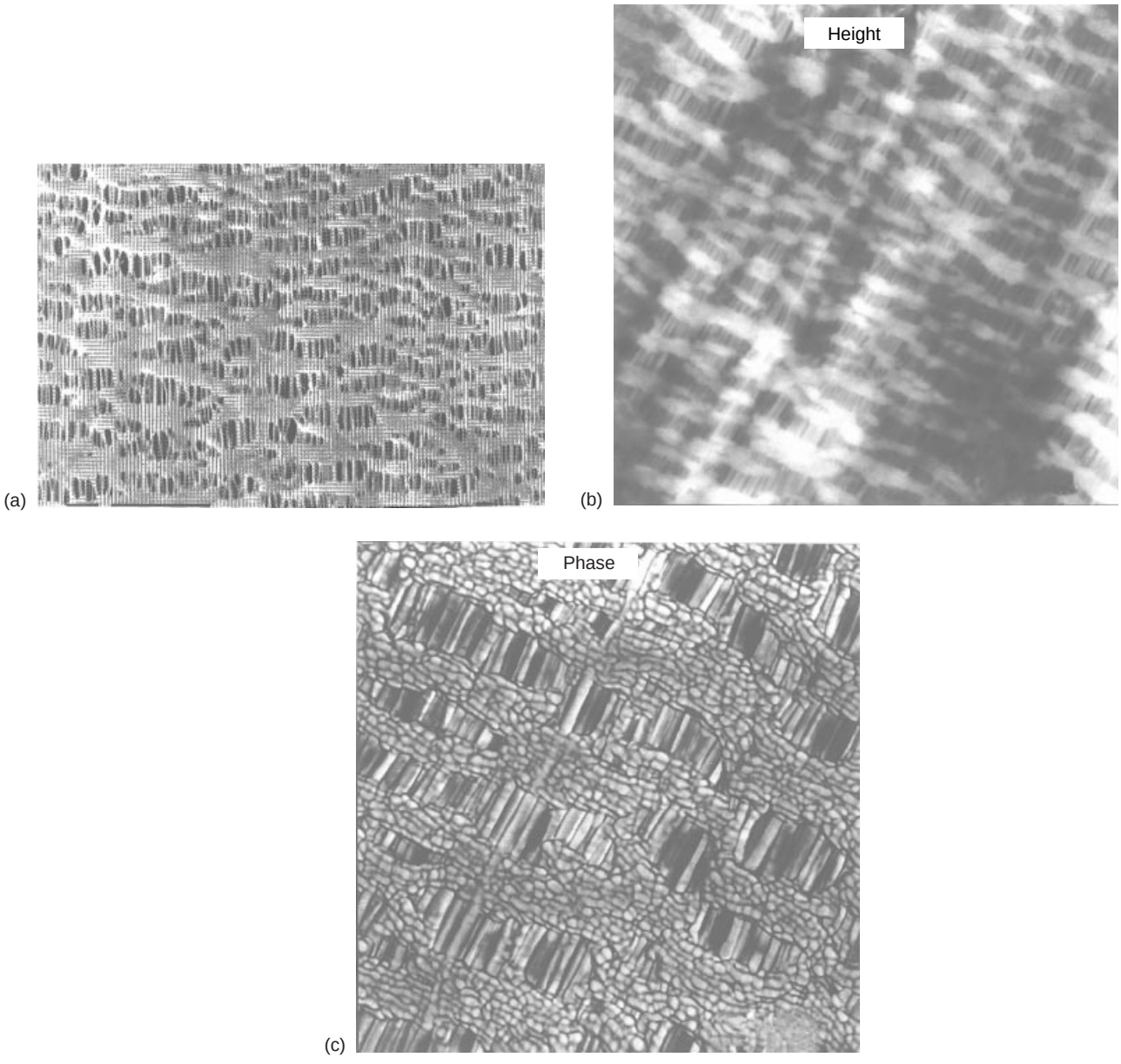

Figure 22 (a) Scanning electron microscopy micrograph of microporous iPP film (Celgard™ 2400). (b) and (c) AFM images of the same film. Scan size 5 μm × 5 μm in (b) and 2 μm × 2 μm in (c).

the layer with an AFM tip in high-force imaging (contact mode will be better for this purpose).[40]

Thin films and layers of block copolymers are routinely examined by AFM. The height and phase images of a thin layer of PS-*b*-PMMA, which was transferred on to a mica substrate from a water subphase in the Langmuir–Blodgett trough, are shown in Figure 25(a–c).[41] Hydrophilic PMMA blocks are spread as a first layer, whereas hydrophobic PS blocks aggregate on top. Imaging in *light tapping* allows evaluation of the thickness of the PS blocks (25–30 nm) and visualization of fine details (5–8 nm) of these blocks, Figure 25(c). The phase separation in the vertical dimension, which exists in this block copolymer film, defines a different response of the PS and PMMA blocks to the AFM probe with *hard tapping*. With the increase of the tip force, PS blocks are deformed more substantially than PMMA blocks. This is seen from the reduced brightness of PS blocks in the height image and dark appearance of these blocks in the phase image (Figure 25b). In other words, PS blocks, which are further from the substrate than PMMA blocks, are compressed by the probe to a larger extent than PMMA blocks, which lie directly on the substrate. When the tip force was reduced again,

the images became the same as in Figure 25(a), which is an indication of the reversible character of the sample deformation.

3.2 Surface Topography in Heterogeneous Polymer Samples

In studies of the topography of thin films which do not cover the substrate completely, and surfaces of other heterogeneous polymer samples, one should be aware of the possible drawbacks of AFM imaging when the topographic differences are a few nanometers in height. These polymer samples, which in addition are soft, are nontrivial for AFM topographic studies. Application of the contact mode for this purpose could induce sample deformation, which will be different for different sample components. Therefore, the height image will not reflect the surface topography correctly. There is also a problem in tapping mode applications. This problem is caused by the shift of the resonant frequency of the cantilever that leads to a drop in the cantilever amplitude measured at the resonant frequency of the free-oscillating cantilever. Therefore, when surface regions exhibit different properties (e.g. adhesive, mechanical, etc.), which cause different frequency shifts, then corresponding surface areas could appear in the height images with a fake topographic contrast, which is in range of the few nanometers. This effect, which is intrinsic to the tapping mode with amplitude feedback, should be considered while studying surface corrugations of a few nanometers in scale.

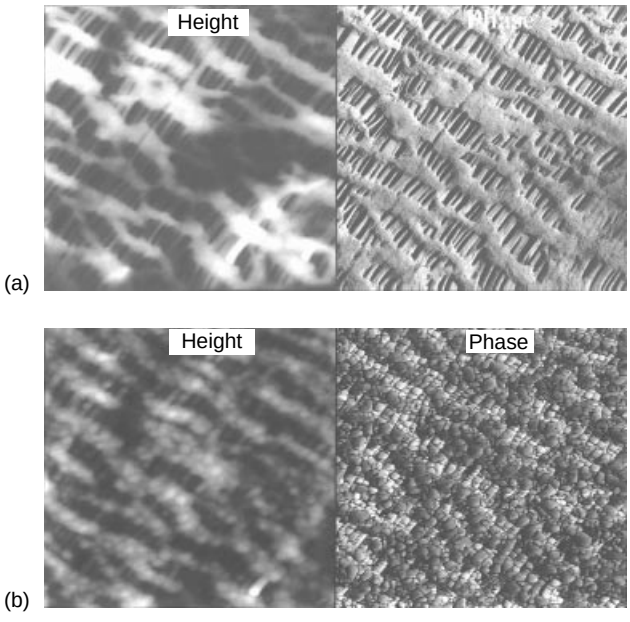

Figure 23 (a) AFM images of microporous polypropylene film Celgard™ 2400. (b) AFM images of the same area as in (a) after Ar+ plasma treatment. Scan size 3 μm × 3 μm in all images.

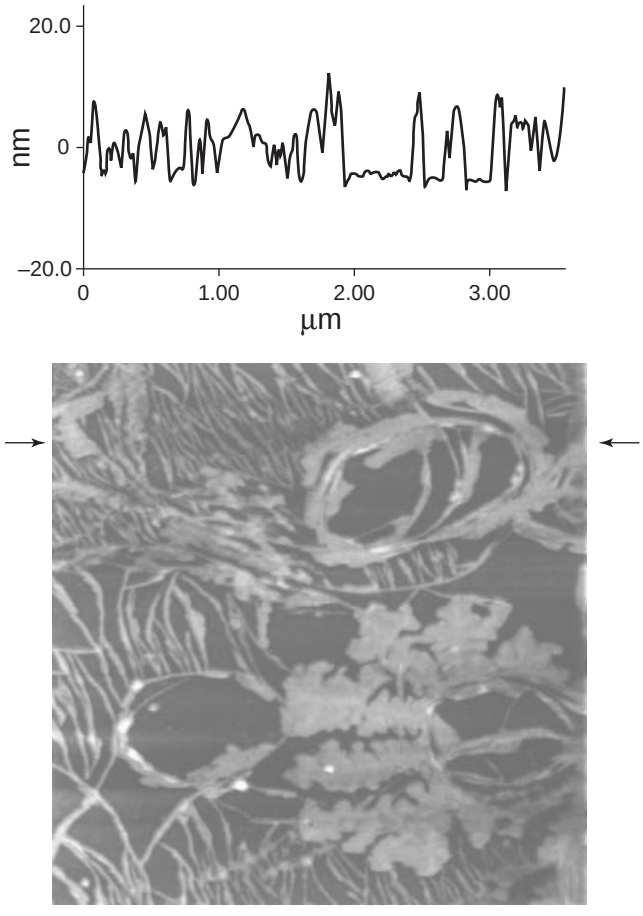

Figure 24 Height image of a 10-nm thick film of ε-PCL grafted to a Si surface. Scan size 4 μm × 4 μm. The cross-section profile shown above the image was taken along the (→ ←) direction.

Height images of an ultrathin film of PS-*b*-PMMA block copolymer, which was prepared by spin-coating on Si substrate, will be discussed to illustrate these effects. The way to minimize these effects will also be shown. Typically, the height images of this film show small height variations of PS and PMMA blocks, which reveal microphase separation in such polymers.[42] The image of the PS-*b*-PMMA film in Figure 26(a) reveals the microphase separation by showing a pattern similar to that usually determined by TEM. Although in TEM micrographs the different contrast in the PS and PMMA blocks is explained by thinning of the PMMA regions caused by electron beam-induced degradation, the reason for the contrast in the AFM images is not well understood. The fact that in Figure 26(a) and (b) a contrast is seen both in the height and phase images indicates that the probe interacts differently with PMMA and PS blocks. This causes variation of the ''heights'' of the corresponding regions, with some being higher than others in the height image recorded with *light tapping*, and lower in the

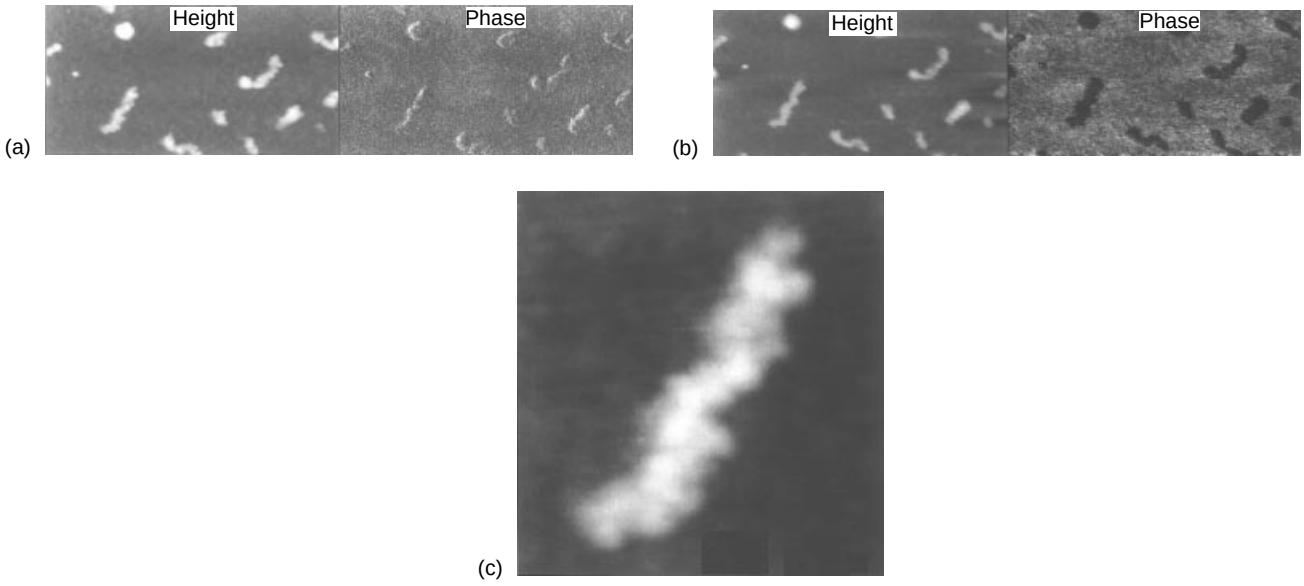

Figure 25 AFM images of a single layer of block copolymer PS-*b*-PMMA recorded in *light tapping* (a) and in *hard tapping* (b). Scan size 1 μm × 0.5 μm. (c) PS blocks shown at high magnification. Scan size 200 nm × 200 nm.

height image recorded with *hard tapping*. True sample topography for this sample was recorded by imaging with a softer Si probe, Figure 26(c). The height image recorded with *light tapping* (Figure 26c, top) reveals no differences in the height of PS and PMMA blocks. The fact that the corresponding phase image is almost featureless indicates that the tip interactions with PS and PMMA blocks are negligible. With *hard tapping*, these interactions become different causing a microphase contrast in the height and phase images (Figure 26c, bottom). This example shows that imaging in different operation regimes using probes of different stiffness is a way towards rational interpretation of AFM images.

3.3 Nanoindentation and Related Techniques

In addition to surface topography studies, AFM can be employed to evaluate mechanical properties of sample surfaces and polymer coatings. It has already been demonstrated that force curves could be employed for this purpose. As an extension of this approach, nanoindentation, scratching or wear experiments with standard Si probes or with special diamond probes can be used.[43] There are definite advantages to these modes in comparison with similar macroscopic procedures. They are related to new mechanical measurement capabilities on the micrometer and submicrometer scales. A broad range of low forces, which can be used for nanoindentation and scratching, allows examination of thin and ultrathin coatings without the need to account for substrate influences. Another advantage of AFM use for such studies is that deformation and fracture marks are made and are evaluated with the same instrument. Although nanoindentation and scratching studies with AFM are performed in a number of industrial applications, the question of the transferability of these data to macroscopic mechanical testing has been raised and needs further investigation. Two examples, which are chosen to demonstrate nanoindentation on polymer samples, are shown below.

A matrix of indents, which were made with a diamond tip on the microlayer PE sample (see section 2.3), is shown in the height image in Figure 27(a). This image was obtained several minutes after the indents were made. From the cross-section profiles taken along the indents in neighboring layers (Figure 27b) it can be seen that the indents in a ULDPE layer are larger at the base but shallower than the indents in the high-density PE layer. The difference in the indents' base is related to differences in the elastic moduli of these layers. The differences in the indents' depth reflect partial recovery of the indents in the rubbery ULDPE layer. The combination of nanoindentation with a heating accessory looks very promising because it could allow the indents' healing to be monitored at different temperatures.

Another example of nanoindentation of two different polyimide films, which were deposited on the Si substrate by spin coating, is given in Figure 27(c) and (d). The indents were made at different tip forces and the size of the indents increases accordingly. These data can be analyzed in terms of microhardness (tens to hundreds of newtons per square millimeter range) by dividing a load on the projection area of the indent. Also the shape of the indents on these films is different. Hill locks on the

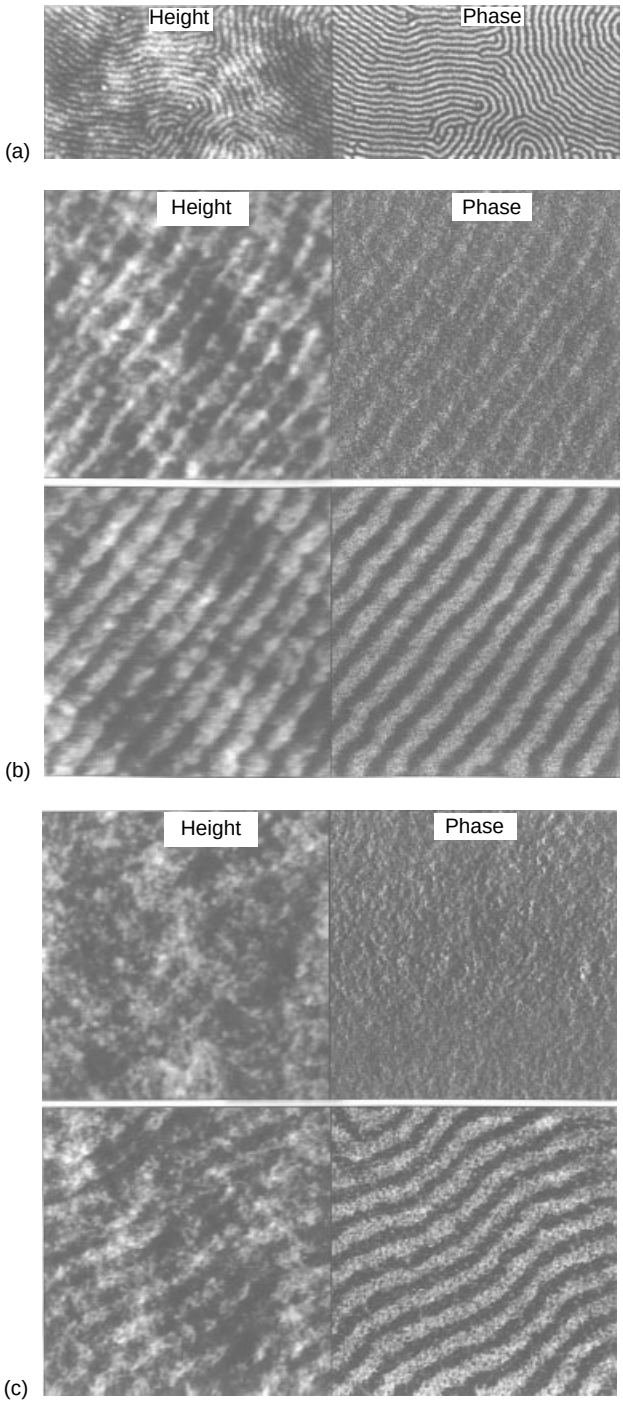

Figure 26 (a) AFM images of block copolymer PS-*b*-PMMA film, which are commonly recorded in tapping mode. Scan size 1 μm × 0.5 μm. (b) and (c) Images of PS-*b*-PMMA film obtained with *light tapping* (top) and with *hard tapping* (bottom) with (b) Si probes $k = 40\,\mathrm{N\,m^{-1}}$ and (c) with Si probe ($k = 2\,\mathrm{N\,m^{-1}}$). Scan size 0.5 μm × 0.5 μm in (b) and 0.6 μm × 0.6 μm in (c). In height images the contrast covers the 0–4-nm range.

borders of the indents, which are most pronounced in Figure 27(c), indicate plastic deformation of the polymer during indentation. Similar effects are observed in AFM images of scratches, which are made in mar studies of polymer coatings.[44]

4 HIGH-RESOLUTION IMAGING OF POLYMER STRUCTURES

Earlier AFM applications to polymer materials included studies of polymer morphologies on the micrometer and tens of micrometers scales and detection of molecular order in oriented samples and single crystals. At present, most efforts are concentrated on the imaging of polymer structures at the submicrometer scale. Such high-resolution imaging substantially complements TEM and diffraction techniques, which are used for analysis of nanometer-scale structures of polymers. In addition to imaging periodic lattices of highly drawn polymer films and single polymer crystals, these studies include visualization of individual polymer chains, the nanostructure of crystalline polymers and oriented samples, and microphase-separated morphologies of block copolymers. Although sample preparation for submicrometer AFM imaging might be nontrivial, in many cases it does not require thin sectioning, etching and staining, which is used for TEM of polymers. Direct visualization of nanometer-scale structures provided by AFM facilitates structural interpretation of the results of other characterization techniques. These data, for example, are invaluable for a correct choice of structural models employed for a rational analysis of diffraction data. However, the question of how surface structures, which are seen in AFM images of polymers, reflect their bulk structure should always be considered.

Several examples of high-resolution images are collected in this section, other examples can be found in following sections.

4.1 Visualization of Single Polymer Chains

Visualization of single DNA macromolecules deposited on different substrates was among the first exciting AFM applications.[45] Because DNA molecules of different lengths are generated in many ways, the sizing of DNA is the subject of many analytical approaches in molecular biology. For AFM studies, DNA molecules should be attached to a solid substrate first, and then imaged in air or in liquid with minimal tip–sample forces. The macromolecules appear as elongated strands in the images, Figure 28(a), and their contour length corresponds to the length of DNA as defined by the number of base pairs. Therefore, histograms that represent the contour length of the macromolecules seen in AFM images shows the molecular weight distribution of DNA. Nowadays, AFM

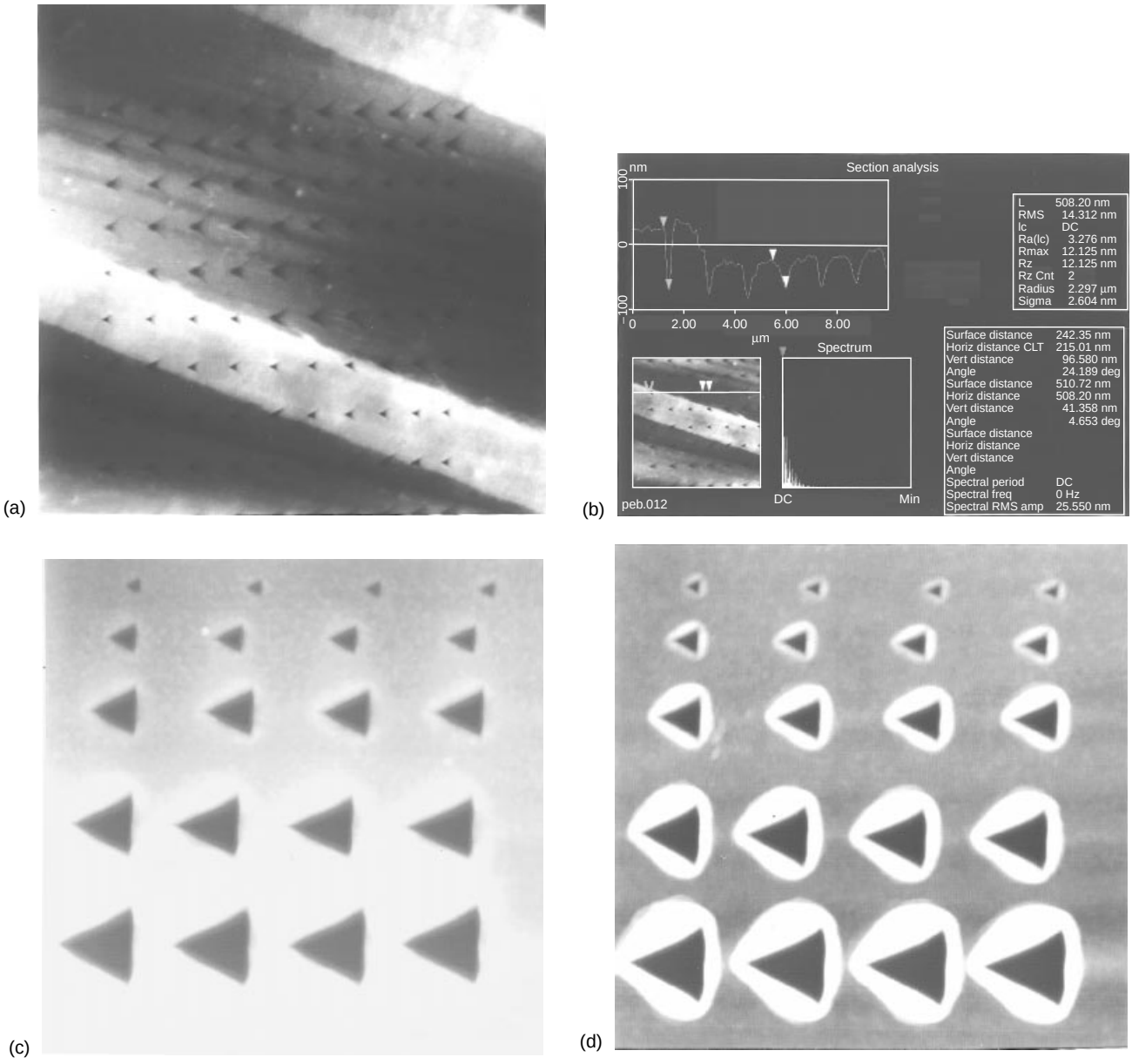

Figure 27 (a) Height image showing indents in multilayer PE obtained with 200-µN force. Scan size 10 µm × 10 µm. (b) Profiles of the indents. (c) and (d) Height images showing indents made in two different polyimide layers with different tip forces. Scan size 3 µm × 3 µm in both images.

is routinely used for examination of the shape and conformation of DNA strands, for observation of DNA kinks[46] and for studies of DNA–protein complexes.[47] Although the DNA length is correctly reproduced in the AFM images, its height and width cannot be determined precisely owing to tip-induced deformation and convolution of the apex shape of the tip on the macromolecule. The same is true for imaging of single polymer molecules. This problem might be resolved in the future with the introduction of sharper AFM probes and further lowering of the tip–sample force interactions.

Although AFM imaging of individual polymer macromolecules was demonstrated earlier,[41] it has only recently attracted a broad attention of researchers. The visualization of single polymer molecules and their transformation from a close-packed arrangement to an extended chain conformation was achieved in studies of single layers of PS-*b*-PMMA block copolymer. The block copolymer layer was transferred on a mica substrate from the water subphase in a Langmuir–Blodgett trough.[41] The AFM images of the complete layer of the block copolymer are shown in Figure 25. When an

incomplete layer of PS-*b*-PMMA was transferred on mica, only ''islands'' consisting of a bottom layer of condensed PMMA blocks and PS blocks on top were seen in AFM images. With an increase in humidity, chains of PMMA blocks underwent a conformational rearrangement from the close packed to the expanded, Figure 28(b). It is most likely that a thin water layer formed on the mica substrate facilitated this chain unfolding.

Visualization of polymer macromolecules is essential for characterization of novel polymers, which are synthesized for possible application in nanotechnology. Molecular conformations of monodendron-jacketed polymers have been examined by AFM.[48] The image of single macromolecules of jacketed PS, which have been deposited on a graphite substrate from a very dilute solution, is seen in Figure 29(a). The macromolecules exhibit a cylindrical shape and worm-like conformation. Alkyl chains, which are present in side groups, lead to

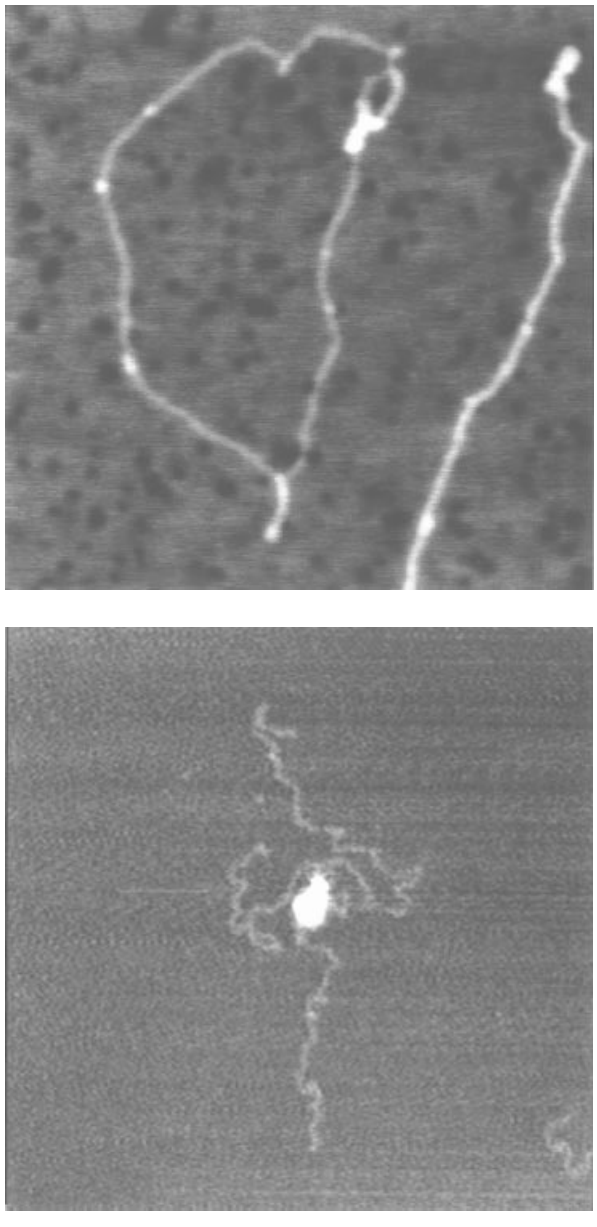

Figure 28 Height images of single macromolecules. (a) Double-stranded DNA deposited on mica, (b) PS-*b*-PMMA block copolymer on mica with unfolded PMMA block chains. Scan size 0.8 μm × 0.8 μm in (a) and 0.5 μm × 0.5 μm in (b).

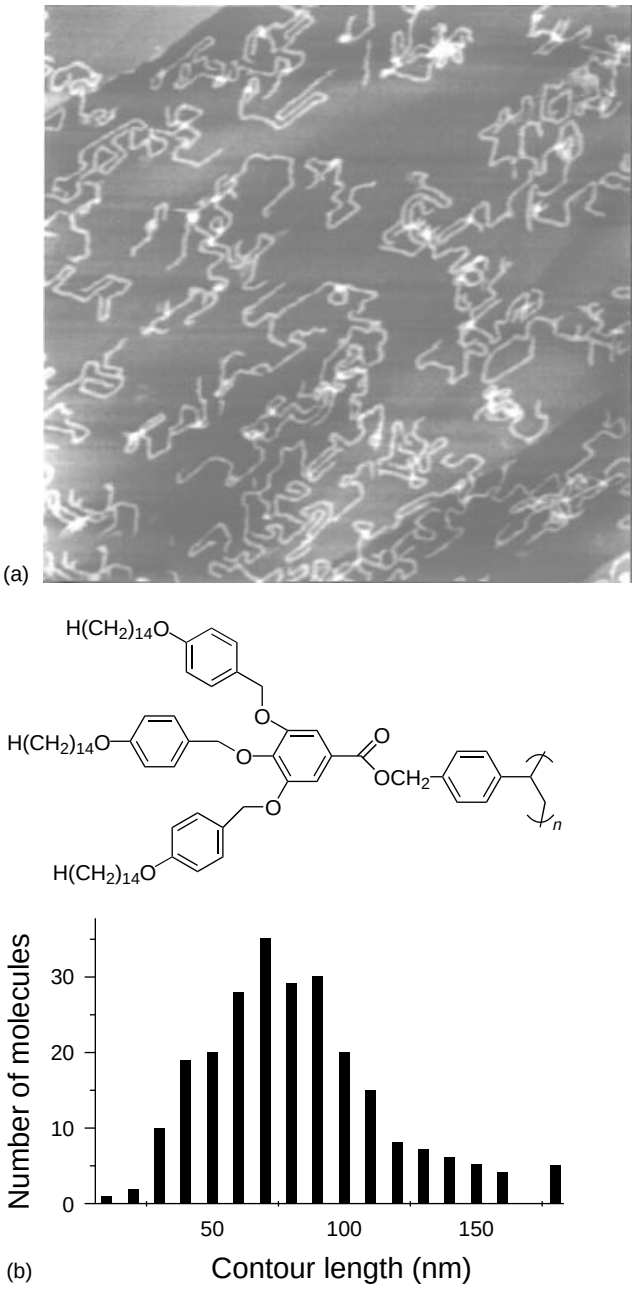

Figure 29 (a) Height image of single chains of monodendron-jacketed PS (structure illustrated) deposited on graphite. Scan size 0.51 μm × 0.51 μm. (b) Histogram of the length distribution of monodendron-jacketed PS.

some orientation order of the macromolecules adsorbed on graphite. A histogram of the length distribution of these macromolecules has been obtained from AFM images and is shown in Figure 29(b). These data complement the molecular weight distribution obtained by size exclusion chromatography and light scattering. In particular, all these methods have been applied to studies of polymacromonomers with a methacrylate main chain and, for the first time, the length per monomer of the cylindrical brush molecules was determined in solution and in the dry state.[49]

Direct visualization of polymer chain branches (some are seen in Figure 29a) is another attractive feature offered by AFM. This analytical capability is important because of a strong dependence of polymer properties on type and amount of branching. Further expansion of AFM visualization of single polymer molecules depends to a large extent on the sample preparation, which includes a proper choice of substrate and on the deposition method for spreading single macromolecules on the substrate.

4.2 Morphology and Nanostructure of Crystalline Polymers

AFM analysis of the morphology and nanostructure of a number of crystalline polymers provides a means for better understanding their structural organization. In particular, AFM data can be employed for the rational interpretation of the results of electron microscopy and diffraction techniques. An optical micrograph and AFM images of banded spherulites of HDPE/LLDPE (linear low-density polyethylene) blends are shown in Figure 30(a–d). These structures were studied in 1959, when only optical microscopy was available.[50] Banded spherulites found in several polymers are seen in optical microphotographs as sets of concentric extension rings caused by a birefringence phenomenon, Figure 30(a). This implies an alternative orientation of PE chains in the spherulites. In other words, the molecular axis (*c* is along the chain in PE) repeatedly changes its orientation at the location of the rings. It has been shown that the chain axis spirals around the radial direction. Surface corrugations associated with the banded spherulitic patterns are seen in an AFM image of the same sample (Figure 30b). It is most likely that the sample topography with elevated ring structures is the result of extraction of amorphous material as a result of permanganate etching applied to this sample. At greater magnification, the height and phase images reveal that the lower regions consist of linear structures with a width of ca. 20 nm. They also exhibit a well-defined pitch along the main direction, Figure 30(c) and (d). These linear structures could be assigned to the edges of the lamellae. Polymer chains in these regions are oriented in a tangential direction with respect to the spherulite radius. Elevated structures actually represent the main surfaces of the lamellae, which left the bulk material and adopted a flat-on orientation on the surface being examined. In these regions PE chains are oriented perpendicular to the surface. Therefore AFM images confirm that the change in PE chain orientation is responsible for birefringence of the banded spherulite. The phase image in Figure 30(d) shows that flat lamellar surfaces exhibit a granular structure, and this structure is responsible for the appearance of the pitch along the lamellar edges.

A granular nanostructure was found in AFM images of other PEs, ranging from single crystals grown from dilute solutions to samples of ultralow-density material.[21] Most of these images were obtained on the surface of melt-crystallized samples, which were not subject to any treatment. Spherulitic morphology and nanostructure of a low-density polyethylene (LDPE) sample, which was melt-crystallized, are presented in Figure 31(a–c). A large-scale height image shows that the spherulite of LDPE is not as well developed as the one shown in Figure 30(a). The height image, which was obtained on a part of this spherulite with *light tapping*, reveals the granular structure of the LDPE surface, Figure 31(b). With the force increased, the same region in the height image in Figure 31(c) is seen as being composed of a number of fibrils. These changes are reversible, and the grain structure is restored when the force is lowered again. Actually, the height image obtained with *hard tapping* resembles typical TEM micrographs of PE samples (Figure 31d), where bright linear strips represent lamellae. Because of chemical staining applied to PE samples in TEM, which is applied to provide a contrast between regions of different density, some details of polymer nanostructure can be omitted. Therefore, AFM images, which show not only a denser core with chain folding but also PE nanostructure with a compliant grain shape that is due to a less ordered exterior, describe the crystalline architecture of polymers more accurately. In other words, the fibrils present a more rigid core of linear sequences of several grains that merge with each other. This core is made from chain-folded polymer molecules. Under appropriate crystallization conditions, chain folding of linear polymer chains will be more extensive thus leading to formation of 2D-lamellae. In particular, this happens in single crystals of polymers, whose surfaces, however, still exhibit grain substructure.[51] A tentative structural model, which describes the nanostructure of crystalline polymers, is presented in Figure 31(e).

For polymer molecules, whose geometry does not facilitate chain folding, the grain nanostructure can be dominant. One of the relevant examples is ULDPE polymer with a large number of branches. AFM images of

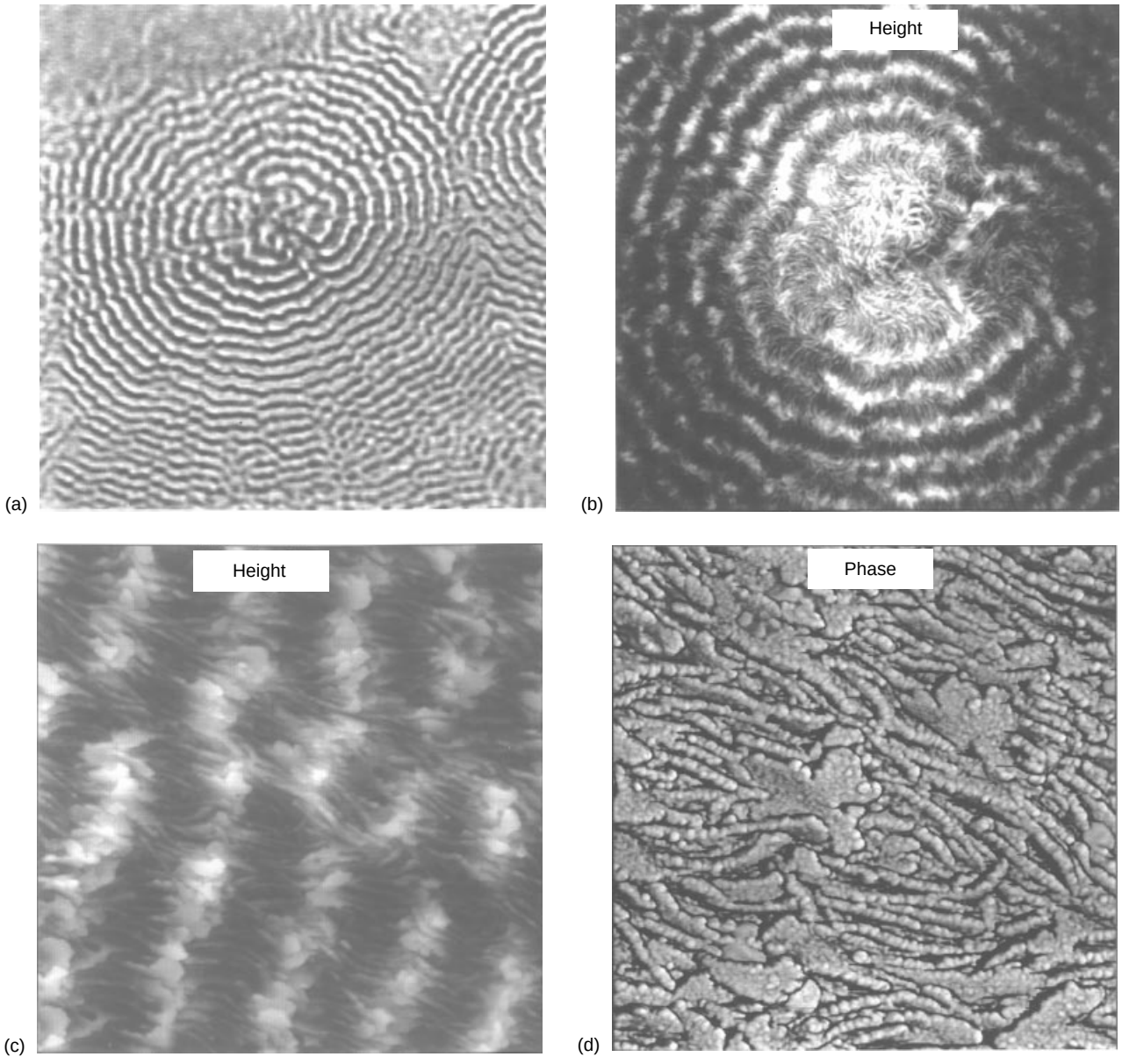

Figure 30 (a) Optical micrograph of banded spherulite of melt-crystallized HDPE/LLDPE blend. Area size 30 μm × 30 μm. (b) AFM height image of banded spherulite of the HDPE/LLDPE blend. Scan size 10 μm × 10 μm. (c) and (d) Height and phase images of a part of the banded spherulite. Scan size 1.5 μm × 1.5 μm.

melt-crystallized ULDPE are shown in Figure 32(a) and (b). Again, as in the case of LDPE, the grain nanostructure is seen mostly in the height image recorded with *light tapping*, Figure 32(a). When the force is increased, the same surface location is represented by a core structure which consists of relatively short lamellae or fibrils arranged into a network, Figure 32(b). This architecture is most likely to be responsible for the rubber-like behavior of this polymer. A similar nanostructure was also observed in PE with a slightly higher density of 0.87 g cm^{-3}. The changes observed in these AFM images have a reversible character.

Further evidence of the grain nanostructure of crystalline polymers is presented in images of melt-crystallized iPP, Figures 33 and 34. A large-scale height image (Figure 33a) shows iPP spherulite of the melt-crystallized sample. Part of this spherulite is presented at higher magnification in the height and phase images shown in Figure 33(b) and (c). Although there are tangential- and radial-aligned structures inside the iPP spherulite, small grains of 10–12 nm in diameter could be considered as elementary blocks of its architecture. These grains are distinctively seen in the phase image and some of them merge with neighbors to form linear tangential

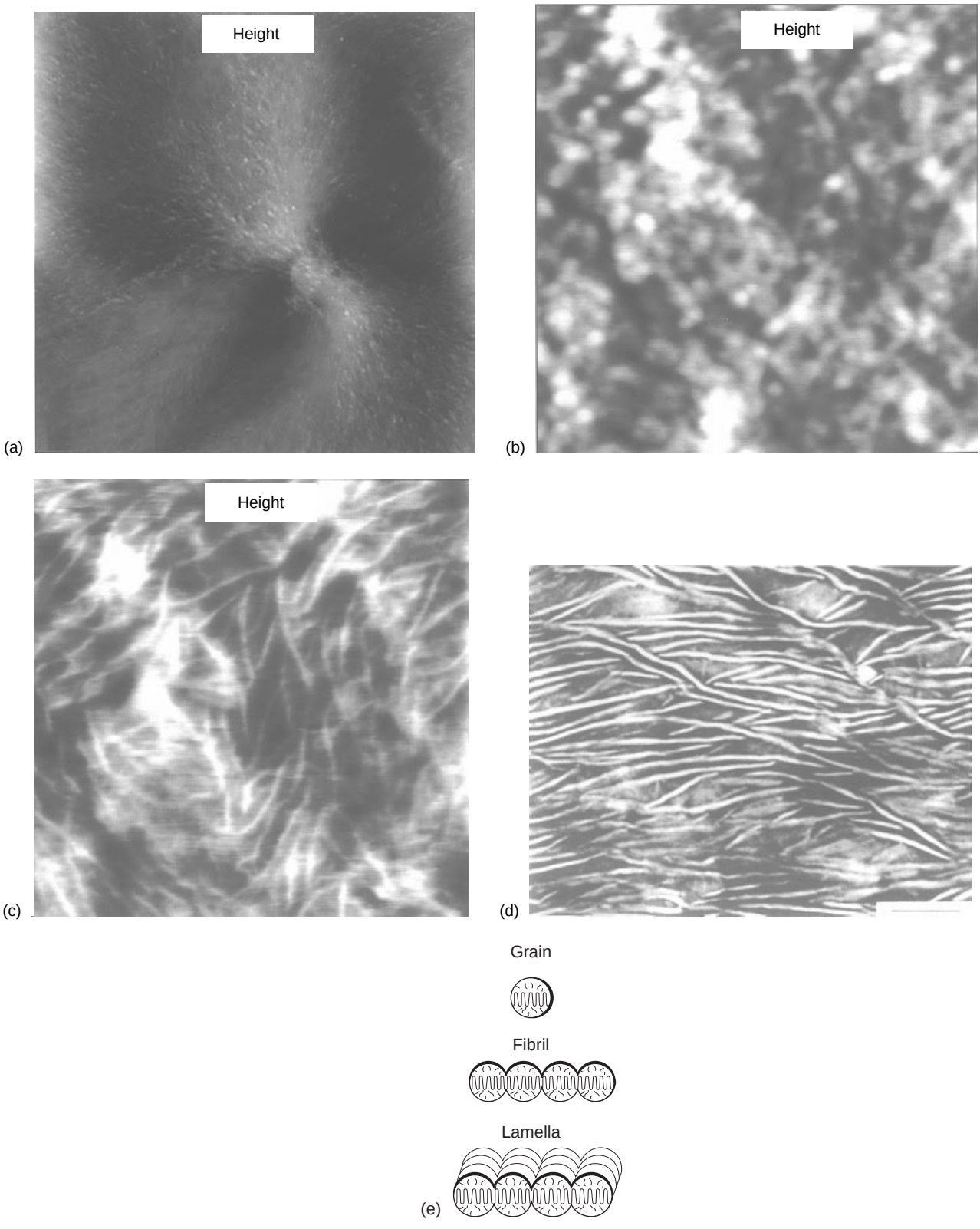

Figure 31 (a–c) AFM height images of melt-crystallized LDPE sample. Scan size 12 μm × 12 μm in (a) and 680 nm × 680 nm in (b) and (c). Images in (b) and (c) were obtained on the same location with light and *hard tapping*, respectively. (d) TEM micrograph of PE lamellae. (Reproduced from R.L. Morgan, M.J. Hill, P.J. Barham, *Polymer*, **40**, 337 (1999).) (e) Tentative model of nanostructure of crystalline polymers.

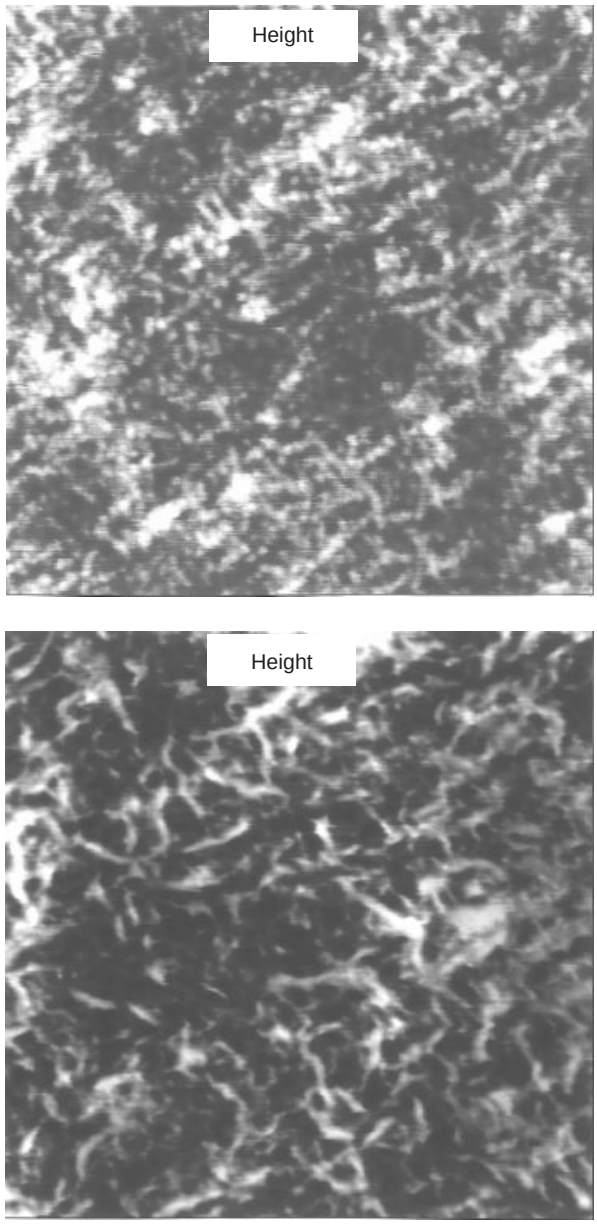

Figure 32 Height images obtained at the same surface location of melt-crystallized ULDPE in *light tapping* (a) and in *hard tapping* (b). Scan size 0.76 μm × 0.76 μm.

and radial sequences. It is worth noting that the structures are best seen in images recorded at relatively small tip–sample interactions. As seen from the phase image, the surroundings of the grain exhibit dark contrast indicating the presence of amorphous polymer in these locations (see section 5.1). With a force increase, the individual grains are not well distinguished because of the increased tip force-induced deformation and larger tip–sample contact area. In such cases, grain sequences are seen as fibrils.

Low-force imaging helps individual grains in iPP spherulites in the crystalline samples prepared not only from melt but also from solution to be visualized. Traces of the grain nanostructure are distinguished in AFM images of epitaxially crystallized iPP samples. The morphology of epitaxially crystallized iPP on anthracene is characterized by linear structures aligned along the main crystallographic axes of the low-molecular weight crystals, Figure 34(a) and (b). At higher magnification, these structures exhibit a pitch along the main direction, Figure 34(c). The presence of granular or block nanostructure was reported in a melt-crystallized sample of syndiotactic polypropylene.[53] The nanoscale surface corrugations, which are related to the grains (blocks) gradually vanished with thermal annealing of this sample. Therefore, the fibrillar structures often present in polymer spherulites can be described as linear sequences of tightly packed grain structures, which have a common dense interior formed from chain-folded molecules. Molecular-scale AFM images obtained in contact mode reveal packing motifs of the molecules.[18]

AFM studies of crystalline polymers suggest that individual grains rather than lamellae and fibrils can be considered as the elementary building blocks of crystalline architecture. For polymer molecules, whose chain structure in not flexible enough for extensive chain folding, grains of various sizes are seen inside spherulitic structures. This is the case for copolyesters (CPESTs),[51] polyketones, poly(ethylene terephthalate)[53] and other polymers. Although grains in the nanometer size range have been found in a large number of TEM micrographs of various polymers, their unambiguous assignment to true polymer structures was not possible owing to the involvement of chemical etching or staining in the sample preparation.[54] This is not a problem for AFM imaging and the existence of grain nanostructure is no longer in question. This viewpoint is not yet the dominant one, however, it is in accordance with already known theoretical approaches and experimental data. In a polymer textbook[55] Strobl suggested that the primary step in polymer crystallization is the unmixing of crystallizable and noncrystallizable chain parts, since this necessarily has to precede the formation and growth of crystallites. This process most probably leads to the formation of an individual grain with chain folding in its interior and with a less ordered exterior. Further crystalline growth could lead to more extended fibrillar and lamellar structures.

4.3 Nanostructure of Oriented Samples

The deformation of polymers is a routine procedure that takes place during different technological processes and is used for preparation of high-performance polymeric materials with outstanding mechanical properties. Therefore, understanding the morphology and nanostructure of

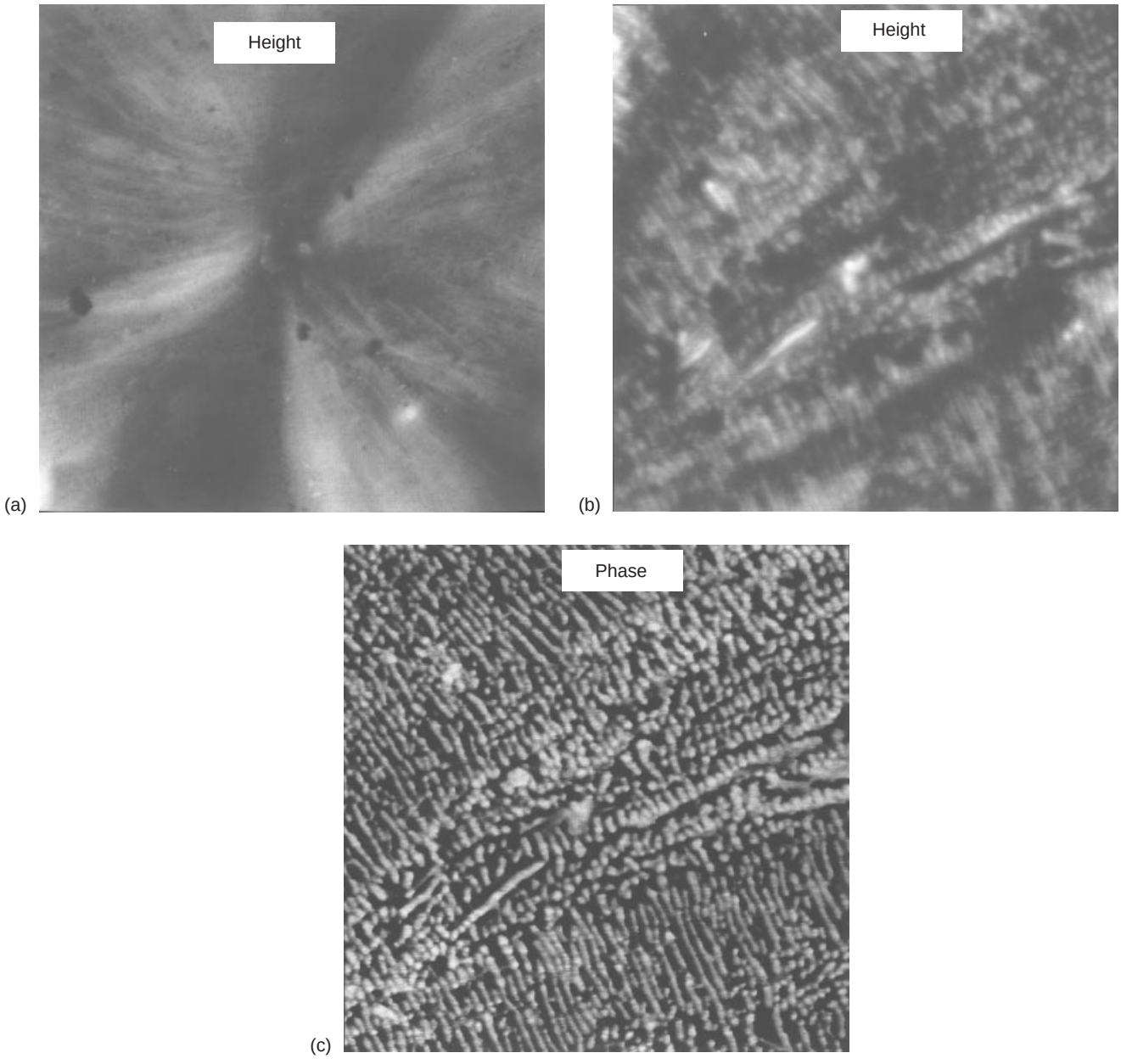

Figure 33 (a) and (b) Height, and (c) phase, images of a melt-crystallized sample of iPP. Height image in (a) was obtained on 8 μm × 8 μm area. Height (b) and phase (c) images were obtained simultaneously on 1 μm × 1 μm area.

oriented materials is of great fundamental and practical importance. Again, AFM could be an invaluable addition to other techniques, which are employed in the study of oriented polymers. AFM measurements of oriented polymers have been conducted for a number of years, starting with the molecular-scale imaging of elongated PE, iPP, polybutene-I and PTFE chains.[14–17] However, the observation of individual polymer chains is not of primary interest in such applications because observation of molecular-scale defects, which define primary events in the mechanical behavior of polymers, are not accessible with AFM. Therefore, AFM studies of the nanostructure of oriented samples, which is not completely characterized by other methods is most important.

Two examples of morphology commonly found in oriented polymers are shown in Figure 35(a) and (b). They show that though the fibrillar morphology dominates in oriented materials (Figure 35a) lamellar regions oriented perpendicular to the drawing direction can often be seen. This is observed in the images of a microporous iPP membrane (Figure 22) and in the images of melt-blown polymer films, Figure 35(b). AFM images

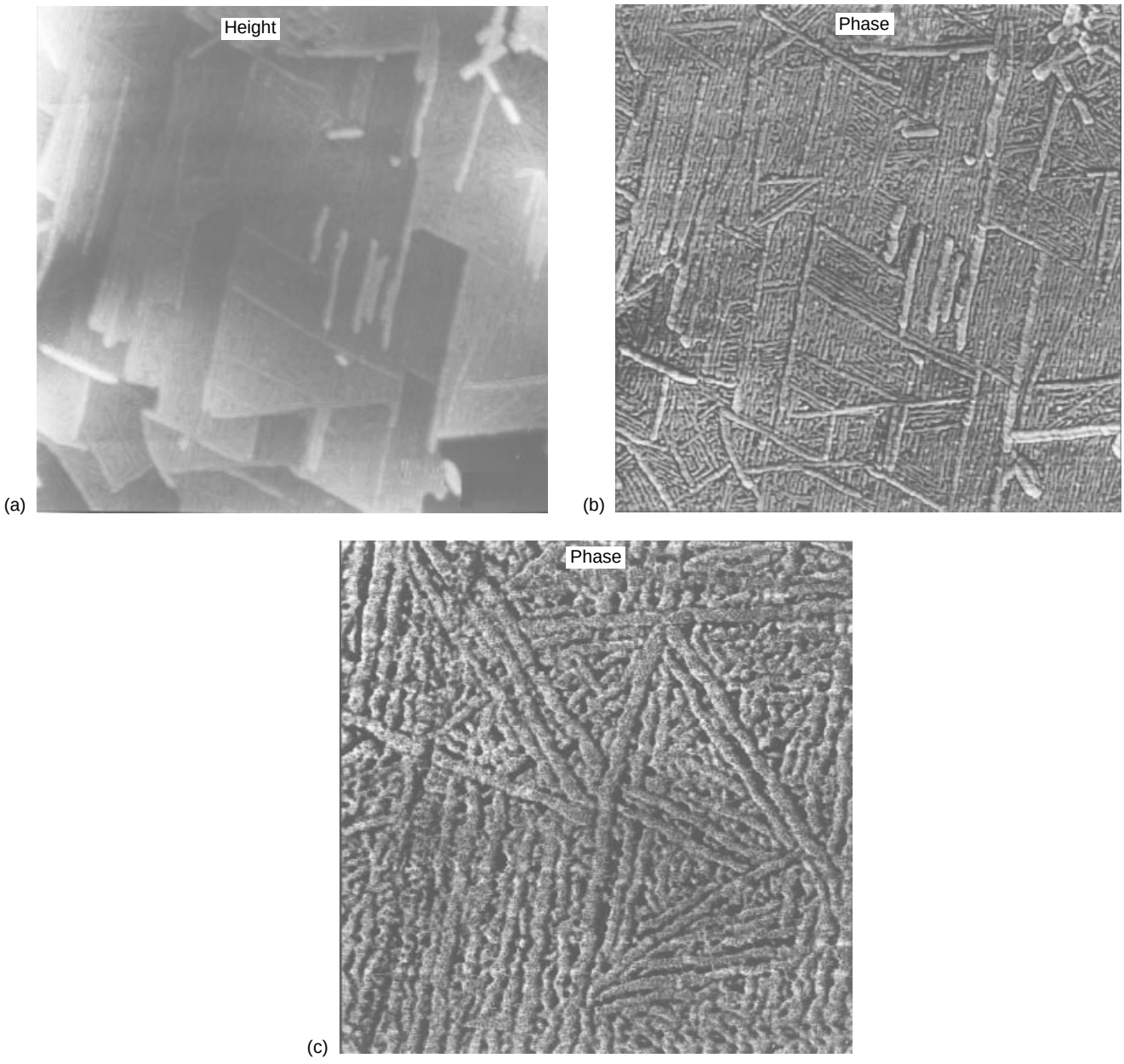

Figure 34 AFM images of iPP epitaxially crystallized on anthracene. Height (a) and phase (b) images were obtained simultaneously on 2.4 μm × 2.4 μm area. Phase image in (c) was obtained on 650 nm × 650 nm area.

of the ultradrawn PE fibers (Spectra™ 1000), whose nanostructure is considered to be fibrillar on the basis of X-diffraction studies, revealed tiny nanofibrils of 5–7 nm in diameter.[56] These nanostructures could be described as crystallites with fully extended chain crystallites, which are responsible for the ultrahigh modulus of these materials. However, in addition to nanofibrils, a large number of different lamellar structures were also detected in the topmost layer. To improve the adhesion of such fibers to the polymer matrix in composites their surface was etched with fluorine gas and the related surface changes were distinguished in AFM images.[56]

Oriented tapes of ultrahigh-molecular weight polyethylene (UHMWPE) with different draw ratios have been extensively studied with AFM in the contact and tapping modes.[57] The fibrillar morphology of these tapes has been characterized by bundles of microfibrils with diameters in the 4–7-micron range, by individual microfibrils with diameters between 0.2 and 2 μm and nanofibrils with diameters in the 5–10-nm range. All these fibrils are oriented along the stretching direction and they are

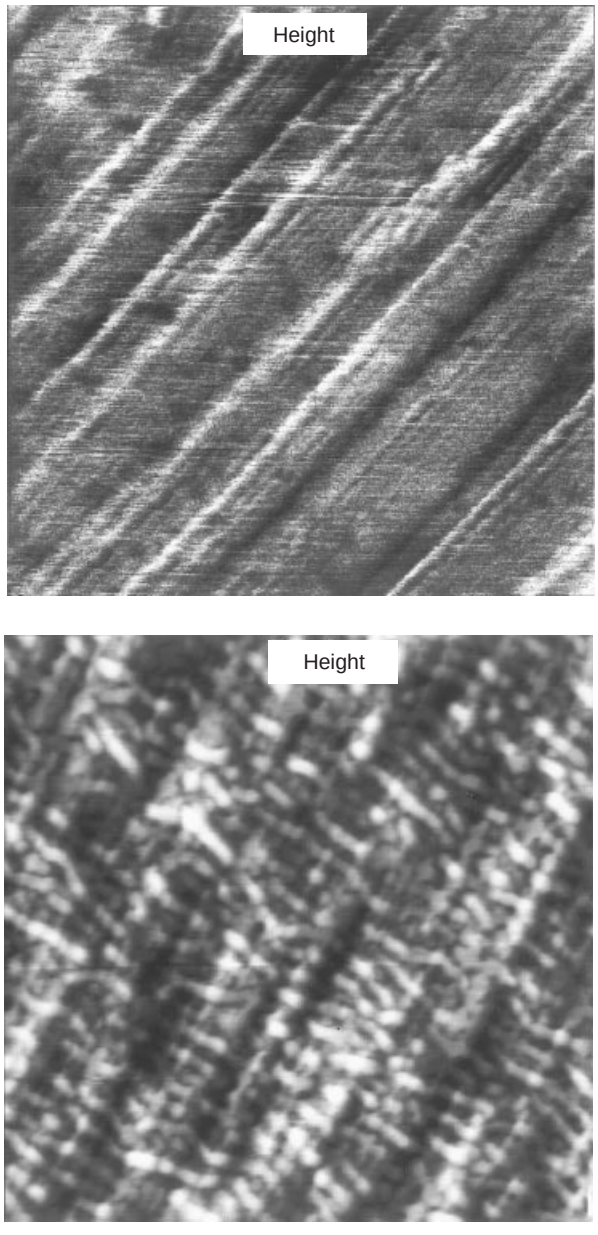

Figure 35 (a) AFM images of uniaxially drawn PE tape (draw ratio 70) and (b) PE film drawn from melt. Scan size 0.5 μm × 0.5 μm in (a) and 1 μm × 1 μm in (b).

present in the sample bulk. Low-force imaging of the UHMWPE tapes, which was achieved in imaging under water, showed the presence of a weakly bound overlayer formed from nanofibrils (3–5 nm in diameter), which are oriented perpendicular to the stretching direction.[58]

AFM images of LLDPE film before and after uniaxial deformation are shown in Figure 36(a) and (b). The original material consists of ''islands'' of lamellar aggregates embedded into an amorphous polymer (see also section 5.1 and Figure 40). During the uniaxial stretching, plastic deformation occurs with the formation of a neck. In the necked part of the LLDPE sample, previously amorphous regions have been transformed into fibrillar domains and the lamellar aggregates were deformed in the draw direction, Figure 36(b). On further deformation all of the material becomes fibrillar.

The evolution of the crystal morphology during deformation proceeds in different ways but clarification of the polymer nanostructure and its changes during orientation is the key to a rational description of structural changes induced by deformation. In addition

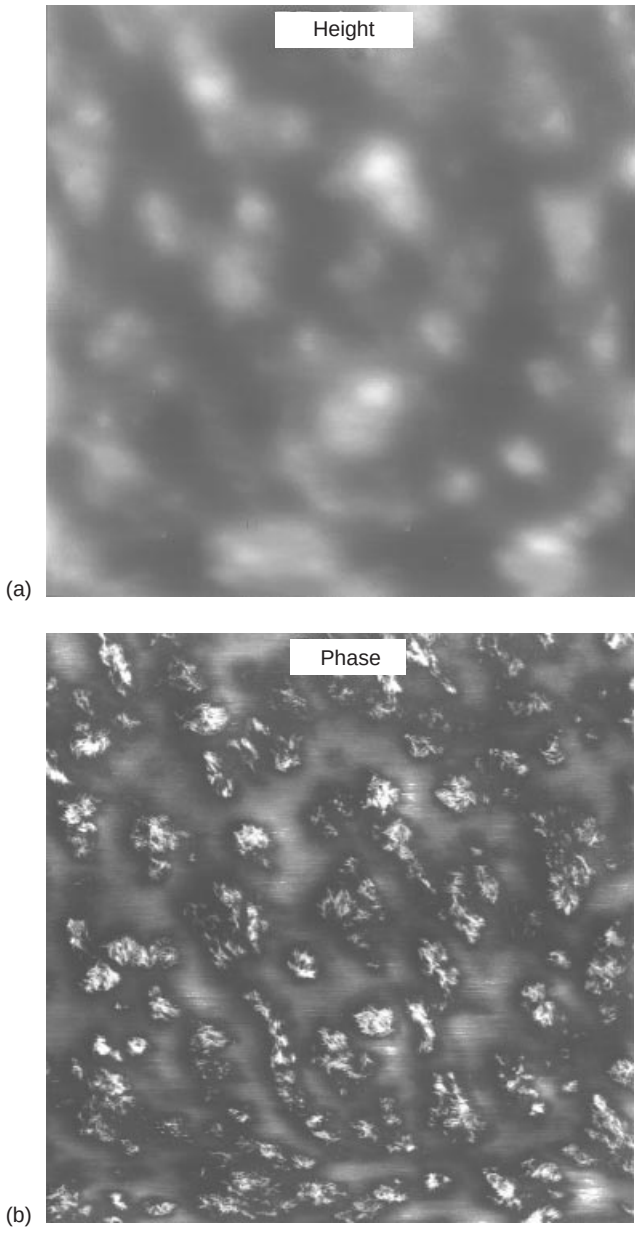

Figure 36 AFM images of LLDPE film before (a) and (b) and after (c) and (d) uniaxial stretching. Scan size 20 μm × 20 μm in all images.

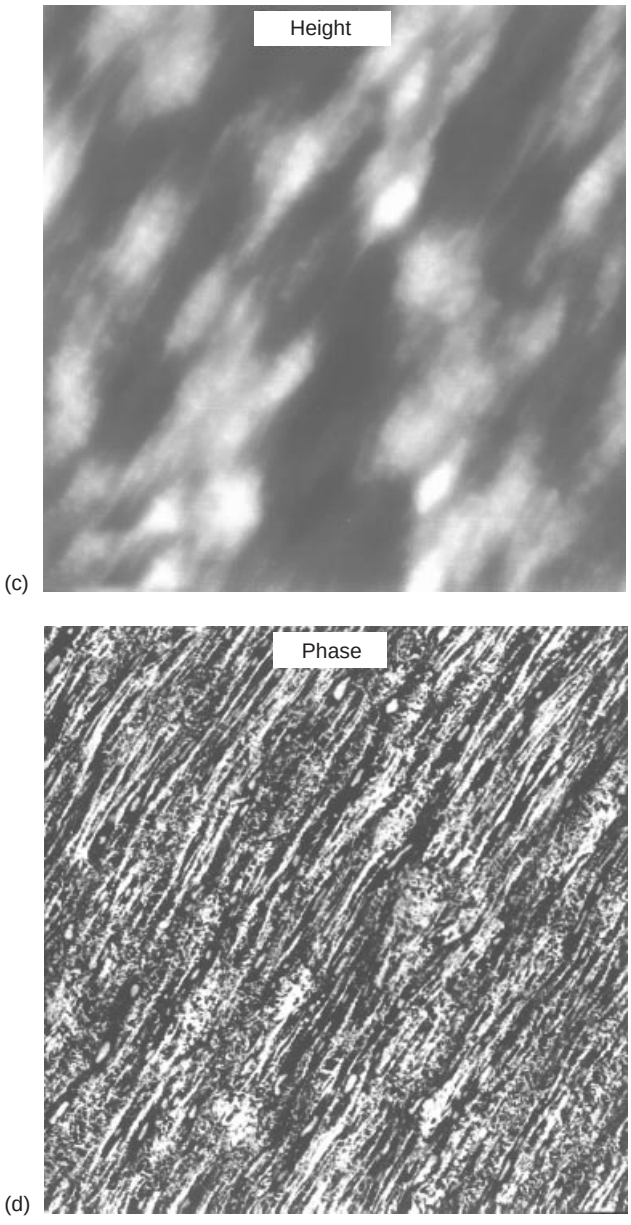

Figure 36 (*Continued*)

to uniaxially drawn polymer materials, biaxially oriented films have been studied with AFM. The morphology and nanostructure of biaxially oriented PVDF film is shown in Figure 37(a) and (b). The morphology is characterized by linear structures, which are aligned in the perpendicular direction. At higher magnification, tiny fibrils (3–5 nm in diameter) and shish-kebab structures have been seen in the phase images. The visualization of lamellar organization in the blown polymer films, which is achieved in AFM images (Figure 38), is of primary importance for the description of polymer orientation and its correlation with the mechanical properties of these films. Diffraction techniques, which are routinely used for these studies, require structural models for the interpretation of data. The most rational models could be built based on the AFM images, which directly show lamellar orientation. Comparative studies of biaxially oriented PE films with X-ray diffraction, electron microscopy and AFM have been performed.[59] Since then, AFM imaging has been further improved and observation of lamella structures in different polymer samples is a routine procedure.[60]

A choice of the parameters for AFM imaging is also important in studies of oriented polymer samples. Even in the contact mode, in AFM studies of oriented polymers

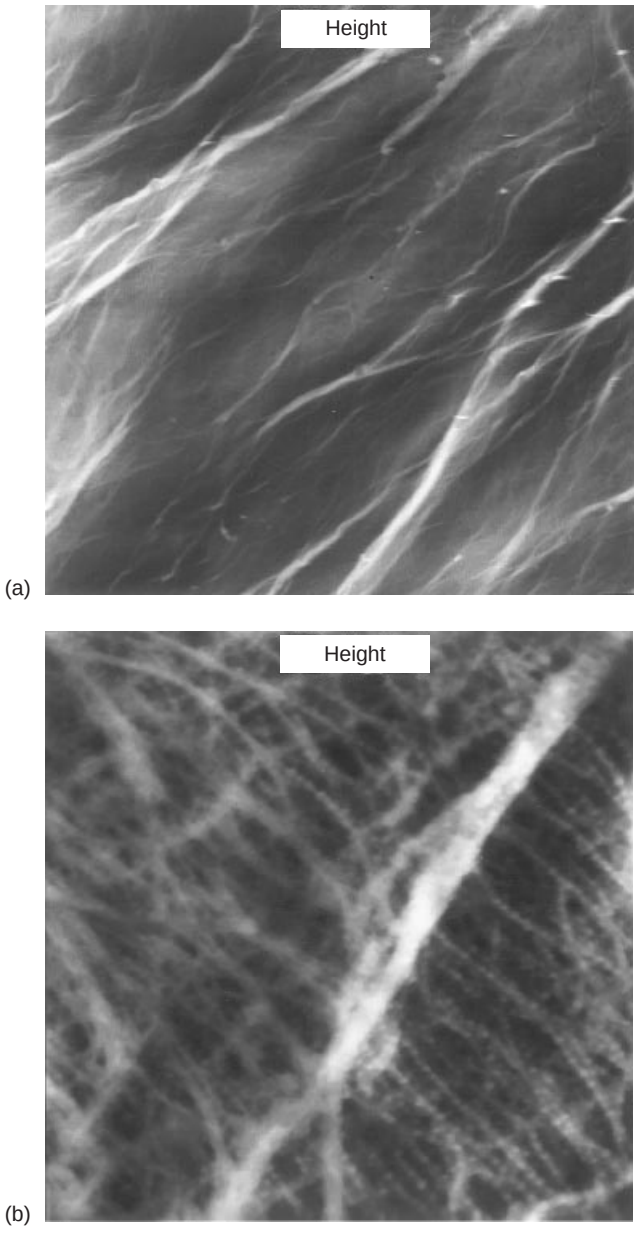

Figure 37 AFM images of biaxially oriented PVDF film. Scan size 25 μm × 25 μm in (a), 1 μm × 1 μm in (b) and 3 μm × 3 μm in (c).

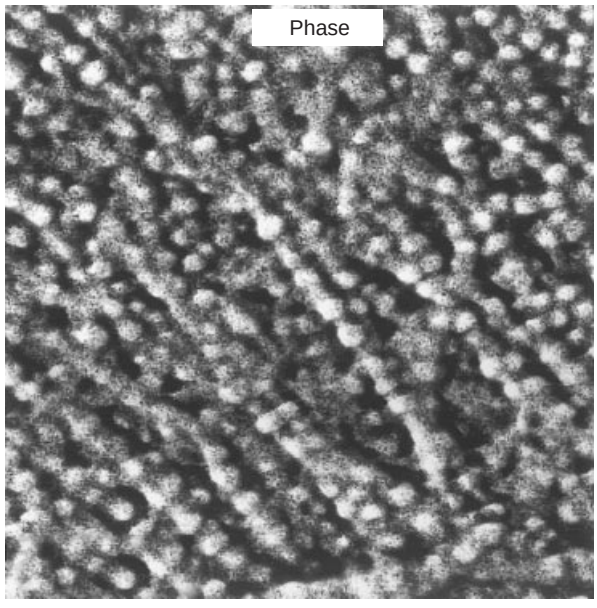

(c)

Figure 37 (*Continued*)

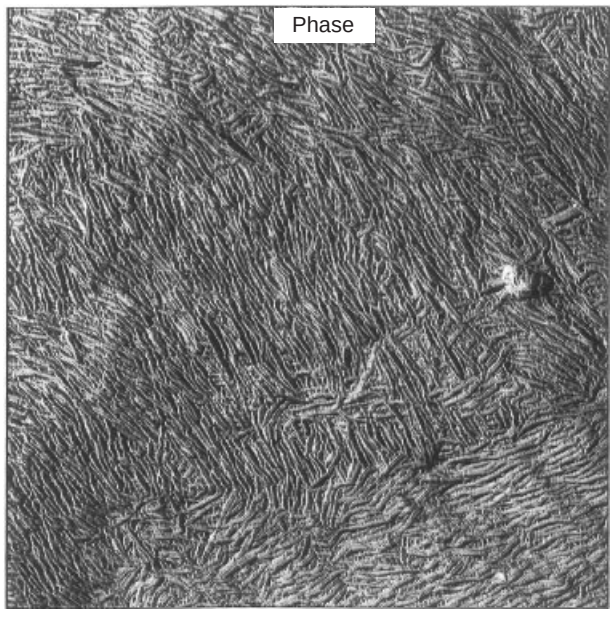

Figure 38 AFM image of biaxially oriented HDPE film. Scan size 3 μm × 3 μm.

which were performed before the appearance of the tapping mode, the important role of the tip–sample force interactions was found. Images of the oriented film of UHMWPE (draw ratio of ~8) obtained at different tip forces are shown in Figure 39. With increased force on the nanofibrils, which are oriented along the drawing direction, regions with alternating contrast can be distinguished. Molecular-scale images taken along the fibrils reveals alternation of bright regions with extended chains with those which exhibit inferior order, Figure 39(c). These variations are of ~25 nm and correlate with the long period found in small-angle X-ray scattering (SAXS) patterns of oriented polymer samples. A structural model, which is commonly used for a description of these features, is shown in Figure 39(d). It is also consistent with AFM data. The contrast in AFM images along the nanofibrils could be explained

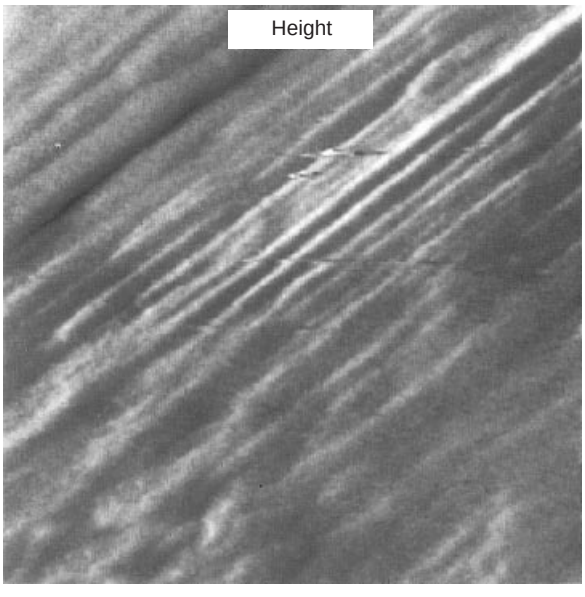

(a)

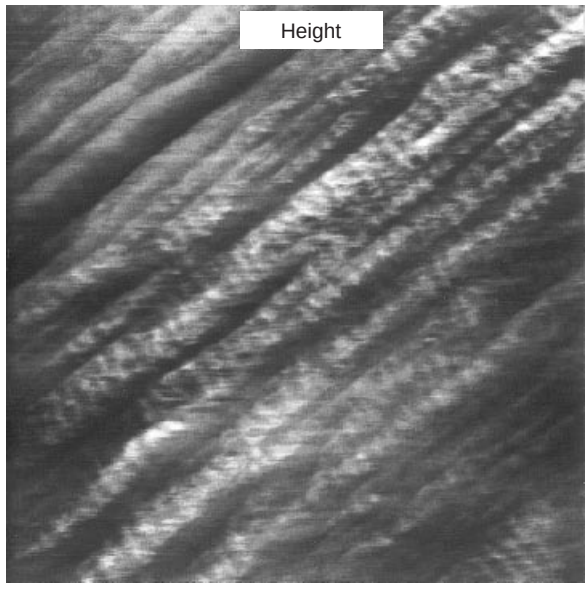

(b)

Figure 39 AFM images of uniaxially drawn PE tape (draw ratio 20). (a) and (b) Images obtained at the same locations at low and high force, respectively. Scan size 0.75 μm × 0.75 μm. (c) High-resolution image revealing sample regions with extended chain packing. Scan size 54 nm × 54 nm. (d) Scheme explaining "*long period*" structures in oriented polymers.

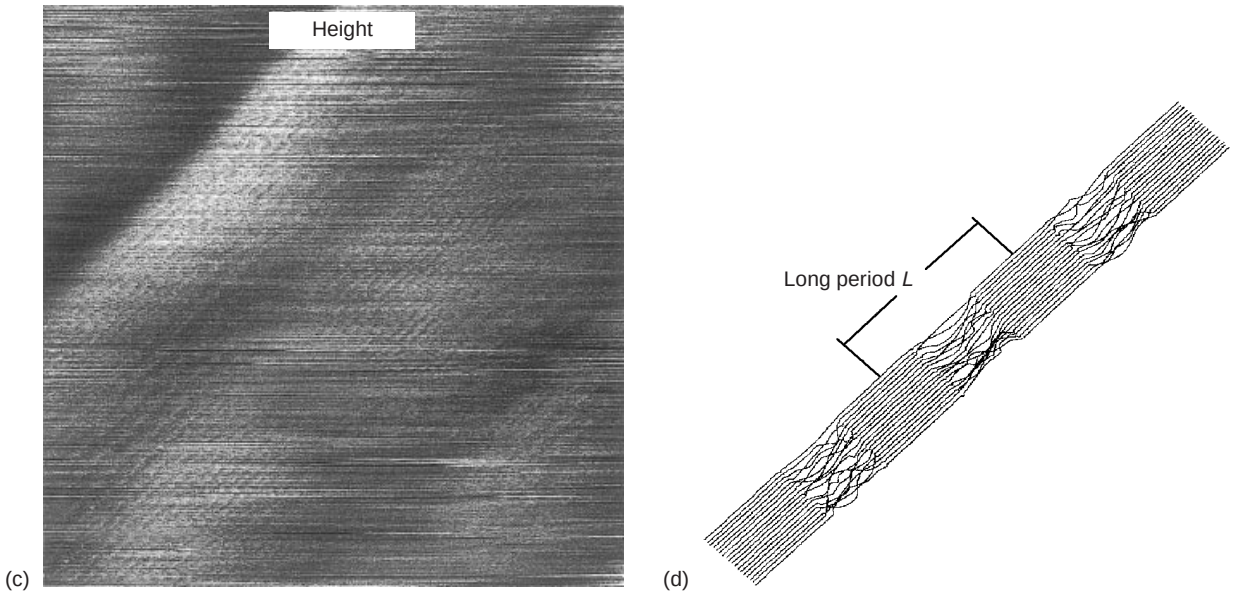

Figure 39 (*Continued*)

by stiffness variations, which are due to more and less regularly arranged PE chains.

The appearance of the long period contrast in high-force AFM images obtained in the contact mode was only achieved by damaging the topmost polymer layer. In tapping mode, the long-period contrast was also observed in *hard tapping*, Figure 40(a) and (b). The image changes observed on transition from *light tapping* to *hard tapping* were reversible. This means that in contrast to variable force imaging in the contact mode, observation of the long period features in the tapping mode occurs without sample damage. Here, it is worth noting that the grain nanostructure, which is seen in AFM images in a number of crystalline polymers, is of the same dimensions as the long period observed in oriented polymer samples. However, the grain nanostructure is observed only in low-force imaging and it is related to the exterior shape of the folded-chain core, whereas the long period variations are related to the alternation in order along the fibrillar core, and they are detected in high-force imaging.

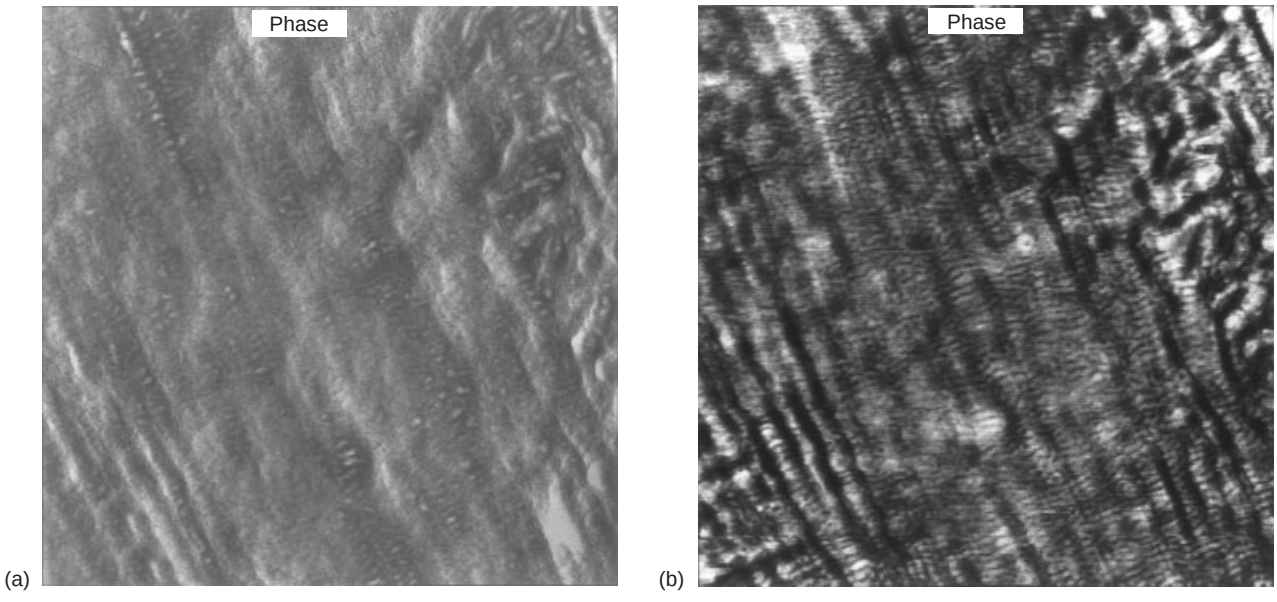

Figure 40 AFM images of oriented PE film obtained on the same location in (a) *light tapping* and (b) *hard tapping*. Scan size 1.22 μm × 1.22 μm.

5 COMPOSITIONAL IMAGING OF HETEROGENEOUS POLYMER SYSTEMS

Structural heterogeneity is a common feature of many polymer samples. This is true not only for polymer blends and block copolymers, but also for most crystalline polymers in which amorphous material is a necessary component. Only amorphous polymers such as atactic PS or PMMA could be considered as homogeneous. In characterizing heterogeneous polymer systems, knowledge of distribution of components is one of the most important properties. When applying AFM to such characterization one should realize that optimization of experimental parameters is essential for high-contrast mapping of multicomponent polymer samples. In many cases, this is achieved with *hard tapping* and the images demonstrate the distribution of individual components on the sample surface. When the probe in a polymer sample passes the viscoelastic top layer, it can penetrate into the sample and subsurface structures can be seen in the images obtained at the elevated tip forces.

The examples, which are given below, show that AFM is a powerful technique for compositional imaging of heterogeneous polymer systems.

5.1 Visualization of Amorphous and Crystalline Components

In analysis of AFM images of crystalline polymers it had been shown that the less ordered exterior surrounds close packed core in grains, fibrils and lamellae. This top layer could be described as an amorphous material bound to the crystalline structure. Traces of free-amorphous material were observed in the phase image of iPP spherulites, Figure 33(c). Additional amorphous components such as wax in industrial PE films, might dominate in surface layers. This is evident from AFM images of LLDPE blown film, which were obtained at different A_{sp}, Figure 41(a–c). In *light tapping*, only a slight surface corrugation is seen, but with the increase in the tip force the fibrillar aggregate appears more and more distinctively. Similar changes have been seen in the phase images. In this case, the probe penetrates through the topmost amorphous layer, which is in the rubber-like state at room temperature (RT). Therefore, the images at elevated forces show subsurface structures. The estimates of the tip penetration, which were made with the force curves, showed that the image in Figure 41(c) was obtained with the probe ca. 20 nm beneath the top surface.

The way a sample is prepared defines its morphology and nanostructure. Comparison of morphology of a blown LLDPE film and a LLDPE sample that was melt-crystallized with a surface facing Si substrate, is given in Figure 41(d) and (e). The phase image of the LLDPE film shows a number of lamellar aggregates embedded in an amorphous polymer. The surface of the melt-crystallized LLDPE sample is completely covered with lamellae, and some of them are seen lying flat. Both images were obtained with *hard tapping* but only for the first sample was the transition from *light* to *hard tapping* accompanied by drastic changes.

Segmented polyurethanes present structures consisting of hard and soft blocks. In such materials, phase separation may precede crystallization or occur simultaneously. AFM reveals crystalline and amorphous components in such systems.[61] The height and phase images obtained with *hard tapping* on a microtomed surface of polyurethane, which was prepared from hard blocks of MDI-butadiol (MDI-diphenylmethane diisocyanate) and soft blocks of poly(ethylene glycol), are shown in Figure 42(a–c). The height image (Figure 42a) shows traces of the microtome knife and several spherulites, which are formed of hard blocks. The spherulites are more pronounced in the phase images (Figure 42b and c), where crystalline structures (bright contrast) are seen

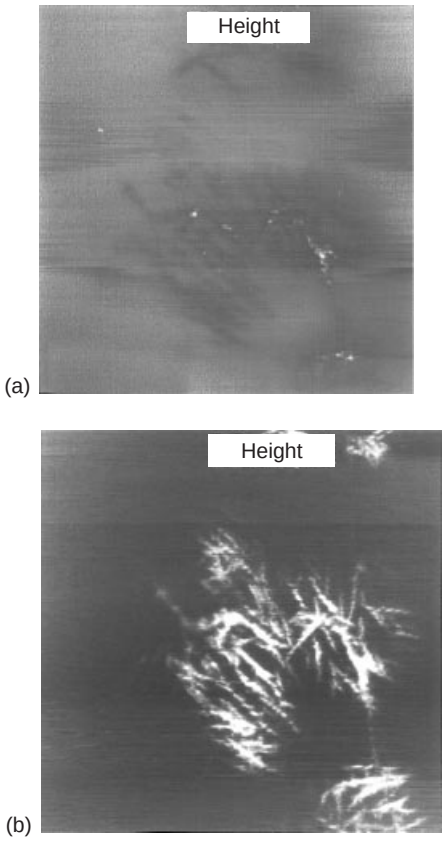

Figure 41 (a–c) AFM images of LLDPE film with *light*, moderate and *hard tapping*, respectively. Scan size 1.5 μm × 1.5 μm. AFM images of (d) LLDPE film and (e) melt-crystallized LLDPE sample obtained with *hard tapping*. Scan size 5 μm × 5 μm in (d) and 10 μm × 10 μm in (e).

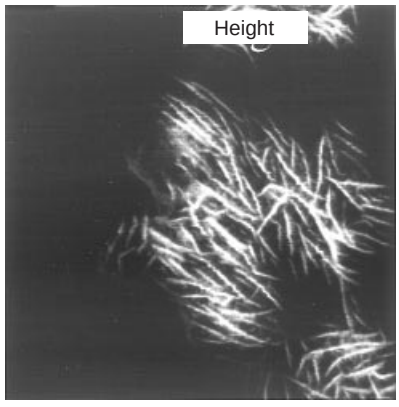

(c)

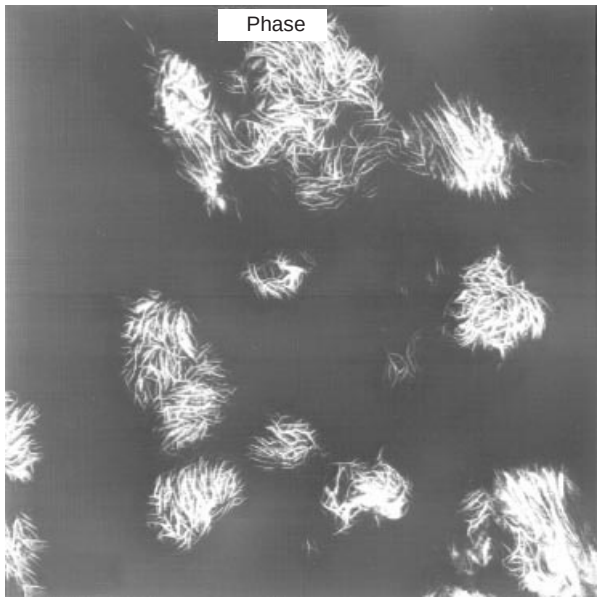

(d)

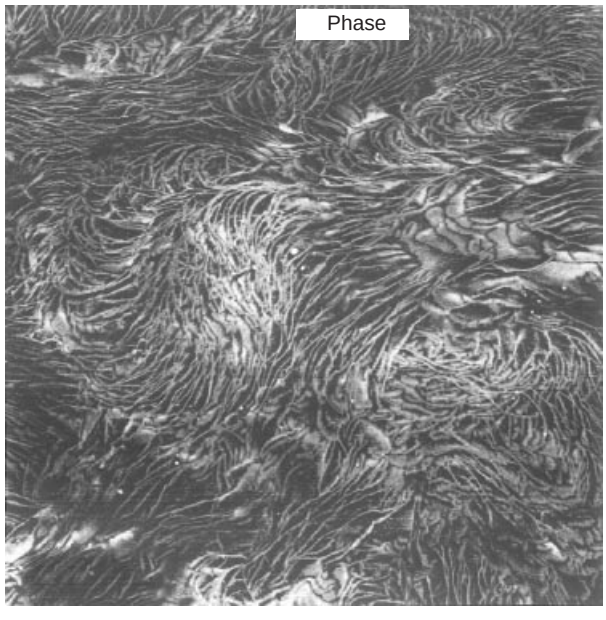

(e)

Figure 41 (*Continued*)

embedded in amorphous material (dark contrast). The phase image at high magnification (Figure 42c) indicates that individual fibrils, which form the spherulites, exhibit a grain substructure.

Besides amorphous and crystalline forms many polymers also exist in a partially ordered or mesomorphic form. Such organization is common for liquid crystalline polymers and polymers with a flexible

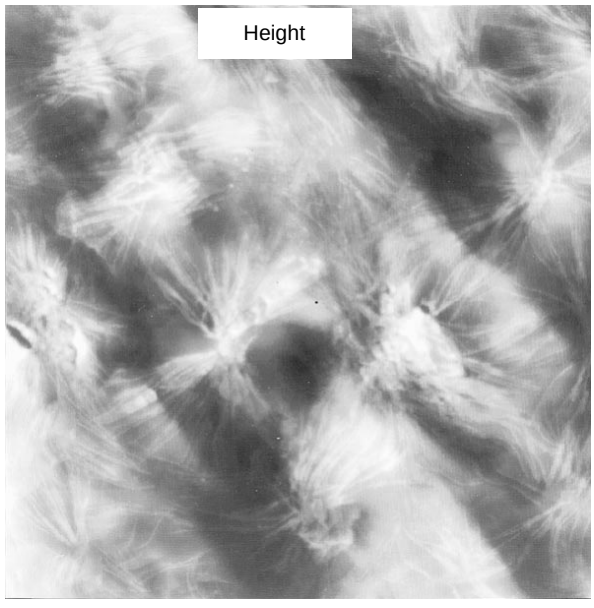

(a)

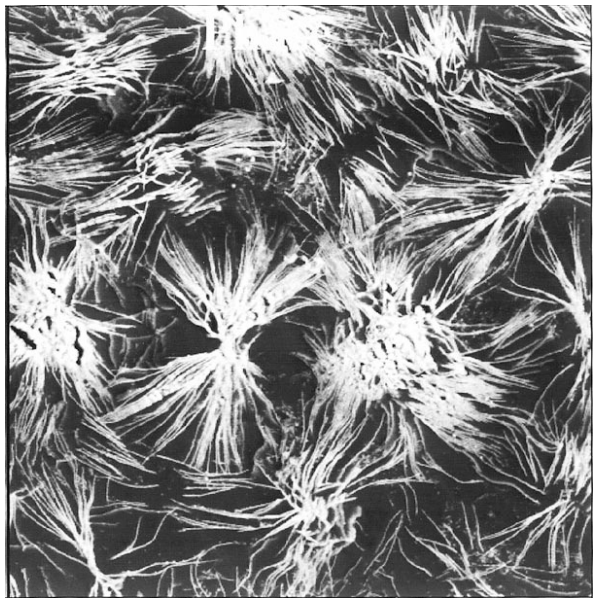

(b)

Figure 42 (a) Height and (b) and (c) AFM phase images of microsegregated polyurethane, which were obtained with *hard tapping*. Height (a) and phase (b) images were obtained simultaneously on 6 μm × 6 μm area. Phase image in (c) was obtained on 0.5 μm × 0.5 μm area.

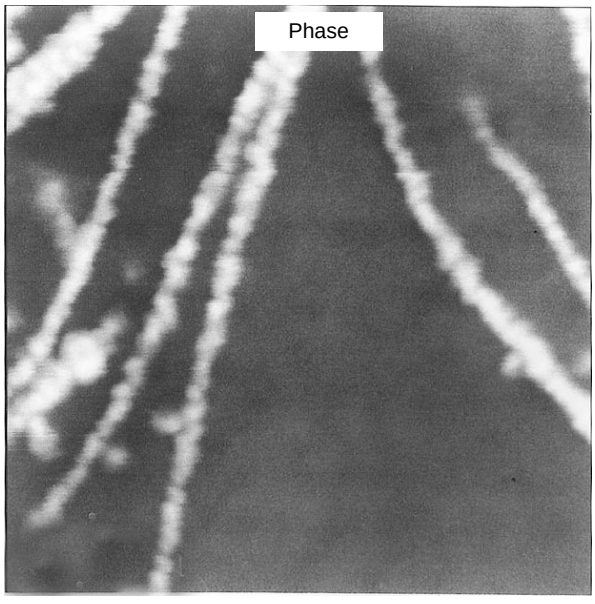

(c)

Figure 42 (*Continued*)

(−Si−O−) backbone.[62] AFM images of several mesomorphic polysiloxanes will be discussed below and also in section 6. The height and phase images of several poly(diethylsiloxane) (PDES) patches, which were deposited on Si substrate by rubbing, are shown in Figure 43(a) and (b). The images were obtained with *hard tapping*. The height image shows striated droplets of polymer that do not wet the hydrophobic Si surface. The phase contrast indicates three different materials present on the surface. The locations of Si substrate are the brightest, whereas dark spots belong to amorphous PDES. Within the polymer droplets brighter lamellar aggregates are seen, which are oriented perpendicular to the rubbing direction. They are practically indistinguishable in the height images. The presence of extended chain organization in PDES lamellae was suggested from X-ray diffraction and electron microscopy data. Structural reorganization that occurs on heating and cooling of these samples is described in section 6.

Heterogeneity of another mesomorphic polymer, cyclolinear polysiloxane[63] is evident from the phase images recorded with *light* and *hard tapping*, Figure 43(c) and (d). Traces of planar structures, which are seen with *light tapping*, can be assigned to lamellar layers lying-flat on the sample surface. With *hard tapping*, multiple linear features of 20–30 nm in width are seen through all polymer patches. It is most likely that these features represent edges of lamellae, which exist in the sample bulk. In this case, transition from *light* to *hard tapping* led to damage of the ordered top layer.

5.2 Imaging Polymer Blends

For illustration of AFM compositional imaging of polymer blends, three types of samples were chosen. They include microlayer samples, rubber toughened plastics and binary latex blends.

A microlayer material with alternating PE layers of different density, which was described in section 2.2, Figure 7 is the most suitable sample for exploring the AFM capabilities for compositional imaging.[21] The

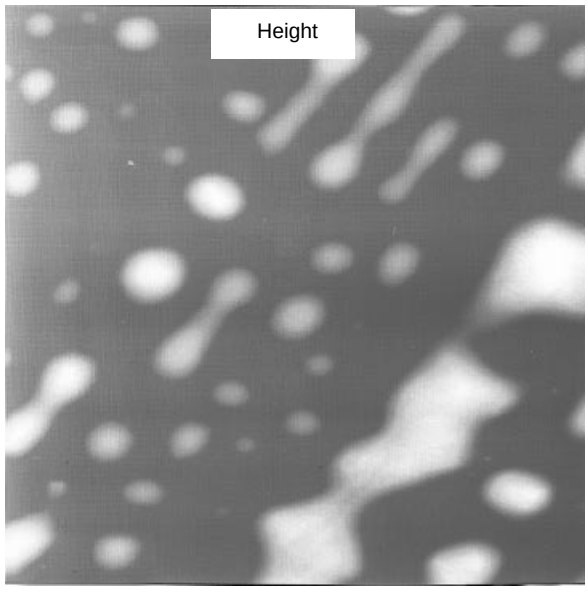

(a)

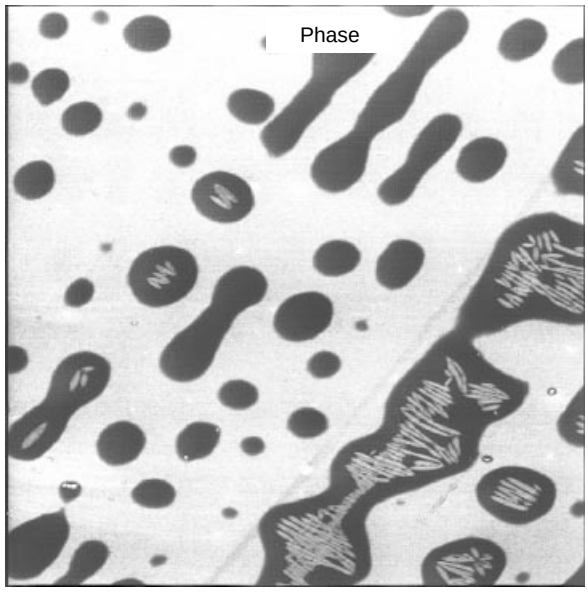

(b)

Figure 43 AFM images of (a) and (b) PDES and (c) and (d) cyclolinear polysiloxane. Height (a) and phase (b) images were obtained simultaneoulsy on 30 μm × 30 μm area. Images in (a) and (b) were obtained with *hard tapping*. Phase images in (c) and (d) were obtained with *light* and *hard tapping*, respectively. Scan size 2.5 μm × 2.5 μm.

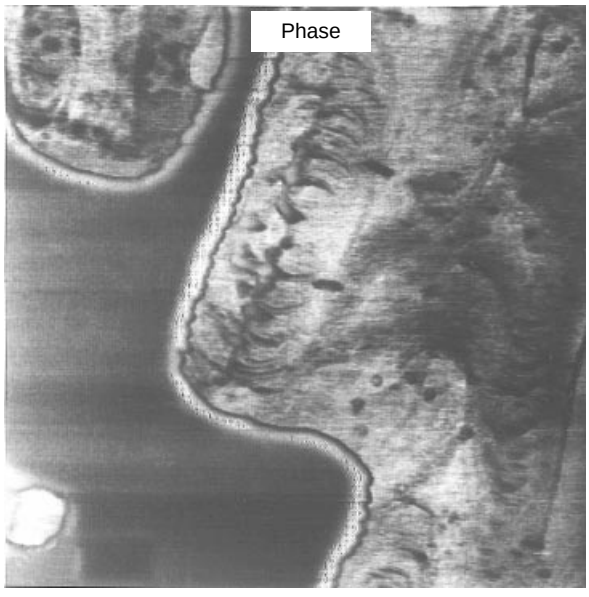

(c)

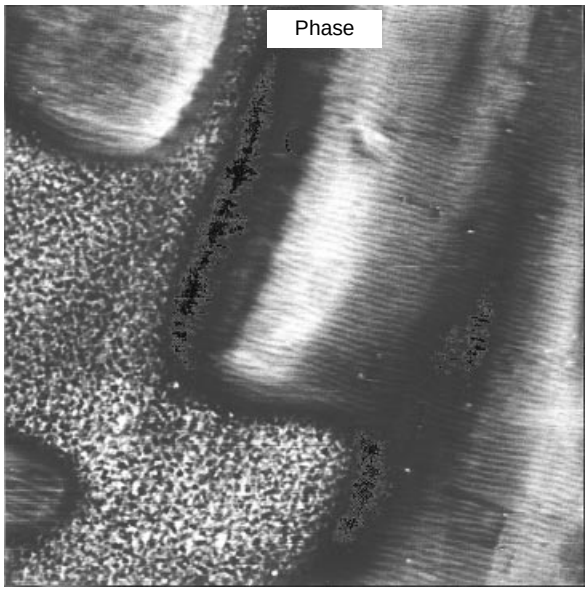

(d)

Figure 43 (*Continued*)

height and phase images of this sample obtained on the microtomed surface are shown in Figure 44(a) and (b). In the height image the interface borders between the wide layers of ULDPE ($\rho = 0.86\,\mathrm{g\,cm^{-3}}$) and narrow layers of PE ($\rho = 0.92\,\mathrm{g\,cm^{-3}}$) are barely distinguishable from other scratches left by an ultramicrotome knife. In the phase image these layers can be easily differentiated. The high-density layers are substantially darker than the ULDPE layers that are seen when small A_0 are used for these images. (In the phase image obtained with *hard tapping*, these layers are darker.) The high sensitivity of the phase contrast to density of PE allows the mixing of the components, which proceeds during high-temperature annealing of this blend to be monitored. The phase image obtained on the annealed sample, Figure 44(c), shows that individual fibrillar and spherulitic structures (both seen as dark strips) appear in the ULDPE layer and, especially at the interlayer interface. ULDPE does not crystallize in this fashion, therefore, this image shows that, during annealing, some polymer from the high-density PE layer diffused into ULDPE layer and crystallized when the sample was cooled.

Other effects were observed in AFM images of a microlayer blend of bisphenol A polycarbonate (PC),

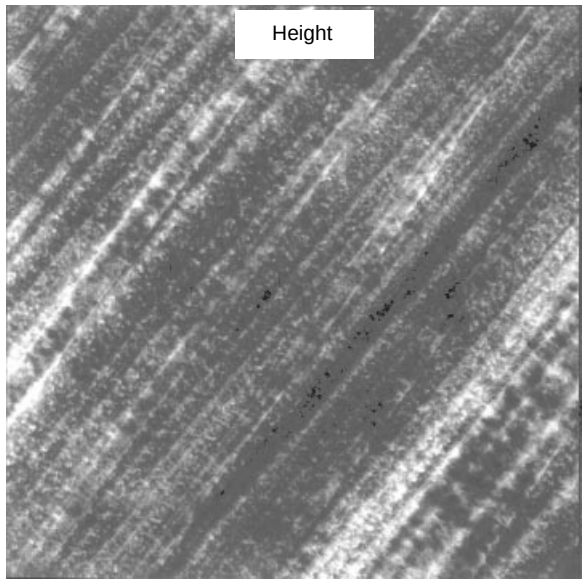

(a)

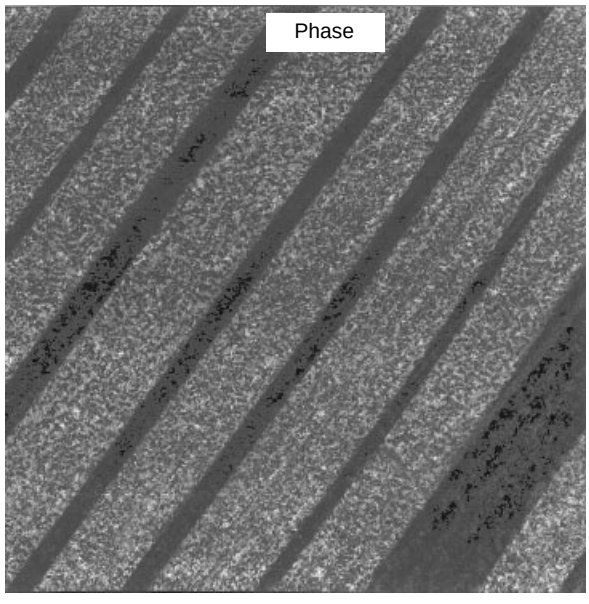

(b)

Figure 44 (a) and (b) AFM height and phase images of a microlayer PE blend, which were obtained simultaneously on 100 μm × 100 μm area. (c) AFM image of the PE blend after 2 h of annealing at 200 °C. Scan size 20 μm × 20 μm.

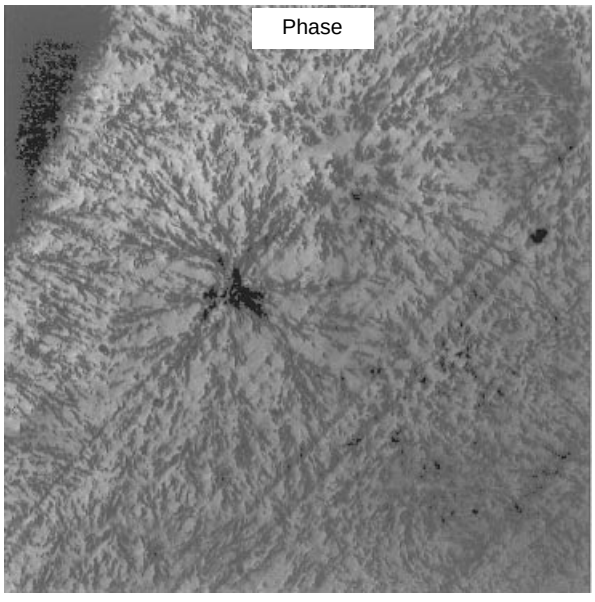

(c)

Figure 44 (*Continued*)

and CPEST (Kodar™).[21] The phase images of this sample as-received after coextrusion did not exhibit any different contrast between the PC and the CPEST layers. This could be explained by the fact that the density of these polymers at RT is almost the same. However, it was possible to assign the layers to different components after thermal annealing at 200 °C. This leads to crystallization of the CPEST whereas the PC remains in an amorphous state. The height and phase images of the annealed microlayer blend are shown in Figure 45(a) and (b). The height image shows alternative layers with height variation. This topography is a consequence of sample microtomy at cryogenic temperatures. Scratches from the microtome knife are seen perpendicular to the polymer layer. Both kinds of feature are also present in unannealed microlayer samples. The main result of sample annealing is spherulitic patterns on elevated layers, which enable their identification as CPEST. The phase contrast between the crystalline and amorphous regions is much stronger than between the PC and CPEST layers (Figure 45b). This is another example showing that phase contrast is related to the density of polymer materials.

Rubber-toughened plastics and filled elastomeric systems are among the multicomponent polymer materials with a broad spectrum of practical applications. Compositional mapping of these materials is an essential part of their characterization, which helps in understanding the structure–property relationships of these materials. Initially, compositional mapping of heterogeneous polymer materials is performed in FM mode[64] and, nowadays, tapping mode with phase imaging is applied. Visualization

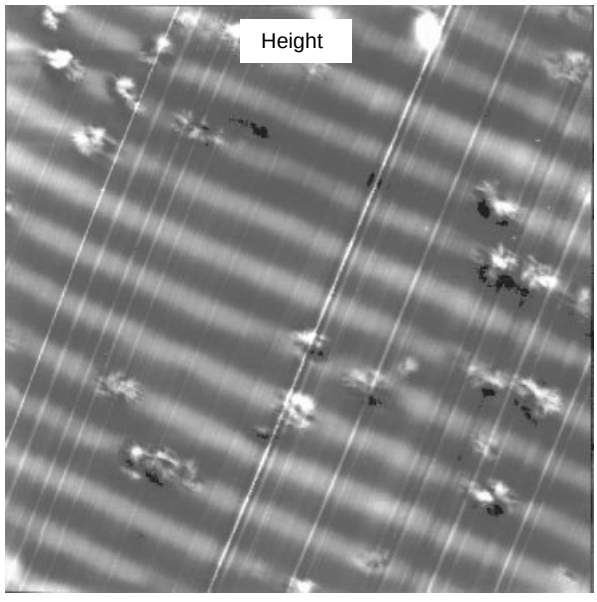

(a)

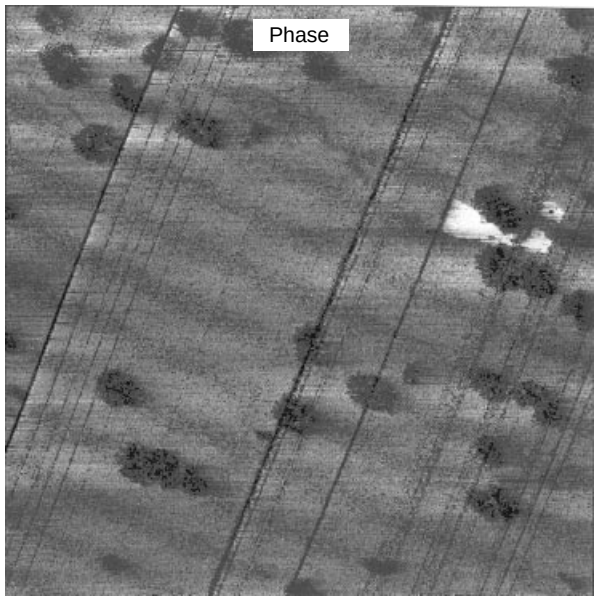

(b)

Figure 45 (a) and (b) AFM height and phase images of a microlayer PC/CPEST blend, which were obtained simultaneously on 50 μm × 50 μm area.

of the component distribution of a rubber-toughened blend of PC and poly(butylene terephthalate) (Xenoy™) is best achieved in images obtained with *hard tapping*, Figure 46(a) and (b). The contrast in the height image, which is mostly related to surface topography, is less informative than the phase image. The latter shows a distribution of two plastic components and rubber particles. The phase image, which was obtained with *light tapping* (not shown here), exhibits almost featureless contrast.

For many samples, the assignment of phase contrast to particular sample components is not trivial, it is rational to

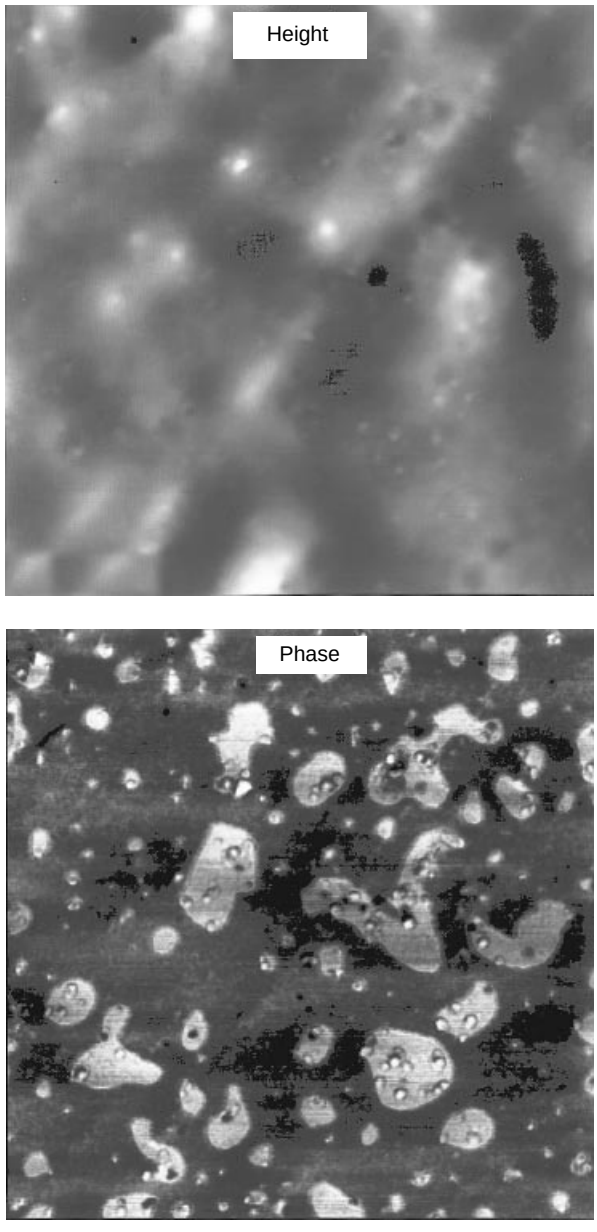

Figure 46 (a) and (b) AFM height and phase images of rubber-toughened plastic, which were obtained simultaneously in *hard tapping* on a flat surface prepared with an ultramicrotome. Scan size 3 μm × 3 μm.

use a correlation installed from studies of model systems, which might have the same components in different ratios. Also a definite correlation between the phase contrast and stiffness of components can be installed from the studies of polymer materials of the same class. For example, in the phase image obtained with *hard tapping*, inorganic fillers and plastics components are typically brighter than synthetic and natural rubbers. The latter are brighter than oil and other liquid components. Identification of components can be more complicated if the number of components increases and when the fillers are in the shape of small particles, which might exhibit a local mechanical response governed by their surroundings. AFM images of a blend of natural and synthetic rubbers filled with carbon black particles are shown in Figure 47(a–c). Aggregates of carbon black, which are seen as bright patches in the phase image (Figure 47b), are the consequence of the high affinity of carbon black for synthetic rubber. Besides these aggregates, individual carbon black particles are homogeneously spread through other areas, and they are distinguished in the phase image, Figure 47(c), for the most part because of their shape.

In AFM studies of binary latex blends, a correlation between the phase contrast and glass transition temperatures (T_g) of individual components was found.[36] These binary blends are made from acrylic latexes which are different in size and exhibit different T_g. These blends were prepared from particles of latex A: 128 nm in size and $T_g = 31\,^{\circ}C$, latex B: 124 nm in size and $T_g = 11\,^{\circ}C$, and latex C: 217 nm in size and $T_g = 20\,^{\circ}C$. The height images of the A_1B_3, B_1C_3, and A_1C_2 blends (subscripts indicate a component part in the blend), which were obtained with *light tapping*, show individual particles of the surface layer, Figure 48(a), (c) and (e). The differences in T_g of the blend components are emphasized in the phase images of the same areas, which were recorded with *hard tapping*, Figure 48(b), (d) and (f). In these images, latex particles with higher T_g values always appear brighter. For

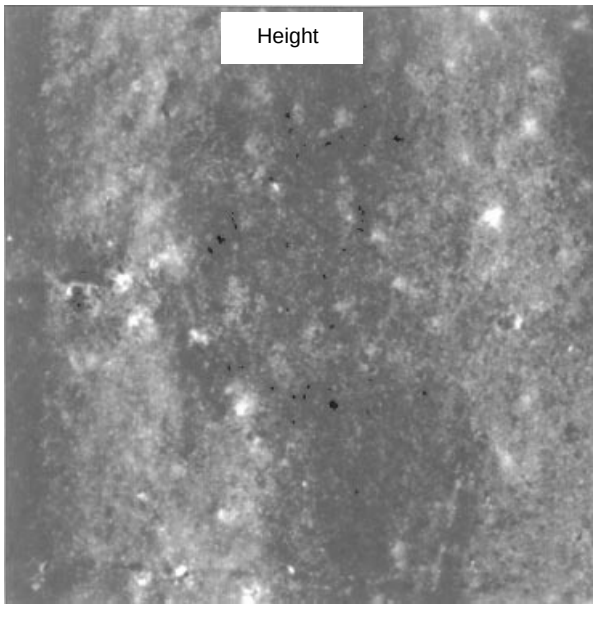

Figure 47 AFM images of a blend of natural and synthetic rubber filled with carbon black. The sample was prepared with a microtome. The images (a) and (b) were obtained simultaneoulsy in *hard tapping* on 35 μm × 35 μm area. Phase image in (c) was obtained in *light tapping*. Scan size 4 μm × 4 μm.

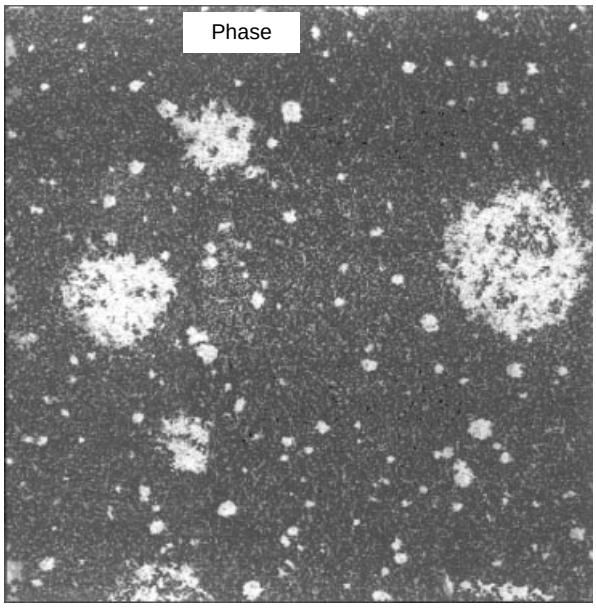

(b)

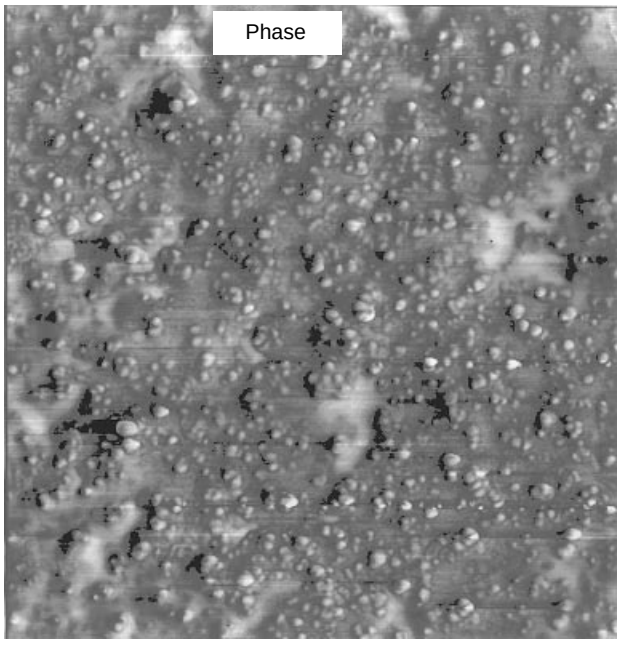

(c)

Figure 47 (*Continued*)

example, in the phase image of the AB blend (Figure 48b), one third of the 100-nm latex particles show up brighter which is consistent with the component ratio in this blend.

5.3 Studies of Microphase Separation in Block Copolymers

Various morphologies with microseparation of components are observed in AFM images of block copolymers. For many block copolymer samples, AFM images show microseparation patterns similar to those obtained with TEM on chemically stained ultrathin sections. With the broadening of such applications, it appears that AFM is invaluable for block copolymers with components for which it is rather difficult to find an appropriate staining agent. Though the preparation of block copolymer films for AFM usually gives a flat surface which can be directly used for imaging, the choice of imaging parameters might be nontrivial. As shown above, differentiation between the components of block copolymers is caused by different tip interactions with their blocks and the same interaction could lead to a faked, non-topography related contrast in the height images obtained in the tapping mode. Although tapping mode is the best alternative for imaging soft samples, such measurements of block copolymers should be performed with special precautions.

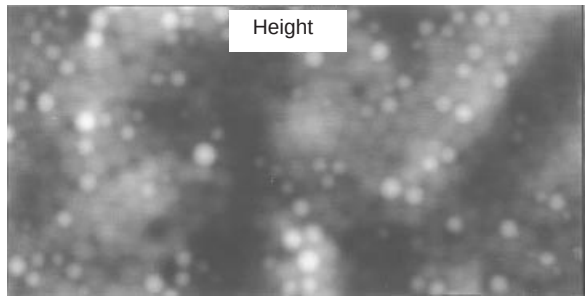

(a)

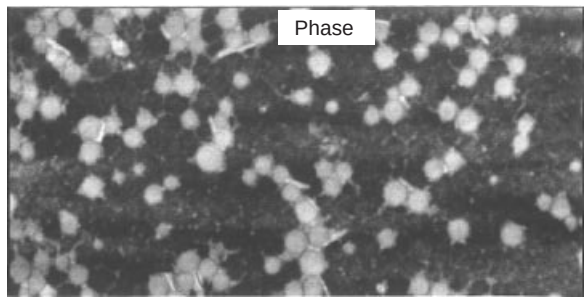

(b)

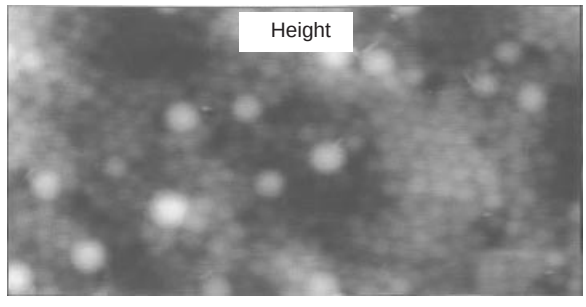

(c)

Figure 48 AFM images of binary latex blends A_1/B_3, B_1/C_3 and A_1/C_2, respectively. The height (a) and phase (b) images were obtained on the same 3 μm × 3 μm area of A_1/B_3 blend in *light tapping* and in *hard tapping*, respectively. The height (c) and phase (d) images were obtained on the same 3 μm × 3 μm area of B_1/C_3 blend in *light tapping* and in *hard tapping*, respectively. The height (e) and (f) images were obtained on the same 3 μm × 3 μm area of A_1/C_2 blend in *light tapping* and in *hard tapping*, respectively.

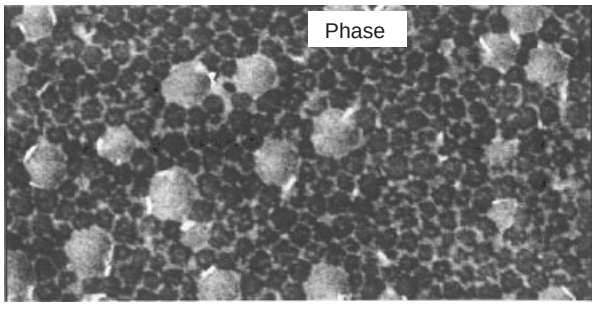

(d)

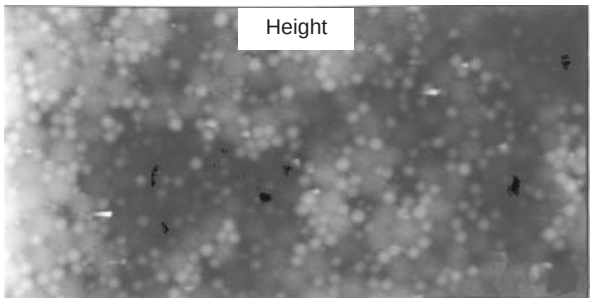

(e)

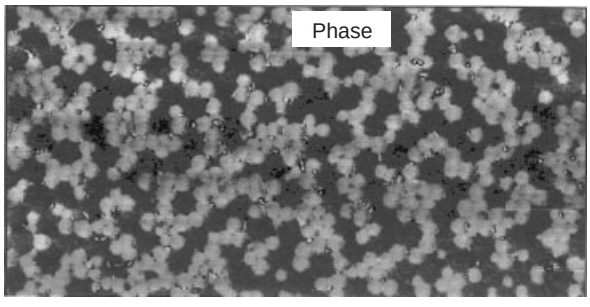

(f)

Figure 48 (*Continued*)

To illustrate this point, AFM imaging of the triblock copolymer systems, PS-*b*-PB-*b*-PS (PB, polybutadiene), will be considered. AFM images of PS-*b*-PB-*b*-PS films prepared by spin casting typically show a microphase separation of the PS and the PB blocks, where PS blocks are seen to be brighter in the height and phase images. These images are similar to TEM micrographs of the same block copolymer, which illustrates bulk film morphology. According to X-ray photoelectron spectroscopy (XPS) data and TEM micrographs, which were taken from the cross-sections cut perpendicular to the film direction, the top surface of PS-*b*-PB-*b*-PS films consists of PB blocks. To make AFM data more representative of the actual case imaging should be performed at different A_{sp}.[65] Then, AFM images obtained with *light tapping* will show the true surface topography of the top PB layer, Figure 49(a). With a decrease in A_{sp} and a corresponding increase in tip–sample force interactions, the probe will gradually penetrate through the top PB layer, which is 15–20-nm thick, and will reveal microphase separation in the subsurface material, Figure 49(b). Bright ''worms'' in this image represent stiff lamellae of PS, which is in a glassy state at RT whereas PB is in a rubber-like state. The image changes are reversible which means fast and complete recovery of the PB top layer after tip penetration. The ability of the tip to penetrate and the layer to recover are due to the rubber-like state of the top layer.

The ability of the AFM probe to penetrate through the top surface layers can be employed for layer-by-layer imaging of more complicated block copolymers. In AFM studies of a fluorinated PS-*b*-PI (PI, polyisoprene) block copolymer, in which fluorosegments are attached to PI blocks, different AFM patterns were observed during

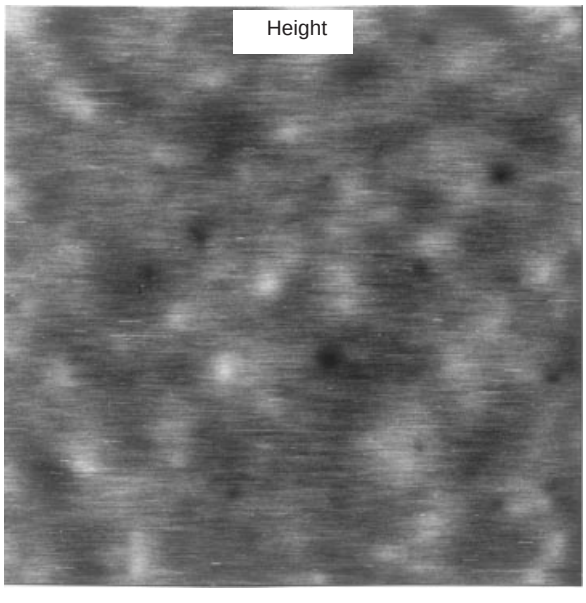

(a)

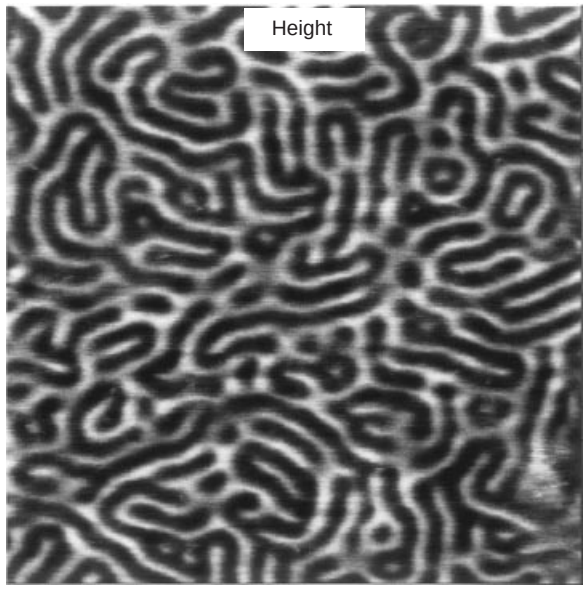

(b)

Figure 49 AFM images of PS-*b*-PB-*b*-PS film in (a) *light* and (b) *hard tapping*. Scan size 1 μm × 1 μm.

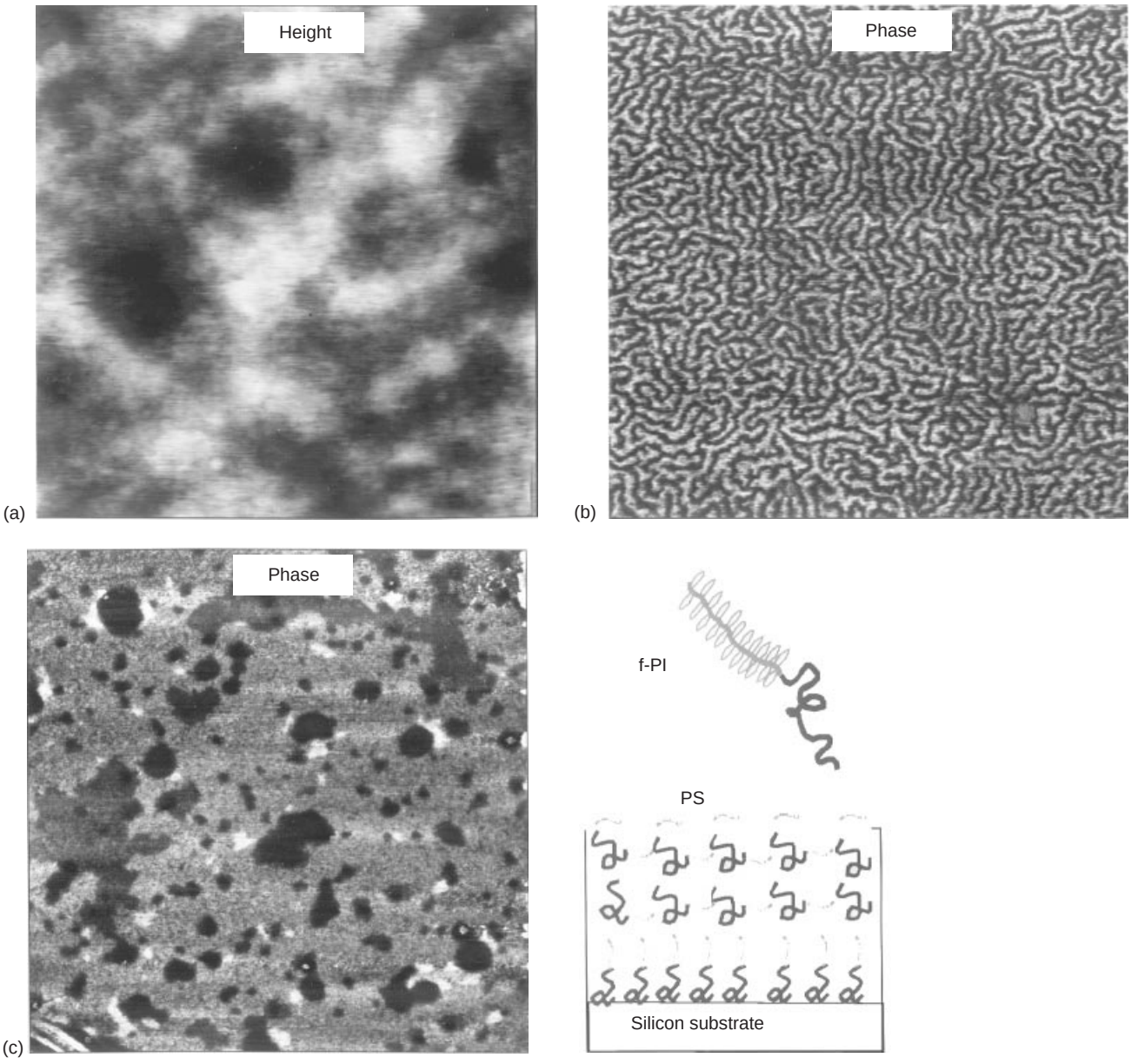

Figure 50 AFM images of fluorinated PI-*b*-PS block copolymer film. The image in (a) was obtained in *light tapping* and the images in (b) and (c) were obtained at higher tip forces. Images in (a) and (b) were obtained on the same 0.6 μm × 0.6 μm area. Image in (c) was obtained on 3 μm × 3 μm area.

imaging at different A_{sp}, Figure 50(a–c).[66] The height image obtained with *light tapping* reveals the topography of the top layer formed by the terminal PI groups, Figure 50(a). The related phase image is featureless and is not shown here. With an increase in the tip force, a more ordered pattern appears, as seen from the phase image in Figure 50(b). The width of individual bright ''worms'' is in the 6–8-nm range. With a further increase in force, the phase image changes again and patches with different contrast are seen on the larger scale, Figure 50(c). These image changes are reversible and they indicate a complex layer-by-layer self-organization in this film. For a tentative explanation of these images, several structures were suggested,[66] however, the final conclusions could be drawn only after combined consideration of AFM images and data from other techniques.[67]

6 EXAMINATION OF POLYMER THERMAL TRANSITIONS

After the first steps in AFM characterization of polymers are done, it is important to extend the measurements

from ambient temperature to temperatures where polymer phase transitions occur. Visualization of structural changes at polymer thermal transitions could be an important addition to a number of macroscopic thermal techniques (DSC (differential scanning calorimetry), TMA (thermal mechanical analysis), DMA, etc.), that are usually employed for studying these transitions. Such studies are in their infancy and they are mostly performed at elevated temperatures.[68–71] Only a few studies were undertaken at subzero (°C) temperatures. There are two approaches towards studies of polymer phase transitions. In one approach, a polymer sample is heated or cooled, and the images at different temperatures are recorded. The images presented below were obtained in such a procedure. In another approach, the AFM probe, which stays in permanent contact with the sample, is heated and the probe deflection is monitored.[72] The probe deflection changes substantially when the probe temperature exceeds glass transition or the melting temperature of the sample. In this way, the polymer material under the probe can be recognized based on its thermal characteristics.

6.1 Imaging Polymers at Elevated Temperatures

In earlier AFM applications to polymers, when measurements at different temperatures were not possible, the effect of thermal annealing on polymer structures was examined at RT. After the initial imaging at RT, the sample was heated to the target temperature for some time and then cooled back to RT. In such procedures, morphology changes of a spin-casted PS-*b*-PB-*b*-PS film were examined. The initial morphology of the film is different from the one shown in Figure 49(b). This morphology and its changes during annealing are seen in the phase images recorded with *hard tapping*, Figure 51(a–d). As-prepared film is characterized by spherical particles and with lamellar morphology develops with annealing time, Figure 51(b–d). Long-term annealing of bulk PS-*b*-PB-*b*-PS samples leads to an appearance of ''single crystals'' formed from extended lamellae arranged into a hexagonal lattice.

With the development of heating accessories, these films were examined at various temperatures, Figure 52(a–d). It appears that in the AFM phase images, recorded under *hard tapping*, the microphase-separated patterns gradually disappear as the sample temperature approaches T_g (~100 °C) of the PS blocks. As mentioned before, the difference in mechanical properties of PS and PB blocks is responsible for the AFM contrast observed in this sample at RT. When both components of this block copolymer are in a rubber-like state, the microphase-separation pattern is not detectable with AFM. However, the presence of the microphase separation in PS-*b*-PB-*b*-PS film at these temperatures is known from the structural data obtained by other techniques.

In the imaging of fluorinated PI-*b*-PS block copolymers at temperatures below and above the smectic phase transition (ca. 50 °C)[66] the manner in which the AFM probe could detect this transition was determined, Figure 53. At temperatures below this transition, the AFM images obtained with *light* and *hard tapping* had practically the same hexagonal pattern, Figure 53(a), which indicates a ''rigidity'' of the topmost layer. The hexagonal pattern gradually disappeared with the temperature increase, Figure 53(b). At temperatures

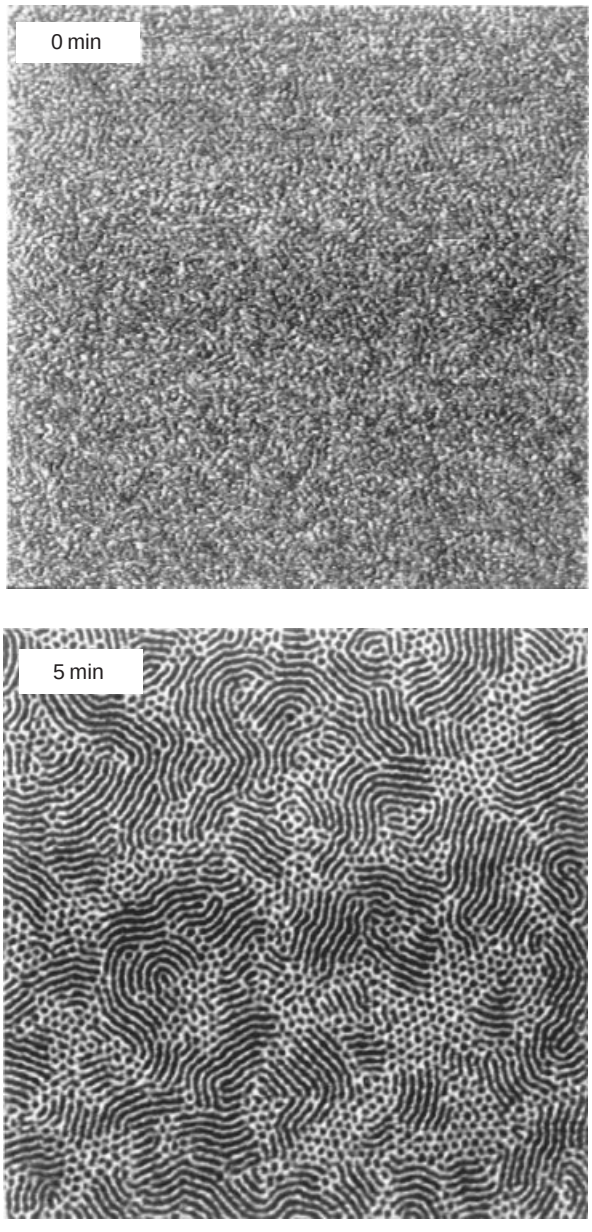

Figure 51 AFM phase images obtained on PS-*b*-PB-PS film directly after spin casing on Si substrate (a), and after annealing at 100 °C after different times (b–d). All images were obtained at RT with *hard tapping*. Scan size 2 μm × 2 μm.

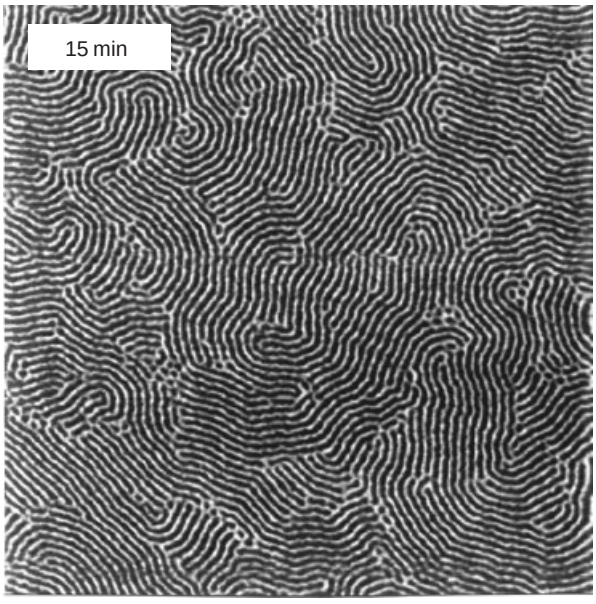

(c)

(d)

Figure 51 (*Continued*)

above 50 °C, the images obtained with *light tapping* were featureless but with an increase in the tip force a well-defined hexagonal pattern appears again, Figure 53(c) and (d). This implies that at the smectic phase transition the mobility of the top layer increases and the AFM tip can penetrate through and reach the more ordered and, therefore, more rigid structure of the underlying layer.

The first AFM studies of polymers at elevated temperatures[68] were aimed at monitoring melting and crystallization. While imaging a polymer sample around its melting point experimental difficulties should be expected related to the increased stickiness and softness of the premelted polymer even in tapping mode. This is indeed the case, and for stable imaging at elevated temperatures larger free amplitudes and lower set-point amplitudes should be used. Unfortunately, these restrictions reduce the flexibility of measurements with *light* and *hard tapping* that exists in imaging at RT.

High-temperature imaging of bulk polymer samples causes more problems than the imaging of thin and ultrathin polymer films on a solid substrate. The presence of the substrate limits of tip penetration into soft materials helps in AFM imaging. An example of such studies is given in Figure 54(a–f), where a sequence of images shows the growth of ε-PCL lamellae at a temperature ($T = 36\,^{\circ}C$) below its melting point.[39] A dendritic structure was growing inside the amorphous polymer, and the change in its dimensions can be used for evaluation of the crystallization rate. In the initial ε-PCL film, the polymer was highly crystalline, Figure 24. The sample crystallized at 36 °C exhibits a semicrystalline morphology. This morphology was stable when the sample was cooled to RT. Full crystallization was achieved only when the sample temperature was below RT, Figure 54(e) and (f). The amorphous material between the dendritic structures crystallized into thin fibrils.

6.2 Visualization of Melting and Crystallization of Polyethylene

AFM studies of melting and crystallization of ultrathin layers of different PE samples have been performed.[73] The crystalline layers of ~20 nm in thickness were formed on a Si substrate after it was dipped into a hot PE solution in xylene and taken out. The morphology of this dip-crystallized PE was characterized by 2D spherulites. The phase image in Figure 55(a) shows a part of such a spherulite formed in the LLDPE film. Almost bare substrate locations are seen in the lower left hand corner of this image. In this spherulite, most of the thin lamellae are edge-on oriented. No structural changes were detected after the sample temperature was raised to 80 °C. A further temperature increase to 125 °C (the melting temperature of this polymer in the bulk state) led to complete disappearance of PE lamellae, Figure 55(b). To detect polymer crystallization in this layer the sample temperature was lowered to 102 °C. During the first scans at this temperature, lamella structures appeared in the lower right hand corner of the imaged area, Figure 55(c). At this temperature, the crystal growth occurs relatively slowly with the formation of semicrystalline morphology, Figure 55(d). Patches of amorphous PE are seen at the edges of the area occupied initially by the spherulite. This morphology was unchanged after cooling the sample to RT.

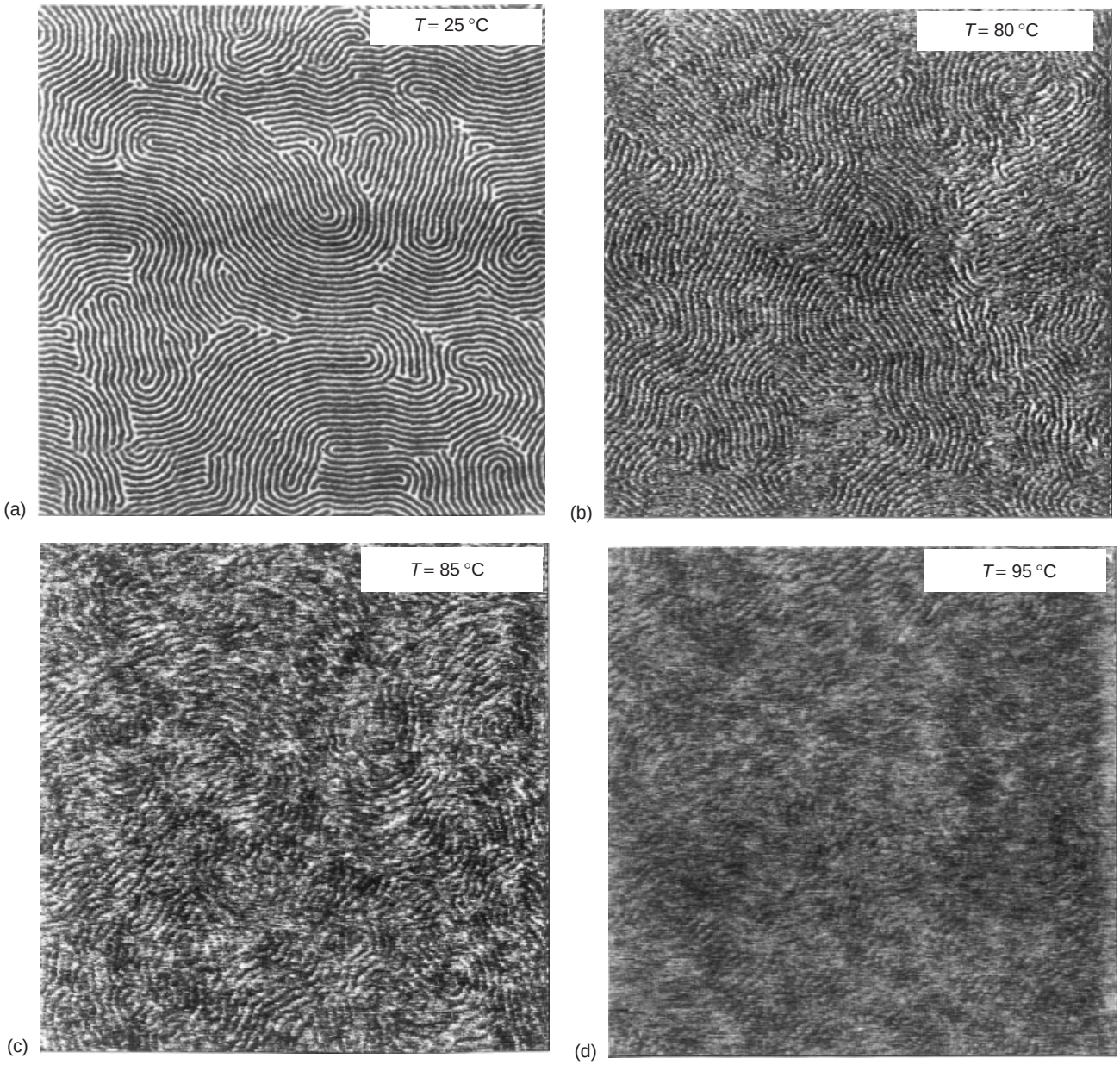

Figure 52 AFM phase images of PS-*b*-PB-*b*-PS film at different temperatures. All images were obtained in *hard tapping*. (a) $T = 25\,^{\circ}C$, (b) $T = 80\,^{\circ}C$, (c) $T = 85\,^{\circ}C$, (d) $T = 95\,^{\circ}C$. Scan size $2\,\mu m \times 2\,\mu m$.

A more systematic AFM study of melting and crystallization was conducted on an ultrathin film of LDPE. During step-like heating of this sample (the sample temperature was increased by $5\,^{\circ}C$ increments followed by imaging) to $90\,^{\circ}C$ or to $100\,^{\circ}C$ (below the LDPE melting temperature of $115\,^{\circ}C$), the 2D spherulites disappeared and large lamella aggregates appeared, Figure 56(a) and (b). This is a consequence of LDPE recrystallization, which leads to the formation of semicrystalline morphology in the ultrathin film. Additionally, numerous holes appeared in this layer which indicates that a dewetting process has accompanied the recrystallization. Estimates of the recrystallization rate from AFM images showed non-linear character for this process, which is in contrast to PE crystallization in bulk. Such a difference is most likely to be caused by a shortage of polymer material in the ultrathin layer. Further heating of the recrystallized sample to the melting temperature of $115\,^{\circ}C$ led to the total disappearance of lamella aggregates, Figure 56(c). After cooling the melted sample to $90\,^{\circ}C$, the semicrystalline morphology with lamellar aggregates appeared again. On further cooling to RT, the

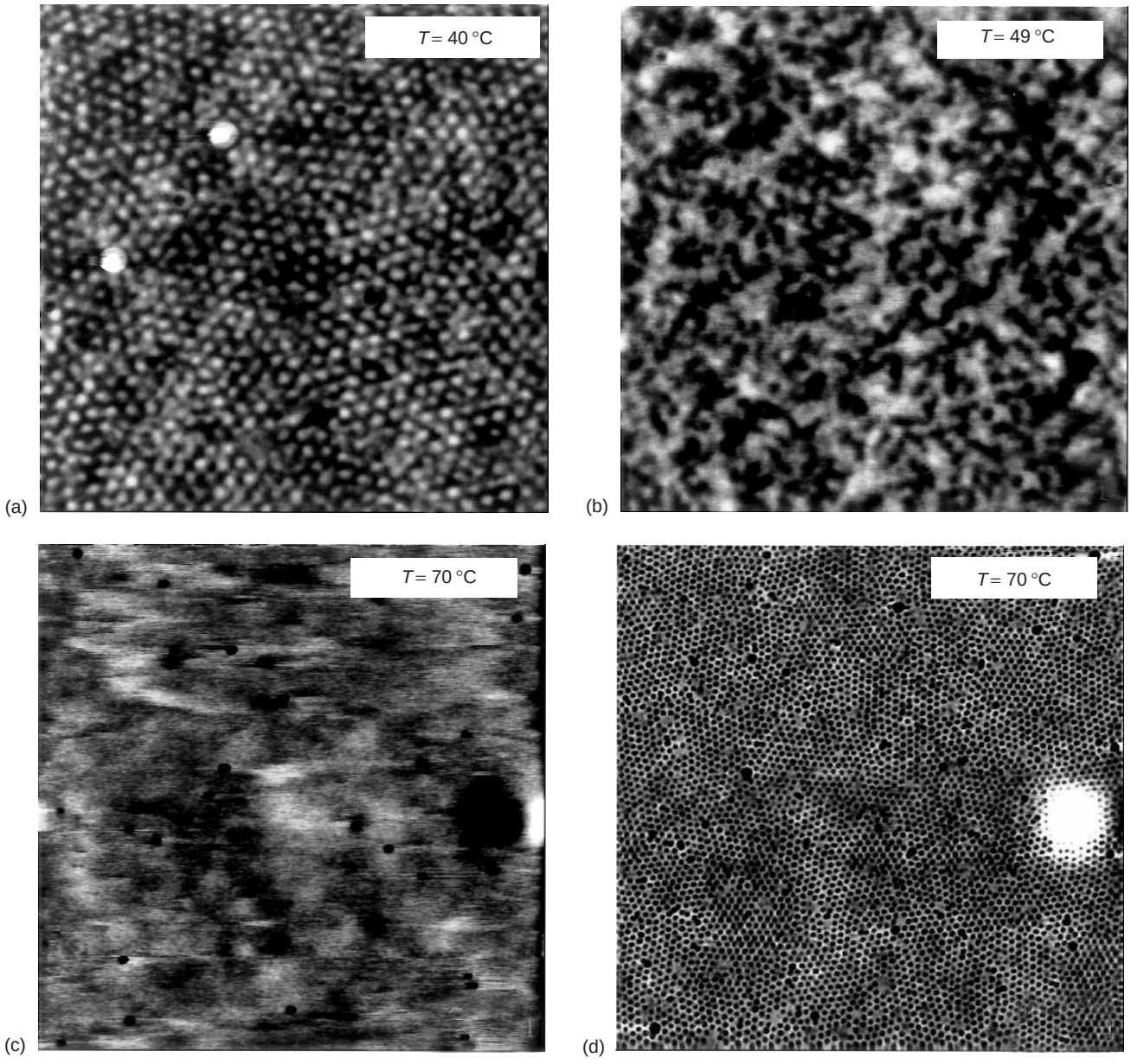

Figure 53 AFM phase images of fluorinated PI-*b*-PS at different temperatures. The images in (a–c) were obtained with *light tapping* and the image in (d) with *hard tapping*. (a) $T = 40\,^{\circ}C$, (b) $T = 49\,^{\circ}C$, (c) and (d) $T = 70\,^{\circ}C$. Scan size $2\,\mu m \times 2\,\mu m$ in (a) and (b) and $5\,\mu m \times 5\,\mu m$ in (c) and (d).

amorphous material surrounding these aggregates crystallized with formation of small lamellae, similar to the low-temperature crystallization of ε-PCL, Figure 54(e) and (f).

Structural changes, which accompany the melting of the ultrathin LDPE layer, substantially depend on the heating rate. This is demonstrated by the images in Figure 57(a) and (b), which show the morphology changes accompanied by rapidly heating the sample from RT to 90 °C. 2D spherulites were converted into sheaf-like patterns, which are embedded in amorphous surroundings. The lamellar structure seen on these sheaf-like patterns at lower temperatures gradually disappeared when the temperature was raised above 115 °C. The sheaf-like morphology remained at higher temperatures and the thickness of the sheaves increased owing to polymer dewetting of the substrate (Figure 57b). When this sample was cooled to RT, its morphology remained the same and crystallization of the sheaves led to lamellar and granular structures. A more detailed description of these changes can be found elsewhere.[73]

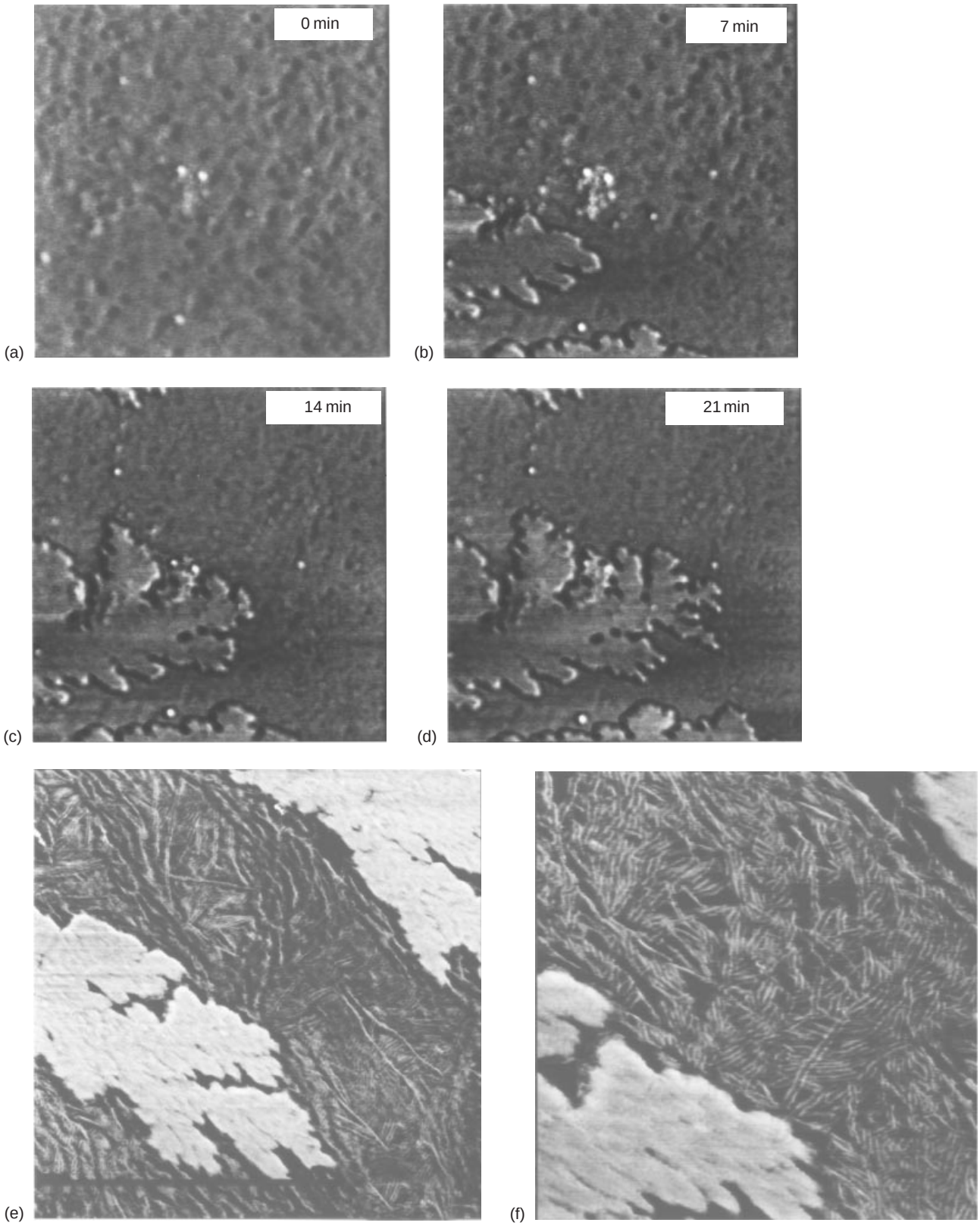

Figure 54 Phase images of ε-PCL layers during crystallization at 40 °C (a–d) and after complete crystallization at −30 °C (e) and (f). Scan size 4 μm × 4 μm in (a–d), 2 μm × 2 μm in (e) and 1 μm × 1 μm in (f).

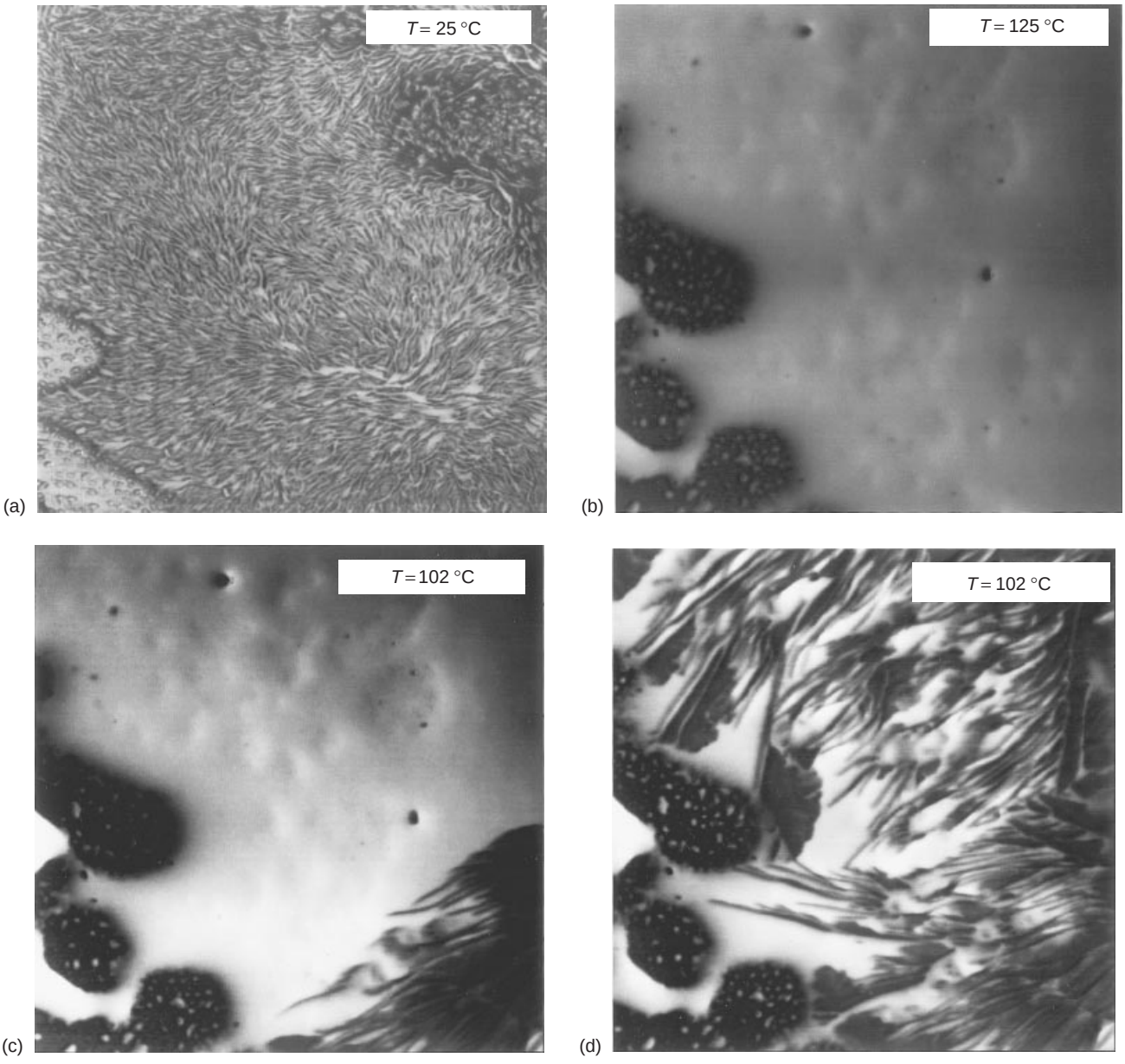

Figure 55 AFM phase images of ultrathin LLDPE film at different temperatures. The image in (d) was obtained 40 min later than the image in (c). (a) $T = 25\,^{\circ}C$, (b) $T = 125\,^{\circ}C$, (c) and (d) $T = 102\,^{\circ}C$. Scan size $12\,\mu m \times 12\,\mu m$.

6.3 Monitoring of Phase Transitions in Mesomorphic Polysiloxanes

The thermal behavior of PDES is described by DSC traces shown in Figure 58. At RT, this polymer is in the mesomorphic state, and it crystallized around $0\,^{\circ}C$. Complete isotropization of this polymer is observed at temperatures above $40\,^{\circ}C$. DSC studies were performed on a bulk sample and it is expected that thermal behavior of PDES patches on Si substrate, which are examined with AFM, may be different.[71] The images in Figure 43(a) and (b) show coexistence of amorphous and lamella structures in PDES patches at RT. When PDES was deposited on the Si substrate, which was treated with argon plasma, a more complete polymer layer was formed. Lamella aggregates, which are embedded in amorphous material and are oriented perpendicular to the rubbing direction, are observed in the phase image of the PDES layer, Figure 59(a). Evident changes in the lamella morphology were observed when the sample was cooled below $0\,^{\circ}C$, Figure 59(b). These changes happened abruptly while the probe was scanning from the bottom of the area. In the higher magnification image (Figure 59c), it can clearly be seen that a linear

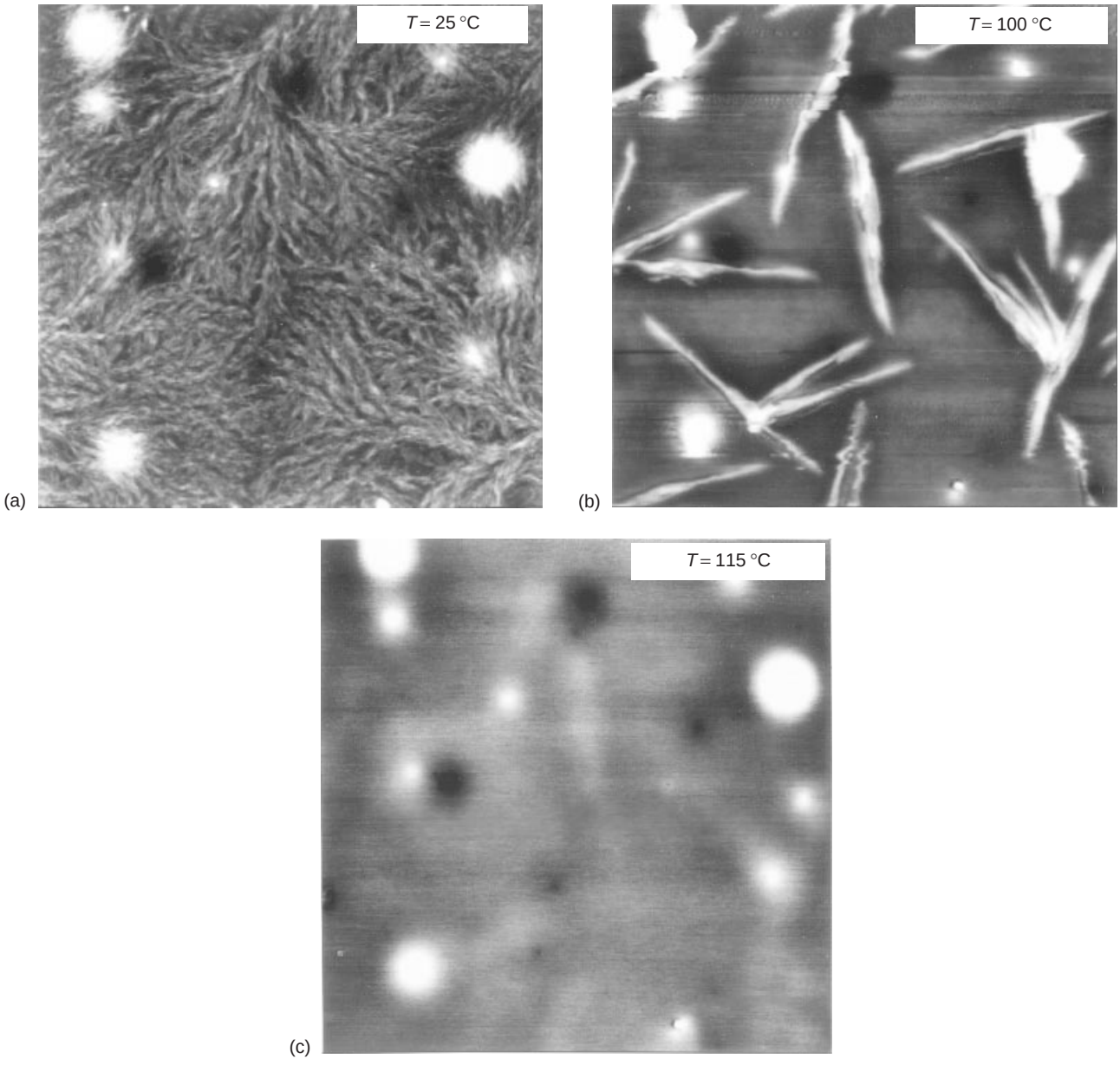

Figure 56 Height images of ultrathin LDPE film at different temperatures. (a) $T = 25\,^{\circ}C$, (b) $T = 100\,^{\circ}C$, (c) $T = 115\,^{\circ}C$. Scan size $10\,\mu m \times 10\,\mu m$.

substructure (with the tens of nanometers wide features) has appeared on the lamellae. The changes in lamella substructure are reversed when the sample is heated above RT, Figure 59(d). Therefore, these changes are associated with the crystallization of PDES. Visualization of these changes during crystallization from the mesomorphic state and back allows better understanding of the structural architecture of PDES. The dimensions of lamella observed in AFM images of a number of mesomorphic polysiloxanes are consistent with the extended chain organization in PDES lamella.[74] However, various nanostructures in the 20–40-nm range, which are seen in AFM images, are not completely understood. For example, the image in Figure 59(e) reveals extended nanostructures on the lamellar aggregates at the temperature of the phase transition ($T = 0\,^{\circ}C$).

The structural changes, which accompany the isotropization of PDES and low-temperature crystallization, are seen in AFM images in Figure 60(a–d). With the increased temperature lamella aggregates of the mesomorphic material gradually melt and finally disappear, Figure 60(a) and (b). The largest thermal effect is related to the crystallization of the mesomorphic material around $0\,^{\circ}C$ (Figure 58). However, this process does not involve

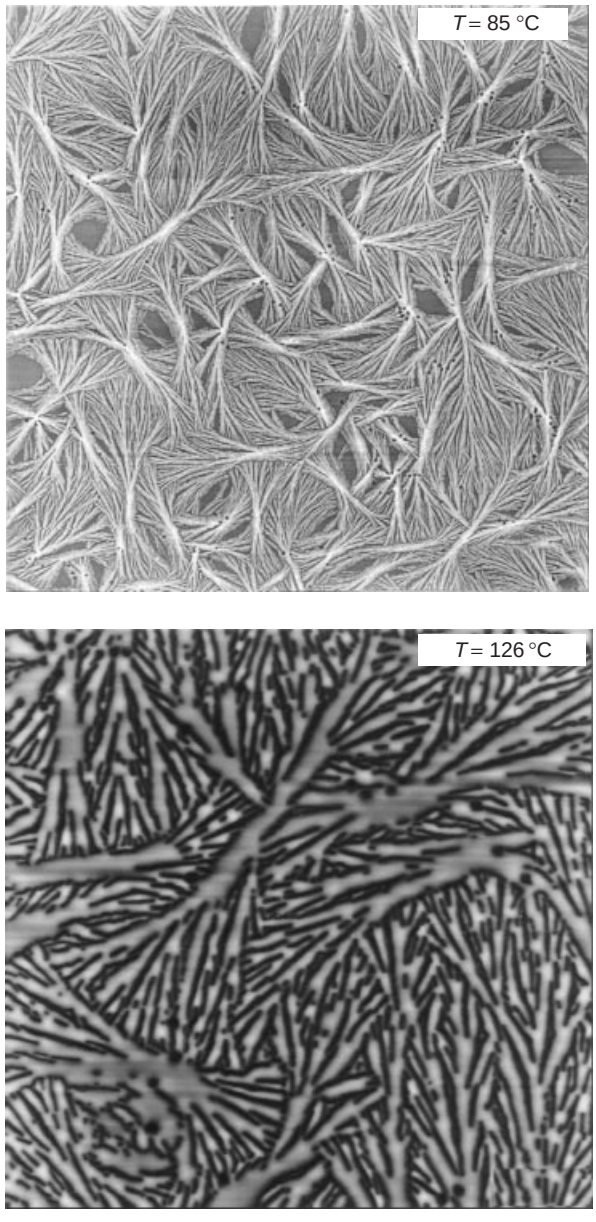

Figure 57 Height images of ultrathin LDPE film at different temperatures. (a) $T = 85\,^{\circ}C$, (b) $T = 126\,^{\circ}C$. Scan size $110\,\mu m \times 110\,\mu m$ in (a) and $40\,\mu m \times 40\,\mu m$ in (b).

amorphous material, which surrounds the lamella. Deep cooling of the PDES sample to $-40\,^{\circ}C$ induced crystallization of the amorphous polymer. The appearance of spherulite-like structures in the bottom right corner of Figure 60(c) and their growth (Figure 60d) can be assigned to the low-temperature crystallization of PDES. Also, new lamellae were grown from the advanced edges of large lamellae.

AFM studies of poly(dipropylsiloxane) (PDPS) at different temperatures reveal structural changes accompanying the reversible crystal–mesophase transition. PDPS differs from PDES in having slightly longer side chains and this change in chemical structure leads to the shift of the thermal phase transition to higher temperatures. At RT this polymer exists in the crystalline state, and the phase image in Figure 61(a) shows crystalline blocks with distinctive edges both of submicrometer size. Upon heating these features disappeared gradually, and in the mesomorphic state all

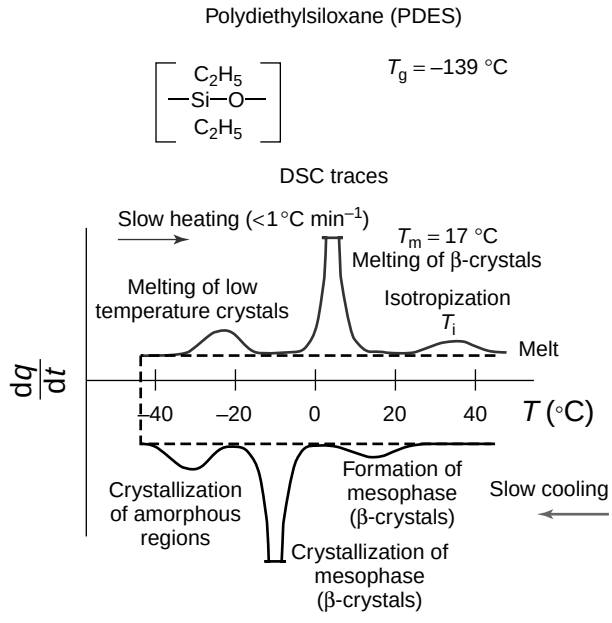

Figure 58 DSC traces of PDES.

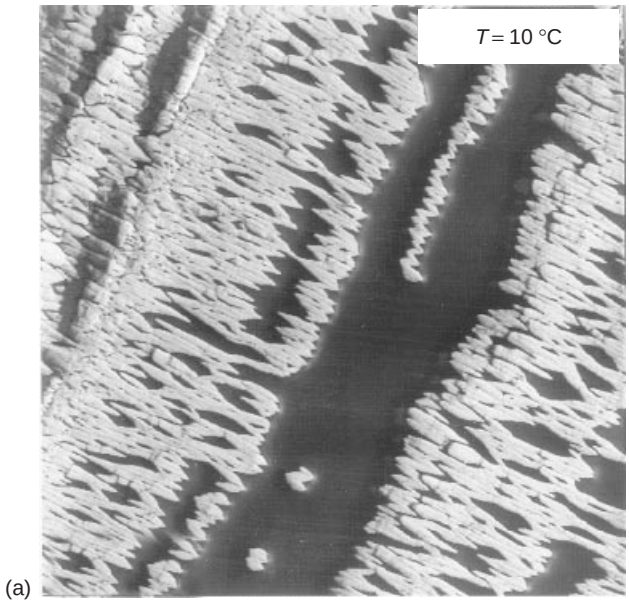

Figure 59 AFM phase images of PDES at different temperatures. Scan size $15\,\mu m \times 15\,\mu m$ in (a) and (b), $6\,\mu m \times 6\,\mu m$ in (c) and (d) and $2.4\,\mu m \times 2.4\,\mu m$ in (e).

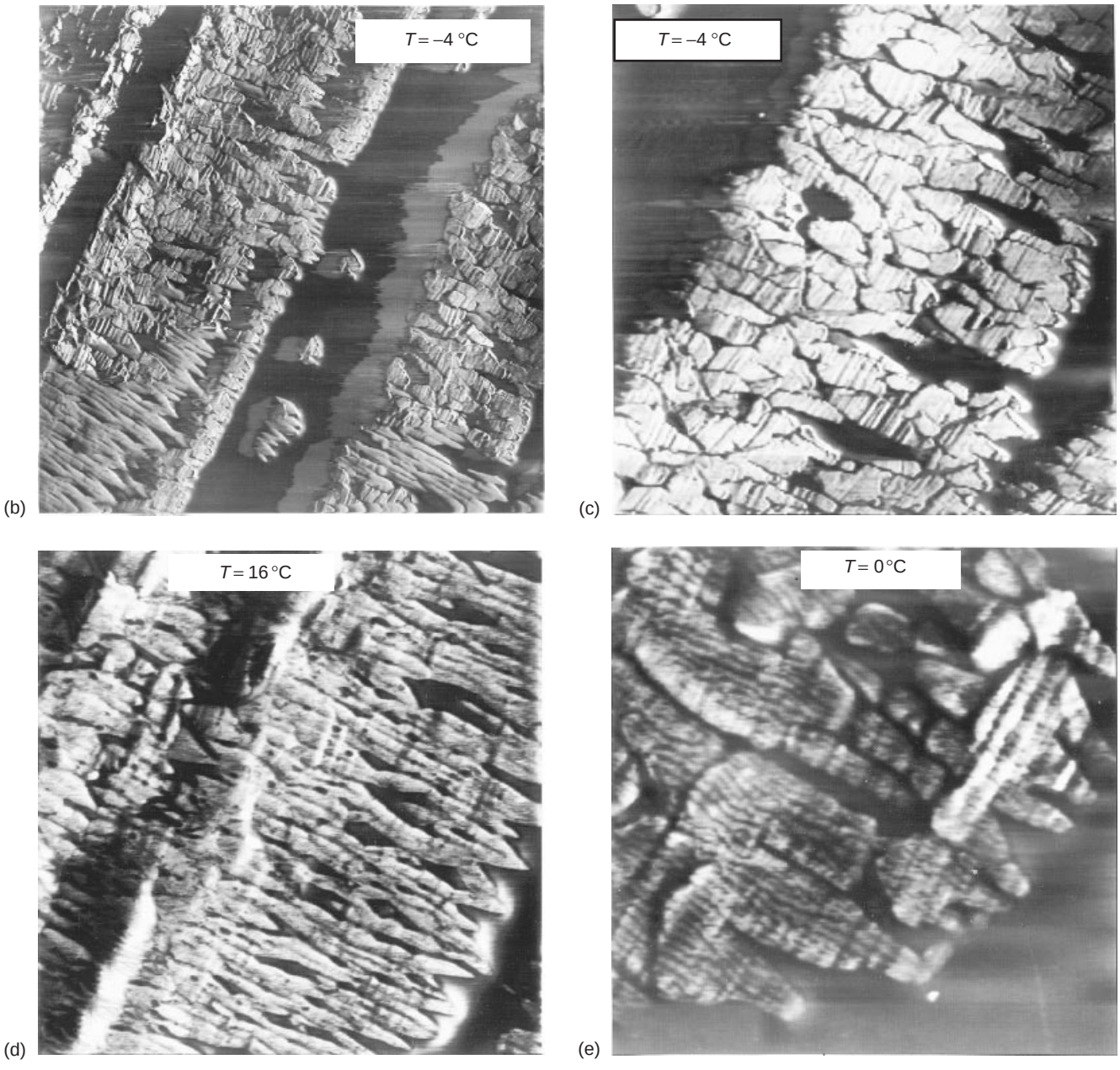

Figure 59 (*Continued*)

of the surface was manifested as large lamellar bundles, which are elongated along the rubbing Y-direction, Figure 61(b) and (c). On cooling the sample back to $T = 40\,^{\circ}\mathrm{C}$, PDPS crystallization is seen as an abrupt transformation of the lamellar bundles to the crystalline blocks.

Visualization of polymer morphology and nanostructure changes at thermal phase transitions, which was demonstrated with the images ε-PCL, PE and mesomorphic polysiloxanes, opens up new capabilities for AFM in the characterization of polymers. The most exciting findings are anticipated in the examination of polymer structural changes on the submicrometer scale. Observations of morphology and nanostructure changes during recrystallization of LDPE substantially broadens our knowledge of these processes.

7 EXPECTATIONS

Applications of AFM to polymer materials have been developing progressively. Many aspects of AFM imaging

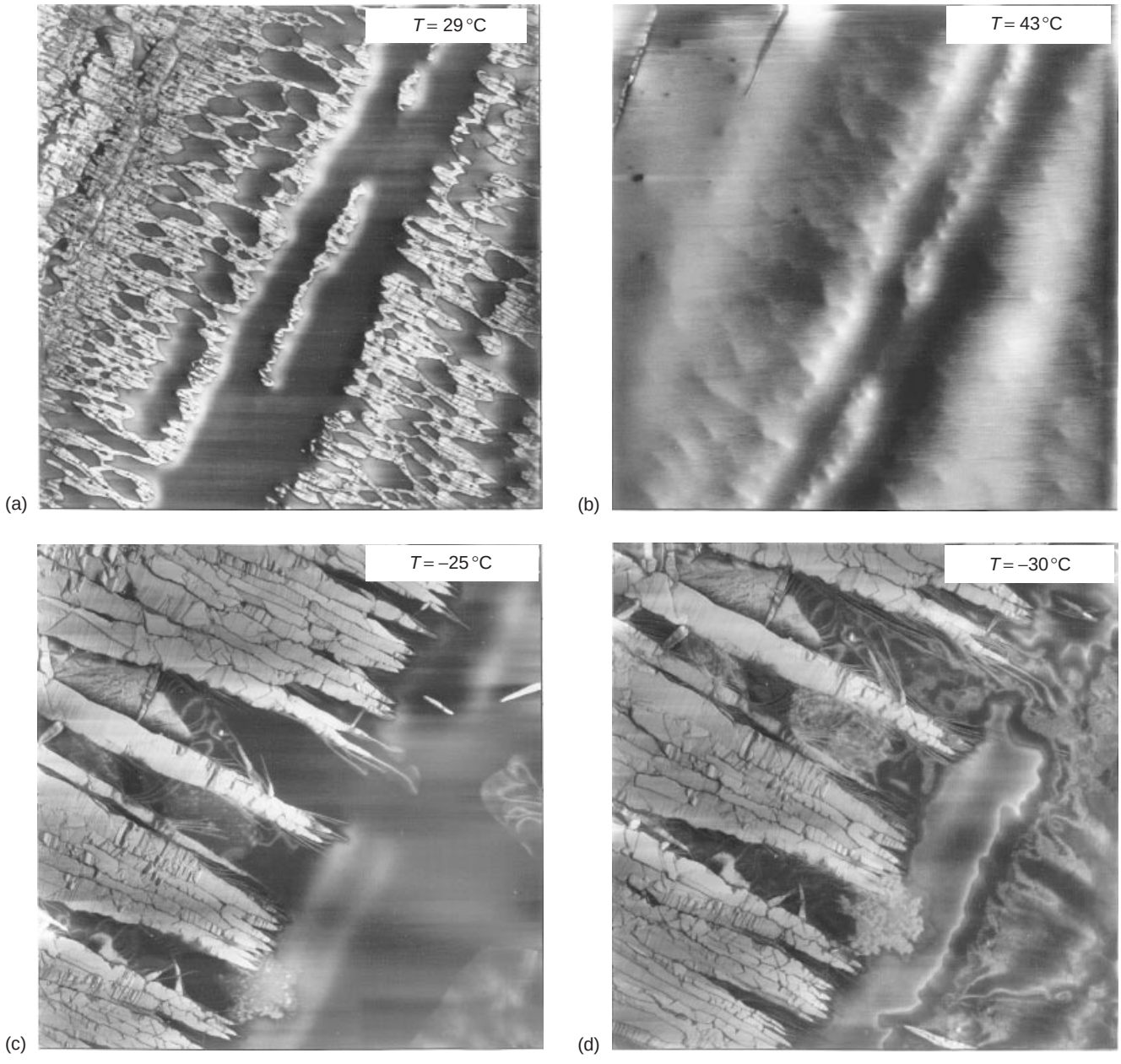

Figure 60 AFM phase images of PDES at different temperatures. Scan size 15 μm × 15 μm in (a) and (b), and 12.5 μm × 12.5 μm in (c) and (d).

of polymer samples are already understood, and the results obtained undoubtedly show the value of this technique. In its basic function, AFM provides high-resolution profiling of a surface allowing visualization of the polymer nanostructure and is an invaluable addition to other characterization techniques. However, its capabilities extend far beyond high-resolution imaging. In addition, AFM provides a means of probing of local sample properties and compositional mapping of heterogeneous polymer systems. For some samples, imaging of subsurface structures can be achieved. To realize these capabilities practically, a more thorough experimental approach should be developed, which includes all aspects of AFM analysis: sample preparation, optimization of imaging parameters and image interpretation. There is a lot of room for improvement in these features and they will be realized because of the strong interest in AFM applications of polymers both in academia and in industry. This interest will grow further because AFM complements a large number of other characterization methods. Also for microscopic studies of some polymer systems, e.g. thin

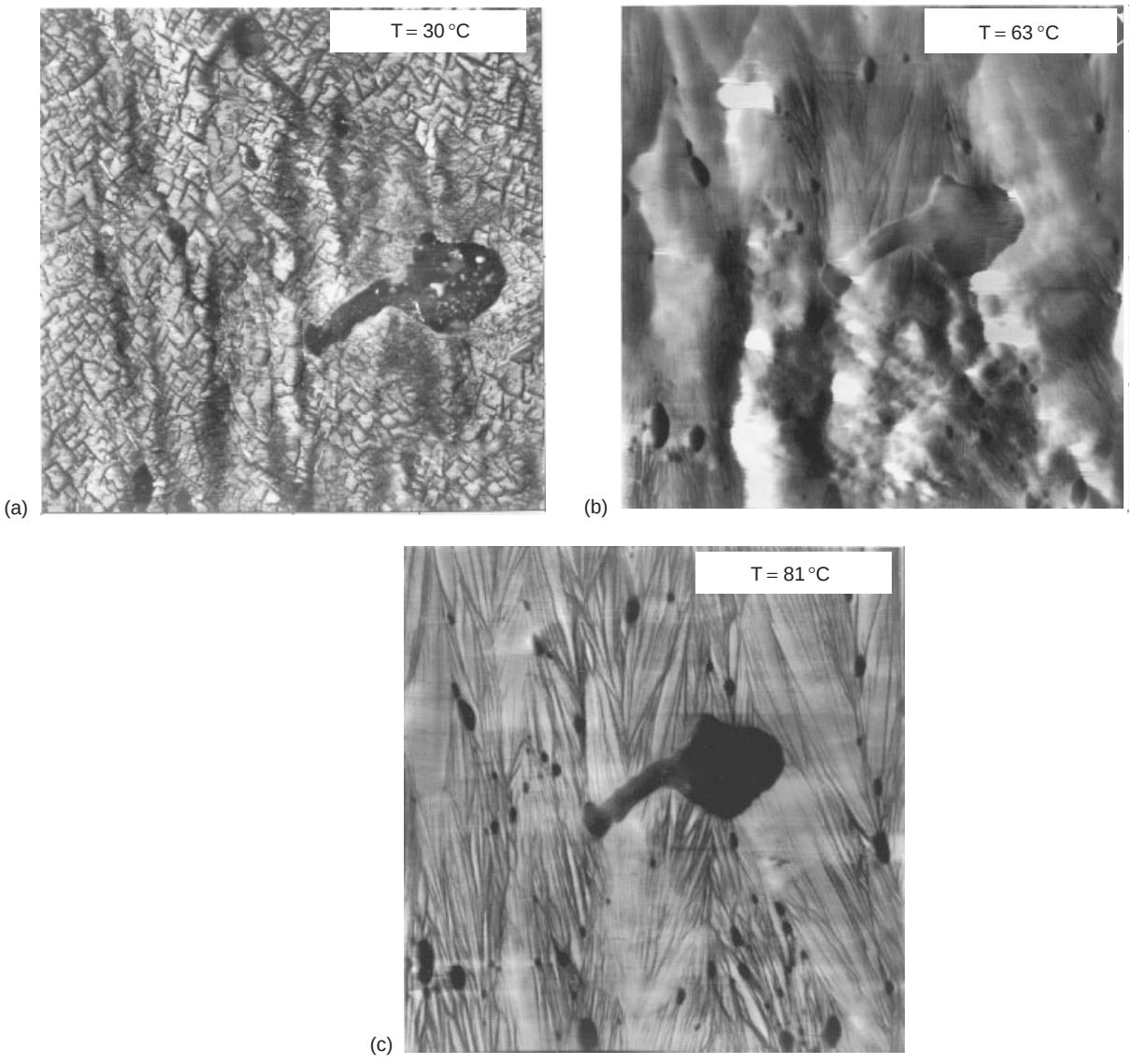

Figure 61 AFM phase images of PDPS at different temperatures. Scan size 10 μm × 10 μm in all images.

and ultrathin films, AFM is in practice the only technique applicable.

Other expectations are related to instrumental developments that broaden AFM studies of polymers. Only the first steps have been taken in AFM studies of polymers at different temperatures. The same is true of monitoring mechanically induced changes in polymer samples. An advanced picture of nanostructure organization in polymer materials, which is emerging from the combination of AFM results with data from other microscopic and diffraction techniques, will motivate researchers to explore thoroughly nanostructural changes during thermal transitions and mechanical loads.

A better understanding of the topography and structure of polymer samples that could be obtained from images recorded at different levels of tip–sample force interactions, is essential for a successful combination of surface profiling with the detection of thermal, electric, magnetic and other properties. It is challenging to map variations of these properties in polymer materials. On one hand, it is interesting to learn about these properties on the micrometer and submicrometer scales. On the other hand, to use them for recognition of individual components in heterogeneous polymer systems is also important. So far in AFM compositional mapping, variations in local mechanical or adhesive properties are employed, however, in some cases the assignment of the image contrast relating to these properties is not straightforward. This situation will improve when variations in thermal or electric properties are also recorded. It will be even more exciting to record the infrared spectrum from different locations on the submicrometer scale and in this

way to obtain ''fingerprints'' of individual polymers or their blocks. Experimental results have shown that such possibilities may be realized practically.[75]

To summarize, in the coming years researchers will explore further the advantages of new developments in AFM itself as well as in different combinations of this method with other characterization techniques. This will promote an advanced characterization of the structure and properties of polymer materials on the micrometer and submicrometer scales.

ACKNOWLEDGMENTS

Most of the AFM results included in this review were obtained at Digital Instruments in cooperation with a number of research groups worldwide.

Many thanks are due to the researchers who kindly provided polymer samples for AFM studies: Profs. E. Baer, A. Hiltner, S. Nazarenko (CWRU, Cleveland, OH), Prof V. Papkov (INEOS, Moscow, Russia), Prof V. P. Shibaev (MSU, Moscow, Russia), Dr R. Callaghan (Celgard™, Charlotte, NC), Dr B. Antrim (Rohm and Haas, Spring House, PA), Prof C. Frank (Stanford University, Stanford, CA), Prof T. Hashimoto (Kyoto University, Kyoto, Japan), Dr Soo Yeol Lee (Seoul National University, Seoul, Korea), Dr G. Capaccio (BP Chemicals, Grangemouth, Scotland), Dr A. Ghanem (Solvay, Brussels, Belgium), Prof J. McClusky (University of Texas, San Antonio, TX), Dr P. Sadhukan (Bridgestone/Firestone, Akron, OH), Dr Ch. Harrison (Princeton University, Princeton, NJ), Dr E. Sivaniah and Prof E. Kramer (UCSB, Santa Barbara, CA).

Dr S. Sheiko (Ulm University, Ulm, Germany) and Dr J. Kumaki (Toray, Japan) are gratefully acknowledged for providing original AFM images from their publications.

I greatly enjoyed my joint work with Prof Yuli Godovsky (Karpov's Institute of Physical Chemistry, Moscow, Russia) in AFM studies of a number of polymers at different temperatures, which is partly described in this paper.

My colleagues at Digital Instruments (Santa Barbara, CA) are acknowledged for sharing their everyday expertise in AFM.

Special thanks are due to Samantha Strausser (Digital Instruments) for her kind help in the preparation of this manuscript.

ABBREVIATIONS AND ACRONYMS

AFM	Atomic Force Microscopy
CPEST	Copolyester
DMA	Dynamic Mechanical Analysis
DSC	Differential Scanning Calorimetry
EFM	Electric Force Microscopy
FFT	Fast Fourier Transform
FM	Force Modulation
HDPE	High-density Polyethylene
iPP	Isotactic Polypropylene
LDPE	Low-density Polyethylene
LLDPE	Linear Low-density Polyethylene
MFM	Magnetic Force Microscopy
NMR	Nuclear Magnetic Resonance
PB	Polybutadiene
PC	Polycarbonate
PDES	Poly(diethylsiloxane)
PDPS	Poly(dipropylsiloxane)
PE	Polyethylene
PI	Polyisoprene
PMMA	Poly(methyl methacrylate)
PS	Polystyrene
PTFE	Polytetrafluoroethylene
PVDF	Poly(vinylidene fluoride)
RT	Room Temperature
SAXS	Small-angle X-ray Scattering
SCM	Scanning Capacitance Microscopy
SNOM	Scanning Near-field Optical Microscopy
SPM	Scanning Probe Microscopy
SThM	Scanning Thermal Microscopy
STM	Scanning Tunneling Microscopy
TEM	Transmission Electron Microscopy
TMA	Thermal Mechanical Analysis
UHMWPE	Ultrahigh-molecular Weight Polyethylene
UHV	Ultrahigh Vacuum
ULDPE	Ultralow-density Polyethylene
XPS	X-ray Photoelectron Spectroscopy
2D	Two-dimensional
3D	Three-dimensional
ε-PCL	Poly(ε-caprolactone)

REFERENCES

1. G. Binnig, C. Quate, *Ch. Gerber, Phys. Rev. Lett.*, **56**, 930 (1986).
2. T.R. Albrecht, M.M. Dovek, C.A. Lang, P. Gruetter, C.F. Quate, S.N.J. Kuan, C.W. Frank, R.F.W. Pease, *J. Appl. Phys.*, **64**, 117 (1988).
3. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, *Phys. Rev. Lett.*, **49**, 57 (1982).
4. J. Tersoff, D.R. Hamman, *Phys. Rev. B*, **31**, 805 (1985).
5. S.N. Magonov, M.-H. Whangbo, 'Surface Analysis with STM and AFM', VCH, Weinheim, 1996.
6. M.-H. Whangbo, S.N. Magonov, H. Bengel, *Probe Microsc.*, **1**, 23 (1997).

7. S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Elings, P.K. Hansma, M. Longmire, J. Gurley, *J. Appl. Phys.*, **164**, 65 (1989).
8. A.L. Weisenhorn, P. Maivald, H.-J. Butt, P.K. Hansma, *Phys. Rev. B*, **45**, 11 226 (1992).
9. Q. Zhong, D. Innis, K. Kjoller, V. Elings, *Surf. Sci. Lett.*, **290**, L688 (1993).
10. K. Babcock, *Data Storage*, September, PennWell Corporation, 43–49, 1994.
11. T. Ohmi, M. Miyashita, M. Itano, T. Imaoka, I. Kawanabe, *IEEE Trans. Electron Devices*, **39**, 537 (1992).
12. K. Babcock, P. Hopkins, *Data Tech.*, ICG Publishing, 155, 1998.
13. A. Noy, D. Vezenov, C. Lieber, *Annu. Rev. Mater. Sci.*, **27**, 381 (1997).
14. S.N. Magonov, K. Qvarnström, V. Elings, H.-J. Cantow, *Polym. Bull.*, **25**, 689 (1991).
15. S.N. Magonov, S. Kempf, M. Kimmig, H.-J. Cantow, *Polym. Bull.*, **26**, 715 (1991).
16. D. Snetivy, G.J. Vansco, *Polymer*, **35**, 461 (1992).
17. K.D. Jandt, T.J. McMaster, M.J. Miles, J. Petermann, *Macromolecules*, **26**, 6552 (1993).
18. R. Patil, D.H. Reneker, *Polymer*, **35**, 1909 (1994).
19. W. Stocker, S.N. Magonov, H.-J. Cantow, J.-C. Wittmann, B. Lotz, *Macromolecules*, **26**, 5915 (1993).
20. H.-Y. Nie, M. Motomatsu, W. Mizutani, H. Tokumoto, *Polymer*, **36**, 183 (1996).
21. S.N. Magonov, Y.K. Godovsky, *Am. Lab.*, **31**, 52 (1999).
22. J. Sugawara, M. Ohta, H. Ueyama, S. Morita, *Science*, **270**, 1646 (1995).
23. J.P. Aime, S. Gauthier, in *Scanning Probe Microscopy of Polymers*, ACS Symposium Series 694, eds. B.D. Ratner, V.V. Tsukruk, ACS, Washington, 1998.
24. E.-L. Florin, V. Moy, H.E. Gaub, *Science*, **264**, 415 (1994).
25. J.P. Aime, C. Elkaakoor, C. Odin, T. Bouhacina, D. Michel, J. Curely, J. Dautant, *J. Appl. Phys.*, **76**, 754 (1994).
26. E.A.- Hassan, W.F. Heinz, M.D. Antonik, N.P.D. Costa, S. Nagaswaran, C.A. Schoenenberger, J.H. Hoh, *Biophys. J.*, **74**, 3 (1998).
27. A. Rosa, S. Hild, O. Marti, *Meas. Sci. Technol.*, **8**, 1 (1997).
28. Y. Martin, C.C. Williams, H.K. Wickramasinghe, *J. Appl. Phys.*, **61**, 4723 (1987).
29. F.M. Serry, P. Neuzil, R. Vilasuso, G.L. Maclay, 'Proceedings Electrochemical Society Symposium', in *Microstructures and Microfabricated Systems II*, eds. P.J. Hesketh, H.G. Hughes, D.D. Denton, D.L. Kendall, Chicago, IL, **95-27**, 83, 1995.
30. S.N. Magonov, M.G. Heaton, *Am. Lab.*, **30**, 9 (1998).
31. J.P. Cleveland, B. Anczykowski, A.E. Schmid, V.B. Elings, *Appl. Phys. Lett.*, **72**, 2613 (1998).
32. L. Wang, *Appl. Phys. Lett.*, **73**, 3781 (1998).
33. G. Bar, R. Brandsch, M.-H. Whangbo, *Surf. Sci. Lett.*, **411**, L802 (1998).
34. Y. Strausser, M. Schroth, J.J. Sweeney, *J. Vac. Sci. Technol., A*, **15**, 1007 (1997).
35. M.C. Goh, D. Juhu'e, O.M. Leung, Y. Wang, M.A. Winnik, *Langmuir*, **9**, 1319 (1993).
36. S.N. Magonov, B. Antrim, *Microscopy of Industrial Polymers*, ed. S.K. Bhattacharya, in press.
37. S. Ponomarenko, N. Boiko, V. Shibaev, S. Magonov, *Abstracts of the 1st International Conference on Scanning Probe Microscopy of Polymers*, eds. E. Kramer, S. Magonov, Santa Barbara, 23, 1999.
38. R.T. Chen, S.T. Saw, M.G. Jamieson, T.R. Alerca, R.W. Callaghan, *J. Appl. Polym. Sci.*, **51**, 471 (1994).
39. S. Gullerud, PhD Thesis, Stanford University, 1999.
40. I. Touzov, K. Craemer, B. Pfannemueller, W. Kreutz, S.N. Magonov, *Adv. Mater.*, **7**, 656 (1995).
41. J. Kumaki, Y. Nishikawa, T. Hashimoto, *J. Am. Chem. Soc.*, **118**, 3321 (1996).
42. T.L. Morkved, W.A. Lopes, J. Hahm, S.J. Sibener, H.M. Jaeger, *Polymer*, **39**, 3871 (1998).
43. M.R. Van Landingham, S.H. McKnight, G.R. Palmese, J.R. Elings, X. Huang, T.A. Bogetti, R.F. Eduljee, J.W. Gillespie, *J. Adhesion*, **64**, 31 (1997).
44. S.M. Setz, R.S. Duran, D.F. Fagerburg, L.T. Germinario, *Polym. Preprints*, **39**, 1151 (1998).
45. W.F. Heinz, J.H. Hoh, *Trends Biotechnol.*, **17**, 143 (1999).
46. W.H. Han, S.M. Lindsay, M. Dlakic, R.E. Harrington, *Nature*, **386**, 563 (1997).
47. W.A. Rees, C. Bustamante, J.P. Vesenka, G. Yang, J.-Y., *Science*, **260**, 1646 (1993).
48. S.A. Prokhorova, S.S. Sheiko, C.-H. Ahn, V. Percec, M. Moeller, *Macromolecules*, **32**, 2653 (1999).
49. M. Gerle, K. Fischer, S. Roos, A.H.E. Mueller, M. Schmidt, S.S. Sheiko, S.A. Prokhorova, M. Moeller, *Macromolecules*, **32**, 2629 (1999).
50. A. Keller, *J. Polym. Sci.*, **39**, 151 (1959).
51. S.N. Magonov, D. Reneker, *Annu. Rev. Mat. Sci.*, **27**, 175 (1997).
52. T. Hugel, G. Strobl, R. Thomann, *Acta Polym.*, **50**, 214 (1999).
53. S.N. Magonov, unpublished data.
54. G. Kanig, *Colloid. Polym. Sci.*, **269**, 1118 (1991).
55. G. Strobl, 'The Physics of Polymers', Springer, Heidelberg, 188–189, 1997.
56. A. Wawkuschewski, H.-J. Cantow, S.N. Magonov, J.D. Hewes, M.A. Kocur, *Polym. Acta*, **46**, 168 (1995).
57. S.N. Magonov, S.S. Sheiko, R.A.C. Deblieck, M. Moeller, *Macromolecules*, **26**, 1380 (1993).
58. A. Wawkuschewski, H.-J. Cantow, S.N. Magonov, S.S. Sheiko, M. Moeller, *Polym. Bull.*, **31**, 693 (1993).
59. M. Kojima, J.H. Magill, J.S. Lin, S.N. Magonov, *Chem. Mater.*, **9**, 1145 (1997).
60. R.S. McLean, B.B. Sauer, *J. Polym. Sci.*, **37**, 859 (1999).
61. R.S. McLean, B.B. Sauer, *Macromolecules*, **30**, 8314 (1997).
62. Yu.K. Godovsky, V.S. Papkov, *Adv. Polym. Sci.*, **129**, 129 (1989).
63. N.N. Makarova, Yu.K. Godovsky, *Progr. Polym. Sci.*, **22**, 1001 (1997).

64. A.A. Galuska, R.R. Poulter, K.O. McElrath, *Surf. Interface Anal.*, **25**, 418 (1997).
65. S.N. Magonov, J. Cleveland, V. Elings, R. Denley, M.-H. Whangbo, *Surf. Sci.*, **85**, 201 (1997).
66. E. Sivaniah, E. Kramer, Ch. Ober, in preparation.
67. H. Ade, D.A. Winesett, A.P. Smith, S. Qu, S. Ge, J. Sokolov, M. Rafailovitch, *Europhys. Lett.*, **45**, 526 (1999).
68. T.J. McMaster, J.K. Hobbs, P.J. Barham, M.J. Miles, *Probe Microscopy*, **1**, 43 (1997).
69. R. Pearce, J. Vansco, *Polymer*, **39**, 1237 (1998).
70. S.N. Magonov, V. Elings, V.S. Papkov, *Polymer*, **38**, 297 (1997).
71. S. Magonov, Yu. Godovsky, *Am. Lab.*, **30**, 11 (1998).
72. A. Hammiche, M. Reading, H.M. Pollock, M. Song, D.J. Hourston, *Rev. Sci. Instrum.*, **67**, 4268 (1996).
73. Yu.K. Godovsky, S.N. Magonov, *Langmuir*, **16**, 3549 (2000).
74. Yu. Godovsky, V.S. Papkov, S.N. Magonov, *Macromolecules*, submitted.
75. A. Hammiche, H.M. Pollock, M. Reading, M. Claybourn, P.H. Turner, K. Jewkes, *Appl. Spectrosc.*, **53**, 810 (1999).

Coupled Liquid Chromatographic Techniques in Molecular Characterization

Peter Kilz
PSS Polymer Standards Service, Mainz, Germany

Harald Pasch
Deutsches Kunststoffinstitut, Darmstadt, Germany

1 Introduction 1
2 Coupled Techniques in Polymer Analysis 3
3 Coupling of Liquid Chromatography with Information-rich Detectors 4
3.1 Introduction 4
3.2 Coupling with Molar Mass-sensitive Detectors 6
3.3 Coupling with Mass Spectroscopy 11
3.4 Coupling with Fourier Transform Infrared Spectroscopy 18
3.5 Coupling with Nuclear Magnetic Resonance Spectroscopy 22
4 Multidimensional Liquid Chromatography 25
4.1 Introduction 25
4.2 Experimental Aspects of Multi-dimensional Separations 28
4.3 Separation Techniques for the First Dimension 31
4.4 Separation Techniques for the Second Dimension 34
4.5 State-of-the-art of On-line Coupled Two-dimensional Chromatography 36
4.6 Conclusions and Future Developments 43
List of Symbols 44
Abbreviations and Acronyms 44
Related Articles 45
References 45

Complex polymers are distributed in more than one direction of molecular heterogeneity. In addition to the molar mass distribution (MMD), they are frequently distributed with respect to chemical composition, functionality, and molecular architecture (see **Size-exclusion Chromatography of Polymers***). For the characterization of the different types of molecular heterogeneity it is necessary to use a wide range of analytical techniques. Preferably, these techniques should be selective towards a specific type of heterogeneity. The combination of two or more selective analytical techniques is assumed to yield multidimensional information on the molecular heterogeneity.*

The present review presents the fundamental ideas of combining liquid chromatography (LC) with other analytical techniques in multidimensional analysis schemes (see **Size-exclusion Chromatography of Polymers; Gas Chromatography in Analysis of Polymers and Rubbers; Field Flow Fractionation in Analysis of Polymers and Rubbers***). The capabilities and limitations of different coupling techniques are discussed and a number of relevant applications are given. It is shown that multidimensional structural information can be obtained when different chromatographic techniques are combined. Another approach is the hyphenation of LC with information-rich detectors. These detectors include molar mass-sensitive detection systems, such as on-line viscometry (VISC) and light scattering (LS). Information on the chemical composition of complex polymers can be obtained when spectroscopic techniques, like Fourier transform infrared (FTIR) (see* **Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships***), nuclear magnetic resonance (NMR) or mass spectrometry (MS) are coupled to LC.*

The basics and applications of multidimensional LC are addressed rather extensively. A brief introduction to different separation mechanisms is given and the particular requirements for the first and second dimensions are discussed. In conclusion, state-of-the-art examples for on-line coupled two-dimensional (2-D) chromatography are demonstrated, and future developments are reviewed.

1 INTRODUCTION

Today's polymeric materials are designed to meet very specific requirements defined by the application. Therefore, most synthetic polymers are highly complex multicomponent materials. They are composed of macromolecules varying in chain length, chemical composition, and architecture. By definition, complex polymers are heterogeneous in more than one distributed property (for example, linear copolymers are distributed in molar mass and chemical composition).

In general, the molecular structure of a macromolecule is described by its size, its chemical structure, and its architecture. The chemical structure characterizes the constitution of the macromolecule, its configuration and its conformation. For a complete description of the constitution the chemical composition of the polymer chain and the chain ends must be known. In addition to the type and quantity of the repeat units their sequence

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of incorporation must be described (alternating, random, or block in the case of copolymers). Macromolecules of the same chemical composition can still have different constitutions due to constitutional isomerism (1,2- versus 1,4-coupling of butadiene, head-to-tail versus head-to-head coupling, linear versus branched molecules). Configurational isomers have the same constitution but different steric patterns (*cis*- versus *trans* configuration; isotactic, syndiotactic and atactic sequences in a polymer chain). Conformational heterogeneity is the result of the ability of fragments of the polymer chain to rotate around single bonds. Depending on the size of these fragments, interactions between different fragments, and a certain energy barrier, more or less stable conformations may be obtained for the same macromolecule (rod-like versus coil conformation).

Depending on the composition of the monomer feed and the polymerization procedure, different types of heterogeneities may become important. For example, in the synthesis of tailor-made polymers frequently telechelics or macromonomers are used. These oligomers or polymers usually contain functional groups at the polymer chain end. Depending on the preparation procedure, they can have a different number of functional endgroups, i.e. be mono-, or bifunctional and so forth. In addition, polymers can have different architectures, i.e. they can be branched (star- or comb-like), and they can be cyclic.

The structural complexity of synthetic polymers can be described using the concept of molecular heterogeneity, see Figure 1, meaning the different aspects of MMD, chemical composition distribution (CCD), functionality type distribution (FTD) and molecular architecture distribution (MAD). They can be superimposed one on another, i.e. bifunctional molecules can be linear or branched, linear molecules can be mono- or bifunctional, copolymers can be block or graft copolymers etc. In order to characterize complex polymers it is necessary to know the MMD within each other type of heterogeneity.

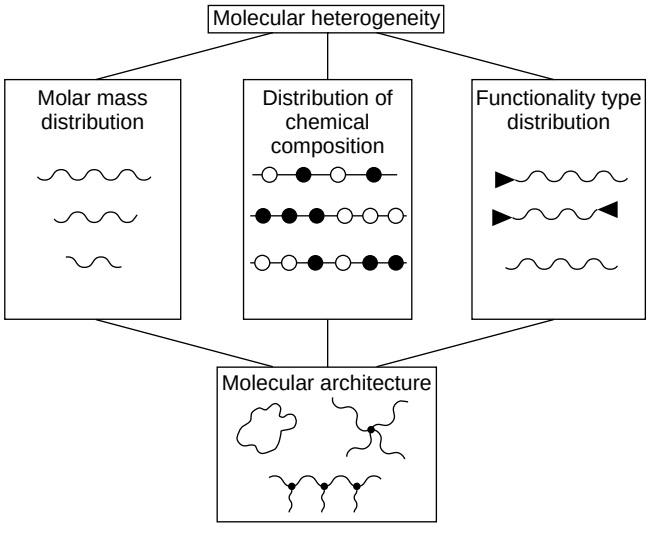

Figure 1 Schematic representation of the molecular heterogeneity of complex polymers.

Using the traditional methods of polymer analysis, such as NMR, one can determine the type and concentration of monomers or functional groups present in the sample. However, the determination of functional endgroups is complicated for long-chain molecules because of low concentration. On the other hand, these methods do not yield information on how different monomer units or functional groups are distributed in the polymer molecule. Finally, these methods in general do not provide molar mass information.

With respect to methods sensitive to the size of the macromolecule, one can face other difficulties. Size exclusion chromatography (SEC), which is most frequently used to separate polymer molecules from each other according to their molecular size in solution, must be used very carefully when analyzing complex polymers. The molecular size distribution of macromolecules can in general be unambiguously correlated with MMD only within one heterogeneity type. For samples consisting of a mixture of molecules of different functionality, the distribution obtained represents a sum of distributions of molecules having a different functionality and, therefore, cannot be attributed to a specific functionality type without additional assumptions. For the analysis of copolymers by SEC either the chemical composition along the molar mass axis must be known or detectors must be used which, instead of a concentration information, can provide molar mass information. To this end, SEC has to be coupled to composition-sensitive or molar mass-sensitive detectors.

Another option for the analysis of complex polymers is the separation with respect to chemical composition or functionality by means of interaction chromatography. In this case, functionally or chemically homogeneous fractions are obtained which then can be subjected to molar mass determination.

To summarize, for the complete analysis of complex polymers a minimum of two different characterization methods must be used. It is most desirable that each method is sensitive towards a specific type of heterogeneity. Maximum efficiency can be expected when, similar to the 2-D distribution in properties, 2-D analytical techniques are used. A possible approach in this respect is the coupling of different chromatographic modes in 2-D chromatography or the coupling of a separation technique with selective detectors, such as molar mass-sensitive or spectroscopic detectors.

2 COUPLED TECHNIQUES IN POLYMER ANALYSIS

Coupled techniques (also termed hyphenated techniques) are very frequently used in low molar mass organic chemistry. Using high-resolution chromatographic techniques, such as capillary gas chromatography (GC), gradient high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE), complex mixtures are separated into single components which are then identified by MS. By hyphenated GC/MS, HPLC/MS, and CE/MS up to several hundreds of different components can be separated and identified in one run with very high sensitivity. This is particularly important for environmental and biological samples, where frequently only very limited sample amounts are available.

Polymers are typically complex mixtures in which the composition depends on polymerization kinetics and mechanism and process conditions. To obtain polymeric materials of desired characteristics, polymer processing must be carefully controlled and monitored. Furthermore, one needs to understand the influence of molecular parameters on polymer properties and end-use performance. MMD and average chemical composition may no longer provide sufficient information for process and quality control nor define structure–property relationships. Modern characterization methods now require multidimensional analytical approaches rather than average properties of the whole sample.[1]

Different from low molar mass organic samples, where single molecules are to be determined, for complex synthetic polymers, the analytical task is the determination of a distributed property. The molecular heterogeneity of a certain complex polymer can be presented in either a three-dimensional (3-D) diagram or a so-called ''contour plot''. For a telechelic polymer these presentations are given in Figure 2. Using appropriate analytical methods, the type and concentration of the different functionality fractions must be determined and, within each functionality, the MMD has to be obtained. To do this, two different methods must be combined, each of which preferably is selective towards one type of heterogeneity. For example, a chromatographic method separating solely with respect to functionality could be combined with a molar mass selective method. Another approach would be the separation of the sample into different molar mass fractions which are then analyzed with respect to functionality.

For copolymers, in particular random copolymers, instead of discrete functionality fractions a continuous drift in composition is present, see Figure 3. To determine this chemical composition drift in interrelation with the MMD, a number of classical methods have been used, including precipitation, partition, and cross-fractionation.[2] The aim of these very laborious techniques was to obtain fractions of narrow composition and/or MMD which are then analyzed by spectroscopy and SEC.

During the last 20 years a number of techniques have been introduced in organic chemistry and applied to

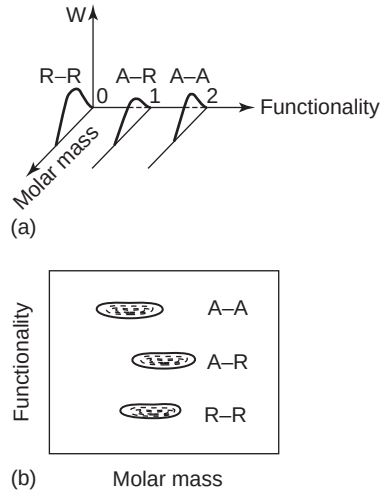

Figure 2 Representation of the molecular heterogeneity of a telechelic polymer in a 3-D diagram (a) and a contour plot (b).

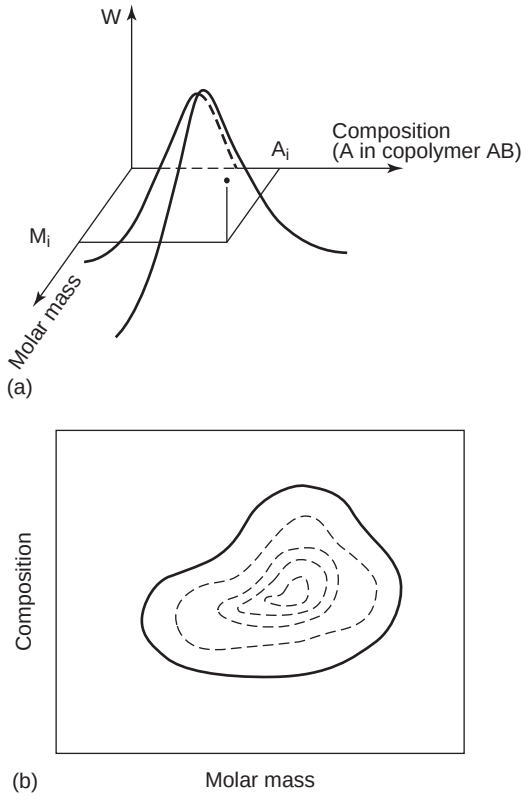

Figure 3 Representation of the molecular heterogeneity of a random copolymer in a 3-D diagram (a) and a contour plot (b).

polymer analysis, combining chromatographic separation with spectroscopic detection.[3] GC/MS has been used in polymer analysis,[4–11] but, due to the low volatility of high molar mass compounds it is limited to the oligomer region. The combination of pyrolysis and gas chromatography/mass spectrometry (GC/MS), however, is of great value for polymer characterization.[12,13] It provides for the analysis of complex polymers with respect to chemical composition. For a number of polymer systems characteristic low molar mass pyrolysis products are obtained, which yield information of the average composition and the ''blockiness'' of the polymer chain. Molar mass information, however, is not available from pyrolysis-GC/MS.

Much more important for polymer analysis than GC are the different techniques of LC. Using SEC, liquid adsorption chromatography (LAC), or liquid chromatography at the critical point of adsorption (LC/CC) polymers can be fractionated with respect to different aspects of molecular heterogeneity, including molar mass, functionality, and chemical composition. The advantage of these techniques over GC is that intact macromolecules are separated and analysed. As will be shown in the next sections, LC can be efficiently coupled to infrared (IR) spectroscopy,[14–19] to MS, and to NMR spectroscopy.[20,21]

Another most efficient approach is the chromatographic separation of complex polymers by combining different separation mechanisms. This can be done by coupling two chromatographs in an off-line or on-line mode. Each of these chromatographs must operate in a mode which is selective towards one type of molecular heterogeneity. This 2-D chromatography has been termed ''orthogonal chromatography'' assuming the selectivity of each separation method with respect to one distribution function, e.g. MMD, FTD, or CCD.[22] The first truly automated 2-D chromatography set-up for polymer analysis was proposed by Kilz et al.,[23] who coupled gradient HPLC and SEC.

The need for such analysis protocols results from the fact that in complex polymers in addition to chemical heterogeneity of the first kind, another type of chemical heterogeneity may exist: chemical heterogeneity of the second kind, in which polymers of different composition and chain length have similar hydrodynamic volumes and, hence, co-elute in SEC, see Figure 4.

A possible separation protocol for a complex polymer mixture is presented in Figure 5. The sample under investigation comprises molecules of different chemical compositions (different colors) and different sizes. In a first separation step this mixture is separated according to composition, yielding fractions which are chemically homogeneous. These fractions are transfered to a size-selective separation method and analyzed with respect to molar mass. As a result of this 2-D separation, information on both types of molecular heterogeneity is obtained.

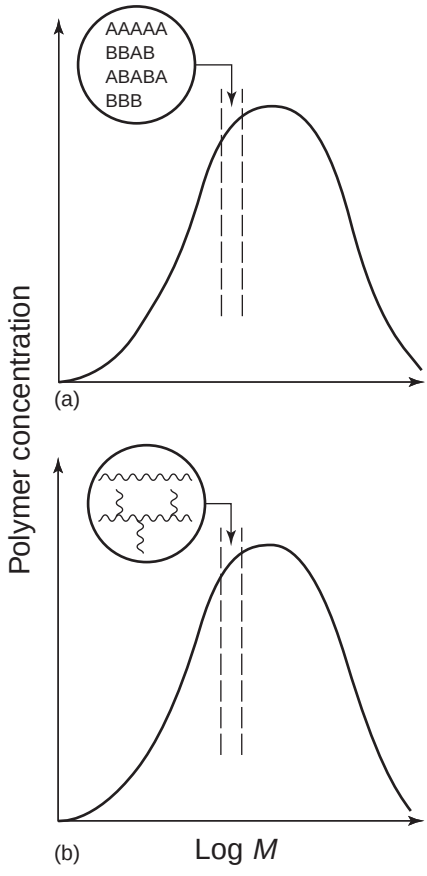

Figure 4 SEC fractionation showing composition or architecture at a given retention volume. (Reproduced by permission from Barth.[1])

3 COUPLING OF LIQUID CHROMATOGRAPHY WITH INFORMATION-RICH DETECTORS

3.1 Introduction

LC of polymers is often understood to be synonymous with SEC. SEC separates polymers according to the size of the macromolecules by entropic interactions and enables the MMD of a sample to be evaluated. However, in addition to size exclusion phenomena, other types of interaction of the macromolecules and the stationary and mobile phases of a chromatographic system can be used for separation. LAC uses enthalpic interactions to separate substances such as copolymers according to chemical composition. Finally, LC/CC can be used for functionality type separation by balancing entropic and enthalpic interactions in the chromatographic system.

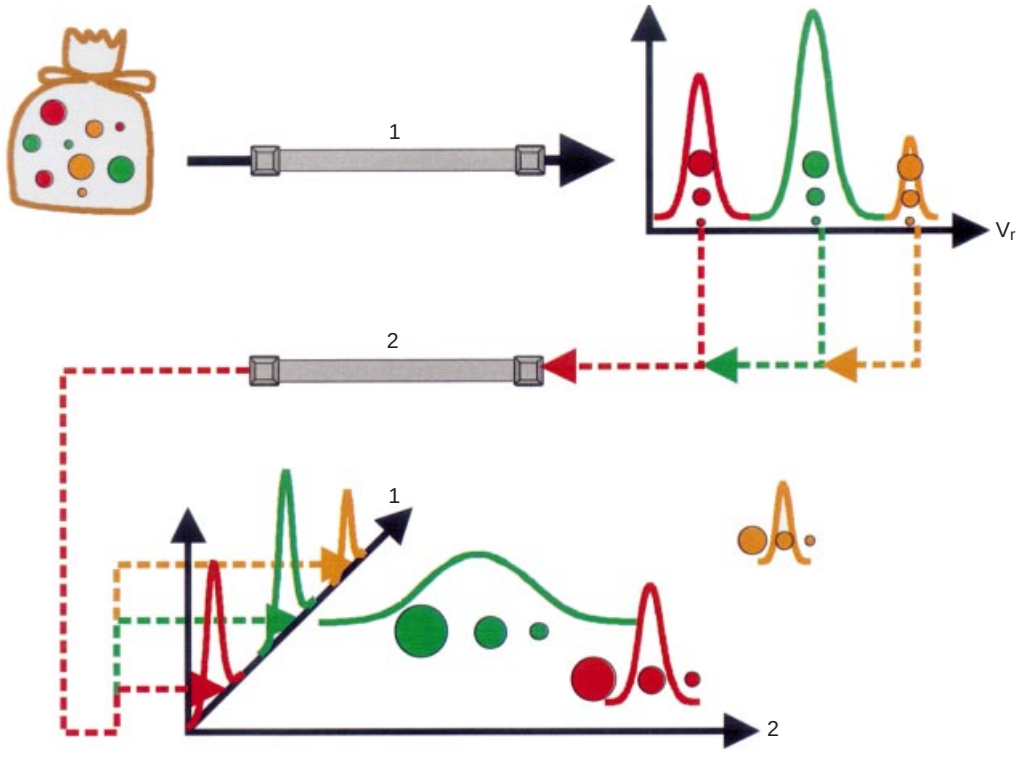

Figure 5 Schematic separation protocol for the analysis of a complex polymer mixture.

SEC is the premier polymer characterization method for determining MMD. In SEC, the separation mechanism is based on molecular hydrodynamic volume. For homopolymers, condensation polymers and strictly alternating copolymers, there is a correspondence between elution volume and molar mass. Thus, chemically similar polymer standards of known molar mass can be used for calibration. However, for SEC of random and block copolymers and branched polymers, no simple correspondence exists between elution volume and molar mass because of possible compositional heterogeneity of these materials. The dimensional distribution of macromolecules can, in general, be unambiguously correlated with MMD only within one heterogeneity type. For samples consisting of molecules of different chemical composition, the distribution obtained represents an average of dimensional distributions of molecules having a different composition and, therefore, cannot be attributed to a certain type of macromolecule.

The inadequacy of using SEC without further precaution for the determination of MMD of polymer blends or copolymers results from the following consideration: for a linear homopolymer distributed only in molar mass, fractionation by SEC results in one molar mass being present in each retention volume. The polymer at each retention volume is monodisperse. If a blend of two linear homopolymers is fractionated then two different molar masses can be present in one retention volume. If now a copolymer is analyzed then a multitude of different combinations of molar mass, composition, and sequence length can be combined to give the same hydrodynamic volume. In this case, fractionation with respect to molecular size is completely ineffective in assisting the analysis of composition or MMD.

Three on-line methods are used to try to characterize copolymers by SEC with respect to MMD and composition:

- conventional SEC utilizing multiple concentration detection
- on-line analysis of SEC fractions with a LS detector.
- VISC.

The experimentally simplest approach is the combination of SEC with multiple concentration detectors. If the response factors of the detectors for the components of the polymer are sufficiently different, the chemical composition of each slice of the elution curve can be determined from the detector signals. Typically, a combination of ultraviolet (UV) and refractive index (RI) detection is used; another possibility is the use of a diode-array detector. In the case of non-UV absorbing polymers, a combination of RI and density detection yields information on chemical composition.[24–26]

The principle of dual detection is rather simple: when a mass m_i of a copolymer, which contains the weight fractions w_A and w_B ($=1 - w_A$) of the monomers A and B, is eluted in the slice i (with the volume ΔV) of the peak, the area $x_{i,j}$ of slice i obtained from detector j depends on the mass m_i (or the concentration $c_i = m_i/\Delta V$) of polymer in the slice, its composition (w_A), and the corresponding response factors $f_{j,A}$ and $f_{j,B}$, wherein j denotes the individual detector, as in Equation (1) below:

$$x_{i,j} = m_i(w_A f_{j,A} + w_B f_{j,B}) \tag{1}$$

The weight fractions w_A and w_B of the monomers can be calculated using Equation (2) below:

$$\frac{1}{w_A} = 1 - \left\{ \frac{(x_1/x_2)f_{2,A} - f_{1,A}}{(x_1/x_2)f_{2,B} - f_{1,B}} \right\} \tag{2}$$

Once the weight fractions of the monomers are known, the correct mass of polymer in the slice can be calculated using Equation (3) as follows:

$$m_i = \frac{x_i}{w_A(f_{1,A} - f_{1,B}) + f_{1,B}} \tag{3}$$

and the molar mass M_C of the copolymer is obtained by interpolation between the calibration lines of the homopolymers[27] which is given in Equation (4):

$$\ln M_C = \ln M_B + w_A(\ln M_A - \ln M_B) \tag{4}$$

wherein M_A and M_B are the molar masses of the homopolymers, which would elute in this slice of the peak (at the corresponding elution volume V_e).

It is clear that the interpolation between the calibration lines cannot be applied to mixtures of polymers (polymer blends): if the calibration lines are different, different molar masses of the homopolymers will elute at the same volume. The universal calibration is not capable of eliminating these errors, either, which originate from the simultaneous elution of two polymer fractions with the same hydrodynamic volume, but different composition and molar mass.

The architecture of a copolymer (random, block, graft) has also to be taken into account, as Revillon[28] has shown by SEC with RI, UV, and viscosity detection. Intrinsic viscosity varies largely with molar mass according to the type of polymer, its composition, and the nature of its components. Tung[29] found that for block copolymers in good SEC solvents the simpler first approach (Equation 4) is more precise.

Further information on quantitative aspects of SEC with dual detection can be obtained from Trathnigg et al.[30] Different applications of dual detection SEC in the analysis of segmented copolymers,[31] block copolymers,[32,33] star polymers,[34] and polymer blends[35,36] are also available. The limitation of SEC with dual detection is that only binary combinations of monomers can be investigated successfully. In the case of ternary combinations, more than two detectors must be used or one of the detectors must be able to detect two components simultaneously.

To overcome the problems related to classical SEC of complex polymers, molar mass-sensitive detectors are coupled to the SEC instrument. Since the response of such detectors depends on both concentration and molar mass, they have to be combined with a concentration-sensitive detector. The following types of molar mass-sensitive detectors are used frequently:[37–40]

- differential viscometer
- low angle laser light scattering (LALLS) detector
- multiangle laser light scattering (MALLS) detector.

3.2 Coupling with Molar Mass-sensitive Detectors

As has been pointed out, for SEC of complex polymers no simple correspondence exists between elution volume and molar mass. It is, therefore, useful to determine the molar mass not via a calibration curve but directly from the SEC effluent. This can be done by using molar mass-sensitive detectors based on Rayleigh LS or intrinsic viscosity measurements.[41]

In a LS detector, the scattered light of a laser beam passing through the cell is measured at angles different from zero. The (excess) intensity of the scattered light at the angle Θ ($R(\Theta)$) is related to the weight-average of molar mass M_w as expressed by Equation (5):

$$\frac{K^*c}{R(\Theta)} = \frac{1}{M_w P(\Theta)} + 2A_2c \tag{5}$$

wherein c is the concentration of the polymer, A_2 is the second virial coefficient, and $P(\Theta)$ describes the scattered light angular dependence. K^* is an optical constant containing Avogadro's constant N_A, the wavelength λ_0, the RI n_0 of the solvent, and the RI increment dn/dc of the sample. Their relationship is described by Equation (6):

$$K^* = \frac{4\pi^2 n_0^2 (dn/dc)^2}{\lambda_0^4 N_A} \tag{6}$$

In a plot of $K^*c/R(\Theta)$ versus $\sin^2(\Theta/2)$, M_w can be obtained from the intercept and the radius of gyration (R_g) from the slope. A multiangle measurement provides additional information.

In most cases the injected concentration is small and A_2 can be neglected. Thus, if the optical properties (n_0 and dn/dc) of the polymer solution are known, the molar mass at each elution volume increment can be determined as expressed by Equation (7):

$$M_{w,i} = \frac{R(\Theta)_i}{K^* P(\Theta)_i c_i} \tag{7}$$

If a low-angle LS instrument is used, $P(\Theta)$ is close to unity and $M_{w,i}$ can be calculated directly. For a multiangle LS instrument, the mean-square radius of gyration $\langle R_g^2 \rangle$ at each elution volume can also be obtained from $P(\Theta)$ as shown in Equation (8):

$$\frac{1}{P(\Theta)_i} = 1 + \frac{q^2 \langle R_g^2 \rangle_i}{3} \qquad (8)$$

$$q = \left(\frac{4\pi}{\lambda_0} \quad \sin \left(\frac{P(\Theta)}{2} \right. \right.$$

In practice, however, the radius of gyration can only be determined for molecules larger than 20 nm in diameter. By measuring radius of gyration as a function of M_w, insight into the molecular conformation of the polymer can be obtained.[1]

Molar mass determination requires the knowledge of the specific RI increment dn/dc which in the case of complex polymers depends on chemical composition. Copolymer RI increments $(dn/dc)_{copo}$ can be calculated accurately for chemically monodisperse fractions, if comonomer weight fractions w_i and homopolymer values are known, as described in Equation (9):

$$\left(\frac{dn}{dc} \right)_{copo} = \sum w_i \left(\frac{dn}{dc} \right)_i \qquad (9)$$

However, in some cases additional effects on $(dn/dc)_{copo}$ must be considered. Due to cooperative interactions between the monomer units in the polymer chain, copolymer RI increments may deviate from the summation scheme. As a result of different sequence length distributions, different $(dn/dc)_{copo}$ can be obtained for the same gross composition. Copolymer $(dn/dc)_{copo}$ values can be obtained by multiple detection SEC providing the chemical composition at each slice of the elution curve.

Unfortunately, LS investigations of copolymers are complicated even further by the fact that SEC does not separate into chemically monodisperse fractions. Accordingly, due to compositional heterogeneity the RI increment of a particular scattering center may be different from the total dn/dc of the corresponding SEC slice. Therefore, in general, only apparent molar masses for copolymers can be measured.[34] Another influencing factor is the RI of the solvent. As has been shown by Kratochvil,[42] the solvent RI should be significantly different from the values of the copolymer fractions and the corresponding homopolymers.

The evaluation of LS detectors for SEC was conducted by Jeng et al. with respect to precision and accuracy[43] and the proper selection of the LS equation.[44] The results obtained for polystyrene (PS) and polyethylene were compared for a low-angle and a multiangle LS instrument. The application of SEC/LS has been discussed in a multitude of papers. In addition to determining M_w values, the formation of microgels has been studied by Pille and Solomon.[45] Mourey and Coll investigated high molar mass PS and branched polyesters, and discussed the problems encountered in molar mass and radius of gyration determination.[46,47] Grubisic-Gallot et al. proved that SEC/LS is useful for analysing micellar systems with regard to determining molar masses, qualitative evaluation of the dynamics of unimer-micelles re-equilibration, and revealing the mode of micelle formation.[48–50]

Another very useful approach to molar mass information of complex polymers is the coupling of SEC to a viscosity detector.[51–56] The viscosity of a polymer solution is closely related to the molar mass (and architecture) of the polymer molecules. The product of polymer intrinsic viscosity $[\eta]$ times molar mass is proportional to the size of the polymer molecule (the hydrodynamic volume). Viscosity measurements in SEC can be performed by measuring the pressure drop ΔP across a capillary, which is proportional to the viscosity η of the flowing liquid (the viscosity of the pure mobile phase is denoted as η_0). The relevant parameter $[\eta]$ is defined as the limiting value of the ratio of specific viscosity $(\eta_{sp} = (\eta - \eta_0)/\eta_0)$ and concentration c for $c \to 0$, as shown by Equation (10):

$$[\eta] = \lim \frac{\eta - \eta_0}{\eta_0 c} = \lim \frac{\eta_{sp}}{c} \quad \text{for } c \to 0 \qquad (10)$$

The viscosity of a polymer solution as compared to the viscosity of the pure solvent is measured by the pressure drop ΔP across an analytical capillary-transducer system. The specific viscosity is obtained from $\Delta P/P$, where P is the inlet pressure of the system. As the concentrations in SEC are usually very low, $[\eta]$ can be approximated by η_{sp}/c.

A simple approach using one capillary and one differential pressure transducer will not work very well, because the viscosity changes $\Delta\eta = \eta - \eta_0$ will typically be very small compared to η_0, which means that one has to measure a very small change of a large signal. Moreover, flow-rate fluctuations due to pulsations of a reciprocating pump will lead to much greater pressure differences than the change in viscosity due to the eluted polymer. Instruments of this type should be used with a positive displacement pump.

A better approach is the use of two capillaries (C1 and C2) in series, each of which is connected to a differential pressure transducer (DP1 and DP2), and a sufficiently large holdup reservoir (HR) in between. With this approach, one measures the sample viscosity η from the pressure drop across the first capillary, and the solvent viscosity η_0 from the pressure drop across the second capillary. Pulsations are eliminated in this set-up, because they appear in both transducers simultaneously. Another design is that of the differential viscometer, in which four

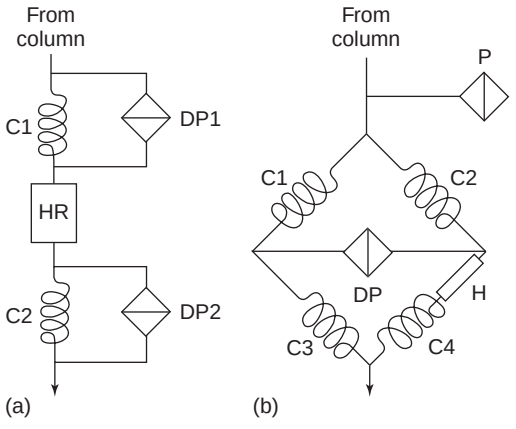

Figure 6 Schematic representation of differential viscometers. P is inlet pressure transducer: C1–C4 are flow restriction capillaries.

capillaries are arranged in a manner similar to that of a Wheatstone bridge. In Figure 6, both designs are shown schematically.

In the ''bridge'' design, a holdup reservoir in front of the reference capillary (C4) makes sure that only pure mobile phase flows through the reference capillary, when the peak passes the sample capillary (C3). This design offers considerable advantages: the detector actually measures the pressure difference ΔP at the differential pressure transducer (DP) between the inlets of the sample capillary and the reference capillary, which have a common outlet, and the overall pressure P at the inlet of the bridge. The specific viscosity $\eta_{sp} = \Delta\eta/\eta_0$ is thus obtained from $\Delta P/P$. One concern with this type of detector is that the flow must be divided in the ratio of 1 : 1 between both arms of the bridge. This is achieved by capillaries C1 and C2, which must have a sufficiently high back pressure. Nevertheless, when a peak passes through the sample capillary, a slight deviation of the 1 : 1 ratio will be observed. A problem of flow-rate variations exists also in a single capillary viscometer: when the polymer peak passes through the measuring capillary, the increased back pressure leads to a peak shift.[57]

Being able to determine $[\eta]$ as a function of elution volume, one can now compare the hydrodynamic volumes (V_h) for different polymers. The hydrodynamic volume is, through Einstein's viscosity law, related to intrinsic viscosity and molar mass by $V_h = [\eta]M/2.5$. Einstein's law is, strictly speaking, valid only for impenetrable spheres at infinitely low volume fraction of the solute (equivalent to concentration at very low values). However, it can be extended to particles of other shapes, defining the particle radius then as the radius of a hydrodynamically equivalent sphere. In this case V_h is defined as the molar volume of impenetrable spheres which would have the same frictional properties or enhanced viscosity to the same degree as the actual polymer in solution.

Assuming the validity of this approach and in agreement with the SEC mechanism, similar elution volumes correspond to similar hydrodynamic volumes, as shown in Equation (11):

$$V_{e,1} = V_{e,2} \longrightarrow M_1[\eta]_1 = M_2[\eta]_2 \tag{11}$$

In a plot of log $(M[\eta])$ versus V_e identical calibration lines should be found for the two polymers 1 and 2, irrespective of their chemical composition. This ''universal calibration'' approach has been predicted and experimentally proved by Benoit et al.[58] As a consequence, using the universal calibration curve established with known calibration standards (for example PS), one can obtain the SEC-molar mass calibration for an unknown polymer sample.

The intrinsic viscosity is a function of molar mass given by the Mark–Houwink relationship (Equation 12), wherein K and a are coefficients for a given polymer in a given solvent at a given temperature.

$$[\eta] = KM^a \tag{12}$$

This leads to Equation (13):

$$K_1 M_1^{a(1)+1} = K_2 M_2^{a(2)+1} \tag{13}$$

If a column has been calibrated with polymer 1 (e.g. PS), the calibration line for polymer 2 can be calculated, provided that the coefficients K and a are known for both polymers with sufficient accuracy. This is shown by Equation (14):

$$\ln M_2 = \left(\frac{1}{1+a_2}\right. \ln\left(\frac{K_1}{K_2}\right. + \left(\frac{1+a_1}{1+a_2}\right. \ln M_1 \tag{14}$$

Thus, the concept of universal calibration provides an appropriate calibration also for polymers for which no calibration standards exist. The limiting factor of this approach is the accuracy of determining K and a. There are very high variations in the values reported in the literature.[59,60] Even for such common polymers as PS and polymethyl methacrylate (PMMA) the values may differ considerably.

If the Mark–Houwink coefficients are not available, a universal calibration curve is established using PS calibration standards and the SEC–viscometer combination. The basic steps involved in the MMD analysis are summarized in Figure 7. First, the universal calibration curve of the SEC separation system has to be established by using narrow molar mass standards as indicated by the top arrow pointing to the right. Once the universal calibration curve is established, one can then reverse the procedure, by going from right to left following the bottom arrow, to

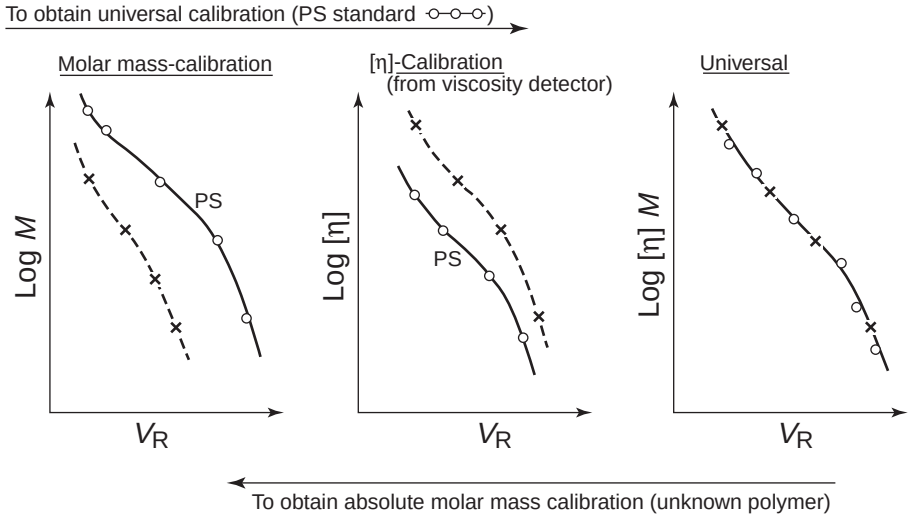

Figure 7 Determination of absolute molar masses via universal SEC calibration.

obtain the molar mass calibration curve of any unknown polymer. The calibration curve is obtained literally by substracting the [η] calibration curve of the unknown sample from the universal calibration curve. The [η] calibration curve for the unknown sample is obtained from the on-line viscometer.[61]

The application of RI and differential viscometer detection in SEC has been discussed by a number of authors.[62–64] Lew et al. presented the quantitative analysis of polyolefines by high-temperature SEC and dual RI–viscosity detection.[65] They applied a systematic approach for multidetector operation, assessed the effect of branching on the SEC calibration curve, and used a signal averaging procedure to define better intrinsic viscosity as a function of retention volume. The combination of SEC with RI and viscosity detectors was used to determine molar mass and functionality of polytetrahydrofurane simultaneously.[66] Long chain branching in ethylene–propylene–diene rubber (EPDM) copolymers by SEC–viscometry was analysed by Chiantore et al.[67]

One of the difficult problems in characterizing copolymers and polymer blends by SEC–viscometry is the accurate determination of the polymer concentration across the SEC elution curve. The concentration detector signal is a function of the chemical drift of the sample under investigation. To overcome this problem, Goldwasser proposed a method where no concentration detector is required for obtaining number-average molar mass (M_n) data.[68] In the usual SEC–viscometry experiment, the determination of the intrinsic viscosity at each slice of the elution curve requires a viscosity and a concentration signal as shown by Equation (15):

$$[\eta]_i = \left(\frac{\ln \eta_{rel}}{c}\right)_i \tag{15}$$

where $\ln \eta_{rel}$ is the direct detector response of the viscometer. One calculates the molar mass averages by the expressions given in Equation (16) and in Equation (17):

$$M_n = \frac{\sum c_i}{\sum [c_i/(V_{h,x}/[\eta])_i]} \tag{16}$$

$$M_w = \frac{\sum c_i (V_{h,x}/[\eta])_i}{\sum c_i} \tag{17}$$

where $V_{h,x} = [\eta]_x M_x$ is the data retrievable from the universal calibration curve. By rearranging Equation (17) using Equation (16) the following expression (Equation 18) is obtained:

$$M_n = \frac{\sum c_i}{\sum (\ln \eta_{rel}/V_{h,x})_i} \quad \text{or}$$

$$M_n = \frac{\text{sample amount}}{\sum (\ln \eta_{rel}/V_{h,x})_i} \tag{18}$$

The sample amount can be determined easily from the injection volume and the sample concentration and no information from a concentration detector is required. With this approach, the M_n value of any polymer sample can be determined by SEC using only the viscosity detector. Other molar mass averages, however, cannot be determined. The advantage of the Goldwasser M_n method is that it can access much wider molar mass ranges than other existing methods like osmometry or endgroup methods.

Due to the problems encountered with SEC/LALLS and SEC–viscometry, a triple-detector SEC technology has been developed, where three on-line detectors are

used together in a single SEC system. In addition to the concentration detector, an on-line viscometer and a LALLS instrument are coupled to the SEC: this arrangement is known as TriSEC. With TriSEC, absolute molar mass determination is possible for polymers that are very different in chemical composition and molecular conformation. The usefulness of the TriSEC approach has been demonstrated in a number of applications. It was shown by Pang and Rudin that only by using both viscometer and LS detection are accurate MMDs obtained.[69] Wintermantel et al. have developed a custom-made multidetector instrument and demonstrated that it has great potential not only for absolute molar mass determinations but also for structure characterization of linear flexible, semiflexible, and branched polymers.[70] Degoulet et al. characterized polydisperse solutions of branched PMMA,[71] while Jackson et al. investigated linear chains of varying flexibility in order to prove universal calibration.[72] Yau and Arora discussed the advantages of TriSEC for the determination of Mark–Houwink coefficients, long-chain branching, and polymer architecture.[73]

Finally, several attempts have been made to develop an absolute molar mass detector based on osmotic pressure measurements. Commercially available membrane osmometers are designed for static measurements, and the cell design with a flat membrane is not suited for continuous flow operation. Yau[61,74] developed a detector which is different from the conventional design; it measures the flow resistance of a column caused by osmotic swelling and deswelling of soft gel particles used for the packing, see Figure 8. With a microbore gel column an M_n sensitive detector with a fast response was obtained which could be coupled to the SEC equipment. However, since the change in flow resistance could not easily be related to the osmotic pressure of the solution, absolute calibration was lost.

Recently, an osmometer based on a concentric design with a capillary-shaped membrane has been developed by Köhler et al.[75] and Lehmann et al.[76] The flow cell volume is 12.2 μL, the response time approximately 15 s, and the molar mass cut-off is below 5.000 g mol^{-1}. The design of the cell is given in Figure 9. The cylinder symmetry and stiffness of the osmometer and the favorable properties of the membrane were combined to meet the requirements for on-line detection. Testing the instrument in both batch and continuous flow operation with PS standards yielded reproducible results and good agreement with the nominal molar masses. However, the osmometer still caused a certain peak broadening, and the pressure noise level still strongly exceeded the noise of the concentration detector.

As has been discussed, the combination of SEC and molar mass-sensitive detectors is a powerful tool for the analysis of complex polymers. However, it is important to distinguish between claimed versus actual capabilities and between potential expectations and demonstrated performance. Tables 1 and 2 below, taken from a critical review of different techniques summarize the information content and additional details of SEC/LS and SEC–viscometry coupling.[61] The information content is classified into two categories. ''Primary'' information is of high precision and accuracy, insensitive to SEC operation variables, and does not require molar mass or universal calibration. ''Secondary'' information is less precise and requires calibration.

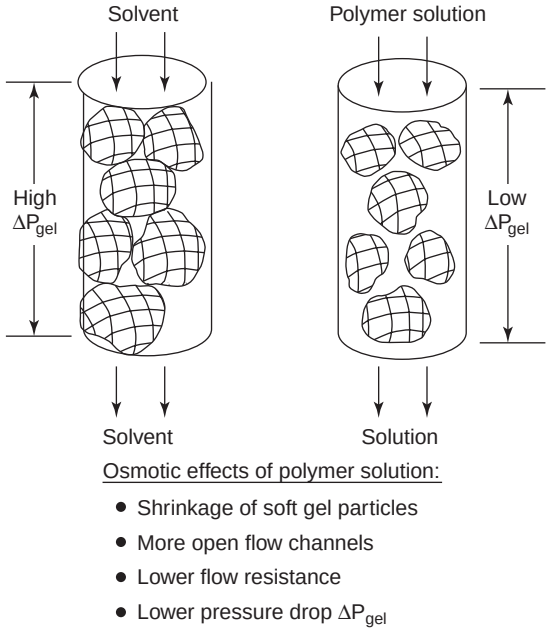

Figure 8 Differential pressure measurement of osmotic effect on a soft gel column.

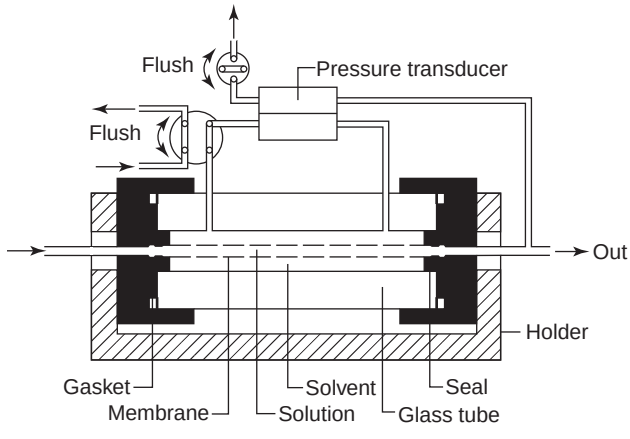

Figure 9 Design of a concentric osmometric flow cell. (Reproduced by permission from Lehmann et al.[76])

Table 1 SEC analysis using molar mass-sensitive detectors

Method	Information content	
	Primary	Secondary
Conventional SEC		MMD
SEC/LALLS	MMD	
SEC/MALLS	MMD	R_g distribution
SEC/VISC	[η] distribution	MMD R_g distribution Copolymer M_n
SEC/VISC/LS	[η] distribution MMD R_g distribution	Copolymer M_n

Table 2 Generalization of molar mass-sensitive detectors

Intended measurements	LALLS/MALLS	Viscometer
MMD	Requires precise n and dn/dc values Not affected by nonexclusion effects	Requires universal calibration and K, a-parameters
[η] distribution		Directly from experiment Not affected by nonexclusion effects
R_g distribution	MALLS only	Calculable from $[\eta]M$
Chain conformation and branching	R_g vs M plot, MALLS only	[η] vs M plot, R_g vs M plot
Chemically heterogeneous polymer analysis	Limited	Better
Noise, particulates, bubbles	Strongly affected	Less affected

In addition, the complex procedures related to SEC/LS and SEC–viscometry coupling are a potential source of error. According to Jackson and Barth[77] these include:

1. Accuracy of the universal calibration curve.
2. Detector configuration: arrangement of multiple detectors in series or in parallel can cause additional peak broadening, flow rate variations, back pressure variations.
3. Interdetector volume: detectors are placed at different physical positions and their signals must be aligned very precisely.
4. Detector sensitivity: LS and viscosity detectors are very sensitive towards higher molar masses, while the RI detector is most sensitive at lower molar masses.
5. Low molar mass fractions: polymer molecules may not adopt random coil conformation, the Mark–Houwink coefficients become functions of molar mass.

To summarize, although the principal limitation of SEC separating according to hydrodynamic volume and not molar mass cannot be overcome, the advantages of multidetector SEC in the accurate characterization of complex polymers are significant. However, in order to generate reproducible and accurate results on a routine basis, special care must be taken regarding the added complexity of the instrumentation. In addition to improving the design of multidetector SEC set-ups, important advances are expected from methods for determining the chemical composition across the MMD by interfacing SEC with FTIR spectroscopy, MS, and NMR.

3.3 Coupling with Mass Spectroscopy

From the very early stages of development of modern MS, the value of its combination with chromatography was quickly recognized. The coupling of GC with MS was a natural evolution since they are both vapor phase techniques, and very quickly GC/MS has been accepted as a standard component of the organic analytical laboratory. It has taken considerably longer to achieve a satisfactory and all-purpose mode of HPLC/MS coupling. The difficulties with HPLC/MS were associated with the fact that vaporization of typically $1\,mL\,min^{-1}$ from the HPLC translates into a vapor flow-rate of approx. $500–1000\,mL\,min^{-1}$. Other difficulties related to the eluent composition as a result of the frequent use of nonvolatile modifiers, and the ionization of nonvolatile and thermally labile analytes. However, during the past several years commercial interfaces have been developed which have led to a broad applicability of HPLC/MS.[78–80] The techniques necessary for the successful introduction of a liquid stream into a mass spectrometer are based on the following principles: electrospray ionization (ESI),[81] atmospheric pressure chemical ionization,[82] thermospray ionization,[83] and particle beam ionization.[84]

From the point of view of polymer analysis, a mass spectrometric detector would be a most interesting alternative to the conventional detectors because this detector could provide absolute molar masses of polymer components.[85,86] Provided that fragmentation does not occur, intact molecular ions could be measured. The measured mass of a particular component could then be correlated with chemical composition or chain length. However, the major drawback of most conventional HPLC/MS techniques is the limited mass range, preventing higher

oligomers (molar mass above 2000–3000 g mol^{-1}) to be ionized without fragmentation.[87–89]

The use of MS for detailed polymer analysis has become increasingly established due to the introduction of soft ionization techniques that afford intact oligomer or polymer ions with less fragmentation.[90–93] One of these techniques, ESI/MS, has been widely applied in biopolymer analysis. Proteins and biopolymers are typically ionized through acid–base equilibria. When a protein solution (the effluent from an HPLC separation) is exposed to an electrical potential it ionizes and disperses into charged droplets. Solvent evaporation upon heat transfer leads to the shrinking of the droplets and the formation of analyte ions. Larger molecules acquire more than one single charge, and, typically, a mixture of differently charged ions is obtained.

Unfortunately, up to now ESI/MS has had limited application in polymer analysis.[94,95] Unlike biopolymers, most synthetic polymers have no acidic or basic functional groups that can be used for ion formation. Moreover, each molecule gives rise to a charge distribution envelope, thus complicating the spectrum further. Therefore, synthetic polymers that can typically contain a distribution of chain lengths and have a variety in chemical composition or functionality furnish complicated mass spectra, making interpretation nearly impossible.

To overcome the difficulties of ESI/MS, Prokai and Simonsick added sodium cations to the mobile phase to facilitate ionization.[96,97] To simplify the resulting ESI spectra, the number of components entering the ion source was reduced. Prokai et al. implemented microcolumn SEC for the separation of polydisperse mixtures prior to ESI detection.[98] They used a 250 × 0.5 mm internal diameter SEC column for the molar mass separation of octylphenoxy polyethylene oxide (PEO). Applying a flow of 4 μL min, they were able to supply the effluent from the column directly into the ESI source. To promote ionization, a sheath liquid of sodium iodide in methanol was delivered to the ESI interface. Figure 10 shows a representative chromatogram and mass spectrum from the SEC/ESI analysis. The mass spectrum was obtained by averaging between 6.9 and 9.2 min. It shows singly and doubly charged molecules in the molar mass range of 1000–2000 g mol^{-1}.

The analysis of PEOs by SEC/ESI/MS with respect to chemical composition and oligomer distribution was discussed by Simonsick.[99] In a similar approach, aliphatic polyesters,[100] phenolic resins,[101] methyl methacrylate macromonomers[101] and polysulfides[102] have been analysed. The detectable mass range for different species, however, was well below 5000 g mol^{-1}, indicating that the technique is not really suited for polymer analysis.

The quantitative analysis of PMMA-butyl acrylate copolymers by coupled LC and particle beam MS has been described by Murphy et al.[103] For separation with respect to chemical composition gradient HPLC was used. The copolymer composition was determined by monitoring several low-mass fragments formed by thermal decomposition and electron impact ionization in the particle beam interface.

Matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF) MS is one of the newest soft ionization techniques that allows desorption and ionization of very large molecules even in complex mixtures. In

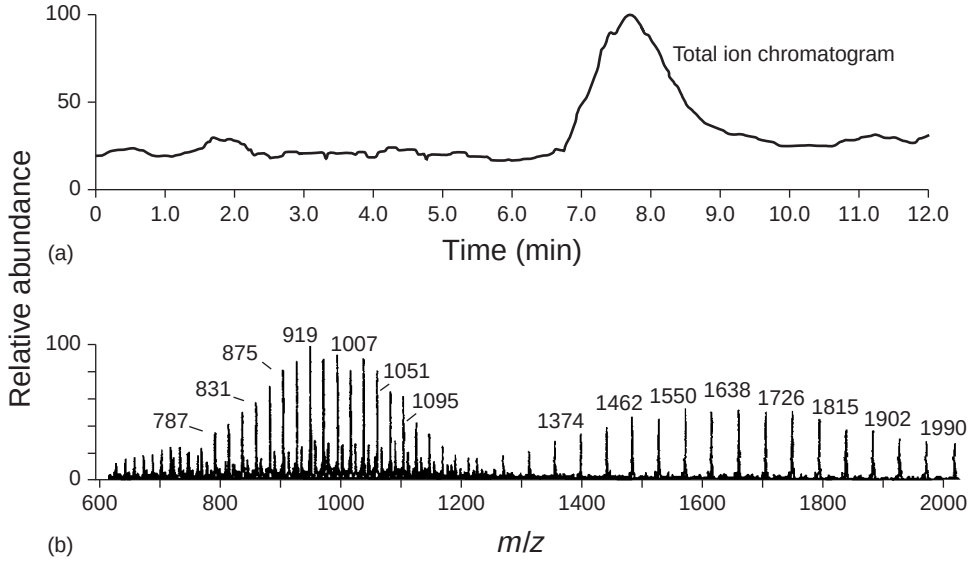

Figure 10 Micro-SEC/ESI analysis of octyloxy PEO, (a) TIC chromatogram, (b) averaged mass spectrum. (Reproduced by permission from Prokai et al.[98])

polymer analysis, the great promise of MALDI/TOF/MS is to perform the direct identification of mass-resolved polymer chains, including intact oligomers within an MMD, and the simultaneous determination of structure and endgroups in polymer samples. This most promising method for the ionization of large molecules and analysis according to their molar mass and functionality has been introduced by Karas and Hillenkamp[104,106] and by Beavis and Chait.[105] Compared to other MS techniques, the accessible mass range has been extended considerably, and the technique is fast and instrumentally very simple. Moreover, relatively inexpensive commercial instrumentation has become accessible. In principle, the sample to be investigated and a matrix solution are mixed in such a ratio that matrix separation of the sample molecules is achieved. After drying, a laser pulse is directed onto the solid matrix to photo-excite the matrix material. This excitation causes the matrix to explode, resulting in the expulsion and soft ionization of the sample molecules. Once the analyte is ionized, it is accelerated and analysed in a TOF mass spectrometer. As a result, the analyte is separated according to the molar mass of its components, and in the case of heterogeneous polymers additional information on chemical composition may be obtained. In a number of papers it was shown that polymers may be analysed up to relative molar masses of about 500 000 Da.[107–111] It was shown in a number of applications that functionally heterogeneous polymers can be analysed with respect to the degree of polymerization and the type of functional groups.[112–115]

The on-line combination of LC and MALDI/TOF/MS would be of great value for polymer analysis. In particular, for chemically or functionally heterogeneous polymers LC could provide separation with respect to chemical composition while MALDI/TOF would analyse the fractions with respect to oligomer distribution or molar mass. Unfortunately, MALDI/TOF is based on the desorption of molecules from a solid surface layer and, therefore, a priori not compatible with LC. In an attempt to take advantage of the MALDI/TOF capabilities, a number of research groups carried out off-line LC separations and subjected the resulting fractions to MALDI/TOF measurements. Although this is laborious, it has the advantage that virtually any type of chromatographic separation can be combined with MALDI/TOF.

The different options for using MALDI/TOF as an off-line detector in LC have been discussed by Pasch and Rode.[116] In SEC of low molar mass samples the separation into individual oligomers and the quantitative determination of the MMD via an oligomer calibration could be achieved, see Figure 11 for oligo(caprolactone). The lower oligomers appeared as well separated peaks at the high retention time end of the chromatogram. For the analysis of the peaks, i.e. the assignment of a certain degree of polymerization (n) to each peak, MALDI/TOF/MS was used. The SEC separation was conducted at the usual analytical scale and the oligomer fractions were collected, resulting in amounts of 5–20 ng substance per fraction in tetrahydrofuran (THF) solution. The solutions were directly mixed with the matrix solution, placed on the sample slide and subjected to the MALDI experiments. As a large number of fractions may be introduced into the mass spectrometer at one time, sample preparation and MALDI/MS measurements take a very short period of time. In total, nine fractions were collected from SEC and measured by MALDI/MS. For the lower oligomers the spectra consisted of a number of peaks of high intensity, having a peak-to-peak mass increment of 114 Da, which equals the mass of the caprolactone repeating unit. These peaks represented the $M + Na^+$ molecular ions, whereas the peaks of lower intensity in their vicinity were due to the formation of $M + K^+$ molecular ions. $M + Na^+$ and $M + K^+$ molecular ions were formed due to the presence of small amounts of Na^+ and K^+ ions in the samples and/or the matrix. Further peaks of low intensity indicated a functional heterogeneity in the samples. From the masses of the $M + Na^+$ peaks the degree of polymerization of the corresponding oligomer was calculated. By this procedure, the first peak in the chromatogram was assigned to $n = 1$, the second peak to $n = 2$, and so on. From the elution time and the degree of polymerization of each oligomer peak an oligomer calibration curve of log molar mass vs elution time was constructed. The conventional calibration curve based on PS standards differed remarkably from this oligomer calibration curve.

A much more demanding task is the analysis of fractions from LC not only with respect to molar mass but also with respect to chemical structure. The separation of a technical fatty alcohol ethoxylate (FAE) by LC, under conditions where the chain length as well as the endgroups direct the separation, is presented in Figure 12. Using this chromatographic technique, the FAE was separated into three main fractions, the first fraction appearing as one peak at a retention time of about 60 s and the second and third fractions showing oligomer separations. Fraction 1 was collected in total, whereas for fractions 2 and 3 the individual oligomer peaks were collected. The MALDI/MS spectra of all three fractions gave a peak-to-peak mass increment of 44 Da, thus indicating that all fractions consisted of species with an ethylene oxide-based polymer chain. From the masses assigned to the peaks and the peak-to-peak mass increment of the ethylene oxide repeating unit the mass of the endgroup for the different fractions was calculated. Provided the sample was a pure FAE, the endgroups of fractions 1–3 could be identified as being polyethylene glycol

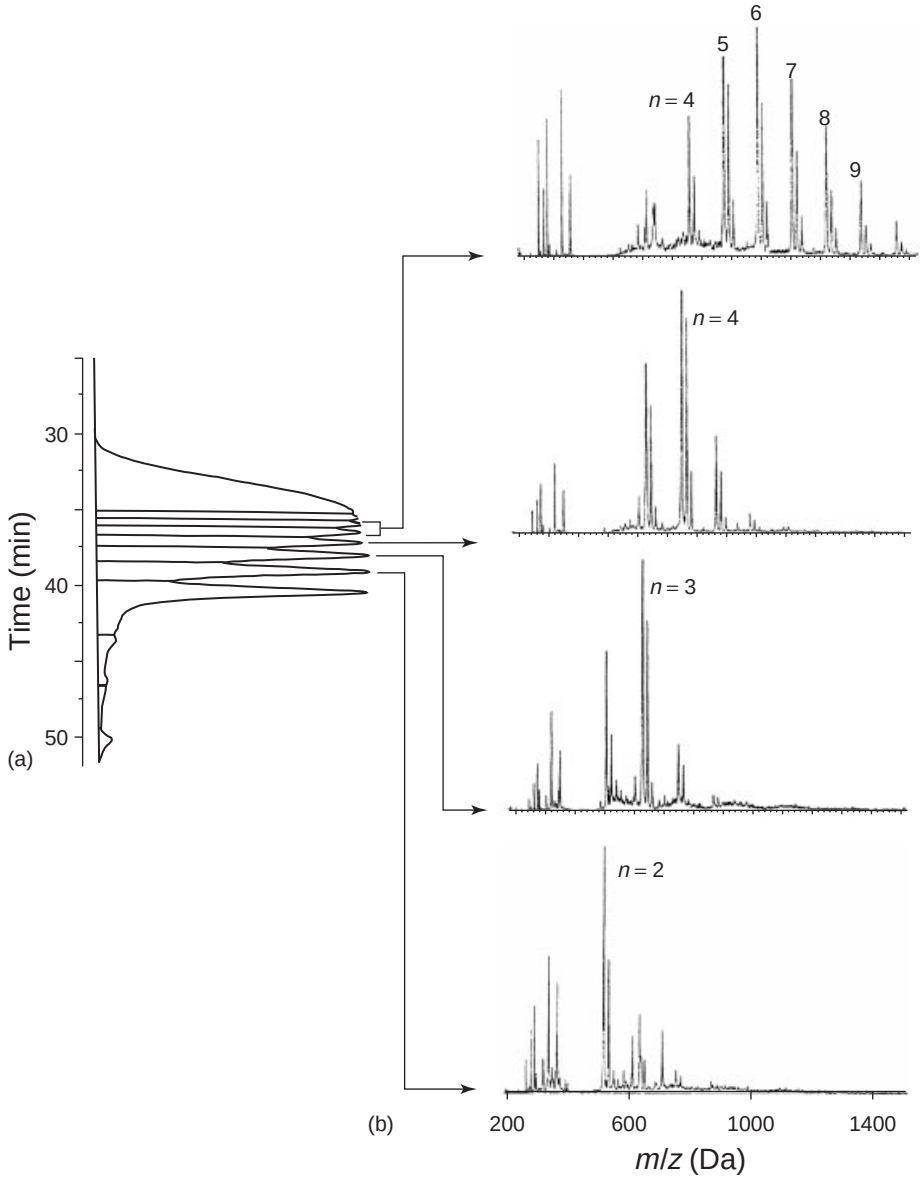

Figure 11 SEC of oligo(caprolactone) and MALDI/TOF analysis of fractions (a) and SEC calibration graphs (b). (Reproduced by permission from Pasch and Rode.[116])

(PEG) (α,w-dihydroxy endgroups), C_{13}-terminated PEO (α-tridecyl-w-hydroxy endgroups) and C_{15}-terminated PEO (α-pentadecyl-w-hydroxy endgroups), respectively. Using MALDI/TOF the oligomer distribution of the PEG fraction was measured directly. For fractions 2 and 3 by determining the degree of polymerization of the oligomer peaks oligomer calibration curves were obtained, which were used for the molar mass calculation of the fractions. Thus, by combining LC and MALDI/MS detection, complex samples can be analysed with respect to chemical structure and molar mass.

Other examples of successful off-line combinations of LC and MALDI/TOF were given by Krüger et al., separating linear and cyclic fractions of polylactides by LC/CC.[117] Just and Krüger were able to separate cyclic siloxanes from linear silanols and to characterize their chemical composition.[118] The calibration of an SEC system by MALDI/TOF was discussed by Montaudo et al.[119] Polydimethyl siloxane (PDMS) was fractionated by SEC into different molar mass fractions. These fractions were subjected to MALDI/TOF for molar mass determination. The resulting peak maximum molar masses were combined with the elution volumes of the fractions from SEC to give a PDMS calibration curve log M vs. V_e. The calibration of SEC by MALDI/TOF/MS for PMMA, polyvinyl acetate and

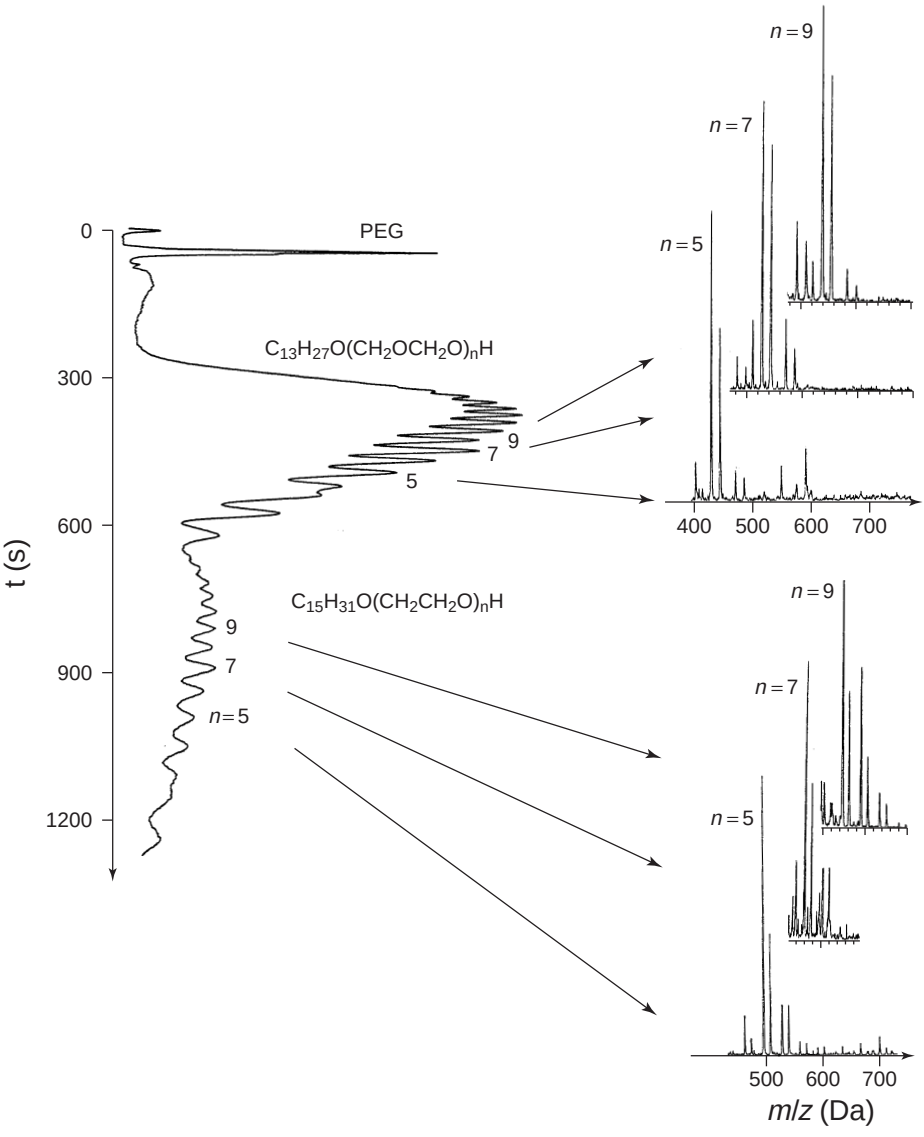

Figure 12 Separation of a technical PEO by LC and analysis of fractions by MALDI/TOF, peak assignment indicates degree of polymerization *n*. (Reproduced by permission from Pasch and Rode.[116])

vinyl acetate copolymers has been discussed by Danis et al. In addition to obtaining proper calibration curves, band broadening of the SEC system was detected.[120] The analysis of random copolyesters has been described recently by Montaudo et al.[121]

To overcome the difficulties of the off-line analysis of SEC fractions, recently interfaces were introduced where the SEC effluent was sprayed onto a moving matrix-coated substrate. Kassis et al. used a modified LC-Transform™ SEC/FTIR interface,[122] while Nielen applied a robotic interface of Bioanalytical Instruments where the effluent was spotted on the MALDI target.[123] A novel interface for coupling SEC and MALDI/TOF has been developed recently by Lab Connections Inc.[124] In this interface, the effluent from the SEC is sprayed through a heated capillary nozzle continuously on a slowly moving MALDI target precoated with the appropriate matrix, resulting in a uniform surface layer of sample fraction and matrix. The matrix can be deposited manually or automatically on the MALDI target from an appropriate solution. When necessary, a salt is added to the matrix solution.

The characterization of PMMA by SEC/MALDI/TOF is shown in Figure 13.[125] Prior to fraction deposition the target was precoated with the matrix dithranol and a small amount of LiCl to enhance the formation of $M + Li^+$ molecular ions. Since the fraction deposition was carried out through a heated capillary nozzle, a solid fraction/matrix film was obtained on the MALDI/TOF target. The MALDI/TOF target had a length of 70 mm

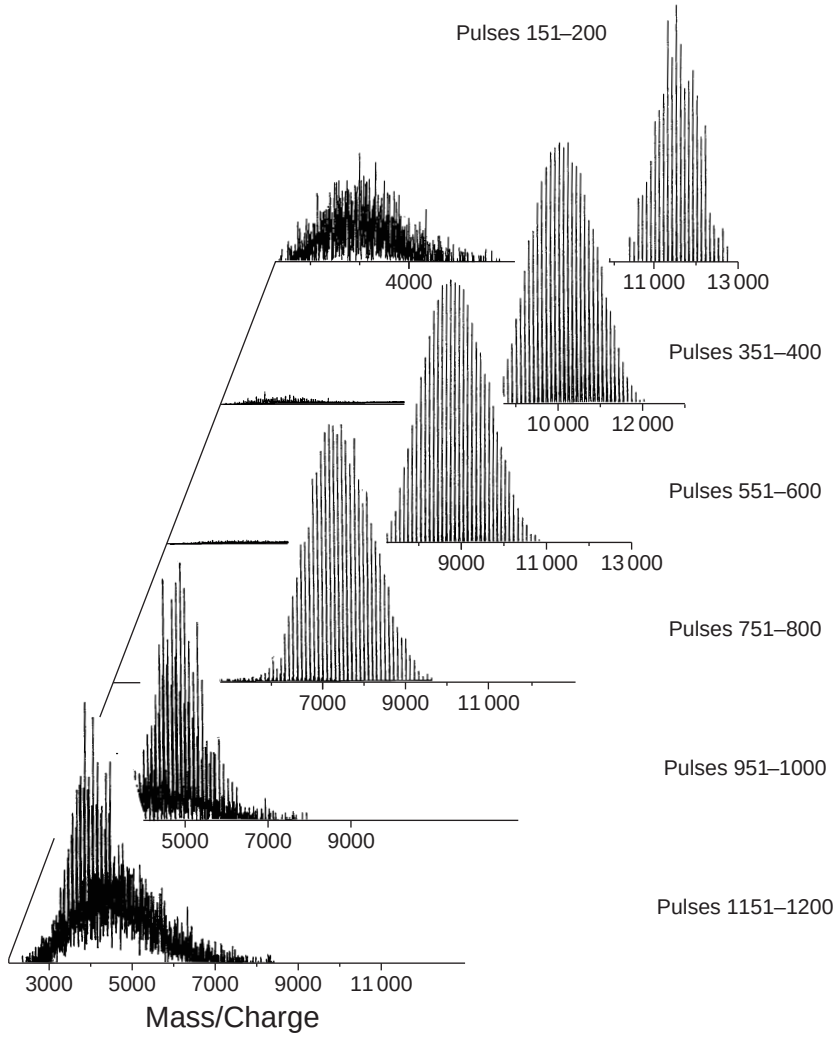

Figure 13 MALDI/TOF spectra of PMMA fractions obtained from SEC/MALDI/TOF analysis. (Reproduced by permission from Pasch.[125])

and was scanned continuously with 3500 laser pulses. Every 50 pulses were summarized to give a complete MALDI/TOF spectrum. With SEC as the preseparation technique, low positions on the target correspond to high molar masses, while high positions are equivalent to low molar masses. Selected spectra from different positions of the polymer/matrix track of the PMMA sample are given in Figure 13. In the present experiment, a sample amount of 10 µg (100 µL of a 0.1 mg mL^{-1} solution) was injected into the SEC. An amount of 10% of the total effluent was sprayed onto the MALDI target, resulting in a total amount of deposited sample of 1 µg. As can be seen, for all fractions high quality spectra were obtained giving the oligomer distributions of the different fractions.

Depending on the complexity of a specific sample, MALDI/TOF is more or less capable of resolving different chemical structures. While this technique is very powerful in determining different endgroups in macromonomers and telechelics, it has its limitations when it comes to analysing copolymers. Due to the fact that the number of possible oligomers increases exponentially with the degree of polymerization, even for low molar masses very complex product mixtures are obtained which cannot be analysed solely by MALDI/TOF. In these cases it is unavoidable to combine a chromatographic prefractionation with a MALDI/TOF analysis. The usefulness of such a combination shall be demonstrated for a diblock copolymer of n-butyl methacrylate and methyl methacrylate, i.e. poly-n-butyl methacrylate (PnBMA)-b-PMMA.

The sample under investigation was prepared by group transfer polymerization (GTP) resulting in structure **(1)** which follows and is typical.

Typical spectra for fractions of different molar masses obtained from the SEC/MALDI/TOF experiment are

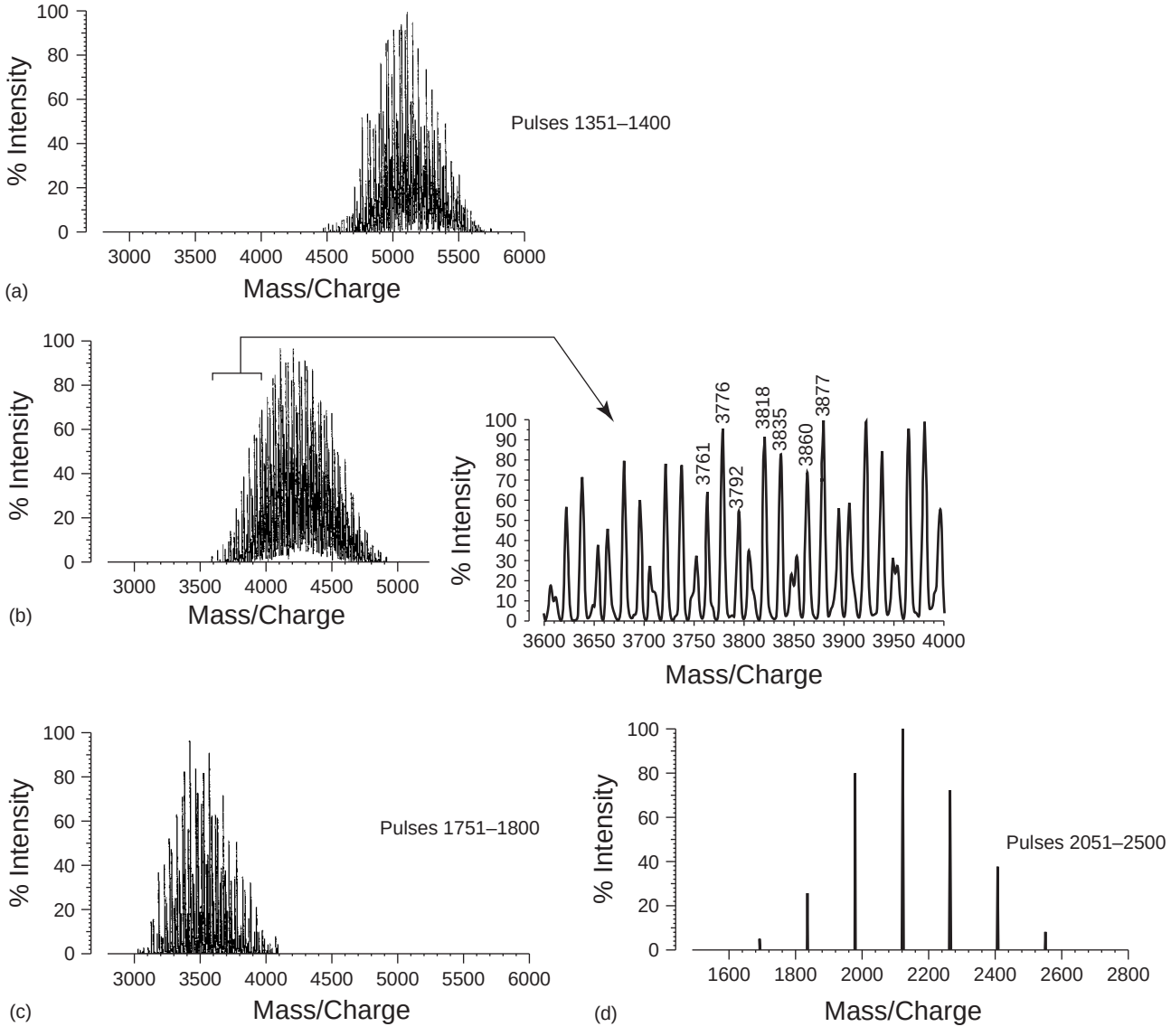

Figure 14 MALDI/TOF spectra of fractions obtained from SEC/MALDI/TOF analysis of a PMMA/PnBMA block copolymer. (Reproduced by permission from Pasch.[125])

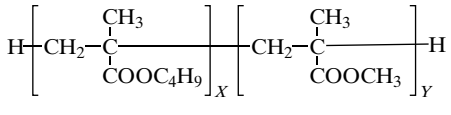

(1)

given in Figure 14. The higher molar mass fractions in Figure 14(a), (b) and (c) are characteristic for copolymer structures exhibiting typical mass increments of 100 Da for the MMA repeat unit and 142 Da for the nBMA repeat unit. Even these narrow disperse fractions exhibit a multitude of different mass peaks (usually more than 100) indicating the high complexity of the fractions. The lower molar mass fraction in Figure 14(d) is very uniform with respect to composition and thus differs from the molar mass fraction in Figure 14(a), (b) and (c). For the fraction in Figure 14(d), only peak-to-peak mass increments of 142 Da were observed which are typically for PnBMA. The chemical composition of the block copolymer was studied in detail by analysing the different mass peaks (see zoomed part of the spectrum in the insert of Figure 14b). Each peak in the spectrum could be assigned to one individual oligomer composition $(nBMA)_X(MMA)_Y$.

Considering the potential of MALDI/TOF in terms of versatility and sensitivity, the direct interfacing of LC and MALDI/TOF would be a highly attractive possibility. Given the experiences with the direct introduction of small matrix-containing liquid streams into high-vacuum

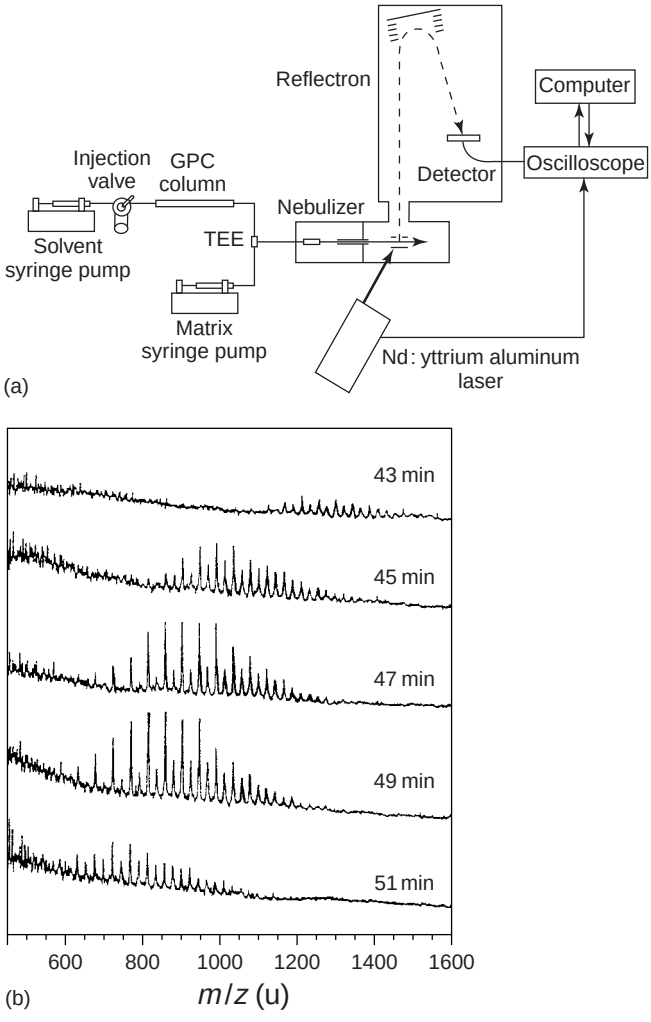

Figure 15 Aerosol MALDI apparatus configured for on-line SEC/MS (a) and five spectra obtained during the separation of PEG 1000 (b). GPC, gel permeation chromatography; TEE, T-shaped flow-splitter. (Reproduced by permission from Fei and Murray.[128])

instruments, it took surprisingly long before a device for liquid introduction to MALDI was described. In some recent papers Murray and Russell[126,127] and Fei and Murray[128] discussed the on-line coupling of SEC and MALDI/TOF/MS. In an aerosol MALDI/SEC experiment, the effluent from the SEC column was combined with a matrix solution and sprayed directly into a TOF/MS. Ions were formed by irradiation of the aerosol particles with pulsed 355 nm radiation from a frequency-tripled Nd:yttrium aluminum garnet laser. The ions were mass separated in a two-stage reflectron TOF instrument, and averaged mass spectra were stored every 11 sec throughout the SEC/MS experiment. Well-resolved MALDI/TOF spectra were obtained from commercial PEG 1000 and poly(propylene glycol) (PPG) 1000, see Figure 15.

3.4 Coupling with Fourier Transform Infrared Spectroscopy

When analysing a complex polymer, very frequently the first step must be the determination of the gross composition. Only when the chemical structures of the polymer components are known can sophisticated separation techniques such as gradient HPLC or LC/CC be adapted to a specific analysis.

The most frequently used techniques for a ''flash'' analysis are IR spectroscopy and SEC. IR spectroscopy provides information on the chemical substructures present in the sample, while SEC gives a first indication of the molar mass range. Information on both molar mass and composition is obtained when SEC or a comparable chromatographic method is combined with an IR detector. In the past, numerous workers have tried to use IR detection of the SEC column effluent in liquid flow cells. The problems encountered relate to obtaining sufficient signal-to-noise (S/N) ratio even with FTIR instruments, flow-through cells with minimum path lengths and mobile phases with sufficient spectral windows. Attempts to use FTIR detection with liquid flow-through cells and high performance columns have not been very successful due to the requirement of considerably less sample concentration for efficient separation.[129–136]

A rather broad applicability of FTIR as a detector in LC can be achieved when the mobile phase is removed from the sample prior to detection. In this case the sample fractions are measured in pure state without interference from solvents. Experimental interfaces to eliminate volatile mobile phases from HPLC effluents have been tried with some success[137–139] but the breakthrough towards a powerful FTIR detector was achieved only by Gagel and Biemann, who formed an aerosol from the effluent and sprayed it on a rotating aluminum mirror. The mirror was then deposited in an FTIR spectrometer and spectra were recorded at each position in the reflexion mode.[140–142]

Recently, Lab Connections Inc. introduced the LC-Transform™, a direct HPLC/FTIR interface based on the invention of Gagel and Biemann and discussed its first applications in polymer analysis.[143–145] The design concept of the interface is shown in Figure 16. The system is composed of two independent modules, the sample collection module and the optics module. The effluent of the LC column is split with a fraction (frequently 10% of the total effluent) going into the heated nebulizer nozzle located above a rotating sample collection disc. The nozzle rapidly evaporates the mobile phase while depositing a tightly focused track of the solute. When a chromatogram has been collected on the sample collector disc, the disc is transferred to the optics module in the

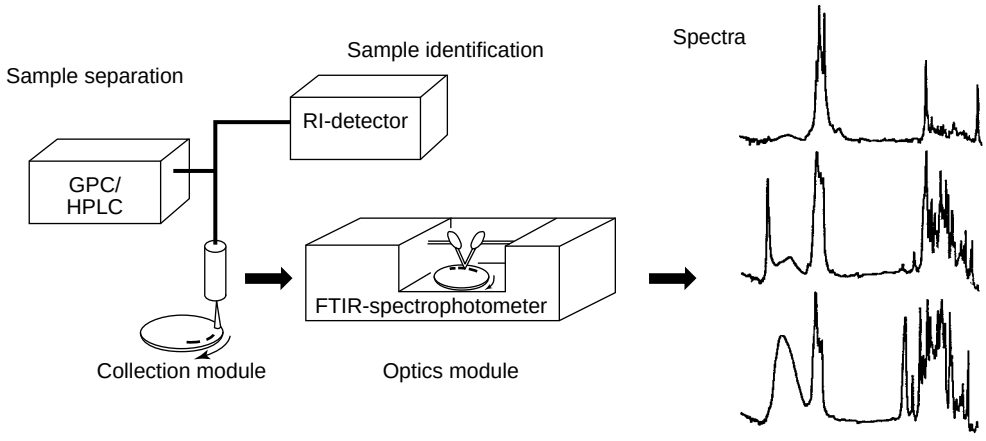

Figure 16 Schematic representation of the principle of coupled LC and FTIR spectroscopy.

FTIR detector for analysis of the deposited sample track. A control module defines the sample collection disc position and rotation rate in order to be compatible with the run time and peak resolution of the chromatographic separation. Data collection is readily accomplished with software packages presently used for GC/FTIR. The sample collection disc is made from germanium which is optically transparent in the range 6000–450 cm^{-1}. The lower surface of the disc is covered with a reflecting aluminum layer.

As a result of the investigation a complete FTIR spectrum for each position on the disc and, hence, for each sample fraction is obtained. This spectrum bears information on the chemical composition of each sample fraction. The set of all spectra can be arranged along the elution time axis and yields a 3-D plot in the coordinates elution time–FTIR frequency–absorbance.

One of the benefits of coupled SEC/FTIR is the ability to identify directly the individual components separated by chromatography. A typical SEC separation of a polymer blend is shown in Figure 17.[146] Two separate elution peaks 1 and 2 were obtained, indicating that the blend contained at least two components of significantly different molar masses. A quantification of the components with respect to concentration and molar mass, however, could not be carried out as long as the chemical structure of the components is unknown.

The analysis of the chemical composition of the sample was conducted by coupled SEC/FTIR using the LC-Transform™. After separating the sample with respect to molecular size, the fractions were deposited on the germanium disc and FTIR spectra were recorded continuously along the sample track. In total, a set of about 80 spectra was obtained which was presented in a 3-D plot, see Figure 18. The projection of the

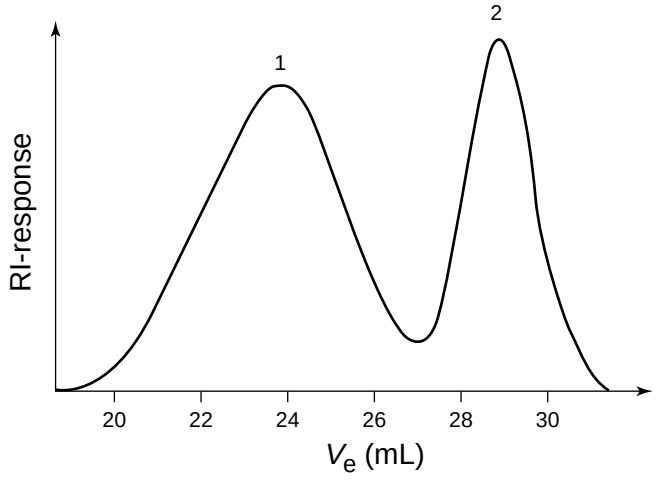

Figure 17 SEC separation of a binary blend, stationary phase: Ultrastyragel 2 × linear + 10^5 Å, eluent: THF.

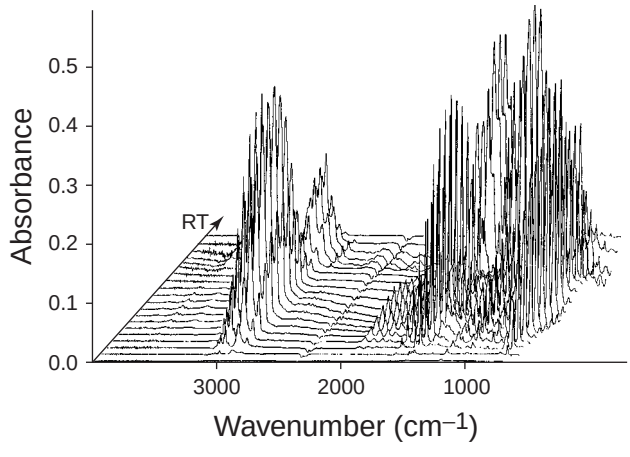

Figure 18 SEC/FTIR analysis of a binary blend, ''Waterfall'' representation.

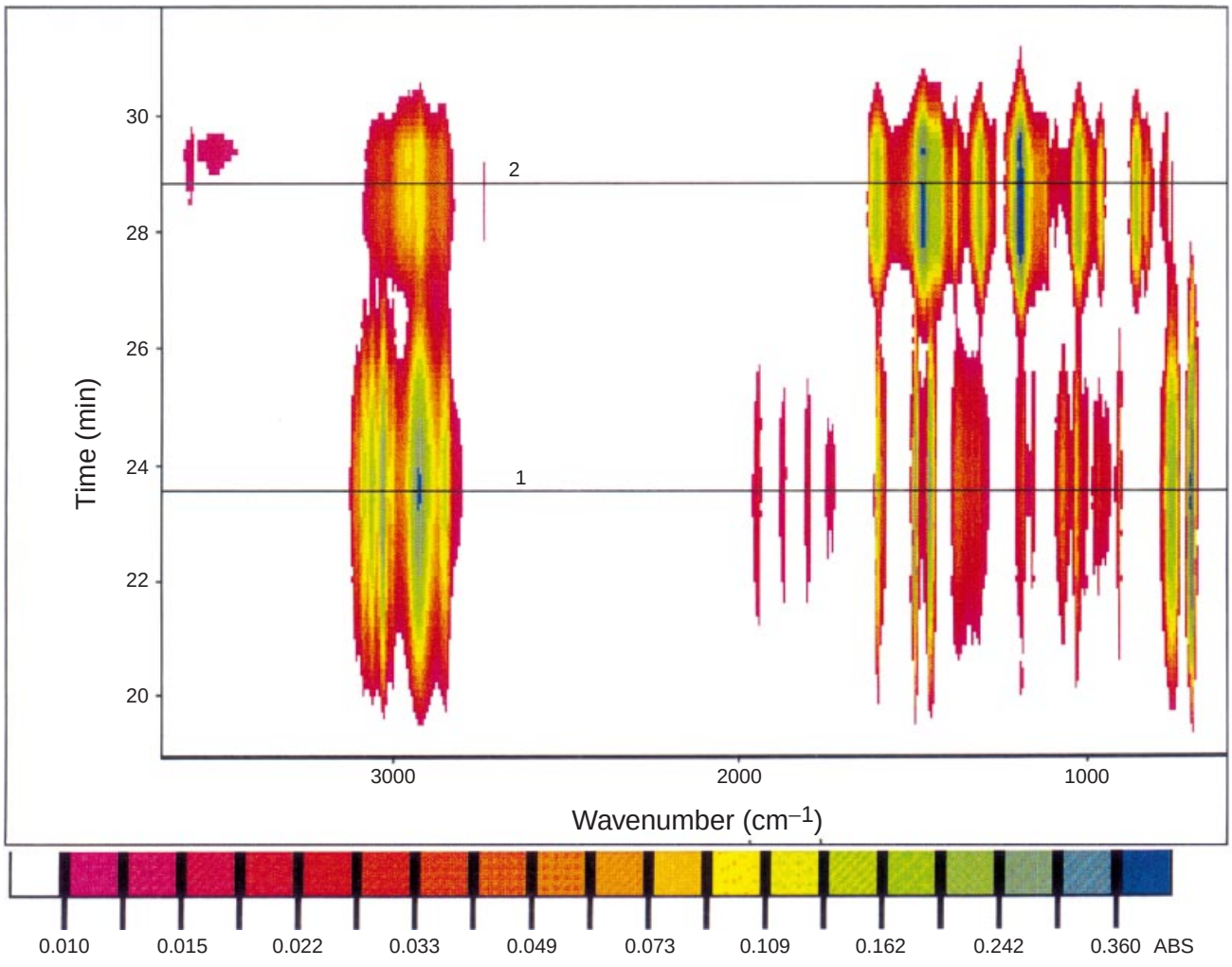

Figure 19 SEC/FTIR analysis of a binary blend, ''Contour plot'' representation.

3-D plot on the retention time–IR frequency coordinate system yielded a 2-D representation, where the intensities of the absorption peaks were given by a color code. Such a ''contour plot'' readily provides information on the chemical composition of each chromatographic fraction, see Figure 19. It was obvious that the chromatographic peaks 1 and 2 had different chemical structures. By comparison with reference spectra which are accessible from corresponding data bases, component 1 could be identified as PS, while component 2 was polyphenylene oxide. With this knowledge, appropriate calibration curves could be used for quantifying the composition and the component molar masses of the blend.

Coupled SEC/FTIR becomes an inevitable tool when blends comprising copolymers have to be analysed. Very frequently components of similar molar masses are used in polymer blends. In these cases the resolution of SEC is not sufficient to resolve all component peaks: see Figure 20 for a model binary blend containing an additive. The elution peaks of the polymer components 1 and 2 overlapped and, thus, the molar masses could not be determined directly. Only the additive peak 3 at the low molar mass end of the chromatogram was well separated and could be quantified.

A first indication of the composition of the present sample could be obtained from the contour plot in Figure 21. Component 3 showed typical absorption peaks of a phenyl benzotriazole and could be identified as a UV stabilizer of the Tinuvin type. Component 2 exhibited absorption peaks which were characteristic for nitrile groups and styrene units, while component 1 showed a strong ester carbonyl peak and peaks of styrene units. In agreement with the peak pattern of literature spectra, component 2 was identified as styrene–acrylonitrile (SAN) copolymer. Component 1 could have been a mixture of PS and PMMA or a styrene–methyl methacrylate copolymer. Since the FTIR

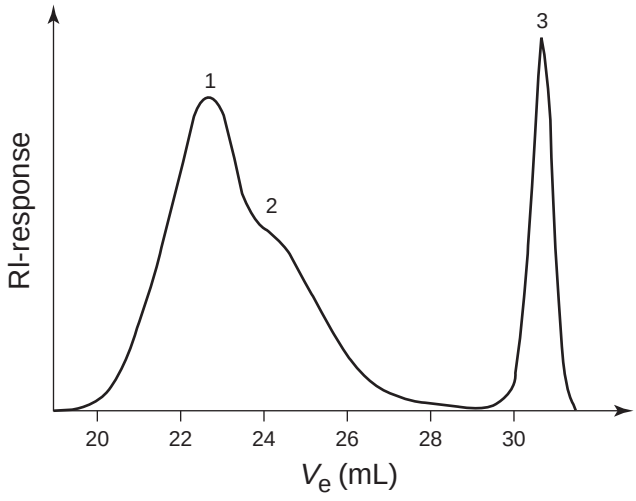

Figure 20 SEC separation of a blend of two copolymers and an additive, chromatographic conditions (see Figure 17).

spectra over the entire elution peak were uniform, it is more likely that component 1 was a copolymer.

One important feature of the SEC/FTIR software is that from the contour plot specific elugrams at one absorption frequency can be obtained. Taking the elugram at 2230 cm^{-1}, which is specific for the nitrile group, the elution peak of the SAN copolymer could be presented individually. For the presentation of component 1 the elugram at the carbonyl absorption frequency was drawn. Thus, via the ''chemigram'' presentation the elution peak of each component is obtained, see Figure 22.

In a relatively short period of time the LC-Transform™ system found its way into a large number of laboratories. Applications of the technique have been discussed in various fields. Willis and Wheeler demonstrated the determination of the vinyl acetate distribution in ethylene–vinyl acetate copolymers, the analysis of branching in high-density polyethylene, and the analysis of the chemical composition of a jet oil lubricant.[147] Provder et al.[148] showed that in powder coatings all additives were positively identified by SEC/FTIR through comparison of the known spectra. Even biocides could be analysed in commercial house paints. The comparison

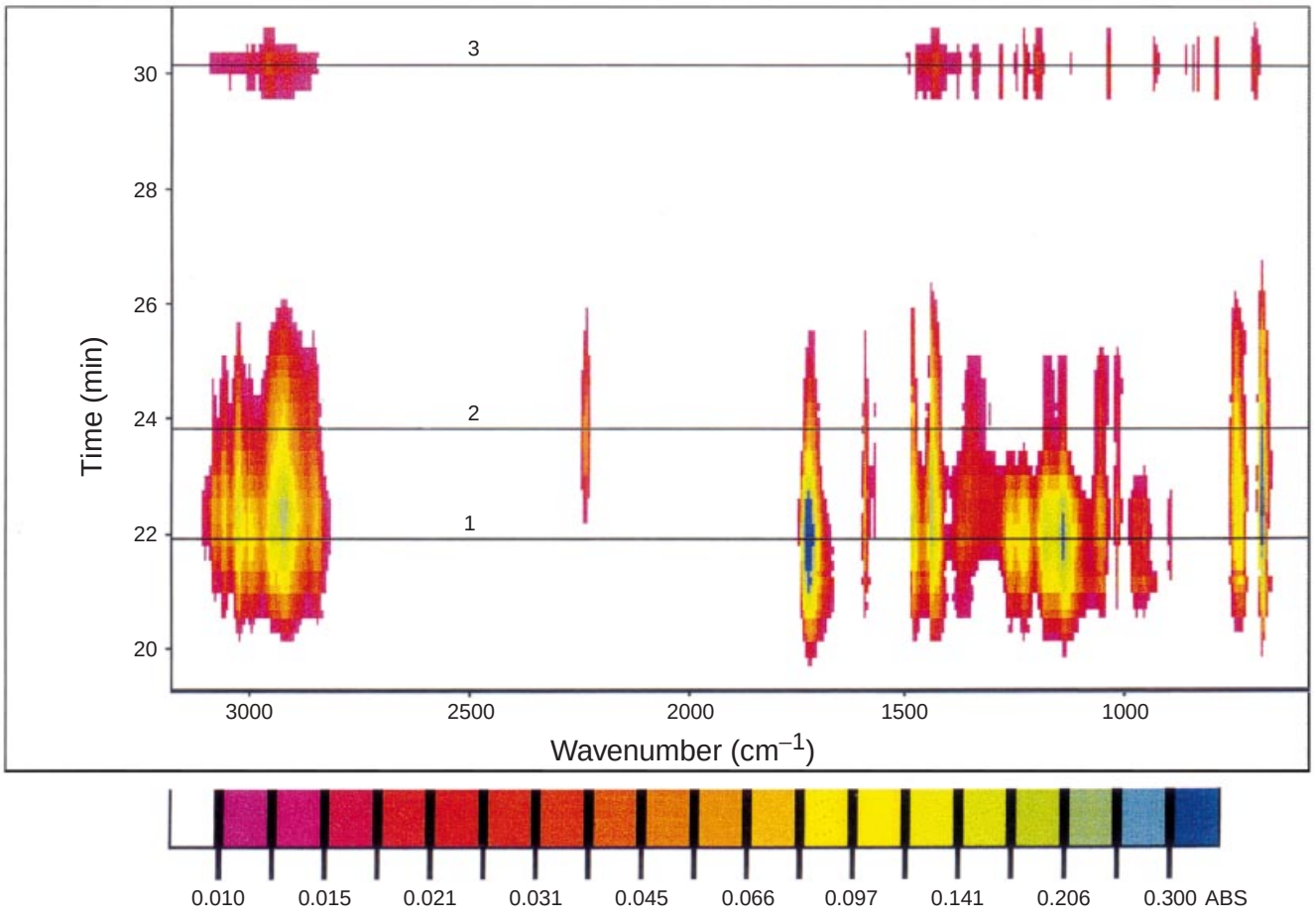

Figure 21 Contour plot of the SEC/FTIR analysis of a blend of two copolymers and an additive.

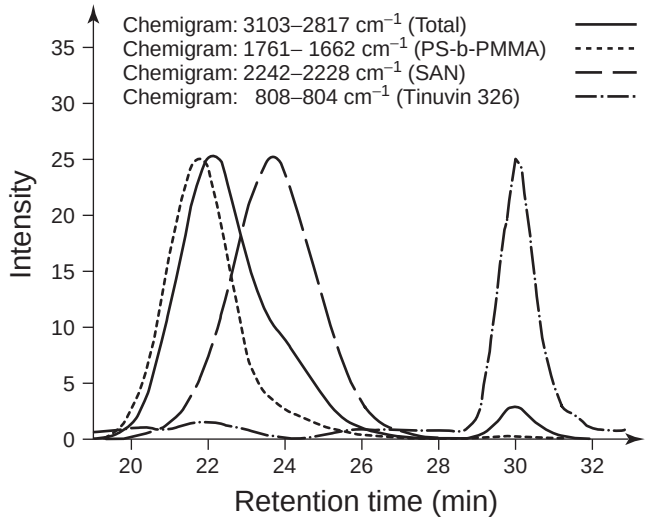

Figure 22 Chemigrams taken from the contour plot in Figure 21.

of a PS/PMMA blend with a corresponding copolymer gave information on the chemical drift. In the analysis of a competitive modified vinyl polymer sample by SEC/FTIR some of the components of the binder could be identified readily (vinyl chloride, ethyl methacrylate and acrylonitrile), and an epoxidized drying oil additive was detected.[148] The analysis of styrene–butadiene copolymers by combining interaction chromatography and FTIR has been demonstrated.[149,150] By using LC/CC it was possible to separate block copolymers and technical rubber mixtures with respect to chemical composition. The determination of the styrene : butadiene ratio and the fine structure of the butadiene units (*cis*/*trans*-, 1,2/1,4-units) was achieved by FTIR spectroscopy.

The quality of the results from SEC/FTIR strongly depend on the surface quality of the deposited sample fractions. Cheung et al. demonstrated that the surface-wetting properties of the substrate dominate the deposit morphology[151] and the spectra fidelity, film quality, resolution and polymer recovery were considered.[152] For different interface designs it was found that the morphology of the deposited polymer film was a key parameter for quantitative measurements.

3.5 Coupling with Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy is by far the most powerful spectroscopic technique for obtaining structural information about organic compounds in solution. Its particular strength lies in its ability to differentiate between most structural, conformational and optical isomers. NMR spectroscopy can usually provide all necessary information to identify unambiguously a completely unknown compound. The NMR detection technique is quantitative with individual areas in spectra being proportional to the number of contributing nuclei. The major drawback of NMR is the relatively low sensitivity in comparison to MS, another is the fact that structure elucidation of mixtures of unknown compounds with overlapping NMR signals is difficult and may be nearly impossible in cases with overcrowding signals in a small chemical shift region of the NMR spectrum. Therefore, in many cases it would be useful that a separation is performed prior to the use of NMR. For more efficient procedures, a direct coupling of separation with NMR detection would be the method of choice.[153]

The direct coupling of LC with proton NMR has been attempted numerous times. Early experiments of coupled HPLC-^{1}H-NMR were conducted in a stop-flow mode or with very low flow rates.[154–156] This was necessary to accumulate a sufficient number of spectra per sample volume in order to improve the S/N ratio. Other problems associated with the implementation of on-line HPLC/NMR have included the need for deuterated solvents. However, with the exception of deuterium oxide the use of deuterated eluents is too expensive for routine analysis. Therefore, proton-containing solvents such as acetonitrile (ACN) or methanol must be used. To get rid of the solvent signals in the spectra, the proton NMR signals of the solvents have to be suppressed.

Recent rapid advances in HPLC/NMR provide evidence that many of the major technical obstacles have been overcome.[157,158] With the development of more powerful NMR spectrometers combined with new NMR techniques for solvent suppression it became much easier to obtain well-resolved spectra in an on-flow mode. In particular, very efficient solvent-suppression techniques significantly improved the spectra during the HPLC/NMR run.[159,160] These techniques combine shaped radio frequency pulses, pulsed-field gradients, and selective ^{13}C decoupling to acquire high-quality spectra at on-flow conditions even with high HPLC gradients. Recently, even the direct coupling of supercritical fluid chromatography (SFC) with ^{1}H-NMR[161–163] together with the monitoring of supercritical fluid extraction[164] as well as the coupling of CE and ^{1}H-NMR[165–167] have been reported. An overview of the applications of on-line HPLC-^{1}H-NMR in organic chemistry was given by Albert.[153]

The first steps of polymer analysis into coupled LC-^{1}H-NMR were performed by Hatada et al.[168] They linked a size exclusion chromatograph to a 500 MHz proton NMR spectrometer and investigated isotactic PMMA. Using deuterated chloroform as the eluent and running the chromatography at a rather low flow-rate of 0.2 mL min^{-1} they were able to accumulate well resolved proton spectra. From the intensities of the proton signals of the endgroups and the monomer units they

determined the number-average molar mass across the elution curve. In further investigations they developed an absolute calibration method for direct determination of molar masses and MMDs by on-line SEC-^{1}H-NMR.[169] Ute reported on the chemical composition analysis of EPDM copolymers as a function of molar mass, and the monitoring of stereocomplex formation for isotactic and syndiotactic PMMA.[170]

The analysis of a technical PEO with respect to chemical composition and degree of polymerization has been performed by Pasch and Hiller.[171] This investigation has been conducted under conditions which are common for HPLC separations, i.e. sufficiently high flow-rate, moderate sample concentration, and on-flow detection. Using an octadecyl-modified silica gel as the stationary phase and an eluent of ACN–deuterium oxide 50:50 (v/v) the sample was separated into different functionality fractions, see Figure 23. The major fraction of the sample eluting between 14 and 25 min exhibited a partial oligomer separation.

For structural identification of the fractions, the ^{1}H-NMR spectrometer was directly coupled via capillary tubing to the HPLC system. The injection of the sample into the HPLC system was automatically initiated by the NMR console via a trigger pulse when starting to acquire NMR data. Using an appropriate pulse sequence, both solvent resonances (ACN at 2.4 ppm and water at 4.4 ppm) could be suppressed simultaneously. As a result of the on-line HPLC/NMR experiment a contour plot of ^{1}H chemical shift vs retention time could be generated, see Figure 24. Owing to the efficient solvent suppression, the obtainable structural information relates to the entire chemical shift region. From the contour plot, four different elution peaks could clearly be identified and analysed with respect to chemical composition. The remarkable feature of this investigation was that even the low concentration components in peaks 1–3 could clearly be identified in the contour plot.

Detailed structural information could be obtained from the individual NMR spectra of the fractions at the peak maximum, see Figure 25. This also gave the relevant structures (2 and 3). The first peak was identified as being PEG while the other fractions were alkylphenoxy PEOs. From the intensities of the endgroups and the ethylene oxide repeat units the average degree of polymerization for each fraction was calculated. Based on the total intensity distribution, a calculated chromatogram (or chemigram) was generated from the NMR contour plot. Comparing the real chromatogram (Figure 23) with the chemigram (Figure 24) an excellent agreement was obtained even recalling the oligomer separation pattern of the major fraction.

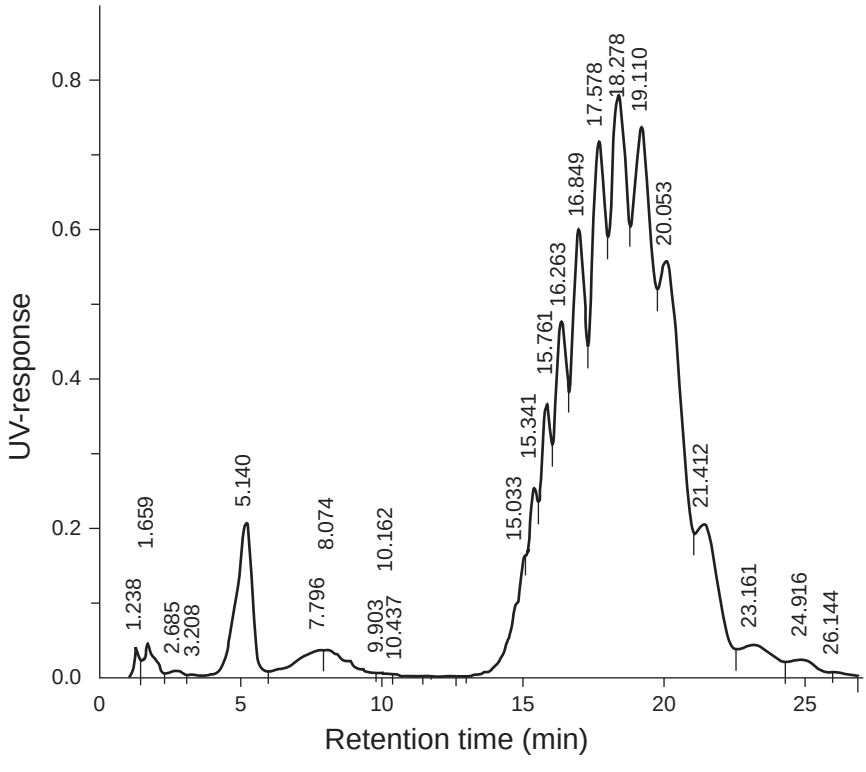

Figure 23 HPLC chromatogram of a technical PEO. Stationary phase: RP-18; eluent: ACN–deuterium oxide 50:50 (v/v). (Reproduced by permission from Schlotterbeck et al.[171])

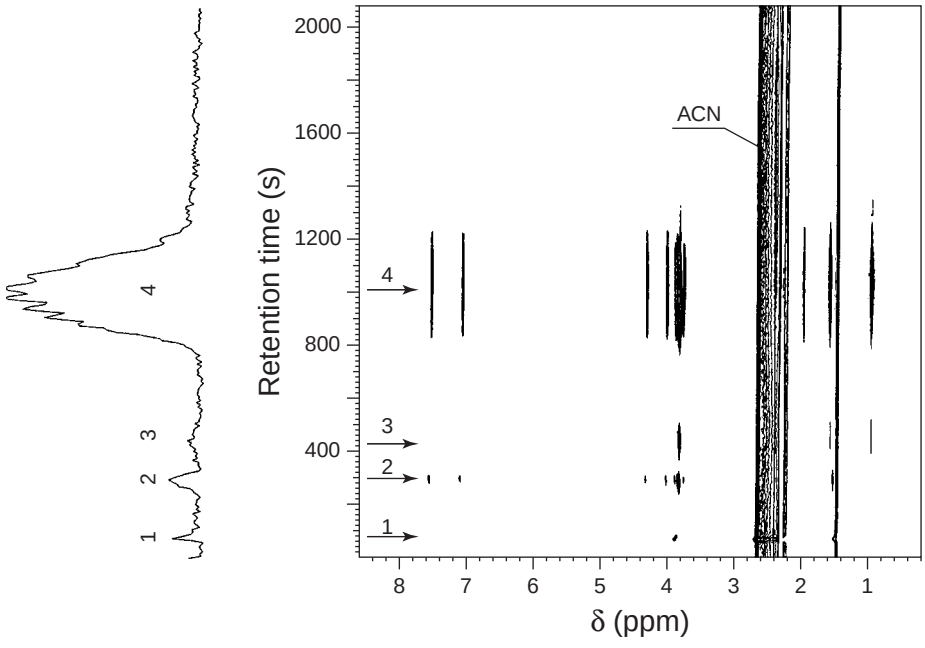

Figure 24 Contour plot of chemical shift vs. retention time and chemigram of the on-line HPLC/NMR analysis of a technical PEO. (Reproduced by permission from Pasch and Hiller.[171])

The analysis of FAE based surfactants by on-line HPLC-^{1}H-NMR has been described by Schlotterbeck et al.[172] Using a reversed stationary phase and ACN–deuterium oxide as the eluent, surfactant mixtures were separated with respect to the fatty alcohol endgroups. ^{1}H-NMR detection revealed the number of components, the chemical structure of the components, endgroups, and the chain length.

Finally, the investigation of the tacticity of oligostyrenes by on-line HPLC-^{1}H-NMR has been reported by Pasch et al.[173] The oligomer separation was carried out by hydrophobic interaction chromatography using isocratic elution with ACN on a reversed phase (RP) column RP-18. The chromatogram of an oligostyrene is shown in Figure 26. The first oligomer peak could be identified as being the dimer ($n = 2$), the next peak was identified as the trimer ($n = 3$) and, accordingly, the following peaks could be assigned to the tetramer, pentamer etc. The dimer peak appeared uniform, whereas for the following oligomers a splitting of the peaks was obtained. For $n = 3$ and $n = 4$ a splitting into two peaks was observed. For $n = 5$ and further, a splitting into three or more peaks occurred, which could be attributed to the formation of different tactic isomers.

The analysis of the isomerism of the oligomers by HPLC/NMR is given in Figure 27.[173] In this experiment conventional HPLC grade ACN was used as the eluent and no deuterium lock was applied. These conditions required high stability of the NMR instrument and a very efficient solvent-suppression technique since 100% ACN must be suppressed. The obtainable structural information related to the entire chemical shift region: however, residual signals of the eluent were obtained at 1.8–2.4 ppm and 1.3 ppm due to ACN and its impurities. The contour plot clearly revealed two signal regions, which could be used for analysis. These were the region of the methyl protons of the *sec*-butyl endgroup at 0.6–0.8 ppm and the aromatic proton region of the styrene units at 6.5–8.0 ppm. For the generation of the contour plot every 8 seconds a complete spectrum was produced by co-adding 8 scans. Accordingly, for the structural analysis 128 spectra were available over the entire retention time range. For the analysis of a separated oligomer, a minimum of four spectra could be used. These spectra bear selective information on the tacticity, even without completely separating the tactic isomers chromatographically.

As has been shown recently by Krämer et al., on-line coupled SEC-^{1}H-NMR can be used to monitor the chemical composition of random copolymers across the molar mass axis.[174] They investigated high-conversion poly(styrene-co-ethyl acrylate)s using dichloromethane as the solvent for SEC. The contour plot for a typical sample with a styrene ethyl acrylate (EA) ratio of 40 : 60 indicates that all characteristic spectral regions are accessible for analysis: see Figure 28. Residual solvent signals at 5.0–5.5 ppm do not overlap with resonances of the polymer molecules. The chemical composition across the elution peak of the copolymers is shown in Figure 29. The data were calculated from the peak

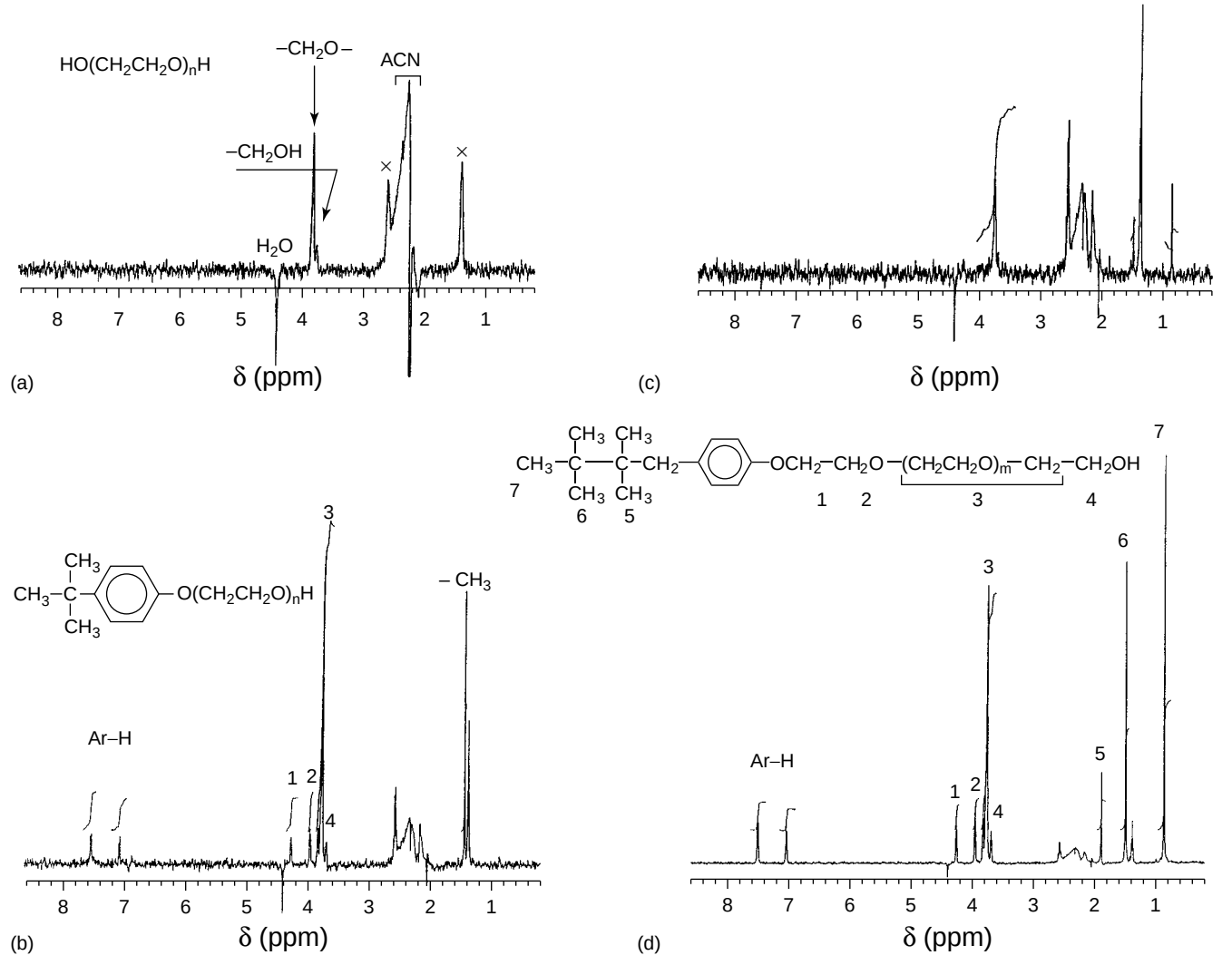

Figure 25 Individual fraction spectra taken from Figure 24. (a) NMR spectrum obtained from peak 1 in Figure 24; (b) NMR spectrum from peak 2 in Figure 24; (c) NMR spectrum from peak 3 in Figure 24; (d) NMR spectrum from peak 4 in Figure 24.

areas of the aromatic protons ($\delta = 6.5–7.3$ ppm) and the oxymethylene protons ($\delta = 3.5–4.2$ ppm) which were recorded during the SEC/NMR experiment. Owing to the low S/N ratio at the start and the end of the SEC elution curves, the chemical composition determination is less accurate than in the peak maximum. As can be seen, most of the copolymers exhibit constant chemical composition across the elution curves. This corresponds to the average chemical composition of the bulk sample. An exception is the EA-richest copolymer (styrene-EA 10.95), where at the high molar mass end of the elution curve an increased EA content is detected. At the same time, at the low molar mass end of the elution curve a strong tailing and an increased EA content is obtained. This strongly indicates that at high concentrations of EA in the monomer feed, copolymers with increased chemical and molar mass heterogeneity are obtained.

4 MULTIDIMENSIONAL LIQUID CHROMATOGRAPHY

4.1 Introduction

Despite the fact that substantial progress has been achieved in recent years in size-exclusion and interaction modes of polymer chromatography, the need and use for multidimensional separation systems has increased. The main reason for that is the fact that nowadays most classes of macromolecules posses property distributions in more than one parameter (e.g. molar mass *and* chemical composition at the same time). It is obvious that n independent properties require n-dimensional methods for accurate (independent) characterization of all those parameters. Moreover, the separation efficiency of any single separation method is limited by the efficiency and selectivity of

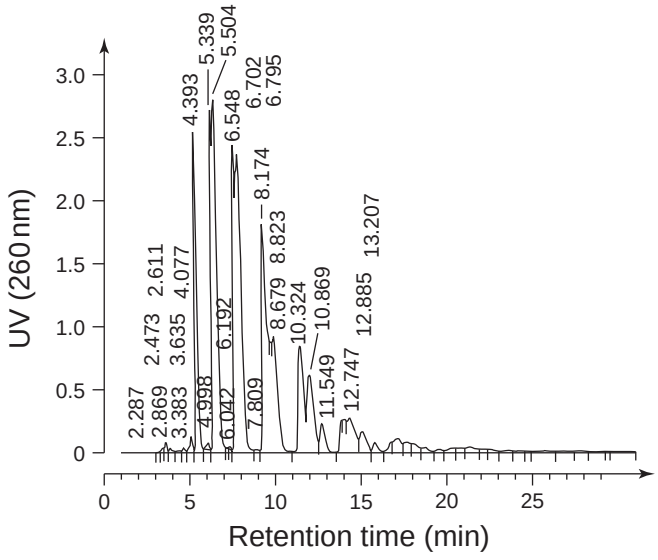

Figure 26 HPLC chromatogram of an oligostyrene PS 530. Stationary phase: RP-18; eluent: ACN. (Reproduced by permission from Pasch et al.[173])

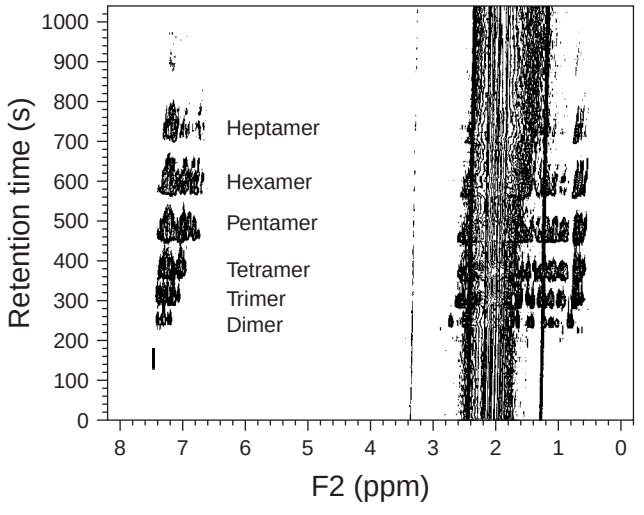

Figure 27 Contour plot of chemical shift vs retention time of the on-line HPLC/NMR analysis of PS 530. (Reproduced by permission from Pasch et al.[173])

the separation mode, i.e. the plate count of the column and the phase system selected. Adding more columns will not overcome the need to identify more components in a complex sample, due to the limitation of peak capacities. The peak capacity in an isocratic separation can be described, following Grushka,[175] as in Equation (19):

$$n = 1 + \frac{\sqrt{N}}{4} \ln \frac{V_p}{V_0} \tag{19}$$

The corresponding peak capacity in an n-dimensional separation is considerably higher due to the fact that

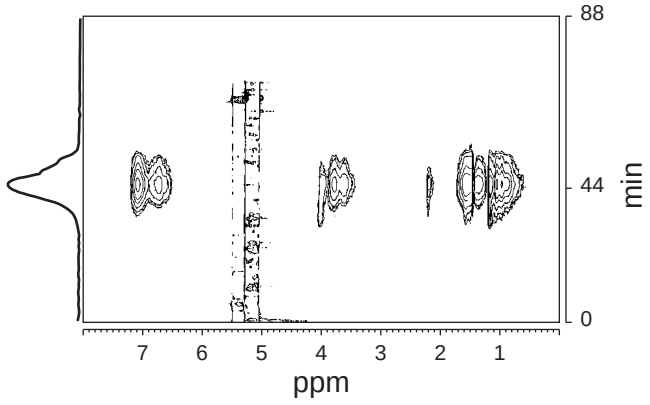

Figure 28 Contour plot of chemical shift vs retention time of the on-line SEC/NMR analysis of a random copolymer of styrene and EA. (Reproduced by permission from Krämer et al.[174])

each dimension contributes to the total peak capacity as a factor and not as an additive term for single dimension methods as described in Equation (20):

$$n_{\text{total}} = \quad n_i \sin^{(i-1)} \vartheta_i \tag{20}$$

where n_{total} represents the total peak capacity, n_i the peak capacity in dimension i and ϑ_i is the "angle" between two dimensions. The angle between dimensions is determined by the independence of the methods; a 90° angle is obtained by two methods, which are completely independent of each other and will, for example, separate two properties solely on a single parameter without influencing themselves.

In the case of a 2-D system the peak capacity is given by Equation (21):

$$\begin{aligned} n_{2D} &= n_1 n_2 \sin \vartheta \\ &= \left(1 + \frac{\sqrt{N_1}}{4} \ln \frac{V_{p,1}}{V_{0,1}} \quad \left(1 + \frac{\sqrt{N_2}}{4} \ln \frac{V_{p,2}}{V_{0,2}} \quad \sin \vartheta \right. \right. \end{aligned} \tag{21}$$

This effect is schematically illustrated in Figure 30.

Multidimensional chromatography separations can be done in planar systems or coupled-column systems. Examples of planar systems include 2-D thin-layer chromatography (TLC),[176,177] where successive 1-D TLC experiments are performed at 90° angles with different solvents, and 2-D electrophoresis, where gel electrophoresis is run in the first dimension followed by isoelectric focusing in the second dimension.[178–180] Hybrids of these systems where chromatography and electrophoresis are used in each spatial dimension were reported nearly 40 years ago.[181]

The main problem using planar methods is the difficulty in detection and collection of fractions among other

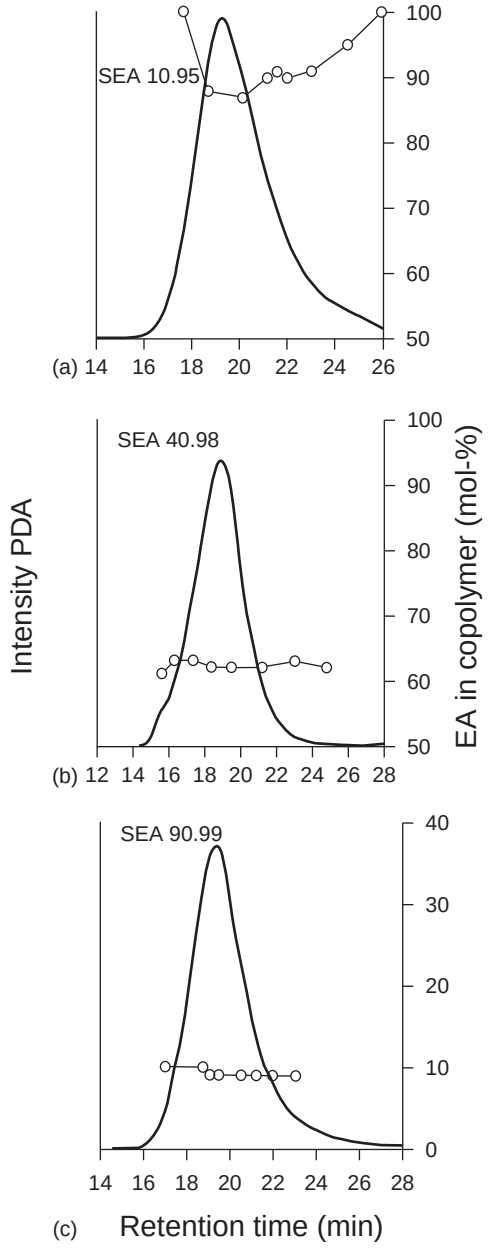

Figure 29 Chemical composition as a function of molar mass for random styrene–EA copolymers of different average composition. (Reproduced by permission from Krämer et al.[174])

less critical problems, such as homogeneous preparation of chromatographic media. However, the detection problem exists also for the coupled-column methods, mainly because of fraction dilution by each stage in a multidimensional separation system. Another aspect is the adjustment of chromatographic time bases between the different dimensions so that first dimension peaks may be sampled an adequate number of times by the next dimension separation system. This aspect has been recently studied in detail.[182]

In 2-D column chromatography systems an aliquot from a column or channel is transferred into the next separation method in a sequential and repetitive manner. Storage of the accumulating eluent is typically provided by sampling loops connected to an automated valve. Many variations on this theme exist which use various chromatographic and electrophoretic methods for one of the dimensions. In addition, the simpler ''heart cutting'' mode of operation takes the eluent from a first dimension peak or a few peaks and manually injects this into another column during the first dimension elution process. A partial compilation of these techniques has been given in several places.[182–190]

The use of different modes of LC facilitates the separation of complex samples selectively with respect to different properties like hydrodynamic volume, molar mass, chemical composition or functionality. Using these techniques in combination, multidimensional information on different aspects of molecular heterogeneity can be obtained. If, for example, two different chromatographic techniques are combined in a ''cross-fractionation'' mode, information on CCD and MMD can be obtained. Literally, the term ''chromatographic cross-fractionation'' refers to any combination of chromatographic methods capable of evaluating the distribution in size and composition of copolymers. An excellent overview on different techniques and applications involving the combination of SEC and gradient HPLC was published by Glöckner.[60]

In SEC mode the separation occurs according to the molecular size of a macromolecule in solution, which is dependent on its chain length, chemical composition, solvent and temperature. Thus, molecules of the same chain length but different composition may have different hydrodynamic volumes. Since SEC separates according to hydrodynamic volume, SEC in different eluents can separate a copolymer in two diverging directions. This principle of ''orthogonal chromatography'' was suggested by Balke and Patel.[191–193] The authors coupled two SEC instruments together so that the eluent from the first one flowed through the injection valve of the second one. At any desired retention time the flow through SEC 1 could be stopped and an injection made into SEC 2. The first instrument was operated with THF as the eluent and PS gel as the packing, whereas for SEC 2 polyether bonded-phase columns and THF–heptane were used. Both instruments utilized SEC columns. However, whereas the first SEC was operating so as to achieve conventional molecular size separation, the second SEC was used to fractionate by composition, utilizing a mixed solvent to encourage adsorption and partition effects in addition to size exclusion. Consequently, independent information on both MMD and CCD could not be obtained from such an experiment.

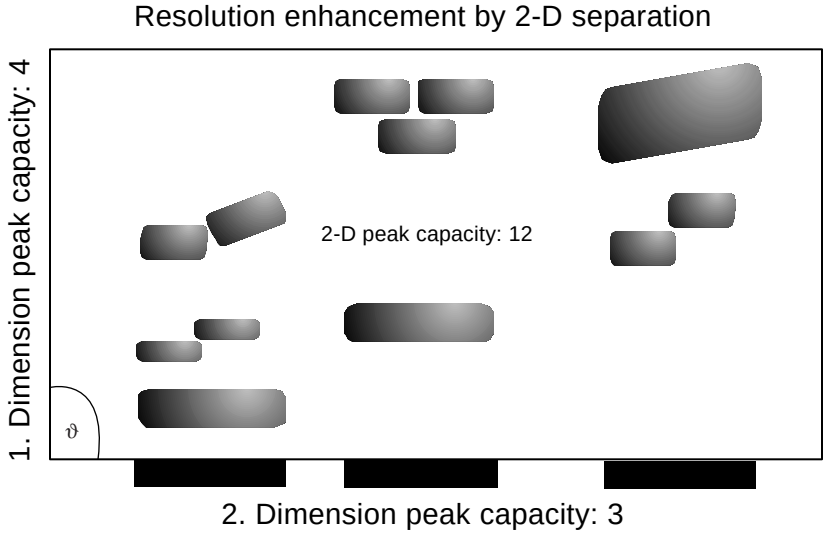

Figure 30 Schematic contour map representation of increased resolution and peak capacity in 2-D separations (peaks in each dimension are indicated by bars at the axes).

Since ''orthogonality'' requires that each separation technique is totally selective towards an investigated property, it seems to be more advantageous to use a sequence of methods, in which the first dimension separates according to chemical composition. In this way quantitative information on CCD can be obtained and the resulting fractions eluting from the first dimension are chemically homogeneous. These homogeneous fractions can then be analyzed independently in SEC mode in the second dimension to get the required MMD information. In such cases, SEC separation is strictly separating according to molar mass, and quantitative MMD information can be obtained.

Examples illustrating the potential of multidimensional separations will be given in section 4.5.

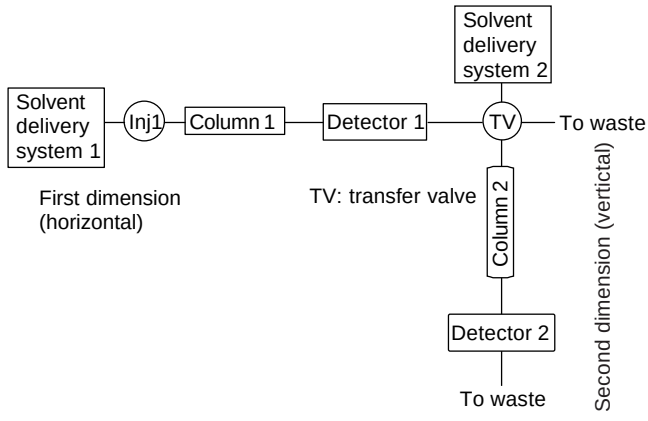

Figure 31 General experimental set-up for an 2-D chromatography system.

4.2 Experimental Aspects of Multidimensional Separations

Setting up a 2-D chromatographic separation system is actually not as difficult as one might think at first. As long as well-known separation methods exist for each dimension the experimental aspects can be handled quite easily in most cases. Off-line systems just require a fraction-collection device and something or someone who reinjects the fractions into the next chromatographic dimension. In on-line 2-D systems the transfer of fractions is preferentially done by automatic injection valves as was proposed by Kilz et al.[23,188,194] Figure 31 shows a general set-up for an automated 2-D chromatography system.

The focal point in 2-D chromatography separations is the transfer of fractions eluting from the first dimension into the second dimension. This can be done in various ways. The most simplistic approach is by collecting fractions from one separation and manually transferring them into the second separation system. Obviously, this approach is prone to many errors, labor intensive and quite time-consuming. A more efficient way of fraction transfer can be achieved by using electrically (or pneumatically) actuated valves equipped with two injection loops. Such a set-up allows one fraction to be injected and analyzed from one loop while the next fraction is collected at the same time in the second loop (see Figure 32). Total mass transfer from the first to the second dimension can be guaranteed by proper selection of flow rates in both dimensions.[195] This is a very beneficial situation as compared to heart-cut transfers, since by-products and trace-impurities can be separated even if they are not visible (VIS) in the first dimension separation. Table 3 shows a summary of potential fraction transfer options.

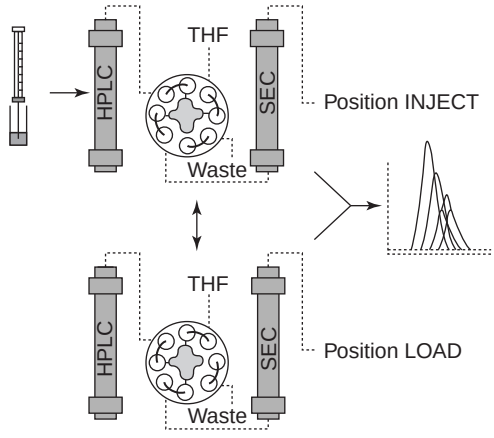

Figure 32 Fraction transfer between chromatography dimensions using a dual-loop 8-port valve.

There are some other important aspects which have to be considered for optimum 2-D experiment design.

4.2.1 Selection of Separation Techniques

Despite the fact that, historically, planar chromatography played an important role in multidimensional separations, this article will not discuss these aspects because they represent the past. The foreseeable future is with column-based techniques, which allow a well-controlled transfer of samples between different methods.

Obviously, destructive methods like GC and SFC, which destroy the chromatographic phase system, play a more limited role in multidimensional separations as they can only be used in the last separation step.

Schure recently published a theoretical paper[196] which discussed different chromatographic method combinations on the basis of efficiency, sample dilution and detectability. He investigated CE, GC, LC, SEC and field-flow fractionation (FFF) in detail, while omitting other methods, which are potential candidates for method hyphenation, such as SFC and temperature-rising elution fractionation (TREF).

Schure highlights several universal experimental factors (including plate count, injection volume, injected mass and injection band dilution), which should be taken into account when designing multidimensional separations. Table 4 summarizes Schure's results for the applicability of a given method in a multidimensional experiment. It is obvious that a low resolution, low injected mass method with high dilution of the

Table 3 Summary of 2-D transfer injection options

Transfer	Mode	Advantages	Disadvantages	Example
Manual	Off-line	Very simple Fast set-up	Time-consuming Not for routine work Not precise No correlation of fraction elution to transfer time Not quantitative	Test tube
Automatic	Off-line	Simple Easy Fast set-up	Less precise No correlation of fraction elution to transfer time Not quantitative	Fraction collector storage valve
Single-loop	On-line	Correct concentrations Correct transfer times Automation	Transfer not quantitative Set-up time	Injection valve (with actuation)
Dual-loop	On-line	Correct concentrations Correct transfer times Quantitative transfer automation	Set-up time Special valve	8-port actuated valve Combination of 2 conventional 6-port injection valves

Table 4 Synopsis of typical conditions and dilution factors in 1-D separations (Schure[196])

Separation mode	Experimental parameters				Result Dilution f
	dimensions (mm)	d_p (μm)	N (1 col^{-1})	V_{inj} (μl)	
GC	2500 × 0.25	n/a	125 000	1	26.2
LC	250 × 4.6	5	10 000	10	25.1
SEC	300 × 8.0	5	10 000	100	6.54
FFF	600 × 2 × 0.002	n/a	651	2	1340
CE	400 × 0.05	n/a	100 000	4.9e−4	12.7

Table 5 Calculated dilution factors for 2-D separations (Schure[196]) assuming splitless transfer injections (experimental conditions similar to those of Table 4)

1-D \ 2-D	GC	LC	SEC	FFF	CE[a]
GC	226	n/a	n/a	n/a	n/a
LC	217	141	113	4120	43.7
SEC	56.4	36.6	29.4	1070	11.4
FFF	11 600	7500	6030	220 000	2230
CE[b]	28	18.2	14.6	533	5.7

[a] Splitless transfer injections very difficult.
[b] Transfer concentrations very small.

injection band is a poor candidate for a multidimensional experiment.

Schure also calculated parameters to estimate the potential of 2-D method combinations. Results for splitless transfer injections are given in Table 5. CE, LC and SEC were rated best. This theoretical result agrees very well with experimental results and the actual number of published papers on 2-D separations. The most widely used method combination currently is that of LC with SEC.

4.2.2 Sequence of Separation Methods

This is an important aspect in order to get the best resolution and most accurate determination of property distributions. It is advisable to use the method with highest selectivity for the separation of one property as the first dimension. This ensures the highest purity of eluting fractions being transfered into the subsequent separation. In the case of gradient HPLC and GPC as separation methods, authors of early publications[191–193,197,198] used GPC as the first separation, because it took much longer than a subsequent HPLC analysis. This is not the best set-up, however, because the GPC fractions are only monodisperse in hydrodynamic volume, not in molar mass, chemical composition, etc. On the other hand, HPLC separations can be fine-tuned using gradients to fractionate only according to a single property, which can then be characterized for molar mass without any bias.

In many cases, interaction chromatography as the first dimension separation method is the best and most adjustable choice. From an experimental point of view, high flexibility is required for the first chromatographic dimension. In general, this is also more easily achieved when running the interaction chromatography mode in the first dimension, because:

(i) More parameters (mobile phase, mobile phase composition, mobile phase modifiers, stationary phase, temperature etc.) can be used to adjust the separation according to the chemical nature of the sample.

(ii) Better fine-tuning in interaction chromatography allows for more homogeneous fractions.

(iii) Sample load on such columns can be much higher as compared to SEC columns, for instance.

4.2.3 Detectability and Sensitivity in the Second Dimension

Because of the consecutive dilution of fractions, detectability and sensitivity become important criteria in 2-D experiment design. If byproducts and trace impurities have to be detected, only the most sensitive and/or selective detection methods can be employed. Evaporative light scattering detection (ELSD), despite several drawbacks, has been used mostly due to its high sensitivity for compounds which will not evaporate or sublime under detection conditions. Fluorescence and diode array UV/VIS are also sensitive detection methods, which can pick up samples at nanogram level. Mass spectrometers have a high potential in this respect too: however, they are currently not developed to a state where they would be generably usable.

Only in rare cases has RI detection, otherwise very popular in SEC, been used in multidimensional separations, because of its low sensitivity and strong dependence on mobile phase composition.

As a general rule, the higher the inject band dilution of a given separation method the more sensitive a subsequent detection method has to be. Such type of model calculations can be done easily; refer to Table 4 in section 4.2.1 and the paper by M. Schure[196] for further details.

4.2.4 Other Experimental Factors Affecting Multidimensional Separations

Depending on the specific type of the multidimensional experimental set-up, there are a number of other parameters to control and care about. Some are listed here, but because they are specific to the method combination, this list reflects only those techniques in most common use.

4.2.4.1 Influence of Eluent Transfer from First to Second Dimension A very important aspect in multidimensional chromatography design is the compatibility of mobile phases which are transferred between the different dimensions. It is a necessity that the mobile phases in two consecutive stages in multidimensional separations are completely miscible. Otherwise the separation in the second method is dramatically influenced and the fraction transfer is restricted or completely hindered. In gradient

systems, this requirement has to be verified for the total composition range.

In SEC separations the transfer of mixed mobile phases can affect molar mass calibration. In order to get proper molar mass results, the calibration curves have to be measured using the extremes of mobile phase composition and tested for changes in elution behavior and pore-size influence in the SEC column packing. The better the thermodynamic property of the SEC eluent, the less influence is expected on the SEC calibration when the transfer of mobile phase from the previous dimension occurs. It has been shown to be advantageous to use the SEC eluent as one component of the mobile phase in the previous dimension to avoid potential interference and mobile phase incompatibility.

4.2.4.2 Time Consumption Time is an important issue when designing multidimensional experiments. Set-up time itself plays only a lesser role, but the time needed for the multidimensional separations themselves can be considerable. This is especially true for 2-D separations using quantitative mass transfer via tandem-loop transfer valves. Heart-cut experiments require much less time and are often sufficient to check out the applicability of the approach. Cutting down on time consumption for multidimensional experiments is currently a heavily investigated topic. Several approaches are investigated and allow investigators to be optimistic and reduce experiment times by a factor of about 10 for complete mass transfer experiments using optimized column sets and flow conditions.

Another time requirement in multidimensional separations is that needed to reduce the amount of data and present them in an instructive way. With several dozen transfers between dimensions, data reduction and presentation can be very time-consuming and has been a real burden for those who performed the first cross-fractionation experiments.[191–193,198] There is a clear need for specialized multidimensional software which does all the data acquisition, fraction transfer, valve switching, data reduction, data consolidation and presentation of results. Currently, there is only one 2-D chromatography system commercially available[23,199] which is widely used. A few laboratories use in-house solutions, which are specific to their own chromatography and data capture hardware and specific also to result calculation and report creation.

4.3 Separation Techniques for the First Dimension

For an in-depth description of individual separation techniques used for multidimensional separations, please refer to the respective sections in this encyclopedia. This chapter deals with the specific aspects of the separation methods, which will help the reader to understand how to select one of them for a given multiple separation experiment.

4.3.1 Liquid Chromatography

This is the most often used technique for multidimensional separations. LC can be performed in normal phase or RP systems using isocratic or gradient elution. There is an abundance of stationary phases with different types of surface modifications of different polarities. This flexibility in experimental parameters is a very important consideration when using LC as a first dimension method, since it can be fine-tuned to separate according to a given property more easily than most other chromatographic techniques.

Gradient HPLC has been useful for the characterization of copolymers.[200–204] In such experiments careful choice of separation conditions is imperative. Otherwise, low resolution for the polymeric sample will obstruct the separation. On the other hand, the separation in HPLC, dominated by enthalpic interactions, perfectly complements the entropic nature of the SEC retention mechanism in the characterization of complex polymer formulations.

LC separation is based on an enthalpic interaction between the solute and the ''surface'' of the stationary phase. In pure interactive LC separations entropic contributions to the retention are absent. The distribution coefficient K_d can be derived from basic thermodynamics and can be measured from the activity of the analyte in the mobile (a_m) and stationary (a_s) phase as shown in Equation (22):

$$K_d(\mathrm{LC}) = \frac{a_s}{a_m} = e^{-\frac{\Delta H}{RT}} \tag{22}$$

The enthalpy change of the analyte corresponds to dispersion, polarization and charge-transfer interactions as well as H-bonding and ion exchange. Obviously, the distribution coefficient is larger than unity in LC separations.

The retention in pure LC separations can be calculated according to Equation (23):

$$V_e = V_0 + K_d(\mathrm{LC})V_{\mathrm{pore}} \tag{23}$$

The absence of entropic contributions to the separation is only possible if the stationary phase consists of nonporous beads or if the analyte molecules cannot penetrate into any pore in the stationary phase because of their size or interaction energy (e.g. ionic repulsion). In general, it will not be possible to avoid entropy changes in LC experiments with samples of different molar masses or sizes. In such cases it is best to select either a column which has very small or very large pores, which will force

the molecules to be excluded totally from the stationary phase or to permeate totally the pores of the packing. In both situations entropic contributions to the separation can be minimized.

In the general case, the distribution coefficient can be written as in Equation (24):

$$K_d = \frac{a_s}{a_m} = e^{-\frac{\Delta G}{RT}} = e^{-\frac{\Delta H}{RT}} e^{\frac{\Delta S}{R}} = K_d(\mathrm{LC})K_d(\mathrm{SEC}) \quad (24)$$

The elution in mixed-mode LC separations can be calculated according to Equation (25):

$$V_e = V_0 + [K_d(\mathrm{LC})K_d(\mathrm{SEC})]V_{pore} \quad (25)$$

This equation describes the general behavior of solutes in porous stationary phases. However, the dominant separation mechanism in LC with entropic contributions is the enthalpy term.

It is interesting to study the retention dependence on molar mass in LC separations. This is especially important when applying this technique to macromolecules, where chain statistics, chemical composition and molecular size play an important role in chromatographic behavior. Figure 33 shows the adsorption characteristics of molecules of different chain length or molar mass. The adsorption process is determined by the interaction enthalpy which itself is governed by active sites on the analyte molecule. If molecules are small only very few interactive sites are present, which may differ in nature and possess different interaction energies. In such cases, retention is relatively small (K_d less than about 10). Larger molecules, especially macromolecules, are composed of repeating units (usually called monomers), which can totally adsorb on the column packing. This is due to the fact that each and every repeating unit can potentially interact with the chromatographic surface. The longer the chains get, the higher the total interaction enthalpy and the higher the retention time to elude from the column (see Figure 33).

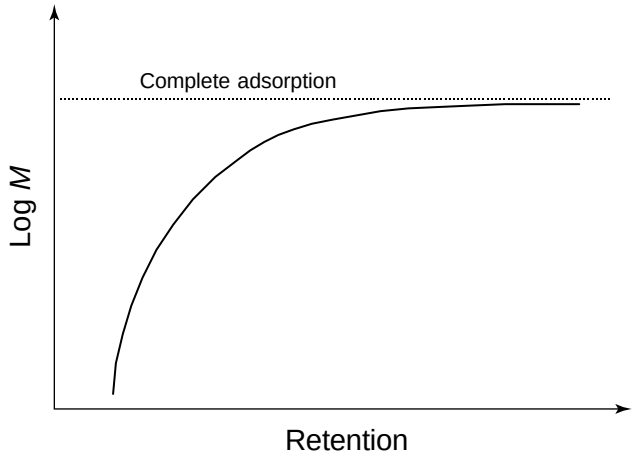

Figure 33 Adsorption of macromolecules of different molar mass on interactive LC columns.

The statistics of chain adsorption are determined by the magnitude of the entropy loss of the adsorbed macromolecule. If the entropy loss of the chain is small as compared to the adsorption enthalpy gain, the molecules will completely attach to the surface and the Gaussian chain will collapse into a 2-D layer. This scenario will dominate in cases where the solutes contain functional groups with strong interactions and where the eluent strength is relatively poor. In cases where the adsorption energy is relatively small, the entropy loss due to adsorption can be larger than the enthalpy gain. The chain only attaches selectively at the surface on the column packing forming loops. In such cases the macromolecules can desorb again. The adsorption/desorption process can be controlled by the nature of the stationary phase or even more easily by the thermodynamic properties of the eluents and gradient composition. Recently, temperature was also used to moderate adsorption behavior in chromatography.[205] Details of chain statistics in adsorption chromatography can be found in the book by Glöckner.[206]

4.3.2 Liquid Chromatography at the Critical Point of Adsorption

In LC and SEC chromatography modes either the enthalpy or the entropy dominate the separation. Several years ago Russian scientists found that homopolymers of different molar mass show exactly identical retention behavior on TLC plates[207,208] and on silica columns,[209] if a special eluent mixture was used for that macromolecule. They found that under ''critical'' conditions the sorbent did not ''see'' the polymeric nature of the chain. The separation was dependent only on the enthalpic interaction of the sample-sorbent pair (so-called ''critical chromatography'' or ''liquid chromatography at critical conditions''.[210–212]

The LC/CC mode relates to a chromatographic situation where the entropic and enthalpic interactions of the macromolecule and the column packing compensate each other, as shown in Equation (26):

$$\Delta G = \Delta H - T\Delta S = 0 \quad (26)$$

Therefore, we can derive Equation (27):

$$K_d = e^{-\frac{\Delta G}{RT}} = 1 \quad (27)$$

Rewriting this result for retention in elution volume terms, we directly get the experimentally observed result that the chromatographic peak position is independent

of the molar mass of the analyte and equal to the accessible volume of the stationary phase, which is stated in Equation (28):

$$V_e = V_0 + K_d(\text{LC/CC})V_{pore} = V_0 + V_{pore} = V_t \quad (28)$$

The Gibbs free energy of the macromolecule remains constant when it penetrates the pores of the stationary phase ($\Delta G = 0$). The distribution coefficient K_d is unity, regardless of the size of the macromolecules, and all macromolecules of equal chemical structure elute from the chromatographic column in one peak. The term ''chromatographic invisibility'' is used to refer to this phenomenon. This means that the chromatographic behavior is not directed by the size, but by the heterogeneities (chemical structure, branching point, endgroup, etc.) in the macromolecular chains.[207–210]

In general, as the Gibbs free energy is influenced by the length of the polymer chain and its chemical structure, contributions G_i for the polymer chain and G_j for the heterogeneity may be introduced, as stated in Equation (29):

$$\Delta G = \sum n_i \Delta G_i + \sum n_j \Delta G_j \quad (29)$$

For a perfectly uniform homopolymer chain the free energy change is determined by the contribution of the repeating units of the polymer chain (Equation 30):

$$\Delta G = \sum n_i \Delta G_i \quad (30)$$

At the critical point of adsorption of the polymer chain of a complex polymer, however, the contribution G_i becomes zero and chromatographic behavior is exclusively directed by imperfections in the macromolecular chain (Equation 31):

$$\Delta G = \sum n_j \Delta G_j \quad (31)$$

This chromatographic effect can be employed to determine imperfections in the polymer chain selectively and without any contribution by the repeating units themselves. LC/CC has been successfully used for the determination of the FTD of telechelics and macromonomers,[213–217] for the analysis of block copolymers,[218–220] macrocyclic polymers,[221] and polymer blends.[222–224]

Thus, LC/CC represents a chromatographic separation technique yielding fractions which are homogeneous in one property (e.g. chemical composition) but polydisperse in a different property (e.g. molar mass). These fractions can readily be analysed by SEC, which for chemically homogeneous fractions provides true MMDs without interference of CCD or FTD. Therefore, the combination of LC/CC and SEC in a 2-D chromatography experiment can be regarded as ''orthogonal'' chromatography in the strict sense provided that LC/CC is used as the first dimension separation mode. Consequently, for functional homopolymers being distributed in functionality and molar mass, coupling LC/CC with SEC can yield combined information on FTD and MMD. Such property information is important, e.g. for the quality control of amphiphilic polyalkylene oxides.

There is another area where the 2-D combination of LC/CC and SEC is extremely useful and can give results no other technique can provide in a single experiment. The 2-D separation of segmented copolymers (such as block- or graft- or comb-shaped copolymers) allows the complete molecular characterization of the copolymer with regard to individual segment molar masses and composition. It has been demonstrated that with such a set-up the polydispersities of copolymer segments can be determined independently.[225]

The thermodynamics of segmented copolymers is based on the idea that free energy change ΔG_{AB} of a segmented copolymer molecule, A_nB_m, is the sum of the contributions of segments A and segments B, ΔG_A and ΔG_B, respectively, which can be expressed by Equation (32):

$$\Delta G = \sum n_A \Delta G_A + \sum n_B \Delta G_B + \chi_{AB} \quad (32)$$

where χ_{AB} is the Flory–Huggins parameter describing the interactions between segments A and B. It has been demonstrated for a number of block copolymers[27,34] that no specific interactions between the heterosegments A and B ($\chi_{AB} = 0$) can be measured by chromatography. Using this assumption the change in the Gibbs free energy is solely dependent on the energy contributions of segments A and B, as is shown in Equation (33):

$$\Delta G = \sum n_A \Delta G_A + \sum n_B \Delta G_B \quad (33)$$

Applying experimental conditions, which correspond to the critical point of homopolymer A, the A segment in the segmented copolymer will become chromatographically invisible, as expressed in Equation (34):

$$\Delta G_A = 0 \quad (34)$$

Consequently, the retention of the segmented copolymer will be determined solely by the chromatographic properties of segment B. This is shown in Equation (35):

$$\Delta G_{AB} = \sum n_B \Delta G_B \quad (35)$$

This also means that the distributions coefficient for this system, K_d^{AB} can be reduced to the very simple term which is used for homogeneous molecules (Equation 36):

$$K_d^{AB} = K_d^B \quad (36)$$

Repetition of such an experiment using critical conditions for segment B allows the determination of molecular parameters for the other segment A in the copolymer. The same equations derived above apply, just the parameters for segments A and B are exchanged. Critical conditions for segment B mean that Equation (37) holds:

$$\Delta G_B = 0 \tag{37}$$

and the copolymer shows only the chromatographic behavior of segment A (Equation 38):

$$\Delta G_{AB} = \sum n_A \Delta G_A \tag{38}$$

The retention is then given by Equation (39):

$$K_d^{AB} = K_d^A \tag{39}$$

The characterization of ABA triblock copolymers can be done in a similar way to the analysis of diblock copolymers. The two possible cases for this type of investigation are:

(a) analysis with respect to the center block B using the critical conditions of the outer blocks A
(b) the analysis of the terminal A segments using the critical conditions of the center block B.

It is particularly useful to carry out experiments at the critical point of A. The separation occurs then with respect to the chain length of B, yielding fractions which are monodisperse with respect to B and polydisperse with respect to A and A′. These fractions can be analyzed selectively with respect to the outer blocks A and A′ in separate experiments.

4.3.3 Temperature-rising Elution Fractionation

This is not, strictly speaking, a chromatographic technique but it uses the same equipment and leads to detector traces which resemble chromatograms. This method is widely used for the short-chain branching characterization of polyolefins in the petrochemical industry.

Many polyolefins are in fact copolymers; the second comonomer introduces short-chain branches into the macromolecule. The copolymer properties depend strongly on the average composition and its sequence length distribution. In the case of polyethylene copolymers, changes in composition change the ease and temperature at which the polymer chains can crystallize. This property allows for the determination of composition distribution by measuring TREF. The TREF techniques rely on the fact that the redissolution temperature of a precipitated sample depends strongly on the number of short-chain branches. A chromatographic pump will transport the redissolved species through the column loaded with precipitated polymer into a concentration detector, which in turn will measure the concentration of the fraction redissolved at a given temperature, T, as shown in Figure 34.

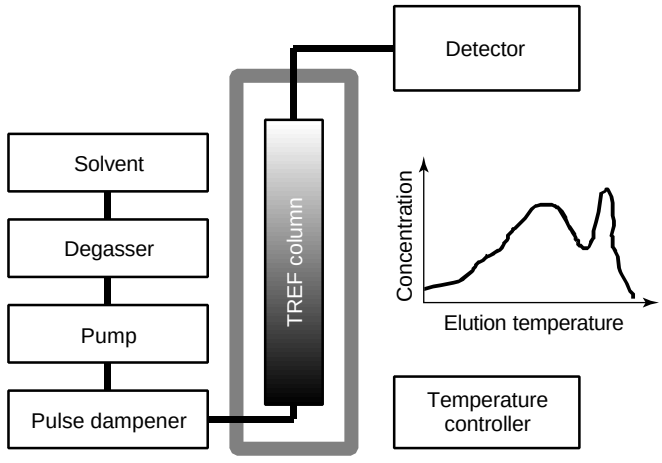

Figure 34 Schematic layout of a TREF instrument with thermostated TREF column.

4.4 Separation Techniques for the Second Dimension

The LC methods described in section 4.3 can also be used in the subsequent separation stage in a multidimensional chromatography set-up. However, as pointed out earlier (cf. section 4.2), it is advantageous to use gradient LC or LC/CC in the first separation dimension. On the other hand, SEC is preferentially used as the second method to retrieve molar mass information. In theory, TREF can also be used in a later separation step; however, for practical reasons, this is not advisable.

There are a number of chromatographic separation methods which can only be used in the last stage of a multidimensional experiment.

4.4.1 Size-exclusion Chromatography

SEC is described in the section of techniques for the second dimension, because of its primary benefits there (cf. section 4.2.2). However, it can also be used with lower efficiency and more biased results in the first separation dimension. This technique was developed for the separation and characterization of large molecules. It is also called Gel Filtration Chromatography for natural and biopolymers and known as GPC for synthetic polymers.

The principles are the same in both cases and rely on the fact that the macromolecule can only partially penetrate the porous packing, depending on its molecular size (not molar mass). The molecular size of a macromolecule in solution, or more accurately its hydrodynamic volume,

will depend on its chain length, the nature of the repeating units, chemical composition, molecular topology and the thermodynamic quality of the solvent. The same dependencies exist for the distribution coefficient in SEC, which can lead to the co-elution in SEC of species having identical hydrodynamic radius, but different composition, molecular architecture, and so forth. Large molecules can only access pores larger than the hydrodynamic radius of the molecule and will elute from the column early. If macromolecules are larger than the biggest pores, they will be totally excluded from the pore volume and will not be separated into fractions of different molar mass. On the other hand, the smallest molecules might be able to penetrate into all pores in the packing. They also will not be separated and will all co-elute at the total permeation volume. This chromatographic behavior is illustrated in Figure 35.

Similarly to the thermodynamic treatment of LC, retention in SEC can be described by basic thermodynamic entities and can be determined by measuring the concentration of the molecule in the stationary and mobile phases, as expressed in Equation (40):

$$K_d(\text{SEC}) = \frac{a_s}{a_m} = e^{\frac{\Delta S}{R}} \tag{40}$$

In ideal SEC separations the retention in the column is only governed by the entropy loss when the macromolecule enters the pore of the packing. No enthalpic interaction should be present in order to allow for accurate molar mass determinations.

The retention in ideal SEC experiments is given by Equation (41):

$$V_e = V_0 + K_d(\text{SEC})V_{\text{pore}} \tag{41}$$

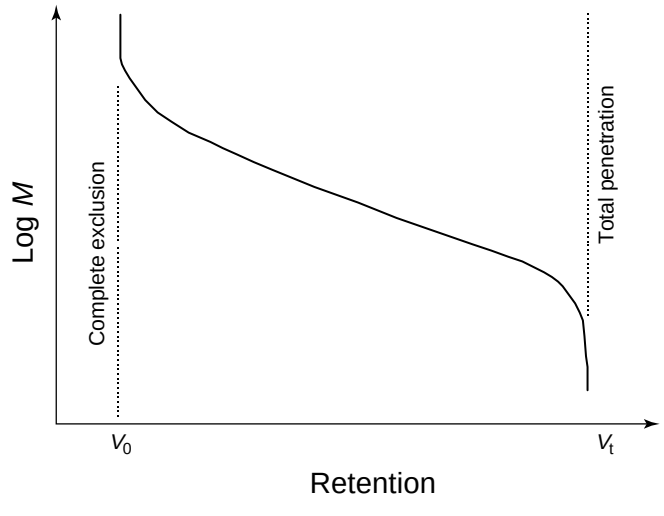

Figure 35 Size-exclusion behavior of macromolecules of different size on a porous column packing.

This equation looks very similar to Equation (23). The only difference is the magnitude of K_d, which is always less than unity for SEC separations and reaches a maximum value of unity for molecules which enter all pores.

Ideal SEC conditions are difficult to obtain for real macromolecules, however. As in the case of LC, there are contributions from both entropic and enthalpic terms to the distribution coefficient (cf. Equation 25).

In such cases, the retention for molecules with enthalpic interactions is higher and they will elute later from the column. This behavior can be described by Equation (42):

$$V_e = V_0 + [K_d(\text{SEC})K_d(\text{LC})]V_{\text{pore}} \tag{42}$$

In this equation the term $K_d(\text{LC})V_{\text{pore}}$ describes the delayed elution from the column as compared to the ideal case. This equation is mathematically identical to the respective equation for LC. However, the dominant parameter is $K_d(\text{SEC})$ and the entropy change governs the penetration of the analyte into the pores.

4.4.2 Capillary Electrophoresis

CE is a very efficient microseparation method (typically $N > 100\,000$), which uses a strong electric field to create an electro-osmotic flow in which the species will migrate. The reason for that is that the surface of the silicate glass capillary contains negatively-charged functional groups that attract positively-charged counterions. The positively-charged ions migrate towards the negative electrode and carry solvent molecules in the same direction. This overall solvent movement is called electro-osmotic flow. During a separation, uncharged molecules move at the same velocity as the electro-osmotic flow (with very little separation). Positively-charged ions move faster and negatively-charged ions move more slowly.

CE can also be used as a first set in multidimensional separation, but its practical use here is very limited due to the minute sample amounts injected.

For further information and details on the CE technique, please consult the articles in this encyclopedia.

4.4.3 Supercritical Fluid Chromatography

SFC is a relatively recent chromatographic technique which was commercialized in the early 1980s.

In SFC, the sample is carried through a capillary or packed column by a supercritical fluid (typically carbon dioxide). The properties of the mobile phase can be modified easily by polar additives and/or pressure programming, just as in gradient HPLC, to optimize selectivity. All three basic modes of chromatography (interaction, size-exclusion and critical conditions) have been verified in SFC separations.[226] SFC is a very

efficient separation technique, which has most of its applications in low molecular weight separations.

SFC has several advantages over conventional chromatographic techniques. SFC separations can be done considerably faster than HPLC separations, because the diffusion of solutes in supercritical fluids is about 10 times greater than that in liquids (and about 3 times less than in gases). This results in a decrease in resistance to mass transfer in the column and allows for fast high resolution separations. Compared with GC, capillary SFC can provide high resolution chromatography at much lower temperatures. This allows fast analysis of thermolabile and nonvolatile compounds. These advantages make SFC a good choice for multidimensional chromatography set-ups. Since SFC is a mobile phase-destroying technique, it can only be used in the last separation step in multidimensional separations.

4.4.4 Gas Chromatography

As with SFC, GC is a mobile phase-destroying technique and can only be used in the last stage of multidimensional separations. It also shares with SFC the high efficiency and speed of separations. However, it is limited to relatively low molecular weight compounds which are volatile without degradation and thermostable.

Most ''multidimensional'' applications reported in the literature use the first dimension for precleaning and removal of high molar mass species and not for complete characterization of the samples. Examples of such applications are the removal of humic acids from pesticides in soil extracts by SEC/GC and of high molecular byproducts from mono, di, and triglycerides.

For an indepth description of the technique, please refer to the *Gas Chromatography* in this encyclopedia.

4.5 State-of-the-art of On-line Coupled Two-dimensional Chromatography

This section will illustrate the current state and future potential of 2-D chromatography by reviewing separations published in the literature. Examples will be given for different separation techniques and combinations of these.

4.5.1 Two-dimensional Separations by Liquid Chromatography and by Size Exclusion Chromatography

Much work on chromatographic cross-fractionation was carried out with respect to the combination of SEC and gradient HPLC. In early experiments SEC was used as the first separation step, followed by HPLC. In a number of early papers the cross-fractionation of model mixtures was discussed. Investigations of this kind demonstrated the efficiency of gradient HPLC for separation by chemical composition. Mixtures of random copolymers of styrene and acrylonitrile were separated by Glöckner et al.[198] In the first dimension an SEC separation was carried out using THF as the eluent and PS gel as the packing. In total, about 10 fractions were collected and subjected to the second dimension, which was gradient HPLC on a CN bonded-phase using iso-octane/THF as the mobile phase. Model mixtures of random copolymers of styrene and 2-methoxyethyl methacrylate were separated in a similar way, the mobile phase of the HPLC mode being iso-octane/methanol in this case.[227] This procedure was also applied to real-world copolymers.[198] Graft copolymers of methyl methacrylate onto EPDM were analyzed by Augenstein and Stickler;[228] whereas Mori reported on the fractionation of block copolymers of styrene and vinyl acetate.[229] For all these experiments the same limitation with respect to the SEC part holds true: when SEC is used as the first dimension, true MMDs are not obtained.

From the theoretical point of view, a better copolymer separation set-up is prefractionation through HPLC in the first dimension and subsequent analysis of the fractions by SEC. HPLC was found to be rather insensitive towards molar mass effects and yielded very uniform fractions with respect to chemical composition.[230,231]

The major disadvantage of all early investigations on chromatographic cross-fractionation was related to the fact that both separation modes were combined with each other either off-line or in a stop-flow mode. Regardless of the separation order SEC vs HPLC or HPLC vs SEC, in the first separation step fractions were collected, isolated, and then subjected to the second separation step. This procedure, of course, is very time-consuming and the reliability of the results, at least to a certain extent, depends on the skills of the operator.

A fully automated 2-D chromatographic system was developed by Kilz et al.[23,188,194] The operation of the column-switching device is automatically driven by the software, which at the same time organizes the data collection from the detector.

One of the very few applications of 2-D gradient HPLC/SEC was published by Kilz et al. and described the analysis of styrene–butadiene star polymers.[188] Four-arm star polymers based on poly(styrene-*b*-butadiene) were prepared by anionic polymerization to give samples with well-known structure and molar mass control. In a first reaction step, a poly(styrene-*b*-butadiene) with a reactive chain end at the butadiene was prepared. This precursor reacted with a tetrafunctional terminating agent to give a mixture of linear (of molar mass M), 2-arm (2M), 3-arm (3M) and 4-arm (4M) species. Four samples with varying butadiene content (about 20%, 40%, 60%, and

80%) were prepared in this way. A mixture of these samples was used for the 2-D experiment. Accordingly, a complex mixture of 16 components, resulting from the combination of four different butadiene contents and four different molar masses (M, 2M, 3M, 4M) had to be separated with respect to chemical composition and molar mass.

Initially, the 16-component star block copolymer was investigated by SEC. As can be seen in Figure 36, four peaks were obtained. They correspond to the four molar masses of the sample consisting of species with one to four arms. The molar masses are defined by the number of arms and were in the ratio M–2M–3M–4M. Despite the high resolution, the chromatogram did not give any indication of the very complex chemical structure of the sample. Even when pure fractions with different chemical composition were investigated, the retention behavior did not show significant changes as compared to the sample mixture. In each case a tetramodal MMD was visible, indicating the different topological species. The SEC separation alone did not show any difference in chemical composition of the samples, which varied from 20% to 80% butadiene content.

Running the sample mixture in gradient HPLC mode gave poorly resolved peaks, which might suggest different composition, but gave no clear indication of different molar mass and topology, see Figure 37.

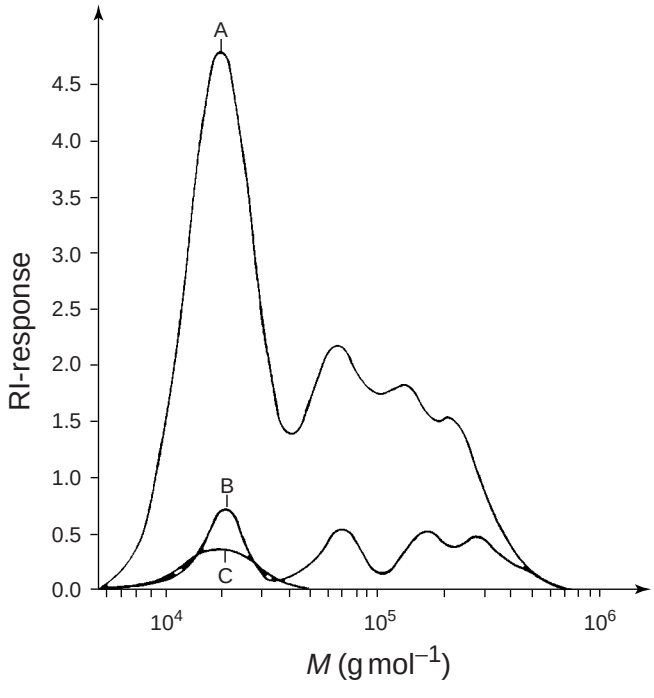

Figure 36 SEC chromatograms of the 16-component star block copolymer mixture (A) overlayed with styrene–butadiene 60:40 (B) star copolymer and 20:80 (C); peaks correspond to molar mass separation of different arm numbers.

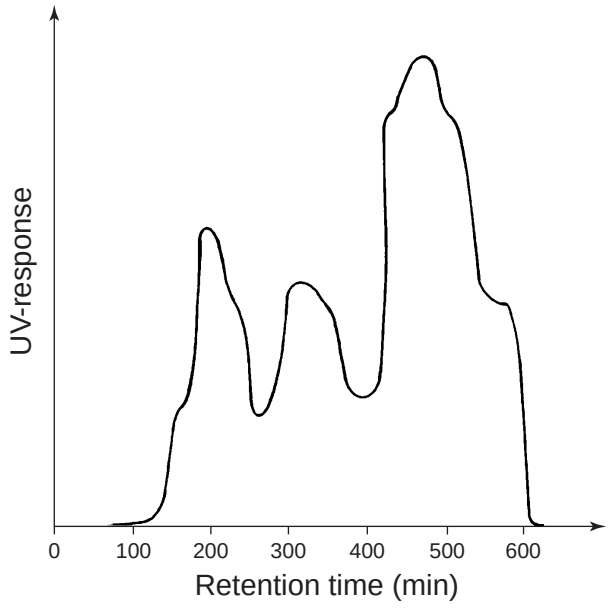

Figure 37 Gradient HPLC chromatogram of the 16-component star block copolymer mixture; separation is dominated by chemical composition and does not reveal the complexity of the sample.

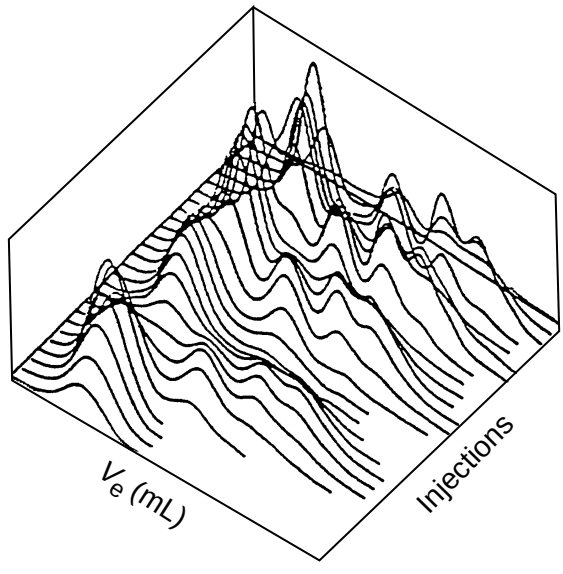

Figure 38 3-D view of the 16-component star block copolymer mixture; each trace represents an SEC chromatogram from a transfer injection.

The combination of the two methods in the 2-D set-up dramatically increased the resolution of the separation system and gave a clear picture of the complex nature of the 16-component sample. A 3-D representation of the gradient HPLC/SEC separation is given in Figure 38. Each trace represents a fraction transferred from HPLC to SEC and reflects the result

of the SEC analysis in the second dimension. Based on the composition of the sample, a contour map with the coordinates' chemical composition and molar mass is expected to show 16 spots, equivalent to the 16 components. Each spot would represent a component which is defined by a single composition and molar mass. The experimental evidence of the improved resolution in the 2-D analysis is given in Figure 39. This contour plot was calculated from experimental data based on 28 transfer injections.

The contour plot clearly revealed the broad chemical heterogeneity (*y*-axis, chemical composition) and the wide MMD (*x*-axis) of the mixture. The relative concentrations of the components were represented by colors. Sixteen major peaks were resolved with high selectivity. These correspond directly to the components. For example, peak 1 corresponds to the component with the highest butadiene content (80%) and the lowest molar mass (molar mass 1M) whereas peak 13 relates also to a molecule with 80% butadiene content but a molar mass of 4M. Accordingly, peak 16 is due to the component with the lowest butadiene content and a molar mass of 4M, representing a 4-arm star block copolymer with a styrene–butadiene content of 80 : 20.

A certain molar mass dependence of the HPLC separation is indicated by a drift of the peaks for components of similar chemical composition: see peaks 1-5-9-13, for example. This kind of behavior is normal for polymers, because pores in the HPLC stationary phase lead to size-exclusion effects which overlap with the enthalpic interactions at the surface of the stationary phase. Consequently, 2-D separations of this type will in general be not orthogonal but skewed, depending on the pore size distribution of the stationary phase and the nature of the sample. The quantitative amount of butadiene in each peak could be determined via an appropriate calibration with samples of known composition. The molar masses could be calculated based on a conventional molar mass calibration of the second dimension.

The mapping of ethoxylated fatty alcohols and ethylene oxide–propylene oxide block copolymers by 2-D chromatography was discussed by Trathnigg et al.[232] They

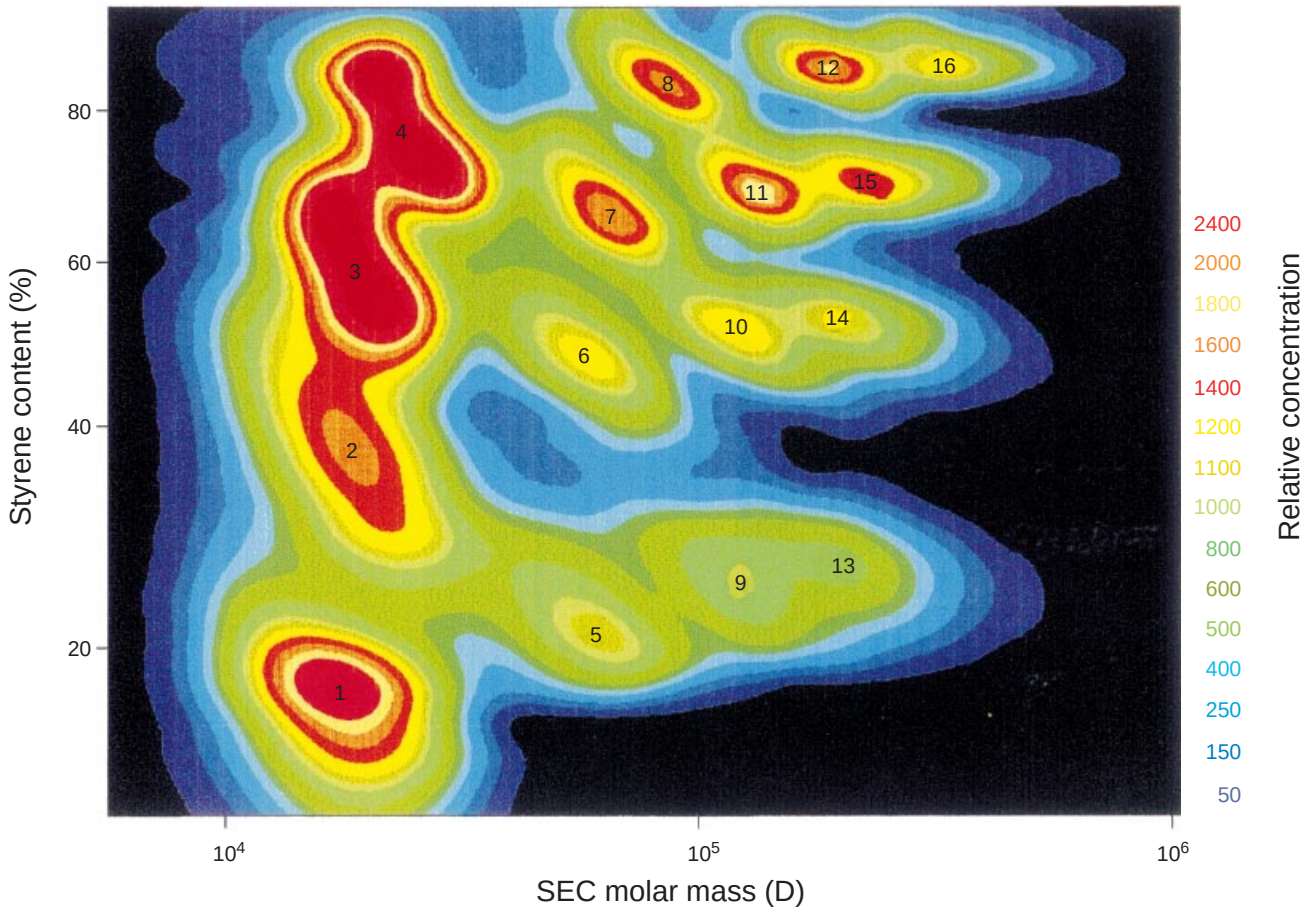

Figure 39 Contour plot of the 16-component styrene–butadiene star block copolymer mixture characterized by the 2-D HPLC/SEC separation.

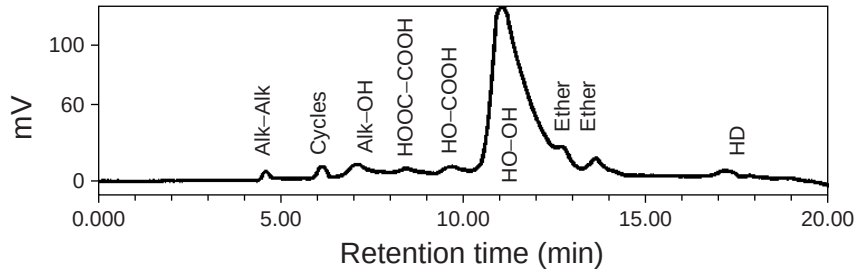

Figure 40 LC chromatogram of an aliphatic polyester sample at critical conditions; peak labels identify endgroups or functionality.

combined LAC and SEC and were able to determine the CCD and the MMD of the polyethers.

4.5.2 Two-dimensional Separations by Liquid Chromatography at Critical Conditions and Size Exclusion Chromatography

The analysis of aliphatic polyesters with respect to FTD and MMD has been demonstrated.[188,233] Polyesters from adipic acid and 1,6-hexanediol are manufactured for a wide field of applications with an output of thousands of tonnes per year. They are intermediates for the manufacture of polyurethanes, and their FTD is a major parameter affecting the quality of the final products. In particular, nonreactive cyclic species are responsible for the ''fogging effect'' in polyurethane foams.

For the separation of the polyesters with respect to functionality, LC/CC was used, the critical point of adsorption of the polymer chain corresponding to an eluent composition of acetone–hexane 51 : 49 (v/v) on silica gel. The critical chromatogram of a polyester sample together with the functionality fraction assignment is given in Figure 40. The ''ether'' peaks could be attributed to the formation of ether structures in the polyester samples by a condensation reaction.

The MMDs of the functionality fractions could be determined by preparatively separating the fractions and subjecting them to SEC. The SEC chromatograms of fractions 1–9 are summarized in Figure 41. For a number of fractions oligomer separations were obtained, which could be used to calibrate the SEC system.

The very complex nature of the sample could be verified in a 2-D experiment using LC/CC as the first dimension to separate for functionality and SEC in the second dimension to determine molar masses. The contour plot in Figure 42 reveals the structural complexity of the sample including the functionality fractions from LC/CC and the oligomer separations from SEC which were well-recognizable. The sample was prepared from an adipic acid-rich reaction mixture resulting in an acid number of about 5. The high content of dicarboxylic acid endgroups is clearly reflected in the contour map. Quantification of

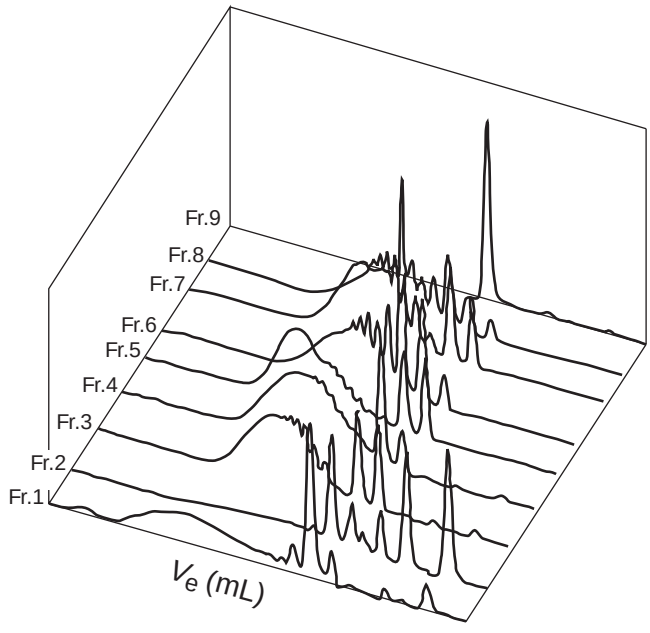

Figure 41 SEC separation of different fractions taken from the LC/CC separation of the aliphatic polyester sample (cf. Figure 40).

the contour plot yielded quantitative information on both FTD and MMD.

The analysis of an octylphenoxy-terminated PEO with respect to FTD and MMD has been demonstrated by Adrian et al.[225] A separation of the sample with respect to the terminal groups could be achieved by LC/CC on an RP-18 stationary phase and a critical eluent composition of methanol–water 86 : 14 (v/v). All peaks in the chromatogram could be identified by MALDI/TOF MS as pure fractions of different functionality, proving that the separation followed the chemical structure of the endgroups. The combination of LC/CC with SEC in a 2-D chromatography experiment resulted in the contour plot shown in Figure 43.[225] At the abscissa, the retention volume of the SEC runs (second dimension) is given, whereas the ordinate gives the retention volume of the LC/CC (first dimension). Relative concentrations are mapped to a color code on a log scale to make small quantities visible.

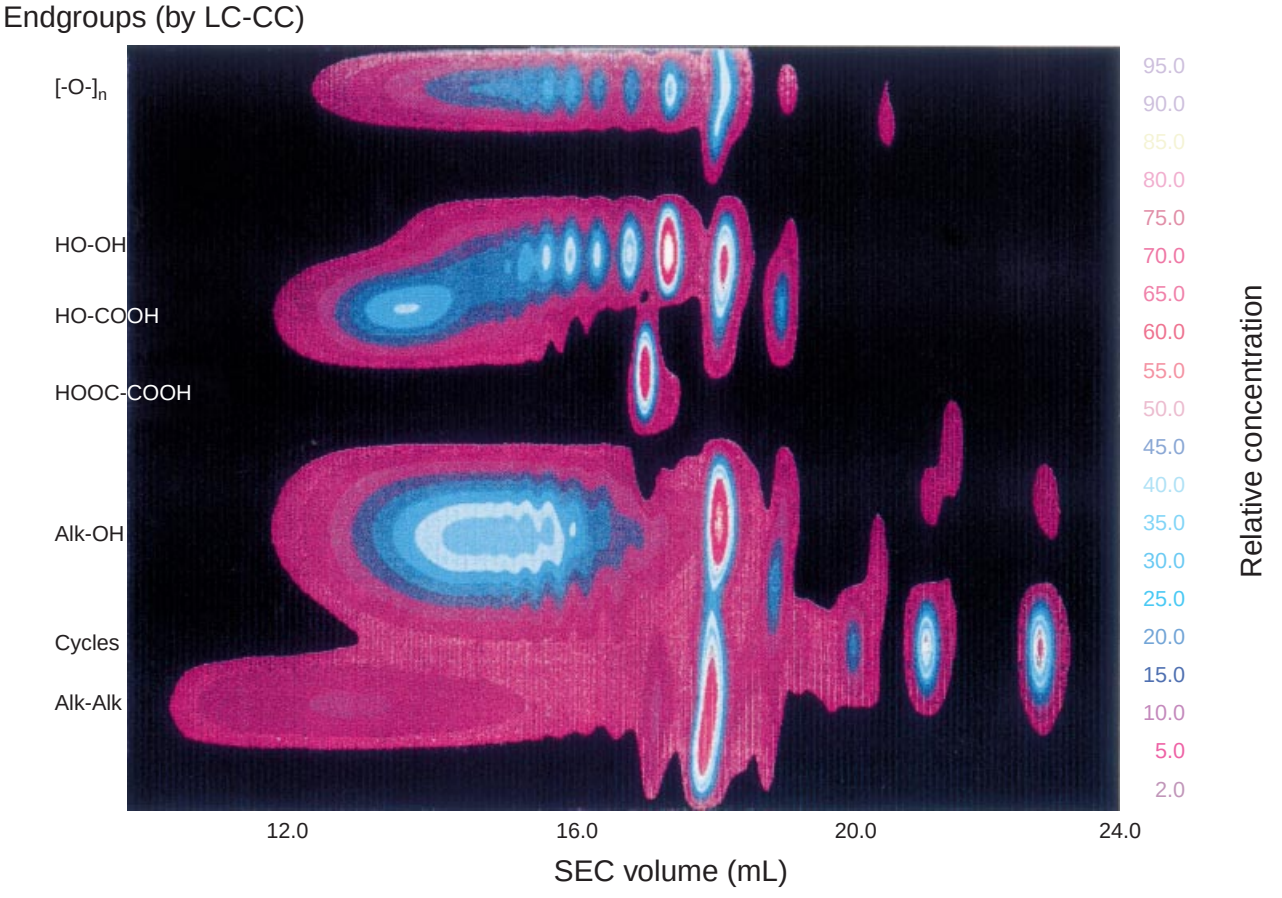

Figure 42 Contour map of the 2-D LC/CC and SEC separation of the aliphatic polyester with acid number of about 5; functionality (on *y*-axis) was calibrated by model compounds.

The contour plot (cf. Figure 43) clearly reveals five spots corresponding to the five functionality fractions, fraction 2 being the main fraction containing the α-octylphenoxy-w-hydroxy oligomers. In addition, α,w-di(octylphenoxy) oligomer fractions and fractions having butylphenoxy endgroups are identified.

The 2-D experiment yielded separation with respect to functionality and molar mass, and FTD and MMD could be determined quantitatively. For calculating FTD, the relative concentration of each functionality fraction must be determined. These concentrations are equivalent to the volume of each peak in the contour plot. With the appropriate software this can be done easily. Determination of the MMD for each fraction was possible after calibrating chromatograph 2 with PEO calibration standards. Calculation of the MMD could then be achieved in the usual way, taking one chromatogram for each functionality fraction, preferably from the region of the highest peak intensity.

In similar approaches other polyalkylene oxides have been analysed by 2-D chromatography. Murphy et al.[234] separated PEGs and Brij type surfactants according to chemical composition and molar mass by RP HPLC versus SEC.

The analysis of methacryloyl-terminated PEOs by LC/CC versus SEC was described by Krüger et al.[233] The functionality-type separation was conducted on an RP system at a critical eluent composition of ACN–water 43 : 57 (v/v). The functionality fractions, including PEG, α-methoxy-w-hydroxy, α-methoxy-w-methacryloyloxy, and α,w-di(methacryloyloxy) PEO, were identified by MALDI/TOF MS.

Finally, a technical C_{13},C_{15}-alkoxy-terminated PEO was analysed by Pasch and Trathnigg using LC/CC vs SEC.[235]

4.5.3 Two-dimensional Combination of Different High Performance Liquid Chromatography Modes

The deformulation of alcohol ethoxylates, which are an important class of nonionic surfactants, by 2-D chromatography has been published by Murphy et al.[236] This class of products possesses a polydispersity of

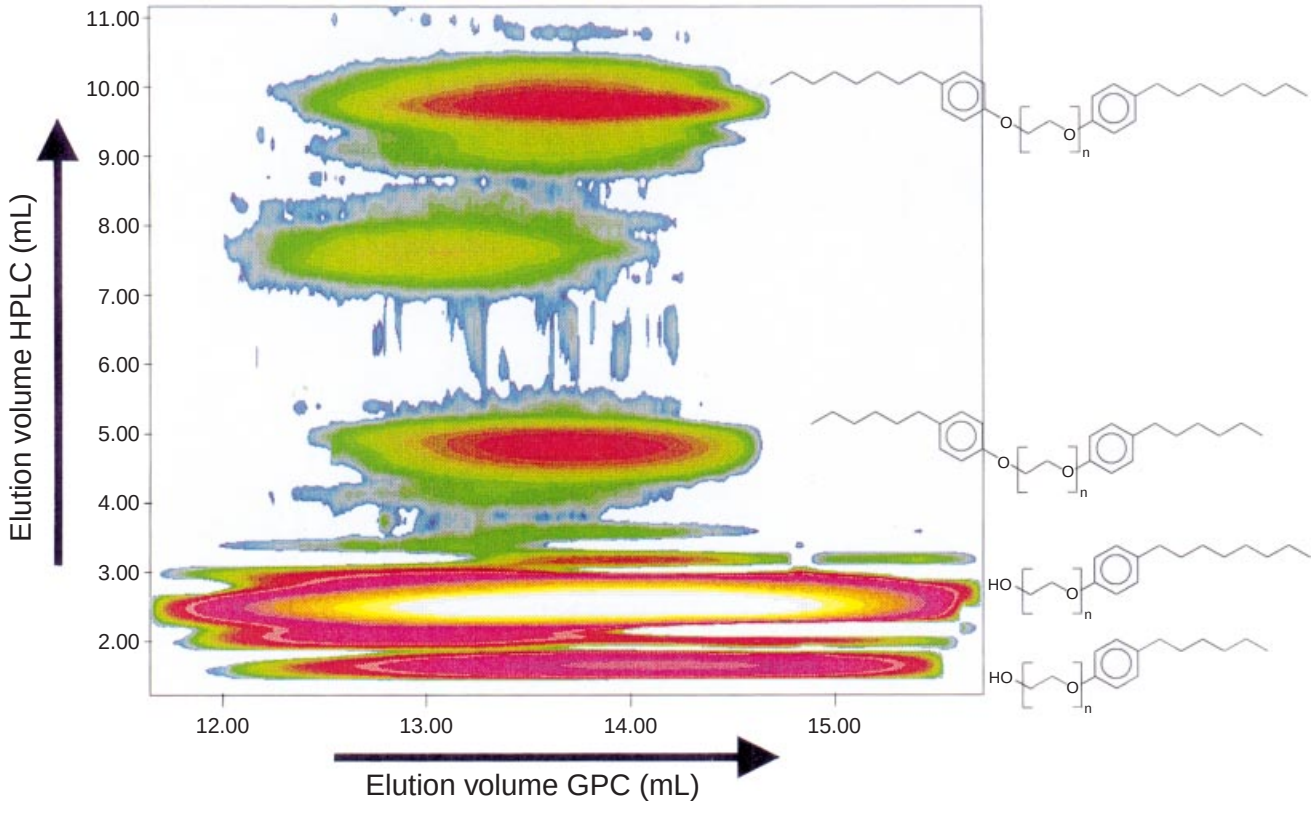

Figure 43 2-D separation of octylphenoxy-terminated PEO shown as a contour plot; the labels indicate the molecular structure verified by MALDI/MS.

methylene groups and a distribution of ethylene glycol units in the PEG segment of the nonionic surfactant. The authors established chromatographic methods for both heterogeneities independently and then combined them later for the 2-D investigations. They used an RP system with a 3 µm C-18 column with 100 Å pore width to separate the alcohol alkyl chains using an isocratic mobile phase (methanol–water (95 : 5)) (see Figure 44). The PEG segments were separated on a normal phase system comprising a 3 µm nonmodified silica with 70 Å pores run with a concave water–ACN gradient (see Figure 45). Both chromatograms show high resolution separations for the alcohol groups or the PEG MMD, without *any* indication of any other unresolved property.

The combination of both techniques in a 2-D experiment revealed the complex nature of the surfactant. The authors applied the gradient normal phase liquid chromatography (NPLC) separation of PEG as the first dimension with the reversed phase liquid chromatography (RPLC) as the second dimension using ELSD as a highly sensitive means of detection. Figure 46 shows the 2-D contour map which displays both the alkyl chain separation on the *x*-axis and the ethylene oxide chain length

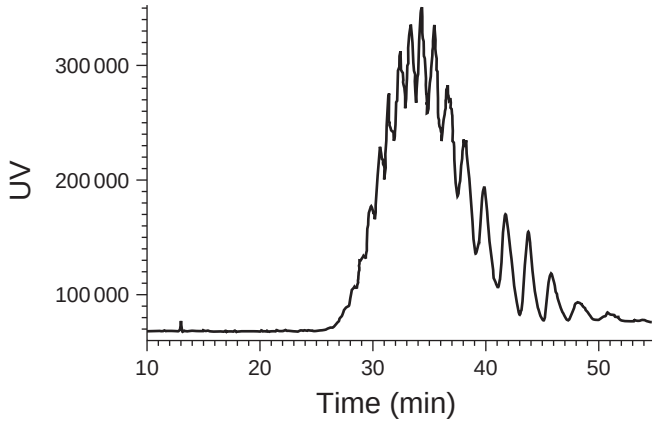

Figure 44 RP separation of FAE non-ionic surfactant for ethylene oxide content determination.

on the *y*-axis. The separation is not truly orthogonal, which the authors attribute to the concave gradient in the NPLC separation. This separation clearly demonstrates the increased peak capacity of 2-D experiments. The high resolution separation in NPLC showed very high peak capacities of about 15, whereas the RPLC experiment resulted in an extremely good separation into the four

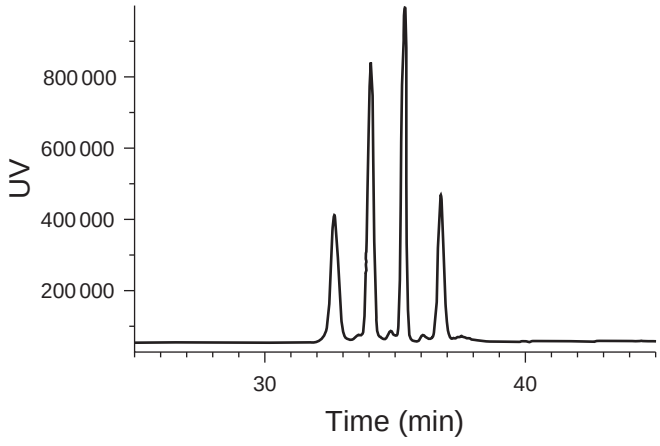

Figure 45 High resolution normal-phase chromatogram of non-ionic surfactant; separation occurs by alkyl chain length without indication of ethylene oxide polydispersity.

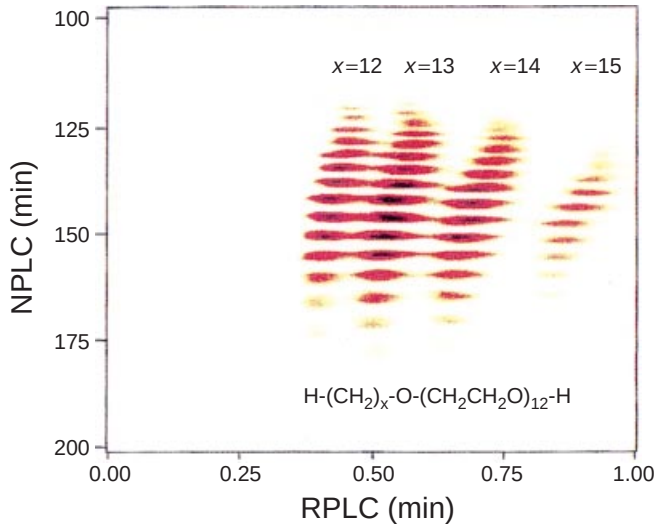

Figure 46 2-D contour plot of FAE characterized in an NPLC/RPLC experiment with very high peak capacity (about 70).

alkyl alcohol chains. The total peak capacity in the 2-D experiment is about 70, which agrees very favorably with the theoretical value of about 60 (cf. section 4.1).

The authors were able to distinguish between different samples of one manufacturer and samples of different producers using this technique. They used the polydispersity of the PEG segment and the presence of various C_n chains from the 2-D contour plot as indicators for identifying sample, synthetic process and manufacturer.

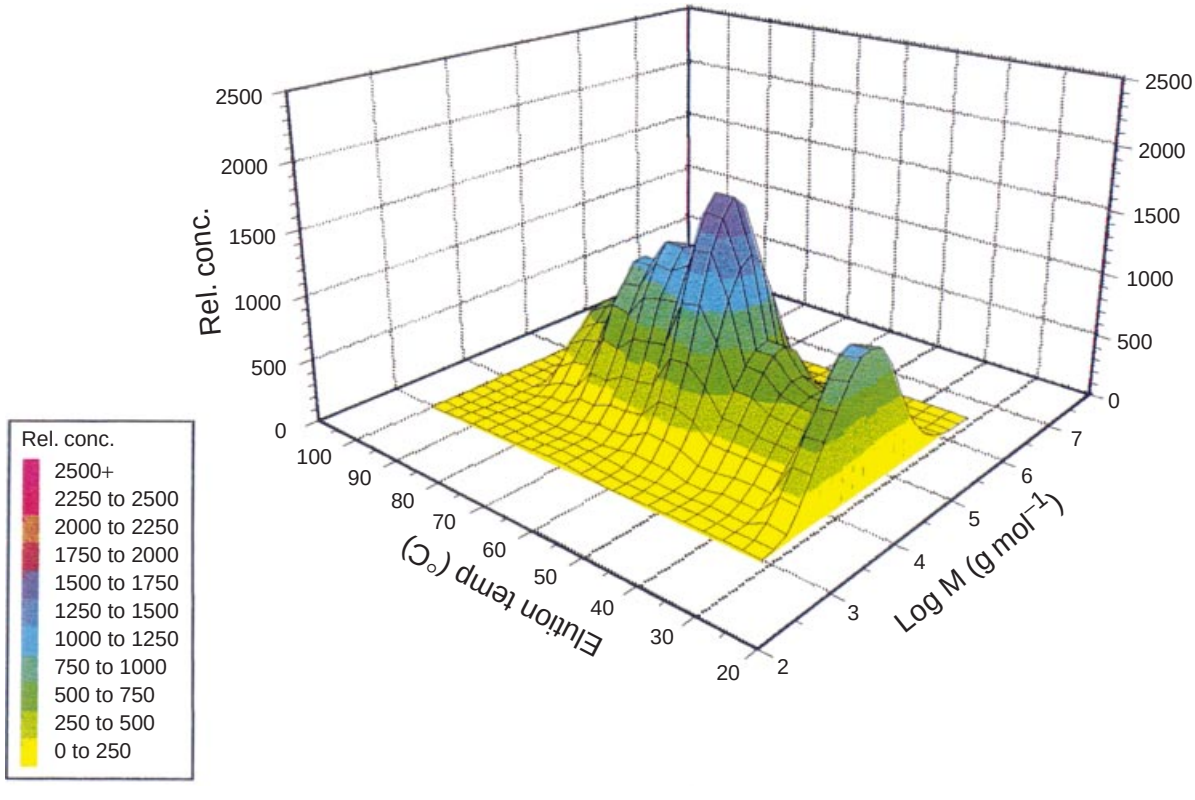

Figure 47 Surface plot of the 2-D TREF/SEC characterization of the ethylene–butene copolymer (Ziegler–Natta catalyst). The temperature axis reflects short-chain branching, the molar mass axis shows the broad MMD of the sample.

4.5.4 Temperature-rising Elution Fractionation–Size Exclusion Chromatography Two-dimensional Separations

This method combination has high potential for the study of the physical properties of polyolefins and can relate them to their molecular structure. The thermal properties and the performance of many polyolefin applications depend on the formation of crystalline domains, which are controlled by the type and the degree of branching of the macromolecule.

TREF (section 4.3.3) is a technique which is widely used for the short-chain branching investigation of polyolefins. Micklitz et al. reported on the 2-D characterization of poly(ethylene-co-butene) using different initiator systems.[237]

The 2-D system comprised a preparative TREF unit running in xylene from which samples were collected and manually transfered into a high-temperature SEC system running in 1,2,4-trichlorobenzene at 135 °C. Figure 47 shows the surface plot of the TREF/SEC characterization of the ethylene–butene copolymer, which has been synthesized from a traditional Ziegler–Natta type catalyst. Both the branching distribution (indicated by the measured property elution temperature by TREF) and the MMD (by SEC calibration) are very broad, as would be expected from a Ziegler–Natta type catalyst. The average molar mass of the sample was about 100 000 g mol^{-1}.

Figure 48 shows an ethylene–butene copolymer formed from a metallocene catalyst using otherwise identical conditions (comonomer ratio, molar mass). Metallocene catalysts marked a new step in polyolefin synthesis, because they allow a much improved control of the structure of the growing macromolecule. The high degree of structure and molar mass control with a metallocene catalyst can be seen from this figure. The width of the 3-D peak with respect to the elution temperature and the molar mass axis is much smaller than that of the Ziegler–Natta product. This indicates that the degree of branching and the insertion of new monomers into the growing polymeric chain can indeed be controlled in a way not possible before.

4.6 Conclusions and Future Developments

Multidimensional chromatography separations are currently one of the most promising and powerful methods for the fractionation and characterization of complex

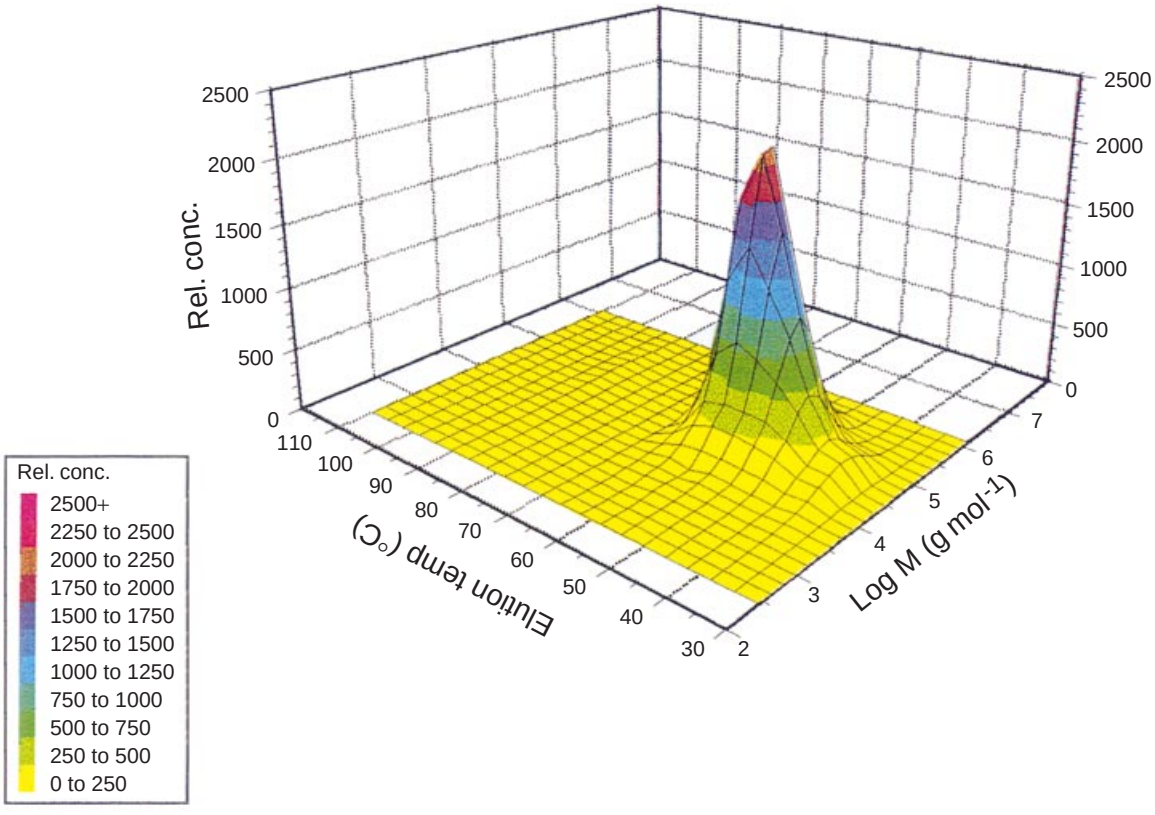

Figure 48 3-D surface plot from a TREF/SEC experiment of a ethylene–butene copolymer formed by a metallocene catalyst (refer to Figure 47 for details).

sample mixtures in different property coordinates. It combines extraordinary resolution and peak capacity with flexibility and overcomes the limitations of any given single chromatographic method. This is the basis for the identification and quantification of major compounds and byproducts, which might adversely affect product properties if not detected in time.

Using a chromatographic separation which is selective towards functionality or chemical composition in the first dimension and SEC in the second dimension, truly ''orthogonal'' separation schemes can be established. Thus, the combination of gradient HPLC versus SEC yields quantitative information on CCD and MMD, while coupling LC/CC and SEC is useful for the analysis of functional homopolymers and block copolymers in the coordinates FTD/MMD and CCD/MMD respectively. Even more complex systems, such as graft copolymers and polymer blends, in which each component may be chemically heterogeneous itself, can be analyzed.

Although 2-D LC is experimentally more demanding than other chromatographic techniques, the complete characterization yields much more qualitative and quantitative information about the sample, and results are presented in an impressively simple way. The contour plot of a 2-D separation maps all obtainable information and allows a fast and reliable comparison between two samples. For future development, the automated comparison of the results of different samples can be considered as an important step to improve process control and quality management. Currently, there is much activity to identify the best method combinations for multidimensional chromatography of many applications which are difficult to separate and quantify.

Recent improvements in the efficiency of the interactive chromatography of macromolecules and other separation modes (e.g. SEC, CE) widen the applicability of 2-D analyses. The next step is the on-line identification of fractions using information-rich detection systems such as FTIR and MS described in section 3 of this article. This will advance the state of the art to a still higher level.

LIST OF SYMBOLS

a_m	activity (concentration) in mobile phase
a_s	activity (concentration) in stationary phase
c	concentration
f	response factor
G	Gibbs free energy
K^*	optical constant in light scattering
K_d	distribution coefficient
M	molar mass
M_w	weight-average molar mass
m_i	mass of species i
n_i	peak capacity of the i-th dimension
N	number of theoretical plates
$P(\Theta)$	scattered light angular dependence
R	universal gas constant
$R(\Theta)$	intensity of scattered light at the angle Θ
V_0	exclusion volume
V_e	elution volume
V_h	hydrodynamic volume
V_p	volume at peak maximum
V_{pore}	pore volume
V_t	total penetration volume
w	weight fraction
η	viscosity of a solution
λ	wavelength
ϑ	projection angle between two separation methods
i	portion of copolymers eluted
V	volume
x	area of slice
j	detector
M_C	molar mass of copolymers
A_2	second virial coefficient
N_A	Avogadro's constant
$[\eta]$	intrinsic viscosity
χ_{AB}	Flory–Huggins parameter
R_g	radius of gyration
η_0	viscosity of a solvent (viscosity of the pure mobile phase)
η_{sp}	specific viscosity
P	pressure
a	Mark–Houwink exponent
M_n	number-average molar mass
K	constant factor in the Mark–Houwink equation
η_{rel}	relative viscosity
H	interaction enthalpy
T	absolute temperature
d_p	diameter of column packing
f	dilution factor
M_v	viscosity-average molar mass
n_i	number of species i
S	entropy
V_{inj}	injection volume

ABBREVIATIONS AND ACRONYMS

ACN	Acetonitrile
CCD	Chemical Composition Distribution
CE	Capillary Electrophoresis
EA	Ethyl Acrylate
ELSD	Evaporative Light Scattering Detection
EPDM	Ethylene–Propylene–Diene Rubber
ESI	Electrospray Ionization

FAE	Fatty Alcohol Ethoxylate
FFF	Field-flow Fractionation
FTD	Functionality Type Distribution
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
GTP	Group Transfer Polymerization
HPLC	High-performance Liquid Chromatography
IR	Infrared
LAC	Liquid Adsorption Chromatography
LALLS	Low Angle Laser Light Scattering
LC	Liquid Chromatography
LC/CC	Liquid Chromatography at the Critical Point of Adsorption
LS	Light Scattering
MAD	Molecular Architecture Distribution
MALDI	Matrix-assisted Laser Desorption/Ionization
MALLS	Multiangle Laser Light Scattering
MMD	Molar Mass Distribution
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NPLC	Normal Phase Liquid Chromatography
PDMS	Polydimethyl Siloxane
PEG	Polyethylene Glycol
PEO	Polyethylene Oxide
PMMA	Polymethyl Methacrylate
PnBMA	Poly-n-butyl Methacrylate
PPG	Poly(propylene glycol)
PS	Polystyrene
RI	Refractive Index
RP	Reversed Phase
RPLC	Reversed Phase Liquid Chromatography
SAN	Styrene–acrylonitrile
SEC	Size Exclusion Chromatography
SFC	Supercritical Fluid Chromatography
S/N	Signal-to-noise
THF	Tetrahydrofuran
TLC	Thin-layer Chromatography
TOF	Time-of-flight
TREF	Temperature-rising Elution Fractionation
UV	Ultraviolet
VIS	Visible
VISC	On-line Viscometry
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Field Flow Fractionation in Analysis of Polymers and Rubbers • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Size-exclusion Chromatography of Polymers • Supercritical Fluid Chromatography of Polymers • Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Normal-phase Liquid Chromatography • Supercritical Fluid Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. H.G. Barth, 'Hyphenated Polymer Separation Techniques. Present and Future Role', in *Chromatographic Characterization of Polymers. Hyphenated and Multidimensional Techniques*, eds. T. Provder, H.G. Barth, M.W. Urban, *Adv. Chem. Ser.* 247, American Chemical Society, Washington, DC, Chapter 1, 1995.
2. F. Francuskiewicz, *Polymer Fractionation*, Springer, Berlin, Heidelberg, New York, 1994.
3. H.J. Cortes, *J. Chromatogr.*, **626**, 3 (1992).
4. A. Factor, J.C. Carnahan, S.B. Dorn, P.C. van Dort, *Polym. Degrad. Stab.*, **45**, 127 (1994).
5. M. Ezrin, G. Lavigne, *Annu. Tech. Conf. – Soc. Plast. Eng.*, **50**, 1717 (1992).
6. Y. Liu, H.C. Chou, J.O. Stoffer, *J. Appl. Polym. Sci.*, **53**, 247 (1994).
7. B. Boinon, M. Raihane, J.P. Montheard, *Polym. Degrad. Stab.*, **43**, 27 (1994).
8. Z. Jedlinski, M. Kowalczuk, P. Kurcok, *J. Macromol. Sci.*, **A29**, 1223 (1992).
9. G. Matischek, H. Stoffers, K.H. Ohrbach, A. Kettrup, *Polym. Degrad. Stab.*, **39**, 381 (1993).
10. B. Bell, D.E. Beyer, N.L. Maecker, R.R. Papenfus, D.B. Priddy, *J. Appl. Polym. Sci.*, **54**, 1605 (1994).
11. M. Farina, G. Di Silvestro, P. Sozzani, C.M. Yuan, 'Makromol. Chem.', *Macromol. Symp.*, **47**, 1 (1991).
12. S. Liebman, E. Levy, *Pyrolysis and GC in Polymer Analysis*, Marcel Dekker, New York, 1985.
13. M. Geißler, *Kunststoffe*, **87**, 194 (1997).

14. M. Sabo, *Anal. Chem.*, **57**, 1822 (1985).
15. J.W. Hellgeth, L.T. Taylor, *Anal. Chem.*, **59**, 295 (1987).
16. J.W. Hellgeth, L.T. Taylor, *J. Chromatogr. Sci.*, **24**, 519 (1986).
17. P.R. Griffiths, C.M. Conroy, *Adv. Chromatogr.*, **25**, 105 (1986).
18. J.J. Gagel, K. Biemann, *Anal. Chem.*, **58**, 2184 (1986).
19. L.M. Wheeler, J. Willis, *Appl. Spectrosc.*, **47**, 1128 (1993).
20. J.C. Lindon, *Progr. Nucl. Magn. Reson. Spectrosc.*, **29**, 1 (1996).
21. K. Albert, *J. Chromatogr., A*, **703**, 123 (1995).
22. S.T. Balke, R.D. Patel, *Adv. Chem. Ser.*, **203**, 281 (1983).
23. P. Kilz, *Laborpraxis*, **6**, 628 (1992).
24. B. Trathnigg, *J. Liq. Chromatogr.*, **13**, 1731 (1990).
25. B. Trathnigg, *J. Chromatogr.*, **552**, 505 (1991).
26. B. Trathnigg, X. Yan, *Chromatographia*, **33**, 467 (1992).
27. J.R. Runyon, D.E. Barnes, J.F. Rudel, L.H. Tung, *J. Appl. Polym. Sci.*, **13**, 2359 (1969).
28. A. Revillon, *J. Liq. Chromatogr.*, **3**, 1137 (1980).
29. L.H. Tung, *J. Appl. Polym. Sci.*, **24**, 953 (1979).
30. B. Trathnigg, S. Feichtenhofer, M. Kollroser, *J. Chromatogr., A*, **786**, 75 (1997).
31. C. Johann, P. Kilz, *Proc. 1st Int. Conf. Molar Mass Charact.*, Bradford, 1989.
32. J.V. Dawkins, 'Compositional Heterogeneity of Copolymers by Coupled Techniques with Chromatographic Columns and Multidetectors', in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques*, eds. T. Provder, H.G. Barth, M.W. Urban, *Adv. Chem. Ser.* 247, American Chemical Society, Washington, DC, Chapter 15, 1995.
33. E. Meehan, S. O'Donohue, J.A. McConville, *Polym. Mater. Sci. Eng.*, **69**, 269 (1993).
34. F. Gores, P. Kilz, 'Copolymer Characterization Using Conventional SEC and Molar Mass-sensitive Detectors', in *Chromatography of Polymers, ACS Symp. Ser.* 521, ed. T. Provder, American Chemical Society, Washington, DC, Chapter 10, 1993.
35. H.C. Lee, M. Ree, T. Chang, *Polymer*, **36**, 2215 (1995).
36. H.C. Lee, T. Chang, *Bull. Korean Chem. Soc.*, **17**, 648 (1996).
37. S. Podzimek, *J. Appl. Polym. Sci.*, **54**, 91 (1994).
38. U. Dayal, S.K. Mehta, *J. Liq. Chromatogr.*, **17**, 303 (1994).
39. D. Dayal, *J. Appl. Polym. Sci.*, **53**, 1557 (1994).
40. S. Balke, B. Rao, R. Thiriratsakul, T.H. Mourey, T.C. Schunk, *Prep. 9th Int. Symp., Polym. Anal. Char. (ISPAC 9)*, Oxford, 1996.
41. C. Jackson, H.G. Barth, in *Molecular Weight Sensitive Detectors for Size Exclusion Chromatography*, ed. C.S. Wu, Marcel Dekker, New York, Chapter 4, 1995.
42. P. Kratochvil, in *Classical Light Scattering from Polymer Solutions, Polym. Sci. Libr. 5*, ed. A.D. Jenkins, Elsevier, Amsterdam, 1987.
43. L. Jeng, S.T. Balke, T.H. Mourey, L. Wheeler, P. Romeo, *J. Appl. Polm. Sci.*, **49**, 1359 (1993).
44. L. Jeng, S.T. Balke, *J. Appl. Polm. Sci.*, **49**, 1375 (1993).
45. L. Pille, D.H. Solomon, *Macromol. Chem. Phys.*, **195**, 2477 (1994).
46. T.H. Mourey, H. Coll, *Polym. Mater. Sci. Eng.*, **69**, 217 (1994).
47. T.H. Mourey, H. Coll, *J. Appl. Polym. Sci.*, **56**, 65 (1995).
48. Z. Grubisic-Gallot, Y. Gallot, J. Sedlacek, *Macromol. Chem. Phys.*, **195**, 781 (1994).
49. Z. Grubisic-Gallot, Y. Gallot, J. Sedlacek, *J. Liq. Chromatogr.*, **18**, 2291 (1995).
50. M. Zigon, Z. Grubisic-Gallot, *Proc. 11th Int. Symp., Polym. Anal. Char.*, Santa Margerita Ligure, Italy, 1998.
51. W.W. Yau, S.D. Abbott, G.A. Smith, M.Y. Keating, in *Detection and Data Analysis in Size Exclusion Chromatography, ACS Symp. Ser.* 352, ed. T. Provder, American Chemical Society, Washington, DC, 1987.
52. M. Styring, J.E. Armonas, A.E. Hamielec, in *Detection and Data Analysis in Size Exclusion Chromatography, ACS Symp. Ser.* 352, ed. T. Provder, American Chemical Society, Washington, DC, 1987.
53. M.A. Haney, J.E. Armonas, L. Rosen, in *Detection and Data Analysis in Size Exclusion Chromatography, ACS Symp. Ser.* 352, ed. T. Provder, American Chemical Society, Washington, DC, 1987.
54. L. Brower, D. Trowbridge, D. Kim, P. Mukherjee, R. Seeger, D. McIntyre, in *Detection and Data Analysis in Size Exclusion Chromatography, ACS Symp. Ser.* 352, ed. T. Provder, American Chemical Society, Washington, DC, 1987.
55. S. Mori, *J. Chromatogr.*, **637**, 129 (1993).
56. P. Cheung, R. Lew, S.T. Balke, T.H. Mourey, *J. Appl. Polym. Sci.*, **47**, 1701 (1993).
57. J. Lesec, M. Millequant, T. Havard, in *Chromatography of Polymers, ACS Symp. Ser.* 521, ed. T. Provder, American Chemical Society, Washington, DC, 1993.
58. H. Benoit, P. Rempp, Z. Grubisic, *J. Polym. Sci.*, **B5**, 753 (1967).
59. M. Kurata, Y. Tsunashima, M. Iwama, K. Kamada, in *Polymer Handbook*, eds. J. Brandrup, I. Immergut, Wiley, New York, IV-1, 1975.
60. G. Glöckner, *Gradient HPLC of Copolymers and Chromatographic Cross-fractionation*, Springer, Berlin, Heidelberg, New York, 1991.
61. W. Yau, Chemtracts-Macromol, *Chem.*, **1**, 1 (1990).
62. R.A. Sanayei, K.G. Suddaby, A. Rudin, *Makromol. Chem.*, **194**, 1953 (1993).
63. J.E. Puskas, R. Hutchinson, *Rubber Chem. Technol.*, **66**, 742 (1993).
64. J. Xie, *Polymer*, **35**, 2385 (1994).
65. R. Lew, D. Suwanda, S.T. Balke, T.H. Mourey, *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **52**, 125 (1993).
66. C.A. Harrison, T.H. Mourey, *J. Appl. Polym. Sci.*, **56**, 211 (1995).
67. O. Chiantore, M. Lazzari, S. Caldari, E. Zecchi, *Proc. 11th Int. Symp., Polym. Anal. Char.*, Santa Margerita Ligure, Italy, 1998.

68. J.M. Goldwasser, *Proc. Int. GPC Symp.*, Newton, MA, 1989.
69. S. Pang, A. Rudin, *Polymer*, **33**, 1949 (1992).
70. M. Wintermantel, M. Antonietti, M. Schmidt, *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **52**, 91 (1993).
71. C. Degoulet, T. Nicolai, D. Durand, J.P. Busnel, *Macromolecules*, **28**, 6819 (1995).
72. C. Jackson, Y.J. Chen, J.W. Mays, *J. Appl. Polym. Sci.*, **61**, 865 (1996).
73. W.W. Yau, K.S. Arora, *Polym. Mater. Sci. Eng.*, **69**, 210 (1994).
74. W.W. Yau, *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **48**, 85 (1991).
75. W. Köhler, A. Kühn, A. Motsch, *Acta Polym.*, **44**, 238 (1993).
76. U. Lehmann, W. Köhler, W. Albrecht, *Macromolecules*, **29**, 3212 (1996).
77. C. Jackson, H.G. Barth, *Trends Polym. Sci.*, **2**, 203 (1994).
78. M.P. Balogh, *LC-GC Int.*, **10**, 728 (1997).
79. R. Smits, *LC-GC Int.*, **8**, 92 (1995).
80. W.M.A. Niessen, A.P. Tinke, *J. Chromatogr., A*, **703**, 37 (1995).
81. M. Dole, *J. Chem. Phys.*, **49**, 2240 (1968).
82. E.C. Horning, *J. Chromat. Sci.*, **12**, 725 (1974).
83. M.L. Vestal, G.J. Ferguson, *Anal. Chem.*, **57**, 2373 (1985).
84. R.C. Willoughby, R.F. Browner, *Anal. Chem.*, **56**, 2626 (1984).
85. P. Vouros, J.W. Wronka, in *Modern Methods in Polymer Characterization*, eds. H.G. Barth, J.W. Mays, Wiley, New York, Chapter 12, 1991.
86. M.L. Vestal, *Science*, **226**, 595 (1984).
87. J.D. Vargo, K.L. Olson, *Anal. Chem.*, **57**, 672 (1985).
88. M.L. Vestal, *Mass Spectrom. Rev.*, **2**, 447 (1983).
89. J.D. Vargo, K.L. Olson, *J. Chromatogr.*, **353**, 215 (1986).
90. D.V. Leyen, B. Hagenhoff, *J. Vac. Sci. Technol., A*, **7**, 1790 (1989).
91. I.V. Bletsos, D.M. Hercules, *Anal. Chem.*, **63**, 1953 (1991).
92. R.P. Lattimer, H.R. Schulten, *Int. J. Mass Spectrom. Ion Phys.*, **67**, 227 (1985).
93. K. Rollins, J.H. Scrivens, M.J. Taylor, H. Major, *Rapid Commun. Mass Spectrom.*, **4**, 355 (1990).
94. J.B. Fenn, T.J. Nohmi, *J. Am. Chem. Soc.*, **114**, 3241 (1992).
95. G.J. Kallos, D.A. Tomalia, *Rapid Commun. Mass Spectrom.*, **5**, 383 (1991).
96. L. Prokai, W.J. Simonsick, in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques, Adv. Chem. Ser.* 247, eds. T. Provder, H.G. Barth, M.W. Urban, American Chemical Society, Washington, DC, Chapter 4, 1995.
97. L. Prokai, W.J. Simonsick, *Rapid Commun. Mass Spectrom.*, **7**, 853 (1993).
98. L. Prokai, D.J. Aaaserud, W.J. Simonsick, *J. Chromatogr., A*, **835**, 121 (1999).
99. W.J. Simonsick, *Polym. Mater. Sci. Eng.*, **69**, 412 (1993).
100. W.J. Simonsick, C.W. Ross, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)*, **37**, 286 (1996).
101. L. Prokai, S.W. Myung, W.J. Simonsick, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)*, **37**, 288 (1996).
102. T.J. Kemp, Z. Barton, A. Mahon, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)*, **37**, 305 (1996).
103. R.E. Murphy, M.R. Schure, J.P. Foley, *J. Chromatogr., A*, **824**, 181 (1998).
104. M. Karas, F. Hillenkamp, *Anal. Chem.*, **60**, 2299 (1988).
105. R.C. Beavis, B.T. Chait, *Rapid Commun. Mass Spectrom.*, **3**, 233 (1989).
106. F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, *Anal. Chem.*, **63**, 1193A (1991).
107. U. Bahr, A. Deppe, M. Karas, F. Hillenkamp, U. Giessmann, *Anal. Chem.*, **64**, 2866 (1992).
108. P.O. Danis, D.E. Karr, F. Mayer, A. Holle, C.H. Watson, *Org. Mass Spectrom.*, **27**, 843 (1992).
109. M.S. Kahr, C.L. Wilkins, *J. Am. Soc. Mass Spectrom.*, **4**, 453 (1993).
110. G. Montaudo, M.S. Montaudo, C. Puglisi, F. Samperi, *Rapid Commun. Mass Spectrom.*, **8**, 1011 (1994).
111. P.O. Danis, D.E. Karr, Y. Xiong, K.G. Owens, *Rapid Commun. Mass Spectrom.*, **10**, 862 (1996).
112. H. Pasch, R. Unvericht, *Angew. Makromol. Chem.*, **212**, 191 (1993).
113. H. Pasch, F. Gores, *Polymer*, **36**, 1999 (1995).
114. D. Braun, R. Ghahary, H. Pasch, *Polymer*, **37**, 777 (1996).
115. H. Pasch, A. Deffieux, R. Ghahary, M. Schapacher, L. Riquet-Lurbet, *Macromolecules*, **30**, 98 (1997).
116. H. Pasch, K. Rode, *J. Chromatogr., A*, **699**, 21 (1995).
117. R.P. Krüger, H. Much, G. Schulz, *GIT Fachz. Lab.*, **4**, 398 (1996).
118. U. Just, R.P. Krüger, in *Organosilicon Chemistry*, eds. N. Auner, J. Weis, VCH Publishers, Weinheim, Vol. 11, 1996.
119. G. Montaudo, *Rapid Commun. Mass Spectrom.*, **9**, 1158 (1995).
120. P.O. Danis, D.A. Saucy, F.J. Huby, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)*, **37**, 311 (1996).
121. G. Montaudo, *Proc. 11th Int. Symp., Polym. Anal. Char.*, Santa Margerita Ligure, Italy, 1998.
122. C.E. Kassis, J.M. DeSimone, R.W. Linton, E.E. Remsen, G.W. Lange, R.M. Friedman, *Rapid Commun. Mass Spectrom.*, **11**, 1134 (1997).
123. M.W.F. Nielen, *Anal. Chem.*, **70**, 1563 (1998).
124. J. Dwyer, D. Botten, *Internat. Lab.*, **27**, 13A (1997).
125. H. Pasch, *Phys. Chem. Chem. Phys.*, **1**, 3879 (1999).
126. K.K. Murray, D.H. Russell, *Anal. Chem.*, **65**, 2534 (1993).
127. K.K. Murray, D.H. Russell, *J. Am. Soc. Mass Spectrom.*, **5**, 1 (1994).
128. X. Fei, K.K. Murray, *Anal. Chem.*, **68**, 3560 (1996).
129. J.H. Ross, R.L. Shank, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)*, **12**, 812 (1971).

130. D.W. Vidrine, D.R. Mattson, *Appl. Spectrosc.*, **32**, 502 (1978).
131. D.W. Vidrine, *J. Chromatogr. Sci.*, **17**, 477 (1979).
132. R.S. Brown, D.W. Hausler, L.T. Taylor, B.C. Carter, *Anal. Chem.*, **53**, 197 (1981).
133. C.P. Wang, D.T. Sparks, S.S. Williams, T.L. Isenhour, *Anal. Chem.*, **56**, 1268 (1984).
134. C.C. Johnson, L.T. Taylor, *Anal. Chem.*, **56**, 2642 (1984).
135. M. Sabo, *Anal. Chem.*, **57**, 1822 (1985).
136. J.W. Hellgeth, L.T. Taylor, *Anal. Chem.*, **59**, 295 (1987).
137. R.M. Robertson, J.A. de Haseth, J.D. Kirk, R.F. Browner, *Appl. Spectrosc.*, **42**, 1365 (1988).
138. P.R. Griffiths, C.M. Conroy, *Adv. Chromatogr.*, **25**, 105 (1986).
139. J.W. Hellgeth, L.T. Taylor, *J. Chromatogr. Sci.*, **24**, 519 (1986).
140. J.J. Gagel, K. Biemann, *Anal. Chem.*, **58**, 2184 (1986).
141. J.J. Gagel, K. Biemann, *Anal. Chem.*, **59**, 1266 (1987).
142. J.J. Gagel, K. Biemann, *Mikrochim. Acta*, **11**, 185 (1988).
143. L.M. Wheeler, J.N. Willis, *Appl. Spectrosc.*, **47**, 1128 (1993).
144. J.N. Willis, J.L. Dwyer, L.M. Wheeler, *Polym. Mat. Sci.*, **69**, 120 (1993).
145. J.N. Willis, J.L. Dwyer, M.X. Lui, *Proc. Int. GPC Symp.*, 1994, Lake Buena Vista, FL, 1995.
146. H. Pasch, E. Esser, P. Montag, *GIT Fachz. Lab. Chromatogr.*, **16**, 68 (1996).
147. J.N. Willis, L. Wheeler, in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques, Adv. Chem. Ser.* 247, eds. T. Provder, H.G. Barth, M.W. Urban, American Chemical Society, Washington, DC, Chapter 19, 1995.
148. T. Provder, C. Kuo, M. Whited, D. Huddleston, *Proc. Int. GPC Symp.*, 1994, Lake Buena Vista, FL, 1995.
149. D. Braun, E. Esser, H. Pasch, *Int. J. Polym. Char.*, **4**, 501 (1998).
150. E. Esser, D. Braun, H. Pasch, *Angew. Makromol. Chem.*, (1999) (in press).
151. P.C. Cheung, S.T. Balke, T.C. Schunk, in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques, Adv. Chem. Ser.* 247, eds. T. Provder, H.G. Barth, M.W. Urban, American Chemical Society, Washington, DC, Chapter 19, 1995.
152. P. Cheung, S.T. Balke, T.C. Schunk, T.H. Mourey, *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **52**, 105 (1993).
153. K. Albert, *J. Chromatogr.*, **703**, 123 (1995).
154. H.C. Dorn, *Anal. Chem.*, **56**, 747A (1984).
155. D.A. Laude, C.L. Wilkins, *Trends Anal. Chem.*, **5**, 230 (1986).
156. K. Albert, E. Bayer, *Trends Anal. Chem.*, **7**, 288 (1988).
157. K. Albert, E. Bayer, in *HPLC Detection: Newer Methods*, ed. G. Patonay, VCH Publishers, New York, 1992.
158. M. Hofmann, M. Spraul, R. Streck, I.D. Wilson, A. Rapp, *Labor Praxis*, **10**, 36 (1993).
159. H.S. Smallcombe, L.P. Pratt, P.A. Keifer, *J. Magn. Reson., A*, **117**, 295 (1995).
160. R.J. Ogg, P.B. Kingsley, J.S. Taylor, *J. Magn. Reson., B*, **104**, 1 (1994).
161. K. Albert, U. Braumann, *Anal. Chem.*, **66**, 3042 (1994).
162. K. Albert, U. Braumann, R. Streck, M. Spraul, R. Ecker, *Fresenius' J. Anal. Chem.*, **352**, 521 (1995).
163. K. Albert, *J. Chromatogr., A*, **785**, 65 (1997).
164. U. Braumann, H. Händel, R. Albert, *Anal. Chem.*, **67**, 930 (1995).
165. N. Wu, T.L. Peck, A.G. Webb, R.L. Magin, J.V. Sweedler, *J. Am. Chem. Soc.*, **116**, 7929 (1994).
166. N. Wu, T.L. Peck, A.G. Webb, R.L. Magin, J.V. Sweedler, *Anal. Chem.*, **22**, 3849 (1994).
167. J.V. Sweedler, *Proc. 7th Int. Symp., High Performance Capillary Electrophoresis*, Würzburg, Germany, 1995.
168. K. Hatada, K. Ute, Y. Okamoto, M. Imanari, N. Fujii, *Polym. Bull.*, **20**, 317 (1988).
169. K. Hatada, K. Ute, M. Kashiyama, M. Imanari, *Polym. J.*, **22**, 218 (1990).
170. K. Ute, *Proc. 11th Int. Symp., Polym. Anal. Char.*, Santa Margerita Ligure, Italy, 1998.
171. H. Pasch, W. Hiller, *Macromolecules*, **2**, 6556 (1996).
172. G. Schlotterbeck, H. Pasch, K. Albert, *Polym. Bull.*, **38**, 673 (1997).
173. H. Pasch, W. Hiller, R. Haner, *Polymer*, **39**, 1515 (1998).
174. I. Krämer, H. Pasch, H. Händel, K. Albert, *Macromol. Chem. Phys.*, **200**, 1734 (1999).
175. E. Grushka, *Anal. Chem.*, **42**, 1142 (1970).
176. R. Consden, A.H. Gordon, A.J.P. Martin, *Biochem. J.*, **38**, 224 (1944).
177. N. Grinberg, H. Kalász, S.M. Han, D.W. Armstrong, in *Modern Thin-layer Chromatography*, ed. N. Grinberg, Marcel Dekker Publishing, New York, Chapter 7, 1990.
178. P.H. O'Farrell, *Biol. Chem.*, **250**, 4007 (1975).
179. J.E. Celis, R. Bravo, *Two-dimensional Gel Electrophoresis of Proteins*, Academic Press, New York, 1984.
180. N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter, N.G. Anderson, *Clin. Chem.*, **27**, 1807 (1981).
181. M.L. Efron, *Biochem. J.*, **72**, 691 (1959).
182. R.E. Murphy, M.R. Schure, J.P. Foley, *Anal. Chem.*, **70**, 1585 (1998).
183. S.T. Balke, *Quantitative Column Liquid Chromatography, Journal of Chromatography Library*, Elsevier Publishing, New York, Vol. 29, 1984.
184. M.M. Bushey, J.W. Jorgenson, *Anal. Chem.*, **62**, 161 (1990).
185. H.J. Cortes, *Multidimensional Chromatography: Techniques and Applications*, Marcel Dekker, New York, 1990.
186. J.P. Larmann, A.V. Lemmo, A.W. Moore, J.W. Jorgenson, *Electrophoresis*, **14**, 439 (1993).
187. Z. Liu, S.R. Sirimanne, D.G. Patterson, L.L. Needham, J.B. Phillips, *Anal. Chem.*, **66**, 3086 (1993).
188. P. Kilz, R.-P. Krüger, H. Much, G. Schulz, in *Chromatographic Characterization of Polymers: Hyphenated and Multidimensional Techniques, Adv. Chem. Ser.* 247, eds. T. Provder, M.W. Urban, H.G. Barth, American Chemical Society, Washington, DC, 1995.

189. E. Venema, P. de Leeuw, J.C. Kraak, H. Poppe, R.J. Tijssen, *Chrom. A*, **765**, 135 (1997).
190. R.E. Murphy, M.R. Schure, J.P. Foley, *Anal. Chem.*, **70**, 4353 (1998).
191. S.T. Balke, R.D. Patel, *J. Polym. Sci., Polym. Lett.*, **18**, 453 (1980).
192. S.T. Balke, *Sep. Purif. Methods*, **1**, 1 (1982).
193. S.T. Balke, R.D. Patel, in *Polymer Characterization, Adv. Chem. Ser.* 203, ed. C.D. Craver, American Chemical Society, Washington, DC, 1983.
194. P. Kilz, R.-P. Krüger, H. Much, G. Schulz, *Polym. Mater. Sci. Eng.*, **69**, 114 (1993).
195. *PSS Application Note: 2D Transfer Options*, PSS Polymer Standards Service, Mainz, 1995.
196. M.R. Schure, *Anal. Chem.*, **71**, 1645 (1999).
197. T. Ogawa, M. Sakai, *J. Polym. Sci., Polym. Phys. Ed.*, **19**, 1377 (1982).
198. G. Glöckner, J.H.M. van den Berg, N.L. Meijerink, T.G. Scholte, in *Integration of Fundamental Polymer Science and Technology*, eds. I. Kleintjens, P. Lemstra, Elsevier Applied Science, Barking, 1986.
199. *PSS Win GPC Software Brochure*, PSS Polymer Standards Service, Mainz, 1996.
200. M. Danielewicz, M. Kubin, *J. Appl. Polym. Sci.*, **26**, 951 (1981).
201. G. Glöckner, H. Koschwitz, C. Meissner, *Acta Polym.*, **33**, 614 (1982).
202. H. Sato, H. Takeuchi, Y. Tanaka, *Macromolecules*, **19**, 2613 (1986).
203. T.H. Mourey, *J. Chromatogr.*, **357**, 101 (1986).
204. S. Mori, *J. Appl. Polym. Sci.*, **38**, 95 (1989).
205. T. Chang, H.C. Lee, *Polymer*, **37**, 5747 (1996).
206. G. Glöckner, *Polymer Characterization by Liquid Chromatography*, Elsevier, Amsterdam, 1987.
207. B.G. Belenkii, E.S. Gankina, M.B. Tennikov, L.Z. Vilenchik, *Dokl. Akad. Nauk*, **231**, 1147 (1976).
208. M.B. Tennikov, P.P. Nefedov, M.A. Lazareva, S.J. Frenkel, *Vysokomol. Soedin*, **A19**, 657 (1977).
209. A.M. Skvortsov, B.G. Belenkii, E.S. Gankina, M.B. Tennikov, *Vysokomol. Soedin*, **A20**, 678 (1978).
210. S.G. Entelis, V.V. Evreinov, A.V. Gorshkov, *Adv. Polym. Sci.*, **76**, 129 (1986).
211. B.G. Belenkii, E.S. Gankina, *J. Chromatogr.*, **141**, 13 (1977).
212. A.M. Skvortsov, A.A. Gorbunov, *J. Chromatogr.*, **507**, 487 (1990).
213. A.V. Gorshkov, T.N. Prudskova, V.V. Guryakova, V.V. Evreinov, *Polym. Bull.*, **15**, 465 (1986).
214. A.V. Gorshkov, S.S. Verenich, V.V. Evreinov, S.G. Entelis, *Chromatographia*, **26**, 338 (1988).
215. A.V. Gorshkov, H. Much, H. Becker, H. Pasch, V.V. Evreinov, S.G. Entelis, *J. Chromatographia*, **523**, 91 (1990).
216. H. Pasch, I. Zammert, *J. Liq. Chromatogr.*, **17**, 3091 (1994).
217. H. Pasch, C. Brinkmann, H. Much, U. Just, *J. Chromatogr.*, **623**, 315 (1992).
218. H. Pasch, M. Augenstein, *Makromol. Chem.*, **194**, 2533 (1993).
219. H. Pasch, Y. Gallot, B. Trathnigg, *Polymer*, **34**, 4986 (1993).
220. H. Pasch, M. Augenstein, B. Trathnigg, *Makromol. Chem.*, **195**, 743 (1994).
221. H. Pasch, A. Deffieux, I. Henze, M. Schappacher, L. Rique-Lurbet, *Macromolecules*, **29**, 8776 (1996).
222. H. Pasch, *Polymer*, **34**, 4095 (1993).
223. H. Pasch, K. Rode, *Macromol. Chem. Phys.*, **197**, 2691 (1996).
224. H. Pasch, K. Rode, N. Chaumien, *Polymer*, **37**, 4079 (1996).
225. J. Adrian, H. Pasch, D. Braun, *LC-GC Int.*, **11**, 32 (1998).
226. U. Just, H. Much, *Proc. Int. GPC Symp.*, 861, Milford, 1994.
227. G. Glöckner, M. Stickler, W. Wunderlich, *J. Appl. Polym. Sci.*, **37**, 3147 (1989).
228. M. Augenstein, M. Stickler, *Makromol. Chem.*, **191**, 415 (1990).
229. S. Mori, *J. Chromatogr.*, **503**, 411 (1990).
230. S. Mori, *Anal. Chem.*, **53**, 1813 (1981).
231. S. Mori, *Anal. Chem.*, **60**, 1125 (1988).
232. B. Trathnigg, B. Maier, X. Yan, *Proc. 7th Int. Symp., Polym. Anal. Char.*, 1994.
233. R.-P. Krüger, H. Much, G. Schulz, *Int. J. Polym. Anal. Char.*, **2**, 221 (1996).
234. R.E. Murphy, M.R. Schure, L.P. Mink, J.P. Foley, *Proc. 8th Int. Symp., Polym. Anal. Char.*, 1995.
235. H. Pasch, B. Trathnigg, *HPLC of Polymers*, Springer, Berlin, 203, 1997.
236. R.E. Murphy, M.R. Schure, J.P. Foley, *Anal. Chem.*, **70**, 4353 (1998).
237. W. Micklitz, D. Lilge, H.-J. Müller, M. Lux, F. Langhauser, S. Brode, E. Folga, E. Hädicke, H.-H. Görtz, G. Schweier, *ACS National Meeting*, San Diego, USA, 1994.

Dielectric Spectroscopy in Analysis of Polymers

John W. Schultz
Georgia Tech Research Institute, Atlanta, USA

1 Introduction **1**
2 Instrumentation and Sample Preparation **2**
2.1 Instrument Basics 2
2.2 Measured Parameters 2
2.3 Electrode–Sample Configurations 3
2.4 Potential Pitfalls 5
2.5 Thermally Stimulated Current 6
3 Data Analysis **7**
3.1 Calculated Parameters 7
3.2 Origins of Dielectric Properties in Polymers 7
3.3 Relaxation 9
4 Applications and Methods **10**
4.1 Detecting Relaxations 10
4.2 Cure and Process Monitoring 12
4.3 Polymer Emulsions and Film Formation 14
4.4 Humidity/Environmental Effects 14
4.5 Analysis of Composites 15
Acknowledgments **16**
List of Symbols **16**
Abbreviations and Acronyms **17**
Related Articles **17**
References **17**

Dielectric spectroscopy is the measurement of complex permittivity (or ''dielectric constant'') as a function of frequency. This measurement can also be done as a function of temperature or time at fixed frequencies to determine various physical and chemical properties of a given polymer. The permittivity is usually measured by placing the sample in contact with electrodes and applying a sinusoidal voltage. Permanent and induced dipoles, and also ion or electron conduction, all contribute to the dielectric response of a material. Dielectric spectroscopy can be used to probe molecular relaxation processes in polymers such as the glass transition or phase transitions such as a melt. For many of the relaxations, it is among the most sensitive of the thermal analysis techniques. Dielectric spectroscopy can also be used to follow chemical changes in polymer systems, such as polymer cure or aging. Because of its enormous frequency range (10^{-6}–10^{12} Hz), dielectric spectroscopy can follow a wide variety of phenomena. However, since it requires either dipole rotation or charge transport, there are some molecular processes that may not be dielectrically active. In addition, analysis of the measured data can be difficult when there are different phenomena occurring simultaneously.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The study of dielectric properties in materials can be said to extend back over a 100 years. Some of the earliest measurements were made by Drude,[1] who, among other things, studied dielectric dispersion behavior of materials at optical frequencies. One of the first and best known works to relate experimental dielectric response to molecular phenomena was published by Debye,[2] who won the 1936 Nobel Prize in Chemistry for his work. Debye's text, along with texts such as those written by Smyth,[3] Smith,[4] Fröhlich,[5] and Böttcher et al.,[6,7] dealt primarily with dielectric behavior of small molecules. Considerable effort on the dielectric behavior of macromolecules was not expended until the 1950s and 1960s. Much of this effort was reviewed in the landmark texts by Von Hippel[8,9] and McCrum et al.[10] These texts are still required references for investigators of dielectric properties of polymers. Other more recent texts that specialize in the dielectric characterization of polymers include Hedvig,[11] Riande and Saiz,[12] Runt and Fitzgerald,[13] and Havriliak and Havriliak.[14]

Dielectric spectroscopy is now a well-established method not only for studying electronic properties, but also for providing fundamental understanding of molecular dynamics processes in polymers. In this respect, dielectric spectroscopy is analogous to dynamic mechanical analysis (DMA). However, instead of the mechanical oscillation employed by DMA, an electrical oscillation is applied to the sample. One of the primary advantages of dielectric spectroscopy over DMA and other techniques, is the extreme breadth of the frequency range (from $\sim 10^{-6}$ to $\sim 10^{12}$ Hz). Of course, a number of different experimental techniques are required in order to span this entire range. However, the 18 decades of timescale available allow for a wide variety of molecular processes to be probed. Dielectric analysis is one of the few analytical techniques that can follow transformations from low-viscosity liquids to rubbery solids to hard glassy solids. It can provide information such as mixture miscibility, onset of flow, molecular relaxations, thermal transitions, and polymerization reaction rates. Finally, dielectric measurements can be considerably more sensitive to changes in material properties than other thermal analysis techniques.

This article is concerned with dielectric spectroscopy applied to polymeric systems. It is not possible to include all the theoretical and experimental aspects, so only a broad overview is given. The interested reader may refer to the texts listed above or to the broad base of literature for more detailed information. The next section discusses various aspects of experimental instrumentation, including limitations and pitfalls. The following section covers data analysis and provides definitions and techniques that are commonly used in the dielectric community. The final section gives some examples of measurement applications for which dielectric spectroscopy is particularly effective.

2 INSTRUMENTATION AND SAMPLE PREPARATION

2.1 Instrument Basics

In dielectric spectroscopy, a sample is placed in contact with two or more electrodes and a time-varying voltage is applied across the electrodes. In most cases, these voltages and currents are converted into electrical impedances and the impedances are then related to sample permittivity by incorporating information about the electrode geometry. Because of the extreme frequency range over which dielectric measurements can be made, there is no single instrument that covers the entire range. Table 1 lists some of the more common types of dielectric instruments along with their typical frequency coverage.

In time-domain spectrometry, a voltage step is applied to the sample, and its response is measured as a function of time. Thus it is analogous to stress–relaxation measurements in mechanical analysis. Conversion to frequency domain data is accomplished via a Fourier transform. While the time-domain technique can extend to microwave frequencies, it is most accurate at very low frequencies (sub-hertz to tens of hertz). Unlike the step-function waveform used in time-domain spectrometry, the other measurement techniques listed in Table 1 use sinusoidal voltages to perturb the sample. These other techniques are therefore inherently frequency domain measurements. The impedance analysis technique can be sub-categorized into two primary methods: frequency response analyzers and impedance bridges. Frequency response analyzers use two phase-sensitive voltmeters to measure the voltage and current response of the sample. Impedance bridges use an adjustable compensation impedance to balance the impedance of the sample. Impedance analyzers are the most commonly used instruments for measurements on polymers. They provide reasonable accuracy in the frequency range where most polymer relaxations occur. Additional information on these techniques can be found in reviews by Kremer and Arndt[15] and Pochan et al.[16]

Table 1 Summary of the various types of dielectric measurement techniques

Time-domain spectrometer	$\sim 10^{-6}$–10^{10} Hz
Impedance analysis	$\sim 10^{-4}$–10^{7} Hz
Reflectometry	$\sim 10^{6}$–10^{9} Hz
Network analysis	$\sim 10^{7}$–10^{12} Hz

When accurate measurements are required at frequencies above 10 MHz, reflectometry and network analysis techniques are most appropriate. In reflectometry, the sample electrodes and the line connecting the electrodes to the instrument become important parts of the circuit and must be configured with great precision. The reflectometer uses microwave techniques to measure the reflection coefficient from the sample. Network analysis techniques also use microwave techniques to measure sample permittivity. However, instead of placing samples between electrodes, samples are placed in waveguides (e.g. rectangular or coaxial) or held in free space and their network scattering coefficients are measured. A numerical inversion algorithm must then be employed to calculate the permittivity from the scattering coefficients.[17]

2.2 Measured Parameters

Other than the very high-frequency microwave techniques, dielectric spectroscopy usually involves a sample that is in close proximity to a pair of electrodes. When a voltage is applied across the electrodes, an electric field exists within the sample and it becomes polarized in response to that field. The dielectric apparatus then measures the current generated by the sample. When the applied voltage is sinusoidal, the response current from the sample is also sinusoidal. Figure 1 illustrates such a voltage and current response for a hypothetical polymer sample. It shows that the measured current is the same frequency as the applied voltage, but it is shifted along the time axis. The quantities that relate the current response to dielectric material quantities are the phase shift, δ, and the relative change in amplitude. The real part of the waveforms for the voltage and current can be expressed by Equations (1) and (2):

$$\mathrm{Re}[V(t)] = V_0 \cos(wt) \tag{1}$$

$$\mathrm{Re}[I(t)] = I_0 \cos(wt + \delta) \tag{2}$$

where $V(t)$ and $I(t)$ are the voltage and current as a function of time, V_0 and I_0 are the amplitudes of $V(t)$ and

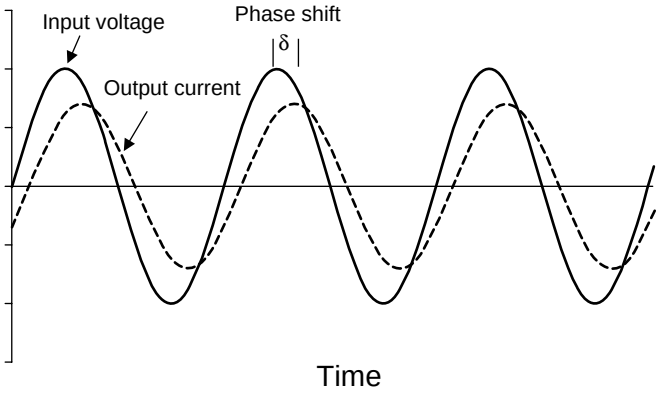

Figure 1 Voltage and current response of a hypothetical polymer material.

$I(t)$, and w is the angular frequency of the sinusoidal waveform. More generally, voltage and current are complex quantities, which are better expressed in terms of impedance (the ratio of voltage to current). In particular, the sample within the electrodes can be represented as a simple circuit element with a characteristic impedance, Z. Again, because the response is measured at a non-zero frequency, the impedance has both a real term and an imaginary term (Equation 3):

$$Z = \frac{V}{I} = Z' + iZ'' \tag{3}$$

where the prime indicates the real part and the double prime indicates the imaginary part. By incorporation of information about the electrode geometry, the impedance values can be converted to permittivities. A detailed description of the conversion of impedance to permittivity is provided in the next section, which covers specific electrode geometries.

Because an impedance bridge must use a specific circuit to balance the unknown impedance of the sample, polymers are usually modeled as being electrically equivalent to a resistor and capacitor connected in parallel to each other. This is shown in Figure 2. Other more complicated circuits have also been used to model dielectric behavior,[18] but the circuit shown in Figure 2 is appropriate for most polymer samples. It should be noted that a sample may also be modeled as a capacitor and resistor in series, and there is a mathematical equivalence between this series circuit and the parallel circuit of Figure 2.[10] Using elementary circuit analysis, the capacitance, C, and resistance, R, of Figure 2 are related to the complex impedance by Equation (4):

$$\frac{1}{Z} = \frac{1}{R} - iwC \tag{4}$$

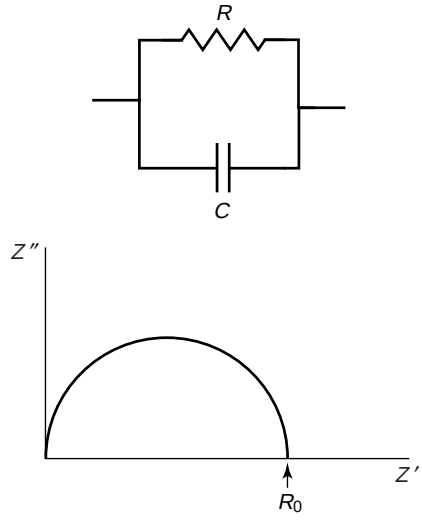

Figure 2 Parallel resistor and capacitor circuit used to model impedance of polymers and the complex impedance diagram that characterizes this circuit.

Suitable rearrangement of Equation (4) along with Equation (3) provides Equations (5) and (6):

$$Z' = \frac{R}{1 + w^2R^2C^2} \tag{5}$$

$$Z'' = \frac{wR^2C}{1 + w^2R^2C^2} \tag{6}$$

Combining Equations (5) and (6) give the following useful result (Equation 7):

$$\left(Z' - \frac{R}{2}\right)^2 + (Z'')^2 = \frac{R^2}{4} \tag{7}$$

which is the equation of a circle of radius $R/2$ and centered at $Z' = R/2$ and $Z'' = 0$. The upper half of this circle (corresponding to $Z'' > 0$) is shown in Figure 2. The semicircle in Figure 2 assumes that C and R are constant as a function of frequency. It turns out that real data have more of a distribution of C and R values which result in semicircles that are skewed and that have centers somewhat below the Z'-axis. As will be shown later, complex impedance plots such as that shown in Figure 2 are useful for providing insight into the processes that contribute to a particular sample's dielectric behavior. They are also useful for situations where complicated electrode geometries must be used so that quantitative calculation of permittivity is not practical.

2.3 Electrode–Sample Configurations

By far the most commonly used electrode geometry is the parallel-plate arrangement. Two planar electrodes are used to sandwich a thin, flat sample, as shown in

Figure 3. The parallel-plate geometry is popular because of the simplicity with which dielectric permittivity can be calculated. Using the area of the electrodes, A, and the spacing between them, d, the permittivity is a simple function of the measured impedance. The equations relating impedance to permittivity are summarized in the data analysis section.

Because of the finite size of the electrode, there can be an adverse effect on the measured impedance caused by fringe fields along the perimeter of electrodes. Designing the electrode spacing to be much smaller than the horizontal dimensions can minimize this. Incorporating a guard ring into one of the electrodes can also minimize it. A typical guard ring configuration is shown in Figure 3. As Von Hippel pointed out,[9] the most effective guard electrode has a width at least twice the thickness of the specimen and the unguarded electrode must extend to the outer edge of the guard. Even with a guarded electrode, it is still best to calibrate the electrode area by first measuring the capacitance in air or vacuum (i.e. without the sample). However, this is not always practical.

When calculating permittivity from unguarded parallel-plate electrode measurements, analytical corrections can be applied to the data. These corrections are described in detail in ASTM D 150.[19] To insure quantitative measurements, the samples should be flat and make good contact with the electrodes. In particular, if there is an air gap between the electrode and the sample, the effective capacitance decreases causing a shift in the calculated permittivity. The electrodes can be constructed of metal and placed in contact with the sample, or they may be evaporated, sputtered, or painted directly onto the sample. There are a number of ASTM standards that provide additional guidance for electrode construction and application.[19–21]

An electrode geometry that is similar to the parallel plate is the concentric cylinder configuration, pictured in Figure 4. These electrodes are useful for measuring liquids such as low molecular weight monomers, polymer emulsions, or polymers dissolved in solvents. As for the parallel-plate electrodes, calibration can be done by measuring the empty cell. Alternatively, calibration can be accomplished by measuring the cell filled with a nonpolar liquid with a known dielectric constant. For liquids, it is also common to use a four-probe arrangement where the electrodes are arranged with two outer electrodes inducing the sinusoidal voltage and two inner electrodes measuring the response. This four-probe electrode is useful when ionic mobility is high enough to cause anomalous polarization effects. These effects, also called electrode-blocking effects, are discussed in more detail in subsequent sections.

A third type of electrode that is useful for measuring polymer systems is the interdigitated comb electrode. This type of electrode was first introduced by Armstrong[22] and greatly refined by Senturia et al.[23,24] and Kranbuehl.[25] A simple schematic diagram of such an electrode is shown in Figure 5. Unlike the parallel-plate electrodes, the geometry of comb electrodes is fairly complicated. In addition, there are large fringe fields that exist in the substrate of the electrodes. Hence numerical methods are required for the quantitative calculation of dielectric permittivity.[26] Their primary advantage is that they can be mounted noninvasively on the surface

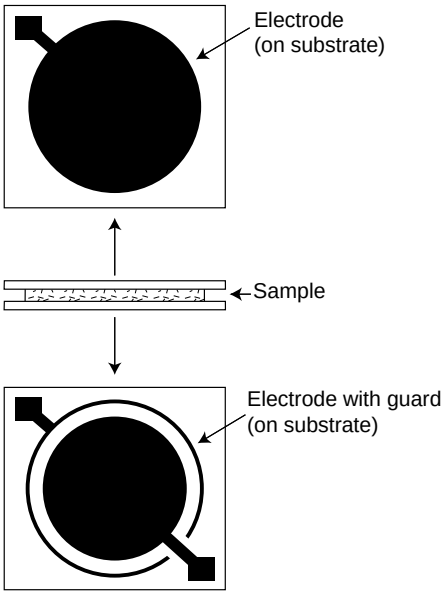

Figure 3 Parallel-plate electrodes. Also shown is a guard ring used to minimize fringe field effects.

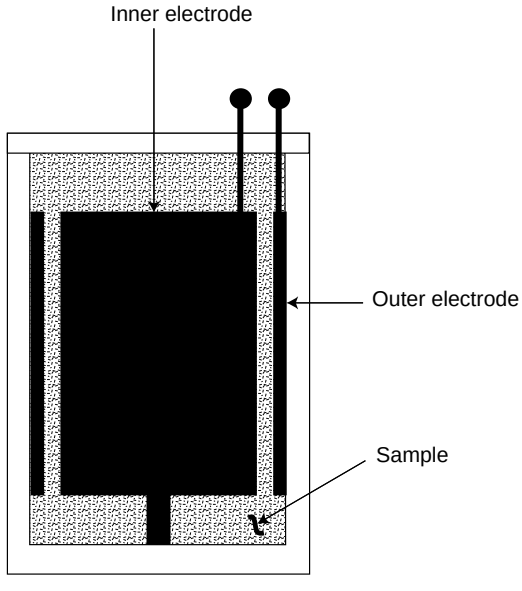

Figure 4 Cylindrical electrodes.

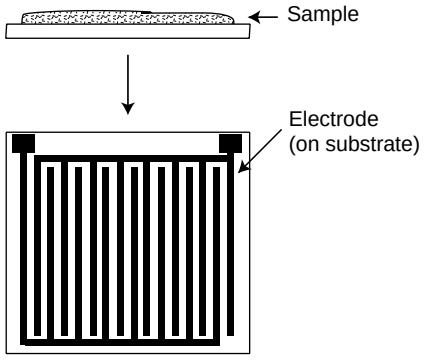

Figure 5 Interdigitated comb electrodes.

of a structure, or can be easily incorporated into laminated structures such as fiber-reinforced composites or adhesive joints.[27] Often valuable information such as molecular relaxation or changes in material properties can be monitored with just the raw impedance data, so that quantitative calculation of permittivity is not necessary.

2.4 Potential Pitfalls

An important aspect of the experimental measurement of materials is the separation of instrument effects from real material behavior. Unfortunately, the scientific literature is full of conclusions that incorrectly attribute instrumental artifacts to the material being measured. Often experimental details, important for correct interpretation of the data, were overlooked during the original research. This section briefly lists the assumptions and instrument effects that can be important when making dielectric measurements.

When making measurements on a sample, two fundamental assumptions are usually made. The first is that the measurement is linear. In other words, the measured impedance is independent of applied voltage. For most conditions, this assumption is true. However, it is possible for dielectrics to experience breakdown at high enough voltages ($>10^6\,\mathrm{V\,cm^{-1}}$). In addition, the measured behavior becomes nonlinear when there are electrochemical reactions resulting from the applied voltage. Again, this is usually restricted to high voltages, and generally does not occur in most polymers systems of interest. The second assumption made when interpreting dielectric data is that the measured property is time invariant. Because the measurement is usually done to follow changes in sample properties as a function of time or temperature, time invariance is not strictly followed. Hence the question of time invariance becomes one of whether the sample is constant within the time it takes to measure the impedance at a given frequency. Typically, most instruments average over several cycles. Hence, the time for a given measurement will take several times the inverse of the frequency. For high frequencies, this measurement time is still less than a fraction of a second. However, for frequencies below 1 Hz, this time can become significant. This is important in experiments such as measurement of changes during chemical reactions.

Another important effect that can occur in ionically conductive polymers is the blocking of ions by the electrodes. Blocking electrode effects can obscure the bulk properties of the sample. This blocking effect is caused by the pile-up of the sample's mobile ions at the electrode/sample interfaces and is especially prevalent in samples where ion mobility is high (e.g. low molecular weight samples or samples at high temperature).[28,29] As the ions accumulate at the electrodes, the sample becomes polarized, so that a large false contribution to the dielectric constant is measured. Besides the sample itself, two major experimental factors may be varied to control the electrode polarization. The first factor is the timescale. As the length of time for the experimental measurement is increased, the ions have more time to accumulate at the electrodes. Thus the blocking effect is minimized by taking data at higher frequencies (shorter timescales). The specific frequency at which blocking becomes important depends on the concentration of ions and the sample viscosity. The second factor that influences electrode polarization is the sample geometry. As the electrode separation decreases, the amount of ions that pile up in a given amount of time increases. Hence increasing the spacing between electrodes can be used to minimize blocking effects. However, increasing the distance between electrodes can have other adverse effects, such as decreased sensitivity and increased fringe fields. Thus optimization of the electrode spacing is often an exercise in compromise, which depends on the sample properties.

Because blocking effects can cause large "false" contributions to the measured data, it is important to be able to detect when this blocking occurs. In order to account for the pileup of ions at the electrodes, the parallel resistor and capacitor model of Figure 2 must be modified. Although more complicated models of blocking effects exist,[18] the simplest model is made by adding a second capacitance in series with the original circuit, as pictured in Figure 6. For this model, the complex impedance is then given by Equations (8a) and (8b):

$$Z' = \frac{R_{\mathrm{bulk}}}{1 + w^2 R_{\mathrm{bulk}}^2 C_{\mathrm{bulk}}^2} \tag{8a}$$

$$Z'' = \frac{w R_{\mathrm{bulk}}^2 C_{\mathrm{bulk}}}{1 + w^2 R_{\mathrm{bulk}}^2 C_{\mathrm{bulk}}^2} + \frac{1}{w C_{\mathrm{blocking}}} \tag{8b}$$

where the subscript bulk indicates material-dependent properties and the subscript blocking indicates the anomalous blocking effects. As in the nonblocking case, Equations (8a) and (8b) can be plotted on a complex

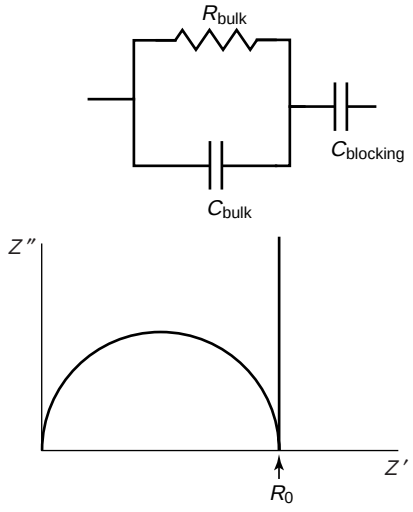

Figure 6 Circuit used to model impedance of polymers that include ion-blocking effects at the electrodes. Also shown is the complex impedance diagram that characterizes this circuit. The ''tail'' is indicative of data that is dominated by the electrode blocking effects.

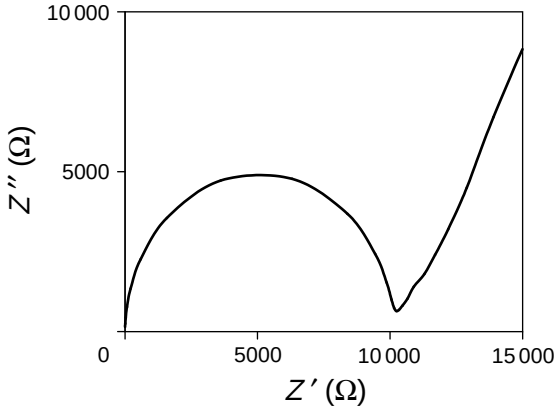

Figure 7 Complex impedance of a polymer latex emulsion that demonstrates the qualitative validity of the circuit model illustrated in Figure 6 for diagnosing ionic blocking problems.

impedance plot and this is also shown in Figure 6. Again, there is a semicircle that represents the impedance of the material being tested, but also, there is a vertical line that is caused by the ions blocking the electrode. R_0 is the resistance of the sample in the limit of zero frequency. Pictured in Figure 7 for comparison are some actual data from a polymer latex emulsion. These data show the usefulness of the model in diagnosing the presence of electrode blocking effects.

The blocking effect is very much geometry dependent, so interpreting electrode-blocked data is difficult if not useless. Because it has to do with how fast the ions can accumulate at the electrode, it is also time (or frequency) dependent. In Figure 7, each data point corresponds to a different measurement frequency. In these data, there are electrode blocking effects at the lower frequencies where the ions have more time to accumulate. However, at higher frequencies, electrode blocking is no longer a problem, and the data are a valid measure of the material characteristics.

One final difficulty that can adversely affect measurement accuracy is the existence of parasitic impedances. Even the most carefully designed electrode fixture will have finite impedances that result from the wires and electrodes that connect the sample to the measurement instrument. These impedances can become important at frequencies above a few hundred kilohertz. Figure 8 shows a schematic diagram of the undesired impedances that can occur. Jonscher[30] has noted that in conductive samples, parasitic inductances in the electrodes can lead to inaccuracies that become worse as the frequency is increased. Schultz[31] has shown that high frequency anomalies can occur from both inductances and resistances that are present in the sample fixtures.

2.5 Thermally Stimulated Current

One final instrumentation topic that should be mentioned is the subject of thermally stimulated current (TSC) spectroscopy. This technique is different than traditional dielectric spectroscopy in that it is a dc (zero-frequency) technique. A polymer sample is placed between two electrodes and a dc voltage is applied in order to polarize the polymer. It is then cooled and the external field is removed. Then, as the temperature is slowly ramped up again, the depolarization current is measured across the same electrodes. This current corresponds to dipole relaxation. There are also variations of this procedure, but the main concept is that the measured current is a probe of the dipole relaxation. For example, if a polymer glass transition is to be measured, the voltage would be applied at a temperature above the glass transition, then, by cooling the sample into the glassy state, the dipole polarization is locked in. As the polymer is heated back through the transition, the dipoles slowly lose their orientation causing a small but measurable current.

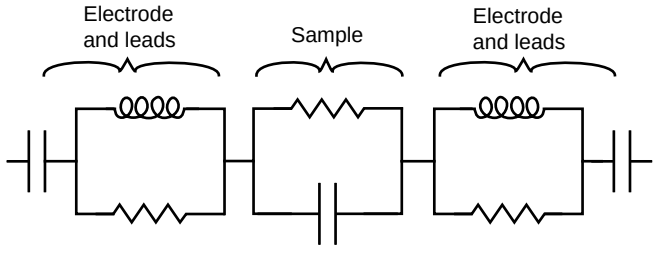

Figure 8 Circuit diagram showing R, L, and C circuit elements from both the sample and electrode/leads that contribute to the overall response.

An advantage of this technique over traditional dielectric spectroscopy is that the dc conditions provide better separation of multiple relaxation peaks. The disadvantage is that the theory and data analysis can be considerably more cumbersome. Comprehensive reviews of TSC spectroscopy have been given by Lavergne and Lacabanne[32] and Ibar et al.[33]

3 DATA ANALYSIS

3.1 Calculated Parameters

Dielectric spectroscopy measures the effect of microscopic phenomena on macroscopic properties. For a material such as a polymer, an applied electric field induces an electric polarization of the material. Usually the magnitude of this polarization is linearly proportional to the applied field. The proportionality constant in this case is called the permittivity and is designated by the symbol ε. A fundamental constant of nature is the permittivity of a vacuum, $\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{J^{-1}\,C^2\,m^{-1}}$. Usually the permittivity is expressed as the relative permittivity, which is the ratio of the material permittivity to the permittivity of a vacuum (Equation 9):

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \tag{9}$$

This can often be a source of confusion since it is common to drop the subscript r from the symbol for relative permittivity. It is usually up to the reader to infer whether ε means permittivity or relative permittivity based on its context. To add to the confusion, the relative permittivity is sometimes also called the ''permittivity'' or the ''dielectric constant''.

In a time-varying or oscillating electric field, the permittivity is best represented by a complex number (Equation 10):

$$\varepsilon = \varepsilon' - i\varepsilon'' \tag{10}$$

In this notation, ε' is the real part of the permittivity, and is often called ''permittivity'' for short; ε'' is the imaginary part of the permittivity, and is usually called the ''dielectric loss factor''. The loss factor is usually associated with energy absorption by the polymer.

Another quantity that is associated with the energy absorption by the sample is the ''loss tangent'' (Equation 11):

$$\tan\delta = \frac{\varepsilon''}{\varepsilon'} \tag{11}$$

This loss tangent is a convenient quantity for qualitative comparison to dynamic mechanical data since it is directly analogous to the mechanical loss tangent. It is also directly calculable from the measured impedance data without knowing the geometrical dimensions of the electrodes. Often it is desirable to compare the permittivity to dynamic mechanical modulus data. However, since it is not directly analogous, a dielectric modulus, M, is usually defined (Equation 12):

$$M = \frac{1}{\varepsilon} \tag{12}$$

Like permittivity, Equation (12) can be separated into real and imaginary dielectric moduli (M' and M'', respectively). In the case where conduction processes are occurring, it is possible to calculate an ''apparent conductivity'' also. This quantity is usually calculated from the dielectric loss factor (Equation 13):

$$\sigma = w\varepsilon'' = w\varepsilon_0\varepsilon_r'' \tag{13}$$

where w is the angular frequency ($w = 2\pi f$).

In order to calculate the quantities defined in Equations (10–12), information about the electrode geometry must be incorporated. Table 2 lists the applicable equations for both parallel-plate electrodes and cylindrical electrodes.

3.2 Origins of Dielectric Properties in Polymers

When analyzing dielectric data, there are two physical phenomena that contribute to the values of complex permittivity: dipole reorientation and conduction. In simple terms, dipoles are manifested by charge asymmetries that exist in molecules or molecular fragments. A molecular fragment that has at least two different atoms can have

Table 2 Calculation of permittivity and loss tangent from measured impedance data

Electrode	Parameter	Equation[a]
Parallel-plate electrode	Dielectric permittivity	$\varepsilon_r' = \dfrac{d}{2\pi f\varepsilon_0\lvert Z\rvert^2 A}Z''$
	Dielectric loss factor	$\varepsilon_r'' = \dfrac{d}{2\pi f\varepsilon_0\lvert Z\rvert^2 A}Z'$
	Loss tangent	$\tan\delta = \dfrac{\varepsilon_r''}{\varepsilon_r'} = \dfrac{Z'}{Z''}$
Concentric cylinder electrode	Dielectric permittivity	$\varepsilon_r' = \dfrac{\ln(b/a)}{4\pi^2 f\varepsilon_0\lvert Z\rvert^2 l}Z''$
	Dielectric loss factor	$\varepsilon_r'' = \dfrac{\ln(b/a)}{4\pi^2 f\varepsilon_0\lvert Z\rvert^2 l}Z'$
	Loss tangent	$\tan\delta = \dfrac{\varepsilon_r''}{\varepsilon_r'} = \dfrac{Z'}{Z''}$

[a] A = area of smallest electrode; d = distance between electrodes; l = length; a = inner electrode radius; b = outer electrode radius.

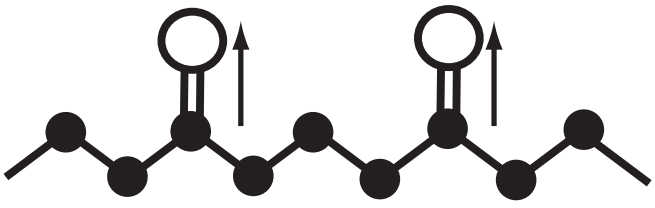

Figure 9 A portion of a hypothetical polymer chain with molecular fragments that have dipole moments.

an intrinsic dipole moment because different atoms have different affinities for electric charge.[12] Example dipoles are illustrated in Figure 9, which shows a hypothetical polymer fragment. Indicated by arrows are the dipole moments, which are vectors that describe the charge distribution in units of [charge × displacement]. Polymer chains are usually made up of thousands of bonds that can rotate in response to external stimuli. Thus in an externally applied electric field, the dipole fragments will realign themselves to partially cancel the effects of the electric field. This is shown schematically in Figure 10(a). The time that it takes for the dipole to realign itself varies according to the properties of the polymer and external conditions such as temperature and pressure. Thus a polymer can be characterized by measuring the relaxation time (or more precisely the distribution of relaxation times) of the intrinsic dipoles. While Figure 10(a) shows an idealized diagram of dipole orientation, in reality steric hindrances prevent dipoles from fully reorienting in response to an external field. This is especially true in semicrystalline and glassy amorphous polymers where the polymer chains are much more rigidly fixed so that dipoles reorient only very slightly, resulting in a lower dielectric constant and loss factor. Conversely, polymers in a rubbery amorphous or viscous flow state have higher dipole mobility and therefore higher dielectric constants and loss factors.

While Figure 9 shows an example of a polymer with permanent dipoles, it is also possible to induce dipoles in polymers that do not have intrinsic permanent dipole moments. In essence, an applied electric field causes the electron cloud to shift relative to the nuclei, thus creating a spatial charge imbalance. Similar to the behavior of permanent dipole moments, this ability of an external field to ''polarize'' the molecular structure is dependent on the chemical properties of the polymer and its environment.

The second way in which a material responds to an electric field, conduction, involves the physical translation of charged species. Charged species can be either ions or electrons. When an electric field is applied, opposites attract, so a positively charged species will be attracted to the negative electrode, while the negatively charged species will be attracted to the positive electrode. More precisely, the applied electric field perturbs the

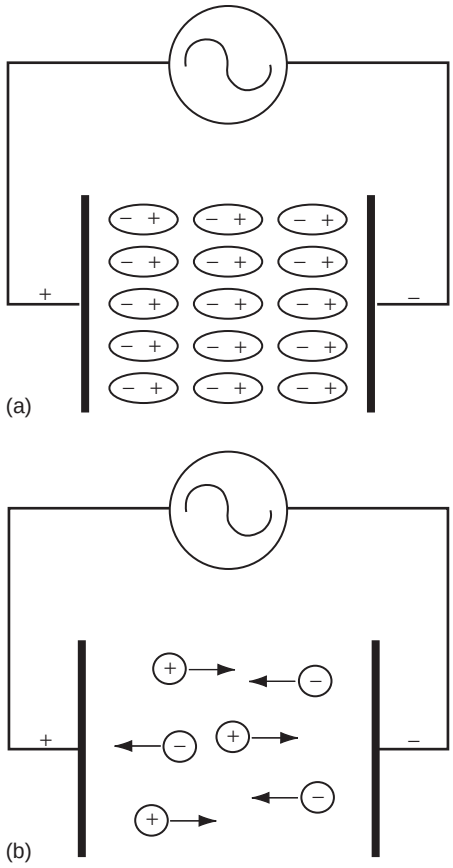

Figure 10 Effect of external electric field on (a) dipole moments and (b) charged species.

Brownian motion of the ions so that they tend to drift towards oppositely charged electrodes. This is illustrated in Figure 10(b). While there has recently been considerable development of electrically conductive polymers, most conduction phenomena in polymers are caused by the presence of ions. Often, ions are simply unwanted impurities in the sample. Since they make up such a low percentage of the material, they are often neglected. However, if a polymer has no intrinsic dipole moments, then the dielectric response can often be dominated by these ion impurities. Starkweather et al. have pointed out that not realizing the importance of ion conduction in these types of systems has led to the occasional misinterpretation of polymer dielectric response, particularly in fluoropolymers.[34]

Similarly to dipole reorientation, ion (and also electron) conduction is also affected by various chemical and environmental variables. As the charged species travel towards an electrode, they are slowed by their surroundings. The slower the ions travel, the more *resistant* the material; the faster the ions travel, the more *conductive* the material. Unfortunately, dielectric spectroscopy measures the cumulative response of both dipole

reorientation and conduction. Sometimes the effects of the two phenomena overlap so that one cannot be distinguished from the other.

3.3 Relaxation

One of the primary functions of dielectric spectroscopy is to characterize relaxation phenomena in polymers. In many polymers there are a number of different relaxations that are detectable by dielectric spectroscopy. In amorphous polymers, the main transition is the glass transition and it is usually labeled as the ''α-transition''. Additional relaxations can often be detected at lower temperatures/higher frequencies and these may be labeled β, γ, and δ, in order of progression. Although there is still considerable debate as to the exact nature of the glass transition, it is generally attributed to cooperative motions of polymer molecules.[35] Conversely, sub-T_g relaxations are usually attributed to specific local motions of molecular segments. In semicrystalline polymers, there is still an amorphous phase, so there are still glass and sub-glass relaxations analogous to wholly amorphous polymers. In addition, there is a high-temperature process that is related to the crystalline phase.

Relaxations are manifested by transitions in the polymer properties (e.g. glass to rubber), so they play an important role in the usefulness of a polymer for a given application. Relaxations are typically temperature dependent. However, as in the case of the glass transition, they can also be time or frequency dependent. Hence it is possible to analyze dielectric relaxation data with respect to either temperature or frequency. In terms of temperature, the glass transition is adequately fitted by the well-known Williams–Landel–Ferry (WLF) equation[10,36] (Equation 14):

$$\log \frac{f(T)}{f(T_g)} = \frac{-C_1(T - T_g)}{C_2 + (T - T_g)} \tag{14}$$

where $f(T)$ is the relaxation rate at temperature T, and C_1 and C_2 are fitted constants. Equation (14) has been used not only for dielectric spectroscopy, but also for various dynamic mechanical and rheological data. Alternatively, the glass transition can be modeled with the Vogel–Fulcher equation[37,38] (Equation 15):

$$\log f(T) = A - \frac{B}{T - T_0} \tag{15}$$

where T_0 is the Vogel temperature (usually 30–70 °C below T_g) and A and B are fitted constants. The form of Equation (15) can be obtained by algebraic rearrangement of the WLF equation. For sub-glass relaxations, the relaxation rate is often better approximated with an Arrhenius relationship (Equation 16):

$$f(T) = C \exp\left(\frac{-E_a}{kT}\right. \tag{16}$$

where C is a fitted parameter and E_a is the apparent activation energy. Equations (14–16) are expressed in terms of relaxation rate; however, they can just as easily be written in terms of relaxation time (the inverse of relaxation rate).

For frequency domain data, there are several models that have been applied and these are briefly discussed in the remainder of this section. For more in-depth information, the interested reader should consult recent reviews by Havriliak and Havriliak.[14,39] Perhaps the simplest model for describing dielectric relaxation is the Debye model, which is based on a ''hindered'' reorientation of dipoles[2] (Equation 17):

$$\varepsilon_r = \varepsilon_U + \frac{\varepsilon_R - \varepsilon_U}{1 + iw\tau} \tag{17}$$

where ε_R and ε_U are the ''relaxed'' and ''unrelaxed'' permittivities, respectively, and τ is the characteristic relaxation time for reorientation of the dipoles. As mentioned in the previous section, both dipole reorientation and conduction mechanisms can be involved in the measured dielectric response so a term that includes conduction is often incorporated in the Debye model. Including this conduction term and splitting into real and imaginary components gives Equations (18) and (19):

$$\varepsilon'_r = \varepsilon_U + \frac{\varepsilon_R - \varepsilon_U}{1 + (w\tau)^2} \tag{18}$$

$$\varepsilon''_r = \frac{\sigma}{w\varepsilon_0} + \frac{(\varepsilon_R - \varepsilon_U)w\tau}{1 + (w\tau)^2} \tag{19}$$

These equations are plotted in Figure 11(a). The curves plotted correspond to the case where there is also some conduction (i.e. $\sigma > 0$), resulting in a low-frequency tail in the dielectric loss. Ignoring for now the conduction, the permittivity data show that ε_R corresponds to the relative permittivity at low frequencies when the dipoles can attain their maximum reorientation and ε_U is the permittivity at high frequencies when the dipoles are unable to keep up with the external field. The loss factor data, which are associated with energy absorption or damping by the material, indicate a peak at $w = 1/\tau$, which is the middle of the transition from ε_R to ε_U. While this model provides important insight into the nature of relaxation phenomena in polymers, it is qualitative at best. In fact, most relaxation phenomena do not have a single relaxation time, τ, as suggested by this model. Rather, they may have a distribution of relaxation times which makes modeling somewhat more complex.

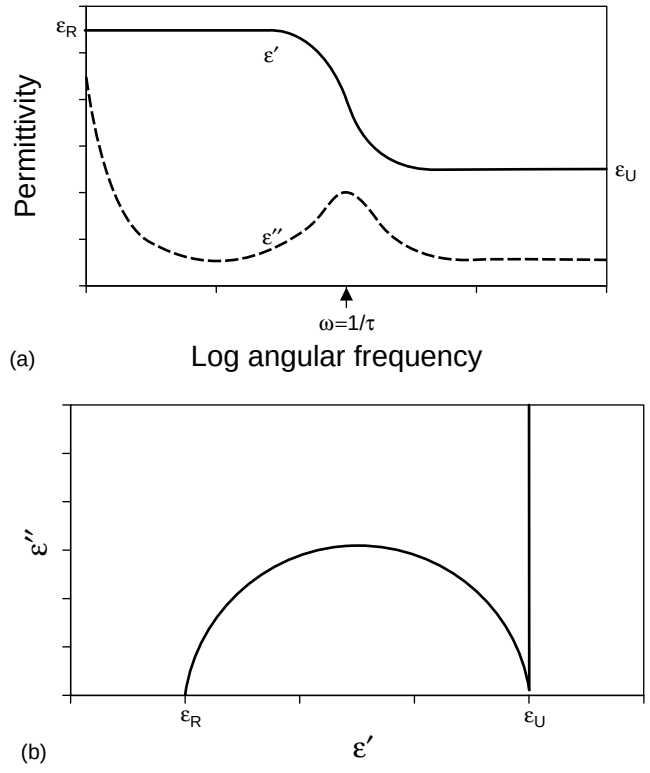

Figure 11 (a) Plot of permittivity and loss factor for the Debye model (these data are plotted as a function of normalized frequency, $w\tau$). (b) Cole–Cole plot of the same data. The ''tail'' in complex permittivity plots indicated ion conduction.

Because few systems follow Debye behavior quantitatively, there have been a number of empirical generalizations made on Equations (17–19). One of the first was a generalization proposed by Cole and Cole.[40] Ignoring for now the contributions from ionic conductivity, Equation (17) can be re-written with an empirical fitting factor, α (Equation 20):

$$\varepsilon_r = \varepsilon_U + \frac{\varepsilon_R - \varepsilon_U}{1 + (iw\tau)^{1-\alpha}} \tag{20}$$

In addition to proposing this model, Cole and Cole also demonstrated the usefulness of a complex plane plot for representing the real and imaginary permittivity. These plots are commonly called Cole–Cole plots. They are similar to the complex impedance plots presented in the earlier section on instrumentation and sample preparation. They represent the data by plotting the imaginary permittivity, ε'' against the real permittivity, ε'. Figure 11(b) shows Cole–Cole plot representations of both the Debye model (Equation 17) and Cole and Cole's generalization (Equation 20). Whereas the Debye model data show a perfect semicircle, the Cole–Cole model allows for a compressed semicircle with its center below the ε' axis.

In addition to Cole and Cole, various researchers have suggested additional generalizations to the relaxation equations. For example, Fuoss and Kirkwood,[41] Davidson and Cole,[42] Jackle,[43] Williams and Watts,[44] Jonscher,[45] and Havriliak and Negami[46,47] have all proposed various generalizations to Equation (17). These equations are empirical in that they include one or two parameters for better fitting measured dielectric data. These equations are also nonlinear with respect to frequency, so nonlinear numerical regression methods must be applied in order to fit measured data to these various functions. If an investigator chooses to fit measured data to one of these models, then the choice of the ''best'' empirical model can depend on the nature of the dielectric relaxation being studied. In a recent review, Havriliak and Havriliak[48] have shown that the Havriliak–Negami equation does fit a wide range of data. It includes two empirical parameters, α and β, which allow for a Cole–Cole type of plot to have both flatness and unsymmetrical skew (Equation 21):

$$\varepsilon_r = \varepsilon_U + \frac{\varepsilon_R - \varepsilon_U}{[1 + (iw\tau)^{\alpha}]^{\beta}} \tag{21}$$

Havriliak and Havriliak also pointed out[48] the importance in providing details about the analysis methods used to fit measured data to a relaxation equation (i.e. convergence criteria, confidence intervals, etc.). In order to gain a better understanding of the sensitivity of the various fitted parameters, it is useful to use numerical experimentation. Such experiments can be accomplished by constructing additional input data sets with ''noise'' that reflects the measurement uncertainties of the original data. Only when these kinds of studies are done can data be intelligently understood in terms of the relaxation models.

4 APPLICATIONS AND METHODS

4.1 Detecting Relaxations

One of the most important aspects to characterizing the physical and mechanical properties of a polymer is its relaxation behavior. For an amorphous polymer, the glass transition temperature will determine if the polymer is to be a stiff glass or a soft rubber at its use temperature. When a polymer is semicrystalline, most engineering applications require that the polymer be used well below its melting temperature. Hence the positions of these transition temperatures are important also for optimizing processing conditions when manufacturing parts.

Other more subtle properties are also determined by relaxation behavior. For example, if a polymer is to be used for absorbing vibration, then it must have

a large amount of damping at the temperature and frequency of use. Conversely, if a polymer is to be used in applications such as transmission of electromagnetic energy, then the damping or loss should be minimized. Since relaxations are associated with energy absorption, the relative positions and intensities of a polymer's relaxation transitions are of great importance in these applications.

Dielectric spectroscopy can be particularly sensitive to molecular relaxations in polymers. Figure 12 shows the glass transition for amorphous PET, measured at 1 kHz.[49] The real part of the permittivity at low temperature is approximately 2.5, but then undergoes a step transition to a value of ~3 at high temperature. Thus, when the polymer is glassy, the dipoles are greatly inhibited in their motion and the permittivity has a relatively low value. When the polymer has softened, the dipoles can reorient much more easily and the permittivity can attain a much higher value. Figure 12 also shows the loss factor which is peaked at the center of the glass transition (~120 °C). The loss factor, which is indicative of the amount of damping or energy absorption by the polymer, shows that the polymer absorbs the most energy at a temperature that corresponds to the middle of the transition. The data in Figure 12 are permittivity as a function of temperature. It is also possible to obtain data at a single temperature as a function of frequency. The frequency-dependent data will show very similar features to the idealized Debye model in Figure 11(a). However, a better fit to the frequency-dependent data is usually obtained when a more general but semiempirical equation is used, such as the Havriliak–Negami equation (Equation 21). A useful resource that contains tables of dielectric relaxation parameters for a wide variety of polymers is the review by Havriliak and Havriliak.[14]

The electrical relaxation behavior measured by dielectric spectroscopy is analogous to the mechanical relaxation behavior measured by DMA. Both techniques show a step transition in the real part of the data and a peak in the imaginary part. However, whereas the real part of the dielectric permittivity goes from a low to a high value, the real part of the dynamic mechanical modulus goes in the opposite direction. Therefore, when dielectric data are compared with DMA data, it is common practice to convert the DMA data into mechanical compliance. Compliance as a function of frequency can also be fitted to relaxation equations such as the Havriliak–Negami function. Alternatively, the permittivity can be converted into "dielectric modulus" as defined in Equation (12). The dielectric modulus for PET measured at 1 kHz is shown in Figure 13. It should be noted that although the dielectric modulus is qualitatively analogous to mechanical modulus, it is by no means quantitatively equivalent. Furthermore, it is possible to have relaxations that are dielectrically strong but mechanically weak. Conversely, it is possible to have relaxations that are dielectrically weak or inactive, but mechanically strong. Hence dielectric spectroscopy and mechanical spectroscopy are complementary techniques for measuring the molecular relaxations of polymers.

The data in Figure 12 show only a single molecular relaxation, but it is possible for polymers to exhibit multiple relaxations. When measured over a broad temperature range or a broad frequency range, a polymer may show not only a glass transition (T_g), but also sub-T_g relaxations and a melting transition. While the relaxation spectra may be thought of as a "fingerprint" of a particular polymer, it is dependent on more than just the chemical structure. Factors such as morphology and thermal history can have a profound effect on the relaxation spectra, so they should be carefully considered or controlled

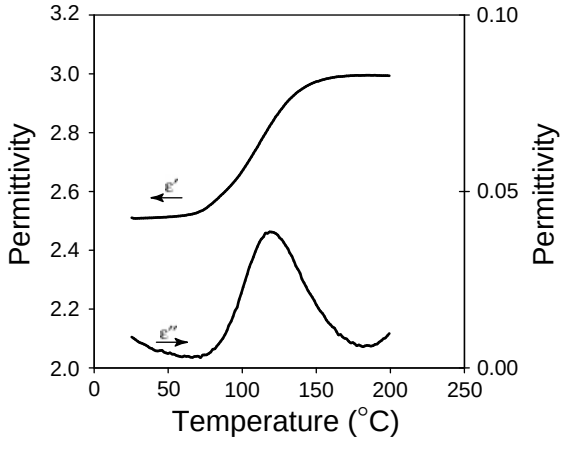

Figure 12 Glass transition of amorphous poly(ethylene terephthalate) (PET): permittivity.

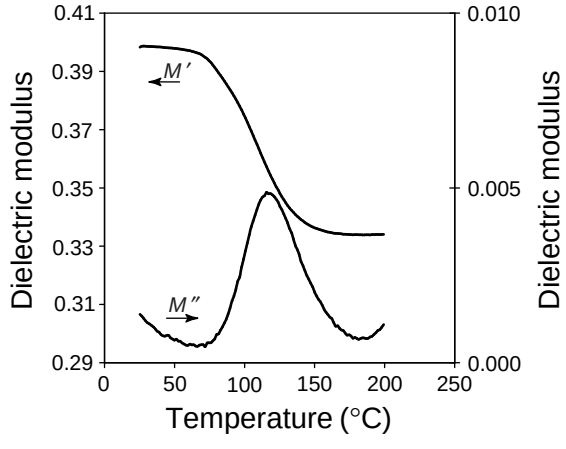

Figure 13 Glass transition of amorphous PET: dielectric modulus.

when making measurements. For example, the relaxation peak positions and intensities are dependent on the temperature scan rate of the dielectric experiment, and also the thermal history of the polymer sample before the experiment was even begun.

Another instance in which there may be multiple relaxations is when a polymer is actually a heterogeneous mixture. Polymer mixtures or blends have become prevalent in modern engineering applications because they can have useful mechanical properties such as increased toughness. Most of these blends are multiphase, meaning that under a microscope, regions that are rich in one polymer are visibly separate from regions that are rich in a different polymer. It is therefore possible for a blend to have multiple glass transitions, each corresponding to a separate phase. There may be some limited miscibility, however, and the result of this miscibility is that there will be a corresponding shift in the position of each glass transition. On the other hand, if a blend is completely miscible, then there will only be a single phase, and therefore only a single glass transition.

The dielectric spectroscopy of partially miscible blends has been studied by Rellick and Runt,[50] and Figure 14 shows some representative data. Plotted are loss tangent curves of a blend of EVA and PVC after different thermal histories. Also shown are vertical lines that correspond to the glass transition temperatures of pure EVA and pure PVC. Because the blend is only partially miscible, the two separate T_gs are still evident, but they are each shifted by a finite amount from their pure values. The degree of mixing in each of these two phases can be calculated by applying the well-known Fox equation[51] (Equation 22):

$$\frac{1}{T_g} = \frac{w_1}{T_{g_1}} = \frac{w_2}{T_{g_2}} \tag{22}$$

where w_1 and w_2 are the respective weight fractions of the two components, T_{g_1} and T_{g_2} are their respective glass transition temperatures (i.e. in pure form), and T_g is the shifted glass transition temperature. Equation (22) is valid for homogeneous blends, but the example in Figure 14 is clearly a two-phase blend. However, if each phase is considered separately, then application of Equation (22) is still valid. Thus, if T_{g_1} and T_{g_2} are known, and since $w_2 = 1 - w_1$, the weight fractions in each phase can be calculated from the shifted glass transition temperatures.

One final aspect of the data in Figure 14 is the fact that the glass transitions shift more and more towards a central value as the mixture is held at higher temperature. Curve (d), which was blended without any supplementary heating, shows very little shift in the respective glass transitions. On the opposite end, the sample of curve (a) was heated at 105 °C for 12 h, and shows a much more pronounced shift in T_g values. Thus heating caused a migration of the polymer chains so that the two phases changed their compositions. Note that the data in Figure 14 are for amorphous polymers. When the polymers are semicrystalline, this technique only indicates the weight fractions in the respective amorphous phases.

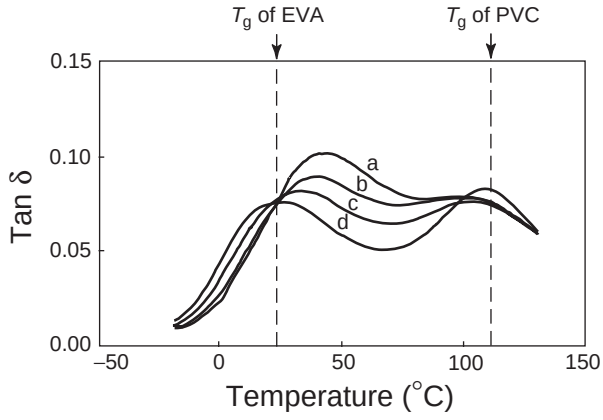

Figure 14 Loss tangent curves at 100 kHz for 50% poly(ethylene-co-vinyl acetate) (EVA) and 50% poly(vinyl chloride) (PVC) blends with different thermal histories: (a) no heating; (b) 55 °C for 24 h; (c) 70 °C for 24 h; and (d) 105 °C for 12 h. [Reproduced from G.S. Rellick, J. Runt, *J. Polym. Sci., Polym. Phys. Ed.*, **24**, 279–302 (1986), Copyright 1986 John Wiley & Sons Inc. Reprinted by permission of John Wiley & Sons Inc.]

4.2 Cure and Process Monitoring

In the last 40 years, there has been substantial work published about dielectric measurements for the monitoring of polymer cure. Much of this work has been stimulated by the increasing importance of polymers and polymer matrix composites in engineering applications. Because of this extensive development, the dielectric technique has been refined to be more than just a laboratory measurement. With suitable electrodes it can be used to monitor polymerization in situ, particularly in thermosets.[52] Thus it is often used as a process control variable in industrial settings such as in batch reactors, presses, autoclaves, ovens, and molding operations.[27,53–55] Dielectric spectroscopy has even been used to monitor the ultraviolet cure of photopolymers.[56] Using data from dielectric sensors combined with sophisticated process models to control actively these various polymer processes is an active area of research.[57]

The dielectric monitoring of cure involves the measurement of the change of ion and dipole mobility in the resin as it polymerizes. The ion and dipole mobility is directly related to the viscosity of the resin, which is, in turn, a function of molecular weight or degree of cure. Dielectric spectroscopy has two distinct advantages over direct mechanical measurement of the viscosity. First, a mechanical viscosity measurement is usually limited to either low viscosities (i.e. viscometer or rheometer) or

high viscosities (i.e. DMA), whereas a dielectric measurement can continue over both regimes. Second, with the small, thin electrode sensors that are available today, dielectric analysis can monitor cure nondestructively, in situ, whereas mechanical techniques require a separate sample with a specific geometry.

Since dielectric monitoring of cure involves measurement of the ion and dipole mobility, the parameter of most interest is the resistance or resistivity. Figure 15 shows both the resistance and viscosity of an epoxy composite as it goes through a thermal cure cycle.[58] It shows that good qualitative agreement exists between the mechanical viscosity and dielectric resistance during the heating and curing of the composite. Physically, as the temperature is increased, viscosity and resistance both decrease because the ion mobility is higher. As the temperature approaches 175 °C, the polymerization reaction begins causing both the viscosity and resistance curves to change direction. This is because as the resin begins to solidify into a polymer, ion mobility decreases again. Although the data are shown as resistance, they could just as easily be plotted as resistivity, since they differ only in geometrical constants dependent on the experimental set-up.

While resistance and resistivity are scalar quantities that are defined only for zero frequency, an "apparent" resistance or resistivity can be computed at arbitrary frequency. More specifically, apparent resistivity, ρ, is the inverse of the apparent conductivity defined in Equation (17) (Equation 23):

$$\rho = \frac{1}{\sigma} = \frac{1}{w\varepsilon_0\varepsilon_r''} \tag{23}$$

An example of temperature-dependent resistivity for a sample of neat epoxy is plotted in Figure 16(a).[49] This sample underwent a linear temperature ramp of 5 °C min^{-1}. Similarly to Figure 15, the resistivity decreases with increases in temperature until the resin begins to polymerize into a solid. Then the resistivity increases again as ion mobility decreases. For completeness, Figure 16(b) shows the corresponding real and imaginary permittivities. As the temperature rises before cure, the loss factor increases owing to increasing ion mobility. As the cure proceeds, both the real and imaginary permittivity drop dramatically, indicating the formation of a polymer network that inhibits ion and dipole motion.

Because of the dramatic changes that can occur to the dielectric properties during cure, dielectric spectroscopy seems a natural technique for determining chemical kinetic rate constants. For example, kinetic information can be obtained by assuming a correlation between extent of cure and various dielectric data, such as logarithm of resistivity or relaxed permittivity. Vitrification can be determined by following the glass transition (e.g. with frequency domain, relaxation data). However, these concepts should be used with caution, since empirical correlations can be an oversimplification. The cure process can be complicated with multiple steps occurring simultaneously and a direct relationship between a

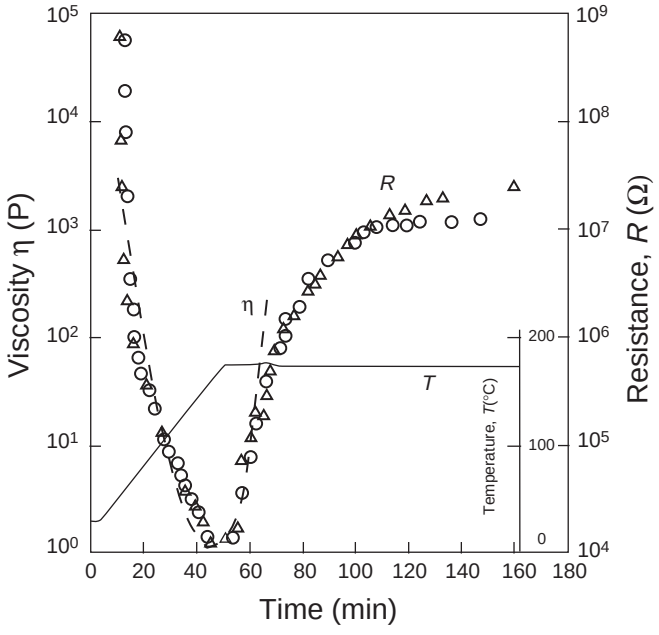

Figure 15 Resistance, R, and viscosity, η, of an epoxy composite during cure. Temperature was ramped from 24 to 175 °C at a nominal heating rate of 3 °C min^{-1}. [Reproduced by permission of the Society of Plastics Engineers from Y.A. Tajima, *Polym. Composites*, **3**(3), 162–169 (1982).]

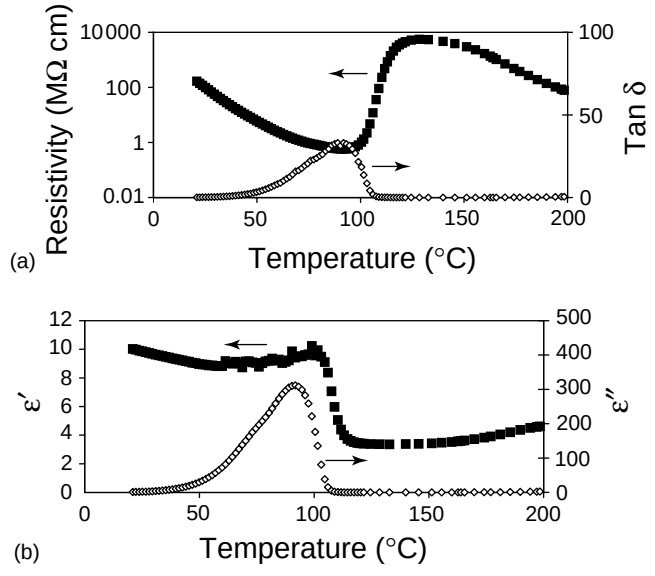

Figure 16 (a) Apparent resistivity and loss tangent of an epoxy thermoset during thermal cure. (b) Real and imaginary permittivity of the same sample.

dielectric parameter and cure kinetics does not always exist. An excellent review of these issues is given by Senturia and Sheppard.[29]

4.3 Polymer Emulsions and Film Formation

Another important area of synthetic polymers where dielectric spectroscopy is useful is in the characterization of polymer latex emulsions. Because of its ease of application and environmental friendliness, the polymer latex has become a popular alternative to traditional solvent-based coatings. Characterization of these systems and their performance is an important problem for a large number of industrial applications. Dielectric spectroscopy is a useful technique for measuring certain aspects of these materials.

One such area were dielectric analysis has been proven effective is in the determination of latex particle size.[59] Figure 17(a) and (b) show the real and imaginary dielectric constant as a function of frequency for three different latex standards with diameters of 109, 497, and 1090 nm. It should be pointed out that these data were obtained with a four-electrode cell. This is because polymer latices contain highly mobile ions (high ionic conductivity), so the four-electrode cell helps to minimize blocking effects. In particular, separation of the ''driving'' electrodes from the ''sensing'' electrodes reduces the effect of ion accumulation on the measured signal. The data show that there is dielectric relaxation associated with these systems and that size is an important variable affecting the frequency where the relaxation is centered. Of course, there are other variables that can also affect both the center frequency and relative amplitude of the relaxation, including size distribution, solids content, temperature, and formulation. However, the data in Figure 17(a) and (b) demonstrate that under controlled conditions, dielectric spectroscopy can be used to measure rapidly particle sizes in polymer latices.

Dielectric analysis is also useful in the study of film formation or drying of polymer latices.[60,61] In recent work, Schultz and Chartoff[62] studied the drying of polymer latex films under various humidities. Figure 18 shows the conductance as a function of drying time for a typical sample film cast on interdigitated comb electrodes. Conductance, as opposed to conductivity defined in Equation (17), does not take the electrode geometry into account. Instead, it is calculated directly from the measured impedances. It is the inverse of the resistance defined in Equation (4). Hence it does not represent quantitative, geometry-independent data. The conductance data in Figure 18 could have been calculated at a fixed frequency. However, increased range and accuracy were obtained by fitting complex impedance plots of frequency scans and extrapolating the data to zero frequency. The data in Figure 18 show that there are several different stages of drying. These different drying stages are associated with different types of drying: fast evaporation, percolation, and diffusion. Hence the experimental data can be related to the underlying physical phenomena responsible for film formation. What is interesting about these data is that the quantitative values are unimportant. Rather, the relative trends and changes are often all that are required to enable a researcher to understand the material better.

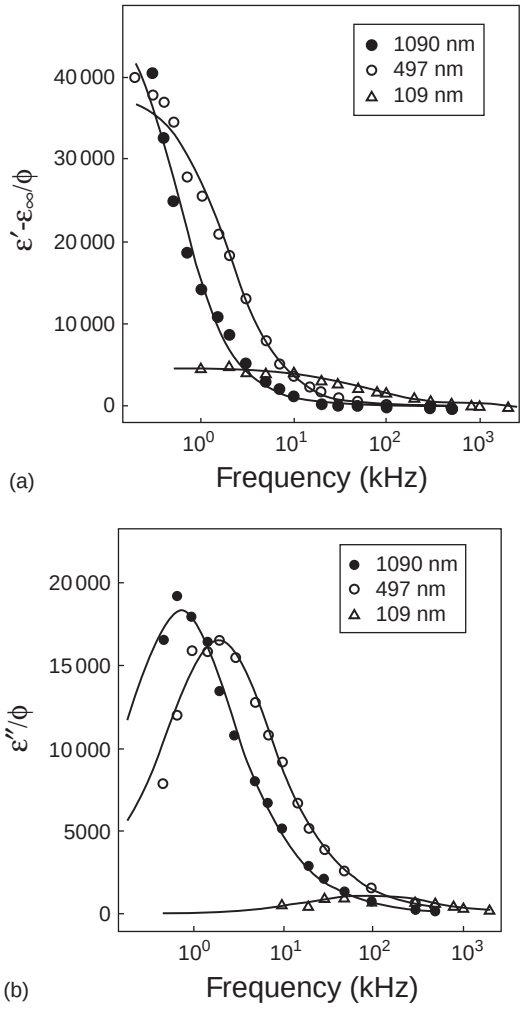

Figure 17 (a) Real and (b) imaginary parts of the dielectric constant normalized by the weight fraction as a function of frequency for latex standards with diameters as indicated. [Reproduced from B.B. Sauer, R.S. Stock, K.-H. Lim, W.H. Ray, *J. Appl. Polym. Sci.*, **39**, 2419–2441 (1990), Copyright 1990 John Wiley & Sons Inc. Reprinted by permission of John Wiley & Sons, Inc.]

4.4 Humidity/Environmental Effects

Polymeric materials often undergo chemical and physical changes as a result of their environment. For example,

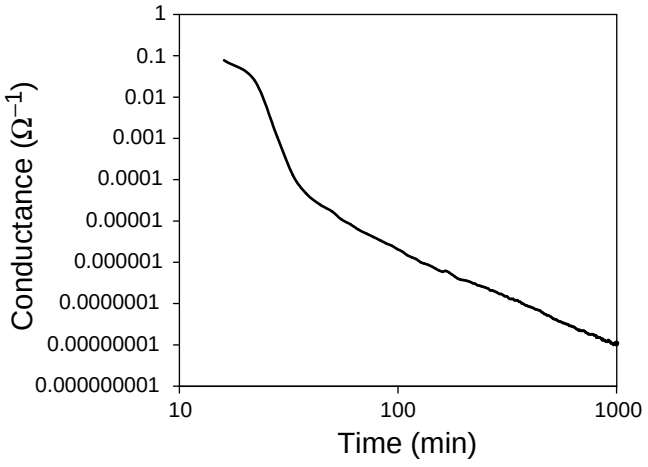

Figure 18 Conductance of a polymer latex film during drying.

radiation may cause cross-linking or chain scission in polymers, humidity may cause swelling because of water sorption, or polymers may age as a result of additive migration. Dielectric spectroscopy can provide a useful way to monitor these changes. In these types of studies, polymer samples may be measured at different times after exposure to different conditions, or a single polymer sample may be monitored continuously as it is undergoing change.

An example of environmentally induced change is shown in polymers studied by Maffezzoli et al.[63] and illustrated in Figure 19. In this case an epoxy matrix composite was immersed in water and an interdigitated comb electrode embedded between plies monitored the

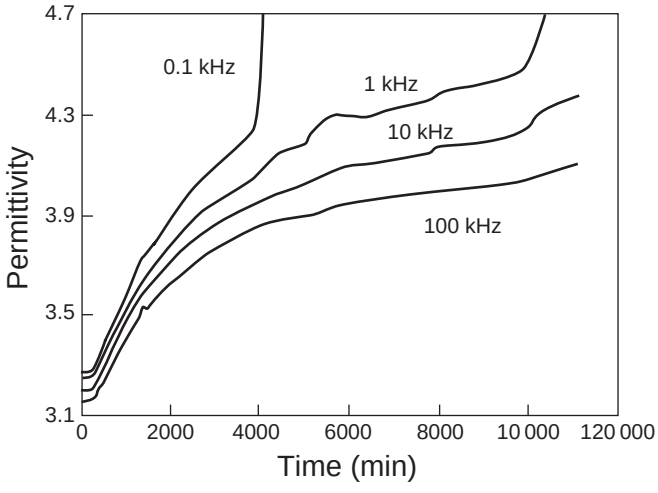

Figure 19 Permittivity as a function of time during water absorption by an epoxy matrix composite at 20 °C. [Reproduced by permission of the Society of Plastics Engineers from A.M. Maffezzoli, L. Peterson, J.C. Seferis, J. Kenny, L. Nicolais, *Polym. Eng. Sci.*, **33**(2), 75–82 (1993).]

permittivity. The data show that the permittivity is initially constant because of the finite amount of time it takes for water to diffuse to the sensor position. After about 400 min, the permittivity begins to increase, and this is attributed to the increasing concentration of water molecules at the electrode. Eventually, the permittivity levels off as the concentration of water reaches saturation.

Interestingly, the curves also indicate that the permittivity experiences a sudden and dramatic increase after a sufficiently long time. The effect is most severe at lower frequencies. While this effect may seem puzzling at first, it is easily explained by electrode-blocking effects. As water is absorbed by the epoxy, the system swells, and ionic impurities that may be present in the material become more mobile. Thus, at low enough frequencies the electrodes may experience ionic polarization that leads to the anomalous upturn in the permittivity curves. While sudden and large changes in the permittivity may be an indicator of electrode-blocking effects, it is best to plot the complex impedance data as described in the instrumentation and sample preparation section of this article, in order to confirm that electrode blocking is occurring.

Other examples of environmental aging measurements include measurement of deterioration in dielectric insulation.[64] Specifically, dielectric loss factor measurements permit the diagnosis of aging from high-voltage stresses. As a final example, Kranbuehl et al.[65] studied the use of an interdigitated electrode as an in situ monitor of the state of polymer coatings exposed to acids. The advantage to this type of in situ dielectric sensing is that measurements can be made without interrupting the experiment or process.

4.5 Analysis of Composites

A final application of dielectric spectroscopy that has received considerable attention is in the evaluation of composites. Polymer matrix composites can be filled with continuous fibers, chopped fibers, platelets, and even simple fillers such as carbon black or sand. The synergistic properties that result from these mixtures result in a wide array of polymer-based materials with advantageous properties. Understanding the composite dielectric behavior is important for both scientific and practical reasons.

Three important issues that must be considered in composites are their homogeneity, microstructure, and anisotropy. Homogeneity concerns the length scales of the constituent materials. When measuring composites with a dielectric apparatus, the sample size (e.g. separation between electrodes and area of electrodes) must be much larger than the size of the largest constituent dimension within the composite, otherwise an ''average'' response

is not being measured. When measuring moderately conductive materials such as carbon black-filled polymers, microstructure becomes an important issue. In particular, the actual arrangement of the carbon black distributed within the polymer matrix may form a network that can create conduction paths and these paths may be anisotropic. The distribution of constituents within a matrix is highly dependent on the processing conditions and thermal history. To address microstructure issues properly, it is often necessary to include some form of microscopy along with the dielectric measurement of composites. This allows an investigator to tie the dielectric response directly to these important physical factors.

The third issue, anisotropy, is concerned with the orientation of the constituents in a composite. For example, a continuous fiber composite will have dielectric properties that are different depending on the orientation of the sample within the electrodes. If there are particles within a polymer matrix with an anisotropic shape, then they may be oriented because of processing and this too will cause an anisotropic response depending on the alignment. It should also be noted that anisotropy can occur in neat polymers in addition to composites. Drawing or pressing processes can result in anisotropic dielectric properties in amorphous and semicrystalline polymers. However, more dramatic anisotropy can be found in liquid crystalline polymers,[66] which are sometimes termed 'in-situ' composites. The anisotropy in the dielectric permittivity can even be used to estimate quantitatively the degree of alignment in these types of systems.[67] When there is dielectric anisotropy in neat polymers or composites, the orientation of the sample with respect to the measuring electric field must be considered. With parallel-plate electrodes, the electric field orientation is perpendicular to the plane of the electrodes. Several samples made with different orientations may be necessary to characterize fully the permittivity in these materials.

In order to design new composites and understand current ones, it is useful to construct mathematical models that describe these materials. There are a number of theoretical equations for relating constituent properties to the average dielectric response of the composite. Because of the complexity of issues described above, most of these theories assume a homogeneous and isotropic material. Two of the best known theories were originally developed by Bruggeman and Maxwell–Garnett. These theories and all their relatives are collectively known as effective medium theories.

The Maxwell–Garnett theory[68] assumes spherical inclusions, phase a, are completely surrounded by a matrix, phase b. The average permittivity (ε) is then related to the permittivities of phase a (ε_a) and phase b (ε_b) by Equation (24):

$$\frac{\varepsilon - \varepsilon_b}{\varepsilon + 2\varepsilon_b} = v_a \frac{\varepsilon_a - \varepsilon_b}{\varepsilon_a + 2\varepsilon_b} \tag{24}$$

where v_a is the volume fraction of phase a. In contrast to this matrix-inclusion formulation, the Bruggeman effective medium theory[69] assumes that the phases are distributed symmetrically, resulting in Equation (25):

$$0 = v_a \frac{\varepsilon_a - \varepsilon}{\varepsilon_a + 2\varepsilon} + v_b \frac{\varepsilon_b - \varepsilon}{\varepsilon_b + 2\varepsilon} \tag{25}$$

A good physical understanding of effective medium theories can be obtained from a review by Aspnes.[70] In addition, Equations (24) and (25) are only a small sample of the various equations for modeling composite dielectric response. There are numerous other theories that account for particle shape, orientation, microstructure, and many other factors.[71]

ACKNOWLEDGMENTS

I would like to express thanks to Mary Galaska of the University of Dayton for her help in acquiring some of the example data. I also wish to express gratitude to Dr Richard Chartoff of the University of Dayton for his various suggestions.

LIST OF SYMBOLS

A	electrode area
a	inner radius of cylindrical electrodes
b	outer radius of cylindrical electrodes
C	capacitance
d	distance between electrodes
E_a	apparent activation energy
f	frequency (in hertz)
i	denotes imaginary number ($i = \sqrt{-1}$)
I	current
I_0	current amplitude
l	length of cylindrical electrodes
M	dielectric modulus ($M = M' + iM''$)
M'	real part of dielectric modulus
M''	imaginary part of dielectric modulus
R	resistance
t	time
T_g	glass transition temperature
V	voltage
V_0	voltage amplitude
v_i	volume fraction
w_i	weight fraction
Z	complex impedance (ratio of voltage to current)

Z'	real part of the impedance
Z''	imaginary part of the impedance
α	empirical parameter used in Havriliak–Negami and Cole–Cole equations for dielectric relaxation
β	empirical parameter used in Havriliak–Negami equation for dielectric relaxation
δ	phase shift
ε	permittivity (also used as shorthand notation for relative permittivity, ε_r)
ε_0	permittivity of vacuum (8.854×10^{-12} F m^{-1})
ε_r	relative permittivity
ε_R	''relaxed'' permittivity (permittivity measured at low frequencies – used in relaxation models)
ε_U	''unrelaxed'' permittivity (permittivity measured at high frequencies – used in relaxation models)
ε'	real part of complex permittivity
ε''	imaginary part of complex permittivity (dielectric loss factor)
η	viscosity
σ	conductivity
w	angular frequency (in rad s^{-1})

ABBREVIATIONS AND ACRONYMS

DMA	Dynamic Mechanical Analysis
EVA	Poly(ethylene-co-vinyl acetate)
PET	Poly(ethylene terephthalate)
PVC	Poly(vinyl chloride)
TSC	Thermally Stimulated Current

RELATED ARTICLES

Polymers and Rubbers **(Volume 9)**
Dynamic Mechanical Analysis of Polymers and Rubbers

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction

General Articles **(Volume 15)**
Microwave Techniques

REFERENCES

1. P. Drude, *The Theory of Optics*, republished by Dover, New York, 1959.
2. P.W. Debye, *Polar Molecules*, Chemical Catalog Co., New York, 1927; republished by Dover, New York, 1945.
3. C.P. Smyth, *Dielectric Behavior and Structure*, McGraw Hill, New York, 1955.
4. J.W. Smith, *Electric Dipole Moments*, Butterworths, London, 1955.
5. H. Fröhlich, *Theory of Dielectrics*, Oxford University Press, New York, 1958.
6. C.J.F. Böttcher, O.C. Van Belle, P. Bordewick, A. Rip, *Theory of Electric Polarization, Vol. 1, Dielectrics in Static Fields*, Elsevier, New York, 1973.
7. C.J.F. Böttcher, P. Bordewick, *Theory of Electric Polarization, Vol. 2, Dielectrics in Time-dependent Fields*, Elsevier, New York, 1978.
8. A. Von Hippel, *Dielectrics and Waves*, John Wiley & Sons, New York, 1959.
9. A. Von Hippel, *Dielectric Materials and Applications*, John Wiley & Sons, New York, 1954.
10. N.G. McCrum, B.E. Read, G. Williams, *Anelastic and Dielectric Effects in Polymeric Solids*, John Wiley & Sons, London, 1967; republished by Dover, New York, 1991.
11. P. Hedvig, *Dielectric Properties of Polymers*, Adam Hilger, Bristol, 1977.
12. E. Riande, E. Saiz, *Dipole Moments and Birefringence of Polymers*, Prentice Hall, Englewood Cliffs, NJ, 1992.
13. J.P. Runt, J.J. Fitzgerald, *Dielectric Spectroscopy of Polymeric Materials*, American Chemical Society, Washington, DC, 1997.
14. S. Havriliak, Jr, S.J. Havriliak, *Dielectric and Mechanical Relaxation in Materials, Analysis, Interpretation, and Application to Polymers*, Hanser Publishers, New York, 1997.
15. F. Kremer, M. Arndt, 'Broadband Dielectric Measurement Techniques', in *Dielectric Spectroscopy of Polymeric Materials*, eds. J.P. Runt, J.J. Fitzgerald, American Chemical Society, Washington, DC, Chapter 2, 67–79, 1997.
16. J.M. Pochan, J.J. Fitzgerald, G. Williams, 'Experimental Methods for Chemists. A Simplified Approach. Dielectric Properties of Polymers and Other Materials', in *Determination of Electronic and Optical Properties*, 2nd edition, eds. B.W. Rossitter, R.C. Baetzold, Physical Methods of Chemistry Series, Vol. VIII, John Wiley & Sons, New York, Chapter 6, 391–436, 1993.
17. J. Baker-Jarvis, M.D. Janezic, J.H. Grosvenor, Jr, R.G. Geyer, 'Transmission/Reflection and Short-Circuit Line Methods for Measuring Permittivity and Permeability', *NIST Tech. Note*, 1355, 1992.
18. R.M. Hill, C. Pickup, 'Barrier Effects in Dispersive Media', *J. Mater. Sci.*, **20**, 4431–4444 (1985).
19. 'Standard Test Methods for AC Loss Characteristics and Permittivity (Dielectric Constant of Solid Electrical Insulation)', ASTM D 150, American Society for Testing and Materials, Philadelphia, 1995.
20. 'Standard Test Method for Permittivity (Dielectric Constant) and Dissipation Factor of Solid Dielectrics at Frequencies to 10 MHz and Temperatures to 500 °C', ASTM D 2149, American Society for Testing and Materials, Philadelphia, 1996.
21. 'Standard Test Methods for Relative Permittivity and Dissipation Factor of Expanded Cellular Polymers Used

for Electrical Insulation', ASTM D 1673, American Society for Testing and Materials, Philadelphia, 1994.
22. R.J. Armstrong, 'Method of Laminating Employing Measuring the Electrical Impedance of a Thermosetting Resin', *US Pat.*, 3 600 247, 1971.
23. S.D. Senturia, S.L. Garverick, 'Method and Apparatus for Microdielectrometry', *US Pat.*, 4 423 371, 1983.
24. N.F. Sheppard, Jr, S.L. Garverick, D.R. Day, S.D. Senturia, 'Microdielectrometry: A New Method for In Situ Cure Monitoring', in *Proceedings 26th SAMPE Symposium*, 65, 1981.
25. D.E. Kranbuehl, 'Method of Using a Dielectric Probe to Monitor the Characteristics of a Medium', *US Pat.*, 4 710 550, 1987.
26. H.L. Lee, 'Optimization of a Resin Cure Sensor', Master's Thesis, Massachusetts Institute of Technology, 1982.
27. D.E. Kranbuehl, 'Cure Monitoring', in *International Encyclopedia of Composites*, ed. S. Lee, VCH, New York, 531–543, 1990.
28. J.J. Fontanella, J.J. Wilson, M.K. Smith, M.C. Wintersgill, C.S. Coughlin, P. Mazaud, S.G. Greenbaum, R.L. Siddon, 'Electrical Relaxation in Poly(propylene oxide) With and Without Alkali Metal Salts', *Solid State Ionics*, **50**, 259–271 (1992).
29. S.D. Senturia, N.F. Sheppard, Jr, 'Dielectric Analysis of Thermoset Cure', *Adv. Polym. Sci.*, **80**, 1–47 (1986).
30. A.K. Jonscher, 'Analysis of the Alternating Current Properties of Ionic Conductors', *J. Mater. Sci.*, **13**, 553–562 (1978).
31. J.W. Schultz, 'Anomalous Dispersion in the Dielectric Spectra of Conductive Materials', *IEEE Trans. Instrum. Meas.*, **47**, 766–768 (1999).
32. C. Lavergne, C. Lacabanne, 'A Review of Thermo-stimulated Current', *IEEE Electr. Insul. Mag.*, **9**, 5–21 (1993).
33. J.P. Ibar, P. Denning, T. Thomas, A. Bernes, C. de Goys, J.R. Saffell, P. Jones, C. Lacabanne, 'Characterization of Polymers by Thermally Stimulated Current Analysis and Relaxation Map Analysis Spectroscopy', *Adv. Chem. Seq.*, **227**, 167–190 (1990).
34. H.W. Starkweather, P. Avakian, R.R. Matheson, J.J. Fontanella, M.C. Wintersgill, 'Low-Temperature Dielectric Behavior of Polymers and Copolymers of Tetrafluoroethylene', *Macromolecules*, **24**, 3853–3856 (1991).
35. S. Matsuoka, *Relaxation Phenomena in Polymers*, Hanser, New York, 1992.
36. J.D. Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, John Wiley & Sons, New York, 1980.
37. H. Vogel, 'The Law of Viscosity Change with Temperature', *Phys. Z.*, **22**, 645 (1921).
38. G.S. Fulcher, 'Analysis of Measurements of the Viscosity of Glasses', *J. Am. Ceram. Soc.*, **8**, 339–355 (1925).
39. S. Havriliak, Jr, S.J. Havriliak, 'Electric Polarization of Polar Time-dependent Rigid Materials', *Adv. Chem. Phys.*, **98**, 217–282 (1997).
40. K.S. Cole, R.H. Cole, 'Dispersion and Absorption in Dielectrics', *J. Chem. Phys.*, **9**, 341–351 (1941).
41. R.M. Fuoss, J.G Kirkwood, 'Electrical Properties of Solids. VIII. Dipole Moments in Polyvinyl Chloride–Diphenyl Systems', *J. Am. Chem. Soc.*, **63**, 385–401 (1941).
42. D.W. Davidson, R.H. Cole, 'Dielectric Relaxation in Glycerine', *J. Chem. Phys.*, **18**, 1417 (1950).
43. J. Jackle, 'Theory of Glass Transitions. New Thoughts About Old Facts', *Philos. Mag. B*, **56**, 113–127 (1987).
44. W. Williams, D.G. Watts, 'Non-symmetrical Dielectric Relaxation Behavior Arising from a Simple Empirical Decay Function', *Trans. Faraday Soc.*, **66**, 80–85 (1970).
45. A.K. Jonscher, *Dielectric Relaxation in Solds*, Chelsea Dielectric, London, 1983.
46. S. Havriliak, Jr, S.J. Negami, 'A Complex Plane Analysis of α-Dispersions in Some Polymer Systems', *J. Polym. Sci., Part C*, **14**, 99–117 (1966).
47. S. Havriliak, Jr, S.J. Negami, 'A Complex Plane Representation of Dielectric and Mechanical Relaxation Processes in Some Polymers', *Polymer*, **8**(4), 161–210 (1967).
48. S. Havriliak, Jr, S.J. Havriliak, 'Unbiased Modeling of Dielectric Dispersions', in *Dielectric Spectroscopy of Polymeric Materials*, eds. J.P. Runt, J.J. Fitzgerald, American Chemical Society, Washington, DC, Chapter 6, 175–200, 1997.
49. J.W. Schultz, unpublished data.
50. G.S. Rellick, J. Runt, 'A Dielectric Study of Poly(ethylene-co-vinyl acetate)–Poly(vinyl chloride) Blends. I. Miscibility and Phase Behavior', *J. Polym. Sci. Polym. Phys. Ed.*, **24**, 279–302 (1986).
51. T.G. Fox, 'Influence of Diluent and of Copolymer Composition on the Glass Temperature of a Polymer System', *Bull. Am. Phys. Soc.*, **1**, 123 (1956).
52. N.T. Smith, D.D. Shepard, 'Dielectric Cure Analysis: Theory and Industrial Applications', *Sensors*, **12**(10), 42–48 (1995).
53. D.E. Kranbuehl, P. Kingsley, S. Hart, G. Hasko, B. Dexter, A.C. Loos, 'In Situ Sensor Monitoring and Intelligent Control of the Resin Transfer Molding Process', *Polym. Composites*, **15**, 299–305 (1994).
54. T.J. Crowley, K.Y. Choi, 'In-line Dielectric Monitoring of Monomer Conversion in a Batch Polymerization Reactor', *J. Appl. Polym. Sci.*, **55**, 1361–1365 (1995).
55. D. Kranbuehl, J. Rogozinski, A. Meyer, 'Monitoring the Changing State of a Polymeric Coating Resin During Synthesis, Cure, and Use', in *Proceedings of the XXIVth International Conference in Organic Coatings, Athens, Greece*, 197–211, 1998.
56. D. Schuele, R. Renner, D. Coleman, 'Monitoring of Photopolymerization Through Dielectric Spectroscopy', *Mol. Cryst. Liq. Cryst.*, **299**, 343–352 (1997).
57. J.F. Maguire, M.A. Miller, S. Venketesan, 'Processing of Soft Matter and Composites: Integration of Material Sensors with Process Models and Intelligent Control Algorithms', *Eng. Appl. Artif. Intell.*, **11**, 605–618 (1998).

58. Y.A. Tajima, 'Monitoring Cure Viscosity of Epoxy Composite', *Polym. Composites*, **3**(3), 162–169 (1982).
59. B.B. Sauer, R.S. Stock, K.-H. Lim, W.H. Ray, 'Polymer Latex Particle Size Measurement through High Speed Dielectric Spectroscopy', *J. Appl. Polym. Sci.*, **39**, 2419–2441 (1990).
60. F. Cansell, F. Henry, C. Pichot, 'Study of Polymer Latex Coalescence by Dielectric Measurements in the Microwave Domain: Influence of Latex Characteristics', *J. Appl. Polym. Sci.*, **41**, 547–563 (1990).
61. D. Kranbuehl, D. Hood, L. McCullough, C. Kellam, J. Yang, 'In Situ Sensing for Monitoring Molecular and Physical Property Changes During Film Formation', *Am. Chem. Soc. Symp. Ser.*, **648**, 96–117 (1996).
62. J.W. Schultz, R.P. Chartoff, 'Dielectric and Thermal Analysis of the Film Formation of a Polymer Latex', *J. Coat. Technol.*, **68**, 97–106 (1996).
63. A.M. Maffezzoli, L. Peterson, J.C. Seferis, J. Kenny, L. Nicolais, 'Dielectric Characterization of Water Sorption in Epoxy Resin Matrices', *Polym. Eng. Sci.*, **33**(2), 75–82 (1993).
64. T. Kwaaitaal, W.M.M.M. van den Eijnden, 'Dielectric Loss Measurement as a Tool to Determine Electrical Aging of Extruded Polymeric Insulated Power Cables', *IEEE Trans. Electr. Insul.*, **EI-22**, 101–105 (1987).
65. D. Kranbuehl, H. Aandahi, N. Haralampus, W. Newby, D. Hood, G. Boiteux, G. Seytre, P. Pascault, A. Maazouz, J.F. Gerard, H. Sautereau, J.F. Chailan, A.C. Loos, J.D. MacRae, 'Use of In-situ Dielectric Sensing for Intelligent Processing and Health Monitoring', *Proc. SPIE*, **2779**, 112–117 (1996).
66. J.K. Moscicki, 'Dielectric Relaxation in Macromolecular Liquid Crystals', in *Liquid Crystal Polymers: from Structures to Applications*, ed. A.A. Collyer Elsevier Applied Science, London, Chapter 4, 1992.
67. J.W. Schultz, R.P. Chartoff, 'Photopolymerization of Nematic Liquid Crystal Monomers for Structural Applications: Molecular Order and Orientation Dynamics', *Polymer*, **39**, 319–325 (1998).
68. J.C.M. Garnett, 'Colours in Metal Glasses and in Metal Films', *Philos. Trans. R. Soc. London, Ser. A*, **203**, 385–420 (1904).
69. D.A.G. Bruggeman, 'Calculation of Various Physical Constants of Heterogeneous Substances. Part I. Dielectric Constant and Conductivity of mixtures of Isotropic Substances', *Ann. Phys.*, **24**, 636–679 (1935).
70. D.E. Aspnes, 'Local-field Effects and Effective-medium Theory: A Microscopic Perspective', *Am. J. Phys.*, **50**, 704–708 (1982).
71. P.S. Neelakanta, *Handbook of Electromagnetic Materials: Monolithic and Composite Versions and Their Applications*, CRC Press, Boca Raton, FL, 1995.

Dynamic Mechanical Analysis of Polymers and Rubbers

Kevin P. Menard
PerkinElmer Life and Analytical Sciences, Shelton, CT, USA

Bryan W. Bilyeu
Xavier University, New Orleans, LA, USA

1 Introduction and History of Dynamic Mechanical Analysis 1
2 Theory and Operating Principles 2
2.1 Forced Frequency Analyzers 2
2.2 Free Resonance Analyzers 3
3 Instrumentation 4
4 Applications 5
4.1 Thermoplastic Solids and Cured Thermosets 5
4.2 Polymer Melts and Solutions 12
4.3 Thermoset Curing 16
4.4 Testing Environmental Effects 19
Acknowledgments 21
List of Symbols 21
Abbreviations and Acronyms 21
Related Articles 21
References 21
Further Reading 24

Dynamic mechanical analysis (DMA) is the technique of applying a stress or strain to a sample and analyzing the response to obtain phase angle and deformation data. These data allow the calculation of the damping or tan δ; the storage, loss, and complex modulus; as well as complex viscosity and complex compliance. Two approaches are used: (i) forced frequency, where the signal is applied at a set frequency and (ii) free resonance, where the material is perturbed and allowed to exhibit free resonance decay. Most modern DMA instruments are of the forced resonance type. In both approaches, the technique is very sensitive to the motions of the polymer chains and it is a powerful tool for measuring transitions in polymers. It is estimated to be 100 times more sensitive to the glass transition than differential scanning calorimetry (DSC) and resolves other more localized transitions not detected in DSC. In addition, the technique allows the rapid scanning of a material's modulus and viscosity as a function of temperature or frequency. DMA may also be referred to as dynamic mechanical thermal analysis (DMTA), dynamic mechanical spectroscopy (DMS), or dynamic thermomechanical analysis (DTMA).

1 INTRODUCTION AND HISTORY OF DYNAMIC MECHANICAL ANALYSIS

DMA is a very important tool in the modern polymer laboratory. Despite that, only a few books have concentrated on the technique.[1–3] The first attempts to carry out oscillatory experiments to measure the elasticity of a material were made by Poynting[4] in 1909. Other early works gave methods to apply oscillatory deformations by various means to study metals,[5] and many early experimental techniques were reviewed by te Nijenhuis[6] in 1978. Miller's book on polymer properties[7] referred to dynamic measurements in this early discussion of molecular structure and stiffness. Early commercial instruments included the Weissenberg rheogoniometer (~1950) and the Rheovibron (~1958). The Weissenberg rheogoniometer, which dominated cone-and-plate measurements for over 20 years following 1955, was the commercial version of the first instrument to measure normal forces.[8] By the time Ferry wrote the *Viscoelastic Properties of Polymers*[9] in 1961, dynamic measurements were an integral part of polymer science and he gives the best development of the theory available. In 1967, McCrum et al.[10] collected the current information on DMA and dielectric analysis (DEA) in their landmark textbook. The technique remained fairly specialized until the late 1960s when commercial instruments became more user friendly. Around 1966, Gillham developed the torsional braid analyzer (TBA),[11] which started the modern period of DMA. In 1971, Macosko and Starita[12] built a DMA instrument that measured normal forces, and from this came the Rheometrics Corporation. In 1976, Bohlin also developed a commercial DMA and started Bohlin Rheologia. Both instruments used torsional geometry. The early instruments were, regardless of manufacturer, difficult to use, slow, and limited in their ability to process data. In the late 1970s, Murayama[13] and Read and Brown[14] wrote books on the uses of DMA for material characterization. Several thermal and rheological instrument companies introduced DMA instruments in the same time period and currently most thermal and rheological vendors offer some type of DMA device.

Polymer Laboratories offered a DMA instrument using an axial geometry in the early 1980s. This was soon

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

followed by an instrument from Du Pont. PerkinElmer developed a controlled stress analyzer based on their thermomechanical analyzer (TMA) technology, which was designed for increased low-end sensitivity. In 1990 and early 2000, the latest wave of DMA devices came out, adding the capability to handle samples in solution, immersion, and exposure to UV light.[15] The competition between vendors has led to easier-to-use, faster, and less expensive instruments. The revolution in computer technology, which has so much affected the laboratory, changed the latter, and DMA instruments of all types became more user-friendly as computers and software evolved.

2 THEORY AND OPERATING PRINCIPLES

2.1 Forced Frequency Analyzers

If a constant load applied to a sample begins to oscillate sinusoidally (Figure 1), the sample will deform sinusoidally. This will be reproducible if the material is deformed within its linear viscoelastic region. For any one point on the curve, the stress applied is described by Equation (1):

$$\sigma = \sigma_0 \sin \omega t \tag{1}$$

where σ is the stress at time t, σ_0 is the maximum stress, ω is the frequency of oscillation, and t is time. The resulting strain wave shape will depend on how much viscous behavior and how much elastic behavior the sample has. In addition, the rate of stress can be determined by taking the derivative of the above-mentioned equation

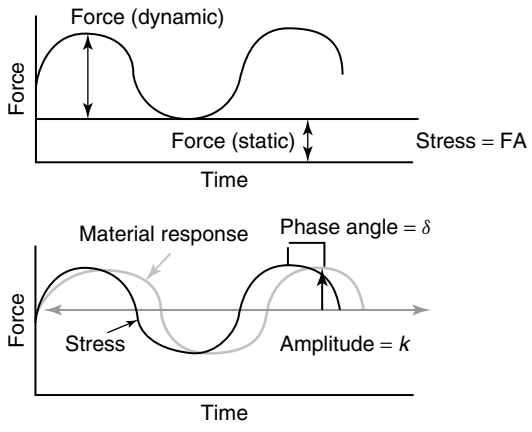

Figure 1 When a sample is subjected to a sinusoidal oscillating stress, it responds with a similar strain wave provided the material stays within its elastic limits. (Reproduced with permission of Taylor & Francis Group LLC.)

with respect to time:

$$\frac{d\sigma}{dt} = \omega\sigma_0 \cos \omega t \tag{2}$$

The two extremes of the material's behavior, elastic and viscous, provide the limits that will sum to give the strain wave. The behavior can be understood by evaluating each of the two extremes. The material at the spring-like or Hookean limit will respond elastically with the oscillating stress. The strain at any time can be written as Equation (3):

$$\varepsilon(t) = E\sigma_0 \sin(\omega t) \tag{3}$$

where $\varepsilon(t)$ is the strain at any time t, E is the modulus, σ_0 is the maximum stress at the peak of the sine wave, and ω is the frequency. Since in the linear region σ and ε are linearly related by E, the relationship is given by Equation (4):

$$\varepsilon(t) = \varepsilon_0 \sin(\omega t) \tag{4}$$

where ε_0 is the strain at the maximum stress. This curve, shown in Figure 2(a), has no phase lag (or no time difference from the stress curve) and is called the *in-phase portion* of the curve.

The viscous limit was expressed as one in which the stress is proportional to the strain rate, which is the first derivative of the strain. This is best modeled by a dashpot, and for that element the term for the viscous response in terms of strain rate is described by Equation (5):

$$\varepsilon(t) = \eta\frac{d\sigma_0}{dt} = \eta\omega\sigma_0 \cos(\omega t) \tag{5}$$

or Equation (6):

$$\varepsilon(t) = \eta\omega\sigma_0 \sin\left(\omega t + \frac{\pi}{2}\right) \tag{6}$$

where the terms are as above and η is the viscosity. Substituting terms as above gives Equation (7):

$$\varepsilon(t) = \omega\sigma_0 \cos(\omega t) = \omega\sigma_0 \sin\left(\omega t + \frac{\pi}{2}\right) \tag{7}$$

This curve is shown in Figure 2(b). Now consider the behavior of the material that lies between these two limits. This curve is shown in Figure 2(c), which is intermediate between the above-mentioned cases. The difference between the applied stress and the resultant strain is an angle δ, and this must be added to the equations. Hence, the elastic response at any time can now be written as Equation (8):

$$\varepsilon(t) = \varepsilon_0 \sin(\omega t + \delta) \tag{8}$$

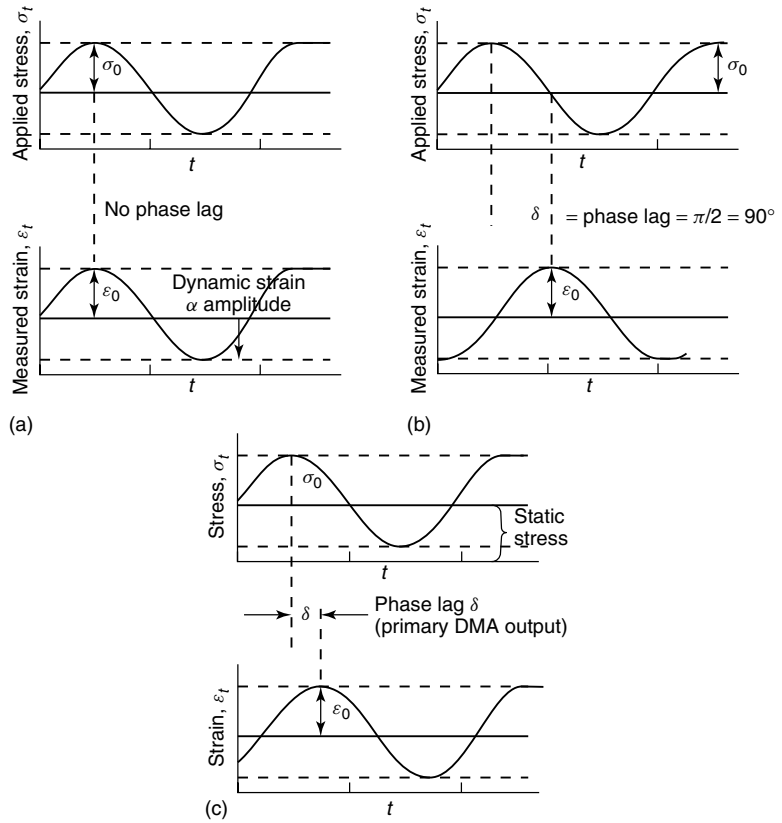

Figure 2 When the material responds to the applied wave perfectly elastically, an in-phase response is seen (a), while a viscous response gives an out-of-phase response (b). Viscoelastic materials fall in between these two lines as shown in (c). (With permissions from PerkinElmer Life and Analytical Sciences, Shelton, CT.)

Using trigonometry, this can be rewritten as Equation (9):

$$\varepsilon(t) = \varepsilon_0[\sin(\omega t)\ \cos\ \delta + \cos(\omega t)\ \sin\ \delta] \tag{9}$$

This equation, corresponding to the curve in Figure 2(c), can be separated into the in-phase and out-of-phase strains that correspond to curves such as those in Figure 2(a) and (b), respectively. These sum to the curve in Figure 2(c) and are given by Equations (10) and (11):

$$\varepsilon' = \varepsilon_0\ \sin(\delta) \tag{10}$$

$$\varepsilon'' = \varepsilon_0\ \cos(\delta) \tag{11}$$

And the vector sum of these two components gives the overall or complex strain on the sample (Equation 12):

$$\varepsilon^* = \varepsilon' + i\varepsilon'' \tag{12}$$

2.2 Free Resonance Analyzers

If a suspended sample can swing freely, it will oscillate like a harp or guitar string until the oscillations gradually come to a stop. The naturally occurring damping of the material controls the decay of the oscillations. This produces a wave, shown in Figure 3, which is a series of sine waves decreases in amplitude and frequency. Several methods exist to analyze these waves and are covered in a review by Gillham.[16,17] These methods have also been applied successfully to the recovery portion of a creep-recovery curve, in which the sample goes into free resonance on removal of the creep force.[18,19]

From the decay curve, the period T and the logarithmic decrement Λ can be calculated. Several methods exist for

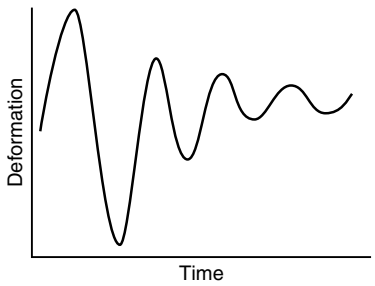

Figure 3 The decay wave from a free resonance analyzer shows the decrease in amplitude of signal with time. (Reproduced with permission of Taylor & Francis Group LLC.)

both manual and digital processing.[16,17,20] Fuller details of the following may be found in McCrum et al.[20] and Gillham.[16,17] Basically, the decay of the amplitude is evaluated over as many swings as possible to reduce error (Equation (13)):

$$\Lambda = \frac{1}{j} \ln \left(\frac{A_n}{A_{n+j}} \right) \tag{13}$$

where j is the number of swings and A_n is the amplitude of the nth swing. For one swing, where $j = 1$, the above equation becomes Equation (14):

$$\Lambda = \ln \left(\frac{A_n}{A_{n+j}} \right) \tag{14}$$

If for a low value of Λ, where $A_n/A_{n+1} \approx 1$, the equation can be rewritten as

$$\Lambda \approx \frac{1}{2} \left(\frac{A_n^2 - A_{n+1}^2}{A_n^2} \right) \tag{15}$$

then from this, since the square of the amplitude is proportional to the stored energy, $\Delta W / W_{\text{st}}$, and the stored energy can be expressed as $2\pi \tan \delta$, this equation becomes Equation (16):

$$\Lambda \approx \frac{1}{2} \left(\frac{\Delta W}{W_{\text{st}}} \right) = \pi \tan \delta \tag{16}$$

which gives us the phase angle δ. The time of the oscillations, the period T, can be found using the Equation (17):

$$T = 2\pi \sqrt{\frac{M}{\Gamma_1}} \sqrt{\frac{1 + \Lambda^2}{4\pi^2}} \tag{17}$$

where Γ_1 is the torque for one cycle and M is the moment of inertia around the central axis. Alternatively, T can be calculated directly from the plotted decay curve as Equation (18):

$$T = \left(\frac{2}{n} \right) (t_n - t_0) \tag{18}$$

where n is the number of cycles and t is the time. From this, the shear modulus, G, can be calculated, which for a rod of length L and radius r is given by Equation (19):

$$G = \left(\frac{4\pi^2 ML}{NT^2} \right) \left(1 + \frac{\Lambda^2}{4\pi^2} \right) - \frac{mgr}{12N} \tag{19}$$

where m is the mass of the sample, g is the gravitational constant, and N is a geometric factor. In the same system, the storage modulus, G', can be calculated using Equation (20):

$$G' = \left(\frac{1}{T^2} \right) \left(\frac{8\pi ML}{r^4} \right) \tag{20}$$

Having the storage modulus and the tangent of the phase angle, the remaining dynamic properties can be calculated.

Free resonance analyzers normally are limited to rod-shaped or rectangular samples or materials that can be impregnated on to a braid. This last approach is how the curing studies on epoxy and other resin systems are done in torsion, which gives these instruments the name TBA.

3 INSTRUMENTATION

One of the biggest choices when selecting a DMA instrument is to decide whether to use stress (force) or strain (displacement) control for applying the deforming load to the sample (Figure 4a and b). Strain-controlled analyzers, whether for simple static testing or for DMA, move the probe to a set distance and use a force balance transducer or load cell to measure the stress. These parts are typically located on different shafts. The simplest version of this is a screw-driven tester, where the sample is pulled one turn. This requires very large motors, so the available force always exceeds what is needed. They normally have better short-time response for low-viscosity materials and can perform

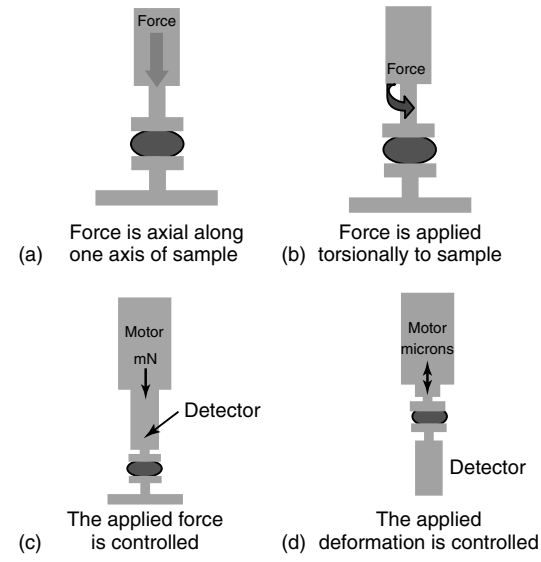

Figure 4 Types of forced resonance DMA instruments: (a) axial, (b) torsional, (c) controlled stress, and (d) controlled strain. (Reproduced with permission of Taylor & Francis Group LLC.)

stress–relaxation experiments easily. They also usually can measure normal forces if they are arranged in torsion. A major disadvantage is that their transducers may drift at long times or with low signals.

Stress-controlled analyzers are cheaper to make because there is only one shaft, but are somewhat trickier to use. Many of the difficulties have been alleviated by software and many strain-controlled analyzers on the market are really stress-controlled instruments with feedback loops, making them act as if they were strain controlled. In stress control, a set force is applied to the sample. As temperature, time, or frequency varies, the applied force remains the same. This may or may not be the same stress: in extension, for example, the stretching and necking of a sample will change the applied stress seen during the run. However, this constant stress is a more natural situation in many cases and it may be more sensitive to material changes. Good low-force control means that they are less likely to destroy any structure in the sample. Long relaxation times or long creep studies are more easily performed on these instruments. Their greatest disadvantage is that their short-time responses are limited by inertia with low-viscosity samples.

Since most DMA experiments are run at very low strains (~0.5% maximum) so as to stay well within a polymer's linear region, it has been reported that both analyzers give the same results.[21] However, when one reaches the nonlinear region, the difference becomes significant as stress and strain are no longer linearly related. Most of the instruments on the market today are stress-controlled analyzers that mimic strain control by a feedback loop. For most users, this is not a problem but if one is interested in the nonlinear regions, it should be considered.

DMA instruments are normally built to apply the stress or strain in two ways (Figure 4c and d). One can apply force in a twisting motion, so that one is testing the sample in torsion. This type of instrument is the dynamic analog of the constant-shear, spinning-disk rheometers. While mainly used for liquids and melts, solid samples may also be tested by twisting a bar of the material. Torsional analyzers normally also permit continuous shear and normal force measurements. Most of these analyzers can also do creep–recovery, stress–relaxation, and stress–strain experiments.

Axial analyzers are normally designed for solid and semisolid materials and apply a linear force to the sample. These analyzers are usually associated with flexure, tensile, and compression testing, but they can be adapted to do shear and liquid specimens by proper choice of fixtures. Sometimes the instrument's design makes this inadvisable, however (e.g. working with a very fluid material in a system where the motor is underneath the sample and has the potential to damage the instrument if the sample spills into the motor). These analyzers can normally test higher modulus materials than torsional analyzers and can run TMA studies in addition to creep–recovery, stress–relaxation, and stress–strain experiments. Recent advances in sample handling techniques have made it possible for these instruments to handle both lower viscosity samples and powders.[1,15,22]

Despite the traditional selection of torsional instruments for melts and liquids and axial instruments for solids, there is really considerable overlap between these types of instruments. With the proper choice of sample geometry and good fixtures, both types can handle similar samples, as shown by the use of both types to study the curing of neat resins. Normally, axial analyzers cannot handle fluid samples below about 500 Pa s and torsional instruments will top out with the harder samples (the exact modulus depending on the size of the motor and/or load cell).

4 APPLICATIONS

4.1 Thermoplastic Solids and Cured Thermosets

The thermal transitions in polymers can be described in terms of either free volume changes[23] or relaxation times.[24] While the latter tend to be preferred by engineers and rheologists in contrast to chemists and polymer physicists who lean toward the former, both descriptions are equivalent. Changes in free volume, v^f, can be monitored as a volumetric change in the polymer, by the absorption or release of heat associated with that change, the loss of stiffness, increased flow, or a change in relaxation time.

The free volume of a polymer, v^f, is known to be related to viscoelasticity,[25,26] aging,[27–30] penetration by solvents,[31] and impact properties.[32] Defined as the space that a molecule has for internal movement, it is shown schematically in Figure 5(a). A simple approach to looking at free volume is the crankshaft mechanism,[33] where the molecule is imagined as a series of jointed segments. From this model, it is possible to describe simply the various transitions seen in a polymer. Other models exist that allow for higher precision in describing the behavior: the best seems to be the Doi and Edwards model.[34] Aklonis and MacKnight[35] give a good summary of the available models, as does Rohn.[36]

The crankshaft model treats the molecule as a collection of mobile segments that have some degree of free movement. This is a very simplistic approach, yet very useful for explaining the behavior. As the free volume of the chain segment increases, its ability to move in various directions also increases (Figure 5b). This increased

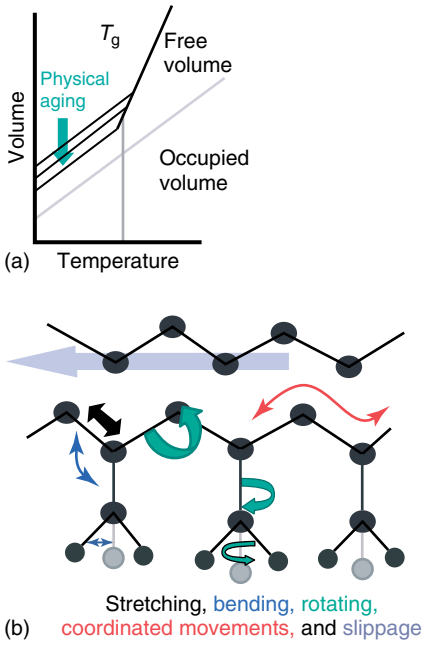

Figure 5 Free volume, v^f, in polymers: (a) the relationship of free volume to transitions and (b) a schematic example of free volume and the crankshaft model. Below the T_g in (a), various paths with different free volumes exist depending on heat history and processing of the polymer, where the path with the least free volume is the most relaxed. (b) The various motions of a polymer chain. Unless enough free volume exists, the motions cannot occur. (Reproduced with permission of Taylor & Francis Group LLC.)

mobility in either side chains or small groups of adjacent backbone atoms results in a greater compliance (lower modulus) of the molecule. These movements have been studied, and Heijboer[37] classified β and γ transitions by their type of motions. The specific temperature and frequency of this softening help drive the end use of the material.

Moving from very low temperature, where the molecule is tightly compressed, to higher temperatures, the first changes are the solid-state transitions. This process is shown in Figure 6. As the material warms and expands, the free volume increases so that localized bond movements (bending and stretching) and side-chain movements can occur. This is the γ transition, at temperature T_γ, which may also involve associations with water.[38] As the temperature and the free volume continue to increase, the whole side chains and localized groups of four to eight backbone atoms begin to have enough space to move and the material starts to develop some toughness.[39] This transition, called the *β transition*, at temperature T_β, is not as clearly defined as described here (Figure 6). Often it is the glass transition, T_g, of a secondary component in a blend or of a specific block in a block copolymer. However, a correlation with toughness is seen empirically.[40]

As heating continues, T_g or the glass transition appears, when the chains in the amorphous regions begin to coordinate large-scale motions (Figure 6). One classical description of this region is that the amorphous regions have begun to melt. Since T_g occurs only in amorphous

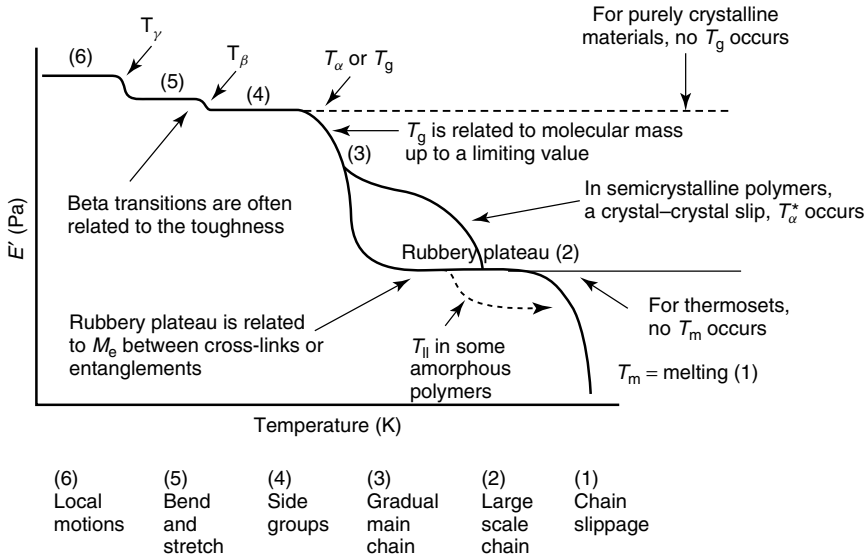

Figure 6 Idealized temperature scan of a polymer. Starting at low temperature, the modulus decreases as the molecules gain more free volume, resulting in more molecular motion. This shows the main curve as divided into six regions that correspond to local motions (6), bond bending and stretching (5), movements in the side chain or adjacent atoms in the main chain (4), the region of T_g (3), coordinated movements in the amorphous portion of the chain (2), and the melting region (1). Transitions are marked as described in the text. (Reproduced with permission of Taylor & Francis Group LLC.)

material, in a 100% crystalline material there would not be a T_g. Continued heating drives the material through the T_α^* and T_{ll}. The former occurs in a crystalline or semicrystalline polymer and is a slippage of the crystallites past each other. The latter is a movement of coordinated segments in the amorphous phase that relates to reduced viscosity. These two transitions are not universally accepted. Finally, the melt is reached and large-scale chain slippage occurs and the material flows. This is the melting temperature, T_m. For a cured thermoset, nothing happens after T_g until the sample begins to burn and degrade because the cross-links prevent the chains from slipping past each other.

This brief overview provides an idea of how an idealized polymer responds. Now a more detailed description of these transitions can be provided with some examples of their applications. The best general collection of this information is still McCrum et al.'s 1967 text.[33]

4.1.1 *Sub-T_g Transitions*

The area of sub-T_g or higher-order transitions has been extensively studied,[41–48] as these transitions are associated with mechanical properties. These transitions can sometimes be seen by DSC and TMA, but they are normally too weak or too broad to be determined by these methods. DMA, DEA, and similar techniques are usually required.[49] Some authors have also called these types of transitions[50,51] *second-order transitions* to differentiate them from the primary transitions of T_m and T_g, which involve large sections of the main chains. Boyer reviewed T_β in 1968[52] and pointed out that while a correlation often exists, T_β is not always an indicator of toughness. Bershtein and Egorov[53] have reported that this transition can be considered the 'activation barrier' for solid-phase reactions, deformation, flow or creep, acoustic damping, physical aging changes, and gas diffusion into polymers, as the activation energies for the transition and these processes are usually similar. The strength of these transitions is related to how strongly a polymer responds to those processes. These sub-T_g transitions are associated with the material properties in the glassy state. In paints, for example, peel strength (adhesion) can be estimated from the strength and frequency dependence of the subambient β transition.[54] Nylon 6,6 shows a decreasing toughness, measured as impact resistance, with declining area under the T_β peak in the tan δ curve. Figure 7 shows the relative differences in T_β compared with T_g for high-impact and low-impact nylon. It has been shown, particularly in cured thermosets, that increased freedom of movement in side chains increases the strength of the transition. Cheng et al.[55] reported for rigid rod polyimides that the

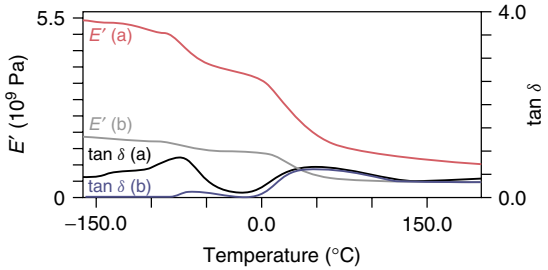

Figure 7 High- and low-impact nylon samples showing how the β transition is related to sample toughness as measured by impact testing. The (a) lines show a material with good impact strength by the falling dart test and the (b) lines show one with poor values by the same test. (Reproduced with permission of Taylor & Francis Group LLC.)

β transition is caused by the noncoordinated movement of the diamine groups, although the link to physical properties was not investigated. Johari et al. reported from both mechanical[56] and dielectric studies[57] that both the β and γ transitions in bisphenol A–based thermosets depend on the side chains and unreacted ends, and that both are affected by physical aging and post cure. Nelson[58,59] reported that these transitions can be related to vibration damping. This is also true for acoustic damping.[60] In both these cases, the strength of the β transition is taken as a measurement of how effectively a polymer will absorb vibrations. There is some frequency dependence involved in this, which will be discussed later in Section 5.7.

Boyer[61] and Heijober[37] showed that this information needs to be considered with care, as not all β transitions correlate with toughness or other properties. This can be due to misidentification of the transition or because the transition does sufficiently disperse energy. A working rule of thumb[62–66] is that the β transition must be related to either localized movement in the main chain or very large side-chain movement to absorb sufficient energy. The relationship of large side-chain movement and toughness has been extensively studied in polycarbonate by Yee and Smith[67] and also in many other tough glassy polymers.[68]

Less use is made of the T_γ transitions and they are mainly studied to understand the movements occurring in polymers. Wendorff and Schartel[69] report that this transition in polyacrylates is limited to inter- and intramolecular motions within the scale of a single repeat unit. Both McCrum et al.[33] and Boyd[70] limited T_γ and T_δ similarly to very small motions either within the molecule or with bound water. The use of what is called *two-dimensional infrared* (IR) *spectroscopy*, which couples Fourier transform infrared (FTIR) or Raman spectroscopy and DMA to study these motions, is a topic of current interest.[71,72] This couple can also be used to

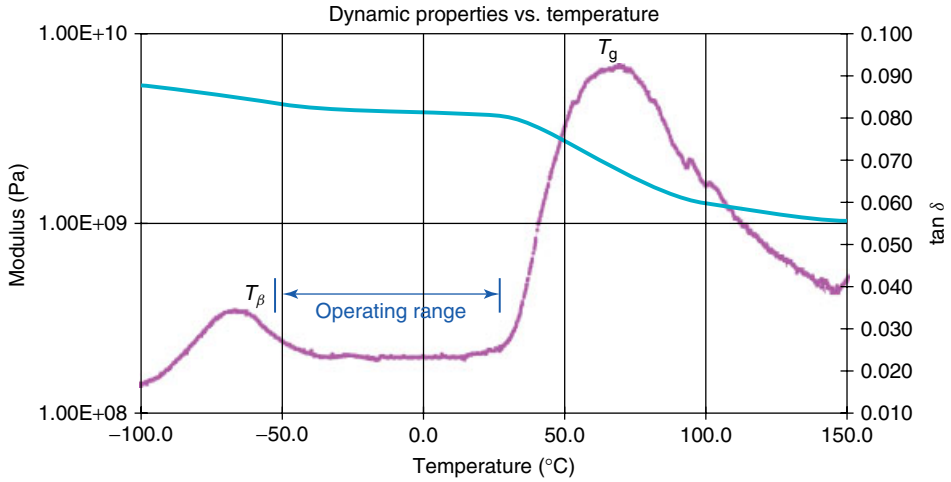

Figure 8 Definition of operating range based on position of T_g and T_β in a cured epoxy silica composite.

study the curing of various materials in which the modulus does not track linearly with the degree of cure.[73]

4.1.2 *The Glass Transition (T_g or T_α)*

As the free volume continues to increase with increasing temperature, the glass transition, T_g, occurs where large segments of the chain start moving. This transition is also called the *α transition*, T_α. T_g is very dependent on the degree of polymerization up to a value known as the *critical* T_g or the *critical molecular weight*. Above this value, T_g typically becomes independent of molecular weight.[74] T_g represents a major transition for many polymers, as the physical properties changes drastically as the material goes from a hard glassy to a rubbery state. It defines one end of the temperature range over which the polymer can be used, often called the *operating range of the polymer*, and an example of this range is shown in Figure 8. Where strength and stiffness are needed, it is normally the upper limit for use. In rubbers and some semicrystalline materials such as polyethylene and polypropylene, it is the lower operating temperature. Changes in T_g are commonly used to monitor changes in the polymer such as plasticizing by environmental solvents and increased cross-linking from thermal or ultraviolet (UV) aging.

T_g of cured materials or thin coatings is often difficult to measure by other methods, and more often than not the initial cost justification for a DMA is measuring a hard-to-find T_g. While estimates of the relative sensitivity of DMA with respect to DSC or differential thermal analysis (DTA) vary, it appears that DMA is 10–100 times more sensitive to the changes occurring at T_g. The T_g in highly cross-linked materials can easily be seen long after it has become too flat and diffuse to be seen in the DSC (Figure 9). A highly cross-linked molding resin

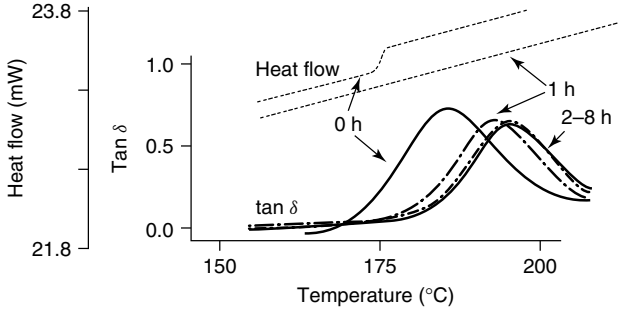

Figure 9 The T_g of a chip encapsulation material as measured by DSC and DMA as a function of postcure time. (Reprinted with permission from K. Menard, *Dynamic Mechanical Analysis: A Practical Introduction*, CRC Press, Boca Raton, FL, 1999.)

used for chip encapsulation was run by both methods, and DMA was able to detect the transition after it was undetectable in DSC.[1] This is also a known problem with certain materials such as medical-grade urethanes and very highly crystalline polyethylenes.[15]

The method of determining T_g in DMA can be a matter for disagreement, as at least five ways are currently in use (Figure 10a). It should be noted that DSC also has multiple methods (Figure 10b). Depending on the industry standards or the background of the operator, the peak or onset of the tan δ curve, the onset of the E' drop, or the onset or peak of the E'' curve may be used. The values obtained from these methods can differ by up to 25 °C from each other in the same run. In addition, a 10–20 °C difference from DSC is also seen in many materials. In practice, it is important to specify exactly how the T_g should be determined. For DMA, this means defining the heating rate, applied stresses (or strains), the frequency used, and the method of determining T_g. For example, the sample will be run at 10 °C min^{-1} under 0.05% strain at 1 Hz in a nitrogen

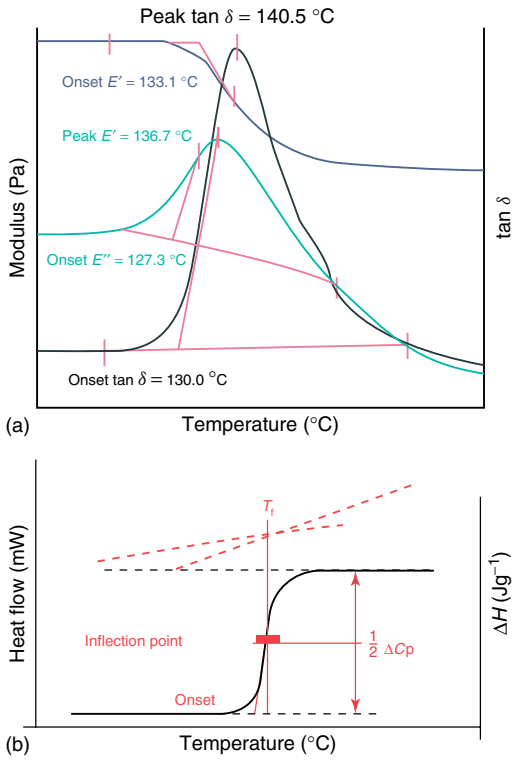

Figure 10 Methods of determining T_g. (a) Multiple methods of determining T_g are shown for DMA. T_g varies by up 10 °C in this example, depending on the value chosen. Differences as great as 25 °C have been reported. (b) Four of the methods used to determine T_g in DSC are shown. The half-height and half-width methods are not included. (Reproduced with permission of Taylor & Francis Group LLC.)

purge (20 $cm^3\,min^{-1}$) and the T_g determined from the peak of the tan δ curve.

It is not unusual to see a peak or hump on the storage modulus directly preceding the drop that corresponds to the T_g. This is shown in Figure 11. This is also seen in the DSC and DTA and corresponds to a rearrangement in the molecule to relieve stresses frozen in below T_g by the processing method. These stresses are trapped in the material until enough mobility is obtained at T_g to allow the chains to move to a lower energy state. Often a material will be annealed by heating it above T_g and cooled to remove this effect. For similar reasons, some experimenters will run a material twice or use a heat–cool–heat cycle to eliminate processing effects.

4.1.3 *The Rubbery Plateau, T_α^* and T_{ll}*

The area above the T_g and below the melt is known as the *rubbery plateau*, and its length and viscosity are dependent on the molecular weight between entanglements (M_e)[75] or cross-links. The molecular weight between entanglements is normally calculated during a stress–relaxation experiment but similar behavior is observed in DMA. The modulus in the plateau region is proportional to either the number of cross-links or the chain length between entanglements. This is often expressed in shear as Equation (21):

$$G' \approx \frac{\rho RT}{M_e} \tag{21}$$

where G' is the shear storage modulus of the plateau region at a specific temperature, ρ is the polymer density, and M_e is the molecular weight between entanglements. In practice, the relative modulus of the plateau region shows the relative changes in M_e or the number of cross-links compared with a standard material.

The rubbery plateau is also related to the degree of crystallinity in a material, although DSC is a better method for characterizing crystallinity than DMA.[76,77] Also as in DSC, there is evidence of cold crystallization in the temperature range above T_g (Figure 12). That is one of several transitions that can be seen in the rubbery plateau region. This crystallization occurs when the polymer chains have been quenched (quickly cooled)

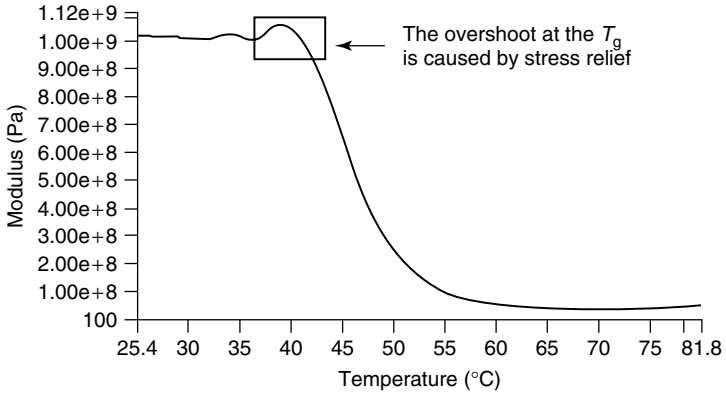

Figure 11 Stress relief at T_g in DMA. The overshoot is similar to that seen in DSC and is caused by molecular rearrangements that occur owing to the increased free volume at the transition. (Reproduced with permission of Taylor & Francis Group LLC.)

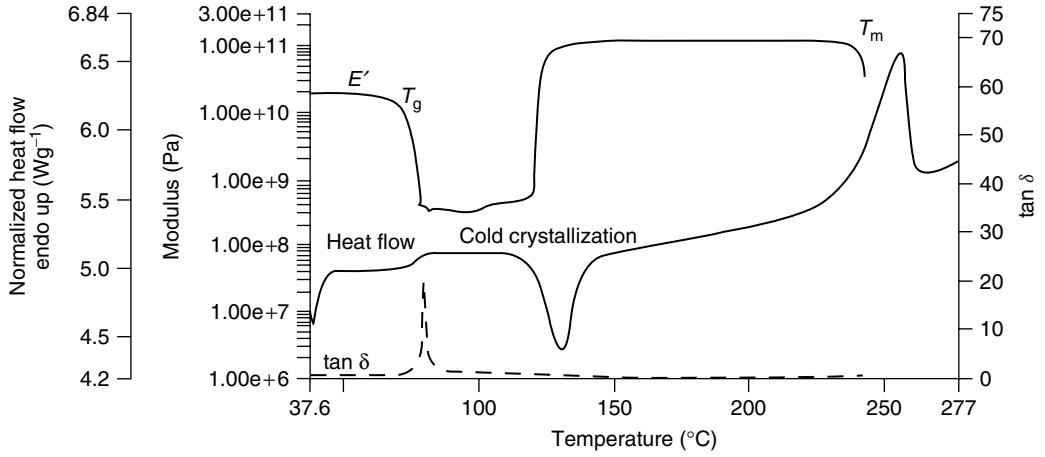

Figure 12 Cold crystallization in poly(ethylene terephthalate) causing a large increase in the storage modulus, E', above T_g. A DSC scan of the same material is included. (Reproduced with permission of Taylor & Francis Group LLC.)

into a highly disordered state. On heating above T_g, these chains gain enough mobility to rearrange into crystallites, which causes sometimes a dramatic increase in modulus. DSC or its temperature-modulated variant, dynamic differential scanning calorimetry (DDSC), can be used to confirm this.[78–81] The α^* transition, T_α^*, the liquid–liquid transition, T_{ll}, the heat set temperature, and the cold crystallization peak are all transitions that can appear on the rubbery plateau. In some crystalline and semicrystalline polymers, a transition called the α^* *transition* (T_α^*) is seen.[82,83] Figure 13 shows this in a sample of polypropylene. The α^* transition is associated with the slippage between crystallites, which helps extend the operating range of a material above T_g. This transition is very susceptible to processing-induced changes and can be enlarged or decreased by the applied heat history, processing conditions, and physical aging.[84,85] T_α^* has been used by fiber manufacturers to optimize properties in their materials.

In amorphous polymers, the T_{ll} transition is seen instead, which is a liquid–liquid transition associated with increased chain mobility and segment–segment

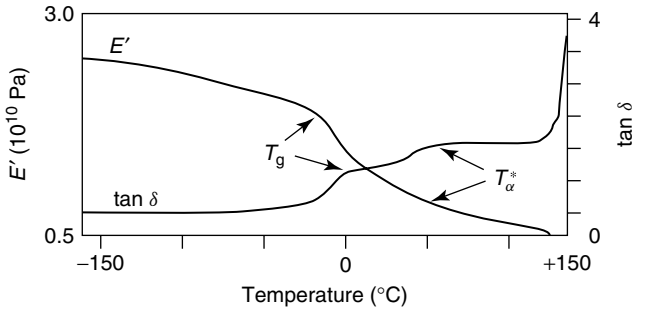

Figure 13 The α^* transition, T_α^*, in polypropylene corresponding to a crystal–crystal slip in the polymer. (Reproduced with permission of Taylor & Francis Group LLC.)

associations.[86] This order is lost when T_{ll} is exceeded and regained on cooling from the melt. Boyer[87–89] reported that, like T_g, the appearance of T_{ll} is affected by the heat history. T_{ll} is also dependent on the number-average molecular weight, M_n, but not on the weight-average molecular weight, M_w. Bershtien et al.[90] suggest that this may be considered as quasimelting on heating or the formation of stable associates of segments on cooling. While this transition is reversible, it is not always easy to see, and Boyer[87] spent many years trying to prove it was real. Not everyone accepts the existence of this transition. This transition may be similar to some of the data from temperature-modulated DSC experiments showing a recrystallization at the start of the melt.[91–94] In both cases, some subtle changes in structure are sometimes detected at the start of melting. Following this transition, a material enters the terminal or melting region.

Depending on its strength, the heat set temperature can also be seen in DMA. While it is normally seen in either a TMA or a constant gauge length (CGL) experiment, it will sometimes appear as either a sharp drop in storage modulus (E') or an abrupt change in probe position. Heat set is the temperature at which some strain or distortion is induced into polymeric fibers to change its properties, such as to prevent a nylon rug from feeling like a fishing line. Since heating above this temperature will erase the texture, and polyesters must be heated above the T_g to dye them, it is of critical importance to the fabric industry. Many final properties of polymeric products depend on changes induced in processing.[95,96]

4.1.4 The Terminal Region

On continued heating, the melting point, T_m, is reached. The melting point is where the free volume has increased

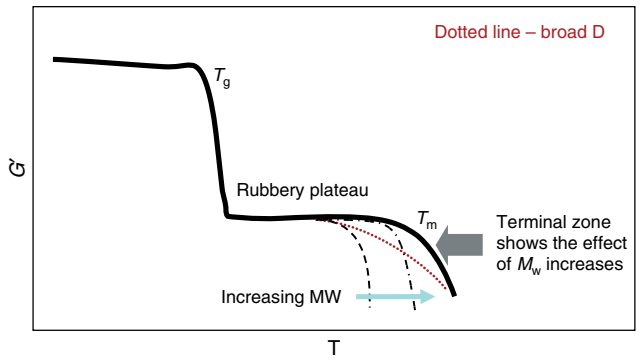

Figure 14 Molecular weight and flow: the terminal zone or melting region follows the rubbery plateau and is sensitive to the M_w of the polymer. (Reproduced with permission of Taylor & Francis Group LLC.)

so that the chains can slide past each other and the material flows. This is also called the *terminal region*. In the molten state, this ability to flow is dependent on the molecular weight of the polymer (Figure 14). The melt of a polymer material will often show changes in temperature of melting, width of the melting peak, and enthalpy as the material changes[97,98] resulting from changes in the polymer molecular weight and crystallinity.

Degradation, polymer structure, and environmental effects all influence the changes that occur. Polymers that degrade by cross-linking will look very different from those that exhibit chain scission. Very highly cross-linked polymers will not melt, as they are unable to flow.

The study of polymer melts and especially their elasticity was one of the areas that drove the development of commercial DMA instruments. Although a decrease in the melt viscosity is seen with temperature increases, DMA is most commonly used to measure the frequency dependence of the molten polymer and its elasticity. The latter property, especially when expressed as the normal forces, is very important in polymer processing.

Recently, methods have been developed to hold and handle powders in the DMA. These Material Pockets allow the holding of materials that can support their own weight. This allows the measurement of the glass transition not measurable in substances such as amorphous lactose, protein excipient mixtures, and amorphous drugs.

4.1.5 *Frequency Dependences in Transition Studies*

The choice of a testing frequency or its effect on the resulting data must be addressed. A short discussion of how frequencies are chosen and how they affect the measurement of transitions is in order. Considering that higher frequencies induce more elastic like behavior, there is some concern that a material will act stiffer than it really is if the test frequency is chosen to be too high. Frequencies for testing are normally chosen by one of three methods. The most scientific method would be to use the frequency of the stress or strain to which the material is exposed in the real world. However, this is often outside the range of the available instrumentation. In some cases, the test method or the industry standard sets a certain frequency and this frequency is used. Ideally, a standard method such as this is chosen so that the data collected on various commercial instruments can be shown to be compatible. Some of the American Society for Testing and Materials (ASTM) methods for DMA are listed in Table 1. Many industries have their own standards, so it is important to know whether the data are expected to match a Mil-spec, an ASTM standard, or a specific industrial test. Finally, one can arbitrarily pick a frequency. This is done more often than not, and 1 Hz and 10 rad s^{-1} are often used. As long as the data are obtained under the proper conditions, they can be compared to highlight material differences. This requires that frequency, stresses, and the thermal program be the same for all samples in the data set.

Lowering the frequency shifts the temperature of a transition to a lower temperature (Figure 15). At one time, it was suggested that multiple frequencies could be used and T_g should then be determined by extrapolation to

Table 1 ASTM tests for DMA[a]

- D4065-06 Standard Practice for Plastics: Dynamic Mechanical Properties: Determination and Report of Procedures[b]
- D6456-99(2004) Standard Specification for Finished Parts Made from Polyimide Resin
- E1640-04 Standard Test Method for Assignment of the Glass Transition Temperature By Dynamic Mechanical Analysis
- E1867-01 Standard Test Method for Temperature Calibration of Dynamic Mechanical Analyzers
- E1867-06 Standard Test Method for Temperature Calibration of Dynamic Mechanical Analyzers
- E1953-07 Standard Practice for Description of Thermal Analysis and Rheology Apparatus
- E2254-03 Standard Test Method for Storage Modulus Calibration of Dynamic Mechanical Analyzers
- E2425-05 Standard Test Method for Loss Modulus Conformance of Dynamic Mechanical Analyzers
- WK278 New Test Method for Glass Transition Temperature (T_g) DMA of Polymer Matrix Composites
- WK3113 Loss Modulus Calibration or Conformance of DMAs
- WK3645 Standard Test Method for Temperature Calibration of DMAs
- WK10344 Standard Terminology: Plastics: Dynamic Mechanical Properties

CTE, coefficient of thermal expansion.
[a] Reproduced with permission of Taylor & Francis Group LLC.
[b] This standard qualifies a DMA as acceptable for all ASTM DMA standards.

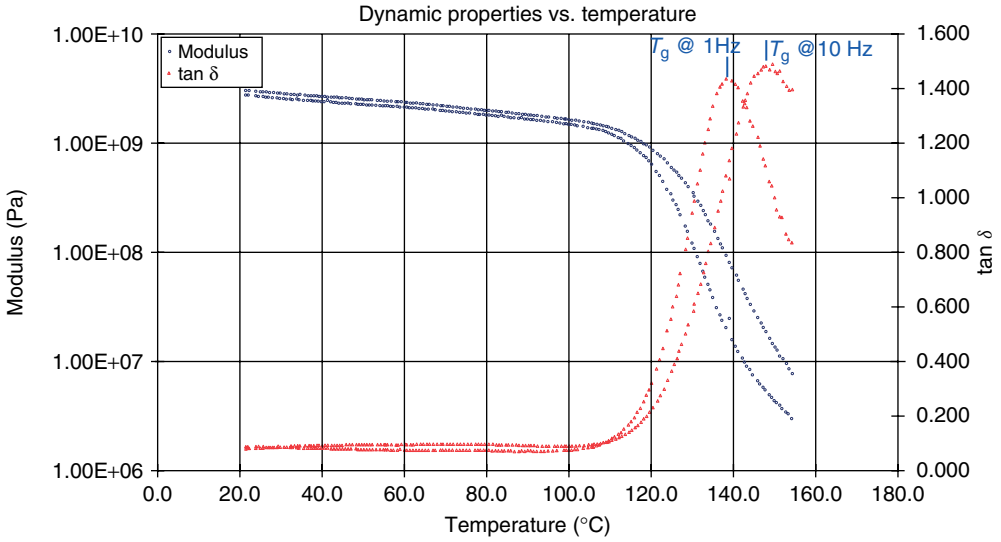

Figure 15 Effect of frequency on transitions: the dependence of T_g in polymethyl methacrylate (PMMA) on frequency.

0 Hz. This has not been accepted, but multiple frequency runs are quite useful. Different types of transitions also have different frequency dependences; McCrum et al.[33] have listed many of these. If one looks at the slope of the temperature dependence of transitions against frequency, one sees that in many cases the primary transitions such as T_m and T_g have a different dependence on frequency than the lower temperature transitions. In fact, the activation energies are different for α, β, and γ transitions because of the different motions required and the transitions can be sorted by this approach.[1–3,10,51] Running a sample at two or three frequencies in a temperature scan tells you about the type of transition measured and also gives you an estimate of the degree of shrift per decade.[15] The frequency scan will be discussed further below.

Frequency scans are often done on these materials to generate master curves using the Williams–Landel–Ferry (WLF) equation.[9] A material will be isothermally scanned across a selected frequency range, normally the full range of the instrument, and then heated to another temperature where the process repeats. One temperature is selected as the reference temperature and the data is shifted around it. The resulting master curve allows estimation of E' and tan δ over a very wide temperature range. An example is seen in Figure 16. The WLF time–temperature superposition makes some specific assumptions and these need to be followed if the data is to be useful.[1,9,15] Superposition can also be preformed with stress[99] as well as using other models.[100,101]

4.2 Polymer Melts and Solutions

A fluid or polymer melt responds to strain rate rather than to the amount of stress applied. The viscosity is one of the main reasons why frequency scans are run. As the stress–strain curves and the creep recovery runs show, viscoelastic materials exhibit some degree of flow or unrecoverable deformation. The effect is strongest in melts and liquids in which frequency versus viscosity plots are the major application of DMA.

Figure 17 shows a frequency scan on a viscoelastic material. In this example, the sample is a rubber above T_g in three-point bending, but the trends and principles apply to both solids and melts. The storage modulus and complex viscosity are plotted on logarithmic scales against log (frequency). In analyzing the frequency scans, trends in the data are more significant than specific peaks or transitions.

Starting with the viscosity curve, η^*, a fairly flat region appears at low frequency, called the *zero shear plateau*.[102] This is where the polymer exhibits Newtonian behavior and its viscosity is dependent on molecular weight, and not the strain rate. The viscosity of this plateau has been shown to be related experimentally to the molecular weight for a Newtonian fluid (Equation 22):

$$\eta_0 \propto cM_v^1 \tag{22}$$

For cases where the molecular weight, M_v, is less than the entanglement molecular weight, M_e and for cases where M_v is greater than M_e (Equation 23)

$$\eta_0 \propto cM_v^{3,4} \tag{23}$$

where η_0 is the viscosity of the initial Newtonian plateau, c is a material constant, and M_v is the viscosity-average molecular weight. This relationship can be written in

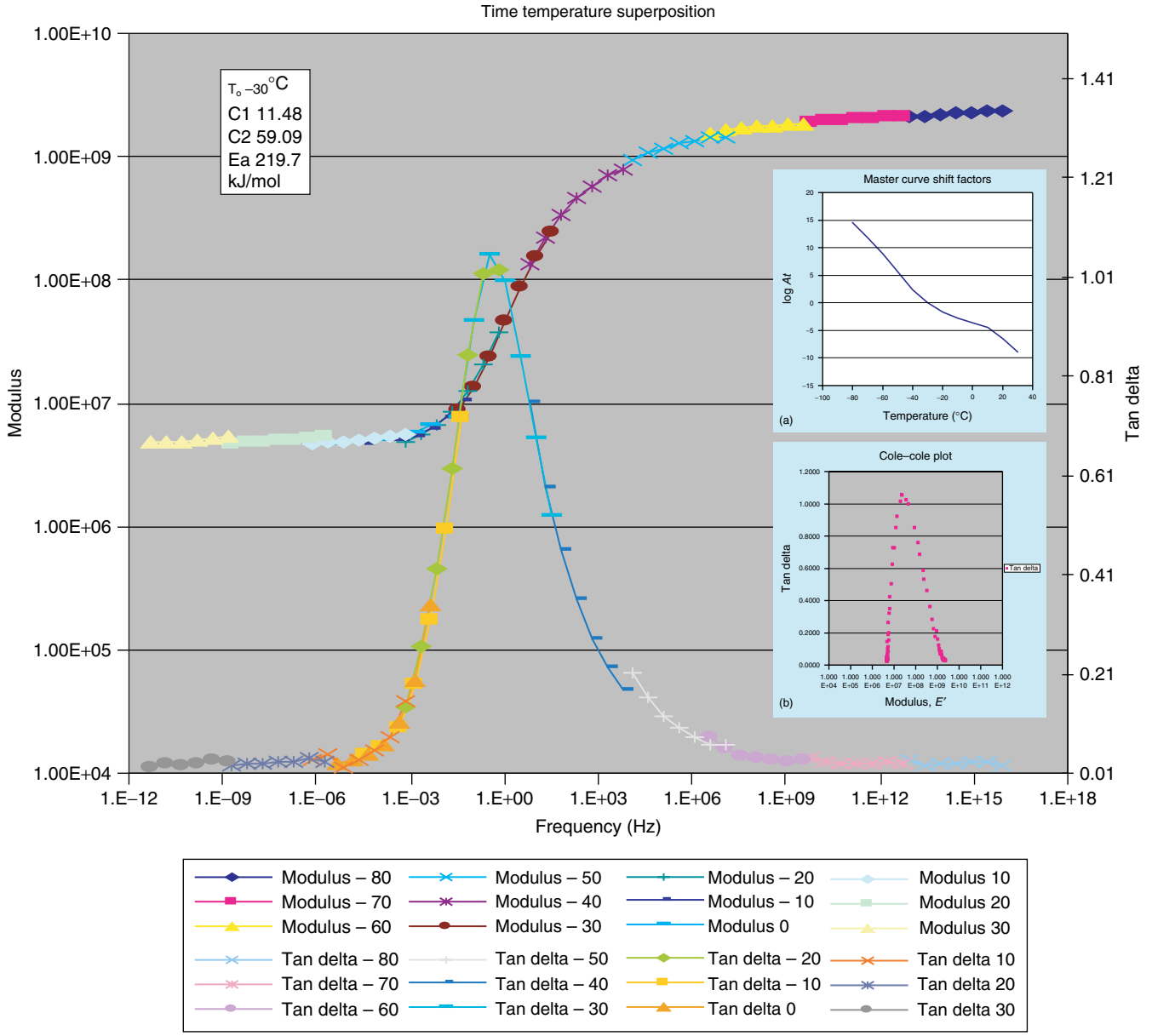

Figure 16 The time–temperature–superposition (TTS) of a rubber is shown. The insert shows the plot of shift constants in (a) and the Cole-Cole Plot in (b).

general terms, replacing the exponential term with the Mark–Houwink constant, a. Equation (23) can be used as a method of approximating the molecular weight of a polymer. The value obtained is closest to the viscosity-average molecular weight obtained by osmometry.[103] In comparison with the weight-average data obtained by gel permeation chromatography (GPC), the viscosity-average molecular weight would be between the number-average and weight-average molecular weights, but closer to the latter.[104] This was originally developed for steady shear viscosity but also applies to complex viscosity.

The relationship between steady shear and complex viscosity is fairly well established. Cox and Merz[105,106] found that an empirical relationship exists between complex viscosity and steady shear viscosity when the shear rates are the same. The Cox–Merz rule is stated as (Equation 24):

$$|\eta(\omega)| = \eta(\dot{\gamma})|_{\dot{\gamma}=\omega} \tag{24}$$

where η is the constant-shear viscosity, η^* is the complex viscosity, ω the frequency of the dynamic test, and $d\gamma/dt$ the shear rate of the constant-shear test. This rule of

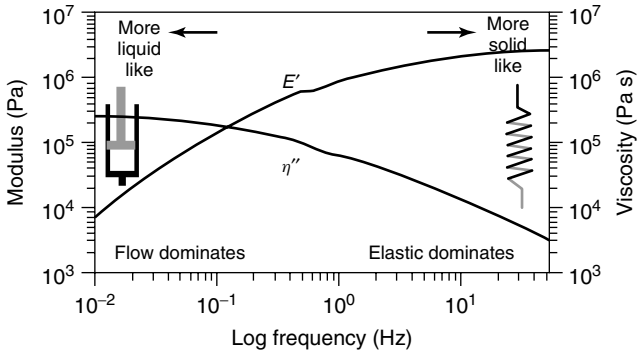

Figure 17 An example of a frequency scan showing the change in a material's behavior as the frequency varies. Low frequencies allow the material time to relax and respond, and hence flow dominates. High frequencies do not allow relaxation and elastic behavior dominates. (Reproduced with permission of Taylor & Francis Group LLC.)

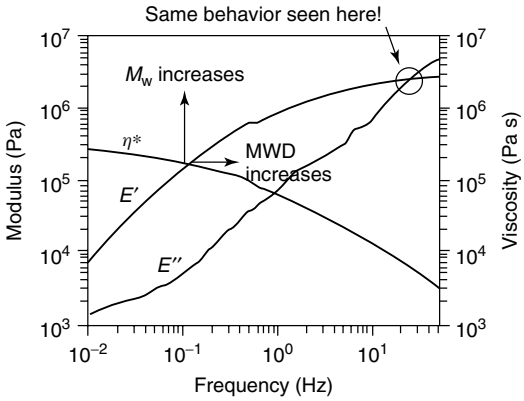

Figure 18 The crossover point between either E' and E'' or between E' and η^* for a material corresponding to the relative molecular weight and molecular weight distribution. (Reproduced with permission of Taylor & Francis Group LLC.)

thumb seems to hold for most materials to within about ±10%. Another approach is the Gleissele[107] mirror relationship, which states (Equation 25):

$$\eta\dot{\gamma} = \eta^{+}(t)|_{t=1/\dot{\gamma}} \tag{25}$$

when $\eta^{+}(t)$ is the limiting value of the viscosity as the shear rate, $\dot{\gamma}$, approaches zero.

The low-frequency range is where viscous or liquid like behavior predominates. If a material is stressed over long enough times, some flow occurs. As time is the inverse of frequency, this means that materials are expected to flow more at low frequency. As the frequency increases, the material will act in a more and more elastic fashion. Silly Putty, the children's toy, shows this clearly. At low frequency, Silly Putty flows like a liquid, whereas at high frequency it bounces like a rubber ball.

This behavior is also similar to what happens with temperature changes. A polymer becomes softer and more fluid as it is heated and it goes through transitions that increase the available space for molecular motions. Over long enough time periods, or small enough frequencies, similar changes occur. Hence one can move a polymer across a transition by changing the frequency. This relationship is also expressed as the idea of time–temperature equivalence.[108] Often stated as low-temperature equivalent to short times or high frequency, it is a fundamental rule of thumb in understanding polymer behavior.

As the frequency is increased in a frequency scan, the Newtonian region is exceeded and a new relationship develops between the rate of strain, or the frequency, and the viscosity of the material. This region is often called the *power-law zone* and can be modeled by Equation (26):

$$\eta^* \approx \eta(\dot{\gamma}) = \mathrm{c}\dot{\gamma}^{n-1} \tag{26}$$

where η^* is the complex viscosity, $\dot{\gamma}$ is the shear rate, and the exponent term n is determined by the fit of the data. It can also be written as Equation (27):

$$\sigma \approx \eta(\dot{\gamma}) = c\dot{\gamma}^{n} \tag{27}$$

where σ is the stress and η is the viscosity. The exponential relationship is why the viscosity versus frequency plot is traditionally plotted on a logarithmic scale. With modern curve-fitting programs, the use of log–log plots has declined and is rather anachronistic. The power-law region of polymers shows shear thickening or thinning behavior. This is also the region in which the E'–η^* or the E'–E'' crossover point is found. As the frequency increases and shear thinning occurs, the viscosity (η^*) decreases. At the same time, increasing the frequency increases the elasticity (E'). This is shown in Figure 18. The E'–η^* crossover point is used as an indicator of the molecular weight and molecular weight distribution.[109] Changes in its position are used as a quick method of detecting changes in the molecular weight and distribution of a material. After the power-law region, another plateau is seen, the infinite shear plateau.

This second Newtonian region corresponds to where the shear rate is so high that the polymer no longer shows a response to increases in the shear rate. At the very high shear rates associated with this region, the polymer chains are no longer entangled. This region is seldom seen in DMA experiments and is usually avoided in use because of the damage done to the chains. It can be reached in commercial extruders and causes degradation of the polymer, which causes the poorer properties associated with regrind.

As the curve in Figure 18 shows, the modulus also varies as a function of the frequency. A material exhibits more elastic-like behavior as the testing frequency increases,

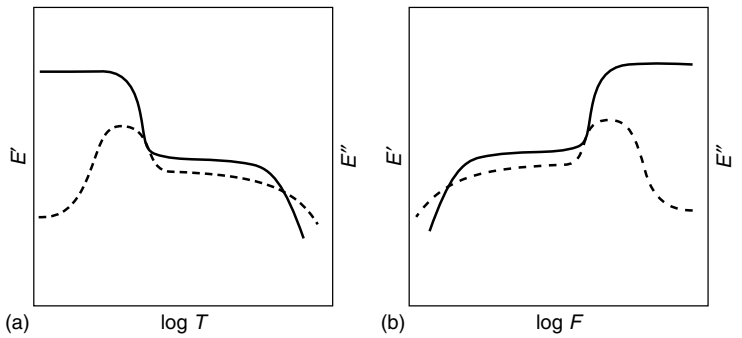

Figure 19 Comparison of a modulus scan taken by scanning at various frequencies and by varying the temperature. This relationship is called *time–temperature equivalency* and is discussed later. (Reproduced with permission of Taylor & Francis Group LLC.)

and the storage modulus tends to slope upward toward higher frequency. The change in storage modulus with frequency depends on the transitions involved. Above T_g, the storage modulus tends to be fairly flat with a slight increase with increasing frequency as it is on the rubbery plateau. The change in the region of a transition is greater. If one can generate a modulus scan over a wide enough frequency range (Figure 19a and b), the plot of storage modulus versus frequency appears like the reverse of a temperature scan. The same time–temperature equivalence as discussed above also applies to modulus, in addition to compliance, tan δ, and other properties.

The frequency scan is used for several purposes, which will be discussed in this section. One very important use that is very straightforward is to survey the material's response over various shear rates. This is important because many materials are used under different conditions. For example, adhesives, whether tape, Band-Aids, or hot melts are normally applied under conditions of low frequency and this property is referred to as *tack*. When they are removed, the removal often occurs under conditions of high frequency called *peel*. Different properties are required in these regimes and to optimize one property may require chemical changes that harm the other. Similarly, changes in polymer structure can show these kinds of differences in the frequency scan. For example, branching affects different frequencies.[36]

For example, in a tape adhesive, sufficient flow under pressure at low frequency is desired to fill the pores of the material to obtain a good mechanical bond. When the laminate is later subjected to peel, the material needs to be very elastic so that it will not pull out of the pores (this is a very simplified version of adhesion; a detailed discussion can be found elsewhere).[110,111] The frequency scan allows the measurement of these properties in one scan, thereby ensuring that tuning one property does not degrade another. This type of testing is not limited to adhesives, as many materials see multiple frequencies in actual use. Viscosity versus frequency plots are used extensively to study how changes in polymer structure or formulations affect the behavior of the melt. Often, changes in materials, especially in uncured thermosetting resins and molten materials, affect a limited frequency range and testing at a specific frequency can miss the problem.

It should be noted that when a solid material is scanned across a frequency range, there are some conditions where the material–instrument system acts like a guitar string and begins to resonate when certain frequencies are reached. These frequencies are either the natural resonance frequency of the sample–instrument system or one of its harmonics. This is shown in Figure 20. Under this set of experimental conditions, the sample–instrument system is oscillating like a guitar string and the desired information about the sample is obscured. Since there is no way to change this resonance behavior, as it is a function of the system (in a free resonance analyzer, this effect is necessary), it is required to redesign the experiment by changing the sample dimensions or geometry to escape the problem. Using a sample with very different dimensions, which changes the mass or changing from extension to three-point bending

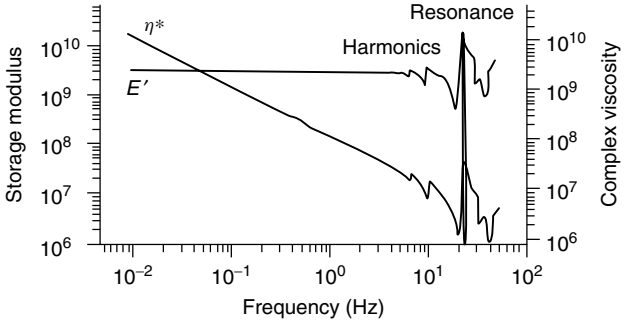

Figure 20 Free resonance occurring during a frequency scan. (With permissions from PerkinElmer Life and Analytical Sciences, Shelton, CT.)

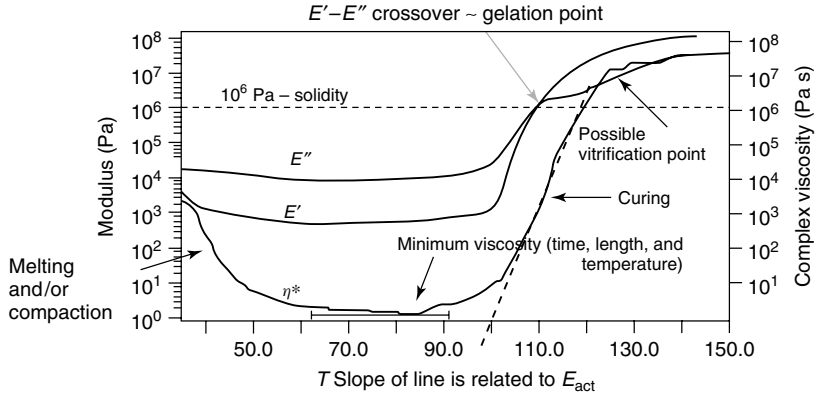

Figure 21 The DMA cure profile of a two-part epoxy showing the typical analysis for minimum viscosity, gel time, vitrification time, and estimation of the action energy. See discussion in text. (Reproduced with permission of Taylor & Francis Group LLC.)

geometry changes the natural oscillation frequency of the sample and hopefully solves this problem.

4.3 Thermoset Curing

The ability of DMA to give viscosity and modulus values for each point in a temperature scan allows the estimation of kinetic behavior as a function of viscosity. This has the advantage of describing how fluid the material is at any given time, so as to determine the best time to apply pressure, what design of tooling to use, and when the material can be removed from the mold. The simplest way to analyze a resin system is to run a plain temperature ramp from ambient to some elevated temperature.[112–116] This ''cure profile'' allows the collection of several vital pieces of information as shown in Figure 21.

4.3.1 Cure Studies in the Dynamic Mechanical Analysis

Before analyzing the cure in Figure 21 in more detail, it should be mentioned that in curing studies all three types of commercial DMA instruments are used. The shape of curve and the temperature of events follow the same pattern. The values for viscosity and modulus often differ greatly. Both types of forced resonance DMA instruments also use samples impregnated into fabrics in techniques that are referred to as *torsional braid*. There are some problems with this technique, as temperature increases it will cause an apparent curing of nondrying oils because thermal expansion increases friction. However, the ''soaking of resin into a shoelace'', as this technique has been called, allows one to handle difficult specimens under conditions where the pure resin is impossible to run in bulk (owing to viscosity or evolved volatiles). Composite materials such as graphite–epoxy composites are sometimes studied in industrial situations as the composite rather than the ''neat'' or pure resin because of the concern that the kinetics may be significantly different. In terms of ease of handling the sample, the composite is often easier to work with.

Another special area of concern is paints and coatings[117,118] where the material is used in a thin layer. This can be addressed experimentally by either a braid as above or coating the material on a thin sheet of metal. The metal is often run first and its scan subtracted from the coated sheet's scan to leave only the scan of the coating. This is also done with thin films and adhesive coatings.

A sample cure profile for a commercial two-part epoxy resin is shown in Figure 21. From this scan, it is possible to determine the minimum viscosity (η^*_{min}), the time to η^*_{min} and the length of time it stays there, the onset of cure, the point of gelation where the material changes from a viscous liquid to a viscoelastic solid, and the beginning of vitrification. The minimum viscosity is seen in the complex viscosity curve, and it is where the resin viscosity is the lowest. The minimum viscosity of a given resin is determined by its chemistry, the previous heat history of the resin, the rate at which the temperature is increased, and the amount of stress or stain applied. Increasing the rate of the temperature ramp is known to decrease the η^*_{min}, the time to η^*_{min}, and the gel time. Not only the resin becomes softer soon but also cures faster. The degree of flow limits the type of mold design and when and how much pressure can be applied to the sample. The time spent at the minimum viscosity plateau is the result of a competitive relationship between the material's softening or melting, as it heats, and its rate of curing. At some point, the material begins to cure faster than it softens, and that is where the viscosity starts to increase.

As the viscosity begins to climb, an inversion is seen of the E'' and E' values as the material becomes more solid like. This crossover point also corresponds to where the tan $\delta = 1$ (since $E' = E''$ at the crossover). This is taken to be the gel point,[15,19,22,73,99–101,119–164] where the cross-links have progressed to forming an

''infinitely'' long network across the specimen. At this point, the sample will no longer dissolve in the solvent. Although the gel point correlates fairly often with this crossover, it does not always do so. For example, for low initiator levels in chain addition thermosets, the gel point precedes the modulus crossover.[122] A temperature dependence for the presence of the crossover has also been reported.[112–116] In some cases where powder compacts and melts before curing, there may be several crossovers.[123] Then, the one following the η^*_{min} is usually the one of interest. Some researchers[124–126] believe that the true gel point is best detected by measuring the frequency dependence of the crossover point. This is done either by multiple runs at different frequencies or by multiplexing frequencies during the cure. At the gel point, the frequency dependence disappears.[126] The value is usually only a few degrees different from that obtained in a normal scan. During this rapid climb of viscosity in the cure, the slope for η^* increase can be used to calculate an estimated E_a (activation energy).[127] This will be discussed in the subsequent text, but the fact that the slope of the curve here is a function of E_a is important. Above the gel temperature, some workers estimate the molecular weight, M_c, between cross-links according to Equation (28):

$$G' = \frac{RT\rho}{M_c} \tag{28}$$

where R is the gas constant, T is the temperature in kelvins, and ρ is the density. At some point the curve begins to level off and this is often taken as the vitrification point, T_{vf}.

The vitrification point is where the cure rate slows because the material has become so viscous that the bulk reaction has stopped. At this point, the rate of cure slows significantly. The apparent T_{vf}, however, is not always real: any analyzer has an upper force limit. When that force limit is reached, the 'topping out' of the analyzer can pass as the T_{vf}. Use of a combined technique such as DMA/DEA (DEA is dielectric analysis, where an oscillating electrical signal is applied to a sample; from this signal, the ion mobility can be calculated, which is then converted to a viscosity (see McCrum et al.[20] for details; DEA will measure up to significantly higher viscosities than DMA) to see the higher viscosities or the removal of a sample from a parallel plate and sectioning it into a flexure beam is often necessary to see the true vitrification point. A reaction can also completely cure without vitrifying and will level off in the same way. One should be aware that reaching vitrification or complete cure too quickly could be as bad as reaching it too slowly. Often, an overly aggressive cure cycle will lead to a weaker material, as it does not allow for as much network development, but gives a series of hard (highly cross-linked) areas among softer (lightly cross-linked) areas.

On the way to vitrification, an important value is 10^6 Pa s. This is the viscosity of bitumen[128] and is often used as a rule of thumb for which a material is stiff enough to support its own weight. This is a rather arbitrary point, but is chosen to allow the removal of materials from a mold and the cure is then continued as a postcure step. As will be seen below, the postcure is often a vital part of the curing process.

The cure profile is both a good predictor of performance and a sensitive probe of processing conditions. A final note on cure profiles is that a volume change occurs during the cure.[129] This shrinkage of the resin is important and can be studied by monitoring the probe position of some DMA instruments and also by TMA and dilatometry.

4.3.2 *Photocuring*

A photocure in a DMA instrument is run by applying a UV light source to a sample that is held at a specific temperature or subjected to a specific thermal cycle.[130–132] Photocuring is done for dental resin, contact adhesives, and contact lenses. UV exposure studies are also run on cured and thermoplastic samples by the same techniques as used in photocuring to study UV degradation. The cure profile of a photocure is very similar to that of a cake or epoxy cement. The same analysis is used and the same types of kinetics are developed as for thermal curing studies. The only difference is that the cure is now driven by light rather than temperature.

While the PerkinElmer DMA-7e had been successfully adapted[133] to use quartz fixtures, newer instruments like the PerkinElmer DMA 8000 are now available with UV curing accessory as complete packages. Besides photocuring, these instruments can also be used for degradation studies under high-intensity UV or visible light.[1]

4.3.3 *Modeling Cure Cycles*

The above discussions are based on using a simple temperature ramp to see how a material responds to heating. In actual use, many thermosets are actually cured using more complex cure cycles to optimize the trade-off between the processing time and the final product's properties.[134,135] The use of two-stage cure cycles is known to develop stronger laminates in the aerospace industry. Exceptionally thick laminates often also require multiple stage cycles in order to develop strength without porosity. As thermosets shrink on curing,

careful development of a proper cure cycle to prevent or minimize internal voids is necessary.

One reason for the use of multistage cures is to drive reactions to completion. Another is to extend the minimum viscosity range to allow greater control in forming or shaping of the material. The development of a cure cycle with multiple ramps and holds would be very expensive if done with full-sized parts in production facilities. The use of DMA gives a faster and cheaper way of optimizing the cure cycle to generate the most efficient and tolerant processing conditions.

4.3.4 *Isothermal Curing Studies*

Often curing is done at a constant temperature for a period of time. This is how the data needed for the kinetic models discussed in the next section are normally collected. It is also how rubber samples are cross-linked, how initiated reactions are run, and how bulk polymerizations are performed. Industrially, continuous processes, as opposed to batch, often require an isothermal approach. UV light and other forms of nonthermal initiation also use isothermal studies for examining the cure at a constant temperature.

4.3.5 *Curing Kinetics by Dynamic Mechanical Analysis*

Several approaches have been developed to studying the chemorheology of thermosetting systems. Halley and MacKay[136] recently reviewed chemorheology and the more common kinetic models. A fundamental method is the WLF model,[137] which looks at the variation of T_g with degree of cure. This has been used and modified extensively.[138,139] A common empirical model for curing has been proposed by Roller et al.[140–145] In the latter approach, samples of the thermoset are run isothermally as described above and the viscosity versus time data are collected. This is plotted as log η^* versus time in seconds, where a change in slope is apparent in the curve. This break in the data indicates that the sample is approaching the gel time. From these curves, the initial viscosity, η_0, and the apparent kinetic factor, k, can be determined. By plotting log (viscosity) versus time for each isothermal run, the slope, k, and the viscosity at $t = 0$ are apparent. The initial viscosity and k can be expressed according to Equations (29) and (30):

$$\eta_0 = \eta_\infty e^{\Delta E_\eta / RT} \tag{29}$$

$$k = k_\infty e^{\Delta E_k / RT} \tag{30}$$

Combining these allows one to set up the equation for viscosity under isothermal conditions (Equation 31):

$$\ln \eta(t) = \ln \eta_\infty + \frac{\Delta E_\eta}{RT} + t k_\infty e^{\Delta E_k / RT} \tag{31}$$

By replacing the last term with an expression that treats temperature as a function of time, the equation becomes Equation (32):

$$\ln \eta(T, t) = \ln \eta_\infty + \frac{\Delta E_\eta}{RT} + \int_0^t k_\infty e^{\Delta E_k / RT}\, dt \tag{32}$$

This equation can be used to describe viscosity–time profiles for any run where the temperature can be expressed as a function of time. The activation energies can now be calculated. The plots of the natural logarithm of the initial viscosity (determined above) versus $1/T$ and the natural logarithm of the apparent rate constant, k, versus $1/T$ are used to give the activation energies, ΔE_η and ΔE_k. Comparison of these values of k and ΔE with those calculated by DSC shows that this model gives larger values.[117,118] The DSC data are faster to obtain, but they do not include the needed viscosity information.

Several corrections have been proposed, addressing different orders of reaction[146] (the above assumes first order) and modifications to the equations.[147,148] Many of these adjustments are reported in Roller's 1986 review[149] of curing kinetics. It is noted that these equations do not work well above the gel temperature. This same equation has been used to predict the degradation of properties in thermoplastics successfully.[150]

4.3.6 *Mapping Thermoset Behavior: the Gillham–Enns Diagram*

Another approach to attempt to understanding fully the behavior of a thermoset was developed by Gillham et al.[151–157] and is analogous to the phase diagrams used by metallurgists. The time–temperature–transformation (TTT) diagram or the Gillham–Enns diagram (after its creators) is used to track the effects of temperature and time on the physical state of a thermosetting material. Figure 22 shows an example. Running isothermal studies of a resin at various temperatures and recording the changes as a function of time can do this. One has to choose values for the various regions and Gillham et al. have done an excellent job of detailing how one picks T_g, the glass, the gel, the rubbery, and the charring regions.[158,159] These diagrams are also generated from DSC data,[160] and several variants,[161,162] such as the continuous heating transformation and conversion–temperature–property diagrams, have been

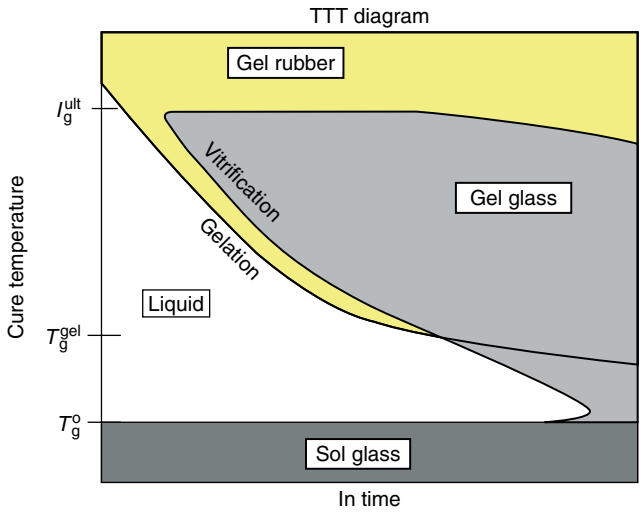

Figure 22 The Gillham–Enns or TTT diagram. (Reprinted from B. Bilyeu's Ph.D. thesis and used with his permission.)

reported. Surprisingly easy to do, although rather slow, they have not yet been accepted in the industry despite their obvious utility. A review[163] summarizes the approach, which has details on these and the continuous heating variant.

4.4 Testing Environmental Effects

One of the major advances recently in DMA is the ability to test in both solution and controlled humidity environments. Many materials' properties change on exposure to solvents and relative humidity and this is a problem for materials in use. The ability to run samples in a solvent bath or in a humidity chamber and to see how these materials change is an important addition to any testing regimen.

4.4.1 Immersion Studies

Figure 23 shows a DMA run on a sample of spaghetti run in a fluid bath with water as the solvent. As the water heats, the spaghetti softens. While this experiment does not take it to *al dente*, it makes a nice example of the approach. In most cases, the solvent acts to plasticize the material as time and temperature increase. Studies of materials in solution are numerous and include

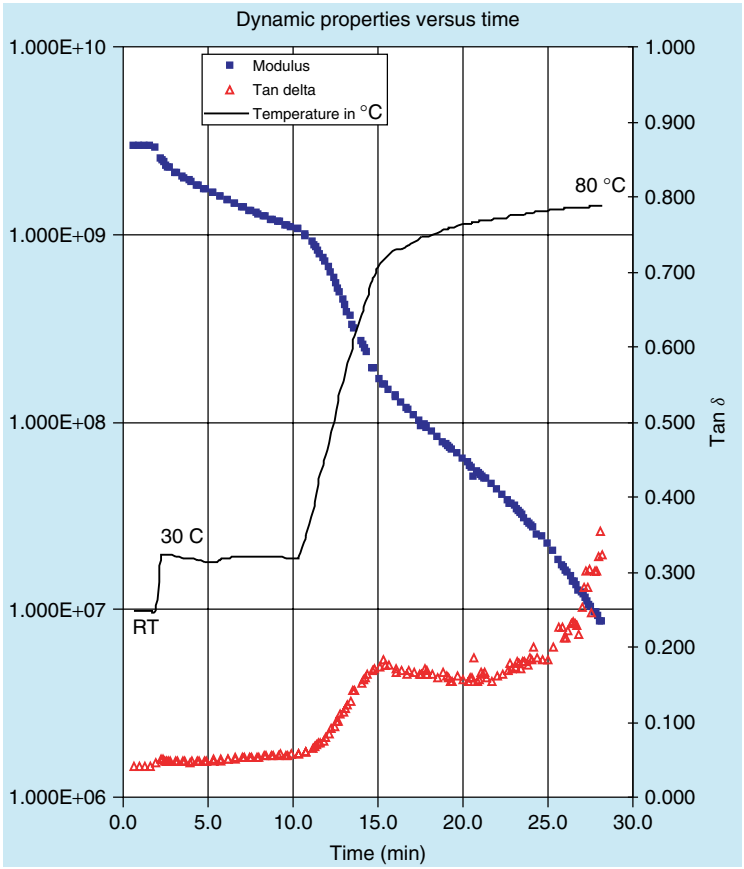

Figure 23 The effect of immersion into warm water on a pasta strand is shown. Modulus decreases and tan delta increases as soak time and temperature increase).

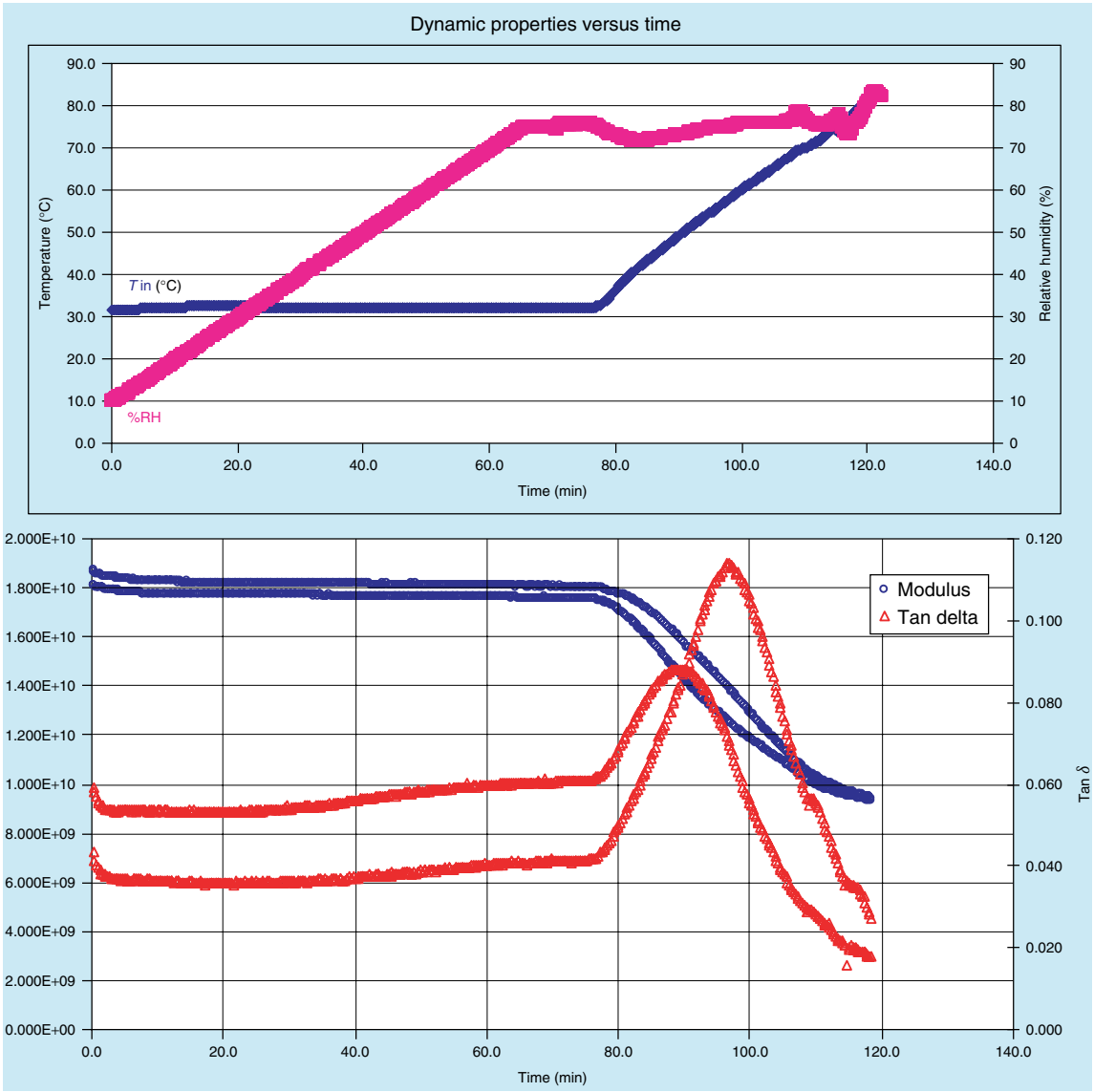

Figure 24 The affect of humidity on a sample of tea leaves run in a metallic Material Pocket shows the decrease in modulus as moisture is absorbed by the tea.

polystyrene resistance to oil,[164] coaxially coated fibers,[165] polymers,[166] aerospace materials[167] as well as dental materials, polyamide composites, epoxy films, and fishing lines. Many biomedical devices are also tested this way.

4.4.2 Humidity Studies

Anyone working with materials like paper, nylon fibers, or sugars is well aware of the effect that moisture levels changes have on T_g and the modulus values. Water is known to act as a plasticizer in polymers and most studies on the effect of water have involved soaking the material in a solution or subjecting it to a humidity chamber and then running it under dry gas. Running the samples in humidity directly allows you to see how water changes the modulus as it is absorbed as well as the effects of the relative humidity level on final properties.

Examples of the effect of humidity are shown in Figure 24. In the first case, samples of dry tea leaves were loaded into a material pocket and allowed to equilibrate at 75% RH. It was then temperature scanned to measure T_g. A second sample was tested at 40% RH. The second example shows the change in the sucrose T_g as a function of humidity. The third example shows a nylon fishing line. A range of materials have been reported in the literature including cellulose films,[168] sunflower protein isolate,[169] Nafion™ 117 films,[170] polyesters,[170] leather,[171] and polyurethanes.[172]

ACKNOWLEDGMENTS

The authors acknowledge the Materials Characterization Division of PerkinElmer Life and Analytical Sciences and the Materials Science Department of the University of North Texas (UNT) for their support and assistance. In addition, the help and advice of Professor Witold Brostow of the Materials Science Department of UNT, and the editorial staff of John Wiley & Sons Ltd is greatly appreciated.

LIST OF SYMBOLS

δ	Phase Angle
$\tan\delta$	Tangent of the phase angle, also called the *damping*
σ	Stress
γ	Shear strain
ε	Tensile strain
$\dot{\gamma}$	Shear strain rate
$\dot{\varepsilon}$	Strain rate
η	Viscosity
η^*	Complex viscosity
η'	Storage viscosity
η''	Loss viscosity
E^*	Complex modulus
E'	Storage modulus
E''	Loss modulus
J	Compliance
k	Deformation
T	Period
ρ	Density
G	Shear modulus
M_e	Entanglement molecular weight
M_c	Molecular weight between cross-links
M_w	Molecular weight
f	Frequency
w	Frequency (Hz)
k	Rate constant
E_a	Activation energy
v^f	Free volume
$T_{\alpha,\beta,\gamma}$	Transition
Λ	Logarithmic decrement
Γ	Torque

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
CGL	Constant Gauge Length
CTE	Coefficient of Thermal Expansion
DDSC	Dynamic Differential Scanning Calorimetry
DEA	Dielectric Analysis
DMA	Dynamic Mechanical Analysis
DMS	Dynamic Mechanical Spectroscopy
DMTA	Dynamic Mechanical Thermal Analysis
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
DTMA	Dynamic Thermomechanical Analysis
FTIR	Fourier Transform Infrared
GPC	Gel Permeation Chromatography
IR	Infrared
TBA	Torsional Braid Analyzer
TMA	Thermomechanical Analyzer
TTS	Time–Temperature–Superposition
TTT	Time–Temperature–Transformation
UV	Ultraviolet
WLF	Williams–Landel–Ferry

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction • Mechanical Properties of Solid Coatings • Rheology in Coatings, Principles and Methods • Thermal Analysis of Coatings

Food **(Volume 5)**
Viscosity of Food: Measurement and Application

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction • Atomic Force Microscopy in Analysis of Polymers

Polymers and Rubbers cont'd **(Volume 9)**
Dielectric Spectroscopy in Analysis of Polymers • Mechanical Properties of Polymers and Rubbers

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction • Differential Scanning Calorimetry and Differential Thermal Analysis • Inorganic Systems, Thermal Analysis Applications to • Simultaneous Techniques in Thermal Analysis • Thermogravimetry

REFERENCES

1. K. Menard, *Dynamic Mechanical Analysis: A Practical Introduction*, Taylor & Francis, New York, 2008.
2. M. Sepe, *Dynamic Mechanical Analysis for Plastic Engineering*, Plastic Design Library, New York, 1998.
3. T. Murayama, *Dynamic Mechanical Analysis of Polymeric Materials*, Elsevier Science, New York, 1977.

4. J.H. Poynting, *Proc. R. Soc. London, Ser. A or B*, **82**, 546 (1909).
5. A. Kimball, D. Lovell, *Trans. Am. Soc. Mech. Eng.*, **48**, 479 (1926).
6. K. te Nijenhuis, in *Rheology. Volume 1. Principles*, eds. G. Astarita, et al., Plenum Press, New York, 263, 1980.
7. M.L. Miller, *The Structure of Polymers*, Reinhold, New York, 1966.
8. J. Dealy, *Rheometers for Molten Plastics*, Van Nostrand Reinhold, New York, 136–137; 234–236, 1992.
9. J. Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, John Wiley & Sons, New York, 1980.
10. N. McCrum, B. Williams, G. Read, *Anelastic and Dielectric Effects in Polymeric Solids*, Dover Publications, New York, 1991.
11. J. Gilham, J. Enns, *Trends Polym. Sci.*, **2**, 406 (1994).
12. C. Macosko, J. Starita, *SPE J.*, **27**, 38 (1971).
13. T. Murayama, *Dynamic Mechanical Analysis of Polymeric Materials*, Elsevier Science, New York, 1977.
14. B.E. Read, G.D. Brown, *The Determination of the Dynamic Properties of Polymers and Composites*, John Wiley & Sons, New York, 1978.
15. J. Duncan, 'Dynamic Mechanical Analysis', in *Principles and Applications of Thermal Analysis*, ed. P. Gabbott, Blackwell Publishing, Oxford, 2008.
16. J. Gillham, *Developments in Polymer Characterizations*, ed. J. Dworkins, Applied Science, Princeton, NJ, 159–227, Vol. 3, 1982.
17. J. Gillham, J. Enns, *Trends Polym. Sci.*, **2**(12), 406 (1994).
18. U. Zolzer, H.-F. Eicke, *Rheol. Acta*, **32**, 104 (1993).
19. R. Ewodlt, G. McKinley, *Rheol. Bull.*, **76**, 4 (2007).
20. N. McCrum, G. Williams, B. Read, *Anelastic and Dielectric Properties of Polymeric Solids*, Dover Publications, New York, 192–200, 1991.
21. D. Holland, *J. Rheol.*, **38**(6), 1941–1943 (1994).
22. K. Van Ness, 'Development of a Co-Centric Cylinder for an Axial DMA', Private communication.
23. P. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York, 1953.
24. R. Bird, C. Curtis, R. Armstrong, O. Hassenger, *Dynamics of Polymer Fluids*, 2nd edition, John Wiley & Sons, New York, 1987.
25. J.D. Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, John Wiley & Sons, New York, 1980.
26. J.J. Aklonis, W.J. McKnight, *Introduction to Polymer Viscoelasticity*, 2nd edition, John Wiley & Sons, New York, 1983.
27. L.C.E. Struik, *Physical Aging in Amorphous Polymers and Other Materials*, Elsevier Science, New York, 1978.
28. L.C.E. Struik, *Failure of Plastics*, eds. W. Brostow, R.D. Corneliussen, Hanser Publishers, New York, 1986.
29. S. Matsuoka, *Failure of Plastics*, eds. W. Brostow, R.D. Corneliussen, Hanser Publishers, New York, 24–59, 1986.
30. S. Matsuoka, *Relaxation Phenomena in Polymers*, Hanser Publishers, New York, 237–264, 1992.
31. J.D. Vrentas, J.L. Duda, J.W. Huang, *Macromolecules*, **19**, 1718 (1986).
32. W. Brostow, M.A. Macip, *Macromolecules*, **22**(6), 2761 (1989).
33. N. McCrum, G. Williams, B. Read, *Anelastic and Dielectric Effects in Polymeric Solids*, Dover Publications, New York, 1967.
34. M. Doi, S. Edwards, *The Dynamics of Polymer Chains*, Oxford University Press, New York, 1986.
35. A.J. Aklonis, W. MacKnight, *Introduction to Viscoelasticity*, John Wiley & Sons, New York, 1983.
36. C.L. Rohn, *Analytical Polymer Rheology*, Hanser-Gardener Publications, New York, 1995.
37. J. Heijboer, *Int. J. Polym. Mater.*, **6**, 11 (1977).
38. N. McCrum, G. Williams, B. Read, *Anelastic and Dielectric Effects in Polymeric Solids*, Dover Publications, New York, 1967.
39. R.F. Boyer, *Polym. Eng. Sci.*, **8**(3), 161 (1968).
40. C.L. Rohn, *Analytical Polymer Rheology*, Hanser-Gardener Publications, New York, 279–283, 1995.
41. J. Heijboer, *Int. J. Polym. Mater.*, **6**, 11 (1977).
42. M. Mangion, G. Johari, *J. Polym. Sci., Part B: Polym. Phys.*, **29**, 437 (1991).
43. G. Johari, G. Mikoljaczak, J. Cavaille, *Polymer*, **28**, 2023 (1987).
44. S. Cheng, F. Arnold, F. Harris, *Polym. Eng. Sci.*, **33**, 21 (1993).
45. G. Johari, *Lect. Notes Phys.*, **277**, 90 (1987).
46. R. Daiz-Calleja, E. Riande, *Rheol. Acta*, **34**, 58 (1995).
47. R. Boyd, *Polymer*, **26**, 323 (1985).
48. V. Bershtein, V. Egorov, L. Egorova, V. Ryzhov, *Thermochim. Acta*, **238**, 41 (1994).
49. B. Twombly, *Proc. North Am. Therm. Anal. Soc.*, **20**, 63 (1991).
50. C.L. Rohn, *Analytical Polymer Rheology*, Hanser-Gardener Publications, New York, 1995.
51. J. Heijboer, *Int. J. Polym. Mater.*, **6**, 11 (1977).
52. R. Boyer, *Polym. Eng. Sci.*, **8**(3), 161 (1968).
53. V. Bershtein, V. Egorov, *Differential Scanning Calorimetry in the Physical Chemistry of Polymers*, Ellis Horwood, Chichester, 1993.

54. B. Coxton, Personal communication.
55. S. Cheng, D. Shen, K. Bruno, F. Arnold, M. Eashoo, C. Lee, F. Harris, *Polym. Eng. Sci.*, **33**, 21 (1993).
56. G. Johari, G. Mikoljaczak, J. Cavaille, *Polymer*, **28**, 2023 (1987).
57. M. Mangion, G. Johari, *J. Polym. Sci., Part B: Polym. Phys.*, **29**, 437 (1991).
58. F.C. Nelson, *Shock Vib. Dig.*, **26**(2), 11 (1994).
59. F.C. Nelson, *Shock Vib. Dig.*, **26**(2), 24 (1994).
60. W. Brostow, Personal communication.
61. R. Boyer, *Polym. Eng. Sci.*, **8**(3), 161 (1968).
62. G. Johari, *Lect. Notes Phys.*, **277**, 90 (1987).
63. J. Heijboer, *Int. J. Polym. Mater.*, **6**, 11 (1977).
64. J. Heijboer, A. Waterman, L. Struik, M. van der Duikeren, *Physics of Non-Crystalline Solids*, ed. J. Prins, Interscience, New York, 214–245, 1965.
65. J. Heijboer, *J. Polym. Sci.*, **C16**, 3755 (1968).
66. L. Nielsen, et al., *J. Macromol. Sci., Phys.*, **9**, 239 (1974).
67. A. Yee, S. Smith, *Macromolecules*, **14**, 54 (1981).
68. G. Gordon, *J. Polym. Sci., Part A2*, **9**, 1693 (1984).
69. J. Wendorff, B. Schartel, *Polymer*, **36**(5), 899 (1995).
70. R.H. Boyd, *Polymer*, **26**, 323 (1985).
71. I. Noda, *Appl. Spectrosc.*, **44**(4), 550 (1990).
72. V. Kien, in *Proceedings of the 6th Symposium on Radiation Chemistry*, P. Hedvig (ed.), Akademiai Kaido, Budapest, Hungary, Vol. 2, 463, 1987.
73. S. Newman, work in progress.
74. L.H. Sperling, *Introduction to Physical Polymer Science*, 2nd edition, John Wiley & Sons, New York, 1992.
75. C. Macosko, *Rheology*, Wiley-VCH, New York, 1994.
76. F. Quinn, et al., *Thermal Analysis*, John Wiley & Sons, New York, 1994.
77. B. Wunderlich, *Thermal Analysis*, Academic Press, New York, 1990.
78. J. Schawe, *Thermochim. Acta*, **261**, 183 (1995).
79. J. Schawe, *Thermochim. Acta*, **260**, 1 (1995).
80. J. Schawe, *Thermochim. Acta*, **271**, 1 (1995).
81. B. Wunderlich, et al., *J. Therm. Anal.*, **42**, 949 (1994).
82. R.H. Boyd, *Polymer*, **26**, 323 (1985).
83. R.H. Boyd, *Polymer*, **26**, 1123 (1985).
84. S. Godber, Personal communication.
85. M. Ahmed, *Polypropylene Fiber–Science and Technology*, Elsevier Science, New York, 1982.
86. A. Lobanov, et al., *Polym. Sci. USSR*, **22**, 1150 (1980).
87. R. Boyer, *J. Polym. Sci., Part B: Polym. Phys.*, **30**, 1177 (1992).
88. J.K. Gilham, et al., *J. Appl. Polym. Sci.*, **20**, 1245 (1976).
89. J.B. Enns, R. Boyer, *Encycl. Polym. Sci.*, **17**, 23–47 (1989).
90. V. Bershtien, V. Egorov, L. Egorova, V. Ryzhov, *Thermochim. Acta*, **238**, 41 (1994).
91. B. Cassel, G. Curran, *PETAN #69 DDSC*, Perkin-Elmer, Norwalk, CT, 1995.
92. W. Sichina, *Proc. North Am. Therm. Anal. Soc.*, **23**, 137 (1994).
93. B. Wunderlich, A. Boller, Y. Jin, *J. Therm. Anal.*, **42**, 307 (1994).
94. B. Wunderlich, *Modulated DSC*, University of Tennessee, Knoxville, TN, 1994.
95. J. Dealy, K. Wissbrun, *Melt Rheology and its Role in Plastic Processing*, Van Nostrand Reinhold, Toronto, 1990.
96. N. Chereminsinoff, *An Introduction to Polymer Rheology and Processing*, CRC Press, Boca Raton, FL, 1993.
97. E. Turi (ed.), *Thermal Characterization of Polymeric Materials*, Academic Press, Boston, MA, 1981.
98. E. Turi (ed.), *Thermal Analysis in Polymer Characterization*, Heyden, London, 1981.
99. W. Brostow, N. D'Sousa, J. Kubat, R. Maksimov, *Int. J. Polym. Mater.*, **43**, 233 (1999).
100. A. Akinay, W. Brostow, R. Maksimov, *Polym. Eng. Sci.*, **41**, 6 (2001).
101. W. Brostow, J. Mano, J. Lopes, R. Silva, *Polymer*, **44**, 4293 (2003).
102. C. Rohn, *Analytical Polymer Rheology*, Hanser Publishers, New York, 1995.
103. M. Miller, *The Structure of Polymers*, Reinhold, New York, 611–612, 1966.
104. S. Rosen, *Fundamental Principles of Polymeric Materials*, Wiley–Interscience, New York, 53–77; 258–259, 1993.
105. W. Cox, E. Merz, *J. Polym. Sci.*, **28**, 619 (1958).
106. P. Leblans, et al., *J. Polym. Sci.*, **21**, 1703 (1983).
107. W. Gleissele, in *Rheology*, ed. G. Astarita, et al., Plenum Press, New York, 457, Vol. 2, 1980.
108. D.W. Van Krevelin, *Properties of Polymers*, Elsevier Science, New York, 289, 1987.
109. C. Macosko, *Rheology*, Wiley-VCH, New York, 120–127, 1996.
110. L.-H. Lee (ed.), *Adhesive Bonding*, Plenum Press, New York, 1991.
111. L.-H. Lee (ed.), *Fundamentals of Adhesion*, Plenum Press, New York, 1991.
112. G. Martin, A. Tungare, J. Gotro, *Polymer Characterization*, American Chemical Society, Washington, DC, 1990.

113. M. Ryan, et al., *ANTEC Proc.*, **31**, 187 (1973).
114. C. Gramelt, *Am. Lab.* **16**, 26 (1984).
115. S. Etoh, et al., *SAMPE J.*, **3**, 6 (1985).
116. F. Hurwitz, *Polym. Compos.*, **4**(2), 89 (1983).
117. M. Roller, *Polym. Eng. Sci.*, **19**, 692 (1979).
118. M. Roller, et al., *J. Coat. Technol.*, **50**, 57 (1978).
119. M. Heise, et al., *Polym. Eng. Sci.*, **30**, 83 (1990).
120. K. O'Driscoll, et al., *J. Polym. Sci., Polym. Chem.*, **17**, 1891 (1979).
121. O. Okay, *Polymer*, **35**, 2613 (1994).
122. M. Hiese, G. Martin, J. Gotro, *Polym. Eng. Sci.*, **30**(2), 83 (1990).
123. K. Wissbrun, et al., *J. Coat. Technol.*, **48**, 42 (1976).
124. F. Champon, et al., *J. Rheol.*, **31**, 683 (1987).
125. H. Winter, *Polym. Eng. Sci.*, **27**, 1698 (1987).
126. C. Michon, G. Curvelier, Launay, *Rheol. Acta*, **32**, 94 (1993).
127. I. Kalnin, et al., *Epoxy Resins*, American Chemical Society, Washington, DC, 1970.
128. H. Barnes, J. Hutton, K. Walters, *An Introduction to Rheology*, Elsevier Science, New York, 1989.
129. A.W. Snow, et al., *J. Appl. Polym. Sci.*, **52**, 401 (1994).
130. T. Renault, et al., *NATAS Notes*, **25**, 44 (1994).
131. H.L. Xuan, et al., *J. Polym. Sci., Part A*, **31**, 769 (1993).
132. W. Shi, et al., *J. Appl. Polym. Sci.*, **51**, 1129 (1994).
133. J. Enns, Personal communication.
134. R. Follensbee, J. Koutsky, A. Christiansen, C. Myers, R. Geimer, *J. Appl. Polym. Sci.*, **47**, 1481 (1993).
135. R. Roberts, *SAMPE J.*, **5**, 28 (1987).
136. P.J. Halley, M.E. MacKay, *Polym. Eng. Sci.*, **36**(5), 593 (1996).
137. J. Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, John Wiley & Sons, New York, 1980.
138. J. Mijovic, B. Schafran, *J. Compos. Mater.*, **23**, 163 (1989).
139. J. Mijovic, T. Ho, *SAMPE J.*, **23**, 51 (1990).
140. M. Roller, *Met. Finish.*, **78**, 28 (1980).
141. M. Roller, *ANTEC Proc.*, **24**, 9 (1978).
142. J. Gilham, *ACS Symp. Ser.*, **78**, 53 (1978).
143. M. Roller, *ANTEC Proc.*, **21**, 212 (1975).
144. M. Roller, *Polym. Eng. Sci.*, **15**, 406 (1975).
145. M. Roller, *Polym. Eng. Sci.*, **26**, 432 (1986).
146. C. Rohn, *Problem Solving for Thermosetting Plastics*, Rheometrics, Austin, TX, 1989.
147. J. Seferis, et al., *Chemorheology of Thermosetting Polymers*, American Chemical Society, Washington, DC, 301, 1983.
148. R. Patel, et al., *J. Therm. Anal.*, **39**, 229 (1993).
149. M. Roller, *Polym. Eng. Sci.*, **26**, 432 (1986).
150. M. Roller, Personal communication, 1998.
151. J. Gillham, et al., *Polym. Compos.*, **1**, 97 (1980).
152. J. Enns, J. Gillham, *J. Appl. Polym. Sci.*, **28**, 2567 (1983).
153. L.C. Chan, et al., *J. Appl. Polym. Sci.*, **29**, 3307 (1984).
154. J. Gillham, *Polym. Eng. Sci.*, **26**, 1429 (1986).
155. S. Simon, G. McKenna, *J. Appl. Polym. Sci.*, **51**, 1741 (1994).
156. G. Palmese, et al., *J. Appl. Polym. Sci.*, **34**, 1925 (1987).
157. J. Enns, et al., in *Polymer Characterization*, ed. C. Craver, American Chemical Society, Washington, DC, 1983.
158. J. Gillham, J. Enns, *J. Appl. Polym. Sci.*, **53**, 709 (1994).
159. J. Enns, J. Gillham, *Trends Polym. Sci.*, **2**(12), 406 (1994).
160. A. Otero, et al., *Thermochim. Acta*, **203**, 379 (1992).
161. J. Gillham, J. Enns, *J. Appl. Polym. Sci.*, **42**, 2453 (1991).
162. B. Osinski, *Polymers*, **34**, 752 (1993).
163. J. Enns, J. Gillham, *Trends Polym. Sci.*, **2**(12), 406 (1994).
164. J. Sosa, K. Menard, *NATAS Proc.*, **25**, 176 (1995).
165. C. Daley, et al., *NATAS Notes*, **26**(2), 56 (1994).
166. J. Rosenblatt, et al., *J. Appl. Polym. Sci.*, **50**, 953 (1993).
167. E. McKaque, et al., *J. Test. Eval.*, **1**(6), 468 (1973).
168. I. Yakimets, et al., *Mech. Mater.*, **39**, 500 (2007).
169. A. Rouilly, et al., *Polym. Eng. Sci.*, **46**, 1635 (2006).
170. F. Bauer, et al., *J. Polym. Sci., Part B: Polym. Phys.*, **43**, 786 (2005).
171. S. Jeyapalina, et al., *J. Soc. Leather Technol. Chem.*, **91**, 102 (2007).
172. R. Broos, et al., *J. Cell. Plast.*, **36**, 207 (2000).

FURTHER READING

J. Carpenter, L. Liu, W. Chonkaew, K. Menard, *J. Therm. Anal. Calorim.*, in press.

G.M. Foster, S. Ritchie, C. Lowe, 'Controlled Temperature and Relative Humidity Dynamic Mechanical Analysis of Paint films', *J. Therm. Anal. Calorim.*, **73**, 119 (2003).

P. Gabbott, P. Clarke, T. Mann, P. Royall, S. Shergill, 'A High-Sensitivity, High-Speed DSC Technique: Measurement of Amorphous Lactose', *Am. Lab.*, **35**, 17 (2003).

P. Gupta, A.K. Bansa, 'Devitrification of Amorphous Celecoxib', *AAPS PharmSciTech.*, **6**(2), 214–225 (2005).

M. Saunders, K. Podluii, S. Shergill, G. Buckton, P. Royall, 'The Potential of High Speed DSC (Hyper-DSC) for the Detection and Quantification of Small Amounts of Amorphous Content in Predominantly Crystalline Samples', *Int. J. Pharm.*, **274**, 35 (2004).

C.M. Warner, *'Evaluation of the DSC for Observation of the Liquid – Liquid Transition'*, Master's Thesis, Central Michigan State University, (1988).

D.S. Katayama, J.F. Carpenter, K.P. Menard, M.C. Manning, T.W Randolph, 'Mixing Properties of Lyophilized Protein Systems: A Spectroscopic and Calorimetric Study', *J Pharm Sci.*, (2008) Jul 11. (E pub).

D.S. Katayama, J.F. Carpenter, M.C. Manning, T.W. Randolph, P. Setlow, K.P. Menard, 'Characterization of Amorphous Solids with Weak Glass Transitions Using High Ramp Rate Differential Scanning Calorimetry', *J Pharm Sci.*, **97**(2), 1013–1024 (2008) Feb.

Field Flow Fractionation in Analysis of Polymers and Rubbers

S. Kim Ratanathanawongs Williams
Colorado School of Mines, Golden, USA

Maria-Anna Benincasa
Universitá di Roma, Rome, Italy

1 Introduction **1**
2 Theory **2**
2.1 Basic Theory of Retention 3
2.2 Molecular Weight Dependence on Physicochemical Parameters 5
2.3 Conversion of a Fractogram to Molecular Weight Distribution 6
2.4 Calibration Methods 6
2.5 Zone Broadening 7
2.6 General Theory of Asymmetric Flow Field-flow Fractionation 8
2.7 Retention in Hollow-fiber Channels 9
3 Instrumentation **9**
3.1 Flow Field-flow Fractionation 9
3.2 Thermal Field-flow Fractionation 10
3.3 Detectors 10
4 Experimental Procedures **11**
4.1 Sample Preparation and Handling 11
4.2 Sample Injection and Relaxation 12
4.3 Operation with Constant or Programmed Field 13
5 Applications **13**
5.1 Organic-soluble Polymers 13
5.2 Water-soluble Polymers 16
6 Method Development **17**
6.1 Determining Experimental Conditions 17
6.2 Effect of Sample Size 18
6.3 Membranes in Flow Field-flow Fractionation 19
Acknowledgments **20**
List of Symbols **20**
Abbreviations and Acronyms **21**
Related Article **21**
References **21**

Field-flow fractionation (FFF) was conceived by J. Calvin Giddings in 1966 as a separation and characterization method for macromolecules, colloids, and particulates. Like chromatography, sample migration is caused by differential interaction with a field acting along an axis orthogonal to that of the transport liquid. Unlike chromatography, where separation is achieved by solutes partitioning between mobile and stationary phases, separation in FFF arises from the distribution of sample components in fluid laminae flowing at different velocities in a single phase. The different flow velocities, described by a parabolic profile, arise from the high aspect ratio of the FFF channel. Different types of fields can be used in FFF as long as they interact with some physicochemical property of the sample. The FFF channel design makes it highly suited for analyses of fragile aggregates, high-molecular-weight polymers, and gels. In comparison with packed columns, shear rates in the channel and the probability of plugging the channel are low. The ability of thermal FFF to differentiate polymers and latexes of different bulk and surface composition is unique among currently used separation techniques. The FFF family of techniques can provide a great deal of information about the sample but an initial time investment is often required for methods development.

In this article, the fundamental mechanism of FFF is shown at play in the separation and characterization of polymers and rubbers by the two techniques par excellence in this field: flow FFF and thermal FFF.

1 INTRODUCTION

FFF is a separation method conceived by J. Calvin Giddings in 1966.[1] It is particularly suitable for macromolecular, colloidal and diverse particulate materials extending from a few hundred[2,3] to 10^{18} Da.[4] It is an elution technique and is often referred to as a chromatography-like technique. In chromatography, the solute's rate of migration is determined by its partitioning between a mobile phase and a stationary phase in a column. In FFF, an externally applied field induces selective distribution of solutes in fluid laminae flowing at different velocities in a single phase inside a channel. The difference in the type of forces used in chromatography and in FFF defines the range of applicability of the techniques. While forces in chromatography are localized at interfaces and very selective, those used in FFF and electrophoresis are more diffuse and weaker. Consequently, the mass transport phenomena occurring in a chromatographic technique, such as high-performance liquid chromatography (HPLC), tend to be slower as the solute molecular weight increases. When the molecule's energy of interaction with the interface is significantly greater than the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

thermal energy kT, adsorption becomes irreversible.[5] However, even before irreversible adsorption appears, structural disruption and denaturation of the macromolecular component may occur because of the strong shear forces present in the irregular flow in the tightly packed chromatography column.[6] By contrast, the FFF separation is carried out in the absence of a stationary phase within an open channel. This channel is obtained by removing a geometrical portion from a Teflon, Mylar, or polyimide spacer and then clamping the spacer between two flat parallel plates. As shown in Figure 1, the removed section of the channel is a parallelepiped with tapered ends that facilitate the flow of liquid and sample in and out of the channel. The most commonly used channel dimensions are 27–87 cm in length L, 1–2 cm in breadth b and 0.0075–0.05 cm in thickness w. Because of the very high aspect ratio of the FFF channel and the frictional drag at the walls, the velocity of a liquid carrier moving in the longitudinal direction has a parabolic profile with a maximum in the center and minima, virtually zero, at the walls. In the normal mode of operation, the field applied perpendicular to the flow direction drives sample components toward one wall, referred to as accumulation wall,[7] with a velocity determined by the particle–field interaction. This field-induced displacement, optimized in the absence of longitudinal flow, is always counteracted by the diffusive flux that originates from the concentration gradient across the channel. The combination of these opposing effects results in a nonuniform distribution of components across the channel, those with a higher rate of back-diffusion being driven further away from the accumulation wall than those with a lower diffusion rate. Because of the parabolic flow velocity profile in the channel, the faster-diffusing component C shown in Figure 1 will be displaced along the channel more rapidly than component B, which has a lower diffusivity. The output signal, collected by a detector sensitive to some solute property, will thus register the elution profile of distinct peaks.

The normal mode of operation described above is the one mostly used for polymer separations because of the molecular size/weight range of these particular materials. Another retention mechanism, steric FFF,[8,9] comes into play when the component size is significant relatively to the channel thickness. This is generally the case for particle dimensions higher than 1 μm. The solute elution velocity is determined, in this case, by the extension of component particles into the flow. Larger particles extend into regions of faster streamlines because of steric exclusion from the accumulation wall and elute earlier than smaller particles. The retention order in this mode is reverse of that in the normal mode of operation.

Given the precise channel geometry and the flow profile that may be described mathematically, retention

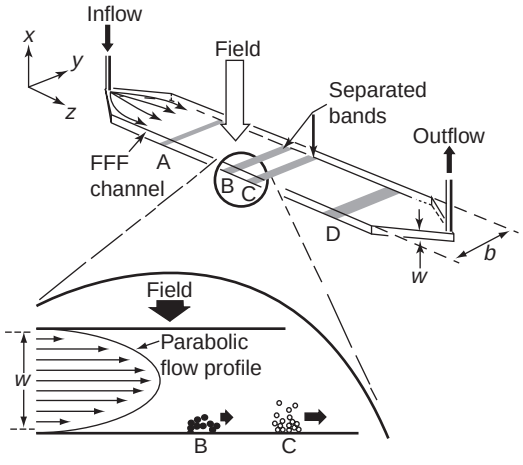

Figure 1 Schematic diagram of a typical FFF channel and the normal mode FFF separation mechanism. (Reproduced from L.F. Kesner, J.C. Giddings, *High Performance Liquid Chromatography*, eds. P.R. Brown, R.A. Hartwick, Chapter 15, 1989. Copyright © 1989, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

in FFF may be accurately calculated in theory and related to various solute physicochemical properties.[10,11] The particular sample property controlling retention depends on the type of the applied field. The use of different fields has generated a number of FFF techniques, such as sedimentation field-flow fractionation (SdFFF)[12] when a centrifugal force is used to induce retention, flow field-flow fractionation (FlFFF)[13] when the field is established by a transverse or crossflow of liquid, thermal field-flow fractionation (ThFFF)[14] when a thermal gradient is used, and electrical field-flow fractionation (ElFFF)[15] when a potential gradient is applied to electrically charged solutes.

FFF is particularly suitable for the separation of macromolecular samples and suspended colloidal particles of various origins because of the minimal surface area of the channel compared with the total surface area of a packed chromatographic column (10^7 cm^2).[16] For this reason, adsorption phenomena are greatly reduced in FFF. The driving force may be accurately adjusted to yield the desired levels of retention without the need to change the column as in chromatography and sample distribution between zones is very fast because no phase boundaries must be crossed. A rich literature is available for a great number of applications from colloids of environmental origin[17] to bacteria and viruses.[18]

2 THEORY

The theory of retention discussed in this section is developed for point particles at infinite dilution, that

is, for species with negligible size with respect to the channel dimensions, particularly its thickness. When this assumption is not satisfied corrections must be made to account for the particle size. One such correction is considered in the retention model for steric FFF.

2.1 Basic Theory of Retention

Let us consider the space between two parallel plates and a field applied orthogonal to them in the x-direction (measured across the channel thickness from the accumulation wall). Under the effect of the field alone, component particles are displaced with a velocity U. The particles' motions give rise to a field-induced flux Uc defined as the number of particles per unit cross-sectional area per unit time directed toward the accumulation wall, that is in the negative x-direction. The field-induced displacement is, however, counteracted by a diffusive flux proportional to the concentration gradient $\mathrm{d}c/\mathrm{d}x$ through the component's diffusion coefficient D and directed in the positive x-axis direction. The mass transport across the channel determines a net flux J_x given by Equation (1):

$$J_x = Uc - D\frac{\mathrm{d}c}{\mathrm{d}x} \tag{1}$$

The concentration profile across the channel thickness reaches a steady state when the field-induced and diffusive fluxes are at equilibrium. At this point the net flux is equal to zero. Equation (1) is integrated and solved, assuming constant U and D, to give an equation that describes the steady state concentration profile (Equation 2):

$$\frac{c(x)}{c_0} = \exp\left(-\frac{|U|}{D}x\right. \tag{2}$$

where c_0 is the concentration at the accumulation wall, x is the distance from the wall, and $|U|$ is negative for consistency with the coordinate system. Concentration decreases exponentially from the accumulation wall. Since it may be shown that an exponential distribution behaves as a much thinner layer positioned at a characteristic height ℓ, Equation (2) may be rewritten as Equation (3):

$$\frac{c(x)}{c_0} = \exp\ -\frac{x}{\ell} \tag{3}$$

where (Equation 4)

$$\ell = \frac{D}{|U|} \tag{4}$$

The mean layer thickness ℓ is a measure of the average distance of a sample component from the accumulation wall. It is apparent from Equation (4) that ℓ depends on the opposing effects of the field that acts to compress the solute layer and the diffusion process that broadens it. The quantity ℓ is frequently expressed in the dimensionless form shown in Equation (5):

$$\lambda = \frac{\ell}{w} = \frac{D}{|U|w} \tag{5}$$

where w is the channel thickness and λ is the retention parameter basic to FFF equations. Since λ may be shown to be dependent on the field strength and on a sample–field interaction constant, each component will have a characteristic λ value.

After the formation of the steady-state zone, a stream of liquid is allowed to flow along the longitudinal axis of the channel. Component particles are then carried downstream with an average velocity that depends on the region where they are found. The sample distribution at this point assumes a Gaussian distribution in the longitudinal direction because of the free diffusive motion of component molecules between regions of different velocities across the channel. The elution time will then be different for different components and may be used to measure the sample elution velocity once the channel void volume is known. Elution time, however, is not the most universal parameter defining the behavior of a species migrating along a chromatographic column or an FFF channel since it always depends on the average fluid velocity. The dimensionless retention ratio R, defined as the ratio of component migration velocity v_p to the mean fluid velocity $\langle v\rangle$ (Equation 6),

$$R = \frac{v_\mathrm{p}}{\langle v\rangle} \tag{6}$$

is a more useful parameter. R is widely used in separation techniques as a measure of the retardation of the solute relative to the liquid carrier flow velocity. It is a more universal measure of retention than elution time and only depends on the field strength and on the particle property responding to the field, regardless of the flow velocity. It then frees the mathematical architecture from a variable. In the expression of R, the solute mean migration velocity is the average of the particle velocities expressed as $v_\mathrm{p} = \langle cv\rangle/\langle c\rangle$.[19] Inserting this expression into Equation (6), substituting the concentration term with Equation (3) and the parabolic function to calculate v, and using Equation (5), one obtains an expression for R whose solution is given by Equation (7):

$$R = 6\lambda\left(\coth\frac{1}{2\lambda} - 2\lambda\right. \tag{7}$$

Equation (7) is the basic retention equation in FFF. It shows that retention is solely dependent on λ, which is characteristic of each eluting species. Approximate forms of Equation (7) may be used for low λ values or high levels of retention. In particular, the approximation $R = 6\lambda$ has

an error of 20% at $R = 0.25$. The empirical relationship $R = 6(\lambda - 2\lambda^2)$ has the greater range of applicability with less than 10% error up to $R = 0.7$. The retention ratio may also be written as $R = t^0/t_r$ since $V = L/t_r$ and $\langle v \rangle = L/t^0$, where t^0 is the residence time of the eluent or of a nonretained component and t_r the average sample residence time or retention time.

In theory, any external field to which sample components are responsive may be used to induce selective retention and fractionation by FFF. The applied field determines the type of interaction and hence the sample property that will be measured. Besides FlFFF, ThFFF, SdFFF and ElFFF, the theoretical foundation has also been laid down for many less developed FFF techniques. One such technique uses a concentration gradient in a solvent mixture to establish a chemical potential gradient capable of driving solutes towards regions of lowest potential [concentration field-flow fractionation (CFFF)].[20] A nonuniform electric field that induces charge polarization may exert selective dielectrophoretic forces on component particles and generate fractionation.[21] Interesting applications are reported using a magnetic field.[22] Photophoretic FFF[23] is based on the transfer of momentum from a photon to a particle. An acoustic wave field[24] may induce retention selective to particle diameter in addition to other parameters.

The present discussion focuses on the FFF techniques that have been applied to polymer separations: FlFFF and ThFFF. These techniques have been extensively used for a great number of different samples. They are the FFF techniques par excellence for polymer fractionation. ElFFF has shown its potential only with a new channel design[25] and may prove to be very useful for the characterization of charged polymers.

In his investigations on the theory of Brownian motion, Einstein[26] showed that in a system of point particles at infinite dilution, i.e. in the absence of flow perturbations due to the motion of one particle affecting another, the diffusion coefficient is inversely proportional to the friction coefficient (Equation 8):

$$D = \frac{kT}{f} \tag{8}$$

If D is replaced with the Einstein equation and the relationship for the field-induced velocity specific for FlFFF ($U = \dot{V}_c/Lb$) is used in Equation (5) one obtains

$$\lambda = \frac{kT}{f}\frac{V^0}{\dot{V}_c w^2} \tag{9}$$

In Equation (9), the relationship $V^0 = Lbw$ is also used to replace Lb with V^0/w. It is worth noting that Equation (9) depends on the friction coefficient f as well as on the instrumental parameters V^0 and w. This is not the case with SdFFF where both D and U depend inversely on f, which then cancels out in the ratio $D/|U|$. The dependence of the retention parameters λ and R on the friction coefficient in FlFFF is the key to obtaining solute physicochemical parameters from FFF measurements. More than a century before Einstein's work in this area, Stokes[27,28] found a quantitative relationship between the shape and the dimensions of a moving particle and its friction coefficient. For a spherical body of radius R_0 moving in a fluid of viscosity η he showed that the friction coefficient could be expressed by Equation (10):

$$f_0 = 6\pi\eta R_0 \tag{10}$$

Stokes' equation may be used to relate diffusion coefficient to particle dimensions. Correction factors that extend Stokes' relation to nonspherical particles were first introduced by Perrin[29] and Herzog et al.[30] for solids of revolution such as prolate (cigar-shaped) and oblate (disk-shaped) ellipsoids. They chose to express the departure of the actual friction coefficient f, for an ellipsoid with axes of symmetry of different length from the friction coefficient of a sphere of same volume f_0. They found a quantitative relation between f/f_0 and the ratio a/b of the semimajor axis a to the semiminor axis b. By combining Equations (8) and (10), Equation (11):

$$D_0 = \frac{kT}{6\pi\eta R_0} \tag{11}$$

is obtained. The Stokes–Einstein relation in Equation (11) is derived considering equivalent particle orientation which applies rigorously only to spherical bodies. The substitution of Equation (10) into Equation (9) gives the expression for λ in FlFFF (Equation 12):

$$\lambda = \frac{kT}{6\pi\eta R_0}\frac{V^0}{\dot{V}_c w^2} \tag{12}$$

which shows explicitly the effective size-based separation in normal-mode FlFFF where samples are retained proportionally to their hydrodynamic dimensions.

The expression for λ in ThFFF is derived using a conceptual procedure similar to that followed for FlFFF. The derivation of this expression is complicated, however, by the distortion of the parabolic flow profile due to changes of viscosity with temperature across the channel.[31,32] In 1856, Fick (and later Soret in 1859) showed that if a salt solution of uniform concentration in a tall container is heated at the top and cooled at the bottom, a flux of matter originates that increases the salt concentration at the cold end of the column. This effect, named after Soret, is known as thermal diffusion since it is clearly a diffusive effect occurring only in the presence of a thermal gradient. The equation of flux (Equation 1)

set for molar thermal and diffusive fluxes gives

$$J_x = -D\left(\frac{\mathrm{d}c}{\mathrm{d}x} + c\gamma\frac{\mathrm{d}T}{\mathrm{d}x}\right. \quad \left. - D_{\mathrm{T}}c\frac{\mathrm{d}T}{\mathrm{d}x}\right. \tag{13}$$

where D_{T} is the thermal diffusion coefficient and γ is the coefficient of thermal expansion. The solution for zero net flux at the steady state yields the retention parameter for ThFFF (Equation 14):

$$\lambda = \left. w \quad \frac{\alpha}{T} + \gamma \quad \frac{\mathrm{d}T}{\mathrm{d}x}\right.^{-1} \tag{14}$$

where α is the thermal diffusion factor equal to $D_{\mathrm{T}}T/D$.[33] Both α and $\mathrm{d}T/\mathrm{d}x$ depend on the local position x and thus on the temperature. The ordinary diffusion coefficient is strongly dependent on the temperature through the T/η term in Equation (11) where $1/\eta$ also increases with T. In the case of flexible-chain macromolecules, it must be borne in mind that, on polymer chemistry theoretical grounds, the molecular coil is expected to expand on increasing temperature as a consequence of the increased excluded volume. This property is treated in terms of the second virial coefficient.[34] The tendency of polymer hydrodynamic dimensions to increase with temperature is reported in light-scattering studies.[35] The local temperature gradient is not independent of T[31,32] but the dependence is relatively small. D_{T} is shown to depend on temperature[36] but to a smaller extent than the ordinary diffusion coefficient. In the standard theory of FFF, however, λ is generally used in an approximate form that is valid at high retention where the sample cloud is very close to the accumulation wall. The local values of $\mathrm{d}T/\mathrm{d}x$ and α may therefore be assumed to be the same as those at the cold wall.[37] The term γ becomes negligible[31] and λ may be rewritten as in Equation (15):

$$\lambda \approx \frac{T_{\mathrm{c}}}{\alpha w\left(\dfrac{\mathrm{d}T}{\mathrm{d}x}\right._{\mathrm{c}}} = \frac{D}{D_{\mathrm{T}}w\left(\dfrac{\mathrm{d}T}{\mathrm{d}x}\right._{\mathrm{c}}} \tag{15}$$

where $(\mathrm{d}T/\mathrm{d}x)_{\mathrm{c}}$ is the temperature gradient at the cold wall.

2.2 Molecular Weight Dependence on Physicochemical Parameters

Synthetic polymers are often flexible-chain molecules whose dimensions cannot be defined precisely and must be considered as average values of all the configurations that the molecules assume. In such a situation it is convenient to introduce a new parameter called the radius of gyration, R_{g}. It is found from polymer theory that molecular weight may be related to polymer molecular dimensions through this statistical parameter according to Equation (16):

$$R_{\mathrm{g}}^2 \propto \frac{M}{M_0} \tag{16}$$

where M is the molecular weight and M_0 is the molecular weight of the repeat unit (constant for synthetic homopolymers). The equivalent radius R_{e}, defined as the radius of an equivalent sphere having the same value of the friction and diffusion coefficients as the polymer, is a convenient parameter that is often used. Since it may be demonstrated that $R_{\mathrm{g}} \propto R_{\mathrm{e}}$,[34,38] equations dependent on R_{g} may be written in terms of the equivalent radius. For approximately spherical molecules, the volume of the molecule is linearly related to molecular weight. This may be expressed as Equation (17):

$$R_0 \propto M^{1/3} \tag{17}$$

where R_0 is the radius of a spherical particle. Globular proteins are an example of a real system that satisfies the model of a rigid spherical body. Even in such a simple case however, correction factors must be applied to Equation (17) to account for the hydration volume and deviation from the perfectly spherical symmetry. When the polymer molecule is a statistical chain that is allowed to meander randomly in a Brownian-like way with no forbidden physical volume, polymer theory shows that $\langle R_{\mathrm{g}}^2\rangle$ is proportional to the number of repeat units in the macromolecule chain given by $N = M/M_0$. It follows that the equivalent radius (or Stokes radius) is given by Equation (18):

$$R_{\mathrm{e}} \propto M^{1/2} \tag{18}$$

If the polymer chain is not allowed to self-cross, there is some space occupied by other polymer segments from which each segment is excluded. Polymer conformations are significantly affected by the excluded volume and the equivalent radius for a three-dimensional polymer chain model becomes (Equation 19):

$$R_{\mathrm{e}} \propto (M^{6/5})^{1/2} \tag{19}$$

Based on the model of the equivalent sphere (Stokes–Einstein relationship), Equation (11) may be applied to polymers in solution, and the diffusion coefficient, at a given temperature, may be related to molecular weight through the polymer equivalent dimensions R_{e} and the friction coefficient.[38,39] Substituting Equation (18) or (19) into Equation (11) and grouping together all the constants, we obtain Equation (20):

$$D = AM^{-b} \tag{20}$$

where A is constant for a given polymer type and b is an exponent that depends on the polymer conformation in

solution. The value of the exponent in Equation (18) is obtained on theoretical grounds, considering zero attractive or repulsive interactions between different polymer segments as well as between polymer molecules and the surrounding medium. It also assumes infinitely dilute systems of neutral, nonfree draining molecules with average spherical symmetry. A system under these conditions is defined as a Θ-system (theta-system) at Θ-temperature. Θ-conditions are set by the type of solvent (which determines the energy of interaction with the polymer molecules) and by the temperature. Generally, only one solvent behaves as a Θ-solvent for a given polymer type. Therefore, the theoretical value of the b exponent is always related to a specific polymer–solvent system at Θ-temperature. The excluded volume has a considerable swelling effect on flexible-chain molecules and on the dependence of particle size on molecular weight. The excluded volume effect is similar to and may also be thought of as that due to a good solvent or to a temperature higher than the Θ-temperature. By relating the diffusion coefficient and molecular weight at a given temperature, Equation (20) allows FFF-measured parameters to be correlated with the polymer molecular weight. Combining Equations (12) and (20), it appears that λ in FlFFF is inversely dependent on molecular weight, that is, the zone mean migration layer decreases as the polymer molecular weight increases (Equation 21):

$$\lambda = AM^{-b}\frac{V^0}{\dot{V}_c w^2} \tag{21}$$

A similar expression for ThFFF is complicated by the D_T term in Equation (15). In the classical theoretical treatment of FFF, D_T is considered independent of molecular weight,[37] as predicted and experimentally observed by some authors[40] (see section 5.1). Other authors,[41,42] however, report a molecular weight dependence of the thermal diffusion coefficient. Assuming that, at a given temperature, we have (Equation 22):

$$D_T = BM^{\beta} \tag{22}$$

and considering Equation (20) for the dependence of the ordinary diffusion coefficient on molecular weight, the overall contribution of molecular weight to λ in ThFFF is given by the D/D_T term as Equation (23):

$$\frac{D}{D_T} = \frac{A}{B}M^{-(b+\beta)} = \Phi M^{-n} \tag{23}$$

where $\Phi = A/B$ and $n = b + \beta$. The parameter Φ must be temperature dependent because of the dependence of D on temperature. Equation (23) substituted into Equation (15) yields a negative exponential correlation between λ and the molecular weight (Equation 24):

$$\lambda = \Phi M^{-n}\frac{1}{w\left(\dfrac{dT}{dx}\right)_c} \tag{24}$$

2.3 Conversion of a Fractogram to Molecular Weight Distribution

To obtain a meaningful molecular weight distribution of a polymer sample, it is necessary that elution occurs in the same mode (normal, steric, etc.) over the entire retention time interval with a monotonic function of some sample property. Transformation of a fractogram from a time-based function to a molecular weight function is readily accomplished but requires a correction to the amplitude to account for the nonlinear relationship between M and t_r. The mass abundance in a time interval δt_r, corresponding to $c(t_r)\dot{V}\delta t_r$, must equal the mass comprised in the corresponding molecular weight interval δM. Integrating the mass distribution $m(M)$ over that interval yields Equation (25):

$$m(M)\,\delta M = c(t_r)\dot{V}\delta t_r \tag{25}$$

where $c(t_r)$ is the detected sample concentration at time t_r and $\dot{V}$ is the volumetric channel flow rate. For $\delta t_r \to 0$, Equation (25) becomes Equation (26):

$$m(M) = c(t_r)\dot{V}\frac{dt_r}{dM} \tag{26}$$

The scale correction function dt_r/dM allows the conversion of the retention timescale to the molecular weight scale. For a normal-mode elution, the molecular weight distribution may be obtained in theory from first principles.[43]

2.4 Calibration Methods

It may be shown that the scale correction function depends on the constants A and b for FlFFF and Φ and n for ThFFF, through the dependence of λ on molecular weight. These constants may be found from theory and are available in literature for a wide number of polymer–solvent systems. However, it is common laboratory practice to obtain them through a calibration procedure with a set of well-characterized, narrowly distributed polymer standards whose molecular weights have been measured by some absolute technique such as light scattering, viscometry or osmometry. To obtain A and b values that can be transferred to an unknown polymer sample, calibration must be performed with standards of the same or similar composition as the unknown, in the same carrier liquid and at the same temperature. Constants A and b for the FlFFF analysis of

polymers are obtained from the intercept and slope of a plot of the measured $\log D$ versus $\log M$ (Equation 27):

$$\log D = \log A - b \log M \tag{27}$$

derived from Equation (20). The plot of $\log \lambda$ versus $\log M$ would also give these constants. A similar $\log \lambda \Delta T$ versus $\log M$ plot provides the constants Φ and n for the calibration in ThFFF. Gao and Chen[44] and Giddings[45] have shown the system transferability of the ThFFF calibration constants. A more detailed treatment of the "universal calibration" in ThFFF[46] also accounts for the change of the cold wall temperature on polymer retention and on the calibration constants. FFF constants allow a more universal calibration than size-exclusion chromatography (SEC) since they are derived from fundamental physical properties of the carrier liquid and of the polymer under investigation and do not depend on any system parameter. In contrast, calibration in SEC has no rigorous theoretical grounds; it is not valid for some pore structures[47] and it is therefore not system transferable. Moreover, it is still under evaluation for some types of polymers[48–50] and does not always apply to different polymer samples[51,52] even when run in the same column under the same conditions.[16] Calibration in FFF may also be performed using a set of polydisperse standards[53,54] or with a single broadly disperse sample when the relative molar mass is measured at two or more retention times by an absolute technique such as light scattering[55] or mass spectrometry.[56] It should be noted that absolute molecular weight (M) measurements made by FlFFF combined with mass spectrometry show an excellent data fit for $\log \lambda$ (and thus D) versus $\log M$ at molecular weight values far lower than those predicted from polymer theory.

2.5 Zone Broadening

The previous discussion on calibration in FFF assumed that broadening of the peak was due solely to molecular mass differences. In contrast, a number of concomitant effects contribute to the increase in peak width. A measure of zone spreading, and hence of resolution, is given by the plate height. As with other separative techniques, the overall plate height in FFF is given by the sum of contributing uncorrelated terms (Equation 28):

$$H = H_l + H_n + H_r + H_p + \sum H_i \tag{28}$$

In Equation (28), the zone spreading due to longitudinal diffusion H_l arises from concentration gradients along the sample plug and increases with the residence time. It is linearly related to the diffusion coefficient as $H_l = 2D/R\langle v \rangle$ and therefore becomes negligible for very slowly diffusing components such as macromolecules. In contrast, the nonequilibrium term H_n arising from the random displacement of component molecules across the channel is one of the dominant contributions to peak broadening in FFF. The band spreading associated with the time needed for the zone to reach the equilibrium position in the presence of the longitudinal flow, H_r, may be minimized by adopting a stop-flow procedure (see later). Other contributions, H_i, associated with system nonidealities such as dead volumes or injection volumes, may be disregarded in a well-designed channel. The spreading due to differences in a characteristic property such as molecular weight H_p is only an apparent broadening. It arises from the fact that individual molecules of a macromolecular or a particulate sample may differ somewhat from one another in their relative mass or size and are retained to a slightly different extent. If the difference in M is not large enough to produce distinct peaks, the sample zone will appear as a broadened peak with different mass elements continuously distributed in a molecular weight interval. It may be shown that (Equation 29):[57]

$$H_p = L S_M^2 (\mu - 1) \tag{29}$$

where $\mu = M_W/M_N$ is the sample polydispersity index which gives a measure of the departure of the weight-average molecular weight M_W from the number-average molecular weight M_N. For monodisperse samples, the two averages coincide. It appears from Equation (29) that the plate height, and hence resolution, strongly depend on the system selectivity S_M, which is defined as the retention volume difference of components of different M relative to the molecular weight difference (Equation 30):[58]

$$S_M = \frac{d \ln V_r}{d \ln M} = \frac{d \ln R}{d \ln M} \tag{30}$$

The subscript M denotes a molar mass-based selectivity whose value is a measure of the system ability to discriminate samples by their mass. Under conditions of high retention, the approximation $R = 6\lambda$ may be used and S_M becomes in the limiting form (Equation 31):

$$S_M = \frac{d \ln \lambda}{d \ln M} \tag{31}$$

Application of Equation (31) to Equations (21) and (24) shows that selectivity values for FlFFF and ThFFF coincide with the exponent of molecular weight, i.e. b in FlFFF and n in ThFFF. Assuming that the thermal diffusion is independent of molecular weight, b and n may be associated with the polymer molecule conformation in a given solvent and temperature system. From polymer theory, values of exponent b are predicted to be 0.33 for solid spheres, 0.5 for random-coil macromolecules[39] in Θ-conditions and about 0.7 for

flexible-chain polyelectrolytes.[59] The theoretical selectivity for FFF is considerably higher (0.5–0.7)[60] than that for SEC, where typical values range between 0.05 and 0.1.[58,61] Differences in selectivity when polymers are analyzed in organic solvents or aqueous solutions are expected considering that water generally behaves as a good solvent for hydrophilic polymers whereas Θ-systems are mostly reported for neutral polymers in organic solvents. The theoretical value of 0.588[38,62] for polystyrene (PS) in tetrahydrofuran is obtained in ThFFF at a specific cold wall temperature.[46] The same polymer, run in ethylbenzene by FlFFF,[63] gives a selectivity ranging between 0.51 and 0.56. Synthetic water-soluble polymers are generally fractionated with a selectivity above 0.6.[64–69] Polyvinylpyridine is an interesting example of a polymer that behaves as a neutral statistical chain in tetrahydrofuran and as a charged coil in water where it is soluble only as a polyelectrolyte. This polymer, bearing a six-term aromatic ring in the repeat unit, is fractionated with a selectivity of 0.51 in tetrahydrofuran by ThFFF and of 0.62 in an aqueous medium at low pH.[65] Systematic studies of polyelectrolytes in aqueous solution by FlFFF indicate a strong effect of the solution ionic strength on the selectivity, which generally decreases with increasing concentration of the added simple electrolyte.[68] Further investigations on the change in selectivity with the ionic strength of aqueous solutions and with the type of added electrolyte have been used to show specific interactions between the polymer and some metal ions.[69] Retention and selectivity in ThFFF of neutral polymers in organic solvents appear to increase with rising solubility parameters of the polymer–solvent system.[70] This effect was first registered in early investigations on ThFFF of polymers.[37] Comparative studies on the separation of copolymers by ThFFF and gel permeation chromatography (GPC) show that only the former yields a good separation of a diblock copolymer of poly(styrene-*co*-isoprene) from a triblock poly(styrene-*co*-isoprene-*co*-styrene).[71] This ThFFF separation is attributed to the difference in thermal diffusion coefficient of the two polymers.

2.6 General Theory of Asymmetric Flow Field-flow Fractionation

The general concepts of FFF were first developed for uniform field strengths and constant flow velocities along the channel. The symmetric FlFFF channel, shown in Figure 2(a), was designed to achieve these characteristics. In 1987, Wahlund and Giddings[72] introduced a new design of a FlFFF channel with only one permeable wall and no independent transverse flow. In this variant of FlFFF named "asymmetric FlFFF", shown in Figure 2(b), the permeable upper wall is replaced by a solid glass plate that conveniently allows visual inspection of the interior of the channel. Both the channel and crossflow originate at the channel inlet but exit the channel as two separate streams at the channel and crossflow outlets. Using this asymmetric channel configuration, a crossflow velocity gradient is established across the channel thickness. The sample concentration in the transverse coordinate in this design may no longer be described by Equation (3), but it approaches the exponential distribution of the standard FFF model for highly retained samples. Expressions for R and λ are found following the same mathematical procedure as in the symmetric configuration but taking into account that the flow velocity profile is linearly decreasing along the channel at a rate determined by the crossflow velocity. For high retention levels ($x/w \approx 0.1$), λ may again be approximated by Equation (9). Rectangular and trapezoidal[73] shaped channels have been used in asymmetric FlFFF. In the latter case, the breadth decreases as a function of channel length. An exponential

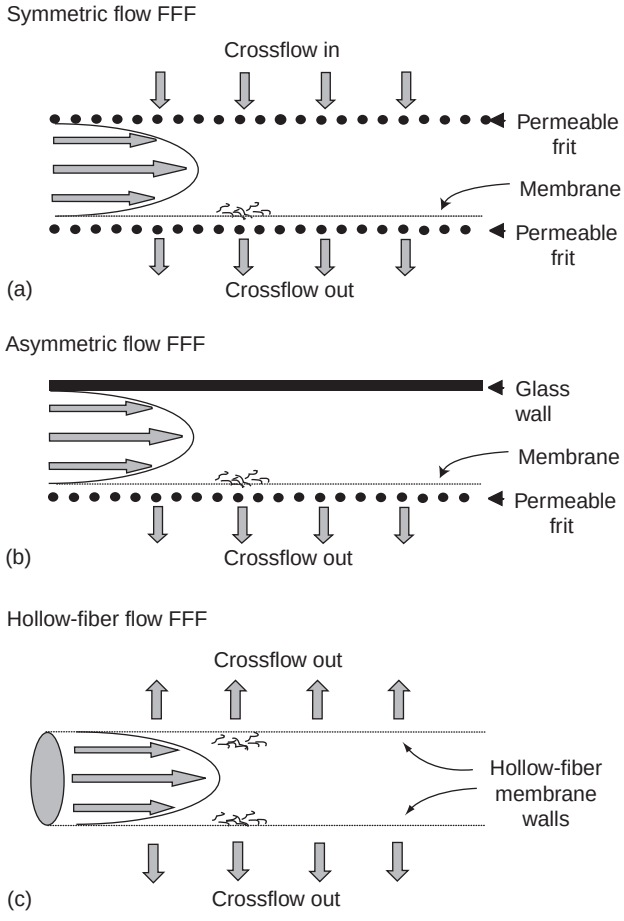

Figure 2 Different configurations of FlFFF channels. The symmetrical (a) and asymmetric (b) channels are commercially available and thus more commonly used than the hollow-fiber configuration shown in (c).

decrement in channel breadth[74] is shown to give a constant average channel flow velocity.

2.7 Retention in Hollow-fiber Channels

An interesting but commercially unavailable configuration of FlFFF utilizes a hollow cylindrical fiber with porous walls in place of the rectangular channel.[75] In this arrangement, illustrated in Figure 2(c), the axial flow is provided by an external pump connected with the inlet of the hollow-fiber membrane while the transverse flow is obtained by removing liquid radially from the channel with a second pump. The two flow velocity profiles in this configuration are different from any of the previously described FlFFF systems, both being variable with the distance Z from the inlet because of the pressure drop along the channel. From studies of fluid dynamics in hollow fibers,[76] it is found that the axial and radial flow velocities decrease exponentially as a function of Z and the permeability of the fiber and the fluid viscosity. Retention time in this type of channel is a function of channel length and Péclet number.

3 INSTRUMENTATION

An FFF system is generally assembled in a manner similar to that of a chromatographic apparatus. It uses most of the ancillary equipment employed in chromatography such as injector valves, pumps for the carrier liquid delivery, detectors, and some data acquisition devices such as chart recorders or more conveniently computers. A generalized FFF system is shown in Figure 3.

3.1 Flow Field-flow Fractionation

FlFFF is the most universally applicable FFF technique because any solute particle is subject to transport in a liquid stream. As shown in Figure 4, the spacer containing the channel form is clamped between two parallel blocks of some material such as Plexiglas, polyethylene, anodized aluminum, or stainless steel that accommodate two porous frit panels with 2–5-μm pores. Ceramic frits are more commonly used while polyethylene,[64,77] polypropylene,[18,78] and stainless steel[63,64,79] are generally employed with clamping blocks of the same material. An important component of the FlFFF channel is a permeable, generally polymeric, membrane placed over the accumulation wall to impede sample loss through the frit. Given the importance of the membrane in the successful application of FlFFF, a specific section (section 6.3) is dedicated to this topic. The design of the asymmetric FlFFF channel is somewhat different from that of the symmetric channel, as shown in Figure 2(b), but besides the dissimilarity of the channel geometry and the absence of the top porous wall it bears few differences from the symmetric FlFFF system. Some operational procedures, such as sample injection and relaxation, however, differ from those routinely used in the symmetric channel. A hollow-fiber channel has to be placed in a mantle, possibly of cylindrical symmetry, such as an empty stainless-steel chromatographic column, and the two coaxial tubes sealed to each other at the ends. The mantle must have a port connected to a crossflow pump that draws fluid out through the wall of the fiber. The channel flow is supplied by another pump connected to one of the hollow-fiber extremities. Sample injection and relaxation also in this case are particular to the set-up and will be dealt with in section 4.2.

Standard FFF theory assumes constant and uniform field strength and average flow velocity in the length and breadth dimension (edge effects not considered). Therefore, accurate control and continuous measurements of both these parameters are necessary when operating an FFF system. In FlFFF, the pump flow rates for both the longitudinal and transverse streams must be kept constant and continuously checked because, after mixing inside the channel, both flows may exit the channel through the outlet with the lower pressure. It is therefore of primary importance to equalize the pressures at the channel and crossflow outlets. This is achieved by placing back-pressure regulators at one or both outlets

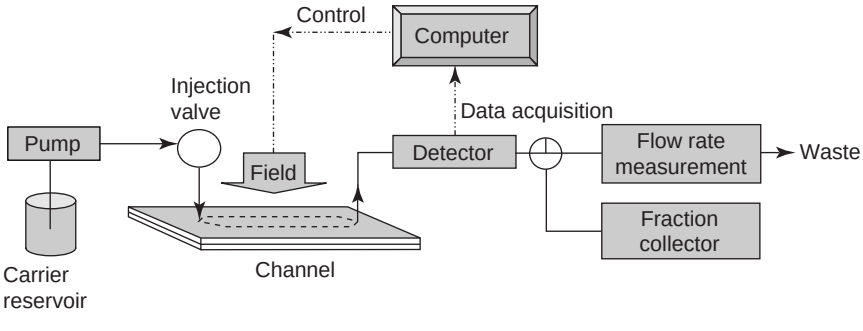

Figure 3 FFF system assembled with the ancillary equipment.

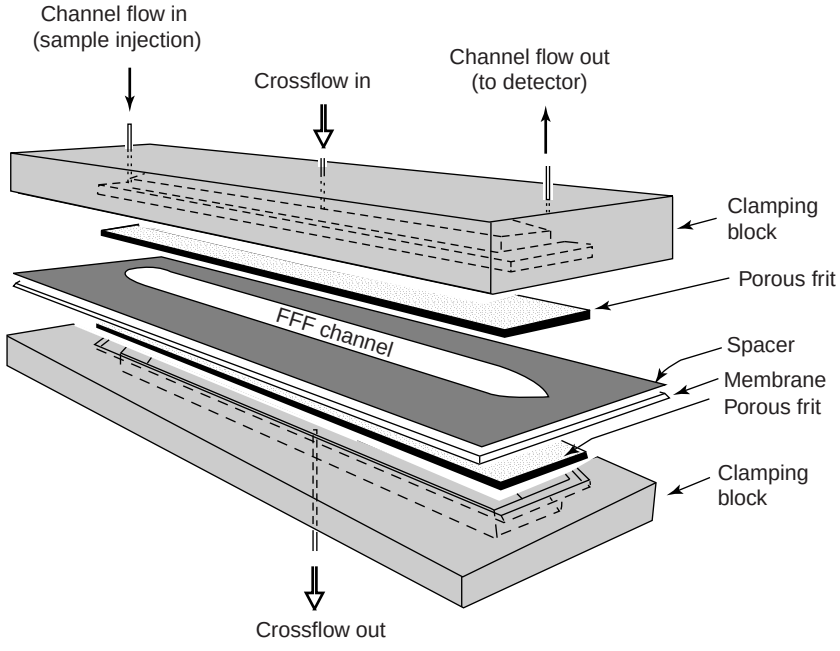

Figure 4 Diagram of the constituent elements of a FlFFF channel. (Reprinted from M.H. Moon, J.C. Giddings, *J. Pharm. Biomed. Anal.*, **11**, 911–920 (1993). Copyright 1993, with permission from Elsevier Science.)

and checking the pressure with in-line pressure gauges. An easy, cheap, but time-consuming method to check the flow rates is to measure the liquid volumes displaced over unit time with a buret and a stopwatch. Alternatively, a balance can be used to measure the weight of the carrier liquid exiting from each outlet as a function of time. The balance may be connected to a computer[80] to store data on time-dependent flow rates that may be used for accurate calculations of the retention parameters. This is particularly useful in programmed runs (see later) where the field strength, and hence the crossflow rate, is varied with some function of time.

3.2 Thermal Field-flow Fractionation

ThFFF channels (Figure 5) are formed by clamping a polyimide spacer (with the FFF channel volume removed) between two copper blocks with highly polished nickel or chromium surfaces. The field in this system is a thermal gradient that is provided by electrically heated metal elements inserted into one block and a stream of cold water flowing through the second block. Holes are drilled into the copper blocks to allow the insertion of temperature-measuring probes at different positions along the channel.

3.3 Detectors

Virtually any of the detectors used in chromatography is compatible with FFF apparatus. The HPLC separation

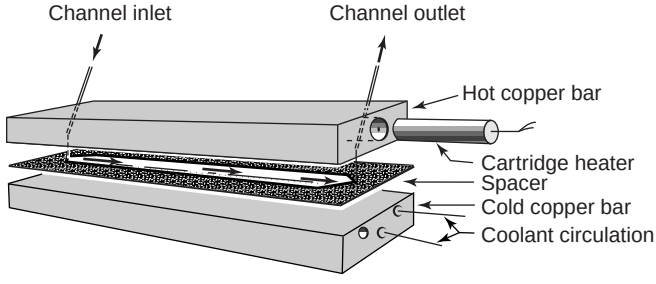

Figure 5 ThFFF channel. (Reprinted with permission from J.C. Giddings, V. Kumar, P.S. Williams, M.N. Myers, *Adv. Chem. Ser.*, **227**, 3–21 (1990). Copyright 1990, American Chemical Society.)

mechanism generally induces a concentration of the initial sample plug so that samples of continuously decreasing concentration may be detected. In FFF, the solute plug undergoes considerable concentration during the relaxation process[81] and considerable dilution during separation and elution. While reduction of the injected sample mass is always sought, particularly for high-molecular-weight polymer samples, the final concentration must be considered and the injected load adjusted to give a good signal-to-noise ratio. Nonspecific detection methods, such as refractive index[61,71,82,83] or viscosity,[84] may be generally used while others may be employed only when the solute is susceptible to specific response. This is the case for spectrophotometric detection, one of the most widely used methods in chromatography and FFF. Spectrophotometric detection

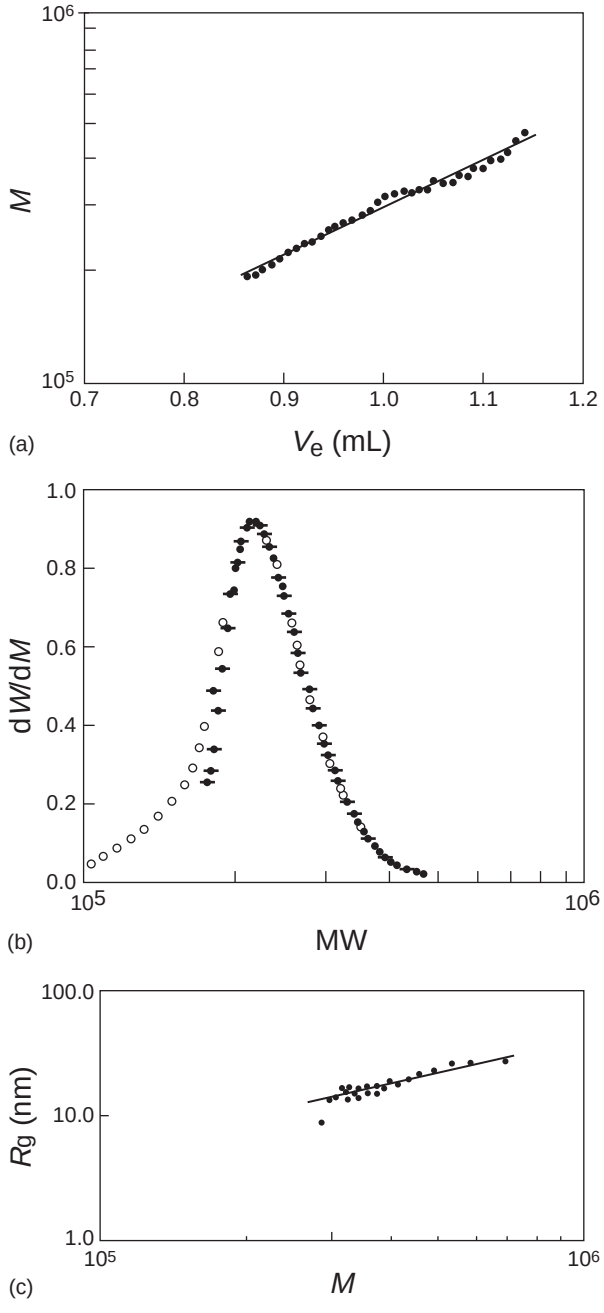

Figure 6 ThFFF/MALS of PS microgel. (a) Increase in molecular weight with increasing elution volume V_e; (b) molecular weight distribution determined from MALS (dashes) and calculated by calibration (open circles); (c) radius of gyration R_g versus molecular weight. (Reproduced by permission of Wiley-VCH from Antonietti et al.[90])

depends on the absorption of radiant energy of specific wavelengths by the chromophores. However, when the solute particle size becomes comparable to the detector wavelength, the output signal is considerably affected by light scattering. Consequently, the measured signal must be corrected in order to obtain an accurate mass distribution.[85] The Mie theory[86] used in these corrections takes into account the dependence of the scattered intensity on the particle size, the refractive index, and the scattering angles. Fluorescence detection is most often used with derivatized samples.[87] A recent development in polymer characterization is the coupling of FFF with a multiangle light scattering (MALS) detector.[88–91] This hyphenated system combines the high resolution capabilities of FFF with the absolute and independent molecular weight determination of light scattering, thus eliminating the need to calibrate the FFF system. The absolute determination of molar masses by MALS also allows for small nonidealities in the operating conditions such as fluctuations in temperature and flow rate. The best results are obviously obtained when optimum separation conditions are used and in the absence of nonidealities. The high fractionating capability of FFF is registered as an increase in the relative molar mass along the eluted peak as shown in Figure 6(a–c). The integrated FFF MALS apparatus has also allowed an in-depth study of the effect of experimental conditions such as injected sample mass or crossflow rate on the elution and fractionation of polymer samples.[92] A detailed discussion on these topics may be found in sections 6.1 and 6.2. The generally low sensitivity of laser light scattering detection is a problem, particularly for low-molecular-weight components and polyelectrolytes whose molar mass determinations are meaningful only at high ionic strength.[93] This problem has been overcome by coupling an electrospray mass spectrometer to a FlFFF channel[56] as mentioned in section 2.4.

4 EXPERIMENTAL PROCEDURES

4.1 Sample Preparation and Handling

No special treatment is generally required for the preparation of samples to be analyzed by FFF. Specific procedures, such as extraction, purification, and concentration, may be necessary for polymers of natural origin since synthetic polymers undergo sample purification as part of the production process. As mentioned previously, the narrow sample pulse injected in an FFF channel undergoes considerable changes in concentration as it is relaxed and fractionated[81] (see later). The concentration of the injected polymer sample solution is therefore a parameter that must be carefully controlled. Injection of large sample masses affect the plate height[94] even before the effect of overloading is registered by other retention parameters. The sample concentration during relaxation depends on the mass injected and can be one or more orders of magnitude higher than that of the initial sample solution depending on the retention ratio (concentration at the wall is approximately

equal to the sample concentration divided by λ which is expected to be ≤ 0.1). The effect of molecular dimensions on the concentration of polymer solutions is known from theory. Four model polymer solutions are generally identified: the dilute solution corresponding to a concentration of molecules separated by large volumes of solvent, the intermediate regime with a much reduced distance between polymer chains that may touch each other but still do not overlap, the semidilute, and the concentrated regime. In the last two cases, the polymer concentration is so high that chain entanglement dominates and the system may not be regarded as that of individual molecules suspended in a liquid medium. The solution properties are not governed by the properties of individual molecules and such systems will therefore have a behavior very different from that of any concentrated solution of small molecules. For this reason, the concentration c^* corresponding to that of the intermediate regime has been the subject of several theoretical and experimental investigations. In a number of studies, the dependence of c^* on molecular weight was found to follow a power law of the type $c^* \propto M^{-\alpha}$, where α is ~ 0.7.[95–97] From considerations of the polymer coil density and volume fraction Φ, it may be shown that the threshold value Φ^* for the transition between the dilute regime and the semidilute is related to the number of repeat units N in the macromolecular chain and the square of a characteristic parameter d/l (the ratio between the coil thickness d and its length l). An estimate of the volume fraction Φ^* shows that the onset of the semidilute regime occurs at much lower concentration for more elongated macromolecular chains. Theoretical findings are corroborated by experiments when the behavior of flexible chain polymers, for which the d/l value is between 1/2 and 1/3, is compared with that of DNA, which has a value of $\sim 1/50$. The critical concentration for this macromolecule hence decreases 2500-fold. The experimentally determined critical concentration reported for DNA is 2.2–2.6 μg per 100 μL.[98,99] An injection concentration 100 times lower than the critical value is recommended in FFF experiments. For synthetic polymers, concentrations of 0.05–0.50 g L^{-1} and injection volumes of 1–10 μL give easily detectable peaks under conditions of total sample recovery.[68]

4.2 Sample Injection and Relaxation

Sample solutions may be introduced into an FFF channel through an injection valve with a constant-volume loop such as those commonly used in chromatography or through an on-line tee-union fitted with a septum. The latter does not limit the injected volume but the polymer septum should be isolated from the channel with an on-line zero-dead-volume filter. When sample particles are first introduced into the channel, they are dispersed over the entire cross-section and experience the same field strength but different longitudinal migration velocities depending on their distance from the accumulation wall. While under the effect of the field alone the sample plug would concentrate in a narrow layer of exponential concentration. However, with displacement by the longitudinal flow, the plug undergoes a considerable dispersion along the channel length. This happens because molecules starting their migration far from the accumulation wall will take longer to reach the equilibrium position and will be swept ahead of species closer to the accumulation wall. A simple way to circumvent this problem and greatly reduce the relaxation contribution to the plate height is to halt the longitudinal flow as soon as the sample enters the channel. This stop-flow procedure allows the field to complete the sample relaxation process without the undesirable effects of differential migration velocities.[1] The stop-flow time, t_{sf}, depends on the channel thickness, the field strength, and the final transverse position of the zone's center-of-gravity. It has limits for low and high λ of $w^2/2D$ and $w^2/12D$, respectively. In practice, the relaxation process is started when all sample molecules have entered the channel. The delay time between the injection and the beginning of the relaxation is given by the ratio of the volume of the tubing between the injection port and the channel inlet and the volumetric flow rate. Although the stop-flow procedure has proven successful in yielding well-shaped peaks at the expected retention times, it is associated with a number of nonideal phenomena such as sample loss due to adsorption on the accumulation wall, baseline disturbance and increased analysis time. Different channel designs have been adopted to achieve sample relaxation without stopping the axial flow. A well-relaxed sample zone can rapidly form by applying a strong field in a small area localized at the channel inlet or by reducing the channel thickness in the same region. Hydrodynamic relaxation, currently applicable only to flow systems, may be achieved by isolating the inlet portion of the depletion wall and applying a higher field strength only in that area.[100] Sample components entering the channel are rapidly transported to the accumulation wall and relaxed. This system has been mostly used for biological macromolecules.[101] The pinched-inlet channel, applicable in principle to any FFF system,[102] has been tested experimentally on latex particles in an FFF channel under gravitational force.[103] Hydrodynamic relaxation obtained with a thin channel splitter[104] has been examined in SdFFF with latex particle samples.[105] Hybrid split and frit-inlet FlFFF systems have been designed and tested.[106]

The relaxation procedure in asymmetric FlFFF is somewhat different from that in a symmetrical system. It may be achieved virtually at any point along the channel

length by injecting sample material at a chosen distance from the channel inlet when two inflows of carrier liquid are introduced at both the channel inlet and outlet.[72] This procedure is termed ''opposing flows relaxation''. The two flows entering the channel through the inlet and the outlet meet at a focusing point that depends on the ratio between the inlet flow rate and the sum of the inlet and outlet flow rates. The focusing point in an asymmetric channel may thus be adjusted by selecting the flow rates. Alternatively, the flow entering through the inlet end may be eliminated so that only the reverse flow is maintained. This procedure defined ''reverse flow relaxation'' allows the sample to migrate to the very beginning of the channel and relax at the inlet point. When the opposing flows relaxation is used sample material may be loaded through either the inlet or the outlet or even an independent port at a given point along the channel. In this case, an additional pump must be used. The opposing flows approach has given rise to the opposed flow sample concentration procedure which allows very large volumes of dilute sample to be loaded on to a FlFFF channel.[107] As much as 10^5-fold concentration has been reported (1-L sample volume introduction). This sample concentration technique can be applied to symmetrical, asymmetric, and hollow-fiber FlFFF. In the hollow-fiber FlFFF technique, sample relaxation is obtained much as in the asymmetric set-up by pumping opposing flows in from the inlet and outlet while maintaining a constant radial flow.

4.3 Operation with Constant or Programmed Field

When complex mixtures have to be analyzed, differences in the property of interest of the species under investigation may be so large that no constant experimental condition may yield a satisfactory result for all species in a single run. This situation, well known in the chromatography of complex natural mixtures, is circumvented in HPLC with the so-called gradient elution, where the eluent composition is changed with time. The basic concept of gradually decreasing the retention power in order to let the most retained sample elute in a shorter time, thus gaining in detectability, has been applied to FFF.[108] Programmed field strength and/or flow rate are particularly useful for polydisperse samples whose elution would be spread over a wide time interval because of the high system selectivity. In FFF, there is an almost unlimited number of programming choices considering that both the field and the channel velocity may be varied as needed. In principle, the eluent composition may also be programmed but this option has only been applied to the carrier fluid density in SdFFF.[108] Both the fluid velocity and field strength may follow a step, linear, quadratic, parabolic, or exponential time-dependent function. In any of these forms of programming a period at constant conditions is usually applied to allow early-eluting particles to undergo adequate retention and separation before starting the programmed decrease of the field or increase of the fluid velocity. It has been shown that the fractionating power is a convenient universal parameter for comparing the effectiveness of different forms of programming. The diameter-based fractionating power, F_d, or the molecular-weight-based F_M, is defined as the resolution R_s of two sequentially eluting species divided by the relative difference in their size or molecular weight.[109] Linear and parabolic decay programs,[64,110] the first functions to be investigated for polymer separations, resulted in a considerable improvement over a constant field strength separation. However, closer examination showed an initial rise in F_d followed by a rapid decrease as the field strength went to zero.[111] The exponential field decay program introduced by Kirkland et al.[112,113] for particle size analysis by SdFFF has been applied to polymer separations by FlFFF[114] and ThFFF.[110,115–117] This type of program also resulted in an initial increase of the fractionating power followed by a strong decrease.[109] Only a power-law dependence of the field decay proposed by Williams and Giddings[118] yielded a constant molecular-based fractionating power for a wide range of molecular weights.[119]

5 APPLICATIONS

5.1 Organic-soluble Polymers

The high versatility of FFF has proven suitable to so many applications in different fields that a complete survey would be impossible. We therefore choose to report here only selected examples of the most innovative and recent applications. Many others may be found in the literature.

ThFFF with field-strength programming has been used to separate PS standards with molecular weights spanning 4000–7 100 000 Da. In Figure 7(a), the ΔT follows a parabolic function that decreases from 70 to 0 °C.[110] Since this early ThFFF work, the analysis time has been dramatically reduced to 20 min without significant loss of resolution by the advent of new instrumentation and the introduction of the power program function.[119] An application of this form of programmed elution is shown in Figure 7(b).

Although FFF was conceived as a separative technique, its rigorous theoretical background has demonstrated that measurements of fundamental physicochemical properties may be very accurate and in some cases unachievable by other techniques. This is the case for investigations of the thermal diffusion of polymers and copolymers and their correlation with the polymer and solvent chemical composition. Studies in this field showed a linear relationship between the thermal diffusion coefficient D_T and the

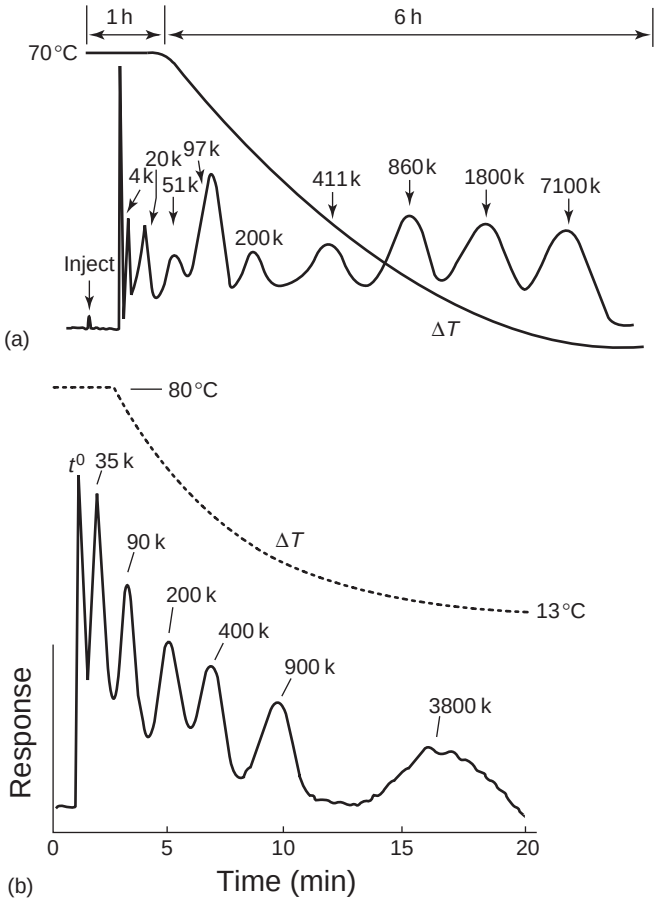

Figure 7 Separation of two mixtures of PS standards in a similar molecular weight range with field strength decay. (k = kDa) (a) Parabolic programming. (Reprinted with permission from J.C. Giddings, L.K. Smith, M.N. Myers, *Anal. Chem.*, **48**, 1587–1592 (1976). Copyright 1990, American Chemical Society.) (b) Power programming. (Reprinted with permission from M. Myers, P. Chen, J.C. Giddings, *ACS Symp. Ser.*, **521**, 47–62 (1993). Copyright 1993, American Chemical Society.)

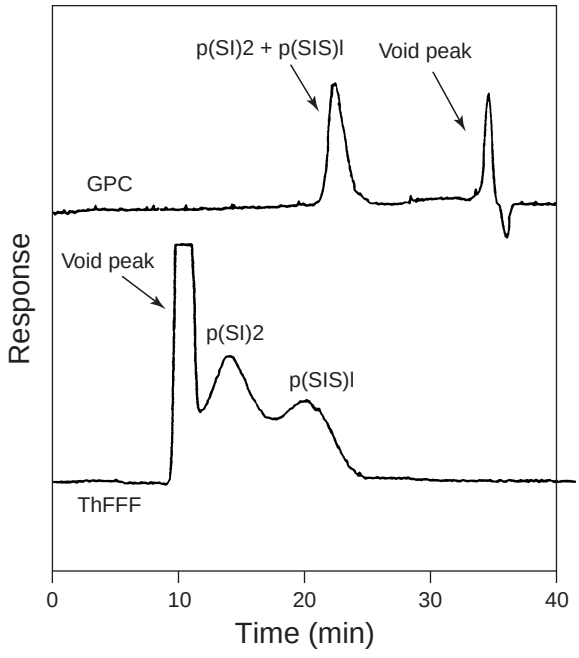

Figure 8 Co-elution of a sample of diblock copolymer of poly(styrene-*co*-isoprene) and triblock poly(styrene-*co*-isoprene-*co*-styrene) from a GPC column and separation of the same mixture by ThFFF based on the difference in thermal diffusion coefficient. (Reprinted from Cho et al.,[71] by courtesy of Marcel Dekker Inc.)

temperature at the center of the sample zone for PS in ethylbenzene[36] and a similar correlation between D_T and the mole fraction of one of the monomers in random copolymers.[120] This finding has a number of implications, one of which is that ThFFF has an additional separating dimension that allows samples to be resolved according to chemical composition as well as hydrodynamic size. This is shown in Figure 8, where the diblock copolymer poly(styrene-*co*-isoprene) is separated from the triblock of same size poly(styrene-*co*-isoprene-*co*-styrene) only by ThFFF. Hydrodynamic chromatography (HDC), another analytical technique for polymer separations, has a higher efficiency than FFF but, like GPC, discriminates samples only by size. Consequently, fractions of PS, polyisoprene and polybutadiene (PB) with similar hydrodynamic dimensions are only partially separated by HDC whereas they are completely resolved by ThFFF[121] because of their different thermal diffusion coefficients. Although retention in ThFFF is dependent on D_T, the evaluation of absolute values of this parameter is not straightforward since the measurable retention parameter λ yields values of D/D_T. D_T alone may therefore be determined only if the diffusion coefficient is measured by an independent technique. One approach is to couple SEC to ThFFF and determine D using light scattering. This multidimensional approach allows the fractionation of polymer samples according to size by SEC and to thermal diffusion by ThFFF.[122,123] The usefulness of the combined SEC/ThFFF technique is demonstrated in the analysis of polydisperse samples of copolymers whose relative composition, which may vary with molecular weight, gives rise to materials with different properties. The preliminary fractionation of a polydisperse sample of PS by SEC may occur with a selectivity of 0.15 in the molecular weight range 150 000–1 000 000 Da and drops to 0.04 for lower M fractions.[122] This selectivity, well below that commonly found in FFF, is obtained after a system calibration with well-characterized PS fractions and the light-scattering determination of diffusion coefficients. The thermal diffusion coefficient, calculated from retention measurements once both molecular weight and ordinary diffusion coefficients are known, is generally independent of molecular weight.[124] Fractionation according to

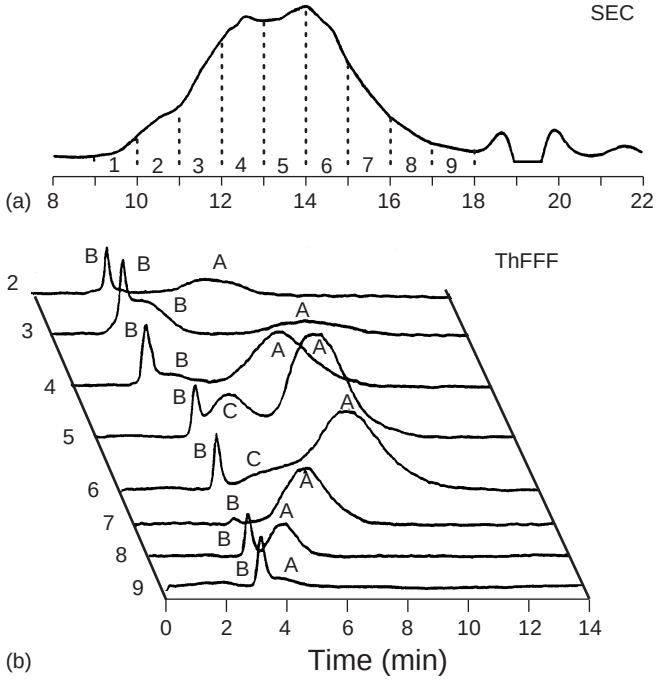

Figure 9 (a) Size exclusion chromatogram of a blend of PS, PB and PTHF and (b) ThFFF of individual fractions taken after the SEC elution. Numbers along the *y*-axis of the ThFFF fractogram correspond to the labeled SEC fractions in (a). (Reprinted from A.C. van Asten, R.J. van Dam, W.Th. Kok, R. Tijssen, H. Poppe, *J. Chromatogr. A*, **703**, 245–263 (1995). Copyright 1995, with permission from Elsevier Science.)

differences in thermal diffusion coefficient alone gives information on the copolymers' chemical composition. Figure 9(a) and (b) show the partial fractionation (a) by SEC of a blend of PS, PB, and polytetrahydrofuran (PTHF) and then (b) by ThFFF of some SEC fractions collected at different retention times. The ThFFF fractogram indicates the presence of different polymer species (fractions 3–6) whose measured D_T are in good agreement with values of the corresponding PB and PTHF homopolymers determined by ThFFF and light scattering, e.g. 0.22×10^{-7} and 0.47×10^{-7} cm^2 s^{-1} K^{-1} versus 0.23×10^{-7} and 0.50×10^{-7} cm^2 s^{-1} K^{-1}.[125]

The measurement of the thermal diffusion coefficient has become a key step in the determination of copolymer relative chemical composition. Two styrene–methyl methacrylate copolymers with different styrene content analyzed by SEC/ThFFF[122] show different D_T versus retention time trends when analyzed by ThFFF, although the SEC traces are almost identical. Using a calibration plot for the SEC column based on PS standards (which is not a rigorous procedure), the fractions may be assigned a molecular weight and D_T may be related to this parameter. The invariant trend of the thermal diffusion coefficient of one sample and the clearly increasing values of the second sample suggest that the styrene percentage is different and that it increases with M in the second case.[122] Increasing D_T with increasing weight fractions of vinyl acetate is reported for poly(ethylene-*co*-vinyl acetate) copolymers.[83] Multidimensional analysis may also be obtained by coupling a ThFFF system with HDC.[126] Samples with the same D/D_T eluting from the ThFFF system are subsequently fractionated according to size by HDC.

ThFFF appears to be the only separation technique suitable for the analysis of gels and rubbers[82,90,127] because irreversible adsorption evident in other methods is minimized.[90,127] In addition, the need to filter samples prior to their injection into a GPC column results in the removal of high-molecular-weight polymer components as well as of the gel. This causes the average molecular weight and molecular weight distributions to be consistently lower in GPC than in ThFFF.[127] The open FFF channel design eliminates the problem of plugging that is encountered in GPC when gel containing samples are not filtered before the analysis. Hence polymer and gels may be completely separated within a single run by ThFFF,[128] as shown in Figure 10.

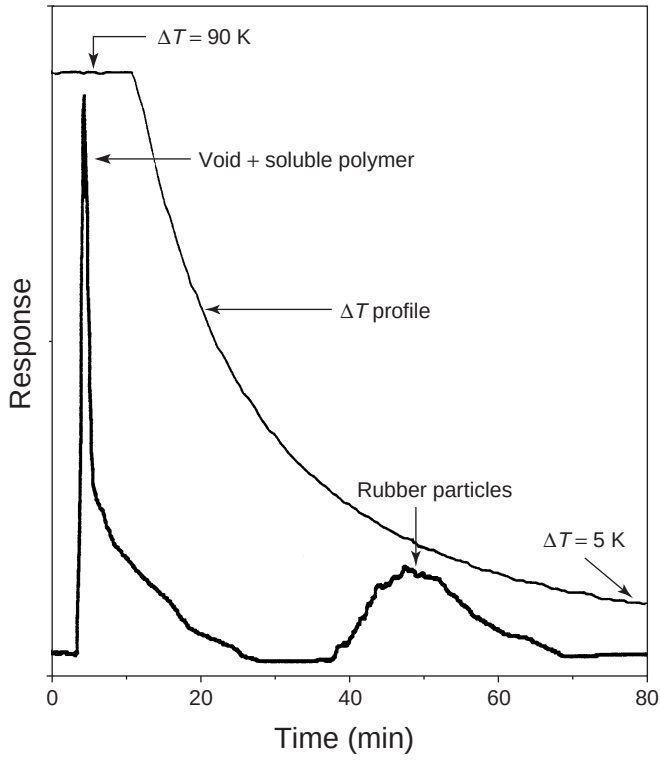

Figure 10 Separation and characterization of the polymer and gel components present in acrylonitrile–butadiene–styrene plastics. (Reproduced from P.M. Shiundu, E.E. Remsen, J.C. Giddings, *J. Appl. Polym. Sci.*, **60**, 1695–1707 (1996). Copyright © 1996, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

Molecular weight and particle size distributions are calculated for both the polymer and the gel components. Although other techniques have been used successfully to analyze gels and rubbers, SEC has been considered the separative methodology par excellence in spite of its limitations.[129,130] It was shown that the different D_T values for different polymers have been exploited in the ThFFF analysis of core-shell latex particles.[131] The retention time is sensitive to the composition of the polymer shell. A calibration curve may be drawn to relate retention time and percentage of methacrylic acid in the shell. The sensitivity of D_T to the particle surface composition is further illustrated using similar sized particles of different polymeric and inorganic surfaces.[132,133]

5.2 Water-soluble Polymers

Early studies in ThFFF[37] showed that different organic solvents had a very similar effect on the retention of polymers. In contrast, when water is used as the carrier liquid, the thermal diffusion factor α is very low and retention is negligible unless a considerable amount of some organic solvent (~60%) is present. Except for a few results confirming the poor retention in such a solvent,[117] water-soluble synthetic polymers have been analyzed by the most versatile technique of the FFF category, FlFFF. This technique was recognized since its first appearance as highly suited to polymer fractionation.[134,135] Although resolution and analysis time were not optimized, the early results allowed the determination of molecular parameters, such as hydrodynamic size, on the basis of theoretical concepts that have subsequently been extensively confirmed. Polyacrylamide (PAAm) is a widely employed polymer in many fields that is difficult to characterize and fractionate. Commercially available fractions of PAAm generally have a broad distribution.[136] The FlFFF fractograms of the three PAAm samples illustrated in Figure 11 consistently show the presence of low-molecular-weight components in each polymer sample.[68] A similar peak asymmetry is also observed for broadly disperse samples in organic solvents.[63] As mentioned earlier (section 2.4), in the absence of an absolute technique for the molar mass determination, the molecular weight distribution may be accurately determined by calibration. The calibration procedure, however, should be performed with narrowly distributed, well-characterized standards of closest possible chemical composition to the unknown. When standards with these characteristics are not available, absolute measurements of the diffusion coefficient may be used to evaluate the polymer hydrodynamic size using the Stokes–Einstein equation. The absence of non-ideal sample–wall interactions on the elution of PAAm is verified by comparing measurements

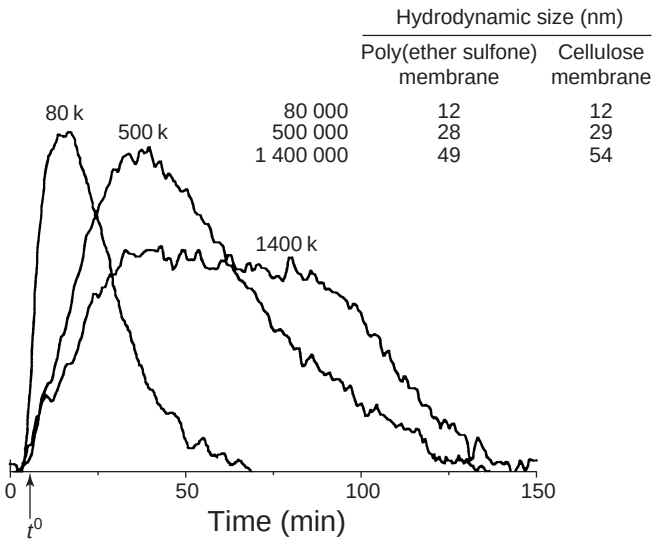

Figure 11 FlFFF separation of three samples of PAAm of different nominal molecular weight. The hydrodynamic diameters were determined using two different membranes. (Reproduced from M.-A. Benincasa, J.C. Giddings, *J. Microcol. Sep.*, **9**, 479–495 (1997). Copyright 1997, John Wiley & Sons, Inc. Reprinted with permission of John Wiley & Sons, Inc.)

made in solutions of identical ionic strength but with membranes of different composition, namely poly(ether sulfone) and cellulose.[137]

As discussed in the theory section, the overall polymer molecular dimensions are anticipated to depend, to a certain extent, on the properties of the surrounding medium. This effect is expected to be enhanced for charged polymers. The effect of solvent and ionic strength was identified in the first study in FlFFF of water-soluble polymers.[135] Poly(ethylene oxide) (PEO), a polymer widely used in biomedical and biotechnological applications because of its non-toxicity, shows an almost ideal correlation between the measured diffusion coefficient and molecular weight in aqueous sodium sulfate solutions.[69] Similar to the observation reported by Hassellöv et al.,[56] this correlation is found for polymers with molecular weights well below the range for which the general polymer scaling laws are expected to hold. The correlation between diffusion coefficient and molecular weight in potassium sulfate is very different to that in sodium sulfate. The authors attribute this difference to the capability of low-molecular-weight PEO fractions to form complexes with some metal cations, as reported in independent studies.[138] The elution profiles of PEO samples in a molecular weight range 250 000–1 000 000 Da may be obtained by FlFFF with good resolution as shown in Figure 12. Charged amphiphilic graft copolymers are a particular type of sample that may dramatically change their conformation in different solvents because of the presence of hydrophilic and hydrophobic

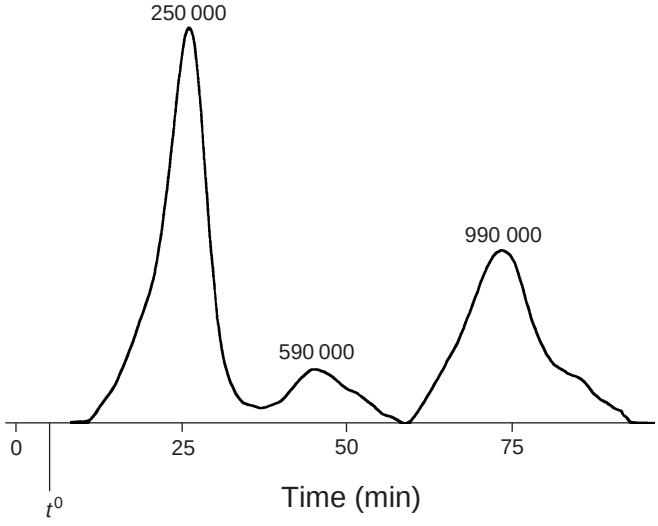

Figure 12 Separation of three PEO samples by FlFFF in aqueous 0.025 M Na_2SO_4.

moieties derived from different homopolymers. In aqueous solutions, they form a hydrophobic core with the hydrophilic moieties on the surface. At low pH and in the presence of salts, aggregation may occur for the copolymer of styrene, methyl methacrylate and maleic anhydride with grafts of poly(ethylene oxide) monomethyl ether (MPEO).[139] The polymer molecular conformation and solubility in water depend on the pH, which affects dissociation of the carboxyl groups, and on the solvent ionic strength. An extended conformation may exist at high pH or aggregation may occur at low pH and high ionic strength. The copolymer hydrodynamic diameter increases with decreasing pH. This effect is attributed to the formation of aggregates of less charged copolymers at low pH, a phenomenon that seemed amplified when long stop-flow times are used. When a ca. 0.002 M buffer solution is used as the carrier liquid, the molecular dimensions are considerably higher (16–40 instead of 2–20 nm) than those measured in unbuffered solutions. It is postulated that the buffer salts cause shielding of charged sites which consequently promote aggregation through hydrophobic interactions of the polymer backbone. Formation of aggregates is evidenced by the particle size distributions shown in Figure 13.[140]

SdFFF is a technique usually associated with particle rather than polymer analysis. The maximum field strength or centrifuge speed of currently available commercial instruments is insufficient to induce transport of the polymers to their equilibrium positions at the accumulation wall. Consequently, retention and separation by SdFFF are poor. However, an instrument with high-field-strength capabilities can be used to characterize high-molecular-weight polymers.[141]

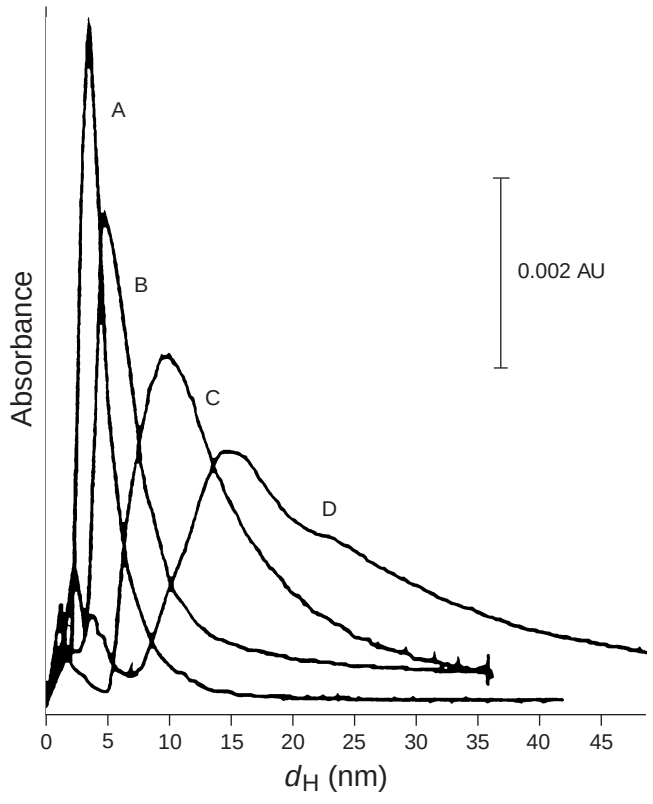

Figure 13 FlFFF-derived particle size distributions of an amphiphilic grafted copolymer poly(styrene-*co*-methyl methacrylate-*co*-maleic anhydride) and MPEO. Aggregation is observed as the sodium sulfate concentration is increased. The carrier liquids used were (A) pure water, (B) 1 μM Na_2SO_4, (C) 10 μM Na_2SO_4, and (D) 100 μM Na_2SO_4. (Reprinted with permission from B. Wittgren, K.-G. Wahlund, H. Dérand, B. Wesslén, *Langmuir*, **12**, 5999–6005 (1996). Copyright 1996, American Chemical Society.)

6 METHOD DEVELOPMENT

6.1 Determining Experimental Conditions

Before starting an analysis, specific requirements may be set for a number of experimental parameters such as analysis time, resolution, and selectivity. A thorough discussion of this topic would require a long navigation through FFF theory and is beyond the scope of this article. Therefore, some practical recommendations are given here with reference to theoretical fundamentals on which they are based. Demonstration of these concepts may be found in the specific literature and will not be dealt with here. A generally desired characteristic is short analysis times. The retention time in FFF is linearly related to the applied field strength, ceteris paribus. This translates into a dependence on the temperature gradient, crossflow rate, solvent viscosity, channel thickness, and void volume.

Although in theory values of these parameters may be accurately determined, common laboratory practice has shown that void volume V^0 and channel thickness determination in FlFFF is not a trivial procedure. A peak breakthrough technique[142] has been proposed that allows the determination of the channel void time, and related volume, from the time needed for an unretained sample to emerge from the channel under high-flow-rate conditions. Experiments must be carried out with great care and in the total absence of a crossflow. A simple way to ensure this is to block the crossflow inlet and outlet. More accurate void volume determinations may be obtained by sandwiching the spacer and the membrane between two glass plates with holes drilled through one of the plates to act as the channel inlet and outlet. The determination of the actual void volume in FlFFF may not be bypassed since it also yields a measure of the channel thickness. This last parameter is the most critical in FlFFF since the retention time varies with w^2. Molecular weight measurements, obtained from retention parameters, are strongly affected by the channel thickness. The crossflow rate $\dot{V}_c$ and channel flow rate $\dot{V}$ have opposing effects on t_r since they contribute as the ratio $\dot{V}_c/\dot{V}$. An increase in the two flow rates which does not alter this ratio would have no effect on t_r. However, λ and t^0 would decrease. This would translate into a decrease of the retention ratio t^0/t_r and a higher compression of the sample zone with a greater probability of overloading and sample interaction with the accumulation wall. Resolution, which depends on $(\dot{V}_c^3/\dot{V})^{1/2}$, would increase with the 3/2 power of the crossflow rate and decrease with the 1/2 power of the channel flow rate. Studies on the effect of field strength and injected sample load on polymer fractionation by FlFFF have shown that the molar mass distribution seems to broaden with increasing crossflow rates or decreasing injected sample loads.[92] Unlike ThFFF, where the flow profile is considerably affected by the field, perturbations due to the crossflow in FlFFF are negligible.[143] In general, axial flow rates of 0.2–2 mL min^{-1} are used for polymer analysis. Low flow rates and velocities protect samples from shear degradation and peak broadening due to the nonequilibrium contribution to the plate height that depends linearly on the flow velocity.

Unlike FlFFF, the retention time in ThFFF does not depend directly on the channel thickness. It depends on the void time t^0, which is related to w. The retention time is a function of the thermal gradient dT/dx, which can be increased by reducing w or increasing the hot wall temperature. The first approach is useful when the working temperature is above the solvent boiling point and further pressurization of the system to elevate the boiling point[3] is not an option. However, the reduction of w in ThFFF is limited by heat transfer that may require substantial heat fluxes between the hot and cold walls. In addition to the previous considerations, w affects resolution and sample dilution in all the FFF techniques. Generally, 90–99% of the FFF channel volume is occupied by pure solvent during elution. The injected sample is hence very diluted on elution from the FFF channel. This effect may be reduced using stream splitters[104] or frit outlet systems.[100] All channel dimensions may in principle be varied. An increase or reduction of either b or L has some advantages and some disadvantages. An extensive discussion on the theoretical and practical aspects of changing these dimensions is reported in the literature.[144] The dimensions of asymmetric FlFFF channels are subject to more constraints.[145] Commonly used channel flow rates in ThFFF are of the same order of magnitude as those used in FlFFF. Data collected over a 15-year period using a number of different channels have shown that retention in ThFFF is affected by the absolute value of the cold wall temperature T_c.[146] Higher T_c values lead to lower retention with the same ΔT. The use of binary solvents in ThFFF has been shown to enhance retention considerably in some cases. This result may be used to extend the range of applicability of ThFFF toward lower molecular weight limits.[147]

6.2 Effect of Sample Size

Sample size effects on retention were described in early FFF studies.[135] These effects, common to other separation techniques, manifest themselves in FFF as distortions of the elution profile and shifts in retention time that cannot be related to any sample physicochemical property but rather to the amount of sample injected. It was also recognized in the early FFF work that the ionic strength changes remarkably the effect of sample load on retention time.[135,148] The effect of sample load depends on the polymer–solvent system rather than on the FFF technique employed. The general trend of increasing retention time with increasing injection amount is observed for FlFFF and ThFFF of PS in ethylbenzene and tetrahydrofuran.[81,149] In this case the role of molecular weight in enhancing the effect of load even at very low ΔT values was shown. Longer retention times are also reported for higher amounts of PTHF analyzed in toluene by ThFFF.[150] The opposite trend is found in aqueous separations of particles by SdFFF[148] and for synthetic[65,68,135] and biological polymers.[151] The dependence of sample overloading on the physicochemical properties of the polymer–solvent system rather than on the analytical technique is substantiated by the findings that aqueous synthetic and biological polymer systems also show a decrease in retention time with increase of load in

hollow-fiber FlFFF.[152,153] The observed decrease in retention time with increasing sample load for particles and charged polymers has been explained on the basis of excluded volume effects. For systems where the sample volume fraction is not negligible, λ values are expected to be higher than those predicted from standard FFF theory, which assumes infinitely small non-interacting species.[154] Since both the interparticle electrostatic repulsion and chain expansion due to intramolecular repulsion contribute to an increase in the effective volume occupied by the sample species, an enhanced overloading effect is expected for charged particles. In contrast, a depression of the phenomenon may be predicted with a reduction of the electrostatic effects. It follows that the ionic strength of the carrier liquid should considerably affect the onset of sample overloading. High ionic strength is expected to reduce both the overall molecular dimensions of flexible-chain macromolecules and the double-layer thickness. The reduction of these parameters decreases the effective sample volume to a greater extent than the lower solvating power of a concentrated salt solution alone. Experiments on synthetic water-soluble polymers show that the ionic strength[65] and the type of electrolyte added to the carrier liquid may affect the relationship between sample size and retention.[69,139]

6.3 Membranes in Flow Field-flow Fractionation

A unique feature of the FlFFF system is the semipermeable membrane that is laid over the frit and serves as the accumulation wall. The possibility of wrinkling or swelling of the membrane and uncertainties in the performance of the different polymer materials used in their fabrication have been the main deterrent in the development of FlFFF. The membrane used in the FlFFF channel must meet a number of specific requirements not necessarily fulfilled by commercially available membranes. The pore size and the pore density, i.e. the number of pores per unit area, have to be uniform. Inhomogeneous pore density leads to regions of nonuniform permeability and crossflow rate. The nominal and effective pore sizes are often given by the manufacturer. While the first is an absolute measurement obtained by electron microscopy, the second must be considered an indication of the membrane performance with respect to the specific probe and to the conditions used for this determination and may not be directly transferred to samples with different physicochemical properties. For example, the Celgard isotactic polypropylene membrane is marketed with a nominal pore size of 50 nm and an effective pore size of 20 nm, but samples of much smaller diameter are retained.[65] It is known that filtration through a membrane is not a purely mechanical process and that a number of parameters contribute to the sample permeation or retention by a membrane. A higher percentage of latex particles of considerably smaller diameter than that of PS samples in tetrahydrofuran are retained by a polytetrafluoroethylene (PTFE) membrane with a nominal 20-nm pore size.[155] A regenerated cellulose membrane from Millipore with a 10 000 molecular weight cut-off (MCO) retains PEO samples in the 4000–1 000 000 molecular weight range whereas a membrane of the same material and from the same supplier with a lower cut-off shows no elution of the PEO samples.[156] Considering the lower MCO, it may be inferred that adsorption occurred in this case. Specific tests to check sample permeation through the membrane are always recommended, however. They may be carried out by connecting the crossflow outlet to a detector and monitoring the eluted carrier liquid for the presence of sample. The general scarcity of membranes capable of withstanding organic solvents has been one of the main limitations in the application of FlFFF to organosoluble polymers. Cellulose nitrate gave a good performance in the first experiments of FlFFF in ethylbenzene,[81] but many more membrane materials compatible with organic solvents such as fluoropolymers, polyvinylidene, polyaramide and PTFE are now commercially available.[155]

Ultrafiltration membranes such as those commonly used in FlFFF may be classified as cellulosic and noncellulosic. Cellulose and its derivatives were one of the first materials employed as a semipermeable membrane and successfully used in the FlFFF analysis of aqueous systems of proteins, synthetic hydrophilic polymers, and latex particles. Cellulosic ultrafiltration membranes are available in a variety of MCOs and are cast on a thicker, more permeable support material. The presence of this support adds robustness and ease of handling to the membrane. Thin-film (5–25 μm) unsupported membranes, most often noncellulosic, are more difficult to handle but their flexibility allows easy positioning on the frit wall. Unlike cellulosic ultrafiltration membranes, these mainly hydrophobic membranes do not allow wicking of carrier liquid out of the area of permeation (no leaking). An ample variety of membranes used with various solvents and samples is listed in Table 1. Besides preliminary considerations on the physicochemical properties of both the membrane material and the sample to be analyzed, an unambiguous answer on the performance of a membrane is given by a test of the absolute sample recovery.[68,137,157] This is accomplished by comparing the peak area of the output signal of a regular FFF run with the area acquired upon injecting the same sample amount through an open tube.

Table 1 Summary of membrane materials and their applications in FlFFF

Membrane material	Sample	Carrier composition
Acrylic copolymer	Proteins, lipoproteins	Aqueous solution
Cellulose	PAAm, humics	Aqueous solution
Cellulose acetate	Viruses, proteins	Aqueous solution
Cellulose nitrate	PS	Ethylbenzene
Fluoropolymer	PS	Tetrahydrofuran
Polyamide	Polyethylene, toner pigment	Xylene Toluene
Polyaramide	PS	Tetrahydrofuran
Polycarbonate	Antibodies, proteins, humic and fulvic acids	Aqueous solution
Polyelectrolyte complex	Proteins, viruses	Aqueous solution
Poly(ether sulfone)	PAAm, PEO, Poly(styrene sulfonate), PS, carbon black, Polyethylene	Aqueous solution Xylene, cyclohexane Xylene
Poly(ethylene terephthalate)	Proteins	Aqueous solution
Poly(phenylene oxide)	PS	Tetrahydrofuran
Polypropylene, isotactic	Poly(styrene sulfonate), polyvinylpyridine, proteins	Aqueous solution
Polysulfone	Minerals, cells, humic and fulvic acids, poly(styrene sulfonate), dextrans, latex	Aqueous solution
PTFE	PS, PEO, latex, silica	Tetrahydrofuran, acetonitrile
Polyvinylidene	PS	Tetrahydrofuran
Regenerated cellulose	Algae, bacteria, amphiphilic copolymers, DNA, hemoglobin, microspheres, lipoproteins, liposomes, nucleic acids, plasmids, pollens, polysaccharides, PEO, ribosomes, silicas, Poly(methyl methacrylate)	Aqueous solution Tetrahydrofuran
Regenerated cellulose, modified	PEO	Ammonium acetate in methanol solution

Literature related to the applications reported here may be found elsewhere.[137,155,157,158]

ACKNOWLEDGMENTS

K.R.W. acknowledges support from the Colorado School of Mines by a start-up grant for this work.

LIST OF SYMBOLS

a	axis of prolate or oblate particles
A	constant defined by DM^b
b	axis of prolate or oblate particles
b	field-flow fractionation channel breadth
b	exponent in diffusion coefficient expression
B	empirical constant in Equation (22)
c	particle or molecule concentration
c_0	particle or molecule concentration at the wall
$c(t_r)$	time-dependent sample mass concentration at elution
d	width of polymer chain
D	diffusion coefficient
D_0	diffusion coefficient at infinite dilution
D_T	thermal diffusion coefficient
f	friction coefficient
f_0	friction coefficient of an isolated sphere
F_d	diameter-based fractionating power
F_m	molecular weight-based fractionating power
H	plate height
H_i	instrumental contribution to plate height
H_l	longitudinal diffusion contribution to plate height
H_n	nonequilibrium contribution to plate height
H_p	polydispersity contribution to the plate height

H_r	relaxation contribution to plate height
J_x	flux of particles
k	Boltzmann constant
ℓ	sample mean layer thickness
l	length of polymer chain
L	channel length
$m(M)$	mass distribution as a function of molecular weight
M	molecular weight
M_0	molecular weight of repeat unit
δM	difference in molecular weight interval
M_w	weight-average molecular weight
M_N	number-average molecular weight
n	exponent in Equation (24)
N	number of repeat units
R	retention ratio in Equation (6)
R_e	equivalent sphere radius
R_g	radius of gyration
R_0	spherical particle radius
R_s	resolution
S_M	molecular weight selectivity
t^0	void time
t_r	retention time
δt_r	retention time difference
dt_r/dM	scale correction function
t_{sf}	stop-flow time
T	absolute temperature
T_c	cold wall temperature
ΔT	temperature difference across the channel
U	field-induced velocity
$\dot{V}$	volumetric flow rate
$\dot{V}_c$	volumetric cross flow rate
V^0	void volume
V_r	retention volume
$\langle v \rangle$	mean fluid velocity
v_p	zone migration velocity
w	channel thickness

Greek characters

α	thermal diffusion factor
β	exponent in thermal diffusion Equation (22)
γ	thermal expansion coefficient
η	fluid viscosity
λ	retention parameter
μ	molecular weight polydispersity
Φ	coefficient in Equation (24) or polymer volume fraction (section 4.1)

ABBREVIATIONS AND ACRONYMS

CFFF	Concentration Field-flow Fractionation
ElFFF	Electrical Field-flow Fractionation
FFF	Field-flow Fractionation
FlFFF	Flow Field-flow Fractionation
GPC	Gel Permeation Chromatography
HDC	Hydrodynamic Chromatography
HPLC	High-performance Liquid Chromatography
MALS	Multiangle Light Scattering
MCO	Molecular Weight Cut-off
MPEO	Poly(ethylene oxide) Monomethyl Ether
PAAm	Polyacrylamide
PB	Polybutadiene
PEO	Poly(ethylene oxide)
PS	Polystyrene
PTFE	Polytetrafluoroethylene
PTHF	Polytetrahydrofuran
SdFFF	Sedimentation Field-flow Fractionation
SEC	Size-exclusion Chromatography
ThFFF	Thermal Field-flow Fractionation

RELATED ARTICLE

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

REFERENCES

1. J.C. Giddings, 'A New Separation Concept Based on a Coupling of Concentration and Flow Nonuniformities', *Sep. Sci.*, **1**, 123–125 (1966).
2. P.J.M. Dycus, K.D. Healy, G.K. Stearman, M.J.M. Wells, 'Diffusion Coefficients and Molecular Weight Distribution of Humic and Fulvic Acids Determined by Flow Field-flow Fractionation', *Sep. Sci. Technol.*, **30**, 1435–1453 (1995).
3. J.C. Giddings, L.K. Smith, M.N. Myers, 'Thermal Field-flow Fractionation: Extension to Low Molecular Weight Separations by Increasing the Liquid Temperature Range Using a Pressurized System', *Anal. Chem.*, **47**, 2389–2394 (1975).
4. S.K. Ratanathanawongs, J.C. Giddings, 'Particle Size Analysis Using Flow Field-flow Fractionation', *ACS Symp. Ser.*, **521**, 30–46 (1993).
5. J.C. Giddings, 'Theoretical Basis of Macromolecular Chromatography', *J. Gas Chromatogr.*, **5**, 413–419 (1967).
6. J.C. Giddings, 'Field-flow Fractionation: Extending the Molecular Weight Range of Liquid Chromatography to One Trillion', *J. Chromatogr.*, **125**, 3–16 (1976).
7. P.S. Williams, J.C. Giddings, R. Beckett, F. Dondi 'Nomenclature for Field-flow Fractionation', to be submitted.

8. J.C. Giddings, M.N. Myers, 'Steric Field-flow Fractionation: a New Method for Separating 1–100 μm Particles', *Sep. Sci. Technol.*, **13**, 637–645 (1978).
9. M.N. Myers, J.C. Giddings, 'Properties of the Transition from Normal to Steric Field-flow Fractionation', *Anal. Chem.*, **54**, 2284–2289 (1982).
10. J.C. Giddings, 'Field-flow Fractionation', *Chem. Eng. News*, **66**, 34–45 (1988).
11. J.C. Giddings, 'Field-flow Fractionation', *Sep. Sci. Technol.*, **19**, 831–847 (1984).
12. J.C. Giddings, F.J.F. Yang, M.N. Myers, 'Sedimentation Field-flow Fractionation', *Anal. Chem.*, **46**, 1917–1924 (1974).
13. J.C. Giddings, F.J.F. Yang, M.N. Myers, 'Flow Field-flow Fractionation: a Versatile New Separation Method', *Science*, **193**, 1244–1245 (1976).
14. G.H. Thompson, M.N. Myers, J.C. Giddings, 'An Observation of a Field-flow Fractionation Effect with Polystyrene Samples', *Sep. Sci.*, **2**, 797–800 (1967).
15. K.D. Caldwell, L.F. Kesner, M.N. Myers, J.C. Giddings, 'Electrical Field-flow Fractionation of Proteins', *Science*, **176**, 296–298 (1972).
16. P.L. Dubin, M. Tecklenburg, 'Size-exclusion Chromatography of Strong Polyelectrolytes on Porous Glass Columns' *Anal. Chem.*, **57**, 275–279 (1985).
17. R.G. Keil, E. Tsamakis, C.B. Fuh, J.C. Giddings, J.I. Hedges, 'Mineralogical and Textural Controls on the Organic Composition of Coastal Marine Sediments: Hydrodynamic Separation Using SPLITT Fractionation', *Geochim. Cosmochim. Acta*, **58**, 879–893 (1994).
18. J.C. Giddings, F.J. Yang, M.N. Myers, 'Flow Field-flow Fractionation: New Method for Separating, Purifying and Characterizing the Diffusivity of Viruses', *J. Virol.*, **21**, 131–138 (1977).
19. J.C. Giddings, 'Nonequilibrium Theory of Field-flow Fractionation', *J. Chem. Phys.*, **49**, 81–85 (1968).
20. J.C. Giddings, F.J.F. Yang, M.N. Myers, 'Criteria for Concentration Field-flow Fractionation', *Sep. Sci.*, **12**, 381–393 (1977).
21. J.M. Davis, J.C. Giddings, 'Feasibility Study of Dielectrical Field-flow Fractionation', *Sep. Sci. Technol.*, **21**, 969–989 (1986).
22. T.C. Schunk, J. Gorse, M.F. Burke, 'Parameters Affecting Magnetic Field-flow Fractionation of Metal Oxide Particles', *Sep. Sci. Technol.*, **19**, 653–666 (1984).
23. V.L. Kononenko, J.C. Giddings, M.N. Myers, 'On the Possibility of Photophoretic Field-flow Fractionation', *J. Microcol. Sep.*, **9**, 321–327 (1997).
24. S.N. Semenov, K.I. Moslow, 'Acoustic Field-flow Fractionation' *J. Chromatogr.*, **446**, 151–156 (1988).
25. K.D. Caldwell, Y.-S. Gao, 'Electrical Field-flow Fractionation in Particle Separation. 1. Monodisperse Standards', *Anal. Chem.*, **65**, 1764–1772 (1993).
26. A. Einstein, *Investigations on the Theory of Brownian Movement*, Dover Publications, New York, Chapter 5, 1956.
27. G.G. Stokes, 'On the Theories of the Internal Friction of Fluids in Motion, and of the Equilibrium and Motion of Elastic Solids', *Cambridge Philos. Soc. Trans.*, **8**, 287–319 (1849).
28. G.G. Stokes, 'On the Effect of the Internal Friction of Fluids on the Motion of Pendulums', *Cambridge Philos. Soc. Trans.*, **9**, 8–106 (1856).
29. F. Perrin, 'Brownian Movement of Ellipsoids. II. Free Rotation and Depolarization of Fluorescence. Translation and Diffusion of Ellipsoidal Molecules', *J. Phys. Radium*, **7**, 1–11 (1936).
30. R.O. Herzog, R. Illig, H. Kudar, 'Diffusion in Molecularly Dispersed Solutions', *Z. Phys. Chem. A*, **167**, 329–342 (1934).
31. J.J. Gunderson, K.D. Caldwell, J.C. Giddings, 'Influence of Temperature Gradients on Velocity Profiles and Separation Parameters in Thermal Field-flow Fractionation', *Sep. Sci. Technol.*, **19**, 667–683 (1984).
32. J.E. Belgaied, M. Hoyos, M. Martin, 'Velocity Profiles in Thermal Field-flow Fractionation', *J. Chromatogr. A*, **678**, 85–96 (1994).
33. H.J.V. Tyrrel, *Diffusion and Heat Flow in Liquids*, Butterworths, London, 1961.
34. C. Tanford, *Physical Chemistry of Macromolecules*, John Wiley & Sons, New York, 1961.
35. V.J. Novotny, 'Temperature Dependence of Hydrodynamic Dimensions of Polystyrene in Cyclohexane by Quasielastic Light Scattering', *J. Chem. Phys.*, **78**, 183–189 (1983).
36. S.L. Brimhall, M.N. Myers, K.D. Caldwell, J.C. Giddings, 'Study of Temperature Dependence of Thermal Diffusion in Polystyrene/Ethylbenzene by Thermal Field-flow Fractionation', *J. Polym. Sci., Polym. Phys. Ed.*, **23**, 2443–2456 (1985).
37. J.C. Giddings, K.D. Caldwell, M.N. Myers, 'Thermal Diffusion of Polystyrene in Eight Solvents by an Improved Thermal Field-flow Fractionation Methodology', *Macromolecules*, **9**, 106–112 (1976).
38. P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, 2nd Edition, Cornell University Press, Ithaca, NY, 1985.
39. P.J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
40. F. Brochard, P.-G. de Gennes, 'Soret Effect of Flexible Macromolecules', *C.R. Acad. Sci., Ser. 2*, **293**, 1025–1027 (1981).
41. G. Meyerhoff, B. Rauch, 'Thermal Diffusion of Polystyrene in Various Solvents', *Makromol. Chem.*, **127**, 214–223 (1969).
42. F.S. Gaeta, G. Perna, G. Scala, 'Characterization of Macromolecules in Liquid Solutions by Thermal Diffusion. II. Dependence of the Soret Coefficient on the Molecular Mass of the Solute', *J. Polym. Sci., Polym. Phys. Ed.*, **13**, 203–222 (1975).
43. J.C. Giddings, K.D. Caldwell, 'Field-flow Fractionation', in *Physical Methods of Chemistry*, eds. B.W. Rossiter,

J.F. Hamilton, John Wiley & Sons, New York, Chapter 8, Vol. 3B, 867–938, 1989.
44. Y. Gao, X. Chen, 'Universal Calibration for Thermal Field-flow Fractionation', *J. Appl. Polym. Sci.*, **45**, 887–892 (1992).
45. J.C. Giddings, 'Universal Calibration in Size Exclusion Chromatography and Thermal Field-flow Fractionation', *Anal. Chem.*, **66**, 2783–2787 (1994).
46. W. Cao, P.S. Williams, M.N. Myers, J.C. Giddings, 'Thermal Field-flow Fractionation Universal Calibration: Extension for Consideration of Variation of Cold Wall Temperature', *Anal. Chem.*, **71**, 1597–1609 (1999).
47. J.C. Giddings, E. Kucera, C.P. Russell, M.N. Myers, 'Statistical Theory for the Equilibrium Distribution of Rigid Molecules in Inert Porous Network. Exclusion Chromatography', *J. Phys. Chem.*, **72**, 4397–4408 (1968).
48. M. Potschka, 'Universal Calibration of Gel Permeation Chromatography and Determination of Molecular Shape in Solution', *Anal. Biochem.*, **162**, 47–64 (1987).
49. S. Hussain, M.S. Metha, J.I. Kaplan, P.L. Dubin, 'Experimental Evaluation of Conflicting Models for Size Exclusion Chromatography', *Anal. Chem.*, **63**, 1132–1138 (1991).
50. P.L. Dubin, 'Problems in Aqueous Size Exclusion Chromatography', *Adv. Chromatogr.*, **31**, 119–151 (1992).
51. S. Mori, 'Secondary Effects in Aqueous Size Exclusion Chromatography of Sodium Poly(styrenesulfonate) Compounds', *Anal. Chem.*, **61**, 530–534 (1989).
52. P.L. Dubin, J.M. Principi, 'Failure of Universal Calibration for Size Exclusion Chromatography of Rodlike Macromolecules versus Random Coils and Globular Proteins', *Macromolecules*, **22**, 1891–1896 (1989).
53. M.T. Nguyen, R. Beckett, 'Calibration of Thermal Field-flow Fractionation Using Broad Molecular Weight Standards', *Polym. Int.*, **30**, 337–343 (1993).
54. T. Nguyen, R. Beckett, 'Calibration Methods for Field-flow Fractionation Using Broad Standards. I. Thermal Field-flow Fractionation', *Sep. Sci. Technol.*, **31**, 291–317 (1996).
55. P.J. Wyatt, 'Multiangle Light Scattering Combined with HPLC', *LC–GC*, **15**, 160–166 (1997).
56. M. Hassellöv, G. Hulthe, B. Lyvén, G. Stenhagen, 'Electrospray Mass Spectrometry as Online Detector for Low Molecular Weight Polymer Separations with Flow Field-flow Fractionation', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2843–2856 (1997).
57. M.E. Schimpf, M.N. Myers, J.C. Giddings, 'Measurements of Polydispersity of Ultra-narrow Polymer Fractions by Thermal Field-flow Fractionation', *J. Appl. Polym. Sci.*, **33**, 117–135 (1987).
58. J.C. Giddings, 'Field-flow Fractionation of Polymers: One Phase Chromatography', *Pure Appl. Chem.*, **51**, 1459–1471 (1979).
59. J. Marra, H.A. van der Schee, G.J. Fleer, J. Lyklema, 'Polyelectrolytes Adsorption from Saline Solutions', in *Adsorption Solution Symposium.*, eds. R.H. Ottewill, C.H. Rochester, A.L. Smith, Academic Press, London, 245–258, 1983.
60. J.C. Giddings, K.D. Caldwell, L.F. Kesner, 'Molecular Weight Distribution from Field-flow Fractionation', in *Molecular Weight Determination*, ed. A.R. Cooper, Chemical Analysis Series, Vol. 103, Wiley-Interscience, New York, Chapter 12, 1989.
61. J.J. Gunderson, J.C. Giddings, 'Comparison of Polymer Resolution in Thermal Field-flow Fractionation and Size Exclusion Chromatography', *Anal. Chim. Acta*, **189**, 1–15 (1986).
62. M. Doi, S.F. Edwards, *The Theory of Polymer Dynamics*, International Series of Monographs on Physics, Vol. 73, Clarendon Press, Oxford, 1986.
63. S.L. Brimhall, M.N. Myers, K.D. Caldwell, J.C. Giddings, 'Separation of Polymers by Flow Field-flow Fractionation', *J. Polym. Sci., Polym. Lett. Ed.*, **22**, 339–345 (1984).
64. K-G. Wahlund, H.S. Winegarner, K.D. Caldwell, J.C. Giddings, 'Improved Flow Field-flow Fractionation System Applied to Water-soluble Polymers: Programming, Outlet Stream Splitting and Flow Optimization', *Anal. Chem.*, **58**, 573–578 (1986).
65. M.-A. Benincasa, J.C. Giddings, 'Separation and Molecular Weight Distribution of Anionic and Cationic Water-soluble Polymers by Flow Field-flow Fractionation', *Anal. Chem.*, **64**, 790–798 (1992).
66. J.C. Giddings, M.-A. Benincasa, M.-K. Liu, P. Li, 'Separation of Water Soluble Synthetic and Biological Macromolecules by Flow Field-flow Fractionation', *J. Liq. Chromatogr.*, **15**, 1729–1747 (1992).
67. J.E.G.J. Wijnhoven, J.P. Koorn, H. Poppe, W.Th. Kok, 'Hollow-fiber Flow Field-flow Fractionation of Polystyrene Sulfonates', *J. Chromatogr. A*, **699**, 119–129 (1995).
68. M.-A. Benincasa, J.C. Giddings, 'Separation and Characterization of Cationic, Anionic, and Nonionic Water-soluble Polymers by Flow FFF: Sample Recovery, Overloading, and Ionic Strength Effects', *J. Micro. Sep.*, **9**, 479–495 (1997).
69. M.-A. Benincasa, K.D. Caldwell 'Flow FFF of Poly(ethylene Oxide). Effect of Carrier Ionic Strength and Composition', *Macromolecules*, in preparation.
70. M.N. Myers, 'Considerations in Thermal FFF', lecture presented at the 7th International Symposium on Field-flow Fractionation, Salt Lake City, UT, 8–11 February 1998.
71. K.-H. Cho, Y.H. Park, S.J. Jeon, W.-S. Kim, D.W. Lee, 'Retention Behavior of Copolymers in Thermal Field-flow Fractionation and Gel Permeation Chromatography', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2741–2756 (1997).
72. K.-G. Wahlund, J.C. Giddings, 'Properties of an Asymmetric Flow Field-flow Fractionation Channel Having One Permeable Wall', *Anal. Chem.*, **59**, 1332–1339 (1987).

73. A. Litzén, K.-G. Wahlund, 'Zone Broadening and Dilution in Rectangular and Trapezoidal Asymmetrical Flow Field-flow Fractionation Channels', *Anal. Chem.*, **63**, 1001–1007 (1991).
74. P.S. Williams, 'Design of an Asymmetrical Flow Field-flow Fractionation Channel for Uniform Channel Velocity', *J. Microcol. Sep.*, **9**, 459–467 (1997).
75. J.Å. Jönsson, A. Carlshaf, 'Flow Field-flow Fractionation in Hollow Cylindrical Fibers', *Anal. Chem.*, **61**, 11–18 (1989).
76. A.A. Kozinski, F.P. Schmidt, E.N. Lightfoot, 'Velocity Profiles in Porous-walled Ducts', *Ind. Eng. Chem. Fundam.*, **9**, 502–505 (1970).
77. T. Schauer, L. Dulog, 'Field-flow Fractionation', *GIT Fachz. Lab.*, **37**, 744–745 (1993).
78. J.C. Giddings, F.J. Yang, M.N. Myers, 'Flow Field-flow Fractionation as a Methodology for Protein Separation and Characterization', *Anal. Biochem.*, **81**, 395–407 (1977).
79. R. Beckett, Z. Jue, J.C. Giddings, 'Determination of Molecular Weight Distribution of Fulvic and Humic Acids Using Flow Field-flow Fractionation', *Environ. Sci. Technol.*, **21**, 289–295 (1987).
80. J. Li, K.D. Caldwell, 'Improved Accuracy in the Determination of Field-flow Fractionation Elution Volumes', *J. Chromatogr.*, **555**, 260–266 (1991).
81. K.D. Caldwell, S.L. Brimhall, Y. Gao, J.C. Giddings, 'Sample Overloading Effects in Polymer Characterization by Field-flow Fractionation', *J. Appl. Polym. Sci.*, **36**, 703–719 (1988).
82. S. Lee, 'Gel-content Determination of Polymers Using Thermal Field-flow Fractionation', *ACS Symp. Ser.*, **521**, 77–88 (1993).
83. S.J. Jeon, D.W. Lee, 'Thermal Diffusion and Molecular Weight Calibration of Poly(Ethylene-*co*-Vinyl Acetate)s by Thermal Field-flow Fractionation', *J. Polym. Sci.: Part B: Polym. Phys.*, **33**, 411–416 (1995).
84. J.J. Kirkland, S.W. Rementer, W.W. Yau, 'Polymer Characterization by Thermal Field-flow Fractionation with a Continuous Viscosity Detector', *J. Appl. Polym. Sci.*, **38**, 1383–1395 (1989).
85. F.-S. Yang, K.D. Caldwell, J.C. Giddings, 'Colloid Characterization by Sedimentation Field-flow Fractionation. II. Particle Size Distribution', *J. Colloid Interface Sci.*, **92**, 81–91 (1983).
86. M. Kerker, *The Scattering of Light*, Academic Press, New York, 1969.
87. S.K.R. Williams, R.G. Keil, 'Monitoring the Biological and Physical Reactivity of Dextran Carbohydrates in Seawater Incubations Using Field-flow Fractionation', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2815–2834 (1997).
88. H. Thielking, W.-M. Kulicke, 'On-line Coupling of Flow Field-flow Fractionation and Multiangle Laser Light Scattering for the Characterization of Macromolecules in Aqueous Solution as Illustrated by Sulfonated Polystyrene Samples', *Anal. Chem.*, **68**, 1169–1173 (1996).
89. D. Roessner, W.-M. Kulicke, 'On-line Coupling of Flow Field-flow Fractionation and Multiangle Laser Light Scattering', *J. Chromatogr. A*, **687**, 249–258 (1994).
90. M. Antonietti, A. Briel, C. Tank, 'Chromatographic Characterization of Complex Polymer Systems with Thermal Field-flow Fractionation', *Acta Polym.*, **46**, 254–260 (1995).
91. J. Lou, M.N. Myers, J.C. Giddings, 'Separation of Polysaccharides by Thermal Field-flow Fractionation', *J. Liq. Chromatogr.*, **17**, 3239–3260 (1994).
92. B. Wittgren, K.-G. Wahlund, 'Effects of Flow-rates and Sample Concentration on the Molar Mass Characterization of Modified Cellulose Using Asymmetrical Flow Field-flow Fractionation–Multiangle Light Scattering', *J. Chromatogr. A*, **791**, 135–149 (1997).
93. K.S. Schmitz, *Dynamic Light Scattering by Macromolecules*, Academic Press, New York, 1990.
94. J.C. Giddings, G. Karaiskakis, K.D. Caldwell, 'Concentration and Analysis of Dilute Colloidal Samples by Sedimentation FFF', *Sep. Sci. Technol.*, **16**, 725–744 (1981).
95. M. Adam, M. Delsanti, 'Dynamical Properties of Polymers Solutions in Good Solvents by Raleigh Scattering Experiments', *Macromolecules*, **10**, 1229 (1977).
96. K. Kubota, K.M. Abbey, B. Chu, 'Static and Dynamical Properties of Polymer Solution with Upper and Lower Critical Solution Points. NBS 705 Polystyrene in Methylacetate', *Macromolecules*, **16**, 137–143 (1983).
97. T.L. Yu, H. Reihanian, J.G. Southwick, A.M. Jamieson, 'Effect of Chain Overlap on Translational Diffusion of Polystyrene in Tetrahydrofuran', *J. Macromol. Sci. Phys.*, **B18**, 771–785 (1981).
98. W.W. Graessley, 'Polymer Chain Dimensions and the Dependence of Viscoelastic Properties on Concentration, Molecular Weight and Solvent Power', *Polymer*, **21**, 258–262 (1980).
99. J.Y. Ostashevsky, C.S. Lange, 'The Effect of Solvent Viscosity and Temperature on DNA Viscoelastic Behavior', *Biopolymers*, **25**, 291–306 (1986).
100. J.C. Giddings, 'Hydrodynamic Relaxation and Sample Concentration in Field-flow Fractionation Using Permeable Wall Elements', *Anal. Chem.*, **62**, 2306–2312 (1990).
101. P. Li, M. Hansen, J.C. Giddings, 'Separation of Lipoproteins from Human Plasma by Flow Field-flow Fractionation', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2777–2802 (1997).
102. J.C. Giddings, 'A Pinched Inlet System for Reduced Relaxation Effects and Stopless Flow Injection in Field-flow Fractionation', *Sep. Sci. Technol.*, **24**, 755–768 (1989).
103. M.H. Moon, M.N. Myers, J.C. Giddings, 'Evaluation of Pinched Inlet Channel for Stopless Flow Injection in

Steric Field-flow Fractionation', *J. Chromatogr.*, **517**, 423–433 (1990).
104. J.C. Giddings, 'Optimized Field-flow Fractionation System Based on Dual Stream Splitters', *Anal. Chem.*, **57**, 945–947 (1985).
105. S. Lee, M.N. Myers, J.C. Giddings, 'Hydrodynamic Relaxation Using Stopless Flow Injection in Split Inlet Sedimentation Field-flow Fractionation', *Anal. Chem.*, **61**, 2439–2444 (1989).
106. M.-K. Liu, P.S. Williams, M.N. Myers, J.C. Giddings, 'Hydrodynamic Relaxation Using Both Split and Frit Inlets', *Anal. Chem.*, **63**, 2115–2122 (1991).
107. H. Lee, S.K.R. Williams, J.C. Giddings, 'Particle Size Analysis of Dilute Environmental Colloids by Flow Field-flow Fractionation Using an Opposed Flow Sample Concentration Technique', *Anal. Chem.*, **70**, 2495–2503 (1998).
108. F.J.F. Yang, M.N. Myers, J.C. Giddings, 'Programmed Sedimentation Field-flow Fractionation', *Anal. Chem.*, **46**, 1924–1930 (1974).
109. J.C. Giddings, P.S. Williams, R. Beckett, 'Fractionating Power in Programmed Field-flow Fractionation: Exponential Sedimentation Field Decay', *Anal. Chem.*, **59**, 28–37 (1987).
110. J.C. Giddings, L.K. Smith, M.N. Myers 'Programmed Thermal Field-flow Fractionation', *Anal. Chem.*, **48**, 1587–1592 (1976).
111. P.S. Williams, J.C. Giddings, R. Beckett, 'Fractionating Power in Sedimentation Thermal Field-flow Fractionation with Linear and Parabolic Field Decay Programming', *J. Liq. Chromatogr.*, **10**, 1961–1998 (1987).
112. W.W. Yau, J.J. Kirkland, 'Retention Characteristics of Time-delayed Exponential Field-programmed Sedimentation Field-flow Fractionation', *Sep. Sci. Technol.*, **16**, 577–605 (1981).
113. J.J. Kirkland, S.W. Rementer, W.W. Yau, 'Time-delayed Exponential Field-programmed Sedimentation Field-flow Fractionation for Particle Distribution Analyses', *Anal. Chem.*, **53**, 1730–1736 (1981).
114. J.J. Kirkland, C.H. Dilks, Jr, 'Flow Field-flow Fractionation of Polymers in Organic Solvents', *Anal. Chem.*, **64**, 2836–2840 (1992).
115. J.J. Kirkland, W.W. Yau, 'Thermal Field-flow Fractionation of Polymers with Exponential Temperature Programming', *Macromolecules*, **18**, 2305–2311 (1985).
116. J.J. Kirkland, S.W. Rementer, W.W. Yau, 'Molecular-weight Distribution of Polymers by Thermal Field-flow Fractionation with Exponential Temperature Programming' *Anal. Chem.*, **60**, 610–616 (1988).
117. J.J. Kirkland, W.W. Yau, 'Thermal Field-flow Fractionation of Water-soluble Macromolecules', *J. Chromatogr.*, **353**, 95–107 (1986).
118. P.S. Williams, J.C. Giddings, 'Power Programmed Field-flow Fractionation: a New Program Form for Improved Uniformity of Fractionating Power', *Anal. Chem.*, **59**, 2038–2044 (1987).
119. J.C. Giddings, V. Kumar, P.S. Williams, M.N. Myers, 'Polymer Separation by Thermal Field-flow Fractionation: High Speed Power Programming', *Adv. Chem. Ser.*, **227**, 3–21 (1990).
120. M.E. Schimpf, J.C. Giddings, 'Characterization of Thermal Diffusion of Copolymers in Solution by Thermal Field-flow Fractionation', *J. Polym. Sci., Part B: Polym. Phys.*, **28**, 2673–2680 (1990).
121. A.C. van Asten, E. Venema, W.Th. Kok, H. Poppe, 'Use of Thermal Field-flow Fractionation for the Fractionation of Polybutadiene in Various Organic Solvents', *J. Chromatogr.*, **644**, 83–94 (1993).
122. A.C. van Asten, R.J. van Dam, W.Th. Kok, R. Tijssen, H. Poppe, 'Determination of the Compositional Heterogeneity of Polydisperse Polymer Samples by the Coupling of Size-exclusion Chromatography and Thermal Field-flow Fractionation', *J. Chromatogr. A*, **703**, 245–263 (1995).
123. S. Jeon, M.E. Schimpf, 'Cross-fractionation of Copolymers Using SEC and Thermal FFF for Determination of Molecular Weight and Composition' in *Chromatography of Polymers: Hyphenated and Multi-dimensional Techniques*, ed. T. Provder, ACS Symposium Series 731, ACS Washington, DC, 141–161, 2000.
124. M.E. Schimpf, J.C. Giddings, 'Characterization of Thermal Diffusion in Polymer Solutions by Thermal Field-flow Fractionation: Dependence on Polymer and Solvent Parameters', *J. Polym. Sci., Polym. Phys. Ed.*, **27**, 1317–1332 (1989).
125. A.C. van Asten, W.Th. Kok, R. Tijssen, H. Poppe, 'Study of the Thermal Diffusion of Polybutadiene and Polytetrahydrofuran in Various Organic Solvents', *J. Polym. Sci., Part B*, **34**, 297–308 (1996).
126. E. Venema, P. de Leeuw, J.C. Kraak, H. Poppe, R. Tijssen, 'Polymer Characterization Using On-line Coupling of Thermal Field-flow Fractionation and Hydrodynamic Chromatography', *J. Chromatogr. A*, **765**, 135–144 (1997).
127. S. Lee, A. Molnar, 'Determination of Molecular Weight and Gel Content of Natural Rubber Using Thermal Field-flow Fractionation', *Macromolecules*, **28**, 6354–6356 (1995).
128. P.M. Shiundu, E.E. Remsen, J.C. Giddings, 'Isolation and Characterization of Polymeric and Particulate Components of Acrylonitrile–Butadiene–Styrene (ABS) Plastics by Thermal Field-flow Fractionation', *J. Appl. Polym. Sci.*, **60**, 1695–1707 (1996).
129. A. Subramanian, 'Gel Permeation Chromatography of Natural Rubber', *Rubber Chem. Technol.*, **45**, 346–358 (1972).
130. D. McIntyre, A.L. Shis, J. Saroca, R. Seeger, A. MacArthur, 'Shear Degradation of Very High Molecular Weight Polymers in Gel Permeation Chromatography', *ACS Symp. Ser.*, **245**, 227–240 (1984).
131. S.K. Ratanathanawongs, P.M. Shiundu, J.C. Giddings, 'Size and Compositional Studies of Core-shell Latexes

Using Flow and Thermal Field-flow Fractionation', *Colloids Surf. A: Physicochem. Eng. Aspects*, **105**, 243–250 (1995).
132. P.M. Shiundu, J.C. Giddings, 'Influence of Bulk and Surface Composition on the Retention of Colloidal Particles in Thermal Field-flow Fractionation', *J. Chromatogr.*, **715**, 117–126 (1995).
133. S.J. Jeon, M.E. Schimpf, A. Nyborg, 'Compositional Effects in the Retention of Colloids by Thermal Field-flow Fractionation', *Anal. Chem.*, **69**, 3442–3450 (1997).
134. J.C. Giddings, M.N. Myers, G.C. Lin, M. Martin, 'Polymer Analysis and Characterization by Field-flow Fractionation (One Phase Chromatography)', *J. Chromatogr.*, **142**, 23–38 (1977).
135. J.C. Giddings, G.C. Lin, M.N. Myers, 'Fractionation and Size Distribution of Water Soluble Polymers by Flow Field-flow Fractionation', *J. Liq. Chromatogr.*, **1**, 1–20 (1978).
136. E.A. Bekturov, Z.Kh. Bakauova, *Synthetic Water-soluble Polymers in Solution*, Hüthig & Wepf, Basle, 1986.
137. M.-A. Benincasa, 'Synthetic Polymers – Water Soluble', in *FFF Handbook*, eds. M.E. Schimpf, K.D. Caldwell, J.C. Giddings, Wiley Scientific, New York, Chapter III-6, 2000.
138. S. Yanagida, K. Takahashi, M. Okahara, 'Metal-ion Complexation of Noncyclic Poly(oxyethylene) Derivatives. I. Solvent Extraction of Alkali and Alkaline Earth Metal Thiocyanates and Iodides', *Bull. Chem. Soc. Jpn.*, **50**, 1386–1390 (1977).
139. B. Wittgren, K.-G. Wahlund, H. Dérand, B. Wesslén, 'Aggregation Behavior of an Amphiphilic Graft Copolymer in Aqueous Medium Studied by Asymmetrical Flow Field-flow Fractionation', *Macromolecules*, **29**, 268–276 (1996).
140. B. Wittgren, K.-G. Wahlund, H. Dérand, B. Wesslén, 'Size Characterization of a Charged Amphiphilic Copolymer in Solution of Different Salts and Salt Concentrations Using Flow Field-flow Fractionation', *Langmuir*, **12**, 5999–6005 (1996).
141. R. Hanselmann, W. Burchard, M. Ehrat, H.M. Widmer, 'Structural Properties of Fractionated Starch Polymers and Their Dependence on the Dissolution Process', *Macromolecules*, **29**, 3277–3282 (1996).
142. J.C. Giddings, P.S. Williams, M.-A. Benincasa, 'Rapid Breakthrough Measurement of Void Volume for Field-flow Fractionation Channels', *J. Chromatogr.*, **627**, 23–35 (1992).
143. J.M. Davis, 'Influence of Crossflow Hydrodynamics on Retention Ratio in Flow Field-flow Fractionation', *Anal. Chim. Acta*, **246**, 161–169 (1991).
144. J.C. Giddings, 'Micro-FFF: Theoretical and Practical Aspects of Reducing the Dimensions of Field-flow Fractionation Channels', *J. Microcol. Sep.*, **5**, 497–503 (1993).
145. A. Litzén, 'Separation, Speed, Retention, and Dispersion in Asymmetrical Flow Field-flow Fractionation as Functions of Channel Dimensions and Flow Rates', *Anal. Chem.*, **65**, 461–470 (1993).
146. M.N. Myers, W. Cao, C.I. Chien, V. Kumar, J.C. Giddings, 'Cold Wall Temperature Effects on Thermal Field-flow Fractionation', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2757–2776 (1997).
147. R. Sisson, J.C. Giddings, 'Effects of Solvent Composition on Polymer Retention in Thermal Field-flow Fractionation: Retention Enhancement in Binary Solvent Mixtures', *Anal. Chem.*, **66**, 4043–4053 (1994).
148. M.E. Hansen, J.C. Giddings, R. Beckett, 'Colloid Characterization by Sedimentation FFF. VI. Perturbations due to Overloading and Electrostatic Repulsion', *J. Colloid. Interface Sci.*, **132**, 300–312 (1989).
149. J. Janca, M. Martin, 'Influence of Operational Parameters on Retention of Ultra-high Molecular Weight Polystyrenes in Thermal Field-flow Fractionation', *Chromatographia*, **34**, 125–131 (1992).
150. A.C. van Asten, W.Th. Kok, R. Tijssen, H. Poppe, 'Thermal Field-flow Fractionation of Polytetrahydrofuran', *J. Chromatogr. A*, **676**, 361–373 (1994).
151. A. Litzén, K.-G. Wahlund, 'Effects of Temperature, Carrier Composition and Sample Load in Asymmetrical Flow Field-flow Fractionation', *J. Chromatogr.*, **548**, 393–406 (1991).
152. A. Carlshaf, J.Å. Jönsson, 'Perturbations of the Retention Parameter Due to Sample Overloading in Hollow-fiber Flow Field-flow Fractionation', *Sep. Sci. Technol.*, **28**, 1191–1201 (1993).
153. J.E.G.J. Wijnhoven, J.-P. Koorn, H. Poppe, W.Th. Kok, 'Influence of Injected Mass and Ionic Strength on Retention of Water-soluble Polymers and Proteins in Hollow-fiber Field-flow Fractionation', *J. Chromatogr. A*, **732**, 307–315 (1996).
154. M. Hoyos, M. Martin, 'Retention Theory of Sedimentation Field-flow Fractionation at Finite Concentrations', *Anal. Chem.*, **66**, 1718–1730 (1994).
155. J.E.G.J. Wijnhoven, M.R. van Bommel, H. Poppe, W.Th. Kok, 'Practical Experience with Organic Solvent Flow Field-flow Fractionation', *Chromatographia*, **42**, 409–415 (1996).
156. M.-A. Benincasa, unpublished results.
157. S.K.R. Williams, J.C. Giddings 'Sample Recovery', in *FFF Handbook*, eds. M.E. Schimpf, K.D. Caldwell, J.C. Giddings, Wiley Scientific, New York, Chapter II-11, 2000.
158. S.K.R. Williams, 'Flow FFF', in *FFF Handbook*, eds. M.E. Schimpf, K.D. Caldwell, J.C. Giddings, Wiley Scientific, New York, Chapter II-7, 2000.

Gas Chromatography in Analysis of Polymers and Rubbers

M. Hakkarainen and S. Karlsson
The Royal Institute of Technology (KTH), Stockholm, Sweden

1 Introduction **2**
2 Gas Chromatographic Analysis of Low-molecular-weight Compounds in Polymers **2**
2.1 Trace Levels of Monomers, Initiators and/or Catalysts 2
2.2 Additives and Degradation Products of Additives 2
2.3 Degradation Products 3
2.4 Aroma and Fragrance in Polymers 3
3 Sample Preparation–Extraction Techniques **4**
3.1 Precipitation 4
3.2 Soxhlet Extraction 4
3.3 Ultrasonication 4
3.4 Microwave-assisted Extraction 4
3.5 Liquid–Liquid Extraction 4
3.6 Solid-phase Extraction 5
3.7 Headspace Gas Chromatography Analysis 5
3.8 Solid-phase Microextraction 6
3.9 Supercritical Fluid Extraction 7
4 Analysis of Low-molecular-weight Products in Polymers **8**
4.1 Synthesis-related Trace Compounds in Polymers 8
4.2 Headspace Analysis of Polymers in Indoor Applications 8
4.3 Thermal Degradation Products of Polymers at Processing Temperatures 9
4.4 Environmental Degradation Products of Polymers 9
4.5 Additive Systems in Plastics and Coatings 11
Abbreviations and Acronyms **12**
Related Articles **12**
References **13**

This chapter gives an overview of gas chromatography (with mass spectrometry) (GC/MS) of polymers and rubbers. Gas Chromatography (GC) analyzes volatile organic compounds, with an upper limit of 350 °C, which means that the compounds to be analyzed must be volatile below this temperature. The technique is able to analyze small quantities of material, which means that it is applicable for example to residual monomers, initiators, catalysts, some additives and degradation products of polymers. It is generally not suitable for analysis of organic compounds at high molecular weight or of low volatility. Care must be taken not to analyze reactive species, which may ruin columns or other parts of the equipment. Proper sample preparation is necessary before GC/MS. Sensitive and selective techniques are used to separate and extract low-molecular-weight organic compounds from polymers. The sample preparation–extraction techniques may be grouped into (1) solvent extraction from solid matrices, (2) solvent extraction of organic compounds from aqueous solutions containing polymer (e.g. biomedical implants in physiological buffers) and (3) solvent-free extraction methods. An ideal extraction method is quantitative, selective, rapid and uses little or no solvent.

Soxhlet is the old and traditional method for solvent extraction from solid sample matrices. Soxhlet is, however, time-consuming (two or three days is not uncommon), nonselective, uses large volumes of solvents and is often not quantitative. Ultrasonication and microwave-assisted extraction (MAE) are instead much more effective. Ultrasonication works by agitating the solution and producing cavitation in the liquid. The technique is useful for example to extract antioxidants from polyethylene (PE). MAE, extracts (semi)volatiles from solid matrices and has been successfully used to extract additives from polyolefins, aroma and flavor compounds from recycled polymers, and oligomers from poly(ethylene terephthalate) (PET). Solvent extractions from aqueous solutions are liquid–liquid extraction (LLE) and solid-phase extraction (SPE). LLE is rapid, but lacks in selectivity, is labor intensive and uses large volumes of organic solvents. SPE is instead suitable for separating volatile and semivolatile compounds and is a physical extraction process involving liquid and a solid phase (sorbent). Examples of separations are degradation products of PE and hydroxyacids in buffer solutions.

Solvent-free extraction methods are headspace gas chromatography (HS/GC), solid-phase microextraction (SPME) and supercritical fluid extraction (SFE). HS/GC determines volatile compounds in liquids and solids. SPME is an inexpensive, rapid and solvent-free technique with applications reported for air samples, water and soil, based on 1-cm long, thin fused silica fiber coated with a polymeric stationary phase mounted in a modified syringe.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The stationary phase is available in four different kinds. SFE uses a supercritical fluid to penetrate a material. Applications are antioxidants in polyolefins, aroma vapors absorbed in PE and surface coatings and their raw materials.

This article reports applications in synthesis-related compounds in polymers, HS (headspace) analysis of polymers in indoor environments, thermal degradation products of polymers at processing temperatures, environmental degradation products, and additive systems in plastics and coatings.

1 INTRODUCTION

GC is the method of choice for the analysis of volatile organic compounds. The upper temperature limit is usually not greater than 350 °C, which means that the compounds to be analyzed must be volatile below this temperature. One of the great advantages of the gas chromatograph is its ability to detect and analyze rapidly extremely small quantities of material. Frequently the polymer chemist is concerned with trace amounts of residual monomer. If solvent processing is employed in polymer production residual solvent as well as monomer are of interest. Impurities present in the monomer may also be detectable at levels comparable with those of monomer in the finished polymer. Reliable quantitative analytical methods for these materials are necessary because they can affect such polymer properties as strength, dimensional stability, vacuum forming and molding characteristics and chemical resistance. Where polymer films are used in packaging food, residual monomer and solvent can affect the taste or odor of the food. Apart from the above-mentioned undesirable elements, certain volatile compounds must frequently be added to improve the properties of polymeric materials. They may be added at low concentrations, for example antioxidants or ultraviolet (UV) absorbers, or at high concentrations, as plasticizers or solvents for paints. For many years gas chromatographic analysis of residual monomer has also been applied to kinetic studies or in determining reaction mechanisms and reactivity ratios during copolymerization.

During the processing, lifetime and after its disposal, low-molecular-weight products are formed in the material due to degradation of the polymer matrix. The changes in the long-term properties of polymers are usually manifested as a decrease or increase in mechanical properties, weight loss, molecular weight or some other property relying on the macromolecular nature of the substrate. In addition to the changes in polymer matrix short polymer fragments and low-molecular-weight degradation products are released. Analysis of these products gives information about the degradation mechanisms and environmental impact of the material. The product analysis may also be used in lifetime predictions. Sensitive and selective techniques must be used to isolate and identify the low-molecular-weight products in polymeric materials and/or the environment where polymers are found. An appropriate extraction method combined with gas chromatographic analysis are tools for identification of wide range of low-molecular-weight products.

2 GAS CHROMATOGRAPHIC ANALYSIS OF LOW-MOLECULAR-WEIGHT COMPOUNDS IN POLYMERS

2.1 Trace Levels of Monomers, Initiators and/or Catalysts

There are in most cases many synthesis-related low-molecular-weight impurities in the polymeric materials. These include monomer rests, oligomers, cross-linking residues, initiators or catalysts, solvent rests and so on. The detection of residual volatiles can yield clues about manufacturing processes; similar materials from different manufacturers usually contain different volatiles. The solvents observed are frequently those used in polymer synthesis or processing. Other impurities can aid in the identification of specific manufacturing processes. Problems such as surface crazing, bubble formation and chemical degradation of the polymer can sometimes be caused by residues. Even if the trace amounts would not affect the performance of the end product, they may act as odorants or irritants in the material. Numerous small molecules (e.g. benzene) are also regulated because of their toxic nature.

2.2 Additives and Degradation Products of Additives

Commercial plastics and rubbers are complex materials. In addition to various polymers, commercial formulations always contain a number of compounding ingredients (additives) that are included to give particular physical and/or chemical properties. These additives include plasticizers, extender oils, carbon black, inorganic fillers, antioxidants, antiozonants, antifatigue agents, heat and light stabilizers, tackifying resins, processing aids, cross-linking agents, accelerators, retarders, adhesives, pigments, smoke and flame retardants, and others. A knowledge of additives is necessary for a number of applications, e.g. the verification of ingredients, reconstruction of recipes in unknown materials, investigation of manufacturing problems, identification of odorants or irritants that evolve from polymeric materials and so on. Because of the environmental concern about

use of chemicals, the total composition of polymeric materials, including additives has become an important issue. Many additives and their degradation products are hazardous and their migration to the surrounding environment is a point of concern. In addition to the environmental aspect, following the fate of the additives in the material gives information about the remaining lifetime of the material. The fate of the additives may be followed by quantification and identification of the products that migrate out of the material. In an opposite approach the additives remaining in the material may be measured.

Most modern polymer additives are compounds of insufficient volatility to be analyzed by GC or they are thermally unstable and decompose under the conditions of gas chromatographic separation. For additives with sufficient volatility, e.g. many alcohols, fatty acids and their esters, hydrocarbons, phthalate plasticizers and lower-molecular-weight aromatic amines and hindered phenols, GC has been set up as a standard technique of analysis.[1] The use of short columns, low-stationary phase loadings and high-temperature packing materials will extend the applicability of GC/MS to some higher-molecular-weight materials. Another approach is to convert nonvolatile components into more volatile derivatives such as trimethylsilyl ethers. However, in many cases, owing to the low volatility or thermal instability, liquid chromatography is a preferred choice for the analysis of additives.[2]

2.3 Degradation Products

The identification of degradation products is a prerequisite for the determination of degradation mechanisms of polymers. In many cases the degradation of polymers results in complex mixtures of degradation products varying in chemical composition and molecular weight, e.g. oxidation of PE may result in hundreds of different products. In contrast, polymers with weak linkages, where specific degradation, like hydrolysis occurs, generally give rise to a few well-defined degradation products. Degradation products can be formed during synthesis, processing, use and after disposal. The processing of plastics takes place at relatively high temperatures up to 300 °C. At these temperatures volatile products are formed and released from most plastics. Some of these compounds can already be dangerous to health at low concentrations. The best way to protect the employees is to identify and quantify these compounds and to make toxicological studies to obtain information about the work environment.

Environmental concerns have promoted the development of degradable plastics. To understand the environmental impact of such polymers fully, their degradation products need to be established. The nontoxicity of the degradation products is important in packaging materials, but even more important when materials are used for medical purposes. The foreign body reaction of medical implants is usually caused by leakage of degradation products or other low-molecular-weight products. The identification of degradation products in vitro is a prerequisite in determining the interaction of degradable polymers in the human body. A basic requirement for use in medical applications is the biocompatibility of both the polymer and its degradation products.

2.4 Aroma and Fragrance in Polymers

Plastic materials are particularly vulnerable to absorbing detectable amounts of contaminants because of their permeable nature. The migration of these compounds out of packaging material into the package filling must not have a detrimental effect on the quality of the filling, therefore recycled materials used for food packaging, for example, require some special consideration to ensure that nonregulated chemicals or contaminants either are not present in the material or do not migrate into food. In the case of plastic packaging waste, wasted packages might have come in contact with substances other than food, e.g. flavor components and other harmful contaminants which diffuse into the package material. There have been several studies on the aroma transport from food to polymeric packaging material.[3,4] In the life of the package, these substances may migrate from the recycled packaging material into the food. This absorption/desorption process observed in plastics makes it difficult to recycle them directly into new packaging.

In addition to typical PE degradation products, e.g. carboxylic acids, ketones, aldehydes and hydrocarbons, aroma and fragrance compounds and amines have been identified in recycled PE.[5] Some of the aroma and fragrance compounds identified are widely used in cosmetics, e.g. isopropyl myristate, isopropyl palmitate. These substances are used in products for personal hygiene, cleaning agents, food and beverages and they are probably absorbed by the packaging material during storage. Aroma and odor compounds such as limonene and 3-carene and some alcohols were detected in recycled high-density polyethylene (HDPE). Alcohols are used as basic compounds in cosmetic products and in manufacturing of household and industrial cleaning agents. In another recycling study, virgin and recycled HDPE, polypropylene (PP), polystyrene (PS) and PET were compared.[6] Generally only a few contaminants were found. Most of the contaminants were aroma compounds and were linked to foodstuffs contained in the package during the first usage.

3 SAMPLE PREPARATION–EXTRACTION TECHNIQUES

The nature of the matrix and the products to be extracted determine the choice of extraction method. Generally different extraction methods are divided into methods that use organic solvents and solvent-free extraction methods. An ideal extraction method is quantitative, selective, rapid and uses little or no solvent. Extracting products directly from the solid matrix is possible directly with an organic solvent, by Soxhlet extraction, ultrasonication or by MAE. The volatile products from the air above the polymer material may be extracted by SPME or by HS/GC. Extraction from aqueous media is best done by LLE, SPE or SPME.

3.1 Precipitation

For the extraction of additives the polymer can first be dissolved completely in a suitable solvent. After dissolution a second solvent is added to precipitate the polymer. Common organic stabilizers, plasticizers and lubricants remain in solution for analysis.[7]

3.2 Soxhlet Extraction

Soxhlet extraction is the traditional method of extracting low-molecular-weight products from solid sample matrices such as polymers[8] or soil. The sample is put in a porous holder or thimble and bathed with fresh hot solvent. The insoluble matrice remains in the thimble, while the analytes are transferred to the solvent reservoir. The extraction solvent is again vaporized, condensed and allowed to percolate through the solid sample matrice. This method is time-consuming (two or three days is not uncommon), nonselective, uses large volumes of solvent and is often not quantitative. Some analytes may be damaged by prolonged exposure to hot solvent.

3.3 Ultrasonication

Ultrasonic extraction works principally by agitating the solution and producing cavitation in the liquid. This increases the rate of transfer across the polymer/liquid boundary layer but does not increase the diffusion of compounds within the polymer. This technique has been applied to extraction of additives from polyolefins with good results,[9] e.g. a hindered amine light stabilizer (HALS) Chimassorb 944, phenolic antioxidant Irganox 1010 and phosphite antioxidant Irgafos 168 were rapidly and quantitatively extracted from low-density polyethylene (LDPE) and medium-density polyethylene (MDPE) with chloroform at 60 °C.[10]

3.4 Microwave-assisted Extraction

MAE is a relatively new method for the extraction of (semi)volatiles from solid matrices. Chemical compounds absorb microwave energy in proportion to their dielectric constants: the higher the value, the higher the level of microwave energy absorption. When using MAE a microwave-absorbing extraction solvent with high dielectric constant is used. The sample and solvent are placed in a closed chemically resistant, nonmicrowave-absorbing pressure-resistant vessel. Microwave radiation can heat the solvent to a temperature higher than its boiling point, accelerating the diffusion through the extracted matrix providing rapid extraction of analyte under moderate pressure. The extraction solvent power for the analytes of choice is an important criterion. Parameters such as heating time, and pulsed heating versus continuous heating are varied to optimize extractions. The advantages compared to Soxhlet are reduced volumes of organic solvent (10 mL versus 300 mL) and a faster extraction time ranging from 10 min to 2 h compared to 24–48 h with Soxhlet extraction. Several samples can be extracted simultaneously increasing the sample throughput. MAE has been successfully used to extract additives from polyolefins,[11] aroma and flavor from recycled polymers,[5] oligomers from PET[12] and degradation products from enhanced degradable HDPE and PP blends aged in soil.[13]

3.5 Liquid–Liquid Extraction

LLE is the traditional method for extracting organic compounds from water.[14] The low-molecular-weight compounds are transferred from one liquid phase to another immiscible or partially immiscible liquid by shaking them in a separation funnel. There are also automated methods. The analyte will go into the phase where it forms a more stable chemical species and is extracted according to its distribution constant for the particular liquid–liquid system. No universal solvent has been found, but diethyl ether, pentane, hexane, isooctane, toluene, methylcyclohexane, chloroform, and dichloromethane are some of the most usual solvents. Most organic compounds are more soluble in organic solvents than in water, exceptions being low-molecular-weight compounds with one or more polar groups, which are solvated through hydrogen bonding. LLE is still a common method, but has several drawbacks such as low selectivity, labor intensivity and the use of a large amount of organic solvent. A concentration step is usually necessary after the extraction and extremely water-soluble substances are difficult to extract. We have used LLE to extract and fractionate degradation products of casein, a biopolymer used in building materials.[15,16] Amines as well as volatile and nonvolatile organic acids were identified by a subsequent GC analysis of the

extracts. The hydrolysis products of degradable polyesters such as poly(lactic acid) and its copolymers have also been extracted from the buffer solution by LLE.[17]

3.6 Solid-phase Extraction

SPE is suitable for separation of volatile and semivolatile substances from water or organic solvent.[18,19] It can also be used to separate complex product mixtures into several fractions, to simplify the GC analysis. The principle of SPE is analogous to LLE, but it gives highly selective extractions with minimum of solvent. SPE is a physical extraction process that involves a liquid and a solid phase. The selectivity of an extraction is a function of the chemical structure of the analyte, the properties of the sorbent and the composition of the sample matrix. The solid phase should have a greater attraction for the analyte than the solvent in which the analyte is dissolved. The liquid sample is passed over a solid or ''sorbent'' that is packed in a medical-grade PP cartridge. As a result of strong attractive forces between the analytes and the sorbent, the analytes are retained on the sorbent. Later, the sorbent is washed with a small volume of a solvent that has the ability to disrupt the bonds between the analytes and the sorbent. The final result is that the analytes are concentrated in a relatively small volume of clean solvent and are ready to be injected into a GC. The sorbent normally consists of a silica substrate bonded to an organosilane compound. The specific properties of a given bonded silica sorbent are a result of the functional group covalently bonded to the silica substrate. Table 1 presents common SPE columns for different matrices and analytes.

Three types of interaction are possible between the analyte and the sorbent material, namely nonpolar, polar and ionic interactions. Nonpolar interactions are those that occur between the carbon–hydrogen bonds of the sorbent functional groups and the carbon–hydrogen bonds of the analyte. In general nonpolar extractions are less selective than polar and ion-exchange extractions. Conversely, nonpolar interactions are very effective for isolating compounds that are dissimilar in structure and is often chosen when the object is to isolate simultaneously the maximum number of analytes with different chemical properties. Most organic molecules have structures which in some part are nonpolar so this mode of interaction is often utilized. Polar interactions include hydrogen bonding, dipole/dipole, induced dipole/dipole, π-π and a variety of other interactions in which the distribution of electrons between individual atoms in the functional groups in unequal, causing positive and negative polarity. This property allows an analyte molecule bearing a polar functional group to interact with a polar group on a sorbent. Examples of possible groups are hydroxys, amines, carbonyls, aromatic rings, double bonds and groups containing heteroatoms.

Table 1 Common SPE columns for different matrices and analytes

Non-polar Extraction
Sorbents: C_{18}(octadecyl), C_8(octyl), C_2(ethyl), cyclohexyl, phenyl, cyanopropyl
Matrix: water, buffers, biological fluids
Analytes: alkanes, alkenes, aromatics, cyclic hydrocarbons, ethers
Polar Extraction
Sorbents: cyanopropyl, diol, silica, aminopropyl
Matrix: nonpolar organic solvents, oils, lipids
Analytes: amines, glycols, alcohols, aldehydes, ketones, phenols
Cation-exchange Extraction
Sorbents: propylsulfonic acid, benzenesulfonic acid, carboxylic acid
Matrix: water, acidic buffers of low ionic strength, biological fluids
Analytes: amines, pyrimidines
Anion-exchange Extraction
Sorbents: quaternary amine, aminopropyl, diethylaminopropyl, primary/secondary amine
Matrix: water, alkaline buffers of low ionic strength, biological fluids
Analytes: carboxylic acids, sulfonates, phosphates, amino acids

The third class of interaction used in sorbent extraction are ionic interactions. They occur between an analyte molecule carrying a charge opposite to that of the sorbent. Groups on analytes and sorbents can be either cationic or anionic. The retention occurs by means of an ion-exchange mechanism. Owing to the limited number of exchange sites, smaller sample volumes can be handled than when employing other interactions. The ionic charge depends on the pH of the surrounding solvent environment and it must be at a pH where both the analyte and the sorbent are charged. Furthermore the solvent should not contain high concentrations of strongly competing ionic species of the same charge as the analyte. The total ionic strength should be low. Figure 1 presents an SPE extraction scheme for aqueous matrices containing complex product mixtures, such as PE degradation products.[20,21] Figure 2 shows an SPE scheme for the extraction of polyester hydrolysis products from a buffer solution.[22]

3.7 Headspace Gas Chromatography Analysis

Headspace gas chromatography/mass spectrometry (HS/GC/MS) is an indirect method for the determination of volatile constituents in liquids and solids. Beside

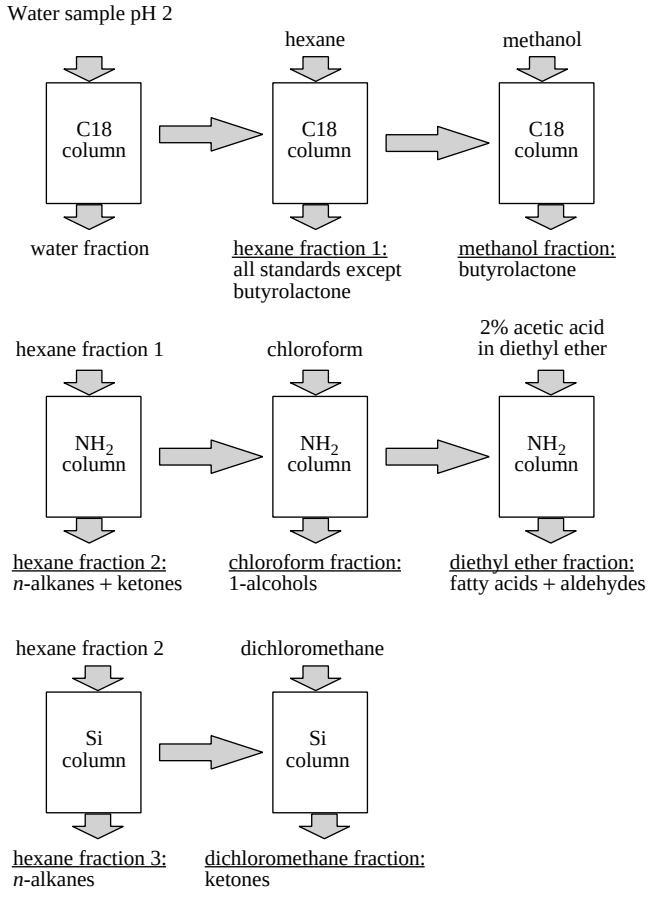

Figure 1 SPE extraction scheme for aqueous matrices with complex analyte mixtures.

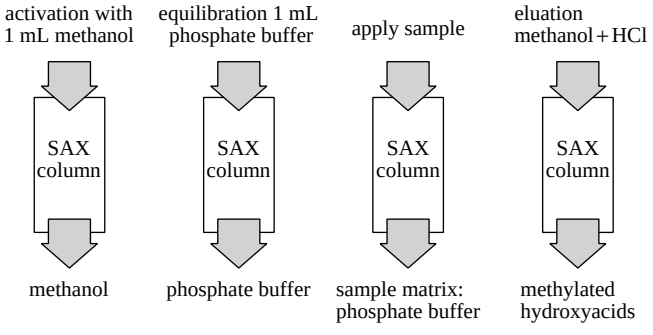

Figure 2 SPE scheme for the extraction of polyester hydrolysis products from buffer solution.

the analysis of volatile compounds in polymers, applications are found in analysis of sewage, water, oil, food and blood samples. The HS analysis can be divided into static and dynamic analysis. Static HS analysis indirectly determines the volatile constituents in liquid or solid samples by analyzing the vapor phase that is in thermodynamic equilibrium with the sample in a closed system.[23,24] The name static implies that there is no flow of either gas or liquid. The sample vial is first pressurized with carrier gas (helium is recommended) until its pressure is equal to the head pressure of the column. This generates a stable gas mixture. Then the carrier gas flow to the column is temporarily shut off, letting the gas flow from the sample vial to the column for a preset time.

At equilibrium the distribution of the individual compounds between the two phases will depend on their saturation pressures at the temperature of the vial and their concentration in the liquid phase. The distribution is highly sensitive to the temperature of the system. It follows that the concentration of the low-boiling sample components will be relatively much higher in the vapor phase than in the original liquid sample. HS analysis is very useful for determining volatile compounds in samples that are difficult to analyze by conventional chromatographic means, e.g. when the matrix is too complex or contains subtances that seriously interfere with the analysis or even damage the column, decomposition may occur on volatilization or dissociation products not originally present can be formed during the sampling procedure. The technique is very convenient, but has some drawbacks, e.g. solutes with very low concentrations or very low vapour pressure cannot be detected. The dynamic HS procedures involve sampling from the flowing gas above the liquid. On-column and splitless sampling are possible, thereby improving the detection limits.

The peak area for equilibrium HS/GC depends on sample volume and the partition coefficient of the compound of interest between the gas phase and matrix. This need to include the partition coefficient and thus the matrix in the calibration procedure causes serious problems with certain sample types, for which no calibration sample can be prepared. Such problems can be handled with multiple headspace extraction (MHE),[25] which is based on repeated HS analysis of the sample at equal time intervals. This procedure was originally developed for the analysis of volatiles in solid samples such as determination of monomers in polymers or for the determination of residual solvents in printed films.[26] The combination of multidimensional GC/MS with dynamic HS analysis is a powerful technique for studying the volatile organic compounds in polymers.[27]

3.8 Solid-phase Microextraction

The SPME technique was recently developed by Pawliszyn et al.[28] It is an inexpensive, rapid, solvent-free technique for isolation of organic compounds. Table 2 summarizes some of the applications where SPME has been used successfully. The applications include extraction of organic compounds from air, water and soil.[29,30] We have also demonstrated the usefulness of SPME

Table 2 Application areas for SPME

SPME applications
Industrial applications such as surfactants
Environmental analysis of e.g. water, pesticides, soil and air
Flavor analysis of food products
Toxicology analysis of blood alcohol and drugs in urine/serum
Trace impurities and degradation products in polymers

in extracting degradation products from polymers.[31,33] SPME is based typically on a 1-cm long, thin fused silica fiber coated with a polymeric stationary phase mounted in a modified syringe. For sampling, an aqueous sample containing organic analytes or a solid containing organic volatiles is placed in a vial, which is closed with a cap that contains a septum. The sheath is pushed through the septum and the fiber is immersed directly into aqueous samples or into the HS over the liquid or solid sample matrice. The analytes are absorbed on the fiber and later the adsorbed analytes are thermally desorbed in the injection port of the GC and transferred to the capillary column. Figure 3 shows the SPME extraction and injection steps. SPME is an equilibrium technique, therefore the analytes are not completely extracted from the matrix. Nevertheless, the method is useful for quantitative work and excellent precision and linearity have been demonstrated. The sensitivity and time required to reach adsorption equilibrium depend on the partition coefficients between the fiber and the analytes and the thickness of the phase. For reproducible results consistent

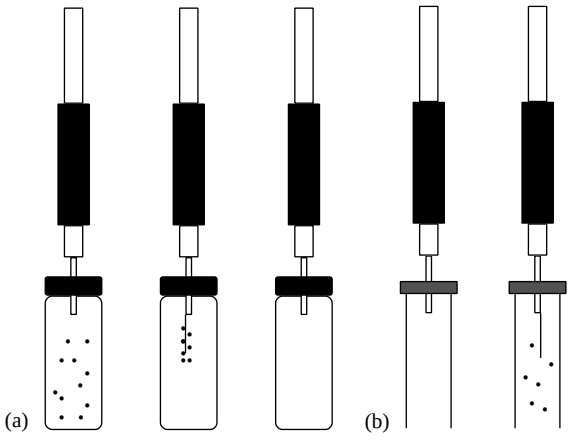

Figure 3 (a) SPME scheme: pass the needle with the retracted fiber through the sample vial. Depress the plunger to expose the fiber to the sample. The fiber can be immersed in the liquid sample placed in the HS above the sample. After 2–30 min of adsorption retract the fiber into the needle and remove it from the sample vial. (b) GC analysis: insert the needle into the GC injector port. Depress the plunger, exposing the fiber in the heated zone in the injector to desorb the analytes on to the column.

Table 3 Commercial SPME fiber materials

Stationary phase	Recommended use
PDMS	Volatiles, nonpolar semivolatiles
PDMS/DVB	Polar volatiles, e.g. amines, nitroaromatic compounds
PDMS/carboxen	Trace-level volatiles, gases, e.g. short olefins
CW/DVB	Polar analytes, e.g. alcohols
Polyacrylate	Polar semivolatiles, e.g. phenols
DVB/carboxen on PDMS	Flavors and odors

PDMS, polydimethylsiloxane; PDMS/DVB, polydimethylsiloxane/divinylbenzene; CW/DVB, carbowax/divinylbenzene.

timing is more important than full equilibrium. Limits of detection and quantitation are often below 1 ppb. Selectivity and sensitivity can be altered by changing the phase type.

Table 3 presents commercial SPME fiber materials. The most common polymeric coating for nonpolar analytes is PDMS; for polar semivolatile analytes, polyacrylate and CW/DVB fibers provide efficient extractions. PDMS/DVB fiber is suitable for extraction of volatile amines. The degradation products of PE were extracted with SPME fibers coated with either nonpolar PDMS or polar CW/DVB.[31,32] The film thicknesses of the fibers were 100 μm and 65 μm, respectively. The fibers were exposed to the HS over the solid polymer for 30 min at 80 °C. After the extraction the products were desorped into the injection port of the GC.

3.9 Supercritical Fluid Extraction

SFE is an extraction technique in which the extracting "liquid" is a supercritical fluid. Supercritical fluid is a compound that has been raised above its critical pressure and temperature. It has the chemical (solvent) properties of a liquid and physical (transport) properties of a gas. This means that a supercritical fluid penetrates a material as though it were a gas, but with the important difference that it has solvent properties approaching those of a liquid. It has high solvent strength, high diffusitivity, no surface tension and low viscosity. Conventional extractions use an assortment of solvents to extract different analytes. The solvent power of a supercritical fluid increases as its density increases. Thus a single supercritical fluid can extract a variety of components depending on the pressure applied. For several reasons carbon dioxide is the most often used supercritical fluid. It has a low critical temperature, is nonflammable, odorless and chemically inert, and is available in good purity at low cost. Even

nitrous oxide and hydrocarbons have been used, but they present fire, explosion or toxicity problems.

An SFE system includes a suitable pump, a heating block or oven, a pressure chamber to contain the sample for extraction, control valves and provision for collecting or trapping extracted components. The extractions can be done in dynamic, static or recirculating mode. In dynamic extraction the fluid is pumped through the sample to a receiver. The fluid passes the sample only once and evaporates leaving the concentrated analytes behind. This is an effective process when the analytes are quite soluble and the matrix is easily penetrated. In the static extraction the sample ''soaks'' in the supercritical fluid. After some time, the fluid is pumped to the receiver where the analytes are concentrated. Static extraction is used when the dissolution of analytes is slow or the matrix is dense. In recirculation the solvent is pumped through the sample repeatedly. After some time the fluid is pumped to the receiver and concentrated. The analytes may be collected into a liquid solvent and presented directly to a GC, high-performance liquid chromatography (HPLC) or supercritical fluid chromatography (SFC). The theory of SFE has been described in several books.[34] It has been used to extract organic compounds from solid materials, e.g. antioxidants in polyolefins,[35] aroma vapors absorbed in PE[36] and surface coatings and their raw materials.[37] The sample is usually a solid such as polymer, soil, sludge or food. SFE is also excellent for extraction of thermally labile and high-molecular-weight components.

4 ANALYSIS OF LOW-MOLECULAR-WEIGHT PRODUCTS IN POLYMERS

Low-molecular-weight compounds in polymeric materials are trace monomers, catalyst and solvent residues, low-molecular-weight additives and degradation products. The latter are formed in the polymer matrix during processing, use and/or after disposal. The molecular structure of the polymer and the conditions it is subjected to may result in everything from one to hundreds of different products.

4.1 Synthesis-related Trace Compounds in Polymers

There are in most cases many synthesis-related low-molecular-weight impurities in polymeric materials. These include monomer rests, oligomers, cross-linking residues, initiators or catalysts, solvent rests and so on. HS/GC/MS analysis showed the presence of residual 1,5-dioxepan-2-one monomer in poly(1,5-dioxepan-2-one).[17] Residual styrene in PS was quantified by MHE.[26] Purge-and-trap together with GC/MS has been used to analyze residual volatiles in high-temperature polymers such as polyimides.[38] The detection of residual volatiles gives clues to the manufacturing processes, e.g. poly(methyl methacrylate) samples from two different vendors yielded considerably different chromatograms.[39] Methyl methacrylate was the main component in both chromatograms, but the other volatiles varied greatly. Methyl *t*-butyl ether, saturated aliphatic hydrocarbons and phenol were identified in the sample from vendor 1, while methyl formate, acetone, methanol and aliphatic amines were identified in the sample from vendor 2.

4.2 Headspace Analysis of Polymers in Indoor Applications

The use of nonsterile casein in building materials, such as self-levelling floor-topping compound, led to various defects, including bad odor and health complaints in buildings where these materials were used. Our studies concluded that the cause of these problems is due to growth of alkali-tolerant *Clostridia* which degrades the casein. Air sampling was carried out above the casein-containing building materials by impinging in weak acidic solution and adsorption on silica gel.[13,14] The GC analysis of the HS samples showed the presence of small amounts of organic acids, amines and ammonia in the indoor air. In another study emissions from commercial floor components, such as carpet glue, poly(vinyl chloride) (PVC) carpet and plasticizers were studied by HS/GC/MS.[40]

Areas in which the use of polymeric materials are of particular interest, with respect to contamination risks, are food and pharmaceutical packaging materials. Technical polymers commonly used as such are PVC, PE and PP. Dynamic HS and GC/MS have been used to analyze contaminants in these materials.[41] Large amounts of plasticizer (bis(2-ethylhexyl)phthalate), and also glycerol, 2-ethylhexanol, butyl hydroxyanisole, and butylated hydroxytoluene were identified in the PVC samples, branched aliphatic hydrocarbons were extracted from PP/PE copolymers as well as some additives, for example 2,6-bis(1,1-dimethylethyl)-4-methylphenol. Residual vinyl chloride, butadiene, acrylonitrile, styrene and 2-ethylhexyl acrylate monomers were determined in their respective polymers by HS analysis.[42]

A glassfiber-reinforced unsaturated polyester was subjected to accelerated ageing at 40 °C or 60 °C and 80% relative humidity. GC/MS and HS/GC/MS were used to analyze the volatile products present in the material after different accelerated ageing times ranging from 0 to 6 years.[43] In most cases the same products but in varying amounts were present after different ageing times. Diallyl phthalate was the most abundant product in all the GC/MS chromatograms, but several other phthalates

were also identified. The concentration of the phthalates decreased as a function of degradation time. Several aromatic degradation products of the different phthalates, e.g. phthalic acid anhydride, benzoic acid and small traces of 2-propenyl ester of benzoic acid were also seen in the product mixture.

The volatile products lost during the extraction and concentration steps before GC/MS analysis were analyzed by HS/GC/MS. Several alcohols were identified in addition to the products already detected by GC/MS. None of the phthalates was volatile enough to be detected by the HS/GC/MS analysis. Isobutanol (2-methyl-1-propanol), allyl alcohol (2-propen-1-ol), 1-butanol and 1,2-propanediol were identified by HS/GC/MS. Temperature had a large influence on the formation of alcohols. A good correlation was found between the amounts of alcohol detected by HS/GC/MS and the corresponding phthalates detected by GC/MS. The most abundant phthalate was diallyl phthalate followed by diisobutyl phthalate and dibutyl phthalate. In the similar manner the most abundant alcohol was allyl alcohol formed by hydrolysis from diallyl phthalate. The fact that the amount of phthalates decreases during ageing, while alcohols increase also supports the idea that they are formed from the corresponding phthalates.

The HS/GC/MS analysis of nitrile rubber showed that large amounts of volatile compounds are formed and are able to diffuse out of the material even at relatively low temperatures such as 60 °C and 80 °C.[44] A large increase in the number and amount of products was seen when the temperature was raised from 60 °C to 80 °C. Trace amounts of 2-butoxyethanol and *N*-methylbenzylamine were already seen in the HS/GC/MS chromatograms before ageing. The area of the 2-butoxyethanol peak increased to 100 times the original peak after 7 days at 80 °C. After 248 days at 80 °C around 20 peaks were present in the chromatograms. Most of the volatiles formed, e.g. 2-butoxyethanol and 1-butanol, are hydrolysis or oxidation products of the phosphate additive.

4.3 Thermal Degradation Products of Polymers at Processing Temperatures

High-temperature processing of polyolefins causes thermal oxidation of the polymer. The resulting low-molecular-weight products can cause health problems to people exposed during processing. In the case of food packaging, the low-molecular-weight products can also migrate to food with potential toxicity effects. Processing conditions were stimulated in laboratory to capture and analyze the low-molecular-weight products formed in PE in the temperature range 260–290 °C[45] and in PP in the temperature range 220–280 °C.[46] Acids and aldehydes formed the majority of the 44 products in PE samples identified by GC/MS. The most toxic products (e.g. acrolein, methyl vinyl ketone and acrylic acid) were formed in relatively low quantities. In PP, 23 products, mainly aldehydes, but also ketones, acids and alcohols, were identified. In an attempt to improve the working environment for polymer manufacturing processes the evolution of gaseous degradation products of unstabilized and stabilized LDPE at 210 °C were compared.[47] In another study, 84 volatiles were identified when PE was thermally oxidized at 150–350 °C for between 5 and 15 min.[48,49] The products included aliphatic hydrocarbons, aldehydes, ketones and olefins.

4.4 Environmental Degradation Products of Polymers

With increased social awareness of the large percentage of plastics entering the waste stream, there is a renewed interest in producing plastics with enhanced degradability. Existing synthetic materials can be modified to increase the degradability or new (e.g. hydrolyzable) materials can be designed. One of the most difficult issues concerning these degradable plastics is their potential burden on the environment. A total fate analysis, including the identification of the degradation products, must be done on the newly developed degradable plastics before they are pronounced environmentally friendly or more adequately environmentally adaptable.

The environmental degradation of PE can be increased by different additives. PE is an example of a polymer with a random degradation mechanism. This means that during the degradation almost no monomer is produced. Instead a complex mixture of products varying in polarity and volatility is formed. It is necessary to characterize the low-molecular-weight products formed during the degradation. We have identified around 200 different degradation products in photooxidized or thermally aged enhanced photo- or biodegradable PE. These products included alkanes, alkenes, ketones, aldehydes, alcohols, mono- and dicarboxylic acids, lactones and ketoacids.[20,29–32,50–53] Several extraction methods for LLE, SPE, direct liquid extraction from the films, HS/GC/MS analysis and SPME were developed.

The products of the low-temperature (80 °C) thermooxidized and photooxidized PE were either extracted directly from the films by diethyl ether[52,53] or SPME was used to extract the volatilized products from the HS above the films.[31,32] SPME in comparison with the traditional HS/GC/MS method resulted in detection and identification of several times more products.[31] Despite having the same extraction temperature SPME was in particular more effective in extracting less volatile compounds. Carboxylic acids, dicarboxylic acids, ketoacids, ketones and hydrocarbons were identified in the ether extract,

while the volatile products in the HS mainly consisted of ketones, furanones and carboxylic acids. Because of the polarity of some of the degradation products the polar CW/DVB fiber was more suitable for the extraction than the nonpolar PDMS fiber.

The water-soluble products formed in PE films aged in water at 80 °C and at 95 °C were extracted from the water phase. In a preliminary study LLE was used,[54] but later an SPE method was developed, which allowed the identification of around three times as many products.[20,21] A nonselective nonpolar C_{18} column was chosen for the extraction. It is an ideal column for the extraction of a large number of products with different polarities and functional groups from water. The column was first activated with methanol (1 mL) and then conditioned with pH 2 water (1 mL). After that the pH of the sample water was adjusted to pH 2 with HCl. Sample water (1 mL) was then allowed to penetrate the solid-phase column. The column was dried slightly and the products were eluted with 1 mL hexane, 1 mL methanol or 1 mL methanol modified with 0.1% HCl. The methanol fractions were evaporated to 50 µL. The fraction with 0.1% HCl in methanol was warmed for 15 min at 60 °C to ensure methylation of the dicarboxylic acids.

The same LLE and SPE methods were used to extract products from preheated PE films aged in biotic and abiotic environments at ambient temperatures.[50,51] The SPE method using the nonpolar C_{18} columns allowed extraction and identification of dicarboxylic acids, which were not detected after the LLE. The extractions revealed the total absence of carboxylic acids in the biotic environment, while in an abiotic environment homologous series of carboxylic and dicarboxylic acids were identified. The results thus showed that the PE degradation products are assimilated by the microorganisms.

In several cases the identification of products was difficult because of the large number of peaks and severe overlapping in the chromatograms. Fractionation of the products by using polar NH_2 SPE columns allowed the identification of around 150 products in thermooxidized PE.[21] The products were eluted and analyzed in four fractions. The hexane-soluble products were dissolved in hexane and subjected to SPE. After the hexane-soluble products were removed, the remains of the products (e.g. dicarboxylic acids) were dissolved in methanolic HCl and analyzed directly by GC/MS. Methanolic HCl was used as a solvent to methylize the dicarboxylic acids, making them more suitable for the GC column. The hexane fraction was subjected to SPE. The nonpolar products (e.g. hydrocarbons) were not retained in the SPE column and they passed through with hexane. The mid-polar products (e.g. ketones, lactones and alcohols) were eluted with chloroform and the most polar fraction (e.g. aldehydes and carboxylic acids) was eluted with diethyl ether modified by 2% acetic acid. The fractionation allowed the use of different GC columns for products with different polarities which further facilitated the analysis.

Polymers that contain hydrolyzable groups as part of the main bone (e.g. aliphatic polyesters, polyanhydrides and polycarbonates) usually produce only a few different degradation products. We have used LLE to extract breakdown products from the buffer solution after hydrolysis of homo- and copolymers of lactide and 1,5-dioxepan-2-one.[17] SPE was used to extract lactic acid and glycolic acid after hydrolysis of lactide/glycolide copolymers in a buffer solution.[22] By using the SAX ion-exchange column we could demonstrate complete conversion of polylactic and glycolic acid (PLG) copolymers to lactic acid and glycolic acid. The results showed that the hydrolysis rate of glycolide units was much faster than that of lactide units giving increasing content of lactide units in the remaining copolymers. The hydrolysis of homo- and copolymers of polylactide (PLA) and polyglycolide (PGA) proceeded in three stages. During the first stage the molecular weight decreased rapidly, but no significant weight loss occurred. In stage two the molecular weight decrease levelled off and mass loss increased. At the time of total mass loss about 50% of the polymer had hydrolyzed to monomers. In stage three the hydrolysis of soluble oligomers continued until the polymers were totally hydrolyzed to monomers.

SPE with hyper cross-linked styrene-divinyl-benzene copolymer (ENV+) nonpolar columns and subsequent GC/MS analysis showed that the product patterns differ after biotic and abiotic hydrolysis of copolyesters of poly(3-hydroxybutyrate) (PHB) and poly(3-hydroxyvalerate) (PHV).[55,56] 3-Hydroxybutyric acid, 3-hydroxyvaleric acid, C_4-C_4 dimer, C_4-C_5 dimer, C_5-C_4 dimer and C_5-C_5 dimers were identified after 10 days of biotic hydrolysis. However, when biotic hydrolysis was continued 3-hydroxybutyric acid and 3-hydroxyvaleric acid disappeared and acetic, butanoic and pentanoic acids were detected instead. Crotonic acid, 2-pentenoic acid, 3-hydroxybutyric acid, 3-hydroxyvaleric acid and the four different dimers were identified after abiotic hydrolysis.

ENV+ SPE columns were also used to extract hydrolysis products of poly-L-lactide (PLLA), after aging in abiotic or biotic mineral medium for different periods of time.[57,58] After the first week acetic and propanoic acid were extracted from the biotic mineral medium; later these products were assimilated by microorganisms and were no longer detected in the biotic medium. The rate of abiotic hydrolysis was quite slow, explaining the absence of hydrolysis products in the abiotic medium. The low-molecular-weight products were also extracted directly from the films with diethyl ether. Lactide, lactic acid and lactoyl lactic acid were extracted from the abiotically aged

films. In the biotic medium lactic acid and lactoyl lactic acid were rapidly assimilated by the microorganisms, but lactide remained in the films.

4.5 Additive Systems in Plastics and Coatings

Chemical additives are frequently used to improve the useful properties and to extend the service life of polymers. Various additives are employed for these purposes, antioxidants and light stabilizers being the most important in protecting and developing desirable chemical and physical properties. The polyolefins industry consumes about half of the antioxidants and three-quarters of the light stabilizers produced for thermoplastic polymers. Besides antioxidants and light stabilizers, other additives are sometimes incorporated into the polymer matrix, e.g. PVC contains large amounts of plasticizers.

The additive packages used for polyolefins can be quite complicated, containing several antioxidants, both primary (such as hindered phenols) and secondary antioxidants, as well as antistats, slips, pigments and so on. The extraction and subsequent separation and quantitation of polymer additives in polyolefins has proven to be a challenge for the analytical chemist. There have been several workers who have investigated the separation of additives from polyolefins,[59] but extraction and recovery (at >90%) in reasonable times has been difficult. Reprecipitation (required time 0.5–2 h) or Soxhlet extraction (12–48 h) have been used for many years. Ultrasonication and MAE have been shown to be the most effective extraction methods.[10,11]

Typical polyolefin antioxidants, such as Irganox 1010, Irgafos 168 and Cyasorb UV 531 were extracted from HDPE, LDPE and PP by ultrasonication and MAE.[11] The extraction times were typically 20 min for microwave oven and 30–60 min for the ultrasonic bath with 90+% recoveries. MAE was also used in another study to extract Irganox 1010, Irgafos 168 and Chimassorb 81 from PE and PP.[60] Rapid extractions were achieved with a mixture of acetone and *n*-heptane and with chloroform alone. 91–100% of the stabilizers were extracted after only 2–6 min of microwave heating. Thick samples such as pellets had to be ground before extraction. A range of analytical techniques including GC/MS and HPLC were used to characterize the conversion products of a common polyolefin processing stabilizer, Irgafos 168, and a common polyolefin antioxidant, Irganox 1076.[61] GC/MS analysis showed that the phosphite stabilizer is oxidized to phosphate under processing conditions. Degradation products of Irganox 1076 (e.g. cinnamate) were different dimeric oxidation products and dealkylation products depending on temperature.

The primary commercial application of plasticizers is to modify stiffness characteristics of PVC resins. In general even small quantities of plasticizer have a noticeable effect. However, PVC is highly plasticized to produce soft and flexible packaging films, wire insulation, tubing, artificial leather and so on. In most flexible formulations plasticizer levels are 20–80 parts of plasticizer per 100 parts of PVC. Phthalate esters are the most commonly used class of plasticizers. Alterations in the properties of plasticized PVC are caused mainly by plasticizer lost, e.g. as a result of contact with liquid or solid media. The plasticizer loss depends on temperature, contact medium, structure and vapour pressure of the plasticizer.

Several investigations have been done on the migration of plasticizer from PVC used in food packaging or in medical articles. GC provides the most effective identification method for monomeric plasticizers in PVC.[62] An ultrasonic bath with hexane as extraction solvent was used to extract bis(2-ethylhexyl)adipate (DEHA) plasticizer from food products wrapped in PVC films.[63] The method was successfully used to study migration of DEHA to several cheeses. Migration of epoxidized soya bean oil (ESBO) plasticizer from PVC gaskets into baby food was studied by GC/MS.[64] ESBO was detected in almost all 81 studied dishes with levels ranging from 1.5 to 50.8 mg kg^{-1}. The plasticizer of 12 flexible PVC foils used as coating materials in drinking water reservoirs was investigated. The plasticizers were leached out by *n*-hexane and analyzed by GC.[65] The diffusion of dioctylphthalate plasticizer from PVC into an oil solvent, with or without stirring, was determined by GC analysis.[66] PVC used in medical and pharmaceutical packaging may contain up to 40% of additives. Both PVC bags and contamination of intravenous solutions from the PVC bags were studied by dynamic HS/GC/MS and LC.[67] Many different compound (e.g. alcohols, aldehydes, cyclohexanone, carboxylic acids, phthalates and other aromatic compounds) were present in the PVC materials. Liquid chromatographic analysis showed that some of these compounds were also dissolved in the intravenous solutions.

Volatile species produced during UV radiation of a rigid PVC construction panel were established to be a function of irradiation time.[68] Several organic and chloroorganic products were identified by GC/MS, including chloroethane, benzene, acetone, dichloropropanes, dichloromethane, chloroform, butanol, butanal, butyl formate, 1-chlorobutane and di-*n*-butyl ether. Powder coatings consist of an intimate mixture of resin, pigment and suitable additives whose function is to promote flow, adhesion or chemical curing of the finish. The growth of powder coatings business is affected by the toxicological effects of raw materials used in formulations. Different chromatographic techniques have been investigated for analyzing surface coatings and many of the raw materials and additives (e.g. plasticizers, curing agents

and surface active agents, oligomeric mixtures and so on).[69,70] Less volatile paint additives such as HALS or UV absorbers have been determined by using HS/GC.[71] GC/MS has been used to characterize the binding media in paintings.[72] Solvents in water-based coatings have been determined by sampling coating HS using SPME and GC.[73] Experimental techniques were developed to investigate the migration of UV absorbers and HALS in acrylic/melamine clearcoats during cure[74] and loss of UV absorbers in automotive coatings.[75] Automotive paint film degradation on exposure to 290–300 °C in a nonoxidizing atmosphere was characterized by thermal desorption GC/MS.[76] The volatile emissions varied depending on the paint system, manufacturer and the thermoplastic substrate.

ABBREVIATIONS AND ACRONYMS

CW/DVB	Carbowax/Divinylbenzene
DEHA	Bis(2-ethylhexyl)adipate
ENV+	Hyper Cross-linked Styrene-divinylbenzene Copolymer
ESBO	Epoxidized Soya Bean Oil
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HALS	Hindered Amine Light Stabilizer
HDPE	High-density Polyethylene
HPLC	High-performance Liquid Chromatography
HS	Headspace
HS/GC	Headspace Gas Chromatography
HS/GC/MS	Headspace Gas Chromatography/Mass Spectrometry
LDPE	Low-density Polyethylene
LLE	Liquid–Liquid Extraction
MAE	Microwave-assisted Extraction
MDPE	Medium-density Polyethylene
MHE	Multiple Headspace Extraction
PDMS	Polydimethylsiloxane
PDMS/DVB	Polydimethylsiloxane/Divinylbenzene
PE	Polyethylene
PET	Poly(ethylene terephthalate)
PGA	Polyglycolide
PHB	Poly(3-hydroxybutyrate)
PHV	Poly(3-hydroxyvalerate)
PLA	Polylactide
PLG	Polylactic and Glycolic Acid
PLLA	Poly-L-lactide
PP	Polypropylene
PS	Polystyrene
PVC	Poly(vinyl chloride)
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
UV	Ultraviolet

RELATED ARTICLES

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Environment: Water and Waste **(Volume 4)**
Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Food **(Volume 5)**
Flavor Analysis in Food • Liquid Chromatography in Food Analysis • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General • Sample Preparation, Headspace Techniques

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization • Infrared Spectroscopy in Analysis of Polymer Degradation • Pyrolysis Techniques in the Analysis of Polymers and Rubbers • Size-exclusion Chromatography of Polymers • Supercritical Fluid Chromatography of Polymers

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Sample Preparation for Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. T.R. Crompton, *Chemical Analysis of Additives in Plastics*, 2nd edition, Pergamon Press, Oxford, 1977.
2. D. Munteany, 'Analysis of Antioxidants and Light Stabilisers in Polymers by Modern Liquid Chromatography', in *Mechanisms of Polymer Degradation and Stabilisation*, ed. G. Scott, Elsevier Science, Oxford, 211–314, 1990.
3. K.S. Miller, J.M. Krochta, 'Measuring Aroma Transport in Polymer Films', *Trans. ASAE*, **41**(2), 427–433 (1998).
4. T.J. Nielsen, I.M. Jagerstad, R.E. Oste, B.O. Wesslen, 'Comparative Absorption of Low-molecular Aroma Compounds into Commonly Used Food Packaging Polymer Films', *J. Food Sci.*, **57**(2), 490–492 (1992).
5. W. Camacho, S. Karlsson, 'Fragrance and Flavour Compounds in Reprocessed Packaging HDPE and PP', submitted for publication.
6. M. Huber, R. Franz, 'Studies on Contamination of Post Consumer Plastics from Controlled Resources for Recycling into Food Packaging Applications', *Dtsch. Lebensm.-Rundsch.*, **93**(10), 328–331 (1997).
7. T. Macko, B. Furtner, K. Lederer, 'Analysis of Aromatic Antioxidants and Ultraviolet Stabilizers in Polyethylene using High-temperature Extraction with Low Boiling Solvent', *J. Appl. Polym. Sci.*, **62**(13), 2201–2207 (1996).
8. D. Dilettato, P.J. Aprino, K. Nguyen, A. Bruchet, 'Investigation of Low Mass Oligomers and Polymeric Additives from Plastics. 2. Application to Polyolefin Soxhlet Extracts', *J. High Resolut. Chromatogr.*, **14**(5), 335–341 (1991).
9. H.J. Vanderburg, A.A. Clifford, K.D. Bartle, J. Carrol, 'Analytical Extraction of Additives from Polymers', *Analyst*, **22**, R101–R115 (1997).
10. N. Haider, S. Karlsson, 'A Rapid Ultrasonic Extraction Technique to Identify and Quantify Additives in Poly(ethylene)', *Analyst*, **124**, 797–800 (1999).
11. R.C. Nielson, 'Extraction and Quantitation of Polyolefin Additives', *J. Liq. Chromatogr.*, **14**(3), 503–519 (1991).
12. C.T. Costley, J.R. Dean, I. Newton, J. Carroll, 'Extraction of Oligomers from Poly(ethylene terephthalate) by Microwave-assisted Extraction', *Anal. Commun.*, **34**, 89–91 (1997).
13. L. Contat-Rodriges, N. Haider, Ribes, S. Karlsson, 'Identification of Degradation Products from HDPE and PP Blends with Enhanced Degradability Aged in Soil', submitted for publication.
14. R.E. Majors, 'Liquid Extraction Techniques for Sample Preparation', *LC/GC*, **14**(11), 936–943 (1996).
15. S. Karlsson, Z.G. Banhidi, A.-C. Albertsson, 'Gas Chromatographic Detection of Volatile Amines found in Indoor Air due to Putrefactive Degradation of Casein-containing Building Materials', *Mater. Struct.*, **22**, 163–169 (1989).
16. S. Karlsson, A.-C. Albertsson, 'The Biodegradation of a Biopolymeric Additive in Building Materials', *Mater. Struct.*, **23**, 352–357 (1990).
17. S. Karlsson, M. Hakkarainen, A.-C. Albertsson, 'Identification by Headspace Gas Chromatography/Mass Spectrometry of in vitro Degradation Products of Homo- and Copolymers of L- and D,L-Lactide and 1,5-Dioxepan-2-one', *J. Chromatogr. A*, **688**, 251–259 (1994).
18. M.-C. Hennion, V. Pichon, 'Solid-phase Extraction of Polar Organic Pollutants from Water', *Environ. Sci. Technol.*, **28**(13), 576A–583A (1994).
19. Z.E. Penton, 'Sample Preparation for Gas Chromatography with Solid-phase Extraction and Solid-phase Microextraction', *Advances in Chromatography*, eds. P.R. Brown, E. Grushka, Marcel Dekker, 205–236, Vol. 37, 1997.
20. A.-C. Albertsson, C. Barenstedt, S. Karlsson, 'Solid-phase Extraction and Gas Chromatographic–Mass Spectrometric Identification of Degradation Products from Enhanced Environmentally Degradable Polyethylene', *J. Chromatogr. A*, **690**, 207–217 (1995).
21. M. Hakkarainen, A.-C. Albertsson, S. Karlsson, 'Solid-phase Extraction and Subsequent Gas Chromatography/Mass Spectrometry Analysis for Identification of Complex Mixtures of Degradation Products in Starch-based Polymers', *J. Chromatogr. A*, **741**, 251–263 (1996).
22. M. Hakkarainen, A.-C. Albertsson, S. Karlsson, 'Weight Losses and Molecular Weight Changes Correlated with the Evolution of Hydroxyacids in Simulated in vivo Degradation of Homo- and Copolymers of PLA and PGA', *Polym. Degrad. Stab.*, **52**, 283–291 (1996).
23. H. Hachenberg, A.P. Schmidt, *Gas Chromatographic Headspace Analysis*, Heyden & Son, London, 1997.
24. B. Kolb, 'Application of an Automated Headspace Procedure for Trace Analysis by Gas Chromatography', *J. Chromatogr. Sci.*, **122**, 553–568 (1976).
25. B. Kolb, 'Multiple Headspace Extraction–A Procedure for Eliminating the Influence of the Sample Matrix in Quantitative Headspace Gas Chromatography', *Chromatographia*, **15**(9), 587–594 (1982).
26. B. Kolb, P. Pospisil, 'A Gas Chromatographic Assay for Quantitative Analysis of Volatiles in Solid Materials by Discontinuous Gas Extraction', *Chromatographia*, **10**(12), 705–711 (1977).
27. A. Hagman, S. Jacobsson, 'Analysis of Volatile Organic Compounds in Polymers by Dynamic Headspace-multidimensional Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **395**, 271–279 (1987).
28. A.A. Boyd-Boland, M. Chai, Y.Z. Luo, Z. Zhang, M.J. Yang, J.B. Pawliszyn, T. Gorecki, 'New Solvent-free Sample Preparation Techniques, Based on Fiber and

Polymer Technologies', *Environ. Sci. Technol.*, **28**(13), 569A–574A (1994).
29. K.D. Buchholz, J. Pawliszyn, 'Optimization of Solid-phase Microextraction Conditions for Determination of Phenols', *Anal. Chem.*, **66**, 160–167 (1994).
30. P. Popp, K. Kalbitz, G. Oppermann, 'Application of Solid-phase Microextraction and Gas Chromatography with Electron-capture and Mass Spectrometric Detection for the Determination of Hexachlorocyclohexanes in Soil Solutions', *J. Chromatogr. A*, **687**, 133–140 (1994).
31. M. Hakkarainen, A.-C. Albertsson, S. Karlsson, 'Solid Phase Microextraction (SPME) as an Effective Means to Isolate Degradation Products in Polymers', *J. Environ. Polym. Degrad.*, **5**(82), 67–73 (1997).
32. F. Khabbaz, A.-C. Albertsson, S. Karlsson, 'Trapping of Volatile Low-molecular Weight Photoproducts in Inert and Enhanced Degradable LDPE', *Polym. Degrad. Stab.*, **61**, 329–342 (1998).
33. F. Khabbaz, A.-C. Albertsson, S. Karlsson, 'Chemical and Morphological Changes of Environmentally Degradable Polyethylene Films Exposed to Thermo-oxidation', *Polym. Degrad. Stab.*, **63**, 127–138 (1999).
34. B. Wenclawiak (ed.), *Analysis with Supercritical Fluids: Extraction and Chromatography*, Springer-Verlag, Berlin, 1992.
35. M. Waldeback, C. Jansson, F.J. Senorans, K.E. Markides, 'Accelerated Solvent Extraction of Antioxidant Irganox 1076 in Linear Low Density Polyethylene (LLDPE) Granules before and after Gamma-irradiation', *Analyst*, **123**(6), 1205–1207 (1998).
36. F. Johansson, A. Leufven, M. Eskilson, 'Quantification of Aroma Vapors Sorbed in Polyethylene Films using Supercritical Carbon Dioxide', *J. Sci. Food Agric.*, **61**(2), 241–244 (1993).
37. M.W. Raynor, K.D. Bartle, 'Supercritical Fluid Chromatography and Extraction in Surface Coatings Analysis–Review of Representative Applications', *J. Supercrit. Fluids*, **6**(1), 39–49 (1993).
38. J.W. Long, R.W. Snyder, J. Spalik, 'Purge and Trap Gas Chromatographic Method for Residual Volatiles in High-temperature Polymers', *J. Chromatogr.*, **450**, 394–398 (1988).
39. W.V. Ligon Jr, M.C. George, 'Analysis of Volatile Constituents in Commercial Polymers by Direct Thermal Desorption and Gas Chromatography/Mass Spectrometry', *J. Polym. Sci., Polym. Chem. Ed.*, **16**, 2703–2709 (1978).
40. C.-A. Eriksson, F. Björk, S. Karlsson, F. Khabbaz, *Svensk rapport Alkalisk nedbrytning av golvkomponenter*, Meddelande No. 177, KTH, Stockholm, 1999.
41. S. Jacobsson, 'Analysis of Volatile Organic Compounds in Polymers by Dynamic Headspace and Gas Chromatography/Mass Spectrometry', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **7**, 185–190 (1984).
42. R.J. Steichen, 'Modified Solution Approach for the Gas Chromatographic Determination of Residual Monomers by Headspace Analysis', *Anal. Chem.*, **48**(9), 1398–1402 (1976).
43. M. Hakkarainen, G. Gallet, S. Karlsson, 'Prediction by Multivariate Data Analysis of Long-term Properties of Glassfiber Reinforced Polyester Composites', *Polym. Degrad. Stab.*, **64**, 91–99 (1999).
44. M. Hakkarainen, S. Karlsson, 'Low Molecular Weight Products in Long-term Aged Nitrile-rubber Identified by GC/MS and Headspace/GC/MS', manuscript.
45. A. Hoff, S. Jacobsson, 'Thermooxidative Degradation of Low-density Polyethylene Close to Industrial Processing Conditions', *J. Appl. Polym. Sci.*, **26**, 3409–3423 (1981).
46. A. Hoff, S. Jacobsson, 'Thermal Oxidation of Polypropylene Close to Industrial Processing Conditions', *J. Appl. Polym. Sci.*, **27**, 2539–2551 (1982).
47. J.O. Madsen, N.B. Olsen, G. Atlung, 'Effectiveness of Antioxidants: Suppression of Evolution of Gaseous Degradation Products from Low-density Polyethylene during Thermo-oxidation', *Polym. Degrad. Stab.*, **12**, 131–140 (1985).
48. A. Bravo, J.H. Hotchkiss, 'Identification of Volatile Compounds Resulting from the Thermal Oxidation of Polyethylene', *J. Appl. Polym. Sci.*, **47**, 1741–1748 (1993).
49. A. Bravo, J.H. Hotchkiss, T.E. Acree, 'Identification of Odor-active Compounds Resulting from Thermal Oxidation of Polyethylene', *J. Agric. Food Chem.*, **40**(10), 1881–1884 (1992).
50. A.-C. Albertsson, B. Erlandsson, M. Hakkarainen, S. Karlsson, 'Molecular Weight Changes and Polymeric Matrix Changes Correlated with the Formation of Degradation Products in Biodegraded Polyethylene', *J. Environ. Polym. Degrad.*, **6**(4), 187–195 (1998).
51. A.-C. Albertsson, C. Barenstedt, T. Lindberg, S. Karlsson, 'Degradation Product Pattern and Morphology Changes as Means to Differentiate between Abiotically and Biotically Aged Degradable Polyethylene', *Polymer*, **36**(16), 3075–3083 (1995).
52. M. Hakkarainen, A.-C. Albertsson, S. Karlsson, 'Susceptibility of Starch-filled and Starch-based LDPE to Oxygen in Water and Air', *J. Appl. Polym. Sci.*, **66**, 959–967 (1997).
53. S. Karlsson, S. Karlsson, M. Hakkarainen, A.-C. Albertsson, 'Dicarboxylic Acids and Ketoacids Formed in Degradable Polyethylenes by Zip Depolymerization through a Cyclic Transition State', *Macromolecules*, **30**, 7721–7728 (1997).
54. A.-C. Albertsson, C. Barenstedt, S. Karlsson, 'Abiotic Degradation Products from Enhanced Environmentally Degradable Polyethylene', *Acta Polym.*, **45**, 97–103 (1994).
55. C. Eldsäter, A.-C. Albertsson, S. Karlsson, 'Impact of Degradation Mechanisms of Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) during Composting', *Acta Polym.*, **48**, 478–483 (1997).

56. C. Eldsäter, A.-C. Albertsson, S. Karlsson, 'Effect of Abiotic Factors on the Degradation of Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in Simulated and Natural Composting Environments', *Polym. Degrad. Stab.*, **64**, 177–183 (1999).
57. M. Hakkarainen, S. Karlsson, A.-C. Albertsson, 'Rapid (bio)degradation of Polylactide by Mixed Culture of Compost Microorganisms–Low-molecular Weight Products and Matrix Changes', *Polymer*, in press (1999).
58. M. Hakkarainen, S. Karlsson, A.-C. Albertsson, 'The Influence of Low-molecular Weight Lactic Acid Derivatives on (bio)degradability of Polylactide', *J. Appl. Polym. Sci.*, in press (1999).
59. A.M. Wims, S.J. Swarin, 'Determination of Antioxidants in Polypropylene by Liquid Chromatography', *J. Appl. Pol. Sci.*, **19**, 1243–1256 (1975).
60. W. Freitag, O. John, 'Fast Separation of Stabilizers from Polyolefins by Microwave Heating', *Angew. Makromol. Chem.*, **175**, 181–185 (1990).
61. J. Scheirs, J. Pospisil, M.J. O'Connor, S.W. Bigger, 'Characterization of Conversion Products Formed during Degradation of Processing Antioxidants', *Adv. Chem. Ser.*, **248**, 359–375 (1996).
62. R.R. Kozlowski, T.K. Gallagher, 'A Practical Guide to the Identification of Monomeric Plasticizers in Flexible PVC Compounds', *J. Vinyl Additive Technol.*, **3**(3), 249–255 (1997).
63. C. Nerin, P. Gancedo, J. Cacho, 'Determination of bis(2-ethylhexyl) Adipate in Food Products', *J. Agric. Food Chem.*, **40**(10), 1833–1835 (1992).
64. L. Hammarling, H. Gustavsson, K. Svensson, S. Karlsson, A. Oskarsson, 'Migration of Epoxidized Soya Bean Oil from Plasticized PVC Gaskets into Baby Food', *Food Additives Contaminants*, **15**(2), 203–208 (1998).
65. D.K. Frensch, D. Schoenen, H.F. Scholer, 'Identification of Leachable Organic Compounds of PVC-foils Used in Drinking-water Reservoirs', *Z. Wasser Abwasser Forsch. (Journal for Water and Wastewater)*, **24**(6), 294–299 (1991).
66. E. Kondili, M. Kontominas, M. Kosmas, 'The Effect of Stirring on the Diffusion of Small Molecules from a Polymer Matrix into a Solution', *Polymer*, **34**(12), 2592–2596 (1993).
67. A. Arbin, S. Jacobsson, K. Hänninen, A. Hagman, J. Östelius, 'Studies on Contamination of Intravenous Solutions from PVC Bags with Dynamic Headspace GC/MS and LC–Diode Array Techniques', *Int. J. Pharm.*, **28**, 211–218 (1986).
68. D.J. Carlsson, M. Krzymien, G. Pleizier, D.J. Worsfold, M. Day, 'Volatiles Release from Photodegrading, Pigmented PVC: Kinetic Changes', *Polym. Degrad. Stab.*, **62**, 413–419 (1998).
69. M.W. Raynor, K.D. Bartle, 'Supercritical Fluid Chromatography and Extraction in Surface Coatings Analysis–Review of Representative Applications', *J. Supercrit. Fluids*, **6**(1), 39–49 (1993).
70. T.P. Wampler, G.A. Bishea, W.J. Simonsick, 'Recent Changes in Automotive Paint Formulations using Pyrolysis Gas Chromatography Mass Spectrometry for Identification', *J. Anal. Appl. Pyrolysis*, **40**(1), 79–89 (1997).
71. Tsuge, S. Kurlyama, K. Ohtani, 'Dynamic Headspace Analysis of Trace Volatile Components in Polymeric Materials by Wide Bore Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **12**(11), 727–731 (1989).
72. A. Casoli, P.C. Musini, G. Palla, 'Gas Chromatographic Mass Spectrometric Approach to the Problem of Characterizing Binding Media in Paintings', *J. Chromatogr. A*, **731**(1–2), 237–246 (1996).
73. A.C. Censullo, D.R. Jones, M.T. Wills, 'Direct VOC Analysis of Water-based Coatings by Gas Chromatography and Solid-phase Microextraction', *J. Coatings Technol.*, **69**(869), 33–41 (1997).
74. G. Haacke, F.F. Andrawes, B.H. Campbell, 'Migration of Light Stabilizers in Acrylic/Melamine Clearcoats', *J. Coatings Technol.*, **68**(855), 57–62 (1996).
75. R. Iyengar, B. Schellenberg, 'Loss Rate of UV Absorbers in Automotive Coatings', *Polym. Degrad. Stab.*, **61**(1), 151–159 (1998).
76. W.R. Rodgers, T.S. Ellis, G.D. Cheever, R. Louisferdinand, D.P. Thorton, N. Somers, 'Chemical Analysis of Painted Thermoplastics by Thermal-degradation GC/MS', *J. Coatings Technol.*, **66**(837), 27–33 (1994).

Infrared Spectroscopy in Analysis of Plastics Recycling

Norbert Eisenreich and Thomas Rohe
Fraunhofer Institut für Chemische Technologie (ICT), Pfinztal, Germany

1 Introduction **1**
1.1 Government Regulations 2
1.2 Plastics Consumption and Waste 3
1.3 Economic and Technical Demands 3
2 Materials **4**
2.1 Waste Streams 4
2.2 Methods of Analysis 5
3 Infrared Bands of Plastic Materials **6**
3.1 Principles of Vibrational Spectroscopy 6
3.2 Spectra of Plastics in the Infrared and Near-infrared Wavelength Range 7
3.3 The Spectra of Industrial Plastic Samples 9
3.4 Near-infrared Spectra 9
3.5 Mid-infrared Spectra 11
3.6 Gas-phase Spectra 12
4 Identification Measurement Systems **13**
5 Identification and Performance **15**
5.1 Identification Algorithms 15
5.2 Applications 17
6 Conclusions **18**
List of Symbols **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Plastics recycling recovers valuable materials to save natural resources which are limited in the long term. Reprocessing of the materials requires the identification of the various types of polymers. Difficulties in applying physical separation techniques led to the modification of methods of chemical analysis which have to be operated fast and reliably under harsh industrial environments favoring infrared (IR) methods.

In general, the plastic products to be handled can be separated into two main fractions. Domestic wastes contain only five relevant polymers in large quantities. Technical products use a broad variety of polymers containing fillers and additives, e.g. flame retardants, which also have to be identified to allow economic recycling.

Solid samples absorb IR radiation strongly, and the fundamental vibrations of the molecules are especially intense. The near-infrared (NIR) spectral range with the overtone and combination bands is appropriate for analyzing bulky materials in reflection and transmission modes. The plastics from domestic wastes can be reliably identified by the first overtone of the C−H bands between 1600 and 1800 nm. Extending this wavelength range to 1000 nm, technical non-black plastics can also be treated. In the case of black plastics, which are often used in audio and video appliances, the mid-infrared (MIR) spectral range has to be applied, owing to the reduced penetration depth of radiation into these materials. Also, larger amounts of additives such as plasticizers and flame retardants can be analyzed.

The identification of plastics from household waste requires fast scanning techniques in the millisecond range, especially if many samples are to be taken for one identification on movement. Mainly acousto-optic tunable filter (AOTF) spectrometers or grating spectrometers with diode arrays can be applied. Germanium (Ge) and indium gallium arsenide (InGaAs) detectors with extended wavelength sensitivity cover the spectral range of interest. The movement excludes Fourier transform infrared (FTIR) spectrometers with e.g. indium antimonide (InSb) and mercury cadmium telluride (HgCdTe) diodes, but its application to technical plastics is possible because these pieces are heavier and can be identified in a fixed position.

Identification for recycling means a qualitative analysis. Simple algorithms such as those forming the differences of normalized spectra are sufficient for comparing measured with stored spectra. The performance of microprocessors also allows the application of more complex algorithms such as multivariate analysis or even neural networks. The use of NIR spectroscopy in industrial recycling started with application to sorting systems for plastics from household wastes or hand-held sensors for cable recycling. The MIR technique still suffers from its sensitivity to harsh industrial environments and has been used only rarely and needs further development.

1 INTRODUCTION

Plastics are of great importance in our modern industrial world. They are used in nearly every industrial branch and can be made from cheap raw materials. A broad variety of types and blends modified by plasticizers, fillers, and pigments account for various properties required by the applications. Moreover, they have a low density (about

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

$0.8–1.5\,g\,cm^{-3}$), can be easily processed, formed, and colored at low temperatures and can be adapted to fulfill low- and high-grade specifications of packaging materials and engineering plastics. Owing to their currently high and still strongly growing production, plastics cause environmental problems when the materials used have to be taken care of after the end of life of the product. Even after decades they do not degrade substantially in landfill sites. Therefore, plastics recycling has become a mandatory task today and will remain challenging in the future.

Recycling of plastics is carried out in different ways, as shown in Figure 1. Material recycling means that the material itself is directly reused for the manufacture of new products. Application with a high technical standard demands, of course, appropriate identification and sorting of the various types including important additives, because mixed plastics can only be processed to low-grade materials with limited use. This step is also necessary for the alternatives: raw-material recycling (feedstock), i.e. the polymer is cracked into its monomers, which are used for the production of virgin polymer granulates, and energy utilization, i.e. use of the heat of combustion. In both cases the used plastic parts have to be characterized to avoid process instabilities or ineffective processes. In the case of halogenated materials such as poly(vinyl chloride) (PVC) or brominated flame retardants, even the various thermal treatments are suspected to produce harmful constituents or to corrode the facility.

In conclusion, this means that recycling requires an effective and reliable separation system. IR spectroscopic methods have proved to be the most powerful identification techniques. They fulfill the specifications for plastics recycling within the framework consisting of

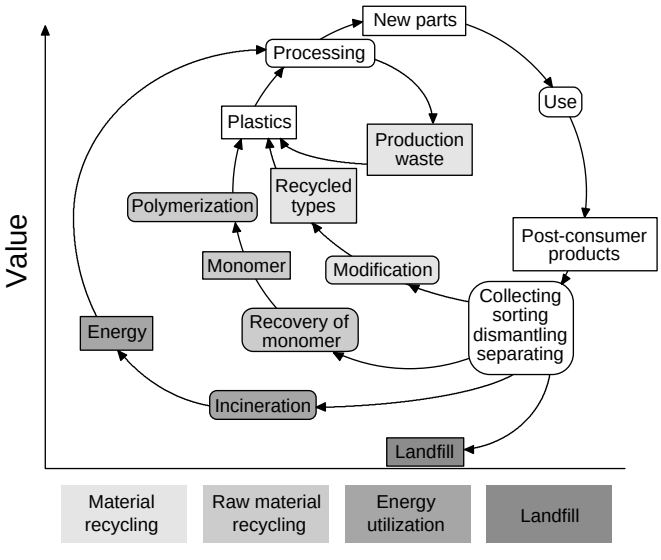

Figure 1 Recycling concepts.

government regulations, amounts and various streams of plastic waste, economic demands, and technical and industrial conditions.

1.1 Government Regulations

In the last two to three decades, the awareness of our environment has grown tremendously. Most countries are aware of the growing problem of pollution and the decrease in natural resources. The consequences force us to economize on these natural resources and to re-use post-consumer products. Meanwhile, the legislative authorities in different countries, especially in Europe, have become aware of these consequences and have taken action.

One of the most progressive countries in this context is Germany, where the Packaging Ordinance was introduced in 1991 with the goal of avoiding packaging waste. This ordinance is based on the take-back duty of packaging materials by trading companies. These companies are not in a position to take care of professional recycling, so an organization was founded which is responsible for this duty, the Duales System Deutschland (DSD). The DSD marks packaging materials with the so-called Green Dot, which demonstrates to the consumer that this material will be recycled and that the manufacturer or trader of this product has a contract with the DSD.

However, not only the packaging waste from households is covered by national legislation. In 1991, the first draft of an Electronic Scrap Ordinance was published in Germany. Since then there have been years of discussion between all stakeholders involved. The first Member States of the European Union (EU) have already issued producer responsibility legislation, and others are preparing corresponding regulations or seem to be waiting for general guidelines given by the EU. The current status of the EU Member States' legislation on take-back of waste of electrical and electronic equipment is as follows:

- (draft) legislation already developed or adopted: Austria, Denmark, Germany, The Netherlands, Sweden, Switzerland;
- draft legislation under development: Belgium, Finland, France, Italy, Norway, United Kingdom;
- no significant official steps taken: Greece, Ireland, Portugal, Spain.

Harmonization within the EU is very likely to take place at the beginning of the next millennium, as it is already standard in other sectors of waste management legislation.[1] Product return legislation has also been established in Japan, and is under construction in the USA.[2–6] The reason for this legislation all over the world is simple: consumers want to return used plastic

material and to be assured that it will be recycled in the most ecological and nonpolluting way.

1.2 Plastics Consumption and Waste

All over the world, plastics production, and hence consumption, have grown tremendously since World War II. World production of plastics in 1997 was about 150 million tons.[7] If all these plastics were to be made as one thread and passed around the earth, it would have a diameter of about 1.8 m, and if this thread were to be tied between the Earth and the Moon, it would have a diameter of about 0.5 m! This demonstrates the huge amount which is produced year by year. The world distribution can be seen in Figure 2; 28% came from North America and 32.5% from Europe. Of the 41 million tons of plastics produced in Western Europe, the largest part of about 26 million tons was the so-called standard plastics, which are polyethylene (PE), PVC, polypropylene (PP), polystyrene (PS), and poly(ethylene terephthalate) (PET) (see Figure 3). These standard plastics are mainly used in packaging, which is the largest application field of plastics today (see Figure 4). However, other branches of industry are now using more

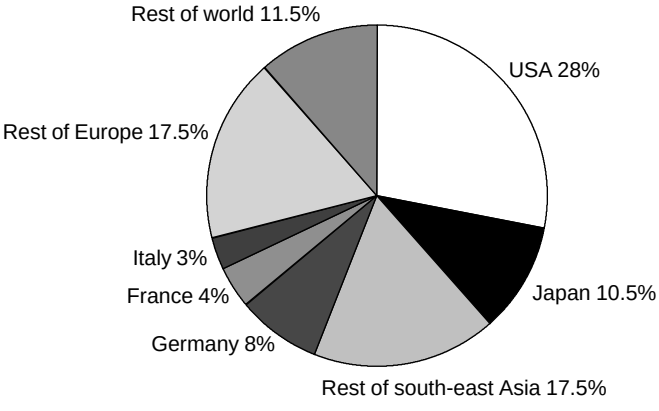

Figure 2 World production of plastics.

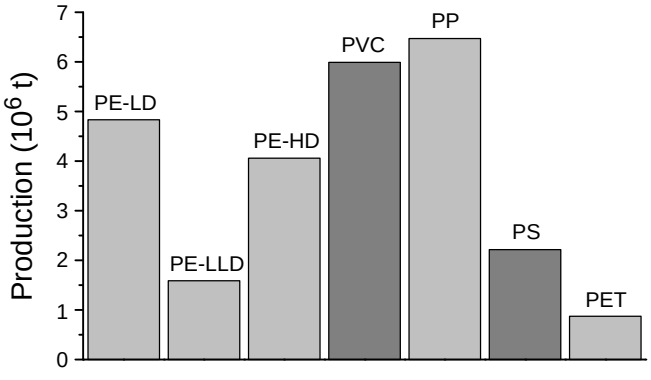

Figure 3 Distribution of plastic production by plastic types.

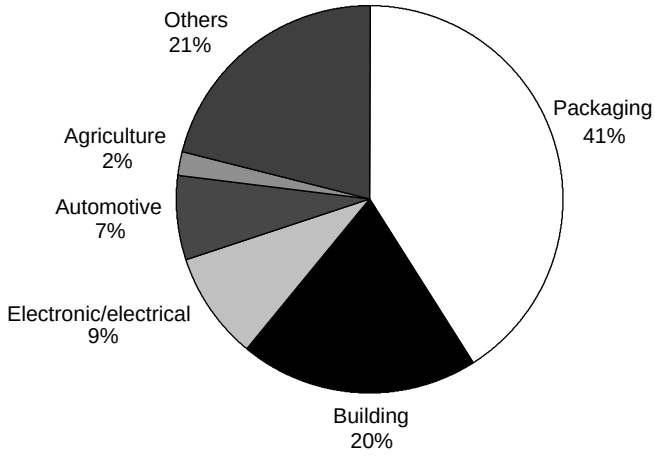

Figure 4 Distribution of plastics according to different industries.

and more plastics for manufacturing. Some examples of the variety of applications for plastics are foils, bottles, cups, window frames, isolation mats, fenders (bumpers), synthetic fibers, housings for monitors, television (TV) sets, inline skates, biker helmets, and tools.

In summary, a huge amount of plastics is produced each year and it is expected that the production and consumption of plastics will have an annual growth of 4–5% per year. This gives rise to the question of what to do with this material after the end of life of the products made out of plastics.

There are many recycling operations all over the world and here the statistics of recycling operations in Germany will be examined.[8] In 1995, about 1 million tons of plastics were taken for this purpose; 54% came from post-consumer products and 46% were production or processing waste. This material was recycled in different ways; 38% were reused for new self-manufactured products. This means, for example, that production waste remaining from injection-molded products, which can be ground and regranulated, is reprocessed directly, because the material is well known and has properties close to those of the virgin material. Some 58% of the collected plastics was delivered to recycling companies and 4% went to waste incineration or landfill. The percentage of old plastics which was recycled to real marketable products was about 70%. This percentage is very impressive, but it has to be kept in mind that most of these products are of only low technical value, e.g. benches or sound-protection walls. The reasons for this low-level reuse will be given in the next section.

1.3 Economic and Technical Demands

Landfill or incineration is in nearly every case cheaper than recycling of mixed plastics. However, there is a

solution for an economic recycling operation which is based on the reuse of high-quality polymers for high-end applications to recover the real value of the plastic material. If the plastics have to be used at a similar grade to the end-of-life product, the various types of polymers need specifying and important additives and fillers have to be identified and sorted out of the waste stream of mixed materials. Only then is rigorous reprocessing to high-tech applications possible, because even small quantities of impurities could spoil the quality of the final product. Only sophisticated diagnostic tools can fulfill the separation requirements. They have to measure directly the chemical composition or the structure of the polymer and the other constituents. Other physical methods, e.g. based on the density, thermal conductivity, or dielectric properties, failed in the separation of polymers, especially polymers from technical applications, with sufficient purity. Therefore, various standard methods of chemical analysis have been discussed in the past whereas qualitative or semiquantitative results are sufficient. Extensive investigations were performed using X-ray fluorescence, X-ray diffraction, pyrolysis mass spectrometry, laser breakdown spectroscopy (LBDS), IR spectroscopy, etc.

The identification methods have to guarantee a high throughput to provide economic operation. The instruments have to work on-line, fast, and reliably in harsh industrial environments. They are often operated automatically or by unskilled staff and should not need much servicing. On the other hand, the instruments have to identify samples which may be lacquered, partially covered and polluted, deformed, and not precisely positioned. The recorded signals are of a quality far removed from laboratory standards. Also considering these constraints, the results obtained favor IR techniques, which are described in more detail later. However, the objects of interest, i.e. the plastic pieces to be recycled, must be examined in more detail, because the different waste streams require different IR techniques.

2 MATERIALS

The different waste streams of plastic materials have to be characterized with respect to volume, type, and identification requirements, because there are considerable differences between, e.g. packaging materials from household waste and technical plastics in audio and video appliances (e.g. portable stereos, video recorders, housings of TV sets), automotive parts [fenders (bumpers)], telecommunication equipment (e.g. telephones, mobile phones, fax machines) and data processing facilities (e.g. copying machines, computers).

2.1 Waste Streams

In general, the plastic products to be handled can be divided into two main fractions:

- the packaging materials mainly in domestic (household) waste;
- the plastics used in technical equipment.

Domestic waste contains about 5–10% plastics,[9] consisting of bottles, beakers, foil, etc. These consist predominantly of five relevant polymers, PE [high-density polyethylene (PE-HD); low-density polyethylene (PE-LD); linear low-density polyethylene (PE-LLD)], PP, PS, PET, and PVC,[9] which have only to be identified and sorted. In Germany, for example, about 535 000 t of packaging materials from household wastes were collected in 1996 and sorted by the DSD.[10] Up to now these plastics parts have been mainly sorted from the waste manually and therefore no differentiation between the different polymers is implemented. However, these plastics, apart from some special-purpose formulations, contain relatively small amounts of additives and hardly any fillers. The pieces are dyed in a broad variety of colors and can contain print and labels on the surface. Screw caps are often made of a different material from the accompanying bottle. Nevertheless, a well-processed, thoroughly graded pure recyclate can closely approach the quality of virgin materials. Owing to the low weight of the materials (60% have a weight of less than 10 g),[10] the identification has to be very fast to guarantee a relevant and economic throughput. In medium-sized recycling enterprises, the total waste throughput is in the range of about 500 kg h^{-1}, which corresponds to a throughput of about 42 000 pieces h^{-1} assuming a medium weight of 12 g. This would demand an identification rate of at least 12 parts s^{-1}. Taking into account the higher price to be obtained for sorted pure plastics, an identification rate of a few parts less per second is actually still acceptable for recyclers. However, this high identification rate means that the plastic pieces have to be identified while moving, e.g. on a conveyor belt.

Compared with the above-mentioned standard polymers, technical products use a much broader variety of plastics, e.g. acrylonitrile–butadiene–styrene (ABS), polycarbonate (PC), poly(oxymethylene) (POM), styrene–acrylonitrile copolymers (SAN), polyamide (PA), polybutadiene (PB), poly(butylene terephthalate) (PBT), polyurethane (PU), poly(methyl methacrylate) (PMMA), poly(phenylene ether) (PPE), and styrene–butadiene (SB). In the case of PA there has also to be a differentiation between the various modifications, the most important being PA6, PA66, PA11, and PA6/12 (the numbers indicate the number of different condensation units). These polymers differ in the configuration

of the two monomers involved. In addition to the pure polymers, there are also many polymer blends such as ABS/PC and SB/PPE. These blends are intended to combine the different properties of the pure polymers to give a new material with advanced properties. For these polymer blends also different amounts of ingredients have to be taken into account.

To the basic polymers there are added numerous additives and fillers. Fillers include reinforcing materials such as glass or carbon fibers to increase the stiffness of polymers. However, other fillers such as talcum, chalk, titanium dioxide (TiO_2), and calcium carbonate ($CaCO_3$), are also used to increase the volume with materials cheaper than the polymer itself. Other additives include dyes (especially carbon black in the automotive and consumer electronics industries), plasticizers, pigments, stabilizers, flame retardants (mostly based on halogens), and other additives necessary for processing or modifying material properties. The surfaces of the materials are often lacquered or treated by special techniques to create, e.g. specular effects or roughness. However, heavy metals such as lead (Pb), cadmium (Cd), chromium (Cr), and mercury (Hg) are also often found in post-consumer products.

An example of the combination of many of these components is a printed circuit board. It consists normally of glass fibers, unsaturated polyester, or phenol resins and contains polybrominated aromatic carbon hydrates as flame retardants and often heavy metals such as cadmium and antimony trioxide as synergists. In many cases the surface is lacquered and partially coated by metals. All the main constituents have to be identified to allow economic recycling.

Owing to the higher weight of technical plastics compared with packaging materials from domestic waste, the identification rate can be lower. A normal TV set has a weight of about 27 kg. The pure plastic material (without printed circuit boards) constitutes about 3 kg.[11] Assuming an average weight of only 500 g for a plastic piece, an identification rate of only 1 part s^{-1} would result in an overall throughput of 1.8 t h^{-1}. However, all important components have to be identified.

2.2 Methods of Analysis

Summarizing, the information which is necessary for recycling of plastic waste comprises:

- the main polymer and composition in the case of blends;
- main additives such as plasticizers, flame retardants, etc.;
- fillers and reinforcing fibers;
- heavy metals such as cadmium (Cd) or tin (Sn).

The most promising methods to obtain this information are spectroscopic ones because they measure directly the chemical composition or the structure of the polymer and the other constituents. From the current point of view, the most successful analytical methods are:

- NIR spectroscopy;
- MIR spectroscopy;
- laser-induced pyrolysis spectroscopy (LIPS) or LBDS;
- sliding spark spectroscopy;
- pyrolysis/mass spectrometry;
- pyrolysis/IR spectroscopy;
- X-ray fluorescence spectroscopy;
- X-ray diffraction.

NIR, MIR and pyrolysis/IR spectroscopy will be described in detail later.

Using LIPS, a high-power laser is focused onto the plastic surface under investigation and generates a plasma plume, which is observed by an ultraviolet/visible (UV/VIS) spectrometer delivering information about the vaporized and ionized atoms. Therefore, this method enables one to identify atoms, predominantly metals (especially heavy metals), chlorine and bromine as indicators for halogenated flame retardants, and silicon pointing to glass fibers as reinforcing material.

Pyrolysis/mass spectrometry uses a mass spectrometer to detect vaporized and ionized molecules and atoms which are produced by a heat source. The main disadvantage is that the material has to be transferred to a vacuum by capillary tube, which tends to become plugged by impurities.

X-ray fluorescence is mainly used for the identification of PVC plastics, because it is sensitive to heavy atoms such as chlorine and insensitive to light atoms such as carbon and hydrogen, which are the main components of polymers.

X-ray diffraction analyzes the crystal structure of materials so it is limited to polymers which crystallize in sufficient amounts.

Particularly important characteristics of the methods include:

- selectivity with respect to polymers and plastics additives;
- response rate, repeating rate of the measurement;
- penetration depth into the material to be detected;
- sensitivity to disturbances and contamination;
- user friendliness, ease of automation, and costs.

Only the IR techniques and X-ray diffraction deliver direct information on the polymer and are favored for identification. Some of the requirements of recycling cannot be fully met by these methods, e.g. the identification

of metals, heavy metals, and their compounds. Especially in the case of plastics from technical products the identification will possibly require a combination of methods of analysis. The combination of, for example, an IR technique with LIPS or LBDS could be a solution for technical plastics.

3 INFRARED BANDS OF PLASTIC MATERIALS

The main polymers, important additives such as plasticizers and flame retardants, and gaseous pyrolysis products in the gas phase are mainly composed of carbon, hydrogen, oxygen, and nitrogen. The polymers comprise aliphatic (e.g. PE) and aromatic structures and various combinations (e.g. PS). Therefore, predominant bands arise from the fundamental vibrations of the related submolecular groups and their overtones and combination bands.

3.1 Principles of Vibrational Spectroscopy

The interpretation of vibrational spectra assumes the molecules to be oscillators which are excited by absorbing IR radiation. In a classical description (Equation 1) the frequencies ν of an oscillator which consists of two atoms of masses m_1 and m_2 are derived from a force constant f and the reduced mass μ:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{f}{\mu}} \qquad \mu = \frac{m_1 m_2}{m_1 + m_2} \tag{1}$$

Following a quantum mechanical description, the energy levels of the oscillator (Equation 2) are characterized by the quantum number n:

$$E_{\text{vib}} = h\nu\left(n + \tfrac{1}{2}\right) \qquad n = 0, 1, 2 \tag{2}$$

In the case of small molecules with the possibility of rotation (e.g. in the gas phase), each vibrational energy level is split up by the energy levels E_{rot} of the quantum mechanical rotator assigned to the molecule. The masses m_1 and m_2 are given by the related atoms and the force constant f by the chemical bondage between them. The simple relationships (Equations 1 and 2) explain that the combination of atoms lead to distinguishable absorption spectra of different molecules.

However, the absorption and emission of real molecules are more complicated. The complex molecules especially of long polymeric chains are assumed to build a system of coupled oscillators: The rigorous derivation of the absorption bands requires a sophisticated mathematical treatment. Various effects must be taken into account, e.g.

- the oscillators of real molecules are inharmonic, the energy levels are modified with respect to Equation (2) and allow transitions between various vibration modes;
- strongly and weakly coupled molecular groups of atoms;
- the true motion of the atoms is composed of normal modes of vibration of these groups;
- group theory determines the modes to be observed, depending on the symmetry of the groups.

The molecules absorb photons of energy ΔE_{vib} (Equation 3) from the electromagnetic field of radiation corresponding to the energy levels of the normal modes and the combinations of them, thus forming IR spectra:

$$\Delta E_{\text{vib}} = h\nu_{\text{photon}} = h\nu_i - h\nu_j \tag{3}$$

The electromagnetic fields interact directly only with polarized molecules, so causing the strong IR bands and the intensities to be proportional to the square of the derivative dipole moment with respect to the coordinate of the normal vibration mode. An effect of second order, the Raman effect (which is not discussed further here), is encountered if the radiation field interacts with the dipole moment which was transiently induced by the electromagnetic field itself. Molecules which reveal a fully symmetric distribution of the electrons such as molecular nitrogen (N_2), molecular oxygen (O_2), sodium chloride (NaCl) or potassium bromide (KBr) show no permanent dipole moment, and are inactive with respect to IR radiation.

To describe the spectra of molecules in the IR region, the vibration frequencies ν (s^{-1}) are expressed by wavenumbers $\bar{\nu}$ (cm^{-1}) and/or wavelengths λ (μm) where Equation (4) holds:

$$\bar{\nu} = \frac{\nu}{c} = \frac{1}{\lambda} \tag{4}$$

The wavelength of IR radiation ranges from 700 nm to 50 μm, corresponding to wavenumbers from 14 000 to 200 cm^{-1}. Wavelengths up to 2.5 μm are assigned to the NIR and wavelengths from 2.5 to 20 μm to the MIR region.

The IR radiation of intensity I penetrating a material of thickness Δx is diminished by ΔI by the vibrations with absorptivity α of component concentration c^* following the simple Beer–Lambert–Bouguer relationship (Equations 5 and 6):

$$\Delta I = -\alpha c^* I \Delta x \tag{5}$$

$$T = \frac{I}{I_0} = e^{-\alpha c^* d} \tag{6}$$

The transmittance T describes the response of the material with a thickness d related to the initial radiation intensity I_0. It is often expressed as a percentage, which means that T is multiplied by 100. Forming the logarithm of T, the absorbance $\alpha c^* d$ is obtained, which is linear in concentration and sample thickness (Equation 7):

$$\alpha c^* d = \ln I - \ln I_0 \tag{7}$$

A sample reflects light owing to various effects. For example, light is partially reflected at phase boundaries of different materials (e.g. air–solid sample) and from the substrate. It is scattered at inhomogeneities, impurities, and additives in the sample and gives rise to diffuse reflectance. The Kubelka–Munk theory models the diffuse reflection described in detail, e.g. by Kortüm.[12] The specular reflectance applied at small angles of incidence is obtained from the sample surface, and is governed by dispersion in the dielectric material. The evaluation of the spectra obtained has to account for the complex refractive index by applying a Kramers–Kronig transformation.

3.2 Spectra of Plastics in the Infrared and Near-infrared Wavelength Range

As mentioned above, plastic materials consist of the macromolecular polymers, fillers, plasticizers, and important additives such as flame retardants. Engineering plastics also include fiber-reinforced composites and polymeric resins and the organic constituents are predominately composed of C, H, O, and N. The constitution of the polymers comprises aliphatic [e.g. PE $(CH_2)_x$, PVC] and aromatic parts (e.g. PS) in various combinations: Structure **(1)**, PE; Structure **(2)**, PS; Structure **(3)**, PVC; Structure **(4)**, TBBA (tetrabromobisphenol A).

The interpretation of the spectra of complicated molecules is based on the concept of group frequencies. It assumes that submolecular groups of atoms are coupled weakly to other submolecular groups. The absorption bands of these groups are called characteristic bands as they always appear at the same wavelength position, only slightly modified by the weak interaction with neighboring groups. The bands characteristic of the individual molecules are called "fingerprint" bands.

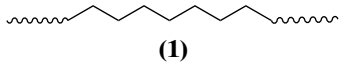

(1)

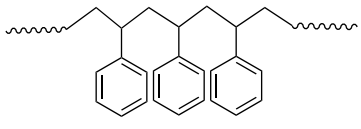

(2)

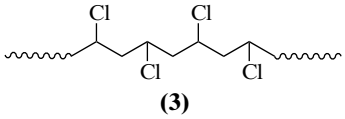

(3)

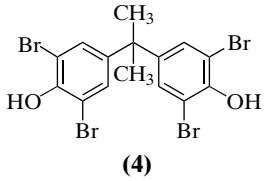

(4)

The most important bands of organic compounds arise from the CH, CH_2, and CH_3 groups. These vibrations are built of the fundamental modes of motion between C and H atoms:

- stretching;
- in-plane deformation;
- out-of-plane deformation;
- wagging;
- twisting; and
- rocking.

The fundamental stretching modes of CH arise near the wavenumber 2900 cm^{-1} (wavelength 3.4 μm). The symmetric and asymmetric modes of PE lie at 2930 and 2850 cm^{-1}. The modes of CH groups neighboring C=C bonds are found at a higher wavenumber of 3100 cm^{-1} (3.2 μm), which is also valid for CH groups in aromatic rings (e.g. PS at 3000–3100 cm^{-1}). The wagging bands lie between 1180 and 1350 cm^{-1}. In chains containing CH_2 the twisting bands appear in the same wavenumber region. The rocking vibrations cause a characteristic band for long chains (e.g. PE) at 720 cm^{-1}. The spectra of a PE and a PS foil (5 μm thickness) in the MIR region are plotted in Figure 5(a) and (b) respectively.

Polymers such as PA and PU contain, in addition, NH groups. The analogous stretching modes of NH absorb near 3300 cm^{-1} (3.0 μm). OH bands are found in polymers such as cellulose acetate (CA) at 3500 cm^{-1} (2.8 μm) (see Figure 13).

In standard chemical analysis, the "fingerprint" region begins close to 5 μm and extends to longer wavelengths and shows a broad variety of bands of various origins, including especially bands with other elements such as Cl (e.g. in PVC). Important groups for identification are those of C–C, C=O, C–O–C, C–N, etc.

Gaseous species with low molecular weight show vibration band profiles broadened by rotation in low resolution. At better resolution a large number of rotational lines accompany the vibrational lines.

If the submolecular groups of polymers were to build perfect harmonic oscillators, there would occur only transitions between adjacent energy levels of the fundamental modes. Combinations of different normal modes would be forbidden. The inharmonicity of the molecular oscillators allows the appearance of overtones and combination bands. They arise when the energy of an absorbed photon is used for two or more energy quanta

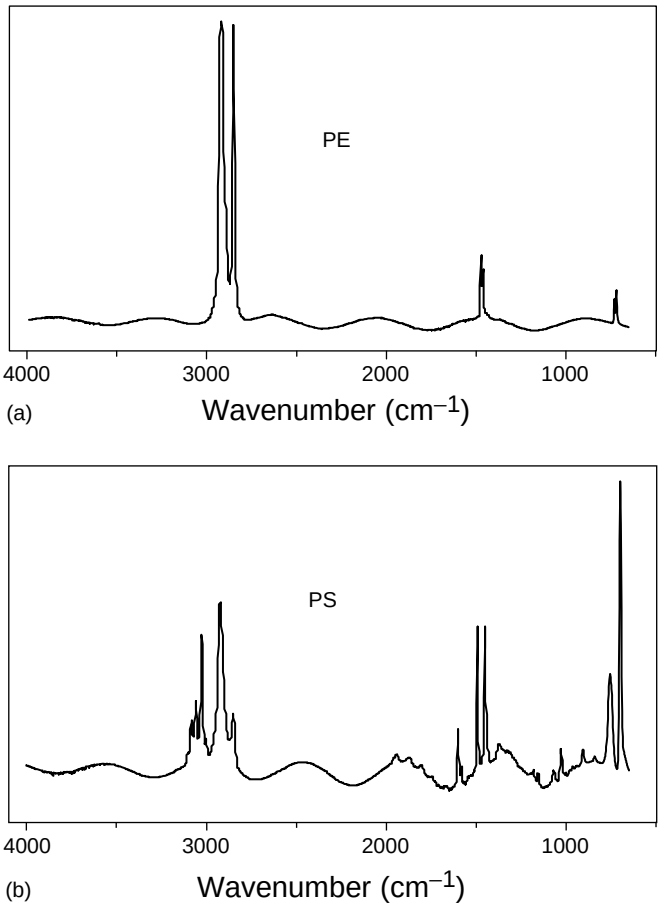

Figure 5 MIR transmission spectra of (a) a PE and (b) a PS foil.

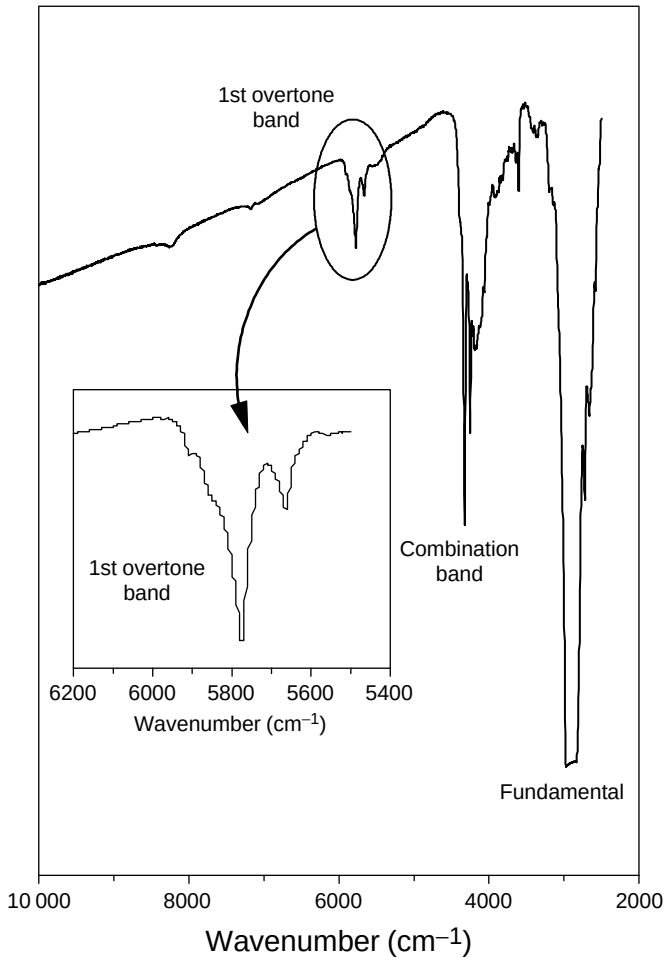

Figure 6 Transmission of an industrial PE/PP foil containing predominantly PE in the MIR/NIR spectral region, showing weak overtones and combination bands in the NIR spectrum (inset).

of the same normal mode or is split up for the excitation of two or more different normal modes. However, the absorbances of the overtones and combination vibrations are still strongly reduced compared with the fundamental vibrations. This effect is shown by the transmission spectrum of a PE/PP foil in Figure 6, where the first overtone is compared with the intensity of the fundamental mode which is in saturation.

In the NIR wavelength region, the vibrational bands are mainly generated by the overtones of stretching vibrations of groups containing hydrogen and combinations of these stretching vibrations with wagging and/or rocking vibrations. This is obvious from Equation (2), which indicates that the bands where atoms of low weight are involved appear at shorter wavelengths in the IR spectrum. A rough estimate of the wavenumber positions of the overtones and combination bands is obtained by adding the wavenumbers of the participating fundamental modes.

The first overtone of the CH stretching mode appears near 1.7 μm and the second near 1.2 μm. Comparable to the fundamental vibration, the overtones are also observed at shorter wavelengths if the CH functional group is in the neighborhood of groups such as C=C or in an aromatic C–C bond. For example, the aromatic CH band of PS or PMMA arises at 1.675 μm compared with the band of PE near 1.72 μm. Combination bands lie between the fundamental vibrations and the various overtones (e.g. 2.3 and 1.4 μm). The first overtone of NH is located near 1.5 μm and the second near 1.1 μm.

The basic concepts of vibrational spectroscopy are described here rather crudely and include only aspects to support the understanding of the classification of the plastic materials and to emphasize advantages of wavelength regions or measurement techniques to be applied. Details are described in standard textbooks of the theory of IR spectroscopy, especially of polymers [e.g. Hertzberg,[13,14] Steele,[15] Siesler and Holland-Moritz,[16] and Burns and Ciurczak[17]]. An overview on the broad variety of polymers can be obtained in collections of spectra (e.g. Aldrich Library).[18] The Sadtler Atlas[19]

and the Atlas of Buback and Vögele[20] for NIR spectra contain only a few spectra of polymers. A comprehensive collection of a variety of NIR spectra of polymers can also be found in Siesler and Holland-Moritz.[16]

3.3 The Spectra of Industrial Plastic Samples

Solid samples of organic compounds, especially their fundamental vibrations, absorb MIR radiation strongly. In addition, an increase in the thickness of the material to be penetrated by the radiation leads to an exponential decay of the transmitted intensity, as is obvious from Equations (6) and (7). Therefore, in the laboratory mainly thin films of thickness well below 100 μm or diluted samples in inactive material, e.g. KBr, are investigated to avoid saturation effects. The MIR spectral region can be applied to materials with small penetration depths of radiation. For example, fillers, pigments, and fibers can reduce the pathlength of radiation drastically by absorption, reflection, and scattering. This is especially true for carbon black, which is an important filler for black plastics, and strongly absorbs radiation of a broad wavelength range (blackbody radiation) already close to the surface.[21–23]

3.4 Near-infrared Spectra

In the NIR spectral range, the overtones and combination bands absorb orders of magnitude (10–100) less light when compared with the fundamental vibrations. Therefore, this wavelength region between 1 and 2.5 μm (wavenumber 10 000–4000 cm^{-1}) is appropriate for analyzing bulky materials in reflection and transmission when radiation penetrates a few millimeters into the material. The situation is fairly simple in the NIR region where mainly diffuse and partially regular reflection and superimposed transmission from substrates opposite to the illumination can contribute to the spectrum.

There is no need to distinguish between the different contributions to the spectra because all give rise to a transmission-like spectrum. For identification only qualitative spectra are needed. Therefore, in the figures only transmission-like spectra are plotted. Owing to the strongly varying conditions in real identification measurements the ordinate is arbitrary as no clear correlation with the illumination I_0 is possible in many cases. These properties have attracted much attention to apply NIR spectroscopy to plastic recycling.[24–55]

Qualitative reflection spectra of various plastic materials are shown in Figure 7(a) and (b), recorded with an AOTF spectrometer from engineering plastics. The polymeric samples differ in the structural composition of aliphatic and aromatic groups. The CH bands absorb in the wavelength regions 1100–1250 nm (second overtone of CH stretching mode), 1300–1450 nm (CH combination band), 1600–1800 nm (first overtone of CH stretching mode), and 2150–2500 nm (combination band). The CH bands show a significant complex shape depending on the molecular structures which allow a clear identification of, e.g. PE, PVC, PP, PS, ABS, PET, PC, PBT, and others.[16,17,19,20,24–31,34–46]

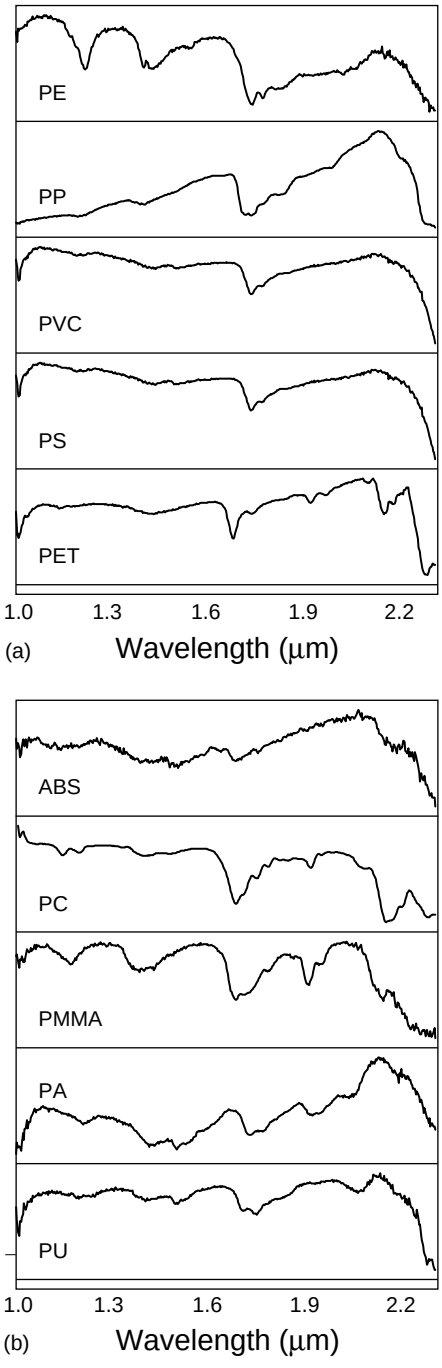

Figure 7 NIR spectra of (a) five plastic materials mainly present in household wastes and (b) five plastic materials present in electronic product wastes, recorded with a scan speed of 150 nm ms^{-1}.

There are only a few functional groups of molecules without hydrogen which show strong NIR absorption, e.g. CO. CO bands can be observed at 1900–2000 nm and are found, e.g. in the spectra of PMMA, PET and PC.[16,17,19,20,24–35]

NH groups are important in polymers such as PU and PA and the related bands appear distinctly in the spectra. The first overtone dominates near 1500 nm and overlaps the CH combination band at this position and a combination band is observed near 2050 nm.

The different polymers can be distinguished clearly, as proved by using multicomponent analysis. The plastics from domestic wastes mainly consist of various types of bottles, beakers, foils, etc. and are made of the five polymers PE, PP, PS, PET, and PVC, the last being used in decreasing amounts. Each piece is produced as a nearly homogeneous material with thickness between 100 μm and 1 mm, contains only few additives, and gives rise to spectra of good quality and high intensity in the NIR region. Dyes and prints do not substantially distort the spectra. Therefore, they differ strongly from engineering plastics. They can be reliably identified by a single band in the NIR spectrum, e.g. by the CH overtone between 1600 and 1800 nm if engineering plastic types are ignored (see Figure 8). Spectra of moving bottles or beakers measured with scan rates of 200 nm ms^{-1} are plotted in Figure 9(a) and (b). Extending the wavelength range from 1.1 to 1.8 or

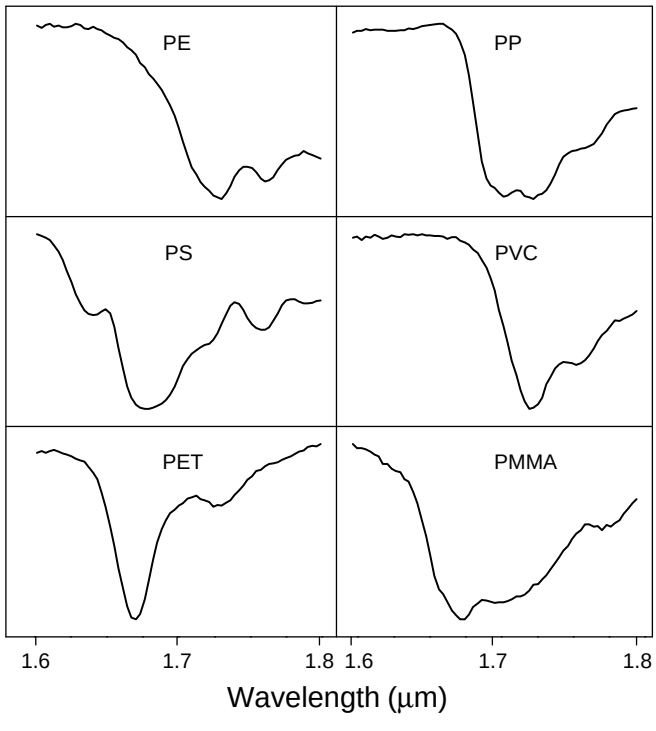

Figure 8 First overtone of plastics predominantly found in household waste.

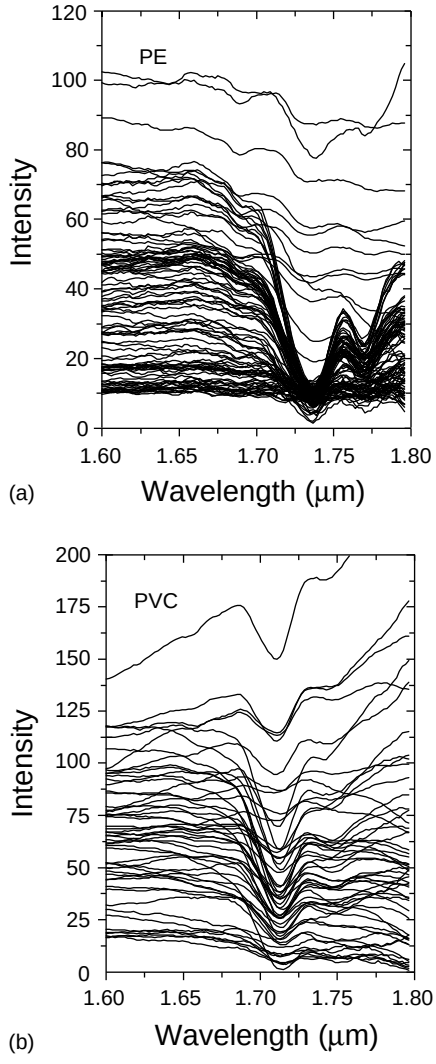

Figure 9 First overtone of CH stretching mode of bottles and beakers passing the sensor head of an AOTF spectrometer at a speed of 1 m s^{-1} (scan rate 200 nm ms^{-1}). (a) PE; (b) PVC.

even to 2.3 μm, nonblack engineering plastics containing fillers and additives (present in higher quantities) can also be identified (see Figure 7). Figure 10 shows NIR spectra in the range of 1.1–1.8 μm and demonstrates that a flame retardant TBBA could be identified in a nondyed ABS granulate. Also the plasticizer in soft PVC, which is often a phthalate, can be detected owing to its aromatic band.

The processing and the surface treatment influence the spectra to be obtained from plastic materials. Fillers also affect the spectra strongly and can absorb, reflect, and scatter NIR light. These effects result in baseline shifting and twisting. The highly transparent plastics PMMA, PC, and PE show flat baselines. PP often contains inhomogeneities which scatter the light. Fillers such as glass fibers, chalk, talcum, and pigments also scatter light. The scattering follows Mie's law, which means that it

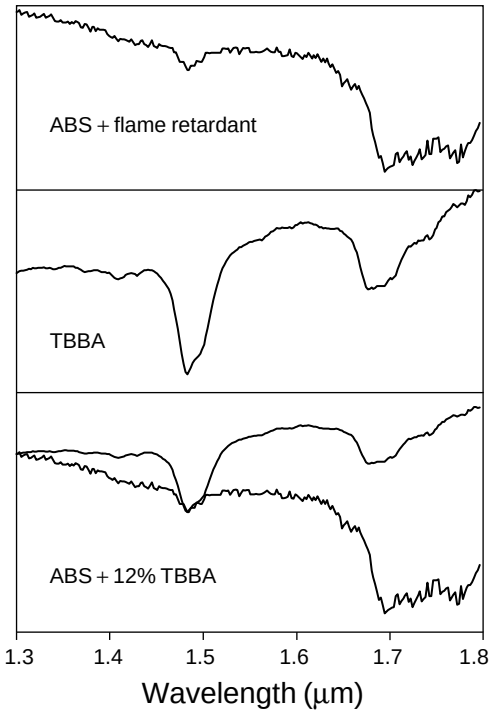

Figure 10 NIR spectrum of ABS and the flame retardant TBBA, indicating that about 12% TBBA is present in the plastic.

increases strongly towards shorter wavelengths, which even allows an estimation of particle size. Blazing effects on the surface reflect light in special directions and can complicate light detection. These effects also result in baseline shift and twisting. Thin, transparent lacquers and prints (<10 μm) do not prevent the acquisition of useful spectra.

3.5 Mid-infrared Spectra

In the case of black plastics which are often used, for example, in audio, video, and automotive applications, the NIR spectral range is inapplicable owing to the reduced penetration depth of radiation into these materials. The MIR spectral range is appropriate for technical plastics in cases when the plastics contain high concentrations of glass fibers or fillers which reduce the transparency, especially in the case of black materials.[58–80]

The spectra are comparable to normal reflectance spectra, revealing the bands of fundamental vibrations of CH and NH stretching motion at 3.3 and 3.4 μm and the highly specific bands from the ''fingerprint'' region at wavelengths above 5 μm, including also CC, C=O, C–O–C and CX groups (X = end group) or chain segments. Many spectra of samples collected from used engineering plastics can be found in the literature.[45,58,66,67,80] Additives for processing[81] and those present in higher amounts such as plasticizers and flame retardants can also be analyzed via the MIR bands.

Carbon black is used to obtain black plastic material. In this case the penetration depth of radiation reaches only to the micrometer range. If thin foils are measured, a wave-like structure generated by internal interferences modulates the baseline of the spectrum (see, e.g. Kampffmeyer and Pfeil[82]). Spectra from reflection measurements are more complicated. Different effects can overlap and distort the spectra (see Figure 11): The reflected light can be composed of specular reflectance from the surface, diffuse reflectance from pigments and inhomogeneities from the interior, surface scattering, and even transflectance.[21–23] The band peaks of the specular reflectance spectra show a shape like a derivative spectrum, owing to refractive index dispersion. A Kramers–Kronig transformation[56,57] produces peak shapes similar to ''normal'' absorption index spectra, which can be compared with spectra obtained by well-defined sample preparation. Spectra of two important black engineering plastics are plotted in Figure 12(a) and (b), showing derivative-like peak shapes of the raw

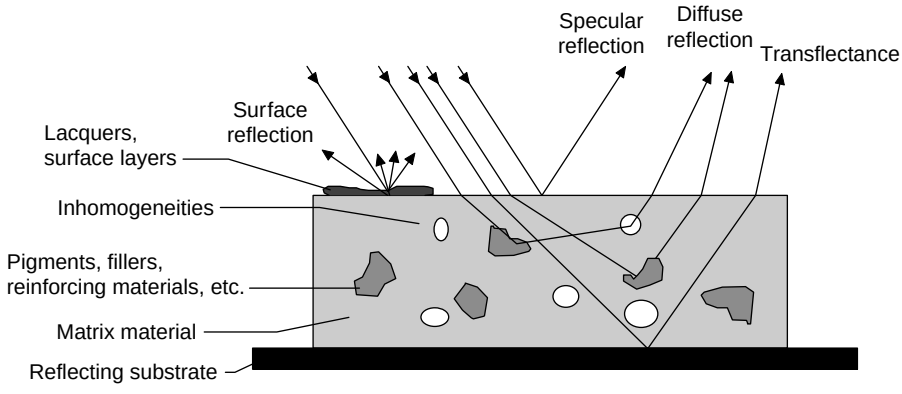

Figure 11 Scheme of the various effects that contribute to reflection spectra.

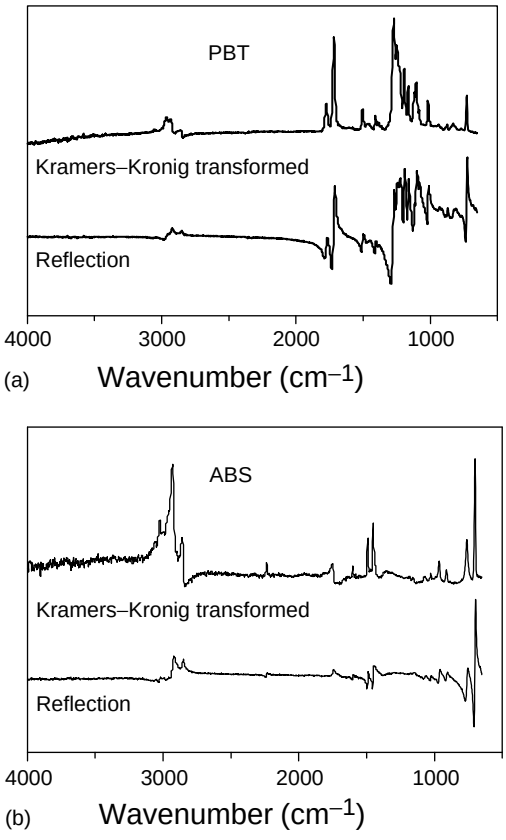

Figure 12 MIR spectra of (a) PBT and (b) ABS as examples of industrial polymer samples.

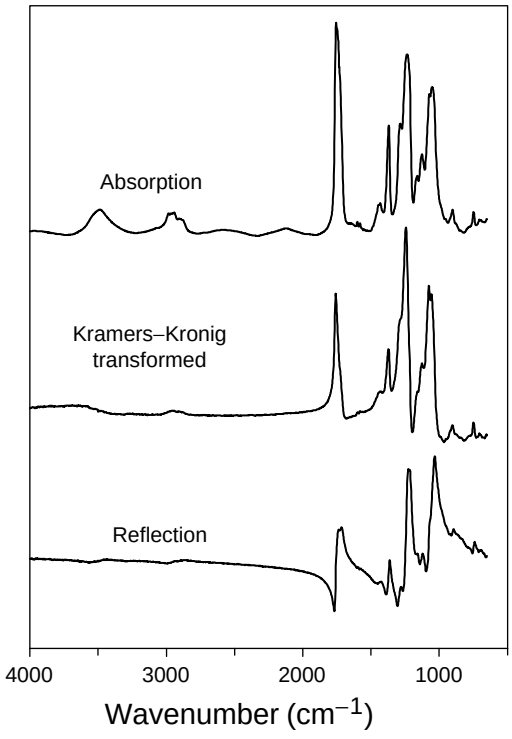

Figure 13 MIR spectra of CA: absorption, Kramers–Kronig transformed reflection, and reflection spectra.

spectra and those obtained after Kramers–Kronig transformation. In addition, these diagrams show that defects in the spectra indicate the distorting influences described above. Figure 7(a) and (b) show the wave-like structure found in the spectra of thin foils. A comparison of the spectra of black CA from a thin foil (5 µm) and bulky material is shown in Figure 13 together with the Kramers–Kronig transformation of the latter. Many spectra obtained from important engineering plastics used in cars and audio and video products can be found in the literature.[56,64,65]

If the main polymer has to be identified and if copolymers do not have to be sorted out, the bands near 3.4 and 3.5 µm are sufficient for identification. A detailed analysis including the determination of additives such as the plasticizers, special fillers such as barium sulfate ($BaSO_4$), and polybrominated flame retardants (e.g. in PBT or ABS) requires the measurement of the "fingerprint region". Also the black plastics show more intense bands in that wavelength region. For some blends (e.g. ABS/PC and PBT/PC) and copolymers, even semiquantitative results can be derived. Materials of similar composition such as PA6, PA66, and PA12 can be distinguished.

The spectra in the MIR region are even more influenced by pollution and lacquers than those in the NIR region. The measurement of reliable spectra from the bulk polymer requires surface cleaning and treatment. It is recommended to remove the surface layer and to treat the measuring spot of the material.

3.6 Gas-phase Spectra

The treatment of materials by various heating methods, such as heating by contact with hot devices, flash heating, or shock heating, produces evolved gases. These are composed of vapors, fragments, and oxidation products of the substance under investigation. The application of controlled heating or pulse heating in air or in closed sample cells is established as a method to analyze the composition of the material and to investigate its thermal behavior.[83] In industrial environments also vapors and fragments of pyrolyzed pollutants, fillers, and additives will overlap the desired gases of the main polymers. In general, the resulting spectra should be indicative of the polymer under investigation but detailed studies on this technique are ongoing.[84,85]

Gaseous species of low molecular weight show vibration frequencies of similar groups such as condensed substances. The rotational lines lead to strong broadening

of the vibrational lines if measured at low wavelength resolution. At higher resolution a large number of rotational lines accompany each vibrational line. In air, water vapor and carbon dioxide from the environment or from partial oxidation often overlap or even dominate the spectra in the MIR region. The MIR region is useful for analyzing gas phases, because the low density or low species concentration would require very long pathlengths if the overtones or combination vibrations were applied (see the description of the principles of IR spectroscopy).

4 IDENTIFICATION MEASUREMENT SYSTEMS

The standard instruments to measure the NIR and MIR spectra of solid or gaseous samples have been widely used in laboratories for detailed chemical analysis. Applications in chemical engineering for on-line/in-line measurements for process control are increasing continuously. The original instruments were not adequate to work under the conditions of rapid qualitative analysis in an industrial environment where the identification systems have to be operated for recycling of plastics. These conditions required the development of new optics, spectrometers, photodetectors, and data evaluation algorithms or at least substantial modification of standard equipment. Appropriate components were not commercially available when the activities started at the beginning of the 1990s.

The identification of plastics from household waste requires fast scanning techniques, especially if many spectra are to be sampled for the identification of one piece while moving. Full spectra, possibly of a limited wavelength region, have to be acquired in a few milliseconds from materials of good spectral quality at a thickness of 100 μm–1 mm. These constraints favor the NIR spectral range. Although some investigations are based on Fourier transform near-infrared (FTNIR) spectrometers,[34–36] measurement during movement of the sample makes it difficult to apply. Mainly the following spectrometer types are used:

- AOTF spectrometers;[24–33,48,49]
- grating spectrometers with diode arrays;[37–46]
- reference sample absorbance at characteristic wavelengths.[53–55]

These methods allow high scan rates of sufficient spectral elements in the range of milliseconds and suitable wavelength resolution. They also fulfill the criteria of robustness in a harsh industrial environment.

The AOTF has only scarcely been used in laboratory equipment so it is described here in more detail. It uses the anisotropic Bragg effect at large angles. This effect is observed on studying birefringent crystals such as tellurium dioxide (TeO_2).[86] Ultrasonic waves generated by piezoelectric transducers form a grating of different material densities in the crystal, which diffracts incident light to an ordinary and an extraordinary beam. If the frequency of the acoustic wave is changed, the wavelength changes owing to the constant sound velocity of the crystal (650 m s^{-1}). In Figure 14 the principle of light diffraction by an AOTF crystal is plotted. The half wavelength corresponds to the lattice distance, which causes the Bragg diffraction. Oversimplified, a spectrum is scanned by an AOTF by varying the distances of the ''grooves'' and measuring the different wavelengths at a fixed angle and not rotating the grating (conventional grating spectroscopy). A detailed theoretical description of the effect was given by Yano and Watanabe[86] and Chang.[87] Based on a fast-scanning AOTF spectrometer,[88–91] the first versions of plastic identification units which met the requirements were presented at fairs (SENSORS 1988, ACHEMA 1988) and first results were published later.[24–27]

Other groups used diode-array spectrometers,[37–46] which are commercially available. The arrays comprise 128–1024 diodes which acquire simultaneously the full spectrum but the signals from the diodes are multiplexed and fed to a single amplifier. The results obtained are similar to those obtained by AOTF spectrometers.

The method to analyze the reflected or transmitted NIR radiation after further transmission through reference samples is an analogous correlation technique.[54,55] The evaluation for identification takes into account that the transmission through an inappropriate reference material reduces integrated intensity more than that through an appropriate material.

The NIR spectral range is partially covered by the following photodetectors. The uncooled germanium (Ge) photodiode is sensitive up to 1.8 μm and can measure the first CH overtone at 1.7 μm. Indium gallium arsenide (InGaAs) detectors record light up to 1.7 μm, and a

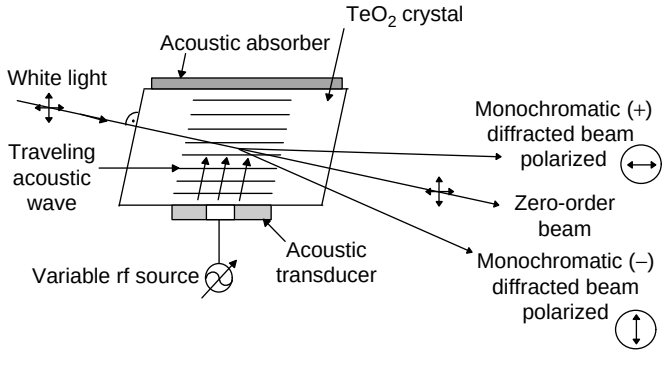

Figure 14 Principle of an AOTF.

spectrometer using it has to work with the second overtone and/or the combination band at 1.4 μm. InGaAs photodetectors are available with extended wavelength sensitivity to cover the spectral range up to 2.4 μm, but only recently, and they are expensive, which is important if arrays are applied. Cooling these types of photodetectors increases strongly the signal-to-noise ratio, often referred to as the detectivity, d^*. Other photodiodes such as lead sulfide (PbS) detectors are in principle appropriate for the NIR spectroscopy but are relatively slow.

Photodetectors for the MIR spectral region are predominantly cooled. For example, the indium arsenide (InAs) (cut-off wavelength 2.7 μm) and cooled indium antimonide (InSb) (5.5 μm) photodiodes can also be used for detection up to 2.5 μm. At higher wavelengths, mercury cadmium telluride (HgCdTe) (14 μm) is a sensitive detector which is often used. In addition, pyroelectrical sensors are applied. The cooling is performed mainly by thermoelectric effects or using Dewar vessels containing liquid nitrogen.

A typical identification unit for plastics from packaging materials in domestic waste has an enlarged measuring area and allows the simultaneous observation of reflected and transmitted light of moving samples. The light source consists of several lamps, which illuminate the measuring area ($\phi = 100$ mm) from different directions in order to increase the probability that an arbitrarily positioned moving sample reflects directly into the collecting optics, which could be, for example, a quartz condensing lens (Figure 15). In Figure 10(a) and (b) a series of spectra are plotted recorded during the fall of a PE and a PVC bottle through the sensitive area of the spectrometer. When the plastic piece enters or leaves the sensitive area

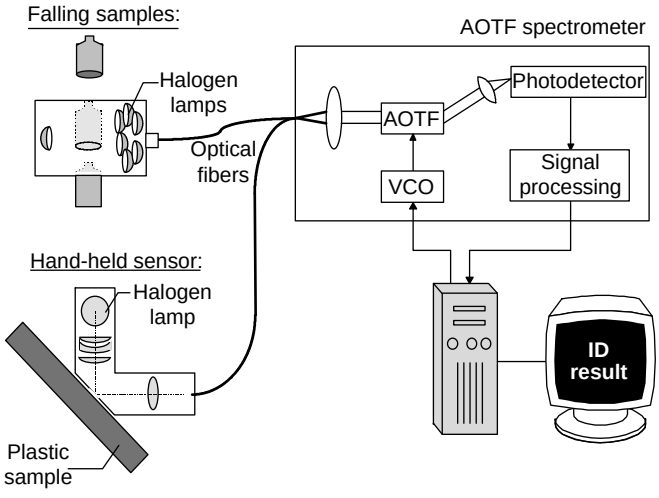

Figure 15 Scheme of an AOTF spectrometer (right) identifying falling samples (upper left) and its application for a hand-held sensor (lower left) VCO = voltage-controlled oscillator.

of the spectrometer, strong arbitrary deviations of the baseline are found. During the time interval when the plastic covers a large area of the sensitive illuminated zone of the detector head, spectra of similar quality are obtained and form a bundle of the characteristic band if plotted in one diagram. For the identification algorithm, for example, 20–50 spectra are used which allow the elimination of spectra strongly disturbed by unwanted effects.[27–32] For a moving piece the different spectra are obtained from different spots on it. This increases the probability of obtaining most of the spectra from the basic material and not arbitrarily from polluted material or inactive labels. Water droplets on the sample surface do not really distort the spectra substantially.[92]

The NIR spectral range is also useful for the identification of non-black engineering plastics.[24,26–32,48–52] In this case, FTNIR spectrometers are also applied.[40,51,52] A typical fiber-optic detector head provides a fixed measuring plane and can be moved by hand or by a robot to the sample. The light is focused on the sample (focal area $\phi = 5$ mm) by an appropriate optic, e.g. a quartz condensing lens or an out-of-plane parabolic mirror. A second optic (e.g. a quartz condensing lens, out-of-plane parabolic or elliptical mirror or Cassegrainian optic) collects the reflected light at an angle of e.g. 90° relative to the incident light and transfers it via a fiber optic to the spectrometer (Figure 15). The spectra plotted in Figure 7(a) and (b) were measured by such an identification unit.

The given examples of identification systems are not the only versions which would work. They only represent some comprehensive solutions which fulfill the following requirements:

- to account for the variable position of the sample and sample surface;
- to collect as much light in reflection or transflection of the sample as possible;
- to deal with samples which are partially contaminated or covered by labels or undefined materials.

The IR spectral range is adequate to identify engineering plastics, especially if these are black. Currently, FTIR spectrometers are used to identify plastics in the MIR spectral region.[56–78]

FTIR spectrometers analyze light by measuring the Fourier transformed intensity of a spectrum with the use of a Michelson interferometer or modified versions of it. The interferogram is obtained by a scanning mirror and the spectrum obtained by a Fourier transformation. A scheme of the basic optical configuration is shown in Figure 16. The main advantage of this technique arises from the fact that the photodetector receives light from all spectral elements simultaneously at each position of the mirror scan. In addition, the radiation

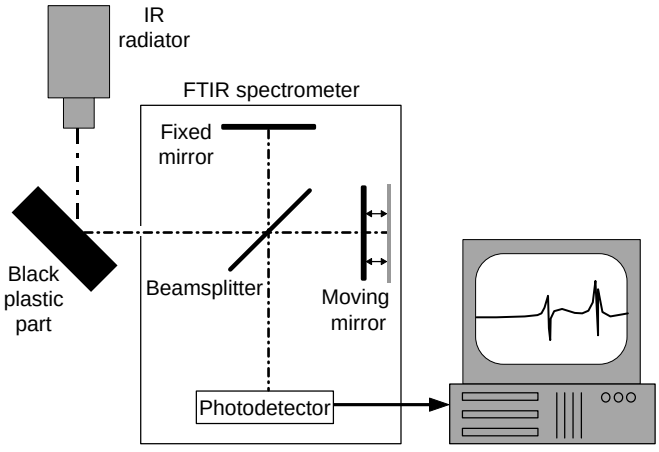

Figure 16 Operation scheme of an FTIR spectrometer based on a Michelson interferometer.

to be analyzed does not have to pass an entrance slit as in the case of a grating spectrometer. However, the Fourier transform technique is sensitive to changes in the spectrum during one scan because these changes are interpreted as a contribution to the complete wavelength range. In addition, the interferometer configuration has to be isolated against external mechanical vibrations or shocks, which is difficult to realize in an industrial environment. Performing the Fourier transformation takes time, hence FTIR spectroscopy is slow, especially if the Kramers–Kronig transformation also has to be applied. Typically up to 4 s are needed to obtain one spectrum. Its application to technical plastics is possible because these pieces are heavier and can therefore be identified in a fixed position. They allow a throughput sufficient for economical identification. However, the identification takes place at one spot on the total plastic piece. Redundant identification would require multiple time intervals.

FTIR spectrometers are suitable instruments to identify plastics in the MIR region, especially in the case of black plastics, although an AOTF spectrometer was used recently.[45] Also, if fillers and flame retardants have to be analyzed, the MIR region, where the full range or at least the ''fingerprint'' wavelength range has to be scanned, gives reliable results.

Identification using FTIR spectrometers is achieved by measuring the diffuse reflection or the specular reflection. The sensor optics are completely built by mirrors. Normally the parallel light beam which has already been transmitted by the spectrometer is focused on the sample and the reflected light is then focused on the detector. Fiber optics are difficult to apply as the fluoride fibers for the MIR region exhibit many disadvantages (brittle, low transparency, limited wavelength range, expensive) which limit their use. Hence the sensor head is normally positioned at a fixed position and provides a plane sensitive area where the sample has to be placed.

The high absorbance of condensed materials in the MIR region by the fundamental modes is of great advantage for the identification of technical engineering plastics where only small penetration depths are possible. However, surface coatings, pollutants, and moisture also absorb strongly and can prevent the acquisition of spectra from the basic material. Lacquers often have thicknesses of more than 10 μm and can complicate the extraction of the desired spectrum.[56,57] Consequently, surface cleaning and repositioning of lacquers or paints as a pretreatment is needed at the spot of identification. Techniques used are milling, brushing, etc.

FTIR spectroscopy is also applied to analyze the gases evolved from pyrolyzed plastic materials.[84,85] The plastic sample under investigation is heated by some method and the evolved gas can accumulate in a cell directly above it or be guided by a tube to an optical cell. The composition of the gas is analyzed via the transmission spectra recorded by the FTIR spectrometer. The spectra are specific for the polymers. The spectrometer configuration can be close to laboratory equipment.

Heavy metals and types of fillers and flame retardants are often difficult to identify by IR techniques. Hence it should be considered whether or not it would be a good idea to combine the IR analysis with another technique, e.g. LBDS.[31,32,39]

5 IDENTIFICATION AND PERFORMANCE

Identification for recycling means that qualitative spectra are compared with reference spectra and some algorithm decides which polymer containing which main additive is actually to be recognized. Identification by IR techniques allows a positive sorting because the algorithms search for the presence of patterns in the spectra definitely assigned to a specific material. The probability of deciding correctly can be increased by decreasing the maximum difference between the reference pattern and the test pattern, which is the defined upper limit for agreement. There arises a strong coupling of the handling techniques of the plastic pieces to be identified, allowing various rates of throughput and user-defined purity (the quality or contamination of the samples) of the recovered materials. Of course, the number of plastic types to be sorted out has to be limited to those available in the waste stream in a sufficient volume for further economic processing.

5.1 Identification Algorithms

The identification algorithms were taken from existing methods of pattern recognition, modeling, and statistical

description of data. Taking into account that only qualitative or semiquantitative information is required at a high speed, simple algorithms such as forming the differences of normalized spectra are sufficient in comparing measured spectra with stored reference spectra. The performance of microprocessors allows also the application of more complex algorithms. The simplest linear methods are described here in more detail.

It is assumed that a spectrum $y_{\text{test}}(x_i)$ is measured at n wavelength positions ($i = 1, \ldots, n$) and compared with each spectrum of a set of m reference samples $y_{\text{ref},j}(x_i)$ ($j = 1, \ldots, m$). Data pretreatment can comprise a baseline subtraction, a normalization of the spectra to 0 and 1, and a subtraction of the average of the spectrum (Equation 8):

$$\chi_j = \sum_{i=1}^{n} |y_{\text{test}}(x_i) - y_{\text{ref},j}(x_i)| \qquad j = 1, \ldots, m \tag{8}$$

Comparison of the sum of the norm of the difference spectra[94] leads to a fast algorithm which contains only additions and subtractions.[30] The decision is made on the smallest χ_j obtained for reference $y_{\text{ref},j}$ but exceeding a user-defined threshold χ_t. Taking into account the nonideal samples of used plastics, it is obvious from this procedure that small values of χ_t produce a high purity of sorted plastic materials and higher values increase the positively sorted material but with an increased number of erroneously identified pieces. The algorithm is remarkably stable and reliable also if small differences in the spectra are to be detected.

Multilinear regression can be derived by minimizing the sum of squared differences (χ^2 function) of reference and test spectra with respect to some values a_j which are obtained by differentiating χ^2 (Equations 9–12):

$$\chi^2 = \sum_{i=1}^{n} \left[y_{\text{test}}(x_i) - \sum_{j=1}^{m} a_j y_{\text{ref},j}(x_i) \right]^2 \tag{9}$$

$$0 = \frac{\partial \chi^2}{\partial a_j} = \sum_{i=1}^{n} \left[y_{\text{test}}(x_i) - \sum_{j=1}^{m} a_j y_{\text{ref},j}(x_i) \right] y_{\text{ref},k}(x_k) \qquad k = 1, \ldots, m \tag{10}$$

$$\alpha_{k,j} = \sum_{i=1}^{n} y_{\text{ref},j}(x_i) y_{\text{ref},k}(x_i);$$

$$\beta_j = \sum_{i=1}^{n} y_{\text{test},j}(x_i) y_{\text{ref},j}(x_i) \qquad k, j = 1, \ldots, m \tag{11}$$

$$a_j = \sum \alpha_{j,k}^{-1} \beta_k \qquad j = 1, \ldots, m \tag{12}$$

When testing a plastic material k, the value of a_k should be close to 1 and all others close to 0. The user can define the level near 1 which is accepted to identify positively material k and therefore define the purity of the sorted output. In comparison with Equation (8), least-squares procedures require a multiplication of two vectors and of a matrix with a vector. The inverse matrix $\alpha - 1$ can be computed from the reference spectra before the operation. If α is not singular, the m reference spectra and therefore also the corresponding plastic types can in principle be distinguished. A modified version is obtained by assuming that the tested spectrum is composed of a linear combination of the reference spectra (Equations 13 and 14):

$$y(x_i) = \sum_{j=1}^{m} a_j y_{\text{ref},j}(x_i) \qquad i = 1, \ldots, n \tag{13}$$

$$a_j = \sum_{i=1}^{n} y_{\text{test}}(x_i) \; y_{\text{ref},j}(x_i)^{-1} \qquad j = 1, \ldots, m \tag{14}$$

The plastics are distinguishable if the rank of the $n \times m$ matrix $[y_{\text{ref},j}(x_i)]^{-1}$ is m. Blends are indicated if two or more a_j are significantly different from 0. The procedures described in Equations (9–14) follow the outline of Press et al.[94] They allow one also to select the wavelength regions needed for a reliable differentiation of the reference plastics by varying them and testing the rank of the $\alpha_{k,j}$ and/or $y_{\text{ref},j}$. They were applied when testing if the above-mentioned types of plastics are distinguishable within selected wavelength regions. However, such a result does not imply that these plastics can be distinguished from all other plastic types under any conditions.

There are various methods available to model NIR and MIR data which were developed to analyze complex mixtures of chemical compounds. The linear ones are summarized by the multivariate data evaluation concepts of chemometrics where principal component analysis (PCA) and partial least squares (PLS) are powerful methods (see, e.g. Martens and Naes).[95] These methods were applied to sorting plastics in household waste and also for engineering plastics. A nonlinear generalization of these linear methods using multilayers is the neural networks,[34,35,42–46,65] which have also been tested in some cases together with fuzzy logic concepts.[44–46]

In reality, the identification has to be reduced to a limited number of plastic types because it is not possible to separate all the materials possibly present in the waste stream. Strongly contaminated, polluted, or impaired plastic pieces are better not sorted out because these parts could complicate the subsequent recycling steps. In this case, the simple methods work reliably and run fast enough for on-line identification. A problem arises when engineering plastics are sorted. They exist in many modifications which also have to be taken into account

as an extra material. The spectra are dominated not only by the composition and structure of the polymers but also by scattering, absorption, light reflection, baseline variations, and other effects. The analysis then needs the more sophisticated identification algorithms or requires extended wavelength regions for acquiring redundant information.

5.2 Applications

The interest in identification developed in parallel with the beginning of activity to recycle plastic materials. Many applications were investigated and realized on an industrial scale. Some are described here. However, some technical barriers had to be overcome before starting broad applications.

In the case of household waste plastics, mechanical separation (singulation) at adequate rates turned out to be very difficult. Actually, mainly bottles are sorted, which are heavier, and sorting about three pieces per second is state of the art. Efforts are being made to develop faster singulation, which is difficult if beakers and damaged pieces are also included. The sorting rates range up to 5–7 pieces s^{-1} and will increase further if more experience is available. The NIR technique itself is faster. As described in detail, the spectrometer can acquire spectra with sufficient quality in 1 ms and fast identification algorithms can keep up with it (the identification of 10–12 pieces s^{-1} with 50 spectra per part is no problem). In the case of bottle sorting the recovery rates are fairly high because the material is presorted. If samples of various sizes and shapes including beakers from real noncleaned waste streams are sorted, the output decreases. Labels, dyes, and inscriptions on household waste did not disturb the identification significantly. Erroneous identifications can lie below 0.1%. Heavy contamination reduces the quantity of identified samples and cleaner materials increase it. The conditions of the samples depend strongly on the presorting treatment and the collections, which vary considerably not only from country to country but also from area to area, and cannot be described adequately. The evaluation of results obtained in a single study cannot be transferred to a general situation. However, it can be estimated that output results of real applications are of the same order of magnitude of about 80–95%.

Automated dismantling, identification, and sorting have been applied to medical products, predominantly tubes and syringes. For recycling, syringes have to be washed, disinfected and dismantled into their two base parts: piston and case (tube). NIR spectroscopy identifies the polymer types before separating them; they are mainly PE, PP, and PVC.[96]

The automotive industries use various types of plastics, ranging from thermoplastic materials to composites produced by PU or reaction injection molding (RIM) technologies. FTIR spectroscopy in the MIR range has been applied to the identification of dismantled plastic pieces.[51,52,58–64,77] These are mostly covered by lacquers. The identification procedure requires pretreatment of the sample surfaces, which could take about 20 s. The identification lasts about 4 s. Studies with e.g. more than 7000 pieces revealed a purity of significantly below 1%. The weight of sorted plastic pieces is in the region of kilograms. Therefore, the time to identify the polymer is adequate as, in addition, the price of engineering plastics is higher by a factor up to five or more compared with packaging plastics. Pyrolysis/IR spectroscopy has also been applied with success.[84,85]

The electric/electronic industries produce mass consumer goods for private and industrial use which contain plastics in many variations. Plastic housings comprise the highest volume. In audio and video equipment the housings are black. In this case the application of FTIR spectroscopy in the MIR spectral region is appropriate.[58,73] The results are similar to those for plastic pieces in the automotive industry. Nonblack plastics are often used in information technologies and data processing where personal computers are produced in high volumes. MIR and NIR spectroscopy identify the polymers with sufficient accuracy and reliability. This is illustrated by an example analyzing the cases of electric homeworker tools (not dyed black) with an AOTF spectrometer for the NIR region.[27–31] The identification error was below 1% (Table 1). The NIR technique was also integrated into an automated dismantling and sorting system which treated keyboards of personal computers.

There are other ongoing investigations to treat plastic material from various industrial branches such as the building and construction industries and cable industries. The predominant aspects of identification are related to the variety of materials and their modifications used similarly as described in previous sections. Automated sorting is the objective of many investigations and progress suffers from the widely varying shapes and sizes of the products, which have to be taken into account

Table 1 Statistical test of plastic identification using an AOTF spectrometer with a hand-held sensor head in the wavelength range 1.1–1.8 μm with a 300 ms identification time and 200 scans accumulated at 300 nm ms^{-1}:[30] test results for a total of 810 homeworker tools

	Plastic type					
	PA	PP	ABS	PC	PBT	Not identified
Number of pieces	615	16	24	81	47	36
Error	4 (= PP)	0	0	1 (= PP)	0	5

to cover adequate volumes of plastics for economic operation.

6 CONCLUSIONS

In the future plastics will be used in many products in ever-increasing volumes. Their recycling saves natural resources of raw materials and areas of deposition sites, and recovers valuable substances.

NIR spectroscopy has proved to be an appropriate method to identify the plastics of packaging materials predominantly found in domestic wastes. The spectral quality of the plastics to be treated is high, allowing the analysis of fast-moving pieces at various sensing positions. The identification systems can be integrated into automated sorting lines.

The situation is more complicated in the case of engineering plastics. NIR and MIR spectroscopy can solve the problem of nonblack pieces but MIR spectroscopy requires a pretreatment in most cases. Black plastics require MIR techniques. Pyrolysis/IR spectroscopy is also applicable to engineering plastics. Integration into automated disassembly and sorting lines is currently under development and has been realized only in special cases.

LIST OF SYMBOLS

a_k	content of material k in the test sample
c^*	component concentration
c	velocity of light ($\mathrm{cm\,s^{-1}}$)
d^*	detectivity
E_{rot}	energy levels of rotations
E_{vib}	energy levels of vibrations
f	force constant
h	Planck's constant
I	radiation intensity
I_0	initial radiation intensity
m_i	mass of an atom
n	quantum number
T	transmittance
x_i	wavelength, wavenumber, or frequency
$y_{\mathrm{ref},j}(x_i)$	spectral intensity of reference sample j, measured at wavelength (wavenumber, frequency) x_i
$y_{\mathrm{test}}(x_i)$	spectral intensity of the test sample, measured at wavelength (wavenumber, frequency) x_i
ΔI	difference in intensities
Δx, d	thickness of material
α	absorptivity
$\alpha_{k,j}$	matrix built from the spectra of the reference samples
$\beta_{k,j}$	matrix built from the spectra of the reference and test samples
χ^2	sum of squared differences
χ_j	sum of normalized differences
χ_t	threshold value
λ	wavelength (μm)
μ	reduced mass
ν	frequency of molecular vibration or photons ($\mathrm{s^{-1}}$)
$\bar{\nu}$	wavenumber ($\mathrm{cm^{-1}}$)

ABBREVIATIONS AND ACRONYMS

ABS	Acrylonitrile–Butadiene–Styrene
AOTF	Acousto-optic Tunable Filter
CA	Cellulose Acetate
DSD	Duales System Deutschland
EU	European Union
FTIR	Fourier Transform Infrared
FTNIR	Fourier Transform Near-infrared
IR	Infrared
LBDS	Laser Breakdown Spectroscopy
LIPS	Laser-induced Pyrolysis Spectroscopy
MIR	Mid-infrared
NIR	Near-infrared
PA	Polyamide
PB	Polybutadiene
PBT	Poly(butylene terephthalate)
PC	Polycarbonate
PCA	Principal Component Analysis
PE	Polyethylene
PE-HD	High-density Polyethylene
PE-LD	Low-density Polyethylene
PE-LLD	Linear Low-density Polyethylene
PET	Poly(ethylene terephthalate)
PLS	Partial Least Squares
PMMA	Poly(methyl methacrylate)
POM	Poly(oxymethylene)
PP	Polypropylene
PPE	Poly(phenylene ether)
PS	Polystyrene
PU	Polyurethane
PVC	Poly(vinyl chloride)
RIM	Reaction Injection Molding
SAN	Styrene–Acrylonitrile Copolymers
SB	Styrene–Butadiene
TBBA	Tetrabromobisphenol A
TV	Television
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Near-infrared Spectroscopy of Polymers and Rubbers • X-ray Scattering in Analysis of Polymers

Atomic Spectroscopy **(Volume 11)**
Laser-induced Breakdown Spectroscopy

Chemometrics **(Volume 11)**
Chemometrics • Clustering and Classification of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Structure Determination, X-ray Diffraction for • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. C. Nagel, 'Reverse Logistic Systems – A Comparative Study', in *Proc. CARE INNOVATION '98 Symposium, Vienna, Austria, November 16–19, 1998*, 53–60.
2. B. Krummenacher, P. Peuch, M. Fisher, M. Biddle, 'Automatic Identification and Sorting of Plastics from Different Waste Streams – A Status Report', APME Technical Report, Association of Plastics Manufacturers in Europe, Brussels, 1998.
3. M. Fisher, 'Plastics Recovery in the USA', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
4. A. Miyake, 'The Outline of Separation Technology of Plastics Waste Developed and Practiced in Japan', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
5. K. Flowers, 'European Community Policy', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
6. N. Russotto, 'Plastics Recovery in Europe', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
7. V. Trautz, 'Die Entwicklung der westeuropäischen Kunststoffindustrie', *Kunststoffe*, **88**(10), 1648–1652 (1998).
8. CONSULTIC, *Kurzfassung Kunststoffrecyclingstatistik 1995*, CONSULTIC Marketing und Industrie beratung, Grossostheim, 1995.
9. F. Mader, Presented at the European Workshop on the Recycling of Plastics', Brussels, Belgium, 18–19 April 1994.
10. J. Brandrup, 'Ökologie und Ökonomie der Kunststoffverwertung', *Müll Abfall*, **30**(8), 492–501 (1998).
11. R. Gabriel, S. Salhofer, 'Treatment of End-of-life Electric and Electronic Equipment', in *Proc. CARE Innovation '98 Symposium, Vienna, Austria, November 16–19, 1998*, 159–166.
12. G. Kortüm, *Reflexionsspektroskopie: Grundlagen, Methodik, Anwendungen*, Springer, Berlin, 1969.
13. G. Hertzberg, *Molecular Spectra and Structure I*, Van Nostrand, Princeton, NJ, 1953.
14. G. Hertzberg, *Molecular Spectra and Structure II*, Van Nostrand, Princeton, NJ, 1960.
15. A. Steele, *Theory of Vibrational Spectroscopy*, Saunders, Philadelphia, PA, 1971.
16. H.W. Siesler, K. Holland-Moritz, *Infrared and Raman Spectroscopy of Polymers*, Marcel Dekker, New York, 1980.
17. D.A. Burns, E.W. Ciurczak (eds.), *Handbook of Near Infrared Analysis*, Marcel Dekker, New York, 1992.
18. Ch.J. Pouchert, *The Aldrich Library of FTIR Spectra*, 1st edition, Aldrich, Milwaukee, WI, Vols. I and II, 1985.
19. *The Atlas of Near Infrared Spectra*, Sadtler Research Laboratories, Philadelphia, PA, 1981.
20. M. Buback, H.P. Vögele, *FT-NIR Atlas*, VCH, Weinheim, 1993.
21. M. Claybourn, P. Colombel, J. Chalmers, 'Characterization of Black Materials by FTIR', in *Proc. Int. Workshop on Fourier Transformed Infrared Spectroscopy, Antwerp, 1990*, 118–127.
22. W.W. Hart, P.C. Painter, J.L. Koenig, M.M. Coleman, 'A Fourier Transform Infrared Method of Studying Elastomers Filled with Carbon Black', *Appl. Spectrosc.*, **31**(3), 220–224 (1977).
23. J.A. Jansens, W.E. Haas, 'Application of Diffuse Reflectance Optics for the Characterization of Polymer Surfaces by Fourier Transform Infrared Spectroscopy', *Polym. Commun.*, **29**, 77–80 (1988).
24. N. Eisenreich, H. Krause, H. Kull, Th. Messer, 'Schnelle Erkennung von Kunststoffen mit AOTF-Reflektionsspektrometer', in *Neue Konzepte für die Autoverwertung,Tagung Wolfsburg 1991*, Verein Deutscher Ingenieure, Düsseldorf, 447–460, 1991.
25. N. Eisenreich, H. Kull, E. Thinnes, 'Schnelle Identifikation von Kunststoffen mit schneller Nahinfrarotspektroskopie', in *Waste Management of Energetic Materials and Polymers, 23rd Int. Ann. Conf. ICT, Karlsruhe, 1992*, 59/1–59/12.

26. N. Eisenreich, J. Herz, H. Kull, 'Schnelle Erkennung von Kunststoffen mit NIR-Spektroskopie', in *Entsorgung von Elektrischen und Elektronischen Geräten 'Elektroschrott '93', DVM, Tag. Berlin, 12–14 May 1993*, 247–255.
27. N. Eisenreich, J. Herz, H. Kull, 'Schnelle Kunststoffidentifizierung mit AOTF-Spektrometer', in *4. Deutsche Informationsionstagung Nahinfrarotspektroskopie und Lichtleitersonden in der Qualitätsicherung und Prozesskontrolle, Haus der Technik, Essen, 21–22 September 1993*.
28. N. Eisenreich, J. Herz, H. Kull, W. Mayer, 'Fast On-line Identification of Plastic Materials by Near Infrared Spectroscopy' in *Proc. R '95 – Recovery, Recycling, Reintegration, Geneva, 1–3 February 1995*, eds. A. Barrage, X. Edelmann, EMPA, Swiss Federal Laboratories for Materials Testing and Research, 230–235, Vol. 3, 1995.
29. N. Eisenreich, T. Rohe, 'Identifying Plastics', *Kunstst. Plast. Eur.*, **86**(2), 31–32 (1996).
30. N. Eisenreich, J. Herz, H. Kull, W. Mayer, T. Rohe, 'Fast On-line Identification of Plastics by Near-infrared Spectroscopy for Use in Recycling Processes', in *Proc. ANTEC '96, Indianapolis, 1996*, 3131–3135.
31. N. Eisenreich, T. Härdle, H. Kull, T. Rohe, 'Fast Identification of Polymers and Additives of Mass Consumer Products', in *Proc. Int. Congress CARE INNOVATION '96, Frankfurt/M, 1996*, 219–224.
32. N. Eisenreich, 'New High Performance Identification Concepts', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
33. U. Eschenauer, O. Henck, M. Huehne, P. Wu, I. Zebger, H.W. Siesler, 'Near-infrared Spectroscopy in Chemical Research, Quality Assurance and Process Control', *Near Infra-Red Spectrosc.*, **7**, 11–18 (1992).
34. M.K. Alam, S.L. Stanton, 'Sorting of Waste Plastics using Near-infrared Spectroscopy and Neural Networks', *Process Control Qual.*, **4**(4), 245–252 (1993).
35. M.K. Alam, S.L. Stanton, G.A. Hebner, 'Near-infrared Spectroscopy and Neural Networks for Resin Identification', *Spectroscopy*, **2**(9), 30–40 (1994).
36. H. Ritzmann , 'NIR: New Ways for the Sorting of Mixed Plastics Waste', in *Recycle '93, Conference Proceedings, 22–26 March 1993*, paper 11/1, Maack Business Services, Davos, 1993.
37. H. Lucht, 'NIR Sorter Takes Close Look at Plastic', *Technol. News*, **49** (Dec.) (1993).
38. H. Lucht, 'Industrielles Verfahren zur sortenreinen Kunststofftrennung', *EUWID Recycling Entsorgung*, **3**(20), 12 (1993).
39. H. Lucht, 'Laser Spectroscopy Plastics Analysis/a Supplement to NIR Systems', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
40. B. Willenberg, W. Kohn, H. Ritzmann, 'Identification of Plastics with Near-IR Spectroscopy', *Plastverarbeiter*, **44**(12), 21–25 (1993).
41. A. Coghlan 'Infrared Makes Light Work of Sorting Plastics', *New Sci.*, **139**(1882), 20 (1993).
42. B. Wienke, D.W. van den Broek, L. Budyens, 'Identification of Plastics Among Non-plastics in Mixed Waste by Remote Sensing Near Infrared Imaging Spectroscopy 2: Multivariate Image Rank Analysis for Rapid Classification', *Anal. Chem.*, **67**(20), 3760–3766 (1995).
43. W. van den Broek, D. Wienke, W.J. Meissen, C.W.-A. de Crom, L. Budyens, 'Identification of Plastics Among Non-plastics in Mixed Waste by Remote Sensing Near Infrared Imaging Spectroscopy 1: Image Improvement and Analysis by Singular Value Decomposition', *Anal. Chem.*, **67**(20), 3753–3759 (1995).
44. B. Wienke, W. van den Broek, L. Buydens, T. Huth-Fehre, R. Feldhoff, T. Kantimm, K. Cammann, 'Adaptive Resonance Theory Based Neural Network for Supervised Chemical Pattern Recognition (FuzzyARTMAP), Part 2: Classification of Post-consumer Plastics by Remote NIR Spectroscopy using an InGaAs Diode Array', *Chemom. Intell. Lab. Syst.*, **32**(2), 165–176 (1996).
45. T. Huth-Fehre, T. Kantimm, R. Feldhoff, H. Freitag, F. Kowol, M. Oleimeulen, 'A Multiplexed NIR Spectrograph or a MIR/AOTF Spectrometer as Two Supplementary Solutions for the Rapid Identification of Plastics from Electronic Devices', in *Proc. Int. Congress CARE INNOVATION '96, Frankfurt/M, 1996*, 214–218.
46. R. Feldhoff, D. Wienke, K. Cammann, H. Fuchs, 'Online Post Consumer Package Identification by NIR Spectroscopy Combined with a FuzzyARTMAP Classifier in an Industrial Environment', *Appl. Spectrosc.*, **51**(3), 362–368 (1997).
47. P. Castro, H. Mark, 'Qualitative Analysis of Thin-film Polymer Samples using Near-IR Spectroscopy for Recycling Applications', *Spectroscopy (Eugene, Oreg.)*, **9**(1), 27–28, 30–32 (1994).
48. T. Amano, 'Identification of Plastics with NIR Technology', in *Proc. Recycle '95, Davos, May 15–19, 1995*, paper 68, 8(13).
49. T. Amano, 'New Practical Type of Identification Method (AOTF/NIR)', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
50. K. Yoneda, 'NIR Automatic Identification and Sorting System', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
51. D.F. Gentle, 'Automatic Identification of Plastics Out of Junk Cars', Automatische Identifikation für Kunststoffe aus Altfahrzeugen, in *Kunststoffe im Automobilbau. Rohstoffe, Bauteile, Systeme*, ed. D.F. Gentle, VDI-Verlag, Dusseldorf, 314–329, 1994.
52. D.F. Gentle, 'Automated Material Identification in Recycling', in *Plastics in Automotive Engineering*, eds. H.G. Haldenwanger, L. Vollrath, Carl Hanser, Munich, 261–271, 1994.
53. B.M. Scott, 'A Two-color Near-infrared Sensor for Sorting Recycled Plastic Waste', *Meas. Sci. Technol.*, **6**(2), 156–159 (1995).
54. *Bühler NIR Factorfilter-Spectrometer System NIRIKS*, Bühler, Uzvil, Switzerland, **262**(15)-9922T, 1994.

55. C. Schudel, W. Koller, 'Economic Solutions to Materials Recycling of Plastics Using the Buhler NIR Spectrometer NIRIKS', in *Proc. R '95–Recovery, Recycling, Re-integration*, Vol. 3, Geneva, 1–3 September 1995, eds. A. Barrage, X. Edelmann, EMPA, Swiss Federal Laboratories for Materials Testing and Research, 257–265, Vol. 3, 1995.
56. K. Ohta, H. Ishida, 'Comparison Among Several Numerical Integration Methods for Kramers–Kronig Transformation', *Appl. Spectrosc.*, **42**(6), 952–957 (1988).
57. M.L. Bortz, R.H. French, 'Quantitative, FFT-based, Kramers–Kronig Analysis for Reflectance Data', *Appl. Spectrosc.*, **43**(8), 1498–1501 (1989).
58. J. Jansen, C.v. Hastenberg, 'Identifikation von Kunststoffteilen', *Kunststoffe*, **84**(1), 51–54 (1994).
59. B. Willenberg, 'Welche Verfahren zum schnellen Erkennen von Kunststoffen?', *Kunststoffe*, **84**(1), 54–55 (1994).
60. A. Willenberg, 'MIR: a Proven Technology for Identification of Engineering Plastics – Experience in Automotive Plastic Parts', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
61. K. Vornberger, B. Willenberg. 'Rapid Identification of Plastics', *Kunstst. Plast Eur.*, **84**(5), 17–19 (1994).
62. K. Vornberger, G. Zachmann, B. Willenberg, 'First Experiences at the BMW Dismantling and Recycling Center', in *Proc. Recycle '94, Davos, 14–18 March 1994*, paper 76, Maack Business Services, Davos, 1993.
63. J. Florestan, A. Lachambre, N. Mermilliod, J.C. Boulou, C. Marfisi, 'Recycling of Plastics: Automatic Identification of Polymers by Spectroscopic Methods', *Resour. Conserv. Recycling*, **10**(1/2), 67–74 (1994).
64. J. Graham, P.J. Hendra, P. Mucci, 'Rapid Identification of Plastics Components Recovered from Scrap Automobiles', *Plast. Rubber Compos. Process. Appl.*, **24**(2), 55–67 (1995).
65. A.M. Scott, R.L. Waterland, 'Identification of Plastic Waste using Spectroscopy and Neural Networks', *Polym. Eng. Sci.*, **35**(12), 1011–1015 (1995).
66. G. Zachmann, 'A Rapid and Dependable Identification System for Black Polymers', *J. Mol. Spectrosc. Struct.*, **348**, 453–456 (1995).
67. G. Zachmann, P. Tumer, 'Fast and Reliable Identification of Black Plastics', *Spectrosc. Eur.*, **9**(1), 18 (1997).
68. M.B. Biddle, 'Recovery of Higher Value Plastics Improved by Advanced Processes', *Mod. Plast. Encycl.*, **73**(12), A37–A39 (1996).
69. M. Fletcher, 'Scanner Progresses Plastic ID Method', *Eureka*, **16**(10), PIE 1/3 (1996).
70. C.E. Tinney, B. Thomson, E. Johnson, 'Mid-infrared Spectroscopic Fingerprinting for Plastics Identification and Recycling', *Spectroscopy*, **11**(6), 50–53 (1996).
71. S. Stuckey, 'Light Relief for Car Recyclers', *Engineer*, **284**(7333/7334), 27 (1997).
72. D. Walters, 'The Benefits of using the FTIR Identification Method within the Recycling Industry', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
73. E. Michiels, 'Rapid Identification of Consumer Electronics, PC and TV Plastics Parts by FTIR', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
74. E.J. Sommer, 'New Techniques for Rapid Identification and Sorting of Plastics', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
75. B.J. Kip, 'Identification and Sorting of Carpet Waste Materials', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
76. G. Kenny, 'High Speed Plastics Sorting Systems', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
77. S.W. Gaarenstroom, P.B. Coleman, C.M. Duranceau, 'Rapid Identification of Automotive Plastics in Dismantling Operations: Evaluation of Specular-reflectance Infrared Spectroscopy Systems', *Soc. Automot. Eng. Spec. Publ.*, SP-1263 (Design for Environmentally Safe Automotive Products and Processes), 47–54 (1997).
78. P.E.R. Mucci, 'Rapid Identification of Plastics using External Beam Mid-infrared Spectroscopy', *Spec. Publ. R. Soc. Chem.*, **199** (Chemical Aspects of Plastics Recycling), 53–70 (1997).
79. B. Eggleston, 'Recycling Loop Widens', *Plast. Rubber Wkly*, (1726), 10 (1998).
80. W. Becker, N. Eisenreich, 'Identification of Black Polymers in the Mid-infrared Spectral Region with an FTIR Spectrometer', in *Proc. Int. Congress CARE INNOVATION '98, Vienna (Austria), 1998*, 88–95.
81. C. Braun, E. Richter, 'Qualitative Analysis of Heat Stabilizers in Poly(vinyl chloride)', *Kautsch. Gummi Kunstst.*, **50**(2), 112–118 (1997).
82. G. Kampffmeyer, A. Pfeil, 'Self-supporting Thin-film Beam Splitter for Far-infrared Interferometers', *Appl. Phys.*, **14**, 313–317 (1977).
83. H. Krause, N. Eisenreich, A. Pfeil, 'Kinetics of Evaporation and Decomposition of Isopropyl Nitrate by Rapid Scan IR Spectroscopy', *Thermochim. Acta*, **149**, 349–356 (1989).
84. A. Murase, 'Rapid Identification of Plastics by Pyrolysis IR and Hand-held Dielectric Instrument', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
85. R. Klebba, 'Plastics Identification by Pyrolysis FTIR', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
86. T. Yano, A. Watanabe, 'New Noncollinear Acousto-optic Tunable Filter using Birefringence in Paratellurite', *Appl. Phys. Lett.*, **24**(6), 256–258 (1974).
87. I.C. Chang, 'Acousto-optic Tunable Filters', *Opt. Eng.*, **20**(6), 824–829 (1981).
88. A. Blanc, N. Eisenreich, H. Kull, W. Liehmann, 'Charakterisierung von Verbrennungsprozessen mittels zeitaufgelöster IR-Spektroskopie im Breich von 1 µm–14 µm', in *Combustion and Detonation Phenomena, 19th Int. Ann. Conf. ICT, Karlsruhe 1988*, 74/1–74/15.
89. N. Eisenreich, H. Kull, Th. Messer, 'A Fast Scanning Spectrometer in the NIR using an Acousto Optic Tunable

Filter and Fiber Optics', in *Optical Measurement Techniques and Fiber Optics (French–German Workshop, Freiburg, 1991)*.

90. T. Messer, H. Kull, N. Eisenreich, 'Untersuchung der Emission von Verbrennungsvorgängen im NIR-Bereich mit einem Spektrometer auf der Basis eines akusto-optischen durchstimmbaren Filters', in *Combustion and Reaction Kinetics, 22nd Int. Ann. Conf. ICT, Karlsruhe, 1991*, 99/1–99/6.
91. M. Hühne, U. Eschenauer, H.W. Siesler, 'Acousto-optic Tunable Filter NIR spectrometer for Rapid Process Control', *Proc. SPIE*, **1575**, 214 (1992).
92. T. Schultz, H. Hoberg, K.W. Plessman, P. Laufens, 'Influence of Surface Contamination on the Identification of Plastics with NIR Spectroscopy', in *Proc. 1st Int. Conf. Identiplast, Brussels, October 27–28, 1997*.
93. N. Eisenreich, W. Engel, 'Difference Thermal Analysis of Crystalline Solids by the Use of Energy-dispersive X-ray Diffraction', *Appl. Crystallogr.*, **16**, 259–263 (1983).
94. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes – The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.
95. H. Martens, T. Naes, *Multivariate Calibration*, Wiley, Chichester, 1989.
96. B. Bohlmann, K.-F. Ziegahn, 'MEDIREMAT – A Recycling System for Medical Single-use Devices', in *Proc. R '99 – Recycling, Recovery, Re-integration, 4th Int. Congress, Genf, 2–5 February, 1999*.

Infrared Spectroscopy in Analysis of Polymer Crystallinity

Surya K. Mallapragada
Iowa State University, Ames, USA

Balaji Narasimhan
Rutgers University, Piscataway, USA

1 Introduction 1
2 Analysis of Orientation in Polymers Using Infrared Spectroscopy 2
2.1 Sampling Methods for Fourier Transform Infrared Spectroscopy 2
2.2 Modes of Measurement 3
2.3 Semicrystalline Polymers 4
2.4 Liquid-crystalline Polymers 5
3 Comparison of Infrared Measurements with Other Techniques 9
Abbreviations and Acronyms 11
Related Articles 11
References 11

This article provides a comprehensive review of the application of infrared (IR) spectroscopy in the analysis of polymer crystallinity. A brief description of the theory behind the analysis of orientation in polymers using IR spectroscopy is followed by a comparison of various sampling methods for Fourier transform infrared (FTIR) spectroscopy. These include transmission, attenuated total reflectance (ATR), external reflection, diffuse reflectance, photoacoustic and emission spectroscopic techniques. Various modes of measurement, such as polarized IR measurements, deuteration and rheo-optical measurements to obtain more information from IR spectra and make the interpretation of spectra easier, are also discussed. The use of IR spectroscopy in analyzing both semicrystalline polymers and liquid crystalline polymers is presented. In the case of semicrystalline polymers, applications involve measuring phase transformations, chain conformation/packing, degree of crystallinity, crystallization kinetics, isomer ratios and chain alignment during mechanical treatment. In the case of liquid crystalline polymers, IR spectroscopy has been used to analyze molecular orientation, segmental mobility, photo-induced reorientation and phase characterization/transformations. A comparison of IR measurements with other techniques used to analyze crystallinity in polymers such as X-ray diffraction (XRD), differential scanning calorimetry (DSC), nuclear magnetic resonance (NMR) spectroscopy, Raman spectroscopy and neutron scattering has been presented.

1 INTRODUCTION

Semicrystalline polymers exhibit short-range as well as long-range ordering of macromolecular chains. Both types of ordering contribute towards the degree of crystallinity of the polymer. Liquid crystals (LCs) are substances that exhibit long-range order in one or two dimensions, but not all three.[1,2] LCs and liquid-crystalline polymers (LCPs) have gained scientific and technological importance because of their use as display materials, materials for optical information storage, high-modulus fibers, and high-temperature plastics. An understanding of the order and structure in both LC and semicrystalline polymer systems, using techniques such as IR spectroscopy, is essential for designing these materials for various applications and predicting their properties. Research on such polymers includes studies of phase characterization/transformations, orientation under the influence of external fields, photo-induced reorientation, dynamics, and segmental mobility.[3]

IR spectroscopy is one of the most commonly used spectroscopic tools because the method is rapid and sensitive with sampling techniques that are easy to use. IR spectroscopy measures the vibrational energy levels of molecules. The characteristic band parameters measured in IR spectroscopy are frequency (energy), intensity (polar character), band shape (environment of bonds), and the polarization of various modes.[4] A very good review of the theory and applications of IR spectroscopy for investigating polymeric systems is provided in the books by Koenig.[4,5]

When a polymer passes from the amorphous to the crystalline phase, either new absorption bands appear in the IR spectrum, or splitting occurs in the bands present in the spectrum of the amorphous material.[6] With certain polymers, several bands have been found to increase in intensity as the polymer passes from a crystalline to an amorphous phase. The former bands are termed ''crystalline bands'' and the latter are called ''amorphous bands''. Since the symmetry of a molecule in a crystal field is usually lower than the symmetry of a free molecule, certain types of vibrations, which are inactive in an amorphous polymer, may be active in the unit cell of a three-dimensional or one-dimensional crystal. Crystalline bands may also appear, caused by the change in the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

configuration of the chain in the crystal cell of the polymer compared with its configuration in the amorphous state. The vibration of an individual molecule in a crystal field, combined with the interactions of vibrations between chains in the crystal cell, presents a complete picture of the process. Intermolecular interactions such as those that occur in the crystalline state will influence the detailed structure of the absorption band associated with a given mode, as well as its intensity,[7] but are not likely to greatly affect the frequencies of group normal vibrations. A special case of an intermolecular interaction with a large influence on group frequency is hydrogen bonding.

This article discusses some background theory on the use of IR to investigate crystallinity of polymers, describes some of the experimental methods and sampling techniques commonly used, and outlines applications related to structure elucidation of semicrystalline and liquid crystalline polymers that IR spectroscopy has been used for. The article concludes with a comparison of IR versus other commonly used techniques for characterizing such polymers.

2 ANALYSIS OF ORIENTATION IN POLYMERS USING INFRARED SPECTROSCOPY

The fundamental modes of the single chain are split into various spectral components in the crystal if the intermolecular forces between polymer chains are sufficient. Physically, such splitting, called the correlation field splitting or factor group splitting, occurs as a result of the resonance interactions of analogous vibrations undergone by individual chains in the unit cell. The number of theoretically expected bands depends on the number of molecules in the unit cell.[4] For an infinite, ordered polymer, symmetry reduces the vibrational analysis to the analysis of the translational repeat unit. This is generally not a bad assumption for high-molecular-weight polymers, but there are usually some secondary IR bands that cannot be assigned to fundamentals, combinations or end-groups because of the semicrystalline nature of even the most highly regular polymer.

Vibrational spectroscopy can be used to study conformational defects and to distinguish between stereoisomers. Mass configuration plays a dominant role in rotational isomerism. Crystallization usually results in stabilization of the lowest energy isomer, whereas all forms are present in the liquid state. Since the amorphous state in polymers is essentially a disordered state comparable to liquids, the spectrum of a partially crystalline polymer in which rotational isomerism can occur is essentially a superposition of the spectra of various isomers. Also, because rotational isomers have different symmetry, the number of modes associated with a given group can be different for different isomers. In certain cases, amorphous absorption bands are formed on passing from the crystalline to the amorphous phase, which may be due to the breakdown of selection rules operating in the crystal or to the existence of rotational isomers which are absent in the crystalline phase.[6] The intensity of amorphous bands is usually low.

Polyethylene, having closer chains than any other polymer, exhibits the largest crystal field splitting.[4] One of the limitations of IR spectra for determination of polymer crystallinity is that very few polymers exhibit the crystal field splittings observed in specific cases, such as polyethylene, since the splitting is very sensitive to the separation of the polymer chains. However, even if most polymers do not exhibit crystal field splittings due to intermolecular or crystalline packing, band pairs can arise from intramolecular helical splitting, such as in the case of polypropylene. Polymers can exhibit resonance splitting of bands associated with interaction between analogous groups along the polymer chain. Polymers having a helical chain structure, for example, show splitting of certain vibrations into components parallel to the axis of the helix and perpendicular to the plane of the helix. Most IR correlations of spectral features with crystallinity or density are based on the amount of rotational isomer preferred in the crystalline state, which, on average, may correlate with crystallinity or density. Empirical correlations of IR bands to the vibrations of partially ordered polymers have been reported for a number of polymers.

The sensitivity of IR spectroscopy has been improved tremendously by the use of multiplex optical devices, such as the Michelson interferometer, which allow continuous, simultaneous detection of all transmitted energy. This has led to the development of FTIR spectroscopy.

2.1 Sampling Methods for Fourier Transform Infrared Spectroscopy

IR spectra depend on sample preparation techniques, and, depending on the specimen to be investigated, various sampling techniques can be used. Transmission spectroscopy is the most commonly used technique, has the highest signal-to-noise ratio, and is the easiest to quantify. For transmission IR spectroscopy, the primary dimensional requirement on the polymer sample is that it be sufficiently thin. Various techniques are used to produce thin films, including solvent casting, melt casting, and microtoming. Orientation in polymer specimens can be achieved in several ways, including stretching the film or fiber after heating or swelling if it is necessary. Stretching and rolling produce different kinds of chain

orientation in a polymer film. In a stretched specimen, a particular crystallographic axis will become aligned along the stretch direction, but all orientations of the crystallites about this axis will be equally probable.[7] In rolled specimens, double orientation is produced due to the axis aligning with the rolling direction and the crystallographic planes orienting parallel to the plane of the film.

Another mode of operation that can be used to obtain spectra, especially when it is difficult to obtain very thin films or if the polymer samples have a low transmission, is the ATR mode. This is a contact sampling technique that involves a crystal such as germanium with a high refractive index and low IR absorption in the IR region of interest. The ATR mode is more general than the transmission mode and also applies to samples irrespective of their thickness or their absorbance.[8] The ATR technique also has some disadvantages in that it shows a great deal of uncertainty depending on contact between the sample and the high-index crystal.[9] ATR has been very useful in the studies of aqueous systems because it is possible to subtract the water absorption in a reproducible fashion. In transmission spectroscopy, a comparison of two bands of a spectrum poses no problems because the analyzed thickness is the sample thickness itself. However, in ATR/FTIR such a comparison is not possible because the analyzed thickness now depends on the wavelength. Therefore, two distinct absorption bands corresponding to remote spectral zones can be compared in transmission but not in ATR spectroscopy. The transmission method helps analyze the core of the sample, while ATR analyzes the surface. Therefore the methods are complementary.

The external reflection technique, involving a sample deposited on the surface of a smooth mirror-like surface, is very useful in the analysis of coatings and adhesives on the surface of metals. The polarization sensitivity of this technique adds to its usefulness. However, the major difficulties with this technique are the appearance of asymmetric absorption modes and the shifting of some of the mode frequencies.

Sample preparation is often the most time-consuming part of the IR analysis. Diffuse reflectance spectroscopy has been proposed as a ''no preparation IR technique''. However, it works well only on highly scattering and weakly absorbing samples.[4] In addition, symmetrically shaped absorption bands are represented as asymmetric ''differential-like'' band shapes. The use of IR absorption spectra to investigate the conformation of polymers in solution is difficult because most solvents themselves absorb in the range investigated, and it is difficult to choose compensating conditions for solvents such as dichloroacetic and trifluoroacetic acids.[6]

Photoacoustic spectroscopy is based on the principle that modulated IR radiation striking the surface of a sample will cause the surface to alternately heat and cool with IR absorption.[4,9] This technique requires minimal sample preparation and provides the capability of studying opaque samples and depth profiling the sample. However, it has a low signal-to-noise ratio and high absorption related to photoacoustic saturation.

Emission spectroscopy can be used on samples that are opaque and cannot be studied by transmission spectroscopy. But high temperatures are usually required and sample decomposition could be a problem, especially in the case of polymers. IR microspectroscopy involves using IR microscopes for both transmission and reflection measurements and can have a spatial resolution of about 50 μm. It is developing into a valuable tool for polymeric samples, especially because it is relatively easy to obtain a minute polymer sample.[4]

2.2 Modes of Measurement

Vibrational spectroscopy is very sensitive to the degree of order present in a polymer sample, but the interpretation of results is not always straightforward. A particular sample can be envisaged to be a mixture of several phases: ordered conformations, regular or irregular chain folds, and a bulk amorphous phase. Experimentally, the major problem arises from the fact that it is often difficult to separate the contributions of the various phases.[4,7] The amorphous component also contributes to the spectrum of a semicrystalline material. Therefore, by subtracting the spectrum of an amorphous component, appropriately scaled, from that of a semicrystalline material, a difference spectrum is obtained, which is usually characteristic of the crystalline phase. Also, various modes of measurement, such as the ones listed here, can be used to get information that can make the interpretation of spectra easier and provide more information about structure and order in the specimen.

2.2.1 Polarization

The intensity of an absorption band is proportional to the square of the scalar product of the transition moment and electric field vectors. As a result, if the electric field vector is perpendicular to the transition moment vector, no absorption will occur. This fundamental relationship is the basis for utilization of polarized radiation to study spectra of oriented polymers. The study of oriented polymers with polarized IR radiation allows, with certain restrictions, the determination of orientation (with respect to the molecular structure) of the transition moment for a given normal mode. This makes it possible to classify bands in the spectrum and establish their origin.[7]

Polarized IR measurements permit determination of the direction that the transition moment associated with

a given band makes with the fiber axis, which is often significant in establishing structural parameters. The dichroic ratio of the corresponding absorption band is the experimental quantity related to the orientation of the transition moment for a given mode. When a specimen is stretched, the chains do not all become oriented in the same direction, but assume a distribution of orientations about this direction. In general, it is not possible to obtain the form of this distribution based on dichroic ratio measurements. However, some general conclusions about the nature of the distribution can be obtained by tilting an oriented specimen at various angles to the polarized beam. Individual spectra of polychloroprene were obtained using light alternately polarized parallel and perpendicular to the direction of elongation, and the dichroic ratio was evaluated.[10] In cases where the polymer chain exhibits symmetry, a qualitative knowledge of an absorption band helps associate the band with certain of the symmetry species, and thereby with certain normal modes of the molecule. A number of studies have been conducted to obtain IR spectra and measure dichroic ratios using polarized radiation[8,10–17] and to compare the results with the spectra obtained using unpolarized radiation.[18]

With respect to the sample, several problems could arise.[7] First, reflection at the surface and scattering in the interior will modify the incident intensity value from that observed in the absence of a sample. These losses can be corrected by subtracting the optical density in a nearby region that is free of absorption bands from the peak optical density. Secondly, for birefringent specimens, the propagation of a beam of plane polarized radiation can be different in the interior of the specimen and in air. Because of the symmetry of oriented polymer specimens, although the maximum absorption occurs when the electric vector is parallel to the transition moment, for the gross specimen the maximum absorption occurs with the electric vector along or perpendicular to the fiber axis. Thirdly, the specimen may exhibit form dichroism, which might have to be subtracted from the observed dichroism to obtain the structural dichroism.

2.2.2 Deuteration

The use of partially deuterated chains in the analysis of the conformational structure of polyethylene was investigated[5] because of the problems associated with identifying and assigning bands to localized vibrations. This problem is especially significant when there is an overlap of such bands in a specific spectral region, and the vibrations are not entirely localized. When an isotopic substitution is made in a molecule, the forces between the atoms remain unchanged, and only the masses of the hydrogen atoms change in magnitude. The result is that the normal modes, which involve primarily the motion of hydrogen atoms, will be lowered in frequency upon deuterium substitution.[7] Moreover, if a molecule contains different kinds of hydrogen atoms it is possible to deuterate these selectively, enabling a more detailed identification. Finally, in cases where there is interaction between hydrogen modes and other modes, such interaction will usually disappear upon deuterium substitution, indicating its presence in the original molecule.

A method of detecting the presence of inter-chain interactions is to study the spectra of mixtures of deuterated and normal polymers. When a neighboring chain has different frequencies for comparable normal modes, any band splitting due to inter-chain interactions will disappear because the interaction between different frequencies is much smaller. This change in band contour qualitatively identifies the existence of an interaction.[7] IR spectra have been obtained using deuterated samples for a number of polymers, especially for polyethylene.[19,20]

2.2.3 Rheo-optical Measurements

Rheo-optical FTIR spectroscopy is based on the simultaneous acquisition of stress–strain data and FTIR spectra and is used for characterization of transient structural changes during deformation and stress-relaxation of polymers.[11] This technique has been extensively used for studying phenomena such as strain-induced crystallization in polymers.[10–12,22–24]

2.3 Semicrystalline Polymers

FTIR spectra of various polymers have been obtained to elucidate the order present in the specimen. IR spectra can be used for measurements of stereoregularity, conformation, characterization of polymer blends, deformation in polymer systems, dichroic measurements of orientation, trichroic measurements of orientation (for samples which have orientation with a three-dimensional character), measurements of intermolecular and intramolecular interactions and morphological units and structural changes as a result of thermal annealing or mechanical orientation.[4]

The most common applications have involved studying phase transformations in semicrystalline polymers. IR spectroscopy has been used to measure orientation in the crystalline and noncrystalline phases using polarized radiation[25] to follow changes from one crystal form to another and changes in degree of crystallinity (Figure 1) when the specimen is subjected to thermal annealing.[26–31] The intensity of selected peaks is directly proportional to the degree of crystallinity of

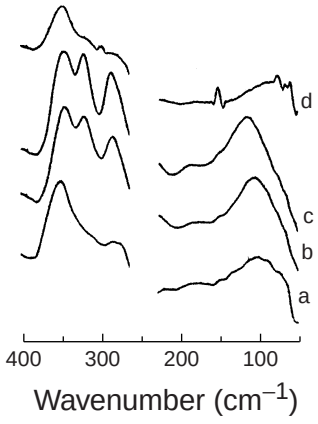

Figure 1 Far-IR spectra of differently ordered polyurethane samples measured at room temperature[13] (annealed at the indicated temperature (°C) for 2 h). [Reprinted with permission from D.Y. Shen, S.K. Pollack, S.L. Hsu, *Macromolecules*, **22**, 2564–2569 (1989). Copyright 1989 American Chemical Society.]

the polymer. IR spectroscopy can also provide valuable information about chain conformation/packing using hydrogen bond vibrational modes and can elucidate microphase separation between hard and soft segments in polyurethanes.[13] It has been used to measure the degree of crystallinity of semicrystalline polymers[11,32–37] as well as the temperature coefficient of crystallinity.[38] Crystallization kinetics during thermal annealing can also be followed using this technique, and Avrami coefficients have been estimated based on IR spectra.[39–41] IR spectra can be used to distinguish between intermolecular and intramolecular effects[42,43] and elucidate chain conformations[20] in the polymers. Phase separations in blends of polymers[44] and photoinduced orientation of groups[15] have also been interpreted from IR spectra. The conformation of poly(ethylene oxide) chains in a unit cell has been elucidated using IR.[14]

Another area where IR spectra can provide valuable information is in distinguishing between various isomers in the same polymer. Crystal structures of two samples with different fractions of *cis*- and *trans*-components have been compared.[45,46] Similarly, IR spectra have been used to compare samples with different tacticities,[47] such as atactic and isotactic polystyrene (Figure 2), where the tacticity strongly influences the crystallizability.[48] Evolution of crystalline *trans* and gauche conformations during thermal treatment has also been investigated.[9,49] Orthorhombic to hexagonal phase transitions in ultra-drawn polyethylene have been elucidated using IR.[18] IR spectra can help determine the extent of isotopic fractionation in deuterated samples.[19]

Rheo-optical spectroscopy helps measure chain alignment during mechanical treatment.[10–12,22,23] Most

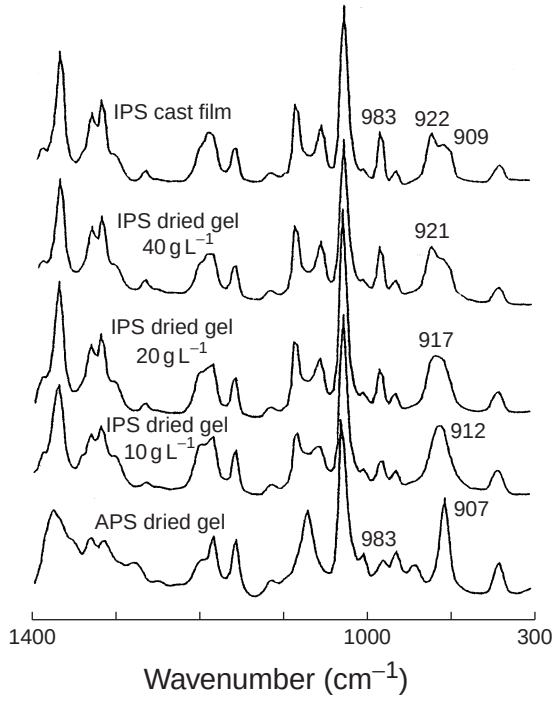

Figure 2 Comparison of IR spectra of atactic polystyrene dried gel, isotactic polystyrene gel dried from various polymer concentrations and solution-cast isotactic polystyrene film.[48] [Reprinted from X.M. Xie, A. Tanioka, K. Miyasaka, 'Fourier-transform Infrared Spectroscopic Studies of Gelation and Crystallization of Polystyrene Carbon-disulfide Solutions', *Polymer*, **34**, 1388–1392, Copyright 1993, with permission from Elsevier Science.]

semicrystalline polymers undergo uniaxial stretching at small deformations, and exhibit an increase in the degree of orientation or crystallinity with time. The strain-induced crystalline-sensitive absorption bands were found to be accentuated when the sample was subjected to strain-induced crystallization. Crystal phase transformations under elongation (Figure 3) have been followed in situ using IR spectroscopy.[11]

Finally, the distribution of additives in different phases of a semicrystalline polymer has also been studied using IR spectroscopy.[50] It can also provide valuable information about crystal structure changes induced by doping a given polymer with a compound such as iodine.[16,51] Therefore, valuable information about the conformation structure of semicrystalline polymers can be obtained from IR spectra. Table 1 provides a list of IR spectroscopy studies involving some common semicrystalline polymers.

2.4 Liquid-crystalline Polymers

IR spectroscopy has been widely used to study interactions between chains and conformational order in

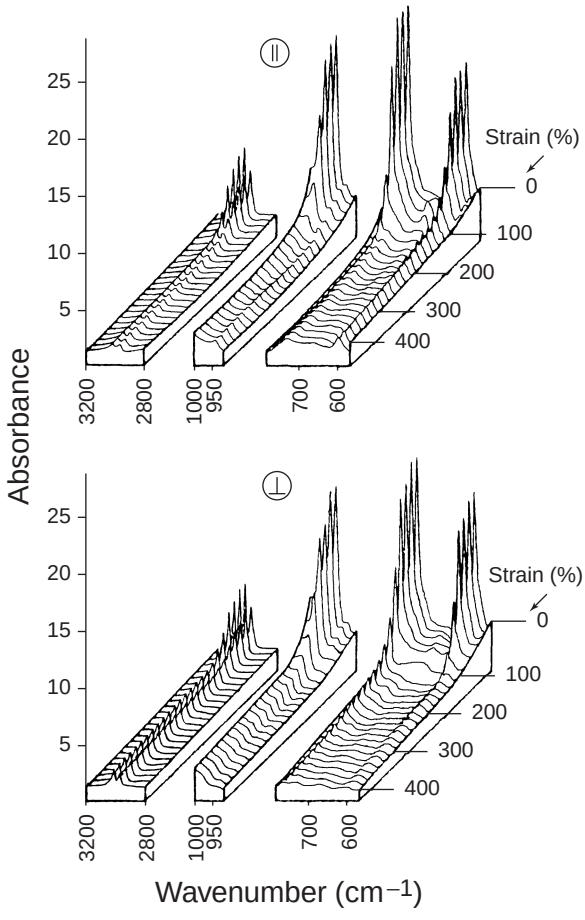

Figure 3 FTIR polarization spectra taken during elongation at 348 K of a poly(vinylidene fluoride) film showing a decrease in its initial II(α) crystal structure as it transforms into the all-*trans* I(β) form during elongation.[11] [Reprinted by permission from H.W. Siesler, *J. Polym. Sci. Polym. Phys. Ed.*, **23**, 2413–2422 (1985). Copyright © 1985 John Wiley & Sons Inc.]

polymers.[4] The molecular interactions are a function of the nature of the phase, which in turn determines the relative arrangement of functional groups. Hence, IR can be used to study the dependence of the mesophase microstructure (in LCs and LCPs) on molecular interactions. Specifically, analysis of absorption bands can provide information about the degree of order/disorder, molecular orientations under the influence of external fields, etc. Modern rapid-scan FTIR spectrometers can follow the kinetics of molecular motion with a time resolution of the order of nanoseconds per spectrum. The inherent limitation is the ability to record spectra of periodic events. Orientation relaxation and segmental mobility within LCs and LCPs are good examples of reproducible phenomena, and time-resolved FTIR can be used to investigate these phenomena at a molecular level. Examples of various molecular phenomena in LCs and LCPs investigated using IR spectroscopy are presented.

2.4.1 Molecular Orientation

Shibaev et al.[55,56] have shown that polarized IR spectroscopy can be used to study orientation in a mechanical field of LC polyacrylates and polymethacrylates involving cyanobiphenyl mesogenic groups in the side chains. The length of the mesogenic group was varied by varying the length of the methylene sequence within the side chain. Their results established that the type of orientation depends on the molecular weight of the polymer. The axes of the mesogenic groups were oriented along the shear direction for low-molecular-weight polymers whose degree of polymerization (DP) was <100, while they were normal to the shear direction when DP was >100. By estimating the degree of orientation as a function of temperature, Shibaev et al. were able to show that the content of gauche isomers in the different mesophases was different, thereby leading to different degrees of orientation.

Another example of mechanically induced orientation in side-chain LCPs is shown in the work of Zhao and Lei,[57] in which a thin LCP film was cast on the surface of a supporting polyvinyl alcohol film, which could be prepared and stretched over a wide temperature range. They obtained an order parameter of *ca.* 0.4 by IR dichroism measurements for the mesogenic groups at a draw ratio of 1.5, i.e. at a deformation of only 50%, indicating rapid macroscopic alignment of the nematic domains along the

Table 1 IR spectroscopic investigations of some common semicrystalline polymers

Polymer	Comments	Refs.
Polyethylene	Crystal splitting and chain conformations extensively investigated	7, 18–20, 24, 38, 50
Polypropylene	Orientation in crystalline and noncrystalline phases	7, 25, 52
Poly(ethylene terephthalate)	Evaluation of crystallinity	7–9, 22, 29, 42, 44, 53
Poly(vinyl alcohol)	Hydrogen bonding–intermolecular interactions investigated	7, 17, 28, 33–37, 47, 54
Polystyrene	Effect of tacticity on crystallinity	7, 23, 30, 40, 48
Polyurethanes	Chain conformations	12, 13
Poly(tetrafluoroethylene)	Crystallinity measurements	7, 32
Polyamides	Crystalline phase transformations	31, 49

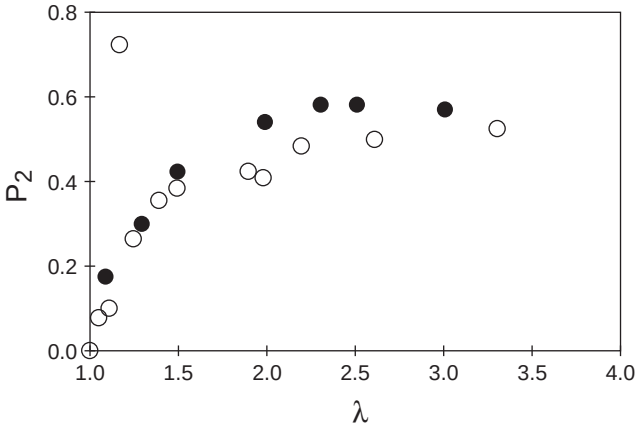

Figure 4 Order parameter as a function of draw ratio for side-chain LCPs. The closed circles represent the case where the polymer stretched at 85 °C, while the open circles represent polymer stretching at 102 °C.[57] [Reprinted from Y. Zhao, H. Lei, 'A Polarized IR Spectroscopic Study of Mechanically Induced Orientation in Side-chain Liquid Crystalline Polymers', *Polymer*, **35**, 1419–1424, Copyright 1994, with permission from Elsevier Science.]

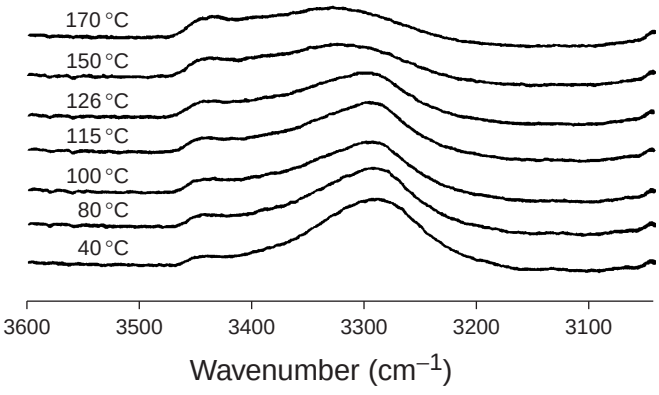

Figure 5 IR spectra of liquid crystalline polyurethane recorded from 40 to 170 °C in the N–H stretching region.[59] [Reprinted with permission from W. Tang, W.J. MacKnight, S.L. Hsu, *Macromolecules*, **28**, 4284–4289 (1995) Copyright 1995 American Chemical Society.]

direction of the applied mechanical field (Figure 4). They also quantified the effect of the flexible spacer (methylene sequence) group and demonstrated that a shorter spacer group provides a stronger coupling between the mesogenic groups and the backbone chain. This leads to favorable interaction between the nematic domains and the applied mechanical field.

Segmented block copolymers with alternating liquid crystalline hard and amorphous soft segments are a new class of thermoplastic elastomers.[58] The hard LC domains behave like physical cross-links and become macroscopically anisotropic during shear. Thus, highly oriented thermoplastic materials can be obtained. MacKnight et al.[59] investigated the effect of strain on orientation in liquid crystalline polyurethanes, which have applications as high-strength fibers. It was shown that mechanical strain orients the nematic mesophase in the direction of stretching. The N–H stretching vibrations in the high-temperature IR spectra (Figure 5) point towards microstructural changes during thermal transitions.

Similar molecular orientation studies have been performed[60] with photochromic Langmuir–Blodgett films, which show promising applications in the area of optical information storage media. The system under consideration was a polyion complex with azobenzene-containing fatty acid and polycation with LCs mixed with stearic acid-d_{35}. Studies were performed using FTIR transmission and reflection–absorption spectroscopy. The results revealed that the orientation of molecules in LCs is controlled by the *trans–cis* photoisomerization of the polyion complex at elevated temperatures, but the orientation is unaffected at room temperature. The state of the LC is different at these temperatures and this explains the difference in orientation.

Some LCPs exhibit a unique molecular structure at the surface, and this structure is completely different from that in the bulk. ATR/FTIR spectroscopy using polarized radiation has been used to characterize molecular orientation at the surface of stretched polymer sheets.[61–65] In addition, polarized specular reflection FTIR spectroscopy has been used[66,67] to characterize the surface orientation of extrusion-molded thermotropic LCPs. The reflection IR method is reliable for the analysis of the surface structure, but caution must be exercised for intense absorption bands because of the effects of the refractive index dispersion on the ATR spectra. The effect of the draw ratio, the flow characteristics, and the morphological structure on the orientational distributions at the surface of the LCPs has been studied.[67]

2.4.2 Segmental Mobility

The study of segmental mobility as a function of external (electric or electromagnetic) perturbation is fundamental in order to understand the dynamics of such molecular processes within LCs and LCPs. Siesler et al.[68–71] have pioneered studies of segmental mobility in liquid crystalline materials using FTIR spectroscopy. Some key results from their studies are discussed in this section.

Time-resolved FTIR spectroscopy was used[69] to study the segmental mobility of low-molecular-weight nematic LCs and a nematic LCP that had one of the LC units in its side-chain. The difference between the reorientation of the free LC and the bound LC was studied using two-dimensional correlation analysis. It was shown that molecules of the LCs reorient as

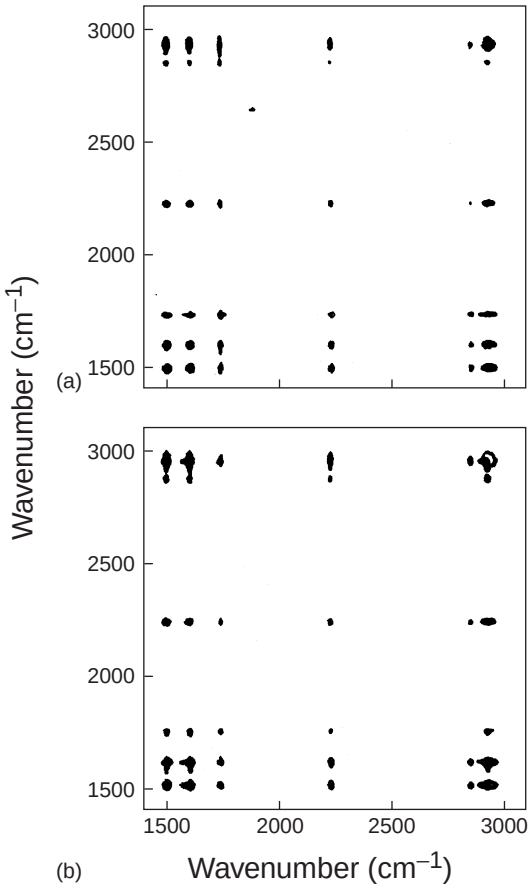

Figure 6 Synchronous two-dimensional correlation spectra[69] of the relaxation process of prealigned nematic LCs at (a) 45 °C and (b) 48 °C. [Reprinted with permission from M.A. Czarnecki, S. Okretic, H.W. Siesler, *J. Phys. Chem. B*, **101**, 374–380 (1997) Copyright 1997 American Chemical Society.]

rigid units (Figure 6). Within the nematic LCP, only the mesogen and a small part of the spacer directly attached to the mesogen actually reorient; the rest of the spacer and the backbone chain do not move under the influence of the electric field. The low-molecular-weight LCs were found to be unaffected by temperature and prealignment, while increasing temperature and prealignment significantly reduced the relaxation time of the nematic LCP.

In a related study, the laser-induced anisotropy in thin films of cyanoazobenzene side-chain LCPs based on polyesters selectively deuterated at different positions was investigated[71] using polarized FTIR spectroscopy. Analysis of the segmental orientation based on dichroic ratios of characteristic absorption bands showed that in the polyesters with long main-chain spacing, the light-sensitive azo chromophore, the methylene segment, and to a lesser extent the flexible spacer, were all preferentially oriented normal to the laser polarization. The degree of orientation increased with increasing spacer length. For shorter polyesters, only the chromophore and the spacer were aligned. On-line rapid-scan FTIR analyses revealed that the alignment of the aliphatic segments arose simultaneously with the chromophore orientation. The preservation of the photo-induced anisotropy was directly related to the polyester phase behavior by temperature-dependent IR studies.

Segmental mobility of a ferroelectric LCP was studied[70] using time-resolved FTIR spectroscopy and two-dimensional correlation analysis. It was shown that under the influence of an electric field, only the mesogen and a fragment of the spacer reorient with the attached terminal group. The remainder of the spacer and the backbone chain do not reorient. Similar results from IR spectroscopy studies on other side-chain LCPs lead to the conclusion that this behavior is a general phenomenon in LCPs in which the mesogen is attached to the polymer backbone by a flexible spacer. Estimated relaxation times for selected bands reveal that the reorientation rate increases with temperature. This effect was shown to be due to the decrease of the viscosity with temperature and hence temperature was shown not to have an effect on the reorientation mechanism. Based on such results from IR spectroscopy, new liquid crystalline systems can be made readily accessible for rapid correlation between chemical structure and electro-optical properties.

2.4.3 Photoinduced Reorientation in Dye-containing Liquid-crystalline Polymers

The photo-induced isomerization of dyes in LCPs has been widely investigated by researchers. In particular, mesogenic LCPs containing azobenzene moieties show a unique set of properties and have been proposed as media for reversible optical data storage.[72] Continuous irradiation with polarized light leads to a reorientation of the optical axes. Knowledge of the interactions between the isomerizing dye and the highly anisotropic LC environment can provide valuable information for designing optical storage media.

Spiess et al.[73,74] investigated two statistical LC side-group copolymers with 30 and 50 mol % of a mesogenic cyanoazobenzene-containing moiety as one of the comonomers and a cyanophenyl benzoate-containing moiety as the other. IR spectroscopy was used to simultaneously study the orientational behavior of these copolymers. Under continuous irradiation with polarized light above and below the glass transition temperature, T_g, of the polymers, both the azobenzene and phenyl benzoate side groups preferentially reorient normal to the plane of the film, leading to a strongly biaxial orientation distribution (Figure 7). The kinetics of reorientation are revealed to be a single exponential process,

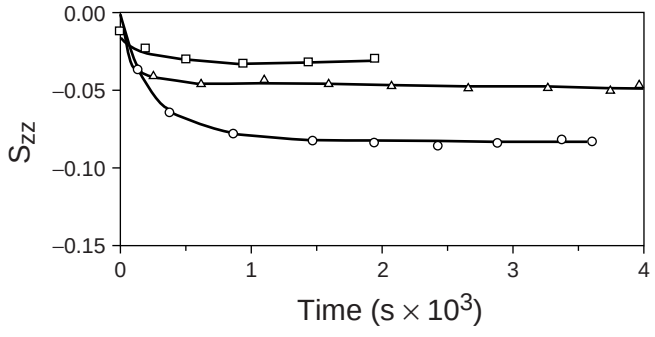

Figure 7 Order parameter for the azobenzene unit in a copolymer as a function of irradiation time.[74] The squares represent 298 K, the triangles 293 K, and the circles 263 K. All temperatures are below the T_g, which is 303 K. [Reproduced from U. Wiesner, N. Reynolds, C. Boeffel, H.W. Spiess, *Liq. Cryst.*, **11**, 251–267 (1992) by permission of Taylor & Francis Ltd.]

and an additional term is included for the azobenzene dye and is assigned to the initial *trans*-to-*cis* isomerization step. Studies performed on isotropic copolymer systems with azobenzene-containing side groups show that temperature-dependent anisotropy can be induced through irradiation with polarized light.

2.4.4 Phase Characterization/Transformations

The microstructural transformation and dynamics that accompany both monotropic[75] and enantiotropic[76,77] LCs have been of great interest to researchers. The use of vibrational spectroscopy has provided important information at the molecular level on phase transformations and dynamics in LCPs. MacKnight et al.[75] have used IR spectroscopy together with DSC and wide-angle X-ray scattering (WAXS) to study microstructural changes occurring in the different phases of mesogen-containing polyurethane elastomers. Their studies show that the mesophase-to-crystal exothermic transition is accompanied by an increase of H-bonding between the C=O and the NH groups. The crystal melting transition is accompanied by the disappearance of ordered H-bonded amide peaks.

IR spectroscopy has been used to study H-bonding effects[78] in a nematic liquid crystalline poly(ester amide). The NH stretching vibrations were shown to be stable and resulted in H-bonded structures over a wide range of temperatures. It was however shown that the strength of these linkages was very poor and did not affect the mechanical properties of the material significantly.

IR spectroscopy has been used to obtain important information regarding phase behavior in new classes of materials. For example, polymer-dispersed liquid crystals (PDLCs) are a new class of electro-optic material, with applications ranging from architectural glass to active matrix color projection televisions. PDLCs are composites of low-molecular-weight LC droplets dispersed in a polymer binder sandwiched between two transparent, conducting electrodes. In the absence of an electric field, the symmetry axes of the bipolar droplets are randomly oriented, and the refractive index of the LC droplet is mismatched with that of the polymer, resulting in an opaque appearance. In the presence of an electric field, the director orientation rotates to align the droplet with the field. The refractive index of the droplet and the polymer match and the film is transparent.

Various researchers have used IR spectroscopy to quantify the phase-separated microstructure within PDLCs. West and Ondris-Crawford[79] have incorporated dyes to improve the contrast of PDLC shutters. The dichroic IR absorption of the cyano stretch of the low-molecular-weight LC was used to calculate the composition of the droplets. This information can be used to optimally design films for visible operation.

Koenig et al.[80] used FTIR microspectroscopy to compare the polymer and liquid droplet regions within PDLC films made of three different LCs (E7, 5CB, and ZLI-1957/5) and polyvinylpyrrolidone and poly(methyl methacrylate). Their results showed that thermoplastic PDLC matrixes contain higher amounts of LC droplets than thermoset PDLC matrixes. IR functional group images of a droplet were used to show characteristic textures corresponding to the visual images of the same droplet (Figure 8). Conventional FTIR spectroscopy studies revealed that the C≡N and the CH_2 chain require less energy than the C=C chain to align with the field.

The dynamic behavior of PDLCs under the influence of an electric field has been analyzed[81] by static and two-dimensional IR spectroscopy. The PDLC samples were prepared by polymerization-induced phase separation of a mixture of nematic LC E7 and acrylate. Two-dimensional IR correlation analysis confirmed that the rigid core of the LC reorients as a unit. In addition, the authors showed that the polymer side chain, which exists at the interface between the polymer and the LC, reorients by interacting with the reoriented LC molecules.

3 COMPARISON OF INFRARED MEASUREMENTS WITH OTHER TECHNIQUES

XRD is the most commonly used and relatively accurate method of measuring polymer crystallinity.[11,14,16,18,24,29,30,32,42,43,46,51] XRD measures long-range order or intermolecular order as a result of chain packing. IR measures short-range order or intramolecular phenomena.[4] These two methods do

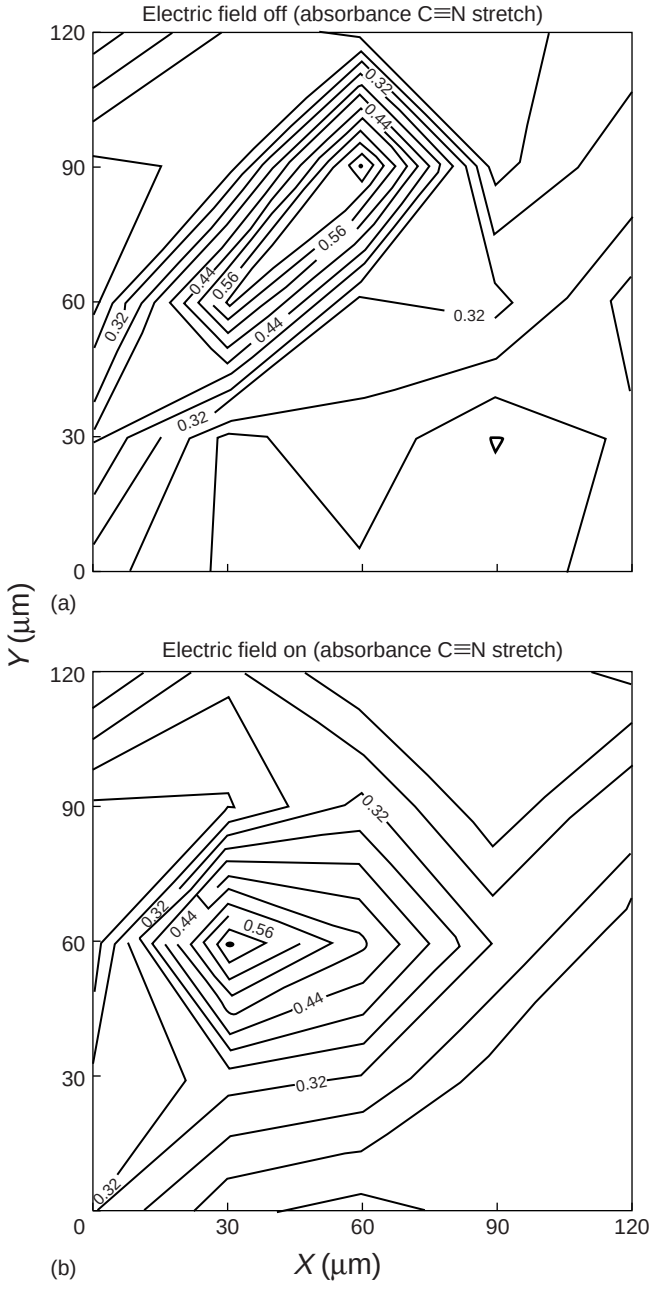

Figure 8 Two functional group images demonstrating the effect of an applied electric field. The orientation within the droplet changes with the field.[80] [Reproduced from C.A. McFarland, J.L. Koenig, J.L. West, *Appl. Spec.*, **47**(3), 321–329 (1993) by permission of the Society for Applied Spectroscopy.]

not necessarily measure the same crystallinity, since short-range intramolecular order is a necessary condition for long-range order to be present, but not vice versa. Therefore XRD and IR measure two different kinds of crystallinity, and IR can detect repeat units in a crystalline domain that can be undetected by XRD.

DSC measures the amount of heat required to melt a sample, and by using knowledge of heat of fusion of a 100% crystalline sample, the degree of crystallinity can be calculated.[4,25,27,31,39,44,46,49] DSC measures the crystallinity based on the meltable portion of the polymer, and represents a degree of order less than that of long-range XRD. However, the degree of crystallinity measured using DSC does not include the contributions of isolated ordered chains to short-range order. Therefore the method used to measure crystallinity needs to be specified for the reported results to be meaningful.

Raman spectroscopy has been used to characterize crystalline polymers. The Raman effect occurs because of an exchange of energy between the incident photons and the vibrational energy levels of the molecule. In many cases, the magnitudes of Raman shifts in the frequency correspond exactly to shifts in IR frequencies and therefore provide the same information. The differences between Raman and IR spectra provide valuable information, and the greater the symmetry of the molecule, the greater the difference between the two spectra.[4] Nonpolar molecular groupings give rise to strong Raman scattering, and molecules with polar groups give rise to strong IR absorption. Therefore, the vibrations of substituents on the carbon chain of polymers can be most easily studied by IR spectroscopy, and the vibrations of the carbon chain itself can be best studied using Raman spectroscopy.

Solid-state NMR is sensitive to crystal packing environments and has been used to characterize the amide configurations of various polymers such as polyamides.[31] However, these spectra are usually complex to decipher as compared with IR spectra. Along with changes in the carbon chemical shifts, the solid-state nitrogen NMR data show chemical shift changes with respect to chain conformation and packing. Nitrogen analysis, where possible, offers the advantages of simplified spectra and a larger spectral range, leading to better peak separation, but suffers from the disadvantages of low natural abundance and low magnetic susceptibility.

The combined use of neutron scattering and IR spectroscopy on mixed polyethylene crystals provides a very sensitive means of determining chain conformation. This is because the two techniques are complementary to a large extent. Neutron scattering is dependent on chain correlations over a wide range of distances,[19,20,24] while the IR technique is predominantly dependent on nearest-neighbor interactions.[20] These interactions could however extend over a much larger distance depending on the arrangement of the chains. Therefore, in summary, IR is a very valuable tool for analyzing crystalline polymers, and can be used in conjunction

with other techniques to provide complete structural information about the polymer.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflectance
DP	Degree of Polymerization
DSC	Differential Scanning Calorimetry
FTIR	Fourier Transform Infrared
IR	Infrared
LC	Liquid Crystal
LCP	Liquid-crystalline Polymer
NMR	Nuclear Magnetic Resonance
PDLC	Polymer-dispersed Liquid Crystal
XRD	X-ray Diffraction

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Near-infrared Spectroscopy of Polymers and Rubbers • Neutron Scattering in Analysis of Polymers and Rubbers • Supercritical Fluid Chromatography of Polymers • X-ray Scattering in Analysis of Polymers

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Gas Chromatography/Infrared Spectroscopy • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Liquid Chromatography/Infrared Spectroscopy • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. A.M. Donald, A.H. Windle, *Liquid-Crystalline Polymers*, Cambridge University Press, Cambridge, 1992.
2. R.V. Tal'roze, N.A. Plate, 'Structural Transformations of Thermotropic Liquid-crystal Polymers in Electric and Magnetic Fields', in *Liquid-Crystal Polymers*, ed. N.A. Plate, Plenum Press, New York, 1993.
3. P.-G. De Gennes, J. Prost, *The Physics of Liquid Crystals*, 2nd edition, Oxford University Press, Oxford, 1993.
4. J.L. Koenig, *Spectroscopy of Polymers*, American Chemical Society, Washington, 1992.
5. P.C. Painter, M.M. Coleman, J.L. Koenig, *The Theory of Vibrational Spectroscopy and its Application to Polymeric Materials*, Wiley Interscience, New York, 1982.
6. V.N. Nikimin, B.Z. Volchek, 'Influence of Intermolecular and Intramolecular Order on the Infrared Absorption Spectra of Polymers', *Russ. Chem. Rev.*, **37**(3), 225–239 (1968).
7. S. Krimm, 'Infrared Spectra of High Polymers', *Fortschr. Hochpolym. Forsch.*, **2**, 51–172 (1960).
8. R. Belali, J.M. Vigoureux, 'Transposition to Attenuated Total Reflection of Classical Laws Used in Transmission Spectroscopy: Application to Crystallinity Determination of Polyethylene Terephthalate Films', *Appl. Spec.*, **48**(4), 465–471 (1994).
9. L. Quintanilla, J.C. Rodriguez-Cabello, T. Jawhari, J.M. Pastor, 'Structural-analysis of Injection-molded Semicrystalline Polymers by Fourier-transform Infrared-spectroscopy with Photoacoustic Detection and Differential Scanning Calorimetry. 1. Poly(ethylene-terephthalate)', *Polymer*, **34**(18), 3787–3795 (1993).
10. H.W. Siesler, 'Rheo-optical Fourier-transform Infrared-spectroscopy of Polymers. 11. Strain-induced Crystallization of Polychloroprene', *Makromol. Chem., Rapid Commun.*, **6**(10), 699–705 (1985).
11. H.W. Siesler, 'Rheo-optical Fourier-transform Infrared-spectroscopy of Polymers. 9. Stretching-induced II(alpha)–I(beta) Crystal Phase-transformation in Poly (vinylidene fluoride)', *J. Polym. Sci., Part B: Polym. Phys.*, **23**(12), 2413–2422 (1985).
12. H.W. Siesler, 'Rheo-optical Fourier-transform Infrared-spectroscopy of Polymers. 14. Segmental Orientation and Strain-induced Crystallization of a Poly(ether urethaneurea) Elastomer', *Berichte Der Bunsen-Gesellschaft Fur Physikalische Chemie*, **92**(5), 641–645 (1988).
13. D.Y. Shen, S.K. Pollack, S.L. Hsu, 'Far-infrared Study of Hydrogen Bonding in a Semicrystalline Polyurethane', *Macromolecules*, **22**(6), 2564–2569 (1989).

14. P. Damman, J.J. Point, 'Crystalline-structure of the Poly-(ethylene oxide) *p*-Nitrophenol Complex. 2. Fourier-transform Infrared-spectroscopy', *Macromolecules*, **27**(14), 3919–3925 (1994).
15. T. Buffeteau, M. Pezolet, 'Photoinduced Orientation in Azopolymers Studied by Infrared-spectroscopy – Cooperative and Biaxial Orientation in Semicrystalline Polymers', *Macromolecules*, **31**(8), 2631–2635 (1998).
16. K. Tashiro, Y. Minagawa, M. Kobayashi, S. Morita, T. Kawai, K. Yoshino, 'Crystal Structural-change of Poly(3-alkylthiophene) Induced by Iodine Doping as Revealed by X-ray-diffraction and Infrared Raman-spectroscopic Measurements', *Jpn. J. Appl. Phys., Part 2*, **33**(7), L1023–L1026 (1994).
17. H. Tadokoro, S. Seki, I. Nitta, 'X-ray and Infrared Microspectroscopic Measurements on a Highly Double-oriented Specimen of Polyvinyl Alcohol', *J. Polym. Sci.*, **28**(11), 244–246 (1958).
18. K. Tashiro, S. Sasaki, M. Kobayashi, 'Structural Investigation of Orthorhombic-to-hexagonal Phase-transition in Polyethylene Crystal – The Experimental Confirmation of the Conformationally Disordered Structure by X-ray-diffraction and Infrared/Raman Spectroscopic Measurements', *Macromolecules*, **29**(23), 7460–7469 (1996).
19. S.J. Spells, 'Polyethylene Mixed-crystal Infrared-spectroscopy – The Influence of Matrix Isotopic-species on Guest Conformation', *Polymer*, **26**(13), 1921–1928 (1985).
20. S.J. Spells, D.M. Sadler, 'Mixed-crystal Infrared-spectroscopy of Annealed Solution-grown Crystals of Polyethylene', *Macromolecules*, **22**(10), 3948–3954 (1989).
21. H.W. Siesler, 'Rheo-optical Fourier-transform Infrared-spectroscopy of Polymers. 10. Strain-induced Crystallization of Sulfur-crosslinked Natural-rubber During Cyclic Deformation', *Appl. Spectrosc.*, **39**(5), 761–765 (1985).
22. S.S. Sheiko, I.S. Vainilovitch, S.N. Magonov, 'FTIR Spectroscopy of Polymer-films Under Uniaxial Stretching. 2. Amorphous and Semicrystalline Poly(ethyleneterephthalate)', *Polym. Bull.*, **25**(4), 499–506 (1991).
23. S.N. Magonov, I.S. Vainilovitch, S.S. Sheiko, 'FTIR Spectroscopy of Polymer Films Under Uniaxial Stretching', *Polym. Bull.*, **25**(5), 491–498 (1991).
24. E.U. Okoroafor, S.J. Spells, 'The Application of Mixed-crystal IR Spectroscopy to Molecular Deformation in Polyethylene by Rolling', *Polymer*, **35**(21), 4578–4589 (1994).
25. F.M. Mirabella, 'Determination of the Crystalline and Noncrystalline Molecular-orientation in Oriented Polypropylene by Infrared-spectroscopy', *J. Polym. Sci., Part B: Polym. Phys.*, **25**(3), 591–602 (1987).
26. M. Gavish, A.E. Woodward, 'Effects of Temperature on the Infrared Spectrum of Partially Crystalline *trans*-1,4-Polyisoprene', *Polymer*, **30**(5), 905–909 (1989).
27. N. Vasanthan, J.P. Corrigan, A.E. Woodward, 'Infrared Spectroscopic Investigation of Bulk-crystallized *trans*-1,4-Polyisoprene', *Polymer*, **34**(11), 2270–2276 (1993).
28. G. Hubca, M. Tomescu, M. Hillebrand, H. Iovu, 'Spectral Study of Modified Polyvinylic Alcohol Submitted to Chemical and Heat Treatments', *J. Appl. Polym. Sci.*, **58**(7), 1675–1680 (1995).
29. F. Kimura, T. Kimura, A. Sugisaki, M. Komatsu, H. Sata, E. Ito, 'FTIR Spectroscopic Study on Crystallization Process of Poly(ethylene-2,6-naphthalate)', *J. Polym. Sci., Part B: Polym. Phys.*, **35**(16), 2741–2747 (1997).
30. P. Musto, S. Tavone, G. Guerra, C. Derosa, 'Evaluation by Fourier-transform Infrared-spectroscopy of the Different Crystalline Forms in Syndiotactic Polystyrene Samples', *J. Polym. Sci., Part B: Polym. Phys.*, **35**(7), 1055–1066 (1997).
31. C.G. Johnson, L.J. Mathias, 'Structural Characterization of Nylon 7 by Solid-state Nuclear Magnetic Resonance Differential Scanning Calorimetry and Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy', *Polymer*, **35**(1), 67–75 (1994).
32. R.J. Lehnert, P.J. Hendra, N. Everall, N.J. Clayden, 'Comparative Quantitative Study on the Crystallinity of Poly(tetrafluoroethylene) Including Raman, Infrared and F-19 Nuclear-magnetic-resonance Spectroscopy', *Polymer*, **38**(7), 1521–1535 (1997).
33. N.A. Peppas, 'Infrared Spectroscopy of Semicrystalline Poly(vinyl alcohol) Networks', *Makromol. Chem.*, **178**(2), 595–601 (1977).
34. L.V. Smirnov, N.P. Kulikova, N.V. Platonova, 'Infrared Spectra of Polyvinylalcohol', *Vysokomol. Soedin.*, **A9**(11), 2515–2520 (1967).
35. R. Marupov, I.Y. Kalontarov, G.I. Konovalova, 'Infrared Spectra of Dehydrated Polyvinyl Alcohol', *Appl. Spectrosc.*, **8**(4), 401–402 (1968).
36. H. Tadokoro, H. Nagai, S. Seki, I. Nitta, 'Syntheses and Infrared Absorption Spectra of Deuterated Polyvinyl Alcohols', *Bull. Chem. Soc. Jpn.*, **34**(10), 1504–1511 (1961).
37. M.O. Ngui, S.K. Mallapragada, 'Understanding Isothermal Semicrystalline Polymer Drying: Mathematical Models and Experimental Characterization', *J. Polym. Sci., Part B: Polym. Phys.*, **36**(7), 2771–2780 (1998).
38. H. Hageman, R.G. Snyder, A.J. Peacock, L. Mandelkern, 'Quantitative Infrared Methods for the Measurement of Crystallinity and its Temperature Dependence: Polyethylene', *Macromolecules*, **22**(9), 3600–3606 (1989).
39. M. Shibayama, H. Takahashi, H. Yamaguchi, S. Sakurai, S. Nomura, 'Isothermal Crystallization of End-linked Poly(tetrahydrofuran) Networks', *Polymer*, **35**(14), 2944–2951 (1994).
40. T. Kimura, H. Ezure, S. Tanaka, E. Ito, 'In-situ FTIR Spectroscopic Study on Crystallization Process of Isotactic Polystyrene', *J. Polym. Sci., Part B: Polym. Phys.*, **36**(7), 1227–1233 (1998).
41. B.J. Bulkin, M. Lewis, J. Kim, 'Crystallization Kinetics of Poly(propylene terephthalate) Studied by Rapid-scanning Raman Spectroscopy and FT-IR Spectroscopy', *Macromolecules*, **20**(4), 830–835 (1987).

42. C.E. Miller, B.E. Eichinger, 'Determination of Crystallinity and Morphology of Fibrous and Bulk Poly(ethylene-terephthalate) by Near-infrared Diffuse Reflectance Spectroscopy', *Appl. Spectrosc.*, **44**(3), 496–504 (1990).
43. H. Ishida, M.T. Huang, 'Molecular-level Study of the Crystallization of a Thermoplastic Polyimide by Infrared-spectroscopy', *J. Polym. Sci., Part B: Polym. Phys.*, **32**(13), 2271–2282 (1994).
44. J. Jang, K. Sim, 'Spectroscopic Studies of the Crystallization Behavior in Poly(ether imide)/Poly(ethylene terephthalate) Blends', *Polymer*, **38**(16), 4043–4048 (1997).
45. P.G. Wang, A.E. Woodward, 'Infrared Spectroscopic Investigation of Solution-crystallized *trans*-1,4-Polybutadiene Copolymers', *Makromol. Chem., Macromol. Chem. Phys.*, **190**(4), 875–882 (1989).
46. J. Dybal, P. Schmidt, J. Baldrian, J. Kratochvil, 'Ordered Structures in Polycarbonate Studied by Infrared and Raman Spectroscopy, Wide-angle X-ray Scattering, and Differential Scanning Calorimetry', *Macromolecules*, **31**(19), 6611–6619 (1998).
47. K. Fuji, J. Ukida, 'Observations on the Infrared Absorption Band of Polyvinyl Alcohol at 916 cm^{-1}', *Makromol. Chem.*, **65**(7), 74–79 (1963).
48. X.M. Xie, A. Tanioka, K. Miyasaka, 'Fourier-transform Infrared Spectroscopic Studies of Gelation and Crystallization of Polystyrene Carbon-disulfide Solutions', *Polymer*, **34**(7), 1388–1392 (1993).
49. L. Quintanilla, J.C. Rodriguez-Cabello, J.M. Pastor, 'Structural-analysis of Injection-molded Semicrystalline Polymers by Fourier-transform Infrared-spectroscopy with Photoacoustic Detection and Differential Scanning Calorimetry. 2. Polyamide-6,6', *Polymer*, **35**(11), 2321–2328 (1994).
50. A. Singhal, L.J. Fina, 'Dynamic 2-Dimensional Infrared-spectroscopy of the Crystal-amorphous Interphase Region in Low-density Polyethylene', *Polymer*, **37**(12), 2335–2343 (1996).
51. K. Tashiro, M. Kobayashi, T. Kawai, K. Yoshino, 'Crystal Structural-change in Poly(3-alkyl thiophene)s Induced by Iodine Doping as Studied by an Organized Combination of X-ray-diffraction, Infrared/Raman Spectroscopy and Computer-simulation Techniques', *Polymer*, **38**(12), 2867–2879 (1997).
52. N. Kawamoto, H. Mori, K. Nitta, N. Yui, M. Terano, 'Characterization of the Differences in the Crystallinity from Surface to Bulk of Compression-molded Polypropene Sheets Using Attenuated Total-reflection Fourier-transform IR Spectroscopy', *Macromol. Chem. Phys.*, **197**(11), 3523–3530 (1996).
53. B.J. Bulkin, J. Kim, M. Lewin, 'Crystallization Kinetics of Poly(propylene terephthalate) Studied by Rapid-scanning Raman-spectroscopy and FT-IR Spectroscopy', *Macromolecules*, **20**(4), 830–835 (1987).
54. K. Fuji, T. Mochizuki, S. Imoto, J. Ukida, M. Matsumoto, 'Investigation of Stereoregularity of Poly(vinyl alcohol)', *J. Polym. Sci., Part A*, **2**(4), 2327–2347 (1964).
55. Ya.S. Freidzon, A.V. Kharitonov, V.P. Shibaev, N.A. Plate, 'Thermotropic Liquid Crystalline Polymers – 19. Peculiarities of the Liquid Crystalline Structure of Cholesterol-containing Polymers', *Eur. Polym. J.*, **21**, 211–216 (1985).
56. Z.A. Roganova, A.L. Smolyansky, S.G. Kostromin, V.P. Shibaev, 'Thermotropic Liquid Crystalline Polymers – 20. Study by IR-dichroism Spectroscopy of an Oriented State of Comb-like Liquid-crystalline Polymers with Cyanobiphenyl Mesogenic Groups', *Eur. Polym. J.*, **21**, 645–650 (1985).
57. Y. Zhao, H. Lei, 'A Polarized IR Spectroscopic Study of Mechanically Induced Orientation in Side-chain Liquid Crystalline Polymers', *Polymer*, **35**, 1419–1424 (1994).
58. W. Tang, R.J. Farris, W.J. MacKnight, C.D. Eisenbach, 'Segmented Polyurethane Elastomers with Liquid Crystalline Hard Segments. 1. Synthesis and Phase Behavior', *Macromolecules*, **27**, 2814–2819 (1994).
59. W. Tang, W.J. MacKnight, S.L. Hsu, 'Segmented Polyurethane Elastomers with Liquid Crystalline Hard Segments. 3. IR Spectroscopic Study', *Macromolecules*, **28**, 4284–4289 (1995).
60. T. Kawai, 'Molecular Orientation in Liquid Crystals on a Photochromic Polyion Complex LB Film by FTIR Spectroscopy', *Thin Solid Films*, **301**, 225–229 (1997).
61. C.S.P. Sung, 'A Modified Technique for Measurement of Orientation from Polymer Surfaces by ATR-IR Dichroism', *Macromolecules*, **14**, 591–594 (1981).
62. J.P. Hobbs, C.S.P. Sung, K. Krishnan, S. Hill, 'Characterization of Surface Structure and Orientation in Polypropylene and Poly(ethylene terephthalate) Films by Modified ATR-ID Dichroism Studies', *Macromolecules*, **16**, 193–199 (1983).
63. F.M. Mirabella, Jr, 'Surface Orientation of Polypropylenes. I. Theoretical Considerations for the Application of Internal Reflection Spectroscopy', *J. Polym. Sci., Polym. Phys. Ed.*, **22**, 1283–1291 (1984).
64. F.M. Mirabella, Jr, 'Surface Orientation of Polypropylenes. II. Determination for Uniaxially and Biaxially Oriented Films Using Internal Reflection Spectroscopy', *J. Polym. Sci., Polym. Phys. Ed.*, **22**, 1293–1304 (1984).
65. M.K. Gupta, D.J. Carlson, D.M. Wiles, 'Oriented Polyolefin Laminates: A Refractometry and IR Study', *J. Polym. Sci., Polym. Phys. Ed.*, **22**, 1011–1027 (1984).
66. A. Kaito, K. Nakayama, 'Surface Orientation in the Sheet of a Liquid Crystalline Poly(ester amide) Characterized by Reflection IR Spectroscopy', *Macromolecules*, **25**, 4882–4887 (1992).
67. A. Kaito, M. Kyotani, K. Nakayama, 'Orientation Distribution of Liquid-crystalline Polyester Sheets Studied by Polarized IR Spectroscopy', *J. Polym. Sci., Polym. Phys. Ed.*, **31**, 1099–1108 (1993).
68. S.V. Shilov, S. Okretic, H.W. Siesler, M.A. Czarnecki, 'Segmental Mobility of Liquid Crystals and Liquid Crystalline Polymers in an Electric Field: A Study

by Time-resolved Rapid-scan and Step-scan FTIR Spectroscopy', *Appl. Spec. Rev.*, **31**(1&2), 125–165 (1996).
69. M.A. Czarnecki, S. Okretic, H.W. Siesler, 'Reorientation of Nematic Liquid Crystals and Liquid Crystalline Polymers in an Electric Field Studied by FTIP Time-resolved Spectroscopy and 2D-correlation Analysis', *J. Phys. Chem. B*, **101**, 374–380 (1997).
70. M.A. Czarnecki, S. Okretic, H.W. Siesler, 'The Switching Process in a Ferroelectric Liquid Crystalline Side-chain Polymer by Time-resolved Step-scan FTIR Spectroscopy and 2D-correlation Analysis', *Vib. Spectrosc.*, **18**, 17–23 (1998).
71. C. Kullina, S. Hvilsted, C. Hendann, H.W. Siesler, P.S. Ramanujam, 'Selectively Deuterated Liquid Crystalline Cyanoazobenzene Side-chain Polyesters. 3. Investigations of Laser-induced Segmental Mobility by FTIR Spectroscopy', *Macromolecules*, **31**, 2141–2151 (1998).
72. M. Eich, J.H. Wendorff, 'Erasable Holograms in Polymeric Liquid Crystals', *Makromol. Chem., Rapid. Commun.*, **8**, 467–471 (1987).
73. U. Wiesner, N. Reynolds, C. Boeffel, H.W. Spiess, 'Photo-induced Reorientation in Liquid-crystalline Polymers Below the Glass Transition Temperature Studied by Time-dependent IR Spectroscopy', *Makromol. Chem., Rapid. Commun.*, **12**, 457–464 (1991).
74. U. Wiesner, N. Reynolds, C. Boeffel, H.W. Spiess, 'An IR Spectroscopic Study of Photo-induced Reorientation in Dye-containing Liquid-crystalline Polymers', *Liq. Cryst.*, **11**, 251–267 (1992).
75. F. Papadimitrakopoulos, E. Sawa, W.J. MacKnight, 'Investigation of a Monotropic Liquid Crystal Polyurethane Based on Biphenol-2,6-tolylene Diisocyanate, and a Six Methylene Containing Flexible Spacer. 2. IR Spectroscopic Phase Characterization', *Macromolecules*, **25**, 4682–4691 (1992).
76. V. Percec, A. Keller, 'A Thermodynamic Interpretation of Polymer Molecular Weight Effect on the Phase Transitions of Main-chain and Side-chain Liquid Crystalline Polymers', *Macromolecules*, **23**, 4347–4350 (1990).
77. F. Papadimitrakopoulos, E. Sawa, W.J. MacKnight, 'Novel Synthesis of Mesogenic *N*-Methyl Polyurethanes, and Demonstration of the Effect of Hydrogen Bonding on Liquid Crystalline Properties', *Polym. Prepr.*, **31**, 486–487 (1990).
78. H. Zhang, G.R. Davies, D.I. Green, H.V.St.A. Hubbard, I.M. Ward, 'IR Spectroscopic Studies of a Nematic Liquid Crystalline Poly(ester amide) in the Solid Phase', *Polymer*, **37**, 5817–5824 (1996).
79. J.L. West, R. Ondris-Crawford, 'Characterization of Polymer-dispersed Liquid-crystal Shutters by UV/VIS and IR Absorption Spectroscopy', *J. Appl. Phys.*, **70**(7), 3785–3790 (1991).
80. C.A. McFarland, J.L. Koenig, J.L. West, 'Analysis of Polymer-dispersed Liquid Crystals by IR Spectroscopy', *Appl. Spec.*, **47**(3), 321–329 (1993).
81. R. Hasegawa, M. Sakamoto, H. Sasaki, 'Dynamic Analysis of Polymer-dispersed Liquid Crystal by IR Spectroscopy', *Appl. Spec.*, **47**(9), 1386–1389 (1993).

Infrared Spectroscopy in Analysis of Polymer Degradation

Michele Edge
Manchester Metropolitan University, Manchester, UK

1 Introduction: Chemical Transformations during Polymer Degradation 1
1.1 Characterization and Analysis of Functional Group Changes 1
1.2 Changes in Molecular Weight 9
1.3 Structural Order and Degradation 9
2 Quantification of Functional Groups and the Kinetics of Polymer Oxidation 11
3 Role of Antioxidants in Polymer Degradation 11
4 Techniques 12
4.1 Pyrolysis Gas Chromatography and Infrared Spectroscopy 12
4.2 Attenuated Total Reflectance 13
4.3 Diffuse Reflectance Infrared Fourier Transform Spectroscopy 14
4.4 Photoacoustic Spectroscopy 14
4.5 Fourier Transform Infrared Emission 14
5 Advantages and Limitations of Infrared Methods 15
Abbreviations and Acronyms 17
Related Articles 17
References 17

Infrared (IR) spectroscopy locates the spectral position (frequency) and intensity of absorbed (or emitted) radiation arising from molecular transitions between quantum states associated with the vibrations and rotations of specific chemical groups. Using this technique both qualitative and quantitative information can be obtained on changes in the composition of polymers during degradation. The use of mathematical data-processing techniques and instrumental attachments has enabled considerable progress to be made in this field,[1–3] where changes to the composition of the polymer structure are invariably complex. In conventional IR spectroscopy the light is dispersed and each wavelength is examined separately, whereas the main advantage of Fourier transform infrared (FTIR) spectroscopy is that the signal is multiplexed and each point of an interferogram contains information on all the wavelengths present, enabling all frequencies to be determined simultaneously. The increased speed and higher signal-to-noise ratio of FTIR spectroscopy is the prime advantage in the analysis of complex materials. The use of computational software to obtain difference or derivative spectra enhances the dynamic range. This is of particular value when monitoring polymer degradation, where the main interference is from the nondegraded structure of the polymer. The main disadvantage of IR spectroscopy in identifying and quantifying the functional group changes that take place during polymer degradation is that broad and rather featureless bands are frequently produced. As a consequence it is necessary to resolve individual components by subjecting the sample to a prior chemical treatment, or by the use of coupled instrumentation.

1 INTRODUCTION: CHEMICAL TRANSFORMATIONS DURING POLYMER DEGRADATION

Reference sources on polymer degradation are extensive,[4–9] and a number of journals cover current awareness of the subject area, in particular the specialist journal *Polymer Degradation and Stability*.[10]

The degradation in polymers is an irreversible process in which, of all its possible mechanisms, oxidation is usually dominant. Purely thermal degradation is difficult to achieve because it is difficult to exclude oxygen. This is especially true in processing operations when the polymer introduced as powder, granules or pellets will contain oxygen dissolved in, or occupying the spaces between, particles. According to the literature,[11] the incorporation of one molecule of oxygen per hundred polymer chain repeat units can reduce molecular weight by a factor of two to three, and lead to a total embrittlement of the polymer. Figure 1 illustrates inadequately stabilized PVC roofing that is brittle and opaque after outdoor weathering. Embrittlement itself is underpinned by changes to the crystallinity of the polymer. Hence, IR spectroscopy can provide valuable information on polymer degradation kinetics and mechanisms by identifying oxidation products (functional group changes), changes in molecular weight, and changes in structural order during the process of polymer breakdown.

1.1 Characterization and Analysis of Functional Group Changes

Oxidation results in polymer chain scission, crosslinking and secondary oxidation reactions leading to

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

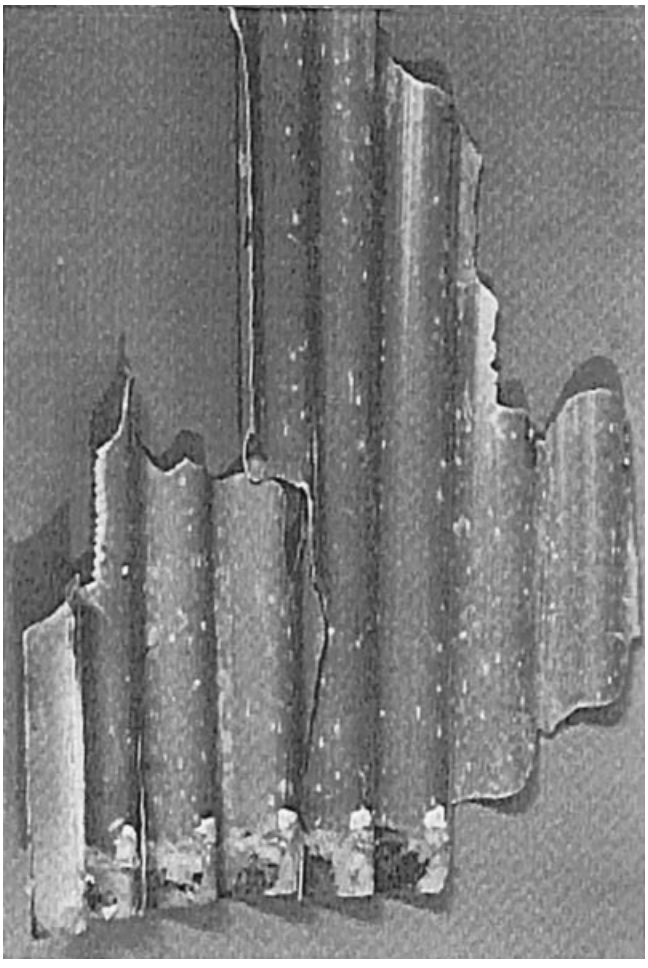

Figure 1 Inadequately stabilized rigid poly(vinyl chloride) (PVC) roofing, brittle and opaque following outdoor exposure.

the formation of new functional groups. The mechanism involves a radical chain reaction, comprising initiation, propagation (chain branching, chain scission) and termination (Scheme 1).

1.1.1 Initiation

There is an ongoing debate as to how the initiation step in polymer oxidation arises, but considering the heterogeneous nature of the process it is highly likely that breakdown begins at any impurity centers, weak links, or is due to shear and thermal scission during processing. The key feature of initiation is that macroalkyl radicals are formed, and if initiation does not greatly influence the rate, course or extent of degradation then the former arguments are of little issue here. Furthermore, impurity centers are frequently present at concentrations as low as 0.5 per 1000 polymer repeat units, and as such IR spectroscopy lacks the necessary sensitivity for their characterization/quantification.

$RH \longrightarrow R\bullet$	Initiation
$R\bullet + O_2 \longrightarrow ROO\bullet$ $ROO\bullet + RH \longrightarrow ROOH + R\bullet$	Chain propagation
$ROOH \longrightarrow RO\bullet + \bullet OH$ $RH + \bullet OH \longrightarrow R\bullet + H_2O$	Chain branching
$RO\bullet \longrightarrow$	Chain scission
Radical combination $\longrightarrow$ Cross-linked products	Termination

RH = Polymer
R• = Polymer alkyl radical
RO• = Polymer alkoxy radical
ROO• = Polymer peroxy radical
ROOH = Polymer hydroperoxide
•OH = Hydroxy radical

Scheme 1

1.1.2 Propagation and the Formation of Peroxides, Cyclic Peroxides and Epoxides

Oxygen reacts with polymer macroalkyl radicals to yield polymer peroxy radicals, which abstract hydrogen by intra- and/or intermolecular processes to form polymer hydroperoxides according to Scheme 2. When hydroperoxides are formed they may be identified by the combined absorptions of hydroxy and hydroperoxy groups in the region 3300–3650 cm^{-1}. Figure 2 shows the changes in this region during the irradiation of polybutadiene over 67 h. When degraded over a similar period of time, thermally oxidized polyolefins show hydroxyl/hydroperoxide bands of about half the absorbance values of photooxidized polyolefins. Quantitative determination of hydroperoxides by IR spectroscopy generally shows good agreement with other chemical methods employed for their determination.[12]

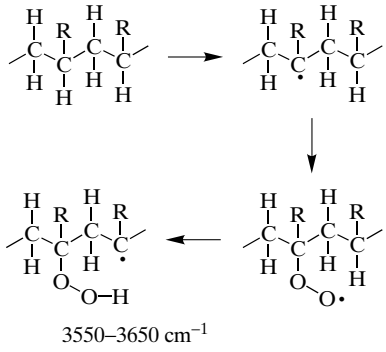

Scheme 2

It should be appreciated that polymer peroxy radicals in close positions in the same chain may terminate to

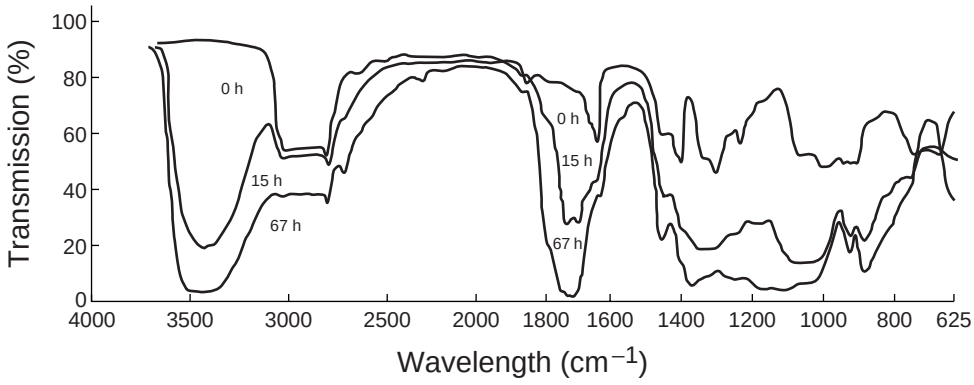

Figure 2 IR spectrum of polybutadiene after 0, 15 and 67 h of irradiation. (Reproduced from Beavan and Philips, *Eur. Polym. J.*, **10**, 593 (1974), Pergamon Press Ltd.)

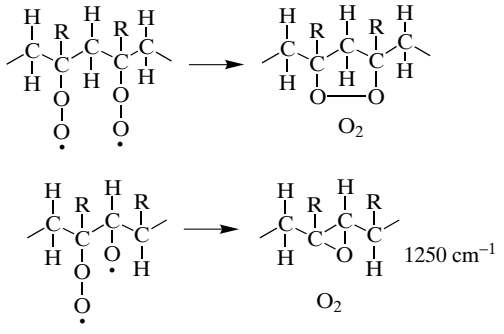

Scheme 3

form stable cyclic peroxides or epoxides according to Scheme 3.

Carlsson et al. have demonstrated that exposure of preoxidized polyolefin films to SO_2 caused >90% loss in hydroperoxide group IR absorbances.[13] This is illustrated by the subtracted spectra of treated and untreated polyethylene (PE) and polypropylene (PP) films given in Figure 3. In both cases new absorptions appear at 900, 1040 and 1195 cm^{-1}. The reaction of SO_2 with hydroperoxide groups proceeds by two consecutive reactions, depending on the relative ratio of SO_2 to the hydroperoxide groups (Scheme 4). When there is an excess of SO_2, hydosulfates are predominantly formed, but when the local peroxide concentration is dominant dialkyl peroxides are formed. This introduces uncertainty in the measurements and so films are exposed to gaseous NH_3 following SO_2 exposure (Scheme 5). This results in the formation of ammonium sulfate, as evinced by new bands at 1420, 1105 and 620 cm^{-1}. The extinction coefficients of these bands are relatively high and hence afford enhanced sensitivity of OOH detection, despite the variable stoichiometry of reaction. Because SO_2 reacts selectively with hydroperoxide groups, this method may be used to differentiate them from hydroxyl groups.

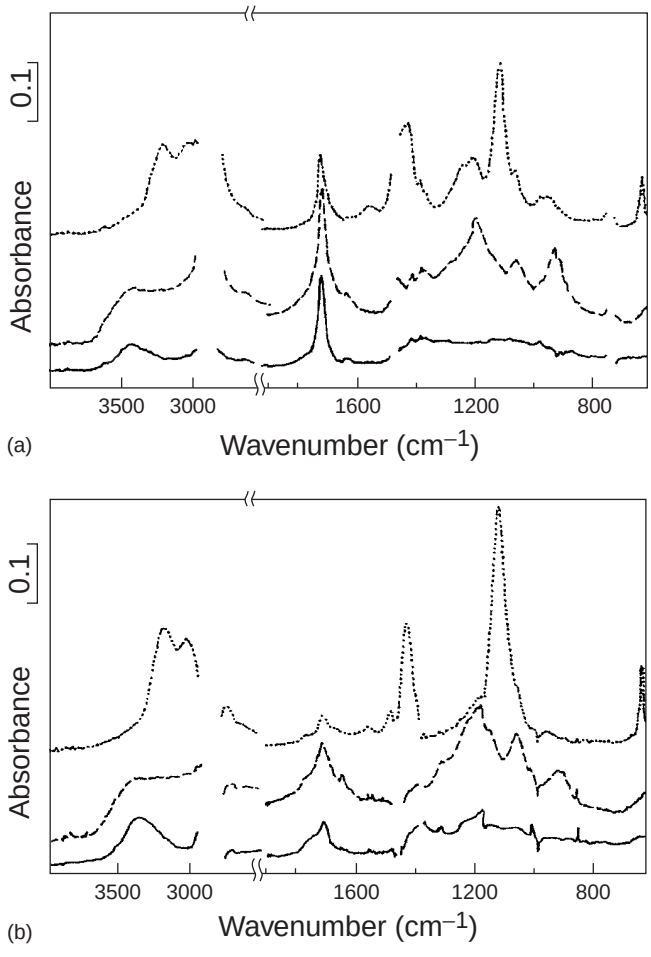

Figure 3 Subtraction IR spectra of PE (a) and PP (b) films: (———) unoxidized; (- - - - -) after γ-irradiation; (· · · ·) after exposure to SO_2 and NH_3. (Reproduced from Carlsson et al., *Polym. Degrad. Stab.*, **15**(1), 70 (1986), Elsevier Applied Science.)

The collective determination of hydroperoxide and hydroxyl groups may be executed by exposure to nitric

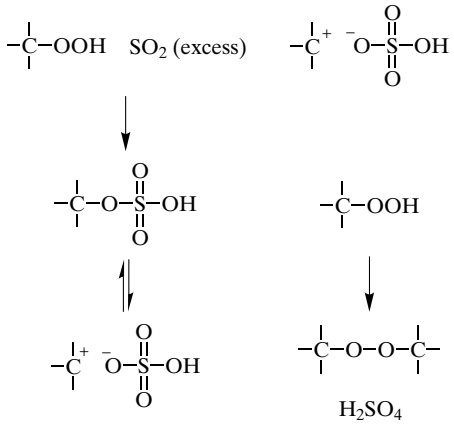

Scheme 4

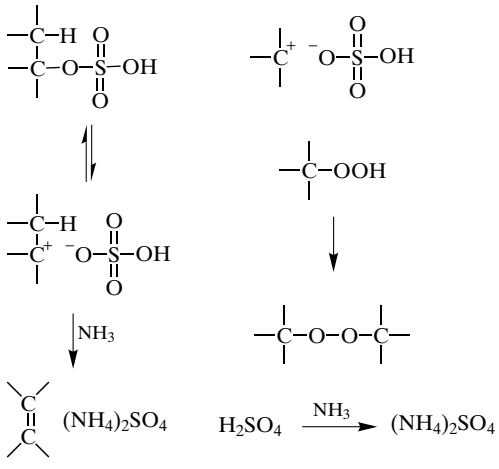

Scheme 5

oxide (NO) for 2–150 h, depending on the specific polymer sample. Because NO alone produces complex spectra, due to the formation of a variety of nitrates and nitrites, it is more often used in conjunction with sulfur tetrafluoride (SF_4).[14–16] After reaction with NO, the OH and OOH groups are separated by the formation of nitrites and nitrates, respectively. In this way primary, secondary and tertiary products can be resolved; moreover, the molar decadic absorptivity of the NO adducts is significantly higher than the original OH and OOH groups, thereby enhancing detection sensitivity. On addition of SF_4, all types of groups containing a hydroxyl component are fluorinated, with OOH and OH groups being converted to fluorides. On the other hand, reaction with NO_2 produces a complex mixture of products, with strong absorptions at 1700, 1300 and 925 cm^{-1} due to the formation of nitric acid.[16]

In addition, absorptions of hydroxyl groups may be separated from hydroperoxide groups by reaction with either diisopropoxytitanium bis(acetylacetonate) or chloroacetyl chloride (Scheme 6).[17]

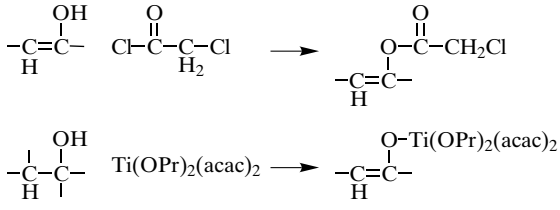

Scheme 6

1.1.3 Chain Branching and the Formation of Carbonyl Groups

Carbonyl groups are formed during the chain-branching steps of oxidation. Thermolysis and/or photolysis of hydroperoxides results in the formation of polymer oxy radicals, which participate in β-scission reactions of the polymer chain to form either ketone or aldehyde groups and a radical chain-end (Scheme 7).

The acids, ketones, aldehydes, esters and lactones produced in such oxidation steps result in broad and rather featureless bands extending from 1650 to 1850 cm^{-1} (Figure 4). Kato et al. have shown that such a typical band may be resolved into six overlapping peaks (Figure 5).[18] For precise quantitative determination of specific groups it is necessary to separate the individual absorptions. The absorptivity of acid groups has been reported to be 2.4 times greater than that of ketones and 3.1 times greater than that of aldehydes.[19] Rugg et al.[20] have demonstrated that during thermal oxidation the carbonyl groups produced are predominantly ketonic, whereas during photooxidation roughly equal quantities of acid ketone and aldehyde groups are formed. Using FTIR

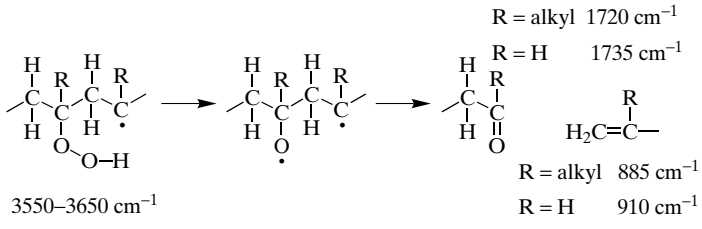

Scheme 7

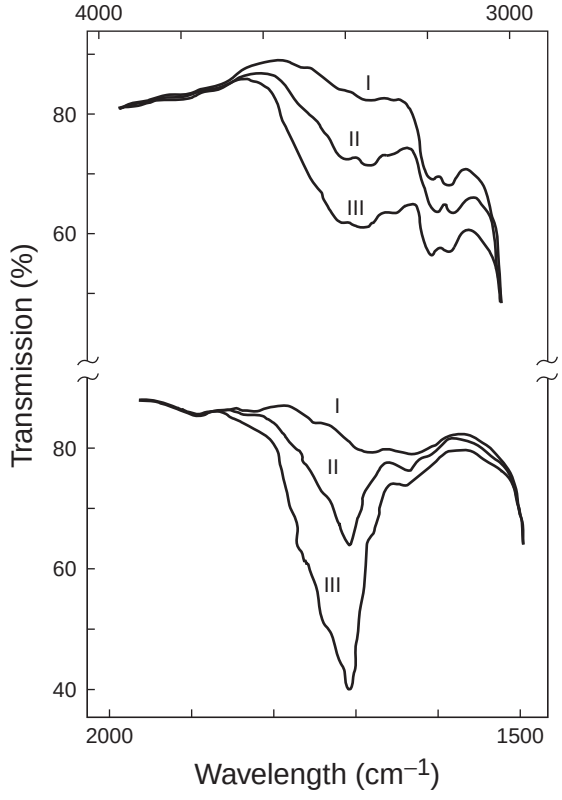

Figure 4 IR spectra of the photooxidation products in PP film after irradiation for 0 (I), 6 (II) and 110 h (III). (Reproduced from Wiles, *Polym. Eng. Sci.*, **13**, 74 (1973), Society of Plastics Engineers.)

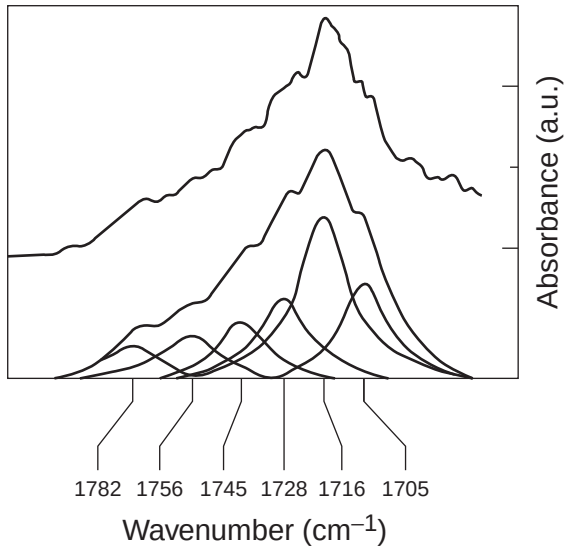

Figure 5 Composition of the carbonyl absorption of photooxidized PP film. (Reproduced from Kato et al., *J. Appl. Polym. Sci.*, **13**, 1447 (1969), John Wiley & Sons.)

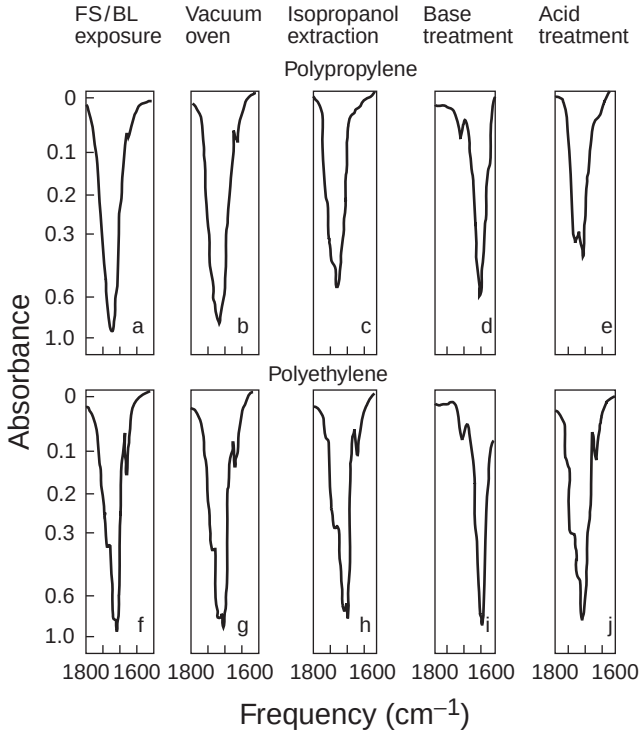

Figure 6 IR spectra of photodegraded PP (a–e) and PE (f–j) following various treatments. FS/BL exposure = fluorescent sun lamp/black lamp. (Reproduced from Adams, *J. Polym. Sci. Polym. Chem. Ed.*, **8**, 1279 (1970), John Wiley & Sons.)

spectroscopy, Tabb et al. have shown that the concentrations of aldehyde groups decrease during the photooxidation of PE, but that the ketonic groups increase.[21]

For polyolefins, the identification and analysis of the various groups comprising the carbonyl band has been achieved by various chemical treatments, as summarized in Figure 6.[22] Films may be placed in a vacuum oven to remove volatile products, but this does not resolve the carbonyl band. Extraction of the polymer with isopropanol removes part of the carbonyl absorption, leaving a narrow band centered at 1740 cm^{-1} that may be assigned to ester groups. The absorption band at 1790 cm^{-1} is attributed to γ-lactone and that at 1715 cm^{-1} to carboxylic acid. Treatment of the films with a base converts the lactones, esters and carboxylic acids to carboxylates, which shifts the main absorption band to 1580 cm^{-1}, leaving a small band at 1720 cm^{-1} due to aldehydes and ketones. Upon acidification of the base-treated polymer samples, the orginal esters absorbing at 1740 cm^{-1} are not regenerated but are converted into free acids. On this basis it was shown that during the photooxidation of PE only ketones were observed, but for PP both ketones and aldehydes were shown to be present. The carbonyl group of a carboxylic acid may be distinguished from that of a ketone by the addition of molar sodium hydroxide, which results in

the formation of a new band due to the carboxylate anion at 1580 cm^{-1} as the carbonyl band at 1715–1720 cm^{-1} is reduced.

The broad carbonyl absorption may be resolved better by chemical derivatization methods. Keto groups may be differentiated from aldehyde groups by tagging with specific reagents, namely pentafluorophenyl hydrazine and 2,4-dinitrophenyl hydrazine (Scheme 8).[23] In each case the ketone absorption at 1720 cm^{-1} is reduced and a new band due to C=N appears at 1650–1690 cm^{-1}. This is useful as a measure of oxidation but to isolate a wider range of carbonyl groups a more sophisticated approach is required.

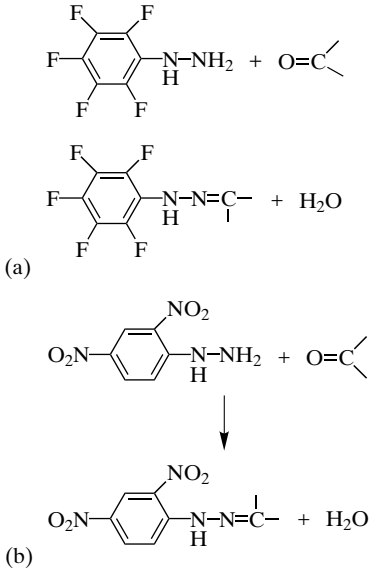

Scheme 8

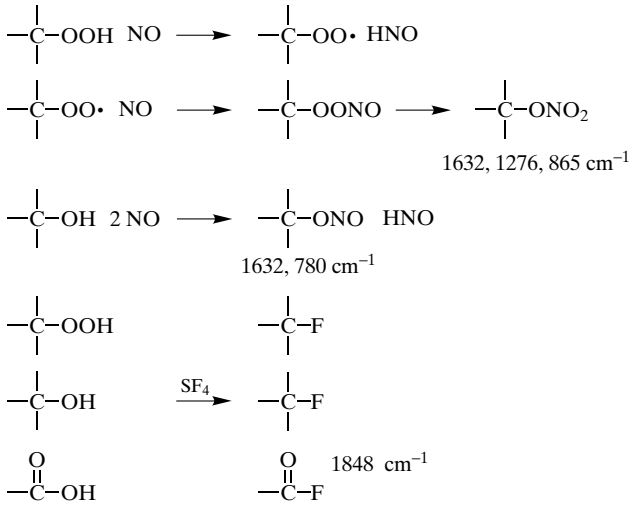

Scheme 9

Lacoste et al. employed γ-irradiation to produce a simple and controlled mixture of polyolefin oxidation products, and compared this with thermally or photo-oxidized polyolefins.[24] Films were treated with NO followed by SF_4 to monitor changes in both carbonyl and hydroperoxide groups (Scheme 9). To enable accurate measurement of IR absorbances, films were tilted at the Brewster angle and a polarizer placed in the beam, which according to Harrick eliminated the interference ripples that arise from polymer films.[25] Figure 7 shows LLDPE films after oxidation and following derivatization. For NO-treated samples, nitrate peaks centered at 1276 and 865 cm^{-1}, corresponding to secondary hydroperoxides, are seen as negative peaks following spectral subtraction. A nitrite peak at 778 cm^{-1}, indicative of secondary hydroxyl, is positive after subtraction. The negative

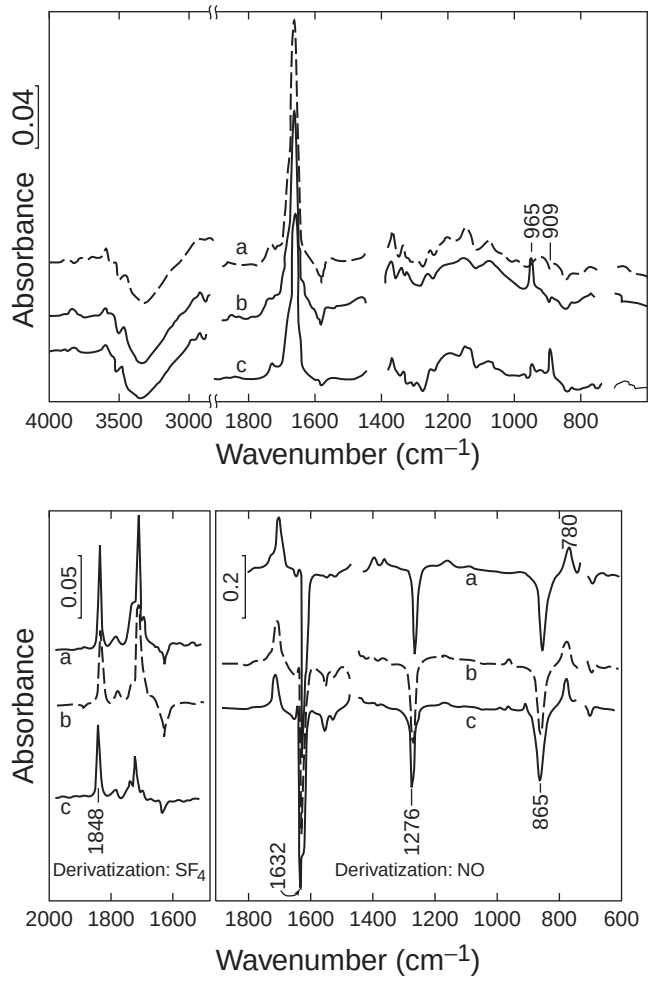

Figure 7 FTIR changes in linear low-density polyethylene (LLDPE) films before and after derivatization: (a) after 96 h of thermal decomposition at 85 °C; (b) after 10 Mrad of γ-irradiation; (c) after 69 h of xenon arc exposure. (Reproduced from Carlsson et al., *Polym. Degrad. Stab.*, **34**, 309–323 (1991), Elsevier Applied Science.)

peaks at 1632 cm^{-1} are characteristic of a mixture of these previous absorptions. On treatment with SF_4, a band due to acid fluoride at 1848 cm^{-1} is produced by reaction with carboxylic acid end-groups. Heacock had previously observed this latter absorption, but centered around 1835 cm^{-1}.[26] This greatly simplifies the carbonyl region, allowing ketone, ester and γ-lactone absorptions to be distinguished.

1.1.4 Unsaturated Groups

If hydroperoxides are incorporated adjacent to unsaturated impurities, then α,β-unsaturated carbonyl groups will result (Scheme 10). The α,β-unsaturated carbonyl groups so formed may isomerize to give β,γ-unsaturated carbonyl groups according to Scheme 11. Carbonyl groups may be photolyzed by Norrish Type I and II processes to yield: in the former case, carbon monoxide and a further radical chain-end; in the latter case, end-chain vinyl groups and carbonyl groups (Scheme 12).[27]

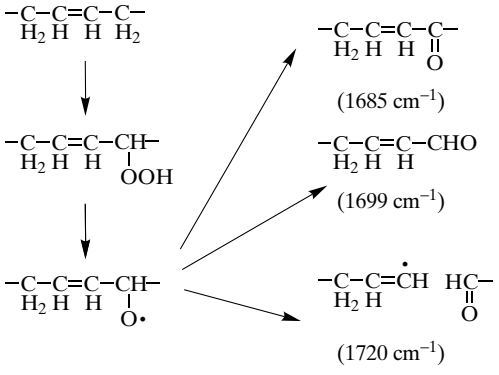

Scheme 10

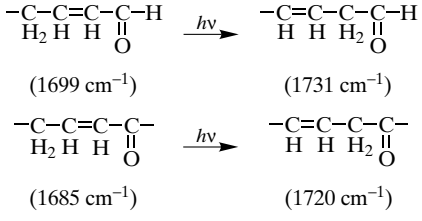

Scheme 11

The formation of unsaturated groups is best illustrated by reference to polyolefin oxidation (Figure 8). Rueda et al. have used IR spectroscopy to measure vinyl, vinylidene and internal *cis* or *trans* olefinic end-groups in PE, which can be used as a basis to identify any changes in these groups.[28] Prior to photooxidation, PE exhibits both vinyl and vinylidene at 909 cm^{-1} and

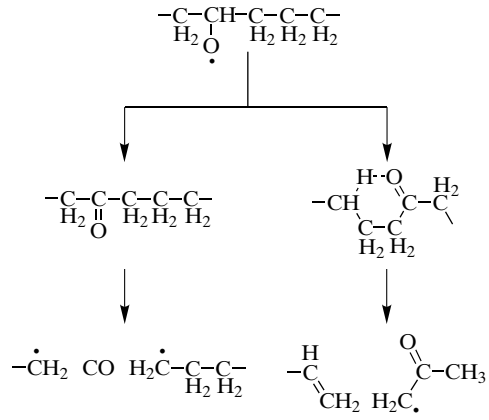

Scheme 12

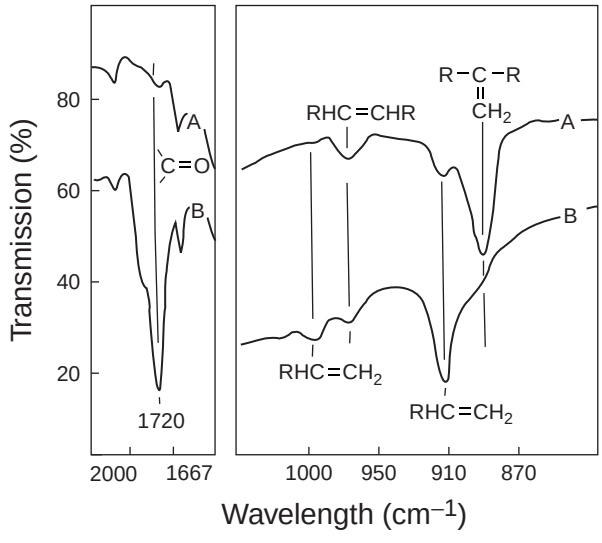

Figure 8 IR spectrum of low-density polyethylene (LDPE) (10-mL film, molecular weight = 10 000) before (A) and after (B) photooxidation. (Reproduced from Rugg et al., *J. Polym. Sci. Polym. Chem. Ed.*, **13**, 535 (1977), John Wiley & Sons.)

887 cm^{-1}. During irradiation the vinylidene groups are destroyed and the vinyl content increases, in parallel with carbonyl formation (Figure 9).[29]

In other polymers, such as PVC, thermal oxidation results in the formation of carbonyl groups and poly-conjugated sequences of double bonds. Figure 10 shows changes in the IR spectrum of a PVC film subjected to thermal oxidation at 140 °C over 160 h.[30] A broad carbonyl band is evident between 1650 and 1850 cm^{-1}, and overlapping this is a peak due to the development of unsaturation centered at 1605 cm^{-1}. Dienes and trienes usually absorb around 1650–1600 cm^{-1}, whereas the band for polyenes is shifted to lower wavenumbers (1650–1580 cm^{-1}). Although IR spectroscopy is able to detect the presence of unsaturation and indicate that

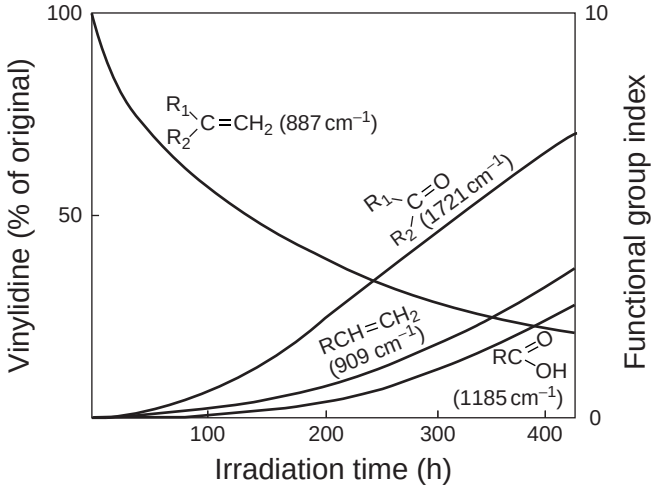

Figure 9 Change in the composition of functional groups in LDPE. (Reproduced from Amin et al., *Eur. Polym. J.*, **11**, 85 (1975), Pergamon Press.)

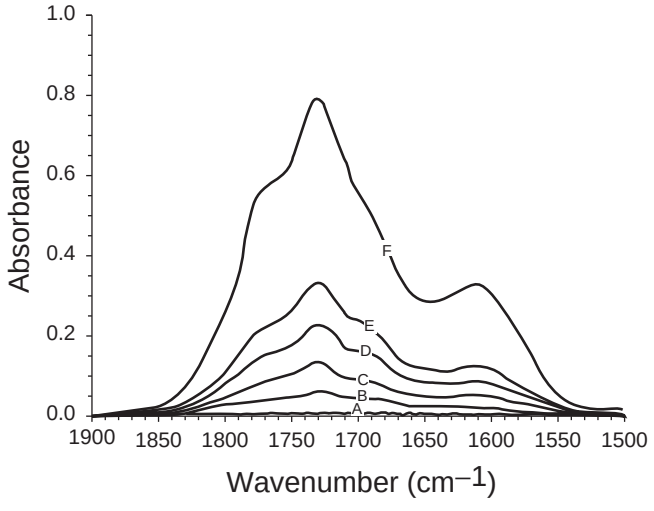

Figure 10 Changes in the IR spectrum of a PVC film before (A) and after being submitted to thermal oxidation at 140 °C for (B) 20 h, (C) 45 h, (D) 70 h, (E) 90 h, (F) 160 h. (Reproduced from Gardette et al., *Polym. Degrad. Stab.*, **34**, 149 (1991), Elsevier Applied Science.)

polyene sequences of varying length are formed, by the broadening of the IR band, it is unable to give information on the length of the polyene sequences.

In contrast, for polymers already containing significant unsaturation, such as polydienes, IR spectroscopy has been useful in distinguishing different types of unsaturation and how important different double bonds are in thermal oxidation mechanisms. In this respect, Brako and Wexler have exposed polybutadiene films to bromine vapor, noting the resulting marked changes (Figure 11).[31] By this treatment, bands at 730, 910, 965

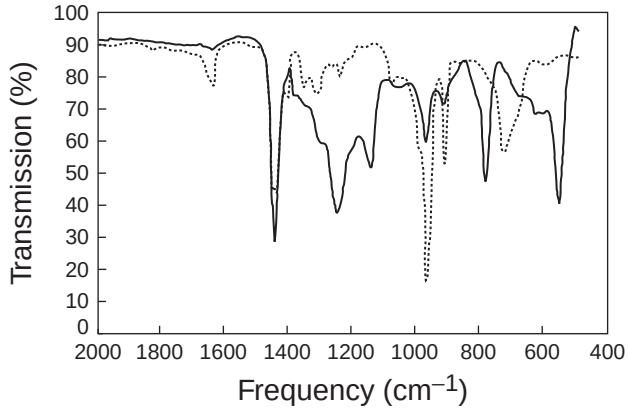

Figure 11 IR spectra of polybutadiene before (·····) and after (———) exposure to bromine vapor. (Reproduced from T.R. Crompton, *Analysis of Polymers: an Introduction*, 92, 1989, Pergamon Press.)

and 1640 cm^{-1} associated with unsaturation are reduced and new bands associated with bromination of double bonds appear at 550, 785, 1145 and 1250 cm^{-1}.

Other workers have also capitalized on this methodology for PE. Unsaturation in LDPE has been estimated as ± 0.003 C=C 10^{-3} carbon atoms by compensating with brominated polymer of the same thickness.[32] Dankovics has determined the total degree of unsaturation in LDPE as the sum of vinyl, vinylene and vinylidene groups, using difference spectra between an unbrominated polymer film (sample) and a brominated film (reference).[33]

1.1.5 Propagation and the Formation of Acids and Peracids

Any end-of-chain aldehyde groups may react with polymer peroxy radicals to produce polymer acyl radicals. The acyl radicals may be oxidized further to peracid radicals, which can hydrogen abstract to form polymer peracids, with IR absorptions centered at 3280, 1785 and 950 cm^{-1}. In turn, cleavage of the peracid groups results in the formation of carboxylic acids at 1710 cm^{-1} (Scheme 13).

1.1.6 Chain Scission and Cross-linking

Chain scission may be determined directly by the production of new functional groups and chain-ends in the polymer, and indirectly as a reduction in the molecular weight of the polymer (see section 1.2). Cross-linking reactions result from termination reactions, and may link polymer chains or chain fragments either without or with the incorporation of oxygen. In the latter case this will result in the formation of a peroxide group or an ether group (Scheme 14). Cross-linked material is frequently intractable in nature, as such direct identification by IR spectroscopy is often not possible and it

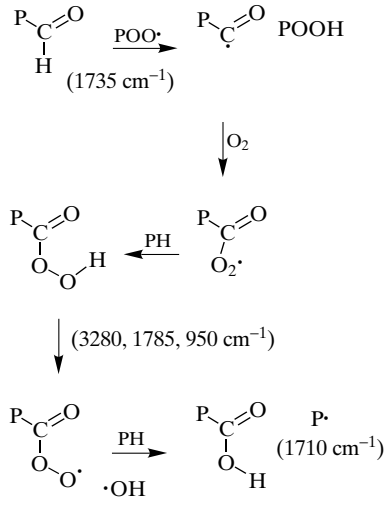

Scheme 13

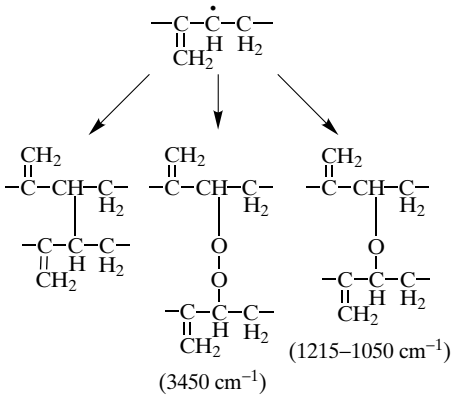

Scheme 14

is necessary to resort to more sophisticated techniques (refer to section 5.1).

1.2 Changes in Molecular Weight

IR measurements of molecular weight depend largely on the detection of end-groups. Here, a number-average molecular weight may be found, provided that the polymer chain is linear with few branch points. This is of particular value when examining the hydrolytic or oxidative degradation of condensation polymers. Low-molecular-weight analogs may be used to calibrate the spectra, because in polymer chains, end-groups are not coupled to the main chain. Figure 12 illustrates the IR spectra of two samples of poly(butylene terephthalate) of different molecular weight.[34]

In some instances end-groups give intense and readily observable vibrational absorptions. This is the case for PE terminal vinyl groups (909 and 990 cm^{-1}). Such groups are detectable because their dipole moment is stronger

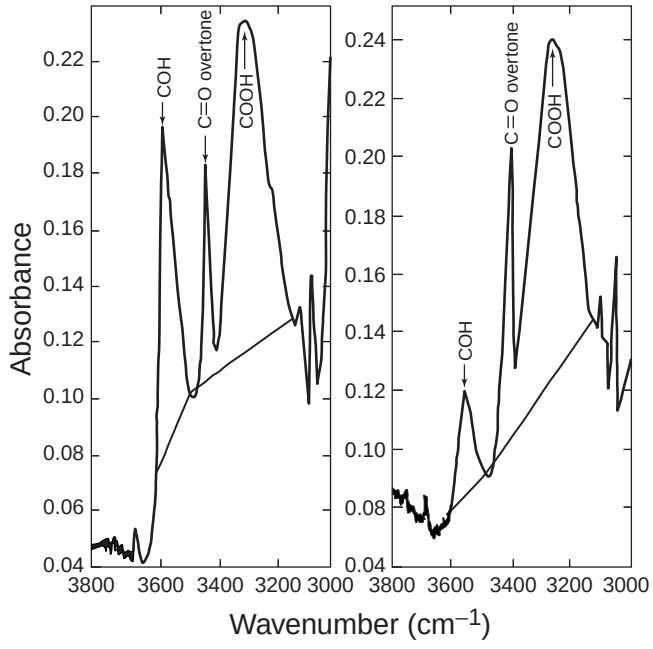

Figure 12 IR spectra of two samples of poly(butylene terephthalate) with different molecular weights. (Reproduced from Kosky et al., *Polym. Eng. Sci.*, **25**, 389, 1985.)

than those of constituent methylene wagging and twisting modes. The absorptions of end-groups can be accentuated by subtracting a high-molecular-weight polymer spectrum from the spectrum of a low-molecular-weight analog.

However, this approach is not applicable to many polymers where absorptions of end-groups may be weak. For example, methyl end-groups at 1379 cm^{-1} not only give rise to a weak absorption but are overlapped by a larger number of amorphous methylene modes located in the same region of the IR spectrum. Furthermore, IR spectroscopy cannot provide direct information on the distribution of species of differing molecular weight that may be present in a degraded sample. To identify the different oligomeric fractions produced during polymer degradation, gel permeation chromatography (GPC), in combination with FTIR spectroscopy, and using a flow-cell technique is useful.[35] Thin flow cells with very small internal volume (10 μL), allowing better solvent transmission and suppressing peak broadening, can be used for detection. Here, the FTIR system transforms and displays spectra in real-time during the chromatographic procedure. Furthermore, absorbance subtraction techniques may be used to explicate the composition of chromatographically unresolved components.

1.3 Structural Order and Degradation

It is widely accepted that the degree of crystallinity of a polymer has a marked affect on its rate of oxidation.

Generally, reactivity between free radicals and oxygen is higher in regions of low crystallinity (amorphous regions) owing to the higher segmental mobility of chains, whereas in highly crystalline regions radical recombination is promoted.

However, great care should be taken when attempting to use IR spectroscopy to elucidate structural order in polymers. Maddams has stated that ''Peaks should not be designated as crystallinity bands unless they disappear on melting, are predicted by group theory and are shown to depend upon the existence of a crystal lattice''.[36] Even when bands may be attributed decidedly to crystallinity, the relative intensity of such bands may be due to orientation in addition to crystallinity. This is the case for PE, where the doublet at 720/730 cm^{-1} is a function of both crystallinity and orientation (Figure 13). In this sense IR spectroscopy is a poor substitute for X-ray diffraction and dielectric loss spectroscopy. Crystallinity is itself influenced by the degree of chain branching and by the length of the branches formed. Again, IR methods are of limited value in comparison with other techniques, notably nuclear magnetic resonance (NMR) spectroscopy, when it comes to assigning specific branches.[2] However, an IR scale of branching is still widely used by polymer manufacturers owing to the relative expense of high-resolution NMR methods.

Morphological changes in semicrystalline polymers, brought about by the application of mechanical stress, depend on their original morphology, in particular the degree of orientation. It is believed that any imposed stress reduces the probability of radical recombination at a site of bond rupture. This has been supported by a study of the rupture of stressed bonds as the first to be broken, as demonstrated by shifts in IR absorption bands to lower wavenumbers.[37–39]

Extensive studies have confirmed that orientation results in a reduction of segmental mobility and the sorption and diffusion of gases in the polymer matrix brought about by closer packing of the amorphous fraction. Figure 14 illustrates the use of IR spectroscopy to study how changes in the orientation of polymer chains – of varying draw ratio – influence the rate of oxidation, as evinced by carbonyl formation at 1715 cm^{-1}. This study revealed that for a given draw ratio, the oxidative degradation of high-density polyethylene (HDPE) was retarded with an increase in drawing temperature.[40]

If orientation of polymer chains has a bearing on oxidative stability then dichroic and trichroic measurements of orientation of polymers are likely to be of value in establishing which orientational isomers confer stability to a given polymer and, more specifically, in establishing relationships between orientation and mechanical properties.[2] Here the polymer axis is aligned parallel, perpendicular or through the film thickness to the spectrometer sampling plane.

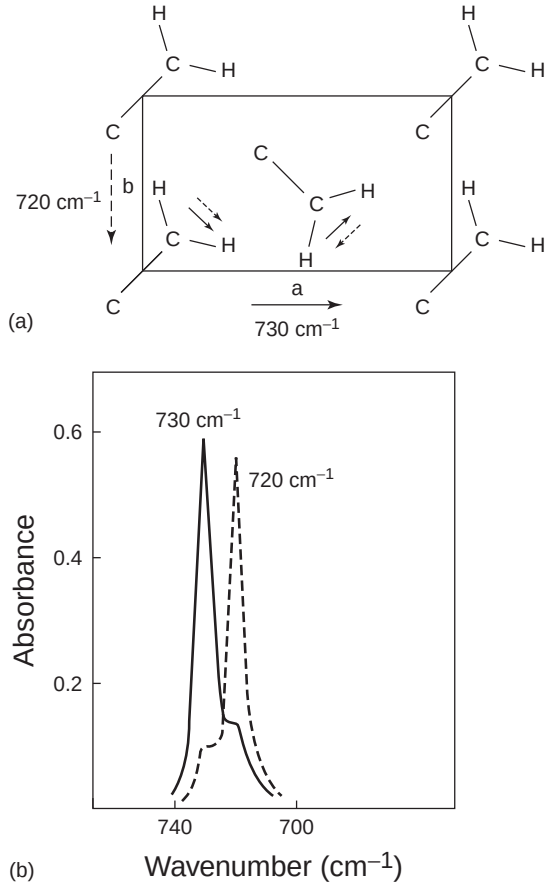

Figure 13 (a) Transition moment directions of the in-phase and out-of-phase CH_2 rocking vibration in PE. (b) The IR spectrum of PE. (Reproduced from Krimm, *Infrared Spectroscopy and Molecular Structure*, ed. M. Davies, 270, 1963, Elsevier, Amsterdam.)

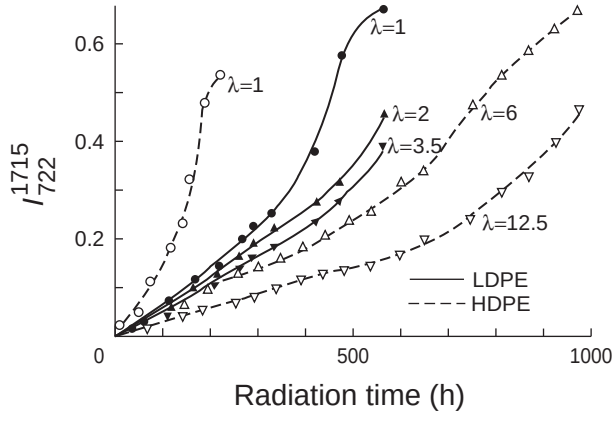

Figure 14 Variation of the carbonyl index (I_{722}^{1715}) for various draw ratios (λ) of PE. (Reproduced from Tincer et al., *Polym. Eng. Sci.*, **26**(7), 479–487, 1986.)

2 QUANTIFICATION OF FUNCTIONAL GROUPS AND THE KINETICS OF POLYMER OXIDATION

IR spectroscopy remains a fundamental technique for the study of the kinetics of polymer oxidation. A considerable number of studies have been undertaken on the kinetics of the thermal and photooxidation of polyolefins. The rates of formation of nonvolatile products under various experimental conditions may be derived from the formation of carbonyl groups as a function of temperature and oxygen concentration. The activation energies for the major oxidation stages of polymers may be estimated from various Arrhenius plots.

Stivala et al. have used IR spectroscopy to study the continuous oxidation of polyolefins.[41] To achieve this, a temperature-controlled cell was constructed to allow constant monitoring of a polymer film undergoing oxidation. The relative concentration of carbonylic products formed when the ratio of oxygen to nitrogen impinging on the polymer film was varied is given at different temperatures. Plots of area under the carbonyl band as a function of time at various temperatures were sigmoidal in shape, the maximum slope giving the rate of carbonyl formation. This is a manifestation of the autocatalytic nature of the oxidative process.

In order to correct for variations in sample thickness, it is usual to measure the intensity of a particular band relative to some other band, the intensity of which does not vary during the course of degradation. The resulting ratio is defined as a specific functional group index. The most widely employed of such indices is the so-called carbonyl index.[9] This index is used for the indirect determination of embrittlement time in polyolefins, which is taken as the time in hours to achieve a carbonyl index of 0.06. The carbonyl absorbance intensity is compensated for by the ratio of its value to film thickness or the intensity of an internal reference peak. In PE, the methylene wagging-mode centered at 1369 cm^{-1} may be used as the reference (Figure 15).

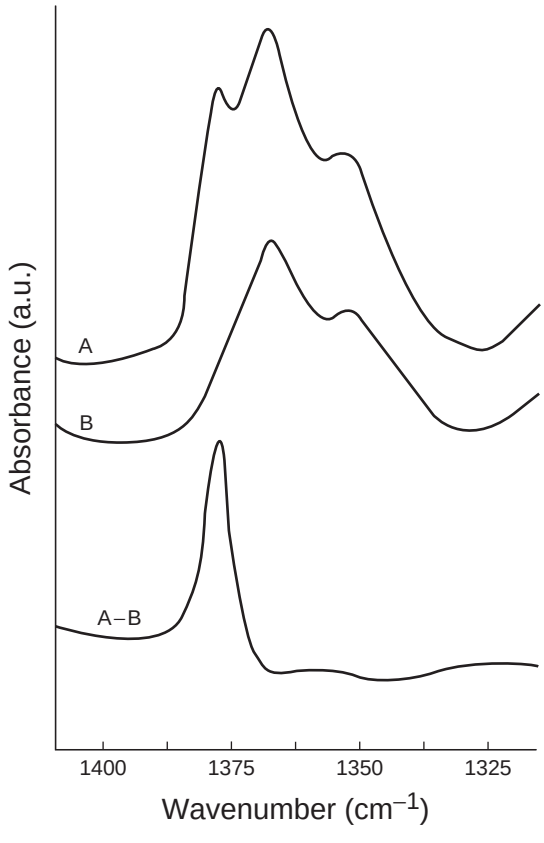

Figure 15 IR wagging bands in samples A and B, and their difference spectrum A – B. (Reproduced from Rabek, *Photostabilization of Polymers: Principles and Applications*, 444, 1990, Elsevier Applied Science.)

3 ROLE OF ANTIOXIDANTS IN POLYMER DEGRADATION

In general, antioxidants are added to polymers to inhibit the radical reactions taking place.[5–9] They may be classified as radical scavengers, peroxide decomposers, metal deactivators, color inhibitors, acid scavengers and ultraviolet absorbers. These compounds are usually added during processing at levels 0.05–0.2% w/w of the polymer. During aging, these additives may experience either facile loss from the polymer or be transformed by reaction with the degrading polymer matrix.

IR spectroscopy is more specific than ultraviolet spectroscopy, but frequently additive levels are too low to give suitable spectra. Despite this it is one of the most commonly used direct methods for additive analysis in polymers. The mid-IR range (4000–400 cm^{-1}) has been most widely used, e.g. the carbonyl groups of phenolic antioxidants at 1740 cm^{-1} and the oxygen phosphorus–phenyl groups of phosphite antioxidants at 890 cm^{-1}. Figure 16 illustrates calibration curves for IR determination of the antioxidants Irganox 1010 and Irgafos 168.[42]

Alternatively, time-consuming chemical separation of multicomponent systems may be replaced by obtaining difference IR spectra. This applies to systems where one or more components occur in small concentrations. Once the spectrum of a standard ''pure'' polymer has been obtained, it can be used to remove the base (polymer) spectrum from that of the modified polymer sample.

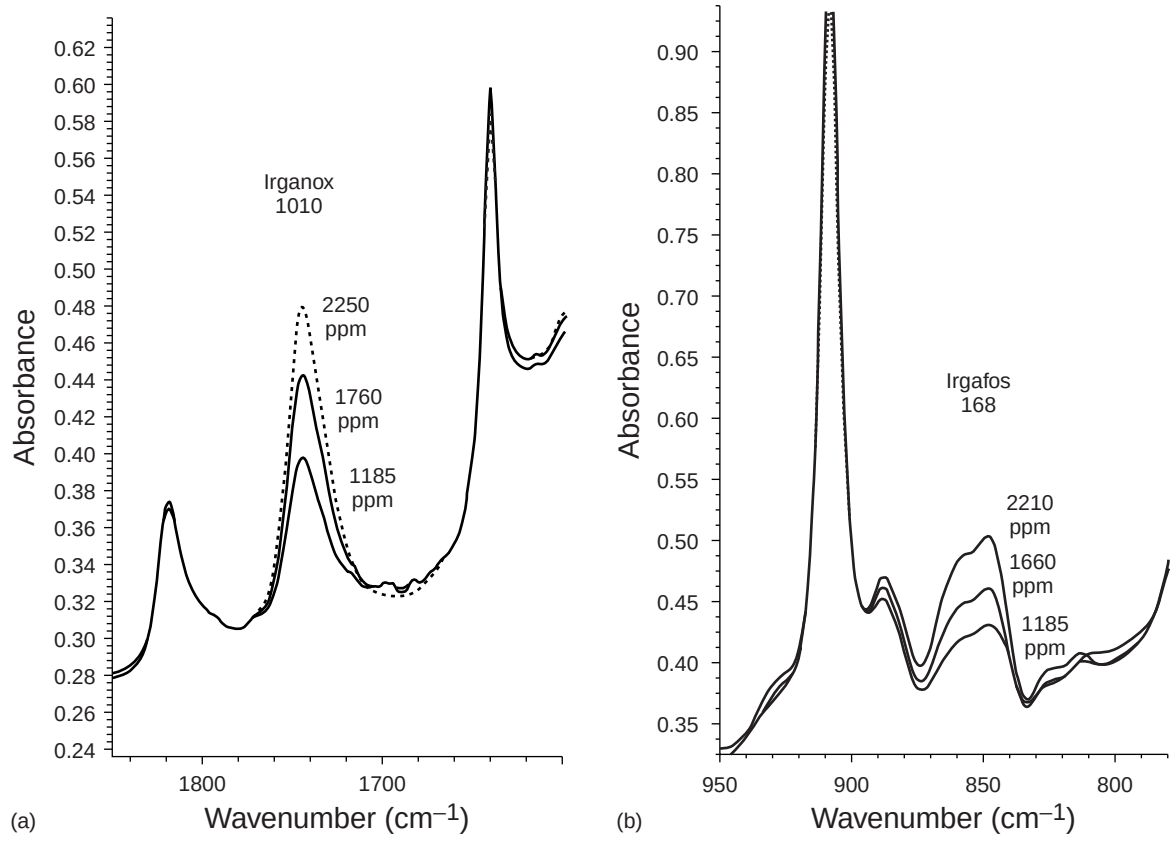

Figure 16 FTIR spectra of a set of PE calibration standards formulated with varying levels of Irganox 1010 and Irgafos 168. (Reproduced from Tikuisis, *Plastics Additives: an A–Z Reference*, 91, 1998, Chapman and Hall.)

Miller and Willis obtained IR spectra of antioxidants in polymer films by compensating with an additive-free polymer in the reference beam.[43]

Owing to interference effects by extraction solvents, other additives and adventitious impurities in the polymer, direct IR spectroscopy of polymer extracts has found very limited application in the determination of additives. However, when IR spectroscopy is preceded by a clean-up or separation technique, such as thin-layer chromatography, then it is an extremely useful method for both characterization and analysis. Spell and Eddy have described IR procedures for the determination of additives, at concentrations up to 500 ppm, in PE pellets following solvent extraction with carbon disulfide at room temperature for 2–3 h.[44]

Development of a calibration curve for the quantitative determination of an antioxidant by IR analysis can be difficult. Unlike chromatographic analysis, where calibration only requires the gravimetric dissolution of the antioxidant in an appropriate solvent, IR analysis requires the antioxidant to be dispersed homogeneously throughout the polymer matrix at the correct concentration.

Near-infrared (NIR) analysis is able to circumvent problems of additive dispersion. NIR has lower absorptivity than conventional mid-IR radiation, which enables it to penetrate more deeply into a polymer sample. Because of this, slightly inhomogeneous additive distributions are less problematic. Furthermore, because glass does not act as a cut-off filter for NIR, fiber optics may be employed to probe antioxidants directly in a polymer melt stream during processing.[45]

Rather than studying the antioxidants themselves, an alternative approach is to study changes to the mechanism or rate of oxidation of the polymer itself in the presence or absence of an antioxidant. The reflectance IR spectroscopic study by Chan and Allara, of the oxidation of PE coated on to a copper substrate in the presence and absence of the metal deactivator *N*,*N*-diphenyloxamide, illustrates this approach.[46]

4 TECHNIQUES

4.1 Pyrolysis Gas Chromatography and Infrared Spectroscopy

The IR spectra of gas chromatographic effluents may be taken from gaseous samples or from liquid samples by

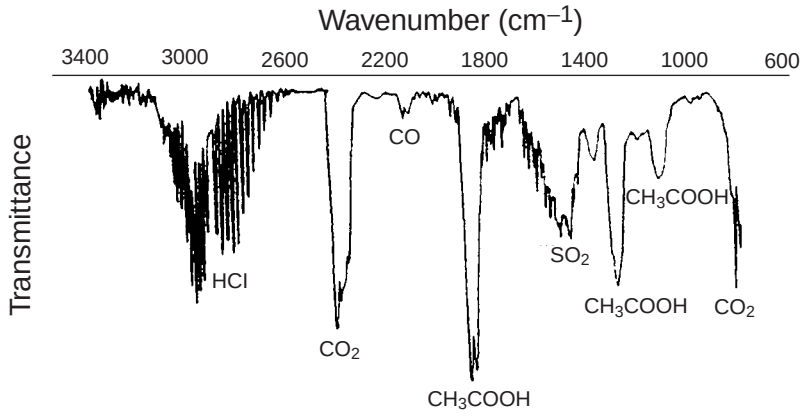

Figure 17 FTIR spectrum of gaseous combustion products of a PVC composite at 648 K in air. (Reproduced from Liebman et al., *Appl. Spectrosc.*, **30**, 355, 1976.)

trapping techniques. Because analysis in the gas stream eliminates the difficulties associated with condensing vapor-phase degradation products, it is considered more desirable. Furthermore, this latter technique holds most value when monitoring the toxic gases associated with the degradation of polymers caused by combustion (Figure 17).[47] Here, fast-scanning FTIR spectrometers may be interfaced to a flow-through heated light-pipe gas cell to allow the continuous measurement of any volatile effluents.

The difficulties associated with the identification of intractable, cross-linked material produced during degradation may be overcome by spectroscopic examination of liquid and gaseous pyrolysis products along with any involatile residues. The IR spectra of gas chromatographic effluents may be taken either from liquid (trapping technique) or gaseous samples. Analysis in the gas phase is desirable because this approach eliminates the difficult process of preferentially condensing the sample from a vapor stream. The advantage of this approach is the selectivity of IR spectroscopy when combined with the separation capacity of gas chromatography. However, when the highly cross-linked material is a vulcanizate, pyrolysis IR methods have limited value.[48] A more useful approach is thermal volatilization analysis, which, being nondestructive of the various product fractions that it produces, lends itself readily to further IR analysis.[49–55]

4.2 Attenuated Total Reflectance

Whether the degradation is typically thermal or thermo-oxidative depends on the relative rates of the two reactions given in Scheme 2. Under processing conditions, because access to air is restricted and at high temperatures oxygen solubility is low, higher concentrations of macroalkyl radicals are likely to be present and reactions such as depropagation (of monomer), radical chain transfer and chain scission dominate. Normally a steady state is rapidly established, but there are two main exceptions to this. The first takes place in very thin films where the rate of diffusion of volatile products results in a nonsteady-state process, preventing side reactions between the volatile products and chain radicals. The second occurs during the extrusion of a hot polymer into air, where oxygen ingress is diffusion-controlled, giving rise to oxidation reactions in a surface layer only a few microns thick. The depth of oxidation in a thick-section material is an important parameter, influencing its mechanical properties. In this respect the ability of IR spectroscopic techniques, such as attenuated total reflectance (ATR) and microtoming, in profiling the extent of degradation is of considerable industrial importance.

Stivala et al. found that the rate of oxidation in atactic PP was constant for films of thickness up to 8.6 μm but decreased for films of 19 μm. These authors found a linear relationship between the band at 1380 cm^{-1} and film thickness. Hence, this band may be used as an internal sample for PP films of varying thickness.[56]

Information on the distribution of chemical structure of polymers in the depth direction may be achieved in several ways.[57,58] Samples may be cut obliquely with a microtome into films of typical thickness 1–2 μm. This increases the apparent depth, i.e. when the cutting angle is 1° the length of the cutting surface is 50 times the corresponding depth (Figure 18). The thin films produced may then be analysed in the depth direction by IR microspectroscopy operating in transmittance mode.

For opaque samples, or when samples are difficult to cut into thin films, perhaps due to embrittlement during degradation, ATR may be employed. Adaptations of the basic ATR method are particularly useful when studying the degradation of highly filled polymers, where the process of multiply-disturbed total internal reflection

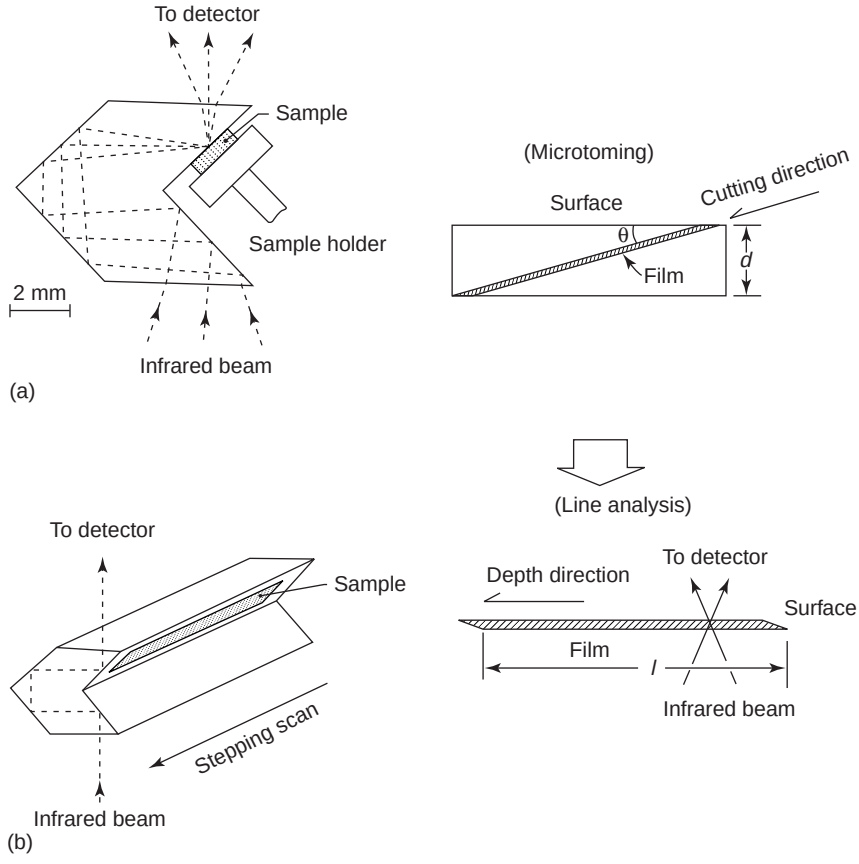

Figure 18 Schematics of the principle of depth analysis and the prism developed for ATR IR microspectroscopy. (Reproduced from Murase et al., *Polym. Degrad. Stab.*, **43**, 416, 1994, Elsevier Science.)

is of value.[59] Here the sample is placed in intimate contact with a crystal of high refractive index and low IR absorption. The IR beam undergoes a series of multiple reflections, which are a function of the length of the crystal. The main advantage of this method is that not only does it permit the analysis of opaque samples but it also allows the signal-to-noise ratio to be optimized. In general, by correct choice of the reflection device, and hence the angle of incidence of IR radiation impinging on the sample, it is possible to record ATR spectra to various depths of penetration of the IR beam into the surface of the sample. Hence, by using this technique a limited ''concentration profile'' of specific groups associated with the polymer or polymer additives may be assimilated.

4.3 Diffuse Reflectance Infrared Fourier Transform Spectroscopy

Recent developments in the diffuse reflectance infrared fourier transform (DRIFT) technique have further enabled the depth profiling of surfaces and in addition the study of interfaces.[60–62] Here, the uppermost surface of the sample is covered with a layer of powdered KBr, such that as the thickness of this powdered layer is increased, the ratio of surface-to-bulk contribution to the spectrum increases.

4.4 Photoacoustic Spectroscopy

This technique has the advantage in that it allows IR spectra to be obtained for samples that are difficult to prepare as transparent films, where the sample has high internal light scattering and for samples that are coated on to opaque or strongly light-scattering substrates. Figure 19 illustrates photoacoustic spectroscopy (PAS) spectra of Kevlar fibres before and after removal of the outer layer by oxidation.[63]

4.5 Fourier Transform Infrared Emission

IR emission spectroscopy enables easy acquisition of spectroscopic information from samples undergoing degradation at elevated temperatures, by detecting the time-dependent IR emission originating from polymers as they deteriorate. DeBlase and Compton provide a detailed review of emission spectroscopy.[64] These

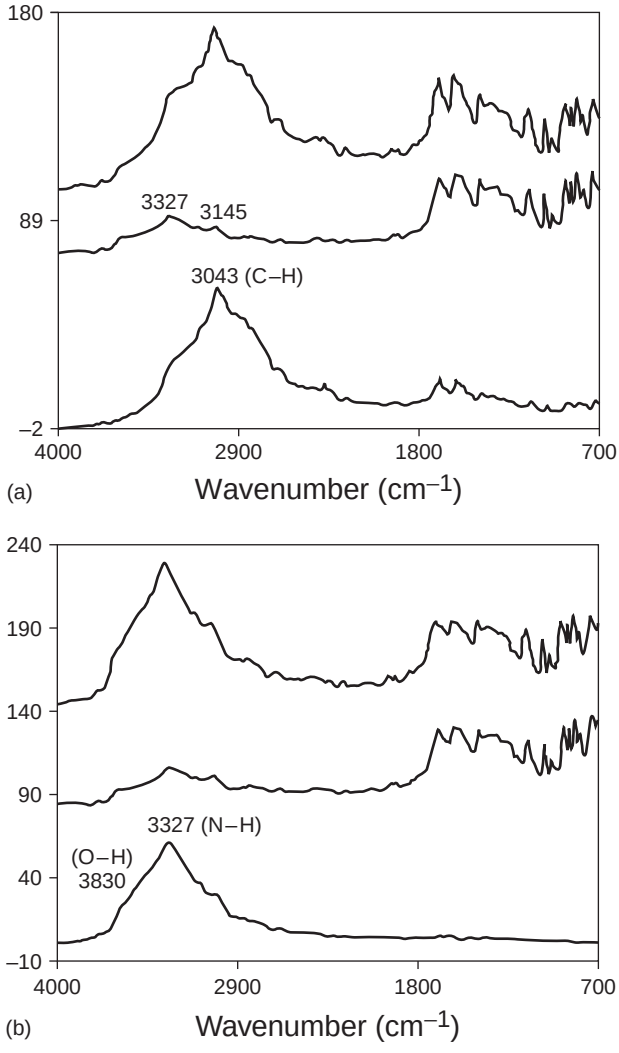

Figure 19 Photoacoustic spectra pertaining to the relative emissivity of Kevlar-49 fibres before (a) and after (b) oxidation. (Reproduced from Koenig et al., *Makromol. Chem.–Makromol. Symp.*, **5**, 99, 1986, Hüthig & Wepf Verlag.)

authors describe a process whereby IR emissions from a moving sample may be collected when the surface of the sample has been quickly heated or cooled by a jet of hot or cold gas. This process induces a thermal gradient across a thin film of sample and, as such, has distinct advantages for industrial applications. George et al. have investigated the thermal oxidation of polyolefins by monitoring carbonyl band growth as a function of degradation time.[65,66] Using this method it was possible to infer the heterogeneous nature of the degradation process. Celina et al. have recently applied FTIR emission spectroscopy to the study of in situ polymer degradation.[67,68] In Figure 20(a) and (b) the emittance and changes therein are depicted for polystyrene oxidized at 220 °C.

It should be noted that for quantitative measurements careful referencing of spectra to an instrumental background is necessary to obtain band-specific relative emission intensities. For bands with high emittances, quantitative measurements are difficult due to the uncertainty in determining values close to band saturation. In addition, this technique is not suited to carbon-black-filled materials, because featureless blackbody emission is observed, although other highly filled systems may be studied.

5 ADVANTAGES AND LIMITATIONS OF INFRARED METHODS

By reference to the various examples discussed above it is obvious that IR spectroscopy is widely used in the study of polymer degradation. One of its main advantages in direct analysis is its speed and convenience. Ironically, it is limited in the study of polymer degradation because it is so very matrix-dependent. The complexity of the polymer degradation mechanism(s) results in a large number of functional group changes that lead to the broadening of IR bands. Fortunately, this lack of selectivity during direct analysis has been largely circumvented. Over the last few decades the development of chemical derivatization methods has enabled overlapping absorptions to be resolved. Furthermore, the use of attachment devices has allowed easier handling of friable samples and the depth profiling of samples, and computational methods such as spectral subtraction have led to enhanced sensitivity. Despite this evolution in methodology, IR spectroscopy remains limited in some respects. In the evaluation of microstructural changes it cannot effectively compete with NMR or Raman spectroscopy, nor is it able to define adequately the morphological changes, as X-ray diffraction can.

Overall, the interfacing of FTIR with other instruments offers the best way forward to the study of polymer degradation. A pertinent example to illustrate this is vapor-phase IR spectroscopy employing fast-scanning techniques, as used by Kiran and Gillham.[69,70] The system, composed of a programmable pyrolyzer, thermal chromatogram, mass chromatogram and fast-scan vapor-phase IR spectrometer coupled to a computer, is able to provide a profile of the thermal history of the sample prior to and during pyrolysis. The integrated system is able to determine the relative amounts of volatile and nonvolatile degradation products from the resulting chromatographic retention times, IR spectra and mass numbers.

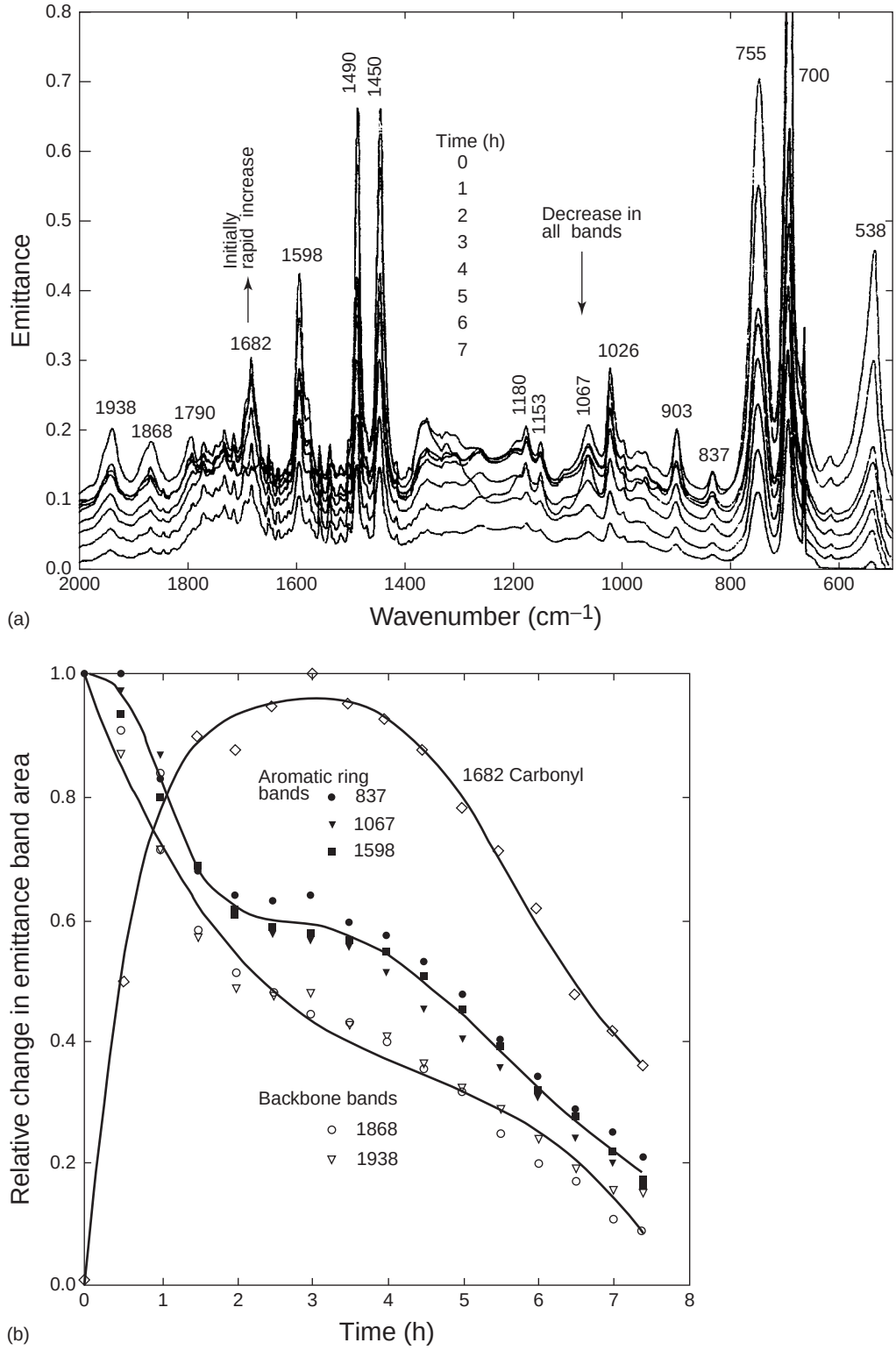

Figure 20 FTIR emission spectra of a polystyrene sample at 220 °C (a) and the relative quantitative changes in functional groups (b). (Reproduced from Celina et al., *Polym. Degrad. Stab.*, **58**, 30, 1997, Elsevier Science.)

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflectance
DRIFT	Diffuse Reflectance Infrared Fourier Transform
FTIR	Fourier Transform Infrared
GPC	Gel Permeation Chromatography
HDPE	High-density Polyethylene
IR	Infrared
LDPE	Low-density Polyethylene
LLDPE	Linear Low-density Polyethylene
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
PAS	Photoacoustic Spectroscopy
PE	Polyethylene
PP	Polypropylene
PVC	Poly(vinyl chloride)

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Near-infrared Spectroscopy of Polymers and Rubbers

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Gas Chromatography/Infrared Spectroscopy • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. S.L. Hsu, 'IR Spectroscopy', in *Comprehensive Polymer Science: The Synthesis, Characterization, Reactions and Applications of Polymers*, eds. G. Allen, J.C. Bevington, Pergamon, New York, 429–468, 1989.
2. J.L. Koenig, *Spectroscopy of Polymers*, American Chemical Society, Washington, DC, 19–112, 1992.
3. T.R. Crompton, *Analysis of Polymers: An Introduction*, Pergamon Press, New York, 75–207, 1989.
4. W. Schnabel, *Polymer Degradation: Principles and Practical Applications*, Hanser, New York, 1981.
5. N.S. Allen, *Degradation and Stabilization of Polyolefins*, Elsevier Science Publishers, New York, 1983.
6. T. Kelen, *Polymer Degradation*, Van Nostrand Reinhold, New York, 1983.
7. G. Scott, *Mechanisms of Polymer Degradation and Stabilization*, Elsevier Applied Science, New York, 1990.
8. J.F. Rabek, *Photostabilization of Polymers: Principles and Applications*, Elsevier Applied Science, New York, 1990.
9. N.S. Allen, M. Edge, *Fundamentals of Polymer Degradation and Stabilization*, Elsevier Applied Science, New York, 1992.
10. *Polymer Degradation and Stability*, Elsevier Science Publishers, New York.
11. N.C. Billingham, P.D. Calvert, 'The Degradation and Stabilization of Polyolefins – An Introduction', in *Degradation and Stabilisation of Polyolefins*, ed. N.S. Allen, Elsevier Science Publishers, New York, 1983.
12. D.J. Carlsson, D.M. Wiles, *Macromolecules*, **2**, 597 (1969).
13. D.J. Carlsson, R. Brousseau, D.M. Wiles, *Polym. Degrad. Stab.*, **15**, 67 (1986).
14. B.B. Cooray, G. Scott, *Eur. Polym. J.*, **17**, 385 (1981).
15. D.J. Carlsson, R. Brousseau, C. Zhang, D.M. Wiles, *Polym. Degrad. Stab.*, **17**, 67 (1986).
16. D.J. Carlsson, R. Brousseau, C. Zhang, D.M. Wiles, *Polym. Degrad. Stab.*, **17**, 303 (1987).
17. K. Kato, *J. Appl. Polym. Sci.*, **18**, 2449 (1974).
18. Y. Kato, D.J. Carlsson, D.M. Wiles, *J. Appl. Polym. Sci.*, **13**, 1447 (1969).
19. Yu. A. Pentin, B.N. Tarasevich, B.S. El'tsefon, *Vestn. Mosk. Univ. Khim.*, **14**(1), 13 (1973).
20. R.M. Rugg, J.J. Smith, R.C. Bacon, *J. Polym. Sci.*, **13**, 535 (1973).
21. D.L. Tabb, J.J. Sevick, S.L. Koenig, *J. Polym. Sci. Polym. Phys. Ed.*, **13**(4), 815 (1975).
22. J.H. Adams, *J. Polym. Sci. Polym. Chem. Ed.*, **8**, 1279 (1970).
23. D. Briggs, C.R. Kendall, *Int. J. Adhes.*, **199**, 13 (1982).
24. J. Lacoste, D.J. Carlsson, S. Falicki, D.M. Wiles, *Polym. Degrad. Stab.*, **34**, 309–323 (1991).
25. N.J. Harrick, *Appl. Spectrosc.*, **31**, 548 (1977).
26. J.F. Heacock, *J. Appl. Polym. Sci.*, **7**, 2319 (1963).
27. N.S. Allen, J.F. McKellar, *Chem. Soc. Rev.*, **4**, 533 (1975).
28. D.R. Rueda, F.J. Balta-Calleja, A. Hidalgo, *Spectrochim. Acta*, **30A**, 1545 (1974).
29. M.U. Amin, G. Scott, L.M.K. Tillakeratne, *Eur. Polym. J.*, **11**, 85 (1975).
30. S. Gaumet, J.L. Gardette, *Polym. Degrad. Stab.*, **34**, 135–167 (1991).

31. F.D. Brako, A.S. Wexler, *Anal. Chem.*, **35**, 1944 (1963).
32. M.J. Barral, *Polym. Sci.*, **13**, 1515 (1975).
33. A. Dankovics, Munat Gumi, **11**(12), 380 (1974); *Chem. Abstr.*, **82**, 86773g (1975).
34. P.G. Kosky, R.S. McDonald, E.A. Guggenheim, *Polym. Eng. Sci.*, **25**, 389 (1985).
35. D.W. Vidrine, D.R. Mattson, *Appl. Spectrosc.*, **32**, 502 (1978).
36. W.F. Maddams, 'Infrared and Raman Spectroscopy', in *Analysis of Polymer Systems*, eds. L.S. Bark, N.S. Allen, Applied Science Publishers, London, 70, 1982.
37. V.R. Regel, A.I. Slutsker, E.E. Tomashevskii, *Sov. Phys. Usp.*, **15**, 45 (1972).
38. V.R. Regel, *Rheol. Acta*, **14**, 19 (1975).
39. R.P. Wool, *J. Polym. Sci. Polym. Phys. Ed.*, **19**, 449 (1981).
40. T. Tincer, I. Cimen, G. Akay, *Polym. Eng. Sci.*, **26**(7), 479–487 (1986).
41. S.S. Stivala, L. Reich, P.G. Kellehmery, *Makromol. Chem.*, **59**, 28 (1963).
42. T. Tikuisis, 'Antioxidants: Their Analysis in Plastics', in *Plastics Additives: An A–Z Reference*, ed. G. Pritchard, Chapman and Hall, London, 88–92, 1998.
43. I.G.J. Miller, H.A. Willis, *Spectrochim. Acta*, **14**, 119 (1959).
44. H.L. Spell, R.D. Eddy, *Anal. Chem.*, **32**, 1811 (1960).
45. R. Spatafore, L. McDermott, 'Near-IR Reflectance Analysis Quantifies Polyolefin Additives', *Plast. Compound.*, **Sept./Oct.**, 68–71 (1991).
46. M.G. Chan, D.I. Allara, *Polym. Eng. Sci.*, **14**(1), 12 (1974).
47. T. Morimoto, K. Takeyama, F. Konishi, *J. Appl. Polym. Sci.*, **20**, 1967 (1976).
48. M. Lerner, R.C. Gilbert, *Anal. Chem.*, **36**, 1382 (1968).
49. D.A. McKillop, *Anal. Chem.*, **40**, 607 (1968).
50. I.C. McNeill, *J. Polym. Sci. Part A-1*, **4**, 2479 (1966).
51. I.C. McNeill, *Eur. Polym. J.*, **3**, 409 (1967).
52. D.L. Gardner, I.C. McNeill, *J. Thermal Anal.*, **1**, 389 (1969).
53. I.C. McNeill, *Eur. Polym. J.*, **6**, 373 (1970).
54. I.C. McNeill, L. Ackerman, S.N. Gupta, M. Zulfiqar, S. Zulfiqar, *J. Polym. Sci., Polym. Chem. Ed.*, **15**, 2381 (1977).
55. I.C. McNeill, in *Developments in Polymer Degradation*, ed. N. Grassie, Applied Science, London, 1, 43, 1977.
56. A.R. Jadrnicek, S.S. Stivala, L. Reich, *J. Appl. Polym. Sci.*, **14**, 2537 (1970).
57. A. Murase, Y. Esaki, M. Sugiura, T. Araga, *Anal. Sci.*, **7**, 415 (1991).
58. A. Murase, M. Sugiura, T. Araga, *Polym. Degrad. Stab.*, **43**, 415–422 (1994).
59. M.T. Bryk, *Degradation of Filled Polymers: High Temperature and Thermal-oxidative Processes*, Ellis Horwood, Chichester, 1991.
60. M.T. McKenzie, S.R. Culler, J.L. Koenig, in *Fourier Transform Infrared Characterization of Polymers*, ed. H. Ishida, Plenum, New York, 1987.
61. M.T. McKenzie, J.L. Koenig, *Appl. Spectrosc.*, **39**, 408 (1985).
62. S.R. Culler, M.T. McKenzie, L.J. Fina, H. Ishida, J.L. Koenig, *Appl. Spectrosc.*, **38**, 791 (1984).
63. E.G. Chatzi, M.W. Urban, J.L. Koenig, *Makromol. Chem.-Makromol. Symp.*, **5**, 99 (1986).
64. F.J. DeBlase, S. Compton, *Appl. Spectrosc.*, **454**, 611 (1991).
65. G.A. George, G.A. Cash, L. Rintoul, *Polym. Int.*, **41**, 169 (1996).
66. G.A. George, M. Celina, A.M. Vassallo, P.A. Cole-Clarke, *Polym. Degrad. Stab.*, **48**, 199 (1995).
67. M. Celina, K.T. Gillen, J. Wise, R.L. Clough, *Radiat. Phys. Chem.*, **48**, 613 (1996).
68. M. Celina, D.K. Ottesen, K.T. Gillen, R.L. Clough, *Polym. Degrad. Stab.*, **58**, 15 (1997).
69. E. Kiran, J.K. Gillham, *J. Macromol. Sci. Chem.*, **8**, 211 (1974).
70. E. Kiran, J.K. Gillham, *Polym. Eng. Sci.*, **19**, 699 (1979).

Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships

Brian D. Pennington
Read-Rite Corporation, Milpitas, CA, USA

1 Introduction 1
2 History 2
3 Fourier Transform Sampling Methods 3
3.1 Sample Preparation 3
3.2 Transmission Fourier Transform Infrared Spectroscopy 3
3.3 Internal Reflectance Fourier Transform Infrared Spectroscopy 4
3.4 Reflection–Absorbance Fourier Transform Infrared Spectroscopy 5
3.5 Diffuse Reflectance Fourier Transform Infrared Spectroscopy 6
3.6 Photoacoustic Fourier Transform Infrared Spectroscopy 6
3.7 Step-scan Photoacoustic Fourier Transform Infrared Spectroscopy 7
3.8 Two-dimensional Fourier Transform Infrared Spectroscopy 8
4 Polymer Blends 8
4.1 Phase Structure Determination 8
4.2 Intermolecular Interactions 9
5 Polymer Adhesion 10
5.1 Coupling Agents 11
5.2 Composite Interfaces 12
5.3 Polymer–Polymer Interdiffusion and Adsorption 12
5.4 Film Formation 12
6 Polymer Surface Characterization 13
6.1 Plasma Methods 13
6.2 Chemical Methods 14
7 Polymer Bulk Characterization 14
7.1 Thermoset Curing 15
7.2 Film Formation and Stability 17
7.3 Orientation Measurements 18
Acknowledgments 19
Abbreviations and Acronyms 19
Related Articles 20
References 20

Infrared (IR) spectroscopy is the measurement of the intensity of IR radiation absorbed at specific wavelengths, which can be used for the characterization of solid, liquid, gas, fiber, and powder samples. This technique allows molecular level identification of surface, interfacial, and bulk regions owing to its high selectivity and sensitivity. Quantitative measurement of molecular functionalities is accomplished by comparison with standards of known concentrations. The IR spectroscopic methods covered in this article include transmission, internal reflectance, reflection–absorption, diffuse reflectance, photoacoustic (PA), and step-scan techniques. Common applications of IR spectroscopy include polymer blending, adhesion, surface analysis, kinetics, and multidimensional studies.

The molecular specific information IR spectroscopy provides is one of the most useful aspects of this technique when compared to other common analytical methods. Compared with gas chromatography/mass spectrometry (GC/MS) which reaches parts per billion detection limits, IR spectroscopy, which is much less sensitive, provides nondestructive analysis. The depth resolution of X-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES) is several orders of magnitude better than IR spectroscopy, yet the ultrahigh vacuum (UHV) operating conditions of these techniques would destroy volatile samples. While the lateral resolution of scanning electron microscopy (SEM), energy-dispersive X-ray analysis (EDX), and transmission electron microscopy (TEM) techniques is superior to IR spectroscopy, these methods require tedious sample preparation. IR spectroscopy offers sensitivity to a wide range of samples with limited sample preparation needed.

1 INTRODUCTION

Vibrational spectroscopy is a powerful analytical tool for identification and characterization of organic molecules. IR spectroscopy measures the amount of radiation absorbed at specific wavelengths corresponding to distinct vibrational motions of molecular bonds. By analyzing the IR radiation absorbed or transmitted by a sample, the characteristic functional group vibrations can be deduced, and the molecular composition can be identified from the resulting IR spectrum. This ''fingerprint'' spectrum provides unique information about molecular structure and chemical bonding. There are numerous publications on vibrational spectroscopy which include IR

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

theory, instrumentation, data interpretation, and polymer characterization.[1–8]

The energy levels in IR spectroscopy are commonly represented as wavenumbers ($\bar{\nu} = 1/\lambda$), expressed in units of cm^{-1}. The IR region is divided into three sections: the near-IR ($\bar{\nu} = 12\,820–3333\,cm^{-1}$), the mid-IR ($\bar{\nu} = 3333–333\,cm^{-1}$), and the far-IR ($\bar{\nu} = 333–33\,cm^{-1}$) regions. The near-IR region involves overtone and combination bands of the mid-IR fundamental vibrations, which are often overlapping, making interpretation difficult. The far-IR region is difficult to study owing to energy limitations of Nernst glower sources, beamsplitters, and detectors, as well as background gas absorptions. The majority of the useful molecular information in IR spectroscopy is contained in the mid-IR region, although the near- and far-IR regions have found useful applications as a result of new developments in chemometric analysis.

IR spectroscopy is straightforward in terms of experimental design, instrumental maintenance, and operating expenses. Owing to high selectivity and sensitivity at the molecular level, IR spectroscopy has become a very capable analytical technique for understanding polymeric structure.[9–11] IR analysis produces molecular-specific identification of a wide range of samples, including solids, liquids, gases, fibers, and powders, at a minimum measurable absorbance of 0.001 AU (absorbance unit).[12] With proper sampling techniques, the surface, interfacial, and bulk regions of a sample can be studied.

The main limitation of IR spectroscopy is quantitative analysis. Absolute measurement of specific functionalities requires extensive sample preparation and is often prone to artifacts due to optical effects.[13] This can be overcome by comparison with standards of known concentrations, but sample preparation is crucial to success analytical analysis.

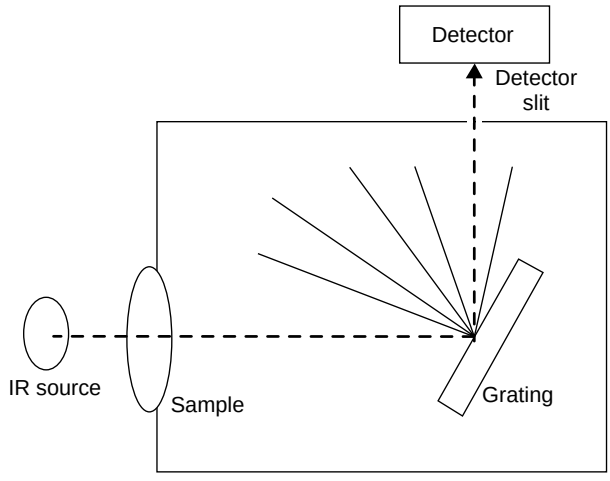

Figure 1 Dispersive IR spectrometer.

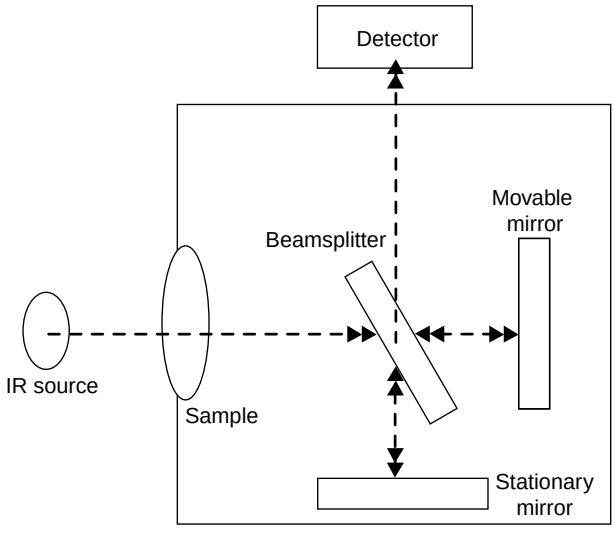

Figure 2 Michelson interferometer.

2 HISTORY

IR spectroscopy has been used in polymer characterization for many years. In the simplest form, IR spectroscopy is based on transmitting IR radiation through a sample and measuring the absorption. This was initially accomplished with a dispersive instrument using a prism or grating to scatter the IR radiation, as shown in Figure 1. By using a calibrated scanning mechanism, specific frequencies of interest can be isolated and the dispersed radiation recorded. From this method, the energy transmitted through a sample as a function of frequency can be obtained.

Dispersive IR analysis was replaced by Fourier transform infrared (FTIR) spectroscopy, where IR energy is collected in the time domain and Fourier transformed to the frequency domain. As shown in Figure 2, this interferometric method involves dividing the IR light into two separate paths by passing through a beamsplitter. By reflecting the IR light back along the same optical path, the energy is recombined in a constructive or destructive manner at the beamsplitter, forming the basis for interferometry. The detected energy as a function of optical path difference, also known as an interferogram, is Fourier transformed to the frequency domain. The resulting FTIR spectrum represents IR energy transmitted through a sample as a function of frequency.

The introduction of the Michelson interferometer, and subsequent FTIR instruments, vastly improved IR spectroscopy owing to several factors: the Fellgett (multiplex advantage), Jacquinot (throughput advantage), and Connes (frequency advantage).[14] The Fellgett

advantage arises from the fact that the FTIR spectrometer examines the entire spectrum in the same time period required for a dispersive instrument to examine a single spectral element. For similar measurements, the signal-to-noise ratio (S/N) of spectra from FTIR will be $n^{1/2}$ times greater than a grating instrument, where n is the number of resolution elements being examined during the measurement. This equates to a 60-fold increase in S/N for FTIR instruments. Spectroscopists have taken advantage of this factor in the study of polymer kinetics,[15] degradation,[16] and other time-dependent processes,[17,18] where S/N improvements are critical.

The Jacquinot advantage results from more effective throughput of the radiant power of the source with a Michelson interferometer, although the throughput for FTIR instruments is limited by the size of the mirrors.[19,20] Higher throughput is particularly important in trace analysis where the IR response is typically weak.

The Connes advantage arises from the fact that the frequencies are internally calibrated with an He–Ne laser for an FTIR instrument, whereas calibration of gratings in dispersive instruments is less reliable and has a tendency to drift.[21] Since co-adding of spectra is routinely done to improve S/N, frequency accuracy is an absolute requirement.

3 FOURIER TRANSFORM SAMPLING METHODS

Although many analytical techniques are suitable for surface analysis, there are inherent limitations to be aware of. The UHV operating conditions of secondary ion mass spectrometry (SIMS), XPS, and AES will destroy volatile samples unless special precautions are taken, such as a low-temperature sampling stage. Likewise, techniques such as SEM, EDX, and TEM require tedious sample preparation. FTIR offers surface sensitivity for a wide range of samples, without limitations on sample analysis. The various sampling methods commonly used in FTIR studies of polymers are shown in Figure 3. These spectroscopic methods include transmission, internal reflectance, reflection–absorption, diffuse reflectance, PA, and step-scan.[22–27] The sampling capabilities of these techniques are summarized in Table 1.

3.1 Sample Preparation

In order to analyze polymer specimens fully with FTIR, proper sample preparation is essential. Common preparative techniques have been reviewed by Haslam and Willis.[4] The preparation method is generally determined by the type of information required and the sample stability. A solid polymer sample can be prepared for transmission FTIR by mulling in liquid paraffin (Nujol), cast from the melt or a solution, or grinding with a powder and compressing into a disk. Disks are prepared with IR-transparent halide salts such as KBr, CsI, and NaCl, which will lose less spectroscopic information. Although these techniques yield films of high spectral quality, the process is destructive and the processing history of the sample is lost. An alternative method is a diamond cell where the solid sample is placed between two flat diamond surfaces.[28] A screw closes the diamond surfaces, compressing the sample to a suitable sampling thickness.[29] The diamond cell is used to study fiber and powder samples.[30]

3.2 Transmission Fourier Transform Infrared Spectroscopy

In spite of the development of reflective methods, transmission measurements remains the most commonly used technique, particularly for inorganic and organic substrates which can be prepared in a suitably transparent form. The transmission technique is based on passing IR energy of intensity I_0 directly through the sample where a portion of the energy is absorbed (Figure 3a). By detecting the remaining transmitted energy as I_T, a transmittance spectrum is obtained. Transmittance spectra are defined by $T = I_T/I_0$, where T is transmittance. Although transmission measurements are very accurate and reproducible for transparent materials, metals and other opaque specimens are troublesome to analyze. Koenig first recognized that transmission spectroscopy could be used to study polymer surface species.[31] By spectral subtraction of a bulk sample spectrum from a surface-treated sample of the same bulk composition, contributions of the bulk phase will be removed and the resultant difference spectrum will represent the IR response of the molecular structure of the surface. The difference spectrum can then be used to measure quantitatively the surface species.[31] For suitable samples, IR transmission measurements are superior to reflection techniques owing to high sensitivity, quantitative measurement, and relative simplicity.

While transmission measurements are commonly employed, poor sample preparation can lead to erroneous results. A typical problem in sample preparation is the adverse effect of surface smoothness, creating interference fringes. This problem can be avoided by mechanical roughening of the surface. Another common problem occurs when samples are too thick or concentrated for accurate FTIR analysis. Since an increase in sample thickness will produce a nonlinear relationship between the IR signal and pathlength, saturation behavior can be corrected by preparing a thinner sample or increasing the proportion of nonabsorbing material.

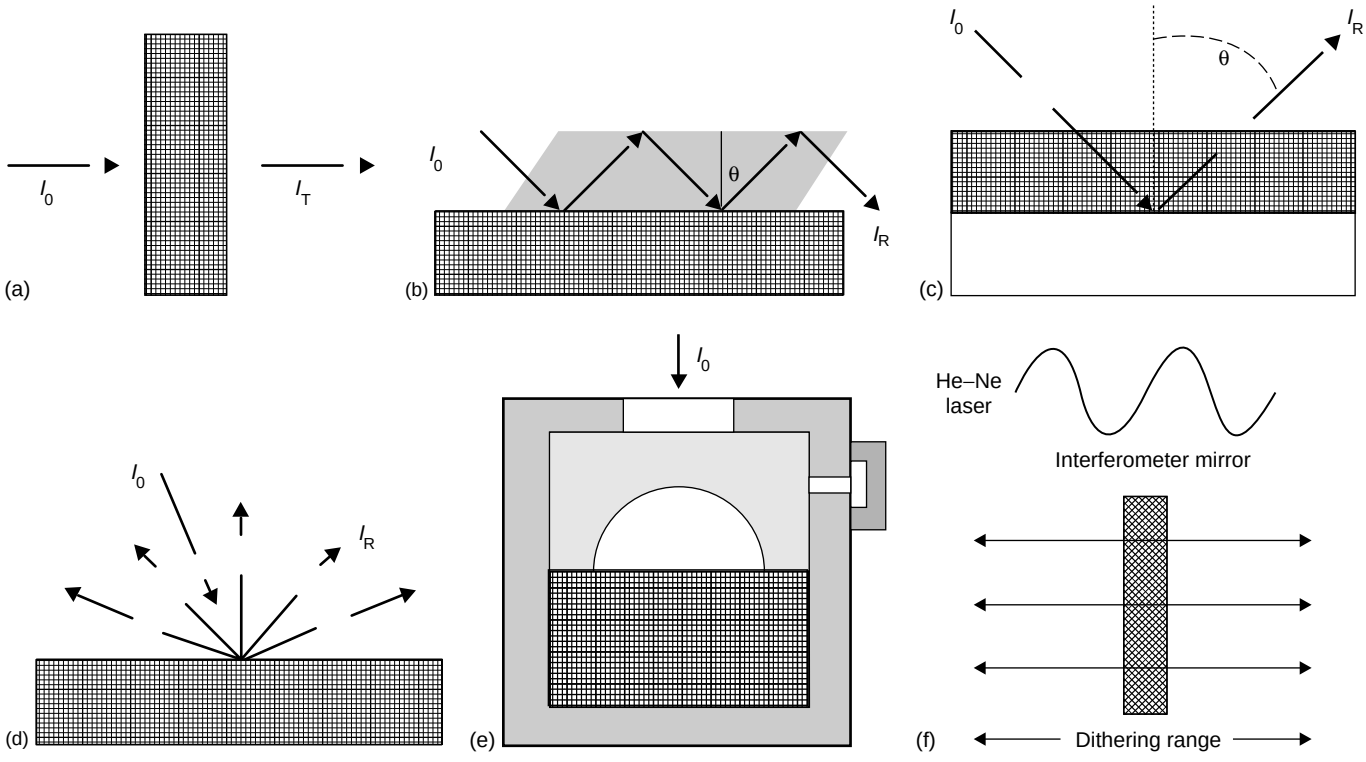

Figure 3 FTIR techniques: (a) transmission; (b) internal reflectance; (c) reflection–absorption; (d) diffuse reflectance; (e) PA; (f) step-scan.

Table 1 Sampling capabilities of FTIR techniques[12]

Technique	Sampling depth	Sample preparation	Limit of detection	Sample type
Transmission	N/A	Extensive	1×10^{-5} M	Solid, liquid, gas
IRS/ATR	40 Å–3 μm	Minimal		Film, powder, fiber
R–A	Submonolayer–2 μm	Moderate		Reflective substrate
DRIFT	Monolayer–2 μm	Minimal		Powder, fiber
PAS	200 Å–50 μm	Nil	1×10^{-7} M	All
IR microscopy	10 μm cross-section	Extensive		All

DRIFT, diffuse reflectance Fourier transform infrared; R–A, reflection–absorbance; PAS, photoacoustic Fourier transform infrared spectroscopy.

Liquid samples are analyzed with a sealed transparent sampling cell consisting of two windows with a spacer between them to define the pathlength. Several designs are available to accommodate various liquid properties, such as viscous or volatile samples. By increasing the area of the liquid sampling cell and using cell windows of ZnSe or ZnS, analysis throughput can be greatly improved.[32]

When only small amounts of sample are available, microsampling becomes necessary. Microsampling accessories include a broad range of techniques, including transmission and reflectance methods. Microsampling equipment should accommodate the IR beam geometry to ensure that the entire sample is analyzed, while maximizing the sample response. Beam condensers are used to focus the size of the FTIR spectrometer beam to facilitate analysis of microsamples larger than 1 mm. Sample analysis is extended further through the use of FTIR microscopy.

3.3 Internal Reflectance Fourier Transform Infrared Spectroscopy

For samples which are difficult to prepare as transparent films, or if a concentration profile is needed, reflectance techniques are probably more suitable than transmission. Internal reflectance spectroscopy (IRS) can be used to measure optical spectra of a sample in contact with an IR-transparent crystal of higher refractive index, as shown

in Figure 3(b). As light passes through the crystal, a standing wave is formed at the sample–crystal interface. If the sample absorbs a portion of this radiation, the propagating wave will be attenuated. The reflected IR energy (I_R) is ratioed against a background scan to generate the spectra of the sample surface in contact with the crystal. This process is described as attenuated total reflectance (ATR) spectroscopy, which is a contact sampling method used to obtain surface compositional information from polymeric samples.[33,34] In the IRS procedure, as depicted in Figure 4, the IR beam (I_0) is directed at an incident angle (θ) through an IRE of refractive index n_1 on to a sample of refractive index n_2.

The passage of IR light through the IRE depends on the critical angle (θ_c), which is defined as $\sin^{-1}(n_2/n_1)$. Owing to the different refractive indices of the IRE and sample, light reflects inside the crystal when $\theta > \theta_c$ and refracts out of the crystal when $\theta < \theta_c$. The internal reflection element and physical principles of IRS form the basis of numerous configurations from which a variety of sample types can be studied.[22,33]

One of the main advantages of IRS is that it is possible to conduct surface depth-profiling experiments by varying the angle of incidence or using IRE crystals with different refractive indices.[23,33,35] This is expressed by Equation (1):

$$d_p = \frac{\lambda}{2\pi n_1(\sin^2\theta - n_{21}^2)^{1/2}} \quad (1)$$

where d_p (μm) is the distance the evanescent wave extends into the surface and decays to 1/e of the original value, n_{21} is the ratio of the refractive index of the sample to the refractive index of the IRE, and λ (μm) is the wavelength of IR radiation. Typical values for d_p range from 0.2 to 4 μm. Using the surface depth profiling concept, IRS has been applied to analysis of polymer coatings, including melamine polyesters,[36] alkyds,[37] urethanes,[38–41] and latexes.[42–52]

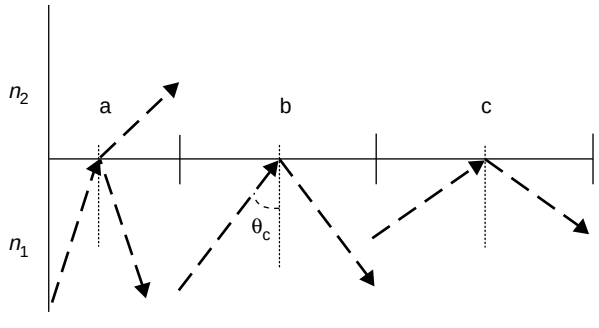

Figure 4 Schematic diagram of the effect of incidence angle on the light path through an IRE (internal reflectance element): (a) $\theta < \theta_c$ refracts; (b) $\theta = \theta_c$; (c) $\theta > \theta_c$ reflects.

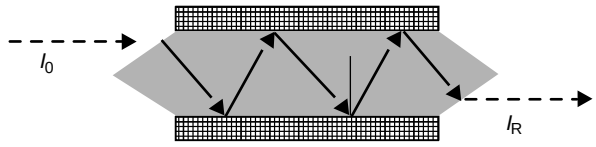

Figure 5 Schematic diagram of circular IRE configuration.

The surface sensitivity of IRS has proven to be useful for qualitative and quantitative understanding of polymer substrates.[53,54] For solid polymer films, the most common IRE is rectangular in shape. By applying even pressure to the sample film, intimate contact is made with the IRE and a significant improvement in the S/N of the analysis is seen. A circular crystal was designed for the study of liquid samples which are not readily analyzed with the rectangular IRE. A schematic of the circular cell is shown in Figure 5.

3.4 Reflection–Absorbance Fourier Transform Infrared Spectroscopy

As illustrated in Figure 3(c), R–A is based on the reflection of incident radiation at a specific incidence angle (θ).[55] When the surfaces are highly reflecting, R–A spectroscopy can be used successfully.[56] For optimum intensity of the reflection bands of films 0.5–20 μm thick, angles of incidence between 10 and 60° are used. Films in the nanometer range of thicknesses can best be measured with grazing angle R–A.

In R–A spectroscopy, there are two electromagnetic components present at the surface, the s-polarized component (perpendicular to the plane of incidence) and the p-polarized component (parallel to the plane of incidence). The p-polarized component has an enhanced electromagnetic field at the surface, while the s-polarized component's field approaches zero. For the surface analysis of thin films, only bonds with dipole components normal to the surface can absorb IR radiation. In order not to interfere with the incoming beam, angles of incidence near 80° are typically used.[57] At this angle, very thin films on the surface of a highly reflecting metal surface can be studied in the IR region, making grazing angle R–A spectroscopy a very important technique. Although R–A/FTIR spectroscopy is strongly surface sensitive, more than half of the energy is lost through the reflection optics for a highly reflective sample.

The use of polarized light in IR spectroscopy is an important tool in establishing structure–property relationships in polymer films. Polarization modulation of IR light greatly enhances the sensitivity and characterization capabilities of surfaces and thin films. This is the main premise of the polarization modulation infrared reflection–absorption spectroscopy

(PM/IRRAS) technique.[58] In PM/IRRAS, a photoelastic modulator is used to provide a differentially polarized IR light source which impinges on the sample surface in a near-grazing angle configuration. By employing a lock-in amplifier with a dual-channel input spectrometer, simultaneous collection of the differential polarization and reference signals is made. This technique greatly improves spectral quality, since polarization modulation effectively discerns near-surface absorptions from stray isotropic sample absorptions. Since the sample and reference signals are collected simultaneously, the interfering effect of CO_2 and H_2O absorbances is avoided. Compared with R–A spectroscopy, PM/IRRAS gives a much higher surface sensitivity with added in situ sampling capabilities.[59]

3.5 Diffuse Reflectance Fourier Transform Infrared Spectroscopy

DRIFT spectroscopy differs from other reflective techniques in that the reflectance collected from the sample surface is diffuse, not specular. Samples which strongly scatter and weakly absorb, such as powders or fibers, are ideal for DRIFT analysis.[60] As shown in Figure 3(d), the incident IR light (I_0) is diffusely reflected at all angles with an intensity I_R. Collection optics initially used an integrating sphere to capture the reflected light, but improved cells have been designed for FTIR instrumentation.[61,62]

Diffuse reflectance is defined for samples where the reflected light is isotropic. With a powder sample, both scattering and absorption occur, producing angularly distributed scattered radiation which may be anisotropic. This factor decreases for a larger number of particles such that the reflected light will be diffuse.[63] In order to relate signal to analyte concentration, the Kubelka–Munk theory must be applied. This relates the absorption coefficient (k) and the scattering coefficient (s) by Equation (2):[64–66]

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} \qquad (2)$$

where R_∞ is the absolute reflectance of an infinitely thick layer. In practice, the sample response is ratioed to a KBr standard.[62]

For powdered samples, diffuse reflectance is advantageous since the sample morphology remains intact. Specular reflectance remains a limitation of DRIFT, since strong absorptions lead to Reststrahlen bands in the reflectance spectrum, where the bands can appear inverted at their center. This effect makes quantitation of samples with strong absorptivity very difficult.

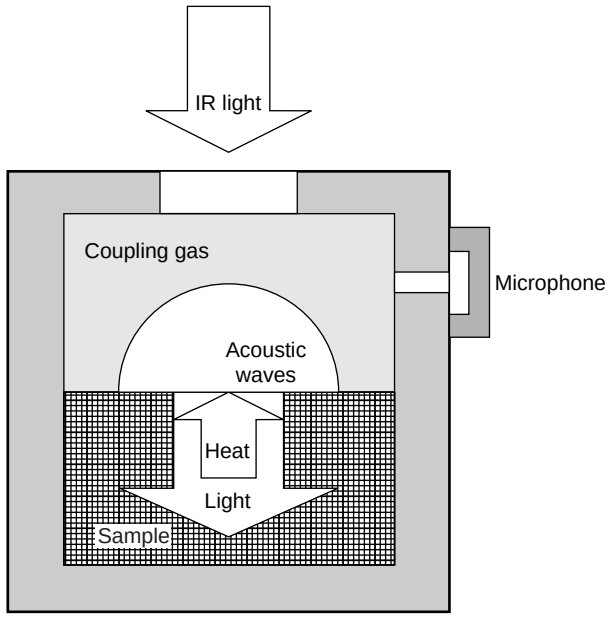

Figure 6 Diagram of PA cell and related processes.

3.6 Photoacoustic Fourier Transform Infrared Spectroscopy

Unlike reflective techniques, PA spectroscopy is based on the direct detection of the energy absorbed by the sample and does not require contact with an optical element (Figure 3e). As a result, this technique is suited to the analysis of opaque samples of essentially any shape or morphology.[24,67,68] As shown in Figure 6, the PA effect is a result of absorption of modulated IR light, producing corresponding heat fluctuations at the sample surface. The coupling gas phase above the sample surface will undergo subsequent pressure changes, producing a periodic acoustic signal. This acoustic response is measured with a sensitive microphone, and the electrical signal obtained is Fourier transformed, producing the PA spectra.

The basis of PA spectroscopy depends on the two-step process of radiation absorption and thermal diffusion. The amount of IR energy absorbed corresponds to the distribution of IR-active functional groups throughout the specimen. The process of energy conversion to heat is in the nanosecond range, so absorbed energy is released as heat which is transferred to the sample surface. The physics of this process has been mathematically detailed by Rosencwaig, in which the absorbed IR energy is represented by a thermal signal defined by periodic boundary conditions which decreases exponentially as it approaches the surface.[68]

Owing to the dissipative transfer of thermal energy through the solid and gas phases, the concept of thermal diffusion length (μ_{th}) is introduced.[69] Over the distance μ_{th}, the thermal signal is attenuated by e^{-1} of the

initial value, which is also the distance corresponding to 2π/modulation frequency ($2\pi/w$). Therefore, the PA spectra represents the IR functionalities from the surface to a depth of μ_{th} into the solid, with any response below μ_{th} being lost. This relationship is represented as Equation (3):

$$\mu_{th} = \left(\frac{2\alpha}{w}\right)^{1/2} \tag{3}$$

where α is the thermal diffusivity of the sample.[70–72]

Since the intensity of the optical or thermal signal decreases with distance, the photothermal response depends on sample thickness (l), optical absorption length (l_β), and μ_{th}.[69,73,74] These factors allow surface depth profiling experiments, since by altering the modulation frequency, μ_{th} will change, thus providing a different depth of penetration which may range from 0 to 150 μm. PA/FTIR depth profiling experiments are very useful in surface and interfacial polymer analysis,[72,75–77] polymer stratification,[78] time-dependent processes,[79–81] and polymer film formation.[38,39,82,83]

In the last 20 years, PAS has been based on the rapid-scan mode in which the interferometer mirror moves at a continuous and constant velocity to produce a change in retardation. The constantly changing retardation interferometrically modulates the intensity of each wavelength of the IR beam at a specific Fourier modulation frequency, w. By varying the interferometer mirror velocity (V), surface depth profiling is achieved by differences in the modulation frequency of the IR radiation.[72,84] This relationship is shown by Equation (4):

$$w = 2\bar{\nu}V \tag{4}$$

where $\bar{\nu}$ is the spectral wavenumber in cm^{-1}.[85]

One of the drawbacks of the rapid-scan mode is that the depth of penetration is dependent on wavenumber. Hence, low-wavenumber regions provide spectral information from greater depths than the high-wavenumber regions. To eliminate this dependence and obtain a constant penetration depth over the entire spectral range, step-scan photoacoustic (SSPA) FTIR spectroscopy was introduced.

3.7 Step-scan Photoacoustic Fourier Transform Infrared Spectroscopy

Although the majority of PA studies are conducted in the continuous, rapid-scan mode, the use of step-scan analysis overcomes the wavenumber dependence of the depth of penetration. In the step-scan mode (Figure 3f), the interferometer mirror moves incrementally, and the data are acquired at each retardation point, which eliminates the Fourier modulation frequency dependence. As a result, the SSPA/FTIR method allows measurements at constant sampling depths, because a single modulation frequency is applied over the entire spectral range. Furthermore, precise timing of the signal collection and mirror position permits collection of in-phase (I; in-phase with the incoming signal) and in-quadrature (Q; 90° out-of-phase with the incoming signal) components of the PA signal.

Methods of detection in the step-scan mode require a modulated signal to achieve optimal response; therefore, the IR light must be altered by either amplitude modulation (AM) or phase modulation (PM). In AM, the IR beam is chopped to produce a modulated PA signal.[70,86] AM techniques are seldom used, since S/N decreases by 50% as a result of the chopping device, producing a signal with a large dc component. In the PM mode, such as shown in Figure 7, modulation results from dithering the mirror at each retardation during data collection, which modulates the IR beam as a result of constructive and destructive interference.

The major advantages of PM over AM techniques in SSPA/FTIR spectroscopy is a higher source throughput

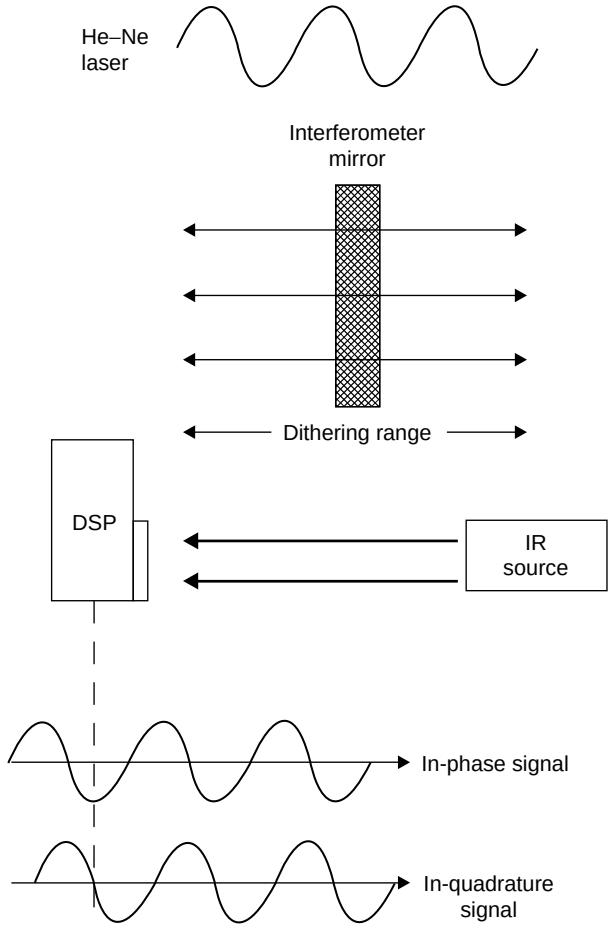

Figure 7 Schematic diagram of phase-modulation step-scan.

and fewer baseline drift problems since the interferogram is a first derivative in PM. PM step-scan spectroscopy provides a constant modulation frequency for all wavelengths and allows phase-sensitive detection, providing precise control of the probe depth and convenient extraction of the signal phase.[87,88]

Since PA measurements indirectly measure the time lag of the thermal signal to reach the surface, the introduction of phase analysis is useful in determining the depth of absorbing species in the specimen surface. Bertrand applied phase analysis to PA detection by the analysis of sorbed water at the surface of polyethylene (PE).[89] These studies showed that the depth resolution using phase analysis was much higher than could be obtained by changing the modulation frequency. By implementing the phase analysis of the PA signal, the distinction between the surface and bulk signals can be improved by two orders of magnitude. Improvements to the depth profiling resolution of $\leq 1\,\mu m$ are realized from utilizing the PA phase spectra, as shown by Palmer et al.[90,91]

The combination of modulation frequency and phase analysis can be utilized to distinguish multiple layers. Bertrand demonstrated that the optical absorption coefficient, β, of polymeric materials can be measured from phase analysis.[89] Application of basic phase relationships to PA analysis allows the isolation of the surface and bulk signals.

As shown by Poulet et al.,[92] the PA phase (Ψ) for a homogeneous, thermally thick sample can be represented as a function of β and μ_{th} by Equation (5):

$$\Psi = -\pi + \tan^{-1}(\beta\mu_{th} + 1) \tag{5}$$

Since μ_{th} is held constant in the step-scan mode, β can be abstracted.

3.8 Two-dimensional Fourier Transform Infrared Spectroscopy

The advent of two-dimensional (2D) IR correlation spectroscopy extended the application of FTIR spectroscopy into new areas.[27,93,94] In a typical 2D correlation experiment, an external perturbation is applied to the system while changes to the spectroscopic response are recorded as spectral fluctuations.[95] By applying correlation analysis to these signal fluctuations, a 2D plot can be generated which shows spectral features generally not observable in one-dimensional analysis. Using this approach, one can take advantage of the extreme sensitivity of vibrational bands to changes in the local environment of functional groups. As indicated in previous 2D/FTIR studies,[96–100] a wavenumber correlation map allows interpretation of synchronous changes between specific functional groups and improved spectral resolution. An example of a 2D synchronous correlation spectrum is illustrated in Figure 8. This contour plot shows the in-phase changes which occur between functional groups A and B.

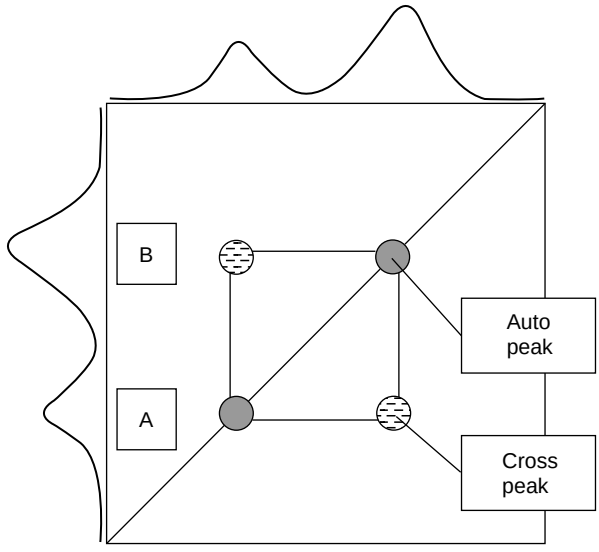

Figure 8 2D synchronous correlation spectrum derived from dynamic spectra.

Step-scan 2D/FTIR analysis can be applied to virtually any system in which spectroscopic changes occur as a result of some perturbation applied to the sample. By applying forces, such as pressure or tensile stress, fluctuations in the spectral intensity are induced in the specimen under analysis.

4 POLYMER BLENDS

Polymer blending has proven to be a practical method to improve the macroscopic properties of two otherwise nonideal materials. For example, polypropylene (PP), which has a poor low-temperature impact strength, can be improved by blending with elastomers, such as ethylene–propylene (EPR) copolymers or ethylene–propylene–diene (EPDM) terpolymers.[101,102] By polymer blending, changes are made to the polymer microstructure which ultimately controls macroscopic properties. FTIR has been found to be indispensable to the determination of phase structure in polymer blends, as well as associating intermolecular interactions with spectroscopic changes. These specific topics will be discussed in detail.

4.1 Phase Structure Determination

While polymer blends can be readily studied with such techniques as differential scanning calorimetry (DSC), wide-angle X-ray scattering (WAXS), dynamic mechanical thermal analysis (DMTA), TEM, or SEM, IR

spectroscopy enables us to elucidate structural information in a nondestructive manner at the molecular level. For an incompatible blend, phase separation will take place, leaving the two phases inhomogeneously distributed throughout the sample. The FTIR spectrum of incompatible polymers blends can be represented as the sum of the spectra of the two pure components, while compatible polymer systems can be found as a single phase. In this phase, molecular-level mixing of individual polymer chains takes place from chemical interactions between the two phases. As a result, considerable differences are seen in the IR spectra of polymers in a compatible blend and the pure component spectrum. The molecular-level interaction leads to spectroscopic differences which can easily be shown by subtracting out the pure component contribution, isolating the interaction spectrum. Analysis of a compatible blend was illustrated in the case of blending poly(phenylene oxide) (PPO) and polystyrene (PS). The spectral data were obtained from transmission FTIR of films cast from a chloroform solution on to KBr plates.[103] Substitution of 4-chloro(polystyrene) in place of PS induces the blend to become incompatible, allowing phase separation to be observed. Factor analysis of the FTIR spectra for the incompatible system revealed only homopolymer spectral components, whereas the compatible blend showed additional spectral responses from chemical interactions between the pure components. These additional spectral features can be used to estimate the relative interaction of PPO and PS blended at various ratios or when cured at higher temperatures. From least-squares analysis, the maximum interaction was seen at a 30 : 70 PPO to PS ratio. From these studies, the authors determined that the PPO/PS blend has a lower critical solution temperature of $>200\,^{\circ}C$.

A polymer blend which shows unique properties is poly(vinylidene fluoride) (PVDF) and atactic poly(methyl methacrylate) (PMMA). The exceptional compatibility of these two polymers is attributed to complex formation between polymer chains, even though PVDF is highly crystallizable. Coleman et al. reported on the spectral results and component interactions of blending these polymers.[104] The samples were prepared by dissolving the polymers in acetone and casting on KBr windows for transmission analysis. At 75 : 25 PVDF to PMMA ratio, the blend becomes incompatible, with PVDF forming a crystalline phase. The IR spectra can be closely approximated by adding the pure component spectra at the same weighting ratio. Subsequent analysis of a 39 : 61 PVDF to PMMA blend showed that the IR spectrum can no longer be approximated by the pure component spectra. This is a possible indication of chemical or conformational changes in the system, typical of a compatible blend. From subtraction of spectra of the blended system and pure components, a small residual band at $1718\,cm^{-1}$ was observed, indicating a distinct interaction of the carbonyl groups with a spectral shift of $12\,cm^{-1}$. It was noted that the PVDF spectra show an induced conformational change at a low percentage of PVDF. From these studies, PVDF/PMMA blend compatibility is attributed to chemical interactions between polymer chains of the two components. Another explanation to account for these results is the occurrence of an optical effect.[105,106] Based on the differences between the real parts of the refractive index of the pure components, the blending process can alter the band shapes and cause shifts to lower frequencies. As a comparison, Coleman studied a blend of poly(ε-caprolactone) (PCL) and poly(vinyl chloride) (PVC) in which the refractive indices are identical.[107] In this study, the author correlated frequency shifts and broadening of the carbonyl band to the concentration of PVC, attributing 9% of the carbonyl peak to the frequency shift. As a comparison, an incompatible blend system was studied by substituting poly(β-propiolactone) (PPL) for PCL. This experiment showed no chemical interactions in the IR spectra, confirming that the optical effect does not solely account for the interaction response.[108] A blend of polycarbonate (PC) and PCL was studied with FTIR.[109,110] By increasing the concentration of PC, the PCL crystallinity decreases. This illustrates the plasticizing effect of PCL in a PC matrix and establishes the mobility of PC in the blend.

The processing of a blended system has been shown to influence the distribution of pure components. To determine the effects of processing, Pennington et al. utilized PA depth profiling to identify the distribution of components in injection-molded thermoplastic olefin (TPO).[111] TPO is a rubber-toughened polymer blend of PP, EPR, PE, pigments, fillers, and additives.[112] Because the distribution of TPO components may have significant effects on such properties as adhesion and durability, understanding the distribution of individual components is particularly important. Using various modulation frequencies, talc was detected predominantly at the surface, crystalline PP just below the surface, and the EPR elastomer layer extends into the bulk, as shown in Figure 9. PAS and DMTA analysis indicates that phase separation occurs in the TPO polymer matrix. This phenomena of phase separation of organic coating components has been observed spectroscopically for many systems, including melamine polyesters,[36] alkyds,[37] urethanes,[38,39] and latexes.[51,52]

4.2 Intermolecular Interactions

FTIR theory is used to associate spectroscopic changes with intermolecular interactions, such as hydrogen bonding. Frequency shifts and intensity changes are common

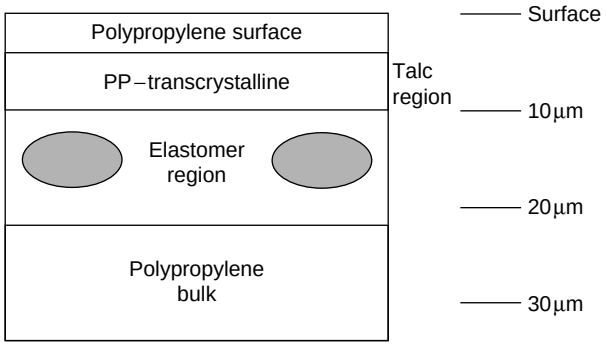

Figure 9 Side profile of proposed stratification profile of TPO.

indicators of the vibrational state of molecular structures. Measurement of these spectroscopic features is useful in determining a molecular origin of thermodynamic parameters.

Characterization of intermolecular interactions has contributed to a better understanding of structure–property relationships in polyurethane chemistry. The exceptional mechanical properties of polyurethanes can be attributed to the interaction of the elastic behavior of the soft segments and the high modulus of the hard segments which phase separate into well-defined regions. In these polymers, IR bands can easily distinguish between the hard segments (isocyanates) and soft segments (polyesters, polyethers or polybutadiene).[113,114] If the proper spectroscopic techniques are used, one may obtain information about the mixing behavior of hard and soft segments through interpretation of spectroscopic features associated with strong intermolecular interactions, such as hydrogen bonding, at the interface between the two components.[115] Previous polyurethane studies examined the effect of mold substrates on molecular interactions at the film–substrate interface.[40] The authors showed that urethane hydrogen-bonded C=O groups exhibit preferential orientation as a function of substrate surface tension. Likewise, substrate surface tension was shown to influence the mobility of additives in latexes[24,116] and film formation of water-borne coatings.[117] Other areas of research rely on surface tension gradients to form self-stratifying coatings.[118,119]

These studies are particularly relevant to polyurethanes because chemical interactions occurring between hydrogen-bonding functional groups will strongly contribute to the mechanical properties of the polymer structure.[120–123] From FTIR spectra, the amide II stretching modes of polyurethane and polyurea are detected at 1535 cm^{-1}. The bands in the 1700 cm^{-1} region, assignable to the amide I C=O stretching vibrations, can be separated into two components. The first, at 1760 cm^{-1}, is assigned to free C=O groups, while the second, at 1690 cm^{-1} is assignable to the C=O groups participating in hydrogen bonding. Since the origins of fundamental vibrations are well understood, both the orientation and relaxation behavior of chain segments can be used to identify the structural component responsible for the toughness of polyurethanes.[115]

PAS is useful in the study of interactions between poly(ethylene terephthalate) (PET) and poly(butylene terephthalate) (PBT).[70] By examination of the crystallization behavior of PET/PBT blends, the kinetic profile was obtained. The crystallization half-life obtained from transmission measurements was in close agreement with the PAS results. PAS analysis, however, did not require extensive sample preparation, unlike transmission FTIR. The SSPA method has been used to examine stratification processes in TPOs by monitoring molecular-level processes that occur near the substrate surface, which inherently influences adhesion to organic coatings.[25] It was shown that the interactions between the substrate components and the coating are strongly influenced by stratification over the sample thickness. Based on rheo-photoacoustic (RPA) FTIR results, correlations between stratification processes and work of adhesion can be determined.

The orientational behavior of blends was studied in a PPO/PS system.[124] It was observed that the PS chains increased orientation with concentrations up to 24% PPO and then remained constant whereas the PPO chains did not exhibit orientational behavior as a function of concentration. These results are explained in terms of the different relaxation behavior of the two types of chains.

5 POLYMER ADHESION

Although macroscopic evaluation of adhesion has important implications, adhesion can be further understood from a molecular-level perspective. Mechanical test methods provide information concerning only the forces that hold two substances together, but do not provide the origin of physicochemical processes responsible for adhesion.[125,126] In contrast, spectroscopic methods are capable of monitoring molecular-level interactions responsible for adhesion and thus further enhance knowledge of the factors promoting or disrupting adhesion.[35,127–129] As FTIR provides a correlation between the spectral changes and interfacial properties at the coating–substrate interface, it has potential for studying the work of adhesion.[130]

Since the characterization techniques previously outlined previously provide information about interactions at interfacial regions, we shall focus on the spectral features that allow measurement of adhesion. By probing interfacial regions, factors which lead to adhesive failure can be identified. This information will allow one

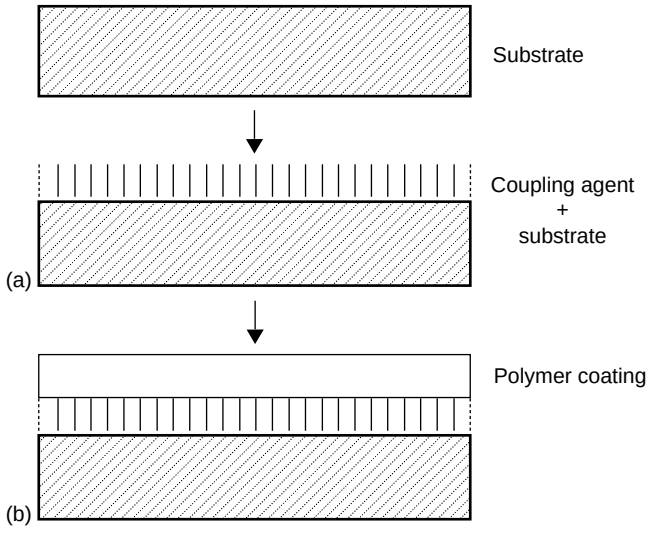

Figure 10 Schematic diagram of adhesion promotion: (a) substrate pretreated with coupling agent; (b) polymeric coating applied to pretreated substrate.

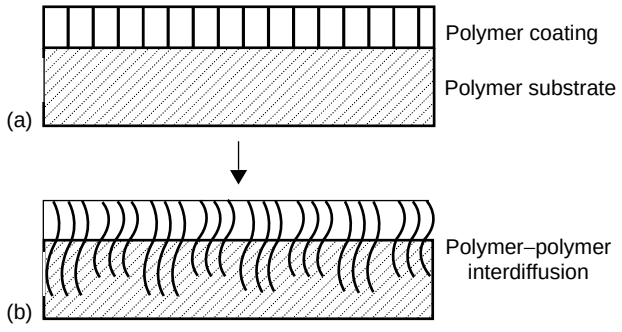

Figure 11 Schematic diagram of adhesion promotion: (a) polymeric coating applied to substrate; (b) interdiffusion of polymeric coating into polymeric substrate.

to modify interfacial chemistry, ultimately leading to the improvement of adhesion. The promotion of adhesion can be achieved by increasing the affinity of the adhesive and adherend. This can be accomplished in many ways: by use of a coupling agent (Figure 10) and through the process of interdiffusion (Figure 11).[131–135]

5.1 Coupling Agents

The adhesive performance of two materials, such as a coating on a substrate, is strongly affected by the nature of the interfacial region. As two homogeneous phases are brought into contact with each other, the presence of incompatibilities at the adjoining surfaces will prevent or limit adhesive interactions. To improve surface compatibility, a primer can be applied to the substrate before topcoat application. The primer, also known as a coupling agent, is a bifunctional molecule which has one functionality capable of reaction with the substrate and a second functionality suitable for reaction with the coating, thus allowing two surfaces to be ''coupled'' together, as shown in Figure 10. The most commonly used coupling agents include silanes, titanates, zirconates, and other bifunctional molecules.[136,137] Since the investigation of primer chemistry is fairly extensive, the focus of this section will be on spectroscopic research applied to promotion of adhesion on substrate surfaces.

Using characterization techniques described previously, silane coupling agent reactions occurring at the interfacial region can be investigated. By using silanes as a composite primer, Chiang and Koenig[138] demonstrated by FTIR studies that amino-substituted silanes are ineffective as coupling agents for anhydride-cured epoxy composites. They showed that anhydrides, instead of reacting with the epoxide groups, undergo a side reaction with primary amine groups of the coupling agent to form cyclic imides, limiting potential cross-linking at the interphase region. Plueddemann later showed that silane coupling agents with chloropropyl, epoxy, or cationic styryl functionalities were more compatible with anhydride–epoxy composites.[139] To characterize further the thermal reactivity of silanes, Culler et al. studied amino-functional silanes placed on a KBr substrate and exposed to air.[140] Subsequent IRS analysis showed the formation of hydrogencarbonates below 95 °C and imines above 125 °C. Carton studied the anhydride and epoxy reaction on Ge at 150 °C for 3 h in the presence of aminosilane using IRS.[141,142] These studies showed that, while water remained at the interface, the silane amino groups were able to react with the anhydride to form amide groups, without the cyclic imide side reactions. Likewise, the conversion of γ-aminopropyltriethoxysilane (γAPS) amino groups into amides was observed on KRS-5 and sapphire crystals to depend on temperature.[143] Subsequent FTIR analysis by Chiang et al. determined that γAPS hydrolyzes almost immediately in aqueous solutions to form stable zwitterion structures.[144]

The prediction of primer thickness and stability is imperative to forming a suitable interphase. As a silane coupling agent is applied to a glass surface, Koenig showed several distinct molecular-level structures are formed.[145] The mechanism of coupling was shown to occur by hydrolyzed silanes hydrogen bonding with silanol groups on the glass surface, forming monolayers up to 300 monolayer equivalents. Using FTIR, the vinyltriethoxysilane adsorbed layer on glass was shown to increase with increasing silane concentration and treatment time.[146,147] Similar spectroscopic studies of γAPS showed much faster adsorption, owing to the autocatalytic effect of the amino group.[148]

5.2 Composite Interfaces

In composites, interfaces between fibers and a polymer matrix are essential in achieving mechanical integrity. FTIR analysis provides an excellent alternative, since measurements can be made without actual contact to the sample surface, allowing studies of fibers and also other nonuniform sample shapes. Urban and Koenig showed that the orientation of surface species could be determined by comparison of PA spectra, one collected with a highly polarizable coupling gas and another with a nonpolarizable coupling gas.[149] Chatzi et al. applied this method to the analysis of Kevlar fibers and showed that orientation of phenyl groups was parallel to the surface of the fiber skin, while the radial orientation of the polymer chains in the fiber core was observed from the N–H stretching modes.[84,150] Tidrick and Koenig investigated the effects of epichlorohydrin and sodium hydride modification on the surface of Kevlar fibers.[151] Improved adhesion of the treated Kevlar fiber in composites was attributed to polyether branching from the polymer backbone, as shown by PA/FTIR spectra. Adhesive improvement was attributed to enhanced compatibility with the composite matrix, without the formation of covalent interfacial bonds.

Nextel™ ceramic fiber/polyimide composite interfaces were investigated using PAS, allowing monitoring of interactions between the fiber and matrix phases.[152] The necessity for a coupling agent was found to depend on the properties of the polyimide. Degradation of Nextel™ fibers in acidic, neutral, and basic aqueous environments was observed using PAS.[153] Using a subtraction method, the presence of positive bands at 1300 and 850 cm^{-1} indicated an increase in borates [$B(OH)_3$] at the surface of the fibers, possibly as a result of B_2O_3 reacting with water. PAS was used to study the effects of oxidation on polyacrylonitrile (PAN)-derived carbon fiber composites with poly(phenylene sulfide) (PPS), poly(ether ketone) (PEK), and epoxy resins.[154] As the degree of oxidation was increased, the band intensities at 2920 cm^{-1} (CH stretch) and 2250 cm^{-1} (CN stretch) decreased, indicating structural changes at the carbon fiber–matrix interface. Studies of the mode and mechanisms of thermal oxidation in glass–nylon 66 composites was also observed with PAS.[155] Thermal oxidation of the composite exposed to 150 °C for 5000 h resulted in transamination reactions and formation of α,β-unsaturated carbonyl species with chain scission.

5.3 Polymer–Polymer Interdiffusion and Adsorption

Adhesion typically takes place in the form of chemical bonding, such as covalent or van der Waals interactions, yet physical bonding plays an important role in the adhesive bonding strength. Physical bonding may result from diffusion of adhesive molecules across the interface to achieve close contact with the substrate or by the forces of physical adsorption between the adhesive and substrate.

Diffusion at polymer–polymer interfaces, such as shown in Figure 11, is important to the adhesive integrity of the bonded system. Voyutskii first suggested that adhesion results from the interdiffusion of polymer segments across the interface.[156] IR spectroscopy plays an important role in the study of interdiffusion, since deuterium labeling of each phase is not required and the diffusion process can be measured in situ.[157] The interdiffusion of a PS–poly(vinyl methyl ether) (PVME) composite was used as a model of adhesion at the interfaces of the two domains.[158] Measurements of interdiffusion of PS/PVME on a Ge IRE at temperatures above and below the T_g of PS allowed the determination of the interdiffusion coefficient (1.1×10^{-12} cm^2 s^{-1}). It was found that interdiffusion is not dominated by either component; instead, it is controlled by the rate of the PVME swelling of PS. Since the extent of adhesion is directly related to the extent of diffusion and the interfacial thickness, this analysis can be applicable to the design of rubber-toughened plastics or bioadhesives.

Adsorption is a phenomenon associated with the exudation or removal of species at an interface. This process is usually governed by van der Waals forces or donor–acceptor interactions. In general, the greater the polarity of two molecules, the greater is the molecular attraction between them. Cook et al. studied the absorption of iron oxide particles to various polymer substrates using PAS.[159] Using subtraction of the iron oxide bands from the spectra, preferential adsorption to more polar substrates, such as a phenoxy (PKKH®), was observed. Similar PA studies were performed to study the adsorption of γ-iron oxide powders on polyurethanes for possible application as magnetic recording media.[160] Surface reactions were apparent on the polyurethane substrates as a result of preferential absorbance to polar functionalities. Characterization of the reactivity of polycyclic aromatic hydrocarbons (PAH) on silica and alumina surfaces was researched by Saucy et al.[161] Folkers et al. illustrated the extensive control over surface wettabilities through the use of self-assembled monolayers (SAMs) of alkanethiol on gold surfaces.[162] By competitive adsorption between short- and long-chain alkanethiols, the structure ordering can be properly established with polarized IR external reflectance spectroscopy.[163] From these analyses, the interfacial region was discovered to be significantly disordered as a result of the relationship between the molecular structure and wetting.

5.4 Film Formation

During the cure of an adhesive, shrinkage usually occurs as a result of cross-linking, solvent loss, or cooling of

the adhesive material. Since the contact area of the adhesive to the substrate remains essentially constant, internal stresses will build up, which ultimately reduce the fracture strength of the adhesive bonding. Greenblatt et al. observed spectroscopically the cohesive failure of polyimide coatings on metal substrates in the presence of the γAPS.[164] The failure was attributed to increased internal stresses due to the thermal mismatch between polymer and metal layers, leading to chain scission of the polyimide. By reducing the imidization temperature from 350 to 320 °C, the adhesive bond strength was shown to increase. In contrast, copper substrates inhibit imidization by the formation of copper carboxylate. In this instance, IRS studies show chemical bonding between copper and polyimide.[165]

An understanding of the chemistry during the film formation process is an invaluable tool in the design of durable coatings. The research of Jang and Lee indicated that the interfacial adhesion between benzophenone tetracarboxylic dianhydride and oxydianiline (BTDA and ODA) polyimide and aluminum depends highly on the imidization temperature.[166] As observed with IRS, increase of the cure temperature leads to the formation of ester and carboxylic groups, at the expense of imide ring cleavage. Shifting of the carboxylic acid band from 1725 to 1710 cm^{-1}, and a decrease from 1670 to 1660 cm^{-1} for the carbonyl group, indicate chemical interaction with the surface hydroxyl groups. The band intensities of these chemical interaction increase with increase in temperature, correlating with a maximum peel strength at 320 °C. Likewise, Flament et al. studied the PMDA/ODA polyimide–aluminum interface (PMDA = pyromellitic dianhydride) using XPS analysis and verified that direct bonding does occur.[167]

The use of vibrational spectroscopy to correlate interfacial forces with work of adhesion has long been a goal of scientists. This approach takes advantage of one of the major strengths of vibrational spectroscopy, namely the extreme sensitivity of vibrational spectra to the changes in the local environment. To accomplish this, a unique method of monitoring adhesion of a coating to a substrate was developed by combining tensile elongation and PA detection as a RPA/FTIR cell (Figure 12).[129,130] With this device, one is able to monitor molecular changes upon applying stress to the substrate. Using this technique, the adhesion of a polydimethylsiloxane (PDMS) coating to a PE substrate was investigated.[168] These studies showed that the intensity of the PE C–H stretching bands at 2925 and 2850 cm^{-1} as a result of elongation of the substrate increases. Although the initial increases are small, possibly owing to thinning of the substrate, a large increase at 17% elongation is attributed to interfacial failure in the form of microvoids. The microvoid formation was independently confirmed by SEM analysis.

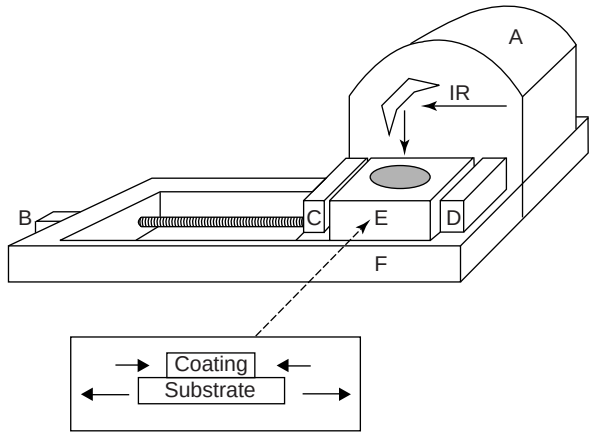

Figure 12 Schematic diagram of cell adapted for RPA measurements: A, microphone compartment; B, leadscrew for sample elongation; C,D, movable and stationary clamping blocks; E, sealed sample area; F, gas inlet for purging the cell.

6 POLYMER SURFACE CHARACTERIZATION

Many polymers are unsuitable for bonding to other materials because of the low surface energy of the surface structure. The pretreatment of surfaces is a commonly used method to tailor the surface chemistry of polymers and achieve adequate bond strength. FTIR studies have often been employed in characterizing these surface changes, since the region of interest is typically less than 2 μm from the surface. Transmission FTIR has shown limited success in this area owing to the relatively ineffective surface sensitivity of this method. Reflectance techniques, on the other hand, have been used extensively, since these techniques employ nondestructive depth-profiling measurement with preferred sensitivity to surfaces. In the following sections, FTIR studies of surface pretreatments will be presented.

6.1 Plasma Methods

Plasma treatment of polymers consists of placing a material in a vacuum chamber and introducing an appropriate gas which is ionized to form a plasma. During the plasma treatment, adsorbed material on the polymer surface is removed and reactive groups are created through a free-radical formation process. Determination of the changes which take place during this process provides important information for subsequent coating operations.

Plasma pretreatment of polymers has shown promising application in the automotive, medical, and microelectronics fields. The elastomeric nature and relatively inert composition of polymers make this material suitable for biomedical substrates. Plasma polymerization methods have been the focus of recent studies in achieving surface modifications. These modifications are accomplished through the deposition of a thin film or by reaction with surface functionalities. Plasma polymerization provides a well-controlled process for the application of films ranging from 10 to 500 Å using a wide range of monomers. A complete overview of plasma polymerization reactions and characterization of these systems was provided by Yasuda.[169]

Plasma polymerization can be used to obtain uniform polymer coatings on a variety of surfaces. Gaboury and Urban examined the effect of both a nitrogen and argon gas plasma on PDMS films, monitoring surface changes with FTIR/IRS.[170] After plasma exposure, FTIR detected bands at 2158 and 912 cm^{-1}, indicating the formation of Si–H groups on the polymer surface. Waldman et al. used FTIR to characterize polymer deposition rates for controlled adhesion of graphite fibers to PEEK® (polyether-ether-ketone).[171] By monitoring processing parameters, such as composition of the monomer mix, flow rate, pressure, and power of the rf field, the plasma deposition rate was found to be the smallest for 100% ethylene and increased threefold when ammonia was added to the monomer mixture. From the FTIR results, the polymer coatings were found to be of uniform thickness and exhibited a complex cross-linked structure.

In a similar study, hydroxyl-rich surfaces were obtained from a plasma of allyl alcohol on various polyolefin substrates.[172] R–A spectroscopy was used to obtain structural information of the plasma polymerized layers, which were found to be less than 50 nm thick. FTIR analysis verified that the substrate did not seriously affect the chemical structure of the plasma polymer, but the adhesive properties were found to depend significantly on the specific substrate. The difference in adhesive properties between the substrates may be explained as an effect of vacuum ultraviolet (UV) emission present during the plasma deposition. Radicals formed in the plasma near the surface of the substrate cause chain scission reactions, lowering the cohesive strength of the substrate owing to the formation of weak boundary layers. Savage et al. reported an improvement of molecular compositional control by use of a variable duty cycle pulsed rf plasma.[173] The plasma-generated film was monitored spectroscopically using FTIR and XPS. Using SSPA/FTIR, Palmer et al. indicated that, by varying the modulation frequency, depth information of plasma-deposited films of cobalt tetraphenylporphrin and hexamethyldisilazane on PP could be obtained.[71,174]

6.2 Chemical Methods

The surface energy of polymer substrates can be strongly influenced by exposure to specific types of surface treatment. The chemical processes include chromic–sulfuric acid, ozone, and gas-phase fluorination methods. Morphology changes from the chemical treatment of TPO have been observed with PAS.[111,175] FTIR spectra of TPO indicated that stratification of PP crystalline components can be identified by monitoring the bands at 1000 and 843 cm^{-1}. The 1000 cm^{-1} band is attributed to the combination of CH_3 rocking, CH_2 wagging, and CH bending motions, while the 843 cm^{-1} represents the coupled CH_3/CH deformation and CH_2/CH_3 rocking normal vibrations.[176] Previous studies on PP blend properties concentrated on polymer composition,[177] processing conditions,[178] morphology,[179] mechanical response,[180] and paintability. Ryntz showed that changes in morphology of TPO are seen upon treatment of surfaces with solvents.[181] Studies have shown that a relatively small change in cure time or temperature will have a profound effect on film adhesive properties.[182]

Chemical vapor deposition was examined as an alternative technique for preparation of poly(phenylene vinylene) (PPV) films, a promising candidate for the active layer in light-emitting diodes. By depositing a chlorinated xylene layer, polymerization on the substrate surface will give a Cl precursor polymer, which on heating forms the conjugated PPV structure.[183] In situ infrared reflection–absorption spectroscopy (IRRAS), ex situ transmission IR, and polarized IRS spectra were used to characterize these films.

7 POLYMER BULK CHARACTERIZATION

The typical time resolution of FTIR provides spectroscopists with the opportunity to obtain an IR spectrum in the millisecond timescale, enabling the observation of reaction changes to be studied. By plotting the absorbance of a characteristic group frequency, a kinetic profile of the process can be obtained. A kinetic profile can also be obtained by the use of differential spectra by subtraction of each spectrum from the preceding one with respect to time.[184] The rapid sampling provided by FTIR allows the monitoring of short-lived spectroscopic changes for many polymer processes, including curing kinetics,[15,185] orientation,[186] and relaxation.[186] Recent developments in step-scan interferometry pushes the time resolution into the nanosecond range, since the limiting factor is the response time of the detector, not the mirror movement.

7.1 Thermoset Curing

In characterizing polymer curing, it is important to evaluate how both the microstructure and the superstructure change as a function of time. The kinetics of the curing process for polymeric materials can be carried out using FTIR.[15,187,188] By collecting spectra at designated intervals, kinetic profile measurements can be made, where the time intervals of collection are determined by the reaction speed. Accurate study of polymerization kinetics requires rapid collection in a rapid time frame with an extremely fast spectrometer. The relatively new class of spectrometers which are capable of following reaction kinetics have been termed real-time Fourier transform infrared (RTFTIR) spectrometers. Recent advances in FTIR instrumentation have allowed up to 204 spectra per minute to be obtained.[189]

Spectroscopic monitoring of adhesive cure rates is a useful application of IR spectroscopy. RTFTIR was utilized to examine the kinetic profile of an anaerobic adhesive, Loctite® 642.[190] These studies examined the effect of surface-activated polymerization on copper, aluminum, galvanized steel, zinc-plated steel, stainless steel, and passivated stainless steel. The copper surface was the most active surface for initiating the polymerization, while passivated stainless steel specimens showed a higher ultimate degree of conversion and cure speed over the surface without passivation. The increase in surface activity was attributed to the passivation process, which results in roughening and removal of inert oxide layer of the surface. The surface reactivity was evaluated based on the IR spectra of the anaerobic cure in a real-time mode and then correlated with the metal type and content in the interface region. The surface texture and elemental analysis of the metal surfaces were measured using SEM and EDX.

Using RTFTIR, the mechanisms for photochemical reactions of polymer systems can be defined. UV cure reactions typically occur very rapidly, making the spectroscopic monitoring of these reaction difficult. Yang et al.[190] studied the UV-initiated anionic polymerization of ethyl cyanoacrylate. The reaction was monitored using the C=C stretching band at 1617 cm^{-1}. From the RTFTIR data, the highest cure levels were obtained with a high intensity light source with the reaction occurring in the first seconds of the reaction. Other reaction parameters can be studied, such as oxygen inhibition or photoiniator concentration. This information is critical in assessing the structure–property relationships in the adhesive formulation.

FTIR spectroscopy was used to study the photopolymerization of 4,4′-bis(acryloyl)biphenyl (BAB), dispersed in a liquid crystal matrix (E7).[191] The IR absorbance of the carbon–carbon double bond of the acrylate group was monitored to follow the course of the polymerization and to determine the extent of cure. It was observed that a major change in the character of the IR absorbance band shape and position of the acrylate carbonyl group accompanied the decrease in the carbon–carbon double bond absorbance. Curing was found to be rapid, of the order of minutes. It was found that the reaction did not proceed appreciably after a brief initial exposure to UV radiation. This result implies that this type of system may be time-stable with programmable degrees of cure. It was also discovered that single continuous doses of UV exposure produced a higher degree of cure than equivalent UV exposure times administered in multiple shorter doses. This observation indicates the existence of a relatively long induction period in this system. The long induction period may be caused by a UV screening effect due to the other components of this system.

Polyurethanes are high-performance polymers known for abrasion resistance, hardness, and flexibility properties.[192–195] Owing to these factors, polyurethanes are becoming more widely used as coatings. In order to understand the process of polyurethane film formation, FTIR can be used to track cross-linking reactions in polyurethanes[196] and related rheological factors.[197,198] The primary interest is monitoring of the cross-linking reactions when NCO reacts with the active hydrogen of a hydroxy-functional compound and a urethane linkage is formed. A typical polyurethane system is illustrated in Scheme 1. When isocyanate reacts with water by a condensation reaction, an unstable intermediate, carbamic acid, is formed. This immediately dissociates to evolve carbon dioxide gas and an amine, which continues to react with free isocyanate to form urea, as shown in Scheme 2.

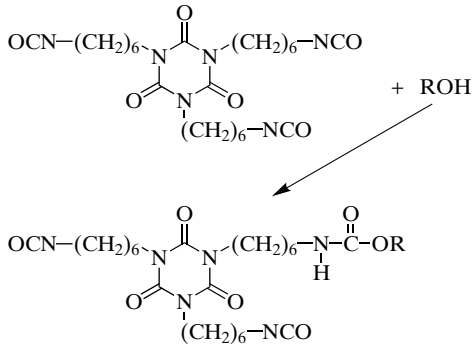

Scheme 1

The NCO reaction with water can significantly affect film formation, as the evolution of carbon dioxide gas may result in undesirable property changes. Furthermore, accelerated reaction rates between isocyanates and

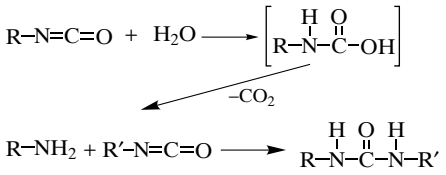

Scheme 2

amines may cause a rapid increase in molecular weight and viscosity, thus reducing pot life. Since amines are more nucleophilic than alcohols, urea reactions tend to be faster and may predominate the cross-linking process over the urethane network. Under these circumstances, the presence of water is undesirable and may promote hydrolytic cleavage of polyester groups. Using ATR/FTIR and DMTA, Allison demonstrated the influence of humidity on polyurethane film formation near the film–air and film–substrate interfaces.[199,200] Polyurethane cross-linking reactions have been monitored and quantified by following the decrease of the reactant NCO band intensity along with the increase of the product CO_2 band.[38,40,41] As cross-linking reactions progress, the isocyanate band intensity detected at 2271 cm^{-1} decreases, while the band at 2337 cm^{-1} attributed to CO_2 gas formation increases. The results presented in Figure 13 show Circle™ ATR spectra recorded from 10 to 50 min. These cross-linking reactions leading to polyurethane formation can be quantified by using the Beer–Lambert law (Equation 6):

$$A = \varepsilon bc \tag{6}$$

From the concentration profile shown in Figure 14, the molar absorption coefficient of isocyanate and CO_2 can be determined. During this time, from 10 to 50 min, CO_2 increases much faster than the decrease in NCO.

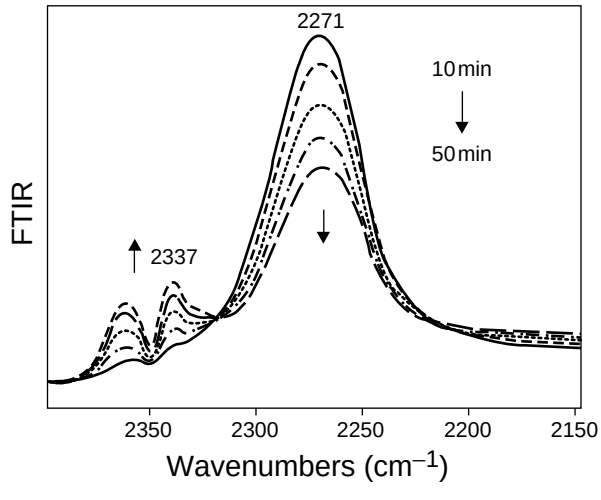

Figure 13 Circle™ ATR/FTIR spectra of water-borne polyurethane recorded from 10 to 50 min.

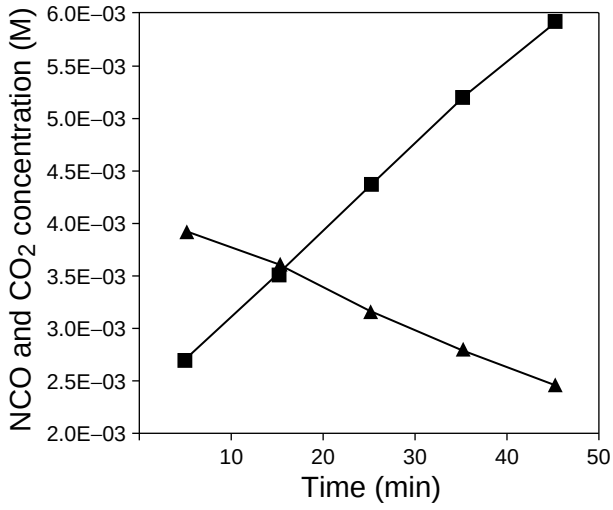

Figure 14 Concentration changes of ( ) NCO and (■) CO_2 species as a function of reaction time.

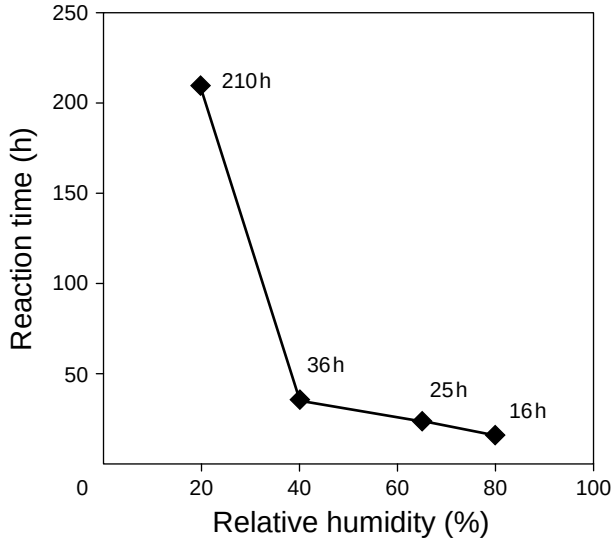

Figure 15 Cross-linking reaction time plotted as a function of RH.

Because NCO functional groups can react with water from the surrounding environment, relative humidity (RH) is expected to have a significant effect on the rate of isocyanate consumption, as illustrated in Figure 15. Because solvent evaporation and partial vapor pressure influence cross-linking reactions, the concentration of NCO was monitored as a function of RH, time, distance from the interface, and polarization. By DMTA analysis of films cured at various humidity levels, the RH influence on glass transition temperature (T_g) and cross-link density (XLD) can be seen. The DMTA results are shown in

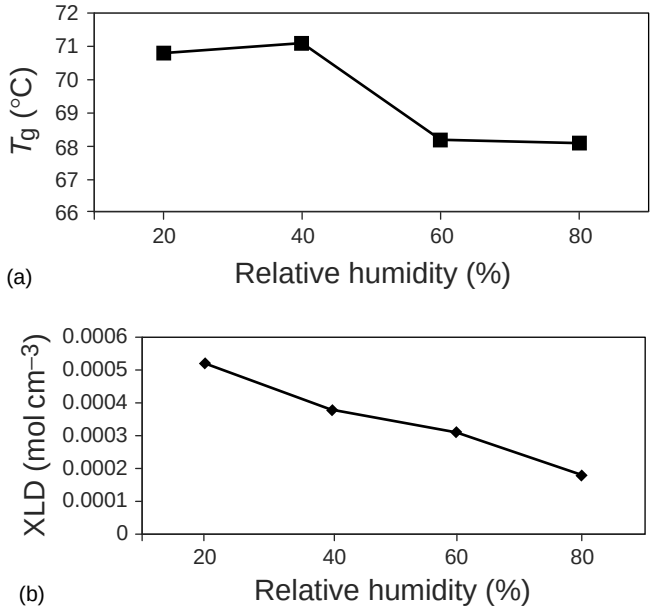

Figure 16 (a) T_g and (b) XLD of polyurethane films cross-linked at 20, 40, 60, and 80% RH.

Figure 16(a) and (b) for the polyurethane specimens cross-linked at 20–80% RH.

Component stratification in polymer films has been successfully investigated with ATR/FTIR spectroscopy. Hamada et al. studied the curing process of a polyester and hexamethoxymethylmelamine (HMMM) resin system, in which a concentration gradient was observed.[201] The degree of surface enrichment strongly depended on the HMMM to polyester monomer ratio. IRS analysis indicated that as the monomer ratio was increased above a 29 : 71 HMMM to polyester ratio, an increase in HMMM at the surface resulted. As the monomer ratio was lowered, HMMM became more pronounced in the cured material bulk. The authors explained these results as a selective activation of the curing catalyst at the surface region leading to a gradient in component concentrations. The melamine molecules are able to diffuse and the concentration gradient induces an HMMM current from the inner portions towards the surface. Further studies with butylated resins show very high concentrations of the butylated component on the surface of the film, yet remaining constant in the bulk. Butylated melamine enriched the surface because of the low surface tension compared with that of polyester and solvents in the formulation.

7.2 Film Formation and Stability

As shown previously, the mode and extent of film formation can be monitored readily with FTIR analysis. Hirotsu investigated the protective effects of coatings on polyolefin substrates from oxidative plasma etching using FTIR.[202] PP was coated with an ultrathin polymer layer of hexamethyldisiloxane. After exposure to an O_2 plasma, the IRS results suggested that the cross-linked network of poly(hexamethyldisiloxane) was contributing to an improved etching resistance. FTIR spectra of the PP surface after O_2 plasma treatments indicated an increase in the peak intensity at 1023 cm^{-1}, reflecting Si–O formation. Etching was correlated to the weight loss of the PP substrate after the exposure to an oxidative plasma of O_2 or air. The etching resistance of the protective polymer coating was investigated with respect to the type of polymer, protective layer thickness, and oxidative plasma etching conditions. The weight loss of the coated substrate decreased compared to the uncoated PP film. The etching resistance was attributed to the formation of the cross-linked network in the protective layer.

IRS/FTIR spectroscopy is a common choice for the analysis of interfaces and orientation effects that occur during film formation of water-borne polyurethanes. Previous studies of water-borne systems examined the concentration changes of isocyanate to elucidate factors which influence film formation processes.[41] By examining the amide II stretching modes in the 1530–1560 cm^{-1} region, the influence of RH was correlated with the urethane and urea concentrations near interfacial regions.

Antoon and Koenig studied the reaction mechanism of the copolymerization of cyclic anhydrides and epoxy resins using absorbance difference spectra.[15] Solutions of Epon 828 epoxy resin (Shell), methylbicyclo-[2.2.1]heptene-2,3-dicarboxylic acid anhydride (NMA), and dimethylbenzylamine (BDMA) were analyzed on NaCl plates. The proposed tertiary amine (BDMA) catalysis was not observed, since there were no absorption increases at 1600 or 3500 cm^{-1}, which would indicate changes to C=C or OH groups for the postulated allylic alcohol intermediate. The role of the tertiary amine was observed to differ drastically in the presence or absence of the anhydride functionalities. The preferred reaction mechanism involved activation by the hydroxyl groups in which carboxylamine and hydroxylamine act as active centers during the polymerization. Further kinetic studies of the reaction in the presence of fillers showed that E-glass increased the reaction rates, whereas fumed silica decreased it.

The film formation mechanism of ZrO_2 deposited on a titanium support was followed as a function of the firing temperature.[203] It was found that an intramolecular hydrolysis takes place upon heating, resulting in the liberation of hydrogen chloride. FTIR and Fourier transform Raman investigations revealed that polymorphism occurs in the oxide film. The upper surface of the film is of tetragonal structure, whereas the bulk is a cubic form stabilized by TiO_2 formed by oxidation of the support.

Boerio et al. used the R–A technique to study silanes and epoxy resins on metal surfaces at thicknesses of 15 Å.[204,205] The results indicated that the molecules were absorbed on the reflective surface in a conformation with the oxirane oxygen in contact with the surface.[205]

Characterization of the degradation processes of polymers is essential in establishing the integrity of films. Factors such as temperature, humidity, corrosive elements, acidity, and UV radiation can be integrated into weathering environments for IR analysis. Film degradation studies have been made of thin films of PMMA on gold, nickel, and zinc surfaces using external reflection spectroscopy (ERS).[206] If the optical constants of a film are known, the thickness changes of the film as a result of degradation can be calculated based on reflective theory.[207] The solar-induced degradation of PC on metal surfaces was studied by exposure of the film to a solar source for various time intervals. Increases in absorbance in the 3500–3250, 1690, 1620, 1590, 1490, 1340, and 1260 cm^{-1} regions suggest the formation of hydroxyl, a portion of which is hydrogen-bonded, a salicylate-like carbonyl, and a conversion of carbonate C–O bonds into phenolic C–O bonds.[208] Using an environmental R–A cell, where UV radiation, thermal, and environmental exposure are controlled, the photodegradation of PC on gold mirrors was studied by Webb.[208]

PAS and thermogravimetric analysis (TGA) were used by Paroli et al. to study the effect of UV radiation and aqueous environments on polyurethane sealants.[209] By monitoring the chemical composition, the formulation and filler content were shown to have a substantial effect on the polymer degradation. The most prominent spectroscopic effect was detected by PAS is the 3200–3600 cm^{-1} OH stretching region which results from the cleavage of urethane cross-links.

FTIR spectroscopy is commonly used in assessing the lifetime and degradation pathways of protective coatings under corrosive environmental conditions. The durability of 2-undecylimidazole on a copper substrate was studied by R–A spectroscopy.[210] Since 2-undecylimidazole is commonly used as a corrosion inhibitor, the reactivity with copper was studied at elevated temperatures. From the spectral analysis, cleavage of the imidazole ring was noted in the decomposition products. This information is useful for determination of the integrity of films exposed to corrosive surroundings.

The durability of organosilicon coatings was determined by static SIMS and DRIFT.[211] Hygroscope tests and contact angle measurements prior to spectroscopic surface investigations of a propyl/octylsilane and a methyl/octylsiloxane revealed an improvement in long-term stability with the protective silsesquioxane coatings. Investigations on the weathered coatings after 8 years showed the presence of silsesquioxanes only in case of the propyl/octylsilane system. DRIFT measurements on both fresh and exposed material confirmed the presence of the coating in case of propyl/octylsilane, but not in case of methyl/octylsiloxane.

An IRE cell was used to model the effects of thermal aging in enamel-coated wire.[212] Polyamide–polyimide (SX-81002) was cured on a zinc selenide IRS element coated with a thin film of metallic copper. The coated crystal was inserted in a Circle™ ATR cell housed in a heating jacket and set at 250 °C to simulate thermal aging of enamel-coated wire. Evaluation of the IR spectra in the fingerprint region suggested that the polymer experienced chemical degradation within 23 days of thermal aging. Through comparisons with controls containing copper-free coatings, and aging studies carried out at 28 °C, it was determined that aging at elevated temperatures caused more pronounced chemical changes in the polymer than did exposure to the copper. These results indicate that FTIR/IRS may be a useful tool to detect fatigue after thermal exposure.

The photooxidative stability of acrylic latex films was studied with reflectance FTIR to determine the influence of processing parameters.[213] The nature of impurities and oxidation products generated were characterized and related to photooxidative degradation. IRS and UV spectrophotometric analysis confirmed that hydroperoxides are generated by photochemical process. Initial chemical changes and the subsequent influence on the physical and chemical properties of the acrylic films during the later stages of photooxidation are found to be closely related. The main photooxidation pathways were identified as de-esterification and hydroperoxide formation. While hindered piperidine stabilizers or the addition of a comonomer improved photostability, the incorporation of simple terminal dialkyl acrylamide/methacrylamide groups was shown to reduce photooxidative degradation.

7.3 Orientation Measurements

Recent developments in polymer processing technology have contributed to the development of high-performance materials, by forming fibers or films of ordered polymer chains from their anisotropic states. In these materials, an understanding of polymer chain orientation is important for the evaluation of the material properties of the polymers under study. By employing polarized IR, orientational effects can be characterized by several methods: dichroic ratio,[214] spectral subtraction,[215] sample tilting, or FTIR/IRS.[216]

One of the most interesting recent developments is the use of polarization modulation in the measurement of segmental orientation. The traditional techniques of characterizing molecular orientation include WAXS, small-angle X-ray scattering (SAXS), light scattering,

birefringence and FTIR spectroscopy. Even though orientation information from X-ray scattering can be very accurate and complete, it is still limited to the characterization of crystalline regions. Light-scattering experiments utilize long wavelengths (in the micrometer range), which limits the information obtained to large-scale morphologies. Birefringence techniques are subject to artifacts caused by improper measurements.[217] With improvements in FTIR spectroscopic technology, highly polarized beams can be obtained, making polarized IR absorption experiments a robust tool in orientation measurements. Because of the high selectivity of this spectroscopic method, the orientations of individual structural units can be followed with great accuracy. The usual polarization measurements involve two separate measurements, one parallel and the other perpendicular. From the calculation of their difference, the dichroic ratio is obtained, as well as calculation of the orientation function. For nearly isotropic materials, this type of measurement is limited by the dynamic range of the FTIR instrument. However, with the polarization modulation technique, it is possible to measure the differences directly, thus greatly improving the S/N. This technique was first used with dispersion instruments to measure deformation-induced orientation changes in PE.[218] Structural information is inferred from the relative absorption of intensities of incident IR radiation polarized perpendicular and parallel to the plane of incidence. Polarization techniques are only useful if the bands are previously assigned with well-defined transition moments. Details of polarization experiments and possible errors have been described in a number of studies.[219,220]

Recent improvements in the processing technology of polymers have led to an influx of high-performance materials. In these materials, an understanding of structure–property relationships is critical to the development of new polymers. Garton examined the orientation induced by draw processing of PET fibers.[215] In this case, where uniaxial orientation can be presumed, it is possible to calculate the relative amounts of the trans and gauche isomers. The measurements indicate that increased orientation of the amorphous phase leads to an increase in the trans isomer.

Since developments in spectroscopy allow measurements on the microsecond timescale, Nakano et al. applied step-scan capabilities to time-resolved Fourier transform infrared spectroscopy (TRS) to study reorientational dynamics of liquid crystalline materials.[221] While one-dimensional TRS was used to measure real-time orientation dynamics of nematic liquid crystals under the influence of an external electric field, 2D/TRS showed that the rigid and flexible segments of the liquid-crystalline molecules exhibit different reorientational responses to the applied electric field.

The molecular orientation of surface species at the interfacial region plays an important role in determining the interaction of a coating with the interfacial layers. The interaction of mercaptoester coupling agents with gold surfaces was studied using polarized R–A/FTIR at a grazing angle incident angle.[222] Formation of strong chemical bonding is observed as the thioester bonds to the gold surface through sulfur–metal bonds. Only C_{12} (dodecyl thioglycolate) and C_{18} (stearyl thioglycolate) functional coupling agents self-assemble and form a close packed structure on the surface. The C_{12} alkyl chains of dodecyl thioglycolate are densely packed and oriented perpendicular to the substrate surface. In stearyl thioglycolate, the C_{18} alkyl tail chains are also closely packed and oriented, but the chains tilt from the normal surface, whereas the carbonyl group of the C_{18} thioester is parallel to the surface. For short-chain mercaptoesters, such as octyl thioglycolate, the alkyl chains do not self-assemble or orient on the gold surface. Further studies showed that the interfacial shear strength to steel surfaces is enhanced by mercaptoester coupling agents.

ACKNOWLEDGMENTS

I acknowledge my colleagues (Claudia Allison, Anneke Kaminski, Bor-Jiunn Niu, Heung-Soo Kim, Joe Stegge) for their contributions to this article. I also express my gratitude to Read-Rite Corporation management for supporting this effort.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AM	Amplitude Modulation
ATR	Attenuated Total Reflectance
BAB	4,4′-Bis(Acryloyl)biphenyl
BDMA	Dimethylbenzylamine
BTDA	Benzophenone Tetracarboxylic Dianhydride
DMTA	Dynamic Mechanical Thermal Analysis
DRIFT	Diffuse Reflectance Fourier Transform Infrared
DSC	Differential Scanning Calorimetry
EDX	Energy-dispersive X-ray Analysis
EPDM	Ethylene–Propylene–Diene
EPR	Ethylene–Propylene
ERS	External Reflection Spectroscopy
FTIR	Fourier Transform Infrared
GC/MS	Gas Chromatography/Mass Spectrometry

HMMM	Hexamethoxymethylmelamine
I	In-phase
IR	Infrared
IRE	Internal Reflectance Element
IRRAS	Infrared Reflection–Absorption Spectroscopy
IRS	Internal Reflectance Spectroscopy
NMA	Methylbicyclo-[2.2.1]heptene-2,3-dicarboxylic acid anhydride
ODA	Oxydianiline
PA	Photoacoustic
PAH	Polycyclic Aromatic Hydrocarbons
PAN	Polyacrylonitrile
PAS	Photoacoustic Fourier Transform Infrared Spectroscopy
PBT	Poly(Butylene Terephthalate)
PC	Polycarbonate
PCL	Poly(ε-Caprolactone)
PDMS	Polydimethylsiloxane
PE	Polyethylene
PEEK	Polyether-ether-ketone
PEK	Poly(Ether Ketone)
PET	Poly(Ethylene Terephthalate)
PM	Phase Modulation
PMDA	Pyromellitic Dianhydride
PM/IRRAS	Polarization Modulation Infrared Reflection–Absorption Spectroscopy
PMMA	Poly(Methyl Methacrylate)
PP	Polypropylene
PPL	Poly(β-Propiolactone)
PPO	Poly(Phenylene Oxide)
PPS	Poly(Phenylene Sulfide)
PPV	Poly(Phenylene Vinylene)
PS	Polystyrene
PVC	Poly(Vinyl Chloride)
PVDF	Poly(Vinylidene Fluoride)
PVME	Poly(Vinyl Methyl Ether)
Q	In-quadrature
R–A	Reflection–Absorbance
RH	Relative Humidity
RPA	Rheophotoacoustic
RTFTIR	Real-time Fourier Transform Infrared
SAM	Self-assembled Monolayer
SAXS	Small-angle X-ray Scattering
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectrometry
S/N	Signal-to-noise Ratio
SSPA	Step-scan Photoacoustic
TEM	Transmission Electron Microscopy
TGA	Thermogravimetric Analysis
TPO	Thermoplastic Olefin
TRS	Time-resolved Fourier Transform Infrared Spectroscopy
UHV	Ultrahigh Vacuum
UV	Ultraviolet
WAXS	Wide-angle X-ray Scattering
XLD	Cross-link Density
XPS	X-ray Photoelectron Spectroscopy
2D	Two-dimensional
γAPS	γ-Aminopropyltriethoxysilane

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Photoacoustic Spectroscopy in Trace Gas Monitoring

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Near-infrared Spectroscopy of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis

Pulp and Paper **(Volume 9)**
Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. D.I. Bowers, W.F. Maddams, *Vibrational Spectroscopy of Polymers*, Cambridge University Press, Cambridge, 1989.
2. G. Socrates, *Infrared Characteristic Group Frequencies*, Wiley, Chichester, 1980.

3. K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, Wiley, New York, 1986.
4. J. Haslam, H.A. Willis, *Identification and Analysis of Plastics*, 2nd edition, Iliffe, London, 1972.
5. N.B. Colthup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, 2nd edition, Academic Press, New York, 1975.
6. N.P. Roeges, *A Guide to the Complete Interpretation of Infrared Spectra of Organic Structures*, Wiley, New York, 1994.
7. B.C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton, FL, 1996.
8. R.A. Nyquist, R.O. Kagel, C.L. Putzig, M.A. Leugers (eds.), *Handbook of Infrared and Raman Spectra of Inorganic Compounds and Organic Salts*, Academic Press, New York, 1996.
9. A. Garton, *Infrared Spectroscopy of Polymer Blends, Composites, and Surfaces*, Hanser, Munich, 1992.
10. E. Schroder, G. Muller, K. Arndt, *Polymer Characterization*, Hanser, Munich, 1989.
11. C.D. Craver, T. Provder (eds.), *Polymer Characterization: Physical Property, Spectroscopic, and Chromatographic Methods*, Advances in Chemistry Series, American Chemical Society, Washington, DC, Vol. 227, 1990.
12. T.D. Harris, 'High-Sensitivity Spectrophotomerty', *Anal. Chem.*, **54**(6), 741A–750A (1982).
13. J.L. Koenig, 'Effect of Nonuniform Distribution of Absorbing Material on the Quantitative Measurement of Infrared Band Intensities', *Anal. Chem.*, **36**, 1045–1046 (1964).
14. P.R. Griffiths, J.A. DeHaseth, *Fourier Transform Infrared Spectrometry*, Chemical Analysis Series 83, Wiley, New York, 1986.
15. M.K. Antoon, J.L. Koenig, 'Crosslinking Mechanism of an Anhydride-cured Epoxy Resin as Studied by Fourier Transform Infrared Spectroscopy', *J. Polym. Sci., Polym. Chem. Ed.*, **19**, 549–570 (1981).
16. S.A. Liebman, D.H. Ahlstrom, P.R. Griffiths, 'On-line Fourier Transform Infrared Analysis of Pyrolysis and Combustion Products', *Appl. Spectrosc.*, **30**(3), 355–357 (1976).
17. K. Holland-Moritz, W. Stach, I. Holland-Moritz, 'Possibilities of FTIR Instrumentation for the Investigation of Fast Processes: Elongation and Relaxation Phenomena in Polymers', *J. Mol. Struct.*, **60**, 1–6 (1980).
18. K. Holland-Moritz, W. Stach, I. Holland-Moritz, 'Investigation of Time-dependent Phenomena in Polymers by Fourier Transform Infrared Spectroscopy', *Prog. Colloid Polym. Sci.*, **67**, 161–164 (1990).
19. P. Jacquinot, 'New Developments in Interference Spectroscopy', *Rep. Prog. Phys.*, **13**, 267–312 (1960).
20. P.R. Griffiths, H.J. Sloane, R.W. Hannah, 'Interferometers vs. Monocromaters: Separating the Optical and Digital Advantages', *Appl. Spectrosc.*, **31**, 485–495 (1977).
21. J. Connes, P.J. Connes, 'Near-infrared Planetary Spectra by Fourier Spectroscopy. I. Instruments and Results', *J. Opt. Soc. Am.*, **56**(7), 896–910 (1966).
22. J.L. Koenig, *Spectroscopy of Polymers*, American Chemical Society, Washington, DC, 1992.
23. M.W. Urban, *Vibrational Spectroscopy of Molecules and Macromolecules on Surfaces*, Wiley-Interscience, New York, 1993.
24. M.W. Urban, C.D. Craver (eds.), *Structure–Property Relationships in Polymers*, American Chemical Society, Washington, DC, 1993.
25. B.D. Pennington, R.A. Ryntz, M.W. Urban, 'The Effect of Thermoplastic Olefin Stratification on Adhesion; Step-scan Photoacoustic FTIR Spectroscopic Studies', *Polym. Mater. Eng. Sci.*, **77**, 591–592 (1997).
26. B.D. Pennington, M.W. Urban, 'High Pressure Photoacoustic Step-scan FTIR Spectroscopy of Polymers', American Chemical Society Abstracts, *Polym. Matl. Sci. Eng.*, **75**, 39–40 (1996).
27. I. Noda, A.E. Dowrey, C. Marcott, 'Recent Developments in Two-dimensional Infrared (2D IR) Correlation Spectroscopy', *Appl. Spectrosc.*, **47**, 1317–1323 (1993).
28. J.F. Reynolds, 'Effect of Pressure on the Infrared Spectra of Some Hydrogen-bonded Solids', *J. Chem. Phys.*, **41**(1), 47–50 (1964).
29. E.R. Lippincott, 'Studies of Infrared Absorption Spectra of Solids at High Pressures', *Spectrochim. Acta*, **16**, 58–73 (1960).
30. D.J. Garton, D.M. Wiles, 'Infrared Spectroscopy of Polyethylene Terephthalate Fibers – Uses and Limitations', *Text. Res. J.*, **51**(1), 28–34 (1981).
31. J.L. Koenig, 'Application of Fourier Transform Infrared Spectroscopy to Chemical Systems', *Appl. Spectrosc.*, **29**(4), 293–308 (1975).
32. J. Coates, 'A Diamond-based Sampling Accessory for Solid and Liquid Sampling in Infrared Spectroscopy', *Am. Lab.*, **29**(8), C22–27 (1997).
33. N.J. Harrick, *Internal Reflection Spectroscopy*, Interscience, New York, 1967.
34. J. Fahrenfort, 'Attenuated Total Reflection, A New Principle for the Production of Useful Infrared Reflection Spectra of Organic Compounds', *Spectrochim Acta*, **17**, 698–709 (1961).
35. M.W. Urban, 'Surface and Interface Vibrational Spectroscopy – Relevance to Adhesion', *J. Adhes. Sci. Technol.*, **7**(1), 1–47 (1993).
36. T. Hirayama, M.W. Urban, 'Distribution of Melamine in Melamine Polyester Coatings – FTIR Spectroscopic Studies', *Prog. Org. Coat.*, **20**(1), 81–96 (1992).
37. E.M. Salazar-Rojas, M.W. Urban, 'Curing of Non-pigmented Alkyd Coatings Detected by In Situ Photoacoustic Fourier Transform Infrared Spectroscopy (PA FTIR)', *Prog. Org. Coat.*, **16**(4), 371–386 (1989).
38. B.W. Ludwig, M.W. Urban, 'The Effect of Polyurethane Curing Conditions on the Delamination of Pressure-sensitive Films', *J. Coat. Technol.*, **66**, 59–67 (1994).

39. B.W. Ludwig, M.W. Urban, 'Quantitative Determination of Isocyanate Concentration in Crosslinked Polyurethane Coatings', *J. Coat. Technol.*, **68**, 93–97 (1996).
40. A.M. Kaminski, M.W. Urban, 'Interfacial Studies of Crosslinked Polyurethanes. I. Quantitative and Structural Aspects of Crosslinking Near Film–Air and Film–Substrate Interfaces in Solvent-borne Polyurethanes', *J. Coat. Technol.*, **69**, 55–70 (1997).
41. A.M. Kaminski, M.W. Urban, 'Interfacial Studies of Crosslinked Polyurethanes. II. The Effect of Humidity on Waterborne Polyurethanes; A Spectroscopic Study', *J. Coat. Technol.*, **69**, 115–123 (1997).
42. M.W. Urban, K.W. Evanson, 'Factors Governing Surfactant Assembly at the Latex Film Interfaces – FTIR Spectroscopic Study', *Polym. Commun.*, **31**(7), 279–282 (1990).
43. K.W. Evanson, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. I. Surfactant–Copolymer Interactions', *J. Appl. Polym. Sci.*, **42**, 2287–2296 (1991).
44. K.W. Evanson, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. II. Surfactant–Copolymer Compatibility and Mobility of Surfactants', *J. Appl. Polym. Sci.*, **42**, 2297–2307 (1991).
45. B.J. Niu, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. IX. The Effect of Homopolymer and Copolymer Structures on Surfactant Mobility in Sty/BA Latices', *J. Appl. Polym. Sci.*, **56**, 377–385 (1995).
46. C.L. Zhao, Y. Holl, T. Pith, M. Lambla, 'FTIR-ATR Determination of the Distribution of Surfactants in Latex Films', *Colloid Polym. Sci.*, **265**(9), 823–829 (1987).
47. T.A. Thorstenson, K.W. Evanson, M.W. Urban, 'Forces Governing Mobility of Surfactants in Latices', American Chemical Society Abstracts, *Polym. Mater. Sci. Eng.*, **64**, 195–196 (1991).
48. K.W. Evanson, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. III. The Effects of Substrate Surface Tension and Elongation on Exudation of Surfactants', *J. Appl. Polym. Sci.*, **42**, 2309–2320 (1991).
49. T.A. Thorstenson, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. IV. The Effect of Surfactant Structure on the Copolymer–Surfactant Interactions', *J. Appl. Polym. Sci.*, **47**, 1381–1386 (1993).
50. T.A. Thorstenson, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. V. The Effects of Copolymer Composition on Surfactant Exudation', *J. Appl. Polym. Sci.*, **47**, 1387–1393 (1993).
51. B.J. Niu, M.W. Urban, 'Surface and Interfacial FTIR Spectroscopic Studies of Latexes. XIII. Surfactant Exudation in Silicone-modified Latex Films', *J. Appl. Polym. Sci.*, **60**, 389–397 (1996).
52. B.J. Niu, M.W. Urban, 'Surface and Interfacial Fourier Transform Infrared Spectroscopic Studies of Latexes. XVI. Quantitative Analysis of Surfactant in Multilayered Films', *J. Appl. Polym. Sci.*, **62**, 1903–1911 (1996).
53. D.J. Carlsson, D.M. Wiles, 'Photooxidation of Polypropylene Films. V. Origin of Preferential Surface Oxidation', *Macromolecules*, **4**(2), 173–184 (1971).
54. J.R. Webb, 'Reaction Profiles for Kinetic Studies of Inhomogeneous Reactions within Solid Organic Films', *J. Polym. Sci., Polym. Chem. Ed.*, **10**, 2335–2348 (1972).
55. R.G. Greenlet, R.R. Rahn, J.P. Schwartz, 'Effect of Index of Refraction on Position, Shape, and Intensity of Infrared Bands in Reflection–Absorption Spectra', *J. Catal.*, **23**(1), 42–50 (1971).
56. H.G. Tompkins, 'Some Reflections on Infrared Spectroscopy', *Appl. Spectrosc.*, **30**(4), 377–383 (1976).
57. S.A. Francis, A.H. Ellison, 'Infrared Spectra of Monolayers on Metal Mirrors', *J. Opt. Soc. Am.*, **49**(2), 131–138 (1959).
58. T. Buffeteau, B. Desbat, J.M. Turlet, 'Polarization Modulation FTIR Spectroscopy of Surfaces and Ultra-thin Films: Experimental Procedure and Quantitative Analysis', *Appl. Spectrosc.*, **45**, 380 (1991).
59. M.D. Schlotter, M.D. Porter, T.B. Bright, D.L. Allara, 'Formation and Structure of a Spontaneously Adsorbed Monolayer of Arachidic on Silver', *Chem. Phys. Lett.*, **132**, 93 (1986).
60. G. Kortum, H. Delfts, 'Reflexionsspektrometrische Messungen im Infraroten Spektralgebiet', *Spectrochim. Acta*, **20**, 405–413 (1964).
61. R.R. Willey, 'Fourier Transform Infrared Spectrophotometer for Transmittance and Diffuse Reflectance Measurements', *Appl. Spectrosc.*, **30**(6), 593–601 (1976).
62. M.P. Fuller, P.R. Griffiths, 'Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry', *Anal. Chem.*, **50**, 1906–1910 (1978).
63. G. Kortum, *Reflectance Spectroscopy Principles. Methods and Applications*, Springer-Verlag, Heidelberg, 1964.
64. K. Moradi, C. Depecker, J. Corset, 'Diffuse-reflectance Infrared Spectroscopy Experimental Study of Nonabsorbing Materials and Comparison with Theory', *Appl. Spectrosc.*, **48**(12), 1491–1497 (1994).
65. P. Kubelka, 'New Contributions to the Optics of Intensely Light-scattering Materials Part I', *J. Opt. Soc. Am.*, **38**(5), 448–457 (1948).
66. H.G. Hecht, 'Interpretation of Diffuse Reflectance Spectra', *J. Res. Natl. Bur. Stand., A*, **80**(4), 567–583 (1976).
67. A. Rosencwaig, 'Theoretical Aspects of Photoacoustic Spectroscopy', *J. Appl. Phys.*, **49**, 2905–2910 (1978).
68. A. Rosencwaig, *Photoacoustics and Photoacoustic Spectroscopy*, Wiley, New York, 1980.
69. A. Rosencwaig, A. Gersho, 'Photoacoustic Effect with Solids – Theoretical Treatment', *Science*, **190**, 556–557 (1975).
70. R.M. Dittmar, R.A. Palmer, R.O. Carter, 'Fourier Transform Photoacoustic Spectroscopy of Polymers', *App. Spectrosc. Rev.*, **29**, 171–231 (1994).

71. R.A. Palmer, R.M. Dittmar, 'Step-scan FTIR Photothermal Spectral Depth Profiling of Polymer Films', *Thin Solid Films*, **223**, 31–38 (1993).
72. M.W. Urban, J.L. Koenig, 'Depth-profiling Studies of Double-layer PVF2-on-PET Films by Fourier Transform Infrared Photoacoustic Spectroscopy', *Appl. Spectrosc.*, **40**(7), 994–998 (1986).
73. A. Rosencwaig, 'Photoacoustic Spectroscopy of Solids', *Opt. Commun.*, **7**(4), 305–308 (1973).
74. A. Rosencwaig, A.J. Gersho, 'Theory of Photoacoustic Effect with Solids', *Appl. Phys.*, **47**(1), 64–69 (1976).
75. M.W. Urban, 'Photoacoustic Fourier-transform Infrared Spectroscopy – A New Method for Characterization of Coatings', *J. Coat. Technol.*, **59**, 29–34 (1987).
76. A. Rosencwaig, 'Photoacoustic Spectroscopy', *Annu. Rev. Biophys. Bioeng.*, **9**, 31–54 (1984).
77. D.W. Vidrine, 'Photoacoustic Fourier Transform Infrared Spectroscopy of Solid Samples', *Appl. Spectrosc.*, **34**(3), 314–319 (1990).
78. E.M. Salazar-Rojas, M.W. Urban, 'Probing Organic–Inorganic Interactions and Curing Processes in Coatings by Photoacoustic Fourier Transform Infrared Spectroscopy', *J. Polym. Sci., Polym. Chem. Ed.*, **28**, 1593–1613 (1990).
79. J.B. Huang, M.W. Urban, 'Photoacoustic Fourier Transform Infrared Spectroscopy of Gases: A Nonequilibrium Approach', *J. Chem. Phys.*, **98**, 5259–5268 (1993).
80. J.B. Huang, M.W. Urban, 'Nonequilibrium Mass and Heat Transport in Polymers: Spectroscopic Approach', *J. Chem. Phys.*, **100**, 4509–4517 (1994).
81. R.A. Palmer, J.L. Chao, R.M. Dittmar, V.G. Gregoriou, S.E. Plunkett, 'Investigation of Time Dependent Phenomena by use of Step-scan FTIR', *Appl. Spectrosc.*, **47**, 1297–1310 (1993).
82. M.W. Urban, S.R. Gaboury, 'On the Sensitivity of Photoacoustic Fourier Transform Infrared Spectroscopy to Cross-linking Reactions', *Macromolecules*, **22**, 1486–1487 (1989).
83. S.R. Gaboury, M.W. Urban, 'Crosslinking of Polyester Styrene Detected by Thermal Photoacoustic Fourier-transform Infrared Spectroscopy', *Polym. Mater. Eng. Sci.*, **197**, 194–195 (1989).
84. E.G. Chatzi, M.W. Urban, H. Ishida, J.L. Koenig, 'Determination of the Accessibility of N–H Groups of Kevlar 49 Fibres by Photoacoustic FTIR Spectroscopy', *Polymer*, **27**, 1850–1854 (1986).
85. P.R. Griffiths, *Transform Techniques in Chemistry*, Plenum Press, New York, 122, 1978.
86. R.A. Palmer, 'Step-scan FTIR, A Versatile Tool for Time- and Phase-resolved Vibrational Spectroscopy', *Spectroscopy*, **8**, 26–33 (1993).
87. C.J. Manning, R.M. Dittmar, R.A. Palmer, L.J. Chao, 'Use of Step-scan FTIR to Obtain the Photoacoustic Photothermal Response Phase', *Infrared Phys.*, **33**(1), 53–62 (1992).
88. R.M. Dittmar, J.L. Chao, R.A. Palmer, 'Photoacoustic Depth Profiling of Polymer Laminates by Step-scan Fourier Transform Infrared Spectroscopy', *Appl. Spectrosc.*, **45**(7), 1104–1110 (1991).
89. L. Bertrand, 'Advantages of Phase Analysis in Fourier Transform Infrared Photoacoustic Spectroscopy', *Appl. Spectrosc.*, **42**(1), 134–138 (1988).
90. R.A. Palmer, E.Y. Jiang, J.L. Chao, 'Phase Analysis and Its Application in Step-scan FTIR Photoacoustic Depth Profiling', *Proc. SPIE*, **2089**, 250–251 (1993).
91. R.A. Palmer, E.Y. Jiang, 'Proceedings of the 8th International Topical Meeting on Photoacoustic and Photothermal Phenomena', *J. Phys. (Paris) IV, C and C7, Suppl. J. Phys.*, **4**, C7 (1994).
92. P. Poulet, J. Chambron, R.J. Unterreiner, 'Quantitative Photoacoustic Spectroscopy Applied to Thermally Thick Samples', *Appl. Phys.*, **51**(3), 1738–1742 (1980).
93. I. Noda, 'Generalized Two-dimensional Correlation Method Applicable to Infrared, Raman, and Other Types of Spectroscopy', *Appl. Spectrosc.*, **47**, 1329–1336 (1993).
94. I. Noda, 'Transient 2D IR Correlation Spectroscopy of the Photopolymerization of Acrylic and Epoxy Monomers', *Appl. Spectrosc.*, **47**(9), 1337–1344 (1993).
95. I. Noda, 'Two-dimensional Infrared Spectroscopy', *J. Am. Chem. Soc.*, **111**(21), 8116–8118 (1989).
96. B.O. Budevska, C.J. Manning, P.R. Griffiths, R.T. Boginski, 'Step-scan Fourier Transform Infrared Study on the Effect of Dynamic Strain on Isotactic Polypropylene', *Appl. Spectrosc.*, **47**(11), 1843–1851 (1993).
97. C. Marcott, A.E. Dowery, I. Noda, 'Instrumental Aspects of Dynamic Two-dimensional Infrared Spectroscopy', *Appl. Spectrosc.*, **47**, 1324–1328 (1993).
98. V.G. Gregoriou, I. Noda, A.E. Dowrey, C. Marcott, J.L. Chao, R.A. Palmer, 'Dynamic Rheo-optical Characterization of a Low-density Polyethylene/Perdeuterated High-density Polyethylene Blend by Two-dimensional Step-scan FTIR Spectroscopy', *J. Polym. Sci., Polym. Phys. Ed.*, **31**, 1769–1777 (1993).
99. A. Jayaraman, 'Diamond Anvil Cell and High Pressure Physical Investigations', *Rev. Mod. Phys.*, **55**(1), 65–108 (1983).
100. B.A. Weinstein, R. Zallen, 'Pressure Raman Effects in Covalent and Molecular Solids', *Top. Appl. Phys.*, **54**, 462–527 (1984).
101. B.Z. Jang, D.R. Uhlmann, J.B. Vander Sande, 'Crystalline Morphology of Polypropylene and Rubber-modified Polypropylene', *J. Appl. Polym. Sci.*, **29**(12), 4377–4393 (1984).
102. K.C. Dao, 'Rubber Phase Dispersion in Polypropylene', *Polymer*, **25**, 1527–1533 (1984).
103. J.L. Koenig, M.J.M. Tovar Rodriquez, 'Factor Analysis and Fourier Transform Infrared Spectra of Compatible Blends of Polyphenylene Oxides and Polystyrene', *Appl. Spectrosc.*, **35**, 543–548 (1981).

104. M.M. Coleman, J. Zarian, D.F. Varnell, P.C. Painter, 'A Fourier Transform Infrared Study of Polymer Blends. I. Poly(Vinylidene Fluoride)-Poly(Methyl Methacrylate) System', *J. Polym. Sci., Polym. Lett. Ed.*, **15**, 745–750 (1977).
105. D.L. Allara, 'Infrared Spectra of Polymer Mixtures – Problems Caused by Dispersion Effects in Difference Spectra', *Appl. Spectrosc.*, **33**, 358–361 (1979).
106. D.L. Allara, A. Baca, C.A. Pryde, 'Distortions of Band Shapes in External Reflection Infrared Spectra of Thin Polymer Films on Metal Substrates', *Macromolecules*, **11**(6), 1215–1220 (1978).
107. M.M. Coleman, J. Zarian, 'Fourier-transform Infrared Studies of Polymer Blends. II. Poly(ε-Caprolactone)–Poly(Vinyl Chloride) System', *J. Polym. Sci., Polym. Phys. Ed.*, **17**, 837–850 (1979).
108. M.M. Coleman, D.F. Varnell, 'Fourier-transform Infrared Studies of Polymer Blends. III. Poly(β-Propiolactone)–Poly(Vinyl Chloride) System', *J. Polym. Sci., Polym. Phys. Ed.*, **18**, 1403–1412 (1980).
109. M.M. Coleman, D.F. Varnell, J.P. Runt, 'Fourier Transform Infrared Studies of Polymer Blends: IV. Poly(ε-Caprolactone)–Poly(Bisphenol A Carbonate) System', *Contemporary Topics in Polymer Science*, ed. W.J. Bailey, Plenum Press, New York, 807–828, Vol. 4, 1982.
110. D.F. Varnell, J.P. Runt, M.M. Coleman, 'Fourier Transform Infrared Studies of Polymer Blends. VI. Further Observations on the Poly(Bisphenol A Carbonate)–Poly(ε-Caprolactone) System', *Macromolecules*, **14**, 1350–1356 (1981).
111. B.D. Pennington, J.C. Grunlan, M.W. Urban, 'Crosslinking of Epoxy-modified Phenol Novolac (EPN) Powder Coatings: Particle Size and Adhesion', *J. Coat. Technol.*, **71**, 135–142 (1999).
112. A.C. Ramamurthy, R.A. Ryntz, J.W. Holubka, B. Hartmann, 'Dynamic–Mechanical Characterization of Thermoplastic Olefins (TPO) over Eight Decades of Frequency', *Polym. Mater. Sci. Eng.*, **69**, 28–29 (1992).
113. R.W. Seymour, G.M. Estes, S.L. Cooper, 'Infrared Studies of Segmented Polyurethane Elastomers. I. Hydrogen Bonding', *Macromolecules*, **3**(5), 579–583 (1970).
114. H. Ishihara, L. Kimura, K. Saito, H.J. Ono, 'Infrared Studies on Segmented Polyurethane–Urea Elastomers', *Macromol. Sci., Phys.*, **10**(4), 591–618 (1974).
115. S.B. Lin, K.S. Hwang, S.Y. Tsay, S.L. Cooper, 'Segmental Orientation Studies of Polyether Polyurethane Block Copolymers with Different Hard Segment Lengths and Distributions', *Colloid Polym. Sci.*, **263**(2), 128–140 (1985).
116. B.J. Niu, M.W. Urban, 'The Effect of Homopolymer and Copolymer Structures on Surfactant Mobility in Sty/BA Latices – ATR FTIR Spectroscopic Study', *Polym. Mater. Sci. Eng.*, **69**, 196–197 (1993).
117. J. Schwartz, 'The Importance of Low Dynamic Surface-tension in Water-borne Coatings', *J. Coat. Technol.*, **64**, 65–74 (1992).
118. D.J. Walbridge, 'Self-stratifying Coatings – An Overview of a European Community Research Project', *Prog. Org. Coat.*, **28**(3), 155–159 (1996).
119. V.V. Verkholantsev, 'Heterophase and Self-stratifying Polymer Coatings', *Prog. Org. Coat.*, **26**(1), 31–52 (1995).
120. V.W. Srichatrapimuk, S.L. Cooper, 'Infrared Thermal Analysis of Polyurethane Block Polymers', *J. Macromol. Sci., Phys.*, **B15**(2), 267–311 (1978).
121. R.W. Seymour, G.M. Estes, S.L. Cooper, 'Infrared Studies of Segmented Polyurethane Elastomers. I. Hydrogen Bonding', *Macromolecules*, **3**, 579–583 (1970).
122. G. Oertel (ed.), *Polyurethane Handbook*, Hanser, Munich, 1985.
123. J.H. Saunders, K.C. Frisch, *Polyurethanes Chemistry and Technology*, Interscience, New York, Vol. 1, 1962.
124. C. Bouton, V. Arrondel, V. Rey, P. Sergot, J.L. Manguin, B. Jasse, L. Monnerie, 'Fourier Transform Infrared and Birefringence Studies of Orientation in Uniaxially-stretched Poly(2,6-Dimethyl 1,4-Phenylene Oxide) Atactic Polystyrene Blends', *Polymer*, **30**(8), 1414–1418 (1989).
125. S.G. Croll, 'Adhesion and Internal Strain in Polymeric Coatings', *Adhesion Aspects of Polymeric Coatings*, ed. K.L. Mittal, Plenum Press, New York, 107–129, 1983.
126. R.O. Carter, R.A. Dickie, J.W. Holubka, N.E. Lindsay, 'Infrared Studies of Interfacial Reactions of Dicyandiamide on Zinc', *Ind. Eng. Chem. Prod. Res. Dev.*, **29**(1), 48–51 (1989).
127. K.L. Mittal, 'Role of Interface in Adhesion Phenomena', *Polym. Eng. Sci.*, **17**(7), 467–473 (1977).
128. W.F. McDonald, 'Interfacial Photoacoustic FTIR Spectroscopic Studies of Polymers', PhD Thesis, North Dakota State University, 1991.
129. W.F. McDonald, H. Goettler, M.W. Urban, 'A Novel Approach to Photoacoustic FTIR Spectroscopy Rheophotoacoustic Measurements', *Appl. Spectrosc.*, **43**, 1387–1393 (1989).
130. M.W. Urban, H. Goettler, 'Rheo-photoacoustic FTIR Spectroscopic Method and Apparatus', *US Pat.*, 5 036 708, 1991.
131. S. Wu, *Polymer Interface and Adhesion*, Marcel Dekker, New York, 1982.
132. J. Comyn, 'Silane Coupling Agents', in *Structural Adhesives; Developments in Resins and Primers*, ed. A.J. Kinloch, Elsevier, New York, 269–312, 1986.
133. S. Yoshida, H. Ishida, 'A FTIR Reflection–Absorption Spectroscopy Study of an Epoxy Coating on Imidazole Treated Copper', *J. Adhes.*, **16**(3), 217–232 (1984).
134. C.S. Richard, G. Beranger, J. Lu, J.F. Flavenot, 'The Influences of Heat Treatments and Interdiffusion on the Adhesion of Plasma-sprayed NiCrAlY Coatings', *Surf. Coat. Technol.*, **82**, 99–109 (1996).
135. D.R. Paul, S. Newman (eds.), *Polymer Blends*, Academic Press, New York, 1978.
136. B. Arkles, 'Tailoring Surfaces with Silanes', *Chemtech*, **7**(12), 766–778 (1977).

137. E.P. Plueddemann, *Silane Coupling Agents, Plenum Press*, New York, 1982.
138. C.H. Chiang, J.L. Koenig, 'Crosslinking Mechanism of an Anhydride-cured Epoxy Resin as Studied by Fourier Transform Infrared Spectroscopy', *Polym. Compos.*, **1**, 88–102 (1980).
139. E.P. Plueddemann, P.G. Page, 'Methods for Improving the Performance of Silane Coupling Agents', *J. Adhes. Sci. Technol.*, **5**(10), 831–842 (1991).
140. S.R. Culler, H. Ishida, J.L. Koenig, 'The Silane Interphase of Composites – Effects of Process Conditions on Gamma-aminopropyltriethoxysilane', *Polym. Compos.*, **7**(4), 231–238 (1986).
141. A. Carton, 'A Novel Technique for the Study of Coupling Agent Phenomena', *Polym. Compos.*, **5**(4), 258–263 (1984).
142. A. Carton, 'The Crosslinking of Epoxy Resins at Interfaces. I. Germanium, Germanium Dioxide, Silane-treated Germanium', *J. Polym. Sci., Polym. Chem. Ed.*, **22**, 1495–1506 (1984).
143. N.H. Sung, A. Kaul, I. Chin, C.S.P. Sung, 'Mechanistic Studies of Adhesion Promotion by Gamma-aminopropyltriethoxysilane in Alpha-Al_2O_3-Polyethylene Joint', *Polym. Eng. Sci.*, **22**(10), 637–644 (1982).
144. C.H. Chiang, H. Ishida, J.L. Koenig, 'The Structure of Gamma-aminopropyltriethoxysilane on Glass Surfaces', *J. Colloid. Interface Sci.*, **74**(2), 396–404 (1980).
145. J.L. Koenig, 'Fourier Transform Infrared Spectroscopy of Polymers', *Adv. Poly. Sci.*, **54**, 87–154 (1983).
146. M.E. Schrader, I. Lerner, F.J. D'Oria, 'Radioisotope Study of Coupling Agents in Reinforced Plastics', *Mod. Plast.*, **45**, 195–282 (1967).
147. H. Ishida, J.L. Koenig, 'Fourier-transform Infrared Spectroscopic Study of Structure of Silane Coupling Agent on E-Glass Fiber', *J. Colloid. Interface Sci.*, **64**(3), 565–576 (1978).
148. H. Ishida, S. Naviroj, S.K. Tripathy, J.J. Fitzgerald, J.L. Koenig, 'The Structure of an Aminosilane Coupling Agent in Aqueous Solutions and Partially Cured Solids', *J. Polym. Sci., Polym. Phys.*, **20**(4), 701–718 (1982).
149. M.W. Urban, J.L. Koenig, 'Enhancement of the Surface Modes in Photoacoustic FTIR Spectroscopy Using a Highly Polarizable Inert Gas', *Appl. Spectrosc.*, **39**(6), 1051–1056 (1985).
150. E.G. Chatzi, M.W. Urban, J.L. Koenig, 'Characterization of Kevlar Fiber Surfaces Using a Newly Developed Infrared Photoacoustic Technique', *Makromol Chem., Makromol. Symp.*, **5**, 99–104 (1986).
151. S.L. Tidrick, J.L. Koenig, 'Photoacoustic FTIR Analysis of Surface Modified Kevlar-49 Fibers', *J. Adhes.*, **29**, 43–51 (1989).
152. W.F. McDonald, M.W. Urban, 'Chemical Structures at the Nextel Fiber Polyimide Matrix Interface Detected by Photoacoustic FTIR Spectroscopy', *Composites*, **22**(4), 307–318 (1991).
153. A.M. Tiefenthaler, M.W. Urban, 'Stability of Nextel Fibers in Aqueous Environments Detected by Photoacoustic Fourier-transform Infrared Spectroscopy', *Composites*, **20**(6), 585–588 (1989).
154. J.R. Simms, C.Q. Yang, 'Infrared Spectroscopy Studies of the Petroleum Pitch Carbon Fiber. II. The Distribution of the Oxidation Products Between the Surface and the Bulk', *Carbon*, **32**(4), 621–626 (1994).
155. E.M. Murty, T.W. Yehl, 'Adaptation of Photoacoustic Fourier Transform Infrared Spectroscopy for Studying the Thermal Oxidation of Nylon 66 at 150 °C Correlated to Mechanical Properties', *Polym. Eng. Sci.*, **30**(24), 1595–1598 (1990).
156. S.S. Voyutskii, 'Some Comments on Interfacial Contact and Bonding in Autohesion', *J. Adhes.*, **3**(1), 69–79 (1971).
157. E. Jabbari, N.A. Peppas, 'Use of ATR-FTIR to Study Interdiffusion in Polystyrene and Poly(Vinyl Methyl Ether)', *Macromolecules*, **26**, 2175–2186 (1993).
158. B.B. Sauer, D.J. Walsh, 'Use of Neutron Reflection and Spectroscopic Ellipsometry for the study of the Interface between Miscible Polymer Films', *Macromolecules*, **24**(22), 5948–5955 (1991).
159. L.E. Cook, S.Q. Luo, J.F. McClelland, 'Fourier Transform Infrared Photoacoustic Spectroscopy of Polymers Adsorbed from Solution by Gamma-iron Oxide', *Appl. Spectrosc.*, **45**(1), 124–126 (1991).
160. Y. Nishikawa, K. Kimura, A. Matsuda, T. Kenpo, 'Surface Characterization of Gamma-iron Oxides for Magnetic Memory Media Using Fourier Transform Infrared Photoacoustic Spectroscopy', *Appl. Spectrosc.*, **46**(11), 1695–1698 (1992).
161. D.A. Saucy, G.E. Cabaniss, R.W. Linton, 'Surface Reactivities of Polynuclear Aromatic Adsorbates on Alumina and Silica Particles Using Infrared Photoacoustic Spectroscopy', *Anal. Chem.*, **57**, 876–879 (1985).
162. J.P. Folkers, P.E. Laibinis, G.M. Whitesides, 'Self-assembled Monolayers of Alkanethiols on Gold: The Adsorption and Wetting Properties of Monolayers Derived from Two Components with Alkane Chains of Different Lengths', *Contact Angle, Wettability and Adhesion: Festschrift in Honor of Professor Robert J. Good*, ed. K.L. Mittal, VSP, Netherlands, 547–560, 1993.
163. C.E. Chidsey, D.N. Loiacono, 'Chemical Functionality in Self-assembled Monolayers: Structural and Electrochemical Properties', *Langmuir*, **6**(3), 682–691 (1990).
164. J. Greenblatt, C.J. Araps, H.R. Anderson, 'Aminositane–Polyimide Interactions and their Implications in Adhesion', *Polyimides*, ed. K.L. Mittal, Plenum Press, New York, 573–588, Vol. 1, 1984.
165. M.C. Burrell, P.J. Codella, J.J. Chera, 'Interfacial Reactions at Copper Surfaces Coated with Polymer Films', *J. Vac. Sci. Technol.*, **A7**(3), 1778–1783 (1989).
166. J. Jang, J.H. Lee, 'Effect of Imidization Temperature on the Adhesion of Polyimide on Aluminum', *J. Appl. Polym. Sci.*, **62**, 199–205 (1996).

167. O. Flament, J. Russat, E. Druet, 'X-ray Photoelectron Spectroscopy Study of the Interface Between Aluminum and Spin-coated Polyamic Acid', *J. Adhes. Sci. Technol.*, **4**(2), 109–117 (1990).
168. W.F. McDonald, M.W. Urban, 'A New Method to Monitor Adhesion and Interfacial Failure Between Polymers – Rheophotoacoustic FTIR Spectroscopy', *J. Adhes. Sci. Technol.*, **4**(9), 751–764 (1990).
169. H. Yasuda, *Plasma Polymerization*, Academic Press, Orlando, FL, 1985.
170. S.B. Gaboury, M.W. Urban, 'Spectroscopic Evidence for Si–H Formation During Microwave Plasma Modification of Poly(Dimethylsiloxane) Elastomer Surfaces', *Polym. Commun.*, **32**(13), 390–392 (1991).
171. D.A. Waldman, Y.L. Zou, A.N. Netravali, 'Ethylene Ammonia Plasma Polymer Deposition for Controlled Adhesion of Graphite Fibers to PEEK. I. Characterization of Plasma Formed Polymers', *J. Adhes. Sci. Technol.*, **9**(11), 1475–1503 (1995).
172. A. Nihistrand, T. Hjertberg, K. Johansson, 'Plasma Polymerization of Allyl Alcohol on Polyolefin-based Materials: Characterization and Adhesion Properties', *J. Adhes. Sci. Technol.*, **10**(2), 123–150 (1996).
173. C.R. Savage, R.B. Timmons, J.W. Lin, 'Spectroscopy Characterization of Films Obtained in Pulsed Radiofrequency Plasma Discharges of Fluorocarbon Monomers', *Adv. Chem. Ser.*, **236**, 745–768 (1993).
174. V.G. Gregoriou, M. Daun, M.W. Schauer, J.L. Chao, R.A. Palmer, 'Modification of a Research-grade FTIR Spectrometer for Optional Step-scan Operation', *Appl. Spectrosc.*, **47**, 1311–1316 (1993).
175. T. Miyazawa, K. Fukushima, Y.J. Ideguchi, 'Far Infrared Spectra and their Vibrational Assignments of Isotactic Polypropylene', *J. Polym. Sci., Polym. Lett.*, **1**, 385–387 (1963).
176. H. Tadokoro, M. Kobayashi, M. Ukita, K. Yasufuku, S. Murahashi, 'Normal Vibrations of the Polymer Molecules of Helical Conformation. V. Isotactic Polypropylene and Its Deuteroderivatives', *J. Chem. Phys.*, **42**(4), 1432–1449 (1965).
177. X.M. Xie, Y. Chen, Z.M. Zhang, A. Tanioka, M. Matsuoka, K. Takemura, 'Controls of Gradient Morphology and Surface Properties of Polymer Blends', *Macromolecules*, **32**(13), 4424–4429 (1999).
178. P. Ning, T.M. Ko, 'Influence of Miscibility on the Morphology and Properties of Polycarbonate/(Maleic Anhydride)–Grafted-polypropylene Blend', *Polym. Eng. Sci.*, **37**(7), 1226–1237 (1997).
179. D.R. Fitchmun, Z. Mencik, 'Morphology of Injection-molded Polypropylene', *J. Polym. Sci., Polym. Phys.*, **11**, 951–971 (1973).
180. J. Bowman, N. Harris, M.J. Bevis, 'Investigation of Relationships between Processing Conditions, Microstructure, and Mechanical Properties of an Injection Molded Semi-crystalline Thermoplastic', *J. Mater. Sci.*, **10**(1), 63–76 (1975).
181. R.A. Ryntz, 'The Effects of Thermoplastic Polyolefin (TPO) Morphology on Subsequent Paintability and Thermal Shock Performance', *Prog. Org. Coat.*, **27**, 241–254 (1996).
182. D.R. Fitchmun, S. Newman, R. Wiggle, 'Electroplating on Crystalline Polypropylene. I. Compression Molding and Adhesion', *J. Appl. Polym. Sci.*, **14**(10), 2441–2455 (1970).
183. K.M. Vaeth, K.F. Jensen, 'Chemical Vapor Deposition of Thin Polymer Films Used in Polymer-based Light Emitting Diodes', *Adv. Mater.*, **9**(6), 490–500 (1997).
184. J.O. Lephardt, G. Vilcins, 'Application of the Rapid Scan Fourier Transform Infrared Spectrometer to Kinetics', *Appl. Spectrosc.*, **29**(3), 221–225 (1975).
185. H. Siesler, 'Fourier Transform Infrared (FTIR) Spectroscopy in Polymer Research', *J. Mol. Struct.*, **59**, 15–37 (1980).
186. H.W. Siesler, 'Characterization of Deformation Phenomena in Polymers by Rapid-scanning Fourier Transform IR (FTIR) Spectroscopy and Mechanical Measurements. II. Effect of Stress-relaxation on the Reversible, Stress-induced Conformational Transition in Poly(Butylene Terephthalate)', *Makromol. Chem., Makro. Chem. Phys.*, **180**(9), 2261–2263 (1979).
187. R.W. Snyder, S.J. Fuerniss, 'ATR/IR Spectroscopic Method for Following Photopolymer Curing', *Appl. Spectrosc.*, **46**(7), 1113–1116 (1992).
188. T.M. Don, J.P. Bell, 'Fourier Transform Infrared Analysis of Polycarbonate/Epoxy Mixtures Cured with an Aromatic Amine', *J. Appl. Polym. Sci.*, **69**(12), 2395–2407 (1998).
189. N.S. Allen, S.J. Hardy, A.F. Jacobine, D.M. Glaser, B. Yang, D. Wolf, F. Catalina, S. Navaratnam, B.J. Parsons, 'Photochemistry and Photopolymerization Activity of Perester Derivatives of Benzophenone', *J. Appl. Polym. Sci.*, **42**, 1169–1178 (1991).
190. D.B. Yang, D. Wolf, T. Wakamatsu, M. Holmes, 'Characterization of Cure Profiles of Anaerobic Adhesives by Real-time FTIR Spectroscopy. II. Surface Activation', *J. Adhes. Sci. Technol.*, **9**(10), 1369–1379 (1995).
191. B.G. Wall, J.L. Koenig, 'Studying the Curing Kinetics of a Diacrylate by Using Infrared Spectroscopy', *Appl. Spectrosc.*, **51**(10), 1453–1459 (1997).
192. P.B. Jacobs, P.C. Yu, '2-Component Waterborne Polyurethane Coatings', *J. Coat. Technol.*, **65**, 45–50 (1993).
193. W.J. Blank, V.J. Tramontano, 'Properties of Crosslinked Polyurethane Dispersions', *Prog. Org. Coat.*, **27**, 1–15 (1996).
194. V.J. Tramontano, W.J. Blank, 'Cross-linking of Waterborne Polyurethane Dispersions', *J. Coat. Technol.*, **67**, 89–99 (1995).
195. J.W. Rosthauser, K. Nachtkamp, *Advances in Urethane Science Technology*, Technomic, Westport, CT, 121–162, Vol. 10, 1987.
196. C.R. Hegedus, A.G. Gilicinski, R.J. Haney, 'Film Formation Mechanism of Two Component Waterborne

Polyurethane Coatings', *J. Coat. Technol.*, **68**(852), 51–61 (1996).
197. Z.M. Wang, D.B. Gao, J.W. Yang, Y.L. Chen, 'Synthesis and Characterization of UV-curable Waterborne Polyurethane–Acrylate Ionomers for Coatings', *J. Appl. Polym. Sci.*, **73**(14), 2869–2876 (1999).
198. D.L. Lundberg, R.G. Brown, J.E. Glass, R.R. Eley, 'Synthesis Characterization, and Solution Rheology of Model Hydrophobically-modified, Water-soluble Ethoxylated Urethanes', *Langmuir*, **10**(9), 3027–3034 (1994).
199. C.L. Allison, 'Water-borne Polyurethanes: Structure–Property Relations', M.S. thesis, North Dakota State University, 1998.
200. L. Hill, K. Kozlowski, 'Cross-link Density of High-solids MF-cured Coatings', *J. Coat. Technol.*, **59**, 63–71 (1987).
201. T. Hamada, H. Kanai, T. Koike, M. Fuda, 'FTIR Study of Melamine Enrichment in the Surface Region of Polyester/Melamine Film', *Prog. Org. Coat.*, **30**(4), 271–278 (1997).
202. T. Hirotsu, P. Nugroho, 'Surface Coating Protection Against Oxidative Plasma Etching of Polypropylene', *J. Appl. Polym. Sci.*, **66**, 1049–1057 (1997).
203. J. Mihaly, J. Kristof, J. Mink, L. Nanni, D. Patracchini, A. De Battisti, 'Study of ZrO_2 Coatings by Thermoanalytical and Fourier Spectroscopic Methods', *Mikrochim. Acta, Suppl.*, **14**, 617–619 (1997).
204. F.J. Boerio, L.H. Schoenien, J.E. Greivenkamp, 'Adsorption of γ-Aminopropyltriethoxysilane onto Bulk Iron from Aqueous Solution', *J. Appl. Polym. Sci.*, **22**(1), 203–213 (1978).
205. F.J. Boerio, S.L. Chen, 'Reflection–Absorption Infrared Spectroscopy and Ellipsometry of Epoxy Films on Metals', *Appl. Spectrosc.*, **33**, 121–126 (1979).
206. R.G. Greenler, 'Infrared Study of Adsorbed Molecules on Metal Surfaces by Reflection Techniques', *J. Chem. Phys.*, **44**(1), 310–315 (1966).
207. D.L. Allara, 'The Study of Thin Polymer Films on Metal Surfaces Using Reflection–Absorption Spectroscopy', *Characterization of Metal and Polymer Surfaces*, ed. L.H. Lee, Pergamon Press, New York, 193–206, Vol. 2, 1972.
208. J.D. Webb, P. Schissel, A.W. Czanderna, A.R. Chughtai, D.M. Smith, 'An Experimental Approach for Studies of the Photochemical, Thermal, and Environmental Degradation of Polymer Coated Metals', *Appl. Spectrosc.*, **35**, 598–603 (1981).
209. R.M. Paroli, K.C. Cole, A.H. Delgado, 'Evaluating the Weatherability of Polyurethane Sealants', *Multidimensional Spectroscopy of Polymers*, ed. M.W. Urban, ACS Symposium Series, American Chemical Society, Washington, DC, Chapter 7, 117–136, Vol. 598, 1995.
210. S. Yoshida, H. Ishida, 'An Investigation of the Thermal Stability of Undecylimidazole on Copper by FTIR Reflection–Absorption Spectroscopy', *Appl. Surf. Sci.*, **89**(1), 39–47 (1995).
211. M.L. McKelvy, T.R. Britt, B.L. Davis, J.K. Gillie, L.A. Lentz, A. Leugers, R.A. Nyquist, C.L. Putzig, 'Infrared Spectroscopy', *Anal. Chem.*, **68**, 93R–160R (1996).
212. P.J. Bremer, G. Geesey, 'Determination of the Feasibility of Using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy to Evaluate Thermal Ageing of Enamel-coated Magnet Wire', *J. Mater. Sci.*, **32**(1), 141–146 (1997).
213. N.S. Allen, C.J. Regan, R. McIntyre, B.W. Johnson, W. Dunk, 'The Photooxidative Degradation and Stabilization of Water-borne Acrylic Coatings Systems', *Macromol. Symp.*, **115**, 1–26 (1996).
214. K. Krishnan, 'Resolution of Complex Band Shapes by Infrared Polarization Measurements', *Appl. Spectrosc.*, **32**, 549–553 (1978).
215. A. Garton, 'Trans–Gauche Rotational Isomers in Partially Oriented Poly(Ethylene Terephthalate) Fibers', *J. Appl. Polym. Sci.*, **25**, 1505–1507 (1980).
216. C.S.P. Sung, 'A Modified Technique for Measurement of Orientation from Polymer Surfaces by Attenuated Total Reflection Infrared Dichroism', *Macromolecules*, **14**, 591–594 (1981).
217. R.S. Stein, S. Onogi, D.A. Keedy, 'The Dynamic Birefringence of High Polymers', *J. Polym. Sci. B. Polym. Phys.*, **34**(9), 1507–1527 (1996).
218. R.S. Stein, 'A Procedure for the Accurate Measurement of Infrared Dichroism of Oriented Film', *J. Appl. Polym. Sci.*, **5**(13), 96–99 (1961).
219. R. Zbinden, *Infrared Spectroscopy of High Polymers*, Academic Press, New York, 1964.
220. N.J. Everall, A. Bibby, 'Improvements in the Use of Attenuated Total Reflection Fourier Transform Infrared Dichroism for Measuring Surface Orientation in Polymers', *Appl. Spectrosc.*, **51**, 1083–1091 (1997).
221. T. Nakano, T. Yokoyama, H. Toriumi, 'One- and Two-dimensional Infrared Time-resolved Spectroscopy Using a Step-scan FTIR Spectrometer: Application to the Study of Liquid Crystal Reorientation Dynamics', *Appl. Spectrosc.*, **47**, 1354–1366 (1993).
222. D.B. Yang, T. Wakamatsu, 'Polarized Grazing Angle FTIR Study of Molecular Orientation and Bonding of Thioglycolates at the Metal Interface', *Surf. Interface Anal.*, **24**(12), 803–810 (1996).

Infrared Spectroscopy in Analysis of Polymers and Rubbers

John M. Chalmers
University of Nottingham, Nottingham, UK

1 Introduction 1
2 Sample Types 2
3 Sampling Techniques 2
3.1 Transmission 2
3.2 Reflection Methods 6
3.3 Diffuse Reflectance 7
3.4 Emission 8
3.5 Photoacoustic 10
4 Sample Presentation 10
5 Qualitative Analysis 10
6 Quantitative Analysis 11
7 Polymer Spectra 12
7.1 Isomerism, Conformation, Tacticity, Regularity, Configuration and Crystallinity 12
7.2 Infrared Dichroism and Molecular Orientation 14
7.3 End Groups, Branching 17
7.4 Hydrogen Bonding 19
7.5 Blend, Miscibility Studies 20
7.6 Copolymer Composition 21
7.7 Polymerization and Cure Studies 23
7.8 Degradation and Oxidation 26
8 Chromatography and Fourier Transform Infrared 29
9 Pyrolysis and Infrared 29
10 Evolved Gas Analysis, Thermogravimetric Analysis and Fourier Transform Infrared 32
11 Optical Properties and Constants 33
12 Two-dimensional Infrared Spectroscopy 33
13 Process Measurements 35
14 Microspectroscopy 35
14.1 Fourier Transform Infrared Microscopy: Sample Presentation 35
14.2 Fourier Transform Infrared Microscopy Applications 36
Acknowledgments 44
Abbreviations and Acronyms 44
Related Articles 45
References 45

The infrared spectrum of a polymer or rubber is a profile of its absorption characteristics plotted against infrared wavenumber (or frequency). It is generated when infrared radiation interacts with the molecular moieties that constitute the polymer or rubber material. It is a distinctive property (a ''fingerprint'') of the polymer or rubber sample in the form, manner and environment in which it is being examined. It is based on the absorption of infrared radiation at frequencies that match those of the normal modes of vibration within the macromolecule. These absorption features are characteristic of the molecular configuration, sequencing and conformation, and state of order. Absorptions by internal vibrations predominate in the mid-infrared region (ca. 4000–400 cm^{-1}); they involve a few selected atoms in a molecular (functional) grouping, which is a sub-set of those constituting the macromolecule. External vibrations, such as lattice vibrations, which involve segments of macromolecules in crystalline regions, tend to occur at low wavenumbers (<400 cm^{-1}). The intensity of an absorption band is, to a first approximation, proportional to the number density of vibrating species giving rise to that band. The intensity of an absorption band is related to the dipole moment change associated with the molecular vibration; if it is large then the band intensity will be high. Infrared spectroscopy is highly specific. For some quantitative analyses, it can be very precise and sensitive (<0.1%). It is complemented with Raman spectroscopy, for which the selection rules are different and relate band intensity to the change of polarizability occurring during a molecular vibration. Infrared spectroscopy can be used to identify and analyse the molecular structure, composition, order and morphology, both qualitatively and quantitatively, of polymers, copolymers and rubbers and their products. However, for heavily filled and plasticized formulations the spectral interferences may be too high to allow direct or adequate interpretation of a spectrum, and separation of the components will be necessary for unambiguous identifications.

1 INTRODUCTION

Infrared spectroscopy has a long tradition and remains one of the most widely used spectroscopic techniques in the analysis and characterization of polymers and rubbers, and their products. Sampling techniques are available to examine them in their various forms; they cover both macro- and micro-analysis methods for both bulk and surface layer measurements. It is used to both

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

identify materials and probe the molecular microstructure and morphology of complex macromolecules. There are many reviews and books on the subjects of polymer study and analysis. They cover both vibrational spectroscopy fundamentals through to experimental practice.[1–13]

This article begins with a short discussion on the range of sample forms and properties. This is followed by concise descriptions of sampling techniques. Discussions of qualitative and quantitative analysis and the peculiarities associated with polymer infrared spectra are then followed by more specific applications sections concerned with chemical, physical, morphological, performance and optical property study and characterization. A short consideration is given to industrial process analysis, before the article text closes with an account of the use of Fourier transform infrared (FTIR) microspectroscopy in the analysis of polymers and rubbers. Since its introduction in the 1980s, FTIR microscopy has made major contributions to advancing the analysis and study of macromolecules. This article does not consider near-infrared analysis (NIRA), since these are covered in article **Near-infrared Spectroscopy of Polymers and Rubbers** in this Encyclopedia, nor does it discuss Raman analyses of polymers and rubbers.

The application references incorporated within this article are not exhaustive, but selected to exemplify a diversity of analyses, studies and polymer types; reference lists within these publications will lead those with particular interests to more comprehensive information.

2 SAMPLE TYPES

As produced, commercial polymers may, for example, be entirely aliphatic or aromatic, or contain both entities. They extend from high through to low molecular weight grades and can be in the form of powders, granules, waxes, dispersions or lattices. Solid forms may be tacky or rigid, ductile or brittle; they may be low-melting (as with some hot-melt adhesives, such as some ethylene/vinyl acetate (EVA) copolymers) or melt at temperatures in excess of 300 °C (such as poly(aryl ether ether ketone), (PEEK), an engineering thermoplastic). A polymer may have different polymorphic or stereoregular forms; some polymers may be amorphous, others semi-crystalline or crystalline. Copolymers may be sequenced in differing ways. Polymer end group balances and functionality may differ throughout a commercial grade range or between manufacturers. The polymer or rubber sample may have been processed into a fiber, film or foam, or molded into an article such as a bottle. It may have been formulated or filled with a range of additives, to improve processability, stability, impact strength, or aesthetic appearance. It may have been derived from a thermoset formulation, or be cross-linked. It may be in the form of a composite with inorganic fibers, such as glass or carbon. It may be part of a laminated structure or a surface coating.

A wide and diverse range of sample preparation procedures and sampling techniques with supporting accessories are therefore available, which enable polymers, polymer composites and rubbers, and their formulated and fabricated products, in their many forms, uses and applications, to be analysed and investigated by using infrared spectroscopy.

3 SAMPLING TECHNIQUES

The infrared spectroscopic analysis of organic polymer and rubber samples is concerned firstly with handling, preparing and presenting materials of many different forms and with many different physical properties in an appropriate way for spectroscopic examination. The infrared spectrum recorded is then used for characterizing their chemical structure or measuring their composition, sometimes investigating their morphology, occasionally determining their optical properties.

Since the majority of experiments are concerned in some way with the absorption characteristics of a polymer or rubber, almost the whole gamut of infrared spectroscopy condensed phase sampling techniques are used to record infrared spectra of polymers and rubbers for analytical purposes. The choice of sample presentation method will depend on whether the measurement is to be made directly from a bulk sample, or from a formulated product, or from a fabricated end-use article, or from material which has been coated onto a substrate. Some approaches will be peculiar to kinetic studies and process investigations. The choice of method will likely not only depend on the physical nature of the substance, but also be governed by the output requirement of the measurement; see section 4 following. The fine detail of a polymer spectrum may well reflect the method of sample preparation.

3.1 Transmission

The optimum (or standard) macro-sample method for a high spectral contrast, high sensitivity, ambient temperature, full range mid-infrared spectrum of a bulk polymer for qualitative and many quantitative or semi-quantitative purposes is a transmission spectrum recorded from a thin, continuous, homogeneous, uniform (interference-fringe free) thickness film. The optimal thickness of the film will depend on the polymer and measurement purpose. To observe the positions of all the peak maxima

in the infrared spectrum of a strongly infrared absorbing polymer, such as a polysiloxane, perfluoro-polymer or a polyaromatic, it is necessary to prepare a film of a thickness of much less than 10 μm. A thin film will also be required to observe precisely the positions of the νCH bands (near 2900 cm^{-1}) for an aliphatic hydrocarbon polymer. However, such a spectrum will probably only be useful for generic fingerprinting. Better qualitative classification will likely be achieved for these aliphatic materials from spectra recorded from film specimens 0.1–0.2 mm thickness, to increase the intensity of the weaker fingerprint bands. Optimal thickness specimens for quantitative purposes will depend on the absorption coefficients of the analyte band(s) to be measured. They may vary from 10 μm or less to several millimeters or more (if a weak or overtone band is used, or the band is characteristic of a low concentration species). In all cases, the film's surfaces must be of a size that is sufficient to completely cover the cross-section of the infrared beam focus in the spectrometer sample compartment (or wherever the sample is placed). If it is smaller, then it must be mounted in a holder that will exclude radiation that has not passed through the specimen reaching the detector, since stray light will cause spectral contrast to deteriorate significantly and have disastrous consequences for quantitative determinations.

3.1.1 Hot Pressed or Melt Cast Film

Thin films from some low-melting, thermally stable polymers may be prepared as a sandwich between two hard infrared transparent windows. For example, a few milligrams of the sample is placed centrally between a pair of appropriately sized polished KBr plates; the assembly is then heated gently until the polymer sample softens or just melts. This may be accomplished on a gradient temperature hot-plate (e.g. Kofler heated bar) or under a heating lamp. When softened or molten the polymer is made to flow evenly into a thin uniform thickness film between the plates by applying gentle pressure on to the top surface of the uppermost KBr window. A safe practice must be instigated to achieve this in order to avoid burning one's fingers; use of a cork to apply the pressure has been recommended.[12] The assembly is then allowed to cool gradually to room temperature, during which time the film sample preparation solidifies. This procedure can be particularly suitable for amorphous materials such as waxes, tarry solids and hot-melt adhesives. Since essentially only hand pressure is applied, rubbery materials prepared in this way may have a tendency to retract on solidification; to prevent this the assembly may therefore need to be clamped at the edges.

For higher-melting polymers, particularly thermoplastics, hot compression molding is a very convenient and effective means of preparing a free-standing film appropriate to an infrared transmission examination, which can be especially well-suited for quantitative analysis, although experience has shown that large surface area film samples with a thickness of 10 μm or less will likely be difficult to achieve routinely by this method.

Various examples of press apparatus are available commercially for hot, compression molding of polymers and rubbers.[14–16] These include small ''constant thickness'' film-making kits, which incorporate water-cooled, heated and temperature controlled platens, for use with a manually operated, hydraulic press, such as that used for preparing alkali halide disks. Larger, hand-pumped, laboratory hydraulic presses are, for example, capable of applying loads of up to 20 tons (~20 000 kg) on a ram of 4 inches (10.16 cm) diameter, operating on square 8-inch (20-cm) platens. These platens may be electrically heated to temperatures in excess of 350 °C; they may be cooled rapidly with water, which flows through them; additionally, some presses allow for programmable heating and cooling cycles, to facilitate the preparation of film specimens with specific thermal histories. In these larger presses, film specimens may be prepared by pressing samples between metal plates (press-plates) of a few millimeters thickness and of the same dimensions as the press platens.

Uniform thickness polymer films with smooth, shiny surfaces on both sides, particularly thinner ones (<0.2 mm) will likely give rise to a transmission spectrum that exhibits interference (channel) fringes. That is, superimposed on the polymer's infrared absorption spectrum will be a sinusoidal wave, which will be most evident in regions of low absorbance. This will act to the detriment of interpretative observations and almost certainly negate high precision quantitative measurements. These interference fringe contributions may, in some circumstances, be minimized or eliminated by gentle rubbing of the polymer film surfaces with wire wool or sandpaper. Press-plates for use with the larger laboratory presses may be cut, for example, from polished stainless steel sheets. Where these are intended for preparing polymer film specimens, then their surfaces may be pre-conditioned in such a way as to induce into the film's surface a finish which nullifies the conditions for interference fringe production. This can be achieved, for example, by grit-blasting (20–30 mesh) the sample contact surfaces of the press-plates, which are then spray over-coated with a very thin layer of a poly(tetrafluoroethylene) (PTFE). The surface roughness introduced is then generally sufficient to negate interference fringe pattern occurrence, but without detracting from uniformity of film thickness. The PTFE coating, which acts as a mold-release agent, *must not* be applied if it is intended to heat the press-plates to temperatures much in excess of 300 °C, since the

decomposition products of PTFE are toxic, some of them highly so. Films from thermoplastics, which melt near to or above 300 °C, can be prepared as a sandwich between aluminum foil, which can be peeled or dissolved away (in NaOH solution) to reveal the polymer film. Film specimens from rubbery and brittle samples may be prepared and subsequently examined as a sandwich between thin sheets of AgCl.

Control of sample thickness during molding may be achieved in one of several ways. Thick specimens (> ca. 1 mm) are most likely best prepared using mold templates, appropriately sandwiched between the press-plates. Thinner films may be molded from polymer material sandwiched between the coated surface of the press-plates. Feeler gauges may be inserted beside the polymer specimen to help define film thickness. In all cases, final film thickness will be dependent to varying extents not only on press temperature, compression pressure and spacer thickness, but also on the amount of material sandwiched between the press-plates. Although perhaps not always possible for exploratory type research samples, samples should ideally be dry and essentially free from monomer, catalyst and solvent residues, otherwise hydrolysis or additional reaction may take place, or bubbles may form in the prepared film. The optimum temperature, pressure and time at which to compression mold a film from a particular polymer will need to be found empirically,[15] although the temperature will likely be close to its melting point. Too high a temperature or for too long may lead to thermal degradation and consequent discoloration, or give rise to bubbles in the specimen; too low a temperature or insufficient pressure may give rise to a sample that has not coalesced fully. Controlled heating and cooling rates may be used to induce certain morphologies into the prepared compression molded film, e.g. a rapid quench (into ice) from the melt will likely "freeze" the specimen in an amorphous state, whereas slow cooling (annealing) may lead to a more crystalline morphology.[15]

3.1.2 Microtomed Film Sections

In some circumstances, thin sections appropriate to a transmission measurement may be microtomed from samples such as polymer moldings. However, this practice is more usually employed and commonplace in the preparation of specimens for FTIR microscopy, see section 14.1.

3.1.3 Cold Pressed Film

PTFE and many of its copolymer powders will coalesce under pressure at ambient temperature. A translucent PTFE disk may be prepared by pressing neat powder in an alkali halide disk apparatus. Some linear polyethylenes may also compact at room temperature into forms suitable for examination; indeed, finely powdered polyethylene has been used as a substitute for an alkali halide as the matrix material for dispersion-disk type preparations for the far-infrared region. Compression by high pressure, such as that achieved in a diamond anvil cell,[16] usually yields a thin film suitable for producing a high spectral contrast fingerprint spectrum from a very small amount of an unfilled polymer or rubber sample. As with microtoming, this approach is most commonly used in the preparation of specimens for measurements by FTIR microscopy, see section 14.1.

3.1.4 Capillary Layer Films

For liquid polymers and some low molecular weight amorphous polymers or polymer fractions, bubble-free films suitable for infrared examination may be easily prepared as a capillary layer by squeezing a small amount of the sample between a pair of infrared transparent windows.

3.1.5 Solution

The number of good solvents for polymer or rubber samples, which exhibit wide transmission windows in the mid-IR region, is very limited. Consequently, solution sampling is used relatively very little for qualitative analyses. Additionally, working with high solute concentration solutions, many of which will be highly viscous, is often too troublesome for routine applications; thin cells will be difficult to both fill and clean. Furthermore, the solubility of a particular polymer may well depend on its crystallinity or molecular weight. This notwithstanding, solution sampling in thicker pathlength cells can be particularly suitable for some quantitative determinations on polymer and copolymer systems, providing solute–solvent interactions are minimal and do not give rise to band shifts or changes of band shape. It does have the advantage of inherent pathlength normalization for standard concentration solutions, and also that the solute, the polymer, will be in a reproducible state.

3.1.6 Solvent or Dispersion/Latex Cast Film

A thin film from a latex, dispersion or soluble resin may be readily prepared, with practice, by casting onto an infrared transparent support for direct examination. Alternatively, it may be cast onto another substrate, such as a glass microscope slide or thick aluminum foil, from which it may peeled off prior to measuring its characteristic infrared spectrum. Film thickness will be determined by solution concentration, deposition volume and spread. Solvent evaporation should be undertaken in a fume cupboard or other appropriate containment environment,

particularly if it is a toxicological hazard. Volatile solvents may evaporate readily at ambient temperature, and the transmission infrared spectrum of the cast film can be checked periodically to ascertain whether all the solvent has volatilized. The aim is to produce an even thickness, bubble free, nonscattering film and to avoid the formation of rings, i.e. where essentially a resin deposit ring rather than a film remains after the solvent has evaporated. This tendency can be alleviated if, towards the final stages of solvent evaporation, the film is gently evenly spread across its support by, for example, using the side of the deposit end of a capillary pipette. Final traces of solvent may require removal by gentle heating, for example, by placing the cast film on its support under an infrared heating lamp or, alternatively, placing the preparation into a vacuum oven. Complete removal of less volatile solvents from cast films or from films cast from solutions in which the resin is strongly solvated can often be very difficult or time-consuming. Prolonged heating under an infrared heating lamp, on a hot-plate or in a heated vacuum oven will likely be required, which must be carried out with due attention to both safe handling and evolved vapor containment. Also, care must be taken to ensure that no polymer degradation occurs. Some polymers may crystallize as the solvent evaporates, producing films that are highly scattering and not optimal for the examination purpose; others may be brittle and prone to cracking. Films from water solutions may be cast onto a nonhygroscopic window such as AgCl foil.

Films cast onto supports that are not transparent to infrared radiation, such as glass or aluminum, will need to be removed from their substrate. This method has been found to be particularly useful for polyamides (nylons) and linear polyurethanes (PUs). Films from these polymers may be cast from formic acid solutions onto glass microscope slides, which after the solvent has evaporated are well washed with water, dried and then peeled from their glass support prior to infrared examination. Films that adhere strongly to aluminum foil may be released by dissolving away the aluminum in caustic solution. Hummel[12] gives information on solvents and solubility for a wide range of polymers and resins.

Spin-coating and metered spreading are two methods utilized for creating uniform, thin (usually sub-micron thickness) layers on substrates. The former may be used to generate thin coatings on transmission windows or polished metal surfaces for reflection–absorption measurements; the latter is usually used for laying down surface coatings on polymer films, and can be useful for creating standard samples for attenuated total reflectance (ATR) investigations, see section 3.2.1. Polymers that are soluble in a convenient solvent can be spin cast from a 1–2 wt% solution. A sample stub is mounted on a low speed rotor (ca. 100 Hz), and a cleaned 13 mm diameter support is attached to the stub with double-sided adhesive tape. The support may be a suitable infrared transmission window, such as KBr or ZnSe, or a gold-coated disk appropriate to a transflectance measurement. Several drops of the polymer solution are then pipetted onto the support and the rotor is spun for a few seconds. This produces a uniform polymer film (of typically a few microns thickness) on the flat substrate. Metered spreading may be accomplished with a hand or controlled speed and pressure coater. Wired bars (K bar™) deposit wet films 4–500 μm thick onto the film substrate.[17]

3.1.7 Alkali Halide Disk and Mulls

The major requirement for a good dispersion preparation is a finely divided analyte powder, with a particle size below that of the wavelength of the infrared radiation. Ideally the refractive index of the analyte should match or be close to that of the matrix medium. In these circumstances, spectra that are free of anomalous dispersion, have a high contrast, and have a good signal-to-noise ratio may be recorded, with backgrounds which are essentially non-scattering. The most common matrix materials have been liquid paraffin oil (Nujol™) and powdered alkali halide, most often KBr. In mull preparations, a small amount (10–50 mg) of sample is first ground vigorously by using a pestle and mortar, and then mulled (intimately mixed) with a small amount of the oil. This suspension is then placed between the faces of two infrared transparent windows and squeezed to form a bubble-free, uniform, thin layer, which commonly has a translucent appearance. In disk preparations, an intimate mixture is made between an excess of dry, powdered alkali halide with the finely powdered sample. For a 13 mm diameter disk, this will typically be between 300 and 200 mg of KBr and 1 mg of sample. This mixture in a die is then pressed under high pressure, using apparatus that is commonly available from many accessory manufacturers. The KBr, which is transparent to mid-IR radiation, coalesces to form a clear disk containing the suspended, uniformily dispersed sample. The KBr disk pressing is undertaken usually with the die connected to a vacuum system for improved disk clarity.[16]

The requirement for preparing a finely ground sample precludes these presentation methods as being suitable for sampling many polymers and rubbers, because of their mechanical properties. Some brittle resins may respond well to normal grinding procedures, and finely divided powder may be removed from hard samples by rubbing them with a diamond-powder face coated spatula. Sometimes the addition of a small amount of solvent to soak and swell the polymer or rubber sample will facilitate sample preparation, since the swollen gel may break down more readily; the solvent must then be

removed by evaporation. Another alternative, which has been used for softer, rubbery samples, is to use a low temperature (liquid nitrogen) grinding accessory.

In general, nowadays neither the mull nor the alkali halide disk method is commonly used for the infrared analysis of polymer and rubber samples, particularly since the availability of diffuse reflectance and photoacoustic-FTIR (PA-FTIR) sampling techniques, see sections 3.3 and 3.5 respectively.

3.2 Reflection Methods

Reflection methods are appropriate to surface layer and coating characterizations, and they may be called into play as being more convenient when specimen preparation for a transmission experiment proves difficult. The principles underlying each technique have been described in detail in the article **Infrared Reflection–Absorption Spectroscopy** in this Encyclopedia and will only be described briefly here. FTIR microscopy applications are discussed later, in section 14.

3.2.1 Internal Reflection

Infrared internal reflection spectroscopy (IRS) is concerned with recording the spectrum of a sample by placing it in intimate contact with the surface of a higher refractive index element (crystal).[18,19] The IRS element is essentially transparent to infrared radiation, and infrared radiation must pass through the element so as to strike the sample at an angle of incidence that is greater than the critical angle at the IRS element–sample boundary. The internally reflected infrared beam emerging from the IRS element will have been attenuated by the absorption properties of the sample. This will yield a spectrum which resembles that recorded from a transmission measurement, but one in which, by comparison, the relative band intensities are skewed, since effectively an increasing pathlength of sample is probed with increasing wavelength (decreasing wavenumber) of the infrared radiation. This technique is commonly referred to as ATR spectroscopy. The evanescent wave created at the element–sample interface, which decays into the sample, also has a dependency on the critical angle at the boundary and the ratio of the refractive indices of the IRS element and sample. The probe depth is lower for higher angles of incidence, and lower for a greater ratio of IRS refractive index to that of the sample. In typical experimental set-ups, IRS can be described as producing an absorption spectrum which is characteristic of the sample surface in contact with the IRS element over a surface layer depth from a few tenths of a micron at short wavelengths to a few microns at longer wavelengths. Effective path length and thereby spectral intensity is increased through the use of multiple internal reflection (MIR) elements. It must be remembered that, in regions where it absorbs infrared radiation the refractive index of the sample and thus its critical angle has anomalous behaviour. As a consequence, particularly for the more strongly absorbing bands, band shifts and profile distortions may result. These may be minimized by using either or both a higher refractive index IRS element and a greater angle of incidence, both of which will, however, reduce the overall intensity of the recorded spectrum.

Techniques in IRS are used extensively in the polymer, plastics and rubber industries.[5,20,21] They are essential tools in identifying, characterizing and maintaining the quality assurance of surface layers, surface treatments and coatings of solid products such as films, fibers and moldings. In addition, polarized infrared radiation IRS studies may be used to gain information on surface structure anisotropy. Horizontal IRS set-ups particularly enable the ready study of lattices, aqueous dispersions and resin coating solutions, while, cylindrical (rod-like) or similar MIR accessories may be used as the key part of a process or polymerization monitoring probe.

Perhaps the classic use of the IRS technique has been the identification of the outer layers of polymer film laminates, for which the thickness of each outer layer was of the order of 0.5 μm or greater. This remains a major use, and is illustrated in Figure 1. Other applications include polymer weathering or irradiation studies, surface contamination investigations, such as additive bleed, surface coating morphology or cure monitoring, blend stratification observations, and polymer surface molecular orientation correlations with process variables. To record an unhindered fingerprint spectrum of a surface coating by the IRS technique, it will normally require that the coating must be of a thickness of 0.5 μm or greater. It is possible, however, to calibrate quality assurance methods that will monitor a coat weight or thickness to a significantly lower amount, especially if the coating has an absorption band in a region in which the polymer substrate shows no absorption features.

3.2.2 External Reflection

The near normal angle of incidence techniques of specular reflectance and reflection–absorption are used primarily as qualitative approaches for fingerprinting bulk samples. Grazing-angle of incidence reflection–absorption is a more specialized approach that optimizes absorption signal intensity for very thin layers of resins on reflective metal substrates. External reflection measurements made with polarized infrared radiation at the Brewster angle can be useful. At this angle, with parallel polarized radiation, reflectance is zero for films on a reflective substrate, so the recorded spectrum displays essentially only absorption features and there is no overlay of dispersion.

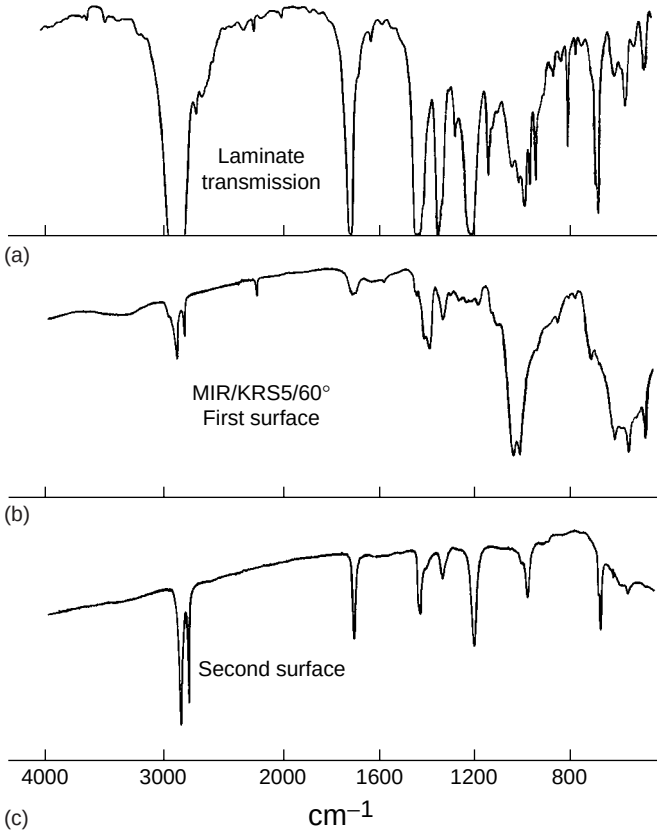

Figure 1 Infrared spectra recorded from a food packaging film. (a) Transmission spectrum of multi-layer film; (b) MIR (KRS5 prism, 60° incidence angle) spectrum of outer surface (barrier layer); (c) MIR (KRS5 prism, 60° incidence angle) spectrum of inner surface (heat-seal layer). The packaging film comprises a PP central layer, with an EVA copolymer heat-seal layer (second surface), and a vinylidene chloride/acrylonitrile/ester terpolymer barrier layer (first surface). The spectrum characteristic of each of the surface layers has been clearly isolated from that of the polypropylene (PP) in the internal reflection spectra.

Claybourn[22] is recommended for an overview of external reflectance principles and applications to polymers.

3.2.2.1 Specular Reflectance A pure specular reflectance spectrum will have the appearance of a dispersion curve; i.e. it will have the appearance of that of the first derivative of an absorption spectrum. From such a spectrum it is possible by subjecting it to a Kramers–Kronig transformation to extract the more analytically useful absorption index spectrum.[23,24] To produce a specular reflectance suitable for this approach, the sample must be homogeneous, have a flat surface and be essentially nonscattering. It should also be of a thickness such that no ray that has entered the sample returns to the detector; i.e. it should be ''optically thick'' so that it is free from both reflection–absorption and interference fringe features. An example is shown in Figure 2. (It should be noted that the weak CH stretching bands near 3000 cm^{-1} in Figure 2 show an inverse behaviour compared with the bands below 2000 cm^{-1}. They are not attributable to the polymer and they most likely arise from low levels of surface contaminant and occurred as a result of a transflectance type measurement. The conditions for pure specular reflectance can sometimes be difficult to achieve for weak bands.)

3.2.2.2 Reflection–Absorption, ''Transflectance'' In this approach, practically the infrared spectrum recorded is collected usually from a ''near'' normal incidence (typically between 10° and 30°) measurement made of a thin polymer, resin or rubber sample that is mounted on a reflective metal substrate. The spectrum is essentially equivalent to that recorded in a transmission experiment from a film of about twice the thickness, but superimposed on which will be a weaker specular reflectance spectrum.

A common application of this measurement technique is spectral fingerprinting of coatings on beverage containers, such as cans. It may also prove useful for qualitatively identifying a film that has been cast onto an appropriate substrate.

3.2.2.3 Grazing Incidence Reflection–Absorption The acronym RAIRS (reflection–absorption infrared spectroscopy) is commonly used to define external RAIRS measurements undertaken with p-polarization (electric vector in the plane of incidence) at high angles of incidence (ca. 80° or greater) on thin layers of materials (in the thickness domain 1000 Å or less) on reflective metal substrates. Associated with this technique is the metal surface selection rule. Under the experimental conditions used there is an intensified electric field normal to the metal surface, but a zero field parallel to the surface. Consequently, molecular vibrations with transition moments perpendicular to the surface will be excited and appear enhanced, while those vibrations with transition moments parallel to the surface will be undetected or give rise to only very weak bands. For reasonably intense infrared absorption bands, chemical information and structural ordering data are accessible for monolayer coverage.

3.3 Diffuse Reflectance

For polymer and rubber samples that are diffusely scattering, the diffuse reflectance infrared Fourier transform (DRIFT) measurement technique can provide a useful minimal or no sample preparation method.[5,25] Although complex[16,26–28] in that practically most DRIFT measurements involve detecting to varying extents Fresnel reflection components, both specular and diffuse, together

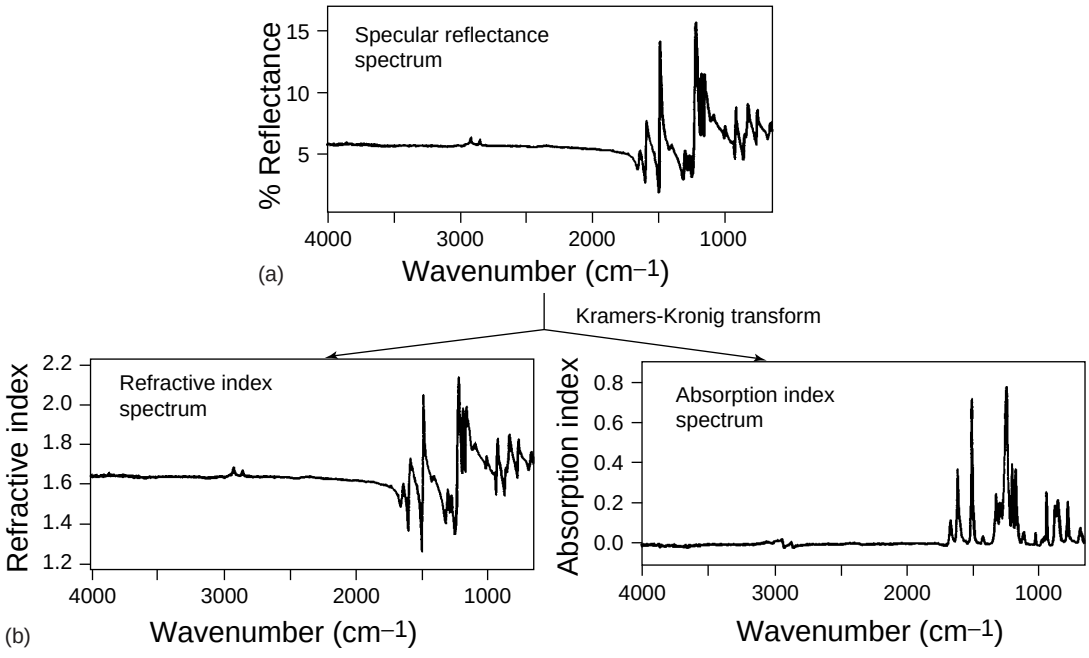

Figure 2 Schematic illustrating application of the Kramers–Kronig algorithm. (a) Recorded specular reflectance spectrum from a PEEK molding, 0.6 mm thick. (b) Absorption and refractive index spectra generated by applying the Kramers–Kronig algorithm to the recorded specular reflectance spectrum. (Reproduced from Chalmers et al.,[336] by kind permission of The Royal Society of Chemistry.)

with that radiation which has interacted more intimately with the sample, the sampling technique can be a very convenient qualitative tool. The diffuse reflectance of analytical interest results from penetration of the infrared radiation into the sample, by for example transmission through a powder particle or particles, and subsequent scatter by the sample matrix. However, in many circumstances reproducible sample presentation is a problem, and quantitative comparisons are often very likely to be only of low precision. Quantitative determinations are usually made from DRIFT spectra displayed using the Kubelka–Munk function,[16,27,28] which relates analyte concentration to the diffuse reflectance characteristics of the sample; it is the analog of the Beer's law for absorbance spectra.

To record a high contrast diffuse reflectance infrared spectrum from a powder, which is much more similar to its transmittance spectrum than its bulk reflectance spectrum, requires usually that it be finely divided. It should have a mean particle diameter of the order of or less than the interrogating wavelengths, i.e. $\leq 10\,\mu m$. Additionally, it is common practice to disperse it uniformly into a finely divided powdered, dry, non-absorbing matrix, such as KCl. A specimen of the mixture is then examined which is of ''infinite depth'', i.e. typically 2–3 mm or greater. The combination of dilution, typically in the ratio of approximately 10 : 1 diluent to analyte, together with using finely divided powders, reduce significantly Fresnel reflectance contributions from the sample. For continuous samples, such as films and fibers, and composites a KBr powder overlayer technique has been promoted to both minimize Fresnel reflection artifacts and enhance the relative contributions of surface layers.[1] For hard, intractable or large objects, a convenient approach is to abrade from the article's surface a fine powder with some silicon carbide or a diamond powder abrasive pad, which are available commercially. A DRIFT spectrum may then be recorded from the abraded powder, either in situ on the abrasive pad or removed and dispersed in dried, powdered KCl.

As an alternative to IRS, DRIFT has been demonstrated to be a very convenient method of recording characteristic spectra directly from a polymer foam,[29] regardless of whether it is rigid, flexible, open- or close-structured, see Figure 3.

3.4 Emission

The emittance, ε, at a particular wavenumber (ν) of a sample at a particular temperature may be defined as $\varepsilon_\nu = E_s|E_{bb}$, where E_s is the radiant energy emitted per unit solid angle by the sample at a particular temperature and E_{bb} is the radiant energy emitted per unit solid angle by a blackbody source at the same temperature. It is dependent on both sample composition and sample

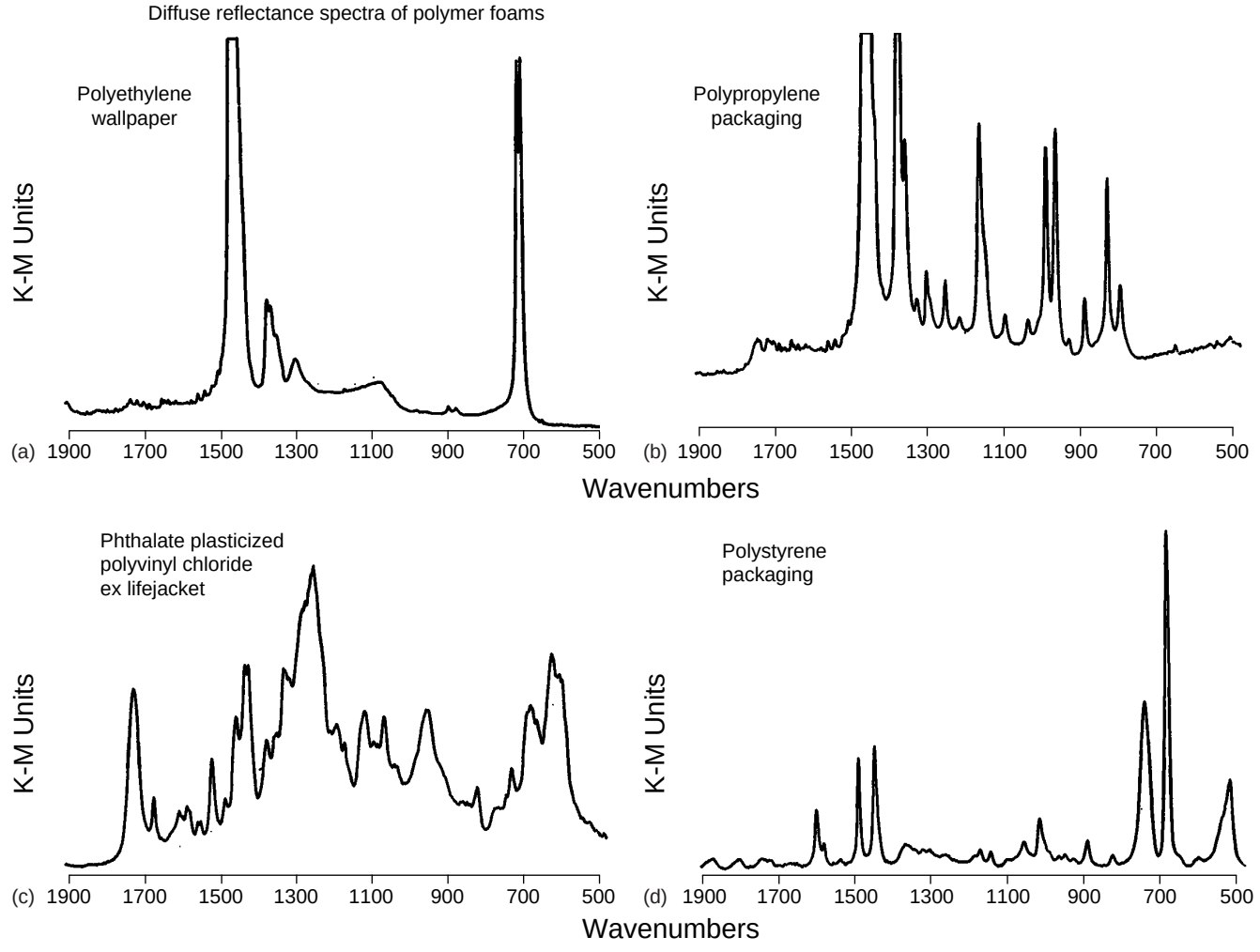

Figure 3 Diffuse reflectance spectra recorded from a range of polymer foams. (a) A polyethylene wallpaper; (b) PP packaging material; (c) phthalate plastcized poly(vinyl chloride) (PVC) from a buoyancy aid; (d) polystyrene packaging foam.

geometry.[30] The infrared emission spectrum of a sample is a plot of its emittance versus wavenumber. Following from Kirchoff's law, for a sample in equilibrium at a given temperature, $\varepsilon_v = a_v$ where, a_v, is the sample absorptance. Therefore, the emission spectrum of a thin unfilled polymer film will be similar to that of its inverted transmission spectrum, and as such may be used to fingerprint the polymer material. It is related to the polymer film's transmission (t) and surface reflectance (r) by $\varepsilon_v = 1 - (t_v + r_v)$, as given by Willis.[31]

A common practice is to record the emission spectrum from a thin polymer film on a reflective metal substrate. In this instance, the spectrum resembles more closely that recorded from a transflectance type measurement and emission spectroscopy may be used successfully to examine semi-quantitatively samples that are amenable to analysis by the reflection–absorption technique. For a thin partially transmitting sample on a totally reflective metal substrate, the emittance expression becomes:[30,32]

$$\varepsilon_v = \frac{(1 - r_v)(1 - t_v^2)}{1 - r_v t_v^2}$$

FTIR experiments for determining the emission spectrum of a thin polymer film must encompass measurements that remove background effects from radiating surfaces other than the sample when a cooled detector is used.[27,33]

Transient infrared emission spectroscopy (TIRES),[34] can be used to acquire analytically useful emission spectra from polymer samples that are optically thick. It largely circumvents the problem of self-absorption by examining a thin, heated surface layer of a moving sample, and is appropriate to on-line process monitoring, see section 13.

3.5 Photoacoustic

A wide range of polymer and rubber sample forms have been successfully fingerprinted by PA-FTIR. These include powder, chip, film, fiber, foam, composite, prepreg, and heavily carbon filled materials, such as a sliver from an automobile tire.

For solid samples, the conventional method of recording a PA-FTIR spectrum is by placing the sample in a cup-like holder inside an acoustically isolated, small-volume sealed chamber, the PA cell.[5,27,35–37] The cell is filled with an infrared transparent gas, which is usually helium. Photoacoustic detection involves direct measurement of the absorption of infrared radiation. Modulated infrared radiation is focused through an infrared transparent window onto the sample in the PA cell. The absorbed radiation is converted to thermal energy within the sample, which propagates to the sample surface and is released to cause pressure fluctuations in the surrounding gas, which generate the acoustic waves within the sample chamber. These modulated waves are detected by a sensitive microphone, located in the cell body, and the signals processed to produce an FTIR spectrum.

The relative intensity of absorption bands in a PA-FTIR spectrum is not solely proportional to their relative absorptivities. The key parameters to be considered are the thickness of the specimen, the optical absorption length (which is the inverse of the absorptivity) and the thermal diffusion length. Generalized theories for differing relationships between these parameters have been developed.[38–40] The thermal diffusion length is related to the sample thermal diffusivity and modulation frequency. If, for a particular band, it is greater than the optical absorption length, then saturation will occur. Two types of PA-FTIR measurements are commonly undertaken. The first involves signal detection undertaken when the FTIR is operating in its more conventional rapid-scan mode. For a Michelson interferometer operating in this mode, the modulation frequency (f) is both wavenumber (ν) and scan velocity (V) dependent, being related by $f = 2V\nu$, so increasing the spectrometer mirror scan speed will reduce the sampling depth, although this will still be wavenumber dependent, increasing with decreasing wavenumber. For a constant modulation frequency, which is independent of wavenumber, then the interferometer has to be operated in step-scan mode,[16,41,42] which when coupled with phase modulation can enable the extraction of characteristic information from both surface and submerged layers within a multi-layered sample.[41–45]

4 SAMPLE PRESENTATION

The quality, spectral contrast and quantitative precision of an infrared spectrum recorded from a polymer or rubber sample will likely be very dependent on the choice of either or both the sample preparation method and sampling technique. The choices will therefore likely not be solely governed by considerations such as sample form and measurement convenience but most importantly on the purpose of the analytical study. The requirements for generic typing may not need to be as stringent as for those for spectral matching against a computerized database, which may be less than those necessary for a compositional assurance analysis. Certain investigations, such as those associated with product morphology, will require that the integrity of material physical properties be maintained throughout the measurement, so specimen preparative methods such as dissolution, melting or grinding will not be viable.

5 QUALITATIVE ANALYSIS

The qualitative evaluation of the infrared spectrum of a polymer or copolymer may proceed on several levels. These depend on the intent of the analysis. The most simplistic approach, and the one perhaps most commonly practiced in industry, is that of generic identification followed by empirical or computerized pattern matching. Here, the polymer or copolymer is considered as a collection of functional groups or sub-structural units, for which the observed absorption bands are assigned using well established group frequency correlation tables.[46–48] This will likely type a polymer as, for example, an aliphatic or aromatic or alkyl–aryl hydrocarbon polymer, a polyamide or a polyimide, a polyester or a polycarbonate, a poly(aryl sulphone) or a poly(aryl sulphide). This may then be followed by pattern matching of the recorded spectrum against those in a hard-copy commercial or in-house reference collection, which will yield the next level of detail; for example, differentiating single number nylons, distinguishing PP from polybutene or poly(ethylene terephthalate) (PET) from an orthophthalate resin, or indicating the approximate composition of a copolymer. Alternatively, this end result may be achieved through spectral searching (pattern matching) against a computerized database of reference spectra. Many commercial polymeric materials and their products are formulations, however. Plastics and rubbers may often contain additive recipes with organic components such as antioxidants, UV stabilizers, processing aids, impact modifiers, plasticizers and lubricants. They may contain inorganic fillers for fire retardency or to add mechanical strength. Unless separated, these additives and modifiers may well contribute bands of significant intensity to the recorded spectrum.

The next level of investigation may be considered as correlating specific spectral features with particular

molecular conformations or phases. It might involve attributing certain bands or groups of bands to chemical structure differences associated with, for example, comonomer sequence lengths or modes of monomer insertion (e.g. 1,2- or 1,4-polybutadiene), end groups or, in the case of a thermoset, degree of cure. In any detailed interpretation of the infrared spectrum of a polymer one must take into account both the chemical and physical state of order of the system, see section 7.

Further interpretation of a polymer's absorption spectrum may then proceed by measuring the response of individual bands to polarized radiation. This aids both the assignments of bands to their symmetry species and in distinguishing possible chain structures. Much more complex, and a task infrequently undertaken by many practicing analytical infrared spectroscopists, is band prediction and assignment by a normal coordinate analysis (NCA).[3–6,8,9,49] The number of normal modes may be calculated from symmetry and group theory considerations. Iterative procedures based on force fields, derived from knowledge of bond distance and angles, are then used to predict the vibrational modes.

6 QUANTITATIVE ANALYSIS

To a first approximation, the absorption of an infrared band is linearly proportional to the number density of species giving rise to that band. As a consequence, infrared spectroscopy is a very valuable, cost-effective and widely used tool for the quantitative analysis of polymers and rubbers.[1,2,7,16] Although measurement procedures need to be calibrated, quantitative applications are diverse. For example, they extend from determining additive and residual monomer levels and copolymer composition, through end-group concentration determinations and cure monitoring, to dynamic and kinetic studies, to conformational analysis and measurements of molecular orientation and packing, to optical property definition. They are applied routinely in support of research, product development, competitive analysis, at-line quality assurance and in-line process measurements. Whatever the purpose, it must be remembered in developing any quantitative method that polymer infrared spectra are sensitive to changes in copolymer sequencing, tacticity, molecular conformation, molecular orientation and configuration.

The basic law, Beer's law, gives the relationship $A = \log_{10} I/I_o = kcx$, where A is the absorbance, I and I_o represent the amount of radiation transmitted and incident upon the sample respectively; c is the concentration of the absorbing species; k is a constant specific to the species and the wavenumber; x is the sample path length/thickness.

For thicker continuous, uniform polymer films (typically 0.1 mm or greater) in transmission, a quantitative method may only require measurement of the absorbance of a single band attributable to the analyte species, this being normalized against sample thickness, which has been determined using, for example, a dial gauge micrometer. The film's average thickness can be calculated by taking and averaging an appropriate series of micrometer readings over the area examined by the infrared beam. Most other univariate methods will be based on a measure of an absorption band ratio. If one of these bands may be considered as an ''internal thickness reference'', that is its absorption intensity is independent, essentially, of changing concentration, then it may be simply used to normalize the absorption intensity of the analyte band. In a two component system, A + B, in which the level of B is in large excess, it may be possible, for example, to consider the concentration of B as being essentially constant. The absorbance of a band due to B will therefore be indicative essentially of sample path length. An absorbance band ratio, A_A/A_B may then be plotted against the property value of A. If both A and B vary significantly, the absorbance ratio A_A/A_B should be plotted versus the property concentration ratio of A/B. A half-bandwidth rather than a peak intensity measurement may be more appropriate for some determinations. For instance, increasing molecular packing (crystallinity) will likely lead to band narrowing.

For more complex analyses, where, for example, extensive band overlap may preclude developing a high precision univariate quantitative method, multicomponent or multivariate analysis procedures may be used. Commonly used amongst these are principal component regression (PCR) and partial least squares (PLS). In addition, these chemometric methods are also frequently used in industry to analyse more simple systems, particularly for process measurements, since they can provide more robust, more precise, faster and convenient procedures for at- and on-line analysis of polymer and rubber properties.

Examples of quantitative applications of infrared spectroscopy in the analysis of polymers and rubbers feature later in this article and in many of the references cited, so consequently will not be illustrated specifically here. Key to many of these will be feature separation or highlighting through linear absorbance subtraction processes; contrast enhancement may be achieved in others through application of band resolution techniques, such as curve-fitting, or spectra enhancement processes, such as derivative spectroscopy or Fourier self-deconvolution (FSD).[1,7,16,50]

7 POLYMER SPECTRA

The infrared spectrum of a polymer can be a complex superimposition of many effects, but as a consequence analytically very valuable and informative. When interpreting the fine detail of the infrared spectrum of a polymer it should borne in mind that a polymer is not a single compound, and many of its qualities will affect its spectrum. Polymers have a molecular weight distribution, and chains have terminal groups which may be chemically significantly different from those in the main chain, and whose concentration will differ with chain length, as may their balance. Chains may also have side branches, which may be of differing lengths. They may be cross-linked. Polymer chains may contain irregularities through, perhaps, different insertions of monomer units, e.g. unsaturated aliphatic monomer units may polymerize via head-to-head or head-to-tail insertion, or through a vinyl group by 1,2- or 1,4- addition, with the latter existing in either a cis or trans conformation. Many polymers at room temperature may exist in equilibrium with as much as a few percent of absorbed water, which has characteristic bands in the infrared spectrum recorded from the polymer. Similarly, particularly for commercial materials, some polymer spectra may exhibit absorption bands that are directly attributable to catalyst or monomer residues. Copolymers may not only be manufactured with varying compositions, but also with differing sequences of the co-monomer units. For example, propylene/ethylene copolymers are sold commercially in many different grades. These may be obtained by varying the monomer feed concentrations. Materials are produced for commercial applications in which the copolymerized ethylene units are all present essentially as isolated units within the copolymer chain. Alternatively, they are also manufactured as grades in which the ethylene units are present mostly as longer (block copolymer) sequences within the copolymer chain, or sometimes with long sequences at the copolymer chain end (end-block). Some grades even contain distributions of each of these. Many solid phase polymer spectra are also particularly sensitive to state of order; they will be characteristic of stereoregularity, molecular conformation and molecular orientation and ordering. Many polymers are supplied commercially as formulated products (composites, plastics), the other constituents of which will likely contribute additional characteristic absorption bands to the infrared spectrum recorded from the bulk material. These additives may include stabilizers and property enhancers, such as antioxidants, UV-stabilizers, slip-agents, impact modifiers, processing aids or fire retardents; or they may be there for aesthetic purposes, such as pigments or dyes; or they may be fillers or reinforcing agents. Plastics may also be constituted as a blend of polymers; thermoset resins may be supplied as unreacted constituents.

7.1 Isomerism, Conformation, Tacticity, Regularity, Configuration and Crystallinity

Since infrared spectra are particularly sensitive to chemical group configuration and conformational isomerism, they play an important role in defining the structure and chemical composition of many polymers and rubbers.[1,4,5,9,11,51] The monomers butadiene and isoprene can, for example, each be incorporated into a growing polymer chain in one of several isomeric forms (see Figures 4 and 5), depending on the synthesis conditions. Stereo-isomeric structures in polybutadienes may be distinguished and in some cases reliable estimations of their composition determined from the γ-vibrations (out-of-plane CH deformation vibrations) of their hydrogen atoms attached to the olefinic bond (see also, polyethylene unsaturation in section 7.3). The 1,2-addition will be characterized by absorption features at 910 cm^{-1} and 990 cm^{-1}; the *trans*-1,4- and *cis*-1,4-polybutadienes

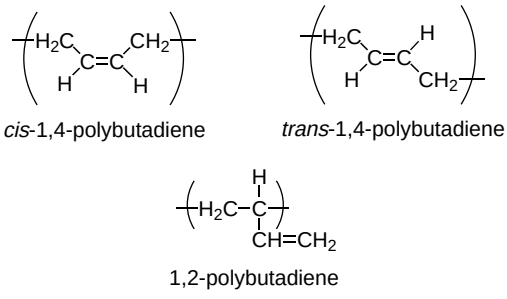

Figure 4 Polybutadiene chain units.

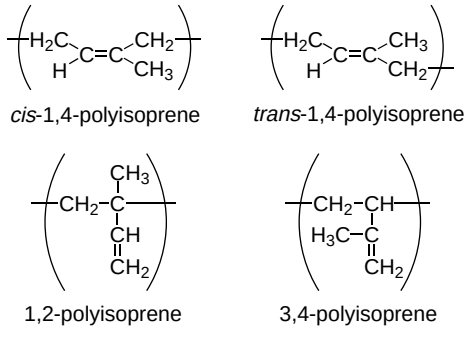

Figure 5 Polyisoprene chain units.

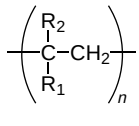

Figure 6 Polymer chain unit.

will give rise to absorption bands at 965 cm^{-1} and 735 cm^{-1} respectively.[2,9,12] *cis*-1,4-Polyisoprene, which occurs naturally in an essentially pure form known as hevea (or natural rubber), and *trans*-1,4-polyisoprene, existing naturally as gutta percha and balata, however, are not readily distinguishable by this vibration, since both show an absorption band near 830 cm^{-1}.[2,5,12] The 1,2-addition polyisoprene stereoisomer has absorption bands in similar positions to its polybutadiene counterpart, while, the 3,4-configured polyisoprene shows a band characteristic of the pendant methylene group near 888 cm^{-1}.

Polymers like 1,2-polybutadiene, PP, PVC, poly(methyl methacrylate) and polystyrene, have the basic structure shown in Figure 6, where $R_2 = H$ for 1,2-polybutadiene, PP, PVC and polystyrene; $R_2 = CH_3$ for poly(methyl methacrylate). R_1 is $CH{=}CH_2$ for 1,2-polybutadiene, $-CH_3$ for PP, $-Cl$ for poly(vinyl chloride), aryl ring for polystyrene, and $C(=O)-O-CH_3$ for poly(methyl methacrylate).

These are examples of polymers that have the possibility of forming stereoregular structures. Successive repeat units may follow each other in a regular configurational sequence. If they have the same configuration they are termed isotactic, if they alternate it is known as a syndiotactic sequence, see Figure 7. Atactic polymers exhibit no overall regularity in the configuration incorporation of successive repeat units when subjected to X-ray diffraction examination, although short-period regularity within polymer chains may still be evident in a nuclear magnetic resonance (NMR) spectrum.[1] They give rise to spectra that are very similar to those of the amorphous polymer.

Many stereoregular polymers that are substantially pure tend to take up preferred conformations. It is common for the backbone atoms of isotactic α-olefin polymers to assume a helical arrangement. An isotactic PP chain will coil into a 3_1 helix, i.e. each turn of the helix involves three repeat units. Polybutene (polybut-1-ene) and poly (4-methyl pent-1-ene) also form isotactic helical structures. PVC favors a planar zig-zag conformation.

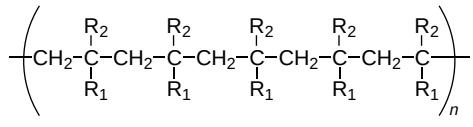

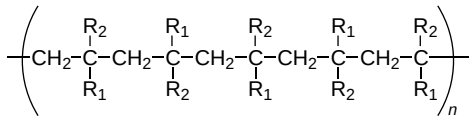

Figure 7 Schematic showing *isotactic* and *syndiotactic* chain unit sequences.

Conformationally regular polymer chains may pack together in a regular arrangement to form a three-dimensional crystalline phase, and highly stereoregular polymers tend to be crystalline. While stereoregularity in vinyl polymers is a prerequiste for the formation of crystalline domains, crystallinity *per se* in a polymer does not imply stereoregularity, it only implies conformational regularity, which may also be associated with amorphous regions. Helical isotactic PP is a semicrystalline polymer, as are many polycondensation products, such as PET and PEEK. Syndiotactic PP may be crystallized in either a 4_1 helix or planar zig-zag conformation. For each of these crystalline arrangements of PP chains there is a distinct fingerprint infrared spectrum[52,53] associated largely with the conformational regularity. Distinct bands may also be associated with the trans and cis isomers of PET, but whereas the latter only resides in the amorphous regions, trans sequences may be present in both amorphous and crystalline phases.[54,55] A poly(but-1-ene) film prepared from the melt will likely exist in the Type II crystalline modification, but with time this will revert to the higher density Type I form. This change, from a material with tetragonal 4_1 helix crystalline phases to one with 3_1 helix rhombohedral crystalline domains, may be observed readily over a few days by changes in the fingerprint infrared spectrum. Polytetrafluoroethylene, which also forms helical structures, can be very highly crystalline as manufactured, in the unsintered state. Above 19 °C PTFE has 15 $-CF_2-$ groups in the identity period, and the 15_7 helix unit cell arrangement is hexagonal; below 19 °C PTFE has 13 $-CF_2-$ groups in the repeat period and the 13_6 helical conformation packs in a triclinic structure.[56] Estimating PTFE crystallinity from an ''amorphous'' band at 778 $cm^{-1}$is the basis for a standard method implemented within industry.[57,58] Linear polyethylene can also have a high degree of crystallinity, and the orthorhombic unit cell contains two polymer chains. In the crystalline domains, the planes of the all trans zig-zag chains alternate perpendicular to each other. Vibrations on adjacent chains may then be in-phase or out-of-phase, and intermolecular coupling changes lead to correlation-splitting of certain bands,[6,8,9] notably the CH_2 rocking mode, which can give rise to a characteristic sharp doublet at ca. 730 and 720 cm^{-1}. It should be noted that the picture of a polymer as a two-phase (crystalline/amorphous) mixture is too simplistic for many systems.[9] For example, polyethylene requires three phases at least to describe its morphology. Raman spectroscopy is a valuable complementary technique to these investigations, and can sometimes be a superior tool for elucidating morphological effects.[59] Furthermore, under certain conditions, polyethylene can undergo a morphological change to a monoclinic structure, which features only a single CH_2 rocking mode band near

720 cm^{-1}. The unit cell of crystalline syndiotactic zig-zag PVC traverses two polymer chains, but as manufactured, although having a preponderance of planar syndiotactic sequences, comprises a mixture of other conformational sequences.[4,8,9,60] Resolving and assigning the component νC–Cl vibrations (750–550 cm^{-1}) has been the subject of several extensive investigations.

Although polymorphism is common among polymers that are crystalline, it is usually only one particular tactic form that has commercial significance as a crystalline or semi-crystalline material, e.g. isotactic PP.

The room temperature, solid phase mid-infrared 1300–800 cm^{-1} region spectra of a stereoregular aliphatic polymer usually contains many weak to medium intensity relatively sharp bands, see Figure 8. This pattern of bands is characteristic of essentially a single polymer chain in a preferred conformation and arises from conformational regularity, i.e. the chain spatial geometry, which is a consequence of the stereoregular structure, i.e. the bands are strictly indicative of regularity rather than being true crystallinity bands. In the melt, their spectra resemble closely those of their amorphous or atactic counterparts.

The far-infrared region (<400 cm^{-1}) is more appropriate for direct monitoring of polymer crystallinity, rather than the largely implicit observations associated with the mid-infrared region. Most of the low-frequency vibrations involve intermolecular associations, and are therefore influenced significantly by lattice packing of chains.[61] For example, differences observed in the far-IR spectra of isotactic PPs, particularly at 100 K, reflect the environments of the chains within particular unit cell (α, β, γ and smectic) configurations.[62]

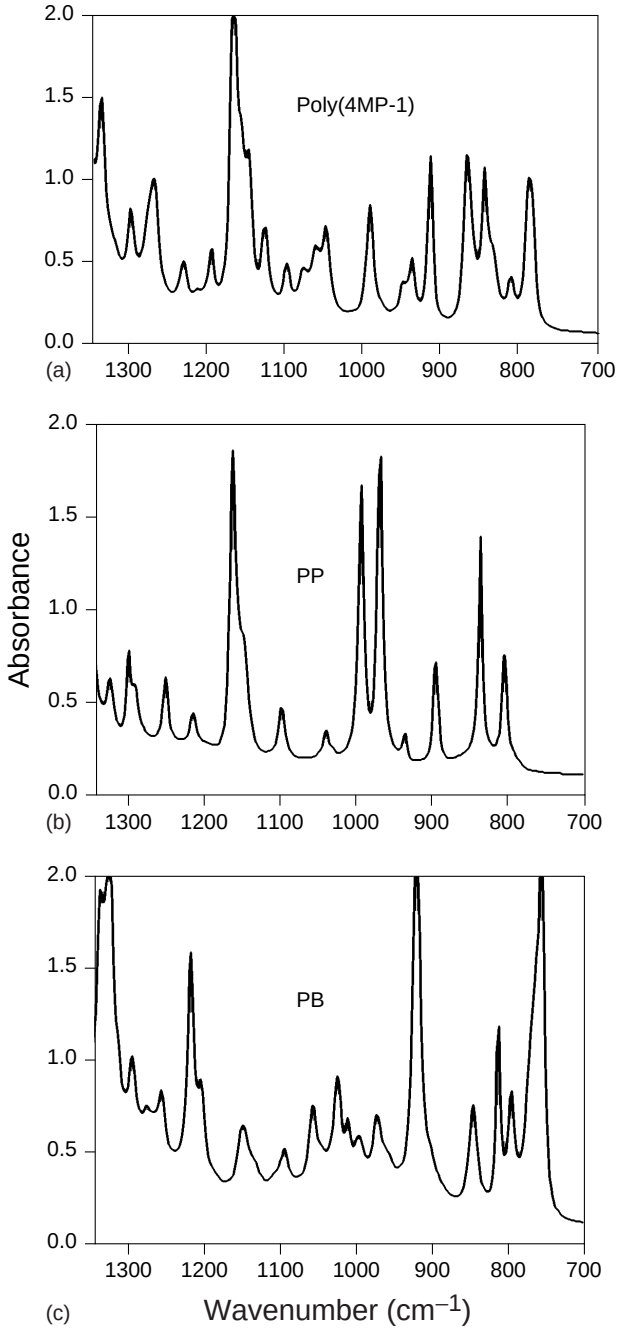

Figure 8 Infrared absorbance spectra of some poly(α-olefins), ca. 0.25 mm thickness. (a) Poly(4-methyl pent-1-ene); (b) polypropylene; (c) poly(but-1-ene).

7.2 Infrared Dichroism and Molecular Orientation

Many polymers are supplied as products in which the polymer chains are oriented (aligned) to some controlled extent during the fabrication process; for example, fibers, films and bottles. This imparts certain desirable end-use physical properties, such as directional stiffness. Polarized infrared spectroscopy is a particularly useful and practicable tool for investigating polymer anisotropy.

In the next two sections, summaries are presented of the principles underlying measurements of molecular orientation by infrared spectroscopy; rigorous derivations and treatments of orientation distribution functions (ODFs) and their measurement may be found in several of the cited references and references quoted therein.

7.2.1 Dichroic Ratio

Absorption of infrared radiation occurs when there is a change of dipole moment during a characteristic vibration of a polymer. Depending on the nature of the vibrations, they may have transition moments (directional changes in dipole moment) parallel, perpendicular or at some other angle to the major symmetry axis of the polymer, i.e. the chain axis. The directional character of the transition moment forms the basis of polarized infrared measurements.[13,63–67]

When investigated by using plane-polarized infrared radiation, the absorbance, A, of a polymer chain group is proportional to the square of the scalar product of the transition dipole moment vector, μ, of the group vibrational mode and the electric vector, E, of the incident polarized radiation. It is therefore also proportional to the square cosine of the angle, θ_μ, between the transition moment dipole and the electric vector, see Figure 9, $A \propto \vec{\mu}^2 \cdot \vec{E}^2 \propto \cos^2\theta_\mu$.

In general, oriented samples are examined with the infrared radiation beam polarized horizontally or vertically by reference to the axes of the sampling plane in the infrared spectrometer. The sample reference directions to be used may be chosen under crossed polarizers to coincide with the principal optical axes of the refractive index ellipsoid, although for commercial samples these may not always correspond exactly to the geometric axes of the product. For a fiber or uniaxially oriented film the axes of interest correspond to the direction of draw, OX_3, and the direction perpendicular to the draw. For a biaxially oriented film the axes in the film plane are often referred to as the forwards draw (FD) (or machine draw (MD)), OX_3, or sideways draw (SD) (or transverse draw (TD)), OX_1, directions, respectively. For a film sample, OX_2 is normal to the plane of the film, i.e. in the thickness direction. (Biaxially oriented films may be manufactured by either the stenter or a bubble process. For the former, the principal orientation directions will likely lie close to the MD and TD in the film center, but may well be displaced from these at the edges.)

For a drawn system of uniaxial symmetry, the dichroic ratio, R, is measured as:

$$R = \frac{A_\parallel}{A_\perp} \quad \left(\text{or } R = \frac{A_\pi}{A_\sigma}\right.$$

where, $A_\parallel$ and $A_\perp$ represent the absorbances of the investigated band measured with the infrared radiation polarized parallel and perpendicular to the draw direction, respectively. The dichroic ratio may have values from zero to those approaching infinity, depending on the angle β, see Figure 9. If $R > 1$, the mode is said to be a parallel mode (π mode); if $R < 1$ then the mode is said to be a perpendicular mode (σ mode).

For a sample with uniaxial orientation, then $R_o = 2\cot^2\beta$, where R_o is the dichroic ratio of an idealized, perfectly uniaxially oriented sample, for which the transition moment vector makes a cone with a semi-angle β with the local chain axis of the polymer chain (or segment of chain),[13,63] see Figure 9. (Note: if $\beta = 54°44'$, $R = 1$ and the absorption will be independent of the direction of the electric vector of the infrared radiation!)

For a simple, but physically unrealistic, description of a one-way stretched polymer sample with axial symmetry

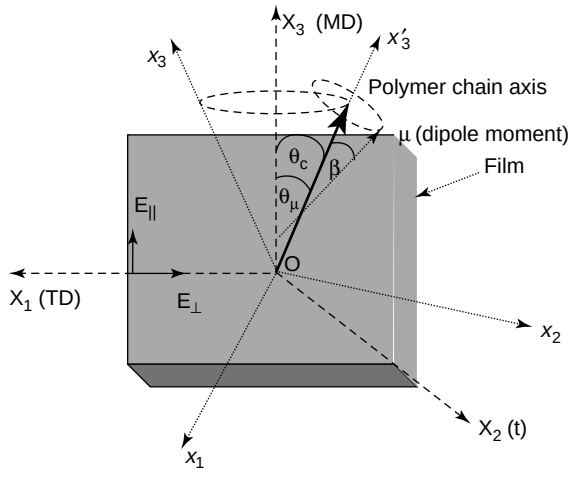

Figure 9 Schematic showing geometric relationship between sample axes, $OX_1X_2X_3$, the polymer chain axis, x_3, the transition moment dipole, μ, and associated angles for an oriented polymer film sample.

in which f represents the fraction of molecules that are perfectly aligned parallel to the symmetry axis, and $(1-f)$ represents the fraction of polymer chains that are perfectly random:[63,64,68]

$$f = \frac{(R-1)(R_o+2)}{(R+2)(R_o-1)} = \frac{(R-1)(2\cot^2\beta+2)}{(R+2)(2\cot^2\beta-1)}$$
$$= \frac{2}{3\cos^2\beta-1} \quad \frac{(R-1)}{(R+2)}$$

For a mode for which the transition dipole moment is along the chain axis, then $\beta = 0$ and the equation above simplifies to:

$$f = \frac{R-1}{R+2} \equiv \frac{A_\parallel - A_\perp}{A_\parallel + 2A_\perp}$$

For a vibration for which the transition dipole moment is perpendicular to the chain axis, then $\beta = 90°$ and:

$$f = \frac{-2(R-1)}{R+2}$$

In reality, there will be a three-dimensional distribution of chains with different values of θ. f is numerically identical to the Hermans orientation function,[69] which is a function of the average squared cosine of θ:

$$f = \tfrac{1}{2}\langle(3\cos^2\theta_\mu - 1)\rangle$$

The orientation function f describes the average of the quantity and $\langle\cos^2\theta_\mu\rangle$ denotes the average cosine squared value of the angle θ_μ between the reference direction in the sample and the transition dipole moment direction.

7.2.2 Orientation Distribution Functions

The quantity $\langle(3\cos^2\theta_\mu - 1)\rangle$ is the second moment of an ODF (the second-order Legendre polynomial in $\cos\theta_\mu$).[66,67,70,71] $\langle P_2(\cos\theta_\mu)\rangle$ (or P^μ_{200}) may therefore be determined from infrared spectroscopy measurements:

$$P^\mu_{200} = \langle P_2(\cos\theta_\mu)\rangle = \tfrac{1}{2}\langle(3\cos^2\theta_\mu - 1)\rangle$$

The range of values for this expression varies from $-1/2$ to 1. If the dipoles are aligned along the draw direction, $\theta_\mu \to 0$, and $\langle P_2(\cos\theta_\mu)\rangle$ is unity; for perpendicular alignment, then $\langle P_2(\cos\theta_\mu)\rangle$ becomes $-1/2$; for random orientation, $\langle P_2(\cos\theta_\mu)\rangle$ becomes 0.

While the dichroic ratio may be used as a convenience for comparative purposes, P_2 values are generally more analytically useful parameters to determine from the infrared spectrum.[66,67,70] They may be compared with P_2 values obtained from other measurement techniques, such as NMR and birefringence, and correlated more effectively with process variables, such as draw ratio. However, it is important to stress that the P_2 information derived from infrared is very limited.[72] It measures at best an average orientation; it says nothing about the shape of the distribution function. (More informative order parameters may be determined from Raman (P_2 and P_4) or NMR (P_2, P_4 and P_6)).

$\langle P_2(\cos\theta_c)\rangle$ defines the average alignment of a molecular axis or segment of a chain axis rather than that of a dipole moment transition. In practice, particularly for correlating product performance with infrared spectroscopic observations, it is more informative to determine the alignment of a specific molecular axis rather than of a dipole moment. In order to determine P_{200} (uniaxial symmetry about the draw axis) for the chain axis the dichroic ratio, R, must be determined for an absorption band with known dipole moment angle, β, to the chain axis, since β will vary for different bands. The orientation function is related to β and the dichroic ratio by the expression:[66,67,71,72]

$$P_{200} = \langle P_2(\cos\theta_c)\rangle = \frac{1}{P_2(\cos\beta)}\langle P_2(\cos\theta_\mu)\rangle$$
$$= \frac{2}{(3\cos^2\beta - 1)}\frac{(R-1)}{(R+2)} = \frac{1}{2}\langle(3\cos^2\theta_c - 1)\rangle$$

Four P_2 orientation averages that may potentially be determined from polarized IR measurements have been defined for different orientation symmetry cases.[66,67,70]

The objective of most investigations is to characterize the effects of process variables on molecular orientation.[71] Determining the dichroic ratio and then the orientation function for some one-way drawn samples requires simply that the absorbances of appropriate orientation sensitive band(s) be measured with the infrared radiation electric vector parallel and perpendicular to the sample draw axis. For weakly absorbing bands or for thin samples these may be made in transmission. Planar orientation may be characterized similarly for some biaxially drawn (or similar) samples. However, both uniaxial lengthening and biaxial stretching processes may induce molecular orientation anisotropy in the direction normal to the sample surface plane, and measurements in three dimensions are required for determining the orientation functions. In these calculations an important relationship is:

$$A_o = \frac{A_{X_1} + A_{X_2} + A_{X_3}}{3}$$

where A_o represents the absorbance of a band exclusive of contributions due to orientation of the polymer,[65,71] and A_{Xi} is the absorbance along the OX_i axis. It may be used as a normalization parameter to determine comparative measures of orientation in a sample, i.e. A_{X_1}/A_o, A_{X_2}/A_o and A_{X_3}/A_o.

To determine all the average second-order orientation parameters that may be calculated from infrared measurements for oriented samples that have other than uniaxial symmetry about the draw axis, e.g. a biaxially oriented film, then measurements must be made for which there is a component of the electric vector parallel to the normal of the sample surface, which in the case of a polymer film is its thickness direction.[67,70] This has been achieved in a number of ways. Transmission and calculation procedures on tilted thin films have been developed,[13,64–67,70,71,73,74] as have surface layer measurements using IRS approaches.[67,71,75–78] Early developed ATR methods relied on non-standard accessory designs, assumptions regarding refractive index and optical path lengths, and was practically difficult. Many of these limitations may be circumvented if measurements made on dichroic bands are normalized relative to a non-dichroic internal band in the sample spectrum.[79] A similar normalization approach has been used successfully for external (specular) reflectance measurements.[80,81] Providing specimen preparation does not induce or alter the orientation characteristics, an alternative approach is to microtome thin orthogonal sections through a sample and examine each by transmission FTIR microscopy.[82,83]

The average cosine squared values for the angles between the sample and polymer chain axes systems are another set of useful and frequently determined descriptive parameters.[67,72] For example, for random orientation, where $\langle P_2(\cos\theta_\mu)\rangle$ and $\langle P_2(\cos\theta_c)\rangle$ are both zero, $\langle\cos^2\theta_c\rangle$ is 1/3; whereas, for perfect uniaxial orientation $\langle P_2(\cos\theta_\mu)\rangle$, $\langle P_2(\cos\theta_c)\rangle$ and $\langle\cos^2\theta_c\rangle$ all equal unity.

7.2.3 Applications

Orientation measurement experiments may be made on pre-stretched samples or series of samples, i.e. in a static mode, or they may be undertaken in a dynamic sense, in which stretching and infrared measurements are made essentially simultaneously, see also section 12.

7.2.3.1 Static Measurements of Molecular Orientation One of the most extensively studied polymers has been the commercially important PET[65,66,70,73,76,77,79–81,84–90] In PET, as the polymer chains are extended then the glycol unit can undergo a conformational change from the gauche to the more extended trans form. Specific $\langle \cos^2 x_i X_j \rangle$ values may be interpreted to describe preferential orientation of the aromatic rings into the sample plane, or preferential orientation of the chain axis with respect to the sample axes. Figure 10 shows some of the key bands (see Table 1) used in studies of molecular orientation in PET.

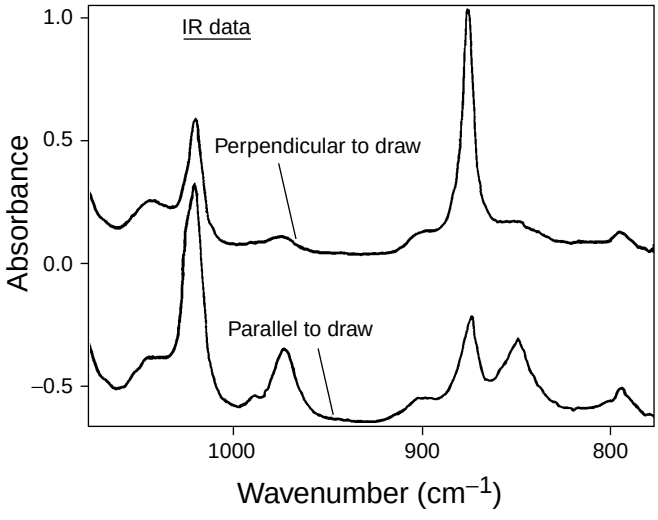

Figure 10 Polarized infrared absorbance spectra recorded from a uniaxially drawn PET. Upper spectrum, electric vector aligned perpendicular to film draw direction; lower spectrum, electric vector aligned parallel to film draw direction. The spectra have been offset for clarity.

Table 1 Some key bands in molecular orientation and crystallinity studies on PET

Band (cm^{-1})	Polarization	Assignment
1410		non-dichroic (aryl ring mode), reference band – 'insensitive' to orientation and conformation
1019	∥	aryl ring
975	∥	*trans* conformer
896		*gauche* conformer
875	⊥	aryl ring
848		*trans* sensitive mode
795	∥	aryl ring

Table 2 Some orientation sensitive bands in polyethylene

Band (cm^{-1})	Polarization	Assignment (A, amorphous; C, crystalline)
2016	π	A + C
1894	σ	C
1368	π	A
1352	π	A
~1305	π	A
1078	σ	A

Since in PET extended trans sequences of chains may crystallize, for measures of molecular orientation it is important to decouple the effects on the infrared spectrum of orientation from those of crystallization. If different conformers may be associated with different phases, then the orientation of individual phases can be measured. For example, in a study of orientation development in low density poly(ethylene) (LDPE) blown films crystalline order has been associated with bands at 1050 cm^{-1} and 1175 cm^{-1};[91] the modes giving rise to these bands have transition moments along the chain axis. For studies on thicker samples, such as sheet and tube walls, then other weak bands have been used to determine orientation levels for both amorphous and crystalline phases,[73,92–94] see Table 2. Many studies of both uniaxial and biaxial molecular orientation have been undertaken on a diverse range of polymers, polymer systems and rubbers. Some additional references are given in Table 3.

7.2.3.2 Dynamic (Rheo-optical) Measurements Rheo-optics is a term used to define a combined mechanical test with an optical measurement.[111,112] Simultaneous stress–strain and infrared data are recorded during polymer elongation in rheo-optical FTIR experiments. The sample, usually in film form, is clenched between the jaws of a tensile stretcher device mounted in a temperature controlled accessory sited into the sample compartment of an FTIR spectrometer. Sequential FTIR spectra are then recorded coincident with stress–strain data as a function of strain (and/or time). Interferograms over short time intervals are acquired with radiation polarized alternately parallel and perpendicular to the elongation direction. Deformation during elongation and subsequent relaxation under constant strain processes may both be monitored, as may structural changes during loading–unloading cycles.[9] Some references to rheo-optical FTIR studies are given in Table 3.

7.3 End Groups, Branching

End group balance or type and short side-chain branches are important characteristics of a polymer. They can

Table 3 Some references to studies of molecular orientation of polymers

Polymer	Orientation (static measurement) references	Rheo-optical measurement references
PET	65, 66, 70, 73, 76, 77, 84–90	111, 112 (films), 113 (yarn)
Polyethylene	73, 91–94	
PP	95	
Nylons	96, 97 (films), 98 (fibers)	
PVC	99	
PVdF	100	112
Poly(dimethyl siloxane)	101, 102	
PTFE	103	
Polystyrene	104	117
Polystyrene-PPO blend	105	
cis-1,4-Polysiprene	106	
Natural rubber	106	
cis-1,4-Polybutadiene	107	
PUs	108–110	111, 114–116
Epoxy resin		118
Liquid-crystalline elastomer		119

PVdF, Poly(vinylidene fluoride).

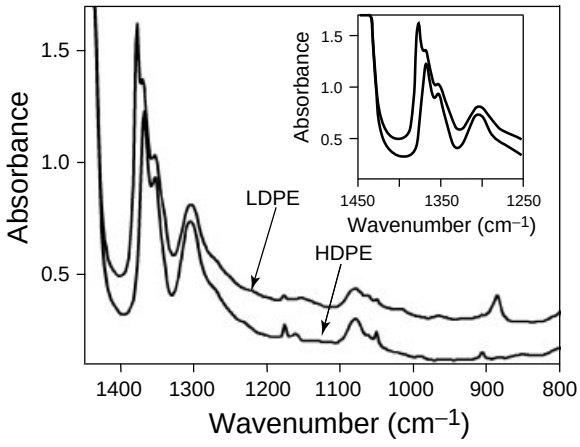

Figure 11 Infrared absorbance spectra of LDPE and HDPE films. Spectra have been offset for clarity. The difference in methyl group concentration is evident from the different band intensities at 1378 cm^{-1}.

have a major influence on polymer morphology and properties, affecting stability, processability and application. Their detection, identification and quantification may sometimes be simply accomplished by infrared spectroscopy. (Long chain branching cannot usually be appropriately studied by infrared spectroscopy.) If not readily discernible or measurable, the contrast of bands characteristic of end groups may need to be enhanced. This may be achieved in a number of ways. A common approach is the use of difference spectroscopy, in which the spectrum of a higher molecular weight polymer is spectrally subtracted from that of a lower molecular weight polymer, or two polymers of similar molecular weight but with different end group types. Other approaches may invoke derivatization or, particularly for hydrogenated end groups, the use of isotope exchange. Two classic examples of end group measurements by infrared spectroscopy are the determination of the methyl group concentration in poly(ethylenes) and the carboxyl and hydroxyl end group concentrations in PET.

LDPE and high density poly(ethylene) (HDPE) may be distinguished readily in their infrared spectra by the intensity of the 1378 cm^{-1} absorption band, which is attributable to the methyl symmetric bending mode, $\delta_{sym}(C-CH_3)$. In HDPEs the methyl groups are present essentially as end groups.[120,121] They are therefore characterized by relatively low concentrations of methyl groups, typically <10 methyl groups per 1000 C atoms; whereas in LDPEs, which contain a much higher concentration of side-chains, the methyl group concentration may typically lie between 20 and 30 per 1000 C atoms. Figure 11 illustrates this typical distinguishing feature between an HDPE and an LDPE. The traditional way of determining the intensity of this band has been by spectral subtraction (difference spectroscopy), in which the overlapping absorption bands at lower wavenumbers due to methylene group wagging modes, $\gamma_w(CH_2)$, are minimized, using a polymethylene or HDPE of very low methyl content for the compensation.[120,121] Further differentiation between the infrared spectra of HDPE and LDPE may often be confirmed in the pattern of bands between 1000 and 850 cm^{-1}, which are the out-of-plane deformation modes of the hydrogen atoms associated with the olefinic unsaturated groups at the chain end or in the polymer chain. Their relative intensities vary with the mode of manufacture and catalyst system used. In many LDPEs the absorption band at 888 cm^{-1} due to vinylidene (pendant methylene) groups is the more intense. This is in contrast to many HDPE spectra, in which the most prominent band is the 910 cm^{-1}, which arises from the terminal vinyl group, $-CH=CH_2$, which has a weaker partner at 990 cm^{-1}. The other key band in this region is that due to the chain unsaturation group $-CH=CH-$ which occurs at 965 cm^{-1}. These simplistic comparisons between LDPE and HDPE become much less straightforward when the infrared spectra of linear low density poly(ethylenes) (LLDPEs) are considered. In LLDPEs short-chain branches are introduced into the polymer chain by copolymerizing ethylene with low levels of an α-olefine, such as propene, butene, hexene, octene or 4-methyl pentene-1. The position, half-width and absorption coefficient of the methyl symmetrical deformation band near 1378 cm^{-1} is dependent on the

branch type.[122,123] The position of the methyl rocking band that occurs in the region 880–940 cm^{-1} is also characteristic of the type of branch. However, these bands are overlapped by those of the out-of-plane deformation modes of the hydrogen atoms of the olefinic unsaturated groups. Their position may be ascertained through following careful and safe bromination, which saturates the double bonds.[122–124]

Specific end-capping of a polymer may be the key to imparting thermal stability. Such processes may in some circumstances be monitored by infrared analysis. In nascent copolymers of tetrafluoroethylene/hexafluoropropylene (TFE/HFP) unstable end groups may be identified and their concentrations quantified through their infrared absorption bands.[125–127] Monomeric and dimeric $-CO_2H$ groups, which give rise to absorption bands at 5.52 μm and 5.64 μm (1812 $cm^{-1}$and 1773 cm^{-1}), respectively, may be decarboxylated by appropriate treatment to the more stable $-CF_2H$ end, which absorbs near 3.32 μm (3012 cm^{-1}). Vinyl ends ($-CF{=}CF_2$), absorbing in the infrared at 5.57 μm (1795 cm^{-1}), are also a potential cause of thermal instability in these copolymers. These may be oxidized during processing to $-COF$ ends, which are characterized by an absorption band at 5.31 μm (1883 cm^{-1}).

The sensitivity of infrared spectroscopy to $-OH$ vibrations makes it particularly well suited to monitoring or determining these types of end groups. A well established quality assurance method is that of measuring the end group balance in PET, which may be synthesized commercially by a polycondensation between ethylene glycol and terephthalic acid. The resulting hydroxyl and carboxyl end groups may be determined from measurements of the (sample thickness) normalized absorption intensities at 3542 cm^{-1} and 3256 cm^{-1}, respectively.[128–130] These may be correlated against absolute determinations from NMR and chemical titration methods, respectively. Their absorption maxima may also be highlighted by deuterium exchange, see Figure 12. Similarly, infrared methods have been developed for determining the hydroxyl and carboxyl end group concentration and therefore the number-average molecular weight in poly(butylene terephthalate) (PBT),[131] and poly(ethylene naphthalene-2,6-dicarboxylate) (PEN).[132]

7.4 Hydrogen Bonding

The properties of many polymers and copolymers depend on their physical network. Since infrared spectroscopy is an excellent tool for observing hydrogen bonding and studying its temperature dependence behavior, many detailed studies have been undertaken on polymer and rubber systems. In particular, several studies have focused on self-association mechanisms in polyamide,[133–137]

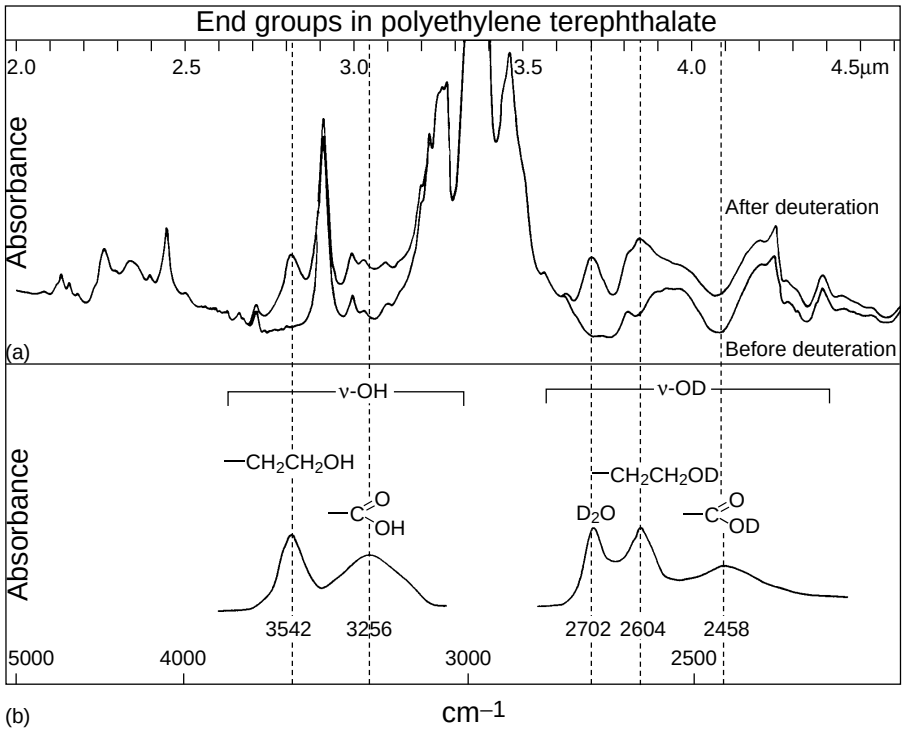

Figure 12 End groups in PET. (a) Overlaid spectra of dried PET film before and after deuteration. (b) Difference spectra between the films before and after deuteration highlighting postions of hydroxyl and carboxyl end groups.

PU[138–143] and polyurea[144,145] systems, many of which have included related work on model compounds. Also, study of inter-association bonding processes between two different polymers has been the subject of considerable investigation, with particular reference to polymer–polymer mixing and miscibility. Coleman et al.[146] provide a comprehensive monograph on the latter. Blend miscibility studies are discussed in more detail in the next section.

Temperature dependence studies of the νNH region of polyamides,[133–137] coupled with restricted curve-fitting procedures,[136,137] have led to the overlapped band envelope observed in this region being assigned as being composed of contributions from two species. These were a ''free'' N–H and hydrogen-bonded N–H stretching modes at room temperature, occurring at ca. 3445 cm^{-1} and ca. 3300 cm^{-1}, respectively. Systematic changes were observed for these bands with increasing temperature, with a shift to higher wavenumber noted for the hydrogen-bonded band. For the Amide I region (νC=O), three bands were discerned at room temperature. These are, a band ascribed to ''free'', non-hydrogen-bonded, carbonyls and a band from disordered hydrogen-bonded carbonyl groups; and for semi-crystalline nylons, an additional band arising from ordered hydrogen-bonded carbonyl groups, showing an observable sensitivity of the Amide I band to conformation. Again, systematic temperature dependent intensity changes and wavenumber shifts were observed for these bands. The spectra shown in Figure 13 illustrate the data recorded from such an experimental study. Similar approaches have led to the assignments of the various bands contributing to the νN–H and νC=O band profiles observed with PUs[138–143] and polyureas.[144,145]

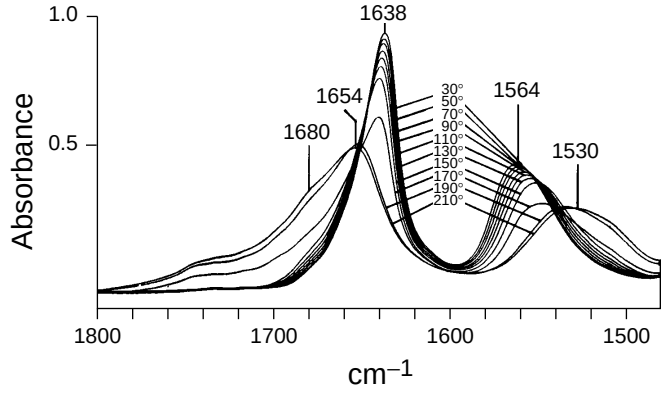

Figure 13 Infared absorbance spectra of Amide I and Amide II regions of nylon 12 as a function of temperature. (Reproduced from Coleman et al.[137] by kind permission of Wiley VCH Verlag GmbH, Weinheim.)

7.5 Blend, Miscibility Studies

Enhancing product ranges through polymer and/or rubber blending has a long commercial history. Blend properties are strongly dependent on the state of mixing. If the polymers are not miscible, which implies that they do not interact at a molecular level to form a compatible mixture, then they will likely phase separate and exhibit poor mechanical performance.[1] Infrared spectroscopy is a prime tool for detecting and probing the interactions and structure of polymer and rubber blends, and FTIR studies are reported extensively in the literature. The reader is recommended to the books by Garton[5] and Coleman et al.,[146] and the study by Coleman & Painter[147] for detailed discussions.

The infrared spectrum of a blend of two incompatible homopolymers will be equivalent essentially to the summation in appropriate proportion of the individual infrared spectra of the two components. Specific intermolecular interactions, such as hydrogen-bonding or polar coupling, may be detected through band position shifts and intensity modifications. The most widely reported are those associated with the perturbation of a carbonyl stretching vibration of one polymer by hydrogen-bonding with the other polymer, which leads to a reduction in the wavenumber position of this absorption band in the infrared spectrum of the blend, see example in Figure 14.

Infrared temperature dependent experimental observations have been made on several blend systems over a range of compositions and associated homopolymers designed to aid elucidating inter-polymer associations from self-association absorption contributions. The experimental data have been analysed with data manipulation procedures such as curve fitting. From such investigations, the fractions of hydrogen-bonded groups have been estimated and thence, for some, equilibrium constants of association have been calculated. These data have then been used to map phase diagrams and predict miscibility windows for polymer blends.[148–154]

The data processing techniques of factor analysis, difference spectroscopy and least-squares curve fitting have been applied to the infrared spectra recorded from cast films of a series of PVdF and poly(vinyl acetate) (PVA) blends.[155] Factor analysis from temperature experiments was used to establish compatibility or incompatibility. An ''interaction spectrum'' was then derived through a least-squares difference spectroscopy procedure, by subtracting the contributions to the infrared spectrum of the pure homopolymers. The percentage contribution of the ''interaction spectrum'' appropriate to each blend composition was determined by a least-squares analysis and followed throughout heat treatment ageing processes.

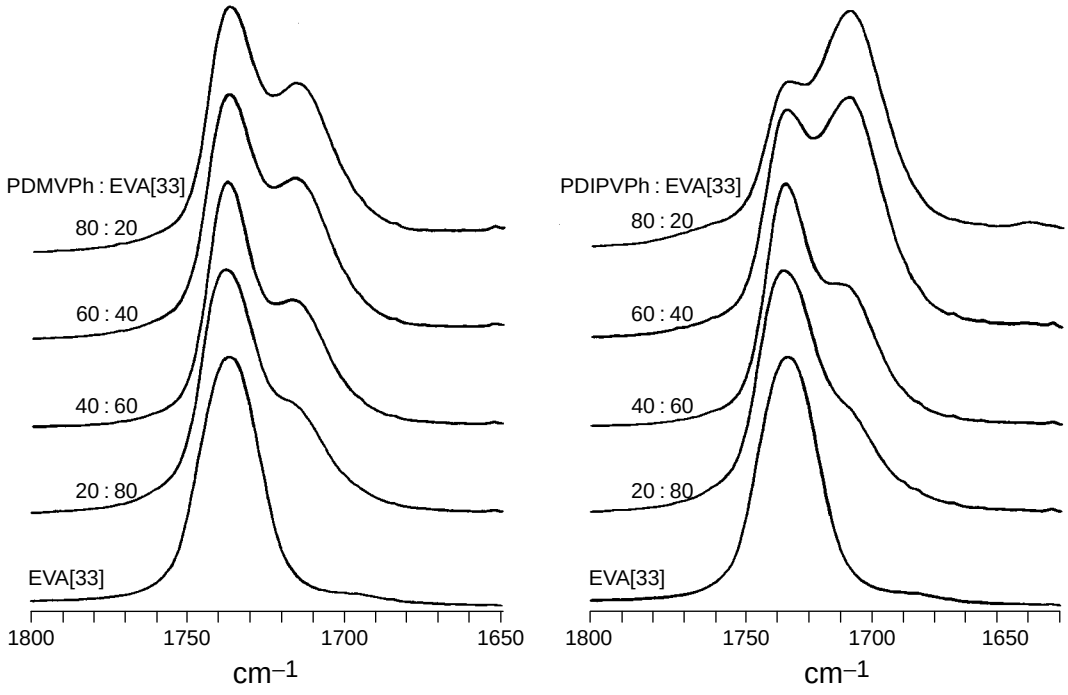

Figure 14 Infrared absorbance spectra at room temperature recorded in the carbonyl stretching region from films of pure EVA copolymer (33 wt% VA vinyl acetate) and blends containing 80, 60, 40 and 20 wt% PDMVPh, poly(2,6-dimethyl-4-vinyl phenol) and PDIPVPh, poly(2,6-diisopropyl-4-vinyl phenol). The lower wavenumber band is a consequence of hydrogen bonding between the blended materials. (Reprinted from G.J. Pehlert, X. Yang, P.C. Painter, M.M. Coleman, 'Self-association Versus Interassociation in Hydrogen Bonded Polymer Blends: 2. Comparison of Theoretical and Experimental Miscibility Windows for Poly(2,6-dialkyl-4-vinyl phenol) Blends', *Polymer*, **37**, 4763–4771 © (1996), with permission from Elsevier Science.)

The strong modulus in the rubbery zone of heat processed acrylamide (AA)-grafted silicone and maleic anhydride modified ethylene–propylene–diene monomer (EPDM) blends has been attributed to chemical interactions from observations made from the νC=O region of their infrared spectra.[156] In a study in which differential scanning calorimetry (DSC) and temperature dependent FTIR were combined to investigate blends of polyisoprene with poly(*cis*-butadiene-co-1,2-vinyl-butadiene), FTIR experimental data were used to detect microstructures that relate to transition temperatures of sequences in the polymers, and postulate that polyisoprene segmental movement onset occurs before that of segments associated with the other polymer, accounting for the broadened DSC observed transitions for the miscible blends.[157]

7.6 Copolymer Composition

Copolymers may be thought of as polymers with chain structures in which chemically different monomer units are linked in different concentrations in a "non-uniform" manner;[12] to a large extent this "non-uniformity" will likely have been designed to produce products with differing chemical or physical properties. An example, that of propylene/ethylene copolymers, has already been outlined in the introduction to section 7. Determining quantitatively copolymer component concentrations by infrared spectroscopic methods is a major analytical application area. Routine, cost-effective methods are widely developed and used for product composition quality assurance within industry. At the simplest level, it involves recording the infrared spectra in an identical manner from a series of copolymers of differing but known compositions, which cover the specification range of interest. From these, it may be possible to assign uniquely non-overlapped absorption bands to a particular copolymer component. The normalized absorption intensity of such a band arising normally from the presence of the lower concentration component is then correlated with comonomer composition. If the method is to be used over a wide concentration range and/or for copolymers prepared possibly by different synthetic routes, then it will be particularly important to ascertain whether the absorption bands are position and shape sensitive to co-monomer sequencing and sequence lengths.[1] For precise, robust procedures, particular care will also have to be exercised in the choice and specification of the method of sample preparation and presentation. These must be reproducible, in order to minimize any band intensity variations that may occur

with changes in copolymer morphology or conformation, or that may be directly related to optical artifacts associated with the sampling technique. Many reliable, simple to operate, multicomponent analyses involving overlapping infrared absorption bands have now been developed and implemented through the use of multivariate regression procedures, such as PLS analysis. While these analyses may be more complex, similar criteria still apply, in that due consideration must still be given to effects associated with sequencing, crystallinity, and sampling.

Approaches to compositional analyses of copolymers for quality assurance will depend on many factors, including specimen presentation method and spectral complexity, but particularly concentration range and precision requirement. For example, in a standard method by the American Society for Testing and Materials (ASTM)[158] copolymerized ethyl acrylate content in the range 1–25 wt% present in ethylene–ethyl acrylate copolymers is calculated from a determination of the absorbance at 11.6 μm (862 cm^{-1}) per mm of film thickness for films of thickness between 0.18 and 0.5 mm; whereas in another published method[159] a ratio of absorbance integrals of the ν–C=O first overtone band at 3450 cm^{-1} to that of 2750–2450 cm^{-1} (CH) were determined over an extended range from films of similar optical path length. Both methods report high precision, e.g. ±3 mol%.[159] Four ASTM standard methods have been described as being required to deal with commercial ethylene–propylene copolymers and ethylene–propylene–diene terpolymers in the range 35–85 mass % ethylene.[160] For an ethylene–propylene–diene terpolymer system containing diene in the range 0.1–10 wt%, diene compositions were determined from normalized second-derivative spectra.[161] Many, simple to implement fit-for-purpose

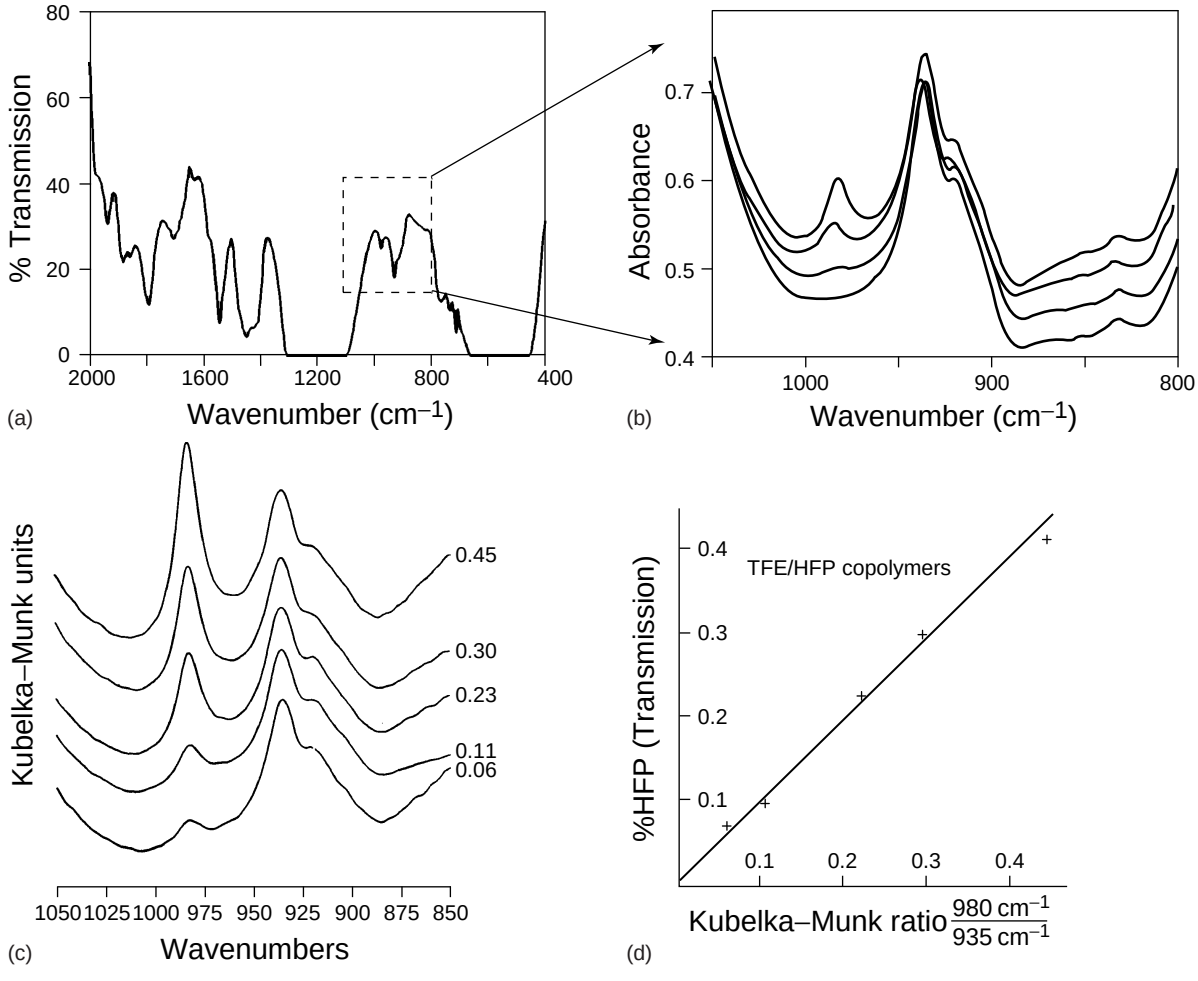

Figure 15 Infrared analysis of TFE/HFP (hexafluoropropylene) copolymers. (a) Transmission spectrum recorded from a cold pressed disk, ca. 0.2 mm thickness, formed from a 0.1 wt.% HFP copolymer powder. (b) Scale-expanded absorbance spectra of a PTFE and three TFE/HFP copolymers containing 0.01, 0.05 and 0.1 wt% HFP. (c) Diffuse reflectance spectra recorded directly from a range of TFE/HFP copolymer powders. (d) A comparison between transmission and diffuse reflectance measurements of band ratios.

copolymer compositional methods have been described by Haslam et al.[2] and Zichy.[162]

Figure 15 shows a band ratio DRIFT method evaluated for dealing directly with TFE/HFP copolymer powders.[27,163] It was developed from the more routine transmission approach, which made the measurements on coalesced disks, see section 3.1.3. Figures 16–18 outline some simple examples of quantitative procedures that have been developed for copolymer composition. Figure 16 illustrates a transmission procedure for determining from cast films the composition of copolymers of aryl–ether–sulphone and aryl–ether–ether–sulphone.[164,165] In Figure 17 an example is shown of a compositional measurement made on a compression-molded EVA copolymer film.[166] A number of potential simple quantitative PA-FTIR methods for rubber compositions have been reported.[167] Figure 18 shows some ethylene oxide/propylene oxide (EO/PO) copolymer spectra recorded on a ZnSe internal reflection element, which were the subject of a successful PLS treatment.[165] Infrared absorbance measurements of bands characteristic of unsaturated species have been used to study the degree of hydrogenation in nitrile-butadiene rubbers[168,169] and composition of butyl and halo-butyl rubbers.[170] In each the primary method used was NMR.

Many additive concentrations and filler levels may be determined in polymer and rubber formulations by similar analytical procedures to that used for copolymer composition.

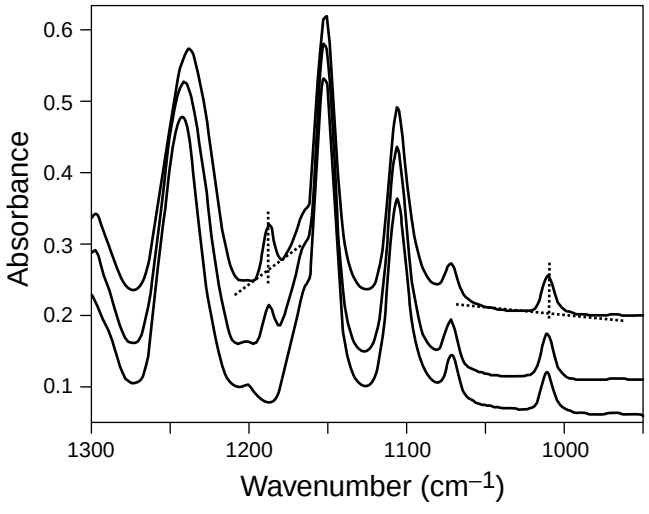

Figure 16 Infrared absorbance spectra recorded from solution (1,1,2,2-tetrachloroethane) cast thin films onto KBr polished plates of a poly(aryl-ether-sulphone) (PES), bottom trace, and two aryl-ether-sulphone/aryl-ether-ether-sulphone copolymers, middle and upper traces. Copolymer compostion may be determined from a measurement of the absorbance ratio, 1190 cm^{-1}/1010 cm^{-1}.

7.7 Polymerization and Cure Studies

Infrared spectroscopy is a valuable tool for studying and monitoring polymerization processes and cure mechanisms. In some cases, time-dependent or temperature-dependent kinetic profiles may be derived by simply using a dispersive or filter spectrometer set to monitor continuously at a characteristic fixed wavenumber the absorption intensity of a functional group, which is either associated with a particular monomer starting material or attributable separately to a product. However, it is now more usual to use a rapid scan FTIR system to record sequential time-lapse spectra, from which measurements may be determined of functional group concentrations at predefined time intervals.

Although prior to ca. 1990 there were frequent reports of the use of infrared spectroscopy for monitoring bulk polymerization processes, these presented considerable

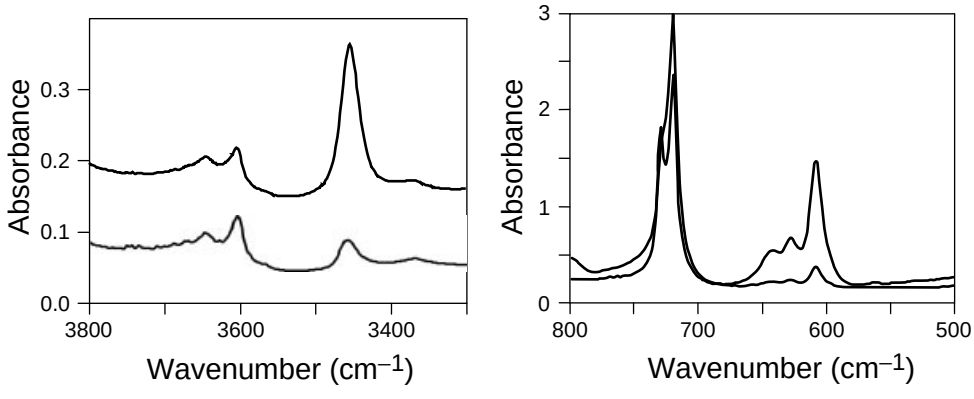

Figure 17 FTIR absorbance spectra recorded from compression molded films (ca. 100 μm thickness) of two EVA copolymers containing 7.7 wt.% and 28.6 wt.% VA units. Potential bands for compositional determinations are the VA absorptions near 3460 cm^{-1} or near 610 cm^{-1}.

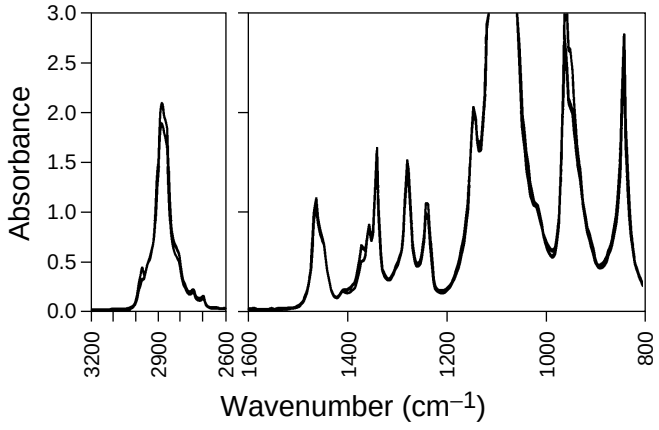

Figure 18 Overlaid FTIR/ATR absorbance spectra, each in duplicate, recorded from two different EO/PO copolymers, containing ca. 72 wt.% and 84 wt.% EO. (Reproduced from J.M. Chalmers, N.J. Everall, 'FTIR, FT-Raman and Chemometrics: Applications to the Analysis and Characterisation of Polymers', *Trends in Analytical Chemistry*, **15**, 18–25 © (1996), with permission from Elsevier Science.)

practical difficulties. Not least of these was that one had to try to mimic conventional laboratory or production synthesis conditions while undertaking the polymerization and spectroscopic measurements in temperature and pressure controlled narrow path length transmission or internal reflection liquid cells. Many of these problems were overcome with the introduction of immersion probes for use in the mid-infrared region. These probes, based on the internal reflectance principle, can be coupled to FTIR spectrometers through transmitting conduit (hollow metallic light guides). They may be used for the real-time analysis of liquid or melt, either batch or continuous, reactions, and adapted for use with both laboratory-scale or plant vessels.[171] Figure 19(a) and (b) show schematics of a probe sensing head construction and an experimental laboratory scale arrangement.[171]

By contrast, many thermal cure mechanisms may be investigated readily by transmission or internal reflection infrared measurements from thin films cast from resin solutions. Thermoset resins based on cure of

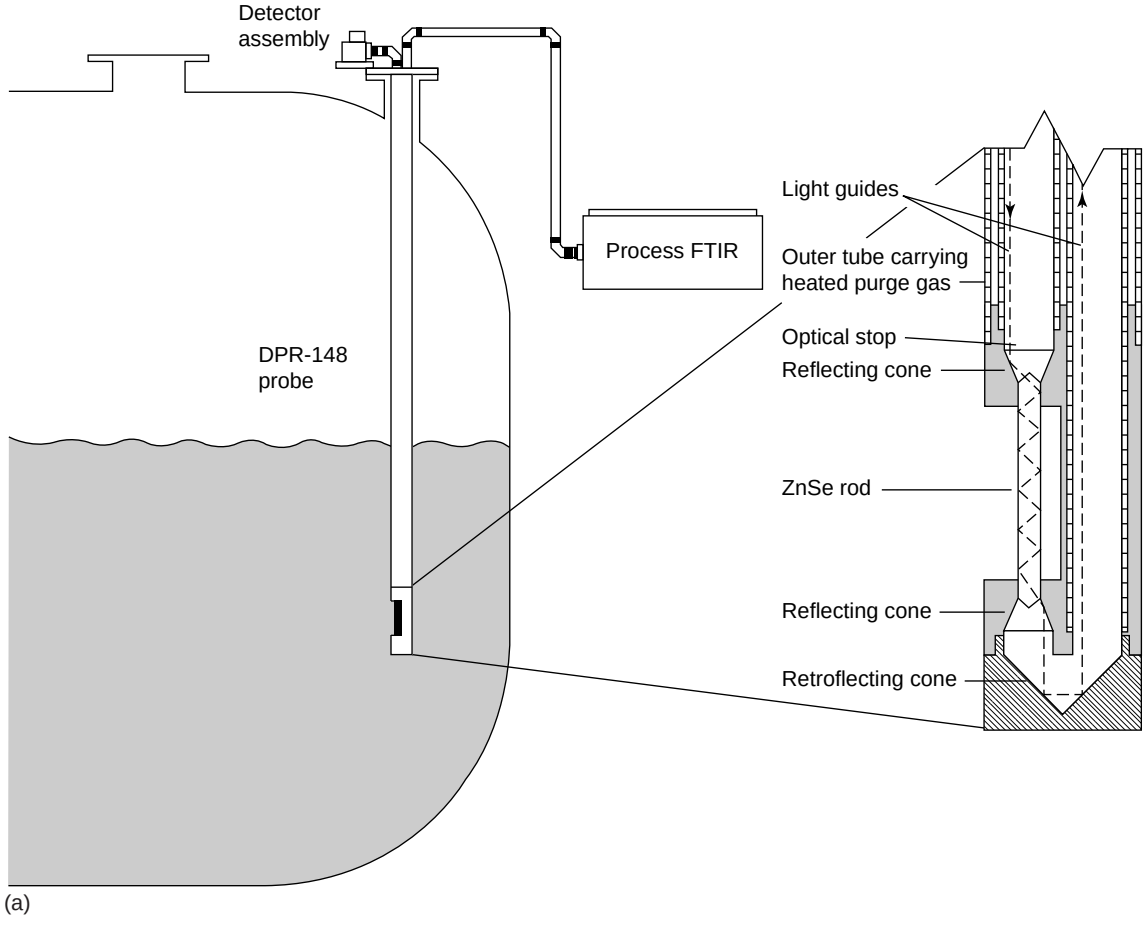

Figure 19 (a) Schematic of an immersion probe head and an immersion probe mounted in a batch reaction vessel. (b) Schematic of an immersion probe in a laboratory scale experimental arrangement. (Reproduced from Doyle[171] by kind permission of the International Society for Measurement and Control. Copyright © ISA, 1990. All rights reserved.)

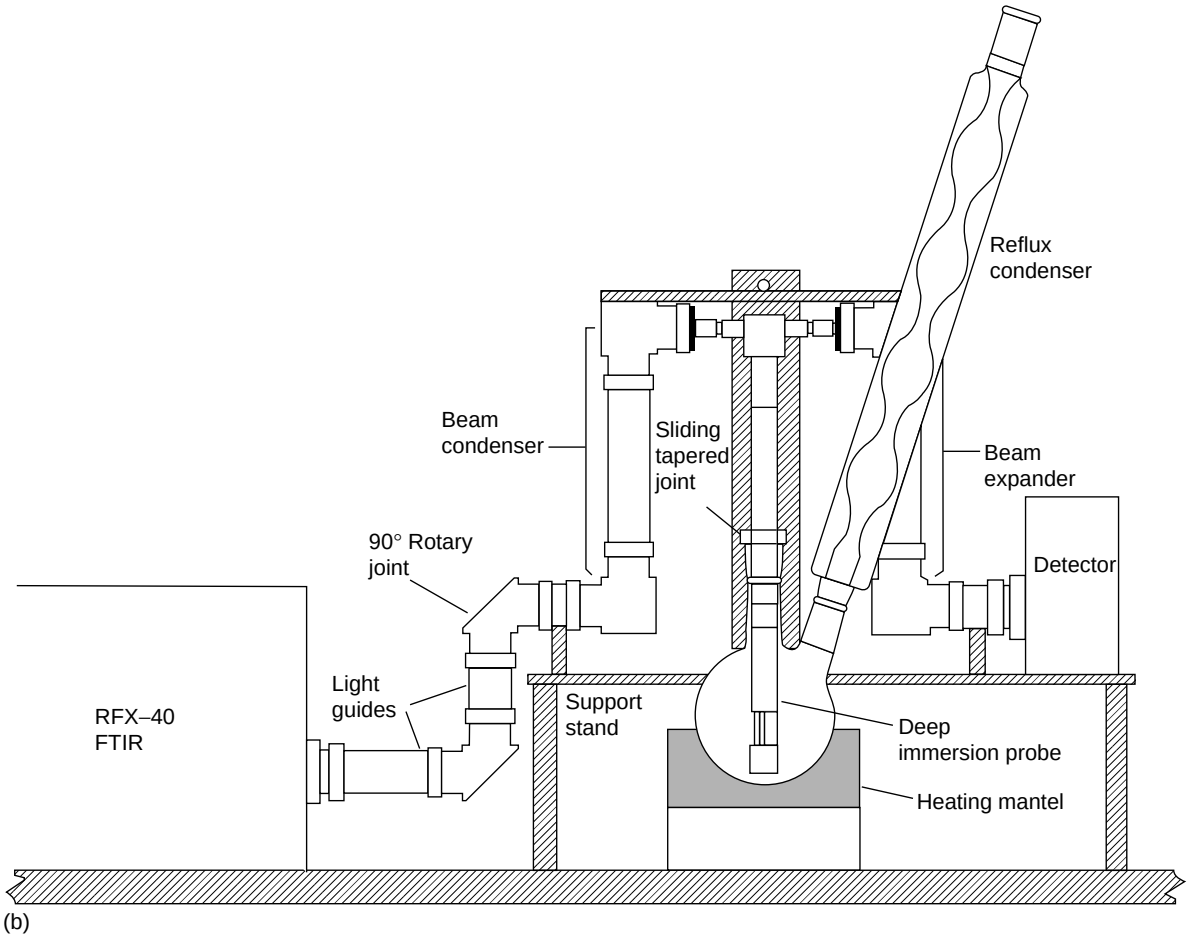

Figure 19 (*Continued*)

epoxy resin formulations are a very important class of materials industrially, particularly as matrices in advanced materials such as fiber-reinforced composites. Their properties depend on the degree of cure and structure of the epoxy matrix, both of which may be investigated beneficially by infrared spectroscopy.[172,173] Homopolymerization and cross-linking reactions may be observed through decrease of the absorption intensity of bands associated with the glycidyl group, most commonly that near 910 cm^{-1} attributed to the epoxy ring,[174–176] although the position of this characteristic vibration of the oxirane ring varies with ring substitution. Additionally, many other changes of absorption intensities of characteristic functional group bands, such as those of hydroxyl, amine, ether, anhydride and ester may be readily observed and related to the mechanisms of cure and cross-linking imparted by the resin formulation components.[177,178]

Polyimides are another important class of high-temperature resistant polymers. As with epoxy systems, because of their functionality infrared spectroscopic analysis can provide considerable insight into polyimide formation and cure. Reaction mechanisms of an aromatic dianhydride or multi-basic aromatic acid (e.g. pyromellitic acid) with an aromatic diamine to form the polyamic acid may be characterized readily. Cyclization to imide formation during thermal curing of this prepolymer and subsequent cross-linking may also be observed; increasing ring closure to imide is readily seen by increasing intensity of its $\nu_{sym}C{=}O$ band near 1780 cm^{-1},[179–182] see Figure 20.

PUs are formed essentially from the reaction of di- (or poly-) isocyanates with polyhydric alcohols, or, in the case of polyester or polyether urethanes, with polymers having terminal alcoholic groups. Their commercial range extends from simple linear aliphatic elastomers, through two-component cross-linkable aromatic–aliphatic coating formulations, to complex resins involving many ingredients designed when mixed to produce foams with different cellular structures and physical properties. Infrared has been used extensively to both analyse the molecular structure of PUs and probe the

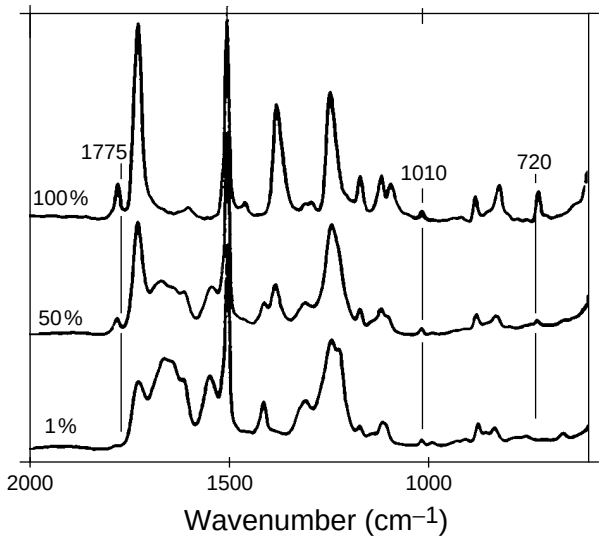

Figure 20 Infrared absorbance spectra of a spin-coated thin film onto a NaCl disk of a polyimide (pyromellitic dianhydride + 4.4′diamino diphenyl ether) at three stages of cure. (Reproduced from Snyder[179]. (Permission applied for).)

chemical reaction rates and morphology associated with cross-linking and foaming processes. Residual isocyanate may be determined quantitatively from a measure of the band intensity of the N=C=O stretching band at 2270 cm^{-1}.[183] In addition to this absorbance, infrared analyses usually focus on three other spectral regions: the νN−H region between 3500 and 3200 cm^{-1}, the Amide I (or carbonyl) region between 1800 and 1600 cm^{-1}, and the Amide II region below 1600 cm^{-1}.[140,184,185] These regions, while very informative, may not be straightforward to evaluate.[140,185] They often consist of many overlapping bands, since, for example, the reaction processes and products of a water-blown foam may include species such as urethanes, ureas, amides, allophanates, biurets, carbodiimides and isocyanurates, which associate eventually to form a polymeric network with a multiphase molecular morphology.[184] Similar complexities also apply for urea-based resins with formaldehyde, for which similar infrared analysis considerations can aid elucidation of structure and cure mechanism.[186]

Special cells that incorporate multiple IRS crystals have been used to monitor the chemical species changes that occur during PU foaming in order to help understand foam behavior.[187,188] Time correlations of isocyanate conversion and evolution of urethane, urea and hydrogen-bonded species were developed from infrared absorbance measurements.[187] Experimental measurements have also been made in which the sensor is a declad portion of a mid-IR transmitting fiber-optic. This is coated or surrounded by resin and thermoset cure monitored. The unclad portion of the fiber provides for evanescent wave (internal reflection) spectroscopy.[189–192] A variety of fiber materials have been used, and several resin, composite and prepreg cure systems sensed in situ. These include both polyimide/graphite and epoxy/graphite systems[191,192] and epoxy adhesives and a urethane resin.[192] A similar approach was utilized to provide for measuring in situ, in-mold kinetics in reaction injection molding (RIM) processes.[193]

Another key area for infrared analysis of cure chemistry is that of measuring the degree of cure and studying cure behaviour of UV curable resins. Real-time (dispersive IR, set at a fixed wavenumber)[194–197] or quasi-continuous (FTIR)[198–200] in situ radiation cure monitoring has been reported for many acrylate based resin formulations, in which both the unreacted monomer and/or residual photoinitiator levels may be determined. Degree of cure and conversion profiles may be assessed by the decrease in intensity of absorption bands of the acrylic monomer vinyl group, notably the νC=C near 1640 cm^{-1} and the C−H deformation mode near 810 cm^{-1}. Transmission,[198] internal reflection[199] and reflection–absorption (transflectance)[200] approaches have been reported, in which the UV cure was carried out in situ in the spectrometer sample compartment. The time-slice resolution of FTIR observations was of the order of 0.5 s.

7.8 Degradation and Oxidation

The sensitivity of mid-infrared spectroscopy to oxygenated functional groups makes it a particularly useful technique for monitoring the molecular species formed or undergoing change in polymers as a consequence of thermal, thermal-oxidative, photo, photo-oxidative and radiation induced damage or modification, whether they be surface or bulk effects.[201] Particularly rewarding has been the study of the degradation processes of aliphatic hydrocarbon polymers, especially polyethylene. For polyethylene, hydroperoxide formation, a primary step (induction period) in the autocatalytic process of polyolefine oxidation, may be observed as an absorption feature at 3555 cm^{-1}, prior to the generation of a variety of carbonyl and hydroxyl species.[202–205] These have absorption band envelopes in the 1750–1690 cm^{-1} (νC=O) and 3600–2800 cm^{-1} (ν−OH) regions respectively. Accompanying these oxidative pathways are changes in the relative intensities of absorption bands associated with unsaturation groups, most notably absorption bands that occur between 1000 and 880 cm^{-1}. Thermal oxidation leads primarily to a reduction of the band characteristic of vinylidene unsaturation (>C=CH_2) at 888 cm^{-1}. Photo-oxidation is characterized by an increase in vinyl end groups (−CH=CH_2), which is associated with a pair of bands at ca. 990 and 910 cm^{-1}, the latter being about

three times the intensity of the former. Absorbance subtraction has been employed to highlight subtle differences in degradation behavior for polyethylene when γ-irradiated in oxygen and nitrogen atmospheres.[206] Quantitative estimation of the absorbance species summing to the ν−C=O band envelope in oxidatively degraded polyethylene has been based on a band deconvolution procedure.[207] Further confirmation of some of the species formed or elucidation of certain of the mechanisms of degradation may be achieved through derivatization or treatment with particular reagents. For example, acid carbonyls in oxidized polyethylene may be converted to carboxylate salts by alkali treatment, with a consequent shift of the saturated carboxylic acid peak at ca. 1710 cm^{-1} to a peak characteristic of a $-C(=O)O^-$ at ca. 1610 cm^{-1}, while SO_2 treatment has shown that hydroperoxidation probably takes place in the chain in the α-position to a pendant methylene (vinylidene) group.[204] Ammonia, SF_4 and NO were all used to derivatize thermo-oxidatively-degraded heterophasic ethylene–propylene copolymers to facilitate understanding of the degradation pathways,[208] and NH_3, SF_4 and NaCl chemical treatments were all used to aid analysis of structure modifications introduced during photo-oxidation of anhydride-cured epoxy resins.[209]

An FTIR analysis, including the use of FTIR-microscopy, of photo-oxidized films of polybutadienes with differing relative concentrations of stereoisomers suggested that stereospecificity per se had little effect on the kinetics of photo-ageing. Irradiation of the film sets up extensively oxidized, cross-linked surface layers, which prevent further ingress (diffusion) of oxygen, and the core of the films remain largely unmodified.[210] However, all these films had high concentrations (>55%) of 1,4-double bonds. In contrast, for a 1,2-polybutadiene (84% 1,2-) no photo-passivation occurred and a much lower rate of photo-oxidation was observed, accounting for its photo-degradability.[211] These examinations involved monitoring and profiling by infrared spectroscopy absorption bands due to carbonyl, hydroxyl and unsaturated groups. In another study of thermo-oxidative degradation, FTIR/ATR was employed to examine thick-walled cylinders of vulcanized natural rubber materials.[212] Again, the observed ageing processes were dominated by the formation of an adventitious oxidized layer, which having low permeability to oxygen protected the interior material. Derivatization was used to aid an investigation of the γ-irradiated initiated oxidation of a random acrylonitrile–butadiene copolymer rubber.[213] SF_4 converted all carboxylic acids to acid fluorides, resulting in a νC=O band shift from ca. 1740 cm^{-1} to 1841 cm^{-1}; NO converted −COH groups to nitrites, while hydroperoxide groups were converted to nitrates, with corresponding changes observed in the infrared absorption features.

In an FTIR investigation into the influence of the radiation dose and the hydrolysis conditions on the chemical structure of PTFE, transmission, diffuse reflectance and photoacoustic sampling were used.[214] The authors concluded that changes observed in the relative intensities of bands between the transmission (bulk phase) and those of the more near-surface sensitive techniques demonstrated that hydrolysis of the −COF groups to −COOH is preferred in the near-surface regions of the irradiated PTFE particles. The former groups were observed at 1884 cm^{-1}, the latter at 1776 cm^{-1} (associated dimer) and 1792 cm^{-1} (H-bonded to water).

The *in vacuo* thermal degradation up to 500 °C behaviours have been compared for a PVA, a poly(methacrylic acid) (PMAA), and a vinyl acetate/methacrylic acid (VA/MAA) copolymer having 65% alternating units.[215] Films of the sample ca. 50 μm thickness were cast onto reflecting stainless steel plates, and their spectra measured by reflection–absorption (transflectance). For PVA, changes in the infrared spectrum beginning at 160 °C may be associated with the formation of −C=C− (3020 and 1657 cm^{-1}) as acetic acid is eliminated. With increasing temperature polyenes are formed, then cross-linking and aromaticization occurs (bands observed at 3020, 1600, 1560, 1520, 1450, 760 and 700 cm^{-1}), with the ester (1740 cm^{-1}) being converted to a ketone (1706 cm^{-1}) above 440 °C. For PMAA, programmed heating leads to formation of anhydride peaks in the infrared spectrum at ca. 1804 and 1765 cm^{-1}. Continued heating leads eventually at 440 °C to a condensed ring aromatic char. For the VA/MAA copolymer, anhydride rings begin to form at 200 °C and then decompose as for PMAA. At 280 °C, lactone formation (1775 and 1203 cm^{-1}) between adjoining VA and methacrylic acid (MAA) units is observed. These have decomposed by 480 °C when the material is again essentially a highly substituted aromatic ring char.

Sequential changes observed in an FTIR thermal oxidative degradation study of nylon at 200 °C showed similar changes for both nylon 6 and nylon 6,6, with a decrease in intensity of every peak related to the amide groups and production of new carbonyl groups.[216] In a photo-oxidative study of these nylons similar spectral changes were observed, suggesting similar degradation mechanisms.[217] Difference spectroscopy was used in both studies, with infrared spectroscopy being used also to identify the gases evolved from the photo-initiated degradations. Comparative infrared studies on thin films of the thermal degradation under reduced pressure of polyacrylonitrile (PACN) and acrylonitrile copolymers with VA, MAA and AA made apparent that the rate of degradation is dependent upon the specific type of comonomer introduced into the acrylonitrile polymer chain.[218] The species monitored was the production

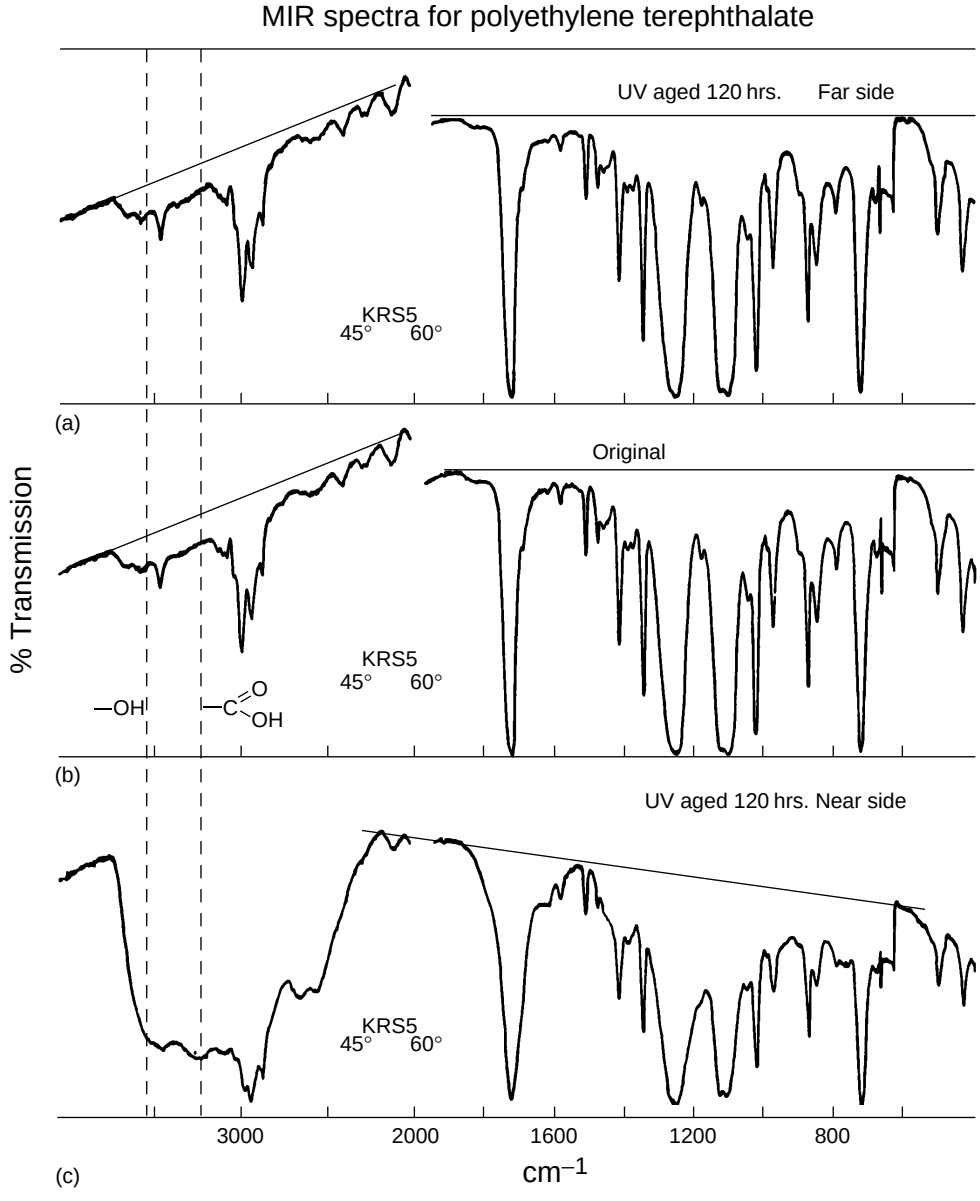

Figure 21 MIR spectra recorded from the surfaces of an artificially UV aged (120 h) PET film. (a) Spectrum of the film surface facing away from the UV source. (b) Spectrum of a surface before exposure to the UV radiation. (c) Spectrum of the film surface facing the UV source. Two KRS5 MIR prisms were used: a 45° angle of incidence set-up to enhance the intensity in the high wavenumber region, and a 60° angle of incidence prism for the region 2000–400 cm^{-1}.

of the cyclized pyridone structure, characterized by the appearance and growth of a doublet at 1610/1580 cm^{-1}, and loss of ν–C≡N at 2240 cm^{-1}. The rate of degradation increased as PACN < ACN/VA < ACN/AA ≈ ACN/MMA. Degradation process mechanisms were elucidated for each copolymer from further infrared studies: ACN/MAA,[219] ACN/VAc[220] and ACN/AA.[221]

The photoproducts of polystyrene irradiated by both short (253.7 nm) and long wavelength (≥300 nm) in atmospheric oxygen have been identified by infrared spectroscopy[222] and profiled through the photoxidized films by infrared microspectroscopy.[223] The products were characterized by a combination of approaches that included difference spectroscopy, derivatization with SF_4 and NH_3, solvent extraction with methanol, thermolysis and photolysis, coupled with model compound studies. Although the products formed at the two irradiation treatments were the same, their relative concentrations were different.[222] A wide range of –OOH, –OH and –C=O containing photoproducts were reported as being identified, both evolved as molecular compounds or as features of the degraded polymer chain. For photo- and

thermo-initiated oxidation of high-impact polystyrene, changes occurring in the FTIR spectra suggested that the first step in the degradation process involved preferential oxidation of the polybutadiene segments that had been grafted onto the polystyrene units.[224]

For surface layer specific products, providing the sample retains a reasonably smooth texture, internal reflectance spectroscopy may be used to highlight surface degradation chemical products.[20] An example is shown in Figure 21. For sealant or protective polymer coatings on metal substrates, reflection–absorption[225] or photoacoustic[226,227] methods may be used conveniently to observe degradation process and accelerated weathering induced infrared spectral changes. Analysis by diffuse reflectance can facilitate certain degradation studies of a polymer–carbon or polymer–glass composite.[228]

8 CHROMATOGRAPHY AND FOURIER TRANSFORM INFRARED

The combination of separation techniques such as liquid chromatography (LC)[229–232] or gel permeation chromatography (GPC) and size (sometimes termed steric) exclusion chromatography (SEC)[232–236] with FTIR analysis has particular value to investigations associated with determining the formulation or compositional heterogeneity of plastic and rubber materials.[229–236]

Organic additives, along with low molecular weight polymer, may be separated from the bulk material by solvent extraction. This extract may then be separated into its components according to their polarity by LC or fractionated according to hydrodynamic volume in solution by GPC. A polymer, a polymer blend or a copolymer composition distribution may be investigated by using SEC/GPC, in which the order of elution is a function of molecular mass. Coupling of these selective techniques with a molecular spectroscopy specificity technique such as infrared spectroscopy can enhance significantly the characterization of a polymer sample, although, because of the infrared absorption intensity of solvent bands, by far the commonest qualitative procedure in each case involves examining the separated fractions after the solvent has been eliminated, rather than in situ in a flow-through cell arrangement. Also, studies are usually limited to separations that do not require the addition of non-volatile buffers, since most of those used are also strongly infrared absorbing.

Various solvent elimination–analyte deposition systems have been developed. The coupling requires that the material presented for IR analysis is deposited onto an appropriate substrate, in a suitable continuous form, in a controlled, synchronized manner, and that it is free from solvent and has not been degraded. Both temperature-controlled spray-jet nozzles and ultrasonic nebulizers have been used as deposition tips. Their effluents are directed to either a linearly translating or rotating, transmission or transflectance support, on which the solute is deposited ideally as a localized, continuous track or series of well-defined spots, with the eluent solvent volatilizing rapidly during the deposition process. The deposited solute is then characterized using either an infrared beam-condensing optics module or under an FTIR-microscope. For transmission measurements, a ZnSe window has been commonly employed as the substrate; while a transflectance system has been commercialized, in which the sample disk is a 2 mm thick Ge disk, which has an aluminium reflective coating on its reverse side.[233,235,236] A schematic of this is shown in Figure 22. Software algorithms, such as the Gram–Schmidt reconstruction, may be used to generate time-based representations of integrated absorption intensities (amount of sample deposited). These will resemble the chromatogram outputs from detectors such as differential-refractive index or UV, but with different relative intensities of the peaks within the profile. Similarly, functional-group profiles as functions of time may be generated. Figure 23 illustrates the types of complementary information that may be recovered from a combined chromatography with infrared spectroscopy analysis of plastics and rubbers. In SEC/FTIR, a drawback is that quantitation at the molecular weight extremes is likely to be very imprecise and problematical, since deposition can change from a thin, continuous film to discontinuous polymer droplets.[235]

In addition to the LC/FTIR characterization of additives,[229–232] the potential has been demonstrated of capillary SFC-FTIR[237] for identifying similarly the components of additive packages, to similar levels of ca. 100 ng. With SFC-FTIR, the volatility of supercritical CO_2 makes solvent elimination simple; deposited fractions may be examined in transmission with an FTIR-microscope system. Diffuse reflectance procedures from material transferred from thin-layer chromatography (TLC) separated fractions may also be employed to fingerprint the solvent extracted components of additive formulations from polymer samples,[238] to levels of 20 μg or less. On-line thermal desorption in series gas chromatography (GC)/FTIR and GC/MS (mass spectrography) has been used to identify building blocks (depolymerization products, residual monomer) in plastic optical fibers.[239]

9 PYROLYSIS AND INFRARED

Implying the chemical structure of a polymer or rubber through depolymerization or decomposition by

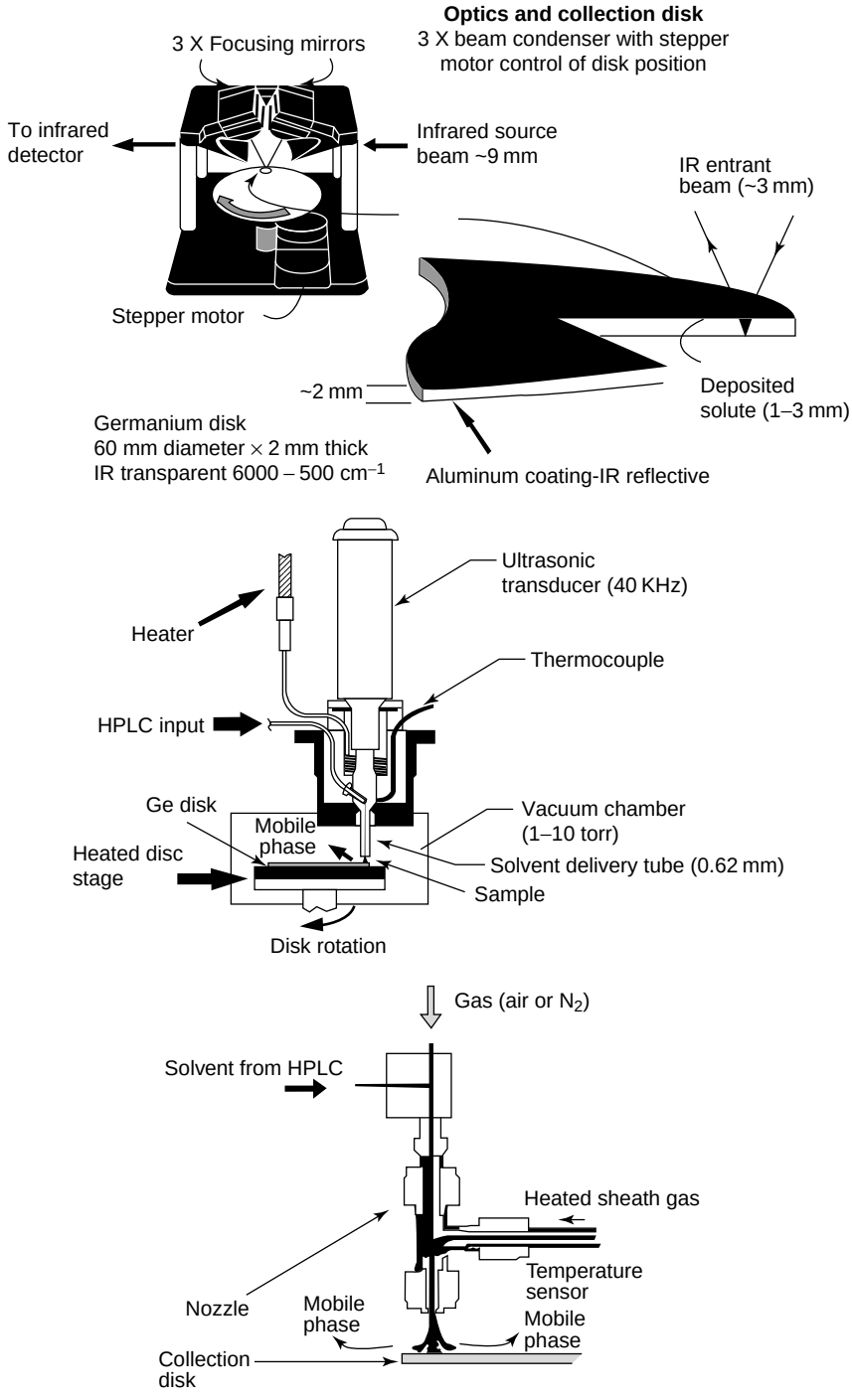

Figure 22 Schematics of a commercial LC/FTIR (LC-transform) optics module and two sample deposition systems, a pneumatic nozzle and an ultrasonic nozzle/vacuum system. HPLC, high-performance liquid chromatography. (Figures reproduced from Willis et al.[236] by kind permission of Gordon and Breach Publishers, Lausanne, Switzerland, © 1997.)

examining the infrared spectra of its volatilization or pyrolysis products is a well-established method. Under controlled conditions, for example, a polyester urethane will cleave to yield the diamine, diol and diacid.[9] While pyrolysate spectra have not always been reproducible, and in many instances their spectra are not explicit enough for unambiguous material identification, these methods have particular value towards identifying intractable, heavily inorganically filled, or cross-linked polymers and rubbers.[2,12] Two approaches are essentially employed.

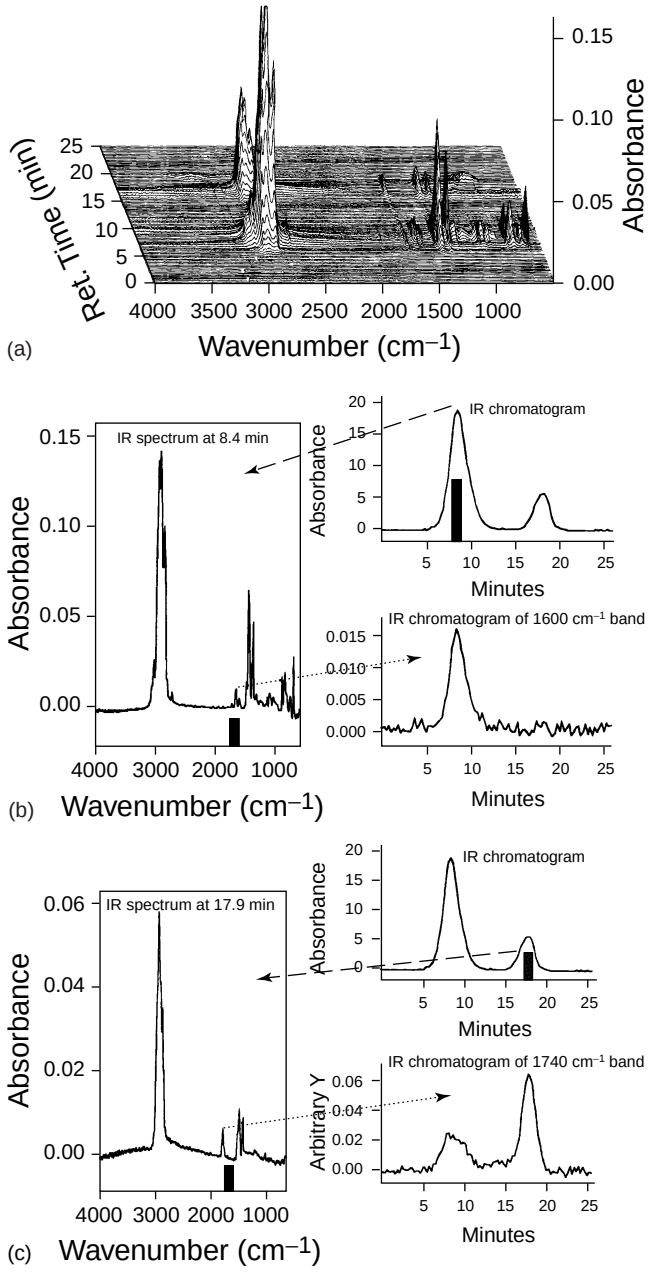

Figure 23 SEC/FTIR. (a) 3D display of absorbance infrared spectra recorded from components separated by SEC of a commercial adhesive. (b) The material that eluted at 8.4 min was identified from its infrared spectrum as a styrene/isoprene block copolymer. This figure also compares an IR chromatogram reconstructed from total absorbance values with that integrated over only the 1600 cm^{-1} band. (c) Similar data for the peak eluting at 17.9 min, characterized as essentially an aliphatic ester. (Figures reproduced from Willis et al.[236] by kind permission of Gordon and Breach Publishers, Lausanne, Switzerland, © 1997.)

The first captures the pyrolysate for a condensed phase analysis; the second examines the evolved gases, either directly or following separation in a GC.

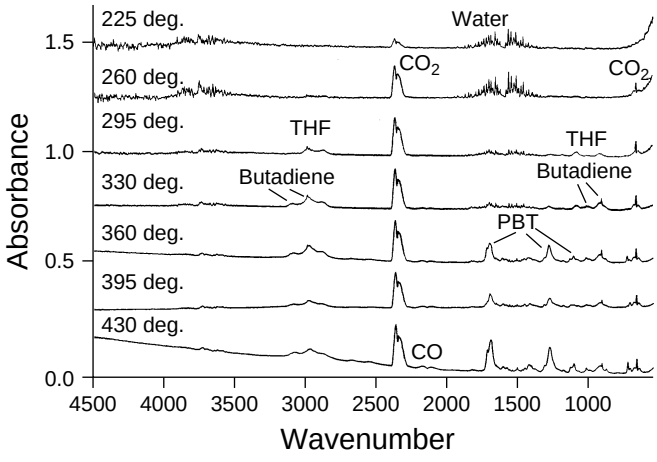

Figure 24 Pyrolysis–FTIR. Comparison of absorbance spectra of pyrolysis products from a compounded PBT, containing 11% of a flame retardent, at different temperatures. Spectra have been offset for clarity and scaled to similar maximum intensity values. (Reproduced from M. Pellow-Jarman, M. Hetem, 'Comparison of the Thermal Degradation Products of Poly(butylene Terephthalate) and Flame Retardant Poly(butylene Terephthalate) Formulations Using a Pyrolysis FTIR Cell', *Polymer Degradation and Stability*, **47**, 413–421, © (1995), with permission from Elsevier Science.)

Several simple experimental set-ups for generating and trapping evolved volatiles and decomposition products have been reported. These have included filament heating at ca. 750 °C in an appropriate tube immersed in an ice bath. Other approaches, with appropriate attention to safe-handling and protection, have been the heating of a small amount of a sample placed at the bottom of either a test-tube[16,240,241] or a glass melting-point tube;[242] the vapors condense at the walls of the cooler ends of these tubes. Pyrolyzed or volatilized, condensed material is then transferred and transmission sampled in an appropriate way such as a KBr disk or as a deposit on a salt window. In the capillary tube method, a high spectral contrast FTIR micro-KBR disk spectrum was reported of the pyrolysate from 100 µg of a black polyester thread.[242] In such experiments, an infrared spectrum recorded from the involatile residues may be useful in producing a spectrum characteristic of any inorganic filler present in the original sample. Fractional distillation has been employed to separate elastomer pyrolysates from organic additives, prior to identifying both the resin and additives by infrared analysis.[243]

Direct-pyrolysis FTIR, in which the pyrolysis probe is positioned inside an appropriate infrared transmission gas cell sited in the sample compartment of an FTIR spectrometer, allows for in situ analysis of the volatiles.[244–246] External shut-off valve ports on the cell allow for purge, atmosphere or sweep gas entry and exit. Both pulse (instantaneous) pyrolysis and time resolved analysis using

programmed heating experiments are possible. Examples of the infrared data generated from a pyrolysis experiment are shown in Figure 24.

Pyrolysers may also be interfaced to a light-pipe for direct evolved gas analysis (EGA) or via a GC column for a GC/FTIR measurement of volatiles.[247,248] For example, the effluent from a thermal decomposition of 1 mg of polymer sample pyrolysed at 700 °C for 10 s in a Curie-point pyrolyser following programmed GC separation may be fed through a FTIR light-pipe system and then onto a GC flame ionization detector (FID), to accomplish in-series combined analyte information.[248] The potential of such a set-up has been illustrated in a comparison of two butadiene–acrylonitrile copolymers of identical acrylonitrile content, one a block copolymer, the other a statistical copolymer.[248] Examination of the FTIR spectral data showed that the two materials were distinguishable, since formation of acrylonitrile dimer was only observed from the block copolymer, whereas higher relative concentrations of mixed dimers and mixed trimers were obtained from the other material, in accord with its more alternating comonomer sequences.

10 EVOLVED GAS ANALYSIS, THERMOGRAVIMETRIC ANALYSIS AND FOURIER TRANSFORM INFRARED

EGA by FTIR is concerned with the real-time identifying and monitoring by infrared spectroscopy of vapors or gases evolved from certain combustion processes. These may, for example, be evolutions from thermal breakdown experiments in controlled environments,[249–251] or, as mentioned in the last section, pyrolysis vapors, or, even fire gases from direct burning of rubbers or plastics.[252,253] In the study by Kallonen[252] the FTIR spectrometer was equipped with a long path gas cell connected to a cone calorimeter. Thermogravimetric analysis (TGA) is a quantitative technique that measures the change of mass of a sample as it is heated. It is commonly used in the polymer industry to determine weight loss profiles against temperature or time of materials, in controlled atmospheres. Gases or vapors evolving from either an EGA or a TGA[254–260] experiment may be passed through a heated flow-through infrared transmission gas cell for characterization. Typically, this cell will be maintained at a temperature between 200 °C and 300 °C, have a length between 6 and 10 mm; it may be multi-passing, and it will be designed to be inert and minimize condensation of materials onto the cell windows. The time-resolved programmed heating experiment described in the direct pyrolysis discussion in the last section could equally be categorized as an EGA measurement. In the development given by McClennen et al.,[258] the

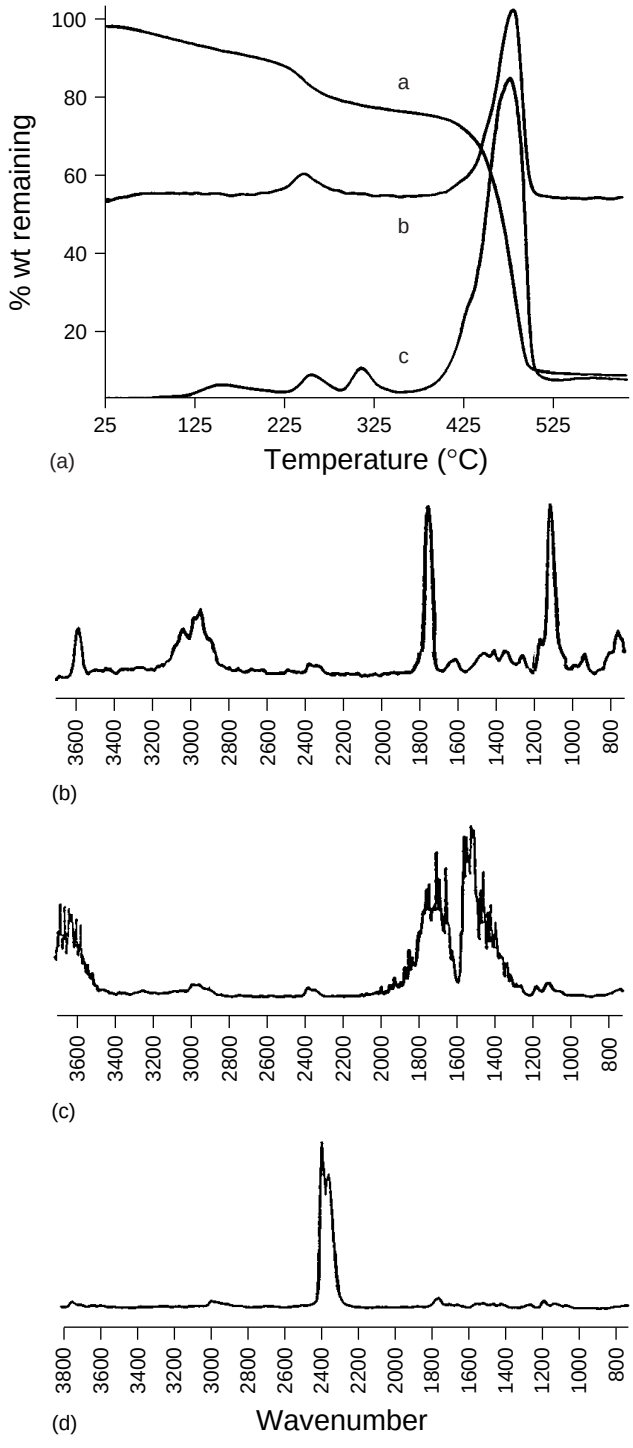

Figure 25 TGA/FTIR. (a): (Curve a) TGA weight loss curve for PMMA heated at 10 °C min^{-1}; (curve b) first derivative of the weight loss curve; (curve c) evolved gas profile for the volatile products generated from total IR absorbances. (b–d) Spectra of low temperature decomposition products from PMMA: (b) MAA at 199 °C; (c) Water at 256 °C; (d) Carbon dioxide at 312 °C. (Figures reproduced from Schild,[256] *Journal of Polymer Science*, © (1993) by permission of John Wiley & Sons, Inc.)

commercial TG apparatus was interfaced to a standard GC/MS/FTIR light-pipe cell system.

An example of the output from a TGA/FTIR analysis is shown in Figure 25.

11 OPTICAL PROPERTIES AND CONSTANTS

Polymers are used extensively in textiles and in applications such as greenhouse and house glazing, thermal screening, horticulture and as coatings to protect reflector surfaces. Some polymers are fabricated as optical components such as lenses and windows for use in the far-infrared region ($300\,cm^{-1}$ to $1\,cm^{-1}$). For these uses it is important to be able to determine the material optical properties (transmittance, reflectance, absorptance) and optical constants (n, absorption index; k, refractive index). Straightforward, practical methods have been derived for determining many of these; others require more specialized instrumental approaches.

The integrated optical properties with respect to thermal radiation have been determined for a variety of minimal-scattering polymer films using conventional transmission and reflection mid-infrared spectroscopy.[261] For measurements on scattering materials an integrating sphere would have been necessary. The radiative properties of natural and synthetic fibers in the near-infrared (0.25–2.5 μm) region[262,263] and PP fibers[263,264] and powders[264] in the mid-infrared have been studied using an integrating sphere. The mid-IR optical constants for films of PET,[265] PACN,[266] polycarbonate,[267,268] polyimide[269] and PVC, polystyrene and poly(methyl methacrylate)[270] have been determined from FTIR experimental reflectance measurements. For determining the far-IR optical constants of polymer samples the technique of dispersive Fourier transform spectrometry can be utilized.[271] In these measurements, the sample is placed in one arm of a two-beam interferometer. Example results have been reported for poly (4-methyl pentene-1), TPX,[272] and a sailcloth material, a woven PET dipped in PU.[273]

Spectroscopic ellipsometry is a well-proven tool for thin and multilayer solid film sample analysis. The optical properties, n and k, as a function of wavenumber may be derived from analysis of the intensity and phase relationship between the s- and p-components of polarized reflected radiation. In FTIR ellipsometry, a sample surface is normally illuminated in a special accessory at an incidence angle of about 70° with radiation polarized at an angle of 45°. The characteristics (geometry) of the elliptically polarized reflected infrared radiation, i.e. the ellipsometric angles Ψ and Δ, can be used to determine the sample dielectric function and film thickness.[22,274,275] Some examples of applications relating to polymers may be found in the studies by Claybourn,[22] Roseler,[274,275] and Vallon & Drevillon.[276]

12 TWO-DIMENSIONAL INFRARED SPECTROSCOPY

Generalized two-dimensional infrared (2D-IR) correlation spectroscopy is a relatively recent, but powerful, method for analyzing time-dependent variations within a set of spectra.[277–279] It is used to effectively highlight the pertinent spectral features associated with fluctuating or transient infrared data. It can provide enhanced spectral resolution, and simplify the interrogation of a complex set of infrared spectra consisting of overlapping peaks, although presently observations are limited in that they may not be interpreted quantitatively.

Dynamic (rheo-optical) 2D-IR spectroscopy using sinusoidal mechanical perturbation has been used to study excitation–relaxation processes in many polymer systems. Examples are given by Marcott et al.,[278] Meier,[279] Steeman et al.[280] and Singhal & Fina.[281,282] In other 2D-IR studies on polymer systems, the perturbations have included the photopolymerization of acrylic and epoxy monomers,[283] in which transient spectra were recorded with a time resolution of 78 ms, and, equal temperature intervals in a thermal 2D FTIR analysis of a PU elastomer.[284]

2D-IR correlation plots are constructed from perturbation-induced time-dependent spectroscopic responses.[277] The perturbation may lead to spectral changes, such as variations in band intensities, shifts in band positions, or changes in band shape. Individual spectral elements may respond differently to the applied perturbation, and their behaviour with respect to each other is highlighted in a 2D-IR correlation map or correlation spectrum. The output from a 2D-IR correlation analysis is commonly presented as a complementary pair of contour plots of 2D spectra constructed from a dynamic spectrum.[277] This dynamic spectrum is usually displayed as a one-dimensional reference spectrum along the two axes of the correlation plot. The dynamic spectrum is typically selected to be the time-averaged, sometimes the static or original, spectrum of the data set. The synchronous 2D correlation spectrum is a representation of infrared signals that are changing in-phase with each other; in the asynchronous plot, independent and uncoordinated fluctuations of the infrared signal are displayed. Simultaneous or coincident spectral changes measured at ν_1 and ν_2 are located in the synchronous 2D correlation spectrum on the diagonal, as autopeaks, or in off-diagonal positions, as cross peaks. Correlation

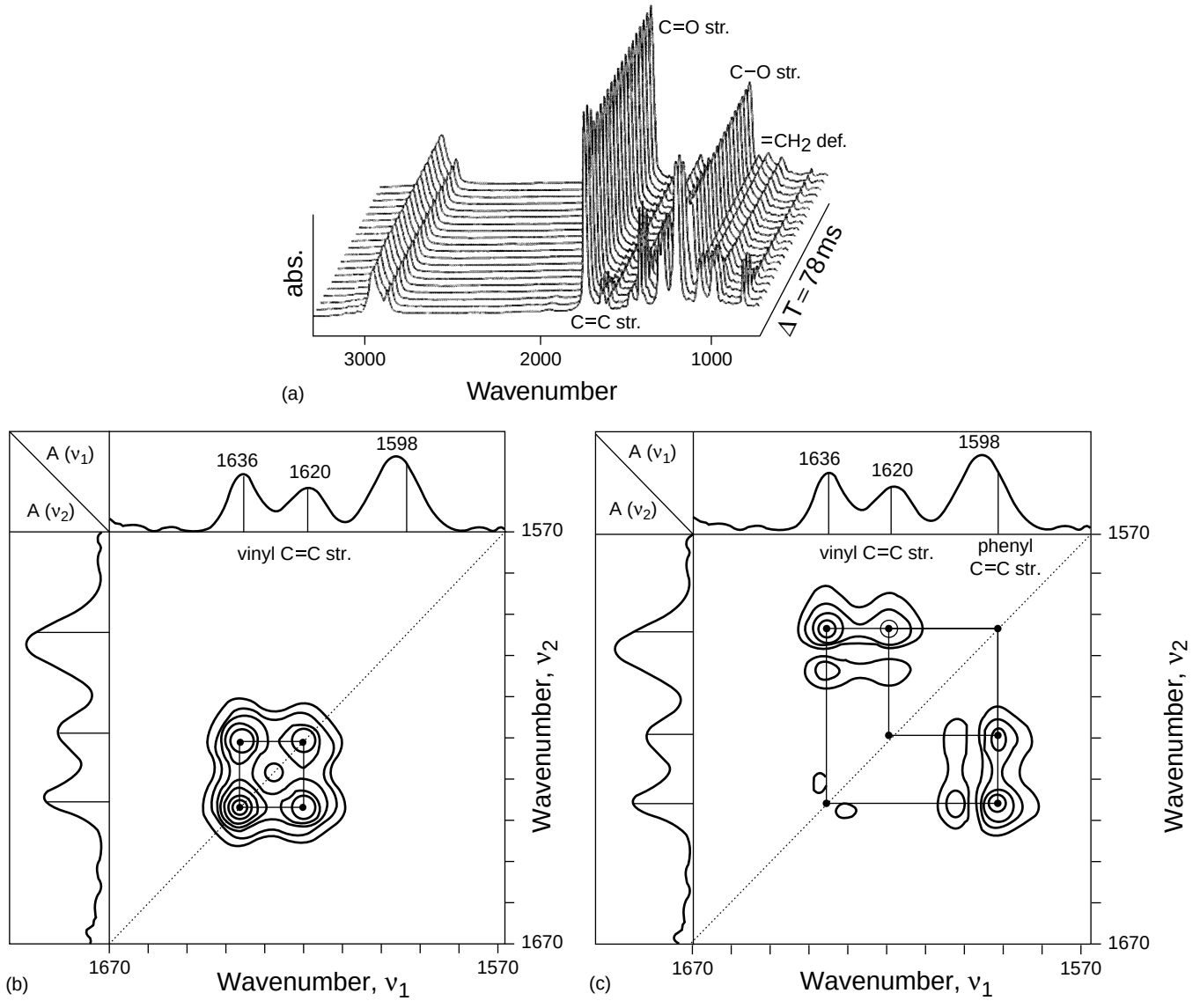

Figure 26 2D-IR correlation spectroscopy. (a) Real-time FTIR spectra collected during the initial stage of the UV curing of 1,6-hexandiol diacrylate (HDDA). (b) and (c) synchronous and asynchronous transient 2D-IR spectra respectively of HDDA obtained in the C=C stretching region. (Reproduced from Nakano et al.,[283] *Applied Spectroscopy*, by kind permission of the Society for Applied Spectroscopy.)

squares may be constructed by joining pairs of cross peaks through corresponding autopeaks, in order to highlight coherent spectral variations.[277] A positive sign for a cross peak shows that spectral variations at corresponding wavenumbers are changing in the same manner, e.g. they are either increasing or decreasing in intensity together; a negative sign highlights opposing changes. Sequential or unsynchronized variations are depicted in the asynchronous 2D correlation spectrum. Cross peaks in an asynchronous spectrum develop when two spectral features vary out of phase with each other. A positive sign indicates that the variation in ν_1 occurs predominantly before that of ν_2, the sign is reversed if the ν_1 change occurs after that of ν_2,[277] where ν_1 is displayed as the horizontal axis and ν_2 as the vertical axis. A simple illustratation taken from Nakano et al.[283] is shown in Figure 26. The real time FTIR spectra were collected during the initial stage of the UV curing of HDDA. The transient 2D infrared correlation synchronous plot in the C=C stretching region shows that the two peaks corresponding to the C=C stretching vibrations at 1636 cm^{-1} and 1620 cm^{-1} of the vinyl group change their intensities simultaneously, while, as expected, the asynchronous plot shows that these vinyl group changes are independent of changes in intensity of the band at 1598 cm^{-1}, assigned to the C=C stretching band of the photo-initiator, since its

radical generation and that of the vinylester chain reaction are occurring at different rates. This was a preliminary experiment in advance of studying a more complex system of the photopolymerization of a 1 : 1 mixture of acrylic and epoxy monomers, for which 2D-IR correlation analysis clearly showed that the radical-induced polymerization of HDDA progressed more rapidly than the cation-induced polymerization of the epoxy.[283]

13 PROCESS MEASUREMENTS

Quantitative mid-IR methods have been used extensively within industry for many years for both off-line (remote) and at- or near-line product quality assurance and production monitoring. Typical examples include co- and ter-polymer composition and formulation determinations, end group levels or balance for polycondensations, and residual monomer concentrations, whereas for other than specialized applications, such as film thickness gauging,[285] until recently on-line methods were a rarity. However, several advances have seen mid-infrared instrumentation become increasingly effective and utilized as a means for on- or in-line continuous measurements of polymer production processes (see article **Infrared Spectroscopy in Process Analysis** in this encyclopedia and D.W. Vidrine[286]). There has been a consolidation of multivariate data analysis techniques into the portfolio of quantitative methods commonly applied to infrared spectroscopic data. In addition, there have been important developments in sampling techniques, such as the introduction of a diverse range of sampling probes, although lack of robust, low attenuation fiber optics for the mid-infrared region still poses a major limitation. (A schematic of an immersion probe[171,287] is shown in Figure 19, see section 7.7.) Another commercial probe design uses a diamond ATR sensing head.[288] Also, spectrometers targeted for use at production environments have been increasingly marketed.

Although many mid-infrared techniques have been developed and applied to monitoring industrial processes,[289] the fine details of many process spectroscopy applications remain hidden or buried in patents or are of such propriety importance that they are never disclosed. On-line flow cell FTIR systems have been described and applied to the compositional analysis of polymer melts.[290–294] One such demonstrated the feasibility for the on-line compositional analysis of propylene/ethylene copolymer melt-streams by a process FTIR spectroscopy system coupled with PLS data analysis.[293] In another, silica and erucamide additive levels were determined in LDPE compounds extruded through the on-line analyser.[294] The potential of TIRES[34] has already been mentioned briefly, see section 3.4, and a range of possible on-line applications have been reported.[295–298]

14 MICROSPECTROSCOPY

Many analytical problems associated with polymer and rubber chemistry and technology and product satisfactions benefit from investigation with infrared micro-spectroscopy.[16,82,83,299] FTIR-microanalysis combines the attributes of light microscopy with the advantages of infrared spectroscopy. Chemical structure or physical characterization of small samples or domains selected by light microscopy may be investigated as seen by FTIR, since in an FTIR-microscope the visual and infrared optical trains are collinear and parfocal. Masks in remote image planes define the area identified for spectroscopic analysis. Their aperture size is limited practically by diffraction to a spatial resolution of ca. 10 μm. (By comparison, dispersive Raman microprobe systems are capable of better lateral spatial resolution, typically ca. 1–2 μm.) Maps of inhomogeneity or anisotropy may be built up point by point and displayed as contour or axonometric plots of functional group absorbance intensity.[300] An alternative approach, which utilizes an infrared array detector coupled with a step-scan FTIR spectrometer, allows for spectroscopic chemical imaging.[301] Probe microanalysis infrared spectra may be recorded both by transmission and reflection techniques, including internal reflection and grazing-incidence reflection–absorption.

14.1 Fourier Transform Infrared Microscopy: Sample Presentation

The same criteria that were discussed for macro-sampling in section 3.1 apply for probe micro-transmission measurements by FTIR microscopy. That is, a high spectral contrast spectrum free from optical aberrations will only be obtained from a thin, continuous, homogeneous, uniform thickness sample of an area greater than that of the defined sampling aperture; although very flat, optically transparent specimens may well suffer from some inteference fringing. Samples may be supported on an infrared transparent window mounted on the microscope sampling stage. Frequently used methods for preparing specimens include microtoming and flattening in a compression cell.

Microtoming is commonly employed to generate an appropriate thickness cross-section specimen from a multi-layered or laminate structure. The infrared microanalysis spectrum may then be recorded from each layer successively, after appropriate aperture masking. Best specimens are created using either glass or diamond

blades. Diamond cleaving knives and diamond-edge or carbide-steel micro-planes may be used for scraping off or shaving off respectively thin sections from a sample. Diamond anvil and other types of compression cells provide a very effective means to reducing the thickness of small samples, and flattening and containing them. A roller-knife, which has a polished steel wheel head, may be used to flatten some samples such as certain fibers. It may also be used sometimes successfully for transferring to its polished surface a material that is contaminating the surface of a polymer product, such as an additive bloom; the transferred material may then be characterized in situ on the roller knife by transflectance infrared microanalysis.

Reflection–absorption (transflectance) is a commonly used microanalysis approach, particularly for identifying thin coatings on metal surfaces, such as those on beverage cans. Other thin materials may be supported on a gold-coated metal disk or microscope slide. Specular reflectance microanalysis can also be a very convenient sampling procedure for some samples.[23] The area examined, however, should be essentially optically flat and the depth should be compositionally homogeneous and ''optically thick''. Microscope objectives are available which allow for grazing-incidence reflection–absorption and IRS measurement.[302] The latter is commonly known as an ATR objective; ATR is a commonly used abbreviation for an IRS measurement, mostly applied to a single internal reflection accessory. A range of ATR objective elements is available. Not all are visually transparent and do not therefore allow microscopic visual inspection of the sampling area. Ge elements are normally used with a pressure (contact) alert in order to ascertain intimate sample surface contact for ATR infrared microanalysis measurement.

14.2 Fourier Transform Infrared Microscopy Applications

FTIR microanalysis measurements fall essentially into two types – those concerned with analyzing limited amounts of sample, and those concerned with spatially resolved data. Examples of the former include contaminant analysis, laminate structure characterizations and fiber identifications. The latter may be involved with determining chemical structure gradients, such as those associated with degradation profiles, or measuring physical property anisotropy, such as molecular orientation differences through a thick film or along the axes of a polymer bottle, or determining crystallinity variations across the surface of a molding.

In the following sections some examples will be given of the wide range of analytical uses to which FTIR microscopy has been applied. Many examples of applications to polymer analysis along with discussions of sample presentation methods and criteria may be found in the study by Katon & Sommer,[303] and the guides by the ASTM[304] and Humecki.[305]

14.2.1 Fourier Transform Infrared Microscopy: Contaminant Analysis

Performance criteria and aesthetics both require that polymer and rubber products are free from contamination. FTIR microscopy is a key tool in identifying ''foreign material'' in or on the surface of commercial products. These imperfections must, of course, be of a size appropriate to FTIR microanalysis. Isolated contaminants may be mounted on a KBr window or flattened and contained in a compression cell for a transmission measurement. For a buried contaminant it may be expedient to microtome a thin section through it, so that it becomes exposed within the host material. Surface contaminants may be examined by FTIR/ATR microscopy.

14.2.2 Fourier Transform Infrared Microscopy: Defect Analysis

As with contamination analysis, characterizing the structure of visible defects is important to failure analyses and quality control. Chemical structure or form difference of localized process- or fabrication-induced defects or in-use property deterioration may often be successively determined by infrared microanalysis. The imperfections may arise typically from poor constituent dispersion in a formulation or blend, differences in molecular weight or end groups from the bulk, morphology differences, degraded, uncured or cross-linked material, all of which may often be characterized readily by FTIR microscopy examinations.

Three examples illustrating the benefits of infrared spectroscopic analysis to this type of investigation are shown in Figure 27. All three imperfections were evident as gel-like defects in bulk film; locally their visual appearances were variously described as ''nibs'', ''blobs'', ''gels'', or some similar term. Each arose from material that, because of a different rheological property, processed differently to the bulk. The first arose because of poor dispersion of a rubber component in a polymer blend.[16] In the second investigation a deficiency in an aryl–Cl end group concentration in a PES was highlighted from 100 μm spatially resolved spectra recorded from a microtomed section from a PES film.[82] The third investigation showed that the defect was associated with amorphous polymer in a matrix of crystalline PEEK.[16] The defect was isolated from the bulk sample and flattened in a diamond window micro-compression cell for infrared analysis.

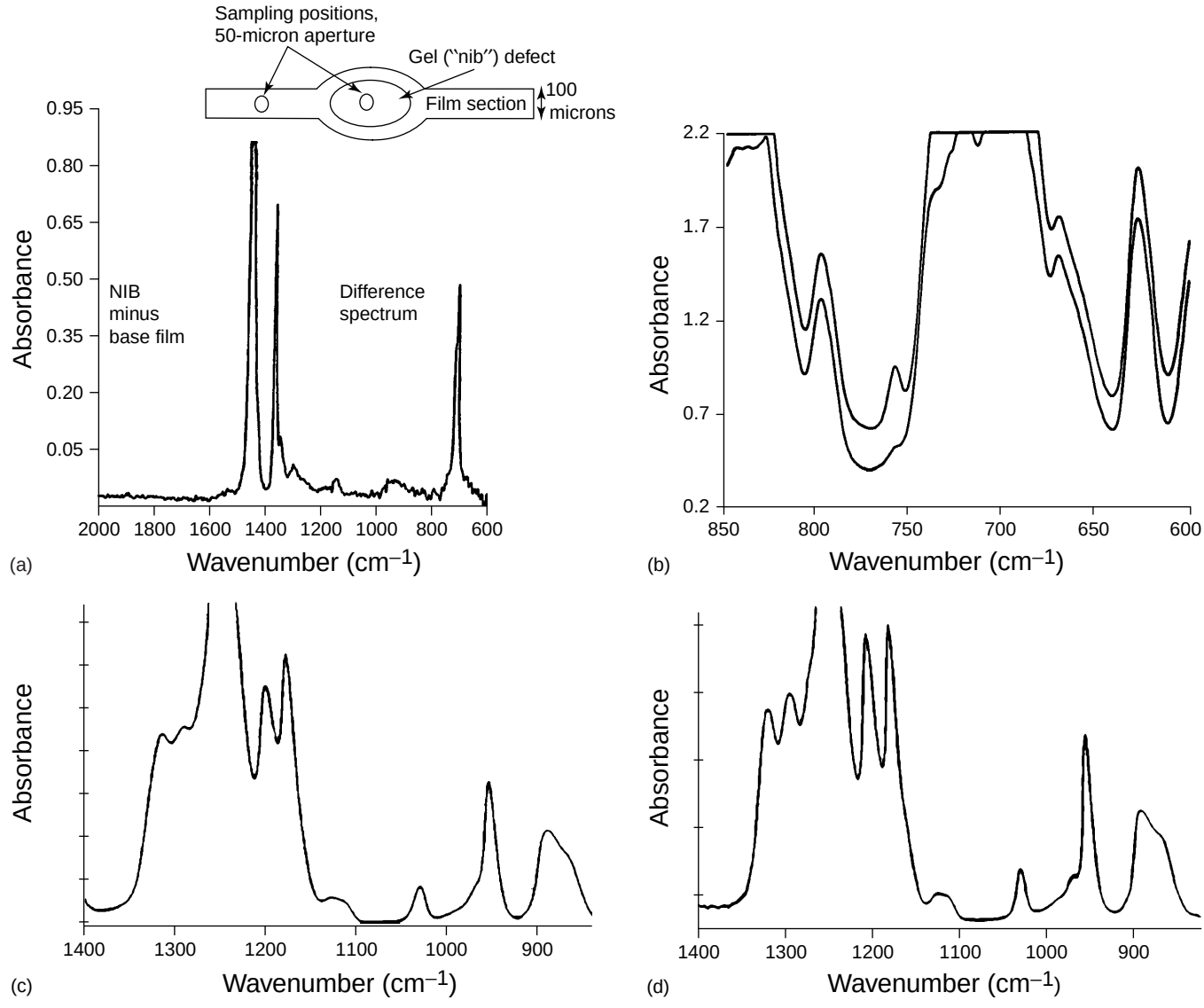

Figure 27 FTIR microscopy, defect analysis. (a) Schematic of a microtomed section from a 100 μm thickness film made from a blend of a PP and an ethylene/propylene rubber, which had a poor profile, exhibiting prominences ("nibs"). Infrared transmission spectra were recorded from the areas indicated using 50 μm aperture masking. The absorbance difference spectrum generated between these spectra, shown in the figure, shows clearly that the "nibs" were a consequence of poor dispersion of the rubber phase. (b) FTIR absorbance spectra recorded from a microtomed section of a PES film, aperture 100 μm. The lower spectrum is representative of the defect area, the upper spectrum is that of the base film. The band at 760 cm^{-1} is attributable to the aryl-Cl end group. (c) FTIR absorbance spectrum recorded from a defect isolated from a PEEK polymer film and compressed in a diamond window micro-compression cell. The spectrum is close to that of amorphous PEEK. (d) Reference spectrum recorded similarly from a section of the matrix semi-crystalline PEEK polymer film.

14.2.3 Fourier Transform Infrared Microscopy: Laminate Structure Analysis

Spectroscopic fingerprinting of the layers in laminated film products is perhaps one of the most advertised and a commonly used application of FTIR microscopy in polymer analysis. The procedure requires that a thin cross-section is microtomed at right angles to the surface of the multi-layer film.[16,82,83,306] (Oblique sectioning may be used to expose wider layers.) Successive layers of the film are then examined, after appropriate remote aperture masking, by transmission FTIR microscopy. Diffraction effects usually limit the recording of "pure" layer spectra to those layers with a width of 10 μm or greater. Thinner adhesive layers may sometimes be spread wider by squeezing the section in a compression cell.[16] An illustration of an analysis of part of a multilayer sample is given in Figure 28.

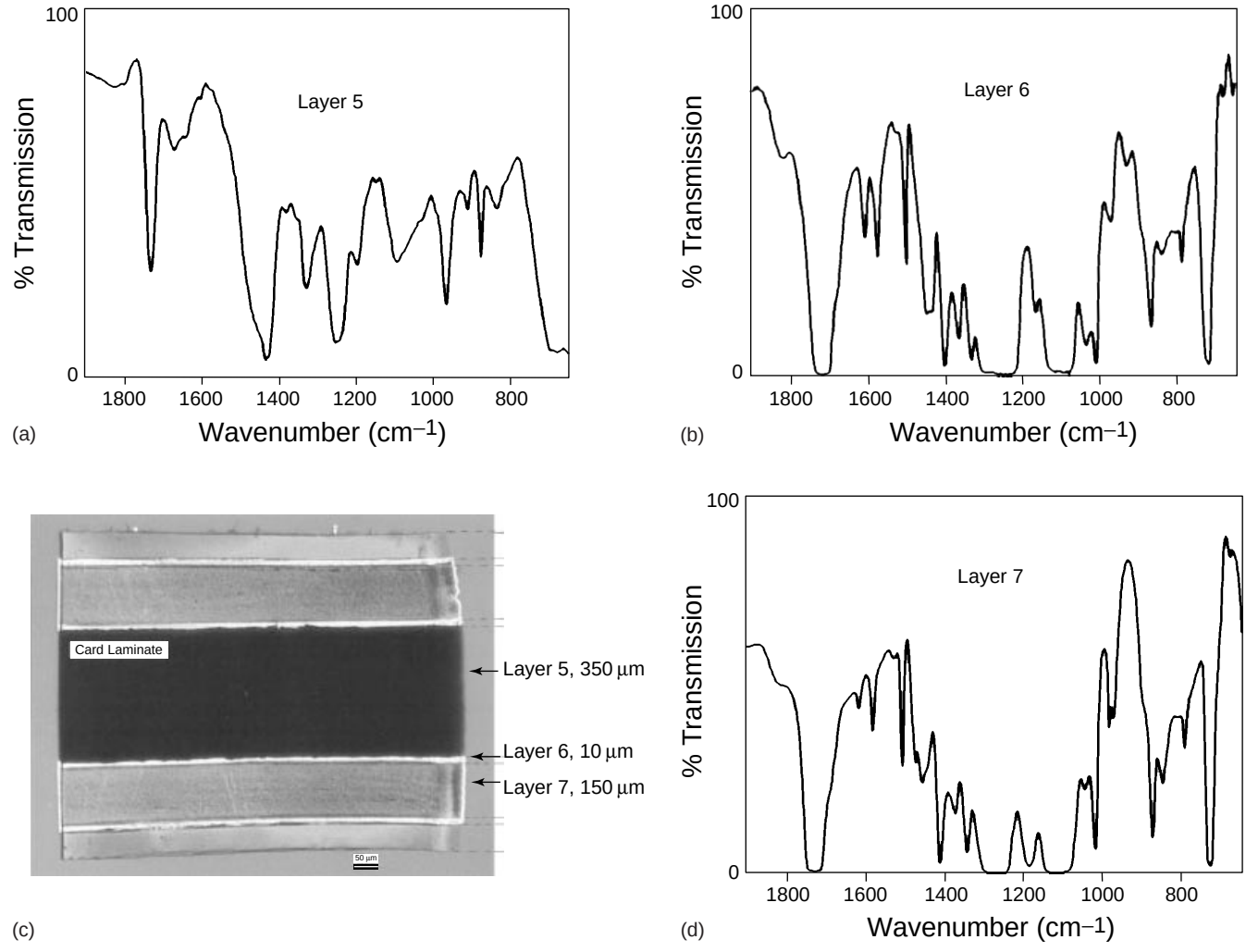

Figure 28 FTIR microscopy of a 20 µm cross-section from a multilayer polymer card sample. (a) Transmission spectrum of layer 5, an ester plasticized PVC, probably filled with TiO_2; (b) transmission spectrum of layer 6, an amorphous terephthalate/isophthalate copolyester; (c) image as seen in transmitted visible light through crossed polars; (d) transmission spectrum of layer 7, a barium sulphate filled PET.

14.2.4 Fourier Transform Infrared Microscopy: Failure Surface Analysis

Surface layer defect and failure surface analysis may be accomplished with FTIR/ATR microscopy. For example, ATR microanalyses performed on failure regions resulting from a mechanical shearing test on two steel sheets bonded with a PU adhesive revealed a distinct inhomogeneity in the distribution of the hard PU segments in the soft PU segment rich matrix through the adhesive layer thickness.[307] A greater concentration of hard segments was observed in a rupture region near the adhesive/steel interface compared with that of a cohesive failure surface localized in the bulk of the adhesive layer. Evidence for this was given by the relative intensity of the νCH peaks to those of νNH and νC=O in the two failure regions, and the differences rationalized as perhaps resulting from hydrogen bonding developed between urethane groups and metallic oxides and hydroxides on the metal surface.

FTIR/ATR microscopy spectra from another multilayer film failure analysis are shown in Figure 29. The surface layer spectra from the initial delamination region (''mechanical shock induced'') show clearly that adhesive failure is taking place at the interface between the heat seal layer, as evidenced by the copolyester peak at 930 cm^{-1}. In contrast, the surface layer spectra recorded from each exposed surface from the propagated delamination region (''tension failure'') support cohesive failure, since the two spectra are essentially identical and that of PET. This diagnosis is shown schematically in Figure 29.

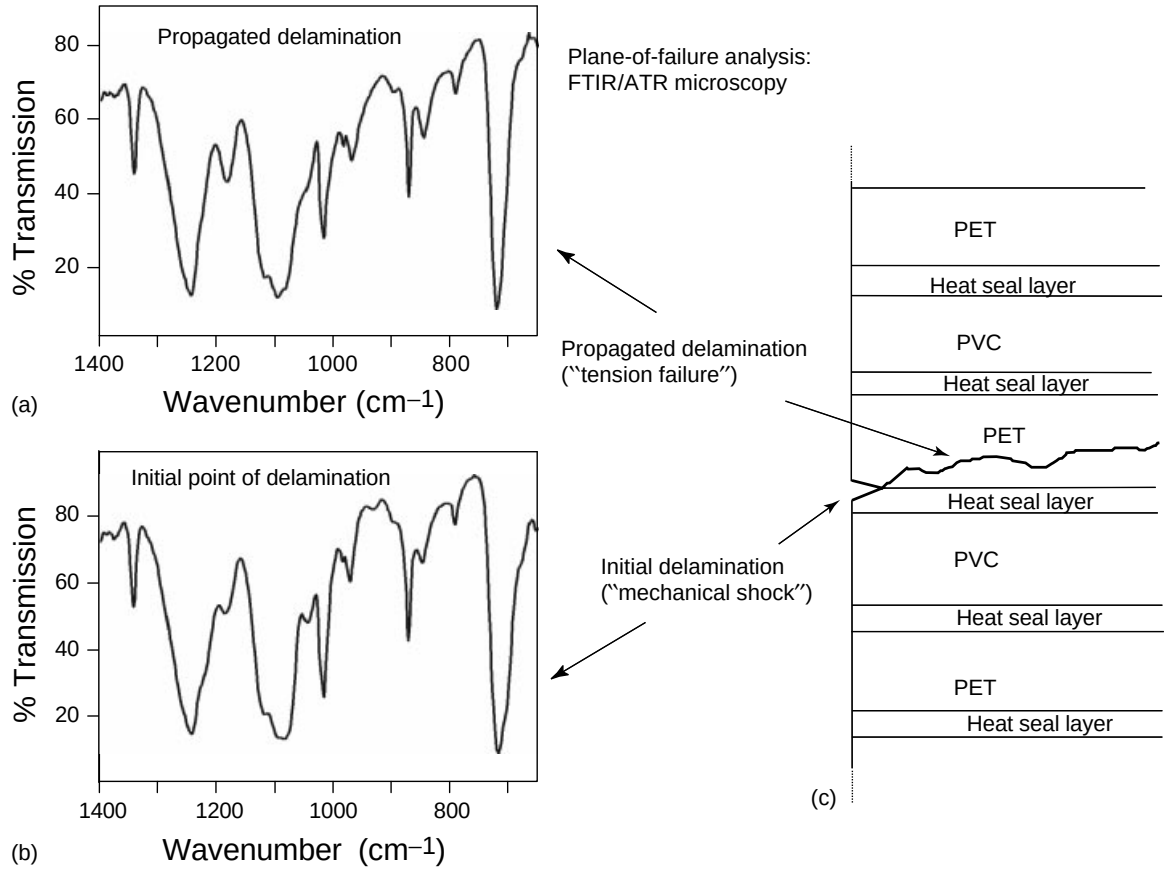

Figure 29 FTIR microscopy, plane of failure analysis of ID card. (a) Infrared transmission spectra of exposed surfaces at section of propagated delamination; neither of these spectra show the band at ca. 930 cm^{-1}, indicating that tension failure has continued through the PET layer, and not at the layer interface. (b) Infrared transmission spectra of exposed surfaces at initial point of delamination; the weak absorption band in the upper spectrum at ca. 930 cm^{-1} that is not present in the lower spectrum indicates that the initial delamination occurred between the PET and the copolyester heat seal layer. (c) Schematic of multilayer card structure.

14.2.5 Fourier Transform Infrared Microscopy: Fiber Analysis

It is possible to characterize fibers by FTIR/ATR microscopy; indeed, by carefully controlling the contact pressure both a thin sheath polymer layer and the polymer core of a heterofil fiber have been identified.[16] It is much more usual, however, to record transmission FTIR microscopy spectra from fiber specimens that have been flattened by either a tool like a roller-knife[308,309] or metal probe[310] or contained within a compression cell.[309,311] Such preparation procedures are, however, very likely to alter the physical characteristics of the fiber, such as level of crystallinity and degree of molecular orientation. However, if the fiber is not flattened its transmission FTIR microscopy spectrum will likely be severely distorted and of poor spectral contrast. Spectral variations will also be minimized if fibers that are oriented are mounted always in a similar direction in relation to the microscope horizontal axes.

14.2.6 Fourier Transform Infrared Microscopy: Chemical Composition Profiling and Mapping

Gradient profiling or mapping of chemical functional group intensity variations through or on the surface of a polymeric product is a widely used application for FTIR microscopy. Anisotropy may be visualized from stack plots generated from point-by-point measurements along a line profile or from axonometric or contour plots generated from normalized intensity values from consecutive point-by-point readings taken across a pre-defined area. These intensity variations may be processed and displayed as false-colour or gray-scale images to aid visualization. Point-by-point, equidistant step displacements can be achieved with an automated motorized $x-y$ microscope stage.

Figure 30 illustrates the preparation procedure for creating a specimen from an aged polymer sample. Figure 30 shows also a degradation profile generated from the illustrated stack plot obtained from examining

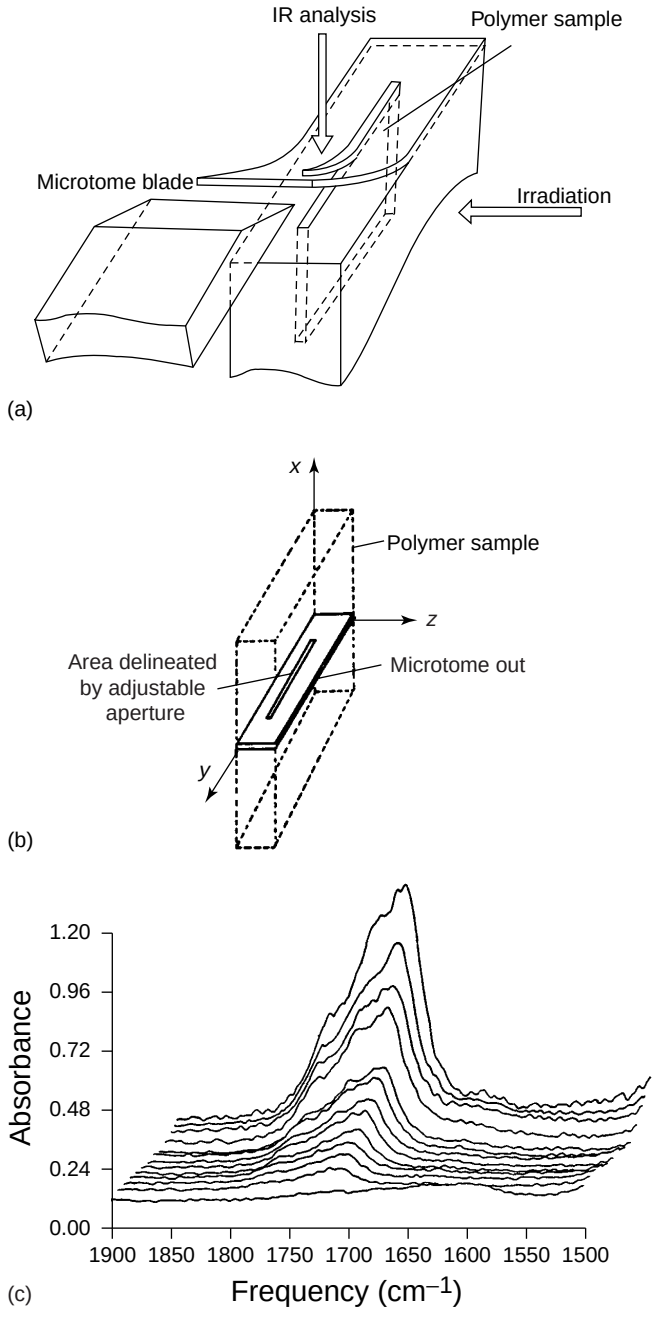

Figure 30 FTIR microscopy of aged polymer films. (a) Schematic showing preparation of thin microtome cross-sections. (b) Schematic representation of the analysis method. (c) Infrared absorbance spectra in the carbonyl stretching region of the first layers of a photo-oxidized PP sample. (Reproduced by permission of *Spectroscopy Europe* from Jean-Luc Gardette, 'Micro-FT-IR Spectroscopic Profiling of Aged Polymer Films', *Spectroscopy Europe*, **5**(2), 28–32 (1993), © *Spectroscopy Europe*.)

15 μm wide sections of a PP sample irradiated in the presence of air under artificial photo-aging conditions.[312] Similar analyses have been undertaken for PVC [313–315] and PMMA.[315] The spatial resolution of an FTIR microscope has been used to monitor in a similar way the time-dependent diffusion-in profiles of a UV stabilizer into a PP plaque.[316] Automated line-scan measurements were again made on microtomed sections. The sampling aperture width was set to 26 μm, with a stage step size advancement of 26 μm of the microscope stage between recordings of spectra. Differences in polymer chain structural entities may also be located. For example, FTIR microscopy has been used to highlight differences between end group (hydroxyl and carboxyl) concentrations between the edges and centers of sections microtomed from a series of PET granules taken from various stages within a solid-phase polymerization cycle.[83,293] Significant concentration differences were observed for the two positions following the solid stage polymerization step.

In another utilization of FTIR microscopy to polymer analysis, an essentially time-profiling room temperature cure study was made of the interphase regions of a PE and Poly(vinyl Alcohol) (PVAL) fiber/epoxy matrix composites.[317] Spatially resolved spectra of dimensions ca. 20–40 μm × 80–150 μm, perpendicular and parallel to the fiber axis respectively, recorded from close to the fiber and 200 μm from the fiber axis were compared. From infrared analysis, coupled with DSC measurements, among the deductions made was that moisture from the fiber surface increased the anhydride consumption in an epoxy/anhydride matrix, but the final extent of reaction (ester production) is reduced near to the fiber. This was in contrast to an epoxy/amine matrix study, in which the final cure state was unaffected by water from the fiber surface.

14.2.7 Fourier Transform Infrared Microscopy: Physical Characteristic Profiling and Mapping

Physical property anisotropy may be mapped in similar ways to those used for profiling gradients in chemical composition. For example, localized measurements have been made by FTIR microscopy of polymer conformation and morphology through fusion welds.[318–320] Figure 31 is a schematic diagram of a hot plate pipe weld, which indicates a microtomed section taken for an analysis (200 μm for PE; 50 μm for PP) and the 188 × 50 μm FTIR microscopy analysis points.[318] In two independent but similar investigations on PE,[318,321] both of which used the intensity of the 1303 cm^{-1} band of polyethylene as an indicator of amorphous content, but which used different absorbance ratio procedures, the crystallinity in the weld region was shown to be less (by ca. 2.5%) than that in the parent pipe, and minimizing at the weld centerline, see Figure 31. In contrast, in similar experiments, the helical content of PP has been determined to increase through the weld region.[318–320]

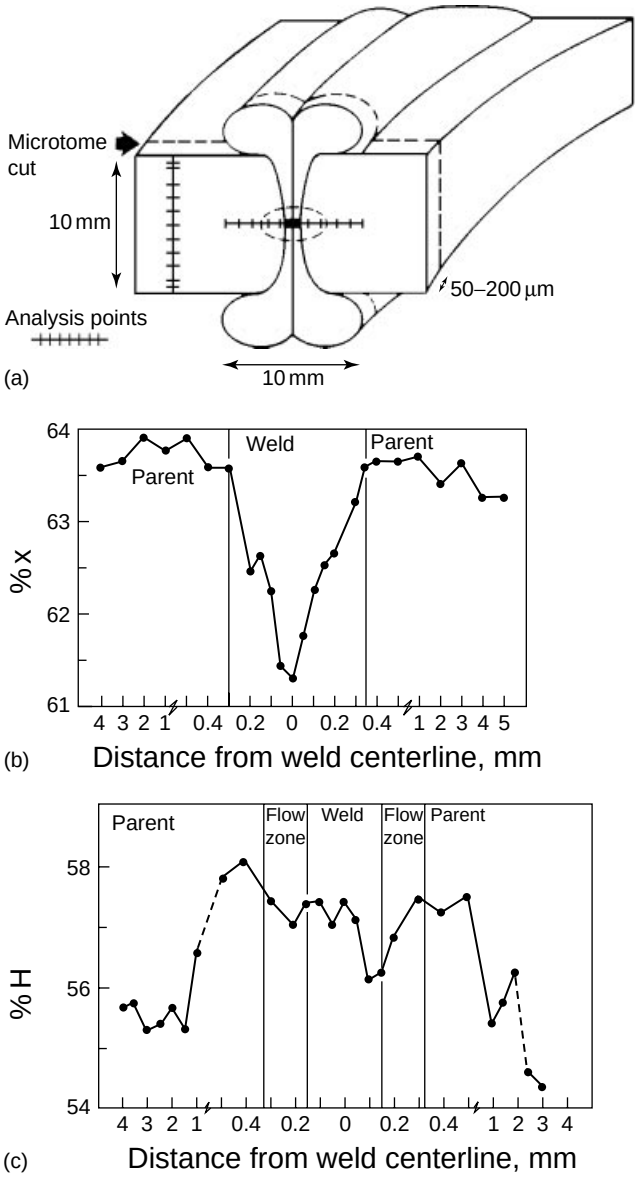

Figure 31 FTIR microscopy of polymer welds. (a) Schematic diagram of a hot plate weld. (b) Infrared determined crystallinity profile through polyethylene weld. (c) Infrared determined helical content profile through PP weld. (Figures reproduced from Stevens[318] by kind permission of TWI, Cambridge, UK.)

Polymorph microdomains in PVdF films cast from different solvents have been characterized by FTIR microscopy.[322] Different zones for infrared analysis were selected from optical micrographs, and spectra associated with the α, β and γ forms of PVdF obtained, although the presence of all three was not determined for all the films; for example, the film cast from tetrahydrofuran (THF) appeared to be completely constituted by the α form. Polarized infrared radiation FTIR microscopy analysis of drawn extruded strands of a liquid-crystalline polyester revealed that for a thicker strand (diameter 1.52 mm) the dichroic ratio was low near the strand core, but increased markedly towards the surface.[323] For thinner strands, the molecular orientation function (dichroic ratio) increased significantly compared to that of the thicker strand, and increased with decreasing diameter. However, for the thinner strands the radial differences were small. The sections examined were microtomed in the extrusion direction to a thickness of 10 μm, from strands that had been embedded in an epoxy resin. The aperture length for the FTIR microscopy examination was 200 μm in each case, having a width of 40 μm for the narrowest strand and 80 μm for the thickest strand. FTIR microscopy dichroic ratio measurements in two transverse mutually perpendicular sections along sample geometric axes of biaxially oriented polymer products may be exploited to gain insight into molecular orientation anisotropy. FTIR microscopy molecular orientation measurements, together with molecular conformation and molecular configuration determinations, have been used to investigate PET thick film curl, and profile PET bottle wall properties.[16,82,83]

14.2.8 Fourier Transform Infrared Microscopy: Imaging Applications

As mentioned in section 14.2.6 and implied in section 14.2.7, equidistant functional group normalized absorbance intensity measurements may be used to generate maps to aid visualization of composition or property variations. These maps (images) may be presented in a variety of ways. They can be pseudo three-dimensional (axonometric) plots of normalized absorbance intensity versus $x-y$ plane distances. These can be reduced to planar line contour maps, which may essentially be gradient false-colour or gray-scale coded to better highlight features and improve contrast. These maps are generated from automated, equidistant point-by-point measurements across a pre-defined sample area step-scanned by means of a motorized microscope $x-y$ stage.[324,325] Alternatively, direct gray-scale or color infrared images may be generated from spectra recorded from a global examination of the pre-defined sample area by using an FTIR step-scan spectrometer fitted with a focal plane multichannel array detector.[326,327]

A combination of difference spectra and constant thickness normalized contour plots generated from FTIR microscopy measurements have been used to make observations on the interphase region of cured epoxy resin–glass fiber reinforced composites.[328,329] Samples thin enough for transmission measurements were prepared from single 10 μm thick glass fiber monofilaments embedded in epoxy resin; the total composite sample thickness was 12–14 μm. In Figure 32

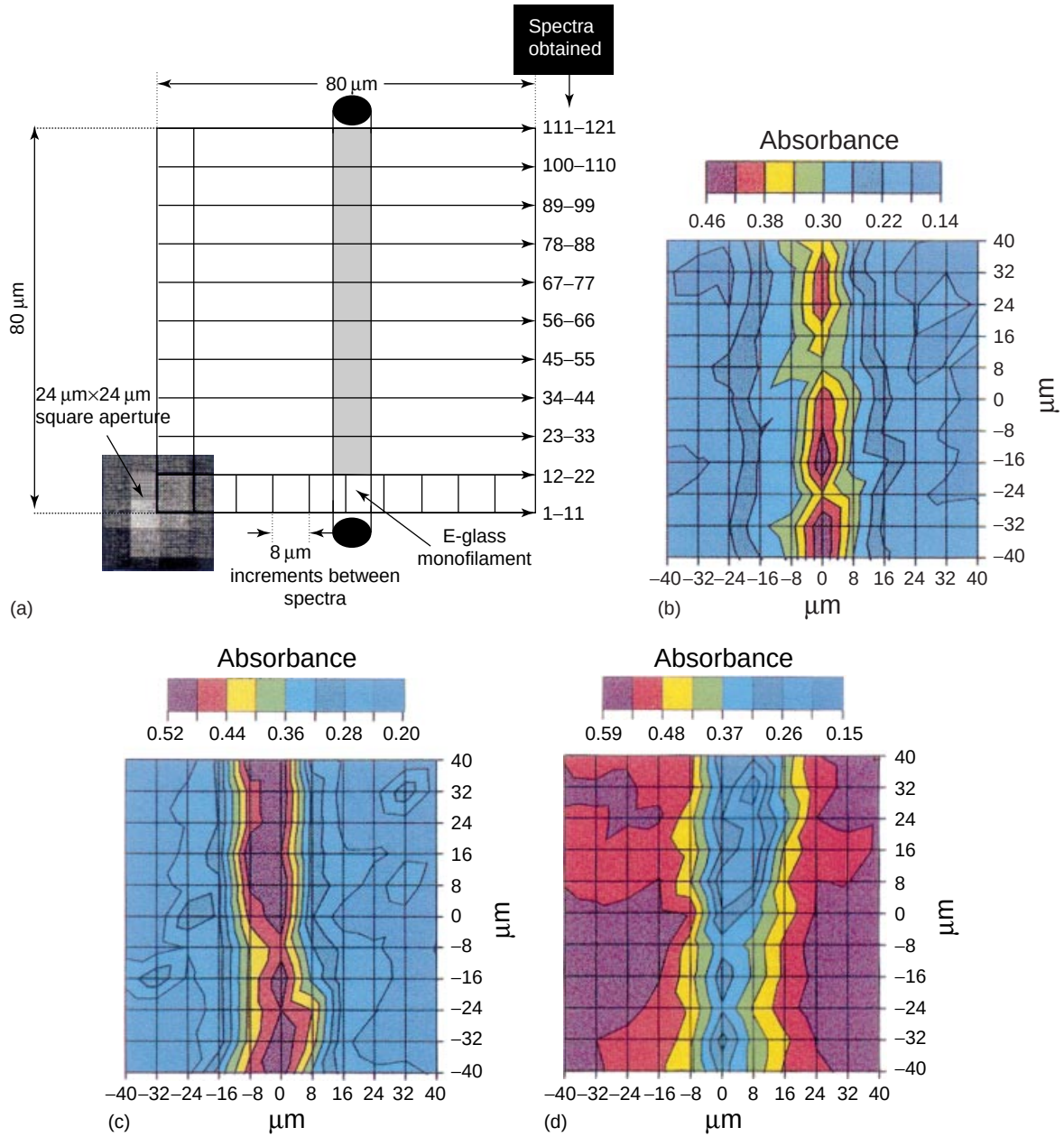

Figure 32 (a) Schematic representation of the geometry of a mapping experiment. (b) Contour plot for the 841 cm^{-1} peak for samples prepared with as-received glass fibers. (c) Contour plot for the 842 cm^{-1} peak for samples prepared with heat-cleaned glass fibers. (d) Contour plot for the 841 cm^{-1} peak for samples prepared with glass fibers treated with γ-aminopropyl-triethoxysilane (γ-APS) coupling agent. (Figures reproduced from Arvanitopoulos and Koenig,[329] from *Applied Spectroscopy*, by kind permission of the Society for Applied Spectroscopy.)

a schematic representation is shown of the geometry used for the localized transmission mapping measurements together with example contour plots generated for NH wagging mode absorbance at ca. 841 cm^{-1}. This peak revealed a higher intensity along the fiber surface for the as-received and heat-cleaned glass fibers, but showed a lower intensity value at the interfacial region for a fiber treated with the coupling agent, γ-APS. In accord with these observations, difference spectra for which the spectrum of the bulk matrix (40 μm away from the fiber center) was subtracted from that of the fiber embedded in the matrix showed a negative feature at ca. 841 cm^{-1} for γ-APS system, whereas it was positive for the fibers not treated with the coupling agent. These suggested that the uncoated glass fiber surfaces attracted the amine curing agent of the epoxy agent, but a preferential

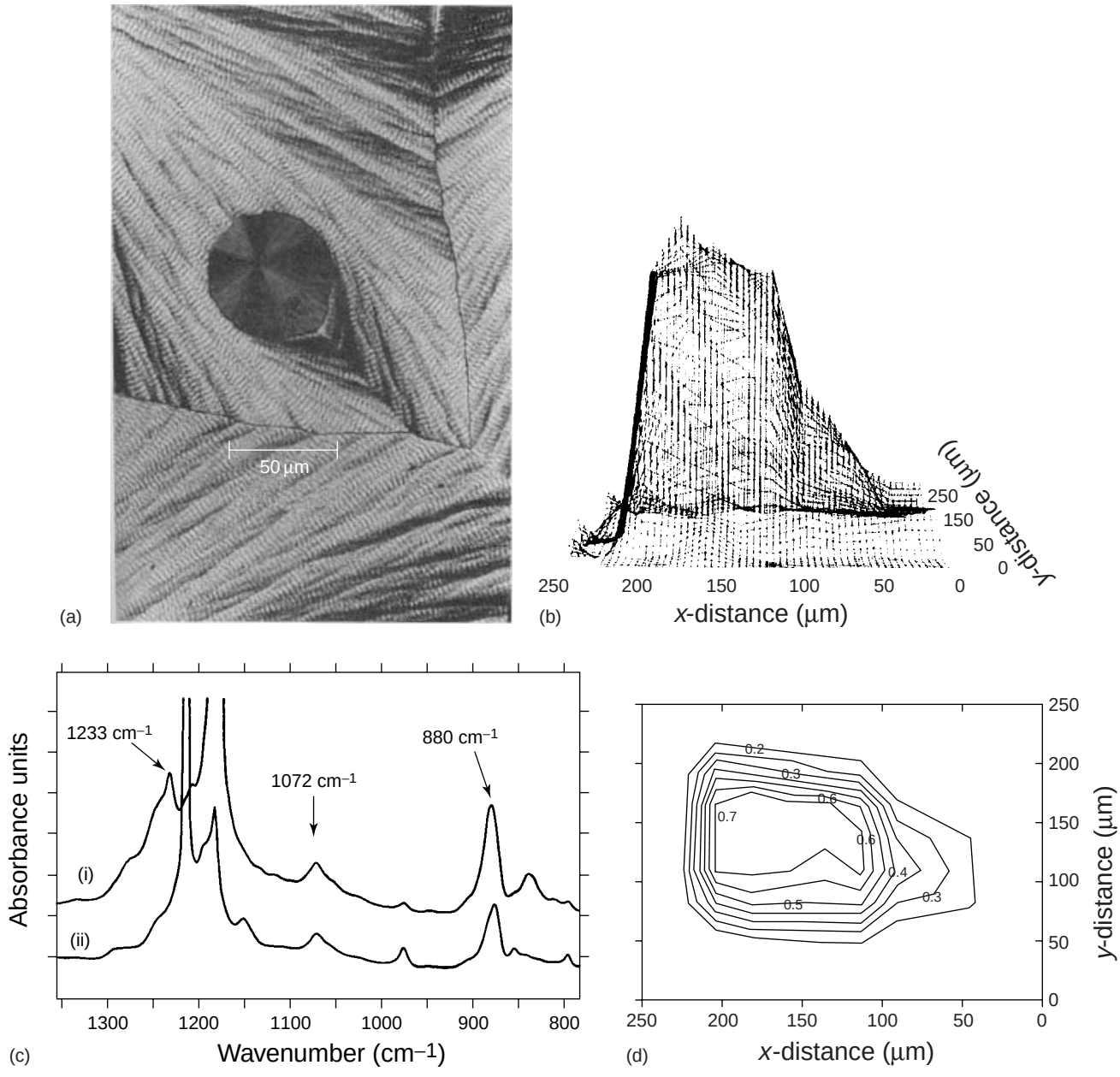

Figure 33 (a) Polarized optical micrograph of area of PVdF spherulites used for FTIR microspectroscopy mapping. (b) Axonometric (functional group image) plot of the CF_2 stretching vibration at 1233 cm^{-1}. (c) (i) Point FTIR microscopy absorbance spectra taken in the area of the small spherulite, and (ii) point FTIR microscopy absorbance spectra taken in the area of the large banded spherulite. (d) Contour (functional group image) plot of the CF_2 stretching vibration at 1233 cm^{-1}. (Figures reproduced from Kressler et al.,[330] from *Applied Spectroscopy*, by kind permission of the Society for Applied Spectroscopy.)

repulsion occurred in the presence of γ-APS. Other specific interaction features of the interphase regions were also observed, which were associated with the different fiber surfaces. When exposed to moisture, the lowest water uptake at the interfacial region was shown to be at the silane-treated fiber, and at a certain level water was found to catalyse a polymerization reaction at the interphase.[329]

Point-by-point mapping images have been used to characterize the morphology and phase behavior of polymers[330] and polymer blends.[330,331] Figure 33 shows the polarized optical micrograph of PVdF spherulites isothermally crystallized at 160 °C.[330] This figure shows also functional group images as axonometric and contour plots, generated by the 1233 cm^{-1} band, a characteristic of the γ-modification. The step size for mapping

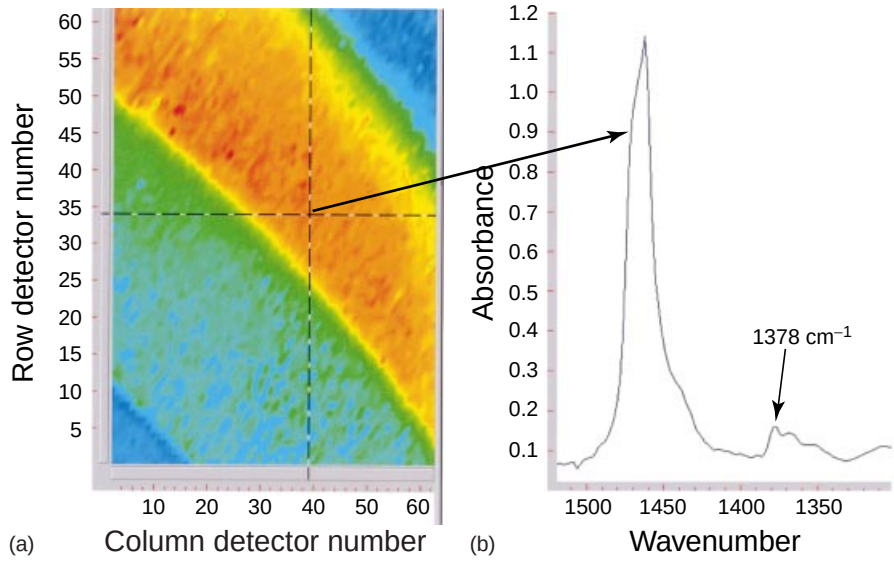

Figure 34 (a) FTIR false-colour image of bilayer poly(ethylene) tube section based on absorbance intensity at 1378 cm^{-1}. (b) Absorbance spectrum associated with pixel at cross-wires.

was 25 μm. The large banded spherulite belongs to the α-modification. Characteristic single-point FTIR spectra of the two regions are shown also in Figure 33. Spatially resolved images have been generated for polymer-dispersed liquid crystal systems[332] and the time dependent concentration profiles of the liquid crystal (E7) into poly(*n*-butyl methacrylate)[333] from FTIR microscopy point-by-point mapping measurements. The diffusion images were constructed from the normalized intensity of the cyano peak (2226 cm^{-1}) of the liquid crystal,[333] whereas the carbonyl intensity of the polymer was used to differentiate matrix and droplet regions.[332] Both spatially resolved, multipoint dichroic ratio measurements around a crack and images from FTIR microscopy measurements have been combined in an assessment of molecular structure in the crack tip region of unfilled natural rubber.[334,335]

Figure 34 shows the 1378 cm^{-1} FTIR microscopy image plane recorded from a 20 μm thick cross-section microtomed from a bi-layer polyethylene tube. The image was generated from data recorded with a step-scan FTIR spectrometer fitted with a focal plane array detector.[336] The detector was a 64 × 64 mercury cadmium telluride (MCT) array detector. The image area displayed is about 450 μm × 450 μm. Also shown in the figure is a spectrum associated with a pixel (ca. 7 μm × 7 μm) from the outer layer region. The image clearly separates the two polyethylenes according to their methyl content; there is a distinct boundary between the inner and outer layers, which have methyl contents per 1000 C atoms of ca. 6 and ca. 25, respectively.

ACKNOWLEDGMENTS

I owe a huge debt of gratitude to many past and present colleagues, friends and collaborators at ICI plc and the vibrational spectroscopy community at large. They have over the last 30+ years taught, explained, inspired, demonstrated, shared experiences and discussed and debated with me on a multitude of topics related to the infrared analysis and characterization of polymers.

ABBREVIATIONS AND ACRONYMS

AA	Acrylamide
ASTM	American Society for Testing and Materials
ATR	Attenuated Total Reflectance
DRIFT	Diffuse Reflectance Infrared Fourier Transform
DSC	Differential Scanning Calorimetry
EGA	Evolved Gas Analysis
EO/PO	Ethylene Oxide/Propylene Oxide
EPDM	Ethylene–Propylene–Diene Monomer
EVA	Ethylene/Vinyl Acetate
FD	Forwards Draw
FID	Flame Ionization Detector
FSD	Fourier Self-deconvolution
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GPC	Gel Permeation Chromatography

HDDA	1,6-Hexandiol Diacrylate
HDPE	High Density Poly(ethylene)
HFP	Hexafluoropropylene
HPLC	High-performance Liquid Chromatography
IRS	Internal Reflection Spectroscopy
LC	Liquid Chromatography
LDPE	Low Density Poly(ethylene)
LLDPE	Linear Low Density Poly(ethylene)
MAA	Methacrylic Acid
MCT	Mercury Cadmium Telluride
MD	Machine Draw
MIR	Multiple Internal Reflection
MS	Mass Spectrography
NCA	Normal Coordinate Analysis
NIRA	Near-infrared Analysis
NMR	Nuclear Magnetic Resonance
ODF	Orientation Distribution Function
PACN	Polyacrylonitrile
PA-FTIR	Photoacoustic-FTIR
PBT	Poly(butylene terephthalate)
PCR	Principal Component Regression
PEEK	Poly(aryl ether ether ketone)
PEN	Poly(ethylene naphthalene-2,6-dicarboxylate)
PES	Poly(aryl-ether-sulphone)
PET	Poly(ethylene terephthalate)
PLS	Partial Least Squares
PMAA	Poly(methacrylic acid)
PP	Polypropylene
PTFE	Poly(tetrafluoroethylene)
PU	Polyurethane
PVA	Poly(vinyl acetate)
PVAL	Poly(vinyl alcohol)
PVC	Poly(vinyl chloride)
PVdF	Poly(vinylidene fluoride)
RAIRS	Reflection–Absorption Infrared Spectroscopy
RIM	Reaction Injection Molding
SD	Sideways Draw
SEC	Size Exclusion Chromatography
TD	Transverse Draw
TFE/HFP	Tetrafluoroethylene/Hexafluoropropylene
TLC	Thin-layer Chromatography
TGA	Thermogravimetric Analysis
THF	Tetrahydrofuran
TIRES	Transient Infrared Emission Spectroscopy
VA	Vinyl Acetate
VA/MAA	Vinyl Acetate/Methacrylic Acid
2D-IR	Two-dimensional Infrared
γ-APS	γ-Aminopropyl-triethoxysilane

RELATED ARTICLES

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis • Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization • Gas Chromatography in Analysis of Polymers and Rubbers • Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Inverse Gas Chromatography in Analysis of Polymers • Near-infrared Spectroscopy of Polymers and Rubbers • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers • Pyrolysis Techniques in the Analysis of Polymers and Rubbers • Supercritical Fluid Chromatography of Polymers

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Gas Chromatography/Infrared Spectroscopy • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Liquid Chromatography/Infrared Spectroscopy • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

REFERENCES

1. J.L. Koenig, *Spectroscopy of Polymers*, 2nd edition, American Chemical Society, Washington, 1999.
2. J. Haslam, H.A. Willis, D.C.M. Squirrell, *Identification and Analysis of Plastics*, 2nd edition, Butterworth, London, 1972.
3. P.C. Painter, M.M. Coleman, J.L. Koenig, *The Theory of Vibrational Spectroscopy and its Application to*

Polymeric Materials, John Wiley & Sons, New York, 1982.
4. D.I. Bower, W.F. Maddams, *The Vibrational Spectroscopy of Polymers*, Cambridge University Press, Cambridge, 1989.
5. A. Garton, *Infrared Spectroscopy of Polymer Blends, Composites and Surfaces*, Carl Hanser Verlag, Munich, 1992.
6. D.O. Hummel (ed.), *Polymer Spectroscopy*, Verlag Chemie GmbH, Weinheim, 1974.
7. J.M. Chalmers, N.J. Everall, 'Vibrational Spectroscopy', in *Polymer Characterisation*, eds. B.J. Hunt, M.I. James, Blackie Academic, Glasgow, 69–114, 1993.
8. S. Krimm, 'Infrared Spectra of High Polymers', *Fortschr. Hochpolym.-Forsch.*, **2**, 51–172 (1960).
9. H.W. Siesler, K. Holland-Moritz, *Infrared and Raman Spectroscopy of Polymers*, Macel Dekker, Inc., New York, 1980.
10. H. Ishida (ed.), *Fourier Transform Infrared Characterization of Polymers*, Plenum Press, New York, 1987.
11. B. Jasse, 'Fourier Transform Infrared Spectroscopy of Synthetic Polymers', in *Developments in Polymer Characterisation-4*, ed. J.V. Dawkins, Applied Science Publishers, 91–129, 1983.
12. D.O. Hummel, *Infrared Analysis of Polymers, Resins and Additives: An Atlas. Vol. I, Plastics, Elastomers, Fibers and Resins. Part 1: Text*, eds. D.O. Hummel, K. Scholl, John Wiley & Sons, Inc, New York, 1971.
13. R. Zbinden, *Infrared Spectroscopy of High Polymers*, Academic Press, New York, 1964.
14. H.A. Willis, 'Preparation of Polymer Samples for IR Examination', in *Laboratory Methods in Vibrational Spectroscopy*, 3rd edition, eds. H.A. Willis, J. van der Maas, R.G.J. Miller, John Wiley & Sons Ltd, Chichester, 217–227, 1987.
15. J.M. Chalmers, 'Infrared Spectroscopy – Theory and Instrumentation. Sample Preparation', in *Encyclopedia of Analytical Science*, ed. A. Townshend, Academic Press Ltd, London, 2176–2188, Vol. 4, 1995.
16. J.M. Chalmers, G. Dent, *Industrial Analysis with Vibrational Spectroscopy*, The Royal Society of Chemistry, Cambridge, 1997.
17. Meter bar (K bar™) coaters are supplied by R.K. Print-Coat Instruments Ltd, Litlington, Royston, Herts., UK. Historically, this approach was sometimes referred to as Meyer-bar coating.
18. N.J. Harrick, *Internal Reflection Spectroscopy*, Wiley Interscience, New York, 1967.
19. F.M. Mirabella, N.J. Harrick, *Infrared Reflection Spectroscopy: Review and Supplement*, Harrick Scientific Corp., New York, 1985.
20. H.A. Willis, V.J.I. Zichy, 'The Examination of Polymer Surfaces by Infrared Spectroscopy', in *Polymer Surfaces*, eds. D.T. Clark, W.J. Feast, Wiley, Chichester, 287–307, 1978.
21. F.M. Mirabella, 'Attenuated Total Reflection Spectroscopy', in *Modern Techniques in Applied Molecular Spectroscopy*, ed. F.M. Mirabella, John Wiley & Sons, Inc., New York, 127–184, 1998.
22. M. Claybourn, '*Infrared Reflectance Spectroscopy of Polymers. Analysis of Films, Surfaces and Interfaces*, Global Press, USA, 1998.
23. M. Claybourn, P. Colombel, J.M. Chalmers, 'Characterization of Carbon-filled Polymers by Specular Reflectance', *Appl. Spectrosc.*, **45**(2), 279–286 (1991).
24. J.M. Chalmers, N.J. Everall, S. Ellison, 'Specular Reflectance: A Convenient Tool for Characterisation by FTIR-microscopy?', *Micron*, **27**(5), 315–328 (1996).
25. M.B. Mitchell, 'Fundamentals and Applications of Diffuse Reflectance Infrared Fourier Transform (DRIFT) Spectroscopy', in *Structure–Property Relations in Polymers: Spectroscopy and Performance*, eds. M.W. Urban, C.D. Craver, Advances in Chemistry Series 236, American Chemical Society, Washington, DC, 351–375, 1993.
26. J.P. Blitz, 'Diffuse Reflectance Spectroscopy', in *Modern Techniques in Applied Molecular Spectroscopy*, ed. F.M. Mirabella, John Wiley & Sons, Inc., New York, 185–219, 1998.
27. J.M. Chalmers, M.W. Mackenzie, 'Solid Sampling Techniques', in *Advances in Applied Fourier Transform Infrared Spectroscopy*, ed. M.W. Mackenzie, John Wiley & Sons, Chichester, 105–188, 1988.
28. M.P. Fuller, P.R. Griffiths, 'Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry', *Anal. Chem.*, **50**(13), 1906–1910 (1978).
29. J.M. Chalmers, M.W. Mackenzie, 'Some Industrial Applications of FT-IR Diffuse Reflectance Spectroscopy', *Appl. Spectrosc.*, **39**(4), 634–641 (1985).
30. F.J. deBlase, S. Compton, 'Infrared Emission Spectroscopy: A Theoretical and Experimental Review', *Appl. Spectrosc.*, **45**(4), 611–618 (1991).
31. H.A. Willis, 'The Infrared Emission Spectrum', in *Laboratory Methods in Vibrational Spectroscopy*, 3rd edition, eds. H.A. Willis, J. van der Maas, R.G.J. Miller, John Wiley & Sons Ltd, Chichester, 473–482, 1987.
32. E. Rytter, 'A comment on Emission, Reflection and Transmission Spectroscopy on Thin Films', *Spectrochim. Acta*, **43A**(4), 523–529 (1987).
33. D.B. Chase, 'The Sensitivity and Limitations of Condensed Phase Infrared Emission Spectroscopy', *Appl. Spectrosc.*, **35**(1), 77–81 (1981).
34. R.W. Jones, J.F. McClelland, 'Transient Infrared Emission Spectroscopy', *Anal. Chem.*, **61**, 650–656 (1989).
35. J.F. McClelland, S.J. Bajic, R.W. Jones, L.M. Seaverson, 'Photoacoustic Spectroscopy', in *Modern Techniques in Applied Molecular Spectroscopy*, ed. F.M. Mirabella, John Wiley & Sons, Inc., New York, 221–265, 1998.

36. M.W. Urban, 'Photoacoustic Fourier Transform Infrared Response to Thermal and Interfacial Changes in Polymers', in *Structure–Property Relations in Polymers: Spectroscopy and Performance*, eds. M.W. Urban, C.D. Craver, Advances in Chemistry Series 236, American Chemical Society, Washington, DC, 645–659, 1993.
37. J.F. McClelland, R.W. Jones, S. Luo, L.M. Seaverson, 'A Practical Guide to FT-IR Photoacoustic Spectroscopy', in *Practical Sampling Techniques for Infrared Analysis*, ed. P.B. Coleman, CRC Press, Boca Raton, 107–163, 1993.
38. A. Rosencwaig, A. Gersho, 'Theory of the Photoacoustic Effect with Solids', *J. Appl. Phys.*, **47**, 64–69 (1976).
39. A. Rosencwaig, A. Gersho, 'Photoacoustic Effect with Solids: A Theoretical Treatment', *Science*, **190**, 556–557 (1975).
40. A. Rosencwaig, *Photo-acoustics and Photo-acoustic Spectroscopy*, J. Wiley, New York, 1980.
41. R.A. Palmer, 'Step-scan FT-IR. A Versatile Tool for Time- and Phase-resolved Vibrational Spectroscopy', *Spectroscopy*, **8**(2), 26–36 (1993).
42. R.A. Palmer, J.L. Chao, R.M. Dittmar, V.G. Gregoriou, S.E. Plunkett, 'Investigation of Time-dependent Phenomena by Use of Step-scan FT-IR', *Appl. Spectrosc.*, **47**(9), 1297–1310 (1993).
43. C.J. Manning, R.A. Palmer, J.L. Chao, 'Step-scan Fourier-transform Infrared Spectrometer', *Rev. Sci. Instrum.*, **62**(5), 1219–1229 (1991).
44. E.Y. Jiang, R.A. Palmer, J.L. Chao, 'Development and Applications of a Photoacoustic Phase Theory for Multilayer Materials: The Phase Difference Approach', *J. Appl. Phys.*, **78**(1), 460–469 (1995).
45. M.C. Wahls, J.C. Leyte, 'Fourier Transform Infrared Spectroscopy of Polymeric Laminates', *J. Appl. Phys.*, **83**(1), 504–509 (1998).
46. G. Socrates, *Infrared Characteristic Group Frequencies*, 2nd edition, J. Wiley, New York, 1994.
47. D. Lin-Vien, N.B. Colthrup, W.G. Fateley, J.G. Grasselli, *The Handbook of Infrared and Raman Characteristic Group Frequencies of Organic Molecules*, J. Wiley & Sons, New York, 1991.
48. N.B. Colthup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, 3rd edition, Academic Press, Inc., San Diego, USA, 1990.
49. B. Fanconi, 'Molecular Vibrations of Polymers', *Ann. Rev. Phys. Chem.*, **31**, 265–291 (1980).
50. J.L. Koenig, 'Fourier Transform Infrared Spectroscopy of Polymers', *Adv. Polym. Sci.*, **54**, 87–154 (1983).
51. J.L. Koenig, *Chemical Microstructure of Polymer Chains*, John Wiley & Sons, Inc., New York, 1980.
52. R.G. Snyder, J.H. Schachtschneider, 'Valence Force Calculation of the Vibrational Spectra of Crystalline Isotactic Polypropylene and some Deuterated Polypropylenes', *Spectrochim. Acta.*, **20**, 853–869 (1964).
53. J.H. Schachtschneider, R.G. Snyder, 'Valence Force Calculation of the Vibrational Frequencies of Crystalline Syndiotactic Polypropylene', *Spectrochim. Acta.*, **21**, 1527–1542 (1965).
54. L. D'Esposito, J.L. Koenig, 'Application of Fourier Transform Infrared Spectroscopy to the Study of Semicrystalline Polymers: Poly(ethylene terephthalate)', *J. Poly. Sci.: Polym. Phys. Ed.*, **14**, 1731–1741 (1976).
55. S.-B. Lin, J.L. Koenig, 'Spectroscopic Characterization of the Rotational Conformations in the Disordered Phase of Poly(ethylene terephthalate)', *J. Polym. Sci.: Polym. Phys. Ed.*, **20**, 2277–2295 (1982).
56. R.H. Lehnert, P.J. Hendra, N. Everall, N.J. Clayden, 'Comparative Quantitative Study on the Crystallinity of Poly(tetrafluoroethylene) Including Raman, Infrared and ^{19}F Nuclear Magnetic Resonance Spectroscopy', *Polymer*, **38**(7), 1521–1535 (1997).
57. R.E. Moynihan, 'The Molecular Structure of Perfluorocarbon Polymers. Infrared Studies on Polytetrafluoroethylene', *J. Am. Chem. Soc.*, **81**, 1045–1050 (1959).
58. H.W. Starkweather, R.C. Ferguson, D.B. Chase, J.M. Minor, 'Infrared Spectra of Amorphous and Crystalline Poly(tetrafluoroethylene)', *Macromol.*, **18**, 1684–1686 (1985).
59. G.H. Strobl, W. Hagedorn, 'Raman Spectroscopic Method for Determining the Crystallinity of Polyethylene', *J. Polym. Sci.: Polym. Phys.*, **16**, 1181–1193 (1978).
60. S. Krimm, V.L. Folt, J.J. Shipman, A.R. Berens, 'Infrared Spectra and Assignments for Polyvinyl Chloride and Deuterated Analogs', *J. Polym. Sci., A.*, **1**, 2621–2650 (1963).
61. V.A. Bershtein, V.A. Rhyzov, 'Far-infrared Spectroscopy of Polymers', *Adv. Polym. Sci.*, **114**, 43–121 (1994).
62. D.R. Beckett, J.M. Chalmers, M.W. Mackenzie, H.A. Willis, H.G.M. Edwards, J.S. Lees, D.A. Long, 'The Far-infra-red Spectra of Crystalline Isotactic Polypropylene Polymorphs', *Eur. Polym. J.*, **21**(10), 849–852 (1985).
63. R.J. Samuels, *Structured Polymer Properties*, John Wiley & Sons, New York, 1974.
64. G.L. Wilkes, 'The Measurement of Molecular Orientation in Polymeric Solids', *Adv. Polym. Sci.*, **8**, 91–136 (1971).
65. P.G. Schmidt, 'Polyethylene Terephthalate Structural Studies', *J. Polym. Sci., A*, **1**, 1271–1292 (1963).
66. I.M. Ward, 'Determination of Molecular Orientation by Spectroscopic Techniques', *Adv. Polym. Sci.*, **66**, 81–115 (1985).
67. D.I. Bower, 'Infrared Dichroism, Polarized Fluorescence and Raman Spectroscopy', in *Structure and Properties of Oriented Polymers*, ed. I.M. Ward, Chapman & Hall, London, 181–233, 1997.
68. R.D.B. Fraser, 'Interpretation of Infrared Dichroism in Fibrous Proteins – the 2μ Region', *J. Chem. Phys.*, **24**(1), 89–95 (1956).
69. J.J. Hermans, P.H. Hermans, D. Vermaas, A. Weidinger, 'Quantitative Evaluation of Orientation in Cellulose

Fibers from the X-ray Fiber Diagram', *Rec. Trav. Chim.*, **65**, 427–447 (1946).
70. D.A. Jarvis, I.J. Hutchinson, D.I. Bower, I.M. Ward, 'Characterization of Biaxial Orientation in Poly(ethylene terephthalate) by Means of Refractive Index Measurements and Raman and Infra-red Spectroscopies', *Polymer*, **21**, 41–54 (1980).
71. B. Jasse, J.L. Koenig, 'Orientational Measurements in Polymers Using Vibrational Spectroscopy', *J. Macromol. Sci. – Rev. Macromol. Chem.*, **C17**(1), 61–135 (1979).
72. N.J. Everall, 'Using Polarised Vibrational Spectroscopy to Characterise Molecular Orientations in Polymers: An Introduction – A Tutorial', Internet Journal of Vibrational Spectroscopy, 1999. (http://www.ijvs.com/volume3/edition2/section1.htm).
73. J.L. Koenig, S.W. Cornell, D.E. Witenhafer, 'Infrared Technique for the Measurement of Structural Changes During the Orientation Process in Polymers', *J. Polym. Sci., A2*, **5**, 303–313 (1967).
74. L.J. Fina, J.L. Koenig, 'Analysis of Trichroic Infrared Absorption in Solids. I. Description of Methods', *J. Polym. Sci., Part B: Polym. Phys.*, **24**, 2509–2524 (1986).
75. P.A. Flourney, W.J. Schaffers, 'Attenuated Total Reflection Spectra from Surfaces of Anisotropic, Absorbing Films', *Spectrochim. Acta*, **22**, 5–13 (1966).
76. J.P. Hobbs, C.S.P. Sung, K. Krishnan, S. Hill, 'Characterization of Surface Structure and Orientation in Polypropylene and Poly(ethylene terephthalate) Films by Modified Attenuated Total Reflection IR Dichroism Studies', *Macromol.*, **16**, 193–199 (1983).
77. P. Yuan, C.S.P. Sung, 'Angular Characterization of Polymer Surfaces by FTIR/ATR Dichroism with a Rotatable Truncated Hemispheric Crystal.', *Macromol.*, **24**, 6095–6103 (1991).
78. R. Belali, J.-M. Vigoureux, J. Morvan, 'Infrared Dichroism and Polymer Orientation in Attenuated Total Reflection. Theoretical Considerations', *J. Polym. Sci. B: Polym. Phys.*, **35**, 1361–1372 (1997).
79. N.J. Everall, A. Bibby, 'Improvements in the Use of Attenuated Total Reflection Fourier Transform Infrared Dichroism for Measuring Surface Orientation in Polymers', *Appl. Spectrosc.*, **51**(8), 1083–1091 (1997).
80. K.C. Cole, J. Guevremont, A. Ajji, M.M. Dumoulin, 'Characterization of Surface Orientation in Poly(ethylene terephthalate) by Front-surface Reflection Infrared Spectroscopy', *Appl. Spectrosc.*, **48**(12), 1513–1521 (1994).
81. N.J. Everall, J.M. Chalmers, A. Local, S. Allen, 'Measurement of Surface Orientation in Uniaxial Poly(ethylene terephthalate) Films Using Polarised Specular Reflectance Fourier Transform Infrared Microscopy', *Vib. Spectrosc.*, **10**, 253–259 (1996).
82. J.M. Chalmers, L. Croot, J.G. Eaves, N. Everall, W.F. Gaskin, J. Lumsdon, N. Moore, 'Vibrational Spectroscopy-microscopy: Applications in Materials Science and Characterisation', *Spectrosc. Int. J.*, **8**, 13–42 (1990).
83. J.M. Chalmers, N.J. Everall, 'The Role of Vibrational Spectroscopy–microscopy Techniques in Polymer Characterisation', *Macromol. Symp.*, **94**, 33–49 (1995).
84. I.J. Hutchinson, I.M. Ward, H.A. Willis, V. Zichy, 'Infrared Measurements on One-way Drawn Poly(ethylene terephthalate) Films Subjected to Constant Strain', *Polymer*, **21**, 55–65 (1980).
85. A. Cunningham, I.M. Ward, H.A. Willis, V. Zichy, 'An Infra-red Spectroscopic Study of Molecular Orientation and Conformational Changes in Poly(ethylene terephthalate)', *Polymer*, **15**, 749–756 (1974).
86. M. Ito, J.R. Pereira, S.L. Hsu, R.S. Porter, 'Fourier-transform Infrared Study of Coextrusion-drawn Poly-(ethylene terephthalate)', *J. Polym. Sci.: Polym. Phys.*, **21**, 389–400 (1983).
87. L.J. Fina, J.L. Koenig, 'Analysis of Trichroic Infrared Absorption in Solids. II. One-way Drawn Poly(ethylene terephthalate)', *J. Polym. Sci., B: Polym. Phys.*, **24**, 2525–2539 (1986).
88. K.H. Lee, C.S.P. Sung, 'Effect of Biaxial Drawing Conditions on Orientation and Structure of Poly(ethylene terephthalate)', *Macromol.*, **26**, 3289–3294 (1993).
89. E.A. Lofgren, S.A. Jabarin, 'Polarized Internal Reflectance Spectroscopic Studies of Oriented Poly(ethylene terephthalate)', *J. Appl. Polym. Sci.*, **51**, 1251–1267 (1994).
90. F. Rietsch, B. Jasse, 'A Study of Molecular Orientation in One-way Drawn Poly(ethylene terephthalate) Films by Means of Polarized Infra-red Spectroscopy and Birefringence Measurements', *Polym. Bull.*, **11**, 287–292 (1984).
91. M. van Gurp, B.J. Kip, J.P.C. van Heel, S. de Boer, 'On the Development of Orientation in LDPE Blown Films', *J. Plast. Film & Sheeting*, **10**(2), 156–176 (1994).
92. A.R. Wedgewood, J.C. Seferis, 'Structural Orientation of Linear Polyethylene by Infrared Spectroscopy', *Pure & Appl. Chem.*, **55**(5), 873–892 (1983).
93. J.M. Chalmers, N.J. Everall, A. Local, 'Interrogating the Morphology of Poly(ethylene) Tubing/Sheathing by Vibrational Spectroscopy–microscopy Techniques', Microbeam Analysis, Proc. 29th Ann. Conf. MAS, August 6–11, Breckenridge, CO, USA, 123–124, 1995.
94. N. Everall, J. Chalmers, P. Mills, 'Use of Polarized Resonance Raman Spectroscopy of a Polyene Probe, and FT-IR Dichroism, to Probe Amorphous-phase Orientation in Uniaxially Drawn Poly(ethylene)', *Appl. Spectrosc.*, **50**(10), 1229–1234 (1996).
95. Y.V. Kissin, 'Orientation of Isotactic Polypropylene in Crystalline and Amorphous Phases: IR Methods', *J. Polym. Sci.: Polym. Phys. Ed.*, **21**, 2085–2096 (1983).
96. A. Garton, D.J. Carlsson, D.M. Wiles, 'Molecular Orientation Measurements in Polymers by Infrared Spectral Subtraction', *Appl. Spectrosc.*, **35**(4), 432–435 (1981).
97. J.P. Sibilia, 'Orientation in Nylon 6 Films as Determined by the Three-dimensional Polarized Infrared Technique', *J. Polym. Sci.: A-2*, **9**, 27–42 (1971).

98. S.R. Samanta, W.W. Lanier, R.W. Miller, M.E. Gibson, 'Fiber Structure Study by Polarized Infrared Attenuated Total Reflection Spectroscopy', *Appl. Spectrosc.*, **44**(2), 286–289 (1990).
99. M. Theodorou, B. Jasse, 'Fourier-transform Infrared Study of Orientation in Pure and Plasticized Poly(vinyl chloride)', *J. Polym. Sci., B: Polym. Phys.*, **24**, 2643–2654 (1986).
100. L.J. Fina, J.L. Koenig, W.L. Gordon, 'Analysis of Trichroic Infrared Absorption in Solids. III. Phase I Poly(vinylidene fluoride)', *J. Polym. Sci., B, Polym. Phys.*, **24**, 2541–2551 (1986).
101. T. Buffeteau, B. Desbat, S. Besbes, M. Nafati, L. Bokobza, 'Molecular Orientation Studies in Polymer Films by Polarization Modulation FTIR Spectroscopy', *Polymer*, **35**(12), 2538–2541 (1994).
102. S. Besbes, L. Bokobza, L. Monnerie, I. Bahar, B. Erman, 'Molecular Orientation in Deformed Bimodal Networks. 2. Fourier Transform Infrared Measurements on Poly(dimethylsiloxane) Networks and Comparison with Theory', *Macromol.*, **28**, 231–235 (1995).
103. C.K. Yeung, B. Jasse, 'Fourier-transform Infrared Study of Uniaxially Oriented Polytetrafluoroethylene', *J. Appl. Polym. Sci.*, **27**, 4587–4597 (1982).
104. D. Lefebvre, B. Jasse, L. Monnerie, 'Orientation and Relaxation in Uniaxially Stretched Atactic Polystyrene', *Polymer*, **24**, 1240–1244 (1983).
105. C. Bouton, V. Arrondel, V. Rey, Ph. Sergot, J.L. Manguin, B. Jasse, L. Monnerie, 'Fourier Transform Infrared and Birefringence Studies of Orientation in Uniaxially-stretched Poly(2,6-dimethyl 1,4-phenylene oxide) – Atactic Polystyrene', *Polymer*, **30**, 1414–1418 (1989).
106. B. Akram, L. Bokobza, J.P. Quesiel, L. Monnerie, 'Fourier-transform Infra-red Dichroism Study of Molecular Orientation in Synthetic High *cis*-1,4-Polyisoprene and in Natural Rubber', *Polymer*, **27**, 877–882 (1986).
107. B. Amram, L. Bokobza, L. Monnerie, J.P. Quesiel, 'Fourier-transform Infra-red Dichroism Study of Molecular Orientation in High *cis*-1,4-Polybutadiene', *Polymer*, **29**, 1155–1160 (1988).
108. H.S. Lee, S.L. Hsu, 'Structural Changes and Chain Orientational Behavior During Tensile Deformation of Segmented Polyurethanes', *J. Polym. Sci., B: Polym. Phys.*, **32**, 2085–2098 (1994).
109. N. Reynolds, H.W. Spiess, 'Structure and Deformation Behaviour of Model Poly(ether-urethane) Elastomers, 1', *Macromol. Chem.*, **195**, 2855–2873 (1994).
110. W. Tang, W.J. MacKnight, S.L. Hsu, 'Segmented Polyurethane Elastomers with Liquid Crystalline Hard Segments. 3. Infrared Spectroscopic Study', *Macromol.*, **28**, 4284–4289 (1995).
111. H.W. Siesler, 'The Characterization of Polymer Deformation by Rheo-optical Fourier-transform Infrared Spectroscopy', *Makromol. Chem., Macromol. Symp.*, **53**, 89–103 (1992).
112. U. Hoffmann, F. Pfeifer, S. Okretic, N. Volkl, M. Zahedi, H.W. Siesler, 'Rheo-optical Fourier Transform Infrared and Raman Spectroscopy of Polymers', *Appl. Spectrosc.*, **47**(9), 1531–1539 (1993).
113. C.J.M. Van Den Heuvel, H.M. Heuvel, W.A. Faasen, J. Veurink, L.J. Lucas, 'Molecular Changes of PET Yarns During Stretching Measured with Rheo-optical Infrared Spectroscopy and Other Techniques', *J. Appl. Polym. Sci.*, **49**, 925–934 (1993).
114. H.S. Lee, S.L. Hsu, 'Structural Changes and Chain Orientational Behavior During Tensile Deformation of Segmented Polyurethanes', *J. Polym. Sci., B: Polym. Phys.*, **32**, 2085–2098 (1994).
115. W.B. Fischer, P. Potschke, G. Pompe, K.-J. Eichorn, H.W. Siesler, 'Rheo-optical Fourier Transform Infrared Spectroscopy of Polyurethanes and their Blends with Polyolefins', *Macromol. Chem. Phys.*, **198**(7), 2057–2072 (1997).
116. N. Reynolds, H.W. Spiess, H. Hayen, H. Nefzger, C.D. Eisenbach, 'Structure and Deformation Behaviour Model Poly(ether-urethane) Elastomers, 1', *Macromol. Chem. Phys.*, **195**, 2855–2873 (1994).
117. W.J. Walczak, R.P. Wool, 'Investigation of Polymer Melt Relaxation Mechanisms via Dynamic Infrared Dichroism', *Macromol.*, **24**, 4657–4665 (1991).
118. T. Scherzer, 'Rheo-optical FTIR Spectroscopy of Epoxy Resins', *J. Appl. Polym. Sci.*, **58**, 501–514 (1995).
119. M. Brauchler, C. Boeffel, H.W. Spiess, 'Dynamic Fourier Transform Infrared Spectroscopy of Liquid-crystalline Elastomers', *Makromol. Chem.*, **192**, 1153–1176 (1991).
120. A.H. Willbourn, 'Polymethylene and the Structure of Polyethylene: Study of Short-chain Branching, Its Nature and Effects', *J. Polym. Sci.*, **XXXIV**, 569–597 (1959).
121. ASTM Designation: D 2238-92, 'Standard Test Method for Absorbance of Polyethylene Due to Methyl Groups at 1378 cm^{-1}', 1998 Annual Book of ASTM Standards, Vol. 8.01, American Society for Testing and Materials, West Conshohocken, PA, 533–539.
122. T. Usami, S. Takayama, 'Identification of Branches in Low-density Polyethylenes by Fourier Transform Infrared Spectroscopy', *Polymer J.*, **16**(10), 731–738 (1984).
123. A. Solti, D.O. Hummel, P. Simak, 'Computer-supported Infrared Spectrometry of Polyethylene, Ethene Copolymers and Amorphous Poly(alkyl ethylene)s', *Makromol. Chem., Macromol. Symp.*, **5**, 105–133 (1986).
124. J.M. Chalmers, M.W. Mackenzie, H.A. Willis, 'FT-IR Spectroscopy in an Industrial Laboratory', *Appl. Spectrosc.*, **38**(6), 763–773 (1984).
125. U.S. Patent 3,085,083, 'Stabilized Tetrafluoroethylene-fluoro-olefin Copolymers Having $-CF_2H$ End Groups', 1963.
126. U.S. Patent 4,742,122, 'Melt-processible Tetrafluoroethylene/Perfluoroolefin Copolymers and the Processes for Preparing Them', 1988.

127. European Patent Application 0 150 953, 'Extrusion Finishing of Perfluorinated Copolymers', 1985.
128. D. Patterson, I.M. Ward, 'The Assignment of the Carboxyl and Hydroxyl Absorptions in the Infra-red Spectrum of Polyethylene Terephthalate', *Trans. Faraday Soc.*, **411**, 53(3), 291–294 (1957).
129. R.L. Addleman, V.J.I. Zichy, 'Accurate Measurement of Carboxyl and Hydroxyl End-group Concentrations in Poly(ethylene terephthalate) Film by Infra-red Spectroscopy', *Polymer*, **13**, 391–398 (1972).
130. D.A.S. Ravens, I.M. Ward, 'Chemical Reactivity of Polyethylene Terephthalate', *Trans. Faraday Soc.*, **57**, 150–159 (1961).
131. P.G. Kosky, R.S. McDonald, E.A. Guggenheim, 'Determination of End-group Concentrations and Molecular Weight of Poly(butylene terephthalate) by Solid-state Fourier Transform Infrared Spectroscopy', *Polym. Eng. Sci.*, **25**(7), 389–394 (1985).
132. H. Zhang, A. Rankin, I.M. Ward, 'Determination of the End-group Concentration and Molecular Weight of Poly(ethylene naphthalene-2,6-dicarboxylate) Using Infra-red Spectroscopy', *Polymer*, **37**(7), 1079–1085 (1996).
133. L.R. Schroeder, S.L. Cooper, 'Hydrogen Bonding in Polyamides', *J. Appl. Phys.*, **47**(10), 4310–4317 (1976).
134. M.M. Coleman, D.J. Skrovanek, S.E. Howe, P.C. Painter, 'On the Validity of a Commonly Employed Infrared Procedure Used to Determine Thermodynamic Parameters Associated with Hydrogen Bonding in Polymers', *Macromolecules*, **18**(2), 299–301 (1985).
135. D.J. Skrovanek, S.E. Howe, P.C. Painter, M.M. Coleman, 'Hydrogen Bonding in Polymers: Infrared Temperature Studies of an Amorphous Polyamide', *Macromolecules*, **18**(9), 1676–1683 (1985).
136. D.J. Skrovanek, P.C. Painter, M.M. Coleman, 'Hydrogen Bonding in Polymers. 2. Infrared Studies of Nylon 11', *Macromolecules*, **19**(3), 699–705 (1986).
137. M.M. Coleman, D.J. Skrovanek, P.C. Painter, 'Hydrogen Bonding in Polymers: III Further Infrared Temperature Studies of Polyamides', *Makromol. Chem., Macromol. Symp.*, **5**, 21–33 (1986).
138. V.W. Srichatrapimuk, S.L. Cooper, 'Infrared Thermal Analysis of Polyurethane Block Polymers', *J. Macromol. Sci. – Phys.*, **B15**(2), 267–311 (1978).
139. M.M. Coleman, K.H. Lee, D.J. Skrovanek, P.C. Painter, 'Hydrogen Bonding in Polymers. 4. Infrared Temperature Studies on a Simple Polyurethane', *Macromolecules*, **19**(8), 2149–2157 (1986).
140. V. Zharkov, A.G. Strikovsky, T.E. Verteleskaya, 'Amide I Absorption Band: Description of the Urethane Group Association Scheme in Polyether Urethane Elastomers', *Polymer*, **34**(5), 938–941 (1993).
141. F.-S. Yen, J.-L. Hong, 'Hydrogen-bond Interactions Between Ester and Urethane Linkages in Small Model Compounds and Polyurethanes', *Macromolecules*, **30**(25), 7927–7938 (1997).
142. F.C. Wang, M. Feve, T.M. Lam, J.-P. Pascault, 'FTIR Analysis of Hydrogen Bonding in Amorphous Linear Aromatic Polyurethanes, I. Influence of Temperature', *J. Polym. Sci., Part B: Polym. Phys.*, **32**, 1305–1313 (1994).
143. L.-S. Teo, C.-Y. Chen, J.-F. Kuo, 'Fourier Transform Infrared Spectroscopy Study of Temperature on Hydrogen Bonding in Amine-containing Polyurethanes and Poly(urethane-urea)s', *Macromolecules*, **30**(6), 1793–1799 (1997).
144. M.M. Coleman, M. Sobkowiak, G.J. Pehlert, P.C. Painter, 'Infrared Temperature Studies of a Simple Polyurea', *Macromol. Chem. Phys.*, **198**, 117–136 (1997).
145. A. Marcos-Frenandez, A.E. Lozano, L. Gonzalez, A. Rodriguez, 'Hydrogen Bonding in Copoly(ether-urea)s and its Relationship with the Physical Properties', *Macromolecules*, **30**(12), 3584–3592 (1997).
146. M.M. Coleman, J.E. Graf, P.C. Painter, *Specific Interactions and the Miscibility of Polymer Blends*, Technomic Publishing Co., Inc., Pennsylvania, 1991.
147. M.M. Coleman, P.C. Painter, 'Fourier Transform Infrared Spectroscopy: Probing the Structure of Multicomponent Polymer Blends', *Appl. Spectrosc. Rev.*, **20**(3&4), 255–346 (1984).
148. M.M. Coleman, D.J. Skrovanek, J. Hu, P.C. Painter, 'Hydrogen-bonding in Polymer Blends. 1. FTIR Studies of Urethane-ether Blends', *Macromolecules*, **21**(1), 59–65 (1988).
149. J. Hu, P.C. Painter, M.M. Coleman, T.D. Krizan, 'Toward a Prediction of the Phase Behavior of Polyamide Blends', *J. Polym. Sci., Part B: Polym. Phys.*, **28**, 149–176 (1990).
150. D.E. Bhagwagar, P.C. Painter, M.M. Coleman, T.D. Krizan, 'Mapping the Phase Diagram of an Aromatic Polyamide Blend Using Infrared Spectroscopy', *J. Polym. Sci., Part B: Polym. Phys.*, **29**, 1547–1588 (1991).
151. M.M. Coleman, G.J. Pehlert, X. Yang, J.B. Stallman, P.C. Painter, 'Self-association Versus Interassociation in Hydrogen Bonded Polymer Blends: 1. Determination of Equilibrium Constants from Miscible Poly(2,6-dialkyl-4-vinyl phenol) Blends', *Polymer*, **37**(21), 4753–4761 (1996).
152. G.J. Pehlert, X. Yang, P.C. Painter, M.M. Coelman, 'Self-association Versus Interassociation in Hydrogen Bonded Polymer Blends: 2. Comparison of Theoretical and Experimental Miscibility Windows for Poly(2,6-dialkyl-4-vinyl phenol) Blends', *Polymer*, **37**(21), 4763–4771 (1996).
153. Y. Xu, J. Graf, P.C. Painter, M.M. Coleman, 'Miscibility Windows for Poly(styrene-co-vinyl phenol) Blends with Poly(*n*-butyl methacrylate) and Poly(*n*-hexyl methacrylate): a Comparison of Theoretical Predictions with Fourier Transform Infra-red Experimental Data', *Polymer*, **32**(17), 3103–3119 (1991).

154. M.M. Coleman, Y. Xu, P.C. Painter, J.R. Harrell, 'Miscibility Maps for Copolymer–Copolymer Blends: A Comparison of Theoretical Predictions to Experimental Studies of Poly(styrene-co-vinyl phenol) Blends with Poly(ethylene-co-methyl acrylate)', *Makromol. Chem., Macromol. Symp.*, **52**, 75–89 (1991).
155. M. Sargent, J.L. Koenig, 'Fourier Transform Infrared Spectroscopic Analysis of Poly(vinylidene fluoride) – Poly(vinyl acetate) blends', *Vibrational Spectroscopy*, **2**, 21–28 (1991).
156. S. Kole, S. Roy, A.K. Bhowmick, 'Influence of Chemical Interaction on the Properties of Silicone-EPDM Rubber Blend', *Polymer*, **36**(17), 3273–3277 (1995).
157. K. Yamada, Y. Funayama, 'FTIR Spectroscopic Studies of Miscible Polymer Blends', *Rubber Chem Technol.*, **63**(5), 669–682 (1990).
158. ASTM Designation: D 3594-93, 'Standard Test Method for Copolymerized Ethyl Acrylate in Ethylene-ethyl Acrylate Copolymers', 1998 Annual Book of ASTM Standards, Vol. 8.03, American Society for Testing and Materials, West Conshohocken, PA, 388–391.
159. M. Buback, M. Busch, T. Droge, F.-O. Mahling, C. Prellberg, 'Quantitative IR Spectroscopic Analysis of Ethylene–acrylate Copolymers', *Eur. Polym. J.*, **33**(3), 375–379 (1997).
160. ASTM Designation: D 3900-95, 'Standard Test Methods for Rubber Raw-determination of Ethylene Units in EPM (Ethylene-propylene Copolymers) and EPDM (Ethylene-propylene-diene Terpolymers)', 1998 Annual Book of ASTM Standards, Vol. 9.01, American Society for Testing and Materials, West Conshohocken, PA, 616–631.
161. Analytical Method-specification Prepared by Analytical Sciences Laboratory Exxon Research and Engineering Company Annandale, New Jersey, 'Determination of 5-Ethylidenenorbornene (ENB) by Fourier Transform Infrared Spectroscopy (FTIR)', Materiaux et Techniques, Nov.–Dec., 69–72, 1991.
162. V.J.I. Zichy, 'Quantitative Infrared Analysis of Polymeric Materials', in *Laboratory Methods in Infrared Spectroscopy*, 2nd edition, eds. R.G.J. Miller, B.C. Stace, Heyden & Son Ltd, 48–70, 1979.
163. H.A. Willis, J.M. Chalmers, M.W. Mackenie, D.J. Barnes, 'Novel Quantitative Polymer Analysis Through Computerized Infrared Spectroscopy', *ASTM Spec. Tech. Publ.*, **934**, 58–77 (1987).
164. G. Ellis, A. Sanchez, P.J. Hendra, H.A. Willis, J.M. Chalmers, J.G. Eaves, W.F. Gaskin, K.-N. Kruger, 'Fourier Transform Vibrational Spectroscopy in the Study of Poly(aryl ether sulphone), Poly(aryl ether ether sulphone) and their Copolymers', *J. Mol. Struct.*, **247**, 385–395 (1991).
165. J.M. Chalmers, N.J. Everall, 'FTIR, FT-Raman and Chemometrics: Applications to the Analysis and Characterisation of Polymers', *TRAC*, **15**(1), 18–25 (1996).
166. J.M. Chalmers, V.J.I. Zichy, H.A. Willis, ICI Plastics Division, Unpublished work and private communication.
167. J.R. Parker, W.H. Waddell, 'Quantitative Characterization of Polymer Structure by Photoacoustic Fourier Transform Infrared Spectroscopy', *J. Elastomers and Plastics*, **28**, 140–160 (1996).
168. D. Bruck, 'IR-spectrometric Determination of the Proportions of Acrylonitrile, Butadiene and Hydrogenated Butadiene in Hydrogenated Acrylonitrile-butadiene Rubbers (HNBR)', *Kautsch. Gummi. Kunstst.*, **42**(3), 194–197 (1989).
169. A.J. Marshall, I.R. Jobe, T. Dee, C. Taylor, 'Determination of the Degree of Hydrogenation in Hydrogenated Nitrile-butadiene Rubber (HNBR)', *Rubber Chem. Technol.*, **63**(2), 244–255 (1990).
170. D.M. Cheng, I.J. Gardner, H.C. Wang, C.B. Frederick, A.H. Dekmezian, 'Spectroscopic Studies of the Structures of Butyl and Bromobutyl Rubbers', *Rubber Chem. Technol.*, **63**(2), 265–275 (1990).
171. W.M. Doyle, 'Chemical Composition Monitoring Using an in-situ Infrared Probe', *Adv. Instrum. Control*, **45**(2), 721–738 (1990).
172. E. Mertzel, J.L. Koenig, 'Application of FT-IR and NMR to Epoxy Resins', *Adv. Polym. Sci.*, **75**, 73–112 (1986).
173. R.O. Allen, P. Sanderson, 'Characterization of Epoxy Glues with FTIR', *Appl. Spectrosc. Reviews*, **24**(3&4), 175–187 (1988).
174. R.E. Smith, F.N. Larsen, C.L. Long, 'Epoxy Resin Cure. II. FTIR Analysis', *J. Appl. Polym. Sci.*, **29**, 3713–3726 (1984).
175. H. Ishida, M.E. Smith, 'Kinetic Parameters of the Thermoplastic/Thermoset Epoxy Reaction by Fourier Transform Infrared Spectroscopy', *Polym. Eng. Sci.*, **32**(2), 136–145 (1992).
176. A. Garton, W.T.K. Stevenson, S.P. Wang, 'The Crosslinking of Epoxy Resins at Interfaces. V. Amine-cured Resins at Carbon and Graphite Surfaces', *J. Polym. Sci.*, **26**, 1377–1391 (1988).
177. R. Tadros, D.C. Timm, 'Fourier-transform Infrared Analysis of a Linear, Anhydride-cured Epoxy', *Macromolecules*, **28**, 7441–7446 (1995).
178. Y.T. Liao, J.L. Koenig, 'Applications of Fourier Transform Infrared Spectroscopy to the Study of Fiber-resin Composites', in *Developments in Reinforced Plastics-4*, ed. G. Pritchard, Elsevier Applied Science Publishers, Barking, 31–87, 1984.
179. R.W. Snyder, 'In Situ FT-IR Analysis of Polyimide Curing', Polyimides: Mater; Chem. Charact., Proc. Int. Conf. Polyimides, 3rd, 363–369, 1989.
180. R.W. Snyder, P.C. Painter, 'Dynamic FT-IR Analysis of Cure Reactions and Kinetics of Polyimides', *Polym. Mater. Sci.*, **59**, 57–60 (1988).
181. R.W. Snyder, B. Thomson, B. Bartges, D. Czerniawski, P.C. Painter, 'FTIR Studies of Polyimides: Thermal Curing', *Macromolecules*, **22**, 4166–4172 (1989).

182. D. Garcia, T.T. Serafini, 'FTIR Studies of PMR-15 Polyimides', *J. Appl. Polym. Sci.*, **25**, 2275–2282 (1987).
183. K.C. Cole, P. Van Gheluwe, M.J. Hebrard, J. Leroux, 'Flexible Polyurethane Foam. I. FTIR Analysis of Residual Isocyanate', *J. Appl. Polym. Sci.*, **34**, 395–407 (1987).
184. R.D. Priester, Jr, J.V. McClusky, R.E. O'Neill, R.B. Turner, M.A. Harthcock, B.L. Davis, 'FT-IR-A Probe into the Reaction Kinetics and Morphology Development of Urethane Foams', *J. Cellular Plastics*, **26**, 346–367 (1990), and, Polyurethanes 90, Proc. SPI – 33rd Ann. Polyurethane Tech./Mark. Conf., Orlando, 527–539, 1990.
185. R. Priester, 'Fourier Transform Infra-red – a Useful Tool to Look at Cure and Morphology', *Urethanes Technol.*, **8**(4), 29–32 (1991).
186. S.S. Jada, 'The Structure of Urea–formaldehyde Resins', *J. Appl. Polym. Sci.*, **35**, 1573–1592 (1988).
187. M.J. Elwell, A.J. Ryan, H.J.M. Gruenbaurer, H.C. Van Lieshout, J.A. Thoen, 'A Correlated View of Reaction and Structure Development During the Reactive Processing of Water Blown, Flexible Polyurethane Foam', Proc. Conf. Cellular Polymers II, Paper 9, 1–10, 1993.
188. K.-H. Dorner, 'Foam Chemistry by FTIR', *Urethanes Technology*, June/July 22 (1993).
189. D.C. Compton, S.L. Hill, N.A. Wright, M.A. Druy, J. Piche, W.A. Stevenson, D.W. Vidrine, 'In Situ FT-IR Analysis of a Composite Curing Reaction Using a Mid-infrared Transmitting Optical Fiber', *Appl. Spectrosc.*, **42**(6), 972–979 (1988).
190. E. Margalit, H. Dodiuk, E.M. Kosower, A. Katzir, 'Silver Halide Infrared Fiber-optic Evanescent Wave Spectroscopy for In Situ Monitoring of the Chemical Processes in Adhesive Curing', *Surface and Interface Analysis*, **15**, 473–478 (1990).
191. M.A. Druy, L. Elandjian, W.A. Stevenson, 'Composite Cure Monitoring with Infrared Transmitting Optical Fibers', SPIE Vol. 986, Proc. Fiber Optic Smart Structures and Skins, 130–134, 1988.
192. M.A. Druy, P.R. Young, W.A. Stevenson, D.A.C. Compton, 'In-situ Composite Cure Monitoring Using Infrared Transmitting Optical Fibers', *SAMPE J.*, **25**(2), 11–16 (1988).
193. R.D. Priester, Jr, J.V. McClusky, D.I. Cortelek, P.S. Carleton, J.R. Porter, J.A. De Haseth, 'In-mold RIM Kinetics via Mid-infrared Fiber Optic FTIR', Proc. 34th Ann. Polyurethanes Conf., 21–24 October, New Orleans, 206–215, 1992.
194. C. Dekker, K. Moussa, 'A New Method for Monitoring Ultra-fast Photopolymerisations by Real-time Infrared (RTIR) Spectroscopy', *Makromol. Chem.*, **189**, 2381–2394 (1988).
195. C. Dekker, K. Moussa, 'Real-time Monitoring of Ultra-fast Curing by UV-radiation and Laser Beams', *J. Coatings Technology*, **62**(786), 55–61 (1990).
196. C. Dekker, K. Moussa, 'Recent Advances in UV-curing Chemistry', Proc. RadTech. Conf., 260–274, Vol. 1, 1992.
197. K.S. Anseth, C. Dekker, C.N. Bowman, 'Real-time Infrared Characterization of Reaction Diffusion During Multifunctional Monomer Polymerisations', *Macromolecules*, **28**, 4040–4043 (1995).
198. J.M. Julian, A.M. Millon, 'Quality Control Testing of UV Curable Coatings Using FTIR', *J. Coatings Technol.*, **60**(765), 89–95 (1988).
199. D.R. Hahn, J.W. Hellgeth, M.A. Zumbrum, T.C. Ward, 'Real Time Mechanical and Spectroscopic Studies of Photocured Polyacrylates', Proc. 14th Ann. Meeting Adhesion Soc., 71–72, 1991.
200. D.B. Yang, 'Kinetic Studies of Photopolymerization Using Real Time FT-IR Spectroscopy', *J. Polym. Sci., Part A: Polym. Chem.*, **31**, 199–208 (1993).
201. S.H. Hamid, W.H. Prichard, 'Application of Infrared Spectroscopy in Polymer Degradation', *Polym.-Plast. Technol. Eng.*, **27**(3), 303–334 (1988).
202. J.P. Luongo, 'Infrared Study of Oxygenated Groups formed in Polyethylene During Oxidation', *J. Poly. Sci.*, **XLII**, 139–150 (1960).
203. D.L. Wood, J.P. Luongo, 'Infra-red Spectrum of Polyethylene', *Modern Plastics*, March 1961, 132–144, 198–209 (1961).
204. T.J. Henman, 'Characterisation of Oxidised Polyolefines by Reaction with Sulphur Dioxide', in *Developments in Polymer Degradation-6*, ed. N. Grassie, Elsevier Applied Science, Barking, 107–145, 1985.
205. R. Arnaud, J.-Y. Moisan, J. Lemaire, 'Primary Hydroperoxidation in Low-density Polyethylene', *Macromolecules*, **17**(3), 332–336 (1984).
206. D.L. Tabb, J.J. Sevcik, J.L. Koenig, 'Fourier Transform Study of the Effects of Irradiation on Polyethylene', *J. Polym. Sci.*, **13**, 815–824 (1975).
207. W.-D. Domke, H. Steinke, 'Oxidative Structures in Polyolefins: FT-IR Method of Quantitative Determination', *J. Polym. Sci, A: Polym. Chem.*, **24**, 2701–2705 (1986).
208. R.P. Singh, R. Mani, S. Sivaram, J. Lacoste, J. Lemaire, 'Thermo-oxidative Degradation of Heterophasic Ethylene–propylene Copolymers and their Fractions', *Polym. Int.*, **32**, 189–196 (1993).
209. V. Ollier-Dureault, B. Gosse, 'Photoxidation of Anhydride-cured Epoxies: FTIR Study of the Modifications of the Chemical Structure', *J. Appl. Polym. Sci.*, **70**, 1221–1237 (1998).
210. C. Adam, J. Lacoste, J. Lemaire, 'Photo-oxidation of Elastomeric Materials. Part 1: Photo-oxidation of Polybutadienes', *Polym. Degrad. and Stab.*, **24**, 185–200 (1989).
211. C. Adam, J. Lacoste, J. Lemaire, 'Photo-oxidation of Elastomeric Materials. Part IV: Photo-oxidation of 1,2-Polybutadiene', *Polym. Degrad. and Stab.*, **29**, 305–320 (1990).

212. B. Mattson, B. Stenberg, S. Persson, E. Ostman, 'Thermo-oxidative Degradation of Thick-walled Rubber Materials Studied by IR-technique and Computed X-ray Tomography Scanning', *Rubber Chemistry and Technology*, **63**(1), 23–31 (1990).
213. D.C. Carlsson, S. Chmela, D.M. Wiles, 'Chemical and Spectroscopic Methods for Investigating Polymer Degradation and Stabilization: Oxidative Degradation of a Hydrogenated Nitrile Rubber', *Makromol. Chem., Macromol. Symp.*, **27**, 139–153 (1989).
214. D. Fischer, U. Lappan, I. Hopfe, K.-J. Eichorn, K. Lunkwitz, 'FTIR Spectroscopy on Irradiated Polytetrafluoroethylene', *Polymer*, **39**(3), 573–582 (1998).
215. I.C. McNeill, S. Ahmed, L. Memetea, 'Thermal Degradation of Vinyl Acetate–methacrylic Acid Copolymer and Homopolymers. I. An FTIR Spectroscopic Investigation of Structural Changes in the Degrading Material', *Polym. Degrad. and Stab.*, **47**, 423–433 (1995).
216. C.H. Do, E.M. Pearce, B.J. Bulkin, H.K. Reimschuessel, 'FT-IR Spectroscopic Study on the Thermal and Thermal Oxidative Degradation of Nylons', *J. Polym. Sci, A: Polym. Chem.*, **25**, 2409–2424 (1987).
217. C.H. Do, E.M. Pearce, B.J. Bulkin, H.K. Reimschuessel, 'FT-IR Spectroscopic Study on the Photo- and Photooxidative Degradation of Nylons', *J. Polym. Sci, A: Polym. Chem.*, **25**, 2301–2321 (1987).
218. M.M. Coleman, G.T. Sivy, 'Fourier Transform IR Studies of the Degradation of Polyacrylonitrile Copolymers – I', *Carbon*, **18**, 123–126 (1981).
219. G. Sivy, M.M. Coleman, 'Fourier Transform IR Studies of the Degradation of Polyacrylonitrile Copolymers – II', *Carbon*, **18**, 127–131 (1981).
220. M.M. Coleman, G.T. Sivy, 'Fourier Transform IR Studies of the Degradation of Polyacrylonitrile Copolymers – III', *Carbon*, **18**, 133–135 (1981).
221. G. Sivy, M.M. Coleman, 'Fourier Transform IR Studies of the Degradation of Polyacrylonitrile Copolymers – IV', *Carbon*, **18**, 137–139 (1981).
222. B. Mailhot, J.-L. Gardette, 'Polystyrene Photooxidation. 1. Identification of the IR-absorbing Photoproducts formed at Short and Long Wavelengths', *Macromolecules*, **25**, 4119–4126 (1992).
223. B. Mailhot, J.-L. Gardette, 'Polystyrene Photooxidation. 2. A Pseudo Wavelength Effect', *Macromolecules*, **25**, 4127–4133 (1992).
224. Y. Israel, J. Lacoste, J. Lemaire, R.P. Singh, S. Sivaram, 'Photo- and Thermoinitiated Oxidation of High-impact Polystyrene. I. Characterization by FT-IR Spectroscopy', *J. Polym. Sci., A: Polym. Chem.*, **32**, 485–493 (1994).
225. D.M. Smith, W.F. Welch, S.M. Graham, A.R. Chughtai, 'Metal Substrates and the Photodegradation of Polymers II: An FTIR Study of Polyethylene Terephthalate and Polyvinyl Fluoride', *Solar Energy Materials*, **19**, 111–120 (1989).
226. D.R. Bauer, M.C. Paputa Peck, R.O. Carter, 'Evaluation of Accelerated Weathering Tests for a Polyesterurethane Coating Using Photoacoustic Infrared Spectroscopy', *J. Coatings Tech.*, **59**(755), 103–109 (1987).
227. R.M. Paroli, K.C. Cole, A.H. Delgado, 'Weatherability of Polyurethane Sealants: A Spectroscopic Approach', *Polym. Mater, Sci. Eng.*, **71**, 435–436 (1994).
228. K.C. Cole, I.G. Casella, 'Fourier Transform Infra-red Study of the Thermal Degradation in Poly(ether ether ketone)-carbon Composites', *Polymer*, **34**(4), 740–745 (1993).
229. G.W. Somsen, E.J.E. Rozendom, C. Gooijer, N.H. Velthorst, U.A.Th. Brinkman, 'Polymer Analysis by Column Liquid Chromatography Coupled Semi-on-line with Fourier Transform Infrared Spectrometry', *Analyst*, **121**, 1069–1074 (1996).
230. J. Sidwell, 'Identification of Additives in Rubber and Plastic Materials by Combined Liquid Chromatography/Infrared Spectroscopy', Paper 6, Proc. Conf. Polymer Testing '96, Shawbury, 5–6 Sept., 1–6, 1996.
231. A.M. Robertson, D. Littlejohn, M. Brown, C.J. Dowle, 'Assessment of a Thermospray Interface for the Coupling of High-performance Liquid Chromatography and Fourier Transform Infrared Spectrometry', *J. Chromatog.*, **588**, 15–24 (1991).
232. J.A. Jansen, 'On-line Liquid Chromatography – Fourier Transform Infrared Spectroscopy for the Analysis of Polymers and Additives', *Fresenius J. Anal. Chem.*, **337**, 398–402 (1990).
233. L.M. Wheeler, J.N. Willis, 'Gel Permeation Chromatography/Fourier Transform Infrared Interface for Polymer Analysis', *Appl. Spectrosc.*, **47**(8), 1128–1130 (1993).
234. T. Provder, M. Whited, D. Huddleston, C-Y. Kuo, 'Characterization of Compositional Heterogeneity in Copolymers and Coatings Systems by GPC/FTIR', *Prog. Org. Coat.*, **32**, 155–165 (1997).
235. M.X. Liu, J.L. Dwyer, 'Quantitative Analysis of Copolymers by GPC/FT-IR', *Appl. Spectrosc.*, **50**(3), 349–356 (1996).
236. J.N. Willis, J.L. Dwyer, M.X. Liu, 'Polymer Characterization Using SEC-FTIR', *Int. J. Polym. Anal. Charact.*, **4**(1), 21–29 (1997).
237. M.W. Raynor, K.D. Bartle, I.L. Davies, A. Williams, A.A. Clifford, J.M. Chalmers, B.W. Cook, 'Polymer Additive Characterization by Capillary Supercritical Fluid Chromatography/Fourier Transform Infrared Microspectrometry', *Anal. Chem.*, **60**, 427–433 (1988).
238. J.M. Chalmers, M.W. Mackenzie, J.L. Sharp, R.N. Ibbett, 'Characterization of Thin-layer Chromatographically Separated Fractions by Fourier Transform Infrared Diffuse Reflectance Spectrometry', *Anal. Chem.*, **59**, 415–418 (1987).
239. P.V.E. Krusemann, J.A.J. Jansen, 'Application of On-line Thermal Desorption–gas Chromatography–Fourier-transform Infrared Spectroscopy–mass Spectrometry

for the Characterisation of Polymers', *J. Chromatog., A*, **819**, 243–248 (1998).
240. 'Standard Test Methods for Rubber – Identification by Infrared Spectrophotometry', Designation: D3677-90, 1998 Annual Book of ASTM Standards, Vol. 9.01, American Society for Testing and Materials, West Conshohocken, PA, USA, 566–586.
241. D.L. Harms, 'Identification of Complex Organic Materials by Infrared Spectra of Their Pyrolysis Products', *Anal. Chem.*, **25**(8), 1140–1155 (1953).
242. I.A. Degen, 'A Simple Microtechnique for Obtaining Pyrolyzate Infrared Spectra from Polymeric Materials', *Appl. Spectrosc.*, **44**(9), 1587–1588 (1990).
243. B. Cleverly, R. Herrmann, 'Rapid Identification of Elastomers and Their Additives', *J. Appl. Chem.*, **10**, 192–195 (1960).
244. J.W. Washall, T.P. Wampler, 'Direct-pyrolysis Fourier Transform-infrared Spectroscopy for Polymer Analysis', *Spectroscopy*, **6**(4), 38–42 (1991).
245. M. Pellow-Jarman, M. Hetem, 'Comparison of the Thermal Degradation Products of Poly(butylene terephthalate) and Flame Retardant Poly(butylene terephthalate) Formulations Using a Pyrolysis FTIR Cell', *Polym. Degrad. and Stab.*, **47**, 413–421 (1995).
246. M. Beltran, A. Marcilla, 'Fourier Transform Infrared Spectroscopy Applied to the Study of PVC Decomposition', *Eur. Polym. J.*, **33**(7), 1135–1142 (1997).
247. S.A. Liebman, D.H. Ahlstrom, P.R. Griffiths, 'On-line Fourier Transform Infrared Analysis of Pyrolysis and Combustion Products', *Appl. Spectrosc.*, **30**(3), 355–357 (1976).
248. D. Weber, 'Pyrolysis-GC-FTIR Spectroscopy of Butadiene–acrylonitrile Copolymers', *Int. Lab.*, September, 50–54 (1991).
249. J.O. Lephardt, R.A. Fenner, 'Fourier Transform Evolved Gas Analysis: Additional Considerations and Options', *Appl. Spectrosc.*, **35**(1), 95–101 (1981).
250. C.-Y. Kuo, T. Provder, 'Fourier Transform Infrared Spectroscopic Method for Evolved-gas Analysis', in *Polymer Characterization: Physical Property, Spectroscopic, and Chromatographic Methods*, Advances in Chemistry Series 227, American Chemical Society, Washington, DC, 343–355, 1990.
251. V. Rossiter, R. Millward, M. Simpson, 'Direct Evolved Gas Analysis/FTIR', *Int. Labmate*, **XIV**(IV), 27–29, July (1989).
252. R. Kallonen, 'Smoke Gas Analysis by FTIR Method. Preliminary Investigation', *J. Fire Sciences*, **8**, 343–360 (1990).
253. K.T. Paul, 'Use of FTIR to Analyse Fire Gases from Burning Rubbers and Plastics', Flame Retardants 94 Conf. Proc., London 27–28 January, 251–262, 1994.
254. M.L. Mittleman, D. Johnson, C.A. Wilkie, 'TGA/FTIR: A Synergistic Combination', *Trends in Polym. Sci.*, **2**(11), 391–398 (1994).
255. H.G. Schild, 'Application of TGA/FTIR to the Thermal Degradation Mechanism of Tetrafluoroethylene-propylene Copolymers', *J. Polym. Sci., A: Polym. Chem.*, **31**, 1629–1632 (1993).
256. H.G. Schild, 'Thermal Degradation of Poly(methacrylic acid): Further Studies Applying TGA/FTIR', *J. Polym. Sci., A: Polym. Chem.*, **31**, 2403–2405 (1993).
257. D. Boschel, M. Fedtke, W. Geyer, 'Investigation of Modified Phenolic Hardeners and Curing of an Epoxy Resin by TG-FTIR', *Polymer*, **38**(6), 1291–1296 (1997).
258. W.H. McClennen, R.M. Buchanan, N.S. Arnold, J.P. Dworzanski, H.L.C. Meuzelaar, 'Thermogravimetry/Gas Chromatography/Mass Spectrometry and Thermogravimetry/Gas Chromatography/Fourier Transform Infrared Spectroscopy: Novel Hyphenated Methods in Thermal Analysis', *Anal. Chem.*, **65**, 2819–2823 (1993).
259. R.C. Wieboldt, S.R. Lowry, R.J. Rosenthal, 'TGA/FT-IR: Thermogravimetric Analysis with Fourier Transform Infrared Detection of Evolved Gases', Proc. 6th Int. Conf. Four. Trans. Spectrosc. Vol. II, Mikrochim. Acta (Wien), I, 179–182, 1988.
260. K. Ichimura, H. Ohta, T. Tajima, T. Okino, 'Combined System and Its Application of Thermogravimetric Analysis/FT-IR', Proc. 6th Int. Conf. Four. Trans. Spectrosc. Vol. II, Mikrochim. Acta (Wien), I, 157–161, 1988.
261. B.J. Stay, J.M. Chalmers, M.W. Mackenzie, D.R. Moseley, 'A Method for Determining the Optical Properties of Materials with Respect to Thermal Radiation and Radiation from a Number of Other Sources', *Appl. Spectrosc.*, **39**(3), 412–417 (1985).
262. M. Papini, 'Study of the Radiative Properties of Natural and Synthetic Fibers in the 0.25–2.5 μm Region', *Appl. Spectrosc.*, **43**(8), 1475–1481 (1989).
263. M. Papini, 'Influence of the Orientation of Polypropylene Fibers on their Radiative Properties', *Appl. Spectrosc.*, **48**(4), 472–476 (1994).
264. M. Papini, 'Study of the Radiative Properties of Powdered and Fibrous Polymers', *Vibrational Spectroscopy*, **11**(1), 61–68 (1996).
265. C.A. Sergides, A.R. Chughtai, D.M. Smith, 'Determination of Optical Constants Using a Fourier Transform Infrared Reflection Method; Polyethylene Terephthalate (PET)', *Appl. Spectrosc.*, **41**(1), 154–157 (1987).
266. C.A. Sergides, A.R. Chughtai, D.M. Smith, P. Schissel, 'Relationship between the Functional Group Concentrations and the Infrared Reflection–absorbance of Polyacrylonitrile Films', *J. Polym. Sci.: Polym. Phys.*, **23**, 1573–1584 (1985).
267. J.D. Webb, G. Jorgensen, P. Schissel, A.W. Czanderna, A.R. Chughtai, D.M. Smith, 'Effect of Absorber Concentrations on IR Reflection–Absorber of Polymer Films on Metallic Substrates', in *Polymers for Solar Energy Utilization*, eds. G.C. Gebelein, D.J. Williams, R. Deanin, American Chemical Society, Washington, DC, 143–167, 1983.

268. J.D. Webb, A.W. Czanderna, P. Schissel, 'Photodegradation of Polymer Films on Reflecting Substrates', *Degrad. Stab. Polym.*, **2**, 373–431 (1989).
269. Z.M. Zhang, G. Lefever-Button, F.R. Powell, 'Infrared Refractive Index and Extinction Coefficient of Polyimide Films', *Int. J. Thermophys.*, **19**(3), 905–916 (1998).
270. R.T. Graf, J.L. Koenig, H. Ishida, 'Optical Constant Determination of Thin Polymer Films', *Appl. Spectrosc.*, **39**(3), 405–408 (1985).
271. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York, 1986.
272. J.R. Birch, E.A. Nicol, 'The FIR Optical Constants of the Polymer TPX', *Infrared Physics*, **24**(6), 573–575 (1984).
273. J.R. Birch, G.W. Chantry, D.B. Shenton, R.E. Hills, M.E.A. Cudby, H.A. Willis, 'Sailcloth, a Low Loss Material for Millimetre and Submillimetre Wavelengths', *Infrared Physics*, **17**, 425–426 (1977).
274. A. Roseler, *Infrared Spectroscopic Ellipsometry*, Akademie-Verlag, Berlin, 1990.
275. A. Roseler,'Spectroscopic Ellipsometry in the Infrared', *Infrared Physics*, **21**, 349–355 (1981).
276. S. Vallon, B. Drevillon, 'Adhesion of Dielectric Thin Films on Polymer Substrates Studied by UV–Visible and Infrared Ellipsometry', Proc. Euradh 94 Conf., 378–381, 1994.
277. I. Noda, 'Generalized Two-dimensional Correlation Method Applicable to Infrared, Raman, and Other Types of Spectroscopy', *Appl. Spectrosc.*, **47**(9), 1329–1336 (1993).
278. C. Marcott, A.E. Dowrey, I. Noda, 'Dynamic Two-dimensional IR Spectroscopy', *Anal. Chem.*, **66**(21/1), 1065A–1075A (1994).
279. R. Meier, 'Two-dimensional Infrared Spectroscopy – An Exciting New Tool for Polymer Characterization', *Trends in Polym. Sci.*, **2**(5), 153–158 (1994).
280. P.A.M. Steeman, R. Meier, A. Simon, J. Gast, 'The Structure of a Styrene–acrylonitrile/butadiene Polymer Studied by 2D-i.r. Spectroscopy', *Polymer*, **38**(21), 5455–5462 (1997).
281. A. Singhal, L.J. Fina, 'Dynamic Two-dimensional Infrared Spectroscopy. Part I: Melt-crystallized Nylon 11', *Appl. Spectrosc.*, **49**(8), 1073–1078 (1995).
282. A. Singhal, L.J. Fina, 'Dynamic Two-dimensional Infrared Spectroscopy of the Crystal–amorphous Interface Region in Low-density Polyethylene', *Polymer*, **37**(12), 2335–2343 (1996).
283. T. Nakano, S. Shimada, R. Saitoh, I. Noda, 'Transient 2D IR Correlation Spectroscopy of the Photopolymerization of Acrylic and Epoxy Monomers', *Appl. Spectrosc.*, **47**(9), 1337–1344 (1993).
284. D.K. Graff, D.A. Wrobleski, A.L. Marsh, E.M. Kober, M.E. Smith, J.R. Schoonover, 'Analysis of a Polyurethane Elastomer Using Thermal 2D FT-IR', *Polym. Sci. Eng.*, **78**, 24 (1998).
285. IR film thickness gauge, e.g. Infragauge 55, Infrared Engineering, NDC Infrared Engineering, Maldon, UK.
286. D.W. Vidrine, 'Mid-infrared Spectroscopy in Chemical Process Analysis', in *Spectroscopy in Process Analysis*, ed. J.M. Chalmers, Sheffield Academic Press, Sheffield, UK, 96–138, 2000.
287. W.M. Doyle, N.A. Jennings, 'Fourier Transform-infrared Chemical Reaction Monitoring Using an In Situ Deep Immersion Probe', *Spectroscopy*, **5**(1), 34–38 (1990).
288. ReactIR™, Asi Applied Systems, Mettler Toledo GmbH, Schwerzenbach, Switzerland.
289. W.E. Blaser, R.A. Bredeweg, R.S. Harner, M.A. LaPack, A. Leugers, D.P. Martin, R.J. Pell, J. Workman, L.G. Wright, 'Process Analytical Chemistry', *Anal. Chem.*, **67**(12), 47R–70R (1995).
290. G. Weis, M. Volgmann, 'On-line Quality Assurance in Plastics Production', *Kunstoffe German Plastics*, **81**(5), 25–27 (1991).
291. R. Stengler, G. Weis, 'FTIR Measurement on Polymer Melts for On-line Quality Control', *Kunstoffe*, **79**(10), 955–960 (1989).
292. R.A. Fidler, D. Rowe, G. Weis, 'On-line Infrared Process Control of Molten Polymers via a High Pressure, High Temperature Flow Cell', Proc. ANTEC 91 Conf., 850–855, 1991.
293. J.M. Chalmers, N.J. Everall, 'Polymer Analysis and Characterisation by FTIR, FTIR-microscopy, Raman and Chemometrics', *Int. J. Polym. Anal. Charact.*, **5**, 223–245 (1999).
294. R.A. Fidler, 'Online FTIR Analyzes Additive Levels', *Plast. Compd.*, **15**(6), 47, 48, 50 (1992).
295. R.W. Jones, J.F. McClelland, 'Transient Infrared Emission Spectroscopy', *Anal. Chem.*, **61**, 650–656 (1989).
296. R.W. Jones, J.F. McClelland, 'Quantitative Analysis of Solids in Motion by Transient Infrared Emission Spectroscopy Using Hot-gas Jet Excitation', *Anal. Chem.*, **62**(19), 2074–2079 (1990).
297. R.W. Jones, S. Luo, J.F. McClelland, 'Real-time Infrared Spectroscopy of Moving Solids for On-line Analysis', 8th Int. Conf. Four. Trans. Spectrosc., eds., H.M. Heise, E.H. Korte, H.W. Siesler, SPIE, 1575, 220–221, 1992.
298. R.W. Jones, J.F. McClelland, 'On-line Analysis of Solids and Viscous Liquids by Transient Infrared Spectroscopy', *Process Control Qual.*, **4**(4), 253–260 (1993).
299. J.A. Reffner, 'Fourier Transform Infrared Microanalysis', *Rubber & Plastics News*, February 4, 15–18, 1991.
300. K. Krishnan, J.R. Powell, S.L. Hill, 'Infrared Microimaging', in *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, Inc., New York, 85–110, 1995.
301. P. Colarusso, L.H. Kidder, I.W. Levin, J.C. Fraser, J.F. Arens, E.N. Lewis, 'Infrared Spectroscopic Imaging: From Planetary to Cellular Systems', *Appl. Spectrosc.*, **52**(3), 106A–120A (1998).

302. J.A. Reffner, W.T. Wihlborg, S.W. Strand, 'Chemical Microscopy of Surfaces by Grazing Angle and Internal Reflection FTIR Microscopy', *Am. Lab.*, April, 46–50 (1991).
303. J.E. Katon, A.J. Sommer, 'Infrared Microspectroscopy', *Appl. Spectrosc. Rev.*, **25**(3&4), 173–211 (1989–1990).
304. *The Design, Sample Handling, and Applications of Infrared Microscopes*, ed. P.B. Roush, ASTM Special Technical Publication 949, American Society for Testing and Materials, 1987.
305. H.J. Humecki (ed.), *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York, 1995.
306. 'Standard Practice for Identification of Polymer Layers or Inclusions by Fourier Transform Infrared Microspectroscopy (FT-IR)', ASTM Designation: D5477-95, 1998 Annual Book of ASTM Standards, Vol. 8.03, American Society for Testing and Materials, West Conshohocken, PA, 511–516.
307. T. Buffeteau, B. Desbat, D. Eyquem, 'Attenuated Total Reflection Fourier Transform Infrared Microspectroscopy: Theory and Application to Polymer Samples', *Vib. Spectrosc.*, **11**(1), 29–36 (1996).
308. M.W. Tungol, E.G. Bartick, A. Montaser, 'Forensic Analysis of Acrylic Copolymer Fibers by Infrared Microscopy', *Appl. Spectrosc.*, **47**(10), 1655–1658 (1993).
309. M.W. Tungol, E.G. Bartick, A. Montaser, 'The Development of a Spectral Data Base for the Identification of Fibers by Infrared Microscopy', *Appl. Spectrosc.*, **44**(4), 543–549 (1990).
310. S.P. Bouffard, A.J. Sommer, J.E. Katon, S. Godber, 'Use of Molecular Microspectroscopy to Characterize Pigment-loaded Polypropylene Single Fibers', *Appl. Spectrosc.*, **48**(11), 1387–1393 (1994).
311. E.G. Bartick, 'Considerations for Fiber Sampling with Infrared Microspectroscopy', 8th Int. Conf. Four. Trans. Spectrosc., eds. H.M. Heise, E.H. Korte, H.W. Siesler, SPIE, 1575, 64–73, 1992.
312. J.-L. Gardette, 'Micro-FT-IR Spectroscopic Profiling of Aged Polymer Films', *Spectroscopy Europe*, **5**(2), 28–32 (1993).
313. J.-L. Gardette, 'Microspectrophotométrie IR-TF: Application à l'Analyse du Vieillisement des Polymères', *Analusis Magazine*, **21**(5), M17–M21 (1993).
314. X. Jouan, J.-L. Gardette, 'Development of a Micro (FTIR) Spectrophotometric Method for Characterization of Heterogeneities in Polymer Films', *Polymer Communications*, **28**(12), 329–330 (1987).
315. J. Lemaire, J.-L. Gardette, J. Lacoste, 'Use of Micro-FTIR Spectrophotometry in the Determination of Polymer Photoageing Mechanisms', *Makromol. Chem., Macromol. Symp.*, **70/71**, 419–431 (1993).
316. S.C. Hsu, D. Lin-Vien, R.N. French, 'Probing the Concentration Profiles of Additives in Polymers by IR Microspectroscopy: The Diffusion of Cyasorb UV531 in Polypropylene', *Appl. Spectrosc.*, **46**(2), 226–228 (1992).
317. B. Van Mele, E. Verdonck, 'Physico-chemical Characterization of the Fiber/matrix Interaction in Polyethylene Fiber/epoxy Matrix Composites. Part 2. Characterization of the Interphase with FT-IR Microspectroscopy', *Composite Interfaces*, **3**(2), 101–119 (1995).
318. S.M. Stevens, 'The Use of FTIR-microspectrometry for the Measurement of Crystallinity in Polyethylene Welds and Helical Content in Polypropylene Welds', in *Analytical Applications of Spectroscopy II*, eds. A.M.C. Davies, C.S. Creaser, Royal Society of Chemistry, Cambridge, 79–84, 1991.
319. S.M. Stevens, 'Measurement of Crystallinity in Polyethylene Welds and Helical Content in Polypropylene Welds Using FTIR-microscopy', *SPIE 1320 Infrared Technology and Applications*, 18–25 (1990).
320. S.M. Stevens, 'The Evaluation of Molecular Orientation and Structural Changes in Polypropylene Welds', 51st Ann. Tech. Conf. SPE, 9–13 May, New Orleans, 378–380, 1993.
321. D. Gueugnaut, P. Wackernie, J.P. Forgerit, 'Infrared Microspectrometric Study of the Thermal History of Fusion Welding of Polyethylene', *J. Appl. Polym. Sci.*, **35**, 1683–1694 (1988).
322. E. Benedetti, S. Catanorchi, A. D'Alessio, G. Moggi, P. Vergamini, M. Pracella, F. Ciradelli, 'FTIR-microspectroscopy and DSC Studies of Poly(vinylidene fluoride)', *Polymer Int.*, **41**, 35–41 (1996).
323. A. Kaito, M. Kyotani, K. Nakayama, 'Orientation Profiles in the Strand of Thermotropic Liquid–crystalline Polymer Studied by Polarized Fourier Transform Infrared Microspectroscopy', *Macromolecules*, **24**(11), 3244–3249 (1991).
324. M.A. Harthcock, S.C. Atkin, 'Imaging with Functional Group Maps Using Infrared Microspectroscopy', *Appl. Spectrosc.*, **42**(3), 449–455 (1988).
325. M.A. Harthcock, S.C. Atkin, 'Infrared Microspectroscopy: Development and Applications of Imaging Capabilities', in *Infrared Microspectroscopy. Theory and Applications*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, Inc., New York, 21–39, 1988.
326. L.H. Kidder, I.W. Levin, E.N. Lewis, V.D. Kleiman, E.J. Heilwell, 'Mercury Cadmium Telluride Focal-plane Array Detection for Mid-infrared Fourier-transform Spectroscopic Imaging', *Optics Letters*, **22**(10), 742–744 (1997).
327. C.M. Sniveley, J.L. Koenig, 'Characterizing the Performance of a Fast FT-IR Imaging Spectrometer', *Appl. Spectrosc.*, **53**(2), 170–177 (1999).
328. C.D. Arvanitopoulos, J.L. Koenig, 'FT-IR Microspectroscopic Investigation of the Interphase of Epoxy Resin–glass Fiber Reinforced Composites', *Appl. Spectrosc.*, **50**(1), 1–10 (1996).
329. C.D. Arvanitopoulos, J.L. Koenig, 'Infrared Spectral Imaging of the Interphase of Epoxy–glass Fiber Reinforced Composites Under Wet Conditions', *Appl. Spectrosc.*, **50**(1), 11–18 (1996).

330. J. Kressler, R. Schäfer, R. Thomann, 'Imaging of Semicrystalline Polymers and Polymer Blends by FT-IR Microspectroscopy', *Appl. Spectrosc.*, **52**(10), 1269–1273 (1998).
331. R. Schäfer, J. Zimmermann, J. Kressler, R. Mülhaupt, 'Morphology and Phase Behavior of Poly(methyl methacrylate)/Poly(styrene-*co*-acrylonitrile) Blends Monitored by FTIR Microscopy', *Polymer*, **38**(15), 3745–3752 (1997).
332. S.R. Challa, S.-Q. Wang, J.L. Koenig, 'Characterization of Polymer-dispersed Liquid Crystal Systems by FT-IR Microspectroscopy', *Appl. Spectrosc.*, **49**(3), 267–272 (1995).
333. S.R. Challa, S.-Q. Wang, J.L. Koenig, 'In Situ Diffusion and Miscibility Studies of Thermoplastic PDLC Systems by FT-IR Microspectroscopy', *Appl. Spectrosc.*, **51**(3), 297–303 (1997).
334. L. Neumeister, J.L. Koenig, 'Crack Analysis of Unfilled Natural Rubber Using Infrared Microspectroscopy', *Rubber Chem. Technol.*, **70**(2), 271–282 (1997).
335. L. Neumeister, J.L. Koenig, 'Crack Analysis of Unfilled Natural Rubber Using Infrared Microspectroscopy', *Polym. Mater. Sci. Eng.*, **74**, 266–267 (1996).
336. J.M. Chalmers, N.J. Everall, K. Hewitson, M.A. Chesters, M. Pearson, A. Grady, B. Ruzicka, 'Fourier Transform Infrared Microscopy: some Advances in Techniques for Characterisation and Structure–property Elucidations of Industrial Materials', *Analyst*, **123**, 579–586 (1998).

Inverse Gas Chromatography in Analysis of Polymers

James Guillet
University of Toronto, Canada

Zeki Y. Al-Saigh
Columbus State University, USA

1 Introduction 1
2 Conventional Gas Chromatography 2
3 Inverse Gas Chromatography 3
4 The Glass Transition 5
5 Surface Studies 6
5.1 Adsorption Isotherms 6
5.2 Surface Areas 7
5.3 Surface Energy of Polymers 8
6 Crystallinity 8
6.1 Degree of Crystallinity 8
6.2 Rates of Crystallization 10
7 Diffusion 10
7.1 Determination of Diffusion Constants 11
8 Thermodynamic Studies 13
8.1 Activity Coefficients, Heats of Mixing and Heats of Solution 13
8.2 Finite Concentration Gas Chromatography 14
8.3 Henry's Law Constants (Solubilities) 14
8.4 Solubility Parameters 15
8.5 Hydrogen Bonding 16
9 Polymer Blends 17
9.1 Experimental Studies on Polymer Blends 19
9.2 Amorphous Plasticizer Blends 19
10 Inverse Gas Chromatography Studies on Natural and Synthetic Fibers 20
11 Biological Molecules 20
12 Applications 20
12.1 Amorphous and Semicrystalline Polymer Blends 21
12.2 Comparison with Other Methods 21
13 Experimental 22
13.1 Thermodynamic Measurement on Blends 22
13.2 Column Preparation 23
13.3 Coating the Support: the Pile Method 24
13.4 Determination of Degree of Crosslinking of Polymer Networks 24
Abbreviations and Acronyms 26
Related Articles 26
References 26

This article describes the application of inverse gas chromatography (IGC) to the study of synthetic and natural polymers. The word ''inverse'' is used to indicate that the component of interest is the stationary phase, either as a finely divided powder or coating, dispersed on a suitable inert support and packed into a chromatographic column. The time required for a probe molecule to pass through the column gives a measure of the molecular interactions between the probe and the polymer which can be quantified with the help of chromatographic theory. Historically, most thermodynamic studies on polymers were carried out in dilute solution, but IGC provides information on condensed phases under conditions much more similar to those under which polymers are actually used.

Using this technique, melting, glass and other solid-phase transitions can be studied and quantified. Degrees of crystallinity can be determined in an unambiguous procedure without calibration by other methods. Solubility, permeability and diffusion constants can be determined for probe molecules. IGC is also used extensively in determining the permeability of additives such as antioxidants in polymers, as well as thermodynamic quantities such as activity coefficients, heats of solution, Flory–Huggins interaction parameters (χ), and solubility parameters. Surface areas can be measured at any desired temperature by the determination of the adsorption isotherms of suitable probes. Measurements on polymer blends can yield important information on polymer–polymer interactions and can be used to predict the compatibility of the components over a wide range of temperatures. The use of finite concentration IGC provides a rapid method of determining the density of crosslinks in rubber-like polymers.

Almost any commercial gas chromatograph can be easily modified to carry out IGC experiments by procedures described in this report. The method provides a wealth of information of both fundamental and practical importance in the study of polymeric materials.

1 INTRODUCTION

Much of what is now known about the structure and chemical interactions of macromolecules comes from

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

studies in dilute solution where the polymer molecules are more or less isolated from each other. However, in most practical applications of polymers, the polymer is highly concentrated and usually represents 90% or more of the bulk phase. Experimental techniques developed for dilute solution studies are often inapplicable under these conditions. Furthermore, there is a large and increasingly important category of polymers that are insoluble in all known solvents and hence cannot be studied at all in dilute solution.

The inverse gas chromatographic procedure (sometimes called the molecular probe technique) eliminates both of these difficulties. The polymer is studied in the solid phase under conditions which can be made to approximate conditions of use, or processing and fabrication. Furthermore, although polymers may be insoluble in solvents, virtually all small organic molecules will have measurable solubilities in solid organic polymers, even when the latter are crosslinked or highly crystalline. Hence the range of interactions which can be probed by the technique is virtually unlimited.

In view of the general availability of gas chromatographic equipment, the experimental simplicity, and the ease with which large amounts of data can be collected, IGC is becoming the method of choice for the study of thermodynamic interactions of small molecules with polymers in the solid phase. However, the method is not limited to equilibrium measurements in the bulk phase. It can also be used to measure surface areas and adsorption isotherms, glass and other solid-phase transitions in polymers, degrees of crystallinity, crosslink density and diffusion constants for small molecules in polymeric materials.

2 CONVENTIONAL GAS CHROMATOGRAPHY

Gas chromatography (GC), like any other chromatographic technique, is based on the distribution of a probe molecule between two phases. In gas–solid chromatography (GSC) the phases are gas and solid; as the injected compound is carried by the gas phase through a column filled with solid phase, partition occurs via the sorption–desorption of the solute as it travels past the solid. Superimposed upon the forward velocity is radial motion of the probe molecules caused by random diffusion throughout the stationary phase. Separation of two or more components injected simultaneously is due to their differing affinities for the stationary phase. In gas–liquid chromatography (GLC) the stationary phase is a liquid coated onto a solid support. The mathematical treatment is equivalent for GLC and GSC.

There are two mechanisms of gas–solid interaction, namely *sorption* of the solute into the *bulk* stationary phase (absorption) or on the *surface* of the stationary phase (adsorption), or a combination of both. For conventional GC, the surface adsorption term is considered to be negligible. The net volume required to move the probe molecules through the column is V_N, the total volume of gas needed minus the "dead" volume in the column, found by sending an inert probe such as methane through the column (Figure 1). V_N is equal to $(t_R - t_M) \times U_T$, where t_R is the probe retention time, t_M is the marker retention time and U_T is the flow rate of carrier gas at the temperature at which it is measured. The parameter used in further calculations is V_g, the specific retention volume, Equation (1)

$$V_g = \frac{273.16}{T} \frac{V_N}{w} \frac{760}{P_o} J_2^3 \tag{1}$$

where T (K) is the temperature of the column at which the flow rate is measured, P_o is the column outlet pressure, w is the number of grams of polymer in the column and J_2^3 is a small correction due to the pressure drop along the column equal to $(2/3)\{[(P_i/P_o)^2 - 1]/[(P_i/P_o)^3 - 1]\}$ where P_i and P_o are the inlet and outlet pressure on the column.

Usually the retention volume is obtained using the experimental peak maximum to define the retention time. In this treatment for conventional GC, only bulk adsorption is assumed, and band broadening effects and the existence of a nonlinear sorption isotherm are not considered.

Everett[1] developed the thermodynamic analysis for a binary solution of components 1 (probe) and 2 (stationary phase) in the presence of a gas which is insoluble in the solution. Assuming that the molar volume of the probe, V_1, does not vary greatly with pressure, that the gas phases are only slightly imperfect, that the system is in equilibrium, and that the solute is infinitely dilute in both phases, then the infinite dilution activity coefficient of component 1 (γ_1^∞) at temperature T and total pressure P

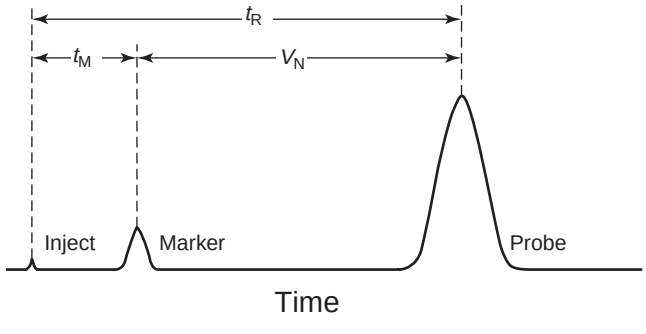

Figure 1 Information obtained for a probe interacting with a polymer-coated stationary phase at temperature T.

can be written as Equation (2)

$$\ln \gamma_1^{\infty} = \ln \left(\frac{n_L RT}{K V_L p_1^0} - \frac{(B_{11} - V_1) p_1^0}{RT} + \frac{(2B_{13} - V_1^{\infty}) P}{RT} \right) \quad (2)$$

where n_L is the number of moles of component 2 occupying volume V_L on the column, p_1^0 is the partial pressure of 1 in the vapor phase, R is the gas constant, B_{11} is the second virial coefficient for the probe, B_{13} is the mixed virial coefficient of the solute vapor and carrier gas, V_1^{∞} is the partial molar volume of 1 at infinite dilution, P is the total pressure and K is the equilibrium partition coefficient, defined as the ratio of concentration of solute in the stationary phase, q, to that in the gas phase, c, i.e. $K \equiv q/c$.

Mixed virial coefficients are difficult to find: if the carrier gas is used at moderate pressures (a pressure of less than 2 atm or 202.65 KN m^{-2}) the last term can be ignored. Rewriting Equation (2) in terms of the specific retention volume gives Equation (3)

$$\ln \gamma_1^{\infty} = \ln \left(\frac{273.16R}{V_g p_1^0 M_1} - \frac{p_1^0 (B_{11} - V_1)}{RT} \right) \quad (3)$$

where M_1 is the molar volume of the probe. Other thermodynamic quantities can be calculated once the activity coefficient is known, for example the excess free energy of mixing, Equation (4)

$$\Delta G_m^e = RT \ln \gamma_1^{\infty} \quad (4)$$

and the excess enthalpy of mixing, Equation (5)

$$\frac{\partial \ln \gamma_1^{\infty}}{\partial (1/T)} = \frac{\Delta H_m^e}{R} \quad (5)$$

3 INVERSE GAS CHROMATOGRAPHY

In an experiment using IGC, the species of interest is the stationary phase which is usually made up of a polymer-coated support, or finely ground polymer mixed with an inert support. A GC column can be packed with such a mixture, or coated on the inside with polymer, to give a capillary column. Some studies have been done using pure polymer in the column[2,3] and using capillary columns.[4] In work on the latter, Gray and Guillet[4] found that V_g values for polystyrene (PS) were slightly higher for the open than for the packed column, possibly because of the higher specific surface area available in the open column. In another study, Lichtenthaler et al.[5] found that the capillary system is more sensitive to carrier gas flow rate, and that the V_g values differed from those obtained using the packed column by as much as 20% for poly(isobutylene) (PIB) and poly(dimethylsiloxane) (PDMS), and even more for poly(vinylacetate) (PVAc). In both cases, the difference decreased as the temperature increased. Both studies reported that the basic disadvantage of the capillary method was the difficulty in calculating the amount of polymer present.

Most of the work discussed here concerns data obtained with columns containing a polymer-coated support. The amount of polymer can be determined from Soxhlet extraction or by calcination of both coated and uncoated support, giving the weight percent of volatile material, and hence the weight of polymer. Errors involved in using IGC to do thermodynamic calculations have been examined.[6] The largest source of error is in the determination of the amount of polymer present. It was concluded that calcination was preferred over extraction owing to the presence of extractable inorganic materials in common supports.

For the characterization of new polymers the most informative IGC data are obtained from a *retention diagram*. In this experiment the value of V_g is determined over a range of temperatures and $\log V_g$ is plotted as a function of $1/T$. An instrument designed to carry out this process automatically has been described,[7–9] and conventional GCs can be readily modified to obtain these data in a similar manner.

A typical curve for a semicrystalline polymer such as polypropylene (PP) is shown in Figure 2. The slope reversals are indicative of phase transitions. Such transitions were noted[10] as early as 1965 for polyethylene (PE) and PP, but the first comprehensive study of polymer structure using IGC was reported in 1969 by Smidsrød and Guillet[11] on poly(*N*-isopropyl acrylamide). The information obtained in such an experiment depends on the temperature region of study. In the sample retention diagram shown,[12] segment AB represents the polymer below its glass transition temperature (T_g). Retention of the probe in this region arises from condensation and adsorption onto the polymer surface, because the probe is unable to diffuse into the bulk of the polymer within the time of passage of the probe peak. The slope of this straight segment is given by $(\Delta H_v - \Delta H_a)/2.3R$, where ΔH_v is the latent heat of vaporization of the probe and ΔH_a is the enthalpy of adsorption of the probe on the polymer surface.

In a study of PS[13] Braun and Guillet examined the change in shape of the retention diagram with respect to the surface-to-volume ratio of the stationary phase. As the amount of polymer on the support decreased, reversal from normal linear behavior became less pronounced (Figure 3) and at low loadings ($\leq$0.36%) the minimum was no longer detectable. However, the data collected

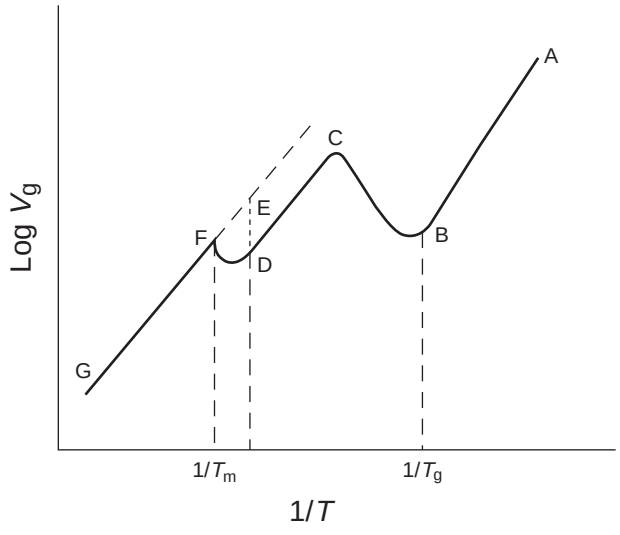

Figure 2 Retention diagram for a semicrystalline polymer. T_g is the glass transition temperature. T_m is the melting temperature. Area A–C represents a nonequilibrium absorption of the probe into the bulk phase; area C–D represents equilibrium absorption of the probe molecules in the amorphous polymer phase; area D–F represents the transition that occurs in crystalline polymers – the melting process; area F–G represents the equilibrium absorption of the probe molecules in the amorphous polymer phase.

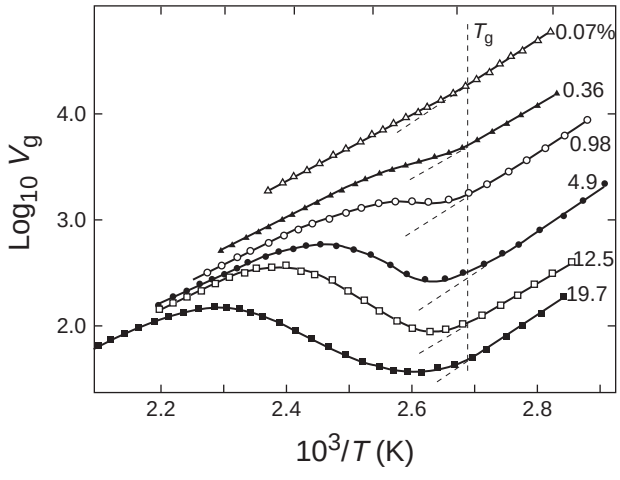

Figure 3 Effect (experimental) of loading (%) on retention diagram for *n*-hexadecane on PS. T_g is taken as the first deviation from linearity of the plot.

on an automatic molecular probe machine is sufficiently precise that in this lower concentration region, T_g can be easily determined from plots of both the first and second derivatives of the experimental plots of log V_g versus $1/T$ (see Figure 4).

The results for PS (Figure 3) showed that the temperature of first deviation from linearity remained constant (96–98 °C). This temperature is interpreted as the glass

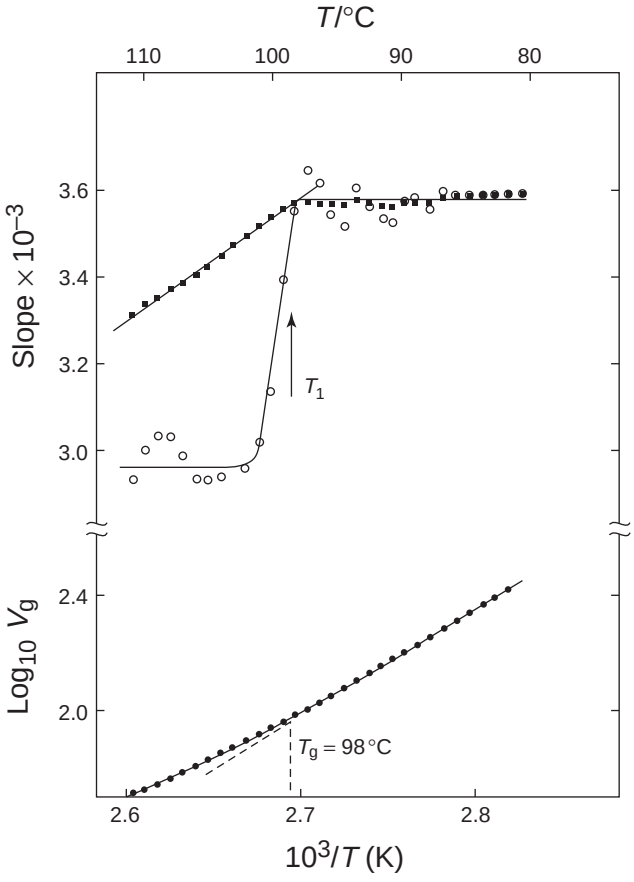

Figure 4 (■) Overall slope and (○) derivative of (●) retention diagram for *n*-hexadecane on PS.

transition T_g. Theoretically this represents the first detectable contribution of bulk sorption to the total retention volume, attributed to the diffusion of the probe into the bulk polymer which reflects the mobility of small 20–40 carbon atom segments, leading to a large increase in its free volume. This is the temperature at which the amorphous polymer changes from the glassy to the rubber-like state. These authors also derived model calculations of retention diagrams using a simple bimodal absorption mechanism which matched the experimental curves (Figure 5).

The next region, corresponding to a change in the slope of the diagram (B–C in Figure 2), represents a nonequilibrium absorption of the probe into the bulk phase. The diffusion rate is slow in this region, hence the molecules injected as a pulse at this temperature would not penetrate through the entire bulk of the polymer during the time of passage of the solute peak. The solute does not have time to reach an equilibrium partition between polymer and carrier gas, hence the V_g value obtained is flow rate dependent. Bulk contribution to the retention volume increases up to point C, which is a

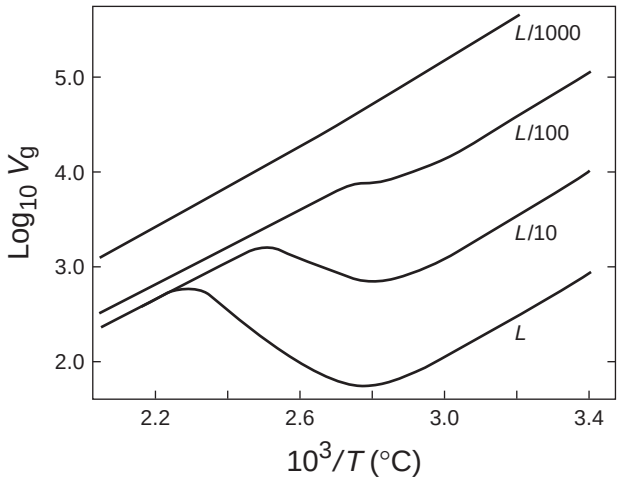

Figure 5 Model calculations of the shape of the retention diagram for a glassy polymer (such as PS) as a function of the thickness of the polymer layer (L).

function of the film thickness. This maximum temperature is reached when the increase in bulk sorption caused by the increase in the probe diffusion constant, D, is balanced by the decrease in retention caused by increased volatility of the probe. In this nonequilibrium region it is possible to obtain information about the probe diffusion coefficient, D, and its dependence on temperature. For noncrystalline polymers, normal liquid-like behavior is achieved at C. Section C–D represents equilibrium absorption of the probe molecules in the amorphous polymer phase. Contributions to V_g come from surface adsorption, bulk absorption and condensation on the surface. The surfaces available to the probe are the polymer–gas and polymer–support interface, and any additional surfaces between crystalline and amorphous regions in the bulk polymer. The slope here is given by Equation (6)

$$\frac{\Delta H_v - \sum_i b_i \Delta H_{a_i} - \Delta H_m}{2.3R} \qquad (6)$$

The b_i values are weighting factors for each surface and ΔH_a are the enthalpies of adsorption; these can be found experimentally[14] and are related to the surface-to-polymer ratio for the polymer film (ΔH_m is the enthalpy of mixing of the polymer and probe.)

The following section of the curve (D–F) represents the next transition that occurs in crystalline polymers – the melting process. Since the crystalline phase is usually impermeable, this section of the plot can give information about the size, shape and distribution of crystalline regions. The final, linear section of the plot represents the wholly amorphous polymer. Extrapolation of this line to lower temperatures and comparison of these values with the true experimental values gives information about the percentage crystallinity of the sample at any temperature, reflected by the difference in specific volume (caused by bulk sorption) between a totally amorphous and partially crystalline product. This percentage crystallinity is given by Equation (7)

$$\%\text{Crystallinity} = 1 - \frac{V_{g(\text{experimental})}}{V_{g(\text{extrapolated})}} \times 100 \qquad (7)$$

Since the use of GC data to obtain thermodynamic quantities is based on the bulk absorption model, it is only in the latter region that experimental data can be used to obtain activity coefficients, and so on.[13] To ensure that the V_g values do represent pure bulk absorption they must be independent of the carrier gas flow rate. Once this is confirmed, the presence of bulk sorption only is verified by measuring V_g for the same probe at several percentage loadings of polymer. If V_g changes, it should be extrapolated to infinite percentage loading. This step is essential before doing any thermodynamic calculations.

4 THE GLASS TRANSITION

Numerous studies have been carried out concerning the effect of experimental variables on the glass transition. Llorente et al.[15] studied poly(cyclohexyl methacrylate) using good and bad solvents as probes; both showed a transition in the retention diagram, but only the nonsolvents gave T_g in quantitative agreement with values from differential scanning calorimetry (DSC) measurements.

Studies have been reported on the glass transitions of PVAc,[16] PP, poly(1-butene), and ethylene–propylene copolymers,[17] as well as polymer blends including PS–butadiene,[18] polyacrylonitrile,[2,19] poly(ethylene terephthalate) (PET)[20] and poly(ε-caprolactam),[21] blends of poly(methyl methacrylate) (PMMA) and poly(vinyl chloride) (PVC),[22] block copolymers of styrene and tetrahydrofuran[23] and graft polymers of methyl methacrylate and stearyl methacrylate.[24] In general, compatible systems are detectable if the mixture demonstrates only a single glass transition.

Braun and Guillet[25] investigated the failure of some probes to detect T_g. They found that the solubility of a probe in the polymer was not enough to characterize the polymer–probe interaction near T_g. Interactions of the probe with both the bulk and the surface of the stationary phase must be considered. To maximize reversal in the $\ln V_g$ versus $1/T$ plot, the bulk contribution should be as large as possible and the surface term as small as possible, hence the bulk to surface ratio, not the solubility of probe in polymer, is the critical parameter.

5 SURFACE STUDIES

5.1 Adsorption Isotherms

For the case of bulk absorption, when the retention volume is independent of probe sample size, the equilibrium concentrations of the probe in the polymer and gas phase are linearly related for low concentrations of the sample.[26] When surface adsorption is present, the isotherm relating these two concentrations is often curved owing to surface heterogeneity and saturation of available sites. In addition, the experimental concentrations used may not be low enough to ensure a linear isotherm; under these conditions the shape of the isotherm can be used to investigate the adsorbate–adsorbant interaction.

There are two approaches using GC to obtain adsorption isotherms. The ''frontal'' technique takes into account kinetic factors and gas phase volume changes due to vapor adsorption. Here[27] the sample is continuously fed into the column. For substances with a Langmuir isotherm (adsorption) the result is a single sharp step, produced at the first exit of the substance. From the time needed for breakthrough, the amount of substance retained can be determined. Usually, the ''elution'' technique[28] is used, where a pulse of material is injected and the shape of the isotherm is found from a single unsymmetrical peak. Using the gas phase concentration of eluted probe vapor, c, the retention volume may be calculated,[29] Equation (8)

$$a = \frac{1}{m} \int_0^c V_c \, dc \tag{8}$$

where a is the amount of probe vapor adsorbed on mass m of adsorbant. If the only peak broadening effect present is nonlinearity of the isotherm, then one side of the elution peak should be vertical and V_c can be determined from the other, diffuse side, as in Figure 6.

Figure 6 shows the change in peak size with varying amounts of n-decane (nC$_{10}$) traveling through a column filled with PMMA-coated support.[26] As the concentration of nC$_{10}$ increases, the front profiles become more diffuse but fall on a common curve while the rear profiles remain almost vertical, fulfilling the experimental requirements needed to use the elution technique in finding the adsorption isotherm. The amount of nC$_{10}$ adsorbed on the column can be described[30] as a function of its partial pressure, p, Equation (9)

$$p = \frac{m_{cal} q R T}{S_{cal} \dot{V}} h \tag{9}$$

where h is the peak height, S_{cal} the calibration peak area on the recorder chart, m_{cal} the number of moles of nC$_{10}$, q the recorder chart speed and $\dot{V}$ the carrier gas flow rate at temperature T. The amount of nC$_{10}$ sorbed per unit

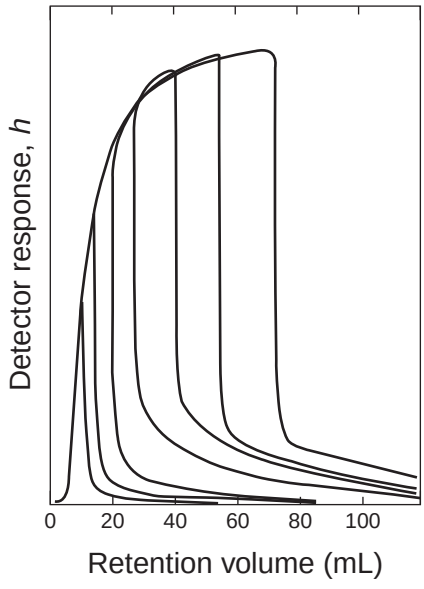

Figure 6 Chromatographic peak shapes for larger injections of decane (0.03, 0.06, 0.15, 0.3, 0.5, 0.7, 1.0 μL) on a column packed with PMMA beads.

weight of sorbant corresponding to the partial pressure p is, Equation (10)

$$a = \frac{m_{cal} S_{ads}}{m S_{cal}} \tag{10}$$

where m is the mass of sorbant and S_{ads} is the chart area bounded by the diffuse profile of the chromatogram. If different areas (S_{ads}) corresponding to different values of h are measured, the isotherm relating a and p may be found, using the above two equations.

Several polymers were investigated by Gray and Guillet[26] using nC$_{10}$ as the probe. For PMMA beads, and PS- and PVAc-coated glass beads, the isotherms represented multilayer adsorption (type II in Brunauer's classification[31]); here the heat of adsorption is equal to or less than the heat of liquification of the adsorbate. As the amount of adsorbate on the surface was increased, the partial vapor pressure approached (as a limit) the vapor pressure of pure adsorbate at T. Apparently, for nC$_{10}$ interaction with glassy polymers, the shape of the isotherm at high coverages is governed primarily by surface saturation. Using the frontal curve in Figure 6 one can construct[26] adsorption isotherms, as shown in Figure 7 at various temperatures.

For nC$_{10}$ on poly(styrene-co-divinylbenzene) beads, and on PS-coated glass beads, the peak retention times increased with decreasing sample size. The rear profiles did not fall on a common curve, implying nonequilibrium conditions and that some kinetic process (probably penetration of the probe into bulk polymer) was the cause of the peak broadening. Bulk sorption below T_g

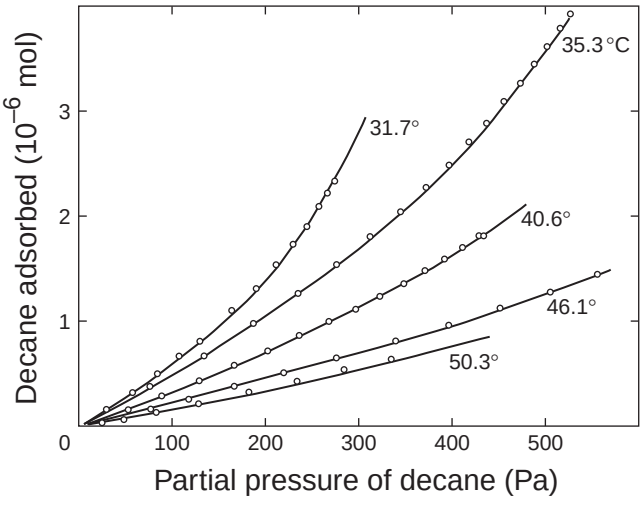

Figure 7 Adsorption isotherms at indicated temperatures for *n*-decane on PMMA blends.

is known, but it is slower and more complex than for above T_g.[32] This explanation was substantiated by a comparison study of nC_{10} with PE (where bulk sorption is rapid) and with glass beads (no bulk sorption). For the latter case, at low concentrations of probe the rear slopes were almost superimposable, hence the factor causing peak broadening was absent; it therefore seems probable that peak tailing was due to limited bulk sorption.

5.2 Surface Areas

Usually it is assumed that bulk and surface processes make independent contributions to the retention volume (V_R),[33] Equation (11)

$$V_R = K_b w_L + K_a A_L \tag{11}$$

where K_b and K_a are the bulk and surface partition coefficients, respectively, w_L is the mass and A_L is the surface area of the stationary phase. If w_L is varied, a plot of V_R/w_L against $1/w_L$ will give an intercept of K_b and a slope of K_a/A_L. To ensure that equilibrium data are being used, the retention volumes must be extrapolated to infinite dilution of probe, zero flow rate of carrier gas, then infinite percentage loading of the column. Knowing A_L as a function of V_L, K_a can be found. This approach assumes negligible contribution from other adsorption processes. In addition, small sample sizes are used so that V_R, K_b and K_a can be assumed to be independent of sample concentration. The surface contribution is expected to be important if the vapor has low solubility (therefore low bulk concentration) in the polymer, or if there are strong dipole–dipole or hydrogen bonding interactions between the probe and the polymer. Below T_g the bulk contribution is often assumed to be negligible, leaving V_R equal to $K_a A_L$. The surface partition coefficient can therefore be found by using a support of known surface area; conversely, by comparison of retention volumes for the same polymer–probe system the surface area of the stationary phase can be found. The surface partition coefficient can be rewritten as Equation (12)

$$\ln K_a = \ln K_{a,0} - \left(\frac{\Delta H_a}{RT}\right. \tag{12}$$

where ΔH_a is the enthalpy of adsorption. If the specific retention volume, V_g, is used, the expression for the surface area is, Equation (13)

$$\ln A_L = \ln V_g + \ln w_L - \ln K_{a,0} + \left(\frac{\Delta H_a}{RT}\right. \tag{13}$$

Another expression for surface area involves the surface area of the uncoated stationary phase and the known amounts of polymer and inert support, Equation (14)

$$A_L = w_L S_{ps} \quad \frac{100}{\chi} - 1 \tag{14}$$

where χ is percentage loading of polymer on the support and S_{ps} is the accessible surface area, i.e. that covered by the polymer, as seen by the probe. From studies on PS[13] it was found that for greater than 1% loading, film thicknesses of 10^3 Å and larger were obtained; for loadings of 19%, the film thickness was 17 000 Å. Below 0.02% loading, it appeared that there was not sufficient material to give monolayer coverage; hence, even above T_g, surface contributions would have to be taken into account when dealing with very thin films.

In a study on PS by Galin and Rupprecht[34] it was found that the value of accessible surface area available to the probe decreased with its molecular weight; this was thought to be related to the relative dimensions of the probe molecules and the smallest pore size in the support. Branched PS[34] and cellulose fibers[35] have also been studied using this technique.

A third method of finding surface areas is the Brunauer–Emmett–Teller (BET) approach, which uses the experimental isotherm. A two-parameter BET equation is Equation (15)

$$\frac{p_1/p_1^0}{v(1-p_1/p_1^0)} = \frac{1}{v_m c} + \quad \frac{c-1}{v_m c} \quad \frac{p_1}{p_1^0} \tag{15}$$

where p_1^0 is the saturated vapor pressure of the solute, v the volume of the solute on the surface, v_m the volume of the solute on the surface supposing monolayer coverage, and c is a constant. Using values for v and p_1/p_1^0 from experimental data, the left side of Equation (15) is plotted against p_1/p_1^0 to give a straight line from whose slope and

intercept v_m can be calculated. Knowing the surface area covered by the probe molecule, the surface area in the column may be found. Results using this approach[26] were found to be in good agreement with geometric surface areas of PMMA beads and of polymers coated on glass beads.

An early example of the use of this technique was carried out by Tremaine and Gray[36] who calculated an adsorption isotherm from the variation of the peak maximum retention volume with sample size on cellulose. Surface areas, enthalpies and entropies of adsorption were found.

In another work[37] by these authors the retention volumes were fitted directly to a modified BET equation to find the monolayer capacity without deriving the adsorption isotherm. Surface area results agreed to within 5% of the values calculated using the adsorption isotherm. The GC method is particularly advantageous using materials like cellulose where the dry and wet polymer have very different properties. By saturating the carrier gas with water vapor (or any other vapor of interest), measurements can be obtained under conditions in which the material is used.

5.3 Surface Energy of Polymers

To quantify the interaction of solute in the gaseous form with the polymer layer, the surface energy, γ_s, may be obtained. The surface energy describes interaction caused by dispersive forces or a combination of dispersive forces with H-bonding or with dipole–dipole forces. Fowkes[38] first reported this method of characterization and determined the surface energy of several components. Generally, the contribution of all forces can be expressed as the energy of adhesion, Equation (16)

$$\gamma_a = \gamma^d + \gamma^{sp} \tag{16}$$

where γ^d is the contribution of dispersive forces and γ^{sp} is the contribution of specific interaction forces such as H-bonding, dipole–dipole, acid base, and so on. From the gas chromatographic measurements, V_g^0 is determined by using Equation (17). V_g^0 relates to the equilibrium constant K between the adsorbed solute and the polymer surface:

$$V_g^0 = K_s A \tag{17}$$

where K_s is the surface partition coefficient and A is the total surface area of the polymer powder in the chromatographic column. Thermodynamically, the molar free energy of adsorption, ΔG_1^a, of solute on the polymer layer can be related to V_g^0 by Equation (18)

$$\Delta G_1^a = -RT \ln V_g^0 + C \tag{18}$$

where C is a constant depending on A. Equation (19) relates the energy of adhesion to the free energy of adsorption:

$$RT \ln V_g^0 + C = 2Na\sqrt{\gamma_s^d \gamma_i^d} \tag{19}$$

where γ_s^d and γ_i^d are the dispersive components of the solid surface and the interactive solute phase, respectively. N is Avogadro's number and a is the area of the adsorbed molecules (solutes). In IGC experiments, a series of interactive solutes such as alkanes can be injected into the chromatographic column in order to determine the dispersive surface energy, γ_s^d. A plot of ΔG_1^a or $RT \ln V_g^0$ versus the number of carbons in the alkane series is linear and the slope of the straight lines will account for the incremental contribution of ΔG_1^a.

The molar enthalpy of adsorption can also be calculated from ΔG_1^a data directly, Equation (20)

$$\Delta H_1^a = -T^2 \frac{\delta}{\delta T} \frac{\Delta G_1^a}{T} \tag{20}$$

Combining Equations (19) and (20), the dispersive surface energy is, Equation (21)

$$\gamma_s^d = \frac{1}{4_{\gamma_{CH_2}}} \left(\frac{\Delta G_a^{CH_2}}{Na_{CH_2}} \right)^2 \tag{21}$$

where γ_{CH_2} is the surface energy of a hydrocarbon consisting only of *n*-alkanes, and a_{CH_2} is the area of one $-CH_2-$ group. Equation (21) usually tests the IGC method for obtaining the dispersive surface energy of polymers.

6 CRYSTALLINITY

The crystallinity of a polymer can be studied using X-ray, infrared (IR), calorimetry, density, and other techniques, but these methods require knowledge of the properties of a 100% crystalline polymer and involve the assumption that the variation of the chosen property is linear with the degree of crystallinity. The gas chromatographic technique only assumes nonpenetration of the probe into the crystal phase below T_m.

6.1 Degree of Crystallinity

Measurements of polymer crystallinity can be readily made using the molecular probe apparatus described in section 13. The polymer can be coated from solution onto a suitable support such as Chromosorb W. However, with many polymers the thermal history of the coating process may alter the crystallinity of the sample. In this case, cryogenic grinding of the polymer, followed by screening

to remove particles larger than about 0.1 mm, will give particles which can be blended mechanically with the support prior to packing the chromatography column.

Figure 8 shows the IGC retention diagram for linear PE using decane as a probe, acquired using an automated molecular probe instrument. The column was heated at a rate of 0.50 °C min^{-1} through the melting transition while very small samples of decane ($<1.5 \times 10^{-6}$ mol) vapor and nitrogen were injected into the helium carrier gas at intervals of approximately 2.5 min. The 1-m column contained 0.16 g of polymer coated on 60/80 mesh glass beads.[39]

Above the crystalline melting point, the 32 data points are exactly linear with a standard deviation for the slope of less than 0.2%. This gives some confidence in a linear extrapolation to lower temperatures to give $\log V_g'$, where V_g' is the expected value of the retention volume of the completely amorphous polymer at each temperature in question. Thus the percent crystallinity at any temperature below T_m is given by Equation (22)

$$\%\text{Crystallinity} = 100\left(1 - \frac{V_g}{V_g'}\right) \qquad (22)$$

where V_g' is the value for the 100% amorphous polymer obtained by extrapolation.

Figure 9[39] shows a plot of percentage crystallinity against temperature, comparing GC results with DSC results. The latter were obtained by integrating the heat of fusion curve for a sample put through the same thermal cycle as the sample used in the GC. The latent heat of crystallization of perfectly crystallized PE was assumed to be 68.5 cal g^{-1}. The small differences in the two curves could be due to PE being a thin layer coated on glass when studied by DSC and being the bulk phase when studied by GC. Retention times were found to decrease as the polymer was cooled from melt to below T_m, a reflection of the decrease in the amount of amorphous material. Although the heating rate in the DSC experiments was more than twice that used in the IGC run, the close agreement between two entirely different experiments is remarkable, and is a strong confirmation of the validity of the assumptions involved in the use of both procedures.

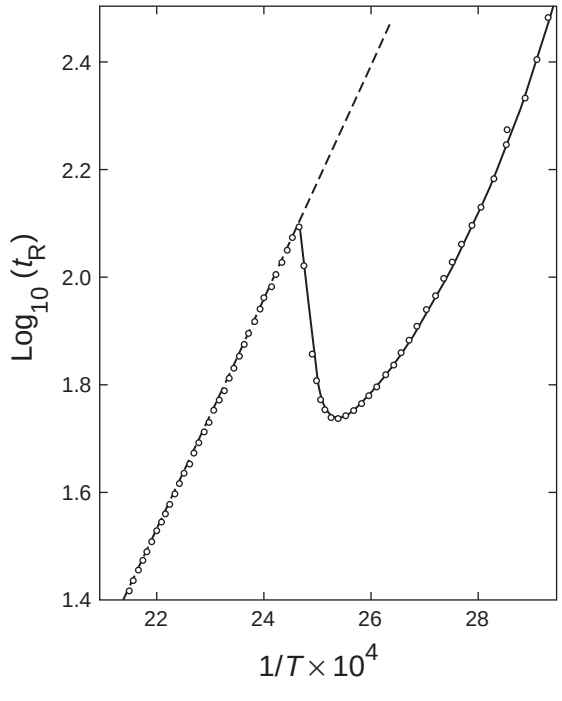

Figure 8 Retention diagram for decane on high-density PE.

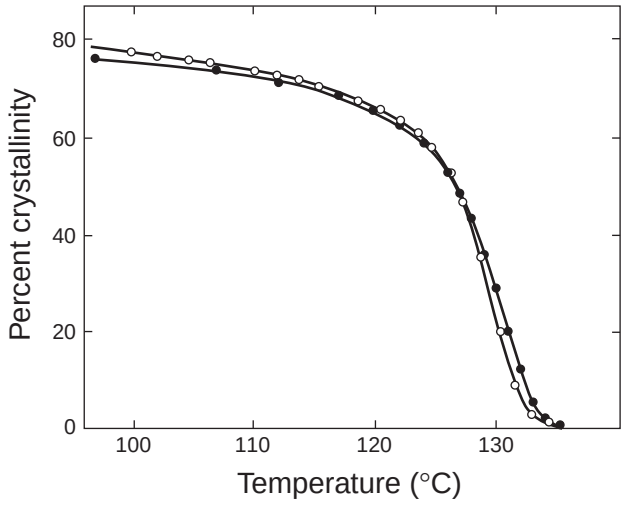

Figure 9 Percentage crystallinity versus temperature for linear PE using DSC and the molecular probe techniques: (●) DSC data, annealed at 1.25 °C min^{-1}, heated at 1.25 °C min^{-1}; (○) molecular probe data, annealed at 1.0 °C min^{-1}, heated at 0.5 °C min^{-1}.

A further improvement in crystallinity measurements by IGC is known as curvilinear extrapolation. In a study on semicrystalline PP, Braun and Guillet[40] found that the linear extrapolation gave a crystallinity of 36% at 50 °C, while using curvilinear extrapolation the curve leveled off near 66%, which was near the value of 64% obtained from density measurements. Besides giving a more accurate extension of the curved part of the retention diagram, the applicability of this approach is less dependent on the temperature range chosen than the linear method.

Experiments can also be done using mixed support with ground-up polymer; this latter method is preferred over using the solution-coated support, as the coating procedure changes the thermal history and thus the crystallinity of the sample. In a study on PE powder Hudec[41] found that a correction was necessary to account for contributions to V_g from adsorption on the support used; when performing the experiments using support alone, it was found that the retention times for

the probes were inversely dependent on temperature and represented 10% of the net retention time for the PE-coated support using the same probes. Tetradecane was most effective as a probe for particles of less than 0.07-mm diameter. Braun and Guillet[42] also found in their study on PE that for alkane probes C_8-C_{12} particles of diameter less than 0.1 mm were most effective. Equation (22) can be approximated by Equation (23)

$$\%\text{Crystallinity} = 100\ \ 1 - \frac{t_c}{t_a} \qquad (23)$$

where t_a is the retention time expected for the totally amorphous sample from the extrapolation and t_c is the retention time obtained from the melting or cooling curve.

6.2 Rates of Crystallization

In an analogous experiment, rates of crystallization for a polymer sample can be determined by melting the polymer coating and measuring its percent crystallinity as a function of time at constant temperatures below T_m. Typical plots on high-density PE are shown in Figure 10.[39]

The plots of percentage crystallinity against log time show some curvature, possibly indicating a complex kinetic scheme; the nucleation and growth process may also change with time and the amount of crystallinity. The fact that the sample is a thin polymer film coated on glass could cause this behavior, as the surface-to-volume ratio has been shown[43] to affect polymer crystallization rates. It is possible that the glass surface can initiate or stabilize nucleation on sites at a temperature where no crystallinity is detectable.

A number of early studies helped to define the scope of the IGC crystallinity procedures. In work done on

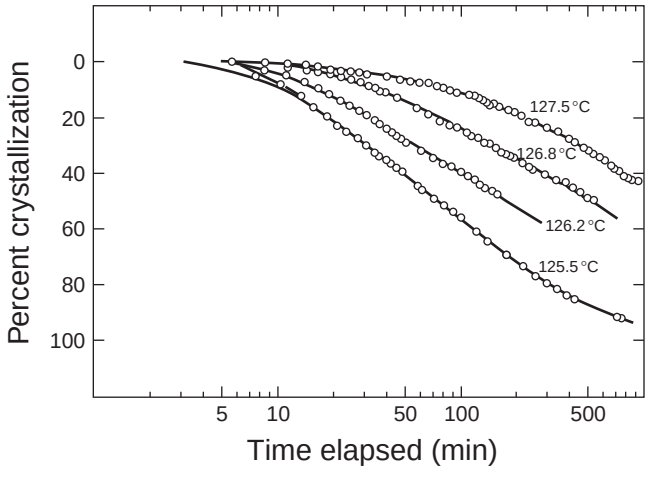

Figure 10 Percentage isothermal crystallization at the indicated temperatures as a function of time for high-density PE.

poly(ethylene glycol) adipate[44] Lipatov and Nesterov found that the degree of crystallinity of the polymer decreased with the film thickness. This was explained by hypothesizing the formation of an adsorption polymer layer at the solid interface. The proximity of this layer to the interface would result in smaller packing density, diminishing the mobility of the polymer molecules. This effect was found to be independent of the nature of the support.

In some systems there will be a surface contribution to the retention volume below T_m, leading to a low apparent crystallinity. Courval and Gray[33] studied the interactions of hydrogen-bonding probes on low-molecular-weight poly(ethylene oxide) (PEO). They found it necessary to extrapolate the retention volumes to infinite percentage loading of polymer, where surface sorption would be negligible.

Crystallinity measurements have been made on a variety of polymers, including PP,[32] PEO,[32] and linear and branched PE,[42] and copolymers of PE with vinyl acetate (VAc) and carbon monoxide (CO).[42]

In later work, Al-Saigh and Chen[45] proposed that below T_m a second adsorption term due to solute adsorption on the surface of small crystallites needs to be considered. Since the crystalline regions in semicrystalline polymers can be very small, their surface area may be significantly larger than either that of the support or the coated polymer.

7 DIFFUSION

The ability of a small molecule to diffuse through a solid polymer phase is of great importance when considering polymers for industrial applications. The diffusion of trace contaminants into a bulk polymer, and the rate at which stabilizers migrate from containers to the food inside, are two such applications where diffusion constants for small molecules at infinite dilution in the polymer system are needed. These conditions are reproduced closely in IGC experiments, making it the preferred method for obtaining diffusion constants under actual use conditions.

In theory, a sample injection that can be represented as a delta function should give a chromatogram which can also be described as a delta function. In fact, at best, the eluted peak will be broadened to give a Gaussian. Often the peak is skewed as well. Various factors combine to produce these effects, including instrumental imperfections such as dead volume in the detector, and these can be minimized by proper design.

Nonequilibrium conditions in the column and the existence of a nonlinear adsorption isotherm will result in skewed peaks. Other peak broadening factors are

diffusion of the solute through the stationary phase perpendicular (axial) to the direction of flow, and eddy diffusion, which is a result of the flow of moving gas around the particles of the column packing. Van Deemter et al.[46] related some of these effects to the plate height of the column, H, Equation (24)

$$H = A + \frac{B}{u} + Cu \tag{24}$$

where u is the linear velocity of the carrier gas in the column and A is a constant which accounts for broadening due to instrumental factors. B/u describes the diffusional spreading of the vapor as it flows through the column where, Equation (25)

$$B = 2wD_g \tag{25}$$

D_g is the diffusion coefficient of the vapor in the gas phase and w is the eddy diffusion constant (<1).

C is a constant related to the finite amount of time needed for the gas and stationary phases to reach equilibrium. Using a simple model, C can be expressed by Equation (26)

$$C = \frac{8}{\pi^2}\frac{d_f^2}{D_1}\frac{k}{(1+k)^2} \tag{26}$$

where d_f is the thickness of the stationary phase, D_1 is the diffusion coefficient of the vapor in the stationary phase and k the partition coefficient (the experimental retention time minus the retention time for an inert marker, i.e. methane).

7.1 Determination of Diffusion Constants

To find diffusion coefficients for solutes in the polymer phase, H values are calculated from the chromatographic peak shapes using Equation (27)

$$H = \frac{l}{5.54}\left(\frac{d}{t_r}\right)^2 \tag{27}$$

where l is the column length and d the measured peak width at half-height. By choosing experimental conditions carefully (minimizing dead volume and experimental response time) one factor can be made to predominate, namely diffusion in the stationary phase.

At high flow rate, B/u will approach zero, so a plot of H against u (Equation 24) should produce a linear section with slope C.[45] The stationary phase thickness can be found using Equation (28)

$$d_f = \frac{w}{\rho}\frac{3v}{r} \tag{28}$$

where w is the weight and ρ the density of the polymer on the column, and v is the volume occupied by the stationary phase, i.e. glass beads of average radius r (for support consisting of spherical beads). Figure 11 shows experimental data of H versus u for a number of probes in low-density PE and natural rubber[47] at various temperatures. In later work, measurements were made on the diffusion of antioxidants and other solvents in molten PE. The values of D and E_a, the activation energy for diffusion, are shown in Table 1.[48]

Giddings[49] worked out an expression for C under nonequilibrium conditions for a thin film distributed evenly over a solid surface, and found that the constant in Equation (26) changed from $8/\pi^2$ to 2/3. In addition, he divided C into two terms where C_g represents the gas phase mass transfer and C_s the stationary phase mass transfer. The expression for C (Equation 26) accounts only for C_s, but for a highly loaded system (>0.9%) C_g is only 3–5% of C; for a system of 0.2% loading C_g would be 15% of C and hence would have to be subtracted from C before accurate diffusion coefficients could be calculated.[50]

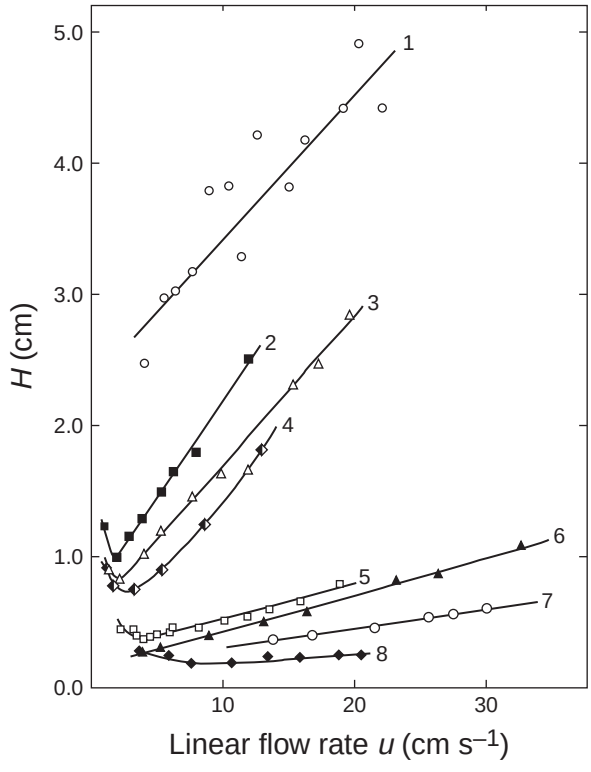

Figure 11 Van Deemter curves for various columns: (1) benzene/low-density PE at 27 °C; (2) benzene/natural rubber at 27 °C; (3) benzene/low-density PE at 25 °C; (4) benzene/natural rubber at 25 °C; (5) nonane/low-density PE at 50 °C; (6) decane/natural rubber at 40 °C; (7) dodecane/natural rubber at 85 °C; (8) hexadecane/PE at 200 °C.

Table 1 Diffusion coefficients and activation energies for selected solutes in low-density PE

Solute	Temperature (°C)	Diffusion coefficient $D_\ell \times 10^8$ ($cm^2\ s^{-1}$)	Activation energy, E_a ($kcal\ mol^{-1}$)
n-Dodecane	121.8	9.15	8.9 ± 1.3
	125.0	10.5	
	132.0	12.0	
Tetralin	115.0	9.44	10.2 ± 0.5
	125.0	13.9	
	132.0	16.1	
	140.0	21.3	
cis-Decalin	121.8	9.12	9.0 ± 1.0
	125.0	10.6	
	132.0	12.3	
trans-Decalin	121.8	10.6	9.2 ± 0.7
	125.0	11.9	
	132.0	13.5	
	140.0	18.1	
BHA	136.8	7.38	10.0 ± 0.2
	143.3	9.20	
	158.9	13.8	
	169.2	18.3	
BHT	136.8	4.70	9.4 ± 0.2
	143.3	5.83	
	158.9	8.51	
	169.2	11.1	

BHA = Butylated hydroxyaniline, BHT = butylated hydroxytoluene.

Giddings revised Equation (26) to Equation (29)

$$C = \frac{2}{3}\frac{d_1^2}{D_1}\frac{k}{(1+k)^2} \tag{29}$$

Knowing C, d_f and k, the diffusion coefficient D_1 can be calculated. According to this model a decrease in D_1 should result in greater peak spread but no change in the peak maximum.

The temperature dependence of D can be determined by monitoring peak shape changes with T. Gray and Guillet[47] found that for tetradecane in PE melt, D_1 increased only ten-fold between 125 and 170 °C and hence there was little change in peak shape or maximum. For the same probe in PS melt,[51] calculated D_1 values dropped sharply as T approached T_g. The peak distortion as T approaches T_g is related to the temperature dependence of D_1. At 130 °C no peak is detectable, and the solute is eluted with the carrier gas. This total absence of peak is not uncommon, and occurs because the thick polymer layer gives poor column performance (peak spreading is large compared to retention volume). Work has also been done on the effect of diffusion on the shape of chromatographic peaks in the region of the glass transition.[51]

There are advantages to finding diffusion coefficients of solutes in polymers using IGC. Polymers, being viscous, form a stable even coating on the support and rates of diffusion through polymers are slower than through other stationary phases, so the contribution to peak spreading from diffusion is more important.

In traditional analytical GC, the peak broadening is directly related to the resolving power of the columns and as such has received extensive theoretical attention.[52–55] There are two major factors that contribute to peak broadening: diffusion of the injected solute in the carrier gas and diffusion of the solute in the stationary phase. The first factor is known as the gas-phase mutual diffusion coefficient, D_g, and the second factor is related to the liquid-phase mutual diffusion coefficient, D_L.

In IGC experiments, where a polymer is a stationary phase, D_L is the polymer–solute diffusion coefficient. A large body of data is accumulated using PIB coated on Chromosorb W as the stationary phase and linear alkanes starting from methane to undecane as solutes.[56] Munk et al.[52] varied the temperature, flow rate of the carrier gas, and column loading systematically. They were able to design a method of evaluating the peak broadening data from which both diffusion coefficients, D_g and D_L, were obtained as follows, Equation (30)

$$H = A + \frac{2\gamma D_g^\infty}{u_0} + R_f(1-R_f)u_0\left(\frac{J\gamma_L}{D_L} + \frac{\gamma_g}{D_g^0}\right. \tag{30}$$

where H is the average height equivalent to one theoretical plate which is the ratio of the length of the column to the number of theoretical plates (N), u_0 is the carrier gas flow rate measured at column outlet, J is the pressure correction factor defined in Equation (1), R_f is the retardation factor, γ, γ_L and γ_g are the tortuosities of the structural factor (L and g refer to the liquid and gas phase, respectively), A is a constant factor that is independent of the flow rate, and D_g and D_L are the diffusion coefficients of the solute in the gas and liquid phase, respectively. Thus the average height equivalent to one theoretical plate of the column depends on the solute-related parameters D_g and D_L and on the column-dependent parameters A, γ, γ_g, and γ_L. The latter parameters are the same for all solutes, but they vary from column to column. If a set of columns is prepared in a similar way, one would expect the column parameters to have similar values. Therefore, for a set of columns differing only in column loading, one would expect that the parameters A, γ, γ_g, and γ_L that depend only on the column packing will be similar for all columns. In this case γ_L will be strongly dependent on the polymer loading. The R_f value can be calculated as follows, Equation (31)

$$R_f = (V_0 - V_d)(V_r - V_d) \tag{31}$$

where V_0 is the empty volume of the chromatographic column equal to the retention time of the marker, V_r

is the retention time of the solute, and V_d is the dead volume in the column measured as the retention time of the marker.

In order to calculate both diffusion coefficients, it is necessary to correct Equation (30) for gas compressibility, Equation (32)

$$\frac{(H-A)u_0}{f} = 2\gamma D_g^0 + \left(\frac{\gamma_L}{D_L} + \frac{f\gamma_g}{D_g^0} \quad R_f(1-R_f)u_0^2\frac{J}{f} \tag{32}$$

where f in most circumstances is within a few percent of unity and is a correction for gas compressibility. It is defined as follows, Equation (33)

$$f = \left(\frac{P_i}{P_o}\right)^2 + 1 \quad \frac{J^2}{2} \tag{33}$$

where P_i and P_o are inlet and outlet pressures on the column. A plot of $[(\overline{H}-A)u_0/f]$ versus $[R_f(1-R_f)u_0^2J]$ should yield a straight line with a slope of γ_g/D_g. Plots constructed in this manner using uncoated columns yielded negligible slopes within the experimental error. Considering γ_g is negligible, then plots of $[(\overline{H}-A)u_0/f]$ versus $[R_f(1-R_f)u_0^2J/f]$ using coated columns should be linear and should yield $2\gamma D_g$ as the intercept and γ_L/D_L as the slope.

8 THERMODYNAMIC STUDIES

8.1 Activity Coefficients, Heats of Mixing and Heats of Solution

In conventional GC the infinite dilution activity coefficient is expressed in terms of, among other variables, the molecular weight of the stationary phase, M_2. This approach has been used to find M_2, but in IGC the quantity of interest is γ_1^∞. When the stationary phase is a polymer, M_2 becomes difficult to quantify, especially in the case of a polydisperse sample. In addition, γ_1^∞ is defined with respect to the mole fraction of component 1 in a binary solution of small molecules. In IGC the two components are very different, both in molecular size and concentrations present. These problems are rationalized by the use of the weight fraction infinite dilution activity coefficient. Patterson et al.[57] defined the infinite dilution activity coefficient with respect to the weight fraction of component 1 (w_1), giving, Equation (34)

$$\ln\left(\frac{a_1}{w_1}\right)^\infty = \ln\left(\frac{273.16R}{p_1^0V_gM_1} - \frac{p_1^0(B_{11}-V_1)}{RT}\right) \tag{34}$$

In this manner the activity coefficient of the probe in the polymer can be found using experimental data and pure component properties alone. Using the weight fraction activity coefficient, partial molar heats of mixing (ΔH_1^∞) can be calculated (Equation 5), equivalent to ΔH_{mix}^e. DiPaola-Baranyi et al.[58] tabulated these for alkane and butanol probes at infinite dilution in PS, poly(ethylene–carbon monoxide), PE, poly(ethylene–vinyl acetate) (P(E-VAc)), PVC, poly(1-butene), poly(methylacrylate) and poly(*N*-isopropyl acrylamide). Equation (34) does not include the correction for gas phase nonideality, but using an inert carrier gas and low column pressures, the temperature coefficient of the correction should be small. Polymer–solvent systems showed small activity coefficients $(a_1/w_1)^\infty$ and excess heats of mixing (ΔH_1^∞). Nonsolvent systems had much larger activities and heats of mixing.

Heats of solution can be calculated using Equation (35)

$$\frac{-\Delta H_s}{R} = \partial \ln \frac{V_g^0}{\partial}\left(\frac{1}{T}\right. \tag{35}$$

In a study on ethylene–vinylacetate copolymers,[59] Dincer and Bonner divided ΔH_s into contributions from dispersion, dipole and induced dipole interactions and from specific interactions such as hydrogen bonding and charge-transfer complexes, Equation (36)

$$-\Delta H_s = f(\alpha, \mu) + X \tag{36}$$

where $f(\alpha, \mu)$ is a function of the probe's polarizability (α) and dipole moment (μ). A plot was made of $-\Delta H_s$ against α for alkanes and aromatics. These points were fitted to quadratic equations, and the difference between the fitted and experimental values was interpreted as the contribution due to permanent dipole and special interactions. This was taken to be quantitatively equivalent to the enthalpy of hydrogen bonding.[60] Plots were also done of $-\Delta H_s$ against μ; the form of $f(\alpha, \mu)$ was deduced from the two plots.

The free energy of mixing, ΔG_m, can be expressed using volume fractions, Equation (37)

$$\Delta G_m = kT[N_1 \ln \phi_1 + N_2 \ln \phi_2 + \chi\phi_1\phi_2(N_1 + rN_2)] \tag{37}$$

Mixing is unfavorable for $\Delta G_m \leq 0$, therefore, Equation (38)

$$\chi_{critical} \leq \frac{1}{2} \quad 1 + \left(\frac{1}{\sqrt{r}}\right)^2 \tag{38}$$

For large molecular weights r is much larger than 1, giving $\chi_{critical}$ as approximately 0.5. Since χ is a free energy parameter, it can be separated into an enthalpic (χ_H) and an entropic (χ_S) component, Equation (39)

$$\chi = \chi_H + \chi_S \tag{39}$$

χ_H can be calculated from pure component properties using the Hildebrand–Scatchard regular solution theory for nonpolar substances,[61] Equation (40)

$$\chi_H = \frac{V_1}{RT}(\delta_1 - \delta_2)^2 \tag{40}$$

where the δ values are solubility parameters, Equation (41)

$$\delta = \left(\frac{\Delta E_{vap}}{V}\right)^{1/2} \tag{41}$$

The quantity on the right is known as the cohesive energy density and is an indication of the strength of intermolecular forces in the pure component.

Experimental studies[62–64] have shown χ_S to be of the order of 0.35 and it is sometimes left as a constant.

χ can be related to the weight fraction activity coefficient. For $\phi_2 \to 1$, Equation (42)

$$\ln\left(\frac{a_1}{w_1}\right)^{\infty} = \ln\left(\frac{v_1}{v_2}\right) + 1 - \frac{V_1}{(\overline{M}_2)_n v_2} + \chi \tag{42}$$

where v_1 and v_2 are the specific volumes of the probe and polymer, respectively.

In polymers $(M_2)_n$, the number average molecular weight is usually large enough so that the second term in the last parentheses is very small. Substituting for $\ln(a_1/w_1)^{\infty}$ gives Equation (43)

$$\chi = \ln\left(\frac{273.16Rv_2}{p_1^0 V_g V_1}\right) - \frac{p_1^0(B_{11} - V_1)}{RT} - 1 \tag{43}$$

Using Equation (43), χ can be determined directly from measurements of V_g under appropriate conditions.

Most experimental work uses the original Flory–Huggins and/or the modified approach using core volumes and segments, giving χ and/or χ^*. χ is sometimes taken to have a temperature dependence of the form of Equation (44)

$$\chi = \alpha + \frac{\beta}{T} \tag{44}$$

From the original Flory–Huggins definition, a plot of χ versus $1/T$ should give a straight line. In fact, this inverse trend is only true for part of the plot. The $\chi - T$ curve is more parabolic in shape and care should be used in doing any extrapolation of χ data.

According to Flory–Huggins theory, χ should show a composition dependence but studies[65] on PS and high-density PE with benzene and decane showed no such dependence. There is an extensive literature on this subject including studies on PIB,[66–68] PDMS,[69–72] PS, amorphous PE and PP,[73] *cis*-polyisoprene,[74] linear and branched PE,[75] poly(methyl sulfoxide),[70] and poly(ethylene glycol).[76]

8.2 Finite Concentration Gas Chromatography

Work has also been done at finite concentration using GC. It involves the use of a carrier gas with a significant concentration of probe molecule in the vapor phase. A series of articles by Condor and Purnell explain the theory and experimental set-up for these studies.[77–80] Several elution techniques can be used; all start from Equation (1) and one technique is described below.

Brockmeier et al.[73] studied PS, amorphous PE and atactic PP at both infinite and finite dilution. The latter experiments were carried out using the technique of elution-on-a-plateau. The column is equilibrated with a stream of solute gas at steady concentration c. A sample concentration infinitesimally greater or less than c is injected, yielding a Gaussian peak. The first quantity of interest is the distribution isotherm which (at pressure P) is given by Equation (45)

$$q(P) = \frac{j}{w_2}\int_0^{y_0}\left(\frac{V_g}{1} - \psi\right) dc \tag{45}$$

where y_0 is the mole fraction of solute in the gas phase at pressure P_0 in a zone of concentration c, and j is a compressibility correction which approaches J_3^2 as values were always positive and decreased with increasing temperature in a complicated manner. There was a slight composition dependence, with χ'_{23} becoming lower for the PS-rich copolymers. Patterson and Robard[81] reviewed the theory and its application to mixed stationary phases. Other studies have been carried out by Olabisi,[82] Galin and Rupprecht,[83] and others.[84,85] An apparatus for performing these studies is shown in Figure 15.

8.3 Henry's Law Constants (Solubilities)

Henry's Law describes a two-component system where one of the components (1) is in very low concentration relative to the other (2). Under this condition the solubility of 1 in 2 (i.e. X_1) can be related to the partial pressure of 1 in the vapor (p_1) by Equation (46)

$$H_1 = \frac{p_1}{X_1} \tag{46}$$

where H_1 is Henry's law constant. Liu and Prausnitz[86] expressed this in terms of the fugacity of component 1, Equation (47)

$$H_1 = \lim_{x \to 0}\left(\frac{f_1}{X_1}\right) \tag{47}$$

Assuming ideal behavior of the gas in the column, and that f_1 is proportional to the weight fraction of solute in the polymer phase, then under equilibrium conditions,

Equation (48)

$$H_1 = \frac{RT}{V_g M_1} \tag{48}$$

Infinite dilution weight fraction Henry's constants have been found for *n*-alkane solutes in long-chain alkanes[87,88] ($C_{28}H_{58}$–$C_{36}H_{74}$), for volatile solutes in PVAc,[89] for liquid low-density PE at high pressures[90] and for PE[86,91] and P(E-VAc)[86] at ambient pressures.

8.4 Solubility Parameters

Solubility parameters provide another way of quantifying the thermodynamic information obtained from an IGC experiment. Previous methods for finding δ_2 values have involved swelling or viscosity measurements, graphical methods and the group-additive method outlined by Small.[92] The IGC approach uses pure component properties and experimental data only, and gives δ_2^∞ values for solutes at infinite dilution in the solid polymer phase.

DiPaola-Baranyi and Guillet[93] related the solubility parameter for probe (δ_1) and polymer (δ_2) to several thermodynamic variables. From regular solution theory, assuming zero volume change on mixing, Equation (49)

$$\Delta\overline{H}_1^\infty = V_1(\delta_1 - \delta_2)^2 \tag{49}$$

At constant pressure this can be rewritten as Equation (50)

$$\Delta\overline{G}_1^\infty = V_1(\delta_1 - \delta_2)^2 \tag{50}$$

or, combining regular solution theory with the Flory–Huggins theory, as Equation (51)

$$\chi = \frac{V_1}{RT}(\delta_1 - \delta_2)^2\chi_s \tag{51}$$

Working with hydrocarbon probes in PMMA and PS, plots were made of $(\delta_1^2/RT - A/V_1)$ against δ_1 for $A = \Delta H_1^\infty, \Delta G_1^\infty, \chi$, respectively. The plot using χ showed the least scatter of points and gave best values of δ_2^∞, and this was the method chosen for data analysis.

In such a plot the slope would be $+2\delta_2^\infty/RT$ and the intercept would be $-\delta_2^{\infty 2}/RT - \chi_s/V_1$. A typical example is shown in Figure 12.[94] Errors quoted for the slope or intercept values are the standard deviation for slope/intercept in a least-squares plot. By definition[95] the error in the intercept would be greater than that in the slope by a factor of $[(\ {}_N\delta_1^2)\ /N]^{1/2}$, where N is the number of data points.

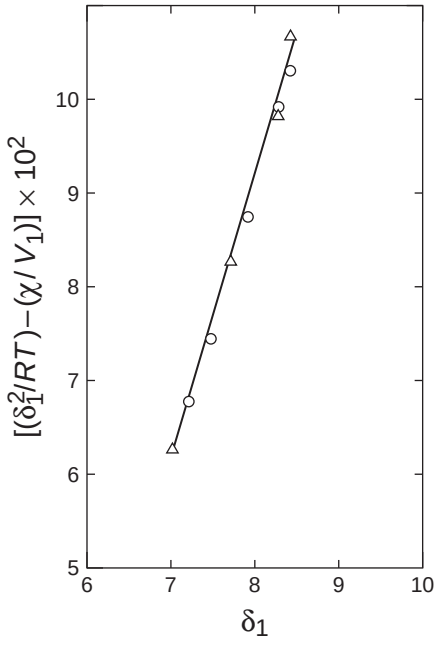

Figure 12 Infinite dilution solubility parameters for poly(butadiene–acrylonitrile) at 75 °C. (Δ) Polar probes; (○) nonpolar probes. δ_2'(slope) = 10.00 ± 0.29.

Infinite dilution solubility parameters have been found by this method for PS and PMMA,[93] PVAc,[96] *cis*-polyisoprene, amorphous PP and an ethylene–propylene copolymer (40% w/w ethylene),[64] and also for polychloroprene and poly(butadiene–acrylonitrile), 34% w/w acrylonitrile, *cis*-1,4 polybutadiene and P(E-VAc) (40% w/w VAc).[94] Infinite dilution solubility parameters determined in this way for a number of polymers are shown in Table 2.

Experimental δ_2' values obtained by IGC are usually at temperatures greater than 25 °C, which is the temperature that literature values are quoted for. To extrapolate the experimental χ results down to 25 °C, Equation (44) is

Table 2 Infinite dilution solubility parameters

Polymer	Temperature (°C)	δ_2^∞	Ref.
PS	193	7.6 ± 0.2	93
PMMA	100	8.5 ± 0.3	93
PVAc	135	8.5 ± 0.4	96
Ethylene–propylene rubber	73	7.18 ± 0.11	97
Polyisoprene	30	7.96 ± 0.10	97
PP	30	7.67 ± 0.16	97
Polychloroprene	75	8.83 ± 0.22	94
Poly(butadiene–acrylonitrile)	75	10.00 ± 0.29	94
P(E-VAc)	75	8.26 ± 0.17	94
cis-1,4-Polybutadiene	75	7.90 ± 0.14	94

sometimes used.

$$\chi = \alpha + \frac{\beta}{T} \tag{44}$$

The validity of this procedure can be tested by measuring χ values for a range of temperatures and constructing a χ versus $1/T$ plot. If the temperature range of interest is in the downward curved section of the plot, Equation (44) can be used with care. When T_g of the polymer is 50 °C or more below room temperature, an extrapolation is not necessary and the easiest method is to perform an experiment at 25 °C.

Note that the experimental δ_2^∞ values represent infinite dilution solubility parameters, where other values quoted arise from finite concentration solution studies. δ_2^∞ values usually fall well within the literature range for a polymer, but the exact relationship between δ and δ^∞ is still not certain.

For industrial application δ is sometimes divided into components representing different contributions to the net interaction. Hansen[97] proposed that the energy of vaporization could be divided, Equation (52)

$$\Delta E^{vap} = \Delta E_d^{vap} + \Delta E_p^{vap} + \Delta E_h^{vap} \tag{52}$$

where ΔE_d^{vap} is from polar (dipole–dipole and dipole–induced dipole) forces and ΔE_h^{vap} is from hydrogen bonding or charge-transfer complexes (specific interactions). This equation leads to Equation (53)

$$\delta^2 = \delta_d^2 + \delta_p^2 + \delta_h^2 \tag{53}$$

The nonpolar part, δ_d, can be approximated using Small's method and this is subtracted from an experimental δ to leave the polar and hydrogen bonding contributions.[62,98] There is some question about how well this division represents the forces experienced by a probe in the polymer phase. Olabisi[82] concluded that the contribution of various forces could not be equally divided, and that a different unknown proportion is involved for each polymer–probe interaction.

Lipson and Guillet[94] calculated δ_2^∞ values for polybutadiene, P(E-VAc), polychloroprene and poly(butadiene–acrylonitrile) using polar, then nonpolar, then a combination of probes, and found no difference within experimental error. A factor affecting nonpolar interactions such as α, the polarizability, is also a factor in dipole-induced dipole interactions. Dipole moments (used in calculating energies of dipole–dipole interactions) will also be a partial indication of a molecule's ability to hydrogen bond. The division of δ into nonpolar/polar/specific interactions is controversial. In addition, the magnitude of the energy contribution arising out of these interactions is questionable, since IGC experiments often involve a majority of hydrocarbon probes, the most polar usually being alcohols or ketones. The net result is an experiment where a smaller fraction of the interactions are specific or polar. For δ_2^∞ calculations the division of δ may not be as applicable as for the more qualitative industrial applications.

8.5 Hydrogen Bonding

DiPaola-Baranyi et al.[76] used IGC to investigate the thermodynamics of hydrogen bonding from heats of mixing. In a binary solution of small molecules, contributions to hydrogen bonding arise from self-association in the pure solute or solvent, and from hydrogen bonding between the two components. The correction for self-association is found by measuring the heat of mixing (ΔH_1^∞) for the compound in question at infinite dilution in a nonpolar solvent, such as hexane. When the solvent in question is a polymer, amorphous PE or PP could be used.

Alternatively, heats of solution (ΔH_s) can be used in these studies, since, Equation (54)

$$\Delta H_1^\infty = \Delta H_{vap} + \Delta H_s \tag{54}$$

ΔH_{vap} is a function of the solute and experimental temperature and not the polymer, so differences in ΔH_s for a given solute at fixed temperature will reflect changes in ΔH_1^∞.

In studying the hydrogen bonding between a small molecule (solute and polymer solvent) using ΔH_s, two corrections are needed: one accounts for the nonhydrogen-bonding interactions (dipole–dipole, dispersion) that the probe is capable of, and is estimated by subtracting ΔH_s^{model} for a model, nonhydrogen-bonding probe in the polymer from ΔH_s^{probe}. The other correction accounts for the nonhydrogen-bonding interactions that the polymer is capable of, and is estimated by measuring ΔH_s values in the polymer and then in a reference polymer (such as low-density PE). The net estimate of the enthalpy change arising from hydrogen bonding between the probe and polymer is, Equation (55)

$$\Delta H_f = (\Delta H_s^{probe} - \Delta H_s^{model})_{polymer} - (\Delta H_s^{probe} - \Delta H_s^{model})_{reference\ polymer} \tag{55}$$

The experimental data needed would be ΔH_s for the probe in the polymer, then in the reference polymer, and ΔH_s for the model in the polymer, then in the reference polymer.

Some results from this procedure[76] are given in Table 3 for stationary phases P(E-VAc) and PVC using probes chosen for their hydrogen-bonding ability. Average bond energies were found to correlate well with literature values.

Table 3 Estimates of hydrogen-bonding contribution to the heat of solution[a] at 135 °C

Hydrogen donor	Model	ΔH_f (kcal mol^{-1})	
		PVAc	P(E-VAc)
Chloroform	Carbon tetrachloride	-1.3 ± 0.2	-1.0 ± 0.3
Butyl alcohol	Butyl chloride	-1.1 ± 0.3	-0.3 ± 0.4
Cyclohexanol	Cyclohexyl chloride	-2.3 ± 0.3	-1.0 ± 0.2
Phenol	Chlorobenzene	–	-1.7 ± 0.3

[a] Ref. 76.

9 POLYMER BLENDS

In polymer blends, the key term in the miscibility of a polymer–polymer pair is the free energy of mixing, ΔG_m, Equation (56)

$$\Delta G_m = \Delta H_m - T\Delta S_m \tag{56}$$

where ΔS_m is the combinatorial entropy of mixing and ΔH_m is the molar heat of mixing. Flory[99] attributed the combinatorial entropy of mixing to the mixing of the segments on a lattice of fixed volume. Because entropy depends on volume, an additional contribution to the entropy of mixing may be necessary in Equation (56). Sanchez and Lacombe[100] developed a theory to allow for this effect by considering that all mixtures obey the equation-of-state when appropriate reducing parameters, such as pressure and temperature, are used for volume. Other equation-of-state theories of mixtures yield a combinatorial entropy of mixing similar to that of Flory. However, the combinatorial entropy becomes very small as the molecular weight of the polymer becomes high. Therefore, in the case of high-molecular-weight polymers, only the value of ΔH_m determines the miscibility of the polymer pairs. Flory and Huggins first introduced the volume fraction term ϕ_i in their theory which described polymer solutions with a reasonable success.[100] The free energy of mixing as described by the Flory–Huggins theory is, Equation (57)

$$\Delta G_{mix} = RT(n_1 \ln \phi_1 + n_2 \ln \phi_2 + n_1\phi_2\chi_{12}) \tag{57}$$

where n_i is the number of moles of the ith component, RT has its usual meaning, and χ_{12} is a parameter which is inversely proportional to absolute temperature. χ_{12} is the same parameter introduced in Equation (37) where it is an enthalpic contact parameter. The two logarithmic terms represent the (combinatory) entropy of mixing. A second parameter that can be used as an indicator for mixing is B_{12} as a binary interaction energy parameter or excess cohesive density of the polymer–solute system. B_{12} is related to the parameter χ_{12}, as in Equation (58)

$$\frac{B_{12}}{RT} = \frac{\chi_{12}}{V_1} \tag{58}$$

Subscripts 1, 2, and 3 refer to the solute, polymer 2 and polymer 3, respectively. While the sign of the combinatorial entropy always favors mixing, it is clear that its magnitude is greatly diminished as molar volumes become very large. Since ΔS_m is small, exothermic heat of mixing is a requirement for miscibility in high-molecular-weight polymer blends. (As a result of the exothermicity principle, a number of miscible polymer blends have been identified.)[101,102] Utilizing the specific retention volume, V_g^0, the polymer–polymer interaction coefficient χ_{23} can be derived. When a polymer pair is used as a stationary (liquid) phase in a chromatographic column, subscripts 2 and 3 will be used to represent polymers 2 and 3, respectively. Subscript 1 refers to the test solute. The interaction between the two polymers is expressed in terms of the free energy of mixing ΔH_{mix} which has the same form as Equation (5), only the subscripts change to 2 and 3. The first two (entropic) terms in this equation are negligible for polymer blends. Thus, for the polymer blend to be miscible (ΔG_{mix} being negative), χ_{23} must be negative. When considering the IGC of polymer blends, the free energy of mixing must be written for a three-component system. It is usually expressed as in Equation (59)

$$\Delta G_{mix} = RT[n_1 \ln \phi_1 + n_2 \ln \phi_2 + n_3 \ln \phi_3 + n_1\phi_2\chi_{12} + n_1\phi_3\chi_{13} + n_2\phi_3\chi_{23}] \tag{59}$$

Equation (59) is generally satisfactory for nonpolar mixtures. The interaction coefficients χ are largely independent of the blend composition. The derivatives of the ΔG_{mix} are, Equation (60)

$$\left(\frac{\delta \Delta G_{mix}}{\delta n_1}\right)_{n=1,p,T} = RT\left(\ln \phi_1 + 1 - \frac{V_1}{V_2}\phi_2 - \frac{V_1}{V_3}\phi_3 + \phi_2\chi_{12} + \phi_3\chi_{13} - \frac{V_1}{V_2}\phi_2\phi_3\chi_{23}\right. \tag{60}$$

In Equation (60), v_2 should be replaced by $(w_2v_2 + w_3v_3)$ where w_2 and w_3 are the weight fractions and v_2 and v_3 are the corresponding specific volumes of the two polymers in the blend. Thus, one can easily derive Equation (61):

$$\ln \frac{273.15R(w_2v_2 + w_3v_3)}{V_g^0 V_1 P_1^0} - 1 - \frac{B_{11} - V_1}{RT}P_1^0 = \phi_2\left(\chi_{12} - \frac{V_1}{M_2v_2}\right. + \phi_3\left(\chi_{13} - \frac{V_1}{M_3V_3}\right. - \frac{V_1}{V_2}\phi_2\phi_3\chi_{23} \tag{61}$$

Usually the parameter χ'_{23} is introduced as in Equation (62)

$$\chi'_{23} = \frac{V_1}{V_2}\chi_{23} \tag{62}$$

The relation for χ_{12} is given by Equation (63)

$$\chi_{12} = \ln\frac{273.15Rv_2}{V_g^0 V_1 P_1^0} - 1 + \frac{V_1}{M_2 v_2} - \frac{B_{11} - V_1}{RT}P_1^0 \tag{63}$$

Comparison of Equations (61) and (63) suggests that to obtain χ_{23} for blends, χ_{12} and χ_{13} should also be known. Thus, three columns are usually prepared, two from homopolymers and a third from a blend. The three columns should be studied under identical conditions for column temperature, carrier gas flow rate, inlet pressure, and the same solutes should be used. This keeps all auxiliary parameters (P_1^0, T, M_2, M_3, V_1, v_2, v_3, and B_{11}) identical for the three experiments, and a combination of Equation (61) (taken twice for two homopolymers) and Equation (63) for the blend will yield Equation (64)

$$\chi_{23} = \frac{\ln\dfrac{V^0_{g,\text{blend}}}{W_2v_2 + W_3v_3} - \phi_2\ln\dfrac{V^0_{g,2}}{v_2} - \phi_3\ln\dfrac{V^0_{g,3}}{v_3}}{\phi_2\phi_3} \tag{64}$$

Here, the second subscript of V_g^0 identifies the nature of the column. From Equation (64), χ'_{23} may be calculated even for solutes for which the parameters P_1^0, B_{11}, and V_1 are not known or are known with insufficient accuracy. χ_{23} can also be related to B_{23}, Equation (65)

$$B_{23} = RT\frac{\chi'_{23}}{V_1} \tag{65}$$

B_{23} determines the miscibility of the blend with a negative value being necessary for mixing. Munk et al.[103] introduced this thermodynamic parameter for the evaluation of the miscibility of blends. B_{23} may also be evaluated directly from the experimental retention volumes and weight fractions w_i by combining Equations (64) and Equation (65)

$$B_{23} = \left(\frac{RT}{V_1\phi_2\phi_3}\right)\left(\ln\frac{V^0_{g,\text{blend}}}{W_2v_2 + W_3v_3} - \phi_2\ln\frac{V^0_{g,2}}{v_2} - \phi_3\ln\frac{V^0_{g,3}}{v_3}\right) \tag{66}$$

Munk et al.[103] showed that B_{23} values evaluated from experiments depended on the chemical nature of the solute, in contradiction to theory. To correct for this effect, a phenomenological relation suggested by Pouchley et al.[104] and developed by Al-Saigh and Munk[105] was introduced.

For a complex blend involving a semicrystalline diluent, it is possible to obtain the polymer–polymer interaction coefficient χ_{23} and the interaction energy parameter B_{23} experimentally in a different way than explained earlier. This technique involves measuring the melting point depression of a polymer mixture (blend) containing a semicrystalline polymer to quantify the heat of mixing of the polymer pair.

When crystals of the semicrystalline polymer are in equilibrium with the amorphous part of the semicrystalline polymer and the amorphous counter-polymer, the melting point of the semicrystalline polymer is lower than when the equilibrium is with the amorphous part of the semicrystalline polymer only. This is known as a melting point depression, resulting in mixing amorphous polymer with a semicrystalline polymer. In most cases, the amount the melting point is depressed depends on the weight fraction of the diluent polymer. The melting point depression can be used as an indicator for polymer blend miscibility. Most of the work performed thus far using melting point depression is based on calorimetry, and there have only been a few studies using IGC.[45] The thermodynamic analysis of polymer–polymer mixtures using melting point depression can be based on Flory–Huggins theory.[99,106] Information can only be obtained at a temperature close to the pure semicrystalline polymer melting point. However, a number of factors influence the measured melting point (perfection, size, and environment of crystals) which must be accounted for.

Equation (67) is an expression of the extent of the depression of the melting point of the semicrystalline polymer after mixing,

$$\frac{1}{T_m} - \frac{1}{T_m^0} = -R\left(\frac{\Delta H_{2u}}{V_{2u}}\right)\left(\frac{\rho_2\ln\phi_2}{M_2} + \left(\frac{\rho_2}{M_2} - \frac{\rho_3}{M_3}\right)v_3 + \left(\frac{B_{23}}{RT_m}\right)\phi_3^2\right) \tag{67}$$

where T_m and T_m^0 are the melting points of polymer 2 (semicrystalline) in the mixture and pure semicrystalline polymer, respectively. The quantity $\Delta H_{2u}/V_{2u}$ is the heat of fusion of the semicrystalline polymer per unit volume. The quantities ϕ_2, ϕ_3, ρ_2, ρ_3, M_2, and M_3 are the volume fractions, densities, and molecular weight of the semicrystalline (2) and diluent polymer (3), respectively, B_{23} is the polymer–polymer interaction energy related to the Flory–Huggins interaction parameter χ'_{23} as in Equation (68)

$$B_{23} = RT_m^0\frac{\chi_{23}}{V_3} \tag{68}$$

where V_3 is the molar volume of component 3 (the amorphous diluent polymer).

The first two terms in Equation (67) are the entropic contribution and the third term is the enthalpic contribution. If the molecular weights of the semicrystalline and the diluent polymers are high, then the entropic contribution represents less than 1 °C which would play only a minor role in the melting point depression. Therefore, the values of the first two terms can be neglected, giving Equation (69)

$$\frac{1}{T_m} - \frac{1}{T_m^0} = -\frac{V_{2u}}{\Delta H_{2u}} \frac{B_{23}}{T_m} \phi_3^2 \tag{69}$$

Rearrangement yields Equation (70)

$$\frac{T_m^0 - T_m}{T_m T_m^0} = -\frac{V_{2u}}{\Delta H_{2u}} \frac{B_{23}}{T_m} \phi_3^2 \tag{70}$$

which can be rewritten in the form of Equation (71)

$$\Delta T_m = T_m^0 - T_m = -T_m^0 \frac{V_{2u}}{\Delta H_{2u}} B_{23} \phi_3^2 \tag{71}$$

ΔT_m in Equation (71) is determined experimentally from IGC, and is used for calculating the polymer–polymer interaction energy B_{23} for semicrystalline–amorphous diluent blends.

9.1 Experimental Studies on Polymer Blends

Most of the work performed in this area has focused on polymer–polymer interaction parameters. This stemmed from the fact that knowledge of compatibility is most significant when studying polymer blends. Most measurements of χ_{23} are reported in the temperature range in which the two polymers are in a molten state; the solid state is avoided even though the possibility exists that the polymer pair can be compatible in a solid or semisolid state. Deshpande et al.[107] studied a blend of PDMS/*n*-tetracosane and di-*n*-octyl phthalate (DnOP)/*n*-tetracosane. Although these blends consist of low-molecular-weight polymers, their results indicated that IGC is useful in characterizing a mixture of low-molecular-weight polymer, and applicable to a ternary stationary phase. They calculated χ_{23} by using the Prigogine–Flory theory and showed that the interaction parameters depended on the chemical nature of the solute.

Olabisi,[82] applied IGC to a blend of several blends of poly(ε-caprolactone) and PVC at 120 °C. These experiments involved mainly low-molecular-weight oligomers ranging from 2000 to 40 000 g mol^{-1}.

Su and Patterson[108] studied a blend of PS ($M_n = 600$) with poly(vinylmethyl ether) ($M_w = 10\,000$). This mixture was found to be compatible. They avoided the difficulties associated with the high glass transition temperature of PS by selecting a PS oligomer. Polymer–polymer coefficient χ_{23} parameters were obtained in the range of −0.10 to 0.30. However, isopropyl alcohol, a poor solvent for both polymers, yielded values as high as 1.47, while chloroform yielded 0.65. However, when higher molecular weight PS was used, the blends were incompatible. The highest average value for χ_{23} was 0.42 when 30% of PS was used, and the lowest average value was 0.23 when 35% PS was used. DiPaola Baranyi et al.[109,110] and Al-Saigh and Munk[105] applied IGC to a variety of polymer blends. A blend of poly(methyl acrylate)–poly(epichlorohydrin) yielded measured values of χ_{23} that varied between −0.09 and 0.49 with an experimental error of about 0.01 (see section 5).

Values within each family of solutes were clustered together. Thus, the dependence of the interaction parameters on the chemical nature of solutes was real. The literature on the IGC of polymer blends is now extensive and beyond the scope of the present report. Much of this work is summarized in a recent review by Al-Saigh.[111]

9.2 Amorphous Plasticizer Blends

There are few studies that use IGC to characterize blends containing amorphous plasticizers. Su et al.[112] applied IGC to a blend of PVC plasticized by DnOP over a temperature range of 100–130 °C. The interaction parameters were strongly negative, indicating high compatibility. However, the interaction parameters became less negative and finally positive at 0.55 vol fraction of DnOP, suggesting a lower compatibility limit. The composition dependence of χ_{23} reflects a nonrandom solution of the solute in the stationary phase and/or nonrandom mixing of PVC/DnOP, particularly for DnOP in the limited compatibility range.

Corrective measures have recently been introduced to eliminate several instrumental and procedural artifacts. It was shown that polymer–solute and polymer–polymer interaction parameters contain other solute-dependent contributions to the free energy of mixing that are not properly accounted for by the polymer solution thermodynamic theories[105] and that the variation of polymer–polymer interaction coefficient from solute to solute is partly due to experimental error and artifacts. A series of critical and corrective procedures that yielded a considerable reduction in experimental error was developed by Card et al.[113]

In order to improve IGC data, new treatments to the polymer solution theories were developed. The reader is referred to original work by Munk et al.,[103] Prolongo and co-workers,[114,115] Shi and Schreiber,[116] Farooque and Deshpande,[117] and El Hibri and Munk.[118]

In another attempt to minimize the chemical dependence on the interaction parameters, Tan and Vancso[119] proposed an alternative route for experimentalists to

study the microstructure of polymeric mixtures using the IGC method. They applied the concept of the cluster integrals to the nonrandom partitioning of solutes in a polymeric blend. Their application yielded an equation that relates the structure factor of a blend, the nonrandom distribution of solutes in a compatible blend, and the experimental quantity obtained from the IGC experiment. This equation delineates the polymer interaction parameters in terms of the structure factor of the blend.

10 INVERSE GAS CHROMATOGRAPHY STUDIES ON NATURAL AND SYNTHETIC FIBERS

IGC is particularly useful in the study of surface interactions of fibers. In these experiments it is only necessary to pack the fibers directly in the open column. Care should be taken that the fibers are uniformly dispersed in the column and not packed in such a way as to prevent access by the carrier gas to the polymer surface. Column temperatures should not exceed the melting point of the fibers and the data are analyzed by the surface energy techniques described in sections 5.1, 5.2 and 5.3.

The IGC method is used to characterize carbon[120] and cellulose fibers,[121] high-solid melamine–alkyd formulations,[122] epoxy resins,[123–125] wood,[126] Kevlar®,[127] and PE fibers,[128,129] and eucalypti Kraft pulp fibers.[130] The surface properties of these fibers were the prime interest in these studies which include the effects of surface extractives and lignin on the surface energy,[130] comparison of surface chemical properties of chlorine-dioxide-bleached pulps with peroxides and ozone-bleached pulps and to study the influence of a latex binder,[131] the thermodynamic surface properties,[129] the properties of the skin layer of ultrahigh modulus PE fibers,[128] mapping of the surface energy levels of Kevlar® fibers,[127] solvent–epoxy interaction parameters,[123–125] solubility parameters of glass fibers,[124] the nature of interactions within pigmented melamine–alkyd coating assemblies,[122] and the dispersive component of the free energy and acid–base properties of cellulose fiber surfaces.[121] Several of these studies correlated their IGC results with those obtained by other methods such as X-ray photoelectron spectroscopy (XPS), electrical conductance, electron microscopy, thermal analysis, and surface wetting studies.

11 BIOLOGICAL MOLECULES

There are only a few applications of IGC to food compared with applications to polymeric systems. These include studies on food that can easily coat a column, such as soybean,[132,133] dehydrated foods like potato starch, egg albumin and wheat gluten,[134] food packaging,[135] pasta and others.[132,136–139] King[133] has determined the solubility parameters of soybean oil δ_2 as well as the oil–solute interaction parameters of 22 solutes which indicated complete miscibility of the solutes with the oil under conditions of infinite dilution. He compared δ_2 at 59, 79, and 123 °C with the interaction parameters, and over this temperature range the δ_2 decreased from 7.9 to 6.9 $cal^{1/2}\,cm^{-3/2}$ with decreasing temperature, a trend consistent with the loss of cohesive energy density in the liquid soybean oil. Abraham and Whiting[132] characterized soybean oil in terms of polarity and polarizability. The solubility of a variety of solutes in soybean oil was factored into components that show the compound–soybean oil interactions that favor solubility. Dispersion forces dominate and provide the largest solute–oil interaction in all cases, while ethanol is the least soluble solute in soybean oil because it is small and lacks dispersive forces in spite of the acidity of the hydrogen bond.

Dry food has also been studied using IGC.[140] Using IGC, one could predict the equilibrium moisture content of potato starch, egg albumin and wheat gluten in the entire range of water activity (0.0–0.8). The isosteric heat of sorption increases with decreasing moisture content, suggesting a physical sorption on highly active polar sites. The heat of sorption of potato starch is greater than that for proteins.

Food packaging is another application of IGC, particularly the study of aroma escaping through thermosealable polymers used to packaged food.[135] Researchers determined ''scalping'' of aroma components in polymers via IGC solubility experiments. Thermosealable polyesters appear to be the best choice to minimize aroma loss by sorption in the packaging inner layer. Sorption of aroma components was shown to be selective; limonene was preferentially sorbed in linear low-density PE. This selectivity may lead to an imbalance in the flavor and may be more important than the prevention of overall loss of aroma.

12 APPLICATIONS

Applications of IGC to polymer systems focus on the determination of partition and activity coefficients (infinite dilution[82,92,108,141–143] and finite concentration[18,24,25,57,58,68,69,74,75,107,112,144–152]), polymer–solvent or polymer–polymer interaction parameters,[153–155] enthalpy of mixing,[44,58,142,156] solubility parameters,[156–161] glass-transition and melting temperatures,[8,93,156,162–167] crystallinity of polymer phase,[33,40,156,163,168–170] solute

diffusion coefficients,[47,52,55,171–176] crosslinking studies,[177–180] liquid crystals,[8] interfacial studies,[181,182] methodology,[103,114,116,183–188] hydrogen-bonding studies,[132] acidity studies,[189] food,[131,136–138] plasticizers,[190] surfactants,[191] epoxy-resin curing,[192] paints and coatings,[193] microstructure studies,[194] and the characterization of conducting polymers.[195–197] In addition, there are several excellent reviews on IGC.[111,151,181,185,198–207]

12.1 Amorphous and Semicrystalline Polymer Blends

There has been much work done on one-component stationary phases, mostly amorphous, using IGC. Such data are useful for polymers because they measure the interactions in the regions of concentration which are the most interesting in polymer science. Patterson et al.[57] resolved the difficulty of applying the usual thermodynamic equation of GC to polymeric stationary phases by introducing activity coefficients based on weight fraction. They arrived at a relationship linking IGC parameters with Flory–Huggins parameters. Many workers cautioned that anomalous behavior frequently observed in IGC in the temperature region associated with transitions often alters the resolution of the column and makes interpretation of the retention time difficult owing to changes in peak shape. Polymers at glass transition and melting points, however, offer a wealth of information about the structures and interactions in polymer stationary phases if the retention mechanism is studied quantitatively. Braun and Guillet[13] reanalyzed the retention curves in the glass retention region. They assumed that if the bulk sorption is the only retention mechanism, specific retention volumes (V_g) should be independent of the mass of the stationary phase and carrier gas flow rate. They concluded that bulk sorption and surface absorption mechanisms are operative within polymer stationary phases near T_g. They corrected the surface absorption to eliminate the dependence of the apparent thermodynamic data on the column loading. Many authors called attention to the effects of surface absorption and the necessity of performing a careful analysis of the thermodynamic data from IGC. Other authors claimed a reasonable agreement between IGC data and data obtained by other procedures.

12.2 Comparison with Other Methods

Several IGC researchers have compared their results with those obtained using other methods. The general consensus is that IGC is capable of producing data comparable with other techniques, although there has been disagreement in individual cases. Nevertheless, where there is a general agreement of results, workers have concluded that IGC is superior in speed, selectiveness in the range of experimental conditions, versatility and accuracy. Tait and Abushihada[208] among other groups compared IGC results to those obtained by vapor sorption techniques to measure polymer–solvent interaction parameters. They concluded that the IGC method is capable of obtaining data in the region (concentration and temperature) above which vapor pressure measurements became difficult, and IGC generally required less experimental time. Al-Saigh and Munk[105] concluded that IGC is sufficiently accurate to be of great value in measuring interaction parameters at infinite solvent dilution, and that solvent dependency of the measured polymer–polymer interaction parameter is a result of weakness in the thermodynamic theories used, and not a fault of the technique itself.

Al-Saigh and Chen[45] conducted experiments in a temperature range below the melting temperature of the semicrystalline diluent. As expected, a depression of the melting point of the crystalline component had occurred. They used the melting point depression to measure the miscibility by calculating B_{23}, and applied it to a blend of poly(ethyl methacrylate) (PEMA) and semicrystalline poly(vinylidene fluoride) (PVF_2). Thermodynamic quantities were in excellent agreement with those obtained by calorimetry on the same blend. For example, the B_{23} parameter for PVF_2/PEMA using DSC was $-3.18\,cal\,mL^{-1}$,[209] compared with $-3.15\,cal\,mL^{-1}$ using IGC.

Studies show that IGC agrees well with the DSC in detecting the immiscibility of poly(isobutyl methacrylate)/styrene–acrylic acid copolymer blends.[210] IGC gave positive values of the polymer–polymer interaction parameters, while DSC showed two separate glass transition temperatures. Both methods also agreed well on the miscibility of poly(styrene-co-acrylic acid) with a series of polycarbonates, showing negative values for χ_{23} and a single glass transition temperature. Although PS is immiscible with poly(isobutyl methacrylate) owing to the absence of specific interactions, it has been shown that a blend of a random copolymer (poly(styrene-co-acrylic acid)) and a homopolymer (PMMA) can be miscible in the absence of specific interactions, provided that there is sufficient repulsion between the two comonomers.[210] Both methods, IGC and DSC, have confirmed the miscibility of a blend of poly(isobutyl methacrylate) with poly(styrene-co-acrylic acid) containing 32 mol% acrylic acid and of poly(isobutyl methacrylate-co-acrylic acid) containing 22 mol% acrylic acid with poly(styrene-co-*N*,*N*-dimethylaminoethyl methacrylate) containing 23 mol% of basic comonomer.

Djadoun and co-workers have used both IGC and DSC to study the compatibility of a variety of copolymers with homopolymer[210–212] and have reported a good agreement between the two methods. Some of their

blends showed positive interaction coefficients using IGC and two glass transition temperatures with DSC. When a blend showed a negative interaction coefficient by IGC, the DSC confirmed the miscibility by showing one single glass transition temperature.

Guillet et al. used several methods, namely IGC, X-ray, IR and DSC,[9] to study the degree of crystallinity of poly(β-hydroxybutyrate-co-β-hydroxy) copolymers. The degree of crystallinity was roughly 60%. The results of IGC were quite similar to those of X-ray diffraction analysis. They found that DSC and density measurements were inappropriate for these copolymers since it is invalid to make a prior assumption about the density and the heat of fusion of the cocrystalline phase for these copolymers. IGC, however, does not depend on assumptions about the properties of the 100% crystalline polymer.

13 EXPERIMENTAL

Virtually any gas chromatograph can, with minor modifications, be used for IGC studies of polymers. Most of the studies published by Guillet et al. were carried out using commercially available GC instruments equipped with flame detectors, whereas groups such as those of Munk and Al-Saigh prefer to work with larger columns and thermal conductivity detectors which they claim give better data on the thermodynamics of polymer blends. The much higher sensitivity of the flame detector allows studies to be carried out much closer to infinite dilution and thus avoids the need to extrapolate from finite concentrations of injected probe.

For many studies, manual injection by microsyringe is satisfactory, but it is tedious if many data points are required. For the determination of retention diagrams early studies were made with automated syringes. A simpler procedure involves the use of a Hewlett-Packard headspace injector (Model 19395A) which injects a small sample of probe vapor mixed with methane as the marker gas. This gives excellent peaks with the equivalent of less than 0.01 µL of probe. A typical procedure is given below.

The polymer was deposited from $CHCl_3$ solution onto the Chromosorb G support by the pile coating method (section 13.3) to a level of 5.33 wt%. A total of 0.864 g of polymer was packed into a 6.35-mm (OD, outer diameter) $\times$ 1.2-m copper tubing column. A column containing 0.372 g polymer at 2.99 wt% was also studied in order to confirm the invariance of specific retention volume with column loading. IGC measurements were performed on a Hewlett-Packard 5890 Series II gas chromatograph equipped with a flame ionization detector. A precision mass flow controller with a range of 0–20 mL min^{-1} was used to control the nitrogen carrier gas at a nominal flow rate of 10 mL min^{-1}. Specific retention volumes were confirmed to be independent of the carrier flow rate at low flow rates. Barometric room pressure was measured to $\pm$6.67 Pa (0.05 mmHg), and column pressures were measured with a mercury manometer. Carrier gas flow rates were measured from the end of the column with a water-jacketed soap-bubble flow meter. Injections of probe vapor at infinite dilution (equivalent to $\leq$0.01 µL of liquid), mixed with methane gas as a marker, were made from a Hewlett-Packard headspace injector (Model 19395A). Data were collected, recorded and analyzed using a Hewlett-Packard Vectra ES-12 personal computer with the associated GC software package HP 3365 Chemstation. The column temperature was continuously measured to 0.1 °C with a thermocouple interfaced to the computer. The column was equilibrated for several minutes for each temperature increment of 1 or 2 °C; the necessary equilibration time was confirmed by the invariance of the retention times with the equilibration times. Equilibration is not necessary for transition temperature determinations. In this case injections are usually made at intervals during a continuous slow heating rate of 0.5 °C min^{-1}. However, thermal equilibration is usually necessary to obtain good activity coefficient data. The retention times t_r and t_m were determined from the positions of the peak maxima and recorded in the computer.

13.1 Thermodynamic Measurement on Blends

Figure 13 shows the design of an automated inverse gas chromatograph used for thermodynamic studies of blends. With certain modifications, it uses the plumbing and circuitry for any conventional gas chromatograph. For thermodynamics studies, a thermal conductivity detector is preferred because of the ease of continuous monitoring of the flow rate, however, a flame ionization detector can be modified by adding a three-way valve for diverting the flow when the flow rate is ready to be measured. Nitrogen or helium can be used as a carrier gas, although the reader should be aware of the problem of helium diffusion through the bubble in the flow meter. The design allows for continuous monitoring of the flow rate, temperature, and inlet and outlet pressures, quantities that must be measured with great precision. The design makes it possible to control these quantities electronically by means of controllers or a personal computer, and to adjust these quantities as needed. The software allows researchers to monitor and adjust these quantities as well as to collect and calculate the chromatographic quantities needed for the determination of the thermodynamic parameters.

Continuous monitoring of carrier gas flow and gas pressure is important to limit the fluctuation in the flow

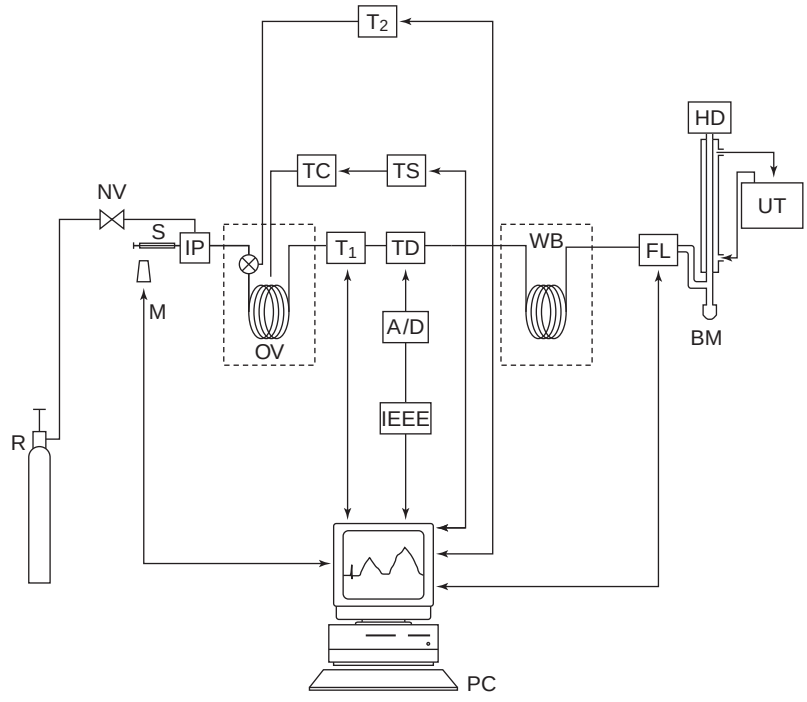

Figure 13 Inverse gas chromatograph set-up. A/D, analog/digital convertor; BM, bubble meter; FL, flow meter; HD, hood; IEEE, IEEE 488 board; IP, injection port; M, mouse; NV, needle valve; OV, oven; PC, personal computer; R, two-stage regulator; S, syringe; T_1, transducer 1; T_2, transducer 2; TC, temperature control; TD, thermal detector; TS, temperature sensor; UT, ultrathermostat; WB, water bath.

to less than 0.3% in a period of 8 h of operation. The flow rate is varied within a range of 1–30 mL min^{-1} to explore its effect on V_g^0 values.

To maintain electronic communication, precision communication boards such as an IEEE 488 are interfaced with the apparatus. To measure the flow rate in milliliters per minute, an electronic temperature insensitive mass flow meter is used. The flow meter is calibrated using a thermostated soap bubble meter equipped with a glass hood (Figure 14). Thermostating is important in order to eliminate any temperature gradient along the bubble meter. The bubble meter must be fabricated with a cap as a hood to ensure a complete helium atmosphere inside the meter. If nitrogen is used as carrier gas, such a hood is not necessary. The thermostating procedure and the hood design have proved effective in lowering the error in the carrier gas flow measurements. The pressure of the gas exiting the thermal conductivity detector is close to atmospheric and passes through a coil immersed in a water bath thermostated at room temperature. The gas then enters the soap bubble meter. The atmospheric pressure is monitored with an electronic barometer that transmits readings back to the computer. A barometric pressure transmitter is available commercially as a silicon capacitive pressure sensor which offers high accuracy and long-term stability. This type of transmitter offers an accuracy of ±0.3 mbar per year. It sends a linear 0–5 V DC output signal to the data acquisition system. If electronic instruments for pressure measurements are not available, the inlet and outlet pressures should be monitored frequently using a precision mercury manometer. The outlet pressure is generally atmospheric while the inlet pressure is in the range of 900 mmHg in conventional IGC experiments.

Temperature control is a critical factor in IGC experiments. Conventional chromatographs often do not provide sufficient precision in temperature control over 8 h of operation to be of value in IGC experiments. It is important to interface a precision electronic thermometer with 0.01 °C accuracy. Fluctuation of more than 3% in temperature measurements may create a considerable error in V_g^0.

In IGC experiments, markers such as methane or air are used to account for the dead volume in a column. For the thermal conductivity detector, air or nitrogen are often used as a marker, but for a flame ionization detector, methane is the least noninteractive solute that can be used.

13.2 Column Preparation

Each sample must be weighed carefully and dissolved in the appropriate solvent (about 100 mL). The polymer

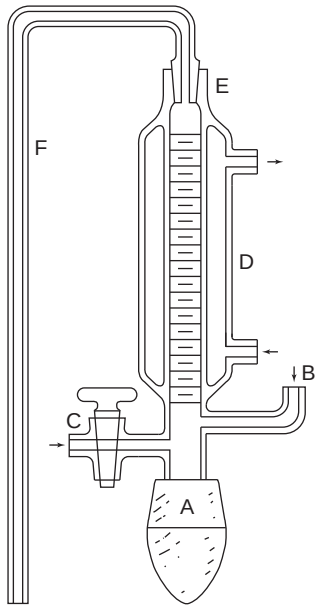

Figure 14 Flow meter is calibrated using a thermostated soap bubble meter equipped with a glass hood. A = soap solution reservoir; B = incoming helium gas from the column; C = inlet valve for fast flushing by helium gas; D = thermostated bubble meter; E = ground joint; F = inverted U-tube 1/8 in ID.

solution can then be used to coat the solid support using the pile method. The loading of the polymer sample can be calculated relative to the weight of the solid support (wt%). Chromosorb W with 60/80 mesh is often treated with dimethylchlorosilane to deactivate the hydroxyl groups on the surface. Teflon™ powder is often used when the solid polymer particles are simply blended with the support.

A typical loading in a chromatographic column varies between about 3 and 10%.

13.3 Coating the Support: the Pile Method

In this method the polymer is dissolved in a solvent. The support is piled on a watch glass and a small amount of the solution is applied on the top of the support pile. Care is taken to wet the pile as much as possible without letting the solution touch the surface of the dish, either around or under the pile. The solvent is allowed to evaporate and the pile is thoroughly mixed. Then the next portion of the solution is applied and the whole procedure is repeated until all of the solution (including several rinsings of the solution flask) is consumed. It takes typically 10–20 applications and requires only a few hours. No polymer is left on the surface of the dish. Then the support is dried in an oven and quantitatively transferred into the column.

13.4 Determination of Degree of Crosslinking of Polymer Networks

A particularly valuable application of IGC is the determination of crosslink density in rubbers and other crosslinked polymers. Its simplicity and rapidity make it useful in industrial quality control and similar applications. It involves the use of finite concentration IGC as described in section 8.2.

13.4.1 Theory

The full treatment of the elution-on-a-plateau method has been given by Conder and Purnell[77–80] and by Price and Guillet.[214] It involves the equilibration of the polymer with a stream of inert carrier gas that contains a known concentration of solvent vapor. The retention volume, V_N, of a small amount of solvent injected into the stream is then measured from Equation (72)

$$V_N = (t_r - t_m)F \tag{72}$$

where t_r and t_m are the times taken for the solvent and an inert marker to traverse the column and F is the flow rate of carrier corrected to standard conditions. Under these conditions, it may be shown that the partition isotherm for the solvent between the vapor and solution phases is given by Equation (73)

$$\frac{\partial q}{\partial c} = \frac{jV_N}{v_p}(1 - \psi) \tag{73}$$

In Equation (73), q and c are the concentrations of solvent in the liquid and vapor phases, respectively, and v_p is the volume of polymer in the column. ψ is the mole fraction of solvent in the vapor stream, corrected by the factor j for gas-phase nonideality and compressibility effects from the pressure gradient along the column.

Integration of Equation (73) and introduction of the weight of polymer in the column, w, allow calculation of the concentration of solvent in the liquid phase, q, expressed as moles per gram of polymer, Equation (74)

$$q = \frac{j}{w} \int_0^c \frac{V_N}{1 - \psi}\, dc \tag{74}$$

The concentration in the vapor stream may be calculated from the first terms of a virial expansion, Equation (75)

$$c = \frac{\psi P_A}{jRT + \psi^2 B_{11} P_A} \tag{75}$$

where P_A is the total pressure of carrier gas at the column outlet and B_{11} is the second virial coefficient of the solvent vapor at the column temperature T.

From the definition of q given above, the weight fraction of solvent in the solution phase, W_1, is, Equation (76)

$$W_1 = \frac{qM_1}{1 + qM_1} \tag{76}$$

where M_1 is the molecular weight of the solvent (probe). The activity of the solvent in solution may be calculated from the partial pressure of solvent in the carrier stream, P_1, and the saturated vapor pressure of the solvent, P_1^0. Hence, the activity coefficient Ω_1 may be found, Equation (77)

$$\Omega_1 = \frac{a_1}{W_1} = \frac{P_1}{P_1^0} W_1 \tag{77}$$

The weight fraction of solvent in the liquid phase can be calculated from Equation (76) and converted to a volume fraction by using the densities ρ, Equation (78)

$$\phi_1 = \left\{ \frac{W_1/\rho_1}{(W_1/\rho_1)(W_2/\rho_2)} \right\} \tag{78}$$

The Flory–Huggins interaction parameter is then calculated from Equation (79)

$$\chi = \left\{ \frac{\ln a_1 - \ln \phi_1 - [1 - (1/r)]\phi_2}{\phi_2^2} \right\} \tag{79}$$

the size ratio being calculated from the ratio of the molar volumes of the polymer and solvent.

13.4.2 Experimental

The apparatus used is shown schematically in Figure 15.[178]

A known partial pressure of the solvent probe *n*-hexane was introduced into the helium carrier gas by a saturator consisting of a flask containing *n*-hexane near its boiling point and a condenser through which thermostated water was passed. The partial pressure of hexane in the carrier gas was controlled by changing the temperature of the water in the saturator condenser.

The carrier gas passed over the polymer packing contained in one-quarter-inch (6.25-mm) copper columns as in conventional GC, then to a thermal conductivity detector, and finally, via a cold trap, to a soap bubble flow meter for accurate measurement of the carrier flow rate. The columns were contained in an oil bath maintained at $77.0 \pm 0.02\,^{\circ}C$.

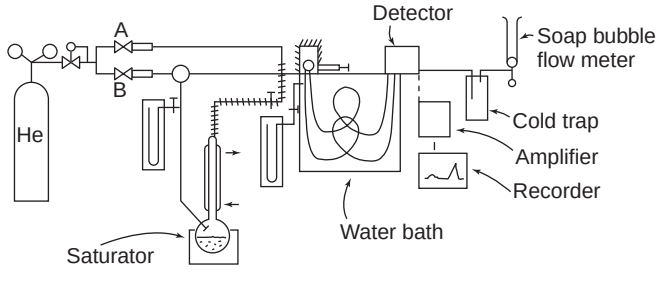

Figure 15 Schematic diagram of finite concentration GC apparatus.

After equilibration of the detector at a known concentration of solvent vapor, the retention times of air (the inert marker) and solvent were recorded. An average of at least three values agreeing within 1% was taken. Injections of 0.1–0.5 μL were needed to give an adequate signal.

Since the saturated vapor pressure of the solvent is required for the calculation of results, solvent purity is of great importance. The *n*-hexane used was a BDH Assured grade of certified purity > 99.5%. The polymer used was an Aldrich ethylene–propylene rubber copolymer of low average molecular weight and an ethylene content of 70%.

Coating the polymer onto a solid support of 30–45 mesh size Chromosorb P was carried out in the usual manner employing hot toluene as the solvent. The percentage loading was determined by duplicate calcination to be 8.8% after correction for volatiles in the uncoated solid support.

Approximately 50 g of the polymer-coated support was placed in Pyrex glass ampules, and these were evacuated on a glass-vacuum line to a pressure of $<10^{-5}$ Torr for 2 days to degas the polymers thoroughly. The ampoules were then sealed and placed in a ^{60}Co γ-ray source (Gammacell, Atomic Energy of Canada Ltd.). The approximate dose rates were estimated from the activity and known decay rates of the source supplied by the manufacturer. After irradiation, the ampoules were opened and the packing was loaded into columns.

In this experiment, three columns were prepared with the same amounts of polymer, sample A irradiated with 140 Mrad, sample B at 200 Mrad, and sample C (unirradiated).

A plot of χ versus concentration for the uncrosslinked polymer showed a linear variation of χ with volume fraction. Linear regression of the results leads to Equation (80) for χ:

$$\chi = 0.344 + 0.004\phi_1 \tag{80}$$

For the crosslinked samples, the effective number of crosslinks was calculated from the rearranged Flory–Rehner equation, Equation (81)

$$v_e = v_2 \left\{ \frac{\ln a_1 - \ln \phi_1 - [1 - (1/r)]\phi_2 - \chi\phi_2^2}{V_1[\phi_2^{1/3} - (\phi_2/2)]} \right\} \tag{81}$$

and the value of χ determined by Equation (79). If the Flory–Rehner equation were to hold accurately, v_e should be independent of concentration. However,

Figure 16 clearly shows that the value of v_e increases with concentration but rapidly reaches a plateau. The explanation for this behavior is that v_e cannot be determined until the network is fully swollen. The values of v_e and the molecular weight between crosslinks M_c taken from the graphs are shown in Table 4. The experimental error of the results is estimated to be 5–8%.

It is clear from these data that in the routine measurement of crosslink density it is only necessary to measure two retention times, namely V_n^0 and V_n^p, where V_n^0 is the retention time at zero concentration of hexane (probe) in the carrier gas and V_n^p is the retention time at a probe concentration above that at which the polymer sample is fully swollen. Since the latter concentration can be determined experimentally as shown in Figure 16, the procedure could result in a rapid method for the determination of crosslink density in polymer samples by IGC.

Although there is some debate about the theoretical aspects of the Flory–Rehner equation, it is still the basis of the most commonly used experiments to determine crosslink density in polymers. In has been shown[180] that the IGC procedure with crosslinked PDMSs gave results similar to those obtained from classical swelling measurements, but consistently about 30% higher in crosslink density. Further research on this useful technique should lead to an improved theory to account for these differences.

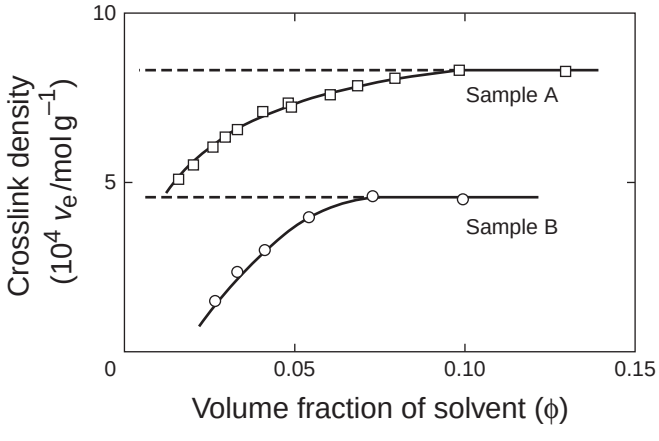

Figure 16 Plot of apparent crosslink density versus concentration for irradiated ethylene–propylene rubber samples.

Table 4 Degree of crosslinking for ethylene–propylene rubber from finite concentration results

Sample	Radiation dose (Mrad)	$10^4 v_e$	M_c
A	140	4.5	2500
B	200	8.3	1205

ABBREVIATIONS AND ACRONYMS

BET	Brunauer–Emmett–Teller
DnOP	Di-*n*-octyl Phthalate
DSC	Differential Scanning Calorimetry
GC	Gas Chromatography
GLC	Gas–Liquid Chromatography
GSC	Gas–Solid Chromatography
IGC	Inverse Gas Chromatography
IR	Infrared
OD	Outer Diameter
PDMS	Poly(dimethylsiloxane)
PE	Polyethylene
PEMA	Poly(ethyl methacrylate)
PEO	Poly(ethylene oxide)
PET	Poly(ethylene terephthalate)
P(E-VAc)	Poly(ethylene–vinyl acetate)
PIB	Poly(isobutylene)
PMMA	Poly(methyl methacrylate)
PP	Polypropylene
PS	Polystyrene
PVAc	Poly(vinylacetate)
PVC	Poly(vinyl chloride)
PVF_2	Poly(vinylidene fluoride)
VAc	Vinyl Acetate
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers

Gas Chromatography **(Volume 12)**
Liquid Phases for Gas Chromatography

REFERENCES

1. D.H. Everett, 'Effect of Gas Imperfection on GLC Measurements: a Refined Method for Determining

Activity Coefficients and Second Virial Coefficients', *Trans. Faraday Soc.*, **61**, 1637–1645 (1965).
2. J.E. Guillet, M. Galin, 'Studies of Polyacrylonitrile Structure by Inverse Chromatography', *J. Polym. Sci., Polym. Lett. Ed.*, **11**, 233–237 (1973).
3. G. Perrault, M. Tremblay, M. Bédard, G. Duchesne, R. Voyzelle, 'Applications de la Sonde Moleculaire aux Composites a Teneur Elevee en Solides. I. Determination des Temperatures de Transition Vitreuse et Etudes des Phenomenes de Surface', *Eur. Polym. J.*, **10**, 143–148 (1974).
4. D.G. Gray, J.E. Guillet, 'Open Tubular Columns for Studies on Polymer Stationary Phases by Gas Chromatography', *J. Polym. Sci., Polym. Lett. Ed.*, **12**, 231–235 (1974).
5. R.N. Lichtenthaler, D.D. Liu, J.M. Prausnitz, 'Polymer–Solvent Interactions from Gas–Liquid Chromatography with Capillary Columns', *Macromolecules*, **7**, 565–570 (1974).
6. R.J. Laub, J.H. Purnell, P.S. Williams, M.W.P. Harbison, D.E. Martire, 'Meaningful Error Analysis of Thermodynamic Measurements by Gas–Liquid Chromatography', *J. Chromatogr.*, **155**, 233–240 (1978).
7. J.E. Guillet, 'Molecular Probe Apparatus', U.S. Patent 3,717,040, Feb. 20, 1973.
8. M. Romansky, J.E. Guillet, 'The Use of Inverse Gas Chromatography to Study Liquid Crystalline Polymers', *Polymer*, **35**, 584–589 (1994).
9. W.J. Orts, M. Romansky, J.E. Guillet, 'Measurement of Crystallinity of Poly(β-hydroxybutyrate-co-β-hydroxyvalerate) Copolymers by Inverse Gas Chromatography', *Macromolecules*, **25**, 949–953 (1992).
10. V.R. Alishoev, V.G. Berezkin, Y.V. Mel'nikova, 'Effect of Phase Transitions in the Stationary Phase on the Chromatographic Behavior', *Russ. J. Phys. Chem.*, **39**, 105–107 (1965).
11. O. Smidsrød, J.E. Guillet, 'Study of Polymer–Solute Interactions by Gas Chromatography', *Macromolecules*, **2**, 272–277 (1969).
12. J.E. Guillet, 'Study of Polymer Structure and Interactions by Inverse Gas Chromatography', in *New Developments in Gas Chromatography*, ed. J.H. Purnell, Wiley, New York, 187–235, 1973.
13. J.-M. Braun, J.E. Guillet, 'Studies of Polystyrene in the Region of the Glass Transition Temperature by Inverse Gas Chromatography', *Macromolecules*, **8**, 882–888 (1975).
14. J.R. Condor, D.C. Locke, J.H. Purnell, 'Concurrent Solution and Adsorption Phenomena in Chromatography. I. General Considerations', *J. Phys. Chem.*, **73**, 700–708 (1969).
15. M.A. Llorente, C. Menduiña, A. Horta, 'Inverse Gas Chromatography of Poly(cyclohexyl methacrylate)', *J. Polym. Sci., Polym. Phys. Ed.*, **17**, 189–197 (1979).
16. D.D. Deshpande, O.S. Tyagi, 'Gas Chromatographic Behavior of Poly(vinyl acetate) at Temperatures Encompassing T_g: Determination of T_g and χ', *Macromolecules*, **11**, 746–751 (1978).
17. J.-M. Braun, J.E. Guillet, 'Glass Transitions of Polyolefins by Inverse Gas Chromatography', *J. Polym. Sci., Polym. Chem. Ed.*, **14**, 1073–1081 (1976).
18. J. Klein, H. Widdecke, 'The Problem of Solute Retention in Mixed Stationary Phases in Gas–Liquid Chromatography', *J. Chromatogr.*, **147**, 384–387 (1978).
19. R. Hayakawa, T. Nishi, K. Arisawa, Y. Wada, 'Dielectric Relaxation in the Paracrystalline Phase in Polyacrylonitrile', *J. Polym. Sci., Part A-2: Polym. Phys.*, **5**, 165–177 (1967).
20. P.L. Hsiung, D.M. Cates, 'Study of the Glass Transition of Poly(ethylene terephthalate) by Gas Chromatography', *J. Appl. Polym. Sci.*, **19**, 3051–3060 (1975).
21. I.A. Schneider, E.-M. Calugaru, 'The Estimation of Phase Modifications in Polymers by the Method of 'Molecular Probes' in Gas Chromatography. V. The Study of First and Second-order Transitions in Poly(ε-caprolactam)', *Eur. Polym. J.*, **13**, 833–835 (1977).
22. I.A. Schneider, E.-M. Calugaru, 'The Estimation of Phase Modifications in Polymers by the Method of 'Molecular Probes' in Gas Chromatography. III. Glass Transition Behaviour and Compatibility in Mixtures of Poly(methyl methacrylate) with Poly(vinyl chloride)', *Eur. Polym. J.*, **11**, 861–863 (1975).
23. K. Ito, H. Sakakura, Y. Yamashita, 'Characterization of Styrene–Tetrahydrofuran Block Copolymers by Inverse Gas Chromatography', *J. Polym. Sci., Polym. Lett. Ed.*, **15**, 755–759 (1977).
24. K. Ito, N. Usami, Y. Yamashita, 'Syntheses of Methyl Methacrylate–Stearyl Methacrylate Graft Copolymers and Characterization by Inverse Gas Chromatography', *Macromolecules*, **13**, 216–221 (1980).
25. J.-M. Braun, J.E. Guillet, 'Inverse Gas Chromatography in the Vicinity of T_g: Effects of the Probe Molecule', *Macromolecules*, **9**, 340–344 (1976).
26. D.G. Gray, J.E. Guillet, 'A Gas Chromatographic Method for the Study of Sorption of Polymers', *Macromolecules*, **5**, 316–321 (1972).
27. D.H. James, C.S.G. Phillips, 'The Chromatography of Gases and Vapours. Part III. The Determination of Adsorption Isotherms', *J. Chem. Soc.*, 1066–1070 (1954).
28. E. Glueckauf, 'Theory of Chromatography', *J. Chem. Soc.*, 1302–1329 (1947).
29. D. DeVault, 'The Theory of Chromatography', *J. Am. Chem. Soc.*, **65**, 532–540 (1943).
30. A.V. Kiselev, Le. Zung, Yu.S. Nikitin, 'Adsorption–Chromatographic Study of Porous Polyacrylates', *Kolloid Zh.*, **33**, 224–229 (1971).
31. S. Brunauer, *The Absorption of Gases and Vapors*, Princeton University Press, Princeton, NJ, Vol. 1, 1945.
32. J. Crank, G.S. Park (eds.), *Diffusion in Polymers*, Academic Press, London, 1968.
33. G. Courval, D.G. Gray, 'The Effect of Surface Adsorption on Gas Chromatographic Measurements Near

Polymer Melting Transitions', *Macromolecules*, **8**, 326–331 (1975).
34. M. Galin, M.C. Rupprecht, 'Study by Gas–Liquid Chromatography of the Interactions between Linear or Branched Polystyrenes and Solvents in the Temperature Range 60–200 °C', *Polymer*, **19**, 506–512 (1978).
35. U.-B. Mohlin, D.G. Gray, 'Gas Chromatography on Polymer Surfaces Adsorption on Cellulose', *J. Colloid Interface Sci.*, **47**, 747–754 (1974).
36. P.R. Tremaine, D.G. Gray, 'Adsorption of Non-swelling Vapours on the Surface of Cellulose', *J. Chem. Soc., Faraday I*, **71**, 2170–2185 (1975).
37. P.R. Tremaine, D.G. Gray, 'Determination of Brunauer–Emmett–Teller Monolayer Capacities by Gas–Solid Chromatography', *Anal. Chem.*, **48**, 380–382 (1976).
38. F.M. Fowkes, 'Quantitative Characterization of the Acid–Base Properties of Solvents, Polymers and Inorganic Surfaces', *J. Adhes. Sci. Technol.*, **4**, 669–691 (1990).
39. D.G. Gray, J.E. Guillet, 'The Application of the Molecular Probe Technique to a Study of Polymer Crystallization Rates', *Macromolecules*, **4**, 129–133 (1971).
40. J.-M. Braun, J.E. Guillet, 'Determination of Crystallinity by Gas Chromatography: Effect of Curvilinearity of Retention Diagrams', *Macromolecules*, **10**, 101–106 (1977).
41. P. Hudec, 'Determination of Crystallinity of Polyethylene Powder by Inverse Gas Chromatography', *Makromol. Chem.*, **178**, 1187–1193 (1977).
42. J.-M. Braun, J.E. Guillet, 'Determination of Crystallinity of Olefin Copolymers and Polyolefin Powders by Inverse Gas Chromatography', *J. Polym. Sci., Polym. Chem. Ed.*, **13**, 1119–1131 (1975).
43. R. Buchdahl, R.L. Miller, S. Newman, 'Crystallization Kinetics and Mechanical Properties of Polyethylene', *J. Polym. Sci.*, **36**, 215–231 (1959).
44. Y.S. Lipatov, A.E. Nesterov, 'The Influence of Thickness of Polymeric Stationary Phase on Its Properties Determined by Gas Chromatography', *Macromolecules*, **8**, 889–894 (1975).
45. Z.Y. Al-Saigh, P. Chen, 'Characterization of Semicrystalline Polymers by Inverse Gas Chromatography. 2. A Blend of Poly(vinylidene fluoride) and Poly(ethyl methacrylate)', *Macromolecules*, **24**, 3788–3795 (1991).
46. J.J. Van Deemter, F.J. Zuiderweg, A. Klinkenberg, *Chem. Eng. Sci.*, **5**, 271 (1956).
47. D.G. Gray, J.E. Guillet, 'Studies of Diffusion in Polymers by Gas Chromatography', *Macromolecules*, **6**, 223–227 (1973).
48. J.-M. Braun, S. Poos, J.E. Guillet, 'Determination of Diffusion Coefficients of Antioxidants in Polyethylene by Gas Chromatography', *J. Polym. Sci., Polym. Lett. Ed.*, **14**, 257–261 (1976).
49. J.C. Giddings, 'Advances in the Theory of Plate Height in Gas Chromatography', *Anal. Chem.*, **35**, 439–449 (1963).
50. W. Millen, S.J. Hawkes, 'Determination of Diffusion Coefficients in Polyethylene by Gas Chromatography', *J. Polym. Sci., Polym. Lett. Ed.*, **15**, 463–465 (1977).
51. D.G. Gray, J.E. Guillet, 'Gas Chromatography on Polymers at Temperatures Close to the Glass Transition', *Macromolecules*, **7**, 244–247 (1974).
52. P. Munk, T.W. Card, P. Hattam, M.J. El-Hibri, Z.Y. Al-Saigh, 'Inverse Gas Chromatography. 4. The Diffusion Phenomena on the Column', *Macromolecules*, **20**, 1278–1285 (1987).
53. O.E. Schupp, 'Gas Chromatography', in *Technique of Organic Chemistry*, eds. E.S. Perry, A. Weissberger, Interscience, New York, Chapter III, 1979.
54. J.R. Conder, C.L. Young, *Physicochemical Measurement by Gas Chromatography*, Wiley, New York, 1979.
55. J.S. Vrentas, C.M. Vrentas, I.H. Romdhane, 'Analysis of Inverse Gas Chromatography Experiments', *Macromolecules*, **26**, 6670–6672 (1993).
56. T.W. Card, Z.Y. Al-Saigh, P. Munk, 'Diffusion in the Bubble Flow Meter in IGC Experiments', *J. Chromatogr.*, **301**, 261–264 (1984).
57. D. Patterson, Y.B. Tewari, H.P. Schreiber, J.E. Guillet, 'Application of Gas–Liquid Chromatography to the Thermodynamics of Polymer Solutions', *Macromolecules*, **4**, 356–359 (1971).
58. G. DiPaola-Baranyi, J.-M. Braun, J.E. Guillet, 'Partial Molar Heats of Mixing of Small Molecules with Polymers by Gas Chromatography', *Macromolecules*, **11**, 224–227 (1978).
59. S. Dinçer, D.C. Bonner, 'Thermodynamic Analysis of an Ethylene and Vinyl Acetate Copolymer with Various Solvents by Gas Chromatography', *Macromolecules*, **11**, 107–113 (1978).
60. S. Dinçer, D.C. Bonner, 'Solubility Criteria for a Thermally Stable Polymer Determined by Gas–Solid Chromatography', *J. Appl. Polym. Sci.*, **22**, 3235–3247 (1978).
61. D. Patterson, 'Thermodynamics of Non-dilute Polymer Solutions', *Rubber Chem. Technol.*, **40**, 1–35 (1967).
62. R.F. Blanks, J.M. Prausnitz, 'Thermodynamics of Polymer Solubility in Polar and Nonpolar Systems', *Ind. Eng. Chem. Fundam.*, **3**, 1–8 (1964).
63. R.L. Scott, M. Magat, 'Thermodynamics of High Polymer Solutions. III. Swelling of Cross-linked Rubber', *J. Polym. Sci.*, **4**, 555–571 (1949).
64. K. Ito, J.E. Guillet, 'Estimation of Solubility Parameters for Some Olefin Polymers and Copolymers by Inverse Gas Chromatography', *Macromolecules*, **12**, 1163–1167 (1979).
65. N.F. Brockmeier, R.W. McCoy, J.A. Meyer, 'Gas Chromatographic Determination of Thermodynamic Properties of Polymer Solutions', *Macromolecules*, **5**, 130–132 (1972).
66. W.E. Hammers, C.L. DeLigny, 'The Thermodynamics of Solutions of Some Normal and Branched Alkanes in Polyisobutylene: an Investigation by Gas–Liquid

Chromatography', *J. Polym. Sci., Part C: Polym. Symp.*, **39**, 273–280 (1972).
67. Y.-K. Leung, B.E. Eichinger, 'Gas Liquid Chromatography on Polymers. I. Polyisobutylene-hydrocarbons at 25 °C', *J. Phys. Chem.*, **78**, 60–64 (1974).
68. W.E. Hammers, C.L. DeLigny, 'A Gas Chromatographic Investigation of the Thermodynamics of Solutions of Some Tetrachloromethane in Polyisobutylene', *Recl. Trav. Chim. Pays-Bas*, **90**, 912–931 (1971).
69. W.R. Summers, Y.B. Tewari, H.P. Schreiber, 'Thermodynamic Interaction in Polydimethylsiloxane–Hydrocarbon Systems from Gas–Liquid Chromatography', *Macromolecules*, **5**, 12–16 (1972).
70. M. Galin, 'Gas Chromatographic Investigation of the Thermodynamic Interactions of Poly(dimethylsiloxane) or Poly(diethylsiloxane) with Some Solvents between 60 and 180 °C', *Macromolecules*, **10**, 1239–1244 (1977).
71. W.E. Hammers, B.C. Bos, L.H. Vaas, Y.J.W.A. Loomans, C.L. DeLigny, 'Partial Molar Excess Free Enthalpies of *n*-Alkenes in Apiezon M and Alkanes in Squalane and Polydimethylsiloxane, and the Solution Theories of Prigogine and Flory: an Investigation by Gas Chromatography', *J. Polym. Sci., Polym. Phys. Ed.*, **13**, 401–417 (1975).
72. W.E. Hammers, C.L. DeLigny, 'Determination of the Flory–Huggins χ Parameter, the Heat of Dilution, and the Excess Heat Capacity of Some Alkanes in Polydimethylsiloxane by Gas Chromatography', *J. Polym. Sci., Polym. Phys. Ed.*, **12**, 2065–2074 (1974).
73. N.F. Brockmeier, R.W. McCoy, J.A. Meyer, 'Gas Chromatographic Determination of Thermodynamic Properties of Polymer Solutions. I. Amorphous Polymer Systems', *Macromolecules*, **5**, 464–470 (1972).
74. Y.B. Tewari, H.P. Schreiber, 'Thermodynamic Interactions in Polymer Systems by Gas–Liquid Chromatography', *Macromolecules*, **5**, 329–332 (1972).
75. H.P. Schreiber, Y.B. Tewari, D. Patterson, 'Thermodynamic Interactions in Polymer Systems by Gas–Liquid Chromatography. III. Polyethylene-hydrocarbons', *J. Polym. Sci., Polym. Phys. Ed.*, **11**, 15–24 (1973).
76. G. DiPaola-Baranyi, J.E. Guillet, H.-E. Jeberien, J. Klein, 'Thermodynamics of Hydrogen Bonding Polymer–Solute Interactions by Inverse Gas Chromatography', *Makromol. Chem.*, **181**, 215–226 (1980).
77. J.R. Condor, J.H. Purnell, 'Gas Chromatography at Finite Concentrations. Part 1. Effect of Gas Imperfection on Calculation and the Activity Coefficient in Solution from Experimental Data', *Trans. Faraday Soc.*, **64**, 1505–1512 (1968).
78. J.R. Condor, J.H. Purnell, 'Gas Chromatography at Finite Concentrations. Part 2. A Generalized Retention Theory', *Trans. Faraday Soc.*, **64**, 3100–3111 (1968).
79. J.R. Condor, J.H. Purnell, 'Gas Chromatography at Finite Concentrations. Part 3. Theory of Frontal and Elution Techniques of Thermodynamic Measurement', *Trans. Faraday Soc.*, **65**, 824–838 (1969).
80. J.R. Condor, J.H. Purnell, 'Gas Chromatography at Finite Concentrations. Part 4. Experimental Evaluation of Methods for Thermodynamic Study of Solutions', *Trans. Faraday Soc.*, **65**, 839–848 (1969).
81. D. Patterson, A. Robard, 'Thermodynamics of Polymer Compatibility', *Macromolecules*, **11**, 690–695 (1978).
82. O. Olabisi, 'Polymer Compatibility by Gas–Liquid Chromatography', *Macromolecules*, **8**, 316–322 (1975).
83. M. Galin, M.C. Rupprecht, 'Gas Chromatographic Study of the Interactions in Styrene–Dimethylsiloxane Block Copolymers and Blends', *Macromolecules*, **12**, 506–511 (1979).
84. J.S. Aspler, D.G. Gray, 'Gas Chromatographic and Static Measurements of Solute Activity for a Polymeric Liquid Crystalline Phase', *Macromolecules*, **12**, 562–566 (1979).
85. N.F. Brockmeier, R.E. Carlson, R.W. McCoy, 'Gas Chromatographic Determination of Thermodynamic Properties of Polymer Solutions at High Temperatures', *AIChE J.*, **19**, 1133–1139 (1973).
86. D.D. Liu, J.M. Prausnitz, 'Solubilities of Gas and Volatile Liquids in Polyethylene and in Ethylene–Vinyl Acetate Copolymers in the Region 125–225 °C', *Ind. Eng. Chem. Fundam.*, **15**, 330–335 (1976).
87. T. Sugiyama, T. Takeuchi, Y. Suzuki, 'Thermodynamic Properties of Solute Molecules at Infinite Dilution Determined by Gas–Liquid Chromatography. 1. Intermolecular Energies of *n*-Alkane Solute in C_{28}–C_{36} *n*-Alkane Solvents', *J. Chromatogr.*, **105**, 265–272 (1978).
88. T. Sugiyama, T. Takeuchi, Y. Suzuki, 'Thermodynamic Properties of Solute Molecules at Infinite Dilution Determined by Gas–Liquid Chromatography. 2. Interaction Energies of Contact Pairs of Methyl and Methylene Segments', *J. Chromatogr.*, **105**, 273–277 (1978).
89. D.D. Liu, J.M. Prausnitz, 'Solubilities of Volatile Solutes in Poly(vinyl acetate) from 125 to 200 °C', *J. Polym. Sci., Polym. Phys. Ed.*, **15**, 145–153 (1977).
90. D.P. Maloney, J.M. Prausnitz, 'Solubility of Ethylene in Liquid, Low-density Polyethylene at Industrial Separation Pressures', *Ind. Eng. Chem. Process Res. Dev.*, **15**, 216–220 (1976).
91. D.P. Maloney, J.M. Prausnitz, 'Solubilities of Ethylene and Other Organic Solutes in Liquid Low-density Polyethylene in the Region 124° to 300 °C', *AIChE J.*, **22**, 74–82 (1976).
92. P.A. Small, 'Some Factors Affecting the Solubility of Polymers', *J. Appl. Chem.*, **3**, 71–80 (1953).
93. G. DiPaola-Baranyi, J.E. Guillet, 'Estimation of Solubility Parameters by Gas Chromatography', *Macromolecules*, **11**, 228–235 (1978).
94. J.E.G. Lipson, J.E. Guillet, 'Studies of Polar and Nonpolar Probes in the Determination of Infinite-dilution Solubility Parameters', *J. Polym. Sci., Polym. Phys. Ed.*, **19**, 1199–1209 (1981).

95. W. Mendenhall, R. Schaeffer, *Mathematical Statistics with Applications*, Duxbury Press, North Scituate, MA, 1973.
96. G. DiPaola-Baranyi, J.E. Guillet, H.-E. Jeberien, J. Klein, 'Estimation of Solubility Parameters for Poly(vinyl acetate) by Inverse Gas Chromatography', *J. Chromatogr.*, **166**, 349–356 (1978).
97. C.M. Hansen, A. Beerbower, 'Solubility Parameters', *Kirk–Othmer Encyclopedia of Chemical Technology*, 2nd edition, suppl., ed. A. Standen, Interscience, New York, 889–910, 1971.
98. R. Kumar, J.M. Prausnitz, 'Solvents in Chemical Technology', in *Solutions and Solubilities, Part I*, ed. M.R. Dack, Wiley-Interscience, New York, 259–326, 1976.
99. P.J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
100. I.C. Sanchez, R.H. Lacombe, 'Statistical Thermodynamics of Polymer Solutions', *Macromolecules*, **11**, 1145–1156 (1978).
101. D.R. Paul, S. Newman (eds.), *Polymer Blends*, Academic Press, New York, 1978.
102. O. Olabisi, L.M. Robeson, M.T. Shaw, *Polymer–Polymer Miscibility*, Academic Press, New York, 1979.
103. P. Munk, P. Hattam, A.A. Abdel-Azim, Q. Du, 'Thermodynamics of Miscible Polymer Blends Studied by Inverse Gas Chromatography', *Makromol. Chem., Macromol. Symp.*, **38**, 205–220 (1990).
104. J. Pouchly, A. Zivny, K. Solc, 'Thermodynamic Equilibrium in the System Macromolecular Coil–Binary Solvent', *J. Polym. Sci., Part C: Polym. Symp.*, **23**, 245–256 (1968).
105. Z.Y. Al-Saigh, P. Munk, 'Study of Polymer–Polymer Interaction Coefficients in Polymer Blends Using Inverse Gas Chromatography', *Macromolecules*, **17**, 803–809 (1984).
106. O.S. Tyagi, S.M. Sajjad, S. Husain, 'Polymer–Polymer Interaction Parameter Determined by Inverse Gas Chromatography', *Polymer*, **28**, 2329–2334 (1987).
107. D.D. Deshpande, D. Patterson, H.P. Schreiber, C.S. Su, 'Thermodynamic Interactions in Polymer Systems by Gas–Liquid Chromatography. IV. Interactions between Components in a Mixed Stationary Phase', *Macromolecules*, **7**, 530–535 (1974).
108. C.S. Su, D. Patterson, 'Determination by Gas–Liquid Chromatography of the Polystyrene–Poly(vinyl methyl ether) Interaction', *Macromolecules*, **10**, 708–710 (1977).
109. G. DiPaola-Baranyi, S.J. Fletcher, P. Degré, 'Gas Chromatographic Investigation of Poly(vinylidene fluoride)–Poly(methyl methacrylate) Blends', *Macromolecules*, **15**, 885–889 (1982).
110. G. DiPaola-Baranyi, P. Degré, 'Thermodynamic Characterization of Polystyrene-Poly(*n*-butyl methacrylate) Blends', *Macromolecules*, **14**, 1456–1460 (1981).
111. Z.Y. Al-Saigh, 'Inverse Gas Chromatography for the Characterization of Polymer Blends', *Int. J. Polym. Anal. Charact.*, **3**, 249–291 (1997).
112. C.S. Su, D. Patterson, H.P. Schreiber, 'Thermodynamic Interactions and the Properties of PVC–Plasticizer Systems', *J. Appl. Polym. Sci.*, **20**, 1025–1034 (1976).
113. T.W. Card, Z.Y. Al-Saigh, P. Munk, 'Inverse Gas Chromatography. 2. The Role of ''Inert'' Support', *Macromolecules*, **18**, 1030–1034 (1985).
114. M.G. Prolongo, R.M. Masegosa, A. Horta, 'Polymer–Polymer Interaction Parameter in the Presence of a Solvent', *Macromolecules*, **22**, 4346–4351 (1989).
115. E.G. Lezcano, C.S. Coll, M.G. Prolongo, 'Thermodynamic Study of Poly(4-hydroxystyrene)/Poly(vinyl acetate) Blends by Inverse Gas Chromatography', *Macromolecules*, **25**, 6849–6854 (1992).
116. Z.H. Shi, H.P. Schreiber, 'On the Application of Inverse Gas Chromatography to Interactions in Mixed Stationary Phases', *Macromolecules*, **24**, 3522–3527 (1989).
117. A.M. Farooque, D.D. Deshpande, 'Studies of Polystyrene–Polybutadiene Blend System by Inverse Gas Chromatography', *Polymer*, **33**, 5005–5018 (1992).
118. M.J. El-Hibri, P. Munk, 'Recent Developments in Inverse Gas Chromatography of Polymer Blends', *Polym. Prepr. (Am. Chem. Soc. Div. Polym. Chem.)*, **28**(1), 262–263 (1987).
119. Z. Tan, G.J. Vancso, 'Molecular Probing of Polymeric Microstructure and Nonrandom Partitioning of Solvents Absorbed in Polymers by Inverse Gas Chromatography', *Macromolecules*, **30**, 4665–4673 (1997).
120. A.B. Garcia, A. Cuesta, M.A. Montes-Moran, A. Martinez-Alonso, J.M.D. Tascon, 'Zeta Potential as a Tool to characterize Plasma Oxidation of Carbon Fibers', *J. Colloid Interface Sci.*, **192**, 363–367 (1997).
121. M.N. Belgacem, G. Czeremuszkin, S. Sapieha, 'Cellulose, Surface Characterization of Cellulose Fibers by XPS and Inverse Gas Chromatography', *Cellulose*, **2**, 145–157 (1995).
122. J.T. Guthrie, M. Kunaver, 'Interactions in Coatings Formulations. Part 3. The Influence of Selected Pigment Paste Assemblies', *Dyes Pigment*. **26**, 201–215 (1994).
123. P.L. Jackson, M.B. Huglin, A. Cervenka, 'Use of Inverse Gas Chromatography to Quantify Interactions in Amine Cured Epoxy Resins', *Polym. Int.*, **35**, 119–133 (1994).
124. P.L. Jackson, M.B. Huglin, A. Cervenka, 'Investigations by Inverse Gas Chromatography Relevant to Coated Glass Fiber Epoxy Composites', *Polym. Int.*, **35**, 145–152 (1994).
125. P.L. Jackson, M.B. Huglin, A. Cervenka, 'Use of Inverse Gas Chromatography to Quantify Interactions in Anhydride Cured Epoxy Resins', *Polym. Int.*, **35**, 135–143 (1994).
126. M. Kazayawoko, J.J. Balatinecz, M. Romansky, 'Thermodynamics of Adsorption of *n*-Alkanes on Maleated Wood Fibers by Inverse Gas Chromatography', *J. Colloid Interface Sci.*, **190**, 408–415 (1997).
127. S. Rebouillat, J.B. Donnet, H. Guo, T.K. Wang, 'Surface Energy Mapping of Kevlar® Fibers by Inverse Gas

Chromatography', *J. Appl. Polym. Sci.*, **67**, 487–500 (1998).
128. L.I. Kuzub, A.I. Efremova, E.N. Raspopova, O.S. Svechnikova, N.I. Shut, V.I. Irzhak, 'Properties of the Surface Layer of Ultrahigh Modulus Polyethylene Fiber', *Vysokomol. Soedin., Ser. B*, **35**, 308–311 (1993).
129. H. Chtourou, B. Riedl, B.V. Kokta, 'Surface Characterizations of Modified Polyethylene Pulp and Wood Pulps Fibers Using XPS and Inverse Gas Chromatography', *J. Adhes. Sci. Technol.*, **9**, 551–574 (1995).
130. W. Shen, I.H. Parker, Y.J. Sheng, 'The Effects of Surface Extractives and Lignin on the Surface Energy of Eucalpt Kraft Pulp Fibers', *J. Adhes. Sci. Technol.*, **12**, 161–174 (1998).
131. A. Lundqvist, L. Oedberg, J.C. Berg, 'Surface Characterization of Non-chlorine-bleached Pulp Fibers and Calcium Carbonate Coatings Using Inverse Gas Chromatography', *Tappi J.*, **78**, 39–42 (1995).
132. M.H. Abraham, G.S. Whiting, 'Hydrogen Bonding. Part 22. Characterization of Soybean Oil and Prediction of Activity Coefficients in Soybean Oil from Inverse Gas Chromatographic Data', *J. Am. Oil Chem. Soc.*, **69**, 1236–1238 (1992).
133. J.W. King, 'Determination of the Solubility Parameter of Soy Bean Oil by Inverse Gas Chromatography', *Food Sci. Technol. (London)*, **28**, 190–195 (1995).
134. M. Lagoudaki, P.G. Demertzis, G. Panagiotis, 'Equilibrium Moisture Characteristics of Dehydrated Food Constituents as Studied by a Modified Inverse Gas Chromatographic Method', *J. Sci. Food Agric.*, **65**, 101–109 (1994).
135. R. Gavara, R. Catala, P. Hernandez-Munoz, 'Study of Aroma Scalping Through Thermosealable Polymers Used in Food Packaging by Inverse Gas Chromatography', *Food Addit. Contam.*, **14**, 609–616 (1997).
136. M. Lagoudaki, P.G. Demertzis, M.G. Kontominas, *Lebensm. Wiss. Technol.*, **26**, 512 (1993).
137. K.A. Riganakos, P.G. Demertzis, M.G. Kontominas, 'Water Sorption by Wheat and Soy Flour: Comparison of Three Methods', *J. Cereal Sci.*, **20**, 101–106 (1994).
138. K.A. Riganakos, M.G. Kontominas, 'GC Study of the Effect of Specific Heat Treatment on the Sorption of Water by Wheat and Soy Flour', *Z. Lebensm. Unters. Forsch.*, **198**, 47–51 (1994).
139. H. Kumagai, M. Iwase, H. Kumagai, A. Mizuno, T. Yano, 'Application of Solution Thermodynamics to the Water Sorption Isotherms of Food Materials', *Biosci. Biotechnol. Biochem.*, **58**, 475–481 (1994).
140. H. Balard, A. Saada, J. Hartmann, O. Aouadj, E. Papirer, 'Estimation of Surface Energetic Heterogeneity of Fillers by Inverse Gas Chromatography', *Macromol. Symp.*, **108**, 63–80 (1996).
141. G. DiPaola-Baranyi, P. Degré, 'Thermodynamic Characterization of Polystyrene–Poly(*n*-butyl methacrylate) Blends', *Macromolecules*, **14**, 1456–1460 (1981).
142. A. Robard, D. Patterson, 'Temperature Dependence of Polystyrene–Poly(vinyl methyl ether) Compatibility in Trichloroethene', *Macromolecules*, **10**, 1021–1025 (1977).
143. M.J. Fernandez-Berridi, G.M. Guzman, J.M. Elorza, L. Garijo, 'Study by Gas–Liquid Chromatography of the Thermodynamics of the Interactions of Poly(vinyl acetate) with Various Solvents', *Eur. Polym. J.*, **19**, 445–450 (1983).
144. C.P. Doubé, D.J. Walsh, 'Studies of Poly(vinyl chloride)/Solution Chlorinated Polyethylene Blends by Inverse Gas Chromatography', *Eur. Polym. J.*, **17**, 63–67 (1981).
145. T. Nishi, T.T. Wang, 'Melting Point Depression and Kinetic Effects of Cooling on Crystallization in Poly(vinylidene fluoride)–Poly(methyl methacrylate) Mixtures', *Macromolecules*, **8**, 909–915 (1975).
146. T.P. Russell, H. Ito, G.D. Wignall, 'Neutron and X-ray Scattering Studies on Semicrystalline Polymer Blends', *Macromolecules*, **21**, 1703–1709 (1988).
147. C.A. Barron, S.K. Kumar, J.P. Runt, 'Crystal–Amorphous Interphases in Semi-crystalline Polymers and Polymer Blends', *Polym. Prepr. (Am. Chem. Soc. Div. Polym. Chem.)*, **33**(2), 610–611 (1992).
148. D.J. Walsh, J.S. Higgins, S. Rostami, K. Weeraperuma, 'Compatibility of Ethylene–Vinyl Acetate Copolymers with Chlorinated Polyethylenes. 2. Investigation of the Thermodynamic Parameters', *Macromolecules*, **16**, 391–396 (1983).
149. M.J. Fernandez-Berridi, G.M. Guzman, J.K. Iruin, J.M. Elorza, 'Determination of the Interaction Parameter χ of Poly(ethylene oxide) by Gas–Liquid Chromatography below the Melting Temperature', *Polymer*, **24**, 417–422 (1983).
150. F.H. Covitz, J.W. King, 'Solute Absorption into Molten Polystyrene', *J. Polym. Sci., Part A-1: Polym. Chem.*, **10**, 689–699 (1972).
151. M.J. El-Hibri, W. Cheng, P. Hattam, P. Munk, 'Inverse Gas Chromatography of Polymer Blends: Theory and Practice', in *Inverse Gas Chromatography, Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 10, 1989.
152. R.D. Newman, J.M. Prausnitz, 'Polymer–Solvent Interactions from Gas–Liquid Partition Chromatography', *J. Phys. Chem.*, **76**, 1492–1496 (1972).
153. W. Merk, R.N. Lichtenthaler, J.M. Prausnitz, 'Solubilities of Fifteen Solvents in Copolymers of Poly(vinyl acetate) and Poly(vinyl chloride) from Gas–Liquid Chromatography. Estimation of Polymer Solubility Parameters', *J. Phys. Chem.*, **84**, 1694–1695 (1980).
154. A. Heintz, R.N. Lichtenthaler, J.M. Prausnitz, 'Solubilities of Volatile Solvents in Polyvinyl Acetate, Polyvinyl Chloride, and Their Random Copolymers', *Ber. Bunsenges. Phys. Chem.*, **83**, 926–928 (1979).

155. A. Voelkel, E. Andrzejewska, R. Maga, M. Andrzejewski, 'Characterization of Dispersive and Acid–Base Properties of Crosslinked Polymers by Inverse Gas Chromatography', *Polymer*, **34**, 3109–3111 (1993).
156. C.T. Chen, Z.Y. Al-Saigh, 'Characterization of Poly(ethyl methacrylate) by Inverse Gas Chromatography', *Polymer*, **31**, 1170–1176 (1990).
157. P. Munk, P.J. Hattam, Q. Du, 'Thermodynamic Interactions in Mixtures', *J. Appl. Polym. Sci.*, **43**, 373–399 (1989).
158. A. Voelkel, J. Janas, 'Inverse Gas Chromatographic Determination of Solubility Parameter and Binary Parameters of α,w-Diamino Oligoethers', *J. Chromatogr., A*, **669**, 89–95 (1994).
159. A. Voelkel, J. Janas, 'The Solubility Parameters of Fluorine-containing Oxyethylates', *J. Fluorine Chem.*, **67**, 75–81 (1994).
160. A. Voelkel, J. Janas, 'Solubility Parameters of Broad and Narrow Distributed Oxyethylates of Fatty Alcohols', *J. Chromatogr.*, **645**, 141–151 (1993).
161. P.L. Jackson, M.B. Huglin, 'Use of Inverse Gas Chromatography to Assess the Compatibility of Polyurethane Foam with Liquids at Different Temperatures', *Polym. Int.*, **33**, 409–411 (1994).
162. P. Munk, Z.Y. Al-Saigh, T.W. Card, 'Inverse Gas Chromatography. 3. Dependence of Retention Volume on the Amount of Probe Injected', *Macromolecules*, **18**, 2196–2201 (1985).
163. J.E. Guillet, A.N. Stein, 'Study of Crystallinity in Polymers by the Use of ''Molecular Probes'' ', *Macromolecules*, **3**, 102–105 (1970).
164. A. Lavoie, J.E. Guillet, 'Estimation of Glass Transition Temperatures from Gas Chromatographic Studies on Polymers', *Macromolecules*, **2**, 443–446 (1969).
165. G. Boiteux-Steffan, J.P. Soulie, D. Sage, J.-M. Letoffe, 'Multiple Transitions in Poly(allylbenzene): 1. Thermophysical Properties', *Polymer*, **26**, 1443–1447 (1985).
166. A.B. Wojcik, 'Porous Bead Aliphatic–Aromatic Methacrylate Copolymers. VII. TG, DTA, DSC, and Inverse Gas Chromatography Studies', *J. Appl. Polym. Sci.*, **39**, 179–187 (1990).
167. R. Sanetra, B.N. Kolarz, A. Wlochowicz, 'Determination of the Glass Transition Temperature of Poly(styrene-co-divinylbenzene) by Inverse Gas Chromatography', *Polymer*, **26**, 1181–1186 (1985).
168. C.T. Chen, Z.Y. Al-Saigh, 'Characterization of Semicrystalline Polymers by Inverse Gas Chromatography. 1. Poly(vinylidene fluoride)', *Macromolecules*, **22**, 2974–2981 (1989).
169. C. Bonnerup, P. Gatenholm, 'The Effect of Surface Energetics and Molecular Interdiffusion on Adhesion in Multicomponent Polymer Systems', *J. Adhes. Sci. Technol.*, **7**, 247–262 (1993).
170. S. Galassi, G. Audisio, 'The Application of the Molecular Probes Gas Chromatographic Technique to the Study of Polypropylene Transitions', *Makromol. Chem.*, **175**, 2975–2982 (1974).
171. L.Q. Xie, 'Measurement of Solute Diffusion in Polymers by Inverse Gas Chromatography Using Fused-silica Open Tubular Columns', *Polymer*, **34**, 4579–4584 (1993).
172. J.S. Vrentas, C.M. Vrentas, 'Solvent Self-diffusion in Glassy Polymer–Solvent Systems', *Macromolecules*, **27**, 5570–5576 (1994).
173. S. Prager, F.A. Long, 'Diffusion of Hydrocarbons in Polyisobutylene' *J. Am. Chem. Soc.*, **73**, 4072–4075 (1951).
174. P. Hattam, P. Munk, 'Inverse Gas Chromatography. 5. Computer Simulation of Diffusion Processes on the Column', *Macromolecules*, **21**, 2083–2090 (1988).
175. D. Arnould, R.L. Laurence, 'Size Effects on Solvent Diffusion in Polymers', *Ind. Eng. Chem. Res.*, **31**, 218–228 (1992).
176. R.E. Balthus, M.A. Alger, T.J. Stanley, 'Solubility and Diffusivity of Cyclic Oligomers in Poly(dimethylsiloxane) Using Capillary Column Inverse Gas Chromatography', *Macromolecules*, **26**, 5651–5656 (1993).
177. M.F. Grenier-Loustalot, G. Mouline, P. Grenier, 'Inverse Gas Chromatography Used to Follow Kinetics of Epoxy-amine Reactions in the Molten State', *Polymer*, **28**, 2275–2281 (1987).
178. G.J. Price, K.S. Siow, J.E. Guillet, 'The Use of Gas Chromatography to Determine the Degree of Crosslinking of a Polymer Network', *Macromolecules*, **22**, 3116–3119 (1989).
179. M.L. Abel, M.M. Chehimi, 'Comments on 'Characterization of Dispersive and Acid–Base Properties of Crosslinked Polymers by Inverse Gas Chromatography' by A. Voelkel et al.', *Polymer*, **35**, 1789–1790 (1994).
180. Z. Tan, R. Jaeger, G.J. Vancso, 'Crosslinking Studies of Poly(dimethylsiloxane) Networks: a Comparison of Inverse Gas Chromatography, Swelling Experiments, and Mechanical Analysis', *Polymer*, **35**, 3230–3236 (1994).
181. J. Guillet, M. Romansky, G.J. Price, R. van der Mark, 'Studies of Polymer Structure and Interactions by Automated Inverse Gas Chromatography', in *Inverse Gas Chromatography: Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 3, 1989.
182. F. Chen, 'Study of Acceptor–Donor Interactions at the Polymer Interface by Inverse Gas Chromatography Data Analysis', *Macromolecules*, **21**, 1640–1643 (1988).
183. O.S. Tyagi, D.D. Deshpande, 'Inverse Gas Chromatography of Poly(*n*-butylmethacrylate): Effect of Flow Rate on Specific Retention Volume and Detection of Glass Transition Temperature', *J. Appl. Polym. Sci.*, **34**, 2377–2388 (1987).
184. J.Y. Wang, G. Charlet, 'Use of Moment Analysis in Inverse Gas Chromatography', *Macromolecules*, **22**, 3781–3788 (1989).

185. P. Hattam, Q. Du, P. Munk, 'Computer Simulation of Elution Behavior of Probes in Inverse Gas Chromatography: Comparison with Experiment', in *Inverse Gas Chromatography: Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 4, 1989.
186. A.E. Bolvari, T.C. Ward, P.A. Koning, D.P. Sheehy, 'Experimental Techniques for Inverse Gas Chromatography', in *Inverse Gas Chromatography: Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 2, 1989.
187. M.J. El-Hibri, P. Munk, 'Marker Retention in Inverse Gas Chromatography Experiments on Polymers', *Macromolecules*, **21**, 264–266 (1988).
188. H. Becker, R. Gnauck, 'Determination of Dead Times for Inverse Gas Chromatography: Measurements by Use of Methane', *J. Chromatogr.*, **366**, 378–381 (1986).
189. R. Zhang, J. Jagiello, J.F. Hu, Z.Q. Huang, J.A. Schwarz, A. Datye, 'Effect of Tungsten Trioxide Loading on the Surface Acidity of WO3/Alumina', *Appl. Catal., A*, **84**, 123–139 (1992).
190. P.J. Kalaouzis, P.G. Demertzis, 'Influence of a Monomeric Plasticizer on the Water-sorption Behavior of Food-grade Plastics Packaging Materials', *Polym. Int.*, **32**, 125–130 (1993).
191. A. Voelkel, J. Janas, J.A. Garcia-Dominguez, 'Inverse Gas Chromatography in Characterization of Surfactants', *J. Chromatogr., A*, **654**, 135–141 (1993).
192. M.F. Grenier-Loustalot, P. Grenier, 'The Mechanism of Epoxy Resin Curing in the Presence of Glass and Carbon Fibers', *Polymer*, **33**, 1187–1199 (1992).
193. H. Balard, E. Papirer, 'Characterization and Modification of Fillers for Paints and Coatings', *Prog. Org. Coat.*, **22**, 1–17 (1993).
194. J. Jagiello, T.J. Bandosz, J.A. Schwarz, 'Study of Carbon Microstructure by Using Inverse Gas Chromatography', *Carbon*, **32**, 687–691 (1994).
195. M.M. Chehimi, M.L. Abel, E. Pigois-Landureau, M. Delamar, 'Characterization of Conducting Polymers by Inverse Gas Chromatography. Part II. Effect of Dopant on the Dispersive and Specific Properties of Polypyrrole', *Synth. Met.*, **60**, 183–194 (1993).
196. E. Pigois-Landureau, M.M. Chehimi, 'Quantitative Assessment of Acid–Base Properties of Chloride Doped Polypyrrole by Inverse Gas Chromatography', *J. Appl. Polym. Sci.*, **49**, 183–186 (1993).
197. M.M. Chehimi, E. Pigois-Landureau, M. Delamar, 'Characterization of Conducting Polymers by Inverse Gas Chromatography. Part 1. Dispersive and Acid–Base Properties of Conducting Polypyrrole and its Related Insulating Form', *J. Chim. Phys.*, 1173–1178 (1992).
198. J.E.G. Lipson, J.E. Guillet, 'Study of Structure and Interactions in Polymers by Inverse Gas Chromatography', in *Developments in Polymer Characterisation – 3*, ed. J.V. Dawkins, Applied Science Publishers, Barking, Chapter 2, 1982.
199. Z.Y. Al-Saigh, 'Recent Advances in the Characterization of Polymers and Polymer Blends Using the Inverse Gas Chromatography Method', *Polym. News*, **19**, 269–279 (1994).
200. J.-M. Braun, J.E. Guillet, 'Study of Polymers by Inverse Gas Chromatography', *Adv. Polym. Sci.*, **21**, 107–145 (1976).
201. P. Munk, 'Polymer Characterization Using Inverse Gas Chromatography', *Chem. Anal. (N.Y.)*, **113**, 151–200 (1991).
202. G. DiPaola-Baranyi, 'Thermodynamics of Polymer Blends by Inverse Gas Chromatography', in *Inverse Gas Chromatography, Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 9, 1989.
203. S.P. Wesson, R.E. Allred, 'Surface Energetics of Plasma-treated Carbon Fiber', in *Inverse Gas Chromatography, Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 15, 1989.
204. A.J. Vukov, D.G. Gray, 'Properties of Carbon Fiber Surfaces', in *Inverse Gas Chromatography, Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 13, 1989.
205. J. Schultz, L. Lavielle, 'Interfacial Properties of Carbon Fiber-epoxy Matrix', in *Inverse Gas Chromatography, Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 14, 1989.
206. E. Osmont, H.P. Schreiber, 'Surface Characteristics of Glass Fibers', in *Inverse Gas Chromatography, Characterization of Polymers and Other Materials*, eds. D.R. Lloyd, T.C. Ward, H.P. Schreiber, ACS Symposium Series no. 391, Washington, DC, Chapter 17, 1989.
207. Z. Tan, G.J. Vancso, 'Polymer–Polymer Interaction Parameters by Inverse Gas Chromatography. A Novel Molecular Interpretation of Nonrandom Partitioning of Solvent Probes in Polymer Blends', *Macromol. Theory Simul.*, **6**, 467–478 (1997).
208. P.J.T. Tait, A.M. Abushihada, 'Use of a Gas Chromatographic Technique for the Study of the Variation of the Interaction Energy Parameter with Temperature', *Macromolecules*, **11**, 918–922 (1978).
209. F.L. Imken, D.R. Paul, J.W. Barlow, 'Transition Behavior of Poly(vinylidene fluoride)/Poly(ethyl methacrylate)', *Polym. Eng. Sci.*, **16**, 593–601 (1976).
210. F. Feraz, A.S.H. Hamou, S. Djadoun, 'Miscibility of Poly(styrene-co-acrylic acid) with Polymethacrylates or Poly(methacrylate-co-vinylpyridine)', *Eur. Polym. J.*, **31**, 665–669 (1995).
211. S. Djadoun, F.E. Karasz, A.S.H. Hamou, 'Blends of Poly(isobutyl methacrylate) with Poly(styrene-co-acrylic

acid) and of Poly(isobutyl methacrylate-co-acrylic acid) with Poly(styrene-co-*N*,*N*-dimethyl aminoethyl methacrylate)', *Thermochim. Acta*, **282/283**, 399–410 (1996).

212. D.E. Cherrak, S. Djadoun, 'Inverse Gas Chromatography of Poly(ethylene oxide) with Poly(ethyl methacrylate) or Poly(ethyl methacrylate-co-methacrylic acid)', *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **51**, 209–219 (1992).

213. D.E. Cherrak, S. Djadoun, 'A Study of Mixtures of Poly(ethylene oxide) with Polystyrene or Poly(styrene-co-acrylic acid) by Inverse Gas Chromatography and Viscosimetry', *Polym. Bull.*, **27**, 289–295 (1991).

214. G.J. Price, J.E. Guillet, 'The Determination of Thermodynamic Properties of Polymer Solutions Using Finite Concentration Gas Chromatography', *J. Macromol. Sci., Chem.*, **A23**, 1487–1502 (1986).

Mechanical Properties of Polymers and Rubbers

Zhanjun (Justin) Gao
Eastman Kodak Company, Rochester, USA

1 Introduction **1**
2 Stress and Strain Relationship **2**
2.1 Hooke's Law 2
2.2 Rubber Elasticity 4
3 Time- and Temperature-dependent Behavior of Polymers **6**
3.1 Viscoelastic Model 6
3.2 Dynamic Mechanical Thermal Analysis 9
3.3 Effects of Temperature on Mechanical Properties 12
3.4 Time–Temperature Superposition 13
4 Ultimate Properties of Polymers and Rubbers **15**
4.1 General Stress–Strain Diagram 15
4.2 Cold Drawing in Crystalline Polymers 15
4.3 Temperature and Pressure Effects on Yield Stress 17
4.4 The Brittle–Ductile Transition 17
4.5 Crazing 18
4.6 Fracture 19
4.7 The Tearing of Rubbers: Extension of Griffith Theory 21
4.8 Temperature and Loading Rate Effects: Viscoelastic Rupture of Elastomers 22
4.9 Fatigue 23
4.10 Effects of Molecular Weight and Crystallinity 24
5 Polymer Blends and Composites **26**
5.1 Polymer Blends 27
5.2 Polymer Composites 28
5.3 Mechanical Behavior of Rubber-toughened Polymers 28
5.4 Polymer Thin Films with Fillers 30
Acknowledgments **33**
Abbreviations and Acronyms **33**
Related Articles **33**
References **33**

One major focus of this article is on models and testing techniques that can be applied readily to the research and development of polymers and rubbers. The article is divided into five sections, with the introduction being the first section. Sections 2 and 3 deal with stress–strain relationships and time- and temperature-dependent behavior. The linear Hooke's law is described first, followed by the molecular mechanism of rubber elasticity, the Mooney–Rivlin model, and the Ogden model. Three viscoelastic models are presented, together with the Boltzmann superposition principle. Various types of apparatus for viscoelastic property measurements are described. The Williams–Landel–Ferry (WLF) equation for time–temperature superposition is illustrated with experimental data.

Section 4 covers the ultimate properties. Topics include: cold drawing in crystalline polymers; temperature and pressure effects on yield stress; brittle–ductile transition; crazing; fracture; tearing of rubbers: extension of Griffith theory; viscoelastic rupture of elastomers; fatigue; and the effects of molecular weight and crystallinity. The ultimate properties are more complex and less well understood. Emphasis is placed on the most important aspects and techniques for understanding and representing yielding and fracture.

Section 5 is devoted to polymer blends and composites. Rubber-toughened polymers are used to discuss the enhancement of the deformation mechanisms in polymers. A general, mechanistic model is then presented to determine the moduli, break strengths, and break strains for polymer films containing liquid and solid microfillers.

1 INTRODUCTION

The emergence of synthetic polymers with their manifold uses has affected modern life so profoundly that our world is hardly imaginable without them. Polymers can encompass a wide range of material properties, of which hardness, deformability, toughness, and ultimate strength are amongst the most significant. Certain features, such as high modulus and impact strength, combined with low creep characteristics are desirable in a polymer if eventually it is to be subjected to loading. Unfortunately these are conflicting properties, because a polymer with a high modulus and a low creep response does not absorb energy by deforming easily, and hence has poor impact strength. This means that a compromise must be sought, depending on the application that the polymer is intended for, and this compromise requires knowledge of the mechanical response and properties in detail.

The practical utility of a polymer depends, to a great extent, on its mechanical attributes. For this reason,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the mechanical properties may be considered the most important of all the physical and chemical properties of high polymers for most applications. Thus, everyone working with such materials needs at least an elementary knowledge of their mechanical behavior and how this behavior can be modified by the numerous structural factors that can be varied in polymers.

The detail and in-depth discussion necessary for a thorough understanding of the mechanical properties of polymers and rubbers cannot be covered fully in the few pages of this article. The major goal is to present a concise review of various key aspects of the topic. Emphasis is placed upon general principles and useful empirical generalizations. One major focus is models and testing techniques that can be applied readily to the research and development of polymers and rubbers. Mathematical derivations are, for the most part, avoided, with only the final practical equations listed and numerous illustrative examples given. Extensive references to published literature enable the reader to find, easily and quickly, important sources of information on numerous topics.

An entire section is devoted to polymer blends and composites. The understanding of the mechanical properties of these materials is becoming more critical due to their increasing utility in structural applications and due to the complicated nature of nonlinear and synergistic effects of the phases in blend/composites in controlling their mechanical performance. The introduction of blends and composites has brought a new dimension to the applications of polymers as major load-bearing components, such as aircraft fuselages, as well as new challenges and higher expectations on understanding and predicting the mechanical properties of these materials. It has been reported[(1)] that nearly 5 million tons of thermoset composites representing a global business worth some $120 billion were used worldwide in 1997 in applications as diverse as bridges, cars, and windmills. A firm understanding of the physical properties of each polymer phase and the synergistic interactions of the different phases in blends and composites is essential in engineering design with these materials.

In the next section, the concepts of stress and strain are introduced, followed by descriptions of one-dimensional and three-dimensional Hooke's laws. The nonlinear deformation behavior of rubbers is then discussed, including the Mooney–Rivlin model and the Ogden model, which have been validated experimentally and widely used in constitutive modeling of rubber materials. The entropy-controlled molecular mechanism responsible for rubber elasticity is discussed and compared with that of metals.

Section 3 deals with the time- and temperature-dependent behavior of polymers. Three viscoelastic models (Maxwell model, Voigt model, and standard linear solid model) are presented first. The Boltzmann superposition principle is then discussed. A large portion of this section is devoted to various experiments and apparatus for viscoelastic property measurements, such as the storage and lost moduli and phase angles. Test equipment discussed includes creep and relaxation transient measurements, torsional pendulum, vibration reed, and forced vibration. The applicable frequency ranges of these tests are outlined. The WLF equation for time–temperature superposition is illustrated with experimental data.

The ultimate properties of polymers and rubbers are the subject of section 4. Various terminologies are first defined with the use of a general stress–strain diagram. Topics include: cold drawing in crystalline polymers; temperature and pressure effects on yield stress; brittle–ductile transition; crazing; fracture; tearing of rubbers: extension of Griffith theory; viscoelastic rupture of elastomers; fatigue; and the effects of molecular weight and crystallinity. The ultimate properties are more complex and less well understood. Emphasis is placed on the most important aspects and techniques for understanding and representing yielding and fracture.

Polymer blends and composites are discussed in section 5. Classifications of polymer blends and composites are given. Rubber-toughened polymers (high-impact polystyrene (HIPS) and rubber-toughened poly(methyl methacrylate) (RTPMMA)) are used as examples to discuss two enhancements of the deformation mechanisms in polymers: extensive shear yielding and crazing. A general, mechanistic model is then presented to determine the moduli, break strengths, and break strains for polymer films containing liquid and solid microfillers. This model, which is based on rigorous continuum mechanics principles, considers the filler–filler interactions, incorporates the nonlinear synergistic effects of fillers, and provides accurate predictions in comparison with experimental data.

2 STRESS AND STRAIN RELATIONSHIP

2.1 Hooke's Law

Forces applied to solids cause deformation. If we pull a rubber band, it stretches. If we compress a cylinder, it shrinks. If we bend a rod, it bends. If we twist a shaft, it twists. Tensile force causes tensile strain. Shear stress causes shear strain. To express these phenomena quantitatively, it is necessary to define strain measures, and the relationships between stress and strain.

Let an isotropic polymer solid bar with an initial length L_0 be subjected to an axial force P as shown in Figure 1.

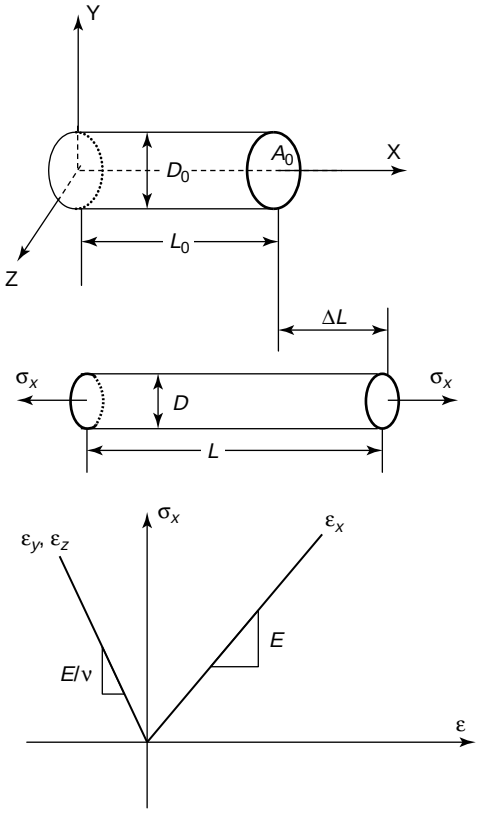

Figure 1 A tensile bar under load.

The axial stress is equal to the force per unit area, according to Equation (1):

$$\sigma_x = \frac{P}{A_0} \tag{1}$$

where A_0 is the cross-sectional area of the bar. If the bar is stretched to a length L, it is natural to describe the change by a dimensionless ratio called strain, according to Equation (2):

$$\varepsilon_x = \frac{(L - L_0)}{L_0} = \frac{\Delta L}{L_0} \tag{2}$$

where ε_x denotes the strain in the x-direction and ΔL is the change of length under the force P. In a similar manner, the strain along the diameter of the bar is as shown in Equation (3):

$$\varepsilon_y = \varepsilon_z = \frac{(D - D_0)}{D_0} = \frac{\Delta D}{D_0} \tag{3}$$

where D is the deformed diameter, D_0 is the initial diameter, and $\Delta D = D - D_0$. This transverse strain is negative as the bar becomes thinner when stretched in the length direction.

The experimental results can be presented as a curve of the tensile stress σ_x plotted against the strain ε_x (Figure 1). It was found that, for most engineering materials subjected to an infinitesimal strain, a linear relation between the stress and strain holds, according to Equation (4):

$$\sigma_x = E\varepsilon_x \tag{4}$$

where E is a material constant called Young's modulus. Equation (4) is called Hooke's law. For an isotropic elastic material, two material constants are needed to characterize the stress–strain relationship: Young's modulus, which is the slope of the σ_x versus ε_x line in Figure 1; and Poisson's ratio, as shown in Equation (5):

$$\nu = -\frac{\text{Transverse strain}}{\text{Longitudinal strain}} = -\frac{\varepsilon_y}{\varepsilon_x} \tag{5}$$

Hence, the slope of ε_y versus ε_x is $-\nu$. Because ε_y is negative, ν is positive. Substituting $\varepsilon_x = -\varepsilon_y/\nu$ from Equation (5) into $E = \sigma_x/\varepsilon_x$ gives Equation (6):

$$\varepsilon_y = -\frac{\nu\sigma_x}{E} \tag{6}$$

To illustrate shear deformation, consider a rectangular block subjected to a balanced pair of forces F (Figure 2). If the base of the solid is fixed, the transverse force F applied causes a distortion, as shown in Figure 2. In this case the angle α can be taken as a strain measure. For a Hookean material, we have (Equation 7):

$$\tau = \frac{F}{A_0} = G\tan\alpha \approx G\alpha = G\gamma \tag{7}$$

where G is the shear modulus, γ is the shear strain, τ is the shear stress, and A_0 is the cross-sectional area of the block.

When a hydrostatic pressure $-p$ is applied to a solid of volume V_0, causing a change in volume ΔV, a bulk modulus K can be defined according to Equation (8):

$$K = \frac{-p}{(\Delta V/V_0)} \tag{8}$$

The four material constants $E, \nu, G,$ and K can be related to each other in a simple manner (Equation 9)

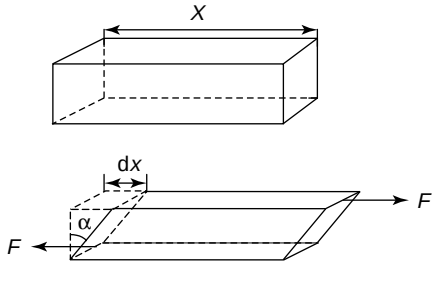

Figure 2 A block under shear.

Table 1 Relation among material constants in isotropic materials

Constants involved	E (Pa)	ν, Dimensionless	G (Pa)	K (Pa)
E, ν, G	$E = 2G(1+\nu)$	$\nu = \dfrac{E-2G}{2G}$	$G = \dfrac{E}{2(1+\nu)}$	
E, ν, K	$E = 3K(1-2\nu)$	$\nu = \dfrac{3K-E}{6K}$		$K = \dfrac{E}{3(1-2\nu)}$
E, G, K	$E = \dfrac{9GK}{3K+G}$		$G = \dfrac{3EK}{9K-E}$	$K = \dfrac{GE}{3(3G-E)}$
ν, G, K		$\nu = \dfrac{3K-2G}{2(3K+G)}$	$G = \dfrac{3K(1-2\nu)}{2(1+\nu)}$	$K = \dfrac{2G(1+\nu)}{3(1-2\nu)}$

because an isotropic solid has only two independent material constants. For example, we have:

$$K = \frac{E}{3(1-2\nu)} \qquad G = \frac{E}{2(1+\nu)} \tag{9}$$

For reference, Table 1 gives all possible equations relating the four constants E, ν, G, and K.

Now consider the general state of stress at a point, as illustrated in Figure 3. A complete description consists of normal stress in three directions, σ_x, σ_y, σ_z, and shear stresses on three planes, τ_{xy}, τ_{xz}, and τ_{yz}. The normal stress is considered first. If it is assumed that small strain theory applies, the strain caused by each component of stress can simply be added together. A stress in the x-direction causes a strain in the x-direction of σ_x/E. This σ_x also causes a strain in the y-direction (from Equation 6) of $-\nu\sigma_x/E$ and the same strain in the z-direction. Similarly, stresses in the y- and z-directions each cause strain in all three directions.

The situation can be summarized in Table 2. Adding the columns in this table to obtain the total strain in each direction gives Equation (10):

$$\varepsilon_x = \frac{1}{E}[\sigma_x - \nu(\sigma_y + \sigma_z)] \quad \varepsilon_y = \frac{1}{E}[\sigma_y - \nu(\sigma_x + \sigma_z)]$$
$$\varepsilon_z = \frac{1}{E}[\sigma_z - \nu(\sigma_x + \sigma_y)] \tag{10}$$

The shear strains that occur on the orthogonal planes are each related to the corresponding shear stress by the shear modulus G, according to Equation (11):

$$\gamma_{xy} = \frac{\tau_{xy}}{G} \quad \gamma_{xz} = \frac{\tau_{xz}}{G} \quad \gamma_{yz} = \frac{\tau_{yz}}{G} \tag{11}$$

Note that the shear strain on a given plane is unaffected by the shear stresses on other planes. Hence, for shear strain, there is no effect analogous to Poisson contraction. Equations (10) and (11) are called the generalized Hooke's law.

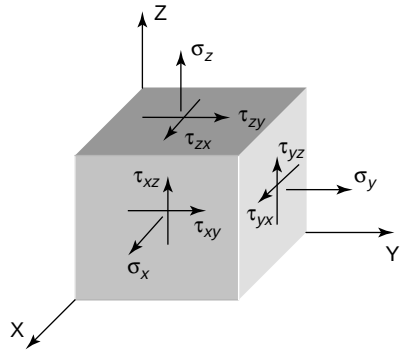

Figure 3 The six components needed to describe the state of stress.

Table 2 Stress–strain relations

Stress	Resulting strain in each direction		
	x	y	z
σ_x	$\dfrac{\sigma_x}{E}$	$-\dfrac{\nu\sigma_x}{E}$	$-\dfrac{\nu\sigma_x}{E}$
σ_y	$-\dfrac{\nu\sigma_y}{E}$	$\dfrac{\sigma_y}{E}$	$-\dfrac{\nu\sigma_y}{E}$
σ_z	$-\dfrac{\nu\sigma_z}{E}$	$-\dfrac{\nu\sigma_z}{E}$	$\dfrac{\sigma_z}{E}$

2.2 Rubber Elasticity

Hooke's law describes the linear stress–strain relation limited to small deformation. However, many polymers experience large deformation, which gives rise to nonlinear stress–strain responses. One example is elastomers. Elastomers are typified by natural rubber but also include a variety of synthetic polymers with similar mechanical behavior. In fact, most noncrystalizable polymers with reasonably high molecular weight are capable of being obtained in the rubber-like state. At low temperatures, the polymer is glassy with a high modulus ($\sim 10^9$ Pa). The modulus then falls through the glass transition region where the polymer is viscoelastic and the modulus becomes very rate- and temperature-dependent. At

a sufficiently high temperature the polymer becomes rubbery. However, all rubbery polymers do not show useful elastomer properties because they will flow irreversibly on loading unless they are cross-linked. The modulus of a linear polymer drops effectively to zero at a sufficiently high temperature when it behaves like a liquid. For cross-linked polymer the modulus will reach a plateau at about 10^6 Pa as the temperature increases.

Cross-linked rubbers possess the remarkable ability of being able to stretch to up to ten times their initial length and then retract rapidly to near their original dimensions when the force is removed. The molecular mechanism responsible for rubber elasticity is based on changes in chain conformation brought about by the overall strain (Figure 4). Clearly, the number of possible chain conformations must be fewer in case (c) than in case (a), resulting in a reduction of entropy.[2] The decreased conformational entropy results in a positive free energy for the stretched state, providing a driving force that causes the material to snap back when the stretching force is released. In this respect, elastomers differ fundamentally from metal springs. The latter develop retractive forces primarily because of increased interatomic distances, which correspond primarily to a change in enthalpy rather than entropy. Thus, in terms of free energy, the entropy term ΔS is primarily responsible for the free energy change ΔF in elastomers, according to Equation (12):

$$\Delta F = \Delta H - T\Delta S \tag{12}$$

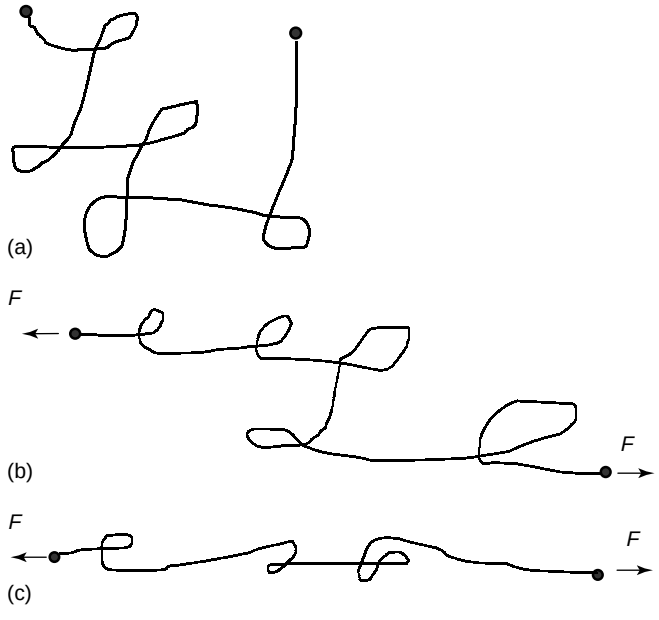

Figure 4 Illustration of entropy reduction on application of force F to network chain: (a) chain at equilibrium; (b) chain as force is being applied; (c) final state with force applied.

whereas the enthalpy term ΔH is of prime importance in metal springs. The fundamental difference between elastomers and metal springs results in a quite different dependence of mechanical behavior on temperature. For elastomers, Equation (13) gives the temperature dependence of the shear modulus as

$$G = nRT = \frac{\rho RT}{M_c} \tag{13}$$

where n represents the number of network chains per unit volume, ρ is the density and M_c is the number-averaged molar mass of the chain length between cross-links. Equation (13) is analogous to the equation of state for ideal gases, $PV = nRT$, which is also based solely on entropy consideration. Neither theory provides any details of chemical structure, but rather involves the modes of molecular motion.

As shown in Equation (13), the shear modulus G increases as the length of chain between the cross-links is reduced. This means that the material becomes stiffer as the cross-link density increases and the network becomes tighter. Equation (13) also shows a surprising phenomenon that has been substantiated experimentally. The modulus of an elastomer increases with temperature, which is in contrast with almost every other material. The modulus increases with temperature because of the increased thermal or Brownian motion, which causes the stretched molecular segments to tug at their ''anchor points'' and try to assume a more probable coiled-up shape. This is a direct consequence of the deformation of elastomers being dominated by changes in entropy rather than internal energy.

The theory of rubber elasticity leads to a nonlinear stress–strain expression (Equation 14) for unidirectional stretching:

$$\sigma = C_0 \left(\lambda - \frac{1}{\lambda^2} \right. \tag{14}$$

where σ is the stress, $\lambda = L/L_0$ is the ratio of the final to the original length, called the stretch ratio, and C_0 is a constant. The stretch ratio λ is a different strain measurement of materials, and is related to the strain defined in Equation (2), i.e. $\lambda = 1 + \varepsilon$.

Equation (14) is valid for small extensions only as shown by Flory[2] and Treloar[3] through experiments. However, it suffers from the same defects as afflict the ideal gas equation, which makes no allowance for such factors as molecular attractive forces or spatial considerations. An important semi-empirical improvement is given by Equation (15), the Mooney–Rivlin equation:[4–6]

$$\sigma = C_1 \left(\lambda - \frac{1}{\lambda^2} \right. + C_2 \left(1 - \frac{1}{\lambda^3} \right. \tag{15}$$

Table 3 Mooney–Rivlin constants

Elastomer	Natural rubber	Butyl rubber	Styrene–butadiene rubber	Ethene–propene rubbers	Polyacrylate rubbers	Silicon rubbers	Polyurethanes
C_1	2.0	2.6	1.8	2.6	1.2	0.75	3
(10^5 N m^{-2})	(0.9–3.8)	(2.1–3.2)	(0.8–2.8)	(2.1–3.1)	(0.6–1.6)	(0.3–1.2)	(2.4–3.4)
C_2	1.5	1.5	1.1	2.5	2.8	0.75	2
(10^5 N m^{-2})	(0.9–2)	(1.4–1.6)	(1.0–1.2)	(2.2–2.9)	(0.9–4.8)	(0.3–1.1)	(1.8–2.2)

where C_1 and C_2 are empirical constants whose values are given in Table 3 for several elastomers.[7]

In the early 1970s Ogden[8] extended Mooney–Rivlin's model and proposed a more general form for the strain energy (Equation 16) through which the stress–strain relation can be derived:

$$W = \sum_{i=1}^{I} \frac{2\mu_i}{\alpha_i}(\lambda_1^{\alpha_i} + \lambda_2^{\alpha_i} + \lambda_3^{\alpha_i}) \qquad (16)$$

where λ_1, λ_2 and λ_3 are three principal stretch ratios representing general three-dimensional deformation, and I is the number of terms used in the series defined by constants μ_i and α_i to fit experimental data. A nonlinear stress–strain relation is obtained by equating the stresses to the derivatives of the strain energy in Equation (16) with respect to the stretch ratios. Ogden,[8] as well as Treloar,[9] Twizell and Ogden,[10] have demonstrated that Equation (16) is capable of representing elastomers under large, general three-dimensional deformation. The Ogden equation (Equation 16) has also been modified[11,12] to model compressible rubber (such as foam).

3 TIME- AND TEMPERATURE-DEPENDENT BEHAVIOR OF POLYMERS

3.1 Viscoelastic Model

A distinctive feature of the mechanical behavior of polymers is that the way in which they respond to an applied stress depends very much on the rate and duration of loading. This is in direct contrast to the behavior of elastic solids such as metals and ceramic, for which, at least at low temperatures and small deformation, the stress is proportional to the strain and independent of loading rate. The behavior of most polymers can be considered as being somewhere between that of elastic solids and liquids. At low temperature and high strain rate they display elastic behavior, whereas at high temperature and low strain rate they behave in a viscous manner. Polymers are therefore termed viscoelastic because they exhibit aspects of both viscous and elastic types of behavior. Because of their chain-like structure, polymers are not perfectly elastic bodies and deformation is accompanied by a complex series of long- and short-range cooperative molecular rearrangements. Consequently, the mechanical behavior is dominated by viscoelastic phenomena, in contrast to materials such as metal and glass where atomic adjustments under stress are more localized and limited. Some polymers show viscoelastic behavior in a striking way. For example, silicon putty flows when it is left alone but bounces like a rubber ball when loaded quickly.

The early work on viscoelasticity was performed on silk, rubber, and glass, and it was concluded that these materials exhibited a ''delayed elasticity''. This delayed elasticity is manifested in the observation that the imposition of a stress resulted in an instantaneous strain that continued to increase more slowly with time. It is this delay between cause and effect that is fundamental to the observed viscoelastic response. The three major examples of this hysteresis effect are: creep, where there is a delayed strain response after the rapid application of a stress; stress relaxation, in which the material is quickly subjected to a strain and a subsequent delay of stress is observed; and dynamic response of polymer to the imposition of a steady sinusoidal stress, which produces a strain oscillating with the same frequencies but out of phase with the stress.

To be of any practical use, an object made from a polymer material must be able to retain its shape when subjected to a small force over a long period of time. This dimensional stability is an important consideration in choosing a polymer for a polymer product. No one wants a plastic telephone receiver that sags after sitting in its cradle for several weeks. Creep can be considered as a progressive increase in strain, observed over an extended time period, in a polymer subjected to a constant stress. Stress relaxation involves the decay of stress required to maintain a constant deformation.

Mechanical models are often used to discuss the viscoelastic behavior of polymers. In Figure 5 are shown three mechanical models of linear viscoelastic behavior, namely: the Maxwell model, the Voigt model, and the ''standard linear'' model, all of which are composed of combinations of linear springs with spring constant E and dashpots with coefficient of viscosity η. A linear spring

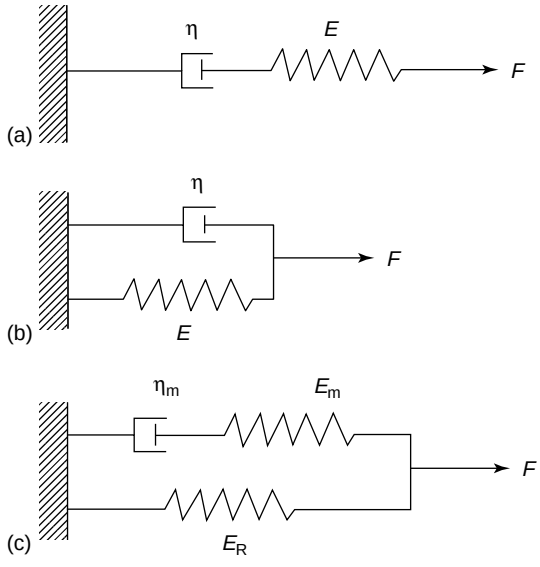

Figure 5 Models of linear viscoelasticity: (a) Maxwell; (b) Voigt; (c) standard linear solid.

produces instantaneous deformation proportional to the load. A dashpot produces a velocity proportional to the load at any instant. Under creep conditions, a constant load is applied to the polymer, and the strain response of the material for the Maxwell model is given according to Equation (17):

$$\varepsilon(t) = \left(\frac{1}{E} + \frac{1}{\eta}t \quad \sigma \right. \tag{17}$$

The creep function (compliance function) is defined as the strain under a unit stress, i.e. $\sigma = 1$. For instance, the creep function for the Maxwell model is given by Equation (18):

$$c(t) = \frac{\varepsilon(t)}{\sigma} = \left(\frac{1}{E} + \frac{1}{\eta}t \right. \tag{18}$$

The strain response and creep functions for the Voigt model and the standard linear model are given by Equations (19) and (20), respectively:

$$\varepsilon(t) = \frac{1}{E}(1 - \mathrm{e}^{-(E/\eta)t})\sigma$$

$$c(t) = \frac{\varepsilon(t)}{\sigma} = \frac{1}{E}(1 - \mathrm{e}^{-(E/\eta)t}) \tag{19}$$

$$\varepsilon(t) = \frac{1}{E_\mathrm{R}} \quad 1 - \left(1 - \frac{\tau_\varepsilon}{\tau_\sigma} \quad \mathrm{e}^{-t/\tau_\sigma} \quad \sigma \right.$$

$$c(t) = \frac{\varepsilon(t)}{\sigma} = \frac{1}{E_\mathrm{R}} \quad 1 - \left(1 - \frac{\tau_\varepsilon}{\tau_\sigma} \quad \mathrm{e}^{-t/\tau_\sigma} \right. \tag{20}$$

where $\tau_\varepsilon = \eta_\mathrm{m}/E_\mathrm{m}$ and $\tau_\sigma = \eta_\mathrm{m}(1/E_\mathrm{m} + 1/E_\mathrm{R})$.

Interchanging the roles of stress and strain, we obtain the relaxation function, which is the response stress corresponding to constant strain. The relaxation function $k(t)$ is the stress that must be applied in order to produce a strain that changes at $t = 0$ from zero to unity and remains at unity thereafter. Equations (21–23) give the respective relaxation functions for the Maxwell, Voigt, and standard linear models:

$$\sigma(t) = E\mathrm{e}^{-(E/\eta)t}\varepsilon \quad k(t) = \frac{\sigma(t)}{\varepsilon} = E\mathrm{e}^{-(E/\eta)t} \tag{21}$$

$$\sigma(t) = [\eta\delta(t) + E]\varepsilon \quad k(t) = \frac{\sigma(t)}{\varepsilon} = [\eta\delta(t) + E] \tag{22}$$

$$\sigma(t) = E_\mathrm{R} \quad 1 - \left(1 - \frac{\tau_\sigma}{\tau_\varepsilon} \quad \mathrm{e}^{-t/\tau_\varepsilon} \quad \varepsilon \right.$$

$$k(t) = \frac{\sigma(t)}{\varepsilon} = E_\mathrm{R} \quad 1 - \left(1 - \frac{\tau_\sigma}{\tau_\varepsilon} \quad \mathrm{e}^{-t/\tau_\varepsilon} \right. \tag{23}$$

Here, we have used the symbol $\delta(t)$ to indicate the unit-impulse function that is infinity at $t = 0$ and zero elsewhere.

The creep and relaxation function for these three models are shown in Figures 6 and 7. For the Maxwell solid, a sudden application of a load induces an immediate deflection by the elastic spring, which is followed by "creep" of the dashpot. On the other hand, a sudden

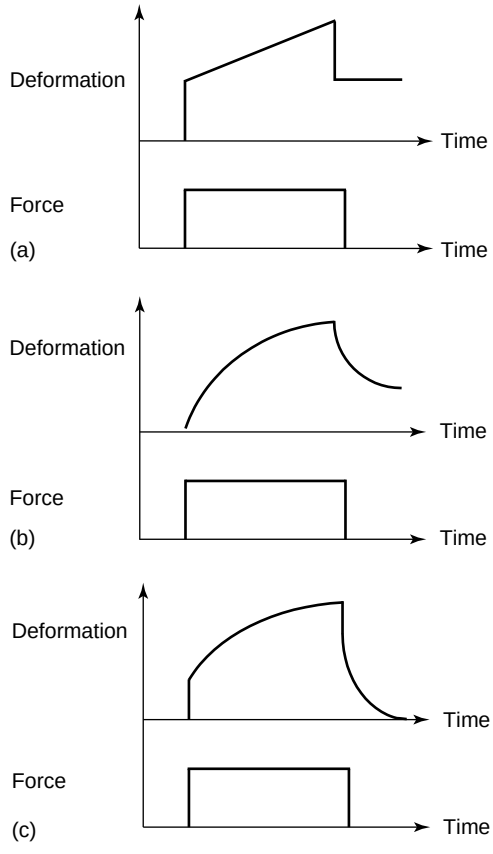

Figure 6 Creep functions of: (a) Maxwell; (b) Voigt; (c) standard linear solid.

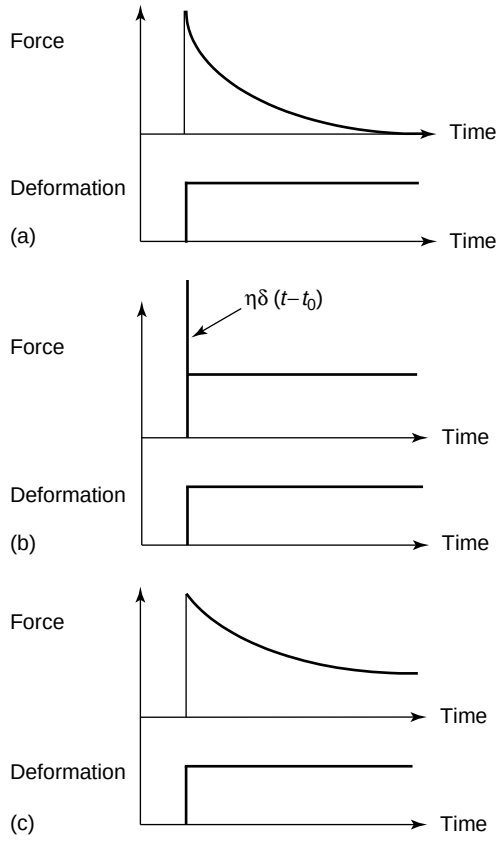

Figure 7 Relaxation functions of: (a) Maxwell; (b) Voigt; (c) standard linear solid.

deformation produces an immediate reaction by the spring, which is followed by stress relaxation according to an exponential law (Equation 21). The factor η/E, with the dimensions of time, may be called the relaxation time. It characterizes the rate of decay of the force.

For a Voigt solid, the sudden application of force will produce no immediate deflection because the dashpot, arranged in parallel with the spring, will not move instantaneously. Instead, as shown by Equation (19) and Figure 6(b), a deformation will be built up gradually while the spring takes a greater and greater share of the load. The dashpot displacement relaxes exponentially. Here, the ratio η/E is again a relaxation time. It characterizes the rate of relaxation of the deflection.

For the standard linear solid, a similar interpretation is applicable. The constant τ_ε is the time of relaxation of load under the condition of constant deflection (see Equation 23), whereas the constant τ_σ is the time of relaxation of deflection under the condition of constant load (see Equation 20). As t approaches infinity, the dashpot is completely relaxed, and the stress–strain relation becomes that of the springs, as is characterized by the constant E_R, in Equations (20) and (23). Therefore, E_R is called the relaxed elastic modulus.

Maxwell introduced the model represented by Equation (17) with the idea that all fluids are elastic to some extent. Lord Kelvin showed the inadequacy of the Maxwell and Voigt models in accounting for the rate of dissipation of energy in various materials subjected to cyclic loading. Kelvin's model is commonly called the standard linear model because it is the most general relationship to include the load, the deflection, and their first (commonly called "linear") derivatives. More general models may be built by adding more and more elements to the Kelvin model. Equivalently, we may add more and more exponential terms to the creep function or to the relaxation function.

The most general formulation under the assumption of linearity between cause and effect is due to Boltzmann. In the one-dimensional case, we may consider a simple bar subjected to a stress $\sigma(t)$ and strain $\varepsilon(t)$. The strain $\varepsilon(t)$ is caused by the total history of the loading up to time t. If the function $\sigma(t)$ is continuous and differentiable, then in a small time interval $d\tau$ at time τ the increment of loading is $(d\sigma/d\tau)\,d\tau$. This increment remains acting on the bar and contributes an element $d\varepsilon(t)$ to the strain at time t, with a proportionality constant c depending on the time interval $t-\tau$. Hence, we may write Equation (24):

$$d\varepsilon(t) = c(t-\tau)\frac{d\sigma(\tau)}{d\tau}\,d\tau \tag{24}$$

Let the origin of time be taken at the beginning of motion and loading. Then, on summing over the entire history, which is permitted under the Boltzmann hypothesis, we obtain Equation (25):

$$\varepsilon(t) = \int_0^t c(t-\tau)\frac{d\sigma(\tau)}{d\tau}\,d\tau \tag{25}$$

A similar argument, with the role of σ and ε interchanged, gives Equation (26):

$$\sigma(t) = \int_0^t k(t-\tau)\frac{d\varepsilon(\tau)}{d\tau}\,d\tau \tag{26}$$

The functions $c(t-\tau)$ and $k(t-\tau)$ are the creep function (compliance function) and relaxation function defined earlier, and are zero when $t-\tau$ is less than zero.

The Maxwell, Voigt, and standard linear solid models are special examples of Boltzmann formulation. More generally, we can write the relaxation function in the form of Equation (27):

$$k(t) = \sum_{n=0}^{N} \alpha_n e^{-\upsilon_n t} \tag{27}$$

which is a generalization of Equation (23). If we plot the amplitude α_n associated with each characteristic frequency υ_n on a frequency axis, we obtain a series of

lines that resemble an optical spectrum. Hence, $\alpha_n(\upsilon_n)$ is called a spectrum of realization function. A generalization to a continuous spectrum may be desired. In some cases, experimental results on relaxation, creep, and hysteresis cannot be reconciled unless a continuous spectrum is utilized.

3.2 Dynamic Mechanical Thermal Analysis

Many experimental techniques are used to measure viscoelastic properties. Measurement of the response in deformation of a material to periodic stress, e.g. during forced vibration, shows that stress and strain are not in phase, i.e. the strain lags behinds the stress by a phase angle δ. Thus, the variation of stress and strain with time can be given by expressions of the type shown in Equation (28):

$$\varepsilon = \varepsilon_0 \sin wt \quad \sigma = \sigma_0 \sin(wt + \delta) \tag{28}$$

which can be rewritten to give Equation (29):

$$\begin{aligned} \sigma &= \sigma_0 \sin wt \cos\delta + \sigma_0 \cos wt \sin\delta \\ &= \varepsilon_0 E_1 \sin wt + \varepsilon_0 E_2 \cos wt \end{aligned} \tag{29}$$

where the dynamic moduli $E_1 = (\sigma_0/\varepsilon_0)\cos\delta$ and $E_2 = (\sigma_0/\varepsilon_0)\sin\delta$ are in phase and 90° out of phase with the strain, respectively; E_1 and E_2 are also called storage and loss moduli. The phase angle δ is obtained according to Equation (30):

$$\tan\delta = \frac{E_1}{E_2} \tag{30}$$

It is more convenient to use complex notation. For such a purpose, let $\varepsilon^* = \varepsilon_0 e^{iwt}$ and $\sigma^* = \sigma_0 e^{i(wt+\delta)}$, then Equation (31) is formed:

$$E^* = \frac{\sigma^*}{\varepsilon^*} = \frac{\sigma_0}{\varepsilon_0} e^{i\delta} = \frac{\sigma_0}{\varepsilon_0}\cos\delta + i\frac{\sigma_0}{\varepsilon_0}\sin\delta = E_1 + iE_2 \tag{31}$$

The shear modulus may be written as a complex modulus in the same way, giving Equation (32):

$$G^* = G_1 + iG_2 \tag{32}$$

where the real and imaginary parts G_1 and G_2 are the storage and loss shear moduli, respectively. The imaginary parts of the complex moduli are damping terms determining the dissipation of energy as heat when the material is deformed; this is why they are called loss moduli. The real parts of the complex moduli are terms determining the amounts of recoverable energy stored as elastic energy; hence, they are called storage moduli.

There are five main experimental approaches for measuring the dynamic mechanical properties of a sample: transient measurements (creep and stress relaxation); low-frequency vibrations (free oscillation methods); high-frequency vibrations (resonance methods); forced vibrations (nonresonance methods); and wave propagation methods. For a full understanding of the viscoelastic behavior of polymers, data are required over a wide range of frequency (or time) and temperature. The number of experiments required can be reduced sometimes by using the equivalency of creep and stress relaxation, the dynamic mechanical data, or the equivalency of time and temperature. Nevertheless, the advantages and limitations of various methods need to be recognized. The approximate frequency scale for each of the five techniques mentioned above is indicated in Figure 8.[13] The first four approaches are reviewed below. The reader is referred to the literature for various wave propagation methods to measure dynamic mechanical properties at kilohertz and megahertz frequency ranges.[14,15]

3.2.1 Transient Measurements: Creep and Stress Relaxation

Because of the simplicity of a creep experiment, a wide variety of experimental techniques have been used in different laboratories, and many have never been described in a publication. The constant force is often applied by a dead weight, and the displacement is measured by a differential transformer, traveling microscope, or deflection of a light beam. The measured displacement gives the strain variation with time and the creep function, $c(t)$, such as that in Equations (18–20) and (25), is obtained.

Any of the experimental techniques for creep experiments can, in principle, be used for studying stress relaxation if the force (stress) can be measured as a function of time while the deformation is held constant. There are, in general, three ways of keeping constant

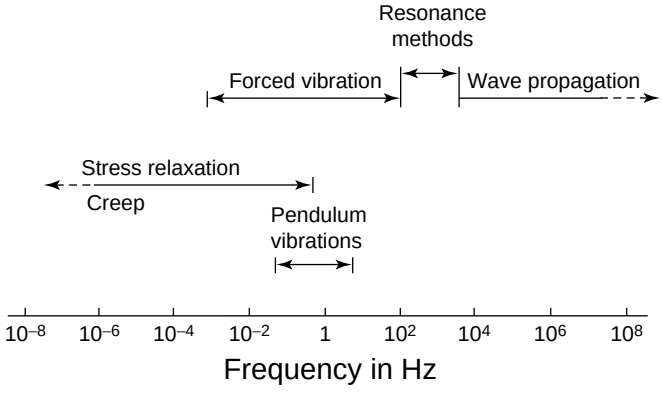

Figure 8 Approximate frequency scales for different experimental techniques.

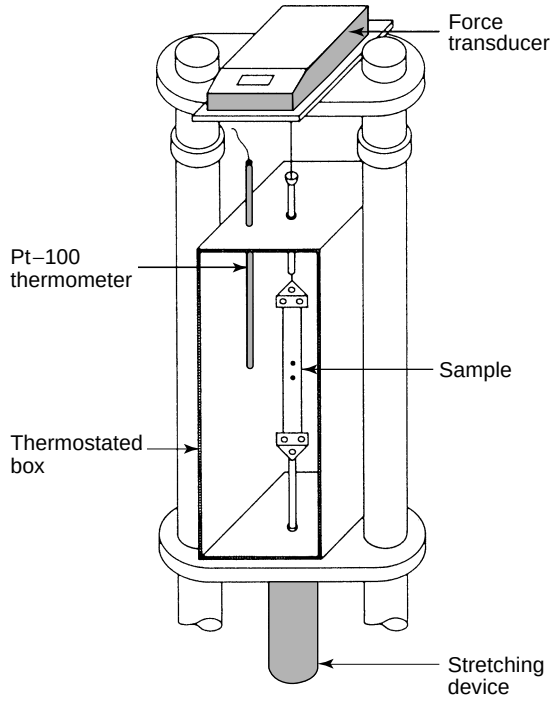

Figure 9 Apparatus for measuring stress relaxation in simple extension. The stretching device can be operated at a controlled rate of extension prior to the relaxation at constant extension. (Reproduced by permission of John Wiley & Sons from J.D. Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, Wiley, New York, 1980.)

strain: by continuous adjustment of the force as the stress in the sample relaxes, which can be done by an automatic servomechanism;[16] by applying more than enough force, with a stop to limit the deformation;[17] and by straining a stiff spring element in series with the sample. The apparatus used for stress relaxation measurements varies in complexity with the physical nature of the sample: simplest for an elastomer and becoming more sophisticated when the polymer is more rigid. One type of experimental set-up[18] is shown in Figure 9, where a force transducer records the stress with an automatic adjustment to zero displacement. Experimental methods and procedures for stress relaxation have been reviewed in detail by Bergen.[19]

3.2.2 Torsional Pendulum: Free Vibration

A study of the mechanical damping and shear modulus under free vibration can be made using a torsional pendulum (Figure 10).[20] The specimen is firmly fixed at one end and the other end is clamped to a disc, with a large moment of inertia, which can move freely. Because the polymer sample should not be under a tensile stress, the suspension wire supporting the disc is passed over a pulley and the weight of the disc and sample are counterbalanced by loading the end. If the disc is subjected to an angular displacement and then released, the sample will twist backwards and forwards about the vertical axis. The oscillations stimulated in the sample are picked up by an arm attached to the rigidly fixed end held in torsion bars, and transmitted to a recorder by a linear variable differential transformer. The sample movements are traced as a series of oscillations whose frequency is a function of the physical state of the sample.

A quantitative measure of the damping is provided by the logarithmic decrement Δ, defined as the logarithmic decrease in amplitude per cycle. It is calculated from the ratio of amplitudes of any two successive oscillations using Equation (33):

$$\Delta = \ln\left(\frac{A_1}{A_2}\right) = \ln\left(\frac{A_2}{A_3}\right) = \cdots = \ln\left(\frac{A_n}{A_{n+1}}\right) \qquad (33)$$

For a specimen with circular cross-section of radius a and length L, the phase angle, storage modulus and loss shear modulus can also be derived from the data[21] according

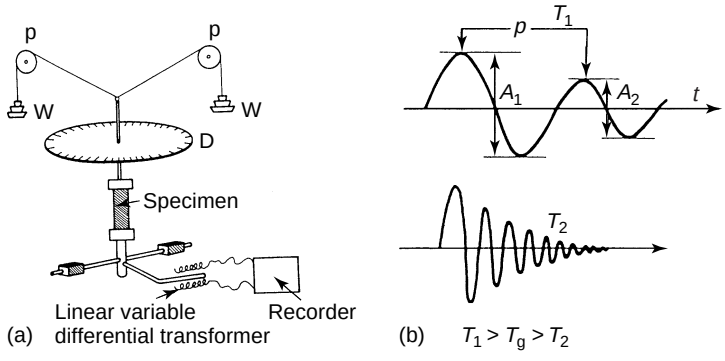

Figure 10 Schematic diagram of a torsion pendulum and the damping of the oscillations for a viscoelastic material. (Reproduced by permission of Stanley Thornes Publishers from J.M.G. Cowie, *Polymers: Chemistry and Physics of Modern Materials*, Blackie Academic & Professional, London, 1977.)

to Equation (34):

$$\tan\delta = \frac{\Delta/\pi}{1-\Delta^2/4\pi^2} \quad G_1 = \frac{2LIw^2}{\pi a^4}\left(1-\frac{\Delta^2}{4\pi^2}\right.$$

$$G_2 = \frac{2LIw^2}{\pi a^4}\frac{\Delta}{\pi} \tag{34}$$

where I is the polar moment of inertia of the disc. Usually Δ is much less than unity so that Equation (34) can be simplified to give Equation (35):

$$\tan\delta = \frac{\Delta}{\pi} \qquad G_1 \approx \frac{2LIw^2}{\pi a^4} = G \tag{35}$$

When the cross-section of the sample is not circular, a geometric correction factor needs to be introduced in the above expressions for G_1, G_2, and δ. The values of the geometric factor for various geometries are given by Rodriguez.[22]

A torsion pendulum is a particularly useful device for examining the modulus and internal friction of polymers. When the sample geometry is in the form of a cone-and-plate or concentric cylinders, the pendulum can be used for very weak gels, which are hard to characterize in conventional testing apparatus. The method can cover the complete range of moduli encountered in polymeric systems but is confined to a relatively narrow frequency range of 0.01–10 Hz.

3.2.3 Vibration Reed: Resonance

This formerly popular industrial method employs a small sample in the form of a thin strip that is clamped firmly at one end, leaving the other end free. The clamped end of the system is then vibrated laterally at a given frequency and the amplitude of the vibration induced at the free end of the sample is recorded. A range of frequencies (200–1500 Hz[23,24]), wide enough to ensure that it encompasses the resonant frequency of the sample f_r, is then examined. The resonant frequency is detected as the maximum of a graph of amplitude against frequency.

A complete treatment of the viscoelastic behavior of the vibrating reed has been given by Bland and Lee.[25] If mechanical loss is small, the modulus E can be obtained from the elasticity solution, and $\tan\delta$ from the resonance curve. For a cantilever of density ρ, length L and thickness h, the resonant frequency of the fundamental mode is given by Equation (36):

$$f_r = \frac{1}{2\pi}\frac{1.875^2}{L^2}\quad\frac{\overline{E}}{\rho}\frac{h}{2\sqrt{3}} \tag{36}$$

from which E is obtained, as shown in Equation (37):

$$E = 38.24\frac{L^4\rho}{h^2}f_r^2 \tag{37}$$

The sensitivity to sample dimensions can give rise to large uncertainty in interspecimen comparisons. This sensitivity is similar to that observed with free torsional vibration.

An analogy between the viscoelastic reed and an equivalent electrical circuit[26] gives Equation (38):

$$\tan\delta = \frac{\Delta f}{f_r} \tag{38}$$

where Δf, the bandwidth, is the difference between the frequencies at which the amplitude of vibration is $1/\sqrt{2}$ of the maximum value (see Figure 11). This technique is not as useful as the torsional pendulum but covers the higher frequency range of 10–1000 Hz.

3.2.4 Forced Vibration: Nonresonance

Several types of instrument can be used for this type of test, and these are usually limited to measurements on rigid polymers or rubbers. One such instrument is shown in Figure 12. The sample (C) is attached firmly at each end to a strain gauge; one of these is a force transducer measuring the applied sinusoidal force and the other records the sample deformation. A sinusoidal tensile stress of a given frequency can be generated in the vibrator (A). Typically, strains are measured by an unbonded strain-gauge transducer. The signals are then fed to a phase meter, which provides a direct reading of the relative amplitudes and the phase difference. This

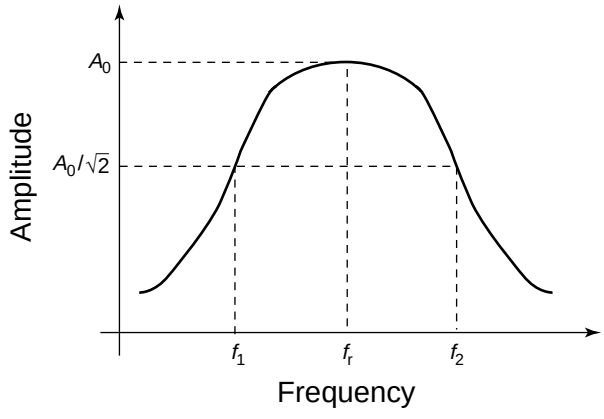

Figure 11 Resonance curve for the vibrating reed.

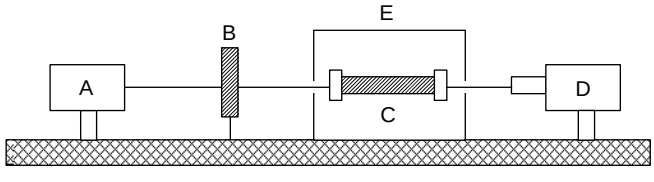

Figure 12 Schematic diagram of apparatus for measuring the dynamic mechanical response using a nonresonance technique.

provides values of the tensile storage and loss moduli as well as the phase angle.

A second version now widely used for these measurements is the Polymer Laboratories DMTA instrument. Several damping arrangements are available for the sample so that measurements may be made in the bending, shear, or tensile modes. In the bending mode the sample, in the form of a small bar, is clamped firmly at both ends and the central point is vibrated by means of a ceramic drive shaft. This can be driven at frequencies selected from the range 0.01–200 Hz. The applied stress is proportional to the AC current fed to the drive shaft and the strain is detected using a transducer that measures the displacement of the drive clamp. Temperature can be controlled over the range 120–770 K, and experiments can be done either isothermally or nonisothermally by ramping the temperatures up and down at various fixed rates.

3.3 Effects of Temperature on Mechanical Properties

In considering the effects of temperature on the mechanical properties of polymers, we will follow closely the discussion by Aklonis and MacKnight.[27] A polymer sample of unit cross-sectional area is subjected to an instantaneous tensile strain that is thereafter maintained constant. The stress is monitored as a function of time, and the tensile stress relaxation modulus is obtained. Let t be any arbitrary time, perhaps 10 s. Next, the stress is removed, allowing the sample to relax, and the temperature is changed. The same experiment is carried out yielding the Young's modulus, E (10 s), at this new temperature. The experiment is repeated at many temperatures to yield the ''ten-second tensile relaxation modulus'' as a function of temperature.

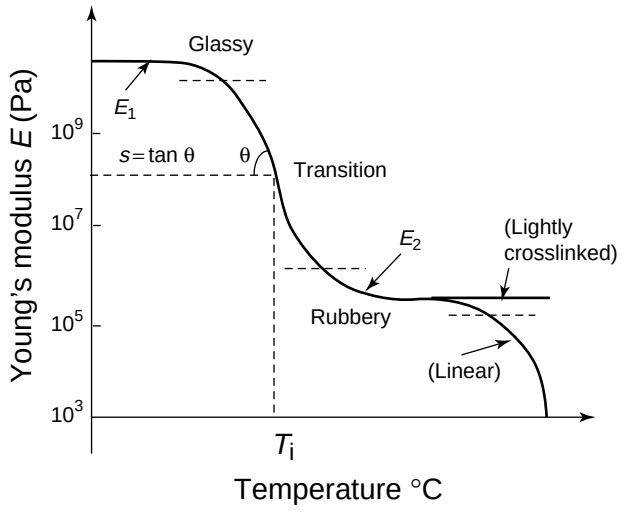

Figure 13 Schematic modulus–temperature curve showing various regions of viscoelastic behavior. (Reproduced by permission of John Wiley & Sons from J.J. Aklonis, W.J. MacKnight, *Introduction to Polymer Viscoelasticity*, 2nd edition, Wiley, New York, 1983.)

The types of behavior most often observed are shown in Figure 13, which shows idealized modulus–temperature curves for typical linear and cross-linked amorphous polymers. In this plot, four regions of viscoelastic behavior are defined. At low temperatures, where the modulus is higher than 10^9 Pa, the polymer is hard and brittle. This is the glassy region. The glassy modulus, E_1, is a slowly decreasing function of temperature and is a useful parameter for characterizing polymeric behavior. In this glassy region, thermal energy is insufficient to surmount the potential barriers for rotational and translational motions of segments of the polymer molecules. The chain segments are essentially ''frozen'' in fixed positions on the sites of a disordered quasi-lattice with their segments vibrating around these fixed positions much like low-molecular-weight molecules in a molecular crystal. With increasing temperature, the amplitude of vibrational motion becomes greater and eventually the thermal energy becomes roughly comparable to the potential energy barriers to segment rotation and translation. In this temperature region, the polymer is at the glass transition temperature, where short-range diffusional motions begin. Segments are free to ''jump'' from one lattice site to another and the brittle glass becomes a resilient leather.

This phenomenon is accompanied by a modulus decrease of several orders of magnitude, as indicated in Figure 13. The breadth of this transition region ranges from 5 °C to more than 20 °C, depending on the nature of the polymer in question. The transition region may be characterized by two parameters. The temperature at which the 10-s modulus reaches 10^8 Pa is termed the inflection temperature, T_i, and the negative slope of the curve ($\tan\theta$) at this point is termed s. Values of these characteristic viscoelastic parameters for several amorphous polymers are listed in Table 4.[27]

As the temperature is increased further, the modulus again reaches a plateau region. This rubbery plateau is characterized by the modulus E_2, as shown in Figure 13. In this temperature interval, the short-range diffusional motions of the polymer segments that initially gave rise to the glass transition occur much faster than the measurement time of 10 s. On the other hand, the long-range cooperative motion of chains that would result in translational motions of complete molecules is still greatly restricted by the presence of strong local interactions between neighboring chains. In the case of the cross-linked material, these interactions consist of primary chemical bonds. In the linear polymer, they are known as entanglements, and their precise nature is not clear. In any case, in the rubbery plateau region, segments

Table 4 Characteristic viscoelastic parameters for selected polymers

Polymer	$3G_1$ (Pa)	$3G_2$ (Pa)	T_i (°C)	s (°C^{-1})
Crystalline polymers				
Polyethylene ($M_w = 2.6 \times 10^5$)	3×10^9	3×10^6	75	0.01
Polyvinyl chloride	4	2	78	0.25
Amorphous linear polymers				
Polyisobutylene	3.5	1	−62	0.15
Natural rubber (unvulcanized)	2.5	4	−67	0.2
Polystyrene	2	0.5	101	0.2
Poly(methyl acrylate)	3	1.5	16	0.2
Poly(methyl methacrylate)	1.5	2	107	0.15
Poly(butyl acrylate)	1.5	0.5	−53	0.2
Poly(butyl methacrylate)	1	1	31	0.15
cis-Polybutadiene	2	1	−106	0.2
Poly(ethylene tetrasulfide)	2	4	−24	0.15
Atactic polypropylene	2	2.5	−16	0.2
Polyacenaphthalene	–	–	264	0.1
Poly(ethylhexyl acrylate)	–	0.5	−70	0.2
Selenium	6	–	45	0.15
Sulfur	–	–	−28	0.2
Bisphenol A polycarbonate ($M_w = 4 \times 10^5$)	1.5	5	150	0.3
Poly(ethyl ether disulfide)	0.5	2	−53	0.25
Poly(ethyl formal disulfide)	1	0.5	−58	0.35
Amorphous lightly cross-linked polymers				
Natural rubber (vulcanized)	3.5	4	−57	0.2
Polytetrahydrofuran	3	10	−73	0.15
Poly(2-hydroxylethyl methacrylate)	3	2.5	96	0.1
Poly(*t*-butylaminoethyl methacrylate)	2.5	1	41	0.1
Poly(ethyl methacrylate)	2.5	2.5	77	0.1
Poly(*n*-propyl methacrylate)	2	2	56	0.1
Poly(methoxyethyl methacrylate)	2	1.5	23	0.1
Poly(diethylaminoethyl methacrylate)	2	0.5	26	0.1

of chains reorient relative to each other but large-scale translational motion does not occur.

The viscoelastic responses of linear and cross-linked polymers through the rubbery plateau region are essentially identical. As the temperature is increased further, however, differences between these two categories of polymers become evident (Figure 13). First consider a cross-linked network. As temperature is increased, the cross-links consisting of primary chemical bonds remain intact, preventing the chains from translating relative to one another. Although the modulus changes slightly with temperature in the rubbery plateau region of a cross-linked polymer, this change is small compared to changes exhibited during the glass transition. Thus, to a first approximation, the modulus will remain constant for a cross-linked rubber up to temperatures where chemical degradation begins to occur. The situation is quite different for a linear polymer. In this case, increasing temperature causes molecular motions to become more and more large scale until eventually whole polymer molecules begin to translate. When the temperature is high enough, local chain interactions are no longer of sufficiently high energy to prevent molecular flow. This area of the 10-s modulus temperature curve is called the flow region.

3.4 Time–Temperature Superposition

It is thought that there is a general equivalency between time (frequency) and temperature. For instance, a polymer that displays rubbery characteristics under a given set of testing conditions can show glass behavior by either reducing the temperature or increasing the testing rate or frequency. Leaderman[28] was the first to suggest that in viscoelastic materials time and temperature are equivalent, meaning that data at one temperature can be superimposed upon data taken at a different temperature merely by shifting curves. Williams et al.,[29] Tobolsky,[30] and Ferry[18] have worked out this suggestion and demonstrated the validity of the principle. With their procedures it is possible to convert short-term stress–relaxation data at widely different temperatures to a single curve covering many decades of time at some reference temperature. This is accomplished by translating the small curves collected at different temperatures along the $\log t$ axis until

they are all superimposed to form a large composite curve spanning a long period of time.

Figure 14[31,32] is chosen as an example showing the curves of polyisobutylene. These curves are first corrected (reduced) for density and temperature. An arbitrary temperature T_s is selected as the reference temperature. The reduced modulus values are calculated using Equation (39):

$$E(t)_{\text{reduced}} = \frac{T_s}{T}\frac{\rho_s}{\rho}E(t) \tag{39}$$

where $E(t) = \sigma(t)/\varepsilon$ is the modulus in a stress relaxation test. The correction comes from the kinetic theory of rubber elasticity; it is relatively small and is sometimes neglected. After this reduction the experimental curves are replotted, as shown on the left side of Figure 14. These reduced curves can now be shifted, one at a time, with respect to the reference curve (at $T_s = 25\,^{\circ}\text{C}$), until portions of the curves superimpose to give a master curve, as shown on the right side of Figure 14. The amount by which each reduced modulus has to be shifted along the logarithmic time axis in making the master curve – the so-called shift factor – is a function of temperature (see upper right corner of Figure 14). The curve obtained at each temperature is shifted by an amount as shown in Equation (40):

$$(\log t - \log t_s) = \log\left(\frac{t}{t_s}\right) = \log a_T \tag{40}$$

The parameter a_T is called the shift factor and is positive if the movement of the curve is to the left of the reference and negative for a move to the right. The shift factor is a function of temperature only and decreases with increasing temperature. It is, of course, unity at the reference temperature, T_s. The superposition principle also applies to creep data.

Williams et al.[29] proposed that a_T could be given by an equation of the form shown in Equation (41):

$$\log a_T = \frac{-C_1(T - T_s)}{C_2 + (T - T_s)} \tag{41}$$

where C_1 and C_2 are constants and T_s is the reference temperature. This is normally called the WLF equation and was originally developed empirically. It holds extremely well for a wide range of polymers in the vicinity of the glass transition. If T_s is taken as T_g – the glass transition temperature measured by a static method such as dilatometry – then, according to Equation (42):

$$\log a_T = \frac{-C_1^g(T - T_g)}{C_2^g + (T - T_g)} \tag{42}$$

and the new constants C_1^g and C_2^g become "universal", with values of 17.4 and 51.6 K, respectively. In fact, the constants vary somewhat from polymer to polymer, but it is often quite safe to assume the universal values because they usually give shift factors that are close to measured values.

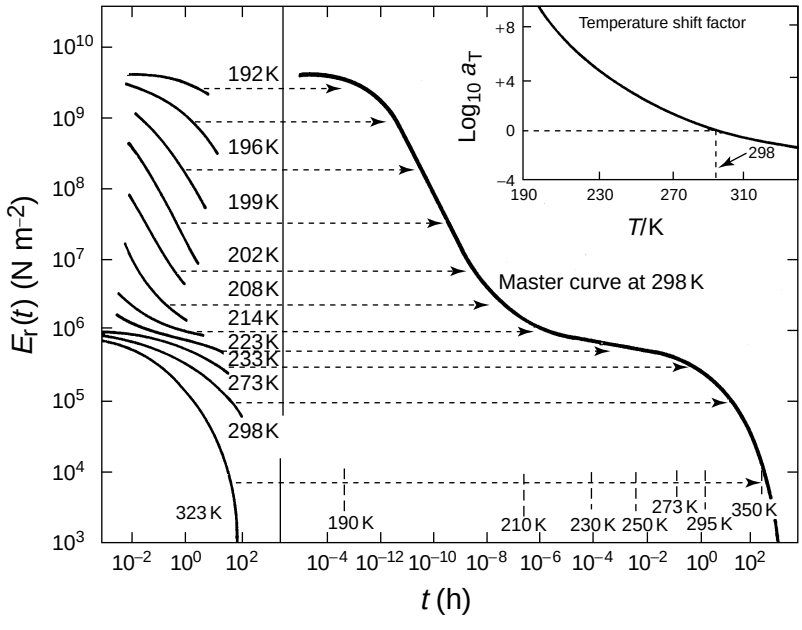

Figure 14 Time–temperature superposition principle illustrated with polyisobutylene data. The reference temperature of the material curve is 25 °C. The inset graph gives the amount of curve shifting required at the different temperature. (Reproduced by permission of Academic Press from E. Catsiff, A.V. Tobolsky, *J. Colloid Sci.*, **10**, 375 (1955).)

The WLF master curve produced by superposing data obtained at different temperatures is a very useful way of presenting the mechanical behavior of a polymer. The WLF equation is also useful in predicting the mechanical behavior of a polymer outside the range of temperature and frequency (or time) for which experimental data are available. Although the WLF equation was developed originally by curve fitting, it is possible to justify it theoretically from considerations of free volume.[29]

4 ULTIMATE PROPERTIES OF POLYMERS AND RUBBERS

4.1 General Stress–Strain Diagram

The strength properties of solids are illustrated most simply by a stress–strain diagram, which describes the behavior of a homogeneous specimen of uniform cross-section subjected to uniaxial tension. Three important types of stress–strain curves are illustrated in Figure 15. The brittle, plastic stress–strain curve is linear up to fracture at about 1–2% elongation. Typical stresses at break are of the order of 6×10^7 Pa. Ordinary polystyrene behaves in this way. An example of a tough plastic is polyethylene, which is semicrystalline, with the amorphous portions above T_g. Its Young's modulus, given by the initial slope of the stress–strain curve, is somewhat lower than that of the brittle, glassy plastic (see Figure 15). Typically, this class of polymer exhibits a yield point, followed by extensive elongation at almost constant stress. This is called the plastic flow region. Stresses in the range of 2–5 $(\times 10^7)$ Pa are commonly exhibited in this range. Extension at constant stress in the plastic flow region is often referred to as cold drawing.

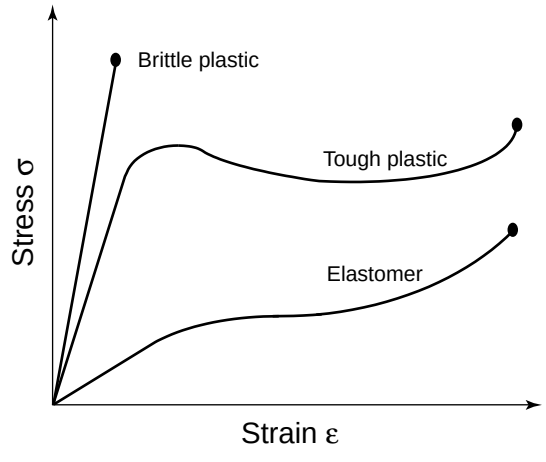

Figure 15 Stress–strain behavior of three types of polymers. (Reproduced by permission of John Wiley & Sons from E. Catsiff, A.V. Tobolsky, *J. Polym. Sci.*, **19**, 111 (1956).)

Finally, the polymer strain hardens and then ruptures. The third type of stress–strain curve is that exhibited by elastomers. The equation of state for rubber elasticity governs here, with its peculiar nonlinear curve.

Elongation to break for both the tough plastic and the elastomer may be of the order of several hundred percent and is indicated by the dot at the end of the curve. For crystallizing elastomers such as natural rubber, the curve swings upward sharply at the point of crystallization, and tensile strengths of 14–34 MPa are common. Noncrystallizing elastomers have much lower tensile strengths, often below 7 MPa. With the addition of reinforcing fillers, such as finely divided carbon black, the tensile strength is much increased. Of course, this latter material is widely used for automobile tires, one of the toughest materials known. The moduli and strengths of some typical polymers are listed in Table 5.[33]

There are several terms in use that describe the "strength" of a polymer. The tensile strength describes the stress to break the material in a simple tension test. The tensile strength is certainly important, but in engineering practice a polymer is rarely stressed so greatly that it breaks immediately. Toughness of the polymer is frequently a more useful parameter. The toughness is measured by the area under the stress–strain curve. This area has the units of energy per unit volume and is the work expended in deforming the material. The deformation may be elastic and recoverable, or permanent (irreversible deformation). Elastic energy is stored in the sample in terms of energy per unit volume.

4.2 Cold Drawing in Crystalline Polymers

The tough plastic in Figure 15 is shown to have a yield point, followed by a region of cold drawing at almost constant stress. There are two basic causes for this phenomenon. First, for rubber-toughened amorphous plastics, the region of cold drawing is where extensive orientation of the chains takes place, accompanied by significant viscoelastic flow. Second, for semicrystalline polymers with amorphous portions above T_g, cold drawing rearrangement of the chains takes place in a characteristic complex manner that begins with necking. A neck is a narrowing down of a portion of the stressed material to a smaller cross-section. This is a form of shear yielding. The neck grows at the expense of the material at either end, eventually consuming the entire specimen.

In the region of the neck, a very extensive reorganization of the polymer is taking place. Spherulites are broken up, and the polymer becomes oriented in the direction of stretch. The number of chain folds decreases, and the number of tie molecules between the new fibrils is increased. The crystallization is usually enhanced by

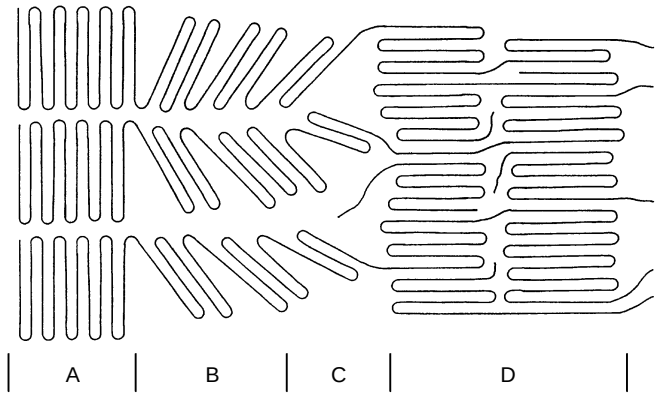

Figure 16 Proposed mechanism of reorientation of crystalline structures during necking. (Reproduced by permission of John Wiley & Sons from A. Peterlin, *J. Polym. Sci.*, **9C**, 61 (1965).)

the chain alignment. At the end of the reorganization, a much longer, thinner, and stronger fiber or film is formed.

This phenomenon is easily demonstrated at home or in the classroom. A strip of polyethylene film, perhaps with print on it, is satisfactory. The strip is slowly stretched between the hands at room temperature, and the neck will form after several percent elongation. Such a material can then be elongated by 2–5 times its original length. Draw ratios can be significantly higher under more scientific conditions.

The molecular reorganization illustrated in Figure 16[34] has attracted the attention of the theoreticians.[35,36] Does the polymer actually melt under stress, and then reorganize, or do the crystals themselves rotate? At this point the evidence seems to be divided between the two possibilities.

It must be pointed out that drawing operations are critically important in the manufacture of fibers such as nylon and rayon. In the case of nylon, the polymer is spun in the melt state. The fiber is then cooled until it crystallizes, and finally stretched with a draw ratio of 4 : 8.

The case of rayon is slightly complicated because an alkaline solution of sodium cellulose xanthate is spun. The nascent fiber is spun into an acid bath, which removes the xanthate groups, thus precipitating the polymer. As it precipitates, it crystallizes. The fiber is then stretched

Table 5 Mechanical properties of polymers (unmodified) (moduli in 10^9 Pa; strengths in 10^7 Pa)

Polymer	Tensile strength at yield	Elongation at yield (%)	Tensile strength at break	Elongation at break (%)	Tensile modulus	Flexural strength	Flexural modulus	Compressive strength	Poisson's ratio
1 Polyethylene (LD)	0.8	20	1.0	800	0.2				0.49
2 Polyethylene (HD)	3.0	9	3.0	600	1	4.5	0.8	2	0.47
3 Polypropylene	3.2	12	3.3	400	1.4	4.9	1.5	4.5	0.43
4 Poly(1-butene)			3.0	350	0.75				0.47
5 Polystyrene			5.0	2.5	3.4	8	3.3	9.5	0.38
6 Poly(vinyl chloride)	4.8	3	5.0	30	2.6	9	3.5	7	0.42
7 Poly(chlorotrifluoro-ethylene)	3.0	10	3.5	175	1.9	5.5	2	4	0.44
8 Poly(tetrafluoroethylene)	1.3	62.5	2.5	200	0.5		0.35	0.8	0.46
9 poly(methyl methacrylate)			6.5	10	3.2	11	3	10.5	0.40
10 Poly(methylene oxide)			6.5	40	2.7		2.5	12	0.44
11 Poly(phenylene oxide)			6.5	75	2.3				0.41
12 Poly(phenylene sulfide)			6.5	3	3.4	11			
13 Poly(ethylene terephthalate)		6	5.4	275	3.0		2.9	9	0.43
14 Poly(tetramethylene terephthalate)			5.0		2.5			8	0.44
15 Nylon 66	5.7	25	8.0	200	2.0		2.3	10	0.46
16 Nylon 6	5.0	30	7.5	300	1.9		2.0	9	0.44
17 Poly(bisphenol carbonate)	6.5	30	6.0	125	2.5	9	2.5	7	0.42
18 Polysulfone			6.5	75	2.5	10		8	0.42
19 Polyimide			7.5	7	3.0	10			0.42
20 Cellulose acetate	4	6	3	30	2	5	1.25		
21 Phenol formaldehyde resin			5.5	1	3.4	9	4	13	0.35
22 Uns. polyester resin			6.0	3	5.0	9	5.0	15	0.36
23 Epoxy resin			5.5	5	2.4	11	2.5	13	0.37

HD = high density; LD = low density.

wet, approximately 4–7 times its original length. In both cases, the fibers are highly oriented, as shown by X-ray studies. The high degree of orientation contributes to the high strength of such fibers.

4.3 Temperature and Pressure Effects on Yield Stress

For polymers, rather unusual features are revealed when the yield stress of the same polymer is measured in simple uniaxial tension and compression. Conventional yielding criteria such as the Tresca and von Mises criteria[21] predict that the yield stress should be the same in both cases. This is what is found for metals, but for polymers the compressive yield stress is usually higher than the tensile yield stress. This difference between the compressive and tensile yield stress can be taken as an indication that the hydrostatic component of the applied stress is exerting an influence upon the yield process. This can be demonstrated more directly by measuring the tensile yield stress under the action of an overall hydrostatic pressure. The results of such measurements on different polymers are shown in Figure 17.[37] It can be seen that there is a clear increase in the yield stress with hydrostatic pressure. In general, it is found that the yield stresses of amorphous polymers show larger pressure dependence than those of crystalline polymers. There is no significant change in the yield stresses of metals at similar pressures, but it is quite reasonable to expect that the physical properties of polymers will change at these relatively low pressures. They tend to have much lower bulk moduli than metals and so undergo significant volume changes on pressurization.

Plastic deformation, like other aspects of the mechanical behavior of polymers, has a strong dependence upon the testing rate and temperature. Typical behavior is illustrated in Figure 18[21] for a glassy thermoplastic deformed in tension. At a given strain rate the yield stress drops as the temperature is increased, and the yield stress falls approximately linearly to zero at the glass transition temperature when the polymer glass becomes a rubber. If the strain rate is increased and the temperature held constant, the yield stress increases.

The behavior of semicrystalline polymers is rather similar, with the main difference being that the yield stress drops to zero at the melting temperature of the crystals rather than at the T_g (glass transition temperature). Between T_g and T_m (melting temperature), the noncrystalline areas are rubbery and the material gains its strength from the crystalline regions that reinforce the rubbery matrix. If the temperature of a semicrystalline polymer is reduced below T_g, it behaves more like a glassy polymer and the crystals do not have such a significant strengthening effect.

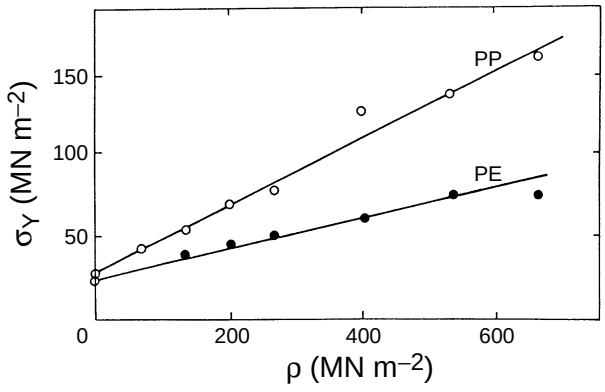

Figure 17 Variation of yield stress of polyethylene and polypropylene with hydrostatic pressure p. (Reproduced by permission of the American Institute of Physics from D.R. Mears, K.D. Pae, J.A. Sauer, *J. Appl. Phys.*, **40**, 4229 (1969).)

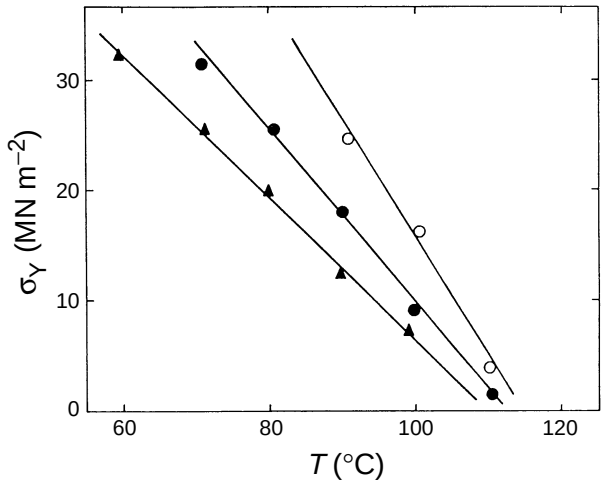

Figure 18 Variation of the yield stress of poly(methyl methacrylate) with temperature. Strain rates are ( ) 0.002 min^{-1}; (●) 0.02 min^{-1}; (○) 0.2 min^{-1}. (Reproduced by permission of Stanley Thornes Publishers from R.J. Young, P.A. Lovell, *Introduction to Polymers*, 2nd edition, Chapman & Hall, London, 1995.)

4.4 The Brittle–Ductile Transition

Many materials are ductile under certain testing conditions, but when these conditions are changed, e.g. by reducing the temperature, they become brittle. This type of behavior is also encountered in steels and is very common with polymers. In this section, we follow to a large extent the discussion given by Young and Lovell.[21] Figure 19(a)[21] shows how the strength of poly(methyl methacrylate) (PMMA) varies with testing temperature. At low temperatures the material fails in a brittle manner dictated by brittle break strength σ_b. When the testing temperature is raised to just above room temperature the polymer undergoes general yielding (indicated by yielding strength σ_Y) and brittle fracture is suppressed. This type

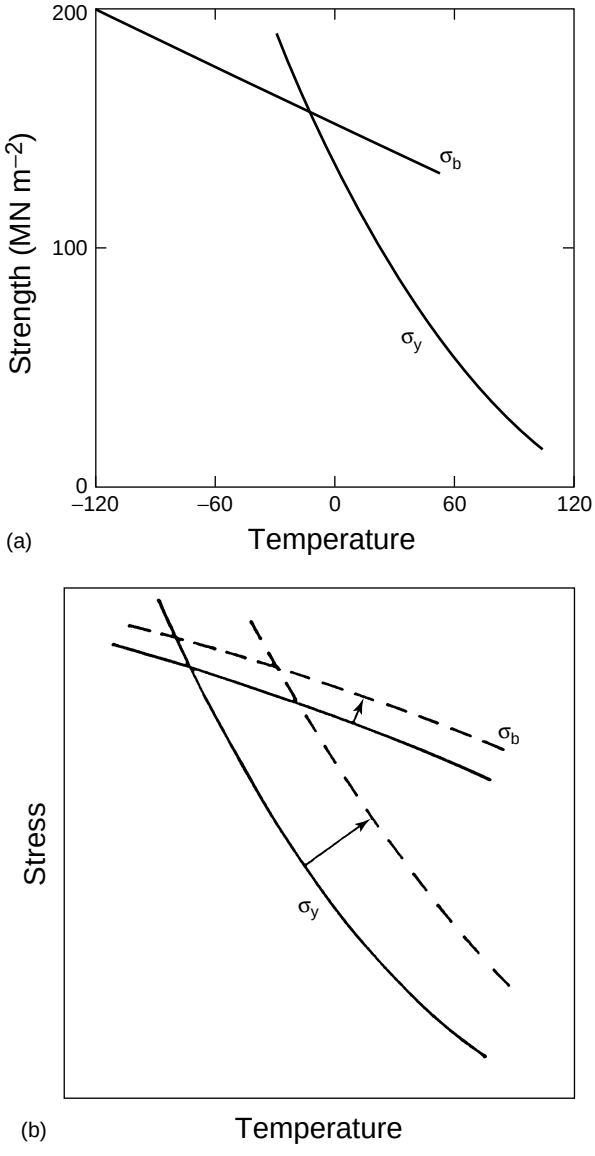

Figure 19 Ductile to brittle transition: (a) variation with temperature of the brittle strength measured in flexure (σ_b) and yield strength measured in tension (σ_y) for PMMA; (b) effect of strain rate on the transition. Curves for a higher strain rate are drawn as dashed lines. (Reproduced by permission of Stanley Thornes Publishers from R.J. Young, P.A. Lovell, *Introduction to Polymers*, 2nd edition, Chapman & Hall, London, 1995.)

of brittle–ductile transition has been explained in terms of brittle fracture and plastic deformation being independent processes that have a different dependence on temperature, as shown schematically in Figure 19(b).[21] The yield stress increases more rapidly as the temperature is reduced, and at a critical temperature it becomes higher than the stress required to cause brittle fracture. The temperature at which this occurs corresponds to that of the brittle–ductile transition, T_b, and it is envisaged that the process that occurs during deformation is that which can take place at the lowest stress, i.e. brittle fracture at $T < T_b$ and general yielding at $T > T_b$.[21]

There is an added complication for polymers, i.e. the strong dependence of the mechanical properties upon strain rate as well as upon temperature. For example, nylons can be cold-drawn at room temperature when relatively low strain rates are used but they become brittle as the strain rate is increased. It is thought that this is caused by a strain-rate-induced shift (to the right in Figure 19b) of the yield stress and brittle stress curves. This results in a temperature increase of the brittle–ductile transition.

It is known that the temperature at which the brittle–ductile transition occurs for a particular polymer is sensitive to both the structure of the polymer and the presence of surface flaws and notches. The temperature of the transition is generally raised by cross-linking and increasing the crystallinity. Both of these factors tend to increase the yield stress without affecting the brittle stress significantly. The addition of plasticizers to a polymer has the effect of reducing yield stress, thus lowering T_b. Plasticization is widely used in order to toughen polymers such as poly(vinyl chloride), which would otherwise have a tendency to be brittle at room temperature.

The presence of surface scratches, cracks or notches can have the effect of both reducing the tensile strength of the material and also increasing the temperature of the brittle–ductile transition. In this way the presence of a notch can cause a material to fail in a brittle manner at a temperature at which it would otherwise be ductile. As one might expect, this phenomenon can play havoc with any design calculations that assume the material to be ductile. This type of ''notch sensitivity'' is normally explained by examining the state of stress that exists at the root of the notch. When an overall uniaxial tensile stress is applied to the body there will be a complex state of triaxial stress at the tip of the crack or notch. This can cause a large increase in the stress of the material at the crack tip. As shown in Figure 20,[23] the brittle fracture stress will be unaffected and so the overall effect is to raise T_b. If the material is used between T_b for the bulk and T_b for the notched specimen, it will be prone to suffer from ''notch brittleness''.

4.5 Crazing

Crazing involves the formation of an intricate network of fibrils connecting the upper and lower surfaces of the craze. The craze itself forms at right angles to the applied stress. Shear yielding, on the other hand, involves molecular slip, usually at 45° to the applied stress, assuming uniaxial stress–strain relationships. If the polymer crazes, it usually fails after 1% or 2% extension and is said to be brittle. If it undergoes yielding, the

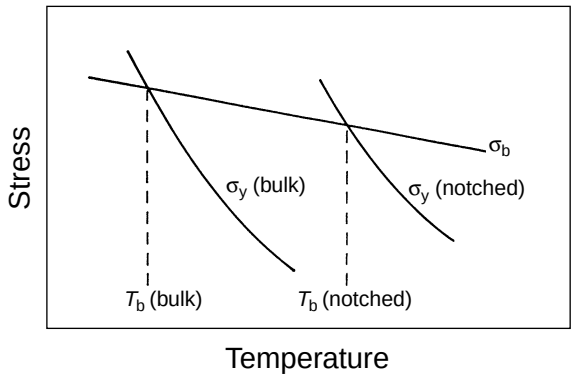

Figure 20 Schematic representation of the effect of notch on the ductile to brittle transition. (Reproduced by permission of John Wiley & Sons from I.M. Ward, D.W. Hadley, *An Introduction to the Mechanical Properties of Solid Polymers*, John Wiley & Sons, Chichester, 1996.)

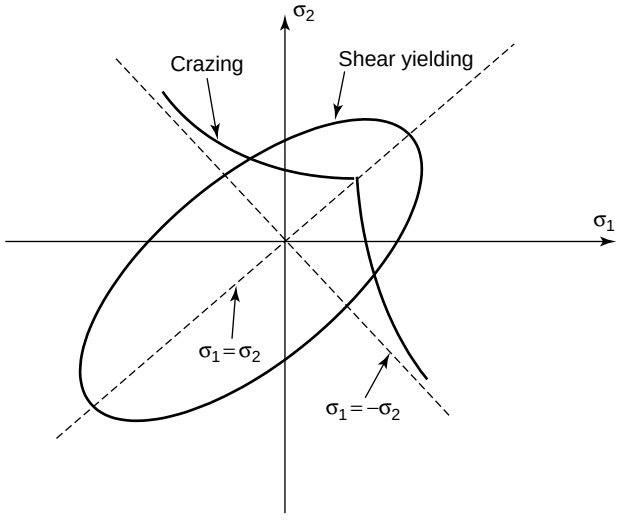

Figure 21 Failure envelope for PMMA under biaxial stress at room temperature. (Reproduced by permission of the American Chemical Society from S.S. Sternstein, L. Ongchin, *Polym. Prepr.*, **10**(2), 1117 (1969).)

polymer usually can be stretched by at least 10% or 20% and is said to be ductile.

What criteria determine which mechanism will prevail? Polymers sometimes change from brittle to ductile on raising the temperature, as discussed earlier. Similarly, if the polymer is tested under compression, or under hydrostatic pressure, it tends to fail by shear yielding. Sternstein and Ongchin[38] studied the formation of crazes under biaxial stress conditions, and found that the stress conditions for crazing involved both principal stresses. They calculated the biaxial stress envelopes for PMMA. Figure 21 shows separately the envelopes for craze initiation and shear yielding. In the first quadrant of stress space, the crazing envelope is everywhere inside the shear yielding envelope, which implies that all combinations of tensile biaxial stress produce crazes prior to shear yielding. When the sample is under compression, third-quadrant yielding will always take place. In the second and fourth quadrants, the two envelopes intersect each other. Generally, the failure mechanism that takes place is the one requiring the lower stress. The 45° line running through the second and fourth quadrants represents pure shear deformation, marking the boundary between hydrostatic compression and hydrostatic tension. Crazing does not take place below this line because the pressure component of the stress tends to reduce rather than increase the volume.

The most physically acceptable explanation of Sternstein and Ongchin's results was proposed by Powden and Oxborough,[39] who suggested that crazing occurs when the extensional strain in any direction reaches a critical value. Unfortunately, there are several pieces of experimental evidence[40–42] that contradict this assumption, so there is still no completely satisfactory stress criterion for craze initiation.

There is, however, a theory for the growth of crazes, which is consistent with all the experimental evidence. Argon et al.[43] have proposed that the craze front advances by a meniscus instability mechanism in which craze tufts are produced by the repeated break-up of the concave air/polymer interface at the crack tip (as illustrated in Figure 22). A theoretical treatment of this model predicted that the steady-state craze velocity would relate to the five-sixths power of the maximum principal tensile stress, and support for this result was obtained from experimental results on polystyrene and PMMA.[43] More discussions on crazing can be found in Kausch.[44]

4.6 Fracture

Modern understanding of the fracture behavior of brittle materials stems from the seminal research of Griffith[45] on the brittle fracture of glass. The Griffith theory of fracture, which is the earliest statement of linear elastic fracture mechanics, has been applied extensively to the fracture of glass and metals, and more recently to polymers. The Griffith's criterion was initially conceived to describe the propagation of a crack in a perfectly elastic material at small elastic strains (hence, linear elastic). Subsequent work has shown that it is still applicable for situations including localized plastic deformation at the crack tip, which does not lead to general yielding in the specimen.

Griffith considered that fracture produces a new surface area and postulated that for fracture to occur the increase in energy required to produce the new surface must be balanced by a decrease in elastically stored energy.

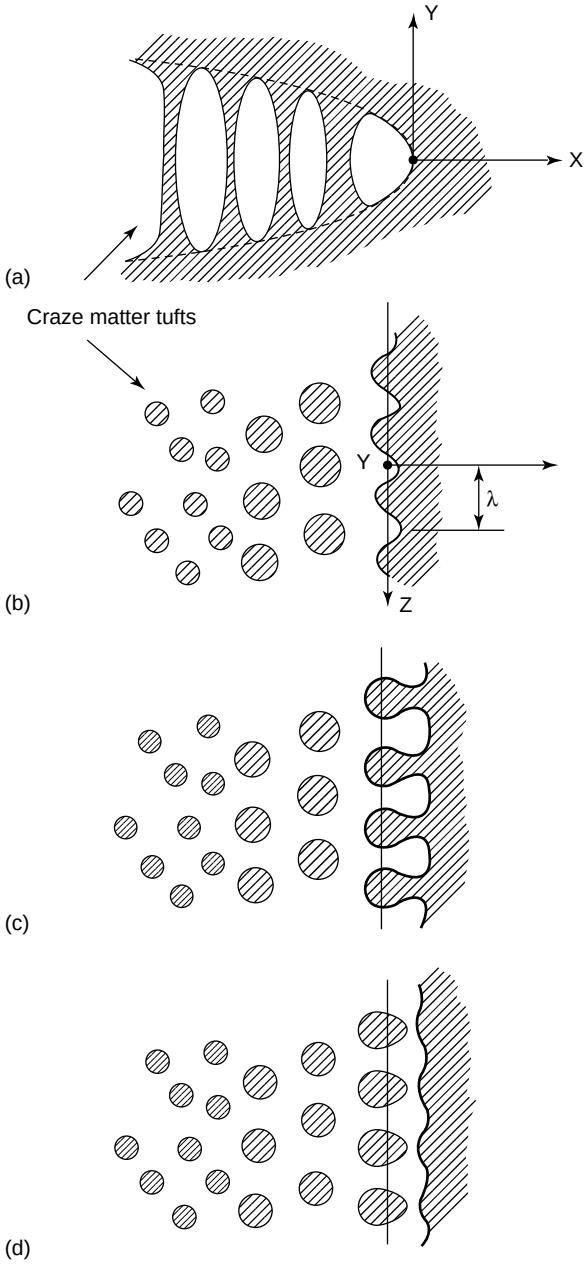

Figure 22 Schematic diagram showing craze matter production by the mechanism of meniscus instability: (a) outline of a craze tip; (b) cross-section in the craze plane across craze matter tufts; (c, d) advance of the craze front by a completed period of interface convolution. (Reproduced by permission of Pergamon Press from A.S. Argon, J.C. Hannoosh, M.M. Salama, in *Fracture 1977*, Pergamon Press, Waterloo, Canada, 445, Vol. 1, 1977.)

To explain the large discrepancy between the measured strength of materials and those based on theoretical considerations, he proposed that the elastically stored energy is not distributed uniformly throughout the specimen but is concentrated in the neighborhood of small cracks. Fracture thus occurs due to the spreading of cracks that originate in pre-existing flaws.

Griffith calculated the change in elastically stored energy using a solution obtained by Inglis[46] for the problem of a plate pierced by a small elliptical crack. The plate is stressed at right angles to the major axis of the crack. He proposed that material fracture occurs when the stress reaches the break stress level as specified in Equation (43):

$$\sigma_b = \sqrt{\frac{2\gamma_s E^*}{\pi a}} \tag{43}$$

where γ_s is the intrinsic surface energy; a is the half-crack length; and E^* is the ''reduced modulus'', equal to Young's modulus E for a thin sheet in plane stress, and to $E/(1-\nu^2)$ for a thick sheet in plane strain. Here, ν is Poisson's ratio.

Equation (43) was specifically derived for a center-notched panel made of inorganic glasses for which viscoelastic motions are almost nonexistent. Experimentally determined values for the surface energy for polymers are, however, 100–1000 times the values calculated on the basis of bond breakage alone.[47,48] The increase is largely due to viscoelastic flow, which absorbs large amounts of energy. A more general form (Equation 44) was derived by Orowan,[49] who replaced γ_s with the term $\gamma_s + \gamma_p$, where γ_p accounts for the energy involved in plastic deformation:

$$\sigma_b = \sqrt{\frac{2(\gamma_s + \gamma_p)E^*}{\pi a}} \tag{44}$$

Irwin[50] considered the fracture of solid by studying the stress near the crack tip, which is given in the form shown by Equation (45):

$$\sigma = \frac{K}{\sqrt{2\pi r}} f(\theta) \tag{45}$$

where r is the distance to the crack tip and $f(\theta)$ is a function of the polar angle, θ (Figure 23). The value of Irwin's approach is that the stress field around the crack is identical in form for all types of loading situations normal

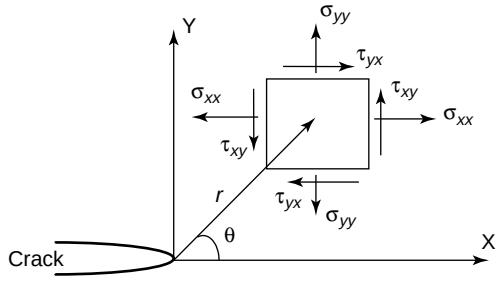

Figure 23 Definition of the coordinate axis ahead of a crack tip.

to the crack with the magnitude of the stresses (i.e. their intensity) determined by K, which is constant for given loads and geometry; K is called the stress intensity factor. As we approach the crack tip, the stresses become infinite in magnitude and r tends to zero, but the product $\sigma\sqrt{r}$ and hence K remain finite.

For an infinite sheet with a central crack of length $2a$ subjected to a uniform stress σ, it was shown (Equation 46) by Irwin that:

$$K = \sigma\sqrt{\pi a} \tag{46}$$

Irwin postulated that when σ reaches the fracture stress σ_b, K has a critical value given by Equation (47):

$$K_c = \sigma_b\sqrt{\pi a} \tag{47}$$

In linear elastic fracture mechanics it is useful also to consider the strain energy release rate, G, available for unit increase in crack length. It is assumed that fracture occurs when G reaches a critical value G_c, which is related to Griffith's surface energy as shown by Equation (48):

$$G = 2(\gamma_s + \gamma_p) \tag{48}$$

Comparison of Equations (44), (47) and (48) gives Equation (49):

$$K_c^2 = G_c E^* \tag{49}$$

There are three types of loading that a crack can experience, as shown in Figure 24. Mode I loading, where the principal load is applied normal to the crack plane, tends to open the crack. Mode II corresponds to in-plane shear loading and tends to slide one crack face with respect to the other. Mode III refers to out-of-plane shear. A cracked body can be loaded in any one of these modes, or a combination of two or three modes.

Both fracture toughness, K_c, and the critical energy release rate, G_c, are intrinsic material properties of polymers, like yielding strength and moduli. They form the central part of a great deal of the fracture and fatigue literature. Once the fracture toughness or critical energy release rate is known, it is possible to determine the value of the load that would cause failure. Values of K_c and G_c for some polymers are listed in Table 6.

Table 6 Young's modulus (E), fracture toughness (K_c), and critical energy release rate (G_c) for different polymers

Polymer	E (GPa)	K_c ($MN \cdot m\,m^{-2}$)	G_c ($kJ\,m^{-2}$)
Polystyrene	3.0	1.1	0.4
High-impact polystyrene	2.1	5.8	16
PMMA	3.0	1.2	0.5
Rubber-toughened PMMA	2.1	2.4	2.6
Epoxy resin	2.8	0.5	0.1
Rubber-toughened epoxy resin	2.4	2.2	2.0
Silica-filled epoxy resin	7.5	1.4	0.3

4.7 The Tearing of Rubbers: Extension of Griffith Theory

Rivlin and Thomas[51] and Thomas[52] recognized that the reduction in elastically stored energy due to crack propagation may be balanced by changes in energy other than that due to an increase in surface energy. Their approach was to define a quantity termed the "tearing energy", which is the energy expended per unit thickness per unit increase in crack length. The tearing energy includes surface energy, energy dissipated in plastic flow processes and energy dissipated irreversibly in viscoelastic processes. If all these changes in energy are proportional to the increase in crack length and are primarily determined by the state of deformation in the neighborhood of the tip of the crack, the total energy will still be independent of the shape of the test piece and the manner in which the deforming forces are applied.

In formal mathematical terms,[23] if the crack increases in length by an amount da, an amount of work TB da must be done, where T is the tearing energy per unit area

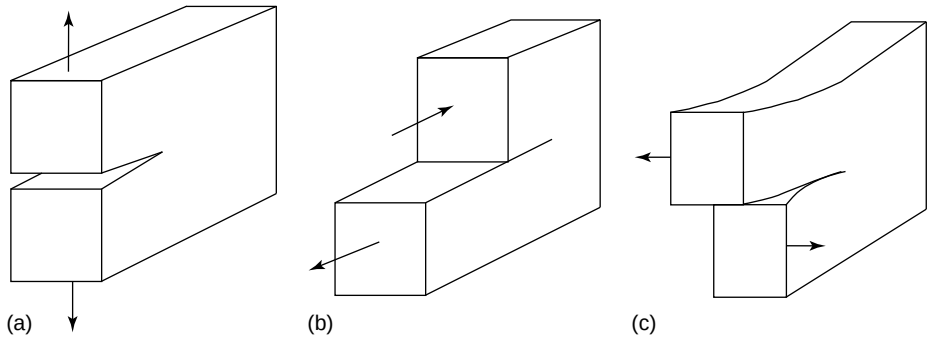

Figure 24 The three modes of loading that can be applied to a crack.

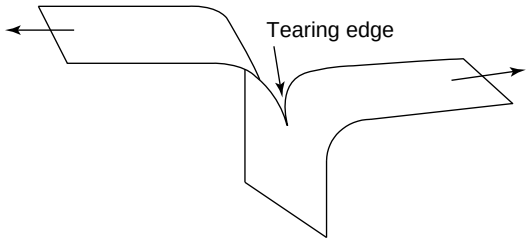

Figure 25 The standard "trouser tear" experiment.

and B is the thickness of the sheet. If no external work is due, this can be equated to the change in elastically stored energy, U, according to Equation (50):

$$-\left(\frac{\partial U}{\partial a}\right)_d = TB \tag{50}$$

The suffix d indicates that differentiation is carried out under conditions of constant displacement of the parts of the boundary that are not force-free. In Equation (50), T is defined for unit thickness of specimen. As in the case of glassy polymers, T is not to be interpreted as a surface-free energy, but involves the total deformation in the crack tip region as the crack propagates.

The so-called "trouser tear" experiment shown in Figure 25 is a particularly simple case where the equation can be evaluated immediately. After making a uniform cut in a rubber sheet the sample is subjected to tear under the applied force F. The stress distribution at the tip of the tear is complex but, provided that the legs are long, it is independent of the depth of the tear.

If the sample tears a distance da under the force F, and changes in extension of the material between the tip of the tear and the legs are ignored, the work done is given by $\Delta W = 2F\Delta a$. Because the tearing energy $T = \Delta W/B\ \Delta a$, $T = 2F/B$ and can be measured easily.

Rivlin and Thomas[51] found that two characteristic tearing energies could be defined, one for very slow rates of tearing ($T = 37\,\mathrm{kJ\,m^{-2}}$) and one for catastrophic growth ($T = 130\,\mathrm{kJ\,m^{-2}}$), and that both of these quantities were independent of the shape of the specimen.

The tearing energy is the energy required to extend the rubber to its maximum elongation and does not relate directly to tensile strength, but depends on the shape of the stress–strain curve together with the viscoelastic nature of the rubber. For example, we may contrast two different rubbers, the first possessing a high tensile strength but a very low elongation to fracture and very low viscoelastic losses, and the second possessing a low tensile strength but a high elongation to fracture and high viscoelastic losses. In spite of its comparatively low tensile strength, the second rubber may still possess a higher tearing energy.

4.8 Temperature and Loading Rate Effects: Viscoelastic Rupture of Elastomers

When simple cross-linked elastomers are stretched and held at high elongation, they will undergo a smooth relaxation period followed by sudden failure. This sequence is depicted in Figure 26 for poly(styrene–co-butadiene) rubber.[22,53–55]

Our interest lies in the loci of failure points as a function of temperature. The family of curves of stress to break σ_b at various temperatures is plotted schematically in Figure 27(a).[53] Use may be made of the time–temperature superposition principle and the WLF equation to construct a master curve, as shown in Figure 27(b).

Any amorphous polymer that is stretched undergoes some orientation of polymer segments. In this oriented state, crystallization may occur. This will increase the effective number of cross-links. If crystallization does not occur, the behavior of the sample to rupture may be described by the same equation as that for small deformations. The mechanical test most often used is one that measures stress at a constant strain rate. The ultimate tensile stress and strain at rupture vary with temperature and rate of strain. For a cross-linked unfilled rubber, which does not crystallize on stretching, the shift factors of the WLF equation (see Equation 51) allow a reasonable superposition over a wide range of conditions. In Figure 28, the stress at break σ_b is multiplied by the ratio of a reference temperature ($T_s = 263\,\mathrm{K}$) to the test temperature in accordance with the theory of rubber elasticity. The shift factors a_T, used to multiply the rate of extension R (meters per meter per second), is given

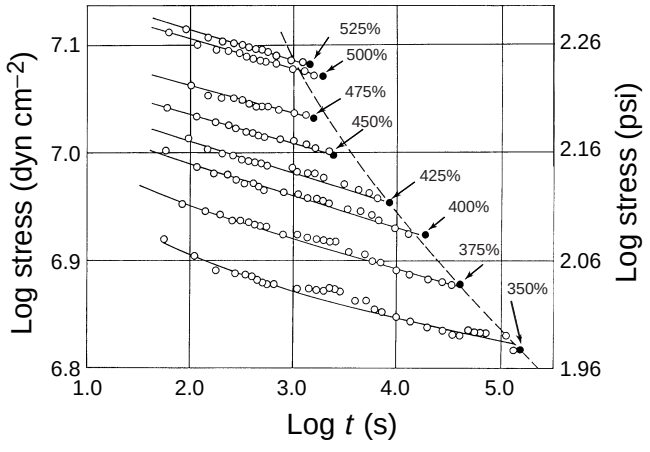

Figure 26 Stress relaxation of a cross-linked poly(styrene–co-butadiene) rubber (at 1.7 °C) at elongation from 350 to 525%. Solid points indicate rupture; the dashed line gives the ultimate strength. (Reproduced by permission of the Society of Plastics Engineers from K.W. Scott, *Polym. Eng. Sci.*, **7**, 158 (1967).)

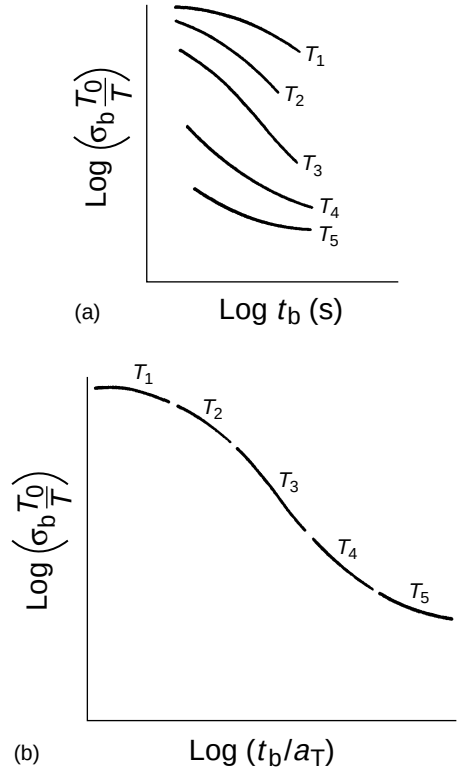

Figure 27 (a) The loci of failure point for a cross-linked poly(styrene–co-butadiene) rubber obtained as the temperature is increased from T_1 to T_5 through the glass–rubber transition. (b) Master curve of reduced stress to break vs reduced time to break for cross-linked poly(styrene–co-butadiene) rubber. (Reproduced by permission of the Society of Plastics Engineers from K.W. Scott, *Polym. Eng. Sci.*, **7**, 158 (1967).)

approximately by the WLF equation in modified form, because T_s is not T_g (Figure 29):

$$\log a_T = -\frac{8.86(T - T_s)}{101.6 + T - T_s} \qquad (51)$$

A similar reduced curve can be produced for the strain at break ε_b (Figure 30) using the same shift factors.

4.9 Fatigue

The case of creep in a polymer subjected to a constant stress has already been mentioned before. Cyclic or repeated stresses are also important in actual service, and give rise to the phenomenon of fatigue.[56–61]

Materials frequently fail by fatigue caused by the cyclic application of stresses that are below that required to cause yield or fracture when a continuously rising stress is applied. Thousands or tens of thousands of cycles may elapse with little appearing to happen. This is followed by catastrophic failure. In some cases a so-called ''endurance limit'' is observed below which fatigue failure never occurs (see Figure 31[61]). The endurance limit of PMMA shown in Figure 31 is about 2700 psi, below which no fatigue failure is observed. On the other hand, Nylon 6 does not have an endurance limit. In general, the higher the stress, or the higher the temperature, the shorter the lifetime. In terms of molecular processes, fatigue failure involves two mechanisms: weakening by adiabatic heating due to the presence of energy-dissipating modes of relaxation, and mechanical initiation and propagation of a crack. The former mechanism is dominant at high frequencies.[61] and the latter at low frequencies.[56,59]

Hertzberg and Manson[60] have identified several important molecular aspects of the fatigue process:

- High molecular weights and narrow molecular-weight distributions are generally more fatigue resistant.
- Chemical changes such as bond breakage are to be avoided or minimized.
- Elastic, inelastic, and viscoelastic deformation of the chains is desired. These motions absorb energy and tend to prevent crack growth, provided that the sample does not heat up excessively.
- Morphological changes such as drawing, orientation, and crystallization also absorb energy and are desired.
- Samples undergoing cyclic stressing at frequencies and temperatures near the glass transition temperature or secondary transitions will tend to heat up more than otherwise, causing softening or degradation.
- Adiabatic heating caused by such hysteretic effects is clearly undesirable.
- Inhomogeneous deformations, such as crazing and shear banding, absorb energy and are desired. It is especially desirous to have shear banding occur before crazing, i.e. the former mechanism should have a lower free energy of initiation.

Of course, several of these factors may be operative simultaneously.

The presence of a crack can significantly reduce the strength of a polymer due to brittle fracture as already discussed in section 4.6. A common situation is that a small flaw that was initially present develops into a crack and then grows until it reaches the critical size for brittle fracture. Crack growth can be caused by cyclic loading, a behavior called fatigue crack growth. Analysis of fatigue crack growth can be done using the stress intensity concept of fracture mechanics. Figure 32 is a schematic log–log plot of da/dN versus ΔK, which illustrates typical fatigue crack growth behavior. Here, a is the crack length, N is the number of cycles, and ΔK is the range of the stress intensity factor, i.e. the difference between the maximum stress intensity factor K_{max} and the minimum stress intensity factor K_{min} during fatigue.

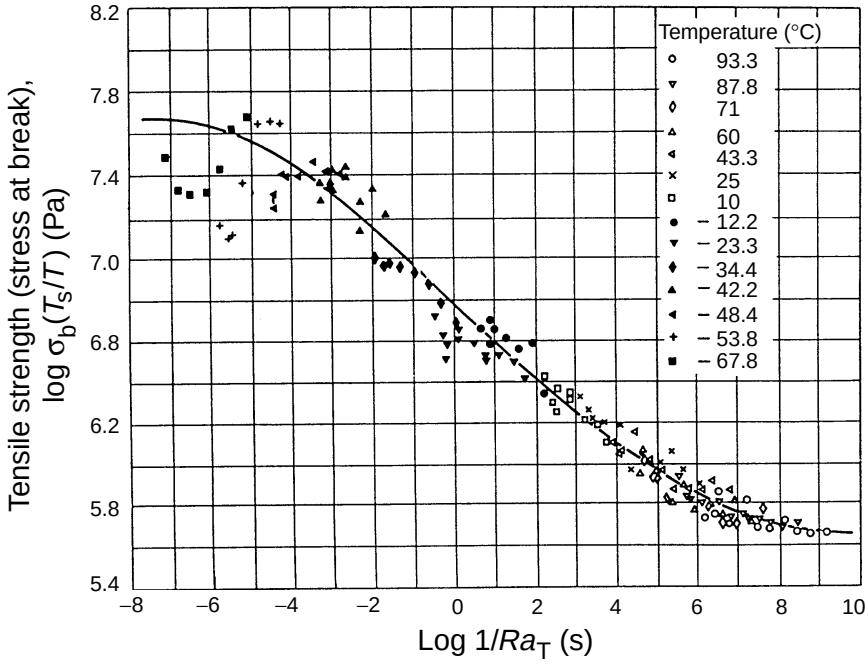

Figure 28 Variation of the tensile strength for a cross-linked poly(styrene–co-butadiene) rubber with reduced strain rate. Reference temperature is 263 K. (Reproduced by permission of John Wiley & Sons from T.L. Smith, *J. Polym. Sci.*, **32**, 99 (1958).)

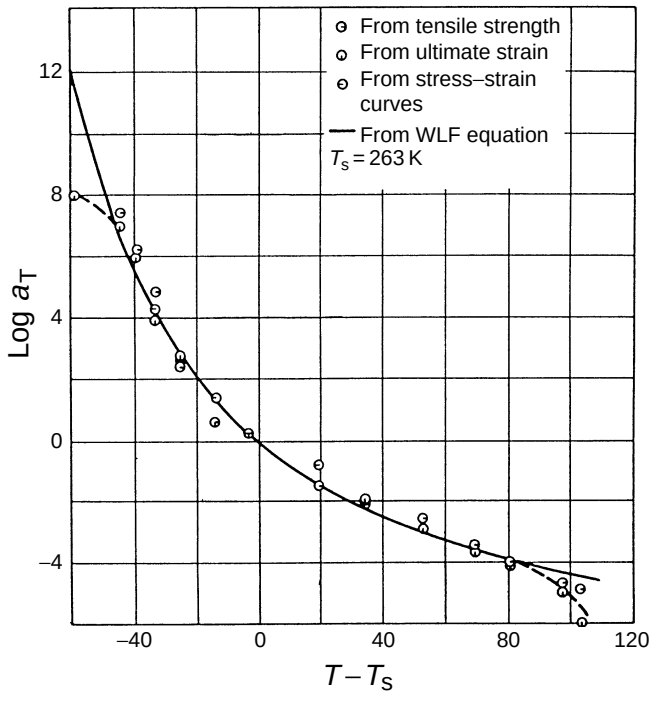

Figure 29 Experimental values of shift factor a_T from experiment and from a WLF-type equation in which the reference temperature is 263 K. (Reproduced by permission of John Wiley & Sons from T.L. Smith, *J. Polym. Sci.*, **32**, 99 (1958).)

The sigmoidal curve contains three distinct regions. At intermediate ΔK values the curve is linear, but the crack growth rate deviates from the linear trend at high and low ΔK levels. In the former case, the crack growth rate accelerates as K_{max} approaches K_c, the fracture toughness of the material. At the other extreme, da/dN approaches zero at a threshold ΔK.

The linear region of the log–log plot in Figure 32 can be described by the power law shown in Equation (52):

$$\frac{da}{dN} = A(\Delta K)^m \tag{52}$$

where a is the crack length, N is the number of cycles, ΔK is the range of the stress intensity factor (i.e. $K_{max} - K_{min}$), and A and m are constants that depend on the material and test conditions.

Equation (52) is also the most general form of the law proposed by Paris[62] and Paris and Erdogen[63] for predicting fatigue crack growth rates in metals. The general situation for glassy polymers is illustrated with some typical results shown in Figure 33,[64] where c, instead of a, is used for the crack length. A comprehensive review of the application of the Paris equation and its modified form to the fatigue behavior of polymers has been given by Manson and Hertzberg.[60]

4.10 Effects of Molecular Weight and Crystallinity

The molecular weight of the polymer has significant effects on mechanical properties. At very low molecular weight, a polymer is a viscous liquid if the T_g is below

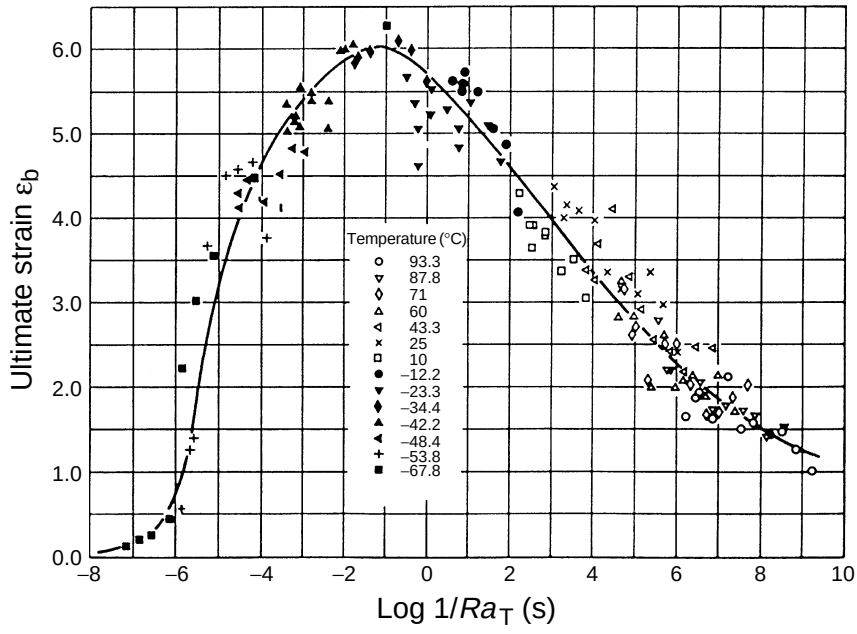

Figure 30 Variation of the break strain for a cross-linked poly(styrene–co-butadiene) rubber with reduced strain rate. Reference temperature is 263 K. (Reproduced by permission of John Wiley & Sons from T.L. Smith, *J. Polym. Sci.*, **32**, 99 (1958).)

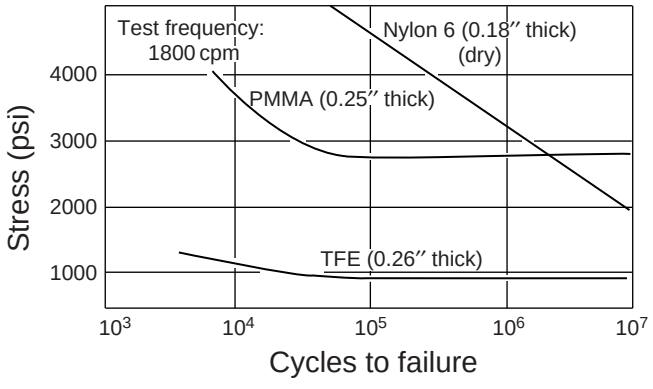

Figure 31 Fatigue life curve for nylon, PMMA, and polytetrafluoroethylene. (Reproduced by permission of the Society of Plastics Engineers from N.M. Riddell, G.P. Koo, J.L. O'Toole, *Polym. Eng. Sci.*, **6**, 363 (1966).)

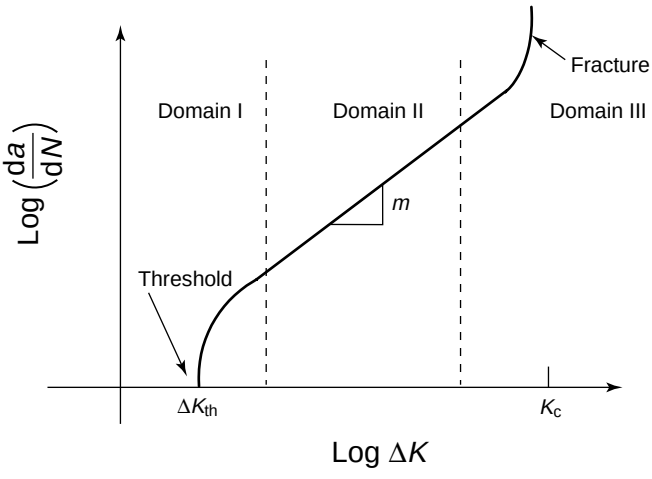

Figure 32 Typical fatigue crack growth behavior.

ambient temperature. At higher molecular weight, a polymer may become an elastomer, having low strength and a high elongation to break. Above a molecular weight of about 10^5, polymers become entangled enough to show true rubbery behavior above T_g and to exhibit a greatly increased elongation to break. If the polymer is of very low molecular weight and the T_g is higher than ambient temperature, the material is brittle. Hence, molecular entanglement must take place before the polymer can become strong enough to sustain the load. It is also true that chain ends act as imperfections in the polymer network that can adversely affect the strength properties. These same ends have little effect on the elastic moduli.

Nunes et al.[65] have reviewed the literature on polymer mechanical properties as a function of molecular weight and molecular weight distribution published from 1972 to 1982. Thermal properties, stress–strain properties, impact, fracture, fatigue, creep, stress relaxation, and cracking and crazing were examined for a wide variety of homopolymers and a limited number of copolymers. In general, mechanical properties increase as the molecular weight increases. Above some limiting molecular weight the mechanical property is usually unaffected. Although much work has been done to describe the effects of molecular weight on mechanical properties, little

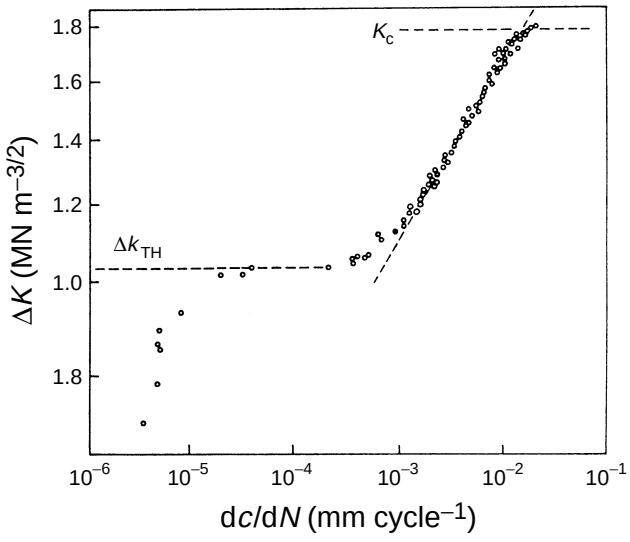

Figure 33 Fatigue crack growth data for a vinyl urethane polymer. (Reproduced by permission of Kluwer Academic Publishers from J.S. Harris, I.M. Ward, *J. Mater. Sci.*, **8**, 1655 (1973).)

quantitative correlation exists. The available equations to predict such properties as cracking and crazing, T_g, T_m and tensile strength from molecular characteristics were discussed in detail by Nunes et al.[65] A more quantitative description incorporating a wider range of mechanical properties would be more useful. This would facilitate the use of the vast amount of information available and enable it to be applied more readily to new polymer systems.

Crystallization improves the strength and modulus of polymers by reducing the degree of molecular randomization.[66] The basic crystal structure of a semicrystalline polymer consists of chain-folded lamella radiating outward from a nucleation site to form a spherulite. A molecular chain can exist in more than one lamella and more than one spherulite. The intracrystalline and intercrystalline molecular segments are called tie molecules. The molecules provide the load-bearing strength of the bulk polymer. High-molecular-weight polymers have more tie molecules and consequently higher strengths. The modulus of a polymer, however, is not strongly affected by molecular weight.

Orienting semicrystalline polymers further improves the strength and modulus in the orientation direction. Orientation is used to produce a high degree of crystal alignment in high-strength nylon, poly(ethylene terephthalate) and polypropylene textile fibers. The crystal structure in an oriented fiber still contains regions between the chain folds that cannot carry a load. This reduces the overall strength and modulus of the fiber. When the crystal segments are lengthened there are fewer chain fold defects, resulting in a stronger and stiffer fiber. High-density polyethylene has an ultimate strength of 19 GPa and a modulus of 300 GPa because its backbone is the very strong carbon–carbon bond.[67] The primary reason why these high mechanical properties are not achieved is that thermodynamic considerations favor the formation of the spherulitic rather than extended crystal structures as the polymer is quenched from the melt.

Orientation of a semicrystalline polymer below its melting point produces extended chain crystals having fewer chain folds and defects. Because of the high ultimate strength of the carbon–carbon bond, a fully aligned polymer has a very high strength and modulus. There are a number of solid-state processes capable of producing this type of crystal structure realignment. Melt processing is the most common polymer processing technique, because it allows high rates of production, has modest power requirements, and is capable of producing a wide variety of product shapes. Highly oriented crystal structures can, however, only be formed by rearranging the spherulitic crystal structure below the melting point of the polymer. There are several solid-state processing techniques capable of producing the extended chain crystals for the production of very-high-modulus polymers: drawing, extrusion, rolling, and high-pressure crystallization. A review of these techniques with emphasis on the process parameters and properties of the resulting materials has been given by Bigg.[68]

5 POLYMER BLENDS AND COMPOSITES

In this article, polymer blends are defined as combinations of two or more kinds of polymers, and composites are defined as systems combining two or more materials that are mutually insoluble by mixing or bonding them in such a way that each maintains its integrity. Polymer blends and composites can be tailored to meet special needs such as high strength and stiffness combined with light weight. The resulting high-performance materials are being increasingly used in aircraft, space, and defense applications. More economical composites, such as glass-reinforced plastics, are continually finding new uses in a wide range of products, such as automotive components, boat hulls, sports equipment, and furniture.

Polymers are often thought of as being mechanically weak and their mechanical properties have consequently been somewhat ignored in the past. With the increasing use of polymer blends and composites as major load-bearing components, however, an understanding of their mechanical properties and performance is becoming an essential part of engineering design.

Polymer blends and composites display a broad gamut of behavior, ranging from toughened elastomers through

impact-resistant plastics to fiber-reinforced thermosets and polymer-impregnated concrete. Such materials are of practical importance because their unique multiphased structure often allows for nonlinear and synergistic behavior. In this respect, polymer blends and composites are useful because of their complexity, not in spite of it. Blends and composites are toughened because many modes of resistance to failure are available.

A tremendous amount of research and development has been conducted in the past three decades on the mechanical properties of polymer blends and composites because of their importance for structural applications, and the complicated nature of nonlinear and synergistic effects of phases in controlling their mechanical performance. The flood of important research in this area has resulted in many symposia, edited collections of papers, reviews and contributions to scientific journals.[69–76] A systematic approach has been established from which the mechanical performance of a composite system, such as stiffness, strength, and failure, can be accurately predicted in a closed form or a semiclosed form based on the mechanical properties of each constituent and the interface.[77–83] A comprehensive review of the subject of composites in six print volumes has been published by Pergamon Press.[84]

5.1 Polymer Blends

The principal methods of preparing polymer blends include mechanical blending, graft copolymerization, block copolymerization, and interpenetration of two networks. The last two are often considered as subgroups of the graft method.

5.1.1 Mechanical Blends

Polymer blends may be defined as intimate mixtures of two or more kinds of polymers, with no covalent bonds between them. Historically, the oldest and simplest method involves mechanical blending, where a plastic and a noncross-linked elastomer are blended either on open rolls or through extruders.[85] Materials prepared in this manner usually contain several percent of elastomer dispersed in a plastic matrix.

In simple mechanical blends the plastic component usually predominates, with the dispersed elastomer having dimensions of the order of several micrometers. The shear action of mechanical blending also generates free radicals through polymer degradation reactions. The free radicals thus induced by mechanochemical action subsequently react to form a small number of true chemical grafts between the two components. The quantity and importance of such grafted material obviously depend on the exact mode of blending.[86] Significant improvements in impact resistance and toughness are usually noted for such blends even in cases where no particular amount of grafting is noted.

5.1.2 Graft Copolymers

Further improvement in mechanical behavior can be obtained by graft copolymerization. In the graft copolymerization method[87] the first polymer (usually the rubber) portion is dissolved in the plastic monomer, and polymerization is effected. During the polymerization, some or the entire second polymer becomes joined to the first. Often the polymerization is carried out with stirring, to bring about phase inversion. This results in a much finer dispersion of the rubber phase, and a far more complex morphology. Polymer blends and grafts designed for impact resistance have been discussed in detail by Manson and Sperling.[88]

5.1.3 Block Copolymers

In block copolymers, the individual components are joined at their ends. Block copolymers have been synthesized by several methods, but perhaps the most elegant procedure follows the ''living polymer'' anionic polymerization process.[89] The unusual features of this reaction include simultaneous nucleation and uniform growth rates of all chains, and lack of termination reactions. After exhaustion of a first monomer the polymer chains remain alive, and addition of a second monomer results in a block copolymer of the form shown in Equation (53):

$$\mathrm{A} - (\mathrm{A})_{n-2} - \mathrm{A} - \mathrm{B} - (\mathrm{B})_{m-2} - \mathrm{B} \qquad (53)$$

where n and m are the degrees of polymerization of the A- and B-mer units, respectively. More blocks can be added in the same manner, if desired. Block copolymerization results in extremely fine phases, called domains, in the form of characteristic spheres, cylinders, or lamellae. The size and shape of the phase depend on composition. An important class of block copolymers is made up of the triblock ABA thermoplastic elastomers[90] discussed by Manson and Sperling.[88]

5.1.4 Interpenetrating Polymer Networks

This novel class of polymers, together with simultaneous interpenetrating networks (SINs) and interpenetrating elastomeric networks (IENs),[91] form another important class of two-phase polymer systems. Interpenetrating polymer networks (IPNs) can be formed[92] by preparing a cross-linked polymer network, swelling in a second monomer together with activator and cross-linking agent, and polymerizing in situ. This second reaction forms

a cross-linked polymeric network that interpenetrates the first network. A SIN differs from an IPN in that both networks are formed more or less simultaneously. An example might involve concurrent noninterfering addition and condensation polymerizations, each reaction having appropriate cross-linking agents.[93] IPNs and SINs also form distinct and characteristic domain-type structures, with the restriction that both components must be continuous throughout the macroscopic mass. IENs are prepared by mixing and coagulating two different kinds of polymeric latexes, followed by a single cross-linking reaction. The result is a three-dimensional mosaic structure. The IPNs and IENs have been discussed by Manson and Sperling.[88]

5.2 Polymer Composites

Composites can be placed into three categories – particulate, fiber, and laminar – based on the shapes of the reinforcement materials.

5.2.1 Particulate Composites

Many engineering polymers that contain fillers and extenders are particulate composites. A classic example is carbon black in vulcanized rubber. Carbon black consists of tiny carbon spheroids only 5–500 nm in diameter. The carbon black improves the strength, stiffness, hardness, wear resistance, and heat resistance of the rubber.

Extenders, such as calcium carbonate, solid glass spheres, and various clays, are added so that a smaller amount of the more expensive polymer is required. The extenders may stiffen the polymer, increase the hardness and wear resistance, increase thermal conductivity, or improve resistance to creep. Strength and ductility, however, normally decrease. Introducing hollow glass spheres may impart the same changes in properties while significantly reducing the weight of the composite.

5.2.2 Fibrous Composites

Strong and stiff fibers can be made from ceramic materials that are difficult to use as structural materials in bulk form. Examples include glass, graphite (carbon), boron, and silicon carbide. When fibers of these materials are embedded in a matrix of a ductile material, such as a polymer, the resulting composite can be strong, stiff, and tough. The fibers carry most of the stress, whereas the matrix holds them in place. Good adhesion between fibers and matrix is important because this allows the matrix to carry the stress from one fiber to another. Fiber diameters are typically in the size range 1–100 μm. Fibers are used in composites in a variety of different configurations. Short, randomly oriented fibers result in a composite that has similar properties in all directions. Chopped glass fibers used to reinforce thermoplastics are of this type. Whiskers are a special class of short fiber that consist of tiny elongated single crystals that are very strong because they are dislocation free. Diameters are 1–10 μm or smaller, and lengths are 10–100 times larger than the diameter.

Long fibers can be woven into a cloth or made into a mat of intertwined strands. Glass fibers in both of these configurations are used with polyester resins to make common fiberglass. High-performance composites are often made using long, straight, continuous fibers. Continuous fibers all oriented in a single direction provide maximum strength and stiffness parallel to the fibers. Because such a material is weak if stressed in the transverse direction, several thin layers with different fiber orientations are usually stacked into a laminate. For example, composites with a thermosetting plastic matrix, often epoxy, are assembled in this manner. This is done using partially cured sheets, which are called prepregs because they have been previously impregnated with the epoxy resin. Appropriate heat and pressure are applied to complete the cross-linking reaction, while at the same time bonding the layers into a solid laminate. Fibers commonly used in this manner with an epoxy matrix include glass, graphite, boron, and the aramid polymer Kevlar.

5.2.3 Laminated Composites

A material made by combining layers is called a laminate. The layers may differ as to the fiber orientation, or they may consist of different materials. As already noted, unidirectional composite sheets are frequently laminated. The recently developed aramid–aluminum laminate has layers of an aluminum alloy and a composite with unidirectional Kevlar fibers in an epoxy matrix.

Where stiffness in bending is needed along with light weight, layers of a strong and stiff material may be placed on either side of a lightweight core. Such sandwich materials include aluminum or fibrous composite sheets bonded on each side of a core that is made of a stiff foam. Another possibility is a core made of a honeycomb of aluminum or other material.

5.3 Mechanical Behavior of Rubber-toughened Polymers

The incorporation of rubber particles into a brittle polymer has a profound effect upon the mechanical properties, as shown from the stress–strain curves in Figures 34 and 35, and the discussion provided by Young and Lovell.[21] Included in Figure 34 are the stress–strain curves for polystyrene and high-impact polystyrene (HIPS), which is a blend of polystyrene and polybutadiene. The curve for polystyrene shows

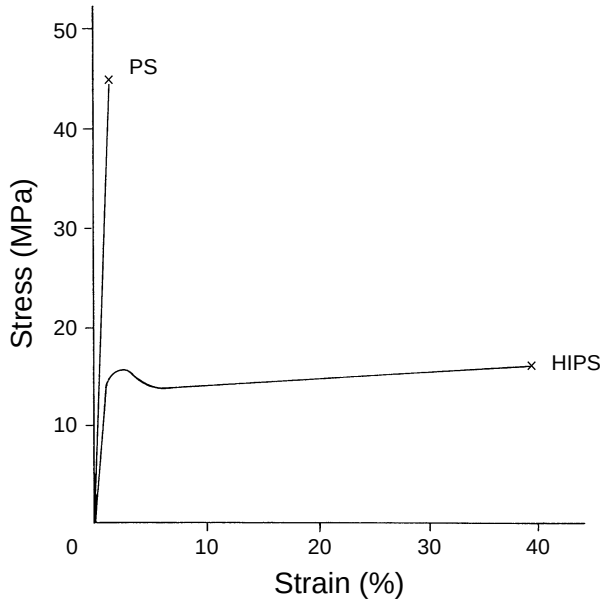

Figure 34 Stress–strain curves for polystyrene and HIPS. (Reproduced by permission of Stanley Thornes Publishers from R.J. Young, P.A. Lovell, *Introduction to Polymers*, 2nd edition, Chapman & Hall, London, 1995.)

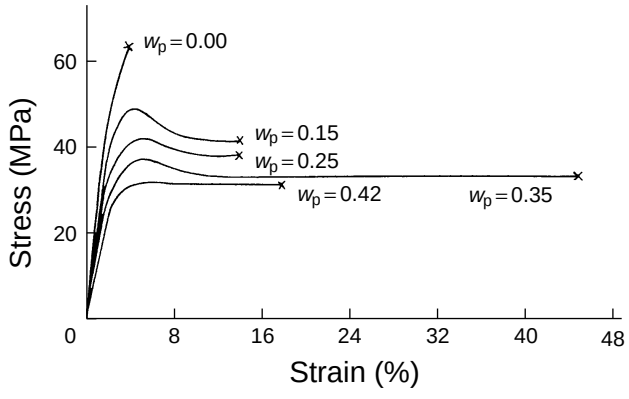

Figure 35 Stress–strain curves for RTPMMA showing the effect of the weight fraction of rubber particles w_p. (Reproduced by permission of Stanley Thornes Publishers from R.J. Young, P.A. Lovell, *Introduction to Polymers*, 2nd edition, Chapman & Hall, London, 1995.)

brittle behavior. The inclusion of the rubbery phase causes the material to undergo yield and the sample to deform plastically to about 40% strain before eventually fracturing. The plastic deformation is accompanied by "stress-whitening", whereby the necked region becomes white in appearance during deformation. As will be explained later, this is due to the formation of a large number of crazes around the rubber particles in the material.

Figure 35 shows a series of stress–strain curves for samples of rubber-toughened poly(methyl methacrylate) (RTPMMA) containing different weight fractions, w_f, of toughening particles. It can be seen that PMMA is relatively brittle, but following the addition of rubbery particles the material is able to undergo yield and deform plastically. A maximum elongation of over 45% is obtained for $w_f = 0.35$. It is found that, as with HIPS, stress-whitening is obtained following yield. In the case of RTPMMA this is thought to be due to the particles undergoing voiding rather than crazing taking place. The reduction in elongation for $w_f > 0.35$ is because at high values of w_f the particles become so close that they touch and interact with each other and thereby reduce the efficiency of toughening.

As well as toughening thermoplastics such as polystyrene and PMMA, it is also possible to toughen brittle network polymers such as epoxy resins by the addition of a rubbery second phase. Figure 36 shows the effect of the volume fraction of rubber particles, V_p, upon the fracture energy release rate, G_c, for a rubber-toughened epoxy resin.[94] The samples were prepared by the addition of 8.7% by weight of a carboxyl-terminated poly(butadiene–co-acrylonitrile) rubber to the epoxy prepolymer and curing agent. The rubber is soluble in the prepolymer but precipitates in the form of particles as the resin increases in molar mass during curing. The variation of V_p in Figure 36 was obtained by varying the cure conditions employed. Although the weight of rubber used in each case was constant, it can be seen from Figure 36 that it is the volume fraction of rubber precipitated in the form of particles that controls the toughness of the material. It should be noted that the rubber employed had reactive carboxylic acid end-groups that are capable of reacting with the epoxy prepolymer, thus ensuring a

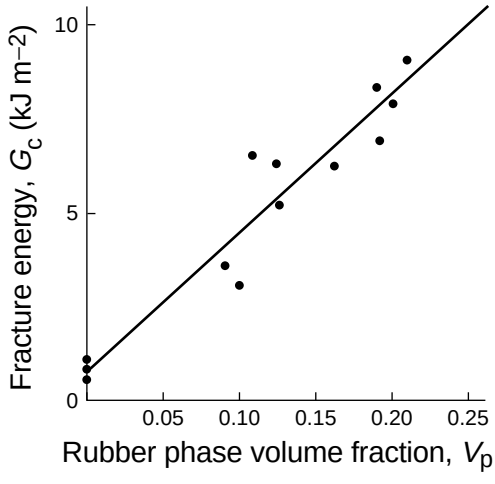

Figure 36 Dependence of G_c on rubber phase volume fraction for a rubber-toughened epoxy resin. (Reproduced by permission of Elsevier from C.B. Bucknall, M. Yoshi, *Br. Polym. J.*, **10**, 53 (1978).)

strong interface between the rubber particles and the epoxy matrix.

Both crazing and shear yielding involve the absorption of energy, and most methods of toughening brittle polymers involve modifying the polymer such that controlled high levels of crazing or shear yielding are able to take place. The second-phase spherical particles of the rubber have a Young's modulus about three orders of magnitude lower than that of the glassy matrix (Table 7[95]). This leads to a stress concentration at the equators of the particles during mechanical deformation, which is similar to the stress concentration found around holes and notches in plates. The presence of the stress concentration can lead to shear yielding or crazing around every particle and hence throughout a large volume of material rather than just at the crack tip. The polymer, therefore, absorbs a large amount of energy during deformation and is toughened.

The exact mechanisms of deformation around the rubber particles depend upon the type of polymer and the conditions of testing (rate and temperature). The structure of rubber-toughened polymers is generally visualized by transmission electron microscopic examination of thin microtomed sections. For HIPS, relatively large crazes emanate from the equators of the rubber particles. In the case of RTPMMA a large number of smaller particles generally are employed. The particles are typically of the order of 0.2–0.3 μm in diameter and have a complex morphology. It is thought that the particles in the RTPMMA toughen the material by inducing the formation of shear bands around the particles. Because they do not lead to cavitation in the matrix or a change in polymer density, they cannot be seen by electron microscopy. The material does, however, undergo some stress-whitening, which is thought to be caused by some limited cavitation within the particles.

A further insight into the deformation mechanisms that take place in rubber-toughened polymers can be obtained by measuring the change in specimen volume during deformation. In the case of HIPS, there is a significant increase in specimen volume. Plastic deformation normally occurs at constant volume and so this large increase in volume shows that crazing takes place on a large scale during deformation of the material. The behavior of HIPS can be contrasted with that of RTPMMA, for which there is no change in specimen volume during deformation. This is further confirmation that the toughening mechanism in the RTPMMA is shear yielding because shear yielding takes place at approximately constant volume.

More recently, Lovell et al.[96] have studied the effects of acrylic-matrix properties, toughening-particle composition, morphology, size, and level of inclusion on mechanical properties and fracture behavior of a range of rubber-toughened acrylic polymers.

Table 7 Typical values of Young's modulus (E) and break strength (σ_f) at room temperature and moderate testing rate

Polymer	E (Pa)	σ_f (MPa)
Rubber	$\sim 10^6$	~ 10
Semicrystalline polymer, above T_g	$\sim 10^8$	~ 30
Semicrystalline polymer, below T_g	$\sim 10^9$	~ 50
Glassy polymer	$\sim 10^9$	~ 70
Semicrystalline polymer fiber	$\sim 10^{10}$	~ 500
Polymer single-crystal fiber	$\sim 10^{11}$	~ 2000

5.4 Polymer Thin Films with Fillers

Some polymers and polymer thin films contain fillers (particulate) for a purpose other than mechanical reinforcement. In fact, they may degrade the mechanical properties. One example is photographic film, most of which consist of gelatin layers containing fillers.[97] Fillers such as silver halide crystals and coupler/coupler solvent droplets are in the photographic emulsions. Couplers are organics that react with the oxidized developer during photoprocessing to form dyes of various colors.

Although the needs for understanding the effects of fillers on mechanical behavior of polymer thin films and coatings are slowly being recognized, experimental determination remains difficult. Some polymer films are too thin and well adhered to the substrate to be readily available as a single layer for testing. Some of these additives, such as the coupler/coupler solvent dispersions or latex, are prepared in solutions and could not be extracted easily to form solid films for evaluation. Thus, analysis by means of analytical models becomes indispensable in evaluating the mechanical testing results of filler-containing films and coatings to understand the filler effects. These models may also be applied to study the effects of filler size, shape, concentration, and mechanical behavior on coating properties and, thus, potentially be utilized to optimize and design the filled system for mechanical strength without many costly experiments.

The approaches to filler effects on polymers may be divided into three categories: empirical approach, kinetic approach, and micromechanical approach. The empirical approach is based on interpolation of certain experimental data for specific filler/polymer systems. Such a formulation usually involves some curve-fitting parameters, but their physical implications are generally vague. The accuracy of the empirical approach may be very poor, especially when the formulation is used in a material system that is quite different from the one used to establish the model.

The kinetic approach considers molecular structures, network chains of filler, polymer and filler-to-polymer

linkages, and provides a rigorous treatment of mechanical (physical) properties of filler/polymer systems. In the case of filler-reinforced elastomers, kinetic theory predicts that the equilibrium stress in an unfilled vulcanizate is proportional to the number of elastically effective network chains supporting the load.[2] This number depends on the primary molecular weight of the rubber, the number of cross-links forming between the network, and the number of chain entanglements isolated between these cross-links.[98,99] For a swollen rubber a simple method exists for estimating the number of network chains using the concept of equilibrium swelling.[100]

The numerous difficulties encountered in the application of kinetic theory are discussed by Bueche.[101] They may be summed up by stating that a truly rigorous extension of the kinetic theory to filled polymers is still lacking. The result is that estimates of the number of filler-to-polymer linkages based on modulus considerations can be regarded as realistic only within an order of magnitude.[102] Thus, although the basic idea of this approach is undoubtedly sound, it sometimes does not lead to very useful results.[102]

The micromechanical approach is a continuum mechanics-based approach that utilizes a level larger than molecular scale but is still small enough to represent filler–filler interactions and filler–polymer interaction. It is based on the premise that the mechanical properties of a filled polymer are the result of the mechanical and synergistic effects of the filler, the host polymer, and the interactions through interfacial bonding. The emphasis is placed on the mechanics involving filler, host polymer, and interface because the mechanical properties of the filled system are the quantities of interest. The filler-to-polymer linkages come into play through mechanics representation of adhesion in the micromechanics treatment.

5.4.1 Moduli of Polymer Composites

The exact theoretical evaluation of effective elastic properties of a filled polymer film is, in general, very difficult because it requires a precise knowledge of the stress and strain fields everywhere in the filled polymer film. Therefore, many micromechanics-based models have been introduced to make such a problem mathematically tractable. The simplifying assumptions used in these models may, however, give rise to considerable differences in the predictions, as shown by Christensen[103] and Christensen et al.[104] The methods that are widely used include the self-consistent method,[105–107] the differential scheme,[108] the Mori–Tanaka method,[109,110] the generalized self-consistent method,[111,112] and the composite cylinder assemblage model.[113]

Gao and Tsou[114] have formulated expressions for moduli of a material containing three-dimensional, randomly distributed, ellipsoidal inclusions (fillers). When the inclusions are spherical, these expressions are simplified as (Equations 54–56):

$$K_c = K_m + \frac{V_f(K_f - K_m)}{(1 - V_f)\alpha(K_f - K_m) + K_m} K_m \tag{54}$$

$$G_c = G_m + \frac{V_f(G_f - G_m)}{(1 - V_f)\beta(G_f - G_m) + G_m} G_m \tag{55}$$

where

$$\alpha = \frac{1}{3}\frac{(1 + \nu_m)}{(1 - \nu_m)} \qquad \beta = \frac{2}{15}\frac{(4 - 5\nu_m)}{(1 - \nu_m)} \tag{56}$$

V_f is the volume fraction of the fillers; the subscripts, "c", "f", and "m" denote quantities related to the filled polymer film (composite), the filler, and the matrix material, respectively. The bulk and shear moduli, K and G, are related to Young's modulus and Poisson's ratio through conventional relationships for moduli of isotropic materials given in Table 1.

There are various models in the literature for the prediction of the moduli of polymers containing fillers.[115] Most of them work well for a low volume fraction of fillers, whereas others require that the stiffness of the fillers be much lower or much higher than that of the matrix. The Voigt model is based on the rule of mixture formula (Equation 57):

$$E_c = E_m(1 - V_f) + E_f V_f \tag{57}$$

It assumes that the monomer and polymer molecules are lined up in series and experience the same stress. The Reuss model has the form of Equation (58):

$$\frac{1}{E_c} = \frac{(1 - V_f)}{E_m} + \frac{V_f}{E_f} \tag{58}$$

which assumes that all molecules are lined up in parallel and so experience the same strain. For fiber-reinforced composites, Equations (57) and (58) can be used to determine the composite moduli in fiber and transverse direction, respectively.[23] The effect of particle shape on the moduli of filled polymers has been reviewed by Chow.[116]

The situation is considerably more complicated with more random structures such as spherulitic polymers. The measured modulus of such a structure is an average of that of the crystalline and amorphous regions. Even when it is assumed that the amorphous material is isotropic, there are problems in determining how to estimate the effective average modulus of the crystals and how to combine this with the modulus of the amorphous material to give an overall modulus for a polymer sample. This latter problem can be overcome by extrapolating

measurements of modulus as a function of crystallinity to 100% crystallinity.[21] The structure of such a material can be thought of as a randomly oriented polycrystalline structure not unlike a polycrystalline metal. The Reuss and Voigt models can again be applied with their respective assumptions of uniform stress and uniform strain.

5.4.2 *Strength of Polymer Composites*

Despite its great practical importance, the dependence of the strength of polymer films containing fillers on the filler's properties is not as clearly understood as that of the moduli of such materials. The fracture and failure processes of filled polymers are extremely complicated. This is due not only to the existence of local inhomogeneity in the materials, but also to several possible failure modes that are closely related to interfacial strength, dewetting, stress concentration at the interface, and relative brittle or ductile nature of the filler and the matrix materials. An approach often used to predict the strength of random short-fiber composites utilizes a laminate-analysis procedure. In this approach, the strength of an isotropic laminate constructed from unidirectional plies (with fiber aligned in one direction) is used to approximate the strength of random-fiber composites. Kardos[117] compared experimental results obtained by Lavengood[118] with the predictions of isotropic laminate analogy calculation. For the purpose of calculating ultimate strength, a modified maximum strain theory of failure was applied. Very little work can be found for strength prediction for composites containing cubic, spherical or elliptical fillers.

Unlike the stiffness of a material, the strength is a property that depends directly on local stress concentration and ductility of the material, and therefore is extremely difficult to predict. The addition of fillers can significantly change the mechanical characteristics of the material. The ability to predict the break strength and break strain is very critical for many polymer films to determine the physical performance of the materials. In most cases, the films are too thin to be tested mechanically for break strength and break strain.

Using the energy consideration, Gao and Tsou[114] have proposed the break stress and strain formulations shown in Equations (59) and (60):

$$\tilde{\sigma}_c = \left(\frac{E_c}{E_m}\right)^{\tau} \tilde{\sigma}_m \tag{59}$$

$$\tilde{\varepsilon}_c = \left(\frac{E_m}{E_c}\right)^{w} \tilde{\varepsilon}_m \tag{60}$$

where E_c and E_m are the Young's modulus of the filled polymer composite and matrix polymer, respectively; ε_c and σ_c are the average strain and stress of the composite in the loading direction, respectively; and $\tilde{\sigma}_m$ and $\tilde{\varepsilon}_m$ are the break stress and break strain of the matrix material, respectively. The parameters τ and w are to be determined from experimental data.

The material system used in the tests for validation of the composite modulus model is strippable 20-μm thick coatings of gelatin containing various concentrations of tricresyl phosphate solvent droplets. Uniaxial tensile properties of these coatings were measured using a SYNTEC tensile tester with specimens of dimension 15 mm × 100 mm at 21 °C and 50% relative humidity. The strain rate was 0.5 in. in.$^{-1}$ min^{-1}. The modulus and break strength of the pure gelatin (without fillers)

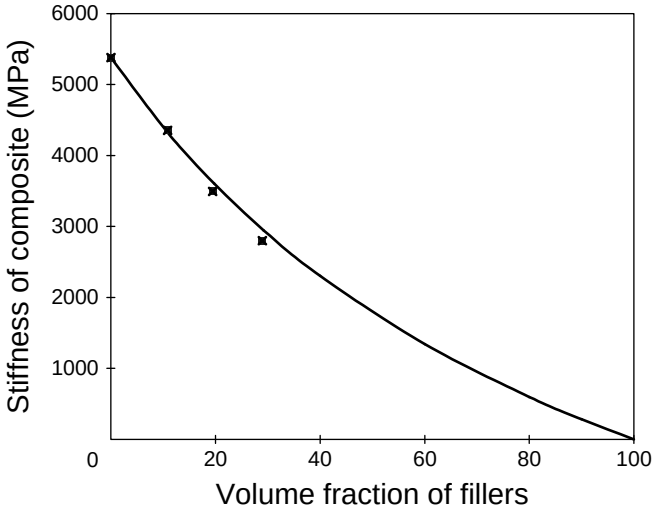

Figure 37 Comparison of the present model with experimental data of Young's modulus.

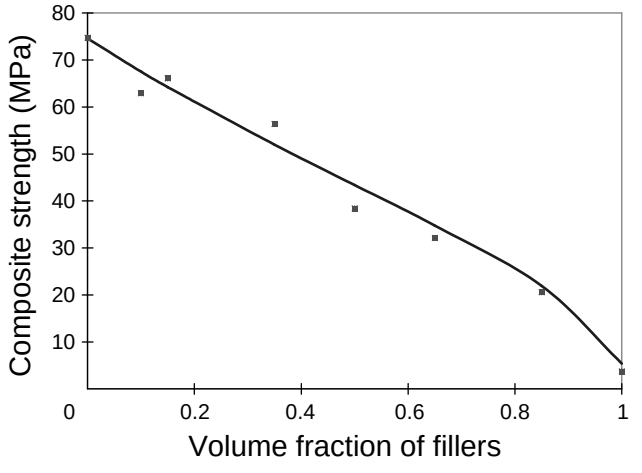

Figure 38 Comparison between the proposed model (solid line) and experimental data (dots) for the break strength of a gelatin/latex system.

were measured as 5.38 GPa and 101.97 MPa, respectively. The average size of the droplets was measured as 0.55 μm by confocal microscopy and transmission electron microscopy.

In Figure 37, the experimental data for the 20-μm-thick coatings of gelatin containing various concentrations of tricresyl phosphate solvent droplets are compared with Equation (55). The predictions for tensile break strength from the proposed model (Equation 59) with $\tau = 0.5$ are compared with experimental data for gelatin films containing various volume fractions of latex filler (Figure 38). The latex filler is poly(*n*-butylacrylate–co-2-aminoethyl methacrylate hydrochloride–co-2-hydroxyethyl methacrylate) (60 : 30 : 10) made by emulsion polymerization. As shown in Figures 37 and 38, the predictions agree extremely well with the experimental data for various volume fractions of fillers.

ACKNOWLEDGMENTS

I wish to acknowledge my managers and colleagues at Eastman Kodak Company who have encouraged and supported my work on the mechanics of polymers. These include especially Dr John Pochan, Dr Allen Caton, Mr William Hunt, Mr Richard Gammons, Dr Glenn Pearson, Mr David R Smith, Mr Charles Hura, and Dr Yeh-Hung Lai. Dr Pochan has read through the article and has given many insightful comments. Finally, and most importantly, particular appreciation is due to my family, especially my wife, Ying, for her inspiration, encouragement and support.

ABBREVIATIONS AND ACRONYMS

HIPS	High-impact Polystyrene
IEN	Interpenetrating Elastomeric Network
IPN	Interpenetrating Polymer Network
PMMA	Poly(methyl methacrylate)
RTPMMA	Rubber-toughened Poly(methyl methacrylate)
SIN	Simultaneous Interpenetrating Network
WLF	Williams–Landel–Ferry

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction • Atomic Spectroscopy in Coatings Analysis • Gas Chromatography in Coatings Analysis • Infrared and Raman Spectroscopy and Imaging in Coatings Analysis • Mechanical Properties of Solid Coatings • Microscopy of Coatings • Nuclear Magnetic Resonance of Coating and Adhesive Systems • Rheology in Coatings, Principles and Methods • Thermal Analysis of Coatings

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction • Atomic Force Microscopy in Analysis of Polymers

Polymers and Rubbers cont'd **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization • Dielectric Spectroscopy in Analysis of Polymers • Dynamic Mechanical Analysis of Polymers and Rubbers • Field Flow Fractionation in Analysis of Polymers and Rubbers • Gas Chromatography in Analysis of Polymers and Rubbers • Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Inverse Gas Chromatography in Analysis of Polymers • Near-infrared Spectroscopy of Polymers and Rubbers • Neutron Scattering in Analysis of Polymers and Rubbers • Nuclear Magnetic Resonance, Imaging of Polymers • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers • Positron Annihilation Spectroscopy of Polymers and Rubbers • Pyrolysis Techniques in the Analysis of Polymers and Rubbers • Size-exclusion Chromatography of Polymers • Supercritical Fluid Chromatography of Polymers • Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation • Thermogravimetry of Polymers • X-ray Scattering in Analysis of Polymers

Pulp and Paper **(Volume 10)**
X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction • Differential Scanning Calorimetry and Differential Thermal Analysis • Inorganic Systems, Thermal Analysis Applications to • Simultaneous Techniques in Thermal Analysis • Thermogravimetry

REFERENCES

1. A. Weaver, 'Composites: World Markets and Opportunities', *Mater. Today*, **2**(1), 3–6 (1999).
2. P.J. Flory, *Principles of Polymers*, Cornell University Press, Ithaca, Chapter 11, 1953.

3. L.R.G. Treloar, *The Physics of Rubber Elasticity*, Clarendon Press, New York, 1985.
4. M. Mooney, 'A Theory of Large Elastic Deformation', *J. Appl. Phys.*, **11**, 582–592 (1940).
5. R.S. Rivlin, 'Large Elastic Deformation of Isotopic Materials, Part III', *Trans. R. Soc. London*, **A240**, 509–525 (1948).
6. R.S. Rivlin, 'Large Elastic Deformation of Isotropic Materials, Part IV, Further Developments of the General Theory', *Trans. R. Soc. London*, **A241**, 379–397 (1948).
7. R. Blokland, 'Elasticity and Structure of Polyurethane Network', Thesis, Delft University of Technology, 1968.
8. R.W. Ogden, 'Large Deformation Isotropic Elasticity – on the Correlation of Theory and Experiment for Incompressible Rubber like Materials', *Proc. R. Soc. London*, **A326**, 565–584 (1970).
9. L.R.G. Treloar, 'The Mechanics of Rubber Elasticity', *Proc. R. Soc. London*, **A351**, 301–330 (1976).
10. E.H. Twizell, R.W. Ogden, 'Non-linear Optimization of the Material Constants in Ogden's Stress-Deformation Function for Incompressible Isotropic Elastic Materials', *J. Aust. Math. Soc. B*, **24**, 424–434 (1983).
11. R. Hill, 'Aspects of Invariance in Solid Mechanics', *Adv. Appl. Mech.*, **18**, 1–75 (1978).
12. B. Storakers, 'On Materials Representation and Constitutive Branching in Finite Compressible Elasticity', *J. Mech. Phys. Solids*, **34**(2), 125–145 (1986).
13. G.W. Becker, 'Advanced Experimental Methods for Mechanical Properties of Polymers', *Mater. Plast. Elast.*, **35**, 1387–1394 (1969).
14. H. Kolsky, 'The Propagation of Stress Waves in Viscoelastic Solids', *App. Mech. Rev.*, **11**, 465–468 (1958).
15. O.K. Chan, F.C. Chen, C.L. Choy, I.M. Ward, 'The Elastic Constants of Extruded Polypropylene and Polyethylene Terephthalate', *J. Phys. D*, **11**, 617–630 (1978).
16. R.S. Stein, H. Schaevitz, 'An Autographic Stress Relaxometer', *Rev. Sci. Instrum.*, **19**, 835–839 (1948).
17. R.D. Andrews, N. Hofman-Bang, A.V. Tobolsky, 'Elastoviscous Properties of Polyisobutylene. I. Relaxation of Stress in Whole Polymer of Different Molecular Weights at Elevated Temperatures', *J. Polym. Sci.*, **3**, 669–692 (1948).
18. J.D. Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, Wiley, New York, 1980.
19. R.L. Bergen, Jr, 'Stress Relaxation Tests', in *Testing of Polymers*, ed. J.V. Schmitz, Wiley, New York, 1–31, Vol. 2, 1966.
20. J.M.G. Cowie, *Polymers: Chemistry and Physics of Modern Materials*, Blackie Academic & Professional, London, 1977.
21. R.J. Young, P.A. Lovell, *Introduction to Polymers*, 2nd edition, Chapman & Hall, London, 1995.
22. F. Rodriguez, *Principles of Polymer Systems*, 3rd edition, Hemisphere Publishing, New York, 1989.
23. I.M. Ward, D.W. Hadley, *An Introduction to the Mechanical Properties of Solid Polymers*, Wiley, Chichester, 1996.
24. D.W. Robinson, 'An Apparatus for the Measurement of Dynamic Mechanical Properties of Polymers over a Wide Temperature Range', *J. Sci. Instrum.*, **32**, 2–6 (1955).
25. D.R. Bland, E.H. Lee, 'Calculation of the Complex Modulus of Linear Viscoelastic Materials from Vibrating Reed Measurements', *J. Appl. Phys.*, **26**, 1497–1503 (1955).
26. A.W. Nolle, 'Methods for Measuring Dynamic Mechanical Properties of Rubber-like Materials', *J. Appl. Phys.*, **19**, 753–774 (1948).
27. J.J. Aklonis, W.J. MacKnight, *Introduction to Polymer Viscoelasticity*, 2nd edition, Wiley, New York, 1983.
28. H. Leaderman, *Elastic and Creep Properties of Filamentous Materials and Other High Polymers*, The Textile Foundation, Washington, DC, 1943.
29. M.L. Williams, R.F. Landel, J.D. Ferry, 'The Temperature Dependence of Relaxation Mechanism in Amorphous Polymers and Other Gas-forming Liquids', *J. Am. Chem. Soc.*, **77**, 3701–3707 (1955).
30. A.V. Tobolsky, *Properties and Structures of Polymers*, Wiley, New York, 1960.
31. E. Catsiff, A.V. Tobolsky, 'Stress-relaxation of Polyisobutylene in the Transition Region (1,2)', *J. Colloid Sci.*, **10**, 375–392 (1955).
32. E. Catsiff, A.V. Tobolsky, 'Elastoviscous Properties of Polyisobutylene (and Other Amorphous Polymers) from Stress-relaxation Studies. IX. A Summary of Results', *J. Polym. Sci.*, **19**, 111–121 (1956).
33. D.W. Van Krevelen, *Properties of Polymers*, 3rd edition, Elsevier, Amsterdam, 1997.
34. A. Peterlin, 'Crystalline Character in Polymers' *J. Polym. Sci.*, **9C**, 61–89 (1965).
35. J.D. Hoffman, 'Region III Crystallization in Melt-crystallized Polymers: the Variable Cluster Model of Chain Folding', *Polymer*, **24**, 3–27 (1983).
36. J.D. Hoffman, 'Role of Reptation in the Rate of Crystallization of Polyethylene Fractions from the Melt', *Polymer*, **23**, 656–670 (1982).
37. D.R. Mears, K.D. Pae, J.A. Sauer, 'Effects of Hydrostatic Pressure on the Mechanical Behavior of Polyethylene and Polypropylene', *J. Appl. Phys.*, **40**, 4229–4237 (1969).
38. S.S. Sternstein, L. Ongchin, 'Yield Criteria for Plastic Deformation of Glassy High Polymers in General Stress Fields', *Polym. Prepr.*, **10**(2), 1117–1124 (1969).
39. P.B. Powden, R.J. Oxborough, 'A General Critical-strain Criterion for Crazing in Amorphous Glassy Polymers', *Philos. Mag.*, **28**, 547–559 (1973).
40. K. Matsushige, S.V. Radcliffe, E. Baer, 'The Mechanical Behaviour of Polystyrene under Pressure', *J. Mater. Sci.*, **10**, 833–845 (1975).

41. R.A. Duckett, B.C. Goswami, L.S.A. Smith, I.M. Ward, A.M. Zihlif, 'The Yielding and Crazing Behaviour of Polycarbonate in Torsion Under Superposed Hydrostatic Pressure', *Br. Polym. J.*, **10**, 11–16 (1978).
42. M. Kitagawa, 'Craze Initiation of Glassy Polymers Under the Action of Crazing Agent', *J. Polym. Sci. Polym. Phys. Ed.*, **14**, 2095–2103 (1976).
43. A.S. Argon, J.G. Hannoosh, M.M. Salama, 'Initiation and Growth of Crazes in Glassy Polymers', in *Fracture 1977*, Pergamon Press, Waterloo, Canada, 445–470, Vol. 1, 1977.
44. H.H. Kausch, *Crazing in Polymers*, Advances in Polymer Science Series 91/92, Springer-Verlag, Berlin, Vol. 2, 1990.
45. A.A. Griffith, 'The Phenomena of Rupture and Flow in Solids', *Philos. Trans. R. Soc. London, Ser. A*, **221**, 163–198 (1920).
46. G.E. Inglis, 'Stresses in a Plate Due to the Presence of Cracks and Sharp Corners', *Trans. Inst. Naval Architect.*, **55**, 219–241 (1913).
47. J.P. Berry, 'Fracture Processes in Polymeric Materials. I. The Surface Energy of Poly(methyl methacrylate)', *J. Polym. Sci.*, **50**, 107–115 (1961).
48. B. Rosen, *Fracture Processes in Polymeric Solids, Phenomena and Theory*, Interscience, New York, 1964.
49. E. Orowan, 'Fracture and Strength of Solids', *Phys. Soc. Rep. Prog. Phys.*, **12**, 185–232 (1948–1949).
50. G.R. Irwin, 'Analysis of Stresses and Strains Near the End of a Crack Traversing a Plate', *J. Appl. Mech.*, **24**, 361–364 (1957).
51. R.S. Rivlin, A.G. Thomas, 'Rupture of Rubber. I. Characteristic Energy for Tearing', *J. Polym. Sci.*, **10**, 291–318 (1953).
52. A.G. Thomas, 'Rupture of Rubber. II. The Strain Concentration at an Incision', *J. Polym. Sci.*, **18**, 177–188 (1955).
53. K.W. Scott, 'Rupture of Elastomers as a Viscoelastic Process', *Polym. Eng. Sci.*, **7**, 158–160 (1967).
54. T.L. Smith, 'Dependence of the Ultimate Properties of a GR-S Rubber on Strain Rate and Temperature', *J. Polym. Sci.*, **32**, 99–113 (1958).
55. T.L. Smith, P.J. Stedry, 'Time and Temperature Dependence of the Ultimate Properties of an SBR Rubber at Constant Elongations', *J. Appl. Phys.*, **31**(11), 1892–1898 (1960).
56. E.H. Andrews, *Fracture in Polymers*, Elsevier, Amsterdam, 1968.
57. E.H. Andrews, 'Fatigue in Polymers', in *Testing of Polymers*, ed. W. Brown, Interscience, New York, 237–296, Vol. IV, 1969.
58. E.H. Andrews, 'Fracture of Polymers', in *Macromolecular Science*, (MTP International Review of Science, Physical Chemistry Series One), ed. C.E.H. Bawn, Butterworths, London, 227–276, Vol. 8, 1972.
59. J.A. Manson, R.W. Hertzberg, 'Fatigue Failure in Polymers', *CRC Crit. Rev. Macromol. Sci.*, **1**, 433–500 (1973).
60. J.A. Manson, R.W. Hertzberg, 'Fracture of High-impact Polystyrene Under Fatigue Loading', *J. Polym. Sci., Polym. Phys. Ed.*, **11**, 2483–2488 (1973).
61. N.M. Riddell, G.P. Koo, J.L. O'Toole, 'Fatigue Mechanisms of Thermoplastics', *Polym. Eng. Sci.*, **6**, 363–368 (1966).
62. P.C. Paris, 'The Fracture Mechanics Approach to Fatigue', in *Fatigue – an Interdisciplinary Approach*, Syracuse University Press, Syracuse, New York, 107–132, 1964.
63. F. Erdogan, P.C. Paris, 'A Critical Analysis of Crack Propagation Laws', *J. Basic Eng., Trans. ASME, Ser. D*, **85**, 528–534 (1963).
64. J.S. Harris, I.M. Ward, 'Fatigue-crack Propagation in Vinyl Urethane Polymers', *J. Mater. Sci.*, **8**, 1655–1665 (1973).
65. R.W. Nunes, J.R. Martin, J.F. Johnson, 'Influence of Molecular Weight and Molecular Weight Distribution on Mechanical Properties of Polymers', *Polym. Eng. Sci.*, **22**(4), 205–228 (1982).
66. E.H. Andrews, 'The Influence of Morphology on the Mechanical Properties of Crystalline Polymers', *Pure Appl. Chem.*, **31**(1/2), 91–111 (1972).
67. P.J. Lemstra, R. Kirschbaum, 'Speciality Products Based on Commodity Polymers', *Polymer*, **26**, 1372–1384 (1985).
68. D.M. Bigg, 'Mechanical Property Enhancement of Semicrystalline Polymers – a Review', *Polym. Eng. Sci.*, **28**(13), 830–841 (1988).
69. R.J. Ceresa, *Block and Graft Copolymerization*, Wiley–Interscience, New York, Vol. 1, 1973.
70. L.H. Sperling, *Recent Advances in Polymer Blends, Grafts, and Blocks*, Plenum Press, New York, 1974.
71. L.H. Sperling, 'Mechanical Behavior of Polymer Blends, Blocks, Grafts and Interpenetrating Networks', *J. Polym. Sci., Symp.*, **60**, 175–188 (1977).
72. M. Hara, J.A. Sauer, 'Synergism in Mechanical Properties of Polymer/Polymer Blends', *J. Macromol. Sci. Rev. Macromol. Chem. Phys.*, **38**(2), 327–362 (1998).
73. M.M. Dumoulin, 'Polyblends'97/SPE RETEC, International Symposium on Polymer Blends, Alloys and Filled Systems', Industrial Materials Institute, National Research Council Canada, Boucherville, Quebec, Canada, 1997.
74. R.M. Jones, *Mechanics of Composite Materials*, McGraw-Hill Book Company, New York, 1975.
75. S.W. Tsai, H.T. Hahn, *Introduction to Composite Materials*, Technomic Publishing Company, Lancaster, PA, 1980.
76. B.D. Agarwal, L.J. Broutman, *Analysis and Performance of Fiber Composites*, 2nd edition, Wiley, New York, 1990.
77. J.D. Eshelby, 'The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems', *Proc. R. Soc., London, Ser. A*, **241**, 376–396 (1957).

78. T. Mori, K. Tanaka, 'Average Stress in Matrix and Average Elastic Energy of Materials with Misfitting Inclusions', *Acta Metall.*, **21**, 571–574 (1973).
79. C.C. Chamis, 'Simplified Composite Micromechanics Equations for Strength, Fracture Toughness and Environmental Effects', *SAMPE Q.*, **15**(4), 41–55 (1984).
80. T. Mura, *Micromechanics of Defects in Solids*, 2nd revised edition, Martinus Nijhoff, Dordrecht, 1987.
81. Z.J. Gao, 'Numerical Modeling of Damage, Property Degradation and Life Prediction in Fatigue of Composite Materials', in *Numerical Analysis and Modeling of Composite Materials*, ed. J.W. Bull, Chapman and Hall, New York, 247–286, 1995.
82. Z.J. Gao, 'A Circular Inclusion with Imperfect Interface: Eshelby's Tensor and Related Problems', *J. Appl. Mech.*, **62**, 860–866 (1995).
83. S. Nemat-Nasser, M. Hori, *Micromechanics: Overall Properties of Heterogeneous Materials*, 2nd edition, Elsevier, Amsterdam, 1999.
84. A. Kelly, C. Zweben, *Comprehensive Composite Materials*, Pergamon Press, Oxford, Vols. 1–6, 2000.
85. M. Matsuo, 'Polymerography, A New Field – Micromorphologies and Properties of Two-phase Polymer Systems', *Jpn. Plast.*, **2**, 6–16 (1968).
86. A. Casale, R.S. Porter, 'Mechanochemistry of High Polymers', *Rubber Chem., Technol.*, **44**, 534–577 (1971).
87. H.A.J. Battaerd, G.W. Tregear, *Graft Copolymers*, Interscience, New York, 1967.
88. J.A. Manson, L.H. Sperling, *Polymer Blends and Composites*, Plenum Press, New York, 1976.
89. J.F. Henderson, M. Szwarc, 'Living Polymers in the Preparation of Polymer Structures of Controlled Architecture', *Macromol. Rev.*, **3**, 317–401 (1968).
90. G. Holden, E.T. Bishop, N.R. Legge, 'Thermoplastic Elastomers', *J. Polym. Sci.*, **26C**, 37–57 (1969).
91. D. Klempner, H.L. Frisch, K.C. Frisch, 'Topologically Interpenetrating Elastomeric Networks', *J. Polym. Sci. A-2*, **8**, 921–935 (1970).
92. L.H. Sperling, D.W. Friedman, 'Synthesis and Mechanical Behavior of Interpenetrating Polymer Networks: Poly(ethyl acrylate) and Polystyrene', *J. Polym. Sci. A-2*, **7**, 425–427 (1969).
93. L.H. Sperling, R.R. Arnts, 'Simultaneous Interpenetrating Networks', *J. Appl. Polym. Sci.*, **15**, 2317–2319 (1971).
94. C.B. Bucknall, T. Yoshi, 'Relationship Between Structure and Mechanical Properties in Rubber-toughened Epoxy Resins', *Br. Polym. J.*, **10**, 53–59 (1978).
95. A.J. Kinloch, R.J. Young, *Fracture Behavior of Polymers*, Applied Science, London, 1983.
96. P.A. Lovell, M.M. Sherratt, R.J. Young, 'Mechanical Properties and Deformation Micromechanics of Rubber-toughened Acrylic Polymers', *Adv. Chem. Ser.*, **252**, 211–232 (1996).
97. T.H. James, *The Theory of the Photographic Process*, 4th edition, Macmillan Publishing, New York, 1977.
98. L. Mullins, 'Determination of Degree of Crosslinking in Natural Rubber Vulcanizates. Part III', *J. Appl. Polym. Sci.*, **2**, 1–7 (1959).
99. G. Kraus, 'Quantitative Characterization of Polybutadiene Network', *J. Appl. Polym. Sci.*, **7**, 1257–1263 (1963).
100. G. Kraus, 'Interactions Between Elastomers and Reinforcing Fillers', in *Reinforcement of Elastomers*, Wiley, New York, 125–152 (1965).
101. A.M. Bueche, 'An Investigation of the Theory of Rubber Elasticity Using Irradiated Polydimethylsiloxanes', *J. Polym. Sci.*, **19**, 297–306 (1956).
102. B. Bueche, 'Network Theories of Reinforcement', in *Reinforcement of Elastomers*, ed. G. Kraus, Wiley, New York, 1–22, 1965.
103. R.M. Christensen, 'A Critical-evaluation for a Class of Micromechanics Models', *J. Mech. Phys. Solids*, **38**, 379–404 (1990).
104. R.M. Christensen, H. Schantz, J. Shapiro, 'On the Range of Validity of the Mori-Tanaka Method', *J. Mech. Phys. Solids*, **40**, 69–73 (1992).
105. D.A.G. von Bruggeman, 'Berechnung Verschiedener Physikalischer Konstanten von Heterogenen Substanzen', *Ann. Phys.*, **24**, 636–679 (1935).
106. B. Budiansky, 'On the Elastic Moduli of Some Heterogeneous Materials', *J. Mech. Phys. Solids*, **13**, 223–227 (1965).
107. R. Hill, 'A Self-consistent Mechanics of Composite Materials', *J. Mech. Phys. Solids*, **13**, 213–222 (1965).
108. A.N. Roscoe, 'The Viscosity of Suspensions of Rigid Spheres', *Br. J. Appl. Phys.*, **3**, 267–269 (1952).
109. Y. Benveniste, 'A New Approach to the Application of Mori-Tanaka's Theory in Composite Materials', *Mech. Mater.*, **6**, 147–157 (1987).
110. T. Mori, K. Wakashima, 'Successive Iteration Method in the Evaluation of Average Fields in Elastically Inhomogeneous Material', in *Micromechanics and Inhomogeneity*, eds. G.J. Weng, M. Taya, H. Abe, Springer, New York, 269–282, 1992.
111. J.K. Mackenzie, 'The Elastic Constants of a Solid Containing Spherical Holes', *Proc. Phys. Soc. London, Sect. B*, **63**, 2–11 (1950).
112. R.M. Christensen, K.H. Lo, 'Solutions of Effective Shear Properties in Three Phase Sphere and Cylinder Models', *J. Mech. Phys. Solids*, **27**, 315–330 (1979).
113. Z. Hashin, B.W. Rosen, 'The Elastic Moduli of Fiber-reinforced Materials', *J. Appl. Mech.*, **31**, 223–232 (1964).
114. Z.J. Gao, A.H. Tsou, 'Mechanical Properties of Polymers Containing Fillers', *J. Polym. Sci., Part B: Polym. Phys.*, **37**, 155–172 (1999).

115. L.E. Nielsen, *Mechanical Properties of Polymers and Composites*, Marcel Dekker, New York, Vol. 2, 1974.
116. T.S. Chow, 'The Effect of Particle Shape on the Mechanical Properties of Filled Polymers', *J. Mater. Sci.*, **15**(8), 1873–1888 (1980).
117. J.L. Kardos, 'Structure–Property Relations for Short-Fiber Reinforced Plastics', Divisional Technical Meeting, Engineering Properties and Structure Division, SPE, Akron, OH, 1975.
118. R.E. Lavengood, 'Strength of Short-fiber Reinforced Composites', *Polym. Eng. Sci.*, **12**, 48–52 (1972).

Near-infrared Spectroscopy of Polymers and Rubbers

Jerry J. Workman, Jr
The Kimberly-Clark Corporation, Neenah, USA

1 Introduction **1**
2 Review of Near-infrared Spectroscopy in Polymer and Rubber Analysis **7**
2.1 Sorting Polymers 7
2.2 Polymer Reaction Monitoring 8
3 Sample Preparation **8**
4 Interpretive Spectroscopy of Organic Compound Spectra for Polymers and Rubbers in the Near-infrared Region **9**
4.1 Aliphatic Hydrocarbons (*n*-Alkanes) including Polymeric Forms 9
4.2 Alkenes (Olefins) and Alkynes (Acetylenes) 10
4.3 Olefinic C−H Stretch 10
4.4 Alkyne C−H Stretch 11
4.5 Olefinic C−H Bend 11
4.6 Aromatic Systems 11
4.7 Carbonyl Compounds (>C=O) 12
4.8 Amides (R−C=ONH−X) 12
4.9 Hydrogen Bonding (X−H···Y) 13
4.10 The O−H Stretch 13
4.11 N−H Group Frequencies and Nitrogen-containing Compounds 14
4.12 Amines (R−NH−X) 15
4.13 Amides (−C=O−NX′X″, also shown as −NHCO−) 15
5 Polymers **15**
5.1 Types of Polymer 15
5.2 Spectral Correlation Chart for Polymers 17
5.3 Qualitative Analysis of Polymers 17
5.4 Quantitative Analysis of Polymers 19
6 Rubbers **21**
6.1 Types of Rubber 21
6.2 Spectral Correlation Chart for Rubbers 21
6.3 Qualitative Analysis of Rubbers 21
6.4 Quantitative Analysis of Rubbers 22
7 Method Development **23**
7.1 Qualitative Analysis 24
7.2 General Qualitative Practices 24
7.3 Spectral Searching 24
7.4 Near-infrared Quantitative Calibration 24
7.5 Laboratory Methods 25
7.6 At-line Measurements 25
7.7 Process and On-line Measurements 25
8 Comparison of Near-infrared with Infrared and Raman Spectroscopy **26**
Abbreviations and Acronyms **26**
Related Articles **27**
References **27**

Molecular spectroscopy as provided using the near-infrared (NIR) measurement technique is valuable for polymer identification, characterization, and quantitation. NIR spectroscopy can be completed for in situ process applications where no sample preparation is required, and where rugged optical systems are a necessity. The NIR region is a complimentary band of the electromagnetic spectrum to the mid-infrared (MIR) region (4000–500 cm^{-1}), encompassing 13 333–4000 cm^{-1} or 750–2500 nm (nanometers, 10^{-9} m).

NIR and infrared (IR) spectroscopy are routinely used to qualify monomers prior to polymerization reactions. NIR is used to measure the kinetics of polymer onset and can be used to detect end-point completion and initiator compound levels in polymerization reactions. NIR spectroscopy can also be used to sort polymers and to control the quality of incoming raw monomers and finished polymeric materials. Molecular spectroscopy using the NIR and IR measurement techniques is often used for competitive analysis and to determine thermal or photoinduced oxidation or degradation reactions in polymers. This article delineates the background, theory, band assignments, and applications of NIR spectroscopy for the measurement of polymers and rubbers.

1 INTRODUCTION

H. Staudinger[1] was one of the first investigators into macromolecular compounds. His original work resulted in the synthesis and test method development of numerous macromolecules which we now refer to as polymers, meaning poly (many) + mers (parts). Many chemists disagreed with Staudinger's pioneering work and would not accept the widespread existence of macromolecules held together by covalent bonds. Herman Mark[2,3] provided support for Staudinger's thesis by confirming that cellulose consisted of a number of carbon, hydrogen, and oxygen atoms held together by covalent bonds. This work of Mark's led to the discovery that macromolecules as

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

polymers do not have specific molecular weights, but rather exhibit a range of molecular weights that are statistically averaged and reported in units termed daltons, where 1 dalton is equivalent to 1 amu (atomic mass unit). Thus, the molecular weights of polymer systems are reported as an average molecular weight in X daltons, with a standard deviation of $\pm Y$ daltons.

Since the discoveries of Herman Mark, polymer science has depended much on the use of instrumentation for confirmation and characterization of molecular weights and structures of polymers, particularly, DSC (differential scanning calorimetry), TGA (thermal gravimetric analysis), proton and carbon-13 NMR (nuclear magnetic resonance) spectrometry, MS (mass spectrometry), LC/MS (liquid chromatography/mass spectrometry), X-ray diffraction, GPC (gel permeation chromatography), and GC (gas chromatography). Spectroscopic techniques often used in polymer analysis and characterization include hyphenated methods such as GC/IR and TGA/IR, as well as ATR (attenuated total reflectance)/IR, DRIFTS (diffuse reflectance Fourier transform spectroscopy), specular reflectance, transmission IR, and Raman spectroscopy. Each of these analytical measurement techniques is capable of yielding complementary information relative to a macromolecule of interest, such as modern polymer systems.

NIR and IR are routinely used to qualify monomers prior to polymerization reactions. NIR is used to measure the kinetics of polymer onset and can be used to detect end-point completion and initiator compound levels in polymerization reactions. NIR spectroscopy can also be used to sort polymers and to control the quality of incoming raw monomers and finished polymeric materials. Molecular spectroscopy using the NIR and IR measurement techniques is often used for competitive analysis and to determine thermal or photoinduced oxidation or degradation reactions in polymers. In summary, molecular spectroscopy as provided using the NIR measurement technique is valuable for polymer identification, characterization, and quantitation. NIR spectroscopy can be completed for in situ process applications where no sample preparation is required, and where rugged optical systems are a necessity.

Some of the earliest work published was in applying IR and NIR spectroscopy to polymer characterization.[5–14] The NIR region is a complimentary band of the electromagnetic spectrum to the MIR region (4000–500 cm^{-1}). The NIR region is generally defined as 13 333–4000 cm^{-1}, or 5750–2500 nm. The majority of analytical work in the NIR region is reported by the literature in nanometers and so this chapter will follow the same convention. The most notable and earliest NIR work was completed by J.W. Ellis and published in 1928.[4] Ellis' review work reported on absorption studies of organic liquids and cited 44 separate papers published prior to June, 1929. Ellis reported bands for carbon–hydrogen-based organic liquids to be present at 2.3–2.2, 1.7, 1.4, 1.2, 1.0, and 0.9 μm. W. Kaye provided a review of NIR spectroscopic publications from the late 1920s up until April 1954. His work cites 106 references demonstrating the ''hydrogenic'' stretching vibrations for C−H, N−H, and O−H containing molecules from 0.7 to 3.5 μm. The most notable review papers describing the history and use of NIR molecular spectroscopy for organic compound and polymer characterization are found in work by Kaye,[5] Goddu,[6] Whetsel,[7] and Stark et al.[8] Excellent references are directly related to polymer analysis using IR and NIR spectroscopy.[9–14]

IR energy is the electromagnetic energy of molecular vibration. The energy band is defined for convenience as the NIR (0.75–2.50 μm); the IR 2.50–40.0 μm; and the far-IR (FIR) (40.0–1000 μm). Table 1 illustrates the region of the EMR (electromagnetic radiation) spectrum referred to as the NIR region. The table shows the molecular interactions associated with the energy frequencies (or corresponding wavelengths) of the various regions. Dominant NIR spectral features include methyl C−H, methylene C−H, methoxy C−H, carbonyl C−H, aromatic C−H, hydroxyl O−H; N−H from primary amides, secondary amides (both alkyl, and aryl group associations), N−H from primary, secondary, and tertiary amines; and N−H from amine salts.

Light has both particle and wave properties; quantum theory tells us that the energy of a light ''particle'' or photon (E_p) is given by Equation (1)

$$E_p = h\nu \tag{1}$$

where h is Planck's constant (or 6.6256×10^{-27} erg-s) and ν is the frequency of light (or the number of vibrations per second in units of s^{-1}). Thus the energy for any specific photon can be quantified, and it is this energy which interacts with the vibrating bonds within *IR active molecules*.

In this text the wavelength from 350 to 3000 nm (28 500–3333 wavenumbers) will be addressed. First, a

Table 1 Spectroscopic regions of interest for polymer analysis

Region	Wavelength (nm)	Characteristic measured
Visible	360–780	Electronic transitions: color measurements
NIR	780–2500	Overtone and combination bands of fundamental molecular vibrations
IR	2500–40 000	Fundamental molecular vibrations: stretching, bending, wagging, scissoring

review of the relationships between micrometers, centimeters, and nanometers to wavenumbers and frequency is given using Equations (2–5):

$$\bar{\nu} = \frac{\nu}{(c/n)} \tag{2}$$

$$\bar{\nu} = \frac{1}{\lambda} \tag{3}$$

and

$$\nu = \frac{(c/n)}{\lambda} \tag{4}$$

$$\lambda = \frac{(c/n)}{\nu} \tag{5}$$

where $\bar{\nu}$ is the wavenumber (number of waves per unit length), ν is the frequency (number of vibrations per unit time), c is the velocity of light in a vacuum (approximately 2.998×10^{10} cm s^{-1}), n is the refractive index of material in which c is measured (air is 1.0003) and λ is the wavelength (length of one complete wave); note that λ in nanometers $= 10^7/\bar{\nu}$.

Each of the above symbols is in the following units:

- $\bar{\nu}$ in cm^{-1} *or* 1/cm
- ν in s^{-1} *or* 1/s *or* hertz (Hz)
- λ in cm.

Light from a NIR spectrophotometer is directed to strike a polymer or rubber matrix consisting of one or more types of macromolecule. If the molecules do not interact with the light then the light passes through the matrix unaffected. If molecules interact with the light in a very specific way (i.e. molecular absorption) we refer to them as *active* or *IR active*. For NIR energy, the overtones of X–H bonds, that is N–H, C–H, and O–H, stretching and bending are of the greatest interest. IR active molecules can be seen as consisting of mechanical models with vibrating dipoles. Each dipole model vibrates with a *specific frequency* and *amplitude* as shown in the simple model in Figure 1.

Note that the term *frequency* refers to the number of vibrations per unit of time, designated by the Greek letter ν (nu) and generally specified in units of s^{-1} or hertz. *Amplitude* is defined by the interatomic distance covered at the extremes of the vibrating dipole, and is dependent upon the amount of energy absorbed by the IR active bond. When incoming photons from a spectrophotometer source lamp (after they have passed through the monochromator or interferometer) strike different molecules in a sample, two direct results

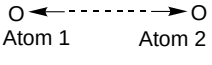

Figure 1 Model of IR active molecule as a vibrating dipole between two atoms.

may occur: (1) the disturbing energy does not match the natural vibrational frequency of the molecule; or (2) the disturbing frequency does match the vibrational frequency of the molecule. When there is a match between the disturbing frequency of the illumination energy and the natural vibrational frequency of a molecule in the sample, the molecule absorbs this energy, which in turn increases the vibrational amplitude of the absorbing dipoles. However, regardless of the increase in amplitude, the frequency of the absorbing vibration remains constant.

Another name for the dipole model from Figure 1 is an *ideal harmonic oscillator* (HO). The frequency at which the dipole (or ideal HO) vibrates (stretches or bends) is dependent upon the bond strength and the masses of the atoms bonded together. When the HO vibrates, the vibrational energy is continuously changing from kinetic to potential and back again. The total energy in the bond is proportional to the frequency of the vibration. The use of Hooke's law (in our case referring to the elasticity properties of the HO), is applied to illustrate the properties of the two atoms with a well-behaved spring-like bond between them. The natural frequency of vibration for a bond (or any two masses connected by a spring) is given by the well-known relationship, Equation (6)

$$\nu = \frac{1}{2\pi} \sqrt{\frac{K}{\left(\frac{1}{m_1} + \frac{1}{m_2}\right.}} \tag{6}$$

where K is a force constant that varies from one bond to another; m_1 is the mass of atom 1 and m_2 is the mass of atom 2. Note that as a first approximation the force constant (K) for single bonds is one-half times that of a double bond, and one-third that of a triple bond. Also note that as the mass of the atoms increases, the frequency of the vibration decreases.

The effects of quantum mechanics on a simple HO indicate that we cannot treat the bond between two atoms quite as simply as two masses connected by a spring. This is no surprise since quantum mechanical evidence has shown that vibrational energy between atoms in a molecule is quantized into discrete energy levels. When the conditions are right, vibrational energy in a molecule "jumps" from one energy level to another. The discrete vibrational energy levels for any molecule (E_{VIB}) are given by Equation (7)

$$E_{\mathrm{VIB}} = h\nu \left(\upsilon + \frac{1}{2}\right. \tag{7}$$

where h is Plank's constant, ν is the vibrational frequency of the bond and υ (represented by the Greek upsilon) is

the quantum number (which can only have the integer values of 0, 1, 2, 3, and so on).

The concept of an *anharmonic oscillator* allows for the more realistic calculations of the positions of the allowed overtone transitions. The energy levels for these overtones are not found as the product of exact integer multiples times the fundamental frequency. In fact, the following expression defines the relationships between wavenumber (for a given bond) and the vibrational energy of that bond. The relationship is calculated from the Schrödinger equation to yield Equation (8)

$$\bar{\nu} = \frac{E_{\text{VIB}}}{hc} = \bar{\nu}_1 \upsilon - x_1 \bar{\nu}_1 (\upsilon + \upsilon^2) \tag{8}$$

where υ is an integer number, that is, 0, 1, 2, 3, . . ., n and $x_1\bar{\nu}_1$ is the unique anharmonicity constant for each bond.

Calculations of band positions using Equation (8) will more closely approximate observed band positions than those calculated from the ideal HO expression found in Equation (6). As a rule of thumb, the first overtone for a fundamental can be calculated as a 1–5% shift due to anharmonicity, thus the expression, Equation (9)

$$\lambda = \frac{\lambda_1}{\kappa} + \frac{\lambda_1}{\kappa}(0.01, 0.02, \ldots, 0.05) \tag{9}$$

is used to illustrate the occurrence of overtones absorption using a fundamental absorption occurring at approximately $2632\,\text{cm}^{-1}$; converting to nanometers using $10^7/\nu(\text{in cm}^{-1}) = 3800\,\text{nm}$; the first overtone would occur at, Equation (10)

$$\nu = \frac{3800}{2} + (3800 \times 0.01) = 1938\,\text{nm} \quad \text{to}$$
$$\nu = \frac{3800}{2} + (3800 \times 0.05) = 2090\,\text{nm} \tag{10}$$

and thus one would expect the first overtone to occur somewhere between 1938 and 2090 nm rather than the 1900 nm calculated using a simple HO model.

Molecules that absorb NIR energy vibrate in two fundamental modes, *stretching* and *bending*. Stretching is defined as a continuous change in the interatomic distance along the axis between two atoms, and bending is defined as a change in the bond angle between two atoms. Figure 2 illustrates the often repeated stretching and bending interactions defining IR active species within IR active molecules. Note that NIR spectral features arise from the molecular absorptions of the overtones and combination bands from fundamental vibrational bands found in the MIR region. For fundamental vibrations there exists a series of overtones with decreasing intensity as

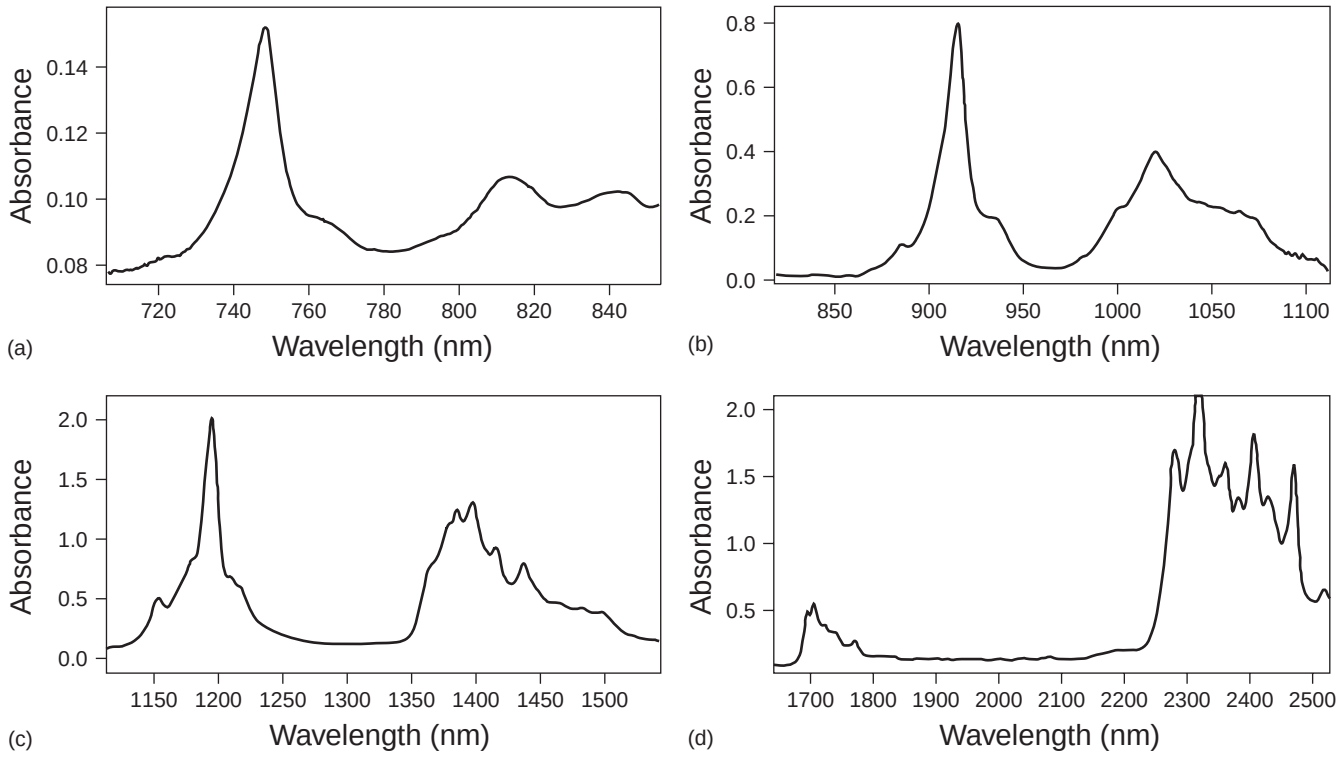

Figure 2 Illustration of aliphatic C–H (as predominantly methyl C–H str.) from 2,2,4-dimethyl pentane (iso-octane). (a) The fourth overtone (5ν) with combination bands, (b) the third overtone (4ν) with combination bands, (c) the second overtone (3ν) and combination bands, (d) the first overtone (2ν) and combination bands and a portion of the fundamental band.

the transition number (overtone) increases. Combination bands arise as the summation of fundamental bands, their intensity decreasing with an increase in the summation frequency.

Fundamental and overtone absorptions arise when molecules are excited from the ground state to higher energy excited states. Fundamental vibrations will change in amplitude when absorbing energy of the same frequency from an outside source that strikes the vibrating bond(s). Variations in hydrogen bonding manifest themselves as changes in the force constants of the X−H bonds. Generally bands will shift in frequency and broaden due to the formation of hydrogen bonding. Since combination bands result as the summation of two or more fundamental vibrations, and overtones occur as the result of the multiples of fundamental vibrations, frequency shifts related to hydrogen bonding have a greater relative effect on combination and overtone bands than on their corresponding fundamentals. This feature of the NIR region alerts one to the importance of the relative hydrogen bonding effects brought about by solvent and temperature variations.

Precise band assignments are difficult in the NIR region because a single band may be attributable to several possible combinations of overlapping overtone and combination bands. The influence of hydrogen bonding results in band shifts to lower frequencies (higher wavelengths); a decrease in hydrogen bonding due to dilution and higher temperatures results in band shifts to higher frequencies (lower wavelengths). Band shifts of the magnitude of 10–100 cm^{-1}, corresponding to from a few nanometers to 50 nm may be observed. The substantial effect of hydrogen bonding should be kept in the forefront of thought when composing calibration sample sets and experimental designs for NIR experiments.

NIR spectra contain information relating to differences in bond strengths, chemical species, electronegativity, and hydrogen bonding. For solid samples, information with respect to scattering, diffuse reflectance (DR), specular reflectance, surface gloss, refractive index, and polarization of reflected light are all superimposed on the NIR vibrational information. Aspects related to hydrogen bonding and hydronium ion concentration are included within the spectra. Light can interact with the sample as reflection, refraction, absorption, scattering, diffraction, and transmission. Signal losses from the sample can occur as specular reflection, internal scattering, refraction, complete absorption, transmission loss during reflectance measurements, and trapping losses. Spectral artifacts can also arise as offset or multiplicative errors caused by coloration of the sample, variable particle sizes and resultant variability in apparent pathlength, refractive index changes in clear liquids relative to temperature changes, and pathlength differences due to temperature-induced density changes.

The types of vibration found in NIR and IR spectroscopy are designated by ν (Greek letter nu) with a subscript designating whether the vibration is stretching (s) or bending (b), for example, stretching is designated as $\tilde{\nu}_s$, and bending vibrations as $\tilde{\nu}_b$. Combination bands resulting from the sum of stretching and bending modes are designated as $\tilde{\nu}_b + \tilde{\nu}_s$ and harmonics are designated as $k\tilde{\nu}_s$, where k is an integer number as 2 (first overtone), 3 (second overtone), ..., k ($k-1$th overtone), and $\tilde{\nu}_s$ is the frequency of the fundamental stretch vibration for a specific functionality.

As an exercise, apply this brief lesson in quantum mechanics to a problem using *n*-decane. Typical NIR spectra of *n*-decane, $CH_3(CH_2)_8CH_3$, appear in Figure 3. *n*-Decane, a member of a class of hydrocarbon compounds known as paraffins (alkanes), contains two methyl groups and eight methylene groups. Thus we would expect to see strong methylene C−H stretch absorptions as well as notable methyl C−H stretch absorptivity. Indeed, the spectrum of decane shows combination bands for methyl and methylene C−H stretch plus C−H bend as well as overtone bands. To calculate the expected positions for methyl C−H bands we relate the following calculations, assuming an ideal HO model.

Methyl C−H combination bands appear near the sum of two or three fundamental bands. For example (in the case of *n*-decane), a C−H methyl combination band should occur near 2262.4 nm. To illustrate, 2960 cm^{-1} (C−H stretch) + 1460 cm^{-1} (C−H bend) = 4420 cm^{-1}; converting to wavelength we calculate $(10^7/4420\,cm^{-1}) = 2262.4$ nm.

As previously noted, overtones (also termed harmonics) occur at integer multiples of fundamental bands. For example, using the simple HO model, C−H stretch overtones should occur at 1689 nm, 1126 nm, and 845 nm. For the example of *n*-decane we note the following:

- first overtone at 2960 cm^{-1}(C−H stretch) × 2 = 5920 cm^{-1} = 1689 nm;
- second overtone at 2960 cm^{-1}(C−H stretch) × 3 = 8880 cm^{-1} = 1126 nm;
- third overtone at 2960 cm^{-1}(C−H stretch) × 4 = 11 840 cm^{-1} = 845 nm.

When laboratory measurements are taken for *n*-decane it can be noted that band positions do not match the calculations for ideal. In fact, the actual positions are approximately 2310 nm, 1725 nm, 1212 nm, and 930 nm, respectively. As noted, an application of the anharmonic oscillator expression (Equations 8 and 9) allows better approximations of actual band positions. In fact using a

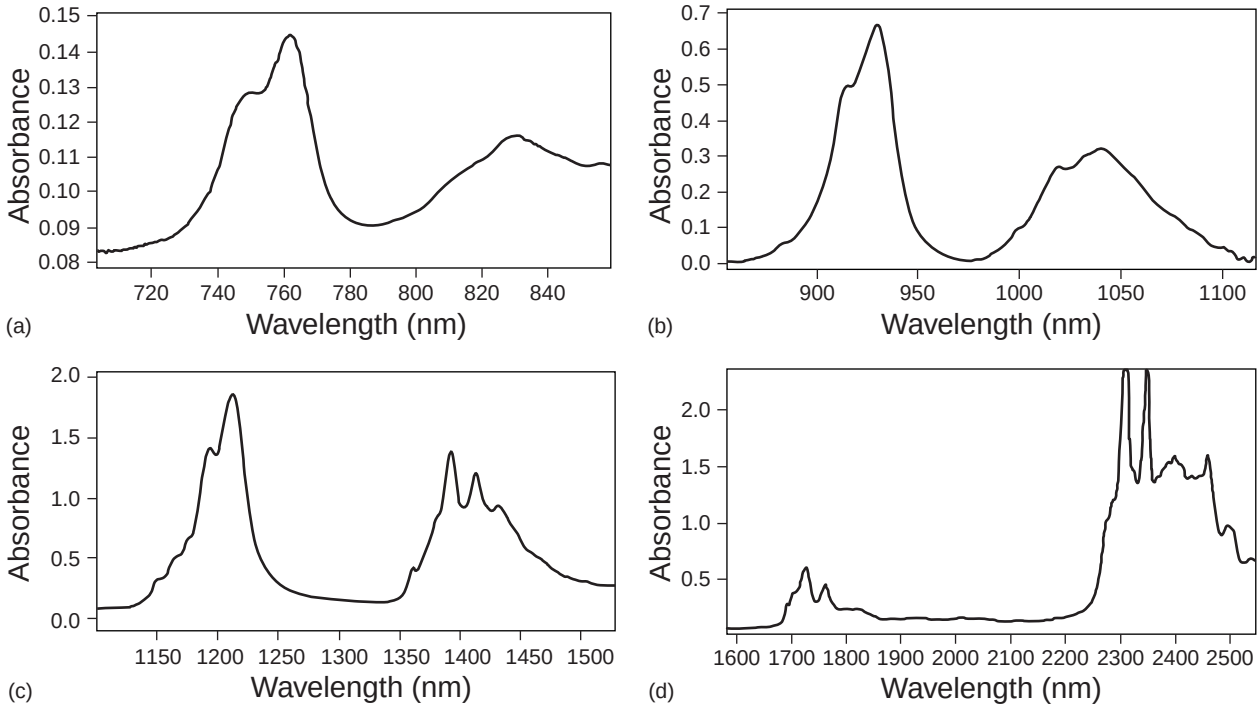

Figure 3 Illustration of aliphatic C−H (as predominantly methylene C−H str.) from *n*-decane. (a) The fourth overtone (5ν) with combination band, (b) the third overtone (4ν) with combination bands, (c) the second overtone (3ν) and combination bands, (d) the first overtone (2ν) and combination bands and a portion of the fundamental.

2.0% anharmonicity yields theoretical band positions as 2308 nm, 1723 nm, 1149 nm, and 862 nm (slightly closer to actual than the simple HO calculations).

The energy absorbed by a matrix consisting of polymeric compounds depends upon the chemical composition of the matrix, defined by the species (or types) of molecules present, the concentration of these individual species, and the interactions between the molecules in the matrix. In order for NIR (or any other vibrational spectroscopic measurement technique) to be valid one must be absolutely assured that different types of molecule absorb at unique frequencies. Owing to the broad-band nature of NIR spectra, consisting of overlapping combination and overtone bands, the individual species are not well resolved as they are in the MIR region. In addition, many compounds absorb NIR energy throughout the entire wavelength region making it difficult, if not impossible, to resolve a usable baseline clearly for simple peak height or peak area quantitative methods. This brings one to the same conclusion drawn by early investigators that the NIR region is not especially useful as a quantitative measurement technique. At the very least, novel techniques for spectral manipulation were required to interpret the poorly resolved bands and compensate for background interferences. The first assumption in spectroscopic measurements is that Beer's law relationship applies between a change in spectrometer response and the concentration of analyte material present in a sample specimen. The Bouguer, Lambert, and Beer relationship assumes that the transmission of a sample within an incident beam is equivalent to 10 exponent the negative product of the molar extinction coefficient (in mol^{-1} cm^{-1}), times the concentration of a molecule in solution (in mol^{-1}) times the pathlength (in cm) of the sample in solution. There are some obvious (and not so obvious) problems with this assumption. The main difficulty in the assumed relationship is that the molecules often interact, and the extinction coefficient may vary due to changes in the molecular configuration of the sample. The obvious temperature, pressure, and interference issues also create a less than ideal situation for the analyst. However, for many (if not most) analytical problems the relationship holds well enough.

Properties of the Bouguer, Lambert and Beer (Beer's law) relationship, Equation (11):

$$T = \frac{I}{I_0} = 10^{-\varepsilon cl} \tag{11}$$

where T is the transmittance (T), I_0 is intensity of incident energy, I is the intensity of transmitted light, ε is the molar extinction coefficient (in L mol^{-1} cm^{-1}), C is the concentration (in mol^{-1}), and l is the pathlength (in cm).

To simplify Equation (11) into its more standard form showing absorbance as a logarithmic term used to linearize the relationship between spectrophotometer response and concentration, gives the expression, Equation (12), as the relationship between absorbance (A) and concentration

$$A = -\log\left(\frac{I}{I_0}\right) = -\log(T) = \varepsilon cl \qquad (12)$$

The following statements hold true for what is most often termed Beer's law: (1) the relationship between T and concentration is nonlinear, yet (2) the relationship between absorbance and concentration is linear. Beer's law is the common basis for *quantitative* analysis. Knowledge of Beer's law allows us to calculate the maximum theoretical dynamic range for an instrument using a few simple mathematical relationships. A measured spectrum is the ratio of the sample and reference channels in transmission. As in the case of single beam instruments, in practice the final T spectrum from this device is given as, Equation (13)

$$T = \frac{\text{Sample}_T - \text{Dark}}{\text{Reference}_T - \text{Dark}} \qquad (13)$$

To convert this to an absorbance spectrum is completed using the relationship, Equation (14)

$$A = \log_{10}\left(\frac{1}{T}\right) \qquad (14)$$

2 REVIEW OF NEAR-INFRARED SPECTROSCOPY IN POLYMER AND RUBBER ANALYSIS

Table 2 shows the broad range of applications which are possible using the NIR spectral region. The list of applications which have been successfully implemented using the NIR and SW (short wave)/NIR techniques is vast. The NIR technique offers a possible solution to many other process analytical measurement requirements for polymers and rubbers than those listed in Table 2.

Table 2 Common applications for NIR in polymers

Matrix	Constituents
Copolymers	Vinyl acetate, PVC, PVA
Epoxy resins	Terminal CH_2 group
Ethylene–propylene blends	Percent polymer
Glycols	Water, OH number
Polymers (in general)	Cure kinetics, chemical/physical properties, phase separation, hydroxyl number
Polymer oxidation	Unsaturation and polyene formation due to thermal or photoinduced oxidation
Polymer mixtures	In situ quantitative analysis of individual polymer types
Polyolefins	Antioxidants
Polyols	Hydroxyl number
Polyurethane foams	Physical properties
Poly (ether urethane)	Bulk composition
Poly (ethylene terephthalate)	Crystallinity/morphology
Proteins	Water binding

PVC, polyvinyl chloride; PVA, polyvinyl alcohol.

2.1 Sorting Polymers

Polymer sorting for recycling operations has received attention. This technique provides automatic recognition of polymer type in a recycling operation and demonstrates the powerful capability of molecular spectroscopy for this type of application. A method has been developed for sorting plastics from nonplastics in waste recycling. Partial least squares (PLS) is used for data reduction in the classification of NIR spectral images.[15] A method for the discrimination of plastics from nonplastics in household waste on the basis of images taken in six wavelength ranges between 1100 and 2500 nm has been reported.[16] The use of this technique provided correct classification in 80% or better of the test cases.

Identification of plastics among nonplastics in mixed waste by remote sensing NIR imaging spectroscopy has been described.[17] In this application household waste was characterized by a sequence of images taken in four wavelength regions using NIR imaging spectrometry. Each sample was represented by a three-dimensional stack of NIR images. The method is insensitive to sample size and relative position within the camera cone of focus. The method was useful even with slight sample movement. An improved data processing method for reduction of experimental artifacts in a multivariate stack of remotely sensed NIR images is described.[18] The images provide data for real-time plastic identification and classification. Reference and dark current spectral images were used to correct the raw spectral images for normal fluctuations in background, illumination geometry, lamp source intensity, and optical transmission.

A device for on-line postconsumer package identification using NIR spectroscopy combined with a FuzzyARTMAP classifier has been demonstrated in an industrial environment.[19] A 30-cm wide conveyor belt moving at $1\,\text{m}\,\text{s}^{-1}$ carried the waste through an optical path defined by a 220-V, 500-W halogen lamp and collection optics mounted at one side of the belt and a metal reflector (60 cm from the collection optics) at the other. The collection optics were coupled by an optical fiber to a computerized f/2.8 grating spectrometer with an InGaAs detector. Radiation backscattered from opaque objects on the belt was detected, together with radiation

reflected back through transparent objects; the reflector provided the reference signal when the optical path was not obstructed. Spectra ($158\,s^{-1}$) were recorded in the range 900–1700 nm. Use of a neural-network classifier gave an accuracy of >97% in the classification of objects made of poly(ethylene terephthalate), polystyrene, PVC and cardboard, but was slightly less reliable at distinguishing between polyethylene and polypropylene, owing largely to the frequently composite nature of objects made from these materials.

The direct examination of NIR absorption bands can provide information about the inter- and intramolecular hydrogen bonding and other interactions of polymers, which are not only used to study, for example, the curing of epoxy resins, but can also be used in process control in the production of polymers and their blends.[20]

An optoacoustic FTIR (Fourier transform infrared spectroscopy) gas analyzer was used on-line with a fixed-bed pyrolysis reactor that consisted of a vertical quartz tube (1-cm i.d.) heated by a Lindberg furnace. Each sample (50–100 mg) was divided into portions that were packed in the tube separated from each other by quartz wool. A stream of He ($0.6\,L\,min^{-1}$) entered at the bottom of the tube and exited at the top, passing through a filter before it reached the analyzer. The reactor was also coupled to a gas chromatograph fitted with FID (flame ionization detection) or thermal conductivity detection. The results are tabulated. The major toxic products of polyacrylonitrile were HCN and NH_3; nylons produced NH_3, but only minor amounts of HCN. Both types of polymer produced CO. These by-products were monitored using the commercial device.[21]

Three classifiers, namely, adaptive resonance theory-based artificial neural network (ART-2a), multilayer feedforward backpropagation of error neural networks (MLFBP) and SIMCA (soft independent modeling of class analogies), were compared for sorting postconsumer plastics into classes (poly(ethylene terephthalate), poly(propylene), polyethylene, PVC, polystyrene) using NIR data.[22] The spectra were recorded using an advanced type of fast NIR spectrometer equipped with a InGaAs diode-array multichannel detector. With a minimum integration time of 6.3 ms, 158 complete NIR spectra (825–1700 nm) can be recorded per second. The performance of the classifiers was quantitatively evaluated by a repeated validation procedure based on the method of medium overlap between data clusters. MLFBP exhibited a slightly better discrimination power than ART-2a. Both types of neural network performed significantly better than SIMCA.

2.2 Polymer Reaction Monitoring

A variety of applications referring to the use of NIR for monitoring polymer reactions is available within the literature. Several examples of the current work in this area includes work by Siesler.[23] A review is presented on applications of optical-fiber remote NIR spectrometry, the examples quoted being the monitoring of the consumption of methyl methacrylate during its polymerization, the study of structural changes during the formation of poly(ethylene terephthalate) film, monitoring the reactive extrusion of styrene/maleic anhydride copolymer with 6-aminohexanoic acid at up to 300 bar and up to 400 °C, and the rapid identification of polymer waste.

The severe conditions (temperatures greater than 260 °C with turbulence) that obtained during the manufacture of polyoxyethylene and polyoxypropylene glycols necessitate remote analysis of the samples taken in situ; this can be accomplished by use of a commercial NIR polyol analyzer, which is connected by optical fibers to a stainless-steel immersion-sampling absorption cell (pathlength 6 mm). Thirty-two scans are co-added for each sample and corrected internally for absorption by the fibers. The establishment of calibration equations for hydroxyl number and acid value by multilinear least-squares regression on the basis of laboratory-analyzed samples is described; these gave standard errors of calibration of 0.41 and 0.28, respectively. A result was obtained in 1 min, and showed a relative error of 1.5% with respect to the laboratory value.[24]

On-line remote monitoring of thermoset resin cure using NIR external reflection spectroscopy is described by Xu et al.[25] NIR analysis in polymer reactions is given by Dallin.[26] The challenge of introducing NIR on-line in polyol production is described by Stromberg.[27] In-line process monitoring on polymer melts by NIR spectroscopy is given by Fischer et al.[28] In-line measurement of copolymer composition and melt index is explained by Hansen and Vedula.[29] In situ real-time monitoring of epoxy/amine kinetics by remote NIR spectroscopy is demonstrated by Mijovic et al.[30] A patent describing the process for measurement of the degree of cure and percent resin of glass-fiber-reinforced epoxy resin is given by Moe and Konicek.[31]

3 SAMPLE PREPARATION

A variety of sample presentation methods are available to the analyst. These include transmission (straight and diffuse), reflectance (specular and diffuse), transflectance (reflectance and transmission), and interactance (a combination of reflectance and transmission). In the selection of a pathlength for measurement, the following rules of thumb generally apply: for the ultraviolet (UV) region of 190–350 nm, a pathlength of 1 mm–10 cm is used, noting that 1 cm is most often used for hydrocarbon

Table 3 Sample presentation methods for polymers and rubbers

Sample type	Sample presentation	Comments
Liquids (lower viscosity)	T and DT	Maintain instrument within linear measurement range, that is generally less than 1.5 Au
Slurries (high viscosity)	T and DT	Maintain instrument within linear measurement range, that is generally less than 1.5 Au
Solids and smaller particles	DR	For DR use infinite optical thickness
Webs	DR and DT	For DR use infinite optical thickness
Pellets (large particles)	DR and DT	For DR use infinite optical thickness

Au = absorbance units.

measurements. For the SW/NIR region of 800–1100 nm, pathlengths ranging from 5 to 10 cm are typically used. For longwave (LW) NIR (LW/NIR) of 1100–2500 nm, common pathlengths for hydrocarbons include 0.1–2 cm.

The sample presentation geometries most often used in NIR measurements are given in Table 3. For this table the abbreviations used designate transmittance (T), diffuse transmittance (DT), and diffuse reflectance (DR). T and DT have identical sample presentation geometry, but the sample for DT is a series of particles allowing light to penetrate without significant backscatter. DR is used for samples of infinite optical thickness that will not transmit light if T or DT is used. The ideal surface for DR is a Lambertian or isotropic surface where light is reflected diffusely or evenly throughout a hemisphere above the surface.

Samples exhibit different optical characteristics which must be considered for optimum spectroscopic measurement. Standard clear samples or solutions are measured using transmission spectroscopy. Highly colored samples are generally measured using transmission spectroscopy unless the optical density exceeds the linear range of the measuring instrument. At this point either dilution or reducing the pathlength is desirable.

Fine scattering particulates of 5 to 25 times the measuring wavelength are measured using DT or DR methods. The scattering produced by some of the reflected light creates a pseudopathlength effect. This effect is compensated for by using scatter correction data processing methods for quantitative measurements.

Large scattering particulates that are in the order of 100 times the measuring wavelength in diameter present a challenge for measurements as the particles intercept the optical path at random intervals. Signal averaging can be employed to compensate for random signal fluctuations. Reflectance spectroscopy can be used to measure the size, velocity, and concentration of scattering particulates within a flowing stream.

Very high absorptivity (optically dense) materials with absorbances above 4–6 Au (absorbance units) are difficult to measure accurately without the use of double monochromator instruments which have stray light specifications below 0.0001% T. Measurements can be made with extremely slow scanning speeds and by opening the slits during measurement. These measurements should be avoided by the novice unless high-performance instrumentation and technical support are available.

4 INTERPRETIVE SPECTROSCOPY OF ORGANIC COMPOUND SPECTRA FOR POLYMERS AND RUBBERS IN THE NEAR-INFRARED REGION

Several tables and charts within this chapter describe the approximate band positions for organic compound spectra. Polymers and rubbers exhibit these band positions which will vary depending upon the molecular structure and the associated or attached chemical groups. Table 4 demonstrates the positions of major chemical groups encountered in analysis of polymer and rubber materials. The accompanying text describes the individual molecular vibrations and band assignments in much greater detail.

4.1 Aliphatic Hydrocarbons (*n*-Alkanes) including Polymeric Forms

The number of normal vibrations occurring within the IR region for saturated hydrocarbons equals $3N - 6$, where N is the sum of carbon plus hydrogen atoms, for example, *n*-hexane (C_6H_{14}) exhibits $3(20) - 6 = 54$ normal vibrational modes. Internal rotation and overlapping of bands complicate the interpretation of alkane (paraffinic) group spectra. C–H stretching vibrations can be expressed to a first approximation by the straightforward use of Hooke's law describing the vibration between two masses connected by a vibrating spring as previously presented.

The first consideration in the description of alkane group spectral interpretation would include the C–H stretching vibrations for methyl ($-CH_3$) and methylene ($-CH_2-$) groups. The methyl C–H stretching vibrations are found in two basic modes: the asymmetric (degenerate or out-of-phase) methyl C–H stretch near 3377 nm, and the symmetric (in-phase) methyl C–H stretch near 3482 nm. Three C–H stretching modes are observed

Table 4 C–H, N–H, and O–H stretch absorption bands for specific LW/NIR (1100–2500 nm) functional groups (first to fourth overtones)[32]

Structure	Bond vibration	Location (in nm) of first overtone	Location (in nm) of second overtone	Location (in nm) of third overtone	Location (in nm) of fourth overtone
ArCH (Aromatics)	C–H stretching	1680–1684	1143	875	715
. CH_2–CH_2. (Methylene)	C–H stretching	1726–1727	1212	927–930	761–764
. CH_3 (Methyl)	C–H stretching	1743–1762	1191–1194	911–915	745–748
. CH_3 (Methyl)	C–H combination	2020–2050	1347–1367	–	–
>C=O (Carbonyl)	>C=O stretching	2915	1920–1945	1420–1460	1140–1170
R–OH (Alcohols)	O–H stretching	1410–1455	940–970	–	–
ArOH (Phenols)	O–H stretching	1421–1470	940–980	–	–
HOH (Water)	O–H stretching	1440–1485	960	–	–
Starch	O–H stretching	1451	967	–	–
Urea	Sym. N–H stretching	1460	973–993	–	–
. $CONH_2$ (Primary amides)	N–H stretching	1463–1484	975–989	–	–
. CONHR (Secondary amides)	N–H stretching	1472	981	–	–
Cellulose	O–H stretching	1490	993	–	–
$ArNH_2$ (Aromatic amines)	N–H stretching	1493	995	–	–
. NH (Amines, general)	N–H stretching	1500	1000	–	–
Protein	N–H stretching	1511	1007	–	–

corresponding to three coupled C–H stretching vibration oscillators. Note that the absorption coefficient for methyl C–H stretch remains constant irrespective of increasing alkane chain length. The methylene C–H stretching vibrations are also found in two basic modes: the asymmetric (degenerate or out-of-phase) methylene C–H stretch near 3418 nm, and the symmetric (in-phase) methylene C–H stretch near 3505 nm.

Figure 2 illustrates the repetitive aliphatic (as predominantly methyl C–H stretching) information present within the NIR region for the fourth through first overtone spectra of isooctane. The repetitive molecular vibrational information within the spectra allow selection of the optimal pathlength for transmission work in the NIR spectral region. The fourth and third overtone measurements are taken using a 10.0-cm pathlength cell. The second and first overtone measurements were taken using 2.0-cm and 1.0-mm pathlength cells, respectively.

Figure 3 illustrates the repetitive aliphatic (as predominantly methylene C–H stretching) information present within the NIR region for the fourth through first overtone spectrum of *n*-decane. The repetitive molecular vibrational information within the spectra allow selection of the optimal pathlength for transmission work in these spectral regions. The fourth and third overtone measurements are taken using a 10.0-cm pathlength cell. The second and first overtone measurements were taken using 2.0-cm and 1.0-mm pathlength cells, respectively.

The two methylene stretching modes result from two coupled C–H stretching vibration oscillators. Increasing the chain length within alkane groups brings about a linear increase in the molecular absorption of methylene C–H stretching bands.

The carbon–hydrogen vibrational modes for alkanes also contain both methyl and methylene bending vibrations. The methyl C–H bending vibrations are found in two modes: the asymmetric (degenerate or out-of-phase deformation) methyl C–H bending near 6850 nm, and the symmetric (in-phase) methyl C–H bending near 7273 nm. This functional group vibrational frequency is dependent upon the atom to which the methyl group is attached. There is no coupling of methyl C–H bending modes to the rest of the system.

The methylene C–H bending vibrations are found in four basic modes: the symmetric (scissoring) methylene C–H bending with no coupling near 6873 nm, and the in-phase twisting vibration, where coupling is observed, near 7693 nm. The in-phase rocking vibration for alkanes with four or more carbons is not found within the NIR region. In addition other bending modes of methyl and methylene can be observed. The alkane carbon-to-carbon stretching, and carbon-to-carbon bending vibrations also occur as weak group frequencies.

4.2 Alkenes (Olefins) and Alkynes (Acetylenes)

The olefins include bands for saturated hydrocarbons as well as the olefinic C=C stretching and olefinic C–H stretching and bending vibrations.

4.3 Olefinic C–H Stretch

Fundamental olefinic C–H stretch occurs near 3280 nm. C–H stretch for an unsaturated carbon occurs below 3333 nm. The saturated carbon C–H bond stretching fundamental occurs above 3333 nm. Olefinic C–H stretch

can occur as an end group, for example R−R−C=C; or as an embedded group associated with other carbons on either end, for example, R−C=C−R. The absorption bands for these two olefinic types occur for the end group-type from 3226 to 3268 nm. Note that compounds containing the vinyl group ($-CH=CH_2$) have weak absorption bands near 3273 nm.

4.4 Alkyne C−H Stretch

The sp-type C−H stretch for C−H occurs at approximately 3030 nm. The sp^1-type C−H stretch associated with aromatic and olefinic carbons occurs near 3280 nm. The sp^1-type stretch for saturated carbons is observed at the high frequency ranges between 3344 and 3571 nm.

4.5 Olefinic C−H Bend

Two modes of bending occur in olefinic hydrogens, in-plane and out-of-plane. These bands do not appear with any importance within the NIR region and the following two paragraphs are for information only. In-plane bend represents a change in the C−H bond angle along the same plane, occurring near 6920–8333 nm as a weak and not very useful group frequency for the presence of olefinic hydrogens. Out-of-plane bending occurs as the C−H bonds vibrate outside the flat olefinic group plane. Out-of-plane bending occurs at approximately 10 000–15 385 nm and presents excellent IR bands for identifying the presence of olefinic groups; however the overtones do not occur within the NIR spectral region. The out-of-plane C−H bend related bands are group frequencies which are distinct, not being affected by conjugation or coupling with other vibrations in the olefinic molecule.

There is only one IR active C−H bend of the four possible modes for olefinic bending vibrations. The four bending modes include *twisting* or *rotational motion*; opposite twisting at the individual olefinic carbons producing a *torsion* at the C=C bond; wagging, or a motion of the hydrogen bonds; and two forms of wag, one where the two groups of two each hydrogens move in opposite planes causing a rotational motion around a common center of axis and the second wag where all four hydrogens move in or out of the plane together. The first type of wag is Raman only active and the second type of wag is the only IR active type, though not observed in the NIR region.

4.6 Aromatic Systems

Aromatic systems produce strong and sharp IR bands owing to the relatively rigid molecular structures (see Figure 4). There are multiple vibrational modes present

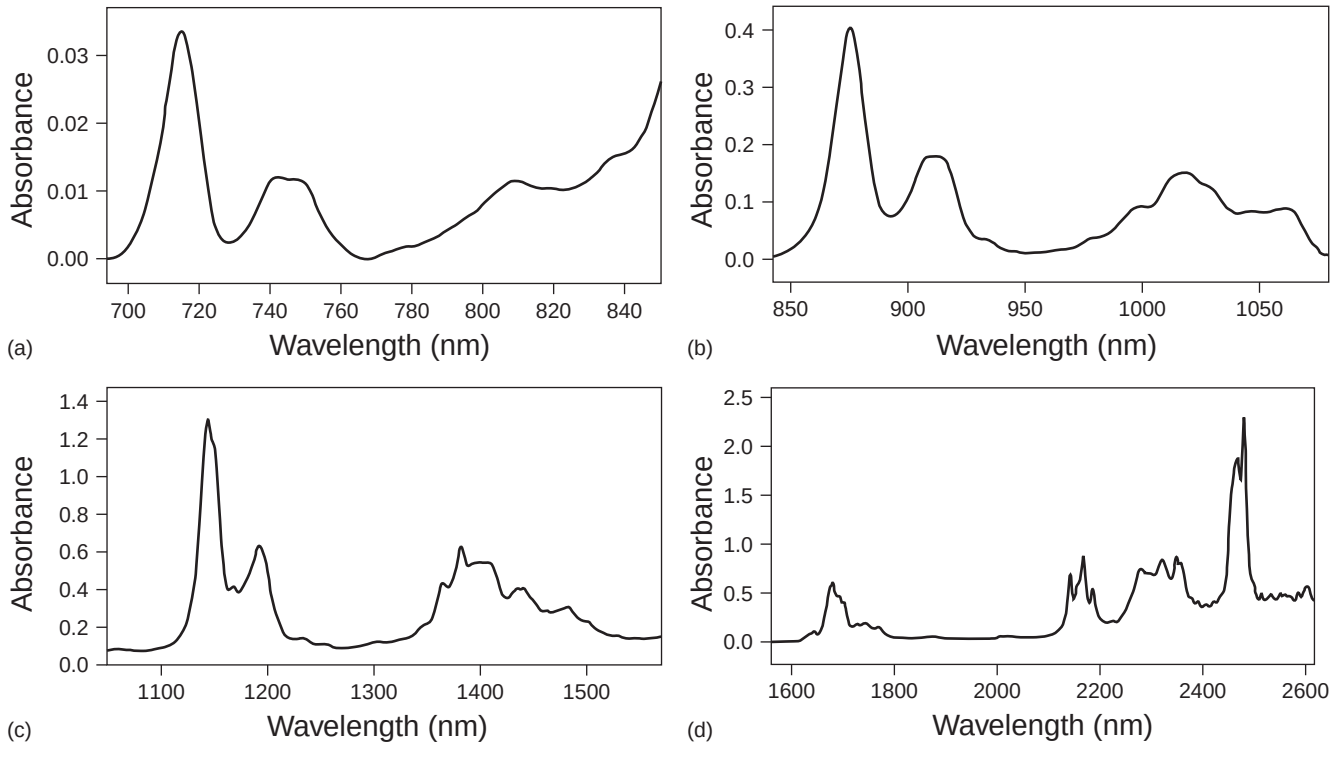

Figure 4 Aromatic C−H stretching as illustrated using toluene NIR spectra. (a) The fourth overtone (5ν) with combination bands, (b) the third overtone (4ν) with combination bands, (c) the second overtone (3ν) with combination bands, (d) the first overtone (2ν) and combination bands and a portion of the fundamental vibrational band.

in aromatic systems including carbon–hydrogen stretching, in-plane bending, and out-of-plane bending. Carbon–carbon interactions include ring stretching and ring bending modes. Internal vibrations caused by groups associated with the aromatic system also exhibit vibrational group frequencies. The total vibrational modes for aromatic systems is equal to $3N - 6$ vibrations, where N is the number of carbons in the ring structure. Carbon–hydrogen stretching vibrations occur from approximately 3226 to 3333 nm. IR spectra of aromatics using high-resolution instrumentation show a pair of composite bands near 3257 nm. These composite bands arise as the combined interactions of several stretching modes.

The carbon–carbon ring stretching occurs as a vibrational pair near 6250 nm. This pair of frequencies is termed a degenerate pair and results from two stretching modes of the carbon–carbon ring stretching. However owing to Fermi resonance, one of the vibrations (near 6270 nm) is not observed in the IR spectrum. When there is substitution on the aromatic ring, the degeneracy is lost and the pair of bands occur near 6250 nm. The 6329-nm band will be more intense if there is conjugation with the aromatic system by an electron withdrawing group, or if substitution by O or N has occurred. Note that the occurrence of overtones for this band are observed near 3228–3481 nm (first overtone), 2173–2426 nm (second overtone), and 1645–1899 nm (third overtone). Paradisubstitution of a benzene ring by two identical groups yields to the center of symmetry rule, and thus the carbon-to-carbon ring stretching is IR forbidden. It follows that if the paradisubstituted groups are not identical, some weak bands will occur near the 6250-nm region.

Figure 4 illustrates the repetitive aromatic (as predominantly aromatic C–H stretching with some slight methyl C–H stretching) information present within the NIR region for the fourth through first overtone spectrum of toluene. The repetitive molecular vibrational information within the spectra allow selection of the optimal pathlength for transmission work in these spectral regions. The fourth and third overtone measurements are taken using a 10.0-cm pathlength cell. The second and first overtone measurements were taken using 2.0-cm and 1.0-mm pathlength cells, respectively.

4.7 Carbonyl Compounds (>C=O)

Compounds containing carbonyl groups exhibit frequencies over a broad range from 5128 to 6452 nm (1950–1550 cm^{-1}), although most often carbonyl associated groups occur near 5831 nm (1715 cm^{-1}). Overtone associated bands for carbonyl compounds occur near 2916 nm (first overtone), 1944 nm (second overtone), and 1458 nm (third overtone). The position or frequency of the carbonyl C=O stretching vibration is affected by the isotope effects and mass change of substituted groups, bond angles of substituted groups, electronic (resonance and inductive) effects, and interactions of these effects. Substituents with higher mass decrease the C=O stretch frequency; increasing the mass or bond angles of substituents also decreases the frequency by up to 40 cm^{-1} at 1715 cm^{-1}; this is equivalent to an increase of 139 nm at 5831 nm. Similarly decreasing the substituent mass, or bond angles between the carbonyl carbon and its substituents increases the frequency by 25 cm^{-1} above the nominal 1715 cm^{-1} carbonyl C=O stretch frequency, equivalent to a decrease in wavelength of 84 nm at 5831 nm. More electronegative (electron withdrawing) substituents will increase the carbonyl carbon–oxygen stretch frequency by up to 100 cm^{-1} above the 1715 cm^{-1} nominal frequency: a decrease of 321 nm at 5831 nm. Conjugation of the carbonyl group to aromatic or olefinic groups tends to lower the frequency for both C=O and C=C by 30–40 cm^{-1} depending on the ring size of the substituent. Electronic effects such as resonance and inductive type produce a double-bond character and increase the frequency of the carbonyl stretch.

4.8 Amides (R–C=ONH–X)

Amides are a special case of carbonyl compounds which are subject to frequency shifts due to hydrogen bonding. The frequencies for dilute solutions of amides can vary by as much as 250–275 nm (75 cm^{-1}) at ~5880 nm compared with the same molecule with hydrogen bonding (such as a Nujol mull preparation). Dilute solutions of amides exhibit carbonyl frequencies as follows. Primary amides ($R–C=O–NH_2$) occur at 5764 nm (fundamental at 1735 cm^{-1}), 3054 nm (first overtone), 2094 nm (second overtone), 1498.6–1729.2 nm (third overtone). Secondary amides (R–C=O–NHR) occur at 5865 nm (fundamental at 1705 cm^{-1}), 2991.2–3225.8 nm (first overtone), 2013.7–2248.3 nm (second overtone), 1524.9–1759.5 nm (third overtone). Tertiary amides, $R–(C=O)–NR_2$, bands are found at 6042 nm (fundamental at 1655 cm^{-1}), 3081.4–3323.1 nm (first overtone), 2074.4–2316.1 nm (second overtone), and 1570.9–1812.6 nm (third overtone).

These respective hydrogen-bonded vibrations occur at shifted frequencies from those above. For primary amides vibration occurs at 5882–6061 nm (fundamental at 1700–1650 cm^{-1}), 2999.8–3333.6 nm (first overtone), 2019.5–2323.4 nm (second overtone), and 1529.3–1818.3 nm (third overtone). For secondary amides vibrations occur at 5952–6116 nm (1680–1635 cm^{-1}), 3035.5–3363.8 nm (first overtone), 2043.5–2344.5 nm (second overtone), and 1547.5–1834.8 nm (third overtone). For tertiary amides, vibrations occur at 6024 nm

(1660 cm^{-1}), 3072.2–3313.2 nm (first overtone), 2068.2–2309.2 nm (second overtone), and 1566.2–1807.2 nm (third overtone). The N–H stretch for amides occurs as two bands between 2976 and 3150 nm (3360 and 3175 cm^{-1}). Geometrically, two forms of amide exist as either the *trans* or *cis* forms. The *trans* form demonstrates a frequency of 3008–3053 nm (3325–3275 cm^{-1}), 1534.1–1679.2 nm (first overtone), 1032.7–1170.3 nm (second overtone), and 782.1–915.9 nm (third overtone). The *cis* form demonstrates a frequency of 3140–3190 nm (3185–3135 cm^{-1}), 1601.4–1754.5 nm (first overtone), 1078.1–1222.8 nm (second overtone), and 816.4–957.0 nm (third overtone). The geometries included are as R–(C=O)–(N–R′)–H where the R and R′ are either in the *trans* or *cis* configuration. The term "amide I" band refers to the C=O carbonyl stretch and the nominal C=O stretch frequency is 5831 nm (1715 cm^{-1}), 2973.8–3207.1 nm (first overtone), 2002.0–2235.2 (second overtone), and 1516.1–1749.3 nm (third overtone). The amide form of the carbonyl has an electronegative N atom, lowered by resonance with the nitrogen electrons and lowered by hydrogen bonding. The C=O nominal frequency for the amide I stretch is approximately 6061 nm (1650 cm^{-1}), 3091.1–3333.6 nm (first overtone), 2080.9–2323.4 nm (second overtone), and 1575.9–1818.3 nm (third overtone).

The term "amide II" refers to the NH_2 scissoring associated with primary amides, R–(C=O)–NH_2. The amide II frequency results from the combined interactions of the C–N stretch and the N–H in-plane bend; these vibrations interact to form bands in the 6494 nm (1540 cm^{-1}) region with overtones at 3311.9–3571.7 nm (first overtone), 2229.6–2489.4 nm (second overtone), and 1688.4–1948.2 nm (third overtone). In addition bands are formed in the 7634–8197 nm (1310–1220 cm^{-1}) region with overtones at 3893.3–4508.4 nm (first overtone), 2621.0–3142.2 nm (second overtone), and 1984.8–2459.1 nm (third overtone). A band at 6410–6515 nm (1560–1535 cm^{-1}) indicates a *trans* substituted amide; this band is lost in the *cis*-amide configuration. Overtones for this band occur at 3269.1–3583.3 nm (first overtone), 2200.8–2497.4 nm (second overtone), and 1666.6–1954.5 nm (third overtone).

4.9 Hydrogen Bonding (X–H···Y)

A hydrogen-bonded molecular system has four possible modes: O–H stretch found at 2941–4000 nm (3400–2500 cm^{-1}) with overtones occurring at 1499.9–2200 nm (first overtone), 1009.7–1533.3 (second overtone), and 764.7–1200 nm (third overtone). R–O–H in-plane bend at 7143–7692 nm (1400–1300 cm^{-1}) with overtones occurring at 3642.9–4230.6 nm (first overtone), 2452.4–2948.6 nm (second overtone), and 1857.2–2307.6 nm (third overtone). The R–O–H out-of-plane bend at 15 152 nm (660 cm^{-1}); no vibrations are found in the NIR region. The –O–H···O stretch is at 57 143 nm (175 cm^{-1}); no vibrations are found within the NIR region.

Specific frequencies for molecules containing O–H with hydrogen bonding include alcohols (R–O–H) with n(O–H) = 2981 nm (3355 cm^{-1}) with overtones at 1520.3–1639.6 nm (first overtone), 1023.5–1142.7 nm (second overtone), and 775.1–894.3 nm (third overtone); carboxylic acids (–COOH) with n(O–H) = 3333 nm (3000 cm^{-1}) with overtones at 1699.8–1833.2 nm (first overtone), 1144.3–1277.7 nm (second overtone), and 866.6–999.9 nm (third overtone); amines (R–NH–X) with n(N–H) = 3000 nm (3300 cm^{-1}), with overtones occurring at 1530–1650 nm (first overtone), 1030–1150 nm (second overtone), and 780–900 nm (third overtone).

In general hydrogen bonding has several main effects on frequency of X–H bands: (1) it lowers the frequency of the X–H stretch compared with the gas phase and dilute liquid phase of a molecule; (2) it raises the X–H bend frequency; (3) it broadens all O–H stretch bands by up to 300–475 nm (300–500 cm^{-1}); (4) it intensifies all bending and stretching X–H bands.

4.10 The O–H Stretch

It is commonly accepted among IR spectroscopists that any fundamental band occurring above 2857 nm (3500 cm^{-1}) can be confidently assigned to the O–H stretch group. A sharp band exists for the stretching frequency of O–H for the IR spectrum of O–H containing systems as delineated in the following text: primary alcohols (–CH_2–OH) exhibit fundamental stretching vibration of (O–H) = 2740 nm (3650 cm^{-1}), 1397.4–1507.0 nm (first overtone), 940.7–1050.3 nm (second overtone), and 712.4–822.0 nm (third overtone); secondary alcohols (R–CHOH–R′) exhibit fundamental stretching vibration of (O–H) = 2762 nm (3620 cm^{-1}), 1408.6–1519.1 nm (first overtone), 948.3–1058.8 nm (second overtone), and 718.1–828.6 nm (third overtone). Tertiary alcohols (R–CR′OH–R″) demonstrate fundamental stretching vibration of (O–H) = 2770 nm (3610 cm^{-1}), 1412.7–1523.5 (first overtone), 951.0–1061.8 nm (second overtone), and 720.2–831.0 nm (third overtone). Phenols fundamental stretching vibration of (O–H) = 2762–2793 nm (3620–3580 cm^{-1}), 1408.6–1519.1 nm (first overtone), 948.3–1070.7 nm (second overtone), and 718.1–837.9 nm (third overtone). Carboxylic acids have fundamental stretching vibration of (O–H) = 2817–2857 nm (3550–3500 cm^{-1}), 1436.7–1571.4 nm

(first overtone), 967.2–1095.2 nm (second overtone), and 732.4–857.1 nm (third overtone).

To distinguish between primary, secondary, and tertiary alcohols using IR spectra requires the use of these bands as well as the fundamental stretching vibration of (C−O) at 8547–9615 nm (1170–1040 cm^{-1}), 4903.7–5288.3 nm (first overtone), 3301.2–3685.8 nm (second overtone), and 2499.9–2884.5 nm (4 n); this band is not observed in the NIR region. The O−H in-plane fundamental bend occurs as (O−H) at 7143–7692 nm (1400–1300 cm^{-1}), 3642.9–4230.6 nm (first overtone), 2452.4–2948.6 nm (second overtone), and 1857.2–2307.6 nm (third overtone). The C−O−H out-of-plane fundamental bend as (O−H) occurs at 13 158–15 385 nm (760–650 cm^{-1}); it is not observed in the NIR. Carboxylic acids (−COOH) form dimers (hydrogen bonding between molecules as (−C=O···H−O−C−) in condensed states that exhibit four distinct frequencies: (1) a very broad, very strong O−H stretch at approximately 3333 nm (3000 cm^{-1}, at a lower frequency than alcohols), 1699.8–1833.2 (first overtone), 1144.3–1277.7 (second overtone), and 866.6–999.9 nm (third overtone); (2) summation tones as one to three weak bands near 3846 nm (2600 cm^{-1}), 1961.5–2115.3 nm (first overtone), 1320.5–1474.3 nm (second overtone), and 1000–1153.8 nm (third overtone); (3) carbonyl (C=O) stretch vibration near 5797 nm (1725 cm^{-1}), 2956.5–3188.4 nm (first overtone), 1990.3–222.2 nm (second overtone), 1507.2–1739.1 nm (third overtone); (4) a weak broad band indicating a dimer condition near 10 471 nm (955 cm^{-1}) is not observed in the NIR.

4.11 N−H Group Frequencies and Nitrogen-containing Compounds

The average frequencies for free N−H groups include those given in detail in the following paragraph.

Amide groups include primary amides, −(C=O)−NH_2, fundamental stretching vibration of (N−H) = 2857 nm (3500 cm^{-1}), 1457.1–1571.4 nm (first overtone), 980.9–1095.2 nm (second overtone), 742.8–857.1 nm (third overtone). For primary amides a band also occurs at 2933 nm (3410 cm^{-1}), 1495.8–1613.2 nm (first overtone), 1007.0–1124.3 nm (second overtone), and 762.6–879.9 nm (third overtone); and secondary amides, −(C=O)−NHC, fundamental stretching vibration of (N−H) = 2899 nm (3450 cm^{-1}), 1478.5–1594.5 nm (first overtone), 995.3–111.3 nm (second overtone), 753.7–869.7 nm (third overtone).

Alkyl (aliphatic-NH-X) groups include alkyl-NH_2, fundamental stretching vibration of (N−H) = 2946 nm (3395 cm^{-1}), 1502.5–1620.3 nm (first overtone), 1011.5–1129.3 nm (second overtone), and 766–883.8 nm (third overtone); and 3003 nm (3330 cm^{-1}), 1531.5–1651.7 nm (first overtone), 1031.0–1151.2 nm (second overtone), and 780.8–900.9 nm (third overtone); alkyl-NH-X, fundamental stretching vibration of (N−H) = 2981 nm (3355 cm^{-1}), 1520.3–1639.6 nm (first overtone), 1023.5–1142.7 nm (second overtone), 775.1–894.3 nm (third overtone).

For aryl (aromatic-NH-X) groups, the N−H stretch fundamentals occur as aryl-NH_2, fundamental stretching vibration of (N−H) = 2869 nm (3485 cm^{-1}), 1463.2–1578.0 nm (first overtone), 985.0–1099.8 nm (second overtone), and 745.9–860.7 nm (third overtone); and a second band is found near 2946 nm (3395 cm^{-1}), 1502.5–1620.3 nm (first overtone), 1011.5–1129.3 nm (second overtone), and 766–883.8 nm (third overtone). Aryl-NH-X, fundamental stretching vibration of (N−H) = 2911 nm (3435 cm^{-1}), 1484.6–1601.1 nm (first overtone), 999.4–1115.9 nm (second overtone), 756.9–873.3 nm (third overtone). Note that the N−H frequency is lower and less intense than the O−H frequency; the hydrogen-bonding effects for N−H are weaker and of less effect than in O−H.

Primary amines (R−NH_2) demonstrate two N−H stretching bands due to coupling: the in-phase (or symmetric), and the out-of-phase (or asymmetric); aliphatic amines, (C_nH_{2n+1})-NHX, demonstrate medium strength in-phase frequencies at 2999–3053 nm (3335–3275 cm^{-1}), 1529.5–1613.2 nm (first overtone), 1029.7–1124.3 nm (second overtone), and 779.7–879.9 nm (third overtone). The out-of-phase medium strength frequency occurs at 2933–2985 nm (3410–3350 cm^{-1}), 1495.8–1641.8 nm (first overtone), 1007.0–1144.3 nm (second overtone), 762.2–895.5 nm (third overtone).

Aromatic amines exhibit two medium strength frequencies, an in-phase frequency at 2920–3003 nm (3425–3330 cm^{-1}), 1489.2–1651.7 nm (first overtone), 1002.5–1151.2 nm (second overtone), and 759.2–900.9 nm (third overtone), and an out-of-phase frequency at 2849–2950 nm (3510–3390 cm^{-1}), 1453.0–1622.5 nm (first overtone), 978.2–1130.8 nm (second overtone), 740.7–885.0 nm (third overtone). The scissoring for primary amines occurs as a medium to strong broad band at 6042–6289 nm (1655–1590 cm^{-1}), 3081.4–3459.0 nm (first overtone), 2074.4–2410.8 nm (second overtone), and 1570.9–1886.7 nm (third overtone). The NH_2 wagging is a strong broad band near 12 500 nm (800 cm^{-1}) and is not observed in the NIR region.

Secondary amines (R′R″−NH) exhibit a single broad N−H stretch vibrational frequency near 3021 nm (3310 cm^{-1}), 1540.7–1661.6 nm (first overtone), 1037.2–1158.1 nm (second overtone), and 785.5–906.3 nm (third overtone). An N−H bend occurs as a weak-to-medium strength band near 6667 nm (1500 cm^{-1}), 3400.2–3666.9 nm (first overtone),

2289.0–2555.7 nm (second overtone), 1733.4–2000.1 nm (third overtone).

To separate spectral features for the various amines, that is primary, secondary, and tertiary, it is helpful to point to obvious differences between the groups. For example, the number of N–H stretches for the three types are primary (2), secondary (1), and tertiary (0). For N–H bending vibrations, primary amines have a band near 6042–6289 nm (1655–1590 cm^{-1}); secondary amines, near 6667 nm (1500 cm^{-1}); and this feature is absent in tertiary amines.

The C–H stretch vibration for CH_2 is to be found for all three types of amines from 3503 to 3683 nm (2855–2715 cm^{-1}), 1786.5–2025.7 nm (first overtone), 1202.7–1411.8 nm (second overtone), and 910.8–1104.9 nm (third overtone). For primary amines, the C–H stretch will be found at higher frequencies nearer to 3448 nm (2900 cm^{-1}), 1758.5–1896.4 nm (first overtone), 1183.8–1321.7 nm (second overtone), and 896.5–1034.4 nm (third overtone).

Amine salts as $R-NH_3^+$ and $R'R''NH_2^+$ exhibit several weak bands near 3571–4545 nm (2800–2200 cm^{-1}), 1821.2–2499.8 nm (first overtone), 1226.0–1742.3 nm (second overtone), 928.5–1363.5 nm (third overtone). These salts exhibit a weak-to-medium, very broad band, centered near 4000 nm (2500 cm^{-1}), 2040–2200 nm (first overtone), 1373.3–1533.3 nm (second overtone), and 1040.0–1200 nm (third overtone). This broad band is lower in frequency than the C–H stretch and up to 314 nm (500 cm^{-1}) in width at the baseline.[33]

4.12 Amines (R–NH–X)

Amines are polar compounds, and both primary and secondary amines can form intermolecular hydrogen bonding. All classes of amines are water soluble (up to six carbon atoms) and will form hydrogen bonding in aqueous solutions. NIR absorptions for amines occur in decreasing wavelength order as:

- 2020 nm — primary N–H combination band
- 2000–2050 nm — N–H stretch and N–H bend combination band
- 1990–2020 nm — free and bonded primary amines
- 1960 nm — combination of N–H stretch and NH_2 bending, for example meprobamate
- 1500–1550 nm — primary amines first overtone aliphatic N–H stretch
- 1530 nm — N–H stretch overtone for aliphatic-NH system
- 1450 nm — cyclic amines, e.g. pyrroles, indoles, and carbazoles show intense first overtone N–H stretch
- 1010–1020 nm — aliphatic (alkyl-) N–H stretch second overtone
- 995–1015 nm — aromatic (aryl-) N–H stretch second overtone.

Whetsel[7] provided NIR spectra for primary aromatic amines. He provides spectra for aniline and 39 of its ring-substituted derivatives, giving special attention to N–H bands located between 1000 and 2000 nm.

4.13 Amides (–C=O–NX′X″, also shown as –NHCO–)

Amide absorptions are as follows:

- 2817–3279 nm — fundamental N–H for acyl- group of amide (3550–3050 cm^{-1}) as (R–C=O–)
- 5917–6061 nm — the C=O stretch fundamental (1690–1650 cm^{-1})
- 2000 nm (5000 cm^{-1}) — combination band
- 1500 nm (6667 cm^{-1}) — first overtone
- 1000–1050 nm — primary amides (–C=O–NH_2)
- 1450–1500 nm — primary amides
- 2850–2950 nm — primary amides

5 POLYMERS

5.1 Types of Polymer

Polymers are large molecules consisting of repeating units. The repeating units are referred to as monomers. Natural polymers exist such as starches and polysaccharides. Synthetic polymers are commonly termed plastics and are used for myriad products. Polymers vary in molecular formula and molecular weight due to variation in the number of repeating units. Polymer backbone structures often have attached molecular groups. The molecular arrangement of these groups determines the stereochemical configuration for any given polymer. If all the attached groups are in the identical position along the polymer backbone chain the polymer is in an *isotactic* configuration. If the attached groups alternate in their attached positions with a regular pattern the *syndiotactic* configuration is ascribed. When attached groups are randomly attached to the polymer backbone the polymer is

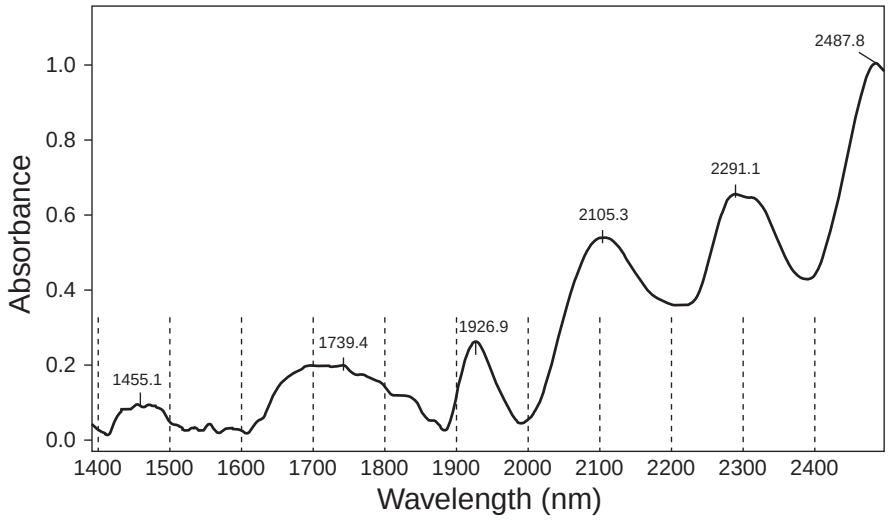

Figure 5 Natural product polymers exist as proteins (amino acid polymers), cellulose (carbohydrate polymers) with high molecular weights, and starches (carbohydrate polymers with intermediate molecular weight). The NIR spectrum shown is that of a plant starch (32 cm^{-1} resolution).

said to have an *atactic* configuration. The isotactic configuration represents the most crystalline (rigid) of the configuration types.

Copolymers involve the use of two or more monomer types in a single backbone structure to achieve specific material performance properties. Structures are classified as *alternating* (–A–B–A–B–A–B–), *random* (–A–A–B–A–B–B–A–), *block* (–A–A–B–B–B–B–A–A–), and *graft* (–A–A–A–A–A–A–A–).

B B
B B

Polymerization involves the reaction of the monomer building blocks to make polymers. The polymerization reaction types involve *addition* reactions and *condensation* reactions. Addition reactions typically involve the use of ethene to form polyethene and polyethylene. Condensation reactions typically involve different reaction products reacting to form a heteropolymer and a small molecular by-product. The reaction of 1,6-diaminoethane and hexanedioic acid to form nylon and water is a classic example. Polymers made from one monomer type are termed *homopolymers* and those formed from two different monomers are referred to as *copolymers*.

Typical structures of polymers and their respective NIR spectra are shown.

Dacron **(1)** is a polyester of ethylene glycol and terephthalic acid, known as 'Mylar'. Delrin **(2)** is a paraformaldehyde polymer. Melmac or Bakelite **(3)** is a phenolic copolymer with formaldehyde. Molecular structure **(4)** is a ureaformaldehyde resin.

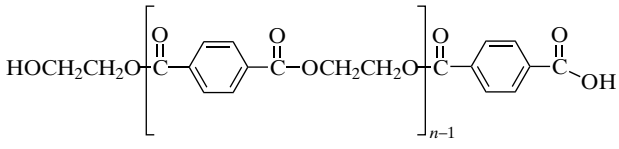

(1)

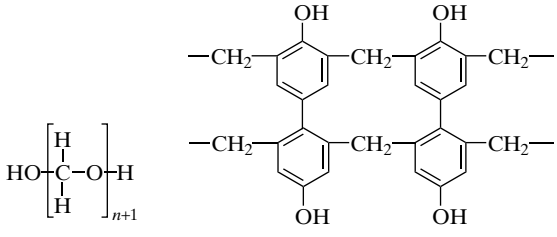

(2) **(3)**

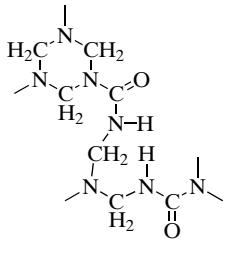

(4)

Figure 5 shows the NIR spectrum of a natural product plant starch and Figure 6 shows the NIR spectrum of pure Rayon.

Neoprene **(5)** is a polymeric chloroprene. Nylon 66 **(6)** is a polyamide. Orlon **(7)** is composed of multiple units of

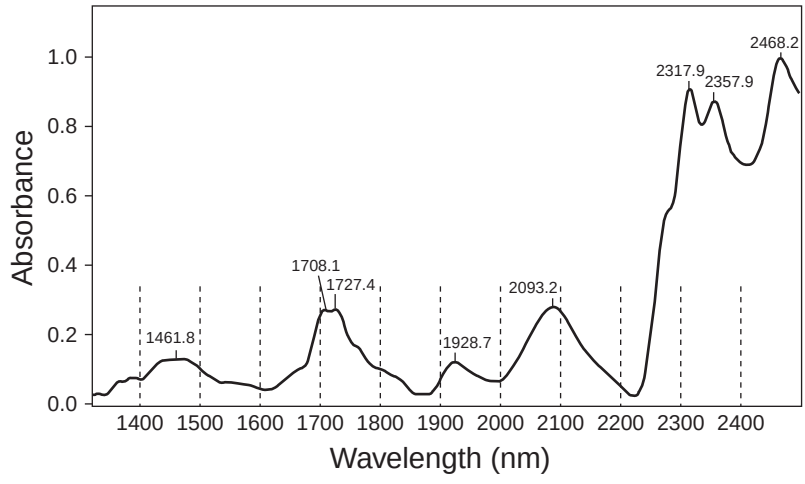

Figure 6 NIR spectrum of pure rayon (as cellulose) at 32 cm^{-1} resolution.

acrylonitrile (vinyl cyanide). Polyethylene **(8)** is a typical polyolefin. Polyethylene oxide **(9)** and polypropylene **(10)** are polyolefins.

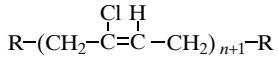

(5)

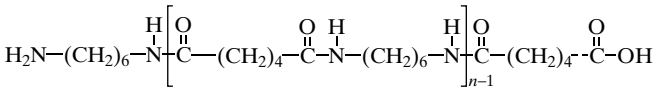

(6)

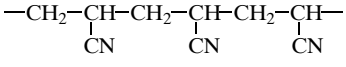

(7)

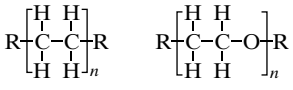

(8) **(9)**

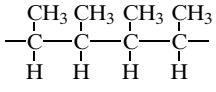

(10)

Figure 7 shows the NIR spectrum of atactic polypropylene and Figure 8 shows the NIR spectrum of polypropylene plus polyethylene copolymer.

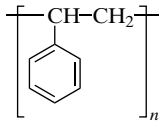

(11)

Structure **(11)** is polystyrene monomer with Figures 9 and 10 showing the NIR spectra of atactic polystyrene and crystalline polystyrene, respectively.

Teflon **(12)** is a polyfluorocarbon. Polyacrylic acid **(13)** and polymethacrylic acid **(14)** are water-soluble polymers with Figure 11 depicting the NIR spectrum of **(13)**.

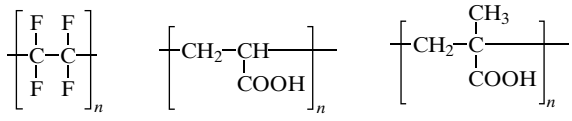

(12) **(13)** **(14)**

5.2 Spectral Correlation Chart for Polymers

Typical polymer spectra exhibit similar NIR bands and positions which are represented by the chart in Table 5.

5.3 Qualitative Analysis of Polymers

Qualitative polymer analysis using NIR spectroscopy involves the identification or verification of unknown substances for quality control and problem solving; and the measurement of reactants and reaction products during batch reactions; and the determination of degree of mixing and of quality index of manufactured products. Identification or verification of an unknown material involves sampling the unknown, measuring the NIR spectrum of the material, and comparing the resultant spectrum to a standard library of known spectra. The

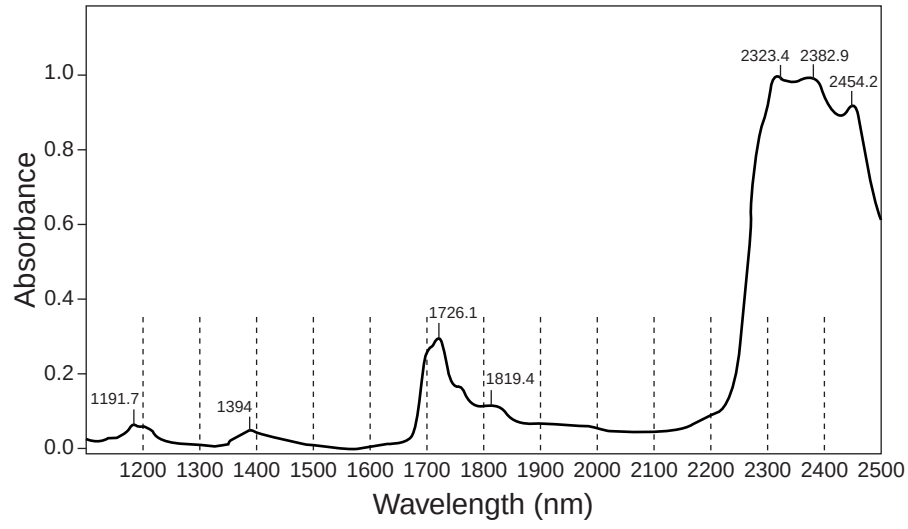

Figure 7 NIR spectrum of atactic poly(propylene) at 32 cm^{-1} resolution.

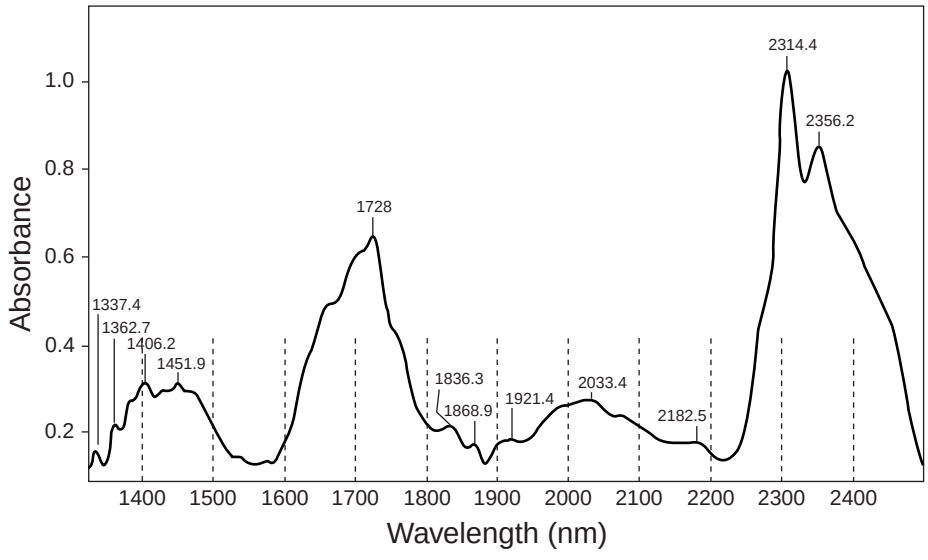

Figure 8 NIR spectrum of the poly(propylene) plus poly(ethylene) copolymer at 32 cm^{-1} resolution.

method of comparison often termed *spectral matching* or *spectral searching* is described in more detail within section 7 of this article. Once the unknown is identified the spectrum from it can be archived for future retrieval and comparison, or simply retained for quality reasons.

Measurement of reactants during processing is accomplished for polyethylene and polypropylene (olefins) production by measuring the solution polymerization mixture for the disappearance of double bonds from the polyethene materials. Workers have reported the occurrence of a terminal double bond combination band at 2110 nm and a first overtone band at 1620 nm.[9] Note that the NIR technique is useful for measuring the ratios of methyl, methylene, and methine (weak) groups although their location within the polymer structure cannot be determined using this technique. Proton NMR would prove to be more useful if configuration studies were required.

Activators in condensation reactions can be monitored using NIR when the reactants are in a liquid form. During the production of PVC the activator molecules can be monitored. IR-active reactants and reaction products are often monitored using NIR. The solids content or degree of conversion for organic molecules can be monitored as the reaction progresses. This type of measurement allows reaction kinetic studies or end-point monitoring.

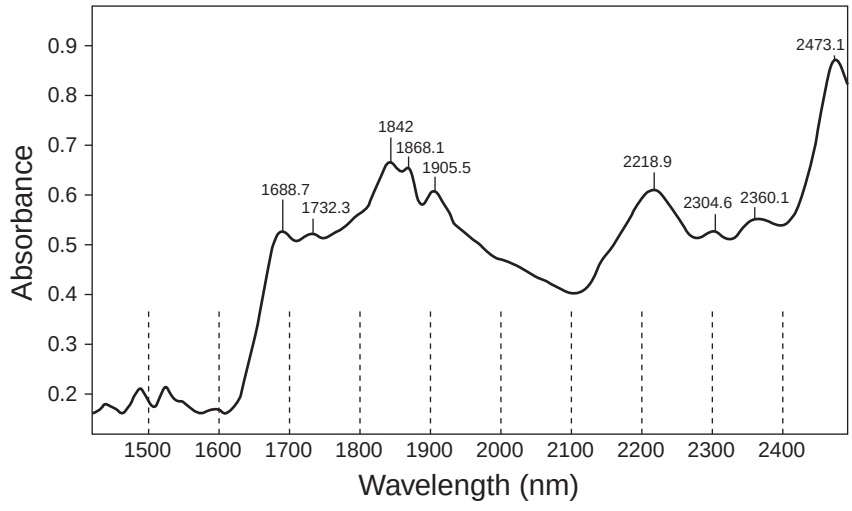

Figure 9 NIR spectrum of atactic poly(styrene) at 32 cm^{-1} resolution.

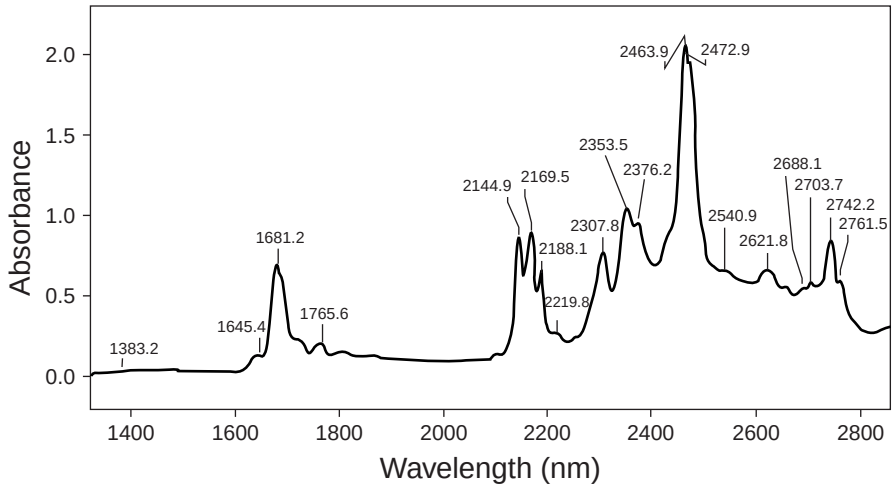

Figure 10 NIR spectrum of crystalline poly(styrene) at 16 cm^{-1} resolution.

Degree of mixing in a batch reaction is determined by monitoring a point or multiple points within a mixing chamber for a change in spectral signal. When the mixing is just beginning there will be large spectral variation; after the materials are well mixed the measured spectral variations will become slight. When the components are well mixed the spectral variation during further mixing will be unchanged.

The determination of the quality index of manufactured products is accomplished using spectral searching or comparison. A library of high-quality finished product spectra are maintained for comparison purposes. The sample to be tested is compared during the manufacturing process to the reference library set. When the predetermined match index (MI) value is achieved during manufacturing, the quality index is recorded. Further testing is sometimes completed for verification purposes, but the concept of rapid nondestructive quality index measurement allows more efficient production costs with higher quality products.

5.4 Quantitative Analysis of Polymers

Quantitative analysis of polymers involves the use of classic spectroscopic quantitative practices or the use of multivariate calibration methods described in section 7. The technique is useful in the production of many polymers as a quantitative monitoring of reactants and reaction products during the manufacturing process. The polymerization process can be optimized for specific yields of reaction products using quantitative methods. The literature describes NIR methods for monitoring or measuring components from a variety of polymerization reactions including polyolefins, polyesters, paraformaldehydes,

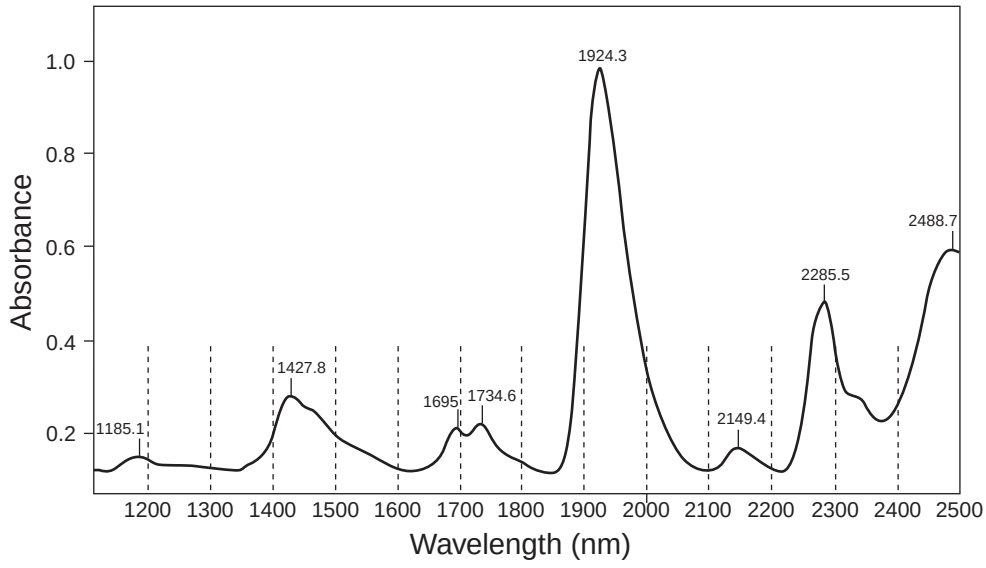

Figure 11 NIR spectrum of poly(acrylic acid) at 32 cm^{-1} resolution.

Table 5 Spectral correlation for polymers

Polymer	Band location (nm)	Band assignment
Poly(propylene)	1192	Asymmetric methyl (C−H) stretch – second overtone
	1220	Asymmetric methylene (C−H) stretch – second overtone
	1394	Methyl and methylene (C−H) combination
	1700	Asymmetric methyl (C−H) stretch – first overtone
	1726	Asymmetric methylene (C−H) stretch – first overtone
	1820	Symmetric methyl (C−H) stretch – first overtone
	2323	C−H bend – second overtone
	2383	C−H stretch and C−C stretch combination
	2454	C−H combination band
Poly(propylene) + poly(ethylene) copolymer	1728	Asymmetric C−H stretch – first overtone
	2314	C−H bend – second overtone
	2356	C−H stretch and C−C stretch combination
Poly(styrene)	1142	Aromatic (C−H) stretch – second overtone
	1684	Aromatic (C−H) stretch – first overtone
Starch	1456	O−H stretch – first overtone
	1700	Asymmetric methyl (C−H) stretch – first overtone
	1739	Asymmetric methylene (C−H) stretch – first overtone
	1927	O−H stretch + HOH deformation + O−H bend combination
	2105	O−H stretch + HOH deformation combination
	2291	C−H stretch + CH_2 deformation combination
	2488	C−H stretch + C−C stretch + C−O−C stretch combination
Poly(acrylic acid)	1185	Asymmetric methylene (C−H) stretch – second overtone
	1428	O−H stretch – first overtone
	1695	Asymmetric methyl (C−H) stretch – first overtone
	1735	Asymmetric methylene (C−H) stretch – first overtone
	1924	O−H stretch – first overtone + C=O stretch – second overtone + O−H stretch/HOH deformation combination + O−H bend second overtone
	2149	C−H stretch/C=O stretch combination + symmetric C−H deformation
	2286	C−H stretch + CH_2 deformation + C−H bend – second overtone
	2489	C−H stretch + C−C stretch + C−O−C stretch combination

phenolics, natural products (polysaccharides), chloroprenes, polyamides, acrylonitriles, polystyrenes, polyfluorocarbons, and polyacrylates. Kradjel and McDermott[9] describe analytical parameters for measurements in olefin production, PVC, PVA, polyvinyl acetate, polystyrene, acrylic and methacrylic polymers, polyamides, cellulose esters, starch, alcohols, polyurethanes, polyesters, epoxy resins, phenolics, and melamine formaldehydes.

6 RUBBERS

6.1 Types of Rubber

Rubber or elastomer is a natural or synthetic material having the ability to undergo deformation under the influence of an applied force and regain its original shape once the applied force is removed. Natural rubbers are polymeric materials originally obtained from the rubber tree (*Hevea brasiliensis*). The sap of this tree is a latex material which is processed by coagulation and drying the sap, and then vulcanizing and adding filler compounds. The basic natural rubber compound is isoprene which contains the repeating unit of $-CH_2C(CH_3):CHCH_2-$. Synthetic rubbers exist such as nitriles, butadienes, neoprene, butyl rubbers, polysulfide rubbers, PVC, and silicone rubbers whose structures and NIR spectra are shown.

Structures **(15)** (*trans* form) and **(16)** (*cis* form) show typical repeating units found in natural rubber (isoprene). Structure **(17)** shows a repeating unit of nitrile rubber.

Structure **(18)** is a monomer unit of *cis* or *trans* polybutadiene rubber and structure **(19)** is a repeating unit of neoprene rubber. Structure **(20)** is a monomer unit of butyl rubber and **(21)** is a repeating unit of synthetic natural rubber (*z*-polyisoprene). Structure **(22)** shows the general formula for polysulfide rubbers where R represents −C−C, −C−C−O−C−C, or −C−C−O−C−O−C−C− groups. Structure **(23)** is a building block of PVC. The methyl group in silicone rubbers (polysiloxane) **(24)** is often substituted with a variety of alkyl groups. Structure **(25)** shows SB (styrene–butadiene) rubber. Urethane rubbers **(26)** are depicted where R and R′ are −C− and −C−C−.

Figures 12–14 shows the NIR spectra of styrene, ethylene butylene, styrene copolymer, styrene–isoprene–styrene and silicone DMS, respectively.

6.2 Spectral Correlation Chart for Rubbers

Typical rubber spectra exhibit similar NIR bands and positions which are represented by the chart in Table 6.

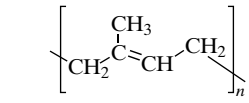

(15)

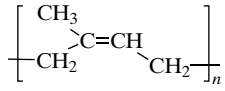

(16)

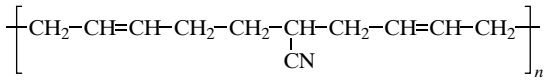

(17)

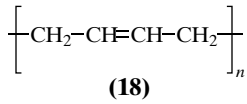

(18)

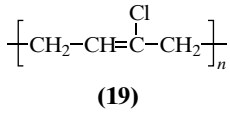

(19)

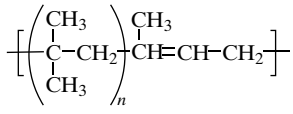

(20)

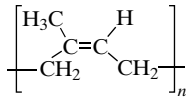

(21)

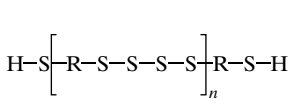

(22)

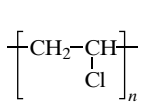

(23)

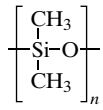

(24)

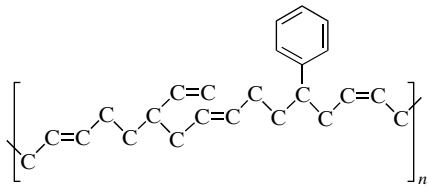

(25)

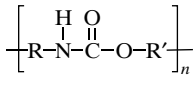

(26)

6.3 Qualitative Analysis of Rubbers

Rubber analysis can be completed using NIR spectra for verification of rubber quality and type. This analytical method utilizes spectral searching methods as described in section 7 with comparisons against known quality reference standards. Batch mixing conditions and reaction kinetics are also measured using NIR. The description of qualitative techniques found in section 5.3 describes some of the general methods used for rubber (or elastomer) measurements. Final product quality and evaluation is becoming a more common practice for qualifying polymer and elastomer production quality.

The traditional use of molecular spectroscopy for qualitative analysis of elastomeric materials involved the use of IR microspectroscopy. This technique was

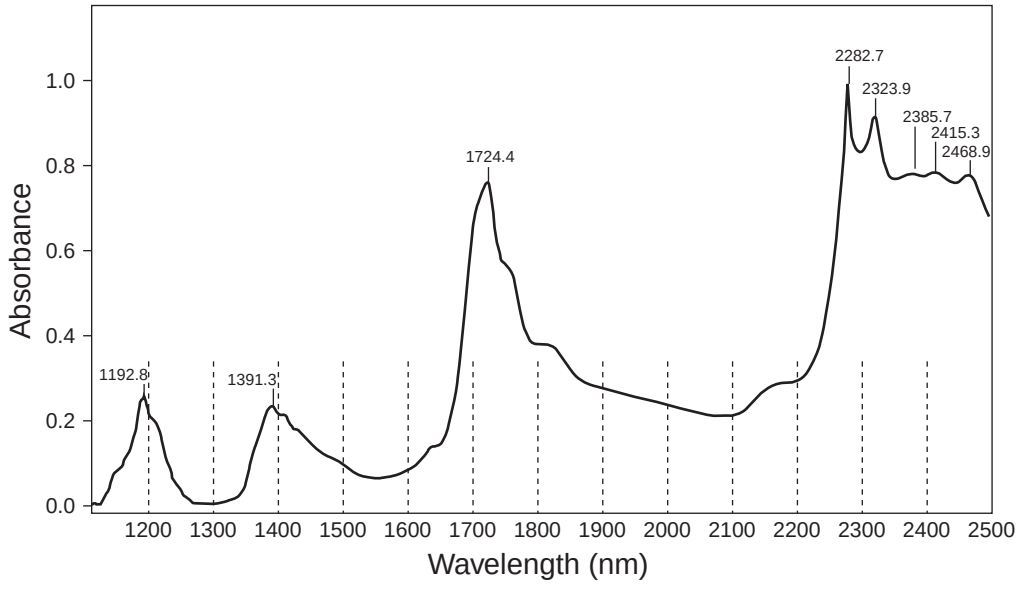

Figure 12 NIR spectrum of styrene, ethylene butylene, styrene copolymer at 32 cm^{-1} resolution.

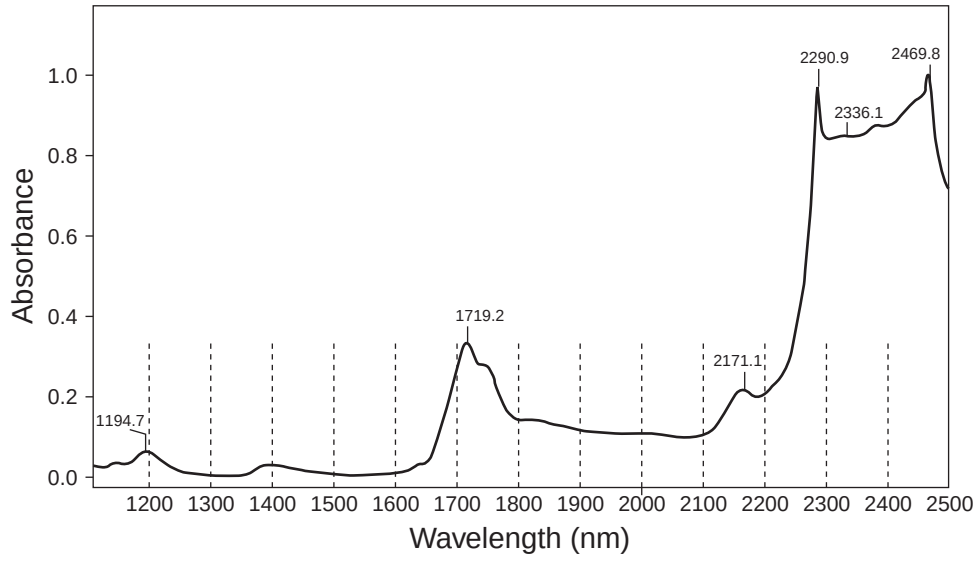

Figure 13 NIR spectrum of styrene–isoprene–styrene at 32 cm^{-1} resolution.

used for analysis of minute amounts of material for forensic or synthetic organic chemistry. This technique is still used for cases involving a small quantity of analyte. Samples as small as 5–20 μm can be measured using microspectroscopic techniques. The use of NIR for assignment of functional groups present in an unknown compound is generally not performed. For identification of the presence of various functional groups, FTIR and Raman spectroscopy are generally employed due to the greater sensitivity and resolution associated with these techniques (see section 8 for further details). NIR can be used for bulk qualitative analysis when sample identification or qualification are necessary.

6.4 Quantitative Analysis of Rubbers

As is described in detail in section 4 of this article, functional groups containing C–H, N–H, and O–H within elastomer compounds are most easily characterized using NIR spectroscopy. The stretching vibrations of these molecules can generally be quantified using multivariate calibration and NIR spectra as described in section 7. The unique attributes of an NIR spectrum

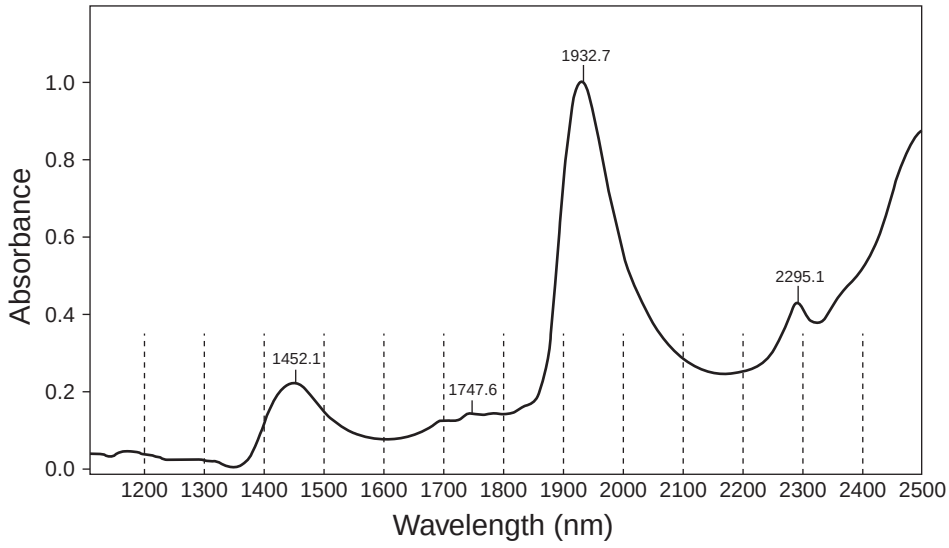

Figure 14 NIR spectrum of silicone (DMS, dimethyl siloxane) at 32 cm^{-1} resolution.

Table 6 Spectral correlation for rubbers

Rubber	Band location (nm)	Description
Styrene, ethylene butylene, styrene copolymer	1193	Methyl (C−H) stretch – second overtone
	1391	C−H stretch mode combination
	1724	Asymmetric methylene (C−H) stretch – first overtone + aromatic C−H stretch – first overtone
	2180	C−H bend – second overtone
	2283	C−H stretch + CH_2 deformation combination
	2324	C−H stretch + CH_2 deformation combination
	2386	C−H stretch + C−C stretch + C−O−C stretch combination
Styrene–isoprene–styrene	1195	Methyl (C−H) stretch – second overtone
	1400	C−H stretch mode combination
	1719	Asymmetric methylene (C−H) stretch – first overtone + aromatic C−H stretch – first overtone
	2171	C−H bend – second overtone
	2291	C−H stretch + CH_2 deformation combination
	2336	C−H stretch + CH_2 deformation combination
	2470	C−H stretch + C−C stretch + C−O−C stretch combination
Silicone (DMS)	1452	Si−O stretch – first overtone
	1748	Methyl (C−H) stretch – first overtone
	1933	Si−O−H stretch + Si−O−Si deformation combination
	2295	C−H bend – second overtone

allow quantification of carbon–carbon double bonds, methyl, methylene, and methine (in some cases) carbon–hydrogen stretching. Bands due to the N−H stretch of amines and amine salts are also measurable. Carbonyl C=O stretch and associated C−H stretch are also quantifiable, as is information from the C−O−C and C−C stretching vibrations. A large signal is obtained when measuring O−H, H−O−H, O−Si and Si−O−Si stretching vibrations. With this basic knowledge the analyst can determine the appropriate spectral regions for measuring reactants and reaction products before, during, and after the elastomeric forming reaction(s). The tables and reference spectra demonstrated in this article are useful for identifying the appropriate spectral regions.

7 METHOD DEVELOPMENT

Method development for NIR spectroscopy in the analysis of polymers and rubbers often involves the use

of chemometrics. Chemometrics as it pertains to NIR spectroscopy involves the preprocessing of spectroscopic measurement data to maximize the signal-to-noise characteristics and the use of multivariate calibration algorithms to develop predictive or comparative models for analysis. Method development for polymer and rubber analysis involves the generalized steps described in the following text for both qualitative and quantitative analysis.

7.1 Qualitative Analysis

The ''alikeness'' of one test polymer spectrum (or series of spectra) to a reference set of polymer spectra can be determined by calculating a point-by-point correlation between absorbance data for each test and reference spectrum. The correlation matching can be accomplished for all data points available or for a preselected set only. The more alike the test and reference spectra are, the higher (closer to 1.00) are the r (correlation coefficient) and R^2 (coefficient of determination) values. A perfect match of the two spectra would produce r or R^2 values of 1.00000. The sensitivity of the technique can be increased by pretreating the spectra first to higher order derivatives and then calculating the correlation between the test and reference spectra. Full spectral data can also be truncated (or reduced) to include only spectral regions of particular functional group interest, a practice which will further improve matching sensitivity for a particular spectral feature of interest.

7.2 General Qualitative Practices

The ASTM (American Society for Testing and Materials) has published a *Standard Practice for General Techniques for Qualitative Analysis* (Method E 1252-88). The method[34] describes techniques useful for qualitative evaluation of liquids, solids, and gases using the spectral measurement region of $4000–50\,cm^{-1}$ (above 2500 nm). A revised method is being drafted to encompass the newer chemometrics techniques in current research practice.[35] Several primary references describe the use of this mathematical approach.[36–39]

7.3 Spectral Searching

Techniques for searching and matching test spectra include the use of Mahalanobis distance[36–39] and cross-correlation techniques (correlation matching) described above. The general method for comparing two spectra (test versus reference), where the reference is a known compound, or the mean spectrum of a set of known spectra, is given as the MI. The MI is computed by comparing the vector dot products between the test and reference spectra. The theoretical values for these dot products range from -1.0 to $+1.0$, where -1.0 is a perfect negative (inverse) correlation, and $+1.0$ is a perfect match. Since for NIR spectroscopy only positive absorbance values are used to compute the dot products, the values for the MI must fall within the 0.0 to $+1.0$ range. The method, or variations thereof, is described in the literature.[40–44] The mathematics are straightforward and are given below.

The MI is equal to the cosine of the angle (designed as α) between two row vectors (the test and reference spectra) projected on to a two-dimensional plane, and is equivalent to the correlation (r) between the two spectra (row vectors), Equation (15)

$$\mathrm{MI} = \cos\alpha = \frac{T \bullet R}{|T||R|} \tag{15}$$

where T is the test spectrum row matrix, and R is the reference spectrum row matrix.

Note the following, Equation (16)

$$T \bullet R = \sum_{i=1}^{n} T_i R_i \tag{16}$$

where T_i represents the individual data points for the test spectrum (designated as the absorbance values of spectrum T from wavelengths $i = 1$ through n), and R_i represents the individual data points for the reference spectrum (designated as the absorbance values of spectrum R from wavelengths $i = 1$ through n) and where, Equation (17)

$$|T||R| = \left(\sum_{i=1}^{n} T_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} R_i^2\right)^{\frac{1}{2}} \tag{17}$$

Note the angle (α), in degrees, between two vectors can be determined from the MI using, Equation (18)

$$\alpha = \cos^{-1}(\mathrm{MI}) \tag{18}$$

7.4 Near-infrared Quantitative Calibration

When NIR spectroscopy is applied as a quantitative analysis technique a computer is used to collect the measurement data, to calculate multivariate calibration models, and to apply those models to the prediction of the chemical composition of unknown samples. Table 7 describes the general quantitative calibration procedures. Sample sets are selected for teaching, model validation, and analyzer monitoring control. The purpose of the calibration model is to relate the concentration of some analyte found in a sample to the spectral data collected from that sample. The teaching sets must be comprised of a wide survey of sample chemistries as well as a variety of instrument response types. The instrumental artifacts are convoluted into the NIR spectra, and thus unless

Table 7 Critical procedures in calibrating (and maintaining) an NIR spectrophotometer

1. Selecting the calibration set using experimental design criteria.
2. Determination and verification of standard concentrations using verifiable reference methods.
3. Collection and verification of optical response data.
4. Final validation for the integrity of X-block (instrument) and Y-block (reference) data.[47–49]
5. Optimization for data preprocessing method.
6. Calibration modeling.
7. Statistical validation of the calibration model.
8. Routine analysis and calibration monitoring (using outlier statistics and monitoring samples).
9. Calibration model updating for sample or instrument changes.
10. Calibration model transfer optimization.

instruments have identical response across systems (or following major repair) the instrument function must be considered when developing a calibration model.

The complexity of calibration is dependent upon sample and instrument complexity, as well as the desired model performance. There is generally a trade between calibration performance and calibration robustness. Robustness here refers to the characteristic of a calibration where it can be transferred to other instruments (or to the same instrument after repair) without loss of performance. The use of multivariate quantitative analysis to describe the relationships between the concentration of an analyte and its NIR instrument response (viz. absorbance, T, transflectance, or reflectance) is common practice in analytical chemistry.[45] The practice of relating spectrophotometric response to concentration has been accomplished for a number of applications using Beer's law combined with C-matrix, K-matrix, MLR (multiple linear regression), PCR (principal components regression), PLSR (partial least squares regression), LWR (locally weighted regression), and NN (neural networks). Each application, and each set of experimental conditions, will produce slight differences in the selection of calibration channels (due to instrument response) as well as the weighting vectors (or coefficients) used for the final calibration model.

The use of NIR and chemometrics for *official* analytical methods is still in its early stages. The ASTM has adopted *Standard Practices for IR, Multivariate, Quantitative Analyses (Practice E1655-98)*,[34] which describes the peer-reviewed use of chemometrics for calibrating IR and NIR instruments. The perceived "barriers" to a more rapid deployment of chemometrics-based spectroscopy for chemical analysis include (1) a broad-based natural resistance to newer technologies, (2) the technical complexity and non-intuitive nature of chemometrics, and (3) difficulty in implementation (and maintenance) of chemometric methods. Since many of the calibration techniques are nonautomated, they require continuous expert manual labor for maintenance; this is accomplished through highly skilled chemometricians and application scientists. Table 7 shows the steps required to establish and maintain NIR quantitative methods.

7.5 Laboratory Methods

Typical laboratory methods for polymer and rubber analysis include chromatographic techniques, thermal analysis, microscopy, elemental, structural, rheometry, and various chemical identification tests. Many of these methods which relate directly to the presence of molecular functional groups can be measured using NIR spectroscopy. It is important to remember in laboratory analysis of polymer components that NIR is a correlation technique which can relate chemical or physical property information to NIR spectra. The capability of NIR to predict polymer components accurately depends on the true correlation between the measured physical or chemical property within a sample and that sample's NIR spectrum.

Most NIR techniques can be applied either at-line (near the actual manufacturing point), on-line (automated and at the manufacturing site), or in situ (measured in site or where a reaction or chemical/physical phenomenon is occurring). Moving a NIR measurement technique from the laboratory to the production line can be accomplished using remote measurement optics, such as specially configured optical collection heads or fiber optic cables or bundles. These will be discussed briefly in the next two sections.

7.6 At-line Measurements

At-line measurements involve taking the analytical measurement(s) using a dedicated and ruggedized spectrophotometer very near the production line, reaction vessel, or batch mixer. The advantage of having the measurement at the line is that rapid information used for making manufacturing decisions becomes readily available in time to make process adjustments related to quality and production efficiency. Slower laboratory measurements are available only after the fact and do not enable real-time adjustments during the process.

7.7 Process and On-line Measurements

Process analyzers are associated with such terms as airborne, at-line, automatic/automated, fieldable, hyperspectral, imaging/image analysis, in situ, in-line, near-line, noninvasive/nondestructive, on-line, open path, portable/hand-held, process monitoring/production control, quality control/quality assurance, quality monitoring,

rapid, real-time, and remote. In this discussion the term on-line is referred to as a NIR measurement made on the manufacturing process during production. The on-line measurement provides real-time or near real-time information for control of manufacturing parameters during production. On-line measurements are often made for batch reactions of most polymer products including glycols, polymers, polymer oxidation reactions, polymer mixing, polyolefin production, polyurethane foam manufacturing, polyether urethanes, polyethylene terephthalate, and biotechnology-related polymers.

8 COMPARISON OF NEAR-INFRARED WITH INFRARED AND RAMAN SPECTROSCOPY

NIR spectroscopy is used where multicomponent molecular vibrational analysis is required in the presence of interfering substances. The NIR spectra consist of overtones and combination bands of the fundamental molecular absorptions found in the MIR region. A detailed description of NIR spectral bands is given by Workman.[46] NIR spectra consist of generally overlapping vibrational bands that are nonspecific and poorly resolved. Chemometric mathematical data processing can be used to calibrate in quantitative analysis despite these apparent spectroscopic limitations. Traditional NIR spectroscopy was used in agricultural product analysis for lignin polymers (2270 nm), paraffins and long-alkane chain polymers (2310 nm), glucose-based polymers such as cellulose (2336 nm), amino acid polymers as proteins (2180 nm), carbohydrates (2100 nm), and moisture (1440 and 1940 nm).

Dominant NIR spectral features include the methyl C−H stretching vibrations, methylene C−H stretching vibrations, aromatic C−H stretching vibrations, O−H stretching vibrations, methoxy C−H stretching, and carbonyl associated C−H stretching. In addition, N−H from primary amides, secondary amides (both alkyl and aryl group associations), N−H from primary, secondary, and tertiary amines, and N−H from amine salts predominate in NIR spectral features for polymers and organic compounds.

MIR spectroscopy provides a measurement technique for intense, isolated and reliable absorption bands of fundamental molecular vibrations from polymers and other organic compounds. The spectrometric methodology allows for univariate calibration with higher signal strength (absorptivities) required for solid, liquids or gas phase measurements. Relatively small pathlengths of 0.1–1.0 mm are required for hydrocarbon liquids and solids. The technique is generally incompatible with the use of fiber optics, but specialized fiber materials exist. Instrumentation costs more than NIR spectrophotometers for the most part.

Dominant MIR spectral features include the C−H (methyl, methylene, aromatic, methoxy, and carbonyl) fundamental stretching and bending molecular vibrations, O−H (hydroxyl) stretch fundamental vibrations, N−H (amine) stretching, C−F (fluorocarbon) stretching, −C=N (nitrile) stretching, −C=O (carbonyl) stretch from esters, acetates, and amides, C−Cl stretch from chlorinated hydrocarbons, and $-NO_2$ from nitro-containing compounds.

Raman spectroscopy can be used for a variety of measurements on samples that are aqueous in nature or where glass sample holders are present. Carbon dioxide, water, and glass (silica) are weak scatterers and thus there is generally no problem in analyzing samples that have these properties. There is typically no sample preparation involved in samples measured using Raman spectra. Raman spectroscopy is complimentary to MIR spectroscopy in the measurement of fundamental molecular vibrations. Raman measurements are compatible with fiber optics. Raman measurements exhibit high signal-to-noise and represent a reasonable cost for instrumentation. The dominant Raman spectral features are acetylenic C=C stretching, olefinic C=C stretch at 1680–1630 cm^{-1}, N=N (azo-) stretching, S−H (thio-) stretching, C=S stretching, C−S stretching, and S−S stretching bands. Raman spectra also contain such molecular vibrational information as CH_2 twist and wagging, carbonyl C=O stretch associated with esters, acetates, and amides, C−Cl (halogenated hydrocarbons) stretching, and $-NO_2$ (nitro-/nitrite) stretching. In addition, Raman yields information about phenyl-containing compounds at 1000 cm^{-1}.

ABBREVIATIONS AND ACRONYMS

ART-2a	Adaptive Resonance Theory-based Artificial Neural Network
ASTM	American Society for Testing and Materials
ATR	Attenuated Total Reflectance
DMS	Dimethyl Siloxane
DR	Diffuse Reflectance
DRIFTS	Diffuse Reflectance Fourier Transform Spectroscopy
DSC	Differential Scanning Calorimetry
DT	Diffuse Transmittance
EMR	Electromagnetic Radiation
FID	Flame Ionization Detection
FIR	Far-infrared

FTIR	Fourier Transform Infrared Spectroscopy
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HO	Harmonic Oscillator
IR	Infrared
LC/MS	Liquid Chromatography/Mass Spectrometry
LW	Longwave
LWR	Locally Weighted Regression
MI	Match Index
MIR	Mid-infrared
MLFBP	Multilayer Feedforward Backpropagation of Error Neural Networks
MLR	Multiple Linear Regression
MS	Mass Spectrometry
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
NN	Neural Networks
PCR	Principal Components Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
PVA	Polyvinyl Alcohol
PVC	Polyvinyl Chloride
SB	Styrene–Butadiene
SIMCA	Soft Independent Modeling of Class Analogies
SW	Short Wave
T	Transmittance
TGA	Thermal Gravimetric Analysis
UV	Ultraviolet

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared • Spectral Data, Modern Classification Methods for • Theory of Infrared Spectroscopy

REFERENCES

1. H. Staudinger, *Ber. Bunsenges. Phys. Chem.*, **53**, 1073 (1920).
2. H. Mark, R. Raff, *Z. Phys. Chem. B*, **31**, 275 (1936).
3. H. Mark, G. Saito, *Mh. Chem.*, **68**, 237 (1936).
4. J.W. Ellis, 'Molecular Absorption Spectra of Liquids Below 3 Microns', *Trans. Faraday Soc.*, **25**, 888–898 (1928).
5. W. Kaye, 'Near-infrared Spectroscopy; A review. I. Spectral Identification and Analytical Applications', *Spectrochim. Acta*, **6**, 257–287 (1954).
6. R.F. Goddu, 'Near-infrared Spectrophotometry', *Adv. Anal. Chem. Instrum.*, **1**, 347–424 (1960).
7. K.B. Whetsel, 'Near-infrared Spectrophotometry', *Appl. Spectrosc. Rev.*, **2**(1), 1–67 (1968).
8. E. Stark, K. Luchter, M. Margoshes, 'Near-infrared Analysis (NIRA): A Technology for Quantitative and Qualitative Analysis', *Appl. Spectrosc. Rev.*, **22**(4), 335–399 (1968).
9. C. Kradjel, L. McDermott, 'NIR Analysis of Polymers', in *Handbook of Near-infrared Analysis*, eds. D. Burns, E. Ciurczak, Marcel Dekker, New York, 1992.
10. D.O. Hummel, *Infrared Spectroscopy of Polymers*, Wiley-Interscience, New York, 1966.
11. R. Zbinden, *Infrared Spectroscopy of High Polymers*, Academic Press, New York, 1964.
12. J.C. Henniker, *Infrared Spectroscopy of Industrial Polymers*, Academic Press, New York, 1967.
13. J. Haslam, H.A. Willis, *Identification and Analysis of Plastics*, Van Nostrand, Princeton, New Jersey, 1965.
14. J.L. Koenig, *Spectroscopy of Polymers*, American Chemical Society, Washington, D.C., 1992.
15. W. van den Broek, E.P.P.A. Derks, E.W. van de Ven, D. Wienke, P. Geladi, L.M.C. Buydens, 'Plastic Identification by Remote Sensing Spectroscopic NIR Imaging Using Kernel Partial Least Squares (KPLS)', *Chemom. Intell. Lab. Syst.*, **35**(2), 187–197 (1996).
16. D. Wienke, W. van den Broek, W. Melssen, L. Buydens, R. Feldhoff, T. Huth-Fehre, T. Kantimm, F. Winter, K. Cammann, 'Near-infrared Imaging Spectroscopy (NIRIS) and Image Rank Analysis for Remote Identification of Plastics in Mixed Waste', presented at ANAKON '95 held in Schliersee, Germany, 24–26 Apr. 1995; *Fresenius' J. Anal. Chem.*, **354**(7–8), 823–828 (1996).
17. D. Wienke, W.H.A.M. Van den Broek, L. Buydens, 'Identification of Plastics Among Non-plastics in Mixed Waste by Remote Sensing Near-infrared Imaging Spectroscopy. 2. Multivariate Image Rank Analysis for Rapid Classification', *Anal. Chem.*, **67**(20), 3760–3766 (1995).
18. W. van den Broek, D. Wienke, W.J. Melssen, C.W.A. de From, L. Buydens, 'Identification of Plastics among Non-plastics in Mixed Waste by Remote Sensing Near-infrared Imaging Spectroscopy. 1. Image Improvement and Analysis by Singular Value Decomposition', *Anal. Chem.*, **67**(20), 3753–3759 (1995).

19. R. Feldhoff, D. Wienke, K. Cammann, H. Fuchs, 'A Device for On-line Post-consumer Package Identification Using NIR Spectroscopy Combined with a Fuzzy-ARTMAP Classifier has been Demonstrated in an Industrial Environment', *Appl. Spectrosc.*, **51**, 362–368 (1997).
20. G. Lachenal, I. Stevenson, 'Interaction and Hydrogen-bonding Studied by Near-FTIR Spectroscopy', *NIR News*, **7**(6), 10–12 (1996).
21. M. Nielsen, P. Jurasek, J. Hayashi, E. Furimsky, 'Formation of Toxic Gases During Pyrolysis of Polyacrylonitrile and Nylons', *J. Anal. Appl. Pyrolysis*, **35**(1), 43–51 (1995).
22. D. Wienke, W. van den Broek, W. Melssen, L. Buydens, R. Feldhoff, T. Kantimm, T. Huth-Fehre, L. Quick, F. Winter, K. Cammann, 'Comparison of an Adaptive Resonance Theory Based Neural Network (ART-2a) against other Classifiers for Rapid Sorting of Post Consumer Plastics by Remote Near-infrared Spectroscopic Sensing Using an InGaAs Diode Array', *Anal. Chim. Acta*, **317**(1–3), 1–16 (1995).
23. H.W. Siesler, 'Near-infrared Spectroscopy for Polymer Reaction and Process Control: Synthesis, Processing and Recycling', *NIR News*, **6**(1), 3–6 (1995).
24. P. Brush, 'The Near-infrared Analysis of Polyols: Process Monitoring in a Hostile Environment', *NIR News*, **5**(5), 14–15 (1994).
25. L. Xu, Z. Shen, J.R. Schlup, 'Online Remote Monitoring of Thermoset Resin Cure Using Near-infrared External Reflection Spectroscopy', *J. Adv. Mater.*, **28**(4), 47–54 (1997).
26. P. Dallin, 'NIR Analysis in Polymer Reactions', *Process Control Qual.*, **9**(4), 167–172 (1997).
27. K. Stromberg, 'Challenge of Introducing NIR Online in Polyol Production', *Process Control Qual.*, **9**(4), 179–183 (1997).
28. D. Fischer, T. Bayer, K.J. Eichhorn, M. Otto, 'Inline Process Monitoring on Polymer Melts by NIR Spectroscopy', *Fresenius' J. Anal. Chem.*, **359**(1), 74–77 (1997).
29. M.G. Hansen, S. Vedula, 'In-line Measurement of Copolymer Composition and Melt Index', *Polym. Process Eng.*, **97**, (Pap. Int. Conf.), 89–102 (1997).
30. J. Mijovic, S. Andjelic, J.M. Kenny, 'In Situ Real-time Monitoring of Epoxy/Amine Kinetics by Remote Near-infrared Spectroscopy', *Polym. Adv. Technol.*, **7**(1), 1–16 (1996).
31. A.L. Moe, J.D. Konicek, US Patent No. 5 457 319 A, Oct. 10, 7, 1995.
32. J. Workman, A. Springsteen (eds.), 'Applied Spectroscopy: A Compact Reference for Practitioners', Academic Press, Boston, 46–47, 1998.
33. J. Grasselli, R. Hannah, D.W. Mayo, F.A. Miller, 'Infrared Spectroscopy, Interpretation of Spectra', Bowdoin College, Brunswick, Maine, 1994.
34. ASTM (American Society for Testing and Materials) Method E 1252, 'Standard Practice for General Techniques for Qualitative Infrared Analysis', ASTM Committee E 13, 1916 Race Street, Philadelphia, PA 19103, 1988. The current address for ASTM is 100 Barr Harbor Drive, West Conshohocken, PA 19428–2959, USA.
35. ASTM Committee E13.11, 'Practice for Near-infrared Qualitative Analysis', ASTM Committee E 13, 1916 Race Street, Philadelphia, PA 19103, 1998.
36. H.L. Mark, D. Tunnell, 'Qualitative Near-infrared Reflectance Analysis Using Mahalanobis Distances', *Anal. Chem.*, **57**, 1449–1456 (1985).
37. H. Mark, 'Normalized Distances for Qualitative Near-infrared Reflectance Analysis', *Anal. Chem.*, **58**, 379–384 (1986).
38. R.G. Whitfield, M.E. Gerber, R.L. Sharp, *Appl. Spectrosc.*, **41**, 1204–1213 (1987).
39. P.C. Mahalanobis, *Natl. Inst. Sci. Proc.*, **2**, 49 (1936).
40. J.C. Reid, E.C. Wong, 'Data-reduction and -Search System for Digital Absorbance Spectra', *Appl. Spectrosc.*, **20**, 320–325 (1966).
41. P.M. Owens, T.L Isenhour, 'Infrared Spectral Compression Procedure for Resolution Independent Search Systems', *Anal. Chem.*, **55**, 1548–1553 (1983).
42. K. Tanabe, S. Saeki, 'Computer Retrieval of Infrared Spectra by a Correlation Coefficient Method', *Anal. Chem.*, **47**, 118–122 (1975).
43. L.V. Azarraga, R.R. Williams, J.A. de Haseth, 'Fourier Encoded Data Searching of Infrared Spectra (FEDS/IRS)', *Appl. Spectrosc.*, **35**, 466–469 (1981).
44. J.A. de Haseth, L.V. Azarraga, 'Interferogram-based Infrared Search System', *Anal. Chem.*, **53**, 2292–2296 (1981).
45. H. Mark, J. Workman, *Statistics in Spectroscopy*, Academic Press, Boston, 1991.
46. J. Workman, 'Interpretive Spectroscopy for Near-infrared', *Appl. Spectrosc. Rev.*, **31**(3), 251–320 (1996).
47. D.C. Hoaglin, R.E. Welsch, 'The Hat Matrix in Regression and ANOVA', *Am. Statistician*, **32**(1), 17–22 (1978).
48. N.R. Draper, J.A. John, 'Influential Observations and Outliers in Regression', *Technometrics*, **23**(1), 21–26 (1981).
49. R.D. Cook, S. Weisberg, 'Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression', *Technometrics*, **22**(4), 495–508 (1980).

Neutron Scattering in Analysis of Polymers and Rubbers

D.G. Bucknall
University of Oxford, Oxford OX1 3PH, UK

1 Introduction 1
2 Some Basic Concepts of Neutron Scattering 2
2.1 Scattering and Structure 7
2.2 Molecular Motion and Dynamics 10
3 Neutron Production and Measurement 10
4 Applications of Neutrons to Polymeric Materials 11
4.1 Effects of Deuteration 11
4.2 Determination of Molecular Parameters 11
4.3 Phase-separated Systems 12
4.4 Shear and Deformation 13
4.5 Homopolymer Interfaces 14
4.6 Polymer Interdiffusion 15
4.7 Copolymers at Interfaces 16
4.8 Polymer Dynamics 16
5 Conclusions 18
Abbreviations and Acronyms 18
Related Articles 19
Further Reading 19
References 20

The neutron is in many ways an ideal probe for the investigation of the structure and dynamics of materials. Elastically scattered neutrons can be used to determine the structure of the sample from the atomic to macromolecular length scales, whilst dynamic information describing the vibration, rotation and translation of these molecules is obtained from inelastically scattered neutrons. The use of isotopic substitution provides an easy route to highlighting whole or parts of a molecule in the neutron scattering experiment. Isotopic labeling has been extensively used in polymer studies owing to the large difference between hydrogen and deuterium. The combined use of neutrons often with isotopic labeling schemes has provided a wealth of detailed information about polymer materials, which could not be derived so easily by other methods.

1 INTRODUCTION

Neutrons have been used widely in scattering experiments to elucidate not only the structure but also the dynamics of materials. The neutron scattering techniques that have been generally applied to study polymeric material include small-angle neutron scattering (SANS) and wide-angle neutron scattering (WANS), neutron reflection (NR), quasielastic neutron scattering (QENS) and inelastic neutron scattering (INS) and neutron spin-echo (NSE). These techniques can each provide different information about the material. Elastic scattering (WANS, SANS and NR) provide structural and compositional information, whereas inelastic scattering techniques (QENS, INS and NSE) allow determination of molecular motion. In many ways therefore the neutron is an ideal probe for materials studies. The neutrons used for these studies have wavelengths given by de Broglie's relationship, $\lambda = h/mv$, where $m = 1.6 \times 10^{-27}$ kg, and v is the neutron velocity. With v of the order $10^3\,\mathrm{m\,s^{-1}}$ this gives neutron wavelengths of order 10^{-10} m (i.e. 1 Å), which is comparable to atomic dimensions allowing the possibility of structural determination on molecular length scales. The energies of the neutron equal to $\frac{1}{2}mv^2$ are of the order $k_\mathrm{B}T$, i.e. in the meV (10^7–10^{15} Hz) regime. This dynamic range allows the study of molecular rotation, vibration and translation and is comparable to the methods of infrared (IR), Raman or nuclear magnetic resonance (NMR) spectroscopy.

Although structural studies using neutrons covers the same spatial range provided by X-rays, the major difference in the techniques is in the interaction with the probing radiation and the sample. Unlike X-rays, which are scattered by the electron cloud surrounding the atom, neutrons are scattered either by the nucleus (nuclear scattering) or for the case of polarized neutrons also by the magnetic moment of unpaired electrons (magnetic scattering). The rest of this article will confine the discussion to nuclear scattering since magnetic scattering has not been widely exploited.[1–6] The strength of the neutron–nucleus interaction is characterized by the scattering length, b, which is not a function of atomic number but varies seemingly randomly across the periodic table. Of most interest to the polymer scientist is the difference between hydrogen and its isotope deuterium, which have large but different values of b ($b_\mathrm{H} = -3.739 \times 10^{-15}$ m and $b_\mathrm{D} = +6.671 \times 10^{-15}$ m). By deuterating parts or whole molecules, a huge contrast is produced between the labeled and unlabeled species. It is perhaps this property of the neutron above all others that has made neutron scattering such a powerful technique. The neutron is also highly penetrating and this has great advantages for the types of sample environment that can

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

be used. A number of useful engineering materials are almost transparent to the neutron, allowing the use of cryostats, pressure vessels, shear cells and simpler sample confinement systems to be used with little problems of beam loss or parasitic scattering. Finally, it should also be noted that unlike IR and Raman spectroscopy, owing to the neutron–nucleus interaction there are no selection rules which apply to INS.

Although there are only a limited number of facilities worldwide which can provide neutrons for research purposes, it may seem surprising to discover the vast wealth of literature describing results gained from neutron studies. However, neutron scattering has and continues to be arguably the only way to answer many fundamental and applied questions, in which case the amount of literature is not disproportionate with the number of instruments that allow such studies. In trying to summarize the huge amount of available literature on the use of neutrons in polymer studies, this article is necessarily limited to providing an overview, with a few examples by way of illustration. References to a large number of studies not mentioned directly in the text are given in a number of extended tables (see Tables 1–4). These are intended to provide a starting point for further literature searches because it has not been possible to provide a comprehensive list of all the available literature. Indeed, in the limited space available, it has also been necessary to limit the discussion and references mainly to polymers in the solid state. Further reading suggestions to general references and reviews have also been provided at the end of the article.

2 SOME BASIC CONCEPTS OF NEUTRON SCATTERING[7]

The features of a generalized neutron scattering experiment are shown in Figure 1.[8] The incident neutron beam is characterized in terms of a wavelength, λ_0, wavevector, κ_0, and energy, E_0. When passing through the sample, neutrons that are not absorbed are either transmitted without interaction or scattered via nuclei–neutron interactions. The scattered neutrons are analyzed in terms of their wavelength, λ, wavevector, κ, and energy, E. Both changes in direction of travel, $Q = \kappa - \kappa_0$, and energy, $\Delta E = E - E_0$, can occur. If the scattering process is elastic, then there is no energy change ($\Delta E = 0$) and only changes in direction of travel are observed. The momentum transfer, Q, is expressed by Equation (1):

$$Q = |Q| = \frac{4\pi}{\lambda} \sin \frac{\theta}{2} \tag{1}$$

where θ is the scattering angle. Therefore, elastic scattering only contains spatial information on the system.

If the scattering nuclei undergo motion with energies of the order of the neutron energies, then energy exchange between the neutron and nuclei may occur. In these circumstances the momentum transfer is now given by

Table 1 Structure and morphology[a]

Polymer	Description	Refs.
Homopolymers (including isotopic blends)		
BPA/PC	Short-range local order	155
PA	Ordering in LCPs	156–161
PAT	Morphology	162
h-PB + d-PB	Isotopic phase behavior	34, 163–168
	Chain orientation and conformation	169–173
	Self-diffusion	174
PBMA	Side-chain conformation in LCPs	175
h-PDMS + d-PDMS	Miscibility	35
	Gaussian chain statistics	176, 177
	Network morphology	178–180
h-PE + d-PE	Phase behavior	44, 62–64, 181–187
	Local structure	55, 188–191
	Effects of deformation	192–194
	Carbon black composites	195
h-PEE + d-PEE	Composition effects on χ	44
PEEK	Structural determination	196
PEs	Chain morphology	84, 197, 198
PEK	Structure and morphology	199
h-PET + d-PET	Scattering behavior	200–204
h-PIB + d-PIB	Phase behavior	205
h-PMMA + d-PMMA	Deviation from Gaussian statistics	206, 207
	Local chain conformation	208

Table 1 (*continued*)

Polymer	Description	Refs.
Polyolefins (general)	Morphology	62, 163, 173, 209
h-PP + d-PP	Phase behavior	210
	Chain conformation	211, 212
PPV	Morphology	213
PPy	Bulk conformation	214
h-PS + d-PS	Thermodynamics	44, 52, 185, 215–221
	Thin-film behavior	222, 223
	Local chain configuration	54, 224, 225
	Effects of cross-linking	226
	Orientation effects by extrusion	69, 75, 76, 227–230
PSO	Morphology in smectic phases	231
PTFE	Local chain morphology	12, 188, 232
h-PU + d-PU	Ionomer dimensions	233
PVA	Chain morphology	234, 235
PVC	Structure and morphology	236
h-PVE + d-PVE	Composition effects on χ	44
Na- SPS	Electrostatic persistence length	237
Blends containing PS		
PS + Pα-MS	Determination of χ	238
PS + PB	Isotopic phase behavior	173, 221, 239–241
PS + PB + PS-*co*-PB	Phase behavior	242
PS + PBA	Nodule coalescence	243
PS + PBMA	Miscibility	244–247
PS + PC	Spinodal decomposition	248
PS + PCHA-*stat*-PBMA	Spinodal decomposition	98, 249, 250
PS + PDMS	Interpenetrating network morphology	251
PS + PDPO	Phase behavior	252
PS + PMMA + PS-*co*-PMMA	Phase domain size effects	253
PS + PPO	Thermodynamics	254
PS + PI	Phase behavior	255–257
PS + P*p*MS	Determination of χ	40
PS + PPMS	Phase separation	219
	Pressure dependence	50, 51
PS + PVME	Phase behavior and isotopic effects	60, 61, 167, 221, 240, 258–260
	Determination of R_g and χ	39, 42, 258, 261–264
	Chain orientation	265–267
	Pressure dependence	49–51, 268
PS + SPS	Phase behavior	269
Blends containing PMMA		
PMMA + PC	Spinodal decomposition	270
PMMA + PECH	Thermodynamic behavior	271
PMMA + PEO	Composition and molecular weight effects of χ	45, 266
	Deformation behavior	272
PMMA + PS$_x$-*co*-PMMA$_{(1-x)}$	Phase behavior	273, 274
PMMA + SAN	Q dependence on χ	41
PMMA + SCPE	Phase behavior	275
	Spinodal decomposition	57–59
Miscellaneous blends		
EPDM + PSVP	Domain dimensions	276
$LiClO_4$ + PPO	Structural properties	277
PA + PVME	Miscibility and phase behavior	278
PAEK + PEI	Structural determination	279, 280
PAn + nylon 6	Morphology and fractal geometry	281, 282
PAS + epoxy	Phase diagram prediction	283
PB + PI	Phase behavior near T_g	167
PB + PS-*co*-PB	Phase behavior	284
PB + SBR	Phase behavior	285

(*continued overleaf*)

Table 1 (*continued*)

Polymer	Description	Refs.
PB + SBS	Blend morphology	286
PC + PCL	Critical phase behavior	287, 288
PC + PU + PVP	Microphase network structure	289
PE + PE-*co*-PEA	Blend morphology	290
PE + PEP	Phase behavior	71
PEB + PMB + PEB-*co*-PMB	Thermodynamic interactions	291–293
PEB + PMB	Pressure effects	294
PEG + PE-*co*-PVA	Thermodynamic behavior	295
PEO + PS-*co*-PMMA	Phase behavior	274
PI + PP	Thermodynamic interactions	296
PVME + PS-*co*-PMMA	Phase behavior	274

[a] Abbreviations (other than those already given in the text): BPA, bisphenol A; EPDM, ethylene–propylene–diene rubber; LCP, liquid crystal polymer; PA, polyacrylate; PAEK, poly(aryl ether ketone); PAn, polyaniline; PAS, poly(aryl sulfone); PAT, polyalkylthiophene; PB, 1,4-polybutadiene; PBA, poly(butyl acrylate); PBMA, poly(butyl methylacrylate); PC, polycarbonate; PCHA, poly(cyclohexyl acrylate); PCL, poly(ε-caprolactone); PDMS, polydimethylsiloxane; PDPO, poly(dimethylphenylene oxide); PE, polyethylene; PEA, poly(ethyl acrylate); PEB, polyethylbutene; PECH, polyepichlorohydrin; PEE, poly(ethylethylene); PEEK, poly(ethyl ether ketone); PEG, poly(ethylene glycol); PEI, poly(ether imide); PEK, poly(ethyl ketone); PEO, poly(ethylene oxide); PEP, polyethylpropylene; PEs, polyester; PET, poly(ethylene terephthalate); PI, polyisoprene; PIB, polyisobutene; PMB, polymethylbutene; PMMA, poly(methyl methacrylate); PP, polypropylene; P*p*MS, poly(*p*-methylstyrene); PPMS, polyphenylmethylsiloxane; PPO, poly(phenyl oxide); PPV, polyphenylenevinylene; PPy, polypyrrole; PS, polystyrene; PSO, polysiloxane; PSVP, poly(styrene–vinylpyridine); PTFE, polytetrafluoroethylene; PU, polyurethane; PVA, poly(vinyl acetate); PVC, poly(vinyl chloride); PVE, polyvinylethylene; PVME, poly(vinyl methyl ether); PVP, polyvinylpyridine; SAN, poly(styrene–acrylonitrile); SBR, styrene–butadiene rubber; SBS, styrene–butadiene–styrene copolymer; SCPE, solution-chlorinated polyethylene; SPS, sulfonated polystyrene.

Table 2 Interfaces and surfaces[a]

Polymer	Description	Refs.
Homopolymers		
EF/PDMS	Surface segregation of end groups	297
EF/PS	End group segregation	298, 299
EF/PS + PS	Surface segregation profile	110, 298, 300–302
Li SPS/*m*PA	Kinetics of interdiffusion	303
Pα-MS + PS	Surface segregation and wetting	304
PBrS/PS	Effect of bromination on width	96, 97
	Interdiffusion kinetics	305, 306
PB + PS	Segregation profiles	307, 308
PBTP	Thin-film composition profile	309
PC + PMMA	Phase separation in thin-films	310
	Interface structure	311
PE/PP	Interfacial width and χ determination	104, 312
PE/PS	Interfacial width and χ determination	105, 312
h-PEP + d-PEP	Surface-directed spinodal segregation	37, 313–316
h-PMMA + d-PMMA	Early stages of interdiffusion	120
	Surface enrichment	317–319
	Thin-film densification	320–322
PMMA/PS	Interfacial width and shape	94, 95, 99, 323
	Effect of molecular weight	306, 324
PS/PVME	Interfacial entanglement density	118, 325
PMMA/PVC	Interfacial width	119
	Surface segregation	326
P*n*BMA/PS	Interfacial width as function of T	98, 103
PPh	Thin-film microstructure	327
PPl	Thin-film composition profile	328
PPV	Layer composition of self-assembling layers	329
h-PS + d-PS	Surface segregation profile	185, 318, 330–332
	Instability of thin-films	333
	Effect of film thickness on surface segregation	123, 330, 331
	Interdiffusion molecular-weight effects	112

Table 2 (*continued*)

Polymer	Description	Refs.
	Interdiffusion	113, 114, 334–338
	Brush penetration and formation	109, 110
	Polymer-network interfacial profile	339
	Triblock-copolymer interdiffusion ripple experiment	340, 341
PS/PVP	Determination of χ	342
SCPE/PMMA	Interdiffusion measurements from interfacial width	116, 117, 343, 344
Copolymer systems		
PE/PS/PS-*co*-PB	Interfacial behavior of copolymer	345
PMMA/PS/PS-*co*-PMMA	Copolymer distribution	125, 126, 346, 347
	Effect of copolymer on interfacial width	126, 344
	Effect of random copolymer	348, 349
PMMA + PS-*co*-PMMA	Copolymer surface segregation	350
	Brush profile	351
PMMA + PS-*co*-PVPh	Surface segregation in thin-films	352
	Interfacial width	353
Polyolefin copolymers	Surface segregation	354, 355
P*p*MS/PS/PS-*co*-P*p*MS	Copolymer segregation effects	121, 356
PS/PCHA-*co*-PBMA	Determination of χ from interfacial width	98, 250, 306
PS + PS-*co*-PBrS	Surface enrichment effects	357, 358
PS/PVP/PS-*co*-PVP	Copolymer interfacial segregation profile	359

[a] Abbreviations (other than those already given in the text and Table 1): EF, end functionalized; *m*PA, poly(*N*,*N*′-dimethylethylene sebacamide); Pα-MS, poly(α-methylstyrene); PBrS, polybromostyrene; PBTP, polybithiophene; P*n*BMA, poly(*n*-butyl methacrylate); PPh, polyphenylene; PPl, polyphenol; PVPh, polyvinylphenol.

Table 3 Polymer dynamics[a]

Polymer	Description	Refs.
Homopolymers		
$LiClO_4$–PPO	Chain dynamics	277
Nylon 6	Molecular vibrations	360
PA	Liquid crystal spacer/mesogenic unit dynamics	361
PB	Microscopic dynamics	150, 362–369
	Pressure effects on dynamics	370
PC	Vibration relaxation crossover	371, 372
PCP	Dynamics near T_g	373
PDMS	Dynamics	374
	Diffusion behavior (cyclics and linear)	375
PE	Molecular motions	154
PEs	Chain dynamics	376
	Determination of shear constants	377
PEB	Dynamic structure factor	378
PEP copolymer	Chain dynamics	368
PIB	Microscopic effects	362, 379–381
PI	Chain dynamics	368, 369, 382
PI-*co*-PB	Segmental dynamics well above T_g	383
PMMA	Me group dynamics	148
	Relaxation processes	384, 385
a-PP	Chain dynamics below T_g	362, 386, 387
PP	Methyl torsion dynamics	388
PPO	Chain dynamics in complexes	389
PPV	Low-frequency vibration modes	390
PS	Dynamics well above T_g	391
	Star arm relaxation	16
PVA	Me group dynamics	392–394
PVC	Segmental dynamics	395–398
	β Relaxation dynamics	399
PVME	Me group dynamics	148, 396, 397, 400, 401
	α Process dynamics	402

(*continued overleaf*)

Table 3 (*continued*)

Polymer	Description	Refs.
Blends		
PMMA + SCPE	Me group dynamics	147, 149, 403, 404
PS + PVME	Me group dynamics	404

[a] Abbreviations (other than those already given in the text and Tables 1 and 2): *a*-, atactic; PCP, polychloroprene.

Table 4 Copolymer morphology[a]

Copolymer	Description	Refs.
PAl-*b*-PEO	Crystallinity effects on morphology	405
PB-*b*-PE	Shear-induced ordering	406
PB-*b*-PS	Bulk morphology	65
	Effects of shear	407
PDMS isotopic copolymers	Triblock bulk morphology	374
PDMS-*co*-PEP	Phase behavior	408, 409
PE-*b*-PEE	ODT behavior	410
PE-*b*-PEP	ODT behavior	410
PEA-*g*-PS	Lamella morphology	411
PEE-*b*-PEP	Thin-film morphology	412, 413
	ODT behavior	410
PEO-*b*-PIB	Triblock morphology	414
PEO-*b*-PPO	Triblock phase behavior	415
PEO-*b*-PS	Bulk morphology	416
PI-*b*-PS	Thermodynamic properties	67, 417, 418
	Triblocks	418–421
	Multiblocks	422, 423
PMMA-*b*-PS	Thin-film lamella ordering	95, 137–142, 144, 347, 424, 425
	T dependence of ODT	138
	Morphology	426
	Effects of thin-film confinement	140–142, 427
PMMA-*r*-PS	Thin-film behavior	428, 429
P*p*MS-*b*-PS	Microphase separation	430, 431
	Effect of pressure	432
PS isotopic copolymers	Triblock copolymer interdiffusion	340
	Multiblock morphology	433
PS-*b*-PVP	Effects on thin-film confinement	70, 434–436
	Bulk morphology	419, 437
	Triblock lamella morphology	438, 439
PU copolymers	Morphology	440, 441
SAN	Surface segregation	442, 443

[a] Abbreviations (other than those already given in the text and Tables 1–3): PAl, polyalkane; ODT, order–disorder transition.

Equation (2):

$$Q^2 = \kappa^2 + \kappa_0^2 - 2\kappa\kappa_0 \cos\theta \qquad (2)$$

and the energy change $\Delta E = (\hbar^2/2m)(\kappa^2 - \kappa_0^2)$. Neutrons which are scattered inelastically either gain or lose energy to the molecular motion. The energies associated with these motions are quantized, and in the case of vibrational motion the energy level spacing is large compared with the neutron energy. The changes in neutron energy are therefore observed as distinct peaks in the neutron energy spectra a distance equal to ΔE either side of the elastic peak (see Figure 2a). The difference in these inelastic peak intensities is due to the difference in occupancy of the two energy levels involved. The energy level spacings for rotational and translational motion are small compared with the neutron energy, which has the effect of broadening the elastic peak (see Figure 2b). Inelastic scattering therefore holds information not only about the spatial but also the dynamic character of the system.

The atomic scattering length, b, determines the amplitude of the wave scattered from a nucleus with respect to the amplitude of the incident wave. As seen for hydrogen

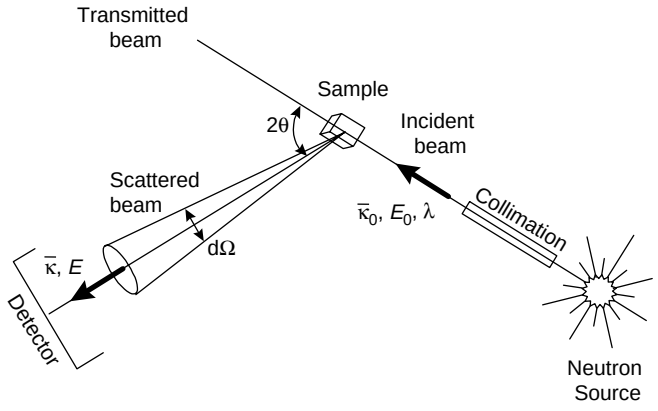

Figure 1 Schematic diagram of a generalized neutron scattering experiment. (Reproduced by permission from Arrighi et al.[8])

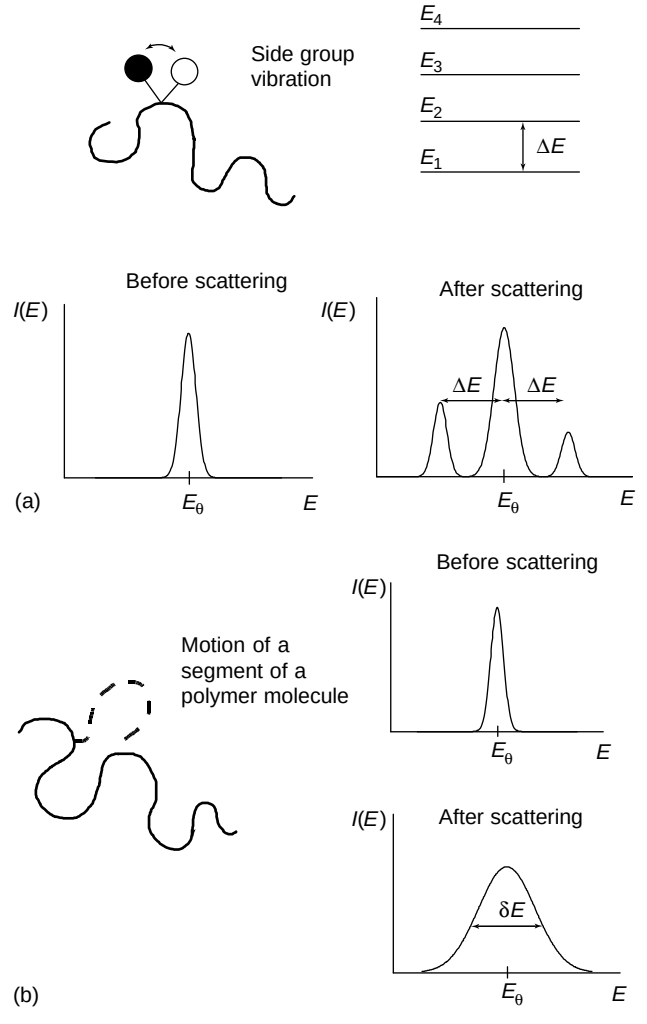

Figure 2 Schematic representation of the effects of inelastic scattering resulting from (a) molecular vibrations and (b) rotational or translation of molecules. (Reproduced by permission from Higgins and Benoit.[11])

and deuterium, it is therefore possible to have both positive and negative values for b. A negative value indicates a half wavelength phase shift on scattering. Tabulated values of b can be found in the literature.[9,10] The square of the amplitude determines the probability that the neutron is scattered somewhere in space. With 4π sr in a sphere this probability is given by $4\pi b^2$, which is called the scattering cross-section, σ. The way in which the neutrons interact with the nuclei depends on their relative spins and type of isotope. This means that the total scattering cross-section of an assembly of nuclei, σ, is made up of coherent (σ_{coh}) and incoherent (σ_{inc}) cross-sections. The coherent scattering arises from interference effects between waves scattered from different nuclei and consequently contains all the structural information on the sample. The incoherent scattering, on the other hand, arises from scattered neutrons which are unaware of any correlation between nuclei. This means that incoherent scattering carries no structural information but is able to be exploited to study motion of atoms and molecules.

Experimentally, what is measured is the double differential scattering cross-section, $\partial^2\sigma/\partial\Omega\partial E$, which represents the number of neutrons scattered per second into a solid angle $d\Omega$ with an energy change between E and $E + dE$, with respect to the incident neutron flux. If the scattering is elastic (with no energy change) then the differential scattering cross-section, $\partial\sigma/\partial\Omega$, is measured. This is usually expressed as $d\Sigma/d\Omega$ (cm^{-1}), which is the scattering per sample unit volume.

2.1 Scattering and Structure

For elastic small-angle scattering the experimentally measured intensity, $I(Q)$, contains all the information on the size, shape and interactions between scattering nuclei. It is possible to write a generalized expression for SANS as in Equation (3):

$$I(Q) = c\frac{d\Sigma}{d\Omega}(Q) = cNV^2(\Delta\rho)^2P(Q)S(Q) + B \quad (3)$$

where c is a collection of instrumental- and sample-dependent constants,[11] N is the number concentration of scattering centers, V is the volume of one scattering center, $\Delta\rho = (N_i\Sigma b_i - N_j\Sigma b_j)$ is the contrast factor equal to the differences in scattering length densities, and B is the background signal and contains all the incoherent scattering intensity. By using isotopic substitution, where for instance H is replaced by D, the contrast between differing segments of a molecule or whole molecules can be enhanced. This enables the experimentalist to choose which part of the system is to be highlighted. The term NV^2 can be rewritten as $(\phi M)/(N_A\delta)$, where ϕ is the weight fraction, M is the effective molecular weight and

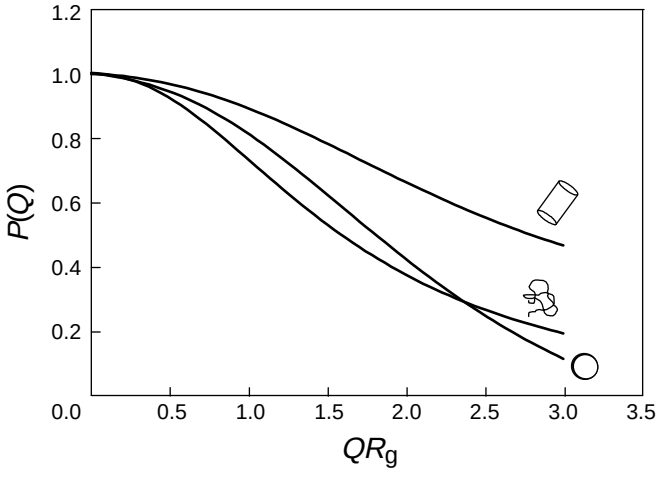

Figure 3 Plot of calculated form factor $P(Q)$ as a function of scattering vector illustrating the differences for sphere, a random coil and a rod-like structure. (Reproduced by permission from Arrighi et al.[8])

δ is the bulk density. The terms $P(Q)$ and $S(Q)$ are the form and structure factors, respectively, and contain all the structural information about the sample. The form factor, $P(Q)$, is related to the average conformation of an individual polymer chain. Model scattering functions have been developed for a variety of particle shapes, a number of which are shown in Figure 3. The structure factor, $S(Q)$, is a probability that the neutrons will be scattered as a function of Q.

The spatial range available with SANS varies between a few and thousands of ångstroms, making it ideal for studying both polymers and blends. The technique is not only suitable to studies of polymer blends in the homogeneous one-phase region, but can also be employed to investigate the kinetics of the phase separation process, the morphology of the two-phase systems which are formed as well as domain sizes. Owing to the length scales over which the polymer can be observed, different features of their size and shape manifest themselves in different regions of Q. It is therefore necessary to analyze the data in a number of ways in order to extract the data required. Many of the analysis methods used in SANS had been developed for small-angle X-ray scattering (SAXS) or light scattering (LS) and can be applied directly to SANS with only minor modifications. As discussed above, the major difference between neutron scattering and other techniques is the contrast between different species that can be induced by deuteration schemes.

The measured intensity in a scattering experiment is proportional to the differential scattering cross-section (see above). In very high Q regimes this is in turn related to the spatial Fourier transform of the atomic pairwise correlation functions. In the simple case of a system composed of identical nuclei, the differential cross-section in the elastic limit is given by Equation (4):

$$\frac{\partial\sigma}{\partial\Omega}(Q) = \sum_{i,j} b_i b_j \exp[iQ(\mathbf{r}_i - \mathbf{r}_j)] \tag{4}$$

where b_i and b_j are the scattering amplitudes of the nuclei in positions $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively, and the angle brackets indicate an ensemble average. The double summation can be separated into self and pairwise terms. The self term contributes only to the incoherent background of the signal and contains no structural information and needs to be removed from the data to obtain the pairwise term, which contains all the structural information of interest. One way to remove the incoherent scattering is by use of deuteration but this is not always necessary.[12]

2.1.1 Single-phase Systems

For a one-phase system it can be shown generally that the form factor for low Q values can always be written as in Equation (5):

$$P(Q) = 1 - \frac{Q^2 R_g^2}{3} + \cdots \tag{5}$$

This relationship is valid for both rigid particles and flexible molecules and is an easy way to determine the size of the molecule. Arguably the two most important form factors for the polymer scientist are for a homogeneous sphere of radius R and that of a Gaussian segment distribution characterized by a radius of gyration, R_g. Assuming spherical geometry the scattering intensity given by Equation (3) can be written in terms of the Guinier approximation by Equation (6):

$$\frac{d\Sigma}{d\Omega}(Q) \approx NV^2(\Delta\rho)^2 \exp\left(-\frac{Q^2 R_g^2}{3}\right. \tag{6}$$

The radius of gyration is determined by the initial slope of a plot of $\ln[d\Sigma/d\Omega(Q)]$ versus Q^2 and the effective molecular weight from the intercept. Similar approximations also exist for a number of simple mathematical geometries such as discs, rods, stars and rings.[11] The closely related Zimm approximation also provides the means of obtaining the values of R_g and M from the initial slope and intercept $d\Sigma/d\Omega(Q)$ at $Q = 0$. The concentration dependence can also be incorporated into the Zimm approximation by including the second virial coefficient, A_2, giving Equation (7):

$$\frac{1}{\dfrac{d\Sigma}{d\Omega}(Q)} \approx \frac{1}{k_1 cM}\left(1 + \frac{Q^2 R_g^2}{3} \quad + 2A_2 k_2 c + \cdots \right. \tag{7}$$

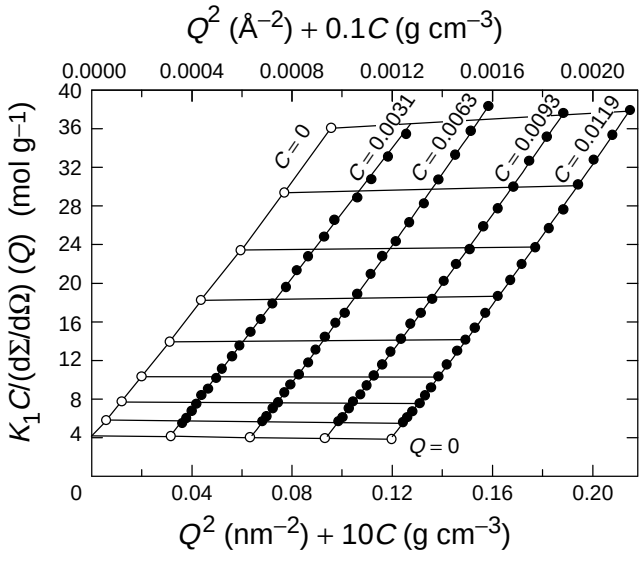

Figure 4 Zimm plot of SANS data from a mixture of various concentrations of h-PMMA dissolved in d-PMMA. The open circles are the result of extrapolation to zero concentration and zero Q. (Reproduced by permission from Kirste et al.[13])

where $k_1 = (\Delta\rho)^2/(N_A\delta^2)$, $k_2 = 1/(k_1 c)$ and c is the concentration. A Zimm plot of $k_1 c/(d\Sigma/d\Omega)(Q)$ versus $c + \alpha Q^2$ where α is an arbitrary constant gives the radius of gyration from the ratio of the slope to intercept. Normally the scattering is collected from a number of sample concentrations and plotted in the Zimm format, extrapolation to zero concentration and Q allows determination of the molecular weight. This value is an important check on the validity of the scattering data. A nice example of this is shown in Figure 4 for low concentrations of hydrogenated (h) PMMA in a matrix of the fully deuterated form (d-PMMA).[13]

Debye demonstrated that the intermediate Q scattering from chains, which exhibit a Gaussian distribution of segment densities about a center of mass, such as observed for polymers in a theta state either in solution or in a melt, is given by Equation (8):[14]

$$\frac{d\Sigma}{d\Omega}(Q) \approx \frac{2NV^2(\Delta\rho)^2}{Q^2R_g^2} \tag{8}$$

A plot of $d\Sigma/d\Omega(Q)Q^2$ versus Q (called a Kratky plot) should asymptotically approach a plateau at high Q for chains which exhibit Gaussian statistics. Non-Gaussian chain behavior such as branched or star polymers produce characteristic deviations to a normal Kratky plot.[11,15,16]

2.1.2 Two-phase Systems

The scattering behavior arising from composition fluctuations in phase-separated polymer blends or copolymer melts can be interpreted in the context of the random phase approximation (RPA).[17] The scattering is therefore given by Equation (9):

$$\frac{1}{\dfrac{d\Sigma}{d\Omega}(Q)} = \frac{1}{(\Delta\rho)^2} \sum_i \frac{1}{n_i v_i \phi_i P(Q, R_{g,i})} + \frac{2\chi_{eff}}{\sqrt{v_1 v_2}} \Bigg\} \tag{9}$$

where n_i is the segment volume for polymers $i = 1$ or 2, v_i is the segment volume, ϕ_i is the volume fraction, χ_{eff} is the effective Flory–Huggins interaction parameter and the form factor is given by $P(Q, R_{g,i}) = 2[\exp(-Q^2R_{g,i}^2) + Q^2R_{g,i}^2 - 1](Q^2R_{g,i}^2)^{-2}$. In the low-$Q$ limit this reduces to an Ornstein–Zernicke function where $(d\Sigma/d\Omega)^{-1}(Q) \propto (Q\xi)^2$, where ξ is the average composition fluctuation length of the composition. Assuming randomly distributed two-phase systems of indefinite size and shape also allows ξ to be obtained from the Debye–Bueche function where $(d\Sigma/d\Omega)(Q) \propto \xi^3/(1 + Q^2\xi^2)^2$.

If the polymer blend consists of two phases with arbitrary shape and orientation but sharp boundaries, as may be expected in immiscible polymer blends, then the high-Q scattering decays as Q^{-4} in accordance with Porod's law as given by Equation (10):[18]

$$I(Q)Q^4 = \frac{c_p}{\rho_p} 2\pi(\Delta\rho)^2 \left(\frac{S_p}{V_p} \right. + I_{inc}(Q)Q^4 \tag{10}$$

where c_p is the particle concentration, ρ_p is the density of the particle, S_p/V_p is the surface area to volume ratio and $I_{inc}(Q)$ is the incoherent scattering which is observed as a flat nonzero background at high Q. As seen below, deviations from true Porod's law arise if the interphase region is diffuse and can be accounted for by suitable changes to Equation (10).

2.1.3 Interfaces and Surfaces

NR is an elastic scattering technique that measures the reflectivity, $R(Q)$, of thin-films and surfaces. $R(Q)$ is the ratio of reflected to incident intensity and is related to the Fourier transform of the scattering length density profile, $\rho(z)$, which is the scattering length density variation as a function of depth, z, perpendicular to the surface. In all but a very limited set of examples, where a kinematic approximation can be applied,[19] the $R(Q)$ data cannot be Fourier transformed into the desired $\rho(z)$ profile. Analysis is usually made by calculation of reflectivity profiles from models, where the thickness, scattering length density and interfacial roughness of each layer within the model are fitting parameters.[20] The reflectivity can be calculated exactly for any model using the equations of Fresnel in addition to those of Snell.[21–23] For all but the simplest case of an infinitely sharp interface between two bulk media, the application

of these equations becomes cumbersome and to aid computational speed matrix methods have been used for calculating reflectivity profiles.[21]

2.2 Molecular Motion and Dynamics

Noting that the neutron energy can be written in terms of angular frequency, so that $\Delta E = hw$, then the experimentally measured double differential scattering cross-section can be written according to Equation (11) in terms of a dynamic scattering structure factor, $S(Q, w)$:

$$\frac{\partial^2\sigma}{\partial\Omega\partial w} = \hbar\frac{\partial^2\sigma}{\partial\Omega\partial E} = \frac{\kappa}{\kappa_0}b^2 N_T S(Q, w) \tag{11}$$

where N_T is the ratio of volume of the sample to the volume of the scattering unit. The measured neutron intensity as a function of both Q and w, $I(Q, w) = c(\partial^2\sigma/\partial\Omega\partial w)$, contains information not only about the motions of the individual nuclei and their motion relative to one another, but also structural information contained within the elastic scattering component. The quasielastic broadening of the elastic peak arising from translational or rotational motion can be described in terms of a self-diffusion coefficient, D_s.[24–26] For a simple center of mass diffusion the scattering law is given by Equation (12):

$$S(Q, w) = \frac{\pi D_s Q^2}{w^2 + D_s^2 Q^4} \tag{12}$$

A Lorenzian shape with a half-width proportional to $D_s Q^2$ therefore describes the scattering. Departures from simple diffusion have also been proposed such as jump diffusion.[26]

Scattering from other motions can be described by models that specify the way in which cooperative disturbances are transmitted along the backbone. Such treatments based on the ideas of Rouse and Zimm use bead and spring models to represent the connectivity of the polymer chain.[27–30] The Rouse model assumes disturbances are transmitted along the chain and has been used to describe molecular motion in the melt. In the Zimm model, the presence of a surrounding medium, such as solvent molecules, via hydrodynamic interactions modifies the behavior of the chain segments. The Zimm model has had particular application in studies of polymers in solution.[31]

Another technique used to measure the dynamics of materials is NSE. This technique was conceived by Mezei[32] and utilizes the polarizable nature of the neutron through its spin ($I = 1/2$) to analyze molecular motion. In a magnetic field the polarized neutron will precess at the Lamor frequency. This Lamor precession is used as an internal clock to monitor the scattering process of the neutron and permits a very precise determination of the energy difference before and after scattering.[33] Any molecular motion is therefore directly observed by changes in the polarization, P, as given by Equation (13):

$$P = \frac{S(Q, t)}{S(Q, t_0)}\frac{\int_0^\infty I(\lambda)\,\mathrm{d}\lambda S(Q, w)\cos(wt)\,\mathrm{d}w}{\int_0^\infty I(\lambda)\,\mathrm{d}\lambda S(Q, w)\,\mathrm{d}w} \tag{13}$$

where λ is the neutron wavelength, w is the angular frequency and t is time. Such polarization changes can only be measured by dedicated instruments that can detect the neutron polarization state.

3 NEUTRON PRODUCTION AND MEASUREMENT

Although the method of neutron production may at first be of little importance to the experimentalist the energy, intensity and time structure of the neutrons are important factors affecting the measurements that can be performed. For this reason, a brief explanation of neutron production and measurement is included here. Regardless of the method of production, the current generation of neutron sources are all relatively weak. The most intense neutron sources provide beams with fluxes at the sample of 10^7 neutrons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ compared with a typical flux from a laboratory laser of 10^{15} photons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This means that in order to make the most of these weak beams, care has to be given to optimize neutron transport from the source and to the construction of the instruments themselves.

Methods of neutron production are basically divided between nuclear reactors and spallation sources. Nuclear reactors for neutron production such as the ILL reactor at Grenoble, France, are not the same as power-generating reactors since these latter reactors are not designed to allow radiation to be emitted from the reactor core. In neutron production reactors, holes in the core container are essential to allow the neutron to be emitted. Within the reactor core, fission reactions often from enriched uranium (^{235}U) produce high-energy neutrons (1–2 MeV). These neutrons are too highly energetic for using in scattering experiments and are therefore passed through moderators most probably containing D_2O, H_2O or graphite, which cause the neutrons to be multiply scattered, losing energy with each collision. The emerging neutrons will therefore have energies dependent on the moderator temperature and described by a Maxwell–Boltzmann (MB) velocity distribution. For a water moderator at 50 °C this gives a peak flux with a wavelength of 1.4 Å. Owing to the nature of the fission

process, reactors produce a continuous source of neutrons as a function of time. Neutron flux is limited to a large extent by heating effects of the fission process within the reactor, since the maximum flux is produced as a balance of the heat dissipation needed within the core. Owing to the continuous nature of neutron production, one specific wavelength is chosen for a specific neutron spectrometer depending on the types of measurements required to be performed. Wavelength and energy selection is often achieved by velocity selecting rotating choppers or by reflection from a single plane of a crystal.

By contrast, spallation neutron production uses a beam of highly energetic particles such as protons which chip off neutrons from the heavy nuclei of the target material. The neutron yield produced per proton is fairly high with minimal heating effects, but is set against the extremely high cost of producing the protons. At the ISIS neutron spallation source at the Rutherford Appleton Laboratory, UK, 800-MeV protons are produced by combining a linear accelerator with a synchroton. The resulting protons and hence the neutrons are produced in short pulses with a peak flux which is larger than reactor sources. However, between each pulse the neutron flux falls to zero so the time-integrated average flux is less than for a reactor. As with reactors, the neutrons produced are highly energetic and have to be moderated. At the ISIS facility three types of moderator are used, H_2O at ambient temperatures, liquid methane and liquid hydrogen. The neutrons passing through the moderators are characterized by a wide MB velocity distribution giving a ''white'' beam of neutron wavelengths within each pulse. Velocity and hence wavelength determination in such white beams are determined by time-of-flight (TOF) methods. Here the time taken for the neutron to travel the known distances from the target–moderator assembly to the sample and detectors is measured, giving the velocity and through de Broglie's relationship the wavelength. Energy selection for white-beam instruments can be performed either before or after scattering from the sample and again depends on the spectrometer design and the information required.

To study the conformation, morphology and dynamics of polymeric systems, a wide range of both momentum and energy transfer must be covered. This is not possible with one spectrometer alone but requires dedicated instruments, each of which is optimized to solve specific structural or dynamic problems.

4 APPLICATIONS OF NEUTRONS TO POLYMERIC MATERIALS

The following discussion outlines some of the many uses that neutrons have been applied to for studying polymer systems. Although specific systems are mentioned in the text, this represents only a small number of those that have been studied. In order to supplement the information in the text, Tables 1–4 contain a more comprehensive listing of systems and references found in the literature, mostly within the last 20 years. Although the intention was to be as thorough as possible, these tables certainly do not contain all the references on polymer systems studied by neutrons, which would run to several thousand for the past two decades alone.

4.1 Effects of Deuteration

Although deuteration is widely used in neutron scattering, it does have an influence on the behavior of the polymer mixture. At higher concentrations of labeled polymer it was observed that for isotopic blends of deuterated and hydrogenous analogs of the same polymer phase separation occurs. This was first observed using SANS for PB,[34] PS[34] and PDMS.[35] Surface segregation has subsequently been observed using NR for isotopic blends in thin-films.[36,37] The segregation in these isotopic mixtures occurs due to a weakly unfavorable χ interaction parameter originating from the difference in the polarizability between the C−H and C−D bonds.[36] The effect is only observed if $\chi N > 2$ is satisfied, where N is the degree of polymerization. Since χ is typically very small for isotopic mixtures [χ for h-PS/d-PS is 2×10^{-4}[38]), N must be large before separation occurs.

High levels of labeling which can increase the signal intensity can lead to difficulties due to these isotopic effects. By working at low levels of deuteration it is possible effectively to remove the effects of labeling. Another possibility is to measure a number of samples with different deuteration levels and extrapolate to zero concentration of labeled polymer.

4.2 Determination of Molecular Parameters

As seen from the equations above, detailed analysis of the Q dependence of the SANS profiles provides information on the polymer chain conformation and also the interaction parameter. Using the Debye or RPA equations, the radius of gyration of the polymer can be calculated. This also applies to blends of polymers in the one-phase region where radii of gyration for each of the blend components can be obtained. Analysis of SANS from blends of PS and PVME[39] and PS and P*p*MS,[40] for instance, have used weighted average radii of gyration as fitting parameters of the data.

The radius of gyration has been studied as a function of blend composition and temperature. In a PMMA blend with SAN annealed at 140 °C, the average radius of gyration goes through a minima at an approximate 35%

volume fraction of PMMA.[41] Excess scattering was also observed and has been interpreted in terms of effective conformational changes of the PMMA due to the affect of rotational states on intermolecular interactions or compositional variation in the interaction parameter. Similar composition and temperature dependence has also been observed in other blends such as PS–isotactic(*i*)-PVME deriving from *i*-PVME self-correlations.[42]

Isotopic polymer mixtures have proved to be ideal model systems for exploring the limitations of the widely used Flory–Huggins theory and for testing and improving new theories of polymer thermodynamics. These isotopic blends lack the structural asymmetries and specific interactions encountered in blends of chemically distinct species. The Flory–Huggins interaction parameter, χ, between deuterium-labeled and unlabeled segments of the same species should in principle be independent of concentration (ϕ), but SANS experiments have shown that it exhibits a minimum at $\phi \approx 0.5$ for PVE, PEE[43] and PE.[44] By comparison, isotopic mixtures of PS show a maximum in χ as shown in Figure 5.[44]

Some blends have demonstrated a χ dependence on not only composition but also molecular weight as seen from the SANS obtained from PS/PVME[39] and PEO/PMMA.[45] The origin of this composition dependence is still unclear but may be due to volume changes on mixing and concentration fluctuations.[46–48]

Blend miscibility can be affected dramatically under the influence of pressure, for instance upward shifts of 25–30 °C kbar^{-1} of the spinodal have been observed for PS/PVME blends.[49] Typically the pressure dependence of the χ parameter is determined, and separated into a pressure-independent enthalpic contribution and an entropic component which decreases with increasing pressure. Such interpretations have been applied to SANS data obtained from blends of PS/PVME and PS/PPMS which exhibit lower and upper critical solution temperatures, respectively.[50,51] Pressure has also been shown to have an effect on the chain dimensions, where an increase in pressure results in a reduction in chain dimensions.[49,52,53]

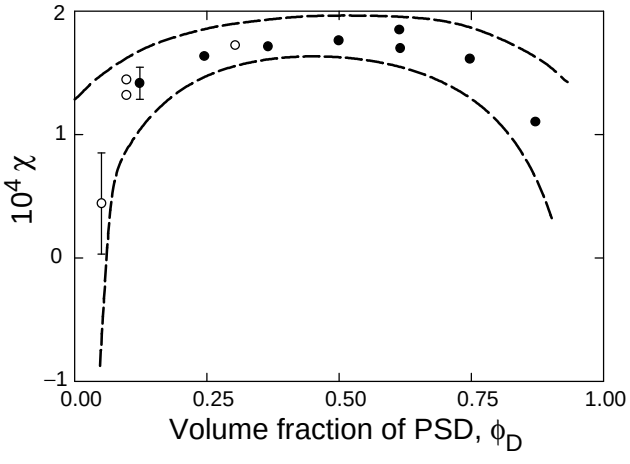

Figure 5 Plot of Flory–Huggins interaction parameter versus volume fraction of d-PS in a blend with h-PS calculated from SANS data taken at 160 °C. The degree of polymerization of d-PS is 11 500 and that of h-PS is 8700 (solid circles) and 15 400 (open circles). The dotted lines represent error limits on the interaction parameter due to systematic error in degree of polymerization and absolute intensity. (Reproduced by permission from Londono et al.[44])

The study of short-range order in disordered polymers is particularly challenging owing to both the structural complexity and absence of periodic lattices in these materials. The complexities arise from the variety of both intra- and interchain correlations. WANS has been used to obtain quantitative information on the nature of the chain segments, in terms of local chain conformation and chemical configuration, and on the spatial and orientational correlations of segments.[12,54,55]

WANS experiments on *a*-PS have shown that true quantitative neutron scattering data to be obtained over a large scattering vector range for hydrogenous polymers. Results for samples of both fully protonated PS in comparison with its deuterated analog have shown that investigating the local structure of hydrogenous polymers provides detailed information about the polymer without the need for deuteration, making it an important tool for commercial systems.[12] Despite this, isotopic substitution is an important tool for such studies. The cation environment in molten lithium iodide-doped PEO has been studied using SANS utilizing ^{6}Li and deuterated PEO.[56] The experimental results of these studies have been confirmed using molecular dynamic simulations and show that the Li^+ has a coordination number with oxygen of 3.5, and changes of the coordination with variations in Li–O bond length have been mapped.

4.3 Phase-separated Systems

The scattering behavior from a two-phase blend is distinctly different from that observed when the system is in the one-phase regime. As with LS, this distinction provides a means of locating the phase boundary. The phase separation process can also be followed and because the scattering behavior is different for the nucleation and growth and spinodal decomposition mechanisms, it is therefore possible to easily distinguish between the two. The phase separation process in polymers is slow enough that time-resolved scattering experiments can be performed. In the very early stages of spinodal decomposition the development of regular co-continuous

structures give rise to a maximum in the scattering pattern. The maximum occurs at a specific Q value (Q_m) which is indicative of a preferred length scale in the phase separated morphology, given by $\Lambda_m = 2\pi/Q_m$. Initially the characteristic size of the fluctuations remains constant despite the increasing difference between the two phase compositions, and therefore the position of the characteristic spinodal peak at these early times remains constant.[57–60] At longer times the position of Q_m moves to lower Q,[61] resulting in morphology which becomes indistinguishable from that developed during nucleation and growth. Nucleation and growth mechanisms do not produce morphologies with regular structure and thus lack a maximum as seen for spinodal decomposition. The phase separation is, however, characterized by a significant increase in scattering, and is analogous to the cloud point observed in LS. Owing to the differences in domain sizes measured by LS and SANS the onset of these cloud points may differ.

In general, the use of SANS to study two-phase morphology is not the technique of choice, since other more available techniques can provide the information required. However, in a number of cases the use of neutrons is the only way of obtaining the information. Such cases arise, for example, when the samples are too thick or opaque for other scattering methods or when no contrast can be obtained other than by deuteration, as may be experienced for polyolefin blends.[62–64] SANS has been widely applied to studies of polymer micelles and also block copolymer morphology.

The morphology of block copolymers has been investigated extensively both in the bulk and in concentrated solutions. If the blocks of the copolymer are incompatible microphase separation may occur leading to phases with spherical, cylindrical, gyroid or lamellar morphology depending on the block lengths and matrix. For instance, the morphology of a range of PS/PB diblock copolymers has been investigated which show well-resolved Bragg peaks originating from a macrolattice of body-centered cubic packing (BCC) of PB spheres.[65,66] Studies of PS-block-PI (PS-*b*-PI) using SANS has also shown spherical,[67] lamellar and cylindrical morphologies (see Figure 6).[68,69] Lamellar domain spacings of copolymer blends with their homopolymers and conformations of block chains in their microdomains have been studied for PS and PVP homopolymers and copolymer.[70] SANS studies of PE/PEP diblock copolymers blended with PE and PEP homopolymers have also been used to test theories where melt behavior is controlled by a common interaction parameter.[71]

4.4 Shear and Deformation

By using large-area detectors, SANS instruments are able to observe any deviation from isotropic scattering,

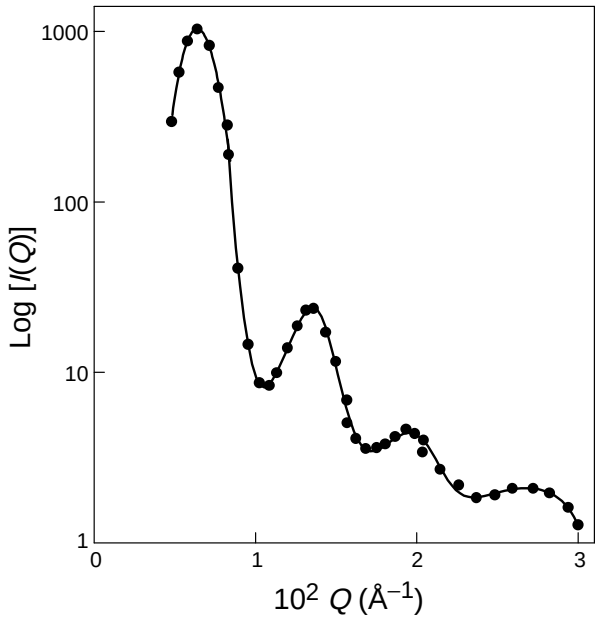

Figure 6 SANS profile of a styrene–isoprene block copolymer with $M_n = 178\,100$ and polydispersity 1.06. The weight fraction of styrene is 0.73. The interdomain distance obtained from the position of the first Bragg peak maxima is 95.2 nm. (Reproduced by permission from Richards and Thomason.[67])

as may be experienced by systems with defined order, or preferred orientation, from either single chains or particulates. In this respect, SANS has been used to study deformation and relaxation processes of polymer chains.[72–76] The scattering in these oriented polymer systems now displays an anisotropic scattering pattern which is typically ellipsoidal in shape, although other shapes have been observed. Relative to the extension or shear propagation axis, the radius of gyration can then be analyzed with respect to these orthogonal directions.

There is considerable interest in the behavior of polymer solutions under shear and SANS has been applied to help interpret the morphological changes that occur under such external force fields.[77] The high penetrability of the neutron makes in situ rheological measurements possible in any required geometry, including Couette,[78,79] parallel plate[80] and Poiseuille.[81,82] Chain deformation by the shear flow creates changes in the polymer chain conformation, which can be observed, by changes in the scattering intensity parallel and perpendicular to the flow creating anisotropic scattering. Polymeric micelles formed by copolymers at high concentrations can form high degrees of ordering such as BCC or face-centered cubic (FCC) polycrystalline structures.[79,80] The steady-state ordering behavior can be disrupted by application of shear such as observed for PS-*b*-PI copolymers in decane where the FCC structure shows a transition to (111) sliding layers. As the shear rate is increased these (111) planes

are oriented normal to the shear gradient. In contrast the BCC polycrystals form (110) layers normal to the shear gradient.[79] Similar long-range ordering has also been observed for diblock copolymers of poly(styrene-*b*-ethylene-*co*-propylene) in the selective solvent dodecane at concentrations above 3 wt%, where the shear rate affects the lattice ordering.[83] Another method of alignment is in magnetic fields and a number of studies have shown anisotropic scattering of LCPs. For example, a main-chain semiflexible copolyester has been shown to align using a magnetic field of 0.6 T in the nematic phase between 160 and 250 °C.[84,85] Using SANS, the chain anisotropy, radius of gyration, average degree of polymerization and persistence length of these oriented polymers have been obtained.

Not all scattering from oriented polymers is so straightforward and unusual anisotropic scattering such as figure of eight "butterfly" patterns has been observed for uniaxial extension of polymer networks or swollen gels.[73,75,76,86] This anomalous scattering behavior has been studied using deuteration contrast variation methods and indicate that the butterfly scattering patterns from rubbers originate from the interchain structure factor. This implies that there are spatial concentration fluctuations in labeled free chains on length scales that are larger than the characteristic mesh size of the system. This contradicts the classical hypothesis of molecular rubbers and has led to the postulation of the cross-linking process introducing random large-scale spatial fluctuations in elastic modulus, so-called hard and soft zones. The butterfly patterns would then result from anisotropic spatial separation of the hard zones, which would amplify concentration fluctuations in the stretching direction.[87,88] In gels and swollen networks, theories based on thermodynamic and frozen concentration fluctuations have been suggested to explain the appearance of butterfly patterns.[89,90] These theories have been critically tested by analyzing the butterfly patterns obtained by randomly cross-linked PS gels swollen by deuterated toluene.[86]

In recent work, McLeish et al. used SANS to help understand the rheological behavior of PI with H-shaped architecture,[91] which can expect to show an intriguing combination of features of star and linear polymers. H-polymers have four star-like arms which must relax by linear pathlength fluctuations, but are connected by a "cross-bar" which cannot move at all in the entangled state. Until the dangling arms retract to their branch points the crossbar cannot move and then motion occurs in the same way that star polymers diffuse. By measuring the anisotropic scattering (see Figure 7), and therefore morphology obtained from stretched and instantly quenched samples of carefully deuterated H-shaped PIs, the theories relating to their retraction dynamics have been studied.[91]

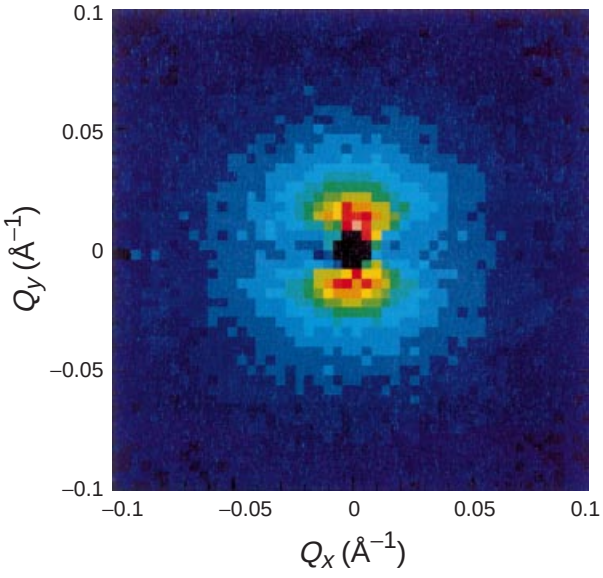

Figure 7 SANS pattern from an H-shaped polymer taken after a extensional stretch of 2.3. The direction of stretch is in the vertical direction. A pronounced asymmetry is observed consistent with theoretical calculations. (Reproduced by permission from McLeish et al.[91])

4.5 Homopolymer Interfaces

Many homopolymer systems are characterized by narrow interfacial widths owing to the nature of their immiscibility. The width of these interfaces can be described in terms of self-consistent mean-field theories and is simply determined by the Flory–Huggins interaction parameter, χ, and defined in Equation (14):[92]

$$w_\infty = \frac{2a}{\sqrt{6\chi}} \tag{14}$$

where a is an average segment length of the two polymers and the ∞ symbol indicates that this equation strictly applies only to infinite molecular weights. A common immiscible polymer pair of PMMA and PS with a χ of 6×10^{-2}[93] gives $w = 2.6$ nm. With its exceptional spatial resolution, NR is ideally suited to studying surfaces and interfaces on such length scales and has been widely used for such studies.[94–99] The shape of the interfacial width as described by theory to be a hyperbolic tangent function has been confirmed using NR.[100] Early NR measurement of these immiscible homopolymer systems always obtained interfacial widths larger than that predicted by theory.[94,95] This discrepancy can be accounted for by considering perpendicular fluctuations to the interface. The apparent broadening observed by NR can be attributed to thermally induced capillary waves.[96,99,101,102] The measured interfacial width w is simply given by the quadrature addition of the capillary wave amplitude and the theoretical width, so that

$w = (w_0^2 + w_c^2)^{1/2}$, where w_0 is the intrinsic (unbroadened by capillary waves) interfacial width as predicted by theory and w_c is the capillary wave amplitude. The capillary wave component of the interfacial width can be defined as in Equation (15):[99]

$$w_c^2 = \frac{k_B T}{2\gamma_0} \ln \frac{(2\pi/\Delta_0)^2}{(2\pi/\lambda_c)^2 + (2\pi/a_d)^2} \qquad (15)$$

where $\Delta_0 = w_c/\sqrt{2\pi}$. The capillary wave broadening is therefore dependent on the in-plane coherence length of the neutron, $\lambda_c \approx 20\,\mu m$, the dispersive capillary length, $a_d = 4\pi\gamma_0 d^4 A^{-1}$ (d is the layer thickness), the Hamaker constant, A, and the interfacial tension $\gamma_0 = av^{-1}k_B T\sqrt{\chi/6}$ (v is the monomer volume). As can be seen from Equation (15), determination of the interfacial width using NR is therefore a route to determining the Flory–Huggins interaction parameter, χ, and γ_0.

This approach to determining interfacial parameters has been applied to a number of homopolymer systems: PS/PMMA,[94,95] PS/PBrS,[96,97] and PS/P*n*BMA.[98,103] Most studies have been carried out at room temperature on samples that have been annealed to equilibrium at temperatures above the glass transition temperature, T_g, and rapidly quenched. The interfacial properties at equilibrium are therefore assumed to be trapped and immobile at the room temperature. This approach does not allow investigation of crystalline polymers where the macroscopic roughness associated with the crystallinity prevents room temperature measurements.[104] By using a heated cell which confines the polymer films it has been possible to measure crystalline film interfaces in situ at elevated temperatures. A number of such systems have been studied using various grades and tacticities of PE, and PP, PMMA and PS.[104–106] For blends of PE and PP the interfacial width is seen to be larger for those containing (*a*-PP) in comparison with to its *i* form (*i*-PP).

Polymer adsorbed layers are widely used for many applications, such as viscosity modifiers and dispersion stabilizers in colloidal dispersions.[107] SANS has been used to study the surface-adsorbed polymers on spherical polymer lattices, inorganic oxide particles or emulsion droplets, whilst NR utilizes flat surfaces.[108–110] These techniques are able to provide the component volume fraction profiles at the interface and the amount of adsorbed polymer. This requires careful design of the experiment since meaningful interpretation from SANS data is only possible if the scattering is from simple geometries, with minimal interlayer interactions, excluded volume effects and multiple scattering effects. If the layers become too extended into the melt or solution then NR will also not be able to distinguish the polymer from the matrix and the volume fraction profile in these tail regions.

4.6 Polymer Interdiffusion

The observation of the dynamics of the interdiffusion process is ideally suited to investigation by NR, where it is possible to resolve the early stages where only chain segments move across the initially sharp interface.[103,111–120] At times smaller than the reptation time of the molecule, several time dependences predicted by theory have been observed using NR (see Figure 8).[111] Only after the interface has become sufficiently large, comparable to the polymer chain dimensions, does it achieve essentially bulk toughness.[121]

Using multiple alternating layers of PS and d-PS, the effect of chain confinement has been observed when the layer thicknesses are on length scales approaching or below the overall dimensions of the polymer.[122] The multilayer composition creates Bragg peaks in the reflectivity profile which disappear with interdiffusion of the layers. The change in the composition profile is seen to be diffusive even for times less than the reptation time, with a Q-dependent diffusion coefficient that decreases with increasing Q. Related work on this system studying the effects of confined geometry and substrate interaction on the short-time interdiffusion in polymer films has also been studied. Only a slowing of interdiffusion dynamics for thin films sandwiched between two thicker ones is detected as well as a still smaller interdiffusion for thin films placed directly on the substrate, which is attributed to polymer–substrate interaction.[123]

An alternative approach to measuring diffusion using neutrons is, of course, with SANS. One such example is the diffusion observed in miniemulsified lattices of PS and d-PS.[124] Samples containing a small fraction of

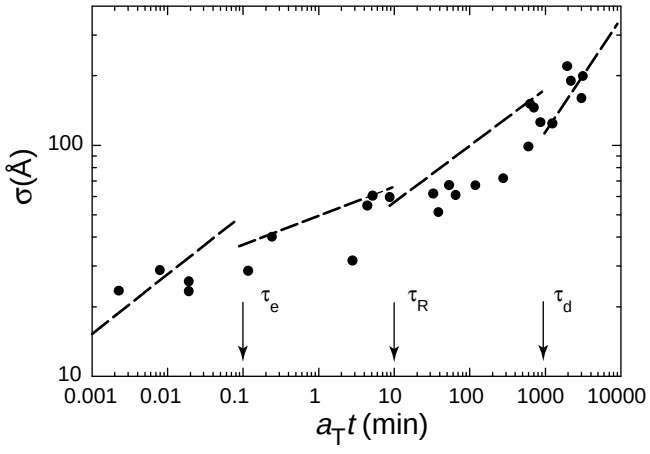

Figure 8 Interfacial width (root-mean-square half-width) between h-PS and d-PS as a function of reduced time at a reference temperature of 120 °C. The arrows indicate the times corresponding to the Rouse entanglement (τ_e), Rouse relaxation (τ_R) and reptation (τ_r) times. (Reproduced by permission from Karim et al.[111])

deuterated particles were annealed well above T_g and the average depth of penetration of the deuterated chains in the protonated matrix was measured during the annealing process.

4.7 Copolymers at Interfaces

Reflectivity studies have contributed greatly to the understanding of the behavior of copolymers at homopolymer interfaces.[125,126] Studies using NR have shown that it is possible to determine quantitatively the amount of copolymer segregated to the interface and also the distribution of copolymer segments at the interface. By extensive use of deuterium labeling, the precise distribution of a PS-*b*-PMMA diblock copolymer at the interface between PS and PMMA homopolymers has been determined.[125] The width of the PS/PMMA interface including the copolymer is 25% larger than that of the bare interface in the absence of copolymer, indicating a significant overlap of the homopolymers.

When the interface between two homopolymers becomes diffuse, such as is the case when a compatibilizing copolymer is added to the mixture, SANS data can no longer be analyzed by the Porod law described above. Boundaries of finite width can be incorporated by convolution of a smoothing function[127,128] expressed in terms of the Fourier transform function, $H^2(Q)$, which is related to the density profile across the interface, allowing the determination of the interface thickness, d_I. A relatively good approximation for this Fourier transform function is a Gaussian distribution.[129] The Porod scattering from a two-phase blend with diffuse interfaces is given by $I(Q) = P(Q)H^2(Q)$ with the analytical expression for $P(Q)$ varying depending on the phase morphology.[130] The scattering from this type of system can therefore be expressed by Equation (16).[130]

$$\ln[I(Q)Q^{\eta}] = \ln \kappa_m - \sigma_d^2 Q^2 \qquad (16)$$

where κ_m is a parameter describing the morphology of the particle shapes according to the value of η, which takes a value of 4 for spherical, 3 for cylindrical and 2 for lamellar morphology. The interface thickness, d_I, can be determined from the second moment of the density variation, o_d^2, from the relationship $d_I = \sqrt{12\sigma_d}$.[130]

This method of analysis has been applied to blends of PS and PMMA compatibilized with the deuterated low molecular weight diblock copolymer d(PS-*b*-PMMA), where Porod scattering indicates that lamellar structures are present ($\eta = 2$).[131] This behavior can be understood in terms of the interface curvature and packing requirements of the copolymer. The interface thickness as a function of copolymer volume fraction is shown in Figure 9. These values of d_I can be compared with the theoretical predictions of Helfand and Wasserman.[132–135]

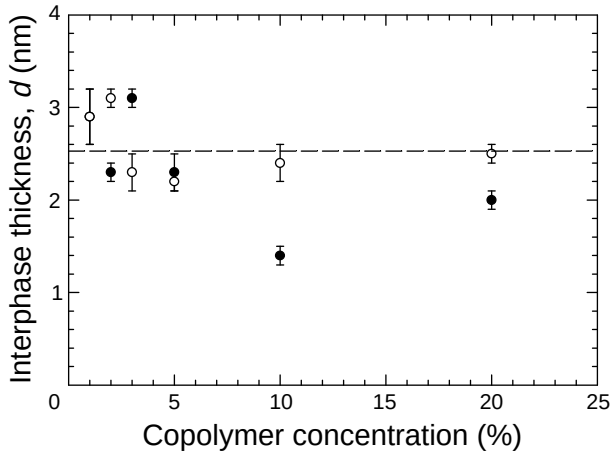

Figure 9 Interphase thickness width between PS and d-PMMA in a blend as a function of concentration of diblock copolymer of PS-*b*-PMMA, obtained from fits to a modified Porod plot of SANS data. (Reproduced by permission from Bucknall.[131])

In thin-films, the microdomain morphology of diblock copolymers is greatly influenced by the interactions with the surfaces. NR studies have been able to provide detailed understanding of these systems. Oscillatory concentration depth profiles are expected to be observed perpendicular to the surface owing to the segregation of the chemically distinct but molecularly connected segments of the copolymer. Such ordering has been observed when diblock copolymer films are on silicon,[95,136–139] or confined between solid walls.[140–142] The lamellar ordering is due to the preferential segregation of one of the blocks to the surface. These lamellae have been shown using NR to have a very regular structure provided that the film is not so thick as to allow disorder to be introduced into the bulk of the film.[138] If the total film thickness is not a multiple of the lamellar period then frustration can result in island formation at the free surface.[140,143] In films which are confined between solid surfaces, this frustration cannot be removed by island formation. Under these circumstances, the frustration is distributed throughout the sample by slight increases in the lamellar period, and as the total thickness increases there is a point at which the frustration, rather than distributing throughout the film, creates another lamellar layer. If the lamellar thickness is plotted against the total film thickness, a series of discrete steps is observed (see Figure 10).[144]

4.8 Polymer Dynamics

The information obtainable by INS and QENS overlaps considerably with other experimental techniques, in particular NMR,[145] IR and Raman spectroscopy, photon correlation spectroscopy and dielectric measurements.

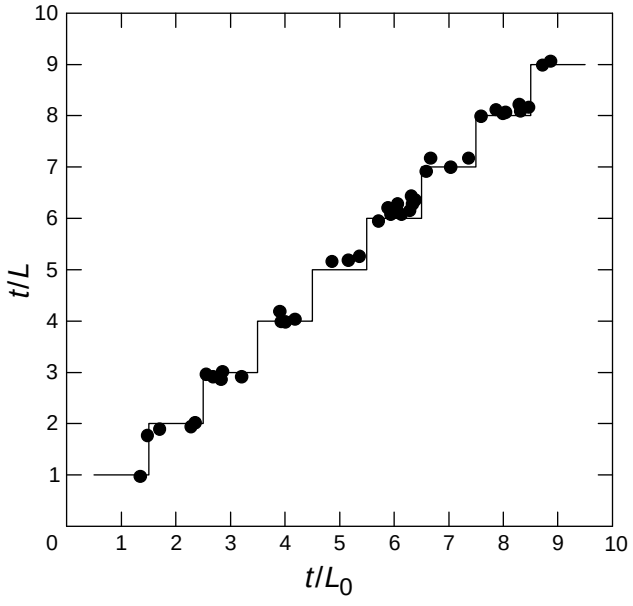

Figure 10 Number of periods confined between two solid surfaces, t/L, for a d-PS-*b*-h-PMMA copolymer, as a function of reduced film thickness, t/L_0. (Reproduced by permission from Russell.[144])

This has meant that INS and QENS have not achieved the widespread application to polymers as perhaps SANS and NR have enjoyed. However, in certain circumstances neutrons provide distinct advantages over other techniques which have necessitated their use, such as lack of selection rules governing the observable motions and contrast enhancement with deuteration schemes. In addition, the technique inherently conveys direct information about both spatial and dynamic information, which cannot be obtained by other techniques and can be used to discriminate between inter- and intramolecular effects. Reviews of applications to polymers have been published.[11,146]

Even in the solid state, polymer chains are very mobile, and as a consequence analysis of their dynamics relies on decoupling the various kinds on motion that the chains may undergo. By assuming that the timescales over which these motions occur are very different, it is possible to separate translations, vibrations and rotations. The total scattering law is then a convolution of the respective translational, vibrational and rotational laws. These can be separated experimentally by choosing the appropriate energy and time domains to observe the motion required. In general, translational motion broadens the whole of the elastic peak, while rotation gives a sharp elastic peak with a broadened foot. Vibrational motion, on the other hand, gives sharp peaks either side of the elastic peak that are analogous to the Stokes and anti-Stokes lines observed in Raman spectroscopy. Experimentally, many inelastic translations may be observed corresponding to the many vibrational modes in the sample.[11]

The observed inelastic and quasielastic scattering behavior are characteristic of the polymer.[11] The INS from the crystalline polymer PTFE shows an intense unbroadened elastic peak and well-defined inelastic structure that varies strongly with scattering angle and is due to phonons within the crystalline regions. By contrast, for an amorphous polymer such as PMMA, the inelastic phonon scattering is much less apparent with only a broad inelastic peak, characteristic of molecular vibration. In this case this has been attributed to torsional motion of the α-methyl group. A slight broadening of the elastic peak (quasielastic broadening) at its base is also observed deriving from molecular rotations and can be assigned to deriving from the ester methyl group. An INS spectrum of PDMS rubber shows strong inelastic scattering which is Q dependent arising from the methyl group torsional mode. Quasielastic broadening is also present and is due to a main-chain ''wriggling'' motion which is characteristic of rubbery state polymers.

To date only a limited number of studies have been reported on dynamic studies on polymer blends. Blends of selectively deuterated PMMA (d_5-PMMA) and SCPE and d-PS/PVME have been studied in detail to investigate the effect of blending on the local motions below T_g.[147,148] Measurements on these systems have been performed on these blends to investigate the effect blending has on the rotational motion of the ester methyl group in PMMA and the OCH_3 moiety in PVME. In the d_5-PMMA/SCPE blend the scattering from the SCPE is purely elastic owing to the absence of side-groups. The d_5-PMMA contributes to both the elastic and quasielastic components, where the latter is described by rotational hopping of the CH_3 groups over the threefold potential barrier.[147–149] The quasielastic broadening, which can be simply modeled by a Lorentzian function convoluted with the instrumental resolution, is considerably reduced in the blend compared with pure PMMA. The activation energy that characterizes the motion of the CH_3 increases with increasing SCPE content, indicating that the rotational motion is hindered in by the SCPE. By comparison for the d-PS/PVME system the presence of the d-PS had no effect on the rotational motion on the OCH_3 group.[148]

NSE has been used to check the validity of the reptation model at the molecular level, since it is able to observe individual molecules in crowded surroundings. Using this technique, many studies have been made to understand polymer reptation.[5,6] Using blends of poly(tetrahydrofuran) (PTHF) of high and low molecular weight, the effects of chain entanglements were investigated. Two samples were studied, where high molecular weight PTHF was blended in a matrix of either low or high molecular weight deuterated PTHF. Since the matrix

polymer was deuterated it is possible to measure the intermediate structure factor $S(Q,t)$, which derives from the Fourier transform of $S(Q,w)$. The NSE data for the low molecular weight matrix could be explained using simple Rouse-type behavior, but could not be used to explain the data for the high molecular weight matrix sample. Using a reptation model these data could be explained and gave a tube diameter, D_r, for the reptating polymer of 3.0 nm. Improvements in instrumental resolution of the spin-echo instruments will allow more detailed analysis of reptation theory to be tested.

The use of inelastic scattering has been pursued for investigations into the motion responsible for the α-relaxation process associated with the glass transition temperature. Among other techniques applied to these studies, incoherent QENS may be able to provide insights into the local motions of this α-relaxation process. Using simple polymers such as PB, main-chain motions such as conformational transitions have been shown to occur alongside cooperative motion.[150,151] Recently, the mode coupling theory of glass transition has attracted the attention of neutron scatterers since the predicted singularity in the density fluctuations can be measured directly by incoherent QENS. The divergence observed in the vibrational amplitudes of the Debye–Waller factors are as predicted by the mode coupling theory, although a number of questions remain open as to the validity of the model for all glassy polymers.

INS has also been applied to the study of microphase separation in ionomers of PS.[152] The differences in *a* and *i*-PS in comparison with sodium sulfonated PS (Na SPS) were observed and vibrational modes due to ring and chain moieties identified by selective deuteration. A typical set of INS spectra for these polymers is shown in Figure 11. Sulfonation was shown to have an effect upon the spectrum of *i*-PS very similar to that observed with *a*-PS. Observation of a new mode at low wavenumbers indicates that the process of sulfonation results in a release of a low-frequency ''breathing mode'' between the chains. Detailed studies of both randomly oriented and stretched PE have been used to understand the crystallization behavior and to test quantitative spectra calculations using ab initio methods.[153,154]

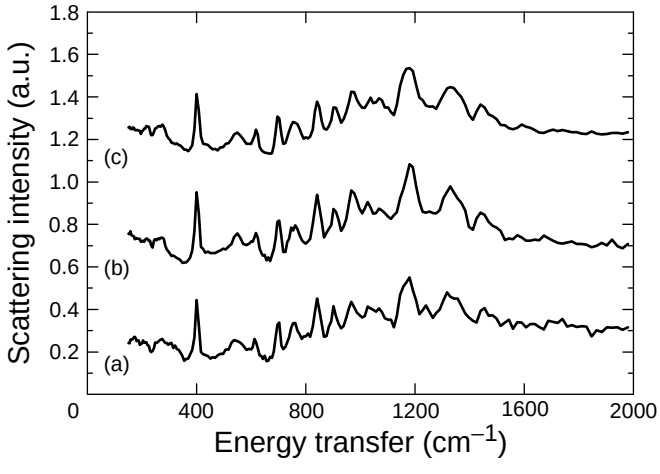

Figure 11 Energy-transfer spectra obtained from an INS instrument for (a) *i*-PS, (b) 10.5 mol% sulfonated *i*-PS and (c) *a*-PS. (Reproduced by permission from Gabrys et al.[152])

5 CONCLUSIONS

Since the 1950s when neutrons were first applied to condensed matter physics, there has been an enormous growth in publications where neutrons have been applied to experimental studies. This overview has only been able to give a flavor of the huge amount of published literature available on polymer studies with neutrons, and it is hoped that it is clear that in many ways the neutron is the ideal probe for studying polymers. Combined with the use of deuteration, neutron scattering has provided information about both structure and dynamics not obtainable by other techniques.

The development of improved instruments at existing neutron sources means that the application of neutrons to answer the many problems of polymer science has a positive future. However, with the possibility of the next generation of neutron sources that are at present either planned or under construction, another explosion of interest in neutrons is anticipated. These new sources, with the potential for much higher fluxes, will make possible a whole host of experiments that have been waiting for such advances. Indeed, there are bound to be many experiments performed on these new sources which have not yet even been conceived. Of the large number of significant advantages such high flux sources will provide, the possibilities of smaller sample volumes, better resolution and even true time-dependent studies are exciting prospects for the future.

ABBREVIATIONS AND ACRONYMS

a	Atactic
BCC	Body-centered Cubic Packing
BPA	Bisphenol A
EF	End Functionalized
EPDM	Ethylene–Propylene–Diene Rubber
FCC	Face-centered Cubic
i	Isotactic
INS	Inelastic Neutron Scattering
IR	Infrared
LCP	Liquid Crystal Polymer
LS	Light Scattering

MB	Maxwell–Boltzmann
*m*PA	Poly(*N*,*N*′-dimethylethylene sebacamide)
NMR	Nuclear Magnetic Resonance
NR	Neutron Reflection
NSE	Neutron Spin-echo
ODT	Order–Disorder Transition
Pα-MS	Poly(α-methylstyrene)
PA	Polyacrylate
PAEK	Poly(aryl ether ketone)
PAl	Polyalkane
PAn	Polyaniline
PAS	Poly(aryl sulfone)
PAT	Polyalkylthiophene
PB	1,4-Polybutadiene
PBA	Poly(butyl acrylate)
PBMA	Poly(butyl methylacrylate)
PBrS	Polybromostyrene
PBTP	Polybithiophene
PC	Polycarbonate
PCHA	Poly(cyclohexyl acrylate)
PCL	Poly(ε-caprolactone)
PCP	Polychloroprene
PDMS	Polydimethylsiloxane
PDPO	Poly(dimethylphenylene oxide)
PE	Polyethylene
PEA	Poly(ethyl acrylate)
PEB	Polyethylbutene
PECH	Polyepichlorohydrin
PEE	Poly(ethylethylene)
PEEK	Poly(ethyl ether ketone)
PEG	Poly(ethylene glycol)
PEI	Poly(ether imide)
PEK	Poly(ethyl ketone)
PEO	Poly(ethylene oxide)
PEP	Polyethylpropylene
PEs	Polyester
PET	Poly(ethylene terephthalate)
PI	Polyisoprene
PIB	Polyisobutene
PMB	Polymethylbutene
PMMA	Poly(methyl methacrylate)
P*n*BMA	Poly(*n*-butyl methacrylate)
PP	Polypropylene
PPh	Polyphenylene
PPl	Polyphenol
PPMS	Polyphenylmethylsiloxane
P*p*MS	Poly(*p*-methylstyrene)
PPO	Poly(phenyl oxide)
PPV	Polyphenylenevinylene
PPy	Polypyrrole
PS	Polystyrene
PSO	Polysiloxane
PSVP	Poly(styrene–vinylpyridine)
PTFE	Polytetrafluoroethylene
PTHF	Poly(tetrahydrofuran)
PU	Polyurethane
PVA	Poly(vinyl acetate)
PVC	Poly(vinyl chloride)
PVE	Polyvinylethylene
PVME	Poly(vinyl methyl ether)
PVP	Polyvinylpyridine
PVPh	Polyvinylphenol
QENS	Quasielastic Neutron Scattering
RPA	Random Phase Approximation
SAN	Poly(styrene–acrylonitrile)
SANS	Small-angle Neutron Scattering
SAXS	Small-angle X-ray Scattering
SBR	Styrene–Butadiene Rubber
SBS	Styrene–Butadiene–Styrene Copolymer
SCPE	Solution-chlorinated Polyethylene
SPS	Sulfonated Polystyrene
TOF	Time-of-flight
WANS	Wide-angle Neutron Scattering

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Light Scattering, Classical: Size and Size Distribution Characterization

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Near-infrared Spectroscopy of Polymers and Rubbers • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers • X-ray Scattering in Analysis of Polymers

Surfaces **(Volume 10)**
Ellipsometry in Analysis of Surfaces and Thin Films

Nuclear Methods **(Volume 14)**
Rutherford Backscattering Spectroscopy

FURTHER READING

J.F. Ankner, C.F. Majkrzak, S.K. Satija, 'Neutron Reflectivity and Grazing Angle Diffraction', *J. Res. Nat. Inst. Stand. Technol.*, **98**, 47–58 (1993).

V. Arrighi, M.L. Fernandez, J.S. Higgins, 'Looking at Polymers with Neutrons', *Sci. Prog.*, **77**, 71–90 (1993/94).

D.G. Bucknall, V. Arrighi, 'Neutron Scattering and Polymer Blends', in *Polymer Blends*, eds. D.R. Paul, C.B. Bucknall, John Wiley and Sons, New York, 349–378, Chapter 12, 1999.

D.G. Bucknall, J.S. Higgins, 'Neutron Reflection Studies of Polymer–Polymer Interfaces', in *Polymers and Surfaces – A Versatile Combination*, ed. H. Hommell, Research Signpost, Trivandrum, 161–199, 1998.

B. Ewen, D. Richter, 'Neutron Spin-echo Investigations on the Segmental Dynamics of Polymers in Melts, Networks and Solutions', *Adv. Polym. Sci.*, **137**, 1–129 (1997).

B. Gabrys, 'Quasielastic and Inelastic Neutron Scattering from Polymers', in *Scattering Methods in Polymer Science*, ed. R.W. Richards, Ellis Horwood, Chichester, 153–192, 1995.

B. Hammouda, 'Small-angle Neutron Scattering from Homogenous Polymer Mixtures', *Adv. Polym. Sci.*, **106**, 87–133 (1997).

H.J.M. Hanley, 'Structure of Macromolecular Systems Under Shear: Applications of Neutron Scattering', *Curr. Opin. Colloid Interface Sci.*, **2**, 635–640 (1997).

J.S. Higgins, H.C. Benoit, *Polymers and Neutron Scattering*, Oxford University Press, Oxford, 1994.

J.S. Higgins, 'Contribution of Neutron Scattering Experiments to the Understanding of Conformation and Dynamics of Polymer Molecules', *Inst. Phys. Conf. Ser.*, **64**, 311–315 (1983).

J.S. Higgins, 'Dynamics of Polymer Molecules Using Neutron Spin-Echo', *Br. Polym. J.*, **19**, 103–109 (1987).

J.S. Higgins, 'Formation of Microstructure in Polymer Blends', *Philos. Trans. R. Soc., London, Ser. A*, **348**, 131–147 (1994).

J.S. Higgins, 'Polymer Conformation and Dynamics', *Treatise Mater. Sci. Technol.*, **15**, 381–422 (1979).

T. Lodge, 'Characterisation of Polymer Materials by Scattering Techniques, with Applications to Block Copolymers', *Mikrochim. Acta*, **116**, 1–31 (1994).

D.J. Lohse, 'The Application of Small Angle Scattering to Elastomers', *Rubber Chem. Technol.*, **67**, 367–383 (1994).

A. Maconnachie, R.W. Richards, 'Neutron Scattering and Amorphous Polymers', *Polymer*, **19**, 739–762 (1978).

K. Mortensen, 'Small-angle X-ray and Neutron Scattering Studies from Multiphase Systems', *Curr. Opin. Solid State Mater. Sci.*, **2**, 653–660 (1997).

R.W. Richards, 'Small-angle Neutron Scattering from Block Copolymers', *Adv. Polym. Sci.*, **71**, 1–39 (1985).

R.W. Richards, 'Small-angle Neutron Scattering from Interpenetrating Networks', *Makromol. Chem. Macromol. Symp.*, **40**, 209–222 (1990).

T.P. Russell, 'On the Reflectivity of Polymers: Neutrons and X-rays', *Physica B*, **221**, 267–283 (1996).

T.P. Russell, 'X-ray and Neutron Reflectivity for the Investigation of Polymers', *Mater. Sci. Rep.*, **5**, 171–271 (1990).

D. Richter, 'Neutron Spin-echo Investigations on Molecular Motion in Polymers', *Physica B*, **180**, 7–14 (1992).

T.P. Russell, 'The Characterisation of Polymer Interfaces', *Annu. Rev. Mater. Sci.*, **21**, 249–268 (1991).

L.H. Sperling, J.N. Yoo, S.I. Yang, A. Klein, 'Small-angle Neutron Scattering – Recent Applications to Multicomponent Systems', *Adv. Chem. Ser.*, **227**, 455–470 (1990).

T. Springer, 'Recent Contributions of Neutron Scattering for Condensed Matter Research – Polymers', *Phys. Scr.*, **T49A**, 11–16 (1993).

M. Stamm, 'Investigation of the Interface Between Polymers – A Comparison of Scattering and Reflectivity Techniques', *J. Appl. Crystallogr.*, **24**, 651–658 (1991).

M. Stamm, D.W. Schubert, 'Interfaces Between Incompatible Polymers', *Annu. Rev. Mater. Sci.*, **25**, 325–356 (1995).

R.K. Thomas, 'Neutron Reflectometry in Solid-state and Materials Science', *Curr. Opin. Solid State Mater. Sci.*, **1**, 636–644 (1996).

REFERENCES

1. B. Gabrys, O. Scharpf, 'Scattering From Polymers Using Polarized Neutrons – A New Development', *Physica B*, **180**(A), 495–498 (1992).
2. B. Gabrys, O. Scharpf, D.G. Peiffer, 'Short-range Order in Noncrystalline Polymers Examined with Spin-polarized Neutrons', *J. Polym. Sci. B, Polym. Phys.*, **31**(13), 1891–1895 (1993).
3. B. Gabrys, O. Scharpf, D.G. Peiffer, 'Polymers Studied with Spin-polarized Neutrons', *J. Macromol. Sci. Pure Appl. Chem.*, **A31**(S6–7), 1069–1076 (1994).
4. A. Arbe, D. Richter, J. Colmenero, B. Farago, 'Merging of the Alpha-relaxation and Beta-relaxation in Polybutadiene – A Neutron Spin-echo and Dielectric Study', *Phys. Rev. E*, **54**(4B), 3853–3869 (1996).
5. B. Ewen, D. Richter, 'The Dynamics of Polymer Melts as Seen by Neutron Spin-echo Spectroscopy', *Macromol. Symp.*, **90**, 131–149 (1995).
6. B. Ewen, D. Richter, 'Neutron Spin-echo Investigations on the Segmental Dynamics of Polymers in Melts, Networks and Solutions', *Adv. Polym. Sci.*, **134**, 1–129 (1997).
7. V.F. Sears, *Neutron Optics*, Oxford University Press, Oxford, 1989.
8. V. Arrighi, M.L. Fernandez, J.S. Higgins, 'Looking at Polymers with Neutrons', *Sci. Prog.*, **77**(1–2), 71–90 (1993).
9. http://ne43.ne.uiuc.edu/n-scatter/n-lengths.
10. V.F. Sears, 'Neutron Scattering Lengths and Cross-sections', *Neutron News*, **3**(3), 26–37 (1992).
11. J.S. Higgins, B.C. Benoit, *Polymers and Neutron Scattering*, Clarendon Press, Oxford, 1993.
12. B. Rosi-Schwartz, G.R. Mitchell, 'A Detailed Single-chain Model for Molten Poly(tetrafluoroethylene) from a Novel Structure Refinement Technique on Neutron-scattering Data', *Polymer*, **35**(15), 3139–3148 (1994).
13. R. Kirste, W.A. Kruse, K. Ibel, 'Determination of the Conformation of Polymers in the Amorphous Solid State and in Concentrated Solution by Neutron Diffraction', *Polymer*, **16**(2), 120–124 (1975).
14. P. Debye, 'Molecular-weight Determination by Light Scattering', *J. Phys. Colloid Chem.*, **51**, 18–32 (1947).
15. H. Benoit, J.F. Joanny, G. Hadziioannou, B. Hammouda, 'Scattering by Linear, Branched, and Copolymer

Chain Molecules for Large Scattering Vectors', *Macromolecules*, **26**(21), 5790–5795 (1993).
16. D. Richter, B. Farago, J.S. Huang, L.J. Fetters, B. Ewen, 'A Study of Single-arm Relaxation in a Polystyrene Star Polymer by Neutron Spin-echo Spectroscopy', *Macromolecules*, **22**(1), 468–472 (1989).
17. P.G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY, 1985.
18. G. Porod, 'Die Rontgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. I. Teil', *Kolloid Z.*, **124**(2), 83–114 (1951).
19. R.K. Thomas, 'Neutron Reflection from Polymer-bearing Interfaces', in *Scattering Methods in Polymer Science*, ed. R.W. Richards, Ellis Horwood, Chichester, 39–102, 1995.
20. D.G. Bucknall, J.S. Higgins, 'Neutron Reflection Studies of Polymer–Polymer Interfaces', in *Polymers and Surfaces – A Versatile Combination*, ed. H. Hommell, Research Signpost, Trivandrum, 1998.
21. O.S. Heavens, *Optical Properties of Thin-films*, Dover Publications, New York, 1991.
22. J. Lekner, *Theory of Reflection*, Martinus Nijhoff, Dordrecht, 1987.
23. J. Lekner, 'Reflection Theory and the Analysis of Neutron Reflection Data', *Physica B*, **173**(1–2), 99–111 (1991).
24. W. Marshall, S.W. Lovesey, *Theory of Thermal Neutron Scattering*, Clarendon Press, Oxford, 1971.
25. A. Maconnachie, R.W. Richards, 'Neutron Scattering and Amorphous Polymers', *Polymer*, **19**, 739–762 (1978).
26. G. Allen, J.S. Higgins, 'Physicochemical Aspects of Neutron Studies of Molecular Motion', *Rep. Prog. Phys.*, **36**, 1073–1133 (1973).
27. P.G. de Gennes, 'Quasielastic Scattering of Neutrons by Dilute Polymer Solutions: I. Free-draining Limit', *Physics*, **3**(1), 37–45 (1967).
28. E. Dubois-Violette, P.G. de Gennes, 'Quasielastic Scattering by Dilute, Ideal, Polymer Solutions: II. Effects of Hydrodynamic Interactions', *Physics*, **3**(4), 181–198 (1967).
29. P.E. Rouse, 'A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers', *J. Chem. Phys.*, **21**(7), 1272–1280 (1953).
30. B.H. Zimm, 'Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss', *J. Chem. Phys.*, **24**(2), 269–278 (1956).
31. A.K. Akcasu, M. Benmouna, C.C. Han, 'Interpretation of Dynamic Scattering from Polymer Solutions', *Polymer*, **21**, 866–890 (1980).
32. F. Mezei, 'Neutron Spin-echo: A New Concept in Polarized Thermal Neutron Techniques', *Z. Phys.*, **255**, 146–160 (1972).
33. L.K. Nicholson, 'The Neutron Spin-echo Spectrometer – a New High-resolution Technique in Neutron Scattering', *Contemp. Phys.*, **22**(4), 451–475 (1981).
34. F.S. Bates, G.D. Wignall, W.C. Koehler, 'Critical Behavior of Binary-liquid Mixtures of Deuterated and Protonated Polymers', *Phys. Rev. Lett.*, **55**(22), 2425–2428 (1985).
35. A. Lapp, C. Picot, H. Benoit, 'Determination of the Flory Interaction Parameters in Miscible Polymer Blends by Measurement of the Apparent Radius of Gyration', *Macromolecules*, **18**(12), 2437–2441 (1985).
36. R.A.L. Jones, E.J. Kramer, M.H. Rafailovich, J. Sokolov, S.A. Schwarz, 'Surface Enrichment in an Isotopic Polymer Blend', *Phys. Rev. Lett.*, **62**(3), 280–283 (1989).
37. R.A.L. Jones, L.J. Norton, E.J. Kramer, F.S. Bates, P. Wiltzius, 'Surface-directed Spinodal Decomposition', *Phys. Rev. Lett.*, **66**(10), 1326–1329 (1991).
38. F.S. Bates, G.D. Wignall, 'Isotope-induced Quantum-phase Transitions in the Liquid State', *Phys. Rev. Lett.*, **57**(12), 1429–1432 (1986).
39. C.C. Han, B.J. Bauer, J.C. Clark, Y. Muroga, Y. Matsushita, M. Okada, T.C. Qui, T.H. Chang, I.C. Sanchez, 'Temperature, Composition and Molecular-weight Dependence of the Binary Interaction Parameter of Polystyrene–Poly(Vinyl Methyl Ether) Blends', *Polymer*, **29**(11), 2002–2014 (1988).
40. J.D. Londono, G.D. Wignall, 'The Flory–Huggins Interaction Parameter in Blends of Polystyrene and Poly(*p*-methylstyrene) by Small-angle Neutron Scattering', *Macromolecules*, **30**(13), 3821–3824 (1997).
41. K. Hahn, B.J. Schmitt, M. Kirschey, R.G. Kirste, H. Salie, S. Schmittstrecker, 'Structure and Thermodynamics in Polymer Blends – Neutron Scattering Measurements on Blends of Poly(methyl methacrylate) and Poly(styrene-*co*-acrylonitrile)', *Polymer*, **33**(24), 5150–5166 (1992).
42. G. Beaucage, R.S. Stein, 'Tacticity Effects on Polymer Blend Miscibility. 3. Neutron-scattering Analysis', *Macromolecules*, **26**(7), 1617–1626 (1993).
43. F.S. Bates, M. Muthukumar, G.D. Wignall, L.J. Fetters, 'Thermodynamics of Isotopic Polymer Mixtures – Significance of Local Structural Symmetry', *J. Chem. Phys.*, **89**(1), 535–544 (1988).
44. J.D. Londono, A.H. Narten, G.D. Wignall, K.G. Honnell, E.T. Hsieh, T.W. Johnson, F.S. Bates, 'Composition Dependence of the Interaction Parameter in Isotopic Polymer Blends', *Macromolecules*, **27**(10), 2864–2871 (1994).
45. I. Hopkinson, F.T. Kiff, R.W. Richards, S.M. King, T. Farren, 'Isotopic Labeling and Composition Dependence of Interaction Parameters in Polyethylene Oxide–Polymethyl Methacrylate Blends', *Polymer*, **36**(18), 3523–3531 (1995).
46. M. Muthukumar, 'Thermodynamics of Polymer Solutions', *J. Chem. Phys.*, **85**(8), 4722–4728 (1986).
47. S.K. Kumar, 'Chemical Potentials of Polymer Blends From Monte-Carlo Simulations – Consequences on SANS-determined Chi-parameters', *Macromolecules*, **27**(1), 260–271 (1994).

48. S.T. Cui, H.D. Cochran, P.T. Cummings, S.K. Kumar, 'Computer Simulations of the Static Scattering from Model Polymer Blends', *Macromolecules*, **30**(11), 3375–3382 (1997).
49. B. Hammouda, M. Benmouna, 'Neutron Scattering from Polymer Blends Under Pressure', *J. Polym. Sci. B, Polym. Phys.*, **33**(17), 2359–2364 (1995).
50. S. Janssen, D. Schwahn, K. Mortensen, T. Springer, 'Pressure-dependence of the Flory–Huggins Interaction Parameter in Binary Polymer Blends Investigated by SANS', *J. Phys. IV*, **3**(C8), 17–20 (1993).
51. S. Janssen, D. Schwahn, K. Mortensen, T. Springer, 'Pressure-dependence of the Flory–Huggins Interaction Parameter in Polymer Blends – a SANS Study and a Comparison to the Flory–Orwoll–Vrij Equation of State', *Macromolecules*, **26**(21), 5587–5591 (1993).
52. S. Janssen, D. Schwahn, T. Springer, K. Mortensen, 'Coil and Melt Compressibility of Polymer Blends Studied by SANS and PVT Experiments', *Macromolecules*, **28**(7), 2555–2560 (1995).
53. S. Janssen, D. Schwahn, T. Springer, K. Mortensen, H. Hasegawa, 'investigation of the Pressure-dependence of the Gibbs Potential for Polymer Blends by Means of SANS', *Physica B*, **213**, 691–693 (1995).
54. D.M. Sadler, S.J. Spells, A. Keller, J.M. Guenet, 'Wide-angle Neutron Scattering from Isotactic Polystyrene – the Fold Arrangement in Solution Grown Crystals', *Polym. Commun.*, **25**(10), 290–293 (1984).
55. B. Rosi-Schwartz, G.R. Mitchell, 'A Complete Atomistic Model of Molten Polyethylene from Neutron Scattering Data – a New Methodology for Polymer Structure', *Polymer*, **35**(25), 5398–5407 (1994).
56. J.D. Londono, B.K. Annis, A. Habenschuss, O. Borodin, G.D. Smith, J.Z. Turner, A.K. Soper, 'Cation Environment in Molten Lithium Iodide Doped Poly(ethylene oxide)', *Macromolecules*, **30**(23), 7151–7157 (1997).
57. R.G. Hill, P.E. Tomlins, J.S. Higgins, 'Preliminary Study of the Kinetics of Phase Separation in High Molecular-weight Poly(methyl methacrylate) Solution-chlorinated Polyethylene Blends', *Macromolecules*, **18**(12), 2555–2560 (1985).
58. J.S. Higgins, 'Polymer Blends', *Makromol. Chem. Macromol. Symp.*, **15**, 201–213 (1988).
59. J.S. Higgins, H.A. Fruitwala, P.E. Tomlins, 'Experimental Evidence for Slow Theory of Mutual Diffusion Coefficients in Phase Separating Polymer Blends', *Br. Polym. J.*, **21**(3), 247–257 (1989).
60. D. Schwahn, S. Janssen, T. Springer, 'Early State of Spinodal Decomposition Studied with Small-angle Neutron Scattering in the Blend Deuteropolystyrene and Polyvinylmethylether – a Comparison with the Cahn–Hilliard–Cook Theory', *J. Chem. Phys.*, **97**(11), 8775–8788 (1992).
61. D. Schwahn, H. Yee-Madeira, 'Spinodal Decomposition of the Polymer Blend Deutereous Polystyrene (d-PS) and Polyvinylmethylether (PVME) Studied with High-resolution Neutron Small-angle Scattering', *Colloid Polym. Sci.*, **265**(10), 867–875 (1987).
62. C. Schipp, M.J. Hill, P.J. Barham, V.M. Cloke, J.S. Higgins, L. Oiarzabal, 'Ambiguities in the Interpretation of Small-angle Neutron Scattering from Blends of Linear and Branched Polyethylene', *Polymer*, **37**(12), 2291–2297 (1996).
63. R.G. Alamo, J.D. Londono, L. Mandelkern, F.C. Stehling, G.D. Wignall, 'Phase Behavior of Blends of Linear and Branched Polyethylenes in the Molten and Solid States by Small-angle Neutron Scattering', *Macromolecules*, **27**(2), 411–417 (1994).
64. R.G. Alamo, W.W. Graessley, R. Krishnamoorti, D.J. Lohse, J.D. Londono, L. Mandelkern, F.C. Stehling, G.D. Wignall, 'Small-angle Neutron Scattering Investigations of Melt Miscibility and Phase Segregation in Blends of Linear and Branched Polyethylenes as a Function of the Branch Content', *Macromolecules*, **30**(3), 561–566 (1997).
65. F.S. Bates, R.E. Cohen, C.V. Berney, 'Small-angle Neutron Scattering Determination of Macro-lattice Structure in a Polystyrene Polybutadiene Diblock Co-polymer', *Macromolecules*, **15**(2), 589–592 (1982).
66. C.V. Berney, R.E. Cohen, F.S. Bates, 'Sphere Sizes in Diblock Copolymers – Discrepancy Between Electron Microscopy and Small-angle Scattering Results', *Polymer*, **23**(8), 1222–1226 (1982).
67. R.W. Richards, J.L. Thomason, 'A Small-angle Neutron Scattering Investigation of Block Co-polymers of Styrene and Isoprene in the Solid State', *Polymer*, **22**(5), 581–589 (1981).
68. R.W. Richards, J.L. Thomason, 'Small-angle Neutron Scattering Study of Block Co-polymer Morphology', *Macromolecules*, **16**(6), 982–992 (1983).
69. G. Hadziioannou, L.H. Wang, R.S. Stein, R.S. Porter, 'Small-angle Neutron Scattering Studies on Amorphous Polystyrene Oriented by Solid State Coextrusion', *Macromolecules*, **15**(3), 880–882 (1982).
70. N. Torikai, N. Takabayashi, I. Noda, S. Koizumi, Y. Morii, Y. Matsushita, 'Lamellar Domain Spacings of Diblock Copolymer/Homopolymer Blends and Conformations of Block Chains in Their Microdomains', *Macromolecules*, **30**(19), 5698–5703 (1997).
71. W.W. Maurer, F.S. Bates, T.P. Lodge, K. Almdal, K. Mortensen, G.H. Fredrickson, 'Can a Single Function for Chi Account for Block Copolymer and Homopolymer Blend Phase Behavior?', *J. Chem. Phys.*, **108**(7), 2989–3000 (1998).
72. F. Boue, M. Nierlich, G. Jannink, R. Ball, 'Polymer Coil Relaxation in Uniaxially Strained Polystyrene Observed by Small-angle Neutron Scattering', *J. Phys. (Paris)*, **43**(1), 137–148 (1982).
73. J. Bastide, L. Leibler, J. Prost, 'Scattering by Deformed Swollen Gels – Butterfly Isointensity Patterns', *Macromolecules*, **23**(6), 1821–1825 (1990).

74. F. Boue, J. Bastide, M. Buzier, A. Lapp, J. Herz, T.A. Vilgis, 'Strain-induced Large Fluctuations During Stress-relaxation in Polymer Melts Observed by Small-angle Neutron Scattering – Lozenges, Butterflies, and Related Theory', *Colloid Polym. Sci.*, **269**(3), 195–216 (1991).
75. A. Ramzi, F. Zielinski, J. Bastide, F. Boue, 'Butterfly Patterns – Small-angle Neutron Scattering from Deuterated Mobile Chains in a Randomly Cross-linked Polystyrene Network', *Macromolecules*, **28**(10), 3570–3587 (1995).
76. A. Ramzi, A. Hakiki, J. Bastide, F. Boue, 'Uniaxial Extension of End-linked Polystyrene Networks Containing Deuterated Free Chains Studied by Small-angle Neutron Scattering: Effect of the Network Chains and the Size of the Free Chains', *Macromolecules*, **30**(10), 2963–2977 (1997).
77. P. Lindner, 'Steady State SANS Experiments Under External Constraints', in *Neutron, X-Ray and Light Scattering*, eds. P. Lindner, Th. Zemb, Elsevier, Amsterdam, 261–278, 1991.
78. F. Boue, P. Lindner, 'Semidilute Polymer Solutions under Shear', *Europhys. Lett.*, **25**(6), 421–427 (1994).
79. G.A. Mcconnell, M.Y. Lin, A.P. Gast, 'Long-range Order in Polymeric Micelles under Steady Shear', *Macromolecules*, **28**,(20), 6754–6764 (1995).
80. I.W. Hamley, J.A. Pople, J.P.A. Fairclough, N.J. Terrill, A.J. Ryan, C. Booth, G.E. Yu, O. Diat, K. Almdal, K. Mortensen, M. Vigild, 'Effect of Shear on Cubic Phases in Gels of a Diblock Copolymer', *J. Chem. Phys.*, **108**(16), 6929–6936 (1998).
81. S.M. King, R.K. Heenan, V.M. Cloke, C. Washington, 'Neutron Scattering from a Poly(oxyethylene)–Poly (oxypropylene)–Poly(oxyethylene) Copolymer in Dilute Aqueous Solution Under Shear Flow', *Macromolecules*, **30**(20), 6215–6222 (1997).
82. V.M. Cloke, J.S. Higgins, C.L. Phoon, S.M. Richardson, S.M. King, R. Done, T.E. Cooper, 'Poiseuille Geometry Shear-flow Apparatus for Small-angle Scattering Experiments', *Rev. Sci. Instrum.*, **67**(9), 3158–3163 (1996).
83. C.L. Phoon, J.S. Higgins, G. Allegra, P. Vanleeuwen, E. Staples, 'Shear-induced Crystallization and Melting of a Micellar Solution', *Proc. R. Soc. London, Ser. A*, **442**, 221–230 (1993).
84. V. Arrighi, J.S. Higgins, R.A. Weiss, A.L. Cimecioglu, 'A Small-angle Neutron Scattering Study of a Semiflexible Main-chain Liquid-crystalline Copolyester', *Macromolecules*, **25**(20), 5297–5305 (1992).
85. L. Abis, V. Arrighi, A.L. Cimecioglu, J.S. Higgins, R.A. Weiss, 'Phase Behavior and Orientational Order of a Main-chain Nematic Polyester – A Combined SANS and NMR Study', *Eur. Polym. J.*, **29**(2–3), 175–181 (1993).
86. E. Mendes, R. Oeser, C. Hayes, F. Boue, J. Bastide, 'Small-angle Neutron Scattering Study of Swollen Elongated Gels – Butterfly Patterns', *Macromolecules*, **29**(17), 5574–5584 (1996).
87. J. Bastide, E. Mendes, F. Boue, M. Buzier, P. Lindner, 'Enhancement of Inhomogeneities in Gels Upon Swelling and Stretching', *Makromol. Chem. Macromol. Symp.*, **40**, 81–99 (1990).
88. E. Mendes, P. Lindner, M. Buzier, F. Boue, J. Bastide, 'Experimental Evidence for Inhomogeneous Swelling and Deformation in Statistical Gels', *Phys. Rev. Lett.*, **66**(12), 1595–1598 (1991).
89. A. Onuki, 'Scattering from Deformed Swollen Gels with Heterogeneities', *J. Phys. II*, **2**(1), 45–61 (1992).
90. Y. Rabin, R. Bruinsma, 'Strain Effects on Thermal Fluctuations in Soft Solids – Butterfly Patterns', *Europhys. Lett.*, **20**(1), 79–85 (1992).
91. T.C.B. McLeish, J. Allgaier, D.K. Bick, G. Bishko, P. Biswas, R. Blackwell, B. Blottiere, N. Clarke, B. Gibbs, D.J. Groves, A. Hakiki, R.K. Heenan, J.M. Johnson, R. Kant, D.J. Read, R.N. Young, 'Dynamics of Entangled H-polymers: Theory, Rheology and Neutron Scattering', *Macromolecules*, **32**(20), 6734–6758 (1999).
92. E. Helfand, A.M. Sapse, 'Theory of Unsymmetric Polymer–Polymer Interfaces', *J. Chem. Phys.*, **62**(4), 1327–1331 (1975).
93. T.P. Russell, R.P. Hjelm, P.A. Seeger, 'Temperature-dependence of the Interaction Parameter of Polystyrene and Poly(methyl methacrylate)', *Macromolecules*, **23**(3), 890–893 (1990).
94. M.L. Fernandez, J.S. Higgins, J. Penfold, R.C. Ward, C. Shackleton, D.J. Walsh, 'Neutron Reflection Investigation of the Interface Between an Immiscible Polymer Pair', *Polymer*, **29**(11), 1923–1928 (1988).
95. S.H. Anastasiadis, T.P. Russell, S.K. Satija, C.F. Majkrzak, 'The Morphology of Symmetric Diblock Copolymers as Revealed by Neutron Reflectivity', *J. Chem. Phys.*, **92**(9), 5677–5691 (1990).
96. W. Zhao, X. Zhao, M.H. Rafailovich, J. Sokolov, T. Mansfield, R.S. Stein, R.C. Composto, E.J. Kramer, R.A.L. Jones, M. Sansone, M. Nelson, 'Neutron and X-ray Reflectivity Measurements of Polystyrene–Polybromostyrene (PS–PBrS) Interfaces', *Physica B*, **173**(1–2), 43–46 (1991).
97. B. Guckenbiehl, M. Stamm, T. Springer, 'Interface Properties of Blends of Incompatible Polymers', *Physica B*, **198**(1–3), 127–130 (1994).
98. D.W. Schubert, V. Abetz, M. Stamm, T. Hack, W. Siol, 'Composition and Temperature-dependence of the Segmental Interaction Parameter in Statistical Copolymer Homopolymer Blends', *Macromolecules*, **28**(7), 2519–2525 (1995).
99. M. Sferrazza, C. Xiao, R.A.L. Jones, D.G. Bucknall, J. Webster, J. Penfold, 'Evidence for Capillary Waves at Immiscible Polymer/Polymer Interfaces', *Phys. Rev. Lett.*, **78**(19), 3693–3696 (1997).
100. K. Binder, 'Collective Diffusion, Nucleation, and Spinodal Decomposition in Polymer Mixtures', *J. Chem. Phys.*, **79**(12), 6387–6409 (1983).

101. K.R. Shull, A.M. Mayes, T.P. Russell, 'Segment Distributions in Lamellar Diblock Copolymers', *Macromolecules*, **26**(15), 3929–3936 (1993).
102. M. Muller, K. Binder, 'Computer Simulation of Asymmetric Polymer Mixtures', *Macromolecules*, **28**(6), 1825–1834 (1995).
103. D.F. Siqueira, D.W. Schubert, V. Erb, M. Stamm, J.P. Amato, 'Interface Thickness of the Incompatible Polymer System PS/PnBMA as Measured by Neutron Reflectometry and Ellipsometry', *Colloid Polym. Sci.*, **273**(11), 1041–1048 (1995).
104. H.E. Hermes, D.G. Bucknall, S.A. Butler, J.S. Higgins, 'Neutron Reflectivity Studies of Hot Polymer Melt Interfaces', *Macromol. Symp.*, **126**, 331–342 (1997).
105. H.E. Hermes, J.S. Higgins, D.G. Bucknall, 'Investigation of the Melt Interface Between Polyethylene and Polystyrene Using Neutron Reflectivity', *Polymer*, **38**(4), 985–989 (1997).
106. S. Butler, H. Hermes, D. Bucknall, J.S. Higgins, 'Polymer Interfaces Studied by Neutron Reflection.', *Abstr. Pap. Am. Chem. Soc.*, **213**(2), 476-POLY, (1997).
107. G.J. Fleer, M.A. Cohen Stuart, J.M.H.M. Scheutjens, T. Cosgrove, B. Vincent, *Polymers at Interfaces*, Chapman and Hall, London, 1993.
108. S.K. Satija, P.D. Gallagher, A. Karim, L.J. Fetters, 'Neutron Reflectivity Study of a Chemically End-grafted Polystyrene Brush in a Binary Solvent Mixture', *Physica B*, **248**, 204–207 (1998).
109. C.J. Clarke, 'The Kinetics of Polymer Brush Penetration into a High-molecular-weight Matrix', *Polymer*, **37**(21), 4747–4752 (1996).
110. C.J. Clarke, R.A.L. Jones, J.L. Edwards, A.S. Clough, J. Penfold, 'Kinetics of Formation of Physically End-adsorbed Polystyrene Layers from the Melt', *Polymer*, **35**(19), 4065–4071 (1994).
111. A. Karim, A. Mansour, G.P. Felcher, T.P. Russell, 'Short-time Relaxation at Polymeric Interfaces', *Phys. Rev. B*, **42**(10), 6846–6849 (1990).
112. G.P. Felcher, A. Karim, T.P. Russell, 'Interdiffusion at the Interface of Polymeric Bilayers – Evidence for Reptation', *J. Non-cryst. Solids*, **131**, 703–708 (1991).
113. M. Stamm, S. Huttenbach, G. Reiter, T. Springer, 'Initial Stages of Polymer Interdiffusion Studied by Neutron Reflectometry', *Europhys. Lett.*, **14**(5), 451–456 (1991).
114. G. Reiter, U. Steiner, 'Measurements of Polymer Diffusion Over Small Distances – a Check of Reptation Arguments', *J. Phys. II*, **1**(6), 659–671 (1991).
115. G. Agrawal, R.P. Wool, W.D. Dozier, G.P. Felcher, T.P. Russell, J.W. Mays, 'Short-time Interdiffusion at Polymer Interfaces', *Macromolecules*, **27**(15), 4407–4409 (1994).
116. M.L. Fernandez, J.S. Higgins, J. Penfold, C. Shackleton, 'Interfacial Structure for a Compatible Polymer Blend for Short and Intermediate Annealing Times', *J. Chem. Soc., Faraday Trans.*, **87**(13), 2055–2061 (1991).
117. M.L. Fernandez, J.S. Higgins, J. Penfold, C. Shackleton, D.J. Walsh, 'Evidence for a Sharp Interface Between Partially Miscible Polymers from a Study by Neutron Specular Reflection', *Polymer*, **31**(11), 2146–2151 (1990).
118. B.B. Sauer, D.J. Walsh, 'Use of Neutron Reflection and Spectroscopic Ellipsometry for the Study of the Interface Between Miscible Polymer Films', *Macromolecules*, **24**(22), 5948–5955 (1991).
119. K. Kunz, M. Stamm, 'Interface Formation Between Compatible Polymers as Determined by Neutron Reflectometry – PMMA and PVC', *Macromol. Symp.*, **78**, 105–114 (1994).
120. K. Kunz, M. Stamm, 'Initial Stages of Interdiffusion of PMMA Across an Interface', *Macromolecules*, **29**(7), 2548–2554 (1996).
121. R. Schnell, M. Stamm, C. Creton, 'Direct Correlation Between Interfacial Width and Adhesion in Glassy Polymers', *Macromolecules*, **31**(7), 2284–2292 (1998).
122. E. Sivaniah, M. Sferrazza, R.A.L. Jones, D.G. Bucknall, 'Chain Confinement Effects on Interdiffusion in Polymer Multilayers', *Phys. Rev. E*, **59**(1), 885–888 (1999).
123. T. Kuhlmann, J. Kraus, P. Mullerbuschbaum, D.W. Schubert, M. Stamm, 'Effects of Confined Geometry and Substrate Interaction on the Initial Stages of Interdiffusion in Polystyrene Films', *J. Non-cryst. Solids*, **235**, 457–463 (1998).
124. K.D. Kim, L.H. Sperling, A. Klein, G.D. Wignall, 'Characterization of Film Formation from Direct Mini-emulsified Polystyrene Latex Particles via SANS', *Macromolecules*, **26**(17), 4624–4631 (1993).
125. T.P. Russell, S.H. Anastasiadis, A. Menelle, G.P. Felcher, S.K. Satija, 'Segment Density Distribution of Symmetrical Diblock Copolymers at the Interface Between 2 Homopolymers as Revealed by Neutron Reflectivity', *Macromolecules*, **24**(7), 1575–1582 (1991).
126. T.P. Russell, A. Menelle, W.A. Hamilton, G.S. Smith, S.K. Satija, C.F. Majkrzak, 'Width of Homopolymer Interfaces in the Presence of Symmetrical Diblock Copolymers', *Macromolecules*, **24**(20), 5721–5726 (1991).
127. W. Ruland, 'Small-angle Scattering of Two-phase Systems: Determination and Significance of Systematic Deviations from Porod's Law', *J. Appl. Crystallogr.*, **4**, 70–73 (1971).
128. L. Auvray, P. Auroy, 'Scattering by Interfaces: Variations on Porod's Law', in *Neutron, X-ray and Light Scattering*, eds. P. Lindner, Th. Zemb, Elsevier, Amsterdam, 199–221, 1991.
129. J.T. Koberstein, B. Morra, R.S. Stein, 'The Determination of Diffuse-boundary Thicknesses of Polymers by Small-angle X-ray Scattering', *J. Appl. Crystallogr.*, **13**(1), 34–45 (1980).
130. R.W. Richards, J.L. Thomason, 'Small-angle Neutron Scattering Measurement of Block Co-polymer Interphase Structure', *Polymer*, **24**(9), 1089–1096 (1983).

131. D.G. Bucknall, 'Segregation Behaviour of Diblock Copolymers in Immiscible Polymer Blends', PhD Thesis, Imperial College, London, 1991.
132. E. Helfand, 'Block Copolymer Theory. III. Statistical Mechanics of the Microdomain Structure', *Macromolecules*, **8**(4), 552–556 (1975).
133. E. Helfand, Z.R. Wasserman, 'Block Copolymer Theory. 4. Narrow Interphase Approximation', *Macromolecules*, **9**(6), 879–888 (1976).
134. E. Helfand, Z.R. Wasserman, 'Block Copolymer Theory. 5. Spherical Domains', *Macromolecules*, **11**(5), 960–966 (1978).
135. E. Helfand, Z.R. Wasserman, 'Block Copolymer Theory. 6. Cylindrical Domains', *Macromolecules*, **13**(4), 994–998 (1980).
136. T.P. Russell, G. Coulon, V.R. Deline, D.C. Miller, 'Characteristics of the Surface-induced Orientation for Symmetric Diblock Ps/Pmma Copolymers', *Macromolecules*, **22**(12), 4600–4606 (1989).
137. S.H. Anastasiadis, T.P. Russell, S.K. Satija, C.F. Majkrzak, 'Neutron Reflectivity Studies of the Surface-induced Ordering of Diblock Copolymer Films', *Phys. Rev. Lett.*, **62**(16), 1852–1855 (1989).
138. A. Menelle, T.P. Russell, S.H. Anastasiadis, S.K. Satija, C.F. Majkrzak, 'Ordering of Thin Diblock Copolymer Films', *Phys. Rev. Lett.*, **68**(1), 67–70 (1992).
139. A.M. Mayes, R.D. Johnson, T.P. Russell, S.D. Smith, S.K. Satija, C.F. Majkrzak, 'Distributions of Chain Ends and Junction Points in Ordered Block Copolymers', *Macromolecules*, **26**(5), 1047–1052 (1993).
140. T.P. Russell, P. Lambooy, G.J. Kellogg, A.M. Mayes, 'Diblock Copolymers Under Confinement', *Physica B*, **213**, 22–25 (1995).
141. G.J. Kellogg, D.G. Walton, A.M. Mayes, P. Lambooy, T.P. Russell, P.D. Gallagher, S.K. Satija, 'Observed Surface-energy Effects in Confined Diblock Copolymers', *Phys. Rev. Lett.*, **76**(14), 2503–2506 (1996).
142. P. Lambooy, T.P. Russell, G.J. Kellogg, A.M. Mayes, P.D. Gallagher, S.K. Satija, 'Observed Frustration in Confined Block Copolymers', *Phys. Rev. Lett.*, **72**(18), 2899–2902 (1994).
143. G. Coulon, B. Collin, D. Ausserre, D. Chatenay, T.P. Russell, 'Islands and Holes on the Free Surface of Thin Diblock Copolymer Films. 1. Characteristics of Formation and Growth', *J. Phys. (Paris)*, **51**(24), 2801–2811 (1990).
144. T.P. Russell, 'On the Reflectivity of Polymers – Neutrons and X-rays', *Physica B*, **221**(1–4), 267–283 (1996).
145. G. Fleischer, F. Fujara, 'NMR as a Generalised Incoherent Scattering Experiment', in *Solid State NMR I. Methods*, ed. B. Bluemich, Springer-Verlag, Heidelberg, 161–207, Vol. 30, 1994.
146. B. Gabrys, 'Quasi-elastic and Inelastic Neutron Scattering from Polymers', in *Scattering Methods in Polymer Science*, eds. R.W. Richards, Ellis Horwood, Chichester, 153–192, 1995.
147. G. Floudas, J.S. Higgins, 'Ester Methyl-group Rotation in Poly(methyl methacrylate) and in the Blend Solution Chlorinated Polyethylene/Poly(methyl methacrylate) – a Quasielastic Neutron Scattering Study', *Polymer*, **33**(19), 4121–4128 (1992).
148. V. Arrighi, J.S. Higgins, A.N. Burgess, W.S. Howells, 'Rotation of Methyl Side-groups in Polymers – a Fourier Transform Approach to Quasielastic Neutron Scattering. 1. Homopolymers', *Macromolecules*, **28**(8), 2745–2753 (1995).
149. B. Gabrys, J.S. Higgins, K.T. Ma, J.E. Roots, 'Rotational Motion of the Ester Methyl-group in Stereoregular Poly(methyl methacrylate) – a Neutron Scattering Study', *Macromolecules*, **17**(4), 560–566 (1984).
150. T. Kanaya, K. Kaji, K. Inoue, 'Local Motions of *cis*-1,4-Polybutadiene in the Melt – A Quasielastic Neutron-scattering Study', *Macromolecules*, **24**(8), 1826–1832 (1991).
151. B. Frick, D. Richter, W. Petry, U. Buchenau, 'Study of the Glass-transition Order Parameter in Amorphous Polybutadiene by Incoherent Neutron Scattering', *Z. Phys. B: Condens. Matter*, **70**(1), 73–79 (1988).
152. B. Gabrys, H.A. Dong, F. Nardi, D.G. Peiffer, J. Tomkinson, 'Dynamics of Sulfonated Ionomers in the Solid State – Inelastic Neutron-scattering Studies', *Macromolecules*, **26**(8), 2007–2011 (1993).
153. S.F. Parker, 'Inelastic Neutron Scattering Spectra of Polyethylene', *J. Chem. Soc., Faraday Trans.*, **92**(11), 1941–1946 (1996).
154. S.F. Parker, W.F. Maddams, M.E. Vickers, K.P.J. Williams, G.W. Downs, 'Order in Nascent Polyethylene', *Polymer*, **37**(13), 2755–2757 (1996).
155. C. Lamers, C. Schonfeld, S.M. Shapiro, J. Batoulis, R. Timmermann, J.W. Cable, D. Richter, 'Short-range Order Effects in Amorphous Polycondensates as Studied by Spin-polarized Diffuse Neutron Scattering and Simulation', *Colloid Polym. Sci.*, **272**(11), 1403–1419 (1994).
156. S. Lecommandoux, M.F. Achard, F. Hardouin, A. Brulet, J.P. Cotton, 'Are Nematic Side-on Polymers Totally Extended? A SANS Study', *Liq. Cryst.*, **22**(5), 549–555 (1997).
157. S. Lecommandoux, L. Noirez, M.F. Achard, F. Hardouin, 'Thermodynamic and Conformational Study of Side On/Side End Fixed Liquid-crystalline Copolyacrylates', *J. Phys. II*, **6**(9), 1231–1242 (1996).
158. S. Lecommandoux, L. Noirez, H. Richard, M.F. Achard, C. Strazielle, F. Hardouin, 'Effect of the Spacer and Aliphatic Tail Length on the Conformation of Side-on Fixed Liquid-crystal Polyacrylates – SANS Experiments', *J. Phys. II*, **6**(2), 225–234 (1996).
159. N. Leroux, P. Keller, M.F. Achard, L. Noirez, F. Hardouin, 'Small-angle Neutron Scattering Experiments on Side-on Fixed Liquid-crystal Polyacrylates', *J. Phys. II*, **3**(8), 1289–1296 (1993).

160. L. Noirez, P. Davidson, W. Schwarz, G. Pepy, 'Side-chain Liquid-crystal Polymers – a Study by Neutron-diffraction of the Backbone Distribution Profile in the Smectic A Phase', *Liq. Cryst.*, **16**(6), 1081–1092 (1994).
161. S. Poser, M. Arnold, H. Fischer, 'Liquid-crystalline Side-group Block-copolymers with a Uniform Polymer Backbone', *Eur. Polym. J.*, **32**(10), 1169–1173 (1996).
162. M. Lemaire, W. Buchner, R. Garreau, J.P. Guette, J.P. Cotton, 'Study of Poly(3-alkylthiophene) by Small-angle Neutron Scattering (SANS)', *New J. Chem.*, **16**(10), 1009–1012 (1992).
163. W.W. Graessley, R. Krishnamoorti, N.P. Balsara, L.J. Fetters, D.J. Lohse, D.N. Schulz, J.A. Sissano, 'Effect of Deuterium Substitution on Thermodynamic Interactions in Polymer Blends', *Macromolecules*, **26**(5), 1137–1143 (1993).
164. W.W. Graessley, R. Krishnamoorti, N.P. Balsara, R.J. Butera, L.J. Fetters, D.J. Lohse, D.N. Schulz, J.A. Sissano, 'Thermodynamics of Mixing for Blends of Model Ethylene–Butene Copolymers', *Macromolecules*, **27**(14), 3896–3901 (1994).
165. H. Jinnai, H. Hasegawa, T. Hashimoto, R.M. Briber, C.C. Han, 'Effect of Cross-links on the Miscibility of a Deuterated Polybutadiene and Protonated Polybutadiene Blend', *Macromolecules*, **26**(1), 182–188 (1993).
166. H. Jinnai, H. Hasegawa, T. Hashimoto, C.C. Han, 'Inversion of the Phase Diagram from UCST to LCST in Deuterated Polybutadiene and Protonated Polybutadiene Blends', *Macromolecules*, **25**(22), 6078–6080 (1992).
167. H. Takeno, S. Koizumi, H. Hasegawa, T. Hashimoto, 'Small-angle Neutron-scattering Study of Anomalous Mixing Behaviors in Deuterated Polystyrene Poly(vinyl methyl ether) Mixtures Near the Glass-transition Temperature', *Macromolecules*, **29**(7), 2440–2448 (1996).
168. A.R. Nesarikar, 'Rheology of Polymer Blend Liquid–Liquid Phase Separation', *Macromolecules*, **28**(21), 7202–7207 (1995).
169. A.M. Fernandez, L.H. Sperling, G.D. Wignall, 'Conformation of Polybutadiene in the Network State by Small-angle Neutron Scattering', *Macromolecules*, **19**(10), 2572–2575 (1986).
170. W. Gronski, F. Forster, W. Pyckhouthintzen, T. Springer, 'Molecular Aspects of Polymer Network Deformation – Small-angle Neutron Scattering and NMR Studies', *Makromol. Chem. Macromol. Symp.*, **40**, 121–137 (1990).
171. S. Sakurai, H. Hasegawa, T. Hashimoto, C.C. Han, 'Chain Dimension Determination of Deuterated Polybutadiene by Small-angle Neutron Scattering on the Basis of Random Phase Approximation', *Polym. Commun.*, **31**(3), 99–102 (1990).
172. S. Sakurai, H. Hasegawa, T. Hashimoto, I.G. Hargis, S.L. Aggarwal, C.C. Han, 'Microstructure and Isotopic Labeling Effects on the Miscibility of Polybutadiene Blends Studied by the Small-angle Neutron Scattering Technique', *Macromolecules*, **23**(2), 451–459 (1990).
173. N.P. Balsara, L.J. Fetters, N. Hadjichristidis, D.J. Lohse, C.C. Han, W.W. Graessley, R. Krishnamoorti, 'Thermodynamic Interactions in Model Polyolefin Blends Obtained by Small-angle Neutron Scattering', *Macromolecules*, **25**(23), 6137–6147 (1992).
174. B. Crist, 'Polymer Self-diffusion Measurements by Small-angle Neutron Scattering', *J. Non-Cryst. Solids*, **131**, 709–714 (1991).
175. I.W. Hamley, J.P.A. Fairclough, S.M. King, J.S. Pedersen, R.M. Richardson, C.T. Imrie, A.A. Craig, 'A Small-angle Neutron Scattering Study of the Conformation of a Side Chain Liquid Crystal Poly(methacrylate) in the Smectic C Phase', *Liq. Cryst.*, **22**(6), 679–684 (1997).
176. G. Beaucage, S. Sukumaran, S.J. Clarson, M.S. Kent, D.W. Schaefer, 'Symmetric, Isotopic Blends of Poly (dimethylsiloxane)', *Macromolecules*, **29**(26), 8349–8356 (1996).
177. M.H. Wagner, 'Analysis of Small-angle Neutron Scattering Data on Poly(dimethylsiloxane) Network Unfolding', *Macromolecules*, **27**(18), 5223–5226 (1994).
178. A.N. Falcao, J.S. Pedersen, K. Mortensen, 'Neutron Scattering Experiments on Swollen, Uniaxially Stretched Polymer Networks', *J. Mol. Struct.*, **383**(1–3), 69–74 (1996).
179. A.N. Falcao, J.S. Pedersen, K. Mortensen, 'Structure of Randomly Cross-linked Poly(dimethylsiloxane) Networks Produced by Electron Irradiation', *Macromolecules*, **26**(20), 5350–5364 (1993).
180. A.N. Falcao, J.S. Pedersen, K. Mortensen, F. Boue, 'Poly-dimethylsiloxane Networks at Equilibrium Swelling – Extracted and Nonextracted Networks', *Macromolecules*, **29**(3), 809–818 (1996).
181. L. Mandelkern, R.G. Alamo, G.D. Wignall, F.C. Stehling, 'The Phase Structure of Blends of the Different Polyethylenes', *Trends Polym. Sci.*, **4**(11), 377–383 (1996).
182. K. Tashiro, K. Imanishi, M. Izuchi, M. Kobayashi, Y. Itoh, M. Imai, Y. Yamaguchi, M. Ohashi, R.S. Stein, 'Cocrystallization and Phase Segregation of Polyethylene Blends Between the D-species and H-species. 8. Small-angle Neutron Scattering Study of the Molten State and the Structural Relationship of Chains Between the Melt and the Crystalline State', *Macromolecules*, **28**(25), 8484–8490 (1995).
183. B. Crist, 'Ambiguities in the Interpretation of Small-angle Neutron Scattering from Blends of Linear and Branched Polyethylene – Comments', *Polymer*, **38**(12), 3145–3147 (1997).
184. B. Crist, 'Concentration and Chain-length Dependence of Thermodynamic Interactions in Polyethylene Isotope Blends', *J. Polym. Sci. B, Polym. Phys.*, **35**(17), 2889–2899 (1997).
185. R.S. Stein, 'Some Scattering Studies of Polymer Microstructures, Surfaces, and Interfaces', *Pure Appl. Chem.*, **63**(7), 941–950 (1991).

186. G.D. Wignall, J.D. Londono, J.S. Lin, R.G. Alamo, M.J. Galante, L. Mandelkern, 'Morphology of Blends of Linear and Long-chain-branched Polyethylenes in the Solid State – A Study by SANS, SAXS, and DSC', *Macromolecules*, **28**(9), 3156–3167 (1995).
187. W. Wu, G.D. Wignall, 'Small-angle Neutron Scattering of Partially Segregated Blends of Polyethylene and Deuteropolyethylene', *Polymer*, **26**(5), 661–666 (1985).
188. B. Rosi-Schwartz, G.R. Mitchell, 'Extracting Force Fields for Disordered Polymeric Materials from Neutron Scattering Data', *Polymer*, **37**(10), 1857–1870 (1996).
189. D.M. Sadler, S.J. Spells, 'A Neutron Scattering Study of Slowly Crystallized Bulk Polyethylene', *Polymer*, **25**(9), 1219–1226 (1984).
190. M. Stamm, 'Influence of the Conformation of Polyethylene on Wide-angle Neutron Scattering Patterns', *J. Polym. Sci., Polym. Phys. Ed.*, **20**(2), 235–244 (1982).
191. G.D. Wignall, L. Mandelkern, C. Edwards, M. Glotin, 'Wide-angle Neutron Scattering Studies of Mixed-crystals of Polyethylene and Deuteropolyethylene', *J. Polym. Sci., Polym. Phys. Ed.*, **20**(2), 245–258 (1982).
192. W. Wu, G.D. Wignall, L. Mandelkern, 'A SANS Study of the Plastic-deformation Mechanism in Polyethylene', *Polymer*, **33**(19), 4137–4140 (1992).
193. S.J. Spells, E.U. Okoroafor, 'A Neutron Scattering Study of Polyethylene Deformed by Rolling', *Polymer*, **35**(21), 4590–4594 (1994).
194. B.K. Annis, J. Strizak, G.D. Wignall, R.G. Alamo, L. Mandelkern, 'A Small-angle Neutron Scattering Study of the Plastic-deformation of Linear Polyethylene', *Polymer*, **37**(1), 137–140 (1996).
195. D.W.M. Marr, M. Wartenberg, K.B. Schwartz, M.M. Agamalian, G.D. Wignall, 'Void Morphology in Polyethylene–Carbon Black Composites', *Macromolecules*, **30**(7), 2120–2124 (1997).
196. A. Mahendrasingam, Y. Alhayalee, V.T. Forsyth, P. Langan, W. Fuller, R.J. Oldman, D.J. Blundell, S. Mason, 'Neutron-diffraction Studies of the Structure of PEEK', *Physica B*, **180**(A), 528–530 (1992).
197. M.H. Li, A. Brulet, J.P. Cotton, P. Davidson, C. Strazielle, P. Keller, 'Study of the Chain Conformation of Thermotropic Nematic Main-chain Polyesters', *J. Phys. II*, **4**(10), 1843–1863 (1994).
198. M.H. Li, A. Brulet, P. Keller, C. Strazielle, J.P. Cotton, 'Study of the Transesterification of a Main-chain Mesomorphic Polyester by Small-angle Neutron Scattering', *Macromolecules*, **26**(1), 119–124 (1993).
199. R.G. Kirste, H. Salie, P. Lindner, 'The Structure of Polyetherketones', *Physica B*, **180**(A), 552–554 (1992).
200. W.S. Dubner, J.M. Schultz, G.D. Wignall, 'Estimation of Incoherent Backgrounds in SANS Studies of Polymers', *J. Appl. Crystallogr.*, **23**, 469–475 (1990).
201. M. Imai, K. Kaji, T. Kanaya, Y. Sakai, 'Chain Conformation in the Induction Period of Crystallization of Poly(ethylene-terephthalate)', *Physica B*, **213**, 718–720 (1995).
202. M. Imai, K. Kaji, T. Kanaya, Y. Sakai, 'Ordering Process in the Induction Period of Crystallization of Poly(ethylene-terephthalate)', *Phys. Rev. B: Condens. Matter*, **52**(17), 12 696–12 704 (1995).
203. K.P. Mcalea, J.M. Schultz, K.H. Gardner, G.D. Wignall, 'Ester Interchange Reactions in Poly(ethylene-terephthalate) – Observation Using Small-angle Neutron Scattering', *Polymer*, **27**(10), 1581–1584 (1986).
204. K.P. Mcalea, J.M. Schultz, K.H. Gardner, G.D. Wignall, 'Small-angle and Intermediate-angle Neutron Scattering from Amorphous and Semicrystalline Polyethylene Terephthalate', *J. Polym. Sci. B, Polym. Phys.*, **25**(3), 651–661 (1987).
205. R. Krishnamoorti, W.W. Graessley, L.J. Fetters, R.T. Garner, D.J. Lohse, 'Anomalous Mixing Behavior of Polyisobutylene with Other Polyolefins', *Macromolecules*, **28**(4), 1252–1259 (1995).
206. J.M. O'Reilly, D.M. Teegarden, G.D. Wignall, 'Small-angle and Intermediate-angle Neutron Scattering from Stereoregular Poly(methyl methacrylate)', *Macromolecules*, **18**(12), 2747–2752 (1985).
207. I. Hopkinson, F.T. Kiff, R.W. Richards, S.M. King, H. Munro, 'Thermodynamics of Isotopic Mixtures of Syndiotactic Poly(methyl methacrylate) from Small-angle Neutron Scattering', *Polymer*, **35**(8), 1722–1729 (1994).
208. D.J. Ward, G.R. Mitchell, 'Modeling the Structure of a Polymer Glass Poly(methyl methacrylate) Through Neutron Scattering Experiments', *Phys. Scr.*, **T57**, 153–160 (1995).
209. W.W. Graessley, R. Krishnamoorti, G.C. Reichart, N.P. Balsara, L.J. Fetters, D.J. Lohse, 'Regular and Irregular Mixing in Blends of Saturated-hydrocarbon Polymers', *Macromolecules*, **28**(4), 1260–1270 (1995).
210. G.C. Reichart, W.W. Graessley, R.A. Register, R. Krishnamoorti, D.J. Lohse, 'Anomalous Attractive Interactions in Polypropylene Blends', *Macromolecules*, **30**(10), 3036–3041 (1997).
211. M. Stamm, J. Schelten, D.G.H. Ballard, 'Determination of the Chain Conformation of Polypropylene in the Crystalline State by Neutron Scattering', *Colloid Polym. Sci.*, **259**(2), 286–292 (1981).
212. A. Zirkel, V. Urban, D. Richter, L.J. Fetters, J.S. Huang, R. Kampmann, N. Hadjichristidis, 'Small-angle Neutron Scattering Evaluation of the Temperature Dependence of Atactic Polypropylene and Poly(1-butene) Chain Dimensions in the Melt', *Macromolecules*, **25**(23), 6148–6155 (1992).
213. G. Mao, J.E. Fischer, F.E. Karasz, M.J. Winokur, 'Nonplanarity and Ring Torsion in Poly(*p*-phenylene vinylene) – a Neutron-diffraction Study', *J. Chem. Phys.*, **98**(1), 712–716 (1993).
214. F. Vicentini, L. Noirez, G. Pepy, M. Mauzac, 'Bulk Conformation of a Conductive Polymer Substituted with Lateral Liquid-crystalline Groups', *Europhys. Lett.*, **32**(8), 657–662 (1995).

215. K.M. O'Connor, J.M. Pochan, P. Thiyagarajan, 'Small-angle Neutron Scattering Studies of Interactions and Chain Dimensions in Bimodal Polystyrene Blends', *Polymer*, **32**(2), 195–201 (1991).
216. D. Schwahn, K. Hahn, J. Streib, T. Springer, 'Critical Fluctuations and Relaxation Phenomena in the Isotopic Blend Polystyrene–Deuteropolystyrene Investigated by Small-angle Neutron Scattering', *J. Chem. Phys.*, **93**(11), 8383–8391 (1990).
217. M.R. Landry, 'Scaling Behavior of Swollen Chains in Bimodal Molecular Weight Blends', *Macromolecules*, **30**(24), 7500–7510 (1997).
218. G. Muller, D. Schwahn, H. Eckerlebe, J. Rieger, T. Springer, 'Deviation of Early Stage of Spinodal Decomposition from the Cahn–Hilliard–Cook Theory Observed in an Isotopic Polymer Blend', *Physica B*, **234**, 245–246 (1997).
219. G. Muller, D. Schwahn, H. Eckerlebe, J. Rieger, T. Springer, 'Effect of the Onsager Coefficient and Internal Relaxation Modes on Spinodal Decomposition in the High-molecular Isotopic Blend Polystyrene–Deuteropolystyrene Studied with Small-angle Neutron Scattering', *J. Chem. Phys.*, **104**(13), 5326–5337 (1996).
220. D. Schwahn, H. Eckerlebe, E. Hadicke, T. Springer, 'Observation of Internal Relaxation Modes During Early State of Spinodal Decomposition in the Isotopic Polymer Mixture D-Ps/Ps with SANS', *J. Phys. IV*, **3**(C8), 13–16 (1993).
221. D. Schwahn, T. Springer, S. Janssen, E. Hadicke, 'Investigation of Equilibrium and Nonequilibrium States of Polymer Blends by Small-angle Neutron Scattering', *J. Appl. Crystallogr.*, **24**, 685–691 (1991).
222. K. Shuto, Y. Oishi, T. Kajiyama, C.C. Han, 'Aggregation Structure of a 2-Dimensional Ultrathin Polystyrene Film Prepared by the Water Casting Method', *Macromolecules*, **26**(24), 6589–6594 (1993).
223. K. Shuto, Y. Oishi, T. Kajiyama, C.C. Han, 'Preparation of 2-Dimensional Ultra Thin Polystyrene Film by Water Casting Method', *Polym. J.*, **25**(3), 291–300 (1993).
224. B. Rosi-Schwartz, G.R. Mitchell, A.K. Soper, 'Neutron Scattering from Hydrogenous Polymers', *Physica B*, **180**(A), 561–563 (1992).
225. B. Rosi-Schwartz, G.R. Mitchell, A.K. Soper, 'A New Approach to Neutron Scattering for Noncrystalline Hydrogenous Polymers', *Polymer*, **33**(17), 3744–3747 (1992).
226. R.M. Briber, B.J. Bauer, 'Small-angle Neutron Scattering of Blends of Cross-linked and Linear Polystyrene', *Macromolecules*, **24**(8), 1899–1904 (1991).
227. C. Hayes, L. Bokobza, F. Boue, E. Mendes, L. Monnerie, 'Relaxation Dynamics in Bimodal Polystyrene Melts – a Fourier Transform Infrared Dichroism and Small-angle Neutron Scattering Study', *Macromolecules*, **29**(14), 5036–5041 (1996).
228. C. Hayes, E. Mendes, L. Bokobza, F. Boue, L. Monnerie, 'Analysis of Orientational Relaxation in Binary Blends of Long and Short Polystyrene Chains by Fourier Transform Infrared Dichroism and Small-angle Neutron Scattering', *Macromol. Symp.*, **94**, 227–242 (1995).
229. R. Muller, C. Picot, 'Chain Conformation in Polymer Melts During Flow as Measured by Small-angle Neutron Scattering', *Makromol. Chem., Macromol. Symp.*, **56**, 107–115 (1992).
230. J.R. Schroeder, B. Hammouda, R.A. Bubeck, J.W. Chang, 'SANS from Hot-stretched Polystyrene Revisited', *J. Polym. Sci. B, Polym. Phys.*, **29**(13), 1607–1612 (1991).
231. S. Lecommandoux, L. Noirez, M. Mauzac, F. Hardouin, 'Small-angle Neutron Scattering Experiments in Smectic-A and Smectic-B Phases of Diluted Liquid-crystal Copolysiloxanes', *J. Phys. II*, **4**(12), 2249–2255 (1994).
232. G.R. Mitchell, B. Rosi-Schwartz, 'The Local Structure of Molten Poly(tetrafluoroethylene)', *Physica B*, **180**(A), 558–560 (1992).
233. S.A. Visser, G. Pruckmayr, S.L. Cooper, 'Small-angle Neutron Scattering Analysis of Model Polyurethane Ionomers', *Macromolecules*, **24**(25), 6769–6775 (1991).
234. M. Shibayama, H. Kurokawa, S. Nomura, S. Roy, R.S. Stein, W.L. Wu, 'Small-angle Neutron Scattering Studies on Chain Asymmetry of Coextruded Poly(vinyl alcohol) Film', *Macromolecules*, **23**(5), 1438–1443 (1990).
235. M. Sugiyama, Y. Maeda, 'Small-angle Neutron Scattering Study of Annealed and Stretched PVA Films', *J. Phys. Soc. Jpn.*, **64**(3), 1002–1010 (1995).
236. R. Scherrenberg, H. Reynaers, K. Mortensen, W. Vlak, C. Gondard, 'Structural Aspects of Suspension Poly (vinyl chloride) – Small-angle Neutron Scattering of Rigid and Plasticized Suspension PVC', *Macromolecules*, **26**(12), 3205–3211 (1993).
237. K. Nishida, H. Urakawa, K. Kaji, B. Gabrys, J.S. Higgins, 'Electrostatic Persistence Length of NaPSS Polyelectrolytes Determined by a Zero Average Contrast SANS Technique', *Polymer*, **38**(24), 6083–6085 (1997).
238. A. Rameau, Y. Gallot, P. Marie, B. Farnoux, 'Deuterated Polystyrene–Poly(alpha-methylstyrene) Blends – Range of Miscibility and Determination of the Interaction Parameter by Small-angle Neutron Scattering', *Polymer*, **30**(3), 386–392 (1989).
239. S. Janssen, D. Schwahn, T. Springer, 'Mean-field Ising Crossover and the Critical Exponent-γ, Exponent-ν, and Exponent-η for a Polymer Blend – d-PB/PS Studied by Small-angle Neutron Scattering', *Phys. Rev. Lett.*, **68**(21), 3180–3183 (1992).
240. D. Schwahn, T. Springer, S. Janssen, 'Equilibrium and Nonequilibrium States of Polymer Blends Investigated by Small-angle Neutron Scattering', *Physica B*, **174**(1–4), 159–163 (1991).
241. P.E. Tomlins, J.S. Higgins, 'Molecular-conformation and Interactions in Oligomeric Mixtures of Polystyrene

with Polybutadiene – a Small-angle Neutron Scattering Study', *Polymer*, **26**(10), 1554–1560 (1985).
242. L. Sung, A.I. Nakatani, C.C. Han, A. Karim, J.F. Douglas, S.K. Satija, 'The Role of the Copolymer Additives on Phase Behavior of a Polymer Blend', *Physica B*, **241**, 1013–1015 (1997).
243. M. Hidalgo, J.Y. Cavaille, B. Cabane, Y. Chevalier, J. Guillot, L. Rios, R. Vassoille, 'Coalescence of Polystyrene Nodules in Soft Poly(butyl acrylate)-based Films', *Polym. Adv. Technol.*, **6**(5), 296–300 (1995).
244. B. Hammouda, B.J. Bauer, T.P. Russell, 'Small-angle Neutron Scattering from Deuterated Polystyrene–Poly (butyl methacrylate) Homopolymer Blend Mixtures', *Macromolecules*, **27**(8), 2357–2359 (1994).
245. M.J. He, Y.M. Liu, F. Yi, J. Ming, C.C. Han, 'Spinodal Decomposition in a Hydrogen-bonded Polymer Blend', *Macromolecules*, **24**(2), 464–473 (1991).
246. E.K. Hobbie, G. Merkle, B.J. Bauer, C.C. Han, 'Spinodal Decomposition in Hydrogen-bonded Polymer Blends', *Mod. Phys. Lett. B*, **10**(25), 1219–1226 (1996).
247. G. Merkle, B.J. Bauer, C.C. Han, 'Relaxation After a Temperature Jump within the One-phase Region of a Polymer Mixture', *J. Chem. Phys.*, **104**(23), 9647–9650 (1996).
248. V. Abetz, 'Determination of Dynamic and Thermodynamic Quantities from the Scaling Behavior During the Later Stages of Spinodal Decomposition', *Macromolecules*, **27**(16), 4621–4624 (1994).
249. T. Hack, V. Abetz, M. Stamm, D.W. Schubert, K. Mortensen, W. Siol, 'Spinodal Decomposition of a Polystyrene/Poly(cyclohexyl acrylate–stat-butyl methacrylate) Blend', *Colloid Polym. Sci.*, **274**(4), 350–355 (1996).
250. D.W. Schubert, M. Stamm, 'Neutron Reflectometry and Small-angle Neutron Scattering – Two Complementary Techniques for Polymer Blend Investigations', *Physica B*, **234**, 286–288 (1997).
251. B. McGarey, R.W. Richards, 'The Size of Phase Separated Regions in Poly(dimethylsiloxane) Polystyrene Sequential Interpenetrating Networks Determined by Small-angle Neutron Scattering', *Polymer*, **27**(9), 1315–1324 (1986).
252. A. Maconnachie, R.P. Kambour, R.C. Bopp, 'Compatibility and Neutron Scattering Studies of Mixtures of Polystyrene with Poly(2,6-dimethyl-1,4-phenylene oxide) and its Brominated Derivatives', *Polymer*, **25**(3), 357–364 (1984).
253. D.G. Bucknall, J.S. Higgins, S. Rostami, 'The Compatibilizing Effect of Diblock Copolymer on the Morphology of Immiscible Polymer Blends', *Polymer*, **33**(20), 4419–4422 (1992).
254. L.J. An, D.Y. He, J.K. Jing, Z.G. Wang, D.H. Yu, B.Z. Jiang, Z.H. Jiang, R.T. Ma, 'Effects of Molecular Weight and Interaction Parameter on the Glass Transition Temperature of Polystyrene Mixtures and its Blends with Polystyrene/Poly(2,6-dimethyl-*p*-phenylene oxide)', *Eur. Polym. J.*, **33**(9), 1523–1528 (1997).
255. H. Hasegawa, S. Sakurai, M. Takenaka, T. Hashimoto, C.C. Han, 'Small-angle Neutron Scattering and Light Scattering Studies on the Miscibility of Protonated Polyisoprene–Deuterated Polybutadiene Blends', *Macromolecules*, **24**(8), 1813–1819 (1991).
256. T. Hashimoto, H. Jinnai, H. Hasegawa, C.C. Han, 'Late-stage Spinodal Decomposition in Binary Critical Fluids – Scaling Function Obtained Over a Wide Q-Space of 4 Orders of Magnitude', *Physica A*, **204**(1–4), 261–276 (1994).
257. S. Sakurai, H. Jinnai, H. Hasegawa, T. Hashimoto, C.C. Han, 'Microstructure Effects on the Lower Critical Solution Temperature Phase Behavior of Deuterated Polybutadiene and Protonated Polyisoprene Blends Studied by Small-angle Neutron Scattering', *Macromolecules*, **24**(17), 4839–4843 (1991).
258. H. Yang, G. Hadziioannou, R.S. Stein, 'The Effect of Deuteration on the Phase Equilibrium of the Polystyrene–Poly(vinyl methyl ether) Blend System', *J. Polym. Sci. B, Polym. Phys.*, **21**(1), 159–162 (1983).
259. S. Choi, X.D. Liu, R.M. Briber, 'Small-angle Neutron Scattering of Polymer Blends of Polyvinylmethylether at Dilute Concentration in Deuterated Polystyrene', *J. Polym. Sci. B, Polym. Phys.*, **36**(1), 1–9 (1998).
260. M. Shibayama, H.J. Yang, R.S. Stein, C.C. Han, 'Study of Miscibility and Critical Phenomena of Deuterated Polystyrene and Hydrogenated Poly(vinyl methyl ether) by Small-angle Neutron Scattering', *Macromolecules*, **18**(11), 2179–2187 (1985).
261. G. Beaucage, R.S. Stein, 'Mechanical Behavior and Morphology of Tactic Poly(vinyl methyl ether)–Polystyrene Blends', *Polymer*, **35**(13), 2716–2724 (1994).
262. F.B.C. Larbi, S. Leloup, J.L. Halary, L. Monnerie, 'Phase Diagrams of Poly(vinyl methyl ether)–Selectively Deuterated Polystyrene – An Original Route to the Precise Nature of Polymer–Polymer Interactions', *Polym. Commun.*, **27**(1), 23–25 (1986).
263. B. Hammouda, R.M. Briber, B.J. Bauer, 'Small-angle Neutron Scattering from Deuterated Polystyrene–Poly(vinylmethyl ether)–Protonated Polystyrene Ternary Polymer Blends', *Polymer*, **33**(8), 1785–1787 (1992).
264. J.W. Rhee, B. Crist, 'Thermodynamics and Phase Separation in Melt Blends of Polyethylene and Model Copolymers', *Macromolecules*, **24**(20), 5663–5669 (1991).
265. E. Abtal, R.E. Prudhomme, 'Study of the Orientation of Poly(styrene)–Poly(vinyl methyl ether) Blends Relative to the Glass-transition Temperature', *Macromolecules*, **27**(20), 5780–5786 (1994).
266. Q. Trancong, Y. Ishida, A. Tanaka, T. Soen, 'Further Evidence for the Effects of Critical Concentration Fluctuations on the Diffusion-controlled Reaction Kinetics in Binary Polymer Mixtures', *Polym. Bull.*, **29**(1–2), 89–96 (1992).

267. Q. Trancong, K. Meisyo, Y. Ishida, O. Yano, T. Soen, M. Shibayama, 'Effects of Critical Concentration Fluctuations on the Photocyclization of a Bichromophoric Molecule in the One-phase Region of Polystyrene Poly(vinyl methyl ether) Blends', *Macromolecules*, **25**(9), 2330–2335 (1992).
268. B. Hammouda, B.J. Bauer, 'Compressibility of 2 Polymer Blend Mixtures', *Macromolecules*, **28**(13), 4505–4508 (1995).
269. N.C.B. Tan, X. Liu, R.M. Briber, D.G. Peiffer, 'Immiscibility in Polystyrene–Sulfonated Polystyrene Blends', *Polymer*, **36**(10), 1969–1973 (1995).
270. M. Motowoka, H. Jinnai, T. Hashimoto, Y. Qiu, C.C. Han, 'Phase Separation in Deuterated Polycarbonate Poly(methylmethacrylate) Blend Near Glass-transition Temperature', *J. Chem. Phys.*, **99**(3), 2095–2100 (1993).
271. J.N. Clark, J.S. Higgins, C.K. Kim, D.R. Paul, 'Reexamination of Poly(methyl methacrylate)–Poly(epichlorohydrin) Blends by Small-angle Neutron Scattering and Differential Scanning Calorimetry', *Polymer*, **33**(15), 3137–3144 (1992).
272. J.M.R. Lefebvre, R.S. Porter, G.D. Wignall, 'On the Deformation of Poly(methyl methacrylate)–Poly (ethylene oxide) Blends – A Molecular Characterization By Small-angle Neutron Scattering', *Polym. Eng. Sci.*, **27**(6), 433–439 (1987).
273. E.H. Hellmann, G.P. Hellmann, A.R. Rennie, 'Chain Fragment Diffusion in Liquid and Glassy Polymer Blends', *Macromolecules*, **24**(13), 3821–3827 (1991).
274. T. Sato, K. Katayama, T. Suzuki, T. Shiomi, 'UCST and LCST Behavior in Polymer Blends Containing Poly(methyl methacrylate-stat-styrene)', *Polymer*, **39**(4), 773–780 (1998).
275. J.N. Clark, M.L. Fernandez, P.E. Tomlins, J.S. Higgins, 'Small-angle Neutron Scattering Studies of Phase Equilibria in Blends of Deuterated Poly(methyl methacrylate) with Solution Chlorinated Polyethylene', *Macromolecules*, **26**(22), 5897–5907 (1993).
276. J.N. Clark, J.S. Higgins, D.G. Peiffer, 'Reduction in Domain Size in Polymer Mixtures via Charge-transfer Interactions', *Polym. Eng. Sci.*, **32**(1), 49–56 (1992).
277. P. Carlsson, J. Swenson, R.L. Mcgreevy, B. Gabrys, W.S. Howells, L. Borjesson, L.M. Torell, 'Structural and Dynamical Properties of Polymer Electrolytes PPO–$LiClO_4$', *Physica B*, **234**, 231–235 (1997).
278. S.S. Chakraborty, N. Maiti, B.M. Mandal, S.N. Bhattacharyya, 'Miscibility and Phase Diagrams for Poly(vinyl methyl ether) and Polyacrylate Blend Systems', *Polymer*, **34**(1), 111–114 (1993).
279. J. Caspar, B. Dunges, R.G. Kirste, T. Heitz, A. Wiedenmann, 'Preparation and Determination of the Structure of High-performance Polymer Blends by Small-angle Neutron Scattering', *Macromol. Rapid Commun.*, **16**(10), 771–778 (1995).
280. J. Caspar, B. Dunges, R.G. Kirste, T. Heitz, A. Wiedenmann, 'The Structure of High-performance Polymer Blends by Small-angle Neutron Scattering', *Physica B*, **234**, 240–241 (1997).
281. B.K. Annis, G.D. Wignall, A.R. Hopkins, P.G. Rasmussen, R.A. Basheer, 'Structural Investigation of Polyaniline/Nylon 6 Blends by Small-angle Neutron Scattering and Small- and Wide-angle X-ray Scattering', *J. Polym. Sci. B, Polym. Phys.*, **35**(17), 2765–2774 (1997).
282. A.R. Hopkins, P.G. Rasmussen, R.A. Basheer, B.K. Annis, G.D. Wignall, 'Comparison of Phase Separation Structures in Blends of Polyaniline/Camphor Sulfonic Acid and Nylons 6 and 12', *Synth. Met.*, **95**(3), 179–184 (1998).
283. S. Elliniadis, J.S. Higgins, N. Clarke, T.C.B. Mcleish, R.A. Choudhery, S.D. Jenkins, 'Phase Diagram Prediction for Thermoset/Thermoplastic Polymer Blends', *Polymer*, **38**(19), 4855–4862 (1997).
284. S. Sakurai, T. Izumitani, H. Hasegawa, T. Hashimoto, C.C. Han, 'Small-angle Neutron Scattering and Light Scattering Study on the Miscibility of Poly(styrene-ranbutadiene)–Polybutadiene Blends', *Macromolecules*, **24**(17), 4844–4851 (1991).
285. T.R. Maier, A.M. Jamieson, R. Simha, 'Phase Equilibria in SBR Polybutadiene Elastomer Blends – Application of Flory–Huggins Theory', *J. Appl. Polym. Sci.*, **51**(6), 1053–1062 (1994).
286. J.P. Baetzold, I. Gancarz, X. Quan, J.T. Koberstein, 'Mechanical Property Modification and Morphology of Poly(styrene-*b*- Hydrogenated butadiene-*b*-styrene)–Poly(hydrogenated butadiene) Blends', *Macromolecules*, **27**(19), 5329–5340 (1994).
287. Y.W. Cheung, R.S. Stein, 'Critical Analysis of the Phase Behavior of Poly(Epsilon-caprolactone) (PCL)/Polycarbonate (PC) Blends', *Macromolecules*, **27**(9), 2512–2519 (1994).
288. Y.W. Cheung, R.S. Stein, G.D. Wignall, H.E. Yang, 'Small-angle Scattering Investigations of Poly(Epsilon-caprolactone)–Polycarbonate Blends. 1. Small-angle Neutron and X-ray-scattering Study of Crystalline Blend Morphology', *Macromolecules*, **26**(20), 5365–5371 (1993).
289. A. Brulet, M. Daoud, P. Zhou, H.L. Frisch, 'Small-angle Neutron Scattering by Interpenetrating Polymer Networks of Poly(carbonate–urethane) and Polyvinyl Pyridine', *J. Phys. II*, **3**(7), 1161–1168 (1993).
290. D.W.M. Marr, 'Neutron Scattering Studies of a Semicrystalline Homopolymer and Copolymer Blend', *Macromolecules*, **28**(25), 8470–8476 (1995).
291. N.P. Balsara, S.V. Jonnalagadda, C.C. Lin, C.C. Han, R. Krishnamoorti, 'Thermodynamic Interactions and Correlations in Mixtures of 2 Homopolymers and a Block Copolymer by Small-angle Neutron Scattering', *J. Chem. Phys.*, **99**(12), 10 011–10 020 (1993).
292. C.C. Lin, H.S. Jeon, N.P. Balsara, B. Hammouda, 'Spinodal Decomposition in Multicomponent Polymer Blends', *J. Chem. Phys.*, **103**(5), 1957–1971 (1995).

293. C.C. Lin, S.V. Jonnalagadda, N.P. Balsara, C.C. Han, R. Krishnamoorti, 'Thermodynamic Interactions in Multicomponent Polymer Blends', *Macromolecules*, **29**(2), 661–669 (1996).
294. B. Hammouda, N.P. Balsara, A.A. Lefebvre, 'Small-angle Neutron Scattering from Pressurized Polyethylbutylene/Polymethylbutylene Blends', *Macromolecules*, **30**(18), 5572–5574 (1997).
295. N.E. Clough, I. Hopkinson, R.W. Richards, T. Ibrahim, S.M. King, 'Calorimetric and Small-angle Neutron Scattering Investigation of an Ethylene–Vinyl Acetate Blend', *Polymer*, **36**(22), 4245–4252 (1995).
296. G.C. Reichart, W.W. Graessley, R.A. Register, R. Krishnamoorti, D.J. Lohse, 'Measurement of Thermodynamic Interactions in Ternary Polymer Blends by Small-angle Neutron Scattering', *Macromolecules*, **30**(11), 3363–3368 (1997).
297. C. Jalbert, J.T. Koberstein, A. Hariharan, S.K. Kumar, 'End Group Effects on Surface Properties of Polymers: Semiempirical Calculations and Comparison to Experimental Surface Tensions for Alpha, Omega-functional Poly(dimethylsiloxanes)', *Macromolecules*, **30**(15), 4481–4490 (1997).
298. J.F. Elman, B.D. Johs, T.E. Long, J.T. Koberstein, 'A Neutron Reflectivity Investigation of Surface and Interface Segregation of Polymer Functional End-groups', *Macromolecules*, **27**(19), 5341–5349 (1994).
299. T.F. Schaub, G.J. Kellogg, A.M. Mayes, R. Kulasekere, J.F. Ankner, H. Kaiser, 'Surface Modification via Chain-end Segregation in Polymer Blends', *Macromolecules*, **29**(11), 3982–3990 (1996).
300. I. Hopkinson, F.T. Kiff, R.W. Richards, D.G. Bucknall, A.S. Clough, 'Equilibrium Concentration Profiles of Physically End Tethered Polystyrene Molecules at the Air–Polymer Interface', *Polymer*, **38**(1), 87–98 (1997).
301. R.A.L. Jones, L.J. Norton, K.R. Shull, E.J. Kramer, G.P. Felcher, A. Karim, L.J. Fetters, 'Interfacial Segment Density Profiles of End-anchored Polymers in a Melt', *Macromolecules*, **25**(9), 2359–2368 (1992).
302. F.T. Kiff, R.W. Richards, H.L. Thompson, D.G. Bucknall, J.R.P. Webster, 'Influence of Matrix Molecular-weight on a Brush-like Surface Layer Formed by a Polymer Functionalized at Both Ends', *J. Phys. II*, **7**(12), 1871–1891 (1997).
303. Y. Feng, R.A. Weiss, A. Karim, C.C. Han, J.F. Ankner, H. Kaiser, D.G. Peiffer, 'Compatibilization of Polymer Blends by Complexation. 2. Kinetics of Interfacial Mixing', *Macromolecules*, **29**(11), 3918–3924 (1996).
304. M. Geoghegan, R.A.L. Jones, D.S. Sivia, J. Penfold, A.S. Clough, 'Experimental Study of Surface Segregation and Wetting in Films of a Partially Miscible Polymer Blend', *Phys. Rev. E*, **53**(1B), 825–837 (1996).
305. B. Guckenbiehl, 'Oberflächenanreicherung und Grenzflächendiffusion in Binären Polymermischungen', PhD Thesis, Universität Mainz, 1993.
306. M. Stamm, 'The Correlation Between Segment Interaction Parameter and Interface Width of Polymer Blends', *Macromol. Symp.*, **112**, 25–30 (1996).
307. M. Geoghegan, R.A.L. Jones, R.S. Payne, P. Sakellariou, A.S. Clough, J. Penfold, 'Lamellar Structure in a Thin Polymer Blend Film', *Polymer*, **35**(10), 2019–2027 (1994).
308. M. Geoghegan, R.A.L. Jones, A.S. Clough, J. Penfold, 'The Morphology of As-cast Films of a Polymer Blend – Dependence on Polymer Molecular-weight', *J. Polym. Sci. B, Polym. Phys.*, **33**(8), 1307–1311 (1995).
309. S.J. Roser, R.M. Richardson, M.J. Swann, A.R. Hillman, 'In Situ Neutron Reflectivity Studies of Polybithiophene', *J. Chem. Soc., Faraday Trans.*, **87**(17), 2863–2864 (1991).
310. A.P.Y. Wong, A. Karim, C.C. Han, 'Neutron Reflection Studies of Phase Separation and Transesterification in Thin-film Polymer Blends', *Physica B*, **221**(1–4), 301–305 (1996).
311. H. Yoon, C.C. Han, 'Interfacial Structure in Reacting Poly(methyl methacrylate)/Polycarbonate Polymer Blends', *Polym. Eng. Sci.*, **35**(18), 1476–1480 (1995).
312. D.G. Bucknall, S.A. Butler, H.E. Hermes, J.S. Higgins, 'Neutron Reflectivity Investigations of Semi-crystalline Polymer Interfaces', *Physica B*, **241**, 1071–1073 (1998).
313. R.A.L. Jones, 'Correction', *Phys. Rev. Lett.*, **66**(23), 3087–3087 (1991).
314. G. Krausch, C.A. Dai, E.J. Kramer, J.F. Marko, F.S. Bates, 'Interference of Spinodal Waves in Thin Polymer-films', *Macromolecules*, **26**(21), 5566–5571 (1993).
315. E. Kim, G. Krausch, E.J. Kramer, J.O. Osb, 'Surface-directed Spinodal Decomposition in the Blend of Polystyrene and Tetramethyl-bisphenol-A Polycarbonate', *Macromolecules*, **27**(20), 5927–5929 (1994).
316. L.J. Norton, E.J. Kramer, R.A.L. Jones, F.S. Bates, H.R. Brown, G.P. Felcher, R. Kleb, 'Resonantly Enhanced Neutron Intensity in a Surface Segregated Polymer Blend', *J. Phys. II*, **4**(2), 367–376 (1994).
317. I. Hopkinson, F.T. Kiff, R.W. Richards, S. Affrossman, M. Hartshorne, R.A. Pethrick, H. Munro, J.R.P. Webster, 'Investigation of Surface Enrichment in Isotopic Mixtures of Poly(methyl methacrylate)', *Macromolecules*, **28**(2), 627–635 (1995).
318. S. Tasaki, H. Yamaoka, F. Yoshida, 'Neutron Reflectivity Study on Thin-film Polymer Surface', *Physica B*, **180**(A), 480–482 (1992).
319. E.K. Lin, W.I. Wu, S.K. Satija, 'Polymer Interdiffusion Near an Attractive Solid Substrate', *Macromolecules*, **30**(23), 7224–7231 (1997).
320. M.L. Fernandez, J.S. Higgins, J. Penfold, C.S. Shackleton, 'Structure of Thin Deuterated Poly(methyl methacrylate) Films Studied by the Specular Reflection of Neutrons', *Polym. Commun.*, **31**(4), 124–127 (1990).
321. W.L. Wu, W.J. Orts, J.H. Vanzanten, B.M. Fanconi, 'Density Profile of Spin Cast Polymethylmethacrylate

Thin-films', *J. Polym. Sci. B, Polym. Phys.*, **32**(15), 2475–2480 (1994).
322. W. Wu, C.F. Majkrzak, S.K. Satija, J.F. Ankner, W.J. Orts, M. Satkowski, S.D. Smith, 'The Density Profile at a Polymer Solid Interface', *Polymer*, **33**(23), 5081–5084 (1992).
323. M. Sferrazza, M. Heppenstallbutler, R. Cubitt, D. Bucknall, J. Webster, R.A.L. Jones, 'Interfacial Instability Driven by Dispersive Forces: The Early Stages of Spinodal Dewetting of a Thin Polymer Film on a Polymer Substrate', *Phys. Rev. Lett.*, **81**(23), 5173–5176 (1998).
324. D.W. Schubert, M. Stamm, 'Influence of Chain-length on the Interface Width of an Incompatible Polymer Blend', *Europhys. Lett.*, **35**(6), 419–424 (1996).
325. B.B. Sauer, D.J. Walsh, 'Effect of Solvent Casting on Reduced Entanglement Density in Thin-films Studied by Ellipsometry and Neutron Reflection', *Macromolecules*, **27**(2), 432–440 (1994).
326. K. Kunz, M. Stamm, M. Hartshorne, S. Affrossman, 'Surface Enrichment in Thin-films of PVC and PMMA', *Acta Polym.*, **47**(5), 234–242 (1996).
327. A.N. Burgess, A. Nevin, H.B. Stanley, J. Penfold, 'A Neutron Reflectometry Study of the Production of Thin Polyphenylene Films from Precursor Polymer', *Polymer*, **33**(19), 4141–4144 (1992).
328. S.J. Roser, D.J. Caruana, M. Gerstenberg, 'Neutron Reflection Studies of Poly(phenol) Films', *J. Electroanal. Chem.*, **411**(1–2), 153–160 (1996).
329. M. Tarabia, H. Hong, D. Davidov, S. Kirstein, R. Steitz, R. Neumann, Y. Avny, 'Neutron and X-ray Reflectivity Studies of Self-assembled Heterostructures Based on Conjugated Polymers', *J. Appl. Phys.*, **83**(2), 725–732 (1998).
330. A. Hariharan, S.K. Kumar, M.H. Rafailovich, J. Sokolov, X. Zheng, D.H. Duong, S.A. Schwarz, T.P. Russell, 'The Effect of Finite Film Thickness on the Surface Segregation in Symmetrical Binary Polymer Mixtures', *J. Chem. Phys.*, **99**(1), 656–663 (1993).
331. A. Hariharan, S.K. Kumar, T.P. Russell, 'Reversal of the Isotopic Effect in the Surface Behavior of Binary Polymer Blends', *J. Chem. Phys.*, **98**(5), 4163–4173 (1993).
332. R.A.L. Jones, L.J. Norton, E.J. Kramer, R.J. Composto, R.S. Stein, T.P. Russell, A. Mansour, A. Karim, G.P. Felcher, M.H. Rafailovich, J. Sokolov, X. Zhao, S.A. Schwarz, 'The Form of the Enriched Surface-layer in Polymer Blends', *Europhys. Lett.*, **12**(1), 41–46 (1990).
333. G. Reiter, P. Auroy, L. Auvray, 'Instabilities of Thin Polymer-films on Layers of Chemically Identical Grafted Molecules', *Macromolecules*, **29**(6), 2150–2157 (1996).
334. G. Reiter, S. Huttenbach, M. Foster, M. Stamm, 'X-ray and Neutron Reflectometry for the Investigation of Polymer Diffusion', *Fresenius' J. Anal. Chem.*, **341**(3–4), 284–288 (1991).
335. W.L. Wu, W.E. Wallace, J.H. Vanzanten, B.J. Bauer, D.W. Liu, A. Wong, 'Diffusion of Linear Polystyrene into Crosslinked Polystyrene', *Polymer*, **38**(11), 2583–2594 (1997).
336. A. Karim, G.P. Felcher, T.P. Russell, 'Interdiffusion of Polymers at Short Times', *Macromolecules*, **27**(23), 6973–6979 (1994).
337. D. Brautmeier, M. Stamm, P. Lindner, 'Study on Interdiffusion of Polystyrenes of High Molecular-weight by SANS', *J. Appl. Crystallogr.*, **24**, 665–671 (1991).
338. D.G. Bucknall, S.A. Butler, J.S. Higgins, 'Real Time Measurement of Polymer Diffusion Coefficients Using Neutron Reflection', *Macromolecules*, **32**, 5453–5456 (1999).
339. M. Geoghegan, F. Boue, G. Bacri, A. Menelle, D.G. Bucknall, 'A Neutron Reflectometry Study of Polystyrene Network Interfaces', *Eur. Phys. J. B*, **3**(1), 83–96 (1998).
340. G. Agrawal, R.P. Wool, W.D. Dozier, G.P. Felcher, J. Zhou, S. Pispas, J.W. Mays, T.P. Russell, 'Interdiffusion of Polymers Across Interfaces', *J. Polym. Sci. B, Polym. Phys.*, **34**(17), 2919–2940 (1996).
341. K.A. Welp, R.P. Wool, S.K. Satija, S. Pispas, J. Mays, 'Dynamics of Polymer Interdiffusion: The Ripple Experiment', *Macromolecules*, **31**(15), 4915–4925 (1998).
342. C.J. Clarke, A. Eisenberg, J. Lascala, M.H. Rafailovich, J. Sokolov, Z. Li, S. Qu, D. Nguyen, S.A. Schwarz, Y. Strzhemechny, B.B. Sauer, 'Measurements of the Flory–Huggins Interaction Parameter for Polystyrene–Poly(4-vinylpyridine) Blends', *Macromolecules*, **30**(14), 4184–4188 (1997).
343. M.L. Fernandez, J.S. Higgins, J. Penfold, 'Neutron Reflection Studies at Polymer–Polymer Interfaces', *Makromol. Chem. Macromol. Symp.*, **62**, 103–118 (1992).
344. D.G. Bucknall, M.L. Fernandez, J.S. Higgins, 'Neutron Reflection Studies of Interfaces Between Partially Miscible and Compatibilized Immiscilble Polymers', *Faraday Discuss. Chem. Soc.*, **98**, 19–30 (1994).
345. H.E. Hermes, D.G. Bucknall, J.S. Higgins, R.L. Scherrenberg, 'The Ordering of Semi-crystalline PS-b-hPB Copolymers at a PS/PE Interface and Their Effects on Interfacial Strength', *Polymer*, **39**(14), 3099–3108 (1998).
346. A.M. Mayes, T.P. Russell, M.F. Toney, S. Smith, S.K. Satija, C.F. Majkrzak, 'Chain-end and Junction Point Distributions in Lamellar Microdomains', *Abst. Pap. Am. Chem. Soc.*, **204**(2), 89-PMSE, (1992).
347. A.M. Mayes, T.P. Russell, 'Neutron Reflectivity Studies of Ordered Copolymer Films', *J. Phys. IV*, **3**(C8), 41–47 (1993).
348. R. Kulasekere, H. Kaiser, J.F. Ankner, T.P. Russell, H.R. Brown, C.J. Hawker, A.M. Mayes, 'Neutron Reflectivity Measurements of Homopolymer Interfaces Reinforced with Random Copolymers', *Physica B*, **221**(1–4), 306–308 (1996).
349. N.N. Pellegrini, M. Sikka, S.K. Satija, K.I. Winey, 'Segregation of a Random Copolymer from Miscible Blends', *Macromolecules*, **30**(21), 6640–6644 (1997).

350. D.G. Bucknall, J.S. Higgins, J. Penfold, S. Rostami, 'Segregation Behavior of Deuterated Poly(styrene-block-methyl methacrylate) Diblock Copolymer in the Presence of Poly(methyl methacrylate) Homopolymer', *Polymer*, **34**(3), 451–458 (1993).
351. R. Oslanec, P. Vlcek, W.A. Hamilton, R.J. Composto, 'Crossover of a Block Copolymer Brush in a Polymer Melt from a Stretched to Collapsed Conformation', *Phys. Rev. E*, **56**(3A), R2383–R2386 (1997).
352. L. Jong, E.M. Pearce, T.K. Kwei, W.A. Hamilton, G.S. Smith, G.H. Kwei, 'Neutron Reflectivity Study of Thin-films of Hydrogen-bonded Polymer Blends – Influence of Annealing on Composition Profile', *Macromolecules*, **25**(10), 2619–2623 (1992).
353. J. Lei, E.M. Pearce, T.K. Kwei, W.A. Hamilton, G.S. Smith, G.H. Kwei, 'Hydrogen-bonded Polymer Blends – the Influence of Annealing on the Interface Composition of Thin-films', *Macromolecules*, **25**(25), 6770–6774 (1992).
354. M. Sikka, N. Singh, A. Karim, F.S. Bates, S.K. Satija, C.F. Majkrzak, 'Entropy-driven Surface Segregation in Block Copolymer Melts', *Phys. Rev. Lett.*, **70**(3), 307–310 (1993).
355. M. Sikka, N. Singh, F.S. Bates, A. Karim, S. Satija, C.F. Majkrzak, 'Surface Segregation in Model Symmetrical Polyolefin Diblock Copolymer Melts', *J. Phys. II*, **4**(12), 2231–2248 (1994).
356. R. Schnell, M. Stamm, 'The Self-organization of Diblock Copolymers at Polymer Blend Interfaces', *Physica B*, **234**, 247–249 (1997).
357. J. Genzer, A. Faldi, R. Oslanec, R.J. Composto, 'Surface Enrichment in a Miscible Polymer Blend – an Experimental Test of Self-consistent-field and Long-wavelength Approximation Models', *Macromolecules*, **29**(16), 5438–5445 (1996).
358. B. Guckenbiehl, M. Stamm, T. Springer, 'Surface Enrichment in Blends of Miscible Polymers', *Colloids Surf. A*, **86**, 311–316 (1994).
359. K.H. Dai, L.J. Norton, E.J. Kramer, 'Equilibrium Segment Density Distribution of a Diblock Copolymer Segregated to a Polymer–Polymer Interface', *Macromolecules*, **27**(7), 1949–1956 (1994).
360. P. Papanek, J.E. Fischer, N.S. Murthy, 'Molecular Vibrations in Nylon-6 Studied by Inelastic Neutron Scattering', *Macromolecules*, **29**(6), 2253–2259 (1996).
361. B. Cvikl, J. Peternelj, 'On the Spacer and Mesogenic Unit Qns Reorientational Motion Investigation in Liquid-crystal Polyacrylate', *J. Phys. II*, **2**(1), 87–92 (1992).
362. J. Bartos, J. Kristiak, T. Kanaya, 'Free Volume and Microscopic Dynamics in Amorphous Polymers', *Physica B*, **234**, 435–436 (1997).
363. J. Baschnagel, 'Analysis of the Incoherent Intermediate Scattering Function in the Framework of the Idealized Mode-coupling Theory – a Monte-Carlo Study for Polymer Melts', *Phys. Rev. B: Condens. Matter*, **49**(1), 135–146 (1994).
364. S. Rathgeber, A. Zirkel, L. Willner, D. Richter, A. Brulet, B. Farago, 'Dynamics of Bimodal Polymer Melts in the Crossover Region from Rouse- to Reptation-like Behaviour – A Study with NSE Spectroscopy', *Physica B*, **234**, 258–259 (1997).
365. T. Kanaya, T. Kawaguchi, K. Kaji, 'Local Dynamics of *cis*-1,4-Polybutadiene Near the Glass-transition Temperature, T_g', *Physica B*, **182**(4), 403–408 (1992).
366. R. Zorn, 'Deviation from Gaussian Behavior in the Self-correlation Function of the Proton Motion in Polybutadiene', *Phys. Rev. B: Condens. Matter*, **55**(10), 6249–6259 (1997).
367. D. Richter, 'The Glass-transition in Polymer Melts – Results from Neutron Scattering', *J. Phys.: Condens. Matter*, **8**(47), 9177–9190 (1996).
368. D. Richter, R. Butera, L.J. Fetters, J.S. Huang, B. Farago, B. Ewen, 'Entanglement Constraints in Polymer Melts – a Neutron Spin-echo Study', *Macromolecules*, **25**(23), 6156–6164 (1992).
369. R. Zorn, A. Arbe, J. Colmenero, B. Frick, D. Richter, U. Buchenau, 'Neutron Scattering Study of the Picosecond Dynamics of Polybutadiene and Polyisoprene', *Phys. Rev. E*, **52**(1B), 781–795 (1995).
370. B. Frick, C. Albasimionesco, J. Hendricks, L. Willner, 'Incoherent Inelastic Neutron Scattering on Polybutadiene Under Pressure', *Prog. Theor. Phys. Suppl.*, **126**, 213–218 (1997).
371. U. Buchenau, C. Schonfeld, D. Richter, T. Kanaya, K. Kaji, R. Wehrmann, 'Neutron Scattering Study of the Vibration–Relaxation Crossover in Amorphous Polycarbonates', *Phys. Rev. Lett.*, **73**(17), 2344–2347 (1994).
372. G. Floudas, J.S. Higgins, G. Meier, F. Kremer, E.W. Fischer, 'Dynamics of Bisphenol-A Polycarbonate in the Glassy and Rubbery States as Studied by Neutron Scattering and Complementary Techniques', *Macromolecules*, **26**(7), 1676–1682 (1993).
373. T. Kanaya, T. Kawaguchi, K. Kaji, 'Microscopic View of Glass Transition Dynamics: A Quasielastic Neutron Scattering Study on *trans*-1,4-Polychloroprene', *J. Chem. Phys.*, **105**(10), 4342–4349 (1996).
374. M. Benmouna, U. Maschke, B. Ewen, 'Properties of Polymer Blends and Triblock Copolymers Obtained by Neutron Scattering', *J. Polym. Sci. B, Polym. Phys.*, **34**(13), 2161–2168 (1996).
375. J.S. Higgins, K. Ma, L.K. Nicholson, J.B. Hayter, K. Dodgson, J.A. Semlyen, 'Studies of Cyclic and Linear Poly(dimethyl siloxanes). 12. Observation of Diffusion Behavior by Quasielastic Neutron Scattering', *Polymer*, **24**(7), 793–799 (1983).
376. V. Arrighi, J.S. Higgins, 'Molecular Dynamics of a Main-chain Liquid Crystalline Polyester Below the Crystalline to Nematic Phase Transition', *J. Chem. Soc., Faraday Trans.*, **93**(8), 1605–1612 (1997).
377. D. Heyer, U. Buchenau, M. Stamm, 'Determination of Elastic Shear Constants of Polyethylene at Room

Temperature by Inelastic Neutron Scattering', *J. Polym. Sci., Polym. Phys. Ed.*, **22**(8), 1515–1527 (1984).
378. A.P. Chatterjee, R.F. Loring, 'Calculation of the Dynamic Structure Factor in Polymer Melts', *J. Chem. Phys.*, **101**(2), 1595–1606 (1994).
379. B. Frick, D. Richter, 'Change of the Vibrational Dynamics Near the Glass-transition in Polyisobutylene – Inelastic Neutron Scattering on a Nonfragile Polymer', *Phys. Rev. B: Condens. Matter*, **47**(22), 14 795–14 804 (1993).
380. B. Frick, D. Richter, S. Trevino, 'Inelastic Fast Relaxation in a Weakly Fragile Polymer Glass Near T_g', *Physica A*, **201**(1–3), 88–94 (1993).
381. D. Richter, A. Arbe, J. Colmenero, M. Monkenbusch, B. Farago, R. Faust, 'Molecular Motions in Polyisobutylene: A Neutron Spin-echo and Dielectric Investigation', *Macromolecules*, **31**(4), 1133–1143 (1998).
382. R. Zorn, D. Richter, B. Farago, B. Frick, F. Kremer, U. Kirst, L.J. Fetters, 'Comparative Study of the Segmental Relaxation in Polyisoprene by Quasielastic Neutron Scattering and Dielectric Spectroscopy', *Physica B*, **180**(A), 534–536 (1992).
383. G. Fytas, G. Meier, D. Richter, 'Segmental Mobilities in an Athermal Diblock Copolymer Melt Far Above T-G by Incoherent Neutron Scattering', *J. Chem. Phys.*, **105**(3), 1208–1213 (1996).
384. A. Mermet, E. Duval, N.V. Surovtsev, J.F. Jal, A.J. Dianoux, A.F. Yee, 'Localized Fast Relaxation in Poly(methyl methacrylate) Glass', *Europhys. Lett.*, **38**(7), 515–520 (1997).
385. A. Mermet, N.V. Surovtsev, E. Duval, J.F. Jal, J. Dupuyphilon, A.J. Dianoux, 'Inelastic Neutron and Raman Scattering of Poly(methyl methacrylate) – Nanostructure of Polymer Glasses', *Europhys. Lett.*, **36**(4), 277–282 (1996).
386. J. Bartos, M. Klimova, T. Kanaya, K. Kaji, 'Microscopic Dynamic Aspects of the Free Radical Decay Acceleration in Sub-T-G Region in Atactic Poly(propylene)', *Polymer*, **39**(5), 1107–1112 (1998).
387. T. Kanaya, K. Kaji, J. Bartos, M. Klimova, 'Onset of the Fast Process in Amorphous Polypropylene as Detected by Quasielastic Neutron Scattering and Electron Spin Resonance Techniques', *Macromolecules*, **30**(4), 1107–1110 (1997).
388. H. Takeuchi, J.S. Higgins, A. Hill, A. Maconnachie, G. Allen, G.C. Stirling, 'Investigation of the Methyl Torsion in Isotactic Polypropylene – Comparison Between Neutron Inelastic Scattering Spectra and Normal Coordinate Calculations', *Polymer*, **23**(4), 499–504 (1982).
389. W. Zajac, B.J. Gabrys, R. McGreevy, B. Mattssons, 'Dynamics of the Polymer Matrix in PPO·$NaClO_4$ and PPO·$LiClO_4$ Complexes Studied by QENS', *Physica B*, **226**(1–3), 144–151 (1996).
390. J.L. Sauvajol, P. Papanek, J.E. Fischer, A.J. Dianoux, G. Mao, M.J. Winokur, F.E. Karasz, 'Densities of State of the Vibrational Low-frequency Modes for Poly(*p*-phenylene vinylene) – a Study by Inelastic Neutron Scattering', *J. Chim. Phys. Phys. Chim. Biol.*, **92**(4), 955–958 (1995).
391. T. Kanaya, T. Kawaguchi, K. Kaji, 'Process of Amorphous Polystyrene Below and Above the Glass Transition Temperature T-G as Studied by Quasielastic Neutron Scattering', *J. Chem. Phys.*, **104**(10), 3841–3850 (1996).
392. R. Mukhopadhyay, A. Alegria, J. Colmenero, B. Frick, 'Effect of Blending on the Methyl Side Group Dynamics in Poly(vinyl methyl ether)', *J. Non-cryst. Solids*, **235**, 233–236 (1998).
393. R. Mukhopadhyay, A. Alegria, J. Colmenero, B. Frick, 'Methyl Group Dynamics in Poly(vinylacetate): A Neutron Scattering Study', *Macromolecules*, **31**(12), 3985–3993 (1998).
394. J. Colmenero, R. Mukhopadhyay, A. Alegria, B. Frick, 'Quantum Rotational Tunneling of Methyl Groups in Polymers', *Phys. Rev. Lett.*, **80**(11), 2350–2353 (1998).
395. R. Zorn, A. Alegria, A. Arbe, J. Colmenero, D. Richter, B. Frick, 'Fast-dynamics in Plasticized Poly(vinyl chloride)', *J. Non-cryst. Solids*, **235**, 169–172 (1998).
396. J. Colmenero, 'Segmental Dynamics in Polymer Melts by Relaxation Techniques and Quasielastic Neutron Scattering', *Phys. Scr.*, **T49A**, 227–232 (1993).
397. J. Colmenero, A. Alegria, A. Arbe, B. Frick, 'Temperature and Momentum-transfer Dependence of the Dynamics of the Alpha-relaxation in Polymer Melts – a Quasielastic Neutron Scattering Study', *Physica B*, **182**(4), 369–375 (1992).
398. J. Colmenero, A. Arbe, G. Coddens, B. Frick, C. Mijangos, H. Reinecke, 'Crossover from Independent to Cooperative Segmental Dynamics in Polymers: Experimental Realization in Poly(vinyl chloride)', *Phys. Rev. Lett.*, **78**(10), 1928–1931 (1997).
399. J. Colmenero, A. Arbe, A. Alegria, 'Comparative Study of Beta-relaxations in a Glass-forming Polymer (PVC) by Dielectric Spectroscopy and Quasielastic Neutron Scattering', *Physica A*, **201**(1–3), 447–452 (1993).
400. A. Chahid, J. Colmenero, A. Alegria, 'Fast Dynamics Below and Around the Glass-transition in a Side-group Polymer (PVME)', *Physica A*, **201**(1–3), 101–105 (1993).
401. A. Chahid, A. Alegria, J. Colmenero, 'Methyl-group Dynamics in Poly(vinyl methyl ether) – a Rotation Rate Distribution Model', *Macromolecules*, **27**(12), 3282–3288 (1994).
402. J. Colmenero, A. Alegria, J.M. Alberdi, B. Frick, 'Dynamics of the Alpha-process of Polymer Systems on a Microscopic Timescale – Neutron and Nuclear Magnetic Resonance Study', *J. Non-cryst. Solids*, **131**, 949–954 (1991).
403. G. Floudas, J.S. Higgins, 'Quasielastic Neutron Scattering from a Polymer Blend in the 1-Phase and 2-Phase Region', *Physica B*, **182**(4), 361–364 (1992).
404. V. Arrighi, J.S. Higgins, A.N. Burgess, W.S. Howells, 'Rotation of Methyl Side-groups in Polymers – a Fourier-transform Approach to Quasielastic Neutron Scattering.

2. Polymer Blends', *Macromolecules*, **28**(13), 4622–4630 (1995).
405. M.A. Hillmyer, F.S. Bates, 'Influence of Crystallinity on the Morphology of Poly(ethylene oxide) Containing Diblock Copolymers', *Macromol. Symp.*, **117**, 121–130 (1997).
406. I.W. Hamley, J.A. Pople, C. Booth, Y.W. Yang, S.M. King, 'A Small-angle Neutron Scattering Study of Shear-induced Ordering in the Cubic Phase of a Block Copolymer Gel', *Langmuir*, **14**(12), 3182–3186 (1998).
407. A.I. Nakatani, F.A. Morrison, C.L. Jackson, J.F. Douglas, J.W. Mays, M. Muthukumar, C.C. Han, 'Shear-induced Changes in the Order–Disorder Transition Temperature and the Morphology of a Triblock Copolymer', *J. Macromol. Sci. Physi.*, **B35** (3–4), 489–503 (1996).
408. K. Mortensen, K. Almdal, D. Schwahn, H. Frielinghaus, 'Phase Behavior of Diblock Copolymers; Pressure and Temperature Dependence Studied by Small-angle Neutron Scattering', *Macromol. Symp.*, **121**, 245–262 (1997).
409. D. Schwahn, H. Frielinghaus, K. Mortensen, K. Almdal, 'Temperature and Pressure-dependence of the Order-parameter Fluctuations, Conformational Compressibility, and the Phase Diagram of the PEP/PDMS Diblock Copolymer', *Phys. Rev. Lett.*, **77**(15), 3153–3156 (1996).
410. J. Rosedale, F.S. Bates, K. Almdal, K. Mortensen, G.D. Wignall, 'Order and Disorder in Symmetrical Diblock Copolymer Melts', *Macromolecules*, **28**(5), 1429–1443 (1995).
411. W.D. Dozier, P. Thiyagarajan, D.G. Peiffer, M. Rabeony, M.Y. Lin, G. Agrawal, R.P. Wool, 'Small-angle Neutron Scattering and Neutron Reflectometry Study of a Model Graft Copolymer', *Polymer*, **35**(14), 3116–3118 (1994).
412. M.D. Foster, M. Sikka, N. Singh, F.S. Bates, S.K. Satija, C.F. Majkrzak, 'Structure of Symmetrical Polyolefin Block Copolymer Thin-films', *J. Chem. Phys.*, **96**(11), 8605–8615 (1992).
413. A. Karim, N. Singh, M. Sikka, F.S. Bates, W.D. Dozier, G.P. Felcher, 'Ordering in Asymmetric Poly(ethylene–propylene)–Poly(ethylethylene) Diblock Copolymer Thin-films', *J. Chem. Phys.*, **100**(2), 1620–1629 (1994).
414. K. Mortensen, Y. Talmon, B. Gao, J. Kops, 'Structural Properties of Bulk and Aqueous Systems of PEO/PIB/PEO Triblock Copolymers as Studied by Small-angle Neutron Scattering and Cryo-transmission Electron Microscopy', *Macromolecules*, **30**(22), 6764–6770 (1997).
415. K. Mortensen, W. Brown, E. Jorgensen, 'Phase Behavior of Poly(propylene oxide)–Poly(ethylene oxide)–Poly(propylene oxide) Triblock Copolymer Melt and Aqueous Solutions', *Macromolecules*, **27**(20), 5654–5666 (1994).
416. K. Mortensen, W. Brown, K. Almdal, E. Alami, A. Jada, 'Structure of PS/PEO Diblock Copolymers in Solution and the Bulk State Probed Using Dynamic Light Scattering and Small-angle Neutron Scattering and Dynamic Mechanical Measurements', *Langmuir*, **13**(14), 3635–3645 (1997).
417. N.P. Balsara, C.C. Lin, H.J. Dai, R. Krishnamoorti, 'Effect of Saturation on Thermodynamics of Polystyrene–Polyisoprene Block Copolymers', *Macromolecules*, **27**(5), 1216–1220 (1994).
418. C.C. Lin, S.V. Jonnalagadda, P.K. Kesani, H.J. Dai, N.P. Balsara, 'Effect of Molecular Structure on the Thermodynamics of Block Copolymer Melts', *Macromolecules*, **27**(26), 7769–7780 (1994).
419. Y. Matsushita, I. Noda, N. Torikai, 'Morphologies and Domain Sizes of Microphase Separated Structures of Block and Graft Copolymers of Different Types', *Macromol. Symp.*, **124**, 121–133 (1997).
420. R.W. Richards, J.L. Thomason, 'Small-angle Neutron Scattering Study of the Structure of a Triblock Copolymer of Styrene and Isoprene During Extension', *Polymer*, **24**(3), 275–278 (1983).
421. R.W. Richards, G. Welsh, 'Deformation of Matrix Macromolecules in a Uniaxially Extended Styrene–Isoprene–Styrene Linear Triblock Copolymer', *Eur. Polym. J.*, **31**(12), 1197–1206 (1995).
422. Y. Matsushita, Y. Mogi, H. Mukai, J. Watanabe, I. Noda, 'Preparation and Morphology of Multiblock Copolymers of the (AB)n Type', *Polymer*, **35**(2), 246–249 (1994).
423. J. Samseth, K. Mortensen, J.L. Burns, R.J. Spontak, 'Effect of Molecular Architecture on Microstructural Characteristics in Some Polysiloxaneimide Multiblock Copolymers', *J. Appl. Polym. Sci.*, **44**(7), 1245–1256 (1992).
424. N. Koneripalli, N. Singh, R. Levicky, F.S. Bates, P.D. Gallagher, S.K. Satija, 'Confined Block Copolymer Thin-films', *Macromolecules*, **28**(8), 2897–2904 (1995).
425. A.M. Mayes, T.P. Russell, V.R. Deline, S.K. Satija, C.F. Majkrzak, 'Block Copolymer Mixtures as Revealed by Neutron Reflectivity', *Macromolecules*, **27**(25), 7447–7453 (1994).
426. K.F. Freed, J. Dudowicz, 'On the Large Entropic Contribution to the Effective Interaction Parameter of Polystyrene–Poly(methyl methacrylate) Diblock Copolymer Systems', *J. Chem. Phys.*, **97**(3), 2105–2109 (1992).
427. G. Coulon, B. Collin, D. Chatenay, Y. Gallot, 'Kinetics of Growth of Islands and Holes on the Free Surface of Thin Diblock Copolymer Films', *J. Phys. II*, **3**(5), 697–717 (1993).
428. D.G. Walton, G.J. Kellogg, A.M. Mayes, P. Lambooy, T.P. Russell, 'A Free-energy Model for Confined Diblock Copolymers', *Macromolecules*, **27**(21), 6225–6228 (1994).
429. P. Mansky, T.P. Russell, C.J. Hawker, J. Mays, D.C. Cook, S.K. Satija, 'Interfacial Segregation in Disordered Block Copolymers: Effect of Tunable Surface Potentials', *Phys. Rev. Lett.*, **79**(2), 237–240 (1997).

430. V.T. Bartels, V. Abetz, K. Mortensen, M. Stamm, 'Microphase Separation of a Symmetrical Poly(styrene-*b*-paramethylstyrene) Diblock Copolymer', *Europhys. Lett.*, **27**(5), 371–376 (1994).
431. V.T. Bartels, M. Stamm, V. Abetz, K. Mortensen, 'Direct Measurement of Chain Stretching in Diblock Copolymers at the Microphase Separation Transition', *Europhys. Lett.*, **31**(2), 81–86 (1995).
432. V.T. Bartels, M. Stamm, K. Mortensen, 'Change of Phase Behaviour of Diblock Copolymers Upon Application of Pressure', *Polym. Bull.*, **36**(1), 103–110 (1996).
433. G. Hadziioannou, H. Benoit, W. Tang, K. Shull, C.C. Han, 'Neutron Scattering by Multiblock Copolymers of Structure (A–B)(N)–A', *Polymer*, **33**(22), 4677–4681 (1992).
434. N. Koneripalli, R. Levicky, F.S. Bates, J. Ankner, H. Kaiser, S.K. Satija, 'Confinement-induced Morphological Changes in Diblock Copolymer Films', *Langmuir*, **12**(26), 6681–6690 (1996).
435. N. Torikai, Y. Matsushita, I. Noda, A. Karim, S.K. Satija, C.C. Han, 'Studies on the Interfaces of Microphase Separated Structures of Block Copolymers by Neutron Reflectivity', *Physica B*, **213**, 694–696 (1995).
436. N. Torikai, I. Noda, A. Karim, S.K. Satija, C.C. Han, Y. Matsushita, T. Kawakatsu, 'Neutron Reflection Studies on Segment Distribution of Block Chains in Lamellar Microphase Separated Structures', *Macromolecules*, **30**(10), 2907–2914 (1997).
437. N. Koneripalli, R. Levicky, F.S. Bates, M.W. Matsen, S.K. Satija, J. Ankner, H. Kaiser, 'Ordering in Blends of Diblock Copolymers', *Macromolecules*, **31**(11), 3498–3508 (1998).
438. W.H. Dejeu, P. Lambooy, I.W. Hamley, D. Vaknin, J.S. Pedersen, K. Kjaer, R. Seyger, P. Vanhutten, G. Hadziioannou, 'On the Morphology of a Lamellar Triblock Copolymer Film', *J. Phys. II*, **3**(1), 139–146 (1993).
439. Y. Matsushita, M. Nomura, J. Watanabe, Y. Mogi, I. Noda, M. Imai, 'Alternating Lamellar Structure of Triblock Copolymers of the ABA Type', *Macromolecules*, **28**(18), 6007–6013 (1995).
440. J.A. Miller, S.L. Cooper, C.C. Han, G. Pruckmayr, 'Small-angle Neutron Scattering from a Polyurethane Block Copolymer', *Macromolecules*, **17**(5), 1063–1071 (1984).
441. S. Naylor, N.J. Terrill, G.E. Yu, S. Tanodekaew, W. Bras, S.M. King, C. Booth, A.J. Ryan, 'Temperature Dependence of Chain Conformations in a Model Block Copolyurethane', *Polym. Int.*, **44**(3), 371–379 (1997).
442. E. Kim, E.J. Kramer, P.D. Garrett, R.A. Mendelson, W.C. Wu, 'Surface Segregation in Blends of Styrene–Acrylonitrile Copolymers', *Polymer*, **36**(12), 2427–2433 (1995).
443. T. Mansfield, R.S. Stein, R.J. Composto, M.H. Rafailovich, J. Sokolov, S.A. Schwarz, 'Interface Segregation in Copolymer Blends Measured by Neutron Reflection and Secondary Ion Mass Spectrometry', *Physica B*, **173**(1–2), 207–210 (1991).

Nuclear Magnetic Resonance, Imaging of Polymers

Siegfried Hafner
MPI für Polymerforschung, Mainz, Germany

1 Introduction **1**
2 Principles of Nuclear Magnetic Resonance Imaging in Materials **2**
2.1 Nuclear Magnetic Resonance Spectroscopy 2
2.2 Nuclear Magnetic Resonance Imaging 5
2.3 Contrast in Nuclear Magnetic Resonance Imaging of Polymers 8
2.4 Instrumentation for Nuclear Magnetic Resonance Imaging 10
3 Nuclear Magnetic Resonance Imaging in Soft Polymers and Rubbers **10**
3.1 Imaging of Morphologies 11
3.2 Imaging of Material Properties 12
3.3 Imaging of Processes in Polymers 14
3.4 Imaging of Mobile Components in Polymers 15
4 Nuclear Magnetic Resonance Imaging of Rigid Polymers **18**
4.1 Solid-state Imaging Techniques 18
4.2 Stray-field Imaging 20
4.3 Magic-echo Imaging 21
5 Special Techniques **23**
5.1 Surface-scan Techniques 23
5.2 Magnetic Resonance Force Microscopy 23
5.3 Volume-selective Spectroscopy 23
6 Summary and Future Perspectives **24**
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **24**
References **25**

Nuclear magnetic resonance (NMR) imaging is a noninvasive technique which is based on the phenomenon of NMR. The resonance frequencies of the nuclear spins can be made spatially dependent by introducing a magnetic field gradient into the NMR experiment. The acquired NMR signal is thus spatially encoded and a spin-density image of the sample can be reconstructed.

While originally developed for medical applications, NMR imaging is now increasingly applied for materials investigations. The main advantages of the technique are its nondestructivity, which allows processes to be investigated in situ, and the possibility of introducing NMR spectroscopic information and relaxation contrast into the NMR images. Limitations, on the other hand, are the relatively low sensitivity and the broad NMR lines of polymers, both of which limit the spatial resolution. Whereas the latter can be dealt with by suitable methods, the sensitivity ultimately restricts the achievable spatial resolution of conventional NMR imaging techniques to about 10 μm^3. NMR imaging therefore must not be considered as a competitor of other microscopic techniques such as electron microscopy but as a characterization technique that provides supplementary information. Examples illustrating all these aspects are given in this article.

1 INTRODUCTION

Since its discovery 50 years ago, NMR spectroscopy has undergone successful development, resulting in a wealth of applications in various research fields, including chemistry, medicine and material science. One of the most intriguing developments in this field is NMR imaging which has evolved from early test experiments on phantom samples to its present importance as an indispensable tool for medical diagnosis. Its importance derives from several advantages which for many investigations make it superior to other diagnostic techniques. NMR imaging is sensitive to soft tissue and gives an insight into the human body in a completely noninvasive way. Moreover, as a unique advantage, NMR contrast parameters that specifically reflect particular tissue differences can be introduced. A wealth of methods has been developed on this basis, making NMR imaging a rapidly growing field.

In material science, useful applications of NMR imaging have been shown only during the last few years, although the potential of this technique was recognized early. For a long time progress was hampered by the broad NMR lines of materials. Broad lines result in a degradation of the spatial resolution and in case of rigid polymers in a complete failure of conventional NMR imaging techniques. Moreover, the ultimately achievable resolution of around 10 μm^3 is not competitive with that of other microscopy techniques such as light or electron microscopy. As will be presented below, however, such high spatial resolution is not required for many interesting applications in polymer science while the nondestructivity, flexibility and spectroscopic selectivity of NMR imaging are indispensable advantages. Strong efforts have therefore been made to overcome the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

technical and methodological problems of NMR imaging in materials, resulting in a rapid growth of methods and applications in science and in industry.

It is the main intention of this article to give an overview of the methods and possibilities of NMR imaging for polymer characterization, concentrating on some representative examples. Supplementary details can be found in books[1–5] and in earlier reviews on imaging of solids,[6–8] rubbers and other materials.[9–15]

The article is divided into six sections. Section 2 gives an introduction to NMR imaging and discusses the problems and possibilities of its application to materials. The section is divided into sub-sections so that the reader might skip those parts with which he/she is already familiar. Section 3 then discusses applications to soft materials for which standard NMR imaging techniques can be used. Imaging of rigid polymers, which is technically and methodologically more demanding, is described in section 4. The various methodological approaches are briefly reviewed before applications are presented. Section 5 provides a brief overview of techniques which at present must be considered as being somewhat "exotic" but nevertheless show promise for the future. The article concludes with a summary in section 6.

2 PRINCIPLES OF NUCLEAR MAGNETIC RESONANCE IMAGING IN MATERIALS

This section provides the necessary background for the discussion of the imaging techniques and applications in the following sections. Readers without NMR knowledge are nevertheless advised to consult additional literature.[3–5]

2.1 Nuclear Magnetic Resonance Spectroscopy

2.1.1 Basics of Nuclear Magnetic Resonance

NMR is possible for nuclei with a non-vanishing nuclear spin. The most interesting nuclei with respect to NMR imaging of polymers are spin-1/2 nuclei such as protons (^{1}H). For these nuclei there are two possible states in a magnetic field, one with spin-operator component I_z parallel and the other with I_z antiparallel to the static magnetic field direction which by convention is chosen in the z-direction. Whereas the I_z components of the spins are well defined, the transverse components are not and in a semiclassical picture the spins can be considered to precess around the magnetic field $\boldsymbol{B}_0$ with a characteristic resonance frequency, the Larmor frequency [Equation 1]:

$$w_0 = -\gamma|\boldsymbol{B}_0| \tag{1}$$

where γ is the magneto-gyric ratio of the nucleus.

The two states have different energies, and in thermal equilibrium there is a surplus of spins with I_z parallel to the magnetic field so that a net magnetization in the z-direction results. Transitions between the two energy levels are excited by radiofrequency (rf) pulses of the resonance frequency w_0. Since NMR spectroscopy is a coherent technique, a description in terms of energy levels is not sufficient. For most NMR imaging experiments, however, the coherent nature can be taken into account by a simple magnetization-vector picture where the pulses act on the net magnetization $\boldsymbol{M}$ (see Figure 1a and b). The effect of the rf pulses on the magnetization vector is most conveniently described in a new reference frame (x', y', $z = z'$), the so-called rotating frame, which rotates with the angular frequency w_0 around the magnetic-field direction (see Figure 1a). Relative to this frame, the rf field $\boldsymbol{B}_1$ can be considered to be static, pointing for instance, in the $-y'$ direction. It causes the magnetization to precess around its direction analogous to the precession about $\boldsymbol{B}_0$ in the static (laboratory) reference frame. At the end of the rf pulse, the magnetization therefore is tilted with respect to the z-axis by an angle $\theta = -\gamma|\boldsymbol{B}_1|t$, where t is the duration of the pulse. An important flip angle for all NMR experiments is 90° where the magnetization is flipped into the plane transverse to the magnetic field $\boldsymbol{B}_0$. For instance, after a 90° pulse with phase $-y'$ the magnetization vector points in the x' direction of the rotating frame (see Figure 1a). Viewed from the laboratory frame, however, the magnetization is rotating with a frequency w_0 around $\boldsymbol{B}_0$ (Figure 1b). In a coil fixed in the x direction of the laboratory frame, therefore, the magnetic flux varies periodically and a voltage is induced as a function of time which is detected as the NMR signal (Figure 1c). It decays because of relaxation effects (see below) and is called free induction decay (FID). As will be discussed below, the FID contains all of the spectroscopic information, and the NMR spectrum (intensity as a function of frequency) is obtained from it by simple Fourier transformation.

2.1.2 Relaxation

Two main relaxation mechanisms are responsible for the decay of the signal:

- *Longitudinal relaxation (T_1)* describes the re-establishment of the equilibrium magnetization in the z direction after a certain time. The transverse magnetization, on the other hand, is reduced by this process. The energy is dissipated to the so-called lattice, i.e. to degrees of freedom outside of the spin system. In Figure 1(d), the build-up of the z magnetization by longitudinal relaxation is represented by the component $\boldsymbol{M}_z$.

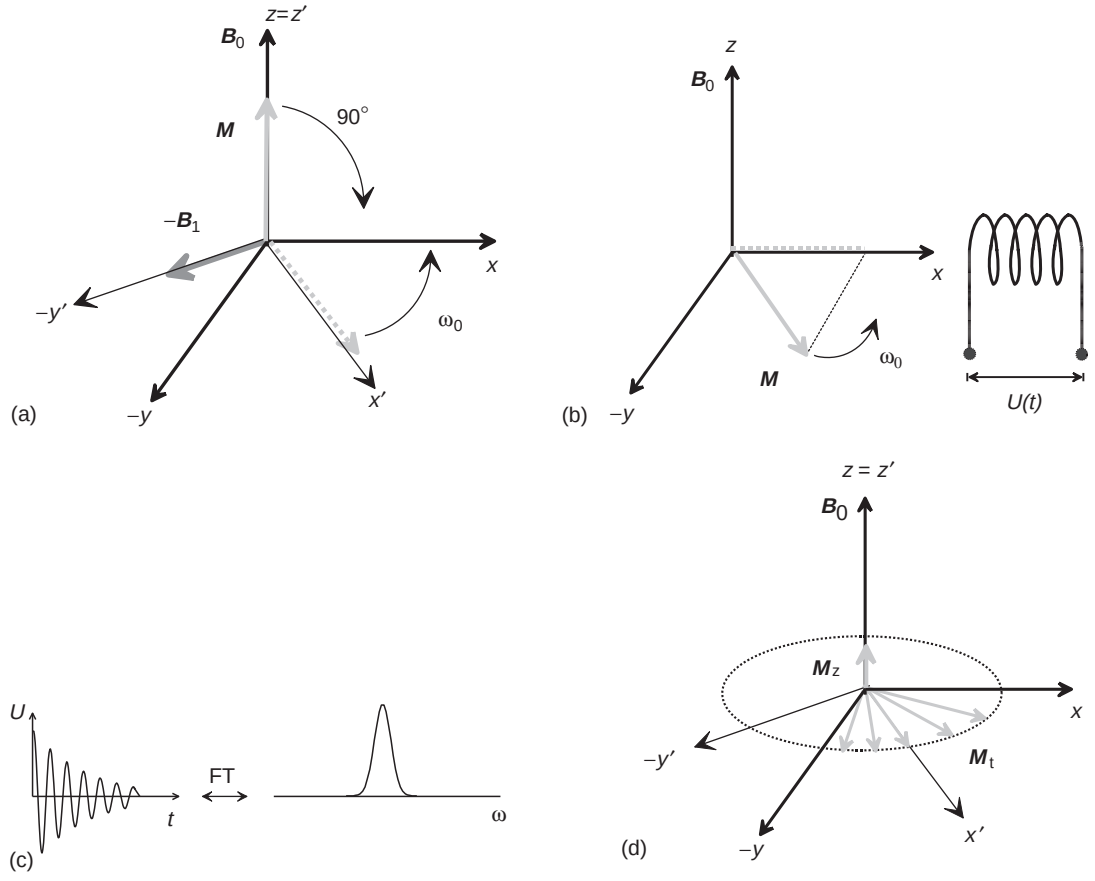

Figure 1 Sketch illustrating the detection of the NMR signal. (a) The net nuclear magnetization $\boldsymbol{M}$ is aligned along the magnetic-field direction $\boldsymbol{B}_0$ (z direction). A resonant rf field is applied perpendicular to it. In a frame rotating with resonance frequency w_0 around $\boldsymbol{B}_0$, this rf field can be represented by a static vector $\boldsymbol{B}_1$ aligned, for instance, along the $-y'$ direction. The magnetization $\boldsymbol{M}$ precesses around $\boldsymbol{B}_1$ in the rotating frame as indicated. For a 90° excitation pulse, $\boldsymbol{B}_1$ is switched off as soon as $\boldsymbol{M}$ is aligned along the x' direction. (b) Viewed from the static frame, the magnetization precesses with frequency w_0 around $\boldsymbol{B}_0$ and thus induces a voltage $U(t)$ in the coil surrounding the sample. (c) The induced voltage as a function of time represents the NMR signal, the free induction decay (FID). The amplitude of the FID is proportional to the nuclear magnetization and the modulation represents the resonance frequencies of the detected nuclei. Fourier transformation provides the NMR spectrum. (d) The signal decay is caused by the loss of coherence (dephasing of the single magnetization vectors) due to field inhomogeneities and spin interactions. Also, the relaxation of the magnetization back to the z direction contributes to some extent to the decay.

- *Transverse relaxation (T_2^*)* describes the loss of phase coherence in the transverse plane (see Figure 1d, components $\boldsymbol{M}_t$). The dephasing results from the distribution of Larmor frequencies due to $\boldsymbol{B}_0$ inhomogeneities and from interactions between spins. It leads to a decrease in the acquired signal because the change of the magnetic flux through the coil is reduced.

In general (and particularly in solid materials), T_2^* relaxation is faster than T_1 relaxation and therefore dominates the FID. Thus, for NMR imaging experiments, longitudinal relaxation is important mainly because the time required for the re-establishment of the z magnetization determines the repetition time of the experiment.

2.1.3 Spin Interactions

As already stated, there are two mechanisms responsible for the loss of phase coherence, magnetic-field inhomogneities and internal spin interactions. The latter dominate in solid-state NMR whereas in liquids the large anisotropic part of the spin interactions is averaged out by fast thermal motions of the molecules and therefore does not contribute much to T_2^* relaxation. The FID of liquids thus is comparatively long. Fourier transformation leads to spectra with narrow lines while in solids broad and often featureless lines are found.

The spin interactions which are important with respect to materials imaging are outlined below.

Chemical shift (anisotropy): The chemical shift results from the shielding of the applied magnetic field by the

electron environment. The shielding causes a slightly different local field at the nucleus and, therefore, a slightly different precession frequency. Since the electron environment depends on the chemical bonds, the measured precession frequency reflects the local chemical environments present in the investigated substance. The chemical shift also depends on the orientation of the molecule with respect to the magnetic field. For powders, noncrystalline or polycrystalline samples, therefore, a broad shift distribution is obtained, the so-called powder pattern. In liquids, motional averaging leads to an isotropic chemical-shift value so that (in the absence of other spin interactions) each chemical site is represented by one line at a characteristic position in the spectrum. The importance of (high-resolution) NMR spectroscopy in chemistry derives mainly from this effect.

Dipolar interaction: The magnetic moments of neighboring spins interact with each other. The resulting dipolar interaction in a strong magnetic field is described in good approximation by the Hamiltonian (given in frequency units), according to Equation (2):

$$H_D = \frac{\mu_0}{4\pi} \sum_{i>j} \frac{\gamma_i \gamma_j \hbar}{r_{ij}^3} \frac{(3\cos^2\theta_{ij} - 1)}{2} (\boldsymbol{I}_i \boldsymbol{I}_j - 3 I_{iz} I_{jz}) \quad (2)$$

where γ_i, γ_j are the magneto-gyric ratios of the spins i, j. The Hamiltonian consists of a spin part which includes the spin operators $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ (and their z-components) and a geometric part which is a function of the magnitude of the internuclear vector $\boldsymbol{r}_{ij}$ connecting the two spins i, j and the angle θ_{ij} relating $\boldsymbol{r}_{ij}$ to the magnetic field. The two parts can be manipulated by suitable pulse sequences or by mechanical rotation of the sample, respectively. The dipolar interaction normally dominates the solid-state spectrum for ^{1}H nuclei and leads to a line broadening of up to around 50 kHz. This has to be compared with a line width of the order of 10 Hz in liquids where the dipolar interaction is removed by motional averaging.

Quadrupolar interaction: This interaction is present only for nuclei ($I > 1/2$) with a nonvanishing quadrupole moment which interacts with the local electric field gradient present in the sample. The quadrupolar interaction often results in very strong line broadening (>100 kHz). For NMR imaging it is important only in special cases, e.g. for deuterated samples.

2.1.4 Spin Echoes

The evolution of spins under the influence of field inhomogeneity and spin interactions can be refocused by suitable pulse techniques thus generating a spin echo. The most important example with widespread applications in the field of imaging is the Hahn spin-echo sequence shown at the top of Figure 2, by which field inhomogeneities and chemical shifts are refocused.

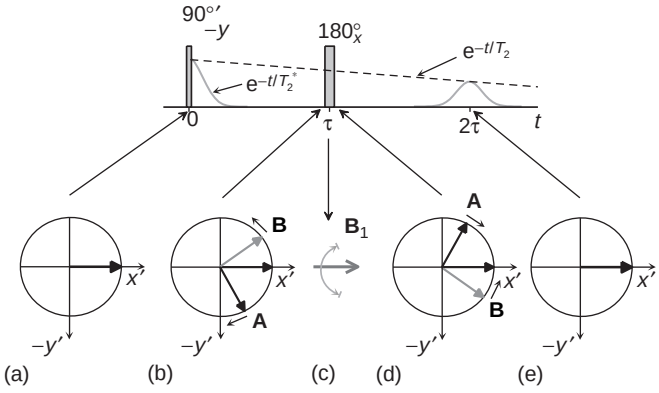

Figure 2 Hahn spin-echo pulse sequence and its explanation with the magnetization-vector model: (a) situation directly after a 90° excitation pulse; (b) the generated transverse magnetization dephases due to field inhomogeneities and chemical shifts (other spin interactions are not considered). The two vectors **A** and **B** represent "slow" and "fast" components, respectively, and the small arrows indicate the rotation directions with respect to the rotating frame. (c) A 180° pulse is applied which acts on the spins as shown. (d) Component **B** and component **A** are interchanged but remain fast and slow, respectively. Thus a refocusing of the original magnetization takes place. (e) After a time 2τ the original state is re-established and an echo signal is obtained (see upper line). The amplitude of the echo is reduced with respect to the original signal by T_2 relaxation, while the FID is governed by the faster T_2^* relaxation.

As already discussed, the 90°_{-y} pulse creates transverse magnetization in the x' direction of the rotating frame (Figure 2a). The magnetization dephases in the following evolution interval τ due to magnetic-field inhomogeneities and chemical shifts (Figure 2b). At the end of this interval, a 180°_x pulse is applied which flips the spins as shown in Figure 2(c). The spins keep their precession frequencies and directions in the second evolution interval (Figure 2d) and, therefore, refocus after a total time 2τ in the x' direction (Figure 2e). As a result, an echo signal is detected with maximum at 2τ.

Such echo sequences are invaluable for NMR spectroscopy and NMR imaging. They are used, for instance, to overcome the receiver deadtime after a strong rf pulse and, even more important, to gain sufficient time for the spatial encoding during the evolution of the spins (see section 2.2.2). In case of solids, Hahn spin echoes are important only if the decay of the signal is dominated by field inhomogeneities and chemical shifts. The dipolar interaction and relaxation effects due to molecular motions are not refocused by the 180° pulse. The amplitude of the Hahn echo is therefore always reduced compared with the original signal. This so-called T_2 *relaxation* of the spin-echo amplitudes as a function of the echo time has to be distinguished from the T_2^* *relaxation* that characterizes the FID (see pulse sequence, Figure 2).

There are also echo sequences that refocus bilinear interactions, such as the dipolar couplings, generating so-called solid echoes. These sequences will be described later in connection with the corresponding imaging experiments.

2.2 Nuclear Magnetic Resonance Imaging

2.2.1 The Principle

In the following, a brief summary of the basic ideas of NMR imaging is given. More detailed introductions can be found, for instance, in the reference list.[3–5]

NMR imaging is based on the application of a well-defined (constant) magnetic field gradient in addition to the homogeneous main magnetic field $\boldsymbol{B}_0 = (0, 0, B_z)$. Since the latter is much stronger than these additional field contributions, it is sufficient to consider to first order only the three gradient components which have a magnetic-field contribution along the z direction, i.e. $G_x = \partial B_z/\partial x, G_y = \partial B_z/\partial y, G_z = \partial B_z/\partial z$. They are called x, y and z gradients, respectively, and are generated by three perpendicular gradient coils.

The principle of spatial encoding is most easily understood by considering the Larmor condition, Equation (1). If a magnetic field gradient, for instance an x gradient G_x, is applied (see Figure 3a), the magnetic field strength becomes spatially dependent, i.e. $B(x) = B_0 + G_x x$. The precession frequency of the spins therefore varies as a function of their position along the x-axis according to Equation (1) [Equation 3]:

$$w(x) = -\gamma B(x) \tag{3}$$

The signal acquired under influence of an x gradient thus corresponds to a spectrum $I(w) = I[w(x)]$ that represents the spatial distribution of the spins, the one-dimensional projection, in the x direction (see Figure 3b). The spatial information in the NMR signal is encoded in form of frequencies and, therefore, the encoding principle is called *frequency encoding*.

In some experiments, however, the spatial information must be acquired in an indirect way using so-called *phase encoding* (see section 2.2.2 for such an experiment). In this case, the gradient is not applied during the signal acquisition but during an evolution interval τ (see Figure 3c). The coherences thus dephase during the evolution time τ with the frequency $w(x)$. A series of experiments is performed by incrementing τ and acquiring the signal (one point!) at the end (see Figure 3c). The signal that represents the image information is thus acquired pointwise in a series of successive experiments and the resulting data set represents a virtual FID which is spatially encoded in the y direction. Fourier transformation provides the spatial profile which is completely equivalent to that obtained by a frequency encoding procedure. For readers familiar with the concept of 2D NMR spectroscopy, it should be noted that the phase encoding procedure corresponds to the indirect-detection principle of 2D NMR spectroscopy. The only difference is that the gradient is under full control of the experimentalist whereas the spin interactions that are probed in spectroscopy are not. Since phase encoding depends only on the product of phase encoding time τ and gradient strength G, usually the gradient amplitude is incremented instead of the time. This excludes sampling of undesired relaxation effects, for instance the loss of phase coherence by the spin interactions.

A third way to use gradients for spatial selection is *slice selection*. For slice selection, a pulse or a series of pulses (pulse train) is applied in the presence of a gradient. Usually the pulses (or the train) are shaped (designed) such that they have a relatively narrow excitation bandwidth. Therefore, they act only on spins which are located at a narrow frequency range within the wide distribution of precession frequencies $w(x)$ generated by the gradient. The frequency selection now translates into a slice that is perpendicular to the gradient direction.

2.2.2 Multi-dimensional Imaging

Based on the encoding principles discussed above, a wealth of imaging methods have been developed,[3–5] which differ mainly in the way that two- or three-dimensional images are generated. The extension of one-dimensional imaging to two or more dimensions is not as trivial as it might appear at first sight. Simultaneous application of frequency encoding gradients in the x and y directions, for instance, does not result in a two-dimensional image but in a one-dimensional profile along the diagonal as a consequence of the vector superposition.

The most important techniques for overcoming this problem are Fourier imaging and back-projection imaging, which are described below. We confine our discussion to two dimensions but extension to three dimensions is straightforward. Moreover, reduction to two dimensions is often possible using slice selection.

Back-projection imaging is based on the acquisition of a series of one-dimensional projections under different gradient orientations from 0° to (usually) 180°. The reconstruction is performed numerically by sophisticated procedures which involve transformation from cylindrical to Cartesian coordinates and filters for the correction of artifacts. The profiles are usually acquired with a Hahn spin-echo sequence to overcome the receiver deadtime. The frequency encoding gradient is preferably switched off during the application of the pulses to ensure that the whole sample is excited. However, sufficiently fast switching of the gradient is often not possible in solids due to the

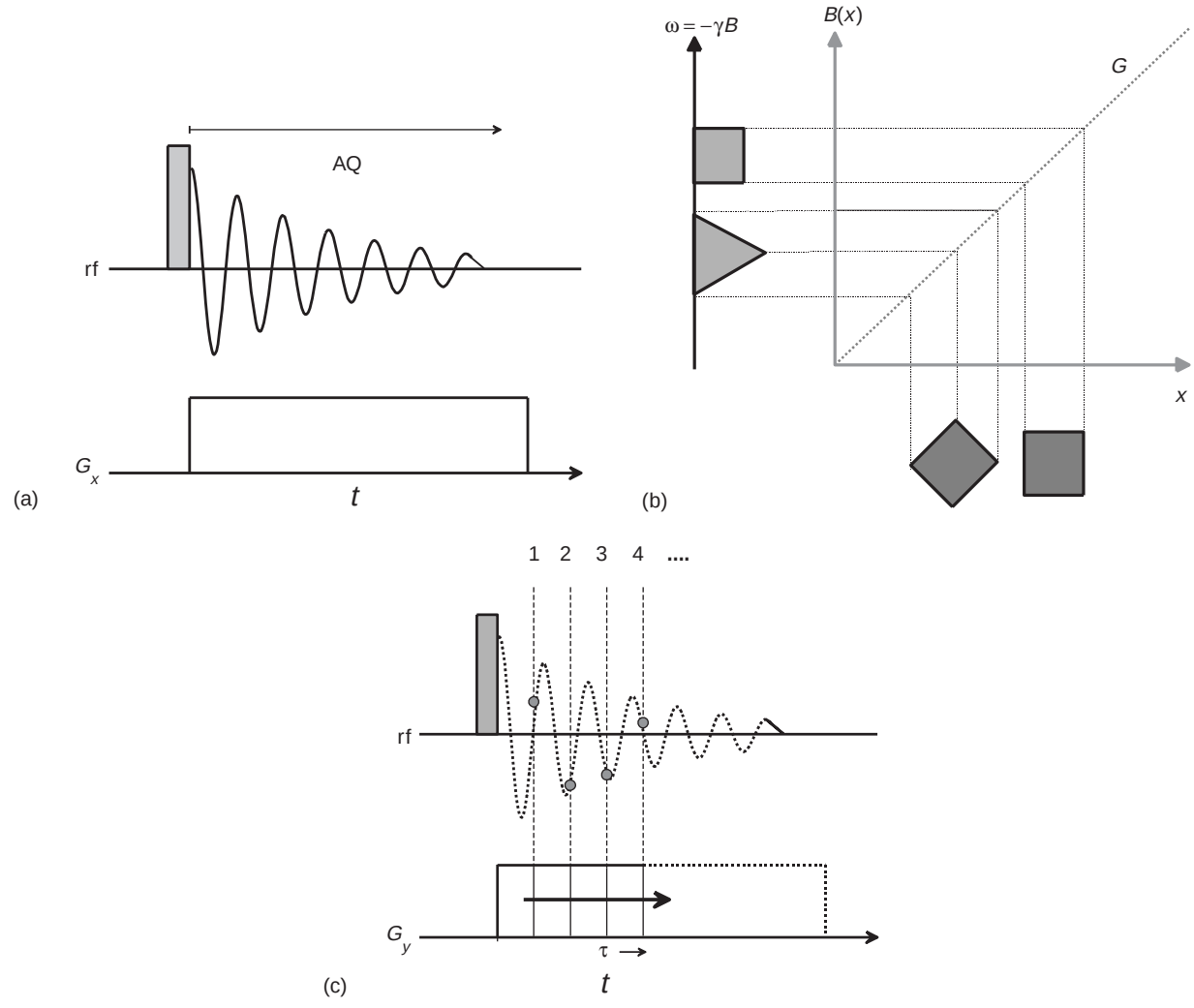

Figure 3 Spatial encoding principles in NMR imaging. (a) Frequency encoding: after a 90° excitation pulse, a constant magnetic field gradient is applied and the magnetic-field strength becomes a function of space, i.e. $B_0 = B_0(x)$. (b) The field gradient causes the precession frequency of the spins in the test sample (dark gray) to become spatially dependent along the gradient direction (here the x direction) according to the Larmor condition $w_0(x) = -\gamma B_0(x)$. The measured NMR spectrum (shaded, bright gray) thus corresponds to a spatial profile of the sample in the gradient direction. (c) Phase encoding: after the excitation pulse, the spins evolve in the presence of a gradient for a time τ which is incremented in a series of experiments. The evolution in the presence of the gradient is sampled at the end of each of the evolution periods in the sequence of experiments. The resulting data set thus is acquired point by point in subsequent experiments (1, 2, 3, 4, ...). It is equivalent to that obtained by direct frequency encoding in one shot.

rapidly decaying signal and the finite rise time of the gradient pulses. In this case, the gradients can be left on during the whole sequence provided that the pulses are broad band enough to excite the entire sample irrespective of the presence of the gradient. This makes back-projection attractive for the imaging of solids while in medical imaging these techniques are nowadays seldom used.

Fourier imaging is based on the combination of frequency encoding (direct detection) and phase encoding (indirect detection). As a typical example, Figure 4 shows a 2D Fourier spin-echo sequence. For frequency encoding, a so-called read gradient in the x direction is applied during the signal acquisition (second line, G_x). An x gradient is also present in the first time interval to compensate for the loss of phase coherence by the read gradient. Full compensation is achieved if the shaded areas under the compensation gradient G_1 and the first part of the read gradient G_2 are the same (see Figure 4). In the y direction a phase gradient is applied during the evolution of the spins in the first τ interval (third line, G_y). The amplitude of the gradient is stepped in successive experiments for phase encoding of the spin echoes. Whereas each single echo is frequency encoded in the x direction, the envelope of the echo series

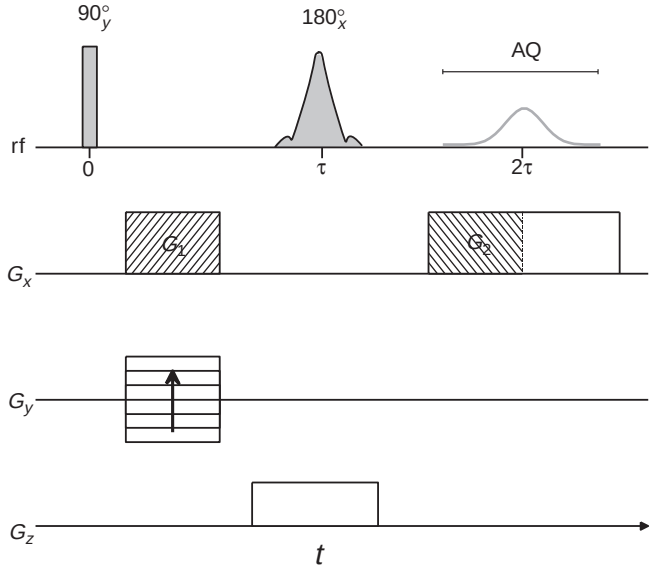

Figure 4 2D Fourier-imaging sequence as a typical example for a two-dimensional imaging technique. It is based on a spin-echo sequence and combines frequency encoding and phase encoding principles (see Figure 3 and text). In the first period of the spin-echo sequence, phase encoding in the y direction takes place while during the acquisition of the echo a frequency encoding gradient in the x direction is applied. In order to avoid undesired dephasing of the coherences by the gradient G_2 before the echo maximum, the gradient G_1 is applied in the first period for compensation (the shaded areas must be equal for full compensation). For the imaging of three-dimensional objects, a slice selection in the z direction can be performed using a frequency-selective 180° pulse applied under influence of a z gradient (see text for further details).

thus contains the information on the phase encoding in the y direction and 2D Fourier transformation directly provides the 2D image. Extension to three dimensions is possible by an additional phase encoding gradient in the z direction. In contrast to frequency encoding gradients, two simultaneously applied phase gradients do not interfere with each other. However, the two gradients must be stepped independently, resulting in a long acquisition time and huge data sets. As an alternative, a slice selection as discussed above can be performed by applying a selective 180° pulse in the presence of a z gradient G_z (see Figure 4).

2.2.3 Spatial Resolution

A detailed discussion of the resolution of NMR imaging was given by Kimmich.[3] Two types of resolution are to be distinguished, *digital resolution* and *physical resolution*. *Digital resolution* or *pixel resolution* can be defined as the separation of two neighboring data points in the frequency domain (which represents the spatial dimension). The frequency interval represented by one data point is $\Delta w_{\mathrm{pix}} = w_{\mathrm{max}}/N$ (where N is the number of data points and w_{max} is the full frequency bandwidth) or, translated in real space, $\Delta x_{\mathrm{pix}} = x_{\mathrm{max}}/N = w_{\mathrm{max}}/(\gamma GN)$, where G is the gradient strength. The pixel resolution thus can be improved by increasing the gradient strength and the number of data points or by decreasing the receiver bandwidth. These possibilities, however, are not independent of each other. For instance, the applicable gradient strength is limited by the available bandwidth into which the profile of the sample has to fit.

More critical than the pixel resolution is the *physical resolution*, particularly in the case of solids. Since NMR is a resonance technique, the physical resolution is determined by the width of the resonance lines, which so far has been neglected. This is well justified in the case of medical imaging where the line widths are only around 100 Hz or less. In polymer materials, however, they extend from several hundred hertz up to around 50 kHz depending on the degree of motional averaging of the anisotropic spin interactions. The measured spatial profile is thus a convolution of the ideal frequency profile representing the image of the sample and the (undesired) intrinsic line width (see Figure 5a and b). The physical resolution in case of frequency encoding is given by Equation (4):

$$\Delta x = \frac{\Delta w}{\gamma G} \tag{4}$$

where Δw is the line width (or the width of the entire spectrum in case of multiple lines). The resolution

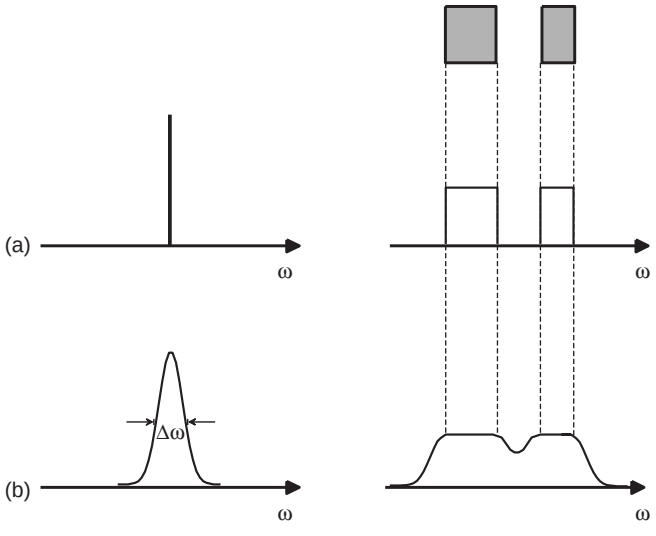

Figure 5 Resolution problem of solid-state NMR imaging. (a) Spectrum (left) and spatial profile (right) of the sample in case of a liquid (see gray-shaded sketch). A well-resolved profile is obtained due to the narrow NMR line. (b) Spectrum and spatial profile in case of a solid. The broad solid-state line is convoluted with the ideal shape of the profile [see (a)] which leads to a strong degradation of the resulting spatial resolution.

therefore increases if the intrinsic line width is reduced or if the applied gradient strength is increased. In contrast to most other spectroscopic techniques, reduction of the line width is possible in NMR since the internal spin Hamiltonians can be manipulated by suitable techniques such as sample spinning or rf pulses. As will be discussed in section 4, however, these techniques are sophisticated and simply increasing the gradient strength is often the best solution, in particular for soft solids such as elastomers.

The final resolution of an NMR image is determined by the most unfavorable of the two, pixel resolution and physical resolution, and they should therefore be well balanced. In practice, the pixel resolution should be set to be slightly better than the achievable physical resolution. Both can be improved simultaneously by increasing the gradient strength. However, by increasing the gradient strength (and hence bandwidth), the total signal intensity is distributed over a larger frequency bandwidth. Correspondingly, the intensity within a given frequency interval decreases and the signal-to-noise ratio (S/N) decreases until finally no signal can be detected. The ultimate resolution limit of a volume is therefore given by the S/N. It can be increased to some extent by accumulating a large number n of experiments according to $S/N \propto \sqrt{n}$. In principle, the resolution is thus limited by the available measuring time.

Since NMR is a relatively insensitive technique, the best achievable resolution in conventional NMR imaging experiments is about $10\,\mu m^3$ so that only macroscopic polymer structures are accessible. As already addressed in the Introduction, however, there are several advantages that compensate for the modest resolution, in particular the noninvasiveness of NMR imaging and the possibility of acquiring spatially resolved spectroscopic information.

2.3 Contrast in Nuclear Magnetic Resonance Imaging of Polymers

NMR spectroscopy and relaxometry provide a variety of contrast parameters for the NMR imaging experiment.[16] They can be conveniently combined with NMR imaging to so-called spectroscopic imaging or parameter-selective imaging. Such techniques enable one to observe macroscopic distributions of microscopic material properties and thus to correlate the different length scales.

In spectroscopic imaging, spatial and spectroscopic information are treated as equivalent dimensions. A typical example is a sequence that uses pure phase encoding in two spatial dimensions while the spectroscopic information is acquired in the direct dimension. Spectroscopic imaging is well established in the medical field but so far has only rarely been applied to bulk polymers, where highly resolved spectra are difficult to obtain.

More attractive with respect to polymer investigations is parameter-selective imaging. NMR parameters such as relaxation times are usually related to polymer properties such as molecular dynamics so that the corresponding contrasts directly reflect sample properties. For parameter weighting, a suitable NMR sequence, a so-called filter sequence, precedes the imaging sequence or is incorporated into it. A well known and often used contrast mechanism is T_2 relaxation, i.e. transverse relaxation during the Hahn spin-echo sequence. This sequence filters out predominantly contributions from immobile parts for which motional averaging of the spin interactions is less effective. The filter strength can be conveniently adjusted by choice of the echo time.

With parameter weighting alone, however, it is not possible to distinguish between spatial variations of the contrast parameter and variations of the spin density. In order to discriminate between the two possibilities, a series of images are usually acquired in which the weighting time, for instance, the spin-echo time, is incremented. In analogy with conventional NMR relaxometry, the relaxation time can then be determined for each of the pixels in the image and an image of the T_2 parameter is obtained.

Other filter sequences commonly used for parameter imaging are T_1, $T_{1\rho}$, $T_{2\mathrm{eff}}$, self-diffusion, flow and spin-diffusion sequences, as outlined below.

T_1 can be measured, for instance, by first saturating the initial z magnetization with a 90° pulse and then observing the build-up of new z magnetization with another 90° pulse as a function of the time between the two pulses. T_1 relaxation is a measure of molecular motions of the order of the Larmor frequency, i.e. with correlation times of around 10^{-8} s.

$T_{1\rho}$ is the relaxation time under influence of an rf field applied along the same direction as the initial transverse magnetization. The pulse sequence for measuring $T_{1\rho}$ thus consists of a 90°_x excitation pulse followed by a rf field applied in the y direction for time t. This so-called spin-lock field prevents the magnetization from being dephased by transverse relaxation. If one acquires the signal as a function of the spin-lock time t, however, a decay is still observed. This $T_{1\rho}$ relaxation is sensitive to motions in the order of the amplitude of the spin-lock field, i.e. in the range up to 150 kHz.

$T_{2\mathrm{eff}}$ is the relaxation time under the influence of a multiple-pulse sequence applied for the refocusing of the dipolar interaction. It is sensitive to motions below 100 kHz (see section 4.3).

Transport properties such as self-diffusion or velocities (in the case of coherent flow) can be determined using a magnetic-field gradient but in a somewhat different way as in imaging. In a typical self-diffusion experiment, for instance, a gradient is applied in the first half of a Hahn

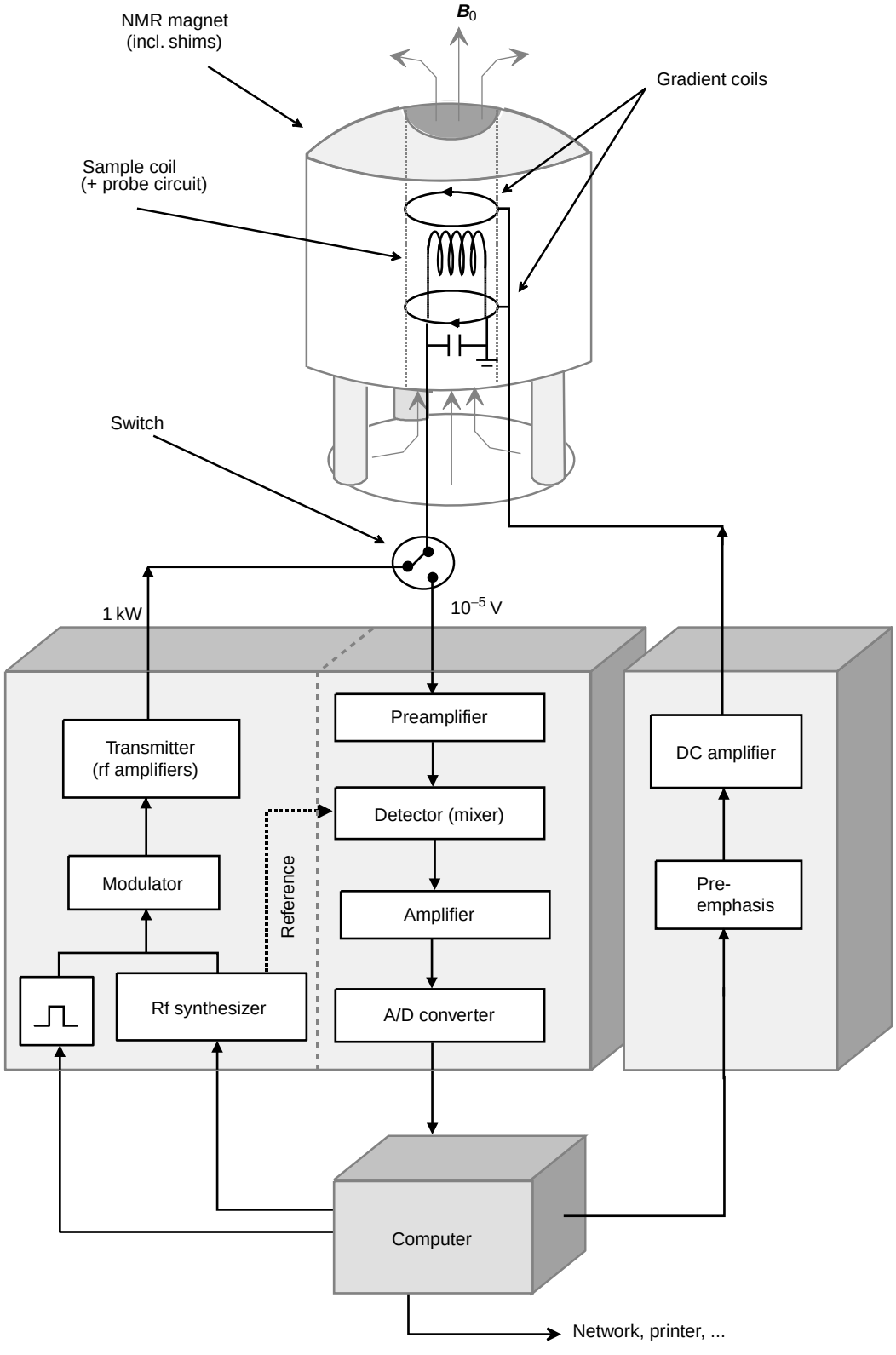

Figure 6 Block diagram of an NMR tomograph. More details on the components and on the flow of information is given in the text.

spin-echo pulse sequence causing the magnetization to dephase. After a refocusing pulse, the same gradient is applied again and leads to a refocusing of the coherences. The refocusing is complete as long as the spins remain at the same place, but if they move to another local magnetic field, full refocusing is prevented and the acquired echo amplitude is correspondingly reduced. From a series of such diffusing-weighted echoes acquired as a function of the gradient strength or the echo time, a decay curve is obtained from which the self-diffusion coefficient can be determined. Similar strategies with gradients can be used to encode coherent motions, i.e. flow velocities.[3]

It should be also noted that some of these contrast parameters are sensitive to external influences such as stress and temperature. This enables one to investigate the influence of such effects on material parameters.

2.4 Instrumentation for Nuclear Magnetic Resonance Imaging

An NMR tomograph is an NMR spectrometer which is supplemented with equipment for generating magnetic-field gradients and with software for image processing. Figure 6 shows a simplified block diagram and a brief discussion of the different parts is given below. More details can be found in several monographs.[1–5]

The *spectrometer computer*, a workstation or a personal computer, is used to control almost all operations of the spectrometer. It is usually integrated in a computer network and most data handling can be performed on external computers.

The *transmitter system* provides the rf pulses. It consists of a synthesizer that produces the rf, a modulator and an amplification system. The modulator generates the rf pulses from the continuous radio frequency and from pulse shapes supplied by a pulse generator and by software. The pulses are amplified to the required power of up to 1 kW by (usually two stages of) amplifiers. In case of double-resonance experiments, two separate channels are required, one for each spin species. The transmitter power is transferred to the probe via a receiver-protection circuit (usually crossed diodes) which ensures that the receiver is decoupled from the intense rf irradiation.

The *probe* consists of an rf circuit with a coil that surrounds the sample and with variable capacitors for tuning the coil to the proper resonance frequency. The sample coil is normally used for both excitation and detection of the NMR signal. In the case of solids, the probe is required to support short pulses, short deadtime, sufficient bandwidth, high sensitivity and a sufficiently large sample volume; a good compromise between these requirements has to be found depending on the problem. Commercial microimaging probes, therefore, use removable inserts which allow the optimal coil for a given sample to be used.

The *receiver system*: the NMR signal is picked up by the coil and transferred via the receiver-protection circuit to a preamplifier. After preamplification, the detector separates the transmitter frequency w_0 from the rest of the signal (this step corresponds to the transformation into the rotating frame as discussed in section 2.1). The resulting low-frequency analog signal then is amplified again and digitized. In order to exploit the full dynamic range of the analog–digital converter, the last amplification step is adjustable. Finally, the digitized signal is stored and can be displayed or evaluated.

The *magnet*, usually a superconducting magnet, is the most expensive part of the spectrometer. For materials imaging normally a field strength between 4.8 T (^{1}H frequency of around 200 MHz) and 9.6 T (400 MHz) is used. The magnet bore is usually between 8.9 and 40 cm. The homogeneity of the magnetic field is improved by shim coils which produce small magnetic fields with various dependences on the spatial coordinates. The currents through these coils are adjusted by maximizing the length of the FID of a water test sample and, thus, minimizing the magnetic-field inhomogeneities. In the case of polymer imaging, a careful shim is not important because of the intrinsically broad lines.

The *gradient system* consists of a gradient control unit, gradient drivers and three coils to generate field gradients in the three orthogonal directions. The gradient strength achieved in commercial microimaging systems is usually up to 1–2 T m^{-1}. Details on the design of gradient coils can be found in the literature.[4,5] The gradients normally are applied in the form of gradient pulses. A critical parameter is the rise time, which is optimized by a so-called pre-emphasis unit.

3 NUCLEAR MAGNETIC RESONANCE IMAGING IN SOFT POLYMERS AND RUBBERS

In this section, NMR imaging of polymers with a moderate NMR line width (up to about 3–5 kHz) and of fluids in polymers is described. A full coverage of the literature is not possible within the limited space of this review, and we confine the discussion to some representative examples of the use of NMR imaging in polymer science. We refer exclusively to polymer work although interesting applications to ceramics (liquids or binders), plants and food have also been published.[1,2] More details on the imaging of soft solids can be found in books[1,2] and in further reviews.[6–14] Applications of NMR imaging in process engineering have also been reviewed.[15]

3.1 Imaging of Morphologies

In spite of the modest resolution of NMR imaging, simple (relaxation-weighted) spin-density imaging of morphologies sometimes already provides useful information. This is the case if there are macroscopic structural variations in the density of the sample or if susceptibility artifacts caused by internal interfaces are present. Susceptibility artifacts often enhance contrasts so that, sometimes, structural details are revealed that are below the spatial resolution. For instance, in recent work on elastomers, interfacial layers that result from folding have been studied by taking advantage of susceptibility effects.[17]

An example with high technical relevance is given in Figure 7(a) and (b) [see also Sarkar and Komoroski[18]] where the NMR images of two pieces of car tire (carbon black-filled rubber) are shown. Both structural details and susceptibility effects are visible. The latter is reflected in the image as white spots and results from an inhomogeneous filler distribution (and from the interfaces). A tire piece with poor dispersion of the filler can be clearly recognized in such images. Similar images of tire pieces reveal layers with different types of rubbers so that the composition of the tire can be studied. A different example of a polymer/filler system with technical relevance is given by materials used for solid rocket propellants. NMR imaging reveals the heterogeneous filler distribution in the different parts of propellant samples,[2,19] and also a thin polymer film around the filler particles is found.

As an example of the imaging of porous polymers, Figure 8 shows the image of a polymer foam which was filled with water. The purpose of this or similar experiments is the determination of the distribution of the pores or channels, which is often not possible otherwise since mechanical slicing with sufficient resolution can destroy the foam structure. With 3D NMR imaging, however, the average pore size and the pore size distribution can be determined completely nondestructively. Another interesting application of 3D imaging to porous organic structures is the coarsening of gelatin foams in situ.[20] The pore size distribution can then be evaluated as a function of time. The results show that the coarsening does not develop in a self-similar way.

It is a well-established strategy in NMR imaging of polymers to use liquids as a probe of sample morphologies if otherwise they cannot be imaged because of broad NMR lines. The image of the foam in Figure 8 is one such example as is the work on swollen rubber samples (see below). In a somewhat complementary approach, a liquid can be also used to increase the molecular mobility of the polymer chains so that the sample can be directly imaged. In such cases, deuterated or fluorinated fluids are typically used so that the fluid does not contribute to the image. A drawback of all these techniques, however, is that the swelling can affect the polymer structure so that the nondestructivity of NMR is sacrificed to some extent.

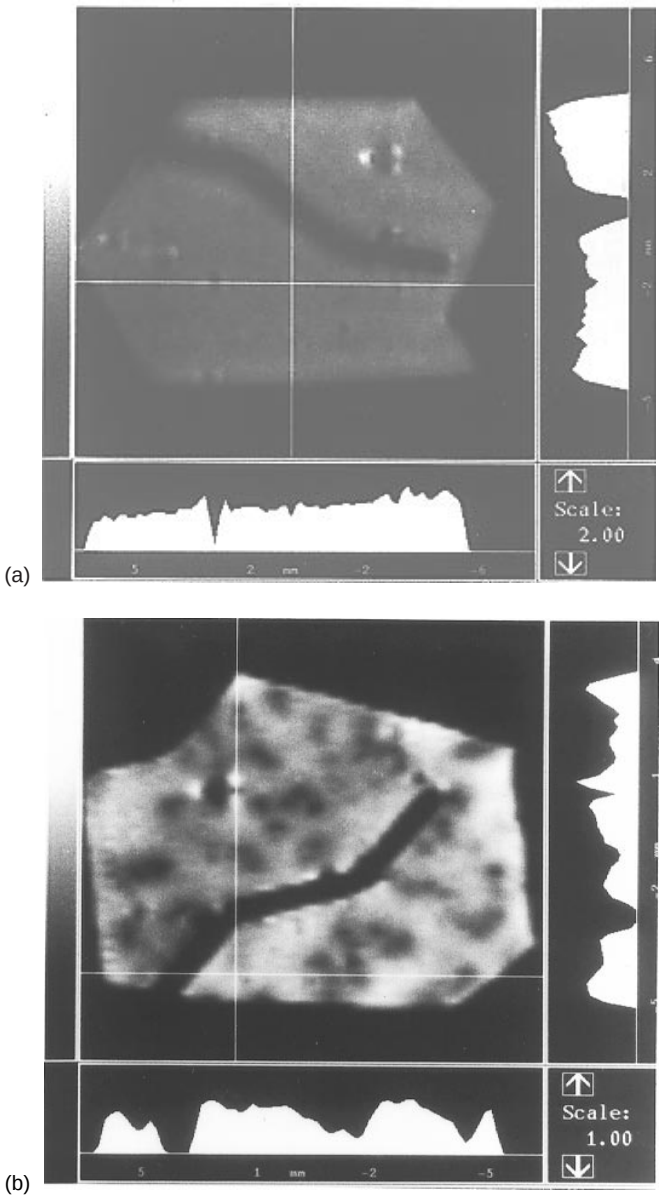

Figure 7 Typical NMR images of tire tread sections, (a) with good dispersion of carbon black and (b) with poor carbon black dispersion. The resolution is 150 × 150 × 350 µm. The spin-echo time was 1.3 ms, the repetition time was 1 s and the total acquisition time of one image was 2 h. Tread grooves are well defined in both sections. The highly uneven appearance in (b) reflects regions of poor dispersion of several hundred micrometers in diameter. [Reprinted with permission from S.N. Sarkar and R.A. Komoroski, *Macromolecules*, **25**, 1420–1426 (1992). Copyright 1992, American Chemical Society.]

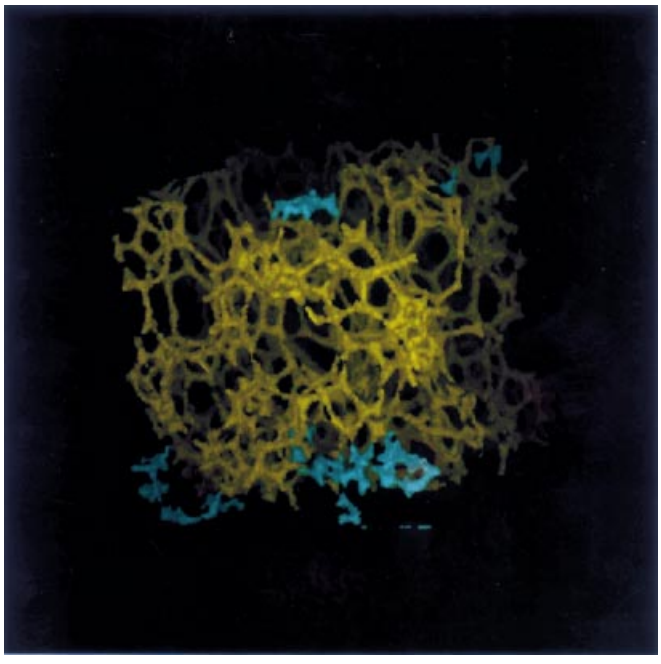

Figure 8 3D spin-echo image of a polyurethane foam with an open-pore structure. Before the image was acquired, the foam was immersed in water until all air was replaced. The spin-echo time of the image was set such that only the water signal could contribute to the image. The image intensity was then inverted, so that the water volumes appear black and the foam structure shows up as high intensity. A 3D surface-reconstruction algorithm was applied to the data set to visualize the foam network. The resolution of the image is $30\,\mu m^3$. (Reproduced by Permission of D. Gross, Bruker Analytik, Rheinstetten.)

3.2 Imaging of Material Properties

As already outlined in section 2.3, there are contrast mechanisms other than spin density in NMR imaging. In particular the various relaxation times and the chemical shift enable one to exploit the full strength of NMR spectroscopy for imaging.

As long as it is sufficient simply to discriminate among different parts of a sample, relaxation-weighted imaging can be performed. The easiest way to do this is to prolong the spin-echo time to increase T_2 contrast. Additional contrast can be obtained by applying the more sophisticated relaxation filters that are discussed in section 2.3 and elsewhere.[1,2] In general, relaxation weighting is a convenient means to study mobility contrast. For instance, the phase separation of a blend of polyisoprene (mobile) and polystyrene (more rigid) was studied showing the increase of the more mobile domains with increasing temperature.[21] Also, the segregation of a stearic acid and paraffin model mixture could be studied using simple spin-echo imaging.

For more quantitative information, however, parameter-selective imaging (see section 2.3) has to be used. Such methods not only provide an image of the corresponding parameters but also the true spin density. In particular, NMR relaxation times are sensitive probes of the molecular dynamics and provide excellent contrast in polymer images. Using T_2 relaxation, for instance, inhomogeneities in rubbers have been investigated, with partial swelling of the samples being used to increase contrast.[22] These inhomogenities were then studied as a function of the cure time.[23] Voids and other inhomogeneities that cause susceptibility differences were found as well as morphological defects such as cross-link density inhomogeneities.

The spatial distribution of the cross-link density was studied extensively in swollen[23,24] and nonswollen samples[25–27] using relaxation-time imaging based on T_2 or $T_{1\rho}$ relaxation times. The basis of the image contrast in these studies is that the molecular mobility decreases with increasing cross-link density and thus the relaxation times that are sensitive to slow chain dynamics also decrease. Spatially inhomogeneous cross-link density distributions can be revealed by such NMR imaging experiments.

Oxidative aging is related to changes in the effective cross-link density and can therefore be studied using the same techniques.[21,25,28–30] Early work showed the thickness of the aged layer as a function of the aging time based on relaxation-weighted spin-density and relaxation-time images.[21,25] More recent studies transform these images directly into cross-link density images (see below).[26–30] An extensive (one-dimensional) investigation of the aging of samples with and without aging protectant was described by Fülber et al.[30]

Stress has been investigated using T_2 relaxation of swollen poly(methacrylic acid)–water gel.[31] It was found that the spatial distribution of the spin density and of the water mobility changes under stress, and the mobility of water in the compressed regions was found to be reduced. Nonswollen polysiloxane rubbers were also studied under stress, and filler inhomogeneities and the stress distribution could be imaged.[32] Stress was also investigated in deuterated samples using a contrast based on the generation of double-quantum coherences so that areas with high stress are reflected by stronger signal intensities.[33] Also, the effect of sample heating by external perturbation has been studied using T_2 as a contrast parameter (see section 3.3).

Material scientists, however, are normally not very interested in images of NMR parameters, the interpretation of which often requires a profound knowledge of NMR. Rather, they are interested in the implications of such images with respect to material properties. In early work, therefore, a semiquantitative interpretation was given for T_2 images of swollen samples.[24,31] Following

this line, the measured NMR parameter images can also be directly transformed into images of the corresponding material parameters.[27–32]

The conceptually simplest way to do this is to perform a calibration measurement on a well-characterized series of samples where the investigated material property is varied. This has been exemplified by determining the stress distribution in stretched elastomer samples.[32] For calibration, the transverse relaxation time was measured as a function of the applied stress. The T_2 image of a sample in which a nonuniform stress distribution was induced by a cut could thus be transformed into a stress image (see Figure 9a–c).

Such measurements require a series of calibration samples, which often are not available. In order to avoid calibration measurements, models that relate NMR parameters and material parameters can be used to derive directly the material-property image from the measured NMR-parameter image.[26,27] This approach has been tested on differently cross-linked rubber samples using T_2 and $T_{1\rho}$ imaging sequences. The resulting images have been interpreted using a single-chain model for the T_2 data and a defect-diffusion model for $T_{1\rho}(w)$. Both provide parameters that correspond to the cross-link density and correlation times that characterize slow chain dynamics.[26,27]

As an example, Figure 10(a) and (b) show the cross-link density image of a sample consisting of three pieces of rubber with different cross-link densities. While the different cross-link densities cannot be distinguished in the normal spin-density image (not shown), they are readily distinguishable in the material-property image. The distribution of the cross-link density was analyzed using the corresponding histogram (Figure 10b). Similar investigations were also made for filled rubber samples and for oxidatively aged rubbers,[28–30] where the different aging conditions of the sample parts could be well discriminated by different cross-link densities. Other material properties may be imaged likewise by the use of suitable theories to relate them to measurable NMR parameters. It is expected that this principle will be applicable to a wide variety of problems in the future, building a bridge between NMR spectroscopists and material scientists.

Spectroscopic imaging has seldom been applied to polymers so far because broad lines prevent a satisfactory resolution in most cases. An example where it was nevertheless used was reported by Traub et al.[34] In this work, the lamellar orientation in a sheared block copolymer was investigated in a spatially resolved way. The shear geometry was such that the applied shear amplitude increased along the piece of sample that was used for imaging. From the acquired spectroscopic image it was found that the lamellar orientation changes as a function of the shear amplitude. In another example,[33] the quadrupole splitting of drawn deuterated samples was used as a contrast parameter. By spectroscopic imaging the line splitting in the different areas of the drawn sample could be determined as a function of position and thus a stress map could be obtained. Spectroscopic imaging

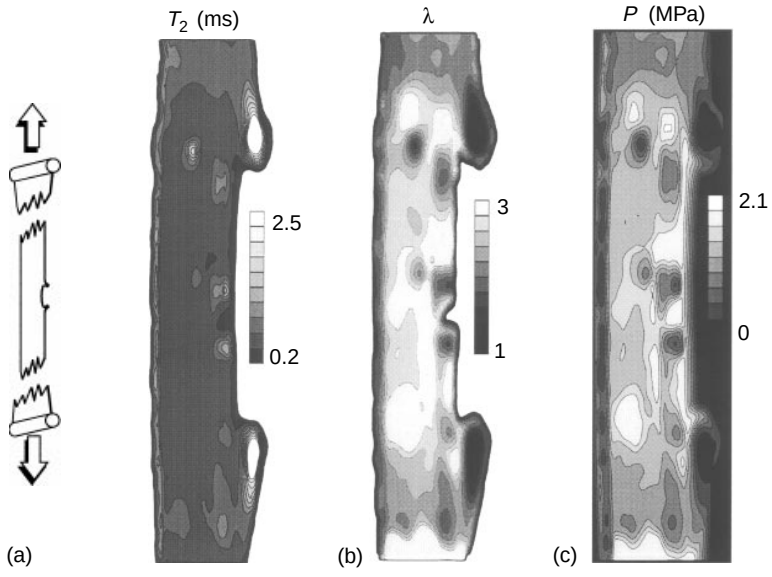

Figure 9 Imaging of a rubber sample drawn at $\lambda = 2$ as shown in the sketch. (a) T_2-image as determined from a series of relaxation-weighted images; (b) "strain" map (parameter λ) of the sample obtained from a calibration curve (the contour levels mark changes from $\lambda = 0$ to $\lambda = 3$ in increments of 0.25); (c) "stress" image after re-calibrating with the stress–strain curve of the sample. The stress varies from 0 to 2.1 MPa in increments of 0.3 MPa (adapted from reference 32 with permission).

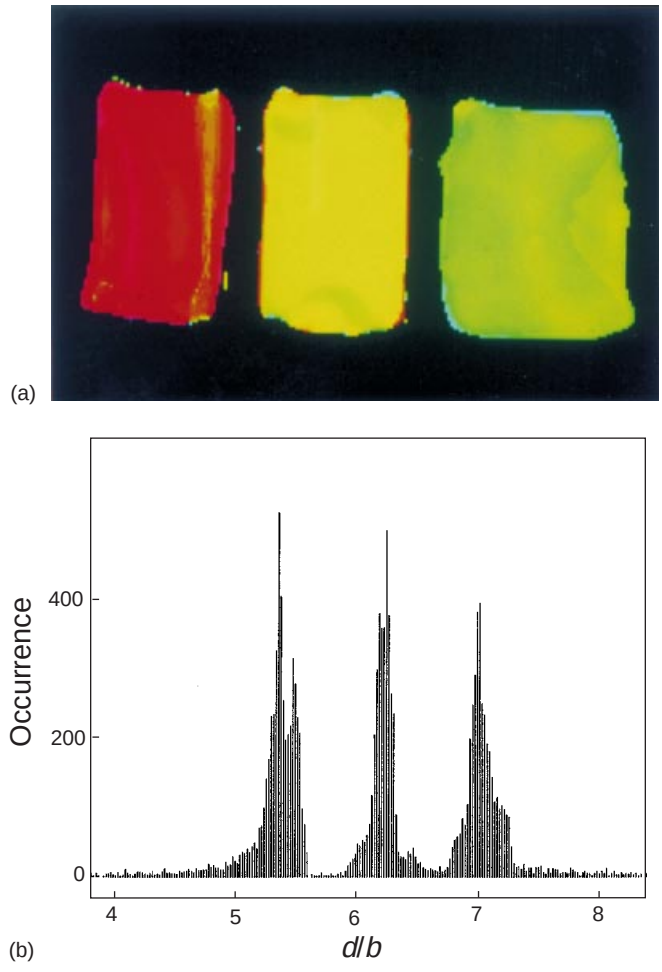

Figure 10 (a) Cross-link density image of a test sample of three differently cross-linked pieces of rubber evaluated from a series of $T_{1\rho}$ relaxation-time images. Displayed is the parameter d/b which corresponds to the cross-link density in a defect diffusion model (d is the diffusion length between two cross-links, b the width of the defect). (b) Histogram analysis of the d/b distribution of image (a). The three peaks correspond to the different cross-link densities of the three rubber pieces and their widths reflect the cross-link density inhomogeneity of each of them.

has also been applied to more rigid samples, and the corresponding experiments are discussed in section 4.

3.3 Imaging of Processes in Polymers

One of the most fascinating aspects of NMR imaging is that it allows the in situ monitoring of processes in a completely noninvasive way as long as the process takes place within a reasonable time window for the NMR experiment. Examples of processes that have been investigated by NMR imaging are the coarsening of a gelatin foam (see section 3.1), polymerization, vulcanization or cross-linking, aging and the ingress of solvents in polymers (which will be discussed in section 3.4). Some of these investigations are performed on a series of samples which represent various stages of an on-going process, but many are actual in situ investigations.

There have been several studies of the spatial aspects of the polymerization process by NMR imaging. For instance, an investigation of the photopolymerization of hexamethylene dimethacrylate as a function of the photoinitiator concentration has been reported.[35] Also, a series of samples with a different light-exposure time were prepared and studied. Chemical shift resolved imaging techniques were used in this work, but the progress of polymerization was monitored by T_2 contrast reflecting the reduction in molecular mobility due to polymerization. It was found that the transition between polymerized and nonpolymerized regions becomes sharper with increasing exposure time.

The polymerization of methyl methacrylate was studied in situ for two samples with different accelerator contents.[36,37] The progress of polymerization was again monitored by relaxation contrast, in particular T_2 contrast. The image intensities contain qualitative information on the reaction rates, the viscosity and the extent of polymerization in the various parts of the bulk sample.

Using similar techniques, methacrylic acid polymerization was studied together with the progress of a wave front that results from the coupling of molecular diffusion and the chemical reaction.[38] The wave-front reaction zone was found to approximate a δ-function. After an induction period where the polymer front proceeds slowly, the reaction suddenly accelerates and settles down into a constant-velocity wave.

Cross-linking was studied in situ for the case of sulfur vulcanization.[39] A vulcanization device was constructed that could be used in a standard microimaging probe. The field inhomogeneities introduced by the device were dealt with by using a spin-echo sequence and pure phase encoding. Only 1D profiles were acquired to allow the vulcanization process to be followed with a reasonable time resolution. As in many other cases, the experiment could be designed such that 1D profiling is sufficient to obtain the relevant information. Figure 11(a) and (b) show the geometry of the arrangement and the resulting profiles as a function of time after the start of vulcanization. More details on the results of the experiment are given in the caption.

The curing of a continuous carbon fiber-reinforced epoxy resin was studied in situ with 2D imaging.[40] It was started by raising the temperature to the appropriate value and the increase in viscosity as a function of time was monitored via T_2 changes. The curing process was found to be inhomogeneous starting at the sample edges.

Another class of processes that can be studied in situ are those caused by external time-dependent perturbations.

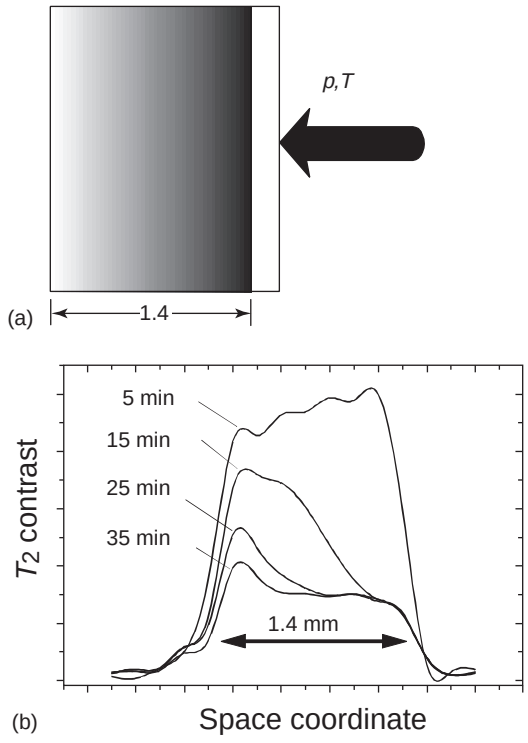

Figure 11 1D in situ imaging of the vulcanization process. An elastomer sample is vulcanized in the NMR spectrometer from the right to the left by applying heat and pressure from this side (see sketch). The profiles are acquired as a function of vulcanization time using a T_2-relaxation-weighted imaging sequence. The first profile (after 5 min) is somewhat larger on the right side as a consequence of the higher temperature and corresponds to the situation before the start of the vulcanization. As soon as the vulcanization takes place, T_2 becomes shorter and thus the signal intensity in the corresponding parts of the sample is reduced. This provides a good means to study the progress of vulcanization as a function of time. [Reproduced by permission from Fülber et al.[39]]

An example is the heating of a filled rubber sample under periodic mechanical load which could be imaged as a function of the shear frequency and the filler content.[41] The image data were acquired synchronized at a given phase of the sine-like mechanical perturbation.

For studies where 1D profiling is not sufficient, fast imaging techniques[3–5] are often useful. These techniques are based on the generation of so-called gradient echoes by reversing the gradient direction. Gradient echoes, however, are limited by T_2^* and are therefore applicable only to relatively mobile systems. An example of the use of fast imaging experiments in material science is the study of the local concentration of suspended particles during a falling-ball experiment.[42] The experiment shows that the concentration of the suspension is higher ahead of the falling ball and that the depleted zone behind the ball never fully recovers in the case of larger ball diameters. The downward velocity was found to decrease as the ball falls owing to the accumulation of suspended particles in front of the ball.

For less mobile cases, a simple one-pulse imaging sequence with back-projection reconstruction has been proposed.[43] As in other fast imaging techniques, a low-angle excitation pulse is applied so that the repetition time can be made very short, of the order of milliseconds. Apart from in situ monitoring of fast processes, such sequences can also be applied for material testing purposes where often it is necessary to have full three-dimensional spatial resolution within a reasonable time. For instance, the sequence described above was first tested on a jellybaby (''gummy bear'') which has a line width of about 2 kHz. A 3D image with 64^3 pixels could be acquired within about a 2 min total acquisition time, in contrast to the hours required for conventional techniques.

3.4 Imaging of Mobile Components in Polymers

One of the strengths of NMR imaging is that it is selective with respect to mobility. The use of fluids for NMR imaging of materials is manifold, and below we discuss a selection of the relevant work.

3.4.1 Distribution of Liquids in Polymers

The distribution of liquids in a polymer can be used as a probe of sample properties. For instance, Figure 12(a) shows an NMR image of the water distribution in the polymer matrix of a microchip which was saturated in a water steam atmosphere at about 390 K for 168 h.[44] The water was found to be distributed homogeneously within the sample. Figure 12(b) shows a slightly different cross-section of the same chip after drying for 11 h. The image reveals that the water distribution now is inhomogenous with little water in the center part. On comparing the two images, it seems that the water does not leave the chip through the surface but through small gaps around the connectors.

Microscopic or mesoscopic properties can be also studied by imaging the distribution of a liquid in a polymer. In early work in this field, the heterogeneous water distribution in polymer composites was shown by 2D NMR imaging.[45] As another example of a heterogeneous water distribution, Figure 12(c) shows the water content of a linoleum sample after 24 h exposure at room temperature. The white spots in the image reflect accumulations of wood fibers whose water uptake is much faster than that of the surrounding polymer matrix. Also, a ''swelling front'' is visible in the image and its progress can be studied as a function of time.

Swelling of a polymer by a solvent is not always motivated by the study of the solvent uptake or the solvent distribution. Nonprotonated solvents are often used to improve the S/N of the image or to enlarge

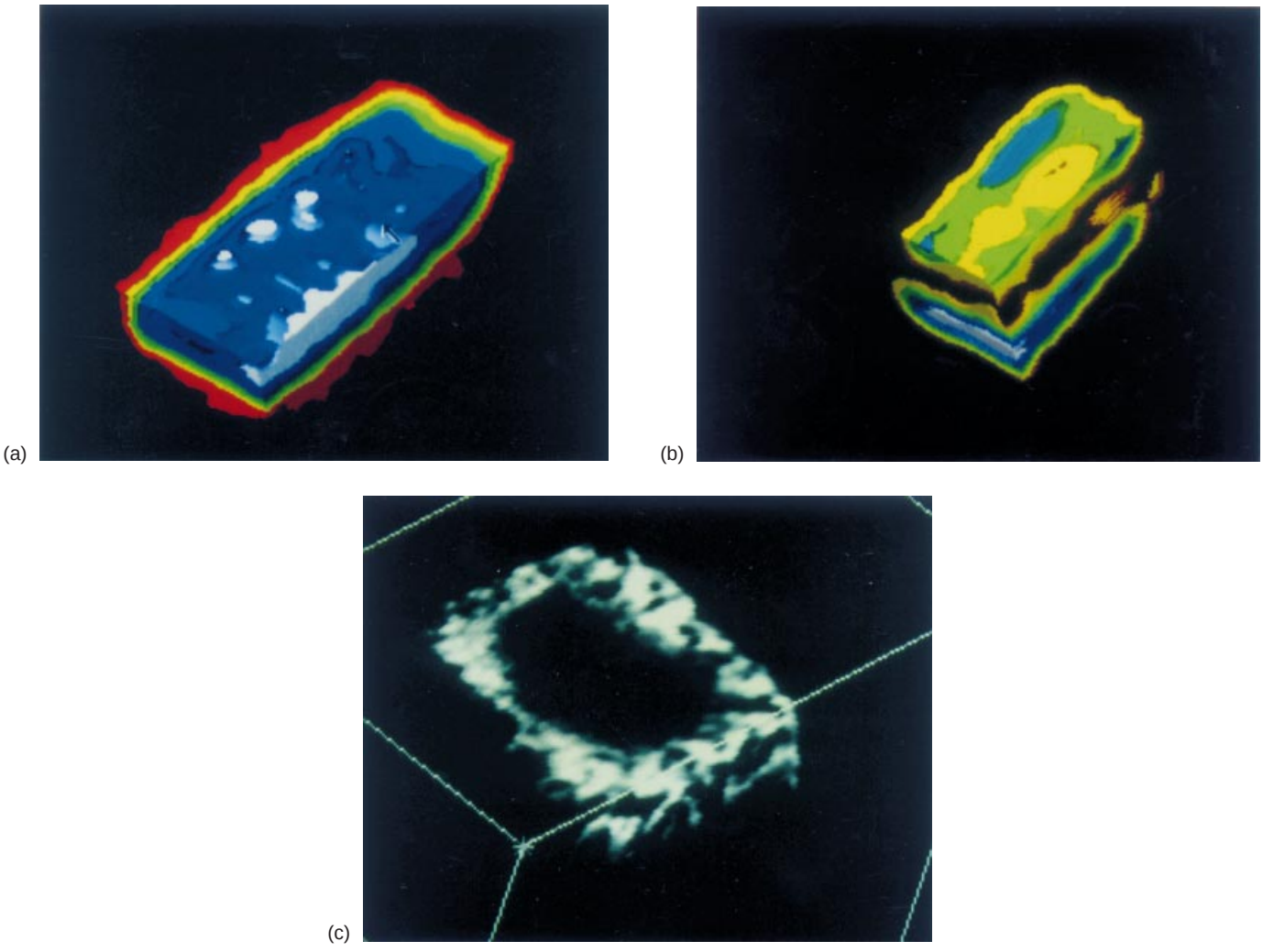

Figure 12 Images of liquids in a polymer matrix. (a) Image of a silicon microchip saturated with water (3D cross-section). The water is distributed relatively homogeneously within the polymer matrix of the chip (color code from high to low signal: white, blue, green, yellow, red, black). (b) Cross-section of the same sample after drying for 11 h. The water is removed predominantly from the center part. It obviously leaves the chip through small gaps around the connectors [the images are from the investigation described by Hafner and Kuhn[44]]. (c) Water distribution in a piece of linoleum swollen for 24 h at 25 °C. The spots reflect regions with a high concentration of wood fibers. The swelling front can be followed in a series of such images.

structural details by taking advantage of the increased molecular mobility of chains in a swollen state. For instance, the spatial heterogeneity of the molecular mobility of the competitive vulcanization of rubber and polybutadiene blends has been studied using swollen samples.[46] The resulting images showed heterogeneous network formation, which probably is a consequence of the different solubilities of the curatives in the two components.

Not only liquids can be used as probes for material properties but also gases, in particular for the investigation of (internal) surfaces. The use of spin-polarized gases allows one to overcome the S/N problem. An example of such investigations is the imaging of aerogels using ^{129}Xe which was polarized by optical pumping.[47] The application of this technique to polymers was also discussed.

3.4.2 Studies of the Ingress of a Solvent into Polymers

The time dependence of solvent fronts has been studied numerous times for many polymers and solvents.[48–54] Again, we restrict ourselves to representative examples.

In most investigations, the signal of the polymer matrix is filtered out by a spin-echo experiment which takes advantage of the longer T_2 relaxation time of the solvent. The progress of a solvent front can thus be followed easily. In studies of rubbery polymers,[48,49] a diffusion-controlled process was found, that is, normal Fickian

diffusion (case I). Studies of glassy polymers, on the other hand, exhibit case II diffusion.[50] The latter is characterized by a linear uptake of the solvent with time and a sharp front which progresses with constant velocity. In a careful temperature-dependent study it was shown that a change in the transport kinetics from case II (low temperature) to case I takes place at around the effective glass-transition temperature of the system.[51]

Multiple-component diffusion (two solvents) has also been investigated.[48,49] Using the chemical shift to discriminate between the two solvents, the ingress of each of the two components can be studied separately.[48] As an alternative, one of the solvents can also be deuterated.[49] In these investigations it was found that, depending on the mixture and type of the two solvents, they either diffuse at nearly the same rate or two separate diffusion fronts form.

The progress of a solvent front can also be studied indirectly using nonprotonated solvents,[52] while the ^{1}H image of the polymer is acquired. The progress of the solvent is then monitored by its effect on the chain mobility. For this, the echo time is adjusted such that the signal from the nonswollen polymer is suppressed whereas the signal from the swollen parts is preserved because of its higher molecular mobility. This not only allows the study of the progress of a solvent front but also gives direct insight into the changes in molecular mobility of the polymer. Moreover, T_1 relaxation time profiles can also be acquired as a measure of segmental motion and have shown that even after 2 weeks a uniform dynamic equilibrium state has not been reached in a polystyrene sample.[52]

Such studies of the ingress of a solvent have also been made in situ.[52,53] This is particularly simple if nonprotonated solvents are used but discrimination between a protonated solvent and the polymer is also possible if one exploits the differences in relaxation time.[53] In studies with nonprotonated solvents, the polymer and the solvent can be imaged separately. As an example of the latter case, ^{19}F images of a fluorinated solvent where acquired in a polymer.[54] In addition to the investigation of solvent ingress or the swelling of polymers, polymer dissolution can be studied for noncross-linked samples. Such studies have important applications in the investigation of drug release by dissolution of a polymer coating or tablet.[55,56]

3.4.3 Imaging of Transport Properties

In the previous section, transport properties were derived from images of a progressing solvent front. As already discussed in section 2.3, however, there are also NMR techniques for the direct measurement of transport properties such as self-diffusion or flow (i.e. velocities and acceleration).[3,4] These techniques for the measurement of transport parameters can be easily combined with NMR imaging by introducing the corresponding contrast filters, for instance, the self-diffusion filter described in section 2.3. From the pixelwise evaluation of the resulting self-diffusion-weighted imaging series, a self-diffusion-constant image is then obtained.

In such experiments, usually only the self-diffusion constants of mobile components can be determined since the self-diffusion of a polymer chain is too slow to be measurable with the comparatively weak imaging gradients. Polymer properties can therefore be investigated only indirectly using the self-diffusion of the solvent as a probe. The anisotropy of self-diffusion and the restrictions felt by the diffusing molecules provide valuable information on mesoscopic structures in materials. As an example, Figure 13 shows the self-diffusion image of water in a superabsorber particle which consists of a hard shell (highly cross-linked) and a soft core.[57] In the self-diffusion-constant image, the outer part is characterized by lower self-diffusion constants, indicating reduced translational mobility of the water. Further examples of self-diffusion imaging can be found in the literature.[3,4,50]

In a similar way, coherent motion can be also imaged.[58] For instance, the shear of polymer solutions has been investigated by flow imaging (velocity profiles).[59] Transitions from Poiseuille flow to power-law

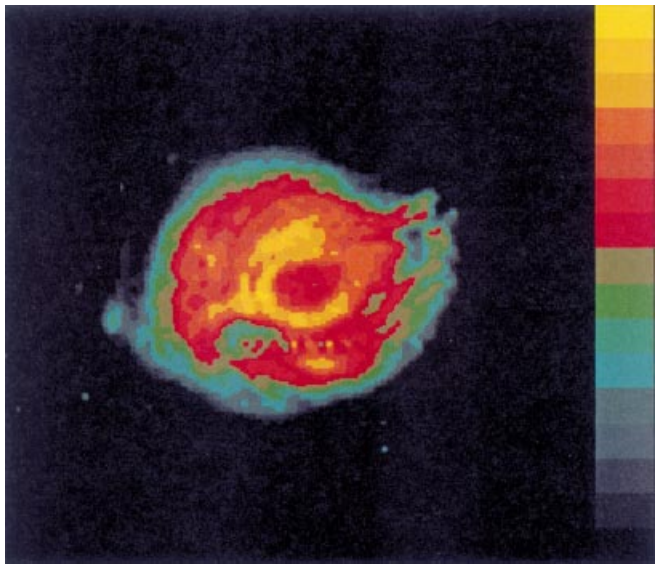

Figure 13 Self-diffusion image of the water in a fully swollen superabsorber particle. The self-diffusion constant in the shell of the particle is reduced by a factor of about two. It reflects the fact that the outer part of the sample has been additionally cross-linked for mechanical stability. The process that is responsible for the reduction in translational mobility of the water is currently being investigated. [Reproduced by permission from Traub.[57]]

shear thinning have been studied as a function of the polyethylene oxide concentration.[59] In a semidilute solution of high molecular mass polystyrene, strong shear thinning (non-Newtonian behavior) was found and could be interpreted by using a power-law fluid model exhibiting M^3 molecular mass scaling.[60] Flow imaging was also applied to solid–liquid suspensions, which are models for rheological characteristics found during polymer processing.[61] Moreover, the flow of water through mixers and extruders has been monitored by NMR imaging, providing an indirect means to study the geometries used for the extrusion of polymers.[62]

4 NUCLEAR MAGNETIC RESONANCE IMAGING OF RIGID POLYMERS

4.1 Solid-state Imaging Techniques

In the NMR imaging of rigid solids, most efforts have focused on methodological developments to overcome the very broad NMR lines. Not only is the spatial resolution reduced in rigid solids, but also conventional techniques such as the Hahn spin-echo fail and solid-state techniques[63] must be applied. It was therefore only recently that solid-state imaging has reached a state-of-the-art where more advanced techniques can be applied to questions relevant for material science. Because solid-state imaging is still at the stage of development, we will first give a brief overview of the techniques before discussing applications in the following sections. More details can be found in reviews.[6–8,10–14]

4.1.1 Techniques Without Line Narrowing

As already discussed in section 2.2.3, spatial resolution can be increased by using extremely strong gradients or by application of line narrowing techniques. In this section we discuss the first case, that is, techniques that rely on the strength of spatial encoding by the gradient. As a consequence, the total signal intensity is distributed over a large frequency bandwidth so that the S/N is particularly critical in this approach.

4.1.1.1 Strong Gradients and Stray-field Imaging The straightforward approach to increase spatial resolution is the application of an extremely strong gradient. In stray-field imaging,[64–68] the strong magnetic-field gradient in the fringe field of a superconducting magnet is used for this purpose. Stray-field imaging is comparatively simple and therefore attractive with respect to applications. It will be discussed in more detail in section 4.2. As an alternative to such extremely strong static-field gradients, designs for very large pulsed-field gradients have also been realized, and their use has been demonstrated on phantom samples and on plant stems.[69]

4.1.1.2 Multiple-quantum Imaging This approach makes use of coherences that involve correlated flips of two or more spins, so-called multiple-quantum coherences.[70] What makes them interesting for solid-state NMR imaging is that an n-quantum coherence is influenced n times more by an applied gradient than a single-quantum coherence. The spatial resolution should therefore in principle be correspondingly higher.[70] Although conceptually interesting, the technique is so sophisticated that it is not of much practical relevance for ^{1}H imaging in solids so far.

4.1.2 Techniques Based on Line Narrowing

Problems related to the use of extremely strong gradients can be avoided by line narrowing techniques. A trivial way to narrow lines is to raise the sample temperature.[37,71] The line narrowing achieved can be considerable if the sample is heated above its glass-transition temperature. However, other sample properties also change dramatically at this point so that the noninvasiveness of NMR imaging is sacrificed to some extent. The less invasive alternative for reducing the line widths is the use of well-known solid-state line narrowing techniques. Several such line narrowing techniques are presently in use,[63] and all of them have been already applied for the imaging of rigid solids.[6–8,10,13]

4.1.2.1 Imaging with Magic-angle Spinning[72–75] The geometric part of anisotropic spin interactions can be manipulated by rapid sample spinning.[63] The dipolar interaction, for instance, depends on the angle θ_{ij} of the internuclear vector $\boldsymbol{r}_{ij}$ with respect to the magnetic field [Equation 2]. In a powder, all angles θ_{ij} are present and thus a broad line results. For the magic-angle spinning (MAS) experiment, the sample is rotated rapidly so that all interaction components perpendicular to the rotation axis are averaged out for full rotor periods. Choosing the rotation axis along the so-called magic angle of 54.7°, the dipolar interaction vanishes and so does the corresponding line broadening.

MAS line narrowing can be combined with imaging techniques using gradients that are (electronically) rotated in synchrony with the spinning sample.[72–75] In practice, however, ^{1}H MAS imaging is confined to rubbers and other relatively weak solids since the rotor frequencies achieved so far in imaging experiments are not sufficient for efficient line narrowing in rigid solids. Imaging with MAS line narrowing is therefore most useful in combination with other line narrowing techniques or for the imaging of dilute spins such as ^{13}C (see below).

Further drawbacks of the technique are the restriction to small, preferably axially symmetric, samples and possible sample deformation. On the other hand, sample deformation of polymers under centrifugal forces can be studied conveniently by MAS imaging, as was exemplified with a polyisoprene sample.[73] Blends of polystyrene and polybutadiene were also investigated in this work using MAS imaging.

4.1.2.2 Imaging with the Effective Field Along the Magic Angle[76,77] This technique is the spin-space analog to MAS. By applying a suitable off-resonance rf field, the effective field in the rotating frame is aligned along the magic angle in spin space. In a first variant of this technique,[76] the effective field plays the same role in the rotating frame as the static magnetic field does in the laboratory frame. In analogy to the rf pulses, audio-frequency pulses are now applied and their frequency is adjusted to the amplitude of the effective field. They excite coherences that are not affected by the dipolar interaction as long as the effective field is along the direction of the magic angle. Spatial encoding now takes place by a suitable combination of static and rf gradients.

A simpler variant[77] follows the original Lee–Goldburg scheme in which a suitable off-resonance pulse is applied.[63] The component of the initial magnetization which is perpendicular to the effective field evolves in the rotating frame under influence of a combination of a static and an rf gradient. The spatial information is acquired stepwise by incrementing the evolution period, i.e. the length of the off-resonance pulse, which makes this variant a pure phase-encoding technique.

4.1.2.3 Multiple-pulse Imaging[78–84] The dipolar interaction can be averaged in spin space using a suitably designed series of pulses applied on-resonance.[63] These pulses flip the interaction Hamiltonian around the magic angle in spin space such that after cycles of three such flips the average dipolar interaction vanishes to zeroth order and line narrowing is achieved.

Spatial information can be encoded by applying a gradient in the evolution windows between the pulses.[78,79] Although conventional multiple-pulse techniques are demanding, multiple-pulse imaging has already been applied to polymers[80] and to the study of chemical reactions.[81]

A less demanding variant of the multiple-pulse technique[82] is based on the magic-echo sequence.[83,84] We shall discuss this technique in more detail together with its application to polymers in section 4.3. Multiple-pulse line narrowing and MAS line narrowing have also been combined for line narrowing in solid-state imaging,[85,86] but useful applications to polymers are rare because of the high technical sophistication of this experiment.

4.1.2.4 Field-modulation Imaging[87] Line narrowing can be also achieved with a combination of rf irradiation and magnetic-field modulation.[63] The off-resonance sensitivity of this technique leads to slice selection in the presence of a gradient. By moving the sensitive slice across the sample, one can map the spin density.

4.1.2.5 Constant-time Phase-encoding Imaging This approach relies on pure phase-encoding of the spatial information. The phase-encoding gradient is incremented in successive experiments for spatial encoding while the signal is acquired at a fixed time. Thus no signal decay due to undesired line broadening is sampled.[88] The main advantages of constant-time phase-encoding techniques is that they are robust, easy to implement and can be conveniently combined with other NMR techniques.

The drawback of the method in its original form[88] is that in rigid solids the signal decays too fast for reasonable phase-encoding. For modestly rigid solids, however, the technique can provide well-resolved images, as is shown by some applications.[89] The S/N in these experiments can be improved by using low-angle excitation. In this case, T_1 relaxation is no longer a strong limitation, and the repetition time can be made much shorter.[90]

The constant-time phase-encoding technique can also be applied to rigid solids. In this case the fast decay due to the dipolar interaction has to be refocused by generating dipolar echoes.[91,92] Here again, the magic-echo sequence provides the best results,[93–95] and is therefore treated in more detail below together with its polymer applications.

Constant-time phase-encoding techniques were also used in combination with ^{13}C spectroscopy exploiting the fact that with pure phase-encoding techniques the direct dimension can be used for spectroscopic purposes. The spatial distribution of molecular order in drawn polypropylene has been studied by this technique.[96] Deuteron NMR has also been shown to provide useful information in a spectroscopic imaging experiment. For this, constant-time phase-encoding imaging was combined with multiple-quantum imaging.[97] The technique was used for spatially resolved studies of polymer orientation and molecular dynamics taking advantage of the sensitivity of deuteron line shapes to these material properties.

4.1.2.6 Imaging of Dilute Nuclei[75,97,98] The problem of dipolar broadening can be avoided by imaging dilute nuclei such as ^{13}C.[98] In this case, the main line broadening results from the interaction with the abundant proton spins which can be easily removed by heteronuclear dipolar decoupling. As already outlined above, the high spectroscopic information of ^{13}C spectroscopy can be exploited either in static[97,98] or MAS[75] experiments.

The main disadvantage of this approach is the low S/N, which usually prevents satisfactory spatial resolution.

At present, the most interesting approaches with respect to applications are stray-field imaging and constant-time phase-encoding/multiple-pulse line narrowing techniques, in particular those based on the magic-echo pulse sequence. Therefore, we shall discuss them in more detail below and show some representative applications.

4.2 Stray-field Imaging

4.2.1 Principle

In stray-field imaging, the sample is placed in the stray field of a conventional superconducting magnet. The gradient in the stray field of such a magnet is extremely high (of the order of $60\,T\,m^{-1}$). It leads to such a broad distribution of resonance frequencies that even a broadband rf pulse of length 10 μs can only excite part of it. That is, only the signal from a slice perpendicular to the gradient direction is measured. By mechanically moving the sample in the gradient direction, the sensitive slice is moved across the sample and the corresponding signal intensities can be determined as a function of slice position. That is, step by step a spin density profile of the sample is acquired. Two- or three-dimensional images are obtained by stepwise rotation of the sample to determine the profiles in the other directions. For the reconstruction of the image, back-projection techniques can be used (see section 2.2.2).

Since the signal decays rapidly in the strong magnetic field gradient, the magnetization is usually observed in the form of an echo or an echo train. Thus, relaxation contrast can be introduced in stray-field imaging. As an alternative, the S/N can be improved to some extent by adding the echo intensities of an echo train. However, there are problems with such echo trains in strong field gradients. These problems have been extensively discussed in the literature together with strategies to partially overcome them.[66,67]

As in conventional imaging, the resolution of stray-field imaging is ultimately limited by the S/N but also other limitations are present. Apart from technical limitations such as the precision of the stepping motor and the curvature of the sensitive plane, the minimum slice width is determined by the length of the excitation pulse which in turn is limited by the T_2 of the sample. Thus, the resolution in more mobile materials is better than in the case of solid polymers, although even there it is of the order of tens of micrometers. For very mobile samples, self-diffusion limits the resolution so that the best resolution is obtained for quasi-solids such as elastomers.

Several other variants of the stray-field experiment have been developed.[66,67] In the frequency-sweep technique, the sample remains static and the slice is scanned across the sample by changing the rf excitation frequency. A Fourier-transform variant of stray-field imaging also exists. It corresponds to conventional (frequency encoding) imaging except that the static stray-field gradient is used. Since this gradient is extremely strong, a broad receiver bandwidth is required, which limits the application of the technique to very small samples.

4.2.2 Applications

In the following, a brief overview of stray-field imaging of polymers is given with representative examples. More complete recent reviews that include also applications to other materials, can be found in the literature.[66,67] A selection of applications to polymers and other materials, in particular studies of the solvent ingress and of the hardening of materials, can be also found in some of the articles in the book edited by Blümler et al.[2]

One of the early images of a plastic object obtained by stray-field imaging is shown in Figure 14(a). This clearly demonstrates that 3D imaging is feasible with stray-field techniques and provides comparatively good resolution. Nevertheless, most applications of stray-field imaging so far have been one-dimensional. In many cases, the imaging geometry can be arranged such that a one-dimensional experiment already provides the required information.

Stray-field imaging was applied predominantly to topics that were also studied using conventional imaging techniques, e.g. solvent ingress into polymers (see also section 3.4). Unlike these studies, however, stray-field images can represent both the solvent signal and the polymer signal because the technique is largely insensitive to T_2 relaxation. Early studies of swelling were performed on poly(vinyl chloride) swollen by acetone. The solvent ingress was found to follow non-Fickian diffusion laws. Chudek et al.[99] investigated solvent ingress into poly(methyl methacrylate) and the results can be compared with previous results obtained by conventional NMR imaging (see section 3.4).

The opposite of swelling, the drying of films or paints, can be also studied by stray-field techniques, again taking advantage of the insensitivity to T_2 relaxation. It is therefore possible to follow drying processes of initially viscous liquids until they are transformed into brittle glass. For instance, the drying of a latex (paint) film has been studied this way.[68] While the signal remains liquid-like at early stages, T_2 suddenly shortens drastically, preventing other imaging techniques from being used.

Other applications exploit the property that stray-field imaging is fairly resistant to the effects of susceptibility. This can be used to study such unfavorable cases as the

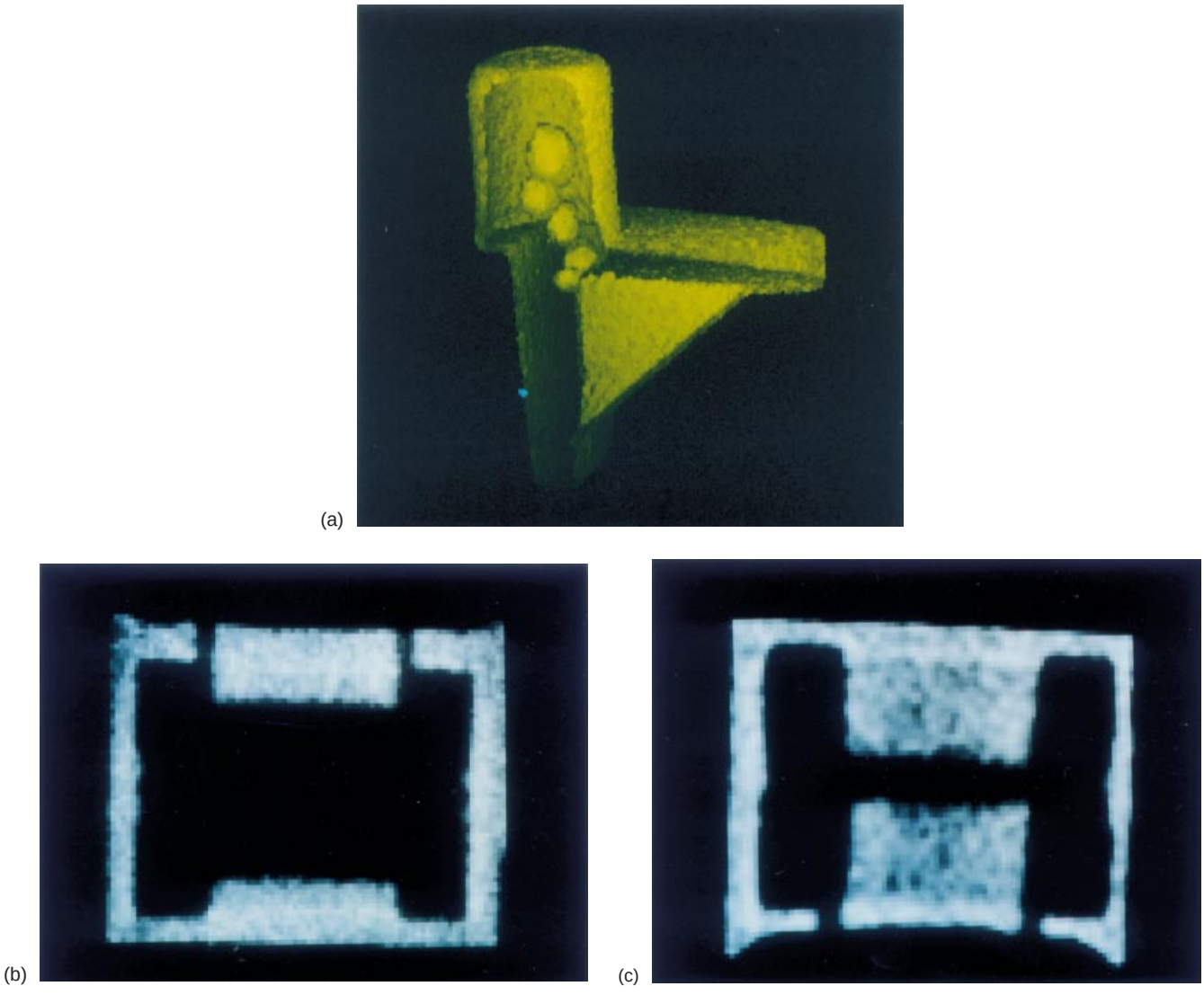

Figure 14 (a) 3D stray-field image of a polycarbonate plug showing that the manufacturing process left air bubbles in the inner parts. The other two images show a mica capacitor before (b) and after (c) burnout. In (b), the intact capacitor plates shield the rf so that no signal from the center parts is acquired. After burnout these plates are destroyed. This example shows that stray-field imaging can be applied to samples that contain metallic parts. (Reproduced by permission of K. Zick, Bruker Analytik, Rheinstetten.)

polymer matrix in an electrical device. As an example, Figure 14(b) and (c) show a mica capacitor imaged before and after burnout, respectively.

4.3 Magic-echo Imaging

4.3.1 Methods

The magic-echo sequence[83,84] is designed such that during a part of the sequence [the *magic sandwich*[84]] the dipolar Hamiltonian has a negative sign. The coherences that have first evolved under influence of the normal dipolar Hamiltonian thus refocus completely during this time and a so-called magic echo appears. In contrast to other solid echoes, the magic echo can be generated for very long echo times, in principle. Although in practice there are technical limitations, the technique nevertheless allows a comparatively long free evolution period for spatial encoding and thus represents the most convenient approach for line narrowing in solid-state imaging. Two variants have been developed on this basis, as follows.

Magic-echo frequency-encoding solid imaging uses homonuclear line narrowing based on the stroboscopic observation of multiple magic echoes.[82] Spatial encoding takes place by gradient pulses of constant strength applied in the windows of the multiple-magic-echo sequence. The phase of the last pulse of the sandwich is changed from $-y$ to y to compensate for the influence of chemical shifts

and field inhomogeneities. The gradient and the receiver phase must then be inverted in every second window. The technique has been demonstrated on solid-state phantom samples of adamantane and hexamethylbenzene (the proton line width at half-intensity of both materials is about 15 kHz) with a reported resolution of better than 100 μm. It was also applied in combination with MAS line narrowing.[86]

Magic-echo phase-encoding solid imaging[93–95] is based on the constant-time approach and uses the magic-echo sequence to prolong the phase-encoding time. As in all constant-time techniques, the residual line broadening does not affect the resolution. The resolution is therefore solely determined by the pixel resolution achievable within the coherence lifetime. Magic-echo phase-encoding is less demanding with respect to instrumentation than the frequency encoding variant. It was demonstrated that images of a remarkably good quality can be achieved even with gradients that are stationary throughout the pulse sequence, so that standard imaging equipment can be used. A resolution of about 100 μm could be achieved with an adamantane test sample. Demonstration experiments with polymers include the imaging of a rubber sample below the glass-transition temperature (the line width at half-intensity exceeds 35 kHz) and the imaging of a plug.[94,95]

The two variants can be combined to form a multiple-magic-echo 2D-Fourier sequence which is the solid-state analog of the conventional 2D Fourier imaging sequence and combines the advantages of the two approaches. A contrast filter preceding the imaging sequence is usually also applied to introduce parameter contrast. The $T_{2\text{eff}}$ relaxation filter was found to be particularly useful for polymer investigations since it is sensitive to molecular motions in the tens of kilohertz regime.

4.3.2 Applications

Despite the sophistication of such experiments, several polymer investigations have been reported. In all of them, a combination of magic-echo imaging with spectroscopic or parameter-selective sequences is essential.

4.3.2.1 Spectroscopic Imaging of Chain Orientations As a first example of spectroscopic magic-echo imaging, drawn polytetrafluoroethylene samples were investigated using ^{19}F-NMR.[100] The ^{19}F spectra show both crystalline and amorphous line contributions. The latter, however, is filtered out under imaging conditions. Drawing of the sample changes the chain orientation and, hence, the chemical-shift anisotropy. The intensity at a certain chemical-shift value can be used as a measure of the applied stress and, applying a chemical-shift imaging sequence, the stress distribution can be mapped as a function of space.

The imaging sequence used in this investigation was a pure phase-encoding multiple-magic-echo sequence with short gradient pulses inserted in the free-evolution intervals. The acquisition of the chemical-shift-resolved spectroscopic dimension takes place in the free-evolution windows of a magic-echo multiple-pulse sequence that eliminates the dipolar interaction but preserves the chemical-shift anisotropy. As a result, a three-dimensional data set is acquired with two spatial dimensions and one spectroscopic dimension. In the spectroscopic dimension, the chemical-shift anisotropy is present and can be exploited for the investigation of orientation dependencies.

An experiment was carried out with a polytetrafluoroethylene sheet in which a cut had been made on one side. After applying tensile stress, a ^{19}F spin-density image and a (normalized) chemical-shift selective image (78 ppm) were recorded. The difference image shows only the regions corresponding to the oriented material. The highest values in the difference image indicate the regions of the strongest strain where the crack eventually occurred.[100]

4.3.2.2 Parameter Imaging of Structure and Mobility in Polymers Various sequences for parameter weighting have been combined with magic-echo imaging, including sequences for measuring the proton spectrum with or without multiple-pulse line narrowing, and filters for T_2, T_1, $T_{2\text{eff}}$ or spin diffusion.[94]

The combination of spin-diffusion experiments with magic-echo imaging allows the measurement of domain-size distributions in an inhomogeneous sample.[101] Thus, macroscopic and microscopic spatial structures can be correlated. As an example, this strategy was applied to cable-insulator material which had been damaged by a high-voltage breakthrough. The spatially resolved spin-diffusion measurement indicated changes of the domain sizes in the affected parts of the cable.

The influence of tensile stress on molecular mobility was investigated using a $T_{2\text{eff}}$ filter sequence. As an example, the molecular mobilities in the necking region of a cold-drawn polycarbonate sample were studied.[102,103] While simple spin-density images show no spatial variations in the material density, an immobilization is found in the necking region. Increasing the stress until the sample starts to break yields an even more pronounced immobilization in two half-moon like regions close to the break (see Figure 15a).

In addition, shear bands can be investigated by the 2D imaging sequence.[103,104] The immobilization of the chains in the shear band is reflected in the $T_{2\text{eff}}$ image while the shear band is not visible in a simple spin-density image. Two crossed shear bands were also visualized by this technique, showing that the immobilization in the intersection of the two is not significantly higher than that in each of them alone (see Figure 15b).

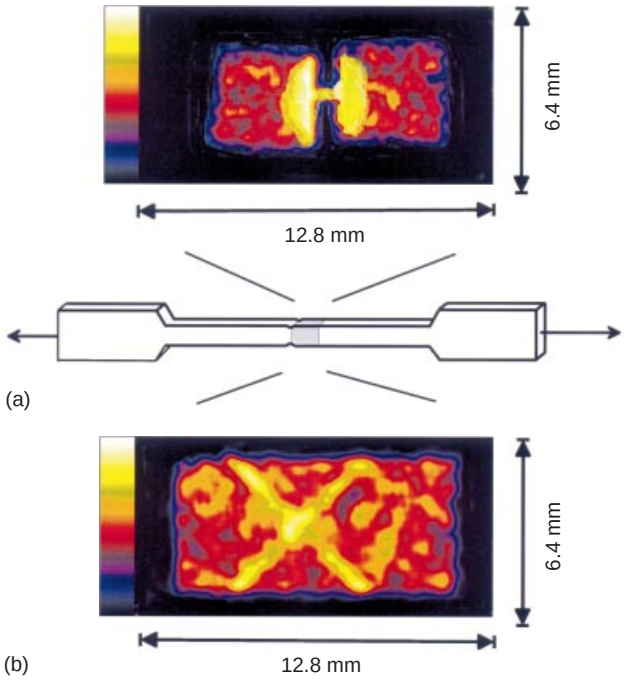

Figure 15 T_{2eff}-relaxation-time images of two pieces of drawn polycarbonate acquired with the magic-echo solid-state imaging sequence. (a) T_{2eff} image of a sample that has been drawn close to the point where it breaks into two pieces; (b) T_{2eff} image of a sample in which two crossing shear bands have been generated. The parameter T_{2eff} clearly reveals the shear band as a region of reduced molecular mobility while the shear bands are not visible in the corresponding spin-density image (not shown). In both cases the rate $1/T_{2eff}$ is displayed. [Reproduced by permission from Traub et al.[34]]

These results demonstrate that, by parameter-selective ^{1}H imaging, material-relevant information can be obtained even in rigid solids. This suggests the applicability of NMR imaging to a wide range of problems in polymer science.

5 SPECIAL TECHNIQUES

5.1 Surface-scan Techniques

Since strong rf pulses are required in solids, there are severe technical limitations on the size of the coil and hence on the sample size. Typical coil diameters used in microimaging are between 5 and 20 mm depending on the required pulse power, that is, on the rigidity of the investigated material.

Considerable efforts have been made to overcome the sample size restriction. The use of surface coils has been proposed for the imaging of larger objects[105] and also devices where the sample can be placed outside the magnet.[106–110] As a prominent example of the latter case, a mobile surface scanner has been described.[107–110] It consists of two permanent magnets separated by a gap where the rf coil is located. The sensitive spot, which consists of a strip of about 3.5 mm along the symmetry axis of the gap, can be scanned over the sample. As in conventional imaging experiments, relaxation studies can be performed. The cross-link density of natural rubber samples has been investigated,[109] and also the stress of a household rubber band. Entire automobile tires can also be investigated despite the presence of ferromagnetic reinforcing steel belts.[107]

5.2 Magnetic Resonance Force Microscopy

Magnetic resonance force microscopy (MRFM) has been proposed for the detection of very small volumes of spins for which conventional NMR techniques are not sensitive enough.[111] This technique promises to be sensitive to the presence of few spins or even to a single spin in the far future so that real microscopic resolution would be achievable.[111–113] The driving motivation behind this technique is thus the perspective to determine molecular structures of biomolecules by direct 3D imaging.[114] In the following, we again restrict ourselves to a brief overview and refer to the literature[114] for more details.

In contrast to the inductive detection used in conventional NMR, MRFM uses mechanical detection of NMR in a force microscope. For this, the sample is fixed on a cantilever and placed in the magnetic field of a permanent magnet. If unpaired electrons or nuclear moments are present in the sample, a force acts on the cantilever. By rf irradiation of the appropriate strength, the magnetization of the sample now is modulated at the resonance frequency of the cantilever so that it starts to oscillate. This oscillation is detected by an interferometry device as the NMR signal.

The applied rf field is able to influence only those spins for which the resonance condition is fulfilled. In a suitably designed inhomogeneous field this is the case only for a given volume of the sample. Shifting the resonance frequency or the position of the sample in the magnetic field thus permits one to scan the resonant volume across the sample.

The technique is currently in an early stage of development but first images have already been recorded. For instance, the one-dimensional image of an ammonium nitrate sample was acquired with a spatial resolution of 2.6 μm.[112] Although ammonium nitrate was chosen because of its favorable relaxation behavior, the experiment is also promising for polymer imaging.

5.3 Volume-selective Spectroscopy

This approach is an alternative to spectroscopic imaging. Whereas in the latter case the full sample is spatially

encoded, spatial information is now introduced by selecting a defined volume element from which the required information can then be extracted by normal NMR spectroscopy.[3] Volume-selective spectroscopy is particularly useful if only selected parts of a sample are of interest or if the quality of the resulting spectrum is of prime importance. Although volume-selective spectroscopy is well established in medicine, applications to materials are rare. Slice selection is particularly difficult for rigid materials since the amplitude of the rf field must exceed the dipolar-coupling strength. Shaped pulses, however, which are normally used for slice selection (see section 2.2.1), are too weak for that purpose.

Nevertheless, first approaches to slice selection in solids have been demonstrated using sequences of broad-band pulses. One technique uses multiple-pulse line narrowing during a frequency-selective multiple-pulse train for selective excitation.[115] The Lee–Goldburg sequence has also been used for slice selection in an imaging experiment.[116] Another approach takes advantage of the frequency selectivity of a spin-lock pulse to select a slice or, applying this technique in all three dimensions, for full volume-selective spectroscopy.[117] As an application of this technique, the orientation of Kevlar fibers in a model sample was determined using the chemical shift as a probe. Finally, the possibility of slice-selective ^{13}C spectroscopy has been demonstrated by taking advantage of the off-resonance sensitivity of the cross-polarization procedure.[118] An overview with more details on these techniques can be found elsewhere.[3]

6 SUMMARY AND FUTURE PERSPECTIVES

During the past 10 years, there has been considerable success in adapting NMR imaging techniques to the investigation of materials. Most of the efforts in the field have focused on method development but, more recently, these efforts were rewarded by a wealth of applications, some even in solids. NMR imaging of polymers is currently in a exciting state where it can contribute already to material research but where future technical advances still hold promise for new breakthroughs.

Despite all methodological efforts, the achievable resolution of the classical techniques will never extend much beyond 10 μm, which is relatively modest compared with other microscopic methods. However, as already discussed, there are several advantages that compensate for this. NMR imaging therefore should not be considered as a competitor of conventional microscopy but as a complement to it. Whereas microscopic techniques are predominantly surface techniques, NMR imaging is a noninvasive bulk technique with high spectroscopic information content. Nondestructivity is certainly not as important for materials investigations as it is for medical applications; nevertheless, it is indispensable for some specialized applications such as the in situ investigation of processes. Moreover, it allows the same object to be investigated further using other techniques.

The main advantage of NMR imaging, however, is the possibility of introducing spectroscopic information. In principle, all information accessible from conventional NMR techniques can also be acquired in a spatially resolved manner and only a few of these possibilities have been realized so far. At present, the high sophistication of some of these methods hampers their accessibility to research groups which specialize in material science rather than in NMR. Nevertheless, interest in these techniques from the polymer industry is noticeable, and some of the examples shown were actually acquired in industrial research laboratories or in cooperation with them. As more material scientists become familiar with NMR imaging, a new step with respect to applications is expected.

ACKNOWLEDGMENTS

I thank R. Komoroski, P. Blümler, B. Blümich, D. Gross and K. Zick for providing figures and the permission to use them in this article. Further, I thank my colleagues B. Traub, F. Weigand, P. Barth, R. Kimmich, D.E. Demco and H.W. Spiess for their excellent collaboration in the field of NMR imaging. Some of the examples used in this article are from this work. I am indebted to S. De Paul and P. Blümler for carefully checking the manuscript and for helpful suggestions.

ABBREVIATIONS AND ACRONYMS

FID	Free Induction Decay
MAS	Magic-angle Spinning
MRFM	Magnetic Resonance Force Microscopy
NMR	Nuclear Magnetic Resonance
S/N	Signal-to-noise Ratio

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance in Medicine, High Resolution Ex Vivo • Multinuclear Magnetic Resonance Spectroscopic Imaging

Food **(Volume 5)**
Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Mechanical Properties of Polymers and Rubbers • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Relaxation in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. B. Blümich, W. Kuhn (eds.), *Magnetic Resonance Microscopy*, VCH, Weinheim, 1992.
2. P. Blümler, B. Blümich, R. Botto, E. Fukushima (eds.), *Spatially Resolved Magnetic Resonance*, Wiley-VCH, Weinheim, 1998.
3. R. Kimmich, *NMR-Tomography, Diffusometry, Relaxometry*, Springer-Verlag, Berlin, 1997.
4. P.T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Clarendon Press, Oxford, 1991.
5. P.G. Morris, *Nuclear Magnetic Resonance Imaging in Medicine and Biology*, Clarendon Press, Oxford, 1986.
6. P. Jezzard, J.J. Attard, T.A. Carpenter, L.D. Hall, 'Nuclear Magnetic Resonance Imaging in the Solid State', *Prog. Nucl. Magn. Reson. Spectrosc.*, **23**, 1–41 (1991).
7. P. Blümler, B. Blümich, 'NMR Imaging of Solids', in *NMR Basic Principles and Progress*, eds. P. Diehl, E. Fluck, H. Guenther, R. Kosfeld, J. Seelig, Springer-Verlag, Berlin, 209–281, Vol. 30, 1994.
8. D.G. Cory, 'Solid State NMR Imaging', *Annu. Rep. NMR Spectrosc.*, **24**, 87–180 (1992).
9. P. Blümler, B. Blümich, 'NMR Imaging of Elastomers: a Review', *Rubber Chem. Technol.*, **70**(3), 469–518 (1997).
10. B. Blümich, P. Blümler, K. Saito, 'NMR Imaging and Spatial Information', in *Solid-state NMR of Polymers*, eds. I. Ando, T. Asakura, Elsevier, Amsterdam, 123–163, 1997.
11. B. Blümich, P. Blümler, 'NMR Imaging of Polymer Materials', *Makromol. Chem.*, **194**, 2133–2161 (1993).
12. P. Jezzard, C.J. Wiggins, T.A. Carpenter, L.D. Hall, P. Jackson, N.J. Clayden, N.J. Walden, 'Nuclear Magnetic Resonance Imaging of Polymers and Polymer Composites', *Adv. Mater.*, **4**, 82–90 (1992).
13. J.H. Strange, M.R. Halse, 'Imaging Techniques for Solids and Quasi-solids', in *Encyclopedia of NMR*, eds. D. Grant, R.K. Harris, John Wiley & Sons, Chichester, 2472–2481, 1996.
14. J.B. Miller, 'NMR Imaging of Materials', *Prog. Nucl. Magn. Reson. Spectrosc.*, **33**, 273–308 (1998).
15. L.F. Gladden, P. Alexander, 'Applications of Nuclear Magnetic Resonance Imaging in Process Engineering', *Meas. Sci. Technol.*, **7**(3), 423–435 (1996).
16. B. Blümich, 'Contrast in Solid-state NMR Imaging: 1. Principles', *Concepts Magn. Reson.*, **10**(1), 19–31 (1998).
17. P. Blümler, V. Litvinov, H.G. Dikland, M. van Duin, 'Detection of Interfacial Layers in Elastomer Samples by NMR Imaging', *Kautsch. Gummi Kunstst.*, **51**, 865–867 (1998).
18. S.N. Sarkar, R.A. Komoroski, 'NMR Imaging of Morphology, Defects and Composition of Tire Composites and Model Elastomer Blends', *Macromolecules*, **25**, 1420–1426 (1992).
19. W.E. Maas, L.H. Mervin, D.G. Cory, 'Nuclear Magnetic Resonance Imaging of Solid Rocket Propellants at 14.1 T', *J. Magn. Reson.*, **129**, 105–108 (1997).
20. C.P. Gonatas, J.S. Leigh, A.G. Yodh, J.A. Glazier, B. Prause, 'Magnetic Resonance Images of Coarsening Inside a Foam', *Phys. Rev. Lett.*, **75**(3), 573–576 (1995).
21. P. Blümler and B. Blümich, 'Aging and Phase-separation of Elastomers Investigated by NMR Imaging', *Macromolecules*, **24**, 2183–2188 (1991).
22. A.G. Webb, P. Jezzard, L.D. Hall, 'Detection of Inhomogeneities in Rubber Samples Using NMR Imaging', *Polym. Commun.*, **30**, 363–366 (1989).
23. M.A. Rana, J.L. Koenig, 'Observation of Spatial Inhomogeneities in *N-tert*-Butylbenzothiazolesulfenamide-sulfur Cured High-vinyl Polybutadiene Using NMR Imaging', *Macromolecules*, **27**, 3727–3734 (1994).
24. S.R. Smith, J.L. Koenig, 'Observation of Cross-link Density Dispersion in Sulfur Donor-vulcanized Polybutadiene Using NMR Imaging', *Macromolecules*, **24**, 3496–3504 (1991).
25. W. Kuhn, E. Köller, I. Theis, 'Investigation of Aging Processes and Crosslinking in Elastomers by Parameter Selective NMR Microscopy', in *Magnetic Resonance*

Microscopy, eds. B. Blümich, W. Kuhn, VCH, Weinheim, 217–235, 1992.
26. W. Kuhn, P. Barth, S. Hafner, G. Simon, H. Schneider, 'Material Properties Imaging of Cross-linked Polymers by NMR', *Macromolecules*, **27**, 5773–5779 (1994).
27. P. Barth, S. Hafner, P. Denner, 'Material Property NMR Imaging of Crosslink Densities Based on Longitudinal Relaxation in the Rotating Frame', *Macromolecules*, **29**, 1655–1659 (1996).
28. S. Hafner, P. Barth, 'Aging of Polymer Networks as Studied by Material Property NMR Imaging', *Magn. Reson. Imag.*, **13**, 739–744 (1995).
29. M. Knörgen, U. Heuert, H. Schneider, 'Application of NMR Imaging to Industrial Polymers', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, Weinheim, 211–220, 1998.
30. C. Fülber, B. Blümich, K. Unseld, V. Herrmann, 'NMR Imaging of Thermal-oxidative Aging in SBR', *Kautsch. Gummi Kunstst.*, **48**, 254–259 (1995).
31. H. Yasunaga, H. Kurosu, I. Ando, 'Spatial Information on a Polymer Gel as Studied by ^{1}H NMR Imaging. 1. Image Analysis of Stress–Strain', *Macromolecules*, **25**, 6505–6509 (1992).
32. P. Blümler, B. Blümich, 'Investigation of Stress Distribution in Filled Polysiloxane by NMR Imaging', *Acta Polym.*, **44**, 125–131 (1993); P. Blümler, B. Blümich, in *Magnetic Resonance Microscopy*, eds. B. Blümich, W. Kuhn, VCH, Weinheim, 167–185, 1992.
33. M. Klinkenberg, P. Blümler, B. Blümich, '^{2}H-NMR Imaging of Stress in Strained Elastomers', *Macromolecules*, **30**(4), 1038–1043 (1997).
34. B. Traub, S. Hafner, D. Maring, H.W. Spiess, 'NMR Imaging of Mechanically Treated Polymers', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, Weinheim, 179–193, 1998.
35. U. Günther, K. Albert, 'Monitoring of Photopolymerization Processes by NMR Imaging', *J. Magn. Reson.*, **98**, 593–598 (1992).
36. P. Jackson, N.J. Clayden, N.J. Walton, T.A. Carpenter, L.D. Hall, P. Jezzard, 'Magnetic Resonance Imaging Studies of the Polymerization of Methylmethacrylate', *Polym. Int.*, **24**, 139–143 (1991).
37. T.A. Carpenter, L.D. Hall, P. Jezzard, C.J. Wiggins, N.J. Clayden, P. Jackson, N. Walton, 'MRI Studies of in situ Polymerization Reactions at Elevated Temperatures', in *Magnetic Resonance Microscopy*, eds. B. Blümich, W. Kuhn, VCH, Weinheim, 267–275, 1992.
38. B.J. Balcom, T.A. Carpenter, L.D. Hall, 'Methacrylic Acid Polymerization. Travelling Waves Observed by Nuclear Magnetic Resonance Imaging', *Macromolecules*, **25**, 6818–6823 (1992).
39. C. Fülber, K. Unseld, V. Herrmann, K.H. Jakob, B. Blümich, 'In Situ Investigation of SBR Vulcanization: a combined Study of ^{1}H-NMR and Vulcametry', *Colloid Polym. Sci.*, **274**(3), 191–196 (1996).
40. P. Jackson, 'Curing of Carbon-fiber Reinforced Epoxy Resin; Non-invasive Viscosity Measurement by NMR Imaging', *J. Mater. Sci.*, **27**, 1302–1306 (1992).
41. D. Hauck, P. Blümler, B. Blümich, 'NMR Imaging of Technical SBR Vulcanizates Under Dynamic Mechanical Load', *Macromol. Chem. Phys.*, **198**, 2729–2742 (1997).
42. S.W. Sinton, A.W. Chow, J.W. Iwamiya, 'NMR Imaging as a Tool for Rheology', *Macromol. Symp.*, **86**, 299–309 (1994).
43. S. Hafner, 'Fast Imaging in Liquids and Solids with the Back-projection Low Angle ShoT (BLAST) Techniques', *Magn. Reson. Imag.*, **12**, 1047–1051 (1994).
44. S. Hafner, W. Kuhn, 'NMR Imaging of Water Content in the Polymer Matrix of Silicon Chips', *Magn. Reson. Imag.*, **12**(7), 1075–1078 (1994).
45. W.P. Rothwell, D.R. Holecek, J.A. Kershaw, 'NMR Imaging: Study of Fluid Absorption by Polymer Composites', *J. Polym. Sci. Polym. Lett. Ed.*, **22**, 241–247 (1984).
46. B. Klei, J.L. Koenig, 'NMR Imaging of Competitive Vulcanization of Natural Rubber and Polybutadiene Blends', *Acta Polym.*, **48**, 199–207 (1997).
47. Y. Song, H.C. Gaede, T. Pietrass, G.A. Barrall, G.C. Chingas, M.R. Ayers, A. Pines, 'Spin-polarized ^{129}Xe Gas Imaging of Materials', *J. Magn. Reson., Ser. A*, **115**, 127–130 (1995).
48. A.G. Webb, L.D. Hall, 'An Experimental Overview of the Use of Nuclear Magnetic Resonance Imaging to Follow Solvent Ingress into Polymers', *Polymer*, **32**, 2926–2938 (1991).
49. R.A. Grinsted, J.L. Koenig, 'Study of Multicomponent Diffusion into Polycarbonate Rods Using NMR Imaging', *Macromolecules*, **25**, 1229–1234 (1992).
50. L.A. Weisenberger, J.L. Koenig, 'NMR Imaging of Diffusion Processes in Polymers: Measurement of the Spatial Dependence of Solvent Mobility in Partially Swollen PMMA Rods', *Macromolecules*, **23**, 2445–2453 (1990).
51. M. Ercken, P. Adriaensens, G. Reggers, R. Carleer, D. Vanderzande, J. Gelan, 'Use of Magnetic Resonance Imaging to Study Transport of Methanol in Poly(methyl methacrylate) at Variable Temperature', *Macromolecules*, **29**, 5671–5677 (1996).
52. F. Tabak, M. Corti, 'Spatially Resolved Nuclear Magnetic Resonance Study of Polymer Chain Mobility During the Swelling Process', *J. Chem. Phys.*, **92**, 2673–2675 (1990).
53. T.M. Hyde, L.F. Gladden, 'Simultaneous Measurement of Water and Polymer Concentration Profiles During Swelling of Poly(ethylene oxide) Using Magnetic Resonance Imaging', *Polymer*, **39**, 811–819 (1998).
54. M. Ilg, K. Albert, W. Rapp, E. Bayer, 'First Applications of Nonmedical ^{19}F NMR Imaging', *J. Magn. Reson.*, **90**, 370–376 (1990).

55. B.J. Fahie, A. Nangia, S.K. Chopra, C.A. Fyfe, H. Grondey, A. Blazek, 'The Use of NMR Imaging in the Optimization of a Compression-coated Regulated Release System', *J. Controlled Release*, **51**, 179–184 (1998).
56. K. Mader, G. Bacic, A. Domb, O. Elmalak, R. Langer, H.M. Swartz, 'Non-invasive In Vivo Monitoring of Drug Release and Polymer Erosion from Biodegradable Polymers by EPR Spectroscopy and NMR Imaging', *J. Pharm. Sci.*, **86**, 126–134 (1997).
57. B. Traub, 'NMR Bildgebung an porösen Polymeren', Diploma Thesis, Universität Mainz, 1995.
58. J.M. Pope, S. Yao, 'Quantitative NMR Imaging of Flow', *Concepts Magn. Reson.*, **5**, 281–302 (1993).
59. Y. Xia, P.T. Callaghan, 'Study of Shear Thinning in High Polymer Solution Using Dynamic NMR Microscopy', *Macromolecules*, **24**, 4777–4786 (1991).
60. B. Manz, P.T. Callaghan, 'NMR Investigations of Self-diffusion and Shear Thinning for semidilute Polymer Solutions Near the Demixing Transition', *Macromolecules*, **30**, 3309–3316 (1997).
61. S.W. Sinton, A.W. Chow, 'NMR Flow Imaging of Fluids and Solid Suspensions in Poiseuille Flow', *J. Rheol.*, **35**(5), 735–772 (1991).
62. K. Rombach, S. Laukemper-Ostendorf, P. Blümler, 'Application of NMR Flow Imaging in Materials Science', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, Weinheim, 517–529, 1998.
63. M. Mehring, *Principles of High Resolution NMR in Solids*, 2nd edition, Springer-Verlag, Berlin, 1983.
64. A.A. Samoilenko, D.Y. Artemov, A.L. Sibel'dina, 'Formation of Sensitive Layer in Experiments on NMR Subsurface Imaging of Solids', *JETP Lett.*, **47**(7), 417–419 (1988).
65. A.A. Samoilenko, K. Zick, 'Stray-field Imaging of Solids (STRAFI)', *Bruker Rep.*, **1**, 40–41 (1990).
66. P.J. McDonald, 'Stray Field Magnetic Resonance Imaging', *Prog. Nucl. Magn. Reson. Spectrosc.*, **30**, 69–99 (1997).
67. P.J. McDonald, B. Newling, 'Stray Field Magnetic Resonance Imaging', *Rep. Prog. Phys.*, **61**(11), 1441–1493 (1998).
68. P. Glover, P.J. McDonald, B. Newling, 'Stray-field Imaging of Planar Films Using a Novel Surface Coil', *J. Magn. Reson.*, **126**, 207–212 (1997).
69. C.J. Rofe, J. Van Noort, P.J. Black, P.T. Callaghan, 'NMR Microscopy Using Large Pulsed Magnetic-field Gradients', *J. Magn. Reson., Ser. B*, **108**, 125–136 (1995).
70. A.N. Garroway, J. Baum, M.G. Munowitz, A. Pines, 'NMR Imaging in Solids by Multiple-quantum Resonance', *J. Magn. Reson.*, **60**, 337–341 (1984).
71. P. Jezzard, T.A. Carpenter, L.D. Hall, P. Jackson, N.J. Clayden, 'Simple NMR Imaging of Solid Polymers at Elevated Temperature', *Polym. Commun.*, **32**(3), 74–76 (1991).
72. W.S. Veeman, D.G. Cory, '^{1}H Nuclear Magnetic Resonance Imaging of Solids with Magic Angle Spinning', *Adv. Magn. Reson.*, **13**, 43–56 (1989).
73. D.G. Cory, J.C. de Boer, W.S. Veeman, 'Magic Angle Spinning ^{1}H NMR Imaging of Polybutadiene/Polystyrene Blends', *Macromolecules*, **22**(4), 1618–1621 (1989).
74. G. Schauss, B. Blümich, H.W. Spiess, 'Conditions for Generating Rotating Gradients in MAS NMR Imaging', *J. Magn. Reson.*, **95**, 437–441 (1991).
75. U. Scheler, G. Schauss, B. Blümich, H.W. Spiess, 'Spatially Resolved Solid-state MAS-NMR Spectroscopy', *Solid State NMR*, **6**, 375–388 (1996).
76. F. de Luca, B. Maraviglia, 'Magic Angle NMR Imaging in Solids', *J. Magn. Reson.*, **67**, 169–172 (1986).
77. F. de Luca, D.C. de Simone, N. Lugeri, B. Maraviglia, 'NMR Imaging of Solids by Spin Nutation in the Rotating Frame. A Comparative Analysis', *J. Magn. Reson., Ser. A*, **102**, 287–292 (1993).
78. G.C. Chingas, J.B. Miller, A.N. Garroway, 'NMR Images of Solids', *J. Magn. Reson.*, **66**, 530–535 (1986).
79. D.G. Cory, J.B. Miller, R. Turner, A.N. Garroway, 'Multiple-pulse Methods of ^{1}H NMR Imaging of Solids: Second Averaging', *Mol. Phys.*, **70**(2), 331–345 (1990).
80. D.G. Cory, J.B. Miller, A.N. Garroway, 'Chemically Resolved Solid State NMR Images of Polymers', *Macromol. Symp.*, **86**, 259–269 (1994).
81. L.G. Butler, D.G. Cory, K.M. Dooley, J.B. Miller, A.N. Garroway, 'NMR Imaging of Anisotropic Solid-state Chemical Reactions Using Multiple-pulse Line Narrowing Techniques and ^{1}H T_1 Weighting', *J. Am. Chem. Soc.*, **114**(1), 125–135 (1992).
82. S. Matsui, 'Solid-state NMR Imaging by Magic Sandwich Echoes', *Chem. Phys. Lett.*, **179**, 187–190 (1991).
83. H. Schneider, H. Schmiedel, 'Negative Time Development of a Nuclear Spin System', *Phys. Lett. A*, **30**, 298–299 (1969).
84. W.-K. Rhim, A. Pines, J.S. Waugh, 'Violation of the Spin-temperature Hypothesis', *Phys. Rev. Lett.*, **25**, 218–220 (1970).
85. D.G. Cory, A.M. Reichwein, J.W.M. van Os, W.S. Veeman, 'NMR Images of Rigid Solids', *Chem. Phys. Lett.*, **143**, 467–470 (1988).
86. Y. Sun, J. Xiong, H. Lock, M.L. Buszko, J.A. Haase, G.E. Maciel, 'Solid-state ^{1}H CRAMPS NMR Imaging with Pulsed Rotating Magnetic Field Gradients', *J. Magn. Reson., Ser. A*, **110**, 1–6 (1994).
87. R.A. Wind, C.S. Yannoni, 'Selective Spin Imaging in Solids', *J. Magn. Reson.*, **36**, 269–272 (1979).
88. S. Emid, J.H.N. Creyghton, 'High Resolution NMR Imaging in Solids', *Physica B*, **128**, 81–83 (1985).
89. D.E. Axelson, A. Kantzas, A. Nauerth, '^{1}H Magnetic Resonance Imaging of Rigid Polymeric Solids', *Solid State NMR*, **6**, 309–321 (1996).

90. S. Gravina, D.G. Cory, 'Sensitivity and Resolution of Constant Time Imaging', *J. Magn. Reson., Ser. B*, **104**, 53–61 (1994).
91. P.J. McDonald, J.J. Attard, D.G. Taylor, 'A New Approach to the NMR Imaging of Solids', *J. Magn. Reson.*, **72**, 224–229 (1987).
92. E. Rommel, S. Hafner, R. Kimmich, 'NMR Imaging of Solids by Jeener–Broekaert Phase Encoding', *J. Magn. Reson.*, **86**, 264–272 (1990).
93. S. Hafner, D.E. Demco, R. Kimmich, 'Magic-echo Solid-state NMR Imaging', *Meas. Sci. Technol.*, **2**, 882–885 (1991).
94. S. Hafner, D.E. Demco, R. Kimmich, 'Magic Echoes and NMR Imaging of Solids', *Solid State NMR*, **6**, 275–293 (1996).
95. D.E. Demco, S. Hafner, R. Kimmich, 'Spatially Resolved Homonuclear Solid-state NMR. Magic-echo and Rotary-echo Phase-encoding Imaging', *J. Magn. Reson.*, **96**, 307–322 (1992).
96. E. Günther, B. Blümich, H.W. Spiess, 'Molecular Order and Density of Skin and Core in Drawn Polyproylene Investigated by Spectroscopic ^{13}C NMR Imaging', *Macromolecules*, **25**(12), 3315–3316 (1992).
97. E. Günther, B. Blümich, H.W. Spiess, 'Deuteron Double-quantum NMR Imaging of Molecular Order and Mobility in Solid Polymers', *Mol. Phys.*, **71**(3), 477–489 (1990).
98. N.M. Szeverenyi, G.E. Maciel, 'NMR Spin Imaging of Magnetically Dilute Nuclei in the Solid-state', *J. Magn. Reson.*, **60**, 460–466 (1984).
99. J.A. Chudek, G. Hunter, F. Mohd Som, P.J. McDonald, B. Newling, 'Stray Field Imaging and Magnetic Resonance Microimaging Studies of High Impact Polystyrene, an Elastomer-toughened Material', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, Weinheim, 235–240, 1998.
100. M.A. Hepp, J.B. Miller, 'Mapping Molecular Orientation by Solid-state NMR Imaging', *J. Magn. Reson., Ser. A*, **111**, 62–69 (1994).
101. F. Weigand, D.E. Demco, B. Blümich, H.W. Spiess, 'Spatially Resolved NMR Spin Diffusion in Solid Polymers', *J. Magn. Reson., Ser. A*, **120**, 190–200 (1996).
102. F. Weigand, H.W. Spiess, 'Influence of Tensile Stress on the Molecular Mobility in Polycarbonate Visualized by Localized ^{1}H NMR Spectroscopy', *Macromolecules*, **28**, 6361–6364 (1995).
103. B. Traub, S. Hafner, U. Wiesner, H.W. Spiess, 'Investigation of Mechanical Deformation in Rigid Polymers by 2D Solid-state NMR Imaging', *Macromolecules*, **31**, 8585–8589 (1998).
104. F. Weigand, S. Hafner, H.W. Spiess, 'Magic-echo Parameter Imaging of Shearbands in Solid Polymers', *J. Magn. Reson., Ser. A*, **120**, 201–205 (1996).
105. J.B. Miller, A.N. Garroway, 'Dipolar-decoupled Inversion Pulses for NMR Imaging of Solids', *J. Magn. Reson.*, **85**, 432–437 (1989).
106. G.A. Matzkanin, 'A Review of Nondestructive Characterization of Composites Using NMR', in *Nondestructive Characterization of Materials*, eds. P. Höller, V. Hauck, C.O. Rund, R.E. Green, Springer-Verlag, Berlin, 655–669, 1989.
107. G. Eidmann, P. Savelsberg, P. Blümler, B. Blümich, 'The NMR MOUSE, a Mobile Universal Surface Explorer', *J. Magn. Reson., Ser. A*, **122**, 104–109 (1996).
108. G. Zimmer, A. Guthausen, U. Schmitz, K. Saito, B. Blümich, 'Weathering Investigation of PVC Coatings on Iron Sheets by the NMR MOUSE', *Adv. Mater.*, **9**, 987–989 (1997).
109. A. Guthausen, G. Zimmer, P. Blümler, B. Blümich, 'Analysis of Polymer Materials by Surface NMR via the MOUSE', *J. Magn. Reson.*, **130**, 1–7 (1998).
110. A. Guthausen, G. Zimmer, R. Eymael, U. Schmitz, P. Blümler, B. Blümich, 'Soft Matter Relaxation by the NMR MOUSE', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, Weinheim, 195–209, 1998.
111. D. Rugar, C.S. Yannoni, J.A. Sidles, 'Mechanical Detection of Magnetic Resonance', *Nature (London)*, **360**, 563–566 (1992).
112. D. Rugar, O. Züger, S. Hoen, C.S. Yannoni, H.-M. Vieth, R.D. Kendrick, 'Force Detection of Nuclear Magnetic Resonance', *Science*, **264**, 1560–1563 (1994).
113. O. Züger, D. Rugar, 'Magnetic-resonance Detection and Imaging Using Force Microscope Techniques', *J. Appl. Phys.*, **75**(10), 6211–6216 (1994).
114. J.A. Sidles, J.L. Garbini, K.J. Bruland, D. Rugar, O. Züger, S. Hoen, C.S. Yannoni, 'Magnetic Resonance Force Microscopy', *Rev. Mod. Phys.*, **67**(1), 249–265 (1995).
115. D.G. Cory, J.B. Miller, A.N. Garroway, 'DANTE Slice Selection for Solid State NMR Imaging', *J. Magn. Reson.*, **90**, 544–550 (1990).
116. F. de Luca, S. Motta, N. Lugeri, B. Maraviglia, 'Plane Selection in Indirect MARF Solid-state Imaging', *J. Magn. Reson., Ser. A*, **121**, 114–120 (1996).
117. S. Hafner, E. Rommel, R. Kimmich, 'Volume-selective Solid-state NMR Spectroscopy', *J. Magn. Reson.*, **88**, 449–456 (1990).
118. S. Hafner, D.E. Demco, R. Kimmich, 'Slice-selective Polarization Transfer (or Cross-polarization) for Localized Solid-state NMR Spectroscopy', *Chem. Phys. Lett.*, **187**, 53–59 (1991).

Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Peter A. Mirau
Bell Laboratories, Lucent Technologies, Murray Hill, USA

1 Introduction **1**
1.1 Fundamental Principles 1
1.2 Solid-state Nuclear Magnetic Resonance Methods 5
2 High-resolution Solid-state Nuclear Magnetic Resonance Studies of Polymers **12**
2.1 Introduction 12
2.2 Semicrystalline Polymers 12
2.3 Amorphous Polymers 15
2.4 Solid–Solid-phase Transitions 16
2.5 Elastomers 16
2.6 Other Studies 19
3 The Solid-state Organization of Polymers **21**
3.1 Introduction 21
3.2 Semicrystalline Polymers 23
3.3 Multiphase Polymers 24
3.4 Polymer Blends 25
4 The Dynamics of Polymers **31**
4.1 Introduction 31
4.2 Semicrystalline Polymers 35
4.3 Amorphous Polymers 38
4.4 Multiphase Polymers 42
5 Summary and Outlook **43**
Abbreviations and Acronyms **44**
Related Articles **45**
References **45**

Solid-state nuclear magnetic resonance (NMR) has emerged as one of the premier methods for polymer characterization in the solid-state. High-resolution spectra can be obtained with cross-polarization and magic-angle sample spinning using equipment that is commonly available on modern NMR spectrometers. Information about the chain structure, conformation, dynamics, intermolecular interactions and the morphology can be obtained by measuring the chemical shifts, line shapes and relaxation times. New experiments have been developed to filter the NMR response such that the spectrum from a particular component can be measured. The resolution has been increased with the development of multidimensional NMR methods for measuring the structure and dynamics. Solid-state NMR experiments can provide the link between the chemical structure and the macroscopic properties of polymers.

1 INTRODUCTION

1.1 Fundamental Principles

NMR spectroscopy is a method of great interest and importance for every aspect of the structure and properties of macromolecules. The first studies were reported by Alpert[1] only about a year after the discovery of nuclear resonance in bulk matter.[2,3] It was observed that natural rubber at room temperature gives a proton line width more like that of a mobile liquid than of a solid, but that the resonance broadens markedly at temperatures approaching the glass temperature. This was recognized as being related to the presence (and cessation) of micro-Brownian motion. NMR methods developed rapidly after these initial observations, and solution-state analysis of polymer microstructure has evolved into the one of the primary methods for the chemical characterization of polymers. With the exception of some early wideline proton NMR studies of polymer dynamics, the development of solid-state NMR for polymer analysis is a more recent development. However, since most polymers are used as solids, the solid-state NMR analysis of polymers has become increasingly important. These studies are of current interest since they can often provide the connection between the molecular structure and dynamics and the macroscopic properties. The methods have become so popular that most modern NMR spectrometers have the capability to perform routine solid-state NMR experiments. The development of new NMR methods is a topic of intense interest that has been driven in large part by the desire to relate the structure and dynamics of polymers measured by NMR to the macroscopic properties.

The NMR phenomenon is possible because in addition to charge and mass, many isotopes possess spin, or angular momentum. Since a spinning charge generates a magnetic field, there is a magnetic moment associated with this angular momentum. According to a basic principle of quantum mechanics, the maximum experimentally observable component of the angular momentum of a nucleus possessing a spin (or of any particle or system having angular momentum) is a half-integral or integral multiple of $h/2\pi$, where h is Planck's constant. This maximum component is I, which is called the spin quantum number or simply "the spin". In general, then, there are $2I + 1$ possible orientations or states of the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

nucleus, so if I is $\frac{1}{2}$, the possible magnetic quantum numbers are $+\frac{1}{2}$ and $-\frac{1}{2}$.

When placed in a magnetic field, nuclei having spin undergo precession about the field direction. The frequency of this so-called Larmor precession is designated w_0 in radians per second or ν_0 in hertz (Hz), cycles per second ($w_0 = 2\pi\nu_0$). The nuclei can be made to flip over, i.e. reverse the direction of the spin axis, by applying a second magnetic field, designated B_1, at right-angles to B_0 and causing this second field to rotate at the precession frequency ν_0.

One might suppose that the Larmor precession of the nuclear moments could itself be detected by some means without the need to invoke a resonance phenomenon. This is not possible because each nucleus precesses with a completely random phase with respect to that of its neighbors and there is therefore no macroscopic property of the system that changes at the Larmor frequency. By a well-known relationship, the Larmor precession frequency is given by Equation (1):

$$w_0 = \gamma B_0 \tag{1}$$

where γ is the magnetogyric ratio. The two quantities that determine the observation frequency for NMR signals are the magnetogyric ratio γ and the magnetic field strength B_0. Table 1 list some of the important nuclear properties of spins that are of interest to polymer chemists. The sensitivity depends both on the magnetogyric ratio and the natural abundance of the NMR-active nuclei. Protons have the highest sensitivity because they have the highest magnetogyric ratio and natural abundance. At a field strength of 11.7 T (1 tesla = 10^5 gauss) the NMR signals are observed at 500 MHz. Fluorine is the second most sensitive nucleus, but it is not a common element in polymers. Most polymers of interest contain carbon, and Table 1 shows that the sensitivity is very low compared with that of protons. However, the sensitivity of a modern NMR spectrometer is such that carbon can still be routinely observed. Nitrogen is also a common element in polymers but it is not studied extensively because of its low magnetogyric ratio and natural abundance. ^{15}N-NMR studies are possible, but only after site-specific isotopic labeling. The sensitivity of silicon and phosphorus is intermediate between that of protons and carbons, and provides a valuable probe of those polymers containing these elements.

With the exception of deuterium, all of the nuclei listed in Table 1 are spin-$\frac{1}{2}$. These are the easiest to study because there are line-narrowing methods that make it possible to obtain high-resolution spectra. Deuterium is a spin-1 nucleus with a quadrupole moment. This causes very efficient relaxation and gives rise to broad lines that cannot be narrowed using the common solid-state NMR methods. Thus, while it is difficult to obtain high-resolution deuterium spectra, the characteristic line shapes make this a valuable nucleus for the study of polymer dynamics. Isotopic labeling is required for deuterium NMR studies.

As in all spectroscopy, the signal-to-noise ratio is determined by the energy separation between the levels, and in the magnetic fields employed in NMR instruments, the separation of nuclear magnetic energy levels is very small. For example, for protons in a field of 14.1 T (500 MHz) it is only about 0.5 J. Even in the absence of the radiofrequency (rf) field there is usually a sufficiently rapid transfer of spins from the lower to the upper state and vice versa, such that an equilibrium population distribution is attained within a few seconds after B_0 is applied. For protons this corresponds to a difference of only ca. 10^{-5}. The population difference will be even smaller for nuclei with smaller magnetogyric ratios.

Spin–lattice relaxation is often termed *longitudinal relaxation* because it involves changes of energy and therefore involves the component of the nuclear moment along the direction of the applied magnetic field. It also shortens the lifetimes of the spins in both the upper

Table 1 The NMR properties of nuclei of interest to polymer chemists

Isotope	Abundance (%)	Spin	$\gamma \times 10^{-8}$[a]	Relative sensitivity[b]	ν_0 at 11.7 T (MHz)
^{1}H	99.98	$\frac{1}{2}$	2.6752	1.0	500
^{19}F	100	$\frac{1}{2}$	2.5167	0.83	470.2
^{29}Si	4.7	$\frac{1}{2}$	−0.5316	0.078	99.3
^{31}P	100	$\frac{1}{2}$	1.0829	0.066	202.3
^{13}C	1.1	$\frac{1}{2}$	0.6726	0.0159	125.6
^{2}H	0.015	1	0.4107	0.00964	76.7
^{15}N	0.365	$\frac{1}{2}$	−0.2711	0.0010	50.6

[a] Magnetogyric ratio in SI units.
[b] For an equal number of nuclei at constant field.

and lower energy states. This leads to an uncertainty in the energies of these states and to a broadening of the resonance lines. This can be estimated from the Heisenberg relation, $\delta E\,\delta t \approx h/2\pi$, from which the uncertainty in the frequency of absorption is given by Equation (2):

$$\delta\nu \approx \frac{1}{2\pi\delta t} \tag{2}$$

Although limiting the lifetime of a spin in a given energy state might in principle lead to line broadening, a more important source is the so-called ''dipolar'' broadening. To understand this effect, let us consider a proton-containing solid. In the case where the protons are sufficiently removed from each other that they do not feel the effects of each other's magnetic fields, the resonant magnetic field at the nucleus will be essentially equal to B_0 (actually very slightly smaller because of the shielding effects of local electrons). Therefore, if the field B_0 can be made very homogeneous over the volume occupied by the sample, the width of the absorption peak may be less than 10^{-4} G (10^{-8} T), i.e. of the order of a few tenths of a hertz, or 1 part in 10^9. In most such substances, the protons are actually near enough to each other so that each is appreciably influenced by the magnetic fields of its neighbors. Let us first imagine that we are dealing with a static system of isolated pairs of protons, i.e. each member of the pair experiences the field of the other member but not those of the other pairs. The field felt by each proton will be made up of the applied field B_0, plus this small additional field, B_{loc}. The sign and magnitude of this increment will depend upon the distance separating the nuclei (r) and upon θ, the angle between the line joining the nuclei and the vector representing B_0. Equation (3) expresses the functional dependence of the separation of the protons' magnetic energy levels on the distance r and the relative orientation θ:

$$\Delta E = 2\mu\,(B_0 \pm B_{\text{loc}}) = \; B_0 \pm \tfrac{3}{2}\mu r^3\left(3\cos^2\theta - 1\right) \tag{3}$$

The $\pm$ sign results from the fact that the local field may add to or subtract from B_0 depending upon whether the neighboring dipole is aligned with or against the direction of B_0. We recall that the net polarization of nuclear moments along B_0 is ordinarily only a few parts per million, and so in a large collection of such isolated pairs the probabilities of a neighboring dipole being aligned with or against B_0 are almost equal. Equation (3) expresses the fact that the spectrum of this proton-pair system will consist of two equal lines whose separation, at a fixed value of θ, will vary inversely as r^3. Only when the orientation of the pairs is such that $\cos^2\theta = \frac{1}{3}(\theta = 54.7^\circ)$ will the lines coincide to produce a single line. The spectrum for a hypothetical system is shown in Figure 1(a). Most materials of interest are not macroscopically ordered, so all values of θ are present in the sample, and the spectrum corresponds to an average over all values of θ, as shown in Figure 1(b). The observed line shape is known as a ''Pake'' pattern.

Protons are abundant in most polymers so it is unlikely that pairs of protons are completely separated from their neighbors. A more likely situation is that a proton has several neighbors that differ in both distance and orientation. Each dipolar coupling gives rise to a Pake pattern, but since they all overlap, the spectrum commonly appears as a broad, unresolved line, as shown in Figure 1(c).

We have so far assumed that the nuclei are fixed in position. If *molecular motion* is allowed to take place, i.e. by raising the temperature of the solid, by melting it, or by dissolving it in a mobile solvent, the variables

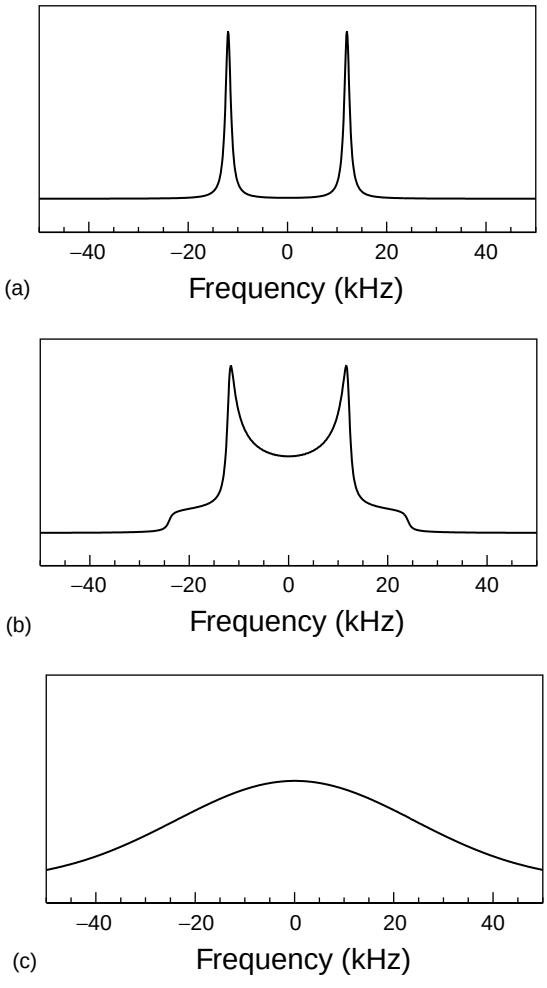

Figure 1 Schematic diagram of the spectrum from dipolar couplings for (a) an isolated spin pair with a single orientation relative to the magnetic field, (b) a collection of isolated spin pairs with a random orientation relative to the field, and (c) a distribution of distances and angles.

in Equation (3) become functions of time. If we assume that r is constant and only θ varies with time, as would be true for pairs of protons with a rigid molecule, then the time-averaged local field will be given by Equation (4):

$$B_{\text{loc}} = \frac{\mu r^{-3}}{T_2} \int_0^{T_2} \left(3\cos^2\theta - 1\right) \mathrm{d}t \tag{4}$$

where T_2 is the time that the nucleus resides in a given spin state. If θ can vary rapidly over all values, this time average can be replaced by a space average[4] (Equation 5):

$$B_{\text{loc}} = \mu r^{-3} \int_0^{\pi} \left(3\cos^2\theta - 1\right) \sin\theta \, \mathrm{d}\theta = 0 \tag{5}$$

Thus, if the correlation time τ_c is so short that this space averaging is valid, then the net effect of the neighboring magnetic nuclei is effectively erased, and the line will be drastically narrowed (by a factor of 10^4–10^5) compared with its rigid lattice value. If we begin with a solid at a temperature so low that it is virtually a rigid lattice, and then permit molecular motion (chiefly rotation) to occur with increasing frequency by raising the temperature, a narrowing of the resonance line will observed when (Equation 6):

$$\frac{1}{\tau_c} \geq 2\pi\delta\nu \tag{6}$$

where $\delta\nu$ is the static line width (in hertz). In liquids of ordinary viscosity, molecular motion is so rapid that Equation (5) holds; the local variations in magnetic field strength have become so short lived that *motional averaging* is complete.

There is another aspect of the interaction of neighboring magnetic dipoles that is closely related to the effects we have just considered and must also be considered in relation to line broadening. We must recall that our nuclear spins are not merely small static magnetic dipoles, but that even in a rigid solid they are precessing about the field direction. We may resolve a precessing nuclear moment into a static component along the direction of B_0, which we have considered so far, and into a rotating component, whose effect we must now also consider. This rotating component constitutes the right type of magnetic field to induce a transition in a neighboring nucleus if this is precessing at the same frequency. If this spin exchange or flip-flop occurs, the nuclei will exchange magnetic energy states with no overall change in the energy of the system but with a shortening of the lifetime of each. The magnitude of local field variations may be taken as μ/r^3, and consequently the relative phases of the nuclei will change in a time of the order of hr^3/μ^2, the "phase memory time". From Equation (6), we expect an uncertainty broadening of about μ^2/hr^3, i.e. of the same form and order of magnitude as that produced by the interaction of the static components of the nuclear moments. It has become customary to include both effects in the quantity T_2 (Equation 7):

$$T_2 = \frac{1}{\pi\delta\nu} \tag{7}$$

which we defined as the spin lifetime or the spin–spin or transverse relaxation time. A detailed theory of its dependence on molecular correlation time has been given by Bloembergen et al.[5] At short correlation times, as in mobile liquids (where "phase memory" is short), it becomes equal to T_1, but after T_1 passes through its minimum, T_2 continues to decrease as molecular motion becomes slower and finally levels out as the system begins to approach a rigid lattice.

Spin–spin exchange and dipolar broadening should not be considered as merely two alternative ways of looking at the same phenomenon, closely interrelated though they are. For example, in a lattice composed of magnetic nuclei α containing a dilute distribution of a different magnetic nuclear species β, spin exchange between nuclei α and β cannot occur since they precess at greatly different frequencies, but nevertheless, dipolar broadening will be present. Since the spin–spin exchange is very efficient in rigid solids, magnetization can be exchanged over a rather large length scale on the timescale of the spin–lattice relaxation. This process is known as *spin diffusion* and has been successfully used to study the length scale of phase separation in polymers.

From what we have said so far, it might be supposed that at any particular rf w_0 all nuclei of a given species, for example, all protons, would resonate at the same value of B_0. The total range of variation of B_0 for protons is about 13 ppm or ca. 26 500 Hz in a 11.7 T field (500 MHz). This spread in chemical shifts allows for the discrimination based on the chemical structure (methyl, methylene, etc.) and, in favorable cases, for the identification of defect structures. The corresponding ^{13}C spectrum is much larger. In both spectra the intensities of the peaks are proportional to the number of nuclei of each type.

The origin of this variation in resonant field strength is the cloud of electrons about each of the nuclei. When a molecule is placed in a magnetic field B_0, orbital currents are induced in the electron clouds, and these give rise to small, local magnetic fields, which, by a well-known physical principle, are always proportional to B_0 but opposite in direction. Such behavior is common to all molecules and gives rise to the universally observed diamagnetic properties of matter. Each nucleus is, in effect, partially shielded from B_0 by the electrons and requires a slightly higher value of B_0 to achieve resonance.

This can be expressed as Equation (8):

$$B_{\text{loc}} = B_0 (1 - \sigma) \tag{8}$$

where B_{loc} is the actual local field experienced by the nucleus and σ is the *screening constant*, expressing the reduction in effective field. The value of σ is independent of B_0 but highly dependent upon chemical structure.

The causes of variations in nuclear shielding may be thought of as arising from variations in electron density from nearby electron-withdrawing or -donating groups. They are also affected by special shielding effects produced by certain groups and structures that allow circulation of electrons only in certain preferred directions within the molecule. Benzene rings, for example, show this behavior very strongly, and many other groups do so in varying degrees, including even carbon–carbon double bonds. These effects are in general smaller than inductive effects, but nonetheless can be easily observed and can provide valuable structural clues. Protons, because of their low density of screening electrons, show smaller variations than do other nuclei.

It has been observed that the carbon chemical shift depends not only on the type of carbon and attached groups, but also on the chain conformation through the γ-gauche effect. The γ-gauche effect is believed to arise from the proximity of a carbon atom with substituents separated by three intervening bonds (the γ-neighbor). While there is no universal agreement on the origin of the γ-gauche effect, it has been suggested that it arises from van der Waals interactions between the attached hydrogens leading to polarization of the CH bonds.[6] These atoms are closest in the gauche (3 Å) conformation and more distant (4 Å) in the trans conformation. In polymers in solution there is relatively free rotation about the carbon–carbon bonds in the main chain, so many gauche and trans conformations are sampled on the timescale of the carbon chemical shifts and the measured shift depends on the average number of gauche and trans conformers. The measured chemical shift has contributions from all of the γ-gauche interactions. Such conformational averaging is not possible in solids, and the polymer signals are often broadened from a distribution of conformations. This is one of the reasons that the signals for amorphous polymers are often broader than their crystalline counterparts. As might be expected for van der Waals or other types of through-space interactions, the magnitude of the induced shift depends on the group at the γ-position, and more polarizable groups like chlorine and hydroxyl groups have larger effects than do carbon atoms.

The molecular screening constant is actually *anisotropic* or directional, as it depends on the orientation of the molecule with respect to magnetic field direction. It is expressed as a tensor, a mathematical quantity having both direction and magnitude, and is composed of three principal components, σ_{ii} (Equation 9):

$$\sigma = \lambda_{11}^2 \sigma_{11} + \lambda_{22}^2 \sigma_{22} + \lambda_{33}^2 \sigma_{33} \tag{9}$$

where λ_{ii} are the direction cosines of the principal axes of the screening constant with respect to the magnetic field. The principal axis systems may lie along the bond direction, but this is not necessarily so. The orientation of the axis system cannot be predicted a priori, and must be determined experimentally. By convention, the lowest field resonance is taken as σ_{33}. In some cases σ_{xx}, σ_{yy}, and σ_{zz} are used in place of σ_{11}, σ_{22}, and σ_{33}.

If the molecule is oriented so that the principal axis is along the field direction, then the observed chemical shift will correspond to that screening constant. For any arbitrary orientation the chemical shift is prescribed by Equation (9). In solution, the screening constant is given by the isotropic average of this equation. Since this is $\frac{1}{3}$ for each λ_i^2, we have (Equation 10):

$$\sigma = \tfrac{1}{3} (\sigma_{11} + \sigma_{22} + \sigma_{33}) \tag{10}$$

As we shall see, such averaging also occurs when the solid is rotated at the magic-angle.

Most organic solids are polycrystalline, the crystals and their constituent molecules being oriented at all angles with equal probability. Under these circumstances, the screening constant – or chemical shift – takes on a continuum of values forming the line shapes shown in Figure 2(a) and (b). The curve with sharp edges represents the theoretical line shape and the smooth curves represent the experimental line shape, in which a Lorentzian broadening function has been incorporated. Figure 2(b) shows a general, asymmetric powder line shape where the three principal values of the screening tensor all differ. In Figure 2(a) the chemical shift is axially symmetric; the screening constants σ_{11} and σ_{22} are equal and may designated as $\sigma_{\perp}$, σ_{33} being then designated as $\sigma_{\parallel}$. If instead $\sigma_{22} = \sigma_{33}$, as may also happen, the pattern will be reversed left to right. σ_{33} is customarily taken as the smallest value. The observation of such patterns does not actually tell us the orientations of the principal axis of the tensor with respect to the molecular framework. To determine this one must study single crystals or consider symmetry questions and relationships to other known molecules.

1.2 Solid-state Nuclear Magnetic Resonance Methods

1.2.1 Magic-angle Spinning

The NMR spectra of solids are fundamentally different from those in solution, primarily because the polymer matrix inhibits the molecular motion that averages the

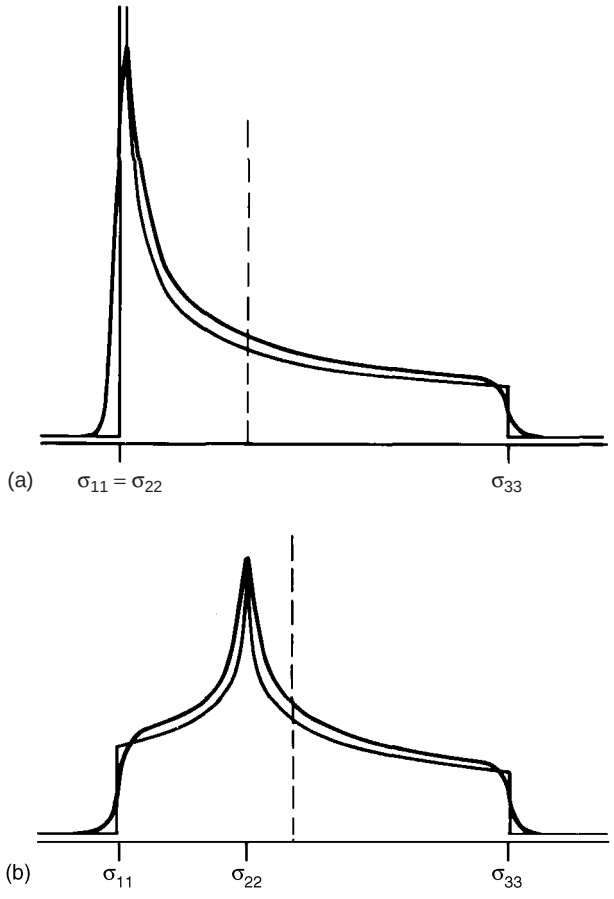

Figure 2 Chemical shift anisotropy pattern for (a) axially symmetric and (b) nonaxially symmetric atoms.

chemical shift anisotropy and dipolar interactions and results in sharp lines. Chemical shift anisotropy is a particularly severe problem when observing some nuclei, such as carbon, nitrogen and phosphorus, because they have a large chemical shift range. Most polymer samples are not oriented relative to the field and chemical shifts take on a continuum of values and form the line shapes shown in Figure 2(a) and (b).

The chemical shift anisotropy arises from a non-spherical distribution of electrons that gives rise to an orientation-dependent chemical shift in isotropic samples. In molecules of any degree of complexity there will be several such patterns that may strongly overlap, producing a broad, unresolved spectrum. Under these circumstances it usually becomes necessary to sacrifice the anisotropy information in order to observe a high-resolution spectrum. Under rapid sample rotation the orientations and chemical shifts become time dependent in the rotor period. Taking the time average under rapid rotation gives Equation (11):

$$\sigma = \tfrac{1}{2}\sin^2\beta\,(\sigma_{11} + \sigma_{22} + \sigma_{33}) + \tfrac{1}{2}\left(3\cos^2\beta - 1\right) \qquad (11)$$

where β is the angle between the rotation axis and the magnetic field direction. When β is equal to the so-called magic-angle (54.7°, or the body diagonal of a cube), $\sin^2\beta$ is $\frac{2}{3}$ and the first term becomes equal to one-third of the trace of the tensor (i.e. the isotropic chemical shift) and the $(3\cos^2\beta - 1)$ term is equal to zero. Thus, under magic-angle rotation the chemical shift pattern collapses to the isotropic average, giving the high-resolution spectrum.

The magic-angle spinning spectrum is obtained by taking the sample and packing it in a rotor that can be spun at several kilohertz. The effect of magic-angle spinning on the chemical shift anisotropy pattern is illustrated in Figure 3 for hexamethylbenzene. In addition to the isotropic peaks, spinning sidebands are observed if the sample rotation is not fast compared with the width of

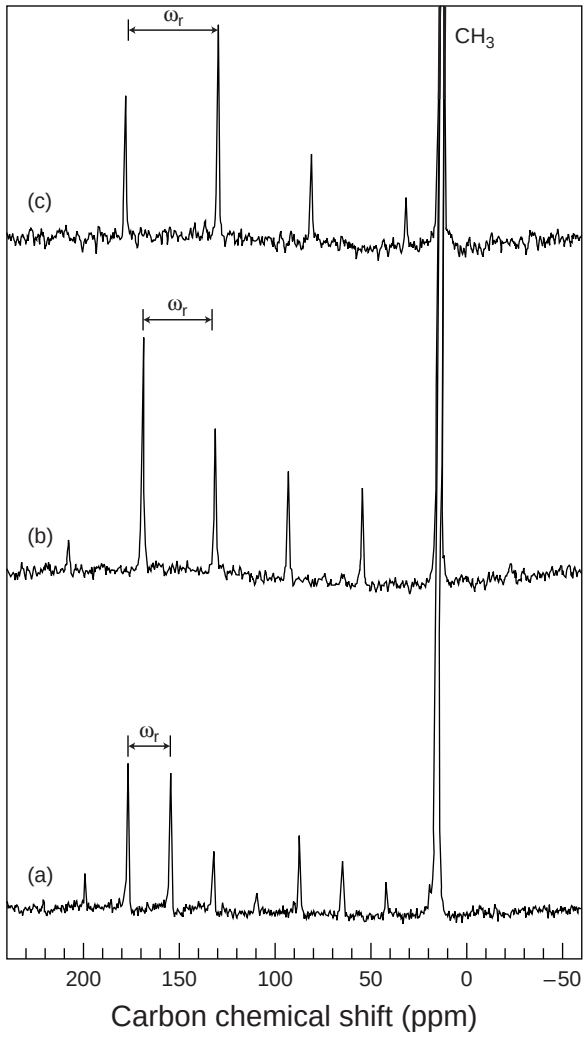

Figure 3 Effect of spinning on the carbon spectrum of hexamethylbenzene at three different spinning speeds. The spinning sidebands for the aromatic carbon are shifted from the isotropic position by the spinning speed.

the chemical shift anisotropy, and these sidebands are offset from the isotropic peak by spinning frequency. The intensities of the sidebands depend on both the spinning speed and the chemical shift anisotropy pattern, and at slow spinning speeds the sidebands trace out the chemical shift anisotropy pattern.

1.2.2 Cross-polarization

A single 90° pulse applied at the resonant frequency produces a free induction decay (FID), and the peak intensities contain quantitative information if we wait for five times the T_1 between scans for the spin system to return to equilibrium. In solids, however, molecular motion is limited and spin–lattice relaxation can be very inefficient. Even in solids that contain rapidly rotating methyl groups, the T_1 values for the other carbons can be prohibitively long. If the T_1s are of the order of hundreds of seconds it becomes difficult to acquire a spectrum with good signal-to-noise ratio if we are forced to wait five times the T_1 between scans.

Cross-polarization is the method most commonly used for signal enhancement in solid-state NMR. It takes advantage of the fact that proton spin diffusion generally causes all of the protons in a solid to have the same T_1, and that the proton T_1s are usually short compared with the carbon T_1s. In the absence of other factors, magnetization transfer between carbons and protons is not very efficient in solids because they have very different resonant frequencies. Cross-polarization works by forcing the proton and carbon signals to precess at the same frequency in the rotating frame even though they do not have the same frequency in the laboratory frame. The means of doing this was demonstrated by Hartmann and Hahn in 1962,[7] when it was shown that energy transfer between nuclei with widely differing Larmor frequencies can be made to occur when Equation (12):

$$\gamma_C B_{1C} = \gamma_H B_{1H} \tag{12}$$

the so-called *Hartmann–Hahn condition* is satisfied. Since γ_H is four times γ_C, the Hartmann–Hahn match occurs when the strength of the applied carbon field B_{1C} is four times the strength of the applied proton field B_{1H}. When the proton and carbon rotating frame energy levels match, polarization is transferred from the abundant protons to the rare carbon-13 nuclei. Because polarization is being transferred from the protons to the carbons, it is the shorter T_1 of the protons that dictates the repetition rate for signal averaging.

Figure 4 shows the pulse sequence diagram the carbon–proton cross-polarization experiment. The pulse sequence begins with a 90_x pulse to the protons to tip them along the y'-axis. The phase of the proton B_1 field is

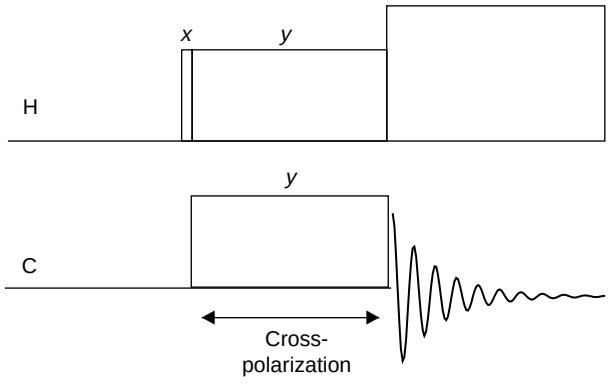

Figure 4 Pulse sequence diagram for cross-polarization.

then shifted by 90° and the protons are *spin-locked* along the y'-axis for the duration of the spin-locking pulse, and are forced to precess about the y'-axis of their rotating frame with a frequency $w_H = \gamma_H B_{1H}$. Meanwhile, the carbons are put into contact with the protons by turning the carbon field B_{1C} on during the spin-lock time. This causes the carbon magnetization to grow in the direction of the spin-lock field. The carbons are now precessing about their y'-axis with a frequency $w_C = \gamma_C B_{1C}$. Because of this common time dependence, mutual spin flip-flops can occur between protons and carbons, a process that can be visualized as a flow of polarization from the abundant proton spins to the rare carbon spins. It can be shown that under these circumstances the rare spin signal intensity is enhanced by as much as the ratio of magnetogyric ratios of the abundant and rare spins. In the ^1H–^{13}C case this factor is γ_H/γ_C or 4. This is in addition to the great advantage of not having to deal with long carbon-13 spin–lattice relaxation times in spectrum accumulation.

The signal intensity during cross-polarization represents a compromise between several competing processes, including the buildup of intensity from carbon–proton cross-polarization and the decay of the proton and carbons magnetization under the spin locking fields. The intensity at any given cross-polarization contact time t is given by Equation (13):

$$M = \frac{M_0}{T_{1C}} \frac{e^{-t/T_{1\rho}(H)} - e^{-t[1/T_{CH}+1/T_{1\rho}(C)]}}{1/T_{CH} + 1/T_{1\rho}(C) - 1/T_{1\rho}(H)} \tag{13}$$

where T_{CH} is the carbon–proton cross-polarization time constant and $T_{1\rho}(C)$ and $T_{1\rho}(H)$ are the carbon and proton rotating frame spin–lattice relaxation time constants. In most cases $T_{1\rho}(C) \gg T_{CH}$, so Equation (13) simplifies to Equation (14):

$$M = \frac{M_0}{T_{1C}} \frac{e^{-t/T_{1\rho}(H)} - e^{-t/T_{1C}}}{1/T_{1C} - 1/T_{1\rho}(H)} \tag{14}$$

The time course of the magnetization is a rapid buildup in signal intensity due to T_{CH} followed by a slower decay due

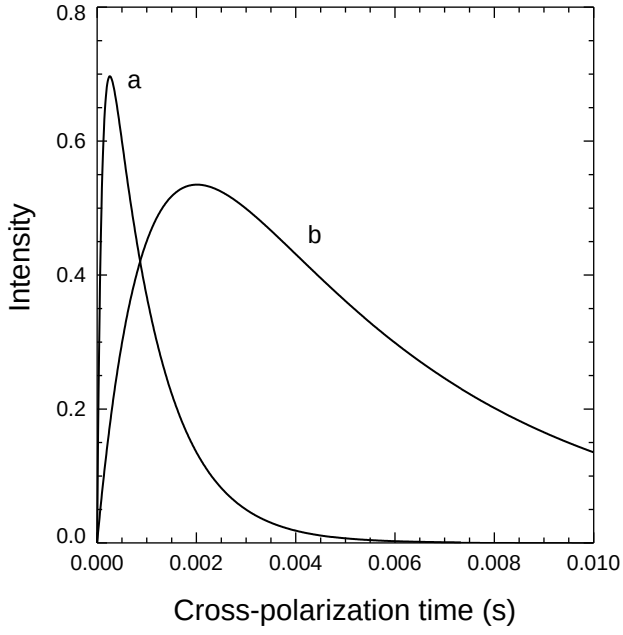

Figure 5 Buildup and decay of magnetization in a cross-polarization experiment for samples with (a) short T_{CH} and $T_{1\rho}$(H) and (b) longer T_{CH} and $T_{1\rho}$(H) values corresponding to crystalline and amorphous materials.

to $T_{1\rho}$(H). The signal intensity depends on the relaxation rate constants, which in turn depend on the chain dynamics. It is often observed that the buildup and decay of magnetization are different for crystalline, amorphous and rubbery polymers, and Figure 5 shows the time course of cross-polarization for two very different polymers. It is clear from Figure 5 that the time dependence of the cross-polarization is one way to separate the signals from polymers with different chain dynamics.

1.2.3 Decoupling

The carbon spectrum is often of interest in NMR studies of polymers, but the carbon lines are broadened from dipolar interactions with nearby protons. The dipolar coupling is often strong, and the line widths are of the order of 30–50 kHz. A practical method for the removal of proton dipolar broadening is to employ a high-power proton decoupling field, i.e. *dipolar decoupling*. This is analogous to the decoupling commonly used in solutions to remove the $^{13}C-^{1}H$ scalar couplings, except that much greater power is required.

The line narrowing from dipolar decoupling can be increased by using stronger rf fields, and decoupling fields of 100 kHz are now becoming routine. The limit on the decoupling field is often not the amplifiers, but rather the ability of the probe to take the high power. It has recently been reported that better decoupling can be obtained using lower powers through phase modulation of the decoupling field with the so-called two-pulse phase modulation (TPPM).[8]

1.2.4 Nuclear Magnetic Resonance Relaxation Measurements in Solids

The molecular dynamics of polymers are frequently studied by measuring the NMR relaxation rates that are sensitive the rate and amplitude of molecular motions. The spin–lattice relaxation times of carbons and nitrogens in polymers are of particular interest, but they can be extremely long (of the order of hundreds of seconds). This makes it very difficult to measure using the standard inversion–recovery method, where a waiting time of five times T_1 is required between scans. The carbon T_1s can be more efficiently measured using the cross-polarization T_1 experiment shown in Figure 6.[9] This experiment begins with the standard cross-polarization followed by a 90° carbon pulse to tip the carbon magnetization that was aligned along the y'-axis along the $-$z-axis. Following a relaxation delay (τ) the magnetization is returned to the y'-axis with another carbon 90° pulse, where it is observed with high-power proton decoupling. The greatest advantage of this experiment is that the delay between repetitions is determined not by the carbon T_1s but by the proton T_1s, which are of the order of 1–2 s. In addition, the signal intensity is enhanced by cross-polarization. However, the starting point in the relaxation experiment (after the first carbon 90° pulse) is a nonequilibrium state due to the signal enhancement from cross-polarization. The data are therefore recorded as a difference experiment by changing the phase of the first proton pulse by 180° every other scan and subtracting the FIDs to remove the effects of cross-polarization.

The proton T_1s are also of interest in solids. They are typically not as easy to interpret in terms of the molecular dynamics because of proton spin diffusion, but they provide valuable information about the length scale of phase separation. The proton T_1s can be measured

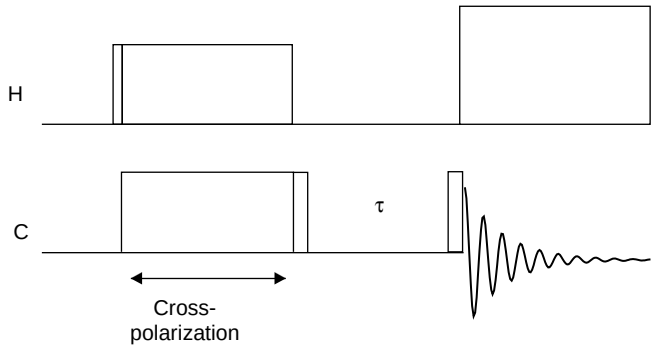

Figure 6 Pulse sequence diagram for measuring the carbon spin–lattice relaxation using cross-polarization.

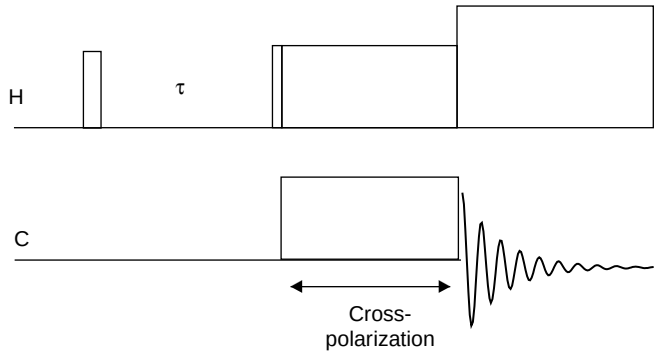

Figure 7 Pulse sequence diagram for measuring the proton spin–lattice relaxation rate using cross-polarization.

in the modified cross-polarization experiment shown in Figure 7. This pulse sequence differs from the normal cross-polarization sequence by the application of a 180° proton pulse and relaxation delay period (τ) before the cross-polarization begins. The 180° pulse inverts the proton magnetization and causes the spectrum to appear inverted relative to the equilibrium spectrum. The proton magnetization returns to equilibrium and the relaxation times are extracted using the normal analysis for the inversion–recovery experiment.

A related but distinct experiment enables one to measure the lifetime of the carbon and proton magnetization in the spin-locking field, a quantity characterized by $T_{1\rho}$, the rotating-frame spin–lattice relaxation time. As with the proton T_1s, spin diffusion makes it difficult to interpret the proton $T_{1\rho}$s in terms of the correlation times, but they can provide information about the length scale of phase separation. Under favorable conditions, the carbon $T_{1\rho}$s can provide information about the dynamics of polymers in the kilohertz frequency range that are often directly related to the mechanical properties of polymers.

Figure 8(a) and (b) show the pulse sequences for measuring the proton and carbon $T_{1\rho}$s. There are actually two closely related pulse sequences that can be used to measure the proton $T_{1\rho}$s. One method is to measure carbon intensity as a function of spin-locking time using the cross-polarization pulse sequence shown in Figure 4. The magnetization initially builds up to a maximum due to the carbon–proton dipolar interactions, and the signal decays exponentially to equilibrium with the time constant $T_{1\rho}$(H). Thus the $T_{1\rho}$(H) can be obtained from a semilogarithmic plot of the intensity as a function of spin-locking time after the signals have reached their maximum intensity. An alternative method is shown in Figure 8(a).[10] This pulse sequence differs from the normal cross-polarization in that after the 90° proton pulse the proton spin-locking field is turned on, but the carbon field is not. During this variable period τ, the proton magnetization decays with the time constant

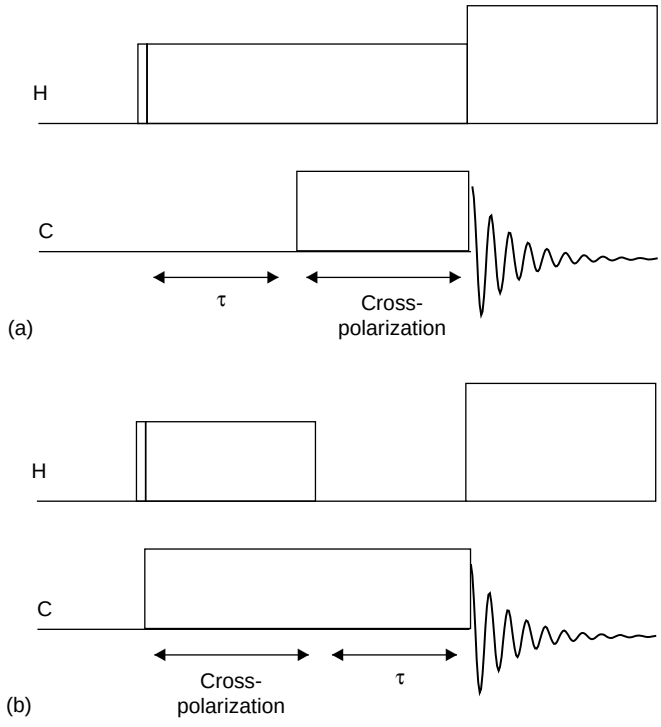

Figure 8 Pulse sequence diagram for measuring the (a) proton and (b) carbon $T_{1\rho}$.

$T_{1\rho}$(H). This decay is monitored by turning on the carbon spin-locking field to initiate cross-polarization followed by signal acquisition. The advantage of this sequence is that the $T_{1\rho}$(H) is obtained from a semilogarithmic plot of the signal decay and nonlinearities in the decay are easier to observe.

The carbon $T_{1\rho}$s are obtained in an analogous manner, as shown in the pulse sequence diagram in Figure 8(b). This pulse sequence begins as the normal cross-polarization with the 90° proton pulse and the spin-locking of the carbons and protons under the Hartmann–Hahn condition. After the carbon signal builds up from cross-polarization, the proton spin-locking field is turned off and the magnetization decays exponentially with the time constant $T_{1\rho}$(C).

1.2.5 Spectral Simplification in Solids

The high-resolution carbon spectra acquired with cross-polarization and magic-angle spinning can contain a large number of signals that can be difficult to assign by simple inspection and comparison to the solution spectra. In addition, there may be spinning sidebands from carbonyl or aromatic groups with large chemical shift anisotropies that overlap with the signals of interest. Hence it is often desirable to simplify the spectra. One simple method used for spectral simplification is *dipolar dephasing*, which is performed using the pulse

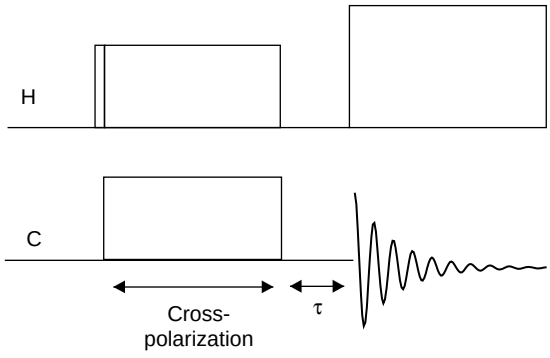

Figure 9 Pulse sequence diagram for dipolar dephasing.

sequence shown in Figure 9.[11] Dipolar dephasing differs from cross-polarization only in the variable delay period τ between the cross-polarization and data acquisition periods during which both the proton and carbon spin-locking fields are turned off. Since there is no proton decoupling during this period, the effective line width (and T_2) are those due to the carbon–proton dipolar coupling. The carbons with directly bonded protons have much stronger dipolar couplings and will decay much faster than those without directly bonded protons.

It is frequently not possible to spin polymer samples rapidly enough to remove all the spinning sidebands. Hence the spectra of polymers with several of these signals may be so complex that it is difficult to separate the isotropic signals from the spinning sidebands. While spinning sidebands can contain important information about the chemical shift anisotropy tensor, it is often desirable to simplify the spectrum by removing all but the signals at the isotropic chemical shift. This can be accomplished with the total suppression of sidebands (TOSS) pulse sequence shown in Figure 10.[12] This pulse sequence uses four 180° pulses applied at specific points during two rotor cycles to form echos leading to the

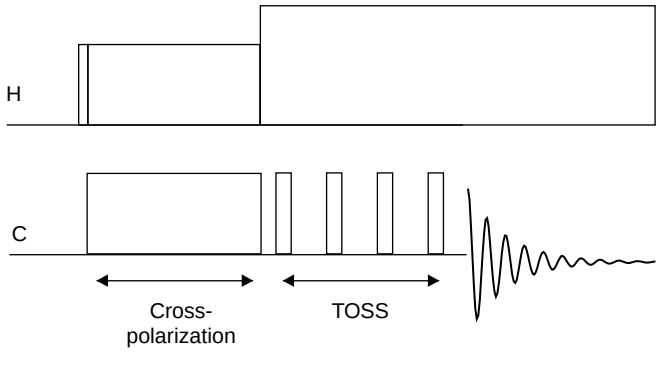

Figure 10 Pulse sequence diagram for sideband suppression using TOSS.

cancellation of all signals except those arising from the isotropic chemical shifts.

1.2.6 Two-dimensional Nuclear Magnetic Resonance

One of the fundamental limitations in solid-state NMR studies of polymers is signal overlap. While good resolution can be obtained with dipolar decoupling and magic-angle spinning, there is often overlap due to the repeating sequence nature of polymer chains and the features of interest, such as defect sites, etc. may not give resolved signals. One way to overcome these limitations is to expand the frequency information into two or more frequency dimensions (*n*D NMR). There are many kinds of *n*D NMR experiments that can be categorized as (a) *correlated* or (b) *resolved* experiments. In correlated experiments the resonance frequency of one signal is related to those of its neighbors and molecular connectivities or distances between atoms can be determined. In resolved experiments the frequency axes show two different interactions. In one kind of resolved experiment, for example, the carbon chemical shift may appear along one axis and the proton line width along the other. Extension of these same principles leads to (3D) NMR experiments in which there are three independent frequency axes. This has the potential of providing still greater resolution for those cases where the (2D) spectrum exhibits overlaps.

*n*D NMR experiments are related to the more familiar experiments in that they consist of a series of pulses and delays. They differ from the more common experiments in that we allow the spin system to *evolve* instead of immediately transforming the FID following the pulse. The pulse sequence for 2D NMR experiments can divided into four periods, *preparation*, *evolution*, *mixing*, and *detection*. During the preparation period, the spins are allowed to come to equilibrium, i.e. the populations of the Zeeman levels are allowed to equilibrate with their surroundings. This interval allows the establishment of reproducible starting conditions for the remainder of the experiment. During the evolution or t_1 period, the x, y, and z components evolve under all the forces acting on the nuclei, including the interactions between them. One of the things that commonly happens during the t_1 period is *frequency labeling*, where the spins are labeled (i.e. prepared in a nonequilibrium state) by their frequency in the NMR spectrum. The mixing period may consist of either pulses or delays and results in the transfer of magnetization between spins that have been frequency labeled in the t_1 period. The final period, t_2, is the signal acquisition that is common to all pulsed NMR experiments. The second frequency is introduced by systematically incrementing the t_1 period. Thus, data collection in a 2D experiment consists in

gathering many FIDs, each obtained with to a different value of t_1. The FIDs are transformed with respect to t_2 to obtain a set of spectra in which the peak intensities or phases are modulated as a function of the t_1 delay. Fourier transformation with respect to t_1 converts the frequency modulation into peaks in the 2D spectrum, which is actually a surface in 3D space and can be represented either as a stacked plot or as a contour plot. *n*D NMR experiments are performed by repeating the evolution and mixing periods $n-1$ times before acquisition. In a 3D experiment, for example, the sequence might be preparation–evolution (t_1)–mixing–evolution (t_2)–mixing–detection (t_3).

1.2.7 Wideline Nuclear Magnetic Resonance

Most solid-state NMR spectra are acquired using cross-polarization/magic-angle spinning (CPMAS) to obtain high-resolution. In some cases, however, it is desirable to measure the wideline spectrum. Little chemical information is obtained in these studies because the lines from different sites overlap, but the averaging of the line shape provides important information about the chain dynamics. Deuterium NMR is the wideline NMR method most often used to study the dynamics because the line shape depend on the rate and amplitude of molecular motion. Deuterium lines have a width of ca. 100 kHz and provide information about molecular motion on this timescale.

The wide lines in deuterium NMR present special experimental difficulties that require the use of special NMR methods. The signals have a very short T_2 and decay quickly after the rf pulses. After the application of a high-powered rf pulse there is a period of time known as the dead time that is required for probe ring down before the signals can be acquired. This is not a problem in high-resolution NMR because the T_2s are long compared with the dead time. However, the deuterium T_2s are of the same order of magnitude as the dead times and most of the signal will decay before the receiver is turned on, leading to an unacceptable distortion in the line shape.

Deuterium spectra can be acquired using the quadrupolar echo pulse sequence shown in Figure 11. The spectra are recorded using an echo so there is a delay period between the pulse and data acquisition to permit probe ring down. The echo delay times are typically of the order of 30 μs, and the spectrometer is adjusted so that the second half of the echo is recorded so the signal can be Fourier transformed in the normal way. In some cases the data are collected before the echo maximum and left shifted prior to Fourier transformation to make sure that data collection is not begun too late. Care must be taken to insure that the line shape and intensity do not change as a function of the echo delay time. This can happen in those cases where the polymer undergoes motion on a timescale near the echo delay time. The spectra are usually gathered at several values of the echo delay time to insure that the line shape is not affected by chain dynamics on this timescale.

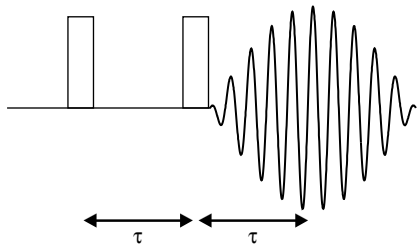

Figure 11 Pulse sequence diagram for the quadrupolar echo. The data are acquired after the second delay time τ.

1.2.8 Solid-state Proton Nuclear Magnetic Resonance

Important information about the chain conformation of polymers and the length scale of polymer mixing is contained in the proton solid-state NMR spectra. However, it is difficult to observe the spectra because the lines are broadened by strong dipolar interactions. The proton spectra can be acquired, but only by simultaneously decoupling and observing the proton spectra. This can be accomplished with the method known as CRAMPS,[13] using the pulse sequence shown in Figure 12. Proton decoupling is accomplished with a series of high-power rf pulses that rotate the magnetization quickly along different axes such that on average the magnetization is at the magic-angle. The dipolar couplings have the same functional form as the chemical shift anisotropy and disappear at the magic-angle. If the eight-pulse cycle is fast compared with the dipolar coupling (40 kHz), then the dipolar interactions are averaged and relatively sharp lines can be observed. The data are acquired by sampling a data point after every eight-pulse sequence cycle.

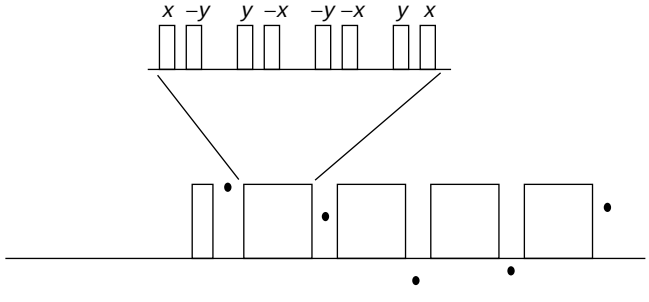

Figure 12 Pulse sequence diagram for combined rotation and multipulse spectroscopy (CRAMPS). The dots show how the data is collected during delays in the multiple pulse sequence.

2 HIGH-RESOLUTION SOLID-STATE NUCLEAR MAGNETIC RESONANCE STUDIES OF POLYMERS

2.1 Introduction

Most uses for polymers are in the solid-state, and understanding the properties of solid polymers has been an important factor driving the development of solid-state NMR methods. These methods have been developed over the past two decades and are now an important part of polymer science. The focus of many of these studies is a molecular level understanding of polymers in their functional state. The resolution of solid-state NMR is less than for solutions, but it has been shown that much useful information can be gathered by this method. As with solution NMR, there have been dramatic improvements in our ability to characterize materials by solid-state NMR with the introduction of 2D NMR and improvements in the NMR spectrometers. Solid-state NMR has become such an important method for polymer characterization that it is now considered a routine characterization tool.

Solid-state NMR has been used for polymer characterization because it can provide information about polymers over a wide range of length scales. On the most detailed level is the information about chain conformation that can be extracted from the chemical shifts. Unlike solution NMR, where the chemical shifts are averaged by molecular motion, the chains are more rigid in solid polymers below T_g, so the chemical shifts [through the γ-gauche effect[14]] reflect the distribution in chain conformation rather than the average chain structure. The organization of polymer chains on a longer length scale (20–200 Å) can be measured using proton *spin diffusion*. In such an experiment the proton signals in one section of the sample (such as the crystalline domains) are excited and magnetization transfer to other domains can be measured. This provides quantitative information because the rate of magnetization transfer depends on the length scale of phase separation.

2.2 Semicrystalline Polymers

A wide variety of NMR methods have been used to study semicrystalline polymers, including the acquisition of carbon, silicon, nitrogen, phosphorus, and proton spectra, NMR relaxation studies and multipulse and other NMR methods. Chain conformation is most commonly studied by carbon NMR with CPMAS. Although carbon has a relatively low sensitivity, it is abundant in most polymers and the spectra can be routinely acquired. Also, since the γ-gauche effect has been extensively studied in polymers,[15] the chemical shifts can provide important information about the chain conformation.

Polyethylene is a commercially important polymer that has been extensively studied by solid-state NMR. Polyethylene exists in an all-trans conformation in the crystalline phase and the spectra have been recorded as a function of crystallinity and sample preparation.[16] Figure 13(a) and (b) show the solid-state spectrum of polyethylene acquired using magic-angle spinning with and without cross-polarization.[17] The largest peak in the cross-polarized spectrum is a sharp resonance at 33.6 ppm that is assigned to crystalline polyethylene in the all-trans conformation, and the broader resonance located at 2.5 ppm to higher field is assigned to the amorphous material. Note that in the spectrum acquired without cross-polarization that the largest peak appears to be the amorphous material, since the relaxation times for the crystalline material are very long and the spectra do not accurately reflect the number of nuclei in the crystalline and amorphous phases. These spectra illustrate how the spectra of a particular phase in a complex material can be emphasized by the proper choice of NMR parameters and experiments, and that extreme care must be taken in the quantitative analysis of these materials. The sample studied in Figure 13(a) and (b) was 68% crystalline by X-ray diffraction.

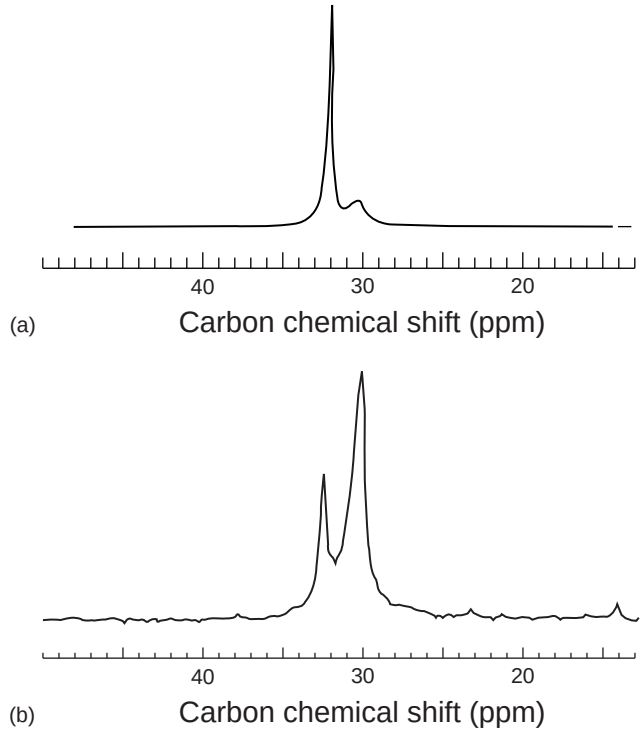

Figure 13 ^{13}C-NMR spectra of polyethylene acquired with magic-angle spinning and high-power decoupling (a) with and (b) without cross-polarization. (Adapted from McBrierty and Packer[17] by permission of Cambridge University Press, New York. Copyright 1993 American Chemical Society.)

The chemical shifts for the crystalline and amorphous fractions of polyethylene are in accord with those calculated from the γ-gauche effect.[14] In polyethylene in the melt or in solution, about 40% of the bonds are expected to be in the gauche conformation so an upfield shift of 5 ppm is expected. While this prediction is in the right direction, the magnitude of the shift is considerably less than expected. Among the possible explanations for this observation are that chains in the amorphous fractions have a different fraction of gauche bonds than expected on the basis of energy calculations, or that other factors, such as intermolecular interactions, also affect the chemical shifts in the solid-state. It has been observed that the chemical shifts for methylene carbons in *n*-alkanes can vary as much as 1.3 ppm even though they exist in an all-trans conformation.[18]

More recently, 2D double-quantum NMR has been developed as a method to determine the chain conformation in the solid-state.[19] The experiment is performed without sample rotation using ^{13}C-labeled spin pairs, and is similar in principle to the 2D incredible natural abundance double quantum transfer experiment (INADEQUATE) experiment used to measure carbon–carbon connectivities in solution.[20]

The idea behind this experiment is to correlate the chemical shift anisotropy patterns between pairs of carbons that are close in space. As noted in section 1.1, the singularities in the line shape are related to principal-axis system of the chemical shift tensor. For nearby carbons, there exist dipolar couplings that can be measured in the double-quantum spectrum, and the ridge pattern observed in the 2D double–single-quantum correlation spectra will depend on the relative orientations of the tensors in the labeled sites. This leads to a direct determination of the torsional angles.

Figure 14 compares the experimental double-quantum spectrum of polyethylene with the calculated spectra in the trans and gauche conformations.[19] The sample contained 4% doubly labeled polyethylene to minimize intermolecular magnetization transfer, and the spectra were recorded with a rather complex pulse sequence to remove both heteronuclear and homonuclear dipolar couplings during the detection period. The data clearly show the presence of the trans conformation for polyethylene. The amorphous polyethylene signal has been suppressed in these spectra.

The sensitivity of the NMR spectra to chain conformation is particularly striking for isotactic and syndiotactic polypropylene. Both polymers are crystalline, but the isotactic polymer adopts a *...gtgtgt...* 3_1 helical conformation while the syndiotactic polymer forms a 2_1 helix with a *...ggttggtt...* conformation. Half of the methylene groups lie along the interior of the helix in the syndiotactic polymer and are in a gauche arrangement with their

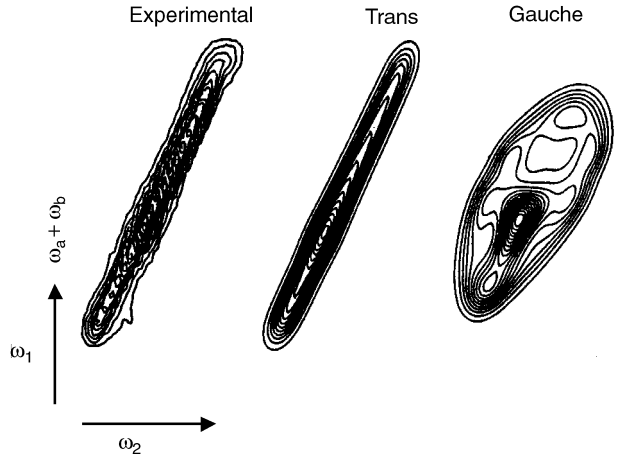

Figure 14 Comparison of the experimental single–double quantum correlation spectrum for polyethylene with simulated data for the trans and gauche conformations. (Adapted from Schmidt-Rohr[19] by permission of Academic Press, Inc. Orlando, FL, USA.)

γ-neighbor, and half the methylene groups lie on the exterior of the helix and are trans to their γ-neighbor.[14] In isotactic polypropylene, the methylene groups are trans to one γ-neighbor and gauche to another. The effects of such conformations are evident in Figure 15(a) and (b) which compare the CPMAS spectra for isotactic and syndiotactic polypropylene.[21] A single resonance is observed for the methylene groups in the isotactic polymer while two resonances separated by 8.7 ppm are observed for the syndiotactic material. This difference in chemical shift for the methylene carbons in the syndiotactic polymer is approximately as large as two γ-gauche effects.[14] The methylene resonance for isotactic polypropylene appears midway between the two peaks in syndiotactic polypropylene, as expected for a methylene group that has one γ-gauche interaction.

The crystal structure in some polymers depends on the crystal growth conditions, and NMR is an efficient way to monitor the state of the sample. This is illustrated in the 20 MHz natural abundance nitrogen NMR spectra for nylon 6 shown in Figure 16(a–c).[22] It is well known that nylon 6 can be prepared in several crystal forms, and it can be seen that significant chemical shift differences are observed between the α- and γ-forms, and that these forms can be identified by their chemical shifts in samples containing a mixture of the two forms. In this case, ^{15}N-NMR is more sensitive to the crystal structure than is ^{13}C-NMR, since it was not possible to distinguish between the two forms in the carbon spectra. Results for many crystalline mixtures have been reported, including those for poly(diethyloxetane)[15] by ^{13}C-NMR and polysilanes[23] by ^{29}Si-NMR.

Chain packing can also effect the chemical shifts of polymers in the crystalline state. Figure 17(a) and (b)

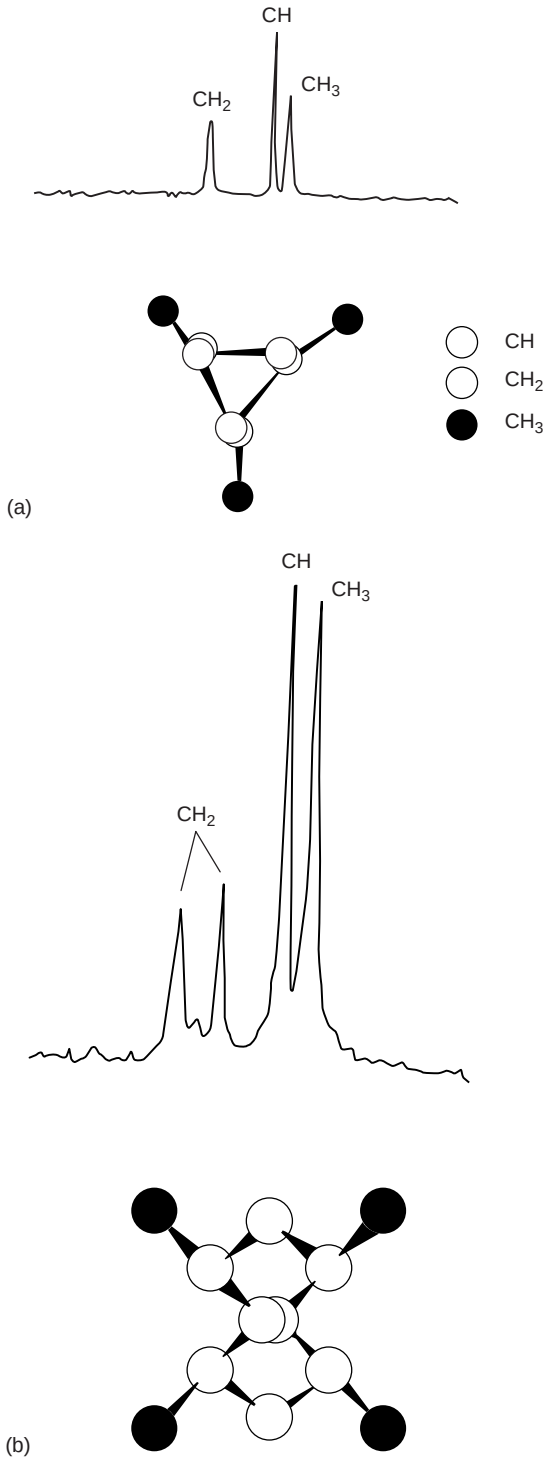

Figure 15 Solid-state carbon spectra of (a) isotatic and (b) syndiotactic polypropylene. (Adapted with permission from Bunn et al.[21] Copyright 1981 American Chemical Society.)

show the effect of crystal packing on the CPMAS carbon spectra of isotactic polypropylene.[24] When polypropylene is annealed above 150 °C, the stable α-polymorph is obtained, while unidirectional crystallization under a

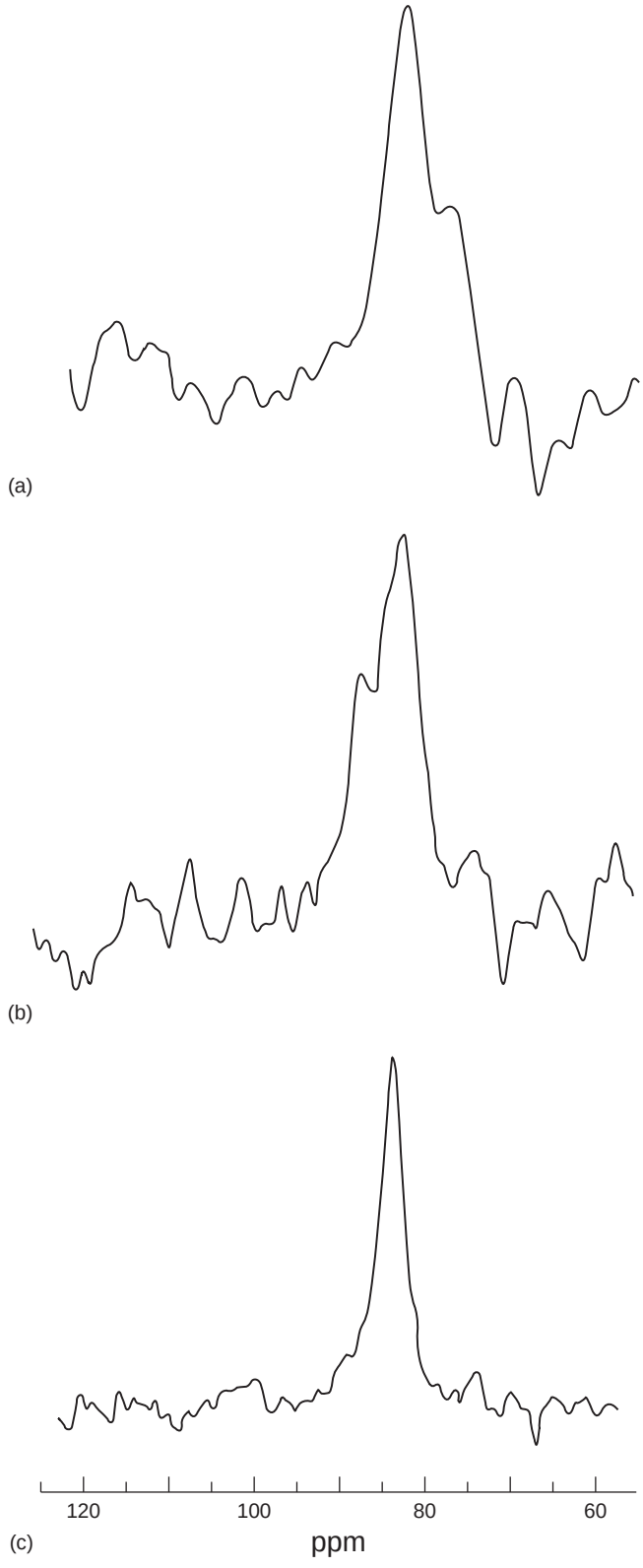

Figure 16 Solid-state ^{15}N spectra of nylon 6 in (a) predominantly the γ-form, (b) predominantly the α-form and (c) the α-form. (Adapted with permission from Powell et al.[22] Copyright 1988 American Chemical Society.)

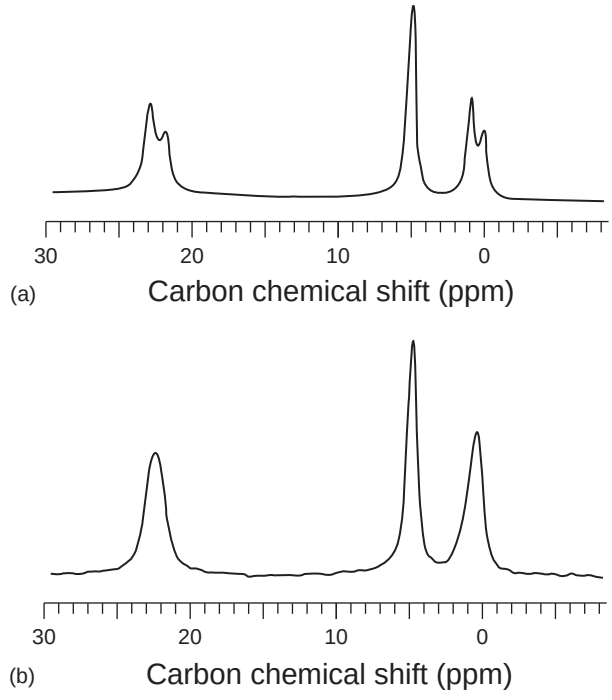

Figure 17 Solid-state carbon NMR spectrum of (a) the α- and (b) the β-form of isotactic polypropylene. (Adapted with permission from Gomez et al.[24] Copyright 1987 American Chemical Society.)

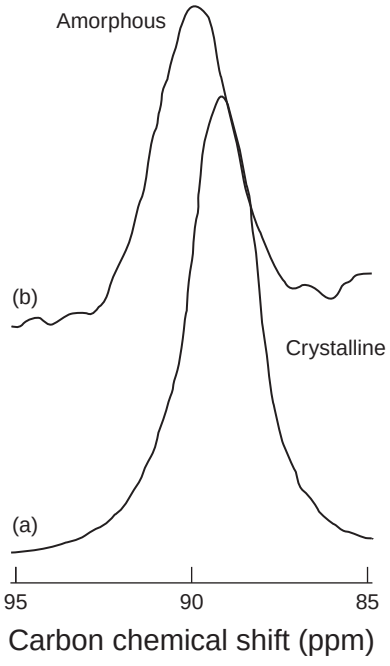

Figure 18 Solid-state carbon NMR spectra of poly(oxymethylene) obtained with magic-angle spinning and (a) cross-polarization and (b) dipolar dephasing. (Adapted with permission from Taylor et al.[123] Copyright 1992 American Chemical Society.)

strong temperature gradient yields the β-form. In both forms the chains adopt a 3_1 helix with a *...gtgtgt...* conformation, so the differences in the NMR spectra must result from chain packing effects. Such packing effects are a consequence of the packing of helicies with different handedness in alternate rows in the α-form, while helicies of the same handedness are clustered in the β-form. The 2 : 1 peak intensity ratios for the methylene and methyl carbons in the α-form corresponds to the ratio of nonequivelent packing sites. The two sites have interchain packing distances of 5.28 and 6.14 Å in the α-form and 6.36 Å in the β-form.

Several methods have been developed to differentiate between the crystalline and amorphous materials based on the differences in the cross-polarization dynamics, the T_1s, the $T_{1\rho}$s or the rates of dipolar dephasing in those polymers that have similar chemical shifts in the crystalline and amorphous phases.[25] One of the simplest methods is to use the differences in the spin–lattice relaxation times to increase the relative intensities of the crystalline or amorphous phases. This is accomplished either by choosing a short delay time between acquisitions, such that the crystalline component will be saturated, or by using the CPT_1 pulse sequence and choosing the delay time such that one of the signals is nulled.

Dipolar dephasing can also be used to distinguish between the crystalline and amorphous phases, since one consequence of the more rigid crystalline environment is that the dipolar couplings are much stronger. The important part of the dipolar dephasing pulse sequence is the period after cross-polarization during which the signals decay due to the proton–carbon dipolar interactions.[11] Figure 18 compares the CPMAS spectrum of poly(oxymethylene) with that obtained using dipolar dephasing.[25] In this case, the chemical shift difference between the crystalline and amorphous phases is too small to be observed without such methods. Very similar results have also been reported for poly(ethylene oxide), where the crystalline and amorphous phase can be distinguished on the basis of dipolar dephasing, the T_1 or the $T_{1\rho}$ relaxation times.[26]

2.3 Amorphous Polymers

NMR has been less extensively used to study chain conformation in amorphous polymers because by their nature they are conformationally disordered. However, NMR studies have provided a fundamental insight into the properties and reactivities of amorphous polymers.

The spectral features of amorphous polymers are usually more poorly resolved than are those of crystalline materials, as illustrated in the comparison of the crystalline and amorphous phases of polyethylene as shown in Figure 13(a) and (b). The broadening is due primarily to conformational effects, since there is

a distribution of conformations that are slowly interchanging on the NMR timescale, so the lines are inhomogeneously broadened by the γ-gauche effect. The broadenings are inhomogeneous so the resolution is not improved with higher magnetic fields, as might be expected for small molecules in solution.

Although conformational information is difficult to extract from the CPMAS spectra of amorphous polymers, there are many possible applications for solid-state NMR in the study of amorphous polymers. Solid-state NMR is an important tool to study the curing and reactivity in polymers, such as poly[(phenylsilylene)ethynylene-1,3-phenyleneethynylene]**(1)** which contains Si–H and acetylenic bonds and produces a highly thermally stable polymer by curing above 150 °C.[27] Figure 19 compares the ^{13}C cross-polarization spectra with magic-angle spinning for polymers cured at temperatures up to 400 °C. It is clear from these spectra that the acetylinic carbons with peaks at 87.1 and 107.6 ppm are converted to aromatic material, presumably from a Diels–Alder reaction. ^{29}Si-NMR has also been used to monitor the cure of these polymers.[27]

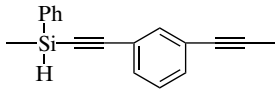

(1)

2.4 Solid–Solid-phase Transitions

Some polymers are stable in multiple crystalline conformations, and solid-state NMR is an effective means to monitor solid–solid-phase transitions because the changes in conformation are often accompanied by changes in chemical shifts due to the γ-gauche effect or crystal packing effects. 1,4-*trans*-Polybutadiene has long been known to adopt two crystalline forms that can be interconverted with temperature.[28] The form-I conformation is stable at ambient temperature and is well defined by X-ray diffraction. Form-I is reported to have an $s^{\pm}$–t–$s^{\pm}$ conformation for the bonds between the double bonds, where $s^{\pm}$ represents the skew ($\pm 109^{\circ}$) conformation. Form-II is observed above 75 °C, where the chains are packed in a hexagonal array but with a lower density and the equitorial reflections are blurred. Figure 20 shows the effect of temperature on the CPMAS spectrum of 1,4-*trans*-polybutadiene **(2)**, where the conversion from form-I to form-II can be easily followed by NMR.[29] The form-II resonances appear upfield from the form-I peaks, and the mixtures of the two forms can be easily observed at intermediate temperatures. Several models for the form-II conformation have been proposed, including one in which the skew angle is

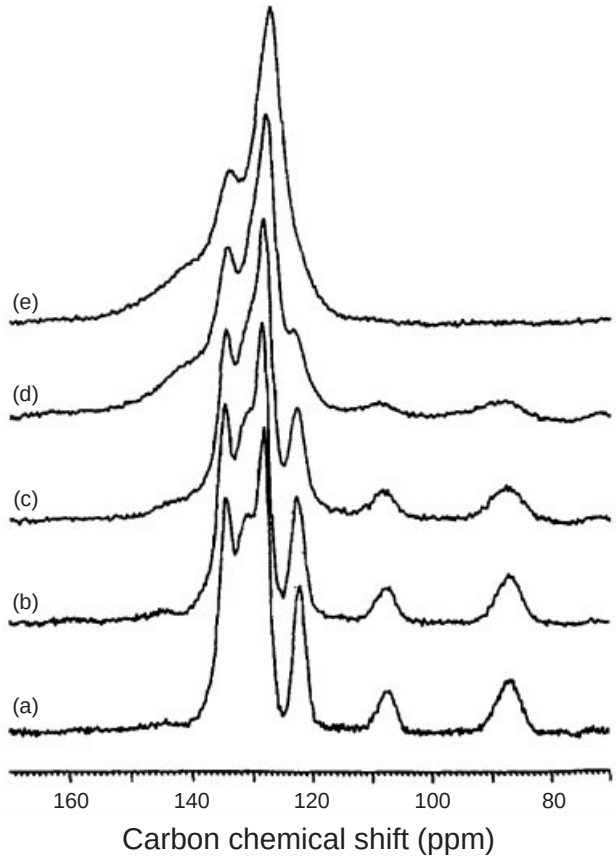

Figure 19 Solid-state carbon NMR spectrum of poly[(phenylsilylene)ethynylene-1,3-phenyleneethynylene] (a) before curing and cured at (b) 150, (c) 200, (d) 300, and (e) 400 °C. (Adapted with permission from Kuroki et al.[27] Copyright 1998 American Chemical Society.)

decreased from $\pm 109^{\circ}$ to $\pm 80^{\circ}$ to account for the contraction in chain-axis repeat distance observed in the X-ray diffraction.[30] It has also been suggested that the chains are conformationally disordered or that they are rapidly rotating about their long axes.[31] The NMR spectra and relaxation time measurements are most consistent with a model in which the chains are disordered in the form-II conformation.[29] The solid–solid-phase transitions for a number of polymers have been reported, including carbon NMR studies of diacetylenes[32] and silicon NMR studies of polysilanes.[33]

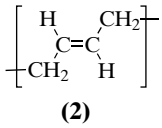

(2)

2.5 Elastomers

The NMR studies of elastomers are fundamentally different from those of other polymers studied by

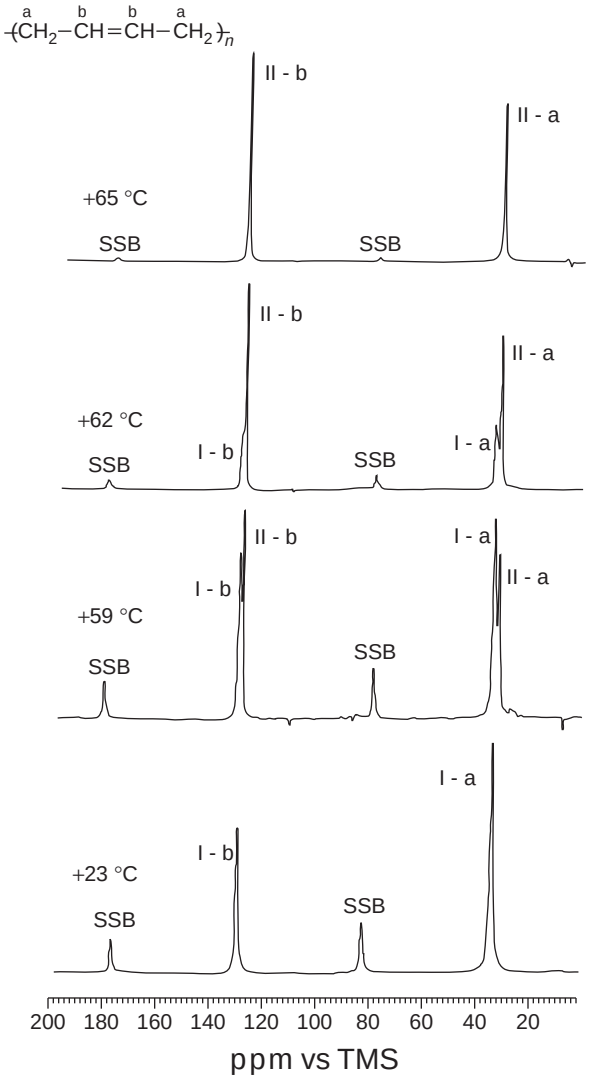

Figure 20 50 MHz carbon spectra of 1,4-*trans*-polybutadiene as a function of temperature showing the form-I to form-II transition. The spinning sidebands are marked. (Adapted with permission from Schilling et al.[29] Copyright 1987 American Chemical Society.)

solid-state NMR because they are usually studied at temperatures far above their T_g. Under these conditions, the dipolar interactions and chemical shift anisotropy that leads to line broadening are partially averaged by chain motion, and sharp lines are often observed with magic-angle spinning alone. In many cases the motional averaging is such that high-resolution signals can be observed with spinning, and peak assignments can be established using the traditional solution methods. In addition, the lines are often narrowed enough by molecular motion that it is possible to image the samples using magnetic resonance imaging (MRI).

The behavior of elastomers above T_g is illustrated in the proton spectra of *cis*-1,4-polybutadiene shown in

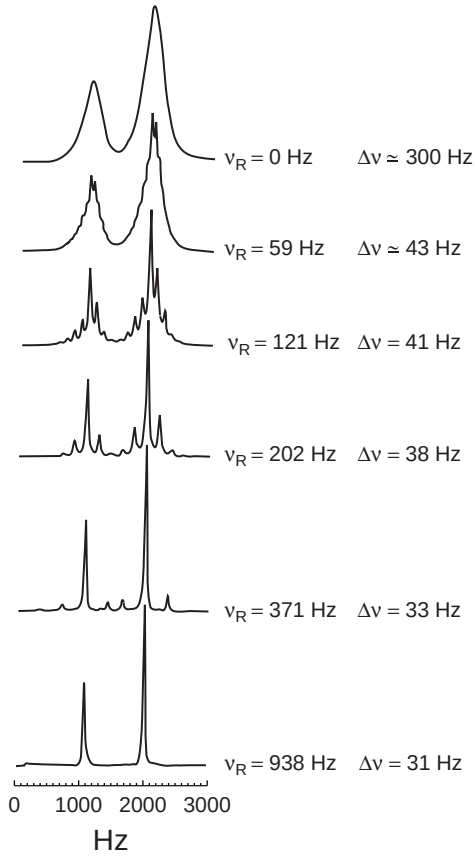

Figure 21 300 MHz magic-angle spinning proton spectra of *cis*-1,4-polybutadiene at several spinning speeds. (Adapted with permission from English.[34] Copyright 1985 American Chemical Society.)

Figure 21.[34] The top spectrum shows that two groups of signals can be resolved in the static proton spectrum, and that each line has a width of ca. 300 Hz. The residual broadening from dipolar couplings and chemical shift anisotropy can be further averaged with magic-angle spinning to 31 Hz. Once the lines have been narrowed by this method, these polymers can be analyzed by any of a number of solution NMR methods. The carbon assignments in elastomers can often be established using the spectral editing methods developed for solution NMR that are based on the number of attached protons and the carbon multiplicity.[35]

The line narrowing in elastomers makes it possible to study their structure by MRI. Imaging is very difficult for most polymers because the resolution in imaging experiments is determined both by the magnetic field gradient strength and the line width. Since the line width for polymers below T_g is of the order of 40 kHz and the T_2s are very short, the signals often decay during the delay periods in an imaging experiment.

NMR imaging is a promising a method that has been extensively used in the medical community, and is now

being used to study the structure of polymers. It has the advantage of being nondestructive and it is not limited to surface analysis or optically clear materials, but has some severe spatial resolution problems in comparison with the imaging of biological samples.[36] The resolution in imaging experiments is limited by the spin–spin relaxation times and line widths, which are much greater for polymers than for the water or fat that is typically imaged in biological samples.

MRI differs from most other NMR methods in that strong magnetic field gradients are used. This is in contrast to most high-resolution NMR studies where the field is made as homogeneous as possible to obtain the highest resolution. The magnetic field gradients are applied so that the resonance frequency for the same atoms can be distinguished if they are at different points in space.

The spatial resolution is obtained by applying field gradients that are deliberately inhomogeneous along the different field directions. The gradient is the spatial derivative of the field and is given by Equation (15):

$$G_l = \frac{\partial Bl}{\partial l} \tag{15}$$

where $l = x, y, z$. The resonance frequency for a spin in the magnetic field is given by Equation (16):

$$w = -\gamma|(1 - \sigma)B_0| \tag{16}$$

where B_0 is the magnetic field, σ is the isotropic chemical shift, and γ is the magnetogyric ratio. In the presence of a magnetic field gradient the frequency becomes a function of spatial position [$\mathbf{r} = f(x, y, z)$] and is given in vector notation by Equation (17):

$$w = -\gamma|(1 - \sigma)B_0 + \boldsymbol{Gr}| \tag{17}$$

This equation shows how the frequency of a resonance in a gradient is modified by its spacial position. This behavior is shown schematically in Figure 22(a) and (b), which compare the spectra for two samples in the presence and absence of a field gradient.[36] The resolution in MRI is determined by the line width and the gradient strength. Two points in space can be resolved if the line width Δw is less than the product of the magnetogyric ratio, the gradient field strength, and the spatial position (Equation 18):

$$\Delta w < \gamma \boldsymbol{Gr} \tag{18}$$

The success of biological imaging is due in large part to the narrow line widths of the water or fat that is imaged. The resolution in polymers will not be as great as in the biological samples, but much useful information can still be obtained.

A variety of pulse sequences have been developed that attempt to overcome the intrinsic imaging limitations

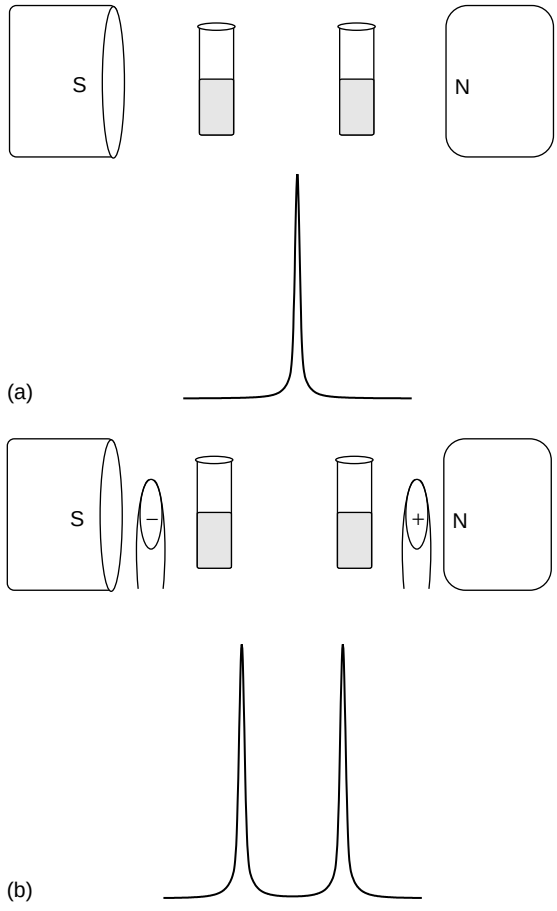

Figure 22 Schematic drawing showing two water samples in (a) a homogeneous field and (b) a homogeneous field with magnetic field gradients. Also shown are the resulting spectra for the two cases.

in solid materials. The choice of the pulse sequence depends on the characteristics of the sample and its NMR properties. The contrast in materials imaging is usually based on differences in relaxation times across the sample to enhance the signals of interest.[36]

Imaging in most materials is performed using 2D NMR methods. The major differences between these experiments and the ones introduced in section 1.2.6 are that MRI uses field gradients during the t_1 period and the acquisition time and the final spectrum is a plot of the intensity as a function of two spatial coordinates. It is possible to obtain 3D images, but more typically selective pulses and gradients are combined to image a plane through the sample.

The line widths in solid polymers are much larger than those for the fat or water signals that are used in biological imaging, and one of the important areas of research is finding new methods for imaging in materials with larger line widths.[37,38] There are many instances, however, where the line widths of polymers

are sufficiently small that the methods used for biological imaging can be used without modification. Among these are the study of liquids in a solid matrix, partially polymerized polymers or polymers at temperatures above T_g. Such studies also provide a valuable insight into voids within solid polymers, rates of solvent diffusion, and polymer flow. Another important area of research is the MRI of rigid solid polymers. Several approaches are currently being explored, including the use of magic-angle spinning[37] and multipulse decoupling.[38] In certain favorable cases, differences in the molecular dynamics or orientations of static samples can be used to obtain useful images.[36]

The quality of the image depends on the instrumental parameters, most notably the time it takes for the spectrometer to recover after a gradient is turned on or off, as well as the line widths and the relaxation times. The data acquisition is often limited by the gradient recovery time, the noisy period following the switching of a gradient on or off during which it is difficult to record a noise-free spectrum. Images are frequently acquired using a Hahn spin echo to detect the signals since the gradient recovery period is placed in the echo delay. This is known as a T_2-weighted image.

Contrast in materials imaging is obtained by taking advantage of the differences in relaxation times across the sample. Differences in the T_2s are most frequently used to enhance the contrast, although differences in T_1s and $T_{1\rho}$s can also be used in favorable circumstances.[39] Figure 23(a–d) show cross-sections through images of a tire tread containing natural rubber and styrene–butadiene rubber using the differences in the relaxation times to enhance the contrast of one material over the other.[39] The contrast differences are due to differences in relaxation times that are related to the differences in the molecular dynamics of the natural rubber and the styrene–butadiene rubber. The differences in the T_1s are not large enough to give contrast enhancement of one component over the other. In experiments where the image is weighted by the T_2 or the $T_{1\rho}$ relaxation times, more intensity is associated with the natural rubber relative the styrene–butadiene rubber, and natural rubber appears brighter in the final image.

The NMR imaging of elastomers has been used to investigate the morphology, composition, and defects in composite materials such as non-steel belted radial tires. Figure 24 shows an image of a cross-section through a tire in which the elastomer can be distinguished from the reinforcing fibers that have a much shorter T_2 and appear as dark circles in the T_2-weighted image.[40] The resolution in this experiment is ca. 200 μm and the image has some distortions due to the magnetic susceptibility variations associated with the voids and

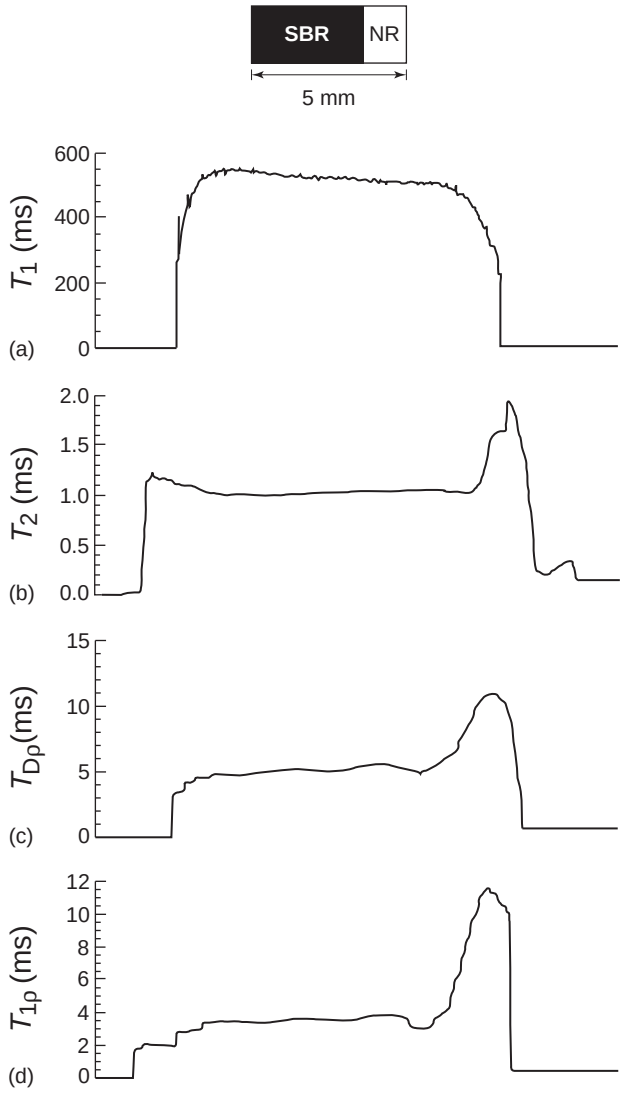

Figure 23 Cross-sections through an image of styrene–butadiene rubber and natural rubber obtained with different filtering pulse sequences applied before the imaging pulse sequence. The data are shown for (a) the T_1 filter, (b) the T_2 filter, (c) the dipolar filter and (d) the $T_{1\rho}$ filter. (Adapted from Blumler et al.[39] by permission of Elsevier Science, Oxford, UK.)

filler particles. It was also shown that the variations in intensity of the elastomeric portions are associated with composition variations in the rubber, and it is estimated that a 20% variation in the composition of the *cis*-polybutadiene/styrene–butadiene rubber can be identified in the T_2-weighted image.

2.6 Other Studies

Solid-state NMR methods are used to study a wide variety of polymeric materials in addition to the crystalline, amorphous and rubbery polymers discussed above. Some of these materials have molecular dynamics like

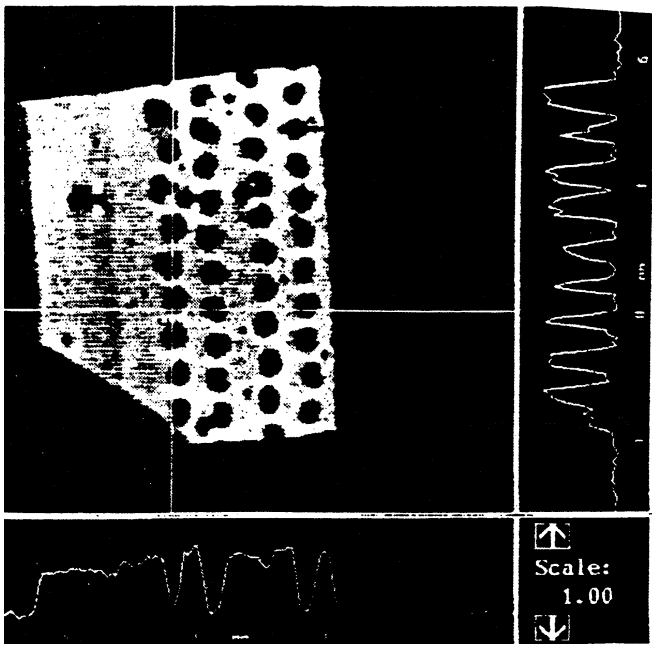

Figure 24 NMR image of a non-steel belted radial tire section. The bright portions of the image are due to the elastomer. (Adapted with permission from Sarkar, Komoroski.[40] Copyright 1992 American Chemical Society.)

elastomers and can be studied using the methods discussed in section 2.5. Among the materials of interest are gels, inclusion compounds, cross-linked polymers, polymers above T_g, melts, and ionomers. The NMR method of choice depends mainly on the molecular dynamics of the material under study. Since many of these materials are more mobile than rigid solids, cross-polarization and magic-angle spinning may not be the methods of choice.

Swollen cross-linked polymer gels are a class of materials that have chain mobilities intermediate between those of solids and liquids. One important example of these materials are the polystyrene derivatives that are often used for solid phase synthesis of peptides and other biomolecules. Figure 25(a) and (b) compare the static and magic-angle spinning proton NMR spectra of a derivatized poly(styrene–divinylbenzene) cross-linked gel swollen in chloroform.[41] The lines are effectively narrowed by magic-angle spinning and the resolution is comparable to that obtained for polystyrene in solution without cross-linking. A similar sharpening in the carbon spectra was also reported.[41] Since a liquid-like spectrum is observed with spinning, these materials can be characterized using the 1D (one-dimensional) and 2D methods developed for polymer characterization in solution.

The inclusion compounds formed between some polymers and certain matrix materials are another class of materials with mobilities intermediate between

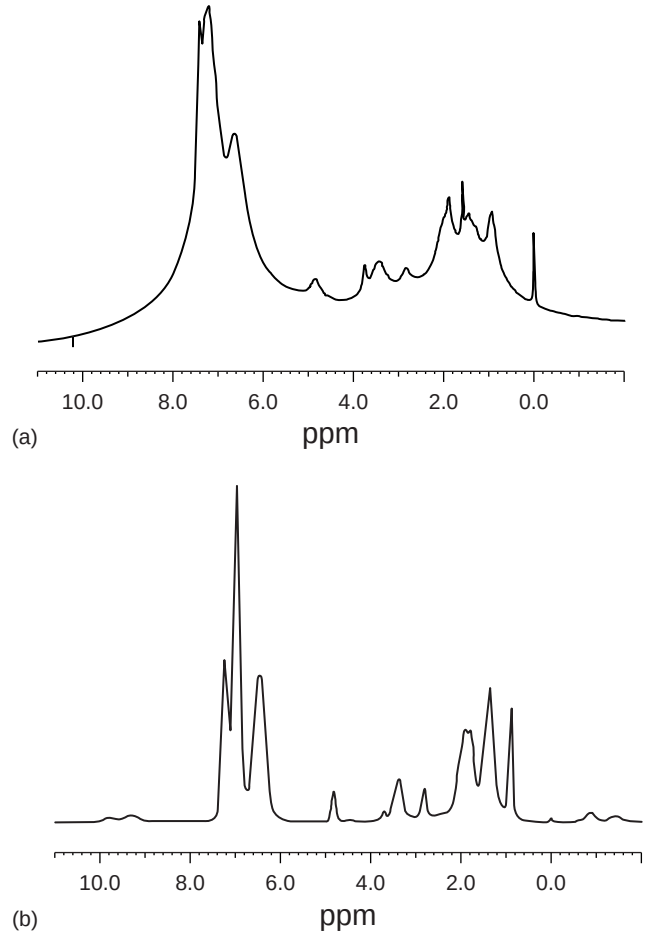

Figure 25 (a) Static and (b) magic-angle spinning proton NMR spectra of a derivatized poly(styrene–divinylbenzene) swollen in chloroform. (Adapted with permission from Stover and Frechet.[41] Copyright 1989 American Chemical Society.)

solutions and rigid solids. It has been observed that some small molecules, such as perhydrotriphenylene or urea, form stable crystal structures in which long-chain polymers are included into narrow channels. Solid-state carbon NMR has been used to study a number of such inclusion compounds, including those formed between perhydrotriphenylene and polyethylene,[42] *trans*-1,4-polybutadiene[43] and *trans*-1,4-polypentadiene.[44] Figure 26 shows the CPMAS spectrum of the perhydrotriphenylene – polyethylene inclusion compound.[42] As noted in section 2.2, polyethylene exists in an all-trans conformation in the crystalline phase and the methylene carbon resonance in the amorphous phase is shifted upfield by 2.5 ppm due to the presence of ca. 40% of the bonds in the gauche conformation. The chemical shift for the polyethylene inclusion compound (33.6 ppm) is the same as for the crystalline phase, demonstrating that the polymer exists in an all-trans conformation in the narrow channel. Since the relaxation

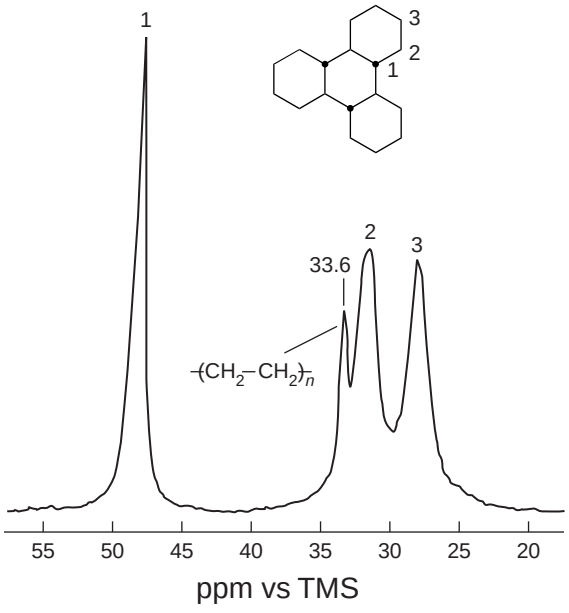

Figure 26 100 MHz carbon NMR spectrum of the polyethylene – perhydrotriphenylene complex obtained with magic-angle spinning and dipolar decoupling. The peaks labeled 1, 2 and 3 are assigned to the matrix. (Adapted with permission from Sozzani et al.[42] Copyright 1991 American Chemical Society.)

times for the polymer and the matrix material differ, it is possible to choose experimental conditions to enhance the signals from either the polymer or the matrix.[42]

3 THE SOLID-STATE ORGANIZATION OF POLYMERS

3.1 Introduction

NMR studies have been valuable not only to study the local structure or conformation of polymer chains, but also to measure the order over larger length scales. These measurements are made primarily by selective excitation of one portion of the sample (i.e. the rigid or mobile phase) and measuring the rate of spin diffusion to the other phase. If the intrinsic spin diffusion rate is known, then the average distance between the phases can be calculated for an assumed morphological model.

In most cases the selective excitation is based on a difference in the relaxation rates between the phases. In semicrystalline polymers, for example, there is typically a large difference in the spin–spin relaxation rates between the rigid crystalline and the more mobile amorphous phases. Therefore, if the magnetization from atoms in both phases are tipped into the xy plane, the signals from the rigid material with the short T_2 decays rapidly, leaving only the signals from the mobile phase. This is the basis

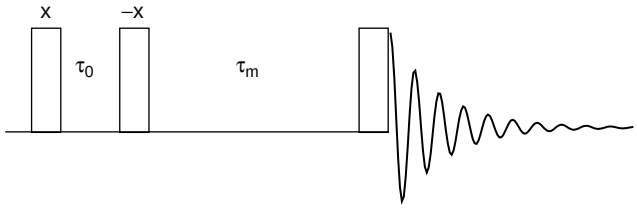

Figure 27 Pulse sequence diagram for the Goldman–Shen experiment.[45]

for the most commonly used pulse sequence to study the domain structure, the Goldman–Shen pulse sequence shown in Figure 27.[45] Following the initial 90° pulse, the magnetization from the rigid phase decays to zero due to T_2 relaxation during the delay period τ_0 whereas the magnetization from the mobile phase does not. The magnetization from the mobile phase is returned to the $+z$-axis with a pulse of opposite phase to the first one, creating a spin state in which the mobile phase is close to equilibrium, while the rigid phase is saturated. If the rigid and mobile material are in close proximity then the rigid phase can relax by magneitzation transfer to the mobile phase. The signals are observed by the final 90° pulse, and the data are typically acquired for several values of the spin diffusion delay time τ_m. In such an experiment the rigid signals will increase as a function of τ_m whereas mobile signals will decrease as a result of spin diffusion from the rigid to the mobile phase.

Information about the domain structure of polymers is obtained by solving the diffusion Equation (19):

$$\dot{m}(r,t) = D\nabla^2 m(r,t) \tag{19}$$

where D is the diffusion constant ($cm^2\,s^{-1}$) and $m(r,t)$ is the local magnetization density. The solutions to this equation are expressed in terms of the response function $R(t)$, which can be obtained from the measured intensities as a function of time[46] (Equation 20):

$$R(t) = 1 - \frac{M(t) - M(\infty)}{M(0) - M(\infty)} \tag{20}$$

where M is the intensity of the rigid portion of the material. The expression for $R(t)$ depends on the boundary conditions and on the dimensionality. The simplest cases are for 1D, 2D, and 3D diffusion, where $\bar{b}$ is taken to be the average domain dimension. For 1D diffusion, $R(t)$ is given by Equation (21):

$$\begin{aligned} R(t) &= 1 - \phi(t) \\ &= \left(\frac{2}{\sqrt{\pi}}\right)\left(\frac{Dt}{\bar{b}^2}\right)^2 \qquad \text{for } t \ll \frac{\bar{b}^2}{D} \\ &= 1 - \sqrt{\pi}\left(\frac{\bar{b}^2}{Dt}\right) \qquad \text{for } t \gg \frac{\bar{b}^2}{D} \end{aligned} \tag{21}$$

For 2D diffusion the expression for $R(t)$ is given by Equation (22):

$$\begin{aligned} R(t) &= 1 - \phi(t)^2 \\ &= \left(4\sqrt{\pi}\right)\left(\frac{Dt}{\bar{b}^2}\right)^{1/2} \qquad \text{for } t \ll \frac{\bar{b}^2}{D} \\ &= 1 - \sqrt{\pi}\left(\frac{\bar{b}^2}{\pi Dt}\right) \qquad \text{for } t \gg \frac{\bar{b}^2}{D} \end{aligned} \tag{22}$$

and for 3D diffusion the result is Equation (23):

$$\begin{aligned} R(t) &= 1 - \phi(t)^3 \\ &= \left(6\sqrt{\pi}\right)\left(\frac{Dt}{\bar{b}^2}\right)^{1/2} \qquad \text{for } t \ll \frac{\bar{b}^2}{D} \\ &= 1 - \left(\frac{\bar{b}^2}{\pi Dt}\right)^{3/2} \qquad \text{for } t \gg \frac{\bar{b}^2}{D} \end{aligned} \tag{23}$$

Physically, 1D diffusion applies to layer-like or lamellar structures, 2D diffusion is most applicable to rod-like structures, and 3D diffusion applies to spheres or cube-like structures. Figure 28 shows a plot of the response function against $\sqrt{\tau}$, which is equal to $\sqrt{Dt}/\bar{b}$, for 1D, 2D, and 3D diffusion.[46]

Spin diffusion is an important tool for measuring the length scale of phase separation in semicrystalline polymers, blends and a wide variety of phase-separated materials. These studies have increased our understanding of phase separation in polymers, but it should be noted that the domain sizes measured in these experiments can only be considered estimates because the values depend on estimates of the diffusion coefficients and the morphological models. Real polymers, for example, might not correspond to the 1D, 2D, or 3D models listed above. It is sometimes necessary to use more sophisticated morphological models that include an interface with a rate constant for spin diffusion that is intermediate between the rigid and amorphous phases.[47] These models, of course, have a larger number of adjustable parameters. It is also possible to derive analytical solutions for spin diffusion in defined geometries, but again, the solutions can be complex and it may not be possible to distinguish between the various models from the shape of the spin diffusion recovery curves. It has been suggested that the initial rate measurements are an easy and reliable way to measure domain sizes in binary mixtures.[48]

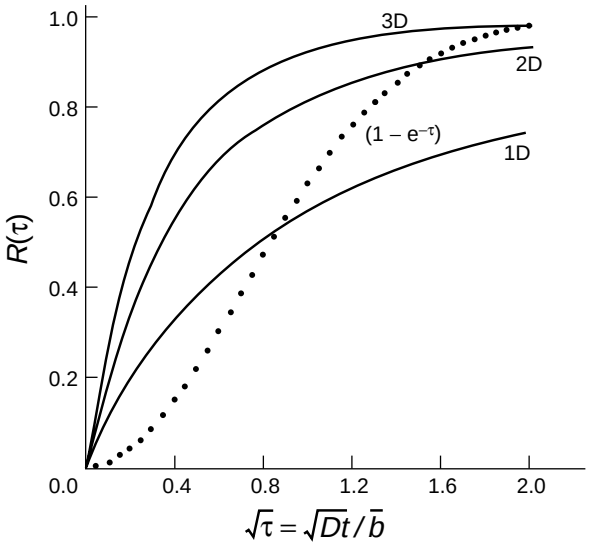

Figure 28 Response function $R(\tau)$ for the recovery of magnetization in the Goldman–Shen experiment for 1D, 2D, and 3D diffusion as a function of the diffusion time. (Adapted with permission from Cheung and Gerstein.[46] Copyright 1981 American Chemical Society.)

The determination of the diffusion constant used to calculate the domain sizes is an important consideration. In most experiments the rigid phase has a shorter T_2 and is saturated during the τ_0 delay in the Goldmen–Shen pulse sequence, so the rate-limiting step in spin diffusion is the spread of magnetization in the mobile phase. In such cases the spin diffusion depends only on the rate of spin diffusion in the mobile phase.

In the earliest studies the rate constants for spin diffusion were estimated from T_2 relaxation rates for the more mobile phase. The spin diffusion rate constant can be estimated from Equation (24):[49]

$$D = \frac{2r_0^2}{T_2} \tag{24}$$

where r_0 is the radius of the hydrogen atom. This value was successfully used to estimate the domain sizes in polyurethanes that were independently measured by X-ray scattering.[49] The upper bound for D can also be estimated by Equation (25):[46]

$$D = \frac{13a^2}{T_2} \tag{25}$$

where a and T_2 are the average distances between protons and the spin–spin relaxation time for the less mobile phase. This estimate was validated in a study of semicrystalline polymers.[46]

In some instances it is possible to determine the spin diffusion constants by measuring spin diffusion in a system where the length scales have been determined by some other experiment. This was reported for symmetric diblock copolymers of styrene and methyl methacrylate that are known to form lamellar structures.[48] In this study the rates of spin diffusion were measured and the diffusion constant was treated as an adjustable parameter to obtain the best agreement with the domain sizes measured by transmission electron microscopy and small-angle X-ray scattering. Excellent agreement between

the NMR and scattering data was obtained using a value of $0.8 \times 10^{-12}\,cm^2\,s^{-1}$ for the glassy phases of both the styrene and methyl methacrylate blocks. The relationships between the line widths or T_2s and the spin diffusion coefficients are known to be approximately correct for rigid materials, but seriously in error for very mobile materials.[50]

In the early measurements of domain sizes by NMR, the spin diffusion data were acquired using proton wideline NMR and fitting the FID to two decay processes that were assigned to the rigid and mobile phases. More recently, variants of the Goldman–Shen pulse sequence have been developed to measure spin diffusion in high-resolution spectra. The simplest application is to follow the spin diffusion period with a cross-polarization pulse sequence to detect the high-resolution carbon spectrum. Pulse sequences such as the dipolar filter,[51] chemical shift filter[52] or chemical shift gradients[53] can be used to create a difference in spin temperature for the two components under study. Selective excitation[54] and 2D CRAMPS NMR[55] have also been used for these studies, as have various types of solid-state carbon–proton 2D NMR methods. Several of these methods will be discussed in the following sections.

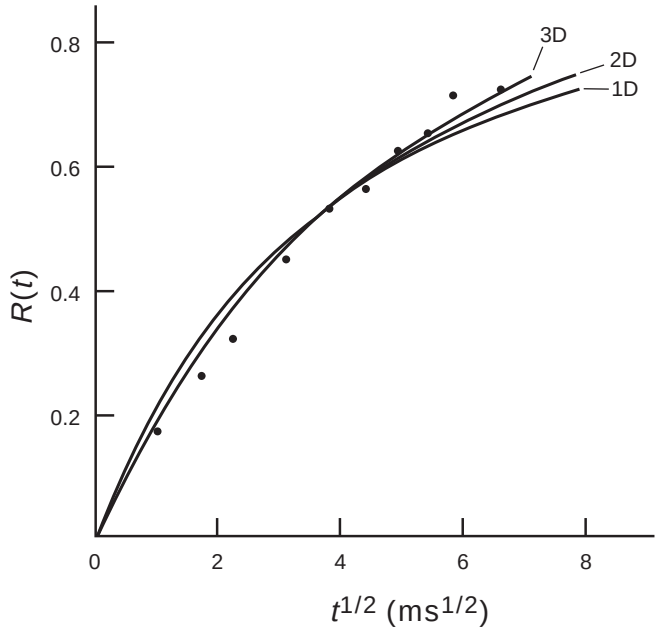

Figure 29 Response function $R(t)$ plotted as a function of $\sqrt{t}$ for polyethylene. The solid lines are calculated for 1D, 2D, and 3D diffusion. (Adapted with permission from Cheung and Gerstein.[46] Copyright 1981 American Chemical Society.)

3.2 Semicrystalline Polymers

Solid-state NMR methods have made important contributions to our understanding of the organization of semicrystalline polymers, in part because the length scales that can be measured by spin diffusion and other techniques are smaller than those that can be measured by microscopy. These methods are also nondestructive and are relatively easy to perform.

The domain sizes of the crystalline and amorphous fractions of polyethylene have been characterized by several spin diffusion methods. In an early study the domain sizes in polyethylene were measured using the Goldman–Shen pulse sequence.[46] The difference in spin polarization between the crystalline and amorphous fractions was produced with a 42 μs delay between the first two 90° pulses of the Goldman–Shen pulse sequence, and the data were acquired as a function of τ_m. The FIDs were fitted to Gaussian and Lorentzian components that were assigned to the crystalline and amorphous phases. The recovery factor $R(t)$ is plotted in Figure 29 along with the best fits from the 1D, 2D, and 3D morphological models. Although polyethylene is known to have a lamellar morphology, the spin diffusion data for the noncrystalline domains appear to be better described by a 2D or 3D model. Using the diffusion coefficient of $8.3 \times 10^{-12}\,cm^2\,s^{-1}$ the average domain spacings for the 2D and 3D models were calculated to be 90 and 144 Å, respectively. As discussed below, more complex morphological models that incorporate interfacial regions might provide a more realistic model for spin diffusion in lamellar semicrystalline polymers.

The structures of other polyethylene samples have been investigated more recently using the dipolar filter pulse sequence to create the polarization difference between the crystalline and amorphous fractions, which can then be monitored by the carbon spectrum using magic-angle spinning and cross-polarization.[51] The pulse sequence used in these studies is shown in Figure 30. The dipolar filter part of the pulse sequence is a multipulse sequence designed to average the dipolar couplings. However, in the dipolar filter experiment the delay time τ between pulses is longer (7–20 μs) than for multipulse decoupling experiments. The dipolar couplings in the rigid phase are not averaged under these conditions, and these signals are saturated by applying the sequence several times, while the signals from the mobile phase are perturbed to a much smaller degree. Following the spin diffusion time, the proton magnetization is sampled with the cross-polarization. This approach has the advantage that the carbon spectrum is observed, where the resolution is typically much greater than in the proton spectrum.

Both low- and high-density polyethylene were studied by this using the dipolar filter. Figure 31 shows a plot of the spin diffusion data for low-density polyethylene.[56] The data could not be adequately fitted to a simple two-phase model, indicating that there may be a substantial interfacial component that was not detected in the earlier

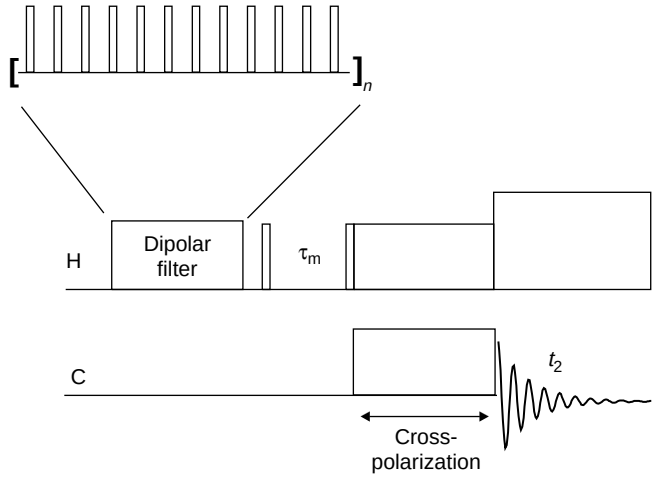

Figure 30 Pulse sequence diagram for the dipolar filtering experiment detected by cross-polarization. The dipolar filter is a multipulse sequence that averages the dipolar couplings of the mobile phase while saturating the signals of the rigid phase. After a spin diffusion period (τ_m) the magnetization is measured using cross-polarization.[51]

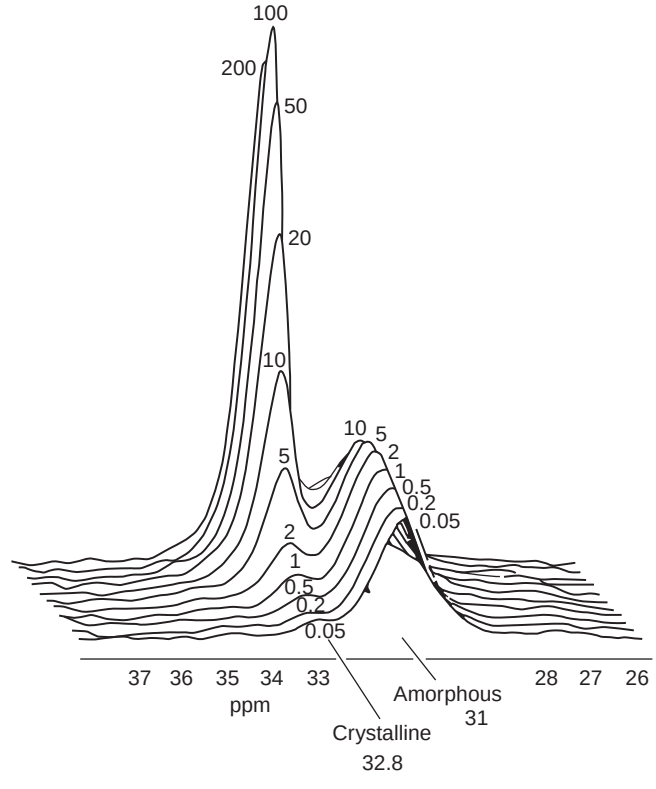

Figure 31 Spin diffusion in low-density polyethylene measured by the dipolar filter. Note the rapid spin diffusion from the more mobile to the less mobile amorphous materials in the first 5 ms followed by a slower spin diffusion to the crystalline region. (Adapted from Blumich et al.[56] by permission of Wiley-VCH.)

experiments. Direct evidence for the presence of an interfacial layer was obtained by monitoring the intensity of the amorphous component as a function of the spin diffusion time. Following saturation of the mobile protons with the dipolar filter pulse sequence, the intensity of the amorphous phase reached a maximum after 5 ms, indicating that there is spin diffusion between the mobile amorphous material to a less mobile amorphous material, the interface. At longer spin diffusion times (100 ms) there is extensive spin diffusion to the crystalline phase. The diffusion behavior was modeled with three components corresponding to the crystalline, amorphous, and interfacial domains. Despite the large differences in the crystallite thicknesses for the high- (40 ± 10 nm) and low-density (9 ± 2.5 nm) polyethylene, the thickness of the interfacial layer was shown to be 2.2 ± 0.5 nm in both samples.

The Goldman–Shen or the dipolar filter pulse sequences can be used to measure domain sizes via spin diffusion only when there are large differences in the spin–spin relaxation rates of the phases under study. This is often a serious limitation since the relaxation times for the crystalline, interfacial, and amorphous phases are too close to be easily separated. For this reason, other methods have been developed to create a difference in polarization between the phases. In some materials there are large differences in the proton $T_{1\rho}$s that can be used to create the polarization difference,[57] but this method is not ideal since the rate of spin diffusion is reduced only by a factor of two during the spin-lock period. A better approach is to create the difference in spin polarization using multipulse NMR methods that quench spin diffusion during the preparation period. Such methods have been used to investigate the phase structure of poly(ethylene terephthalate)[58] and polyethylene.[56]

3.3 Multiphase Polymers

Solid-state NMR is an extremely useful method for investigating the structure of phase-separated polymers. Among the key questions in such systems are how large the phases are and what the degree of phase mixing at the interface is. These questions can be investigated using the proton spin diffusion and by observing differences in the relaxation times for the material in the different phases.

Polyurethane is an example of a commercially important material that has been extensively investigated by solid-state NMR. Several polyurethanes were investigated in an early study using the Goldman–Shen pulse sequence. The magnetization recovery curve for the rigid phase of one such material is shown in Figure 32.[49] The fit to the data points is calculated for a lamellar model that accounts for some spin diffusion during the preparation period (75 μs) and is in excellent agreement with the domain sizes measured by small-angle X-ray scattering. The observation that the T_2 of the mobile phase changes

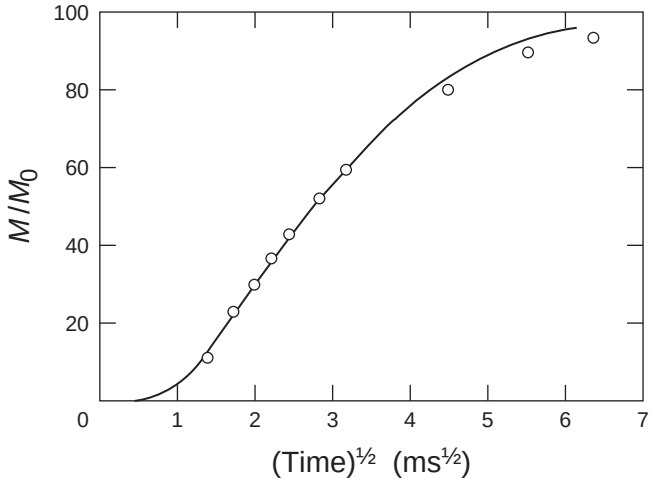

Figure 32 Recovery curve for the rigid phase in polyurethanes measured by the Goldman–Shen experiment. The solid line is a fit to a lamellar model. (Adapted with permission from Assink.[49] Copyright 1978 American Chemical Society.)

with the spin diffusion time was taken as evidence for the presence of an interfacial region.[49]

The NMR studies of polyurethanes have been used not only to measure the length scale of phase separation, but also to measure the degree to which the blocks are separated into the "hard" **(3)** or "soft" **(4)** domains. Deuterium NMR is an effective way to make such measurements because the line shapes for the material in the soft domains are much sharper than those for the hard domains, and the contributions of the hard and soft material to the deuterium wideline spectra can be easily quantified. Figure 33(a–i) show the 55 MHz deuterium NMR spectra for a series of polyurethanes containing 50, 60, and 70% hard segments that have been specifically labeled in the butanediol section of the hard segment.[59] The relative contribution of the deuterated butanediol dissolved in the hard and soft domains is obtained by subtracting the spectra for a polymer containing only the hard segments from the observed spectra, and the material with the motionally averaged line shape is assigned to the amorphous and interfacial material. The NMR results compare favorably

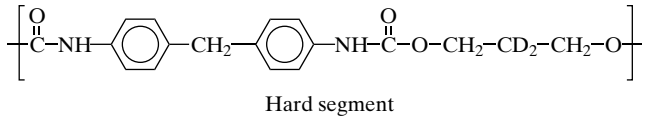

Hard segment

(3)

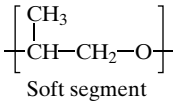

Soft segment

(4)

with those measured by small-angle X-ray scattering for the material containing 60 and 70 wt% hard segments. Similar results were reported for model polyurethanes containing deuterated piperazine rings in the hard segment and poly(tetramethylene oxide) ($M_n \approx 2000$) for the soft segment, where 85% of the labeled rings were observed in the hard domains.[60]

The structure of phase-separated diblock copolymers has been extensively investigated by spin diffusion using the Goldman–Shen[45] and dipolar filtering[51] pulse sequences. In one study the structure of a series of the poly(styrene-*b*-methylphenylsiloxane) diblock copolymers with different molecular weights, compositions, and morphologies was examined. The magnetization recovery curves for some of the samples are shown in Figure 34.[47] The spin diffusion curves for the three of the polymers show that clear differences in the morphology can be inferred from the raw data. It is well known that the morphology of block copolymers depends on the block molecular weights and compositions. It is possible to vary these parameters to obtain different morphologies for the same chemical system. The data in Figure 34 were modeled as a three-phase system, a polystyrene-rich phase, a poly(methylphenylsiloxane)-rich phase and an interface in which diffusivity changes linearly between the values of the two blocks. The dimensions of the phases and the morphologies for six samples are in close agreement with the domain sizes measured by small-angle X-ray scattering.[47]

In the experiments discussed thus far, the polarization difference between the phases is created by a difference in the relaxation times that are related to differences in the molecular dynamics. In some cases of interest, however, the differences in dynamics may not be large enough to study these materials by spin diffusion. One novel way to study such materials is the *chemical shift filter*, which selects signals for cross-polarization based on their proton chemical shift in a multipulse line-narrowing experiment.[61] This method has been used to study a series of blends, blocks, and statistical copolymers of styrene and methyl methacrylate.[62] Polystyrene and poly(methyl methacrylate) are immiscible and the block copolymers with equal block length are known to form lamellar structures. The domain sizes calculated from the spin diffusion data span a range between 0.5 and 100 nm and are in close agreement with those expected theoretically.

3.4 Polymer Blends

Solid-state NMR has been an important tool for the characterization of polymer blends and mixtures. Among the most important questions in this field to which solid-state NMR has made important contributions are what is

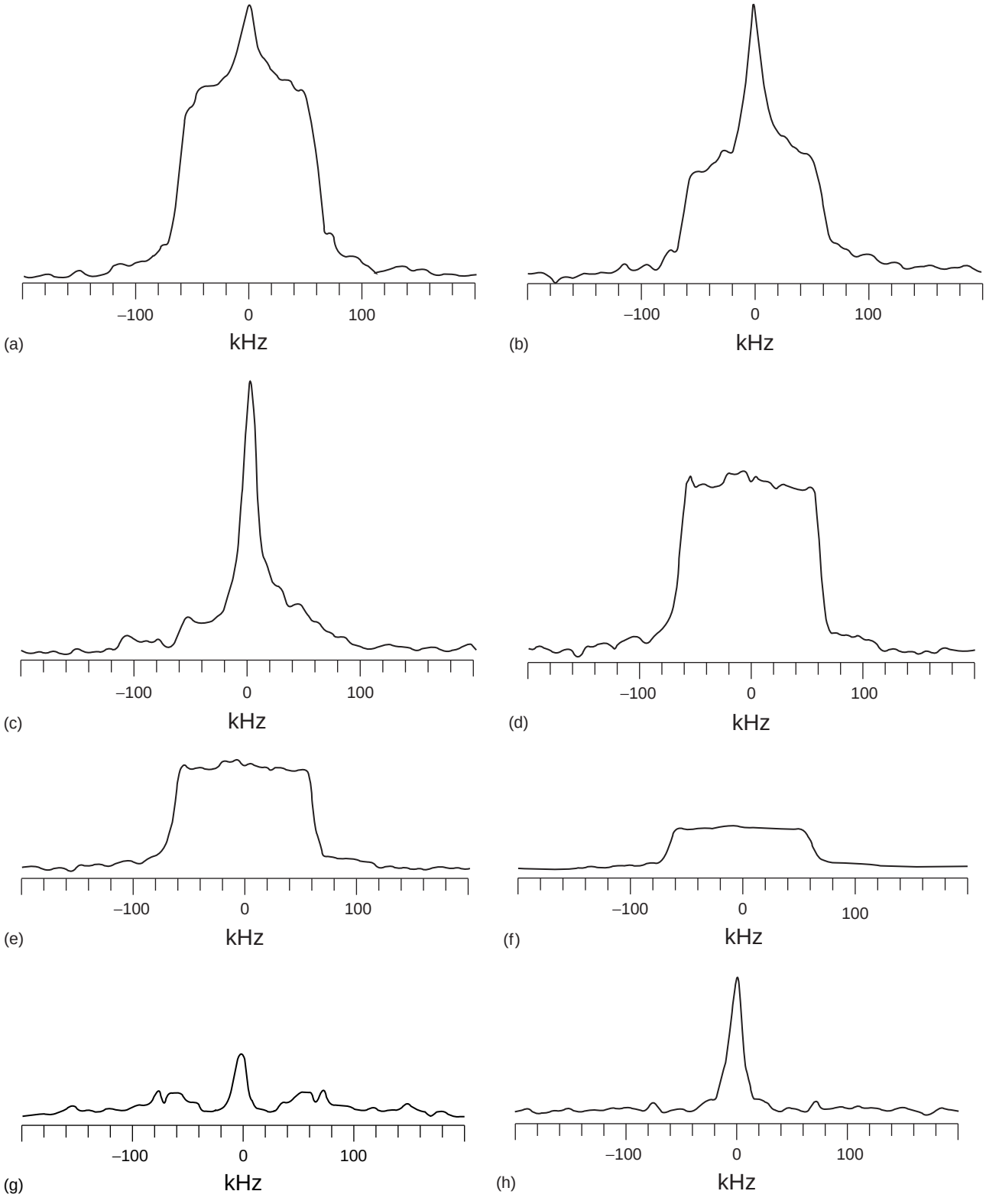

Figure 33 55 MHz deuterium NMR spectra of hard-segment labeled polyurethanes. Spectra (a)–(c) are for polymers containing 70, 60, and 50% hard segments, respectively. Spectra (d)–(f) show the fraction of hard segment contributing to spectra (a)–(c) and spectra (g)–(i) show the result of subtracting the hard-segment spectra from the observed spectra. (Adapted with permission from Dumais et al.[59] Copyright 1985 American Chemical Society.)

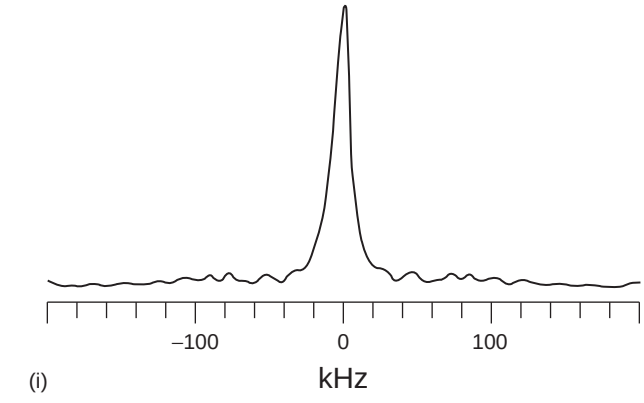

Figure 33 (*Continued*)

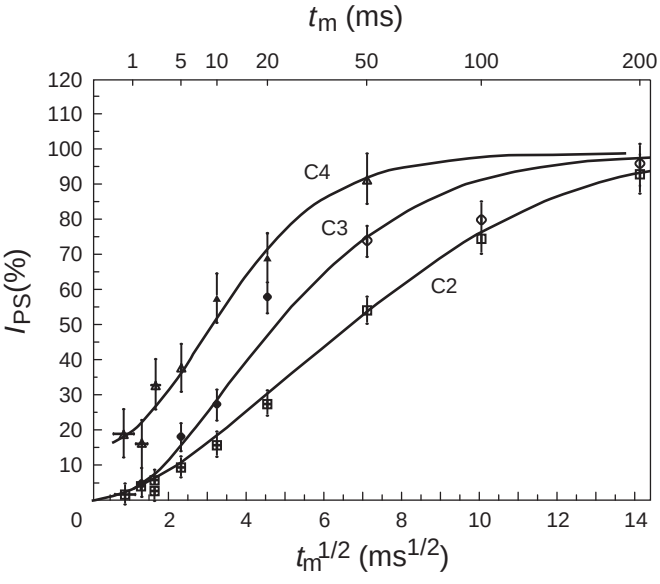

Figure 34 Magnetization recovery curves for poly(styrene-*b*-methylphenylsiloxane) copolymers containing (C2) 37%, (C3) 21% and (C4) 16% methylphenylsiloxane. The solid lines are fits to the morphological models. (Adapted from Cai et al.[47] by permission of Elsevier Science, Oxford, UK.)

the length scale of mixing in polymer blends, and what are the specific intermolecular interactions that lead to miscibility. NMR is one of the methods of choice for the determination of miscibility because the length scale measured using spin diffusion is much smaller than that inferred from differential scanning calorimetric or other studies. With the introduction of new 2D NMR methods and isotopic labeling, it is now possible to measure the local structure in polymer blends. From such studies it is possible to identify the interacting groups in miscible blends and to gain a deeper insight into the factors that determine miscibility.

The molecular level mixing of polymer chains can be investigated by measuring the chemical shifts and relaxation rates. On the smallest length scale, strong intermolecular interactions due to hydrogen bonding, charge-transfer complex formation, or ionic interactions can cause chemical shift changes that are large enough to be measured in the solid-state spectra. On a longer length scale, spin diffusion experiments can be used to measure the average distances between the polymer chains. Blend formation is also known to affect the molecular dynamics.

Most binary mixtures of polymer are not mixed on the molecular level because the entropy of mixing is extremely small for high-molecular-weight polymers, so miscibility is usually a consequence of favorable intermolecular interactions.[63] If these interactions are strong enough it is sometimes possible to observe such effects directly in the solid-state NMR spectra. One such example is shown in Figure 35(a–c), which compare the 50 MHz carbon CPMAS spectra of two polymers [**(5)** and

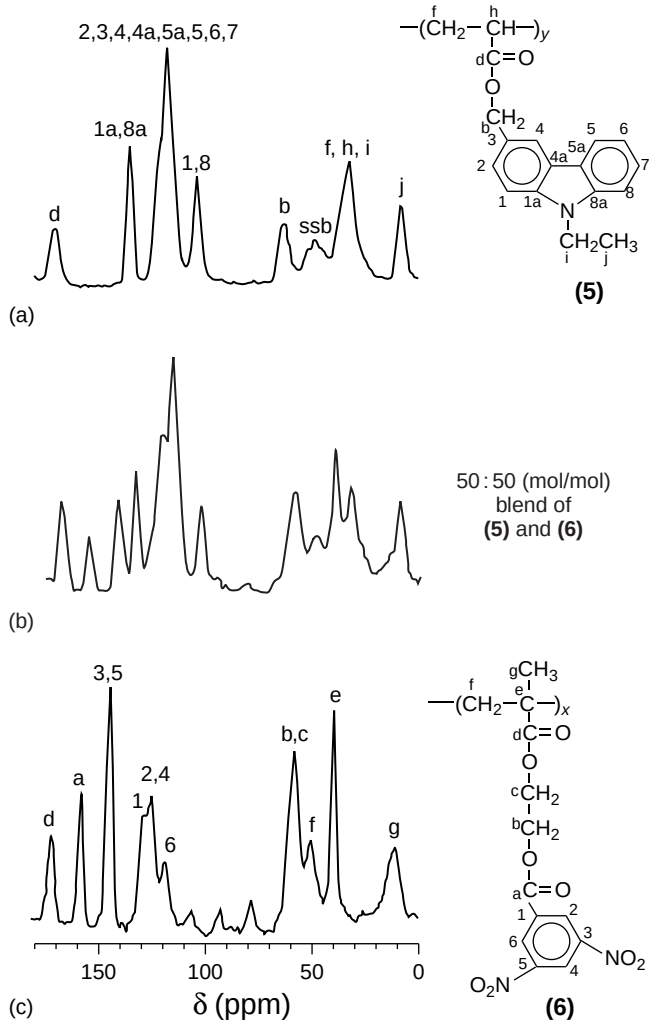

Figure 35 Carbon NMR spectra of (a) **(5)**, (b) a 50 : 50 blend of **(5)** and **(6)**, and (c) **(6)**. (Adapted with permission from Simmons and Natansohn.[64] Copyright 1992 American Chemical Society.)

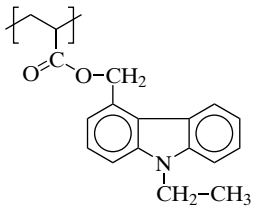

(5)

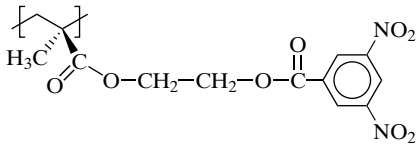

(6)

(6)] labeled with electron-donor and -acceptor moieties in a miscible blend.[64] The miscibility is due to charge-transfer complex formation between the aromatic rings on the donor and acceptor polymers, and complex formation causes chemical shift changes from ring current effects. The largest chemical shift differences (ca. 1.2 ppm) are observed for carbons 1, 3 and 5 of the poly {[2-(3,5-dinitrobenzoyl)oxy]ethyl methacrylate} moiety.

Ionic interactions are also strong enough to influence the chemical shifts of polymers in blends,[65] as are hydrogen bonds.[66] However, unless these interactions are very strong and the peaks well resolved, only broadened lines from the interacting groups are observed.

These experiments show some cases where blend formation is driven by strong intermolecular interactions that lead to easily observable changes in the solid-state carbon NMR spectra. It is much more common, however, that blend formation is caused by weak intermolecular interactions and that the chemical shift changes induced by blend formation are smaller than the solid-state line widths, and are therefore undetectable. In such cases, the length scale of chain mixing is measured by the spin diffusion experiments discussed above.

Intermolecular cross-polarization is perhaps the most direct method to determine molecular level mixing in polymer blends. This experiment is performed on a blend of protonated and deuterated polymers, and takes advantage of the fact that the deuterated polymer has no directly bonded protons and cannot be cross-polarized. If the blends are mixed on a length scale of less than 20 Å, then the deuterated polymer can be cross-polarized from the protons on the other polymer chain.[67] This behavior is illustrated in Figure 36(a–d), which compare the cross-polarization behavior for the blend of deuterated polystyrene and protonated poly(methyl vinyl ether) with the mechanical mixture.[68] The polystyrene aromatic signals at 125 ppm are only weakly cross-polarized with a 24 ms cross-polarization time in the mechanical mixture

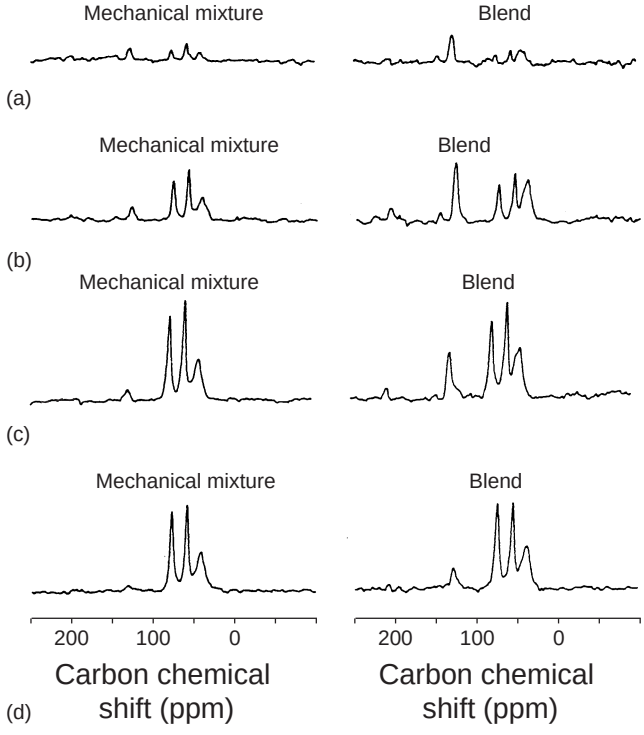

Figure 36 Carbon CPMAS spectra of a mechanical mixture and a blend of polystyrene-d_8 and poly(methyl vinyl ether). The data are shown for the mechanical mixture and the blend with cross-polarization times of (a) 25, (b) 10, (c) 2 and (d) 0.5 ms (Adapted from Gobbi et al.[68] by permission of John Wiley & Sons, Inc.)

while they are strongly cross-polarized in the blend. The introduction of the deuterated polymer is also a useful way to simplify the spectra of the blend when there are many overlapping signals. Since the deuterated polymer has no directly attached protons it can be observed with the nonprotonated carbons using the dipolar dephasing experiment.[69]

The length scale of polymer mixing is often measured via T_1 and $T_{1\rho}$ relaxation rate measurements.[70] If the chains are mixed on a length scale smaller than the length scale of spin diffusion, then magnetization transfer from the slower to the more rapidly relaxing chains is a very effective relaxation mechanism. In such a case the relaxation times of both chains are equal to a weighted average of the values for the individual chains and are given by Equation (26):

$$k = k_a \frac{N_a \varphi_a}{N_a + N_b} + k_b \frac{N_b \varphi_b}{N_a + N_b} \tag{26}$$

where k, k_a, and k_b are the relaxation rates (either T_1 or $T_{1\rho}$) for the blend and the pure polymers a and b, N_T, N_a, and N_b are the total number of protons and those contained in the polymers a and b, and φ_a and φ_b are the respective mole fractions. In those cases where the

length scale of mixing is longer than the spin diffusion length scale, the measured relaxation rates will be the same as for the pure polymers. In cases where there is partial mixing the relaxation behavior may be more complex, and simple exponential relaxation will not be observed.[67] One limitation of this approach to measuring polymer mixing is that the relaxation rates for the pure polymers must differ by more than a factor of two to measure accurately the averaging of the relaxation rates.

The length scale of spin diffusion L measured by such experiments depends on the relaxation time and the diffusion coefficient D, and is approximately given by Equation (27):

$$L = \sqrt{\frac{6D}{k}} \tag{27}$$

where, as above, k is the relaxation rate.[70] Assuming a diffusion coefficient of ca. $10^{-12}\,\mathrm{cm^2\,s^{-1}}$, the length scale of spin diffusion for a polymer with a proton spin–lattice relaxation time of 0.5 s corresponds to ca. 170 Å. Rotating-frame relaxation rates are typically much faster and are therefore affected by a shorter spin diffusion length scale. For a proton $T_{1\rho}$ value of 5 ms, the length scale of chain mixing that can be detected by such measurements is of the order of 17 Å. Exact measurement of the length scale of phase separation is somewhat uncertain because of the uncertainties in estimating D.

Figure 37 shows the effect of blending on the proton $T_{1\rho}$ relaxation times for poly(dimethylphenylene oxide) and poly(4-methylstyrene).[71] These polymers are ideal candidates for the study of polymer mixing because the relaxation times for the pure polymers differ from each other by an order of magnitude, so changes in the relaxation times due to the effects of blend formation can be easily measured. These data show that the relaxation rates measured for the 75:25 mixture are intermediate between those measured on the pure polymers. In addition, the rates are identical for both chains, indicating homogeneous mixing. Single exponential relaxation rates are observed for the 50:50 mixture, but the relaxation rates for the poly(4-methylstyrene) and poly(dimethylphenylene oxide) are no longer equal to each other. At a ratio of 25:75 multiexponential relaxation behavior is observed for the poly(dimethylphenylene oxide). These data show that the mixtures of poly(4-methylstyrene) and poly(dimethylphenylene oxide) are not homogeneous on the length scale measured by $T_{1\rho}$ (30 Å) for the samples prepared with a 50:50 or 25:75 ratio. This approach was initially used to study blends of polystyrene and poly(dimethylphenylene oxide), where it was reported that these polymers do not form homogeneous mixtures.[67]

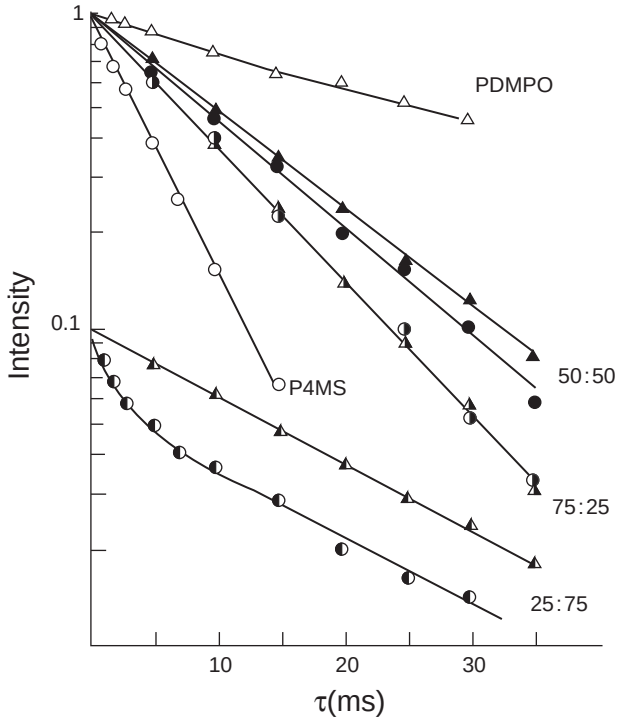

Figure 37 Proton rotating-frame relaxation data for poly(2,6-dimethylphenylene oxide), poly(4-methylstyrene) and 50:50, 75:25, and 25:75 blends. The data for the 25:75 blend are offset from the other data to show the nonexponential behavior. The triangles represent measurements for carbons on poly(2,6-dimethylphenylene oxide) and the circles for poly(4-methylstyrene). (Adapted with permission from Dickinson et al.[71] Copyright 1987 American Chemical Society.)

2D proton solid-state exchange NMR has also been used to study the mixing of the polystyrene and poly(methyl vinyl ether) blends from toluene and chloroform. The broad lines in the proton spectra that result from the dipolar interactions require the use of both magic-angle spinning and multipulse decoupling during both the t_1 and t_2 periods, and the 2D exchange pulse sequence modified for solid-state proton NMR.[55] The resulting 1D spectra and the 2D correlation maps for the blends are shown in Figure 38(a) and (b).[55] Three groups of signals can be observed in the high-resolution proton NMR spectra that are assigned to the polystyrene aromatic protons at 7 ppm, the poly(methyl vinyl ether) methine and methoxy protons at 3.5 ppm, and a band containing the remaining aliphatic protons at higher field. Two sets of cross peaks are observed in the 2D spectra acquired for the blend cast from chloroform that can be assigned to the intramolecular polystyrene aromatic to aliphatic interactions and intramolecular methine/methoxy to aliphatic interactions in poly(methyl vinyl ether). In the blend cast from toluene, additional cross peaks are observed between the polystyrene

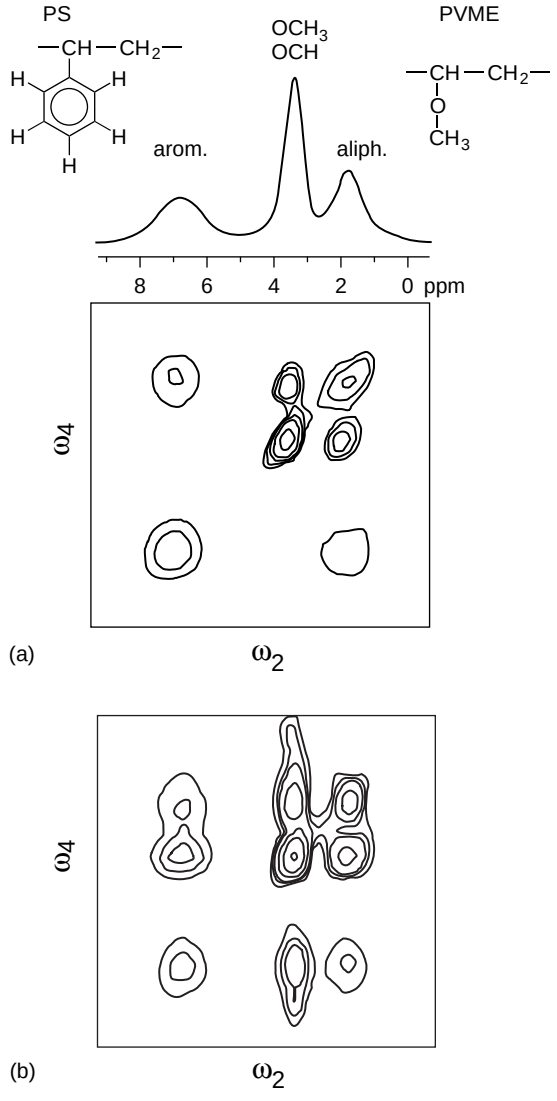

Figure 38 Solid-state proton 1D NMR spectra of a polystyrene–poly(methyl vinyl ether) (PVME) blend and the 2D spin exchange spectra for the blends from (a) chloroform and (b) toluene. (Adapted with permission from Carawatti et al.[55] Copyright 1985 American Chemical Society.)

aromatic protons and the poly(methyl vinyl ether) methine/methoxy peak. Since the only mechanism for magnetization transfer in these experiments is through-space dipolar interactions, these spectra demonstrate that a large fraction of the chains are in molecular contact.[55]

The rate of intermolecular spin diffusion and the fraction of polymer in the mixed phase for the polystyrene and poly(methyl vinyl ether) blends can in principle be calculated from the intensities of the diagonal and cross peaks as a function of mixing time.[55] A more practical way to make such measurements, however, is to use selective excitation followed by detection of the proton

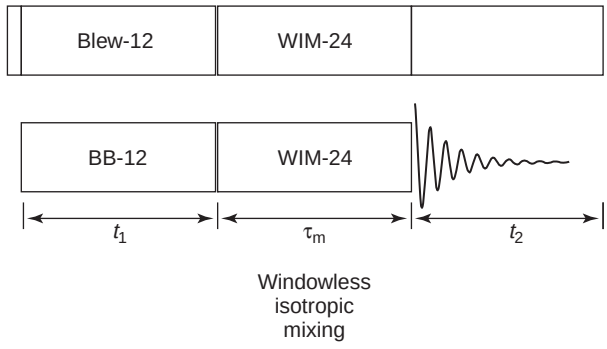

Figure 39 Pulse-sequence diagram for 2D heteronuclear correlation in the solid state. Multipulse decoupling is applied to both the carbons and protons during the t_1 period and selective cross-polarization without spin diffusion occurs during the 24 pulse WIM mixing period.[75]

signals using CRAMPS.[54] These studies showed that the miscible blend was apparently nanoheterogeneous, and that 79% of the polymer was in a mixed phase. The blend has also been studied by wideline separation (WISE) NMR[72] and by the solid-state nuclear Overhauser effect (NOE).[73]

The miscibility of polymer blends has also been investigated using the 2D heteronuclear correlation experiment shown in Figure 39.[74,75] The purpose of this experiment is to correlate the carbons and protons with strong dipolar couplings. Multipulse decoupling is applied during the t_1 period to the protons and carbons with the BLEW-12 and BB-12 pulse sequences to remove both the proton–proton and carbon–proton dipolar couplings, so that the high-resolution proton spectrum can be indirectly measured.[76] Magnetization is transferred from protons to carbons with the WIM-24 pulse sequences, which quenches proton–proton spin diffusion during the cross-polarization.[74] This is in contrast to the normal cross-polarization by spin-locking, where proton spin diffusion proceeds at a half the rate observed in the absence of the rf field.[77] If a spin diffusion delay is inserted between the evolution and mixing periods, correlation via spin diffusion is observed not only between nearby protons and carbons, but also between carbon and more distant protons.

In certain favorable circumstances, it is possible to use the heteronuclear correlations to measure the geometry of the intermolecular contacts in miscible blends. Such an example is illustrated in the heteronuclear correlation spectrum of the blend of poly(vinylphenol) and poly(methyl acrylate) shown in Figure 40.[78] The strongest peaks in the spectrum arise from the correlation of the carbons with their directly bonded protons, while weaker correlations are noted between the nonprotonated carbons and nearby protons. Figure 40 shows a weak correlation between the carbonyl carbon

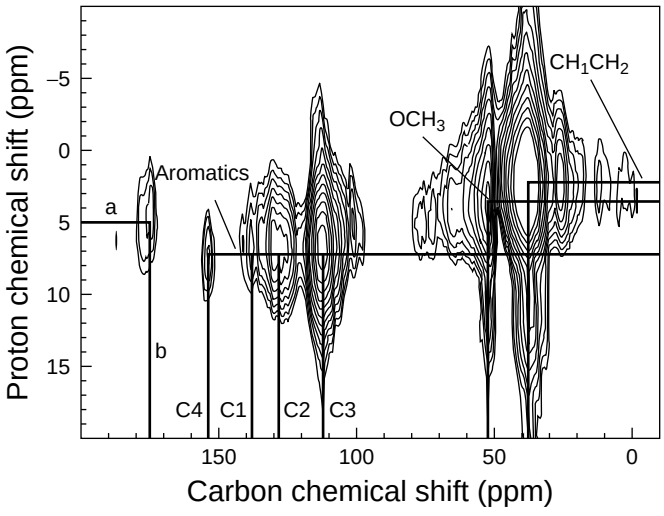

Figure 40 2D heteronuclear correlation spectra of the blend of poly(methyl acrylate) and poly(vinylphenol). Note the correlation between the carbonyl peak and the protons at 5 ppm. (Adapted with permission from White and Mirau.[78] Copyright 1994 American Chemical Society.)

of poly(methyl acrylate) and protons resonating at ca. 5 ppm. Since this correlation is not observed in the spectrum with a poly(vinylphenol) sample in which the hydroxyl protons have been exchanged for deuterons, this correlation must arise from the close contact between the carbonyl of poly(methyl acrylate) and the hydroxyl protons of poly(vinylphenol) due to intermolecular hydrogen bonding. Heteronuclear correlation experiments are a promising new tool to investigate the structure of polymer blends, and have been used to study polycarbonate blends with aromatic diamines,[79] polystyrene–poly(dimethylphenylene oxide)[80] and polycaprolactone–polycarbonate blends.[81]

4 THE DYNAMICS OF POLYMERS

4.1 Introduction

NMR has long been used to observe molecular motion and has been extensively applied to the complex dynamic problems presented by polymer chains. Such studies have customarily been divided into two regimes, polymer solutions, in which the chain motion is relatively fast, and polymer in the solid state, where the molecular motion is slower and may vary over many orders of magnitude. In some crystalline and glassy polymers this may reach the extreme of complete immobility. Since many synthetic polymers are useful because of their physical and mechanical properties in the solid state that are ultimately related to the molecular level dynamics, the study of the molecular dynamics of polymers has received much attention. The studies in solution primarily reveal information about *intramolecular* forces, while the molecular dynamics in the solid state are determined by the combination of *intra-* and *intermolecular* forces.

The molecular dynamics of polymers in solution are typically measured via relaxation rate measurements, such as the T_1, T_2 and NOE measurements that are most sensitive to rapid molecular motions. These same measurements are used in the solid state along with several others, including the $T_{1\rho}$, T_{CH}, and some recently developed 2D NMR methods for studying the low-frequency molecular dynamics in polymers. The choice of experiment depends to a large degree on the timescale of the polymer molecular dynamics, since each of these methods is sensitive to the dynamics over a range of correlation times. In addition, the measurement and interpretation of NMR line shapes in solids can provide additional information about both the rate and amplitude of atomic fluctuations.

Nuclear relaxation is caused by fluctuating local fields that result from the motions of nearby atoms. These fluctuations in the local magnetic fields can occur over a broad range of frequencies. If these fluctuations have components at the resonance frequency for the nuclei of interest, they will cause relaxation. The distribution of motional frequencies is given by the *spectral density function* which is defined by Equation (28):

$$J(w) = \frac{1}{2}\int_{-\infty}^{+\infty} G(\tau)\, e^{-iw\tau}\, d\tau \qquad (28)$$

where $G(\tau)$ is the autocorrelation function of the time-dependent relation expressing the orientation of the internuclear vector, such as a ^{13}C–^{1}H vector, in the laboratory frame and is given by Equation (29):

$$G(\tau) = \langle F(t)F^*(t+\tau)\rangle \qquad (29)$$

The angle brackets denote the ensemble average over a collection of nuclei and F represents a function, related to spherical harmonics, describing the position and motion of the molecule. $J(w)$ may be thought of as expressing the power available at frequency w to relax the spins in question. The spectral densities and autocorrelation functions are Fourier inverses of each other in the time and frequency domains, respectively. If $G(\tau)$ decays to zero in a short time, this corresponds to a short *correlation time* τ_c, which means that molecular motion is rapid and that the molecules have only a short memory of their previous state of motion.

In order to give these ideas specific meaning, we must adopt a dynamic model for the polymer molecule. The simplest model views it as a rigid sphere immersed in a viscous continuum that is reoriented in small diffusive

steps. The correlation time can be thought of as the interval between these alterations in the state of motion of the molecule. For such a molecule the loss of memory of the previous motional state is exponential with the time constant τ_c (Equation 30):

$$G(\tau) = e^{-(\tau/\tau_c)} \tag{30}$$

and the spectral density function becomes Equation (31):

$$J(w) = \frac{\tau_c}{1 + w^2\tau_c^2} \tag{31}$$

Figure 41 shows plots of the spectral density function as a function of frequency for three values of τ_c. A long value of τ_c corresponds to the molecular motion of a large molecule, a stiff chain, or a small molecule in a viscous medium, and a short value of τ_c corresponds to the rapid motion of a small molecule or a very flexible polymer chain. Also shown in Figure 41 is the plot for an intermediate case, when $\tau_c w_0 \approx 1$, where w_0 is the resonant frequency of the observed nuclei. The areas under the curves are the same for all three correlation times, which means that the molecular power available to cause relaxation is the same, and only the distribution varies with τ_c. For long correlation times the component at w_0 is weak, and for the short τ_c the frequency spectrum is broad so no one component, particularly that at w_0, is very intense. At some intermediate value of τ_c the intensity at w_0 with be at a maximum. Since the relaxation rate depends on the component at w_0 it will be at a maximum for the intermediate value of τ_c, and the relaxation rates will be much slower for very fast or very slow motion. The exact dependence of the relaxation rate on the correlation time will depend on which experiment is performed and which relaxation rates are measured. A critical part of measuring the molecular dynamics of polymers is choosing the proper model for the correlation functions and spectral densities. It is well established that a single exponential correlation function is not a good model for the molecular motion of an object as complex as a polymer chain.[82]

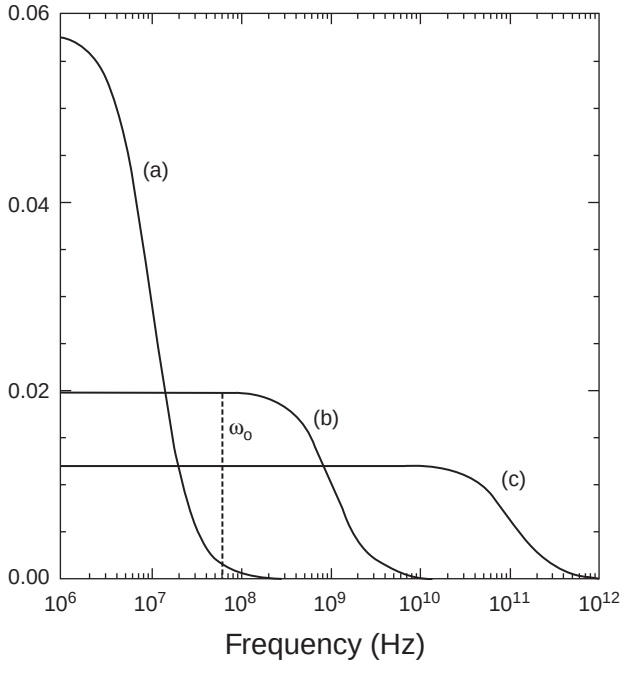

Figure 41 Plot of the spectral density functions as a function of frequency for (a) short, (b) intermediate, and (c) long correlation times. The dotted line at w_0 shows the power of the spectral density at the spectrometer frequency.

NMR studies of the chain dynamics of solid polymers utilize the relaxation measurements used for solution studies as well as several others. Solids differ from solutions in that there is not enough molecular motion to average the dipolar interactions, the chemical shift anisotropy or the quadrupolar interactions, and information about the molecular dynamics can also be extracted from the line shapes. One dramatic manifestation of this effect is the observation of the spectrum above and below the glass transition temperature, where it is often observed that the lines are sharpened above T_g by molecular motion.[83] Other types of motions that are fast on the timescale of the chemical shift anisotropy or the quadrupolar couplings, such as methyl rotation, lead to characteristic changes in the line shape. In deuterium NMR, for example, the quadrupolar line shapes resulting from aromatic ring flips or methyl group rotations can be easily recognized.[84]

More recently, 2D NMR experiments analogous to the solution 2D exchange experiments have been developed to measure the ultraslow molecular dynamics of polymers.[62] If a polymer undergoes molecular motion during the mixing time (milliseconds to seconds) then a characteristic off-diagonal ridge pattern develops in the solid-state deuterium or carbon spectra. The pattern of ridges depends both on the rate and angle of reorientation, and the molecular dynamics can be mapped out to high precision directly from the experimental data.

The methods for studying the dynamics of polymers over a wide timescale can be classified as line shape methods, relaxation rate measurements, or exchange spectra measured by 2D NMR. Each of these methods is sensitive to the dynamics over a range of timescales that depends on the type of interactions. Figure 42 shows a plot of the timescale over which each of the methods is most sensitive. The fastest motions can be measured most efficiently by the relaxation times T_1 and T_2. Motions in the intermediate timescale (10^{-7} s $> \tau_c > 10^{-3}$ s) can be measured using a variety of methods, including changes in the deuterium, dipolar, or chemical shift anisotropy line shape, the $T_{1\rho}$, or the echo intensity in deuterium

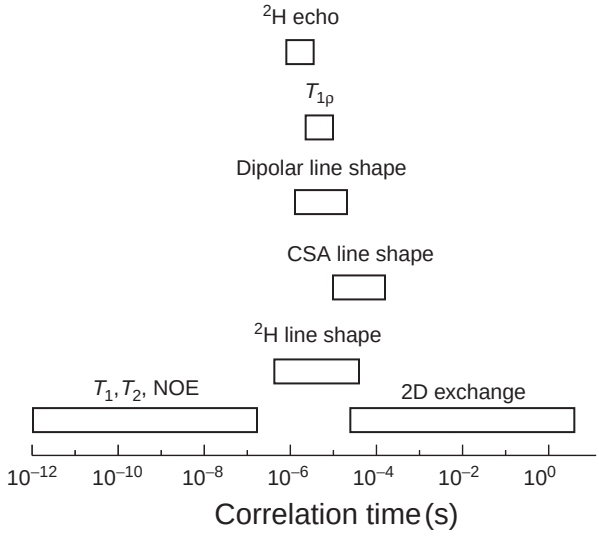

Figure 42 Plot showing the dynamics that can be measured by line shape measurements, relaxation rate measurements, and 2D exchange.

NMR. Motions in the slowest regime can be measured using 2D exchange measurements of the deuterium or chemical shift anisotropy line shape. The upper limit on the dynamics that can be studied by the 2D spin alignment or spin exchange is the T_1 of the material under study. For some crystalline polymers, this may be of the order of hundreds of seconds.

The observation of NMR line shapes, particularly those due to quadrupolar coupling in deuterated polymers or the chemical shift anisotropy in the carbon, silicon, phosphorus or nitrogen spectra, are a rich source of information about the molecular dynamics of polymers. The line shapes have a characteristic appearance in the absence of molecular motion and the changes in line shape due to the molecular dynamics depend on the type of molecular motion and the timescale over which the motion occurs. If the motion has a large amplitude and a nearly isotropic angular distribution, as for polymers above T_g, then solution-like spectra are obtained.

Deuterium NMR is often used to study polymer dynamics because the relaxation is dominated by the quadrupolar coupling, so the relaxation mechanism is well characterized.[84] Deuterium has a low sensitivity, so polymers with site-specific labels are required. This is a disadvantage in that new polymers must be synthesized, but an advantage in the sense that the labels are incorporated into the polymer at specific sites, such as the main-chain or the side-chain aromatic rings. In the absence of molecular motion, the frequency of a deuteron is given by Equation (32):

$$w = w_0 \pm \delta(3\cos^2\theta - 1 - \eta\sin^2\theta\cos 2\phi) \qquad (32)$$

where w_0 is the resonance frequency, $\delta = 3e^2qQ/8\hbar$, $e^2qQ/\hbar$ is the quadrupolar coupling constant, η is the asymmetry parameter and the orientation of the magnetic field in the principal axis of the electric field gradient tensor is specified by the angles θ and ϕ. For rigid solids $\delta/2\pi = 62.5\,\text{kHz}$ and $\eta \approx 0$ for C–D bonds. Thus, two lines are observed corresponding to the transitions for each deuteron. In isotropic samples, averaging over all possible orientations gives rise to the well known Pake spectrum.[85]

Molecular motion of the polymer will cause the line shapes to change in a way that depends on the geometry and timescale of the molecular motion. If the motion is rapid on the timescale defined by the coupling constant, $1/\delta$, it is said to be in the fast motion limit ($\tau_c < 10^{-7}\,\text{s}$). This leads to a characteristic change in line shape given by Equation (33):

$$w = w_0 \pm \bar{\delta}(3\cos^2\theta - 1 - \bar{\eta}\sin^2\theta\cos 2\phi) \qquad (33)$$

where $\bar{\delta}$ and $\bar{\eta}$ are the coupling constant and asymmetry parameter for the averaged electric field gradient tensor. It is important to note that $\bar{\eta}$ may be different from zero even though $\eta = 0$. This is illustrated in Figure 43(a–c) for kink three-bond rotation, crankshaft five-bond motion, and aromatic ring flips.[84] Such motions give rise to $\bar{\eta}$ values of 1, 0, and 0.6 and illustrate how different types of molecular motion result in characteristic deuterium line shapes.

Deuterium NMR is a powerful method for the study of the molecular dynamics of polymers, but caution must be used since it is possible for different motions to give rise to the same values for $\bar{\delta}$ and $\bar{\eta}$, and the same line shape. For example, the same $\bar{\eta} = 1$ pattern is obtained for a two-site jump model with a jump angle of 90° when the population of site two is twice that of site one, a two-site jump with equal populations and an angle of 109°, or with restricted diffusion about a tetrahedral axis with a distribution width of 70°.[86] Such models can be distinguished via the anisotropic relaxation of the line shape using the Jeener–Broekaert pulse sequence to measure T_{1Q}-distorted line shape.[86,87]

For those cases where the molecular dynamics are on the intermediate timescale, $\tau_c \approx 1/\delta$, the line shapes cannot be so easily calculated. However, information about the molecular dynamics can be obtained using echo methods. Deuterium spectra are typically very broad and cannot be acquired using the 90° pulse–acquire pulse sequences because much of the data is lost during the period after the pulse during which the electronics recover after the strong rf pulse. These data are usually acquired with the quadrupolar echo pulse sequence shown in Figure 11. If molecular motion occurs during the echo delay times, it will not be refocused and changes in the

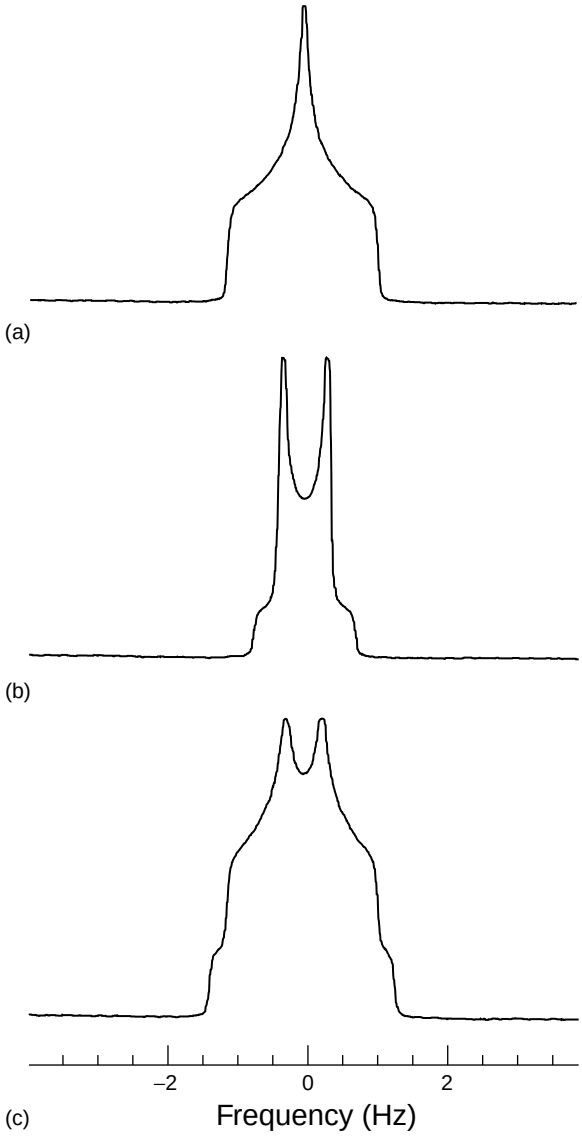

Figure 43 Theoretical deuterium NMR line shapes for deuterons undergoing (a) kink three-bond motion, (b) crankshaft five-bond motion and (c) 180° flips, as in aromatic rings. The rate of molecular motion is assumed to be in the fast motion limit. (Adapted from Spiess[84] by permission of Springer-Verlag New York, Inc.)

line shape and intensity may be observed as a function of the echo delay time.[84]

If the motion is in the slow motion limit, $\tau_c < 10^{-4}$ s, then the line will not be sensitive to the molecular motion. If the correlation time is less than the spin–lattice relaxation time then characteristic changes in the line shape can be observed with pulse sequences that resemble the 2D exchange pulse sequence but which differ significantly in the details. As in the 2D exchange spectra, the frequencies of the deuterons in the t_1 dimension are correlated with those in the t_2 dimension. If no molecular motion occurs during the mixing time then the signals for a particular deuteron are the same in both dimensions and the signals appear along the diagonal. Thus, for mixing times that are short compared with the correlation time the normal "Pake" pattern appears along the diagonal. If the deuterons reorient during the mixing time then they have different frequencies in the t_1 and t_2 dimensions and a ridge is observed in the 2D plane that depends on the angle of reorientation. Such a ridge pattern is observed because the frequency depends on the orientation of the deuteron in the magnetic field.

The study of the molecular dynamics using the chemical shift anisotropy of carbon, nitrogen, phosphorus, or silicon is analogous to the deuterium NMR studies. The chemical shift depends on the orientation of the principal axis of the chemical tensor in the magnetic field and is given by Equation (34):

$$w = w_0 + \tfrac{1}{2}\delta(3\cos^2\theta - 1 - \eta\sin^2\theta\cos 2\phi) \qquad (34)$$

where δ is the strength of the anisotropic coupling given by $\delta = \frac{2}{3}\Delta\sigma\gamma B_0$. The chemical shift anisotropy is typically much smaller than the quadrupolar coupling with deuterium (0–4 vs 62 kHz) and therefore is sensitive to molecular motions on a different timescale.

Information about the molecular dynamics of polymers can also be obtained from the heteronuclear dipolar couplings in 1D or 2D NMR experiments. High-resolution solid-state NMR spectra are often recorded with high-power (50–100 kHz) proton decoupling and magic-angle spinning to average the dipolar interactions and the chemical shift anisotropy. However, if the molecular motion is on the same timescale as the decoupling field, interference with the decoupling will be observed and the spin–spin relaxation rate will be give by Equation (35):[88]

$$\frac{1}{T_2} = \frac{\gamma_C^2\gamma_H^2\hbar^2}{5r^6}\left(\frac{\tau_c}{1 + w_2^2\tau_c^2}\right. \qquad (35)$$

where w_2 is the decoupler field strength. When $w_2^2\tau_c^2 \approx 1$, the line width will be at a maximum and the signal will be broadened beyond detection. In a similar way, if the frequency of molecular motion is on the same scale as the magic-angle spinning, the lines will also be broadened.

The carbon and proton spin–lattice relaxation times provide gross information about the molecular dynamics of polymers, but it is frequently difficult to interpret the relaxation times in terms of the correlation times as has been effectively done for polymers in solution. It is particularly difficult for protons because proton spin diffusion is very efficient in the solid state so the measured relaxation time represents an average for the entire spin system. If the polymer contains methyl groups, then methyl group rotation is the most effective mechanism to cause relaxation and will dominate the relaxation of

the entire spin system, limiting the information about the molecular dynamics that can be extracted from such measurements. Proton T_1 measurements are very useful for studying the length scale of mixing in polymers.

NOEs can sometimes be observed for solid polymers, particularly for those near or above the glass transition temperature. Such measurements are also difficult to interpret in terms of the molecular motions because they are also mainly due to methyl group rotation.[89] These measurements are most useful for studying the structure of polymer blends and other multiphase polymer mixtures.[73]

Information about the molecular dynamics of polymers can also be obtained from the cross-polarization dynamics. As noted in section 1.2.2, the time course of the magnetization in the cross-polarization experiment represents a compromise between the buildup due to proton–carbon cross-relaxation (T_{CH}) in the spin-locking field and the decay due to carbon and proton rotating-frame relaxation rates $T_{1\rho}$(C) and $T_{1\rho}$(H).

The proton rotating-frame relaxation rates can be measured from the decay of cross-polarized magnetization or using the related pulse sequence shown in Figure 8(a) and (b). However, spin diffusion during the spin-locking pulse sequence averages the relaxation times making it difficult to relate the $T_{1\rho}$(H) relaxation times to the molecular dynamics. The $T_{1\rho}$(H) relaxation times are often very useful for measuring the length scale of phase separation in polymer blends and multiphase polymers systems.

In favorable cases, the rotating-frame carbon spin–lattice relaxation time $T_{1\rho}$(C) can be related to the molecular dynamics of polymers. The $T_{1\rho}$(C) is measured using the pulse sequence shown in Figure 8(b) in which the carbon magnetization is spin-locked for a variable contact time in the absence of the proton spin-locking field. This relaxation parameter is often of interest because the relaxation is due to molecular motions in the 15–100 kHz regime that may be related to the mechanical properties of polymers. Interpretation of the $T_{1\rho}$(C) relaxation is complicated by the possibility of cross-relaxation between the spin-locked carbons and the unlocked protons. The relative contributions of the two pathways can be estimated by measuring the relaxation time as a function of the spin-lock power, since the relaxation by the spin–spin pathway has a quadratic dependence on w_1, whereas relaxation by $T_{1\rho}$(C) has an exponential dependence. It has been reported that the relaxation in glassy polymers is primarily due by the $T_{1\rho}$(C) pathway and the relaxation times can be interpreted in terms of chain dynamics over the 15–100 kHz timescale.[90] while the relaxation in semicrystalline polymers has a substantial contribution from cross-relaxation and cannot be so easily interpreted.[91]

4.2 Semicrystalline Polymers

A variety of NMR methods, including relaxation rate measurements, deuterium NMR, and 2D NMR methods, have been used to investigate the molecular dynamics of semicrystalline polymers. Multiple signals and relaxation rates are often observed in the NMR spectra that are assigned to the crystalline, amorphous, and interfacial materials. The measurement of the molecular dynamics plays an important role in the assignment of the signals and in understanding structure–property relationships.

Polymers in the crystalline phase are in a rigid environment. Their high-resolution solid-state NMR spectra are most efficiently observed using cross-polarization and magic-angle spinning. The signals from the crystalline phase can often be distinguished from the amorphous and interfacial material by choosing a short cross-polarization contact time or in the CPT_1 experiment (Figure 6) by choosing a delay time longer than the relaxation time of the more mobile material.

Solid-state ^{2}H-NMR has been used to study the molecular dynamics of the crystalline and amorphous phases of nylon 66 **(7)** in polymers labeled with deuterium at several sites. This approach is illustrated in Figure 44, which shows the wideline deuterium spectrum at 97 °C of polymers labeled at carbons C1 and C6, C2 and C5 and C3 and C4 of the diamine moiety (NY16NHME, NY25NHME, and NY34NHME) and at the C2 and C5, and C3 and C4 positions of the adipoyl moiety (NY25COME and NY34COME).[92] These spectra show that the molecular dynamics of the various methylenes are not identical under these conditions.

The molecular dynamics of the crystalline and amorphous phases of nylon 66 can be individually measured since the differences in the dynamics of the crystalline and amorphous phases give rise to large differences in the spin–lattice relaxation rates.[93] The spectra for the labeled polymers were acquired and several models for the molecular dynamics of nylon 66 were examined to explain the temperature dependence of the line shapes.[93] The best fit was obtained for models incorporating libration motions that change in amplitude as a function of

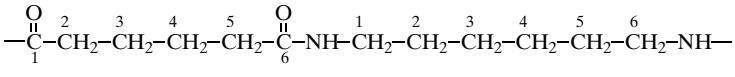

Nylon 66

(7)

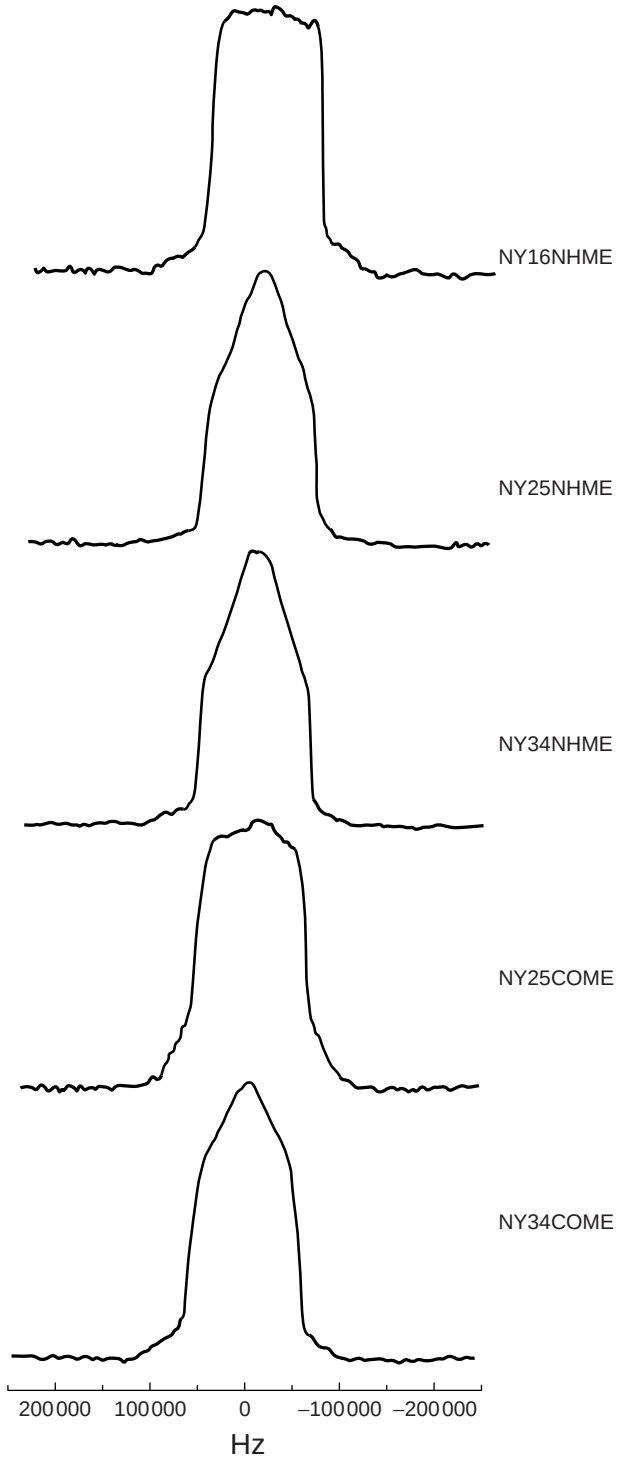

Figure 44 Fully relaxed ^{2}H-NMR spectra at 97 °C of nylon 66 polymers selectively deuterated on the diamine moiety at the C1 and C6 carbons (NY16NHME), the C2 and C5 carbons (NY25NHME) and the C3 and C4 carbons (NY34NHME), and on the adipoyl moiety at the C2 and C5 carbons (NY25COME) and the C3 and C4 carbons (NY34COME). (Adapted with permission from Miura and English.[92] Copyright 1988 American Chemical Society.)

temperature, and Figure 45 compares the experimental and simulated line shape for nylon 66 labeled with deuterium at the C_3 and C_4 positions of the adipoyl moiety along with the distribution of librational angles [$P(\Delta\theta)$] required to give a good fit between the experimental and simulated data. Further evidence for the librational model was obtained from the T_{1Q}-distorted line shape. The line shape could be modeled either with a two-site jump model or the librational model, but the T_{1Q}-distorted line shape is much closer to that expected for librational motion, even though the signal-to-noise ratio is low in these experiments.

The analysis of the deuterium relaxation for the nylon 66 polymers shows that the hydrogen-bonded NH groups act as ''pinning points'' that restrict the mobility of the nylon chains.[93,94] The mean correlation times for the ND groups varies between 200 and 500 ps over the temperature range 185–225 °C. It is also reported that above 160 °C the methylene groups in both the diamine and adipoyl moieties have similar correlation times ($\tau_c \approx$ 40 ps) that are much shorter than that for the ND groups. Below this temperature, however, the adipoyl groups experience larger amplitude fluctuations than do the methylenes in the diamine moiety. These experimental data are in close agreement with the motions observed in molecular dynamics simulations of nylon 66 crystals.[94]

The lower frequency dynamics of crystalline polymers has also been investigated by solid-state NMR using methods related to 2D exchange NMR. In such experiments the change in orientation of the atoms in the polymer chain is translated into a change in frequency during the mixing time of a 2D experiment and the frequencies before and after the motion can be directly correlated. These studies can be used to study the dynamic processes that are faster than the spin–lattice relaxation time.

Polyethylene is a semicrystalline polymer that has been extensively studied by solid-state NMR. An important feature in the relaxation spectrum of polyethylene is the so-called α transition that has been attributed to solid-state chain diffusion. It is possible to observe directly this diffusion between the crystalline and amorphous domains because their chemical shifts differ by ca. 2 ppm because of the differences in chain conformation.[14] Figure 46(a) and (b) show solid-state 2D exchange spectra obtained with magic-angle spinning that demonstrates chain diffusion between the crystalline and amorphous material.[95] These spectra clearly show the cross peaks (ca and ac) that arise from the transfer of material between the two phases. The jump rates and activation energies (105 kJ mol^{-1}) agree well with the values previously reported for the α process and show that chain diffusivities between 10^{-17} and 10^{-21} m s^{-1} can be measured by such NMR methods. These data also explain the NOEs observed for the crystalline and amorphous phases, the

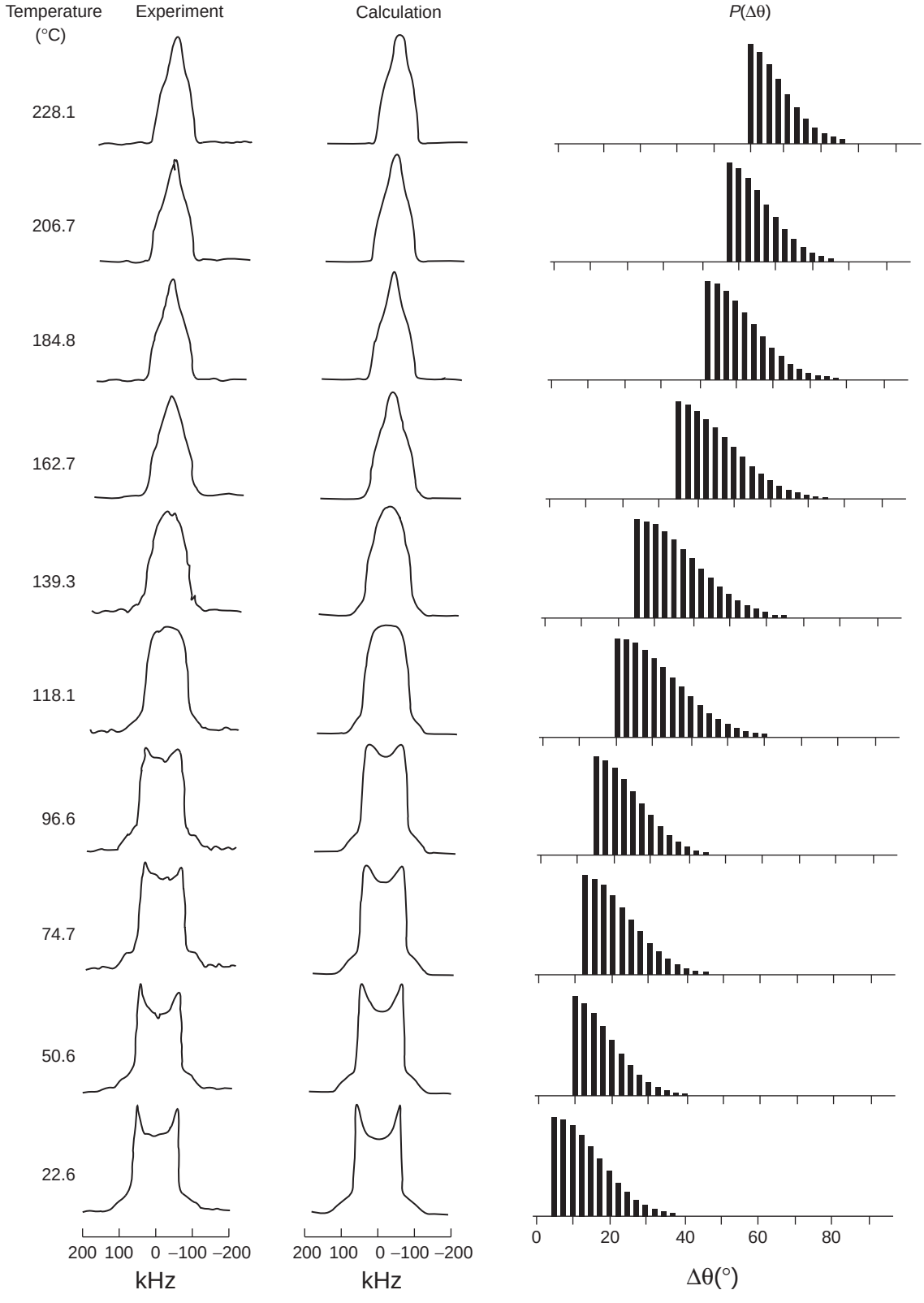

Figure 45 Comparison of the experimental and simulated line shapes for nylon 66 labeled with deuterium at the C3 and C4 positions of the adipoyl moiety as a function of temperature. Also shown is the distribution of librational angles required to fit the spectra as a function of temperature. (Adapted with permission from Hirschinger et al.[93] Copyright 1990 American Chemical Society.)

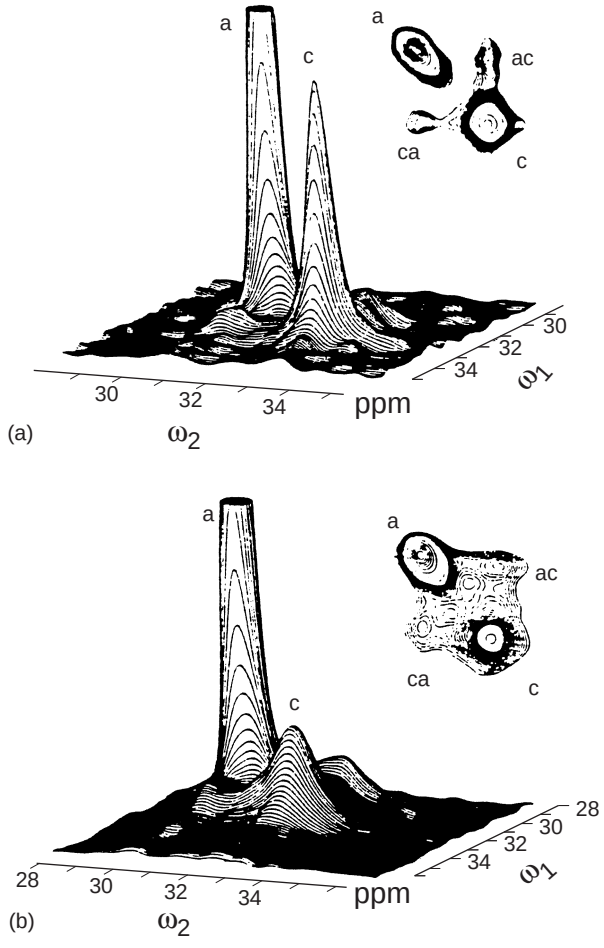

Figure 46 2D magic-angle spinning exchange spectra for high-density polyethylene at (a) 363 K with a mixing time of 1 s and (b) 373 K with a mixing time of 2 s. The cross peaks between the crystalline and amorphous phases are labeled ca and ac. (Adapted with permission from Schmidt-Rohr and Spiess.[95] Copyright 1991 American Chemical Society.)

nonexponential spin–lattice relaxation, and the dependence of the relaxation times on the crystallite size.

Deuteron 2D exchange NMR has been used to study the ultraslow motion in the crystalline phase of poly(vinylidene fluoride) to understand the unusual α relaxation that occurs at ca. 370 K. Figure 47 shows the experimental 2D spectra for a mixing time of 0.2 s. The best fit between the experimental and simulated spectra is obtained for a model in which there are either 67° or 113° jumps between sites in the crystalline lattice. These data are only consistent with one crystal structure of poly(vinylidene fluoride),[96] and from these data it can be concluded that chain motion is characterized by an electric dipole moment transition only along the molecular direction and by a conformational change between the $tgtg^-$ and the g^-tgt conformations. The same methods have been used to identify the 113° jumps of the monomer is the crystalline phase of isotactic polypropylene.[97]

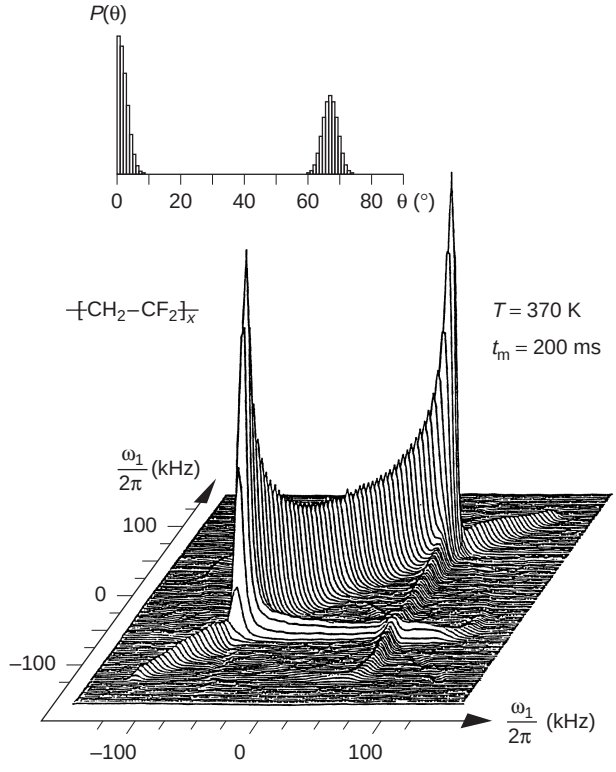

Figure 47 Experimental wideline deuterium exchange spectra of poly(vinylidine fluoride) at 370 K with a mixing time of 0.2 s. (Adapted with permission from Hirschinger et al.[124] Copyright 1991 American Chemical Society.)

4.3 Amorphous Polymers

Solid-state NMR has been also used to study the molecular dynamics of amorphous polymers in an attempt to gain a molecular level understanding of the relationship between the structure of polymers and their macroscopic properties. Included in these studies are the use of NMR relaxation to obtain molecular level assignments for the relaxation peaks observed in dielectric and dynamic mechanical spectroscopy and the use of these methods to understand the impact strength, gas permeability, and the other macroscopic properties of amorphous polymers.

As in the case of crystalline polymers, the molecular dynamics of glassy polymers span a range of timescales from the very rapid methyl group rotation to the very slow-motion chain diffusion. Because amorphous polymers are by their very nature disordered, the relaxation times are often nonexponential because the chains exist in a variety of molecular environments. In the previous section we saw that chain motion can often be described by a two-site jump model in the crystalline lattice. In amorphous polymers the motion is more

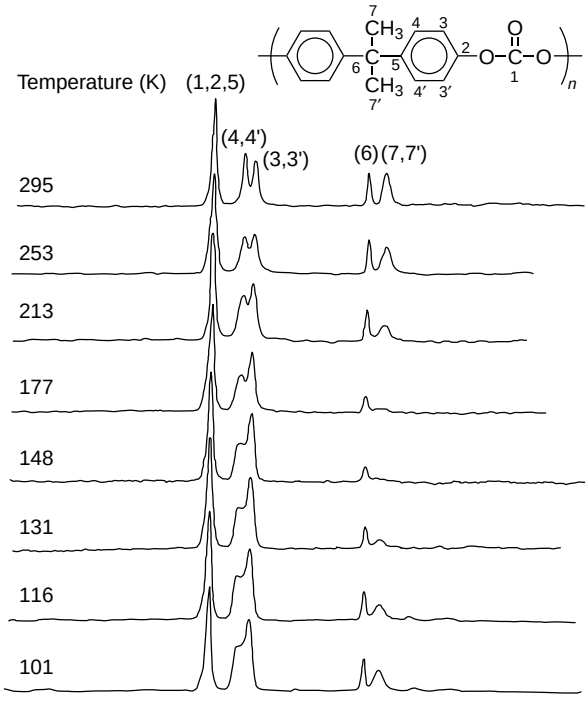

Figure 48 ^{13}C CPAMS of polycarbonate as a function of temperature. (Adapted with permission from Lyerla.[83] Copyright 1986 American Chemical Society.)

accurately described by a distribution of correlations times spanning several orders of magnitude.

The most rapid motions observed in polymers are the localized dynamics such as methyl group rotation that can interfere with dipolar decoupling and magic-angle spinning. Such an effect is demonstrated in Figure 48, which shows the effect of temperature on the magic-angle spinning spectra of polycarbonate.[83] At high temperature the spectrum of polycarbonate shows resolved resonances for several of the peaks, including the protonated aromatic carbons, the quaternary carbon, and the methyl carbon. Similar, but slightly broadened, spectra are observed at lower temperatures, while at intermediate temperatures the methyl signal broadens and disappears from the spectra, presumably owing to interference with the dipolar decoupling. The protonated aromatic carbons also show changes as a function of temperature. This line shape results from aromatic carbons in a variety of magnetic environments that lead to inhomogeneous broadening from conformational and ring current effects. At 295 K there is a rapid averaging among these conformations and sharp lines are observed.

The relaxation of amorphous polymers has been studied for a variety of polymers, mostly by carbon NMR, and it has been reported that the carbon T_1, $T_{1\rho}$, and T_{CH} relaxation times can be used to monitor the chain dynamics over a range of frequencies.[90] An analysis of the $T_{1\rho}$(C) relaxation in a variety of glassy polymers has shown that the relaxation is primarily due to spin–spin rather than spin–lattice relaxation. Therefore, $T_{1\rho}$(C) is a valuable probe of chain dynamics in the 10–50 kHz region while the T_1 relaxation provides information about chain motion in the 5–100 MHz regime.

The relaxation data for several glassy polymers, including poly(methyl methacrylate), polycarbonate, poly(ether sulfone), poly(2,6-dimethylphenylene oxide), polystyrene, and poly(vinyl chloride), were reported in a classical early study of the molecular dynamics by NMR.[90] The results showed both how the relaxation times provide information on the molecular dynamics and how such relaxation times can be related to important macroscopic properties. A correlation has been reported between the dynamics measured by NMR and the notched impact strength for several glassy polymers.[90]

Because of their commercial importance, the molecular dynamics of polystyrene and polycarbonate have been the focus of much attention in the polymer community. The molecular dynamics of polystyrene have been investigated by deuterium NMR for both main-chain and side-chain-labeled samples.[84] Two dynamic populations were observed for ring-labeled polystyrene on the basis of the spin–lattice relaxation studies. The line shapes for the two components can be separated using the differences in the relaxation times. Figure 49 shows the wideline deuterium spectra of atactic polystyrene-d_5 below the glass transition temperature. Using a long relaxation delay, a fully relaxed spectrum is observed that appears to have contributions both from rigid and motionally averaged materials. The line shape for the motionally averaged material can be observed with a short relaxation delay (0.05 s), and is typical of that expected for aromatic rings undergoing 180° flips. It is estimated that the correlation time for ring flipping must be of the order of 10^{-7}–10^{-8} s to give the motionally averaged spectra for the mobile material. The dynamics of polystyrene have also been studied by solid-state ^{13}C-NMR relaxation and rotational dipolar spin echo for a variety of ring- and main-chain-substituted polystyrene derivatives.[98]

These same techniques, and others, have been used to investigate the molecular dynamics of polycarbonate. Figure 50(a–d) show the wideline deuterium NMR spectra of methyl- and ring-labeled polycarbonate at 293 K.[84] The methyl deuterons give a Pake spectrum at both temperatures that can be fitted to a model containing only methyl group rotation, demonstrating that the motion of the C–CH_3 bond vector in the glassy state is in the slow-motion limit on the deuterium NMR timescale. The ^{2}H spectra for the ring-labeled polycarbonate show the characteristic line shape observed for aromatic rings undergoing 180° flips. These spectra are best described by a model in which the rings undergo

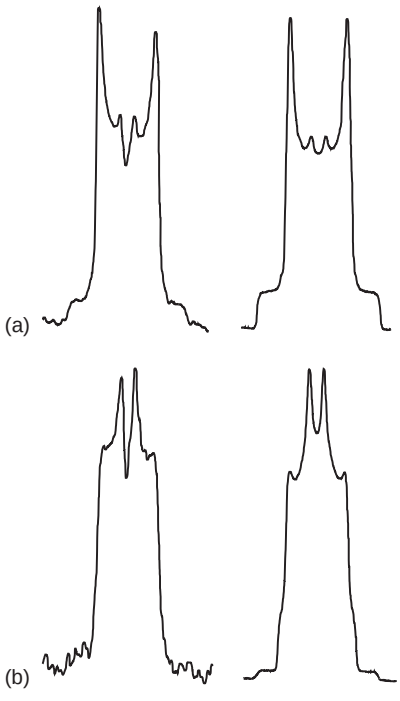

Figure 49 Experimental (a) and simulated (b) wideline ^{2}H-NMR spectra of polystyrene-d_5 with a recycle time of (a) 3 s and (b) 0.05 s at 373 K. The line shapes were simulated assuming that 20% of the polystyrene rings undergo 180° flips. (Adapted from Spiess[84] by permission of Springer-Verlag New York, Inc.)

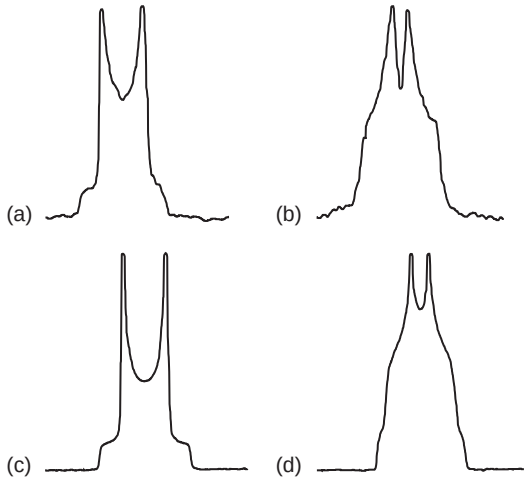

Figure 50 Experimental [(a) and (b)] and simulated [(c) and (d)] deuterium line shapes spectra for [(a) and (c)] methyl-deuterated and [(b) and (d)] ring-deuterated polycarbonate at 293 K. (Adapted from Spiess[84] by permission of Springer-Verlag New York, Inc.)

small-amplitude oscillations in addition to the ring flips. The amplitude of the oscillations increases from ±15° at room temperature to ±35° at 380 K with a Gaussian distribution of amplitudes with a variance of 10°.

The chemical shift anisotropy line shape of polycarbonate labeled with ^{13}C at the ring position ortho to the carbonate group has been also been used to characterize the molecular dynamics of polycarbonate over a range of temperatures. Good agreement between the experimental and simulated data was obtained for models containing combinations of ring flips and restricted rotations.[99,100] These data were combined with measurements of the proton T_1 and $T_{1\rho}$ relaxation times in an attempt to relate the NMR relaxation with the relaxation peaks measured by dynamic mechanical and dielectric spectroscopy. The correlation between the data measured by NMR and other methods shows that such NMR experiments can be used to obtain a molecular level understanding of the relaxation peaks observed by other spectroscopic methods.[100] Rotational dipolar spin echo has also been used to investigate the molecular dynamics of bisphenol polycarbonate and derivatives modified on the links between the aromatic rings and for ring-substituted polycarbonates. Unlike the ^{2}H data reported for polycarbonate,[84] it was found that 20° amplitude oscillations of the methyl groups are required for the best fit of the experimental data. It is also reported that chlorine substitution on the aromatic rings quenches ring flips and substitutions at the linking groups between the rings can also inhibit chain motion.

The molecular dynamics of amorphous polymers have also been investigated as a function of carbon dioxide and hydrostatic pressure and aging. The effect of carbon dioxide on the dynamics of polymers is of interest since the molecular level processes leading to gas permeability in polymers are not well understood. The pressure of carbon dioxide (up to 35 atm) has only a small effect on the deuterium line shape of deuterated polycarbonate, polystyrene, or poly(methyl methacrylate).[101] The T_1s for polystyrene showed the greatest temperature sensitivity while polycarbonate was intermediate and poly(methyl methacrylate) showed no effect. The effect of pressure on the T_1s for polycarbonate and polystyrene can be correlated with the pressure dependence of the carbon dioxide diffusivity. These data suggest that either the polymer dynamics measured by deuterium NMR are altered by the gas, or that both the NMR parameters and the diffusivity have the same dependence of some other process, such as swelling or plasticization. Hydrostatic pressure has been shown to affect ring flips in polycarbonate[102] and polystyrene,[103] but has only a small effect on the main-chain dynamics of polystyrene.[104]

The effect of aging on the dynamics of polycarbonate and poly(ester carbonates) derived from bisphenol-A polycarbonate and tere- or iosphthalic acid has also been studied. Physical aging has little effect on bisphenol-A polycarbonate, but leads to small line shape changes in the terephthalate-labeled rings in poly(ester carbonates).[105]

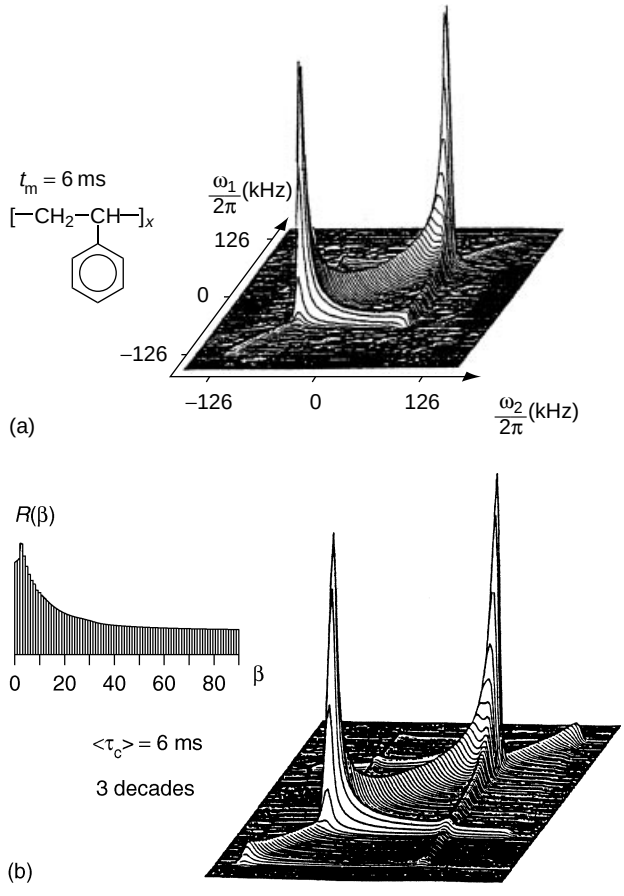

Figure 51 The experimental and simulated 2D deuterium exchange spectra for chain-deuterated polystyrene at 391 K. (Adapted with permission from Kaufmann et al.[106] Copyright 1990 American Chemical Society.)

Two-dimensional wideline exchange NMR has been used to investigate the molecular dynamics of several amorphous polymers, including polystyrene, atactic polypropylene, and polyisoprene, at temperatures above T_g where the chain dynamics are in the range that can be measured by 2D NMR. Figure 51(a) and (b) show the deuterium 2D experimental and simulated 2D exchange spectra for main-chain-labeled polystyrene at 391 K, 10 K above the T_g.[106] The 2D ridge pattern from exchange is more diffuse than in the crystalline polymers, as expected for a more heterogeneous material. The key features of this pattern include extended ridges in the 2D plane at the frequency of the singularities and the decay and broadening of the center of the powder pattern along the diagonal. The pattern of cross-peak intensity in the 2D spectra as a function of mixing time and temperature cannot be described by a model with a single correlation time, but is better fitted by a model with a broad distribution of correlation times. For polystyrene at this temperature the mean correlation time is 0.006 s and the width of the distribution is 3.5 decades. The inset plot

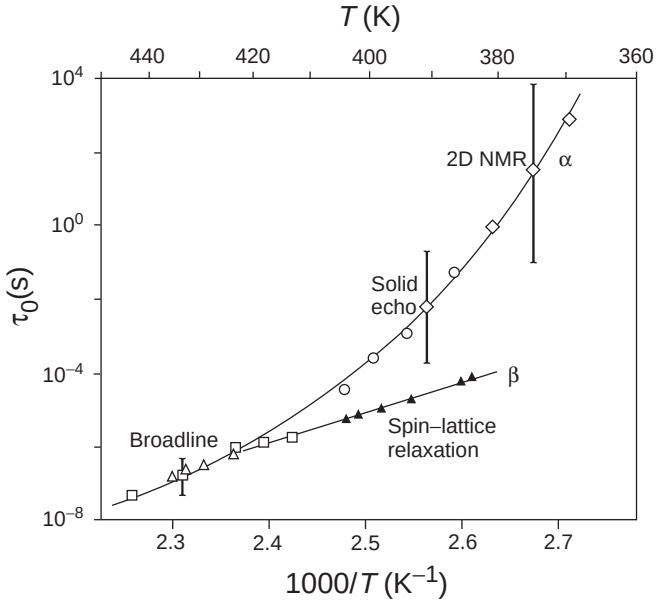

Figure 52 Arrhenius plot of the correlation times for the main-chain dynamics of polystyrene measured by several NMR methods, including 2D exchange NMR, the solid echo, T_1 measurements and broadline NMR. (Adapted with permission from Pschorn et al.[107] Copyright 1991 American Chemical Society.)

in the simulation shows the final angular distribution of C–D bond vectors after the mixing time. Note that the differences in the simulated and experimental data at the edges of the spectra that arises from pulse imperfections.

The 2D NMR results for polystyrene can be combined with other data measured on the same polymer for a more global understanding of structure–property relationships in amorphous polymers. Figure 52 shows a plot of the correlation times for polystyrene acquired by 2D NMR, solid echo, broadline NMR and T_1 relaxation rate measurements that spans 12 orders of magnitude.[107] The data for the 2D exchange, the solid echo, and broadline NMR experiments lie on a single line that can be fitted to the Williams–Landel–Ferry equation[108] describing the temperature dependence of polymers above T_g. This shows that these data are a measure of the dynamics associated with the α transition. Conversely, the T_1 data deviate from this temperature dependence and are associated with the β transition. Similar behavior is observed for atactic polypropylene, where a single curve over many orders of magnitude was constructed from the 2D exchange data just above T_g and the spin–lattice relaxation measurements at higher temperatures.[97] Two-dimensional deuterium NMR has also been used to study the dynamics of 1,4-*cis*-polyisoprene above the glass transition temperature,[109] and multidimensional carbon and deuterium NMR have been reported for poly(methyl methacrylate)[110] and poly(ethyl methacrylate).[111]

4.4 Multiphase Polymers

Microphase-separated polymer systems, including diblock copolymers, immiscible blends, and multiblock thermoplastic elastomers, are commercially important materials that derive their function from the structure and dynamics of their phase-separated natures. These materials have been extensively studied by NMR in order to characterize the structure and dynamics of the phases and the interfaces between them. Among the important questions are how the dynamics change with temperature, how the dynamics are affected by the chemical structure of the monomers, and how the mechanical properties are related to the dynamics measured by NMR.

The structure and dynamics of polyurethanes and phase-separated segmented copolymer have been investigated by wideline deuterium NMR and magic-angle spinning cross-polarization methods. In many of these materials there is a large difference in the molecular dynamics of the segments in ''hard'' and ''soft'' domains that result in large differences in the line shapes and relaxation times. These differences have been used to characterize the length scale of phase-separation[49] and the fraction of polymer in each phase in microphase-separated polymers.[59]

The molecular dynamics of the hard segments of polyurethanes have been studied by wideline deuterium NMR in polymers site-specifically labeled at the butanediol sites in polymers based on bis(4-isocyanophenyl) methane **(8)** and 2,4-toluenediyl diisocyanate **(9)**.[112] Figure 53(a–e) show the deuterium spectra as a function of temperature and recycle delay time to separate the more rapidly and slowly relaxing components. These spectra do not show the classical Pake patterns observed for rigid polymers, but rather show a line shape consistent with a wide distribution of correlation times. Further evidence for a wide distribution of correlation times is obtained from the multiexponential spin–lattice relaxation behavior. From the studies as a function of temperature it is clear that not only are the molecular dynamics changing, but also a fraction of polymer in the hard and soft domains is changing with temperature. The line shape for the more mobile fraction can be fitted to a two-site jump model with a jump angle near 109°, corresponding to gauche–trans isomerizations, combined with small-amplitude librational motion, although this model is by no means the only one which will fit the data.

This same approach has been used to study the dynamics of the deuterium-labeled butanediol unit in poly(butylene terephthalate) and segmented copolymers containing 96 and 87% hard segments.[113] A motionally averaged line shape is observed for poly(butylene terephthalate) at room temperature that can be fitted to a two-site jump with an angle of 103° and a rate of $1.4 \times 10^5\ s^{-1}$. Similar line shapes are observed for the hard segments in the copolymers, but an additional resonance is also observed that can be assigned to mobile deuterons in the soft segments that undergo nearly isotropic motion that is rapid on the deuterium NMR timescale. These same polymers were also studied by ^{13}C static and magic-angle spinning NMR where the chemical shift anisotropy line shapes were reported.[114] These data confirm the results from the deuterium NMR studies and show that the aromatic rings in the hard segments undergo 180° ring flips.

Similar studies have been used to probe the dynamics in model polyurethane polymers containing five piperazine rings separated by carbonyloxytetramethyleneoxycarbonyl spacers. The piperazine rings are deuterium labeled at the outer (1,5), intermediate (2,4), or center position (3) to measure the dynamics at different parts of the hard segment domains.[60] As in the other polyurethanes, the line shape is a composite of the hard and soft domains. At sufficiently low temperature all motion is frozen out and a Pake spectrum is observed, whereas at high temperature all of the domains are mobile and a motionally narrowed spectrum is observed. The relative fraction of hard and soft domains depends strongly on the position of the deuterium label, with more rigid domains observed for the polymer labeled at the 3-position relative to the polymer labeled at the 2,4-positions and the 1,5-positions. Thus there appears to be a range of molecular environments for the rings in the hard segments, ranging from the most rigid at the center to the more mobile at the surface.

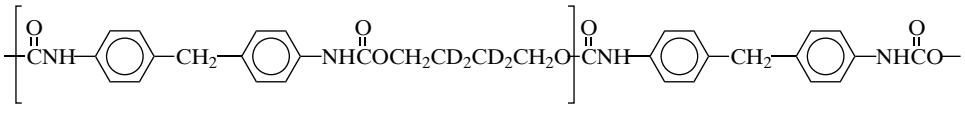

(8)

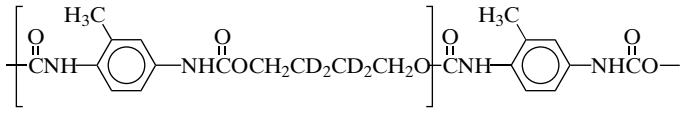

(9)

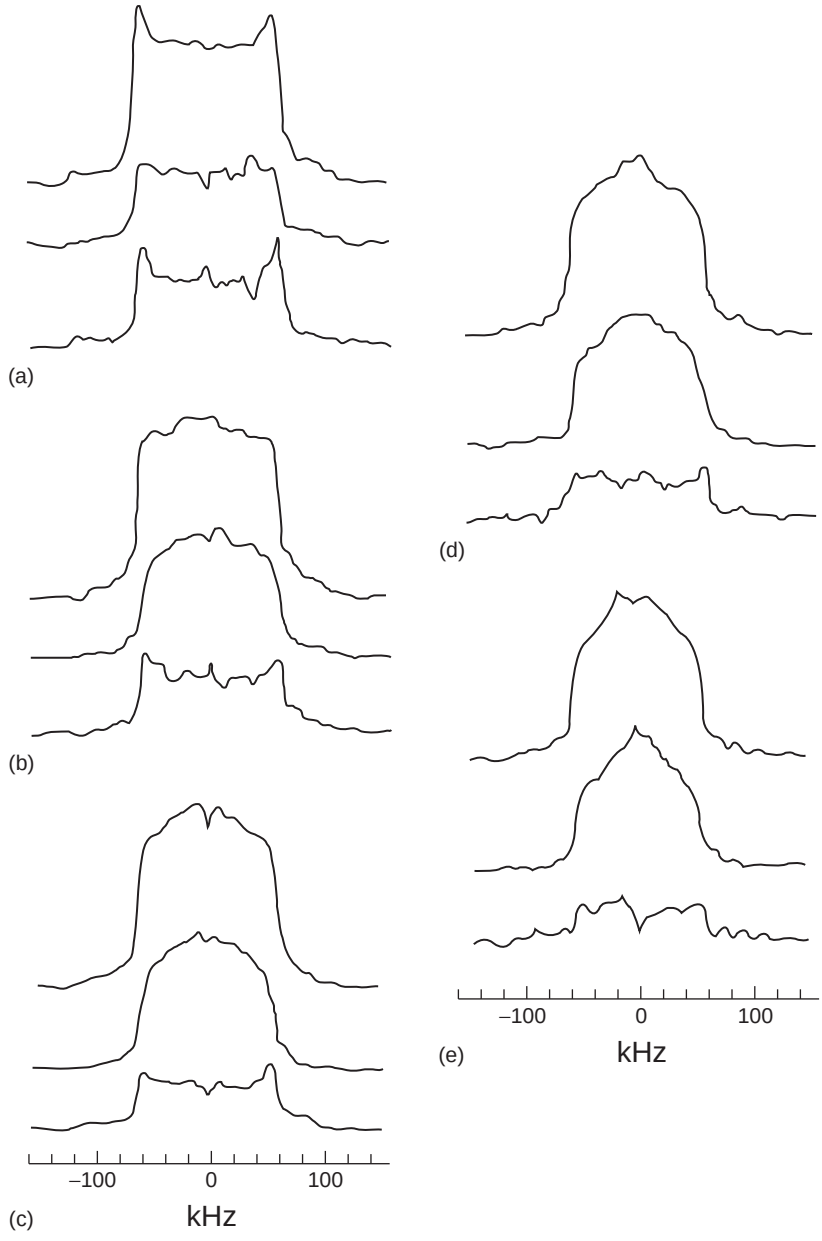

Figure 53 Solid-state ^{2}H-NMR spectra of deuterium-labeled polyurethanes as a function of temperature and recycle delay time. The spectra were recorded at (a) −45, (b) −2, (c) 22, (d) 55 and (e) 73 °C. The top row shows the spectra with a long delay (2–4 s), the middle row with a short delay (100–160 ms) and the bottom row is the difference. (Adapted with permission from Kintinar et al.[112] Copyright 1986 American Chemical Society.)

Further evidence for this interpretation is obtained from the temperature dependence of the line shape for the rings labeled at the three positions.

5 SUMMARY AND OUTLOOK

NMR has emerged as one of the premier methods for the analysis of polymers, and this impact can be judged in part by the many NMR articles published each year in polymer journals. As noted in the Introduction, NMR began primarily as a method for polymer analysis in solutions. The drive to understand the behavior of polymers in their native state has led to the extensive use of solid-state NMR. The solid-state NMR of polymers has benefited from continued advances in NMR methods and technologies, and in some cases these advances have been driven by the desire to better understand the properties of polymers in the solid state.

This article has reviewed many of the recent advances in solid-state NMR that make possible a detailed, atomic-level analysis of the properties of crystalline, amorphous, and rubbery polymers. NMR methods continue to evolve, and the prospects are bright for the development of new methods leading to a more quantitative analysis of polymers in the solid state. The advances will come in part from the development of new hardware that will make new experiments possible.

The solution NMR analysis of biomolecules is a sophisticated and rapidly evolving field, and many of the multidimensional NMR methods that are currently employed were developed to overcome the limitations of standard solution NMR methods. The NMR of solids is undergoing a similar evolution, and many of the experiments developed to study molecules in solutions are being adapted for studies of solids. We have seen some examples of this in this review, including 2D[115] and 3D[110] exchange NMR and the heteronuclear correlations in solids.[78]

Advances in NMR technologies will also fuel the development of new NMR methods to probe the structure of solid polymers. One exciting advance that promises to have many implications is the recent introduction of fast magic-angle sample spinning. It is now possible to buy probes that are capable of spinning samples at 35 kHz. This opens a whole new area of experiments because such spinning speeds are of the same order of magnitude as the dipolar couplings, and it may be possible to obtain high-resolution spectra of polymers from spinning alone. In favorable cases, 35 kHz spinning is fast enough to obtain high-resolution spectra without multipulse decoupling.[116,117] Proton NMR has many advantages, including high sensitivity and natural abundance, that have not been fully utilized because of the difficulty in obtaining proton NMR spectra using multipulse NMR. Fast magic-angle spinning opens up new areas for measuring the structure of polymers because proton spin diffusion can be more fully utilized to measure distances in solids. This will also make an important contribution to one of the most fundamental problems in the solid state NMR analysis of polymers, assigning the resonances. Multiple-quantum NMR with fast spinning could have an important impact on the polymer assignment problem.[116]

Recent advances in decoupling will also have an important impact on the NMR of polymers in the solid state. The combination of new probes that are able to take higher decoupling power and new decoupling schemes, such as TPPM,[8] lead to more efficient decoupling and better resolution, particularly in crystalline polymers where the dipolar couplings are strongest. Better decoupling leads to improvements in many 2D experiments where signal modulations in the indirectly detected dimensions often decay rapidly when the decoupling field is not strong enough.

The success of solution NMR for the analysis of complex materials has advanced in part because of the abundant through-bond scalar couplings that make complex magnetization transfer pathways possible. It is now becoming possible to observe these couplings in solids with new decoupling schemes,[118] such as frequency-switched Lee–Goldberg (FSLG) decoupling.[119] These advances may make it possible to transfer many of the pulse sequences used for solution analysis to the solid-state, leading to much better assignment methods.

The NMR imaging of polymers is a very rapidly evolving field that will have an important impact on the analysis of many types of polymers. To date, much of the imaging of polymers has been on elastomers, because they are easily imaged because molecular motion above T_g narrows the lines to such a degree that high-resolution images can be obtained. New pulse sequences and methods are being developed to overcome this limitation and make possible the NMR imaging of rigid solid materials.[36] The development of mobile NMR imaging systems, such as the NMR mobile universal surface explorer (MOUSE),[120] will make it possible to study a much wider variety of samples.

The advances in multidimensional solution NMR have been made possible in part by the use of pulse-field gradients for coherence selection and the suppression of unwanted resonances. These methods have recently been introduced for the study of solids with magic-angle spinning,[121] and may have a similar affect on solid-state NMR.

In addition to the advances that are driven by the development of new hardware, new methods for the analysis of NMR signals promise to increase the amount of information that we can extract from NMR experiments. One example is the new transforms being developed to analyze the signals in the rotational echo double resonance (REDOR) experiments that are used to measure distances in solids.[122] These simplify the data analysis and may make it possible to measure distances in solids routinely.

In summary, it is fair to say that solid-state NMR has an important impact on many areas of polymer science, ranging from materials characterization to understanding structure–property relationships. Given the enormous interest in NMR and polymers and the rapid growth in NMR methods and technology, the future is very bright.

ABBREVIATIONS AND ACRONYMS

CPMAS	Cross-polarization/Magic-angle Spinning

CRAMPS	Combined Rotation and Multipulse Spectroscopy
FID	Free Induction Decay
FSLG	Frequency-switched Lee–Goldberg
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
MOUSE	Mobile Universal Surface Explorer
MRI	Magnetic Resonance Imaging
*n*D	*n*-dimensional
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PVME	Polystyrene–poly(methyl vinyl ether)
REDOR	Rotational Echo Double Resonance
rf	radiofrequency
TOSS	Total Suppression of Sidebands
TPPM	Two-pulse Phase Modulation
WISE	Wideline Separation
1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Multinuclear Magnetic Resonance Spectroscopic Imaging

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Nuclear Magnetic Resonance, Imaging of Polymers

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Relaxation in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance

REFERENCES

1. N.L. Alpert, 'Studies of the Solid-state by Means of Nuclear Magnetism', *Phys. Rev.*, **72**, 637–638 (1947).
2. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments', *Phys. Rev.*, **69**, 37–38 (1946).
3. F. Bloch, W.W. Hansen, M.E. Packard, *Phys. Rev.*, **69**, 127 (1946).
4. F. Bloch, *Phys. Rev.*, **70**, 460 (1946).
5. N. Bloembergen, E.M. Purcell, R.V. Pound, 'Relaxation Effects in Nuclear Magnetic Resonance Absorption', *Phys. Rev.*, **73**, 679–712 (1948).
6. E. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy*, 3rd edition, VCH, Weinheim, 1987.
7. S.R. Hartmann, E.L. Hahn, 'Nuclear Double Resonance in the Rotating Frame', *Phys. Rev.*, **128**, 2042–2053 (1962).
8. A.E. Bennett, C.M. Rienstra, M. Auger, K.V. Lakshimi, R.G. Griffin, 'Heteronuclear Decoupling in Rotating Solids', *J. Chem. Phys.*, **103**, 6951–6958 (1995).
9. D.A. Torchia, 'The Measurement of Carbon-13 T_1 Values by a Method that Suppresses Artifacts', *J. Magn. Reson.*, **30**, 613–616 (1978).
10. E. Stejskal, J. Schaefer, M. Sefcik, R. McKay, 'Magic-angle Carbon-13 NMR Study of the Compatibility of Solid Polymeric Blends', *Macromolecules*, **14**, 275–279 (1981).
11. S.J. Opella, M.H. Frey, 'Selection of Nonprotonated Carbon Resonances in Solid-state Nuclear Magnetic Resonance', *J. Am. Chem. Soc.*, **101**, 5854–5856 (1979).
12. W.T. Dixon, 'Spinning-sideband Free and Spinning Sideband Only NMR Spectra in Spinning Samples', *J. Chem. Phys.*, **77**, 1800–1809 (1982).
13. B.C. Gerstein, R.G. Pembleton, R.C. Wilson, L.M. Ryan, 'High-resolution NMR in Randomly Oriented Solids with Homonuclear Dipolar Broadening: Combined Multiple-pulse and Magic-angle Spinning', *J. Chem. Phys.*, **66**, 361–362 (1977).
14. A. Tonelli, *NMR Spectroscopy and Polymer Microstructure: the Conformational Connection*, VCH, New York, 1989.
15. A.E. Tonelli, M.A. Gomes, H. Tanaka, M.H. Cozine, *Solid-State ^{13}C-NMR Studies of the Structures, Conformations, and Dynamics of Semicrystalline Polymers*, Plenum Press, New York, 1991.
16. W.L. Earl, D.L. vander Hart, 'Observations in Solid Polyethylenes by ^{13}C-NMR with Magic-angle Spinning', *Macromolecules*, **12**, 762–767 (1979).
17. V.J. McBrierty, K.H. Packer, *Nuclear Magnetic Resonance in Solid Polymers*, Cambridge University Press, New York, 1993.
18. D.L. Vander Hart, 'Influence of Molecular Packing on Solid-state ^{13}C Chemical Shifts: the *n*-Alkanes', *J. Magn. Reson.*, **44**, 117–125 (1981).

19. K. Schmidt-Rohr, 'Complete Dipolar Decoupling of ^{13}C and its Use in Two-dimensional Double-quantum Solid-state NMR for Determining Polymer Conformations', *J. Magn. Reson.*, **131**, 209–217 (1998).
20. A. Bax, R. Freeman, T. Frenkiel, 'An NMR Technique for Tracing Out the Carbon Skeleton of an Organic Molecule', *J. Am. Chem. Soc.*, **103**, 2102–2104 (1981).
21. A. Bunn, E. Cudby, R. Harris, K. Packer, B. Say, *Chem. Commun.*, 15 (1981).
22. D. Powell, A. Sikes, L. Mathias, '^{15}N Solid-state NMR Characterization of Aramid-Containing nylon-6 by In situ Polymerization with Benzoyl Caprolactam Derivatives', *Macromolecules*, **21**, 1533–1536 (1988).
23. F.A. Bovey, F.C. Schilling, 'The Solid-state ^{29}Si and ^{13}C-NMR of poly(di-*n*-alkyl silanes)', in *Solid-state NMR of Polymers*, ed. L.J. Mathias, Plenum Press, New York, 295–304, 1991.
24. M.A. Gomez, H. Tanaka, A.E. Tonelli, *Polymer*, **28**, 2227 (1987).
25. D.E. Axelson, *Carbon-13 Solid-state NMR of Semicrystalline Polymers*, VCH, Deerfield Beach, FL, 1986.
26. J.J. Dechter, '^{1}H and ^{13}C Relaxation for the Resolved Crystalline and Amorphous Phases of Polyethylene Oxide', *J. Polym. Sci. Poly. Lett. Ed.*, **23**, 261–266 (1985).
27. S. Kuroki, K. Okita, T. Kakigano, J. Ishikawa, M. Itoh, 'Thermosetting Mechanism Study of Poly[(phenylsilylene)ethynylene-1,3-phenyleneethynylene] by Solid-state NMR Spectroscopy and Computational Methods', *Macromolecules*, **31**, 2804–2808 (1998).
28. G. Natta, P. Corradini, *Nuovo Cimento Suppl.*, **1**, 40 (1960).
29. F.C. Schilling, M.A. Gomez, A.E. Tonelli, F.A. Bovey, A.E. Woodward, 'Variable Temperature, High-resolution Solid-state Carbon-13 NMR Study of 1,4-*trans*-Polybutadiene', *Macromolecules*, **20**, 2954–2957 (1987).
30. J. Suehiro, M. Takayangagi, *Macromol. Sci.*, **B4**, 39 (1970).
31. C. De Rosa, R. Napolitano, B. Pirozzi, *Polymer*, **26**, 2039 (1986).
32. H. Tanaka, M.A. Gomez, A.E. Tonelli, M. Takur, 'Thermochromic Phase Transition of a Polydiacetylene, Poly(ETCD), Studied by High-resolution Solid-state ^{13}C-NMR', *Macromolecules*, **22**, 1208–1215 (1989).
33. F. Schilling, F. Bovey, A. Lovinger, J. Zeigler, 'Characterization of Poly(di-*n*-hexylsilane) in the Solid-state. 2. ^{13}C and ^{29}Si Magic-angle Spinning NMR Studies', *Macromolecules*, **19**, 2660–2663 (1986).
34. A.D. English, 'Solid-state NMR Investigation of the Contribution of Anisotropic Segmental Motion to Chain Dynamics in *cis*-1,4-Polyisoprene', *Macromolecules*, **18**, 178–181 (1985).
35. N.C. Nielsen, H. Bildsoe, H.J. Jakobsen, 'Spectral Editing in Solid-state ^{13}C-NMR of Elastomers', *Macromolecules*, **25**, 2847–2853 (1992).
36. B. Blumich, 'NMR Imaging of Polymer Materials', *Makromol. Chem.*, **194**, 2133–2161 (1993).
37. D.J. Cory, J.C. de Boer, W.S. Veeman, 'Magic-angle Spinning ^{1}H-NMR Imaging of Polybutadiene/Polystyrene Blends', *Macromolecules*, **22**, 1618–1621 (1989).
38. D.G. Cory, J.B. Miller, A.N. Garroway, 'Chemically Resolved Solid-state NMR Images of Polymers', *Macromol. Symp.*, **86**, 259–269 (1994).
39. P. Blumler, B. Blumich, 'Magnetization Filters: Applications to the NMR Imaging of Elastomers', *Magn. Reson. Imaging*, **10**, 779 (1992).
40. S.N. Sarkar, R. Komoroski, 'NMR Imaging of Morphology, Defects, and the Composition of Tire Composites and Model Elastomer Blends', *Macromolecules*, **25**, 1420–1426 (1992).
41. H.D.H. Stover, J.M.J. Frechet, 'Direct Polarization ^{13}C and ^{1}H Magic-angle Spinning NMR in Characterization of Solvent-swollen Gels', *Macromolecules*, **22**, 1574–1576 (1989).
42. P. Sozzani, F.A. Bovey, F.C. Schilling, 'Characterization of Isolated Polyethylene Chains in the Solid-state', *Macromolecules*, **24**, 6764–6768 (1991).
43. P. Sozzani, R.W. Behling, F.C. Schilling, S. Bruckner, E. Helfand, F.A. Bovey, L.W. Jelinski, 'Traveling Defects in 1,4-*trans*-Polybutadiene as a Inclusion Complex in Perhydrotriphenylene Canals and a Comparison with Molecular Motions in the Crystalline Solid-state', *Macromolecules*, **22**, 3318–3322 (1989).
44. S. Bruckner, P. Sozzani, C. Boeffel, S. Destri, G. Di Silvestro, 'Chain Mobility of Isotactic 1,4-*trans*-Poly(penta-1,3-diene) Included in Perhydrotriphenylene. A Broadline ^{2}H-NMR Study', *Macromolecules*, **22**, 607–611 (1989).
45. M. Goldman, L. Shen, 'Spin–spin Relaxation in LaF_3', *Phys. Rev.*, **144**, 321–331 (1966).
46. T. Cheung, B. Gerstein, '^{1}H-NMR Studies of Domain Structures in Polymers', *J. Appl. Phys.*, **52**, 5517–5528 (1981).
47. W.Z. Cai, K. Schmidt-Rohr, N. Egger, B. Gerharz, H.W. Spiess, 'A Solid-state NMR Study of Microphase Structure and Segmental Dynamics of Poly(styrene-*b*-methylphenylsiloxane) Diblock Copolymers', *Polymer*, **34**, 267–276 (1993).
48. J. Clauss, K. Schmidt-Rohr, H.W. Spiess, 'Determination of Domain Sizes in Heterogeneous Polymers by Solid-state NMR', *Acta Polym.*, **44**, 1–17 (1993).
49. R. Assink, 'Nuclear Spin diffusion Between Polyurethane Microphases', *Macromolecules*, **11**, 1233–1237 (1978).
50. S. Spiegel, K. Schmidt-Rohr, C. Boeffel, H.W. Spiess, '^{1}H Spin diffusion Coefficients of Highly Mobile Polymers', *Polym. Commun.*, **34**, 4566–4569 (1993).
51. N. Egger, K. Schmidt-Rohr, B. Blumich, W.D. Domke, B. Stapp, 'Solid-state NMR Investigation of Cationic Polymerized Epoxy Resins', *J. Appl. Polym. Sci.*, **44**, 289 (1992).

52. C. Schmidt-Rohr, H.W. Spiess, 'Determination of Domain Size in Heterogeneous Polymers by Solid-state NMR', *Acta Polym.*, **44**, 1–17 (1993).
53. D.L. Vanderhart, G.C. Campbell, R.M. Briber, 'Phase-Separation Behavior in Blends of Poly(benzimidazole) and Poly(ether imide)', *Macromolecules*, **25**, 4734–4743 (1992).
54. P. Caravatti, P. Neuenschwander, R. Ernst, 'Characterization of Polymer Blends by Spin Diffusion NMR Measurements', *Macromolecules*, **19**, 1889–1995 (1986).
55. P. Caravatti, P. Neuenschwander, R. Ernst, 'Characterization of Heterogeneous Polymer Blends by Two-dimensional Proton Spin diffusion Spectroscopy', *Macromolecules*, **18**, 119–122 (1985).
56. B. Blumich, A. Hagemeyer, D. Schaefer, K. Schmidt-Rohr, H.W. Spiess, 'Solid-state NMR Spectroscopy in Polymer Science', *Adv. Mater.*, **2**, 72–81 (1990).
57. T.T.P. Cheung, B.C. Gerstein, L.M. Ryan, R.E. Taylor, C.R. Dybowski, 'High-resolution ^{1}H-NMR Studies of Polyethyleneterephthalate', *J. Chem. Phys.*, **73**, 6059–6067 (1980).
58. R.R. Havens, D.L. VanderHart, 'Morphology of Poly-(ethylene terephthalate) Fibers as Studied by Multipulse NMR', *Macromolecules*, **18**, 1663–1676 (1985).
59. J.J. Dumais, L.W. Jelinski, L.M. Leung, I. Gancarz, A. Galambos, J.T. Koberstein, 'Phase-separation in Polyurethanes – A Deuterium NMR Study', *Macromolecules*, **18**, 116–119 (1985).
60. J.A. Kornfeld, H.W. Spiess, H. Nefzger, H. Hayden, C.D. Eisenbach, 'Deuteron NMR Measurement of Order and Mobility in the Hard Segments of Model Polyurethanes', *Macromolecules*, **24**, 4787–4795 (1991).
61. K. Schmidt-Rohr, J. Clauss, B. Blumich, H.W. Spiess, 'Miscibility of Polymer Blends Investigated by ^{1}H Spin Diffusion and ^{13}C Detection', *Magn. Reson. Chem.*, **28**, S3 (1990).
62. H. Spiess, 'Structure and Dynamics of Polymers from 2D and 3D NMR', *Chem. Rev.*, **91**, 1321–1338 (1991).
63. D. Walsh, S. Rostami, 'The Miscibility of High Polymers: The Role of Specific Interactions', *Adv. Polym. Sci.*, **70**, 119–169 (1985).
64. A. Simmons, A. Natansohn, '^{13}C CPMAS NMR Chemical Shift and Molecular Dynamics Study of Charge-transfer Interactions in Blends of Poly(2-[(3,5-dinitrobenzoyl)oxy]ethyl methacrylate) with Poly([*N*-ethylcarbazole-3-yl)methyl methacrylate or acrylate]', *Macromolecules*, **25**, 3881–3889 (1992).
65. L.A. Belfiore, T.J. Lutz, C. Cheng, 'Solid-state NMR Detection of Molecular Level Mixing Phenomena in Strongly Interacting Polymer Blends and Phase-separated Copolymers', in *Solid-state NMR of Polymers*, ed. L.J. Mathias, Plenum Press, New York, 146–165, 1991.
66. J. Grobelny, D. Rice, F. Karasz, W. MacKnight, 'High-resolution Solid-state Carbon-13 NMR Study of Polybenzimidazole/Polyimide Blends', *Macromolecules*, **23**, 2139–2144 (1990).
67. E. Stejskal, J. Schaefer, M. Sefcik, R. McKay, 'Magic-angle Carbon-13 NMR Study of the Compatiblity of Solid Polymeric Blends', *Macromolecules*, **14**, 275–279 (1981).
68. G. Gobbi, R. Silvestri, R. Thomas, J. Lyerla, W. Flemming, T. Nishi, 'Intermolecular Polarization Transfer Study of Polymer Blend Compatibility', *J. Polym. Sci. Part C, Polym. Lett.*, **25**, 61–65 (1987).
69. J. Schaefer, M.D. Sefcik, E.O. Stejskal, R.A. McKay, 'Magic-angle Carbon-13 NMR Analysis of the Interface Between Phases in a Blend of Polystyrene with a Polystyrene–Polybutadiene Block Copolymer', *Macromolecules*, **14**, 188–192 (1981).
70. V. McBrierty, D. Douglass, T. Kwei, 'Compatiblity in Blends of Poly(methyl methacrylate) and Poly(styrene-co-acrylonitrile)', *Macromolecules*, **11**, 1265–1267 (1978).
71. L. Dickinson, H. Yang, C. Chu, R. Stein, J. Chein, 'Limits to Compatibility in Poly(*x*-methylstyrene)/Poly(2,6-dimethylphenylene oxide) Blends by NMR', *Macromolecules*, **20**, 1757–1760 (1987).
72. K. Schmidt-Rohr, J. Clauss, H. Spiess, 'Correlation of Structure, Mobility, and Morphological Information in Heterogeneous Polymer Materials by Two-Dimensional Wideline Separation NMR Spectroscopy', *Macromolecules*, **25**, 3273–3277 (1992).
73. J.L. White, P.A. Mirau, 'Probing Miscibility and Intermolecular Interactions in Solid Polymer Blends Using the Nuclear Overhauser Effect', *Macromolecules*, **26**, 3049–3054 (1993).
74. P. Caravatti, G. Bodenhausen, R.R. Ernst, 'Heteronuclear Solid-state Correlation Spectroscopy', *Chem. Phys. Lett.*, **89**, 363–367 (1982).
75. P. Caravatti, L. Braunschweiler, R.R. Ernst, 'Heteronuclear Correlation Spectroscopy in Rotating Solids', *Chem. Phys. Lett.*, **100**, 305–310 (1983).
76. D.P. Burum, A. Bielecki, 'An Improved Experiment for Heteronuclear Correlation 2D NMR in Solids', *J. Magn. Reson.*, **94**, 645–652 (1991).
77. A. Abragam, *Principles of Nuclear Magnetism*, Oxford University Press, New York, 1961.
78. J.L. White, P.A. Mirau, 'Heteronuclear Correlation in Solid Polymers: Identification of Hydrogen Bond Donors and Acceptors in Miscible Polymer Blends', *Macromolecules*, **27**, 1648–1650 (1994).
79. S. Kaplan, 'Compatibility of TPD/Polycarbonate Blends by ^{13}C–^{1}H HETCOR in the Solid-state', *Macromolecules*, **26**, 1060–1064 (1993).
80. S. Li, D.M. Rice, F.E. Karasz, 'Two-dimensional NMR Characterization of Short-range Order in a Miscible Blend of Polystyrene and Poly(2,6-dimethyl phenylene oxide)', *Macromolecules*, **27**, 2211–2218 (1994).
81. P.A. Mirau, J.W. White, 'Solid-state NMR Studies of Intermolecular Interactions in Solid Polymer Blends', *Magn. Reson. Chem.*, **32**, S32–S29 (1994).

82. F. Heatley, 'Nuclear Magnetic Relaxation of Synthetic Polymers in Dilute Solution', *Prog. Nucl. Magn. Reson. Spectrosc.*, **13**, 47–85 (1979).
83. J.R. Lyerla, 'High-resolution NMR of Glassy Amorphous Polymers', in *High-resolution NMR Spectroscopy of Polymers in Bulk*, ed. R. Komoroski, VCH, Deerfield Beach, FL, 63–120, 1986.
84. H. Spiess, 'Molecular Dynamics of Solid Polymers as Revealed by Deuteron NMR', *Colloid. Polym. Sci.*, **261**, 193–209 (1983).
85. G.E. Pake, *J. Chem. Phys.*, **16**, 327 (1948).
86. J. Hirschinger, A.D. English, 'Differentiation of Models of Rapid Molecular Motion by ^{2}H-NMR Line Shape Simulation', *J. Magn. Reson.*, **85**, 542–553 (1989).
87. H. Speiss, 'Deuteron Spin Alignment: A Probe for Studying Ultraslow-motions in Solids and Solid Polymers', *J. Chem. Phys.*, **72**, 6755–6762 (1980).
88. W.P. Rothwell, J.S. Waugh, 'Transverse Relaxation of Dipolar Coupled Spin Systems Under RF Irradiation: Detecting Motions in Solids', *J. Chem. Phys.*, **74**, 2721–2732 (1981).
89. A. Naito, S. Ganapathy, K. Akasaka, C.A. McDowell, 'Spin–Lattice Relaxation of ^{13}C Amino Acids Using the CP-MAS Technique', *J. Magn. Reson.*, **54**, 226–235 (1983).
90. J. Schaefer, E.O. Stejskal, R. Buchdahl, 'Magic-angle ^{13}C Analysis of Motion in Solid Glassy Polymers', *Macromolecules*, **10**, 384–405 (1977).
91. J.R. Lyerla, C.S. Yannoni, 'High-resolution Carbon-13 NMR of Polymers in the Solid-state', *IBM J. Res. Dev.*, **27**, 302–312 (1983).
92. H. Miura, A. English, 'Segmental Dynamics of Nylon 66', *Macromolecules*, **21**, 1543–1544 (1988).
93. J. Hirschinger, H. Miura, K.H. Gardner, A.D. English, 'Segmental Dynamics in the Crystalline Phase of Nylon 66: Solid-state ^{2}H-NMR Studies', *Macromolecules*, **23**, 2153–2169 (1990).
94. J. Wendoloski, K. Gardner, J. Hirschinger, H. Miura, A. English, 'Molecular Dynamics in Ordered Structures: Computer Simulations and Experimental Results for Nylon 66 Crystals', *Science*, **247**, 431–436 (1990).
95. K. Schmidt-Rohr, H. Spiess, 'Chain Diffusion Between Crystalline and Amorphous Regions in Polyethylene Detected by ^{2}D Exchange ^{13}C-NMR', *Macromolecules*, **24**, 5288–5293 (1991).
96. Y. Takahashi, H. Miyaji, K. Asai, 'Long-range Order Parameters of Form II of Poly(vinylidine fluoride) and Molecular Motion in the α Relaxation', *Macromolecules*, **16**, 1789–1792 (1983).
97. D. Schaefer, H. Spiess, U. Suter, W. Flemming, 'Two-dimensional Solid-state NMR Studies of Ultraslow Chain Motion: Glass Transition in Atactic Poly(propylene) versus Helical Jumps in Isotactic Poly(propylene)', *Macromolecules*, **23**, 3431–3439 (1990).
98. J. Schaefer, M.D. Sefcik, E.O. Stejskal, R.A. McKay, W.T. Dixon, R.E. Cais, 'Molecular Motion in Glassy Polystyrenes', *Macromolecules*, **17**, 1107–1118 (1984).
99. P.T. Inglefield, R.A. Amici, J.F. O'Gara, C.C. Jung, A.A. Jones, 'Solid-state Carbon and Proton Line Shapes for the Characterization of Phenylene Group Motion in Polycarbonates', *Macromolecules*, **16**, 1552–1554 (1983).
100. J.F. O'Gara, A.A. Jones, C.C. Hung, P.T. Inglefield, 'Temperature Dependence of Local Motions in Glassy Polycarbonate from Carbon and Proton NMR', *Macromolecules*, **18**, 1117–1123 (1985).
101. P.B. Smith, D.J. Moll, '^{2}H-NMR Investigation of the Plasticization Effects Induced by High-pressure Carbon Dioxide Gas on the Molecular Dynamics of Polymers', *Macromolecules*, **23**, 3250–3256 (1990).
102. M.T. Hansen, A.S. Kulik, K.O. Prins, H.W. Spiess, 'Effect of Hydrostatic Pressure on the Phenylene Motion in Polycarbonate as Revealed by ^{2}H Spin–Lattice Relaxation', *Polymer*, **33**, 2231 (1992).
103. A.S. Kulik, K.O. Prins, '^{2}H-NMR Study of High Pressure Effects on the Molecular Dynamics of Polystyrene: 2. Phenyl Group Motion', *Polymer*, **34**, 4635–4641 (1993).
104. A.S. Kulik, K.O. Prins, '^{2}H-NMR Study of High Pressure Effects on the Molecular Dynamics of Polystyrene: 1. Main Chain Motion', *Polymer*, **34**, 4629–4634 (1993).
105. P.B. Smith, R.A. Bubeck, S.E. Bales, '^{2}H-NMR Analysis of the Molecular Origin of the Thermal Relaxations in Polycarbonates', *Macromolecules*, **21**, 2058–2063 (1988).
106. S. Kaufmann, S. Wefing, D. Schaefer, H.W. Speiss, 'Two-dimensional Exchange Nuclear Magnetic Resonance of Powder Samples. III. Transition to Motional Averaging and Application to the Glass Transition', *J. Chem. Phys.*, **93**, 197–214 (1990).
107. U. Pschorn, E. Rossler, H. Sillescu, S. Kaufmann, D. Schaefer, H. Speiss, 'Local and Cooperative Motions at the Glass Transition of Polystyrene: Information from One- and Two-dimensional NMR as Compared with Other Techniques', *Macromolecules*, **24**, 398–402 (1991).
108. M.L. Williams, R.F. Landel, J.D. Ferry, 'Temperature Dependence of Relaxation in Amorphous Polymers and Glass Forming Liquids', *J. Am. Chem. Soc.*, **77**, 3701–3707 (1955).
109. D. Schaefer, H.W. Spiess, 'Two-dimensional Exchange NMR of Powder Samples. IV. Distribution of Correlation Times and Line Shapes in the Intermediate Dynamic Range', *J. Chem. Phys.*, **97**, 7944–7954 (1992).
110. K. Schmidt-Rohr, A.S. Kulik, H.W. Beckham, A.O. Ohlemacher, U. Pawelzik, C. Boeffel, H.W. Spiess, 'Molecular Nature of the β Relaxation in Poly(methyl methacrylate) Investigated by Multidimensional NMR', *Macromolecules*, **27**, 4733–4745 (1994).
111. A.S. Kulik, H.W. Beckham, K. Schmidt-Rohr, D. Radloff, U. Pawelzik, C. Boeffel, H.W. Spiess, 'Coupling of

the α and β Processes in Poly(ethyl methacrylate) Investigated by Multidimensional NMR', *Macromolecules*, **27**, 4746–4754 (1994).

112. A. Kintinar, L. Jelinski, I. Gancarz, J.T. Koberstein, 'Complex Molecular Motions in Bulk Hard-segment Polyurethanes: A Deuterium NMR Study', *Macromolecules*, **19**, 1876–1884 (1986).
113. L.W. Jelinski, J.J. Dumais, A.K. Engel, 'Solid-state ^{2}H-NMR Studies of Molecular Motion. Poly(butylene terephthalate) and Poly(butylene terephthalate)-containing Segmented Copolymers', in *NMR and Macromolecules: Sequences, Dynamics and Domain Structure*, ed. J. Randall, American Chemical Society, Washington, DC, Vol. 247, 55–65, 1984.
114. L.W. Jelinski, J.J. Dumais, A.K. Engel, 'Multitechnique Solid-state NMR Approach to Assessing Molecular Motion: Poly(butylene terephthalate) and Poly(butylene terephthalate)-containing Segmented Copolymers', *Macromolecules*, **16**, 403–409 (1983).
115. K. Zemke, B.F. Chmelka, K. Schmidt-Rohr, H. Spiess, 'Conformational Exchange Near the Glass Transition: 2D ^{13}C-NMR Study of Atactic Polypropylene', *Macromolecules*, **24**, 6874–6876 (1991).
116. I. Schnell, A. Lupulescu, S. Hafner, D.E. Demco, H.W. Spiess, 'Resolution Enhancement in Multiple-quantum MAS NMR Spectroscopy', *J. Magn. Reson.*, **133**, 61–69 (1998).
117. P.K. Isbester, T.A. Kestner, E.J. Munson, 'High-resolution Variable-temperature MAS ^{19}F NMR Spectroscopy of Fluorocarbon Polymers', *Macromolecules*, **30**, 2800–2801 (1997).
118. A. Lesage, S. Steuernagel, L. Emsley, 'Carbon-13 Spectral Editing in Solid-state NMR Using Heteronuclear Scalar Couplings', *J. Am. Chem. Soc.*, **120**, 7095–7100 (1998).
119. A. Bielecki, A.C. Kolbert, H.J.M. De Groot, R.G. Griffin, M.H. Levitt, 'Frequency-switched Lee–Goldberg Sequences in Solids', *Adv. Magn. Reson.*, **14**, 111–124 (1990).
120. G. Eidmann, R. Savelsberg, P. Blumler, B. Blumich, 'The NMR Mouse, a Mobile Universal Surface Explorer', *J. Magn. Reson.*, **A122**, 104–109 (1996).
121. W.E. Maas, F.H. Laukien, D.G. Cory, 'Gradient, High-resolution, Magic-angle Spinning NMR', *J. Am. Chem. Soc.*, **118**, 13085–13086 (1996).
122. K.T. Mueller, T.P. Jarvie, D.J. Aurenta, B.W. Roberts, 'The REDOR Transform: Direct Calculation of Internuclear Couplings from Dipolar-dephasing NMR Data', *Chem. Phys. Lett.*, **242**, 535–542 (1995).
123. S.A. Taylor, J.L. White, N.C. Elbaum, R.C. Crosby, G.C. Campbell, J.F. Haw, G.R. Hatfield, 'Morphology, ^{31}P Spin diffusion, and Phase Transitions in a Representative Semicrystalline Polyphosphazene by Solid-state NMR', *Macromolecules*, **25**, 3369–3376 (1992).
124. J. Hirschinger, D. Schaefer, H. Spiess, A. Lovinger, 'Chain Dynamics in the Crystalline Alpha-phase of Poly(vinylidine fluoride) by 2D Exchange ^{2}H-NMR', *Macromolecules*, **24**, 2428–2433 (1991).

Positron Annihilation Spectroscopy of Polymers and Rubbers

Josef Bartoš
Polymer Institute of the Slovak Academy of Sciences, Bratislava, Slovak Republic

1 Introduction 1
2 Historical Overview 1
3 Principles of Positron Annihilation Spectroscopy 2
3.1 Positron and Positronium 2
3.2 Positronium Formation Models 3
4 Positron Annihilation Lifetime Spectroscopy Method 4
4.1 Positron Source 4
4.2 Sample Preparation 4
4.3 Data Acquisition and Evaluation 4
4.4 Data Interpretation 5
4.5 Limitations 6
5 Typical Applications of Positron Annihilation Lifetime Spectroscopy Method 6
5.1 Finite-term Lifetime Analysis 6
5.2 Continuous Lifetime Analysis 13
6 Empirical Correlations 15
Acknowledgments 15
Abbreviations and Acronyms 15
Related Articles 16
References 16

Positron annihilation spectroscopy (PAS) is a common name for a set of techniques based on measurements and analysis of the annihilation characteristics of a positron e^+ *and its bound form with an electron* e^-*, so-called positronium* $(e^+e^-) \equiv Ps$*. PAS includes three basic techniques:*

1. *Positron annihilation lifetime spectroscopy (PALS)*
2. *Angular correlation of annihilation radiation (ACAR)*
3. *Doppler broadening spectroscopy (DBS).*

These techniques measure the time of emission, energies and momenta of positron–electron annihilation photons, respectively. PALS measures positron and positronium lifetimes and their relative intensities, giving information about the structural–physical properties of condensed materials such as metals, ceramics and polymers, while ACAR and DBS provide information about the electron momentum and energy distribution in a material. The first technique is the most developed and it serves to characterize the microstructural features of the disordered phase of condensed matter, such as defects or free-volume holes.

1 INTRODUCTION

The main goal of chemistry, physics and material science of condensed matter, especially of polymers (rubbers, thermoplastics and thermosets), is to reveal the basic structure–property relationships. For this reason, the analytical characterization of the physical structure is of fundamental importance. While analysis of the physical structure of ordered systems is the subject of classical diffraction techniques, in the case of disordered, that is amorphous polymers and nonperfect ordered (i.e. semi-crystalline), ones, PAS methods are of great importance. PAS methods are able to provide information about the physical microstructure of polymers in terms of free-volume quantities such as the mean free-volume hole size, the free-volume hole size distribution, the anisotropy of free-volume holes and the free-volume hole fraction as a function of internal variables such as chemical composition, molecular weight and crystallinity, and of external variables such as temperature, pressure and mechanical stress. In spite of this progress, however, PAS methods are continuously being developed and are close to becoming a routine analytical tool for polymer characterization. Of the above-mentioned PAS techniques, the PALS method is the most developed for microstructural characterization of disordered materials. Dimensions and timescales pertaining to existing experimental techniques for the characterization of free-volume holes, defects and voids are presented in Figure 1.

2 HISTORICAL OVERVIEW

The theoretical prediction of the positron by Dirac in 1930 and its experimental discovery by Anderson in 1932 and of positronium (Ps) by Deutsch in 1951, as well as the first application to polymers by Benedetti and Richings in 1952, led to the relatively rapid development of a new spectroscopic method for the study of the structure of matter. The initial stage in the study of the structure and properties of positron and positronium in the 1950–1960 period was followed in the 1960s by several application studies on condensed matter by physicists and in the 1970s by similar studies, this

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

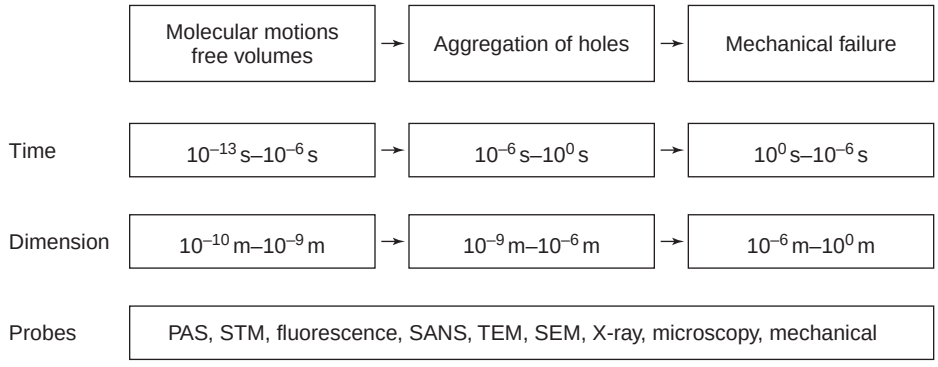

Figure 1 Characteristic dimension and timescales pertaining to the various experimental techniques for the characterization of the open space in condensed matter, such as free-volume, defects and voids. (Reproduced from Jean,[9] by permission of Societa Italiana di Fisica.) SANS, small-angle neutron scattering; SEM, scanning electron microscopy; STM, scanning transmission microscopy; TEM, transmission electron microscopy.

time by physical chemists. Earlier analyses of positron annihilation spectra in terms of two components were replaced, after discovery of the third component by Tao and Green in 1965, by a more precise method in terms of standard three-term analysis. Several reviews can be mentioned.[1–3] In the case of polymers, the earliest studies were in the 1950s, continuing into the 1970s, although the information they provided was rather limited often because of the use of ill-defined polymer samples.[4,5] In the 1980s a renaissance in PAS applications in the field of polymers was connected with the accessibility of characterized polymer materials, as well as with methodological progress which consisted of raw data evaluation and subsequent interpretation allowing free-volume data quantification and its physical explanation; see various reviews.[6–12]

3 PRINCIPLES OF POSITRON ANNIHILATION SPECTROSCOPY

3.1 Positron and Positronium

The positron, e^+, is an antiparticle of the electron, e^-, with the same mass $m = 9.11 \times 10^{-31}$ kg, but with the opposite charge $e = +1.602 \times 10^{-19}$ C and with the spin $s = \frac{1}{2}$. If the positron encounters the electron, according to the Einstein formula, $E = mc^2$, the mass of the particles is converted into radiation (photons) of specific energy 0.511 MeV. Positrons are generated by the radioactive decay of certain unstable isotopes, one of the most utilized being $^{22}_{11}$Na, which decays to $^{22}_{10}$Ne with the simultaneous emission of e^+ and a γ photon of 1.28 MeV. The positrons emitted are characterized by an energy distribution between zero and 0.540 MeV, with a maximum at around 0.200 MeV. When a positron is injected into a substance it loses its kinetic energy, that is it is thermalized by a sequence of inelastic collisions with the particles of the substance. The thermalization stage for positrons occurs very rapidly, within 10 ps,[13] and is connected to the creation of excess electron-positive ion pairs along the track. The last part of this process takes place at a short distance from the thermalized positron, that is, the positron spur. Before the injected positron is annihilated by interaction with one of the electrons, it can take various physical and chemical states. Whereas, in conducting materials such as metals, positrons diffuse into the interstitial sites of perfect crystals or are trapped in vacancy-type defects in a real nonperfect crystal, in nonconducting insulating materials, such as polymers, positrons can exist in a free nonbounded state or as a bound positron–excess electron pair e^+e^-, that is positronium (Ps) with a binding energy 6.8 eV. Positronium is the lightest atom and is similar to the hydrogen atom in having the same size ($r_{Ps} = 0.53$ Å). Two ground states of Ps exist, depending on the mutual spin orientation of the positron–electron pair: singlet *para*-positronium (*p*-Ps) with antiparallel spins where the total spin of the system $S = 0$, and triplet *ortho*-positronium (*o*-Ps) with parallel spins and where $S = 1$. The corresponding formation probabilities are 25% *p*-Ps and 75% *o*-Ps, respectively.

Both the positron and positronium annihilate by emitting γ rays. The selection rule for γ photon emission is derived from the conservation laws of energy, momentum and parity considerations. Typically, two annihilation modes consist of (1) singlet ^1S annihilation for e^+ and e^-, with antiparallel spins producing two γ photons with energies of 0.511 MeV in opposite directions 180° to each other, and (2) a triplet ^3S annihilation process for parallel spin orientation, giving three γ photons with a continuous energy spectrum ranging from 0 to 0.511 MeV. In the case of Ps in a vacuum, the intrinsic annihilations of *p*-Ps and *o*-Ps are singlet or triplet annihilation

respectively, and the corresponding intrinsic lifetime of *p*-Ps is $\tau_{p-\mathrm{Ps}} = 0.125$ ns and that of *o*-Ps $\tau_{o-\mathrm{Ps}} = 142$ ns. In the condensed phase, the *p*-Ps lifetime is hardly changed by the matrix, but *o*-Ps is repelled into the regions of low-electron density, the so-called free-volume holes, where it exists for some time. Here *o*-Ps can interact with one electron of the surrounding medium which has antiparallel spin with respect to the spin of the positron in the *ortho*-positronium system. Consequently, a spin exchange or pick-off annihilation takes place, which leads to a reduction in the *o*-Ps lifetime to 0.5–5 ns by emitting two γ photons. A ''free'' nonbounded positron has a lifetime of about 0.3–0.4 ns and decays by interaction with the valence electrons of the media. Thus, a typical annihilation picture of an insulating condensed material in which *o*-Ps is formed will contain at least three components: a short-lived component from *p*-Ps with $\tau_1 \cong 0.1$–0.2 ns, an intermediate component from ''free'' positron annihilation with $\tau_2 \cong 0.3$–0.4 ns, and a long-lived component from *o*-Ps with $\tau_3 \cong 0.5$–5 ns.

3.2 Positronium Formation Models

In general, any application of PALS to determine the physical properties of polymers requires a basic understanding of the formation of Ps in polymeric materials. In positron annihilation research, three basic models of Ps formation have been formulated.

3.2.1 Ore Energy Gap Model of Ps Formation

In this model it is assumed that Ps is formed during the stage when the positron slows down and passes through a certain energy range E called the Ore gap, where e^+ can effectively pick up an electron from the molecule M to form Ps. Ps formation is represented by the energy condition: $E_i - E_{Ps} < E < E_{exc}$, where E_i is the first ionization potential of the molecule, E_{Ps} is the binding energy of Ps (6.8 eV in vacuum) and E_{exc} is the lowest electronic excitation energy of the surrounding molecules.[14] The modified version of this model[15] assumes that Ps is formed by the Ore gap process, but that Ps can undergo a set of reactions via ''hot'' Ps before it thermalizes, Equations (1–3)

$$e^{+*} + M \longrightarrow Ps^* + M^+ \quad (1)$$

$$Ps^* + M^+ \longrightarrow e^+ + M \quad (2)$$

$$e^+ + e^- \longrightarrow Ps \quad (3)$$

In polymers, the energy gap is modified by the dielectric constant, which narrows this gap, and the zero-point energy caused by Ps localization in holes, which reduces the Ps binding energy. Thus, Ps is formed in the bulk prior to localization in the hole, and Ps formation depends on the physicochemical properties of matrixes.

3.2.2 Free-volume Model of Ps Formation

In the original free-volume model,[16] Ps was assumed to be localized in an interstitial position between the lattice points which model the rigid solid matrix. In principle, this simplest version was able to explain the basic features of the temperature dependence of *o*-Ps annihilation behavior. An extended ''soft'' version of this model[17] included the effect of lattice vibrations on the annihilation process and was used to explain some temperature anomalies. These initial free-volume models have formed useful guidelines for the present understanding of Ps formation and annihilation in terms of the free-volume hole model.

According to a modern free-volume hole version of the model,[18–23] Ps is formed in an open space, a defect in the structure such as vacancy, hole or free-volume, when a positron captures an electron from the surface of the defect and subsequently Ps is trapped in this microstructural defect, Equation (4):

$$e^+ + e^-(\text{defect surface}) \longrightarrow Ps\,(\text{thermalized}) \quad (4)$$
$$\downarrow\rightarrow Ps^* + \text{defect surface} \rightarrow\uparrow$$

The probability of the positron capturing electrons from the inner surface may be energy dependent. Evidently, Ps formation will be more favorable for a positron with a lower energy than one with a higher energy. The epithermal Ps formed will continue to lose energy by collisions with the inner surface of the defects until it reaches thermalization. According to this model, the Ps formation probability is a measure of the fraction of free-volume holes, if the probability of escaping from the trapped defect to another one during its lifetime is negligible. This has been shown to be the case according to Ps diffusion data[24,25] and quantum-mechanical calculations.[26] Quantitative details will be given in section 4.4.

3.2.3 Spur Reaction Model of Ps Formation

In the spur reaction model[27] it is assumed that Ps is formed by a combination reaction of the thermalized positron and one of the excess electrons released from the molecule by the positron itself in the terminal positron spur. In this case, Ps formation must compete with other processes such as electron-ion combination, electron and positron escape from the spur, positron and electron scavenging, and so on. The Ps formation process can be expressed by the following scheme, Equation (5):

$$e^+(\text{thermalized}) + e^-(\text{spur}) \longrightarrow Ps\,(\text{thermalized}) \quad (5)$$

In a medium of static dielectric permittivity ε with thermal energy kT, a stable ion-pair is formed when the critical distance between the charged particles is $r_{cr} = e^+e^-/2\pi\varepsilon_0\varepsilon kT$, where ε_0 is the dielectric permittivity of free space. Consequently, the Ps formation probability is directly related to the spur size at the end of the positron spur.

3.2.4 Positron and Positronium Chemistry

Positron and positronium, the latter being the lightest atomic radical similar to hydrogen, can undergo various chemical interactions with specific atoms or atomic groups with large electron affinities, such as halogens, carbonyls and radicals (such as the oxygen molecule O_2) or ions, and change positron states. For example, a positron can react with an ion: $e^+ + M^- \longrightarrow [e^+M^-]$, *ortho*-positronium can undergo spin conversion: *o*-Ps + $M^\bullet \longrightarrow$ *p*-Ps + $M^\bullet$, complex formation: Ps + M $\longrightarrow$ PsM or oxidation reaction: Ps + M $\longrightarrow e^+ + M^-$. Generally, chemical reactions are transitions between different positron and positronium states, and they cause a change in the annihilation spectra (most often shortening the lifetime or the presence of additional components owing to the different properties of the positron states). In the author's opinion, the present relative autonomy of physics- and chemistry-oriented positron annihilation research will be overcome and both branches of positron annihilation research will converge, leading to a consistent and unified approach.

3.2.5 Comparison of Current Ps Formation Models

Comparing the above-mentioned models of Ps formation, we can conclude that a key difference between them consists in the positron energy at the moment of Ps formation. In the free-volume model the positron can be at both the epithermal and thermalized stage, while the Ore gap hot Ps reaction model is at the epithermal stage, and finally, the positron is mainly in the thermalized stage for the spur reaction mechanism. As for the ability to explain the maximum amount of observed experimental features, the free-volume model is able to address the problems not only of the lifetime but also of the Ps formation probability, via the relative intensity of particular positron states. On the other hand, the Ore gap model and the spur reaction concern Ps formation only. Therefore the PALS method together with the free-volume interpretation is the most developed research area in the field of PAS of polymers. For this reason, in this review we mainly focus on the principal achievements of this method in interpreting results in terms of the free-volume hole model, keeping in mind that the *o*-Ps lifetime τ_3 is determined by the free-volume hole size, while the relative intensity I_3 is dependent on both the positronium formation probability and the density of the free-volume hole sites.

4 POSITRON ANNIHILATION LIFETIME SPECTROSCOPY METHOD

4.1 Positron Source

All positron annihilation studies of polymers have been made with positron-emitting radionuclides, usually ^{22}Na in the form of ^{22}NaCl. ^{22}Na isotope emits positrons with kinetic energy up to 0.540 MeV. A typical source consists of a small spot of ^{22}NaCl prepared by evaporating the water from a droplet of ^{22}NaCl solution and sealed between two thin (usually 8 μm) special polyimide Kapton® foils. Typically, the positron strength is from a few microcurie (μCi) up to 50 μCi. This radiation level is relatively low and no special precautions are needed for protection.

4.2 Sample Preparation

The polymer to be investigated can be in any physical state, that is amorphous or semicrystalline material, the former in the glassy or rubbery state, and in any technical form: bulk, film, or powder. The sample includes two pieces of material, one placed on each side of the positron source. The most important requirement of the sample is that all positrons emitted from the source must stop and annihilate in the sample. In practice, the typical geometry of the sample is a square of 10×10 mm or a cylinder of $\Phi = 10$ mm. Since the emitted positrons have a spectrum of energies, they penetrate to various depths in the sample. In polymers these positrons can penetrate to a depth on the order of 1 mm before reaching thermal energy. For this reason, the typical thickness is 2 mm.

4.3 Data Acquisition and Evaluation

The PALS method measures the time a positron spends in the sample between injection and annihilation. The ^{22}Na isotope emits almost simultaneously (within a few picoseconds) γ photons of energy 1.28 MeV and a positron. PALS measurement is based on the detection of this γ photon, creating the start (birth) signal in one detector and, on the detection of a γ annihilation photon (0.51 MeV), giving the stop (death) signal detected by the other detector. The time difference between the two detector signals, that is the lifetime of a positron, is measured by the time-to-amplitude converter (TAC), and this information is stored in a multi-channel analyzer (MCA). Typically, as annihilation is a statistical process, 10^6–10^7 annihilation events are registered and the output

is obtained in the form of a histogram of distribution of positron annihilation lifetimes, that is the lifetime spectrum. If the positrons all annihilate in a homogeneous environment, each positron has the same annihilation probability and the lifetime spectrum consists of only one exponential. Thus, if N_0 positrons were formed at the same time $t = 0$, the number of positrons remaining at time t will be $N_0 \exp(-t/\tau)$, where τ is the mean lifetime. In the general case, when the material is heterogeneous, different positrons can have different environments and create different positron states (e^+, Ps) after they have slowed down. Consequently, if positrons and positronium are present in a sample, there are several distinct populations of positron states and the positron lifetime spectrum is a sum of several exponentially decaying components, each with its own characteristic lifetime value. In principle, two phenomenological approaches to the annihilation data analysis can be used: finite-term lifetime analysis and continuous-lifetime analysis.

In the first evaluation method, an experimental datum $Y(t)$ is expressed as a convolution (symbol *) of the spectrometer resolution function $R(t)$ and of a finite number n, usually three or more, of negative exponentials with the particular mean characteristic annihilation rate $\lambda_i = 1/\tau_i$, Equation (6)

$$Y(t) = R(t) * \quad N_t \sum_{i=1,n} \alpha_i \lambda_i \exp(-\lambda_i t) + B \qquad (6)$$

where $N(t)$ is the normalized number of counts, $I_i = \alpha_i \lambda_i$ is the relative intensity of the ith spectral component and B is the background. In practice, by fitting the model function to a measured lifetime spectrum, the values of τ_i and I_i can be extracted by using PATFIT program software.[28]

In the second evaluation method it is assumed that each component has a certain distribution of the corresponding characteristic lifetimes, so that a lifetime spectrum reads, Equation (7)

$$Y(t) = R(t) * \quad N_t \int_0^\infty \lambda \alpha(\lambda) \exp(-\lambda t)\, d\lambda + B \qquad (7)$$

in which the annihilation decay integral function is a Laplace transformation of the decay probability density function pdf $[\lambda\alpha(\lambda)]$.[29] In practice, the CONTIN[30,31] or MELT[32,33] computer programs for the numerical solution of integral equations are used.

4.4 Data Interpretation

According to quantum mechanics, the observed pick-off annihilation rate λ of positronium is given by the integral of positron $\rho^+(r)$ component of the Ps wavefunction and the density $\rho^-(r)$ of the electrons bound to the molecule at the annihilation position r in the matrix, Equation (8):

$$\lambda = \frac{1}{\tau} = \text{const} \quad \rho^+(r)\rho^-(r)\, d^3r \qquad (8)$$

At present, the solution of both the densities is an intractable problem, so that an approximated approach is used, based on a combination of a theoretical expression for a simple quantum-mechanical model of Ps confined in a hole and the existing experimental data. In this simple model, Ps is located in a spherical well with radius R_0 having an infinite potential barrier of a homogeneous electron layer with the thickness $\Delta R = R_0 - R$ inside the wall. Assuming that the annihilation rate of o-Ps inside the electron layer is $2\,ns^{-1}$ (the spin averaged annihilation rate of o-Ps and p-Ps) being close to the annihilation rate of positronium anion Ps^-,[34] the final result for the annihilation rate λ_3 of o-Ps as a function of a free-volume hole radius R_h is,[18–20] Equation (9)

$$\lambda_3 = \frac{1}{\tau_3} = 2 \quad 1 - \frac{R_h}{R_0} + \left(\frac{1}{2\pi} \quad \sin 2\pi \left(\frac{R_h}{R_0} \right. \right. \qquad (9)$$

where ΔR is the empirical parameter determined by fitting the observed lifetimes in molecular compounds, such as zeolites and molecular crystals, with the known hole and cavity sizes. The best fitted value of $\Delta R = 1.656$ Å[20] (see Figure 2). A modified free-volume model of an elliptical shape of the holes has been formulated.[35]

According to the free-volume hole model, the relative intensity of o-Ps, I_3, is proportional to the number of holes N_h in the matrix, so that the phenomenological equation for the free-volume fraction is given

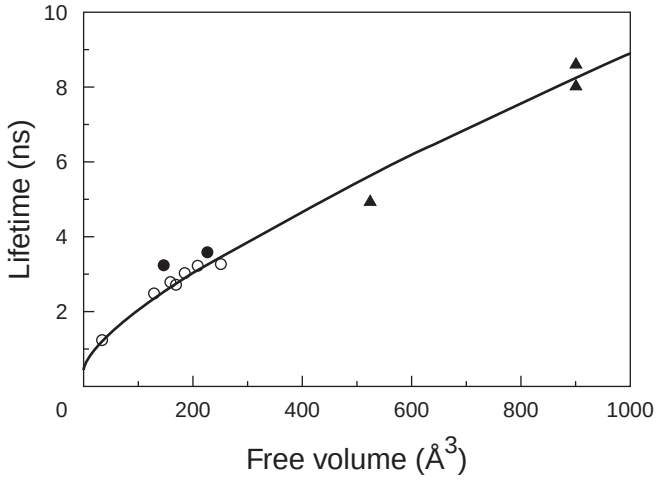

Figure 2 Correlation curve between the observed o-Ps lifetimes τ_3 and the known hole sizes in molecular solids and zeolites. ○, ● molecular solids; , zeolites. (Reproduced from Jean,[9] by permission of Societa Italiana di Fisica.)

by,[21,22] Equation (10)

$$f_h(T) = K_{Vh} V_h(T) I_3(T) \quad (10)$$

where K_{Vh} is the proportionality coefficient which can be determined by various calibration procedures using other physical properties such as the macroscopic volume expansion coefficient[21,36] or by combining the macroscopic volume expansion coefficient in the liquid state above T_g with the microscopic free-volume hole fraction expansion coefficient above T_g.[37] Comparing the approaches in Refs 21 and 36, the first two are based on a direct adjustment procedure by identifying f_h from Equation (10) with the difference between the macroscopic volume expansion coefficients above and below the glass–liquid temperature T_g, while the latter makes the same adjustment via the initial temperature situated below T_g at which the dynamic free-volume holes in the liquid phase set in.

If one-to-one correspondence of the *o*-Ps lifetime and the free-volume hole radius is assumed to be valid, the lifetime distribution can be transformed into the free-volume hole radius distribution,[31] Equation (11)

$$f(R_h) = -2\Delta R_h \left\{ \cos \left(2\pi \left(\frac{R_h}{R_{0,h}} - 1 \right) \right) \right\} \frac{\alpha(\lambda_3)}{R_{0,h}^2} \quad (11)$$

The fraction of *o*-Ps annihilation in the holes with radii between R_h and $R_h + dR_h$ is $f(R_h)\,dR_h$. This assumption about one-to-one correspondence is supported by the experimental results concerning *o*-Ps diffusion in several polymers[24,25] as well as the theoretical results from quantum mechanical calculations.[26] Finally, $f(R_h)$ can be converted into a free-volume distribution function, Equation (12)

$$g(V_h) = \frac{f(R_h)}{4\pi R_h^2} \quad (12)$$

The volume fraction of free-volume holes as determined by *o*-Ps annihilation in holes with volume between V_h and $V_h + dV_h$ is $g(V_h)\,dV_h$.

4.5 Limitations

In connection with the free-volume characterization of polymers by means of the PALS method it is important to note that this technique has a dynamic character because of the existence of both intrinsic dimension and time limits. In fact, there is a lower space limit given by the smallest free-volume hole sizes detectable by *o*-Ps caused by its finite size (*o*-Ps radius is 0.53 Å) as well as a time limit connected with the timescale of molecular motions with characteristic times $\tau < 10^{-9}$ s. In addition to these basic physical aspects, chemical aspects linked to the presence or absence of the electron affine groups, which can influence the lifetime or/and relative intensity, might play the role. For these reasons, the PALS method is able to measure the *effective* free-volume microstructure determined by the mutual interplay of the above-mentioned physical and chemical factors, not the total free-volume as represented by the so-called empty free-volume $V_f^{empty}(T) = V(T) - V^w$, where $V(T)$ is the macroscopic volume and V^w is the van der Waals volume of the constituents.

5 TYPICAL APPLICATIONS OF POSITRON ANNIHILATION LIFETIME SPECTROSCOPY METHOD

In this section we present in a systematic and illustrative way the most typical applications of the PALS method in analysis of polymers. In general, there are the two basic evaluation procedures (see section 4.3) and, simultaneously, a number of internal and external variables influencing the annihilation response of polymer, such a chemical composition and molecular weight or temperature, pressure, stress, and time, respectively. In order to obtain a concise idea of the possibilities of the PALS method, we have chosen a presentation approach based on classification according to the evaluation method used and, further on, the basic external parameters, allowing for variation of the internal parameters. Thus, we demonstrate the most important effects of the *o*-Ps annihilation characteristics in rubbers and plastics, which are dependent on the basic external variables together with the revealed empirical correlations as well as important microstructural information related to various physicochemical phenomena such as transitions, relaxations and penetration phenomena.

5.1 Finite-term Lifetime Analysis

5.1.1 Thermal Effects

5.1.1.1 Amorphous Polymers The most elementary applications of the PALS method consist in investigations of the *o*-Ps annihilation data dependent on temperature T. The first results on polymers were published in Bell and Graham in 1953.

Typical courses of the *o*-Ps data, that is the *o*-Ps lifetime, τ_3, and the relative intensity, I_3, over a wide temperature range as a function of heating or cooling are shown for the illustrative case of the amorphous polymer, polyisobutylene (PIB) in Figures 3 and 4.

The first annihilation quantity, *o*-Ps lifetime, τ_3, in Figure 3 follows a generally sigmoidal course exhibiting several effects. The first pronounced effect on the relatively low temperature side is an abrupt change

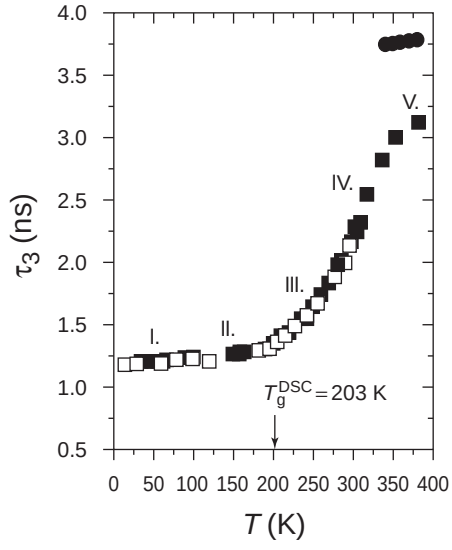

Figure 3 The *o*-Ps lifetime τ_3 as a function of temperature in PIB in a heating (■) and cooling ( ) measurement regime, with the bubble model (●). (Reproduced from Krištiak et al.[38] by permission of Trans Tech Publications Ltd.)

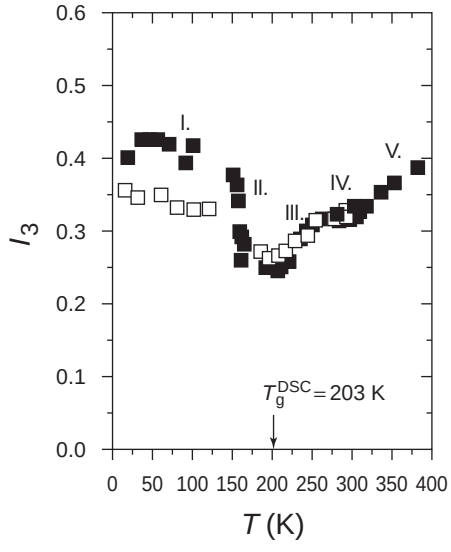

Figure 4 The *o*-Ps relative intensity I_3 as a function of temperature in PIB in a heating (■) and cooling ( ) measurement regime. (Reproduced from Krištiak et al.[38] by permission of Trans Tech Publications Ltd.) (I_3 is a dimensionless quantity.)

of the slope at the first crossing temperature T_g^{PALS}, which lies in the vicinity of the glass–liquid temperature, T_g, from classical macroscopic thermodynamic measurements such as dilatometry T_g^{DIL} or calorimetry T_g^{DSC}. Analogous courses in the vicinity of T_g have been found for a further series of rubbers of the diene type,[38] such as 1,4-polybutadiene, (1,4-PBD) 1,4-polyisoprene (1,4-PIP) as well as for amorphous thermoplastics of very different chemical structure,[39] such as a series of polyalkylmethacrylates,[40] of polycarbonates (PCs)[41] and of epoxy thermosets.[42] For a series of diene, vinylidene and vinyl-homoatomic, backbone polymers, as well as heteroatomic backbone amorphous polymers, an empirical correlation between the τ_3 value at T_g^{DSC}, $\tau_{3,g}^{DSC}$, and the corresponding T_g^{DSC} values, the latter ranging over 300 K, has been found[43] (see Figure 5). The relationship between the mean free-volume hole size at T_g, $V_{h,g}$ and the van der Waals volume of the basic structural unit V_{mon}^{w} is useful in discussing the motional mechanism responsible for the glass–liquid transition.[43]

The second annihilation characteristic, the *o*-Ps relative intensity, I_3, exhibits a generally more complicated dependence on temperature, which is linked to both the above-mentioned dynamic and chemical aspects of the PALS method (see Figure 4). A course similar to that for PIB shows not only further elastomers such as 1,4-PIP,[44] but also thermoplastics such as PVAc,[21] PS,[45] PMMA[46] and thermosets.[42] For the amorphous elastomers cited (1,4-PIP, PIB) investigated in detail so far, the general feature is that I_3 increases with increasing temperature above the particular glass–liquid

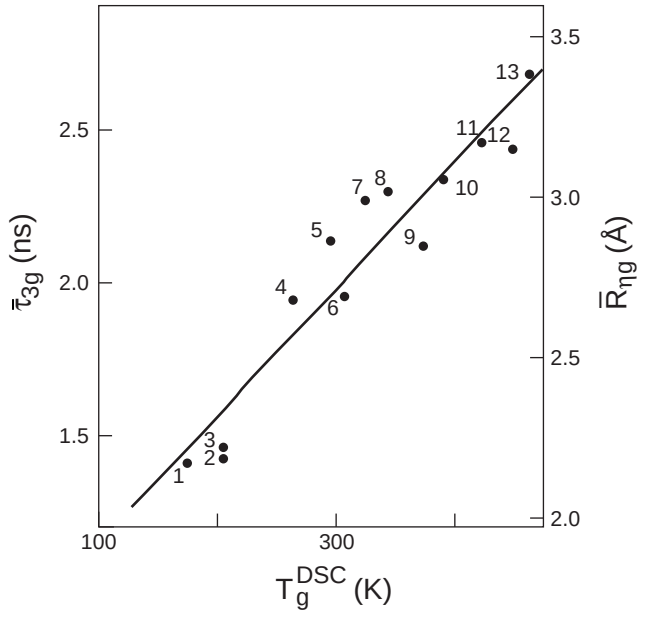

Figure 5 The mean *o*-Ps lifetime at T_g^{DSC} vs appropriate T_g^{DSC} values for a series of 13 amorphous polymers. (1) *cis*-1,4-PBD, (2) PIB, (3) *cis*-1,4-PIP, (4) atactic-polypropylene, (5) poly n-butylmethacrylate, (6) polyvinylacetate (PVAc), (7) polyisobutylmethacrylate, (8) polyethylmethacrylate, (9) polystyrene (PS), (10) polymethylmethacrylate (PMMA), (11) bisphenol A polycarbonate, (12) bisphenol Z polycarbonate, (13) tetramethylbisphenol A polycarbon. (Reproduced from Bartoš[43] by permission of Steinkopff Verlag, Darmstadt.)

transition temperature, indicating the dominant role of increasing the number of free-volume holes in the rubbery state of amorphous polymers. In our illustrative case of PIB, an acceptable correlation between the onset temperature of the decrease in the relative intensity I_3 with the onset temperature of the so-called fast motion from the temperature dependence of the mean square displacement $\langle u^2 \rangle$ of atomic motion, as measured by using the neutron scattering (NS) method,[47] has been found. This suggests that releasing this type of microscopic mobility reduces, at least partly, the total number of free-volume hole sites at the *o*-Ps localization. In some other amorphous plastics such as PVAc,[21] PS,[45] PMMA[46] and PC,[41] more complicated courses for I_3 versus T plots have been found, sometimes with several effects, such as minima below T_g or plateau regions in the vicinity of T_g. Generally speaking, these effects might reflect the simultaneous action of physical effects such as motional degrees of freedom, and in some specific cases also chemical degrees of freedom, such as various reactions (section 3.2). For these reasons, actual interpretations of the particular polymers are currently the subject of continued intense research.

In the previous section, we mentioned the most fundamental feature of annihilation behavior of τ_3 as a function of temperature which is connected with the glass–liquid transition being dependent on the most important internal variable, that is the chemical composition of polymers. Most polymers cited were characterized by a sufficiently high molecular weight M above the so-called entanglement molecular weight M_e. In general, many physical properties of polymers depend on their molecular weight, so that some attention has been focused on revealing the influence of the second most important internal variable, molecular weight on annihilation characteristics. The most studied system has been monodisperse PS, where several research groups[48–50] found the following regularities: (1) τ_3 versus T plots exhibit a slope change at the crossing temperature $T_g^{PALS}(M)$ which grows with molecular weight M in a similar way to the calorimetric T_g^{DSC} values; (2) the lifetime at the corresponding T_g^{PALS}, $\tau_{3,g}$, is independent of M; (3) $\tau_3(T)$ below T_g^{PALS} is essentially independent of M, while $\tau_3(T)$ differs substantially, being higher for lower M at constant T above $T_g^{PALS}(M)$ as a consequence of the increase in $T_g^{PALS}(M)$; (4) at room temperature $\tau_3(300\,K)$ sharply decreases with M and is independent of M within the range $M > 10^3$, that is above the entanglement molecular weight; and finally (5) relative intensities $I_3(T)$ are higher for lower M over 20–120 °C, but all curves saturate above around 55 °C.

Phenomenologically, the low-temperature part of the sigmoidal τ_3 versus T curve around the glass–liquid transition is similar to the macroscopic volume V versus temperature T dependence from dilatometry.[51,52] However, the corresponding expansion coefficients for the macroscopic volume $\alpha_V = (1/V_g)(\Delta V/\Delta T)$ and that for the mean free-volume from Equation (9), $\alpha_{Vh} = (1/V_{h,g})(\Delta V_h/\Delta T)$, differ by one and even two orders of magnitude in the glassy $T < T_g$ or rubbery $T > T_g$ state, respectively. This is due to the fact that the former represents the global macroscopic property of both the occupied and free-volume, while the latter is a mean local property of the local regions of lowered density. Thus, the typical α_V values are about $2 \times 10^{-4}\,K^{-1}$ in the glassy state and around $5–10 \times 10^{-4}\,K^{-1}$ in the rubbery state, while α_{Vh} is an order of magnitude of $10^{-3}\,K^{-1}$ below T_g, but even an order of magnitude of $10^{-2}\,K^{-1}$ above T_g.[53] On the other hand, the microscopic free-volume hole expansion coefficient, α_{Vh}, is comparable with that for the mean square displacements $\alpha^2_{\langle u \rangle}$ of the atomic motion as detected by NS measurements, indicating the common origin of both phenomena.[53,54]

On the highest temperature side, the second dramatic change in the slope of the τ_3 versus T plot is evident at the intercept temperature $T_b \cong T_g + 150\,K$ (see Figure 2) and Jean et al.[42] and Bartoš et al.[54] Earlier, this effect was attributed to the occurrence of a specific positronium state in the soft matter – the so-called bubble positronium state.[55] It was believed[42,54] that the situation is similar to that for low-molecular weight liquids.[56] The bubble state is considered to be a result of competition between the quantum zero-point motion of *o*-Ps and the matrix resistance. The bubble size is determined by the balance between outward pressure of the zero-point motion of *o*-Ps and inward pressure from surface tension, γ, and internal pressure, p_{int}, of the medium,[56,57] Equation (13)

$$\frac{(\pi h)^2}{4m_e r_b^2} = 4\pi r_b^2 \gamma + \left(\frac{4\pi}{3} \quad r_b^3 p_{int} \right. \tag{13}$$

where m_e is the mass of the positron and r_b is the bubble radius. From application of this bubble model to our case of PIB using pVT data,[58] we can conclude that the bubble state of *o*-Ps does not appear in PIB, at least.

In addition to the two most dramatic effects in the τ_3 versus T plot mentioned above, further rather weaker effects can sometimes be found both below and above the glass–liquid transition temperature T_g.

In detailed studies of the glassy states of amorphous polymers below T_g, at some characteristic temperature(s), a slight change(s) in the slope of the τ_3 versus T plot takes place, for example in elastomers such as PIB (see Figure 3), in thermoplastics such as polycarbonate (PC),[59] PMMA[46,60] as well as in thermosets such as epoxies.[42] In some polymers, the anomalous effects are more evident via the pronounced decrease in the I_3 versus

T plot.[44–46,60,61] Within the physical framework, these empirical findings can be, from the thermodynamic point of view, related to similar effects in the corresponding $V-T$ dependences from dilatometry[52,59] and/or, from the dynamic point of view, related to the characteristic dynamic features at the microscopic and macroscopic level as detected by NS methods[47] or by the standard mechanical or dielectric relaxation techniques,[62] respectively. The relationship to microscopic dynamics is based on the observation of the temperature correlation between the subglass effect and the onset of the so-called fast motion below the corresponding T_g elastomers.[47,54] At the dynamic macroscopic level, the above-mentioned effects have been found to correlate with the presence of the so-called secondary $\beta, \gamma, \delta \ldots$ relaxations in the glassy state of polymer matrixes.[45,46,60,61]

On the higher temperature side above T_g, detailed investigations of the rubbery state of amorphous polymers have revealed a similar slight change in the slope of τ_3 versus T dependence at about 1.2–1.3 T_g. In this case, approximate agreement with the existence of the so-called liquid–liquid transition temperature T_{ll} from other experimental techniques[63] occurs (see Figure 3).

5.1.1.2 Semicrystalline Polymers According to the general free-volume idea of condensed matter, the free-volume is a measure of structural disorder and thus, it exists only in amorphous polymers or in the disordered regions of semicrystalline ones, that is, in the amorphous and interface regions and in the defect regions of the crystalline phase. In the earlier reports on semicrystalline polymers, some trends between the *o*-Ps formation probability and the degree of crystallinity have been found, but not at the quantitative correlation level, owing to various factors such as different thermal treatments, specific features connected to inhomogeneous primary and secondary crystallization, problems in crystallinity determination, and finally, ambiguity in the assignment of the positron lifetime components. Under strict control of these factors, however, using a series of polyetheretherketone (PEEK) samples as an example it has been found that (i) the *o*-Ps lifetime is independent of the degree of crystallinity x_c, (ii) the *o*-Ps intensity I_3 decreases with increasing crystallinity, and (iii) the normalized free-volume hole fraction correlates linearly with the crystallinity, giving an intercept close to 1 at $x_c = 0$.[64] These findings support the argument that all Ps annihilate at free-volume sites localized in the amorphous phase of the polymers, and indicate that the probability of Ps formation is proportional to the whole sample free-volume population in PEEK, at least. On the other hand, the detailed combined structural study of polyethyleneterephthalate (PET), by means of small-angle X-ray scattering and wide-angle X-ray scattering (SAXS, WAXS) and PALS techniques, showed that the decrease of I_3 with increasing crystallinity cannot be completely described if one assumes that *o*-Ps localizes in the amorphous phase only.[65]

It has been suggested that *o*-Ps can be formed in the crystalline regions, owing to the less homogeneous electron density of the PET crystalline lattice in contrast to the more homogeneous one in PEEK. In some other semicrystalline polymers with a relatively high degree of crystallinity, the annihilation picture appears to be more complicated. First, some researchers use four-component analysis instead of usual three-component data evaluation. Typical effects of both the long-term annihilation components on temperature are demonstrated in Figures 6 and 7 on examples of linear and branched PEs.[66]

The first two shorter lifetimes τ_1 and τ_2 reach common values and have the usual meaning (*p*-Ps and free positron), but the origin of the first of the longer components $\tau_3 \cong 0.8–1$ ns is the subject of continuing debate. It has been proposed that this additional component is connected with the existence of the e^+ or Ps molecular complex species and/or is related to the complicated morphology of the multiphase system, containing not only purely disordered amorphous phase and purely ordered crystalline phase but also the defect crystalline phase and the interface phase between the amorphous and crystalline phases. Two assignments have been proposed, but in the first case the nature of the molecular complex(es) is unknown,[67] while the other is partially supported by some relationships with the morphological features.[68] The fourth component τ_4 is comparable with that for the amorphous polymers, so that it is attributed to *o*-Ps in the free-volume holes of the amorphous regions. In addition to an abrupt change at the corresponding T_g's, the second effect is localized in the vicinity of the melting temperature, T_m, of the ordered phase.[66] Similar results from four-component analysis have been found for isotactic polypropylene (i-PP) of low and high density.[69] On the other hand, some researchers have reported three-component analysis of semicrystalline polymers such as PE, where the second component from a third-term fit is considered to be a mixture of the second and third components from a four-term fit: $\tau_2 = 0.44$ ns versus $\tau_2 = 0.3$ ns and $\tau_3 = 0.8$ ns.[70] In this case, the second component is because the temperature independence over a very wide temperature range exceeding T_m is attributed to the free positron annihilation mode.[70] Like amorphous polymers, complicated effects of the relative intensity of the longest *o*-Ps lifetime component (I_3 or I_4) versus temperature T dependence have been reported.[67–70] In

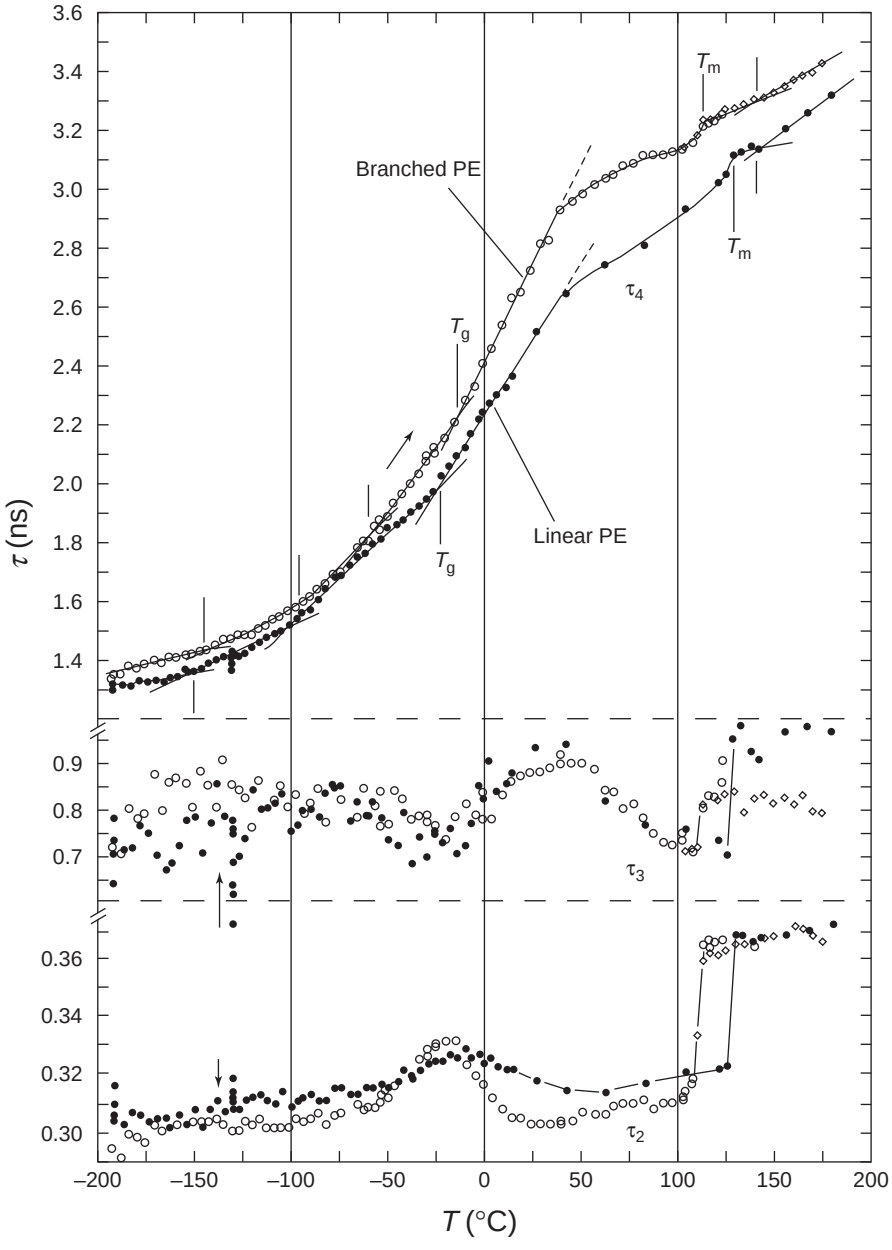

Figure 6 Positron lifetimes from fourth-component fit in linear polyethylene (PE) compared with branched PE as a function of temperature. (Reproduced from Reiter and Kindl[66] by permission of Wiley-VCH Verlag, Berlin.)

some cases, however, quite acceptable interpretations in terms of local secondary β, γ relaxations, and of global primary α relaxation connected with glass–liquid transition, have been proposed.[66,67,70]

5.1.2 Mechanical Effects

The second external variable includes mechanical action on condensed matter in the form of pressure p or of stress σ on the various types of mechanical loading of the sample, such as static uniaxial deformation or cyclic loading. The first study of positron annihilation was performed by Wilson et al. in 1963 on a series of one amorphous polymer PMMA and two semicrystalline polymers: PE and polytetrafluoroethylene (PTFE).[71] It was found that at a fixed temperature the long lifetime component decreases exponentially with the change in volume caused by increasing pressure. Further detailed studies have been carried out on amorphous epoxy thermosets under uniaxial[72] or isotropic pressure.[72,73] Generally, the increase in pressure has an opposite effect on both annihilation quantities compared to increasing temperature. It has been found

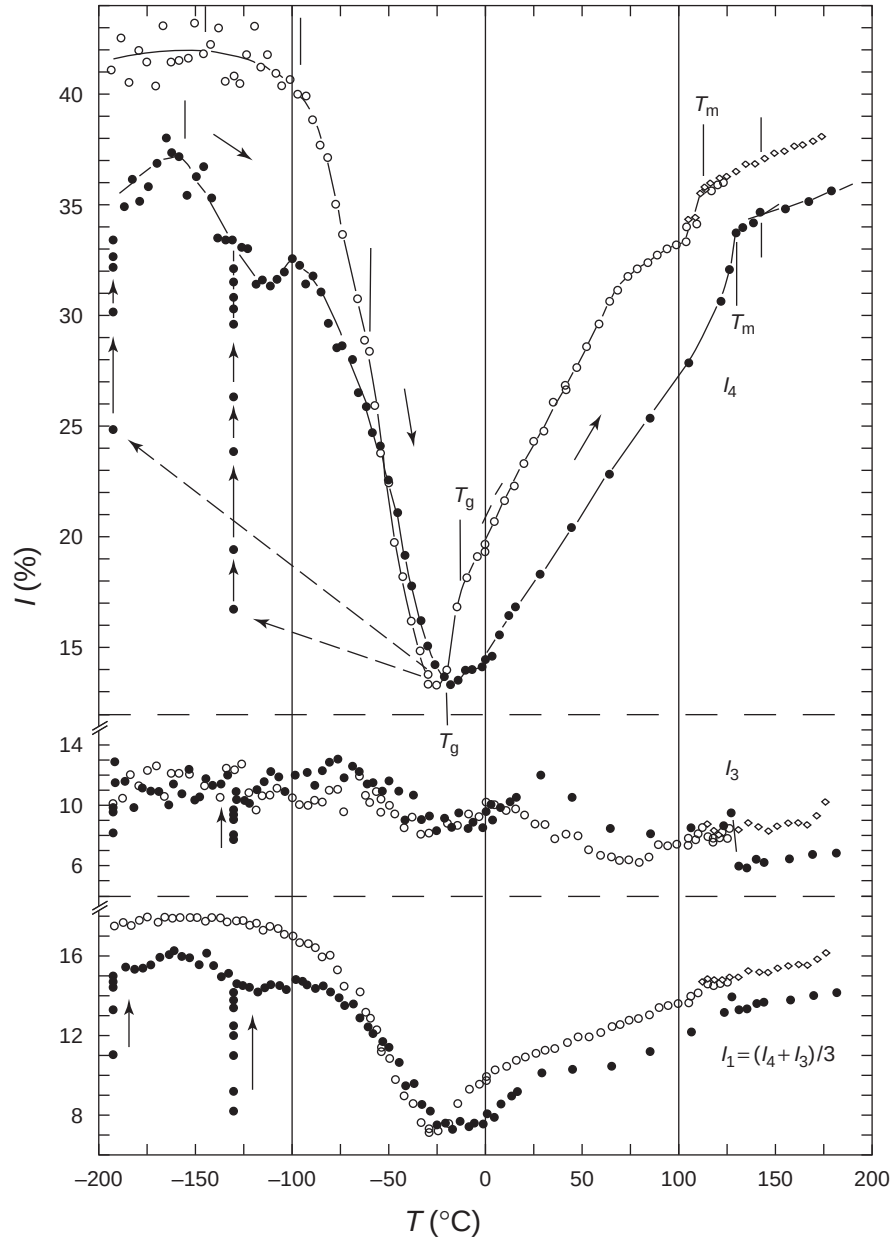

Figure 7 Relative intensities of positrons from fourth-component fit in linear PE compared with branched PE as a function of temperature. (Reproduced from Reiter and Kindl[66] by permission of Wiley-VCH Verlag, Berlin.)

that *o*-Ps lifetime and *o*-Ps relative intensity at temperatures below and above T_g decrease with increasing pressure, in the latter case more markedly than in the former one. Moreover, the pressure dependence of the mean free-volume hole size can be described by the negative exponential relationship. Similarly to the isobaric expansion coefficient of the free-volume hole size α_{Vh}, the isothermal compressibility of the mean free-volume hole size $\beta_{Vh} = -(1/V_h)(\Delta V_h/\Delta p)$ is one to two orders of magnitude higher than that of the macroscopic volume $\beta_V = -(1/V)(\Delta V/\Delta p)$. All these findings are connected to the above-mentioned fact that PALS probes only free-volume regions, while the dilatometric measurements involve both the free and occupied volume.

The deformation properties of polymers and the influence of various types of deformation on the microstructure of polymers form one of the most important topics of polymer science and technology. In a typical stress σ–strain $\varepsilon = (l - l_0)/l_0$ experiment, where l_0, l are the original and actual length of sample, several regimes of stretching behavior can be distinguished dependent on temperature and strain rate $d\varepsilon/dt$.[74] Usually, the stress

σ- versus elongation ε curve exhibits a yield point ε_y marking the boundary between elastic reversible deformation under smaller stress at $\varepsilon < \varepsilon_y$, and the inelastic irreversible one at $\varepsilon > \varepsilon_y$, which is characterized by essential structural reorganization of the material. For this reason, several PALS studies have focused on measuring these structural changes via the changes in *o*-Ps annihilation characteristics. These include both morphological types of polymer, such as amorphous PEEK,[75] PMMA,[75,76] and PC,[77,78] as well as semicrystalline PE,[79] and PTFE,[80] which have amorphous regions in the elastic state, and semicrystalline PEEK with glassy disordered regions.[81]

The first study[75] of the influence of cold tensile deformation on thermoplastics by the PALS method comes from investigation of PEEK, where a slight increase in τ_3 and a very slight decrease in I_3 have been found, the latter attributed to stress-induced crystallization in the stretched sample compared to that of the isotropic material. The anisotropy of free-volume holes has been determined by a one-dimensional ACAR method.[75] It was found that uniaxial deformation causes a change in the shape of free-volume holes from spherical to ellipsoidal in PEEK as well as in PMMA.

The preyielding regime of deformation in PC has been studied by using an in situ experimental set-up.[77] At room temperature, τ_3 grows with tensile strain up to a level of 4%, in agreement with the increase of the macroscopic volume up to 3.5% and in accord with the linear part of the stress–strain curve showing typical macroscopic yielding phenomenon at $\varepsilon_y \cong 6\%$. I_3 remains constant so that the free-volume hole fraction given by the product $V_h \cdot I_3$ increases with tensile deformation. However, the relative increase of the fractional free-volume is larger than the fractional increase in macroscopic volume, indicating that the free-volume increase is larger than that generated by an affinity expansion of the free-volume holes.[77] The detailed results of in situ study of the whole yielding regime under tension and compression in PC can be found in Xie et al.[78] The preyielding region is similar to the previous work,[78] but after crossing the yield point τ_3, I_3 decreased in such a way that the product $V_h \cdot I_3$ initially increased quite linearly, but then began to decrease at the level depending on the strain rate, being higher for a higher deformation rate. Similar results have been obtained on PMMA, but under zero-stress at the corresponding residual strain level.[76]

Semicrystalline polymers with an amorphous phase in the elastic state are discussed in Refs 4, 79 and 80. For PE,[4] a very large deformation up to a draw ratio $\lambda = l/l_0 = 12$ has influenced the lifetime spectrum only modestly, with slight decreases in τ_3 and I_3 remaining practically constant. It reflects the fact that *o*-Ps is formed in between the crystalline phase and hence the crystallite reorganization caused by deformation has no impact on the free-volume microstructure of the stretched matrix. More detailed four-component analysis[79] of hot-deformed high-density PE and draw ratios $\lambda \in \langle 1; 5 \rangle$ revealed that the two longer lifetime components are sensitive to the deformation-induced change in microstructure. Initially, τ_3, I_3 increase, and τ_4, I_4 decrease up to $\lambda \cong 2$, and at $\lambda > 3$ all parameters reach constant values. The third component has been attributed to *o*-Ps localization in the deformed paracrystalline phase, while the fourth component has been attributed to *o*-Ps localization in the increasingly ordered fibrous amorphous phase and/or in the elongated free-volume hole entities. The annihilation properties of uniaxially deformed PTFE[80] have been studied as a function of residual strain over the whole deformation range. Applying four-component fit, according to I_4 the stretching process can be divided into three stages, that is, an elastic stage, a plastic-flow region and finally, a strain-hardening regime, with the product $V_h \cdot I_3$ being constant in each stage. On the other hand, τ_4 increases with elongation in the elastic range, reaches a maximum in the plastic-flow regime, and finally tends to decrease and level off in the strain-hardening region.

The majority of free-volume interpretations of the τ_3 data have been realized by using a simple model considering a spherical shape for the free-volume hole (section 4.4). Although this approximation is plausible for isotropic materials, as confirmed by a one-dimensional ACAR experiment,[75] a nonspherical form of the free-volume holes can be expected for stretched samples. An ellipsoidal model of the free-volume hole, relating *o*-Ps lifetime to the anisotropic dimension of an ellipsoidal shape of the free-volume hole has been formulated.[35] The model predicts that the *o*-Ps lifetime in such an ellipsoidal hole will be shorter than that in a spherical one with the same volume. Application to a series of in situ uniaxially stretched semicrystalline PEEK samples with an amorphous phase in the glassy state confirmed the theoretical prediction of a large reduction of *o*-Ps lifetime at a slightly increasing *o*-Ps intensity, indicating a significant shape change after deformation. The last finding for the anisotropy of free-volume hole dimensions has been supported by the two-dimensional ACAR technique.[81]

5.1.3 Time Effects

The most important parameter in any physicochemical measurement of material properties is time, as it is an independent variable. In general, polymers are viscoelastic materials exhibiting complex temperature and time-dependent property behavior.[51,82] From the thermodynamic point of view, amorphous polymers or the amorphous phase of semicrystalline polymers

can exist in an equilibrium supercooled liquid, in the rubbery state above the glass transition temperature T_g or in the nonequilibrium glassy state below T_g. The latter is characterized by two basic phenomena: (1) formation history, that is the glassy state properties, depend on the preparation mode performed, for example, by rapid cooling (quenching) or by slow cooling of the sample, by formation pressure and so on, and (2) structural relaxation and physical ageing, that is, the evolution over time of thermodynamic properties such as macroscopic volume[51] and enthalpy, or of various mechanical and other physical properties represented by various characteristic relaxation times.[82]

One of the concepts most used to explain these often complicated formation and time dependences is based on the free-volume concept,[51,82] that is on the idea of the excess free-volume over the equilibrium free-volume level and its time evolution. For this reason, the PALS method has been often used to detect the formation history and the time evolution of the free-volume via the *o*-Ps annihilation characteristics.

On the other hand, the PALS method is characterized by two timescales. The first timescale is connected with the duration of its own annihilation process (a few nanoseconds) and the second is linked to the duration of the entire experiment, because a certain time window is necessary for data collection (several tens of minutes, up to a few hours), depending on the radiation activity of the positron source. Consequently, the final experimental response of a given material is the result of an often complicated interplay between both the intrinsic physicochemical features of the material investigated and the two above-mentioned characteristic timescales of the PALS method.

The first attempt to address the structural relaxation or physical ageing of polymers via the PALS method came from McGervey et al., using PS as an example.[83] In this and further earlier studies on amorphous polymers such as PC,[84,85] PS,[45] PVAc,[21] and PMMA,[86] as well as semicrystalline ones such as i-PP,[87] the time evolution of the *o*-Ps annihilation characteristics below T_g was believed to reflect the long-term structural relaxation, that is, that I_3 is determined by a physical factor, dominated by the density of the free-volume hole sites.

The general features of thermal cycling measurements over a wide temperature range are demonstrated for PIB in Figures 3 and 4. Two types of PALS experiment were used: one in a slow heating regime during a stepwise increase in the temperature after relatively rapid cooling down to the cryogenic temperature range, and the other one in a cooling regime during a stepwise decrease in the temperature. Two basic findings are evident: (1) the mean lifetime τ_3 is independent of the thermal cycle or the means of measurement, respectively. The constant value of the mean free-volume hole size reflects the independence of "clustering" of the constituents of the polymer matrix forming the hole defect. (2) On the other hand, the relative intensity, I_3, representing the *o*-Ps formation probability and the density of the free-volume hole sites, depends on the thermal treatment of the sample and the method of measurement. Evidently, the lower I_3 level at very low temperatures in the slowly cooled sample reflects, at least partially, the course of structural relaxation in the glassy state below T_g during the cooling procedure. Detailed studies[88–92] have revealed that the extent to which the *o*-Ps relative intensity, I_3, changes with positron exposure time, depends on the chemistry of polymer and on such external factors as the activity of the positron source and the method of annealing the sample – that is continuously together with irradiation or discontinuously outside the source. These experiments indicate the role of additional physical factors such as electric field and/or electron or positron scavenging reactions. For these reasons, in order to study structural relaxation by the PALS method, one must perform investigations under very well-defined and controlled experimental conditions that include annealing outside the measurement cell, using the same location of the sample for subsequent measurements, making a compromise between the source activity and the duration of data collection and lastly eliminating the electric charging effect by earthing the source.

5.2 Continuous Lifetime Analysis

In general, in disordered regions of polymers, the nonuniform distribution of electron density forms the heterogeneity of the local molecular environment in which annihilation takes place. Thus, the free-volume distribution is expected to generate distributions of lifetimes (sections 4.3 and 4.4). The first application of continuous lifetime analysis was carried out by Gregory in 1991, on PTFE.[93] Since then, many studies on all types of polymers, dependent on chemical structure, morphology and practically all important external variables such as temperature, pressure and mechanical stress, have been performed because of the greater information content of this sort of annihilation data analysis compared to a finite term of evaluation. These results have had a great impact, not only on polymer analytical characterization but also on the various branches of polymer chemistry and physics such as dynamics, penetration and so on.

5.2.1 Temperature Effects

The first systematic studies of the free-volume microstructure of polymers in terms of the free-volume hole radius distribution function $f(R_h)$ and the free-volume

hole distribution $g(V_h)$ and their dependence on temperature were carried out on amorphous polymers such as epoxy resin,[94] PS,[95] PC,[96] as well as on semicrystalline polymers like i-PP.[97] Following these pioneering studies, further systematic work has focused on the effects of variation in chemical structure within a certain class of polymers such as PCs[98] (at room temperature) and polyalkylmethacrylates[99] (over a wide temperature range) or on the effect of morphological change caused by phase transformation from the amorphous state to the semicrystalline state in PET.[100]

Typical results of the temperature dependence of both $f(R_h)$ and $g(V_h)$ over a wide temperature range are demonstrated by our illustrative case for PIB in Figure 8. The most important features can be summarized as follows:

1. The maxima of the $f(R_h)$ and $g(V_h)$ distribution functions shift to higher values with increasing temperature, consistent with the τ_3 results of a finite lifetime term analysis.
2. The widths of both $f(R_h)$ and $g(V_h)$ distribution functions broaden with increasing temperature; the broadening is more pronounced in the elastic state above T_g than in the glassy state below T_g.
3. Quantitatively, $f(R_h)$ distribution functions are approximately symmetric, so that they may be approximated by a Gaussian distribution function, while $g(V_h)$ distribution functions are highly asymmetric, with a pronounced tail on the high-value side, and they can be described by a log normal distribution function, Equation (14)

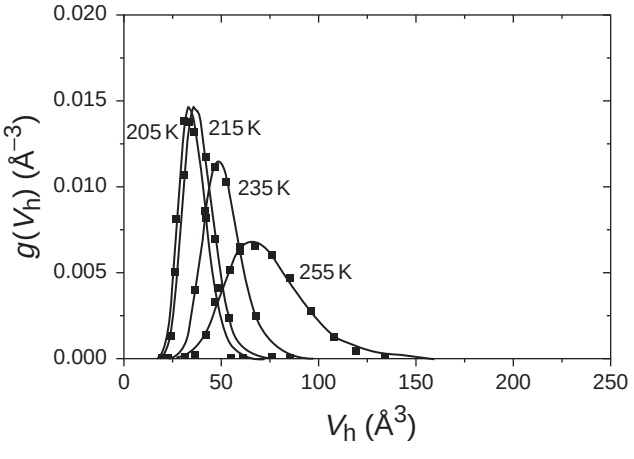

Figure 8 Free-volume hole distribution functions $g(V_h)$ as a function of temperature in PIB fitted by the log normal functions in Equation (14).

$$g(V_h) = \left(\frac{1}{\sqrt{2\pi w}} \quad \exp \quad \frac{-[\ln^2(V_h/V_{h,m})]}{2w^2} \right\} \qquad (14)$$

where $V_{h,m}$ and w are two basic parameters of the distribution function: the mean free-volume hole size and the standard deviation, respectively.

4. From the $f(R_h)$ and $g(V_h)$ results it follows that the R_h values are of the order of magnitude of interatomic distances, while the V_h values reach a part or a few times the occupied volume of the basic structural unit as represented by the particular van der Waals volume V^w_{mon}.

5.2.2 *Mechanical Effects*

The first analyses of the pressure dependence of $f(R_h)$ and $g(V_h)$ distribution functions on two systems studied so far, amorphous epoxy[101] and semicrystalline i-PP,[102] have revealed the following features:

1. At a given temperature the maxima of both the distribution functions shifts to lower values with increasing pressure.
2. The width of the free-volume distribution function narrows with increasing pressure.

Both these findings indicate compressing and collapsing of the free-volume holes, with a quasi-isotropic compression of the sample.

The change in the lifetime distribution with deformation has been studied on examples of hot-drawn PE.[79] The bimodal character of the distribution functions corresponding to the two long-lifetime components τ_3 and τ_4 has been found in both undeformed and deformed samples. The maximum of the longest lifetime component from the amorphous phase, shifts to lower values owing to deformation, and the width of the distribution increases slightly. The maximum of the third component increases slightly, together with an increase in the width, and was attributed to an increase in the paracrystalline disorder of the originally crystalline phase.

5.2.3 *Time Effects*

In section 5.1.3, we mentioned that the mean *o*-Ps lifetime τ_3 is independent of time for both amorphous and semicrystalline polymers. On the other hand, a special investigation of the crystallization process in PET[100] has shown that while the mean *o*-Ps lifetime τ_3 slightly increases and the relative intensity I_3 decreases with increasing degree of crystallinity x_c, the free-volume hole distribution undergoes a rather dramatic change: the maximum of $g(V_h)$ shifts to slighter higher values,

with simultaneous narrowing of the $g(V_h)$ distribution with continuing phase transformation.

6 EMPIRICAL CORRELATIONS

Many scientifically and technologically important properties like dynamic properties, that is relaxation, rheological[103] and diffusion properties,[104] have been often interpreted within the free-volume concept. In general, there are many operationally defined equations for the free-volume, $V_f(T, p)$, which can be expressed in the following unified form, Equation (15)

$$V_f(T, p) = V(T, p) - V_{occ}(T, p) \quad (15)$$

where $V(T, p)$ is the macroscopic volume from dilatometry and $V_{occ}(T, p)$ is the so-called occupied volume.[105] The latter has been postulated on a thermodynamic basis to be van der Waals volume, V^w_{mon}, giving the empty free-volume, V_f^{empty}, or the extrapolated hypothetical macroscopic volume at 0 K, V (0 K, p) providing the expansion free-volume, V_f^{exp}, and finally, in the case of semicrystalline polymers, the crystalline volume, V_{cryst} (T, p) giving a measure of structural disorder caused by phase change. Another approach based on dynamic properties consists in fitting various empirical formulae which have a free-volume interpretation, such as the Vogel–Fulcher–Tamman–Hesse (VFTH) equation for viscosity or relaxation time data or, equivalently, the Williams–Landel–Ferry (WLF) equation for the shift factor results. Although the free-volume concept has been successfully used in correlating and explaining many physical quantities such as viscosity, relaxation time, correlation time and diffusion coefficient, this success was achieved at the expense of the existence of many different free-volume definitions. On the other hand, the present progress in direct measurement of the *effective* free-volume properties by the PALS method opens up a unique opportunity to directly relate the corresponding microstructure free-volume properties to the relevant dynamic and diffusion properties. Several empirical correlations of this kind will be presented.

In the case of dynamic properties of polymer liquids, an empirical correlation between the expansivity of the mean free-volume hole $\beta_h = \Delta V_h/\Delta T$ and the fragility m_g, as well as the nonexponentiality β_g, has been found for a series of seven polymers.[106] The fragility m_g is a measure of the non-Arrhenius character of the temperature dependence of viscosity or relaxation time, and the nonexponentiality coefficient β_g is a measure of the non-Debye time behavior of relaxation time. These correlations indicate a close relationship between the presence of the free-volume holes detectable by PALS and the basic dynamic temperature and time features of polymer liquids. Consequently, a direct empirical correlation between the effective free-volume data and the viscosity data in PIB via the WLF–Doolittle type equation has been established.[107]

In the case of diffusion behavior of small gas molecules in polymer matrices, several correlative studies have been presented.[98,108–111] The first systematic study of the mutual relationship between these characteristics was performed for diffusion of carbon dioxide and methane in nine rubbery polymers at temperatures from T_g up to $T_g + 70$ K.[108] The free-volume Fujita–Doolittle[112] model for diffusion was tested and it was found that a correlation between the diffusion coefficient D and the mean free-volume hole size V_h is better than that between the D values and the product $V_h \cdot I_3$. In a following work, the same authors studied the relationship between sorption behavior and annihilation characteristics in a series of polyimides in the glassy state below T_g.[109] In a further work,[110] an empirical correlation between the expansion free-volume V_f^{exp} and the free-volume hole size V_h at room temperature was revealed. Consequently, acceptable correlations between the mean free-volume hole sizes V_h in polymer matrices and the corresponding diffusion coefficients for a variety of gases, such as argon, nitrogen and oxygen, in a series of seven polymers with very different chemical structures have been presented via the Fujita–Doolittle-type expression. On the other hand, a correlation between the empty free-volume fraction and the measured free-volume hole fraction has been found for a series of four PCs.[98] Consequently, a number of free-volume correlations for the diffusion coefficients of oxygen and carbon dioxide in these materials have been presented. Finally, similar relationships between the permeability coefficient P, the product of the *o*-Ps lifetime and the relative intensity for carbon dioxide and methane in a series of seven polyetherimides have been found.[111]

ACKNOWLEDGMENTS

The author wishes to thank Drs J. Krištiak, O. Šauša, P. Bandžuch, K. Krištiaková and J.Zrubcová for their very close collaboration and enthusiasm during our investigations of polymeric materials, as well as to the Slovak Grant Agency VEGA for supporting this research with grant 2/4008/97.

ABBREVIATIONS AND ACRONYMS

ACAR	Angular Correlation of Annihilation Radiation

DBS	Doppler Broadening Spectroscopy
i-PP	Isotactic Polypropylene
MCA	Multi-channel Analyzer
NS	Neutron Scattering
PALS	Positron Annihilation Lifetime Spectroscopy
PAS	Positron Annihilation Spectroscopy
PBD	Polybutadiene
PC	Polycarbonate
PE	Polyethylene
PEEK	Polyetheretherketone
PET	Polyethyleneterephthalate
PIB	Polyisobutylene
PIP	Polyisoprene
PMMA	Polymethylmethacrylate
PS	Polystyrene
PTFE	Polytetrafluoroethylene
PVAc	Polyvinylacetate
SANS	Small-angle Neutron Scattering
SAXS	Small-angle X-ray Scattering
SEM	Scanning Electron Microscopy
STM	Scanning Transmission Microscopy
TAC	Time-to-Amplitude Converter
TEM	Transmission Electron Microscopy
WAXS	Wide-angle X-ray Scattering
WLF	Williams–Landel–Ferry

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Dielectric Spectroscopy in Analysis of Polymers • Dynamic Mechanical Analysis of Polymers and Rubbers • Neutron Scattering in Analysis of Polymers and Rubbers • X-ray Scattering in Analysis of Polymers

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction • Differential Scanning Calorimetry and Differential Thermal Analysis

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Structure Determination, X-ray Diffraction for

REFERENCES

1. P.R. Wallace, *Positron Annihilation in Solids and Liquids*, Solid State Physics, Academic Press, New York, 1–69, Vol. 10, 1960.
2. A.T. Stewart, L. Roelling (eds.), *Positron Annihilation*, Academic Press, New York, 1967.
3. R.N. West, 'Positron Studies of Condensed Matter', *Adv. Phys.*, **22**, 263–383 (1973).
4. A.E. Hamielec, M. Eldrup, O. Mogensen, P. Jansen, 'Positron Annihilation Techniques in Polymer Science and Engineering', *J. Macromol. Chem.-Rev. Macromol. Chem.*, **C9**, 305–337 (1973).
5. J.R. Stevens, 'Positron Annihilation' in *Methods of Experimental Physics*, ed. R.A. Fava, 371–403, Vol. 16A, 1980.
6. W. Brandt, A. Dupasquier (eds.), *Positron Solid State Physics*, North Holland, Amsterdam, 1983.
7. D.M. Schrader, Y.C. Jean (eds.), 'Positron and Positronium Chemistry', Elsevier, Amsterdam, 1988.
8. Y.C. Jean, 'Positron Annihilation Spectroscopy for Chemical Analysis: A Novel Probe for Microstructural Analysis of Polymers', *Microchem. J.*, **42**, 72–102 (1990).
9. Y.C. Jean, 'Characterizing Free Volumes and Holes in Polymers by Positron Annihilation Spectroscopy', in *Positron Spectroscopy of Solids*, eds. A. Dupasquier, A.P. Mills, Jr, IOS Press, Ohmsha, Amsterdam, 563–580, 1995.
10. Y.C. Jean, 'Positron Annihilation in Polymers', *Mat. Sci. Forum*, **175–178**, 59–70 (1995).
11. O.E. Mogensen, *Positron Annihilation in Chemistry*, Springer-Verlag, Berlin, 1995.
12. R.A. Pethrick, 'Positron Annihilation – A Probe for Nanoscale Voids and Free Volume?', *Prog. Polym. Sci.*, **22**, 1–47 (1997).
13. C.L. Wang, K. Hirata, J. Kawahara, Y. Kobayashi, 'Electric-field Dependence of Positronium Formation in Liquids and Polymers', *Phys. Rev. B*, **58**, 14 864–14 869 (1998).
14. A. Ore, University I Bergen Arbok. Naturvitenskap., Rekke, No. 12, 1949 [thesis].
15. J.E. Jackson, J.D. McGervey, 'Mean Lives of Positrons in Aqueous Solutions', *J. Chem. Phys.*, **38**, 300–303 (1963).
16. W. Brandt, S. Berko, W.W. Walker, 'Positronium Decay in Molecular Substances', *Phys. Rev.*, **120**, 1289–1295 (1960).
17. W. Brandt, I. Spirn, 'Positron Lifetime Spectra in Molecular Substances', *Phys. Rev.*, **142**, 231–237 (1966).
18. S.J. Tao, 'Positron Annihilation in Molecular Substances', *J. Chem. Phys.*, **56**, 5499–5510 (1972).
19. M. Eldrup, D. Lightbody, J.N. Sherwood, 'The Temperature Dependence of Positron Lifetimes in Solid Pivalic Acid', *Chem. Phys.*, **63**, 51–58 (1981).
20. H. Nakanishi, S.J. Wang, Y.C. Jean, in *Positron Annihilation Studies of Fluids*, ed. S.C. Sharma, World Scientific, Singapore, 292–298, 1988.
21. Y. Kobayashi, W. Zheng, E.F. Meyer, J.D. McGervey, A.M. Jamieson, R. Simha, 'Free Volume and Physical Aging of Polyvinylacetate Studied by Positron Annihilation', *Macromolecules*, **22**, 2302–2306 (1989).

22. Y.Y. Wang, H. Nakanishi, Y.C. Jean, T.C. Sandreczki, 'Positron Annihilation in Amine-cured Epoxy Polymers – Pressure Dependence', *J. Polym. Sci., Part B: Polym. Phys.*, **28**, 1431–1441 (1990).
23. H. Cao, R. Zhang, J.P. Yuan, C.M. Huang, Y.C. Jean, R. Suzuki, T. Ohdaia, B. Nielsen, 'Free-volume Hole Model for Positronium Formation in Polymers: Surface Studies', *J. Phys. Condens. Matter*, **10**, 10429–10442 (1998).
24. K. Hirata, Y. Kobayashi, Y. Ujihara, 'Diffusion Coefficients of Positronium in Amorphous Polymers', *J. Chem. Soc., Faraday Trans.*, **92**, 985–988 (1996).
25. Y.C. Jean, 'Comments on the Paper: Can Positron Annihilation Lifetime Spectroscopy Measure the Free-Volume Hole Size Distributions in Amorphous Polymers?', *Macromolecules*, **29**, 575–577 (1996).
26. A.H. Baugher, W.J. Kossler, K.G. Petzinger, 'Does Quantum Mechanical Tunneling Affect the Validity of Hole Distributions Obtained from Positron Annihilation Lifetime Measurements?', *Macromolecules*, **29**, 7280–7284 (1996).
27. O.E. Mogensen, 'Spur Model of Positronium Formation', *J. Chem. Phys.*, **60**, 998–1004 (1974).
28. P. Kirkegaard, N.J. Pederson, M. Eldrup, 'PATFIT-88: A Data-Processing System for Positron Annihilation Spectra on Mainframe and Personal Computers', Tech. Rep. Risø-M-2740, Risø National Laboratory, DK-4000 Roskilde, Denmark, 1989.
29. D.M. Schrader, in *Positron Annihilation*, eds. P.G. Coleman, S.C. Sharma, L.M. Diana, North-Holland, Amsterdam, 912–914, 1982.
30. S.W. Provencher, 'CONTIN (Version 2)' User's Manual, Tech. Rep. EMBL-DA07, Max-Planck-Institut für Biophysikalische Chemie, Göttingen, 1984.
31. R.B. Gregory, 'CONTIN (PAL-2): A Modification of CONTIN (Version 2) for the Determination of Positron Annihilation Lifetime Distributions', Kent State University, Kent, 1993.
32. L. Hoffmann, A. Shukla, M. Peter, B. Barbiellini, A.A. Manuel, 'Linear and Non-linear Approaches to Solve the Inverse Problem: Applications to Positron Annihilation Experiments', *Nucl. Instrum. Methods, Sect. A*, **335**, 276–287 (1993).
33. C. Wästlund, F.H.J Maurer, 'Positron Lifetime Distributions in Polymers Determined with the Maximum Entropy Method', *Nucl. Instrum. Methods, Sect. B*, **117**, 467–473 (1996).
34. A.P. Mills, Jr, 'Observation of the Positronium Negative Ion', *Phys. Rev. Lett.*, **46**, 717–721 (1981).
35. Y.C. Jean, H. Shi, 'Positronium Lifetime in an Ellipsoidal Free Volume Hole of Polymers', *J. Non-Cryst. Solids*, **172–174**, 806–814 (1994).
36. H.A. Hristov, B. Bolan, A.F. Yee, L. Xie, D.W. Gindley, 'Measurement of Hole Volume in Amorphous Polymers', *Macromolecules*, **29**, 8507–8516 (1996).
37. J. Bartoš, J. Krištiak, 'Free Volume Microstructure from the Positron Annihilation Lifetime Spectroscopy Method and its Relationships with the Microscopic and Macroscopic Dynamics in *ortho*-Terphenyl', *J. Phys.: Condens. Matter*, **11**, A371–A377 (1999).
38. J. Krištiak, J. Bartoš, O. Šauša, P. Bandžuch, 'The Expansion of Free-volume in Main Chain Polymers Seen by Positron Annihilation Lifetime Spectroscopy and the Vogel–Fulcher Temperature T_0', *Mater. Sci. Forum*, **255–257**, 35–39 (1997).
39. B.C. Groseclose, G.D. Loper, 'Temperature Dependence of Positron Mean Lives in Polymers', *Phys. Rev. A*, **137**, 939–941 (1965).
40. B.D. Malholtra, R.A. Pethrick, 'Positron Annihilation Studies of the Glass–Rubber Transition in Poly(alkylmethacrylates)', *Macromolecules*, **16**, 1175–1179 (1983).
41. J.E. Kluin, Z. Yu, S. Vleeshouwers, J.D. McGervey, A.M. Jamieson, R. Simha, K. Sommer, 'Ortho-Positronium Lifetime Studies of Free Volume in Polycarbonates of Different Structure: Influence of Hole Size Distribution', *Macromolecules*, **26**, 1853–1861 (1993).
42. Y.C. Jean, T.C. Sandreczki, D.P. Ames, 'Positronium Annihilation in Amine-cured Epoxy Polymers', *J. Polym. Sci., Part B: Polym. Phys.*, **24**, 1247–1258 (1986).
43. J. Bartoš, 'Free Volume Microstructure of Amorphous Polymers at Glass Transition Temperatures from Positron Annihilation Spectroscopy Data', *Colloid Polym. Sci.*, **274**, 14–19 (1997).
44. F.H. Hsu, P.K. Tseng, S.Y. Chuang, Y.L. Chong, 'Temperature Dependence of Positron Lifetimes in *trans*- and *cis*-1,4-Polyisoprene', in *Positron Annihilation*, eds. R.R. Hasiguti, K. Fujiwara, Aoba Aramaki, Sendai, 581–584, 1981.
45. A.D. Kasbekar, P.L. Jones, A. Crowson, 'Positron Annihilation Lifetime Evaluation of Thermal Cycling Effects in Atactic Polystyrene', *J. Polym. Sci., Part A: Polym. Chem.*, **27**, 1373–1382 (1989).
46. S. Millan, U. Geckle, B. Levay, H. Ache, 'Temperature Dependence of Molecular Motions in Polymethylmethacrylate Studied by Positron Annihilation Techniques', *Ber. Bunsen-Ges. Phys. Chem.*, **94**, 781–785 (1980).
47. D. Richter, A. Arbe, J. Colmenero, M. Monkenbusch, B. Farago, R. Faust, 'Molecular Motions in Polyisobutylene: A Neutron Spin-echo and Dielectric Investigation', *Macromolecules*, **31**, 1133–1143 (1998).
48. J.R. Stevens, A.C. Mao, 'Glass Transition of Atactic Polystyrene by *ortho*-Positronium Study', *J. Appl. Phys.*, **41**, 4273–4275 (1970).
49. D.H.D. West, V.J. McBrierty, C.F.G. Delaney, 'Positron Decay in Polymers. Molecular Weight Dependence in Polystyrene', *Appl. Phys.*, **7**, 171–174 (1975).
50. Z. Yu, U. Yahsi, J.D. McGervey, A.M. Jamieson, R. Simha, 'Molecular Weight Dependence of Free Volume in Polystyrene Studied by Positron Annihilation

Measurements', *J. Polym. Sci., Part B: Polym. Phys.*, **32**, 2637–2644 (1994).
51. A.J. Kovacs, 'Transition Vitreuse Dans Les Polymeres Amorphes. Etude Phenomenologique', *Fortsch. Hochpolym. Forsch.*, **3**, 394–507 (1963).
52. J.M. Roe, R. Simha, 'Thermal Expansivity and Relaxation Behavior of Amorphous Polymers at Low Temperatures', *Int. J. Mater. Sci.*, **3**, 193–227 (1974).
53. J. Bartoš, J. Krištiak, T. Kanaya, 'Free Volume and Microscopic Dynamics in Amorphous Polymers', *J. Non-Cryst. Solids*, 234–236, 435–436 (1997).
54. J. Bartoš, P. Bandžuch, O. Šauša, K. Krištiaková, J. Krištiak, T. Kanaya, W. Jenninger, 'Free Volume Microstructure and Its Relationship to the Chain Dynamics in *cis*-1,4-Polybutadiene as Seen by PALS', *Macromolecules*, **30**, 6906–6912 (1997).
55. R.A. Ferrell, 'Long Lifetime of Positronium in Liquid Helium', *Phys. Rev.*, **108**, 167–168 (1957).
56. L.O. Roelling, T.M. Kelly, 'Cavity Formation in Helium', *Phys. Rev. Lett.*, **18**, 387–390 (1967).
57. O.E. Mogensen, 'Information on the Structure of Liquids and Solutions Obtained by Positron Annihilation', *Electrochim. Acta*, **33**, 1203–1210 (1988).
58. S. Beret, J.M. Prausnitz, 'Densities of Liquid Polymers at High Pressure', *Macromolecules*, **8**, 536–839 (1975).
59. J. Krištiak, J. Bartoš, K. Krištiaková, O. Šauša, P. Bandžuch, 'Free-volume Microstructure of Amorphous Polycarbonate at Low Temperatures Determined by Positron Annihilation Lifetime Spectroscopy', *Phys. Rev.*, **49**, 6601–6607 (1994).
60. C.L. Wang, T. Hirade, F.H.J. Maurer, M. Eldrup, N.J. Pedersen, 'Free-volume Distribution and Positronium Formation in Amorphous Polymers: Temperature and Positron Irradiation Time Dependence', *J. Chem. Phys.*, **108**, 4654–4661 (1998).
61. A. Uedono, T. Kawano, S. Tanigawa, M. Ban, M. Kyoto, T. Uozomi, 'Study of Relaxation Processes in Polethylene and Polystyrene by Using Positron Annihilation', *J. Polym. Sci., Part B: Polym. Phys.*, **34**, 2145–2151 (1996).
62. N.G. McCrum, B.E. Read, G. Williams, *Anelastic and Dielectric Effects in Polymeric Solids*, Wiley, London, 1967.
63. R.F. Boyer, 'T_{ll} from the Temperature Dependence of ^{13}C NMR Correlation Times τ_c', *J. Polym. Sci., Part B: Polym. Phys.*, **26**, 893–910 (1988).
64. H. Nakanishi, Y.C. Jean, E.G. Smith, T.C. Sandreczki, 'Positronium Formation at Free-volume Sites in the Amorphous Regions of Semi-crystalline PEEK', *J. Polym. Sci., Part B: Polym. Phys.*, **27**, 1419–1424 (1989).
65. L. Xie, D.W. Gindly, H.A. Hristov, A.F. Yee, 'Positronium Formation in Semi-crystalline Polyethyleneterephthalate', *Polymer*, **35**, 14–17 (1994).
66. G. Reiter, P. Kindl, 'Positron Lifetime Investigations on Linear Polyethylene Compared to Branched Polyethylene', *Phys. Status Solidi A*, **118**, 161–168 (1990).
67. T. Suzuki, Y. Oki, M. Numajiri, T. Miura, K. Kondo, Y. Ito, 'Positron Annihilation in Nonirradiated and Irradiated Polyethylenes', *J. Polym. Sci., Part B: Polym. Phys.*, **30**, 517–525 (1992).
68. G. Dlubek, K. Saarinen, H.M. Fretwell, 'The Temperature Dependence of the Local Free Volume in Polyethylene and Polytetrafluroethylene: A Positron Lifetime Study', *J. Polym. Sci., Part B: Polym. Phys.*, **36**, 1513–1528 (1998).
69. Y. Ohko, A. Uedono, Y. Ujihara, 'Thermal Variation of Free-volume Size Distribution in Polypropylenes Probed by Positron Annihilation Lifetime Techniques', *J. Polym. Sci., Part B: Polym. Phys.*, **33**, 1183–1190 (1995).
70. A. Uedono, T. Kawano, S. Tangawa, M. Ban, M. Kyoto, T. Uozumi, 'Transition and Relaxation Processes of Polyethylene, Polypropylene and Polystyrene Studied by Positron Annihilation', *J. Polym. Sci., Part B: Polym. Phys.*, **35**, 1601–1609 (1997).
71. R.K. Wilson, P.O. Johnson, R. Stump, 'Variation in Positron Lifetime with Pressure', *Phys. Rev.*, **129**, 2091–2095 (1963).
72. Y.Y. Wang, H. Nakanishi, Y.C. Jean, T.C. Sandreczki, 'Positron Annihilation in Amine-cured Epoxy Polymers: Pressure Dependence', *J. Polym. Sci., Part B: Polym. Phys.*, **28**, 1431–1441 (1990).
73. Q. Deng, C.S. Sundar, Y.C. Jean, 'Pressure Dependence of Free-volume Hole Properties in an Epoxy Polymer', *J. Phys. Chem.*, **96**, 492–495 (1992).
74. I.M. Ward, *Mechanical Properties of Solid Polymers*, Wiley-Interscience, London, 1971.
75. Y.C. Jean, H. Nakanishi, L.Y. Hao, T.C. Sandreczki, 'Anisotropy of Free-volume Hole Dimensions in Polymers Probed by Positron Annihilation Spectroscopy', *Phys. Rev. B*, **42**, 9705–9708 (1990).
76. Q.A. Hasan, M.C. Boyce, X.S. Li, S. Berko, 'An Investigation of the Yield and Post-yield Behavior and Corresponding Structure of Polymethylmethacrylate', *J. Polym. Sci., Part B: Polym. Phys.*, **31**, 185–195 (1993).
77. M.Y. Ruan, H. Moaddel, A.M. Jamieson, R. Simha, J.D. McGervey, 'Positron Annihilation Lifetime Studies of Free Volume Changes in Polycarbonate under Static Tensile Deformation', *Macromolecules*, **25**, 2407–2411 (1992).
78. L. Xie, D.W. Gidley, H.A. Hristov, A.F. Yee, 'Evolution of Nanometer Voids in Polycarbonate under Mechanical Stress and Thermal Expansion', *J. Polym. Sci., Part B: Polym. Phys.*, **33**, 77–84 (1995).
79. G. Dlubek, T. Lübke, H.M. Fretwell, M.A. Alam, A. Wutyler, H.J. Radusch, 'The Local Free Volume in Deformed Polyethylene: A Positron Lifetime Study', *Acta Polym.*, **49**, 236–243 (1998).
80. C.L. Wang, B. Wang, S.Q. Li, S.J. Wang, 'Effects of Deformation on the Microstructure of PTFE Studied by Positron Annihilation', *J. Phys.: Condens. Matter*, **5**, 7515–7520 (1993).

81. Y.C. Jean, Y. Rhee, Y. Lou, H.L. Yen, H. Cao, K. Cheong, Y. Gu, 'Anisotropy of Hole Structures in Polymers Probed by Two-dimensional Angular Correlation of Annihilation Radiation', *Phys. Rev. B*, **54**, 1785–1790 (1996).
82. L.E. Struik, 'Physical Aging in Amorphous Polymers and Other Materials', Elsevier, Amsterdam, 1978.
83. J.D. McGervey, N. Panigrahi, R. Simha, A.M. Jamieson, 'A Positron Lifetime Study of Physical Aging in Polystyrene' in *Positron Annihilation*, eds. P.C. Jain, R.M. Singru, K.P. Gopinatlan, World Scientific, Singapore, 690–691, 1985.
84. A.J. Hill, P.L. Jones, J.H. Lind, G.W. Pearsall, 'A Positron Annihilation Lifetime Study of Isothermal Structural Relaxation in Bisphenol–A Polycarbonate', *J. Polym. Sci., Part A: Polym. Chem.*, **26**, 1541–1552 (1988).
85. W.J. Davis, R.A. Pethrick, 'Positron Annihilation Studies of Physical Ageing in Polycarbonate', *Europ. Polym. J.*, **34**, 1747–1754 (1998).
86. W.J. Davis, R.A. Pethrick, 'Investigation of Physical Aging in PMMA Using Positron Annihilation, Dielectric Relaxation and Dynamic Mechanical Thermal Analysis', *Polymer*, **39**, 255–266 (1998).
87. J.H. Lind, P.L. Jones, G.W. Pearsall, 'A Positron Annihilation Lifetime Study of Isotactic Polypropylene', *J. Polym. Sci., Part A: Polym. Chem.*, **24**, 3033–3047 (1986).
88. P.L. O'Connor, C.W. Kocher, B.G. Laudes, S.L. Anderson, P.T. DeLassus, E.A. Grulke, 'Physical Aging Artifacts in Polymers as Measured by Positron Annihilation Lifetime Spectroscopy', *ACS Polym. Prepts*, **33**, 308–309 (1992).
89. M. Welander, F.H. Maurer, 'Decrease in *o*-Ps Formation in Polymers during Positron Lifetime Measurements', *Mater. Sci. Forum*, **105–110**, 1811–1814 (1992).
90. X.S. Li, M.C. Boyce, 'On the Measurement of Structural Relaxation in Polymers Using Positron Annihilation Lifetime Spectroscopy', *J. Polym. Sci., Part B: Polym. Phys.*, **31**, 869–873 (1993).
91. T. Suzuki, Y. Oki, M. Numajiri, T. Miura, K. Kondo, N. Oshima, Y. Ito, 'Radiation Effect on Polypropylene Studied by the Relaxation Behavior at Low Temperatures Using Positron Annihilation', *Polymer*, **37**, 5521–5524 (1996).
92. Z.L. Peng, B.G. Olson, J.D. McGervey, A.M. Jamieson, 'Temperature and Time Dependence of *ortho*-Positronium Formation in Polystyrene', *Polymer*, **40**, 3033–3040 (1999).
93. R.B. Gregory, 'Free-volume and Pore Size Distributions Determined by Numerical Laplace Inversion of Positron Annihilation Lifetime Data', *J. Appl. Phys.*, **70**, 4665–4670 (1991).
94. Q. Deng, F. Zandiehnadem, Y.C. Jean, 'Free-volume Distributions of an Epoxy Polymer Probed by Positron Annihilation: Temperature Dependence', *Macromolecules*, **25**, 1090–1095 (1992).
95. J. Liu, Q. Deng, Y.C. Jean, 'Free-volume Distributions of Polystyrene Probed by Positron Annihilation: Comparison with Free Volume Theories', *Macromolecules*, **26**, 7149–7155 (1993).
96. J. Krištiak, K. Krištiaková, O. Šauša, P. Bandžuch, J. Bartoš, 'Temperature Dependence of Free Volume Distributions in Polymers Studied by Positron Annihilation Lifetime Spectroscopy', *J. Phys. IV*, C4, **3**, 265–270 (1993).
97. Y. Ohko, A. Uedono, Y. Ujihara, 'Thermal Variation of Free-volume Size Distribution in Polypropylenes Probed by Positron Annihilation Lifetime Spectroscopy', *J. Polym. Sci., Part B: Polym. Phys.*, **33**, 1183–1190 (1995).
98. Y.C. Jean, J.P. Yuan, J. Liu, Q. Deng, H. Yang, 'Correlation Between Gas Permeation and Free-volume Hole Properties Probed by Positron Annihilation Spectroscopy', *J. Polym. Sci., Part B: Polym. Phys.*, **33**, 2365–2371 (1995).
99. H.L. Li, Y. Ujihara, S. Tanaka, T. Yamashita, K. Horie, 'Variation of Free Volume Parameters of Polyalkyl Methacrylate Derivatives in the Range from 15 to 300 K Probed by Positron Annihilation Lifetime Measurements', *J. Radioanal. Nucl. Chem.*, **210**, 543–553 (1996).
100. H.F.M. Mohamed, Y. Ito, M. Imai, 'Change of Distribution of Free Volume Holes during Crystallization of Polyethyleneterephthalate Revealed by PALS', *J. Chem. Phys.*, **105**, 4811–4855 (1996).
101. Q. Deng, Y.C. Jean, 'Free-volume Distributions of an Epoxy Polymer Probed by Positron Annihilation: Pressure Dependence', *Macromolecules*, **26**, 30–34 (1993).
102. Y.C. Jean, Q. Deng, 'Direct Measurement of Free-volume Hole Distributions in Polymers by Using a Positronium Probe', *J. Polym. Sci., Part B: Polym. Phys.*, **30**, 1359–1364 (1992).
103. J.D. Ferry, 'Viscoelastic Properties of Polymers', 3rd edition, Wiley, New York, 1980.
104. J. Crank, G.S. Park (eds.), 'Diffusion in Polymers', Academic Press, New York, 1968.
105. Y. Lipatov, 'The Iso-free-volume State and Glass Transition in Amorphous Polymers: New Development of the Theory', *Adv. Polym. Sci.*, **26**, 64 (1978).
106. J. Bartoš, J. Krištiak, 'Free-volume Aspects of the Strong–Fragile Classification of Polymer Liquids', *J. Non-Cryst. Solids*, 235–237, 293–295 (1998).
107. J. Bartoš, J. Krištiak, 'Experimental Verification of the WLF–Doolittle Equation for Supercooled Liquid Dynamics by PALS Method. Polyisobutylene', in *Progress and Trends in Rheology*, eds. I. Emri, R. Cvelbar, Steinkopff, Darmstadt, 246–247, 1998.
108. K. Okamoto, K. Tanaka, M. Katsuke, H. Kita, O. Suoka, Y. Ito, 'Correlation between Positron Annihilation and Gas Diffusion Properties of Various Rubbery Polymers', *Polym. J.*, **25**, 275–284 (1993).

109. K. Tanaka, K. Okamoto, H. Kita, Y. Ito, 'Correlation between Positron Annihilation and Sorption of CO_2 in Glassy Polymers', *Polym. J.*, **25**, 577–584 (1993).
110. Y. Kobayashi, K. Haraya, S. Hattori, T. Sasuga, 'Evaluation of Polymer Free Volume by Positron Annihilation and Gas Diffusivity Measurements', *Polymer*, **35**, 925–928 (1994).
111. G. Eastmond, J.H. Daly, A.S. McKinnon, R.A. Pethrick, 'Polyetherimides: Correlation of Positron Annihilation Lifetime Studies with Polymer Structure and Gas Permeability', *Polymer*, **40**, 3605–3610 (1999).
112. H. Fujita, 'Diffusion in Polymer–Diluent Systems', *Adv. Polym. Sci.*, **3**, 1–47 (1961).

Pyrolysis Techniques in the Analysis of Polymers and Rubbers

Thomas P. Wampler
CDS Analytical, Inc., Oxford, PA, USA

1 Introduction **1**
2 Chemistry: Degradation Mechanisms **2**
2.1 Chain Bonds are Weakest 2
2.2 Side Bonds are Weakest 3
2.3 Elimination of Small Molecules 5
2.4 Microstructure 5
3 Instrumentation **8**
3.1 Constant Heating: Isothermal Microfurnaces 9
3.2 Inductive Heating: Curie Point 10
3.3 Resistive Heating: Filaments 10
3.4 Other Heating Techniques 11
4 Interfacing **11**
4.1 Pyrolysis/Gas Chromatography 11
4.2 Pyrolysis/Mass Spectrometry 11
4.3 Pyrolysis/Fourier Transform Infrared Spectroscopy 12
4.4 General Considerations 12
5 Applications **12**
5.1 Polyolefins 12
5.2 Polyamides 12
5.3 Polyesters 12
5.4 Polyurethanes 13
5.5 Vinyl Polymers 13
5.6 Acrylics 13
5.7 Copolymers 15
Abbreviations and Acronyms **19**
Related Articles **20**
References **20**

Pyrolysis is defined as a chemical degradation reaction induced by thermal energy only,[1] and analytical pyrolysis is the study of materials based on the results of such reactions. Since the overall result of a pyrolysis reaction is the production of small molecules from larger ones, it is extensively utilized in the analysis of polymers, especially as a sample introduction tool since it permits the application of techniques such as gas chromatography (GC) to such macromolecules by breaking them into fragments which are compatible with the technique. Interpretation of the results of a pyrolytic analysis depend on an understanding of the degradation mechanisms involved, which reflect relative bond strengths within the macromolecule. Since specific polymers are chemically different from each other, the degradation products they make are different, providing characteristic information about the original macromolecule. Analytical pyrolysis, coupled with GC, mass spectrometry (MS), gas chromatography/mass spectrometry (GC/MS), Fourier transform infrared (FTIR) spectroscopy and other techniques, has been used in the study of a wide range of polymer types, including polyolefins, vinyl polymers, polyurethanes, acrylics, rubbers, polyamides and polyesters.

Analytical laboratories employ one of a variety of pyrolysis techniques (microfurnace, resistively or inductively heated filament) interfaced to a detection instrument (GC, MS, FTIR) to create an analytical system. These combined systems have strengths and weaknesses based on the limitations of the individual components and must be optimized for a particular analysis. Systems have been devised which permit the routine analysis of such difficult materials as cured and cross-linked paints and rubbers, textile fibers, filled composites, laminates, dried adhesives, copolymer systems, coatings and packaging. Practical applications are found in laboratories charged with the analysis of forensic evidence, museum artifacts, ink, paper, lubricants, automobile tires, clothing, varnish, wood products, furniture and food containers.

1 INTRODUCTION

Because so much of the chemistry observed in pyrolysis reactions involves free-radical reactions, the early history of pyrolysis largely parallels the development of free-radical theory. What is considered by many to be the first intentional application of pyrolysis for an analytical purpose, however, was the analysis of natural rubber by Williams[2] in 1862, well predating the development of free-radical theories. By heating a sample of rubber (polyisoprene) to the point of destruction, Williams produced both the monomer and dimer of the polymer. Free-radical mechanism theories developed in the early years of the 20th century, and the relationship between pyrolytic processes and free-radical generation was shown in 1929 in the well-known experiments by Paneth and Lautsch.[3] This work demonstrated that, first for tetramethyllead and then for tetraethyllead, high temperatures produced elemental lead and alkyl free radicals, which could subsequently react with a second source of the metal. By 1931, in referring to the work of Paneth, Rice[4] stated that, ''We can proceed

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

with some confidence to examine the mechanism of the decomposition of organic compounds from this point of view because the actual existence of the simple alkyl radicals has now been demonstrated''. Rice then went on to publish widely concerning free-radical degradation mechanisms experienced by organic molecules at high temperatures, including products from paraffins[5] and chain reaction mechanisms.[6] The evolution of early pyrolysis applications has been reported in some detail by Irwin, first in a 1979 review article[7] and later in a text[8] on analytical pyrolysis.

From an analytical standpoint, pyrolysis must be coupled with some other device for the identification of the products made, and by far the most common is GC. Analysts realized early on in the development of GC that pyrolysis/gas chromatography (Py/GC) provided a means of analyzing materials not suitable for GC in their native state. As early as 1962,[9] pyrolysis had been coupled to GC for the analysis of hydrocarbons, which led Dietz[10] to comment that the results suggested that, ''many different types of organic substances can be characterized by pyrolysis techniques.'' By the early 1970s Py/GC was already being applied to the analysis of such complex problems as the microstructure of ethylene–propylene copolymers[11] and monomer inversion in polypropylene.[12] A review by Brauer[13] in 1970 listed 200 references to polymer characterization using Py/GC, including qualitative analysis, quantitative estimation of copolymer composition and studies on thermal stability. In subsequent years, applications continued to grow, with the introduction of commercial instruments for pyrolysis, and the publication of additional texts on pyrolysis[14,15] and review articles.[16–18]

2 CHEMISTRY: DEGRADATION MECHANISMS

A considerable amount of work has been published on the degradation mechanisms experienced by polymers when heated to pyrolysis temperatures. This includes general reviews of mechanisms[19,20] and also studies into the degradation of specific polymers, including polybutadiene,[21] polystyrene,[22] polyesters,[23] polyurethanes[24] and fluoropolymers.[25]

Although any particular material may be studied for years in an effort to arrive at a definitive understanding of the degradation processes experienced during its pyrolysis, some fairly simple generalities may be proposed for an overall appreciation of the chemistry involved. Degradation mechanisms are frequently grouped into four categories, i.e. random scission, in which a polymer dissociates at multiple points along the chain to produce a variety of oligomeric products, unzipping, which causes the polymer to revert to monomer, side-group scission, which eliminates groups attached to the polymer chain, and char formation, which results in a complex nonvolatile residue. From the standpoint of analytical pyrolysis, which usually deals only with the volatile products of pyrolysis, the problem may be reduced to just two broad categories based on a simple model of bond types.

Many synthetic polymers may be seen to be variations of the simple model **(1)**, in which there is a long carbon chain (—C–C—) attached to which are side-groups, shown here as a, b, d and e. To make up the polymer, this simple unit is repeated many times, and different polymers are created by changing the identity of the side groups. For example, if all groups a–e are hydrogen atoms, the polymer is polyethylene. If a, b and d are hydrogen and e is a methyl group, the polymer is polypropylene. If e is Cl, the polymer is poly(vinyl chloride) (PVC), if both b and e are Cl, the polymer is poly(vinylidine chloride), and so on. Seen this way, there are just two kinds of bonds to be concerned with – those which hold the chain together (C–C) and those which

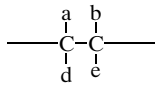

(1)

attach the side groups (C–a, C–b, C–d and C–e). Since pyrolysis is defined as a degradation mechanism induced by thermal energy only, bond dissociation is the principle pathway to the generation of products. As the polymer molecule is heated to pyrolysis temperature, the weakest bonds break first, and these bond dissociations will dictate most of what is seen as product. Consequently, two very broad degradation mechanisms arise. First are cases in which the C–C bonds holding the chain together are the weakest, and second are cases in which bonds attaching a side group are weaker than the C–C bonds of the chain.

2.1 Chain Bonds are Weakest

If the bonds holding the side groups to the chain are stronger than the bonds holding the polymer chain together, pyrolysis causes chain scission, producing fragments of the polymer with the side groups intact. This bond dissociation produces two free radicals, and the products which result from the pyrolysis depend on the nature and structure of these free radicals. Polyethylene, for example, dissociates to form two primary free radicals as shown in Scheme 1, which may then undergo any processes normal for free radicals, including β-scission, hydrogen transfers and so on. Immediate β-scission will

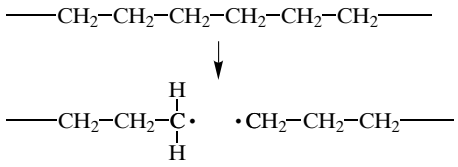

Scheme 1

produce a molecule of ethylene, the monomer, and regenerate the free radical on the chain.

Transfer of a hydrogen from the next carbon, followed by β-scission, will produce a molecule of propylene. The free-radical may abstract a hydrogen from a neighboring molecule and simply become a saturated chain end. A hydrogen may be transferred from the fifth carbon in the free radical, via a six-membered ring, to produce the secondary free radical shown in Scheme 2.

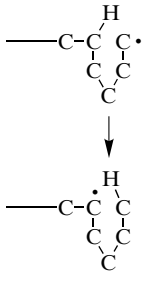

Scheme 2

This 1–5 hydrogen transfer is an important mechanism in understanding the product distribution in the pyrolyzate of many polymers. By transferring the hydrogen to the first position, a secondary, more stable free radical is formed at the fifth carbon. When this free radical undergoes β-scission, 1-hexene is formed, which is the trimer of ethylene, and the free radical continues on the chain of the polymer. It is important to recognize that the three monomer units which are eliminated together as the trimer are connected to each other in the same way that they were in the original polymer, and such oligomeric fragments therefore convey much microstructural information.

Since all of the carbon–carbon bonds along the chain have the same bond strength, at pyrolysis temperatures any of them could dissociate. Two dissociations near each other will produce an oligomeric fragment which may be small enough to be analytically significant. Since the free radical may take a hydrogen from a neighboring molecule, creating a chain ending in CH_3, eliminate a hydrogen or undergo β-scission, producing a chain ending in $C{=}CH_2$, the resulting hydrocarbons could have no double bonds, a double bond at one end or double bonds at both ends. Consequently, the pyrolyzate generated from polyethylene consists of chain fragments which are normal hydrocarbons, a mixture of alkanes, alkenes and dienes. All of these processes are evident in a pyrolysis chromatogram (pyrogram) of polyethylene, such as that shown in Figure 1. The triplets of peaks represent the diene, alkene and alkane of increasing chain length, with the C_{12} hydrocarbons eluting at about 14 min. The 1–5 hydrogen transfer accounts for the increased amount of hexene (at 2 min) over the amounts of pentene or heptene, eluting just before and after it. Since the hydrogen transfer may take place more than once, there is an increased amount of the pentamer (decene) and the heptamer (tetradecene) in addition to the trimer, so the pattern always shows larger peaks for the C_6, C_{10} and C_{14} hydrocarbons. The fact that all normal hydrocarbons up to C_{30} are represented indicates that the chain bonds are all equally likely to dissociate.

The presence of side groups other than hydrogen significantly affects product formation, even in cases where the chain-length bonds are the weakest. Poly(methyl methacrylate) (PMMA), for example, undergoes bond dissociation to produce structure **(2)**.

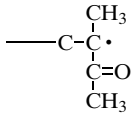

(2)

Unlike polyethylene, this is already a tertiary free radical, so internal rearrangements such as hydrogen transfers cannot produce a more stable species. Further, the carbon fifth from the end is not bonded to a hydrogen, so the 1–5 hydrogen transfer mechanism is impossible. As a result, the free radical simply undergoes β-scission, which produces a molecule of methyl methacrylate (MMA) monomer and regenerates the free radical on the polymer chain. This process continues almost to the exclusion of any other reaction, and the end result is essentially regeneration of the monomer. Consequently, the pyrogram produced from PMMA is very simple, showing little more than a single peak for the MMA monomer, as shown in Figure 2. In polymers where there is a hydrogen available for a 1–5 transfer, this mechanism usually occurs, resulting in the generation of trimer.[26] Poly(methyl acrylate), then, does produce methylacrylate trimer whereas PMMA does not, polystyrene does, but poly(α-methylstyrene) does not, and so on.

2.2 Side Bonds are Weakest

If the bond attaching a side group to the polymer chain is weaker than the bonds of the chain itself, the side group will be removed from the polymer before the

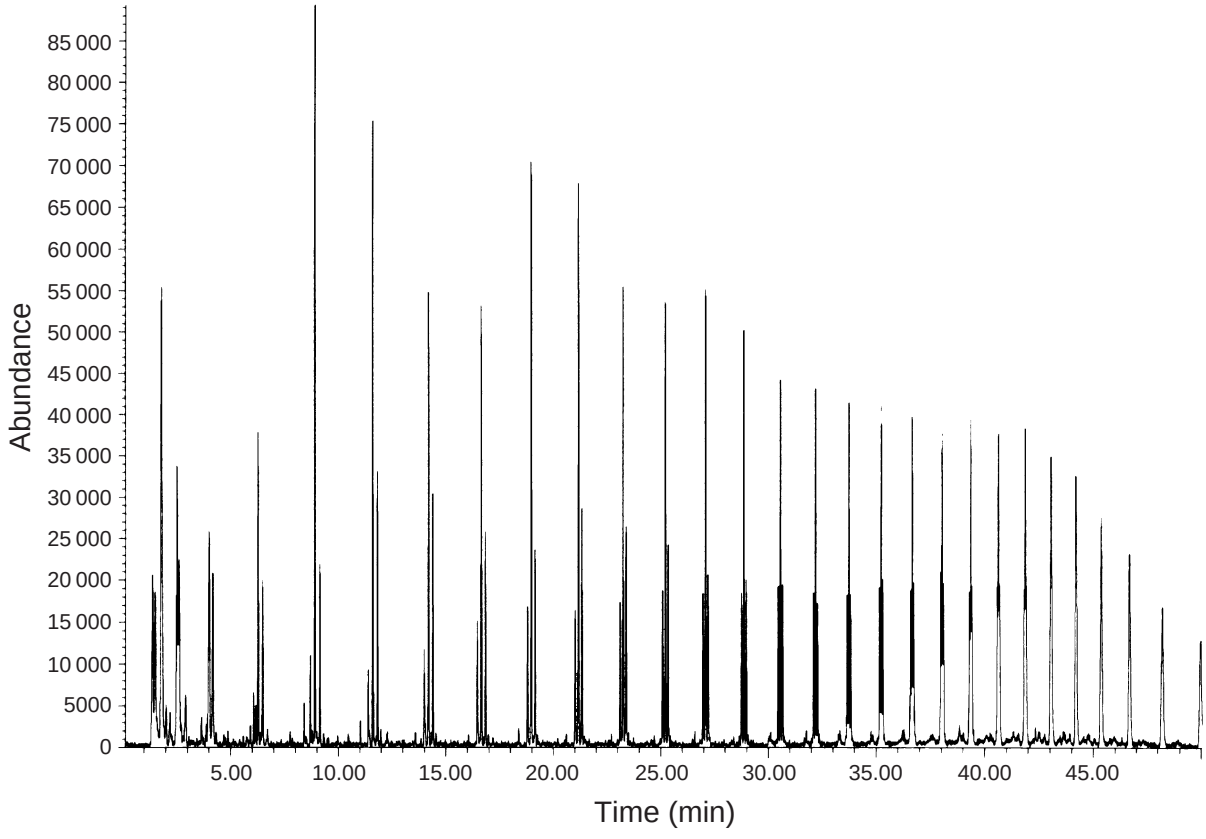

Figure 1 Pyrolysis of polyethylene, showing triplet *n*-hydrocarbon peaks.

chain breaks apart. Consequently, the monomer identity of the polymer is altered and no monomer appears in the pyrolyzate, whereas some monomer is almost always present if the chain-length bonds are the weakest. The absence of monomer does not hinder the ability to identify the polymer from the pyrolyzate, however, since the pyrolysis products are as characteristic and reproducible for this kind of polymer as for any other. PVC, for example, does not produce vinyl chloride monomer when it is pyrolyzed, because the C–Cl bond is weaker than the C–C bond of the chain. At relatively low temperatures (below 400 °C) the C–Cl bond dissociates, making a Cl free radical. This Cl free radical quickly picks up hydrogen from a neighboring carbon and forms HCl, leaving the polymer chain unsaturated. As the polymer is heated further, the unsaturated chain breaks apart, generating aromatics, including benzene, toluene and naphthalene. The combination of aromatics and HCl in a pyrolyzate (see Figure 13, described in section 5.5) is always characteristic of PVC. Poly(vinyl acetate) (PVA) degrades in a similar way, producing acetic acid plus aromatics, and a copolymer of PVA and PVC will produce aromatics plus both HCl and acetic acid.

Poly(vinylidene chloride) (PVDC) provides a good example of how relative bond strengths are reflected directly in the pyrolyzate. In the original polymer, the C–Cl bonds are weaker than the C–C bonds of the chain, so it too releases HCl, as shown in Scheme 3, leaving a chlorinated, unsaturated backbone of the polymer.

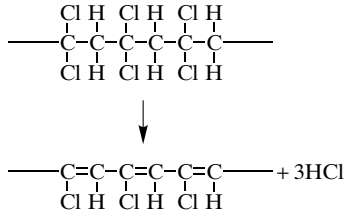

Scheme 3

The resulting material now contains C=C–Cl bonds, which are stronger than the C–Cl bond in the original polymer. Upon further heating, the backbone dissociates and cyclizes, forming aromatics, but the chlorines are retained, so the principal product is 1,3,5-trichlorobenzene. PVC then produces a variety of unchlorinated aromatics, PVDC generates chlorinated aromatics, but neither makes monomer.

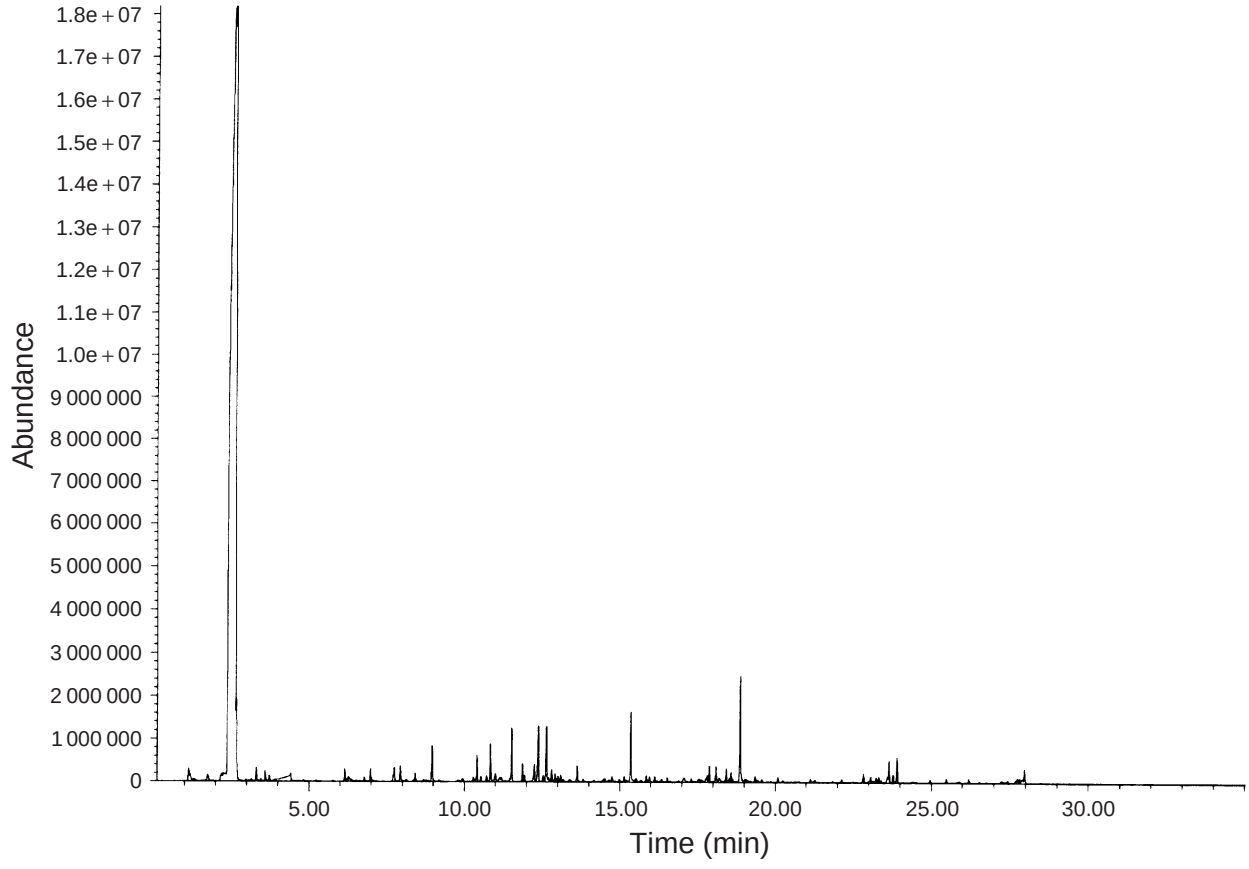

Figure 2 Pyrogram of PMMA. The largest peak is the monomer (MMA).

2.3 Elimination of Small Molecules

In addition to the degradations outlined above, some polymers produce simple gases as they are pyrolyzed, especially polymers including atoms other than carbon and hydrogen. Cellulose and cellulose-based materials including paper, cotton, rayon and cellophane produce considerable CO, CO_2 and water at pyrolysis temperatures[27] in addition to levoglucosan, furans and other oxygen-containing compounds. Ohtani et al.[28] have outlined a scheme accounting for the mono- and dinitriles formed in the pyrolysis of nylons in which the nitriles are formed from the polyamide via a dehydration step. Lytle et al.[29] reported the production of water, HCN and CO as products in the pyrolysis of phenolic urethanes. The products observed in the pyrolysis of polyesters[23] are formed by processes involving considerable decarboxylation.

2.4 Microstructure

Because the fragments generated from pyrolyzing a polymer retain structural information from the original macromolecule, studying these fragments provides considerable information about the nature of the polymer. This is even true regarding the arrangement of the monomeric units in the polymer, revealing structural details such as monomer inversion, tacticity and especially the difference between random and block copolymers. This is a direct result of the fact that differing arrangements of monomers along a chain produce chemically different materials, and these differences survive the pyrolysis event. In a homopolymer, for example, monomers which generally align themselves in a head-to-tail arrangement may form head-to-head or tail-to-tail bonds, which are analytically discernible. In the case of PVC,[30,31] for example, an inversion of this type would result in a structure such as **(3)**, instead of one in which the chlorines are attached to alternate carbons. The effect of having chlorines on neighboring carbons is the production of chlorinated aromatics, particularly chlorobenzene. Likewise for polypropylene,[32] tail-to-tail bonding would place two methyl groups an additional carbon apart, and consequently produce structures which are 2,5-dimethyl instead of 2,4-dimethyl substituted. In addition for polypropylene, since the polymer chain, which is frequently drawn as shown in structure **(4)**, contains asymmetric carbons, there are issues of stereoregularity. The relative positioning of the methyl groups on these asymmetric carbons in the polymer chain dictate the relative

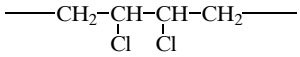

(3)

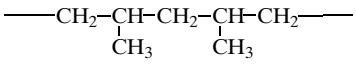

(4)

amounts of various specific degradation products, and this information is reflected in the pyrogram. Figure 3 shows a pyrogram of isotactic polypropylene and Figure 4 shows atactic polypropylene for comparison.

These considerations take on an additional significance in the case of copolymers. Because oligomers formed by pyrolysis, especially trimers and larger, show the positions of monomeric units as they were in the polymer, significant microstructural information about the copolymer is provided. If a copolymeric material is a block copolymer, or a blend of polymers, there are large sections of homopolymer in the system and the oligomeric fragments will be made of only one monomer. If the system is a random copolymer in which two or more monomers are joined together in the same polymer molecule, the oligomeric fragments will show a random mix of monomers, in addition to the homogeneous oligomers. An example of this may be seen in polyethylene–polypropylene copolymers. If the copolymer is a block copolymer, there are sections of polyethylene and sections of polypropylene. When the sample is pyrolyzed, the oligomers seen in the analysis, containing for example 3–10 monomeric units, will be the same as those seen in the pyrolysis of a homopolymer. Figure 1 shows a pyrogram of polyethylene and Figure 3 shows polypropylene, each revealing a repeating pattern of oligomeric peaks, but different from each other. A pyrogram of a block copolymer will look like Figure 1 superimposed on Figure 3, since all the oligomers of both polyethylene and polypropylene are present. The relative amounts of the two sets of oligomers depend directly on the relative amounts of the two monomers used. Figure 5 is a pyrogram of an ethylene (25%)–propylene block copolymer, in which oligomers seen in both Figures 1 and 3

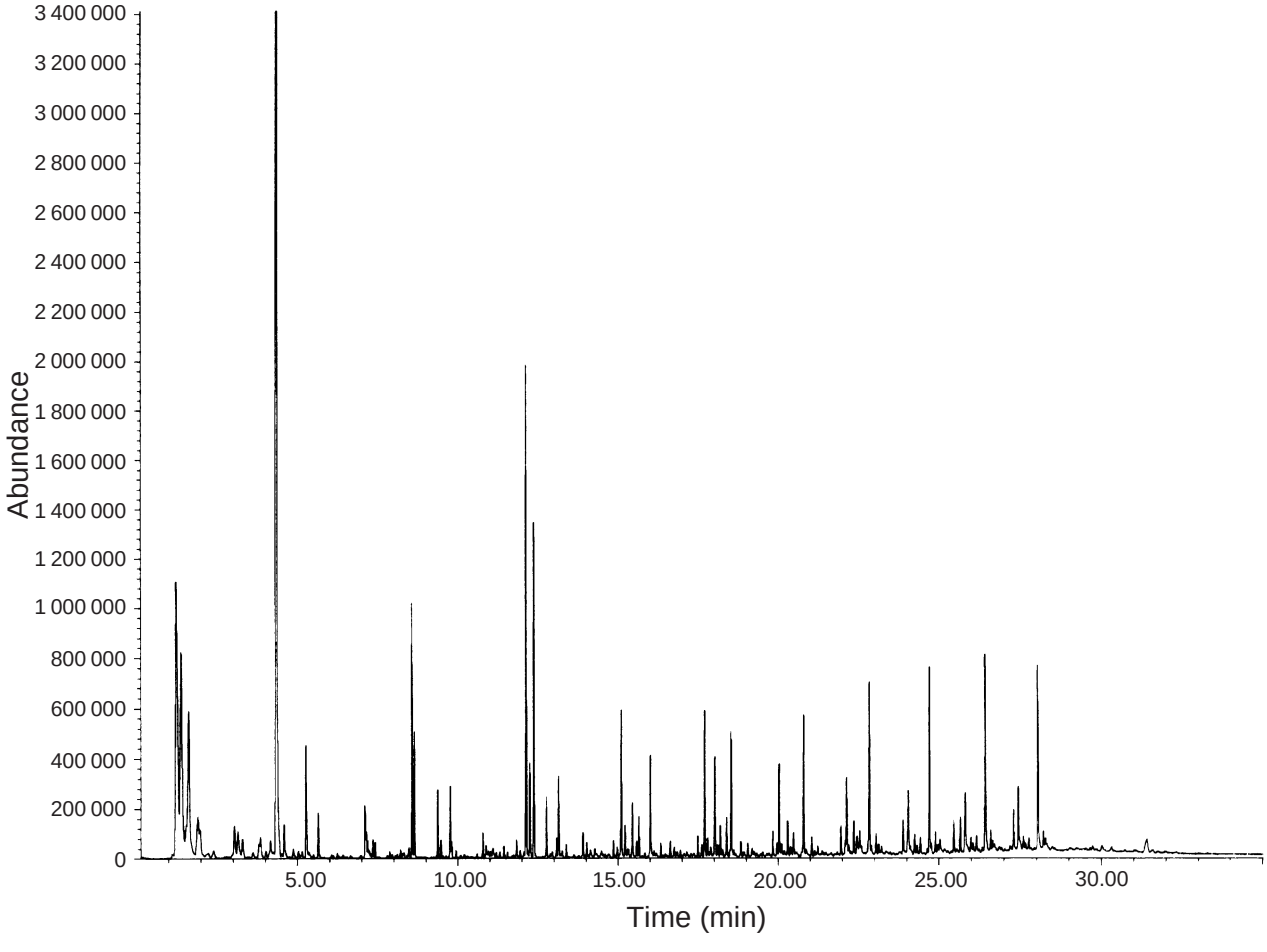

Figure 3 Pyrolysis of isotactic polypropylene.

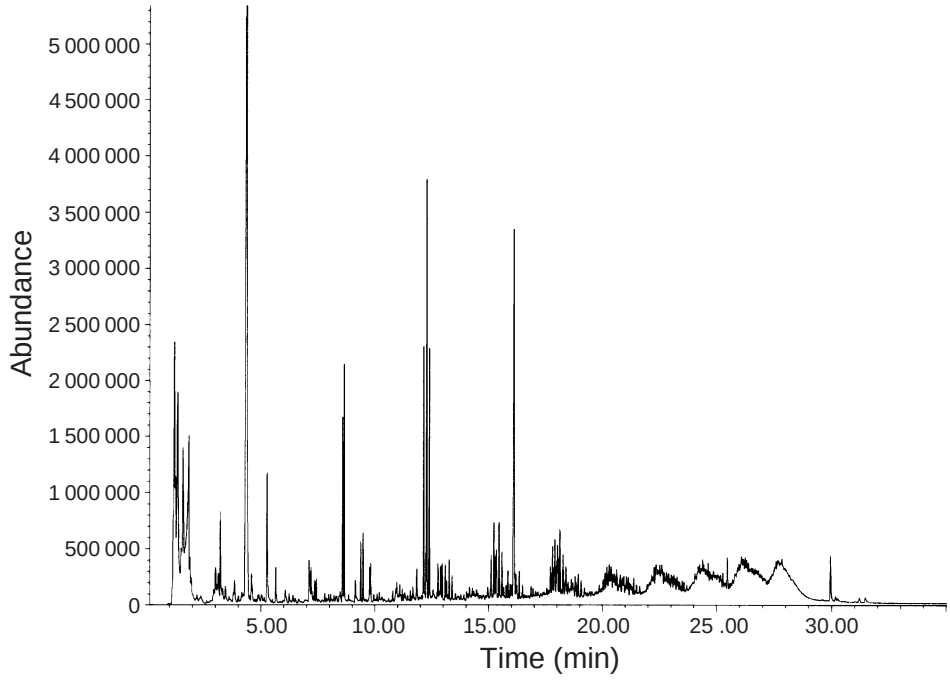

Figure 4 Pyrogram of atactic polypropylene.

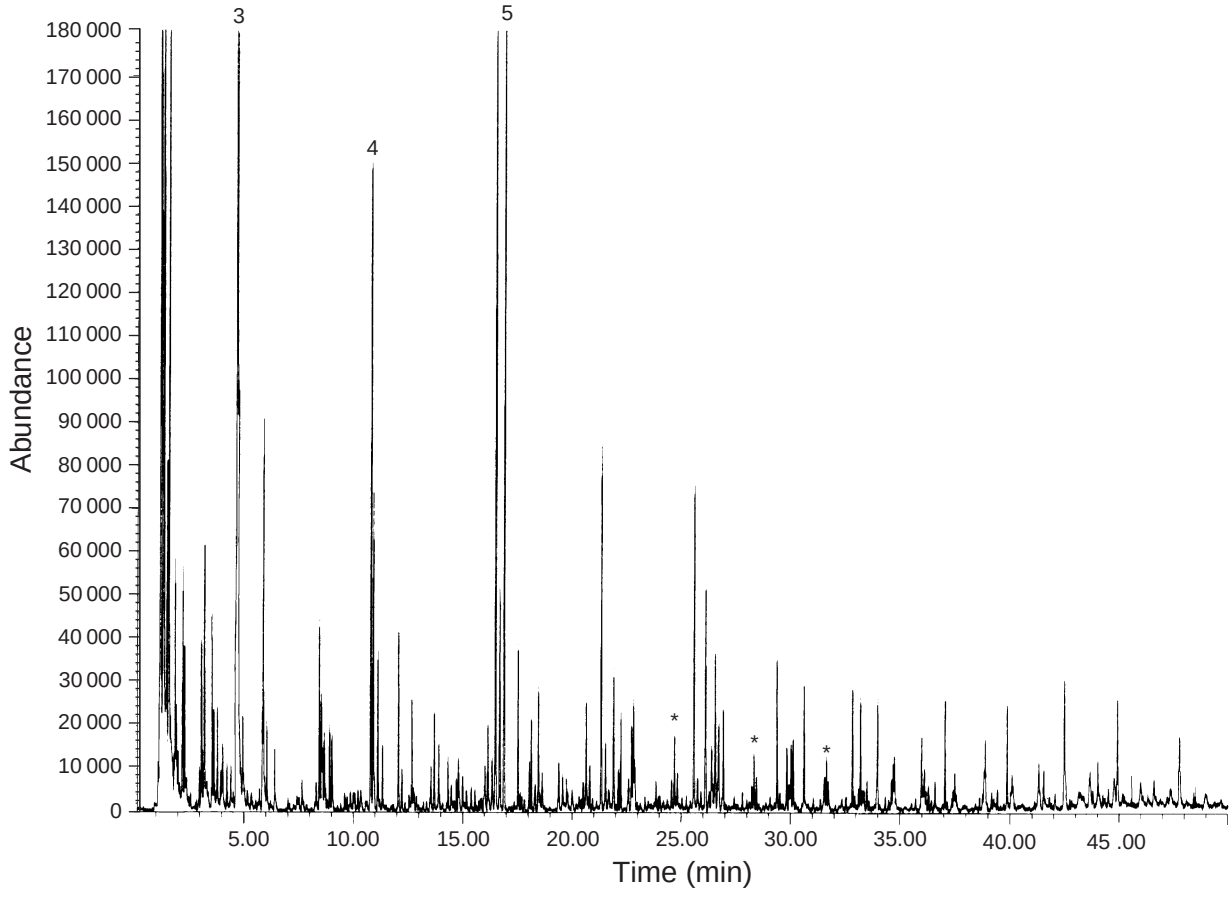

Figure 5 Pyrogram of ethylene–propylene block copolymer.

may be seen. Since the sample is 75% polypropylene, it looks more like Figure 3, and the trimer, tetramer and pentamer of propylene have been marked 3, 4 and 5 respectively. Nevertheless, the triplet normal hydrocarbon peaks from polyethylene are clearly seen throughout the chromatogram, especially after 20 min where they can be seen eluting between the larger propylene oligomers. Three of the polyethylene oligomers have been marked with asterisks to help identify them.

Ethylene–propylene rubber, however, is a random copolymer, so the oligomeric fragments will contain both ethylene and propylene monomers, in a wide and random array of positioning. Consequently, the pyrogram produced from the rubber will look nothing like that from the homopolymers or from the block copolymer. Figure 6 shows a pyrogram for ethylene–propylene rubber, demonstrating the large number of peaks detected at each chain length, because of the varied number and position of the methyl groups introduced by the propylene monomer.

All of this can be seen more easily by expanding the chromatograms of the different homopolymers and copolymers discussed above. In Figure 7(a–d) the region between 20 and 30 min for each pyrogram is represented. Pyrogram (a) shows just polypropylene, and the two sets of oligomers in this region are the hexamers and heptamers of propylene (containing 18–21 carbons). Pyrogram (c) is just polyethylene, and the normal hydrocarbons shown contain 15, 16, 17, 18 and 19 carbons. Again, the triplets arise as the diene, alkene and alkane of each carbon number, eluting in that order. Between these two pyrograms is the analysis for the block copolymer, pyrogram (b). This pyrogram shows the hexamers and heptamers of polypropylene, and also the normal hydrocarbons from polyethylene (the C_{17} and C_{19} triplets are the easiest to see). From this region, then, it is clear that at least seven propylene units must have been adjacent to each other in the original polymer, and at least 10 ethylene units, but, referring to Figure 5, shows evidence of considerably longer sections. Pyrogram (d) is from the random copolymer, showing at least 10 different oligomers for each carbon number and many peaks from mixed oligomeric fragments which are clearly absent from the pyrograms of the homopolymers and the block copolymer.

3 INSTRUMENTATION

Although pyrolysis equipment has been designed with a capacity to handle very large samples, for applications such as recycling scrap polymers, analytical utility generally depends on the ability to heat a small sample very rapidly, generally to a relatively high temperature. This is particularly true for GC interfacing, since the pyrolysis constitutes sample introduction. Slow heating or pyrolysis at low temperatures would introduce the sample in a broad band, not compatible with good chromatography. Consequently, most product development has centered around devising a means to heat a sample very rapidly but still maintain control of the heating

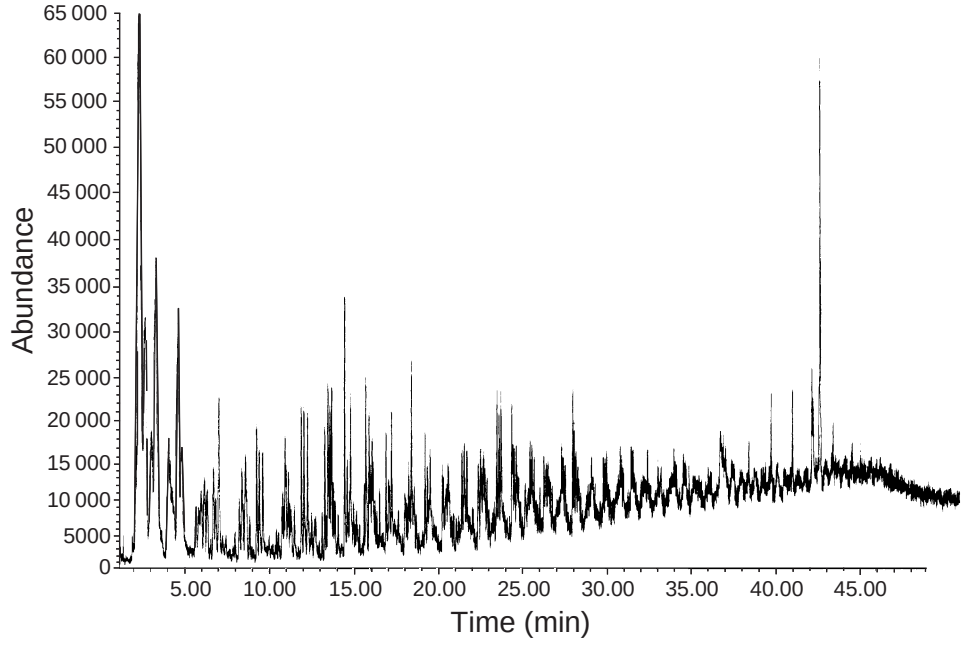

Figure 6 Pyrogram of ethylene–propylene rubber (random copolymer).

rate, temperature and time. Most commercially available instruments achieve this by one of three means: inductive heating, resistive heating or constant (isothermal) heating. Some applications require the ability to heat at slow rates, however, and adaptations of standard instruments supply this capability.

3.1 Constant Heating: Isothermal Microfurnaces

One of the easiest and least expensive ways to heat a sample to a preset temperature is to introduce it into a zone which is already heated to that temperature. The heating rate of the sample then is just the rate at which

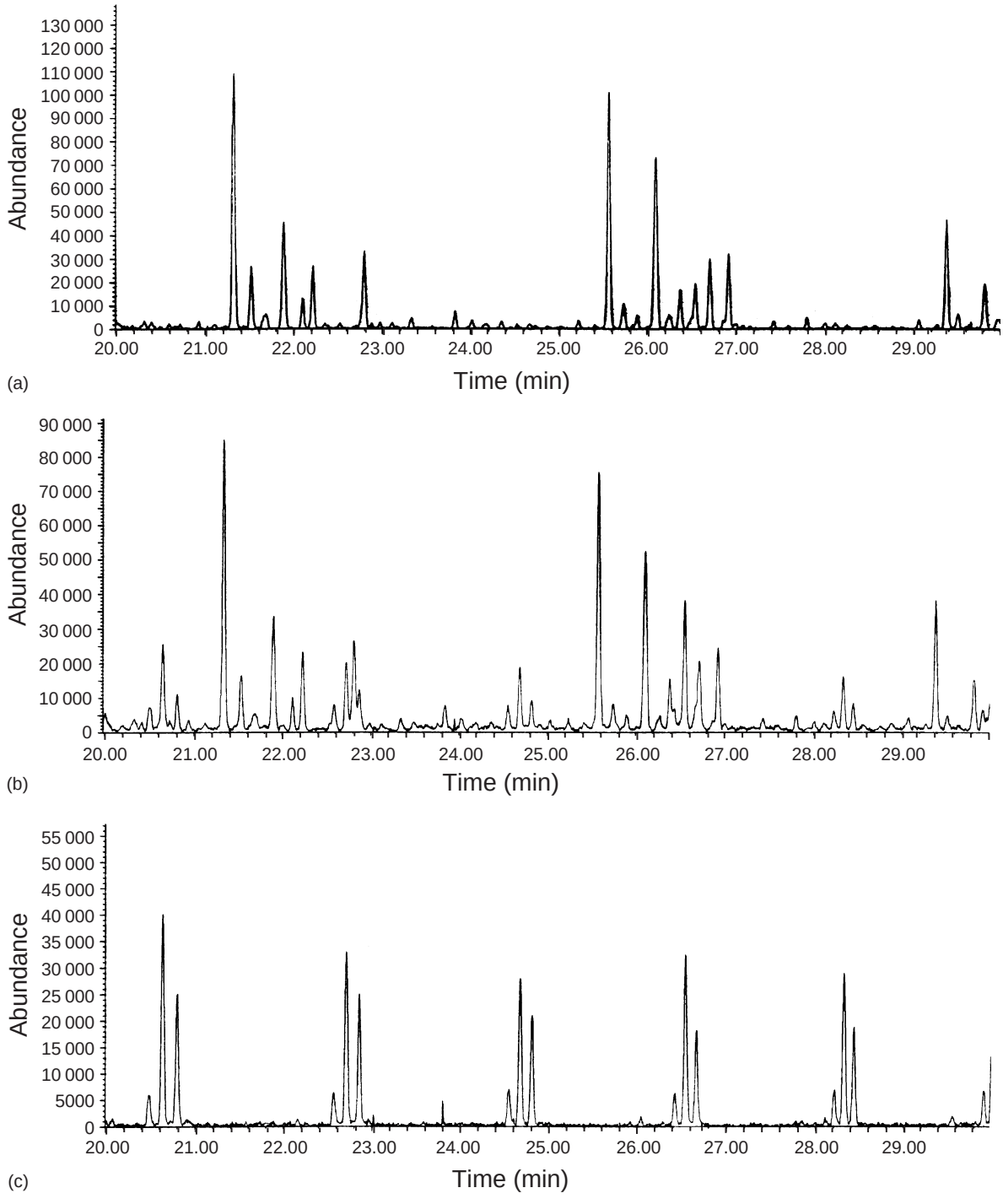

Figure 7 Expanded pyrogram sections for (a) polypropylene, (b) block copolymer, (c) polyethylene and (d) ethylene–propylene rubber.

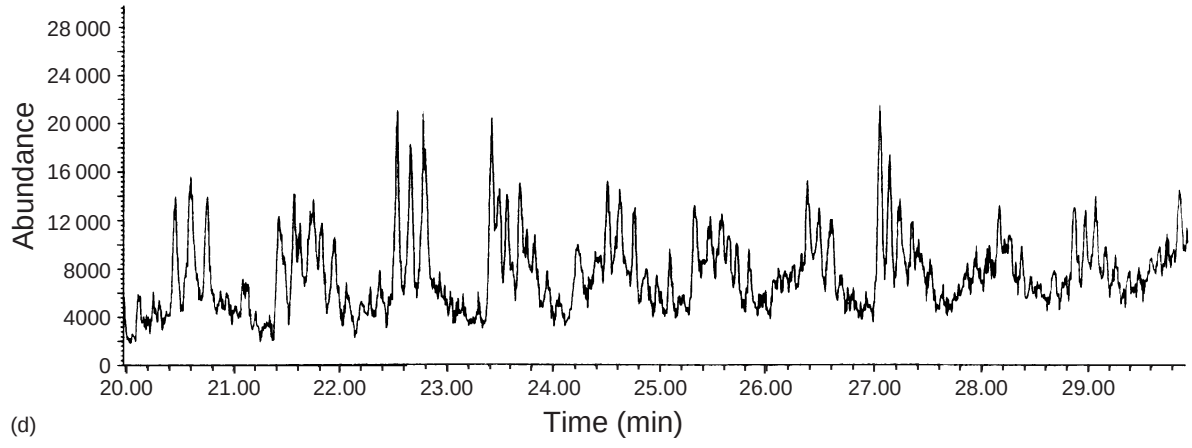

Figure 7 (*Continued*)

the material heats to the new ambient temperature, which is in part dependent on the sample material and mass. Most commercially available microfurnaces are capable of heating the pyrolysis zone to 800–900 °C, which is more than enough to pyrolyze most polymers. Among the advantages of microfurnaces are the relatively low cost, simplicity, stability and ease of use. Disadvantages include a relatively large heated surface area which promotes secondary pyrolysis reactions and difficulty in introducing some types of samples conveniently. Some furnaces permit sample introduction through a septum port only, which means that the sample must be capable of being delivered by a syringe, or plunger-type solids syringe.[33] Other designs incorporate a cooled zone with a sample basket which may be lowered into the heated zone for pyrolysis.

3.2 Inductive Heating: Curie Point

A second way to heat a sample rapidly is to place it on a cool material which may be brought to an elevated temperature quickly. Curie point pyrolyzers are based on the principle that a current may be induced into a ferromagnetic material by use of a high frequency coil. The ferromagnetic metal will be heated rapidly by the current induced in it, until it reaches a temperature at which it is no longer ferromagnetic. At this point, the Curie point for that metal, the heating stops and the temperature stabilizes. The Curie point is different for each ferromagnetic metal or alloy of those metals, so by creating different alloys of iron, cobalt and nickel a range of temperatures may be obtained. As long as the same metal is used each time, the temperature is reproducible. The pyrolysis filament, to which the sample is applied, is usually configured as a wire or thin foil so that it may heat rapidly, and elevated temperatures may be achieved in less than 1 s.

Advantages of the Curie-point approach include the ability to apply the sample to a cool surface then heat it rapidly, the ''self-limiting'' nature of the temperature control, that is, the temperature is not set by the instrument but rather is a function of the metal used, the ability to use a fresh wire for each sample and the availability of autosamplers.[34] Disadvantages include being limited in temperature selection to the alloys available and having no control over the rate of heating.

3.3 Resistive Heating: Filaments

Resistively heated filament pyrolyzers are similar to Curie-point instruments in that the sample is applied to a metal filament cool, then rapidly heated. The filament is made of a resistive material, generally platinum, which may be heated to any temperature up to about 1400 °C. Temperature control is independent of the filament metal, that is, unlike Curie-point instruments, the filament may be used over a wide range of temperatures, not just one. The temperature and heating rate are controllable electronically, so the filament may be heated ballistically as in a Curie-point instrument, but may also be heated more slowly. Since the temperature is not self-limiting as in a Curie-point instrument, it must be monitored, which may be done by optical pyrometry, with a thermocouple or photodiode or by measuring the resistance of the filament.

Advantages of the resistive filament include the ability to use the same filament at a variety of temperatures and control of both the heating rate and temperature. Since the filament may be configured as a coil to accept a small sample tube, sample placement is more convenient for materials such as powders and fibers. As with Curie-point instruments, automated sample introduction is available. Disadvantages of filament pyrolysis include the more complex temperature measurement and effects of

permanent filament mounting. Since the filament must be attached to the electronic controller, replacement of the heating wire or foil is more involved than for a Curie-point instrument.

3.4 Other Heating Techniques

Some specialized heating techniques have also been applied to pyrolysis instrumentation, which depart from the considerations above. One is the programmable furnace, which is not used isothermally as the isothermal microfurnace. Instead, the sample is placed in a cool furnace, then heated gradually to pyrolysis temperatures, with the production or collection of products along the way. Some devices have been made specially for this purpose, but many analysts have obtained similar results by interfacing an instrument for thermogravimetric analysis (TGA) to a mass or infrared spectrometer. A second application involves the use of lasers to heat the sample, which must frequently be applied to or mixed with an adsorbing material to use the laser energy. Lasers are capable of heating a sample very rapidly, but the difficulty in determining the pyrolysis temperature has limited acceptance of the technique.

4 INTERFACING

Analytically, pyrolysis is almost always used as a sample introduction technique, to extend the use of instruments not normally capable of handling polymers. This means that pyrolysis techniques are generally coupled, and the pyrolyzer must be interfaced to the analytical device. By far the most common interfacing is of a pyrolyzer to a gas chromatograph, but direct interfacing to mass spectrometers and FTIR spectrometers is also common.

4.1 Pyrolysis/Gas Chromatography

Regardless of the detector [MS, flame ionization detector (FID), thermal conductivity detector (TCD) or other] Py/GC first involves transfer of the compounds made by pyrolysis to the inlet of the GC column. Once on the column, the constituents of the pyrolyzate should be handled in the same way as any other mix of materials. There are several important considerations in interfacing a pyrolyzer to a gas chromatograph in a way to ensure that the latter can perform properly. The sample should be created rapidly, transferred rapidly through as short a transfer as possible, with good thermal integrity between the two devices. Resolution of the early eluting peaks will be affected adversely by slow product formation or transfer, causing band broadening. Since pyrolysis may create some relatively large molecules (by GC standards), cold spots between the pyrolyzer and GC column will result in either broad peaks or even the loss of analytes, which may be significant to the analysis.

The most important consideration in connecting a pyrolyzer to a gas chromatograph is the recognition that it is an entirely different means of sample introduction. GC inlets are designed to accept a sample injected in solution using a syringe which is inserted into the GC pneumatics. With pyrolysis, a sample is created upstream of the column and delivered to the injection port along with the carrier gas, with the constituents already in the vapor phase. This is why the rapid pyrolysis, transfer and heated connections are emphasized – anything which retards the delivery of the pyrolyzate volatiles as they are delivered to the GC inlet will adversely affect the chromatography. Further, it is important to understand the pnuematics of the injection port to be sure that the analytes are being delivered to the column and not vented, diluted or adsorbed. In general, all of the GC carrier flow usually introduced to the injector should be routed through the pyrolyzer, then back to the injector for efficient sweeping of the pyrolysis zone. Relatively high split ratios help transfer the pyrolyzate rapidly and prevent column overload, since the small samples compatible with microbore capillaries are frequently exceeded by analysts. In cases where slow pyrolysis rates are required for the experiment, or very low flow rates are needed for sensitivity, cryogenic focusing of the pyrolyzate on to the column before analysis compensates for the band broadening otherwise encountered.

4.2 Pyrolysis/Mass Spectrometry

Even though pyrolysis of a sample may create a complex mixture of materials, and because sometimes it does not, some analysts prefer to interface the pyrolyzer directly to a mass spectrometer,[35,36] eliminating the chromatography. The attraction comes from two considerations. First, the analysis may be done much more quickly if it does not require 30–60 min for the chromatography.[37] Second, the pyrolysis reaction is in immediate contact with the mass spectrometer, so species may be observed which are not delivered through a GC column. Further, the pyrolyzer may be heated in a programmed way and the mass spectrometer used to produce a time- or temperature-resolved analysis[38] by scanning continuously during the heating. To help simplify the data, some analysts perform chemical ionization[39] rather than electron impact ionization, but techniques employing field ionization[40] and potassium attachment[41] have also been employed. In addition to the benefits of a short analysis time and time-resolved studies, the spectra obtained may be manipulated by the software of the mass spectrometer for comparison and recognition. Although some success has been demonstrated by interfacing a pyrolyzer

designed for GC applications to the mass spectrometer using a short, inert transfer line, most applications have used specially designed solids probes capable of heating to high temperatures or other direct modifications to the mass spectrometer inlet.

4.3 Pyrolysis/Fourier Transform Infrared Spectroscopy

For many of the same reasons that analysts have interfaced pyrolyzers to mass spectrometers, systems for pyrolysis/Fourier transform infrared (Py/FTIR) spectroscopy have also been devised. These systems may involve a flow of carrier gas from the pyrolyzer into a light pipe in the FTIR instrument,[42] or the design of a specific cell to fit in the FTIR instrument which contains a pyrolysis filament.[43] The ability of the FTIR instrument to scan continuously over a temperature regime permits, as with pyrolysis/mass spectrometry (Py/MS), the display of spectra associated with a particular temperature to help elucidate the degradation pathway of a polymer.[44]

4.4 General Considerations

Regardless of the type of interfacing between the pyrolyzer and the analytical device, the sample to be pyrolyzed and the pyrolyzate it makes must be compatible with the analyzer. Many analytical instruments are designed for sensitivity and are capable of handling only very small samples. The total sample delivered from the pyrolyzer should not exceed that sample capacity. In the case of synthetic polymers, the sample is extensively degraded, so the mass of the original sample should be no more than is usually introduced into the system when not using a pyrolyzer. In GC, for example, a 1-μL injection of a solvent is substantial, representing approximately 1 mg. Polymer samples for introduction should therefore be of the same general amount, and 10–100 μg is considered optimal by many analysts. Consideration must be paid to the materials generated during pyrolysis, since they will be transferred to the analytical device directly. Polymers which produce very large oligomers may eventually contaminate the inlet, requiring more frequent cleaning or the installation of a guard to remove contaminants before they enter the system. Likewise, some polymers generate corrosive or reactive species, and the effect of those on the analytical system must be considered. PVC, for example, generates HCl, which does not preclude its analysis by GC/MS, but does advocate the use of small samples and adequate purging. Finally, since small solid samples are being used, the opportunity to contaminate them and introduce extraneous analytes is always present. When introducing 50 μg of a plastic material, the chance that it contains finger oils, adsorbed volatiles or residues from other samples must always be considered.

5 APPLICATIONS

5.1 Polyolefins

Polyolefins almost always degrade by chain-bond scission to form monomer and higher oligomers, frequently producing considerable amounts of trimer via the 1–5 hydrogen transfer mechanism. Examples of polyethylene and polypropylene are shown in Figures 1 and 3, respectively. Polyolefins made from larger monomers show the same tendency, i.e. a repeating pattern of oligomeric peaks, with the oligomeric materials becoming increasingly larger molecules. Poly(1-butene) therefore produces the pattern seen in Figure 8, in which the largest peak at 9 min is the trimer of 1-butene.

5.2 Polyamides

The synthetic polyamides nylons have been extensively studied via pyrolysis.[28] Nylons may be formed in two ways, either using a single monomer with an acid group at one end and an amine at the other, or by using two monomers, a dicarboxylic acid and a diamine. The single-monomer nylons are given a single number, such as nylon 6, indicating that the monomer has six carbons, and the two-monomer type are given two numbers, such as nylon 6,12, indicating the number of carbons in the diamine and diacid, respectively.

When pyrolyzed, nylons generate a variety of compounds, including nitriles, dinitriles, lactams and hydrocarbons. Nylon 6, for example, produces extensive amounts of caprolactam, producing a pyrogram such as that shown in Figure 9. Nylon 6,6, although structurally similar, produces nitriles and a large peak of cyclopentanone from the C_6 diacid portion of the molecule. This cyclopentanone is a common pyrolysis fragment in nylons synthesized with the six carbon diacid, regardless of the length of the diamine. Figure 10 shows a pyrogram of nylon 6,6 for comparison with nylon 6 in Figure 9.

5.3 Polyesters

The most common polyesters are phthalic acid polymers made with polyfunctional alcohols. Many paints are glycerol phthalate polyesters, including house-paints and the automotive finishes used before PMMA and styrene copolymers were introduced. Poly(ethylene terephthalate) (PET) is the most commonly encountered polyester, comprising the bulk of polyester textile fibers, beverage and food containers and other rugged, inexpensive goods. When phthalate polyesters are pyrolyzed, the most noticeable products are derived from the phthalic acid, and include benzoic acid and other substituted aromatics. Figure 11 shows the pyrogram produced from a piece of overhead projector transparency film, which is PET,

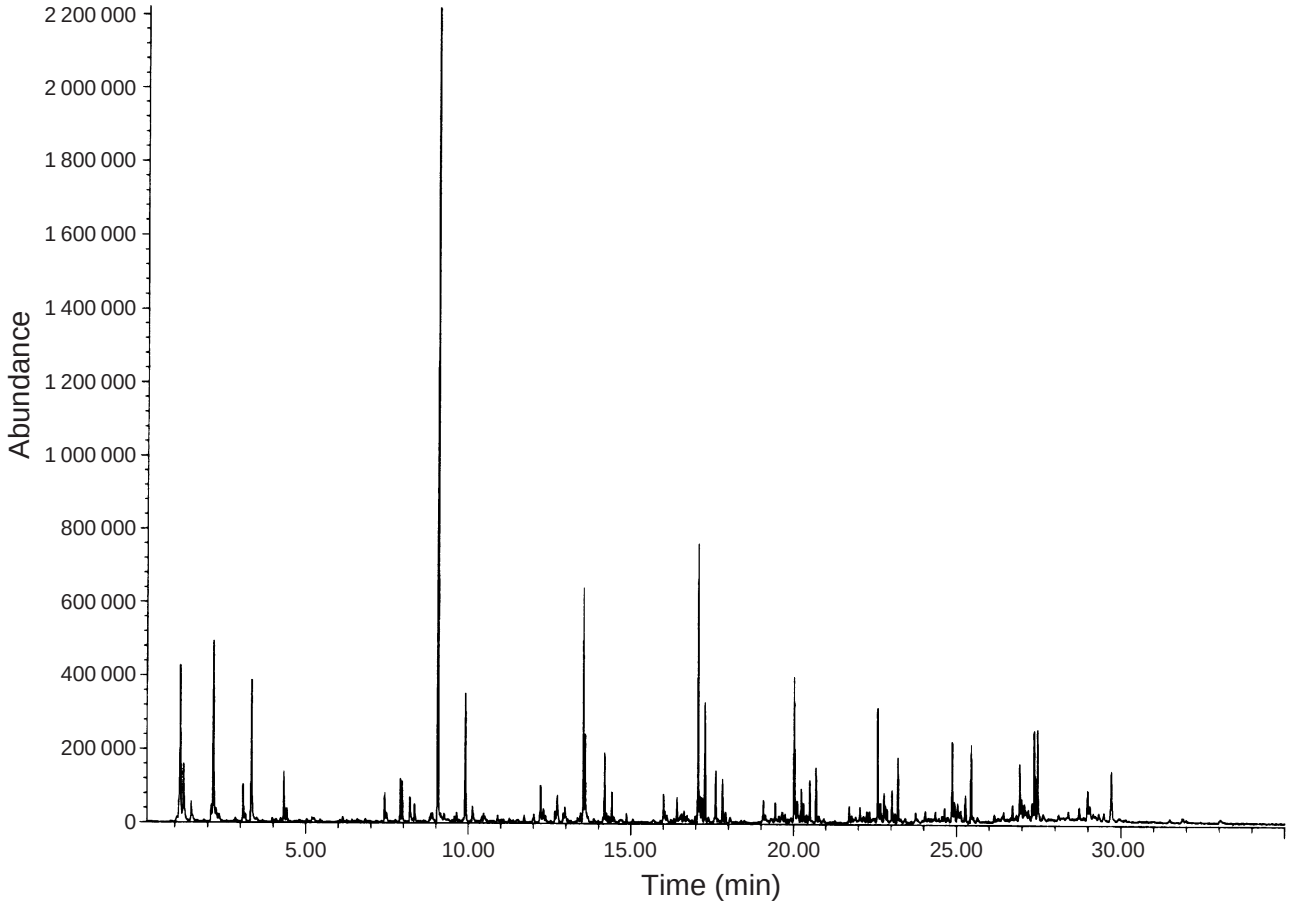

Figure 8 Pyrogram of poly(1-butene).

with the two peaks eluting at about 11 min being vinyl benzoate and benzoic acid.

5.4 Polyurethanes

Polyurethanes are produced using a di- or triisocyanate and a multifunctional alcohol. The OH providing monomer is frequently a ester or ether with two or more OH groups attached, while any of several diisocyanates is commonyl used, including toluene diisocyanate (TDI) and hexane diisocyanate. Polyurethanes are used in protective coatings and synthetic rubber, as in the foam ''rubber'' used in many furniture applications. An interesting property of polyurethanes when pyrolyzed is that they generally regenerate the original diisocyanate used in the polymerization, making the identification of a polyurethane relatively easy. The chromatograms produced from pyrolyzing polyurethanes are usually fairly simple, as in the example in Figure 12. This is a piece of polyurethane foam used for padding, and the large peak at 19 min is TDI.

5.5 Vinyl Polymers

PVC, PVDC and PVA pyrolyze to release HCl and acetic acid respectively, as discussed in section 2.2. For many applications, such as binder covers and plastic tubing, PVC is extensively plasticized, usually with a phthalate plasticizer. The clear, pliable, plastic tubing used in many laboratory applications is generally heavily plasticized PVC, and when a piece of such tubing (or similar material) is analyzed via pyrolysis, the plasticizer is usually seen in the chromatogram along with the pyrolysis products of the PVC. Since the plasticizer is probably volatile at the operating temperature of the pyrolyzer, it is usually desorbed intact before pyrolysis takes place, and the resulting chromatogram looks like Figure 13. The early eluting peaks here include HCl (at 2 min), benzene, toluene and naphthalene, all indications of PVC, while the large peak at 26 min is bis(2-ethylhexyl) phthalate.

5.6 Acrylics

Since the polyacrylates can undergo a 1–5 hydrogen transfer and form trimers, and the polymethacrylates

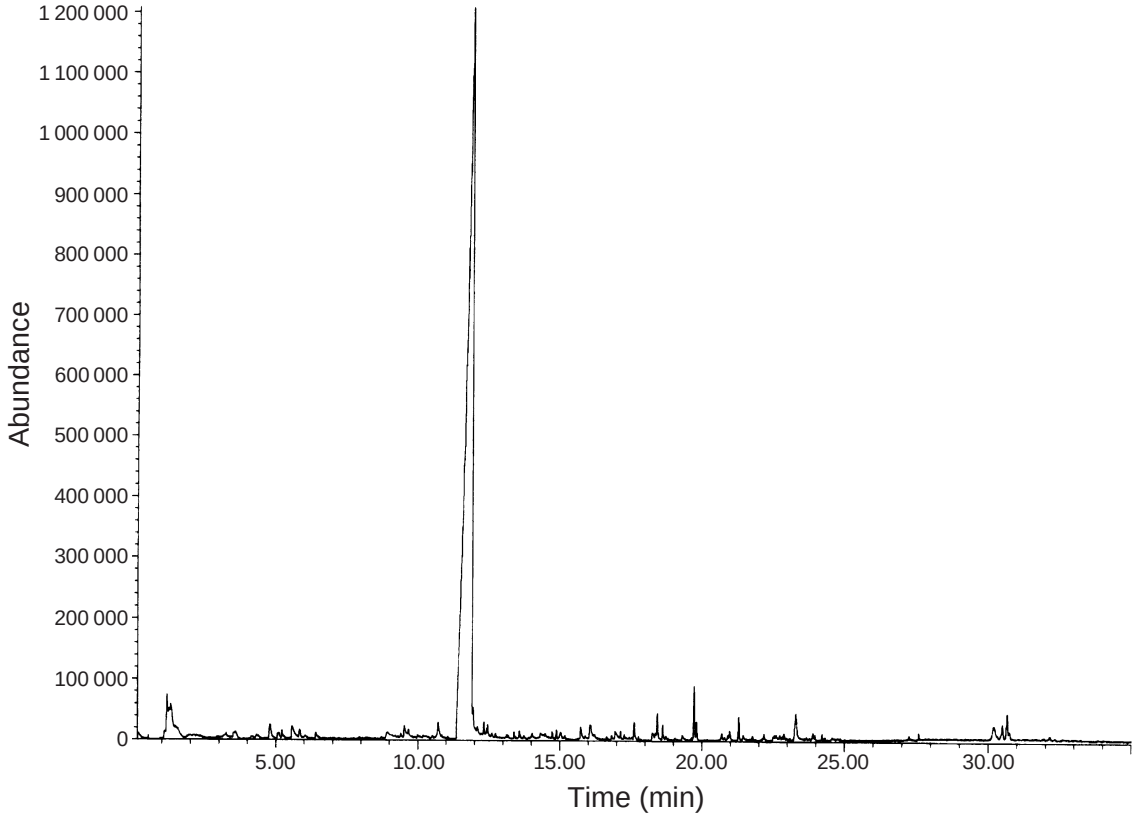

Figure 9 Pyrogram of nylon 6. The large peak is caprolactam (monomer).

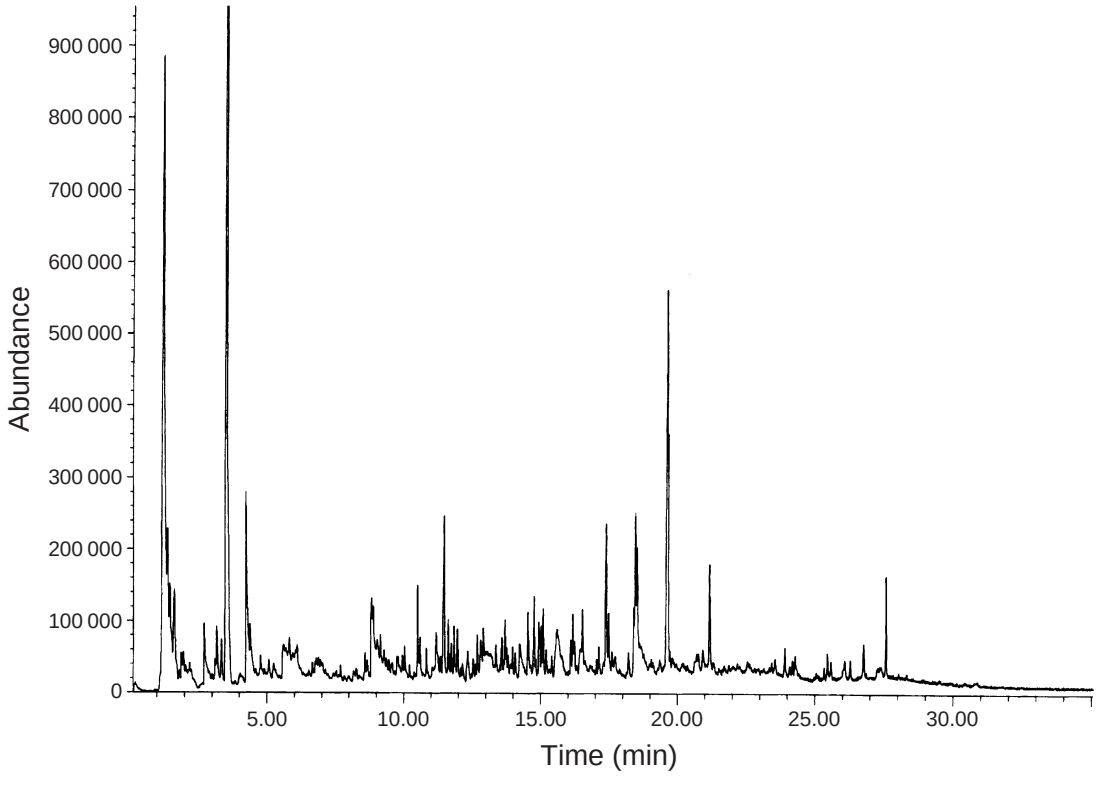

Figure 10 Pyrogram of nylon 6,6. The large peak at 4 min is cyclopentanone.

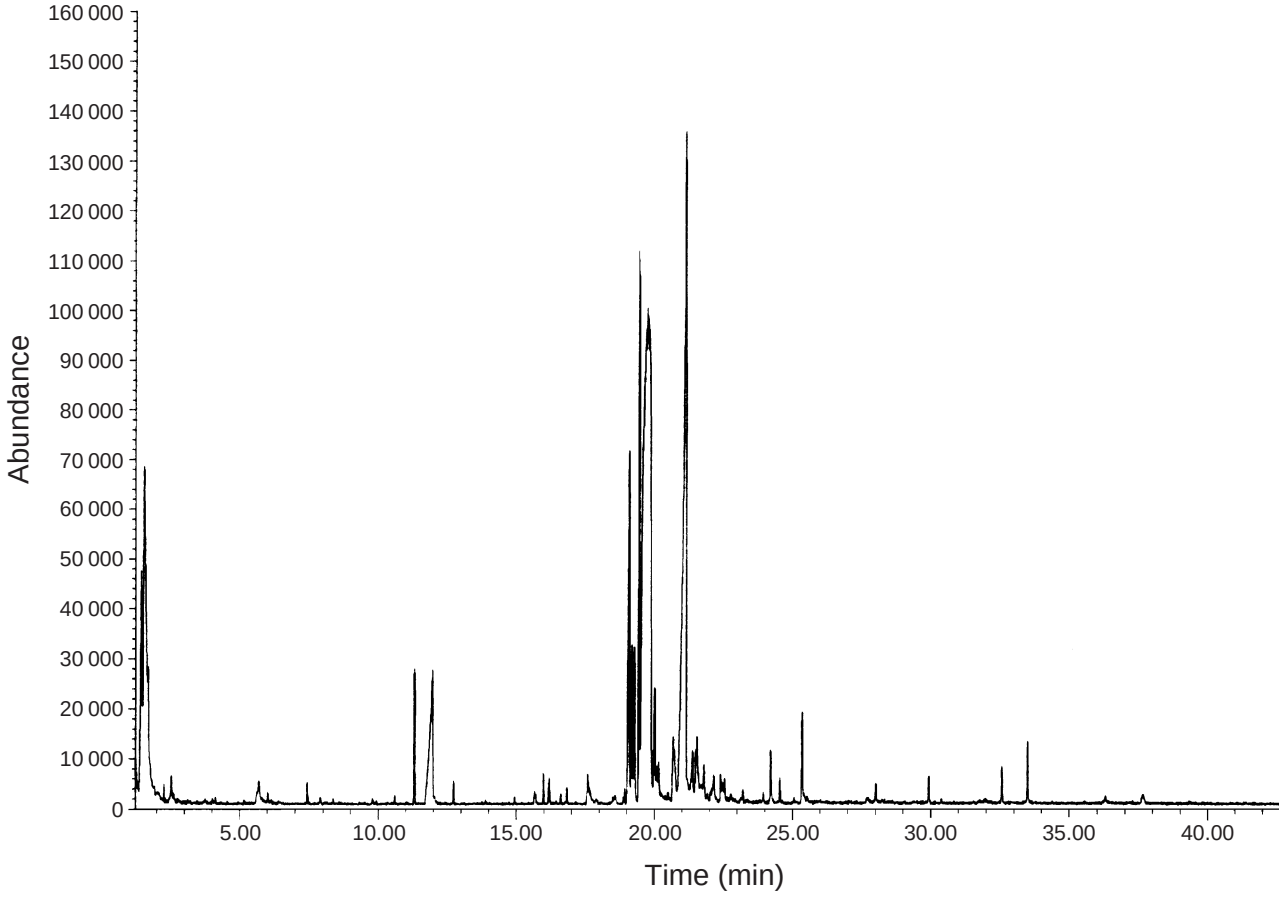

Figure 11 Pyrogram of a piece of PET transparency film.

cannot (section 2.1), the overall pyrograms of these related polymers are easily distinguished. A pyrogram of PMMA is shown in Figure 2, with MMA monomer making the large peak at 3 min. Materials made essentially of PMMA, including Plexiglas™ and countertops such as Corian™, will produce a pyrogram which closely resembles Figure 2. Similarly, poly(ethyl methacrylate) will produce mostly ethyl methacrylate monomer, and so on. Poly(butyl methacrylate) will produce a large peak for monomer, but poly(butyl acrylate) will generate monomer, dimer and trimer. Figure 14 is a pyrogram of poly(butyl acrylate), showing the monomer at 5 min and a large peak for the trimer at 24 min.

5.7 Copolymers

Since most polymeric goods are not simple homopolymers, the utility of pyrolysis is measured by the information it provides on copolymeric systems. Figures 5–7 show both the ability to identify the monomers and the relative amounts of monomers in a ethylene–propylene copolymer, and to distinguish random from block copolymers. The same is true for other copolymers of olefins, and also for the myriad of copolymers using dissimilar monomers. Synthetic rubbers, for example, may be copolymers of styrene, butadiene, acrylonitrile, isoprene, chloroprene and many other monomers. Even in the presence of fillers such as carbon black,[45] however, information is readily generated as to the monomeric composition of these products. Figure 15 is the pyrogram of a piece of automobile tire rubber, extensively filled with carbon black. The rubber portion is a copolymer of butadiene and isoprene, and both monomers and dimers for each can be seen. The two large peaks which elute first (just before 2 min) are butadiene and isoprene, respectively, then the butadiene dimer elutes at 6 min, with the tall peak at 11 min being the isoprene dimer. Many other natural and synthetic rubber formulations have been analyzed,[46] including styrene–butadiene (SBR), styrene–isoprene and acrylonitrile–butadiene–styrene (ABS) copolymers.

Acrylic copolymers, with and without styrene, are used in a wide array of consumer goods from automobile paint to photocopy toner. Since many of the monomers (the methacrylates) produce mostly monomer whereas others (styrene, the acrylates) also produce dimers, trimers and higher oligomers, the product distribution

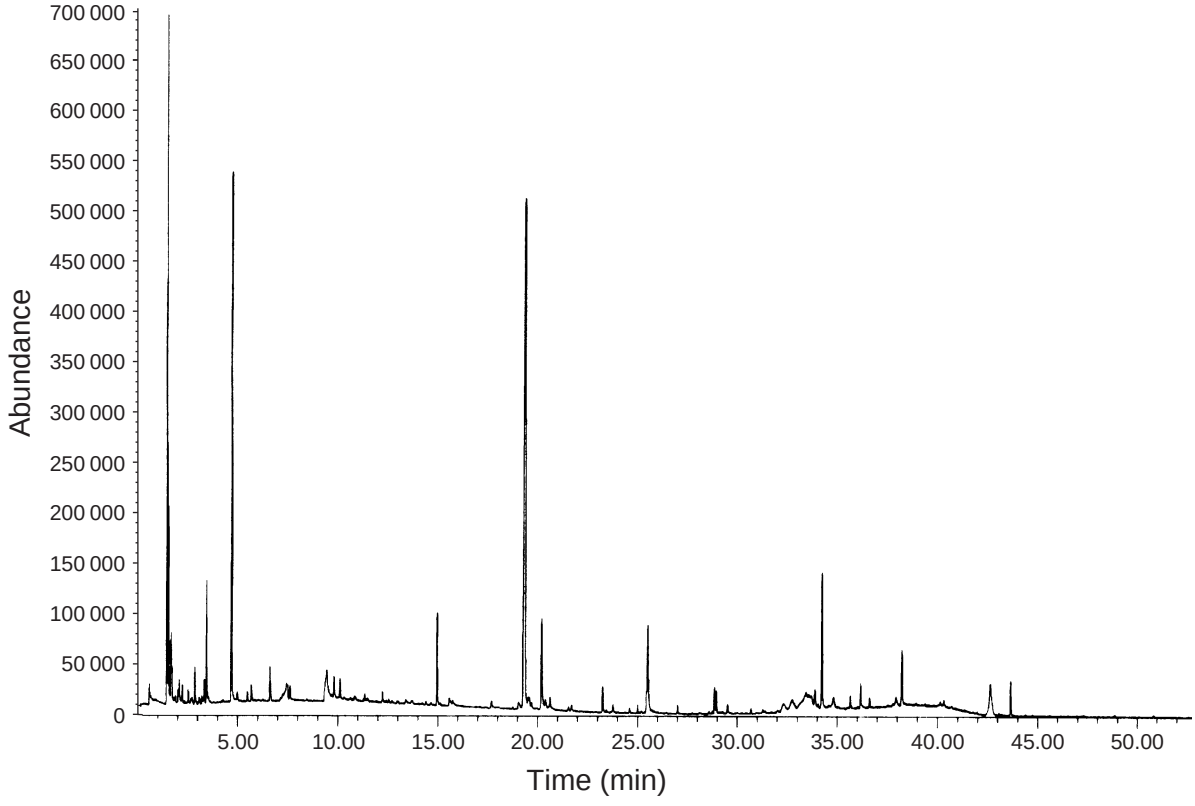

Figure 12 Pyrogram of ''foam rubber'' (TDI-based polyurethane).

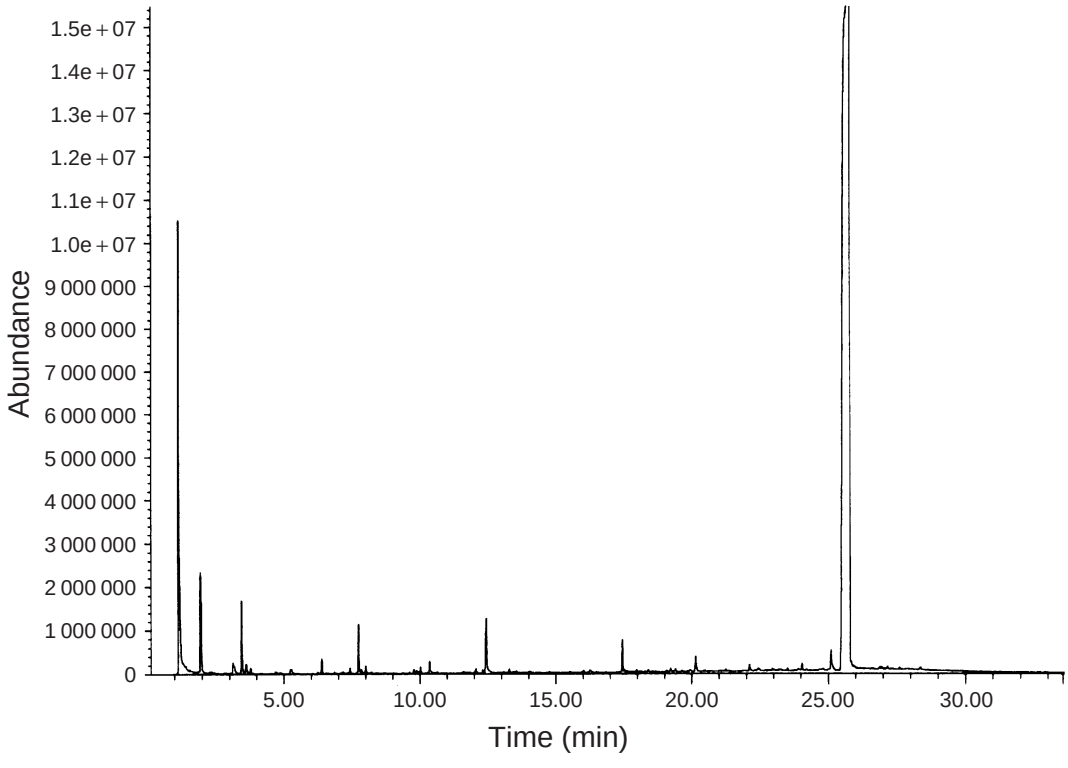

Figure 13 Pyrolysis of PVC with plasticizer. The peak at 26 min is bis(2-ethylhexyl) phthalate.

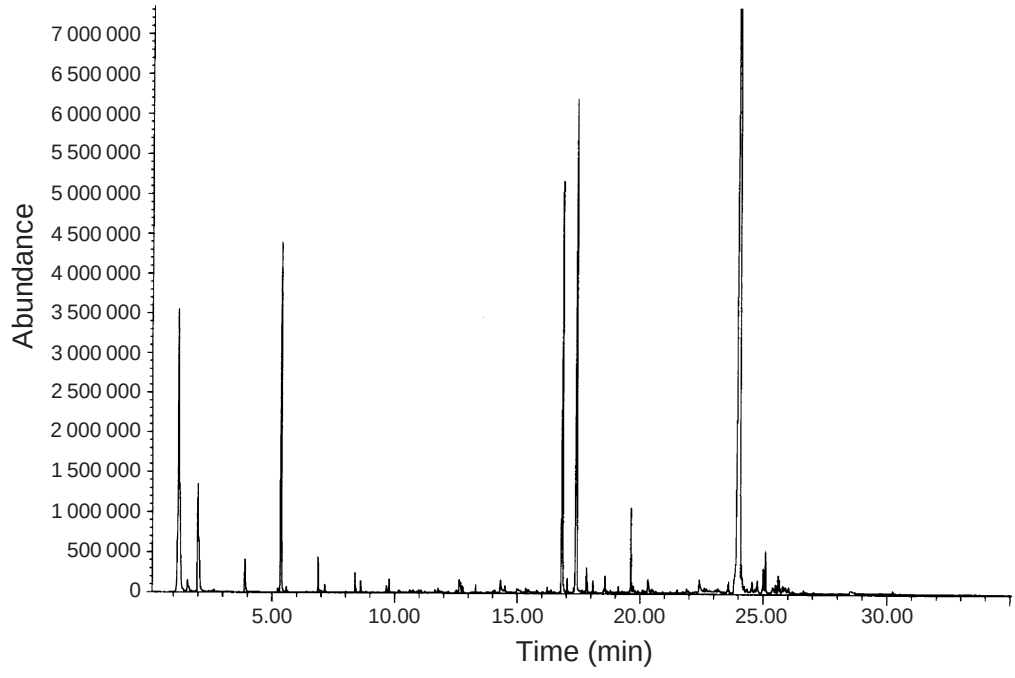

Figure 14 Pyrogram of poly(butyl acrylate). The large peak at 24 min is the trimer of butyl acrylate.

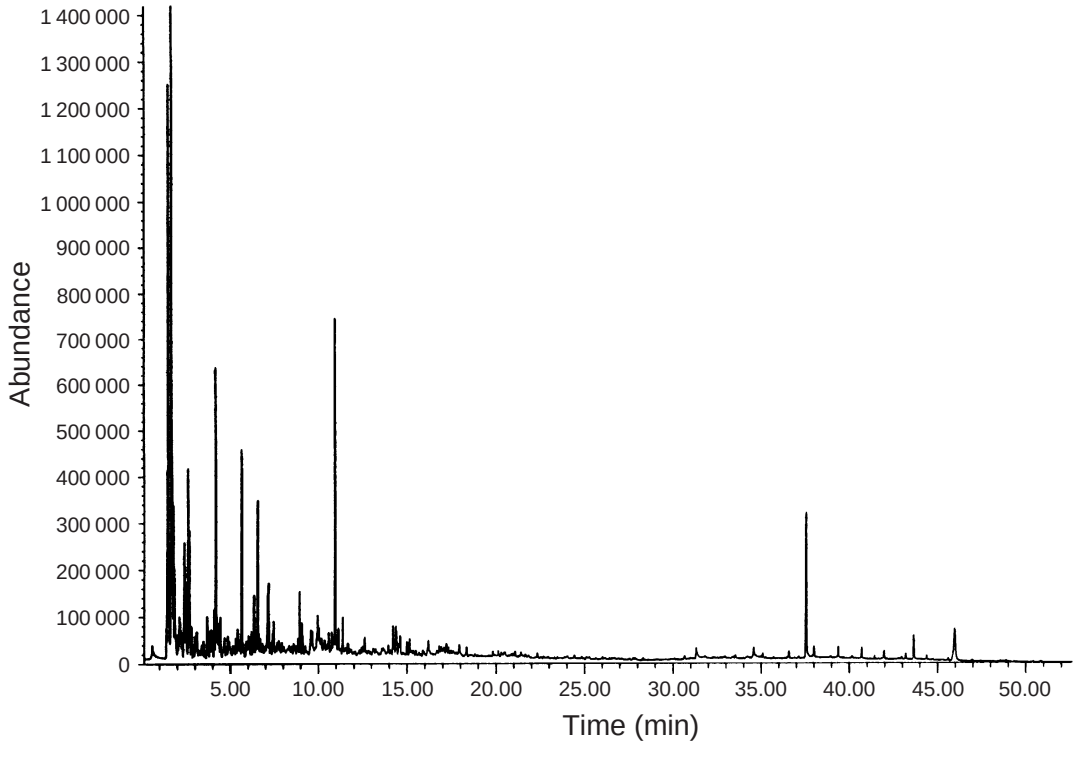

Figure 15 Pyrogram of automobile tire rubber (copolymer of butadiene and isoprene).

depends both on the monomers used[47] and on the type of copolymer (random or block).[48,49] Whether or not there are enough monomeric units together in the chain to produce homotrimers, these copolymers efficiently generate monomer peaks for each of the monomers used, providing both qualitative and quantitative information. Automobile finishes use a wide array of monomers, including styrene, MMA, ethyl acrylate, butyl acrylate,

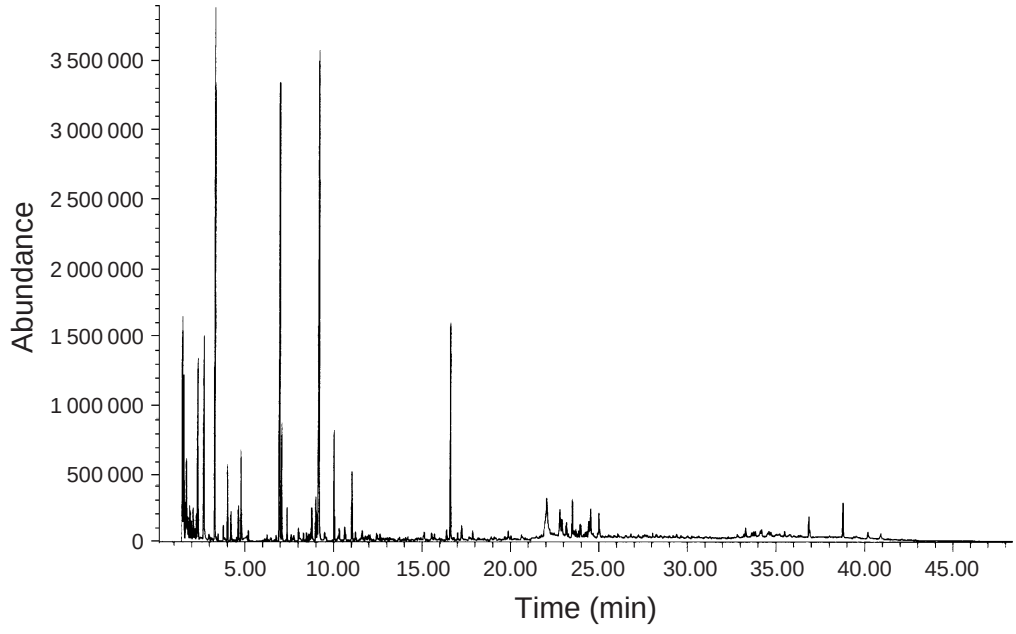

Figure 16 Pyrolysis of automotive paint formulated with MMA, styrene, butyl acrylate, butyl methacrylate and hexyl methacrylate.

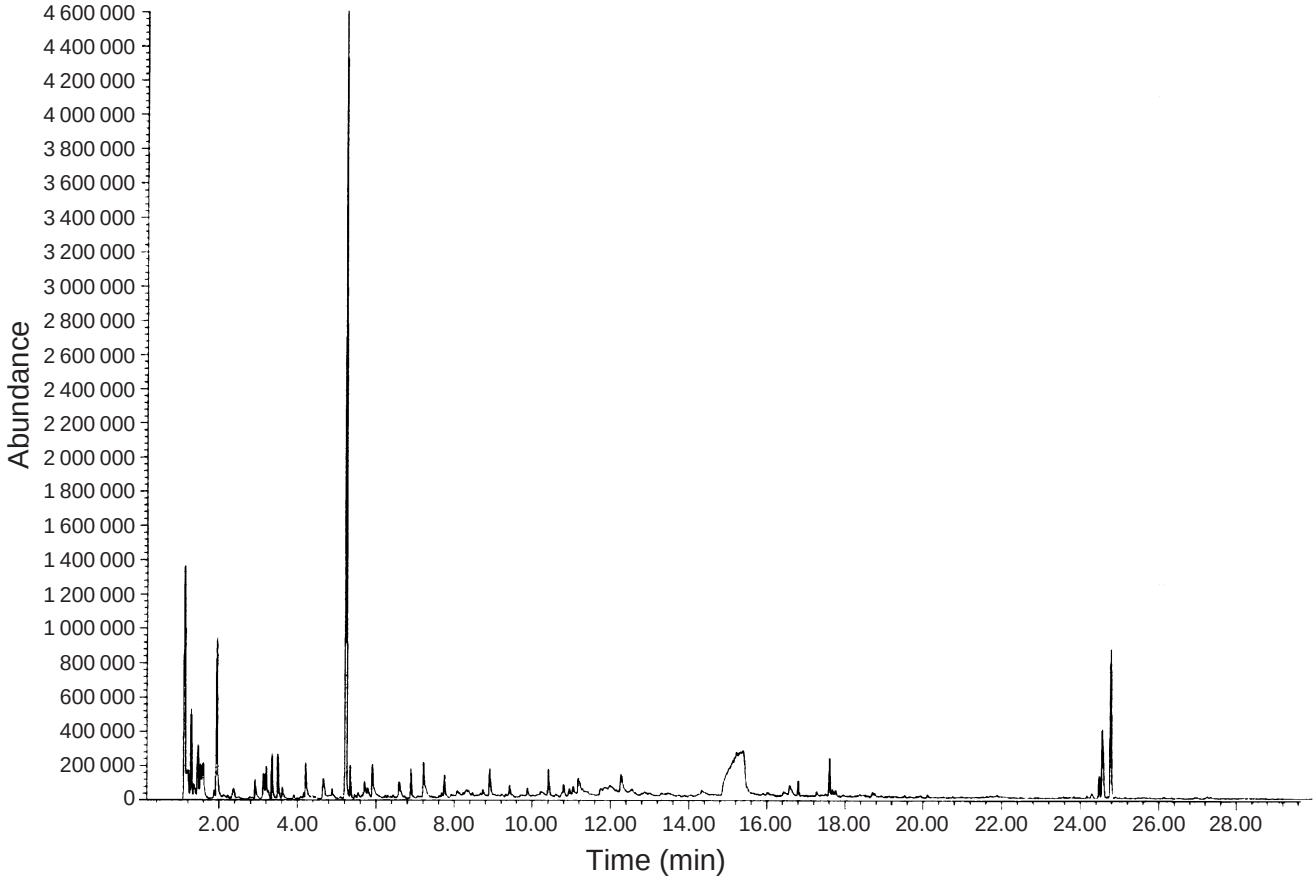

Figure 17 Pyrogram of a photocopy toner showing monomers for styrene and butyl acrylate at 5 min and trimers at 25 min.

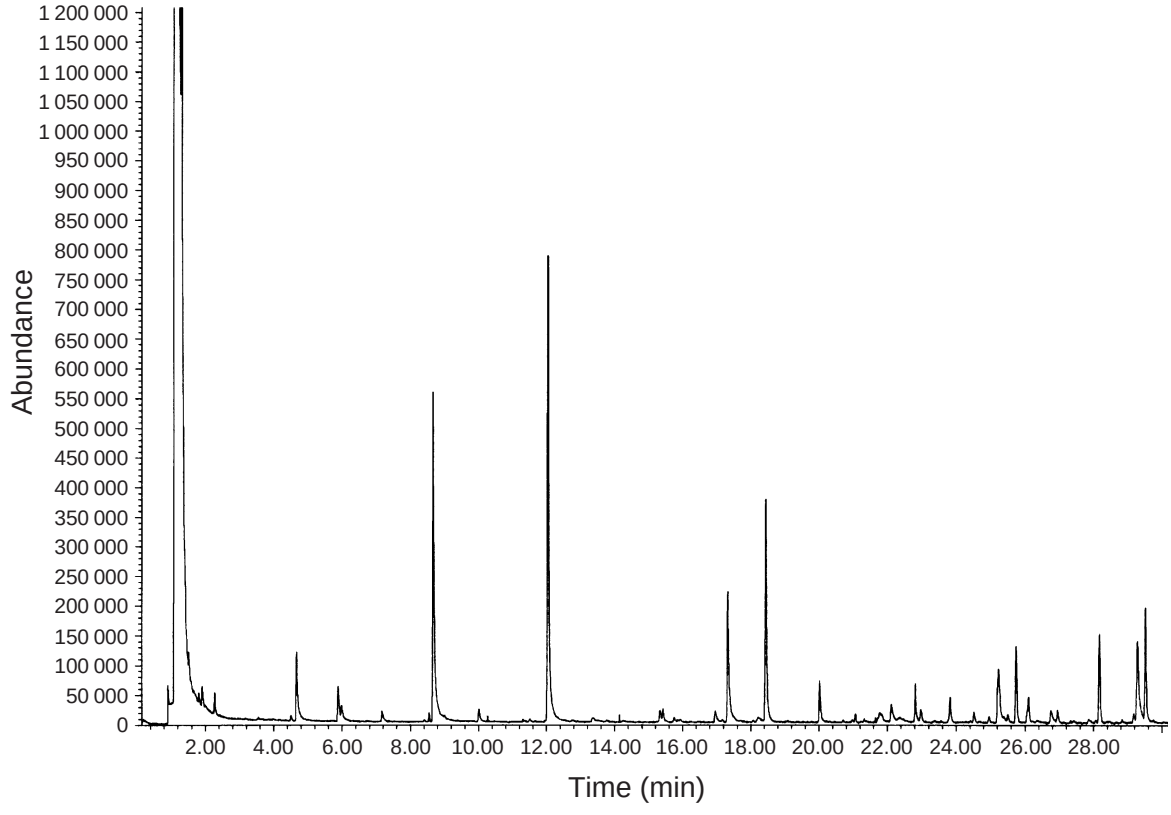

Figure 18 Pyrolysis of a vinyl chloride–vinylidene chloride copolymer.

butyl methacrylate, hydroxy-substituted methacrylates, longer chain methacrylates for flexibility and a variety cross-linking agents.[50] An example of such an automotive paint is shown in Figure 16. Monomers used in the formulation include MMA, styrene, butyl acrylate and butyl methacrylate, which elute at 3.2, 6.9, 7.0 and 9.1 min respectively. In addition, flexibility is gained by adding the monomer hexyl methacrylate, which elutes at 16.5 min.

Since both polystyrene and poly(butyl acrylate) pyrolyze to give monomer, dimer and trimer, styrene–butyl acrylate copolymers should produce these oligomers, and a random copolymer will produce mixed oligomers. The relative abundances of the various homo- and co-oligomers depend on the relative amounts of the two monomers in the copolymer. An example of this is seen in Figure 17, which shows an analysis of a photocopy toner, a copolymer which is mostly styrene with some butyl acrylate. The sytrene monomer and butyl acrylate monomer elute next to each other at about 5 min, with the large styrene peak eluting first and the butyl acrylate the small peak immediately after it. At about 25 min there are three peaks which comprise trimers for the system, the largest of which is the styrene homotrimer, since the amount of styrene in the polymer makes it very likely that there will be three styrenes adjacent to one another to form the trimer.

Vinyl copolymers[51] have also been studied. As discussed in section 2.2, PVC degrades to make HCl and aromatics, including benzene, whereas PVDC gives chiefly HCl and trichlorobenzene. The copolymer of vinyl chloride and vinylidene chloride shown in Figure 18 can be seen to contain more vinylidene chloride than vinyl chloride because of the relative amounts of the primary degradation products. There is almost no benzene (2 min), some chlorobenzene (4.8 min), considerable dichlorobenzene (8.8 min), but mostly trichlorobenzene (12 min), the primary product from PVDC.

In general, then, almost any polymeric material may be characterized using a combination of pyrolysis and a routine analytical device. The availability of automated pyrolysis equipment and inexpensive mass spectrometers makes this an attractive combination for the characterization of polymers, whether new formulations or applications of materials which have been in the marketplace for decades.

ABBREVIATIONS AND ACRONYMS

ABS	Acrylonitrile–Butadiene–Styrene
FID	Flame Ionization Detector

FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
MMA	Methyl Methacrylate
MS	Mass Spectrometry
PET	Poly(ethylene terephthalate)
PMMA	Poly(methyl methacrylate)
PVA	Poly(vinyl acetate)
PVC	Poly(vinyl chloride)
PVDC	Poly(vinylidene chloride)
Py/FTIR	Pyrolysis/Fourier Transform Infrared
Py/GC	Pyrolysis/Gas Chromatography
Py/MS	Pyrolysis/Mass Spectrometry
SBR	Styrene–Butadiene
TCD	Thermal Conductivity Detector
TDI	Toluene Diisocyanate
TGA	Thermogravimetric Analysis

RELATED ARTICLES

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Forensic Science **(Volume 5)**
Pyrolysis Gas Chromatography in Forensic Science

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Thermogravimetry of Polymers

Pulp and Paper **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses

REFERENCES

1. P.C. Uden, 'Nomenclature and Terminology for Analytical Pyrolysis', *Pure Appl. Chem.*, **65**(11), 2405–2409 (1993).
2. C.G. Williams, 'On Isoprene and Caoutchin', *J. Chem. Soc.*, **15**, 110–125 (1862).
3. F. Paneth, A. Lautsch, 'Evidence for the Existence of Ethyl Free Radicals', *Nature (London)*, **125**, 564–576 (1930).
4. F.O. Rice, 'The Thermal Decomposition of Organic Compounds from the Standpoint of Free Radicals', *Johns Hopkins Univ. J.*, **53**, 1959–1972 (1931).
5. F.O. Rice, 'The Thermal Degradation of Organic Compounds from the Standpoint of Free Radicals. III. The Calculation of the Products formed from Paraffin Hydrocarbons', *Johns Hopkins Univ. J.*, **55**, 3035–3042 (1933).
6. F.O. Rice, K.F. Herzfeld, 'The Thermal Degradation of Organic Compounds from the Standpoint of Free Radicals. VI. The Mechanism of Some Chain Reactions', *Johns Hopkins Univ. J.*, **56**, 284–289 (1934).
7. W.J. Irwin, 'Analytical Pyrolysis–An Overview', *J. Anal. Appl. Pyrol.*, **1**, 3–25 (1979).
8. J.W. Irwin, *Analytical Pyrolysis*, Marcel Dekker, New York, 1982.
9. A.I. Keuleman, S.G. Perry, 'Identification of Hydrocarbons by Thermal Cracking', in *Gas Chromatography*, ed. M. van Swaay, Butterworths, London, 356–391, 1963.
10. J.M. Dietz, 'Application of Gas Chromatography to Hydrocarbon Analysis', in *Hydrocarbon Analysis*, ASTM STP 389, ASTM, Philadelphia, PA, 1965.
11. H. Seno, S. Tsuge, T. Takeuchi, 'Pyrolysis–Gas Chromatographic Studies on the Structure of Ethylene/Propylene Copolymers', *Makromol. Chem.*, **161**, 195–205 (1972).
12. H. Seno, S. Tsuge, T. Takeuchi, 'Estimation of Chemical Inversions of Monomer Placement in Polypropylene by Pyrolysis Gas Chromatography', *Makromol. Chem.*, **161**, 185–193 (1972).
13. G.M. Brauer, 'Pyrolysis–Gas Chromatographic Techniques for Polymer Identification', in *Techniques and Methods of Polymer Evaluation, II*, eds. P.E. Slade, L.T. Jenkins, Marcel Dekker, New York, 41–105, 1970.
14. S.A. Liebman (ed.), *Pyrolysis and GC in Polymer Analysis*, Marcel Dekker, New York, 1985.
15. T.P. Wampler (ed.), *Applied Pyrolysis Handbook*, Marcel Dekker, New York, 1995.
16. T.P. Wampler, 'A Selected Bibliography of Analytical Pyrolysis', *J. Anal. Appl. Pyrol.*, **16**, 291–322 (1989).
17. M. Blazso, 'Recent Trends in Analytical and Applied Pyrolysis of Polymers', *J. Anal. Appl. Pyrol.*, **39**, 1–25 (1997).
18. S.A. Liebman, T.P. Wampler, E.J. Levy, 'Developments in Pyrolysis Capillary GC', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **7**, 172–184 (1984).
19. J.H. Flynn, R.E. Flori, 'Degradation and Pyrolysis Mechanisms', in *Pyrolysis and GC in Polymer Analysis*, ed. S.A. Liebman, Marcel Dekker, New York, 149–206, 1985.
20. R. Lehrle, D. Atkinson, S. Cook, P. Gardner, S. Groves, R. Hancox, G. Lamb, 'Polymer Degradation Mechanisms:

New Approaches', *Polym. Degrad. Stabil.*, **42**, 281–291 (1993).
21. I. Ericsson, 'Sequential Pyrolysis Gas Chromatographic Study of the Decomposition Kinetics of *cis*-1,4-Polybutadiene', *J. Chromatogr. Sci.*, **16**, 340–344 (1978).
22. R. Lehrle, R. Peakman, J. Robb, 'Pyrolysis–Gas–Liquid Chromatography Utilized for a Kinetic Study of the Mechanism of Initiation and Termination in the Thermal Degradation of Polystyrene', *Eur. Polym. J.*, **18**, 517–529 (1982).
23. Y. Sugimura, S. Tsuge, 'Studies on Thermal Degradation of Aromatic Polyesters by Pyrolysis–Gas Chromatography', *J. Chromatogr. Sci.*, **17**, 269–272 (1979).
24. K.J. Voorhees, F.D. Hileman, I.N. Einhorn, J.H. Futrell, 'An Investigation of the Thermolysis Mechanism of Model Urethanes', *J. Polym. Sci.*, **16**, 213–228 (1978).
25. L. Hang, J. Wu, S. Xue, 'Mechanisms of Pyrolysis of Fluoropolymers', *J. Anal. Appl. Pyrol.*, **10**, 99–106 (1986).
26. M.J. Matheson, T.P. Wampler, L. Johnson, L. Atherly, L. Smucker, 'A Pyrolysis–GC/MS Library for Polymers', *Am. Lab.*, **29**(7), 24C–24F (1997).
27. G.M. Simmons, M. Gentry, 'Kinetic Formation of CO, CO_2, H_2 and Light Hydrocarbon Gases from Cellulose Pyrolysis', *J. Anal. Appl. Pyrol.*, **10**, 129–138 (1986).
28. H. Ohtani, T. Nagaya, Y. Sugimura, S. Tsuge, 'Studies on Thermal Degradation of Aliphatic Polyamides by Pyrolysis–Glass Capillary Gas Chromatography', *J. Anal. Appl. Pyrol.*, **4**, 117–131 (1982).
29. C.A. Lytle, W. Bertsch, M.D. McKinley, 'Determination of Thermal Decomposition Products from a Phenolic Urethane Resin by Pyrolysis–Gas Chromatography–Mass Spectrometry', *J. High Resolut. Chromatogr.*, **21**, 128–132 (1998).
30. S.A. Liebman, E.J. Levy, 'Advances in Pyrolysis GC Systems: Applications to Modern Trace Organic Analysis', *J. Chromatogr. Sci.*, **21**, 1–10 (1983).
31. S.A. Liebman, D.H. Ahstrom, E.J. Quinn, J.T. Melusky, 'Thermal Decomposition of PVC and Chlorinated PVC', *J. Polym. Sci.*, **9**, 1921–1929 (1971).
32. Y. Sugimura, T. Nagaya, S. Tsuge, T. Murata, T. Takeda, 'Microstructural Characterization of Polypropylenes by High-resolution Pyrolysis–Hydrogenation Glass Capillary Gas Chromatography', *Macromolecules*, **13**, 928–932 (1980).
33. P.A. Dawes, M.K. Cumbers, P.A. Hilling, 'Evaluation of Factors Affecting Pyrolysis in a Microfurnace System', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**, 328–331 (1988).
34. A. Onishi, M. Endo, S. Uchino, N. Harashima, N. Oguri, 'An Autosampler with a Short Transfer Line for Curie-point Pyrolysis Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **16**, 353–357 (1993).
35. J.B. Pausch, R.P. Lattimer, H.L.C. Meuzelaar, 'A New Look at Direct Compound Analysis using Pyrolysis Mass Spectrometry', *Rubber Chem. Technol.*, **56**(5), 1031–1044 (1983).
36. M.R. Holdiness, 'Evolved Gas Analysis by Mass Spectrometry: A Review', *Thermochim. Acta*, **75**, 361–399 (1984).
37. K. Qian, W.E. Killinger, M. Casey, G.R. Nicol, 'Rapid Polymer Identification by In-source Direct Pyrolysis Mass Spectrometry and Library Searching Techniques', *Anal. Chem.*, **68**, 1019–1027 (1996).
38. S.A. Liebman, A.P. Snyder, J.H. Kremer, D.J. Reutter, M.A. Schroeder, R.A. Fifer, 'Time-resolved Analytical Pyrolysis Studies of Nitramine Decomposition with a Triple Quadrupole Mass Spectrometer System', *J. Anal. Appl. Pyrol.*, **12**, 83–95 (1987).
39. S.C. Israel, W.C. Yang, M. Bechard, 'Characterization of Polymers by Direct Pyrolysis/Chemical Ionization Mass Spectrometry', *J. Macromol. Sci.*, **A22**(5–7), 779–801 (1985).
40. R.P. Lattimer, 'Pyrolysis Field Ionization Mass Spectrometry of Hydrocarbon Polymers', *J. Anal. Appl. Pyrol.*, **39**, 115–127 (1997).
41. W.J. Simonsick, 'Mass Spectrometric Techniques for Coatings Characterization', ASTM Technical Publication 1119, ASTM, Philadelphia, PA, 22–38, 1992.
42. S.A. Liebman, D.H. Ahlstrom, P.R. Griffiths, 'On-line Fourier Transform Infrared Analysis of Pyrolysis and Combustion Products', *Appl. Spectrosc.*, **30**(3), 355–357 (1976).
43. T.B. Brill, 'Fast Thermolysis/FT-IR Spectroscopy', *Anal. Chem.*, **61**, 897A–894A (1989).
44. J.W. Washall, T.P. Wampler, 'Direct-pyrolysis Fourier Transform Infrared Spectroscopy for Polymer Analysis', *Spectroscopy*, **6**(4), 24–27 (1990).
45. M.J. Matheson, T.P. Wampler, W.J. Simonsick, 'The Effect of Carbon Black Filling on the Pyrolysis Behavior of Natural and Synthetic Rubbers', *J. Anal. Appl. Pyrol.*, **29**, 129–136 (1994).
46. M. Phair, T.P. Wampler, 'Analysis of Rubber Materials by Pyrolysis GC', *Rubber World*, **215**(5), 30–34 (1997).
47. H. Ohtani, T. Asai, S. Tsuge, 'Characterization of a Multicomponent Alkyl Methacrylate Copolymer by High-resolution Pyrolysis–Gas Chromatography', *Macromolecules*, **18**(6), 1148–1152 (1985).
48. F.C. Wang, P.B. Smith, 'Quantitative Analysis and Structure Determination of Styrene/Methyl Methacrylate Copolymers by Pyrolysis Gas Chromatography', *Anal. Chem.*, **68**, 3033–3037 (1996).
49. K.L. Wallisch, 'Pyrolysis of Random and Block Copolymers of Ethyl Acrylate and Methyl Methacrylate', *J. Appl. Polym. Sci.*, **13**, 203–222 (1974).
50. T.P. Wampler, G.A. Bishea, W.J. Simonsick, 'Recent Changes in Automotive Paint Formulation Using Pyrolysis–Gas Chromatography/Mass Spectrometry for Identification', *J. Anal. Appl. Pyrol.*, **40–41**, 79–89 (1997).
51. F.C. Wang, P.B. Smith, 'Compositional and Structural Studies of Vinylidene Chloride/Vinyl Chloride Copolymers by Pyrolysis Gas Chromatography', *Anal. Chem.*, **68**(3), 425–430 (1996).

Size-exclusion Chromatography of Polymers

Bernd Trathnigg
Karl-Franzens-University, Graz, Austria

1 Introduction 1
1.1 History 2
2 Applications 2
3 Reliability of Size-exclusion Chromatography 2
4 Components of a Size-exclusion Chromatography System 3
4.1 The Mobile Phase 3
4.2 The Pump 3
4.3 The Column(s) 4
4.4 Detectors 5
4.5 Data Acquisition and Processing 7
5 The Separation 8
5.1 Ideal Size Exclusion 8
5.2 Exclusion versus Nonexclusion Effects 8
5.3 The Problem of Peak Dispersion 9
6 Determination of Molar Mass 9
6.1 Size-exclusion Chromatography Calibration 10
7 Quantification in Size-exclusion Chromatography 11
7.1 Homopolymers and Oligomers 11
7.2 Copolymers and Polymer Blends 12
8 Comparison with Other Techniques 13
8.1 Other Types of Chromatography 13
8.2 Mass Spectroscopy 14
9 Hyphenated Techniques 14
9.1 Multidimensional Chromatography 15
9.2 Combination of Size-exclusion Chromatography with Mass Spectroscopy 16
10 Summary 16
Abbreviations and Acronyms 16
Related Articles 16
References 17

Size-exclusion chromatography (SEC) is a standard technique for determining molar mass averages and molar mass distributions (MMDs) of polymers. Sometimes the terms gel permeation chromatography (GPC) or gel filtration chromatography (GFC) are also used, but SEC should be preferred, because this term describes the mechanism much better: polymer molecules are separated according to their hydrodynamic volumes (which can be correlated with molar mass), with the larger size molecules exiting first followed by the smaller. Molar masses are determined either from a calibration or using molar mass sensitive detectors. In the case of copolymers, the knowledge of chemical composition along the MMD is required, which can be obtained from combinations of different concentration detectors. As the hydrodynamic volumes of different polymers are typically somewhat different, molecules with different chemical composition and different molar mass will be eluted in the same slice of the chromatogram. Obviously, a discrimination between such molecules requires a two-dimensional separation, in which one dimension may be SEC, and the other one a chromatographic technique, which separates according to chemical composition rather than molar mass, such as liquid adsorption chromatography (LAC), liquid chromatography at the critical point of adsorption (often also called liquid chromatography under critical conditions, LCCC), supercritical fluid chromatography (SFC), temperature rising elution fractionation (TREF), etc.

In the lower molar mass range, mass spectroscopy competes with SEC. The most frequently used technique is matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (MALDI/TOF/MS), which cannot, however, provide quantitatively accurate MMDs. Due to its excellent resolution in molar mass, it can be combined with chromatographic techniques in order to increase the reliability of the analysis.

1 INTRODUCTION

In the characterization of polymers, SEC has become a standard technique for determining molar mass averages and MMDs of polymers. Depending on the field of application, different terms have been used: in biochemistry and related areas the term GFC is usual, while GPC is commonly used in the analysis of (synthetic) polymers.

The principle of SEC is rather easily understood. Due to limited accessibility of the pore volume within the particles of the column packing, polymer molecules are separated according to their hydrodynamic volumes, with the larger size molecules exiting first followed by the smaller. Residence time can be correlated with molar mass. The correlation obtained then depends upon the type of polymer.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1.1 History

The origins of SEC date back to the early 1960s. In 1959, Porath and Flodin described the separation of water-soluble macromolecules on cross-linked polydextrane gels. As soon as these gels had become commercially avaliable, they were extensively used for separating biomolecules by the new technique, which was called GFC, typically in low pressure systems.[1]

In 1964, J.C. Moore of the Dow Chemical Company disclosed the separation of synthetic polymers on cross-linked polystyrene (PS) gels in organic mobile phases. The new technique was called GPC and very soon became a standard method for the determination of MMDs.

2 APPLICATIONS

Basically, SEC separates according to the size of a species in solution (the hydrodynamic volume). This species may be a single molecule, a polymer coil, an aggregate, a micelle, etc. Hence, SEC can be applied to determine the molar mass of a polymer and also to study aggregation phenomena in solution.

Typically, SEC is applied to the analysis of synthetic polymers and oligomers,[2–7] coal-derived substances,[8–10] lipids,[11,12] and natural macromolecules (such as proteins,[13–15] poly(ethylene glycol) (PEG)-modified proteins,[16,17] glucans,[18,19] cellulose derivatives,[20,21] humic substances,[22] crude-oil alkanes[23]).

SEC may also be used in studying processes accomplished by a change of the hydrodynamic volume of polymers or small molecules (such as lipids[12,24–26]): degradation,[27,28] hydrolysis,[21,29] refolding of proteins,[30] polymerization,[31–35] aggregation,[36,37] etc.

3 RELIABILITY OF SIZE-EXCLUSION CHROMATOGRAPHY

In the last few years several round-robin tests have been performed with different kinds of polymers[38–45] in order to evaluate the reproducibility of SEC and the precision and accuracy of the results thus obtained.

There may be various sources of error responsible for the differences in the results obtained at different laboratories, as can be easily understood from Figure 1, in which the experimental set-up and the basic steps in obtaining an MMD for a polymer sample are shown schematically. An appropriate mobile phase is delivered to a chromatographic column filled with a suitable stationary phase by a pump at a constant and reproducible

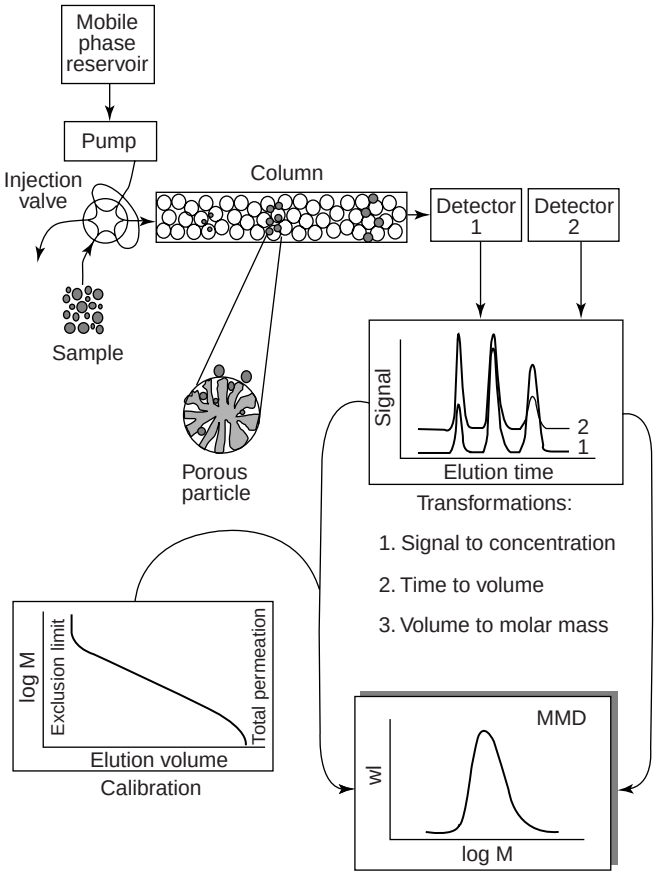

Figure 1 Schematic representation of SEC.

flow rate. Into this solvent stream a small amount (typically 0.01 to 1.0 mg) of the polymer sample is injected.

The separated fractions are detected by at least one detector, the signal of which must represent the concentration of the polymer with good accuracy. From the concentration curve thus obtained the MMD is calculated.

Provided that the separation itself is reliable (which cannot always be taken for granted!), the subsequent transformations are subject to errors:

1. Elution time to elution volume. This requires a highly constant and reproducible flow rate, which means that only high quality pumps should be used.
2. Elution volume to molar mass. The molar mass of a fraction can be obtained either from a calibration or from a molar mass sensitive detector (in addition to the concentration detector).
3. Detector response to polymer concentration. This requires a sufficiently wide linear range, a well defined response of the detector(s) along the entire peak (i.e. for all molar masses within the MMD), and – in the case of copolymers – a second concentration detector.

In the following sections, each step will be referred to in detail. Requirements concerning sample treatment, chromatographic equipment, data acquisition and processing will be discussed and different approaches to the analysis of different types of polymers evaluated.

4 COMPONENTS OF A SIZE-EXCLUSION CHROMATOGRAPHY SYSTEM

As there are considerable differences between SEC and other types of high-performance liquid chromatography (HPLC), the criteria for achieving high performance are somewhat different. In this section, the main components of an SEC system and their influence on the quality of the analysis shall be discussed.

4.1 The Mobile Phase

The mobile phase in SEC must be a good solvent for the polymer in order to avoid nonexclusion effects,[46,47] which will be discussed later on. It is also important to dissolve the sample at appropriate temperature and sufficiently long before injecting it in order to allow the coils to swell in the solvent or to break down aggregates.[48] In some cases, the addition of electrolytes can be required to achieve disaggregation.[49] As some polymers – such as polyolefins – are typically analyzed at high temperatures (140–150 °C) in rather toxic mobile phases (trichlorobenzene, etc.), alternative solvents would be desirable.[50]

An important question concerns preferential solvation: When a polymer is dissolved in a mixed solvent, the composition of the latter within the coils can be different from outside because of different interactions of the polymer with the components of the solvent. When the sample is separated on the column from the zone, where the solvent would elute, a system peak (vacancy peak) appears, which is due to the missing component of the mobile phase. Obviously, the missing amount of solvent in the system peak appears in the peak of the polymer, the area of which is now different from what it would be in absence of preferential solvation. Even though this effect has been known for a long time, it is often neglected by chromatographers, because they consider their mobile phase to be a ''pure'' solvent, which is, however, generally not the case: even HPLC-grade solvents are seldom more than 99.9% pure, and even then the concentration of the sample is in the same order of magnitude as the impurity. Moreover, solvents may take up moisture from the air, form peroxides, etc. (for example, chloroform typically contains 1% of ethanol or 2-methyl-butene as a stabilizer).

Hence it is important to dissolve the sample in the solvent from the reservoir and not from another bottle. If a solvent peak is observed, this is a strong hint for preferential solvation. Preferential solvation is often neglected, which is acceptable if its contribution does not vary along the MMD. If, however, the end groups of the polymer are considerably different from the repeating units, preferential solvation depends on molar mass, as has been shown recently.[51] A similar effect can be expected in copolymers, if their composition varies with molar mass.

4.2 The Pump

As has already been mentioned, a highly constant flow rate has to be maintained during the entire chromatogram. This is very important in SEC: due to the logarithmic relation between molar mass and elution volume a change of the flow rate of only 0.1% can cause an error in molar mass of up to 10%![52] This requires a pump of very good quality or a compensation of flow rate variations. Unfortunately, most pumps can only reproduce the flow rate to 0.2–0.3%, and this precision can be reduced by leakages in the system or increasing back pressure from the column. Moreover, the check valves as well as the pump seals may limit flow rate precision. In-line filters in the solvent reservoir may prevent particles from coming into the pump heads, which might damage the check valves or the pump seals. One should, however, take into account, that even stainless steel filters may corrode in some solvents. It is trivial that rust particles will have the same effect.

There have been attempts to determine the flow rate by measuring the travelling time of a thermal pulse along a capillary, but generally the precision of these devices is not sufficient. The more efficient – and cheaper – approach is the use of a low molecular internal standard in the MMD calibration and in each chromatogram. The corrected flow rate is obtained from the ratio of the elution times of this standard peak.

The absolute flow rate (in the calibration) can also be obtained by measuring the time to fill a calibrated flask or by weighing the solvent passing the system in a defined time.

It must, however, be said, that the knowledge of the absolute flow rate is not absolutely necessary, as long as flow rate variations are compensated by using an internal standard. It is important that such a correction works well only if the flow rate is sufficiently constant within the entire chromatogram!

4.2.1 Types of Pumps

Basically, one has to distinguish between the following types of pumps, the performance of which may differ

considerably (as well as their suitability for high-performance SEC):

- Syringe pumps. This type of pump works like a large syringe, the plunger of which is actuated by a screw-feed drive (usually by a stepper motor). Therefore it delivers a completely pulseless flow, which is especially important for systems using a viscosity detector.
- Reciprocating pumps. This group comprises almost all commercially available pumps: single piston pumps are cheap, but not well suited for SEC; dual piston pumps can have the pistons arranged parallel or in series. The former pumps deliver a smoother flow, the latter are easier to maintain, because they have only two check valves instead of four. The problem of pulsations can be solved by using a pulse dampener.

4.3 The Column(s)

Unlike in other modes of HPLC, the separation efficiency comes only from the stationary phase, while the mobile phase should have no effect. The whole separation occurs within the volume of the pores, which typically equals approximately 40% of the total column volume. This means that long columns or often sets of several columns are required. Therefore, the right choice of the column(s) for a given polymer is the crucial point.

4.3.1 Commercially Available Columns

Basically, there are different types of SEC columns on the market. The typical column diameters are 7.5–8 mm for analytical columns and 22–25 mm for (semi)preparative columns; usual column lengths are 25, 30, 50, and 60 cm. Recently, narrow bore columns with a diameter of 2–3 mm have been introduced, which save time and solvent.

The packings are based on either porous silica or semirigid (highly crosslinked) organic gels, in most cases copolymers of styrene and divinylbenzene. There are, however, other polymer-based packings available, which can be used in different mobile phases.

In general, silica-based packings are rather rugged, while organic packings have to be handled very carefully, as will be pointed out later on.

4.3.2 Selecting Size-exclusion Chromatography Columns

When selecting columns for a given separation problem in SEC, one may choose from a large number of columns from different producers. Many producers offer columns of the same type, which are comparable and sometimes almost equivalent. In general, the following considerations may lead to the choice of an appropriate column or column set:[53]

- The separation range should be selected carefully, as it does not make sense to use a column with an exclusion limit of 10^6 when analyzing low molecular products. On the other hand, the high molecular end of the MMD should still be below the exclusion limit.
- The particle size, which determines the plate height, has also to be taken into account. Small particles (typically 5 μm) provide a better resolution (higher plate numbers) and achieve the same separation with a smaller overall column length than larger ones (10 μm), but produce a higher back pressure for a given column length. Shorter columns save time and solvent. On the other hand, 5 μm (or even 3 μm) packings are more sensitive towards contamination by samples containing impurities.
- Small particle size packings can sometimes result in shear degradation of large polymer molecules because the space between particles is very narrow. Particles as large as 20 μm have been recommended for very high-molecular-weight polymers. However, axial dispersion (band spreading) effects are then increased.
- Combinations of packings with a different separation range can be achieved by using either columns with different porosity or mixed-bed columns, which typically provide a better linear calibration than combinations of columns.
- When combining columns to a set, one should prefer two 60 cm columns to four 30 cm columns, because the column ends as well as the connections increase peak broadening.
- The chemical nature of a column packing can be crucial: some packings must not be used in certain mobile phases or at higher temperatures, which are required in SEC of polyolefins. Moreover, non-exclusion effects can also be due to an inadequate stationary phase. There may be considerable differences between packings with similar specification, which are mostly due to the residual emulsifiers used in their production.

4.3.3 Handling Size-exclusion Chromatography Columns

Unlike with other HPLC columns, several precautions have to be taken in the use of SEC columns.

- A column set in SEC should be always run in the same mobile phase. This is not only because a different solvent will require a new calibration, but mainly because a solvent change can reduce column life and performance. If, however, a solvent change is

necessary (for example, to remove contamination from the packing), this should be done step-wise (using mixtures of solvents 1 and 2) and at a low flow rate (0.5 ml min^{-1} maximum). For some solvents, a direct change should be avoided by using an intermediate solvent. When switching back to the first mobile phase, the column set should be recalibrated, anyway.

- SEC columns should never be operated in a backward direction, because this may destroy the column packing immediately. Some columns will survive such a procedure, but one should not take that for granted.
- Care should also be taken in connecting columns or in sample injection: one single air bubble injected onto the column can damage the packing!
- Replacing a clogged inlet frit is a dangerous operation, which can also considerably reduce column performance. When analyzing samples, which may contaminate a column, one should always use a pre-column.
- Pulsations from the pump, which can be due to air bubbles in the solvent line, a leakage of one pump seal, or a damaged or dirty check valve, can also reduce column life.

4.3.4 Enhancing Separation Efficiency by Recycling

In SEC, the separation efficiency of a given type of packing depends on the column length, i.e. on the number of columns, which can, however, only be increased to a certain limit, which depends on the resulting back pressure. Reducing the flow rate is not a good solution, because at very low flow rates (far away from the optimum in the van Deemter equation) the plate height increases considerably.

A simple approach towards enhanced separation efficiency is recycling using the alternate pumping method, as shown in Figure 2 for a set of four columns, which are connected to a six-port–two-position valve.[7]

When the peak of interest is still in column 4, the valve is actuated (thus changing the order of the columns to 3-4-1-2), and the peak will leave column 4 to go back to column 1 instead of entering the detector. The overall column length is now 6 instead of four (1-2-3-4-1-2).

Before the peak leaves column 2, the valve is switched again, and the overall column length is again increased by two to yield 8 columns. This procedure can be repeated, as long as the entire peak fits into one half of the column set. Typically, three to four switches are allowed, thus making a column set of 10 to 12 out of 4 with the back pressure of only four columns.

Obviously, a good separation is only one part of a good analysis. Another crucial point is the detection of the fractionated sample leaving the column.

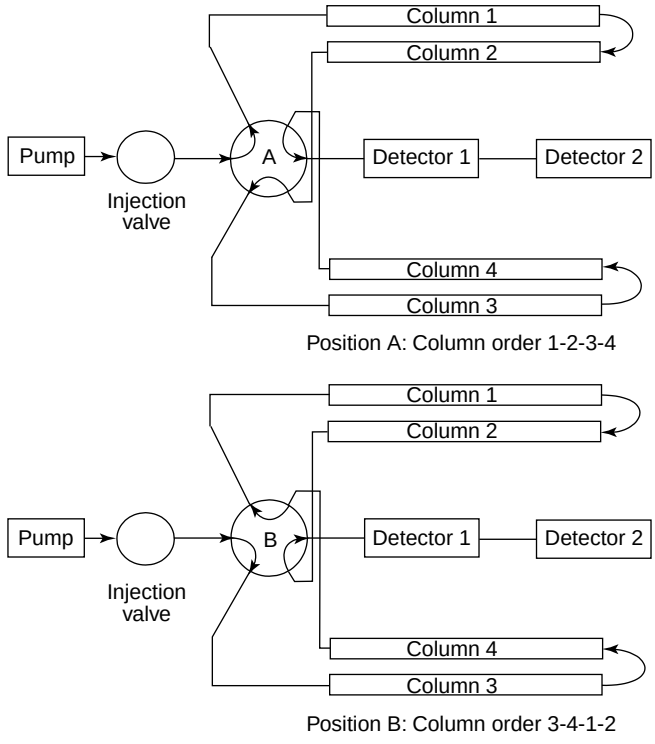

Figure 2 Schematic representation of alternate column recycle SEC.

4.4 Detectors

Among the numerous HPLC detectors, only a limited number can reasonably be applied in SEC. Basically, one has to distinguish the following groups of detectors:

4.4.1 Concentration Sensitive Detectors

It is trivial that at least one concentration sensitive detector has to be used in an SEC system. In the analysis of copolymers, a second concentration sensitive detector is required, the sensitivity of which towards the components of the polymer differs from that of the first detector.

Within the concentration sensitive detectors, one has to distinguish detectors measuring a (bulk) property of the eluate and detectors measuring a property of the solute. Evaporative detectors remove the mobile phase by evaporation prior to detection.

4.4.1.1 Bulk Property Detectors The most familiar instrument in SEC is the refractive index (RI) detector, which exists in various modifications. Its main advantage is that it can be applied in the analysis of almost any polymer.

The density detector, which has been developed in the group of the author, utilizes the principle of the mechanical oscillator and has been described in several publications.[54–56] It can be used in SEC (as an alternative

to the RI detector) and provides valuable information in the analysis of aliphatic polymers, when combined with the RI detector. This instrument is commercially available from CHROMTECH, Graz, Austria. The measuring cell of such an instrument is an oscillating, U-shaped capillary, the period of which depends on its reduced mass, and thus on the density of its content. Period measurement is performed by counting the periods of a time base (an oven-controlled 10 MHz quartz) during a predetermined number of periods of the measuring cell. The signal of such a detector is thus inherently digital, and its response is integrated over each measuring interval.

4.4.1.2 Solute Property Detectors The most familiar solute property detector is the ultraviolet (UV) absorption detector, which exists in different modifications and is available from most producers of HPLC instruments. It can be applied to polymers containing groups with double bonds, such as aromatic rings, carbonyl groups, etc., but not to any other polymers. Typical detection wavelengths are in the range of 180–350 nm, which can, however, be utilized only in solvents with a sufficiently low absorbance. Many typical SEC solvents allow detection only above a wavelength of 250 nm.

Infrared (IR) detectors are limited to certain mobile phases that are sufficiently transparent at the detection wavelength.

4.4.1.3 Evaporative Detectors Evaporative detectors vaporize the mobile phase, and the nonvolatile components of the sample can be detected on-line or off-line.

In the evaporative light scattering detector (ELSD),[12,23,26,57–59] the eluate is nebulized in a stream of pressurized gas and the solvent is evaporated from the droplets. Each droplet containing nonvolatile material forms a particle, which scatters the light of a transversal light beam. The intensity of the scattered light should reflect the concentration of nonvolatile substances in the eluate. There are, however, serious problems in quantification of the signal.[60–63]

It is also possible to use other types of evaporation devices as an interface to a flame ionization detector (FID),[64] a mass spectrometer or a Fourier transform infrared (FTIR) spectrometer.[65–68]

4.4.2 Molar Mass Sensitive Detectors

Molar mass sensitive detectors are very useful in SEC, because they yield the molar mass of each fraction of a polymer peak. As the response of such a detector depends on the concentration as well as the molar mass of the fraction, it has to be combined with a concentration sensitive detector.

Basically, the following types of molar mass sensitive detectors are on the market:

- low angle light scattering (LALS) detectors;[47,69–78]
- multiangle light scattering (MALS) detectors [see references1 21,70,75,77,79–85];
- differential viscometers.[86–90]

The information which can be obtained from such a detector is somewhat different. From light scattering detection, the absolute MMD can be determined directly. With LALS (measuring the scattering intensity at just one angle), no information is obtained on polymer conformation. Using more than one angle, one may also obtain the radius of gyration.

On the other hand, SEC with viscosity detection yields the intrinsic viscosity distribution (IVD). The MMD is, however, determined indirectly (through the universal calibration), and is thus subject to retention errors.

Consequently, it makes sense to combine a light scattering detector with a viscometer detector.[47,69,71–74,76–79] With such a combination, information on branching can be obtained.[89,91–94]

4.4.2.1 Light Scattering Detectors The scattered light of a laser beam passing the measuring cell is measured at angles different from zero. The (excess) intensity $R(\theta)$ of the scattered light at the angle θ is correlated to the weight average of molar mass M_w of the dissolved macromolecules as shown in Equation (1):

$$\frac{K^*c}{R(\theta)} = \frac{1}{M_w P(\theta)} + 2A_2c \tag{1}$$

where c is the concentration of the polymer, A_2 is the second virial coefficient, and $P(\theta)$ describes the scattered light's angular dependence.

K^*, defined in Equation (2), is an optical constant containing Avogadro's number N_A, the wavelength λ_0, RI n_0 of the solvent, and the RI increment dn/dc:

$$K^* = \frac{4\pi^2 n_0^2 (dn/dc)^2}{\lambda_0^4 N_A} \tag{2}$$

Obviously, there will be problems in copolymer analysis if their composition (and thus the RI increment dn/dc) varies within the MMD. In this case, a second concentration detector will be required, which allows a determination of copolymer composition.

A measurement at more than one angle can provide additional information. In a plot of $K^*c/R(\theta)$ versus $\sin^2(\theta/2)$, M_w can be obtained from the intercept and the radius of gyration from the slope.[70,77,79,81,95]

4.4.2.2 Viscosity Detectors[2,47,69,76,77,86–90,92,93,95–111]
A viscosity detector should yield the intrinsic viscosity $[\eta]$, the so-called limiting viscosity number, given by

Equation (3), which is defined as the limiting value of the ratio of specific viscosity ($\eta_{sp} = (\eta - \eta_0)/\eta_0$) and concentration c for $c \to 0$:

$$[\eta] = \lim_{c \to 0} \frac{\eta - \eta_0}{\eta_0 c} = \lim_{c \to 0} \frac{\eta_{sp}}{c} \tag{3}$$

As the concentrations in SEC are typically very low, $[\eta]$ can be approximated by η_{sp}/c. In viscosity detection, one has to determine both the viscosity η of the sample solution as well as the viscosity η_0 of the pure mobile phase, which can be achieved in different ways.

Viscosity measurement in SEC can be performed by measuring the pressure drop across a capillary, which is proportional to the viscosity of the streaming liquid.

Single capillary viscometers (SCVs) using just one capillary and one differential pressure transducer will be strongly affected by the pulsations of a reciprocating pump. Instruments of this type could be used with a syringe pump to eliminate this problem. (This approach is superior to that using additional pulse dampeners.)

A better, but still not perfect approach is the use of two capillaries (C1 and C2) in series, each of which is connected to a differential pressure transducer (DP1 and DP2), and a sufficiently large holdup reservoir (H) in between. The sample viscosity η is thus obtained from the pressure drop across the first capillary, and the solvent viscosity η_0 from the pressure drop across the second capillary. Pulsations are eliminated in this set-up, because they appear in both transducers simultaneously.

A very sophisticated approach is used in another type of differential viscometer, which is commercially available from Viscotek. In this instrument, four capillaries are arranged similar to a Wheatstone bridge.

In Figure 3, both designs are shown schematically. In the Viscotek instrument, a holdup reservoir in front of the reference capillary (C4) ensures that only pure mobile phase flows through the reference capillary, when the peak passes the sample capillary (C3). This design offers several advantages, the most important of which is a higher sensitivity: the detector actually measures the pressure difference ΔP at the differential pressure transducer (DP) between the inlets of the sample capillary and the reference capillary, which have a common outlet, and the overall pressure P at the inlet of the bridge. The specific viscosity $\eta_{sp} = \Delta\eta/\eta$ is thus obtained from $\Delta P/P$.

The main problem in this concept is that the flow in the system must be divided 1 : 1 between both arms of the bridge. This shall be achieved by capillaries 1 and 2, which must have a sufficiently high back pressure. Nevertheless, when a peak passes the sample capillary, a slight deviation of the 1 : 1 ratio will be observed.

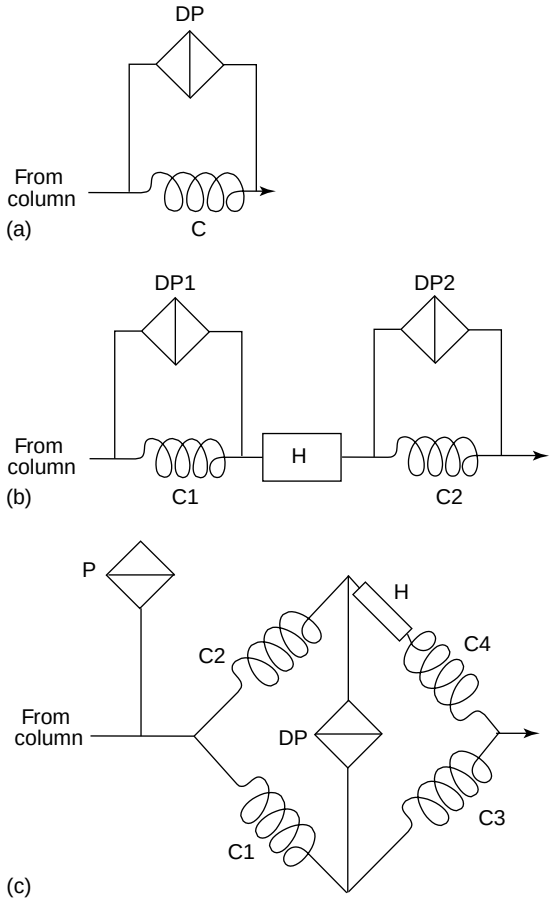

Figure 3 Schematic representation of viscosity detectors: (a) SCV; (b) dual capillary viscometer; (c) Viscotek.

The question of flow rate variations exists, however, also in single or dual capillary viscometers. When the polymer peak passes the measuring capillary, the increasing back pressure leads to a constriction in the system, and thus to a shift of the peak by a weak flow rate fluctuation (Lesec effect).[89,112]

4.5 Data Acquisition and Processing

Software for data acquisition and processing are available from all producers of HPLC equipment. As the requirements of SEC are different from those of other HPLC techniques, standard HPLC software does not fulfill the demands of SEC.

Depending on the nature of samples to be analyzed (whether high or low molecular, homo- or copolymers, etc.) and the equipment used (single or multiple detection), the software should provide special features, which will be discussed in the following sections.

In order to allow calculations not provided by the software, export of data to a spreadsheet or other programs should be possible.

5 THE SEPARATION

In SEC, the separation should be solely governed by size exclusion, which need not always be the case. Aside from an inadequate calibration, nonexclusion effects can cause severe errors. Moreover, low efficiency of the columns or the entire system will cause peak broadening, which also leads to inaccurate results.

5.1 Ideal Size Exclusion

Let us first consider the ideal case, in which size exclusion is governing the separation. As has already been mentioned, the separation in SEC has to be achieved within a volume much smaller than the volume of the column.

It is trivial that no fraction of the sample can be eluted before the interstitial volume V_i (i.e. the volume of the solvent outside the particles of the column packing) has passed the column. This elution volume corresponds to the exclusion limit of the column.

Small molecules, which have access to the entire pore volume V_p, will appear at an elution volume equal to the sum of the interstitial volume V_i and the pore volume V_p.

Molecules of a size between these extremes have access to only a part of the pore volume, hence they will be eluted at an elution volume V_e as shown in Equation (4):

$$V_e = V_i + K_{SEC} V_p \tag{4}$$

where K_{SEC} is the equilibrium constant of a sample in SEC.

The relation between K and the molar mass of a polymer is determined by a calibration, as will be discussed later on.

5.2 Exclusion versus Nonexclusion Effects

The equilibrium constant of a chromatographic separation can be correlated with thermodynamic parameters. The driving force for a separation at the (absolute) temperature T is the change in Gibbs free energy ΔG, defined in Equation (5), which results from the changes in enthalpy and entropy, ΔH and ΔS, respectively:

$$\Delta G = \Delta H - T\Delta S = -RT \ln K \tag{5}$$

In ideal SEC, which should be governed solely by entropy, ΔH should equal zero, and the equilibrium constant K_{SEC} should be given by Equation (6):

$$K_{SEC} = e^{\Delta S/R} \tag{6}$$

where $0 < K_{SEC} < 1$, with $K_{SEC} = 0$ for molecules larger than the largest pore (exclusion limit), $K_{SEC} = 1$ for small molecules, which have access to the entire pore volume V_p.

According to the theory developed by Casassa, the distribution coefficient of a flexible macromolecule with the root-mean-square end-to-end distance R in a slit-like pore with diameter $2d$ will depend on the ratio of sizes of the macromolecule and the pores. Equation (7) shows:

$$K_{SEC} = 1 - \frac{2}{\sqrt{6\pi}} \frac{R}{d} \tag{7}$$

In ideal SEC, elution volumes never exceed the void volume $V_0 = V_i + V_p$.

The opposite is true in LAC, where interactions with the stationary phase occur (whether these interactions are adsorption or partition phenomena is not important). If exclusion phenomena can be neglected (which is the case with nonporous stationary phases or in the case of small molecules and stationary phases with large pores), one may write:

$$V_e = V_i + V_p K_{LAC} \tag{8}$$

The distribution coefficient of LAC is determined by enthalpy:

$$K_{LAC} = e^{-\Delta H/RT} \tag{9}$$

As ΔH (and thus the probability of being adsorbed) increases with the number of groups capable of being adsorbed, K_{LAC} increases exponentially with the degree of polymerization. Consequently, elution volumes typically exceed the void volume considerably (as $K_{LAC} > 1$).

In practice, both exclusion and interaction must be accounted for in LAC. The equilibrium constant K can thus be divided into contributions from ideal size exclusion and adsorption, as shown in Equation (10):

$$V_e = V_i + V_p K_{SEC} K_{LAC} \tag{10}$$

It must be mentioned that even in the absence of adsorption or partition phenomena the separation can be determined by an effect other than (ideal) size exclusion. This effect is called secondary exclusion. It originates from (electrostatic) repulsion of polar groups and has nothing to do with molar mass.[46,47,113]

Mori and Nishimura[49] observed polyelectrolyte effects in SEC of poly(methyl methacrylate) (PMMA) and polyamides in hexafluoro-2-propanol. The addition of sodium trifluoroacetate as an electrolyte suppressed these effects by breaking down hydrogen bonding.

Under special conditions (mobile phase composition, temperature) the enthalpic and entropic terms in Equation (5) may compensate each other, and all polymer chains with the same structure will elute at the same volume (regardless of their number of repeating units),

which means that the polymer chain becomes ''chromatographically invisible''. This situation is utilized in LCCC[96,109,114–130] or liquid chromatography at the critical adsorption point (LCCAP),[131,132] which allows a separation according to other criteria (end groups, branching sites, other blocks in copolymers, etc.).

If a polymer contains different structural units (as is the case in block copolymers or functional oligomers), there may be basically four limiting cases:

1. all components are eluted in ideal exclusion mode;
2. main chain in exclusion mode, (weak) adsorption of end groups;
3. critical adsorption point for main chain, separation of end groups by adsorption;
4. critical adsorption point for main chain, separation of second block by exclusion.

Points 3 and 4 are beyond the scope of this chapter, hence they shall not be discussed in detail. An overview is given in a recent book.[133]

Situation 1 would be the most favorable one, which is, however, rare. In many cases, the calibration functions for different polymer homologous series (with the same repeating unit, but different end groups) can be considerably different. In a systematic investigation, Craven et al.[134,135] have studied the elution behavior of polyoxyethylenes with different end groups (diols, mono- and dimethyl ethers) on a Plgel column in different mobile phases. Considerably different calibration lines were found for the different homologous series in different mobile phases. These differences were explained by combinations of exclusion with partition adsorption effects. In the group of the author similar investigations were performed, which led to very similar results.

5.3 The Problem of Peak Dispersion

When a monodisperse sample is analyzed by chromatography, it will appear as a peak more or less of Gaussian shape and not as a rectangular concentration profile (which it was immediately after injection).

The main reasons for the broadening of peaks are diffusion phenomena in the column, the capillaries, and the detector, which can be minimized, but not completely avoided. Additional broadening can be due to high sample loads, interaction of the sample with the column packing, and an imperfect chromatographic system. Void volumes between the connecting capillaries will lead to a dramatically decreased performance of the system.

It is clear that peak broadening will adversely influence the accuracy of results from SEC, where the peak shape is much more important than the area (which is the interesting parameter in most other HPLC applications).

Basically, a chromatographic peak can be described by the function $F(v)$, the detector response at a given elution volume. It must be mentioned that the actual concentration is not always easily obtained from $F(v)$, as will be discussed later.

This function, shown in Equation (11), results from a convolution of two other functions, $G(v, y)$, which is the shape function of a solute eluting at the mean elution volume y, and $W(y)$, the chromatogram corrected for band spreading:

$$F(v) = \int_0^\infty W(y) G_N(v, y)\, dy \tag{11}$$

This equation is well known in SEC as the Tung axial dispersion equation. It is clear that the deconvolution – the calculation of $W(y)$ from $F(y)$ and $G_N(v, y)$ – can be problematic, because $G_N(v, y)$ is not easily obtained. Sometimes the so-called convolution integral, given in Equation (12), is used instead of the Tung equation:

$$F(v) = \int_0^\infty W(y) G_N(v - y)\, dy \tag{12}$$

Equation (12) is a limiting case of Equation (11), because it explicitly assumes the same normalized shape function for all solutes present and the same spreading (i.e. the same standard deviation in a Gaussian peak). This assumption may not be valid in the SEC of polymers, particularly if very high molecular weight polymers are being analyzed. Different approaches for correcting chromatograms for peak dispersion have been published, which work more or less well.[38,101,111,136–138] Because of the uncertainties in mathematically correcting for axial dispersion, the preferred approach is to utilize a good separation system, which produces low or negligible peak spreading. With today's high resolution columns other sources of error, such as flow variations, an improper baseline, neglect of the molar mass dependence of response factors, etc., are of much more concern.

Mathematical correction of peak spreading makes sense only when molecular weight averages calculated from the chromatograms of standards similar to those of the unknowns to be analyzed do not agree with those known for the standard and provided that other, more likely, sources of error have been minimized.

6 DETERMINATION OF MOLAR MASS

As has already been mentioned, three transformations have to be performed with the chromatographic raw data.

- The first one – time to volume – can be performed very easily using an internal standard, as has already been pointed out.
- The second one – volume to molar mass – requires either a calibration or the use of a molar mass sensitive detector.
- The third one – detector response to concentration (or mass) – will be discussed later. This step is especially important in SEC of copolymers, polymer blends, and oligomers.

6.1 Size-exclusion Chromatography Calibration

As has already been mentioned, the elution volume of a polymer molecule in SEC must be larger than the interstitial volume (exclusion limit) and smaller than the void volume (total permeation). Between these limits, the elution volume increases with decreasing molar mass. Unless a molar mass sensitive detector is used, one has to determine the molar mass of a fraction eluting at the volume V_e from a calibration, which can be obtained in different ways.

6.1.1 Calibration with Narrow Standards

If a series of standards with a narrow MMD is available, their elution volumes have to be determined to establish a calibration, from which the molar mass for a given elution volume is obtained. In classical SEC, a linear relation between $\log M$ and V_e was assumed, which is, however, only a first approximation, the quality of which depends very strongly on the columns used. The calibration function is quite simple in this case, as shown in Equation (13):

$$\log M = A + BV_e \tag{13}$$

where A and B are constants, which can be determined very easily by linear regression. For many columns, the calibration line is, however, sigmoidal rather than linear. In most cases, a polynomial fit can match the experimental points much better, as Equation (14) shows:

$$\log M = A + BV_e + CV_e^2 + DV_e^3 + EV_e^4 + \cdots \tag{14}$$

The coefficients A–E in such a relation have to be determined by regression analysis. This feature is provided by many software packages for SEC. The order of the polynomial fit is, however, critical in some cases: if the number of data points (i.e. the number of standards) is too small, a fit of too high an order may produce an erroneous calibration function. A plot of residuals, i.e. a plot of the percent difference in molecular weight provided by the fitted calibration line compared to the experimental data point at a particular retention volume, plotted versus retention volume is a quick, visual way of evaluating the validity of the fit. The plot reveals whether or not the scatter of data points is random around the fitted line and the magnitude of the difference between the fitted line and the experimental data points.[72]

There can be considerable differences between the calibration lines for different polymers on the same column in the same mobile phase. This is especially important in the analysis of copolymers or polymer blends. Consequently, different molar masses will elute at the same volume when a mixture of two homopolymers is analyzed by SEC. The elution volume of a copolymer should be between the elution volumes of the homopolymers of the same molar mass. If the composition of the copolymer at each point of the peak is known, an approximation will be achieved by interpolation between the calibration lines. The approximation works best for block copolymers.

It must be mentioned that different calibrations for the same polymer will be found on the same column in different mobile phases.

The calibration with narrow standards can be applied to many types of polymers, because appropriate standards have become commercially available for many polymers, and some suppliers provide well characterized standards for speciality polymers.

In the low molecular range, additional data points can be taken from the maxima of oligomer peaks, which are at least partially resolved. If one of these peaks can be identified, this is also possible for the higher oligomers. An extension to even higher molar masses can be achieved by semipreparative separation of oligomers by LAC.[139]

In the analysis of samples for which no narrow MMD standards are available, different approaches have been described in the literature. The most feasible one is the use of molar mass sensitive detectors. Alternatively, mass spectrometric techniques (such as MALDI/TOF/MS) can also be applied in establishing a calibration function.[10,23,140–147]

6.1.2 Calibration with Broad Standards

If a well characterized sample with broad MMD is available, one may use different procedures to establish a calibration fitting these averages. The integral MMD method can be applied if the entire MMD of the standard is known with high accuracy (which is, however, seldom the case). The method may assume that the MMD of the sample can be described by the most probable distribution function, and matches the calibration to this distribution. No assumptions on the shape of the calibration are made; the precision of the method is, however, rather poor at points corresponding to the tails of the distribution.

If only the molar mass averages of the sample are known from independent methods (light scattering or

osmometry), linear calibration methods can be applied. It is clear that with two known parameters only a linear calibration which is defined by two parameters (slope and intercept) can be obtained. However, this method has been expanded to nonlinear calibration curves through the use of more than one different standard. Also, it has been combined with axial dispersion correction theory to provide both a band spreading parameter (i.e. sigma) and a calibration curve.

6.1.3 Universal Calibration

A very elegant approach is based on the fact that in SEC the elution volume V_e of a polymer depends on its hydrodynamic volume, which is proportional to the product of its molar mass M and intrinsic viscosity $[\eta]$.

In a plot of log $(M[\eta])$ versus V_e (on the same column), identical calibration lines should be found for two polymers (1 and 2), which can be considered as universal calibration,[148] as shown in Equation (15):

$$M_1[\eta_1] = M_2[\eta_2] \tag{15}$$

The intrinsic viscosity is a function of molar mass, which is described by the Mark–Houwink relationship, where K and a are constants for a given polymer in a given solvent (at a given temperature), as shown in Equation (16):

$$[\eta] = KM^a \tag{16}$$

Combination of these equations yields Equation (17):

$$K_1 M_1^{a_1+1} = K_2 M_2^{a_2+1} \tag{17}$$

If a column has been calibrated with polymer 1 (e.g. PS), the calibration line for another polymer (2) can be calculated, provided that the constants K and a are known for both polymers with sufficient accuracy, as shown in Equation (18):

$$\ln M_2 = \frac{1}{1+a_2} \ln \frac{K_1}{K_2} + \frac{1+a_1}{1+a_2} \ln M_1 \tag{18}$$

The concept of the universal calibration would provide an appropriate calibration also for polymers for which no narrow standards exist.

For lower molar mass samples the Dondos–Benoit relation,[2,149] shown in Equation (19), is used, which is linear in this region:

$$\frac{1}{[\eta]} = -A_2 + \frac{A_1}{\sqrt{M}} \tag{19}$$

The main problem is the accuracy of K and a, which is rather limited even in the case of polymers for which a sufficient number of well defined standards exists: there are very high variations in the values reported in literature. If one has to rely on these data, there is the question which set of constants would yield an appropriate calibration.

After all, the expense of buying (even costly) narrow standards would be worthwhile in most cases. If such standards are not available, the method of choice will be the use of molar mass sensitive detectors.

7 QUANTIFICATION IN SIZE-EXCLUSION CHROMATOGRAPHY

Once the first two transformations (time to volume and volume to molar mass) have been performed, there remains the third transformation (detector response to amount of polymer in a fraction), which can also be subject to errors, depending on the nature of the samples. In the following section, the particular problems are referred to with respect to the type of polymer to be analyzed.

7.1 Homopolymers and Oligomers

In SEC of polymers, most chromatographers assume a constant response factor within the entire MMD, which is, however, justified only in the analysis of homopolymers with sufficiently high molar mass.

7.1.1 Molar Mass Dependence of Response Factors

The most frequently used detectors in SEC are the UV and the RI detectors. Recently, we have introduced the density detector, which is useful in the analysis of non-UV absorbing polymers.

The UV detector "sees" UV-absorbing groups in the polymer, which may be the repeating unit, the end groups, or both. Basically, there may be two limiting cases:

- If the repeating unit absorbs at the detection wavelength, the signal reflects the weight concentration of the polymer.
- If the end groups can be detected at a wavelength where the repeating units do no absorb, the signal reflects the number concentration of the polymer (provided that the functionality is known). This can be utilized for determining the number of functional groups in oligomers by derivatization with UV-active reagents (as phenyl isocyanate).[150,151]

RI and density detector measure a property of the entire eluate, that means, they are sensitive towards a specific property of the sample (the RI increment or the apparent specific volume, respectively).

It is a well known fact that specific properties are related to molar mass, as shown in Equation (20):

$$x_i = x_\infty + \frac{K}{M_i} \tag{20}$$

where x_i is the property of a polymer with molecular weight M_i, x_∞ is the property of a polymer with infinite (or at least very high) molecular weight, and K is a constant reflecting the influence of the end groups. A similar relation holds for the response factors for RI and density detection, as shown in Equation (21):

$$f_i = f_\infty + \frac{K}{M_i} \tag{21}$$

In a plot of the response factor f_i versus the molecular weight M_i of a polymer homologous series (with the same end groups) one will obtain a straight line with the intercept f_∞ (the response factor of a polymer with very high molecular weight, or the response factor of the repeating unit) and the slope K, which represents the influence of the end groups.[7,152–154]

Different methods can be applied for the determination of f_∞ and K:[155]

- If a sufficient number of monodisperse oligomers is available (as is the case with PEG), linear regression will be the method of choice.
- If at least one sample with very high molecular weight (from which the intercept f_∞ can be obtained) and a polydisperse sample with low molecular weight are available, an iteration procedure can be used to determine K.

Once f_∞ and K are known, the correct response factors for each fraction eluting from an SEC column can be calculated (with the molar mass obtained from the SEC calibration).

Molar mass dependence of response factors – unless compensated – can lead to severe errors, as has been shown in another paper.[7] Ethoxylated fatty alcohols were analyzed using SEC with coupled density and RI detection. While the chromatograms looked quite normal in density detection, the sign of the response for the lower oligomers changed in RI detection: the alkanols and the monoethoxylates appeared as negative peaks, and the diethoxylate was almost invisible.

7.2 Copolymers and Polymer Blends

In the analysis of copolymers, the use of multiple detection is generally inevitable. If the response factors of the detectors for the components of the polymer are sufficiently different, the chemical composition along the MMD can be determined from the detector signals. Typically, a combination of UV and RI detection is used,[156] but other detector combinations have also been described. If the components of the copolymer have different UV spectra, a diode array detector will be the instrument of choice. One has, however, to keep in mind that nonlinear detector response may also occur with UV detection, as Mori and Suzuki[157] have shown. They analyzed PS and copolymers of styrene with methyl methacrylate by SEC with RI and UV detection (at 254 nm) on PS gels in chloroform as mobile phase, and found that the ratio of UV and RI signals increased at the extreme parts of the MMD. Peak dispersion between the detectors, which might have caused a similar effect, was obviously not, or not alone, responsible for the deviations. In a concentration series of PSs, a nonlinear relation between sample size and peak area was found. Lukyanchikov et al.[158] described similar deviations in the analysis of butadiene–styrene copolymers and PS blends with polybutadiene (PB) and poly(dimethylsiloxane) (PDMS) using SEC with UV and refractometric detectors.

In the case of non-UV absorbing polymers, a combination of RI and density detection yields the desired information on chemical composition.[120,124,154,155,159–161] The ELSD cannot be applied because of its poor linearity and its unclear response to copolymers.

The technique can also be applied to oligomers instead of compensating for the molar mass dependence of detector response: in SEC of fatty alcohol ethoxylates or PEG macromonomers, a combination of density and RI detection can be applied as well and yields consistent results.[7,154,161]

The principle of dual detection is rather simple: when a mass m_i of a copolymer, which contains the weight fractions w_A and w_B $(= 1 - w_A)$ of the monomers A and B, is eluted in the slice i of the peak, it will cause a signal $x_{i,j}$ in the detectors, the magnitude of which depends on the corresponding response factors $f_{j,A}$ and $f_{j,B}$, where j denotes the individual detectors. This is shown in Equation (22):

$$x_{i,j} = m_i(w_A f_{A,j} + w_B f_{B,j}) \tag{22}$$

The weight fractions w_A and w_B of the monomers can be calculated using Equation (23):

$$\frac{1}{w_A} = 1 - \frac{(x_1/x_2)f_{2,A} - f_{1,A}}{(x_1/x_2)f_{2,B} - f_{1,B}} \tag{23}$$

Once the weight fractions of the monomers are known, the correct mass of polymer in the slice can be calculated using Equation (24):

$$m_i = \frac{x_i}{w_A(f_{1,A} - f_{1,B}) + f_{1,B}} \tag{24}$$

and the molecular weight M_C of the copolymer is obtained by interpolation between the calibration lines of the homopolymers, as shown in Equation (25):

$$M_C = M_B + w_A(M_A - M_B) \tag{25}$$

where M_A and M_B are the molecular weights of the homopolymers, which would elute in this slice.

The interpolation between the calibration lines cannot be applied to mixtures of polymers: If the calibration lines of the homopolymers are different, different molecular weights of the homopolymers will elute at the same volume. The universal calibration is not capable of eliminating the errors originating from the simultaneous elution of two polymer fractions with the same hydrodynamic volume but different composition and molecular weight![154]

As the molar masses of different polymers eluting at the same elution volume are given by the corresponding constants K and a in the Mark–Houwink equation, one may calculate the molar masses of the homopolymers in a polymer blend, which will be eluted in the same interval, using Equation (26):

$$\ln M = \frac{AV_e}{1+a} + \frac{B - \ln K}{1+a} \tag{26}$$

Basically, in SEC there will always be local polydispersity[162] in each slice of the polymer peak: in the case of homopolymers because of peak spreading, in the case of copolymers and polymer blends because of overlapping chemical composition distribution (CCD) and MMD.[163]

Nevertheless, a discrimination of copolymers and polymer blends is impossible with one-dimensional chromatography! Moreover, the architecture of a copolymer (random, block, graft) has to be taken into account, as Revillon[164] has shown by SEC with RI, UV, and viscosity detection. Intrinsic viscosity varies largely with molar mass according to the type of polymer, its composition, and the nature of its components.

Obviously it is feasible to use a combination of molar mass sensitive detectors, such as a LALS, MALS and viscosity detector with two concentration detectors,[72,163,165] from which the (average) composition for each fraction can be obtained, and thus the amount of polymer in the fraction.[166] When using multiple detection, one has to be aware of errors arising from inaccurate interdetector volume[74,101,108,137,166,167] and peak spreading between the detectors.[133] Bielsa and Meira[136] have studied the influence on instrumental broadening in copolymer analysis with dual-detection SEC, and demonstrated the effect of different corrections. Concentration errors may also influence the reliability of the results.[168] Mourey and Balke[72] have proposed a "systematic approach" for setting up multidetector systems. The approach is needed because, as Mourey and Balke show, in such systems, multiple sources of error are present and often the same error can originate from two different sources. The approach emphasizes the idea of ensuring that each detector alone is functioning correctly by comparing results calculated using only data from that detector with the values known for a standard before using detectors in combination. It also employs a superposition of calibration curves obtained from narrow standards and from molecular weight sensitive detectors to determine the effective volume of tubing between detectors (the effective "inter-detector volume"). This method works very well for broad molelcular weight distribution polymers but not for those with a narrow molecular weight distribution. The configuration of the detector system (whether series or parallel) was not important for broad molecular weight distribution results. It has recently been found that the inter-detector volume as measured from the difference in peak retention volumes of narrow molecular weight distribution standards from one detector to another varied with molecular weight when the detectors were in the parallel configuration and the differential viscometer (DV) was one of the detectors.[169,170] In the series configuration no such dependence was observed. This could partly account for difficulties in analyzing narrow molecular weight distribution polymers in parallel configuration systems and may be due to flow rate variation in different branches of the parallel configuration during elution of a sample.

8 COMPARISON WITH OTHER TECHNIQUES

As the analysis of polymers is a difficult task, different techniques can be applied, some of which yield similar information, while others are rather complementary to SEC.[133,171]

In oligomer analysis, SEC competes with LAC and MALDI/TOF/MS: all three techniques can be applied to determine the MMD and yield comparable results.[172]

8.1 Other Types of Chromatography

Capillary SFC and capillary high-temperature gas chromatography (HTGC) can be applied for the quantitative characterization of nonionic alcohol ethoxylate surfactants[173–176] and other oligomers.[177,178] SFC is also very useful in the analysis of carbohydrates[179] and glycerides,[180] etc.

LAC can be performed in isocratic or gradient mode. While isocratic separations[139,172,181–184] are typically limited to oligomers with a narrow MMD, gradient LAC allows also a separation of higher molar mass samples.

In some cases, chromatograms with fully resolved peaks can be obtained. PEGs can be separated on normal or reversed-phase packings,[185–188] while the separation of surfactants according to their degree of ethoxylation is only possible on normal phases.[189–193] Under similar conditions, polyesters,[194,195] PS[195–197] and other polymers can also be separated according to their degree of polymerization.

On the other hand, LAC is a technique complementary to SEC, which can be used to separate copolymers or polymer blends according to their chemical composition.[61,171,194,198–202]

Gradient elution does not necessarily mean a gradient of solvent composition: recently, temperature gradients have successfully been applied in a new technique called temperature gradient interaction chromatography (TGIC).[203,204]

LCCC allows a separation according to groups (or blocks) different from the polymer chain, which is chromatographically invisible under these special conditions. This technique is highly important in two-dimensional separations, hence it will be discussed there.

TREF can be employed to separate according to quite different criteria: the fractionation process depends on melting temperature, melting enthalpy, average crystallinity, average crystallizable sequence length, and polymer–solvent interaction parameter.[205] It is very useful in the analysis of polyolefins.[42,206] Additional information is obtained by coupling TREF with NMR spectroscopy.[206]

Field flow fractionation in various modifications can also be applied. It has been shown that the results obtained for block copolymers – poly(styrene-*b*-*p*-methoxystyrene-*b*-styrene), poly(styrene-*b*-*p*-methylstyrene-*b*-styrene) and poly(styrene-*b*-*p*-cyanostyrene) – using thermal field-flow fractionation (ThFFF), SEC and light scattering were in satisfactory agreement. ThFFF can also be used to determine the thermal-diffusion coefficients for polydisperse polymers and microgels.[84]

Capillary electrophoresis (CE)[207] can be applied in the separation of PEGs and ethoxylated surfactants.[208] Samples containing no charged group have to be derivatized prior to CE analysis with phthalic anhydride[209–211] or 1,2,4-benzenetricarboxylic anhydride[212] to impart charge and detectability on the neutral polymer.

8.2 Mass Spectroscopy

In the analysis of oligomers (such as nonionic surfactants), fast atom bombardment (FAB), time-of-flight secondary ion mass spectrometry, MALDI, electrospray ionization, and field desorption can be applied.[213] The most frequently used mass spectroscopic technique is MALDI/TOF/MS, which has been applied successfully in the analysis of poly((*R*)-3-hydroxybutanoates),[214] coal-derived liquids[8] and many other oligomers and polymers.

The technique has some considerable advantages. It is rapid, requires very small sample amounts, and its resolution and mass-accuracy are marvellous.

On the other hand, there are serious concerns about the quantitation, for the following reasons:

- Sample preparation and desorption/ionization can introduce serious mass biasing that appears to be due to the characteristics of the MALDI process.[215] There are pronounced effects of solvents, particularly solvent mixtures, used to prepare polymer, matrix, and cationization reagent solutions, on MALDI analysis:[216] solvent mixtures containing a polymer nonsolvent can affect the signal reproducibility and cause errors in average weight measurement. Hence it is important to select a solvent system that will allow matrix crystallization to take place prior to polymer precipitation. If these preconditions are fulfilled, MALDI mass spectrometry can provide accurate molecular weight and molecular weight distribution information for narrow polydispersity polymers.[217]
- Serious problems arise in the analysis of polymers with wide polydispersity: the highest mass molecules in the distribution are not observed unless the more abundant lower mass ions are deflected from reaching the detector.[218]

Polydisperse polymers can be analyzed by a combination of MALDI/TOF/MS with SEC, which can be used to obtain fractions with a narrow MMD.[141,143] Microscale SEC can even be coupled on-line to MALDI/TOF/MS with a robotic interface.[142]

Time-lag focusing MALDI mass spectrometry has been employed to analyse PMMA polymers of industrial relevance.[219] This technique also enables the differentiation of end groups.

9 HYPHENATED TECHNIQUES

The analysis of complex polymers and oligomers is complicated by the fact that there may be several distributions in such samples: MMD, CCD, and type of functionality, eventually also architecture (tacticity, branching, blockiness, etc.). Recently, a combination of SEC with 750 MHz NMR has been successfully applied to determine the MMD and the tacticity of PMMA. The molar mass of the polymer in flowing eluate was determined directly (without a conventional calibration procedure) from the relative intensity of NMR signals due to the end-group and repeating units.[146]

Obviously, a full characterization of such samples is very difficult, if it is possible at all. Anyway, it cannot be achieved by simple analytical techniques.

The goal of a full characterization may be approached in several steps, each of which represents a more or less sufficient approximation and will be subject to particular sources of error, as has already been pointed out in the previous sections.

Concerning the particular case of SEC, the following limitations have to be observed:

- One-dimensional separations with one concentration detector may be applied to homopolymers, where calibration standards are available.
- One-dimensional separations with two concentration detectors may be applied to copolymers, where calibration standards are available for both homopolymers.
- One-dimensional separations with one concentration detector and one molar mass detector may be applied to homopolymers of any type. In the case of copolymers, the chemical composition is required for each molar mass. This can be achieved by a second concentration detector.
- One-dimensional separations with two concentration detectors and one molar mass detector may be applied to copolymers with the same architecture. The determination of molar mass and branching requires, however, one more molar mass detector.
- One-dimensional separations with two concentration detectors and two molar mass detectors (viscometer plus LALS or MALS) may be applied to all copolymers. No discrimination between copolymers and polymer blends is possible even in this case.

Basically, multiple detection always yields only the average composition or molar mass of each fraction: the CCD or type of functionality in addition to the MMD can only be obtained by two-dimensional separations (in some cases, even three or more dimensions would be required, which is, however, not yet possible in practice).

The chromatographic and mass spectroscopic techniques described above (SEC, LAC, LCCC, SFC, field-flow fractionation, and MALDI/TOF/MS), which yield different kinds of information, can be combined in different ways:

- When applied independently, they yield different projections of a three-dimensional surface, which describe complex polymers and oligomers: in the case of copolymers with the axes molar mass, chemical composition, and (weight) fraction (as altitude), in the case of functional oligomers with functionality instead of composition.
- Two-dimensional separations, which allow an independent determination of two distributions, can be achieved by combining different modes of chromatography or by coupling a chromatographic separation to a mass spectrometer (preferably MALDI/TOF/MS).[129]

9.1 Multidimensional Chromatography

The distributions of molar mass and functionality can be determined by orthogonal chromatography.[220,221] This technique was also applied to determine MMD and CCD of poly(ethylene oxide-*b*-propylene oxide)s (with LCCC as the first dimension and SEC or SFC as the second one).[116] The application of SEC and nonexclusion liquid chromatography in the characterization of styrene copolymers was described by Mori.[222] Nonexclusion liquid chromatography for polymer separation can be divided into five separation techniques: adsorption, precipitation (solubility), normal and reversed phases, orthogonal, and adsorption at a critical point.[223]

Methyl methacrylate–methacrylic acid copolymers were analyzed by a combination of normal-phase LAC with gradient elution and SEC.[224]

Random copolymers of *N*-vinylpyrrolidone and 2-methyl-5-vinylpyridine were analyzed by SEC–reversed-phase LAC.[105]

A quantitatively accurate mapping of fatty alcohol ethoxylates can be achieved by a combination of LCCC and SEC with coupled density and RI detection in both dimensions.[225] Alternatively, normal-phase LAC may be used as the second dimension.[226]

On-line coupling of SEC, normal-phase liquid chromatography, and gas chromatography was applied in the characterization of complex hydrocarbon mixtures.[227]

Cross-fractionation of a PS sample blended with a PB, and of butadiene– and styrene–methylmethacrylate copolymers by combining SEC with ThFFF has been described.[228]

PS–poly(ethylene oxide) blends and copolymers were analyzed with respect to CCD and MMD using two-dimensional SEC/ThFFF.[229]

A two-dimensional separation of peptides by SEC/reversed-phase liquid chromatography coupled to mass spectrometry has been described recently.[15]

SEC has also been coupled to anion-exchange chromatography in the analysis of polysaccharides and oligosaccharides.[230]

Coupling of full adsorption–desorption and SEC has been applied to the separation and molecular characterization of polymer blends.[231–234]

9.2 Combination of Size-exclusion Chromatography with Mass Spectroscopy

As has already been pointed out, MALDI/TOF/MS can only be applied to polymers with a narrow MMD. Polydisperse polymers can be analyzed with good accuracy by an SEC fractionation (which yields narrow MMD fractions) prior to mass spectroscopy.[141,143] On the other hand, MALDI/TOF/MS is an excellent tool for establishing SEC calibration functions.[145,147,235] In LCCC of oligomers, it yields information on the type of the functionality as well as on the quality of the chromatographic separation.[129,221]

10 SUMMARY

The potential of SEC in polymer characterization is very high, especially when this technique is combined with other modes (LAC, LCCC, SFC) or with mass spectrometric techniques, such as MALDI/TOF/MS.

Multiple detection is in most cases inevitable: combinations of different concentration detectors provide information on copolymer composition, and with molar mass sensitive detectors one may avoid errors with inadequate calibrations.

For complex polymers (with distributions in molar mass, chemical composition, functionality, etc.) one-dimensional techniques can, however, only provide part of the desired information. For these samples, multidimensional separations will be required. In most cases, one of the dimensions will be SEC, while the other(s) could be (gradient) LAC or LCCC.

ABBREVIATIONS AND ACRONYMS

CCD	Chemical Composition Distribution
CE	Capillary Electrophoresis
DV	Differential Viscometer
ELSD	Evaporative Light Scattering Detector
FAB	Fast Atom Bombardment
FID	Flame Ionization Detector
FTIR	Fourier Transform Infrared
GFC	Gel Filtration Chromatography
GPC	Gel Permeation Chromatography
HPLC	High-performance Liquid Chromatography
HTGC	High-temperature Gas Chromatography
IR	Infrared
IVD	Intrinsic Viscosity Distribution
LAC	Liquid Adsorption Chromatography
LALS	Low Angle Light Scattering
LCCAP	Liquid Chromatography at the Critical Adsorption Point
LCCC	Liquid Chromatography Under Critical Conditions
MALDI/TOF/MS	Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectroscopy
MALS	Multiangle Light Scattering
MMD	Molar Mass Distribution
PB	Polybutadiene
PDMS	Poly(dimethylsiloxane)
PEG	Poly(ethylene Glycol)
PMMA	Poly(methyl Methacrylate)
PS	Polystyrene
RI	Refractive Index
SCV	Single Capillary Viscometer
SEC	Size-exclusion Chromatography
SFC	Supercritical Fluid Chromatography
TGIC	Temperature Gradient Interaction Chromatography
ThFFF	Thermal Field-flow Fractionation
TREF	Temperature Rising Elution Fractionation
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Particle Size Analysis **(Volume 6)**
Field-flow Fractionation in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Polymers and Rubbers **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization • Field Flow Fractionation in Analysis of Polymers and Rubbers • Gas Chromatography in Analysis of Polymers and Rubbers • Infrared

Spectroscopy in Analysis of Polymers and Rubbers • Pyrolysis Techniques in the Analysis of Polymers and Rubbers • Supercritical Fluid Chromatography of Polymers • Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation

Process Instrumental Methods **(Volume 9)**
Chromatography in Process Analysis

Infrared Spectroscopy **(Volume 12)**
Liquid Chromatography/Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Biopolymer Chromatography • Gradient Elution Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Time-of-flight Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. J. Porath, 'From Gel-filtration to Adsorptive Size-exclusion', *J. Protein Chem.*, **16**(5), 463–468 (1997).
2. J.R. Xie, 'Viscometric Constants for Small Polystyrenes and Polyisobutenes by Gel-permeation Chromatography', *Polymer*, **35**(11), 2385–2389 (1994).
3. F. Ysambertt, W. Cabrera, N. Marquez, J.L. Salager, 'Analysis of Ethoxylated Nonylphenol Surfactants by High-performance Size-exclusion Chromatography (Hpsec)', *J. Liq. Chromatogr.*, **18**(6), 1157–1171 (1995).
4. H. Dai, P.L. Dubin, T. Andersson, 'Permeation of Small Molecules in Aqueous Size-exclusion Chromatography Vis-a-vis Models for Separation', *Anal. Chem.*, **70**(8), 1576–1580 (1998).
5. B. Geetha, A.B. Mandal, T. Ramasami, 'Synthesis, Characterization, and Micelle Formation in an Aqueous-solution of Methoxypoly(ethylene Glycol) Macromonomer, Homopolymer, and Graft Copolymer', *Macromolecules*, **26**(16), 4083–4088 (1993).
6. B. Selisko, C. Delgado, D. Fisher, R. Ehwald, 'Analysis and Purification of Monomethoxy-polyethylene Glycol by Vesicle and Gel-permeation Chromatography', *J. Chromatogr.*, **641**(1), 71–79 (1993).
7. B. Trathnigg, D. Thamer, X. Yan, B. Maier, H.R. Holzbauer, H. Much, 'Characterization of Ethoxylated Fatty Alcohols Using Liquid-chromatography with Density and Refractive-index Detection. 1. Quantitative-analysis of Pure Homologous Series by Size-exclusion Chromatography', *J. Chromatogr. A*, **657**(2), 365–375 (1993).
8. M.J. Lazaro, A.A. Herod, M. Cocksedge, M. Domin, R. Kandiyoti, 'Molecular-mass Determinations in Coal-derived Liquids by Maldi Mass-spectrometry and Size-exclusion Chromatography', *Fuel*, **76**(13), 1225–1233 (1997).
9. K. Henning, H.J. Steffes, R.M. Fakoussa, 'Effects on the Molecular-weight Distribution of Coal-derived Humic Acids Studied by Ultrafiltration', *Fuel Process. Technol.*, **52**(1–3), 225–237 (1997).
10. B.R. Johnson, K.D. Bartle, M. Domin, A.A. Herod, R. Kandiyoti, 'Absolute Calibration of Size-exclusion Chromatography for Coal Derivatives Through Maldi-MS', *Fuel*, **77**(9–10), 933–945 (1998).
11. A.I. Hopia, V.I. Piironen, P.E. Koivistoinen, L.E.T. Hyvonen, 'Analysis of Lipid Classes by Solid-phase Extraction and High-performance Size-exclusion Chromatography', *J. Am. Oil Chem. Soc.*, **69**(8), 772–776 (1992).
12. A.I. Hopia, A.M. Lampi, V.I. Piironen, L.E.T. Hyvonen, P.E. Koivistoinen, 'Application of High-performance Size-exclusion Chromatography to Study the Autoxidation of Unsaturated Triacylglycerols', *J. Am. Oil Chem. Soc.*, **70**(8), 779–784 (1993).
13. H. Lakhiari, T. Okano, N. Nurdin, C. Luthi, P. Descouts, D. Muller, J. Jozefonvicz, 'Temperature-responsive Size-exclusion Chromatography Using Poly(N-Isopropylacrylamide) Grafted Silica', *Bba-Gen Subjects*, **1379**(3), 303–313 (1998).
14. G.B. Irvine, 'Size-exclusion High-performance Liquid-chromatography of Peptides – A Review', *Anal. Chim. Acta*, **352**(1–3), 387–397 (1997).
15. G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, '2-Dimensional SEC/RPLC Coupled to Mass-spectrometry for the Analysis of Peptides', *Anal. Chem.*, **69**(13), 2283–2291 (1997).
16. G. Fortier, M. Laliberte, 'Surface Modification of Horseradish-peroxidase with Poly(ethylene Glycol)s of Various Molecular Masses – Preparation of Reagents and Characterization of Horseradish-peroxidase Poly(ethylene Glycol) Adducts', *Biotechnol. Appl. Biochem.*, **17**, 115–130 (1993).
17. J.J. Ratto, S.R. Oconner, A.R. Distler, G.M. Wu, D. Hummel, M.J. Treuheit, A.C. Herman, J.M. Davis, 'Ethanol Sodium-chloride Phosphate Mobile-phase for Size-exclusion Chromatography of Poly(ethylene Glycol) Modified Proteins', *J. Chromatogr. A*, **763**(1–2), 337–344 (1997).
18. A. Huber, W. Praznik, 'Characterization of Branching Characteristics of Starch-glucans by Means of Combined Application of Complexation, Enzymatically Catalyzed Modification, and Liquid-chromatography', *J. Liq. Chromatogr.*, **17**(18), 4031–4056 (1994).

19. B.E. Knuckles, W.H. Yokoyama, M.M. Chiu, 'Molecular Characterization of Barley Beta-glucans by Size-exclusion Chromatography with Multiple-angle Laser-light Scattering and Other Detectors', *Cereal Chem.*, **74**(5), 599–604 (1997).
20. T.E. Eremeeva, T.O. Bykova, 'SEC of Mono-carboxymethyl Cellulose (CMC) and a Wide-range of pH – Mark–Houwink Constants', *Carbohyd. Polym.*, **36**(4), 319–326 (1998).
21. E.Y. Vlasenko, A.I. Ryan, C.F. Shoemaker, S.P. Shoemaker, 'The use of Capillary Viscometry, Reducing End-group Analysis, and Size-exclusion Chromatography Combined with Multi-angle Laser-light Scattering to Characterize Endo-1,4-Beta-D-Glucanases on Carboxymethylcellulose – a Comparative Evaluation of the Three Methods', *Enzyme Microb. Technol.*, **23**(6), 350–359 (1998).
22. S. Tao, 'A Sequential Gel-filtration Chromatographic Method to Estimate the Molecular-weight Distribution of Humic Substances', *Environ. Technol.*, **15**(11), 1083–1088 (1994).
23. L. Carbognani, 'Fast Monitoring of C-20-C-160 Crude-oil Alkanes by Size-exclusion Chromatography Evaporative Light-scattering Detection Performed with Silica Columns', *J. Chromatogr. A*, **788**(1–2), 63–73 (1997).
24. M. Mittelbach, B. Trathnigg, 'Kinetics of Alkaline Catalyzed Methanolysis of Sunflower Oil', *Fett. Wissenschaft Technol. Fat Science Tech.*, **92**(4), 145–148 (1990).
25. B. Trathnigg, M. Mittelbach, 'Analysis of Triglyceride Methanolysis Mixtures using Isocratic HPLC with Density Detection', *J. Liquid Chromatogr.*, **13**(1), 95–105 (1990).
26. I.C. Burkow, R.J. Henderson, 'Isolation and Quantification of Polymers from Autoxidized Fish Oils by High-performance Size-exclusion Chromatography with an Evaporative Mass Detector', *J. Chromatogr.*, **552**(1–2), 501–506 (1991).
27. B. Trathnigg, V. Weidmann, R.M. Lafferty, B. Korsatko, W. Korsatko, 'Poly(beta-hydroxybutyrate) of Low-molecular Weight – Synthesis and Applications', *Angew. Makromol. Chem.*, **161**, 1–8 (1988).
28. D.S.G. Hu, H.J. Liu, 'Effect of Soft Segment on Degradation Kinetics in Polyethylene Glycol/Poly(L-lactide) Block-copolymers', *Polym. Bull.*, **30**(6), 669–676 (1993).
29. G.R. Ponder, 'Arabinogalactan from Western Larch. Part IV. Polymeric Products of Partial Acid-hydrolysis', *Carbohyd. Polym.*, **36**(1), 1–14 (1998).
30. B. Batas, H.R. Jones, J.B. Chaudhuri, 'Studies of the Hydrodynamic Volume Changes that Occur During Refolding of Lysozyme Using Size-exclusion Chromatography', *J. Chromatogr. A*, **766**(1–2), 109–119 (1997).
31. B. Trathnigg, 'Measurement of Density for Determining the Conversion of Polymerization Reactions. 2. Determination of Conversion in the Same Run by Different Methods', *Angew. Makromol. Chem.*, **88**, 127–133 (1980).
32. B. Trathnigg, G. Hippmann, 'Polymerization of Spiro-orthoesters. 2. A Kinetic Approach', *Eur. Polym. J.*, **17**(10), 1085–1087 (1981).
33. K. Ishizu, S. Tsubaki, S. Uchida, 'Free-radical Polymerization of Macromonomers', *J. Macromol. Sci. Pure Appl. Chem.*, **A32**(7), 1227–1234 (1995).
34. S. Cammas, Y. Nagasaki, K. Kataoka, 'Heterobifunctional Poly(ethylene Oxide) – Synthesis of Alpha-methoxy-omega-amino and Alpha-hydroxy-omega-amino PEOS with the Same Molecular-weights', *Bioconjugate Chem.*, **6**(2), 226–230 (1995).
35. M.D. Zammit, M.L. Coote, T.P. Davis, G.D. Willett, 'Effect of the Ester Side-chain on the Propagation Kinetics of Alkyl Methacrylates – An Entropic or Enthalpic Effect', *Macromolecules*, **31**(4), 955–963 (1998).
36. I.V. Berlinova, A. Amzil, N.G. Vladimirov, 'Amphiphilic Graft-copolymers with Poly(oxyethylene) Side-chains – Synthesis via Phthalimidoacrylate Prepolymers', *J. Polym. Sci. A, Polym. Chem.*, **33**(11), 1751–1758 (1995).
37. M. Kunitani, S. Wolfe, S. Rana, C. Apicella, V. Levi, G. Dollinger, 'Classical Light-scattering Quantitation of Protein Aggregates – Off-line Spectroscopy Versus HPLC Detection', *J. Pharmaceut. Biomed. Anal.*, **16**(4), 573–586 (1997).
38. K. Lederer, I. Beytollahiamtmann, J. Billiani, 'Determination and Correction of Peak Broadening in Size-exclusion Chromatography of Controlled Rheology Polypropylene', *J. Appl. Polym. Sci.*, **54**(1), 47–55 (1994).
39. 'Report on Cooperative Determination of Molecular-weight Averages of Polymers by Size-exclusion Chromatography. 1. A Report on the First Round-robin Test (No. 1)', *Bunseki Kagaku*, **44**(6), 497–504 (1995).
40. S. Mori, S. Takayama, Y. Goto, K. Nagata, A. Kinugawa, T. Housaki, M. Yabe, K. Takada, T. Sugimoto, M. Shimizu, I. Nagashima, A. Hasegawa, T. Senba, N. Ooshima, T. Maekawa, H. Sugitani, H. Oozeki, K. Nakahashi, K. Hibi, H. Ootani, S. Nakamura, K. Sugiura, T. Tanaka, S. Ogiwara, Y. Katsuno, T. Ookubo, 'Report on Cooperative Determination of Molecular-weight Averages of Polymers by Size-exclusion Chromatography. 2. A Report on the Second Round-robin Test (No. 1)', *Bunseki Kagaku*, **45**(1), 95–101 (1996).
41. S. Mori, S. Takayama, Y. Goto, M. Nagata, A. Kinugawa, T. Housaki, M. Yabe, K. Takada, T. Sugimoto, M. Shimizu, I. Nagashima, A. Hasegawa, T. Senba, N. Ooshima, T. Maekawa, H. Sugitani, H. Oozeki, K. Nakahashi, K. Hibi, H. Ootani, S. Nakamura, K. Sugiura, Y. Umeda, S. Ogiwara, Y. Katsuno, T. Ookubo, 'Report on Cooperative Determination of Molecular-weight Averages of Polymers by Size-exclusion Chromatography. 4. A Report on the Second Round-robin Test (No. 2) – On the Construction of Base-line', *Bunseki Kagaku*, **45**(9), 879–885 (1996).

42. K. Lederer, N. Aust, 'Molecular Characterization of Polyolefins', *J. Macromol. Sci. Pure Appl. Chem.*, **A33**(7), 927–940 (1996).
43. M. Nagata, S. Mori, Y. Goto, M. Yabe, K. Takada, N. Ooshima, H. Oozeki, K. Sugiura, 'Report on the Cooperative Determination of the Molecular-weight Averages of Polymers by Size-exclusion Chromatography. 6. A Report on the Third Round-robin Test', *Bunseki Kagaku*, **46**(2), 159–166 (1997).
44. S. Mori, S. Takayama, Y. Goto, M. Nagata, A. Kinugawa, T. Housaki, M. Yabe, K. Takada, T. Sugimoto, M. Shimizu, I. Nagashima, A. Hasegawa, T. Senba, N. Ooshima, T. Maekawa, H. Sugitani, H. Oozeki, K. Nakahashi, K. Hibi, H. Ootani, S. Nakamura, K. Sugiura, Y. Umeda, Y. Katsuno, M. Taguchi, 'Report on the Cooperative Determination of Molecular-weight Averages of Polymers by Size-exclusion Chromatography. 5. Dependence of Molecular-weight Averages on Detectors and Mobile Phases', *Bunseki Kagaku*, **46**(1), 81–86 (1997).
45. S. Mori, Y. Goto, M. Nagata, A. Kinugawa, T. Housaki, M. Yabe, K. Mori, T. Sugimoto, M. Shimizu, I. Nagashima, T. Senba, T. Shimizu, T. Maekawa, H. Sugitani, H. Oozeki, S. Takahashi, K. Hibi, H. Ootani, S. Nakamura, K. Sugiura, Y. Umeda, Y. Katsuno, M. Taguchi, T. Nozu, 'Report on the Cooperative Determination of Molecular-weight Averages of Polymers by Size-exclusion Chromatography. 7. A Report on the Fourth Round-robin Test (No. 1)', *Bunseki Kagaku*, **46**(10), 837–844 (1997).
46. H.G. Barth, 'Nonsize Exclusion Effects in High-performance Size Exclusion Chromatography', *ACS Symp. Ser.*, **352**, 29–46 (1987).
47. G. Volet, J. Lesec, 'Nonexclusion Effects in Aqueous SEC – Behavior of Some Polyelectrolytes Using Online Mass Detectors', *J. Liq. Chromatogr.*, **17**(3), 559–577 (1994).
48. N. Manabe, K. Kawamura, T. Toyoda, H. Minami, M. Ishikawa, S. Mori, 'Sample Pretreatment for Aggregate-free PVC Molecular-weight Measurement by Size-exclusion Chromatography', *J. Appl. Polym. Sci.*, **68**(11), 1801–1809 (1998).
49. S. Mori, Y. Nishimura, 'Effects of Addition of an Electrolyte on Size-exclusion Chromatography of Polyamides Using Hexafluoro-2-propanol as Mobile-phase', *J. Liq. Chromatogr.*, **16**(16), 3359–3370 (1993).
50. B. Rao, S.T. Balke, T.H. Mourey, T.C. Schunk, 'Methyl Cyclohexane as a New Eluting Solvent for the Size-exclusion Chromatography of Polyethylene and Polypropylene at 90 °C', *J. Chromatogr. A*, **755**(1), 27–35 (1996).
51. B. Trathnigg, X. Yan, 'Polymer Analysis Using Size-exclusion Chromatography with Coupled Density and Refractive-index Detection. 6. Molecular-mass Dependence of Preferential Solvation of Polyoxyethylenes in Chloroform with Different Ethanol Contents', *J. Chromatogr. A*, **653**(2), 199–206 (1993).
52. L. Letot, J. Lesec, C. Quivoron, 'The Errors in the Treatment of Gel-permeation Chromatography Data – Origins and Corrections', *J. Liq. Chromatogr.*, **3**(11), 1637–1655 (1980).
53. A.A. Gorbunov, A.M. Skvortsov, 'Resolution Obtainable with the Gel-permeation Chromatography Method Applied to Polymers and Proteins', *Polymer*, **32**(16), 3001–3005 (1991).
54. B. Trathnigg, H. Leopold, 'Density-measurement for Detection in Gel-chromatography. 5. Compensation of Temperature-fluctuations by Reference Measurements', *Makromol. Chem. Rapid Commun.*, **1**(9), 569–571 (1980).
55. B. Trathnigg, C. Jorde, 'Densimetric Detection in Gel-permeation Chromatography. 6. An Integrating Densimetric Detector', *J. Chromatogr.*, **241**(1), 147–151 (1982).
56. B. Trathnigg, C. Jorde, 'New Universal Detector for High-performance Liquid Chromatography – The Density Detector', *J. Chromatogr.*, **385**, 17–23 (1987).
57. M. Dreux, M. Lafosse, 'Measurement of Evaporative Light-scattering of Microparticles in Gaseous-phase', *Analusis*, **20**(10), 587–595 (1992).
58. W. Miszkiewicz, J. Szymanowski, 'Analysis of Nonionic Surfactants with Polyoxyethylene Chains by High-performance Liquid-chromatography', *Crit. Rev. Anal. Chem.*, **25**(4), 203–246 (1996).
59. M. Dreux, M. Lafosse, L. Morin-Allory, 'The Evaporative Light Scattering Detector – A Universal Instrument for Non-volatile Solutes in LC and SFC', *LC-GC Int.*, **9**(3), 148–156 (1996).
60. P. Van der Meeren, J. Van der Deelen, L. Baert, 'Simulation of the Mass Response of the Evaporative Light-scattering Detector', *Anal. Chem.*, **64**(9), 1056–1062 (1992).
61. O. Chiantore, 'Compositional Characterization of AES Blends by Size-exclusion and Precipitation-redissolution Liquid-chromatography', *Ind. Eng. Chem. Res.*, **36**(4), 1276–1282 (1997).
62. B. Trathnigg, M. Kollroser, D. Berek, M. Janco, 'Quantitation in the Analysis of Oligomers by HPLC with Elsd', *Abstr. Pap. Am. Chem. Soc.*, **214**, 220-PMSE (1997).
63. B. Trathnigg, M. Kollroser, 'Liquid-chromatography of Polyethers Using Universal Detectors. 5. Quantitative Aspects in the Analysis of Low-molecular-mass Poly(ethylene Glycol)s and their Derivatives by Reversed-phase High-performance Liquid-chromatography with an Evaporative Light-scattering Detector', *J. Chromatogr. A*, **768**(2), 223–238 (1997).
64. G.H.J. Vandoremaele, J. Kurja, H.A. Claessens, A.L. German, 'Flame Ionization Detection for the Determination of the Chemical-composition Distribution of Non-UV Absorbing Copolymers', *Chromatographia*, **31**(9–10), 493–499 (1991).

65. J.N. Willis, L. Wheeler, 'The Use of a GPC-FTIR Interface for Polymer Analysis', *Abstr. Pap. Am. Chem. Soc.*, **206**, 61-PMSE (1993).
66. T.C. Schunk, S.T. Balke, P. Cheung, 'Quantitative Polymer Composition Characterization with a Liquid-chromatography Fourier-transform Infrared Spectrometry Solvent-evaporation Interface', *J. Chromatogr. A*, **661**(1–2), 227–238 (1994).
67. T. Provder, C.Y. Kuo, M. Whited, D. Huddleston, 'GPC FTIR Characterization of Copolymer Compositional Heterogeneity', *Abstr. Pap. Am. Chem. Soc.*, **208**, 186-PMSE (1994).
68. P.C. Cheung, S.T. Balke, T.C. Schunk, 'Size-exclusion Chromatography Fourier-transform IR Spectrometry Using a Solvent-evaporative Interface', *Adv. Chem. Ser.*, **247**, 265–279 (1995).
69. R. Hanselmann, W. Burchard, R. Lemmes, D. Schwengers, 'Characterization of Deae-dextran by Means of Light-scattering and Combined Size-exclusion Chromatography Low-angle Laser-light Scattering Viscometry', *Macromol. Chem. Phys.*, **196**(7), 2259–2275 (1995).
70. L. Jeng, S.T. Balke, T.H. Mourey, L. Wheeler, P. Romeo, 'Evaluation of Light-scattering Detectors for Size-exclusion Chromatography. 1. Instrument Precision and Accuracy', *J. Appl. Polym. Sci.*, **49**(8), 1359–1374 (1993).
71. M. Konas, T.M. Moy, M.E. Rogers, A.R. Shultz, T.C. Ward, J.E. McGrath, 'Molecular-weight Characterization of Soluble High-performance Polyimides. 2. Validity of Universal SEC Calibration and Absolute Molecular-weight Calculation', *J. Polym. Sci., Part B: Polym. Phys.*, **33**(10), 1441–1448 (1995).
72. T.H. Mourey, S.T. Balke, 'A Strategy for Interpreting Multidetector Size-exclusion Chromatography Data. I. Development of a Systematic Approach', *ACS Symp. Ser.*, **521**, 180–198 (1993).
73. M.H. Ottoy, K.M. Varum, B.E. Christensen, M.W. Anthonsen, O. Smidsrod, 'Preparative and Analytical Size-exclusion Chromatography of Chitosans', *Carbohyd. Polym.*, **31**(4), 253–261 (1996).
74. M.G. Pigeon, A. Rudin, 'Correction for Interdetector Volume in Size-exclusion Chromatography (SEC)', *J. Appl. Polym. Sci.*, **57**(3), 287–301 (1995).
75. D.S. Poche, A.J. Ribes, D.L. Tipton, 'Characterization of Methocel™ – Correlation of Static Light-scattering Data to GPC Molar-mass Data Based on Pullulan Standards', *J. Appl. Polym. Sci.*, **70**(11), 2197–2210 (1998).
76. R.L. Qian, R. Mhatre, I.S. Krull, 'Characterization of Antigen – Antibody Complexes by Size-exclusion Chromatography Coupled with Low-angle Light-scattering Photometry and Viscometry', *J. Chromatogr. A*, **787**(1–2), 101–109 (1997).
77. W.F. Reed, 'Data Evaluation for Unified Multidetector Size-exclusion Chromatography – Molar-mass, Viscosity and Radius of Gyration Distributions', *Macromol. Chem. Phys.*, **196**(5), 1539–1575 (1995).
78. M.D. Zammit, T.P. Davis, 'A Comparison of Calibration Procedures for the Analysis of Broad Molecular-weight Distributions Using Size-exclusion Chromatography with Multiple Detection', *Polymer*, **38**(17), 4455–4468 (1997).
79. W.S. Bahary, M.P. Hogan, M. Jilani, M.P. Aronson, 'Size-exclusion Chromatography of Dextrans and Pullulans with Light-scattering, Viscosity, and Refractive-index Detection', *Adv. Chem. Ser.*, **247**, 151–166 (1995).
80. U. Dayal, 'High-temperature SEC Coupled with Malls Detector for Evaluating the End-use Performance of LDPE', *J. Appl. Polym. Sci.*, **53**(12), 1557–1562 (1994).
81. L. Jeng, S.T. Balke, 'Evaluation of Light-scattering Detectors for Size-exclusion Chromatography. 2. Light-scattering Equation Selection', *J. Appl. Polym. Sci.*, **49**(8), 1375–1385 (1993).
82. J.E. Knobloch, P.N. Shaklee, 'Absolute Molecular-weight Distribution of Low-molecular-weight Heparins by Size-exclusion Chromatography with Multiangle Laser-light Scattering Detection', *Anal. Biochem.*, **245**(2), 231–241 (1997).
83. D.J. Nagy, 'Characterization of Poly(vinyl Alcohol) Using SEC Multiangle Laser-light Scattering', *Am. Lab.*, **27**(4), J47+ (1995).
84. M. Nguyen, R. Beckett, L. Pille, D.H. Solomon, 'Determination of Thermal-diffusion Coefficients for Polydisperse Polymers and Microgels by Thfff and SEC-Malls', *Macromolecules*, **31**(20), 7003–7009 (1998).
85. S. Podzimek, 'The Use of GPC Coupled with a Multiangle Laser-light Scattering Photometer for the Characterization of Polymers – On the Determination of Molecular-weight, Size, and Branching', *J. Appl. Polym. Sci.*, **54**(1), 91–103 (1994).
86. W.W. Yau, S.D. Abbott, M.Y. Keating, G.A. Smith, 'A New Stand-alone Capillary Viscometer Used as a Continuous Size Exclusion Chromatographic Detector', *ACS Symp. Ser.*, **352**, 80–103 (1987).
87. S. Mori, 'Design of a Differential Pressure Detector for Use in the Size-exclusion Chromatography of Polymers', *J. Chromatogr.*, **637**(2), 129–135 (1993).
88. R. Lew, P. Cheung, S.T. Balke, T.H. Mourey, 'SEC-Viscometer Detector Systems. 1. Calibration and Determination of Mark–Houwink Constants', *J. Appl. Polym. Sci.*, **47**(9), 1685–1700 (1993).
89. J. Lesec, M. Millequant, T. Havard, 'Single-capillary Viscometer Used for Accurate Determination of Molecular-weights and Mark–Houwink Constants', *ACS Symp. Ser.*, **521**, 220–230 (1993).
90. R. Mendichi, A.G. Schieroni, 'Evaluation of a Single-capillary Viscometer Detector on Line to a SEC System Used with a New Pulse-free Pump', *J. Appl. Polym. Sci.*, **68**(10), 1651–1659 (1998).
91. C.Y. Kuo, A.F. Kah, M.E. Koehler, T. Provder, 'Use of a Viscometric Detector for Size Exclusion

Chromatography – Characterization of Molecular-weight Distribution and Branching in Polymers', *ACS Symp. Ser.*, **352**, 130–154 (1987).
92. S. Pang, A. Rudin, 'Size-exclusion Chromatographic Assessment of Long-chain Branch Frequency in Polyethylenes', *ACS Symp. Ser.*, **521**, 254–269 (1993).
93. C. Jackson, 'Computer-simulation Study of Multidetector Size-exclusion Chromatography of Branched Molecular-mass Distributions', *J. Chromatogr. A*, **662**(1), 1–12 (1994).
94. H.G. Barth, 'Hyphenated Polymer Separation Techniques – Present and Future Role', *Adv. Chem. Ser.*, **247**, 3–11 (1995).
95. J.P. Busnel, C. Degoulet, T. Nicolai, W. Woodley, P. Patin, 'Multiangle Light-scattering and Viscometric Detector for Size-exclusion Chromatography', *J. Phys. III*, **5**(10), 1501–1512 (1995).
96. H. Pasch, K. Rode, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 11. Polymer Blend Separation Using a Reversed Stationary-phase', *Polymer*, **39**(25), 6377–6383 (1998).
97. L. Letot, J. Lesec, C. Quivoron, 'Performance Evaluation of a Continuous Viscometer for High-speed GPC', *J. Liq. Chromatogr.*, **3**(3), 427–438 (1980).
98. J. Lesec, M. Millequant, T. Havard, 'The Use of a Single Capillary Viscometer for Accurate Determination of Molecular-weights and Mark–Houwink Constants', *Abstr. Pap. Am. Chem. Soc.*, **202**, 66-PMSE (1991).
99. S. Pang, A. Rudin, 'Size-exclusion Chromatographic Measurement of PVC Molecular-weight Distributions', *J. Appl. Polym. Sci.*, **49**(7), 1189–1196 (1993).
100. S.T. Balke, R. Thitiratsakul, R. Lew, P. Cheung, T.H. Mourey, 'A Strategy for Interpreting Multidetector Size-exclusion Chromatography Data. II. Applications in Plastic Waste Recovery', *ACS Symp. Ser.*, **521**, 199–219 (1993).
101. P. Cheung, R. Lew, S.T. Balke, T.H. Mourey, 'SEC–Viscometer Detector Systems. 2. Resolution Correction and Determination of Interdetector Volume', *J. Appl. Polym. Sci.*, **47**(9), 1701–1706 (1993).
102. C. Jackson, W.W. Yau, 'Computer-simulation Study of Size-exclusion Chromatography with Simultaneous Viscometry and Light-scattering Measurements', *J. Chromatogr.*, **645**(2), 209–217 (1993).
103. C. Jackson, M.C. Han, H.G. Barth, 'Size-exclusion Chromatography Mobile-phase Optimization for Nylons Using Online Light-scattering and Viscosity Detectors', *Abstr. Pap. Am. Chem. Soc.*, **206**, 133-PMSE (1993).
104. S.T. Balke, T.H. Mourey, C.A. Harrison, 'Number-average Molecular-weight by Size-exclusion Chromatography', *J. Appl. Polym. Sci.*, **51**(12), 2087–2102 (1994).
105. E.K. Fedorov, C.A. Kedik, 'Study of Molecular-mass and Chemical-composition Heterogeneity of Random Copolymers of *N*-vinylpyrrolidone and 2-Methyl-5-vinylpyridine by Liquid-chromatography Methods', *Vysokomol. Soedin.*, **36**(9), 1548–1552 (1994).
106. J. Lesec, M. Millequant, T. Havard, 'High-temperature GPC with a Single Capillary Viscometer', *J. Liq. Chromatogr.*, **17**(5), 1029–1055 (1994).
107. K. Tsutsumi, Y. Okamoto, Y. Tsukahara, 'Radical Polymerization Behavior of Macromonomers. 3. Effect of Macromonomer Concentration', *Polymer*, **35**(10), 2205–2209 (1994).
108. K.G. Suddaby, R.A. Sanayei, K.F. Odriscoll, A. Rudin, 'Eliminating Lag Time-estimation in Multidetector Size-exclusion Chromatography – Calibrating Each Detector Independently', *Adv. Chem. Ser.*, **247**, 79–91 (1995).
109. H. Pasch, K. Rode, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 10. Polymer Blend Analysis Using a Viscosity Detector', *Macromol. Chem. Phys.*, **197**(9), 2691–2701 (1996).
110. F.H. Florenzano, R. Strelitzki, W.F. Reed, 'Absolute, Online Monitoring of Molar-mass During Polymerization Reactions', *Macromolecules*, **31**(21), 7226–7238 (1998).
111. Y. Brun, 'Data Reduction in Multidetector Size-exclusion Chromatography', *J. Liq. Chromatogr. Relat. Technol.*, **21**(13), 1979–2015 (1998).
112. J. Lesec, M. Millequant, 'Star-branched Polymers in Multidetection GPC', *Abstr. Pap. Am. Chem. Soc.*, **206**, 130-PMSE (1993).
113. C.M. Guttman, E.A. Dimarzio, J.F. Douglas, 'Influence of Polymer Architecture and Polymer–Surface Interaction on the Elution Chromatography of Macromolecules Through a Microporous Media', *Macromolecules*, **29**(17), 5723–5733 (1996).
114. A.M. Skvortsov, A.A. Gorbunov, 'Achievements and Uses of Critical Conditions in the Chromatography of Polymers', *J. Chromatogr.*, **507**, 487–496 (1990).
115. A.V. Gorshkov, H. Much, H. Becker, H. Pasch, V.V. Evreinov, S.G. Entelis, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 1. Functionality Type and Composition Distribution in Polyethylene Oxide and Polypropylene Oxide Copolymers', *J. Chromatogr.*, **523**, 91–102 (1990).
116. H. Pasch, C. Brinkmann, H. Much, U. Just, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 2. 2-Dimensional Separations of Poly(ethylene Oxide Block Propyleneoxide)', *J. Chromatogr.*, **623**(2), 315–322 (1992).
117. H. Pasch, C. Brinkmann, Y. Gallot, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 4. Analysis of Poly(styrene-B-methyl Methacrylate)', *Polymer*, **34**(19), 4100–4104 (1993).
118. H. Pasch, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 3. Analysis of Polymer Blends', *Polymer*, **34**(19), 4095–4099 (1993).
119. H. Pasch, M. Augenstein, 'Chromatographic Investigations of Macromolecules in the Critical Range of

Liquid-chromatography. 5. Characterization of Block-copolymers of Decyl and Methyl-methacrylate', *Makromol. Chem. Macromol. Chem. Phys.*, **194**(9), 2533–2541 (1993).

120. H. Pasch, Y. Gallot, B. Trathnigg, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 7. Analysis of the Poly(methyl Methacrylate) Block in Poly(styrene-block-methyl Methacrylate)', *Polymer*, **34**(23), 4986–4989 (1993).
121. H. Pasch, I. Zammert, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. VIII. Analysis of Polyethylene Oxides', *J. Liq. Chromatogr.*, **17**(14–15), 3091–3108 (1994).
122. B. Trathnigg, D. Thamer, X. Yan, B. Maier, H.R. Holzbauer, H. Much, 'Characterization of Ethoxylated Fatty Alcohols Using Liquid-chromatography with Density and Refractive-index Detection. 2. Quantification in Liquid-chromatography Under Critical Conditions', *J. Chromatogr. A*, **665**(1), 47–53 (1994).
123. B. Trathnigg, B. Maier, D. Thamer, 'Analysis of Polyethers by Isocratic HPLC with Universal Detectors. 3. A Study on Reproducibility', *J. Liq. Chromatogr.*, **17**(19), 4285–4302 (1994).
124. H. Pasch, M. Augenstein, B. Trathnigg, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 6. Analysis of the Methyl-methacrylate Block in Block-copolymers of Decyl and Methyl-methacrylate', *Macromol. Chem. Phys.*, **195**(2), 743–750 (1994).
125. V.B. Gancheva, N.G. Vladimirov, R.S. Velichkova, 'Chromatographic Investigations of Polyacetals in the Critical Range of Liquid-chromatography. 1. Analysis of 9-Anthrylmethyl-terminated Poly(1,3,6-trioxacyclooctane)s', *Macromol. Chem. Phys.*, **197**(5), 1757–1770 (1996).
126. V.B. Gancheva, N.G. Vladimirov, R.S. Velichkova, 'Chromatographic Investigations of Polyacetals in the Critical Range of Liquid-chromatography. 2. Determination of Functionality and Molecular-weight Distributions in Alpha,Omega-bis(9-anthrylmethylpoly(1,3,6-trioxacyclooctane)s Obtained by Active Chain-end Mechanism', *Macromol. Chem. Phys.*, **197**(5), 1771–1780 (1996).
127. K. Rissler, 'High-performance Liquid-chromatography and Detection of Polyethers and their Mono(Carboxy)-Alkyl-substituted and Arylalkyl-substituted Derivatives', *J. Chromatogr. A*, **742**(1–2), 1–54 (1996).
128. H. Pasch, K. Rode, N. Chaumien, 'Chromatographic Investigations of Macromolecules in the Critical Range of Liquid-chromatography. 9. Separation of Methacrylate-based Polymer Blends', *Polymer*, **37**(18), 4079–4083 (1996).
129. H. Yun, S.V. Olesik, E.H. Marti, 'Polymer Characterization Using Packed Capillary Size-exclusion and Critical Adsorption Chromatography Combined with Maldi-ToF Mass-spectrometry, *J. Microcolumn. Sep.*, **11**(1), 53–61 (1999).
130. A.M. Skvortsov, A.A. Gorbunov, D. Berek, B. Trathnigg, 'Liquid-chromatography of Macromolecules at the Critical Adsorption Point – Behavior of a Polymer-chain Inside Pores', *Polymer*, **39**(2), 423–429 (1998).
131. D. Berek, M. Janco, K. Hatada, T. Kitayama, N. Fujimoto, 'Separation of Poly(methyl Methacrylate)s According to their Tacticity II – Chromatographic Investigations of Poly(methyl Methacrylate)s with Different Tacticity at the Critical Adsorption Point', *Polym. J.*, **29**(12), 1029–1033 (1997).
132. D. Berek, M. Janco, G.R. Meira, 'Liquid-chromatography of Macromolecules at the Critical Adsorption Point. II. Role of Column Packing – Bare Silica-gel', *J. Polym. Sci. A, Polym. Chem.*, **36**(9), 1363–1371 (1998).
133. H. Pasch, B. Trathnigg, *HPLC of Polymers*, Springer, Berlin, 1997.
134. J.R. Craven, C. Booth, F. Heatley, R.H. Mobbs, G.A. Morris, C.V. Nicholas, R. Webster, D.J. Wilson, 'Molecular Characterization of Oxymethylene-linked Poly(oxyethylene)', *Br. Polym. J.*, **19**(6), 509–516 (1987).
135. J.R. Craven, C. Booth, D. Jackson, S.P.L. Li, H. Tyrer, 'Effect of End Group and Solvent on Gel-permeation Chromatography of Oligo(oxyethylene) Compounds', *J. Chromatogr.*, **387**, 233–240 (1987).
136. R.O. Bielsa, G.R. Meira, 'Linear Copolymer Analysis with Dual-detection Size Exclusion Chromatography – Correction for Instrumental Broadening', *J. Appl. Polym. Sci.*, **46**(5), 835–845 (1992).
137. M. Netopilik, 'Combined Effect of Interdetector Volume and Peak Spreading in Size-exclusion Chromatography with Dual Detection', *Polymer*, **38**(1), 127–130 (1997).
138. J.H. Knox, 'Band Spreading in Chromatography – A Personal View', *Adv. Chromatogr.*, **38**, 1–49 (1998).
139. B. Trathnigg, D. Thamer, X. Yan, 'Analysis of Polyethers by Isocratic HPLC with Universal Detectors. 4. Comparison of Size Exclusion Chromatography and Liquid Adsorption Chromatography for the Analysis of Poly(propylene Glycol)s', *Int. J. Polym. Anal. Char.*, **1**, 35–47 (1995).
140. W.J. Simonsick, L. Prokai, 'Size-exclusion Chromatography with Electrospray Mass-spectrometric Detection', *Adv. Chem. Ser.*, **247**, 41–56 (1995).
141. M.W.F. Nielen, S. Malucha, 'Characterization of Polydisperse Synthetic-polymers by Size-exclusion Chromatography Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass-spectrometry', *Rapid Commun. Mass Spectrom.*, **11**(11), 1194–1204 (1997).
142. M.W.F. Nielen, 'Polymer Analysis by Microscale Size-exclusion Chromatography Maldi Time-of-flight Mass-spectrometry with a Robotic Interface', *Anal. Chem.*, **70**(8), 1563–1568 (1998).
143. M.S. Montaudo, C. Puglisi, F. Samperi, G. Montaudo, 'Application of Size-exclusion Chromatography Matrix-assisted-laser-desorption/Ionization Time-of-flight to the

Determination of Molecular Masses in Polydisperse Polymers', *Rapid Commun. Mass Spectrom.*, **12**(9), 519–528 (1998).
144. M. Nagata, 'Size-exclusion Chromatography with a Multipore Column', *Bunseki Kagaku*, **47**(11), 895–896 (1998).
145. H. Pasch, K. Rode, 'Use of Matrix-assisted Laser-desorption/Ionization Mass-spectrometry for Molar Mass-sensitive Detection in Liquid-chromatography of Polymers', *J. Chromatogr. A*, **699**(1–2), 21–29 (1995).
146. K. Ute, R. Niimi, S.Y. Hongo, K. Hatada, 'Direct Determination of Molecular-weight Distribution by Size-exclusion Chromatography with 750 MHz H-1-NMR Detection (Online SEC-NMR)', *Polym. J.*, **30**(5), 439–443 (1998).
147. B.R. Johnson, K.D. Bartle, M. Cocksedge, A.A. Herod, R. Kandiyoti, 'A Test of the Maldi-MS Calibration of SEC with N-methyl-2-pyrrolidinone for Coal-derived Materials', *Fuel*, **77**(14), 1527–1531 (1998).
148. C. Jackson, H.G. Barth, 'Critical-evaluation of Universal Calibration in Size-exclusion Chromatography of Synthetic-polymers', *Abstr. Pap. Am. Chem. Soc.*, **214**, 5-PMSE (1997).
149. C. Tsitsilianis, A. Dondos, 'Determination of Polymer Molecular-weight by GPC in the Low-molecular-weight Region', *J. Liq. Chromatogr.*, **13**(15), 3027–3037 (1990).
150. C.A. Harrison, T.H. Mourey, 'Number-average Molecular-weight and Functionality of Poly(tetramethylene Glycol) by Multidetector SEC', *J. Appl. Polym. Sci.*, **56**(2), 211–220 (1995).
151. R.O. Bielsa, M.C. Brandolini, L. Akcelrud, G.R. Meira, 'Distributions of Functionality and of Molecular-weights in an Hydroxyl-terminated Polybutadiene by Dual-detection Size-exclusion Chromatography', *J. Appl. Polym. Sci.*, **54**(13), 2125–2134 (1994).
152. H. Nakahara, S. Kinugasa, S. Hattori, M. Mukouyama, T. Hayashi, 'The Molecular-weight and the Refractive-index Measurements of Fractionated Oligo(methyl Acrylate) and Oligo(acrylic acid)', *Bunseki Kagaku*, **42**(5), 273–278 (1993).
153. S.D. Cook, V.S. Sible, 'Detector Response as a Function of Molecular-weight and its Effect on Size-exclusion Chromatography Molecular-weight Determination of Polymer Distributions', *Eur. Polym. J.*, **33**(2), 163–168 (1997).
154. B. Trathnigg, S. Feichtenhofer, M. Kollroser, 'Quantitation in Liquid-chromatography of Polymers – Size-exclusion Chromatography with Dual Detection', *J. Chromatogr. A*, **786**(1), 75–84 (1997).
155. B. Trathnigg, X. Yan, 'Polymer Analysis Using Size-exclusion Chromatography Coupled with Density and Refractive Index Detectors. V. Application to Low-molecular-weight Polyethers', *J. Appl. Polym. Sci.: Appl. Polym. Symp.*, **52**, 193–203 (1993).
156. H.C. Lee, M. Ree, T.Y. Chang, 'Characterization of Binary Polymer Mixtures by Size-exclusion Chromatography with Multiple Detection', *Polymer*, **36**(11), 2215–2218 (1995).
157. S. Mori, T. Suzuki, 'Problems in Determining Compositional Heterogeneity of Copolymers by Size-exclusion Chromatography and UV-Ri Detection System', *J. Liq. Chromatogr.*, **4**(10), 1685–1696 (1981).
158. G.V. Lukyanchikov, T.N. Prudskova, B.M. Prudskov, V.V. Kireev, 'Correctness of the Determination of Composition of Copolymers and Polymer Blends by Gel-permeation Chromatography with Double Detector', *Vysokomol. Soedin.*, **38**(11), 1932–1936 (1996).
159. B. Trathnigg, 'Determination of Chemical-composition of Polymers by SEC with Coupled Density and Ri Detection. 1. Polyethylene-glycol and Polypropylene Glycol', *J. Liq. Chromatogr.*, **13**(9), 1731–1743 (1990).
160. B. Trathnigg, 'Determination of Chemical-composition of Polymers by Size-exclusion Chromatography with Coupled Density and Refractive-index Detection. 3. Polyethylene Oxide and Polytetrahydrofuran', *J. Chromatogr.*, **552**(1–2), 507–516 (1991).
161. B. Trathnigg, D. Berek, R. Murgasova, 'Quantitation in SEC of Polymers by Multiple Detector SEC', *Abstr. Pap. Am. Chem. Soc.*, **214**, 2-PMSE (1997).
162. M. Netopilik, 'Effect of Local Polydispersity in Size-exclusion Chromatography with Dual Detection', *J. Chromatogr. A*, **793**(1), 21–30 (1998).
163. E. Meehan, S. Odonohue, 'Characterization of Block-copolymers Using Size-exclusion Chromatography with Multiple Detectors', *Adv. Chem. Ser.*, **247**, 243–251 (1995).
164. A. Revillon, 'Interpretation of the Gel-permeation Chromatography Behavior of 3 Types of Copolymers', *J. Liq. Chromatogr.*, **3**(8), 1137–1150 (1980).
165. F. Gores, P. Kilz, 'Copolymer Characterization Using Conventional Size-exclusion Chromatography and Molar-mass-sensitive Detectors', *ACS Symp. Ser.*, **521**, 122–148 (1993).
166. L.H. Garciarubio, 'Multiple Detectors in Size Exclusion Chromatography – Signal Analysis', *ACS Symp. Ser.*, **352**, 220–239 (1987).
167. M.D. Zammit, T.P. Davis, K.G. Suddaby, 'Factors Influencing Detector Matching in Multidetector SEC – Solvent and Concentration Effects', *Polymer*, **39**(23), 5789–5798 (1998).
168. M. Guaita, O. Chiantore, 'Simulation of Size Exclusion Chromatograms from Viscometry and Refractometry Detectors – An Analysis of the Influence of Concentration Errors on Reliability of Estimated Parameters', *J. Liq. Chromatogr.*, **16**(3), 633–646 (1993).
169. R. Thitiratsakul, S.T. Balke, T.H. Mourey, 'Combining Detectors in Size Exclusion Chromatography. Part 1. Interdetector Volume', *Int. J. Polym. Anal. Char.*, **2**(4), 345–357 (1996).

170. R. Thitiratsakul, S.T. Balke, 'Combining Detectors in Size Exclusion Chromatography. Part 2. Peak Shape Changes', *Int. J. Polym. Anal. Char.*, **2**, 359–377 (1996).
171. B. Trathnigg, 'Determination of MWD and Chemical-composition of Polymers by Chromatographic Techniques', *Prog. Polym. Sci.*, **20**(4), 615–650 (1995).
172. B. Trathnigg, B. Maier, G. Schulz, R.-P. Krueger, U. Just, 'Characterization of Polyethers Using Different Methods: SFC, LAC, and SEC with Different Detectors, and MALDI-TOF-MS', *J. Appl. Polym. Sci.: Appl. Polym. Symp.*, **110**, 231–240 (1996).
173. A.H. Silver, H.T. Kalinoski, 'Comparison of High-temperature Gas-chromatography and CO_2 Supercritical Fluid Chromatography for the Analysis of Alcohol Ethoxylates', *J. Am. Oil Chem. Soc.*, **69**(7), 599–608 (1992).
174. U. Just, H.R. Holzbauer, M. Resch, 'Molar-mass Determination of Oligomeric Ethylene-oxide Adducts Using Supercritical-fluid Chromatography and Matrix-assisted Laser Desorption-ionization Time-of-flight Mass-spectrometry', *J. Chromatogr. A*, **667**(1–2), 354–360 (1994).
175. F. Guerrero, J.L. Rocca, 'Retention Behavior of Cyanodecyl-bonded Silica for SFC Analysis of Alkyl Polyethyleneglycol Ethers', *Chromatographia*, **40**(9–10), 563–570 (1995).
176. D.J. Pyo, K. Lee, H. Lee, H. Kim, M.L. Lee, 'New Analytical Method Using Polar Modifiers in Supercritical-fluid Chromatography and its Application to the Separation of Fatty Alcohol Ethers', *Bull. Kor. Chem. Soc.*, **17**(6), 496–498 (1996).
177. H. Pasch, H. Kruger, H. Much, U. Just, 'Simultaneous Determination of Functionality Type and Molar Mass-distribution of Oligo(1,3,6-trioxocane)s by Supercritical Fluid Chromatography', *Polymer*, **33**(18), 3889–3892 (1992).
178. X.J.J. Lu, M.R. Myers, W.E. Artz, 'Supercritical Fluid Chromatographic Analysis of the Propoxylated Glycerol Esters of Oleic-acid', *J. Am. Oil Chem. Soc.*, **70**(4), 355–360 (1993).
179. M. Lafosse, B. Herbreteau, L. Morinallory, 'Supercritical-fluid Chromatography of Carbohydrates', *J. Chromatogr. A*, **720**(1–2), 61–73 (1996).
180. S.L. Hansen, M.R. Myers, W.E. Artz, 'Nonvolatile Components Produced in Triolein During Deep Fat Frying', *J. Am. Oil Chem. Soc.*, **71**(11), 1239–1243 (1994).
181. B. Trathnigg, D. Thamer, X. Yan, S. Kinugasa, 'Analysis of Polyethers by Isocratic HPLC with Universal Detectors. 1. Optimization of Chromatographic Conditions', *J. Liq. Chromatogr.*, **16**(12), 2439–2452 (1993).
182. B. Trathnigg, M. Kollroser, A. Gorbunov, A. Skvortsov, 'Liquid Adsorption Chromatography of Polyethers – Theory and Experiment', *J. Chromatogr. A*, **761**(1–2), 21–34 (1997).
183. A. Gorbunov, A. Skvortsov, B. Trathnigg, M. Kollroser, M. Parth, 'Reversed-phase High-performance Liquid-chromatography of Polyethers – Comparison with a Theory for Flexible-chain Macromolecules', *J. Chromatogr. A*, **798**(1–2), 187–201 (1998).
184. B. Trathnigg, B. Maier, A. Gorbunov, A. Skvortsov, 'Theory of Liquid Adsorption Chromatography of Macromolecules Applied to Polyethylene-glycol and Fatty Alcohol Ethoxylates', *J. Chromatogr. A*, **791**(1–2), 21–35 (1997).
185. K. Rissler, U. Fuchslueger, H.J. Grether, 'Separation of Polyethylene-glycol Oligomers on Normal-phase and Reversed-phase Materials by Gradient High-performance Liquid-chromatography and Detection by Evaporative Light-scattering – A Comparative-study', *J. Liq. Chromatogr.*, **17**(14–15), 3109–3132 (1994).
186. K. Rissler, U. Fuchslueger, 'Separation of Polybutylene Glycols on C-18 and C-4 Stationary Phases', *J. Liq. Chromatogr.*, **17**(13), 2791–2808 (1994).
187. K. Rissler, 'Separation of Acetylated Polypropylene Glycol Diamines and Triamines by Gradient Reversed-phase High-performance Liquid-chromatography and Evaporative Light-scattering Detection', *J. Chromatogr. A*, **667**(1–2), 167–174 (1994).
188. C. Sun, M. Baird, J. Simpson, 'Determination of Poly-(ethylene Glycol)s by Both Normal-phase and Reversed-phase Modes of High-performance Liquid', *J. Chromatogr. A*, **800**(2), 231–238 (1998).
189. Y. Mengerink, H.C.J. Deman, S. Van der Wal, 'Use of an Evaporative Light-scattering Detector in Reversed-phase High-performance Liquid-chromatography of Oligomeric Surfactants', *J. Chromatogr.*, **552**(1–2), 593–604 (1991).
190. N. Marquez, R.E. Anton, A. Usubillaga, J.L. Salager, 'Optimizations of HPLC Conditions to Analyze Widely Distributed Ethoxylated Alkylphenol Surfactants', *J. Liq. Chromatogr.*, **17**(5), 1147–1169 (1994).
191. D.F. Anghel, M. Balcan, A. Voicu, M. Elian, 'Analysis of Alkylphenol-based Nonionic Surfactants by High-performance Liquid-chromatography', *J. Chromatogr. A*, **668**(2), 375–383 (1994).
192. C. Sun, M. Baird, H.A. Anderson, D.L. Brydon, 'Separation of Broadly Distributed Nonylphenol Ethoxylates and Determination of Ethylene-oxide Oligomers in Textile Lubricants and Emulsions by High-performance Liquid-chromatography', *J. Chromatogr. A*, **731**(1–2), 161–169 (1996).
193. P. Jandera, J. Urbanek, B. Prokes, H. Blazkovabrunova, 'Chromatographic Behavior of Oligoethylene Glycol Nonylphenyl Ether Anionic Surfactants in Normal-phase High-performance Liquid-chromatography', *J. Chromatogr. A*, **736**(1–2), 131–140 (1996).
194. H.J.A. Philipsen, B. Klumperman, A.L. German, 'Characterization of Low-molar-mass Polymers by Gradient Polymer Elution Chromatography. 1. Practical Parameters and Applications of the Analysis of Polyester Resins Under Reversed-phase Conditions', *J. Chromatogr. A*, **746**(2), 211–224 (1996).

195. H.J.A. Philipsen, H.A. Claessens, H. Lind, B. Klumperman, A.L. German, 'Study on the Retention Behavior of Low-molar-mass Polystyrenes and Polyesters in Reversed-phase Liquid-chromatography by Evaluation of Thermodynamic Parameters', *J. Chromatogr. A*, **790**(1–2), 101–116 (1997).
196. R.A. Shalliker, P.E. Kavanagh, I.M. Russell, 'Separations of High-molecular-mass Polystyrenes on Different Pore-size and Particle-size Reversed-phase Columns in Dichloromethane-acetonitrile', *J. Chromatogr. A*, **664**(2), 221–228 (1994).
197. M. Petro, F. Svec, J.M.J. Frechet, 'Molded Continuous Poly(styrene-co-divinylbenzene) Rod as a Separation Medium for the Very Fast Separation of Polymers – Comparison of the Chromatographic Properties of the Monolithic Rod with Columns Packed with Porous and Nonporous Beads in High-performance Liquid-chromatography of Polystyrenes', *J. Chromatogr. A*, **752**(1–2), 59–66 (1996).
198. S. Mori, 'Liquid-chromatographic Analysis of Styrene-methacrylate and Styrene-acrylate Copolymers', *J. Liq. Chromatogr.*, **13**(15), 3039–3052 (1990).
199. G. Glockner, D. Wolf, H. Engelhardt, 'Separation of Copoly(styrene Acrylonitrile) Samples According to Composition Under Reversed Phase Conditions', *Chromatographia*, **32**(3–4), 107–112 (1991).
200. T.C. Schunk, 'Chemical-composition Separation of Synthetic-polymers by Reversed-phase Liquid-chromatography', *J. Chromatogr. A*, **656**(1–2), 591–615 (1993).
201. T. Kawai, M. Akashima, S. Teramachi, 'Compositional Fractionation of Poly(methyl Methacrylate)-graft-polydimethylsiloxane by Reversed-phase High-performance Liquid-chromatography', *Polymer*, **36**(14), 2851–2852 (1995).
202. S. Teramachi, S. Sato, H. Shimura, S. Watanabe, Y. Tsukahara, 'Chemical-composition Distribution of Poly(methyl Methacrylate)-graft-polystyrene Prepared by a Macromonomer Technique – Effect of Graft Length', *Macromolecules*, **28**(18), 6183–6187 (1995).
203. H.C. Lee, T.Y. Chang, 'Polymer Molecular-weight Characterization by Temperature-gradient High-performance Liquid-chromatography', *Polymer*, **37**(25), 5747–5749 (1996).
204. W. Lee, H.C. Lee, T.Y. Chang, S.B. Kim, 'Characterization of Poly(methyl Methacrylate) by Temperature-gradient Interaction Chromatography with Online Light-scattering Detection', *Macromolecules*, **31**(2), 344–348 (1998).
205. J. Borrajo, C. Cordon, J.M. Carella, S. Toso, G. Goizueta, 'Modeling the Fractionation Process in Tref Systems – Thermodynamic Simple Approach', *J. Polym. Sci., Part B: Polym. Phys.*, **33**(11), 1627–1632 (1995).
206. J.B.P. Soares, A.E. Hamielec, 'Temperature Rising Elution Fractionation of Linear Polyolefins', *Polymer*, **36**(8), 1639–1654 (1995).
207. J.L. Viovy, J. Lesec, 'Separation of Macromolecules in Gels – Permeation Chromatography and Electrophoresis', *Adv. Polym. Sci.*, **114**, 1–41 (1994).
208. T. Okada, 'Nonaqueous Capillary Electrophoretic Separation of Polyethers and Evaluation of Weak Complex-formation', *J. Chromatogr. A*, **695**(2), 309–317 (1995).
209. J. Bullock, 'Application of Capillary Electrophoresis to the Analysis of the Oligomeric Distribution of Polydisperse Polymers', *J. Chromatogr.*, **645**(1), 169–177 (1993).
210. R.A. Wallingford, 'Oligomeric Separation of Ionic and Nonionic Ethoxylated Polymers by Capillary Gel-electrophoresis', *Anal. Chem.*, **68**(15), 2541–2548 (1996).
211. K. Heinig, C. Vogt, G. Werner, 'Separation of Nonionic Surfactants by Capillary Electrophoresis and High-performance Liquid-chromatography', *Anal. Chem.*, **70**(9), 1885–1892 (1998).
212. J.P. Barry, D.R. Radtke, W.J. Carton, R.T. Anselmo, J.V. Evans, 'Analysis of Ethoxylated Polymers by Capillary Electrophoresis in UV-transparent Polymer Networks and by Matrix-assisted-laser-desorption/Ionization Time-of-flight Mass-spectrometry', *J. Chromatogr. A*, **800**(1), 13–19 (1998).
213. D.M. Parees, S.D. Hanton, P.A.C. Clark, D.A. Willcox, 'Comparison of Mass-spectrometric Techniques for Generating Molecular-weight Information on a Class of Ethoxylated Oligomers', *J. Am. Soc. Mass Spectrom.*, **9**(4), 282–291 (1998).
214. H.M. Burger, H.M. Muller, D. Seebach, K.O. Bornsen, M. Schar, H.M. Widmer, 'Matrix-assisted Laser-desorption and Ionization as a Mass-spectrometric Tool for the Analysis of Poly((R)-3-Hydroxybutanoates) – Comparison with Gel-permeation Chromatography', *Macromolecules*, **26**(18), 4783–4790 (1993).
215. D.C. Schriemer, L.A. Li, 'Mass Discrimination in the Analysis of Polydisperse Polymers by Maldi Time-of-flight Mass-spectrometry. 1. Sample Preparation and Desorption/Ionization Issues', *Anal. Chem.*, **69**(20), 4169–4175 (1997).
216. T. Yalcin, Y.Q. Dai, L. Li, 'Matrix-assisted-laser-desorption/Ionization Time-of-flight Mass-spectrometry for Polymer Analysis – Solvent Effect in Sample Preparation', *J. Am. Soc. Mass Spectrom.*, **9**(12), 1303–1310 (1998).
217. H.H. Zhu, T. Yalcin, L. Li, 'Analysis of the Accuracy of Determining Average Molecular-weights of Narrow Polydispersity Polymers by Matrix-assisted Laser-desorption Ionization Time-of-flight Mass-spectrometry', *J. Am. Soc. Mass Spectrom.*, **9**(4), 275–281 (1998).
218. C.N. McEwen, C. Jackson, B.S. Larsen, 'Instrumental Effects in the Analysis of Polymers of Wide Polydispersity by Maldi Mass-spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **160**(1–3), 387–394 (1997).
219. A.T. Jackson, H.T. Yates, C.I. Lindsay, Y. Didier, J.A. Segal, J.H. Scrivens, G. Critchley, J. Brown, 'Utilizing Time-lag Focusing Ultraviolet-matrix-assisted Laser

Desorption/Ionisation Mass-spectrometry for the End Group-analysis of Synthetic-polymers', *Analusis*, **26**(10), M31–M35 (1998).

220. G. Schulz, H. Much, H. Kruger, C. Wehrstedt, 'Determination of Functionality and Molecular-weight Distribution by Orthogonal Chromatography', *J. Liq. Chromatogr.*, **13**(9), 1745–1763 (1990).
221. R.P. Kruger, H. Much, G. Schulz, 'Determination of Functionality and Molar-mass Distribution of Aliphatic Polyesters by Orthogonal Liquid-chromatography. 1. Off-line Investigation of Poly(1,6-hexanediol Adipates)', *J. Liq. Chromatogr.*, **17**(14–15), 3069–3090 (1994).
222. S. Mori, 'Size-exclusion Chromatography and Nonexclusion Liquid-chromatography for Characterization of Styrene Copolymers', *Adv. Chem. Ser.*, **247**, 211–222 (1995).
223. S. Mori, 'Trends of Polymer Analysis by Liquid-chromatography', *Bunseki Kagaku*, **47**(10), 625–643 (1998).
224. T.C. Schunk, 'Composition Distribution Separation of Methyl-methacrylate Methacrylic-acid Copolymers by Normal-phase Gradient Elution High-performance Liquid-chromatography', *J. Chromatogr. A*, **661**(1–2), 215–226 (1994).
225. B. Trathnigg, M. Kollroser, 'Mapping of Polyethers Using Two-dimensional Liquid Chromatography with Coupled Density and RI Detection', *Int. J. Polym. Anal. Char.*, **1**, 301–313 (1995).
226. B. Trathnigg, M. Kollroser, M. Parth, '2D-LC of Functional Polyethers', *Abstr. Pap. Am. Chem. Soc.*, **214**, 65-PMSE (1997).
227. J. Blomberg, E.P.C. Mes, P.J. Schoenmakers, J.J.B. Vanderdoes, 'Characterization of Complex Hydrocarbon Mixtures Using Online Coupling of Size-exclusion Chromatography and Normal-phase Liquid-chromatography to High-resolution Gas-chromatography', *J. High Res. Chromatogr.*, **20**(3), 125–130 (1997).
228. A.C. Vanasten, R.J. Vandam, W.T. Kok, R. Tijssen, H. Poppe, 'Determination of the Compositional Heterogeneity of Polydisperse Polymer Samples by the Coupling of Size-exclusion Chromatography and Thermal Field-flow Fractionation', *J. Chromatogr. A*, **703**(1–2), 245–263 (1995).
229. S.J. Jeon, M.E. Schimpf, 'Molecular-weight and Composition of Styrene/Ethylene Oxide Blends and Copolymers Using 2-Dimensional Separations by Size-exclusion Chromatography and Thermal Field-flow Fractionation', *Abstr. Pap. Am. Chem. Soc.*, **214**, 67-PMSE (1997).
230. T. Suortti, 'Coupled Size-exclusion Chromatography Anion-exchange Chromatography in the Analysis of Polysaccharides and Oligosaccharides', *J. Chromatogr. A*, **763**(1–2), 331–335 (1997).
231. M. Janco, T. Prudskova, D. Berek, 'Liquid-chromatography of Polymer Mixtures Applying Combination of Exclusion and Full Adsorption Mechanisms. 1. Analysis of Polystyrene in its Mixture with Polymethylmethacrylate Single-column Single Eluent Approach', *J. Appl. Polym. Sci.*, **55**(3), 393–397 (1995).
232. M. Janco, D. Berek, T. Prudskova, 'Liquid-chromatography of Polymer Mixtures Applying a Combination of Exclusion and Full Adsorption Mechanisms. 2. Eluent Switching Approach', *Polymer*, **36**(17), 3295–3299 (1995).
233. S.H. Nguyen, D. Berek, 'Liquid-chromatography of Polymer Mixtures by a Combination of Exclusion and Full Adsorption Mechanisms. 3-Component and 4-Component Polymer Blends', *Chromatographia*, **48**(1–2), 65–70 (1998).
234. S.H. Nguyen, D. Berek, 'Full Adsorption-desorption/SEC Coupling in Characterization of Complex Polymers', *Abstr. Pap. Am. Chem. Soc.*, **214**, 93-PMSE (1997).
235. R.P. Kruger, H. Much, G. Schulz, E. Rikowski, 'Characterization of Si Polymers by Liquid-chromatographic Methods in Coupling with Maldi-ToF-MS', *Monatsh. Chem.*, **130**(1), 163–174 (1999).

Supercritical Fluid Chromatography of Polymers

Koichi Ute
Osaka University, Toyonaka, Japan

1 Introduction 1
2 History 2
3 Methods and Techniques 2
3.1 Supercritical Mobile Phase 2
3.2 Columns and Stationary Phase 5
3.3 Detectors 6
4 Supercritical Fluid Chromatography Separations of Oligomers and Low-molecular-weight Polymers 8
4.1 Determination of Molecular Weight and Molecular Weight Distribution 8
4.2 Separation According to Chemical Structures of Polymers 9
4.3 Fractionation of Individual Polymer Homologs 10
5 Supercritical Fluid Chromatography Separations of High Molecular Weight Polymers 13
5.1 Separation of Polymers by Supercritical Fluid Chromatography in the Adsorption Mode 13
5.2 Size-exclusion Chromatography using Supercritical Mobile Phase 14
5.3 Separation of Polymers by Supercritical Fluid Chromatography at the Critical Adsorption Point 14
6 Analysis of Common Polymer Additives by Supercritical Fluid Chromatography 15
Abbreviations and Acronyms 15
Related Articles 16
References 16

The mobile phase in supercritical fluid chromatography (SFC) is a substance raised above its critical temperature and pressure which exhibits greater solvating properties together with reduced viscosities and higher diffusivities than its liquid state. The eluting power of supercritical mobile phase can be varied widely by controlling pressure, temperature and composition. Thus, SFC offers many advantages over high-performance liquid chromatography (HPLC) in the characterization of polymers and oligomers. SFC of polymeric substances has been operated so far in the three modes of chromatography: adsorption chromatography for the separation of oligomer homologs, size-exclusion chromatography (SEC) for the determination of molecular weight distribution, and adsorption chromatography at critical conditions for the characterization of functionality type distribution. The combination of SFC with supercritical fluid extraction (SFE) is useful for the analysis of polymer additives, such as stabilizers and plasticizers, in an industrial polymer material. SFC is compatible with most conventional HPLC and gas chromatography (GC) detectors, and can also be interfaced with infrared (IR) spectroscopy and mass spectrometry (MS), etc.

1 INTRODUCTION

SFC utilizes fluids or mixtures of fluids above their critical temperature and critical pressure. Supercritical fluids have higher solute diffusivities and lower viscosities than liquids and more solvating power than gases. Thus, SFC can be used to separate high molecular weight samples which are not suitable for GC analysis. The mass transport properties of supercritical fluids compared with those of liquids offer an advantage of greater speed or greater numbers of theoretical plates than could be achieved with HPLC. The solubilization advantage of supercritical fluid is augmented by the ability to readily vary the eluting power of the fluid by controlling density through changes in pressure or temperature.

Of the many fields where SFC has attracted considerable attention, the field of polymer analysis is one of the most important. The interest in SFC of polymers focuses on different aspects:

1. The separation of individual homologs in oligomer samples. In this area, the determination of the molecular weight distribution is of particular interest.
2. Assay of parallel homologous series by separation according to chemical structures, for example, end-groups. Such information may be interesting for the characterization of polymer surfactants, technical silicone oils, macromonomers, prepolymers of thermosetting resins, etc.
3. Fractionation of individual homologs on a preparative scale. Because each homolog is perfectly monodisperse (uniform) and has a definite molecular weight, the series of pure homologs is useful as model compounds for the study of polymer structures and properties, and as reference materials for polymer characterizations.
4. The use of supercritical fluids in different modes of chromatography: the separations of polymers in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

adsorption chromatography, SEC, and chromatography at critical conditions of adsorption. The unique properties of supercritical fluids offer new possibilities in this area.

5. The determination of low-molecular weight compounds in polymers, for example, residual monomers and solvents, oligomers, and additives like stabilizers, plasticizers and surface treatment agents. For the determination of these compounds in a polymer matrix, the combination of SFE and SFC is useful.

2 HISTORY

Since its introduction in 1962 by Klesper et al.,[1] SFC has been shown to possess a large potential for the separation of high molecular weight compounds. In 1969, Jentoft and Gouw[2] reported the first application of SFC to polymer separation. They separated polystyrene with an average molecular weight of 578 into the individual homologs from 2-mer to 18-mer by the pressure-programmed SFC using C8-silica as stationary phase and pentane with 5% methanol as the mobile phase (Figure 1). Fractions were also taken and rechromatographed under the same conditions.[2,3] In the same year, nongradient separation of epoxy resin precursors on a packed column SFC was reported by Sie et al.[4] Hartmann and Klesper,[5,6] and Rogers et al.[7,8] continued to develop SFC with packed columns by applying the technique to polystyrene separations.

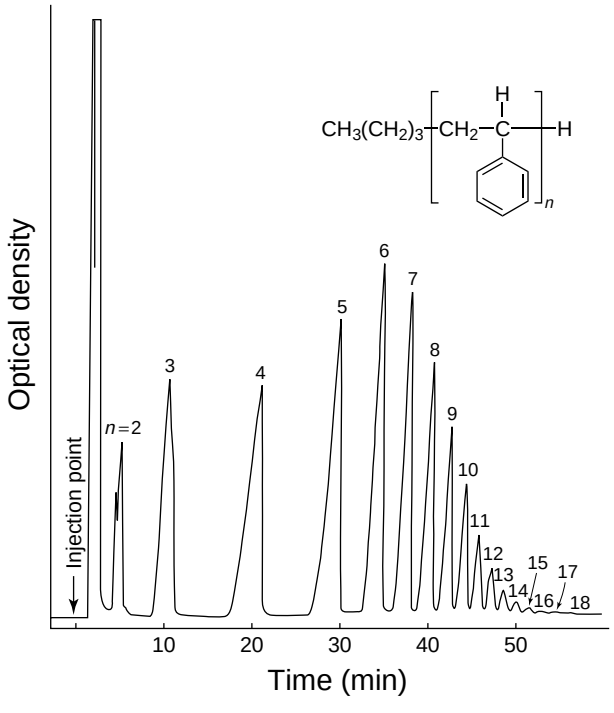

Figure 1 SFC chromatogram of polystyrene with an average molecular weight of 578.[2] (Reproduced with permission from R.E. Jentoft, T.H. Gouw, *J. Polym. Sci. Polym. Lett.*, Copyright © 1969, John Wiley & Sons Inc.)

Until the early 1980s, only a limited number of papers were published on the polymer separation by SFC. One of those papers described the separation of poly(methylphenylsiloxane) (Silicone DC-170) in an analytical[9] or preparative scale.[10] Another interesting example is the separation of polymers by SFC operated in the size-exclusion mode.[11] The exclusion mechanism was demonstrated for polystyrenes with molecular weight up to 20 400 and poly(vinylpyrolidone)s with molecular weight up to 40 000 using 1,1-difluoromethane as the supercritical mobile phase.

Novel combination of supercritical mobile phase with open tubular fused silica capillary column was introduced in 1981.[12] By means of the capillary SFC with supercritical carbon dioxide, poly(oxyethylene) derivatives (so called ''ethoxylated surfactants'') were separated into the individual homologous series.[13] SFC on packed columns was also developed on a new basis in the early 1980s, taking advantage of the numerous improvements of high-pressure chromatographic instrumentation that were primarily achieved in HPLC.[14,15] The explosive growth of SFC applications at the end of the 1980s was mainly due to the introduction of sophisticated commercial instrumentation for capillary and packed column SFC; reviews of commercial SFC instruments are available.[16–18] One of the current research trends in polymer characterization by SFC is the combination with multiple detectors such as the evaporative light scattering detector (ELSD), MS, and Fourier transform infrared (FTIR) spectroscopy.

Progress in SFC separation of polymers and oligomers prior to 1990 has been extensively reviewed by Schmitz and Klesper.[19] White and Houck[20] reviewed the use of SFC for analysis of a variety of complex oligomeric mixtures. Biannual reviews by Chester et al.[21,22] since 1990 provide general information of new SFC applications including polymer analysis. Ute[23] reviewed fractionation of polymer homologs by preparative SFC. Tables 1 and 2 summarize the reported examples of SFC separation of various polymers.

3 METHODS AND TECHNIQUES

3.1 Supercritical Mobile Phase

In the early studies of polymer separation, *n*-pentane (critical temperature 196.6 °C, critical pressure 33.7 bar) was employed as a main component of supercritical mobile phase. The use of propane, butane, hexane,

Table 1 Examples of polymer separation by packed column SFC

Polymer sample	MW range	Column	Mobile phase	Initial conditions	Detection	Ref.
Polyethylene	C18–C64	40 cm × 0.25 mm (Develosil ODS, 7 μm)	CO_2	100 °C, 150 atm	FID	Hirata et al.[24]
cis-1,4-Polyisoprene	$5 < DP < 35$	25 cm × 10 mm (Develosil 100-5, 5 μm)	CO_2–CH_3CH_2OH	130 °C, 200 kg cm^{-2}	UV (210 nm)	Kawahara et al.[25]
Polystyrene	Ave. 810	25 cm × 4.6 mm (LiChrosorb Si 100, 10 μm)	CO_2–1,4-Dioxane	145 °C, 250 bar	UV (254 nm)	Schmitz et al.[26]
Polystyrene	3600–600 000	25 cm × 4.6 mm (LiChrosorb Si 100, 10 μm)	Pentane–1,4-Dioxane	250 °C, 165 bar	UV (254 nm)	Schmitz and Klesper[27]
Poly(4-vinylbiphenyl)		25 cm × 4.6 mm (LiChrosorb Si 60, 10 μm)	Pentane–1,4-Dioxane	270 °C, 45 bar	UV (270 nm)	Schmitz et al.[28]
Poly(2-vinylnaphthalene)		25 cm × 4.6 mm (LiChrosorb Si 60, 10 μm)	Pentane–1,4-Dioxane	270 °C, 51 bar	UV (270 nm)	Schmitz et al.[28]
Poly(*N*-vinylcarbazole)		25 cm × 4.6 mm (LiChrosorb Si 60, 10 μm)	Pentane–1,4-Dioxane	270 °C, 41 bar	UV (295 nm)	Schmitz et al.[28]
Poly(2-vinylpyridine)		25 cm × 4.5 mm (Crosslinked polystyrene PRP-1, 10 μm)	CO_2–CH_3OH	90 °C, 29 MPa,	UV (230 nm)	Gemmel et al.[29]
Poly(4-vinylpyridine)		25 cm × 4.6 mm (LiChrosorb Si 100, 10 μm)	Pentane–1,4-Dioxane	270 °C, 51 bar	UV (254 nm)	Schmitz et al.[30]
Poly(methylphenylsiloxane) [DC-710]	Ave. 2600	25 cm × 1.7 mm (SFPAK ODS, 5 μm)	CO_2–Hexane	180 °C, 212 kg cm^{-2}	UV	Takeuchi and Saito[31]
Poly(oxyethylene) [PEG 1540]	Ave. 1540	25 cm × 4.6 mm (Ultrasphere Si, 5 μm)	CO_2–H_2O–CH_3OH–$(C_2H_5)_3N$	272 atm	ELSD	Brossard et al.[32]
Poly(oxyethylene) [Triton X-165]	$6 < DP < 20$	10 cm × 2 mm (Nucleosil C18, 3 μm)	CO_2–CH_3OH	170 °C, 130 bar	UV (278 nm)	Giorgetti et al.[33]
Poly(oxyethylene)	$1 < DP < 12$	25 cm × 4.5 mm (Crosslinked polystyrene PRP-1, 10 μm)	CO_2–CH_3CN	135 °C, 25 MPa	UV (195 nm)	Gemmel et al.[29]
Poly(propylene oxide)	$\overline{M}_n$ 3010	7 cm × 1 mm (ODS silica Vydac 330 Å, 10 μm)	CO_2	80 °C, 100 atm, 10 atm min^{-1}	FID	Dean and Poole[34]
Poly(oxymethylene)	*DP* 10.1	25 cm × 7.2 mm (Crest SIL, 5 μm)	CO_2–CH_3OH	130 °C, 200 kg cm^{-2}, −2 °C min^{-1}	UV (235 nm)	Ute et al.[35]

(*continued overleaf*)

Table 1 (*continued*)

Polymer sample	MW range	Column	Mobile phase	Initial conditions	Detection	Ref.
Poly(oxyphenylene)	$4 < DP < 7$	400 cm × 2.6 mm (*n*-octane bonded to 120/150 mesh Porasil C)	Pentane–CH_3OH	214 °C, 570 psi, 3.5 psi min^{-1}	UV (275 nm)	Jentoft and Gouw[3]
PMMA	$1 < DP < 100$	25 cm × 10 mm (Develosil 100-5, 5 μm)	CO_2–CH_3OH	125 °C, 200 kg cm^{-2}, −1 °C min^{-1}	UV (235 nm)	Ute et al.[36]
Poly(methyl crotonate)	$5 < DP < 11$	25 cm × 10 mm (Develosil 100-5, 5 μm)	CO_2–CH_2Cl_2	160 °C, 200 kg cm^{-2}, −1 °C min^{-1}	UV (235 nm)	Ute et al.[37]
Poly(ε-caprolactone)	$\overline{M}_n$ 2960	25 cm × 10 mm (Develosil 100-5, 5 μm)	CO_2–CH_3OH	90 °C, 200 kg cm^{-2}, −1.2 °C min^{-1}	UV (200–230 nm)	Ihara et al.[38]
Poly(ε-caprolactone)	Ave. 530	25 cm × 4.5 mm (Crosslinked polystyrene PRP-1, 10 μm)	CO_2–CH_3CN	135 °C, 25 MPa	UV (190 nm)	Gemmel et al.[29]
Poly(D-lactide)	$20 < DP < 48$	25 cm × 10 mm (Develosil 100-5, 5 μm)	CO_2–CH_3OH	55 °C, 200 kg cm^{-2}	UV (210 nm)	Ute[23]
Poly(*n*-butyl isocyanate)	$\overline{M}_n$ 4000, $20 < DP < 60$	25 cm × 10 mm (Develosil 100-5, 5 μm)	CO_2–CH_3CH_2OH	90 °C, 200 kg cm^{-2}	UV (235 nm)	Ute et al.[39]
Epoxy prepolymers		125 mm × 4 mm (Spherisorb CN, 3 μm)	CO_2–CH_3OH	70 °C, 350 bar	UV (280 nm)	Aton et al.[40]
Novolac and Resol	$\overline{M}_n$ 366	25 cm × 1.7 mm (Silica gel-ODS, 5 μm)	CO_2–CH_3CH_2OH	120 °C, 162 kg cm^{-2}, −1 °C min^{-1}	UV (210 nm)	Mori et al.[41]

FID, flame ionization detection; ODS, octadecylsilane; PMMA, poly(methyl methacrylate); UV, ultraviolet; *DP*, degree of polymerization.

Table 2 Examples of polymer separation by capillary SFC

Polymer sample	MW range	Column	Conditions	Detector	Ref.
Polymethylene (*n*-alkanes)	C20–C100 (~1402)	10 m × 50 μm (DB-1)	125 °C, 75 atm, 25 atm min^{-1}	FID, MS	Howthorne and Miller [42]
Poly(2-vinylnaphthalene)	$DP < 10$	10 m × 50 μm (DB-1)	100 °C, 99–400 bar	FID	Schmitz et al.[43]
Polybutadiene	$\overline{M}_n$ 500	5 m × 50 μm (OV)	100 °C, press prog.	MS	Wright et al.[44]
Poly(chlorotrifluoroethylene) [Kel-F Wax]		19.5 m × 100 μm (DB-5)	140 °C, 100 atm, 1.33 atm min^{-1}	FID	White and Houck[20]
Poly(dimethylsiloxane)	$\overline{M}_n$ 2000	10 m × 50 μm (SB-Methyl-100)	100 °C, 1 °C min^{-1}, 0.19–0.945 g ml^{-1}	FID	Later et al.[45]
Poly(ethylene oxide) [Neodol 23]	$0 < DP < 22$	9 m × 100 μm (BP-10)	110 °C	FID	Chester[13]
Poly(propylene oxide)	$\overline{M}_n$ 425	10 m × 63 μm (DB-1)	120 °C, 120 atm, 1.5 atm min^{-1}	FID	White and Houck[20]
Poly(*t*-butyldimethylsilyl acrylate)	900–10 000	10 m × 50 μm (DB-1)	120 °C, 70 atm, 1 atm min^{-1}	FID, MS	Pinkston et al.[46]

In all examples, the mobile phase is neat carbon dioxide.
DB-1 is a 100% dimethyl-polysiloxane stationary phase.
DB-5 is a 95% dimethyl-(5%)-diphenyl-polysiloxane stationary phase.
SE-54 is a 95% dimethyl-(4%)-diphenyl-(1%)-vinyl-polysiloxane stationary phase.

diethyl ether, or ethanol was also examined for the SFC separation of polystyrene.[26]

However, carbon dioxide is currently the fluid of choice in many SFC applications to polymers because of its low critical temperature (31 °C), nontoxic, nonflammable nature, and lack of interference with most detection methods. The solvent strength of carbon dioxide at a density less than 0.25 g cm^{-3} is less than that of perfluorinated alkanes, while at a density of 0.98 g cm^{-3} it exceeds that of hexane, perhaps approaching the solvent strength of methylene chloride.[15] Addition of polar modifiers to carbon dioxide extends its somewhat limited solvent strength, allowing the SFC separation of nonpolar to medium polar polymers. Alcohols, ethers, acetonitrile, and methylene chloride are suitable for modifiers.

When a polymer sample is chromatographed without the use of gradients, i.e. at constant conditions, the retention times of the members of homologous series will increase exponentially with increasing molecular weight which leads to long overall analysis times. Since this increase of retention time is paralleled by an increase of peak widths, detection limits also increase with increasing molecular weight. For compensation of these effects gradient techniques are applied. In SFC, three types of gradients can be used: temperature gradients, pressure/density gradients, and composition gradients. Furthermore the combinations of these gradient types can be used for improving a separation. The gradients chosen were often linear, although a tendency for crowding of the peaks at the end of the program could often be observed. Most of capillary SFC separations are carried out on a density gradient. The gradient conditions in SFC separation of polymers are discussed extensively in the literature.[19,47]

3.2 Columns and Stationary Phase

SFC may be performed either in packed columns or in open tubular capillary columns. Both approaches have been demonstrated numerous times in the literature. In addition, there are a few examples of polymer separation using a hybridization of the two major column types – packed capillary columns.[48,49] There are advantages and disadvantages to each approach.

Packed columns are advantageous when the goal of the separation is to preparatively isolate individual analytes. The sample capacity of packed columns is much greater than capillary columns. It has also been indicated that packed columns may develop more theoretical plates per unit time than capillary columns, which may be important when speed is an important analytical aspect. Another advantage of packed columns is that packed HPLC columns can be used for packed column SFC. Thus, a wide variety of stationary phases are available in packed column form for which no capillary column counterpart currently exists.

Predominantly, nonpolar polymers have been separated on polar as well as nonpolar stationary phases. With increasing polarity of the analyte polar stationary phases like untreated silica become less useful, and better results are obtained on less polar materials. For packed columns, different types of polar and nonpolar stationary phases other than untreated silica are available and used for SFC separation of polymers, e.g. porous glass beads,[50] alumina,[34] surface-modified silica (C_{18}, CN, NH_2, and Diol)[25,31–34,40,41,51,52] and polystyrene beads.[29,53,54] One problem with silica-based materials is the solubility of silica stationary phases in a polar mobile phase which is only slight in liquid polar organic

media but may be pronounced in supercritical eluents. A polystyrene stationary phase is suitable if the polymer is highly crosslinked which precludes swelling and shrinking and gives the material a stable geometry. For example, 2-vinylpyridine oligomers which had been difficult to separate on silica could successfully be chromatographed on a polystyrene stationary phase.[29] As a general rule, the more specific the interactions between analytes and stationary phase, the higher is the selectivity for separation.

The main advantage of capillary column SFC over packed column SFC is that longer capillary columns can be used. This is due to the low pressure drop and open tubular nature of the capillary. Thus, separations using capillary columns can be performed using a greater number of theoretical plates, i.e. increased efficiency. SFC capillary columns are typically fused silica columns of 50–100 μm internal diameter (ID) coated with a bonded and crosslinked polysiloxane stationary phase.

3.3 Detectors

SFC is compatible with many types of conventional HPLC and GC detectors, such as ultraviolet/visible (UV/VIS)

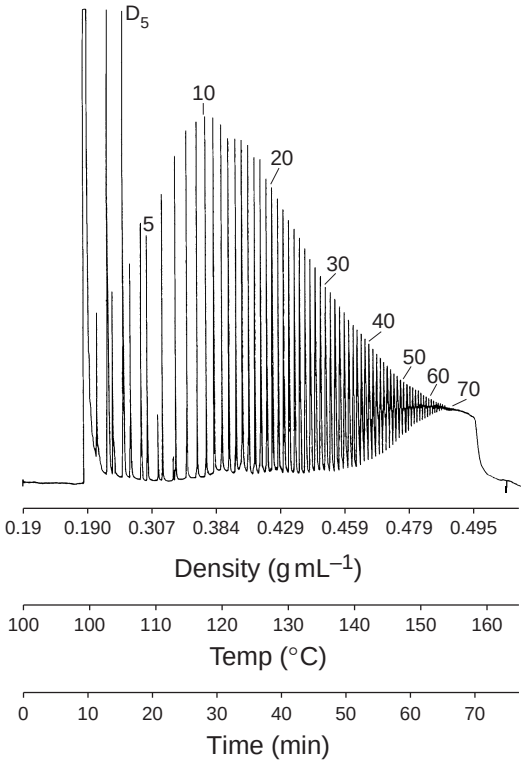

Figure 2 Capillary SFC chromatogram of poly(dimethylsiloxane) using a synchronized asymptotic density/linear temperature program.[45] Conditions: asymptotic density programmed from 0.19 g mL^{-1} (12 min initial hold time) to 0.495 g mL^{-1} with $t_{1/2} = 20$ min; linear temperature programmed synchronously from 100 °C (12 min initial hold time) to 160 °C at 1 °C min^{-1}. (Reproduced by permission of Wiley-VCH.)

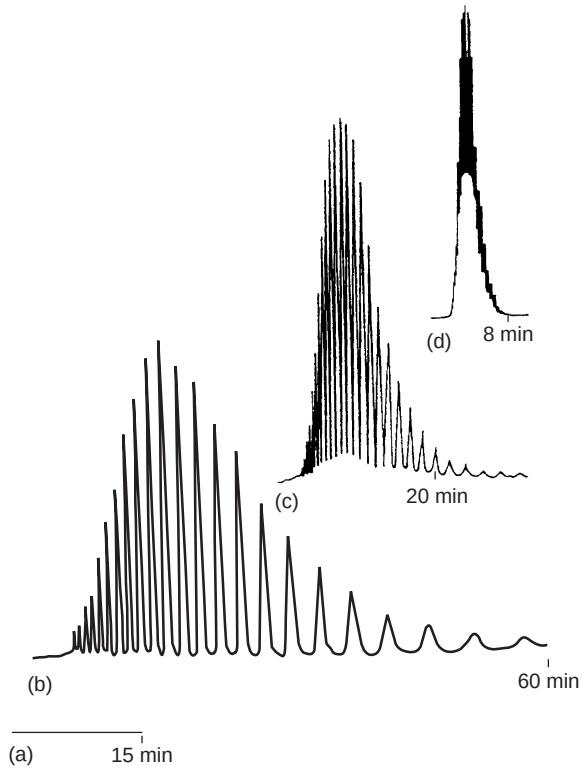

Figure 3 SFC chromatograms of poly(oxyethylene) (PEG 1540) recorded by evaporative light scattering detection.[32] Chromatographic conditions: column, Ultrasphere Si (250 mm × 4.6 mm ID); flow rate, 4.2 mL min^{-1}. (a) $CO_2-CH_3OH-H_2O$ (80 : 18.77 : 1.23, w/w), pressure 272 atm; (b) CO_2-CH_3OH $H_2O-(C_2H_5)_3N$ (80 : 18.60 : 1.22 : 0.18, w/w), pressure 288 atm; (c) $CO_2-CH_3OH-H_2O-(C_2H_5)_3N$ (80 : 18.60 : 1.22 : 0.18, w/w), pressure 288 atm; (d) $CO_2-CH_3OH-H_2O-(C_2H_5)_3N$ (80 : 18.43 : 1.21 : 0.36, w/w), pressure 280 atm. (Reprinted from S. Brossard, M. Lafosse, M. Dreux, 'Comparison of Ethoxylated Alcohols and Polyethylene Glycols by High-performance Liquid Chromatography and Supercritical Fluid Chromatography Using Evaporative Light-scattering Detection', *J. Chromatogr.*, **591**, 149–157 (1992), with permission from Elsevier Science.)

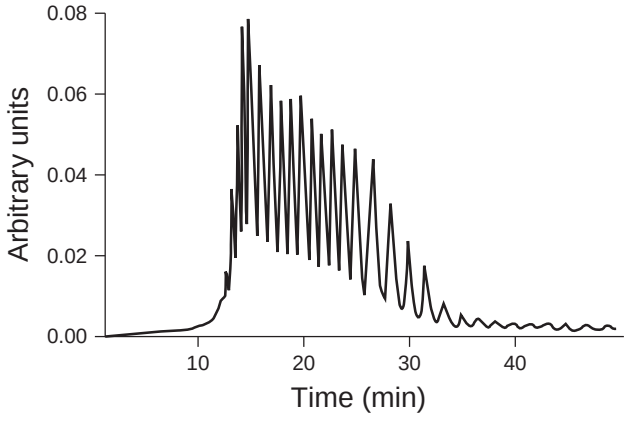

Figure 4 SFC chromatogram obtained by on-line FTIR detection of 3 μg of poly(oxyethylene) (PEG 1500). (Reprinted with permission from D.F. Gurka et al.[76] Copyright 1994, American Chemical Society.)

detection and FID. Typically, packed column SFC system is equipped with an UV detector whereas capillary SFC system is equipped with an FID.

Polymers containing an aromatic chromophore can easily be detected by UV absorbance. High-purity carbon dioxide is transparent over the whole UV range down to 190 nm, and thus polyethers which contain only the weak ether chromophore can also be detected by means of an UV detector.[29] For UV detection at short wavelengths, the choice of appropriate mobile phase is important. Carbon dioxide/methanol as well as carbon dioxide/acetonitrile have successfully been used for UV detected SFC separations of PMMA,[29,33,36,53,55,56] poly(caprolactone),[38,53] poly(oxymethylene),[35] and poly(*n*-butyl isocyanate).[39]

FID has been used numerous times as a universal detector for capillary SFC. However, FID is only applicable when a nonflammable mobile phase such as neat carbon dioxide is used. Therefore, applications in the polymer field are limited to the SFC separations of polymers with low polarity such as paraffin waxes[13,24,42,57,58] or polysiloxanes[24,45,47,58–65] (Figure 2) or separations of low molecular weight polymer adducts such as ethoxylated surfactants.[13,47,57,58,66–73]

ELSD looks promising for a quasi-universal detection in SFC. This method is adequately sensitive (>20 ng) and compatible with modifiers. ELSD was successfully used for the packed column SFC separations of poly(oxyethylene)s[31,32,54] (Figure 3).

Carbon dioxide has IR adsorption that completely blanks out the spectrum from 3800–3500 cm^{-1}, 2500–2200 cm^{-1}, and below 800 cm^{-1}. The remaining regions of the IR spectrum are therefore useful for detection purposes. The application of on-line SFC/FTIR with flow cell interfaces was demonstrated for the separation of poly(dimethylsiloxane)[63] and phenyl-ethoxy-acrylate oligomers.[74] Another approach utilizes FTIR as an off-line detector of SFC. In this approach, analytes are collected uncontaminated on an IR transparent window from the outlet of SFC column as the mobile phase evaporates. IR spectra are measured from the deposited spot using an IR microscope.[59,75,76] Excellent SFC/FTIR chromatograms were obtained for poly(oxyethylene)s of average molecular weights 400, 600, 1000, and 1500 (Figure 4).

Much attention has been focused on SFC systems interfaced to MS detection systems.[42,44,70,73,77–83] Aside from being a very sensitive and highly selective detector, the mass spectra are useful for structure elucidation of analytes. Figure 5 shows the total-ion chromatogram obtained from SFC/MS analysis of polybutadiene with average molecular weight of 500.[44] In this analysis, the

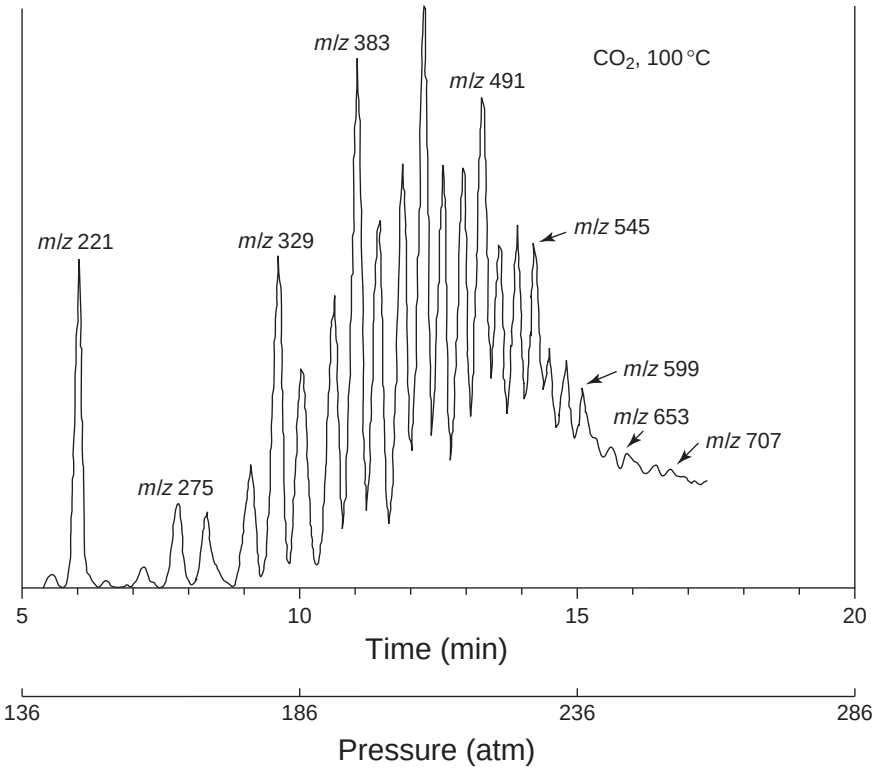

Figure 5 Total-ion chromatogram obtained from the SFC/MS analysis of a 500 average-molecular-weight polybutadiene sample.[44] (Reproduced by permission of Wiley-VCH.)

chromatographic effluent from capillary SFC was directly introduced into the MS source. The isobutane chemical ionization yielded a gentle fragmentation process that produced a dominant protonated molecule, $(M + 1)^+$, for each homolog.

Polymer separation by SFC using a fluorescence[84] or ion-mobility detector[67] has also been reported.

4 SUPERCRITICAL FLUID CHROMATOGRAPHY SEPARATIONS OF OLIGOMERS AND LOW-MOLECULAR-WEIGHT POLYMERS

4.1 Determination of Molecular Weight and Molecular Weight Distribution

As illustrated in Figure 1, SFC of polystyrene separates into 18 individual species. Using the chromatogram and assuming that UV absorption is concentration dependent only (no effects due to molecular weight), one can calculate number average ($\overline{M}_n$) and weight average ($\overline{M}_w$) molecular weights for the polystyrene on the basis of Equations (1) and (2):

$$\overline{M}_n = \left(\frac{\sum W_i}{M_i} \right)^{-1} \quad (1)$$

$$\overline{M}_w = \sum W_i M_i \quad (2)$$

where M_i and W_i denote the molecular weight and weight fraction of species i in the polymer. The results ($\overline{M}_n = 576$, $\overline{M}_w = 649$) were in close agreement with the values by other methods including nuclear magnetic resonance (NMR), vapor pressure osmometry (VPO) and SEC.[85] Close agreement between packed-column SFC/UV and SEC/UV was also observed for PMMAs with $\overline{M}_n$ 900–990.[52]

SFC and MS are complementary methods for characterizing low molecular weight polymers. Both capillary SFC/FID and time-of-flight secondary ion mass spectrometry (TOFSIMS) provide well-resolved signal intensities of low molecular weight poly(dimethylsiloxane)s (Figure 6). To determine the achievable accuracy of these independent analytical methods, a direct comparison was performed on poly(dimethylsiloxane)s in the molecular weight range from 1000 to 10 000. Results obtained by SFC/FID and TOFSIMS fit well up to molecular weight 3000 whereas remarkable differences occur for higher molecular weight regions due to mass discrimination effects. These have been found to be more pronounced for SFC than for TOFSIMS.[64]

A systematic investigation was undertaken as to the performance of different analytical methods, including capillary SFC/FID and matrix-assisted laser desorption/ionization (MALDI) TOFMS (time-of-flight mass spectrometry), in the molecular weight determinations of poly(ethylene glycol) with average molecular weight of 300 and poly(propylene glycol) with average molecular weight of 425. The data from all methods (SFC/FID, MALDI/TOFMS, SEC/density, SEC/refractive index (RI), SEC/ELSD, reversed and normal phase HPLC/density, /RI, /ELSD) agreed quite well.[86]

A poly(ethylene oxide)–poly(propylene oxide) block copolymer, $HO(EO)_{n1}(PO)_m(EO)_{n2}OH$, was characterized with respect to molecular weight distribution and block length of the individual blocks using two-dimensional chromatographic techniques. In the first dimension the block copolymer was separated according to the length of the poly(propylene oxide) block by HPLC at the critical point of adsorption. The

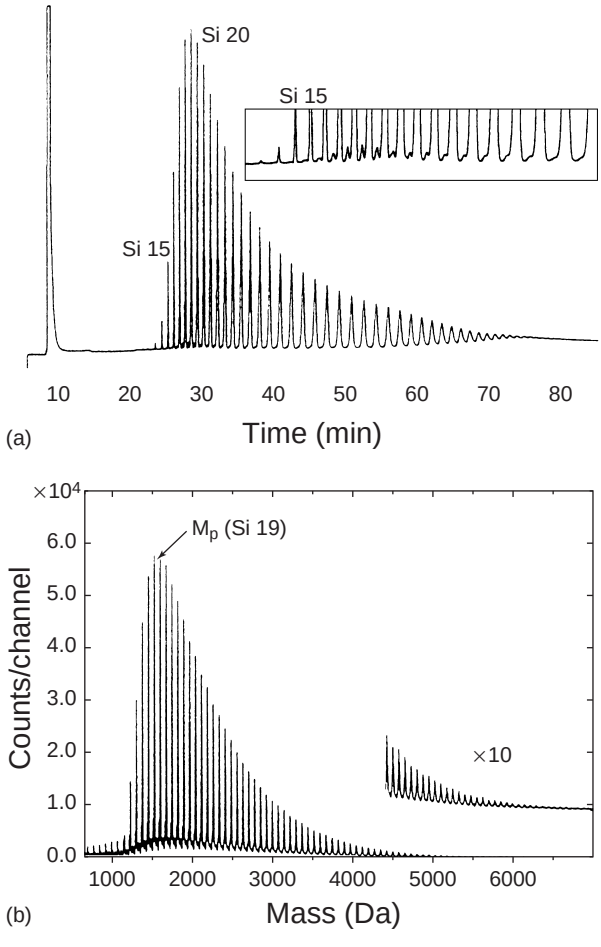

Figure 6 (a) SFC chromatogram and (b) TOFSIMS spectrum (positive secondary ions) of a poly(dimethylsiloxane) sample. (Reproduced with permission from Hagenhoff et al.[64] Copyright 1991, American Chemical Society.)

resulting fractions uniform with respect to the number of PO units ($m = 4–10$) were subjected to SFC and the average length $n_1 + n_2$ and the molecular-weight distribution of the poly(ethylene oxide) blocks were determined for every fraction. Because of the high chemical similarity, SFC did not resolve oligomers of the same gross composition but different ethylene oxide distribution, e.g. $HO(EO)_4(PO)_m(EO)_2OH$ and $HO(EO)_3(PO)_m(EO)_3OH$.[87]

Stem length and stem length distribution in the lamellar crystals of natural and synthetic rubbers were determined by the use of SFC. The stem length was detected as the number of isoprene units per molecule of ozonolysis products of the rubber crystals (the ozonolysis under specific conditions selectively decomposes noncrystalline parts in polyisoprene chains). SFC analysis clearly showed that the ozonolysis products (*cis*-1,4-oligoisoprenes) have bimodal distributions centered at 9–11 ($\overline{M}_w/\overline{M}_n < 1.11$) and 23–25 ($\overline{M}_w/\overline{M}_n < 1.03$) isoprene units respectively (Figure 7).[25]

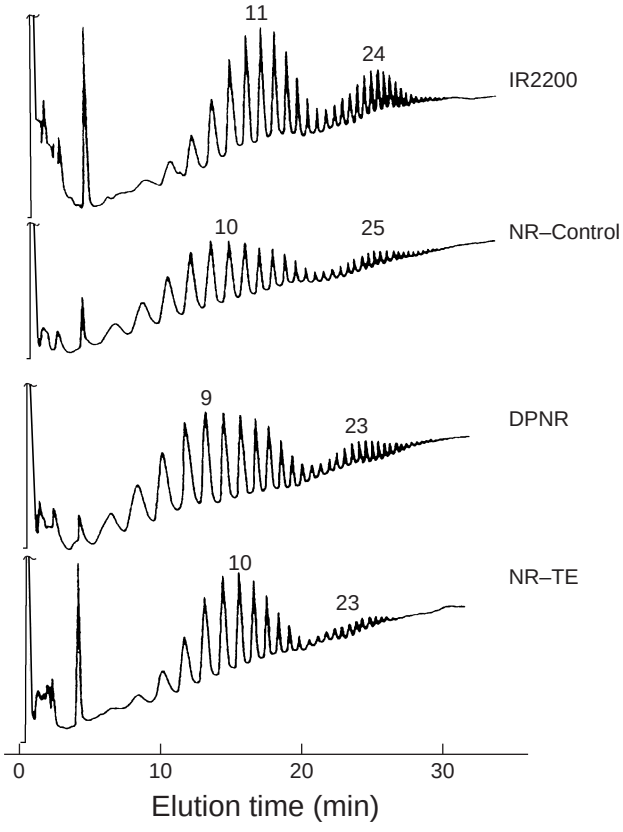

Figure 7 SFC traces of ozonolysis products from *cis*-1,4-polyisoprene crystals. IR2200, commercial synthetic *cis*-1,4-polyisoprene; NR-control, solid natural rubber; DPNR, deproteinized natural rubber; NE-TE, transesterified natural rubber. (Reproduced with permission from Kawahara et al.[25] Copyright 1998, Rubber Division, ACS.)

4.2 Separation According to Chemical Structures of Polymers

Separations by SFC yield different retentions for polymer series depending on their end-groups. This was demonstrated for oligo(1,3,6-trioxocane)s with hydroxy or benzoyloxy functionality.[88] The cationic ring-opening polymerization of 1,3,6-trioxocane in the presence of benzyl alcohol results in the formation of α, w-dihydroxy, α-hydroxy-w-benzyloxy, and α, w-dibenzoyloxy oligo(1,3,6-trioxocane)s. In their SFC separation, two to three peaks were obtained for each *DP*. Molecular weight and functionality type distributions in the oligo(1,3,6-trioxocane)s were determined simultaneously from one SFC chromatogram.[89]

Macromonomers and telechelics are known to be oligomers with exactly one or two functional end-groups. An exact knowledge of the macromonomer and telechelic functionality is necessary for the application. SFC was demonstrated to be useful for monitoring the preparation and subsequent reaction of macromonomers from styrene and from 2-vinylnaphthalene; the end-group included epoxy, hydroxy, or diol functionality. SFC separation of the condensation product from diol-functionalized styrene macromonomer (average *DP* of 13) and sebacyl dichloride was also described.[90]

Oligomeric ethylene oxide adducts which are frequently used as nonionic surfactants are produced via ethoxylation of fatty alcohols. When fatty alcohols are of technical grade, they may contain more than only one alcohol; this leads to the formation of parallel oligomer series. The SFC chromatogram of such a product is shown in Figure 8, where two main series, originating from ethoxylation of dodecanol and tridecanol, are seen along with additional series which are present in minor amounts.[13,70] Because oligo(oxyethylene) surfactants are very readily separated by means of capillary SFC/FID using carbon dioxide as the supercritical mobile phase, the separation of this class of compounds has been exemplified in a large number of publications.[13,47,57,58,66–73,91] Characterization of oligoethers has also been performed on packed column SFC.[33,54]

SFC proved to be useful for analyzing cyclic siloxanes in technical silicone oils (low molecular weight poly(dimethylsiloxane)s). Since the preparation of silicone oils with desired properties such as great range of fluidity and low freezing point requires minimization of cyclic components, the determination of cyclic siloxanes in the presence of linear poly(dimethylsiloxane)s is of special importance. Good results were obtained by SFC providing sufficient separation up to a level of molecular weight nearly 6000.[92] Direct coupling of SFC with MS permitted an unambiguous identification of linear and cyclic poly(dimethylsiloxane)s.[93] By means

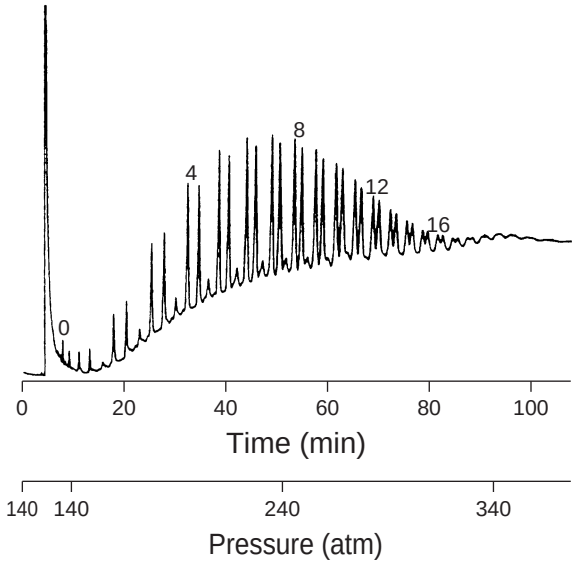

Figure 8 Pressure-programmed SFC/FID chromatogram of Neodol 23-6.5, an alkyl ethoxylate mixture of the formula $H(CH_2)_n(OCH_2CH_2)_eOH$, where n is 12 or 13 and e varies from 0 to at least 22. Some e values are indicated on the chromatogram. Conditions: carbon dioxide mobile phase; column temperature, 110 °C; BP-10 column, 9 m × 100 μm ID, 0.1 μm thick. (Reprinted from T.L. Chester 'Capillary Supercritical-fluid Chromatography with Flame-ionization Detection: Reduction of Detection Artifacts and Extension of Detectable Molecular Weight Range', *J. Chromatogr.*, **299**, 424–431, Copyright 1984, with permission from Elsevier Science.)

of high-temperature GC substances of high molecular weight do not elute, or they elute with great peak broadening. SEC does not provide the resolution required. SFC thus remains the method of choice for solving the above mentioned analytical problem.

Prepolymers of random novolac and resol resins were separated by SFC according to the number of nuclei (phenol groups) and the number of methylol groups attached to the nuclei.[41] Nine oligomers for novolac resins from dihydroxydiphenylmethanes (dimer, dinuclear) to decanuclear oligomers were separated. Seven isomers of trinuclear novolac oligomers were identified. Molecular weight averages were determined without any calibration standards ($\overline{M}_w = 417$ and $\overline{M}_n = 366$ for the sample examined).

Practical advantages of packed-column SFC over HPLC and GC were shown for the analysis of epoxy resins.[40] Analysis times of bisphenol A diglycidyl ether resins and isocyanuric acid triglycidyl ether resins are reduced by a factor of 4 and 5, respectively, with SFC compared with HPLC. For example, the analysis of isocyanuric acid triglycidyl ether resin by HPLC took 40 min, including 10 min reconditioning time, while a packed column SFC chromatogram with a similar resolution was obtained in 8 min, including 2 min reconditioning time. This allowed the use of SFC as an on-line analyzer in a pilot plant to monitor the end of the epoxidation reaction in isocyanuric acid triglycidyl ether.

SFC and SFE were reviewed with references as techniques for analyzing surface coatings and many of the raw materials that go into their manufacture.[94] Reactive oligomeric mixtures such as radiation curing coatings, polyisocyanate curing agents, and surface active agents were separated using capillary SFC with carbon dioxide. Polar oligomeric mixtures such as epoxy, vinyl, and phenol–formaldehyde resins were separated with modified mobile phases on packed column SFC.

4.3 Fractionation of Individual Polymer Homologs

One of the important features of packed column SFC is its use for the separation on a preparative scale, i.e. on a scale which allows us to utilize the fractions further or characterize the fractions by physical-chemical methods. Though the scale-up of SFC in a laboratory may be limited to some extent (≤10–20 mm ID columns) for safety reasons, several tens to hundreds of milligrams of purified fractions can easily be obtained by repeated SFC fractionations. When carbon dioxide is the chosen eluent, preparative SFC has practical advantages of the ease of fraction collection and the simple nature of the solute in the collected state. For example, a simple depressurization of supercritical carbon dioxide drastically reduces its solvating power and the purified sample precipitates.

Fractionation of polymers into the individual homologs by packed column SFC was first demonstrated for the separation of polystyrene with $\overline{M}_n$ of 578. Lower molecular weight fractions were taken and rechromatographed under the same SFC conditions to prove their purity.[2] An instrument for fraction collection was devised.[95] The research for polystyrene fractionation was extended to the isolation of 15 individual homologs from the 1mer to 15mer by the use of pentane–methanol (9 : 1) as the supercritical mobile phase and a 5 mm ID × 6 m column packed with porous silica gel (particle size 37–75 μm); more than 11 h was required for each chromatographic run. The combined amounts of the fractions after 16 consecutive runs ranged from 6 to 20 mg. MS analysis of each fraction was reported.[5,6]

Early efforts for polymer fractionation include the separation of an oligosiloxane (DC-710) which is used as a stationary phase in GC. The separation was optimized on an analytical scale[9] and then carried out on a preparative scale. The separated compounds were investigated with respect to their properties as a stationary phase for GC.[10]

Recent developments in SFC apparatus and packing materials have made it possible to carry out the fractionation in a much shorter time with improved

separation.[23] The applicability of this method to the fractionation of polymer homologs has been enhanced further by the great progress in polymerization chemistry which provides a variety of polymers with controlled molecular weight and well-defined chain structure. This is because the preparation of polymers suitable for SFC separation is essential for effective fractionation of each individual homolog.

Highly isotactic (*it*-) and highly syndiotactic (*st*-) PMMAs prepared by stereospecific living polymerizations were fractionated by the SFC using carbon dioxide containing ethanol or methanol as the mobile phase and a 10 mm ID × 250 mm column packed with nonbonded silica gel (particle size 5 μm, pore size 100 Å). Negative temperature gradients, which increase the density of mobile phase, were applied at a constant fluid-pressure of 200 kg cm^{-2}. Depending on the $\overline{M}_n$ of polymer, 20 to 50 mg was able to be separated in a chromatographic run.[36,55,56] Though the separation was achieved also by the use of ODS-treated silica gel as the stationary phase,[52] nonbonded silica gel was superior in a loadable amount of sample.

Figure 9(a) shows an SFC trace of *it*-PMMA with $\overline{M}_n$ of 4100.[36] Small amounts of authentic samples of the *it*-22mer and *it*-28mer were added to the PMMA as the internal standard. The purified *it*-50mer gave a single peak in a chromatogram recorded under the same conditions (Figure 9b), indicating the uniformity with respect to the number of repeating units in a molecule. Optimization of the operating conditions for SFC made it possible to separate *it*-PMMA with an $\overline{M}_n$ as large as 6650 (Figure 10a). Each individual homolog from 80mer ($C_{404}H_{650}O_{160}$ = 8068) to 100mer ($C_{504}H_{810}O_{200}$ = 10 070) was isolated. The molecular weight of *it*-100mer was confirmed by MALDI/TOFMS.[23,56]

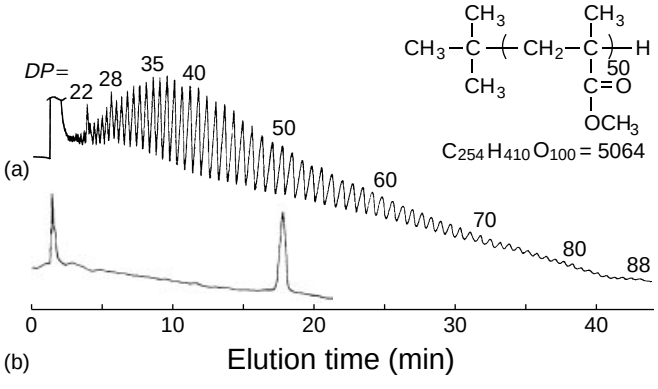

Figure 9 SFC chromatograms of (a) isotactic PMMA (DP = 40.8, $\overline{M}_w/\overline{M}_n$ = 1.12, containing the authentic samples of the 22 and 28mer) and (b) the 50mer isolated from (a).[36] (Reproduced by permission of The Society of Polymer Science, Japan.)

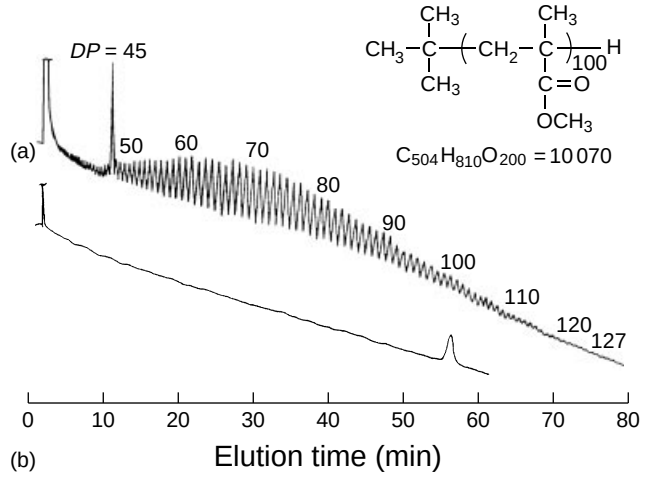

Figure 10 SFC chromatograms of (a) isotactic PMMA ($\overline{M}_n$ = 6.7 × 10^3, $\overline{M}_w/\overline{M}_n$ = 1.07) and (b) the 100mer isolated from (a). Authentic 45-mer was added to the PMMA as an internal standard. Column, Develosil 100-5 (10 mm ID × 250 mm); mobile phase, carbon dioxide–methanol (81 : 19); flow rate, 14.8 mL min^{-1}; initial temperature, 80 °C; temperature, −0.5 °C min^{-1}; pressure, 235 bar. (Reprinted from K. Ute,[23] 349–368, by courtesy of Marcel Dekker Inc.)

The series of purified homologs of *it*- and *st*-PMMAs were useful for the studies of molecular weight dependence of glass transition temperature (T_g) and melting temperature (T_m).[96] Because each homolog is uniform (perfectly monodisperse) with respect to molecular weight, it showed higher T_g and higher degree of crystallinity than nonuniform *it*- or *st*-PMMA with the corresponding (average) molecular weight. Uniform PMMAs with definite molecular weights also served as reference materials for polymer analysis and characterization, e.g. molecular weight standards for SEC[36,56] and MALDI/TOFMS.[97]

SFC is sensitive to the stereostructure of polymer homologs. When a mixture of *it*- and *st*-24mers was subjected to the SFC, *it*-24mer eluted faster than *st*-24mer.[55] The elution order was reversed in the SFC separation on ODS-treated silica gel.[52] The SFC separation according to stereostructure was applied to the fractionation of mixtures of *it*-, *st*-, and stereoblock PMMAs with uniform molecular weights (Figure 11).[98,99]

Low molecular weight poly(D-lactic acid) prepared by the living ring-opening polymerization of D-lactide with Al[OCH(CH$_3$)$_2$] in toluene was fractionated by packed column SFC (Figure 12a).[23] Slight degradation of polymer chain occurred in the SFC separations at above 65 °C. In spite of the fact that the poly(D-lactic acid) was prepared from the cyclic dimer (D-lactide), the polymer consisted of the homologs with both even and odd numbers of repeating units. The formation of odd homologs clearly indicates that

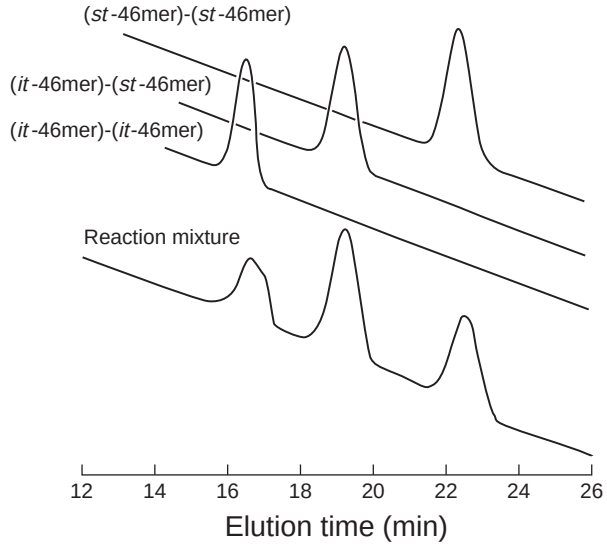

Figure 11 SFC chromatograms of a reaction mixture of hydroxyl-terminated isotactic and syndiotactic 46mers of methyl methacrylate and sebacoyl dichloride and of block PMMAs isolated from the mixture.[98] Methanol, 3.0 mL min^{-1}; carbon dioxide, 7.0 mL min^{-1}; initial temperature, 85 °C; rate of cooling, 2 °C min^{-1}. (Reproduced by permission of The Society of Polymer Science, Japan.)

the polymerization accompanies intermolecular and/or intramolecular transesterification of polymer chains. SFC analysis of poly(D-lactic acid) prepared with *N*,*N*-bis(salicylidene)ethylenediiminoaluminum methoxide [(salen)AlOCH$_3$], however, revealed that the polymer contained only a small amount of odd homologs (Figure 12b). The uniform poly(D-lactic acid)s obtained by SFC fractionation were highly crystalline. The 9mer ($C_{30}H_{44}O_{19} = 708.7$) gave thin-needle crystals from a solution in ethanol. Specific rotation, $[\alpha]_D$, of the uniform poly(D-lactic acid)s in chloroform increased with increasing number of repeating units toward an asymptotic limit which was practically reached at $DP \approx 20$. The $[\alpha]_D$ of 20mer was almost comparable to the value for high molecular weight poly(D-lactic acid) (+151°).

SFC fractionations of poly(ε-caprolactone) and poly(δ-valerolactone) were carried out and the 26mer, 28mer, and 30mer were isolated in pure form. These uniform polymers were analyzed by electrospray and MALDI/TOFMS. Molecular-weight dependences of their T_g and T_m were also reported.[38]

Polyisocyanates have received a great deal of attention due to their chiroptical properties arising from stiff helical conformation of the polymer chain. The living polymerization of *n*-butyl isocyanate was initiated with $TiCl_3OCH_2CF_3$ and terminated with acetic anhydride. Poly(*n*-butyl isocyanate) with the chain structure $CF_3CH_2O(CONC_4H_9)_nCOCH_3$ was obtained. The acetyl end-capping is necessary because low molecular weight

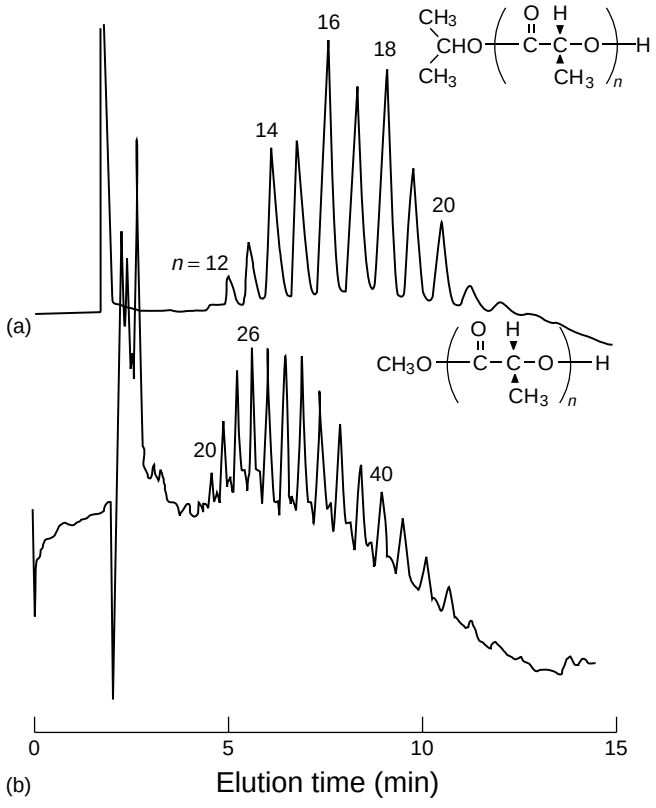

Figure 12 SFC chromatograms of poly(D-lactic acid)s prepared by polymerization with (a) Al[OCH(CH$_3$)]$_3$ and (b) (salen)AlOCH$_3$. Column, Develosil 100-5 (10 mm ID × 250 mm); mobile phase, carbon dioxide–ethanol (a) 97 : 3 (initial) → 90 : 10 (15 min), (b) 90 : 10 (initial) → 85 : 15 (20 min); flow rate, 10 mL min^{-1}; temperature, (a) 65 °C, (b) 55 °C; pressure, 198 bar. (Reprinted from K. Ute,[23] 349–368, by courtesy of Marcel Dekker Inc.)

polyisocyanates with the −NH terminus in the polymer chain readily decompose to cyclic trimers (isocyanurates) under the packed column SFC conditions. The acetyl-endcapped poly(*n*-butyl isocyanate) ($\overline{M}_n = 3970$, $\overline{M}_w/\overline{M}_n = 1.20$) was separated into the uniform polymers from 15mer to 55mer by means of SFC using carbon dioxide/ethanol as the mobile phase at 90 °C and 200 bar (Figure 13).[39] Single crystals were grown from an ethanol solution of the 12mer ($C_{64}H_{113}O_{14}N_{12}F_3 = 1131.7$) and subjected to X-ray crystallographic determination. The backbone of the polyisocyanate chain adopted an 8/3 helix with a repeat distance of 15.6 Å. This chain structure agreed well with the results of X-ray diffraction studies on powder samples and oriented films of high molecular weight poly(*n*-butyl isocyanate). The molecular weight dependence of chiroptical properties of uniform poly(*n*-butyl isocyanate)s in solution were investigated.[23]

In the middle of the 1920s, Staudinger et al. (see Ute et al.[35]) fractionated oligo(oxymethylene) diacetate,

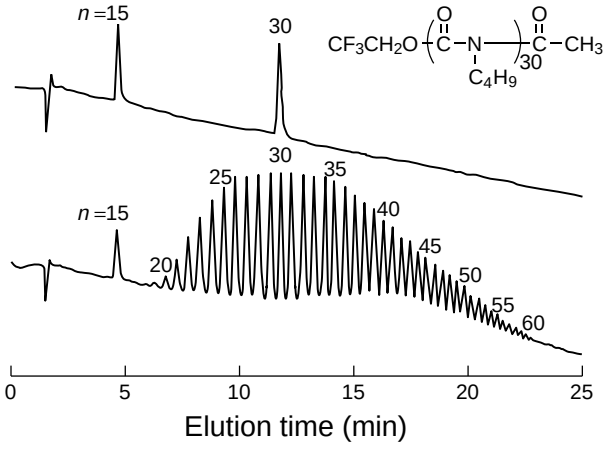

Figure 13 SFC chromatogram of the CF_3CH_2O-initiated, acetyl-terminated oligo(*n*-butyl isocyanate).[39] An authentic sample of the 15mer was added to the oligomer as an internal standard. Chromatographic conditions: Develosil 100-5 (10 mm ID × 250 mm); carbon dioxide–ethanol, 85 : 15 (initial), 70 : 30 (25 min); flow rate, 10 mL min^{-1}; temperature, 90 °C; pressure, 20 MPa. (Reproduced by permission of The Society of Polymer Science, Japan.)

$CH_3CO_2(CH_2O)_nCOCH_3$, into the individual homologs from 1mer to 26mer by rectification and fractional crystallization. These studies provided the basis for the formulation of many concepts essential in macromolecular science. Recently, purely uniform oligo(oxymethylene) diacetates that can be isolated as single crystals have been obtained by SFC fractionation.[35] During the separation, thermal degradation of the substrate occurred slightly (~3%). Each uniform oligo(oxymethylene) homolog showed a sharp melting endotherm. The melting points were significantly higher than those reported in the literature, which demonstrates improved purity of the uniform oligo(oxymethylene) diacetates obtained by the SFC fractionation. The crystal structures determined for the 9mer to 13mer were interesting as models of crystal polymorphysm in high molecular weight poly(oxymethylene).[23,35]

5 SUPERCRITICAL FLUID CHROMATOGRAPHY SEPARATIONS OF HIGH MOLECULAR WEIGHT POLYMERS

5.1 Separation of Polymers by Supercritical Fluid Chromatography in the Adsorption Mode

As described above, SFC has mostly been used by polymer chemists in adsorption chromatography mode to separate oligomers or low molecular weight polymers into their individual molecular species. For high molecular weight polymers, the peaks of the individual members of the homologous series can no longer be separated and give only one, more or less broad peak for the whole sample. The shape and broadness of this peak reflects the molecular-weight distribution of the polymer sample.

Figure 14(a) shows a chromatogram of a mixture of six polystyrene samples, each having relatively narrow molecular weight distribution and $\overline{M}_w$ ranging from 3600 to 600 000. The individual polystyrenes are well resolved from each other. This separation was achieved by the adsorption mode SFC using pentane/1,4-dioxane eluent system and nonbonded silica stationary phase.[27] A proper choice of gradient conditions is essential for this type of separation; the solvent strength must be increased to elute higher molecular weight polymers. In another chromatogram (not shown here), standard polystyrenes with close molecular weight averages ($\overline{M}_w$ = 93 000 and 111 000) were separated almost completely. A comparison of Figure 14(b) with Figure 14(a) shows a very broad molecular weight distribution of an industrial polystyrene ($\overline{M}_w$ = 230 000) which, in addition, is multimodal.

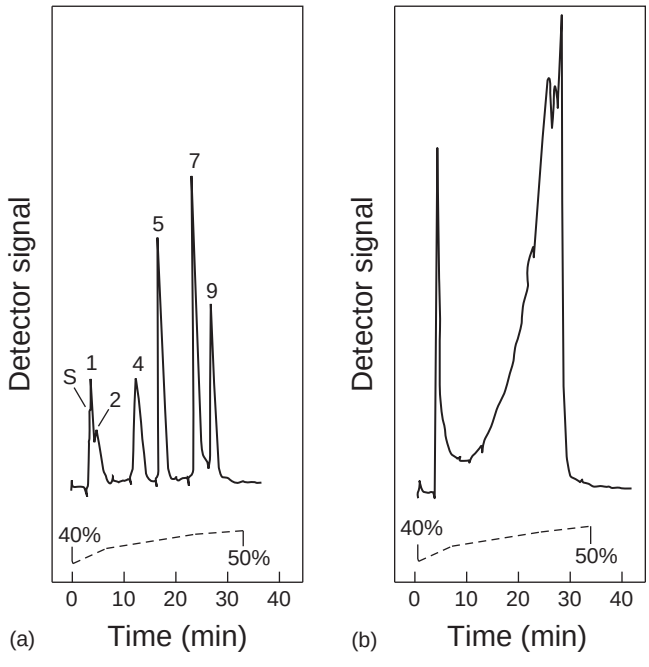

Figure 14 (a) SFC chromatogram of a polystyrene mixture; initial column end pressure: 185 bar. (b) SFC chromatogram of an industrial polystyrene sample; initial column end pressure: 182 bar.[27] Eluents, pentane–1,4-dioxane; sample solvent, 1,4-dioxane; column, 250 mm, ID 4.6 mm, packed with LiChrosorb Si 100, 10 µm; detector, UV, 254 nm; oven temperature, 250 °C; flow rate, 1 mL min^{-1}; (——) detector signal; (- - - -) content of 1,4-dioxane in the eluent mixture (% v/v); S, sample solvent peak; $\overline{M}_w$, 3600 (1), 10 000 (2), 53 700 (4), 93 000 (5), 254 000 (7), 600 000 (9). (Reprinted from F.P. Schmitz, E. Klesper, 'Polystyrene Separation by Supercritical Fluid Chromatography (SFC), *Polym. Commun.*, **24**, 142–144, Copyright 1983, with permission from Elsevier Science.)

5.2 Size-exclusion Chromatography using Supercritical Mobile Phase

Separations according to the size-exclusion mechanism are very popular in liquid chromatography (LC) of polymers. In SEC, elution order is reversed with the adsorption chromatography; the large molecules appear first at the outlet of the column, followed by the smaller molecules in the order of their decreasing size. The use of supercritical fluid in place of liquids for SEC was suggested and first demonstrated in 1977 for polystyrenes up to molecular weight 4000 by using a 40 Å porous silica stationary phase and 1,1-difluoroethane compressed above 1000 kg cm^{-2}.[11] Because the supercritical state is characterized by lower viscosity and faster diffusion than liquids both the resolution and speed of separation would be enhanced. This expectation was confirmed by the separation of standard polystyrenes with SEC using supercritical dichloromethane and pore glass packing (86 Å mean pore size).[100] Figure 15 shows the chromatographic performance derived from liquid and supercritical fluid SEC for separations of three polystyrene samples and benzene. The resolution of higher molecular weight polystyrenes is better in the supercritical case. It was also suggested that the effective dimension or size of polystyrene is smaller in supercritical fluids than in the corresponding liquids.

As an approach closely related to supercritical SEC, "enhanced-fluidity SEC" was proposed recently. An enhanced-fluidity liquid is prepared by mixing a low viscosity liquid, usually liquid carbon dioxide, with a common organic solvent. The use of enhanced-fluidity mixtures of tetrahydrofuran (THF)–liquid carbon dioxide for the SEC separation of standard polystyrenes under room temperature and moderate pressure conditions was investigated.[101,102] Mixtures of up to 40 mol% carbon dioxide with THF resulted in a significant drop in mobile-phase viscosity, without a decrease in solvent strength. As a result, higher efficiency and a shorter analysis time were obtained. Adsorption effects began to play a role at carbon dioxide concentrations above 50 mol%.

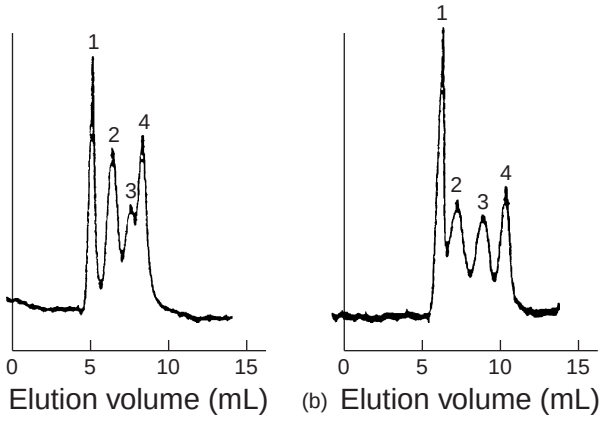

Figure 15 Comparison between (a) supercritical fluid and (b) liquid SEC.[100] 1, MW 110 000 polystyrene (0.3 wt%); 2, MW 17 000 polystyrene (0.8 wt%); 3, MW 4000 polystyrene (0.6 wt%); 4, benzene (0.4 wt%). SFC conditions: column pressure, 320 kg cm^{-2}; column temperature, 224 °C; mobile phase flow rate, 630 μL min^{-1}. LC conditions: mobile phase flow rate, 630 μL min^{-1}; column temperature, 25 °C. (Reproduced from the *Journal of Chromatographic Science*, by permission of Preston Publications, A Division of Preston Industries, Inc.)

5.3 Separation of Polymers by Supercritical Fluid Chromatography at the Critical Adsorption Point

LC at the critical adsorption point (CAP) is a chromatographic method that allows the separation based upon chemical heterogeneities of polymers.[103] Operating at "critical" conditions where size-exclusion effects are balanced by adsorption effects, retention becomes independent of the length of the polymer chain and separation is accomplished exclusively by the chemical heterogeneities of polymers, for example end-groups or branching. This principle can be applied in SFC.

Figure 16 shows the SFC/CAP chromatograms of polystyrene precursor ($\overline{M}_w = 2000$) and the functionalized polystyrenes, synthesized by reaction of the living polystyrene precursor with a trifunctional core compound 2,4,6-trichloro-1,3,5-triazine.[104] The mobile phase was a mixture of 79% dichloromethane and 21% carbon dioxide by volume. The length of arms corresponds to the size of the initial precursor. Separations of hydroxyl- as well as acetyl-terminated polystyrenes could be achieved. The main peak indicates the polystyrene precursor, whereas the following peaks are caused respectively by the hydroxyl and acetyl end-groups of the polystyrene molecules with increasing number of arms.

Finding the exact experimental conditions for the CAP is often a matter of trial and error. In SFC, the determination of the CAP is possible by varying the mobile phase (carbon dioxide mixed with a polar solvent) or the temperature, pressure, and density of the fluid, respectively. The unique benefits of using supercritical fluid for the chromatography at the CAP may be similar to those described for the enhanced-fluidity chromatography at the CAP.[105] The high compressibility of enhanced-fluidity liquid mixtures allows precise and broad ranging control of solvent strength. The optimum method to approach the CAP is to roughly approach by varying the eluent composition, and then precisely determine the critical condition by changing the eluent pressure. The application of enhanced-fluidity chromatography at the CAP was illustrated for the characterization of copolymers of styrene and methyl acrylate.

Just and Much[104] provided a comprehensive analysis of the use of supercritical mobile phases in three

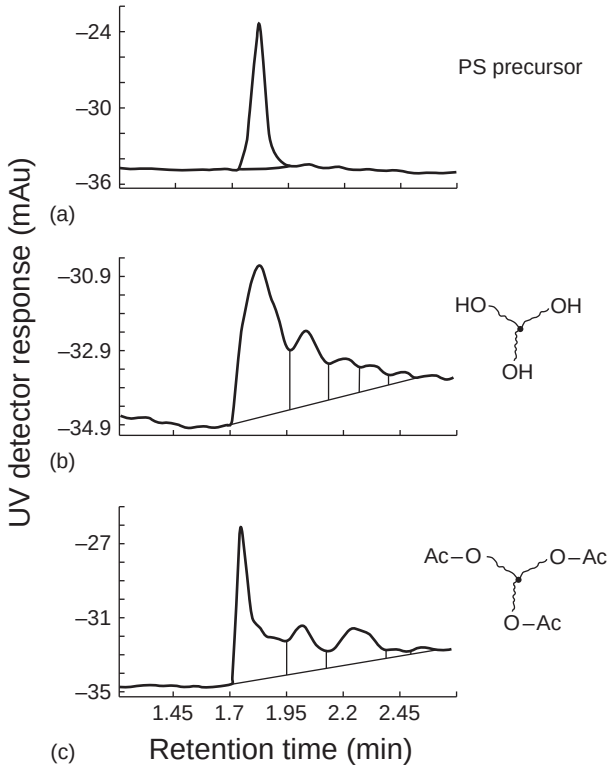

Figure 16 Separation of (a) polystyrene precursor, (b) OH-terminated three-armed polystyrene, and (c) acetyl-terminated three-armed polystyrene at critical conditions of adsorption in SFC.[104] Mobile phase: dichloromethane–carbon dioxide (79 : 21, v/v); temperature, 473 K; pressure, 200 bar. (Reproduced by permission of Gordon and Breach Science Publishers.)

separation modes: adsorption chromatography for the separation of oligomers, SEC for the determination of molecular weight distributions, and adsorption chromatography at critical conditions for the determination of functionality type. The authors provided examples of these modes with separations of polystyrenes.

6 ANALYSIS OF COMMON POLYMER ADDITIVES BY SUPERCRITICAL FLUID CHROMATOGRAPHY

Polymer additives are found in plastics, rubbers, surfactants, textiles, and cosmetics. They are usually added at very low amounts (0.01–1.0%) to polymeric materials to protect them against deterioration during processing at high temperature and/or to improve their aging characteristics. There is a practical need for analytical methods to characterize additives and to determine their concentration in the polymers. In an increasing number of publications including review articles,[40,94,106] the unique performance of SFC and SFE for determination of polymer additives has been mentioned.

A standard mixture of 21 polymer additives (UV stabilizers, antioxidants, metal deactivators, and slip agents) varying in polarity and molecular weight from 225 to 1178 was separated into the individual components by SFC on a nonpolar capillary column with carbon dioxide as the mobile phase and was analyzed by an FTIR detector coupled to the column outlet. Good quality spectra were obtained from samples at levels of the order of 100 ng. A polypropylene extract was analyzed to demonstrate the method and to show the limitation of using retention time data for identifying unknown compounds.[107]

HPLC and SFC separations were compared for characterization of more than 15 polymer additives in polyolefin extracts. Detection in HPLC was by ELSD or RI, in SFC by ELSD and FID. SFC was found to be more universally applicable to these compounds than HPLC. SFC was also superior in terms of both resolution and speed of analysis.[108]

SFE as a powerful technique to eliminate solvent or matrix effects before chromatographic analysis was combined with packed column SFC for determination of polymer additives in various types of polymers. Low-density polyethylene was extracted and analyzed with on-line SFE/SFC with good yields for some of the common polymer additives. The calculated levels of the additives were in good agreement with the assayed values from the polymer manufacturer.[109,110] A procedure to quantitatively analyze additives in various polyethylene and polypropylene samples using on-line SFE/SFC was also described.[111] Either oligomers or polymer additives were selectively extracted depending on the pressure. In other studies, organotin stabilizers and plasticizers in poly(vinyl chloride)[112,113] and polymer additives in polystyrene[110] was analyzed by on-line SFE/SFC.

ABBREVIATIONS AND ACRONYMS

CAP	Critical Adsorption Point
DP	Degree of Polymerization
ELSD	Evaporative Light Scattering Detector
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
ID	Internal Diameter
IR	Infrared
LC	Liquid Chromatography
MALDI	Matrix-assisted Laser Desorption/Ionization
MS	Mass Spectrometry

NMR	Nuclear Magnetic Resonance
ODS	Octadecylsilane
PMMA	Poly(methyl methacrylate)
RI	Refractive Index
SEC	Size-exclusion Chromatography
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extraction
T_g	Glass Transition Temperature
T_m	Melting Temperature
THF	Tetrahydrofuran
TOFMS	Time-of-flight Mass Spectrometry
TOFSIMS	Time-of-flight Secondary Ion Mass Spectrometry
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VPO	Vapor Pressure Osmometry

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Supercritical Fluid Chromatography in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Supercritical Fluid Extraction of Inorganics in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis

Polymers and Rubbers **(Volume 9)**
Size-exclusion Chromatography of Polymers

Liquid Chromatography **(Volume 13)**
Supercritical Fluid Chromatography

REFERENCES

1. E. Klesper, A.H. Corwin, D.A. Turner, 'High Pressure Gas Chromatography Above Critical Temperatures', *J. Org. Chem.*, **27**, 700–701 (1962).
2. R.E. Jentoft, T.H. Gouw, 'Supercritical Fluid Chromatography of a ''Monodisperse'' Polystyrene', *J. Polym. Sci., Polym. Lett.*, **7**, 811–813 (1969).
3. R.E. Jentoft, T.H. Gouw, 'Pressure-programmed Supercritical Fluid Chromatography of Wide Molecular Weight Range Mixtures', *J. Chromatogr. Sci.*, **8**, 138–142 (1970).
4. S.T. Sie, J.P.A. Bleumer, G.W.A. Rijnders, *Gas Chromatography 1968*, ed. C.L.A. Harbourn, Elsevier, Amsterdam, 235, 1969.
5. W. Hartmann, E. Klesper, 'Preparative Supercritical Fluid Chromatography of Styrene Oligomers', *J. Polym. Sci., Polym. Lett.*, **15**, 713–719 (1977).
6. E. Klesper, W. Hartmann, 'Apparatus and Separations in Supercritical Fluid Chromatography', *Eur. Polym. J.*, **14**, 77–88 (1978).
7. J.E. Conaway, J.A. Graham, L.B. Rogers, 'Effects of Pressure, Temperature, Adsorbent Surface, and Mobile Phase Composition on the Supercritical Fluid Chromatographic Fractionation of Monodisperse Polystyrenes', *J. Chromatogr. Sci.*, **16**, 102–110 (1978).
8. J.A. Graham, L.B. Rogers, 'Effects of Column Length, Particle Size, Flow Rate, and Pressure Programming Rate on Resolution in Pressure-programmed Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **18**, 75–84 (1980).
9. J.A. Nieman, L.B. Rogers, 'Supercritical Fluid Chromatography Applied to the Characterization of a Siloxane-based Gas Chromatographic Stationary Phase', *Sep. Sci.*, **10**, 517–545 (1975).
10. F.J. van Lenten, J.E. Conaway, L.B. Rogers, 'Gas chromatographic Retention Characteristics of Different Polysiloxane Oligomers', *Sep. Sci.*, **12**, 1–28 (1977).
11. J.C. Giddings, L.M. Bowman, Jr, M.N. Myers, 'Exclusion Chromatography in Dense Gases: an Approach to Viscosity Optimization', *Anal. Chem.*, **49**, 243–249 (1977).
12. M. Novotny, S.R. Springston, P.A. Peaden, J.C. Fjeldsted, M.L. Lee, 'Capillary Supercritical Fluid Chromatography', *Anal. Chem.*, **53**, 407A–414A (1981).
13. T.L. Chester, 'Capillary Supercritical-fluid Chromatography with Flame-ionization Detection: Reduction of Detection Artifacts and Extension of Detectable Molecular Weight Range', *J. Chromatogr.*, **299**, 424–431 (1984).
14. D.R. Gere, R. Board, D. McManigill, 'Supercritical Fluid Chromatography with Small Particle Diameter Packed Columns', *Anal. Chem.*, **54**, 736–740 (1982).
15. D.R. Gere, 'Supercritical Fluid Chromatography', *Science*, **222**, 253–259 (1983).
16. R.D. Smith, B.W. Wright, C.R. Yonker, 'Supercritical Fluid Chromatography: Current Status and Prognosis', *Anal. Chem.*, **60**, 1323A–1336A (1988).
17. F. Wach, 'The Expanding Horizons of SFE and SFC', *Anal. Chem.*, **66**, 369A–373A (1994).
18. B. Erickson, 'SFC in Flux', *Anal. Chem.*, **69**, 683A–686A (1997).
19. F.P. Schmitz, E. Klesper, 'Separation of Oligomers and Polymers by Supercritical Fluid Chromatography', *J. Supercrit. Fluids*, **3**, 29–48 (1990).
20. C.M. White, R.K. Houck, 'Supercritical Fluid Chromatography and Some of its Applications: a Review',

J. High Resolut. Chromatogr., Chromatogr. Commun., **9**, 4–17 (1986).
21. T.L. Chester, J.D. Pinkston, 'Supercritical Fluid Chromatography', *Anal. Chem.*, **62**, 394R–402R (1990).
22. T.L. Chester, J.D. Pinkston, D.E. Raynie, 'Supercritical Fluid Chromatography and Extraction', *Anal. Chem.*, **64**, 153R–170R (1992); **66**, 106R–130R (1994); **68**, 487R–514R (1996); **70**, 301R–319R (1998).
23. K. Ute, 'Fractionation of Polymer Homologs with Packed Column Supercritical Fluid Chromatography', in *Supercritical Fluid Chromatography with Packed Columns*, eds. K. Anton, C. Berger, Marcel Dekker, New York, 349–368, 1998.
24. Y. Hirata, F. Nakata, M. Horihata, 'Direct Sample Injection in Supercritical Fluid Chromatography With Packed Fused Silica Column', *J. High Resolut. Chromatogr., Chromatogr. Commun.*, **11**, 81–84 (1988).
25. S. Kawahara, Y. Inomata, T. Kakubo, Y. Tanaka, K. Hatada, K. Ute, N. Miyatake, 'Stem Length and Stem Length Distribution of Chain-folded Lamellar Crystal Grown from Dilute Solution of *cis*-1,4-polyisoprene by Ozonolysis-GPC Measurements', *Rubber Chem. Technol.*, **71**, 277–284 (1998).
26. F.P. Schmitz, H. Hilgers, E. Klesper, 'Gradient Elution in Supercritical Fluid Chromatography', *J. Chromatogr.*, **267**, 267–275 (1983).
27. F.P. Schmitz, E. Klesper, 'Polystyrene Separation by Supercritical Fluid Chromatography (SFC)', *Polym. Commun.*, **24**, 142–144 (1983).
28. F.P. Schmitz, H. Hilgers, B. Lorenschat, E. Klesper, 'Separation of Oligomers with UV-adsorbing Side Groups by Supercritical Fluid Chromatography Using Eluent Gradients', *J. Chromatogr.*, **346**, 69–79 (1985).
29. B. Gemmel, B. Lorenschat, F.P. Schmitz, 'Separation of Oligomers of Medium Polarity by Packed Column Supercritical Fluid Chromatography', *Chromatographia*, **27**, 605–610 (1989).
30. F.P. Schmitz, H. Hilgers, B. Gemmel, 'High-performance Liquid and Supercritical Fluid Chromatographic Separation of Vinyl Oligomers by Gradient Elution', *J. Chromatogr.*, **371**, 135–144 (1986).
31. M. Takeuchi, T. Saito, 'Applications of Semi-micro Supercritical Fluid Chromatography with Gradient Elution to Synthetic Oligomer Separation', *J. Chromatogr.*, **515**, 629–638 (1990).
32. S. Brossard, M. Lafosse, M. Dreux, 'Comparison of Ethoxylated Alcohols and Polyethylene Glycols by High-performance Liquid Chromatography and Supercritical Fluid Chromatography Using Evaporative Light-scattering Detection', *J. Chromatogr.*, **591**, 149–157 (1992).
33. A. Giorgetti, N. Pericles, H.M. Widmer, K. Anton, 'Mixed Mobile Phases and Pressure Programming in Packed and Capillary Column Supercritical Fluid Chromatography: a Unified Approach', *J. Chromatogr. Sci.*, **27**, 318–324 (1989).
34. T.A. Dean, C.F. Poole, 'Some Practical Aspects of Column Design for Packed-column Supercritical Fluid Chromatography', *J. Chromatogr.*, **468**, 127–144 (1989).
35. K. Ute, T. Takahashi, K. Matsui, K. Hatada, 'X-ray Single Crystal Analysis of Oligo(oxymethylene) Diacetates', *Polym. J.*, **25**, 1275–1281 (1993).
36. K. Ute, N. Miyatake, Y. Osugi, K. Hatada, 'Isotactic and Syndiotactic Pentacontamers (50mers) of Methyl Methacrylate Obtained by Fractionation with Preparative SFC: the GPC Analysis of the 50mers and Their Stereocomplex', *Polym. J.*, **25**, 1153–1159 (1993).
37. K. Ute, T. Asada, K. Hatada, 'Highly ''Diheterotactic'' Polymerization of *tert*-butyl Crotonate', *Macromolecules*, **29**, 1904–1909 (1996).
38. E. Ihara, M. Tanabe, Y. Nakayama, A. Nakamura, H. Yasuda, 'Chracterization of Lactone Oligomers Isolated by Preparative SFC', *Macromol. Chem. Phys.*, **200**, 758–762 (1999).
39. K. Ute, T. Asai, Y. Fukunishi, K. Hatada, 'Stabilization of Oligo(butyl isocyanate) by Acetyl End-capping', *Polym. J.*, **27**, 445–448 (1995).
40. K. Anton, M. Bach, C. Berger, F. Walch, G. Jaccard, Y. Carlier, 'From Potential to Practice: Relevant Industrial Applications of Packed-column Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **32**, 430–438 (1994).
41. S. Mori, T. Saito, M. Takeuchi, 'Separation of Prepolymers of Phenol-formaldehyde Resins by Supercritical-fluid Chromatography', *J. Chromatogr.*, **478**, 181–190 (1989).
42. S.B. Howthorne, D.J. Miller, 'Analysis of Commercial Waxes Using Capillary Supercritical Fluid Chromatography-Mass Spectrometry', *J. Chromatogr.*, **388**, 397–409 (1987).
43. F.P. Schmitz, B. Gemmel, D. Leyendecker, D. Leyendecker, 'Separation of 2-vinylnaphthalene Oligomers by Open-tubular Capillary Supercritical Fluid Chromatography', *J. High Resol. Chromatogr. Chromatogr. Commun.*, **11**, 339–340 (1988).
44. B.W. Wright, H.T. Kalinoski, H.R. Udseth, R.D. Smith, 'Capillary Supercritical Fluid Chromatography–Mass Spectrometry', *J. High Resolut. Chromatogr., Chromatogr. Commun.*, **9**, 145–153 (1986).
45. D.W. Later, E.R. Campbell, B.E. Richter, 'Synchronized Temperature/Density Programming in Capillary Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**, 65–68 (1988).
46. J.D. Pinkston, T.E. Delaney, D.J. Bowling, 'Characterization of Low Molecular Weight Poly(acrylic acid) Samples by Capillary Supercritical Fluid Chromatography and Capillary Supercritical Fluid Chromatography–Mass Spectrometry', *J. Microcolumn. Sep.*, **2**, 181–187 (1990).
47. D. Leyendecker, 'Selection of Conditions for an SFC Separation', in *Supercritical Fluid Chromatography*,

ed. R.M. Smith, Royal Society of Chemistry, London, 53–101, 1988.
48. Y. Hirata, 'Supercritical Fluid Chromatography of Various Samples with Different Polarities', *J. Chromatogr.*, **315**, 39–44 (1984).
49. Y. Hirata, F. Nakata, 'Supercritical Fluid Chromatography With Fused-silica Packed Columns', *J. Chromatogr.*, **295**, 315–322 (1984).
50. T. Takeuchi, T. Niwa, D. Ishii, 'Porous Glass Beads as Stationary Phase in Supercritical Fluid Chromatography', *Chromatographia*, **25**, 332–334 (1988).
51. T. Takeuchi, D. Ishii, M. Saito, K. Hibi, 'Supercritical Fluid Chromatography with Micro Packed Columns and Carbon Dioxide as a Mobile Phase', *J. Chromatogr.*, **295**, 323–331 (1984).
52. K. Hatada, K. Ute, T. Nishimura, M. Kashiyama, T. Saito, M. Takeuchi, 'Stereoregular Oligomers of Methyl Methacrylate: 4. Supercritical Fluid Chromatography of the Isotactic and Syndiotactic Oligomers of MMA', *Polym. Bull.*, **23**, 157–162 (1990).
53. F.P. Schmitz, B. Gemmel, 'Separation of Different Oligomers by Supercritical Fluid Chromatography Using Gradient Elution', *Fresenius Z. Anal. Chem.*, **330**, 216–217 (1988).
54. H.M. Hagen, K.E. Landmark, T. Greibrokk, 'Separation of Oligomers of Polyethylene Glycols by Supercritical Fluid Chromatography', *J. Microcolumn Sep.*, **3**, 27–31 (1991).
55. K. Ute, N. Miyatake, T. Asada, K. Hatada, 'Stereoregular Oligomers of Methyl Methacrylate: 6. Isolation of Isotactic and Syndiotactic MMA Oligomers from 19-mer to 29-mer by Preparative Supercritical Fluid Chromatography', *Polym. Bull.*, **28**, 561–568 (1992).
56. K. Hatada, K. Ute, T. Kitayama, T. Nishiura, N. Miyatake, 'Preparation and Properties of Isotactic and Syndiotactic Poly(methyl methacrylate)s, Uniform with Respect to Molecular Weight', *Macromol. Symp.*, **85**, 325–338 (1994).
57. S.B. Hawthorne, D.J. Miller, 'An Almost On-column Injector for Capillary Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **27**, 197–202 (1989).
58. J.L. Hensley, H.M. McNair, 'Modification of a Gas Chromatograph for Capillary Supercritical Fluid Chromatography', *J. Liquid Chromatogr.*, **9**, 1985–1996 (1986).
59. S.L. Pentoney, Jr, K.H. Shafer, P.R. Griffiths, R. Fuoco, 'Supercritical Fluid Chromatography/Fourier Transform Infrared Microspectrometry', *J. High Resolut. Chromatogr., Chromatogr. Commun.*, **9**, 168–171 (1986).
60. S.L. Pentoney, A. Giorgetti, P.R. Griffiths, 'Combined Gas and Supercritical Fluid Chromatography for the High Resolution Separation of Volatile and Nonvolatile Compounds', *J. Chromatogr. Sci.*, **25**, 93–98 (1987).
61. G.L. Pariente, S.L. Pentoney, Jr, P.R. Griffiths, K.H. Shafer, 'Computer-controlled Pneumatic Amplifier Pump for Supercritical Fluid Chromatography and Extractions', *Anal. Chem.*, **59**, 808–813 (1987).
62. D.J. Bornhop, S. Schmidt, N.L. Porter, 'Use of Two Simultaneous Detectors in Capillary Supercritical Fluid Chromatography', *J. Chromatogr.*, **459**, 193–200 (1988).
63. R.C. Wieboldt, G.E. Adams, D.W. Later, 'Sensitivity Improvement in Infrared Detection for Supercritical Fluid Chromatography', *Anal. Chem.*, **60**, 2422–2427 (1988).
64. B. Hagenhoff, A. Benninghove, H. Barthel, W. Zoller, 'Supercritical Fluid Chromatography and Time-of-flight Secondary Ion Mass Spectrometry of Poly(dimethylsiloxane) Oligomers in the Mass Range 1000–10 000 Da', *Anal. Chem.*, **63**, 2466–2469 (1991).
65. D.E. Knowles, L. Nixon, E.R. Campbell, D.W. Later, B.E. Richter, 'Industrial Applications of Supercritical Fluid Chromatography: Polymer Analysis', *Fresenius Z. Anal. Chem.*, **330**, 225–228 (1988).
66. B.E. Richter, 'Modified Flame Ionization Detector for the Analysis of Large Molecular Weight Polar Compounds by Capillary Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **8**, 297–300 (1985).
67. R.L. Eatherton, M.A. Morissey, W.F. Siems, H.H. Hill, Jr, 'Ion Mobility Detection after Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr.*, **9**, 154–160 (1986).
68. M.L. Lee, K.E. Markides, 'Current Technological Challenges in Capillary Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **9**, 652–656 (1986).
69. M.W. Raynor, K.D. Bartle, I.L. Davies, A.A. Clifford, A. Williams, 'Preparation of Robust Tapered Restrictors for Capillary Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**, 289–291 (1988).
70. J.D. Pinkston, D.J. Bowling, T.E. Delaney, 'Industrial Applications of Supercritical-fluid Chromatography-mass Spectrometry Involving Oligomeric Materials of Low Volatility and Thermally Labile Materials', *J. Chromatogr.*, **474**, 97–111 (1989).
71. J.A. Crow, J.P. Foley, 'Optimization of Separations Using Short Capillary Columns in Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr.*, **12**, 467–470 (1989).
72. K. Anton, N. Pericles, H.M. Widmer, 'Capillary Column SFC Instrumentation Based on Piston Pumps with a Programmable Back-pressure Regulator for On-line Mixing of Mobile Phases', *J. High Resolut. Chromatogr.*, **12**, 394–399 (1989).
73. U. Just, H.-R. Holzbauer, M. Resch, 'Molar Mass Determination of Oligomeric Ethylene Oxide Adducts Using Supercritical Fluid Chromatography and Matrix-assisted Laser Desorption-ionization Time-of-flight Mass Spectrometry', *J. Chromatogr. A*, **667**, 354–360 (1994).
74. M.W. Raynor, A.A. Clifford, K.D. Bartle, C. Reyner, A. Williams, B.W. Cook, 'Microcolumn Supercritical Fluid Chromatography with On-line Fourier Transform

Infrared Detection', *J. Microcolumn Sep.*, **1**, 101–109 (1989).
75. C. Fujimoto, Y. Hirata, K. Jinno, 'Supercritical Fluid Chromatography–Infrared Spectroscopy of Oligomers: Use of Buffer-memory Technique', *J. Chromatogr.*, **332**, 47–56 (1985).
76. D.F. Gurka, S. Pyle, R. Titus, E. Shafter, 'Direct-deposition Infrared Spectrometry With Gas and Supercritical Fluid Chromatography', *Anal. Chem.*, **66**, 2521–2528 (1994).
77. K. Matsumoto, S. Tsuge, Y. Hirata, 'Development of Directly Coupled Supercritical Fluid Chromatography–Mass Spectrometry with Self-spouting and Vacuum Nebulizing Assisted Interface', *Anal. Sci.*, **2**, 3–7 (1986).
78. R.D. Smith, H.R. Udseth, 'Mass Spectrometer Interface for Microbore and High Flow Rate Capillary Supercritical Fluid Chromatography With Splitless Injection', *Anal. Chem.*, **59**, 13–22 (1987).
79. H.T. Kalinoski, H.R. Udseth, E.K. Chess, R.D. Smith, 'Capillary Supercritical Fluid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **394**, 3–14 (1987).
80. H.T. Kalinoski, H.R. Udseth, B.W. Wright, R.D. Smith, 'Analytical Applications of Capillary Supercritical Fluid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **400**, 307–316 (1987).
81. W. Blum, K. Grolimund, P.E. Jordi, P. Ramstein, 'Simultaneous Coupling of Capillary Supercritical Fluid Chromatography and High Temperature Glass Capillary Gas Chromatography to a Mass Spectrometer', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**, 441–448 (1988).
82. A.T. Kalinoski, L.O. Hargiss, 'Design and Industrial Applications of a Removable Probe Interface for Direct Capillary Supercritical-fluid Chromatography–Mass Spectrometry', *J. Chromatogr.*, **474**, 69–82 (1989).
83. J.D. Pinkston, G.D. Owens, L.J. Burkes, T.E. Delaney, D.S. Millington, D.A. Maltby, 'Capillary Supercritical Fluid Chromatography–Mass Spectrometry Using a ''High Mass'' Quadrupole and Splitless Injection', *Anal. Chem.*, **60**, 962–966 (1988).
84. D.E. Raynie, K.E. Markides, M.L. Lee, S.R. Goates, 'Black-pressure Regulated Restrictor for Flow Control in Capillary Supercritical Fluid Chromatography', *Anal. Chem.*, **61**, 1178–1181 (1989).
85. T. Altares, Jr, 'Supercritical Fluid Chromatography: Molecular Weight Calculations from the Chromatogram', *J. Polym. Sci., Polym. Lett.*, **8**, 761–766 (1970).
86. B. Trathnigg, B. Maier, G. Schulz, R. Krüger, U. Just, 'Characterization of Polyethers Using Different Methods: SFC, LAC and SEC with Different Detectors, and MALDI–TOF–MS', *Macromol. Symp.*, **110**, 231–240 (1996).
87. H. Pasch, C. Brinkmann, H. Much, U. Just, 'Chromatographic Investigations of Macromolecules in the ''Critical Range'' of Liquid Chromatography: II. Two-dimensional Separations of Poly(ethylene oxide-block-propylene oxide)', *J. Chromatogr.*, **623**, 315–322 (1992).
88. H. Pasch, H. Krüger, H. Much, U. Just, 'Analysis of Benzyloxy-terminated Poly(1,3,6-trioxocane)s by Supercritical Fluid Chromatography', *J. Chromatogr.*, **589**, 295–299 (1992).
89. H. Pasch, H. Krüger, H. Much, U. Just, 'Simultaneous Determination of Functionality Type and Molar Mass Distribution of Oligo(1,3,6-trioxocane)s by Supercritical Fluid Chromatography', *Polymer*, **33**, 3889–3892 (1992).
90. F.P. Schmitz, B. Gemmel, 'Supercritical Fluid Chromatography as a Means for monitoring the Preparation of Macromers', *Fresenius J. Anal. Chem.*, **344**, 467–473 (1992).
91. T. Takeuchi, K. Ohta, D. Ishii, 'Temperature Programming Elution in Capillary Supercritical Fluid Chromatography', *Chromatographia*, **25**, 125–128 (1988).
92. U. Just, F. Mellor, F. Keidel, 'Supercritical Fluid Chromatography (SFC) of Linear and Cyclic Components in Technical Silicone Oils', *Fresenius J. Anal. Chem.*, **348**, 745–748 (1994).
93. U. Just, F. Mellor, F. Keidel, 'Supercritical Fluid Chromatography–Mass Spectrometry and Matrix-assisted Laser–Desorption Ionisation Mass Spectrometry of Cyclic Siloxanes in Technical Silicone Oils and Silicone Rubbers', *J. Chromatogr. A*, **683**, 105–113 (1994).
94. M.W. Raynor, K.D. Bartle, 'Supercritical Fluid Chromatography and Extraction in Surface Coatings Analysis–Review of Representative Applications', *J. Supercritical Fluids*, **6**, 39–49 (1993).
95. R.E. Jentoft, T.H. Gouw, 'Apparatus for Supercritical Fluid Chromatography with Carbon Dioxide as the Mobile Phase', *Anal. Chem.*, **44**, 681–686 (1972).
96. K. Ute, N. Miyatake, K. Hatada, 'Glass Transition Temperature and Melting Temperature of Uniform Isotactic and Syndiotactic Poly(methyl methacrylate) from 13mer to 50mer', *Polymer*, **36**, 1415–1419 (1995).
97. N. Sakurada, T. Fukuo, R. Arakawa, K. Ute, K. Hatada, 'Characterization of Poly(methyl methacrylate) by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry. A Comparison with Supercritical Fluid Chromatography and Gel Permeation Chromatography', *Rapid Commun. Mass Spectrom.*, **12**, 1895–1898 (1998).
98. T. Nishiura, T. Kitayama, K. Hatada, 'Intra- and Intermolecular Stereocomplex Formation of Uniform Stereoblock Poly(methyl methacrylate)', *Polym. J.*, **28**, 1021–1023 (1996).
99. K. Hatada, T. Nishiura, T. Kitayama, K. Ute, S. Hirotani, 'Preparation of Uniform Stereoblock Poly(methyl Methacrylate)', *Polym. J.*, **28**, 185–188 (1996).
100. C. Fujimoto, T. Watanabe, K. Jinno, 'Size Exclusion Chromatography of Polystyrenes with Supercritical Dichloromethane', *J. Chromatogr. Sci.*, **27**, 325–328 (1989).

101. H. Yuan, S.V. Olesik, 'Comparison of Enhanced-fluidity and Elevated Temperature Mobile Phases for High-performance Size-exclusion Chromatography', *J. Chromatogr.*, **785**, 35–48 (1997).
102. H. Yuan, I. Souvignet, S.V. Olesik, 'High-performance Size Exclusion Chromatography Using Enhanced-fluidity Liquid Mobile Phases', *J. Chromatogr. Sci.*, **35**, 409–416 (1997).
103. S.G. Entelis, V.V. Evreinov, A.V. Gorshkov, 'Functionality and Molecular Weight Distribution of Telechelic Polymers', *Adv. Polym. Sci.*, **76**, 129–175 (1987).
104. U. Just, H. Much, 'Characterization of Polymers Using Supercritical Fluid Chromatography: Application of Adsorption Chromatography, Size Exclusion Chromatography and Adsorption Chromatography at Critical Conditions', *Int. J. Polym. Anal. Charact.*, **2**, 173–184 (1996).
105. I. Souvignet, S.V. Olesik, 'Liquid Chromatography at the Critical Condition Using Enhanced-fluidity Liquid Mobile Phases', *Anal. Chem.*, **69**, 66–71 (1997).
106. C. Berger, 'Packed Column Supercritical Fluid Chromatography in the Development of Polymer Additives', in *Supercritical Fluid Chromatography with Packed Columns*, eds. K. Anton, C. Berger, Marcel Dekker, New York, 301–348, 1998.
107. M.W. Raynor, K.D. Bartle, I.L. Davies, A. Williams, A.A. Clifford, J.M. Chalmers, B.W. Cook, 'Polymer Additive Characterization by Capillary Supercritical Fluid Chromatography/Fourier Transform Infrared Microspectrometry', *Anal. Chem.*, **60**, 427–433 (1988).
108. T. Greibrokk, B.E. Berg, S. Hoffmann, H.R. Norli, Q. Ying, 'Characterization of Polymer Additives by Supercritical Fluid Chromatography and by Liquid Chromatography', *J. Chromatogr.*, **505**, 283–291 (1990).
109. M. Ashraf-Khorassani, J.M. Levy, 'Quantitative Analysis of Polymer Additives in Low Density Polyethylene Using Supercritical Fluid Extraction/Supercritical Fluid Chromatography', *J. High Resolut. Chromatogr.*, **13**, 742–747 (1990).
110. M. Ashraf-Khorassani, D.S. Boyer, J.M. Levy, 'Optimization of Experimental Parameters for the Determination of Polymer Additives Using Online Supercritical Fluid Extraction. (SFE) – Supercritical Fluid Chromatography (SFC)', *J. Chromatogr. Sci.*, **29**, 517–521 (1991).
111. T.W. Ryan, S.G. Yocklovich, J.C. Watkins, E.J. Levy, 'Quantitative Analysis of Additives in Polymers Using Coupled Supercritical Fluid Extraction–Supercritical Fluid Chromatography', *J. Chromatogr.*, **505**, 273–282 (1990).
112. J.W. Oudsema, C.F. Poole, 'Determination of an Organotin Stabilizer in a Rigid Poly(vinyl chloride) Plastic by On-line Supercritical Fluid Extraction and Chromatography with Formic Acid Modified Carbon Dioxide and Flame Ionization Detection', *J. High Resolut. Chromatogr.*, **16**, 198–202 (1993).
113. T.P. Hunt, C.J. Dowle, G. Greenway, 'Analysis for Poly(vinyl chloride) Additives by Supercritical Fluid Extraction and Supercritical Fluid Chromatography', *Analyst*, **116**, 1299–1304 (1991).

Surface Energetics of Polymers and Rubbers, Characterization of

Richard R. Eley
ICI Paints, Strongsville, USA

Stanislaw Petrash
National Starch and Chemical Company, Bridgewater, USA

1 Introduction **1**
2 Surface Energy of Polymeric Materials **2**
2.1 Extrapolation Methods 2
2.2 Wetting (Contact Angle) Methods 4
2.3 Pendant Drop-shape Method for Polymer Surface Tension 12
3 Instrumentation and Experimental Design **13**
3.1 Experimental Set-up and Methodology 13
3.2 Choice of Test Liquids 14
3.3 Kinetic Effects and Environmental Control 15
4 Method Enhancements **15**
4.1 Digital Data Acquisition and Analysis 15
4.2 Automation of Data Collection and Analysis 15
5 Recent Developments: Contact Mechanics Method of Surface Energy Determination (Johnson, Kendall, Roberts Method) **16**
Acknowledgments **18**
List of Symbols **18**
Abbreviations and Acronyms **19**
References **19**

Adhesion of polymeric materials to various substrates, from metals to glass to other polymers, is a subject that continues to grow in importance, with today's technologies. The biocompatibility of polymer-based implantable devices and prostheses is largely determined by the tendency of the material surfaces to be 'wetted' or adhered to by proteins. Surface and interfacial energy relationships determine in large part whether the adhesive and adherend achieve the intimacy of molecular contact that is a requisite for good adhesion. An aspect of the behavior of polymer blends and of block copolymers is that, depending on the relative surface energy of the components, and on the phase morphology and chain mobility, lower energy components may segregate to the surface to the detriment of interfacial adhesion, or perhaps promoting heterophase nucleation. There are numerous both technological and scientific motivations for the characterization of polymer surface energy.

The objective in surface energy measurement is to characterize the free energy of a solid or a liquid surface (or, in general, an interface). The surface free energy per unit area (or surface tension) of polymers is a fundamental property. It may be measured in the solid, liquid, or solution state, but perhaps most often the intention is to refer the measured value to the solid state. Hence, one approach involves measuring the surface tension of polymer melts and low molecular weight liquid polymers by conventional methods and using one of several equations to extrapolate to values for the polymer in the solid state.

Methods for determining the surface energy of solid polymers usually involve measuring the contact angle of a sessile liquid drop, and this methodology will be discussed in some detail. A relatively new technique for direct polymer surface energy measurement based on Johnson, Kendall, Roberts (JKR) (contact mechanics) theory will be discussed.

1 INTRODUCTION

An *interface* is a part of a multi-component system, where the physicochemical parameters of the system undergo continuous change from the bulk properties of one component to the other. If one of the system components is air or vacuum, the interface becomes a *surface*. By definition, the *surface energy* γ is the work that must be done to create a unit of surface area. Hence, for a closed system [Equation 1]:

$$\gamma = \left(\frac{\partial G}{\partial A}\right)_{T,P} \tag{1}$$

where G is the Gibbs free energy of the total system and A is the surface area. If the work spent to create the surface is reversible, surface energy may be called surface tension. The terms 'surface tension' and 'surface energy' are often used interchangeably. However, this article deals in general with polymers and rubbers, which on a reasonable timescale are typically solid at normal temperatures and pressures. Since the process of creation of the surface of a solid is often irreversible, the more general term surface energy will be more applicable.

For the same reason, methods for the measurement of the surface energy of liquids that are based on

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the reversible creation of surface are impossible to implement for solids. The process of surface creation for polymers and rubbers in the solid state will involve a significant amount of energy spent irreversibly on various viscoelastic processes. Therefore, indirect methods of surface energy determination must be used.

2 SURFACE ENERGY OF POLYMERIC MATERIALS

Several methods for surface energy determination of polymers have been developed. They generally consist of two main approaches: (1) various extrapolation methods and (2) wetting methods based on Young's equation. Recently, determination of surface energy based on JKR theory by direct measurements of molecular adhesion has been reported. Most of these methods will be described here, with more attention given to those that produce most consistent and reliable results.

2.1 Extrapolation Methods

Surface energies of solid polymers may be obtained by measuring the surface tensions of polymers in a liquid state followed by extrapolation of the data to the surface energy of a solid. Variations of molecular weight and the temperature can be used to obtain a set of polymer samples in a liquid state (bearing in mind that surface tension is an increasing function of molecular weight). Conventional methods for measuring the surface tension of a liquid are then employed to obtain γ vs T or γ vs molecular weight relations. These relations are then used to extrapolate to the surface energy of a solid polymer.[1] Both of these extrapolation methods produce accurate results. On the other hand, extrapolation from concentration dependence, which is performed by dissolving the polymer in a solvent, may not give an accurate estimate because of possible solvent segregation effects.

2.1.1 Temperature Dependence Method

Like the surface tension of the low molecular weight liquids, the surface tension of polymers decreases with increase in temperature. For most polymers the surface tension will vary linearly over the normal temperature range (0–200 °C). Therefore, if the dependence of surface tension on temperature for a given polymer is measured experimentally at temperatures where the polymer is in the melt state, such data could be used to extrapolate surface tension to the lower temperature, at which the polymer will be solid. Figure 1 shows the typical dependence of surface tension on temperature for six different polymers from Roe.[2] Table 1 lists the values of surface tension γ at 150 °C and the extrapolated value for 20 °C. Detailed descriptions of experimental apparatus for measuring the surface tension of molten polymers are available in the literature.[2] The standard pendant drop technique is used and surface tension is calculated from the shape of the drop. The method of calculation is identical with that employed to obtain the surface tension of low molecular weight liquids. However, since the viscosity of most polymer melts is relatively high, long times (30 min–1 h) might be necessary for the drop to attain the equilibrium shape.

In general, during the analysis of these kinds of data, one should keep in mind that primary and secondary phase transitions may affect the extrapolated surface

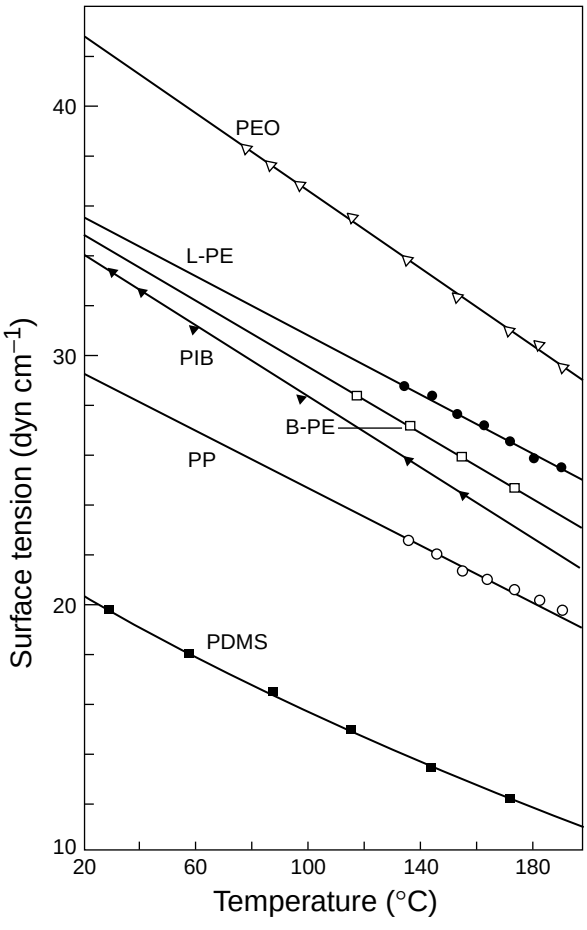

Figure 1 Surface tension values obtained from six polymer samples. Open triangles, poly(ethylene oxide) (Carbowax 600); filled circles, linear polyethylene (Alathon 7050); open squares, branched polyethylene (Bakelite DYLT); filled triangles, polyisobutylene (Oronite 128); open circles, polypropylene (atactic fraction); filled squares, polydimethylsiloxane (Dow Corning silicone 200 fluids), 106 and 6x104 cSt. [Reprinted with permission from R.-J. Roe, *J. Phys. Chem.*, **72**(6), 2013–2017 (1968). Copyright 1968, American Chemical Society.]

Table 1 Surface tension data for various polymer melts, extrapolated to solid state by Macleod's method. [Reprinted with permission from R.-J. Roe, *J. Phys. Chem.*, **72**(6), 2013–2017 (1968). Copyright 1968, American Chemical Society.]

Polymer	γ at 150 °C (dyn cm^{-1})	γ at 20 °C (extrapolated), (dyn cm^{-1})	$-d\gamma/dT$ at 150 °C	Macleod's exponent
Poly(ethylene oxide)	33.0	42.8	0.076	3.0
Linear polyethylene	28.1	35.6	0.058	3.2
Branched polyethylene	26.5	34.8	0.064	3.3
Polybutene	25.1	34.0	0.066	4.1
Polypropylene	22.1	29.4	0.056	3.2
Polydimethylsiloxane	13.6	20.4	0.048	3.5

energy of a polymer. Macleod's equation[3] could be used to analyze quantitatively the effects of melting and glass transitions on polymer surface energy. Although this equation was initially developed for small-molecule liquids, Wu[1,4] and Roe[5] have shown that it could also be applied to polymers. Macleod's equation is usually written as Equation (2):

$$\gamma = \gamma^0 \rho^n \tag{2}$$

where ρ is the density and γ^0 and n are positive, temperature-independent constants. For polymers, n usually varies from 3.0 to 4.4 (see Table 1). Macleod's equation is essentially an acknowledgement of the fact that variation of the surface energy with temperature is due to variations in density.

The analysis of Macleod's equation according to Wu[6] shows that above the glass transition temperature there should be a slight decrease in the slope of the plot of surface tension versus temperature. However, for most polymers this effect of second-order transitions on the surface tension–temperature dependence can often be neglected and the linear extrapolation method can still be used.

First-order transitions generally have a more pronounced effect on the surface energy of polymers and cannot be discounted. The accuracy of surface energy extrapolated from temperature dependence data depends on the degree of crystallinity of the actual polymer surface in question. For polymers that possess some degree of bulk crystallinity, the temperature dependence method could still work since such polymers are usually covered by an amorphous surface layer. This happens because the amorphous phase is less dense than the crystalline phase, and therefore it has lower surface energy. However, various degrees of surface crystallinity can be induced by casting the polymer against certain molds or by surface-localized mechanical stresses. Some high-energy mold surfaces were found to nucleate polyethylene, resulting in formation of transcrystalline surfaces, whereas low-energy mold surfaces gave amorphous polymer surfaces.[7,8]

For those polymers that have crystalline surfaces, Macleod's equation can be used to estimate the influence of crystallinity on surface energy of a polymer.[6] At the crystal–melt transition, the change from the crystalline density ρ_c to amorphous density ρ_a is discontinuous. It follows from Macleod's equation that the surface tension of the crystalline phase γ_c is related to that of amorphous phase γ_a by Equation (3):

$$\gamma_c = \left(\frac{\rho_c}{\rho_a}\right)^n \gamma_a \tag{3}$$

Thus, at the crystal–melt transition, the surface tension should change discontinuously, since γ is proportional to approximately the fourth power of the density [see Equation 2]. Because the crystalline density is usually higher than the amorphous density, the crystalline surface tension is expected to be much higher than the amorphous surface tension. Indeed, for polyethylene the Macleod's exponent n is 3.2, $\gamma_a = 35.7$ dyn cm^{-1}, and $\rho_a = 0.855$ g mL^{-1} at 20 °C. Polyethylene crystalline density ρ_c is 1.000 g cm^{-3}. Thus γ_c, calculated by the above equation, will be 58.9 dyn cm^{-1}, which compares fairly well with an experimental value of 53.6 dyn cm^{-1}.[9]

In general, the temperature-dependence method produces the most accurate results for amorphous polymers with relatively low melting temperature. If there is a possibility that owing to certain processing conditions the surface of the polymer could have some degree of crystallinity, it is recommended that other methods of determination of surface energy, such as wetting methods based on Young's equation, be used.

2.1.2 Molecular Weight Dependence Method

This method is also sometimes called the 'liquid homologues' method, because known surface tensions of liquid low molecular weight homologues are used to extrapolate the surface energy of a high molecular weight solid

Table 2 Comparison of surface energies obtained using the temperature dependence method and liquid homologue method. [Reprinted from S. Wu,[6] p. 80, by courtesy of Marcel Dekker, Inc.]

Polymer	Temperature (°C)	γ (dyn cm^{-1}): temperature dependence method	γ (dyn cm^{-1}): liquid homologue method	
			$\gamma^{1/4}$ vs M^{-1}	γ vs $M^{-2/3}$
Polyethylene	20	35.7	34.75	37.81
Polyisobutylene	24	34.0	34.50	35.62
Poly(dimethylsiloxane)	20	19.8	20.33	21.26
Polytetrafluoroethylene	20	23.9	23.94	25.85
Poly(ethylene glycol)	24	42.5	41.50	44.35
Polystyrene	176	29.5	29.50	29.97

polymer. The surface tension of a series of homologues will tend to increase with increasing molecular weight. LeGrand and Gaines[10,11] found that the surface tension of a homologous series varies with the number-average molecular weight taken to the 2/3 power [Equation 4]:

$$\gamma = \gamma_\infty - \frac{k_e}{M_n^{2/3}} \tag{4}$$

where γ_∞ is the surface tension of the polymer of infinite molecular weight and k_e is a constant and M_n is the number-average molecular weight.

Wu[1,4] also described an alternative expression for molecular weight dependence of surface tension, which could be obtained using Macleod's equation [Equation 5]:[6]

$$\gamma^{1/4} = \gamma_\infty^{1/4} + \frac{k_s}{M_n} \tag{5}$$

where k_s is a positive constant. The denominator 4 in the exponent is actually a value of Macleod's exponent, which in principle could vary from 3.0 to 4.4.

Both equations fit the experimental data for common polymers fairly well but tend to give different values for γ_∞. Wu pointed out that Equation (4) is derived from empirical observations and should be treated as an approximation. Indeed, as can be seen from Table 2, the $\gamma^{1/4}$ vs M^{-1} dependence gives values that are more consistent with values of surface tension obtained from temperature-dependence measurements.

2.2 Wetting (Contact Angle) Methods

2.2.1 Introduction

Van Oss recently stated, "Sessile drop contact angle measurement remains the most accurate method for determining the interaction energy between a liquid and a solid".[12] By use of liquids of known surface energy (or known surface energy components), one can exploit this method to determine accurately the surface energy of solid polymers and also the work of adhesion between a solid polymer and a liquid. The basic relationship among the equilibrium contact angle for a sessile (sitting) drop and the interfacial free energies of the three phases in contact (solid S, liquid L, and vapor V) was described (non-mathematically) by Young, in 1805, as [Equation 6]:[13]

$$\gamma_{LV} \cos\theta = \gamma_{SV} - \gamma_{SL} \tag{6}$$

where γ_{LV} is the free energy of the liquid–vapor interface, γ_{SL} that of the solid–liquid interface, and γ_{SV} that of the solid–vapor interface. Several authors write γ_S for γ_{SV} and γ_L for γ_{LV}. Note that, in order for γ_{LV} to represent the *reversible* work of surface area creation, the liquid is understood to be in equilibrium with its vapor. The surface tension of a liquid is sometimes defined in terms of the energy of cohesion of the liquid in vacuo [see Equation 12] and, in that sense, is defined without reference to a liquid–vapor interface, and hence is given the symbol γ_L. Similarly, the surface free energy of a solid is often symbolized γ_S. We will use both notations in this article. The notation γ_{SV} is used by some to represent the free energy of the solid in equilibrium with its own vapor,[14] but by most for the solid equilibrated with the vapor of the deposited liquid drop, after Bangham and Razouk.[15] The contact angle θ is between the substrate plane and the tangent to the liquid surface at the three-phase contact line. Note that by convention the contact angle is always measured *through the liquid* (Figure 2).

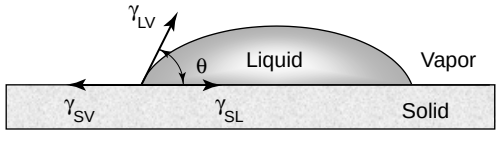

Figure 2 Force balance represented by Young's equation. γ_{LV}, liquid–vapor interfacial tension; γ_{SV}, solid–vapor interfacial tension; γ_{SL}, solid–liquid interfacial tension; θ, Young's contact angle.

Note also that $\gamma_{LV}\cos\theta$ is the projection of the vector γ_{LV} in the plane of the substrate. Cos θ is a measure of the balance of the energy of cohesion of a liquid and the energy of adhesion between a liquid and a solid.[12] Thus, Young's equation represents a force balance and an energy balance.

One issue relative to Equation (6) is whether γ_{SV} represents a 'clean' solid in the presence of vapor only or whether molecules may have adsorbed from the vapor phase to the solid surface. The difference between the free energy of a solid in vacuum γ_{S0} and in the vapor of the test liquid γ_{SV} is the 'equilibrium film pressure' π_e [Equation 7]:[15]

$$\gamma_{SV} = \gamma_{S0} - \pi_e \tag{7}$$

The origin of π_e is the evaporation and recondensation of a liquid on to the solid substrate, leading to a surface of lower energy and a higher contact angle for the drop on such a surface.[16] π_e is considered negligibly small for contact angles $>10^\circ$. Van Oss[17] argues that π_e may be neglected for most situations, with the possible exception of a low-energy liquid vapor over a high-energy substrate. If π_e is negligibly small, then obviously the surface free energy of a solid is essentially the same in vacuo as in contact with the liquid vapor. This will be true for most polymers, owing to their relatively low surface energies.

Figure 2 illustrates the balance of 'forces' or 'tensions' at the three-phase contact line, which determines the contact angle θ. The concept of a 'tension' at a solid interface is not a rigorous one (inasmuch as it cannot be measured as a 'tension'), and so, although the picture in Figure 2 is useful, it may be better to think of the contact angle in terms of the balance of surface or interfacial free energies, hence taking a thermodynamic rather than a mechanical approach.

Dupré[18] defined the work of adhesion between a solid and liquid (expressed here as the work to *separate* the solid and liquid phases, and following the above notation) according to Equation (8):

$$W_{SL}^{adh} = \gamma_{SV} + \gamma_{LV} - \gamma_{SL} \tag{8}$$

W_{SL}^{adh} (the negative of the free energy of adhesion ΔG_{SL}^{adh}) is the sum of contributions for interfaces gained minus that for interfaces lost. The right-hand side of Equation (8) represents the free energy change of the system when the liquid drop in Figure 2 is separated from the substrate, where the first term is the free energy per unit area of a vapor-equilibrated solid surface, the second term that of a freshly created liquid–vapor interface, and the last term that for the lost solid–liquid interface.

Eliminating γ_{SV} between Equations(6) and (8) yields the Young–Dupré equation [Equation 9]:

$$W_{SL}^{adh} = \gamma_{LV}(\cos\theta + 1) \tag{9}$$

giving the work of adhesion in terms of directly measurable quantities.

Cooper and Nuttall[19] devised the concept of a criterion for spreading, S [Equation 10]:

$$S = \gamma_{SV} - \gamma_{SL} - \gamma_{LV} \tag{10}$$

which was termed the *spreading coefficient* by Harkins.[90] If $S > 0$, that means the free energy of the system can be lowered by the liquid spreading over the solid, and it will do so spontaneously. Eliminating γ_{SV} between Equations (6) and (10) yields Equation (11):

$$S = \gamma_{LV}(\cos\theta - 1) \tag{11}$$

which gives the spreading coefficient in experimentally measurable quantities.

2.2.2 *Theoretical Models for Contact Angle Analysis*

The methodology of experimental contact angle measurement is deceptively simple, as the subject has not been lacking in controversy for much of its history and continues to be so. Some of the issues are experimental ones, most are theoretical. We will briefly survey the main theoretical models that have been widely used.

In Equation (6), the experimentally accessible variables are the contact angle θ and the liquid surface tension γ_{LV}. The other two parameters must be obtained by means of some theoretical model. We will discuss two main approaches, which have been broadly classified as *macroscopic* and *microscopic* theories.[20] These approaches differ principally in whether it is considered that the equilibrium contact angle is determined solely by the *total* surface energies of the solid and liquid (macroscopic theory), or whether instead the contact angle is influenced by specific chemical interactions of the liquid and substrate (e.g. hydrogen bonds, acid–base or electron donor–acceptor interactions), which is the premise of microscopic theories. The macroscopic approach, also known as the *Equation of state (ES) theory*, has been principally put forward by Neumann et al.[21–23]

In the view of the macroscopic theory, the contact angle is determined solely by the respective total ('macroscopic') surface tensions of the liquid and solid phases (in equilibrium with the vapor phase).[24] Therefore, two liquids of differing chemical composition but similar surface tensions should show the same contact angle on a given substrate. Opponents of this view claim this can be true only under special circumstances, where polar or Lewis acid–base interactions between liquid and solid are absent, and the approach has been heavily criticized.[25–28] Fowkes, van Oss, Good, and others have chiefly elaborated the microscopic approach, which has come to be known as the *Lifshitz–van der Waals acid–base (LWAB)*

theory. Advocates of the ES theory have criticized the latter in several papers.[22,29,30] The *geometric mean (GM) theory* of Owens and Wendt,[31] and the *harmonic mean (HM)* method of Wu,[6,32] are alternative microscopic approaches that are used by some.

One of the confusing, if not particularly controversial, aspects of the literature is the variability in nomenclature and symbolic formalism used. For example, some authors state the energetic relationships in terms of the Helmholtz free energy, F, while some write the very same equations in terms of the Gibbs free energy, G. Some prefer to include the vapor phase automatically in every surface or interfacial term, as in Equations (6) and (8), which has the merit at least that experimental measurements are normally made under such conditions. Some others, however, prefer to define the surface energies in vacuo, which is appropriate for surfaces of low affinity for the vapor phase.

One definition of the surface tension, for example, is half the free energy of cohesion of a material in a vacuum[33] (since two new interfaces are created – see Figure 3, step 1) (Equation 12):

$$\gamma \equiv \tfrac{1}{2} W^{\mathrm{coh}} \tag{12}$$

whereas elsewhere it has been defined according to Equation (1) for a liquid in equilibrium with its vapor, which could hardly be true in vacuo. (The free energy change is the negative of the work of cohesion, $\Delta G^{\mathrm{coh}} = -W^{\mathrm{coh}}$, and similarly for adhesion.)

2.2.2.1 Apolar Interactions: The Good–Girifalco–Fowkes Model The development of the microscopic approach began with a formulation of the energy of interaction of liquid–liquid interfaces by Girifalco and Good[34,35] in which the work of adhesion between liquid 1 and liquid 2 was given as the root mean square (GM) of their respective energies of cohesion [Equation 13]:

$$W_{12}^{\mathrm{adh}} = \Phi (W_1^{\mathrm{coh}} W_2^{\mathrm{coh}})^{1/2} \tag{13}$$

where Φ is a coefficient that adjusts for the balance of polar and apolar interactions, and is equal to unity for exclusively apolar interactions. The work of adhesion between distinct substances such as two immiscible liquids in contact is the energy cost to reversibly separate them, and hence is related to the surface tensions and the interfacial tension of the two liquids by Equation (14):

$$W_{12}^{\mathrm{adh}} = \gamma_1 + \gamma_2 - \gamma_{12} \tag{14}$$

representing the sum of the energies of the interfaces gained minus that of the interface lost (see Figure 3, step 2).

Combining Equations (13) and (14), together with Equation (12), we obtain Equation (15):

$$\gamma_{12} = \gamma_1 + \gamma_2 - 2\Phi(\gamma_1 \gamma_2)^{1/2} \tag{15}$$

which is the Girifalco–Good equation. In a similar fashion, Fowkes[36] reasoned that $(\gamma_1^{\mathrm{d}}\gamma_2^{\mathrm{d}})^{1/2}$ was equal to the amount by which the surface tension of substance 1 is lowered by the presence of substance 2, and vice versa. Therefore, the tension at the interface of 1 against 2 equals $\gamma_1 - (\gamma_1^{\mathrm{d}}\gamma_2^{\mathrm{d}})^{1/2}$ and of 2 against 1 is $\gamma_2 - (\gamma_1^{\mathrm{d}}\gamma_2^{\mathrm{d}})^{1/2}$. The total interfacial tension is the sum of these tensions, or [Equation 16]:

$$\gamma_{12} = \gamma_1 + \gamma_2 - 2(\gamma_1^{\mathrm{d}}\gamma_2^{\mathrm{d}})^{1/2} = \left(\sqrt{\gamma_1^{\mathrm{d}}} - \sqrt{\gamma_2^{\mathrm{d}}}\right)^2 \tag{16}$$

This is the Girifalco–Good–Fowkes equation, for which the GM 'combining rule' is valid only for interaction forces that are of the dispersion type (γ_i^{d}). [All terms in Equation (16) represent dispersion interactions.] The combining rule [Equations 16 and 13] establishes a quantitative law for the energy of interaction of unlike substances, for systems interacting only by dispersion forces.

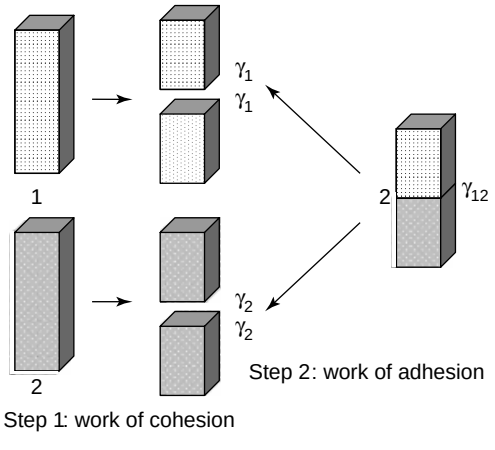

Figure 3 Schematic representations of the works of cohesion and adhesion; the free energies of cohesion and adhesion correspond to the reverse processes and are opposite in sign.

2.2.2.2 Polar Interactions: Geometric and Harmonic Mean Models Fowkes[37] introduced the idea that surface tension could be separated into dispersive and nondispersive components and that these contributions are additive [Equation 17]:

$$\gamma = \gamma^{\mathrm{L}} + \gamma^{\mathrm{D}} + \gamma^{\mathrm{K}} + \gamma^{\mathrm{H}} \tag{17}$$

where the superscripts stand for London–van der Waals dispersive (L), Debye (D) (dipole–induced dipole), Keesom (K) (dipole–dipole), and hydrogen-bonding (H) interactions, respectively. He subsequently collected the

nondispersive components (γ^D, γ^K, γ^H) into a general acid–base interaction term γ^{AB}.[38]

Owens and Wendt[31] and Kaelble[39] extended the GM combining rule to 'polar' interactions γ_i^p (e.g. hydrogen bonding, dipole–dipole, etc.) (although such an extrapolation was explicitly ruled out by Hildebrand[40]; see also Good and van Oss[41]) [Equation 18]:

$$\gamma_{SL} = \gamma_S + \gamma_L - 2(\gamma_S^d\gamma_L^d)^{1/2} - 2(\gamma_S^p\gamma_L^p)^{1/2} \quad (18)$$

which can be written as Equation (19):

$$\gamma_L(1+\cos\theta) = 2[(\gamma_1^d\gamma_2^d)^{1/2} + (\gamma_1^p\gamma_2^p)^{1/2}] \quad (19)$$

and solved for $\cos\theta$. Wu[6] proposed an HM formulation as giving better data for polymers [Equation 20]:

$$\gamma_L(1+\cos\theta) = 4\left(\frac{\gamma_S^d\gamma_L^d}{\gamma_S^d+\gamma_L^d} + \frac{\gamma_S^p\gamma_L^p}{\gamma_S^p+\gamma_L^p}\right. \quad (20)$$

The GM and HM methods can be used to obtain solid surface energies by measuring the contact angles of two liquids, one polar (e.g. water) and one apolar (e.g. diiodomethane), for which γ_L^d and γ_L^p are known. Two equations can then be set up to solve for the two unknowns γ_S^d and γ_S^p. The solid surface tension is then given by Equation (21):

$$\gamma_S = \gamma_S^d + \gamma_S^p \quad (21)$$

Van Oss and Good take strong issue with the assumption that polar interactions can be mathematically treated in the same fashion as apolar interactions, i.e. that polar interactions are 'symmetrical' and therefore subject to the GM combining rule.[42] One notable problem with this approach, van Oss[25] points out, is that by analogy with Equation (16), we have Equation (22):

$$\gamma_{12}^p = \left(\sqrt{\gamma_1^p} - \sqrt{\gamma_2^p}\right)^2 \quad (22)$$

and hence γ_{12}^p can only assume positive values (which means the free energy of interaction of a solute in water with itself can only be negative or at most zero) and consequently the solubility of neutral polymers in water cannot be predicted. Van Oss and Good show[42] that the solubility in water of poly(ethylene oxide) and dextran are underpredicted by several orders of magnitude by the GM method. The van Oss–Chaudhury–Good (or LWAB) model treats 'polar' interactions more generally as Lewis acid–base in nature and recognizes that the acidity and basicity of a material are not, in general, of equal strength. Therefore, such interactions between different molecules will be 'asymmetric' and must be treated according to a new combining rule (see below). In a comparative study, Dalal[43] showed that the GM and HM models give similar results and adequately represent the total solid surface energies of a number of polymers, but found the GM method to be somewhat more consistent. Although the GM or HM approaches have been found useful[44] and in some cases give comparable results to the LWAB model,[45] they have also been shown to be deficient relative to the LWAB method.[20]

2.2.2.3 Generalized Acid–Base Interactions: The Lifshitz–van der Waals Acid–Base Model Van Oss et al.[46–48] modified the Fowkes approach by combining London dispersion forces and Debye and Keesom interactions under the heading Lifshitz–van der Waals (LW) interactions, because the Lifshitz theory shows that all are similarly electrodynamic and obey the same general equations.[49] Thus, the apolar surface energy component γ^{LW} is composed of additive components [Equation 23]:

$$\gamma^{LW} = \gamma^L + \gamma^D + \gamma^K \quad (23)$$

and excludes hydrogen bonding. The total surface tension γ for substance i is now given by Equation (24):

$$\gamma_i = \gamma_i^{LW} + \gamma_i^{AB} \quad (24)$$

where AB represents generalized electron donor–acceptor or Lewis acid–base interaction energies, which include hydrogen bonding. The inherent assumption in the LWAB theory is that, in addition to the dispersive-type interactions, charge-transfer (general acid–base) molecular interactions take place across the liquid–solid interface, which determine the contact angle.

Van Oss et al. recognized that the electron donor–acceptor (generalized acid–base) interactions are asymmetric, i.e. the acidity and basicity of a given molecule are, in general, not equal in magnitude and its acid–base interactions with other substances will likewise not be equal. This requires a different combining rule than for the LW interactions, which are symmetrical. According to van Oss et al.[46,47] the acid–base component of the surface energy for substance i is given by Equation (25):

$$\gamma_i^{AB} = 2\sqrt{\gamma_i^+\gamma_i^-} \quad (25)$$

and the work of adhesion between substances 1 and 2 according to the postulated acid–base combining rule of van Oss et al. is given by Equation (26):[50]

$$W_{12}^{adh,AB} = 2\left(\sqrt{\gamma_1^+\gamma_2^-} + \sqrt{\gamma_1^-\gamma_2^+}\right) \quad (26)$$

where $\gamma^+ \equiv$ Lewis acid (electron pair acceptor) surface free energy parameter and $\gamma^- \equiv$ Lewis base (electron pair donor) surface free energy parameter. Note that the combining rule sums the acid component of substance 1 interacting with the base component of substance 2 and the base component of 1 interacting with the acid

component of 2. By analogy with Equation (24), we write the total interfacial tension as Equation (27):[50]

$$\gamma_{12} = \gamma_{12}^{LW} + \gamma_{12}^{AB} \tag{27}$$

The combining rule for the acid–base component of the interfacial tension, again for substances 1 and 2, is expressed as Equation (28):

$$\gamma_{12}^{AB} = 2\left(\sqrt{\gamma_1^+\gamma_1^-} + \sqrt{\gamma_2^+\gamma_2^-} - \sqrt{\gamma_1^+\gamma_2^-} - \sqrt{\gamma_1^-\gamma_2^+}\right)$$
$$= 2\left(\sqrt{\gamma_1^+} - \sqrt{\gamma_2^+}\right)\left(\sqrt{\gamma_1^-} - \sqrt{\gamma_2^-}\right) \tag{28}$$

which, combined with Equations (16) and (27), leads to the expression for the total interfacial tension [Equation 29]:

$$\gamma_{12} = \left(\sqrt{\gamma_1^{LW}} - \sqrt{\gamma_2^{LW}}\right)^2 + 2\left(\sqrt{\gamma_1^+\gamma_1^-} + \sqrt{\gamma_2^+\gamma_2^-} - \sqrt{\gamma_1^+\gamma_2^-} - \sqrt{\gamma_1^-\gamma_2^+}\right) \tag{29}$$

or [Equation 30]

$$\gamma_{12} = \gamma_1^{LW} + \gamma_2^{LW} - 2\sqrt{\gamma_1^{LW}\gamma_2^{LW}} + 2\left(\sqrt{\gamma_1^+} - \sqrt{\gamma_2^+}\right) \times \left(\sqrt{\gamma_1^-} - \sqrt{\gamma_2^-}\right) \tag{30}$$

The full combining rule for the work of adhesion for both apolar and acid–base components is expressed as Equation (31):[50]

$$W_{12}^{adh} = \gamma_1 + \gamma_2 - \gamma_{12}$$
$$= 2\left(\sqrt{\gamma_1^{LW}\gamma_2^{LW}} + \sqrt{\gamma_1^+\gamma_2^-} + \sqrt{\gamma_1^-\gamma_2^+}\right) \tag{31}$$

Using Equation (9) and writing in terms of a solid–liquid interface, we obtain Equation (32):

$$\gamma_{LV}(\cos\theta + 1) = 2\left(\sqrt{\gamma_S^{LW}\gamma_L^{LW}} + \sqrt{\gamma_S^+\gamma_L^-} + \sqrt{\gamma_S^-\gamma_L^+}\right) \tag{32}$$

Sets of γ_L^{LW}, γ_L^+, and γ_L^- data have been generated for a number of test liquids.[12] That leaves three unknowns, γ_S^{LW}, γ_S^+, and γ_S^-, in Equation (32). By measuring the contact angles θ and total surface tensions γ_{LV} for three different liquids (generally a nonpolar liquid and two polar liquids are used), a set of simultaneous equations can be set up and solved.[50]

Because the acid–base interaction terms γ^+ and γ^- cannot be determined independently by the LWAB method, experimental values for these are referenced to assumed values for water. Van Oss et al. used an experimental value for γ_{LW} of 21.8 mJ m^{-2} for water, and assumed values of $\gamma^+ = \gamma^- = 25.5$ mJ m^{-2}. Whether this assumption of equal acid–base character for water is correct for interfacial interactions has been questioned.[51] However, Good asserts that varying the ratio γ^+/γ^- for water over at least a 2 : 1 range does not materially affect the calculated dependent parameter values.[50]

2.2.2.4 The Equation of State Model The Neumann ES theory postulates the existence of an explicit relation of the form of Equation (33):

$$\gamma_{SL} = f(\gamma_{SV}, \gamma_{LV}) \tag{33}$$

which, when combined with Young's equation [Equation 6], allows the determination of γ_{SV} and γ_{SL} from a single measurement of the contact angle for a liquid of known surface tension γ_{LV}. ES theory holds that it is the *total* surface energy of a liquid and the solid that determines its contact angle on a given solid surface, in contrast to the LWAB theory, which divides the total surface free energy into apolar and acid–base subfactors. In ES theory, only a single liquid is required to characterize the surface free energy of a solid. Neumann et al.[21] obtained the empirical Equation (34) for γ_{SL}:

$$\gamma_{SL} = \frac{\left(\gamma_{LV}^{1/2} - \gamma_S^{1/2}\right)^2}{1 - 0.015(\gamma_{LV}\gamma_S)^{1/2}} \tag{34}$$

A relationship to $\cos\theta$ [Equation 35] was derived from Equation (34) and Young's equation:

$$\cos\theta = \frac{(0.015\gamma_S - 2)(\gamma_{LV}\gamma_S)^{1/2} + \gamma_{LV}}{\gamma_{LV}[0.015(\gamma_{LV}\gamma_S)^{1/2} - 1]} \tag{35}$$

Since Equation (35) is not simple to solve, tables are provided from which γ_S can be obtained, given the value of $\cos\theta$.[52] Equation (35) was later modified,[22] in order to avoid a possible mathematical singularity, to Equation (36):

$$\cos\theta = -1 + 2(\gamma_S/\gamma_{LV})^{1/2}\exp[-0.0001247(\gamma_{LV} - \gamma_S)^2] \tag{36}$$

2.2.2.5 Comparison of Contact Angle Methods for Determination of Surface Energy Schneider[20] did a comparative study of the GM, ES, and LWAB methods on a series of four substrates ranging from high energy to low energy (germanium, stainless steel, polypropylene and Perspex), using four liquids: water, formamide, α-bromonaphthalene, and diiodomethane. The substrates were measured clean and after controlled exposure to various solutions, in order to form a 'conditioning film'. Part of the purpose was to test how well the various theories adhere to their basic premises. A foundational assertion of ES theory is that the contact angles of liquids on solids are wholly defined by their total surface tensions and

that therefore the calculated solid surface energy should be independent of the liquid chosen. Schneider found this approximately true for medium-energy substrates ($\gamma_S \approx 40-50\,mJ\,m^{-2}$), but on high- ($>60\,mJ\,m^{-2}$) and low-energy ($<30\,mJ\,m^{-2}$) substrates observed a dependence on the nature of the probe liquid. The γ_S value obtained on high-energy solids was proportional to the polarity of the test liquid and on low-energy surfaces inversely proportional to liquid polarity. Others[53] have also found liquid-specific results using the ES method. Schneider concluded that the ES model is limited in its ability to characterize a wide range of substrate types, and its results are not consistent with ES theory. He found that the ES, GM, and LWAB methods did produce similar trends in data across substrate types, but the quantitative agreement of data across methodologies was poor. Schneider further concluded that the GM model tended to underestimate the dispersive energy component on medium- to high-energy substrates. Results obtained with the LWAB method were internally consistent, but some sensitivity of data to the order of experimental determination of the acid–base surface energy components was observed. Schneider recommended a specific combination of liquids and experimental protocol to obtain most consistent results.

Dalet et al.[54] examined the performance of the GM, HM, and LWAB models in a unique study where the Lewis basicity of a polymeric substrate was varied in a controlled way by grafting varying amounts of base-functional polymer groups on to comb-like block copolymers. The base-functional group, (4-hydroxyethylbenzoic acid (HEBA)), was grafted on to side chains of poly(chloroethylvinyl) ether. Five levels of HEBA grafting were used: 0, 15, 50, 75, and 100%. Polymer films 1 mm thick were cast from solution on aluminum sheets. The test liquids chosen were water, ethylene glycol, and tricresyl phosphate, which did not react with or swell the polymers.

Dalet et al. found the GM and HM methods to give similar values for the LW component of solid surface energy, but different results for the AB component. The GM and HM work of adhesion data W_{SL}^{LW} and W_{SL}^{AB} (following Dalet's usage) followed the trend in polymer composition well. W_{SL}^{LW} was several times larger than W_{SL}^{AB} but increased only about 10% or less over the range of substitution, whereas W_{SL}^{AB} increased by 250–400%. They concluded, however, that γ_S^{AB} is not a fundamental property of a solid, at least for currently available calculation methods, because its value is liquid dependent (see van Oss's comments on monopolar solids, below.) It is possible, however, that if the procedure and recommendations of Schneider[20] were to be followed, the data might be more internally consistent. It is perhaps reasonable for this system that γ_S^{AB} would not show a trend, since the Lewis acid and base character should be changing in opposite directions, with substitution. The LWAB method should reveal these opposing trends in $\gamma_S{}^+$ and $\gamma_S{}^-$. The data of Dalet et al. show a regular increasing trend in $\sqrt{\gamma_S{}^-}$ with graft content, but mostly negative values for $\sqrt{\gamma_S{}^+}$, the meaning or validity of which is a current theoretical question.[50] Table 3 gives a portion of Dalet's data.

Correia et al.[55] compared the Owens and Wendt (GM), Neumann (ES), and van Oss et al. (LWAB) theories. They used water, glycerol and diiodomethane as test liquids on a side-chain liquid crystalline polymeric substrate. In the process, they demonstrated the significant variability in calculated results that eventuate from using an array of published values of γ^d, γ^p, γ^{LW}, γ^{AB}, γ^+ and γ^- for the liquids, in various combinations. It appears that the variation in values of surface tension components has more of an effect on the LWAB results than on the GM data. Once again, values of polymer surface energy calculated from ES theory were found to depend on the polarity of the test liquid.

Table 4 compares GM and LWAB data for four solid polymers and Table 5 shows the work of adhesion between several liquid–solid polymer pairs for the three models, ES, GM, and LWAB, calculated from literature data by Correia et al.[55] They observed, in general, more

Table 3 Comparison of GM data obtained for grafted poly(chloroethylvinyl) ether using both water and ethylene glycol as the polar test liquid, together with the LWAB acid–base components obtained using water and ethylene glycol as test liquids. [Reprinted from Dalet et al.[54] by permission of VSP International Science Publishers.]

HEBA[a] in polymer (%)	GM model						LWAB acid–base components			
	Water			Ethylene glycol			$\sqrt{\gamma^+}$	γ^+	$\sqrt{\gamma^-}$	γ
	W_{SL}^{LW}	W_{SL}^{AB}	γ_S^{AB}	W_{SL}^{LW}	W_{SL}^{AB}	γ_S^{AB}				
0	54.4	5.8	0.2	62.7	1.7	0.2	+0.01	0	0.57	0.3
15	55.6	9.6	0.5	64.2	1.8	0.1	−0.15	0?	1.03	1.1
50	56.1	14.2	1.0	64.7	2.1	0.3	−0.15	0?	1.57	2.5
75	57.5	20.5	2.0	66.3	2.7	0.3	−0.27	0?	2.29	5.2
100	58.1	26.1	3.4	67.0	4.3	0.5	−0.27	0?	2.85	8.1

[a] HEBA.

Table 4 Values for total, dispersive (γ^d), polar (γ^p), LW (γ^{LW}), acceptor (γ^+), and donor (γ^-) surface energy components for low-energy solids from contact angle measurements. [Reprinted from Correia et al.[55] by permission of Academic Press.]

	γ_{tot}	γ^d	γ^p	γ^{LW}	γ^+	γ^-
Liquids						
Water	72.8	21.8	51.0	21.8	25.5	25.5
Glycerol	63.3	33.6	29.7	34.0	3.92	57.4
Formamide	57.3	28.0	29.3	39.0	2.28	39.6
DMSO[a]	43.5	29.0	14.5	35.6	0.5	32.0
Diiodomethane	50.8	50.42	0.38	50.8	0.72	0.0
Solids						
Polystyrene	42.0	41.4	0.6	42.0	0.0	1.1
PVC[a]	41.5	40.0	1.5	43.0	0.04	3.5
PMMA[a]	44.9	38.95	5.95	43.9	0.03	16.2
Cellulose acetate	40.2	35.0	5.2	35.0	0.3	22.7

[a] DMSO, dimethyl sulfoxide; PVC, poly(vinyl chloride); PMMA, poly(methyl methacrylate).

Table 5 Comparison of work of adhesion data calculated for three models: ES model, GM model, and LWAB model. [Reprinted from Correia et al.[55] by permission of Academic Press; ES model data recalculated by present authors.]

Liquid–solid system	ES model[a]	GM model[b]	LWAB model[c]
Water–polystyrene	45.1	35.6	35.6
Glycerol–polystyrene	47.8	41.5	39.9
Formamide–polystyrene	47.4	38.2	42.1
DMSO–polystyrene	42.7	37.6	39.4
Diiodomethane–polystyrene	45.7	46.2	47.1
Water–PVC	44.7	38.3	41.1
Glycerol–PVC	47.4	43.3	43.4
Formamide–PVC	47.0	40.1	45.0
DMSO–PVC	42.5	38.7	41.6
Diiodomethane–PVC	45.4	45.7	48.3
Water–PMMA	47.1	46.6	52.1
Glycerol–PMMA	50.2	49.5	47.9
Formamide–PMMA	49.5	46.2	48.5
DMSO–PMMA	44.2	42.9	43.4
Diiodomethane–PMMA	47.5	45.8	50.6
Water–cellulose acetate	43.8	43.9	54.4
Glycerol–cellulose acetate	46.4	46.7	48.1
Formamide–cellulose acetate	46.9	43.6	47.6
DMSO–cellulose acetate	41.8	40.5	41.8
Diiodomethane–cellulose acetate	44.5	43.4	46.2

[a] Calculated from $\Phi(\gamma_S\gamma_L)^{1/2}$, where $\Phi = e^{-0.0001247(\gamma_L-\gamma_S)^2}$.
[b] Calculated from $(\gamma_S^d\gamma_L^d)^{1/2} + (\gamma_S^p\gamma_L^p)^{1/2}$.
[c] Calculated from $(\gamma_S^{LW}\gamma_L^{LW})^{1/2} + (\gamma_S{}^+\gamma_L{}^-)^{1/2} + (\gamma_S{}^-\gamma_L{}^+)^{1/2}$.

reasonable results from the GM and LWAB models than for the ES model. For example, in Table 5, the work of adhesion to polystyrene, an essentially dispersive polymer, should be less for water than for diiodomethane. This is true for GM and LWAB, but the reverse for the ES data. Second, considering only water, a bipolar molecule (able to function as either an acid or a base), the work of adhesion should increase with increasing basicity of the polymer (increasing as one goes down the table). This is especially true for the LWAB and also true for the GM data, whereas the ES data are almost constant. For the interaction of water with the most basic polymers (PMMA and cellulose acetate), the LWAB gives the highest values of the adhesion energy, consistent with the expectation of greatest donor–acceptor interactions across the solid–liquid interface.

It appears to be a characteristic of the existing theories that the quality and internal consistency of results are dependent, to one degree or another, on the choice of test liquid or of the particular combination of liquids. It also seems to be true that proponents of a particular model can find experimental cases for which the opposing model fails. This suggests that either the values for certain of the subfactors or surface tension components are in error, or there is something lacking in the physical models, or both. Thus, it is often heard that 'there is no ''correct'' model', and the choice of procedure effectively becomes one of personal taste. Nevertheless, the present authors feel that while there remain issues with it, the LWAB approach is a more convincing and satisfying model overall, being conceptually more complete as well as intuitively more appealing. The methodology will no doubt see continued improvements in coming years.

A final comment on the methodology takes note of van Oss's remark on the value (or lack of it) of measuring the total surface energy of a solid γ_S, especially by the GM or HM method (i.e. undecomposed into its acid–base components).[25] Van Oss states that the total surface energy of a solid 'generally has little correlation with its polarity or hydrophilicity', particularly for the GM and HM methods, because they cannot account for a zero polar surface energy component, which van Oss claims is not an unusual occurrence. Hence, from

$$\gamma_S^{AB} = 2\sqrt{\gamma_S{}^+\gamma_S{}^-} \tag{37}$$

if either $\gamma_S{}^+$ or $\gamma_S{}^-$ are zero (a *monopolar* substance), $\gamma_S^{AB} = 0$ whatever the value of the remaining parameter. The result, from Equation (24), is then Equation (38):

$$\gamma_S = \gamma_S^{LW} + \gamma_S^{AB} = \gamma_S^{LW} \tag{38}$$

This describes the situation where an otherwise quite polar solid can have a surface energy that is solely determined by the dispersive component. Van Oss's argument is that the GM/HM methodology would be insensitive to this state of affairs. The LWAB analysis, in contrast, would presumably correctly account for the monopolar nature of such a solid.

2.2.3 Zisman Method

Zisman found the cosine of the advancing equilibrium contact angle to be a linear function of the liquid surface tension, for different series of chemically similar liquids (homologues), as expressed in Equation (39):[56]

$$\cos\theta = a - b\gamma_{LV} \tag{39}$$

A plot of $\cos\theta$ vs γ_L extrapolates, at $\cos\theta = 1$ (zero contact angle), to what Zisman termed a *critical surface tension* γ_c (Figure 4). The term 'critical' means that a liquid whose surface tension is above that value will not be a 'wetting' liquid for that substrate (i.e. complete wetting requires $\gamma_{LV} \leq \gamma_c$). For a given substrate, different homologous series of liquids yield nearly the same intercept, leading Zisman to propose that this value was characteristic of the substrate.

Zisman's method considers the total surface tension, rather than its components. It has been pointed out[57] that Zisman's correlation works only for liquids having primarily dispersion interactions with the substrate (i.e. for nonpolar liquids). It fails for polar liquids, where hydrogen bonding or acid–base interactions are possible, showing significant curvature in the Zisman plot. In such cases a liquid can spread on a substrate owing to these specific interactions even though its total surface tension may be above the γ_c value determined using apolar liquids.[57] The values of γ_c may be useful for characterizing the energy of surfaces relative to one another, but van Oss[25] states that characterization with a single nonpolar liquid is sufficient, yielding the dispersion component of the substrate surface energy γ_S^{LW}. Although Zisman's method is empirical and not applicable when polar interactions are present, it has been found useful for practical comparison purposes.[44] Table 6[56] lists γ_c values for a number of solid polymers.

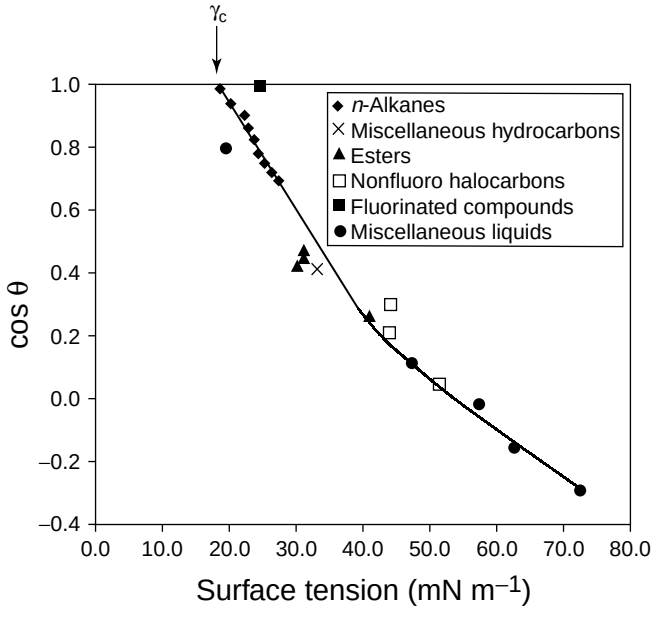

Figure 4 A Zisman plot of $\cos\theta$ vs surface tension for various liquids, sessile drops on polytetrafluoroethylene. [Reprinted with permission from W.A. Zisman, in 'Contact Angle, Wettability, and Adhesion', *Advances in Chemistry Series*, No. 43, American Chemical Society, Washington, DC, Chapter 1, 1964. Copyright 1964, American Chemical Society.]

Table 6 Values of Zisman's critical surface tension (γ_c) for polymeric solids. [Reprinted with permission from W.A. Zisman, in *Advances in Chemistry Series*, No. 43, American Chemical Society, Washington, DC, Chapter 1, 1994. Copyright 1964, American Chemical Society.]

Polymeric solid	γ_c at 20 °C (mN m^{-1})
Polymethacrylic ester of perfluoroctanol	10.6
Polyhexafluoropropylene	16.2
Polytetrafluoroethylene	18.5
Polytrifluoroethylene	22
Poly(vinylidene fluoride)	25
Poly(vinyl fluoride)	28
Polyethylene	31
Polytrifluorochloroethylene	31
Polystyrene	33
Poly(vinyl alcohol)	37
PMMA	39
PVC	39
Poly(vinylidene chloride)	40
Poly(ethylene terephthalate)	43
Poly(hexamethyleneadipamide)	46

Fowkes[36] proposed a modification to Zisman's approach, having a sounder basis in theory. If we rewrite Equation (16) for an S, L, V system, we have Equation (40):

$$\gamma_{SL} = \gamma_{SV} + \gamma_{LV} - 2(\gamma_{SV}^d \gamma_{LV}^d)^{1/2} \tag{40}$$

Equation (6) can be written as Equation (41):

$$\gamma_{LV}\cos\theta = \gamma_{SV} - \gamma_{SL} - \pi_e \tag{41}$$

Eliminating γ_{SV} between Equations (40) and (41) yields Equation (42):

$$\gamma_{LV}\cos\theta = 2(\gamma_{SV}^d \gamma_{LV}^d)^{1/2} - \gamma_{LV} - \pi_e \tag{42}$$

and dividing through by γ_{LV} gives finally Equation (43):

$$\cos\theta = \frac{2(\gamma_{SV}^d \gamma_{LV}^d)^{1/2}}{\gamma_{LV}} - 1 \tag{43}$$

where the π_e term is zero for most liquids on polymers (hence we could write γ_S for γ_{SV}). Therefore, a plot of $\cos\theta$ vs $(\gamma_{LV}^d)^{1/2}/\gamma_{LV}$ should be a straight line intercepting the ordinate at $\cos\theta = -1$, having a slope of $2(\gamma_S^d)^{1/2}$,

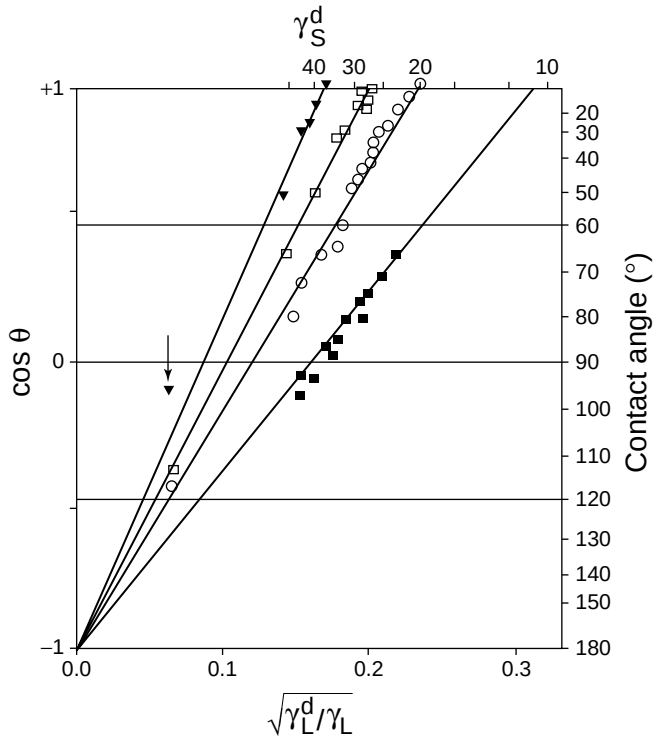

Figure 5 Fowkes alternative plot of Zisman data to obtain solid surface free energies for apolar systems. [Reprinted with permission from F.M. Fowkes, in *Chemistry and Physics of Interfaces*, ed. S. Ross, American Chemical Society, Washington, DC, 1–12, 1965. Copyright 1965, American Chemical Society.]

Table 7 Values of γ_S^d (mJ m^{-2}) for low-energy solids from Fowkes' method of contact angle analysis. [Reprinted with permission from F.M. Fowkes, in *Chemistry and Physics of Interfaces*, ed. S. Ross, American Chemical Society, 1–12, 1965. Copyright 1965, American Chemical Society.]

Perfluorododecanoic acid on Pt	10.4
Perfluorodecanoic acid on Pt	13.1
Polyhexafluoropropylene	18.0
Polytetrafluoroethylene	19.5
n-$C_{36}H_{74}$ wax crystal	21.0
n-Octyldecylamine on Pt	22.1
Paraffin wax	25.5
Poly(trifluoromonochloroethylene) (Kel-F)	30.8
Polyethylene	35.0
Polystyrene	44.0

as shown in Figure 5.[36] Clearly, Equation (43) is fairly successful in correlating data for solid–liquid pairs and provides a means of obtaining the surface free energy of a dispersive solid γ_S^d, which can be used, with Young's equation, to obtain γ_{SL}. Table 7[36] shows values of γ_S^d obtained with Equation (43) for several low-energy solids. Liquids that have 'polar' interactions with the substrate (e.g. hydrogen bonding or electron donor–acceptor interactions) will fall above the line in Figure 5.[54]

2.3 Pendant Drop-shape Method for Polymer Surface Tension

Bashforth and Adams[58] performed a numerical analysis on the Laplace equation for the pressure difference ΔP across a curved interface as a function of surface tension γ and the principal radii of curvature R_1 and R_2 [Equation 44]:[59]

$$\Delta P = \gamma \left(\frac{1}{R_1} + \frac{1}{R_2} \right. \tag{44}$$

from which extended tables were produced relating (among other parameters) sessile and pendant drop shapes to liquid surface tension. A more detailed discussion of the method can be found in Passerone and Ricci.[60] Bashforth–Adams drop shape analysis is now readily accessible via computer video image analysis and look-up tables. Figure 6 shows a typical pendant drop shape analysis performed on an FTÅ200 goniometer–tensiometer instrument. An automated method for both pendant and sessile drop analysis has been described by Hansen and Rødsrud.[61]

Pendant drop methodology provides a fairly simple way of measuring the surface tension of molten polymers. One technique for doing so involves filling a syringe needle with the molten polymer and then allowing it to cool and solidify in the needle. The needle is then inserted into an electrically heated chamber (such as supplied by First Ten Ångstroms, Inc., Portsmouth, VA, USA) where the polymer needle inside the chamber is brought to desired

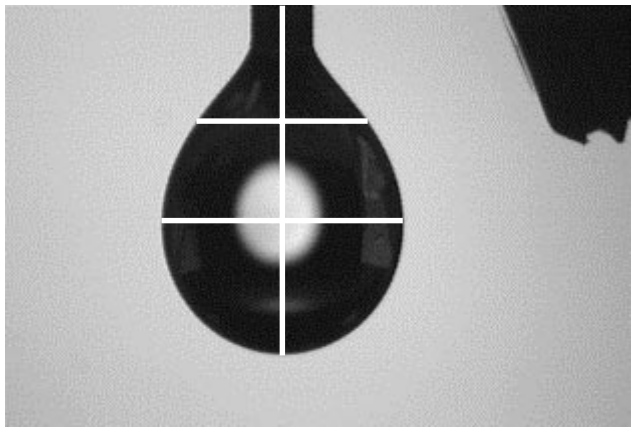

Figure 6 Pendant drop shape analysis by the Bashforth–Adams technique on the FTÅ200 goniometer–tensiometer system. Maximum width of drop and width at a height equal to twice the distance from maximum diameter line to base of hanging drop are used with a look-up table to find the surface tension.

temperature. On melting, the polymer extrudes to form a drop. The drop image can then be photographed or captured electronically for analysis. The density of the polymer melt must be known to complete the analysis.

3 INSTRUMENTATION AND EXPERIMENTAL DESIGN

3.1 Experimental Set-up and Methodology

There are several experimental issues in the measurement of contact angles, which are not yet fully understood in all respects. These issues include, for example, the influence of an adsorbed vapor layer, the effects of substrate roughness, methods of sessile drop formation, and, ultimately, how to obtain a 'true' (thermodynamic) equilibrium contact angle, referred to by some as the 'Young's contact angle', θ_Y.[62] Decker et al.[63] presented a discussion of the complexities of contact angle measurement physics, especially on imperfect surfaces. Zisman[56] has pointed out the importance of a microscopically smooth substrate for accuracy in contact angle measurement, particularly for low contact angles, for which the error goes up significantly for rough substrates. Marmur[64] referred to the contact angle measured on a smooth substrate as the 'intrinsic contact angle', all others as the 'apparent contact angle'.

Contact angle methods require the following:

1. the assumption that Young's equation has thermodynamic validity;
2. formation of a sessile liquid drop on a smooth solid surface (or sessile gas bubble on an inverted surface);
3. an accurate methodology for contact angle measurement.

The thermodynamic rigor of the Young equation has been established and is generally accepted.[50,57] The criterion of necessary smoothness is an important one, because, although roughness at a certain scale can enhance wetting (hence platinum Wilhelmy plates are usually etched to improve wetting), in general a liquid on a rough surface will display a higher contact angle than the thermodynamic equilibrium value.[64,65]

Substrates should be flat (at least locally), smooth, rigid, nonreactive, insoluble and nonabsorptive (nonswelling) toward the test liquid. Smooth polymer coatings can be applied from solution by dip coating, spin coating, or drawdown techniques. Often, smooth mica is used as a support for the polymer film. Insoluble or infusible polymers (e.g. Teflon® FEP) can be used as molded plaques, or formed into films by heat-pressing on to a quartz-plate substrate, for example.[65] Methods of solid surface preparation have been described by Li and Neumann.[22]

3.1.1 Sessile Drop Formation

Timmons and Zisman developed a method for applying liquid drops of repeatable volume to a surface by use of a thin platinum wire.[66] Good[50] recommends dispensing the liquid via a syringe pump and syringe needle, especially where the needle is left in the drop, allowing dynamic contact angle measurements to be made by pumping liquid into the drop, causing advancement of the liquid edge onto the substrate. This can be followed by withdrawal of liquid to measure the receding contact angle. Kwok and Neumann[65] used a method whereby the drop is formed by pumping liquid through a small hole in the substrate. In this manner, the liquid volume can be cycled as above for dynamic contact angle measurement, but since there is no captive syringe needle the drop shape remains undistorted. This is necessary because the axisymmetric drop shape analysis (ADSA) method of contact angle analysis used by Neumann et al. requires an undistorted sessile drop.[67]

Drop deposition by a blunt-tip syringe needle should be done by the 'touch-off' method, whereby the liquid drop, when it has grown to the desired size, is gently touched on the solid substrate. It should then detach from the needle and transfer to the solid surface. The drop should not be allowed to fall from the needle to impact on the substrate, as the drop's kinetic energy can overspread the drop, resulting in too wide a drop and too low a contact angle. This can be conveniently done by using a rack-and-pinion three-axis adjustable specimen stage. The specimen drop may then be lowered away from contact with the needle, or the needle left immersed in the deposited drop to permit further additions or withdrawals of liquid. In cases where a nonwetting liquid is used, detachment may be difficult for small drops but can be facilitated by use of a Teflon® needle.

3.1.2 Equilibrium and Dynamic Contact Angles

A gently deposited, partially wetting sessile drop should reach an essentially constant diameter and contact angle over a period of seconds to minutes, depending on its viscosity, that is, unless there is some active mechanism for continuous spreading, such as absorption into a porous substrate, time-dependent transport of a surfactant to the wetting edge, or perhaps reorganization of the surface structure of the substrate. This equilibrium advancing contact angle θ_e is often used for calculation of the substrate surface energy. Good[50] strongly recommends that contact angles be measured by dynamic methods, i.e. by sequentially forcing the liquid front to advance (giving the advancing dynamic contact angle θ_a), then retreat over

the solid substrate (receding dynamic contact angle θ_r). Not to do so is to sacrifice useful information, according to Good.

Another technique for measuring hysteresis is the 'inclined plate' method, where a drop rests on a tilting plate. At the angle of inclination where the drop just begins to move, the contact angle at the 'leading' edge of the drop is taken as a measure of θ_a and the angle at the trailing edge as θ_r. A drawback of the method is that the measured angles are dependent on drop size.[50]

An alternative method for dynamic contact angle measurement is the Wilhelmy balance technique, in which a liquid is supported on a motor-driven stage and the test specimen is suspended from a sensitive balance (commercial instrumentation is available from Cahn Instruments, Inc., 5225 Verona Road, Bldg. 1, Madison, WI 53711 and also from Krüss USA, 9305-B Monroe Road, Charlotte, NC 28270, USA). The stage moves upward to immerse the specimen, then continues slowly upward, forcing the wetting line to advance. When a specified immersion is reached the stage reverses direction and the wetting line recedes. The advancing and receding angles are obtained indirectly, by calculation from the apparent specimen weight, the dimensions of the specimen and the known value of the liquid surface tension. Limitations of the method are that the specimen must generally be thin, rectangular, and the same in energy on both front and back surfaces. Goniometry offers the advantages that the contact angle is directly measured rather than calculated, and the substrate can be irregular in shape, of nonuniform thickness, and one-sided.

In general, $\theta_a > \theta_r$, the advancing angle is greater than the receding, known as contact angle hysteresis [Equation 45]:

$$H = \theta_a - \theta_r \tag{45}$$

If a substrate is microscopically smooth and uniform in energy, the advancing and receding contact angles will be identical, i.e. showing no hysteresis. Rough substrates, in general, produce contact angle hysteresis. Other causes of hysteresis include nonuniformities in substrate energy, i.e. having high- and low-energy regions or spots. This may be due to substrate contamination or to variations in chemical composition. For example, Bouali et al.[68] found that the receding, but not the advancing, contact angle was sensitive to acid moieties 'buried' just under the surface of a latex polymer film. 'Pinning' of the advancing liquid front can occur on sharp edges of a rough substrate surface and also on low-energy spots, resulting in high values of the advancing angle θ_a, while pinning of the retreating liquid edge on high-energy spots lowers the receding angle θ_r. Such effects can produce significant hysteresis. Large values of H can be an indication that it will be difficult to achieve a meaningful 'equilibrium' contact angle for that substrate. Schwartz[69] and Brandon et al.[70] have numerically simulated these effects for a three-dimensional drop.

3.2 Choice of Test Liquids

The choice of liquids as surface energy probes using contact angle measurement is usually restricted to those which are not soluble in or absorbed by the test substrate, or capable of chemical reaction with it. Schneider[20] surveyed four solid substrates representing a range of surface energy, even deliberately contaminating substrates to form a so-called 'conditioning film'. Using a typical quartet of test liquids, water, formamide, α-bromonaphthalene, and diiodomethane, Schneider tested the GM, ES, and LWAB theories, and found the combination water–formamide–diiodomethane, together with the LWAB method, to give the best results. Schneider recommends against the use of α-bromonaphthalene as a dispersive test liquid because the π-electrons of the phenyl ring apparently impart measurable Lewis base character.

Holländer[71] demonstrated a potential source of significant error in the LWAB methodology due to a mathematical instability. The instability arises when the ratios of the acid and base interaction components γ^+ and γ^- for the test liquids are not different enough in magnitude. In the 'two contact angle' LWAB method, the equation sets that are used to calculate the acid–base terms of the solid are shown as Equations (46) and (47):[50]

$$\gamma^+ = \frac{a_1c_2 - a_2c_1}{b_1c_2 - b_2c_1} \tag{46}$$

$$\gamma^- = \frac{a_2b_1 - a_1b_2}{b_1c_2 - b_2c_1} \tag{47}$$

for which [Equations 48–53]:

$$a_1 = (1 + \cos\theta_1)\gamma_{L1}^{tot} - 2(\gamma_S^{LW}\gamma_{L1}^{LW})^{1/2} \tag{48}$$

$$a_2 = (1 + \cos\theta_2)\gamma_{L2}^{tot} - 2(\gamma_S^{LW}\gamma_{L2}^{LW})^{1/2} \tag{49}$$

$$b_1 = 2\sqrt{\gamma_{L1}{}^-} \tag{50}$$

$$b_2 = 2\sqrt{\gamma_{L2}{}^-} \tag{51}$$

$$c_1 = 2\sqrt{\gamma_{L1}{}^+} \tag{52}$$

$$c_2 = 2\sqrt{\gamma_{L2}{}^+} \tag{53}$$

Holländer showed that when the ratios of the liquid acid–base components for two polar test liquids are too close in magnitude, the denominators in Equations (46) and (47) become small, in which event relatively small errors (order of 2°) in measured contact angle can

result in very large excursions in the calculated γ^+ and γ^- values. Holländer defines $Q_1 = \gamma_{L1}{}^-/\gamma_{L1}{}^+$ and $Q_2 = \gamma_{L2}{}^-/\gamma_{L2}{}^+$, and recommends that $\Delta Q = |Q_1 - Q_2| \geq 3$, and preferably ≥ 15, although an optimal minimum ΔQ had yet to be determined. This presents a new criterion for selecting liquids to be used in conjunction with the LWAB method. By this criterion, formamide and glycerol ($\Delta Q = 2.8$) are not a good pair choice, but water–formamide or water–glycerol are ($\Delta Q \approx 100$).

3.3 Kinetic Effects and Environmental Control

As stated in section 3.1.2, once a relatively inviscid liquid drop is deposited on a solid substrate, it should reach a constant diameter, and hence a constant contact angle, within a short time. This is usually taken as the equilibrium advancing contact angle. The attainment of equilibrium may take seconds to minutes, depending on liquid viscosity. If the contact angle is measured as a function of time (of which some goniometric instruments are capable), one will see a more or less rapid decrease in θ corresponding to the initial spreading phase. The rate of spreading has little to do with the value of the final contact angle, and these initial data should be omitted from the data used to calculate the surface energy. The contact angle θ should then reach a fairly constant value for a partially wetting liquid, in the absence of evaporation, absorption, or other artifacts. For water or aqueous liquids, this may necessitate maintaining the relative humidity at a fairly high value in order to inhibit evaporation for the duration of data collection. For nonaqueous volatile liquids, it may similarly be necessary to maintain a solvent atmosphere around the drop. Goniometers with automated data collection and analysis make it easy to detect the need for environmental control. For example, if the drop width is constant but the contact angle is decreasing over time, the liquid is evaporating. Temperature control may also be an issue, although variations of a few degrees around room temperature will normally have only a slight effect on measured contact angles.

If experiments are performed in a humid environment or solvent vapor atmosphere, there is an issue of a possible effect on contact angle of adsorption of a film of liquid on the solid surface. Adsorbed vapor of the liquid on the substrate leads to higher contact angles, due to the so-called equilibrium film pressure π_e (section 2.2.1). In effect, the presence of a film of the selfsame liquid inhibits spreading of the liquid drop. A high-energy solid in the presence of high relative humidity may have an adsorbed film of water. If the drop liquid is highly compatible with water, as for example is glycerol, a water film may enhance spreading of the glycerol, producing lower contact angles than the thermodynamic equilibrium value.[71]

4 METHOD ENHANCEMENTS

4.1 Digital Data Acquisition and Analysis

Bashforth and Adams described the method of using shape and size of both sessile and pendant liquid drops to measure surface tension.[58] This method has certain advantages in respect that it can be used to measure low interfacial tensions more accurately than conventional methods. It is also well suited for studies of long-term behavior of liquid droplets on a surface. However, mainly because of the laborious nature of this method, it has not gained widespread use. Recent advances in computer hardware and software have led to revolutionary changes in instrumentation for contact angle measurements. Modern instrumentation has greatly increased productivity, automating routine tasks by capturing the image in digital form instead of on a photographic film or using the naked eye. At the same time the speed of processors has reduced the computational time required to calculate the drop shape to just a few seconds. However, these instruments are still affordable by the typical analytical laboratory. Automated methods are available in some form with modern contact angle goniometers.

The heart of an automated contact angle goniometer is the continuous frame grabber connected to a personal computer.[72] The deposition of a test liquid on a surface is accomplished by means of a computer-controlled syringe pump. The live image of the droplet deposited on the surface is acquired by a charge-coupled device (CCD) camera interfaced with a dedicated frame capture card which captures the video signal from the camera and converts it into a digital image. The sequence of digitized images of the droplet is saved for later analysis. Since most combinations of CCD cameras and frame grabbers can capture at the rate of at least 60 frames per second, dynamic processes such as adsorption of liquid by a porous substrate can be studied. More advanced instruments, such as the FTÅ200 (First Ten Ångstroms), offer the capability of triggering the acquisition of a live image at the instant a droplet is deposited on the surface.[72]

4.2 Automation of Data Collection and Analysis

Figure 7 shows the configuration of a typical automated contact angle goniometer system from First Ten Ångstroms. A sessile drop is deposited using a computer-controlled syringe pump. The data acquisition can be initiated either manually or by a signal from the fiber-optic sensor, which detects the moment at which the droplet detaches from the dispensing needle. The latter mode is especially useful for measuring simultaneously both the initial contact angle of the spreading droplet and the quasi-equilibrium contact angle of a sessile droplet.

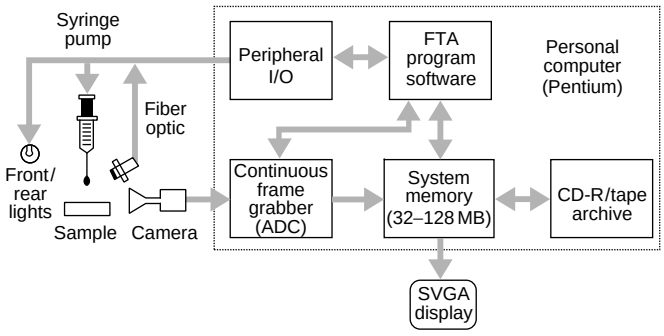

Figure 7 Schematic of configuration of contact angle measurement system used by First Ten Ångstroms. [Reprinted by permission of First Ten Ångstroms, Inc., Portsmouth, VA, Copyright 1996.]

It also allows one to measure dynamic contact angles on substrates that absorb, or are swollen by, the test liquid. Finally, it reduces operator strain and associated experimental errors. Advancing and receding contact angles can also be easily analyzed by pumping liquid in and out of a stationary droplet during the image acquisition. This arrangement typically produces more reliable results than the plate tilting method.[73]

Once the drop image or a sequence of images has been acquired, an image analysis algorithm is used to find the drop edge. Since most CCD cameras acquire images at relatively low resolution (approximately equal to that of a television picture), for the results to be sufficiently accurate, the drop edge in the digital image is determined at a subpixel resolution by special routines.[74,75] One method,[76] for example, detects the drop edge by selecting two pixels along the vertical or horizontal line passing through the drop baseline with intensity above and below the manually selected threshold (usually 50–60% of maximum intensity) and

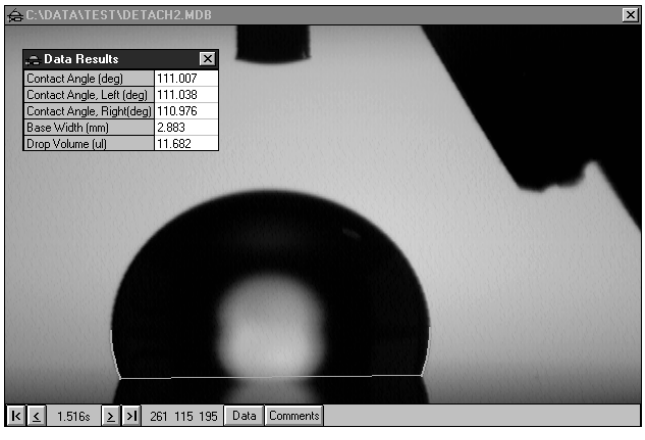

Figure 8 Automated sessile drop contact angle determination with the FTÅ200 goniometer–tensiometer system.

then calculating the exact location of the drop boundary by linear interpolation. This procedure is repeated to obtain the coordinates of the droplet surface, from which the drop shape and contact angles are then calculated. This can be done by fitting a polynomial to the drop edge, and obtaining the contact angle as the derivative of the polynomial at the drop baseline. Figure 8 shows a typical analyzed drop image from the FTÅ200 goniometer–tensiometer instrument. Alternatively, the Laplace–Young equation can be solved for the analytical shape of the drop by using, for example, the Bashforth and Adams technique, or by the method known as axisymmetric drop shape analysis-profile (ADSA-P), developed by Neumann's group.[67]

5 RECENT DEVELOPMENTS: CONTACT MECHANICS METHOD OF SURFACE ENERGY DETERMINATION (JOHNSON, KENDALL, ROBERTS METHOD)

As was discussed in the Introduction, the direct determination of surface energy of an elastomer in general is not possible because one cannot reversibly change the surface area of a solid. The contact angle methods described in this article generally work well, but in some cases they may produce inaccurate results. Such would be the case if the polarities of the test liquids are significantly different from that of a solid in question, or specific liquid–solid interactions exist. The use of temperature or molecular weight extrapolation methods could also become complicated if the elastomer undergoes a glass transition and/or melting processes within the range of extrapolation. These and other factors have led to an increase of interest in the area of direct determination of surface energy using the contact mechanics method. This method could be in principle called 'direct', since it is based on increasing or decreasing the interfacial area between two phases. However, instead of stretching or contracting the interface, as is possible with liquids, one adds or subtracts the amount of interface by changing the normal pressure that is applied to a curved surface of a solid. The phenomena that occur during such processes are described by contact mechanics, which studies the deformation of the interfacial area of solids under the load.[77] In particular, the theory developed by Johnson, Kendall, and Roberts[78] (JKR) derives a set of equations which relate the work of adhesion and area of interfacial contact between two spherical solid surfaces, where at least one of the materials is relatively soft. Since the work of adhesion is directly related to the surface and interfacial energy between the solids, JKR formalism provides a route to the direct measurement of surface energies of

polymers, provided that the elasticity requirement for at least one of the polymers is satisfied.

JKR measurements are relatively easy to implement experimentally. Often it is sufficient to observe an elastomer lens under a microscope and monitor the change in contact area upon application of controlled load by means of a spring connected to a precision micrometer screw.[79] A schematic of typical JKR apparatus is shown in Figure 9.

The goal of the JKR experiment is to determine the contact radius a between the surfaces with GM radius of curvature R (usually an elastic sphere and a flat substrate), as a function of normal pressure P. The experimental data are usually plotted as a^3 vs P and analyzed using the JKR equation [Equation 54]:

$$a^3 = \frac{R}{K}\left\{P + 3W\pi R + \sqrt{[6W\pi RP + (3W\pi R)^2]}\right\} \qquad (54)$$

where K is the mean elastic constant of the materials in contact and W is the work of adhesion. The contact radius vs pressure data could then be fitted to this equation by adjusting either both K and W, or the value of W obtained by using the known value of K. From the work of adhesion W, the surface energy of a material can be obtained, since $W = 2\gamma$ for contact between the two identical materials. For different materials, $W = \gamma_1 + \gamma_2 - \gamma_{12}$, where γ_1 and γ_2 are surface energies and γ_{12} is interfacial energy. Therefore, if the surface energy of one material is known, it is possible to obtain the unknown surface energy of the other material.

Several basic assumptions must be satisfied in order for JKR analysis to be valid.[79] First, deformations have to remain small compared with the dimensions of the sample. This requirement stems from the theoretical assumption that contact between two elastic semi-infinite bodies is considered. This assumption is difficult to satisfy in the usual JKR experiment, which utilizes a small 'elastomeric lens'.[80] Because of the small size of the lens, there is a discontinuity of the modulus between the lens and the upper support of the JKR apparatus. Deruelle et al.[79] pointed out this difficulty and offered a possible solution, which consists in the insertion of a sufficiently thick ribbon of the same elastomer between the lens and support. A second constraint assumes equilibrium conditions at each step of loading or unloading. Specific interactions developing in the contact area or finite velocity of the crack speed upon separating the surfaces could make the total energy required to separate the surfaces larger than the thermodynamic work of adhesion.[81] While hysteresis in the loading–unloading curves could be used to study the specific interactions and/or rearrangements of polymer chains in the contact area,[82] Deruelle et al.[79] also showed that it is important to account for crack speed effects in the analysis of unloading curves.

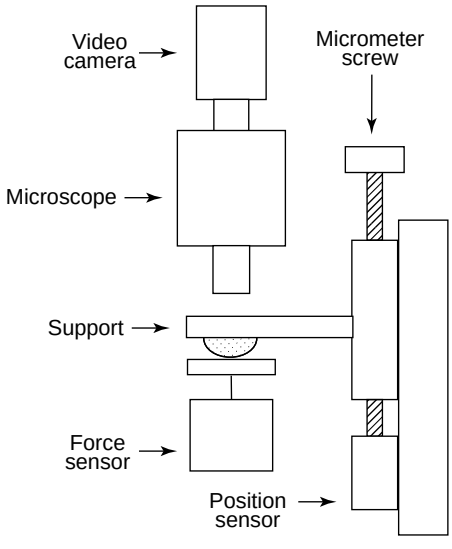

Figure 9 Schematic of JKR contact mechanics apparatus.

One has to keep in mind that JKR theory was originally developed for elastic solids, and therefore is not directly applicable to viscoelastic systems. However, very often chemically or physically cross-linked materials, which embody the majority of commercial elastomers and rubbers, exhibit a plateau in modulus at low deformation rate. For such systems, elastic JKR analysis can be applied to data collected under loads that correspond to this plateau region. For a system that is not at elastic equilibrium, Falsafi et al.[83] incorporated linear viscoelastic effects into the JKR formalism by replacing the elastic modulus with a viscoelastic memory function, which accounts for time and deformation. They have applied this formalism to measure the surface energy of model diblock poly(ethylene)–poly(ethylene–propylene) copolymers. The measured values of surface energies were close to the reported value for a surface-active PEP block, obtained from contact angle data. This indicates that, with proper analysis, reliable values for surface energy of viscoelastic materials could be obtained from contact mechanics experiments.

The range of polymeric materials that can be studied by the classical JKR method is limited by the fact that many polymers do not possess a sufficiently low modulus to produce a measurable change in contact area for small loads. Tirrell[84] found that for glassy or semicrystalline polymers such as polystyrene, PMMA, polyethylene or poly(ethylene terephthalate), the procedure developed by Chaudhury et al.[80,85–87] can be adapted for deposition of thin films of polymer on poly(dimethylsiloxane). In such a procedure, the typical JKR 'probe' is a composite consisting of a thin layer of the polymer of interest

Table 8 Surface energies of polymers[a] measured using contact mechanics compared with contact angle data analyzed using a Zisman plot. [Reprinted with permission from M. Tirrell, *Langmuir*, **12**, 4548–4551 (1996). Copyright 1996, American Chemical Society.]

PVCH	TPX	PE	PS	PMMA	Corona-treated PE	PET	PVP
Contact mechanics measurements							
28 ± 1	26 ± 2	33 ± 1.6	44 ± 2	53 ± 3	52 ± 1	61 ± 2	63 ± 4
Contact angle measurements							
29 ± 1	21.5 ± 1	32 ± 1.6	30 ± 1	40 ± 0.2	33.5 ± 1	38 ± 1	50 ± 2

[a] PVCH, poly(vinylcyclohexane); TPX, poly(4-methyl-1-pentene); PE, polyethylene; PS, polystyrene; PET, poly(ethylene terephthalate); PVP, poly(2-vinylpyridine).

deposited on a soft PDMS backing. One can also use free-standing polymer films for JKR measurements using a surface force apparatus (SFA).[88] A detailed description of these and other methodologies that can be employed in this type of experiment was given by Mangipudi et al.[89]

Table 8 shows a comparison of solid polymer surface energies determined using the 'composite' JKR probe, from Tirrell.[84] In the case of nonpolar surfaces these results agree well with the data obtained using traditional wetting experiments. Tirrell attributed the difference in surface energies of polar surfaces measured by the two methods to rearrangements at the polymer surfaces. Tirrell suggested that polar materials may bury high-energy functionalities in contact with air, but expose those functional groups in contact with another polymer surface. For nonpolar surfaces there would be no energetic benefit to such rearrangement.

Although the JKR technique lacks the simplicity of experimental set-up and analysis inherent in surface energy measurements based on Young's equations, it could serve as a viable alternative when conventional methods of measuring the surface energy of solid polymers are expected to produce inaccurate results.

ACKNOWLEDGMENTS

The authors thank ICI Paints North America for permission to publish this article.

LIST OF SYMBOLS

A	area of surface or interface
$\Delta G_{ij}^{\mathrm{adh}}$	free energy of adhesion between substances i and j
G	Gibbs free energy
$\Delta G_{ii}^{\mathrm{coh}}$	free energy of cohesion of substance i
H	dynamic contact angle hysteresis parameter, $= \theta_{\mathrm{a}} - \theta_{\mathrm{r}}$
K	mean elastic constant of materials in JKR equation [Equation 54]
M_{n}	number-average molecular weight
P	normal pressure of contact in JKR equation [Equation 54]
ΔP	pressure change across a curved interface
Q	ratio of γ^- to γ^+
R	mean radius of curvature of materials in contact, in JKR equation [Equation 54]
R_1, R_2	principal radii of curvature of curved liquid interface in Laplace equation
S	spreading coefficient
T	temperature
W	thermodynamic work; work of adhesion in JKR equation [Equation 54]
W_{ij}^{adh}	work of adhesion between substances i and j
W_{ii}^{coh}	work of cohesion of substance i
a	contact radius in JKR method
d	as superscript, denotes the dispersion interaction component
n	exponent in Macleod's equation [Equation 2]
p	as superscript, denotes the 'polar' interaction component
Φ	coefficient in geometric mean expression for work of adhesion between unlike substances
γ^{AB}	generalized acid–base interaction component of surface tension
γ^{LW}	Lifshitz–van der Waals interaction component of surface free energy
γ_{a}	surface tension of polymer amorphous phase
γ_{c}	Zisman's critical surface tension; surface free energy of polymer crystalline phase
γ_{∞}	surface free energy of polymer of infinite molecular weight
γ_i	surface free energy per unit area or surface tension of substance i

γ_i^d	dispersion interaction contribution to surface free energy of substance *i*
γ_i^p	polar interaction contribution to surface free energy of substance *i*
γ_{ij}	interfacial free energy per unit area between substances *i* and *j*
γ^H	hydrogen-bonding interaction component of surface tension
γ^K	Keesom (dipole–dipole) interaction component of surface tension
γ^L	London–van der Waals dispersive interaction component of surface tension
γ^D	Debye (dipole–induced dipole) interaction component of surface tension
γ_L, γ_{LV}	liquid surface tension or interfacial tension between liquid and its vapor
γ_S, γ_{SV}	solid surface free energy or interfacial tension between solid and vapor phases
γ_{SL}	solid–liquid interfacial free energy
γ_s^0	free energy of a solid in vacuum
γ^+	Lewis acid (electron pair acceptor) surface free energy parameter
γ^-	Lewis base (electron pair donor) surface free energy parameter
θ	contact angle
θ_a	advancing dynamic contact angle
θ_e	advancing equilibrium contact angle
θ_r	receding dynamic contact angle
π_e	equilibrium film pressure
ρ	polymer density
ρ_a	density of amorphous polymer phase
ρ_c	density of crystalline polymer phase

ABBREVIATIONS AND ACRONYMS

ADSA	Axisymmetric Drop Shape Analysis
ADSA-P	Axisymmetric Drop Shape Analysis-profile
CCD	Charge-coupled Device
DMSO	Dimethyl Sulfoxide
ES	Equation of State
GM	Geometric Mean
HEBA	4-Hydroxyethylbenzoic Acid
HM	Harmonic Mean
JKR	Johnson, Kendall, Roberts
LW	Lifshitz–van der Waals
LWAB	Lifshitz–van der Waals Acid–Base
PMMA	Poly(methyl Methacrylate)
PVC	Poly(vinyl Chloride)
SFA	Surface Force Apparatus

REFERENCES

1. S. Wu, 'Surface and Interfacial Tension of Polymer Melts', *J. Colloid Interface Sci.*, **31**(2), 153–161 (1969).
2. R.-J. Roe, 'Surface Tension of Polymer Liquids', *J. Phys. Chem.*, **72**(6), 2013–2017 (1968).
3. D.B. Macleod, 'Relation between Surface Tension and Density', *Trans. Faraday Soc.*, **19**, 38 (1923).
4. S. Wu, 'Interfacial and Surface Tensions of Polymers', *J. Macromol. Sci.*, **C10**(1), 1–73 (1974).
5. R.J. Roe, 'Surface Tension of Polymer Liquids', *J. Phys. Chem.*, **72**, 2013–2017 (1968).
6. S. Wu, *Polymer Interface and Adhesion*, Marcel Dekker, New York, 1982.
7. H. Schonhorn, 'Heterogeneous Nucleation of Polymer Melts on Surfaces. I. Influence of Substrates on Wettability', *Polym. Lett.*, **5**, 919–924 (1967).
8. H. Schonhorn, 'Heterogeneous Nucleation of Polymer Melts on High-energy Surfaces. II. Effect of Substrate on Morphology and Wettability', *Macromolecules*, **1**, 145–151 (1968).
9. H. Schonhorn, F.W. Ryan, 'Wettability of Polyethylene Single Crystal Aggregates', *J. Phys. Chem.*, **70**, 3811–3815 (1966).
10. D.G. LeGrand, G.L. Gaines, Jr, 'Molecular Weight Dependence of Polymer Surface Tension', *J. Colloid Interface Sci.*, **31**, 162–167 (1969).
11. D.G. LeGrand, G.L. Gaines, Jr, 'Surface Tension of Homologous Series of Liquids', *J. Colloid Interface Sci.*, **42**, 181–184 (1973).
12. C.J. van Oss, *Interfacial Forces in Aqueous Media*, Marcel Dekker, New York, 1994.
13. T. Young, 'An Essay on the Cohesion of Fluids', *Philos. Trans. R. Soc. London*, 65–87 (1805).
14. M.E. Schrader, 'Work of Adhesion of a Sessile Drop to a Clean Surface', *J. Colloid Interface Sci.*, **213**, 602–605 (1999).
15. D.H. Bangham, R.I. Razouk, 'Adsorption and the Wettability of Solid Surfaces', *Trans. Faraday Soc.*, **33**, 1459–1463 (1937).
16. B.V. Derjaguin, *Theory of Stability of Colloids and Thin Films*, Consultants Bureau–Plenum Press, New York, 1989.
17. C.J. van Oss, *Interfacial Forces in Aqueous Media*, Marcel Dekker, New York, 22, 1994.
18. A. Dupré, *Theorie Mechanique de la Chaleur*, Gauthier-Villars, Paris, 1869.
19. W.A. Cooper, W.A. Nuttall, 'The Theory of Wetting, and the Determination of the Wetting Power of Dipping and Spraying Fluids Containing a Soap Basis', *J. Agric. Sci.*, **7**, 219–239 (1915).
20. R.P. Schneider, 'Comparative Analysis of Thermodynamic Approaches and Diagnostic Liquids for Determination of Contact Angle-derived Physicochemical Parameters of Solids Coated with Conditioning Films: A

Practitioner's Perspective', *J. Adhes. Sci. Technol.*, **11**(1), 65–93 (1997).
21. A.W. Neumann, R.J. Good, C.F. Hope, M. Sejpal, 'An Equation-of-state Approach to Determine Surface Tensions of Low-energy Solids from Contact Angles', *J. Colloid Interface Sci.*, **49**, 291–304 (1974).
22. D. Li, A.W. Neumann, 'Contact Angles on Hydrophobic Solid Surface and their Interpretation', *J. Colloid Interface Sci.*, **148**, 190–200 (1992).
23. A.W. Neumann, J.K. Spelt (eds.), *Applied Surface Thermodynamics*, Marcel Dekker, New York, 1996.
24. D.Y. Kwok, 'The Usefulness of the Lifshitz–van der Waals/Acid–Base Approach for Surface Tension Components and Interfacial Tensions', *Colloids Surf. A*, **156**, 191–200 (1999).
25. C.J. van Oss, *Interfacial Forces in Aqueous Media*, Marcel Dekker, New York, Chapter IX, 1994.
26. T.G.M. van de Ven, P.G. Smith, R.G. Cox, S.G. Mason, 'The Surface Tension of Solids', *J. Colloid Interface Sci.*, **91**, 298–299 (1983).
27. R.E. Johnson, Jr, R.H. Dettre, 'An Evaluation of Neumann's "Surface Equation of State"', *Langmuir*, **5**, 293–295 (1989).
28. F.M. Fowkes, F.L. Riddle, W.E. Pastore, A.A. Weber, 'Interfacial Interactions between Self-associated Polar Liquids and Squalane used to Test Equations for Solid–Liquid Interfacial Interactions', *Colloids Surf.*, **43**, 367–387 (1990).
29. D.Y. Kwok, D. Li, A.W. Neumann, 'Fowkes' Surface Tension Component Approach Revisited', *Colloids Surf. A*, **89**, 181–191 (1994).
30. D.Y. Kwok, C.N.C. Lam, A. Li, A. Leung, R. Wu, E. Mok, A.W. Neumann, 'Measuring and Interpreting Contact Angles: A Complex Issue', *Colloids Surf. A*, **142**, 219–235 (1998).
31. D.K. Owens, R.C. Wendt, 'Estimation of the Surface Free Energy of Polymers', *J. Appl. Polym. Sci.*, **13**, 1741–1747 (1969).
32. S. Wu, 'Polar and Nonpolar Interactions in Adhesion', *J. Adhes.*, **5**, 39–55 (1973).
33. R.J. Good, 'Physical Significance of Parameters γ_c, γ_s and Φ, that Govern Spreading on Adsorbed Films', *SCI Monogr.*, **25**, 328–350 (1967).
34. L.A. Girifalco, R.J. Good, 'A Theory for the Estimation of Surface and Interfacial Energies. I. Derivation and Application to Interfacial Tension', *J. Phys. Chem.*, **61**, 904–909 (1957).
35. R.J. Good, L.A. Girifalco, 'A Theory for the Estimation of Surface and Interfacial Energies. III. Estimation of Surface Energies of Solids from Contact Angle Data', *J. Phys. Chem.*, **64**, 561–565 (1960).
36. F.M. Fowkes, 'Attractive Forces at Interfaces', in *Chemistry and Physics of Interfaces*, ed. S. Ross, American Chemical Society, Washington, DC, 1–12, 1965.
37. F.M. Fowkes, 'Attractive Forces at Interfaces', *Ind. Eng. Chem.*, **56**, 40–52 (1964).
38. F.M. Fowkes, M.A. Mostafa, 'Acid–Base Interactions in Polymer Adsorption', *Ind. Eng. Chem. Prod. Res. Dev.*, **17**, 3–7 (1978).
39. D.H. Kaelble, 'Dispersion-polar Surface Tension Properties of Organic Solids', *J. Adhes.*, **2**, 66–81 (1970).
40. J.H. Hildebrand, 'A Critique of the Theory of Solubility of Non-electrolytes', *Chem. Rev.*, **44**, 37–45 (1949).
41. R.J. Good, C.J. van Oss, 'The Modern Theory of Contact Angles and the Hydrogen Bond Components of Surface Energies', in *Modern Approaches to Wettability, Theory and Applications*, eds. M.E. Schrader, G.I. Loeb, Plenum Press, New York, 1–27, 1992.
42. C.J. van Oss, R.J. Good, 'Prediction of the Solubility of Polar Polymers by means of Interfacial Combining Rules', *Langmuir*, **8**, 2877–2879 (1992).
43. E.N. Dalal, 'Calculation of Solid Surface Tensions', *Langmuir*, **3**, 1009–1015 (1987).
44. C.K. Schoff, 'Wettability Phenomena in Coatings', in *Modern Approach to Wettability: Theory and Applications*, eds. M.E. Schrader, G. Loeb, Plenum Press, New York, 375–395, 1992.
45. M.H.V.C. Adão, B.J.V. Saramago, A.C. Fernandes, 'Estimation of the Surface Properties of Styrene-acrylonitrile Random Copolymers from Contact Angle Measurements', *J. Colloid Interface Sci.*, **217**, 94–106 (1999).
46. C.J. van Oss, M.K. Chaudhury, R.J. Good, 'Monopolar Surfaces', *Adv. Colloid Interface Sci.*, **28**, 35–64 (1987).
47. C.J. van Oss, M.K. Chaudhury, R.J. Good, 'Interfacial Lifshitz–van der Waals and Polar Interactions in Macroscopic Systems', *Chem. Rev.*, **88**, 927–941 (1988).
48. R.J. Good, M.K. Chaudhury, C.J. van Oss, 'Theory of Adhesive Forces Across Interfaces: 2. Interfacial Hydrogen Bonds as Acid–Base Phenomena and as Factors Enhancing Adhesion', in *Fundamentals of Adhesion*, ed. L.H. Lee, Plenum Press, New York, 153–172, 1991.
49. C.J. van Oss, *Interfacial Forces in Aqueous Media*, Marcel Dekker, New York, 16*ff*, 1994.
50. R.J. Good, 'Contact Angle, Wetting, and Adhesion: A Critical Review', in *Contact Angle, Wettability, and Adhesion*, ed. K.L. Mittal, VSP BV, Utrecht, 3–36, 1993.
51. L.-H. Lee, 'Correlation between Lewis Acid–Base Surface Interaction Components and Linear Solvation Energy Relationship Solvatochromic α and β Parameters', *Langmuir*, **12**, 1681–1687 (1996).
52. A.W. Neumann, D.R. Absolom, D.W. Francis, C.J. van Oss, 'Conversion Tables of Contact Angles to Surface Tensions', *Sep. Purif. Methods*, **9**, 69–163 (1980).
53. B. Janczuk, T. Bialopiotrowicz, 'The Total Surface Free Energy and the Contact Angle in the Case of Low Energetic Solids', *J. Colloid Interface Sci.*, **140**, 362–372 (1990).
54. P. Dalet, E. Papon, J.-J. Villenave, 'Surface Free Energy of Polymeric Materials: Relevancy of Conventional Contact Angle Analyses', *J. Adhes. Sci. Technol.*, **13**(8), 857–870 (1999).
55. N.T. Correia, J.J.M. Ramos, B.J.V. Saramago, J.C.G. Calado, 'Estimation of the Surface Tension of a Solid:

Application to a Liquid Crystalline Polymer', *J. Colloid Interface Sci.*, **189**, 361–369 (1997).
56. W.A. Zisman, 'Relation of the Equilibrium Contact Angle to Liquid and Solid Constitution', *Contact Angle, Wettability, and Adhesion*, Advances in Chemistry Series No. 43, American Chemical Society, Washington, DC, 1–51, 1964.
57. S. Ross, I.D. Morrison, *Colloidal Systems and Interfaces*, John Wiley & Sons, New York, 1988.
58. F. Bashforth, C. Adams, *An Attempt to Test the Theories of Capillary Action*, Cambridge University Press, Cambridge, 1883.
59. A.W. Adamson, A.P. Gast, *Physical Chemistry of Surfaces*, 6th edition, John Wiley & Sons, New York, 6, 1997.
60. A. Passerone, E. Ricci, 'High Temperature Tensiometry', in *Drops and Bubbles in Interfacial Research*, eds. D. Möbius, R. Miller, Elsevier, Amsterdam, 475–524, 1998.
61. F.K. Hansen, G. Rødsrud, 'Surface Tension by Pendant Drop', *J. Colloid Interface Sci.*, **141**, 1–9 (1990).
62. G. Wolansky, A. Marmur, 'Apparent Contact Angles on Rough Surfaces: The Wenzel Equation Revisited', *Colloids Surf. A*, **156**, 381–388 (1999).
63. E.L. Decker, B. Frank, Y. Suo, S. Garoff, 'Physics of Contact Angle Measurement', *Colloids Surf. A*, **156**, 177–189 (1999).
64. A. Marmur, 'Equilibrium Contact Angles: Theory and Experiment', *Colloids Surf. A*, **116**, 55–61 (1996).
65. D.Y. Kwok, A.W. Neumann, 'Contact Angles and Surface Energetics', *Progr. Colloid Polym. Sci.*, **109**, 170–184 (1998).
66. C.O. Timmons, W.A. Zisman, 'The Effect of Liquid Structure on Contact Angle Hysteresis', *J. Colloid Interface Sci.*, **22**, 165–171 (1966).
67. Y. Rotenberg, L. Boruvka, A.W. Neumann, 'Determination of Surface Tension and Contact Angle from the Shapes of Axisymmetric Fluid Interfaces', *J. Colloid Interface Sci.*, **93**, 169–183 (1983).
68. B. Bouali, F. Ganachaud, J.-P. Chapel, C. Pichot, P. Lanteri, 'Acid–Base Approach to Latex Particles Containing Specific Groups Based on Wettability Measurements', *J. Colloid Interface Sci.*, **208**, 81–89 (1998).
69. L.W. Schwartz, 'Hysteretic Effects in Droplet Motions on Heterogeneous Substrates Direct Numerical Simulation', *Langmuir*, **14**, 3440–3453 (1998).
70. S. Brandon, A. Wachs, A. Marmur, 'Simulated Contact Angle Hysteresis of a Three-dimensional Drop on a Chemically Heterogeneous Surface: A Numerical Example', *J. Colloid Interface Sci.*, **191**, 110–116 (1997).
71. A. Holländer, 'On the Selection of Test Liquids for the Evaluation of Acid–Base Properties of Solid Surfaces by Contact Angle Goniometry', *J. Colloid Interface Sci.*, **169**, 493–496 (1995).
72. R.P. Woodward, 'A New Dynamic Contact Angle System', First Ten Ångstroms, Portsmouth, VA, 1995.
73. L.M. Lander, L.M. Siewierski, W.J. Brittain, E.A. Vogler, 'A Systematic Comparison of Contact Angle Methods', *Langmuir*, **9**, 2237–2239 (1993).
74. J. Noordmans, H.J. Busscher, 'Influence of Droplet Volume and Contact Angle on Liquid Surface Tensions by Axisymmetric Drop Shape Analysis-profile', *Colloids Surf.*, **58**, 239 (1991).
75. P. Cheng, D. Li, L. Boruvka, Y. Rotenberg, A.W. Neumann, 'Automation of Axisymmetric Drop Shape Analysis for Measurements of Interfacial Tensions and Contact Angles', *Colloids Surf.*, **43**, 151–167 (1990).
76. F.K. Hansen, 'Surface Tension by Image Analysis: Fast and Automatic Measurements of Pendant and Sessile Drops and Bubbles', *J. Colloid Interface Sci.*, **160**, 209–217 (1993).
77. K.L. Johnson, *Contact Mechanics*, Cambridge University Press, Cambridge, 1985.
78. K.L. Johnson, K. Kendall, A.D. Roberts, 'Surface Energy and the Contact of Elastic Solids', *Proc. R. Soc. London, Ser. A*, **324**, 301–313 (1971).
79. M. Deruelle, H. Hervet, G. Jandeau, L.J. Léger, 'Some Remarks on JKR Experiments', *Adhes. Sci. Technol.*, **12**(2), 225–247 (1998).
80. M.K. Chaudhury, G.M. Whitesides, 'Direct Measurement of Interfacial Interactions between Semispherical Lenses and Flat Sheets of Poly(dimethylsiloxane) and their Chemical Derivatives', *Langmuir*, **7**, 1013–1025 (1991).
81. D. Maugis, M.J. Barquins, 'Fracture Mechanics and the Adherence of Viscoelastic Bodies', *J. Phys. D: Appl. Phys.*, **11**, 1989–2023 (1978).
82. M. Deruelle, L. Léger, M. Tirrell, 'Adhesion at the Solid–Elastomer Interface: Influence of the Interfacial Chains', *Macromolecules*, **28**, 7419–7428 (1995).
83. A. Falsafi, P. Deprez, F.S. Bates, M. Tirrell, 'Direct Measurement of Adhesion between Viscoelastic Polymers: A Contact Mechanical Approach', *J. Rheol.*, **41**(6), 1349–1364 (1997).
84. M. Tirrell, 'Measurement of Interfacial Energy of Solid Polymer Surfaces', *Langmuir*, **12**, 4548–4551 (1996).
85. M.K. Chaudhury, M.J. Owen, 'Adhesion Hysteresis and Friction', *J. Phys. Chem.*, **9**, 29–31 (1993).
86. M.K. Chaudhury, M.J. Owen, 'Correlation between Adhesion Hysteresis and Phase State of Monolayer Films', *J. Phys. Chem.*, **97**, 5277–5726 (1993).
87. M.K. Chaudhury, 'Surface Free Energies of Alkyl Siloxane Monolayers Supported on Elastomeric Polydimethylsiloxanes', *J. Adhes. Sci. Technol.*, **6**, 669–675 (1993).
88. W. Merrill, A.V. Pocius, B.V. Thakkar, M. Tirrell, 'Direct Measurement of Molecular Level Adhesion Forces between Biaxially Oriented Solid Polymer Films', *Langmuir*, **7**, 1975–1980 (1991).
89. V. Mangipudi, E. Huang, M. Tirrell, A.V. Pocius, 'Measurement of Interfacial Adhesion between Glassy Polymers using the JKR Method', *Macromol. Symp.*, **102**, 131–143 (1996).
90. W.D. Harkins, *Chem. Revs.*, **29**, 335 (1941).

Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation

Benjamin Monrabal
Polymer ChAR, Valencia Technology Park, Spain

1 Introduction **1**
1.1 Significance of the Chemical Composition Distribution in Polyolefins 1
1.2 Review of the Analytical Techniques for the Chemical Composition Distribution Analysis of Polyolefins 2
1.3 Theoretical Background 3
2 Temperature Rising Elution Fractionation **4**
2.1 Separation Mechanism 4
2.2 Experimental Set-up in Analytical Temperature Rising Elution Fractionation 5
2.3 Operating Conditions 5
2.4 Solvents 5
2.5 Sample Size 6
2.6 Column 6
2.7 Crystallization Rates 6
2.8 Dissolution and Elution Flow Rate 7
2.9 Temperature Rising Elution Fractionation Automation 7
3 Crystallization Analysis Fractionation **8**
3.1 Separation Mechanism 8
3.2 Experimental Set-up 8
3.3 Solvents 9
3.4 Sample Size 9
3.5 Crystallization Rate 9
4 Comparison of Temperature Rising Elution Fractionation and Crystallization Analysis Fractionation **10**
5 Calculations **10**
6 Calibration **11**
7 Molecular Weight–Composition Dependence **12**
8 Preparative Fractionation **12**
9 Automated Cross-fractionation Apparatus **13**
10 Applications **14**
10.1 Linear Low-density Polyethylene 14
10.2 Low-density Polyethylene and Functional Olefinic Copolymers 15
10.3 High-density Polyethylene 16
10.4 Polypropylene and Copolymers 16
11 Conclusions **18**
Abbreviations and Acronyms **18**
Related Articles **18**
References **18**

Temperature rising elution fractionation (TREF) and crystallization analysis fractionation (CRYSTAF) are separation techniques for the characterization of semicrystalline polymers in terms of chemical composition. TREF has become a powerful tool for preparative polymer fractionation and for the analysis of the chemical composition distribution (CCD) in polyolefins. CRYSTAF, a more recent technique and with a simple hardware design, provides a faster approach for the analysis of the CCD. Both TREF and CRYSTAF separate according to the crystallizability of the polymer molecules, which is influenced by tacticity or the comonomer incorporation into the chain. TREF and CRYSTAF are becoming specially relevant in the characterization of the new and complex polyolefins made with combinations of homogeneous and heterogeneous catalysts.

1 INTRODUCTION

1.1 Significance of the Chemical Composition Distribution in Polyolefins

A polyolefin homopolymer structure such as that of high-density polyethylene (HDPE) is essentially defined by its molecular weight distribution. Most polyolefin products, however, are produced with incorporation of comonomers to reduce crystallinity and extend the range of product performance. To define these copolymers, a chemical composition or comonomer incorporation parameter is required in addition to the molecular weight distribution. This is the case for linear low-density polyethylene (LLDPE), where the incorporation of comonomers such as butene, pentene, hexene or octene results in short-chain branches that influence the crystallinity and final density of the resin. The incorporation of the comonomer into the chains has been shown, however, not to be uniformly distributed in all the molecules of an LLDPE resin. In fact, due to size differences between comonomers and ethylene, and to the multiple-site-type nature of Ziegler (coordination chemistry)-type catalysts used in the manufacturing of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

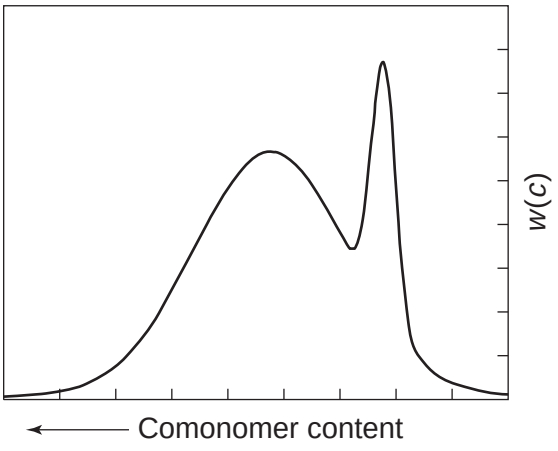

Figure 1 CCD of an LLDPE resin.

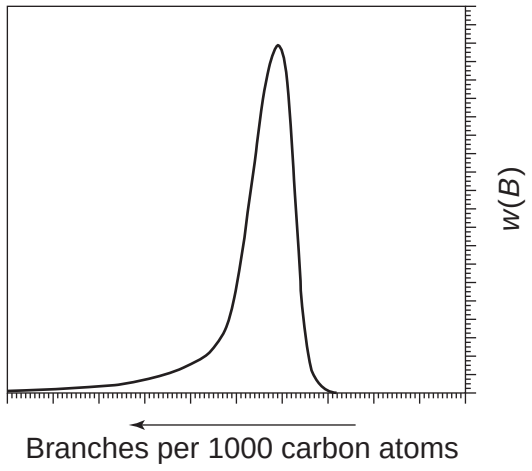

Figure 2 CCD of an LDPE resin.

LLDPE, the CCDs of these resins are quite peculiar, showing a bimodality as depicted in Figure 1.

To define an LLDPE resin structure in terms of composition it is necessary to provide the full CCD, sometimes referred to as the short-chain branching distribution (SCBD), rather than the average composition parameter. The structures of all Ziegler-type polyolefin copolymer resins, from LLDPE up to HDPE, can only be defined uniquely by their CCDs and molecular weight distributions.

The high-pressure polymerization processes also intend to introduce branches through chain transfer reaction or through the addition of selected comonomers. This is the case for low-density polyethylene (LDPE), where the CCD, although more uniform than that of LLDPE, is still not homogeneous, as shown in Figure 2. The reaction of ethylene with functional monomers to produce ethylene vinyl acetate and ethylene acrylic acid also requires the determination of CCD for branches and functional groups.

A special case is that of polypropylene resins, which as a homopolymer can be produced in various configuration forms (Figure 3) that result in crystallinity changes. The addition of small amounts of ethylene will also influence the regularity of the chain. In such cases of homopolymer or propylene–ethylene copolymers, we can refer to the CCD to describe the tacticity and composition changes.

The introduction of metallocene chemistry and other single-site catalysts into polyolefin manufacturing in recent years has provided the possibility of producing homogeneous products in terms of composition and molecular weight, and it would appear that such a simple CCD structure as shown in Figure 4 would make the measurement of CCD unnecessary in future polyolefins.

This in fact has not been the case because the new catalysts are being used to design complex multimodal CCD resins through multireactor–multicatalyst technologies

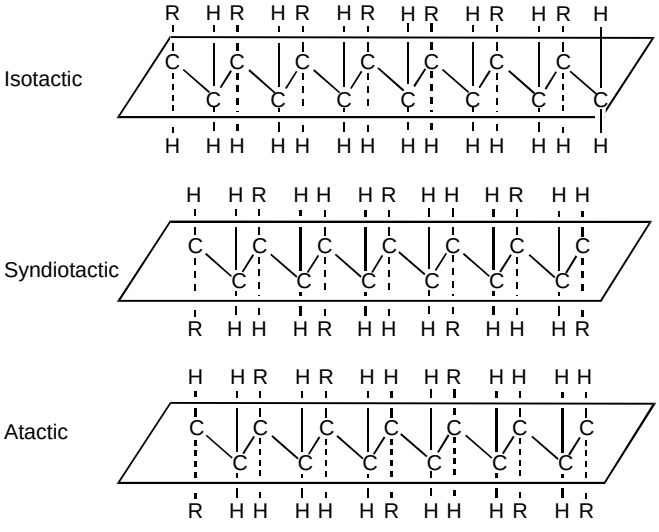

Figure 3 Polypropylene configuration forms.

with structures optimized for specific applications. As will be discussed in the following sections, the CCD of polyolefins is becoming the most important structural parameter to define a resin and it is certainly the most discriminating distribution in complex polyolefins.

1.2 Review of the Analytical Techniques for the Chemical Composition Distribution Analysis of Polyolefins

The structural characterization of polyolefins with broad CCD does necessarily require a fractionation process and, because the incorporation of branches, functional groups or the presence of less-regular configuration forms will all influence the crystallinity, it seems reasonable that potential analytical techniques will require a fractionation according to crystallizability.

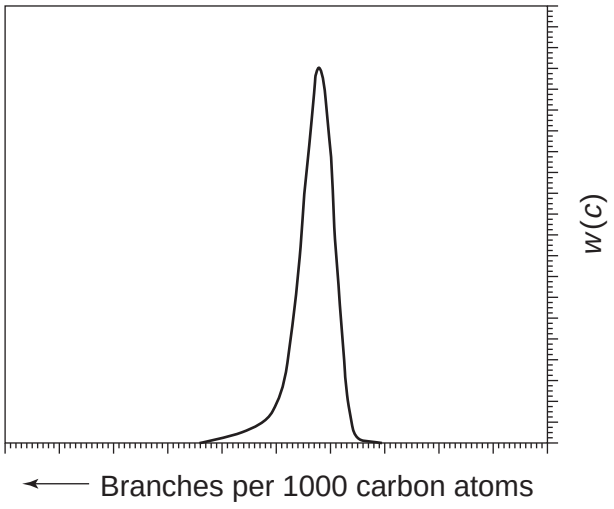

Figure 4 CCD of a metallocene-type resin.

The most comprehensive analytical approach has been TREF, implemented in the polyolefin characterization world by Wild et al. in the late 1970s,[1–3] which led to the understanding of the LLDPE structure in relation to the multiple sites in a Ziegler-type catalyst. TREF analysis resembles a liquid chromatograph separation with a column, an eluent and a detector or collecting device, as will be described in section 2. In TREF, however, the separation requires two temperature cycles (crystallization and dissolution), which makes the analysis process rather slow and requires complex hardware.

Given the long analysis process, the apparatus complexity and the unavailability of commercial TREF apparatus in the past years, many authors[4–9] have made use of the more available differential scanning calorimetry (DSC) technique to obtain qualitative data or parameters that could correlate with the CCD.[4,9,10] It should be clear, however, that DSC, although very powerful in other areas, does not provide the ideal environment for crystallization and does not result in quantitative mass measurements (heat flow instead of concentration detector).

Most significant methods, or parameters, using DSC are: stepwise isothermal segregation,[4] solvated thermal analysis fractionation,[11] DSC index,[10] step crystallization,[9] successive self-nucleation/annealing,[5] and fractional DSC.[12] Comparisons of DSC and TREF have also been reported.[6]

In general, solution crystallization (with higher chain mobility) would provide better resolution[7,13] and less co-crystallization effects than crystallizing in the melt, but at the expense of more tedious experimental work.

The value of the DSC results could be improved by proper correction of the detector response.[14] In DSC, the detection is based on the measurement of heat flow, which is more sensitive to fractions of high crystallinity (large heat of melt). Therefore, the ordinate results expressed in dQ/dT, in a given polymer family, would overemphasize the curve as moving toward the more crystalline fractions.

In spite of possible correction for the nonlinear detector response, the signal-to-noise ratio will decrease with lowering of the crystallinity of the material (a disadvantage that concentration detectors used in TREF or CRYSTAF do not have), and it would not provide a response on the important soluble or noncrystallizable fraction in polyolefins.

When dealing with block copolymers, the calorimetric methods can provide some information not accessible by TREF, where the dominating separation mechanism would be the most crystalline part of the block copolymer. A similar situation will be found with random copolymers if a significant intramolecular heterogeneity exists, as described by Mathot.[15] TREF curves would be narrower than DSC curves, because only the longer sequences in a chain would affect the elution temperatures. Meanwhile, the heat of melt in DSC would not be related to a separation process.

More recently, a new technique referred to as CRYSTAF has become available, which compares very well with TREF but is capable of doing the analysis in a shorter time with a single temperature cycle.[16] In CRYSTAF the samples are not fractionated in a column but in a stirred vessel and only a temperature cycle is required (crystallization), thus speeding up the analysis process and simplifying the hardware requirements.

The three techniques mentioned above share the same fundamental separation mechanisms that are discussed in the following section.

1.3 Theoretical Background

The principles of polymer fractionation by solubility or crystallization in solution have been reviewed extensively on the basis of Flory–Huggings statistical thermodynamic treatment,[17,18] which accounts for melting point depression by the presence of solvents and is expressed according to Equation (1):

$$\frac{1}{T_m} - \frac{1}{T_m^0} = \frac{R}{\Delta H_u}\frac{V_u}{V_1}(v_1 - \chi_1 v_1^2) \qquad (1)$$

where T_m^0 is the melting temperature of the pure polymer, T_m is the equilibrium melting temperature of the polymer–diluent mixtures, ΔH_u is the heat of fusion per polymer repeating unit, V_u and V_1 are the molar volumes of the polymer repeating unit and diluent, respectively, v_1 is the volume fraction of the diluent, and χ_1 is the Flory–Huggings thermodynamic interaction parameter.

Although most effort has been devoted to fractionation of homopolymers in terms of molecular weight, a few

reviews exist in the field of copolymer fractionation.[19–21] For random copolymers the classical Flory equation[22] applies, as shown in Equation (2):

$$\frac{1}{T_m} - \frac{1}{T_m^0} = -\frac{R}{\Delta H_u} \ln(p) \quad (2)$$

where p is the molar fraction of the crystallizing unit. Flory[17] has shown that Equation (1) reduces to the same form as Equation (2) and noncrystallizing comonomer units, diluents and polymer end-groups all have an equivalent effect on melting point depression when the concentration of each is low and they do not enter into the crystal lattice.

Equation (2) can be simplified by placing $p = (1 - N_2)$, where N_2 is the molar fraction of comonomer incorporated (noncrystallizing unit); for low values of N_2, $\ln(1 - N_2) \approx -N_2$, and hence Equation (3) is formed:

$$\frac{1}{T_m} - \frac{1}{T_m^0} \approx \frac{R}{\Delta H_u} N_2 \quad (3)$$

Further simplification is obtained by assimilating $T_m T_m^0 \approx (T_m^0)^2$ and assuming ΔH_u to be constant in the crystallization temperature range; hence, Equation (3) is reduced to Equation (4):

$$T_m \approx T_m^0 - \frac{R(T_m^0)^2}{\Delta H_u} N_2 \quad (4)$$

where the presence of solvent, when crystallizing in solution, is just an additional shift factor; therefore, a linear dependence of crystallization temperature T_m with the amount of comonomer incorporated (N_2) is achieved.

The validity of Equations (1) and (2) has been widely debated;[18,21,23,24] however, because a liquid-crystal phase transition is strongly governed by kinetic factors, the fractionation results are mainly influenced by the experimental procedure[25,26] that has deserved the most attention of researchers devoted to improving the fractionation techniques.[1,26] In experimental practice, a straight-line correlation between temperature and comonomer composition has been obtained by various authors with TREF,[2,27] DSC[28] and CRYSTAF.[16] These correlations are practically independent of molecular weight. The solution crystallization of polyethylene was shown to be independent of molecular weight at M_w above 15 000;[29] experiments in TREF (2) show that when considering the end-groups as noncrystallizing defects, fractionation is independent of molecular weight down to $M_w = 1000$. Similar results have been found in solution[4] and in the melt[12] by DSC.

The importance of co-crystallization in polyethylene has been widely investigated by Alamo et al.[31] Co-crystallization will always be present to a certain degree when crystallizing a polydispersed resin, and specially when crystallizing it in the melt[32] or in concentrated solutions. At the low concentrations used in TREF or CRYSTAF, the co-crystallization effects have been found to be of little significance.[16,33,34] We can conclude that, by using the proper experimental conditions and carrying out the crystallization in dilute solutions, the segregation of crystals will occur on the basis of the comonomer incorporated (assuming intramolecular uniformity).

2 TEMPERATURE RISING ELUTION FRACTIONATION

Fractionation of polyethylene according to composition was first described by Desreux and Spiegels[35] in 1950 by using an extraction technique with a single solvent at increasing temperatures. This was used with success by Hawkins and Smith[36] and Shirayama et al.,[37] who named the technique TREF, but it has been the work of Wild et al.[1–3] in the late 1970s with the development of analytical TREF that established the technique as a standard in the polyolefin industry. Various reviews have been published recently by Wild,[38] Glöckner,[34] Monrabal,[14] Fonseca and Harrison,[39] and Soares and Hamielec.[40,41]

2.1 Separation Mechanism

In TREF the sample is first dissolved in a proper solvent at high temperature and the solution is then introduced into a column containing a support; this is followed by a crystallization step at a slow cooling rate during which polymer fractionation occurs by deposition of layers of decreasing crystallinity, or increasing branch content, as temperature goes down; fractionation takes place within this cycle, which is usually carried out at very low crystallization rates, taking 1–3 days to cool down to room temperature.

Although at this stage the polymer is already segregated in layers or crystalline structures of different composition, the TREF technique still requires a second temperature cycle to quantify or collect these fractions. This is achieved by pumping new solvent while the temperature is being increased. The eluent dissolves fractions of increasing crystallinity, or decreasing branch content, as the temperature rises. These fractions are collected (preparative TREF) or the concentration is monitored with an infrared (IR) detector (analytical TREF) to generate the CCD curve. The name ''TREF'' derives from this second temperature cycle.

2.2 Experimental Set-up in Analytical Temperature Rising Elution Fractionation

A classical TREF apparatus is essentially an HPLC (high-performance liquid chromatography) system with a special oven for performing the crystallization and elution temperature ramps. Typically the samples are dissolved externally and the solution is poured at high temperature into a column filled with a support. The column is capped and introduced into an oil bath to be cooled down very slowly (around $2\,^{\circ}C\,h^{-1}$) (Figure 5a). As this is the slowest process, one can load various columns with different samples and cool them down all at once.

At the end of the crystallization cycle the loaded column is introduced to a programmable oven and connected to the HPLC pump and detector (usually an IR detector measuring the C–H absorption at around 3.5 μm wavelength) (Figure 5b). The pump operation is initiated to recover first the material that did not precipitate, and then with temperature rising (around $40\,^{\circ}C\,h^{-1}$) to monitor the concentration of the dissolved fractions of increasing crystallinity.

The elution curve resembles a chromatogram, with a small peak at the beginning (typically obtained at isothermal elution) corresponding to the fraction that did not crystallize at the lowest crystallization temperature chosen; this is followed by continuous elution of the fractions of increasing crystallinity as temperature rises (Figure 6).

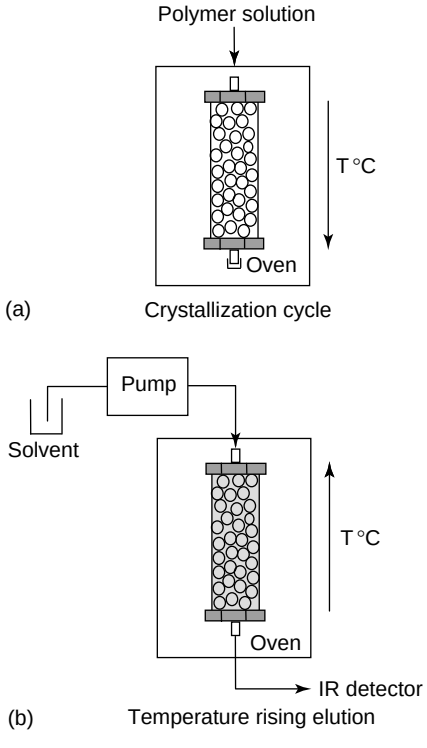

Figure 5 Schematic diagram of a TREF apparatus.

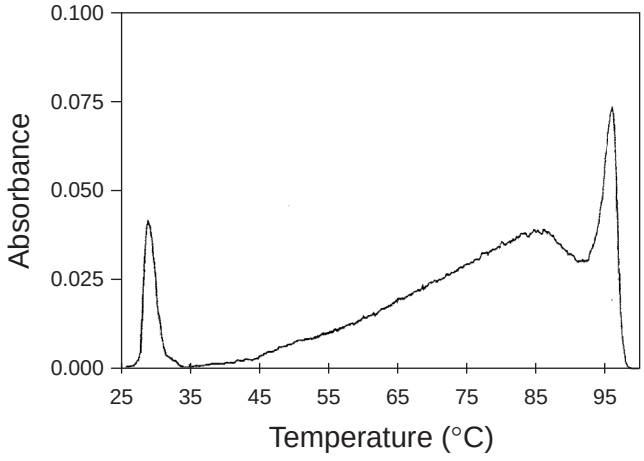

Figure 6 TREF raw data of an LLDPE resin.

2.3 Operating Conditions

TREF has not become commercially available until the early 1990s and most TREF users have developed their own instrumentation; it is not a surprise that a broad range of operating conditions exist in the practice.

Although TREF is not a chromatographic technique, some of the chromatographic theory applies, specially in relation to peak broadening by excessive death volume in the column and lines to detector, and caution should be taken to use a balanced combination of column dimensions, sample size, heating, and flow rate. Compilations of operating conditions discussed in the literature have been presented recently.[39,41]

2.4 Solvents

In analytical TREF the selection of solvents is limited to those transparent in the IR region at the measuring wavelength (around 3.5 μm) because the IR detector is used in virtually all TREF equipment owing to its low dependence on temperature fluctuations and its good sensitivity toward aliphatic C–H-containing polymers. The following solvents have been used: 1,2,4-trichlorobenzene, *o*-dichlorobenzene, perchloroethylene and α-chloronaphthalene. The addition of 0.1% of an antioxidant is recommended to prevent polymer degradation.

The solvent used the most is 1,2,4-trichlorobenzene, with the only inconvenience of the high melting point (17 °C), which prevents it being used at subambient crystallization temperatures. When it is necessary to crystallize the polymer at lower temperatures, *o*-dichlorobenzene and perchloroethylene are recommended.

The solvent does not influence the separation mechanism in the TREF analysis but elution temperatures will

be shifted depending on the solvent power, as discussed by Glöckner;[34] therefore, comparison of TREF analysis obtained with different solvents cannot be made in dissolution temperature units, but results will be comparable when calibrated into the number of branches per 1000 carbon atoms or any other polymer composition units.

Proper safety measurements should be taken in handling these solvents of high toxicity, specially perchloroethylene that has a low boiling point.

2.5 Sample Size

Solution concentrations of 0.1–1% are typical and the amount of polymer introduced in the column is usually in the range 10–100 mg. Sample size, column dimensions, elution flow rate and elution temperature rate are interconnected and all need to be optimized for a given TREF set-up. In general, the lowest concentration (sample size) possible should be used in order to reduce co-crystallization, entrapment and entanglement effects.

2.6 Column

Stainless-steel columns of 6–9 mm o.d. and lengths of 10–15 cm are most common in analytical TREF. The columns are filled with a support, typically diatomateous earth (Chromosorb P 60/80), although glass beads and s.s. shots are used as well.

To avoid the safety implications of filling the column with a hot solution, the crystallization process can be carried out in a separate vessel, with or without a support; the polymer and packing are mixed subsequently and added to the column for the elution step. It has been claimed that the absence of a support during the crystallization results in a resolution improvement.[38,42] The adsorption potential of a support on any functional copolymer that would interfere in the separation mechanism should also be considered.[43]

2.7 Crystallization Rates

The crystallization process is the most important step in the TREF analysis; it is when fractions are segregated

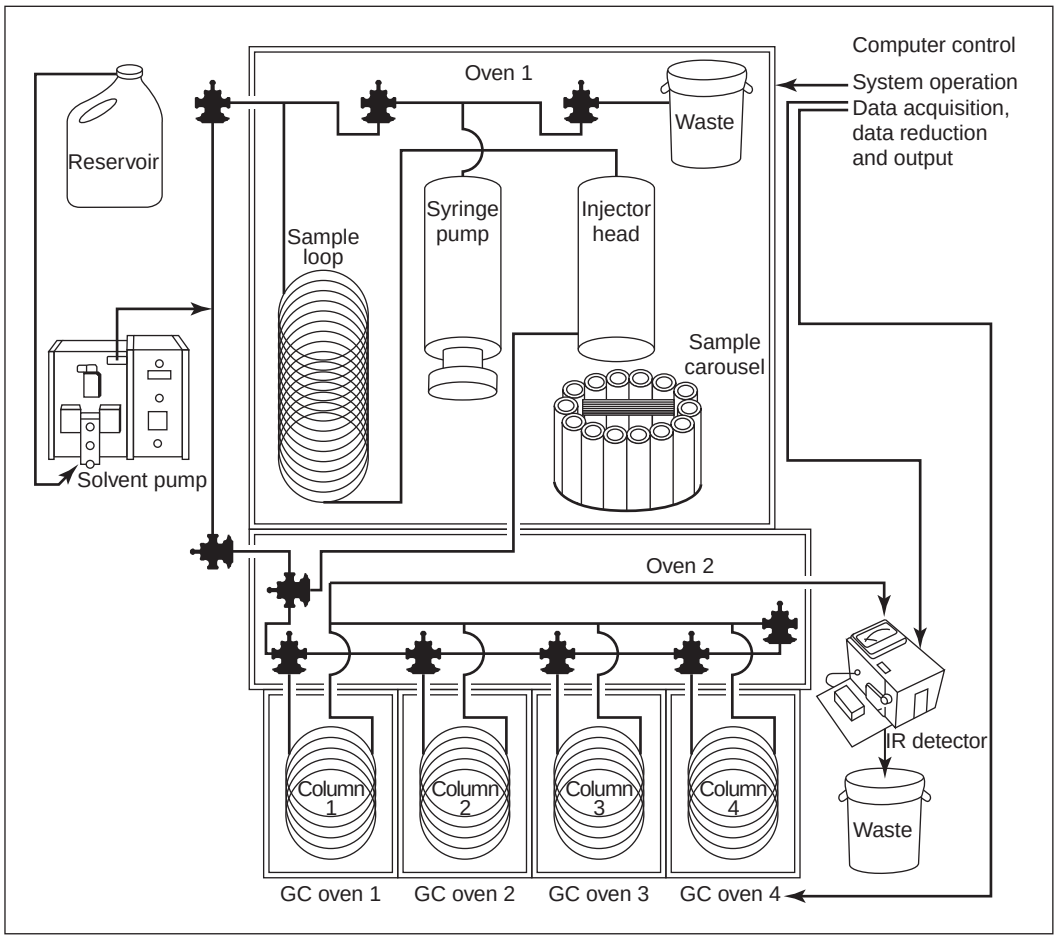

Figure 7 Schematic description of an automated analytical TREF apparatus. GC, gas chromatography. (Reproduced by permission from Hazlitt.[44])

according to their crystallizability and it is usually carried out at a very low cooling rate (typically between 1 and 3 °C h^{-1}). It has been indicated that rates equal to or lower than 2 °C h^{-1} are necessary for good resolution and to minimize co-crystallization and molecular weight interferences.[3]

2.8 Dissolution and Elution Flow Rate

The temperature rising elution step, which gives the name to the technique, is in fact the less important one from a fractionation point of view. Typically heating rates of 10–50 °C h^{-1} are used, but most important is to relate the heating to the flow rate; a high flow rate with a slow heating rate would elute the overall polymer in a large solvent volume and therefore, with a reduced signal-to-noise ratio, a low flow rate and fast heating rate, may result in too concentrated a solution going through the column, which may result in plugging. Typically flow rates of 0.5–4 mL min^{-1} are used, depending on the temperature rising rate, and optimized for column dimensions and sample size.

2.9 Temperature Rising Elution Fractionation Automation

Automation of a TREF apparatus is of significant importance to minimize solvent handling and to reduce manpower involvement. Automation has been associated also with optimized operating conditions to reduce analysis time (on occasions to sacrifice part of the resolution in favor of a larger sample throughput). Automation, however, may result in complex constructions[27] that may demand a high maintenance effort.

The most important steps to automate are: sample preparation (dissolution and column filling); transfer of the column from crystallization to the heating oven; and analysis of multiple samples.

In past years there have been various approaches described to automate the TREF apparatus and speed up the analysis.[44,45] The most ambitious is that of Hazlitt et al.,[44] who reported an automated TREF apparatus that can analyze eight samples per day. It has four columns in independent ovens and column loading is done automatically from vials in a sample carousel (Figure 7).

A new and fully automated TREF apparatus has been introduced recently[46] in which most effort is being done in the sample preparation step, as shown in Figure 8. Five different samples can be introduced at once in different stirred vessels inside a gas chromatography oven where a small column (10 cm and 1/8-in. o.d.) filled with a support is connected to valve B as shown in Figure 8.

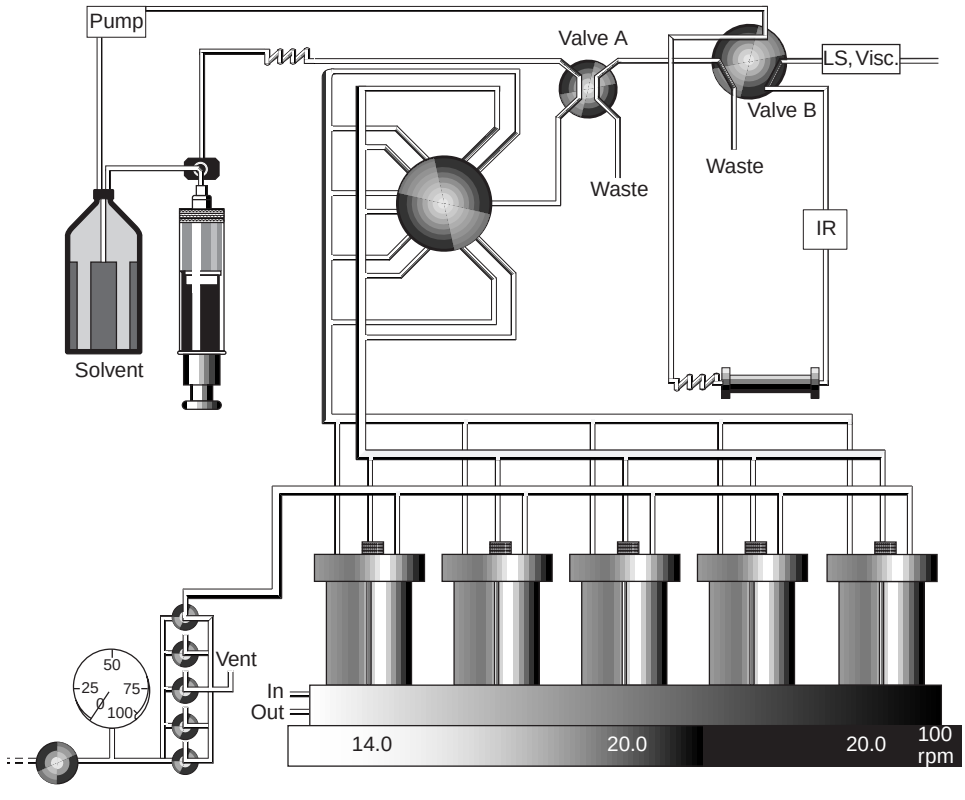

Figure 8 Automated TREF apparatus. Samples are dissolved and filtered in the stirred vessels. Aliquots of the solution are loaded into the column and crystallization begins at 0.1 °C min^{-1}. Temperature rising is done at 60 °C h^{-1} and 0.5 mL min^{-1}.

The whole process is automated from sample dissolution to column loading and temperature rising elution; once the analysis of the first sample has been completed, the equipment continues with dissolution and analysis of the other samples. An air-cooled heat exchanger in the bottom of the stirred vessels, with individual solenoid valves, prevents extended heating of the samples not being analyzed.

3 CRYSTALLIZATION ANALYSIS FRACTIONATION

CRYSTAF was developed by Monrabal[47] in 1991 as a process to speed up the analysis of the CCD, and it shares with TREF the same fundamentals on separation according to crystallizability. The way the analysis is performed – by using a discontinuous sampling process – provides the possibility of automating the technique easily and carrying out simultaneous analysis of multiple samples. Reviews of the technique and applications have been presented.[14,41]

3.1 Separation Mechanism

In CRYSTAF the analysis is carried out in stirred crystallization vessels with no support and by monitoring the polymer solution concentration during crystallization using temperature reduction, as shown in Figure 9. Aliquots of the solution are filtered (through an internal filter inside the vessel) and analyzed by a concentration detector. In fact, the whole process is similar to a classical stepwise fractionation by precipitation with the exception that, in this new approach, no attention is paid to the precipitate but to the polymer that remains in solution.

The first data points, taken at temperatures above any crystallization, provide a constant concentration equal

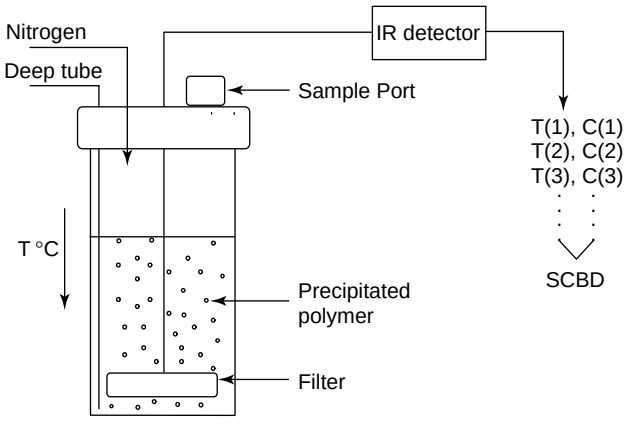

Figure 9 Principles of CRYSTAF.

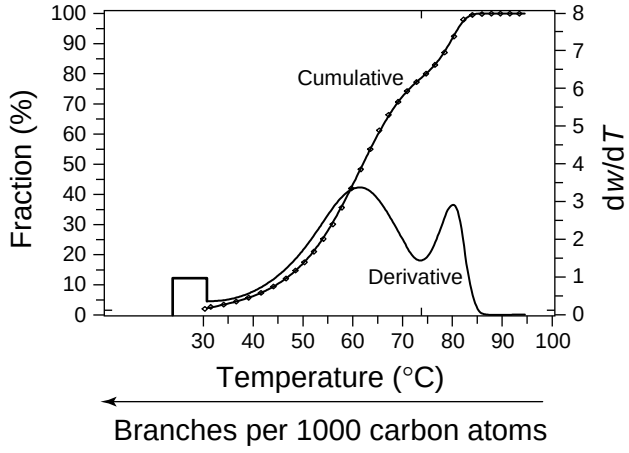

Figure 10 CRYSTAF cumulative and first-derivative curves of an LLDPE resin.

to the initial polymer solution concentration (flat side of cumulative curve in Figure 10); as the temperature goes down the most crystalline fractions – composed of molecules with zero or very few branches (highly crystalline) – will precipitate first, resulting in a steep decrease in the solution concentration on the cumulative plot. This is followed by precipitation of fractions of increasing branch content (or less crystallinity) as temperature continues to decrease; the last data point, corresponding to the lowest temperature of the crystallization cycle, represents the fraction which has not crystallized (mainly highly branched or amorphous material) and remains in solution.

The top curve in Figure 10 corresponds to the cumulative composition distribution curve (in this example an LLDPE resin). The first derivative of this curve can be associated with the CCD, as shown in Figure 10, when the temperature scale is calibrated and transformed to the number of branches per 1000 carbon atoms. With this approach the CCD can be analyzed in a single crystallization cycle without physical separation of the fractions. The term ''CRYSTAF'' stands for this process.

3.2 Experimental Set-up

Crystallization is carried out in stainless-steel stirred vessels of 50-ml volume as shown schematically in Figure 9. Five crystallization vessels are installed in the main oven (a gas chromatography oven) and attached via a rotary valve to a dual-channel optoelectronic IR detector (with 3.5 μm as the measurement wavelength), as shown schematically in Figure 11. Typical crystallization rates are 0.1–0.4 °C min^{-1} (6–24 °C h^{-1}). The IR cell is kept heated isothermally during the whole experiment, typically at 150 °C.

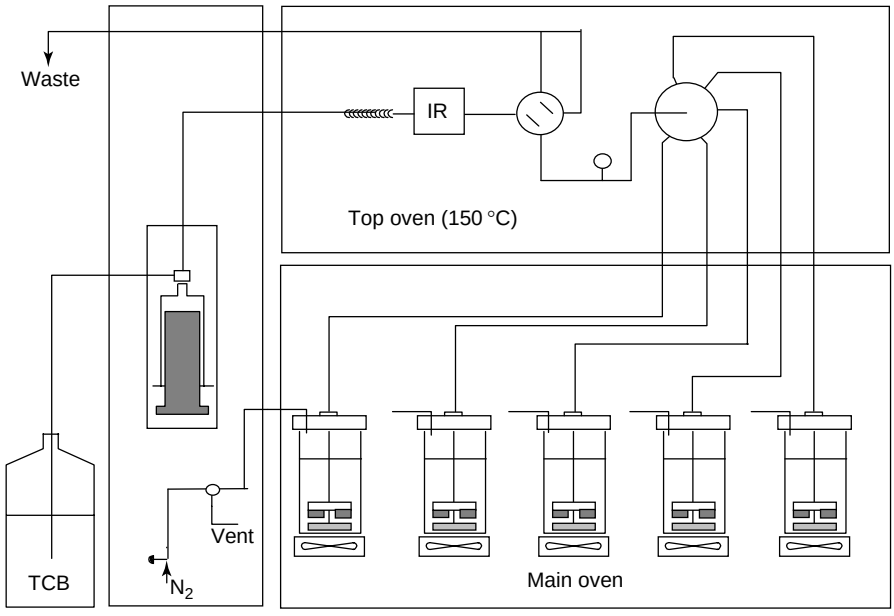

Figure 11 Schematic diagram of the CRYSTAF apparatus with five stirred vessels. TCB, trichlorobenzene.

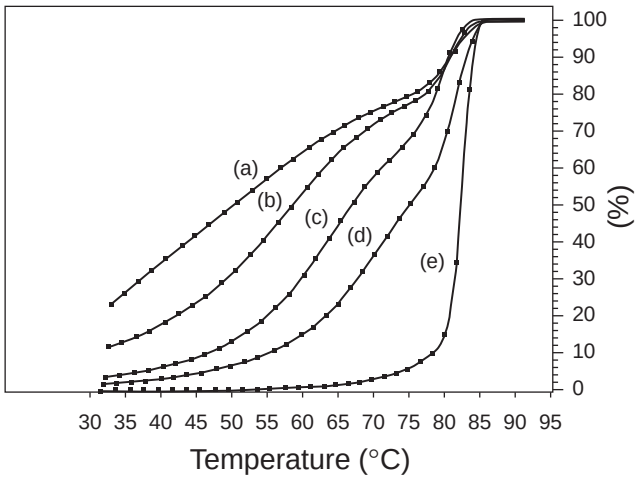

Figure 12 CRYSTAF analysis of five LLDPE samples simultaneously: cumulative curves. Crystallization rate 12 °C h^{-1}. Density of the samples: (a) 0.905; (b) 0.912; (c) 0.920; (d) 0.926; (e) 0.940.

An advantage of this process is the capability of analyzing various samples in different vessels at the same time. During crystallization all the vessels are ''sampled'' many times in a sequential manner and at the end of the analysis there are enough temperature–concentration data points ($T(i)$, $c(i)$) for each sample in order to draw the cumulative curves shown in Figure 12, with the simultaneous analysis of five LLDPE resins of different density. The complete dissolution and crystallization analysis of five samples can be carried out simultaneously in less than 7 h.

3.3 Solvents

The IR detector is also the only detector used in CRYSTAF analysis and therefore the solvent selection is the same as with analytical TREF: 1,2,4-trichlorobenzene, *o*-dichlorobenzene, perchloroethylene and α-chloronaphthalene. The addition of 0.1% of an antioxidant is recommended in order to prevent polymer degradation.

As with TREF, the solvent does not influence the separation mechanism but crystallization temperatures will be shifted depending on the solvent power;[34] therefore, comparison of CRYSTAF analysis obtained with different solvents cannot be made in dissolution temperature units, but can be when calibrated into the number of branches per 1000 carbon atoms or other polymer composition units.

3.4 Sample Size

Solution concentrations of 0.03–0.1% are typical and the amount of polymer being introduced in the stirred vessel is 10–30 mg (with 30 mL of solvent). Good results are obtained at 0.1% but if the polymer is of very high M_w then the concentration should be lowered to prevent filter plugging.

3.5 Crystallization Rate

The crystallization rate is usually between 0.1 and 0.4 °C min^{-1}. There is the possibility of using multiple-step isotherms to carry out the crystallization, with a selectable waiting time at each step; this is of special interest for

very fast analysis (process control) because the solution is brought faster to the point of measurement.[48]

4 COMPARISON OF TEMPERATURE RISING ELUTION FRACTIONATION AND CRYSTALLIZATION ANALYSIS FRACTIONATION

TREF and CRYSTAF share the same principles of fractionation on the basis of crystallizability and through a slow cooling of a polymer solution. TREF is carried out in a packed column and demands two full temperature cycles – crystallization and elution – to achieve the analysis of the composition distribution. In CRYSTAF the analysis is performed in a single step, the crystallization cycle, which results in a faster analysis time and simple hardware requirements. The use of stirred vessels and the fact that only crystallization plays a role in the separation process (in TREF there is an additional melting step where molecules have to come out of the solid state) are probably the factors responsible for achieving in CRYSTAF, at faster crystallization rates, a similar resolution to classical TREF.

TREF has the advantage that a continuous elution signal is obtained (versus the discontinuous sampling of CRYSTAF) and CRYSTAF takes advantage of this discontinuous sampling to analyze multiple samples simultaneously.

Comparison of TREF and CRYSTAF results has been discussed previously[16,49,50] and the most significant difference is in the temperature scale, as seen in Figure 13

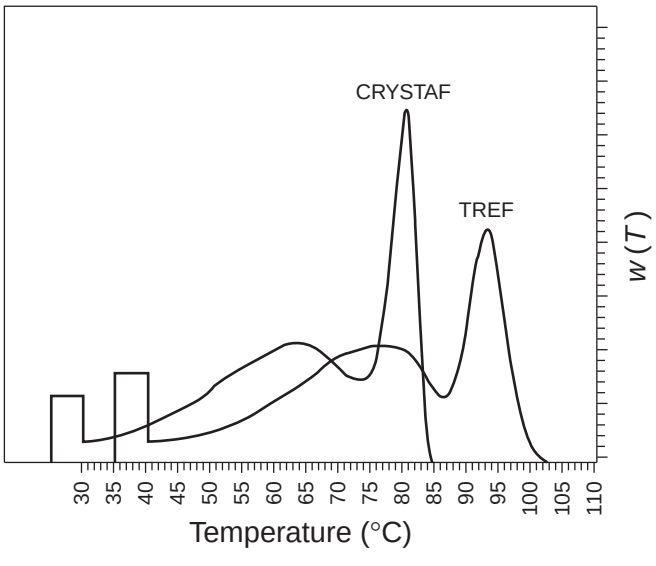

Figure 13 Comparison of TREF and CRYSTAF curves of an LLDPE resin.

where the TREF and CRYSTAF analyses of an LLDPE resin are shown; the temperature difference is due to the undercooling as CRYSTAF data are taken in the crystallization while TREF data are obtained in the melting–dissolution. Both techniques, however, can be calibrated and the results expressed in branches per 1000 carbon units, as has been discussed previously.[2,16,27]

The analysis of polypropylene–polyethylene combinations, which will be discussed in section 10.4, is a special case where both TREF and CRYSTAF may be required to characterize unequivocally such a polymer combination.

5 CALCULATIONS

Most TREF curves described in the literature are shown as raw data with the elution temperature scale without any calibration; this is an indication of the restricted community using the technique, as well as the difficulties and time involved in the calibration. All this is expected to change very soon, given the significance that CCD has on the new polyolefins, the maturity of the technique, the introduction of automated TREF and CRYSTAF apparatus and the availability of standards.

TREF curves, being obtained by a ''quasi-chromatographic'' procedure with an elution peak beginning on the baseline as in Figure 6, are not corrected for the noncrystallizing fraction. This would not present any difficulties if the end crystallization temperature was very low (subambient) or with samples (like HDPE) that result in low, close to zero, noncrystallizing fraction percentages. In most samples, such as LLDPE, high-impact polypropylene and the new multireactor resins, this is not the case and a mathematical correction is necessary. The raw data should be normalized by incorporating the soluble fraction percentage and it should reflect the unknown shape of the initial part of the curve, which in fact was not analyzed because a low enough crystallization temperature was not reached. This problem with TREF curves became rather evident in the initial stages of CRYSTAF development because in CRYSTAF the data are obtained from the cumulative curve by subtraction, as in Figure 12, and if the polymer concentration has not reached a zero value at the lowest crystallization temperature in the experiment (as is very often the case) it is obvious that one could only make a prediction of what it would be at lower temperatures. The large soluble fraction in some LLDPE resins, as shown in Figure 12 (corresponding to the reading of the last point at the lowest temperature), demands this correction.

A reasonable way to reflect the uncertainty of how the curve would look at the temperatures not analyzed is shown in Figure 13. The $w(T)$ curve ends abruptly at the

lowest crystallization temperature (and not necessarily at baseline); the soluble fraction percentage is expressed as a rectangle of constant base and varying height, with a surface area proportional to its value and being part of the overall normalized curve. The soluble fraction (the rectangle in the drawing) is calculated and plotted at the lowest temperature where the temperature scale ends. This procedure should be applied to both CRYSTAF and TREF.

Besides the curve, it is practical to work with some easy-to-use average parameters. In LLDPE the most important ones are the homopolymer and soluble fraction percentages. Calculations similar to M_n and M_w values with gel permeation chromatography (GPC) can be done in terms of elution or crystallization temperatures, as shown with Equations (5–7):

$$T_n = \frac{\sum c_i}{\sum c_i/T_i} \tag{5}$$

$$T_w = \frac{\sum c_i T_i}{\sum c_i} \tag{6}$$

$$r = \frac{T_w}{T_n} \tag{7}$$

With proper calibration, the corresponding comonomer or branching content averages can be calculated with analogous equations, understanding that such parameters do not have a physical meaning as in GPC but provide mathematical averages that emphasize both sides of the CCD to estimate the breadth of the distribution.

Other parameters have been proposed in the patent literature to describe the CCD: the composition distribution breadth index (CDBI), defined as the weight percent of the copolymer molecules having a comonomer content within 50% of the median total molar comonomer content;[51] or the solubility distribution breadth index (SDBI), which is analogous to the standard deviation of the CCD.[52]

6 CALIBRATION

To calibrate the techniques TREF or CRYSTAF in terms of branches per 1000 carbon atoms, the classical approach has been to fractionate a copolymer of the same comonomer type by preparative TREF; this is followed by characterization of the fractions, in terms of branches per 1000 carbon atoms, by nuclear magnetic resonance (NMR) or IR spectroscopy and analyzing them by the technique to be calibrated: TREF[2,27,53,54] or CRYSTAF.[16,55] In Figure 14 a calibration of the TREF technique with this approach is shown. However, the restricted availability of the preparative TREF equipment and the time involved in the preparative fractionation have prevented extensive use of the calibration to represent the CCD curves in TREF and CRYSTAF.

With the commercial development of metallocene-type resins of narrow composition distributions, the calibration work is facilitated and it is expected to be implemented to a broader extent. A collection of metallocene-type resins well characterized by IR and NMR spectroscopy, of densities ranging between 0.86 and 0.94, has been analyzed by CRYSTAF[55] and the results plotted in Figure 15. A reasonably good correlation is obtained that could be of general use in calibrating broad CCD samples of the same comonomer type when analyzed under the same conditions.

Caution should be taken to apply this calibration procedure only to samples of the same type and when it has been shown that intramolecular distribution does not play

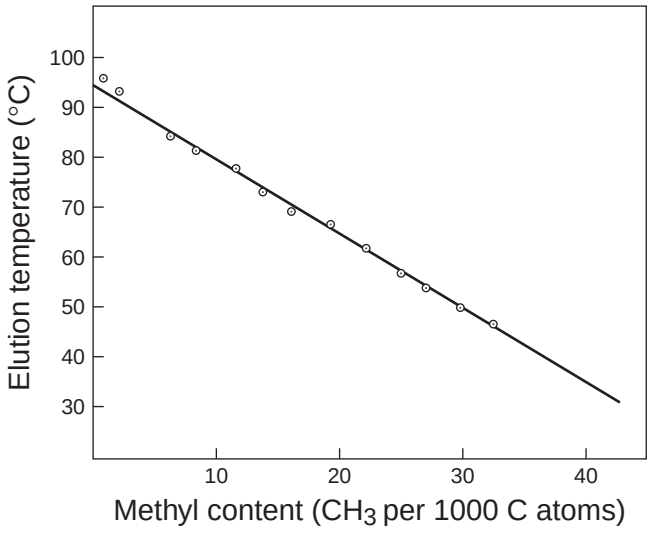

Figure 14 TREF calibration for ethylene–butene copolymers. (Reproduced by permission from Wild et al.[2])

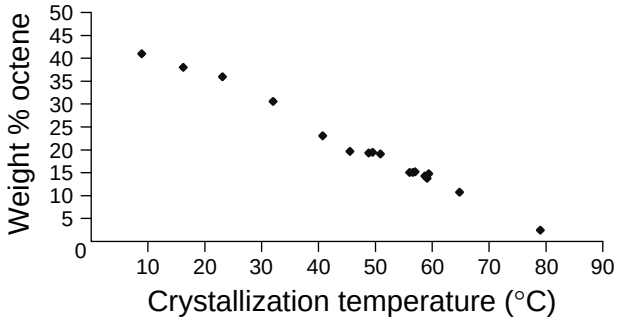

Figure 15 Calibration of CRYSTAF with a series of metallocene-type resins. (Reproduced by permission from Monrabal et al.[55])

a significant role. There are reports[13,56,57] suggesting that sequence distribution (nonuniform intramolecular composition) may have a significant influence in some cases, which would make this type of calibration inappropriate.

7 MOLECULAR WEIGHT–COMPOSITION DEPENDENCE

Although the CCD of a resin with a given comonomer type is a very important parameter, we should not forget that it does not provide the whole structural information. Full characterization of a complex polyolefin resin would require, in addition to the CCD and molecular weight distribution, an analysis of the molecular weight dependence on composition. It would also be necessary to investigate the intramolecular sequence distribution.

There are two possible analytical routes in cross-fractionation: to fractionate the polymer on a molecular weight basis, and to fractionate the polymer on a composition basis. One should choose the fractionation technique that results in the most discriminated fractions.[14] The most general approach is preparative TREF fractionation, because the CCD is more discriminating than the molecular weight distribution in complex polyolefins;[33,58,59] a good example is the characterization of an LLDPE where TREF fractionation is carried out on the bimodal CCD. In HDPE resins, with a small number of branches and showing unimodal CCD, both fractionation approaches (based on composition or molecular weight) should be appropriate. Fractionation is followed by off-line characterization using other techniques such as GPC, NMR and IR spectroscopy, DSC, TREF and CRYSTAF.

The full information of the bivariate distribution is very often presented in three-dimensional plots: contour maps or bird's eye views,[3,58,60,61] as in Figure 16 from the work by Nakano and Goto.[58]

In some cases with well-defined components (multiple reactor–catalyst resins) the average molecular weight dependence on composition, as shown in Figure 17, provides the necessary parameters for resin evaluation.

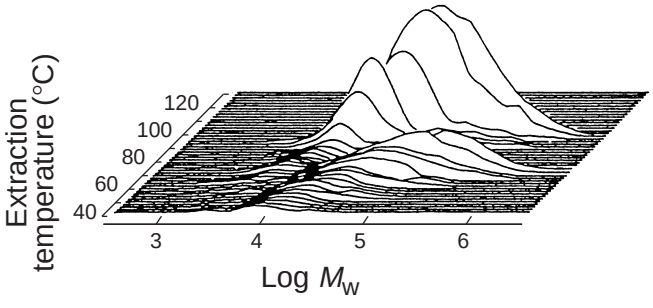

Figure 16 A bird's-eye view of a mixture of LDPE and HDPE. (Reproduced by permission from Nakano and Goto.[58])

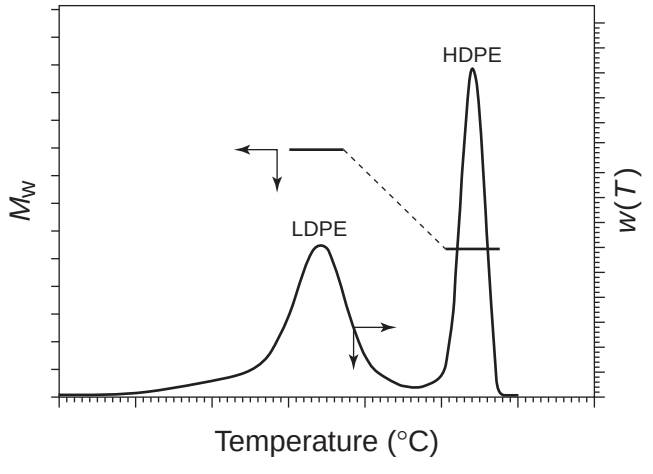

Figure 17 Molecular weight dependence of a blend of an HDPE resin of low molecular weight and an LDPE resin of high molecular weight.

8 PREPARATIVE FRACTIONATION

The first published TREF experiments by Desreux,[35] Shirayama[37] and Wild[1] were done in preparative mode to measure the CCD by weighting the obtained fractions. Later on, the incorporation of a concentration detector[3] popularized the analytical TREF approach that has been discussed previously.

Preparative TREF, although manpower intensive, has been shown to be a powerful tool for off-line characterization of the fractions by GPC, NMR and IR spectroscopy and for mechanical testing. Essentially the technique is a scale-up of the analytical TREF in order to fractionate 1–10 g of polymer through the use of larger columns (10–20 cm long and 1–3 cm o.d.) and higher flow rates (4–50 mL min^{-1}). Crystallization rates are similar to those used in the analytical mode but dissolution is typically carried out in isothermal steps (waiting 10–30 min at each step) corresponding to the number of fractions to be obtained and with the goal of minimizing radial temperature distribution across the column. Wild[42] has suggested performing the crystallization in a flask with no support to improve the fractionation resolution and prevent "tailing" by entrapment of less-crystalline material within the pores of the packing; the polymer and packing are mixed subsequently and added to the column for the elution step. Not requiring an IR detector, the technique has more flexibility in solvent selection, xylene being the solvent most used.

Various authors have described preparative TREF apparatus based on a column scaled up from analytical

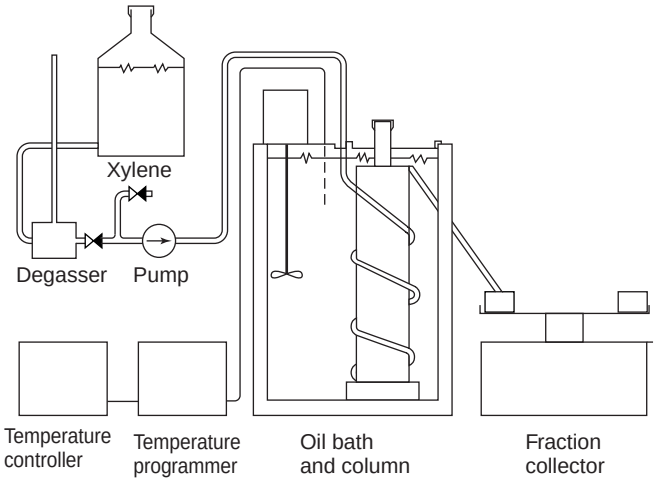

Figure 18 Preparative TREF apparatus. (Reproduced by permission from Bergström and Avela.[64])

TREF: Usami et al.,[62] Kelusky et al.[63] and Bergström et al.[64] Compilations of operating conditions for preparative TREF have been reported[34,38,39,41] and detailed operational procedures described by Francuskiewicz.[65] The apparatus used by Bergström[64] is shown schematically in Figure 18.

Preparative TREF fractionation carried out in vessels, without a column, has been described by Mirabella[66] and Monrabal.[67] A schematic diagram of an upgraded design for the simultaneous fractionation of two samples based on the last reference is shown in Figure 19. Sample dissolution, filtration through an internal filter in the stirred vessel and fraction collection are fully automated. Fractionation of 1 g of sample per vessel is easily achieved. The same apparatus can be used to fractionate in a crystallization mode and on a molecular weight basis as well.

One should be alert to combinations of polypropylene and polyethylene, as will be discussed in section 10.4, because there may be situations where a TREF fractionation would not provide the separation and so a crystallization fractionation would be necessary.[68,69]

9 AUTOMATED CROSS-FRACTIONATION APPARATUS

Preparative fractionation followed by off-line analysis of the fractions is the most flexible approach in the characterization of unknown samples. Nevertheless, there have been various attempts to automate the cross-fractionation apparatus, although this always results in more complex constructions.

The most comprehensive approach is the combination of TREF and GPC as described by Nakano and Goto,[58] which has been widely used in LLDPE to obtain the full molecular weight dependence on composition with three-dimensional plots, as in Figure 16. The equipment can run one sample per day.

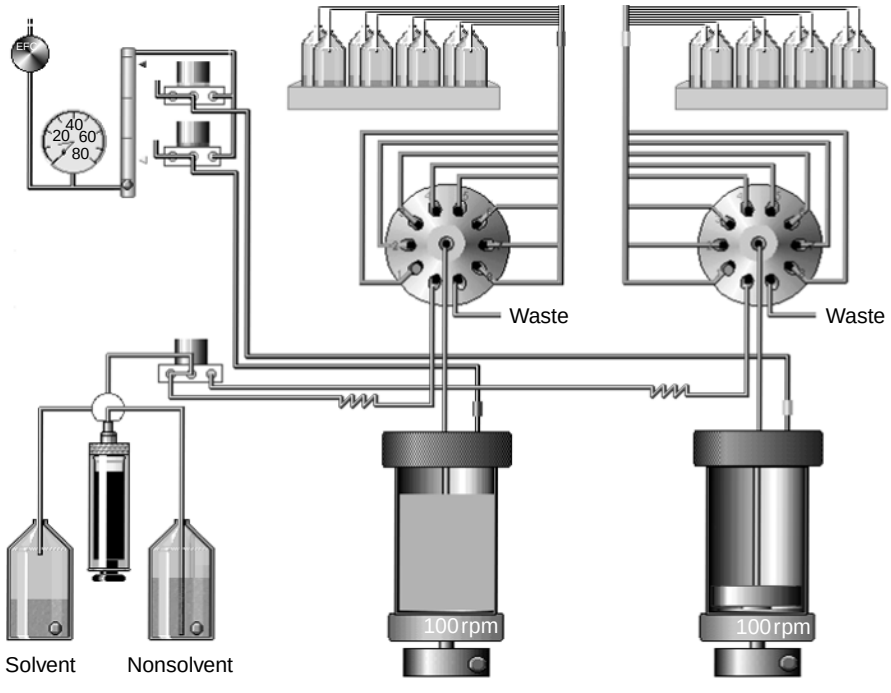

Figure 19 Preparative TREF apparatus with stirred vessels.

The incorporation of IR spectroscopy to GPC through a collecting germanium disc interface[70] has been shown to be of interest for HDPE resins to obtain the branching dependence on molecular weight. A new solid-state IR detector to be used in-line with GPC or TREF/CRYSTAF has been presented with the possibility to measure concentration and composition in functional copolymers.[71]

The incorporation of molecular mass detectors in analytical TREF or CRYSTAF has been explored with interest to achieve simple molecular weight dependence on composition as described in Figure 17. Incorporation of a viscometer[72] in automated TREF and incorporation of a light-scattering detector in TREF[73] and CRYSTAF[68] have been reported, and more recently the addition of both: viscometer and light-scattering detectors in analytical TREF[74] and CRYSTAF[46] have been described to discriminate molecular weight and long-chain branching present in some polyolefins. The potential to use these techniques in the process control of sophisticated multiple catalyst–reactor systems is also discussed, with the capability of analyzing one sample every 2 h.[46]

10 APPLICATIONS

10.1 Linear Low-density Polyethylene

10.1.1 Ziegler-type Linear Low-density Polyethylene Resins

In LLDPE the incorporation of comonomer into the linear polyethylene chains results in side-chain branches that modify the crystallinity and therefore the final morphology and performance of the resin. The intermolecular distribution of short-chain branches, as discussed earlier, is not uniform due to the multiple active site types present in supported Ziegler catalysts, which result in a different selectivity of those sites toward the incorporation of monomer and comonomer into the growing chain. Most interesting in LLDPE is the bimodality, as shown in Figure 1, due to the population discontinuity observed between the fraction of linear molecules (practically excluding the comonomer incorporation) and the remaining fractions with increasing amounts of comonomer incorporated. A considerable effort has been made to understand the bimodal nature of the CCD in commercial Ziegler-type LLDPE resins in terms of catalyst active sites,[62] how to modify the CCD and how it affects the end-product properties.[2,10,60,75–83] Mathematical modeling of the experimental TREF curve has demanded some attention in view of the multiple site activity.[33,41] Broader reviews on the structural characterization of LLDPE have been published.[14,40,60,83]

Cross-fractionation of LLDPE beginning with a TREF step and followed by GPC analysis has been reported by Nakano et al.,[58] Hosoda,[10,60,82] Wild,[3] Mirabella,[84] Wilfong et al.[54] and Usami et al.[62] Characterization of LLDPE and medium-density polyethylene (MDPE) resins beginning with a GPC fractionation have been reported by Mathot et al.[85] and Aust et al.[61]

The significance of the CCD in LLDPE has been discussed by Karbashewski et al.[86] with the analysis of four ethylene–octene copolymers with a similar melt index. The molecular weight distribution of all samples was identical within the limits of the GPC technique; however, the CCD analyzed by TREF presented a large discrimination between samples, as shown in Figure 20.

The CCD, measured by CRYSTAF, of two commercial LLDPE resins with completely different performance, in spite of having identical density and melt index, is shown in Figure 21. In sample (a) the higher *homopolymer*

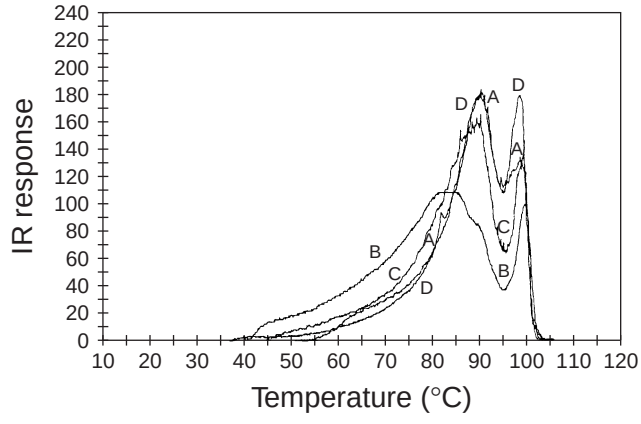

Figure 20 TREF analysis of various ethylene–octene LLDPE resins. (Reproduced by permission from Karbashewski et al.[86])

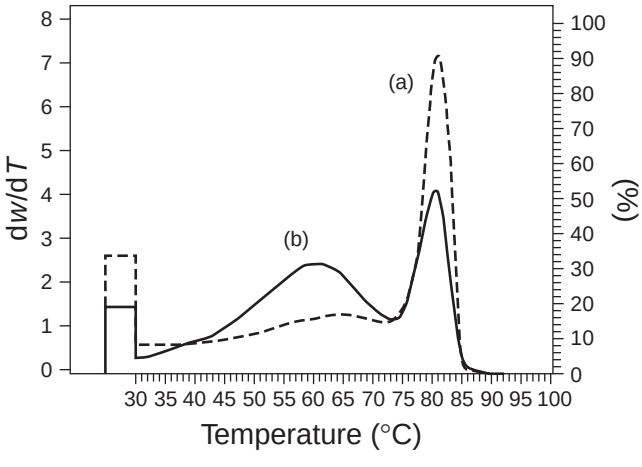

Figure 21 CCD of two LLDPE resins with identical density (0.920) and melt index (3). The analysis was done with CRYSTAF by crystallizing from 95 °C to 30 °C at a crystallization rate of 12 °C h^{-1}.

fraction content is compensated by a higher *soluble fraction* percentage to result in the same density as sample (b) but, overall, having a completely different structure.

10.1.2 Metallocene-type Linear Low-density Polyethylene Resins

CRYSTAF has been used recently in the CCD characterization of single-site catalyst resins.[87] Although most single-site catalyst resins show a narrow CCD, some catalysts result in resins with broad and tailing CCD, especially with supported catalysts.

The CCD of metallocene-type resins has been found experimentally to be broader for the lower M_w resins[55] and this has been shown to fit well with Stockmayer's distribution prediction.[88]

The commercial introduction of homogeneous single-site catalysts (metallocenes), which result in resins with narrow CCD and molecular weight distributions, and their combination with Ziegler-type resins, through blending or multiple catalyst–reactor processes, provides broad designing capabilities for new products and the need to determine the percentages of the various components present. In Figure 22 the analysis of a 50 : 50 combination of two metallocene-type resins (with narrow CCD) is shown as an example of the above discussion.

Some of the metallocene-type resins are produced with significant amounts of comonomer, resulting in very low densities, and it is rather common to use subambient temperature conditions to reach crystallization of the resin, as can be seen in the plot of Figure 15.

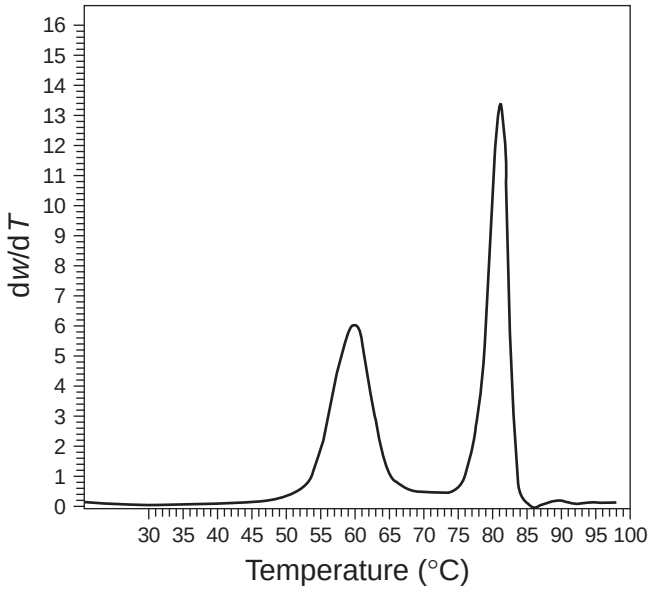

Figure 22 Analysis of a combination of two metallocene-type resins by CRYSTAF. Crystallization rate is 18 °C h^{-1}.

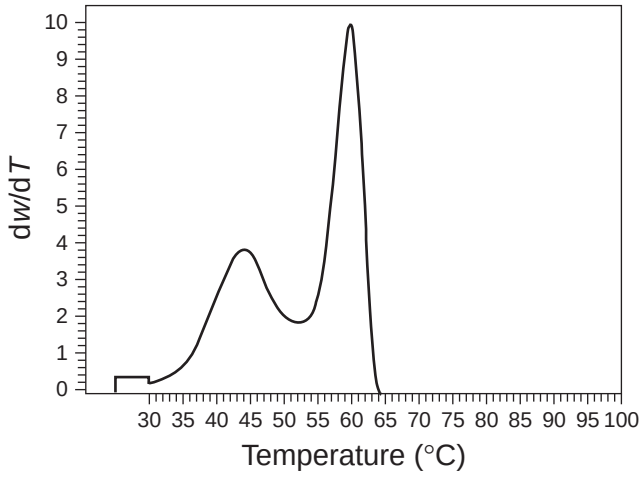

Figure 23 Fast analysis of a blend or dual reactor sample. Crystallization rate is 30 °C h^{-1}.

10.1.3 High-speed Analysis for Process Control

The possibility of using CRYSTAF for process control has been discussed[48] and it is shown in Figure 23 with a hypothetical dual reactor sample (blend of two metallocene-type resins). The analysis is carried out with a dissolution time of 10 min at 140 °C, followed by fast cooling down to 66 °C and a stabilization time of 10 min at this temperature; the crystallization is done in multiple isothermal steps between 66 and 30 °C, with sampling at every 2 °C and waiting isothermally 1 min at each step before sampling. The introduction of preheated solvent in the vessel, the easy dissolution of these products (low density) and the relatively narrow crystallization range make it possible to analyze a sample every 90 min.

10.2 Low-density Polyethylene and Functional Olefinic Copolymers

The structure of LDPE, being fabricated by a free radical polymerization process, results in the least defined and most complex structure of all the polyolefins. Intramolecular chain transfer (back-biting mechanism) results in the presence of short-chain branches, mainly ethyl and butyl; addition of comonomers also results in short-chain branches and influences chain growth; inter- and intramolecular chain transfer reactions result in long-chain branches of similar size to the main chain, and are responsible for the high melt strength of these resins.

Being the first polyethylene ever produced commercially, there is abundant literature on the characterization of LDPE in terms of type and population of short branches, intramolecular characterization by NMR spectroscopy and measurement of long-chain branching and molecular weight distribution.[89–91] Shirayama et al.[37]

analyzed the CCD of LDPE in 1965. Later investigations showed that the SCBD in LDPE is significantly narrower than that in LLDPE products, indicating a more uniform composition distribution of branches,[2,58,64,90] although still far from pure HDPE.[63] Cross-fractionation of LDPE beginning with a TREF fractionation step has been reported by Usami et al.,[62] Kulin et al.[90] and Joskowicz et al.[92]

The CCD curves of three resins of similar melt index and density, but from different suppliers, are shown in Figure 24, indicating a significant difference between samples.

Very little work has been published on functional copolymers typically produced with the same manufacturing process as LDPE. The TREF analysis of ethylene vinyl acetate has been reported[27,63] and it has been shown that the acetate branches are distributed through the polymer in a similar manner to the alkyl branches. The analysis of ethylene acrylic acid by CRYSTAF using a dual-wavelength IR detector at 3.5 and 5.8 μm[68] has shown an unusual CCD with the carboxyl population shifted toward the more crystalline fractions; although no column support is used in CRYSTAF, the potential adsorption of the carboxylic groups in parts of the apparatus should be investigated further.

10.3 High-density Polyethylene

The CCD of a homopolymer HDPE is a narrow peak eluting at around 95–100 °C in TREF and crystallizing at around 80–85 °C in CRYSTAF. These techniques are of little value with this type of resin showing just a single peak, which is practically independent of molecular weight. Most products sold under the name of HDPE, however, are ethylene copolymers with small amounts of α-olefins, such as propylene, butene,

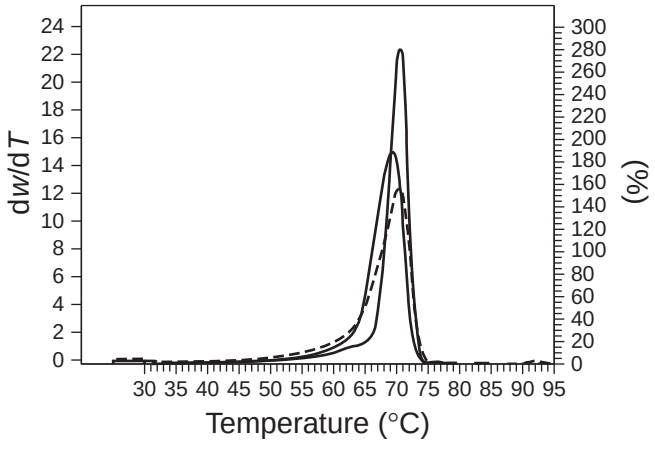

Figure 24 CCD of three LDPE resins from different suppliers analyzed by CRYSTAF.

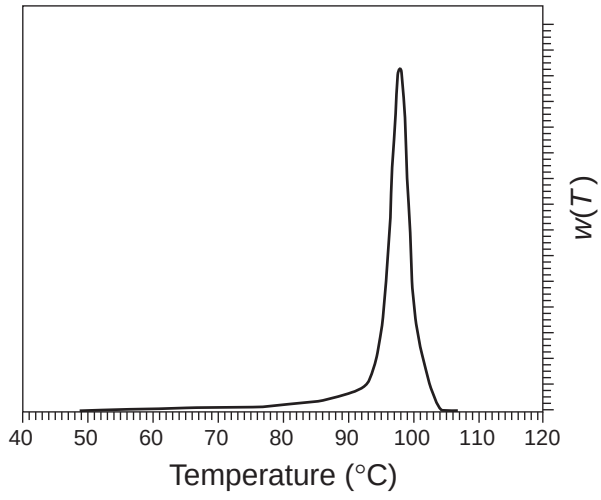

Figure 25 TREF of an HDPE resin.

etc. In those cases the CCD (with Ziegler-type resins, nonhomogenous products) has a tail toward the lower temperature, as shown in Figure 25. This could be considered, in terms of CCD, as the extreme case of an LLDPE resin where no bimodality is shown because the homopolymer fraction is so large in HDPE; the low comonomer content in these polymers also results in close to zero values in the soluble or noncrystallizing fraction.

10.4 Polypropylene and Copolymers

10.4.1 Polypropylene Analysis

Polypropylene homopolymer may be present in three different configurations (as depicted in Figure 3) with different crystallinity, the isotactic being the most crystalline and the atactic being amorphous.

The separation of these configuration forms was already shown by Kamath and Wild in 1966[93] using a step crystallization procedure; since then, quite a few authors have characterized polypropylene tacticity by crystallization techniques,[66,94,95] resulting in curves like that in Figure 26 where, typically, syndiotactic and atactic forms are undesirable fractions of the isotactic polypropylene fabrication process.

The more common polypropylene homopolymer (isotactic form), however, is not fully homogeneous due to the presence of steric defects, and intramolecular characterization, typically by NMR spectroscopy, is also used[66,95] to measure the overall degree of isotacticity (pentad determination). The more irregularities present, the less crystalline the chains will be, and the corresponding peak in Figure 26 will be shifted toward lower crystallization temperatures.[96–98]

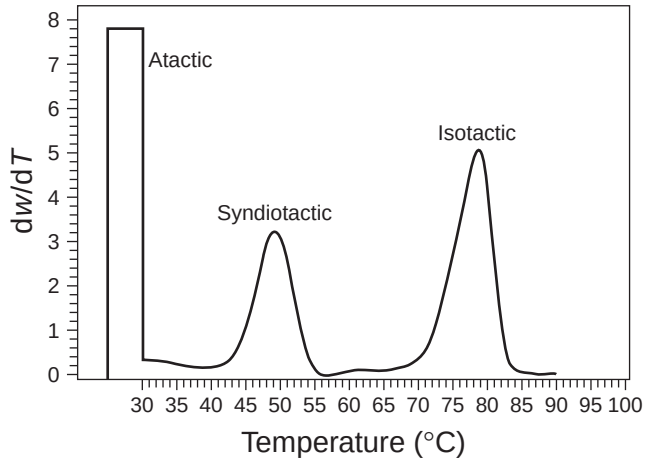

Figure 26 CRYSTAF analysis of a blend of atactic, syndiotactic and isotactic polypropylenes. Crystallization rate is 12 °C h^{-1}.

The incorporation of various levels of ethylene to polypropylene (propylene–ethylene copolymers) will influence the crystallinity as well, and the peak in Figure 26 will also be shifted to the left proportionally to the level of ethylene incorporated.[97,99] Kakugo et al. have shown that the fractionation equipment could be calibrated with the pentad fraction, and this calibration is somewhere general for polypropylene homopolymer and propylene–ethylene copolymers.[96] Usami et al.[94] have shown that in some cases the formation of an LLDPE fraction can take place as well, further complicating the structure of these copolymers.

Fractionation by TREF of polypropylene followed by GPC and NMR analysis has been described by Mingozzi[98] and Xu et al.[100,101] Preparative fractionation by TREF of propylene copolymers have been reported by Usami,[94] Kakugo et al.,[96] Abiru et al.[102] and Feng et al.[103]

10.4.2 Ethylene–Propylene Rubber in High-impact Polypropylene

Characterization of high-impact polypropylene by analytical and preparative TREF has been presented by Mirabella[66] as shown in Figure 27; besides the homopolymer peak and the amorphous rubber fraction, other crystallizable zones from ethylene–propylene fractions were identified.

Analysis of the ethylene–propylene rubber (EPR) content in high-impact polypropylene with TREF has been shown to be possible by measuring the difference between the polymer weight loaded into the column and the crystalline portion eluted.[27] CRYSTAF is especially advantageous in this type of application because of the precision by which the amorphous fraction is calculated. An example is shown in Figure 28 with a resin containing

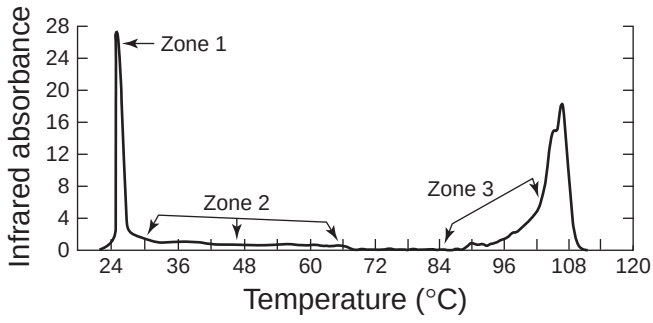

Figure 27 TREF analysis of a commercial impact-grade polypropylene copolymer. (Reproduced by permission from Mirabella.[66])

24% of EPR; it also provides a value of the ethylene content in the propylene–ethylene copolymer by the crystallization temperature of the main peak.

10.4.3 Polypropylene–High-density Polyethylene Combinations

In TREF the elution temperature of polypropylene is significantly higher than that of HDPE ($PP_{TREF} \approx 120\,^{\circ}C > PE_{TREF} \approx 95\,^{\circ}C$) whereas in CRYSTAF the opposite crystallization sequence is observed due to the large undercooling of polypropylene ($PP_{CRYSTAF} \approx 80\,^{\circ}C < PE_{CRYSTAF} \approx 85\,^{\circ}C$). If one needs to analyze a blend of both polypropylene and HDPE, TREF would provide a better peak separation than CRYSTAF.

The presence of ethylene in propylene–ethylene copolymers will decrease the TREF peak elution temperature and, depending on the level of ethylene, it can elute at the same position as HDPE. On the other hand, when analyzing this blend with CRYSTAF, the

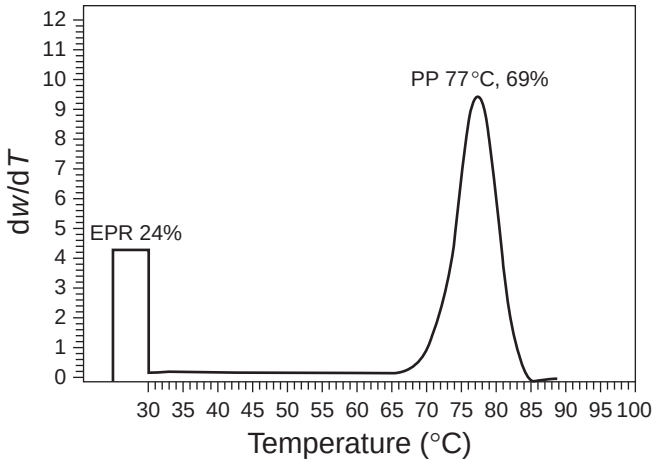

Figure 28 Analysis of EPR in high-impact polypropylene (PP) by CRYSTAF at a crystallization rate of 12 °C h^{-1}.

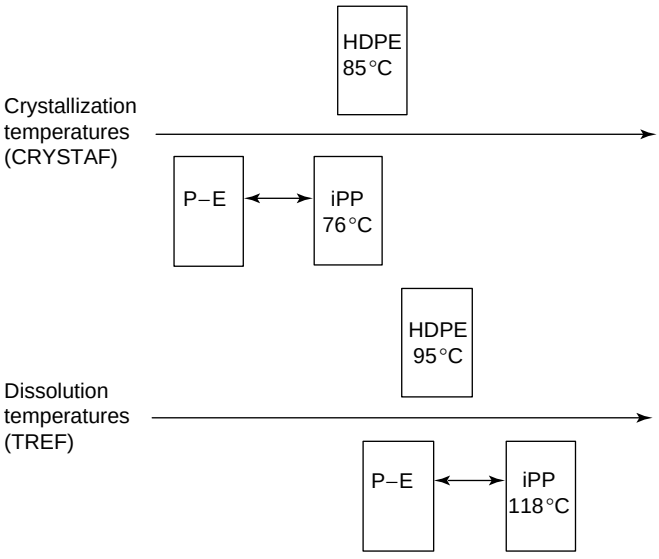

Figure 29 Schematic diagram of the separation of polypropylene (PP) and HDPE by CRYSTAF and TREF. P–E, propylene–ethylene copolymer.

presence of ethylene in polypropylene–ethylene copolymers will decrease the peak crystallization temperature below that of pure polypropylene, therefore increasing the separation from the HDPE peak. In this type of combination, CRYSTAF would provide a better separation than TREF. This phenomenon is described schematically in Figure 29 and will be discussed in further detail with experimental data in a forthcoming publication.[69]

11 CONCLUSIONS

The importance of the CCD in polyolefin characterization has been discussed in terms of the broad and bimodal distributions of polyolefins such as LLDPE. The introduction of single-site catalysts in multireactor–multicatalyst processes opens up the possibility of designing new resins with optimized performance. The CCD of these complex resins is the most important and discriminating characterization parameter but full characterization also requires information on the dependence of molecular weight composition.

TREF has been the most successful technique to analyze the CCD of polyolefin resins in spite of its long analysis time and complex hardware requirements. Simplification of the apparatus and automation of the analysis process, especially in relation to sample preparation, should make TREF more easily available.

The new CRYSTAF technique provides similar results to TREF and offers, with a simpler design, the capability to run multiple samples in a shorter time. It aims to be a workhorse technique in laboratories running hundreds of CCD analyses per month, as well as for quality and process control. Both CRYSTAF and TREF are complementary techniques when analyzing combinations of polyethylene and propylene–ethylene copolymers.

Although preparative TREF appears to be the most powerful route to cross-fractionation, the development of CRYSTAF and TREF apparatus with light-scattering/viscometry detectors attached seems to be a promising technique for the characterization of the composition–mass dependence in polyolefin resins.

ABBREVIATIONS AND ACRONYMS

CCD	Chemical Composition Distribution
CDBI	Composition Distribution Breadth Index
CRYSTAF	Crystallization Analysis Fractionation
DSC	Differential Scanning Calorimetry
EPR	Ethylene–Propylene Rubber
GPC	Gel Permeation Chromatography
HDPE	High-density Polyethylene
HPLC	High-performance Liquid Chromatography
IR	Infrared
LDPE	Low-density Polyethylene
LLDPE	Linear Low-density Polyethylene
MDPE	Medium-density Polyethylene
NMR	Nuclear Magnetic Resonance
SCBD	Short-chain Branching Distribution
SDBI	Solubility Distribution Breadth Index
TREF	Temperature Rising Elution Fractionation

RELATED ARTICLES

Polymers and Rubbers **(Volume 9)**
Field Flow Fractionation in Analysis of Polymers and Rubbers • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymers and Rubbers • Size-exclusion Chromatography of Polymers

REFERENCES

1. L. Wild, T. Ryle, *Polym. Prepr., Am. Chem. Soc., Polym. Chem. Div.*, **18**, 182 (1977).
2. L. Wild, T. Ryle, D. Knobeloch, I.R. Peat, *J. Polym. Sci., Polym. Phys. Ed.*, **20**, 441 (1982).
3. L. Wild, T. Ryle, D. Knobeloch, *Polym. Prepr., Am. Chem. Soc. Polym. Chem. Div.*, **23**, 133 (1982).

4. T. Kamiya, N. Ishikawa, S. Kambe, N. Ikegami, H. Nishibu, T. Hattori, 'A Comonomer Distribution Analysis of LLDPE Using Stepwise Isothermal Segregation Technique by DSC', *Antec'90*, 871–873, 1990.
5. E. Adisson, M. Ribeiro, A. Deffieux, M. Fontanille, *Polymer*, **33–20**, 4337 (1992).
6. S.A. Karoglanian, I.R. Harrison, *Thermochim. Acta*, **212**, 143 (1992).
7. J.J. Mara, K.P. Menard, *Acta Polym.*, **45**, 378 (1994).
8. G. Balbontin, T. Dall'occo, A. Boitchencko, A. Finotti, *Int. J. Polym. Anal. Charact.*, **1**, 217 (1995).
9. P. Starck, *Polym. Int.*, **40**, 111 (1996).
10. S. Hosoda, *Polym. J.*, **20**, 383 (1988).
11. D. Parikh, B.S. Childress, G.W. Knight, *ANTEC Conf. Proc.*, 1543, 1991.
12. M.Y. Keating, E.F. McCord, *Thermochim. Acta*, **243**, 129 (1994).
13. S.A. Karoglanian, I.R. Harrison, *Thermochim. Acta*, **212**, 143 (1992).
14. B. Monrabal, in *New Trends in Polyolefin Science and Technology*, ed. S. Hosoda, Research Signpost, 126, 1996.
15. V.B.F. Mathot, *New Advances in Polyolefins*, ed. T.C. Chung, Plenum Press, New York, 121–146, 1993.
16. B. Monrabal, *J. Appl. Polym. Sci.*, **52**, 491 (1994).
17. P.J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, Chapters XII and XIII, 1953.
18. M.L. Huggings, H. Okamoto, *Polymer Fractionation*, ed. M.J. Cantow, Academic Press, New York, Chapter A, 1967.
19. G.M. Guzman, *Fractionation of High Polymers*, eds. J.C. Robb, F.W. Peaker, Heywood & Company, London, 115–183, 1961.
20. O. Fuchs, W. Schmieder, *Polymer Fractionation*, ed. M.J. Cantow, Academic Press, New York, Chapter D, 1967.
21. G. Riess, P. Callot, *Fractionation of Synthetic Polymers*, ed. L.H. Tung, Marcel Dekker, New York, Chapter 5, 1977.
22. P.J. Flory, *Trans. Faraday Soc.*, **51**, 848 (1948).
23. L. Mandelkern, *Crystallization of Polymers*, McGraw Hill, New York, 1963.
24. J.H. Elliott, *Polymer Fractionation*, ed. M.J. Cantow, Academic Press, New York, 1967.
25. E.F. Casassa, *Fractionation of Synthetic Polymers*, ed. L.H. Tung, Marcel Dekker, New York, Chapter B-2, 1977.
26. A.J. Pennings, *Characterization of Macromolecular Structure*, National Academy of Sciences, Washington, DC, Publication No. 1573, 215, 1968.
27. L. Wild, C. Blatz, *New Advances in Polyolefins*, ed. T. Chung, Plenum Press, New York, 147–157, 1993.
28. R. Alamo, L. Mandelkern, *Macromolecules*, **22**, 1273 (1989).
29. J.G. Fatou, *Encyclopedia of Polymer Science and Engineering*, 2nd edition, eds. H.F. Mark, J.I. Kroschwitz, Wiley, New York, 231, 1989.
30. R.G. Alamo, R.H. Glaser, L. Mandelkern, *J. Polym. Sci., Polym. Phys. Ed.*, **26**, 2169 (1988).
31. M.J. Galante, L. Mandelkern, R.G. Alamo, *Polymer*, **39**, 5105 (1998).
32. C.J. Neves, E. Monteiro, A.C. Habert, *J. Appl. Polym. Sci.*, **50**, 817 (1993).
33. J.B.P. Soares, A.E. Hamielec, *Macromol. Theory Simul.*, **4**, 305 (1995).
34. G. Glöckner, *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **45**, 1–24 (1990).
35. V. Desreux, M.L. Spiegels, *Bull. Soc. Chim. Belg.*, **59**, 476 (1950).
36. S.W. Hawkins, H. Smith, *J. Polym. Sci.*, **23**, 341 (1958).
37. K. Shirayama, T. Okada, S.I. Kita, *J. Polym. Sci. A*, 907 (1965).
38. L. Wild, *Adv. Polym. Sci.*, **98**, 1–47 (1991).
39. C.A. Fonseca, I.R. Harrison, *Modern Techniques for Polymer Characterisation*, ed. R.A. Pethrick, 1–13, 1999.
40. J.B.P. Soares, A.E. Hamielec, *Polymer*, **36**(8), 1639 (1995).
41. J.B.P. Soares, A.E. Hamielec, *Modern Techniques for Polymer Characterisation*, ed. R.A. Pethrick, 15–55, 1999.
42. L. Wild, *Trends Polym. Sci.*, **1**, 50–55 (1993).
43. V. Mierau, D. Voight, E. Bohme, *J. Appl. Polym. Sci.*, **63**, 283–288 (1997).
44. L.G. Hazlitt, *J. Appl. Polym. Sci., Appl. Polym. Symp.*, **45**, 25 (1990).
45. S. Takayama, Y. Goto, US Patent, 4 992 168, 1991.
46. B. Monrabal, A. Ortin, L. Romero, *12th Int. Symp. on Polymer Analysis and Characterization*, La Rochelle, June 28, 1999.
47. B. Monrabal, 'Crystallization Analysis Fractionation', US Patent 5 222 390, 1991.
48. B. Monrabal, *8th Int. Business Forum on Specialty Polyolefins, SPO'98*, Houston, Sept. 23, 1998.
49. B. Monrabal, *International GPC Symp. Proc.*, Orlando, 771–778, 1994.
50. B. Monrabal, *International GPC Symp.*, San Diego, 50–56, 1996.
51. C.F. Stehling, Int. Patent WO 90/03414, 5 April, 1990.
52. C.R. Davey, et al., US Patent 5 322 728, June 21, 1994
53. K. Shirayama, S.I. Kita, H. Watabe, *Makromol. Chem.*, **151**, 97 (1972).
54. D. Wilfong, G.W. Knight, *J. Polym. Sci., Polym. Phys. Ed.*, **28**, 861 (1990).
55. B. Monrabal, J. Blanco, J. Nieto, J.B.P. Soares, *J. Polym. Sci., Chem. Ed.*, **37**, 89–93 (1999).
56. E. Karbashewski, A. Rudin, L. Kale, W.J. Tchir, *Polym. Eng. Sci.*, **33**, 1370 (1993).
57. M. Pigeon, A. Rudin, *J. Appl. Polym. Sci.*, **51**, 303 (1994).
58. S. Nakano, Y. Goto, *J. Appl. Polym. Sci.*, **26**, 4217 (1981).

59. J.M. Vela, A.E. Hamielec, *Polym. Rect. Eng.*, **1**(1), (1992–93).
60. S. Hosoda, *Trends Polym. Sci.*, **3**, 265–297 (1993).
61. N. Aust, I. Beytollahi, K. Lederer, *Int. J. Polym. Anal. Charact.*, **1**, 245 (1995).
62. T. Usami, Y. Goto, S. Takayama, *Macromolecules*, **19**, 2722 (1986).
63. E.C. Kelusky, C.T. Elston, R.E. Murray, *Polym. Eng. Sci.*, **27**, 1562 (1987).
64. C. Bergström, E. Avela, *J. Appl. Polym. Sci.*, **23**, 163 (1979).
65. F. Francuskiewicz, in *Polymer Fractionation*, ed. G. Glöckner, Springer-Verlag, Berlin, 137, 1994.
66. F.M. Mirabella, *Polymer*, **34**, 1729 (1993).
67. B. Monrabal, *International GPC Symp.*, San Diego, 415–419, 1996.
68. B. Monrabal, L. Romero, *International GPC Symp.*, Arizona, 267–274, 1998.
69. B. Monrabal, P. Del Hierro, *J. Appl. Polym. Sci.* (in press).
70. J.N. Willis, J.L. Dwyer, M.X. Liu, *International GPC Symp.*, Orlando, 345–366, 1994.
71. J.V. Bosch, A. Ortin, B. Monrabal, *International GPC Symp.*, Arizona, 633–640, 1998.
72. L.G. Hazlitt, D.G. Moldovan, US Patent 4 798 081, 1989.
73. M.S. Jeansonne, W.W. Yau, *International GPC Symp.*, San Diego, 1996.
74. W.W. Yau, D. Gillespie, *International GPC Symp.*, Arizona, 252–265, 1998.
75. L.D. Cady, *SPE RETEC Proc. Akron Section, Soc. Plast. Eng.*, 107, 1985.
76. F.M. Mirabella, E.A. Ford, S.P. Westphal, P.L. Fernando, *Polym. Prepr.*, **26**, 182 (1985).
77. L.D. Cady, 'LLDPE Properties Tied to Branch Distribution', *Plastics Eng.*, 25–27 (1987).
78. K.K. Dohrer, L.G. Hazlitt, N.F. Whiteman, *J. Plast. Film Sheet.*, **4**, 214 (1988).
79. F. Mirabella, S.P. Westphal, P. Fernando, E. Ford, *J. Polym. Sci., Polym. Phys. Ed.*, **26**, 1995 (1988).
80. J.M. Brady, E. Thomas, *J. Polym. Sci., Polym. Phys. Ed.*, **26**, 2385 (1988).
81. A. Lustiger, N. Ishikawa, *J. Polym. Sci., Polym. Phys. Ed.*, **29**, 1047 (1991).
82. S. Hosoda, *Polym. J.*, **20**, 383 (1988).
83. A. Todo, N. Kashiwa, *Macromol. Symp.*, **101**, 301 (1996).
84. F. Mirabella, E.A. Ford, *J. Polym. Sci.*, **25**, 777 (1987).
85. V. Mathot, T. Pijpers, W. Bunge, *Polym. Mater. Sci. Eng. ACS Polym. Mater. Div.*, **67**, 143 (1992).
86. E. Karbashewski, L. Kale, A. Rudin, W.J. Tchir, D.G. Cook, J.O. Pronovost, *J. Appl. Polym. Sci.*, **44**, 425–434 (1992).
87. M. Atiqullah, H. Hammawa, M.N. Akhtar, J.H. Khan, H. Hamid, *J. Appl. Polym. Sci.*, **70**, 137–147 (1998).
88. J. Soares, B. Monrabal, J. Blanco, J. Nieto, *Macromol. Chem. Phys.*, **199**, 1917–1926 (1998).
89. F.M. Mirabella, J.F. Johnson, *J. Macromol. Sci. – Rev. Macromol. Chem.*, **C12**, 81–108 (1975).
90. L.I. Kulin, N.L. Meijerink, P.S. Starck, *Pure Appl. Chem.*, **60**, 1404 (1988).
91. T. Usami, Y. Gotoh, S. Takayama, *J. Appl. Polym. Sci.*, **43**, 1859 (1991).
92. P.L. Joskowicz, A. Muñoz, J. Barrera, *Macromol. Chem. Phys.*, **96**, 385–398 (1995).
93. M. Kamath, L. Wild, *Polym. Eng. Sci.*, 213 (1966).
94. T. Usami, Y. Gotoh, H. Umemoto, S. Takayama, *J. Appl. Polym. Sci.*, **52**, 145 (1993).
95. F.M. Mirabella, *J. Liq. Chromatogr.*, **17**, 3201 (1994).
96. M. Kakugo, T. Miyatake, K. Mizunuma, Y. Kawai, *Macromolecules*, **21**, 2309 (1988).
97. F.J. Kolpak, J.E. Brady, *International GPC Symp. Proc.*, Waters Corporation, Milford, MA, 491–508, 1994.
98. I. Mingozzi, *Chromatography of Polymers and Related Substances Symp.*, Bratislava, 1995.
99. W.J. Long, M.D. Wolkowicz, *International GPC Symp. Proc.*, Waters Corporation, Milford, MA, 509–521, 1994.
100. J. Xu, L. Feng, S. Yang, Y. Yang, X. Kong, *Polym. J.*, **29**(9), 713–717 (1997).
101. J. Xu, Y. Deng, L. Feng, C. Cui, W. Chen, *Polym. J.*, **30**(10), 824–827 (1998).
102. T. Abiru, A. Mizuno, F. Weigand, *J. Appl. Polym. Sci.*, **68**, 1493–1501 (1998).
103. Y. Feng, J.N. Hay, *Polymer*, **39**(26), 6723–6731 (1998).

Thermogravimetry of Polymers

Duncan M. Price and Douglas J. Hourston
IPTME, Loughborough University, Loughborough, LE11 3TU, UK

Fabrice Dumont
LSEO-UMR 5632, 6 Bvd. Gabriel, 21000 Dijon, France

1 Introduction 1
2 Instrumentation and Operation 2
3 Applications 3
3.1 Thermal Stability Assessment and Compositional Analysis 3
3.2 Sample-controlled Thermogravimetry 5
3.3 Lifetime Prediction and Degradation Kinetics 6
3.4 Combined Techniques 8
Acknowledgments 9
List of Symbols 10
Abbreviations and Acronyms 10
Related Articles 10
References 10

Thermogravimetry (TG) is the study of the relationship between a sample's mass and its temperature. It can be used to study any physical (such as evaporation) or chemical process (such as thermal degradation) that causes a material to lose volatile gases. Polymers have different thermal stabilities and thus the qualitative ''fingerprint'' afforded by TG in terms of temperature range, extent and kinetics of decomposition provides a rapid means to distinguish one polymer from another using only milligram quantities of material. Experiments are most commonly carried out under conditions where the temperature is increased in a linear fashion with time or the sample is held isothermally at an elevated temperature, although more sophisticated temperature profiles are occasionally used for compositional and kinetic analysis. Processes which do not result in a change in sample mass are not detected by TG. Therefore simultaneous measurements by differential scanning calorimetry (DSC) are useful. Volatile decomposition products may be detected and identified (e.g. by infrared (IR) spectrometry or mass spectrometry (MS)) in order to elucidate the mechanism of mass changes. TG is used for quantitative compositional analysis of polymers, lifetime prediction and kinetic studies, making the technique invaluable in all stages of polymer development, fabrication and component testing.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Mass and heat are amongst the oldest concepts known to mankind. However, only in recent times has the utility of recording mass change as a function of temperature been recognized. Many of the earliest experiments using TG were carried out by heating samples to known temperatures and removing them for weighing at regular time intervals.[1]

The concept of weighing a sample continuously while it was being heated appears to have been first used by Urbain and Boulanger in 1912,[2] who constructed an apparatus consisting of a conventional balance adapted for null-point electromagnetic compensation, with the sample hanging from the balance arm into an electrically heated furnace. The instrument was gas tight and had provision for the introduction of absorbent materials for evolved gases. Although this instrument was probably the first thermobalance, it appears not to have received further development, thus it is usually Honda who is credited with the invention of TG in 1915.[3] His equipment was reliable and allowed the heating rate to be reduced during the occurrence of mass losses (sample-controlled thermogravimetry – SCTG).

Much of the early work in this field was to use TG for inorganic chemical analysis. These efforts reached their zenith in Duval's monumental description of over 1000 gravimetric precipitates of nearly 80 elements.[4] In 1949, Jellinek[5] studied the degradation in vacuo of several important polymers including polystyrene and polyethylene. A review by Chiu[6] in 1966 still remains relevant today and covers most of the applications of TG to the study of polymers and predicts its coupling to other techniques such as IR spectrometry and MS for the study of evolved gases.

There is considerable literature dealing with TG and its applications to the characterization of polymers. Almost every scientific journal covering polymer science includes examples of the use of the technique, and two journals, the *Journal of Thermal Analysis & Calorimetry* and *Thermochimica Acta*, are devoted exclusively to thermal analysis. Together, these journals will often describe the latest advances in techniques and instrumentation. There are also a number of textbooks on thermal analysis[7–11] and a two-volume text edited by Turi[12] deals explicitly with the thermal characterization of polymeric materials. Unlike many other publications, the latter is organized by polymer type and applications rather than by techniques. This approach is useful in the context

of presenting an overview at the expense of occasional repetition.

In view of the wealth of published information regarding the applications of TG stretching back over 40 years of study this article can only hope to present an overview of the technique and its applications and the reader is referred to the numerous references and more specialist texts[13] for a fuller coverage of the subject.

2 INSTRUMENTATION AND OPERATION

Gallagher has recently reviewed the design, construction and operation of modern thermobalances.[14] A typical instrument consists of three major components:

1. a sensitive recording balance;
2. a furnace and associated controller/atmosphere management;
3. A data station, recorder, plotter and any other peripheral equipment.

Null-point balances are now used in almost all commercially available thermobalances. Many instruments used the electromagnetic compensation design described by Cahn and Shulz.[15] A beam carrying the sample and counterweight is suspended from the coil of a galvanometer. As the mass of the sample changes, the beam turns on a pivot, causing a variation in the current from a photodetector–shutter–lamp arrangement. This is used in a servoloop to apply a restoring force through to the coil. The change in current (or voltage) required to maintain equilibrium is proportional to the mass change in the sample. This analog signal is amplified and digitized to send to the data station. Typical sample masses range between 1 and 100 mg, although obtaining a truly representative sample of the material under investigation may be difficult as the sample mass decreases. Specialized thermobalances designed to accommodate larger mass ranges have been described.[14]

The furnace is the heart of the thermobalance. Whereas most instruments use a balance mechanism of a standard design, the furnace is often designed for a specific temperature range or response time. Subambient operation is of limited utility for most applications (with the exception of instruments designed for simultaneous thermogravimetry/differential scanning calorimetry (TG/DSC)), although many furnaces have the facility for air or water cooling (by means of a jacket) so that they may be cooled rapidly at the end of an experiment in order to increase sample throughput. Furnaces generally employ noninductively wound electrical resistance elements, although IR and microwave heating have been described. Low thermal mass furnaces are preferred for variable heating rate techniques such as SCTG or modulated-temperature thermogravimetry (mt-TG) where rapid changes in heating rate are required. Typical heating rates range from 50 °C min^{-1} to isothermal operation.

Provision must be made for some means of purging the furnace with a controlled atmosphere. Air or nitrogen are the most common purge gases employed. Operation under an inert atmosphere (such as nitrogen) may not always be possible since the furnace may not be completely airtight and may also retain trapped oxygen. Even high-purity nitrogen may not be completely oxygen-free. This is not, however, a major concern for polymers since they are generally less sensitive to oxidation than other materials commonly examined by TG. Helium may be used for good heat transfer between the furnace and sample particularly for mt-TG. Flow rates are controlled by rotameters or mass flow controllers and there is often some provision for switching between gases for compositional investigations. High-pressure and vacuum thermobalances are sometimes used.[14]

Several different arrangements of the balance and furnace are possible. The sample may hang down from the balance into the furnace, or a top-loading design may be used. Horizontal configurations may also be used in order to reduce the gas flow affecting the apparent weight of the sample during heating by impinging on the sample and thus acting as a piston. Alternatively, the inlet and outlet lines for the purge gas maybe arranged to direct the flow of gas across the sample rather than up or down a vertical furnace. The density of the gas also decreases with increasing temperature; this reduces the upthrust on the sample (Archimedes' Principle) leading to an apparent increase in weight. The gas flow path may also change with temperature. Using small samples (<10 mg) or performing a blank measurement with a similar volume of inert material can reduce these effects.[16] An alternative approach is to use a symmetrical design whereby the sample holder and counterweight both hang down into the same or separate furnaces. Automatic sample changers are available for several commercial thermobalances, thus permitting unattended operation and increased productivity.

Samples are usually contained within small crucibles made of refractory materials, such as quartz or platinum, which are chemically unreactive towards most materials. Less expensive, disposable aluminum holders may be used for polymers since the latter usually decompose below the melting temperature of aluminum (660 °C). In all cases, it is important to allow good interaction between the sample and purge gas so that the build-up of decomposition products in the immediate vicinity of the sample is avoided. Purge gas flow rates should be

sufficient to sweep volatiles from the furnace, but not so fast as to cool the sample.

The temperature of the sample is usually measured by a thermocouple placed nearby. It is unreliable to use the furnace temperature as a measure of sample temperature since this will inevitably be in advance of the sample's true temperature owing to heat transfer considerations. In the event of a chemical reaction occurring, the sample temperature will be affected by the heat of reaction and may be above the furnace temperature for exothermic decompositions. For simultaneous TG/DSC and in some conventional thermobalances, the sample holder sits directly upon a temperature sensor. In other cases it is important to locate the thermocouple as close as possible to the sample without interfering with the operation of the balance. Temperature calibration should be carried out regularly, according to a recommended procedure, and always under the same conditions (e.g. heating rate, purge gas, thermocouple location) as that used for measurements. Calibration of the mass scale of the furnace is less frequently required, but easily checked using standard masses.

Modern thermobalances typically use a personal microcomputer for data acquisition. This allows results to be stored for future reference and subsequent data analysis. It is usual to be able to plot the sample's mass (directly or as a percentage of its initial mass) as a function of time or temperature.[13] Calculation and display of the first derivative of sample mass as a function of time or temperature (dm/dt or $-dm/dT$) is a useful facility in identifying regions of interest.[6] It is common to plot $-dm/dt$ or $-dm/dT$ so that maxima in rate of mass loss appear as peaks. Analysis of such data ranges from simply determining the magnitude and the extrapolated onset of mass changes to sophisticated kinetic modelling of processes although it is essential to have a thorough understanding of any mathematical treatment of data before any reliance is placed on the results. Many computer data stations can often be used to control several thermal analyzers at one time and co-plot the results from different experiments and/or techniques.

Finally, it is important to report all results with sufficient experimental detail (e.g. sample history, measurement conditions, equipment and data treatment) so as to enable their repetition and, if necessary, reconcile any observed differences in outcome with changes in operational procedure.

3 APPLICATIONS

The information provided by TG is confined to the detection of changes in mass of the sample as its temperature is altered. Thus the technique is largely limited to the study of decomposition and oxidation reactions and to such physical processes as vaporization, sublimation and desorption.

3.1 Thermal Stability Assessment and Compositional Analysis

The assessment of thermal stability is one of the most important applications of TG to the study of polymers. Thermogravimetric curves provide information about the decomposition mechanisms for various materials.[17] In addition, the decomposition profiles may be characteristic for each type of polymer and in some cases can be used for identification purposes. The onset of mass loss often defines the upper limit of thermal stability for the material, though it must be appreciated that extensive degradation of the polymer structure by, for example, cross-linking, may have already taken place before the point at which detectable changes in mass occur.

The routes by which polymers degrade can be categorized according to six main mechanisms:[18]

1. main-chain scission
2. side group scission
3. elimination
4. depolymerization
5. cyclization
6. cross-linking.

Cyclization and cross-linking rarely result in any change of sample mass unless they occur in conjunction with 1–4. and are not detected by TG.[17] Routes 1–4 usually result in the evolution of volatile products with an accompanying mass change. In an inert atmosphere, some polymers give an almost quantitative yield of their parent monomers.[19] In air, complete oxidation of the sample to oxides of its constituent elements commonly occurs. Nitrogen-containing polymers usually generate some ammonia or hydrogen cyanide. Halogen-containing polymers yield the respective hydrogen halides.[19]

Figure 1 shows a series of thermogravimetric curves for a number of common polymers in nitrogen. When comparing the thermal stability of polymers it must be noted that such curves are purely procedural. For meaningful correlation of thermal stability with polymer structure it is essential that the experiments are carried out under similar experimental conditions. It is usual to assign a temperature at which degradation begins to occur (e.g. the extrapolated onset of the lowest temperature weight loss) or quote the temperature at which (for example) 5% weight loss has taken place. Often, significant deterioration in polymer properties has occurred below these values.

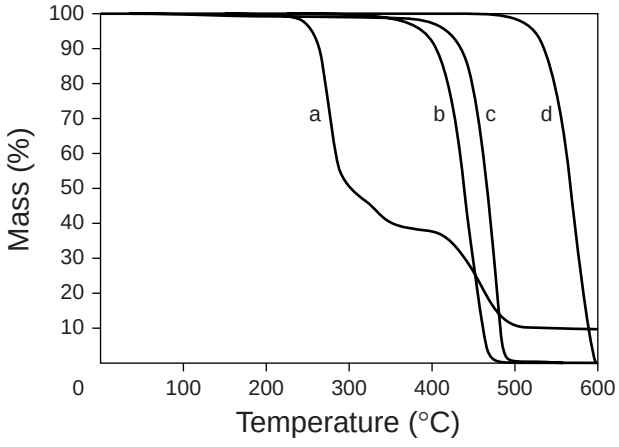

Figure 1 TG curves for several common polymers in nitrogen heated at 5 °C min^{-1} (a = poly(vinyl chloride), b = nylon-6, c = low-density polyethylene, d = polytetrafluoroethylene).

The information provided by TG is inherently quantitative. Provided that the temperature scale is accurately calibrated and the balance provides an output proportional to the mass of the sample across its operating range, then the mass loss profile of a mixture of materials is usually the sum of the individual profiles of each of its components.[13] This affords a means of compositional analysis of materials such as polymer blends and composites.[13] An example of this is shown in Figure 2 for carbon fiber/epoxy composite. In this case, the sample was heated to 825 °C in an inert atmosphere so that degradation of the epoxy resin took place. Then the furnace was purged with air so that oxidation of the carbon fiber took place. The percentage composition of each component in the specimen can be determined after due allowance has been made for any carbonaceous residue arising from degradation of the epoxy resin.

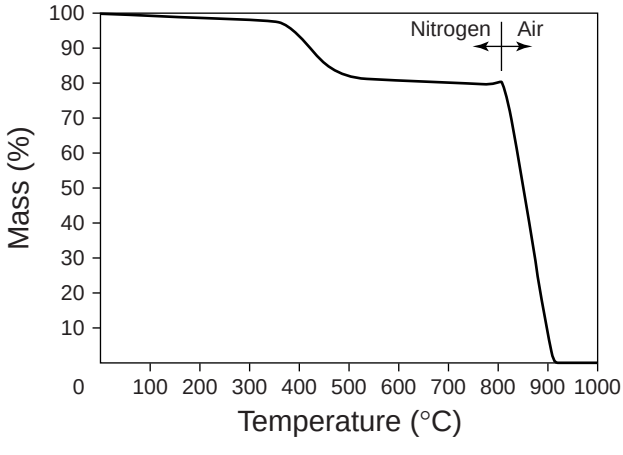

Figure 2 TG curve of carbon fiber/epoxy resin composite heated in nitrogen to 825 °C, then exposed to air.

The degradation of block and graft copolymers rarely corresponds to the sum of their parent homopolymers.[12] Copolymerization may result in a material which is thermally less stable than either of the parent homopolymers since the addition of a heterogeneity in the polymer backbone may incorporate a ''weak link'' into the macromolecular chain. Alternatively, there are a number of co-monomers (such as alkyl acrylates or acrylonitrile) which have been used to stabilize polymers sensitive to depolymerization during processing (such as poly(methyl methacrylate)).[19]

An example of increasing vinyl acetate content on the thermal degradation of poly(ethylene-co-vinyl acetate) is shown in Figure 3. The vinyl acetate component shows a characteristic mass loss associated with elimination of acetic acid before the main chain degradation occurs. In this case TG can be used as a quantitative means of estimating copolymer composition.

Another quantitative use of TG is to study the loss of solvents from polymer dispersions.[20] Such measurements are often performed under isothermal operation and the initial stages of the process often show a linear mass loss profile with time as the solvent evaporates since this is a zero order process. During later stages of drying the rate of mass loss may decrease as the residual solvent becomes trapped in the matrix and its diffusion out of the sample is constrained. For many plasticized polymers, there may be a sudden mass loss as the material is heated through the sample's glass–rubber transition temperature, and the molecular mobility of the polymer increases so allowing the plasticizer to escape. In any drying study, the operator should be aware that the sample may be significantly cooler than the furnace temperature owing to the latent heat of evaporation of the solvent. A study by Mrklić and Kovačić[21] describes

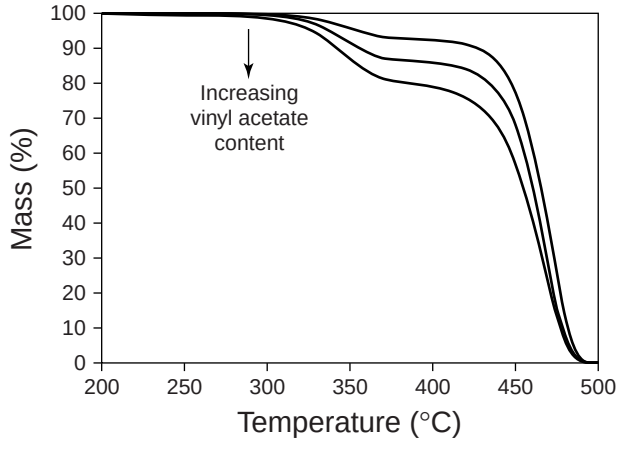

Figure 3 TG curves of poly(ethylene-co-vinyl acetate) copolymers of different vinyl acetate content (top to bottom = 9, 18 and 28 mol.% vinyl acetate).

the use of TG to investigate the volatilization of dioctyl phthalate from plasticized poly(vinyl chloride). Such work is useful in predicting the loss of additives from polymer systems, which has important implications in the lifetime of the formulated product.

3.2 Sample-controlled Thermogravimetry

In many cases the mass loss profiles of polymers are complex and consist of several overlapping steps. One way of overcoming this is to decrease the heating rate so as to achieve more complete separation of each stage of the decomposition process. Such an approach naturally increases the time for the experiment to be carried out. An alternative approach can be used whereby instead of the usual linear temperature ramp a strategy is adopted whereby the rate of rise of temperature is slowed or even suspended as soon as some predetermined rate of mass loss is detected.[22] Once this has occurred, heating is recommenced until a further mass change is detected. This technique is known as sample-controlled TG.

In its simplest implementation, the temperature program can be arranged to alternate between linear heating at a constant rate of temperature rise interspersed by isothermal segments when mass loss is occurring. Sørensen[23] has termed this ''step-wise isothermal'' heating. The advantage with this approach is that the temperature profile can be recorded and used to specify conditions for future experiments on similar samples (for use in quality control). Another approach is to use a dynamic heating rate[24,25] whereby the rate of temperature change is gradually reduced from an initial rate (possibly even resulting in the sample being cooled) as the rate of mass loss increases. This approach has been commercialised as ''high resolution TG''. A review by Reading[26] uses the term ''constrained rate TG'' for this approach.

Figures 4 and 5 show the mass fraction and derivative mass loss curves for a bilayer low density polyethylene/nylon-6 film. Figure 4 was obtained at a constant linear rising temperature profile of 5 °C/min whereas Figure 5 was obtained using the variable heating rate method described by Gill et al.[24] Although there is still incomplete separation of the degradation of the two components in Figure 5, measurement of the mass change between the plateaux either side of the saddle in the derivative mass loss curves allows quantitative analysis of each of the components, which agrees well with theoretical calculations based upon the thickness of each layer. The penalty of this particular approach is the extended experimental time needed to obtain such data indicated by the curves shown in Figure 6.

A final method of temperature control is to dispense with any preconceptions as to the initial choice of temperature alteration, but to arrange a feedback directly from the rate of change of sample mass so as to maintain a constant rate of mass change (or indirectly, by detecting

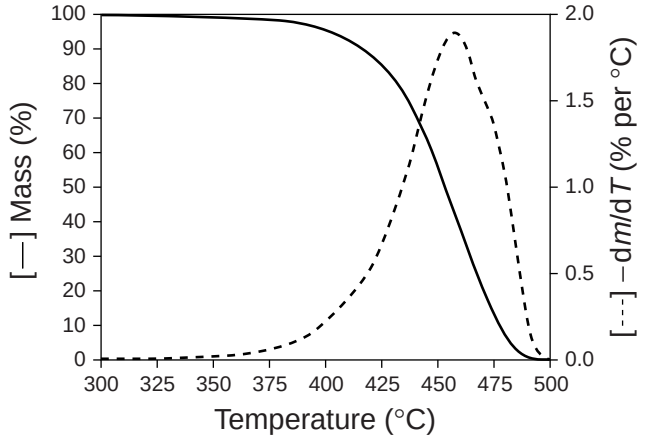

Figure 4 TG curve of low density polyethylene/nylon-6 laminate obtained at a constant heating rate (5 °C min^{-1}).

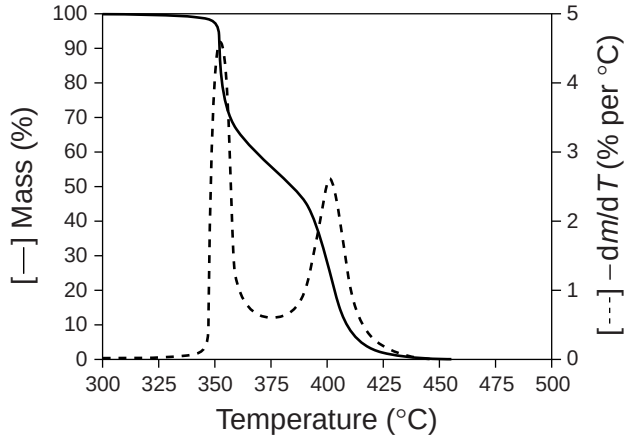

Figure 5 TG curve of low-density polyethylene/nylon-6 laminate obtained at a dynamic heating rate.

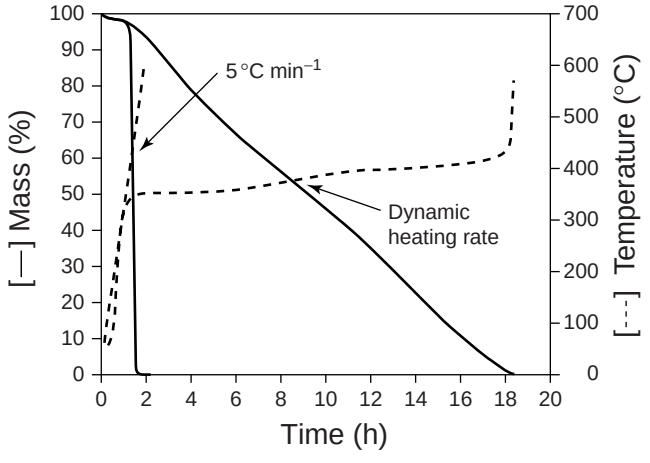

Figure 6 Comparison of Figures 4 and 5 plotted against time.

the rate of evolution of gaseous decomposition products). This area of study has been independently pioneered by Rouquerol[27] and Paulik and Paulik[28] and such approaches are applicable to other thermal methods.

3.3 Lifetime Prediction and Degradation Kinetics

The most obvious use of TG beyond that of thermal stability assessment and compositional analysis is to use the technique to predict the lifetime of polymers under actual service conditions. The simplest approach is to perform isothermal measurements at elevated temperatures and measure the time taken for a certain extent of mass loss to occur. Several experiments may be carried out at different temperatures so as to obtain a table of lifetime versus temperature. Whilst such data are readily obtained, it rarely spans the temperature range of interest (indeed to do so may require inordinately long experiments at lower temperatures), therefore some means of extrapolation is required.

The temperature dependence of chemical processes may be readily expressed in terms of Equation (1), the Arrhenius equation:[29]

$$k = A \exp\left(\frac{-E_a}{RT}\right) \tag{1}$$

where k is the rate constant, R the gas constant and T the thermodynamic (Kelvin) temperature. Values of the Arrhenius parameters (E_a and A) provide measures of the magnitude of the energy barrier to reaction (the activation energy, E_a) and the frequency of the occurrence of a condition that may lead to reaction (the frequency factor, A). The rate constant k is defined by the relationship between the rate of reaction ($d\alpha/dt$) and the extent of conversion or fraction reacted (α). Polymer decompositions are generally heterogeneous reactions since the sample is solid (or molten) and the products are gases. A general relation to describe the relationship between ($d\alpha/dt$) and α is Equation (2), the Ng equation:[30]

$$\frac{d\alpha}{dt} = k\alpha^m(1-\alpha)^n \tag{2}$$

Starting from this basic form, it is possible to derive various subclasses of rate equation such as first-order decay, nucleation and growth, etc. by changing the values of m and n. Isothermal experiments provide the means of determining the form of the kinetic equation though discrimination between different models is not straightforward.[31] A simpler approach is to substitute the reciprocal of the isothermal lifetime for the rate constant in Equation (1) and extrapolate the data to the region of interest. Caution should be placed on the interpretation of such data, especially at lower temperatures where the degradation mechanism may have changed. This is especially important if the original measurements were made above a phase transition e.g. the polymer's melting point or its glass transition temperature.[32]

Isothermal measurements suffer from the drawback of being rather time-consuming. There are also difficulties in bringing the sample and apparatus to the required temperature without some decomposition of the sample having already taken place. One way of avoiding this problem is to change a rate-controlling variable during the course of the measurement. A useful method is temperature jump TG.[33] The rate of decomposition can then be measured either side of the change in temperature and the activation energy determined from Equation (3)

$$E_a = R \ln \frac{(dm/dt)_{T_1}}{(dm/dt)_{T_2}} \quad \frac{1}{T_1} - \frac{1}{T_2} \tag{3}$$

where $(dm/dt)_{T_1}$ and $(dm/dt)_{T_2}$ are the rates of mass loss at T_1 and T_2 either side of the temperature jump. SCTG has also been employed to investigate the kinetics of thermal decomposition of polyesters and it is claimed that the technique offers significant advantages when dealing with polymers.[34]

Another approach is to studying degradation kinetics using data from conventional linear rising temperature TG. Many such methods have been proposed, but the most popular strategy is that described by Ozawa,[35] and Flynn and Wall[36] which has been incorporated into an American Society for Testing and Materials standard method.[37] Essentially, separate measurements are carried out at different linear heating rates and the temperatures at which a set percentage mass loss occurs noted (Figure 7). These are then plotted as a function of heating rate (dT/dt) and the activation energy determined by an iterative process (Figure 8). These algorithms have been

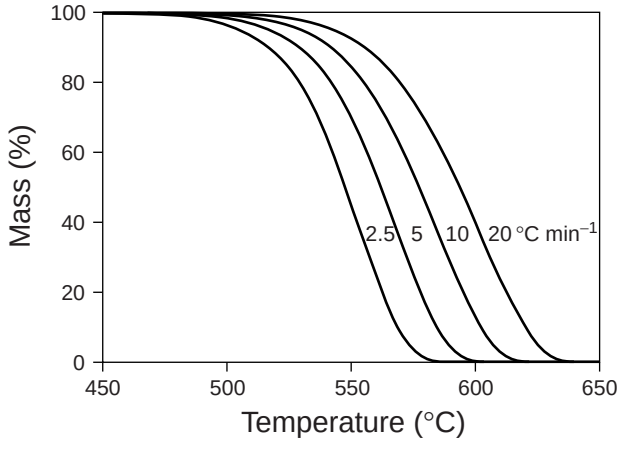

Figure 7 TG curves of poly(tetrafluoroethylene) at different heating rates.

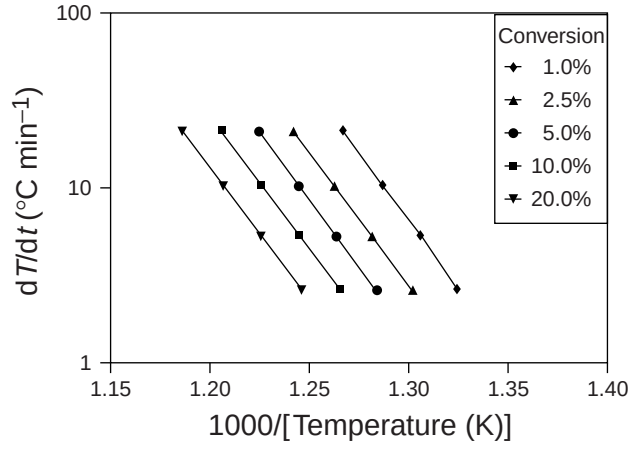

Figure 8 Kinetic analysis of results from Figure 7 to determine E_a.

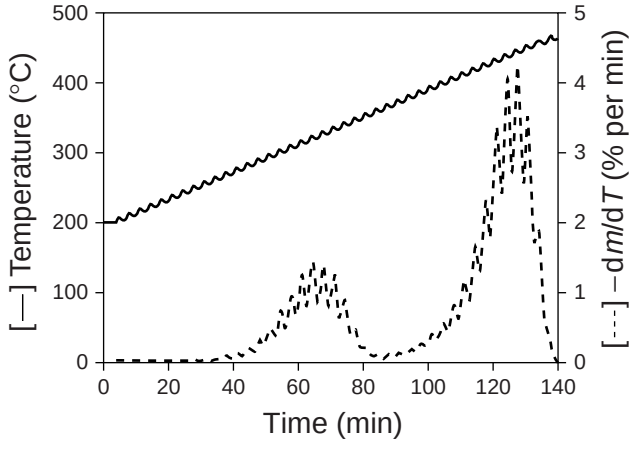

Figure 10 mt-TG curve of poly(ethylene-co-vinyl acetate).

incorporated into a number of commercially available software packages,[38] although the user should always question the predictions of such ''black box'' methods especially since they often assume, without appropriate justification, that the polymer decomposition reactions are ''first-order'' processes, as in Equation (4):

$$\frac{d\alpha}{dt} = k(1-\alpha) \tag{4}$$

Thus, it is possible to predict the lifetime for (say 10% mass loss) as a function of temperature (Figure 9).

A final method for studying polymer degradation kinetics is known as modulated temperature TG. This uses a temperature profile in which a sinusoidal temperature fluctuation is superimposed upon a conventional linear rising temperature program.[39] The raw data from such an experiment are shown in Figure 10 for an ethylene-co-vinyl acetate copolymer. The curves showing the heating rate (dT/dt) and rate of mass loss (−dm/dt) make the effect of the temperature modulations apparent. E_a for the chemical decomposition is then calculated according to Equation 5:

$$E_a = \frac{R[T_{av}^2 - (0.5T_{amp})^2]L}{T_{amp}} \tag{5}$$

where T_{av} is the average thermodynamic temperature, T_{amp} is the amplitude of the temperature modulation and L is the logarithm of the amplitude of the rate of mass loss over one modulation. Plots of mass and E_a as a function of temperature from the data is shown in Figure 11. Although based on the temperature jump method described above, mt-TG is a relatively new technique which promises to simplify the acquisition of kinetic data.

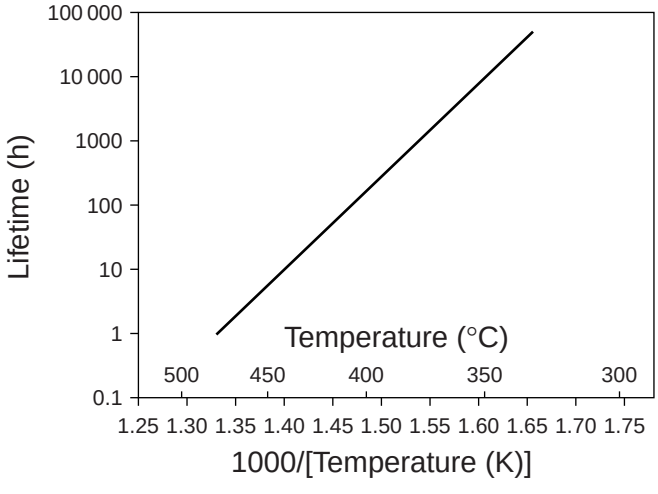

Figure 9 Predicted lifetime (10% weight loss) of poly(tetrafluoroethylene) vs temperature.

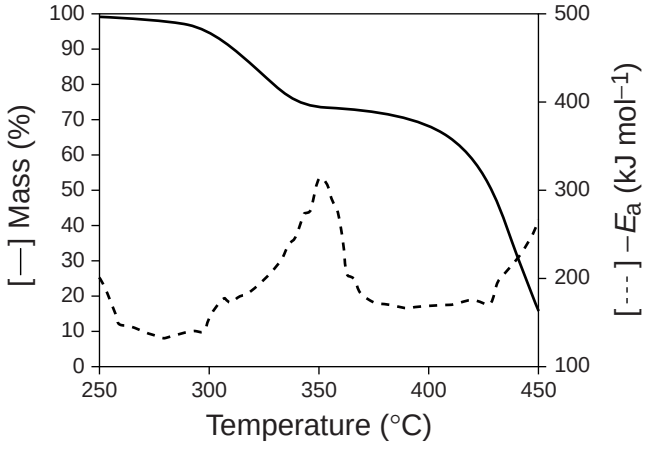

Figure 11 Plot of mass and E_a versus temperature for data from Figure 10.

The kinetic analysis of thermogravimetric data is a veritable ''minefield''[40] and the reader is referred to the excellent review of this area by Galwey and Brown[41] for a more in-depth discussion.

3.4 Combined Techniques

3.4.1 Thermogravimetry/Differential Scanning Calorimetry

DSC is another thermal analysis technique whereby the difference in heat flow between a sample and inert reference substance is measured during a temperature program. Any changes in heat capacity of the sample (such as accompanies the glass–rubber transition in polymers) or processes which consume (endothermic) or liberate (exothermic) heat can be studied. The sample is usually placed in an aluminum pan, with an empty pan used as a reference. Various methods are used to measure the heat flux, but a common arrangement is to place the pans on a metal or ceramic plate containing thermocouples which serve to detect the difference in temperature between the pans and convert this to a thermal energy difference. Several instruments have been designed which incorporate the heat flux plate design into a thermobalance this affording a means of performing simultaneous DSC and TG.[42] There are several advantages to this approach over conventional DSC and TG notwithstanding the saving in experimental time needed to acquire two sets of data. It is useful to be able to determine energy changes associated with thermal decomposition reactions. Because many chemical reactions in polymer systems, such as curing of thermosets, are condensation reactions, water or ammonia may be liberated during the course of the reaction. This will cause a change in mass of the sample in addition to the heat of reaction which can be measured by DSC.[42] TG cannot normally detect physical changes in polymer structure such as those that accompany the glass transition, crystallization or melting so the additional information provided by DSC is useful. Another example of TG/DSC is shown in Figure 12 for a natural polymer (cellulose) containing an excess of water. The total moisture content of the polymer can be found from the mass loss on drying whereas the amount of free water can be determined from its enthalpy of fusion given by the area under the DSC peak around 0 °C. The difference between the two values is the amount of water chemically bound to the polymer. This will vary for cellulose obtained from different sources and will influence its reactivity towards acids and alkalis. This can be important in the preparation of cellulose derivatives such as cellulose acetate.

Differential thermal analysis (DTA) is another variant of DSC whereby the temperature difference between a sample and reference is monitored during the temperature program. This technique is not quantitative in terms of determining energy changes although it is capable of detecting the occurrence and energetic nature of physical and chemical transformations in the same way as DSC. Kettrup et al.[43] have described a simultaneous DTA–thermobalance capable of handling up to 500 g of material for the study of waste products for environmental purposes. Such an instrument has obvious applications for the recycling of plastics and is also capable of being interfaced to a mass spectrometer for analysis of evolved gases.

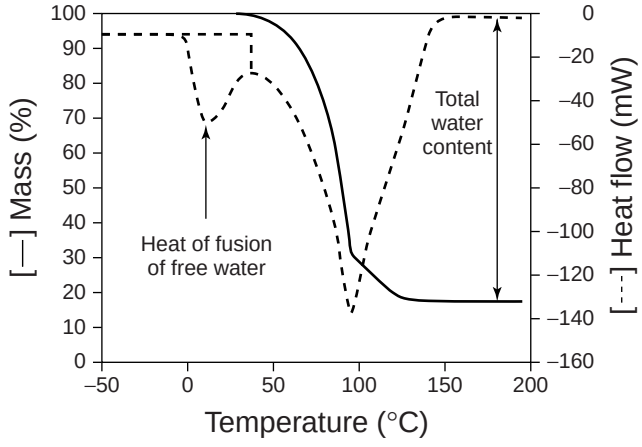

Figure 12 TG/DSC curves for cellulose containing an excess of water.

3.4.2 Thermogravimetry/Evolved Gas Analysis

The capability of TG for materials characterization is greatly increased if other techniques are coupled to the thermobalance in order to identify either the residue or the products evolved during the experiment. Although Wiedemann and Bayer[44] have described an instrument for simultaneous TG with the X-ray diffraction analysis of the residue, it is more common to identify the gaseous degradation products from TG. Nonspecific methods such as coupled TG–photometry[17] afford a means of measuring smoke generation during the degradation of polymers, whereas online titration of evolved gases (e.g. by absorption in dilute alkali solution) is more selective for acidic evolved gases such as hydrogen chloride from the thermal degradation of poly(vinyl chloride)[17] or acetic acid from the thermolysis of vinyl acetate copolymers.[45]

The most usual means of identification of evolved gases employs either IR or MS.[46] In both cases, some form of interface is required to take the evolved gases from the thermobalance into the spectrometer. For thermogravimetry/infrared evolved gas analysis (TG/IR) this comprises a heated transfer line (to prevent the

condensation of less volatile products) and a flow cell through which the beam from the spectrometer is passed. Materazzi[47] has published a review of publications up to 1997 which describes the apparatus and applications of TG/IR. An example of this is shown in Figure 13 which shows the mass loss and derivative mass loss profiles for plasticized cellulose acetate.[48] The total signal from the IR detector is shown in Figure 14 along with a signal derived from the IR absorbance of the plasticizer. Thus, IR evolved gas analysis is a useful means of assigning multiple mass losses to specific degradation mechanisms.

MS is an alternative to IR analysis although the interface design is complicated by the requirement to operate the mass spectrometer under high vacuum. Various splitter designs[49] have been developed in order to reduce the transfer line pressure down to a level suitable for injection into the mass spectrometer. Again,

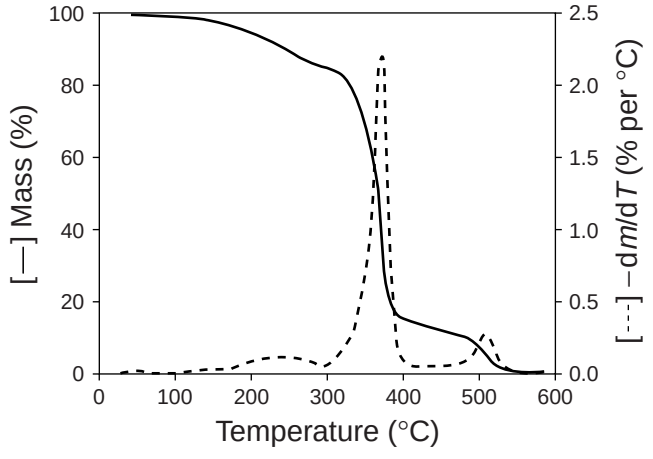

Figure 13 TG/IR curves of cellulose acetate plasticized with dioctyl phthalate.

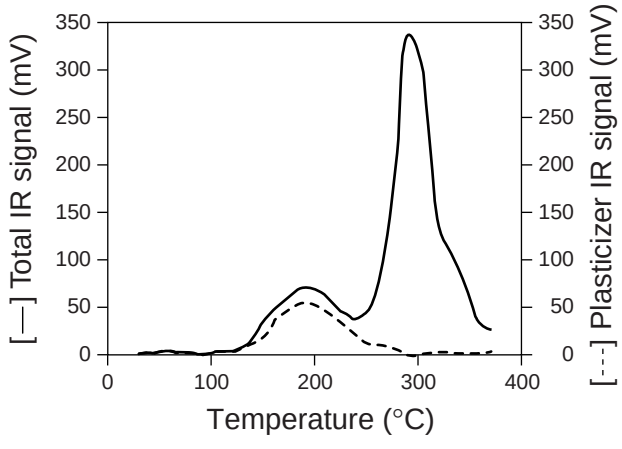

Figure 14 Total IR detector signal and signal due to dioctyl phthalate of plasticized cellulose acetate.

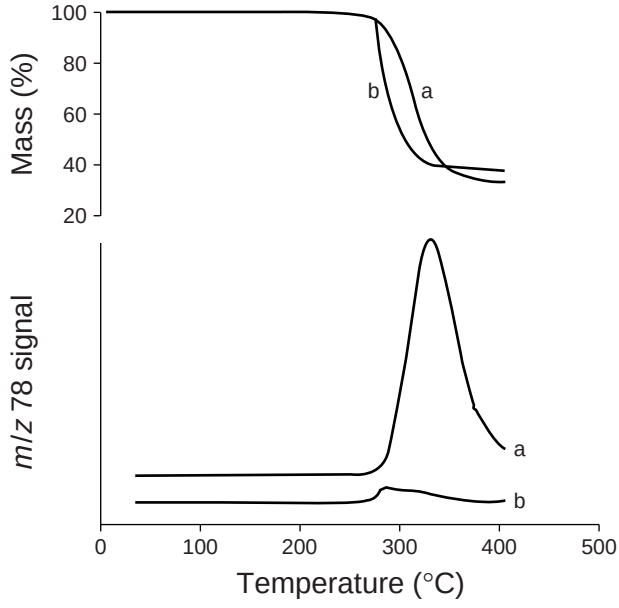

Figure 15 TG/MS curves of poly(vinyl chloride) showing the effect of 4% added MoO_3 on the release of benzene ($m/z = 78$) (a = pure polymer, b = polymer + additive).

these units are heated so as to prevent condensation of less volatile products. Materazzi[50] has also reviewed this area of research to 1998, and Jansen[51] provides an overview of both evolved gas analysis techniques with particular reference to polymers. A simple example of the application of thermogravimetry/mass spectrometry evolved gas analysis (TG/MS) to polymers is given in Figure 15 which shows the effect of adding molybdenum trioxide to poly(vinyl chloride). The reduction in the emission of benzene ($m/z = 78$) brought about by the incorporation of the additive is clearly evident.

Although both IR and MS are valuable techniques for the identification of evolved gases, in many cases a mixture of decomposition products are evolved simultaneously. Several workers have added the capability to perform online[52] or offline[53] gas chromatography to separate the effluent into individual components which can then be identified by MS. Such approaches essentially attach an analytical laboratory to the end of the humble thermobalance and make this an appropriate point at which to conclude this review of the applications of TG of polymers.

ACKNOWLEDGMENTS

The authors wish to thank Mr S.Y. Ng and Dr S.B. Warrington for kindly providing the data used in Figures 9 and 10, and 14, respectively. We also wish to dedicate this

article to the memory of Dr C.J. Keattch (1928–1999), author of Reference 1 and the first secretary of the Royal Society of Chemistry Thermal Methods Group (1965–1999).

LIST OF SYMBOLS

α	extent of conversion or fraction reacted
A	frequency factor
$d\alpha/dt$	rate of reaction
dm/dt	first derivative of mass change with time
dm/dT	first derivative of mass change with temperature
dT/dt	first derivative of temperature with time (i.e. heating rate)
E_a	activation energy
k	rate constant
L	logarithm of the amplitude of the rate of mass loss over one modulation (mt-TG)
m/z	mass to charge ratio
R	ideal gas constant
T	thermodynamic (Kelvin) temperature
T_{av}	average thermodynamic temperature (mt-TG)
T_{amp}	amplitude of temperature modulation (mt-TG)

ABBREVIATIONS AND ACRONYMS

DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
IR	Infrared
MS	Mass Spectrometry
mt-TG	Modulated-temperature Thermogravimetry
SCTG	Sample-controlled Thermogravimetry
TG	Thermogravimetry
TG/DSC	Thermogravimetry/Differential Scanning Calorimetry
TG/IR	Thermogravimetry/Infrared Evolved Gas Analysis
TG/MS	Thermogravimetry/Mass Spectrometry Evolved Gas Analysis

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Polymers and Rubbers cont'd **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Degradation • Pyrolysis Techniques in the Analysis of Polymers and Rubbers

Thermal Analysis **(Volume 15)**
Simultaneous Techniques in Thermal Analysis • Thermogravimetry

REFERENCES

1. C.J. Keattch, *Introduction to Thermogravimetry*, Heyden, London, 1969.
2. G. Urbain, Ch. Boulanger, 'Sur une Balance-Laboratoire à Compensation Électromagnétique Destinée à L'étude des Systèmes qui Dégagent des Gaz avec une Vitesse Sensible', *Compt. Rend.*, **154**, 347–349 (1912).
3. K. Honda, 'On a Thermobalance', *Sci. Rep. Tohuku Univ.*, **4**, 97–103 (1915).
4. C. Duval, *Inorganic Thermogravimetric Analysis*, 2nd edition, Elsevier, 1963.
5. H.H.G. Jellinek, 'Thermal Degradation of Polystyrene and Polyethylene. Part III', *J. Polym. Sci.*, **4**, 13–36 (1949).
6. J. Chiu, 'Applications of Thermogravimetry to the Study of High Polymers', *Appl. Polym. Symp.*, **2**, 25–43 (1966).
7. J.W. Dodd, K.H. Tonge, *Thermal Methods: Analytical Chemistry by Open Learning*, John Wiley & Sons, Chicester, 1987.
8. M.E. Brown, *Introduction to Thermal Analysis*, Kluwer Academic, Dordrecht, 1988.
9. E.L. Charsley, S.B. Warrington, *Thermal Analysis – Techniques & Applications*, The Royal Society of Chemistry, Cambridge, 1992.
10. V.B.F. Mathot, *Calorimetry and Thermal Analysis of Polymers*, Hanser, Munich, 1994.
11. P.J. Haines, *Thermal Methods of Analysis. Principles, Applications and Problems*, Blackie, Glasgow, 1995.
12. E.A. Turi (ed.), *Thermal Characterization of Polymeric Materials*, 2nd edition, Academic Press, San Diego, Vols 1 and 2, 1997.
13. C.M. Ernest (ed.), *Compositional Analysis by Thermogravimetry, ASTM STP 997*, American Society for Testing and Materials, Philadelphia, PA, 1998.
14. P.K. Gallagher, 'Thermogravimetry and Thermomagnetometry', in *Handbook of Thermal Analysis and Calorimetry*, ed. M.E. Brown, Elsevier Science B. V., Amsterdam, Vol. 1, 225–278, 1998.

15. L. Cahn, H.E. Shultz, 'The Cahn Gram Electrobalance', in *Vacuum Microbalance Techniques*, ed. R.F. Walker, Plenum Press, New York, Vol. 2, 7–18, 1962.
16. P.J. Haines, 'Thermogravimetry', in *Thermal Methods of Analysis. Principles, Applications and Problems*, ed. P.J. Haines, Blackie, Glasgow, 22–62, 1995.
17. J. Chiu, 'Thermogravimetry for Chemical Analysis of Polymers', in *Applied Polymer Analysis and Characterization: Recent Developments in Techniques, Instrumentation, Problem Solving*, ed. J. Mitchell, Jr, Hanser, Munich, 175–212, 1987.
18. W. Schnabel, *Polymer Degradation, Principles and Practical Applications*, Hanser, Munich, 25–63, 1981.
19. N. Grassie, G. Scott, *Polymer Degradation and Stabilisation*, Cambridge University Press, Cambridge, 17–67, 1985.
20. D.M. Price, Courtaulds plc., Internal Report, unpublished.
21. Ž. Mrklić, T. Kovačić, 'Thermogravimetric Investigation of the Volatility of Dioctyl Phthalate from Plasticized Poly(vinyl chloride)', *Thermochim. Acta*, **332**, 129–135 (1998).
22. M. Reading, 'Controlled Rate Thermal Analysis and Related Techniques', in *Handbook of Thermal Analysis and Calorimetry*, ed. M.E. Brown, Principles and Practice, Elsevier Science B. V., Amsterdam, Vol. 1, 423–443, 1998.
23. O. Toft Sørensen, 'Quasi-isothermal Methods in Thermal Analysis', *Thermochim. Acta*, **50**, 163–175 (1981).
24. P.S. Gill, S.R. Sauerbrunn, B.S. Crowe, 'High Resolution Thermogravimetry', *J. Thermal Anal.*, **38**, 255–266 (1992).
25. R. Riesen, 'Adjustment of Heating Rate for Maximum Resolution in TG and TMA (MaxRes), *J. Therm. Anal.*, **53**, 365–374 (1998).
26. M. Reading, 'Controlled Rate Thermal Analysis and Beyond', in *Thermal Analysis – Techniques & Applications*, eds. E.L. Charsley, S.B. Warrington, The Royal Society of Chemistry, Cambridge, 127–155, 1992.
27. J. Rouquerol, 'Controlled Transformation Rate Thermal Analysis: The Hidden Face of Thermal Analysis', *Thermochim. Acta*, **144**, 209–224 (1989).
28. F. Paulik, J. Paulik, 'Thermoanalytical Examination Under Quasi-isothermal–Quasi-isobaric Conditions', *Thermochim. Acta*, **100**, 23–59 (1986).
29. K.J. Laidler, 'The Development of the Arrhenius Equation', *J. Chem. Educ.*, **61**(6), 491–498 (1984).
30. W.-L. Ng, 'Thermal Decomposition in the Solid State', *Aust. J. Chem.*, **28**, 1169–1178 (1975).
31. M.E. Brown, A.K. Galwey, 'The Distinguishability of Selected Kinetic Models for Isothermal Solid-state Reactions', *Thermochim. Acta*, **29**, 129–146 (1979).
32. J.H. Flynn, 'A Critique of Lifetime Prediction of Polymers by Thermal Analysis', *J. Therm. Anal.*, **44**, 499–512 (1995).
33. J.H. Flynn, B. Dickens, 'Steady-state Parameter-jump Methods and Relaxation Methods in Thermogravimetry', *Thermochim. Acta*, **15**, 1–16 (1976).
34. T. Arii, S. Ichinara, H. Nakagawa, N. Fujii, 'A Kinetic Study of the Thermal Decomposition of Polyesters by Controlled Rate Thermogravimetry', *Thermochim. Acta*, **319**, 139–149 (1998).
35. T. Ozawa, 'A New Method of Analyzing Thermogravimetric Data', *Bull. Chem. Soc. Jpn*, **38**, 1881–1886 (1965).
36. J.H. Flynn, L.A. Wall, 'A Quick, Direct Method for the Determination of Activation Energy from Thermogravimetric Data', *Polym. Lett.*, **4**, 323–328 (1966).
37. ASTM Test Method E1641, 'Standard Test Method for Decomposition Kinetics by Thermogravimetry', *ASTM Book of Standards 14.02*, American Society for Testing and Materials, 1042–1046, 1994.
38. J. Opfermann, E. Kaisersberger, 'An Advantageous Variant of the Ozawa–Flynn–Wall Analysis', *Thermochim. Acta*, **203**, 167–175 (1992).
39. R.L. Blaine, B.K. Hahn, 'Obtaining Kinetic Parameters by Modulated Temperature Thermogravimetry', *J. Therm. Anal.*, **54**, 695–704 (1998).
40. M.E. Brown, 'Steps in a Minefield: Some Kinetic Aspects of Thermal Analysis', *J. Therm. Anal.*, **49**, 17–32 (1997).
41. A.K. Galwey, M.E. Brown, 'Kinetic Background to Thermal Analysis and Calorimetry', in *Handbook of Thermal Analysis and Calorimetry*, ed. M.E. Brown, Elsevier Science B. V., Amsterdam, Vol. 1, 147–224, 1998.
42. E.L. Charsley, 'Complementary Thermal Analysis Techniques', in *Thermal Analysis – Techniques & Applications*, eds. E.L. Charsley, S.B. Warrington, The Royal Society of Chemistry, Cambridge, 59–83, 1992.
43. A. Kettrup, G. Matushek, H. Utschick, Ch. Namendorf, G. Bräuer, 'A Macro STA-system for Environmental Samples', *Thermochim. Acta*, **295**, 119–131 (1997).
44. H.G. Wiedemann, G. Bayer, 'Investigation of Minerals of Lunar Samples (14163, 14258) by Simultaneous Thermal and X-ray Analysis', *Z. Anal. Chem.*, **266**, 97–109 (1973).
45. S.G. Fischer, J. Chui, 'Analysis of Polymer Blends and Copolymers by Coupled Thermogravimetry and Automatic Titration', *Thermochim. Acta*, **65**, 9–17 (1983).
46. S.B. Warrington, 'Evolved Gas Analysis', in *Thermal Analysis – Techniques & Applications*, eds. E.L. Charsley, S.B. Warrington, The Royal Society of Chemistry, Cambridge, 84–107, 1992.
47. S. Materazzi, 'Thermogravimetry–Infrared Spectroscopy (TG–FTIR) Coupled Analysis', *Appl. Spectrosc. Rev.*, **32**(4), 385–404 (1997).
48. D.M. Price, S.P. Church, 'FTIR Evolved Gas Analysis of the Decomposition Products of Cellulose Diacetate', *Thermochim. Acta*, **294**, 107–112 (1997).
49. G. Szekely, M. Nebuloni, L.F. Zerilli, 'Thermal Analysis–Mass Spectrometry Coupling and its Applications', *Thermochim. Acta*, **196**, 511–532 (1992).
50. S. Materazzi, 'Mass Spectrometry Coupled to Thermogravimetry (TG-MS) for Evolved Gas Characterization: A Review', *Appl. Spectrosc. Rev.*, **33**(3), 189–218 (1998).
51. J.A.J. Jansen, 'Evolved Gas Analysis of Polymers by Coupled Gas Chromatography, Fourier Transform

Infrared Spectroscopy and Mass Spectroscopy', in *Calorimetry and Thermal Analysis of Polymers*, ed. V.B.F. Mathot, Hanser, Munich, 335–352, 1994.

52. W.H. McClennen, R.M. Buchanan, N.S. Arnold, J.P. Dworzanski, H.L.C. Meuzelaar, 'Thermogravimetry/Gas Chromatography/Mass Spectrometry and Thermogravimetry/Gas Chromatography/Fourier Transform Infrared Spectroscopy: Novel Hyphenated Methods in Thermal Analysis', *Anal. Chem.*, **65**, 2819–2823 (1993).
53. M. Webb, P.M. Last, C. Breen, 'Synergic Chemical Analysis – the Coupling of TG with FTIR, MS and GC-MS 1. The Determination of the Gases Released During the Thermal Oxidation of a Printed Circuit Board', *Thermochim. Acta*, **326**, 151–158 (1999).

X-ray Scattering in Analysis of Polymers

Christopher Y. Li, Bojie Wang, and Stephen Z.D. Cheng
The University of Akron, USA

1 Introduction 1
1.1 Origin of X-rays 1
1.2 Absorption, Filtering and Selection of X-rays 2
1.3 X-ray Safety 2
2 Theories of Coherent X-ray Scattering 3
2.1 Scattering from One Electron 3
2.2 Scattering from Many Electrons 3
2.3 Fourier Transformation between Electron Density and Structure Factor 4
3 X-ray Diffraction of Crystals 5
3.1 Crystal Lattices and Indices 5
3.2 X-ray Diffraction and Bragg's Law 6
3.3 Reciprocal Lattice 7
4 Experimental Techniques 7
4.1 Preliminary Considerations 7
4.2 Instrumentation 8
5 Polymer Crystal Structure Analysis 8
5.1 Use of Unoriented Diffraction Patterns 9
5.2 Structure Analyses Using Uniaxially Oriented Diffraction Patterns 9
5.3 Determination of Crystallinity 10
5.4 Application of X-ray Diffraction to Liquid Crystalline Polymers 12
6 Small Angle X-ray Scattering of Polymers 14
6.1 Two-phase Model: Porod's Law 14
6.2 Particle Scattering: Guinier Plot 15
6.3 Analyses of Oriented Polymers 16
6.4 X-ray Diffraction in Thin Films 16
Acknowledgments 17
Abbreviations and Acronyms 18
Related Articles 18
References 18

It has been proved that X-rays play a very important role in the characterization of polymers. A wealth of information concerning polymer structures and morphologies are accessible by using X-ray techniques. X-ray studies in the polymer area are based on the same principle as that in studies of other materials. The present review attempts to give a short summary of the roles that X-rays play in modern polymer science.

The first three sections provide a brief introduction about the origin of X-rays, the principle of the interaction of X-rays and matter, and X-ray diffraction. Section 4 briefly describes X-ray instrumentation. In section 5, crystal unit cell and structural determination methods are discussed. The focus is on the use of uniaxially oriented polymers, as this is the most commonly used method for determination of polymer crystal structures. Crystallinity is another issue, which is specifically important to semicrystalline polymers. A short summary of the concept of crystallinity and its determination based on X-ray diffraction is discussed. Polymer liquid crystals are a new field which has been growing rapidly since the late 1970s. X-rays have again played an important role in liquid crystal phase determination. By combination of X-ray techniques with other structure and thermal property measurements, almost every polymer liquid crystal phase can be identified. Section 6 discusses the small angle X-ray scattering (SAXS) technique. The cause of X-ray scattering in the small-angle region in most polymers is due to the heterogeneous electron density distribution over distances that are large in comparison with the wavelength of the X-ray (5–100 nm). SAXS is a powerful tool in semicrystalline morphology and block copolymer phase-morphology studies. In the thin film field, X-ray diffraction at small angles is also useful in determining the film thickness and interface structure.

1 INTRODUCTION

1.1 Origin of X-rays

X-rays have an approximate range of wavelengths from 0.01 nm (hard X-rays) to 10 nm (soft X-rays). In the electromagnetic spectrum, they lie between ultraviolet light and gamma radiation. There are two ways to generate X-rays, one of which is by rapid deceleration of fast-moving electrons and converting their energy of motion into a quantum of radiation. In this case, the wavelength produced depends on the energy of the electron. X-rays can also be emitted from radioactive isotopes such as ^{55}Fe, which are convenient for testing and calibration.[1]

To generate X-rays, electrons are accelerated by an electric field and directed against a metal target. The target slows them rapidly by multiple collisions and a continuum of radiation is formed. The minimum wavelength (λ_{min}) of this white radiation can be calculated

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

from Equation (1)[1]

$$\lambda_{min} = \frac{1240}{V_{acc}} \text{ nm} \quad (1)$$

where V_{acc} is the acceleration voltage. The greatest intensity occurs at a somewhat longer wavelength. The electron current does not change with respect to the voltage, although the peak moves to short wavelengths and the intensity increases as the voltage increases.

Electrons of each orbital possess certain energy. For the innermost K shell, the energy is around 10 000 eV for elements with atomic numbers of about 30. When bombarding electrons have higher energy than this value, the K shell electrons will be removed from the shell. The resulting vacancies are then filled by the descent of electrons from the next higher shell (the L shell) or the one above that (the M shell). The decrease in potential energy in going from the higher level to the lower one appears as radiation. Each transition gives a nearly monochromatic radiation wavelength (line) as the energies of the shells are well defined. The principal lines of $K_{\alpha 1}$ and $K_{\alpha 2}$ correspond to electrons moving from the L to K shells and those of $K_{\beta 1}$ and $K_{\beta 2}$ correspond to the electrons moving from the M to the K shells. X-ray spectra usually show a number of sharp spikes of high intensities whose positions change from one metal to another. These spikes are the characteristic lines for the target element.[2] As the atomic number of the target element increases, the characteristic lines shift to shorter wavelengths.

1.2 Absorption, Filtering and Selection of X-rays

The radiation used for most diffraction studies should be as nearly monochromatic as possible. The characteristic line fulfills this requirement. However, the interfering radiation needs to be filtered. The absorption properties of materials may be used to accomplish this objective. The absorption of X-rays by a solid follows Equation (2)[1,3]

$$\frac{I}{I_0} = e^{-\mu\tau} \quad (2)$$

where μ is the linear absorption coefficient, τ is the pathlength through the solid and I_0 and I are the intensity of X-rays before and after absorption. Figure 1 shows an X-ray spectrum of characteristic lines and the absorption coefficient change for the same element with respect to the wavelength.[3] The absorption edges (sharp jumps of μ) occur at the wavelength at which the incident X-ray quantum is just energetic enough to knock an electron out of an atomic orbital. In particular, the K absorption edge of an element lies slightly to the short-wavelength side of the K_β lines for that element. The absorption edge also shifts to shorter wavelengths with increasing atomic number. Fortunately, the absorption edge of one element

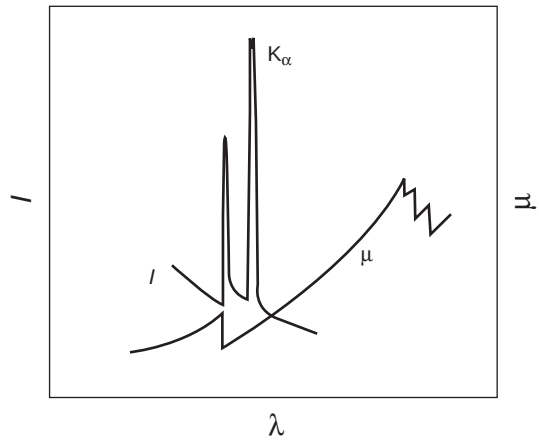

Figure 1 X-ray spectrum showing characteristic peaks and absorption coefficient for the same element as a function of wavelength.

lies between the K_α and K_β lines of the element with one atomic number higher. Therefore, it can be used as a selective filter to remove the K_β line from the spectrum of the next-highest element.

For many years copper K was the conventional choice for use with organic crystals and it is still the best if the diffraction data are to be recorded photographically – it is sufficiently penetrating while not suffering too badly from absorption in the crystal or while passing through air, and it is recorded with reasonable efficiency on films. When used with a diffraction apparatus that can record all of the reflections theoretically accessible, it can provide enough data to determine a structure.

1.3 X-ray Safety

X-ray equipment can be dangerous to the human body if it is mishandled, although the radiation produced by the sources used is comparatively nonpenetrating. The direct beam is particularly dangerous, because it represents a concentrated radiation source (several hundred roentgens per second) and brief exposure can cause serious injuries. Any diffraction equipment must have a beam stop to receive the direct beam that passes through the sample. When aligning cameras, in which case a beam stop cannot be used, the beam intensity should be reduced to the lowest possible value and the experimentor should take particular care to keep clear of the radiation. Shutters should always be provided for covering the exit windows of the X-ray source and should be kept closed whenever diffraction is not being recorded.[1,4,5]

Scattered radiation can also dangerous, although it is less serious than the direct beam. Most scattered radiation originates around the junction between the

X-ray source and the collimator. This junction should be designed with a labyrinth type of coupling so that no radiation can escape. Plexiglas™ is used to construct a surrounding shield to eliminate any risk from scattered radiation. A comprehensive study of this problem and of shutter designs has been made by the Apparatus and Standards Committee of the American Crystallographic Association.[4]

The question of legal responsibility in X-ray diffraction laboratories is uncertain but is becoming increasingly serious. Licensing and inspection requirements vary greatly from state to state and from country to country. Liability in case of any mishaps is potentially a major problem despite the excellent safety record of X-ray diffraction laboratories (X-ray analysis laboratories have not always been so fortunate).

2 THEORIES OF COHERENT X-RAY SCATTERING

Matter absorbs and scatters X-rays while interacting with them. After scattering, some X-rays will change their wavelength, a phenomenon known as incoherent scattering; the rest will retain their wavelength, and this scattering is called coherent scattering. Incoherent scattering, which is also known as Compton scattering, usually gives rise to a continuous background in diffraction experiments. Coherent scattering can be considered as the result of collisions of photons with electrons without any energy loss. It can establish systematic interference effects between rays scattered by different volume elements of the structure, and is therefore of interest in this discussion.[6–13]

2.1 Scattering from One Electron

To establish general relationships between structure and the intensity of the coherently scattered radiation, consider the scattering of a single electron. In Figure 2, the electron sits at the point O and is irradiated by an unpolarized X-ray beam of amplitude E_0. The amplitude of the component of the scattered radiation, which is polarized perpendicular to the plane AOB of the primary and the scattered beam, is given by Equation (3)

$$E_{1\perp} = \frac{E_0(e^2/mc^2)}{R} \tag{3}$$

where R is the distance from the electron (point O), c is the speed of light, and m and e are the mass and charge of the electron. The amplitude of the component polarized

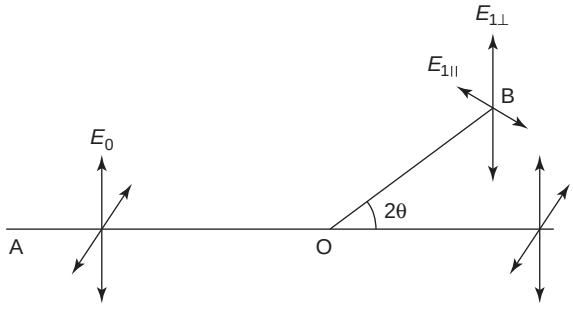

Figure 2 Geometry of scattering from a single electron.

parallel to the plane AOB is given by Equation (4)

$$E_{1\|} = E_0 \left(\frac{e^2}{mc^2} \right) \frac{\cos 2\theta}{R} \tag{4}$$

where 2θ is the angle between the incident and the scattered radiation as shown in Figure 2. The intensity I at B is (Equation 5)

$$I = EE^* \tag{5}$$

The intensity I_1 of the radiation scattered by a single electron then can be written as

$$I_1 = I_0 \left(\frac{e^2}{mc^2} \right)^2 \frac{1 + \cos^2 2\theta}{2R^2} \tag{6}$$

The factor $(e^2/mc^2)^2$ is called Thomson's number. I_0 is the intensity of the primary beam at the position of the electron. Equation (6) shows that, apart from the polarization factor $(1 + \cos^2 2\theta)/2$, which changes with angle in a very gradual way, the scattering from a single electron corresponds to a spherical wave, emanating from the electron.

2.2 Scattering from Many Electrons

The scattering from many electrons can be treated in a straightforward manner, based on the interference between the scattered waves of the individual electrons. For the system containing Q electrons, the maximum scattering amplitude occurs when all the scattered waves are in phase, and will amount to QE_1, where E_1 is given by Equations (3) and (4). However, because of destructive interference, the actual amplitude will in general be much lower, corresponding to a number of F electrons only, with $F < Q$ (Equation 7)

$$E = FE_1 \tag{7}$$

The quantity F is called the structure factor. In structures not containing a center of symmetry it is generally a complex number, whereas in centrosymmetric structures it can be either positive or negative.

The relationship for the intensity I is Equation (8)

$$I = FF^* I_1 \tag{8}$$

where the quantity FF^* is defined as the absolute intensity. In samples of macroscopic dimensions, the diffracted intensity will be proportional to the irradiated volume V. The absolute intensity per unit volume i can be defined (Equation 9) as

$$i = \frac{FF^*}{V} \tag{9}$$

The intensity to be measured in a diffraction experiment can be related to i through Equations (6), (8) and (9). Owing to geometric factors and absorption effects, the intensity of the primary beam is not constant over the whole irradiated part of the sample. Therefore, the contribution $\mathrm{d}I$ per volume element $\mathrm{d}V$ is normally considered which, from Equations (6), (8) and (9), is found to be (Equation 10)

$$\mathrm{d}I = iI_0 \left(\frac{e^2}{mc^2}\right)^2 \frac{1+\cos^2 2\theta}{2R^2}\,\mathrm{d}V \tag{10}$$

Figure 3 shows two electrons at a distance of r from each other. Let the origin of the coordinate system be at e_1, and let the directions of the primary and one of the diffracted beams be indicated by the unit vectors $\mathbf{S}_0$ and $\mathbf{S}_1$. MN and AB indicate the path difference between the rays scattered by e_1 and e_2 in the direction of $\mathbf{S}_1$ and $\mathrm{MN} = r\mathbf{S}_0$ and $\mathrm{AB} = r\mathbf{S}_1$. At a far distant point in the direction of $\mathbf{S}_1$, where the rays from the two electrons converge, these rays will show a phase difference (Equation 11) of

$$\frac{\mathrm{AB} - \mathrm{MN}}{\lambda} = \frac{r\mathbf{S}_1 - \mathbf{S}_2}{\lambda} \tag{11}$$

where λ is the wavelength of the X-rays. The vector (Equation 12)

$$\mathbf{s} = \frac{\mathbf{S}_1 - \mathbf{S}_2}{\lambda} \tag{12}$$

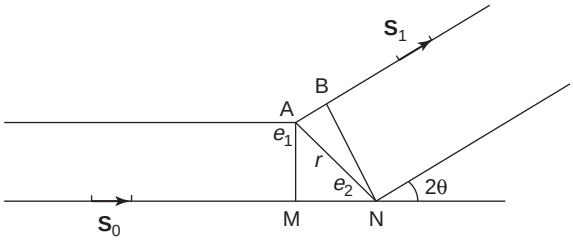

Figure 3 Geometry of scattering from two electrons.

is the scattering vector which has a magnitude (Equation 13) of

$$s = |\mathbf{s}| = \frac{2\sin\theta}{\lambda} \tag{13}$$

The resulting amplitude will be given by Equation (14)

$$E = E_1 e^0 + E_1 e^{2\pi i \mathbf{rs}} \tag{14}$$

Generalizing this result to many electrons, of which the positions are indicated by the vector $\mathbf{r}_k$, fields Equation (15):

$$F = \frac{E}{E_1} = \sum_k e^{2\pi i \mathbf{r}_k \mathbf{s}} \tag{15}$$

Another way to describe the scattering from many electrons is Equation (16)

$$F(s) = \int_{Vr} \rho(r)\, e^{2\pi i \mathbf{rs}}\,\mathrm{d}V_r \tag{16}$$

where the structure is divided into volumes that are much smaller than $1/s$, $\rho(r)$ is the electron density in the volume element $\mathrm{d}V_r$ at the end-point of the vector $\mathbf{r}$, and the integral is to be taken over the irradiated volume.

2.3 Fourier Transformation between Electron Density and Structure Factor

The most striking property of Equation (16) is that it can be ''reversed''. Thus, $\rho(r)$ in Equation (16) can be solved from the very similar equation

$$\rho(\mathbf{r}) = \int_{Vs} F(\mathbf{s}) e^{-2\pi i \mathbf{rs}}\,\mathrm{d}V_s \tag{17}$$

The functions $\rho(r)$ and $F(s)$ are said to be Fourier Transforms of each other (Equations 18 and 19):

$$F(\mathbf{s}) = \mathbf{F}\{\rho(\mathbf{r})\} \tag{18}$$

$$\rho(\mathbf{r}) = \mathbf{F}^{-1}\{F(\mathbf{s})\} \tag{19}$$

The operators $\mathbf{F}$ and $\mathbf{F}^{-1}$ differ only in the minus sign in the exponent. The difference disappears in the case of functions with a center of symmetry.

The integral presented in Equation (17) has to be taken over all elements of the space in which the vector $\mathbf{s}$ can be drawn. At each point in this space the value of F can be defined with the aid of Equation (16). From this equation, it can be seen that there exists a relationship of reciprocity between the function F in this space and the function ρ in real space. Thus, magnifying the structure in real space by a factor μ will result in demagnification of the F^{-1} function in the space of the $\mathbf{s}$ vectors by a factor $1/\mu$. In view of this relationship, the $\mathbf{s}$ space is usually called the reciprocal space, discussed in the next section.

3 X-RAY DIFFRACTION OF CRYSTALS

3.1 Crystal Lattices and Indices

It is convenient to consider a two-dimensional array as an introduction to crystal structure. Figure 4 shows a regular repetition of a simple motif, a five-pointed star, by translation along two directions in the plane. The array of stars can be referred to a lattice or grid system that can be chosen in different ways. For an unlimited lattice there are an infinite number of ways of drawing the grid lines and, in principle, all are equally valid. However, in practice matters of convenience and convention dictate the choice. Figure 5(a) and (b) show two different ways of drawing the grid lines for the array in Figure 4. It is important to note that the grid lines are drawn at equal intervals corresponding to the repeat distances of the array and that as a consequence the surroundings of each grid line intersection are identical. In finite lattices this is not true for points at the edges, but we are dealing with lattices of such sizes that for practical purposes they can be considered infinite. That area in Figure 5(a) and (b) set off by successive grid lines in two grid-line directions is termed a unit cell. In a three-dimensional lattice, the unit cell is a volume. It is the ''unit structure'' referred to above that produces the macroscopic crystal when stacked side to side in three dimensions.[3,10,12,13]

It must be emphasized that the lattice is a purely imaginary construct yet a very necessary and useful one. It functions as a coordinate system to which the structure is referred. The three grid line directions are termed x, y, and z and are chosen to form a right-handed system. The lengths of the unit-cell edges along the x-, y-, and z-axes, respectively, are labeled a, b, and c. These same symbols are often used for the axes as well. The angles between the axes are α, β, and γ, with α between the b- and c-axes, β between the a- and c-axes, and γ between the a- and b-axes.

For use in the subsequent sections, it is necessary to consider sets of parallel lattice planes constructed so that for any given set every lattice point lies on some member of it. Figure 6(a) and (b) show the same two-dimensional lattice with different sets of planes constructed in this way. These sets of planes are identified by three numbers corresponding to three axes. When such planes cut an edge of the unit cell the edge is always divided into an integral number of equal parts. These are common fractions of the unit translation 1, (1/1), 1/2, . . . , $1/n$. The fractional intercepts on the three unit-cell axes are used as a basis for triple numbers, the indices that uniquely

Figure 4 Two-dimensional array of stars.

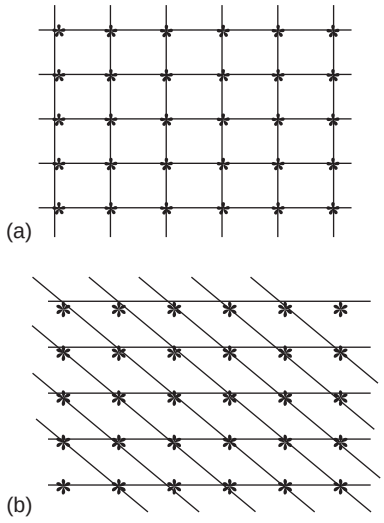

Figure 5 Two different grid systems referred to the array of Figure 4.

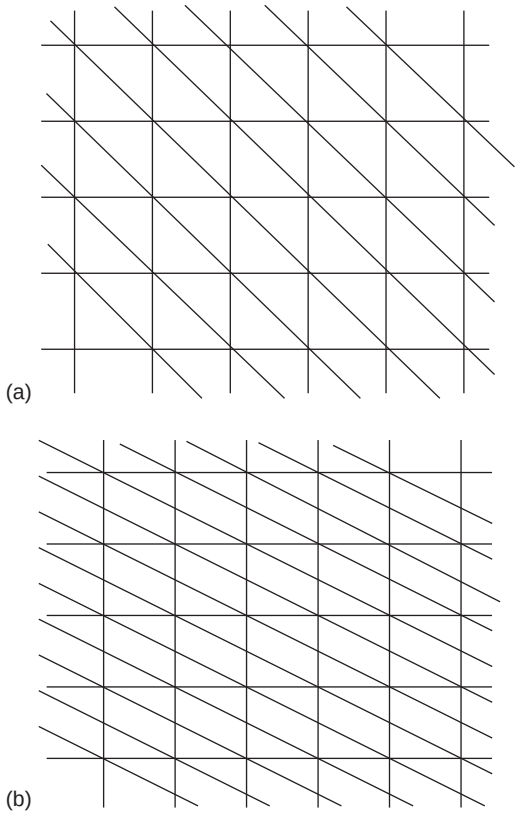

Figure 6 Two sets of lines in a two-dimensional lattice.

characterize each possible set of planes. These indices are obtained by considering some lattice point as the origin and proceeding from it along the axes until the first member of a set of planes is reached. The intercepts of the plane on the axes are expressed as fractions of the unit cell edge, and their reciprocals are the desired indices. In Figure 7, plane 1 has intercepts 1/2, 1/2, and 1 on the *x*-, *y*-, and *z*-axes, respectively, and has indices of (2,2,1). A special case arises when the planes in question are parallel to one or more of the axes. The planes and the axis do not intercept, and the corresponding index is defined as zero.

In classical crystallography, the planes (2,2,1) and (4,4,2) are indistinguishable, and Miller indices can be used. Miller indices refer to the indices of plane that have no common factor. They can be used in any of three ways: to designate a set of lattice planes, a particular member of the set, or the face of a macroscopic crystal parallel to the set. However, [*x*,*y*,*z*] indicates a direction in the lattice, which is a line segment from the origin to the point *x*,*y*,*z*. Thus, the three cell axes are [100], [010], and [001], whereas, (100), (010), and (001) are the *yz*, *xz*, and *xy* planes, respectively. Note that the plane (100) intersects the axis [100], although the two are not necessarily perpendicular, and similarly for the other planes and axes.

A third related notation is {100}, {010}, and {001} which refers to a set of face planes that are equivalent as a result of the symmetry of the crystal. The number of planes included depends on the particular crystal and its symmetry; thus, for a cubic crystal, {100} includes, (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, and, $(00\bar{1})$, whereas for a triclinic crystal it only includes (100). The indices for the various planes in a lattice depends on how the unit cell is chosen. Transformation of the indices in changing from one unit cell to another follows certain well-established rules, which are treated in more detail in specialized texts.

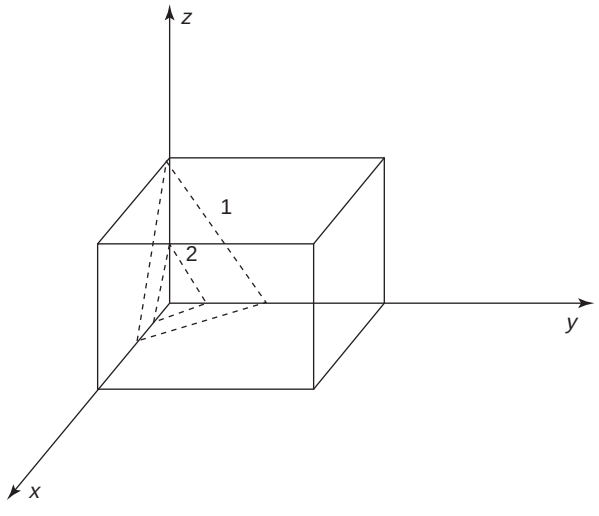

Figure 7 A lattice plane and an interleaving plane.

3.2 X-ray Diffraction and Bragg's Law

X-rays were discovered in 1895 by Roentgen[14] and the diffraction of X-rays by crystals was first observed by M. von Laue in 1912,[15] During the years following their discoveries, a number of determined efforts were made to prove them to be particles or waves. It was the diffraction by crystal experiment that proved the wave character of X-rays. Following the experimental observation of X-ray diffraction early in 1912, von Laue showed that the phenomenon could be described in terms of diffraction from a three-dimensional grating. In the same year, however, W.L. Bragg noticed the similarity of diffraction to ordinary reflection and deduced a simple equation treating diffraction as reflection from planes in the lattice. In order to derive the equation, consider an X-ray beam incident on a pair of parallel planes P_1 and P_2 with interplanar spacing d as shown in Figure 8. The parallel incident rays 1 and 2 have an angle θ with these planes. Electrons assumed at A and B will be forced to vibrate by the oscillating field of the incident beam and, being vibrating charges, will radiate in all directions. For that particular direction where the parallel secondary rays $1'$ and $2'$ emerge at angle θ as if reflected from the planes, a diffracted beam of maximum intensity will result if the waves represented by these rays are in phase. Dropping perpendiculars from A to C and D, respectively, given BC = BD. If BC + BD is an integral number of wavelength λ, ray $1'$ and ray $2'$ will be in phase, and one may thus write

$$\mathrm{BC} + \mathrm{BD} = 2\mathrm{BC} = n\lambda \tag{20}$$

where n is an integer. As $\mathrm{BC}/d = \sin\theta$, substitution in Equation (20) yields

$$2d\sin\theta = n\lambda \tag{21}$$

which is Bragg's law.

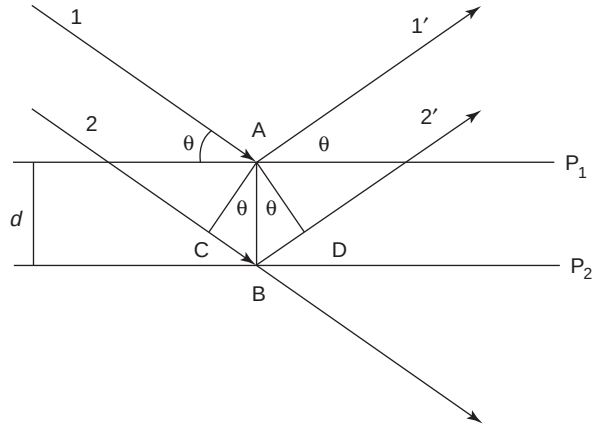

Figure 8 Conditions of diffraction in Bragg's law.

In this derivation reflections from only two planes were considered. In such a case the diffraction maxima would be broad and the various diffracted rays would virtually merge. In crystals many hundreds or thousands of planes make up each of the mosaic blocks that constitute the macroscopic crystal and, under these conditions, the diffraction maximum becomes sharp and occurs only at clearly defined values of θ. Indeed, for real crystals the breadth of a reflection as measured in terms of the range of θ over which it can be observed is usually a small fraction of a degree and due mostly to the slight misalignment of the mosaic blocks.

3.3 Reciprocal Lattice

The reciprocal lattice can be defined as follows. Starting from a lattice point as an origin, normals can be drawn to all possible direct lattice planes (*hkl*). The length of these normals is defined as $1/d_{hkl}$, where d_{hkl} is the perpendicular distance between planes of set (*hkl*). All these end-points of the normals construct the reciprocal lattice. Figure 9 shows a two-dimensional example, in which reciprocal lattice points corresponding to the direct lattice planes (0,1), (1,1), (2,1), and (3,1) are designated by asterisks and assigned the indices of the planes they represent.

The reciprocal lattice points shown as asterisks in Figure 9 derive from planes of direct lattice points, so their indices correspond to Miller indices. However, the reciprocal lattice also includes many points whose indices have common factors, such as 224 and 336. These may be considered as representing parallel families of planes not all of which contain lattice points or, more simply, different orders of diffraction from the fundamental sets of planes. Thus, in Bragg's law the integer n is the order of a particular reflection. It is simply the number of wavelengths difference in pathlength for rays reflected from successive planes. But n need not to be included explicitly when diffraction is considered in terms of the reciprocal lattice.

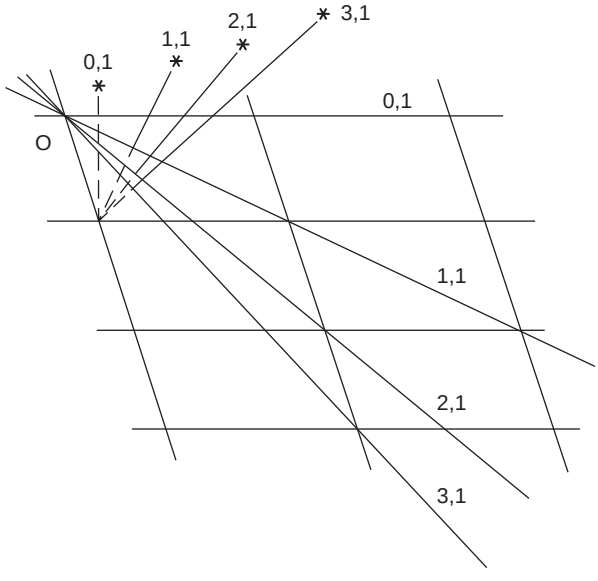

Figure 9 Two-dimensional direct and reciprocal space.

Table 1 Direct and reciprocal space relationship in an orthorhombic unit cell

$a^* = 1/a$	$a = 1/a^*$	$\alpha = \beta = \gamma = \alpha^* = \beta^* = \gamma^* = 90^\circ$
$b^* = 1/b$	$b = 1/b^*$	$V^* = 1/V = a^*b^*c^*$
$c^* = 1/c$	$c = 1/c^*$	$V = 1/V^* = abc$

The relationships between the direct and reciprocal lattices in three dimensions depend on the angles between the axes in the direct lattice. Consider the case of an orthorhombic cell, where the three direct axes are mutually perpendicular to each other but the axial lengths are not the same. The planes (100), (010), and (001) are perpendicular to the *a*-, *b*-, and *c*-axes, respectively, so that their normals are along these axes. Hence, the reciprocal lattice point 100 can be considered to be on the *a*-axis at a distance $1/a$ from the origin. The line connecting this point to the origin is one edge of the unit cell in reciprocal space and is thus a^*. In the same way the reciprocal axes b^* and c^* can be seen to coincide with the corresponding direct axes b and c, and all three reciprocal axes are also mutually perpendicular to each other. The result is a reciprocal unit cell with the relationships summarized in Table 1. The relationships between direct and reciprocal lattices are described in greater detail elsewhere.[1–3,12,13]

4 EXPERIMENTAL TECHNIQUES

Rather than describe the detailed laboratory set-up, this section covers preliminary considerations and general guidelines on the instrumentation involved.[10,12,13]

4.1 Preliminary Considerations

Before carrying out an X-ray experiment, careful attention must be given to the choice of apparatus and procedure – the resultant X-ray diffraction pattern should be one that most readily yields the required information about the specimen. The decision as to the most appropriate type of diffraction pattern depends upon knowledge of the kind of information deducible from each pattern type. Similar knowledge is required for the most suitable choice of experimental equipment (type of X-ray camera), X-ray parameters (wavelength, voltage, current, monochromatization), and the establishment of experimental procedures (method of recording

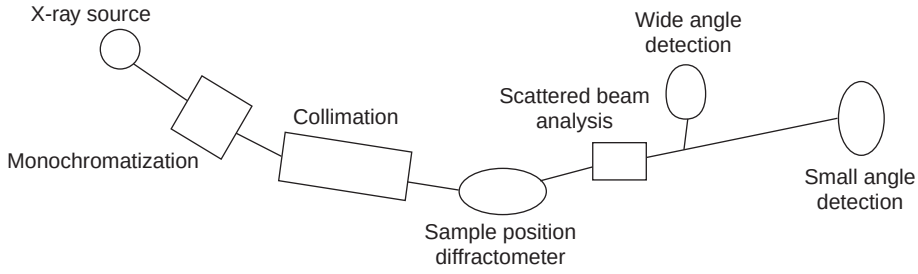

Figure 10 Diagram of an X-ray instrument set-up.

diffraction intensities). Information about the specimen, already obtained by other methods, provides essential data for planning the investigation.

The object of all X-ray diffraction investigations is to find one or more of the following three essential variables, which therefore form the three basic criteria in X-ray diffraction studies. First are the directions (in 2θ) in which the scattered X-rays are diffracted (which necessarily involves measuring the orientations of the specimen and the observation point with respect to the incident X-rays). Second are the features of the diffraction pattern (whether sharp or broad spots, lines, or arcs, or broad halos), and third is the intensity of the diffraction in the various directions (in the case of a continuous pattern, the intensity distribution within the pattern).

The importance of these evaluations, and the precision with which they must be made, will depend upon the objectives of the analysis. Even during the course of an investigation, the ultimate usefulness of a particular X-ray diffraction pattern may come into question, and it may be that the need to verify some tentative conclusion (or a requirement for more detailed information) will dictate the subsequent course of the investigation.

4.2 Instrumentation

Figure 10 is a diagram of the key parts in an X-ray scattering experiment. The key features are components for monochromatization, collimation, sample positioning and environmental control, scattered beam analysis, and detection. Additional components used in some set-ups include mirrors for redirecting or focusing the beam and beam intensity monitors. Either wide angle ($>5^\circ$) information or small angle ($0–5^\circ$) data can be collected, depending on the sample–detector distance. The analysis of wide angle diffraction and small angle scattering is presented in sections 5 and 6, respectively. Measurements of X-ray scattering phenomena require two basic elements, a radiation source and a diffractometer (also called a camera or spectrometer). When appraising the suitability of an instrument to a particular analysis or specifying a system, important issues to consider are resolution, sensitivity, speed of measurements, and cost. For a given resolution, speed is determined by source intensity and detection efficiency, and speed may be gained if one is willing to relinquish resolution.

Two general source types are the tube source and the synchrotron. Common X-ray tubes are of the sealed or rotating target type, with the rotating target offering a factor of 10 increase in intensity at a substantial increase in capital expenditure, system complexity, and maintenance cost. The focus stability of a rotating anode is also not as good as that of the sealed tube. With either tube type various wavelengths may be obtained by securing a selection of targets of various materials, generally at substantial additional cost. Because the focus acts effectively as the first slit in the collimation of the beam, the attainable collimation and intensity is dependent on the focus type. A fine focus in either a slit or point configuration offers the finest collimation and greatest brilliance at a loss of total intensity and stability. The point focus has analytical advantages, but the slit focus allows the total intensity to be maximized, while paying a price in the ease of data analysis by introducing a substantial loss of collimation in one beam dimension. Synchrotron sources, which are available only at large centralized user facilities, offer a combination of high intensity, intrinsically outstanding collimation, and precise wavelength tunability. These capabilities are demanded by more sophisticated and difficult scattering experiments, particularly when time-resolved in situ data are required.

Many similarities exist among the diffractometers used for different types of X-ray scattering measurements, as well as important differences.[10,13,16]

5 POLYMER CRYSTAL STRUCTURE ANALYSIS

Polymer crystals are different from crystals of small molecules. The basic features of polymer crystals are small crystal sizes and anisotropic crystal properties along different unit-cell directions. This is because those linear polymers consist of chemical bonds along the chain

direction, whereas the lateral interactions are attributed to van der Waals forces. The long-chain nature leads to chain folding to form imperfect crystals with small sizes.

5.1 Use of Unoriented Diffraction Patterns

For polymer crystals with very small sizes a definitive structural analysis is difficult using only unoriented diffraction patterns. This results from the small number of observable diffractions and their general broadening, and also from the high-angle diffraction rings obtained, which often consist of the superimposition of several reflections of different indices. In certain special cases, as where a structural similarity to a known substance is suspected on the basis of chemical composition, it may be possible to make an identification by comparison; however, even this procedure contains some pitfalls.[10,12,13]

5.2 Structure Analyses Using Uniaxially Oriented Diffraction Patterns

In most cases, structure analyses of polymer crystals are carried out using uniaxially oriented samples (fibers or films).[12,13,17–24] The basic procedures include: (1) determination of the fiber period, (2) indexing (*hkl*) diffractions and determining unit-cell parameters, (3) determination of the space symmetry group, (4) structure analysis, and (5) Fourier transforms and syntheses and Patterson functions. The first three aspects of the procedure are discussed here, the last two aspects are discussed by Kakudo and Kasai[12] and Alexander.[13]

5.2.1 Determination of the Fiber Period

Polymer chains usually align along the fiber axis during drawing and deformation, and the chain direction is commonly defined as the *c*-axis (there are several exceptions due to hydrogen bonding, smectic domains in liquid crystals, and other reasons). The fiber period for the fully extended conformation of the polymer chains is calculated from the chemical structure of the repeat unit of the polymer using standard bond angle and bond length data.[25] The measured value of the fiber period can be compared with the calculated value. If the value agrees within the limit of experimental error, the chain is considered to be fully extended in the fiber. When the calculated period is slightly greater than the experimental value, this often indicates a twisted or loosely folded conformation of the molecular chains and the degree of twisting or fold can be estimated from the discrepancy between the two values. Another possibility is that the chains are slightly inclined with respect to the principal axis of the unit cell. A large discrepancy often points to a helical conformation of the chains. Computer simulation of the chain conformation with the lowest free energy is a powerful tool with which to examine these discrepancies. However, in some cases, interchain interactions must also be taken into consideration. This may lead to complication of the crystal structure determinations.

5.2.2 Indexing Diffractions and Determining Unit Cell Parameters

In most cases the film is placed on the Bernal chart for the appropriate film radius, and the values of the reciprocal lattice coordinates can be read off directly. Diffraction diagrams made with a microcamera may be enlarged to a suitable size and the value read with the aid of this chart. For samples with perfect orientation, the chain axis (usually the *c*-axis) is aligned along the fiber direction, whereas the a^*- and b^*-axes in reciprocal space are perpendicular to the fiber axis. Therefore, diffractions on the equator come from (*hk*0) diffractions. The first step to determining the unit-cell parameters is to index diffractions on the equator. As in small-molecule single-crystal rotating experiments, several reflections close to the center are chosen, the distances from the reflected points to the center are measured to be $\xi_1, \xi_2, \ldots, \xi_n$. The smallest ξ value (i.e. ξ_1) is assumed to be on the a^*-axis, and is plotted on this axis to the right of the origin as the 100 reciprocal lattice point, at the appropriate distance on the chosen scale. An attempt is also made to find the b^*-axis based on the second closest, ξ_2, which is assumed to be the 010 reciprocal lattice point, and the third closest, ξ_3, which is assumed to be representative of the 110 reciprocal lattice point. A suitable combination of a low-index triangle reciprocal lattice may thus be constructed. Two sides of the triangle lattice are the a^*- and b^*-axes, respectively. A complete two-dimensional lattice can thus be established by translation of the triangle into the two-dimensional space. The cross-over points in the reciprocal lattice, which represent 110, 210, 120, . . ., reciprocal lattice points, must correspond to the diffractions observed along the equator direction (ξ_4, ξ_5, etc.). It should be particularly noted that reflection does not necessarily occur in practice from all sets of planes associated with reciprocal lattice points, although a plane might fulfill these conditions. There are other conditions, connected with the symmetry properties of the crystal, which may be responsible for the systematic absence of certain reflections (the extinction rules, see below).

The next step is to index the general reflections of the type *hkl*, which needs information about the c^*-axis (the fiber period). The first-layer reflections above and below the equator are the *hk*1 and $hk\bar{1}$ reflections, respectively. The indices to be determined in this case are *h* and *k*. If the third axis, i.e. the *c*-axis in this case, is normal to the *a*- and *b*-axes, the c^*-axis will coincide with the *c*-axis, also being perpendicular to both the a^*- and b^*-axes. In this case the ξ values of all the reciprocal lattice

points $hk1$ and $hk\bar{1}$ will be equal to the ξ values of the reciprocal lattice points having the same values of h and k in the equator plane. To index the points, therefore, a point whose ξ value is the same as that of an $hk0$ point is given the same indices h and k, and so becomes $hk1$ or $hk\bar{1}$. The indices h and k can be simply determined using the values of corresponding spots on the equator. If the third axis is not normal to the a^*- and b^*-axes, such as in the case of determination of a triclinic unit cell, the procedure is more tedious. In this case the ξ values of the points on the layer lines are no longer the same as those of the $hk0$ points having the same values of h and k, namely, the projections of the lattice points on the equatorial plane do not coincide with the equatorial lattice points. However, as all the lattice layers are identical, provisional use of this equatorial plane is still possible. The origin (the point where both h and k are zero) of the first layer of the reciprocal lattice is projected on the equatorial plane, and a new lattice is drawn about the projected point as lattice origin. It will therefore be slightly offset from the previous lattice. The original origin of the equatorial plane is taken as the center of the diffraction sphere, which is then rotated. The same result is obtained if the center of rotation of the diffraction sphere is moved from the origin by exactly the same amount as the offset of the points in the first layer, the diffraction sphere then being rotated in the usual manner. By repetition of this procedure for the first layer below the equator, and for the second, third, and other higher-order layers, all the reflections hkl on the film can be indexed. A few examples of detailed determination of uniaxially oriented polymer unit cell parameters have been reported.[26–28]

5.2.3 Determination of the Space Group

X-ray diffractions from a crystal can only occur in directions defined by the reciprocal lattice points, and with intensities which are governed by a structure factor of the form (Equation 22)

$$F(hkl) = \sum_{j}^{N} f_j \exp\{2\pi i(hx_j + ky_j + lz_j)\} \qquad (22)$$

where f_j is the atomic scattering factor of the jth atom in the unit cell, and x_j, y_j, and z_j are the coordinates of this atom in the unit cell. The summation is carried out over all the atoms in the unit cell. The atoms in the unit cell are not independent, but generally fall into groups within which the atoms are related by the symmetry of the crystal. As comparatively simple relationships hold between the coordinates of the atoms in these groups, insertion of these relationships simplifies the equation and reduces the number of terms in the summation. Certain symmetry operation will result in zero intensity for certain (hkl) planes, which is called extinction. As this feature arises from the crystal symmetry, the space group of the crystal may be identifiable from the characteristic systematic absences. In an actual analysis, the indices of all the diffraction spots for the crystal are first tabulated. Regularities are then sought in the indices of observed reflections. This is because sometimes a very weak reflection may not be observed due to the experimental limitations, obscuring the regularity of the indices of the completely missing reflections. The symmetry is determined by both the point groups and space groups, which include rotations, inversions, screw axes, and glide planes. They govern the extinction rules. Finally, the space group can be found by consulting International Tables for X-ray Crystallography.[17]

It is difficult to determine the space group of polymer crystals, because the diffraction diagrams can usually be obtained only for uniaxially oriented specimens. It is often impossible to verify the extinction rules because of indexing difficulties that arise from overlapping of reciprocal lattice points, and the inability to observe the higher-order diffractions. In such cases, one solution is to select a possible space group and carry out the analysis, assuming initially that the symmetry is high. If this fails, other space groups are tried in turn in an attempt to find the correct one. However, this is a rather trial and error method, and it is better to try to obtain as many independent $|F(hkl)|^2$ values as possible, e.g. by preparing a biaxially drawn specimen, making inclined uniaxially oriented fiber photographs or Weissenberg photographs, or using X-rays of shorter wavelength, such as Mo radiation, to obtain data for higher-order diffractions. Structural analysis using Fourier transforms and Patterson functions are relatively complicated and are not discussed here.

5.3 Determination of Crystallinity

The concept of the crystallinity of a semicrystalline polymer is based on the two-phase approximation of the polymer structure, which involves the assumption that uniform crystalline and amorphous regions can be distinguished. This model was adopted in the early years of polymer science and the first serious attempts to determine the ratio of the two phases in a polymer date to the 1930s and 1940s. This model successfully explained the mechanical properties of semicrystalline polymers using continuous mechanics. In the years following the discovery of the folded chain morphology (the chain folding principle), the two-phase concept has been questioned. It was difficult to reconcile the model with the presence of layers containing sharp folds, thought to be present at the interface between adjoining crystalline

lamellae. Instead, the crystal defect model was proposed, according to which the amorphous phase is dispersed in the crystalline phase in the form of defects within the crystalline structure. According to this concept, the degree of crystallinity no longer had a direct obvious meaning, and terms such as the *crystallinity index* or *degree of order* were proposed to quantify the phenomena associated with the occurrence of the defects. However, as a result of increasing experimental evidence, in which both small angle and wide angle X-ray scattering played an important role, confidence in the two-phase model was later restored, and nowadays is generally accepted as the basis for the description of polymer morphology.[10,13,29]

In semicrystalline polymers, the crystallinity can vary widely.[29] Roughly speaking, three classes of polymers can be identified: those having high crystallinity of over 50%, those having medium crystallinity between 20% and 50%, and those having low crystallinity of below 20%. The high-crystallinity polymers can be seen in linear polyethylene, isotactic polypropylene, and poly(ethylene oxide) examples of medium-crystallinity polymers are poly(ethylene teraphthalate), poly(aryl ether ketone)s, and poly(phenylene sulfide); the low-crystallinity polymers include poly(vinyl chloride), some high-temperature aromatic polyimides such as poly(4,4′-oxydiphenylene pyromellitimide), and ultra-low-density short-chain branched polyethylenes. Deviations from the two-phase model seem to occur in the sense that the phase boundary may occupy nonnegligible weight and volume in the materials, in which the structural order gradually changes between those of the crystalline and of the amorphous phase. However, such refinements seem to not invalidate the general idea of the occurrence of two discrete phases. In this section, the width of the phase boundary is ignored.

The crystallinity is a meaningful quantity, and its determination is a matter of paramount importance in the investigation of polymer morphology. It is presented as a fraction either by weight w^c or by volume v^c. Denoting the densities of the whole polymer, the crystalline phase, and the amorphous phase as ρ, ρ_c, and ρ_a, respectively, Equations (23–26) are found to hold:

$$v^c = \frac{V_c}{V} \tag{23}$$

$$w^c = \frac{W_c}{W} \tag{24}$$

$$v^c = \frac{\rho - \rho_a}{\rho_c - \rho_a} \tag{25}$$

$$w^c = \frac{\rho_c}{\rho} v^c \tag{26}$$

The most universal method with which to determine crystallinity was developed by Ruland[30] who introduced a disorder function D (Equation 27):

$$4\pi \int_0^\infty s^2 I_c(s)\,\mathrm{d}s = w^c 4\pi \int_0^\infty s^2 \overline{f^2} D\,\mathrm{d}s \tag{27}$$

In this equation w^c represents the true weight fraction of crystals in the sample. The quantity $\overline{f^2}$ represents the mean-square atomic scattering factor of the polymer summed over all atoms i (where N_i is the number of atoms of type i; Equation (28)):

$$\overline{f^2} = \frac{\sum N_i f_i^2}{\sum N_i} \tag{28}$$

The total diffraction can be expressed (Equation 29) as

$$4\pi \int_0^\infty s^2 I(s)\,\mathrm{d}s = 4\pi \int_0^\infty s^2 \overline{f^2}\,\mathrm{d}s \tag{29}$$

and the weight fraction crystallinity which can be used to evaluate the experimentally obtained data is

$$w^c = \frac{\int_0^\infty s^2 I_c(s)\,\mathrm{d}s \int_0^\infty s^2 \overline{f^2}\,\mathrm{d}s}{\int_0^\infty s^2 I\,\mathrm{d}s \int_0^\infty s^2\,\overline{f^2} D\,\mathrm{d}s} \tag{30}$$

Application of Equation (30) requires either a sample with completely random orientation of crystals or a proper averaging of oriented samples. Averaging techniques have been described by Ruland and Dewaelheyns[31] and by Desper and Stein.[32] Another assumption in the derivation is that three-dimensional order exists in the crystalline regions, so that the disorder is not strongly anisotropic. For the treatment of the anisotropic case additional structural detail is needed to evaluate the amount of material that shows one- or two-dimensional order.[33] An example of an $s^2 I(s)$ curve computed directly from the experimental data of an isotactic polypropylene sample is shown in Figure 11.[30] Before using Equation (30), the incoherent Compton scattering (shadowed area) must be subtracted. The crystalline portion $s^2 I_c(s)$ can be delineated by subtraction of the undulating diffuse scattering area (dotted area). The separation line can be found by drawing a smooth curve through the minimum between the crystalline peaks following the general slope of the continuous scattering. According to Ruland,[30] the procedure restricts the region regarded as being crystalline to sizes bigger than 2–3 nm.

A number of more approximate methods have been devised to avoid much of the involved computation described above and used to estimate the crystallinity from X-ray diffraction data.[13] A diffraction curve of an unoriented sample over a limited range of angles that contains most of the crystalline diffraction peaks and

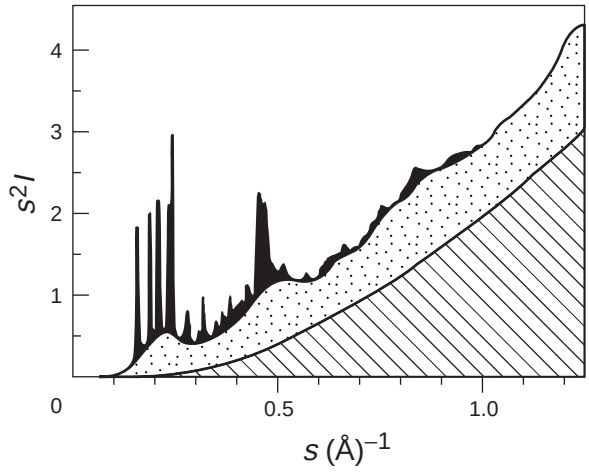

Figure 11 Scattering curve for isotactic polypropylene showing incoherent scattering, amorphous scattering, and crystalline scattering.

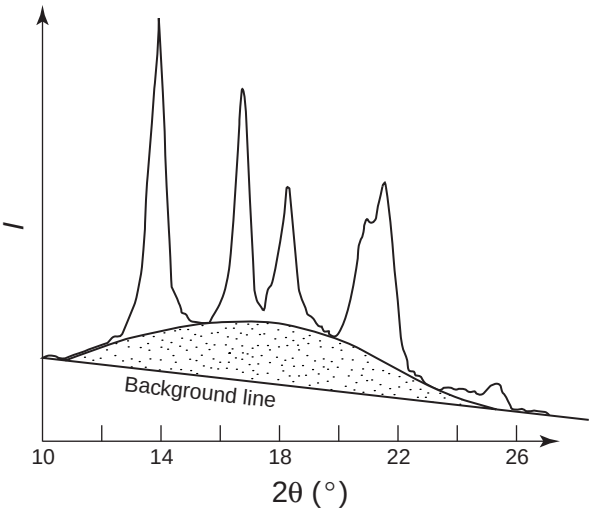

Figure 12 Scattering curve for isotactic polypropylene showing the crystallinity calculation.

the major portion of the amorphous halo can typically be separated, as indicated for the case of isotactic polypropylene in Figure 12.[34] A straight background line is drawn between the limits of the chosen interval of angles which should separate the incoherent scattering and other, largely crystallinity-independent, coherent scattering contributions. A line separating the sharp diffraction peaks is then drawn by connecting the minimum between well-separated crystalline reflections to a shape similar to that found (or expected) for a totally amorphous sample. The proportionality of the separated areas or amplitudes at suitably chosen angles to the crystalline or amorphous fraction is then determined by calibration. Much of the error introduced by integrating the diffraction intensity with respect to θ or using amplitudes at one angle rather than evaluating the integral over all reciprocal space is minimized by the calibration.[35–40]

5.4 Application of X-ray Diffraction to Liquid Crystalline Polymers

Interest in liquid crystalline polymers has grown rapidly both in academia and industry since the late 1970s.[41–44] Like solid-state crystals, liquid crystal structures can also be identified by X-ray diffraction techniques. More than 30 phases have now been identified based on different orders and symmetries in these phases. Figure 13 shows several of the simplest liquid crystal phases (for simplicity, discotic liquid crystals and chiral liquid crystals have been omitted). As the phases are defined by positional, bond orientational, and molecular orientational order, each phase exhibits its own diffraction characteristics identifiable by X-ray. For instance, Figure 14(a) shows a schematic drawing of the smectic A (S_A) phase, each cylinder representing a mesogenic repeating unit. It is obvious that in the S_A phase, all the molecules are aligned towards one direction and there is a layer structure perpendicular to the molecular chain. Within each layer, the molecular lateral packing has liquid-like short-range order. The corresponding uniaxially oriented X-ray pattern and unoriented powder pattern are shown in Figure 14(b). There is a pair of sharp diffractions on the meridian in the low-angle region, which is attributed to the layer structure. A pair of diffuse arcs on the equator

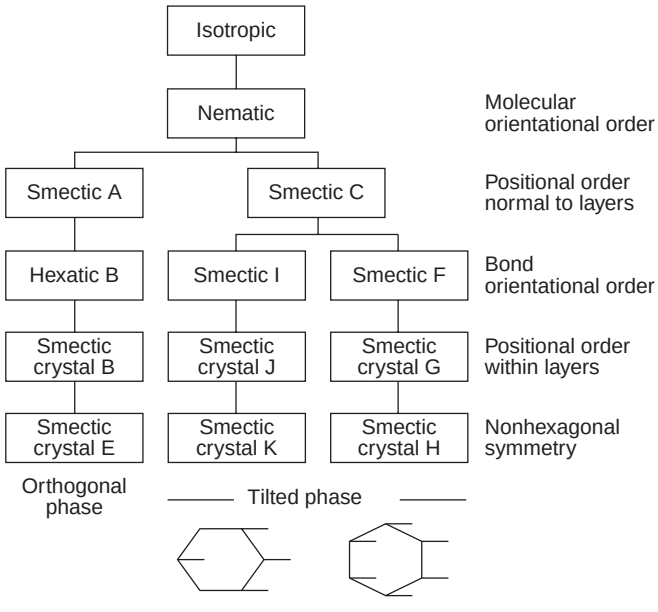

Figure 13 Liquid crystal phases from the nematic phase to highly ordered smectic phases. Only the 12 simplest phases are shown here.

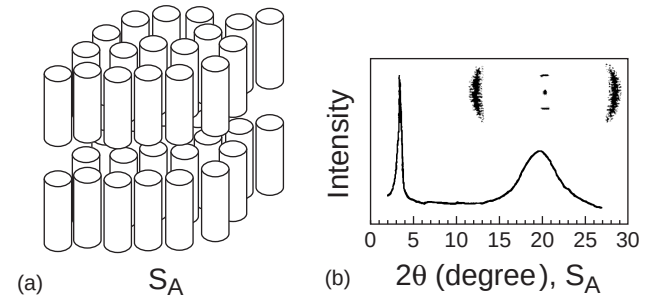

Figure 14 Diagram of the S_A phase (a) and its corresponding X-ray pattern (b).

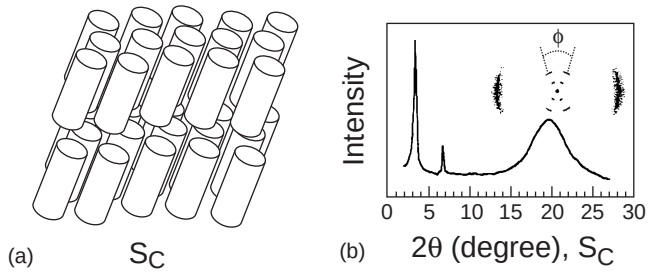

Figure 15 Diagram of the S_C phase (a) and its corresponding X-ray pattern (b).

in wide-angle region correspond to the intermolecular distance within the layer (scattering halo).

Figure 15(a) is a diagram of the smectic C (S_C) phase. The structural features are very similar to those in the S_A phase except that the chain direction is tilted with respect to the layer normal. The resulting uniaxially oriented and unoriented X-ray patterns are shown in Figure 15(b); the sharp diffractions of the layer structure tilted away from the meridian of the oriented direction can be seen, which illustrates the tilt of the molecular chain with respect to the layer normal. With increasing the bond orientational and positional orders, a series of highly ordered liquid crystal phases is approached. Figure 16(a) is a schematic drawing of a smectic F (S_F) phase and its corresponding unoriented X-ray pattern is shown in Figure 16(b) (the uniaxially oriented X-ray pattern varies in the S_F phase depending on the orientation of chain direction with respect to the oriented axis). The difference between S_C and S_F lies in that, in the S_F phase, hexagonal bond orientational order exists within the layer. By increasing the bond orientational order in S_F to positional order, a smectic G (S_G) phase is obtained as shown in Figure 17. Further reduction of the symmetry, to orthorhombic packing of the chain instead of hexagonal packing, changes the S_G to a smectic H (S_H) phase (Figure 18). The S_F, S_G, and S_H phases compose one subcolumn of Figure 13. Other subcolumns of the phases in this figure are similar to this case in the manner of their increasing order. The

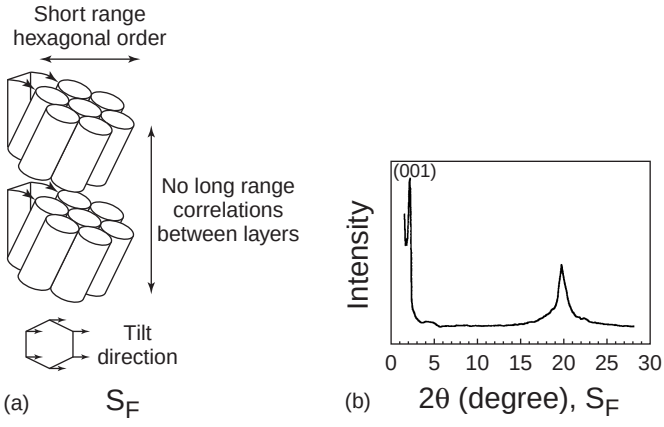

Figure 16 Diagram of the S_F phase (a) and its corresponding X-ray pattern (b).

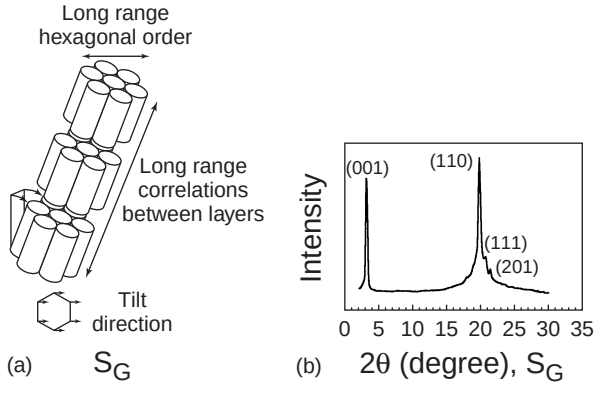

Figure 17 Diagram of the S_G phase (a) and its corresponding X-ray pattern (b).

differences are due to the direction of molecules with respect to the layer normal. More detailed discussions are available.[41–44]

In the liquid crystal phase transition study, differential scanning calorimetry (DSC) usually gives information about the transition temperatures and other thermodynamic transition properties of the liquid crystal phases, but DSC results cannot provide the structural information. In this case, X-ray experiments combined with DSC (and also polarized light microscopy) are a powerful approach to obtaining full information about the liquid crystal phase. An example is given in Figure 19[45] which shows nonisothermal DSC thermograms of a main-chain liquid crystal polyether at a scanning rate of 10 °C min^{-1}. It is obvious that there are three peaks at around temperatures of 147 °C, 129 °C, and 118 °C. The DSC itself shows that there are at least three phase transitions. The insets in this figure are the uniaxially oriented and unoriented WAXD (wide angle X-ray diffraction) patterns at 147 °C, 130 °C, and 115 °C. Detailed analyses show

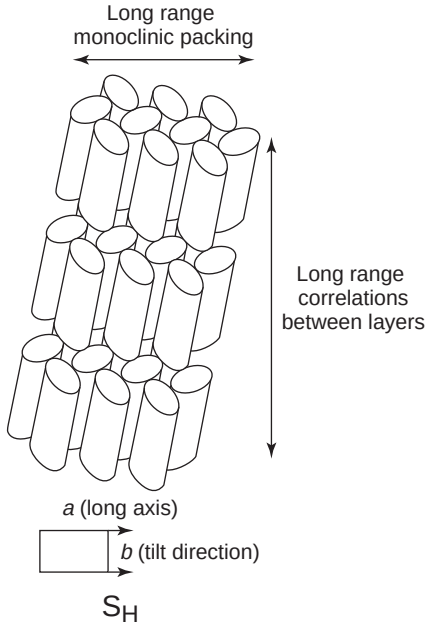

Figure 18 Diagram of the S_H phase.

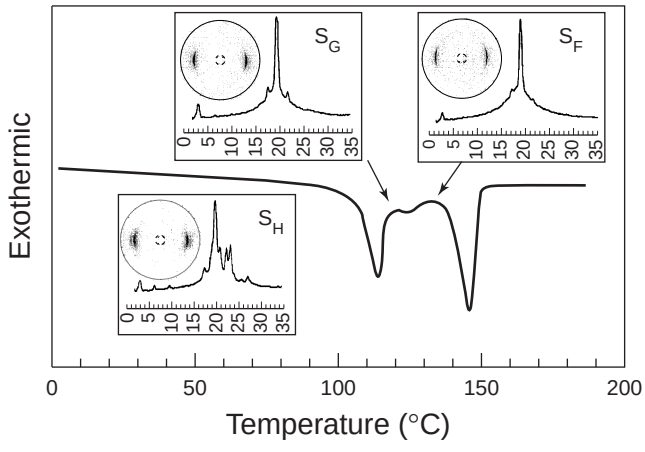

Figure 19 DSC and X-ray unoriented and uniaxial oriented pattern for a polyether showing three different liquid crystal phases at different temperatures. The polyether was synthesized from 1-(4-hydroxy-4′-biphenyl)-2-(4-hydroxyphenyl)propane and 1-15 dibromopentadecane.

that three patterns correspond to the S_F, S_G, and S_H phases.

6 SMALL ANGLE X-RAY SCATTERING OF POLYMERS

SAXS in most polymers results from the electron density distribution being heterogeneous over distances that are large in comparison with the wavelength of the X-rays. The correlation length of these heterogeneities is of the order of 5–100 nm. In homopolymers they arise from differences in mass density, which result from the partial crystallization of these substances; in copolymers or in polymeric solutions they may be caused by local differences in chemical composition.[10,12,13,16,46]

The study of SAXS is based on the same principles as the study of X-ray diffraction in the wide-angle region discussed above. The separate treatment of this phenomenon is needed because the dimensions covered by SAXS relate to supramolecular structure and phase morphology, rather than with the molecular structure of the polymers, and in many cases requires a different approach. The instrumentation also differs in many ways from that used in X-ray crystal diffraction, which often leads to the special status of SAXS in the laboratory. As such, SAXS may be used in other branches of science, such as biology, metallurgy, and colloid chemistry; however, its use in polymer science is one of the largest areas of application, which has led to a formidable amount of information on the supramolecular structures and phases in these materials.

It is to be pointed out that only in a very few cases does SAXS lead to final conclusions by itself. In almost all applications the interpretation of SAXS diagrams depends on the availability of a model of the scattering structure, which is usually based on independent sources of information such as electron microscopy. Thus, in semicrystalline polymers a model, consisting of lamellar-shaped crystals dispersed in an amorphous matrix, is assumed; on the basis of this assumption quantitative details of the morphology can be obtained by SAXS, often with remarkable accuracy.

The treatment of the raw SAXS data usually starts with a correction for the empty cell scattering (called raw data correction) followed by background subtraction. For instruments having a slit collimation system, desmearing is necessary. The Lorenz correction, correlation function, absolute intensities and invariance can all be calculated from the experimental data.[10,16,46]

6.1 Two-phase Model: Porod's Law

In many applications of SAXS to solid polymers, these materials may be considered as two-phase structures, consisting of two or more regions having different electron densities, such as separate crystalline and amorphous regions. In two-phase structures, the specific surface is the most characteristic quantity. This can be obtained from the SAXS curve by observing the way in which the intensity decreases in the region, which is generally indicated as the tail of the curve. Porod[47,48] showed that for an isotropic two-phase structure with a sharp phase boundary Equation (31) holds

$$\lim_{s\to\infty} i = \frac{(\rho_c - \rho_a)^2 S/V}{8\pi^3 s^4} \tag{31}$$

This is Porod's law. The corresponding Equation (32) for the absolute smeared intensity i_s

$$\lim_{s\to\infty} i_s = \frac{(\rho_c - \rho_a)^2 S/V}{16\pi^2 s^3} \tag{32}$$

Provided that $\rho_c - \rho_a$ is known, the specific surface can be obtained from the slope of the line through the data points in a plot of i versus $1/s^4$, or of i_s versus $1/s^3$ of which an example is shown in Figure 20. According to Equation (32) this line should go through the origin. If this is not the case, either the correction for the background was not done correctly, or the phase boundary is not sufficiently sharp. If absolute intensities are not available, the invariant Q may be used to obtain the specific surface (Equations 33–35):

$$Q \equiv 4\pi \int_0^\infty s^2 i(s)\,\mathrm{d}s \tag{33}$$

$$\lim_{s\to\infty} I = \frac{\int_0^\infty s^2 I(s)\,\mathrm{d}s\; S/V}{2\pi^2\varphi(1-\varphi)s^4} \tag{34}$$

$$\lim_{s\to\infty} I_s = \frac{\int_0^\infty s I_s\,\mathrm{d}s\; S/V}{8\pi\varphi(1-\varphi)s^3} \tag{35}$$

The name "invariant" derives from the fact that its value depends only on the mean-squared electron density fluctuation in the sample and not on particulars of how the fluctuations are distributed. The method using the invariant to obtain the specific surface has its own shortcomings because an integral must be evaluated over an s range from zero to infinity.

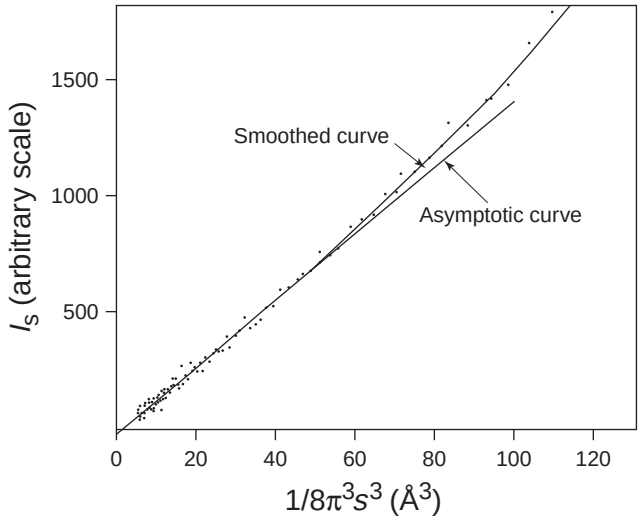

Figure 20 Plot of the smeared intensity I_s versus $1/8\pi^3 s^3$ for a sample showing a true two-phase structure (Pd particles). The small negative intercept of the asymptotic curve from the vertical axis indicates the occurrence of a minor error in the position of the background.

Very few experimental systems are found to follow the expressions corresponding to Porod's ideal two-phase system, and plots of $s^4 I(s)$ versus s or (s^2) are found to increase slightly or decrease slightly with s. These observed deviations are referred to as being positive and negative, respectively. However, more than one structural variation from Porod's assumptions may be present for any given sample, some of which cause positive deviations and some of which cause negative ones. The four causes of deviations are fluctuations in electron density within a phase, interface roughness, interface curvature, and gradual electron density variation across the interface. Generally speaking, detailed study of these deviations for polymer systems attempts to minimize the influence of three effects on the data so that the magnitude of the fourth may be inferred.

6.2 Particle Scattering: Guinier Plot

Particle scattering here applies to polymer systems containing regions of colloidal size, which are dispersed in a matrix of a different electron density. Such systems may be found in di- and tri-block copolymers resulting from phase separation. Examples can also be found in an initial stage of polymer crystallization and polymer solutions. In all these cases, if the electron density inside the particle is uniform, the system will in general be of the two-phase type, as discussed previously. Guinier et al.[46] have shown that one may obtain a characteristic quantity regarding the particles from the scattering near the origin if the shape and/or the electron density distribution of the particles is not known. For dilute systems, Equation (36) holds for small values of the scattering vector:

$$i(s) = \frac{Nm^2}{V} \mathrm{e}^{-4\pi^2 s^2 R_g^2/3} \tag{36}$$

where m is the number of excess electrons per particle, which is the number of electrons in the particle minus the number in the same volume of the matrix. R_g is the radius of gyration of the particle (Equation 37):

$$R_g^2 = \frac{1}{m} \int_V r^2\,\mathrm{d}m \tag{37}$$

where r is the distance of the excess electrons $\mathrm{d}m$ to the electronic center of mass of the particle. This equation is analogous to the one defining the radius of gyration in mechanics, except that in the latter case the distances are measured with respect to an axis of rotation, whereas here they refer to the center of mass. The relationships for mechanics therefore cannot he used here. A compilation

of the values for various shapes, which have to be used in connection with SAXS, is presented in the International Tables for X-ray crystallography.

A plot of $\log i(s)$ versus s^2, usually referred to as a Guinier plot (Figure 21),[12,46] will be in the regions where Equation (36) is valid, and yields a straight line of which the slope is proportional to R_g^2, as shown in Figure 21(b). The intercept on the plot of $\log i(s)$ versus s^2 gives the quantity Nm^2/V. By combining this with data about the concentration of the material constituting the particles, from which in general the quantity Nm/V can be obtained, one may find the value of m, which is directly related to the mass of the particle. It is to be noted that the use of slit-smeared intensities i_s instead of i does not affect the slope in the Guinier plot, and will thus give correct R_g values. However, in that case the intercept can no longer be related to Nm^2/V in a simple way; here, desmearing is necessary for obtaining information on m.

As Guinier's law allows a rapid analysis of intensity data, usually without previous correction, it has found very wide application; however, it has also been much abused. This is often due to a lack of consideration of the deviations from this law (of which the main causes may be the polydispersity of the particle sizes), and by the use of too high concentrations of the particles.[10] Because of the latter effect in particular, the proper application of Guinier's law to the study of synthetic polymers is rather limited. Another improper use of this plot is that the experimental data often is limited to small angles that are not low enough to enter the Guinier region. In the case of particles having a different fixed shape, the corresponding scattering function can be found in the literature.[47,48]

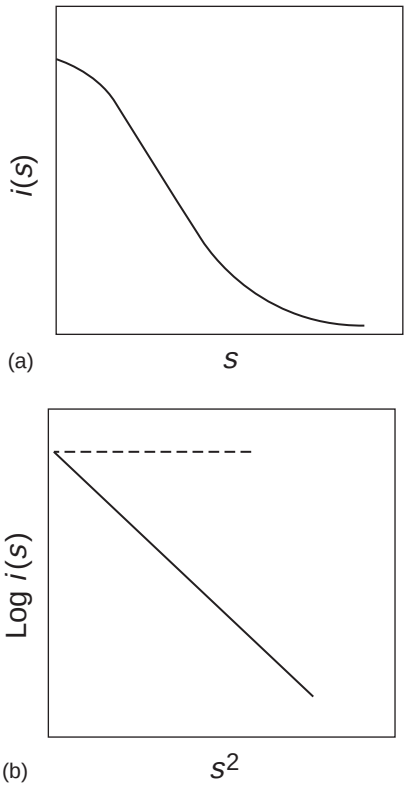

Figure 21 The Guinier plot: (a) scattering intensity versus s, (b) plot of $\log i(s)$ versus s^2.

6.3 Analyses of Oriented Polymers

The study of the morphology of oriented polymers is of great interest to the industry of polymer products. Moreover, the two-dimensional X-ray diffraction pattern of oriented polymers contains more information than the simple one-dimensional diffraction curve of an isotropic sample, and as such may also be studied for gaining an understanding about fundamental polymer properties.[10,12,13] Figure 22 shows a set of common SAXS patterns and their corresponding supramolecular structure and phase morphology in direct space. Figure 22(a) is the statistically spherically symmetrical assemblies of crystallites, such as spherulites or unoriented assemblies of stacked lamellar crystals. Figure 22(b) is the stacked lamellar crystal. Depending on the number and breadth of the texture, the scattering pattern will be different. If the stacked lamellar crystals are inclined with respect to the fiber direction, different scattering patterns will be observed (Figure 22c) If we superimpose two sets of Figure 22(c) with a mirror symmetry relationship between them, another type of scattering pattern appears (Figure 22d). Double oriented structures usually result in scattering patterns like Figure 22(e).[10,12,13]

6.4 X-ray Diffraction in Thin Films

X-ray diffraction techniques have been used extensively to determine the monolayer thickness of thin films such as Langmuir–Blodgett (LB) films.[49] Most of the work has concentrated on long-chain (saturated and unsaturated) fatty acids;[50,51] however, long-chain esters, substituted aromatics, preformed polymers, and biological materials[52] have also been investigated. Multilayer assemblies have even been used for gratings in X-ray spectroscopy systems. As the scattering of X-rays from carbon and hydrogen atoms can be assumed to be very small compared to that from the heavier metal ions, the lattice spacing (normal to the film) measured by X-ray diffraction for simple fatty acids corresponds to the distance between adjacent planes containing metal ions. This led originally to the discovery that some X-deposited films possessed essentially the same structure as Y-type layers, and vice versa. X-deposited films and Y-type layers are different ways for LB film deposition. For a more detailed explanation see Klug and Alexander.[19]

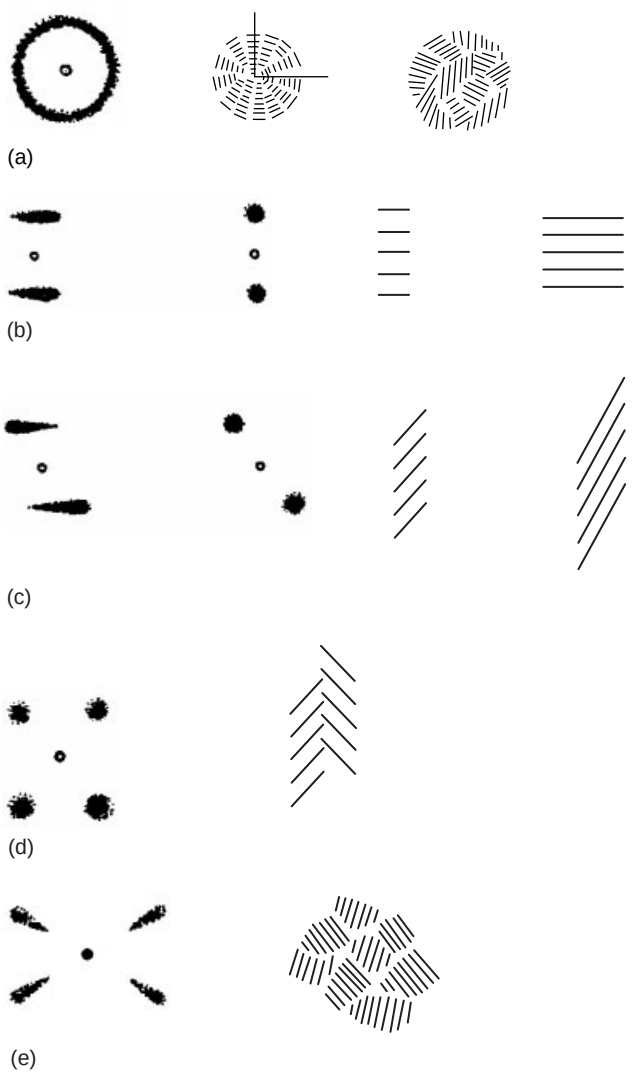

Figure 22 A set of common SAXS patterns and their corresponding supramolecular and phase morphology in the direct space.

Figure 23 shows X-ray diffraction data from a film of 43 layers of perdeuterated manganese stearate on a substrate of single-crystal silicon.[53] The experimental values are shown as points; the solid curve is based on calculation and is displaced from the data points to avoid obscuring them. Many orders of Bragg peaks from the basal planes (001) are clearly visible, giving information about the stacking of the planes of LB layers. However, under certain circumstances it is possible to obtain details of the in-plane structure using X-ray techniques. X-ray studies are usually carried out on multilayer films, although it is possible to observe diffraction data from a single monolayer. For small (<20) numbers of layers, secondary maxima are often observed between the Bragg reflections;[53] the intensity and angular position of these depend on the exact number of deposited layers. This

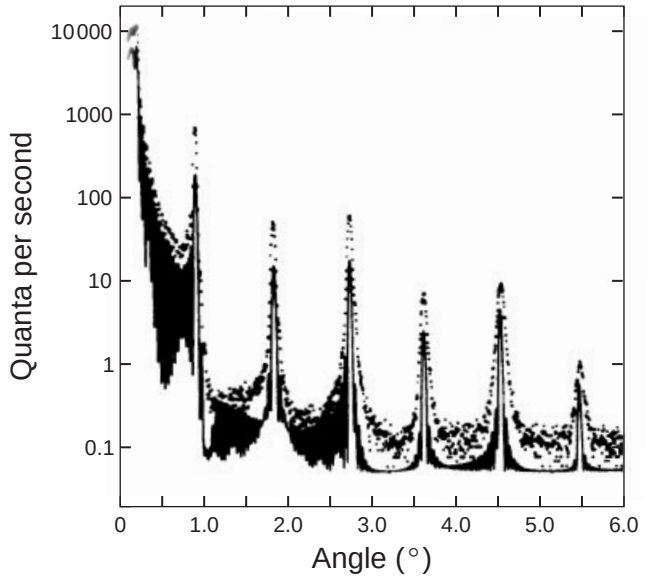

Figure 23 X-ray diffraction from a film of 43 layers of perdeuterated manganese stearate on a Si substrate. Experimental values are shown as points. The solid curve is based on calculation and is displaced from the data points.

phenomenon is analogous to the diffraction of light waves from gratings of only a few slits. A detailed analysis of the subsidiary maxima has been undertaken by Pomerantz and Segmuller.[54] Kapp and Wainfan[55] have reported a further set of X-ray fringes in the nonspecularly scattered radiation from a barium stearate multilayer film; however, no unambiguous explanation of these fringes could be offered. The *d* spacings obtained for simple long-chain fatty acid materials are generally found to correspond to those from known crystalline modifications. Where it has been possible to directly compare the X-ray *d* spacing to values obtained from optical techniques, good agreement within experimental error is usually obtained.[56] Fatty acid films of mixed (within the layer) material have also been investigated using X-ray technique. In one case an intermediate *d* spacing between the individual fatty acids was obtained; but in the other the *d* spacing appeared to be determined by just one of the components. As well as the determination of film thickness, X-ray diffraction data may be used to obtain electron density profiles across multilayer films.[57]

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) (DMR 96-17030), the NSF Science and Technology Center for Advanced Liquid Crystalline Optical Materials (ALCOM) at Kent State University, The University of Akron and Case Western Reserve

University and the NSF/EPIC/Industry Center for Molecular and Microstructure Composites (CMMC) at Case Western Reserve University and The University of Akron.

ABBREVIATIONS AND ACRONYMS

DSC	Differential Scanning Calorimetry
LB	Langmuir–Blodgett
SAXS	Small Angle X-ray Scattering
S_A	Smectic A
S_C	Smectic C
S_F	Smectic F
S_G	Smectic G
S_H	Smectic H
WAXD	Wide Angle X-ray Diffraction

RELATED ARTICLES

Nucleic Acids Structure and Mapping **(Volume 6)**
X-ray Structures of Nucleic Acids

Particle Size Analysis **(Volume 6)**
Diffraction in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
X-ray Crystallography of Biological Macromolecules

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships

Pulp and Paper **(Volume 10)**
X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
X-ray Methods for the Study of Electrode Interaction

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Absorption Techniques in X-ray Spectrometry • Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Structure Determination, X-ray Diffraction for • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. C.H. Macgillavry, G.D. Rieck (eds.), *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham, UK, Vol. III, 1962.
2. J.A. Ibers, W.C. Hamilton (eds.), *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham, UK, Vol. IV, 1974.
3. G.H. Stout, L.H. Jensen, *X-ray Structure Determination A Practical Guide*, 2nd edition, John Wiley & Sons, New York, 1989.
4. E.B.M. Martin, *A Guide to the Safe Use of X-Ray Diffraction and Spectrometry Equipment*, Science Reviews, Leeds, UK, 1983.
5. C.S. Barrett, J.V. Gilfrich, R. Jenkins, D.E. Leyden, J.C. Ross, P.K. Predecki (eds.), *Advances in X-ray Analysis*, Plenum, New York, Vol. 30, 1987.
6. A. Guinier, *X-ray Diffraction*, W.H. Freeman and Company, San Francisco, 1963.
7. J.S. Kaspar, K. Lonsdale (eds.), *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham, UK, Vol. II, 1959
8. R. Hosemann, S.N. Bagchi, *Direct Analysis of Diffraction by Matter*, North Holland Publishing Company, Amsterdam, 1962.
9. R.W. James, *The Optical Principles of the Diffraction of X-Rays*, G. Bell and Sons, London, 1984.
10. F.J. Balta-Calleja, C.G. Vonk, *X-Ray Scattering of Synthetic Polymers*, Elsevier Publishing Company, Amsterdam, 1989.
11. B.K. Vainshtein, *Diffraction of X-Rays by Chain Molecules*, Elsevier Publishing Company, Amsterdam, 1966.
12. M. Kakudo, N. Kasai, *X-Rays Diffraction by Polymers*, Elsevier Publishing Company, Amsterdam, 1972.
13. L.E. Alexander, *X-Ray Diffraction Methods in Polymer Science*, Wiley-Interscience, New York, 1969.
14. Rontgen, stizber Wurzburger physick. *Med. Ges.*, 1895
15. Friedrich, Knipping, Von Laue, *Bayer Acad. Wiss.*, **303** (1912).
16. O. Glatter, O. Kratky, *Small Angle X-Ray Scattering*, Academic Press, London, 1982.
17. N.F.M. Henry, K. Lonsdale (eds.), *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham, UK, Vol. I, 1952.
18. M.J. Buerger, *X-Ray Crystallography*, Wiley, New York, 1942.
19. H.P. Klug, L.E. Alexander, *X-Ray Diffraction Procedures for Polycrystalline and Amorphous Materials*, 2nd edition, John Wiley & Sons, New York, 1974.

20. H. Winick, S. Doniach (eds.), *Synchrotron Radiation Research*, Plenum Press, New York and London, 1980.
21. E.W. Nuffield, *X-Ray Diffraction Methods*, Wiley, New York, 1966.
22. M.J. Buerger, *The Precession Method in X-Ray Crystallography*, Wiley, New York, 1964.
23. C.W. Bunn, *Chemical Crystallography: An Introduction to Optical and X-Ray Methods*, 2nd edition, Oxford University Press, 1961.
24. M.M. Woolfson, *Direct Methods in Crystallography*, Oxford University Press, 1961.
25. *Tables of Interatomic Distances and Configuration in Molecules and Ions*, The Chemical Society, London, 1958.
26. P. Corradini, I.W. Bassi, 'Crystal Structure of Isotactic Poly(vinyl methyl ether)', *J. Poly. Sci. Part C*, **16**, 3233–3244 (1968).
27. M. Eashoo, Z. Wu, A. Zhang, D. Shen, C. Wu, F.W. Harris, S.Z.D. Cheng, K.H. Gardner, B.S. Hsiao, 'High Performance Aromatic Polyimide Fibers. 3. A Polyimide Synthesized from 3,3′,4,4′-Biphenyltetracarboxylic Dianhydride and 2,2′-Dimethyl-4,4′-Diaminobiphenyl', *Macromol. Chem. Phys.*, **195**, 2207 (1994).
28. J.J. Ge, A. Zhang, K.W. McCreight, S.-Y. Wang, X. Jin, F.W. Harris, S.Z.D. Cheng, 'Phase Structures, Transition Behaviors and Surface Alignment in Polymers Containing Rigid-rod-like Backbones with Flexible Side Chains. 1. Monotropic Phase Behavior in a Main-chain/Side-chain Liquid Crystalline Polyester', *Macromolecules*, **30**, 6498–6506 (1997).
29. B. Wunderlich, *Macromolecular Physics*, Academic Press, New York, Vol. I, 1973.
30. W. Ruland, 'X-ray Determination of Crystallinity and Diffuse Disorder Scattering', *Acta Cryst.*, **14**, 1180–1188 (1961).
31. W. Ruland, A. Dewaelheyns, 'Electronic Device for the Spherucal Averaging of X-ray Fiber Diagrams', *J. Sci. Instrum.*, **44**, 236–251 (1967).
32. C.R. Desper, R.S. Stein, 'Randomization of Films and Fibers', *J. Polym. Sci. Part B*, **5**, 893–900 (1967).
33. W. Ruland, 'Rontgenographische Bestimmung der Temperaturebhangigkeit des Ordnungszustandes von Polyathylen', *Faserforsch. Textiltechnik*, **18**(2), 59–63 (1967).
34. J.W. Turley, *X-Ray Diffraction Patterns of Polymers*, The Dow Chemical Company, Midland, Michigan, 1965.
35. J.M. Goppel, 'On the Degree of Crystallinity of Natural Rubber. I. An Improved Method to Determine the Degree of Crystallinity in Rubber. II. The Orientation of the Rubber Crystallinities in Stretched Samples', *J. Appl. Sci. Res. A1*, **3**, 18–26 (1947).
36. S. Krimm, A.V. Tobolsky, 'Quantitative X-ray Determination of Crystallinity in Polyethylene', *J. Polym. Sci.*, **7**, 57–76 (1951).
37. P.H. Hermans, A. Weidinger, 'X-ray Studies on the Crystallinity of Cellulose', *J. Polym. Sci.*, **4**, 135–144 (1950).
38. P.H. Hermans, A. Weidinger, 'On the Determination of the Crystalline Fraction of Polyethylenes from X-ray Diffraction', *Makr. Chem.*, **24**, 44–46 (1961).
39. P.H. Hermans, A. Weidinger, 'On the Determination of the Crystalline Fraction of Isotactic Polypropylene from X-ray Diffraction', *Makr. Chem.*, **50**, 115 (1961).
40. H. Hendus, G. Schnell, 'Rontgenographische und IR-spektroskopische Krostallinitatsbestimmung an Polyathylen', *Kunstst*, **51**, 69–74 (1961).
41. S. Chandrasekhar, *Liquid Crystals*, 2nd edition, Cambridge University Press, New York, 1992.
42. P.G. de Gennes, *The Physics of Liquid Crystals*, 2nd edition, Oxford University Press, Oxford, 1993.
43. G.W. Gray, J.W. Goodby, *Smectic Liquid Crystals: Textures and Structures*, Leonard Hill, Glasgow, 1984.
44. P.S. Pershan, *Structure of Liquid Crystal Phases*, World Scientific Publishing, Singapore, 1988.
45. Y. Yoon, A. Zhang, R. Ho, S.Z.D. Cheng, V. Percec, P. Chu, 'Mesophase Identifications in a Series of Liquid Crystalline TPP Polyethers and Copolyethers Having Highly Ordered Mesophase Structures. 1. Phase Diagrams of Odd-numbered Polyethers', *Macromolecules*, **29**, 294–305 (1996).
46. A. Guinier, G. Fournet, *Small Angle Scattering of X-Rays*, John Wiley & Sons, Chapman & Hall, 1955.
47. G. Porod, 'Die Abhangigkeit der Rontgenkleinwinkelstreuung von form und Grosse der Kolloiden Teilehen in ver Dunnten Systemen IV', *Acta Phys. Austriaca*, **2**, 255–292 (1948).
48. G. Porod, 'Theorie der diffusen tontgenkleinwinkelstreuung an kolloiden systemen', *Z. Naturforsch.*, **A-4**, 401 (1949).
49. G. Robrets, *Langmuir–Blodgett Films*, Plenum Press, New York, 1990.
50. C. Holley, 'X-ray and Optical Measurements of Multimolecular Films', *Phys. Rev.*, **51**, 1000–1007 (1937).
51. M. Von Frieling, H. Bradaczek, W.S. Durfee, 'X-ray Diffraction of Langmiur–Blodgett Films: Establishing a New Method for the Calculation of Electron Density Distributions from a Single Set of Intensity Data', *Thin Solid Films*, **159**, 451 (1988).
52. R.H. Tredgold, A.J. Vickers, A. Hoorfar, P. Hodge, E. Khoshdel, 'X-ray Analysis of Some Porphyrin and Polymer Langmuir–Blodgett Films', *J. Phys. D*, **18**, 1139–1149 (1985).
53. R.M. Nicklow, M. Pomerantz, A. Sefmuller, 'Neutron Diffraction from Small Numbers of Langmuir–Blodgett Monolayers of Manganese Stearate', *Phys. Rev. B*, **23**, 1081–1087 (1981).
54. M. Pomerantz, A. Segmuller, 'High Resolution X-ray Diffraction from Small Numbers of Langmuir–Blodgett Layers of Manganese Stearate', *Thin Solid Films*, **68**, 33–45 (1980).

55. D.S. Kapp, N. Wainfan, 'X-ray Interference Structure in the Scattered Radiation from Barium Stearate Multilayer Films', *Phys. Rev.*, **138**, 1490–1495 (1965).
56. V.K. Srivastava, A.R. Verma, 'Interferometric and X-ray Diffraction Study of ''Built-up'' Molecular Films of Some Long Chain Compounds', *Solid State Commun.*, **4**, 367–371 (1966).
57. B. Belbeoch, M. Roulliary, M. Tournarie, 'Evidence of Chain Interdigitation in Langmuir–Blodgett Films', *Thin Solid Films*, **134**, 89–99 (1985).

Process Analysis: Introduction

James W. Peterson
The Boeing Company, Seattle, USA

Alan H. Ullman
Procter & Gamble Company, Cincinnati, USA

1 Introduction 1
2 Different Approaches to Process Analytical Measurements 2
3 Techniques 5
3.1 Titration and Flow Injection Techniques 5
3.2 Chromatographic Techniques 7
3.3 Spectroscopic Techniques 7
3.4 Sensor-based Techniques 10
3.5 Other Techniques 11
3.6 Technique Selection 11
3.7 Summary 12
4 Sampling and Sampling Preparation Issues 12
5 Data Treatment 14
6 Business and Ownership Issues 15
6.1 Business Case 15
6.2 Initial Costs 15
6.3 Recurring Costs 16
6.4 Ownership 16
Abbreviations and Acronyms 17
Related Articles 17
Further Information 18
I Books on Process Analytical Chemistry 18
II Journals and Magazines with a Significant Coverage of Process Analytical Chemistry 18
III Conferences with a Significant Coverage of Process Analytical Chemistry 18
IV Short Courses on Process Analytical Chemistry 18
V Guides to Manufacturers of Process Analytical Chemistry Instrumentation 18
VI Project Management Resources 19

1 INTRODUCTION

The use of chemical processes in industry has always required measurement techniques which can discern whether or not the processes are producing the desired results. For many years laboratory-based wet chemical analysis of samples taken from the process was the only way to make these measurements. Since the late 1970s we have witnessed the rapidly accelerating development of technologies which enable chemical analysis tools to move out of the laboratory and into the production environment where they are used to make decisions and control processes. The application of chemical analyses directly to manufacturing processes and process control has come to be known as process analytical chemistry (PAC).

So, what is PAC specifically, and how is it different from any other analytical approach?

PAC is the tailored application and support of chemical analysis techniques and equipment to allow their use directly on processes in a manufacturing environment. The fundamental methods of analysis used are no different from those that would be used in the laboratory. The difference lies in adapting the methods and instrumentation to produce the desired information reliably while operating with little or no human intervention in what is usually a less than ideal environment.

Sometimes the adaptation is straightforward. A typical laboratory device may be appropriate. An ion-specific electrode, for instance, might be mounted directly in a protected location in a chemical processing tank to provide continuous real-time information needed for controlling that process. In other cases, extensive modification may be needed. Perhaps a custom-designed analyzer with an automated sampling system may be required, or a special fiber-optic probe may be needed to gather in situ spectroscopic data, which may then require chemometric analysis to interpret. As demonstrated later in this article, even for simple cases like an ion-specific electrode, the on-going support of the installed equipment can create complications. Maintenance and calibration are critical considerations for a reliable application of PAC technology. Without reliability, PAC is not worth much.

What is the driving force behind industry's use of PAC? Why should we bother to move chemical analyses out of the laboratory?

As in nearly all industrial initiatives, the driver is economics. By reducing or eliminating the time required for process sampling and sample transport, PAC methods produce more information in a shorter overall time than is possible with laboratory-based operations. Direct labor and other cost savings can often

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

be achieved. More importantly, the increased volume and quality of information delivered on a more timely basis results in improved process control, increasing the yield and quality of the product and reducing process waste. The availability of more comprehensive process data also allows process engineers and developers to understand the process mechanisms and dynamics more completely. This enables further optimization of the process with associated gains in productivity and quality, and sometimes leads to the development of entirely new processes with even greater productive capability.

With the improved capability of laboratory instrumentation, is it not possible to achieve the same results without putting expensive instrumentation in the production plant environment?

This question must be carefully addressed when considering the use of PAC for a process. Modern laboratory instruments are generally more capable and flexible than PAC equipment. A central laboratory may also have the capability to perform all the types of analysis needed in a chemical production facility; it may be more flexible in responding to some types of change in process requirements; and a single laboratory may involve lower initial capital costs than distributed analyzers. However, the logistics and economics of sampling and transport prevent a laboratory from meeting the needs of processes which require analyses in large numbers, or very frequently, or with short or zero delay times. These information needs are usually driven by process control requirements for complex or rapidly changing processes.

The primary benefits of PAC, and the reasons to consider its use, are that it provides chemical process data in higher volumes, at higher frequency, and with less delay (often in real time) than is possible or economical for a laboratory. An additional benefit is often derived from the short or zero delay times. If sample degradation is a problem, PAC methods can improve the quality of the analysis (or even make an analysis possible) by minimizing sample-to-analysis time or by eliminating sampling altogether with an in situ measurement.

In some cases just one of these benefits may make PAC methods an absolute requirement for controlling a process. In other situations, it may be a combination of several benefits which provides the justification for PAC. And sometimes PAC may not be the best answer, despite its benefits. In all cases where PAC is desirable, though, the keys to a successful implementation are found in these five foundation elements:

- A clear definition of the requirements which the PAC system is to satisfy.
- Careful selection of the analytical technique, sampling method, and the instrumentation to be used.
- Thorough consideration of system reliability and planning for calibration, failure detection, and back-up systems.
- A solid business justification for the system.
- Committed ownership of the system for on-going maintenance.

The remainder of this article, and the other articles in this section of this Encyclopedia, will provide information and resources to help address one or more of these foundation elements.

2 DIFFERENT APPROACHES TO PROCESS ANALYTICAL MEASUREMENTS

Now that we understand the reasons for performing PAC, it is appropriate to discuss the different approaches it may take. Selection of the approach and technique should follow the general problem-solving approach outlined in **Analytical Problem Solving: Selection of Analytical Methods**. For a process analytical problem, we have found it convenient to modify slightly the flow chart discussed there. The modified flow chart is shown in Figure 1.

As you will recall, the first step in approaching a problem is to define it. In defining the problem, a variety of issues must be considered. These include:

- Is the PAC result needed for closed-loop control?
- What is the response time of the process?
- What kind of information is needed?
- Is specific chemical information needed or can the result be inferred from a less specific, but simpler or cheaper or more robust measurement?
- Does the value of the requested measurement justify a high maintenance approach?
- What is the matrix of the sample stream?
- What are the conditions of the sample stream?

As you gather the information needed to answer these questions (and many more), you will be moving towards the first decision: should the approach used be an off-line, at-line, on-line, or in-line one?

The *off-line* approach is the common ''grab a sample and send it to the laboratory for analysis'' style of analytical chemistry. Other than our natural concern that the sample be properly taken from the safety, homogeneity, and so on, standpoints, this is outside the scope of PAC, but it is important that the PAC practitioner understands what the official laboratory methods are. These methods are the baseline against which the PAC results will be judged, and will usually be assumed to be ''correct''.

The *at-line* approach is the overlap zone between laboratory and PAC. Samples are taken manually from

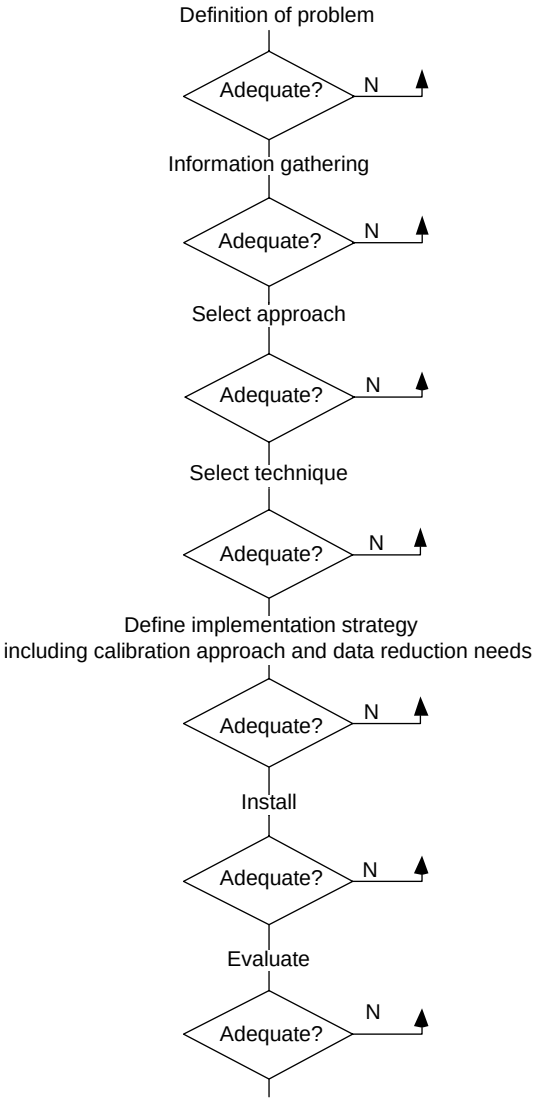

Figure 1 An approach to solving a PAC problem.

the process, but the analysis step is performed ''at the line or process'', perhaps in the control room. Analytical techniques used here range from the simplest to the most complex ones, but are usually straightforward and rapid. In a control room, the personnel are not normally trained analytical technicians and they are busy with their primary responsibility, supervising the process. They need measurements which provide them with information about the process, but these measurements cannot take long to run, or require extraordinary laboratory skills. Figure 2 depicts a typical at-line application of ultraviolet (UV)/visible spectrometry.

So, why choose to perform tests ''at-line''? Perhaps the process has a response time of hours or the sample stream is not amenable to ''simple'' process analysis techniques because it is highly viscous or multiphase. In such cases it

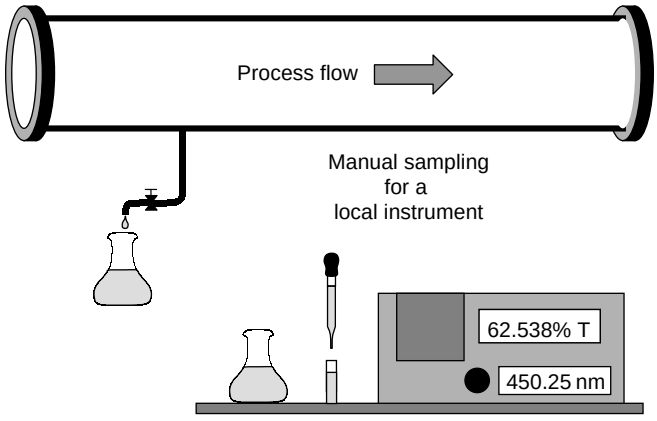

Figure 2 At-line UV/visible measurement.

is sometimes simpler to set up procedures for an operator to take a sample at some interval and perform the test manually. The biggest advantage of at-line analysis is that sampling and measurement are performed manually. This is also the biggest disadvantage – while a human being can handle complex sample preparation steps or complicated instrumental or wet chemical procedures, do they have the time?

The *on-line* approach encompasses any technique in which the measurement is directly coupled to the process, and the sample is withdrawn in some way by the instrument. Figure 3 shows an on-line flow injection analyzer connected to a slip stream (or side stream) of the process.

After appropriate sample conditioning with a filter, which will be discussed later, the instrument analyzes a small portion of the sample stream on a periodic basis. In many cases, it is advantageous to divert a portion of the process to the analyzer. In this side stream one or more instruments can be installed. The stream flow rate, temperature, pressure, etc. can be adjusted to match instrument requirements. Should an analyzer need

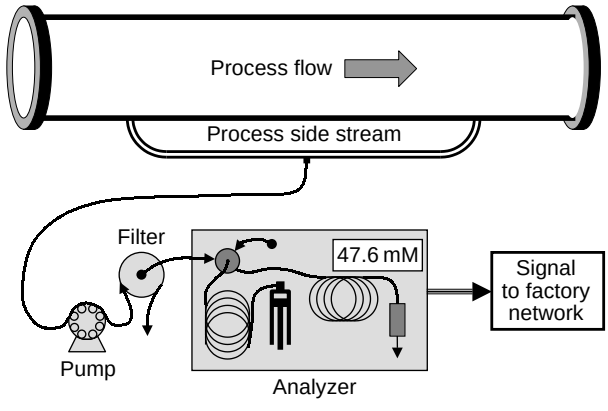

Figure 3 On-line flow injection analysis (FIA).

servicing, the flow to the side stream can be closed off and work on the instrument can be done without shutting down the process. Side streams are not always desirable, however. In sanitary systems, common in pharmaceutical processes, they add complexity to cleaning procedures. With multiphase streams, such as in the paper and pulp industry, small lines may be prone to clogging. In all cases, the slip stream must be representative of the main stream.

The difference between in-line and on-line is simple. *In-line* measurements do not remove a sample from the process stream; the sample is analyzed directly within the process. Perhaps the most common example is pH measurement. As shown in Figure 4, the pH probe is often inserted directly into the main process stream, not a slip stream. The advantage of in-line measurement is the lack of added piping, valves, and so on, to the process. (Keep it simple!) However, the entire process may have to be shut down to service the analyzer – a major disadvantage. For some techniques, including pH, special insertion valves are available which allow removal from the stream without disrupting the process.

In addition to the above terms, process analytical chemists also frequently describe methods as invasive or noninvasive. In some ways these terms overlap with in-line, on-line, and so on. Basically, an invasive technique has the potential to change the process, process stream, or sample in some way. This is not to say that it actually does change the sample, but by modifying the process to include a slip stream, or take a sample, or adjust the temperature, it may change the sample. Most PAC techniques are invasive to some degree. Noninvasive techniques, in many ways the holy grail of process measurements, are less common, but do exist. Examples include clamp-on ultrasonic flow meters, noncontact (toroidal) conductivity meters (especially the style in which the sensors are built into a spool piece of pipe), and photometers which use windows in the pipe.

By now the problem has been defined and, perhaps, redefined. Assuming that it has been concluded that continuous process data are needed, it now remains to be decided which technique is most suitable for continuous analysis and how it should be installed (in-line or on-line). This is the ''Select Technique'' part of the flow chart in Figure 1.

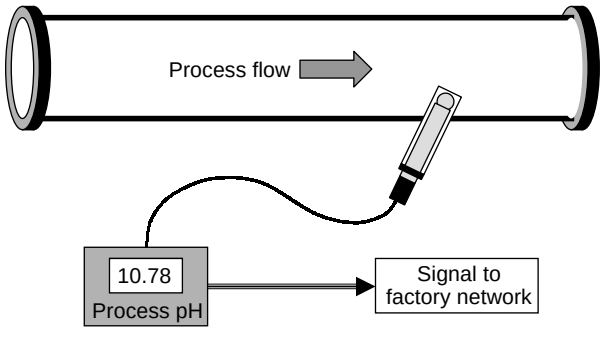

Figure 4 In-line pH measurement.

To help clarify this process, consider two hypothetical examples.

In the first example, a process control engineer calls to explain that if sugar concentration data were available more often than provided by the laboratory, the evaporation process could be improved, saving the company time and money. As this is discussed further, it emerges that the laboratory process for sugar concentration determination is refractometry. The need is for continuous data thus eliminating at-line measurement as an option, although perhaps it would be useful as an additional tool. Considering the techniques, the obvious choice is in-line refractive index. It is the same method that is used in the laboratory, it is also continuous, and not too expensive. Another option under consideration is density measurement, but then the laboratory method and the process method might disagree and cause problems.

In the second example, a different control engineer asks if ways of monitoring the overheads of a distillation process have ever been examined. If a way were found to monitor the distillate, its quality could be optimized. During information gathering you learn that the response of the distillation process is fairly slow. Adjusting the temperature takes time and the control algorithm will not make frequent changes. Since the product of the distillation is not a pure compound, you consider gas chromatography (GC) to be the best option for the analysis, but think that perhaps near-infrared (NIR) or Raman spectroscopy might work as well. A call to the laboratory confirms that GC is the standard method used and a copy of the detailed method is being e-mailed to you. You consider the advantages and disadvantages of the different techniques. GC is the laboratory technique, but is relatively slow. It will require significant piping of sample lines to the gas chromatograph, gas cylinders, fairly high maintenance, and so on. On the other hand, the plant instrument maintenance department already knows gas chromatographs and is able to service them. At this time Raman is an unknown technique to the plant, but NIR is somewhat familiar. Samples tested in the laboratory indicate that NIR might be suitable for the needed measurement, but will require many calibration samples and chemometrics. A fiber-optic system could be installed, negating the sampling issue, but the effort needed to maintain a calibration and the maintenance organization's reticence enable a final decision to be made in favor of GC.

In these two examples, issues involving sampling, laboratory analysis methods, and calibration are considered. These are critical to successful PAC application

and will be addressed further in section 4 on Sampling and Sampling Preparation Issues.

Section 3 will cover technique selection, with a broad survey of the different PAC techniques commonly used. In the flow chart of Figure 1 the ''Define Implementation Strategy'' point has now been reached.

By now a decision has been made about how to approach the PAC problem and a technique has been chosen. It is time to lay out a plan for this PAC system. How will the analyzer be installed? Will it require a minimum flow rate to assure that a homogeneous sample is presented to the analyzer? Will the instrument require sample preparation? How will the instrument be calibrated? What data processing will be needed? These issues are covered in later sections of this article, but in the PAC problem-solving process this is where you should think about them. Often, as these issues are considered, you may decide that the original choice of approach or technique needs to reconsidered and you ''cycle'' back up the flow chart. Perhaps, as the sugar application is discussed with prospective vendors, it is learned that the range covered by your process exceeds the linear range of the commercial instrument, or that your process is operated at a temperature above the limits of the analyzer. In the one case it might be decided to use a different analytical technique or it might be decided to store multiple calibrations in the process computer and apply them externally, thus ''linearizing'' the instrument's response. It is here that these issues should be fleshed out, before purchasing and installing an instrument.

The remaining steps in the flow chart cover installation and evaluation. Installation is straightforward. The approach and the technique have been chosen, the whole system including piping and data handling has been planned, and an instrument has been purchased and installed. (This was probably done in close consultation with the plant engineering and construction departments.) During commissioning the process operators have been asked deliberately to make a product with a little more variation than normal so the instrument's response and calibration can be tested. The plan for drawing laboratory samples at the appropriate time is in place and working. Data will be collected by the plant over the next several weeks and e-mailed to you for study. This is the evaluation period. The analyzer appears to be functioning correctly, but it is not ''approved''. Data are collected from both the analyzer and the laboratory and compared. Working with the control and process engineers, you determine if the data being generated meet the requirements for process control. Is it stable or does it drift? How frequently does the instrument need recalibration or maintenance? On a control chart of the difference between the laboratory and analyzer values, are there unusual events? If you and your colleagues are satisfied with this evaluation, you can safely hand the keys to the analyzer to the operations department. If not, earlier steps is the flow chart will have to be revisited.

3 TECHNIQUES

The selection of an analytical technique is interdependent with the selection of the overall PAC approach, and both depend on the operating conditions and information requirements of the process. The three PAC approaches can be ranked in terms of the degree of proximity of the measurement to the process conditions. In-line measurements are made in the process conditions, while on-line and at-line measurements are increasingly further removed from the process. In general, the closer the measurement is to the process conditions, the fewer will be the number of suitable choices for the analytical technique.

Once the process conditions and information needs are known, it may be evident that one of the PAC approaches (in-line, on-line, or at-line) is preferred, or even required. Selecting this approach may limit the range of choices for analytical techniques.

The majority of analytical techniques commonly applied to PAC can be categorized into four groups, with a fifth miscellaneous category for the remainder. The four primary groups are:

- titration and flow injection techniques (wet chemistry techniques)
- chromatographic techniques
- spectroscopic techniques
- sensor-based techniques.

Each of these categories has generic characteristics which make its members more or less suitable for particular applications.

3.1 Titration and Flow Injection Techniques

This category consists of wet chemical analysis methods which have been automated and scaled down to use small amounts of sample and reagents. Generally they involve monitoring the progress of chemical reactions of a sample with measured amounts of reagents.

Automatic titrators are very flexible and can be configured to perform nearly any titration which is done manually in the laboratory. Method development is straightforward and may be a direct application of an existing laboratory-based titration. The methods can often be readily modified to adapt to changing process requirements. A single titrator can sometimes be used to perform numerous different methods to support a

process or group of processes. On the down side, required sample volumes are usually greater than for other PAC techniques, and samples are likely to require preconditioning such as dilution prior to delivery to the titration vessel. Most systems require a higher than average level of maintenance due to reagent usage and the relatively large number of moving parts. Titrators also create waste from the titrated sample and wash solutions. These characteristics may make titrators well suited to at-line applications where an operator can take full advantage of their flexibility for troubleshooting or support of multiple processes, and their limitations are less of an imposition. Nevertheless, on-line titrators are widely used for a variety of measurements in industry. Figure 5 shows an on-line autotitrator installation for measuring sulfide in dye and pigment processing.

Flow injection analysis (FIA) utilizes small volume streams flowing at precisely controlled rates to accomplish the volumetric measurements required for wet chemical analysis. By transforming volumetric control in a batch environment to rate/time control in a flowing environment, FIA allows the use of smaller samples and reagent volumes than are used in traditional titration, and the analyzer system may be mechanically simpler than a titrator. These characteristics usually result in lower operational and maintenance requirements, making these systems good candidates for on-line applications or for at-line applications where the operator has limited familiarity with analytical methods. Figure 6 shows one such at-line application for measuring dissolved metals in the processing tanks of a metal finishing plant. These benefits come at the expense of flexibility, however. FIA method development is more time-consuming than normal titration, and the resulting methods usually require tailored analyzer hardware, making them less amenable to change in response to changing process requirements.

Titrators and batch mode FIA methods operate on discrete samples from the process, so they do not provide continuous, real-time data on the process. When used in continuous flow mode, FIA can provide continuous data with only a slight time delay in the results. Neither technique can be considered in-line. Analysis times depend on the chemistry, but for an automated on-line instrument the time required for a complete cycle of sampling, analysis, and purging to waste is typically in the range of 5–15 min. For an at-line system the analysis times are similar to those for on-line, but the manual sampling and sample preparation times may be considerably longer. At-line system throughput can be increased through the use of standard laboratory autosamplers, but the delay from the time of sampling to the delivery of the analytical results will still be considerably longer than for an on-line system. More detailed information about the use of titration and flow injection methods can be found in the

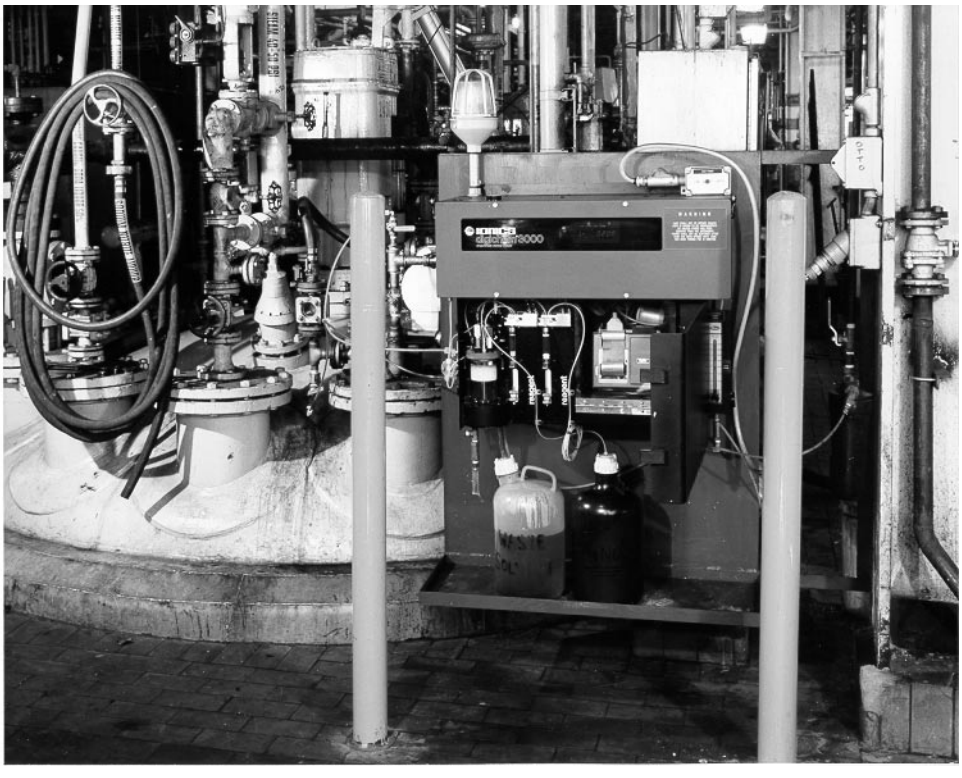

Figure 5 An on-line automated titration for sulfide in dye and pigment processing. (Photo courtesy of Ionics, Inc.)

Figure 6 A flow injection analyzer for at-line use in metal finishing. (Photo courtesy of The Boeing Company.)

articles **Titration Techniques for Process Analysis** and **Flow and Sequential Injection Analysis Techniques in Process Analysis** which appear later in this section.

3.2 Chromatographic Techniques

Chromatographic techniques have been the most widely applied of all PAC techniques. As in all chromatography, a very small volume of sample is injected into a carrier stream and passed through a column of selectively adsorbing material where the sample components are retained to varying degrees and then detected individually as they exit the column. The resulting data provide an indication of the relative concentrations of the detectable components in the process. Contaminants or unwanted components may be detected when present in sufficient concentration, but these techniques are generally not used for trace analysis.

All types of chromatography have been used in PAC. GC is the typical choice for gaseous and volatile organic liquid samples, normal and reversed phase liquid chromatography (LC) for nonvolatile organics, size exclusion chromatography for polymers, ion chromatography (IC) for ionic species in water, and supercritical fluid chromatography (SFC) for large, nonvolatile, or heat-sensitive molecules. Since chromatography, like titration, involves processing discrete samples outside of the process, it is not compatible with an in-line approach and does not provide continuous data. It can only be used where samples can be taken from the process without significant degradation.

GC is the most widely used on-line process analysis technique in industry. Much of its popularity in the petrochemical industry is due to the fact that it is a standard laboratory method for determining chain length distribution (carbon number), and the value of making this measurement on-line is high. A similar application is shown in Figure 7 where a process gas chromatograph is being used to determine the heating value of natural gas on a production platform.

GC provides analysis times typically in the 2–15-min range and a rapid and straightforward sampling interface to the process (especially for gaseous process streams), but it also has its problems. Maintenance issues such as column aging, shifting retention times, injection valve life, carrier gas consumption, and safety issues for flame ionization or electron capture detectors, may make GC a less desirable option if spectroscopic or sensor-based techniques are available.

LC, both normal and reversed-phase, is less widely applied in PAC than GC. LC has most of the same maintenance and safety problems that are associated with GC, only to a greater degree. It has the further disadvantages of relatively long analysis times (typically 10–30 min), less separating power, and higher equipment and operating costs owing to the need for solvents, precision pumps and high-pressure operation. Nevertheless, the pharmaceutical industry, which uses LC as its workhorse technique in the laboratory, is looking increasingly at process LC. Developments in chemometric treatment of LC data are enabling shorter analysis times by providing analytical resolution with far less than perfect peak resolution. These data treatments can be applied to all types of chromatography.

IC and SFC are more specialized than LC and less commonly used. These techniques have analysis times similar to LC, but are less stable in automatic operation and have higher operating and maintenance costs. SFC is sometimes substituted for LC, especially where supercritical carbon dioxide can be used as a safe (nonflammable) mobile phase. In most cases, though, some other analytical method can be found which will be more reliable and cost less. A later article in this section (**Chromatography in Process Analysis**) provides further information on application of these techniques.

3.3 Spectroscopic Techniques

While GC has been the most widely used PAC technique to date, spectroscopic techniques are exhibiting a faster rate of growth in usage. Microwave, infrared (IR), NIR, Raman, UV/visible, and X-ray spectroscopy offer

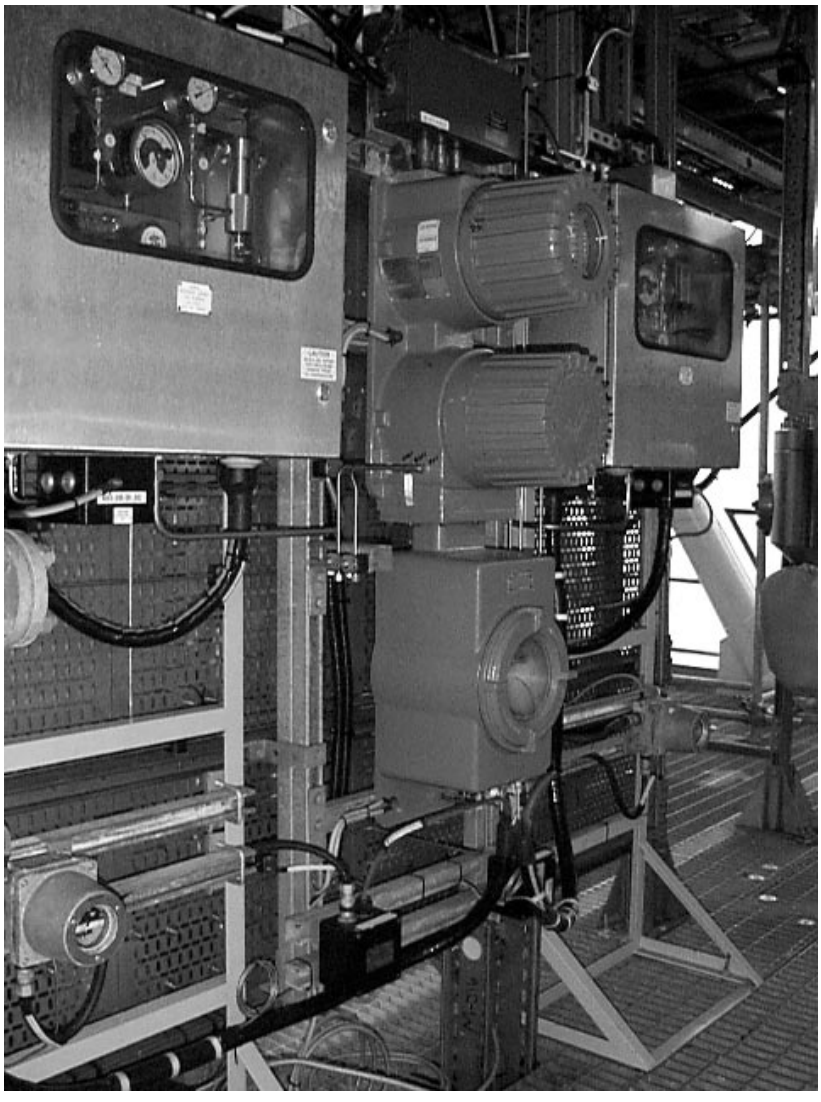

Figure 7 A process chromatograph measures energy content on a natural gas production platform in the North Sea. (Photo courtesy of Rosemount Analytical.)

a host of possible methods for process analysis. Four other articles in this section specifically address the use of IR (**Infrared Spectroscopy in Process Analysis**), NIR (**Near-infrared Spectroscopy in Process Analysis**), Raman (**Raman Spectroscopy in Process Analysis**), and UV/visible techniques (**Ultraviolet/Visible Spectroscopy in Process Analyses**). General advantages of this group of techniques include measurements which are in-line, continuous in real time, and noninvasive (as defined in section 2). The measurements may also offer a high degree of specificity for the analytes of interest, the ability to determine multiple analytes with a single spectrum, and, for simple cases, the equipment may be inexpensive. While all of these advantages are possible in general, they are usually not all true for any specific application.

Spectroscopic techniques can be used for at-line and on-line applications by measuring process samples or a side stream. However, one of the major advantages of these techniques is their ability to make direct process measurements in-line, usually producing a continuous stream of data in real time. If the process is enclosed, this is usually done using windows in pipes or vessels through which the spectrum of interest can freely pass, or using waveguides or fiber-optic probes to transmit radiation into and out of the process stream. Figure 8 shows a NIR fiber-optic system used for monitoring the process in a reactor vessel. If the process is in the open air such as for films, webs, or solid materials on a conveyor, windows or probes may not be needed. Figure 9 illustrates one such application, an open-air NIR measurement of moisture for a papermaking process.

Figure 8 A fiber-optic NIR probe for process monitoring. (Photo courtesy of Foss NIR Systems.)

As with laboratory-based spectroscopy, the ability of a spectroscopic PAC technique to monitor an analyte of interest depends on that analyte having some unique feature or combination of features in the spectral region being measured. This is the primary requirement for the use of these techniques. If an analyte has a single unique and strong spectral feature, it may be monitored with rapid simple methods and inexpensive equipment, such as the continuous measurement of a single or pair of wavelengths using a colorimeter or filter photometer. For multiple analytes, analytes with less pronounced spectral features, or process streams with components which have interfering spectral features, more capable instruments are needed to acquire data from many wavelengths, and more sophisticated data analysis techniques such as chemometrics may be required to extract useful information. Unfortunately, the spectral complexity of many process streams is beyond our current capability to unravel, even with the best equipment and data analysis methods.

When data from multiple wavelengths or the entire spectrum must be collected and processed to produce the desired result, the measurements will be discrete instead of continuous, but the frequency can still be quite rapid, usually in the range of several times per second to several times per minute. When analyte concentrations are low or there is spectral interference, signal averaging may be useful to enhance weak signals which are near the noise threshold. Using signal averaging will also reduce the frequency of the resultant measurements.

Another requirement of spectroscopic techniques is that the probing radiation should penetrate into the process stream and return a spectral signal. A process matrix which is opaque or nearly opaque to the measuring

Figure 9 A NIR filter photometer measures moisture in paper during a drying operation. (Photo courtesy of Moisture Systems, Inc., a Thermedics Detection Company.)

radiation will return a very weak spectral signal which represents only the material in immediate contact with, or very near, the sampling window or probe. If mixing is not excellent in this region, the spectral signal will not be representative of the process stream. Fouling of the sampling windows or probes can also restrict penetration into the process stream, and result in no measurement or a measurement biased by the composition of the fouling material.

The complexity of spectroscopic PAC applications varies widely, and the cost varies widely, as well. Spectroscopic techniques are more stable and mechanically reliable than titrators or chromatographs, and they do not require consumables such as reagents or carrier fluids. The simple example of the inexpensive colorimeter represents the low end of the cost scale. The initial cost of such an instrument is relatively low, and the maintenance cost will also be low if the process is moderately clean and the operating environment not too demanding. In the middle range will be systems such as the moisture meter in Figure 9 used to scan the web of a papermaking machine. The instrument and installation are more complicated, and the operating conditions it must endure are more demanding. At the high end of the range will be systems requiring full spectrum spectrometers with difficult interfaces to the process, advanced data analysis capabilities, and which require protection from severe environments. Not only will the initial costs of these systems be high, they will also require more maintenance from more highly skilled operators and technicians. Building and maintaining the multivariate models needed for these instruments is time-consuming and frequently requires the efforts of a trained chemist, chemometrician, or equivalent skilled personnel.

3.4 Sensor-based Techniques

Chemical sensor-based measurements might be considered the ideal method for implementing PAC. They have the potential to provide instantaneous and continuous chemical information about what is happening inside a process stream at very low cost. This is the kind of information that process engineers need most to operate a process at optimum conditions. Sensors are often used for in-line measurements such as the application shown in Figure 10 where the humidity of drying air is monitored using a polymer film capacitance sensor. They can also be used for on-line measurement of a side stream, or at-line on an extracted sample. Many other instrument types such as titrators, flow injection analyzers, and chromatographs use chemical sensors as detectors in making their measurements.

We can define a chemical sensor as a device which when exposed to a process stream delivers a signal which provides information about the chemical composition of the stream. The range of chemical sensors from which we can select is very broad, and new sensor development continues to be a significant area of PAC research. We will not attempt to provide extensive information about sensor types in this article. Instead we refer you to the further information at the end of the article (especially I, II and V) for references on sensor performance and application information and sources of equipment.

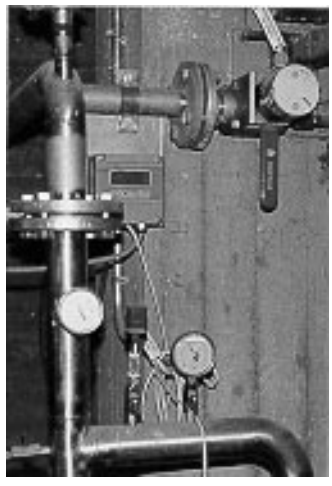

Figure 10 A polymer film capacitance humidity sensor monitors moisture in air for a paper drying operation. (Photo courtesy of Vaisala, Inc.)

The ideal sensor should provide a continuous signal which is highly selective for the analyte of interest and sensitive enough to pick up the smallest meaningful change in the analyte instantly. It should also be very small, impervious to attack by any components or contaminants of the process stream, unaffected by the process temperature or pressure, physically robust, highly reliable, and cheap. While there may not be any truly ideal sensors, there are some which come close. Ion-specific electrodes, humidity, pH, and oxidation–reduction potential probes and sensors are good examples which have seen broad application and are nearly ideal for some processes. Advances in sensor technology are occurring almost daily. Inexpensive and even disposable sensors under development for the medical field may soon become available for PAC, too.

As a closing comment on sensor-based techniques, we note that there are technologies now in use and being developed which tend to blur the distinctions between sensors and other categories of PAC techniques. By our definition a fiber-optic probe for a spectrometer such as the one shown in Figure 8 might be considered to be a sensor because it returns a signal containing chemical information when exposed to a process stream. New devices and technology are putting techniques like

spectroscopy into a small package with characteristics similar to a chemical sensor. Fiber-optic pH probes are an example. Likewise, when microminiaturization makes it possible to put chromatography or FIA on a silicon chip, these techniques will be freed from some of their limitations and will be deployable as sensors.

3.5 Other Techniques

While many techniques used for PAC fit into the four categories already described, there are a variety of measurement techniques which do not. We cannot enumerate and describe them all, but a few examples may be useful. One such technique is mass spectrometry (MS), which ionizes a gaseous sample stream and then uses magnetic and electric fields to separate the ionic species by charge-to-mass ratio so that they can be quantified individually. MS is used primarily for gaseous processes, where it can deliver continuous measurement data in an on-line configuration. Because of the requirement of a gas phase sample stream, MS has not been used broadly in PAC. There is development work underway to extend its application to a broader range of samples. An article **Mass Spectrometry in Process Analysis** later in this section provides further information on the use of this technique.

Nuclear magnetic resonance spectroscopy is another technique which has been little used in process analysis to date because of the requirement to place the sample within a strong magnetic field. But developments with this technique have also made it the subject of increasing interest for process application. Details on these developments are covered in the article **Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis** later in this section.

The various atomic emission and absorption spectrometry techniques do not share the application characteristics of the molecular absorption spectroscopy techniques described earlier in this article. The emission techniques decompose and excite a sample using high levels of energy, usually in a flame or plasma torch, and then detect the radiation emitted by atomic transitions. Atomic absorption also decomposes the sample, but then detects the atomic transitions by measuring radiation absorbed by the sample. All of these techniques are used for identifying and quantifying elements, and will not provide complete analytical information when used on mixtures of compounds which contain the same elements. They are sensitive to low concentrations of metals, and are often used for wastewater monitoring, sometimes in an on-line mode. They are less useful for on-line monitoring of more complex mixtures owing to the interpretation or additional information required to deconvolute interfering species. In addition, these instruments require consumables such as acetylene, hydrogen, or argon to sustain the flame or plasma used to atomize and or excite the sample.

X-ray fluorescence spectrometry is similar to atomic emission spectrometry in that X-ray energy excites a sample and the resultant radiation emitted by fluorescence is measured. However, X-ray fluorescence is more like absorption spectroscopy because no external support gases are needed. X-ray fluorescence can be a continuous, real-time in- or on-line measurement technique. Its primary difficulties are associated with the use and limitations of the X-ray radiation. This techniques finds its use in process analysis for a relatively small number of specific applications, typically in the mineral processing industries and certain web-coating processes.

Another group of techniques uses radiofrequency or microwave radiation to measure dielectric properties which infer chemical information. These methods are on the fringe of what might be considered analytical spectroscopy. Most often used for moisture determination in solids, these techniques can provide continuous data in either in-line or on-line configurations. Many commercially available models are designed for production facility installation, with features like hardened enclosures and built-in capabilities for calibration, fault determination, and self-diagnosis.

3.6 Technique Selection

While the categories just described do have functional value in considering the application of techniques to process analysis, we recognize that there are other criteria for categorization which could have been used. The purpose of the categories is only to aid us in a high level selection of which techniques are reasonable candidates for use in a particular application and which are not. Regardless of how we categorize the techniques, or the category from which the candidates come, the criteria we use for the final selection from among the candidates remains the same – performance in the process. The performance factors which must be considered include at least the following:

- selectivity
- sensitivity
- response time
- resistance to attack by chemicals in the process stream
- temperature and pressure sensitivity
- resistance to physical damage
- compatibility with operating environment
- reliability and reproducibility
- stability and calibration frequency
- aging characteristics
- self-diagnostic capability

- initial cost
- replacement cost
- maintenance costs
- consumable costs.

Other factors may be added to this list to account for the special needs of the process or operating environment. By prioritizing these factors and evaluating each technique by its performance in each area, it is usually possible to narrow the range of candidate techniques to a few, and possibly one. Occasionally a PAC technique can be identified which is a perfect fit to the needs of a process. Usually, though, the selection will involve some performance trade-offs. Knowing the relative value of the different performance factors is the key to evaluating these compromises. It must be stressed that detailed knowledge of the process requirements and operating environment is a prerequisite to prioritizing the factors and making the selection. Adequately defining the problem and gathering information about the process are the first two steps in the flow chart in Figure 1. They are critical to success, and they must be done first or time and money will be wasted.

3.7 Summary

It is obvious that we cannot do justice here to the full scope of options available in PAC techniques. We must refer you to the reference materials and resources at the end of this article, the other articles in this Encyclopedia, and to the marketplace to discover what is available. As we have noted before, PAC technology is a rapidly developing field. In the lists of further information we have made an effort to identify both classic works, which provide a foundation for conceptual understanding, and materials and networking sources which will be updated regularly to reflect the current state-of-the-art.

4 SAMPLING AND SAMPLING PREPARATION ISSUES

One of the biggest issues in PAC is obtaining a representative sample. This is important for the obvious reason that we want to know what is going on in the process and not be misled by stream inhomogeneity, and because calibration of the process instrument and correlation with the laboratory measurement require ''true'' samples. Sampling systems are also needed for some on-line measurements to ''condition'' the sample prior to analysis. As an example, consider the situation in which the moisture on a conveyor belt of grain is being measured using a NIR filter photometer. The photometer must be calibrated against the laboratory measurement, which is a loss-on-drying measurement. The conveyor belt is 4 feet wide (122 cm) and the product depth is 5 inches (approximately 13 cm). Good sampling procedures require that the position across the belt, depth, and timing of the grab sample taken to the laboratory be considered. Obviously, the sampling position must match the position which the NIR light is measuring. If the product is ''turned over'' to expose grain from beneath the surface, that must be sampled. The grab sample should also match the time constant of the NIR measurement. In a liquid measurement such as hydrocarbon distribution by GC, the issues are similar. The stream to the GC must be continuously flowing and the timing of manual sampling must be matched to the GC injection. (What is the process flow rate?) If the instrument is 5 m away from the sample valve, the sample must be drawn taking the lag time into account. If the process is at steady-state (the ideal), this issue is less critical.

Even if sampling issues are carefully considered, the laboratory results may not agree very well with the process instrument results. Most often this is because the two methods are not identical. Even a simple pH measurement may not agree with the laboratory measurement because of temperature differences, dilution differences, and so on. In the grain example, the differences between the laboratory drying method and the process NIR method must be understood. The laboratory method may be a total moisture method, while the process NIR may be measuring only the unbound water. In the sugar example from section 2, this was the prime reason for selection of in-line refractive index for the measurement – it was the same method as used in the laboratory. It is important that operating personnel understand that two methods may not give *identical* values. In addition, many samples undergo changes between sampling the process stream and analysis in the laboratory. Samples may cool down or warm up, water or solvents may evaporate, reactions may begin or continue, the sample may be exposed to air, and so on. For some processes, samples must be taken through special sampling ports to maintain process sterility (e.g. fermentation). Samples drawn from such processes must also be protected from contamination, lest microorganisms change the sample. Any change in the sample, prior to its analysis in the laboratory, can lead to disagreement between the result from the process instrument and the laboratory result. Such disagreement will have deleterious effects on instrument calibration and on operator confidence in the process result.

In some situations it is not possible to match the laboratory and process techniques or sampling procedure so that the results of the two approaches are identical. For process control, this is not necessarily a critical issue, provided that the operating personnel understand it. In this situation, the process measurement still ''tracks''

the process and its results correlate to the laboratory measurements. The process operator or control system can still follow slow process changes and upsets and adjust accordingly. With sufficient data, the process instrument output may eventually be adjusted – perhaps at the control room display – so that the results match the laboratory. The important thing to remember is that people understand that this was done and why.

It is beyond the scope of this introductory article to delve into the details of sampling systems; that is left for a later article in this section, **Sampling and Sample Preparation in Process Analysis**. However, there are several key points the reader should think about:

1. Obtaining a representative sample of a varying process stream is not trivial, whether the sample is intended for analysis in the laboratory or PAC. In some cases a sample removed from the process changes before it can be analyzed in the laboratory. Some examples include thin materials which dry extremely rapidly, solvents evaporating from the sample as they are captured, reactions which are not quenched adequately, and so on. In these instances, it may be easier to bring the analysis to the process than to attempt to improve the sampling procedure for a laboratory sample.
2. A sampling system for PAC has five basic functions:
   - to take representative samples of the flowing system
   - to transport samples to the analyzer
   - to condition samples as needed (filtering, adjusting temperature, pressure, etc.)
   - to handle switching between sample streams, including a calibration stream or tank
   - to return samples to the process or waste.
3. The definition of the term ''representative sample'' depends on the process, analyte of interest, and the technique being used. Therefore, the sampling procedure must be appropriate for the specific PAC application. Representative does not necessarily mean that the sample is identical to the process stream. For example, it may be necessary to filter a sample to remove particulates which would contaminate a GC system. If the concentration of the analyte of interest is unaffected by the filtration, the sample is representative in the analyte, but technically it is not identical to the original.
4. As in the real-estate business, the critical factor for sampling is location. The sampling point must provide a representative sample for the analyzer in a timely manner so that unnecessary delay is not added. Therefore, not only is the location important, so also is the flow rate of a sample line. It may be necessary to install a high speed pump to deliver the sample stream to the analyzer.

The variety of functions which a sampling system must perform often causes these systems to be complex. Figure 11 shows one example of a system used to sample and filter four different petroleum process streams for

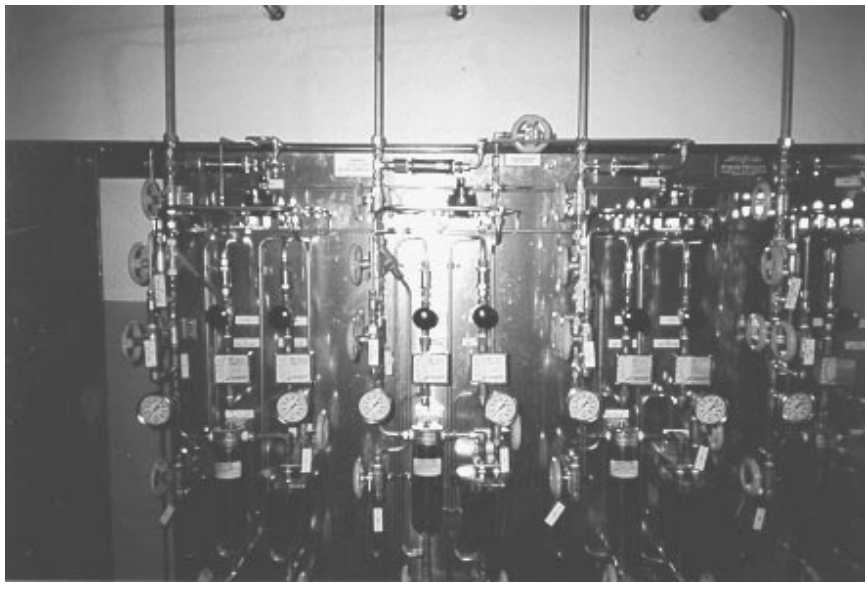

Figure 11 A system for sampling four process streams with pressure and flow control and filtration. (Photo courtesy of Orbital Sciences Corporation.)

delivery to an on-line analyzer. It should be evident that care must be taken in the design and construction of such a sampling system to prevent it from becoming a maintenance headache or the source of frequent overall system failure.

5 DATA TREATMENT

Data treatment begins with data. In its simplest form the data consist of a single value, at a given time, transmitted as an analog, most commonly a 4–20 mA signal. (Generally, an at-line instrument does not transmit its data at all, but has its own display or computer with it. This discussion is intended for in-line and on-line measurements.) In the case of a single value being transmitted as an analog signal, what data treatment issues must be considered? First, the instrument itself must be properly calibrated. In addition, the analog output must be spanned properly, to maximize signal resolution. For example, if the pH of a wastewater stream which is typically at ~6.0 is being monitored, the meter must be calibrated with buffers, say 4 and 7. In addition, the output of the instrument should be set so that a pH of 4 gives 4 mA output and pH 8 gives an output of 20 mA. This provides a reasonable range for the application, with the target or typical value at 50% of the range. If the instrument output ranged from 1 to 14 pH units, resolution would be significantly poorer. Similarly, if the output is going to a process control system (programmable logic controller, PLC, or distributed control system), the input of the control system also must be set properly. In addition, the output and the input should be calibrated for correct current reading.

A more complex situation occurs when the instrument provides more than one signal. In the pH example, perhaps the instrument can transmit both pH and temperature. Now both 4–20 mA signals must be calibrated and correctly spanned.

The most complex situation occurs when an instrument makes multiple measurements or uses its own computer to calculate a value from the raw data. Gas chromatographs and NIR spectrometers are examples of these more complex cases. The gas chromatograph provides data on multiple compounds in the sample stream. It must extract this from the chromatogram peak areas, apply appropriate calibration functions, and then move the data to the control system in a usable form. In doing all this, we must be sure that the instrument correctly recognizes each peak (compound) and that the quantitation is correct. A system must be designed in which calibration standards can be analyzed. In the example of the NIR instrument, it acquires a spectrum of the sample and converts that spectrum to quantitative information on the sample. This could be from several wavelengths or the whole spectrum. If a chemometric model is used to obtain the concentration data, a lot of effort probably went into developing and validating the calibration model. If it is intended to use the same measurement in other locations of the plant or another plant, much effort can be saved by transferring the calibration from one instrument to another. Calibration transfer is one of the biggest areas of research today and should be investigated carefully before proceeding. Calibration model building, and calibration transfer, and other information on the use of chemometric techniques are addressed in a later article in this section on **Chemometric Methods in Process Analysis**.

As discussed in section 4, agreement between the laboratory and process analytical results may not always be perfect. In general, the laboratory result is assumed to be the ''correct'' one. The process instrument is used to monitor and control the process and in many cases need not be ''accurate''. It is more important for a process measurement to be precise and reliable. Of course, in some cases, PAC data may be more accurate because of the sampling issues described in section 4 (samples which change en route to the laboratory, etc.). In this situation, however, how do we know what accuracy is? If the laboratory samples change, but the laboratory method is the ''accepted'' way to make the measurement, how can it be demonstrated that the PAC measurement is accurate?

Operator confidence in PAC instruments comes when the instruments are up and running and not being repaired or calibrated every few days or weeks. It comes from quality results, which are not varying too much (short term) or drifting (long term). Put another way, the PAC instrument needs to have sufficient precision (repeatability, reproducibility) for the data to be useful to the process engineer, operator, and control system. The exact amount of short-term variation (noise) or long-term variation (drift) any process can allow in its instruments, while still operating efficiently, depends on that process and its target value. The target is normally chosen for quality, safety, or cost reasons, so there is no blanket rule for what level of precision a PAC measurement should have. In some instances, it is acceptable for the measurement to be ±10%, in others it must be 25 times better (lower).

An increasingly important development in PAC is instrument fault detection. More and more instruments now have the ability to diagnose problems within the instrument and alert the operator. This can prevent faulty data from being used. For example, some pH analyzers now monitor the impedance of the glass electrode and signal an alarm if it has reached the point of poor pH measurement. In addition, chemometric techniques are being developed to do much the same thing for

multiple measurements on the process. The data from many measurements are examined and ''odd'' results are flagged. In many cases, a faulty result may be traced to an instrument problem; in others, the software can ''validate'' the questionable data and verify that a real, but unusual, process change has occurred. These two developments should become quite significant as PAC moves into the 21st century.

6 BUSINESS AND OWNERSHIP ISSUES

6.1 Business Case

As mentioned at the beginning of this article, industry uses PAC to gain financial advantages. If an application of PAC is to be successful, there must be a business case which solidly justifies it. All business justifications are ultimately based on profit generation, but not all arise directly from the economics of the process being measured. Identification and evaluation of the benefits and costs of the PAC application are necessary to build this justification.

Improvement in process control is the driver for most economic benefits from PAC. The costs of providing measurement tools to improve process control must be balanced by the benefits of what that control can deliver. Increased yields, increased throughput, improved quality, reduced raw material costs, and reduced waste costs are all examples of recurring benefits from improved control. These benefits derive directly from the process economics and are usually straightforward to assess.

Avoiding the costs of process excursions and upsets is another benefit of improved process control. The costs avoided may not be limited to the value of time, material, and labor which is lost when a process goes out of control, but may also involve regulatory fines and the cost of injury or chemical exposure of the workforce. The economic value of cost avoidance is more difficult to quantify than recurring cost reductions, but it is often a critical element in the business case.

Benefits which are further removed from the process economics can also provide justification for PAC applications. Strategic or competitive advantage in the marketplace may be derived from process changes which can only be achieved through the use of PAC. Safety or regulatory requirements may also dictate the application of PAC measurements to monitor emissions, or they may render a process noncompliant without improved process control. When these high level drivers from the business environment are at work the direct impact of PAC on the process economics may become a secondary consideration in the business case. The process economics should nevertheless be fully defined because they are an important element in designing the PAC application and in optimizing the result.

Another important indirect benefit of PAC is the insight into the process which is gained through the new information it provides. It is rare for a new PAC installation not to teach the process engineer or chemist something new. For this reason, PAC should be a key feature of all pilot plants and small-scale facilities, especially where process development and scale-up are involved.

The other half of the business case considers the costs of making the improvement. These costs should be considered in two categories: initial and recurring.

6.2 Initial Costs

It is fortunate that process improvement projects like those associated with implementing PAC technology can be divided into phases. These phases provide a convenient means of delineating and controlling the costs of a project, and of redirecting or halting projects which are not on track to meet the desired objectives. The following is an example of a high level outline which can be used to identify the initial costs of a PAC project.

- Preliminary phase
  1. Determination of process need and business case
  2. Evaluation or development of PAC techniques
  3. Design of PAC system
- Installation phase
  1. Capital acquisition
     (a) instruments and controllers
     (b) sampling system
     (c) data collection, communication, and archiving system
     (d) environmental protection for system components
  2. Modifications to process equipment
  3. Installation of utilities and consumables
- Implementation phase
  1. Training of workforce
     (a) operators
     (b) maintenance technicians
  2. Transition period
     (a) working the bugs out
     (b) learning curve
     (c) acceptance by workforce

Such a project outline (called a work breakdown structure, or WBS, in project management terminology) is the best tool for estimating and accumulating the labor and nonlabor costs and the time period associated with any project. The example of WBS shown here would be adequate only for a very simple project. Most projects will require more detailed levels in the outline to identify clearly all sources of project costs and facilitate accurate estimates of what each cost will be.

6.3 Recurring Costs

It is also convenient to break down recurring costs into categories so that they can be more easily estimated and accumulated. The following list contains typical items which should be considered.

- Scheduled maintenance
  - Consumables
  - Calibration
  - Cleaning
  - Replacement of wearing parts
- Maintaining training and expertise in workforce
  - Operators
  - Maintenance technicians
- Unscheduled maintenance
  - Warranty or repair costs
  - Future upgrade or replacement

The labor and nonlabor costs associated with these items, together with their expected frequency, will provide an annual estimate for the recurring costs of a PAC application. The unscheduled maintenance items are perhaps the most difficult to estimate on an annual basis. The annual cost of instrument repairs can be based on the cost of a service contract if one is offered by the manufacturer, although if high skill levels are available in the local workforce this may be an overestimate. An annual replacement cost may be estimated by averaging the cost of a new instrument over its expected lifetime.

6.4 Ownership

A critical business factor which determines the success or failure of a PAC application is that of ownership. In discussing and planning for the ownership of a PAC system, there are two distinct roles to consider: the role of the champion and the role of the vested owner. Like the other business issues already discussed, these roles are not unique to PAC. They can be universally applied to any project in business.

The champion's role is to act as an agent of change in the business system. This person or persons should have a position which is respected by, and has influence on, both the business decision makers and the people who will implement the project. The role consists of convincing the decision makers of the value of the project, of identifying and enlisting a person as the vested owner, and of overcoming the inevitable resistance to change which may come from any of the parties involved in, or affected by, the project. If the business case is solid, the task of convincing decision makers should be simple. Most of them will recognize a dollar when they see one. However, if the business case is a little more tenuous, the decision makers must rely on their trust of the champion's judgement.

Identifying a vested owner for the completed PAC system is another task which is often done by the champion. The champion and the vested owner may be one and the same person. The vested owner of the PAC application has the role of continuing day-to-day support and maintenance of the system, and planning for future replacements and upgrades of the equipment and software. This person should be vested in the benefits from the system as well as sustaining the costs. If this is not the case, the risk is high that the system will not be adequately maintained in the long term. In some manufacturing organizations, the management structure is not set up to allow direct control of system maintenance by the process owner who benefits most from the system. With costs and benefits assigned separately to two different people, probably in two different organizations, conflicts in priorities often develop. This situation should be avoided whenever possible, or mitigated through the influence of higher level managers who can maintain adequate priority for maintenance and planning for the future.

The most challenging aspect of the champion's role is overcoming resistance to change. The vested owner also often participates in this task. Resistance to change can come from many directions. The workforce or a union may perceive the change to be a threat to job security, and this may be an accurate perception. Local management may resist because of memories of previous difficulties with PAC or other measurement systems not associated with chemical analysis. Technicians or operators may resist because the change was not their idea, or perhaps because they have something different in mind. Change often seems to breed resistance just because it is change. The reasons can be endless, and all of them represent a risk to the success of the project. The champion and vested owner must do everything possible to reduce the risk by winning people over with reasoning, accommodation, and leadership, and then make whatever plans are possible to mitigate the effects of any remaining resistance.

All the business and ownership issues dealt with here are common to successful projects in virtually any business environment. There are many excellent books, periodicals, and training materials available on these subjects, a few of which are given as references in the further information section.

ABBREVIATIONS AND ACRONYMS

FIA	Flow Injection Analysis
GC	Gas Chromatography
IC	Ion Chromatography
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
NIR	Near-infrared
PAC	Process Analytical Chemistry
PLC	Programmable Logic Controller
SFC	Supercritical Fluid Chromatography
UV	Ultraviolet
WBS	Work Breakdown Structure

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Field-portable Instrumentation

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

Food **(Volume 5)**
Near-infrared Spectroscopy in Food Analysis

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Optical Particle Counting • Ultrasonic Measurements in Particle Size Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Fuel Performance Specifications, Mid-infrared Analysis of • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels • Refractive Index Technology as a Real Time Viscosity Technique

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis • Chromatography in Process Analysis • Flow and Sequential Injection Analysis Techniques in Process Analysis • Infrared Spectroscopy in Process Analysis • Mass Spectrometry in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis • Raman Spectroscopy in Process Analysis • Sampling and Sample Preparation in Process Analysis • Titration Techniques for Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher • Signal Processing in Analytical Chemistry • Soft Modeling of Analytical Data

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals

Electronic Absorption and Luminescence **(Volume 12)**
Near-infrared Absorption/Luminescence Measurements

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Spectral Data, Modern Classification Methods for

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Supercritical Fluid Chromatography

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Literature Searching Methodology • Microwave Techniques • Multivariate Image Analysis • Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration

FURTHER INFORMATION

There are many sources of information on PAC, including the other sections of this Encyclopedia. Some valuable books and journals are listed below in sections I and II. Section III lists some of the many conferences sponsored by analytical chemistry and engineering societies which have sessions devoted to process measurements. Short courses on the subject are also available, some of which are listed in section IV. Guides for locating instrument manufacturers are shown in section V, and the manufacturers themselves can be quite helpful. Research consortia such as the Center for Process Analytical Chemistry (University of Washington, Seattle) can also provide a wealth of knowledge. Finally, section VI lists a few of the many resources available for improving the probability of a successful implementation of PAC by planning and managing the project well.

I Books on Process Analytical Chemistry

K.J. Clevett, *Process Analyzer Technology*, J. Wiley & Sons, Chichester, 1986.
D.C. Cornish, G. Jepson, M.J. Smurthwaite, *Sampling Systems for Process Analyzers*, Butterworths, London, 1981.
E.A. Houser, *Principles of Sample Handling and Sampling Systems Design for Process Analysis*, Instrument Society of America, Pittsburgh, PA (now, Research Triangle Park, NC), 1972.
J. Janata, *Principles of Chemical Sensors*, Plenum Press, New York, 1989.
B.G. Liptak (editor-in-chief), *Analytical Instrumentation*, Chilton Book Company, Radnor, PA, 1994.
B.G. Liptak (editor-in-chief), *Instrument Engineers Handbook, 3rd edition, Process Measurement and Analysis*, Chilton Book Company, Radnor, PA, 1995.
F. McLennan, B.R. Kowalski, *Process Analytical Chemistry*, Blackie Academic & Professional, London, 1995.
R.E. Sherman, L. Rhodes (eds.), *Analytical Instrumentation*, Instrument Society of America, Research Triangle Park, NC, 1996.

II Journals and Magazines with a Significant Coverage of Process Analytical Chemistry

Analytica Chimica Acta, Elsevier Science, New York, USA.
Analytical Chemistry, American Chemical Society, Washington, DC, USA (especially the review issues devoted to applications, e.g. June, 1999).
Control Engineering, Cahners, Division of Reed Publishing, Des Plaines, IL, USA.
Control, Putnam Publishing Co., Itasca, IL, USA.
IAN, Instrumentation & Automation News, Cahners, Division of Reed Publishing, Des Plaines, IL, USA.
Instrumentation & Control Systems (I&CS), PennWell, Broomall, PA, USA.
InTech, Instrument Society of America, Research Triangle Park, NC, USA.
Process Control and Quality, Elsevier, New York, USA.
Journal of Near Infrared Spectroscopy, NIR Publications, Chichester, West Sussex, UK.
Measurement and Control News, Measurements & Data Corp., Pittsburgh, PA, USA.
Measurement and Control, Measurements & Data Corp., Pittsburgh, PA, USA.
NIR News, NIR Publications, Chichester, West Sussex, UK.
Sensors, Helmers Publishing, Peterborough, NH, USA.
Spectroscopy, Advanstar Communications, Eugene, OR, USA.
AT-Process, The Journal of Process Analytical Chemistry, InfoScience Services, Northbrook, IL, USA.
Talanta, Elsevier Science, New York, USA.

III Conferences with a Significant Coverage of Process Analytical Chemistry

International Forum for Process Analytical Chemistry, InfoScience Services, Grayslake, IL 60030-7100 (http://www.ifpac.com)
Federation of Analytical Chemistry and Spectroscopy Societies Meeting (FACSS), 1201 Don Diego Av., Santa Fe, NM 87505 (http://FACSS.org/info.html)
Instrument Society of America, 67 Alexander Dr., Research Triangle Park, NC 27709 (http://www.isa.org)
Eastern Analytical Symposium, P.O. Box 633, Montchanin, DE 19710-0633 (http://www.eas.org)
The Pittsburgh Conference, 300 Penn Center Blvd., Pittsburgh, PA 15235-5503 (http://www.pittcon.org)
International Conference on Near Infrared Spectroscopy, (example, NIR-99, Verona, Italy).
International Conference on Flow Injection Analysis.

IV Short Courses on Process Analytical Chemistry

Eastern Analytical Symposium (http://www.eas.org)
The Pittsburgh Conference (http://www.pittcon.org)
American Chemical Society (http://www.acs.org)
American Institute of Chemical Engineers (http://www.aiche.org)
Instrument Society of America (http://www.isa.org)
Also audio and video from the above organizations.

V Guides to Manufacturers of Process Analytical Chemistry Instrumentation

ISA Directory of Instrumentation, Instrument Society of America, Research Triangle Park, NC (http://www.isa.org)

Sensors Annual Buyer's Guide, Helmers Publishing, Peterborough, NH (http://www.sensorsmag.com)

VI Project Management Resources

A.B. Badiru, *Project Management in Manufacturing and High Technology Operations*, J. Wiley & Sons, New York, 1988.

C. Chapman, S. Ward, *Project Risk Management: Processes, Techniques and Insights*, J. Wiley & Sons, New York, 1997.

D.I. Cleland, *Project Management: Strategic Design and Implementation*, 2nd edition, McGraw-Hill, New York, 1992.

J. Davidson Frame, *The New Project Management*, Jossey-Bass Publishers, San Francisco, 1994.

P.C. Dinsmore (ed.), *The AMA Handbook of Project Management*, AMACOM, New York, 1993.

International Journal of Project Management, Elsevier Science, Oxford, UK.

Project Delivery, A System and Process for Benchmark Performance, CH2M Hill and Work Systems Associates, Inc., Denver, Colorado, 1996.

Project Management Journal, Project Management Institute Communications, Sylva, NC.

The Project Management Institute (http://www.pmi.org)

Chemometric Methods in Process Analysis

Karl S. Booksh
Arizona State University, Tempe, USA

1 Introduction **1**
1.1 What is Chemometrics? 1
1.2 Roles of Chemometrics 2
1.3 Soft Modeling Versus Hard Modeling 2
1.4 Chemometric Classification of Data 3
1.5 Nomenclature 3
2 Univariate Analysis **3**
2.1 History and Introduction 3
2.2 Statistical Analysis 4
2.3 Statistical Process Control Charts 4
2.4 Calibration and Prediction 4
2.5 Experimental Design 5
2.6 Figures of Merit 6
3 Multivariate Analysis **7**
3.1 History and Introduction 7
3.2 Factor Analysis 7
3.3 Multivariate Statistical Process Control Charts 11
3.4 Calibration and Prediction 12
3.5 Figures of Merit 14
4 Multiway Analysis **15**
4.1 History and Introduction 15
4.2 Multiway Factor Analysis 16
4.3 Calibration and Prediction 17
4.4 Figures of Merit 18
4.5 Batch Analysis 19
5 Pretreatment of Data **19**
5.1 Centering and Scaling 19
5.2 Background Correction 19
5.3 Instrument Standardization 20
6 Model Validation (Right Versus Wrong Model) **21**
Abbreviations and Acronyms **21**
Related Articles **22**
References **22**

Chemometrics may be loosely defined as the development and application of mathematical and statistical techniques to aid in the analysis of chemical data. The goal of chemometrics is to guide the chemist in the optimum extraction of chemically relevant information from a series of complex chemically oriented observations. To this end, this article addresses a wide array of chemometric tools that are applicable to industrial process analysis. The three main sections present statistical and mathematical tools in the context of univariate, multivariate, and multiway analyses. Each of the three sections discusses the benefits and limitations of qualitative analysis, quantitative analysis, and process monitoring with the class of data in question. Additional sections address the appropriate pretreatment of collected data to aid in efficient analysis and validation of the derived chemometric models. In these sections methods for background correction and instrument standardization are discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

This article surveys the various chemometric data analysis techniques relevant to process analytical chemistry. Recognizing the constraints on space, the focus of this article concentrates primarily on established analytical methods that have been demonstrated to be useful in multiple applications. Section 1 provides the context and rationale for chemometric analysis. Descriptions of the chemometric methods are divided by the complexity of the data being analyzed. Section 2 addresses the techniques available to univariate data. Section 3 addresses qualitative and quantitative analysis with multivariate data. Section 4 describes the methods appropriate for multiway data. Sections 5 and 6 discuss preprocessing of data before analysis and validation of the derived models after analysis, respectively. These references are not comprehensive but are intended to give a representative cross-section of the authors and applications found in the literature. More information on current applications of chemometrics in process analysis chemistry can be found in the ''Application Reviews'' published bi-annually in *Analytical Chemistry*.[1,2]

1.1 What is Chemometrics?

The definition of ''chemometrics'' varies greatly, often depending upon the context of the definition. Wise and Kowalski[3] and Massart et al.[4] place emphasis on the ''chemo'' root and classify chemometrics as a subdiscipline of chemistry that employs mathematical and statistical techniques to analyze chemical systems. Wold stresses that the impact of chemometrics is in problem solving and not data analysis: chemometricians ''must remain chemists and adopt statistics to chemistry instead of vice versa''.[5] Booksh and Kowalski[6] and Brown[7] have placed more emphasis on the ''metrics'' root and

defined chemometrics more as information science that can be applied to many physical science disciplines.

Chemometrics is a truly interdisciplinary science that draws from mathematics, statistics, and information science; however, the tools from these disciplines cannot be directly applied without sound knowledge and understanding of the chemical system in question. Many statistical tools are useless to chemometricians because the underlying assumptions are violated in the chemical system. Concurrently, a chemometrician could develop very useful ''statistical tools'' that are not generalizable beyond the chemical system in question. Furthermore, the distinction between ''statistical significance'' and ''practical significance'' cannot be reliably made without an understanding of both statistics and chemistry.

1.2 Roles of Chemometrics

1.2.1 Analysis

Historically, most attention has been devoted to the analysis side of chemometrics. Further distinction can be drawn to identify two applications of chemometric analysis: factor analysis (FA) and calibration. FA is most closely derived from other ''-metric'' sciences: econometrics and psychometrics. The goal of FA is to take a global view of a chemical process and uncover the underlying trends, or factors, that are imbedded in a collection of data. These factors lend the chemist insights on the factors that have influenced the process and the information content of the observations employed to monitor the process. Calibration implies a more focused perspective on a chemical process. The goal of calibration is to extract a specific piece of information from a collection of data. Examples include measurements of constituent concentrations and product quality. Consequently, calibration methods are more often applied to chemical sensors designed for monitoring specific quantifiable properties (e.g. moisture, octane number, protein) related to product or process quality.

Owing to the importance of calibration of chemical sensors for process monitoring, the majority of this article will be devoted to calibration. However, the implementation of FA for monitoring industrial processes will also be addressed.

1.2.2 Facilitation

In recent years, the role of chemometrics to facilitate the placement of instrumentation in the field for process monitoring has blossomed. This has been fueled by the realization that instruments are not identical and that the performance of an instrument may change dramatically over time. Consequently, much expense and labor are required to develop a calibration model for each instrument. The expense is magnified when the performance of the instrument changes and a new calibration model must be developed. The goal of chemometrics is to develop compact models that ''standardize'' the response of any given instrument to an ideal instrument on which the calibration model was developed. Thus, the calibration model is transportable among instruments.

1.2.3 Design

The most underutilized role of chemometrics in process analysis is that of design. Chemometrics is, at its core, an information science. By knowing exactly which information is needed, and not needed, from an analysis, chemometrics provides insights to the measurements that must be made and the type of instrument to make these measurements. This leads to an economy of experimental design and sensor construction. For example, by designing a sensor that relies partially on chemometric signal deconvolution techniques or multivariate calibration, complete physical separation of the analyte from the sample matrix may be avoided. This, in turn, would lead to a sensor that is placed in a sample stream, removing the need for a sampling loop to be installed in the process.

1.3 Soft Modeling Versus Hard Modeling

Philosophically, chemometrics generally employs a ''soft modeling'' approach to data analysis and calibration. With soft modeling a general class of model is imposed on the data and the parameters of the model are not constrained to be physically meaningful. In contrast, hard models assume that each parameter estimated relates to a physical property. For example, the Beer–Lambert law is a hard model. The parameters estimated (or known) are concentration, path length, and molar absorbtivity. A soft model approach to the Beer–Lambert law problem only says that concentration is somehow related to measured absorbance. Soft modeling has two distinct advantages over hard modeling. First, less needs to be known about the system studied. Soft models can be rapidly developed and tested relative to hard models – if a linear model did not work, try a nonlinear, polynomial, or spline based model. Second, soft models can be applied where no true hard model exists. For example, try to make a model explicitly relating the Raman spectra of gasoline to octane number. Whereas it is difficult to construct a hard model for this relationship, soft model based calibration methods can accurately predict the octane number.[8]

1.4 Chemometric Classification of Data

From a chemometric standpoint, data and instrumentation are often classified based on the dimensionality of the data set obtained. In general, the higher the dimensionality of the data set, the more powerful is the instrument, and consequently more powerful data analysis methods can be applied to higher dimensioned data sets. The different dimensions of data are presented in Figure 1. For a more complete discussion on the interrelationship between data structure and analyzability, see Geladi[9] and Booksh and Kowalski.[6]

The most basic type of data are univariate data. Examples of univariate data are data collected from a pH meter or single-channel photometer. Univariate data are the lowest dimensionality of data – a univariate instrument returns a zero-dimensional (zeroth-order) data tensor. Consequently, a collection of data from a univariate sensor forms a vector and is said to be one-way data, that is, it varies in only one way: sample to sample.

The majority of the literature on chemometrics address the analysis of multivariate data. Examples of multivariate data include chromatograms and spectra. Analysis of a single sample with a multivariate instrument yields a one-dimensional (first-order) data tensor. A collection of samples forms a two-dimensional matrix and is said to be two-way data because it varies from sample to sample and wavelength to wavelength.

Multiway data are formed, for example, by coupled instrumentation such as gas chromatography/mass spectrometry (GC/MS) and excitation–emission matrix spectrometers. Analysis of a single sample yields a two-dimensional (second-order) tensor of data. The key to having true multiway data is that one instrument (or order) must modulate the other instrument (or order). For example, ultraviolet/visible (UV/VIS) spectroscopy coupled with infrared (IR) spectroscopy does not give multiway data because the UV/VIS and IR spectra of a molecule do not modulate each other. A collection of sensor readings during the progression of a batch process do form multiway data (sensors by time). There is no upper limit on the number of ''ways'' that could go into a data set. Conceivably, one could employ on-line high-performance liquid chromatography (HPLC)/UV/VIS spectroscopy to monitor a series of batch processes: a four-way data tensor results (wavelength by chromatographic retention time by time in the batch by batch).

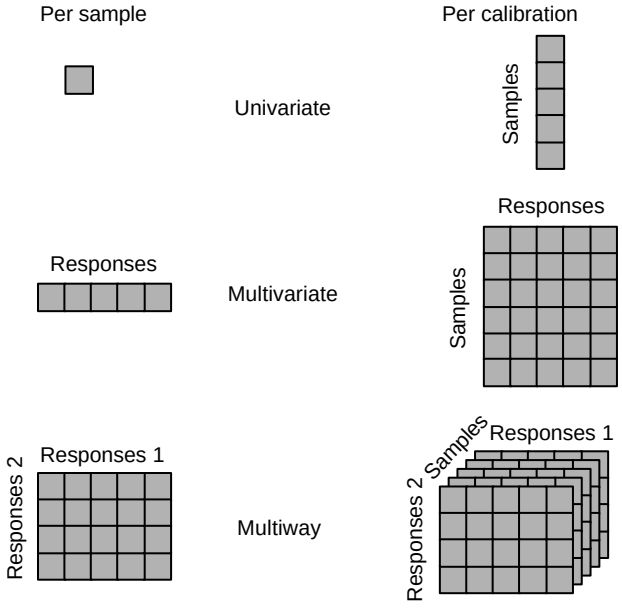

Figure 1 Progression of data structures seen in process analytical chemistry.

1.5 Nomenclature

In general, the nomenclature of referring to data as univariate, multivariate, and multiway is adopted in this article. This avoids confusion with the myriad of other mathematical and statistical definitions for ''dimension'' and ''order''. When data for a specific dimension must be identified, the order will be attached as a prefix, e.g. three-way data.

Throughout the article, scalars are italicized. Upper case italics represent fixed values, e.g. J samples. Lower case italics represent variables, e.g. the jth sample. All vectors are column vectors and represented by lower case bold characters, $\mathbf{r}$. Matrices, second-order tensors, are represented by upper case bold characters, $\mathbf{R}$, and third-order tensors are represented by bold underlined characters, $\underline{\mathbf{R}}$. A character with a subscript (or subscripts) is assumed to be the appropriate element from a higher dimensional data matrix. For example, $\mathbf{R}_k$ is the kth slice of the tensor $\underline{\mathbf{R}}$. The superscript T indicates the transpose of a matrix or vector, e.g. $\mathbf{r}^{\mathrm{T}}$. The presence of a bar over a matrix, e.g. $\overline{\mathbf{V}}$, indicates a reduced space of $\mathbf{V}$ where just the columns that are deemed useful are considered. A carat overlaying any symbol, e.g. $\hat{\mathbf{b}}$, indicates an estimate of the true value. The superscript $+$ indicates the pseudoinverse or Morse–Penrose inverse[10] of a matrix. The character $\otimes$ indicates the Hadamard product, or element-wise multiplication, of two matrices.

2 UNIVARIATE ANALYSIS

2.1 History and Introduction

Chemists have employed formal univariate data analysis since the beginning of the 20th century. Arguably the first ''process chemometrician'', chemist W.S. Gosset,

developed the *t*-test for process quality control while at Guiness in 1908. Since 1908, univariate analysis has matured. Monographs relating statistical methods for chemical process monitoring and optimization have been around since the 1940s,[11] and there have been few fundamental advances in the last 40 years. However, univariate methods form the basis for most multivariate chemometric techniques. For this reason, a brief overview of univariate analysis methods and theory is provided.

2.2 Statistical Analysis

Analysis methods can be divided into two classes, empirical methods and statistical methods. Empirical methods are generally simple and easy to compute. Examples of empirical methods include histograms and box (quartile) plots.[10] While intuitively informative, empirical methods only offer qualitative information regarding the character of a set of observations.

On the other hand, statistical methods are designed to make inferences quantitatively regarding the compliance of the observed data with the assumed nature of the system. This is generally accomplished by making assumptions regarding the probability distribution of random measurement errors. In most popular statistical methods, measurement errors are assumed to follow a normal, or Gaussian, distribution. For Gaussian errors, the mean measurement error is zero and the width of the distribution is defined by the standard deviation. This distribution is employed in Student's *t*-test to compare means (averages) of observations and in χ^2 and *F*-tests to compare variances (squared standard deviations) of observations. However, ''distribution-free'' statistical tests that do not make assumptions regarding measurement errors also exist. Further details of univariate statistical tests can be found in elementary statistics texts[12,13] or monographs aimed at chemists.[14–16]

There are two univariate statistical tools that resurface with multivariate data analysis methods. The first is means testing with a *t*-test. The Gaussian distribution of measurement errors is implicit with the *t*-test. By estimating, or assuming, the standard deviation of the measurement error, σ, a probability can be determined that a measurement, *r*, would be at most the observed difference from a true target value, μ, by random chance alone. The probability is given by the fraction of the area under the probability distribution function that is removed from μ. The *t*-test is used for placing confidence limits on statistical process control (SPC) charts and calibration predictions. The second recurring tool is the *F*-test to compare variances. The *F*-distribution is given by the ratio of two χ^2 distributions. (A χ^2 distribution is the sum of squared normally distributed variables.) If two populations have the same standard deviation (i.e. two instrumental methods have the same precision), the ratio of observed measurement variances is expected to be 1. As with the *t*-test, integrating under the probability distribution function yields the probability that the two otherwise identical populations would be observed to have a particular deviation from their expected value. *F*-tests are commonly employed to test the relative performance of competing calibration models. Tables for *t*-test and *F*-test probabilities as a function of number of observations are given in most statistical texts.

2.3 Statistical Process Control Charts

The goal of SPC charts is to provide the process engineer with a rapid and intuitive visual method to observe the state of a process. In the univariate case, a statistic is continuously plotted against time during a process run. This statistic can be a single measurement (temperature) or a weighted sum of previous measurements (e.g. an exponentially weighted moving average of the last 20 measured temperatures). The expected mean and standard deviation of the statistic are determined from analysis of past process runs that were ''in control''. The probability levels from the statistical analysis of the historical data with a *t*-test serve as warning lines for the process going out of ''control''. For example, the 95% probability limit ($t = \mu \pm 1.96\sigma$) might serve as a warning while the 99% probability limit ($t = \mu \pm 2.81\sigma$) might tell the operator to take action.

Figure 2(a) and (b) presents mock SPC charts for a process leaving and re-entering ''control''. Figure 2(a) presents the SPC chart plotted for a single measurement while Figure 2(b) presents each point as the mean of the previous five measurements. The historical mean, or target value, for the process is 176. For the first 20 min, the process is ''in control''. Note that one of the 20 measurements is outside the 95% probability limit. This is expected owing to normal variation of the process; the next measurement is well within the control limits. However, at time 21 the process begins to creep out of control and begins returning to control at time 40. Note that these trends are evident in both the single measurement and moving window SPC charts; however, the signal averaging with the moving window makes visualization of short-term trends much easier.

2.4 Calibration and Prediction

The typical goal of univariate regression is to model the instrumental response, *r*, as a linear function of the property to be estimated (e.g. concentration, *c*). Including a term in the model that accounts for an instrumental baseline, β_0, yields the classical Equation (1):

$$r = \beta_0 + c\beta_1 + \varepsilon \tag{1}$$

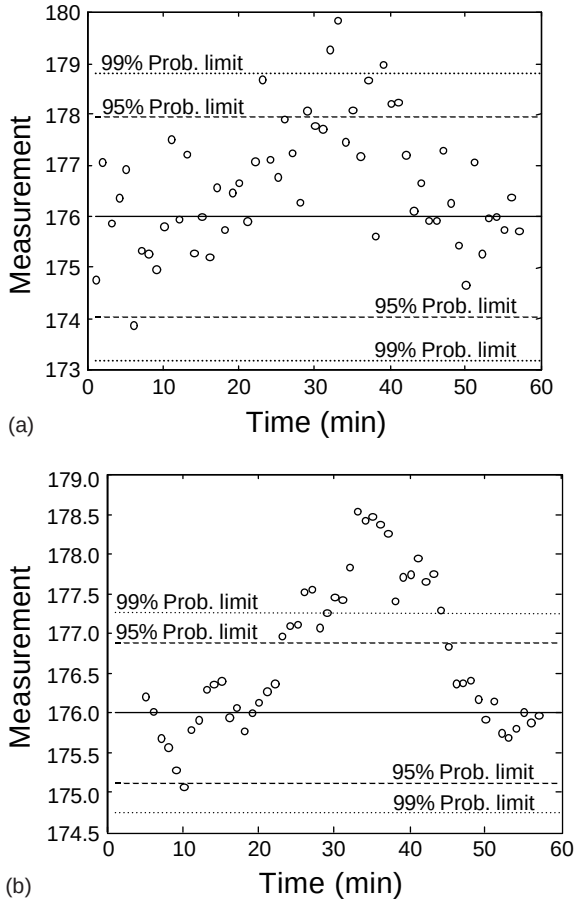

Figure 2 SPC charts with (a) no smoothing and (b) a five-point moving average.

where ε is the error associated with fitting the model to the data. Equation (1) can be expanded as a polynomial to include any number of additional terms (Equation 2):

$$r = \sum_{n=0}^{N} c^n \beta_n + \varepsilon \tag{2}$$

and expressed in matrix notation for multiple samples (Equation 3):

$$\mathbf{r} = [\mathbf{c}^0|\mathbf{c}^1|\ldots|\mathbf{c}^N]\mathbf{b} + \mathbf{e} \tag{3}$$

where $\mathbf{r}$ and $\mathbf{c}$ are $I \times 1$ matrices, the symbol $\mathbf{c}^n$ being shorthand notation for raising each element of $\mathbf{c}$ to the nth power. The $N \times 1$ matrix $\mathbf{b}$ is the regression vector that relates the matrix of expanded concentrations to instrumental response. The regression vector can be estimated by multiplying both sides of Equation (3) by an appropriate inverse of $[\mathbf{c}^0|\mathbf{c}^1|\ldots|\mathbf{c}^N]$. Since $[\mathbf{c}^0|\mathbf{c}^1|\ldots|\mathbf{c}^N]$ is not guaranteed to be a square, nonsingular matrix, either the normal equations or the Morse–Penrose pseudo-inverse of $[\mathbf{c}^0|\mathbf{c}^1|\ldots|\mathbf{c}^N]$ must be employed. Consequently (Equation 4),

$$\hat{b} = [\mathbf{c}^0|\mathbf{c}^1|\ldots|\mathbf{c}^N]^+\mathbf{r} \tag{4}$$

where "+" indicates that the pseudo-inverse of a matrix is employed to estimate the polynomial coefficients of the regression curve. Estimation of the analyte concentration in a future sample is accomplished by solving Equation (3) for $\mathbf{c}$, given $\mathbf{r}$ and $\hat{\mathbf{b}}$. The statistical and numerical properties of Equation (3) are well documented and covered in most textbooks on linear regression.[17]

2.5 Experimental Design

There are two general philosophies for experimental designs in process analysis. The first is exploratory, or discovery, experimental design and the second class is for calibration. For either class of experimental design, the data are collected in only one "way" – by varying the nature of the samples, that is, the same measurement is repeated for a variety of unique samples where each sample has been perturbed in a known manner.

The idea behind exploratory experimental designs is to uncover (and possibly to quantitate), with as few samples as possible, the underlying factors that influence the behavior of the process. Therefore, many factors (e.g. process inputs) are varied for exploratory analysis with the goal of determining which inputs are important. Examples of exploratory experimental designs for modeling the "sample space" include factorial designs and Taguchi designs.[14,18] Exploratory experimental designs, such as Fibbonacci searches[19] and simplex designs,[20] are commonly employed to optimize a system without explicitly modeling its behavior.

Whereas exploratory analysis is aimed at optimizing a process for the future, the goal of calibration is to predict the state of a sample based on past performance. That is, a model is to be constructed that will allow regression of one observation for estimation of a property of the sample. The application of calibration and prediction is covered in section 2.4. Owing to the different goals of exploratory analysis and calibration, experimental designs for calibration are much more restrictive than exploratory experimental designs. Because only one measurement per sample is to be performed, only one property of the sample can be calibrated with a given experimental design. Consequently, all other factors influencing the state of the system must be held constant for the calibration and estimation samples.

Two particular aspects of experimental design recur for multivariate analysis: leverage and sample spaces. Leverage is a measure of the influence each sample has on the model. For linear univariate calibration the leverage of the ith sample is based on the distance of the sample

from the mean of the calibration set, $\bar{r}$ (Equation 5):

$$l_i = (r_i - \bar{r})^2 \tag{5}$$

The greater the total leverage, the more stable is the calibration model, and consequently, the better is the precision of estimation with the calibration model. From this, the astute reader may conclude that it is optimal to have half of the calibration samples at one extreme of the calibration model and half of the calibration samples at the other extreme of the calibration model (Figure 3, circles). However, although this optimizes the experimental design for leverage, the sample space might not be optimally covered. Suppose that the instrumental response to be calibrated is nonlinear. There would be no way to model, or even detect, this nonlinearity with the leverage optimized experimental design. By collecting samples that better fill the sample space, nonlinearities can be detected and modeled (Figure 3, crosses). Optimization of the exact sample placement depends on the formula of the nonlinear curve to be fitted to the data.

It should also be noted that, in many cases of process analysis, the analyst does not have the luxury of applying a truly optimized experimental design. In the extreme case, the analyst has no influence on the experimental design; any variability in the sample space is governed by the inherent variance of the process. Realistically, there is a limited amount of process variance that can be induced by the analyst; however, owing to safety and economic constraints, the process cannot be run outside its normal parameters. In these cases it must be remembered that any models developed are valid only within the sample space. There are no assurances of the applicability of the models outside the range of samples from which the models were based.

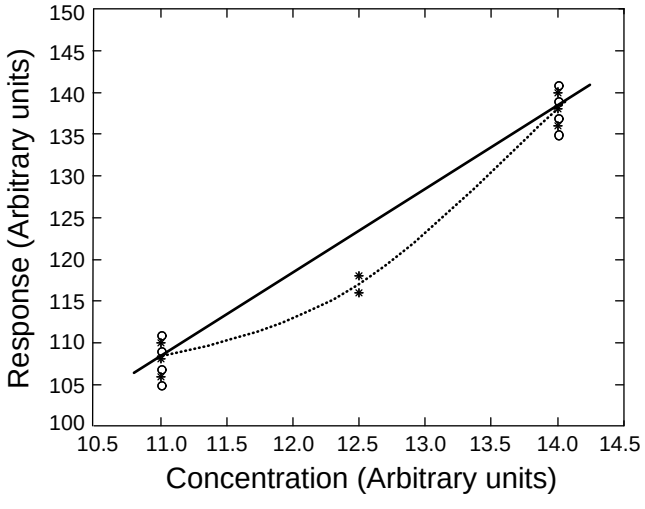

Figure 3 Two potential univariate experimental designs: (○) optimized for linear model and (∗) containing interior measurements to check for nonlinearities.

2.6 Figures of Merit

In order to compare instrumental methods it is useful to have a handful of objective and quantifiable criteria on which discussion is based. Examples include speed, cost, reliability, precision, sensitivity (SEN), selectivity (SEL), and detection limit of analysis.[21] While speed, cost, and reliability weigh heavily in a process analyst's decision on which instrumental technique to employ for a particular application, these figures of merit are not intrinsic to a given instrumental method. For example, the cost and speed of analysis largely depend on the number of samples to be analyzed; with large quantities of samples, economic and time savings can be achieved through bulk purchasing and automation. However, the final four figures of merit are intrinsic to the application of an instrumental method and are directly related to the instrumental response for a particular set of analytes.

The SEL is the fraction of the instrumental signal that is unique to the analyte. Assuming that the instrument is "zeroed" to remove any baseline (Equation 6):

$$\text{SEL} = \frac{r_\text{a}}{r} \tag{6}$$

where r_a is the instrumental signal of just the analyte and r is the instrumental signal of the sample. The SEL is a value between 0 and 1 with SEL = 1 being a fully selective sensor. For univariate calibration an instrument must be fully selective, otherwise a bias will be imbedded in the prediction of future samples. There is no way, based only on statistical analysis of collected data, to determine the contribution (or existence) of nonselective interferents in any given "unknown" sample.

The SEN is the change in instrumental response with respect to changes in analyte concentration (Equation 7):

$$\text{SEN} = \frac{\delta r}{\delta c} \tag{7}$$

For linear calibration, this is the slope of the calibration curve.

The precision of a method is best expressed in the signal-to-noise ratio (S/N) (Equation 8):

$$\text{S/N} = \frac{r_\text{a}}{e} \tag{8}$$

where e is a measure of the reproducibility of replicated measurements. In many cases, the measurement reproducibility is not concentration dependent (e.g. for thermal noise limited analyses). In this case S/N will vary with analyte concentration.

The limit of detection (LOD) is defined by the International Union of Pure and Applied Chemistry

(IUPAC) and the American Chemical Society (ACS) to be the smallest amount of a chemical that can be reasonably detected by a given analytical method. Of the many ways to calculate the LOD, solving Equation (3) for c given a signal, r, equal to three standard deviations of replicated instrumental blanks is the most straightforward. Assuming that the calibration model is linear and there is no instrumental offset, the detection limit can be expressed as Equation (9):

$$\mathrm{LOD} = \frac{3e}{\beta_1} \tag{9}$$

where β_1 is the slope of the calibration.[22]

3 MULTIVARIATE ANALYSIS

3.1 History and Introduction

The success of multivariate chemical analysis is based on three facts. First, most, if not all, chemical processes are multivariate in nature. Consequently, to be able to monitor and model a multivariate process effectively, multivariate data regarding that process must be collected and analyzed. Second, even if only a single piece of information is needed from a chemical system, it is very difficult to design a sensor that is fully selective to the property of interest. Therefore, to circumvent the lack of fully selective sensors, arrays of partially selective sensors can be constructed that rely on multivariate analysis methods to extract the information of interest. Third, there are inherent advantages associated with the redundancy of data when there are many more variables measured per sample than samples collected.

Initially, chemometrics relied heavily on mathematical and statistical methods developed for analyses of biological, economic, and psychological data. In the 1970s, a few academic teams were investigating the benefits of multivariate analysis for instrument calibration. Perhaps overplaying the field, it was envisioned that multivariate data analysis and calibration methods would make most instrumental methods (such as chromatography) obsolete. However, the impact of multivariate data analysis and calibration on process analytical chemistry was negligible.

As personal computers became increasingly powerful and less expensive and chemists began to develop an appreciation for the benefits derived from multivariate analysis and an understanding of the limits of multivariate analysis, multivariate methods slowly became accepted in industry. Not coincidentally, the acceptance of multivariate analysis coincided with the rise of the process analytical chemistry philosophy.[23,24] Most of the sensors designed for on-line, in situ, or noninvasive analysis require multivariate analysis for calibration or data interpretation. This section presents the basic outline for application of the multivariate method useful in process analysis. Particular attention is directed to discussing the utility of selected multivariate methods as they apply to process analysis.

3.2 Factor Analysis

FA is employed to aid in the visualization of sample (time)-dependent trends and measurement (sensor)-dependent trends in a multidimensional data space. In general, FA does not give a physically meaningful model – only correlations among samples and measurements are determined. However, FA methods have been modified to apply constraints and assumptions based on previous knowledge of the chemical system being analyzed. These modified FA methods are useful for determining the underlying instrumental and/or sample (time) profiles of the chemical constituents of a process.

3.2.1 Principal Component Analysis

Perhaps the most commonly applied method of FA is principal component analysis (PCA).[25,26] Most chemometrics-oriented monographs have sections dedicated to PCA. The goal of PCA is to identify the major sources of correlated variance in a collection of data. Once these sources of variance have been identified, they can be exploited to aid in the visualization of the major trends throughout the data collection. The data collection can be reduced from a complicated multidimensional representation (a 1024-channel spectrum is actually a vector in 1024-dimensional space) to a more easily visualized two- or three-dimensional space that describes the majority of the variance (information) in the data collection.

The conceptual idea behind PCA is presented in Figure 4. The largest direction of variance in the data collection is the first principal component (PC). The second PC is defined to describe the maximum amount of variance in the data collection while constrained to be orthogonal to the first PC. Consequently, each additional PC is also defined to maximize variance described while constrained to be orthogonal to all preceding PCs. Note that the PCs are defined as vectors originating at the origin of the coordinate space. Therefore, the PCs are dependent on the average value of the data collection; translating the data cloud to a different point in the coordinate space changes the direction of the PCs. For this reason, the data collection is often translated to be centered about the origin of the coordinate space (see section 5.1). However, the location of the data collection does not affect the ability of PCA to model the data variance. Only the ease of interpreting the model is affected.

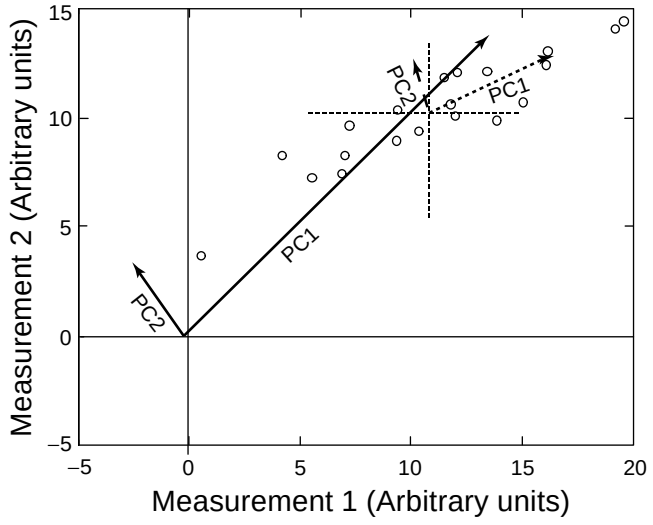

Figure 4 Graphical representation of PCA performed on 20 measurements (○). Solid lines: PCA performed in the measurement space without preprocessing. Dashed lines: PCA performed after adjusting for the mean of each measurement.

3.2.1.1 Determination of Principal Components PCA is a form of eigenanalysis where a matrix of data **R** is decomposed into the product of two (or three) matrices. The most common matrix representations employed for PCA are (Equations 10 and 11)

$$\mathbf{R} = \mathbf{TP}^{\mathrm{T}} \tag{10}$$

and

$$\mathbf{R} = \mathbf{USV}^{\mathrm{T}} \tag{11}$$

where each of the I rows of **R** consists of the J measurements performed on the I samples. Graphically, Equation (11) is illustrated in Figure 5. In practice, the factorization of **R** in Equations (10) and (11) can be performed by iterative algorithms that are found in most numerical analysis toolboxes. The most numerically stable method is the SVD.[19]

The nth column of the matrices **P** and **V** are the loadings of the nth PC and describe the inter-relationships among

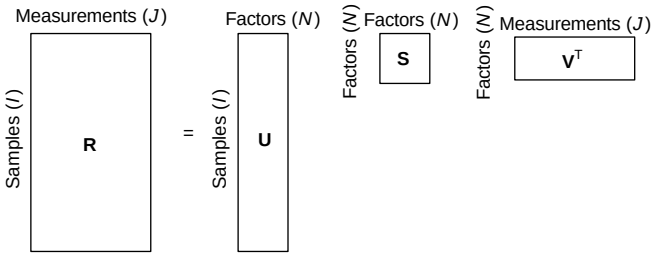

Figure 5 Block representation of the singular value decomposition (SVD) employed in PCA.

the J measurements. The matrices **P** and **V** are identical. The columns of the matrices are the orthonormal eigenvectors of the covariate matrix $\mathbf{R}^{\mathrm{T}}\mathbf{R}$. Consequently, **P** and **V** describe the "measurement space" of **R**. Positive values are all positively correlated with the direction of variance described by a given PC; negative values are negatively correlated with the direction of variance described by the PC. For example, if a PC is found to describe the variance caused by a process upset, sensor readings that increase during (or preceding) the upset will all have the same sign in the loading. Sensors that decrease during the upset will have the opposite sign. Sensors that do not vary as a result of the upset will have loadings near zero on this PC. The absolute sign of the PC is arbitrary; any PC can be equivalently expressed by reversing all signs.

Equations (10) and (11) differ only in the factorization of the "sample space". In Equation (10), the inter-relation among the samples is expressed as the orthogonal column matrix **T**. Because the columns of **P** are normalized to unit length, the "scores" matrix, **T**, also contains information regarding the scale of the data matrix, **R**. By contrast, in Equation (11), the orthonormal columns of **U** contain the inter-relation among the samples that constitute **R**. The columns of **U** are the eigenvectors of the covariance matrix $\mathbf{RR}^{\mathrm{T}}$. The diagonal matrix, **S**, contains the scale of the data matrix **R**. The squared diagonal elements of **S** are the eigenvalues of the covariance matrices $\mathbf{R}^{\mathrm{T}}\mathbf{R}$ and $\mathbf{RR}^{\mathrm{T}}$. The scores of Equations (10) and (11) are related by Equation (12):

$$\mathbf{T} = \mathbf{US} \tag{12}$$

3.2.1.2 Nomenclature of Principal Component Analysis The theory and development of PCA can be traced in the mathematics, statistics, and social science literature. Consequently, the nomenclature employed by any given author will depend greatly on the author's previous background and exposure to PCA. Table 1 presents a compilation of the descriptive terms employed in the various fields. Of note are the social science terms "score"

Table 1 Common nomenclatures for PCA

Field	Singular values^2 ($\mathbf{S}^2$)	Measurement space (**V**)	Sample space (**U**)
Mathematics	Eigenvalues	Eigenvectors	Eigenspace coordinates
	Characteristic roots	Characteristic vectors	Projection coordinates
Statistics	Factor variance	PC vectors	
Social sciences	Factor variance	Loading vectors	Score vectors

and ''loading''. These terms reveal the anthropologist Sir Francis Galton's assumption that there are certain latent factors that are ''loaded'' into each person. These latent factors are common throughout the population in general and give rise to each person's individual personality. The greater an underlying trait is exhibited by a person, the higher the person's ''score'' for that trait would be. As a side note, the term ''regression'' also comes from the psychological literature. It was postulated, and disproved, that children of extraordinary parents tended to ''regress to the mean'' of the population.[27]

3.2.1.3 Interpretation of Principal Component Analysis Models There are two important facts to remember when interpreting PCA models. First, the factorization of **R** is arbitrary; there are infinite factorizations that equivalently model **R**. PCA is one particular factorization that satisfies the constraints of orthogonal factors and progressive modeling of variance. All other equivalent models that can be constructed are rotations of the eigenspace determined by PCA. Therefore, the loadings are not necessarily, for example, spectra and cannot be interpreted as spectra, even though they often resemble spectral features. Furthermore, the factors do not necessarily describe any single underlying latent physical feature. PCA is purely a correlation based model that describes observed variance. Any causality inferred from the model is implied by the analyst.

Second, not all of the factors determined by PCA describe chemically relevant variance. Some factors describe systematic variance that results from instrumental or sampling artifacts. Also, some factors describe random measurement and sampling errors. The success of PCA lies in the ability of the analyst to discriminate between these three sources of variance. In general, the chemically relevant variance and any systematic variance from instrumental and sampling artifacts are convolved in the first few PCs. Derivatives of PCA such as target factor analysis (TFA) (section 3.2.2), curve resolution (CR) methods (section 3.2.3), and evolving factor analysis (EFA) methods (section 3.2.4) aid the analyst in divining the true, chemically relevant, factors. Section 6.1 addresses determination of systematic factors versus random factors.

3.2.2 Target Factor Analysis

As mentioned in the previous section, the latent factors derived from PCA do not necessarily describe single, physically meaningful, effects. That is, while a set of data may consist of the near-infrared (NIR) spectra of hydrocarbon mixtures, the PCs of the data set are not constrained to be NIR spectra of the constituent hydrocarbons. However, the multivariate space defined by the PCs is the same as the multivariate space defined by the pure (true) spectra of the chemical constituents of the data set (plus any other forms of systematic variance).

TFA is a method of testing whether the spectrum of a hypothesized chemical constituent, as defined by an assumed or a priori recorded spectrum, lies in the PC space of the model. If the hypothesized constituent does lie in the PC space, the associated spectrum can be expressed as a linear combination of the PC loadings. That is (Equation 13),

$$\hat{\mathbf{y}} = \overline{\mathbf{V}}\mathbf{t} \tag{13}$$

where $\hat{\mathbf{y}}$ is the estimated spectrum of the hypothesized compound and **t** is the transformation vector [not to be confused with **T** in Equation (10)]. The transformation vector can be calculated by regression of the target spectrum, **y**, on to the loadings **V**. However, the calculation of **t** can also be circumvented by direct projection of **y** into the space defined by the loadings (Equation 14):

$$\hat{\mathbf{y}} = \overline{\mathbf{V}}\overline{\mathbf{V}}^{\mathrm{T}}\mathbf{y} \tag{14}$$

The presence or absence of **y** in **R** is tested by comparing **y** with the projection of **y** into the factor space, $\hat{\mathbf{y}}$. If **y** and $\hat{\mathbf{y}}$ are determined to be sufficiently similar by a statistical or empirical test,[25] analogous equations can be constructed that project sample targets on to the scores of **R**.

3.2.3 Curve Resolution

Where TFA allows the analyst to test for the presence of a hypothesized constituent, TFA is limited in the ability to estimate the spectral profile of any constituents in the data set. TFA requires that the spectral profile of the target is available for target testing. If the profile is unavailable, TFA cannot be performed. On the other hand, CR methods allow for the estimation for both the hypothesized and unknown constituents in the data matrix. The rotational ambiguity of the decomposition in Equations (10) and (11) is circumvented by making assumptions regarding the nature of the true constituent spectral profiles (and sample profiles). These assumptions are translated into constraints applied to the factorization of **R**. Once additional constraints are applied to the factors of **R**, the factors are not true PCs. These factors are properly described as latent variables or intrinsic factors, but not PCs.

3.2.3.1 Constraints Numerous constraints have been applied to the factorization of **R** to enhance the probability that the determined factors will be physically meaningful. Perhaps the most common constraint is non-negativity of estimated spectral and sample profiles.[28–33]

This constraint is based on the commonsense notion that the factorization of Equation (10) should lead to spectral estimates (columns of **P**) and concentration estimates (columns of **T**). In neither case would the true profile likely contain negative values. Another common spectral constraint employs assumptions regarding the content of **R**. If the spectral profile of one or more of the assumed chemical constituents is known, the factorization of Equation (10) can be constrained to contain the assumed spectral profiles in the solution.

It is also possible to employ assumptions regarding the inter-relationship among the samples (rows of **R**). For resolving overlapping chromatographic peaks, Gaussian or unimodal elution profiles are assumed for the columns of **P**.[31,33,34] Concurrently, the presence of samples (rows of **R**) that contain only one compound may be successfully postulated for chromatographic or kinetic data.[31,35] This is referred to as the ''uniqueness'' constraint. If the concentration of one or more compounds is known in any of the particular samples, the resolved profiles can be constrained to reflect this information. For kinetic data, the sample profiles can be constrained to fit a class of differential equations that reflect the postulated reaction pathway.[36,37] The validity of the assumed reaction can be tested based of the ability of the data to fit this model.

Of course, application of other constraints and the combination of multiple constraints are possible. Constraining the FA solutions does not ensure that physically meaningful profiles will be determined. In general, application of constraints only reduces the range of feasible solutions where, ideally, the true profiles will lie within this range. The more constraints are properly applied to the decomposition, the tighter the estimated range of profiles will resemble the true profiles. However, if a constraint is improperly imposed (e.g. nonnegativity when in fact the profile should have negative values) the estimated profiles will yield erroneous profiles.

3.2.3.2 Direct Methods The application of constraints for estimation of spectral profiles can best be seen in the CR method developed by Lawton and Sylvestre.[28] Lawton and Sylvestre CR was originally applied to binary mixtures. After normalization of each collected spectrum to unit area (section 5.1), the data are factored by Equation (10). Plotting the scores of the first PC versus the scores of the second PC yields a straight line for a binary mixture (Figure 6). The extreme points along the array represent the most pure of the analyzed samples. These form the ''inner bounds'' of the estimated profiles. The ''outer bounds'' are calculated by extrapolating the scores line (dotted) and reinflating the scores into the original data space to calculate estimates of the pure spectra (Equations 15 and 16):

$$\hat{\mathbf{y}}_a = p_{1a}\mathbf{t}_1 + p_{2a}\mathbf{t}_2 \tag{15}$$

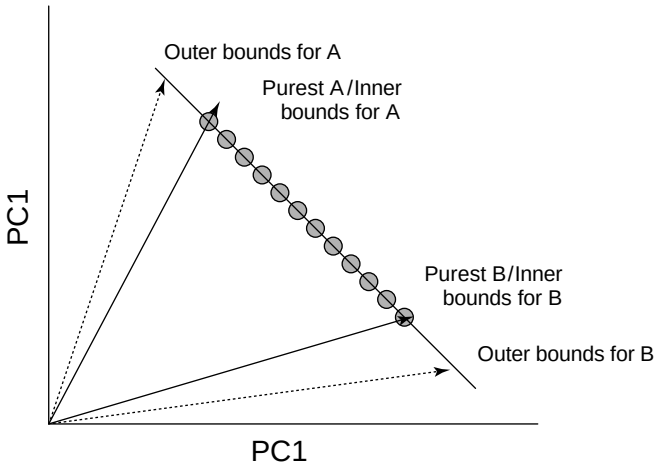

Figure 6 Determination of inner and outer spectral estimates in Lawton and Sylvestre CR.

$$\hat{\mathbf{y}}_b = p_{1b}\mathbf{t}_1 + p_{2b}\mathbf{t}_2 \tag{16}$$

In these equations, values for (p_{1a}, p_{2a}) and (p_{1b}, p_{2b}) are chosen as the extreme points along the extrapolation of the best-fit line in Figure 6 (p_1 vs p_2) that do not result in negative values in the estimates of pure spectra.

Consequently, the only constraint that Lawton and Sylvestre CR applies to the data is nonnegativity of spectral estimates. When there is at least one spectral channel that is unique to the analyte, then the outer bounds will be the true constituent spectral profile. This CR technique has been generalized to three-component mixtures.[38] However, the bounds of the potential spectra are generally too large with tertiary mixtures for this CR method to be of much utility.

3.2.3.3 Iterative Methods Practically, iterative CR methods are capable of resolving much more complicated mixtures than are the direct CR methods. Iterative methods make assumptions regarding both the spectral space profiles of the data and the sample space profiles of the data. The constraints resulting from these assumptions are particularly powerful when well-ordered data, such as kinetics or chromatographic data, are submitted. A description of the potential constraints applied to kinetic or chromatographic systems is covered in section 3.2.3.1.

Iterative CR methods employ the bilinear factorization model in Equation (17):

$$\mathbf{R} = \hat{\mathbf{X}}\hat{\mathbf{Y}}^t + \mathbf{E} \tag{17}$$

where the N columns of $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ are estimates of the true sample (concentration) and spectral profiles of the N constituents of **R**. $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ are determined by an alternating least-squares (ALS) algorithm. This algorithm starts with an estimate of either **X** or **Y**. Assuming an

estimate of $\mathbf{X}$ is employed for initialization, estimates of the spectral profiles are calculated by least squares. Either a constrained least-squares method, e.g. non-negative least squares,[19] is employed or the constraints are imposed after $\widehat{\mathbf{Y}}$ is calculated directly by least squares (Equation 18):

$$\widehat{\mathbf{Y}}^t = \widehat{\mathbf{X}}^{+}\mathbf{R} \tag{18}$$

In this second case, for example, nonnegativity would be imposed by manually setting all negative entries of $\widehat{\mathbf{Y}}$ to zero.

Once the constrained estimates of $\mathbf{Y}$ have been calculated, $\widehat{\mathbf{Y}}$ is employed to update the estimate of $\mathbf{X}$. As with calculating $\widehat{\mathbf{Y}}$, constraints can be imposed during the calculation of $\widehat{\mathbf{X}}$ by a constrained least-squares method, or after the estimation of $\mathbf{X}$ by ordinary least squares (Equation 19):

$$\widehat{\mathbf{X}} = \mathbf{R}\widehat{\mathbf{Y}}^{t+} \tag{19}$$

The method iterates by alternating calculating constrained updates of $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$ until further refinement does not significantly change the model.

One common iterative method is EFA,[29,34,39] which employs the ALS optimization at its core. The unique feature of EFA is the use of eigenanalyses to determine the initial $\widehat{\mathbf{X}}$. The singular values are calculated from analysis of the first j samples. These ''forward eigenvalues'' are plotted against j as j is increased from 1 to J, the number of samples. Concurrently, the ''backward eigenvalues'' are calculated from analysis of the last j samples as j increases from 1 to J. These eigenvalues are plotted against $J - j + 1$.

This plot of eigenvalue versus sample set yields an estimate of the number of significant factors in $\mathbf{R}$ and a rough estimate of the sample profiles. The N significant eigenvalues will be relatively large compared with the remaining eigenvalues and exhibit greater variance throughout the range of j. The area under the intersection on the nth forward and $(N - n + 1)$th backward eigenvalue versus j plot yields the estimate of the nth sample profile. These sample profiles are employed as starting guesses for the ALS loop of the EFA method.

3.3 Multivariate Statistical Process Control Charts

When multiple measurements are collected throughout a process, it is impractical to monitor simultaneously a distinct SPC chart for every measurement. A more optimum course of action is to construct a smaller set of control charts. In this smaller set, each chart would weigh an ensemble of measurements that correlate to important trends in the process. One way to accomplish this is to form multivariate statistical process control (MSPC) charts.[40–42] Whereas univariate SPC charts plot just a single measurement versus time, MSPC charts first employ FA to condense the information in the measurements into a smaller set of significant factors, and then the observed contribution of these factors in each set of measurements is plotted versus time.

The simplest form of MSPC chart would apply PCA to historical runs of a given process to determine the significant factors that occur (loadings). When monitoring the process in the future, each set of measurements, $\mathbf{r}$, is projected on to each loading. The magnitudes of these projections (scores) (Equation 20):

$$t_n = \mathbf{p}^{\mathrm{T}}\mathbf{p}_n(\mathbf{p}_n^{\mathrm{T}}\mathbf{p}_n)^{-1} \tag{20}$$

are plotted for each of the N factors as a function of time. This yields an indication of the magnitude of each factor at a given time during the process run. Probability limits can be assigned based on variance in the factors of the historical data.

Iterative FA techniques can also be applied to construct MSPC charts. These methods can be employed to model a process dynamically as new species (factors) are formed. Each new set of measurements is augmented to the existing collection of data from the process run and the data are reanalyzed. If an additional factor is warranted one is added to the model. The sample profiles are updated as the new measurements are collected.

When constructing control charts from multivariate data, the Hotelling's T^2 statistic provides an important test to check for multivariate normality of the data. The Hotelling's T^2 test is the multivariate analog for the Student's t-test. The T^2 statistic for the ith sample is calculated from the factorized data by the sum of the squared (normalized) scores (Equation 21):

$$T_i^2 = \sum_{n=1}^{N} u_{in}^2 \tag{21}$$

from an N factor model. The scaled statistic $T^2 I(I - N)/[N(I^2 - 1)]$ follows an F-distribution with N and $I - N$ degrees of freedom.

Process upsets and sensor failures can be identified by examining the residuals derived from projecting future samples onto the FA model constructed from archival data (Equation 22):

$$\mathbf{e}_r = \mathbf{r}^{\mathrm{T}}(\mathbf{I} - \overline{\mathbf{V}}\overline{\mathbf{V}}^{\mathrm{T}}) \tag{22}$$

where $\overline{\mathbf{V}}$ represents the set of columns in the loading matrix that are ''significant'' in describing the process and $\mathbf{I}$ is an appropriately dimensioned identity matrix. Whether the sample $\mathbf{r}$ lies outside the space of the model can be investigated by comparing the magnitude of $\mathbf{e}_r$

with the magnitude of the residual error of the model (Equation 23):

$$\mathbf{E}_R = \mathbf{R}^{\mathrm{T}}(\mathbf{I} - \overline{\mathbf{V}}\overline{\mathbf{V}}^{\mathrm{T}}) \tag{23}$$

Adjusting for number of observations and degrees of freedom, an approximate F-statistic (Equation 24)

$$F = \frac{\|\mathbf{e}_r\|_{\mathrm{F}}^2}{(J-N)} \frac{(I-N-1)(J-N)}{\|\mathbf{E}_R\|_{\mathrm{F}}^2} \tag{24}$$

can be constructed with $(J-N)$ and $(I-N-1)(J-N)$ degrees of freedom, where $\| \ \|_{\mathrm{F}}$ is the Froebus norm (root sum of squared elements) of a matrix.

3.4 Calibration and Prediction

The instrumental response of a mixture measured at a single variable relies on the classical least-squares (CLS) model (Equation 25):

$$\mathbf{r} = \mathbf{Cs} \tag{25}$$

where $\mathbf{r}$ is the instrument response from a number of samples, $\mathbf{C}$ is a matrix (samples $\times$ constituents) of the extrinsic property of interest (often concentration) and $\mathbf{s}$ is an intrinsic property (often spectral in nature) that relates the rows of $\mathbf{C}$ to $\mathbf{R}$. This follows a logical Beer's law type of model where the instrument response is expressed as a function of the constituent's concentration and spectra. Equation (25) is readily expanded to multivariate–multianalyte analysis (Equation 26):

$$\mathbf{R} = \mathbf{CS}^{\mathrm{T}} + \mathbf{E} \tag{26}$$

where each row of $\mathbf{R}$ is the spectrum of the mixture of constituents defined by the appropriate row of $\mathbf{C}$. The constituent spectra lie in the columns of $\mathbf{S}$.

Calibration and prediction with Equation (26) are accomplished by multilinear regression (MLR).[4,16,19] For a set of recorded spectra, $\mathbf{R}$, with known constituent extrinsic properties, $\mathbf{C}$, $\mathbf{S}$ can be estimated by ordinary least-squares inversion of $\mathbf{C}$ (Equation 27):

$$\widehat{\mathbf{S}}^{\mathrm{T}} = \mathbf{C}^{+}\mathbf{R} \tag{27}$$

Premultiplication of a future recorded spectrum, $\mathbf{r}_{\mathrm{un}}$, by the (pseudo)inverse of $\mathbf{S}$ affects the estimation of extrinsic properties in the unknown sample (Equation 28):

$$\hat{\mathbf{c}}_{\mathrm{un}} = \widehat{\mathbf{S}}^{+}\mathbf{r}_{\mathrm{un}} \tag{28}$$

Consequently, there are advantages and disadvantages associated with MLR. The primary advantage is that estimates of the true constituent spectra are derived during calibration. However, MLR requires knowledge of the concentrations of all constituents with a spectral profile in the calibration set. From a practical standpoint, this may be an unreasonable requirement. In many cases, e.g. spectroscopic estimation of octane number or Reed vapor pressure, no "spectrum" of the extrinsic property would exist. The second disadvantage is that, in determining $\widehat{\mathbf{S}}$ by least squares in Equation (27), it is assumed that the errors in $\mathbf{R}$ are much greater that the errors in $\mathbf{C}$. This is not usually the case in chemical applications.

Employing an inverse least-squares (ILS) model alleviates the need for complete knowledge of the calibration set constitution. With the ILS model, concentration (or any intrinsic property) is modeled as a function of instrument response (Equation 29):

$$\mathbf{c} = \mathbf{Rb} \tag{29}$$

where the vector $\mathbf{b}$ relates the variance in the observed data, $\mathbf{R}$, to $\mathbf{c}$. The "regression vector" $\mathbf{b}$ is chosen to be correlated with $\mathbf{c}$ and contravariate (orthogonal) to all sources of variance in $\mathbf{R}$ that are not correlated with $\mathbf{c}$. This regression vector is purely a mathematical construct and has no physical meaning. By employing such "soft modeling" calibration based on correlation, knowledge of only the extrinsic property of interest to be estimated is required for calibration.

Figure 7 presents the concept of a contravariate regression vector in a two-dimensional space. Note that the regression vector for analyte A is orthogonal to that for analyte B. Were a different species present instead of B, then the regression vector for A would be slightly rotated to maintain orthogonality. The projection of A on to the regression vector is said to be the net analyte signal (NAS).[43] The NAS is the part of A that is employed

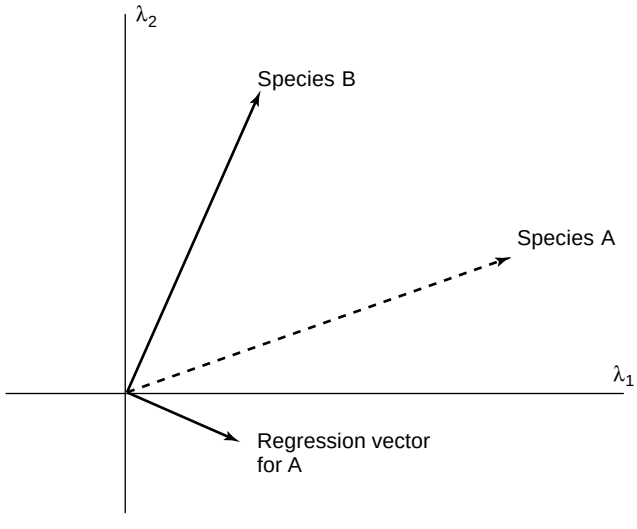

Figure 7 Contravariate representation of regression vectors in two dimensions.

for calibration. In general, the larger the NAS the more stable is the regression model. The role of NAS is further discussed in section 3.5.

This section presents the formulation of common linear and nonlinear calibration methods that employ the ILS procedure. The successful application of the methods often depends on judicious preprocessing of the recorded variables and choice of model complexity. These two topics are covered in sections 5 and 6, respectively.

3.4.1 Linear Methods

3.4.1.1 Principal Components Regression One common and robust ILS calibration method is principal component regression (PCR).[26,44–46] In essence, PCR performs calibration with the scores from the PCA decomposition of **R**. The first step of PCR is factorization of **R** by Equation (11). Once the number of significant factors has been determined, error reduction is accomplished by discarding factors that describe random variance. Thus (Equation 30)

$$\mathbf{R} \approx \overline{\mathbf{U}}\,\overline{\mathbf{S}}\,\overline{\mathbf{V}}^{\mathrm{T}} = \overline{\mathbf{R}} \tag{30}$$

Because PCR [and partial least-squares regression (PLSR) below] discard variance from the calibration (truncated columns of **U** and **V**) set prior to regression, these methods are said to be biased. However, any loss of accuracy is compensated by an improvement in precision. The net result is a general lessening of prediction errors compared with employing all of the variance of **R** in calibration.

The second step in PCR is the estimation of the regression vector **b**. Either a subset of the scores, $\overline{\mathbf{U}}$, or the random error reduced response matrix, $\overline{\mathbf{R}}$, can be related to the dependent variable, **c**. However, both options are equivalent. The orthonormal character of the scores and loadings makes estimation of **b** trivial once **R** has been factorized by Equation (11). Because both $\overline{\mathbf{U}}^{\mathrm{T}}\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}^{\mathrm{T}}\overline{\mathbf{V}}$ are identity matrices (Equation 31):

$$\hat{\mathbf{b}} = \overline{\mathbf{R}}^{+}\mathbf{c} = \overline{\mathbf{V}}\overline{\mathbf{S}}^{-1}\overline{\mathbf{U}}^{\mathrm{T}}\mathbf{c} \tag{31}$$

where the inverse of the square diagonal matrix **S** is the inverse of the elements along the diagonal. Estimation of **c** for future samples is accomplished by multiplication of the regression vector with the recorded instrumental response of the sample in question (Equation 32):

$$\hat{\mathbf{c}}_{\mathrm{un}} = \mathbf{r}_{\mathrm{un}}^{\mathrm{T}}\hat{\mathbf{b}} \tag{32}$$

3.4.1.2 Partial Least-squares Regression PLSR[45–48] has been employed since the early 1980s and is closely related to PCR and MLR.[34,45] In fact, PLSR can be viewed as a compromise midway between PCR and MLR.[49] In determining the decomposition of **R** (and consequently removing unwanted random variance), PCR is not influenced by knowledge of the estimated property in the calibration set, **c**. Only the variance in **R** is employed to determine the latent variables. Conversely, MLR does not factor **R** prior to regression; all variance correlated with **c** is employed for estimation. PLSR determines each latent variable to optimize simultaneously variance described in **R** and correlation with **c**, ρ. Technically, PLSR latent variables are not "principal components". The PLSR factors are rotations of the PCA PCs for a slightly different optimization criterion.

Mathematically, the first PLSR loading is determined to maximize Equation (33):

$$(VAR \times \rho)^2 = (\mathbf{p}_1^{\mathrm{T}}\mathbf{p}_1)^2\frac{(\mathbf{p}_1^{\mathrm{T}}\mathbf{c})^2}{(\mathbf{p}_1^{\mathrm{T}}\mathbf{p}_1)(\mathbf{c}^{\mathrm{T}}\mathbf{c})} = \mathbf{s}^4\frac{(\mathbf{u}_1^{\mathrm{T}}\mathbf{c})^2}{\mathbf{c}^{\mathrm{T}}\mathbf{c}} \tag{33}$$

which can be expressed in terms of Equation (10) or (11). Successive PLSR loadings maximize Equation (33) under the constrained orthogonality to prior PLSR loadings. PLSR scores are determined by projection of **R** on to the PLSR loadings. The PLSR factors are determined by iterative algorithms that are described in most chemometric texts[3,4,16,25] and elsewhere.[44–46] In fact, numerous algorithms exist that are optimized for various sizes of **R**.[50,51]

PLSR has two distinct advantages over PCR. First, PLSR generally provides a more parsimonious model than PCR. PCR calculates factors in decreasing order of **R**-variance described. Consequently, the first factors calculated, that have the least imbedded errors, are not necessarily most useful for calibration. On the other hand, the first few PLSR factors are generally most correlated with concentration. As a result, PLSR achieves comparable calibration accuracy with fewer latent factors in the calibration model. This further results in improved calibration precision because the first factors are less prone to imbedded errors than are lower variance factors.

Second, the PLSR algorithm is often faster to implement and optimize for a given application than is the PCR algorithm. PLSR calculates the factors one at a time. Hence only the latent variables needed for calibration are determined. PCR, employing the SVD, calculates all possible PCs for **R** prior to regression. For large data sets that require relatively few factors for calibration, PLSR can be significantly faster than PCR.

3.4.1.3 Locally Weighted Regression The global linear models calculated by PCR and PLSR are not always the best strategy for calibration. Global models span the variance of all the samples in the calibration set.

If the data are nonlinear (a Beer's law type of relationship between constituents, spectra, and predicted property does not hold.) then the linear PCR and PLSR methods do not efficiently model the data. One option is to employ nonlinear calibration methods. Nonlinear methods employ a global, nonlinear model and are discussed in section 3.4.2. The second option is to employ linear calibration methods on small subsets of the data.

The locally weighted regression (LWR) philosophy assumes that the data can be efficiently modeled over a short span with linear methods.[52–56] The first step in LWR is to determine the N samples that are most similar to the unknown sample to be analyzed. Similarity can be defined by distance between samples in the spectral space,[54] by projections into the PC space,[55] and by employing estimates of the property of interest.[56] Once the N nearest standards have been determined, either PLSR or PCR is employed to calculate the calibration model.

LWR has the advantage of often employing a much simpler, and more accurate, model for estimation of a particular sample. However, there are three disadvantages associated with LWR. First, two parameters must be optimized for LWR, number of local samples and number of factors, compared with just one parameter for PLSR and PCR. Second, a new calibration model must be determined for every new sample analyzed. Third, LWR often requires more samples than PCR or PLSR in order to build meaningful, local calibration models.

3.4.2 Nonlinear Methods

Numerous nonlinear, multivariate calibration methods have been described in the chemometric literature. These methods can be divided into two classes. Alternating conditional expectations (ACE)[57,58] and projection pursuit (PP)[59] seek to transform the nonlinear data such that a linear calibration model is appropriate. Similarly, global linearizing transformations (GLT) are employed to optimally linearize data prior to FA by PCA.[60,61] On the other hand, nonlinear partial least squares (NPLS),[62,63] multivariate adaptive regression splines (MARS),[64,65] and artificial neural network (ANN) methods determine nonlinear global models that span the entire range of samples. Although it is impossible to provide sufficient detail for each method in this article, some general comments regarding the application of these nonlinear methods are warranted.

Specific nonlinear methods have been compared and contrasted over a wide variety of linear and nonlinear calibration applications.[66–68] No single method has demonstrated systematic superiority over the other methods. The safe conclusion is that calibration method superiority is application dependent.[69,70] When the underlying type of nonlinearity implicit in the calibration method matches the latent nonlinearity in the data, the method will optimally model the data. This assertion has been supported by the improvement in calibration performance when theoretical instrument response functions replace the sigmoidal transfer function in ANN calibration.[71]

Nonlinear methods are much more prone to "overfitting" the calibration model. Overfitting occurs when the calibration begins to employ random variance (instrumental errors) for determining calibration parameters. The flexibility of the nonlinear models and the relatively large number of parameters that need to be estimated are the primary causes of this phenomenon. Consequently, the more complicated the model, the more prone the method is to overfitting (e.g. ANN vs PCR). A decision tree based on Occums Razor has been proposed to aid chemists in choosing among the nonlinear methods.[69] Linear and nonlinear calibrations were linked in a hierarchical web. The hierarchy is based on nested models and degrees of freedom required to calculate the model. Simple, linear models are at the top of the hierarchy and complex, nonlinear methods are at the bottom. It is recommended that to guard against overfitting and spurious modeling of the data, the method nearest the top of the hierarchy that provides sufficient calibration reliability for the application be employed. That is, use the simplest model that works.

3.5 Figures of Merit

The figures of merit for multivariate calibration are analogous to those for univariate calibration, with two notable exceptions. First, multivariate figures of merit are based on the ILS calibration model where univariate figures of merit are derived from the CLS calibration model. Therefore, the relationship between the model and the figures of merit may seem inverted when comparing univariate figures of merit with multivariate figures of merit. Second, multivariate figures of merit are based on the NAS compared with univariate figures of merit which are based on the total analyte signal.

The NAS, as defined in section 1.4, is the part of the analyte signal that is used for calibration. This is the analyte signal that is orthogonal to the signal from all other constituents. Mathematically, the NAS can be calculated by projecting the analyte signal on to the space defined by the other constituent signals (Equation 34):

$$\mathrm{NAS} = (\mathbf{I} - \mathbf{R}_{-\mathrm{a}}\mathbf{R}^{+}_{-\mathrm{a}})\mathbf{r}_{\mathrm{a}} \tag{34}$$

where $\mathbf{I}$ is the appropriately sized identity matrix, $\mathbf{R}_{-\mathrm{a}}$ is a matrix of spectra without the analyte, and $\mathbf{r}_{\mathrm{a}}$ is the

spectrum of the analyte of interest. The NAS is related to the regression vector, **b**, by Equation (35):

$$\mathbf{b} = \frac{\text{NAS}}{\|\text{NAS}\|_2^2} \tag{35}$$

with $\| \ \|_2$ being the Euclidian norm, or 2-norm, of a vector. Consequently, multivariate figures of merit can be expressed in terms of either the NAS or **b**. It is important to note that the NAS changes with the constituents of the calibration set. If more constituents are added, or constituents with increasing signal correlation to the analyte are substituted, the magnitude of the NAS will decrease.

In order to determine the figures of merit, it is unnecessary to determine the NAS. In fact, only the Euclidean length of the NAS is needed. This value can be determined from the regression vector by Equation (36):

$$\|\text{NAS}\|_2 = \frac{1}{\|\mathbf{b}\|_2} \tag{36}$$

The figures of merit can be defined either in terms of unit concentration[6,72] or in terms of arbitrary concentration.[73] Here, for the sake of simplicity, they will be defined in terms of unit analyte concentration.

The SEN of a multivariate method, at unit concentration, is the norm of the NAS (Equation 37):

$$\text{SEN} = \|\text{NAS}\|_2 = \frac{1}{\|\mathbf{b}\|_2} \tag{37}$$

because only the NAS is employed for estimation. Similarly, the effective S/N for multivariate calibration (at unit concentration) is given by Equation (38):

$$\text{S/N} = \frac{\|\text{NAS}\|_2}{\|\mathbf{e}\|_2} = \frac{1}{\|\mathbf{e}\|_2\|\mathbf{b}\|_2} \tag{38}$$

where **e** is a vector of independent, identically distributed measurement errors associated with $\mathbf{r}_a$. If three standard deviations of the instrumental blank are recognized as the smallest detectable signal, the IUPAC LOD is calculated as (Equation 39)

$$\text{LOD} = \frac{3\|\mathbf{e}\|_2}{\|\text{NAS}\|_2} = 3\|\mathbf{e}\|_2\|\mathbf{b}\|_2 \tag{39}$$

Actually, these definitions of S/N and LOD are rather optimistic because most real errors increase with **r**. Therefore, for mixtures $\|\mathbf{e}\|_2$ will be larger than would be predicted from the magnitude of $\mathbf{r}_a$ alone.

SEL is defined in univariate calibration as the fraction of the sample signal that comes from the analyte. This value ranges from zero to one. In multivariate calibration, SEL is defined as the fraction of the analyte signal that is useful for calibration. Because the NAS is the part of the useful analyte signal, a ratio of the 2-norm vector yields the SEL (Equation 40):

$$\text{SEL} = \frac{\|\text{NAS}\|_2}{\|\mathbf{r}_a\|_2} = \frac{1}{\|\mathbf{b}\|_2\|\mathbf{r}_a\|_2} \tag{40}$$

The multivariate SEL also ranges from zero to one. A SEL of 1 implies that the analyte response is completely orthogonal to all recorded interfering signals. A SEL of zero implies that the analyte's response is completely correlated with the other constituents and reliable calibration is impossible. Reliable calibration may be accomplished with a SEL greater than zero; however, the larger the SEL, the more stable is the calibration. However, for applications where no true analyte spectrum exists, e.g. octane number, SEL cannot be defined.

4 MULTIWAY ANALYSIS

4.1 History and Introduction

Multiway analysis became popular in the late 1970s in the psychometric literature. These psychologists employed the multiway models primarily for FA – to determine intrinsic factors in large, complex data sets. However, as chemical instrumentation advanced with automated data collection, chemists began to possess large, multiway data collections. In 1980 Hirschfeld listed 66 instruments capable of generating multiway data.[74] Geladi cataloged the manners in which multiway data can be collected in chemical applications.[9]

It is difficult to discern which multiway methods are ''factor analysis'' methods and which methods are ''calibration'' methods because FA is implicit in the multiway calibration philosophy. In this work, the distinction between FA and calibration is made based on the primary applications of each method. Those methods that are commonly employed when factor analytic information is desired are designated FA methods (section 4.2). The methods that are often employed for calibration are detailed in section 4.3. However, it should be stressed that any of the calibration methods can be employed for FA when only CR information is desired.

The underlying model for multiway FA is the trilinear model. Many methods relax the constraints of the model or expand the types of interactions among the factors; however, the trilinear model still provides the foundation of these methods. The trilinear model is a direct extension of the bilinear model employed in multivariate FA. This model can be presented equivalently in statistical form

(Equation 41):

$$\hat{\mathbf{R}}_{ijk} = \sum_{n=1}^{N} \hat{X}_{ni}\hat{Y}_{nj}\hat{Z}_{nk} + E_{ijk} \tag{41}$$

or mathematical form (Equation 42):

$$\hat{\mathbf{R}}_k = \sum_{n=1}^{N} \hat{\mathbf{x}}_n \hat{Z}_{nk} \hat{\mathbf{y}}_n^t + \mathbf{E}_k \tag{42}$$

where N refers to the number of factors employed by the model to describe the $I \times J \times K$ data cube, or the rank of the model.

In general, the number and form of factors are not constrained to be representative of any physical reality. With two-way FA, PCA, this is often referred to as the rotational ambiguity of the factors; there is a continuum of factors that satisfy the PCA model and equivalently describe the data.[26] However, with three-way analysis, when, for an isolated chemical component (1) the true underlying factor in each of the three modes is independent, except for scale, from the state of the other two modes, (2) the true underlying factor in any of the three modes cannot be expressed by linear combinations of the true underlying factors of other components in the same mode, (3) there is linear additivity of instrumental responses among the species present, and (4) the proper number of factors is chosen for the model, then the factors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are unique and are accurate estimates of the true underling factors, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ (except for a scaling constant). This is shown graphically in Figure 8, where the data set $\underline{\mathbf{R}}$ is, in reality, composed of N triads. If the proper number of factors N is chosen, the estimates of the factors in each of the triads, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$, will be accurate estimates of the true underlying factors. In other words, if $\underline{\mathbf{R}}$ were formed by a liquid chromatography (LC)/UV/VIS DAS (diode array spectrometer), each $\hat{\mathbf{x}}_n$ would correspond to one of the true N chromatographic profiles, each $\hat{\mathbf{y}}_n$ to one of the true spectroscopic profiles, and each $\hat{\mathbf{z}}_n$ to the relative concentrations in the K samples.

4.2 Multiway Factor Analysis

4.2.1 Simple Multiway Principal Component Analysis

A simple manner to attempt multiway principal component analysis (MPCA) is a direct extension of multivariate PCA.[75,76] In MPCA, the data cube $\mathbf{R}$ is unfolded sequentially in each of the three ''ways''. This creates three matrices, $\mathbf{R}_x$, $\mathbf{R}_y$, and $\mathbf{R}_z$, where each new matrix has one ''way'' in common from the original cube. For an $I \times J \times K$ tensor $\underline{\mathbf{R}}$, $\mathbf{R}_x$ is an $I \times (JK)$ matrix, $\mathbf{R}_y$ is a $J \times (IK)$ matrix, and $\mathbf{R}_z$ is a $K \times (IJ)$ matrix. PCA is performed on each of the three matrices. The scores from each unfolding are interpreted as the scores in each way of the cube $\underline{\mathbf{R}}$. This philosophy is easily extended to four-way or greater data tensors.

Although this method is fast and simple, it does not create a true three-way model of the data. That is, the data cube $\underline{\mathbf{R}}$ cannot be recreated from the collection of scores. The three sets of scores must be interpreted individually. Often these interpretations lead to contradictory conclusions.

4.2.2 Multiway Curve Resolution

Multiway curve resolution (MCR) methods are three-way (or greater) generalizations of the bilinear CR model presented in section 1.2.3.3.[31,32] The same ALS algorithms drives MCR as drives CR. The least-squares algorithm cycles through each of the N ways, successively updating the estimated parameters and applying any constraints assumed on the model. Nonnegativity, unimodality, unique measurements, and closure are commonly assumed to help derive physically meaningful parameters to the model. When the additional constraint

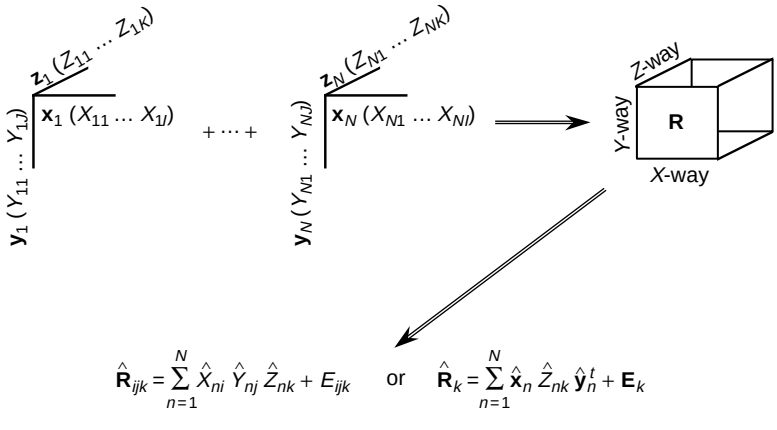

Figure 8 Construction and decomposition of a three-way array via the trilinear parallel factor analysis (PARAFAC) model.

of "trilinearity" is imposed, MCR becomes an alternative route for solving Equations (41) and (42).

It should be noted that the rigorous model of Equations (41) and (42) is not always appropriate for chemical applications. Kinetic, titration, flow injection, and chromatographic data seldom follow the strict trilinear model. For this reason, MCR allows for only the proper constraints being applied to the model.

4.2.3 Multiway Principal Components Analysis (Tucker Models)

The appropriate model for MPCA is the Tucker3 (T3) model.[76–79] The trilinear model of Equations (41) and (42) is a specific case of the T3 model. The T3 model is best understood by viewing a graphical representation such as in Figure 9. A data cube, $\underline{\mathbf{R}}$, is decomposed into three sets of factors, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$, as with the trilinear decomposition of Equations (41) and (42). However, the T3 model differs from the trilinear model in two key ways. The numbers of factors in each order of the T3 model are not constrained to be equal. Also, the T3 model employs a small core cube, $\underline{\mathbf{C}}$, that governs the interactions among the factors. A nonzero element in the pth, qth, and rth position of the core $\underline{\mathbf{C}}$ dictates an interaction between the pth factor in the X-way, the qth factor in the Y-way, and the rth factor in the Z-way. This permits the modeling of two or more factors that might have, for example, the same chromatographic profile but different spectral and concentration profiles.[80,81] If there are the same number of factors in each way, and $\underline{\mathbf{C}}$ is constrained to have only nonzero elements on the super diagonal, then the T3 model is equivalent to the trilinear model. There are other Tucker models; for example, the Tucker2 model is the two-way analog of the T3 model.

The ALS algorithm for the T3 model (without constraints) yields a least-squares model of $\underline{\mathbf{R}}$.[82] Like the PCA model the MPCA model is not unique. An infinite number of equivalent rotations of the factors will fit the data.

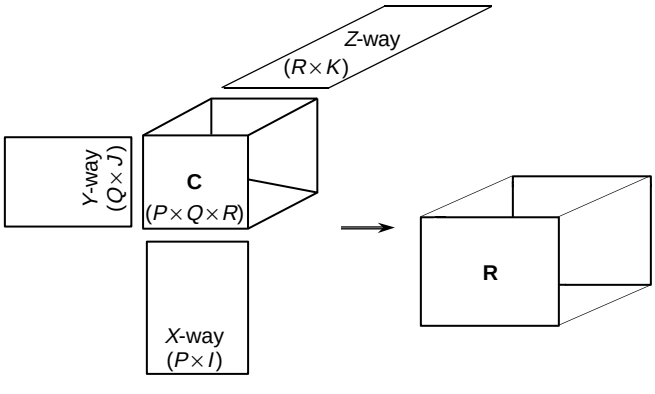

Figure 9 Construction of a three-way array according to the unconstrained Tucker model.

4.3 Calibration and Prediction

One common method for employing multiway analysis for calibration and prediction is parallel factor analysis (PARAFAC). PARAFAC refers both to the parallel factorization of the data set $\underline{\mathbf{R}}$ by Equations (41) and (42) and also to an ALS algorithm for determining $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, and $\hat{\mathbf{Z}}$ in the two equations. The ALS algorithm begins with an initial guess of the X-way and Y-way starting profiles. The initial Z-way profiles are determined by solving Equation (43):

$$\mathbf{R}_{\mathrm{C}} = \mathbf{A}\mathbf{Z}^{\mathrm{T}} \tag{43}$$

such that $\hat{\mathbf{Z}} = \mathbf{C}^{+}\mathbf{R}_{\mathrm{C}}$ with $\mathbf{C}^{+}$ being the generalized inverse of $\mathbf{C}$ that can be calculated from the normal equations or SVD of $\mathbf{C}$. In Equation (43), $\mathbf{R}_{\mathrm{C}}$ is an $I * J \times K$ matrix constructed by unfolding the K slices of $\underline{\mathbf{R}}$ in the IJ plane where $R_{C_{(j-1)I+i,k}} = R_{i,j,k}$. Similarly, $\mathbf{C}$ is an $I * J \times N$ matrix formed from the N columns of $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ where $C_{(j-1)I+i,n} = X_{i,n}Y_{j,n}$.

Updated estimates of the X-way and Y-way profiles are found by solving Equation (44):

$$\mathbf{R}_{\mathrm{A}} = \mathbf{A}\mathbf{X}^{\mathrm{T}} \tag{44}$$

such that $\hat{\mathbf{X}} = \mathbf{A}^{+}\mathbf{R}_{\mathrm{A}}$, and (Equation 45):

$$\mathbf{R}_{\mathrm{B}} = \mathbf{B}\mathbf{Y}^{\mathrm{T}} \tag{45}$$

such that $\hat{\mathbf{Y}} = \mathbf{B}^{+}\mathbf{R}_{\mathrm{B}}$. $\mathbf{R}_{\mathrm{A}}$ and $\mathbf{R}_{\mathrm{B}}$ are constructed analogously to $\mathbf{R}_{\mathrm{C}}$ by unfolding $\underline{\mathbf{R}}$ in the YZ and XZ planes, respectively. This forms a $J * Z \times I$ matrix for $\mathbf{R}_{\mathrm{A}}$ and an $X * Z \times J$ matrix for $\mathbf{R}_{\mathrm{B}}$. Similarly to $\mathbf{C}$, $A_{(k-1)J+k,n} = Y_{j,n}Z_{k,n}$ and $B_{(k-1)I+k,n} = X_{i,n}Z_{k,n}$. The algorithm proceeds iteratively, cycling through Equations (43), (44), and (45) until the convergence criterion is satisfied. At each step the most recent estimates of $\mathbf{X}$ and $\mathbf{Y}$ are used to determine $\hat{\mathbf{Z}}$ (or $\mathbf{Y}$ and $\mathbf{Z}$ to determine $\hat{\mathbf{X}}$, or $\mathbf{X}$ and $\mathbf{Z}$ to determine $\hat{\mathbf{Y}}$, depending on the equation currently being solved). There are, consequently, two important factors that influence the final estimates of the X-way, Y-way, and Z-way profiles: the starting guess for $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$, and the convergence criterion.

The PARAFAC algorithm is sensitive to the starting guess of the solution for $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$. This results from PARAFAC often becoming trapped in local minima and, hence, not converging to the global optimum least-squares solution. Furthermore, the PARAFAC algorithm can become delayed in "swamps" far from the optimum solution.[83] Consequently, the speed of the algorithm is sensitive to the initial guess for $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$. The starting

iteration of $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$ can be provided by a random number generator,[84] direct trilinear decomposition (DTLD) method,[85] or a priori knowledge of analyte profiles. Although this markedly increases the analysis time, when employing a random starting value, multiple initial guesses should be considered. The solution for each starting value will be different; however, if all (or most) of the solutions are similar, it is safe to assume that PARAFAC has converged to near the global optimum solution. The convergence time for PARAFAC can be improved by initializing the algorithm with guesses near the optimum solution. These guesses can come from DTLD or reference spectra of species known (or highly suspected) to be in the data set. Care should be employed when utilizing the DTLD solutions since DTLD often yields significant imaginary components in predicting X-way and Y-way factors. The problems caused by initializing PARAFAC with imaginary components can be circumvented by employing the real component of $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$ from DTLD or the absolute value (complex modulus) of $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$ from DTLD.

Two popular convergence criteria for the PARAFAC algorithm are based on changes in the residuals (unmodeled data) between successive iterations and changes in the predicted profiles between successive iterations. In the first case, the algorithm is terminated when the root averages of the squared residuals between successive iterations agree to within an absolute or relative tolerance, say 10^{-6}. Although such fit-based stopping criteria are conceptually easy to visualize, a faster method for determining convergence relies on the correlation between the predicted X-, Y-, and Z-way profiles between successive iterations. When the product of the cosines between successive iterations in the X, Y, and Z modes approaches arbitrarily close to 1, say within 10^{-6}, the algorithm is terminated. The cosine in the X-way is determined by unfolding the $I \times N$ matrices $\widehat{\mathbf{X}}_{\text{old}}$ and $\widehat{\mathbf{X}}_{\text{new}}$ into column vectors $\hat{\mathbf{x}}_{\text{old}}$ and $\hat{\mathbf{x}}_{\text{new}}$. $\cos\theta_x$ is defined according to Equation (46):

$$\cos\theta_X = \frac{\mathbf{x}_{\text{old}}\mathbf{x}_{\text{new}}}{\sqrt{(\mathbf{x}_{\text{old}}\mathbf{x}_{\text{old}})(\mathbf{x}_{\text{new}}\mathbf{x}_{\text{new}})}} \tag{46}$$

The other two terms, $\cos\theta_Y$ and $\cos\theta_Z$, are defined equivalently. Convergence, when $\cos\theta_X \times \cos\theta_Y \times \cos\theta_Z > 1-10^{-6}$, implies that successive iterations in all three modes are correlated to at least $1-10^{-6}$. Mitchell and Burdick cite, in addition to speed, an additional benefit to correlation based convergence.[83] In cases when two factors are highly correlated in one or more of the three ways, ALS methods may become mired in ''swamps'' where the fit of the model changes slightly but the correlation between the predicted X-, Y-, and Z-ways changes significantly between successive iterations. After many iterations the ALS algorithm will then emerge from the ''swamp'' and the residuals and estimated profiles will then both rapidly approach the optimum. Hence correlation based convergence is more resistant to inflection points in the error response surface when optimizing the model.

4.4 Figures of Merit

In three-way calibration, as with two-way calibration, the figures of merit are similarly derived from the three-way NAS.[6,71] Assuming all calculations are performed at unit analyte concentration, the SEL, SEN, and S/N are the magnitude of the NAS divided by the magnitude of the analyte signal, concentration, and noise, respectively. Mathematically, they can be found from Equations (47–49):

$$\text{SEL} = \frac{\|\text{NAS}\|_{\text{F}}}{\|\mathbf{R}_{\text{A}}\|_{\text{F}}} \tag{47}$$

$$\text{SEN} = \frac{\|\text{NAS}\|_{\text{F}}}{c} \tag{48}$$

$$\text{S/N} = \frac{\|\text{NAS}\|_{\text{F}}}{\|\mathbf{E}\|_{\text{F}}} \tag{49}$$

where $\mathbf{R}_{\text{A}}$ is the response of the analyte at unit concentration, c, $\mathbf{E}$ is a matrix of expected, or estimated, errors, and $\|\|_{\text{F}}$ is the Froebus norm (root sum of squared elements) of a matrix. It should be noted that whereas the NAS is a matrix quantity, SEL, SEN, and S/N are all vector quantities. The LOD and limit of quantitation can also be determined via any accepted univariate definition by replacing $\|\text{NAS}\|_{\text{F}}$ for the analyte signal and $\|\mathbf{E}\|_{\text{F}}$ for the error value.

However, there is still debate over the proper manner to calculate the NAS.[73] In the earliest work by Ho et al.,[86] the three-way NAS is calculated as the outer product of the multivariate NAS from the resolved X-way and Y-way profiles. That is (Equation 50)

$$\mathbf{x}_{\text{NAS}} = \mathbf{x}_{\text{a}}^{\text{T}}(\mathbf{I} - \mathbf{X}_{\text{i}}\mathbf{X}_{\text{i}}^{+}) \tag{50}$$

and (Equation 51)

$$\mathbf{y}_{\text{NAS}} = \mathbf{y}_{\text{a}}^{\text{T}}(\mathbf{I} - \mathbf{Y}_{\text{i}}\mathbf{Y}_{\text{i}}^{+}) \tag{51}$$

so (Equation 52)

$$\text{NAS} = \mathbf{x}\mathbf{y}^{\text{T}} \tag{52}$$

Similarly, Messick et al. suggested the NAS can be found by orthogonal projection of $\mathbf{R}$ following unfolding each $I \times J$ sample and interferent matrix into an $IJ \times 1$ vector. The three-way NAS is the consequent NAS of Equation (34) refolded into an $I \times J$ matrix.[87] The third alternative, propounded by Wang et al., is to construct the NAS from the outer products of the X-way and Y-way

profiles that are unique to the analyte; in this method no projections are explicitly calculated.[88]

4.5 Batch Analysis

Multiway methods, in particular MPCA, can be employed to model batch processes.[89,90] Each historical run of a batch is treated as a ''way'' in the data matrix. This yields a batch × time × sensors matrix of data. Future runs are projected on to the factors obtained from three-way PCA. This yields a condensed data representation of the process variation that can be employed to construct SPC charts. The scores from analysis of the historical data provide confidence limits and control limits. It should be noted that unlike PARAFAC analysis for calibration of multiway sensors, there are no true underlying process ''spectra'' to be determined. The derived scores and loading should be interpreted as correlations (see PCA, section 1.2.1).

5 PRETREATMENT OF DATA

The success of multivariate data analysis often depends on the application of data pretreatment to remove, scale, or standardize the sources of observed variance.

5.1 Centering and Scaling

''Mean centering'' and ''variance scaling'' are often performed on multivariate data without much thought as to the consequences of these actions. Mean centering removes the average, or mean, response of a given variable or sample. This translates the variance of the data set to be centered about the ordinate axis. Variance scaling normalizes each variable, or sample, such that the data's variance becomes unity. This places the data on a unit sphere. When mean centering and variance scaling are both applied to a collection of data, the data are said to be ''autoscaled''. Autoscaling places the data on a unit sphere centered about the origin of the multivariate space of the data.

There are specific instances when mean centering and variance scaling should and should not be applied to a data set. In general, mean centering aids in interpretation of the FA models and construction of calibrations. By removing the mean of the data set, often one less factor is required for analysis. An exception may occur when the data are collected under ''closure''.[91,92] Closure exists when the sum of the variables or concentrations is constrained to equal a preset value. The most common type of closure is seen in mixture analysis when the sum of percentage composition of all detectable species is constrained to equal 100%. Other examples may occur when improper experimental designs are employed. When closure exists, mean centering will not always eliminate a factor. In these instances the errors introduced by estimating the mean of the data set are not offset by the gains associated with a more simple model.

When a data set is variance scaled all variables, or samples, are given equal weight in determining the factors of the model. This may be beneficial when variables with small variance have greater predictive variance than variables with larger variance. A prime example is seen in fusing data measurements with drastically different scales (e.g. physical measurements such as temperature and pressure with spectroscopic data). However, in most spectroscopic, chromatographic, or electrochemical analyses, the measurement is chosen so as to be most sensitive to the analyte of interest. Here, it would not be favorable to give equal weight to background noise in uninformative measurements as is given to measurements with maximum analyte SEN.

A third type of scaling often employed is scaling each variable or sample to unit area. This scaling is successfully applied to samples when matrix or sampling effects alter the measurement efficiency of a method. Examples include sample to sample variance due to sample thickness in NIR reflectance effective path length in other optical methods. Unit area normalization obscures the absolute concentrations of analytes but preserves the relative concentration of constituents between and among samples. Therefore, absolute calibration cannot be performed unless the calibration set is constrained by closure once the data are normalized. Unit area scaling is employed in Lawton and Sylvestre CR (section 3.2.3.2).[28]

5.2 Background Correction

Background correction methods are often employed in spectroscopic applications to remove broad features from the data set. These features hinder calibration as a large source of variance compared with the analyte or as a seemingly random source of variance that consumes many factors in the model. Examples include fluorescence background in Raman spectroscopy and scattering backgrounds in NIR reflectance spectroscopy.

Simple efforts at background correction include derivatives, polynomial curve fitting, and Fourier transform (FT) filtering.[93] Derivatives remove the portion of a background that can be modeled by a low-order polynomial. Taking the first derivative of a spectrum removes the baseline offset. The second derivative removes the linear approximation of the background (and the analyte signal). However, in spite of digital filters for simultaneously smoothing the data while calculating the derivatives,[94]

the S/N rapidly declines with each derivatization. Polynomial curve fitting is useful when there are regions of the spectra that contain only background variance. These regions must be distributed across the entire spectrum such that the background can be modeled. FT filtering removes both low- and high-frequency variance across the spectrum. It is assumed that the lowest frequency signal is the background and the highest frequency signal is random instrumental error. Problems may occur with FT filtering due to poorly chosen apodization functions applied to the signal or insufficient ability to distinguish between the signal and the background. This will lead to distortion of the analyte signal.

Multivariate scattering correction (MSC) has been applied with great success to NIR reflectance and transmittance data.[93,95–97] MSC accounts for the fact that the scattering efficiency of light is a function of wavelength. MSC estimates the amount of scattering in each sample relative to a reference sample. A simple linear model is employed at each wavelength where the scattering background is assumed to have the form (Equation 53)

$$\mathbf{x}_\lambda = a + b\bar{\mathbf{x}}_\lambda + \mathbf{e}_\lambda \tag{53}$$

where $\mathbf{x}_\lambda$ is the NIR measurements for a set of samples at wavelength λ and $\bar{\mathbf{x}}_\lambda$ is the NIR spectrum of a reference sample at wavelength λ. The mean spectrum of the calibration set serves as an excellent reference sample. The additive and multiplicative constants a and b are determined by ordinary least-squares regression. Once these two constants are known, the scatter corrected spectra are found in scaled residuals (Equation 54):

$$\mathbf{x}_\lambda^{\text{corrected}} = \frac{\mathbf{x}_\lambda - \hat{a}}{\hat{b}} = \frac{\hat{\mathbf{e}}_\lambda}{\hat{b}} \tag{54}$$

5.3 Instrument Standardization

One practical concern with multivariate calibration and prediction is the transport and stability of the calibration models. Ideally, a calibration model can be constructed in the laboratory on a benchtop instrument, then the model can be applied to many similar instruments in the field. Also, once a model has been successfully transported to the field, it will be robust to changes in instrumental SEN and alignment. Of course, the goal of a universally transportable and robust instrument/model has not been achieved. Seemingly identical spectrometers have slight wavelength resolution and SEN differences that can prohibit reliable distribution of the calibration model among numerous instruments. Also, time-dependent instrumental drift eventually can render the calibration model obsolete for whichever instrument the model was constructed.

Individual calibration of each instrument is not an acceptable solution to the problem of model distribution. Calibration may be an expensive, time-consuming task when many calibration samples are needed, the calibration samples are not readily transportable, or the instrument is not easily accessible in the process stream. Concurrently, it is also unacceptable to repeat an entire calibration procedure whenever there are minor changes in the instrumental character.

Instrumental standardization strives to solve the problems derived from instrumental differences when constructing one calibration model for multiple instruments. The instrumental standardization philosophy is to construct the best model possible on one instrument then to build a second model that will transform the spectra from other instruments to appear as if they were recorded on the first instrument. Usually, this transfer function can be reliably calculated with less effort.

One standardization method popular in the literature is piecewise direct standardization (PDS).[98–101] With PDS, a set of transfer samples are analyzed on both the original instrument and the instrument to which the calibration model will be transferred. It is best if the transfer samples are a subset of the calibration set; however, other surrogate samples may be employed. A separate transfer function is determined for each wavelength in the spectra by least-squares regression using neighboring wavelengths as the independent variables. That is, a local subset of variables measured on the second instrument is employed to build a model that predicts what each measurement would have been if it were measured with the first instrument. This method accounts for shifts and intensity changes over a small spectral window. The drawback of PDS is that success of the standardization is dependent on choice of the transfer samples. The transfer samples must be identical when measured on each instrument and the set of samples must span the space of all encountered spectral changes between the two instruments. Therefore, the choice and number of transfer samples must be optimized by the analyst.

A more useful method of standardization would not require transfer samples to be analyzed. There have been two approaches to this problem. When it can be safely assumed that the only spectral shifts (e.g. wavelength or retention time) occur, a PCA-based method of standardization may be employed.[102,103] The spectral (or time) indices are shifted such that the projection of each sample into the PC space defined by the original instrument is optimized. A more general method based loosely on MSC has also demonstrated success when there are relatively minor performance differences between the original and second instruments.[104,105] Here a local selection of wavelengths from each spectrum are regressed against the mean spectrum to build a

transfer function. Consequently, the spectra from the second instrument are not transformed to look like the spectra from the first instrument. Instead, spectra from both instruments are transformed to lie in a common multidimensional space.

6 MODEL VALIDATION (RIGHT VERSUS WRONG MODEL)

One topic that has been alluded to in this article, but not adequately discussed, is the validation of the model. Not only must a decision be made in choosing between linear and nonlinear calibration models, but also the number of factors in the FA and calibration models must be determined. If the chosen model is too simple it will not describe properly the variance in the data set and a systematic bias will become imbedded in the qualitative and quantitative interpretation of the model. If the chosen model is too complex, random noise imbedded in the model will be interpreted as chemically meaningful information. The precision of calibration and estimations with the model will be significantly degraded.

The best method for testing the adequacy of the model is to employ three separate sets of data: a calibration set, a test set, and a validation set. It is assumed that the three sets of data span the same measurement (and concentration) space. All of the models to be constructed are determined from the calibration set. The ability of these models to fit the data and/or estimate the property of interest is determined with the test set. The combination of model and number of factors in the model that performs best on the test set is designated as the right model. Note that by the parsimony principle, if a number of models perform equivalently, the simplest model should be employed.[70] Once the right model has been selected, it is further tested on the validation set. This allows for accurate estimation of the future performance parameters of the model.

The three-data-set method for model construction is designed to guard against ''overfitting'' of the model to the data. Adding more factors to a model almost always increases the fit of a model to the data set employed to calculate the model parameters. Because the test and validation sets have different realizations of instrumental errors to the calibration set, parameters that model the noise in the calibration set will not precisely model the noise in the other two sets. Estimation of future performance of the model based on the test set yields an overly optimistic estimation of the model's abilities. After all, that particular model was chosen because it fitted the test set well.

Often it is impractical to collect sufficient data to designate three separate sets for model construction. If only two data sets are employed, the performance of the model on the test set can be estimated by cross-validation (CV).[25] At a pinch, the test set can be eliminated also. However, without a test set it is dangerous to extrapolate the future performance of the model from the calibration set.

CV is a jack-knife method where a number of samples, N, are sequentially left out of the I sample calibration set. A model is constructed with the remaining $I - N$ samples and the fit/predictive ability of the model is tested on the N excluded samples. The N samples are replaced in the calibration set and this procedure is repeated until all of the I samples have been excluded in groups of N. The ensemble performance of the models on the excluded data is interpreted as the performance of the model on a true test set would be interpreted.

ABBREVIATIONS AND ACRONYMS

ACE	Alternating Conditional Expectations
ACS	American Chemical Society
ALS	Alternating Least-squares
ANN	Artificial Neural Network
CLS	Classical Least-squares
CR	Curve Resolution
CV	Cross-validation
DAS	Diode Array Spectrometer
DTLD	Direct Trilinear Decomposition
EFA	Evolving Factor Analysis
FA	Factor Analysis
FT	Fourier Transform
GC/MS	Gas Chromatography/Mass Spectrometry
GLT	Global Linearizing Transformations
HPLC	High-performance Liquid Chromatography
ILS	Inverse Least-squares
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LOD	Limit of Detection
LWR	Locally Weighted Regression
MARS	Multivariate Adaptive Regression Splines
MCR	Multiway Curve Resolution
MLR	Multilinear Regression
MPCA	Multiway Principal Component Analysis
MSC	Multivariate Scattering Correction
MSPC	Multivariate Statistical Process Control
NAS	Net Analyte Signal

NIR	Near-infrared
NPLS	Nonlinear Partial Least Squares
PARAFAC	Parallel Factor Analysis
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDS	Piecewise Direct Standardization
PLSR	Partial Least-squares Regression
PP	Projection Pursuit
SEL	Selectivity
SEN	Sensitivity
S/N	Signal-to-noise Ratio
SPC	Statistical Process Control
SVD	Singular Value Decomposition
TFA	Target Factor Analysis
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis

Chemometrics **(Volume 11)**
Chemometrics • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher • Signal Processing in Analytical Chemistry • Soft Modeling of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. J.R. Workman, D.J. Veltkamp, S. Doherty, B.B. Anderson, K.E. Creasy, M. Koch, J.F. Tatera, A.L. Robinson, L. Bond, L.W. Burgess, G.N. Bokerman, A.H. Ullman, G.P. Darsey, F. Mozayeni, J.A. Bamberger, M.S. Greenwood, 'Process Analytical Chemistry', *Anal. Chem.*, **71**, 121R–180R (1999).
2. W.W. Blaser, R.A. Brederweg, R.S. Harner, M.A. LaPack, A. Leugers, D.P. Martin, R.J. Pell, J. Workman, L.G. Wright, 'Process Analytical Chemistry', *Anal. Chem.*, **67**, 47R–70R (1995).
3. B.M. Wise, B.R. Kowalski, 'Process Chemometrics', in *Process Analytical Chemistry*, eds. F. McLennaan, B.R. Kowalski, Blackie Academic, London, 259–312, 1995.
4. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988.
5. S. Wold, 'Chemometrics; What Do We Mean With It, and What Do We Want From It?', *Chemom. Intell. Lab. Syst.*, **30**, 109–115 (1995).
6. K.S. Booksh, B.R. Kowalski, 'Theory of Analytical Chemistry', *Anal. Chem.*, **66**, 782A–791A (1994).
7. S.D. Brown, 'Has the Chemometrics Revolution Ended? Some Views on the Past, Present and Future of Chemometrics', *Chemom. Intell. Lab. Syst.*, **30**, 49–58 (1995).
8. J.B. Cooper, 'Chemometric Analysis of Raman Spectroscopic Data for Process Control Applications', *Chemom. Intell. Lab. Syst.*, **46**, 231–247 (1999).
9. P. Geladi, 'Analysis of Multiway (Multi-mode) Data', *Chemom. Intell. Lab. Syst.*, **7**, 11–30 (1989).
10. G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd edition, Johns Hopkins University Press, Baltimore, MD, 1996.
11. O.L. Davies, *Statistical Methods in Research and Production with Special Reference to the Chemical Industry*, 3rd edition, Oliver and Boyd, London, 1958.
12. W.L. Hays, *Statistics*, 4th edition, Harcourt Brace Jovanovich, Fort Worth, TX, 1988.
13. M.R. Spiegel, *Schaum's Outline Series: Theory and Problems of Statistics*, 2nd edition, McGraw-Hill, New York, 1990.
14. G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters: an Introduction to Design, Data Analysis, and Model Building*, Wiley, New York, 1978.
15. H. Mark, J. Workman, *Statistics in Spectroscopy*, Academic Press, Boston, 1991.
16. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi, J. Smeyers, *Handbook of Chemometrics and Qualimetrics*, Elsevier, Amsterdam, 1997.
17. N.R. Draper, H. Smith, *Applied Regression Analysis*, Wiley, New York, 1966.
18. S.N. Deming, S.L. Morgan, *Experimental Design: a Chemometric Approach*, Elsevier, Amsterdam, 1993.
19. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vettering, *Numerical Recipies in Pascal: the Art of Scientific Computing*, Cambridge University Press, Cambridge, 1989.
20. F.H. Walters, L.R. Parker, S.L. Morgan, S.N. Deming, *Sequential Simplex Optimization*, CRC Press, Boca Raton, FL, 1991.
21. J.K. Taylor, 'Measurement and Calibration', *ChemTech*, **16**, 763 (1986).
22. G.L. Long, J.D. Winefordner, 'Limit of Detection. A Closer look at the IUPAC Definition', *Anal. Chem.*, **55**, 712A–723A (1983).
23. J.B. Callis, D.L. Illman, B.R. Kowalski, 'Process Analytical Chemistry', *Anal. Chem.*, **55**, 624A–636A (1987).

24. D.C. Hassell, E.M. Bowman, 'Process Analytical Chemistry for Spectroscopists', *Appl. Spectrosc.*, **52**, 18A–29A (1998).
25. E. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, Wiley, New York, 1991.
26. I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.
27. F. Galton, *J. Anthropol. Inst.*, **15**, 246–263 (1885).
28. W.H. Lawton, E.A. Sylvestre, 'Self Modeling Curve Resolution', *Technometrics*, **13**, 617–633 (1971).
29. H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuhler, 'Calculation of Equilibrium Constants from Multiwavelength Spectroscopic Data. III: Model-free Analysis of Spectrophotometric and ESR Titrations', *Talanta*, **32**, 1133 (1985).
30. Y.-Z. Liang, R. Manne, O.M. Kvalheim, 'Constrained Background Bilinearization', *Chemom. Intell. Lab. Syst.*, **14**, 155–184 (1992).
31. R. Tauler, I. Marques, E. Casassas, 'Multivariate Curve Resolution Applied to Three-way Trilinear Data: Study of Spectrofluorometric Acid–Base Titration of Salicylic Acid at Three Excitation Wavelengths', *J. Chemom.*, **12**, 55–75 (1998).
32. J. Saurina, S. Hernandez-Cassou, R. Tauler, A. Izquierdo-Ridorsa, 'Multivariate Resolution of Rank-deficient Spectrophotometric Data from First Order Kinetic Decomposition Reactions', *J. Chemom.*, **12**, 183–203 (1998).
33. W. Windig, 'Self-modeling Mixture Analysis of Spectral Data with Continuous Concentration Profiles', *Chemom. Intell. Lab. Syst.*, **16**, 1–16 (1992).
34. W.C. Bell, K.S. Booksh, M.L. Myrick, 'Monitoring Anhydride and Acid Conversion Supercritical/Hydrothermal Water by In Situ Fiber-optic Raman Spectroscopy', *Anal. Chem.*, **70**, 332–339 (1998).
35. S.P. Gurden, R.G. Bereton, J.A. Groves, 'Resolution of Infrared Spectra of Mixtures of Potential Organic Workplace Pollutants by Factor Analysis', *Chemom. Intell. Lab. Syst.*, **23**, 123–135 (1994).
36. E.A. Sylvestre, W.H. Lawton, M.S. Maggio, 'Curve Resolution Using a Postulated Chemical Reaction', *Technometrics*, **16**, 353–368 (1974).
37. R.I. Shrager, 'Chemical Transitions Measured by Spectra and Resolved Using Singular Value Decomposition', *Chemom. Intell. Lab. Syst.*, **1**, 59–70 (1986).
38. O.S. Borgen, B.R. Kowalski, 'An Extension of the Multivariate Component Resolution Method to Three Components', *Anal. Chim. Acta*, **174**, 1–26 (1985).
39. A. Muller, D. Steele, 'On the Extraction of Spectra of Components from Spectra of Mixtures. A Development in Factor Theory', *Spectrochim. Acta, Part A*, **46**, 817–842 (1990).
40. B.M. Wise, N.L. Ricker, D.J. Veltkamp, B.R. Kowalski, 'A Theoretical Basis for the Use of Principal Components Models for Monitoring Multivariate Processes', *Process Control Qual.*, **1**, 41–51 (1990).
41. P. Nomikos, J.F. MacGregor, 'Multivariate SPC Charts for Batch Processes', *Technometrics*, **37**, 41–59 (1995).
42. C. Wikstrom, C. Albano, H. Eriksson, H. Friden, E. Johnsson, A. Nordahl, S. Rannar, M. Sandberg, N. Kettaneh-Wold, S. Wold, 'Multivariate Process and Quality Monitoring Applied to an Electrolysis Process', *Chemom. Intell. Lab. Syst.*, **42**, 221–232 (1998).
43. A. Lorber, 'Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations', *Anal. Chem.*, **58**, 1167–1172 (1986).
44. J.M. Sutter, J.M. Kalavias, P.M. Lang, 'Which Principal Components to Utilize for Principal Components Regression', *J. Chemom.*, **6**, 217–225 (1992).
45. R. Marbach, H.M. Heise, 'On the Efficiency of Algorithms for Multivariate Linear Calibration used in Analytical Spectroscopy', *Trends Anal. Chem.*, **11**, 270–275 (1992).
46. D.M. Haaland, E.V. Thomas, 'Partial Least-squares methods for Spectral Analysis. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information', *Anal. Chem.*, **60**, 1193–1202 (1988).
47. A. Lorber, L.E. Wangen, B.R. Kowalski, 'A Theoretical Foundation for the PLS Algorithm', *J. Chemom.*, **1**, 19–31 (1987).
48. P. Geladi, B.R. Kowalski, 'PLS Regression: a Tutorial', *Anal. Chim. Acta*, **185**, 1–17 (1986).
49. M. Stone, R.J. Brooks, 'Continuum Regression: Cross-validated Sequentially-constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares, and Principal Component Regression', *J. R. Stat. Soc. B*, **52**, 337–369 (1990).
50. S. de Jong, 'SIMPLS: An Alternative Approach to Partial Least Squares Regression', *Chemom. Intell. Lab. Syst.*, **18**, 251–263 (1993).
51. R. Manne, 'Analysis of Two PLS Algorithms for Multivariate Calibration', *Chemom. Intell. Lab. Syst.*, **2**, 187–197 (1987).
52. T. Naes, T. Isaksson, 'Locally Weighted Regression of Diffuse Near Infrared Transmittance Spectroscopy', *Appl. Spectrosc.*, **46**, 34–43 (1992).
53. K.S. Johnston, S.S. Lee, K.S. Booksh, 'Calibration of Surface Plasmon Resonance Refractometers Using Locally Weighted Parametric Regression', *Anal. Chem.*, **69**, 1844–1851 (1997).
54. W.S. Cleveland, S.J. Devlin, 'Locally Weighted Regression – an Approach to Regression Analysis by Local Fitting', *J. Am. Stat. Assoc.*, **83**, 596–610 (1988).
55. T. Naes, T. Isaksson, B.R. Kowalski, 'Locally Weighted Regression and Scatter Correction for Near-infrared Reflectance Data', *Anal. Chem.*, **62**, 664–673 (1990).
56. Z. Wang, T. Isaksson, B.R. Kowalski, 'New Approach for Distance Measurement in Locally Weighted Regression', *Anal. Chem.*, **66**, 249–260 (1994).

57. L. Brieman, J.H. Friedman, 'Estimating Optimal Transformations for Multiple Regression and Correlation', *J. Am. Stat. Assoc.*, **80**, 580–619 (1985).
58. I.E. Frank, S. Lanteri, 'ACE: a Non-linear Regression Model', *Chemom. Intell. Lab. Syst.*, **3**, 301–313 (1988).
59. J.H. Friedman, W. Steutzle, 'Projection Pursuit Regression', *J. Am. Stat. Assoc.*, **76**, 817–823 (1981).
60. M.M.C. Ferreira, W.C. Ferreira, B.R. Kowalski, 'Rank Determination and Analysis of Non-linear Processes by Global Linearizing Transformations', *J. Chemom.*, **10**, 11–30 (1996).
61. S. Winsberg, J.O. Ramsey, 'Monotone Spline Transformations for Dimension Reduction', *Psychometrika*, **48**, 575–593 (1984).
62. I.E. Frank, 'A Non-linear PLS Model', *Chemom. Intell. Lab. Syst.*, **8**, 109–119 (1990).
63. S. Wold, 'Nonlinear Partial Least Squares Modeling: II. Spline Inner Relation', *Chemom. Intell. Lab. Syst.*, **14**, 71–84 (1992).
64. J.H. Friedman, 'Multivariate Adaptive Regression Splines', *Ann. Stat.*, **19**, 199–216 (1991).
65. S. Sekulic, B.K. Kowalski, 'MARS: a Tutorial', *J. Chemom.*, **6**, 199–216 (1992).
66. S. Sekulic, M.B. Seasholtz, Z. Wang, B.R. Kowalski, S.E. Lee, B.R. Holt, 'Nonlinear Multivariate Calibration Methods in Analytical Chemistry', *Anal. Chem.*, **65**, 835A–844A (1993).
67. I.E. Frank, 'Modern Nonlinear Regression Methods', *Chemom. Intell. Lab. Syst.*, **27**, 1–9 (1995).
68. K.S. Booksh, B.R. Kowalski, 'Calibration Method Choice by Comparison of Model Basis Functions to the Theoretical Instrument Response Function', *Anal. Chim. Acta*, **348**, 1–9 (1997).
69. M. Gerritsen, J.A. van Leeuwen, B.G.M. Vandeginste, L. Buydens, G. Kateman, *Chemom. Intell. Lab. Syst.*, **15**, 171 (1992).
70. M.B. Seasholtz, B.R. Kowalski, 'The Parsimony Principal', *Anal. Chim. Acta*, **277**, 165–177 (1993).
71. Z. Wang, J.N. Hwang, B.R. Kowalski, 'CHEMNETS – Theory and Application', *Anal. Chem.*, **67**, 1497–1504 (1995).
72. K.S. Booksh, Z. Wang, 'Extension and Application of Univariate Figures of Merit to Multivariate Calibration', in *Chemometrics in Environmental Chemistry*, ed. J. Einax, Springer, Berlin, 209–228, 1995.
73. K. Faber, A. Lorber, B.R. Kowalski, 'Analytical Figures of Merit for Tensorial Calibration', *J. Chemom.*, **11**, 419–461 (1997).
74. T. Hirschfeld, 'The Hyphenated Methods', *Anal. Chem.*, **52**, 297A (1980).
75. S. Wold, P. Geladi, K. Esbensen, J. Ohman, 'Multi-way Principal Components and PLS Analysis', *J. Chemom.*, **1**, 41–56 (1987).
76. A.K. Smilde, D.A. Doornbos, 'Three-way Methods for the Calibration of Chromatographic Systems: Comparing PARAFAC and Three-way PLS'. *J. Chemom.*, **5**, 345–360 (1991).
77. P.M. Kroonenberg, *Three-mode Principal Component Analyses. Theory and Applications*, DSWO Press, Leiden, 1983.
78. R.A. Harshman, 'Foundations of the PARAFAC Procedure', UCLA Working Paper on Phonetics, No. 16, 1–84, 1970.
79. R. Henrion, G. Henrion, G.C. Onuoha, 'Multi-way Principal Components Analysis of a Complex Data Array Resulting from Physicochemical Characterization of Natural Waters', *Chemom. Intell. Lab. Syst.*, **16**, 87–95 (1992).
80. A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, 'Multicomponent Determination of Chlorinated Hydrocarbons Using a Reaction Based Sensor. 3. Medium-rank Second-order Calibration with Restricted Tucker Models', *Anal. Chem.*, **66**, 3345–3351 (1994).
81. A.K. Smilde, Y. Wang, B.R. Kowalski, 'Theory of Medium-rank Second-order Calibration with Restricted-Tucker Models', *J. Chemom.*, **8**, 21–36 (1994).
82. P.M. Kroonenberg, J. de Leeuw, 'Principal Component Analysis of Three-mode Data by Means of Alternating Least Squares Algorithm', *Psychometrics*, **45**, 69–97 (1980).
83. B.C. Mitchell, D.S. Burdick, 'Slowly Converging PARAFAC Sequences: Swamps and Two-factor Degeneracies', *J. Chemom.*, **6**, 155 (1992).
84. R.A. Harshman, M.E. Lundy, 'The PARAFAC Model for Three-way Factor Analysis and Multidimensional Scaling', in *Research Methods for Multimode Data Analysis*, eds. H.G. Law, C.W. Snyder, J. Hattie, R.P. McDonald, Praeger, New York, 122–215, 1984.
85. D.S. Burdick, X.M. Tu, L.B. McGown, D.W. Millican, 'Resolution of Multicomponent Fluorescent Mixtures by Analysis of the Excitation–Emission–Frequency Array', *J. Chemom.*, **4**, 15–28 (1990).
86. C.-H. Ho, G.D. Christian, E.R. Davidson, 'Application of the Method of Rank Annihilation to Quantitative Analysis of Multicomponent Fluorescence Data from the Video Fluorometer', *Anal. Chem.*, **50**, 1108–1113 (1978).
87. N.J. Messik, J.H. Kalivas, P.M. Lang, 'Selectivity and Related Measures for *N*th-order Data', *Anal. Chem.*, **68**, 1572–1579 (1996).
88. Y. Wang, O.S. Borgen, B.R. Kowalski, M. Gu, F. Turecek, 'Advances in Second Order Calibration', *J. Chemom.*, **7**, 117–130 (1993).
89. P. Nomikos, J.F. MacGregor, 'Monitoring of Batch Processes Using Multiway Principal Components Analysis', *AIChE J.*, **40**, 1361–1375 (1994).
90. P. Nomikos, J.F. MacCregor, 'Multiway Partial Least Squares in Monitoring Batch Processes', *Chemom. Intell. Lab. Syst.*, **30**, 97–108 (1995).

91. R.J. Pell, M.B. Seasholtz, B.R. Kowalski, 'The Relationship of Closure, Mean Centering, and Matrix Rank Analysis', *J. Chemom.*, **6**, 52–57 (1992).
92. M.B. Seasholtz, B.R. Kowalski, 'The Effect of Mean Centering, on Prediction in Multivariate Calibration', *J. Chemom.*, **6**, 103–111 (1992).
93. Q. Ding, G.W. Small, M.A. Arnold, 'Evaluation of Data Pretreatment and Model Building Methods for the Determination of Glucose from Near-infrared Single Beam Spectra', *Appl. Spectrosc.*, **53**, 402–414 (1999).
94. A. Savitzky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Anal. Chem.*, **36**, 1627 (1964).
95. P. Geladi, D. MacDougall, H. Martens, 'Linearization and Scatter Correction for Near Infrared Reflectance Spectra of Meat', *Appl. Spectrosc.*, **39**, 491–500 (1985).
96. C.E. Miller, S.A. Svendsen, T. Naes, 'Nondestructive Characterization of Polyethylene/Nylon Laminates by Near-infrared Spectroscopy, *Appl. Spectrosc.*, **47**, 346–356 (1993).
97. T. Isaksson, B.R. Kowalski, 'Piece-wise Multiplicative Scatter Correction Applied to Near-infrared Diffuse Transmittance Data from Meat Products', *Appl. Spectrosc.*, **47**, 702–709 (1993).
98. Y. Wang, M.J. Lysaght, B.R. Kowalski, 'Improvement of Multivariate Calibration Through Instrument Standardization', *Anal. Chem.*, **64**, 562–565 (1992).
99. C.S. Chen, C.W. Brown, S.C. Lo, 'Calibration Transfer from Sample Cell to Fiber-optic Probe', *Appl. Spectrosc.*, **51**, 744–748 (1997).
100. J. Lin, 'Near-IR Calibration Transfer between Different Temperatures', *Appl. Spectrosc.*, **52**, 1591–1605 (1998).
101. P.J. Gemperline, J.H. Cho, P.K. Aldridge, S.S. Sekulic, 'Appearance of Discontinuities in Spectra Transformed by the Piecewise Direct Instrument Standardization Procedure', *Anal. Chem.*, **68**, 2913–2915 (1996).
102. K.S. Booksh, C.M. Stellman, W.C. Bell, M.L. Myrick, 'Mathematical Alignment of Wavelength-shifted Optical Spectra for Qualitative and Quantative Analysis', *Appl. Spectrosc.*, **50**, 139–146 (1996).
103. B.J. Prazen, C.E. Bruckner, R.E. Synovec, B.R. Kowalski, 'Second-order Chemometric Standardization for High-speed Hyphenated Gas Chromatography: Analysis of GC/MS and Comprehensive GC × GC Data', *J. Microcol. Sep.*, **11**, 97–107 (1998).
104. T.B. Blank, S.T. Sum, S.D. Brown, S.L. Monfre, 'Transfer of Near-infrared Multivariate Calibrations Without Standards', *Anal. Chem.*, **68**, 2987–2995 (1996).
105. S.T. Sum, S.D. Brown, 'Standardization of Fiber Optic Probes for Near Infrared Multivariate Calibrations', *Appl. Spectrosc.*, **52**, 869–877 (1998).

Chromatography in Process Analysis

Jerry M. Clemons
106 Maplewood Avenue, Ronceverte, WV, USA

1 Introduction 1
2 History 2
3 Definitions 2
3.1 Chromatography 2
3.2 Gas Chromatography 3
3.3 Process Gas Chromatography 3
4 Measurement Principles 3
5 Laboratory versus Process 5
6 Basic Hardware Components of a Process Gas Chromatograph 6
6.1 Analyzer Section 6
6.2 Oven 6
6.3 Carrier-gas 6
6.4 Sample and Column Switching Valves 7
6.5 Columns 9
6.6 Detectors 10
6.7 Carrier-gas Pressure/Flow Control 13
7 Application Engineering 14
8 Calibration Requirements 14
9 Sample Handling and Conditioning System 15
10 Installation Requirements 15
11 Controller/Programmer 16
12 Data Systems 16
13 Other Chromatographic Techniques 17
13.1 Liquid Chromatography 17
13.2 Supercritical Fluid Chromatography 18
13.3 Preparative-scale Chromatography 19
14 Common Applications 19
14.1 Petrochemical/Chemical 19
14.2 Refining 20
14.3 Natural Gas Industry 21
14.4 Environmental Monitoring 22
14.5 Other Industrial Applications 23
Acknowledgments 23
Abbreviations and Acronyms 23
Related Articles 24
References 24

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Process chromatography is an analytical technique which provides for the on-line separation and measurement of the components of a mixture in a chemical process. The chromatographic analyzers utilized for on-line analysis must operate in a continuous and automatic manner, taking a fixed quantity of sample, separating the chemical components using adsorption or partitioning columns, and measuring the concentration of each with a detector. The chromatograph provides both quantitative results, by measuring the area of the peak generated as the components pass through the detector and qualitative results based on the retention time, which is measured from injection to detection.

The most commonly used process chromatographic analyzer is the process gas chromatograph because of its many applications in the hydrocarbon processing industry. A variety of valve, column, and detector configurations are available for these analyses from parts per million to percent levels. All samples must be gas phase or liquids which can be vaporized prior to analysis. The analyzers are controlled using a cyclic timing device which switches valves and columns as the components pass through the system to provide the necessary separation. This time cycle is a limitation of the technique for applications requiring continuous or very fast analysis times. Even with these limitations, the advantages of cost, flexibility, sensitivity, and reliability have made process gas chromatography (PGC) the most widely used analytical technique for on-line process control.

1 INTRODUCTION

Process chromatography appeared on the scene in the mid- to late-1950s. For the first time, we had an analytical technique with the ability to provide a complete compositional analysis, yet one that was simple in principle and application, and broad in scope and analytical power.

Development of the chromatographic analyzer and its use for process measurement and control has increased at a phenomenal rate over the past 50 years. Uses of the chromatographic analyzer range from the simple monitoring of one component in ambient air to providing a computerized output for 50–100 components in a complex mixture. Today there are literally thousands of these analyzers being used, yet there are few references to their applications. This is primarily due to the proprietary nature of many of the chemical processes where these analyzers are being utilized.[1]

Basically, a process chromatograph resembles its laboratory counterpart more closely than most other plant analyzers. The primary differences result from requirements imposed by the plant environment, safety, and

automatic operation. The process analyzer may be considered as having two distinct sections: the analyzer section (where the analysis actually takes place) and the control section or data system (which generates and reduces the data, sends them to the control computer, and provides maintenance functions). The analyzer will automatically take its own sample, separate the components of interest and provide a detector signal output proportional to their concentrations. These signals are then processed in the control section and sent to a recording device or a process control computer. The analyzer will continue to repeat the same cycle, which lasts from a few seconds to several minutes, 24 h per day and 365 days per year until it is interrupted for calibration or maintenance.

2 HISTORY

Chromatography began at the turn of the century when Ramsey[2] separated mixtures of gases and vapors on adsorbent materials. Later a Russian botanist, Michael Tswett,[3] separated plant pigments by passing a liquid mixture through a column of solid adsorbent material. The term ''chromatography'' is identified with this work published by Tswett in 1906. In this publication he described how brilliant bands of color were produced when a mixture of plant pigment and solvent passed through the column. Based on his findings, he coined the term from the Greek words ''chromo'', which means color, and ''graph'', which means writing, to describe the scientific process as color writing. This paper has been translated in English and republished[4] because of its scientific importance to the field.

Based on the same principles, gas chromatography (GC) evolved, which utilized a gas as the mobile phase. It is usually identified with American chemist, D.T. Day,[5] who performed experiments with crude petroleum fractions and separated many of its components by passing it through a column filled with Fuller's earth.[6] The important seminal work was first published in 1952,[7] when Martin and his co-worker James acted on a suggestion made 11 years earlier by Martin himself in a Nobel Prize-winning paper on partition chromatography.[6] This publication described the use of a gas instead of a liquid as a mobile phase and the results were presented at the Analytical Chemistry Congress at Oxford in 1952.[7] It was quickly recognized that GC was simple, fast, and applicable to the separation and measurement of many volatile chemical components, especially petrochemicals, for which distillation was the method of choice for separation at that time.[8] Theories describing the technique were tested and led to still more advanced theories and understanding.[9] Almost immediately the demand for analyzers incorporating the technique gave rise to a new industry that responded quickly by developing commercially available laboratory gas chromatographs.

Essentially simultaneous with the development of gas chromatographs for this use, a similar form of analyzer was developed for industrial applications. An instrument for PGC first appeared in the mid-1950s as a result of work done by the Union Carbide Corporation in South Charleston, WV, USA. Because they were not interested in the actual manufacture of such a product for the commercial market, the concepts were transferred to a small display and control panel manufacturer, Watts Manufacturing, of Ronceverte, WV, USA. This company first manufactured the product on a contract basis for Union Carbide, followed by manufacturing and distribution on a global basis for the hydrocarbon processing industry.[10] In 1957, Beckman Instruments acquired the company and transferred the manufacturing to California. Several of the technical people who elected not to move to California formed a second company known as Greenbrier Instruments. These two companies led the introduction of PGC to its worldwide adaptation as we know it. Today, PGC is the most widely used multicomponent process analysis techniques. For this reason, GC will be the primary emphasis of this article. It is a mature technique and a very important one for process monitoring and control. The worldwide market for PGC analyzers is estimated to be about $100 million or over 2000 analyzers annually.[11]

3 DEFINITIONS

To understand better the basis of this technology, definitions of chromatography, GC, and PGC will be introduced.

3.1 Chromatography

Chromatography is defined as a technique for the separation of components in a mixture. Both qualitative and quantitative measurements are possible with this technique; however, in process analyzers, the quantitative aspects of the application predominate.

Basically, chromatography consists of a two-phase system. One phase is stationary, or static, and consists of the separating media. The second phase is mobile and serves to move the sample being analyzed over or through the separating media. Chromatography can be classified according to the type of moving phase and the static phase, e.g. gas–solid chromatography or liquid–solid chromatography. This describes the chromatographic process according to the physical state of the mobile

phase. Thus, in GC the mobile phase is a gas, and in liquid chromatography (LC) the mobile phase is a liquid.

3.2 Gas Chromatography

In GC, the separating media are contained in columns and consist either of a solid adsorbent material (gas–solid chromatography) or of a liquid held by a solid (gas–LC). The mobile phase consists of a gas which pushes or carries the gaseous sample through the column. The separation of the components in the sample results from a difference in the forces by which the column materials tend to hold or retain each of the sample components. Whether the nature of these forces is adsorption, solubility, chemical bonding, polarity, or molecular filtration, they retain some components longer than others. Consequently, all components pass through the column at varying speeds and emerge in inverse order of their affinity for the stationary column materials.

Normally, the components tend to separate according to their boiling points or vapor pressure, with the low-boiling, high-vapor-pressure components emerging first. However, other physical or chemical properties also play a part in the separation, and it is possible to affect the separation of components whose boiling points or vapor pressures differ only slightly or not at all using properties such as polarity, molecular size, or chemical interactions.

3.3 Process Gas Chromatography

A process gas chromatograph is an analyzer which has been designed for installation and operation on-line with a chemical process. The analyzer will continuously and automatically analyze a flowing process stream in a cyclic, repetitive manner to determine the chemical composition of the flowing process stream. The primary purpose of locating the analyzer on-line or adjacent to the chemical process is to obtain the analytical results with a speed of response that is comparable to process changes. This information will, in turn, be used for process monitoring and control. In general, such an analyzer is designed with the columns, valves, and detectors to perform a particular analysis on a single stream, or at most, a few liquid or gas streams of similar composition. The typical analyzer is designed to measure only key components necessary for process control in the stream. The cyclic nature of analysis which ranges from a few seconds to minutes is viewed by many as a limitation of the technique.[12]

4 MEASUREMENT PRINCIPLES

As a method for separating and measuring the individual components of a complex mixture, GC is unequaled. A qualitative description of the separation process is as simple as is the essential apparatus. An understanding of the separation process is critical for a real appreciation of the power and scope of the technique. More specifically, a description of the processes going on in the separating column, the heart of any chromatographic system, is important to understanding the operation of a gas chromatograph.

For gas–LC, a separation column is prepared using a small-diameter tube (ranging from 0.7 to 4 mm internal diameter) which is packed with inert solid particles coated with up to 20 wt.% of a nonvolatile liquid phase.

The support for the liquid phase can also be the walls of the tube, in which case the tubing diameter will be in the capillary range of micrometers (typically sizes are 100, 250, 320 and 500 μm). The moving phase consists of an inert carrier-gas (N_2, H_2, He) flowing through the column. Into this moving stream is injected the sample mixture, the components of which are in equilibrium in characteristic ratios between the gas and liquid phases. In essence, components more soluble in the liquid move more slowly through the column and a detector at the exit will measure the components as they emerge as a series of peaks. The retention time or time it takes for each component from injection to elution from the column is the qualitative characteristic, whereas the area under the peak is a quantitative measure of its concentration. A chromatogram of a mixture is shown in Figure 1.

There has been a great deal of theory developed about gas–liquid partition principles. Fundamental to the principle is the fact that a given component partitions itself between the gas and liquid phases in a definite ratio. This partition ratio of a component i is illustrated in Figure 2, and given by Equation (1):

$$k_i = \frac{K_i V_l}{V_m} \tag{1}$$

where K_i is the partition coefficient and V_l is the liquid volume in the column and V_m is the gas volume. The value of k_i can vary according to the ratio V_l/V_m.

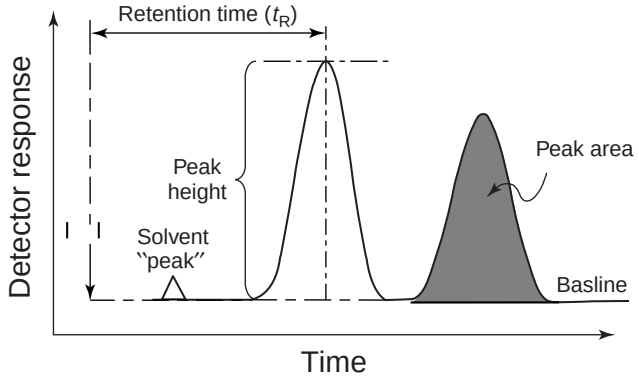

Figure 1 A typical chromatogram.

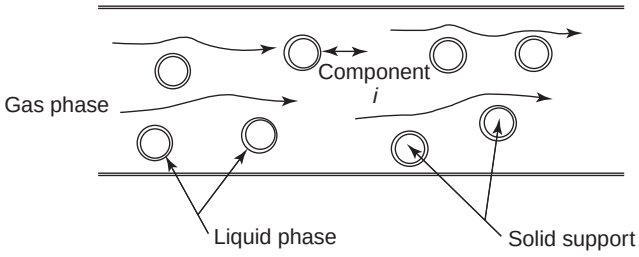

Figure 2 Partitioning between gas and liquid phases.

One can measure k_i from the information illustrated in Figure 3 (using Equation 2) for the partition ratio:

$$k_i = \frac{t_R - t_0}{t_0} \tag{2}$$

The peak t_0 is the retention time of the air peak, which represents the retention of a non-absorbed gas.

The efficiency of a column is measured by using a concept from distillation theory; the total number of theoretical plates, N, or average plate height, H, which is L/N, where L is the column length. In Figure 4 this is given by the equation for efficiency (Equation 3):

$$N = 16\left(\frac{t_R}{w_b}\right)^2 \tag{3}$$

where t_R is the retention time from sample inject and w_b is the width of the base of the peak as shown in Figure 4.

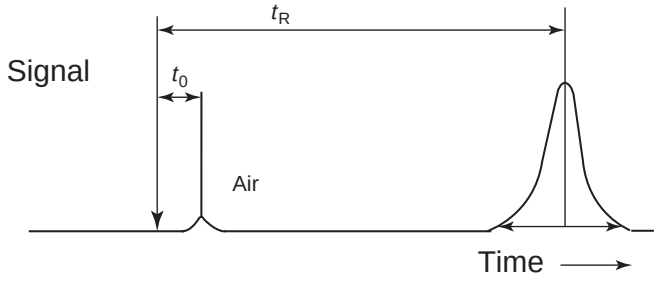

Figure 3 Partition ratio.

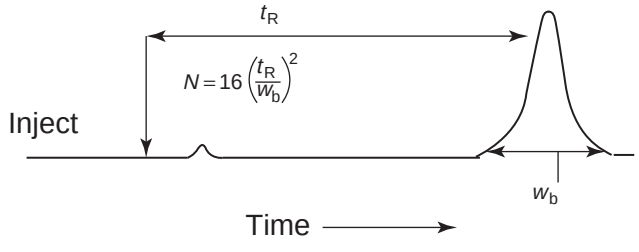

Figure 4 Efficiency: number of theoretical plates, N.

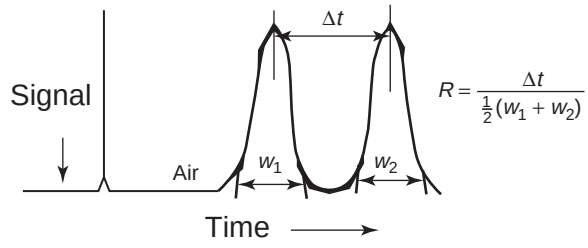

Figure 5 Resolution, R.

This provides a measure of efficiency by showing how much the peak broadens as it moves through the column.

Another parameter often measured is the separation or resolution produced between two components, 1 and 2. This is measured by using the data in Figure 5 applied to the equation for resolution (Equation 4):

$$R = \frac{\Delta t}{\frac{1}{2}(w_1 + w_2)} \tag{4}$$

A resolution of 1 means that two adjacent peaks are just separated at their bases. These three equations allow us to determine the performance of a chromatographic system under different conditions.

The plate theory and its associated equations as presented above provide us with a method to measure column performance. A different theoretical treatment is necessary to understand the experimental factors that affect column performance. To evaluate the reasons for peak broadening under the realistic conditions of column operation, Van Deemter[13] developed a well-known equation which is useful in its simplified form (Equation 5):

$$H = A + \frac{B}{u} + C\bar{u} \tag{5}$$

where A, B and C are constants containing terms related to the factors shown in Figure 6, H is the measured plate height, and u is the carrier linear gas velocity.

From a qualitative perspective, this equation tells us a lot about the behavior of the chromatographic system. The A term is a measure of peak broadening resulting from molecules going down unequal paths in the column. It can be minimized by maintaining uniform particles and packing density in the column. The B term represents the peak broadening caused by molecules diffusing longitudinally along the column length. This effect can be minimized by the choice of carrier and flow rate of the carrier, but we must consider the C term in this region. The C term measures band broadening resulting from the time required to reach equilibrium between the gas and liquid phase, and is a measure of resistance to mass transfer.

The above factors for A, B, and C are a generalized way of looking at column performance under different

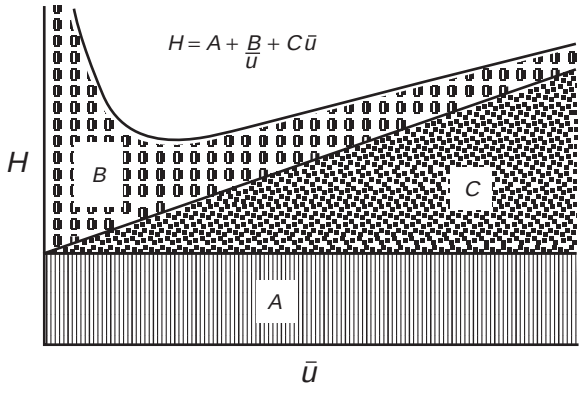

Figure 6 Van Deemter plot. (Reproduced by permission from Dr H.M. McNair, Virginia Polytechnical Institute and State University, Blacksburg, VA, USA.)

operating conditions. For a more complete understanding of the mechanisms involved, we can refer to complete treatment by Giddings.[12]

We now have some understanding of the manner and direction in which variable factors influence the performance of chromatographic columns and methods of measuring their performance. This knowledge is very useful in the design and operation of process gas chromatographs for the wide range of applications for which they are employed.

5 LABORATORY VERSUS PROCESS

PGC is defined as an industrial GC method that can be carried out on an automatic, repetitive basis for the purpose of controlling process streams. Although analytical chemists agree that the principles for a gas chromatograph are the same for laboratory and process analyzers, the requirements and, therefore, the analyzers differ considerably. Although in principle laboratory analyzers can be adopted for on-line use, special process analyzers are almost always used. Because they are operated continuously and with little attention from operators, process gas chromatographs must meet higher standards of reliability.

Process gas chromatographs are expected to provide long, accurate, trouble-free service with minimum maintenance – a difficult task considering the relatively complicated hardware and harsh operating environment. The hardware should be constructed such that when maintenance is required it can be performed quickly and easily. A large number of chemical plants using process gas chromatographs are operated under rigid safety rules requiring protection from corrosive and flammable process materials. Provisions must be made for operation in a relatively unprotected plant environment. The typical process gas chromatograph is designed and the valves, columns, and detectors are configured for one specific analysis from one or more process stream with similar chemical composition.

An important requirement of a process gas chromatograph is that it provides meaningful data in the simplest manner possible. The routine presentation of data of a conventional laboratory gas chromatograph is not practical for a process gas chromatograph. To provide the required data the typical readout system for a process gas chromatograph consists of a bar graph, trend, or digital signal presentation (Figure 7).

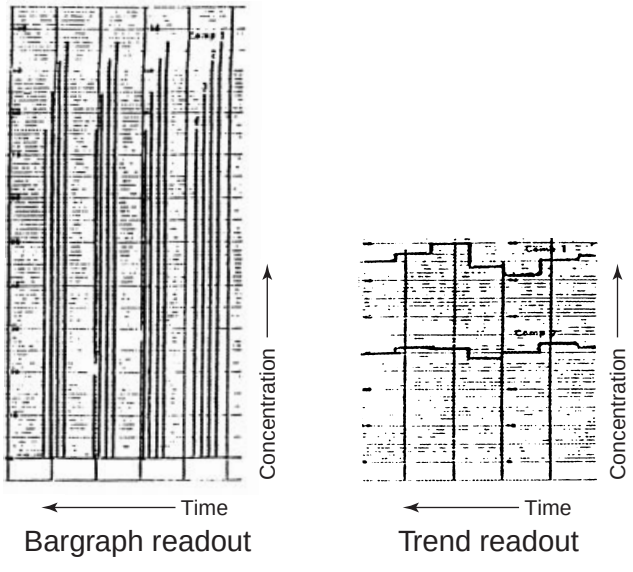

Cal Type= External Standard Mode
Analysis Normalization= NO
Multiplier Exponent = 05
K Factor = 1.000

Name	Units	Time	RF	Cal Conc
		Dev	Calc By	Bench
C6+	%	80.	18.25	0.300
		8.000	Area	0.0000
ETHANE	%	152.	53.35	1.000
		8.000	Area	0.0000
PROPANE	%	210.	35.44	6.000
		12.00	Area	0.0000
ISO-BUTANE	%	280.	30.98	45.75
		12.00	Area	0.0000
N-BUTANE	%	330.	31.60	45.75
		12.00	Area	0.0000
I-PENTANE	%	500.	31.40	1.000
		12.00	Area	0.0000
N-PENTANE	%	565.	36.40	1.000
		12.00	Area	0.0000

Figure 7 Bargraph trend and digital outputs.

Some unique features of a process gas chromatograph include:

1. a constantly fresh sample since the analyzer is connected directly to the process stream;
2. faster analysis times;
3. multiple columns and valves;
4. less manpower required for operation;
5. results immediately available to the process operator;
6. can be incorporated into closed-loop control;
7. isothermal operation is most common;
8. designed for operation in hazardous environments.

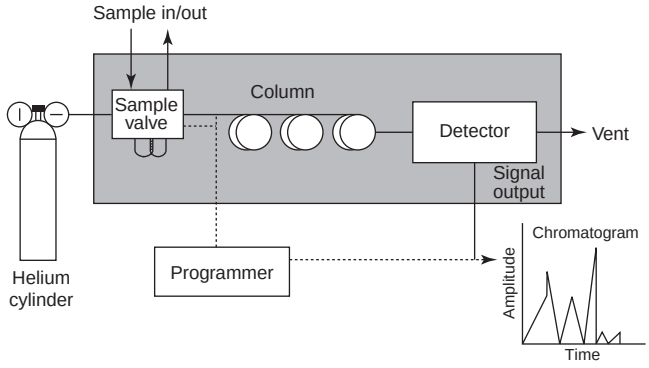

Figure 8 A basic process gas chromatograph. (Reproduced by permission of R. Villalobos. Copyright Instrument Society of America.)

6 BASIC HARDWARE COMPONENTS OF A PROCESS GAS CHROMATOGRAPH

The analyzer one utilizes in a chemical plant will necessarily have fundamental differences from one in the laboratory employing the same principles. The plant environment will require a rugged and usually explosion-proof or air-purged design. The analyzer will obtain its own sample, analyze it, process the information coming from the detector into useful information, and output the data to a recorder, printer, or computer. It must continuously repeat the same cycle to provide analytical results needed for precise process control. Because of the importance of this feature, the automatic performance and reliability of a process gas chromatograph must be very high.

A process gas chromatograph may be thought of as being divided into two sections, an analyzer section and controller or programmer section. These two sections will be presented in detail, starting with the analyzer (Figure 8).

6.1 Analyzer Section

The analyzer section contains the basic hardware components such as sample valves, columns, column switching valves, and detectors, enclosed in a precisely temperature-controlled oven. This section is usually explosion-proof or air-purged to meet the requirements of a hazardous area. In certain applications or designs it may be necessary to locate pneumatic components in the oven to eliminate variances caused by ambient temperature changes.

6.2 Oven

The typical process gas chromatograph employs a single, isothermal temperature zone. Three basic designs have been used, the air bath oven, airless oven which uses resistive heating elements, and the explosion-proof bell-jar design. The most common is the air-bath oven which incorporates an explosion proof heater and detector in a non-explosion proof air bath enclosure. Using dilution purging, these ovens meet the requirements for location in hazardous areas, i.e. Division 1 or Zone 1.

With the air-bath design, 2–3 standard cubic feet min^{-1} or 1.0–1.5 standard $L\,s^{-1}$ of air is passed through an electrical heating element. This oven's design is capable of operating at temperatures from 50 to 200 °C with a stability of ±0.1 °C or better.

Although isothermal operation is the most common, temperature-programmed ovens are available for special applications involving a wide boiling-point range of chemical components. These temperature-programmed designs normally provide isothermal temperature control for sample injection valves and the detector, with only the column being temperature-programmed. This is accomplished using a small temperature-programmed oven within the isothermal oven. This oven-within-oven design provides the speed and precise temperature programming necessary for process control (Figure 9). Another feature utilized to reduce the time required to cool the programmed oven to its initial temperature is a vortex cooler that pumps cool air into the oven following a program to the final temperature. The design of a temperature-programmed oven requires special precautions to ensure that it meets the hazardous area requirements.

6.3 Carrier-gas

Basic to the efficient operation of a process gas chromatograph is the quality of the carrier-gas and other support gases. The carrier-gas must be free of hydrocarbon impurities and moisture for proper operation. Often impure carrier-gases are responsible for reducing the normal useful life of the column system used for specific applications. In many countries it is difficult, if not impossible, to obtain gases of sufficient quality to allow proper

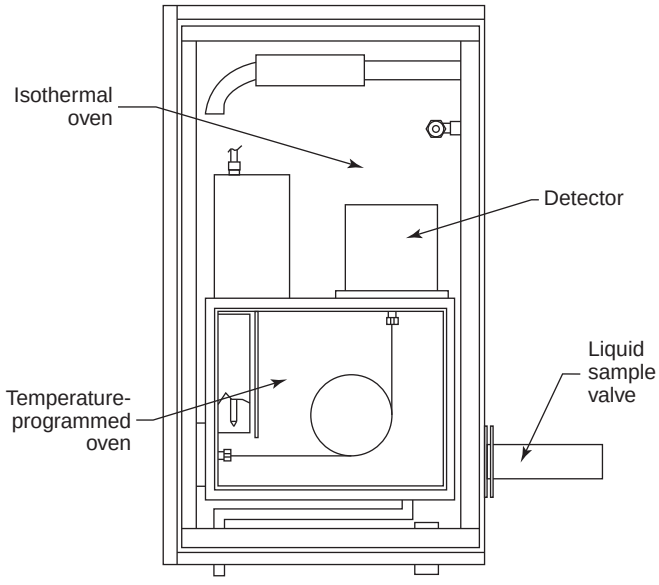

Figure 9 Temperature-programmed gas chromatograph.

operation of the gas chromatograph. This limitation can be overcome to some extent by using cleanup systems such as traps and catalytic converters. These systems are helpful but must be properly maintained to ensure long-term operation.

The type of carrier-gas used for a specific application is normally detector dependent. The thermal conductivity detector (TCD) uses either hydrogen or helium as the carrier-gas. For the flame ionization detector (FID) the carrier-gas most commonly used is nitrogen. Other factors which dictate the choice of carrier-gas include safety issues, cost, speed, availability, purity, and the components being measured.

6.4 Sample and Column Switching Valves

Process gas chromatograph valves are uniquely designed to introduce reliably a fixed volume of gas or liquid sample into the carrier-gas stream and to switch columns into various positions in the carrier-gas stream, while introducing minimum void volume into the system.

Numerous types of valves have been used, but the most common are the sliding plate, diaphragm, rotary, and liquid-injection sample valves. For gases an external sample loop is used to control the volume of sample injected. With liquid samples the volume of sample is metered with an internal hole or channel in the valve slider of the stem. The volume of sample injected in either case is dependent on the sensitivity necessary for the application, but is typically milliliters for gases and microliters for liquids.

The choice of sample valve is a critical element in every application, since it must be compatible with the sample and operate reliably at the operating temperature necessary to perform the application. These valves are available in corrosion-resistant materials such as Hastelloy™, Monel™, and Teflon™. The liquid injection valve design permits injection of the liquid sample through the wall of the analyzer from the ambient outside temperature to the heated interior where the sample is flash vaporized by a temperature-controlled vaporizing zone prior to being transported to the column by the carrier-gas. This is particularly beneficial for liquids with low-boiling components (such as butanes), which will vaporize when exposed to the oven temperatures. This partial vaporization of the liquid sample will result in nonrepeatable sample volumes and subsequent errors in the measurement.

In addition to injecting samples, the various types of valves described are also used for column switching. The primary reason for column switching, sometimes called multidimensional chromatography, with process gas chromatographs is to obtain a speed of response necessary for optimum control of the process being controlled. Other reasons include:

1. All components in the sample must be quantitatively accounted for and removed from the column system in each cycle. Components which are not eluted will accumulate and change the characteristics of the column or elute with a subsequent analysis and interfere with a measurement.
2. The column separation must be designed for upset conditions which may occur in the process in addition to normal conditions.
3. Columns must be protected from components which are irreversibly or strongly absorbed.

Because of these design requirements, process column design has greater emphasis on multicolumn, multivalve switching techniques. While there are literally hundreds of possible column configurations, there are a few basic concepts which are used in combination with all these configurations. These configurations and how they are used are described in the following examples.

6.4.1 Backflush to Measure or to Vent

This is the most widely used technique, appearing as a part of every column application. It consists of two columns, a backflush column and an analytical column in series, with provision for backflushing the first or backflush column to vent while providing carrier-gas to the analytical column during the backflushing operation.

The backflush column makes a partial separation and is then backflushed, rejecting any unmeasured components. The remaining components are separated in the analytical column prior to passing into the detector.

In general, the use of backflushing is good housekeeping practice; it ensures that unknown heavy components are removed from the column system in each cycle, preventing interference in later cycles.

The backflush to measure configuration is similar except that the backflush column is backflushed to the detector, permitting measurement of all the components remaining in the column. A typical application utilizing this technique is the measurement of British thermal units (BTU) or heating value, where the C_{6+} fraction is backflushed to measure.

6.4.2 Dual Column

In this configuration a valve may be used to switch columns so that the components eluted from the first column are passed either into a second column for storage and further separation or directly into the detector. A typical example involves the separation of light or fixed gases, O_2, N_2, CO, and CH_4, on a dual column, such as 5 Å molecular sieves, which is a commonly used gas–solid column packing material.

6.4.3 Heart Cut

This configuration is used most frequently for trace analysis with an FID. The application involves the measurement of trace components on the trailing edge of a major component such as trace acetylene in an ethylene product stream. Two columns are arranged to allow narrow cuts of sample from the first column to be taken into the second column, the bulk of the sample being discarded to vent. The trace component plus the tail of the major component are separated on the second column. The second column only has to separate the trace component from a few percent of the major component instead of the 90+% necessary without a venting or heart cut of the major component.

6.4.4 Selector or Foreflush Valve

This configuration provides the ability to select from two parallel column systems, wherein one at a time will be directed to the detector. This will allow two different types of columns to be used for separation of different fractions of the sample. This configuration may also be used to select between two different detectors to provide the sensitivity necessary to measure different fractions of the sample.[12]

There are other configurations of column switching which have been used over the years, but the above are the most common, and a basic understanding of these will encompass the majority of applications.

The valves used for column switching are the same as those used for sample injection, with the exception of the liquid sample injection valve. These include the sliding plate, rotary, and diaphragm designs. The sliding plate and rotary valves are similar in concept (Figure 10). They utilize a drilled or grooved surface slider or rotor, which is actuated from one position to another, either in a linear or rotary motion, along the finely machined surface of a metal plate or housing. The grooves in the surface connect the valve ports that have been machined in the metal plate. In each position the ports are connected in a different configuration. This provides two different flow paths for the carrier-gas through the valve to allow the valve to perform one of the functions described

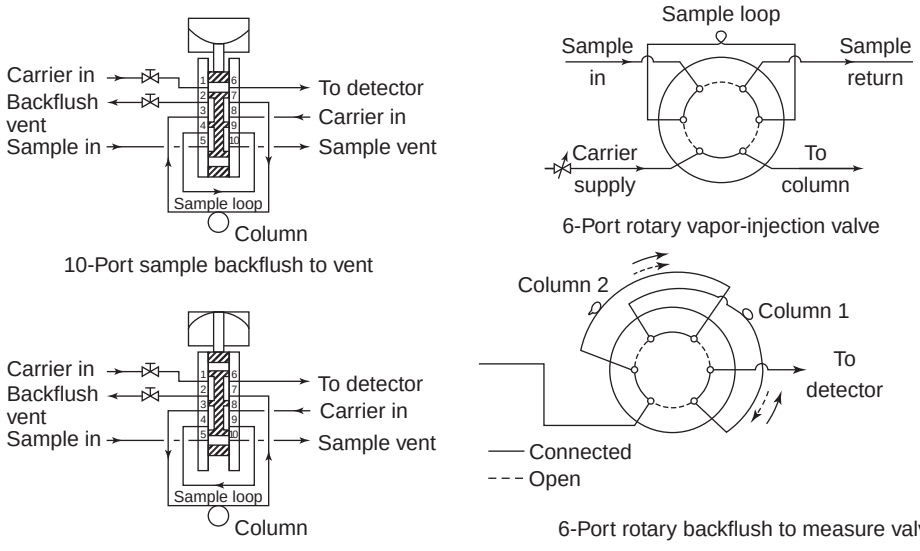

Figure 10 Sliding plate and rotary sample valves. (Reproduced by permission. © Copyright Instrument Society of America. All rights reserved.)

above. Typically, springs are used to press the slider or rotor against the plate. A wide variety of construction materials are available for the slider, rotors, and plates to meet the special requirements of an application based on corrosiveness, pressure, and temperature. Normally air pressure is used to switch the valve between positions.

A different concept is used in design of the diaphragm valve (Figure 11). Instead of a single surface moving against another as used in the rotary or sliding plate design to change the direction of flow of the carrier-gas through the various ports, sets of rods push against a flexible diaphragm to open and close alternating ports. This design is also spring loaded to maintain the tension needed to seal the ports being switched. As with the sliding plate and rotary valves, air is used to actuate the valve or to switch from one position to another. This type of valve is recognized for its fast switching time and low dead volume.

The liquid injection valve or syringe valve is ideally suited for the injection and vaporization of liquid sample for PGC (Figure 12). The valve is mounted through the oven wall with the actuator and sample chamber external to the oven. The vaporizing chamber and carrier connection are located inside the oven. Sample flows through the sample chamber and fills a groove cut around the circumference of the valve stem. This groove provides

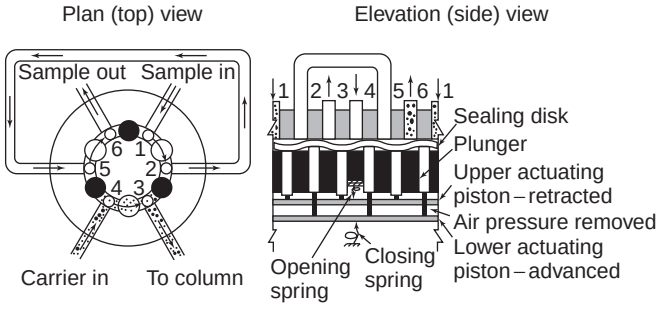

Figure 11 Diaphragm sample valve. (Reproduced by permission of Applied Automation Inc.)

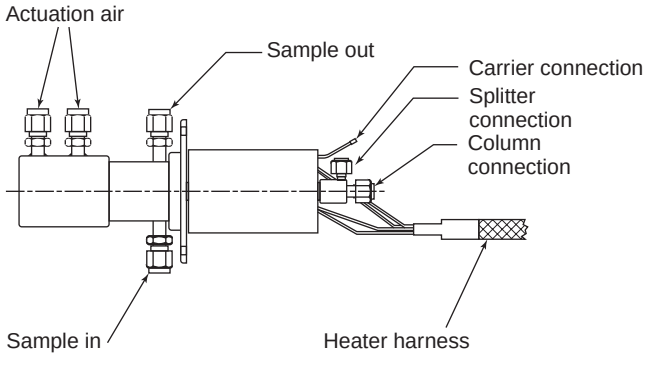

Figure 12 Liquid sample injection valve.

for a fixed volume of sample ranging from 0.25 to 10 µl. The valve is actuated by an air-driven piston which pushes the stem into a vaporizer where the sample is flash vaporized and carrier-gas transports the sample to the column for separation. The operation of the valve is analogous to syringe injection with a laboratory chromatograph where a fixed volume of sample is injected through a septum into a vaporizer and transported into the column by the carrier-gas.

Each of the valves described has advantages and disadvantages, depending on the application in which they are used. The most important criteria for the selection of one valve over another are reliability and ease of maintenance. Since the sample-injection and column-switching valves are essentially the only moving, mechanical parts of the process gas chromatograph, they are at the top of the list of parts most likely to fail. For this reason, a great deal of effort has gone into the design of the valves over the years since PGC was introduced.

6.5 Columns

For many years, packed columns were the only type of columns used for PGC. With evolving demands requiring faster, more complex separations, capillary columns are replacing packed columns for many applications. Because capillary columns are much more efficient, that is, they have more theoretical plates per unit length, they provide simplified separations and faster analysis times. Capillary columns are especially advantageous for high-molecular-weight samples which are liquids at ambient temperature. Lighter gas-phase sample separations are still more easily performed with packed columns. The use of capillary columns in PGC also dictates a change in the design of the hardware components. The dead volume in valves and detectors must be minimized to take advantage of the efficiency gained using these columns.

6.5.1 Packed Columns

Typical column diameters for PGC applications are 1/8, 3/16, or 1/4 in (2, 3, or 4 mm) internal diameter, with 1/16 in (0.7 mm) internal diameter being used in special high-speed applications. The length of the column depends on the difficulty of the application and selectivity of the packing material for components being measured. The packed column length seldom exceeds 20 f (6 m) because of the pressure drop across the column. With large pressure drops it is difficult to obtain the optimum flow rate through the column, resulting in peak broadening and reduced sensitivity. The packing material used in the column is either an absorbent (gas–solid chromatography) or a partitioning material (gas–LC). In the case of adsorbent packings such as

Porapaks, molecular sieves and Durapaks, they are used as purchased with a particular mesh range or particle size (80–100 or 100–120 mesh are most common). For partitioning columns, a stationary liquid phase is coated on a solid support with a large surface area. The weight percent of liquid phase coated on the solid support may be varied to enhance the separation (20 wt.% is the maximum). The solid support is normally 80–100 or 100–120 mesh for 1/8 in (2 mm) internal diameter columns. These packing materials may be purchased with the liquid stationary phase already coated on the solid support. In fact, most users of process gas chromatographs purchase identical prepacked columns as replacements for existing columns. This eliminates or reduces to a minimum the method development step.

6.5.2 Capillary Columns

As the demands on the process gas chromatograph increase relative to speed of analysis, trace measurement, complexity of sample, and component resolution, the inherent advantages of capillary columns become very important. There is a general feeling that capillaries are more producible, expensive, fragile, and difficult if not impossible to work with. The current switch of interest to this type of column for process applications is the result of two significant developments, as follows.

The development of commercially available surface-bonded columns has extended column lifetimes and reduced associated problems with column bleeding. The second development involves the elimination of the mechanical instability problem. The reactivity of metal capillaries and mechanical instability of glass capillaries have been overcome by the introduction of fused-silica columns.[15] Because fused silica has a much higher degree of cross-linking within the silicon–oxygen matrix than ordinary glass, it enjoys a higher tensile strength and permits the construction of thin-walled flexible columns. To protect against surface corrosion, several coatings of a polyimide are applied to the outer wall of the flexible fused-silica column. Most recently, silica-coated steel capillary tubing has become commercially available and the further enhanced mechanical stability has led to the replacement of many fused-silica columns for PGC applications.

Commonly capillaries of 0.25 mm internal diameter and varying lengths (30 m is most common) are used; smaller internal diameters have been used for faster analysis. The introduction of megabore (0.35–0.50 mm) columns allows their use as replacements for packed columns with minimum hardware changes being required. These columns provide higher efficiencies than packed columns using the same stationary phase. Better separation is achieved since longer lengths can be used (because of high permeability).[16]

6.6 Detectors

After a chromatographic column has separated the components in the sample, they pass into the detector, where they are measured. If the chromatographic system has been properly designed, the electronic output from the detector will be proportional to the concentration of the component. Several different types of detectors have been developed, but those most commonly used can be separated into two categories: concentration and mass flow rate. This classification system distinguishes between those detectors which measure the concentration of the sample in the carrier-gas and those which directly measure the absolute amount of sample irrespective of the volume of carrier-gas. Examples of concentration-type detectors would be the TCD and electron capture detector, while the FID is a mass flow rate type. One consequence of this difference is that the peak areas and peak heights are affected by changes in carrier-gas flow rate.[17]

To understand better the reasons for this difference in detector types, consider the effect on a TCD signal if the flow is completely stopped. The detector cell remains filled with a given concentration of sample and its thermal conductivity continues to be measured at a constant level. However, for a mass flow rate detector such as the FID, in which the signal is generated from the ionization process in the burning flame, complete cessation in the flow rate will cause the delivery of sample to the detector to stop and the signal will drop to zero. Consequently, the quantitative results acquired at different flow rates will be affected. This affect emphasizes the need for well-controlled flow rates through the chromatograph to ensure repeatable results.[18]

The chromatogram produced by a detector consists of a series of peaks, each of which corresponds to a different component as it passes through the detector. The area under each peak is proportional to the total mass of that component. Hence the chromatographer may calculate weight percent compositions from area ratios represented by the chromatogram. With the TCD, the peak area is inversely proportional to carrier-gas flow rate, so the flow rate must be kept constant for accurate quantitative analysis. For detectors responding to mass flow rate, such as the FID, the peak area is independent of carrier-gas flow rate.

Generally, the usefulness of a detector, regardless of the type, may be evaluated in terms of the following characteristics:

1. selectivity
2. sensitivity
3. response
4. noise
5. linear range.

Secondary to these aspects for PGC, it is important that the detector of choice is simple, rugged, and insensitive to minor changes in flow rate and temperature.

The most common detector types used for PGC applications are the TCD, because of its universal response, the FID, because of its sensitivity to organic components, and the flame photometric detector, because of its selective response to trace levels of sulfur components. Numerous other detectors have been used because of their specificity towards certain components, but their applications are very few. The three most commonly used detectors will be reviewed relative to their operation, advantages, and limitations.

6.6.1 Flame Ionization Detector

The FID is probably the most popular detector used for PGC because of its wide dynamic range (10^7) and sensitivity to trace-level components (parts per billion levels of many organic components). Current designs of this detector exhibit extremely low dead volumes, making them ideally suited for capillary column applications.

The gas chromatograph column effluent flows into an oxygen-rich hydrogen flame where the combustible components are burned, producing fragment ions. The ions produced are subjected to an electrical field produced by two electrodes, the jet itself acting as one electrode and the other as the collector. A potential of 150–300 V is applied across the electrodes, one negative and the other positive. The ionic species produced in the flame provide a source for current flow between the electrodes. The current flow is directly proportional to the concentration of ions produced, which, in turn, is proportional to concentration of the component in the column effluent. The hydrogen fuel and column effluent are mixed and pass through the jet. A surrounding flow of air is provided to support combustion. The detector response is dependent on the flow rate of the carrier, air, and hydrogen, the flow rate of hydrogen being the most critical. Typical flow rates used are 30 mL min^{-1} of carrier, 30–40 mL min^{-1} of hydrogen, and 300 mL min^{-1} of air (Figure 13).

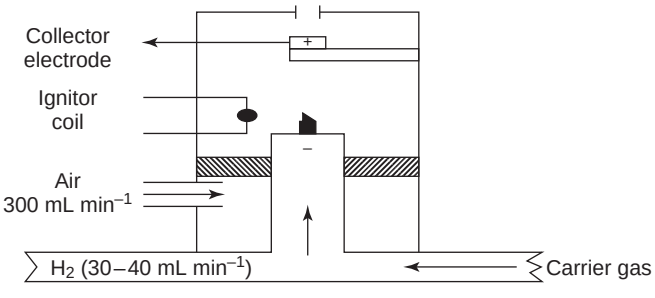

Figure 13 An FID. (Reproduced by permission from Dr H.M. McNair, Virginia Polytechnical Institute and State University, Blacksburg, VA, USA.)

The FID is essentially nonresponsive to the components of combustion, inert gases, and inorganic compounds. It is highly responsive to organic components, the level of which depends on the number of carbon atoms and the other molecules to which they are bonded. For simple molecules there is essentially a one-to-one relationship between the number of carbon atoms and the response. For more complex molecules containing oxygen, halogens, nitrogen, and sulfur atoms, in addition to the level of saturation within the sample molecule there is a corresponding effect on the flame response. For this reason it is necessary to calibrate the FID for each component being measured.

The list of components which cannot be directly measured with an FID includes H_2O, NH_3, CO, CO_2, COS, CS_2, SO_2, H_2S, and the inert gases. In the case of CO and CO_2, a catalytic conversion with hydrogen and a nickel catalyst at 400–450 °C produces methane, which can be measured with the FID.

Although the FID is considered to be a very sensitive detector, its efficiency as an ionization source is limited. A very small percentage of molecules passing through the flame are actually ionized and the high sensitivity possible is a result of its very low noise level (10^{-13}–10^{-14} A) of the output. This low noise level allows high-gain electrometers to be used to amplify the current output generated by the ionization process.

The purity of the air being used is critical for proper operation of the FID. Any traces of hydrocarbons, moisture, or particles that pass into the flame will result in a noisy output. If it is necessary to use plant or instrument air, an air cleanup oxidization catalyst should be used.[10]

A unique feature often incorporated in a process FID is a flame-out sensor. The sensor is normally a thermocouple which senses the flame temperature and provides a signal to the associated electronics when the flame is extinguished. In the event that there is a flame-out, an automatic ignition circuit is energized which provides a current through an ignition coil, resulting in reignition of the flame. Such circuits may be designed to limit the length of time the ignitor is energized because of safety considerations.

6.6.2 Thermal Conductivity Detector

The TCD is a concentration-sensitive detector with moderate sensitivity (low parts per million) and a dynamic range of 10^4. The TCD measures the rate at which heat is conducted away from a hot body. The rate of heat loss depends on the composition of the surrounding gas. This composition varies as the components pass through the detector from pure carrier, to carrier plus sample, to pure carrier. This transition results in a peak being recorded as the output from the detector.

The TCD exhibits a universal response and is inexpensive, rugged, and simple compared with other detector types. The basic design of the TCD consists of a large-mass, metal block made from steel or aluminum. Holes are drilled in the metal block to mount the filaments or thermistors. A carrier flow path with one of three designs (diffusion, semidiffusion, or flow-through) is made in the block to allow the carrier to pass by the filament or thermistor. The filament is a metal coil with a large resistance to temperature coefficient, such as tungsten or tungsten and rhenium alloy, and the thermistors are metal oxides. The filaments are electrically connected in a simple Wheatstone bridge arrangement as shown in Figure 14. Sample plus carrier flow through one side of the detector, called the sample side, and the other side of the detector is the reference side, with carrier-gas only.

An electric current is passed through the filaments, causing them to be heated to a constant temperature depending on the amount of current supplied, the thermal conductivity of the carrier-gas, and the temperature of the thermal conductivity block. With only carrier-gas flowing through the sample and reference sides of the detector, the Wheatstone bridge is balanced and the detector signal is at a minimum. As a component elutes from the column and passes the detector the filament begins to heat up, and the resistance changes because the carrier plus sample has a lower thermal conductivity than the pure carrier. As a result of the change in resistance the Wheatstone bridge is unbalanced and a signal is recorded. For maximum sensitivity a carrier-gas with high thermal conductivity such as hydrogen or helium is selected. Other factors affecting sensitivity are the current through the filaments and temperature of the TCD block. As the current supplied to the filaments is increased, the sensitivity is increased, and unfortunately this also results in shorter filament life. The detector block temperature should be maintained as low as possible, to maintain the maximum differential in temperature between filament and block for maximum sensitivity.

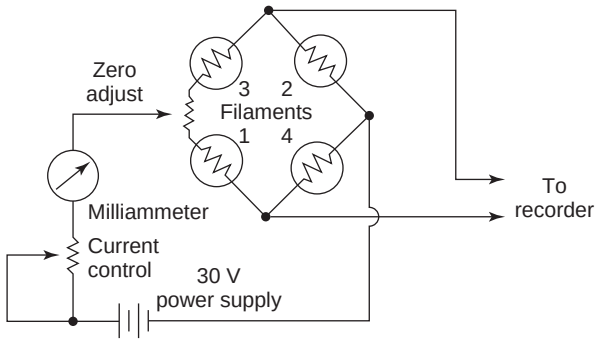

Figure 14 Wheatstone bridge of a TCD. Filaments 1 and 2 are reference filaments. (Reproduced by permission from Dr H.M. McNair, Virginia Polytechnical Institute and State University, Blacksburg, VA, USA.)

The TCD is one of the most popular detectors for PGC because of its universal response. This detector senses any component other than the carrier-gas that passes over the filaments. Consequently, many components which cannot be measured with an FID are possible with a thermal conductivity design.

The primary limitation is the sensitivity, typically in the tens to hundreds of parts per million range, depending on the component being measured. Many attempts have been made to enhance this sensitivity because of the many other favorable attributes (simple, rugged, inexpensive) of this detector. In fact, it has been suggested that for low volume flow rates and small sample sizes required with very fast chromatography, the TCD response will surpass that available from the FID.[19]

6.6.3 Flame Photometric Detector

The flame photometric detector is used in PGC specifically for applications involving trace-level sulfur-containing compounds. Similarly to the FID, it uses a hydrogen flame to burn the sample as it elutes from the column. Unlike the FID, the flame is hydrogen-rich (reducing atmosphere) for the flame photometric detector. Instead of measuring the current flow produced by an ionization process, the flame photometric detector measures the light emitted when the sulfur compound is burned. The emission intensity of specific wavelengths, characteristic of sulfur-containing molecules, is selected using a fixed filter (394 nm) and passes into a photomultiplier tube where the signal is measured and a voltage output proportional to the concentration is generated.

As the sulfur-containing compound is burned in the flame, excited-state S_2^* molecules are formed. As these molecules return to the ground state they give up energy in the form of light of a specific wavelength which may be measured with the photomultiplier tube. The detector has a very high sensitivity (parts per million or billion) for small sulfur-containing molecules such as H_2S, COS, CS_2, SO_2, CH_3SH, and C_2H_5SH. This unique, specific sensitivity for sulfur-containing molecules make the detector ideal for measuring trace levels in natural gas, ethylene, propylene, emissions, and other hydrocarbon-rich streams. Because of the high level of sensitivity it is critical that the support gases used, such as carrier, hydrogen, and air, are also of high purity. The complete chromatographic system must be clean to avoid high background interference and detector noise which results in a loss of detector sensitivity. The signal being measured is nonlinear, based on the formation of the S_2 molecule, which requires special handling of output signal. Either a calibration curve must be generated or

a linearizing circuit must be built into the electronics. Although special precautions must be taken to ensure optimum performance of this detector, it is unique in its ability to measure trace levels of sulfur compounds in a hydrocarbon background. In many situations the purpose for this measurement is the protection of a very expensive catalyst, making the extra effort necessary to operate and maintain the detector worthwhile.

6.6.4 Other Detectors

A number of other detectors such as electron capture, photoionization, and pulsed ionization discharge have been used in PGC applications, but the use is minimal for the trace analysis of compounds which contain molecules specific to these detector types. The details of the operation of these detectors are beyond the scope of this article. Further information and discussion can be found in the list of references.[12]

6.7 Carrier-gas Pressure/Flow Control

The control of the carrier-gas flow rate is of critical importance to PGC to obtain repeatability of peak areas and retention time for the components being measured. Either pressure regulators or flow controllers may be used for flow control, but for PGC pressure regulation is preferred. Pressure regulation can provide the necessary column flow rate when there are downstream stream leaks in the system. Ambient temperature changes are not as critical with pressure regulation. The most important reason for selecting pressure control is the use of valve and column switching to perform the separation. A pressure regulator has a much faster response to downstream upsets than is possible with a flow controller, which is dependent on its flow setpoint. The selection of a pressure regulator for a process chromatograph includes parameters such as operating pressure range, limitations of amount of flow it will pass, cost, ambient temperature, and barometric pressure changes. The most critical of the parameters for automatic, continuous GC operation are the effects of ambient temperature and barometric pressure.

For a pressure drop across a column of approximately 5 bar, the error in peak area will be 1.5 times the relative error in pressure drop. To limit the error to within 0.5% the pressure control must be within 0.03 bar. If the same regulator is referenced to ambient pressure, the error resulting from barometric pressure variation will be 1.4 times the variation. As can be seen, variations in ambient temperature and pressure can cause 0.2–0.5% errors in the results.[19]

The recent introduction of electronic (microprocessor) pressure control corrects for many of the variations.

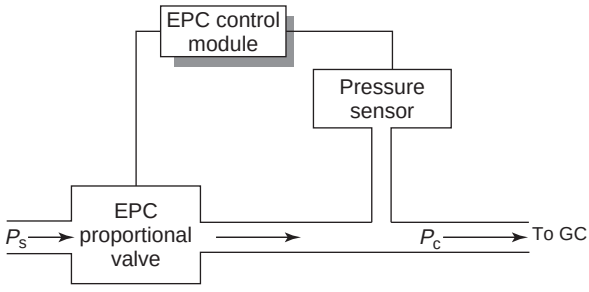

Figure 15 Schematic diagram of electronic pressure control.

In addition, electronic pressure control allows chromatographers to make use of pressure as a variable to improve separations. Pressure programming can be used to reduce cycle times and pressure-switching techniques can be employed to simplify GC multicolumn, multivalve switching configurations. The proportional valves and transducers in the electronic pressure-control system are located in a temperature-controlled housing which isolates the temperature-sensitive components from the ambient environment. This feature eliminates or greatly reduces the ambient temperature effects normally present with pressure regulations. As a result, the repeatability and reliability of the pressure/flow control are greatly enhanced.

With the electronic pressure-control system, pressures are set directly through the microprocessor controller keypad or remotely from the analyzer data control system. The actual pressure readings can be viewed through the microprocessor controller front panel interface. The electronic pressure control package includes control electronics, a manifold block housing, multiple proportional control valves, and pressure sensors. Each pressure sensor is located downstream of the proportional valve (Figure 15). If a downstream pressure change is sensed by the pressure sensor, a signal is transmitted to the control electronics which correspondingly sends a signal to the proportional valve to raise or lower the pressure and corrects for the pressure change at the transducer. The feedback control scheme occurs at a very rapid rate, resulting in very repeatable and smooth pressure control.

This introduction of electronic pressure control provides increased reliability and precision from the PGC system. In addition, microprocessor control of flow settings allows fast and easy implementation of this otherwise tedious process. Pressure programming, which is possible with this type on pressure control, complements traditional temperature programming which will result in even faster analysis cycle times. The true benefits of this capability for PGC applications will continue to evolve as the need for faster and more complex applications arise.

7 APPLICATION ENGINEERING

Unlike its laboratory counterpart, most process gas chromatographs are designed to perform one specific application. In a few situations, multiple streams will be measured with a single chromatograph, but the stream composition must be similar in nature for this to be possible (chemical components and concentration). The first step is to review the sample stream composition. This includes the temperatures, pressures, measured components, concentrations, and the nonmeasured stream matrix. Based on this review, the stream will be sampled as a liquid or a gas. Next, the column and valve configuration is determined to allow the measured components to be separated and measured in the minimum analysis cycle time. A detector and carrier-gas must be selected to provide the required sensitivity and repeatability.

The analysis cycle time is critical for control of the process from which the sample is taken, and is critical to the column/valve configuration selected. There is often more than one configuration possible for a particular application and things to consider other than cycle time include hardware requirements (valves, columns, and detectors), cost, and simplicity. It is very important to the reliability of the measurement to keep the application as simple as possible. This will result in increased reliability and ease of maintenance.

Once these choices have been made the valves and detectors necessary for the application are installed in the analyzer. Next, the columns, flow rates, and temperatures necessary to provide separation of the components of interest are determined. It may be necessary to adjust these parameters to obtain the optimum repeatability of the method. Column selection requires experience and understanding of the basic parameters such as temperature, flow rate, length, diameter, liquid phase, and phase loading which affect separation. The columns and operating conditions are normally provided as a part of the total chromatographic solution from sampling to data reporting by the process gas chromatograph vendor. In the case of a reapplication it may be necessary to go through a great deal of experimentation, refer to references in the literature, or, probably the best approach, consult with someone who has experience with the particular application, in order to select the best columns for a specific analysis.

The most recent innovation in application engineering involves the utilization of more than one detector to perform a specific analysis. This approach allows components with a wide range of concentrations to be measured, provides faster cycle times, and eliminates or reduces interferences from other components. In the author's opinion, this approach should not be utilized for the sole purpose of reducing the number of analyzers necessary for reliable process control. The combining of applications within a single analyzer often results in a complexity of hardware which renders the analyzer unreliable and difficult, if not impossible, to maintain.

When the analyzer has been configured, a sample representative of the stream to be measured is obtained and the analyzer is evaluated for sensitivity and repeatability. The analyzer is calibrated using a calibration mixture with a composition representative of the stream to be measured. With the advent of microprocessor controllers, this process is as simple as entering the concentration of each component in the mixture. The controller, in turn, calculates a response factor for each component. The configured analyzer is ready to be put on-line, automatically taking samples, separating the components, measuring the concentrations, and outputting the results needed for process control.

8 CALIBRATION REQUIREMENTS

The calibration methods employed for PGC utilize external standards, with a comparison with a known concentration of components in a prepared mixture, comparison with a laboratory analysis grab sample, benchmark, and internal normalization.

The most commonly used technique is an external standard where a cylinder of calibration gas or liquid is periodically analyzed by the chromatograph. The concentration of each component is entered in the microprocessor controller and the response factors are automatically determined. The response factors are subsequently used to calculate the concentration of the components measured in the process stream. Peak areas are multiplied by the specific response factors for each component. A less expensive approach is to use grab samples which are taken from the process stream and analyzed by the plant laboratory. Often large quantities of a process sample are taken, analyzed, and used for "benchmark" calibrations. This is an inexpensive alternative to buying expensive, commercially blended calibration standards.

This technique is often used for simulated distillation samples which involve a large number of components which are difficult and expensive to blend. Internal normalization is potentially the most accurate method, but this requires the measurement of all components in the sample and a microprocessor controller. The limitation is the cycle time required to measure and separate all the components in the stream.

A further requirement of a calibration standard is that it be stable over a long period of time. Reactive components or components which absorb strongly on cylinder walls result in erroneous calibrations over time.

Three approaches to overcoming this problem of reactivity of standards involve permeation systems, calibrations with response factors, and surrogate samples. Permeation systems utilize permeation tubes which are liquefied samples contained in a permeable membrane. The permeation rate of these tubes is dependent on the vapor pressure of the liquid, temperature, and permeability of the membrane used. The permeated sample is diluted to the concentration required using inert gas and may be further diluted with air to attain parts per million and billion levels. The second technique involves calibration of a single nonreactive component and the use of response factors for a specific component with a specific detector type obtained from the literature. Although this technique introduces some error, in some cases it is the only option available. The results from the one component used for calibration are used to adjust the literature response factors. Finally, if the components are toxic or very reactive, it is often necessary to find a surrogate sample that has similar chromatographic characteristics to use for calibration. An example of such a sample may be halogen-containing compounds or chemical agents used in chemical warfare.

Regardless of the technique used, calibration is critical to the accuracy of the chromatographic application. The precision of the analyzer is dependent on repeatability, sensitivity, and linearity, but to be truly accurate the analyzer must be properly calibrated. This is a long-term requirement and the responsibility of the user. The frequency of calibration is another commonly asked question and is dependent on the application. The use of a microprocessor controller and automatic calibration with tolerances established for each response factor help to solve this problem. Using these features inherent to the modern GC system the response factors are automatically updated if they change by an operator selected percentage of the previous values.

9 SAMPLE HANDLING AND CONDITIONING SYSTEM

The sampling system employed is the major factor in determining the success or failure of most PGC applications. It should, therefore, be designed as an integral part of the total analyzer and not as an accessory. The sampling system should contain the components necessary to condition and maintain a constant flow of sample into the analyzer. This system may include pressure reducers, driers, filters, vaporizers, cooling units, flow controls, and sample stream-switching valves for multiple stream applications. There are three major aspects of the sampling system that must be considered.

1. *Sample point location*
   The sample point should be located so that the sample is representative of the process stream, but the location should also minimize the complexity and time delay of the sample system.
2. *Sample conditioning*
   In the simple case of a process stream with positive pressure, the sampling system may consist of no more than a filter and pressure reducer. A flow meter for setting up and checking the system is desirable. In many cases additional hardware will be required for heating or cooling and drying the sample, for the removal of mists and droplets from gases or of entrained vapor from liquid samples, and for scrubbing out corrosive or unwanted substances such as particulates. In addition, sample pressure may have to be changed using pumps or pressure reducers. Filters, driers, and heaters should be mounted in pairs with appropriate block isolation valves, so they can be serviced without interrupting the sample flow. It should be pointed out that the sampling system must deal with all process stream compositions and conditions, not only with a stream corresponding to ideal operation of the process. In the case of sample streams that require extensive cleanup, the sample conditioning system may resemble a miniature processing plant. An additional requirement of the sampling system is that it should not change the sample in an undesirable manner through reaction or polymerization at elevated temperatures, through either selective evaporation or condensation, or by adsorption.
3. *Dynamic response of the sample system*
   The performance of a sampling system must also be considered in terms of the response time to a step change in concentration at the stream input. Minimum lag time is critical for the high-speed analyses necessary for closed-loop control. To facilitate this, the number of components necessary for sample cleanup should be kept to a minimum and the sample flow rate maximized.[20]

10 INSTALLATION REQUIREMENTS

The pneumatic and electrical requirements for installation for process gas chromatographs are similar to those required for other process analyzers. With certain detectors there are special support gases needed. There may also be sampling lines which must be heated or cooled to maintain temperatures which will allow a representative sample to the analyzer. The analyzer should be located to protect it from the elements, with a minimum being a three-sided shelter.

Although many of the modern process gas chromatographs can perform reliably under a variety of environmental conditions, it is best to install the gas chromatograph in a climate-controlled analyzer shelter. The actual level of control needed depends on the ambient temperature range and the concentration levels being measured. Each process chromatograph manufacturer provides a published temperature range for the analyzer within which the analyzer may be operated and still meet specifications. If the analyzer is measuring trace levels of components in a process it is operating at maximum sensitivity and wide ambient temperature variation will render the analyzer unreliable without adequate ambient temperature control. Most analyzer shelters provided incorporate both heating and cooling systems for precise temperature control. A process gas chromatograph is a complex integrated analytical system and maintenance must be performed in a variety of weather conditions. The necessary maintenance is much more likely to occur if the analyzer is housed in a climate-controlled analyzer shelter. These special installation requirements for a process gas chromatograph result in an installed cost of 3–4 times the cost of the analyzer, but without the necessary attention to installation the analyzer will not provide the accurate, reliable results needed for process control.

11 CONTROLLER/PROGRAMMER

The controller or programmer (the terms are often used interchangeably) is the electronics normally housed on top of the chromatographic analyzer oven section. The primary purpose of the controller/programmer is to act as a sequencing device and control all the functions of the analyzer. These functions include sample injection and column switching, stream switching, autozeroing, peak gating, integration of peak areas, calculation of concentrations, diagnostic functions, and transmission of data, all of which are done on a cyclic basis. Modern chromatographs use microprocessor- or personal computer (PC)-based electronics coupled with various types and sizes of memory located on plug-in circuit boards. For the typical chromatograph the current or voltage from the detector is transmitted to the detector electronics, where it is amplified and fed to the analog-to-digital conversion electronics. The digitized signal is transmitted to the microprocessor, where resident response factors are used to calculate the concentration of each component in engineering units. The concentration values are output to a printer or another computer system.

The serial data output is often provided as redundant outputs to reduce the possibility of failure of one of the data transmission lines. In the event of a power failure, on-board electrical erasable programmable read-only memory is used to save the latest version of the program used in the application. When power is restored the system automatically reloads the program for normal operation. Other forms of memory backup such as batteries and magnetic tapes have been used to reload the program.

Other features of the control electronics include hardware and software diagnostics and indicating lights on the front panel to signal a problem. Recently introduced capabilities include the settings of temperatures, readout of temperature setpoints, pressure settings, and readout of pressure setpoints. The ability to set and control these two critical parameters through the microprocessor greatly enhances the reliability and precision of the analyzer. Basic calculations using the concentration values in combination with other analog signals are possible in the microprocessor-based controller/programmer. These calculations can be done in real time using an operator-entered program to calculate a desired process variable. An example is the heating value per unit volume of gas for a gas-processing plant. The gas flow rate would be an auxiliary analog input from a gas flow meter.

Finally, the validation of the data is performed in the microprocessor controller/programmer. Based on the results of the analysis such as total peak area, response factor changes, and retention time shifts, validation codes can be assigned to the data prior to transmission to a control computer to allow a decision process to be made on the usefulness of the data for process control.

The front panel of the programmer controller has an operator–machine interface which consists of a keypad and monitoring device such as a cathode-ray tube or liquid-crystal display. This allows the operator to enter and modify the program for a specific application. The monitor is also very useful for diagnostic purposes such as display of alarms and readout of variables such as temperature and pressure. Although a chromatogram is usually available on the monitor, it is a digital version and lacks resolution for troubleshooting purposes when there is a valve, column, or detector problem. Problems with these devices can be analyzed using the recorder output (located on the front panel) connected to a strip-chart recorder. Note: many plant environments would require the maintenance person to obtain a hot work permit prior to connection of such a recorder.

12 DATA SYSTEMS

The modern data system consists of microprocessor- or PC-based computers used to multiplex several process gas chromatographs into single or multiple distributed control system (DCS) computers. Earlier versions served

a dual function of providing data to a DCS and also a maintenance function for the technician responsible for the interconnected analyzers. More modern systems have split the two functions, sending the concentration data, analyzer status, and validation codes to the control room where the DCS is located for use by the process or plant operator. Simultaneously, maintenance and analyzer management data are sent to a separate PC for use by the analyzer technician for troubleshooting and routine maintenance. This PC can be located in the analyzer workshop or on an engineer's desk or be connected through a modem for remote operation. From any of these locations, the program and operating parameters for each of the analyzers can be monitored and modified.

Redundancy within the data system is an important consideration to prevent a single failure in one of the multiplexing components from causing disruption of the entire analyzer system. For this reason, most systems allow redundant hardware at each critical point.

As these systems become more and more complex with redundant hardware and the number of analyzers interconnected continues to grow, the speed of the microprocessor or PC and the cabling for connection are important considerations. The evolution in speed of the microprocessor or PC will continue with time, allowing more and more analyzers with additional features and functions. As these speeds increase, the use of wiring as we know it today will give way to intrinsically safe fiber-optic cables which can transmit signals at the speed of light. Fiber optics are already being incorporated in modern plants and, as the older plants are upgraded with new analyzers and data systems, an expensive portion of the upgrade will be fiber-optic cabling.

Future enhancements in this area will include the use of fieldbus technology. This will allow the various types and vendors of process analyzers to connect together on a common bus to the central process control computer. For complex analyzers such as chromatographs, this seems like a remote possibility, but those who follow the technology are already giving consideration to the idea in their next-generation designs.

13 OTHER CHROMATOGRAPHIC TECHNIQUES

There are other techniques which have been developed to separate and measure components of process streams. Each offers unique applications or sample types which make them complementary to PGC. A limitation of a process gas chromatograph is that the sample must be volatile, that is, we must be able to convert the sample into the gas phase. Any substance, organic or inorganic, which exhibits a vapor pressure of at least 60 torr (8 kPa) GC can be eluted from a gas chromatograph column.[14]

There are many samples which cannot be vaporized because of the temperature levels required, and many that fragment or decompose when raised to a temperature sufficient to vaporize them. These include compounds of higher molecular weight, inorganics, complex molecules, and thermally labile compounds. For these sample types alternative chromatographic techniques may be used. Two of these techniques, LC and supercritical fluid chromatography, will be presented.

13.1 Liquid Chromatography

While only a small fraction of the known organic molecules can be separated using GC techniques (because of volatility limitations), virtually all the known organic molecules could be separated with LC techniques. However, the application and reliability of the technique have been limited because of the complexity of the hardware required. To understand these limitations better, we need to look at differences between GC and LC.

The principles behind the two separation techniques are basically the same, the major difference being that in LC the sample is dissolved in the liquid mobile phase, whereas in GC the sample is carried through the system with a gaseous mobile phase. In both, the separation occurs owing to the solubility or absorption of the sample on the stationary phase contained in the column.

An LC separation occurs by taking a sample with a sample valve (rated for high-pressures). The sample is injected into the mobile phase where it is dissolved and is moved through the column with a high-pressure pump. As the sample is carried through the column, it interacts with the stationary phase in the column and is separated into its various components. The greater the interaction by a specific component, the longer it takes to move through the column. As the components elute from the column, they are measured as they flow through the detector. The peak areas are used to calculate the concentrations of the components and the resulting data are used to control the process being monitored.

Several types of high-pressure pumps have been developed to drive the liquid mobile phase through the column system to the detector. These include syringe-type, pneumatic amplifiers, diaphragm pumps, and positive-displacement piston pumps, all capable of pressurizing to liquid mobile phase to several thousand pounds per square inch. The detectors most commonly used for process chromatography are ultraviolet and refractive index detectors.

The ultraviolet detector has a source of radiation, a method of selecting the specific wavelength, and a

detector for measuring the amount of energy transmitted through the sample stream. This detector type is capable of measurements in the parts per million and billion range. The wavelength selection is performed by either a monochromator or an interference filter.

The refractive index detector measures the change in refractive index that results from the sample dissolved in the mobile phase. It is less sensitive than the ultraviolet detector, but is more universal. The effects of temperature and mobile phase changes in composition limit the applications possible with this type of detector.

The complex design and reliability of the pumps and detectors have limited the process applications for LC. They are advantageous over gas chromatographs for applications involving samples with a very high boiling point or samples that decompose when vaporized. For these applications, the technology has been widely used in pharmaceutical and polymer industries.[21]

The limited application by other industries has led to the discontinuation of all commercially available process LC analyzers. This leaves the potential users of process liquid chromatographs only one choice, to adapt a laboratory system for the particular process application. When considering such adaptation, one must keep in mind environmental and safety constraints, automatic, continuous operation, and the long-term maintenance requirements of such systems.

13.2 Supercritical Fluid Chromatography

The application range of process gas chromatographs is somewhat restricted by the limited volatility and thermal stability of many organic compounds. Mixtures of less volatile compounds can be analyzed by LC (described in the previous section) or high-temperature GC, but these techniques have not found a lot of acceptance as process analyzers.

Another form of chromatography, using supercritical fluids, gained wide popularity in the laboratory during the 1980s. Supercritical fluid chromatography combines several features of GC and LC, which make it an ideal technique for analyzing samples of low volatility and thermal stability. In addition, this technique has been shown to be easily adaptable to ''on-line'' process chromatography.[22]

Supercritical fluid chromatography is a chromatographic technique that uses supercritical fluids as the mobile phase. When a gas is heated and compressed above its critical temperature and pressure it exhibits properties that are intermediate between those of gases and liquids. The most important property is that the density varies as a function of the temperature and pressure; as the pressure increases, the density increases, increasing the solvent strength. Consequently, by varying the pressure, the solvent strength can be adjusted, thus changing the partitioning of the sample between the stationary phase and the supercritical fluid mobile phase. Other properties related to the density include diffusion coefficients less than that of a gas, but greater than that of a liquid, and viscosities similar to those of gases, but much lower than those of liquids.

Taken together, these characteristics lead to a number of differences with respect to GC and LC. For example, because the diffusion coefficient is greater in a supercritical fluid than a liquid, one can achieve higher resolution and faster analysis times for the same conditions when using a supercritical fluid rather than a liquid mobile phase. Viscosity also plays a major role in the transport properties of the supercritical fluid through a column. Since the viscosity of these fluids is similar to that of a gas, the pressure drop across the column is low. This allows the use of capillary columns, which helps to expand the analytical capability of this technique.

There are several supercritical fluids available, but carbon dioxide is most commonly used because of its low critical temperature and pressure. Carbon dioxide has an intermediate polarity, making it the ideal solvent for a wide range of samples. It is nontoxic and has no response on an FID.

There are several key hardware components that make it uniquely different from other process chromatographs. The pump is a pneumatic amplifier, which provides the pressure to keep the system supercritical and generates a programmed pressure ramp to elute the components of interest in the sample matrix. The pump is controlled by a microprocessor and is capable of delivering pressures from 1000 to 7000 psig.

Packed or capillary columns may be used for a separation. For packed columns, 1 mm internal diameters are used and for capillary columns 0.1 mm internal diameters or smaller are common. Because of limited sample capacities, direct injection of sample is not recommended owing to overloading of the column. For GC applications this problem can be overcome by using a splitter. The use of splitters for supercritical fluids has been found to be very unreliable and nonreproducible.[22]

To avoid sample splitters, the process supercritical fluid chromatograph uses a sample diluter, wherein a microliter-size sample is injected into a chamber filled with supercritical fluid. The sample is dissolved and mixed in the chamber and a fixed volume of the diluted sample can be injected on to the column, thus eliminating the need for a splitter.

The final critical component is the column detector interface. An FID requires the column effluent to be depressurized to ambient pressure prior to detection. This is normally accomplished by using a capillary restrictor.

The major application area for the supercritical fluid chromatograph is the petroleum industry. The mild conditions of this technique make it ideal for the analysis of high-boiling petroleum fractions. The samples are dissolved in the supercritical fluid and separated at temperatures of less than 100 °C. This eliminates the fractionation which often results when these samples are vaporized.

13.3 Preparative-scale Chromatography

Often in the literature, preparative scale is referred to as process chromatography. This technique involves a large-scale chromatographic system that uses wide-bore columns of 10–20 cm internal diameter to isolate milliliter volumes of a component from the balance of the sample. This large liquid chromatograph is very similar in operation to a pilot plant used to simulate a process. Although the term ''process'' chromatograph is used to define this analyzer, it is not considered to be a part of the family of chromatographic techniques used to measure and monitor sample streams for process control.

14 COMMON APPLICATIONS

The uses of process gas chromatographs in industry are limited only by the imagination of the process control engineer. This analytical tool is suitable for any application where a sample can be introduced in the gas phase, the components of interest can be suitably separated on a chromatographic column, and reliably measured by a compatible detector.

Some of the more common uses include the following:[12]

- Process control: use information to adjust the process with open- or closed-loop control.
- Process study: obtain information about the process to improve yield or throughout; correlate process variables with product quality.
- Process development: obtain information about the process characteristics as in pilot plants; correlate process variables with reaction products and yields.
- Material balance: use information to calculate material balance for the process unit.
- Product quality specification monitor: monitor impurities in incoming or outgoing product for conformance to specifications.
- Waste disposal monitoring: monitor liquid or gas effluent wastes for loss of valuable product or for presence of toxic compounds.
- Personnel safety–air monitoring: monitor ambient air for the presence of toxic compounds.

Of course, the most common of these uses is the use of information for open- or closed-loop control. In open-loop control, the operator uses the results to make adjustments to the process. In closed-loop control, the results are input to the central process control computer which, in turn, automatically makes adjustments to the process.

Before selecting a process gas chromatograph for on-line analysis in any industry, two important questions must be answered. First, will the analyzer provide information that will allow the plant to save money or provide increased safety of personnel or protect the environment? Savings can come from many sources such as increasing the throughput of a chemical product, reduce the losses of valuable raw materials, upgrade the value of a product, or conserve energy. If one cannot identify an economic benefit, the installation and maintenance of a process gas chromatograph may not be the right decision. Some chromatographs must be installed without direct economic justification for the protection of workers from toxic emissions and to protect the environment in the plant and the surrounding community.

Second, is there a better, less expensive technique to measure the components of a process stream? As technology evolves, alternative instruments such as mass spectrometers, spectrometric analyzers, and electrochemical analyzers are being introduced for process applications on a continuing basis. The benefits of chromatography include the ability to separate the components of interest from complex mixtures and to obtain quantitative and qualitative results at a reasonable cost. The gas chromatograph is also the analyzer of choice when multiple streams and multiple components must be measured in a wide concentration range. Many analytical techniques look promising when applied to binary mixtures, but they fail miserably when installed in real-world process applications.

Considering the more than 30 000 process gas chromatographic analyzers installed since the late-1950s which continue to operate in plants today, one finds very few references describing these applications. Much of this is due to the proprietary nature of the process applications, and another reason is that there are many classes of applications where there are only slight variations.

14.1 Petrochemical/Chemical

Petrochemicals include chemicals or chemical products derived from hydrocarbon raw materials, primarily crude oil and natural gas. They account for the majority of the production of chemical manufacturing industry. Catalysts, thermal cracking, fluid flow, heat transfer, mass transfer technology, equipment, and materials are used to convert hydrocarbon raw materials into primary petrochemicals.

These are used as feedstock for further processing into intermediates or petrochemical products.

The volume plastics business is a good example of petrochemical[23] products produced by this industry. Plastics including several types of polyethylene (PE), polypropylene (PP), and polystyrene (PS) account for approximately 70% of all petrochemicals because of their widespread use in packaging, automotive, construction, electronics, and other markets. Ethylene and propylene plants utilize numerous process gas chromatographs (approximately 60) to control the manufacturing processes. A typical application that utilizes a process gas chromatograph is the ethylene cracking process.[24]

14.1.1 Cracking

Low-molecular-weight feedstock (ethane, propane).

Measurement: furnace inlet, C_1, C_2, C_3, C_4 compounds; furnace outlet, C_2, C_2, and C_3; and ratios of $C_1 : C_2$ and $C_1 : C_3$.

Analysis time: 10 min.

Detector: TCD.

Purpose: furnace inlet. Determine the quality of feedstock to predict yield.

Furnace outlet: the correlating parameters used within industry are severity and selectivity. Severity measures the extent of the pyrolysis (also known as conversion) which can be determined by the methane (C_1)/propylene (C_3) ratio. Selectivity is a measurement of time–temperature history. Coke deposition in the heater coils and effluent exchanges raise the pressure in the coil, favoring the production of methane and liquid products rather than light olefins. The methane (C_1)/ethylene (C_2) ratio is used for correlating selectivity. A gas chromatograph can be used to measure the three compounds, the ratios calculated, and the entire yield structure predicted.[25]

Benefit: feed-forward control of the furnace.

Typical furnace feed and furnace effluent composition: see Table 1.

Table 1 Typical furnace feed and furnace effluent composition

Ethane/propane feed		Furnace effluent	
Component	Concentration (%)	Component	Concentration (%)
Methane	1.0	Hydrogen	30.0
Ethane	69.0	Methane	5.0
Propane	27.0	Ethane	25.0
Isobutane	2.0	Ethylene	35.0
n-Butane	1.0	1,2-Propadiene	1.0
C_{5+}	1.0	Propane	1.0
		Propylene	1.0
		Acetylene	1.0
		Methylacetylene	1.0

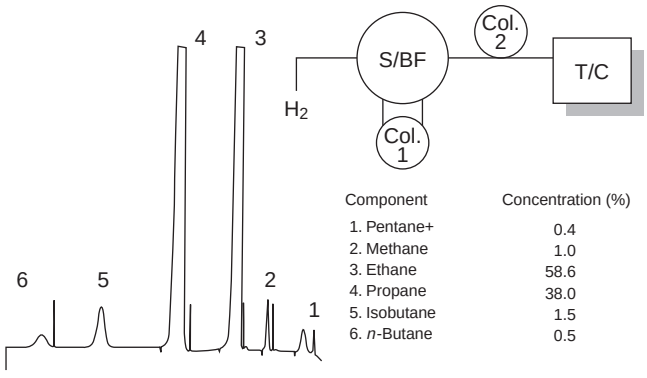

Figure 16 Ethylene furnace feed chromatogram.

The chromatographic separation of the ethylene furnace feed (ethane/propane) is shown in Figure 16. The column–valve configuration utilizes a backflush to measure valve to group pentane plus heavier hydrocarbons followed by the separation of methane, ethane, propane, isobutane and *n*-butane.

The chromatographic separation of ethylene furnace effluent is shown in Figure 17. The column–valve configuration utilizes two sample/backflush valves operated in series and four separation columns.

Other common chemical/petrochemical applications include alcohol production, distillation control, reactor product analysis, ammonia production, vinyl chloride production, aromatics production, chlorine purity, PP production, methyl chloride production, styrene production, methanol, and naphthalenes production.

14.2 Refining

Today society around the world is dependent on motor vehicles as a means of transportation of people and

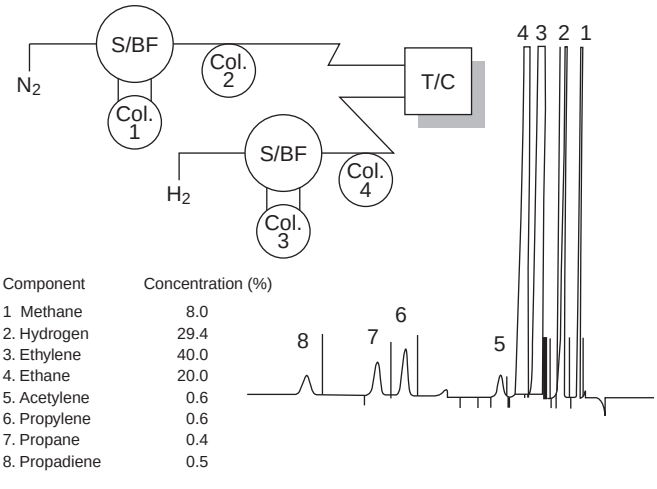

Figure 17 Furnace effluent chromatogram.

products. The demand for fuels to power these motor vehicles continues to grow as our populations and economies increase. Supplying these fuels has grown into one of the largest industries in the world, involving exploration for crude oil, refining, marketing, and distribution of the finished product.

The modern refinery is a complex network of pipes, vessels, and reactors, but if one breaks it down into individual units, it is really fairly simple. The initial step is to take crude oil and separate out the natural gasoline present. Next, the refinery takes the remaining crude and uses various reactions to make more gasoline. For example, the alkylation unit in a refinery is used to make octane from the reaction of two molecules of butane. Other reactors are used to break apart molecules too big to be used in gasoline by catalytic cracking or thermal decomposition in a coker. These reactors must be monitored and controlled, requiring analytical measurements which are easily provided using process gas chromatographs. A typical application utilizing a process gas chromatograph is the crude distillation process.

14.2.1 Crude Distillation

Crude oil.

Measurement: Various fractions including fuel gas (C_1, C_2, C_3), wet gas (C_2, C_3, C_4), light straight-run gasoline (C_5, C_6), heavy straight-run gasoline (C_7–C_{10}), kerosene (C_9–C_{15}), diesel (C_{13}–C_{18}), gas oils (C_{13}–C_{45}), and residuum (C_{40+}).

Method: temperature-programmed PGC, simulated distillation.

Analysis time: 15–30 min.

Detector: FID.

Purpose: the primary purpose of the process gas chromatograph is to analyze the fractions of the crude distillation tower for boiling-point distribution using simulated distillation. Economic incentives exist to upgrade from one fraction to the next by operating as close as possible to the boiling point specifications. As an example, gasoline is more valuable than kerosene, which dictates separating the maximum amount of gasoline possible prior to distilling off the kerosene fraction. Recent environmental mandates for reformulated gasolines also limit the volume of higher boiling components because of their inefficient combustion, leading to increased pollution.

Benefit: improved yields.

Typical measurement range: C_4–C_{12} hydrocarbons.

The chromatographic separation of a gasoline sample using temperature programming is shown in Figure 18. A liquid sample injection valve is used to inject a microliter-size sample into a capillary, nonpolar column. The capillary column is programmed from an initial temperature of 60 °C to a final temperature of 270 °C at

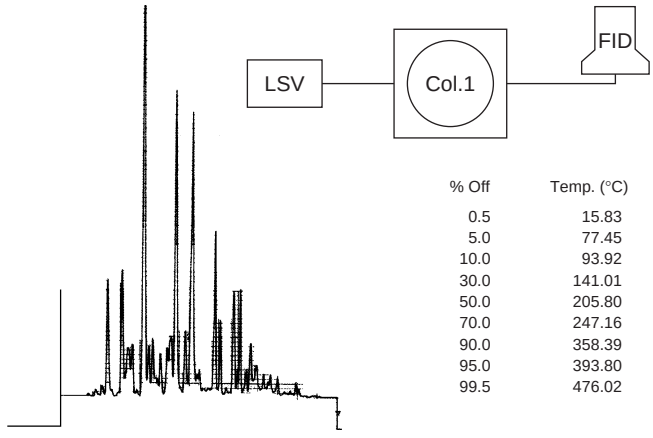

Figure 18 Simulated distillation chromatogram for gasoline.

30 °C min^{-1}. The type of separation is commonly referred to as simulated distillation, separating hydrocarbons in the range from butane to dodecane and higher. The results can be used to calculate the boiling-point distributions of various hydrocarbon fractions.

Other common applications for the refining industry include catalytic reformer unit, aromatics fractionation unit, fluidized bed catalytic cracker unit, HF alkylation unit, sulfuric acid alkylation unit, pentane isomerization unit, butane isomerization unit, hydrotreater, and gasoline blending.

14.3 Natural Gas Industry

The natural gas industry provides a valuable source of energy that is used in homes and industries throughout the world. This industry is composed of four major functions: gathering, processing, transmission, and distribution. First, the gas wells are drilled and the gas is extracted and collected for gas-processing. In the gas-processing plant, the heavier hydrocarbons are removed and methane-rich gas is placed in a pipeline for transport. The heavy fraction which was removed is separated by a natural gas liquids (NGL) plant to produce pure chemical compounds for use by the petrochemical industry.

The gas transmission companies buy the methane-rich natural gas and transport it through an extensive network of pipelines to the gas distribution companies. Finally, the gas distribution companies buy the gas and distribute it to the residential or commercial user.

As the natural gas moves through the various functions, ownership of the gas changes and it is important for the new owner to be able to analyze the purchased gas for heating value, specific gravity or density, and composition. A process gas chromatograph is ideally suited to provide these three measurements. The gas is normally purchased on the heating value per unit volume of flow and as the gas

moves through the various stages the heating value will change as a result of removal of components, dilution, and concentration. Two common contaminants are air and moisture, which have no heating value but change the volume of flow. There are some special requirements for a process gas chromatograph used in these applications, such as limited utilities, simplicity of design, speed of analysis, and reliable operation. A typical application in this industry is the measurement of the heating value, specific gravity, and composition of the natural gas stream.

14.3.1 Heating Value

Natural gas pipeline.

Measurement: N_2, CO_2, C_1, C_2, C_3, i-C_4, n-C_4, i-C_5, n-C_5, C_{6+}; other components such as heavier hydrocarbons, neopentane, and H_2S are sometimes measured.

Method: PGC.

Analysis time: 2–10 min.

Detector: TCD.

Purpose: this measurement is required on custody transfer points where ownership of the gas changes hands. In the past natural gas was sold by volume with a guaranteed heating value. More recently, as the price of gas increased, the need to measure the heating value became more important. A typical pipeline transports large volumes of gas and a small change in heating value can significantly change the price paid for the gas. Using the concentration of each component measured multiplied by the ''ideal'' heating value for each component, the microprocessor integral with the gas chromatograph calculates the heating value of the sample. Because the natural gas is not an ''ideal'' gas, a correction must be made for compressibility; all of these corrections are provided by the microprocessor controller.

Benefit: precise measurement of heating value for custody transfer.

A typical natural gas stream composition is given in Table 2.

The chromatographic separation of a natural gas stream is shown in Figure 19. The column–valve configuration utilizes three valves and three columns. The first valve is used as a backflush to measure valve to group the

Table 2 Typical natural gas stream composition

Component	Concentration (%)	Component	Concentration (%)
Nitrogen	3	Isobutane	0.5
Carbon dioxide	2	Butane	0.5
Methane	85	Isopentane	0.25
Ethane	5	Pentane	0.25
Propane	2.5	Hexane+	1.0

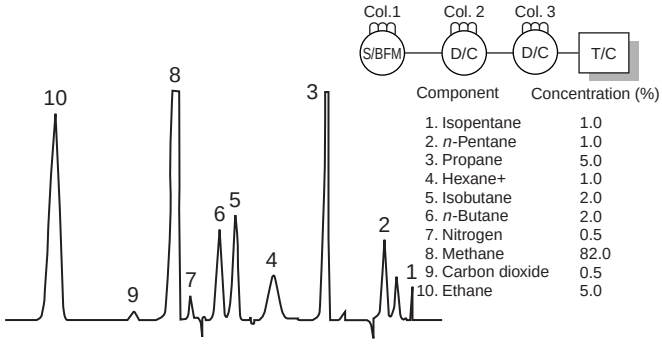

Figure 19 Chromatogram of natural gas stream.

pentane plus heavier hydrocarbons. The second and third valves function as dual column valves to provide enhanced separation of the lighter fractions. Note that the order of elution is not what one would normally expect because of the multiple valve switches.

Other common applications in the natural gas industry include absorption gas plant, expander gas plant, and NGL plant.

14.4 Environmental Monitoring

The application of process gas chromatographs to environmental monitoring or ambient air monitoring is a result of a desire to provide a safer working environment for employees of various industries and to reduce emission into the local environment. Many of the chemicals used in the plants are toxic, explosive, or damaging to the environment in the event of a leak or process upset. Area monitoring allows an early warning of such conditions so that remedial action can be taken prior to disaster. Stack or incinerator monitoring, often dictated by governments to ensure minimum emission of hazardous materials, is controlled to minimum standards.

The limits on emissions of hazardous compounds are established by groups such as the Environmental Protection Agency (EPA), Occupational Safety and Health Administration (OSHA), and local, state and federal governments. Such standards are often set for both air and water pollution. A monitoring program is normally established for hazardous compounds that may be present in the plant environment at regular intervals or on a continuous basis.

Because of the sensitivity of the detectors available for the process gas chromatograph, it is ideal for monitoring plant environments where the concentration levels are in the parts per billion range. In addition, because a chromatograph can be used to measure multiple streams, a single analyzer can monitor 20–30 sample points around the plant. The critical issue relative to the number of sample points is the analysis time of each point. For this

reason, the typical chromatograph used in this type of application only measures one component or, at most, two or three. The system is designed to trigger an alarm if the safe limit for the component being measured is exceeded. Another requirement often provided as a part of such systems is the long-term storage of data for later verification.

The reliability of the analyzer is very important since frequent maintenance would allow a plant to become unsafe without the knowledge of the workers. Analyzers used for environmental monitoring are often maintained by people within the plant who are not familiar with process gas chromatographs, such as the plant safety department. For this reason, the analyzers should be simple to operate and easy to maintain and repair.

14.4.1 Ambient Air Monitoring

Hazardous chemical compounds.

Measurement: trace levels of vinyl chloride, benzene, acrylonitrile, phosgene, hydrogen cyanide.

Method: PGC.

Analysis time: 1–2 min.

Detector: FID, other specific detectors.

Purpose: to protect workers and environment from hazardous compounds on a routine or continuous basis. The process gas chromatograph is usually designed to monitor one or two hazardous compounds in ambient air. If the concentration exceeds specified limits, the analyzer will trigger an alarm. A second alarm at higher specified limits may be used to trigger an evacuation of the area. The analyzer keeps average values for components measured over an 8 h period, on a daily, weekly, and monthly basis, with reports logged at the end of these periods. In addition to providing these reports, the analyzer should be set-up to calibrate itself automatically on a regular basis. This will ensure that the analyzer is always working properly.

Benefit: protection of workers and the environment.

The chromatographic separation of hexane, benzene, toluene, and ethylbenzene in ambient air is shown in Figure 20. A large sample loop (10 mL) on a sample backflush to vent valve is used to inject the sample. Sample concentrations of a few parts per million can be measured with an FID.

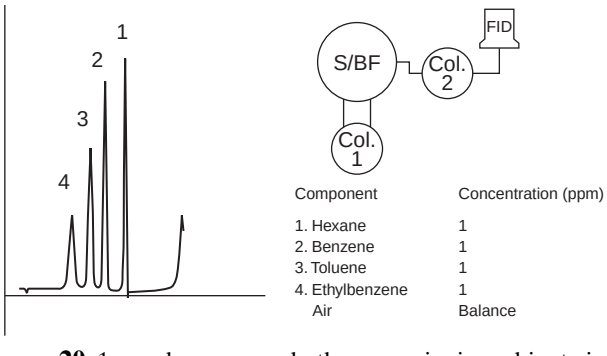

Figure 20 1 ppm benzene and other organics in ambient air.

Other common applications for environmental monitoring include industrial wastewater monitoring, incinerator stack gas monitoring, and fenceline air monitors.

14.5 Other Industrial Applications

Other industrial applications include the following:

- Biotechnology:
  - fermentation off-gas monitoring
- Pharmaceuticals:
  - solvent recovery
  - ethylene oxide sterilization
- Steel industry:
  - blast furnace off-gas analysis
- Food and beverages:
  - alcohol concentration
  - moisture concentration.

ACKNOWLEDGMENTS

The author thanks ABB Process Analytics Inc. for 30 years of employment in the exciting industry of process analyzers. During these many years he has traveled to many interesting places around the world, making presentations on the subject and sharing experiences with countless customers in the hydrocarbon processing industry.

The author also thanks his wife, Linda, for her assistance, encouragement, and patience during the few months in which this article was written. Without her support and understanding, this work would never have been completed.

Finally, the author thanks his secretary, Charlotte, who spent countless hours reading his terrible writing and typing and preparing this article for submission.

ABBREVIATIONS AND ACRONYMS

BTU	British Thermal Units
DCS	Distributed Control System
EPA	Environmental Protection Agency
FID	Flame Ionization Detector
GC	Gas Chromatography
LC	Liquid Chromatography
NGL	Natural Gas Liquids

OSHA	Occupational Safety and Health Administration
PC	Personal Computer
PE	Polyethylene
PGC	Process Gas Chromatography
PP	Polypropylene
PS	Polystyrene
TCD	Thermal Conductivity Detector

RELATED ARTICLES

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers

Pulp and Paper **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Column Theory and Resolution in Liquid Chromatography • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. F.W. Karasek, 'Process Chromatography', *Anal. Instrum. Process Control*, 79–83 (1970).
2. W. Ramsey, *Proc. R. Soc. London, Ser. A*, **76**, 111 (1905).
3. M. Tswett, *Ber. Dtsch. Bot. Ges.*, **24**, 316 (1906).
4. H.H. Strain, J. Sherma, *J. Chem. Educ.*, **44**, 238 (1967).
5. D.T. Day, *Proc. Am. Philos. Soci.*, **36**, 112 (1897).
6. A.J.P. Martin, R.L.M. Synge, *Biochem. J.*, **35**, 1358 (1941).
7. A.T. James, A.J.P. Martin, *Biochem. J.*, **50**, 679 (1952).
8. L.S. Ettre, *Anal. Chem.*, **43**(14), 20A–31A (1971).
9. J.F. Tatera, 'Gas Chromatographic (GC) Analyzers', *Anal. Instrum.*, 619 (1996).
10. Frost and Sullivan Report, 'World Process Analytical Instruments Market', Report 4–28, 1996.
11. R. Villalobos, 'Process Gas Chromatography', *Anal. Chem.*, **47**(11), 983A (1975).
12. J.C. Giddings, *Dynamics of Chromatography*, Marcel Dekker, New York, 1965.
13. J.J. Van Deemter, F.J. Zuiderweg, A. Klinkenberg, *Chem. Eng. Sci.*, **5**, 271 (1956).
14. H.M. McNair, M.W. Ogden, J.L. Hensley, 'Recent Advances in Gas Chromatography', *Am. Lab.*, August, 15 (1985).
15. M.L. Duffy, 'Widebore, Thick-film Capillary Columns in GC', *Am. Lab.*, October, 94 (1985).
16. H.M. McNair, J.M. Miller, 'Detectors', in *Basic Gas Chromatography*, 103, 1997.
17. K.K. Konrad, *Introduction to Process Gas Chromatography*, Bendix Corp, Ronceverte, WV, 21, 1975.
18. R. Annino, R. Villalobos, 'Detectors', in *Process Gas Chromatography*, Instrument Society of America, Triangle Park, NC, 234, 1992.
19. R. Annino, 'Chromatographs–Gas', *Anal. Instrum.*, 100 (1994).
20. J.M. Clemons, 'Process Gas Chromatography', Bendix Corporation Brochure, Ronceverte, WV, 10, 1982.
21. R.E. Cooley, 'Liquid Chromatographic (LC) Analyzers', *Anal. Instrum.*, 673–675 (1996).

22. K. Melda, 'Supercritical Fluid Chromatography', paper presented at the International Forum for Process Analytical Chemistry, 1998.
23. E.J. Swantek, 'Plastics and Resins', in 1998 HPI Market Data, Gulf Publishing Company, Houston, TX, 18, 1998.
24. A. Kania, *On-line Process Chromatograph Applications Notebook*, Applied Automation Application Notebook, Bartlesville, OK, 59–72, 1989.
25. S. Bostic, 'Ethylene Process', ABB Process Analytics Applications Sales Brochure, Lewisburg, WV, E-3, 1998.

Flow and Sequential Injection Analysis Techniques in Process Analysis

Joseph H. Aldstadt III*
University of Wisconsin-Milwaukee, WI, USA
Don C. Olson, Duane K. Wolcott, and Graham D. Marshall
Global FIA, Inc., Gig Harbor, WA, USA
Scott W. Stieg
Zellweger Analytics/Lachat Instruments, Milwaukee, WI, USA

1 Introduction 1
2 Functional Descriptions of Flow Injection Analysis and Sequential Injection Analysis 2
2.1 Flow Injection Analysis 2
2.2 Sequential Injection Analysis 6
3 Instrumentation 7
3.1 Design Considerations 7
3.2 Manifold Hardware 8
3.3 Flow Injection Analysis Software 9
3.4 Method Development 9
3.5 Sampling Systems 10
4 Applications 15
4.1 Chemical (including Petrochemical) Refining 15
4.2 Biotechnology (including Food, Fermentation, Pharmaceutical) 15
4.3 Manufacturing 16
4.4 Environmental 18
5 Strengths and Weaknesses 18
6 Future Directions 19
Acknowledgments 20
Abbreviations and Acronyms 20
Related Articles 21
References 21

The principles of flow injection analysis (FIA) and sequential injection analysis (SIA) are described as applied to industrial process monitoring. A brief history is followed by an overview of the fundamentals of FIA and SIA, with emphasis on injection methods and membrane-assisted sample processing approaches. A discussion of the design of process instruments focuses on simplicity, maintainability, and software development. Hardware issues affecting autosamplers, fittings, pumps, and valves are also discussed. Issues affecting method development, quality control, and method validation are examined using specific examples from the chlor-alkali and phosphate rock industries. The design and operation of on-line sample interfaces is then described in detail. Applications in the major industrial areas are used as examples – chemical, biotechnology, manufacturing, and environmental. Finally, strengths and weaknesses are delineated and prospects for the future are presented.

1 INTRODUCTION

Since the first paper on FIA appeared in 1975,[1] the scope of the technique has grown tremendously, today ranging from routine analyses of wastewater samples in environmental labs to the measurement of the kinetics of drug uptake by single cells. One of the areas where FIA is most successfully applied is the monitoring of streams and samples in the process industries – on-line, at-line, and off-line. The roots of FIA as a process-monitoring technology date back to when the technique was in its infancy. The pioneers of process FIA came from the chemical and petrochemical industries, for example, Amoco, Dow, and Shell, in the late 1970s. The first review article on process FIA was published in 1981,[2] and the first commercial process FIA monitor was introduced in 1983 by FIAtron, Inc. In 1990, a new generation of the technique, SIA, was introduced which offered some important improvements for process analyses.[10]

The more than 9000 papers on FIA published since 1975 are a measure of the extensive activity in the field;[11] however, these publications tell only part of the story. Since the late 1970s, a great deal of research and development on FIA has been conducted in industrial laboratories which, aside from scattered journal papers and presentations at conferences, has never been published. Most of this work has been targeted at the application of FIA to process monitoring, including off-line (i.e. plant laboratory), at-line, and on-line approaches.[2] This article attempts to capture and describe the technology that has been developed by this work, by providing an overview of the application of FIA techniques to process monitoring. The task is made difficult because FIA-based process monitors that are successful are often not described in the open literature by the company for obvious proprietary reasons. Nevertheless, we primarily draw on our experiences and also survey the published research in English

* Corresponding author

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

language journals from the inception of the technique to early 1999. Additionally, we have had discussions with several major industrial users of process FIA and include their valuable input.

It is important to note that a very broad view of the definition of process analysis is taken, ranging from laboratory analysis in plants and refineries to on-line process analysis. The focus is on FIA as a tool for the process analytical chemist versus for the laboratory analytical chemist. Although there is a great deal of overlap, their environment and instrument performance needs differ considerably. The laboratory analytical chemist tends to work with more complex, sophisticated, and delicate instruments, such as ICP/MS (inductively coupled plasma/mass spectrometry), which require a great deal of the attention of an experienced operator. The process analytical chemist needs analytical tools that are simple, robust, have rapid analytical readout, and require very little attention in the often harsh environment of the plant or refinery.

2 FUNCTIONAL DESCRIPTIONS OF FLOW INJECTION ANALYSIS AND SEQUENTIAL INJECTION ANALYSIS

2.1 Flow Injection Analysis

The FIA method is both simple in concept and difficult to define.[12] Because of the inherent versatility of FIA, it is difficult to describe its broad scope by using a single definition. In general, there are two types of definition in use, which might be categorized as the academic and industrial definitions. An example of an academic definition describes FIA as '... information gathering from a concentration gradient formed from an injected, well-defined zone of a fluid, dispersed into a continuous unsegmented stream of a carrier'. A typical industrial definition describes FIA as '... a simple and versatile analytical technology for automating wet chemical analysis, based on the physical and chemical manipulation of a dispersed sample zone formed from the injection of the sample into a flowing carrier stream and detection downstream'. Those most interested in the fundamentals of the FIA experiment will no doubt prefer the academic definition, whereas those interested in the analytical utility of FIA can probably relate better to the industrial definition. It should be noted that FIA has advanced far beyond simply automating existing wet chemical analyses. Entirely new chemical analyses have been created with FIA, many of which are impossible to perform, or not readily performed, by conventional batch techniques.

The FIA technique automatically combines several analytical functions into a method performed in a flowing stream, usually under computer control, in a short period of time (generally 15–60 s), with a minute amount of sample (generally using microliter volumes), with simple and robust hardware, with high precision (typically better than 0.5% relative standard deviation (RSD)), with relatively little waste (of the order of a milliliter per sample), and covering a broad concentration range from ppb (parts per billion) to wt% (weight percent). An FIA method can be generally divided into three stages, all performed in sequence in a flowing carrier stream pumped through narrow bore tubing.

The first stage is *injection*, in which the sample is measured and injected into the flowing carrier stream. This step is usually performed with a two-position injection valve.

The second stage is *sample processing* in which the analyte is transformed into a species that can be measured by the detector and have its concentration manipulated into a range that is compatible with the detector. There are a large number of different processing steps which can be performed on the sample zone as it flows through the FIA system.[13] One common processing step is simple dilution – dilutions up to 1 : 10 000 are possible with high precision. Conversely, enrichment up to a factor of several hundred may be possible in certain cases for trace analysis. For optical detector-based methods in general, some chemistry is performed on the analyte to convert it to a detectable form. One can also do on-line matrix modification and medium exchange. The former refers to moving an analyte from a complex matrix to a simple matrix before detection, whereas the latter refers to moving an analyte from a gaseous sample to a liquid carrier, or vice versa.

The third stage is *detection* where the analyte, or a derivative of it, generates a response that is used for quantitation. A large variety of flow-through detectors are used in FIA, primarily based on ultraviolet–visible optical or electrochemical transducers. It is beyond the scope of this review to describe in detail the various detectors that can be used with FIA. Several excellent monographs are available for this purpose.[14,15]

The value of FIA as an analytical tool lies in its ability to combine these analytical functions in a variety of ways to create a broad range of methodologies, and to perform these methodologies rapidly and automatically with minute (microliter) amounts of sample. The last stage is, to a large extent, based on conventional technology. The first and second stages are the heart of FIA. The following paragraphs look in more detail at these functions, in particular how they are used in a process monitoring environment.

2.1.1 Sample Injection

The device most commonly used to measure and inject the sample into the FIA carrier stream is a two-position six-port commutating sample injection valve, the two positions being commonly termed load and inject. For years, HPLC (high-performance liquid chromatography) valves were used for this purpose, but this is now considered to be unduly expensive because HPLC valves are designed for high pressures, whereas FIA is a low-pressure technique. Low-pressure valves are widely used in FIA work. Important features that valves must have to be suitable for process FIA include high rotary positioning precision, fast switching, pressure limits to ~100 psi, a reliable fitting system, a rugged rotor and stator, and the ability to inject sample volumes from a fraction of a microliter to several thousand microliters.

Recently, a variation of conventional FIA, called carrierless or reverse FIA, was developed by Hach Corporation.[16] In this approach, the sample serves as the carrier stream and the reagent (i.e. for a derivatization reaction) is injected into the sample stream. Obviously, this approach is well-suited to typical process monitoring applications where the quantity of sample is unlimited. The fact that a sample pump is not required is a critical advantage of this type of FIA instrument design.

2.1.2 Sample Processing

Dilution is a very common operation for process FIA monitors. There are several different ways to perform dilution, the most common being:

- dispersion
- electronic dilution
- zone sampling
- gradient chamber
- membrane sampling.

Dilution by dispersion is intrinsic to any FIA experiment. Controlled dispersion is one of the key processes in FIA, which disperses the sample into the carrier, both diluting it and mixing it with reagent. This technique is, without doubt, the simplest and most precise FIA technique for dilution. The smaller the sample, the greater the dilution; therefore, the key to large dilutions is the capability to precisely inject very small samples. The most precise tool for injecting extremely small volumes is a two-position sample injection valve with a fixed internal loop. Volumes as small as 0.2 μL can be injected, creating dilution factors of several hundred.

A disadvantage of the internal loop model of injection valve is that the volume is fixed, so that changing the volume requires changing the valve. A two-position sample injection valve with an external loop allows the volume to be varied conveniently. Generally, the volume is varied by manually changing the length or internal diameter of the external loop. However, valves with micro-electric actuators have the unique capability of timed injection, switching under software control. These valves permit partial-loop variable-volume injection with a fixed sample loop. Volumes from about 1 μL up to the total volume of the loop can be injected reliably.[17]

Electronic dilution is a novel and simple technique for dilution which is also based on controlling dispersion. Electronic dilution simply refers to a technique for measuring the analytical signal along the detector response profile following the maximum where the signal is on-scale, as depicted in Figure 1. Generally, the signal is measured at the peak maximum, depicted as t_1. However, t_2 or t_3, which represent points in the sample zone where the sample is more highly dispersed or diluted, are equally valid points to collect the analytical signal. The timing of the signal measurement must be precisely controlled for this technique to work well. Electronic dilutions up to approximately 100-fold can be achieved. Although this approach is simple, it is probably the least precise of the dilution methods because of errors that are inherent in the measurement of a sloping signal, and errors because of slight shifts in the peak position.

Zone sampling is a relatively simple and powerful technique for dilution. It is similar to the heart-cut technique used in chromatography. It requires two sample injection valves. With the first valve, a relatively large sample is injected that then flows downstream to be dispersed by the reaction coil. As the dispersed sample zone fills the smaller sample loop of the second valve, an aliquot of the zone is injected into a second carrier stream. The second carrier can also pass through a reaction coil for

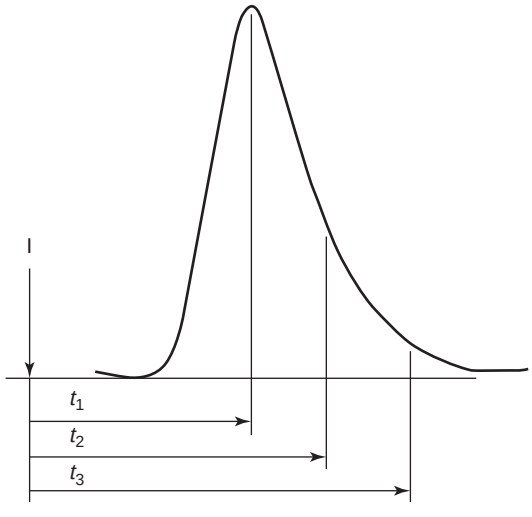

Figure 1 Response profile for the electronic dilution experiment.

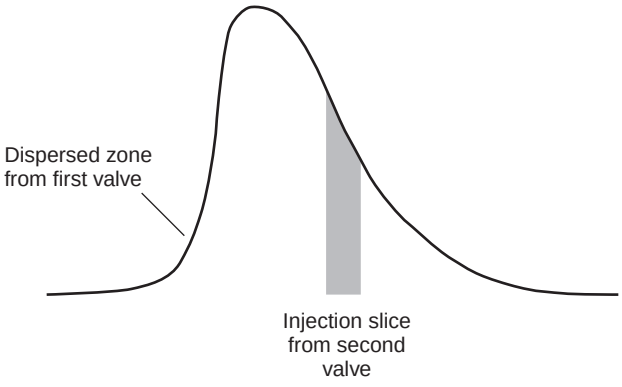

Figure 2 Response profile for the zone sampling experiment.

further dispersion or dilution. Figure 2 depicts how this technique can remove a slice of the peak from the first injection for injection into the second carrier. Dilution by a factor of several hundred can be readily achieved. Precise timing of the two valves is important, requiring software or an electronic timer for optimal precision.

A gradient chamber is a simple device for dilution in FIA and SIA, as depicted in Figure 3. The gradient chamber is a small mixing vessel with a stirring bar and an inlet and outlet for the carrier stream. The volume of the chamber is generally 1 mL or less. The sample is injected into the carrier and disperses to a high degree as it mixes with the much greater volume of the gradient chamber. Dilutions by a factor of several hundred can be achieved with good precision.

The final type of dilution technique is membrane sampling. The use of a membrane sampling device (MSD) for dilution is relatively simple. A donor stream flows on one side of the membrane and an acceptor stream on the other side. The sample is injected into the donor stream and, as the sample segment passes over the membrane, a fraction of it is transported to the acceptor stream. The dilution factor depends on the sample volume injected, the thickness of the membrane, the surface area of the membrane, the channel dimensions in the MSD, and the flow rates of the two streams.

In addition to dilution, MSDs can be used for other sample processing operations, such as matrix modification, sampling of gas streams, solvent extraction, and analyte enrichment.[18,19] The concept of using membrane permeation for sampling dissolved gases and gaseous samples has been applied in continuous flow systems primarily in clinical analysis since the early 1960s. Since then the technique has matured, with advances in membrane materials, sampler design, and flow systems. This approach has been demonstrated for gaseous samples and includes the determination of chlorine, ammonia, carbon disulfide, sulfur dioxide, and hydrogen cyanide.[20] Any gas that crosses a membrane and accumulates on the acceptor side may be suitable for membrane sampling.

Two membrane sampling designs that have been used widely differ in the membrane geometry: the parallel plate (or sandwich) design and the tubular design. In the sandwich design (Figure 4), a planar membrane is securely placed between two inert plates. Two carrier lines, i.e. a donor stream and an acceptor stream, pass through the MSD in separate conduits separated by the planar membrane. The donor stream can be either

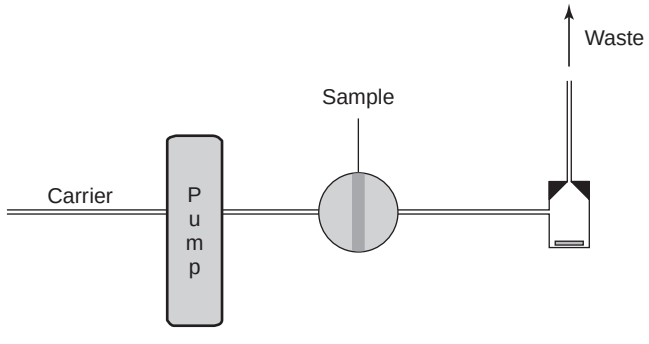

Figure 3 Diagram of a manifold incorporating a gradient chamber.

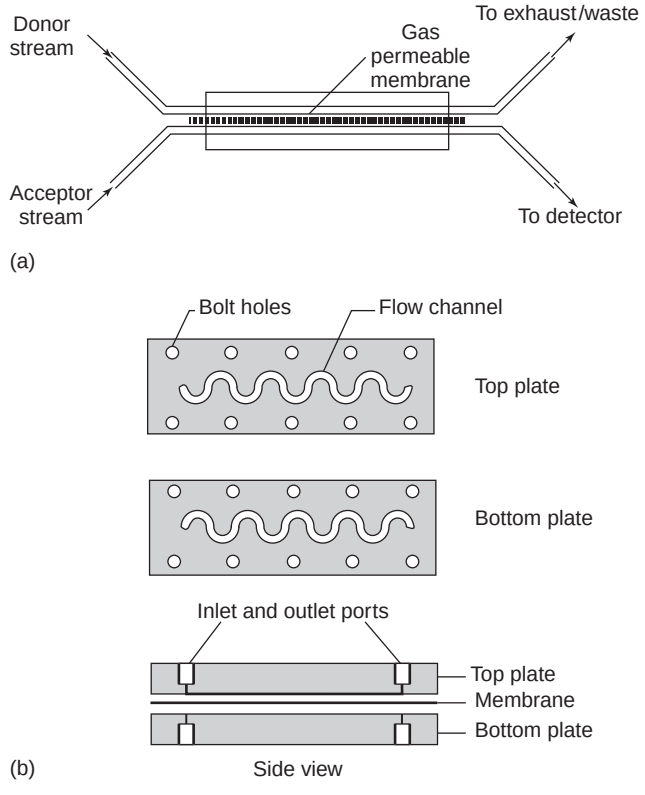

Figure 4 (a) Diagram of a parallel-plate membrane sampler for gas–liquid separation (side view). (b) Detailed views of diffusion plates used in the parallel-plate membrane design.

gaseous or liquid, whereas the acceptor stream is liquid. The plates have engraved channels on their surfaces facing the membrane that define the volume for both the donor and acceptor streams. Separate tubing connections at the ends of each channel provide flow in and out for each stream. For gas–liquid separations, the gaseous analyte is introduced by the donor carrier stream into the planar membrane device where it diffuses through the membrane and dissolves into the liquid carrier stream. Only gases that can permeate the membrane and dissolve in the acceptor stream or react chemically with a reagent in the acceptor stream will be transported across the membrane. Hence, the selectivity advantage is dependent primarily on the membrane material. The acceptor stream is pumped through the planar membrane device and then out toward the detector.

The tubular or tube-in-a-shell membrane sampling design is also used widely (Figure 5). A thin-walled gas-permeable membrane with a tubular geometry is inserted into a larger cylindrical shell made of Teflon®, glass, or some other chemically inert material. The sample passes through the outer shell while the acceptor stream is stopped or flows slowly within the tubing. The tubular design is favored over the sandwich design because it is more resistant to leakage and the tubing length or geometry (linear, coiled, or knitted) can be conveniently varied to alter the contact zone surface area between the donor and acceptor streams. However, the wider availability of planar sheets rather than tubes of membrane material generally favors the parallel plate design for flexibility of use.

Either membrane sampling design can be used in a preconcentration mode by stopping the flow of the acceptor stream during sample collection. The stream to be sampled is kept at a constant flow rate so that the analyte is continuously transported into the acceptor stream. In this mode, equilibrium conditions favor diffusion across the membrane with time. If the analyte reacts with the acceptor stream to form a new compound, a concentration gradient will be maintained that will increase the rate of transport of the analyte across the membrane. That is, analyte that permeates the membrane will be swept off of the inner wall of the membrane into the acceptor stream solution upon reaction. The choice of the acceptor stream depends on the target analyte.

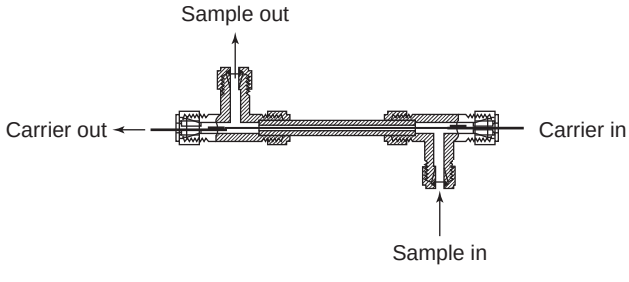

Figure 5 Diagram of a tube-in-a-shell membrane sampler.

A number of different types of membranes have been studied and reported in the literature as MSDs for FIA. Most of them fall into three categories: nonporous (e.g. silicone or natural rubber), microporous (e.g. Teflon®, Goretex®, Zitex®, poly(vinylidene difluoride), poly(vinyl chloride), polypropylene, cellulose esters), and ion exchange (e.g. Nafion®) semipermeable membranes. All are available commercially.

The main difference that is of relevance to FIA among these types of membrane is the mode of mass transport across the membrane wall. With microporous membranes, mass transport occurs by diffusion of the analyte through the pores. Generally, these membranes are used with volatile analytes, such as dissolved CO_2, NH_3, H_2S, and HCN. In the case of nonporous membranes, the analyte actually dissolves in the membrane and diffuses through the wall structure to the acceptor stream side. Only neutral molecules have any appreciable solubility in silicone rubber, the most commonly used nonporous membrane, so this membrane does not work with ionic species. For this type of membrane to effectively transfer analyte from the walls of the membrane to the acceptor stream, the analyte must have a much greater solubility in the acceptor stream than in the membrane, or become converted to a species with this property. Generally, this is achieved using a chemical reaction with a reagent in the acceptor stream that converts the analyte to a soluble ionic species. For example, for transport of NH_3, an acidic acceptor stream will convert the ammonia to NH_4^+. The third type of membrane listed is ionic, where the transport of ions across the membrane wall occurs. The most common membrane of this type is Nafion®, which is a perfluoronated hydrocarbon polymer with pendant sulfonic acid groups. It readily transports small, monovalent cations, such as H^+ and NH_4^+.

The selection of a suitable membrane for the flow system depends on several factors. Optimally, the membrane will have good mechanical and chemical stability, will allow the complete transport of the analyte into the acceptor stream while simultaneously rejecting interferences, and will have an efficient transport rate over short sampling times. For gas-phase sampling, the analyte's volatility, membrane permeation rate, and solubility in the acceptor stream affect the choice of membrane. Membrane porosity, pore size, and thickness will generally determine the collection and separation efficiency. The rate of mass transport across membrane walls is critically dependent upon the wall thickness. Namely, the lag time in analyte transport is proportional to the square of the wall thickness. Thus, thin-walled membranes with

Table 1 Common membrane materials

Material	Properties
Nonporous	
Silicone rubber[81]	high permeation rates chemically and mechanically stable resistant to fouling selectivity influenced by permeability
Microporous	
Polypropylene[82]	more efficient transport than nonporous membranes (~10-fold)
PTFE[83–85]	available with varying porosity, pore size and shape (tubing versus flat)
PVC[86]	not as selective as nonporous membranes selectivity influenced by volatility susceptible to fouling (e.g. particulate)
Ion exchange	
Nafion®[87,88]	efficient for ionic species as well as lipophilic gases good thermal, chemical, and ion transport properties

wall thicknesses in the range 10–200 μm are preferred for FIA work. Table 1 lists a representative number of membranes that have been used in gas-permeable MSDs.

2.2 Sequential Injection Analysis

SIA, which is a novel variation of FIA, has the primary advantages of simpler and more compact hardware, greater method flexibility, and less waste generation. The SIA technique offers an automated approach to sample handling that enables manual wet chemistry procedures to be executed in a rapid, precise, and efficient manner.[10,24] Small solution zones are manipulated under controlled dispersion conditions in narrow-bore tubing. It is readily seen that this definition is quite similar to that given above for FIA. In fact, sharing many characteristics with FIA, many would argue that it is simply an extension of FIA. More than 100 journal articles have been published on SIA since the first paper appeared in 1990.[10] Although it is fundamentally dependent on the controlled dispersion of zones in a flowing stream, the practice of SIA differs from that of FIA.

A simple comparison is useful. Consider a single-line FIA experiment where one injects a sample into a carrier stream containing a reagent, as shown in Figure 6. A sample is pumped into the sample loop of a two-position injection valve and the carrier is flowing constantly through the detector. The length of the sample loop determines the volume of sample injected. When the sample loop is loaded, the valve is switched and the sample is introduced into a flowing carrier stream. The carrier transports the sample through the reactor (usually a coil) to the detector. The sample reacts with the reagent

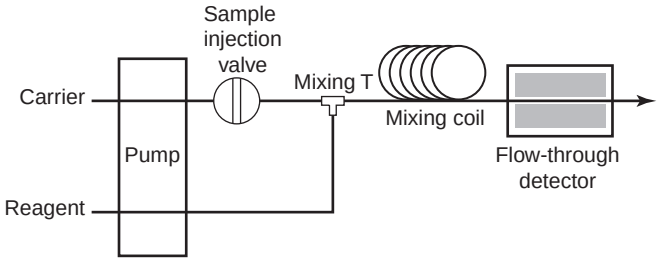

Figure 6 Diagram of a double-line FIA manifold.

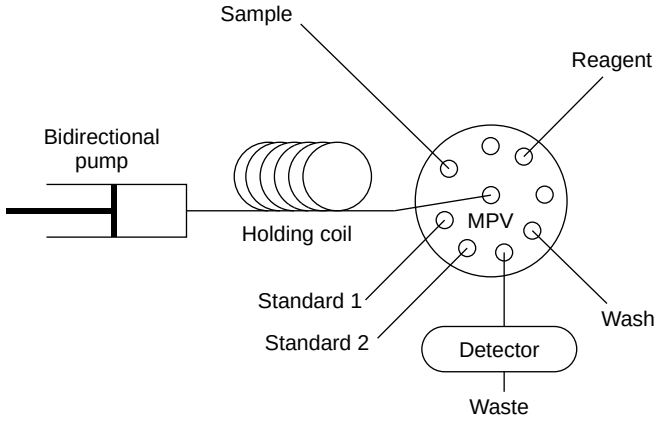

Figure 7 Diagram of an SIA manifold.

en route to form a detectable species. The detectable species gives rise to a peak when it passes through the flow cell of the detector. A calibration curve is then used with the peak height, area, or width to determine the concentration of the analyte in the sample.

However, SIA does not require an injection valve. Rather, a multiposition valve (MPV) replaces the injection valve, as depicted in Figure 7. Usually, the peristaltic pump is replaced with a syringe pump and a coil, called the holding coil, is added between the pump and selection valve. To achieve the same measurement as described above for FIA, the syringe is first filled with a carrier solution that contains the reagent. Then the selection valve is advanced to a port that is connected to the sample line. A small volume of sample is precisely drawn into the holding coil. The volume of sample is determined by the computerized flow program. The selection valve is then advanced to a port that is connected to the detector, the direction of flow is reversed, and the carrier transports the sample through the reactor to the flow cell of the detector. Again, a detectable species is formed and is registered as a peak by the detector. The concentration of the analyte in the sample is determined in a similar manner as for FIA. The flow reversal has a dramatic influence on the mixing of stacked sample and reagent zones.

More often than not, instead of including the reagent in the carrier, an SIA experiment is expanded so that

the reagent is loaded as a separate zone. In this case the syringe is first filled with a simple carrier or buffer. After the sample zone has been drawn into the holding coil, the selection valve is advanced to a port connected to a reagent reservoir and a small reagent zone is drawn into the holding coil. In this way it is possible to construct a stack of well-defined zones which can be mixed together to give rise to a detectable species.[25] Unlike FIA, which requires replumbing when a more complex chemical addition scheme is required, a change to the flow program is the only alteration for SIA. The manifold remains unchanged. The additional advantages of lowering reagent consumption and minimizing the production of potentially hazardous wastes are both important advantages of SIA.

The other ports of the selection valve can be used for calibration standards, additional reagents, and as locations where more sophisticated operations such as dilution, trace enrichment, and incubation of reactants can take place. For example, titrations can be performed by SIA whereby the reactor and detector in Figure 7 are replaced with a stirred titration cell. Appropriate sensors, electrochemical or optical, are placed in the titration cell which acts as the flow cell and titration chamber.

Researchers have applied SIA to the determination of analytes as diverse as trace radionuclides and biopolymers. Although SIA makes use of a simpler flow manifold (this is particularly so for multicomponent chemistries), the development of an SIA method is not as straightforward as in FIA. Careful attention must be given to the design of the measurement sequence to ensure that adequate zone penetration has taken place. Accurate measurement of sample and reagent zones necessitates microprocessor control. Of course, once the method has been developed, the microprocessor ensures slavish repetition of the optimized sequence. To summarize, several advantages can be identified of SIA over conventional FIA in the process environment:

- Reagent use is drastically reduced – typical FIA experiments make use of at least 1 mL of reagent per measurement. SIA typically makes use of 50 µL. This means that in a 24-h period and assuming one measurement per minute, the FIA monitor would consume more than 1.5 L of reagent. The SIA monitor would consume less than 75 mL. It has been noted that the most frequent reason for FIA process monitor failure is running out of reagents.
- Flow manifolds are simple and robust, typically comprising a pump, selection valve, and detector connected by tubing. Unlike FIA, the same manifold can be used for widely different chemistries simply by changing the flow program rather than the plumbing. Monitor maintenance is therefore simplified.
- The selection valve replaces the injection valve and provides a means for selecting different sample streams and calibrants. This enables automated calibration and quality control.
- Components used in an SIA manifold are amenable to laboratory, field, and plant operation.

3 INSTRUMENTATION

3.1 Design Considerations

Because suppliers have often tried to place laboratory equipment directly in an industrial setting, many so-called 'process' monitors have come to pass. However, to be successful, a process monitor must be designed from the ground up. One only needs to examine the major differences that exist between a modern process gas chromatograph and a laboratory gas chromatograph to see a sterling example of this principle. As a case in point, there are custom-designed FIA multi-end-point titrimetric systems at a major petrochemical company, for example, that are still on-line even after 10 years or more of continuous monitoring. In contrast, laboratory systems for this analytical function were tested side-by-side to the custom systems and were inoperable after a 6-month period.

Contrary to popular belief, it is not only as a result of a lack of ruggedness that laboratory instruments are unsuitable for this application. The failure of laboratory systems in the plant can be attributed to the fact that too often the laboratory maintenance philosophy is applied to the plant environment. It is acceptable in the lab (though not desirable) to have a meticulous technician with tissue in hand to pamper the laboratory system, watching it closely for a new source of bubbles, tweaking a malfunctioning component, etc. In the plant, the monitor must be designed and constructed so that the scheduled maintenance for it can be completed in less than 30 min each week. No plant operator will be prepared to hover over the monitor to coax results from it. Where components are chosen and a process monitor is designed with these maintenance requirements in mind, and a disciplined and thorough scheduled maintenance program is established, the success of the monitor increases dramatically.

Another important design consideration is the sustainability of continuous improvements to the design. The literature is full of prototype process monitors which, although ingenious, did not outlast the time it took to generate data for a project report or paper. Industrial process monitoring is not the short-term project beloved of graduate school researchers. Therefore, the industrial process to be monitored should not out-live the process monitor's

manufacturer, otherwise supporting the monitor will become a formidable in-house engineering project. More than a few process monitors languish in storage for the reason that their manufacturers no longer exist.

Additionally, when the process monitor is part of an ongoing commercial endeavor, the design process, sourcing of suitable components, product release, customer consultation, support, and parts availability are all greatly enhanced. The monitor design process must include elements of design review, testing, corrective action, release control, and postrelease support. The ISO 9000 standard series is an excellent summary of a design and manufacturing environment that encourages this process. It is especially appropriate to process monitoring because the application is bound to be long-lived enough for continuous improvement of the process monitor to occur. This design philosophy can be summarized as follows:

1. Simplicity is imperative – the simpler the system, the easier it is to operate and maintain.
2. Design a system that will support a bullet-proof scheduled maintenance plan. Support the process monitoring system for the long haul.
3. Take note of the environment – most plant environments are hostile compared to laboratories. Allow the environment to suggest design constraints. A delicate and temperamental detector is probably not going to survive the rigors of the plant.
4. Carefully design the sampling system. More monitors fail because of an inadequate sampling system than for any other reason.

3.2 Manifold Hardware

Access to a process test bed for initial demonstration and continuous design improvement is important. Samples in process monitoring are often assays of major constituents, in complex matrices, and the challenge is to dilute the sample sufficiently to avoid saturation of the detector and to minimize matrix effects. Fixed-volume sample loops mounted on or in six-port commutating valves are common; generally, the sample loop volume is small, between 2 and 20 μL, to afford greater dispersion of the injected sample. The necessarily small orifice of the loop, however, may present a potential clogging point as the undiluted sample must pass through it during the loading step. This concern may be addressed by use of an efficient sampling system. If this becomes a problem, reducing the flow rate of the filling stream often helps avoid clogging. The precision in the response is typically <0.1% RSD with these types of valves.

The use of certain flangeless fittings can also be a source of blockages in valves. These types of fittings often rely on a ferrule that pinches down on the tubing. This reduces the tube diameter at that point and in the worst case can totally shut off flow. One solution to this problem is to insert a pin into the tube when the fitting is attached. This will keep the tubing from crimping closed and the Teflon® will rather cold-flow around the ferrule. Alternatively, commercially available high- or low-pressure compression fitting systems that use PEEK (polyether ether ketone) ferrules, nuts, bodies, and tubing are rugged, inexpensive, and easy to use. The rigid walls of the PEEK tubing prevent the kinking or pinching that occurs easily in thin-walled Teflon® tubing.

In general, the use of peristaltic pumps in process monitors is a major weakness. Again it comes down to the degree of scheduled maintenance expected. If examination of the peristaltic pump tubing (Tygon®) for wear is required too frequently, the plant operator will quickly lose enthusiasm for the utility of the monitor in controlling the process.

For single-channel monitors, a three-way valve can also be used for variable-volume timed injections. These valves are often less precise because multichannel peristaltic pumps limit the flow precision and thus the volume precision of a time-injected sample. Syringe or displacement pumps will increase the flow precision, but also increase the cost of each channel of flowing reagent or sample.

For a batch (off-line) FIA instrument, an autosampler is generally desirable so that the batches of process samples can be collected and analyzed at one time. The autosampler can contain nonsample vials containing calibration standards, reference samples, check standards, a wash bath for idle periods, and other needed overheads for batch analysis. The autosampler can be one of three types, random access (or XYZ), random access circular design, or sequential. An autosampler will execute a work list of samples usually stored in and controlled by a microprocessor. The autosampler is particularly suited to unattended operation of the monitor – the samples can be collected during the first shift, then the batch analysis can be started at the end of the shift with results available the next day.

For smaller batches or single samples, a manual sampler may be sufficient. These amount to an on-demand sampler, usually consisting of a sample tube or probe that can be placed into the vessel containing the sample or standard. The user then cues the instrument that the probe is in the correct sample. A manual or on-demand sampler may or may not automatically return the sample probe to a wash bath or container. An interesting hybrid batch/stream sampler can be arranged by placing one or more flow-through tubular baths, such as an autosampler wash bath, into the sample tube rack. A process stream can then be flowed through the bath. The autosampler

can thus sample the stream in the bath at any desired frequency.[26]

A well-stocked inventory of spare manifold devices will typically be included as one of the provisions of a good hardware maintenance plan for a process FIA/SIA system. Modular design of the monitor will facilitate rapid replacement in the event of a break down.

3.3 Flow Injection Analysis Software

The rapid development of computer hardware presents an ever-changing spectrum of choices for the control of process monitors. The power of inexpensive computer hardware invites the development of sophisticated software packages, but increased sophistication does not necessarily result in a useful system. On the contrary, device control and data acquisition software for FIA systems should focus on enhancing the capability of the monitor to simply operate without error while unattended. Because hardware changes so rapidly, software must be scaled so that, as new platforms become available, rewriting the software is not necessary.

The FIA/SIA software to be used in a process environment will allow for method and schedule creation, editing, and execution as well as data analysis and quality control. Specifically, it will allow:

- precise timing of devices;
- data acquisition with the ability to filter data;
- automated calibration and quality control;
- results reporting via serial 4–20 mA analog outputs, with controller area network (CAN), FieldBus, or other industrial plant communications standards;
- display of trend data and calibration history;
- annunciation of alarms and warnings.

Because FIA and SIA techniques cover a wide range of unit operations and draw from a large pool of devices and detectors, it is imperative that good software allows a high degree of user configuration. This configuration should extend beyond the accommodation of different devices and detectors to also allow the user to determine the sequence that will be followed, i.e. when calibrations will be initiated, how frequently lines will be flushed, how samples will be presented to the monitor, and when measurements will be carried out.

Current software packages are frequently required to communicate their results to the plant's distributed control system (DCS) where the information is used to control the process either in a closed-loop fashion or as an input that the operator will base certain decisions on. With the expanding use of intranets, it can be expected that these data will also be made available on the company intranet, so that other company personnel – management, accounting, shipping, etc. – can also have access to critical plant data. Hardware is already available to do this using embedded processors and appropriate network connections. The authors are not aware of any commercially available software that provides this as an option. To really exploit this technology though, software should go further and allow remote diagnosis of the process monitor via the company's intranet. The ability to query the status of critical instrument devices, sample streams, and reagent reservoirs will reduce existing scheduled maintenance burdens.

3.4 Method Development

It is important for process monitor developers to understand the chemistry of the industrial process sufficiently to design appropriate analytical methods for the process monitor. This is especially important when replacing older techniques whose analytical chemistry may not be fully understood by plant personnel who are perhaps a generation removed from their invention and installation. For example, in the chlor–alkali industry[27] which produces chlorine and sodium hydroxide by electrolysis of brines, it is important to know the concentrations of sodium chloride, hydroxide, chlorate, and hypochlorite in the electrolysis cells in order to maintain optimal current efficiencies, as electrical power is a major cost to this industry. An FIA process monitor would ideally measure all four cell-liquor analytes simultaneously without dilution. As each of the analytes is a major constituent, the FIA methods for each must be rugged enough to minimize the large changes in matrix concentrations and density across about two orders of magnitude in each analyte.

This particular industry breaks down into two separate applications and the analytical methods must be flexible enough to address both. The older plants have a mercury cell that uses liquid mercury as the cathode; the newer plants have replaced their mercury cells with cells using Nafion® membranes to separate the anodic and cathodic processes. The two processes all have various concentration ranges which need to be anticipated; all ranges require precision of less than 1% RSD. During the design phase of the process monitor, the methods of determination are central. The methods must be developed in close consultation with the plant engineers and quality control chemists. Formerly monitored by titrations, batch colorimetric techniques, and segmented-flow analysis,[28] FIA is replacing these techniques in many chlor–alkali plants.

The applications chemist who designs the method typically works closely with the process monitoring laboratory personnel to obtain representative samples, understands the matrix and concentration range of the analytes in the process, understands the method documentation and

validation used for process monitoring (on-line or laboratory), collects samples for method development, and negotiates and rationalizes any biases discovered between the current and proposed methods. Often the most familiar method to the process engineers and laboratory may not be accurate, but has a legacy of data standing behind it. The applications chemist must respect the use and consequences of the data obtained by the method. In this case, negotiation and diplomacy may count for more than proof of accuracy.

3.4.1 Calibration

Both laboratory (off-line) process monitors and modern on-line or at-line monitors can be calibrated frequently and the calibration standards set can be chosen appropriate to the application. Commonly, the monitor is calibrated once at the beginning of the batch or day, and the calibration is then regularly checked for drift by measuring one or two check standards during the batch, depending on the number of samples in the batch.

Designing calibration standards for a multiparameter process monitoring can be challenging. The expected concentration range in the process must be encompassed at once by a set of standards. For example, the chlor–alkali process requires calibration standards which contain four analytes ranging from 5 ppm to 250 g NaCl per liter, 0.5–2000 ppm sodium chlorate, and 1–150 mg sodium hypochlorite, all in 10–34% (w/w) NaOH. A certain degree of creativity is involved in designing a minimum set of calibration standards to span these ranges. Almost always, the FIA-based method is replacing the earlier method; however, the process client will often want to preserve the calibration standards from the earlier method. In this case, it is important for the method developer to understand fully the composition of these standards, which sometimes may be more venerable than accurate. If the method is to be transferable to other plants and sites within the industry, the calibration standards must be fully characterized.

3.4.2 Quality Control and Method Validation

It is common for an industry to develop their own internal reference samples and standards which must be respected as the final arbiters of accuracy and precision. For example, in the phosphate rock industry, the Chemists of the Florida Phosphate Industry[29] have established a monthly phosphate rock check sample program. The samples are of two types, an analyzed rock sample and a series of pure polyphosphorus compounds. It is important for any newly developed FIA method intended for process monitoring to be included in such programs. In other words, the industry itself must be encouraged to validate the method with familiar reference samples. If the laboratory with the new method does well or is not rejected as an outlier among the various laboratories, then this method can be considered a process monitoring success. If the laboratory is rejected, for whatever reason, the developer must carefully determine the source of bias using the industry's own check samples. The wise developer will obtain these samples and use them to check the new method before proceeding with validation.

Similarly, in the food, beverage, and fertilizer industry, any new FIA methods must perform similarly to the standardized methods they are replacing. For example, the Association of Official Analytical Chemists publishes a compilation of standard methods for the fertilizer industry, each of which has been subjected to one or more collaborative studies.[30] In obtaining acceptance for a new, FIA-based method of analysis, it is important to identify plants and laboratories who historically have led the way in methods innovation as there is considerable inertia in changing from older, approved methods to anything new – even if it offers obvious advantages. Such clients are invaluable in gaining acceptance for a new method and will often fund or organize the necessary collaborative studies. Again, the use of the data and decisions that result from the data need to be respected; these are more important than the methodology itself to process engineers.

3.5 Sampling Systems

This category is divided into samples collected either at-line or on-line.

3.5.1 Grab Sampling

There are many examples of process monitoring in which grab samples are taken from the process and determined at-line, i.e. they are sent to a central laboratory for batch analysis by an off-line laboratory instrument. This is particularly the case in wastewater monitoring situations. Sampling problems can be the most daunting challenge for FIA process monitoring. For example, in the most benign application environment, the biotechnology industry, a continuous withdrawal of process solution would seem most desirable, and this is commonly done for headspace analysis. However, the bioprocess solution has a high suspended solids concentration and requires filtering in order to exclude the organisms and other solids from the withdrawn sample. Another potential obstacle is that withdrawing bioprocess solution continuously may open a path of infection to the bioprocess.

Such sampling considerations apply to the whole range of industries currently benefiting from FIA-based laboratory instruments: for the phosphate rock industry

and chlor–alkali industries discussed above, both have almost insurmountable on-line sampling problems – the sample is a solid or a superheated viscous liquid, in inaccessible sealed cells or large-scale inaccessible reactors, or are sampled seldom enough to preclude a permanent small-scale bypass or sampling stream. Thus, intermittent batch withdrawal of a sample is still the most common and flexible process monitoring sampling method. The batch sample, collected from anything from a shake flask, a 1000-L bioreactor, a tank car, an active electrolytic cell, the packaged product, condensed high-pressure steam, ores, sands, or concentrated brines from wells, can then be centrifuged, filtered, cooled, homogenized, labeled, ashed, crushed, preserved, and batched. The gamut of matrixes, sources, temperatures, and pressures further complicates the sample collection process.

Generally, processes are monitored on a schedule determined by the plant engineers. The sampling frequency varies, from once every few minutes, to perhaps only once or twice a day. Grab samples are obtained through sampling ports by dipping containers, taps, etc., and the samples are bottled, labeled, and sent to the laboratory. This type of monitoring may be designed to validate or to show, after the fact, that a process was under control during the manufacture of the product, or designed to show an imminent need for process maintenance.

This may seem counter to the desire to obtain time-dependent data on the process. However, it is generally inefficient to run each sample individually to the laboratory to obtain a single data point – the entire calibration and quality control overhead must then rest on each single sample. Instead, usually a day or half-day's worth of samples are run at one time, preceded by an instrument calibration, and interspersed with quality control samples. The size and frequency of batches depends on the predictability of the process. For processes under development or improvement, as in a bioprocess, smaller, more frequent batches may be necessary to be able to take corrective action. However, even these processes need a period over which a reliable trend can be established. For processes that have been implemented for some time in large-scale production, the trends have become so predictable that only a post-mortem type of monitoring is necessary. For example, in the chlor–alkali industry, over a period of time, an upward trend in sodium hypochlorite and chlorate concentrations will occur prior to Nafion® membrane replacement.[31] However, this occurs only every 2–3 years. This maintenance schedule is predictable enough that these data serve only to confirm the predictable. Thus, a process that is useful at large scale does not necessarily need to be micromanaged.

3.5.2 On-line Sampling

On-line sampling is certainly more challenging, but the real-time nature of the analytical information provided makes it desirable, especially for pilot-scale process support. Collecting a representative sample from a process stream involves many paradoxes and trade-offs. It is rare that one has available a sample sufficiently pristine that it can be fed directly to the inlet of the selection valve (as in SIA) or loop of the injection valve (as in FIA). To derive meaningful information about the process, one must change the sample without changing the desired measurement variable, or change it in such a fashion that the degree of difference from true process conditions is known. The FIA method is somewhat more vulnerable to sample system influences than some methodologies, as these techniques use relatively small diameter tubing in their internal plumbing. The major advantages of FIA (mechanical simplicity and reliability) far outweigh the potential liability of sampling manipulation.

The chain of events typically involved in obtaining a sample are:

1. collection of the sample from the process, usually a pipe or tank;
2. transport of the collected sample to the location of the monitor;
3. conditioning of the sample for presentation to the FIA instrument;
4. introduction of the sample to the FIA manifold.

Of course, virtually every chemical process that might be subjected to chemical analysis is different from all others. It cannot be overemphasized that, to be successful, one must critically and completely understand the specific sampling situation. Despite the absolute criticality of the sampling system for successful instrument performance, unfortunately the design and engineering of sampling systems are often ignored or done as an afterthought when the performance of the installed instrument does not meet the desired performance goals.

3.5.3 Sample Collection

The first step in the sequence of operations required to obtain a meaningful on-line measurement involves determining how to gain access to the process stream. For practical reasons it is desirable to design the sample handling system to use a minimum number of additional devices. FIA has a significant advantage over many other modes of on-line analysis in that the quantity of sample required for a single determination is quite small – typically less than 100 μL. Thus, the overall amount of material that must be collected and processed can be significantly reduced compared to existing process

sampling methodologies. Four of the most common approaches to process sample collection for FIA are described in the following paragraphs.

Probably the most desirable sample collection approach is to simply tap across an existing process pump. This approach takes advantage of the fact that the capital expenditure (and maintenance responsibility) for the pump has already been made. In such an arrangement, a sample port is added to the exit piping and then routed to the desired monitor location. An output line is routed back to an inlet port on the upstream side of the pump (Figure 8). Samples obtained in this manner are well-mixed by virtue of their passage through the pump. The main disadvantage of this approach is that large (acute) process changes are recirculated through the fast loop and can lead to an increased analysis lag time. If such lag times are a problem, the recycled sample must be reintroduced into the process downstream of the take-off point at a low-pressure point in the process.

If a convenient process pump is unavailable, but the process is under sufficient pressure, a second sample collection approach is to introduce an in-line connection at a high-pressure point. Existing process pressure thereby provides the energy to force the process sample to the desired monitor location and back to a lower-pressure process point. It is tempting to make use of existing process connection points (blind flanges) to provide the connection points. However, extreme care must be taken in such cases to avoid various types of sampling problems. Blinded sidearms on piping runs typically are mounted at right angles to existing piping. If the piping run to which connection is desired runs horizontally, such right-angle mounting may place the sidearm on top of the pipe, on the side of the pipe, or on the bottom of the pipe (Figure 8).

Each of the first two designs may present sampling problems. If connection is to the top of the pipe, gas bubbles will be preferentially routed into the sample line.

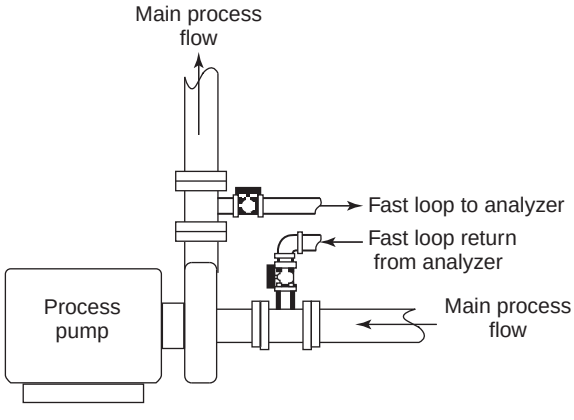

Figure 8 Process sampling by creating a tap across an existing process pump.

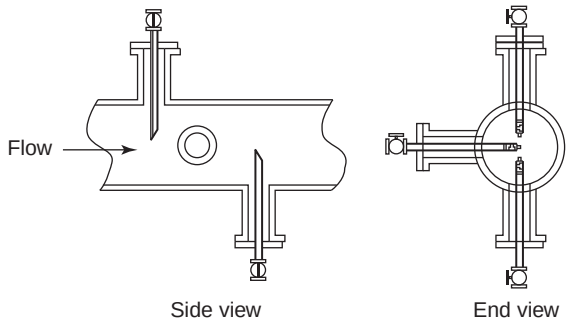

Figure 9 Process sampling by using a stinger.

If connection is to the bottom of the pipe, suspended solids will be routed preferentially into the sample line. The preferred connection is to a sidearm mounted on the side of the pipe. Even then, it is probable that the volume of the sidearm will introduce significant time lags because it represents a large, stagnant volume of process fluid which must be diluted by the sample flow. The sidearm piping orientations are viable sampling points for the third design – the so-called 'stinger' (Figure 9). Note in the figure the shape of the end of the stinger and its orientation with respect to the process stream flow. This arrangement allows some additional discrimination against bubbles and/or particles flowing down the pipe. The stingers illustrated are all for the take-off point of the sample flow; the return point is not shown, being located farther down the process or at some other low-pressure point.

A fourth approach is to use a dedicated pump to provide flow through the fast loop to the monitor. However, this is usually not a popular option because it represents both additional capital and maintenance costs. Also, when such a pump inevitably fails, one typically finds in a plant that its replacement will always be a low-priority item compared to other maintenance emergencies. The net effect is increased monitor downtime and decreased monitor reliability. However, a dedicated pump does provide several advantages because it can be much smaller and less expensive than other options, and it does not require multiple taps into the process for extraction and return points (Figure 10).

The arrangement shown in Figure 11 shows the use of a flat-plate style of cross-washed inertial filter. Systems of this design are usually equipped with piping and a dedicated pump that are sufficiently large to allow the largest solids in the process to completely traverse the fast-loop piping. The fast-loop piping is typically 0.75 in or 1.0 in pipe. With the addition of automated valving and a different stinger style, the stinger system can become a fairly sophisticated means of prescreening the sample to reduce larger particles from entering and thereby permitting the use of a smaller cross-washed filter and

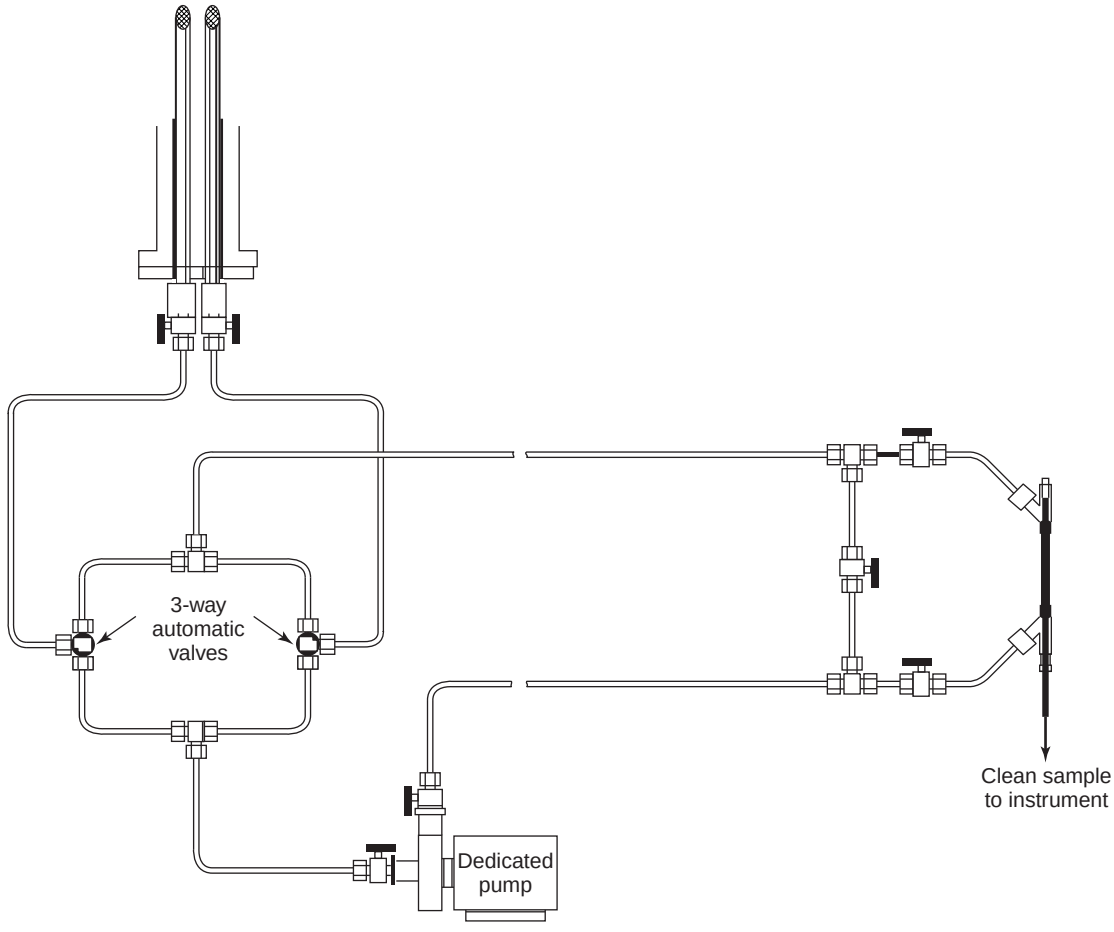

Figure 10 Process sampling using a dedicated pump.

stingers. This design leads to significantly smaller piping in the fast-loop plumbing. Furthermore, the arrangement depicted also provides automatic backwashing of the stingers with the pumped flow. Each stinger alternately serves as a sample take-off and/or return point that depends on the position of the three-way valves. However, flow through the rest of the fast-loop tubing remains monodirectional to and through the smaller, tube-in-tube cross-washed inertial filter.

3.5.4 Sample Transport

Collection of the sample from the process location is only the first step in the chain of events that must occur before quantitative information can be derived from the on-line FIA monitor. Once withdrawn from the process, the sample must be physically transported to the relatively distant instrument. The rate of sample collection must be designed such that transport lag times remain within acceptable limits; that is, the sample transport distance will determine the total amount of sample delivered to the monitor. Whether this transport is done with a process pump, process pressure, stinger, or dedicated monitor pump, several factors must be considered. Chief among them is the representativeness of the sample.

That the sample does not change significantly during transport or that such change is not detrimental to the quality of the analytical results or to the reliability of the instrument is, of course, critical. In the sample transport step, the factor that is most likely to change drastically is temperature. Many processes operate at elevated temperatures, and the collection and transport steps can result in fairly severe decreases in sample temperature. Because many process streams applicable to FIA may be fairly concentrated salt solutions (for instance, plating baths), such temperature drops can result in solid material precipitating from solution – with subsequent catastrophic effects on the sample system! Temperature drops can also result in shifts in chemical equilibria, which results in a measurement bias. It is worth noting that sometimes the reverse situation occurs on a grab sample that is taken from the process and analyzed because such a sample may result in a substantially different value when compared to the on-line monitor.

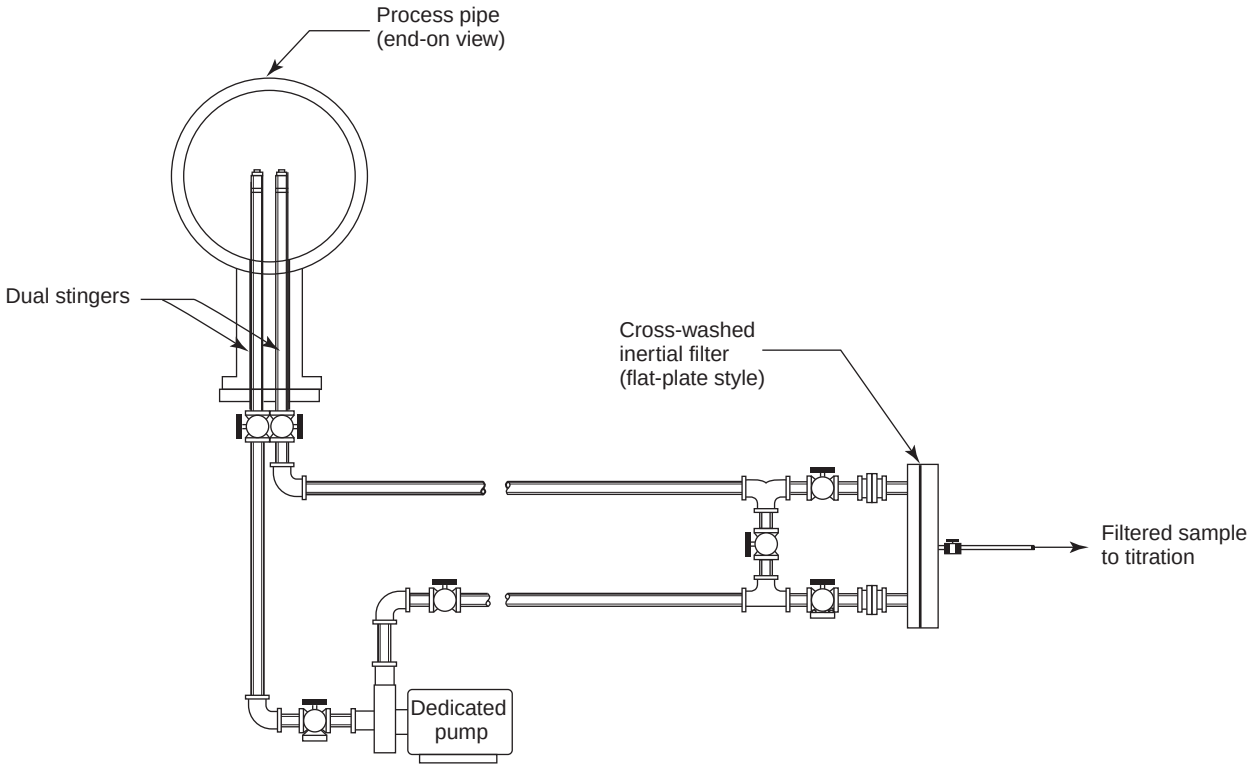

Figure 11 Process sampling approach employing a stinger and a flat plate/cross-washed inertial filter.

Finally, depending on the climatic conditions, cooling in transfer lines may be sufficient to actually allow the samples in the transfer lines to freeze or precipitate. Depending on the specific process and situation, it may be necessary to insulate or provide heat tracing to the transfer lines up to and including the SIA/FIA valve. Finally, the introduction of fast-loop piping (and other sample processing devices) will introduce various time lags in the analytical results. An exhaustive treatment of the possible sources of lag time is beyond the scope of this review, but such time lags must be considered for all parts of the sample handling system.

3.5.5 Sample Conditioning

Filtering is important because the presence of suspended solids is likely. One of the authors recalls an incident in an industrial plant in which a sample collection problem was found to be the result of a mop that had somehow entered the process stream piping! Even high-purity water-based process systems may have particulate matter resulting from piping or pump wear, so the addition of a filtration step is essential in on-line monitoring, and particularly so for FIA instruments because of the relatively small diameter internal tubing that is used. If the sample stream contains very low levels of particulate matter, it is possible to use in-line cartridge filters (available from many vendors), and simply change the filter element based on time interval or by measuring the pressure differential across the filter. However, if the levels of suspended particulate are higher, the cartridge-type filters will clog too frequently and result in unacceptably high maintenance requirements. The criteria for effective filtration must include not only how effective the particular filter is for removing particles above a certain size, but how long the time interval is between required maintenance (e.g. replacement or backwashing). The most useful type of filter is a cross-washed inertial filter (see Figure 11), because these types of filter yield the greatest reliability and require minimal maintenance.

Many sample streams either contain bubbles or dissolved species that will form gas bubbles upon removal from the process. The latter situation is usually attributed to the drop in pressure that occurs from the process stream pressure to the ambient (or nearly ambient) pressure at the instrument. If the initial quantity of sample to be analyzed is determined volumetrically (e.g. by the sample loop in an injection valve), the presence of bubbles will contribute to significant analytical errors, either by changing the size of the volume injected, or by outgassing to form bubbles after sample injection, with subsequent disruption of the normal dispersion patterns in the manifold or detector response. There are basically two approaches

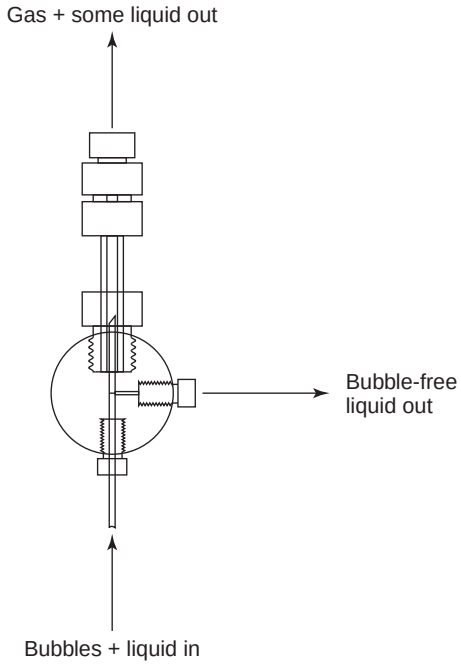

Figure 12 Diagram of a gas–liquid gravity separator.

to removing bubbles: (a) bubble separation based on gravity and/or wettability, and (b) membrane extraction. Gravity-based methods can handle very high quantities of formed gas bubbles, but are relatively ineffective at removing dissolved gases. The inverse is true for membrane extraction methods. Probably the best approach is to combine the two methods in series, with a gravity-separation device to remove the vast majority of formed bubbles, followed by a membrane device to remove any remaining dissolved gas (Figure 12). These two approaches are less suitable for cases where foam is present. Foams can be disrupted or broken by forcing the foam and liquid through a porous element, and then performing a gravity separation and membrane degassing downstream of the foam-breaker unit.

All of the above sample conditioning concerns apply to the determination of a liquid-phase analyte. Of course, FIA is used widely to analyze gaseous analytes as well. In such cases, an MSD is usually incorporated as the interface between the gaseous sample medium and the FIA liquid analytical stream. The gas-phase sample must be successfully collected from the process and physically transported to the instrument, subject to the same types of potential sample processing difficulties as described above for liquid samples (filtration, temperature regulation, pressure regulation, etc.).

Our discussion has attempted to address some of the challenges that are likely to be encountered when designing a process sampling system for FIA. A more exhaustive compilation is available that is applicable to the design of systems for the many other types of analytical instrumentation that are used in process monitoring.[32]

4 APPLICATIONS

Since the late 1970s, process FIA has spread worldwide and is now in use by the chemical, petrochemical, refining, paper and pulp, mining, nuclear, biotech, food, pharmaceutical, brewing, wine, and aircraft industries. Several excellent reviews offer more details.[33–35] The following sections describe examples of FIA process monitoring approaches in several of the major industrial application areas.

4.1 Chemical (including Petrochemical) Refining[33–41]

An example of process FIA developed and implemented in the chemical industry is used to illustrate this application area. This application involved the monitoring of Fe^{II} in a process called SulFerox, an aqueous redox process for scrubbing H_2S from sour gas streams. A number of different technologies have been developed for this purpose, and they meet a great need because much of the natural gas in the world is sour, i.e. it contains H_2S from a few parts per million up to very high percentages. Obviously, this has to be removed before piping it to the customer.

The process chemistry involved an aqueous redox system in which an Fe^{III}–ligand chelate oxidizes H_2S to elemental S in an absorber, followed by removal of the elemental S by filtration in a continuously circulated stream. The absorber chemistry converts the Fe^{III}–ligand to an Fe^{II}–ligand; the stream is circulated to a regenerator where air oxidizes the ferrous chelate back to the ferric chelate, which is then returned to the absorber. The FIA process monitor was developed to monitor the level of the ferrous chelate before and after the regenerator. The FIA instrument is shown schematically in Figure 13, utilizing the well-known analytical chemistry of the formation of a colored complex between Fe^{II} and *o*-phenanthroline. This instrument was used on-line for several years, with process engineers relying on the information provided to optimize the Sulferox process behavior and chemistry. It was very useful to detect process problems before they caused an upset condition.

4.2 Biotechnology (including Food, Fermentation, Pharmaceutical)[42–51,67]

It is significant that much of the on-line process monitoring instrument development activity since 1990 has

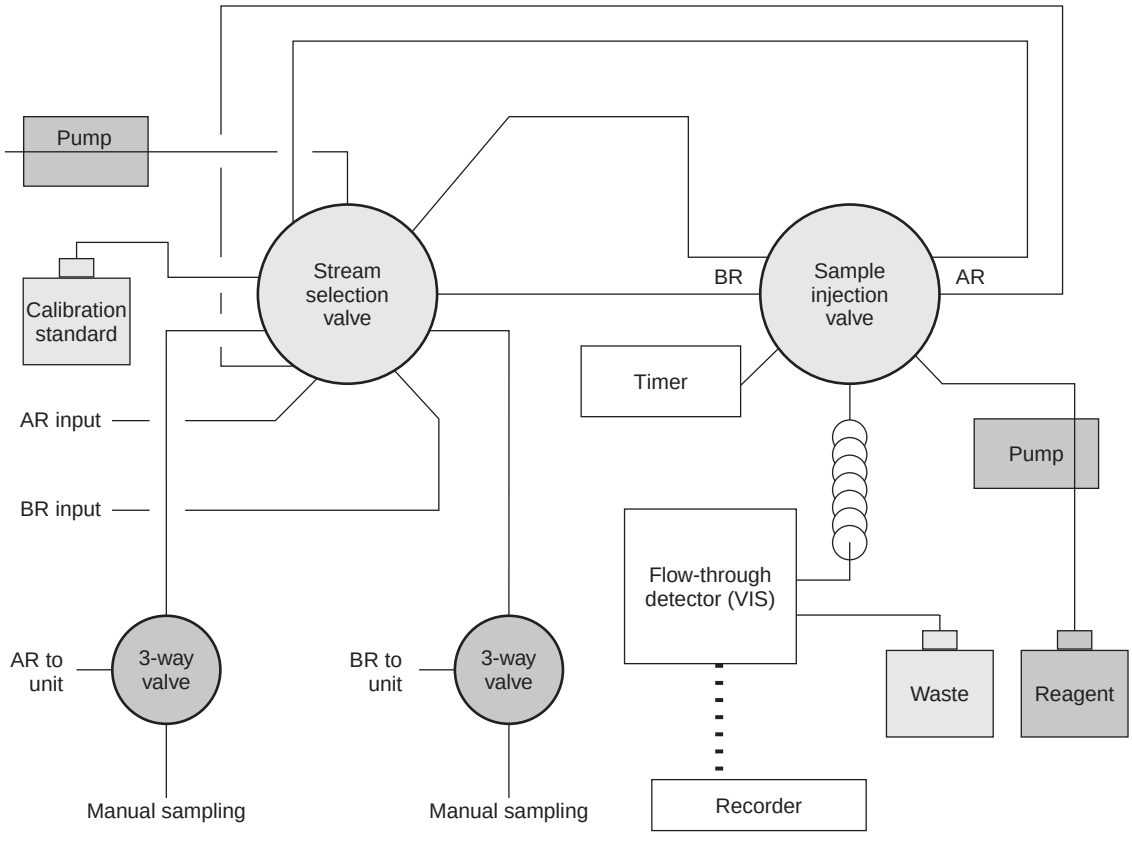

Figure 13 Diagram of an on-line FIA process instrument for monitoring the Sulferox process at Shell Chemical (BR = before regenerator, AR = after regenerator).

been in the biotechnology-related industries. Significantly, several companies have recently released on-line process monitoring products designed for this generally well-funded industry.

Mattisson and colleagues at Lund University recently reported the development of an on-line SIA process monitor for determining D-lactic acid in fermentation broth.[68] The manifold design incorporates a broth-sampling interface and immobilized enzymes on porous glass beads in a minicolumn (Figure 14). The process monitor was tested with a quality control standard on every sixth cycle, and variations in the sample response were normalized (Figure 15). The cycle time was 3 min and the linear range of the method was from 1 to 25 mM – the authors note that the potential exists for automatic dilution by manipulating the sample volume as a function of the previous measurement. The extension of the linear range in this manner would provide flexibility to make the method very useful for the large concentration changes that are typically observed during cell cultivation. The 0.2 μm cellulose acetate membrane on the filtration inlet had to be replaced each day of operation because of the accumulation of cell mass. Reduced costs were realized by adding an expensive enzyme cofactor sequentially rather than continuously. The authors reported excellent agreement between their method and an off-line method that took considerably longer (10×).

4.3 Manufacturing

The application of process monitoring in the manufacturing sector is illustrated by describing a recent example from the aircraft industry. The chemical milling process involves the selective removal of metals from aluminum parts to reduce weight or change the shape. Parts are coated with polymer film, then the film is removed from the areas to be milled. The masked parts are dipped in hot concentrated caustic to rapidly dissolve the aluminum. The parts are removed from the caustic to measure etch rate and then re-immersed until the final dimension is achieved, followed by peeling of the masking polymer. The Boeing Company has recently developed an at-line SIA process monitor to determine the concentrations of NaOH, Al, and Na_2S in the chemical milling solution by titration.[69] The simple, inexpensive instrumentation has the additional advantages of using existing wet chemistry methods, suitability for metals, anions, and acid–base

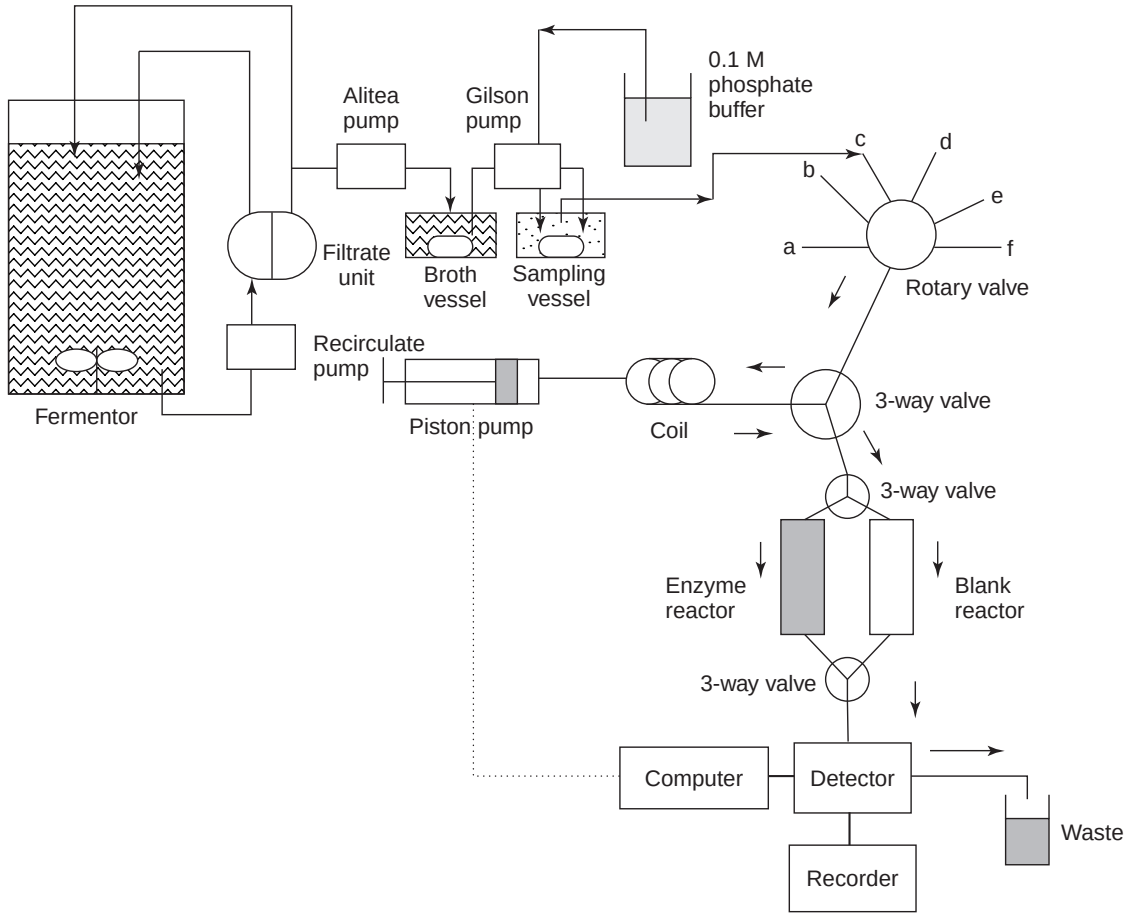

Figure 14 Diagram of an SIA instrument used to monitor a process fermentor. (Reprinted from H.-C. Shu, H. Hakanson, B. Mattiasson, *Anal. Chim. Acta*, **300**, 277–285, Copyright (1995), with permission from Elsevier Science.)

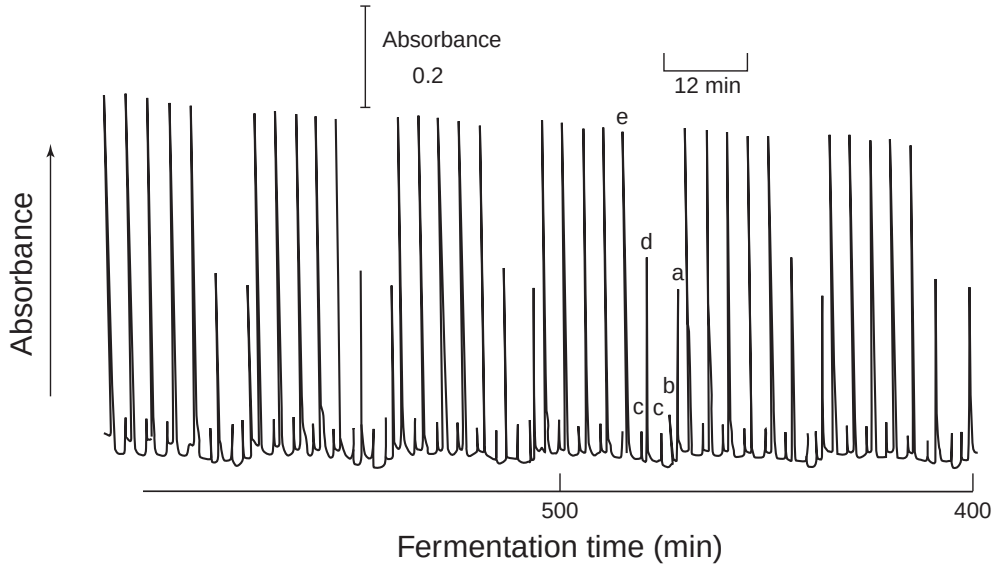

Figure 15 SIA response during on-line monitoring of D-lactic acid production in a batch fermentor following incubation: (a) background peak, (b) washing peak, (c) initial peak, (d) calibration peak, and (e) sample peak. (Reprinted from H.-C. Shu, H. Hakanson, B. Mattiasson, *Anal. Chim. Acta*, **300**, 277–285, Copyright (1995), with permission from Elsevier Science.)

titrations, low reagent usage, and flexibility to configure a single instrument for several methodologies.

4.4 Environmental[70–77]

Process monitoring technology in general provides tighter process control which results in a more efficient process, fewer plant upsets, less off-specification product, and the minimization of waste. Waste minimization is often viewed as primarily an environmental issue, but it can also have a substantial economic pay-off. The following recent example of an at-line SIA process monitor for bioremediation of explosives illustrates this capability.

A spectrophotometric method based on SIA for the determination of 2,4,6-trinitrotoluene (TNT) in soil samples was recently reported.[78] A schematic diagram of the instrument is shown in Figure 16. A derivatization reaction of the analyte with potassium hydroxide and sodium sulfite in a primarily acetone media was the basis of the chemistry of the method. This reaction was found to be sensitive to the amount of acetone in which the reaction took place. The percentage of acetone and water used in the reaction, sample and reagent volumes, mixing coil volume and reaction time were studied with the goal of obtaining optimal working conditions for the method. A percentage of 88% v/v acetone and 12% v/v water was used in the reaction. A study of the response of the method to other explosives (an interference study) demonstrated that the method is selective for TNT. Quantitative results from an analysis of contaminated soil samples were compared to those obtained using a standard HPLC method for TNT. The data showed that there was no statistical difference between the two methods. The precision of the method was established at 5.5% (for replicate soil samples). The limit of detection was $0.5\,\mu g\,mL^{-1}$ for aqueous standards and approximately $80\,\mu g\,mL^{-1}$ for 300 mg soil samples. The extension of the work to contaminated environmental water samples and the means by which samples could be preconcentrated using the sequential injection method were also developed.

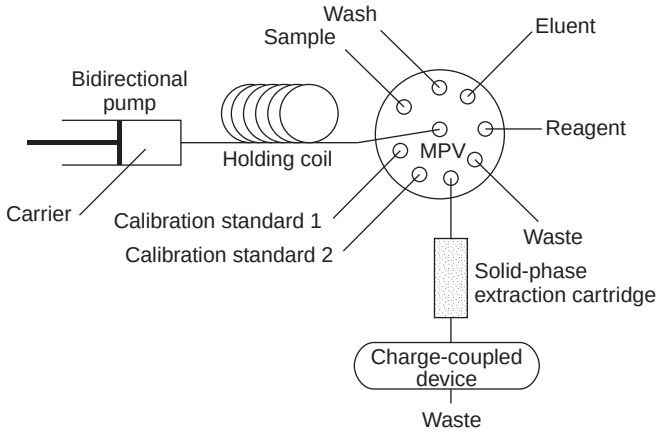

Figure 16 Schematic diagram of an at-line SIA process monitor for determining the secondary explosive TNT in the effluent of a bioreactor.

5 STRENGTHS AND WEAKNESSES

The key strengths of process FIA include versatility, robustness, rapid response, high precision, high reliability, broad concentration range, and application to liquid and gas streams. Its versatility, in particular, provides process FIA with the potential of becoming a modular generic process monitor platform that could replace a large variety of different types of monitors currently in use.

In recent years, there has been a growing emphasis on more and better process analysis technology, and the driving force for this trend is increasingly intense global competition. The name of the game is the production of high-quality products at a minimum cost and in a manner that does not adversely affect the environment. This requires tight process control, and to control something well it is necessary to be able to measure it accurately and reliably. A well-designed and well-placed process FIA or SIA instrument can provide a substantial return on investment. The tighter process control provides a more efficient process, reduction of plant upsets, less off-specification product, and waste minimization.

There are, of course, many different types of process monitor technologies available to address these needs. In fact, it would appear that there are too many different types and this is the source of one of the problems that adds to the high cost of process monitors and the cost of maintaining them. For example, at a typical chemical plant familiar to one of the authors, there are 25 process monitors in place, and of these 25, there are 17 different types. Instrument technicians must be trained on all of the different types, and parts have to be stocked for all of them as well. Because of the great versatility of FIA, a significant number of these monitors could be replaced with a single, modular process FIA platform, simplifying training and maintenance. It may be possible in certain instances to perform more than one method on a single monitor platform, particularly for SIA. Some of the other important strengths of SIA are:

- versatility
- rapid analytical readout
- high reliability
- high precision and calibration stability
- broad concentration range
- low reagent consumption and waste generation

- multicomponent determination capability
- capability to monitor liquid or gaseous streams.

In the early years of process FIA, and this is still the case to a significant extent, a major barrier to the acceptance of the technique in plants was the fact that it is based upon wet chemistry. Plant operators and monitor support staff are very resistant to monitors based on wet chemistry methodologies. They do not like to handle solutions and waste or keep reagent reservoirs full. In fact, one of the major causes of downtime with on-line process FIA in plants is systems running dry because of the failure to refill reservoirs, even though that is generally a very simple task on the routine maintenance list.

As a result of this barrier, plants will, if at all possible, install conventional, well-established process monitor technology that does not involve wet chemistry, such as gas chromatography. The major reason process FIA has made inroads in on-line analyses is that in many important applications it can perform analyses that are impossible or very difficult with other process monitor technologies. If the payback was high enough, a plant would be willing to take a chance on FIA. An example is an application described in a 1994 publication[79] involving on-line measurement of trace hydrogen cyanide. The prototype monitor monitored trace HCN in refinery gas streams from thermal and catalytic cracking units that contained high levels of hydrogen sulfide. Conventional process monitor technology, such as gas chromatography, could not perform this determination reliably on-line. A novel on-line FIA monitor provided a unique solution which could monitor hydrogen cyanide in these streams down to 50 ppbv. Once process FIA has demonstrated its strengths in an application such as this, the barrier to additional applications generally fades.

Another of the strengths of on-line process FIA is its rapid analytical measurement, generally less than a minute for most system designs. Significant changes can occur on relatively short timescales (<30 min) for certain industrial processes. Other strengths include high reliability (typically >98% up-time), high precision (typically 0.5% RSD or better), and calibration stability (typically, calibrations hold for 1–14 days). Most on-line FIA monitors have built-in calibration checks which automatically recalibrate when they are out of specification. Another important strength is a broad concentration range, from a few ppb to several wt% for many instances. For concentrated streams, FIA can perform an automatic on-line dilution. FIA is generally used for liquid streams but, as already shown above for the HCN example, it can also be applied to gas stream monitoring in certain cases using MSDs.

Any critical discussion of the strengths of a process monitor technology should also include some discussion of its weaknesses or barriers to its implementation. These include:

- high-maintenance wet-chemistry technique
- difficult to identify noncorrosive chemistries
- difficult to efficiently introduce the sample
- few commercial suppliers of process FIA/SIA instrumentation.

One weakness that has already been mentioned is that FIA uses wet-chemistry technology and therefore requires some special attention by operators, such as ensuring that reagent reservoirs do not run dry. A dried-out FIA monitor generally requires a major effort to restore and re-equilibrate. Another barrier, also a wet-chemistry problem, is finding chemistries that are stable and benign. There is a rich resource of FIA chemistries to tap, but only a fraction of these are suitable for process FIA. The reagents must be stable for extended periods of time, and they must not be even slightly corrosive to FIA hardware. It can be surprising how many supposedly stable and noncorrosive reagents can, over extended periods of time, slowly corrode valves and pumps, or deposit decomposition products or precipitates that clog FIA lines.

Finally, a well-designed sampling system is critical to a successful on-line FIA, and if it is not included the system is very likely to fail. Although this is the case for most process monitors, it may be particularly true for process FIA because this technique uses narrow-bore tubing and thus is susceptible to particles and bubbles.

6 FUTURE DIRECTIONS

The 1980s brought a surge in innovation, development, and deployment of process FIA. Many large companies had multiple installations; Dow Chemical, for example, at one count had over 30 installed process FIA monitors. Because of the lack of commercially available state-of-the-art process FIA, pioneers in the field, primarily from the chemical and refining industries, developed their own technology in-house during this era.

The end of the 1990s has seen a new wave of interest and use of process FIA, and this is likely to grow into the new millennium. The technology is being embraced by a broader range of process monitoring applications than in the past. A number of factors are at work which account for the increased interest in the technique, in particular the commercial availability of advanced process FIA technology and technical support. The Center for Process Analytical Chemistry at the University of Washington has FIA as one of its research areas, and this has also had an impact on the awareness, development, and application of the technology.

A definite trend toward a preference of SIA over FIA, accompanied by an upswing in the development of new SIA strategies and technology, is likely to accelerate in the future. New components, such as 26-port selection valves and pulseless bidirectional liquid pumps, now allow users to take fuller advantage of the versatility of SIA. Combinations of SIA and FIA are also beginning to appear which simplify a complex FIA technique, and this trend will probably continue.[80]

Future developments will be determined by present research efforts and the authors have noted several new application arenas. The research being conducted in these areas is performance-driven rather than of an esoteric nature. Some examples are given to illustrate this point.

- Some intelligence is currently built into many process FIA monitors, such as self-checking of calibration with decision to recalibrate if out-of-specification. Additional intelligence is likely to be developed and applied in future technology, such as if a measurement falls outside the calibration range to automatically dilute or take a larger sample, preconcentrate, etc.
- Researchers who are developing systems to be used in the environmental field are working at developing portable and robust monitors that can be left unattended for extended periods of time often in rather inhospitable environs. (e.g. at the bottom of lakes, on buoys, and in remote wilderness locations). The design advances used in these rugged instruments would also benefit the industrial process monitoring field.
- Complementary to the push for portability is a simultaneous effort toward miniaturization. Much work is ongoing in various quarters toward producing the laboratory-on-a-chip. Although such completely integrated devices potentially offer much enhanced capability in clinical and environmental analysis, their applicability to process analysis is unlikely to be as successful because of the vast variability in the nature of samples that are encountered. An approach of greater applicability would be integrated unit operations, in which small modules dedicated to specific actions (e.g. filtration) are developed which can be easily mixed and matched to accommodate sample-to-sample variability.
- MSDs, now commercially available but underutilized, are likely to become more widespread to simplify and enhance sampling systems for both liquid and gas streams.
- In situ generation of unstable reagents from stable solutions – either by reaction chemistry or by electrochemistry. This is, of course, already used in ion chromatography to generate eluants. The same approach can be used for many reagents, but particularly oxidants/reductants whose shelf-life is such that their use is normally precluded from process monitoring. The more reagent usage is decreased, the more important the shelf-life variable will become – the true barrier to implementation is not reagent consumption per se, but rather the maintenance interval that is required.
- There is an increasing use of embedded processors and this will continue as these systems become more robust and require less expert intervention.
- The extensive use of intranets, the internet, and extranets, as well as telemetry will increasingly place analytical results instantly in the hands of those who most need them. This will increase the use of analytical data and its conversion into useful information. The speed with which these data will be made available both by virtue of the proximity of the monitor to the sample source and the convenient transfer of analytical results will enable rapid and timely decision-making. We expect this capability to extend to compliance monitoring of industrial pollutant discharges as well.
- Also to be expected, using the same connectivity technology, is remote instrument diagnostics and maintenance. Part of the process of making instruments more robust will be in developing systems to allow remote diagnostics and even repair. No longer will the user be expected to understand the inner workings of the monitor; rather, the instrument manufacturer will be able to perform diagnostic tests from a remote location and then effect changes to the control system or even activate redundant back-up systems while defective components are returned for repair.

In summary, the future will see versatile, small-footprint, robust, and intelligent FIA/SIA monitors that will be deployed in an increasingly wide range of process applications.

ACKNOWLEDGMENTS

We thank Bruce Davis of Boeing for contributing an example of his group's process monitoring research.

ABBREVIATIONS AND ACRONYMS

CAN	Controller Area Network
DCS	Distributed Control System

FIA	Flow Injection Analysis
HPLC	High-performance Liquid Chromatography
ICP/MS	Inductively Coupled Plasma/ Mass Spectrometry
MPV	Multiposition Valve
MSD	Membrane Sampling Device
PEEK	Polyether Ether Ketone
RSD	Relative Standard Deviation
SIA	Sequential Injection Analysis
TNT	2,4,6-Trinitrotoluene

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Flow-injection Techniques in Environmental Analysis

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis • Titration Techniques for Process Analysis

Steel and Related Materials **(Volume 10)**
Automation of Analytical Control in the Steel and Metals Industry

Atomic Spectroscopy **(Volume 11)**
Flow Injection Analysis Techniques in Atomic Spectroscopy

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. J. Ruzicka, E.H. Hansen, 'Flow Injection Analysis. 1. A New Concept in Fast Continuous Flow Analysis', *Anal. Chim. Acta*, **79**(1), 145 (1975).
2. C.B. Ranger, 'Flow Injection Analysis: A New Approach to Near Real-Time Process Monitoring', *Autom. Stream Anal. Process Control*, **1**, 39–67 (1982).
3. R.A. Mowery, Jr, 'Flow Injection Analysis: Potential for Process Composition Measurement', *InTech*, **31**(5), 51–54 (1984).
4. J. Ruzicka, 'Flow Injection Analysis – A Survey of Its Potential for Continuous Monitoring of Industrial Processes', *Anal. Chim. Acta*, **190**, 155–163 (1986).
5. G.D. Christian, J. Ruzicka, 'Analysis Technique Allows On-line Control', *Chem. Eng. (N.Y.)*, **95**(4), 57–60 (1988).
6. M.D. Luque de Castro, 'Continuous Monitoring by Unsegmented Flow Techniques. State of the Art and Perspectives', *Talanta*, **36**(5), 591–599 (1989).
7. E.D. Yalvac, R.A. Bredeweg, 'Status of Process Flow Injection Analysis and Current Trends', *Process Control Qual.*, **6**(2–3), 91–95 (1994).
8. K. Schuegerl, B. Hitzmann, H. Jurgens, T. Kullick, R. Ulber, B. Weigal, 'Challenges in Integrating Biosensors and FIA for On-line Monitoring and Control', *Trends Biotechnol.*, **14**(1), 21–31 (1996).
9. J.J. van der Pol, C.D. de Gooijer, M. Biselli, C. Wandrey, J. Tramper, 'Automation of Selective Assays for On-line Bioprocess Monitoring by Flow Injection Analysis', *Trends Biotechnol.*, **14**(12), 471–477 (1996).
10. J. Ruzicka, G.D. Marshall, 'Sequential Injection: A New Concept for Chemical Sensors, Process Analysis and Laboratory Assays', *Anal. Chim. Acta*, **237**(2), 329–343 (1990).
11. S. Chalk, 'The Flow Analysis Database', University of North Florida, http://www.fia.unf.edu/fad/fad.html.
12. Z. Fang, 'Non-equilibrated Sample Manipulation – The Essence of Flow Injection Analysis', *Microchem. J.*, **45**(2), 137–142 (1992).
13. G.D. Clark, D.A. Whitman, G.D. Christian, J. Ruzicka, 'Sample Handling and Pretreatment using Flow Injection Analysis', *CRC Crit. Rev. Anal. Chem.*, **21**(5), 357–375 (1990).
14. J. Ruzicka, E.H. Hansen, *Flow Injection Analysis*, Wiley, New York, 2nd edition, 1988.
15. Z. Fang, *Flow Injection Separation and Pre-Concentration*, Wiley VCH, Weinheim, 1993.
16. C.H. Pollema, D.L. Campbell, L.E. Moore, 'Carrierless Sequential Injection Analysis', US Patent & Trademark Office, #5849592, 1998.
17. Global FIA, Inc., Technology Note, 'Partial Loop (Timed) Injection', describes how the valve is used in this mode (www.globalfia.com).
18. V. Kuban, 'Gas Diffusion – Permeation Flow Injection Analysis. 1. Principles and Instrumentation', *Crit. Rev. Anal. Chem.*, **23**(5), 323–354 (1992).
19. W. Frenzel, 'Permeation Denuder for Sampling and Continuous Analysis of Gases. 1. System Configuration, Basic Studies and Application to Atmospheric Ammonia and Sulfur Dioxide', *Anal. Chim. Acta*, **291**(3), 305–320 (1994).
20. P.G. Baxter, J. Ruzicka, G.D. Christian, 'Apparatus for the Determination of Volatile Analytes by Stopped-flow Injection Analysis using an Integrated Fibre Optic Detector', *Talanta*, **41**(3), 347–354 (1994).
21. L.C. van Nugteren-Osinga, M. Bos, W.E. van der Linden, 'Hollow-fibre Membrane for Sample Introduction in a Flow Injection System. Determination of Carbon Disulfide in Air', *Anal. Chim. Acta*, **226**(1), 171–175 (1989).

22. V. Kuban, P.K. Dasgupta, 'Selective Determination of Gases by Two-stage Membrane-differentiated Flow Injection Analysis. Determination of Trace Hydrogen Cyanide in the Presence of Large Concentrations of Hydrogen Sulfide', *Anal. Chem.*, **64**(10), 1106–1112 (1992).
23. Z. Genfa, P.K. Dasgupta, 'Fluorometric Measurement of Aqueous Ammonium Ion in a Flow Injection System', *Anal. Chem.*, **61**(5), 408–412 (1989).
24. G.D. Christian, 'Sequential Injection Analysis for Electrochemical Measurements and Process Analysis', *Analyst (Cambridge, UK)*, **119**(11), 2309–2314 (1994).
25. G.D. Marshall, J.F. van Staden, 'Operational Parameters Affecting Zone Penetration in Sequential Injection Analysis', *Process Control Qual.*, **3**(1–4), 251–261 (1992).
26. A. Ivaska, 'Estimation of the Optimal Sampling Frequency for Process Analyzers', *Anal. Chim. Acta*, **190**, 89–97 (1986).
27. K. Wendt, Lachat Instruments, Inc., Internal Company Report, 1996.
28. N. Sherry, Chlorine Institute Proc., Abstract 258-73I/B, 1979.
29. *Methods Used and Adopted*, Association of Florida Phosphate Chemists, 7th edition, Section 14–1, 1991.
30. *Official Methods of Analysis*, 15th edition, Association of Official Analytical Chemists, 1990.
31. D. Pletcher, F. Walsh, *Industrial Electrochemistry*, 2nd edition, Blackie Academic & Professional, 173–209, 1993.
32. E.A. Houser, 'Principles of Sample Handling and Sampling Systems Design for Process Analysis', Instrument Society of America, 1972.
33. T. Miller, T. Stevens, 'A New Process Analysis Approach for Acid or Base Concentration: Differential Conductivity-Ion Exchange', *Adv. Instrum.*, **35**(2), 21–27 (1980).
34. E.A. Jones, D. Barnes, J.F. van Staden, 'Determination of Cyanide by Flow Injection Analysis using a Tubular Solid-state Ion-selective Electrode as a Detector', *S. Afr. J. Chem.*, **42**(1), 25–28 (1989).
35. E.A. Jones, M.J. Hemmings, 'Spectrophotometric Determination by Flow Injection Analysis of Thiocyanate in Metallurgical Process Solutions Containing Free Cyanide and Metal Cyanide Complexes', *S. Afr. J. Chem.*, **42**(1), 6–8 (1989).
36. J. Neubauer, K. Maier, 'Comparison of the FI-hydride System with a Continuous Flow System for On-line Determination of Bismuth', *At. Spectrosc.*, **13**(6), 206–207 (1992).
37. C.K.Y. Lam, Y. Zhang, M.A. Busch, K.W. Busch, 'Design and Performance of a New Continuous-flow Sample Introduction System for Flame Infrared-Emission Spectrometry: Applications in Process Analysis, Flow Injection Analysis, and Ion-exchange High-performance Liquid Chromatography', *Talanta*, **40**(6), 867–878 (1993).
38. Y. Hayashibe, Y. Sayama, 'Determination of Trace Amounts of Cadmium in a Hydrometallurgical Zinc Refining Process Stream by a Flow Injection Method with On-line Preconcentration and Spectrophotometric Detection', *Analyst (Cambridge, UK)*, **121**(1), 7–11 (1996).
39. X.-S. Chai, L.-G. Danielsson, 'Process Monitoring During Eectrochemical Production of Polysulfides by On-line FIA using a Capillary Flow-Through Cell', *Process Control Qual.*, **8**(4), 159–166 (1996).
40. F. Albertus, I. Cortes, L.-G. Danielsson, F. Ingman, 'Determination of Total Ammonia and Carbon Dioxide in Liquors from the Caron Process for Nickel Extraction', *Process Control Qual.*, **8**(4), 185–192 (1996).
41. J.F. van Staden, H. du Plessis, 'Sequential Injection Acid-Base Titration – A Useful Application for Process Analytical Science', *Anal. Commun.*, **34**(5), 147–151 (1997).
42. (a) M. Gisin, C. Thommen, 'Industrial Process Control by Flow Injection Analysis', *Anal. Chim. Acta*, **190**(1), 165–176 (1986); (b) M. Gisin, C. Thommen, 'Contemporary Wet-chemical Flow Analyzers for Process Control', *Trends Anal. Chem.*, **8**(2), 62–66 (1989).
43. I. Ogbomo, U. Prinzing, H.L. Schmidt, 'Prerequisites for the On-line Control of Microbial Processes by Flow Injection Analysis', *J. Biotechnol.*, **14**(1), 63–70 (1990).
44. M.T. Jeppesen, E.H. Hansen, 'Flow Injection Fluorometric Assay of Nitrogen-containing Substrates by On-line Enzymic Generation of Ammonia', *Anal. Chim. Acta*, **245**(1), 89–99 (1991).
45. K. Schuegerl, L. Brandes, T. Dullau, K. Holzhauer-Rieger, S. Hotop, U. Huebner, X. Wu, W. Zhou, 'Fermentation Monitoring and Control by On-line Flow Injection and Liquid Chromatography', *Anal. Chim. Acta*, **249**(1), 87–100 (1991).
46. L.W. Forman, B.D. Thomas, F.S. Jacobson, 'On-line Monitoring and Control of Fermentation Processes by Flow Injection Analysis', *Anal. Chim. Acta*, **249**(1), 101–111 (1991).
47. M. Nilsson, H. Haakanson, B. Mattiasson, 'Flow Injection ELISA for Process Monitoring and Control', *Anal. Chim. Acta*, **249**(1), 163–168 (1991).
48. I. Ogbomo, R. Kittsteiner-Eberle, U. Englbrecht, U. Prinzing, J. Danzer, H.L. Schmidt, 'Flow Injection Systems for the Determination of Oxidoreductase Substrates: Applications in Food Quality Control and Process Monitoring', *Anal. Chim. Acta*, **249**(1), 137–143 (1991).
49. U. Spohn, R. Eberhardt, B. Joksch, R. Wichmann, C. Wandrey, H. Voss, 'Enzymic Multichannel-FIA Methods for On-line Fermentation Monitoring and Control', *GBF Monogr.*, **14**(Flow Injection Anal. Based Enzymes Antibodies), 51–62 (1991).
50. T. Dullau, K. Schuegerl, 'Process Analysis and Control by Enzyme-FIA-Systems', *GBF Monogr.*, **14**(Flow Injection Anal. Based Enzymes Antibodies), 27–39 (1991).
51. S.C. Chung, G.D. Christian, J. Ruzicka, 'Sequential Injection Methodology for Fermentation Monitoring', *Process Control Qual.*, **3**(1–4), 115–125 (1992).
52. M. Nilsson, H. Haakanson, B. Mattiasson, 'Process Monitoring by Flow Injection Immunoassay: Evaluation of a

Sequential Competitive Binding Assay', *J. Chromatogr. B*, **597**(1–2), 383–389 (1992).

53. U. Schrader, V. Denk, H.L. Schmidt, L. Narziss, 'Flow Injection Analysis Based on Biosensors to Improve the Automation and Control of Food Production Processes', *GBF Monogr.*, **17**(Biosens.: Fundam., Technol. Appl.), 119–122 (1992).
54. H. Luedi, M.B. Garn, S.D. Haemmerli, A. Manz, H.M. Widmer, 'Flow Injection Analysis and In-line Biosensors for Bioprocess Control: A Comparison', *J. Biotechnol.*, **25**(1–2), 75–80 (1992).
55. M. Peris, A. Maquieira, R. Puchades, V. Chirivella, R. Ors, J. Serrano, A. Bonastre, 'An Expert System as a Tool for the Specification and Intelligent Control of a Flow Injection Analysis System', *Chemom. Intell. Lab. Syst.*, **21**(2–3), 243–247 (1993).
56. (a) N.W. Barnett, D.G. Rolfe, T.A. Bowser, T.W. Paton, 'Determination of Morphine in Process Streams using Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **282**(3), 551–557 (1993); (b) N.W. Barnett, S.W. Lewis, D.J. Tucker, 'Determination of Morphine in Process Streams by Sequential Injection Analysis with Chemiluminescence Detection', *Fresenius' J. Anal. Chem.*, **355**(5–6), 591–595 (1996); (c) N.W. Barnett, C.E. Lenehan, S.W. Lewis, D.J. Tucker, K.M. Essery, 'Determination of Morphine in Water Immiscible Process Streams using Sequential Injection Analysis Coupled with Acidic Permanganate Chemiluminescence Detection', *Analyst (Cambridge, UK)*, **123**(4), 601–605 (1998).
57. E. Rosenberg, R. Kellner, 'Flow Injection Analysis with Fourier Transform Infrared Detection for Clinical and Process Analysis', *Fresenius' J. Anal. Chem.*, **348**(8–9), 530–532 (1994).
58. M. Nilsson, A.R. Vijayakumar, O. Holst, C. Schornack, H. Haakanson, B. Mattiasson, 'On-line Monitoring of Product Concentration by Flow-ELISA in an Integrated Fermentation and Purification Process', *J. Ferment. Bioeng.*, **78**(5), 356–360 (1994).
59. (a) H. Juergens, R. Kabuss, T. Plumbaum, B. Weigel, G. Kretzmer, K. Schuegerl, K. Andres, E. Ignatzek, F. Giffhorn, 'Development of Enzyme-cartridge Flow Injection Analysis for Industrial Process Monitoring. Part I. Development and Characterization', *Anal. Chim. Acta*, **298**(2), 141–149 (1994); (b) H. Juergens, L. Brandes, R. Joppien, M. Siebold, J. Schubert, X. Wu, G. Kretzmer, K. Schuegerl, 'Development of Enzyme Cartridge Flow Injection Analysis for Industrial Process Monitoring. Part II. Application for Monitoring of Microorganism Cultivations', *Anal. Chim. Acta*, **302**(2–3), 289–296 (1995); (c) H. Juergens, R. Akhnoukh, G. Kretzmer, K. Schuegerl, 'Development of Enzyme-cartridge Flow Injection Analysis for Industrial Process Monitoring. Part III. Application for Monitoring of Recombinant Animal Cell Cultivations', *Anal. Chim. Acta*, **315**(1–2), 153–157 (1995).
60. C. Menzel, T. Lerch, T. Scheper, K. Schuegerl, 'Development of Biosensors Based on an Electrolyte Insulator Semiconductor-Capacitor Structure and Their Application for Process Monitoring. Part I. Development of the Biosensors and their Characterization', *Anal. Chim. Acta*, **317**(1–3), 259–264 (1995).
61. C. Tran-Minh, 'Biosensors in Flow Injection Systems for Biomedical Analysis, Process and Environmental Monitoring', *J. Mol. Recognit.*, **9**(5/6), 658–663 (1996).
62. N.W. Barnett, T.A. Bowser, R.D. Gerardi, B. Smith, 'Determination of Codeine in Process Streams using Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chim. Acta*, **318**(3), 309–317 (1996).
63. F. Lammers, T. Scheper, 'On-line Monitoring of Enzyme-Catalyzed Biotransformations with Biosensors', *Enzyme Microb. Technol.*, **20**(6), 432–436 (1997).
64. X.J. Tang, A. Tocaj, O. Holst, G. Johansson, 'Process Monitoring of Acetic Acid in *Escherichia coli* Cultivation using Electrochemical Detection in a Flow Injection System', *Biotechnol. Tech.*, **11**(9), 683–687 (1997).
65. H.-C. Shu, 'Stop-flow Sequential Injection Analysis using Immobilized D-Lactate Dehydrogenase for On-line D-Lactic Acid Monitoring During a Fermentation Process', *Spec. Publ. – Royal Soc. Chem.*, **167**(Biosensors for Food Analysis), 144–153 (1998).
66. R.W. Min, V. Rajendran, N. Larsson, L. Gorton, J. Planas, B. Hahn-Hagerdal, 'Simultaneous Monitoring of Glucose and L-Lactic Acid During a Fermentation Process in an Aqueous Two-Phase System by On-line FIA with Microdialysis Sampling and Dual Biosensor Detection', *Anal. Chim. Acta*, **366**(1–3), 127–135 (1998).
67. J. Rhee, J. Hagedorn, G. Kretzmer, T. Scheper, K. Schugerl, 'Flow Injection Immunoassay using OPA Derivatization', *Anal. Chim. Acta*, **374**(2–3), 177–183 (1998).
68. H.-C. Shu, H. Hakanson, B. Mattiasson, 'On-line Monitoring of D-Lactic Acid during a Fermentation Process using Immobilized D-Lactate Dehydrogenase in a Sequential Injection Analysis System', *Anal. Chim. Acta*, **300**(1–3), 277–285 (1995).
69. B. Davis, Operations Technology, Boeing Commercial Airplane Group, personal communication, 1999.
70. W. Frenzel, 'Potential of Modified Reverse Flow Injection Analysis for Continuous Monitoring and Process Control', *Fresenius' Z. Anal. Chem.*, **329**(6), 668–674 (1988).
71. R.L. Benson, P.J. Worsfold, F.W. Sweeting, 'Spectrophotometric Flow Injection Techniques for Process Monitoring', *Anal. Proc. (London)*, **26**(11), 385–387 (1989).
72. M.L. Balconi, F. Sigon, M. Borgarello, R. Ferraroli, F. Realini, 'Flow Injection Analysis for Power Plants: Evaluation of Detectors for the Determination of Control Parameters in Conditioned Water-Steam Cycles', *Anal. Chim. Acta*, **234**(1), 167–173 (1990).

73. G.D. Marshall, J.F. van Staden, 'Analyte Enrichment using Sequential-Injection Analysis', *Instrum. Sci. Technol.*, **25**(4), 307–320 (1997).
74. L. Arce, A. Rios, M. Valcarcel, 'Flow Injection-Capillary Electrophoresis Coupling to Automate On-line Sample Treatment for the Determination of Inorganic Ions in Waters', *J. Chromatogr. A*, **791**(1–2), 279–287 (1997).
75. J.F. van Staden, R.E. Taljaard, 'On-line Monitoring of Phosphate in Natural Water and Effluent Streams using Sequential Injection Analysis', *Mikrochim. Acta*, **128**(3/4), 223–228 (1998).
76. J.F. van Staden, T.A. van der Merwe, 'On-line Monitoring of Nitrite in Fertilizer Process Streams, Natural and Wastewater Effluents with Sequential Injection Analysis', *Mikrochim. Acta*, **129**(1–2), 33–39 (1998).
77. J.W. Grate, O.B. Egorov, 'Automating Analytical Separations in Radiochemistry', *Anal. Chem.*, **70**(23), 779A–788A (1998).
78. R.T. Echols, M. Christensen, R. Krisko, J.H. Aldstadt, 'Selective Determination of TNT in Soil Extracts by Sequential Injection Spectrophotometry', *Anal. Chem.*, **71**(14), 2739–2744 (1999).
79. D.C. Olson, S.R. Bysouth, P.K. Dasgupta, V. Kuban, 'New Flow Injection Analyser for Monitoring Trace Hydrogen Cyanide in Process Gas Streams', *Process Control Qual.*, **5**(4), 259–265 (1994).
80. M.D. Luque de Castro, A. Izquierdo, 'Flow Injection Stripping Analysis', *Electroanalysis*, **3**(6), 457–467 (1991).
81. E.R. Gonzalo, J.L. Pavón, J. Ruzicka, G.D. Christian, D.C. Olson, 'Flow Injection Analysis Determination of Phenols in Kerosine and Naphtha by Membrane Extraction-preconcentration', *Anal. Chim. Acta*, **259**(1), 37–44 (1992).
82. F. Zhang, P.K. Dasgupta, Y.-S. Cheng, 'Design of a Straight Inlet Diffusion Scrubber. Comparison of Particle Transmission with Other Collection Devices and Characterization for the Measurement of Hydrogen Peroxide and Formaldehyde', *Atmos. Environ.*, **25A**, 2717–2729 (1991).
83. S. Motomizu, K. Toei, T. Kuwaki, M. Oshima, 'Gas Diffusion Unit with Tubular Microporous Poly(tetrafluoroethylene) Membrane for Flow Injection Determination of Carbon Dioxide', *Anal. Chem.*, **59**(24), 2930–2932 (1987).
84. E.B. Milosavljevic, L. Solujic, J.L. Hendrix, J.H. Nelson, 'Flow Injection Gas Diffusion Method for Preconcentration and Determination of Trace Sulfide', *Anal. Chem.*, **60**(24), 2791–2796 (1988).
85. T.J. Cardwell, R.W. Cattrall, M. Mitri, 'Gas Diffusion with Preconcentration for the Determination of Fluoride in Water Samples by Flow Injection', *Talanta*, **41**(1), 115–123 (1994).
86. M.E. Meyerhoff, Y.M. Fraticelli, 'Flow Injection Determination of Ammonia Nitrogen using Polymer Membrane Electrode-based Gas Sensing System', *Anal. Lett.*, **14**(6B), 415–432 (1981).
87. P.K. Dasgupta, 'Diffusion Scrubber for the Collection of Atmospheric Gases', *Atmos. Environ.*, **18**(8), 1593–1599 (1984).
88. Q. Fan, P.K. Dasgupta, 'Continuous Automated Determination of Atmospheric Formaldehyde at the Parts Per Trillion Level', *Anal. Chem.*, **66**(4), 551–556 (1994).

Infrared Spectroscopy in Process Analysis

John P. Coates
Coates Consulting, Newtown, USA

Paul H. Shelley
Boeing Commercial Airplane Group, Seattle, USA

1 Introduction 1
2 The Basics of Infrared Spectral Measurements 3
3 Instrumentation Basics 4
3.1 Component Technologies 4
4 Process Analytical Instruments 7
4.1 Commercial Process Instrumentation: Chromatographs 8
4.2 Commercial Process Instrumentation: Photometers and Spectrometers 8
4.3 System Requirements for Process Infrared Analyzer Systems 11
4.4 Sample Handling in a Production Environment 17
4.5 Factors Influencing Acceptance of Spectrometer-based Analyzers 18
5 Applications of Process Infrared Spectroscopy 19
6 Process Analysis: A Review 19
6.1 Process Spectroscopy – a Reality Check 21
7 Future Trends and Directions 22
Abbreviations and Acronyms 23
Related Articles 23
References 23

Infrared (IR) instruments have been in use as important measurement and analysis tools for process monitoring and production control for more than half a century. They range in complexity from the simplest filter-based photometers to opto-mechanically complicated devices, such as Fourier transform infrared (FTIR) instruments. Simple nondispersive infrared (NDIR) instruments are in common use for systems where one to a few main analytes are involved. For more complex measurement situations it is necessary to record more spectral information, and often full-spectral analyzers are used. Of the main analytical instrument techniques, IR analysis is the most versatile, where all physical forms of a sample may be considered – gas, liquids, solids and mixed phase materials. A wide range of sample handling options exists, and many of these may be adapted to either near-line/at-line production control or on-line process monitoring applications. Instrumentation used within a manufacturing environment usually requires special packaging to protect the instrument from the environment, and to protect the operating environment from the instrument. This places special constraints on the way that an instrument is designed and the way that it performs. Often, it is necessary to utilize special technologies, normally not required for laboratory instruments, to meet the performance objectives. For most practical installations within a production environment it is a normal requirement for the final instrument to comply with local manufacturing and electrical safety codes. The range of applications of IR spectroscopy is extremely broad and the technique can play an important role in virtually all industries for manufacturing quality and production control, and for environmental monitoring.

1 INTRODUCTION

Process IR analysis and the associated instrumentation are very broad and diverse topics that can be defined in a number of different ways. The instrument may be viewed as something very simple, such as a rudimentary measurement device, or as a complex, multifunctional measurement tool. In the end it depends on the application and the needs of the end-user. For many traditional applications the IR instrument merely provides a single-point measurement, and as such may be considered to be acting as a sensor, similar in context to a pressure transducer or a thermometer. In such cases one is normally not considering the measurement to be a true spectral measurement. Over the past two decades, there has been a trend to use analytical-style instrumentation for process measurements – for analysis, monitoring and control. This has resulted in the use of the term process analytical instruments (PAI).[1] Process spectrometers are an important class of these instruments, and instrumentation used for vibrational spectroscopy, in particular IR spectroscopy, accounts for a significant number of the systems that are implemented for process measurements (both on-line and off-line).[2]

For nearly half a century, vibrational spectroscopy, initially in the form of mid-IR spectroscopy, has been featured extensively in industrial analyses, both for quality control (QC) and process monitoring applications. Next to chromatography, it is the most widely purchased instrumentation for these measurements and analyses.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Spectroscopic methods in general are favored because they are relatively straightforward to apply and to implement. Furthermore, a single IR instrument, in a QC environment may serve many functions, whereas chromatographs (gas and liquid) tend to be dedicated to only a few methods at best.

Historically, the first industrial applications of IR spectroscopy were for quality and production control in the petrochemical industries, primarily for the analysis of fuels, lubricants and polymers. Early instruments were configured solely for mid-IR absorption measurements. Over the years, the other important vibrational spectroscopy techniques, namely Raman and near-infrared (NIR), have grown in popularity. Raman, as an experimental concept, was understood as early, if not earlier than mid-IR. However, there were significant difficulties encountered in its early implementation, and as a result, it has only been in recent years that it can be considered as a viable technique for process-related applications. Early on, NIR was never really considered to be an analytical technique with any practical value, and in fact it was originally addressed as an extension of the visible spectral region, rather than the IR. It was not until work was performed in the 1970s with agricultural and food products that the value of this spectral region was appreciated, in particular for quantitative measurements.[3]

Today, the value, functionality and benefits offered by all three of the major vibrational spectroscopy techniques, IR, both mid-IR and NIR, and Raman, are largely uncontested. Taking into account all of the techniques, virtually any form of sample can be handled, including gases, aerosols, liquids, emulsions, solids, semi-solids, foams, slurries, etc. This is also a major factor when comparing chromatographic methods to spectroscopic methods. In a traditional QC environment, a single spectrometer can be configured to handle a wide range of sample types with a choice of suitable sample-handling devices. The nature of the sample plays an important role in the decision process, and in many ways it is the most demanding element for consideration when assembling a system for an industrial application. This is especially the case if we extend the application to become a full on-line process measurement, where the sample extraction and conditioning can be more expensive than the spectrometer itself.

Since the late 1980s, IR spectroscopy, both in the forms of NIR and FTIR spectroscopy, has been hailed as a major growth area for process spectroscopy. This view assumed that the areas concerned were outside of those covered by traditional laboratory instruments. Overall, this has been very difficult to define in terms of where the boundaries between the laboratory and the process start and end. Some of the confusion arises by the term ''process analysis'' itself. In the most liberal form, it can be considered to be any analysis that contributes in the manufacture of a product. This can range from the quality assessment of incoming raw materials, to the control of the manufacturing process itself, and finally on to the examination of the final product. How this is viewed tends to be industry dependent.

With traditional thinking, process analysis provides feedback to the process control system, thereby ensuring that the product is manufactured correctly. However, in some industries the final quality of the product is the most important controlling parameter, and this requires that analytical controls be placed at all key points in production. Examples here include high-tech manufacturing, such as in the semiconductor industry, and specialty or value-added chemical production, such as in the biotechnology and pharmaceutical industries. In the petrochemical and petroleum industries, which tend to use continuous processes, information that helps to control the process itself tends to be the most important attribute provided by the analysis. Considering all these factors and requirements, it is important to review how the analysis is physically implemented, giving rise to terms such as on-line, off-line, near-line and at-line. In general, these terms should be self-explanatory. However, there are at times some confusion, especially when considering the best way to implement an analysis. Overall, the deciding factors tend to be the turn-around time required for the analysis and the cost and ease of implementation.

In the context of industrial applications, as applied to this article, it is necessary to address the instrumentation as a package or a system. Dependent on the requirements of the analysis and the technique selected, a system can consist of several key components:

- the instrument, which contains a source, analyzing component, detector, analyzer control electronics and data handling (internal or external computer);
- sample handling and interfacing;
- sample extraction and conditioning (mainly a requirement for on-line, continuous monitoring systems);
- environmental enclosure (for hazardous environments); and
- communications to the user/operator and to the process control computer.

Each of these can become a major subject for discussion, and a detailed coverage of all of these items is outside the scope of this article. However, it is intended that this article will provide the reader with an overview of the instrumentation and technology that is available and in common use for process-related measurements featuring IR spectroscopy. Also, general discussion will be provided regarding both the implementation and applications of current instrumentation. It must

be understood that IR spectroscopy does not provide answers for every application, and other techniques, including a technique such as Raman spectroscopy,[4] may provide a better solution. Where appropriate, side references will be provided, in context, to the use of the other vibrational spectroscopy techniques.[5,6]

2 THE BASICS OF INFRARED SPECTRAL MEASUREMENTS

Although IR spectroscopy and IR instrumentation are covered elsewhere in this publication, it is important to define here some of the basic principles of the measurement. This is of particular importance for instruments dedicated for specific applications because it helps to assess the technology and materials that are optimum for a final measurement system. It is important to define the measurement techniques in terms of their spectral range, the units and conventions that are used, and their general utility. While this may seem obvious or basic in context, the boundaries must be defined when discussing measurement technologies and their application in a process environment. In other words, it is essential to match the characteristics of an individual technology to the performance required for the target application.

Initially, the term IR (in the context of instruments) was applied to the spectral region from around 2.5 μm to approximately 16.5 μm. At the time, this was dictated by the optical systems in use, which were based originally on sodium chloride optics. Later, the range was expanded down to 25 μm, and today this is generally accepted as the mid-IR region (2.5 to 25 μm). Instruments can provide measurements out to longer wavelengths; this region is known as the far IR, and generally it has limited utility for routine or industrial applications. Traditionally, an energy scale expressed as wavelength, in micrometers, was used, and is still used in certain contexts. Analytically, a frequency scale, which is linked to molecular vibrations, based on wavenumber (cm^{-1}), is preferred. In the mid-IR region (4000–400 cm^{-1}), the most dominant absorptions are associated with the fundamental vibrations of component chemical bonds from functional groups, the backbone structures and molecular fragments of the material under study.[7,8]

The definition of NIR is in some ways slightly more complex, dependent on context, and also when factoring in analytical and practical issues. By convention, it is the spectral region between the end of the visible, approximately 700 nm (0.7 μm) and the start of the IR 2500 nm (2.5 μm). Traditionally, because of the origin and type of instrumentation used, the units of wavelength are used. However, in recent years, mid-IR instruments have been extended into the NIR region, and in such cases wavenumber scales are sometimes used, ranging from around 14 000 cm^{-1} to 4000 cm^{-1}.

From an analytical perspective, the information content of the NIR varies throughout the spectral region, and different regions are used for different applications. In all cases, the measurements utilize absorptions associated with overtones and combinations from the fundamental vibrations that occur in the mid-IR region.[9] For the most part, these regions are dominated by absorptions from the ''hydride'' functionalities – C–H, O–H, N–H, etc. Three groups of information are observed, and these are nominally assigned to first, second and third overtones of the hydride fundamentals, combination bands, plus some coupling from vibrational frequencies of other functional groups. A progression through the three main overtone regions, from first (longest wavelength) to the third (shortest wavelength), has a significant practical impact in regard to the order of magnitude of the intensity of the signal that is measured. The first overtone being at least one order less intense than the fundamentals (mid-IR), and the second and third each being additional orders less than the first. This impacts the practical nature of the measurement because of the increase in effective pathlength that is required to gain adequate spectrum intensity, plus it influences the success of specific modes of measurement – reflectance vs transmission.

From a technology aspect, relative to the actual measurement, the practical spectral range is impacted by the major components of the instrument – source, detector and energy analyzer. Detectors often have the biggest impact, where some detectors work best in the first overtone and the top-end of the mid-IR (such as InAs and the lead salt detectors), others operate in the first and second overtone regions (such as InGaAs), and others only function in the third overtone region (short-wave NIR, such as silicon and charge-coupled device (CCD) arrays). All of the above issues – information content, measurement mode, and the photonics involved, have a significant influence on what is practical for a particular NIR measurement, and in turn, these can impact a user's perception of the definition and utility of NIR. Finally, one minor issue is that in certain contexts, such as in sensors and certain factory measurement systems (such as moisture meters) the term IR is commonly used even though the measurement is made in the NIR.

Both mid-IR and NIR are established in areas where each one provides unique attributes for application in both QC and process/production measurements. There are many factors that determine the selection of the preferred technique for an application, and these can range from subjective issues, such as personal bias to practical issues, such as ease-of-use, sensitivity, accuracy, size and

cost. One or more of these factors influences the manner in which instruments are applied today. To a large extent this is impacted by the conventional approach that has been adopted for the development of instrumentation, that is, an industrial analyzer tends to evolve from a traditional laboratory instrument. Laboratory instruments, by nature, are flexible and multifunctional. The manufacturers intend that an instrument be used for as many different applications as possible. Although this is a benefit in a research-oriented environment, it can be a handicap in production-oriented instrumentation. An analyzer that has evolved in this manner is often an excellent concept model, but is not ideal as a final dedicated analyzer. For many applications, a ''traditional'' style of instrument is overkill, and it is important to take a ''*keep it simple*'' or a ''*good enough is OK*'' attitude for a practical solution.

3 INSTRUMENTATION BASICS

As indicated earlier, an optical spectrometer can be separated into a set of basic elements: radiation source, energy analyzer or encoder, detector, data acquisition and control electronics, and computer, consisting of hardware and software, with an appropriate user interface (keyboard/control panel, display and graphics). Also, as noted, consideration of the sample, sampling interface and the associated optical geometry are essential if a successful implementation is to be accomplished. Figure 1 provides a typical schematic for an optical spectrometer with all the key elements in place. All spectrometers conform to this configuration, with subtle variations, dependent on the sample and mode of measurement. For example, in emission measurements, the sample is in essence the source. As shown, the configuration implies that the radiation passes through the sample, and this only applies to transmission measurements. In a reflection configuration, the surface of the sample is illuminated, and the reflected radiation is collected from the same side. One other variant of the theme (as illustrated) is the

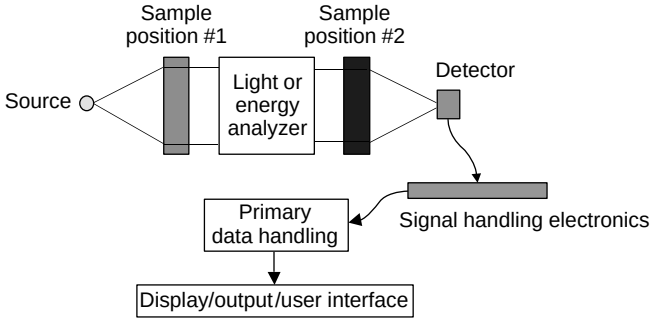

Figure 1 Basic elements of a photometer/spectrometer.

Table 1 Analyzer measurement technologies

Fixed or variable optical filters	NIR, mid-IR, (UV/VIS)
Correlation filters	Mid-IR
Scanning slit-monochromator	NIR, (mid-IR, Raman)
Detector arrays (Si, InGaAs, etc.)	NIR, mid-IR, (UV/VIS)
CCD array	Raman (UV/VIS/NIR)
AOTF[a] (TeO_2)	NIR
Diode laser or tuneable laser	NIR, mid-IR, Raman
Interferometry	(Fourier transform) NIR, mid-IR, Raman

[a] AOTF, acousto-optical tuneable filter; UV/VIS, ultraviolet/visible.

remote sampling approach where the sample is located away from the main spectrometer, and the radiation to and from the sample is carried via optical fibers or some form of light conduit.

There are several measurement techniques that can be considered for use in IR-based instrumentation, and these are summarized in Table 1. Some of these techniques are classical, as in the case of optical filter-based instruments and scanning monochromators. Others, such as AOTF, have been considered for 10 or more years,[10,11] but only now are being utilized to any extent, and currently these are limited to NIR measurements. References to the use of lasers go back more than 30 years, and a few of the most recent applications can be considered as cutting-edge technologies. Their use in the past has been limited by the availability or the cost of the technology involved (examples being diode and tuneable lasers, and certain types of detector arrays).[12–16]

3.1 Component Technologies

This section will be restricted to discussions concerning source and detector technologies. There are other important technologies to consider, especially in the area of data acquisition, control and computer technologies. This is a rapidly changing area, and if viewed generically, serves all forms of instrumentation. Some instrument companies still develop their own proprietary electronics and computer systems. Reasons such as performance optimization, packaging and the need for compliance with certain safety and electronics regulatory codes are often cited. In the latter case, a systems approach is required, especially when attempting to meet the code or performance requirements for compliance with European Certification (CE) mark or electrical and fire safety codes such as National Fire Prevention Association (NFPA) and Cenelec. Often, off-the-shelf electronics can provide the necessary performance characteristics for most applications, and their use can eliminate large expenses related to product development, plus the associated time delays. Furthermore, except in extreme circumstances, nobody would attempt today to develop a custom-built computer

system for a commercial instrument. It is feasible to bring an instrument concept to market for most application areas featuring commercially available components. Only the photonics-related components are addressed in this section because they are the only application-specific components.

The selection of source and detector are interrelated, because obviously the detector must be effective for receiving energy from the source. However, the exact nature of the source is also dependent on the type of sample(s) under consideration, the intended optical geometry, the type of measurement technique, and the final desired performance. In other words, sufficient light must reach the detector to provide a signal-to-noise performance consistent with the required precision for the measurement. This is assuming, of course, that the detector is well matched, optically and performance-wise, and is also capable of meeting the performance specifications.

3.1.1 Polychromatic or Continuum Sources

Most optical spectral measurements, where the measurement of multiple wavelengths is required, will feature some type of polychromatic light. There are a few potential exceptions here, such as tuneable laser sources and source arrays. In such instances, the source is effectively ''*monochromatic*'' at a given point in time. These sources will be covered separately under monochromatic sources.

Almost by definition, sources intended for IR studies are thermal sources, and feature some form of electrically heated element. There are essentially two types of polychromatic source: the open element source and the closed element source. The latter is effectively a light bulb, with the quartz halogen lamp being one of the most important, and is commonly used for NIR applications. This high-temperature source operates around 3000 K and its energy distribution is well optimized for the NIR spectral region. The source also provides a significant amount of energy in the mid-IR region, but the quartz envelope of the bulb does not transmit energy much below 5.5 to 5.0 μm. A variant on the ''*light bulb*'' is an enclosed electrically heated element within a standard electronics package (such as a TO-style (i.e. packaging style) can, as used for detector packaging). The unit is sealed with a low-pressure atmosphere of a protective inert gas, such as nitrogen, by an IR transmitting window, such as potassium bromide (KBr), zinc selenide (ZnSe) or germanium (Ge). The thin filaments used in bulbs and bulb-like sources have a low thermal mass, and this provides the opportunity to pulse the source.

The other form of polychromatic source, used widely in most conventional laboratory-style IR instruments is the exposed, heated element. Various designs have been used, with metal filaments, made from Kanthral® and Nichrome, being simple solutions in lower cost instruments. They tend to be mechanically fragile, especially when subject to vibrations while hot. Electrically heated ceramic sources are popular, and these are more robust and reliable than simple wire sources. These include sources made from silicon carbide, with the Globar® and the Norton sources being common examples. The Norton source is used extensively on modern low-to-medium cost instruments and for instruments used for process applications. It is a simple, rugged source, which was originally used as a furnace igniter. Typical operating temperatures are in the 1100–1300 K range, with the Globar® being an exception at around 1500 K. This higher temperature is seen as a benefit for low light throughput applications, but it can pose thermal problems relative to heat dissipation and sample heating, with the potential for sample degradation.

Most detector systems require that the IR beam be modulated. Modulation of the source energy provides a means of differentiating the measured signal from the ambient background. One of the traditional approaches is to use some form of mechanical ''chopper'', usually in the form of a rotating sector wheel, which modulates the beam by blocking the radiation in one or more sectors during a rotation. An alternative approach is to pulse the source by modulating the electric power to the source, thereby eliminating the need for any form of mechanical chopping device, an important factor in dedicated and process-style instruments. This approach is only practical for low-mass sources, such as wire filament or light bulb style sources described above. Certain measurement techniques, such as interferometry, as used in FTIR instruments, modulate the radiation as a consequence of their mode of operation, and do not require independent modulation of the source.

3.1.2 Monochromatic Sources

The most common ''monochromatic'' source is the laser. Today, this type of source is becoming more popular for a wide range of spectral measurements – in both instruments and dedicated chemical sensors.

Tuneable lasers have always been the dream of spectroscopists. The concept of having a small, high-powered monochromatic IR light source that can be tuned rapidly to other wavelengths would seem to be the ultimate for many applications. Today, there is partial success in this area, and this depends on the application and the spectral region under consideration. Certain solid-state lasers are tuneable over narrow wavelength ranges in the NIR and long-wavelength mid-IR. Tuning is accomplished on a stepwise basis by temperature or pressure programming the device. While the tuning range is typically very small, often no more than a few

nanometers, it is sufficient to be able to scan through the narrow line spectra of certain gas species. The devices are quite useful for the measurement of low concentrations of specific gases and vapors. An array containing several laser elements is feasible for handling gas mixtures. Tuneable laser sources have been used for monitoring species such as hydrogen fluoride, light hydrocarbons, and combustion gases.

Other noncontinuum sources, which are effectively multi-wavelength ''monochromatic'' (because of line width) or multiple line sources, such as gas discharge lamps, have been used, and are often considered to be useful for calibration purposes. Another novel source, which is polychromatic in character, but is *monochromatic* or *species-specific* in behavior, is the molecular equivalent of the electrodeless discharge lamp. The source utilizes a low-pressure gas, the same as the analyte species, and the molecules of the gas are excited within a radiofrequency field. The excitation produces transitions in the vibrational energy states of the molecule, and the source emits radiation in the IR region at the characteristic wavelengths of the molecular species. In this manner, a spectrometric measurement system can be developed where the analytical wavelengths of the analyte are identical and resonantly coupled to that of the source – comparable in nature to an atomic absorption measurement. This is similar in concept to a correlation filter photometer (see later), except that the source provides the filter function. Commercial instrumentation has been developed, based on this technology, for medical applications, but it has potential for use in other applications, such as the measurement of combustion gases.

3.1.3 Single Element Detectors

Traditional spectrometers feature single element detectors. Generally, such instruments utilize only a single detector. Exceptions are certain analyzers that are equipped with detectors dedicated to a single wavelength, either within a spectrograph arrangement, or with dedicated optical filters. The selection of detector is dependent on the spectral region being covered.[12,13,17] There are two main types of detector in use – thermal detectors and photon detectors. In the mid-IR region thermal detectors are used for most routine applications. The most common thermal detectors are the pyroelectrics, such as DTGS (deuterated triglycine sulfate) and lithium tantalate. Lithium tantalate detectors are used widely for low cost sensing devices (including fire alarms), where the detector element can cost as little as $10 (US). They are also used in filter-based instruments and in low-cost scanning instruments. Most routine scanning instruments feature the more expensive and more sensitive DTGS detector, which is somewhere between one and two orders of magnitude more sensitive than lithium tantalate, depending on implementation.

Thermal detectors by nature have a relatively slow response, and are not efficient for fast data acquisition rates. Also, the same applies in certain experiments where there is low energy throughput, as in some reflectance measurements, high-resolution gas analyses, certain fiber optic-based measurements, or where a light-pipe conduit is used for interfacing an instrument to a sampling point. Traditionally, for such applications featuring low light levels or high-speed acquisition, the MCT (mercury–cadmium–telluride) detector is used. This photon responsive detector is operated at sub-ambient temperatures, usually cryogenically cooled to liquid nitrogen temperatures. When operating with liquid nitrogen cooling it is necessary to manually fill a Dewar cold-trap. These normally have an 8-h capacity, but extended fill versions are available, providing up to 24-h usage. Obviously, the requirement to manually fill the Dewar, and the limited capacity, pose limitations on the usage of this style of detector. Although a commercial system exists for autofilling the Dewar, the detector tends to be limited to laboratory-based applications. This type of detector system has been used in a few industrial applications, mainly for high-resolution gas monitoring, chemical reaction studies and specialized semiconductor analyses.

Alternatives to cryogenic cooling include thermo-electric cooling (Peltier devices) and thermo-mechanical cooling (Sterling coolers, recirculating micro-refrigerators and Joule–Thompson (J–T) coolers). None of these solutions are ideal: Peltier coolers have to be operated as three- or four-stage devices, Sterling engine coolers have limited mechanical lifetimes, and J–T coolers often exhibit vibrational coupling problems. Of these, the thermo-electric coolers are often the most straightforward to implement. However, they do not achieve liquid nitrogen temperatures (77 K) which are necessary in order to gain optimum sensitivity. Typically, the devices cool down to about $-75\,^{\circ}C$ (200 K). This limits the spectral range to a band width of around $1000\,cm^{-1}$ in the mid-IR and reduces the detector's performance closer to the DTGS detector. However, the high-speed data acquisition benefits of the detector are retained. Both the three- and four-stage Peltier cooled and Sterling cycle engine cooled MCT detectors have been used for process and environmental monitoring applications with dedicated FTIR instruments.

With the transition into the NIR, the choice of available detector types increases, dependent on the application and the measurement range that is required (as indicated earlier). They are all solid-state photoelectric-style detectors, and they range from the traditional lead salt (PbS and PbSe), germanium and silicon detectors, to binary

and tertiary detectors, such as indium arsenide (InAs), gallium arsenide (GaAs), and indium antimonide (InSb) and indium gallium arsenide (InGaAs). The most popular selections are the lead salts, InGaAs and silicon.

3.1.4 Array Detectors

Until recently, the only detector array elements in common use were the silicon photodiode arrays, and these were initially used for UV/VIS measurements, and later with extended performance into the short-wave NIR. These and CCD arrays increased in availability, thanks to the explosion in the electronic imaging market (fax machines, scanners, digital photocopiers, video and digital cameras, etc.). In recent years, dispersive Raman spectroscopy has gained ground significantly thanks to the availability of high-performance CCD cameras. Recently, there has been a noticeable increase in the commercial availability of miniature spectrometers covering the UV/VIS/NIR spectral region. These are based on relatively low-cost CCD array detectors (with a width as high as 2048 pixels), providing spectral coverage from as low as 190 nm in the UV to 1100 nm (1050 nm) in the short wave NIR. These systems can be purchased for as little as $2000 (US), including provision for fiber-optic coupling.

InGaAs detector arrays operating in the mid-range NIR, developed originally for military imaging applications, have been introduced during the past 5 to 6 years for spectral applications. However, until very recently, these devices lacked consistency (pixel-to-pixel) and were very expensive, even for 128 and 256 element arrays. Note that the number of elements has a direct impact on the relative spectral range and/or the spectral resolution that can be used. Recently, arrays based on lead salts (lead sulfide and lead selenide) have been introduced, and these have provided extended performance into the mid-IR region – down to around 4 μm (2500 cm^{-1}) – which is adequate for measuring the ''hydride'' fundamental (O–H, N–H and C–H) vibrations.

MCT arrays have been used by the military for night vision imaging for many years, and some detector vendors will custom-fabricate these detector arrays. Again, the price of such custom-designed and produced devices is cost-prohibitive, unless there is the opportunity for mass production. So far, a market does not exist for such mass-produced devices, and so their general availability for routine or process instrumentation is currently out of the question. Furthermore, the need for cryogenic cooling has to be considered.

Array detectors operating in the mid-IR spectral region, like tuneable laser devices for the same region, have been a desire for instrumentation producers for some time. It has been known that detector array devices based on DTGS and other pyroelectric materials, including micro-machined devices, have been developed for military applications. Such detectors are expected to become commercially available in the next few years, and these are anticipated to have a significant impact in process-related applications.

4 PROCESS ANALYTICAL INSTRUMENTS

The classical definition of a PAI analyzer is a totally integrated system, implemented on-line, and made up of some or all of the following components:

- instrument, sensor or measurement technology;
- sampling system (if needed);
- signal acquisition and data handling; and
- communications with a process computer.

Often, the term process analyzer is taken too literally, and sometimes nonclassical approaches fail to be considered as practical examples. For example, in the semiconductor industry, traditional analyzers are used as on-line gas and water sensors, but product analyzers are seldom viewed as process analyzers. The semiconductor wafer, the product, is monitored at various stages of the process, effectively *in-line* or *off-line* by FTIR analyzers. It should be pointed out that FTIR is the only accepted method for the QC of the product. In a process that can involve up to 32 critical steps, and where a high yield at each step is essential, it is unreasonable not to consider FTIR as a true process analysis tool. The application is made practical by the use of high-precision wafer-handling robotics.

The traditional industry focus is often on IR photometers, used on-line, and these are mainly used in the environmental market (stack gas and ambient air related applications). There are also a relatively large number of on-line process stream analyzers for liquids and gases (mid-IR) and solids (NIR), that feature either filter (discrete) or full-spectrum analyzers. Many industries are rethinking the role of process analyzers, and applications featuring ''near-line'' or ''at-line'' are now considered to be appropriate especially when the timescale does not require real-time measurements. Many of these analyses are just as legitimate as process measurements as the direct on-line analogues. However, the cost of implementation is often much lower. The final mode of implementation is usually based on the nature of the process, the needs of the end-user, the overall economics, and the required speed of analysis.

4.1 Commercial Process Instrumentation: Chromatographs

While the focus of this article is on IR spectroscopy, it is important to appreciate that spectroscopy is not always the preferred technique for process measurements. To place it in perspective, it is first important to view the technique that has become the process analytical method of choice for many process engineers for nearly 20 years. Process gas chromatography (GC) forms a mainstay for PAI systems in the petrochemicals and chemicals industries, being used for monitoring and control of both gaseous and liquid material production. There are pros and cons – the advantages of GC are that it is a separations technique, it works well for mixtures, and it provides an excellent measurement dynamic range – from ppb/ppm to percent levels. It is also well established, and plant engineers are generally comfortable with the technique. It tends to be a standard for PAI systems throughout most of the petroleum and chemical industries, and it is often the technique of choice, merely on the grounds of familiarity. On the negative side, it is limited to low boiling liquids and gases, and it can have a high cost of maintenance. Sample interfacing and sampling systems can also be an issue, again dependent on application and material type.

Traditional IR process analyzers (filter-based instruments) fail to provide a feasible alternative to GC for most applications. However, modern full-spectrum analyzers, in particular FTIR analyzers, combined with multicomponent analysis software (such as classical least squares and multilinear regression techniques), can provide a good practical alternative for certain applications. Furthermore, removing the need to separate the sample expands the area of applications to mixtures that include nonvolatile and/or thermally labile compounds.

4.2 Commercial Process Instrumentation: Photometers and Spectrometers

As previously noted, spectrometers fall into several categories, from simple optical photometers (single filter devices) to relatively complex full-spectrum devices, such as FTIR instruments. It has been stated, by some process engineers, that there is a preference for using optical methods of measurement rather than chromatographs because of perceived lower maintenance and ease of implementation. However, this may be both system- and application-dependent.

Single component analyzers, in the form of NDIR analyzers, have been in use for decades for both industrial and environmental monitoring applications. These are usually simple filter-based analyzers featuring one or more wavelength specific devices, customized for a specific analyte, such as CO, CO_2, hydrocarbons, water (moisture), etc. Such instruments are inexpensive and easy to maintain. If one reviews the market in terms of the usage and where the most analyzers are being purchased, this class of instrument, both for mid-IR and NIR measurements, is by far the most popular.

To some extent, these simple filter-based instruments are limited to applications where there is simple chemistry, and where the analytes can be differentiated clearly from other species or components that are present. In many ways, these analyzers are viewed as simple sensors or even meters, and are not viewed by the analytical instrument community as true instruments. During the past 10 to 15 years, a new focus on instrumentation has emerged based on more sophisticated technology, and is considered to be a direct evolution from laboratory instruments. To some extent, this started with an initial interest in NIR instruments. In this case, it was demonstrated that because of the nature of the information obtained in the NIR spectral region, that an analyzer capable of measuring multiple wavelengths and preferably a full spectrum was most desirable. From this, the entire market for NIR instruments developed in the late 1980s through to the 1990s for scanning NIR instruments. In some cases filter instruments were used, but because of the need to measure multiple wavelengths, the instruments became more complex than traditional filter photometers.

In parallel to the development of full-spectrum NIR analyzers, interest was developing in the other area of mid-IR. By the mid-1980s, FTIR was becoming accepted as the de facto method for measuring the mid-IR spectrum in a laboratory environment. However, FTIR was considered to be too complex and too fragile to have any utility on the factory floor or on a process line. To some extent this was true based on the state-of-the-art of the technology used in the laboratory-based instruments. Some pioneers in FTIR managed to prove the concept with modified laboratory instruments. However, after a few false starts, it was recognized that it was not practical starting from an instrument designed for the laboratory environment. At least two companies, starting around the mid-1980s, and later others, decided to focus on industrial applications, and designed instruments for operation in more demanding environments.

4.2.1 Photometers

As noted, the largest class of IR analyzer is the filter photometer. These are one of the oldest forms of PAI, with generally simple construction, and are sold at relatively low cost. In the simplest form they consist of a source, a filter, a sample cell and a detector.[18] They are usually single channel devices, and consequently function as a single beam instrument. Miniaturization has lead to the production of small, integrated devices that essentially behave as single species sensors. In each

case, the filter, the source and the detector define the analytical spectral region. Most filters are bandpass or correlation (gas-based), dependent on application – the latter being used for specific low concentration gas applications. Most traditional process photometers are configured for single component measurement, with the most basic featuring a single filter selected to transmit at the analytical wavelength/frequency, as indicated in Figure 2.

With this form of instrument, the baseline is established with a zero sample, and the span (range) is established with a maximum value sample. Although the electronics are simple, and designed for stability and low noise, it is usually necessary to establish zero and span settings on a regular basis, often daily. This can necessitate a calibration overhead on even the simplest photometer. The reason for this comment is that although filter photometers may seem to be simple and low cost, ownership can be more complicated in terms of the need to calibrate and its associated costs.

Conventionally, the first thoughts are towards simple measurement systems featuring one or two optical elements with the simple construction as stylized in Figures 2 and 3. In practice there are many forms of filter device that can be used. These include simple single wavelength filters, clusters of filters or filter arrays (multiple wavelengths), correlation filters (with the analyte), continuously variable filters (circular variable filter, (CVF) and linear variable filter, (LVF)) and tuneable filters (such as the AOTF).

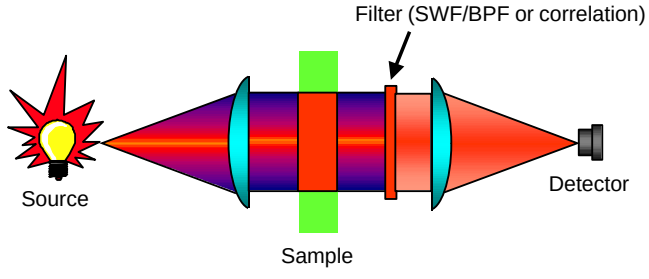

Figure 2 Simple filter photometer featuring straight-through transmission optics.

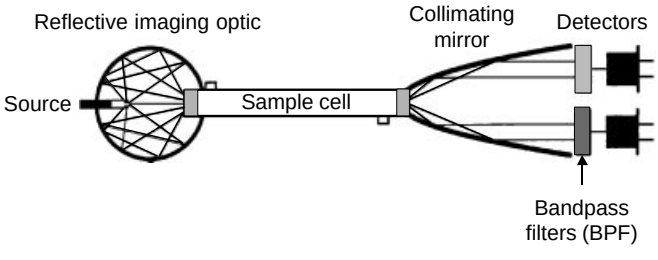

Figure 3 Simple two-channel filter photometer featuring dedicated detectors with individual filters.

The use of continuously variable filters is not new, and scanning instruments featuring single element detectors have been in use for at least 25 years. Likewise, clusters of filters or filter arrays have been used for NIR instruments and some mid-IR instruments with single detectors. In all cases, a mechanical device is required to drive the filter assemblies. The reproducibility and reliability of these devices is cause for concern, and often limits the performance and utility of these approaches. Filter clusters and continuously variable filters (such as an LVF) can combine well with array detector devices, as indicated in Figure 4. Such a combination can lead to a very compact instrument for simple transmission measurements on liquids or gases, and eliminates the mechanically related problems associated with moving parts.

AOTF devices,[10,11] which tend to be limited to the visible and NIR spectral regions, also overcome the mechanical problems associated with multi-wavelength filter instruments, and feature only a single element detector. They provide the benefit of rapid spectral scanning, with a single full-spectrum acquisition in about 200 ms. An alternative mode of operation is to step the device to specific wavelengths. This is an extremely rapid process, and emulates a multi-wavelength filter system, with near-instantaneous output at each wavelength. With these features, these devices would appear to be ideal for rapid NIR measurements. However, they have only received partial acceptance, primarily due to a relatively high cost of the instrumentation, and wavelength calibration and stability issues.

The largest application segment for traditional filter photometers is for combustion gases analysis (for CO, CO_2, hydrocarbons, SO_2, etc.). Other major areas of application include the petrochemical industry, with natural gas and other hydrocarbon process gas streams being important applications. The remainder of the applications are for specialized process analyses, featuring both gas and liquid applications. As measurements become more complex, in general, the more advanced

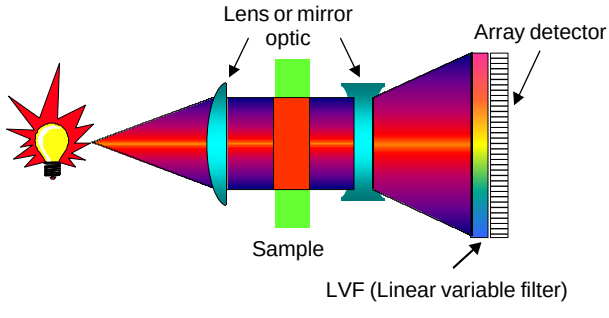

Figure 4 Conceptual design for a simple array-based spectrometer featuring a LVF.

filter-based (LVF and AOTF) analyzers, or full-spectrum FTIR or NIR instruments are considered to be attractive solutions, both in terms of cost and versatility. It is anticipated that this situation may change once array-based systems become available. In such cases, the combination of filter and detector arrays, or LVF devices with arrays, could lead to lower cost, more versatile analyzers.

4.2.2 Monochromators and Polychromators

For many years, the standard configuration for optical spectrometers for mid-IR and NIR featured a scanning monochromator. These included a single source and detector, with a mechanically scanned dispersion element in combination with optical slits for spectral line separation. In many systems, both the dispersion element (a refracting prism or a diffraction grating) and the slits were driven mechanically by relatively complex mechanisms. Although this arrangement functions adequately in a laboratory environment, it is not always ideal for operation in a heavy industrial setting. Reliability, calibration, and overall performance (or lack of) have been factors that that have limited the use of monochromators. Today, a few of the commercial NIR instruments still feature monochromators and all standard commercial mid-IR instruments now feature interferometers (FTIR).

Monochromators, by definition, provide energy of one wavelength (or a few wavelengths) at a time. In other words, the devices scan through the spectrum sequentially. A polychromator allows multiple wavelengths to be measured simultaneously. This arrangement is used in instruments that feature detector arrays, and this principle is illustrated in Figure 5. In this simple example, a concave diffraction grating is featured, usually a holographic element, and this splits the energy spatially into the analytical wavelengths and provides the necessary imaging on the array. The advantage of array-based spectrometers is obvious, based on the fact there are no moving parts, providing the mechanical and wavelength stabilities necessary for highly reproducible measurements in a factory or plant environment. Currently, several commercial instruments, many in a compact form, are available based on this concept, featuring arrays that operate in the NIR. The longest wavelength version of these, based on a lead selenide array, does cover the top-end of the mid-IR region.

One technique that can utilize the output from a monochromator, and can provide some of the benefits of multiplexed data acquisition, is known as a Hadamard transform spectrometer. This type of instrument can in fact feature either a monochromator or a polychromator (equipped with a detector array). Today, such instruments are available as custom-made devices, but none have been commercialized.

4.2.3 Interferometers

By the mid-1980s most of the mid-IR measurement technology moved away from dispersive instrumentation (monochromator-based) to interferometry, with the widespread introduction of FTIR instruments.[19] These instruments provided the performance and flexibility that had been in demand for mid-IR. The basic elements of an FTIR instrument, featuring a Michelson-style of interferometer, are illustrated in Figure 6. Critical elements in this style of instrument are the beam-splitter and the two mirrors (one fixed the other moving). Because the system requires the measurement of an optical interferogram, generated at high optical precision, from this mirror–beam splitter combination, the mechanical tolerances on the construction are extremely demanding.

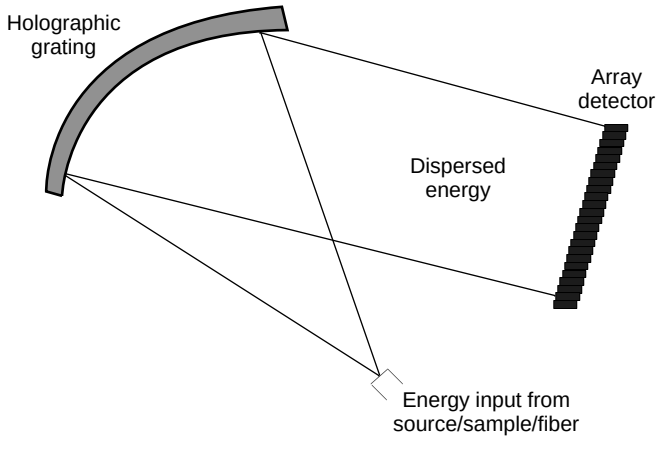

Figure 5 Idealized array-detector spectrometer featuring a concave holographic grating dispersion element.

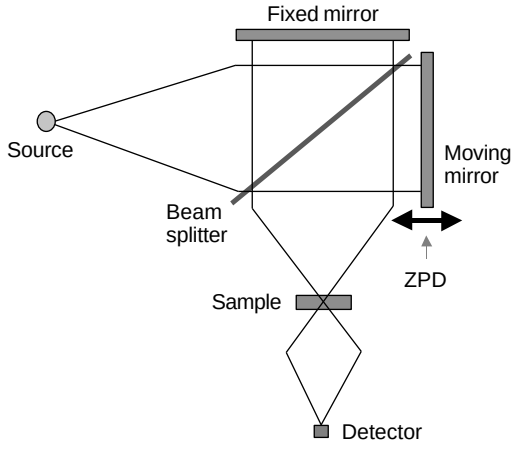

Figure 6 Idealized Michelson interferometer based on conventional fixed mirror, moving mirror and beam splitter optical configuration.

The drive mechanism used for the moving mirror assembly (which provides the spectral scanning) has to be critically controlled, and must be made impervious to external interferences, such as vibration. In the mid-IR region, the beamsplitter is also an environmentally fragile element. This optical element, which is usually constructed from a hygroscopic material such as potassium bromide (KBr), serving as the main support substrate, has to be prepared to high optical flatness, and is coated with multiple layers of high refractive index materials. The element must be protected from the environment, in particular to prevent surface damage from humidity. This requires the instrument to be hermetically sealed, purged, or protected by desiccant. For demanding applications, nonhygroscopic materials, such as barium fluoride and zinc selenide, have been used for beamsplitter construction. One other fall-out from the requirement for high optical tolerances, all round, is the impact of temperature. Interferometers are inherently sensitive to changes in ambient temperature, and in acute situations can lose optical alignment, with consequent losses in stability and energy throughput.

With careful attention to the basic design, it is possible to construct a rugged interferometer suitable for use in a factory or production environment. Several manufacturers have focused only in this area and a range of systems are available, including on-line process monitoring applications. Even in these cases, operational issues can arise, mainly because of the need to maintain good optical alignments, and the fact that the instrument still features key moving parts. An alternative design of interferometer, known as a polarizing interferometer, has no moving parts, and at least two commercial designs have been reported, mainly for operation in the NIR region. However, this concept has not been readily received or encompassed by the major FTIR vendors. The down-side to most interferometers, even those designed for alien environments, is the relatively high cost, and the overall size and system complexity. However, for the moment, in the mid-IR spectral region, with the exception of a few multi-wavelength filter instruments, there is no other practical choice of technology available.

4.3 System Requirements for Process Infrared Analyzer Systems

There are some very clear guidelines that must be followed in the development of instrumentation for industrial process-related measurements. On the surface, there are compelling reasons to transfer traditional instrumentation technology from the laboratory out to the manufacturing line. A common assumption is all that is required in the development of a *process analyzer* is to repackage the laboratory instrument into an industrial-grade enclosure. Many instrument companies have fallen into this trap. While this concept might work initially in the early prototyping phases of an instrument's development, it seldom works for a final process analyzer product. The bottom line is that there is much more to a process analyzer than simply the repackaging of an existing instrument. It is a matter of measurement philosophy, overall system design philosophy and in fact a complete change in operational concepts. Before addressing the different requirements of process instrumentation, it is necessary to classify the mode of operation and the operating environment of the analyzer. Also it is necessary to revisit the term *process analysis*.

First, it is worthwhile to provide a logical separation between the terms instrument and analyzer. Typically, an instrument provides a measurement which requires interpretation of the data generated, and is usually totally open-ended and flexible in terms of sampling and operation. An analyzer is a total system which includes the sampling, the definition of instrument settings, a defined user interface, and a predefined method of measurement and calculation of results. Generally, when focusing on process instrumentation, one is dealing with analyzers. For the purposes of this article, process analysis will be defined as any analysis that is performed to provide feedback to the manufacturing process, in terms of raw material quality (screening), the progress of a process – whether it is batch or continuous – and final product composition/quality. This can range from dedicated laboratory-based instrumental methods for off-line testing to continuous on-line measurements with an environmentally hardened analyzer. The following gives the classes or platforms of process analyzer that can be defined.

4.3.1 Laboratory-based Analyzers

These analyzers are often derived from standard laboratory instruments. However, once well-defined, the system can be provided as a custom-designed or dedicated analyzer. For a spectrometric analysis (IR, NIR or Raman) the main issues are ease of sample handling (usually without modification of the sample), a simple operator interface and automated calculation and reporting of results. Within the laboratory, a single analyzer might be multifunctional, providing analyses for a range of samples. In this case, the analytical methods tend to be "*recipe-based*" and are usually prepared by a separate research and development group, and are typically stored on a disk file. The procedures usually require very little operator input, although a documented audit trail is usually available in the event that a measurement is questioned. Once a method becomes *standardized* there is an opportunity

to develop a standard analyzer just for that measurement. This is particularly beneficial in the case of regulated analytical procedures. Obviously, there are minimum instrument performance requirements for a given analysis, although the most common sources of variability and nonreproducibility in spectrometric measurements originate from the method used for sample handling.

- Most important design issues: simple sampling, ease-of-operation, recipe-based analyses (stored methods), automated data evaluation and reporting of results.
- Typical examples: a dedicated laboratory instrument equipped with a permanent sample handling system (sampling station) and with quantitative and/or qualitative analysis software; a flexible laboratory analyzer that can be ''programmed'' for dedicated analyses; a dedicated analyzer for specific analyses, such as the measurement of unsaturation in polymers, prepolymers and monomers, hydrocarbon characterization (PONA, PIANO, etc.), or hydroxyl value. Other examples outside the traditional chemical industry include semiconductor analyzers focused on established methods used in the semiconductor industry.

4.3.2 Plant-based Off-line Analyzers for Near-line or At-line Measurements

While plant-based off-line analyzers might share a common heritage with laboratory-based instruments, their design requirements are typically more rigorous and demanding. Laboratory-based instruments are normally packaged in a molded plastic or sheet metal housing, with standard power and data cable connections. When working in a plant environment, it is necessary to protect the instrument from the operating environment, and to ensure that the instrument conforms to factory and worker safety requirements. These impact the way that an instrument/analyzer is packaged, and in the nature of the electrical connections used for power and data transmission.

Another critical issue is that a process analyzer is normally required to function 24 h a day, and should work reliably for all work shifts. It is very undesirable for an analyzer to fail unexpectedly during a late night or an early morning shift. This requires a high degree of reliability, which is a design issue – normally achieved by minimizing critical moving parts, and by using components with long mean time between failure (MTBF). This latter point is one of the main arguments away from FTIR, and in favor of newer methods of measurement that by-pass the need for critical moving parts.

Simplicity of sampling and analyzer operation was stressed for the laboratory analyzer. This requirement is even more essential within a true plant environment. Sampling must be highly reproducible, and should require only a minimum of operator skill. One must not second-guess an operator when designing the sample–analyzer interface. The same applies when implementing an operator user interface on the analyzer. Ideally, the analyzer should feature push-button controls, with a minimum number of buttons. If a keyboard input is required, the number of keystrokes for any operation should be kept to a minimum. All visual displays must be kept as simple as possible, they must be unambiguous, and any cautions, warnings or error messages must be clearly displayed. Where practical, universally accepted symbols must be used, and when important messages are presented in written form, ideally they must be presented in the local native language (especially for instruments used in European Community controlled countries).

In terms of packaging, attention must be paid to several issues. Most modern analyzers feature a personal computer (PC) (usually an IBM compatible platform) for data acquisition, control, data analysis, presentation/display of results and communications. There are two approaches that can be adopted when using such a system: either the computer is integrated into the analyzer package, including the visual display and keyboard/control panel (an abbreviated keyboard or a preprogrammed equivalent), or a separate industrial-grade computer is used. An entire industry has been established around industrial-grade computers, from high-performance single-board computers configured for use on a passive back plane for integrated analyzer designs, to rack-mounted PCs and/or environmentally hardened PCs for systems requiring separate computers. This includes keyboards and input devices (mouse, touch-pad or trackball), and displays, which may feature integrated touch-sensitive screens. Touch-screen technologies have advanced to a point where they are robust and quite suitable for use in ''non-ideal'' factory or plant environments.

The main analyzer ideally should be mounted within an environmental enclosure, which is sealed to provide protection from water and dust. It is not uncommon that an analyzer must be required to withstand direct spray from a high-pressure water hose or a steam line. Typical ambient operating specifications require the capability to operate in a wide range of humidity (from zero to 100%) and over a wide temperature range, at minimum from around 0 °C to 40 °C, often wider. If there is the risk of exposure to flammable vapors, either incidentally or continuously, then the analyzer is normally required to conform to fire safety codes. These normally require that the analyzer is purged with positive pressure, and that it is equipped with a power interlock linked to the presence of positive pressure (and flow for Cenelec code compliance). A standard format for such analyzer requirements is

a purged and pressurized stainless steel enclosure. As technologies change, and as miniaturization takes place, this requirement may change because alternative methods of packaging may be used.

Much of the instrument protection and code compliance can be achieved with available hardware, and so implementation can be straightforward. However, designing a process analyzer does not stop at the enclosure. The internal components, the electronics, the spectrometer and all associated optics, must be made to be insensitive to wide variations in temperature. This may be achieved by design, or by providing a temperature controlled and/or stabilized environment within the enclosure. Furthermore, all components must operate reliably independent of the operating environment of the instrument/analyzer, which can include wide variations in humidity and temperature, and the presence of vibration, over a broad range of frequencies.

- Most important design issues: minimum sampling, push-button operation, recipe-based analyses (stored methods), automated data evaluation and reporting of results, high reliability, reproducibility, and stability, and packaging providing ample environmental protection to the analyzer and conformance with local and industry-standardized safety and fire hazard operating codes.
- Typical examples: an analyzer located at a loading dock for screening delivery of raw materials and/or shipment of final products; an analyzer located in a warehouse for raw material screening and inventory control; an analyzer located in an analyzer room or adjacent to a batch reactor for monitoring a process or processes from collected samples; a mobile analyzer used for process optimization or for trouble shooting of batch or continuous processes.

4.3.3 Plant-based On-line Analyzers

The overall requirements of an on-line analyzer in general are similar to the requirements described above for an off-line process analyzer. The main difference is that the analyzer normally operates 24 h a day, unattended, when implemented on continuous processes. This places an even higher dependence on high reliability and safety related issues. Generally, maintenance must be planned and is usually carried out during a scheduled plant shutdown. Any component that is known to have a limited lifetime is usually replaced during such a scheduled maintenance period. This reduces the likelihood of an unexpected failure. For this reason, for such key components, the replacement must be simple, and must not require sophisticated alignment procedures. For example, if a laser in a FTIR instrument is to be replaced, there must be no need for tedious alignment, often associated with laser replacement in laboratory-based instruments. It should be possible for an average electrical technician or maintenance engineer to replace this component.

Most on-line analyzers are installed as permanent fixtures. The only exception is for process monitoring studies, where a spectrometric analyzer is used to characterize and/or optimize a process. For these applications, a semipermanent system, capable of being moved from process to process, with a degree of re-configuring, is installed. Note that overall, such a system must meet all of the normal requirements of a process analyzer, except it will feature greater variability in sampling and configuration options than a permanent system.

Bearing the needs of a permanent system in mind, the environmental issues of temperature and vibration become more critical, and the requirement for conformance to safety standards may be enforced by the needs for system certification. The latter requires an assessment of the working environment, and the potential for fire and/or explosion hazard, relative to the presence of flammable vapors or gases. Note that safety hazard and local electrical design code compliance, which include the CE Mark for Europe, Canadian Standards Association (CSA) for Canada, and possibly Underwriter's Association (UA) for the USA, may be a requirement for both off-line and on-line analyzers for certain applications and/or operating environments. Certification may be performed by recognized organizations such as Factory Mutual (FM – USA) or TÜV (Germany).

System installation requirements can be more demanding for a permanently located system, as these may require a sample conditioning system (featuring some degree of automation, such as automatic cleaning and outlier sample collection) and the need to interface to an existing control system (process computer). The latter may require that the system provides for some standardized form of communications protocol, such as Modbus, for the chemical industry. It should be recognized that certain specialized industries have their own protocols, such as the semiconductor industry, which uses protocols known as SECS and SEC-II. Also, the *Universal* Fieldbus as a method/protocol for process analyzers is gaining ground, albeit fractionated by more than one manifestation of the standard.

With some idealistic thinking, for applications involving the use of optical fiber-based sample interfacing, it has been proposed that the main analyzer can be placed inside the main control room, possibly within a standard 19-inch electronics rack. Although this is certainly not out of the question, and several analyzers have been effectively designed with this format, not all end-users favor this style of installation. The more common approach is to locate the analyzer within an analyzer house. With such

a configuration, it is not unusual to locate the sample cell and sample conditioning system within a hazardous location, and to install the instrument in a general purpose, or a *safe* area, with the optical fibers used for the analyzer–sample cell/sampling system interface.[21] (Note that the use of optical fibers in the mid-IR region covered by most FTIR instruments is very limited. Currently, common choices are zirconyl fluoride glasses, which provide a spectral window down to around 2000 cm^{-1} (5.0 μm) and the chalcogenide glasses, which may be used from around 3300 cm^{-1} (~3 μm) to 800 cm^{-1} (12.5 μm). However, in the case of chalcogenides, the length of fiber that can be used is typically limited to a few meters (sometimes less than 2 m, dependent on quality) because of transmission losses, usually linked to impurities. Note for some applications, a practical alternative to optical fibers is light-pipe technology (optical conduit), where the sampling interface is interconnected to the instrument via a series of light pipes with flexible connections. While this approach works well for certain applications, it is also distance limited, both in terms of bulkiness and light losses.) For such installations, there is less requirement for the analyzer to conform to safety standards, and the rack-mounted format is quite suitable, assuming that a rack assembly is available to house the analyzer. Without a rack in place, a rack-mount style system may still be mounted within an analyzer house, either on a shelf or on rails.

- Most important design issues: sampling system (possibly automated), minimal or no user interface, recipe-based analyses (stored methods), automated data evaluation and reporting of results, standardized communications (such as Modbus), high reliability, reproducibility, and stability, and packaging providing ample environmental protection to the analyzer and conformance with safety and fire hazard operating codes. An exception to this requirement is for analyzers that can be located within a safe environment, such as in a control room, in such cases, a 19-inch rack-mount enclosure format may be preferred.
- Typical examples: a permanently mounted analyzer for on-line reaction or product-line monitoring with an in-line or a side stream sample probe; a fully integrated analyzer for process control with sampling system and communications for a process feedback loop, such as a refinery stream analyzer for fuel component production and blending; an analyzer for monitoring polymer extrusion with the sampling probe integrated into the extruder head; a system configured for gas analysis for production gases, effluent or process gases, and/or combustion gases.

The functional purpose of many on-line process analyzers is often the same as for off-line applications. Consequently, many of the design requirements are similar. A very practical design concept is to use common instrument measurement technology throughout, providing all three analyzer platforms defined above. This should result in a laboratory-based system, with adequate flexibility for methods development, and the ability to handle both off-line and on-line plant applications as required. One of the most important consequences of the use of common technology is that it provides the opportunity to use instrument standardization methods, and to achieve calibration transferability between analyzers, whether installed in a laboratory, or at or on a process line. However, having stated this, the final process analyzer is typically more expensive than the standard laboratory analyzer, and is usually not an attractive or desirable option for calibration. This can make the final analyzer prohibitively expensive for certain applications – FTIR is a case in point. Therefore, there is a trend in certain areas to consider the use of simpler technology, including a filter-based instrument for the final installation, where all of the initial work is performed on the more sophisticated instrument, and then the calibrations are moved over to the simpler instrument with a suitable calibration transfer function.

If the FTIR process analyzer is evaluated in terms of the issues outlined for the different analyzer platforms, it is possible to indicate potential areas of weakness or areas that may need further attention. Traditionally, the main issues are safety, reliability and serviceability, vibration, temperature, optics, sources and detectors, sample interfacing and sampling, and calibration.

4.3.3.1 Safety Within a FTIR instrument there are several potential safety hazards. The source is one – providing local temperatures at the source in the range of 1100 K to 1500 K. The other is the laser and its power supply, which is a potential spark generator. Both are potential fire hazards. Methods for thermally isolating – and possibly sealing – the source are important, especially if an analyzer is to meet fire safety codes. For the laser, there are special power dissipation requirements to ensure adequate discharge of any built-up charge within the capacitance circuitry of the laser power supply.

4.3.3.2 Reliability and Serviceability In the past, FTIR instruments have been considered to be complex and unreliable devices (from a process engineer's view). Attention must be paid to system environmental stability, the reliability of key components, such as optics (non-hygroscopic preferred), electronics, lasers and sources. If a component has to be replaced, then a nonspecialist

technician should be able to perform the task. Complex optical alignments must be avoided at all costs.

4.3.3.3 Vibration Vibration can be a killer for any form of instrumentation. FTIR is a special case because the fundamental measurement is vibration sensitive. Care must be taken in the design of the mirror drive system and the associated optics to ensure that the spectrometer is immune to vibration. A common source of problems can often be traced to mirror mounts that may vibrate if not designed correctly. With a FTIR instrument it is necessary to ensure that the components that are supposed to move do continue to move, and to prevent those that are not supposed to move from moving. Constant vibration can work components loose, and so the extensive use of a screw bonding (adhesive) medium, is strongly recommended, even for noncritical components.

4.3.3.4 Temperature In any opto-mechanical system, temperature changes can result in dimensional changes, which inevitably cause problems if not addressed, especially with FTIR instruments. Temperature compensation is usually required, and careful attention to the materials used for critical components, with regard to expansion coefficients, is essential. This even includes screws and bonding materials. If correctly designed, the optical system should function over the temperature range defined (typically 0° to 40 °C). At issue is the rate of temperature swings – most systems can tolerate gradual swings if designed correctly. Rapid thermal transients can be more problematic, where a rapid rise in temperature may result in thermal shock of a critical optical component. At issue also, relative to thermal design, is the impact on local electronics. Unless designed or specified for high temperature operation, many electronic components can fail or become unreliable at elevated temperatures.

4.3.3.5 Optics Two key issues relative to optics have already been addressed – that is the requirement for nonhygroscopic optics, especially for exposed transmissive optical surfaces (lenses and windows) and the need to pay attention to mirror mounts relative to vibration and/or thermal effects. In terms of nonhygroscopic optics, zinc selenide is a useful material, especially when anti-reflection coated; if a good coating is used, the transmission can be as good as or better than potassium bromide. If potassium bromide has to be used for its lower transmission range (down to 400 cm^{-1}) then a protective coating can be considered. Many forms of coating have been offered, but most have the potential to fail at very high levels of humidity (near 100% or condensing). A process instrument may feature protective windows to help maintain purge and/or for safety reasons. If used, care must be exercised in ensuring that back reflections into the interferometer are avoided, because these will cause photometric errors.

4.3.3.6 Sources and Detectors This critical area is one of the most demanding and also one which is open to new technology. Most sources have a finite lifetime, and hence are considered as replaceable items, also they are a source of considerable heat, which must be successfully dissipated (see comments above regarding temperature). Detectors are of similar concern. For many applications, where the interferometer can scan at relatively low speeds, without any undesirable vibrational/mechanical problems, the traditional lithium tantalate or DTGS detectors may be used. These pyroelectric devices operate nominally at room temperature (do not require cooling to function) and are linear over three or four decades.

For some applications, as a function of interferometer stability, sample interfacing, speed of spectrum acquisition, or sensitivity, it is necessary to resort to a higher performance detector, such as a MCT detector. This detector is capable of high-speed acquisition and has a very high sensitivity (up to two to three orders greater than DTGS). The main issues here are practical, as discussed earlier (section 3.1.3), relative to the need for cooling down to liquid nitrogen temperatures (77 K), and the fact that the detector exhibits nonlinearity with absorbances greater than one (unless electronically corrected).

Note also, that although a DTGS detector normally does not require cooling, it will stop functioning as a detector, because of depolarization, at temperatures above 40 °C. From a design point of view, it is important to note that internal instrument temperatures can be up to 10 degrees higher than the external ambient temperature. Therefore, there is often the need to temperature control or thermally stabilize a DTGS detector in a process analyzer. In this case, a simple, one-stage Peltier (thermoelectric) cooling device will suffice.

4.3.3.7 Transfer Optics Getting the IR beam to the sample, especially within a classified (safety) environment, and ensuring optimum interaction with the beam, can be a major challenge. Gas samples, in particular nonflammable samples, are not necessarily too difficult to deal with. One of the main issues here is paying attention to the corrosivity of the gases, relative to the windows and any internal optics (as with a folded pathlength cell). Liquids offer a completely different set of challenges. Where possible, users prefer to minimize the use of valves, by-pass streams and auxiliary pumps, especially over long distances from the stream to the analyzer. Therefore, there is a need to sample the stream either directly or as close as possible to the stream. As a result there is often the need for transfer optics from the sample cell to the spectrometer. There are a couple of quite elaborate schemes for

''piping'' the light from the spectrometer to the sample and back. While these do work, they are typically cumbersome, and they can be awkward to implement, and maintaining alignment and optical throughput can be a problem.

The use of fiber optics has already been mentioned, and it is worthwhile to consider their use in context. In the mid-IR they may be used for certain applications, but currently, their use is limited when used within the important fingerprint region. The fiber materials of choice have been zirconyl fluoride and chalcogenide glasses. The latter is transparent throughout the fingerprint region down to around 800 cm^{-1}, but with a sacrifice of energy throughput due to high attenuation losses, caused primarily by impurities in the glass. The result is a limitation on fiber length, and on the choice of suitable detectors – MCT is normally required to handle the low light levels. Work is continuing on mid-IR fiber materials, and with the recent introduction of silver halide fibers and capillary light fiber guides, which provide a better transmission range than chalcogenide glasses, and also much lower attenuation losses, the interest in the use of fiber interfacing is expected to continue. Note that the silver halides are only suitable for the mid-IR range because at shorter wavelengths, especially with visible or UV wavelengths, darkening of the silver halide matrix occurs.

The limitation on the use of fibers does not apply in the NIR, where common silica-based optical fibers work well. Note that it is essential to use low-OH fibers to reduce absorption losses due to OH absorptions over the length of the fiber. The use of fiber optics for NIR applications is now well established, and for many applications this is a favored method for interfacing the spectrometer to the process. For certain applications, fiber distances of up to 1 km have been implemented. For most gas or liquid phase applications, a single filament fiber is adequate. However, for applications that involve reflectance measurements, especially with diffusely reflected radiation, fiber bundles are more beneficial in terms of light collection over a large solid angle. The trade between single filament versus multiple fiber bundles is light gathering efficiency versus cost of implementation, as well as coupling efficiency with the spectrometer. Fiber bundles can be very expensive, in terms of cost per meter, and so their use tends to be limited to applications with a short distance between the sampling point and the spectrometer.

While the use of fibers might seem to be a logical choice in terms of applications within a hazardous environment (flammable or toxic vapors), they may not always be the final or preferred choice. Light transmission and attenuation are not the only factors for consideration. Temperature and mechanical shock and/or vibration are other important parameters. The transmission efficiency of a fiber is strongly affected by temperature changes and vibration, and this can lead to poor spectrometry, and unstable baselines. Care must be taken in the way that fibers are mounted and packaged, especially when used externally.[21]

4.3.3.8 Diagnostics There is more than one type of fault that can develop with a process analyzer. A component failure is often relatively easy to diagnose, and so on-line fault-finding can be straightforward. The more insidious problems occur with the analysis itself, or when a component ages, but does not fail, or when a sampling device gives problems (such as corroded or coated windows of a cell). Many of these problems are application dependent, but where possible the failure should be anticipated, and preventative measures put in place to reduce the impact, or at a minimum provide adequate warning of failure. Note that the fact that a FTIR instrument is a full-spectrum device, providing a significant amount of over-sampling, is very beneficial for both application diagnostics and also calibration purposes, when compared to filter-based analyzers.

4.3.3.9 Calibration Most process analyzers are designed to monitor concentration and/or composition of a product or intermediate. In such cases it is necessary to calibrate the analyzer from a set of prepared concentration standards or from well-characterized reference materials. The simple approach should always be adopted first, and for relatively ''simple'' systems the standard approach is to use the Beer–Lambert law relationship, where the measured absorption is directly proportional to the concentration of the analyte.[22] Traditionally, each analyzer is calibrated locally against the standards. In more complex chemical systems, it is necessary to adopt either a matrix approach to the calibration (still relying on the Beer–Lambert law) using simple regression techniques, or to model the concentration and or composition with one or more multivariate methods, an approach known as chemometrics.[23–25]

4.3.3.10 Standardization There is the need to model and compensate for the instrument response function, which will vary from analyzer to analyzer. The ability to address these issues will be essential if calibration transfer is to be achieved across all instrument platforms.[26] Also, there is often a need to be able to perform local standardization and analysis calibration while the instrument is on-line without taking the system off-line or out-of-service. Ideally, on-board calibration/standardization can be integrated into the sample conditioning system. Most commercial NIR analyzers feature some form of standardization and calibration transfer. In fact such schemes are essential to minimize the need to perform frequent calibration – an act that could be cost and time prohibitive in a production

environment. Similarly, many of the modern FTIR (and Fourier transform NIR) systems also include some form of instrument standardization, often based on an internal calibrant, such as a ''standardized'' polymer film. This is becoming an important feature for regulatory controlled analyses, where a proper audit trail has to be established, including instrument calibration.

Overall, most of these requirements are straightforward to implement, but they do require attention at the design level. One other area of importance which is not particularly FTIR specific (although it may be in some areas) is the user interface and the need to provide industry standard data communications. One option here may be to use a standard software package from a company such as Intellution or Wonderware – both recognized products for process interfacing and control. For prototype development, and even for the front-end interface in a stand-alone mode of operation, software products, such as Microsoft's Visual Basic and National Instruments' LabView, are also very flexible tools for instrumentation software. This latter company also provides important computer-based electronics that meet most of the computer interfacing and system control and communications needs.

As previously noted, many of the needs specified are driven directly by the known needs of the traditional process analyzers. When it comes to the high-technology fabrication industries (semiconductors, optics, vacuum deposition, composites, etc.), consumer-associated industries, such as the automotive after market, and environmental, safety and hygiene applications (United States Environmental Protection Agency, United States Occupational Safety and Health Administration, etc.), the needs can be different. Instrument reliability and performance are important in all aspects, and operator safety requirements are much the same throughout, although the actual standards may be different, dependent on the regulating agency. For some of these specialized areas of application, the need to follow well-defined operating procedures and operating protocols is essential, and full documentation of audit trails are often a requirement. In certain cases, ISO 9000 certification may be a prerequisite.

4.4 Sample Handling in a Production Environment

Throughout this article, many references have been made regarding the sample, how it is handled and presented to the IR instrument. Sample handling in the IR has become a major subject for discussion over the past 20 years, and is covered well in several standard texts.[27–31] IR spectroscopy, when compared to other analytical techniques, tends to be unique relative to the large number of sample handling technologies that are available. Selecting the correct or appropriate technique is extremely important relative to the overall success of a process-related application. This applies equally to both off-line and on-line styles of analyzers. The only difference for on-line applications is that the sample handling method must be sufficiently robust to maintain optical performance throughout operation. Attention must be paid to sample corrosion, pressure and temperature effects, and this often requires the use of specialized and even exotic materials for both the optical elements and the cell fabrication materials. Optical materials such as sapphire, cubic zirconia, and even diamond may be used for demanding applications, as long as the optic transmission range matches the needs of the measurement. Likewise, materials such as nickel, Monel® and Hastelloy® may need to be used for cell construction.

Probably the most important factor in selecting the appropriate sample handling technique is to try to work with the sample material as it exists, without any form of chemical or physical modification. For gases and certain liquids, simple transmission cells, often with a flow-through configuration, will work for both NIR and mid-IR measurements. Samples that are strong absorbers of IR radiation will cause problems for mid-IR measurements. In the laboratory, there is a possibility to reduce the sample pathlength for such samples, but in a flow through configuration this becomes impractical. This problem does not usually apply in the NIR, where sample pathlengths are typically one or two orders of magnitude larger (typically between 1 mm and 100 mm) than in the mid-IR.

A common solution to the pathlength problem in the mid-IR is to use one of the many styles of ATR (attenuated total reflectance) sampling devices. These are available in a number of different configurations, providing both methods for manual sample loading (horizontally mounted sampling points and dip probes) and automatic flow through methods. A range of materials, including diamond, are now available for these optical devices. The technique utilizes a surface measurement (to a depth of 1 or 2 μm), and so careful attention must be paid to surface contamination and fouling, especially for on-line, flow-through applications. Fouling of the surface not only causes interferences and/or potential energy losses, but it can also change the photometry of the experiment by modifying the refractive index of the surface. In such cases, it is important to try to implement algorithms for detecting material build-up on the surface. Cleaning schemes can be invoked, and by definition, this implies the incorporation of some form of sampling system. For some applications, a simple flushing or back flushing with a solvent may suffice. If the material adheres strongly to the surface, it may be necessary to implement some form of mechanical cleaning of the surface. One practical alternative approach, that is

available on a commercial FTIR process analyzer, is to use an ultrasonic cleaning system.

It is a general perception that sample handling is easier in the NIR region. To some extent this is true, because longer pathlengths may be used for liquid sample handling, and direct reflectance methods, such as diffuse reflection, may be used for convenient sampling of solids. While diffuse reflectance can be used in the mid-IR, optically related artifacts in the measurement tend to limit its use to purely laboratory applications. Even in off-line applications, where an operator is involved in the sample handling, it is important to keep the sample presentation as simple and as reproducible as possible.

4.5 Factors Influencing Acceptance of Spectrometer-based Analyzers

4.5.1 Low Maintenance and Turnkey Systems

Given the conservative nature of the average process engineer, there is a need to supply the system as a turnkey product. It is expected that the product will be operational ''out-of-the-box'', and will require little technical support at the time of installation. Once installed, it is expected that the system will operate with minimum operator attention. Also, any maintenance requirements should fit into a scheduled maintenance program. Any items that are likely to fail within a given time-frame must be identified, and recommendations for scheduled replacement must be provided. With a FTIR system, components such as sources and lasers are often the most likely to fail, and a knowledge of the MTBF for these is important.

Note that on-board dedicated processors are appropriate; however, integrated hard disk drives are not recommended, and if on-board storage is required, it is recommended that alternative solid-state storage, such as flash random access memory (RAM) or battery-backed RAM is used. Critical moving parts in analyzers are always subject to critical assessment. One factor to bear in mind is that early false starts with FTIR have left a slight stigma against the technology. Much of this is linked to mechanical and environmental (vibration and temperature) issues. It is important to re-enforce design features and attributes that address past concerns of system stability and reliability.

Finally, relative to maintenance issues, the requirement for high cost replacement parts must be kept to a minimum, unless the frequency of replacement is once every 3 to 4 years. For FTIR analyzers, this may not be a problem, as long as components are simple to replace. If an entire sub-assembly needs replacement this could represent a maintenance issue. Key replacement items that need attention, in terms of ease of replacement are sources and lasers. Laser re-alignment after replacement is also an important issue – it must be easy to perform, and should not require a person to perform major optical adjustments. If interferometer alignment is required, then this should be automated via a mechanized auto-align procedure.

4.5.2 Multipoint Monitoring Considerations

One of the benefits of certain types of optical instrumentation, that can be interfaced via fiber optics, is that there is the possibility to use an optical multiplexer, and monitor several points with a single instrument. This approach has been used successfully with NIR instrumentation, where typically up to 8 points can be handled (note that one vendor provides a multiplexer that handles up to 20 inputs). With sequential access on a multiplexer, the elapsed time is still the measurement time multiplied by the number of points – only the overhead with the sample lines does not apply. Having stated this benefit, the use of fiber optics with mid-IR Fourier transform instruments is limited, and relatively exotic fiber materials must be used. One or two commercial multiplexers have been offered for mid-IR fiber systems, but their use has not been widespread.

The opportunity to reduce the cost per analysis point can be very important for certain applications. For the most part, FTIR, when applied to process applications, is often best implemented with a direct optical coupling with the sample. In this case, there is the preferred option to use sample stream multiplexing with a manifold system. For gas-based systems this is very straightforward, and does not impose a high cost overhead. Liquid systems may be more complex, dependent on the nature of the stream involved, and material reactivity, miscibility or viscosity are important issues to address.

4.5.3 General Cost Considerations

If a spectrometer is replacing a service intensive, or possibly more than one service intensive technique, then it is obviously a positive situation. However, if it is necessary to maintain complex chemometric methods, as with some of the NIR applications, and also bear the potential burden of problems associated with calibration transfer, then it can be a large negative factor. Typically, such costs of ownership are estimated over 5- or 10-year time-frames. Note that FTIR systems operating in the mid-IR, or even in the first overtone (Fourier transform NIR) are typically much easier to calibrate, and provide a more robust calibration than traditional NIR instruments. As a result, cost benefits can be demonstrated with FTIR analyzers, which are not necessarily achievable with comparable NIR analyzers.

Calibration is an issue with all spectrometric analyzers. Although the cost of single channel (or multichannel)

photometers is low, they often require frequent calibration, which raises the cost of maintenance, and likewise the cost of ownership. The highest cost of ownership can be experienced with NIR instruments, where chemometric calibrations may need to be maintained on a regular basis. This can add a tremendously high cost of ownership (a qualified chemometrician might be required). Surprisingly, this has not been a major issue in the refining industry, possibly for two reasons: the NIR replaces several traditional analyzers, which in total have a much higher cost of ownership; and in many cases, the burden of maintenance is placed on the instrument vendor (in the form of a service contract).

5 APPLICATIONS OF PROCESS INFRARED SPECTROSCOPY

Having placed IR spectroscopy in perspective relative to fulfilling the needs of the process analysis, it is worthwhile to discuss the technique's strengths relative to specific applications. Also, it is necessary to indicate areas of greatest potential in terms of the physical measurements, and chemical species and materials most amenable to IR analysis. IR provides important information that relates to the molecular structure, of both organic and inorganic materials. Classes of compound that have strong IR signatures, and are important to petrochemical and chemical processing, include aromatic compounds, olefins and unsaturated compounds and branched chain compounds. Beyond these common hydrocarbon classes are virtually all classes of chemical compounds, with most functional groups providing characteristic absorptions throughout the classical mid-IR spectral region. Similarly, many parallel applications exist for inorganic compounds, which typically also provide very characteristic and often intense signatures.

A detailed discussion of potential and actual applications, both industrial and otherwise, is too long to include here. A relatively complete set of lists, that include implemented and potential process, process-related or QC applications, are provided in Tables 2 to 5.

6 PROCESS ANALYSIS: A REVIEW

There are many industrial segments where process spectrometers may be used, ranging from their use in traditional chemicals and petrochemicals manufacturing to fine chemicals and specialized manufacturing, which include pharmaceutical and consumer-oriented products (such as packaged goods and food products). In recent years NIR has become accepted in many of these markets as both a near-line/at-line tool, and as an on-line continuous monitoring technique. There are many areas not covered by NIR, such as gas-phase monitoring.

Table 2 Industrial applications (chemicals and petrochemicals)

Application area	Applications
Refinery production	Distillates, crude oils, reforming, aromatics, hydrofinishing, phenol and furfural finishing, dewaxing
Fuels	Natural gas, LPG, propane/butane, distillates, gasoline, diesel, fuel oil, syn-fuels, additives, octane rating, cetane numbers
Solvents	High purity products, mixtures
Oils and lubricants	Mineral and synthetic oils, greases, coolants, hydraulic fluids
Waxes	Natural and synthetic, crystallinity, phase changes
Specialty gas products	High purity gases, gas mixtures
General chemicals	Organic/inorganic, liquids and solids
Chemical reaction monitoring	Polymerizations, esterifications and other condensation reactions, diazo reactions, oxidation and reduction
Plastics and polymers	Monomers, copolymers, catalysts, fillers, release agents, additives, crystallinity, conformation, monomer units (end groups and degree of polymerization)
Polymer products	Adhesives, adhesive tapes, sealants, latex emulsions, rubber materials, plastic fabrication, etc.
Detergents	Liquids and granules, hard surface cleaners, softeners, dish washing products, bar soap (acids, unsaturates, moisture, glycerol, TiO_2)
Water treatment products	Softening agents, ion-exchange resins
Dyes and pigments	Dye compounds, azo dyes, phthalein dyes, mordants, etc.
Agrochemicals	Herbicides, pesticides, fertilizers
Environmental	Solid and liquid wastes, dump sites control, river water, priority pollutants, polychlorinated biphenyls, gaseous emissions (most types).
Combustion gases	Regulatory monitoring applications, construction materials, furniture and fabrics, fire hazard assessment
Ambient air monitoring	Workplace, ventilation systems

LPG, liquid petroleum gas.

Table 3 Construction and manufactured materials

Application area	Applications
Metal working and manufacturing	Lubrication, surface treatments, coatings
Coatings: general	Resins, waxes, natural resins, oils, inorganics (phosphates, chromates and oxides)
Coatings: paints and varnishes	Solvents, pigments, resins, emulsions
Coatings: wood treatments	Lacquers, varnishes, oil-based treatments, new environmentally friendly water-based products
Wood, pulp and paper	Production and wood by-products – resins, cellulose, leaf-based products
Paper products	Coatings, clays, fiberboard, cardboard
Mining	Ambient air monitoring of dusts, ore yields, mineral composition
Minerals and geological samples	Classification, quartz, asbestos, clays, talc (tremolite content), hydration
Construction and building materials	Concrete, concrete additives, complex silicates, hydration, combustion analysis, flammability testing
Inks and printing materials	Solvents, pigments, formulated inks, ink curing process, carbon paper, carbonless-paper
Abrasives	Al_2O_3, SiC, clays, adhesives
Ceramics and composites	Synthetic materials, bonded materials, resins, fiber glass, carbon-fiber products, clays
Leather products	Treatments, surface finishes

Table 4 Consumer products and finished goods

Application area	Applications
Food products	Unsaturates, sugars and artificial sweeteners, blended products, food additives, meat products
Natural oils and fats	Edible oils, drying oils, unsaturation
Dairy products	Fats, proteins and sugars in milk and cream, protein, fat, unsaturates and moisture in cheese and butter
Flavors and fragrances	Solvents, essential oils, synthetic products (unsaturated alcohols)
Beverages	Alcohol content (also aldehydes and secondary alcohols), CO_2 content, sweeteners
Fermentation reactions	Nutrients, alcohol formation, main product, by-products
Tobacco and tobacco products	Moisture content, leaf assay, paper
Personal care products	Deodorants, shampoos, body lotions, moisturizing creams, toothpaste
Cosmetics	Lipsticks, facial creams, nail products, hair products
Detergents and cleaners	Liquids and granules, hard surface cleaners, softeners, dish washing products, bar soap (acids, unsaturates, moisture, glycerol, TiO_2)
Polishes	Waxes, silicone products, resins, aqueous based emulsions
Pharmaceutical (manufactured products)	Prescription drugs, consumer drugs, analgesics, elixirs and liquid formulations, packaging, ointments
Pharmaceutical development	Intermediates, reaction monitoring, final product assay, dosage forms, dissolution testing
Dental materials	Waxes, plastics, resins (ultraviolet photosetting)
Medical supplies	Packaging, sterilization methods, implants, syringes, dressings
Veterinarian products	Medications, dips, parasite treatments, animal food products
Aerosol products	Packaging, propellants, final product assay
Fabrics and fibers	Synthetic and natural fibers, orientation, blends, surface treatments, nonwoven fabrics
Packaging	Paper, cardboard, plastics, shrink wrap, multilayer laminates, bottles, bottle caps, aluminum foil, cans, coatings, inks
Films	Plastic film, shrink wrap, ionomers, coatings, laminates, photographic products (chemicals, film, photosensitive coatings, dyes, dye-couplers)

For these, mid-IR spectroscopy, used in the form of traditional analyzers, or as FTIR instrumentation, has become a method of choice.

Gas phase applications for dedicated process-related measurements can go beyond product manufacture, and can include industrial health and safety, hazard monitoring, and, as appropriate, the needs of various government agencies, including those that focus on the environment (United States Environmental Protection Agency), energy (Department of Energy) and defense (Department of Defense). In the case of many of these agency requirements, monitoring devices are not required by the agency directly, but are used as tools of compliance for industry.

The petrochemical and chemical industries are extremely diverse, and offer many opportunities for process IR

Table 5 Specialty manufacturing and fabrication

Application area	Applications
Electronics	Circuit board manufacture, photo-etching, coatings, fluxes, components (capacitors, resistors, etc.)
Magnetic and recording media	Magnetic tape production, magnetic media slurries, substrates, hard disk manufacture (lubricant coating), optical disks
Batteries	MnO_2 quality, packaging, electrolytes, membrane media
Semiconductors	Epitaxial film thickness, carbon and oxygen, BPSG, hydride layers, photoresist materials, etc.
Aerospace	Advanced materials and composites, wiring, electronic components, plastics, polymers and resins, fire-retardant materials, fuels, lubricants, metals and metal surface treatments
Electroplating	Plating bath solutions
Glass and optical materials	Optical filters, high-purity silica, quartz optics, laser optics (quartz, ZnSe, etc.), coatings, optical coatings
High-tech materials	Composites, special adhesives, ceramics, etc.
Nuclear and power	D_2O enrichment, nuclear fuels (isotope enrichment), coal, fuel oils, fuel oil residues, transformer oils, polychlorinated biphenyls

BPSG, boron–phosphorus silicate glass.

spectral measurements. Overall, they may be considered as segments of the hydrocarbon processing industry because natural gas and crude oil are considered to be the principal raw materials for a major part of these industries. Products manufactured by the petrochemical and chemical industries range from commodity materials (basic chemicals and raw materials) to fully formulated or compounded end-use products. Value-added products, which are manufactured towards the back end of the product chain, often require an advanced spectral analyzer, such as a process FTIR, whereas commodity chemicals may be adequately served with a traditional analyzer, such as an IR photometer.

In general, IR analyzers are viewed as cost-savers, providing information rapidly, often capable of providing it in ''real-time'', and with a minimal amount of operator intervention. IR spectral-based analyzers are still gaining attention in the process analysis world for measurements on almost all types of material.

6.1 Process Spectroscopy – a Reality Check

It has been assumed for some time that manufacturing companies want to make plants more efficient and wish to maximize profitability during production. In essence, this is true, but it is not always an automatic assumption that a company will actually implement changes, or install expensive equipment to make manufacturing more efficient. With a cost versus benefit analysis, a company might determine that the savings are too small to justify investment in technology to improve production. This in particular may be the case with an older plant, where a major overhaul may be required to make the changes cost-effective. Alternatively, adding new technology might be just a small part of an overall modernization scheme – it depends on the industry, and it depends on the company. Note also that the requirements for return on investment and/or pay-back of plant instrumentation vary significantly from industry-to-industry: some industries require payback within 6 months, for others, 4 or 5 years may be adequate.

The premise for the use of PAI is that the more information that is known about a process and/or a product, the better the opportunity to control the process and to reduce waste. Also, with the move towards improved quality, and the need to define procedures (ISO 9000, installation qualification/operational qualification/performance qualification (IQ/OQ/PQ) issues, good laboratory practice/good manufacturing practice (GLP/GMP), audit trails, etc.) opens the door for the use of improved analytical methodology. Another important factor, in the early 1990s was the initial reaction to the various amendments to the Clean Air Act. The result of these different influences was to create an over-exuberant atmosphere relative to the actual opportunities for process analytical instrumentation.

Bearing the above issues in mind, the following factors must be considered relative to the use of *advanced* process IR instruments, with special focus on advanced analyzers:

- Process photometers are well established, and are standard choices for certain, well defined applications. More advanced analyzers should only be considered if the application goes beyond the capabilities of the simple filter instrument.
- The use of NIR on-line process analyzers has become widespread for many areas of chemical and product analysis. At times, the use of these instruments may be constrained by the limited information content of the NIR spectrum, or by the overhead cost of maintaining a NIR calibration. The ability to extend a FTIR instrument to cover the NIR range has broadened the acceptance of both NIR and FTIR.

- Process FTIR has been implemented since the mid-1980s, but its use has been limited to mainly large chemical companies, and to specialty manufacturing industries (such as gases, semiconductors, pharmaceuticals).
- In traditional petrochemical and chemical production (primaries and intermediates), GC tends to be the PAI technique of choice. IR photometers are used, but mainly for environmental applications. IR spectrometers, including those based on FTIR methods are gaining acceptance for certain applications, such as fuel products (gasoline, diesel, etc.).
- Beyond the intermediates, and into the value-added business areas of the petrochemical and chemical industries, techniques such as NIR and IR are used extensively, but usually in an off-line role, or at the pilot plant level for reaction monitoring on-line.
- Technology champions within an organization have proved to be important for the successful implementation of advanced spectrometer systems.
- New emerging technologies, such as detector arrays and diode laser-based systems, are starting to be used for specialized applications. Scaling-down or miniaturization could lead to more cost-effective analyzers. If this occurs, it is expected to result in a paradigm shift relative to how process IR analyzers are perceived and implemented.

As a general commentary, a focus on a specific application has fostered a significant number of IR (both mid-IR and NIR) instrument developments based on either new or existing technologies. While such devices on the surface are limited to the application, with some modification they can be applied more generally. Many of the developments are either sponsored by governments (for military, aerospace, energy, and environmental applications) or are spawned from projects undertaken by engineering consulting firms for major manufacturing companies for a specific application. Such products come to market, often indirectly, as the result of partnerships between technology companies, instrument vendors, engineering companies and/or end-users.

One example is the refrigerant analyzer developed for the automotive industry for the determination of Freon®/halocarbon type in vehicle air conditioning systems. Specific filters define the analytes, and the packaging is small and inexpensive; such a system can be easily adapted for process applications. Examples such as this are applications driven, and the products are often developed quite independently of the traditional process analyzer business.

Full spectrum analyzers, such as NIR and FTIR spectrometers, have important attributes that can be turned into distinct advantages for certain applications. These include:

- the ability to handle a broad range of sample types;
- in-line sampling for gases and liquids (liquid ATR probes for FTIR);
- the ability to handle even corrosive materials with suitable hardware and sampling aids;
- the selectivity and specificity of IR (mainly mid-IR), with good spectral response for key structural/functional groups.

7 FUTURE TRENDS AND DIRECTIONS

A large number of traditional instruments are sold into industry for QC applications. For applications such as raw material screening this scenario is working adequately. However, there is a growing interest in the ability to make measurements in many areas of a process, with the idea that better production control can lead to a better control of the process and of the quality of the final product. While PAI have been a partial solution in certain industries, the cost of implementation is often very high, and it is impractical to implement more than a couple of instruments on a production line. Also, there is growing concern about the operating environment, worker safety and environmental controls. The ideal scenario would be to have the power of a full-blown vibrational spectrometric analyzer but with the cost and simplicity of implementation of a simple filter device.

There is a potential paradigm shift, with the understanding that if an analyzer becomes simpler and less expensive to implement then the role of PAI instruments could expand significantly. Furthermore, there is the understanding that "*good enough is OK*." This opens the way for spectrometer devices with lower resolution, shorter spectral ranges, and with multiple wavelength selection customized to the analysis. Many such devices have been already contemplated in different arenas, in particular for aerospace, military and defense applications, and in certain medical applications. Many of the technologies outlined in this article can lend themselves to this change in understanding. The use of specialized enabling technologies is expected to provide significant contributions, such as specialized optical coatings, microengineering and micromachining, and alternative light transmission and detection technologies. Most of these come from outside of the traditional instrument industry, and in the future we can expect new generations of instruments and sensor devices that come from sources other than the established instrument vendors. There have been discussions about instrumentation

and, in particular, spectrometers, on a chip. The technology exists, but the market and people are not necessarily ready for this. However, one can expect new generations of small-scale and miniaturized instrumentation in the near future, probably before the year 2005.

ABBREVIATIONS AND ACRONYMS

AOTF	Acousto-optical Tuneable Filter
ATR	Attenuated Total Reflectance
BPSG	Boron–Phosphorus Silicate Glass
CCD	Charge-coupled Device
CE	European Certification
CSA	Canadian Standards Association
CVF	Circular Variable Filter
DTGS	Deuterated Triglycine Sulfate
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GLP/GMP	Good Laboratory Practice/ Good Manufacturing Practice
IQ/OQ/PQ	Installation Qualification/Operational Qualification/Performance Qualification
IR	Infrared
J–T	Joule–Thompson
LPG	Liquid Petroleum Gas
LVF	Linear Variable Filter
MCT	Mercury–Cadmium–Telluride
MTBF	Mean Time Between Failure
NDIR	Nondispersive Infrared
NFPA	National Fire Prevention Association
NIR	Near-infrared
PAI	Process Analytical Instruments
PC	Personal Computer
QC	Quality Control
RAM	Random Access Memory
UA	Underwriter's Association
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Infrared Spectroscopy **(Volume 12)**
Interpretation of Infrared Spectra, A Practical Approach
• Theory of Infrared Spectroscopy

REFERENCES

1. J.J. Workman, Jr, et al., 'Process Analytical Chemistry', *Anal. Chem.*, **71**, 121R–180R (1999).
2. J.J. Workman, Jr, 'Review of Process and Non-invasive Near-infrared and Infrared Spectroscopy', *Appl. Spectrosc. Rev.*, **34**(1&2), 1–89 (1999).
3. D.A. Burns, E.W. Ciurczak (eds.), *Handbook of Near-infrared Analysis*, Practical Spectroscopy Series, Marcel Dekker, New York, Vol. 13, 1992.
4. M.J. Pelletier (ed.), *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Malden, MA, 1999.
5. G. Dent, J.M. Chalmers, *Industrial Analysis with Vibrational Spectroscopy*, Royal Society of Chemistry, Cambridge, 1997.
6. J.P. Coates, 'Vibrational Spectroscopy: Instrumentation for Infrared and Raman Spectroscopy', *Appl. Spectrosc. Rev.*, **33**(4), 267–425 (1998).
7. N.B. Colthrup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, San Diego, 1990.
8. D. Lin-Vien, N.B. Colthrup, W.G. Fateley, J.G. Grasselli, *Infrared and Raman Characteristic Group Frequencies of Organic Molecules*, Academic Press, San Diego, 1991.
9. J.J. Workman, Jr, 'Interpretive Spectroscopy for Near Infrared', *Appl. Spectrosc. Rev.*, **31**(3), 251–320 (1996).
10. X. Wang, J. Soos, Q. Li, J. Crystal, 'An Acousto-optical Tunable Filter (AOTF) NIR Spectrometer for On-line Process Control', *Process Control Qual.*, **5**, 9–16 (1993).
11. X. Wang, 'Acousto-optic Tunable Filters Spectrally Modulate Light', *Laser Focus World*, **28**(5), 173–180 (1992).
12. R.D. Hudson, Jr, 'Infrared System Engineering', in *Pure and Applied Optics*, ed. S.S. Ballard, Wiley-Interscience, New York, 1969.
13. W.L. Wolfe, G.J. Zissis, *The Infrared Handbook*, revised edition, The Infrared Information and Analysis (IRIA) Center, Environmental Research Institute of Michigan, Office of Naval Research, Dept. of the Navy, Washington, DC, 1985.
14. H.R. Schlossberg, P.L. Kelley, 'Infrared Spectroscopy Using Tunable Lasers', in *Spectrometric Techniques*, ed. G.A. Vanasse, Academic Press, New York, 161–238, Vol. II, 1981.
15. J.C. Wright, M.J. Wirth, 'Lasers and Spectroscopy', *Anal. Chem.*, **52**(9), 988A–996A (1980).
16. T.B. Sauke, J.F. Becker, M. Loewenstein, T.D. Gutierrez, C.G. Bratton, 'An Overview of Isotope Analysis Using Tunable Diode Laser Spectrometry', *Spectroscopy*, **9**(5), 34–39 (1994).
17. E.J. Lerner, 'Infrared Detectors Offer High Sensitivity', *Laser Focus World*, **32**(6), 155–164 (1996).
18. P.A. Wilks, 'High Performance Infrared Filtometers for Continuous Liquid and Gas Analysis', *Process Control Qual.*, **3**, 283–293 (1992).
19. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Interscience, New York, Vol. 83, 1986.
20. J.R. Ferraro, K. Krisnan (eds.), *Practical Fourier Transform Infrared Spectroscopy: Industrial and Chemical Analysis*, Academic Press, San Diego, 1989.

21. A. Ganz, J.P. Coates, 'Optical Fibers for On-line Spectroscopy', *Spectroscopy*, **11**(1), 32–38 (1996).
22. J.P. Coates, 'A Practical Approach to Quantitative Methods of Spectroscopic Analysis', in *Computer Methods in UV, Visible and IR Spectroscopy*, eds. W.O. George, H.A. Willis, Royal Society of Chemistry, Cambridge, UK, 95–114, 1990.
23. J.J. Workman, Jr, P.R. Mobley, B.R. Kowalski, R. Bro, 'Review of Chemometrics Applied to Spectroscopy: 1985–1995, Part 1', *Appl. Spectrosc. Rev.*, **31**(1&2), 73–124 (1996).
24. 'Standard Practice for Infrared, Multivariate Quantitative Analysis', The American Society for Testing and Materials (ASTM) Practice E1655–94 (1995). ASTM Annual Book of Standards, West Conshohocken, 755–779, Vol. 03.06, 1996.
25. K.R. Beebe, R.J. Pell, M.-B. Seasholtz, *Chemometrics: A Practical Guide*, John Wiley, New York, 1998.
26. J.J. Workman, Jr, J.P. Coates, 'Multivariate Calibration Transfer: The Importance of Standardizing Instrumentation', *Spectroscopy*, **8**(9), 36–42 (1993).
27. T.J. Porro, S.C. Pattacini, 'Sample Handling for Mid-infrared Spectroscopy, Part I: Solid and Liquid Sampling', *Spectroscopy*, **8**(7), 40–47 (1993).
28. T.J. Porro, S.C. Pattacini, 'Sample Handling for Mid-infrared Spectroscopy, Part II: Specialized Techniques', *Spectroscopy*, **8**(8), 39–44 (1993).
29. P.B. Coleman (ed.), *Practical Sampling Techniques Techniques for Infrared Analysis*, CRC Press, Boca Raton, 1993.
30. N.J. Harrick, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, Ossining, New York, 1987. (Original Publication by John Wiley & Sons, New York, 1967.)
31. F.M. Mirabella, Jr (ed.), *Internal Reflection Spectroscopy: Theory and Applications*, Marcel Dekker, New York, Vol. 15, 1993.

Mass Spectrometry in Process Analysis

José A. Olivares
Los Alamos National Laboratory, Los Alamos, USA

1 Introduction 1
2 History 2
3 On-line Sampling 2
3.1 Instrumentation 2
3.2 Detection Limits and Sensitivity 7
3.3 Software and Instrument Control 7
4 Applications 8
4.1 Process Monitoring in Biotechnology 8
4.2 Process Monitoring of Flue Gases 9
4.3 Process Monitoring in the Petrochemical Industry 10
4.4 Process Monitoring in the Semiconductor Industry 10
5 Quality Control 11
5.1 Interferences 11
5.2 Analysis of Complex Mixtures 12
5.3 Principles of Quantitative Analysis 13
6 Comparison with Other Spectroscopic Methods 13
6.1 Comparison of Technology 13
6.2 Business Issues 14
Acknowledgments 14
Abbreviations and Acronyms 14
Related Articles 15
Further Reading 15
References 15

The mass spectrometer can provide a versatile, highly specific and sensitive tool for the analysis and control of industrial processes. More common techniques for industrial process control are gas chromatography and infrared spectroscopy. Ion mobility spectrometry and chemical sensors are slowly being introduced. Although mass spectrometry is relatively more complex and expensive to implement than these other techniques, the sensitivity and specificity offered is often unmatched. This article gives a general review of the roots of mass spectrometry as it applies to on-line process control. A short discussion of instrumentation commonly used in the practice of this technique has been added and limited only to the more popular and currently commercially available instruments. These include sector, quadrupole and time-of-flight mass spectrometers. The capabilities and specifications of the different classes of mass spectrometers are discussed and compared. Specific sampling methodology for the operation of these instruments in on-line processes is given. Also included is a discussion of common practices in the reduction of interferences and deconvolution of mass spectra from mixtures. In-depth examples of industrial process and control applications have been selected and are discussed in the areas of biotechnology, technical incineration, petrochemical, and semiconductor production.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Mass spectrometry is an analytical technique that has been applied to the study and solution of a diverse number of chemical problems. Owing to the complexity of the instrumentation, the majority of the chemical analysis work performed by mass spectrometry occurs in the laboratory, where controlled conditions exist and time for decision-making can be sacrificed for the ultimate in sensitivity and detectability. Nevertheless, mass spectrometry provides a viable tool for the solution of process control and analysis in many industrial settings. In this case, the chemical and physical constraints, and need for prompt action, require that the instrument sample the process on-line and provide real-time feedback control to the operation. In this article we will review some of the aspects of mass spectrometry related to its use in industrial process analysis as it applies to the control of a production process on-line, i.e. the mass spectrometer is physically attached to the process as a sensor that provides real-time feedback. This is in contrast to the more common uses of mass spectrometry at-line or in analytical laboratories, where the process is sampled through grab samples and feedback is not in real-time. The discussion is strictly limited to this topic, even though other problem categories have often been associated with process analysis. Thus, important topics to which mass spectrometry contributes heavily, such as environmental analysis, medical diagnostics, thermal gravimetric analysis, reaction monitoring, and others, are not considered in this article.

The mass spectrometer can provide a versatile, highly specific and sensitive tool for the analysis and control of industrial processes. More common techniques for industrial process control are gas chromatography and infrared spectroscopy. Ion mobility spectrometry and chemical sensors are slowly being introduced. Although mass spectrometry is relatively more complex and

expensive to implement than these other techniques, the sensitivity and specificity offered is often unmatched. This article gives a general review of the roots of mass spectrometry as it applies to on-line process control. A short discussion of the instrumentation commonly used in the practice of this technique has been added and limited only to the more popular and currently commercially available instruments. These include sector, quadrupole and time-of-flight mass spectrometers. The capabilities and specifications of the different classes of mass spectrometers are discussed and compared. Specific sampling methodology for the operation of these instruments in on-line processes is given. Also included is a discussion of common practices in the reduction of interferences and deconvolution of mass spectra from mixtures. In-depth examples of industrial process and control applications have been selected and are discussed in the areas of biotechnology, technical incineration, petrochemical, and semiconductor production.

2 HISTORY

The history of mass spectrometry as it pertains to industrial process control can be pieced together from a number of classical mass spectrometry texts by Hill,[1] Roboz,[2] and Watson,[3] and reviews by Huber[4] and Cook et al.[5] The first reported use of mass spectrometry to control process streams goes back to the 1940s. In this time period the development of the petrochemical industry required that the precise chemical composition of the feed materials be known. This work was performed off-line with grab samples taken to the laboratory for analysis. The emergence of the first commercial mass spectrometers helped to spark the study of cracking patterns of organic compounds for the petroleum industry. Based on Dempster's single-focusing mass spectrometer (i.e. direction focusing), the Consolidated Engineering Corporation (Pasadena, CA) Model 21-101–103 instruments were responsible for much of the early production of mass spectral data compiled by American Petroleum Institute Research Project 44. These and other commercial instruments could be operated easily by chemists, who focused on the improvement of sample-handling methods. Organic compounds up to 500 Da and with melting points up to 350 °C could be studied using single-focusing mass spectrometers, which provided resolutions of 1000–4000, with mass ranges up to 2000 Da. High-resolution instruments based on double-focusing mass spectrometers (i.e. momentum and velocity focusing) utilizing Nier–Johnson and Mattauch–Herzogg geometries provided an important contribution to the identification of petroleum products. These instruments, with mass resolutions of 5000–100 000, provided the ability to resolve difficult mixtures of compounds in which two or more types of compounds appear at the same nominal mass. Thus, hydrocarbon types in high-boiling-point distillates (160–510 °C) could be identified within minutes. The cycloidal mass spectrometer, introduced in the late 1930s,[6] was first used as a compact residual gas analyzer by Robinson and Hall in 1956.[7] This instrument provides the stability and sensitivity available in magnetic sector instruments in a small package that can be portable and placed easily on an industrial process.

The above-mentioned instruments are termed static spectrometers because the operational parameters of the mass spectrometer remain constant with time, except for the slow change in magnetic and electrical fields required to record a mass spectrum. Similarly, a number of dynamic mass spectrometers have been used to make major contributions to process control and are by far the most popular in use today. Dynamic mass spectrometers are classified as those in which the time dependence of one or more instrument parameters, e.g. electrical field strength, magnetic field strength, or ion movement, is fundamental to the mass analysis. Blauth[8] has published a classical treatise on dynamic mass spectrometry. Some early dynamic mass spectrometers, such as the omegatron, the linear radiofrequency (RF) mass spectrometer, and the farvitron, have enjoyed much success as residual gas mass analyzers for vacuum, rocket, and space exploration, owing to their sensitivity, compactness and lightweight design.[4] Today, the time-of-flight mass spectrometer and the quadrupole mass spectrometer have become the most popular for use in residual gas analysis and process control. The time-of-flight mass spectrometer was first proposed and put into practice in the late 1940s by Stephens.[9] Early on, these instruments achieved partial pressure detection limits of 10^{-12} mbar for residual gas analysis in vacuum processes, with mass resolutions of 10–60.[10,11] The quadrupole mass spectrometer was first introduced by Paul[12] in the early 1950s. The first use of quadrupole mass spectrometers as residual gas analyzers was described by Guenther,[13] and since then they have become by far the most popular process analysis mass spectrometer.

3 ON-LINE SAMPLING

3.1 Instrumentation

The process mass spectrometer consists of several key components: the mass analyzer, which actually does the mass separation of the analytical sample; the ion source, required to produce gas-phase ions of the analyte for separation; the detector, used to collect the separated analyte ions and provide a signal that can be measured

and recorded; and the sampling interface, which actually introduces the analytical sample from the process to be measured into the ion source of the mass spectrometer. These components are described here in some detail, with a focus on currently available commercial mass spectrometers used for process monitoring.

3.1.1 Mass Analyzers

The discussion in this article is limited to some of the most common mass analyzers in use today: cycloidal, time-of-flight, and quadrupole mass spectrometers, yet there are many examples in the literature that use other types: magnetic sector, electrostatic sector, quadrupole ion trap, ion mobility, omegatron, farvitron, and linear RF mass spectrometers. In-depth descriptions of these instruments can be found elsewhere in this publication and in other reviews.[4,5,8,14,15] In principle, any type of mass analyzer can be used for process control. The basic objectives that are most often associated with mass analyzers are mass range, resolution, ion source type, detector type, and sensitivity. Other basic requirements such as size, utility requirements, and system control specifications are more important for site location than actual performance of the instrument, but nevertheless are major contributors to the successful implementation of on-line process analyzers. Both instrument performance and physical specifications make up the basic selection criteria for specific applications.

Table 1 gives some of the basic specifications for common process analyzers available commercially. As can be seen, most residual gas analyzers have been limited in their mass range to 100, 200 or 300 Da. The limitation is due to practicality. Volatility for organic compounds generally decreases with increasing mass. Furthermore, as the volatility decreases, sample introduction into a mass spectrometer becomes more complicated, owing to contamination of the sample inlet, ion source and mass spectrometer vacuum system, resulting in low sensitivity and poor detection. Gas-leak detectors will usually limit the mass range to the leak gas to be detected. Helium is the most useful for this purpose, owing to its low mass, high diffusivity, its inertness and its inflammability, unlike hydrogen. Therefore, gas-leak detectors are usually limited to a mass range of 4 Da. For most residual gas analysis applications involving simple inorganic gases and light hydrocarbons, a 100-Da mass range is sufficient for analysis. In fact the $C_3H_7^+$ peak at mass 43 can be used reliably to estimate the total hydrocarbon partial pressure of a system within a factor of 2.[16] Higher mass ranges may be needed, depending on the application. The trade-off tends to be resolution, sensitivity, and size (physical or power consumption) of the instrument.

The resolution (R) of the mass analyzer is defined as $M/\Delta M$, where M is the mass-to-charge ratio, where the measurement is made, and ΔM is the mass width of the ion intensity peak at 10% of the height. For quadrupole instruments ΔM is typically unity across the mass range. For time-of-flight instruments, R should be specified at a particular mass because this quantity will vary with mass. For magnetic sector instruments, e.g. the cycloidal, R is constant across the mass range because this is fixed by

Table 1 Process mass analyzer specifications

	Quadrupole	Cycloidal	Time-of-flight
Mass range (U)	1–100 1–200 1–300	2–200 10–200	0–350
Resolution ($M/\Delta M$)	$\Delta M \sim 1$ over the mass range	50–100	100 at 56 Da
Sensitivity	10^{-6}–200A mbar^{-1}	10^{-4} A mbar^{-1}	3 μA ppt^{-1} H_2O in argon/nitrogen
Detector type	Electron multiplier or Faraday	Electron multiplier or Faraday	Electron multiplier
Detection limit	10^{-11}–10^{-14} mbar (argon) <500 ppb H_2O in argon	10^{-11} mbar Ratio 100 ppb	20 ppt H_2O in argon/nitrogen
Linearity Range	10^8	10^7	10^5
Useful Pressure Range (no restrictions)	10^{-4}–10^{-9} mbar	5–10^{-6} mbar	Atmospheric (API)
Sampling pressure	10^{-3}–10^3 mbar	10^{-3}–10^3 mbar	Atmospheric (API)
Manufacturer	Pfeiffer Vacuum MKS Instruments Inc. ABB Extrel Corp. Ametek Leybold Inficon Ferran Scientific, Inc. Spectra Instruments, Inc.	Foster-Monitor Group	Sensar Larson-Davis

API, atmospheric pressure ionization.

the geometry of the instrument. Thus, the resolution of different types of instruments should be compared at the masses of interest for each application.

3.1.2 Ion Sources

In order to generate a mass spectrum of a molecule, gas-phase ions of the molecule need to be created. Most process mass analyzers use electron impact ionization. In this technique the sample gas is introduced into a low-pressure region (typically 10^{-4} mbar or lower) near the physical entrance to the mass spectrometer. In this region, sample molecules encounter energetic electrons (10–100 eV) that are emitted from a resistively heated filament. When the electrons collide with the sample molecules, ionization can occur by the removal of an electron from the analyte molecule to form a positive ion. This ion is then accelerated and focused onto the mass spectrometer. The ion source configuration is determined by the required sensitivity, the pressure of the analyte stream, reactivity of the analyte and matrix gases, and interferences that can be caused from these reactions and the walls of the chamber. The filament material tends to be thin tungsten wires. Hydrocarbons react with hot tungsten to form W_2C on the surface of the filament. This affects the electron emissivity of the filament and requires filament conditioning to form a durable W_2C that will not change significantly during the measurement period. Rhenium is another popular filament material that does not form stable carbides, and its oxides are unstable at the required emission temperatures. Furthermore, it does not become brittle at high temperatures and conditioning is not necessary, as it is for tungsten. The filament lifetime tends to be shorter than for tungsten, due to the higher volatility of rhenium. Rhenium is a good material to use when the sample stream contains oxygen or oxygenated compounds. Tungsten reacts with oxygen to form tungsten oxide and carbon monoxide, decreasing the life of the filament and increasing the spectral background relative to rhenium. Yet tungsten is preferred in high-vacuum environments where the volatility of rhenium could be problematic and the long lifetime of tungsten filaments becomes an advantage. Another consideration is the high temperature of $>300\,^{\circ}C$ achieved by the source, which is known to affect the sensitivity and cracking pattern of the analyte. In this case, thoriated tungsten oxide filaments, lanthanum boride on rhenium filaments, and platinum–iridium filaments provide excellent electron emission characteristics at relatively low temperatures.

Electron impact sources vary in configuration but the chamber is kept as gas-tight as possible to minimize the sample size required, reduce background gases, and minimize effects from variations in pumping speed on the signal. Closed-type ion sources accomplish this by physically enclosing the ion formation region near the gas inlet with a metal box or cylinder. The closed-type ion source is best used when the process to be monitored is at a pressure significantly higher than the operating pressure of the mass spectrometer, and the gas can be introduced efficiently into the ion source using an inlet capillary or orifice. Open-type ion sources utilize a grid as the enclosure for the ion formation region, and allow the gas in the vacuum chamber to penetrate freely into the ion source. The open-type ion source is utilized when the mass spectrometer is immersed into a low-pressure process chamber that is to be monitored.

Another type of ion source used in commercial process analyzers is the API source. This source, as the name suggests, operates at atmospheric pressure. Ions are generated in a small corona discharge created when current-limited high voltage is applied to a platinum needle in close proximity (5 mm) to a metal plate. Analyte gas at or near atmospheric pressure is introduced directly into the plasma chamber, where it is ionized. The metal plate contains a small orifice, typically 0.5 mm, that samples the plasma gases and ions into the mass spectrometer's vacuum chamber for analysis. The advantage of this type of source is that it can be used to analyze ambient pressure gas streams or gas streams that can be sampled and swept with an inert gas. Furthermore, this type of source tends to produce mostly molecular ions with very little, if any, fragmentation. The main disadvantages are that it cannot be used to analyze vacuum processes and it requires a differential vacuum pumping system.

3.1.3 Detectors

Process mass analyzers typically will have two types of detectors: the Faraday cup and the electron multiplier. These two detectors often can be found together on systems that require a wide dynamic range in detection. The Faraday cup is a simple small metal electrode; in some instruments it is shaped as a plate but in others it is a small cup that is placed at the end of the analyzer. Ions exiting the analyzer impinge on the metal surface. A high-impedance amplifier or fast electrometer reads the ion current impinging on the metal surface. Often the Faraday plate or cup is shaped or guarded in such a way that the secondary electron emission emitted by the energetic ions impinging on the metal surface is minimized. This is the simplest type of ion detector that could be placed on a mass spectrometer. Depending on the characteristics of the fast-amplifier electrometer system used, ion currents of 10^{-15}–10^{-6} A can be read easily with this system. Although simple in operation, the Faraday cup has some other drawbacks: susceptibility

to vibrational and electrical noise; and, the lower the ion current, the more complex the electrometer and the larger the time constant for proper readout of the ion current.

Electron multipliers offer an additional 100–1000-fold enhancement in sensitivity and detection for process mass analyzers. In this system a resistively coated tube or set of plates amplifies each ion impact on its surface by emitting 10^3–10^6 electrons. This amplified electron current is then read by a fast-amplifier electrometer. The electron multiplier is typically held at 1000–2500 V negative, with respect to the ion beam, for proper electron emission and gain. Although quite sensitive and very fast in response, this detector has limited linearity and is quite susceptible to burnout by large ion currents and high operating pressures. It is best used in conjunction with a Faraday cup, especially when the analyte composition to be measured can vary widely. Electron multipliers also can be operated in single-ion-counting mode, giving the possibility for the extreme in sensitivity and higher linearity. The instrument's linearity is an important measurement parameter for process analytical instruments. In mass spectrometers it is typical to have 6–9 orders of magnitude in linearity when using Faraday cup and electron multiplier detectors. Thus, it is possible to measure trace components in the parts-per-billion range, and major components in the percentage range, with these instruments.

3.1.4 Sampling System

The term ''on-line'' analysis is used to indicate the direct coupling of the mass spectrometer to some process that is to be monitored.[17] The sampling system – considered as being the interface between the process to be measured and the mass spectrometer – is a critical component for successful operation. In a low-pressure (usually $<10^{-3}$ mbar) process the ion source to the mass spectrometer is immersed directly in the chamber to be tested and is called the ''nude'' source. This provides the greatest conductance for the sample to make it to the mass spectrometer, and therefore the greatest sensitivity. If the process to be monitored is at higher pressure, a restriction capillary between the process and the mass spectrometer's ion source must be provided that will minimize the amount of gas that reaches the ion source and the analyzer. In direct inlet sampling, two types of gas flow are usually encountered, depending on the interface design, molecular flow and viscous flow. In molecular flow the restrictions into the ion source are smaller in diameter than the local mean free path of the gas, and there are few interactions, if any, between the gas molecules. In viscous flow, the mean free path of the sampled gas is small compared to the dimension of the sampling pipe or orifice, and there are ample collisions between the molecules. If molecular flow is achieved along long paths with low conductance, the sampled gas is biased (fractionated) in favor of the lighter gases by a factor of $(MW)^{-1/2}$ due to the velocities achieved, where MW is the molecular weight of the molecule.

Molecular flow sampling has the advantage that the partial pressures of the gases in the sampled reservoir can be represented accurately at the ion source, provided that the reservoir is large. A frit can be used to restrict the flow between the sample reservoir and to provide molecular flow at the source. A restriction capillary, 10–200 μm i.d. and several centimeters in length, is typically used to sample gases near atmospheric pressure into the low-pressure ion source region. This sampling system provides for viscous flow throughout most of the capillary, thus minimizing fractionation effects while providing molecular flow in the immediate region of the ion source. The main disadvantage of viscous flow systems is that the flow is controlled by the viscosity of the gas to be sampled and can vary with the sampled mixture. Furthermore, the partial pressure of a component in the source is not proportional to its mole fraction in the sample, and is not independent of other components present in the mixture.[18] A diagram of such a system is shown in Figure 1, where the restriction capillary along with plumbing, pumps, gauges and valves provide an appropriate sampling system into a process analyzer.

Viscous flow can be achieved for the most of such a system, although true viscous flow is never achieved

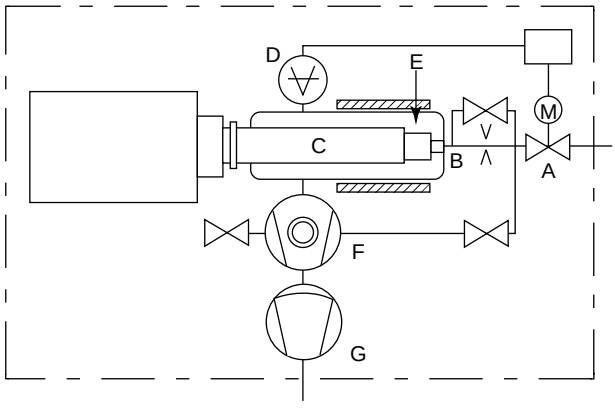

Figure 1 Vacuum diagram of a process analyzer for use in gas analysis. The optimum pressure range is 5×10^{-3} to 1000 mbar. The pressure in the analysis chamber is automatically controlled by a metering valve (A) in front of the restriction capillary (B). Other components include the process mass spectrometer (C), vacuum gauge (D), heated vacuum chamber (E), turbo vacuum pump (F) and rotary vane vacuum pump (G). (Reproduced by permission of Pfeiffer Vacuum from technical note BG800318PE, Copyright Balzers AG.)

because at some point near the ion source the flow in such a restriction must change to molecular flow. Nevertheless, the sampled gas is said to be representative of the gas in the process chamber. Careful calibration can result in corrections for biases that give an accurate measurement of the process. Of course, the exception to the need for a restriction capillary, when sampling ambient pressure processes, is when the ion source to the mass spectrometer operates at high pressure, as is the case for the corona discharge ion source of Sensar's model TOF 2000 (see Table 1). But even in this case there is a 0.5-mm restriction, with differential pumping, between the ion source and the analyzer.

A process analyzer can monitor more than one process, the total number depending on the speed and duration of each process to be monitored. Figure 2 shows a diagram of a process analyzer mass spectrometer connected to a multiple gas stream selector through a heated restriction capillary. Heating and passivation of metal and glass tubes used for sampling a stream are crucial when the gases to be sampled can condense (are adsorbed) and/or react on surfaces. Lining the inside of metal tubes with glass that has been passivated by the method of silanization is a common practice.[19] Notice that the gas in the lines to be sampled is not left stagnant in the stream selector manifold but it is allowed to flow through to waste or back to the process. This ensures continuous flow and rapid response to changes in stream composition.[5] In this manner one process mass analyzer may be multiplexed to many process streams (2–64 are common).

Sampling of gases dissolved in gas and liquid streams can be achieved with the use of a membrane inlet. The basis of this technique is the selective transport of analyte molecules across a semipermeable membrane into the ion source of a mass spectrometer. The technique, termed membrane inlet mass spectrometry (MIMS), has been reviewed extensively.[20,21] In practice a membrane is used to separate the liquid or gas process stream from the vacuum chamber of the mass spectrometer, as shown in Figure 3. The membrane material should have low permeability for the matrix and high permeability for the analytes of interest. Polypropylene, polytetrafluoroethylene, Teflon™, cellulose, silicone rubber, dimethylvinyl silicone, polyethylene and zeolites are all common membrane materials. Analyte in the process stream diffuses into one side of the membrane and is extracted directly into the gas phase on the vacuum side into the mass spectrometer. Physical (temperature, size exclusion, and diffusion) and chemical interactions (affinity) with the membrane can provide selectivity, preconcentration, and in some cases separation of the analytes. Thus, an overall concentration of the analyte with respect to the matrix is achieved that can be up to a factor of 100.[5] This technique is especially useful where complete vaporization and introduction of the sample into the analyzer would not be reasonable, due to large excess of matrix, time considerations, and sensitivity. Yet, the number of volatile components in the liquid matrix is low and they can be resolved. Although detection limits in the order of parts per quadrillion have been reported for some processes,

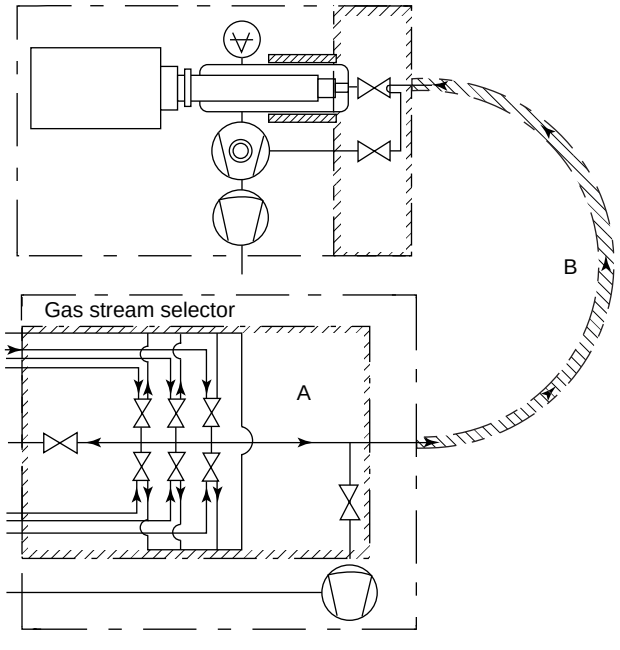

Figure 2 Vacuum diagram of a process analyzer for use in the analysis of multiple gas streams. A heated valve system (A) and transfer line (B) provide the ability to multiplex sample streams and calibration gases into the mass spectrometer. Other components shown are as in Figure 1. (Reproduced by permission of Pfeiffer Vacuum from technical note BG800318PE, Copyright Balzers AG.)

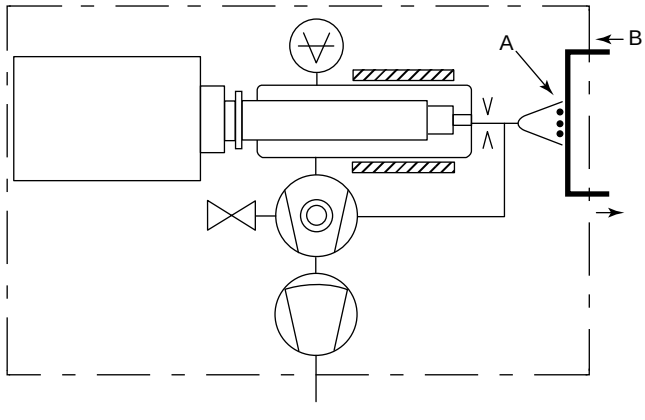

Figure 3 Vacuum diagram of a process analyzer for use in the analysis of volatile compounds, in liquids. A membrane gas inlet (A) provides the interface between the liquid stream (B) and the process chamber. Other components shown are as in Figure 1. (Reproduced by permission of Pfeiffer Vacuum from technical note BG800318PE, Copyright Balzers AG.)

some inherent weaknesses are worth mentioning. The membranes can suffer from long response times arising from the diffusion time constant through the membrane, which can contribute to memory effects, and long cycles for the mass transport of the liquid to the membrane.

Finally, it is worth mentioning at this point that mass spectrometers require ambient pressures of $<10^{-4}$ mbar for proper operation. This is mainly due to the fact that at higher pressures, collisions between the analyte ions and background gases affect the ion paths to a greater extent than the mass spectrometer's electrical or magnetic fields, thus causing loss of sensitivity and resolution. Thus, the mass spectrometer must reside in a vacuum chamber, the complexity of which depends on the application and sample inlet chosen. For example, in semiconductor applications, e.g. vapor deposition, the vapor deposition chamber typically serves as the vacuum system for the mass spectrometer, the ion source is typically nude, and no additional pumping may be required. On the other hand, applications of process mass spectrometry to monitor a high-pressure reactor, e.g. in petrochemical applications, would require a capillary inlet for sampling, and turbomolecular or diffusion pumps along with mechanical forepumps to achieve the required operating pressures. Roboz[2] gives an excellent discussion of vacuum systems for mass spectrometry.

3.2 Detection Limits and Sensitivity

Two measurement parameters are used to specify the detection capability of mass spectrometers used for gas analysis: the limit of detection (LOD) and the sensitivity. Although somewhat related, these two parameters describe two important but different features of the instrument. The sensitivity of the instrument is generally given as the expected signal when a known amount of sample is analyzed. For residual gas analyzers operating in the electron impact mode, the typical units are detector current per unit of pressure ($\mathrm{A\,mbar^{-1}}$). This quantity should be specified for each detector, because more than one may be used on the mass spectrometer. For Faraday cup detection this quantity tends to be around $10^{-3}\,\mathrm{A\,mbar^{-1}}$. Because an electron multiplier detector can easily provide gains of 10^4–10^8, it is not uncommon to see sensitivity in the range of 1–$200\,\mathrm{A\,mbar^{-1}}$. The sensitivity is dependent on the analyte to be measured and its ionization cross-section. For instrument comparisons one should use the same gas (typically nitrogen), detector, ion source, and electron emission current when specifying this quantity, or provide a range of sensitivities. The detector current per analyte gas concentration (e.g. $\mathrm{A\,ppm^{-1}}$ of N_2) is another way of expressing this quantity. Table 1 provides typical sensitivities for common process analyzers.

The LOD of the instrument is a measurement of the minimum quantity of analyte that can be measured statistically above the background noise of an instrument. The units are typically partial pressure of the analyte (e.g. mbar of H_2O), although minimum concentration of the analyte in a particular gas is also widely used (e.g. ppm of H_2O in argon). This quantity is also dependent on the instrument arrangement, and instrument parameters should be very similar when attempting to compare instruments. Typical values of LOD for instruments with Faraday cup detection and electron impact ionization are 10^{-11} mbar of Ar and ca. 1 ppm Ar in N_2. Electron multipliers will lower these quantities by a factor of 10^3 to 10^{-14} mbar and 1 ppb. The atmospheric pressure ion source can be extremely sensitive for some gases, bringing these quantities down to 1 ppt. Furthermore, the use of special inlets and preconcentration, such as MIMS, has been shown to give LODs in the parts-per-quadrillion range. In gas leak detection the LOD is given as the minimum detectable gas leak (helium) per unit of time. This is the leak rate or gas throughput of the leak, i.e. standard cubic centimeters per second ($\mathrm{cc\,s^{-1}}$) or $\mathrm{mbar\,L\,s^{-1}}$. For instruments using a Faraday cup and dedicated towards this type of analysis, this quantity can be extremely low, of the order of $10^{-12}\,\mathrm{mbar\,L\,s^{-1}}$.

3.3 Software and Instrument Control

Integration of the mass spectrometer with a process can result in on-line instantaneous decision-making. In complex processes, usually this is not done just from feedback from the mass spectrometer data by itself, but also from feedback from other physical and chemical sensors on-line. The integration of such a complex system may require more sophistication and customization than any one process analyzer data package can provide by itself. Nevertheless, instrument control for most small process analyzers is improving. Implementations of graphical display languages in process analyzers that allow multicomponent software programming, control and data acquisition are quite common. Most analyzers are now available with a number of digital and analog output and input hardware and software features that go beyond the collection of a mass spectrum. The hardware and software are often capable of controlling the immediate processes surrounding the mass spectrometer, alerting the operator to changes in conditions, and controlling a number of analyzer or analysis streams through multiplexing and multitasking. Yet the traditional components to mass spectrometer software that are essential include data analysis routines, spectral libraries, calibration methods, spectrum acquisition, and data logging.

4 APPLICATIONS

4.1 Process Monitoring in Biotechnology

Process monitoring in the different biotechnology industries can be exemplified by a number of areas in which mass spectrometry is being used to make major contributions. These involve the integrity of containers used to seal medical drugs, monitoring of the production of drugs such as penicillin, and monitoring of alcohol production.

4.1.1 Monitoring Container Integrity

The integrity of vials used for drug containment, transportation, and delivery is critical to the pharmaceutical industry. Microbial challenge testing is used to test the lots statistically for sterile conditions. Assuming that vial sterility is lost if a container is flawed or improperly sealed, rather than from the contamination during manufacturing and cleaning, helium leak testing of vials can be used to ensure sample integrity, even after filling and sealing. Kirsch et al.[22] demonstrated that vials sealed under a helium tracer atmosphere with butyl rubber stoppers and crimps, or alternatively sealed under normal conditions and charged with helium tracer, could be monitored for helium leakage with a sniffing probe, using mass spectrometry. The absolute leak rate of the vials was correlated with the squared nominal leak radius (0.5–10 μm) of micropipettes glued to the sides of vials to produce standard leaks for calibration and reference. The minimal observable leak rate under this mode of operation (sniffing) was about $10^{-6.6}$ cc s^{-1}. This was mostly attributable to leakage through helium permeation of the rubber septa.

4.1.2 Monitoring of Penicillin Production in a Fermentor

Typically, fermentation processes are monitored using headspace gases.[23] In an interesting application of process mass spectrometry, Hansen et al.[24] have demonstrated the ability to monitor directly the production of penicillin in a fermentor. These authors used MIMS to monitor the consumption of phenoxyacetic acid (POAA) during the production of penicillin V in a 20 L reactor. POAA is the side-chain precursor to penicillin V. Accurate monitoring and control of POAA can result in optimum yield. The fermentor broth was sampled through a microfiltration unit, acidified to 0.1 M HCl, heated to 70 °C, and passed by a silicone membrane. Acidification was necessary for efficient diffusion of the POAA through the silicone membrane and increased sensitivity due to the low volatility of POAA, which boils at 285 °C. A cycle time of 20 min for each measurement was necessary due to the long transfer lines, washing cycle, microfiltration apparatus, and standardization cycle. This duty cycle was sufficient to monitor effectively the production of penicillin V over a production time frame of 150–200 h. The loss of POAA as monitored via the 107-Da ion was due to decarboxylation of the molecule. The loss of POAA has a direct correlation to the formation of penicillin V, as measured off-line via grab samples. The detection limit for POAA was 0.1 g L^{-1}. Alternative techniques for monitoring this process include liquid chromatography and flow injection analysis, but these techniques do not have the specificity that is available with MIMS.

4.1.3 Monitoring Alcohol Production in a Fermentor

Monitoring of ethanol production in a pilot plant has been shown by Johnson et al.[25] using MIMS. These authors used a flow injection system to sample 9000-L reactors. The sampling line from the reactor recirculates at 10 gallons per minute, with 3 mL min^{-1} of sample from this flow directed to the MIMS system. A silicone membrane heated to 91 °C with a helium jet separator was used to monitor the aqueous sample for ethanol, acetic acid, and lactic acid. An ion trap mass spectrometer was used to scan the mass range of interest. Each sample injection was followed with an injection of a calibration standard to provide continuous calibration of the system. Continuous operation with 3-min cycles between samples and analysis is possible. The system is relatively accurate (2.36%) when compared to standard methods using liquid chromatography, and has excellent precision over periods of days. The system required some daily maintenance in the form of flushing the injection lines twice daily to avoid broth build-up on the lines. Figure 4 shows the simultaneous display of ethanol signal from a process

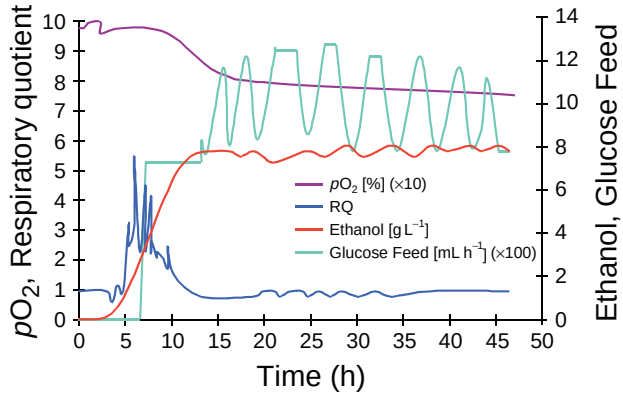

Figure 4 Simultaneous display in a process mass analyzer's data package system of the mass spectrometer data and other relevant data for a fermentation process. (Reproduced by permission of Pfeiffer Vacuum and Novartis from technical note BG800318PE, Copyright Novartis Ltd.)

mass spectrometer along with other fermentor relative data used to control an ethanol fermentation process.

4.2 Process Monitoring of Flue Gases

Technical incineration presents an excellent challenge for process monitoring.[26] In typical technical incinerators fundamental production processes can take place, e.g. metallurgical production, along with energy production. In these plants it is common to monitor flue gases for CO, CO_2, NO_x, and total organic carbon (C_xH_y) production. The CO level, for example, then can be used as an indicator of the burning conditions and the production of possible toxic substances, e.g. polyaromatic hydrocarbons (PAHs) and chlorinated aromatics. Yet no information on the contribution of particular toxic compounds can be obtained and used for control purposes. Typical control parameters in such a system would involve mass flows of air/oxygen, recirculation of exhaust gases, mass flows of the solid and fluid feeds, and temperature profiles in the combustion chambers. Zimmermann et al.[26] have shown the ability to monitor flue gases in a technical incinerator for the presence of benzene, toluene, biphenyls, PAHs, and benzofurans at the parts-per-billion detection level. Furthermore, they demonstrated the ability of these signals to be used as indicators of incinerator conditions showing faults, instabilities, or improper combustion feed processes that can be used for proper control of the incinerator.

Monitoring of trace components in flue gases poses a particular challenge to mass spectrometry due to the complexity of the gas composition to be sampled, the presence of particulates, condensation and reaction of analytes with the sampling lines, and physical blockage of sampling orifices by tarry substances. Furthermore, the apparatus has to be operated in a physically constrained and harsh environment near the furnaces. Zimmermann et al.[26] solved these problems by employing a rugged and mobile time-of-flight mass spectrometer using resonance-enhanced multiphoton ionization (REMPI). The REMPI process utilizes a laser (KrF excimer at 248 nm, 50 Hz, 10-ns pulses, 10 mJ pulse energy) to select optically the analytes of interest and ionize them. The ionization takes place at the end of an effusive molecular beam inlet into the mass spectrometer. The selectivity of such a system can be very high. For example, at this wavelength (248 nm) only aromatic hydrocarbons tend to ionize, whereas aliphatic hydrocarbons, substituted aromatics, carbonyls, and inorganic gases do not. Furthermore, the ionization is soft, i.e. little or no fragmentation of the molecule is observed. The sampling lines to the mass spectrometer were made out of quartz capillary and tubing and heated to 650 K, to minimize reactivity and condensation in the lines. These authors were able to show the ability of this apparatus to detect parts-per-billion levels of PAHs and benzofurans, and correlate their presence to combustion conditions, as shown in Figure 5. In this demonstration the authors made temporary changes in the optimal combustion parameters: the air supply in the incinerator grate is reduced while the air supply in the post-combustion chamber is increased, resulting in unstable combustion conditions. In Figure 5, the post-combustion malfunction was due to accidentally extinguishing, in part, the post-combustion process. Note the time differential in the observation of the anthracene/phenanthrene signal, which differs considerably from the other species.

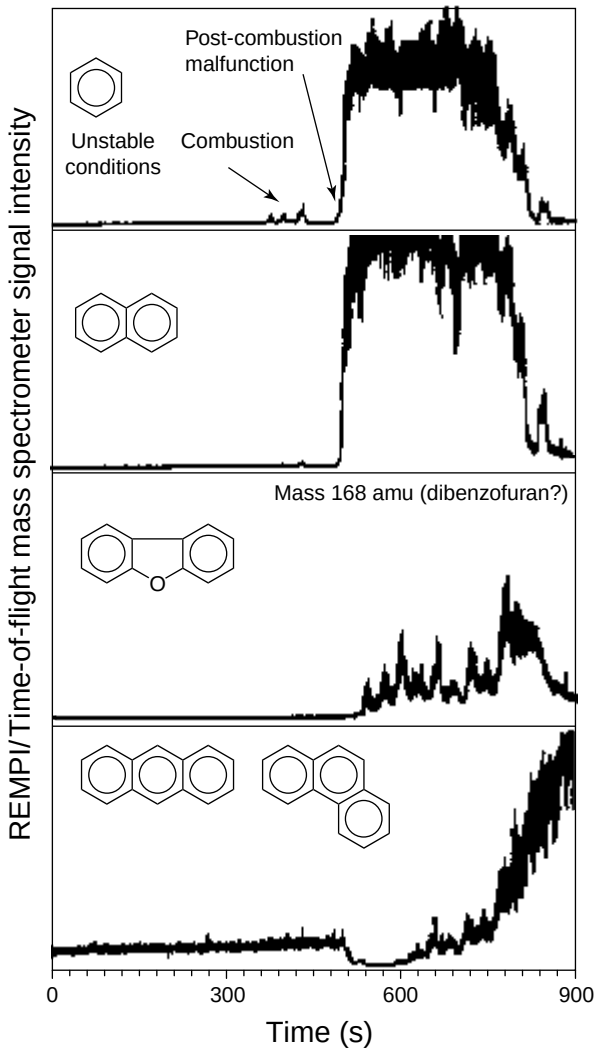

Figure 5 The REMPI/time-of-flight mass spectrometer-generated time profiles of benzene, naphthalene, dibenzofuran, and anthracene/phenanthrene during a change of process parameters in a technical incinerator. (Reproduced by permission of John Wiley & Sons, Ltd., from Zimmermann et al.[26])

4.3 Process Monitoring in the Petrochemical Industry

Process monitoring is used extensively in the petrochemical industry to determine the composition and flow of feed materials and final products. Two typical examples are the production of ethylene oxide and polyethylene.

4.3.1 Ethylene Oxide Production

In the production of ethylene oxide the mass spectrometer will typically determine 10 components simultaneously.[27] These components include feed gases and reaction products from catalytic production. Component concentrations range from parts-per-billion to percentages. Control of this reaction is usually optimized close to its explosive threshold. Monitoring of chlorinated hydrocarbons is important because they can easily poison the reaction. In this system a number of molecules are monitored that contain major ions with isobaric interferences, e.g. ethylene oxide and carbon dioxide at mass 44. Nevertheless, the analytes are easily deconvoluted using minor ion peaks that have no isobaric interferences, e.g. mass 43 for ethylene oxide.

4.3.2 Polyethylene Production

In polyethylene production, light hydrocarbon feed and by-products are monitored with concentrations ranging from a few parts-per-million to percentages.[28] The volatile hydrocarbons are sampled after refluxing in order to eliminate the heavy components. Astle[28] demonstrated that the use of mass spectrometry provided a tighter control of the ethylene furnace when compared to gas chromatography. This was mainly due to the shorter response time for the mass spectrometer: typically 3 s, compared to 7 min for the gas chromatograph. This resulted in a process variance that was 50% lower than with the gas chromatograph, allowing for a 2% increase in the set-point for the furnace, with a commensurate increase of 5–10% in yield.

4.4 Process Monitoring in the Semiconductor Industry

Process monitoring provides the semiconductor industry with a means to detect more effectively the flaws in their process, reduce the amount of scrap material production, and determine more effectively the composition of the end product. All of these factors are critical to cost-effective production. Several examples have been selected from the literature, demonstrating the application of mass spectrometry to semiconductor manufacturing in vapor deposition, ion implantation, plasma-enhanced vapor deposition, and molecular beam epitaxy processes.

4.4.1 Vapor Deposition

Water, nitrogen, oxygen and hydrocarbons cause undesirable effects in the metal grain, reflectivity, conductivity, particles, and adhesion of metals such as aluminum, titanium, and titanium nitride during vapor deposition on wafers. The presence of these gases is indicative of gas leaks in the process chamber and improper removal of the photoresist prior to vapor deposition. Markle et al.[29] have shown the applicability of a micropole quadrupole mass spectrometer to the monitoring of these gases in a physical vapor deposition chamber. Three analyzers were placed on the ''clean chamber'' and two of the sputtering chambers of this device. Wafers that enter these chambers can be monitored for these contaminants, to partial pressures of 10^{-9} mbar. The presence of the contaminants can result in a fault indication that would indicate the need for further removal of photoresist from wafers and/or proper sealing of the chambers.

4.4.2 Ion Implantation

Schneider et al.[30] have shown the ability to determine the presence of normal wafers, wafers with insufficiently cured photoresist, and wafers with no photoresist, during As^+ ion implantation processing using a micropole quadrupole mass spectrometer. Ions at m/z 2, 28 and 44 were used to monitor the process. Release of H_2, C_2H_4, and CO_2 from the interaction of the ion beam with improperly cured photoresist results in a partial pressure increase of these species that is 1–2 orders of magnitude above what is observed with a normally cured wafer. Furthermore, the lack of an initial burst of CO_2 from the wafer at the start of the process is indicative of a wafer without photoresist that has been mistakenly introduced into the system.

4.4.3 Plasma-enhanced Chemical Vapor Deposition

The process resulting in the deposition of nitrogen-rich layers of silicon onto wafer surfaces can be monitored and controlled using mass spectrometry during plasma-enhanced chemical vapor deposition (PECVD). In this technique a mixture of silane and ammonia gases is reacted in an RF plasma. Greve et al.[31] have shown the ability to monitor reactive species in a PECVD system. Because the RF plasma operates at approximately 1 mbar pressure, a 0.25-mm reducing aperture with a differentially pumped chamber is used between the process chamber and the quadrupole mass spectrometer. Reagent gases SiH_4 and NH_3, along with the plasma power, are controlled as a result of the species being monitored. Nitrogen-rich silicon films are deposited when the conditions are selected such that the mass spectrum is rich in $Si(NH_2)_3$ species. The conditions are also selected

such that SiH_x and Si_2H_6 species are minimized. Under conditions of high ammonia flow and/or high power, the triaminosilane species become predominant. As has been shown by Smith et al.,[32] these conditions lead to the deposition of high-quality, nitrogen-rich films because of the excess nitrogen in the triaminosilane molecule over other silicon species. Thus, a multivariate closed-loop control system can be set up to control the reagent gases and plasma conditions in order to increase the formation of the intermediate species of interest. The process works best when the signal for the intermediate species is abundant and low in noise.

4.4.4 Molecular Beam Epitaxy

Semiconductor capabilities realize much of their advantages due, in part, to the heterojunctions formed between various compositions of compounds formed with row III and V elements and II and VI elements. Precise control of the composition of layers can result in well-characterized epitaxial growth of fixed ternary and quaternary layers, or graded transition layers. Celii et al.[33] have shown that the use of in situ quadrupole mass spectrometers improves the control of epitaxial layers in semiconductors by open-loop calibration with short feedback, or by real-time, closed-loop composition control, during vapor deposition in molecular beam epitaxy. In this work the composition of $In_xGa_{1-x}As$ was nearly lattice matched to the InP, and determined with an uncertainty of $\Delta x \sim 0.005$ in real-time. For composition monitoring, the In^+, Ga^+, and As^+ ion signals were monitored individually. It was found

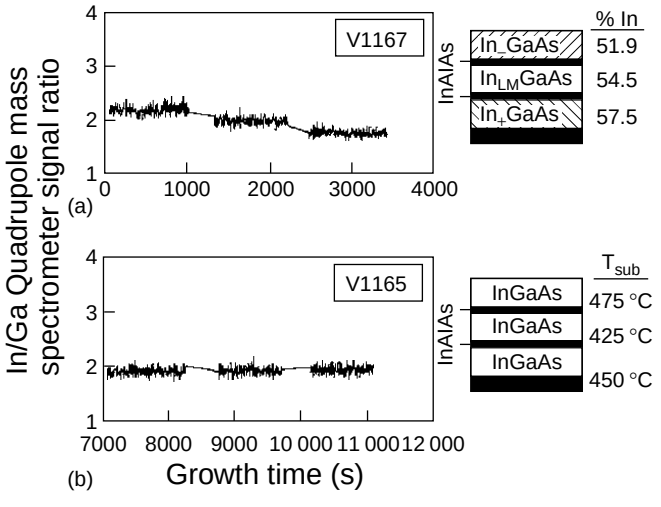

Figure 6 The InGaAs test structures and corresponding In^+/Ga^+ intensity ratios from a quadrupole mass spectrometer following a molecular beam epitaxy process: (a) change of Ga effusion cell temperature; (b) change of substrate surface temperature. (Reproduced by permission of the American Vacuum Society, from Celii et al.[33])

that the In^+/Ga^+ intensity ratio is nearly independent of the As partial pressure in the system. Furthermore, the In^+/Ga^+ intensity ratio is indicative of the In/Ga epilayer composition, as shown in Figure 6. The In^+/Ga^+ intensity ratio from the mass spectrometer and the InGaAs composition of the epilayer, as measured by X-ray diffraction, was linear from a composition of 48–58% InAs in InGaAs and a In^+/Ga^+ intensity ratio of 1.6–2.3, respectively. This linear relationship suggests that the InGaAs composition can be controlled with the use of quadrupole mass spectrometry over this range. Growth periods of approximately 1 week resulted in epilayers of 1000–2000 Å with good precision. Changing the quadrupole mass spectrometer system, opening the chamber, and the position of the Ga cell used resulted in a change in the In/Ga ratio that would require recalibration and short conditioning before proceeding.

5 QUALITY CONTROL

5.1 Interferences

Interferences in mass spectrometry can be classified as spectral and chemical in nature. Both of these have been covered already to some extent in the previous sections, but it is worth consolidating them. Spectral interferences arise from the inability to resolve two or more analytes due to ions with overlapping mass, e.g. nitrogen and carbon monoxide with molecular ions at a nominal mass of 28. Increasing the resolution of the instrument can provide some relief, but usually at the expense of sensitivity and instrument complexity. Fortunately, most molecules will fragment to some extent under electron bombardment. This fragmentation pattern provides a unique fingerprint for the analyte of interest, and may provide additional peaks in the spectrum that can be used to deconvolute the contributions from each component of an analyte mixture.[16] Conversely, if the analyte mixture is very complex, spectral interferences tend to arise because there are too many overlapping fragment ions and deconvolution becomes complex and inaccurate.

Chemical interferences arise from chemical contamination of the sampling system that cannot be resolved spectroscopically. These can also be due to contamination from degassing of system components, back-streaming of vacuum pump oils, and leaks. If unchecked, chemical interferences will introduce a bias and, often, instability in the measurement. As with most vacuum applications, cleanliness in assembling the system and in its operation is paramount. This may be difficult to maintain in a process application, and is often the ''Achilles' heel'' of process mass spectrometers. Nevertheless, with careful

control of temperature, passivation of surfaces, trapping of pumps, and filtering of process gases, adverse effects can be minimized. Finally, the choice of ion source can be used to minimize chemical interferences. For example, an enclosed electron impact ion source tends to be less prone to contamination by the system's vacuum components, but tends to have lower sensitivity. Atmospheric pressure ion sources have the least likelihood of contamination from vacuum components, but suffer from wide responses in sensitivity for analytes, depending on the matrix of the sample.[34]

5.2 Analysis of Complex Mixtures

Several methods may be used for the simplification of complex mixtures of analytes.[35,36] One method is to simplify the spectra to molecular ions by using different ionization techniques. Field ionization, field desorption, and chemical ionization will produce molecular ion spectra with little, if any, fragmentation and rearrangement ions. These techniques complicate on-line instrumentation and are not readily implemented. Instruments with API, as discussed above, will provide molecular ion spectra with almost no fragmentation. Separating the mixtures using chromatographic techniques followed by mass spectrometric identification, when needed, is probably the most powerful method for mixture analysis. A technique that can be implemented easily with on-line process mass analyzers is the low-voltage technique. Furthermore, limiting the analysis to the determination of groups of similar compounds can be the simplest method. This is referred to as compound-type analysis.

5.2.1 Low-voltage Technique

The approach arises from the fact that there is a region of electron bombardment energy (8–10 eV) that results in the formation of molecular ions with little, if any, fragmentation of the molecule for most hydrocarbons. This is also associated with 5–10 times lower sensitivity, but the spectrum is simplified and can be deconvoluted with some knowledge of the components in the analyte mixture. The main limitation to this technique is the resolution of analyte components with the same nominal molecular mass or isotopic interferences. Not all components can be detected using this technique. For example, although aromatic and heteroaromatic analytes are amenable, paraffins, naphthalenes, alkysulfides, and some other compounds are not. Other disadvantages include high susceptibility to small changes in the electron emission energy, and the method is not perfect, i.e. some fragments or rearrangement products may persist that can be mistaken for molecular ions.

5.2.2 Compound-type Analysis

This approach is commonly used for the analysis of complex mixtures of hydrocarbons. It takes advantage of the fact that hydrocarbon spectra are very similar within a homologous series. For example, as can be seen in Table 2, paraffins have a number of similar fragment ions, whereas their molecular ions tend to differ by 14 Da (CH_2) within a continuous series. Thus, the intensities of fragment ions at masses 43, 57, 71, and 85 can be summed to quantitate the total paraffins in a sample. Fragment ions at 41, 55, 69, and 83 are representative of cycloparaffins; 53, 67 and 81 for dinaphthenes, dienes and acetylenes; and 77, 78, 91, 92, 105, and 106 are used for alkylbenzenes. Calibration requires that the gases in the calibration mixture be quite close in composition to the sample. Calibration gases are often prepared from well-separated fractions of petroleum distillates. This form of quantitation can be accurate and precise to within a few percent. The reproducibility is maintained by keeping the ion source conditions such that the ratio of the molecular ion and fragment ions in a pure hydrocarbon gas remains

Table 2 Mass spectra of paraffins, showing the similarities in the ion fragment abundance[43]

Formula	Name	Ion abundance at indicated mass																
		15	27	28	29	39	41	42	43	44	56	57	58	70	71	72	85	86
C_4H_{10}	*n*-Butane		242	278	376	123	330	161	1000	35		37	270					
C_5H_{12}	*n*-Pentane	48	345	65	243	139	396	579	1000			123				87		
C_6H_{14}	*n*-Hexane		454	107	606	197	701	409	812		453	1000						155
C_7H_{16}	*n*-Heptane		393		460	188	519	241	1000		263	478		172	440			
C_8H_{18}	*n*-Octane		292		345	135	381	156	1000		180	342			235		295	
		15	26	27	28	29	39	40	41	42	43	53	55	56	57	69	70	84
C_4H_8	1-Butene		82	251	270	125	346	64	1000			54	181	388				
C_5H_{10}	1-Pentene	58	56	321		271	346	86	443	1000			601				309	
C_6H_{12}	1-Hexene			684		279			1000	751	597		594	860		187		282
C_7H_{14}	1-Heptene			511		706	454		1000	527			598	873	264	267	368	
C_8H_{16}	1-Octene			573		581	457		908	641	1000		762	658		305	612	

constant during the analysis. The disadvantages of this technique include the need for prior knowledge of the sample mixture to be analyzed, and it is restricted to samples with similar origin, chemical treatment, and boiling range.

5.3 Principles of Quantitative Analysis

Because the partial pressures of the components in a mixture are directly proportional to their concentrations and to the molecular and fragment ion composition of the spectrum, the intensities of the analyte ions can be used to determine quantitatively the composition of the mixture. The mass spectrum of each of the components in the mixture must be known within the instrumental parameters. The molecular and fragment ions to be used for analysis are selected from the normalized spectra. Normalized intensities are used because absolute intensities can vary widely with instrumental parameters. The ions selected should have high abundance in the analyte spectrum that they represent and should be absent or have weak intensities in the spectrum of other analytes. The sensitivity for each of the components in the sample mixture is determined with calibration gases. A matrix of n analyte components and $n + i$ ions (where i is equal to or greater than the number of overlapping ions in the selection) is formed.[35] Matrix algebra can be used to solve the unknown concentrations in the matrix. Thus, for a two-component mixture with two ions in each mixture Equations (1) and (2) need to be solved:

$$zA_1 + yA_2 = M_1 \tag{1}$$

$$zB_2 + yB_1 = M_2 \tag{2}$$

where z and y are the weight fraction of the two components in the mixture, A_1, A_2, B_1, and B_2 are the sensitivities for each ion in the component spectra, and M_1 and M_2 are the intensities of each of the ions in the mixture spectrum. The matrix expression is given in Equation (3):

$$\begin{matrix} z \\ y \end{matrix} = \begin{matrix} M_1 & A_1 & B_1 \\ M_2 & A_2 & B_2 \end{matrix} \tag{3}$$

For this simple two-component mixture the matrix solution is given in Equations (4) and (5):

$$z = \frac{M_1B_1 - M_2A_2}{A_1B_1 - B_2A_2} \tag{4}$$

$$y = \frac{M_2A_1 - M_1B_2}{A_1B_1 - B_2A_2} \tag{5}$$

The process can be extended to 30–50 components.[35] Least-squares analysis is the most common quantitation approach. The concentrations are those values that, when multiplied by the corresponding reference spectra and relative sensitivities and then summed for each ion, result in the smallest sum-squares difference between the calculated and the experimentally measured mixture. Other multivariate quantitation approaches include partial least squares, principal component analysis, and chemometrics.[37,38]

6 COMPARISON WITH OTHER SPECTROSCOPIC METHODS

6.1 Comparison of Technology

A number of other analytical instrumentation techniques are available for process control and analysis that can be used to complement mass spectrometry and in some cases rival its capabilities. These techniques include, but are not limited to, gas chromatography, ion mobility spectrometry, infrared spectroscopy, and chemical sensors. In-depth understanding of these techniques can be found elsewhere in this publication. Other reviews of the literature include that by Walsh and LaPack,[39] who review mass spectrometry principles with a good comparison to gas chromatography and infrared spectroscopy, an extensive review of infrared spectroscopy in process control by Workman,[40] ion mobility spectrometry as it pertains to field applications,[41] and advances in chemical sensors.[42] Although these techniques and others are quite suitable for the monitoring of production processes, much like mass spectrometry they all have strengths and weaknesses. The choice of which technique to use will likely be based on the amenability of the process to be sampled, the analytes to be detected under the required conditions, the specificity required, and the ability of the instrument to stand the physical environment to be monitored. When compared to the above-mentioned techniques, mass spectrometry offers a number of advantages. These include the ability to sample and monitor processes in a vacuum, a high degree of specificity, high sensitivity, a large number of analyte components that can be monitored simultaneously, a fast response time, and very low detection limits. However, other methodologies are worth considering when looking for a process analyzer.

Infrared and other spectroscopic techniques offer good specificity for analyte components and can be used for trace analysis of contaminants in a gaseous process. Detection limits for trace components in gases are still limited by the observation path length, to at best parts-per-million levels. This method is easily interfaced onto process streams, with the aid of an appropriate window. Because it is completely nondestructive (unlike mass spectrometry and other techniques), the sample stream can be left undisturbed. Thus, the technique

is easily adapted to process lines, especially lines that are at ambient and high pressure. Infrared spectroscopy requires a high degree of chemometrics for spectral deconvolution of mixtures, or a preseparation step. Finally, infrared process instruments are highly reliable and require very little maintenance.

Gas chromatography offers an affordable solution to monitor slowly changing process streams. The method relies on the physical separation of analyte components, based on their boiling point and affinity toward a stationary phase. Detection limits can be quite low for specific target analytes with the right detector, in the order of nanogram to picogram levels. The technique is easily implemented into gaseous and volatile liquid streams at ambient and high pressure. The method is destructive and requires supplies of carrier and detector gases to be available. The gas chromatography columns and injector port can be fouled up easily and require regular maintenance. The method has little or no specificity because it relies on time separation to identify components and a detector-type response for classes of compounds.

Ion mobility spectrometry is slowly emerging as an alternative detection method to mass spectrometry. The methods are related in that they both rely on ion formation, separation and detection. But, unlike mass spectrometry, ion mobility spectrometry separates components according to their drift times in an electrical field filled with a bath gas. The charge, size, and collision cross-section of the molecule with the bath gas determine the time separation of the analyte. Separations happen in the millisecond time-frame, therefore this method has a fast response time. The resolution of the method is quite poor compared to mass spectrometry, therefore separation of complex mixtures is quite problematic. Nevertheless, ion mobility spectrometry can provide some degree of specificity, with very low detection limits for some analytes. Most typical ion mobility spectrometry modules use electron capture ionization, which allows the technique to be quite sensitive (at ppb levels) for molecules with high electron affinity, e.g. oxygenated compounds, phosphorus and sulfur compounds, and halogen-bearing compounds. The instrument is small, requires no vacuum pumps, and can be applied easily to sample ambient pressure and high-pressure processes.

Chemical sensors or ''chemical noses'' are also emerging as real alternatives to the analysis of volatile streams. These types of detectors are characterized by solid-state electrochemical or transducer-type components that can be integrated easily into small and very inexpensive modules. Each individual module is usually not very specific or has a wide response for a class of analytes. However, with the combination of a number of modules that have different response characteristics, the chemical sensor can be trained to have a specific response to particular analytes. Detection limits tend to be in the parts-per-million range. The method can be applied easily to ambient pressure and high-pressure processes.

6.2 Business Issues

Of the techniques mentioned above, mass spectrometry is probably the most expensive to implement, due to the cost and complexity of the instrumentation. Nevertheless, mass spectrometers are decreasing in price. Residual gas analyzers can be purchased for $2000–5000, but without the vacuum pumps and other essential hardware necessary for operation. A fully operational analyzer with pumps and computer system will cost $15 000 or more. Customization of the analyzer to the particular application will generally determine the majority of the cost for the installation. Installation of these instruments on the process line will require some control of the surrounding environment, because vacuum pumps must be cooled, and vibration and corrosiveness must be minimized. Nevertheless, mass spectrometers can often provide ultimate sensitivity, exceptional specificity, exceptional detection, and unambiguous identification of unknowns which are often unsurpassed by other analytical instrumentation techniques.

ACKNOWLEDGMENTS

The author appreciates the help and support, in gathering figures and information relevant to this topic, from Peter Scoch and Rick Van Vorous of Pfeiffer Vacuum Technology. Support from the Inorganic Trace Analysis Group (CST-9) at Los Alamos National Laboratory during the preparation of this article is greatly appreciated. Los Alamos National Laboratory is operated by the University of California for the US Department of Energy under contract W-7406-ENG-36.

ABBREVIATIONS AND ACRONYMS

API	Atmospheric Pressure Ionization
LOD	Limit of Detection
MIMS	Membrane Inlet Mass Spectrometry
MW	Molecular Weight
PAHs	Polyaromatic Hydrocarbons
PECVD	Plasma-enhanced Chemical Vapor Deposition
POAA	Phenoxyacetic Acid
REMPI	Resonance-enhanced Multiphoton Ionization
RF	Radiofrequency

RELATED ARTICLES

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis • Chromatography in Process Analysis • Flow and Sequential Injection Analysis Techniques in Process Analysis • Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis • Raman Spectroscopy in Process Analysis • Sampling and Sample Preparation in Process Analysis • Titration Techniques for Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Mass Spectrometry **(Volume 13)**
Artificial Intelligence and Expert Systems in Mass Spectrometry • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

FURTHER READING

For further reading see references 1, 2, 4, 5, 20, 21, 25, 26, 28, 29, 35, 37 and 39.

REFERENCES

1. H.C. Hill, 'Introduction to Mass Spectrometry', in *Spectroscopy in Education*, ed. T. Calrus, Heyden & Son Limited, London, Vol. 3, 1966.
2. J. Roboz, *Introduction to Mass Spectrometry, Instrumentation and Techniques*, Kriegler Publishing Co., Huntington, NY, 1979.
3. J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott–Raven Publishers, Philadelphia, PA, 1997.
4. W.K. Huber, 'Partial Pressure Measurements in High and Ultra-high Vacuum Systems, Part I', *Vacuum*, **13**(10), 399–412 (1963).
5. K.D. Cook, K.H. Bennet, M.L. Haddix, 'On-line Mass Spectrometry: A Faster Route to Process Monitoring and Control', *Ind. Eng. Chem. Res.*, **38**(4), 1192–1204 (1999).
6. W. Bleakney, J.A. Hipple, Jr, 'A New Mass Spectrometer with Improved Focusing Properties', *Phys. Rev.*, **53**(7), 521–529 (1938).
7. Ch.F. Robinson, L.G. Hall, 'Small General Purpose Cycloidal-focusing Mass Spectrometer', *Rev. Sci. Instrum.*, **27**(7), 504–508 (1956).
8. E.W. Blauth, *Dynamic Mass Spectrometers*, Elsevier Publishing Co., New York, 1966.
9. M.M. Wolff, W.E. Stephens, 'Pulsed Mass Spectrometer with Time Dispersion', *Rev. Sci. Instrum.*, **24**(18), 616–617 (1953).
10. D.C. Damoth, 'Recent Advances in Time-of-flight Mass Spectrometry', *Adv. Anal. Chem. Inst.*, **4**, 371–410 (1965).
11. E.S. Borovik, S.F. Grishin, 'Determination of the Composition of Residual Gases in Vacuum Obtained by Condensation Pumps', *Zh. Tekh. Fiz.*, **29**(9), 1010–1016 (1959).
12. W. Paul, H. Steinwedel, 'A New Mass Spectrometer without Magnetic Fields', *Z. Naturforsch.*, **8a**, 448–450 (1953).
13. K.G. Guenther, 'A Partial Pressure Vacuum Gauge Working According to the Principle of the Electrical Mass Filter', *Vacuum*, **10**(4), 293–309 (1960).
14. J.R.P. Clarke, 'A Review of On-line Analysis', *Anal. Chim. Acta*, **190**(1), 1–11 (1986).
15. T. Kotiaho, 'On-site Environmental and In Situ Process Analysis by Mass Spectrometry', *J. Mass Spectrom.*, **31**(1), 1–15 (1996).
16. R.D. Craig, E.H. Harden, 'The Interpretation of Mass Spectra in Vacuum', *Vacuum*, **16**(2), 67–73 (1966).
17. P.F. Knewstubb, *Mass Spectrometry and Ion–Molecule Reactions*, Cambridge University Press, New York, 89, 1969.
18. J. Roboz, *Introduction to Mass Spectrometry, Instrumentation and Techniques*, Kriegler Publishing Co., Huntington, NY, 231–235, 1979.
19. R.W. Cheek, J.A. Kelber, J.G. Flemming, R.S. Blewer, R.D. Lujan, 'In-situ Monitoring of the Products from the $SiH_4 + WF_6$ Tungsten Chemical Vapor Deposition Process by Microvolume Mass Spectrometry', *J. Electrochem. Soc.*, **140**(12), 3588–3590 (1993).
20. T. Kotiaho, F.R. Lauritsen, T.K. Choudhury, R.G. Cooks, G.T. Tsao, 'Membrane Introduction Mass Spectrometry', *Anal. Chem.*, **63**(18), A875–A883 (1991).
21. S.J. Bauer, R.G. Cooks, 'MIMS for Trace-level Determination of Organic Analytes in On-line Monitoring and Environmental Analysis', *Am. Lab*, **25**(16), 36–51 (1993).
22. L.E. Kirsch, L. Nguyen, C.S. Moeckly, 'Pharmaceutical Container Closure Integrity. 1. Mass Spectrometry-based Helium Leak Rate Detection for Rubber-stoppered Glass Vials', *PDA J. Pharm. Sci. Technol.*, **51**(5), 187–194 (1997).

23. S. Bohatka, 'Process Monitoring in Fermentors and Living Plants by Membrane Inlet Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 656–661 (1997).
24. K.F. Hansen, F.R. Lauritsen, H. Degn, 'An On-line Sampling System for Fermentation Monitoring Using Membrane Inlet Mass Spectrometry (MIMS): Application to Phenoxyacetic Acid Monitoring in Penicillin Fermentation', *Biotechnol. Bioeng.*, **44**(3), 347–353 (1994).
25. R.C. Johnson, N. Srinivasan, R.G. Cooks, D. Schell, 'Membrane Introduction Mass Spectrometry in a Pilot Plant: On-line Monitoring of Fermentation Broths', *Rapid Commun. Mass Spectrom.*, **11**(4), 363–367 (1997).
26. R. Zimmermann, H.J. Heger, A. Kettrup, U. Boesl, 'A Mobile Resonance-enhanced Multiphoton Ionization Time-of-flight Mass Spectrometry Device for On-line Analysis of Aromatic Pollutants in Waste Incinerator Flue Gases: First Results', *Rapid Commun. Mass Spectrom.*, **11**(10), 1095–1102 (1997).
27. ABB-Extrel, 'Real Time Analysis of Low Level Chloride Components in the Production of Ethylene Oxide', ABB-Extrel Application Note.
28. G.C. Astle, 'Mass Spectrometry for Use in Ethylene Furnace Monitoring and Control', *9th Ethylene Producers Conference*, American Institute of Chemical Engineers, New York, 325–342, Vol. 6, 1997.
29. R.J. Markel, S. Uhlig, G. Magsamen, J. Fidler, R.J. Ferran, 'Volatile Contamination Monitoring for Titanium and Titanium Nitride Sputter Systems', *Semicond. Fabtech.*, **5**(1), 239–243 (1996).
30. T.P. Schneider, P. Krocak, B. Van Eck, 'Real-time In Situ Residual Gas Monitoring During Ion Implantation in High Volume Semiconductor Manufacturing', *Future Fab. Int.*, **1**(4), 237–239 (1998).
31. D.W. Greve, T.J. Knight, X. Cheng, B.H. Krogh, M.A. Gibson, J. LaBrosse, 'Process Control Based on Quadrupole Mass Spectrometry', *J. Vac. Sci. Technol. B*, **14**(1), 489–493 (1996).
32. D.L. Smith, A.S. Alimonda, D.C. Chen, S.E. Ready, B. Wacker, 'Mechanism of Silicon Nitride Hydride (SiN_xH_y), *J. Electrochem. Soc.*, **137**(2), 614–623 (1990).
33. F.G. Celii, Y.-C. Kao, H.-Y. Liu, 'In Situ Composition Monitoring of InGaAs/InP Using Quadrupole Mass Spectrometry', *J. Vac. Sci. Technol. A*, **14**(6), 3202–3207 (1996).
34. A. Hansel, A. Jordan, R. Holzinger, P. Prazeller, W. Vogel, W. Lindinger, 'Proton-transfer Reaction Mass Spectrometry: On-line Trace Gas Analysis at the ppb Level', *Int. J. Mass Spectrom. Ion Processes*, **150**, 609–619 (1995).
35. T. Aczel, 'Applications of Mass Spectrometry in the Petrochemical Industry', *Practical Mass Spectrometry*, ed. B.S. Middleditch, Plenum Press, New York, 289–325, 1979.
36. A. Quayle, 'Industrial Applications of Mass Spectrometry', *Recent Topics in Mass Spectrometry*, ed. R.I. Reed, Gordon and Breach Science Publishers, New York, 267–300 (1971).
37. E.V. Thomas, 'A Primer on Multivariate Calibration', *Anal. Chem.*, **66**(15), 795A–804A (1994).
38. M.H. Kaspar, W.H. Ray, 'Chemometric Methods for Process Monitoring and High-performance Controller Design', *AIChE J.*, **38**(10), 1593–1608 (1992).
39. M.R. Walsh, M.A. LaPack, 'On-line Measurements Using Mass Spectrometry', *ISA Trans.*, **34**(1), 67–85 (1995).
40. J. Workman, Jr, 'Review of Process and Non-invasive Near-infrared and Infrared Spectroscopy: 1993–1999', *Appl. Spectrosc. Rev.*, **34**(1/2), 1–89 (1999).
41. H.H. Hill, G. Simpson, 'Capabilities and Limitations of Ion Mobility Spectrometry for Field Screening Applications Field Analytical Chemistry and Technology', *Field Anal. Chem. Technol.*, **1**(3), 119–134 (1997).
42. J. Janata, M. Josowicz, P. Vanysek, D.M. Devaney, 'Chemical sensors', *Anal. Chem.*, **70**(12), R179–R208 (1998).
43. A. Cornu, R. Massot, *Compilation of Mass Spectral Data*, 2nd edition, Heyden, Philadelphia, Vol. 1, 1979.

Near-infrared Spectroscopy in Process Analysis

Don S. Goldman
Optical Solutions, Inc., Folsom, CA, USA

1 Introduction 1
2 Near Infrared Spectra 2
3 Data Analysis 3
3.1 Partial Least Squares 3
3.2 Multiple Linear Regression 4
3.3 Calibration Transfer between Instruments 4
4 Fiber Optic Instrumentation and Requirements 4
4.1 Scanning Spectrophotometers 4
4.2 Fourier Transform Spectrophotometers 5
4.3 Acousto-optic Tunable Filters 5
4.4 Photodiode Array Spectrophotometers 5
4.5 Fixed Filter Photometers 5
4.6 Fiber Optic Multiplexing 6
4.7 Enclosures for the Process Environment 6
5 Fiber Optics and Fiber Optic Probes 6
5.1 Transmission Probes 7
5.2 Reflectance Probes 7
5.3 Internal Reflectance Probes 7
5.4 Fiber Optics 8
6 Organizational Issues 8
List of Symbols 9
Abbreviations and Acronyms 9
Related Articles 9
References 9

Near infrared (NIR) is a spectroscopic method based upon the absorption of light in the wavelength region between 700 nm and 2500 nm due to vibrations of molecular functional groups in the sample. These spectral features arise from overtones and combinations of the fundamental molecular vibrations of these groups that occur in the infrared (IR) at longer wavelengths. NIR is rich with features arising from vibrations of C−H, O−H and N−H.[1]

NIR spectra are acquired by determining the absorbance, A, *of light over a continuous range of wavelengths in spectrophotometers or at several specific wavelengths in fixed-filter photometers. Absorbance is defined by Equation (1):*

$$A = -\log\left(\frac{I}{I_o}\right) \quad (1)$$

where I *is the signal from the sample relative to the signal from a reference,* I_o. *The reference can be an empty transmission sample cell for liquids, or it can be a broadband spectral reflector, such as a white ceramic, for reflectance measurements of solids.*

Absorbance is usually linearly related to concentration, following Beer's Law (Equation 2):

$$A = \varepsilon C l \quad (2)$$

where ε *represents the molar extinction coefficient and* C *is the concentration of the species, and* l *is the optical pathlength of light through the sample. For a particular species and fixed experimental set-up,* ε *and* l *are constants. Letting a constant,* $k = (\varepsilon l)^{-1}$, *then (Equation 3):*

$$C = kA \quad (3)$$

A principal advantage of process NIR over IR is the ability to use fiber optics to transmit light between the instrument and the sample over considerable distance. This permits the instrument, a potential source of ignition, to be placed in safe areas where no flammable vapors exist. Only the fiber optic probe in contact with the sample and the optical fibers that connect to the instrument need be placed in the hazardous location. A disadvantage of process NIR is the characteristics of the spectra, which are typically comprised of broad, overlapping peaks in comparison to IR spectra. This often requires the use of sophisticated data analysis methods that place additional demands upon vendors to provide stable NIR instruments and upon manufacturers to have highly-trained staff, often at the PhD level, for calibration model development, implementation and maintenance.

1 INTRODUCTION

NIR has been traditionally viewed by spectroscopists trained in laboratory methods as a less-than-useful spectral region that exists to fill the gap between the ultraviolet/visible (UV/VIS) and IR regions. NIR peaks are broad, overlapped and much weaker than their IR counterparts. So, why has NIR become so widespread in process analysis?

Interest in NIR grew quickly in the 1980s due to the advent of fiber optics, bright light sources and sensitive detectors. NIR also solved some of the difficult

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

sampling techniques characteristic of IR, because NIR optical paths through liquid samples may be millimeters or even centimeters, rather than microns. As a result, first commercial fiber optic spectrophotometers were introduced in the 1980s.

The importance of these developments lies in the ability to bring the sampling interface, a fiber optic probe, safely to the process while placing the electronics (ignition sources) remotely, away from hazardous environments. This also meant that the traditional, time-consuming method of collecting (''grab'') samples and bringing them back to the laboratory for analysis could be replaced by continuous ''on-line'' measurements without exposing workers to potentially dangerous chemicals. Continuous on-line measurements, in turn, mean more responsive process control for better production yield, less rework of off-target material, improved product quality and enhanced worker safety. It began to change the paradigm by measuring chemistry directly, rather than inferring it from temperature, pressure, flow and fluid level.

For NIR to be accepted, it had to provide reliable instrumentation requiring minimal maintenance. Otherwise, there would be little advantage over the much more widespread, but more maintenance-intensive methods, such as process gas chromatography (GC). Thus, considerable advancements occurred over the next ten years or so to meet the requirement of instrument reliability.

As mentioned above, NIR peaks are generally overlapped. Gone is the isolated, narrow peak common in GC or IR analysis. As a result, mathematical methods using multivariate statistics were often required to analyze the subtle NIR spectral changes present among overlapped peaks. This field of multivariate statistical methods for chemical analysis is known as chemometrics.[2] Partial least squares (PLS) is perhaps the most widely used chemometric method applied to NIR data.[3] These methods became widespread in the 1980s as computers improved.

Furthermore, as smaller NIR spectral changes were analyzed in regions with overlapping peaks using chemometrics, greater requirements were placed on the wavelength stability of the instrument. Changes of less than 0.050 AU are typically analyzed in PLS models to predict chemical variables. A small shift in a NIR peak along the wavelength scale can produce such changes. Therefore, wavelength drift from the instrument could impair a PLS calibration model. Thus, considerable advancements in wavelength stability were made after the introduction of the first scanning fiber optic NIR instruments. This period also saw the introduction of newer technologies based upon different optical techniques, such as Fourier transform near infrared (FT/NIR), acousto-optic tunable filter (AOTF) and photodiode array (PDA) spectrophotometers.

As can be seen from the evolution described above, considerable advancements have occurred from the introduction of the first process scanning spectrophotometers. All of these approaches assume that a complete spectrum is required to solve the analytical problem. Often, that is indeed the case, when many parameters need to be predicted from each spectrum, or when peaks appear in other regions of the NIR spectrum that should not be there (outlier detection). However, this assumption generally comes with a hefty price tag. Many of these process NIR spectrometers systems with probes and fibers cost well in excess of $100 000 (US). In addition, they require personnel with chemometrics expertise, not only in developing the calibration models, but also in maintaining them.

In many instances, however, only a few wavelengths are required to provide an acceptable calibration. Moisture is one example where fixed filter NIR photometers have been commercially available for many years. Today these photometers are designed for use with optical fibers for the same reasons stated above. The method of multiple linear regression (MLR) can be used to select a few wavelengths from a complete spectrum analysis.[4] A photometer can then be configured with optical wavelength filters to match those wavelengths at 25 to 50% of the cost of its full spectrum counterpart, without the need for high-level internal expertise. The danger is that a contaminant will appear at a wavelength not analyzed by the photometer.

The purpose of this discussion is to provide the reader with an overview of fiber optic NIR process analysis, including instrumentation, fiber optics, probes and mathematical methods.

2 NEAR INFRARED SPECTRA

Spectra of gasoline shown in Figure 1 illustrate the characteristics of NIR spectra described above. These

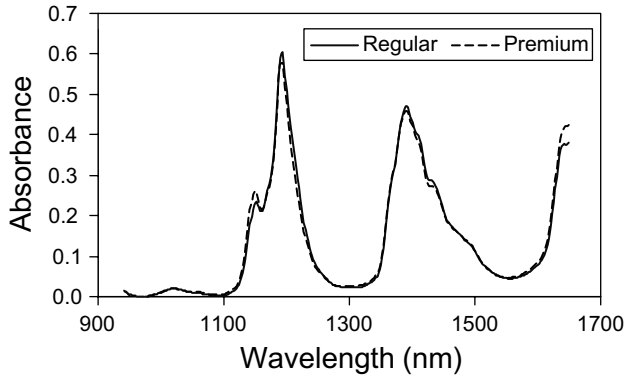

Figure 1 NIR spectra of gasoline.

spectra represent transmission of light through a pathlength of 1 cm of gasoline. C–H overtones occur in the 1100–1200 nm region. Aromatic contributions give rise to the secondary peak near 1150 nm, with premium gasoline showing greater absorbance due to its greater aromatic content. The larger peak centered near 1200 nm arises from C–H, CH_2 and CH_3 groups. The region between the 1300 in 1500 nm represents combinations from C–O–H. Note the small changes in absorbance and subtle changes in the shapes of the peaks throughout these regions. Nevertheless, PLS models of NIR spectra in this region are used to predict the octane number of gasoline on-line in real-time to better than 0.3 units at blending facilities. They can also predict many other components of gasoline at the same time that are important in refinery production, such as various additives.

NIR is commonly used to measure water and moisture due to their relatively strong absorption. For example, NIR is commonly used to measure trace amounts of moisture near 1900 nm, as shown in the spectra of tetrahydrofuran in Figure 2. These spectra show successive additions of 1000 ppm water. The 1900 nm region is unique to molecular water, whereas the region near 1450 nm can represent O–H from both water and hydroxyl. Note that contributions from water in Figure 2 near 1450 nm are present, but are much weaker than the peak at 1900 nm.

NIR reflectance measurements have been widely used for solids in food, agricultural, textile, pharmaceutical and other industries. Unlike the transmission of light through clear liquids described above, reflectance measurements are strongly influenced by particle size, uniformity and size distribution of the sample. Thus, much of the evolution of NIR reflectance has been directed towards mitigating these physical effects using larger areas of illumination, taking the average of many measurement points and using computational corrections. The NIR

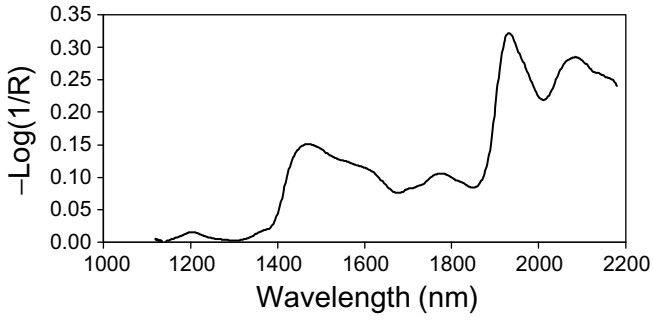

Figure 3 Reflectance spectra of wheat flour.

spectrum of wheat flour is shown in Figure 3. The vertical axis is commonly reported as $-\log(1/R)$ in reflectance spectra, rather than absorbance, where R is reflectance. In the spectrum of wheat, moisture is again analyzed in the 1900 nm region, and protein is measured in the region near 2100 nm.

3 DATA ANALYSIS

3.1 Partial Least Squares

PLS is a powerful, multivariate quantitative method that is commonly used to predict properties from NIR spectra.[2,3] Mathematically, PLS uses eigenvectors and eigenvalues to perform a decomposition of the spectral and constituent concentration data simultaneously. The decomposition process is a systematic means to determine the most important variations in the data. PLS uses constituent concentration information during spectral decomposition, which weights spectra containing higher constituent concentrations more heavily. The term ''factor'' is used to describe a linear combination of spectral data. The first factor that is determined from the

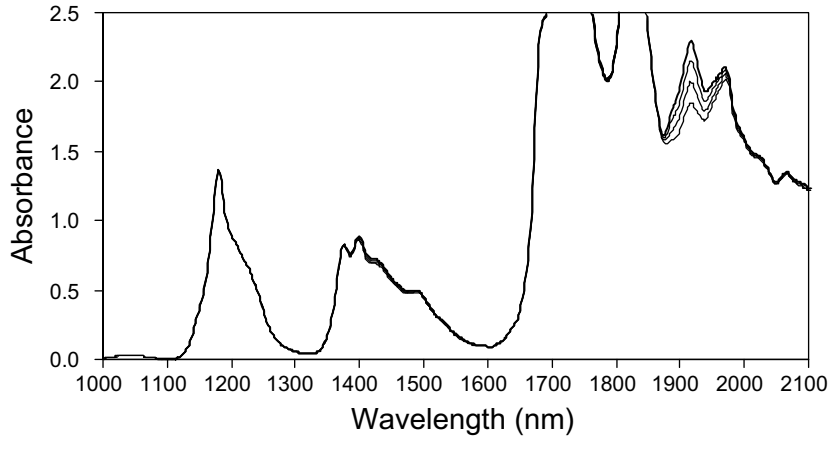

Figure 2 Moisture in tetrahydrofuran.

decomposition contains the most variation in the data. This factor is then removed, and a second factor is then determined, which contains a smaller percentage of the overall variation, and so on, all guided by the constituent concentrations. When it is said that a particular solution is obtained with a ''three factor'' model, this means that the three most important factors explain a sufficient amount of the overall variation in the system, which results in an acceptable calibration. Each factor still contains information from each wavelength in the spectrum, some weighted more heavily than others. Thus, the dimensionality of the problem has been reduced from perhaps 1000 nm of data to three factors. By eliminating less significant factors, PLS is eliminating variations in the data that do not add useful information in establishing a calibration.

PLS reconstructs a spectrum that represents the predicted constituent value. This predicted spectrum is subtracted from the actual spectrum to determine spectral residuals. Spectral residuals are important in outlier detection to ascertain whether the model prediction is believable, i.e. whether the spectrum of the unknown fits within the calibration set used to construct the model.

Disadvantages include the high level of expertise needed to understand and use PLS programs, expertise that many companies lack. Outlier detection may become too sensitive, arising from other factors, such as probe fouling, particles present in the sample and temperature variations in the sample. These problems may lead to spectroscopic anomalies not accounted for in the calibration model. On the other hand, learning of these problems will probably lead to better on-line measurements.

3.2 Multiple Linear Regression

MLR attempts to select a set of wavelengths from the calibration samples that minimizes the sum of the squares of the residuals between the predicted and actual concentrations. All possible combinations of wavelengths are quickly examined with the computational power found in most personal computers today to find the best set. Tests can be used to determine if adding additional wavelengths to the model is statistically significant.[4]

One of the tendencies in using MLR is simply to find the set of wavelengths that produce the ''best'' fit. This may not be the best strategy, because it may be strongly influenced by noise in that particular data set. User guidance is important in arriving at a robust solution. For example, restricting the wavelength search to regions in which a particular species is known to absorb may prevent the computer from finding a wavelength in a less important region, such as an absorbance valley, or in a region with an extremely small absorbance variation. A leave-one-out approach can also help to determine the robustness of an MLR calibration model. A variety of other statistical parameters can help determine the quality of an MLR model. The shape and pattern of residual plots (actual vs. predicted values) may aid in removing outliers from calibration models.

In addition to the wavelengths selected by the MLR model, more wavelengths can be added to a process photometer to help determine if a sample is an outlier. For example, if several wavelengths are required in the 2000 to 2100 nm region to measure the hydroxyl number in a particular resin, perhaps an additional wavelength can be added at 1150 or 1200 nm where the C−H peaks should be relatively constant. The presence of a contaminant elsewhere in the spectrum not analyzed by the photometer may change the absorbance of these C−H peaks outside their normal absorbance range, which could signify an outlier.

3.3 Calibration Transfer between Instruments

Many manufacturers have multiple processes at each plant site and multiple plant sites around the world. As such, they will typically use more than one instrument to measure the same chemistry, presumably using the same model. The problem is that no two instruments are identical. Perhaps the most important difference among instruments is along the wavelength scale. This could involve a small shift of all wavelengths or a change in the wavelength spacing signifying a change in wavelength resolution. As pointed out above, spectral residuals in PLS models can increase quickly even with small absorbance changes over a broad wavelength range that can result from such changes. It becomes impractical to calibrate each new instrument with all of the original calibration samples, which sometimes number in the hundreds. Therefore, calibration transfer methodologies have been developed to help install and adjust models in new instruments without having to re-run all calibration samples. Many commercial PLS software packages now include a calibration transfer module for this purpose. However, their success is apparently quite variable, and it is quite common to hear that calibration samples have been re-run on new instruments.

4 FIBER OPTIC INSTRUMENTATION AND REQUIREMENTS

4.1 Scanning Spectrophotometers

Scanning process instruments rely upon projecting light from an optical fiber or a bundle of optical fibers onto a diffraction grating. The grating disperses the light into its constituent wavelengths. It is mechanically rotated such that a narrow group of wavelengths is sequentially allowed to pass through a narrow slit. In pre-dispersed

instruments, monochromatic light passing through this slit is sent to the sample through a fiber optic cable. The light returning from the sample is then directed to an optical detector. Conversely, in post-dispersed instruments, light is sent directly to the sample and the returning light from the fiber optic cable is directed to the grating where it is dispersed, and passes through a slit placed before the optical detector. Various means are used to relate the position of the grating to the expected wavelength that corresponds to a given signal in constructing a NIR spectrum. The need for faster collection times places a significant burden upon moving the grating quickly and knowing which wavelength corresponds to a particular signal. This is particularly important because PLS models generally required sub-nanometer wavelength stability over time.

4.2 Fourier Transform Spectrophotometers

The process FT/NIR spectrophotometer is an amplitude division interferometer. It is based upon the original Michelson interferometer or one of its derivatives. In all cases, a beamsplitter is used to separate the incoming light beam into two beams, which are later recombined, focused into an optical fiber and sent to the sample.[5] One optical path in the interferometer is kept constant and the other is varied by some mechanical means. An interferogram is produced upon recombining the beams, which contains all the wavelengths in the spectra. Applying a Fourier transformation to the interferogram results in a frequency spectrum (wavenumbers, cm^{-1}), which is readily converted into a wavelength spectrum ($nm = 10^{-7}$ cm). There are a variety of ways by which one optical path is varied, including a moving mirror, moving optical wedges, and moving both reflectors in tandem. These developments focused upon issues present in the process environment in contrast to the laboratory, such as the effects of mechanical vibration and large temperature fluctuations. FT/NIR instruments have a ''multiplex'' advantage over scanning instruments, as well as greater wavelength resolution. The ''multiplex'' advantage, like the PDA discussed below, means that all wavelengths are analyzed at the same time. The greater wavelength resolution is useful in some applications, but is not particularly required for the generally broad peaks found in the NIR. For hot process samples, blackbody radiation is generally eliminated when taking the interferogram. However, FT/NIR technology has a moving part and can still be prone to vibrations in spite of the design advancements.

4.3 Acousto-optic Tunable Filters

In contrast to scanning instruments, AOTF does not require a moving part. Rather, light is directed into a crystal of TeO_2. A high-frequency acoustic wave in the radiofrequency range is coupled into the crystal by the use of a piezoelectric material bonded to the crystal. These acoustic waves quickly propagate through the crystal, interact with the broadband light and generate two monochromatic beams of light, each polarized in a different direction. In essence, the crystal is being made to act as a transmission grating controlled by changes in atomic spacing due to acoustic wave propagation. Either of these monochromatic beams can be used as a source of NIR light, coupled into an optical fiber and sent to the sample. Wavelength switching is very fast compared with scanning instruments, and the AOTF can be programmed to select any sequence of wavelengths. Hence, more time can be spent on specific peaks or in regions of interest, and less time need be spent in spectral regions containing little or no useful information.

4.4 Photodiode Array Spectrophotometers

The PDA is perhaps the most intuitively obvious way of acquiring NIR spectra. A compact, linear array of detectors is placed at an appropriate distance from a diffraction grating to analyze the complete sequence of wavelengths. Thus, light in a narrow spectral region continuously impinges on its own detector element is the array. The integration time, akin to the exposure time on a camera, is easily adjusted to optimize signal-to-noise. PDA has been the latest technology to develop due to the more recent availability of NIR detector material. Although silicon PDAs have been commercially available for some time, they can only be used below about 1000 nm. This region is comprised of even weaker, more overlapped vibrational peaks from higher order overtones than those shown in the previous section. Weaker NIR peaks, in turn, require longer optical paths for liquids and are not particularly practical for reflectance from solids other than for color. The more recent availability of indium–gallium–arsenide (InGaAs) detector material makes the PDA more useful for NIR analysis. InGaAs PDAs currently cover the range from 900–2200 nm. The PDA detector is sealed in an optical enclosure and cooled thermo-electrically. Each photodiode detector in the array is multiplexed and amplified electronically. As with AOTF, PDA is extremely fast, and is able to collect and average a large number of complete spectra in a fraction of a second without moving parts. Another advantage over lead sulfide (PbS) detectors traditionally used in the NIR is the superb low-light sensitivity of InGaAs for highly absorbing liquids or reflectance from solids. Their disadvantage at present is the high price of the detector, although they have decreased by nearly 50% since 1990.

4.5 Fixed Filter Photometers

Traditionally, fixed filter photometers capable of analyzing multiple NIR wavelengths worked by passing light

through a sample cell within the instrument. The sample is continuously extracted from the process and pumped through this cell. Light is then passed through a rotating wheel containing a number of optical interference filters. Interference filters allow light only within a narrow wavelength range, typically 10–20 nm wide, to pass. Different wavelengths are selected in this way and rotated in front of a NIR detector. Thus, traditional photometers had a moving part and still extracted the sample from the process to the instrument. More recently, fiber optics have been integrated into photometers to gain the advantages described above. Also, by using beamsplitters and multiple detectors rather than moving filter wheels, the most recent process photometers have eliminated all moving parts and allow all detectors to analyze their wavelength simultaneously, as in the PDA. This is a subtle, but important point. In process analysis, the sample continuously moves past the fiber optic probe. By analyzing all wavelengths on the same volume of sample at the same time as it moves through the fiber optic probe, potential ''process noise'' can be eliminated. By adding a microprocessor, process NIR photometers can use multiple wavelengths and MLR in a broad range of applications.

4.6 Fiber Optic Multiplexing

Fiber optic multiplexing is another important technological advancement for process analysis. It permits several fiber optic probes to be sequentially analyzed with one instrument, thereby decreasing the cost per point of analysis. One means to accomplish this is to use a pair of angled reflectors at opposite ends of a rotating shaft inside the multiplexer. Light is brought into the center of one side toward the axis of the shaft, reflected up along a rotating arm and back out the same side into another fiber optic cable connected to the first probe. Thus, there is a central fiber optic connection and several connections around the periphery of a circle. Light returning from the probe is connected to the opposite side of the multiplexer at a similar position on the periphery of the circle, reflected down to the shaft and reflected again into the central connector and out toward the detector. Rotating to the next position on the circle examines the next probe, and so on.

One fiber optic cable can be routed from one side to the other of the multiplexer bypassing a fiber optic probe. This can be used to measure changes in the instrument over time, such as the aging of the light source, and correct the spectrum from each probe accordingly. This is particularly important if such changes have different magnitudes in different spectral regions. This procedure is commonly referred to as ''internal'' referencing. It is generally needed in process systems because the reference spectrum is taken with an empty transmission probe or a reflectance standard with a reflectance probe. It is stored in the computer and used for many months. Therefore, the internal reference is used to make corrections when process probes cannot be removed frequently from the process and re-referenced.

4.7 Enclosures for the Process Environment

The presence of process NIR spectrophotometers and photometers in the production environment pose fire or explosion hazards if flammable vapors exist. Therefore, NIR process equipment will likely require special enclosures to reduce or eliminate such risks to personnel and equipment. Hazardous area classifications for North America are divided into Class, Division and Group under the NFPA 496:1993 standard. Common classifications for NIR analyzers are given below.[6]

4.7.1 Class 1, Division 1, Groups B–D

Areas where flammable gases and vapors are likely to be present continuously, intermittently or periodically in quantities sufficient to produce explosive or ignitable mixtures, including most chemical vapors (other than acetylene, Group A). Enclosures for these environments can be bolted and sealed, or may have a pressurized purge systems that will turn off the instrument if pressurization is not maintained. European Zones 0 and 1 correspond to Division 1.

4.7.2 Class 1, Division 2, Groups B–D

Areas where flammable gases and vapors are likely to be present to produce explosive or ignitable mixtures only in the case of accidental rupture or breakdown of equipment, including most chemical vapors (other than acetylene, Group A). Enclosures for these environments generally have pressurized purge systems in which the instrument will continue to function, even if pressurization is lost. Loss of pressurization can be detected electronically by process control computers systems, although such connections require intrinsic safety barriers. European Zone 2 corresponds to Division 2.

4.7.3 General Purpose, NEMA 4

Although there are variety of NEMA classifications, NEMA 4 is a commonly used enclosure, where 4 represents watertight (weatherproof). A comparable European designation would be IP65.

5 FIBER OPTICS AND FIBER OPTIC PROBES

The advancements in NIR instrumentation described above would required concomitant advancements in fiber

optic probe technology. The most stable spectrometer would be useless if the fiber optic probe to which it is attached falls apart in the process. Process probes have to withstand the rigors of the manufacturing environment, including high temperatures, high pressures, aggressive chemicals and mechanical vibrations, and often, combinations of the above.

5.1 Transmission Probes

5.1.1 Insertion

Insertion probes are typically long metal rods, 18–25 mm ($\frac{3}{4}$–1 in) in diameter and commonly 9–28 cm (6–12 in) long that are designed to be inserted through ball valves into the side of a reactor or into a large pipe. These probes generally transmit a collimated beam of light through the sample only once between sapphire windows placed in a cutout along the side of the probe near the tip. Collimated light is used to eliminate the potential effect on the spectrum due to changes in the refractive index of the liquid sample. Light is coupled into and out of the probe with fiber optic connectors. Light propagates within the probe along one side either using a fiber optic cable or a hollow reflective tube. It is reflected back up the other side of the probe using a retroreflector placed inside the probe at the tip. Light passes through a collimating lens and through the sample between the windows and then out of the probe to the analyzer. The windows are generally brazed into the probe and the braze material may be subject to chemical attack. There always seems to be one application in which the chemistry attacks the braze material. Some use primary o-ring seals for the windows selected for specific chemical compatibility rather than using a braze. Insertion probes must be removed using the ball valve for cleaning and re-referencing. These probes can typically withstand temperatures up to 250 °C and 1000 psi (7 MPa). They are more expensive than flow probes by 50–100%, but they are often the only way to continuously monitor reactions in a reactor. Typical optical paths range from 1 to 20 mm for transmission probes in the NIR.

5.1.2 Flow

These probes are designed for slip (side) streams, which are fast bypass liquid streams with smaller diameter piping. Liquid is often pumped through side streams. Flow probes are placed between pipes anywhere from ($\frac{1}{4}$–1 in) (6–25 mm) in diameter in which the liquid flow passes through the metal body of the probe. Most rugged flow probes have sapphire windows and primary o-ring seals. Collimated light passes across the liquid moving through the probe perpendicular to the direction of flow. These probes generally have a removable clean port. The side stream can be closed off and drained using valves and sample ports, and the clean port on the probe can be removed for easy cleaning of the optics. Thus, the probe does not have to be removed from the process piping. This is important because, in hazardous environment, optical fibers are contained in conduits attached directly to the probe making frequent removal difficult. The sample can be also be conditioned in side streams by heat-tracing to control sample temperature and by filtering to remove light scattering particles or bubbles that would adversely impact spectroscopic measurement. Side streams allow samples to be readily taken for laboratory analysis in order to adjust the calibration of the instrument over time.

5.2 Reflectance Probes

Reflectance probes are used similar to the insertion probes described above. They generally use multiple fibers to provide larger areas of illumination and greater intensity for collection of light reflected from the sample. Many configurations are commercially available. The simplest involves placing six optical fibers around one central fiber within the probe. The six fibers are bundled into one optical fiber generally connected to the light source, and single fiber is connected to the detector. The fibers in the probe are generally placed in a $\frac{1}{4}$ in (6 mm) diameter metal tube containing a sapphire window at the tip. Various techniques are used to minimize the specular component of the diffuse reflectance from the sample and to minimize stray light inside the probe. Other designs utilize larger diameters and have reflective tubes inside, thereby eliminating the presence of fibers within the probe. This permits higher temperature operation. Reflectance probes can be inserted directly into light scattering liquids or into moving streams of powder. Sample presentation at the probe tip for solids is critical. Variations in packing efficiency of the powder against the probe will strongly impact the quality of spectral data. Using fiber optic reflectance probes on-line becomes more difficult for larger solids, such as pellets, simply because the optical fibers limit the image spot size relative to the larger size of pellets. Illuminating larger areas of solids from a distance has been used to analyze for moisture for many years. However, these methods do not rely on fiber optics.

5.3 Internal Reflectance Probes

Attenuated total internal reflectance (ATR) is used to measure the liquid phase in multiphase samples containing light scattering centers, such as particles in slurries and bubbles in foams. This method has been used successfully in the IR and the ultraviolet (UV) where the absorption from vibrations and electronic transitions,

respectively, are intense. This is not the case in the NIR. ATR operates by propagating light within an internal reflectance element. In the IR, this is usually a cylindrical rod or flat plate. Light internally reflects off the surface of the element in contact with the sample and ''senses'' the sample only within a micron or so into the sample, even though the propagated beam is contained within the rod or flat plate. Only a few reflections are required in the IR to acquire good quality spectra. The peaks in the NIR are much weaker and require thousands of reflections to produce acceptable spectra. One way has been found to accomplish this in the NIR, and that is to shrink the thickness of the internal reflectance element to less than one micron and propagate light through this thin film. This thin film is deposited on a glass surface at the end of a fiber optic insertion probe that is placed in contact with the liquid sample. Thousands of internal reflections occur resulting in a spectrum of the liquid phase in the slurry approximately equivalent to a transmission cell with a path of 3–4 mm. The danger in using ATR in the process is that surfaces can become coated and prevent the ATR probe from sensing changes in bulk chemistry.

5.4 Fiber Optics

The optical fibers available for NIR are multi-mode with diameters generally ranging from 200 to 600 microns. They have a silica core with a doped silica cladding, and a polyamide protective coating. There are two types of fibers used with process analyzers, referred to a low-OH and high-OH. High-OH fibers are used in the UV/VIS because they are less prone to solarization (darkening) upon exposure to UV light. Their higher OH content precludes their use for most NIR applications. Low-OH fibers are used in the NIR to minimize absorption due to the fiber itself in the important O–H regions near 1450 nm and above 1900 nm. A lens is used to inject light into the optical fiber to match its typical acceptance angle of 22°. Light follows a variety of paths through the fiber and is only attenuated slightly over long distances. For example, a typical low-OH fiber at 1600 nm only absorbs about 0.08 AU over 100 m, which means that about 83% of the light is transmitted in each direction to the fiber optic probe, or 69% round trip. Assuming that the fiber optic probe itself transmits 40% of the light and the sample in the probe, in turn, transmits 40% of the light, that leaves 69% times 40% times 40%, or about 11%, which is easily measured with most NIR detectors. Fibers are jacketed by a variety of protective materials that may also be reinforced by metal coils. Fibers are typically epoxied into metal connectors with their ends exposed and highly polished. Single fibers are generally used over longer distances due to cost, whereas fiber bundles are used over short distances. Fiber bundles may have tens or even hundreds of fibers contained within the cable. Fiber cables are generally secured in manufacturing facilities within conduits or in cable trays. Care must be taken not to bend fibers too sharply. Typically a minimum 6 in (9 cm) bend radius is specified.

6 ORGANIZATIONAL ISSUES

Although this article has focused upon the technologies involved in process NIR, it may not be obvious to many readers that successful implementation is more dependent upon corporate organizational issues than upon the instrumentation itself. This is a bold statement.

Process analytical chemistry is highly multi-disciplinary.[7] It involves process engineers and chemists who know what they want to measure and where it can be measured, instrumentation technicians who will maintain the instrument, technology gatekeepers who look for new instrumentation, analytical chemists who perform the laboratory chemical analyses upon which the NIR calibration is based, and researchers who perform the initial feasibility studies. There are also specialists in data communication, safety and others. Therefore, the person who identifies the instrument may not be the same person who calibrates it, or who is the eventual ''owner'' of the equipment. The person who develops the NIR calibration may not be responsible for its maintenance over time.

Companies that establish the means for this ''internal technology transfer'' among their various groups will have the greatest likelihood of achieving the benefits of on-line NIR technology described above. A fundamental requirement for such success *is visible and consistent upper management support of process analytical chemistry*. Without this, workers will be less inclined to ''stick their neck out'' and try a new technology that may help their organization to improve product quality more efficiently. This is difficult in an era of corporate downsizing where concern for job security does not inspire workers to take such risks. This is also made difficult in most instances because these instruments are often installed in existing, profitable processes. Their benefit is not ''all or nothing'', but rather incremental, which is harder to justify. Sometimes, catching a major process upset early pays for the entire cost of the instrument. In other words, if everything worked perfectly there would be no need for NIR instrumentation in the first place. Lastly, for companies just beginning to consider such on-line measurement technology, it is important to build advocacy with early success by judicious selection of the first application. Selecting a pricey nine-probe NIR spectrophotometer and developing the world's most sophisticated PLS model for the first

on-line NIR application is probably not the best approach to build this advocacy.

With diligence and a consistent focus, NIR process instruments can help manufacturers achieve the objectives of improved product quality, reduced manufacturing costs and enhanced worker safety.

LIST OF SYMBOLS

nm	Nanometer
cm^{-1}	Wavenumber ($= 10^7$/nm)
A	Absorbance
I	Sample signal
I_o	Reference signal
C	Concentration
ε	Molar extinction coefficient
l	Optical pathlength
Si	Silicon
InGaAs	Indium–Gallium–Arsenide
psi	Pounds per square inch
MPa	10^6 Pascal

ABBREVIATIONS AND ACRONYMS

AOTF	Acousto-optic Tunable Filter
ATR	Attenuated Total Internal Reflectance
FT/NIR	Fourier Transform Near Infrared
GC	Gas Chromatography
IR	Infrared
MLR	Multiple Linear Regression
NIR	Near Infrared
PDA	Photodiode Array
PLS	Partial Least Squares
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Polymers and Rubbers **(Volume 9)**
Near-infrared Spectroscopy of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis

REFERENCES

1. D.A. Burns, E.W. Ciurczak, *Handbook of Near-infrared Analysis*, Marcel Dekker, New York, 1992.
2. M.A. Scharaf, D. Illman, B.R. Kowalski, 'Chemometrics', (Volume 82 in *Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications*), John Wiley & Sons, New York, 1986.
3. H. Martens, T. Maes, *Multivariate Calibration*, John Wiley & Sons, New York, 1986.
4. H. Mark, 'Principles and Practice of Spectroscopic Calibration', (Volume 118 in *Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications*), John Wiley & Sons, New York, 1991.
5. W.M. Doyle, 'Principles and Applications of Fourier Transform Infrared (FTIR) Process Analysis', *Process Control Quality*, **2**, 11–41 (1992).
6. K.J. Cleavett, *Process Analyzer Technology*, John Wiley & Sons, New York, 849–851, 1986.
7. S. Jacobs, S. Mehta, 'Process Quality Measurement Systems: An Integrated Approach to a Successful Program in Process Analytical Chemistry', *Am. Lab.*, 15–22, (December 1987).

Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Michael J. McCarthy and Serge Bobroff
University of California, Davis, CA, USA

1 Introduction 1
2 History 2
3 Application Modes 2
3.1 Low Resolution 3
3.2 High Resolution 4
3.3 Particle Size Measurements 4
3.4 Magnetic Resonance Imaging 5
4 Nuclear Magnetic Resonance Applications 6
4.1 Moisture Content 6
4.2 Quality Evaluation 7
4.3 Spectroscopy 7
4.4 Polymer Characterization 10
4.5 Subsurface Earth Hydrocarbon Characterization 10
5 Emerging Applications 10
5.1 Viscosity Measurements 10
5.2 Defect Detection 11
6 Instrumentation Considerations 12
6.1 Magnets 12
6.2 Gradient Coils 13
6.3 Radio-frequency Coils 13
6.4 Process Modifications 14
6.5 Basic Troubleshooting 15
7 Method Development 15
Acknowledgments 15
Abbreviations and Acronyms 15
Related Articles 16
References 16

Nuclear magnetic resonance (NMR) spectroscopy is an experimental technique based on the interaction between nuclear magnetic moments and an applied magnetic field. The nuclei emit and absorb energy at a specific frequency that is proportional to the applied magnetic field. Magnetic resonance imaging (MRI) is the term used to describe the spatial resolution of NMR signals within a sample. NMR/MRI may provide information on the weight, chemical composition, physical properties, and structure of a sample. NMR is a noncontacting and noninvasive measurement. NMR and MRI are suitable for use in process control and have been applied to polymer processing, food processing and refinery processes. The response is linear from the detection limits of approximately 10 ppb to 100%. The accuracy of the technique depends on the signal-to-noise ratio (S/N), which is a function of the sample size, applied magnetic field strength, the radio-frequency (rf) coil and the spectrometer hardware. Typical accuracy of the technique is 0.5% for measurements of composition and 1–5% for viscosity. NMR cannot generally be applied to magnetic or highly conductive materials.

1 INTRODUCTION

Process analytical chemistry is a rapidly expanding field. The need to enhance production efficiency and maximize profits has provided a strong driving force to improve the control of industrial processes. This has resulted in efforts to develop new types of process sensors based on traditional analytical instruments. There has been considerable success in developing on-line analytical instruments based on gas or liquid chromatography, infrared spectroscopy and mass spectrometry, and flow injection analysis.[1–3] Until recently there has been limited success in developing a process analyzer based on NMR. Several companies now market on-line or in-line analyzers based on NMR.

NMR has many properties that are advantageous when considered for process control applications.[4] These advantages include a noncontacting and noninvasive measurement. The NMR signal is directly proportional to the number of nuclei in a specific sample volume. The signal is linear from the detection limits (~10 ppt) to 100%.[4] The signal can be used to measure multiple quality factors like concentration and droplet size. The instrument has no moving parts. Highly specific information can be obtained by analysis of only one nucleus ^{1}H, ^{13}C, ^{19}F, ^{11}B, ^{23}Na or ^{31}P or a combination of nuclei utilizing the same instrument.

Application of NMR for on-line or in-line sensing requires consideration of the influence of motion, temperature and environmental factors. For nuclei with a spin angular momentum, the net magnetization (M) developed in the sample after it is placed in an external magnetic field[5] is given by Equation (1):

$$M = \frac{N\gamma^2\hbar^2 I(I+1)}{3kT} H_0 \qquad (1)$$

Where N is the number of nuclei, γ is the magnetogyric ratio, I is the spin quantum number, H_0 is the external

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

magnetic field strength, $\hbar$ is the Planck constant, k is the Boltzmann constant, and T is the absolute temperature. The time development of the magnetization is most often well approximated by an exponential time constant the spin–lattice relaxation time (T_1). The proportionality of the nuclear magnetization to temperature and time has important implications in the design of on-line NMR sensor systems.

Temperature effects can either be dominant in the measurement or unimportant. For sensing based on relationships to relaxation times or other temperature sensitive parameters, precise control of the sample temperature is necessary. Temperature control is required in the measurement of solid-to-liquid ratio in fats and oils. However, in many applications using low resolution or high resolution spectroscopy the temperature has a negligible effect on the method and temperature variations can be ignored. This is often the case in measurements of the moisture content of products where the field homogeneity of the magnet dominates the decay of the signal.

The influence of motion on a NMR signal is manifest through changes in the effective relaxation times and signal intensity. The influence of motion on the relaxation times[6] is expressed as Equations (2) and (3):

$$\frac{1}{T_1^{\text{observed}}} = \frac{1}{T_1^{\text{static}}} + \frac{1}{\tau} \tag{2}$$

$$\frac{1}{T_2^{\text{observed}}} = \frac{1}{T_2^{\text{static}}} + \frac{1}{\tau} \tag{3}$$

Where the inherent relaxation times are designated with the superscript static, the relaxation times observed during flow are designated with the superscript observed, and τ is the lifetime of the nuclei in the measurement zone. If the sample flows at the same rate through the measurement zone the lifetime will be given as the measurement zone volume divided by the flow rate. Thus, motion of the sample will cause a decrease in the observed time constants for relaxation. This change in relaxation times will result in an increase in the linewidth of the NMR spectra. The sample will also have a residence time in the magnetic field. For full equilibration this should be at least five times the largest spin–lattice relaxation time observed in the sample. If full equilibration is required for the measurement method, the design of the magnetic field, rf probe and sampling system will be critical to successful implementation.

2 HISTORY

NMR spectroscopy is now over 50-years old and has become one of the premier experimental methods for studying the structure and dynamics of condensed matter. The first successful experiments were performed independently by Bloch and Purcell in 1946.[5] Soon after the first experiments, a NMR spectrometer was modified for process control. Nelson, Reilly and Savage[7] developed a NMR spectrometer to record the proton spectra from a flowing stream at 30 MHz. The proton spectra were recorded every 6 s in a 3-mm diameter glass tube. The spectrometer was coupled to a pilot plant process. To overcome the problem of achieving equilibration of the sample, a reservoir was located within the magnet. This demonstration of using a NMR spectrometer for process control was successful however; the authors noted the high cost of the sensor. Additional applications with magnetic fields capable of resolving chemical shifts include measurement of viscosity[8] and fat content in meat.[9]

The largest number of applications that have been proposed, tested and implemented are based on low-cost permanent magnet systems. Two groups, Tri-Valley Research and The Southwest Research Institute made major contributions to the early development of these systems.[9] These NMR analyzers were primarily applied to the measurement of moisture content in agricultural and food products.[6,9–11]

Several companies have begun to offer industrial NMR process analyzers.[12] Auburn Instruments offers a permanent magnet-based system to measure the properties of polymers (http://www.auburnint.com/ (July 30, 1998)). Foxboro Instruments has a spectroscopy-based system oriented towards applications in the oil and chemical industry (The Foxboro Company, http://www.foxboro.com/nmr/index.htm (July 30, 1998)). An MRI based whole-body clinical system has been modified by a Intermagnetics General Corporation/SMIS Ltd joint venture for final product quality monitoring of a packaged food (http://www.industrialmr.com/index.htm (July 30, 1998)).

3 APPLICATION MODES

NMR sensor systems may be designed based on the properties of the NMR equipment or on the requirements of the measurement. The deciding factor is the result of the economic analysis and feasibility testing. Constraints applicable to common types of NMR/MRI systems will be presented in this section. The component that most often defines the range of applications of a NMR/MRI sensor is the magnet. The properties of the magnet, field strength, field homogeneity, and homogeneous volume are the most important for defining the range of applications for a particular NMR/MRI system.

3.1 Low Resolution

The term low resolution indicates that the NMR equipment has a magnet with a low-magnetic field strength. NMR systems with a low-magnetic field strength (<0.5 T) are not usually used to separate the signal from different components by analysis of the frequency content of the signal. This type of analysis has also been referred to as wide-line NMR. Discrimination is usually based on the rates of signal decay governed by either the spin–lattice relaxation time or the spin–spin relaxation time or the magnitude of the signal at specific times. NMR relaxation times characterize the interactions between the nuclei and their surroundings in the sample. The spin–lattice relaxation time characterizes energy exchange between the nuclei and the surrounding lattice. The spin–spin relaxation time (T_2) characterizes the loss of phase coherence between the magnetic moments of the nuclei. If the magnet field in a low-field spectrometer has a high homogeneity the signal from different components can be resolved based on the difference in the frequency response from each component. There have been few applications of these types of system with a highly homogeneous magnetic field of low strength. The applications are similar to those discussed in the next section on high resolution systems.

Low resolution NMR units are often of a small enough size to be located easily on a laboratory bench top. The costs of such a unit are also modest in comparison with commercial research grade NMR systems. Low resolution systems have been available since the early 1970s.[13] These systems have been widely used in the chemical, medical, agricultural and food industries.

The signal from low resolution NMR is analyzed in the time domain. The main focus has been on the use of data from the free induction decay (FID), spin-echo or pulse sequences Carr–Purcell–Meiboom–Gill (CPMG) (as shown in Figure 1). The different types of analysis can be classified into the ratio, absolute, or relaxation time methods.[13]

Ratio applications are where the intensity of the signal at two different times during the signal decay is used to form a ratio (see Figure 1). This ratio is then directly related to the quantity of interest, for example, a moisture content or a solid–liquid ratio in oil. The advantage of this type of measurement is that it is independent of the weight of the sample and the sample volume does not need to be constant from sample to sample. Absolute methods are based on the value of the signal at a specific time. For this type of measurement it is critical that the rf probe have a high homogeneity over the entire sample. The volume of material tested must be precisely controlled from sample to sample and during calibration of the instrument. Relaxation time applications use either the initial amplitude from one component in a multicomponent material or the relaxation data itself to correlate with a quality parameter like moisture content. Simple rf probes like those used for the ratio measurement are appropriate unless a weight-normalized measurement is needed.[13]

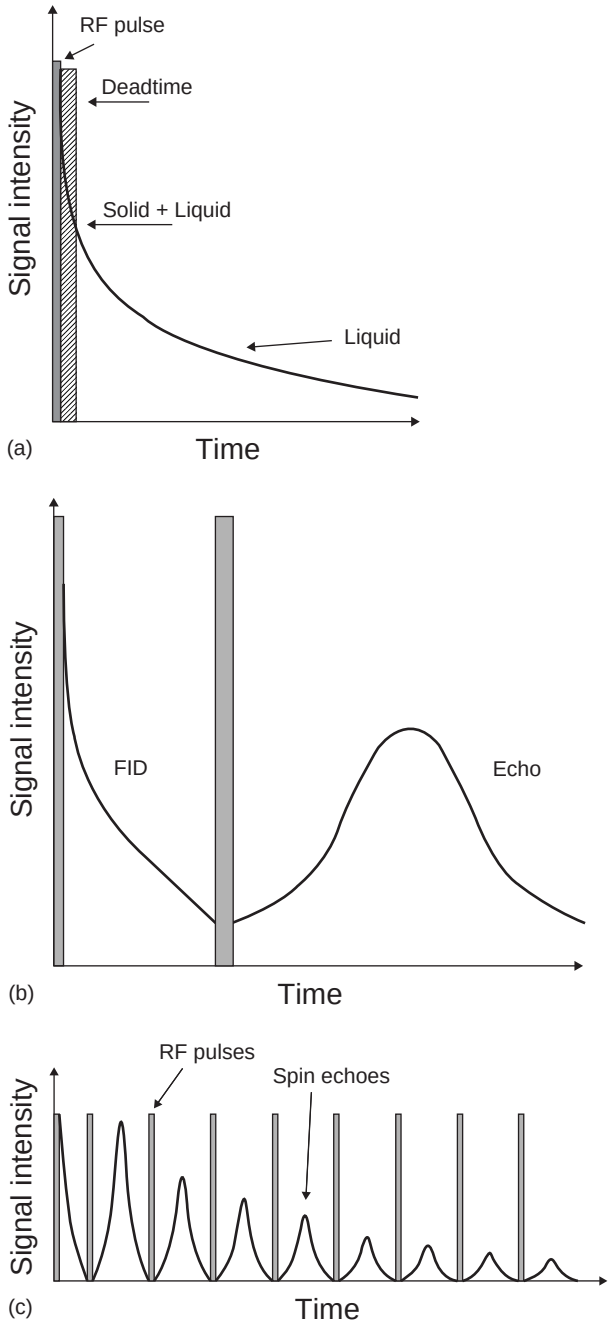

Figure 1 (a) Schematic of the FID from a sample containing solid and liquid components, (b) the spin echo response from a sample, (c) the CPMG response from a sample.

Details of the most common experiments used in low resolution NMR analysis are given in Figure 1.

The FID (Figure 1a) is often used for moisture content measurement or determination of the solid–liquid ratio in fats and oils. The initial signal following the rf pulse is proportional to the total number of protons in the sample. The time decay of the solid and liquid fractions differs by several orders of magnitude, T_2 for solid fats is between 10–20 μs and T_2 for liquid fats is ~100 ms. Therefore, if the signal amplitude at the initial time is compared with the amplitude a short time later, the solid to liquid ratio can be calculated. This is because both the solid and liquid phases contribute signal initially and after a short time only the liquid component contributes to the signal (see Figure 1a), the solid signal having decayed completely. Deadtime in commercial NMR instruments ranges from 7–10 μs after the end of the rf pulse. The liquid signal is generally recorded at 70 μs after the rf pulse. Equation (4) is the expression for the liquid fraction corrected for the deadtime of the NMR receiver:

$$\text{liquid fraction} \quad \frac{w}{w} = \frac{L}{L + f\mathrm{S}} \tag{4}$$

Where f is a correction factor that accounts for the decay in the NMR signal during the time prior to signal acquisition.[9] The other experiments, Figures 1(b) and 1(c), can be used to measure oil content in seeds or moisture and oil content in confectionary products.[13]

3.2 High Resolution

The most important difference between low resolution NMR systems and high resolution NMR systems is the strength of the applied magnetic field. The higher magnetic field permits the separation of spectral components based on their frequency response. A high resolution NMR system is capable of producing data such that nuclei in different local chemical environments can be identified based on their frequency response. The proton resonance frequency for on-line applications of high resolution NMR is between 30–100 MHz. The variation in the frequency response is referred to as the chemical shift. The chemical shift is reported in parts per million referenced to a standard. The most common standard used for proton spectra is tetramethylsilane (TMS). Table 1 demonstrates the range of proton and carbon chemical shifts for different chemical structures. Figure 2 is the proton spectrum taken at a proton resonance of 85.5 MHz from an intact durian fruit. The largest peak is from water; the next two are from sugar and oil components, respectively. The types of data shown in Figure 2 are used to calculate the ratio of the components or a concentration of one of the components, for example, sugar content. Components with at least 1 Hz difference in their resonance frequency are resolvable, as long as the magnetic field homogeneity is significantly better than 1 Hz and the resonance lines are narrow.

Table 1 Representative proton and carbon chemical shifts

Structure	Proton chemical shift (ppm)	Carbon chemical shift (ppm)
Alkyl	1–4	10–50
Alkenic	5–7	100–150
Aromatic	7–9	100–150
Carbonyl	8–10	150–200
Alkynic	3	50–80

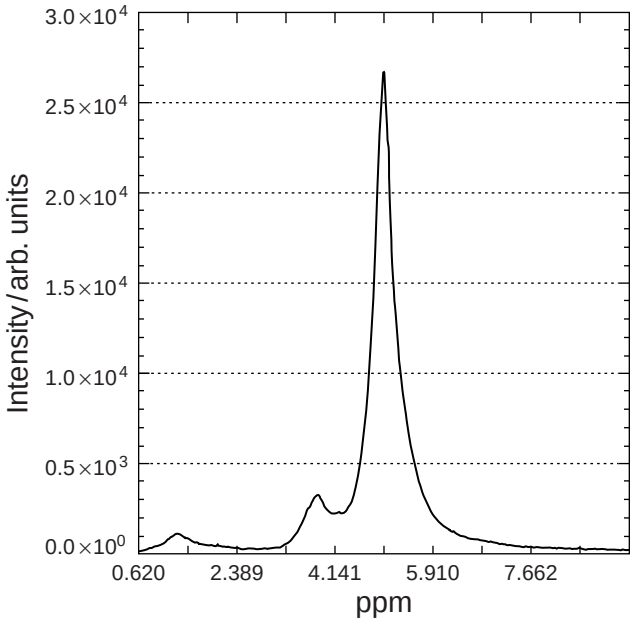

Figure 2 Proton NMR spectra from durian fruit at 85.5 MHz.

3.3 Particle Size Measurements

NMR is one of the premier methods available for measuring the self-diffusion coefficient of molecules. The pulsed field gradient (PFG) NMR technique, as developed by Stejskal and Tanner[14] has been widely applied for measuring the effect of molecular diffusion on the NMR signal in a spin echo experiment. When a system has a heterogeneous structure, the translational motion of diffusing molecules is strongly influenced by the internal structure of the sample. In a confined geometry, the maximal extent of diffusional displacement is limited by the size of the confinement. For example, the geometry and the dimension of the internal droplet boundaries restrict the diffusion of oil molecules within an emulsion droplet. This effect is referred to as ''restricted diffusion'' and has been studied extensively.[15] For molecules bounded by a sphere of radius a, the signal attenuation in a PFG NMR experiment is given by Equation (5)[15]

$$R = \exp\ -2\gamma^2 g^2 \sum_{m=1}^{\infty} \frac{1}{\alpha_m^2(\alpha_m^2 a^2 - 2)} \times \frac{2\delta}{\alpha_m^2 D} - 2 + e^{-\alpha_m^2 D\delta(\Delta-\delta)} - 2e^{-\alpha_m^2 D\delta} - 2e^{-\alpha_m^2 D\Delta} + e^{-\alpha_m^2 D(\Delta+\delta)} \quad (\alpha_m^2 D)^2 \Bigg\} \tag{5}$$

Where α_m is the mth root of the equation $J_{3/2}(\alpha_m a) = \alpha_m a J_{5/2}(\alpha_m a)$, with $J_p(x)$ the Bessel function of order p. When Equation (5) is applied to an emulsion the effect of polydispersity in droplet sizes needs to be incorporated. Packer and Rees[16] have analyzed the observed echo attenuation as a sum of the echo contributions from individual droplets weighted by the volume of each droplet in Equation (6)

$$R_{\text{obs}}(\Delta, g, \delta) = \frac{\int_0^{\infty} a^3 P(a) R(\Delta, g, \delta; D, a)\, da}{\int_0^{\infty} a^3 P(a)\, da} \tag{6}$$

Here, $R(\Delta, g, \delta; D, a)$ is the signal attenuation given by Equation (5) and $P(a)$ is a log–normal distribution. This method is promising as a technique that quickly and accurately quantifies the droplet size distribution in an emulsion, for example, the water droplet size distribution in margarine. Particle sizes on the order of 0.1 μm to tens of micrometers can be measured. Unimodal and bimodal distributions can be characterized.[16]

3.4 Magnetic Resonance Imaging

MRI is a group of NMR methods for spatially localizing the NMR signal from within a sample.[15] This technique can be used to measure the density distribution of water or other components, the velocity distribution of a fluid, the diffusivity of a molecule or other properties. MRI techniques can be subdivided into point, line, and multidimensional imaging protocols. In each of these experiments the signal is obtained from a three-dimensional volume element called a voxel. In the case of a point imaging experiment only the signal from a small volume located within the sample is acquired. For example the signal from inside a peach next to the stone. The volume being investigated is on the order of 0.5 cm^3. The information obtained is similar to that acquired in a high resolution NMR experiment. For line imaging, a series of small volumes next to each other are combined to form a line through the sample, usually chemical shift information is not retained in line-type images. Figure 3 is a proton spin–spin relaxation weighted image of an avocado. The signal intensity a MRI data set can be weighted to provide a variety of information.[15] In the case of relaxation-weighted images, this procedure is often used to highlight specific structure in a material that has different properties from the rest of the sample.

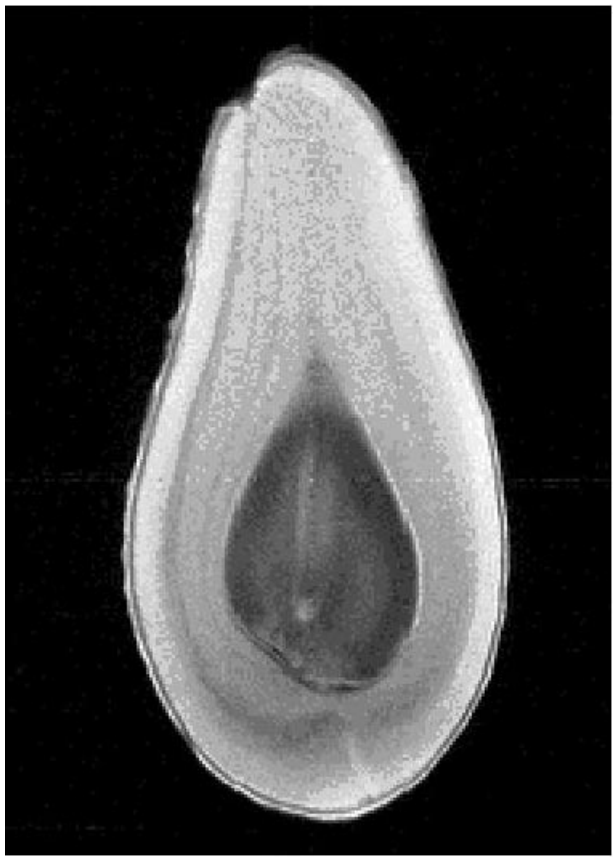

Figure 3 Proton spin–spin relaxation weighted image from an intact avocado fruit. The lighter the gray intensity the higher the MRI signal intensity.

The majority of MRI methods employ a spin echo-based pulse sequence. The signal intensity in each voxel of an image, in the absence of translational motion is given by Equation (7):

$$M(t) = K \sum \rho_i \exp\left(\frac{-TE}{T_{2i}}\right) \times \left(1 - \exp\left(\frac{-PD}{T_{1i}}\right)\right) \exp(-bD_i) \tag{7}$$

Here, K is a constant dependent upon the NMR equipment, ρ_i is the nucleus density of component i, TE is the time-to-echo, b accounts for the effective value of field gradients and PD is the predelay. Equation (7) forms the quantitative basis for contrast in the intensity in MRI data. If the sample is undergoing translational motion this can also be used to encode contrast in the intensity of the MRI data.

4 NUCLEAR MAGNETIC RESONANCE APPLICATIONS

There are a large number of reports on the application of NMR/MRI to characterize the composition, physical properties, and defects in both liquid and solid samples. Many of these have been tested with the goal of determining the feasibility of process control. In Table 2 are listed some of the most common material characteristics that can be measured with NMR/MRI. Listed in Table 3 are examples of applications of these characteristics in different industries. Specific examples for each mode of application will be discussed next.

4.1 Moisture Content

Compared with other sensing methods, NMR offers significant advantages in determining the moisture content of a processed material. Table 4 describes the six most common moisture-sensing techniques that are most often used industrially. In comparison NMR has no inherent limitations, with the exception that the capital expense is usually higher. The NMR method provides a measure of the total moisture and can provide continuous on-line measurements.[10,11] In addition, NMR does not require the assumption of a uniform distribution of moisture within the sample, instead measuring the particular nuclei of interest (hydrogen in the case of moisture) throughout the bulk material. In other words, low resolution NMR measures the total hydrogen content regardless of the distribution and can distinguish between aqueous and nonaqueous forms. This is very advantageous compared with other sensing techniques,

Table 2 Sample characteristics that can be measured with NMR/MRI

Chemical properties	Physical properties	Structural properties
Component saturations	Diffusion coefficients	Pore size distributions
Reaction rates	Thermal properties	Droplet size distributions
Mobilities	Rheological properties	Internal structure
Exchange rates	Phase behavior	''Image'' of interior

Table 3 Examples of NMR/MRI applications in process control

Industry	Material	Measurement
Agriculture	Seeds	Oil and moisture concentration
	Feed	Oil and moisture concentration
Chemical	Powders	Moisture content
	Liquids	Viscosity
		Hydrogen content
		Fluorine content
	Pigments	Particle size
	Detergents	Moisture content
	Emulsions	Droplet size distribution
Food	Fats and Oils	Solid/liquid ratio
	Emulsions	Composition
		Droplet size distribution
		Oil and moisture concentration
	Cereal grains	Moisture content
	Milk products	Fat and moisture concentration
Pharmaceutical	Capsules, tablets	Oil and moisture content
		Component determination
Polymer	General	Plasticizer content
		Hydrogen content
	Copolymer blends	Mixing
		Copolymer concentration
	Resins	Extent of cure
Petroleum	Liquid products	Viscosity
	Waxes	Oil content

Table 4 Common moisture sensing technologies

Technology	Advantage	Disadvantage
rf dielectric constant	Good for homogeneous or uniform materials, such as grains	Accuracy is affected by the material density, bridging, and bulking
Infrared	Very sensitive to surface moisture	Does not accurately measure bulk properties, and reflectance and color of the material can cause errors
Microwave	Very sensitive to moisture on the surface or in pores	Not sensitive to chemically or physically associated moisture, like moisture of hydration
Conductivity	Good for homogeneous or uniform materials, such as grains	Accuracy is affected by the material density, bridging, and bulking
Karl Fisher titration	Very accurate, measures total moisture. Is an accepted primary measurement technique	Measurement is time consuming. Operational costs are high due to requirements of highly trained operators and solvents
Oven drying	Very accurate, measures total moisture. Is an accepted primary measurement technique	Measurement time is very long. Not all moisture is necessarily removed and other volatile compounds may be evaporated

especially with microwave methods, where inhomogeneous distributions can seriously affect the accuracy of the measurement.

NMR has been used to measure moisture in a variety of both food and nonfood products, including chocolate, crackers, potato chips, packaged soups, paper and wood pulp, composite structures, soap and detergent, and coal. Pearson et al. have described the application of NMR to measure moisture in aluminum oxide[17] and in agricultural products.[18,19]

One method of determining the moisture content using NMR is to ratio two points from the time domain signal (often called the FID). The points ratioed are separated in time by several hundred microseconds. This method has several advantages in moisture determination. These include:

- The measurement relies on a ratio which is unaffected by the size of the signal. This method is therefore insensitive to the packing density of the sample volume since air voids in the material will only affect the size of the NMR signal and not the ratio.
- The measurement is extremely fast. The calculation of the moisture content from the signal is very simple, so that it is possible to make as many as 500 measurements per second in most materials.
- The measurement is simple and hence the NMR apparatus required is minimized.

Historically many correlations between NMR parameters and moisture content have been performed by first calculating a relaxation time or a ratio of intensities at different times in the decay.[13] Chemometric techniques have been shown to provide better predictions of quality factors when applied to the NMR time domain data directly.[19,20] Improvements in the area of data processing promise to expand the number of economically viable low resolution measurements.

4.2 Quality Evaluation

The potential uses for quality monitoring with MRI are very broad. MRI images can be produced with a contrast that provides specific information on physical properties, product structure, contaminants, safety, stability, processability and consumer acceptance. An on-line quality evaluation sensor has been designed using a MRI system. The MRI uses a permanent magnet and individual cases of the final product are inspected on a conveyor belt (Figure 4). The measurements are implemented using a modified multislice multiecho pulse sequence. This approach permits the entire three-dimensional sample to be investigated rapidly. The system is designed to inspect 100% of production (The Foxboro Company, http://www.industrialnmr.com/index.htm (July 30, 1998)). Critical to the success of process MRI/NMR units is the design of software to diagnose sensor performance.

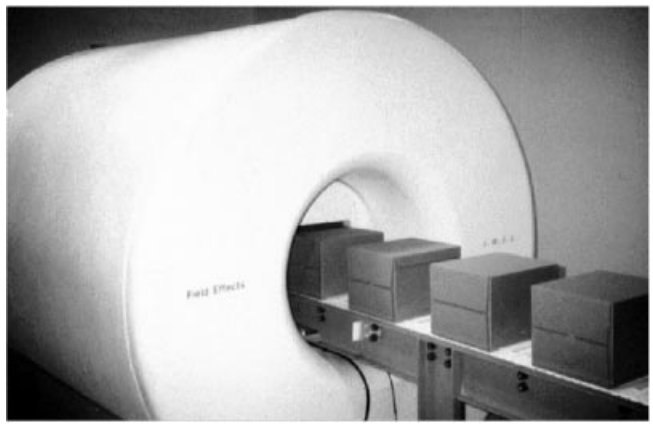

Figure 4 An on-line MRI inspection sensor for measuring the quality of a processed food. (Courtesy of Dr C. Nicholls, SMIS Ltd.)

The diagnostic software developed for this application includes the monitoring of the strength of the magnetic field, the homogeneity of the magnetic field, gradient strength and the operation of the conveyor belt. The sensor is started with a few simple commands shown in Figure 5. During normal operation the screen displays information on the product and sensor diagnostic results. Specific diagnostics include monitoring the magnetic field strength, homogeneity, gradient strengths and conveyor operation. The software automatically corrects for instrumental drifts such as slight changes in magnetic field strength or magnetic field homogeneity.

Variations in the geometric design of the magnet have been demonstrated. For example a single-sided NMR magnet has been implemented by the Southwest Research Institute for the measurement of product moisture content.[10,11] The single-sided system has obvious advantages when applied to conveyor systems both in terms of magnet cost and flexibility of implementation.

4.3 Spectroscopy

The major advantage of using high resolution techniques in comparison to low resolution techniques is that the quantity of different components can easily be separated using the spectral data. This type of data is shown in Figure 2 and the two major constituents are oil and water. The application of high resolution techniques to quantify oil content, moisture content and sugar content in intact fruits, vegetables and foods is straightforward.[9,22,23] Quantification of component concentrations permits grading of food items. For example in US standards for grades of fresh produce the soluble solids content

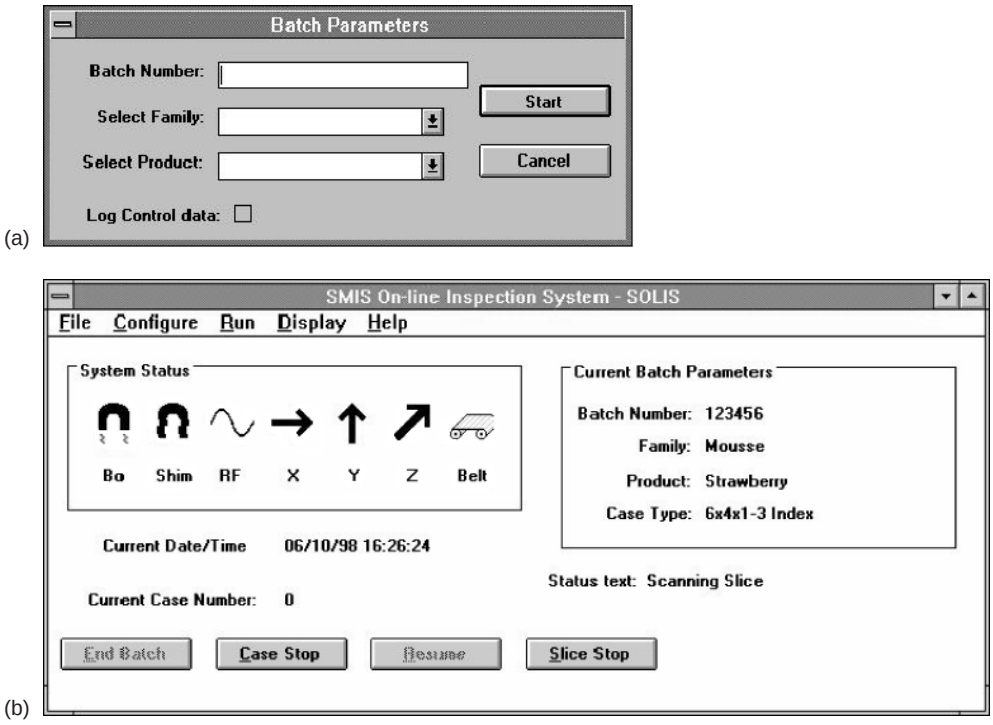

Figure 5 (a) Batch window and (b) main operator windows of the control software for the MRI sensor shown in Figure 4. (Courtesy of Dr C. Nicholls, SMIS Ltd.)

is a key parameter to be evaluated (i.e. juice from cantaloupes should not contain less than 9% soluble solids). Chen et al. have studied the feasibility of grading fruit and determining fruit maturity using high resolution spectra.[22,23] The basic procedure is to record the proton spectra, ratio the spectral peak areas (or peak heights) of the different components and compare with standard destructive analysis. Results obtained from avocado maturity, soluble solids content and sugar content generally have correlation coefficients of 0.9 or better.

The feasibility of making this type of measurement on moving fruit has recently been demonstrated.[23] Spectra were acquired measuring the maturity of avocados on a specially designed conveyor belt. The conveyor was constructed entirely from nonmagnetic parts (see Figure 6). Fruit speed on the conveyor varied between 0–0.25 m s^{-1}. Potential problems such as sample vibration and misalignment of the NMR rf coil were investigated. None of these factors significantly influenced the correlation between NMR spectral analysis and destructive tests.

A commercial sensor is now available to measure the spectra on a liquid side stream (http://www.foxboro.com/nmr/index.htm (July 30, 1998)). The Foxboro Company has recently introduced an on-line, continuous flow-through NMR sensor (see website above). The sensor is capable of multiple analysis of chemical constituents and physical properties. The NMR system is designed to be used on-line in petroleum refineries to perform reformate analysis, crude oil analysis, characterization of naphtha streams, acid alkylation analysis, gasoline analysis and distillate analysis (see website above). This sensor has been installed in a refinery since 1995. The

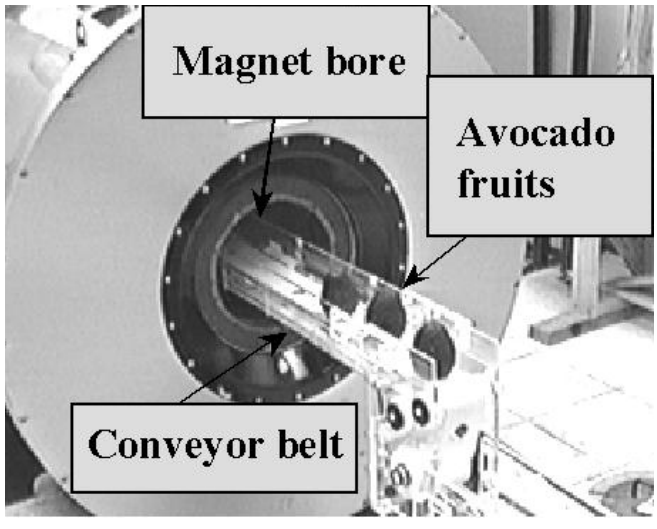

Figure 6 Nonmagnetic pilot scale fruit conveyor coupled to a 2-T superconducting magnet. Avocado fruit are shown on the conveyor belt.

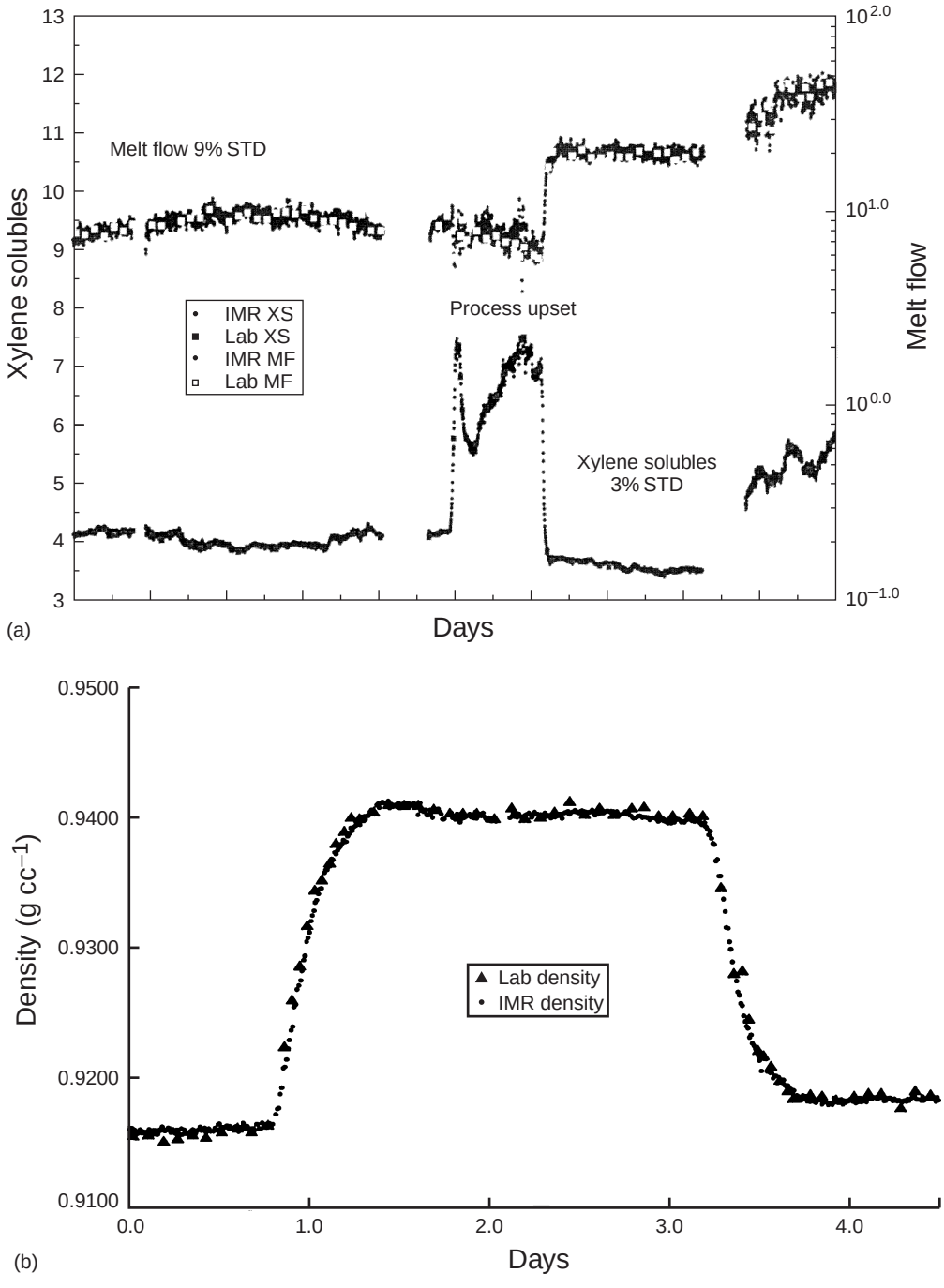

Figure 7 Comparison of NMR process sensor data and laboratory data for (a) the melt index and xylene solubles and (b) gas phase density measurements (Courtesy of Mr Vaughn Davis, Auburn International, Inc.)

system incorporates flow-through rf probes and software with multiple calibration techniques. Advances in the construction of permanent magnets have been critical to the stability and operation of this sensor.[11]

High resolution NMR has also been combined with high-performance liquid chromatography. Techniques using both stopped-flow analysis and continuous flow analysis have been developed. Specific devices are available to interface the NMR spectrometer with either a gas or liquid chromatography system. Combined NMR and liquid chromatography has been used to detect chemical intermediates with short lifetimes and chemical equilibria in batch processing of chemicals.[9]

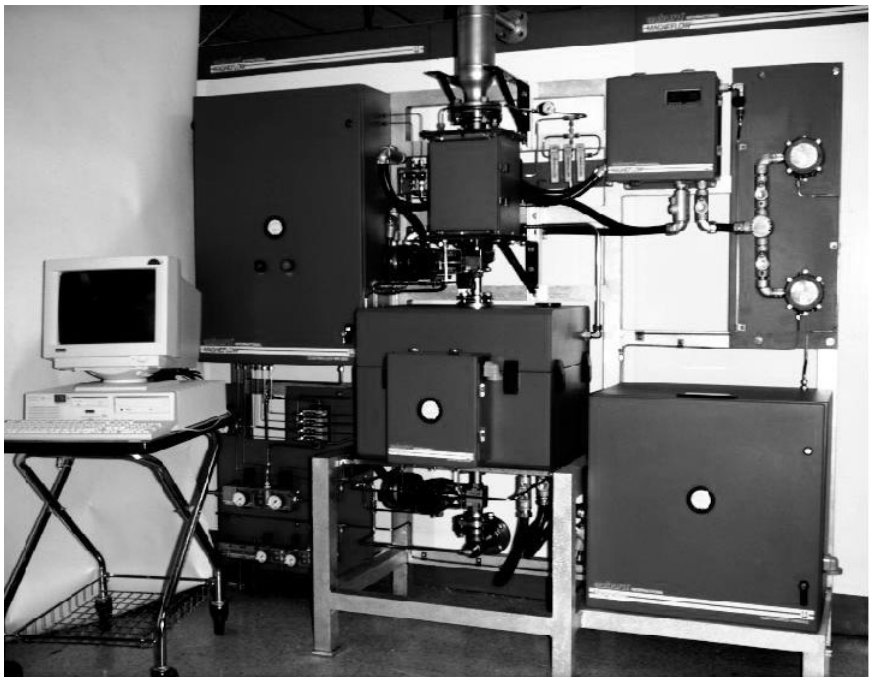

Figure 8 Process NMR analyzer based on a permanent magnet (Courtesy of Mr Vaughn Davis, Auburn International, Inc.)

4.4 Polymer Characterization

Auburn International (http://www.auburnint.com/ (July 30, 1998)) has developed an on-line sensor designed to measure polymer material properties. The sensor is capable of making multiple analyses including density, melt index, copolymer content, solubles, and rubber content. Comparisons of the measurement of the melt index and xylene solubles between the sensor and laboratory procedure are shown in Figure 7(a). NMR may also be used to quantify the properties of gas phase systems, properties such as density can be measured as shown in Figure 7(b). Measurements on polymers can be made on both solid (powders or pellets) and liquid polymers. The sensor is based on a low resolution type NMR spectrometer that has been redesigned for process applications (Figure 8).

4.5 Subsurface Earth Hydrocarbon Characterization

A NMR device has been developed and is now used routinely as one of a number of instruments for characterizing the properties of hydrocarbon reservoirs in subsurface earth formations. The spectrometer has been built to operate at temperatures up to 175 °C and at speeds of up to 10 cm s^{-1}. The NMR sensor is based on measurements performed outside the spectrometer.[24] The measurements are therefore acquired in a very inhomogeneous magnetic field and hence are limited to amplitude and transverse decay of the proton signal. The pulse sequence employed is the CPMG sequence with phase cycling.

The NMR measurements are used to quantify the porosity of the formation, the hydraulic permeability, the fraction of fluid that will not flow (fluid trapped by capillary forces or physisorption on clay materials), and the physical properties of the hydrocarbon. The most important physical property measured is viscosity. The viscosity is measured through a correlation between the distribution of T_2 and hydrocarbon viscosity.[24]

5 EMERGING APPLICATIONS

5.1 Viscosity Measurements

Rheological properties are an important factor relating to processability, safety, stability and quality of many products. Viscosity has been one of the most frequently measured physical properties in NMR/MRI sensor applications. The most frequent method for prediction of viscosity has been through the correlation between changes in relaxation times and solution viscosity. The dependence was predicted and demonstrated in the late 1940s.[25] This method has been applied to a variety of substances including particulate suspensions,[26] polymers[8] and foods.[27] This simple technique is often adequate for process control; however, only a single viscosity at one shear rate or a Newtonian viscosity is obtained. In the

processing of many polymers, foods and fluids exhibiting complex rheology more information is required. For example, in polymer processing the determination of molecular weight distribution is an important parameter and for this determination the shear viscosity versus shear rate over several decades of shear rate is needed. A new MRI based technique has been designed to obtain this data on flowing process streams.[28,29]

The method is based on measuring both the fluid velocity profile ($v(r)$) and simultaneously the pressure drop in the pipe ($\Delta P/L$). The shear rate $\gamma(r)$ is calculated directly from the velocity profile $\gamma(r) = \mathrm{d}(v(r))/\mathrm{d}r$ and the shear stress is calculated from the pressure drop $\sigma(r) = -(\Delta P/2L)/r$. The radial variable is eliminated by combining the equations for the shear stress and shear rate and this permits calculation of $\sigma(\gamma)$ or $\sigma(\gamma)/\gamma$, (as shown in Figure 9). The viscosity data resulting from a single combined measurement of the velocity profile and pressure drop ranges from a shear rate near zero to the maximum at the tube wall ($r = R$). This procedure is essentially an extension of standard capillary viscometry and the shear rate range that is normally obtained is from 10^0–$10^3\,\mathrm{s}^{-1}$. Additional rheological information may also be obtained from the velocity data, including apparent wall slip velocity,[30] and apparent yield stress.[31] The advantages of this MRI-based technique are:

- It can be performed on-line or in-line with only minor modifications to the process line.
- It is applicable to dense suspensions and multilayer systems.
- The same equipment can provide additional quality measurements like component concentrations, or melt index.

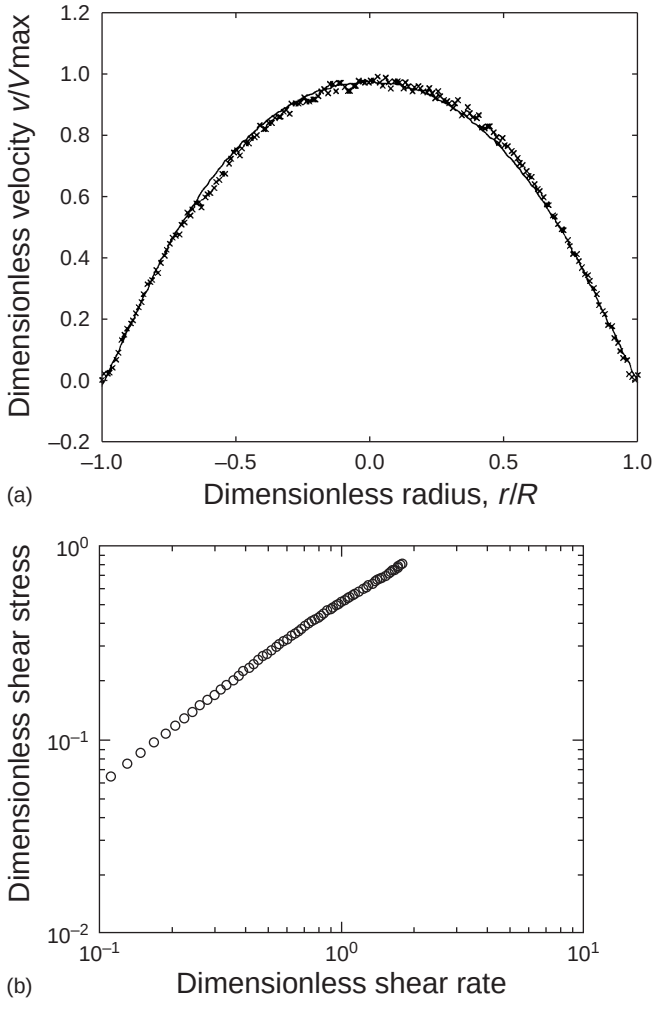

Figure 9 (a) MRI flow image of 50% by weight solution of powdered milk (b) and the resulting shear viscosity shear rate data.

5.2 Defect Detection

Defects such as voids and foreign bodies are frequently detectable using MRI data.[32] MRI can provide a map of the proton density in a material. This density map may be used to monitor a product for defects. For example, in food processing MRI data can be used for pit detection in fruit, quality grading of cheese and inspection of fresh fruit and vegetables for damage.[27] The type of damage that can be detected in fruits and vegetables includes insect damage, bruises, frost damage, voids, internal structure collapse, and the presence of mold.[32] Use of echo planar imaging techniques would enable rapid data acquisition compatible with process line production rates. For objects with a symmetric shape it is sufficient, depending upon the defect, simply to acquire the proton density as a function of position in one direction, a projection. This is demonstrated for detection of pits in processed cherries in Figure 10. The presence of a pit in the cherry changes the expected proton density intensity map. Based on the shape of the profile the presence of a pit can be detected using the projection data.[33]

6 INSTRUMENTATION CONSIDERATIONS

Since the late 1970s NMR has experienced tremendous growth in the clinical and biological arenas. This interest, combined with the immense increase in electronic and computational power, has greatly spurred the development of MRI and spectroscopy. As a consequence, the application of this technique to process analysis is rapidly expanding. The principal components of an NMR/MRI system are summarized in Figure 11.

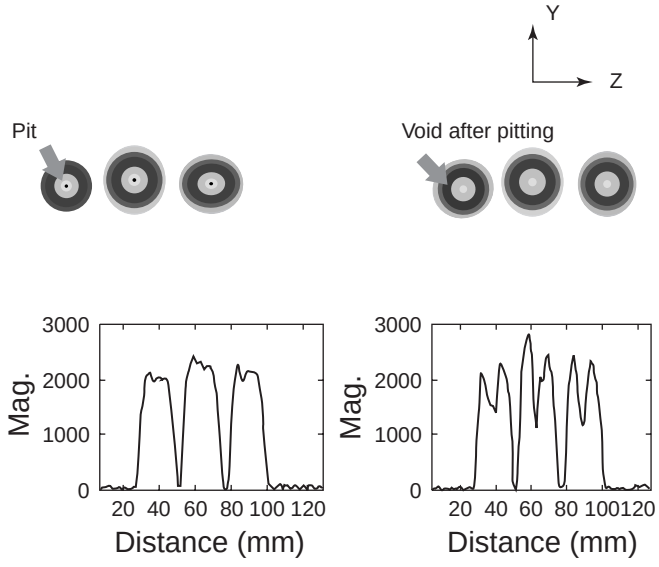

Figure 10 MRI two- and one-dimensional images of whole and pitted cherries acquired while the cherries were moving at 250 mm s^{-1}.

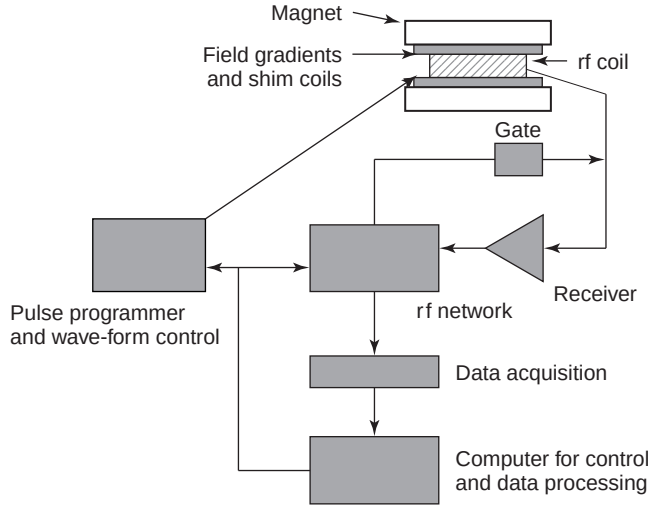

Figure 11 Principal equipment components of a NMR/MRI spectrometer.

Rigorous constraints govern where and how the NMR/MRI system can be located and what type of apparatus will be preferred for a specific application. The choice of the appropriate NMR system is complex and several parameters should be taken into account, the most significant considerations are the magnet, gradient/rf coils and process modifications.

6.1 Magnets

The optimum magnetic field strength will depend upon the application under consideration and the information required. In general the field strength available in NMR process analyzers will provide a proton resonance frequency <100 MHz. The choice of field strength depends on the application under consideration: spectroscopy, imaging or low resolution. Choosing the magnetic field strength may be complicated in process analysis where complex materials are to be investigated and both spectroscopy and imaging information are desired.

Relatively high magnetic fields are essential to magnetic resonance spectroscopy for two basic reasons: chemical discrimination and S/N. Although some applications at low field have shown feasibility for water/fat selectivity,[9,34] the study of more fine detail (i.e. ethanol) or less sensitive nuclei (i.e. ^{2}H, ^{13}C, ^{31}P) favor higher magnetic fields. Magnetic fields producing proton resonance frequency >100 MHz are easily reached only with superconducting electromagnets.[35]

Superconducting electromagnets consist of coils of superconducting materials enclosed in a cryostat. These superconducting materials, when held below a certain temperature (critical temperature) reach almost infinite conductivity. Hence, the materials can carry high currents indefinitely (if maintained at low temperature) producing a very stable and strong magnetic field. Commercially available superconducting magnets operate at liquid helium temperature (4.2 K) and their strength ranges generally from 0.3–11.4 T. The capital cost of the superconducting magnets is high and they require regular refilling with cryogenic liquids. These two factors are generally not considered compatible with process analysis applications. Also, at high fields the intrinsic heterogeneity (internal susceptibility contrasts) of the sample can cause dramatic effects such as broadening of linewidths and decrease in the relaxation times making some applications unworkable. When magnetic field gradients are required, measurements with this type of magnet suffer from strong artifacts resulting from eddy currents created by switching the current in the gradient coils. Using ''shielded'' gradient coil designs where a second set of coils shield the metallic structures within the magnet from the pulsed magnetic fields generated by the switching gradients can minimize eddy current generation.

Considerable effort has gone into permanent magnet design for process applications.[9,12] Permanent magnets are available in a large number of configurations. These can be made as single-sided systems for conveyor applications,[11] with a cylindrical cavity for pipe and conveyor applications or as two parallel plates with the sample passing between the plates. Permanent magnets are made of an assembly of individual magnetized bricks of alnico alloys or rare earth, often mounted around an iron yoke to increase field strength. The principal disadvantage of such magnets is their weight (up to 30 tons for large bore systems). Field strength can range

from ~0.1 T to 2.35 T (proton resonance frequencies from 4–100 MHz). Almost all past and current applications of NMR and MRI for process control have been implemented using permanent magnets.

Resistive magnets are composed of conductive coils (aluminum or copper) that are supplied with constant current (up to 80 kW) to provide the constant magnetic field. In some cases they are mounted around an iron yoke. The main advantage of these types of magnet is their flexibility. They are light in weight (up to 4 tons for large bore systems) and can be turned off, which removes hazardous conditions when transporting and setting up the process analyzer. Their principal disadvantage is the constant power needed to generate the field and the requirements for cooling the coils (water). The operating costs of an electromagnetic are higher than for permanent magnets and the performance is essentially equivalent, hence permanent magnets are the preferred choice.

NMR requires a remarkably homogeneous magnetic field. Field homogeneity is usually measured in parts per million of the main magnetic field. Corrections for the main magnetic field nonhomogeneity are generally done passively (placing metallic or magnetized parts in the inner bore of the magnet) and actively by means of conducting coils called shims. All magnets need a more or less sophisticated package of shim coils to correct for field nonhomogeneities in the region of interest.

In general, distortion of the magnetic field may arise whenever magnetic susceptibility contrasts are present. Susceptibility contrasts can arise from external constituents (probes and parts dedicated to the process application in or close to the measurement region) or from the sample itself. The heterogeneity of the sample can cause multiple artifacts, such as line broadening and image distortion. These effects are generally very slight at the field strengths used in permanent magnets and are not generally important in process control applications.

6.2 Gradient Coils

Gradient coils are designed to create gradients in the main magnetic field to encode local spatial information in the NMR signal. The most common application is imaging which requires a set of three gradient coils to encode all directions in space. Similarly magnetic field gradients are used for diffusion and flow measurements encoding the displacement of molecules. Gradients are also used in a variety of ways in high resolution spectroscopy. These coils are made of large and thick copper conductors which carry intense electric currents (up to 50 A) to reach values of gradient up to 1500 G cm^{-1}.[36] Although standard imaging techniques do not require such large values, current in the coils must switch on and off rapidly, generally in less than 100 μs. Some applications like unidirectional processes (i.e. pipe flow) may require only one high performance gradient coil. Parameters such as linearity, gradient strength, gradient shielding, rise time, and acoustic noise must be optimized depending upon the target application. Various technical issues and performance parameter optimization are thoroughly discussed by Doty.[36]

6.3 Radio-frequency Coils

The rf coil is basically a RLC (resonant circuit) resonator tuned to the resonance frequency, (Larmor frequency) of the nuclei under test. NMR rf coils are designed to confine their field and energy density to a defined volume. During the NMR experiment high power rf pulses have to be transmitted to the sample while weak NMR signals from the sample have to be detected with a sensitive receiver. The NMR probes must have their field perpendicular to the main magnetic field, constraining the possibilities of probe design.

The static magnetic field strength, the electronics utilized in the spectrometer and the sample itself dictate the SNR achievable for a particular application. It is the design and construction of an adequate rf coil that is critical in achieving that dictated SNR. In this sense, the quality of the transmitting and receiving rf coils determines to a large extent the performance of the NMR apparatus. An extensive literature exists on coil theory, design and construction. More detail concerning designs presented in this section can be found in the review on rf coils by Link.[37] rf Coils can be roughly separated in two categories: volume coils and surface coils.

Volume coils are wound around the sample and have the ability to excite and image the whole volume of interest. One of the most popular and efficient probes is the solenoid and its optimized variants[38] (Figure 12a). Unfortunately this design often conflicts with geometric constraints (sample size, magnets with axial main field) and high frequency applications (its high inductance make this design nonoptimal for some applications over 100 MHz). Several other designs have been proposed to compensate for these limitations.

A successful and easy-to-build coil is the optimized slotted tube (also known as Alderman-Grant) resonator which consists of two conductive sheets supported by a cylindrical structure (Figure 12b). More sophisticated designs such as the Bird Cage coils (composed of lumped elements) have become very popular because of their very high field homogeneity.

As their name indicates, surface coils are positioned as close as possible to the sample to ensure high sensitivity (Figure 13). Their utilization is generally combined with a volume coil that is used to excite the sample homogeneously while the surface coil is used as a local

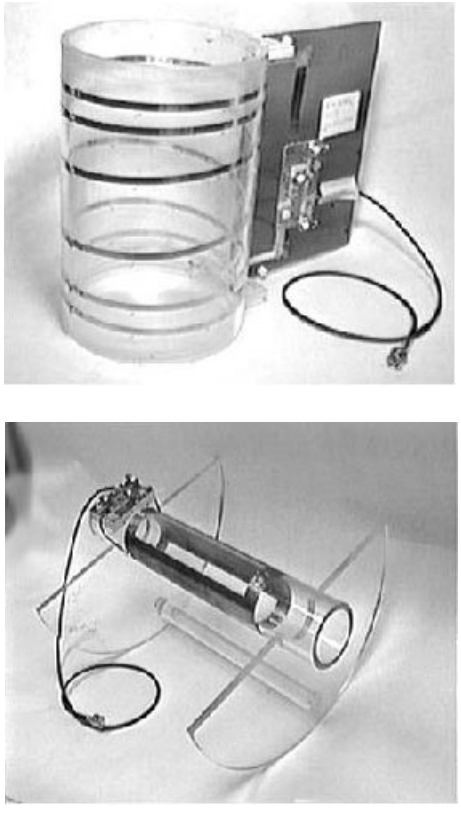

(a)

(b)

Figure 12 rf coils, shown without shields, used for NMR or MRI; (a) solenoid, (b) Alderman-Grant.

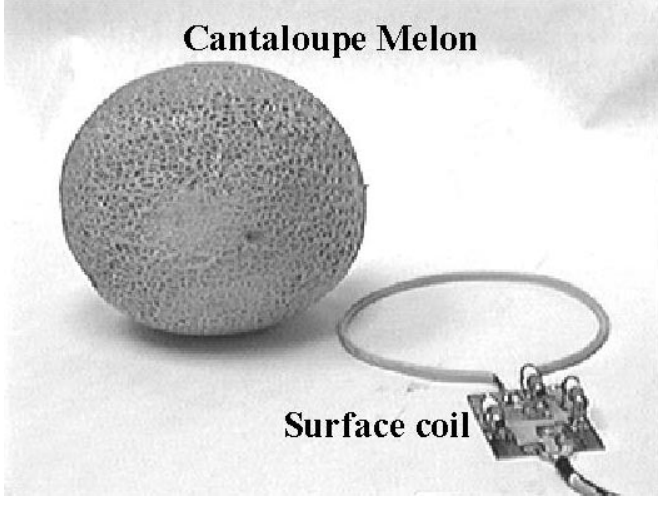

Figure 13 Simple surface coil construction.

sensitive receiver. Although they have an outstanding sensitivity, this type of device suffers from limited field of view and poor field homogeneity. The development of phased array coils (combination of several surface coils) has greatly improved and extended the applications of surface coils.

Significant effort has been devoted to the development of rf superconducting devices. Owing to their low resistivity those devices should increase the SNR of the experiments substantially. Although to date only prototypes have been presented it can be expected that superconducting rf coils will gain importance in the near future.

6.4 Process Modifications

6.4.1 Process Design

It is imperative that the design of the process to be monitored is fully compatible with the NMR experimental constraints, which means basically the use of nonmagnetic parts. Figure 6 shows an example of a fruit conveyor designed and constructed to monitor fruit quality on-line. Common materials used are glass, plastics, and nonmagnetic metals such as copper, aluminum or stainless steel. Since rf fields have to penetrate the sample, the container or holder must be made of nonconductive material such as glass, plastic or ceramic. For low-pressure and low-temperature applications glass-lined probes are adequate.[8] As the temperatures and pressures increase, Teflon®, Kel-F®, or other nonmetallic materials are required. Monitoring canned (iron or aluminum) products or studying the flow in metallic pipes is thus unrealizable unless the rf and gradient coils are placed inside the metallic structure and the metallic structure is nonmagnetic.

Properties of the sample must also be considered when developing a sampling system.[8] For proper loading of the sample narrow bore tubes with diameters between 3–10 m are adequate for low-viscosity fluids. For higher viscosity fluids, fluids with particulates and solid samples, larger tube diameters or piston mechanisms may be required.[8,9] Filters may be required upstream of the magnet to prevent the deposition of magnetic materials from the process in the rf probe and magnet assembly.

6.4.2 Site and Environment

The site and environment of the process analyzer should be reviewed carefully. The first consideration is to study the effect that the fringe magnetic field (magnetic field outside the magnet) may have on devices and activities in surrounding areas. The influence of fringe fields is best controlled by design of the magnet/sensor system. Reciprocally, the effects of devices and activities that could potentially affect the NMR system operation must be quantified. The system must also be isolated from unstable power lines or rf sources that operate at similar frequencies.

6.4.3 Data Sampling, Acquisition and Processing

Most NMR instruments for process analysis are PC-based and offer flexibility and expandability. High-speed digitizing boards (usually up to 10 MHz 12-bit dual-channel) are readily available and cover the majority of applications for high speed NMR. In real-time applications, digital signal processing (DSP) boards permit hundreds of different complex operations at the ''hardware'' level. Processing routines such as two-and three-dimensional FFTs (fast Fourier transforms), binarizing, edge detection, and convolutions are readily available with high-performance independent processors substantially decreasing the computation time. Comprehensive software is furnished with any NMR system and specific applications can be developed with multipurpose software such as StatView®, Matlab® or IDL®. Applications of chemometric techniques for model building have been demonstrated to improve the performance of the sensor systems.[20,21]

6.5 Basic Troubleshooting

When the SNR of the NMR experiment falls below an acceptable point or no signal is detected, basic items can easily be tested to isolate the problem. A basic ''troubleshooting chart'' has been developed by Traficante[39] and a more detailed discussion is given by Fukushima and Roeder.[40]

First, a standard sample should be inserted into the system and the signal acquired. Next, all the cables and connections that are accessible should be inspected and tested. It is also important to note that any changes in the environment (electronic/mechanic equipment) can interfere with a NMR experiment. After external possibilities are tested, the field homogeneity should be checked. A simple one-pulse experiment will provide a good idea of the field homogeneity and indicate if shimming is needed. It is useful to examine the Fourier transform of such signal (it can be only noise) to determine if external signals are affecting the measurement. External noise can be due to improper shielding of the system. If these steps do not isolate the problem, the rf coil should be tested. The coil must be properly matched (usually to 50 Ω) and tuned to the right frequency: this procedure is always accessible and provided by the system manufacturer. If suitable tuning or matching cannot be obtained, the rf coil may require servicing. If the problem is still not apparent, the high-power performance of the probe must be tested. Positioning a small pick-up coil (single loop) in the vicinity of the rf coil and monitoring the pulse with an oscilloscope, during high power operation tests the rf probe performance. If the pulse collapses and a nonrectangular shape is observed, it is possible that the rf probe needs maintenance (e.g. faulty or dirty capacitors, defective soldering points). However, the high power needs to be checked separately (connecting it to an oscilloscope through an appropriate attenuator). If all the points above have been checked and the problem persists the receiver network should be examined.

7 METHOD DEVELOPMENT

In the development of NMR/MRI methods for process analysis, several features of the sample and measurement interaction should be addressed. In NMR/MRI measurements, the signal is a voltage acquired in time from a specific volume. The S/N can be influenced significantly by placing a sample in different regions of the volume or partially outside the volume. When comparing amounts of different materials in NMR spectra, the integration of the peak areas needs to be converted for differing molecular masses and variations in density. Consider the measurement of the percentage of ethanol in whiskey.[41] Integration of the OH and CH_2 peaks and correction for the OH resonance from the ethanol hydroxy proton yields a percentage ethanol that is lower than the expected value. Additional adjustments to the integration values for the mole ratio and density corrects the calculation and the expected result is obtained (40% v/v).[41]

The temperature of samples should be controlled and be reproducible. Relaxation times are often strong functions of both temperature and composition and signal intensity is inversely proportional to temperature. Changes in composition and processing conditions and the influence of these changes should be incorporated into the analysis.

ACKNOWLEDGMENTS

We appreciate the helpful discussions and comments from Dr T. Skloss, Dr I. Pykett, Dr C. Nicholls and Mr Vaughn Davis. The assistance of Ms S. Sadikin and Ms E. Northheimer in preparation of the figures and text is gratefully appreciated.

ABBREVIATIONS AND ACRONYMS

CPMG	Carr–Purcell–Meiboom–Gill
DSP	Digital Signal Processing
FFTs	Fast Fourier Transforms
FID	Free Induction Decay
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
PFG	Pulsed Field Gradient

rf	Radio-frequency
RLC	Resonant Circuit
S/N	Signal-to-Noise Ratio
TMS	Tetramethylsilane

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Imaging, Functional

Chemical Weapons Chemicals Analysis **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation

REFERENCES

1. M.T. Riebe, D.J. Eustace, 'Process Analytical Chemistry an Industrial Perspective', *Anal. Chem.*, **62**, 65A–71A (1990).
2. J.B. Callis, D.L. Illman, B.R. Kowalski, 'Process Analytical Chemistry', *Anal. Chem.*, **59**, 624A–637A (1987).
3. K.R. Beebe, W.W. Blaser, R.A. Bredeweg, J.P. Chauvel, Jr, R.S. Harner, M. LaPack, A. Leugers, D.P. Martin, L.G. Wright, E.D. Yalvac, 'Process Analytical Chemistry', *Anal. Chem.*, **65**, 199R–216R (1993).
4. T.W. Skloss, A.J. Kim, J.F. Haw, 'High Resolution NMR Process Analyzer for Oxygenates in Gasoline', *Anal. Chem.*, **66**(4), 536–542 (1994).
5. A. Abragham, *Principles of Nuclear Magnetism*, Oxford University Press, New York, 1961.
6. C. Tellier, F. Mariette, 'On-line Applications in Food Science', in *Annual Reports on NMR Spectroscopy*, eds. G.A. Webb, P.S. Belton, M.J. McCarthy, Academic Press, 105–122, Vol. 31, 1995.
7. F.A. Nelson, C.A. Reilly, W.E. Savage, 'Process Monitor using High Resolution Nuclear Magnetic Resonance', *Ind. Eng. Chem.*, **52**(6), 487–484 (1960).
8. M.L. Snoddy, 'The Potential of Process NMR on Flowing Streams', *Spectroscopy*, **8**(3), 41–47 (1993).
9. P.J. McDonald, 'The Use of Nuclear Magnetic Resonance of On Line Process Control and Quality Assurance', in *Food Processing Recent Developments*, ed. A.G. Gaonkar, Elsevier, Oxford, 23–36, 1995.
10. C. Nicholls, 'A Hydrogen Transient Nuclear Magnetic Resonance Sensor for Industrial Drying Applications', Proceedings Sensors Expo West, March 13–15, Long Beach California, Helmers Publishing, Peterborough, 201A-1–201A-9, 1990.
11. C.I. Nicholls, A. De Los Santos, 'Hydrogen Transient Nuclear Magnetic Resonance for Industrial Moisture Sensing', *Drying Tech.*, **9**(4), 849–873 (1991).
12. D.E. Podkulski, 'How Do New Process Analyzers Measure Up?', *Chem. Eng. Progr.*, **93**(10), 33–46 (1997).
13. P.J. Barker, H.J. Stronks, 'Application of the Low Resolution Pulsed NMR 'MINISPEC' to Analytical Problems in the Food and Agriculture Industries', in *NMR Applications in Biopolymers*, eds. J.W. Finley, S.J. Schmidt, A.S. Serianni, Plenum Press, New York, 481–498, 1990.
14. E.O. Stejskal, J.E. Tanner, 'Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-dependent Field Gradient', *J. Chem. Phys.*, **42**, 4938–4945 (1968).
15. P.T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Clarendon Press, Oxford, 157–166, 1991.
16. K.J. Packer, C. Rees, 'Pulsed NMR Studies of Restricted Diffusion. I. Droplet Size Distributions in Emulsions', *J. Colloid Interface Sci.*, **40**, 206–218 (1972).
17. R.M. Pearson, T.L. Parker, 'The Use of Small Nuclear Magnetic Resonance Spectrometers as On-line Analyzers for Rotary Kiln Control', *Light Metals*, 81–97 (1984).
18. R.M. Pearson, L.R. Adams, C. Job, 'The Use of Small Nuclear Magnetic Resonance Spectrometers for On-line Process Control', *Cereal Foods World*, **32**(11), 822 (1987).
19. R.M. Pearson, J.Q. Adams, 'Automatic Use of Small Nuclear Magnetic Resonance Spectrometers for Quality Control Measurements', in *NMR Applications in Biopolymers*, eds. J.W. Finley, S.J. Schmidt, A.S. Serianni, Plenum Press, New York, 499–509, 1990.
20. D.N. Rutledge, A.S. Barros, Gaudard, 'ANOVA and Factor Analysis Applied to Time Domain NMR Signals', *Magn. Reson. Chem.*, **35**, S13-S21 (1997).
21. A. Gerbanowski, D.N. Rutledge, M.H. Feinberg, C.J. Ducauze, 'Multivariate Regression Applied to Time Domain-Nuclear Magnetic Resonance Signals: Determination of Moisture in Meat Products', *Sci. Des Aliments*, **17**, 309–323 (1997).
22. B. Zion, P. Chen, M.J. McCarthy, 'Nondestructive Quality Evaluation of Fresh Prunes by NMR Spectroscopy', *J. Sci. Food Agric.*, **67**, 423–429 (1995).

23. P. Chen, M.J. McCarthy, S.M. Kim, B. Zion, 'Development of a High-speed NMR Technique for Sensing Maturity of Avocados', *Trans. ASAE*, **39**(6), 2205–2209 (1996).
24. R.L. Kleinberg, C. Flaun, 'Review: NMR Detection and Characterization of Hydrocarbons in Subsurface Earth Formations', in *Spatially Resolved Magnetic Resonance: Methods, Materials, Medicine, Biology, Rheology, Geology, Ecology, Hardware,*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, New York, 555–573, 1998.
25. N. Bloembergen, E.M. Purcell, R.V. Pound, 'Relaxation Effects in Nuclear Magnetic Resonance Absorption', *Phys. Rev.*, **73**, 679–718 (1948).
26. M. Turney, M.-K. Cheung, R.L. Powell, M.J. McCarthy, 'Hindered Settling of Rod-like Particles Measured with Magnetic Resonance Imaging', *AIChE J.*, **41**, 251–257 (1995).
27. M.J. McCarthy, *Magnetic Resonance Imaging in Foods*, Chapman & Hall, New York, 1994.
28. S.J. Gibbs, D. Xing, T.S. Carpenter, L.D. Hall, 'NMR Flow Imaging of Aqueous Polysaccharide Solutions', *J. Rheol.*, **38**, 1757–1767 (1994).
29. R.L. Powell, J.E. Maneval, J.D. Seymour, K.L. McCarthy, M.J. McCarthy, 'Nuclear Magnetic Resonance Imaging for Viscosity Measurements', *J. Rheol.*, **38**, 1465–1470 (1994).
30. S.J. Gibbs, K.L. James, L.D. Hall, D.E. Haycock, W.J. Frith, S. Ablett, 'Rheometry and Detection of Apparent Wall Slip for Poiseuille Flow of Polymer Solutions and Particulate Dispersions by Nuclear Magnetic Resonance Velocimetry', *J. Rheol.*, **40**(N3), 425–440 (1996).
31. K.L. McCarthy, M.J. McCarthy, 'Yield Stress Measurements of Tomato Concentrates using Magnetic Resonance Imaging', in *Advance in Food Engineering*, eds. G. Narsimhan, M.R. Okos, S. Lombardo, Purdue Research Foundation, West Lafayette, 86–90, 1995.
32. P. Chen, M.J. McCarthy, R. Kauten, 'NMR for Internal Quality Evaluation of Fruits and Vegetables', *Trans. ASAE*, **32**(5), 1747–1753 (1989).
33. B. Zion, M.J. McCarthy, P. Chen, 'Real-time Detection of Pits in Processed Cherries by Magnetic Resonance Projections', *Lebensmi.–Wiss. Technol.*, **27**(4), 457–462 (1994).
34. J. Hennig, H. Friedbourg, 'Fat and Water Separation at 0.23 T Using Simultaneous Shift Selective Imaging', *Magn. Reson. Med.*, **3**, 844–848 (1986).
35. D.D. Laukien, W.H. Tschopp, 'Superconducting NMR Magnet Design', *Concepts Magn. Reson.*, **6**, 255–273 (1993).
36. F.D. Doty, 'MRI Gradient Coil Optimization', in *Spatially Resolved Magnetic Resonance: Methods, Materials, Medicine, Biology, Rheology, Geology, Ecology, Hardware*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, New York, 647–674, 1998.
37. J. Link, 'The Design of Resonator Probes with Homogeneous Radiofrequency Fields', *NMR Basic Principles and Progress*, Springer-Verlag, Berlin Heidelberg, Vol. 26, 1992.
38. S. Idziak, U. Haeberlen, 'Design and Construction of a High Homogeneity rf Coil for Solid-state Multiple Pulse NMR', *J. Magn. Reson.*, **50**, 281–288 (1982).
39. D.D. Traficante, 'Elementary Trouble-shooting for Poor Signal-to-Noise Ratios of NMR Spectrometers', *Concepts Magn. Reson.*, **2**, 63–67 (1990).
40. E. Fukushima, S.B.W. Roeder, *Experimental Pulse NMR, A Nuts and Bolts Approach*, Addison-Wesley, Reading, MA, 1981.
41. D.A.R. Williams, *Nuclear Magnetic Resonance Spectroscopy*, John Wiley & Sons, New York, 1986.

Raman Spectroscopy in Process Analysis

M. Anne Leugers
Dow Chemical Co., Midland, USA

Elmer D. Lipp
Dow Corning Corp., Midland, USA

1 Introduction 1
2 Analyzer Components 2
2.1 Probes 2
2.2 Lasers and Optical Fibers 3
2.3 Spectrographs 3
2.4 Detectors 4
3 Raman as a Process Analytical Tool 5
3.1 Strengths 5
3.2 Weaknesses 5
4 Applications 5
4.1 Distillation Streams 5
4.2 PCl_3 Production 6
4.3 Petrochemicals 6
4.4 Syndiotactic Polystyrene Production 7
5 Summary 7
Abbreviations and Acronyms 8
Related Articles 8
References 8

Raman spectroscopy is a laser light scattering measurement in which changes in the frequency of the incident light as a result of scattering from a sample are measured. Raman spectroscopy is generally used to measure vibrational energy levels of molecules and is complementary in that regard to infrared (IR) spectroscopy which measures vibrational energy levels by the absorbance of IR radiation. Although more difficult in the past to perform than IR spectroscopy, recent technological changes have made Raman spectra easier to measure and with better sensitivity. Currently, Raman spectroscopy has about the same general sensitivity as IR spectroscopy and, in many instances, much simpler and more flexible sampling requirements. Raman spectroscopy is particularly useful for molecules containing some form of unsaturation, various inorganic species, halogenated materials, and homonuclear diatomic gases (such as H_2, N_2, and O_2) which are transparent in the IR.

Increased application of Raman spectroscopy for a variety of problems has led to its use for on-line measurements in industrial manufacturing processes. These applications generally involve remote placement of the spectrometer from the process and the use of fiber optics to take light to the sampling points in the process and return scattered light to the spectrometer for analysis. The use of fiber optics is one of the strengths of Raman spectroscopy for these applications. An additional strength is in the probe designs for focusing the laser light and collecting the scattered light at the sampling point. Probes have been designed which are insertable into processes and can withstand high temperature and pressure, or which can operate from a distance through a viewing window, thus completely avoiding physical contact with the sample. This allows Raman spectroscopy to function in difficult environments where other techniques cannot gain access. Sampling intervals with Raman spectroscopy are generally in the seconds to minutes time regime, well within the range needed for meaningful process control.

The disadvantages of Raman spectroscopy often stem from absorbance of the incident laser light by the sample or impurities in the sample. This can lead to sample decomposition or burning if the absorbance is significant due to the amount of power under the focused laser light. Less severe absorbance can lead to fluorescence (absorbance followed by emission) which can easily produce enough light to obscure the Raman scattering. Some samples may not have distinct enough spectral information for the problem at hand and Raman would be ruled out on that basis.

Raman will compete against IR (both mid- and near-) for many process applications. Mid-IR offers a high level of spectral information, but poor options for fiber optics. Near-IR has much less spectral information for most problems, but good choices for fiber optics. Raman offers the best of both IR approaches, having a high level of spectral information combined with good choices for fiber optics. Specific problems need to be evaluated on an individual basis as to whether Raman, mid-IR, or near-IR offers the best approach using knowledge about the spectral information present by each technique and the sampling options for on-line monitoring of the process.

1 INTRODUCTION

Raman spectroscopy is a laser light scattering measurement which provides information about the molecular structure of a material. Raman scattering refers to the portion of the scattered light which has undergone a wavelength shift from the wavelength of the incident radiation. The energy associated with the wavelength shift corresponds to electronic, vibrational, or rotational energy levels in the molecule. The most common use of Raman

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

spectroscopy is to measure energy shifts associated with vibrational energy levels, although rotational transitions for small molecules such as H_2 are also routinely measured. Raman spectroscopy is useful for determining the chemical composition and certain physical properties of samples for both qualitative and quantitative information, and often provides additional or complementary information to its more common counterpart, IR spectroscopy.

Industrial applications of Raman spectroscopy are relatively sparse when compared to IR spectroscopy. For many years Raman spectroscopy was difficult to perform and suffered from low sensitivity. New technologies such as reliable large-frame laser systems, improved optical filtering designs, more sensitive detectors, and fiber optic sampling probes converged in the early and mid-1990s to make Raman spectroscopy easier to perform and with much better sensitivity. This led to a dramatic increase in the use of Raman spectroscopy in industrial laboratories,[1] which in turn led to the development of Raman spectroscopy for direct on-line analysis of industrial manufacturing processes. Recent references containing material discussing Raman as a process analytical tool include a review article on this topic,[2] a biannual review of Raman spectroscopy,[3] and a book about applications of Raman spectroscopy.[1] It is no exaggeration to say that Raman spectroscopy is now a burgeoning technique which may soon become as widely dispersed as IR spectroscopy.

2 ANALYZER COMPONENTS

The typical design of a fiber optic Raman analyzer is shown in Figure 1. The major components are a laser, fiber optic cables and probe, spectrograph, and detector. The fiber optic cables and probe are an especially important part of the analyzer as they both deliver light to the sample and collect scattered light for analysis. All probes currently in use have a 180° backscattering geometry. A review article on this subject[4] describes many different probe designs. The spectrograph eliminates the residual Rayleigh scattering and disperses the Raman scattering by wavelength across the charge-coupled device (CCD) array detector. A computer then processes the data and displays spectral or other output.

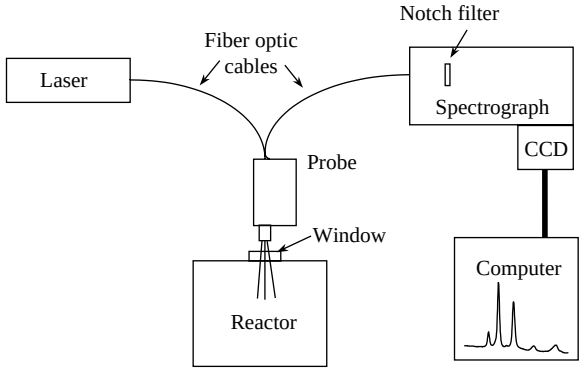

Figure 1 Diagram of components in a process Raman analyzer.

2.1 Probes

Fiber optic probes for Raman spectroscopy are referred to as filtered or unfiltered depending upon whether an optical filter is present within the probe. The earliest designs were unfiltered, and one of the first used a single fiber for both transmitting and collecting light.[5] Despite the complete overlap between excitation and collection fields of view, this design was inadequate for most applications. Sensitivity was low because the single collection fiber gathered too little of the Raman scattered light, and the spectra were fraught with silica artifacts resulting from laser light reflected back into the collection fiber. Therefore, a multifiber probe with different fibers performing the functions of transmitting and collecting, and with some provision for removing silica bands, must be used.

Multifiber probes without filters are frequently referred to as bundle probes. These probes must be inserted into the sample being measured. They are very effective with clear liquids but are generally not useful for solids or liquids which are highly scattering due to the presence of suspended solids or bubbles. A few designs depicting overlap between excitation and collection areas (which ideally is maximized) are shown in Figure 2.

One of the most efficient bundle probe designs consists of a single central illuminating fiber surrounded by a ring of fibers. The central fiber transmits light from a laser to the sample, and the outer fibers collect scattered light and transmit it to the spectrograph. Suppression of silica artifacts is achieved with this type of probe by angling the collection fibers relative to the excitation fiber, and by controlling their spacing. Figure 2(c) shows such a design which was patented by the Dow Chemical Company.[6] The actual probe has six collecting fibers angled around the central fiber. They are fixed in place within a rugged tube and a silica or quartz window tightly seals the end. The Dow Chemical probe has a history of successful process applications, and this includes conditions of high temperature, high pressure, and corrosion.

Another type of bundle probe was developed by Schwab and McCreery[7] who used a close-packed bundle of optical fibers of the same diameter. Bundles of six or 18 fibers naturally form a close-packed cubic geometry and, as such, this type of probe is quite easy to fabricate. By not spacing the fibers, however, the spectra are quite susceptible to silica artifacts from reflected light. This

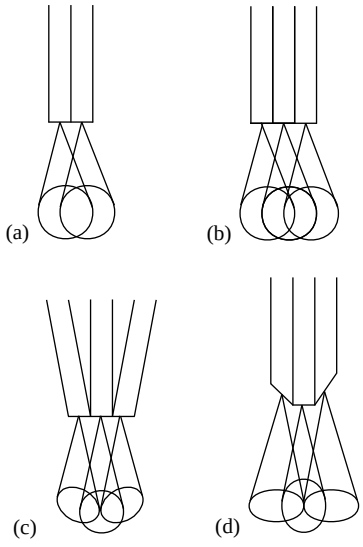

Figure 2 Diagram of some common unfiltered probe designs showing overlap between excitation and collection areas: (a) one excitation fiber, one collection fiber; (b) one central excitation fiber, several collection fibers; (c) collection fibers angled to increase overlap and suppress silica artifacts; (d) collection fibers beveled to suppress silica artifacts.

prevents use of windows in front of the fibers except for very thin films which are thinner than the effective field of view. For short laboratory analyses, these probes worked quite well with a polymer film such as Saran Wrap® in front of the fibers. As such, they are unsuitable for process applications.

A third style of bundle probe, developed by O'Rourke and Livingston,[8] also employs six collecting fibers surrounding the central illuminating fiber. The axes of all of the fibers are parallel but the faces of the collecting fibers at the distal end of the probe are polished into a flattened conical arrangement (Figure 2d). This type of collecting fiber geometry allows rejection of reflected light from the window–sample interface and, like the Dow Chemical design, the end can be sealed with a silica or quartz window.

Filtered probes allow complete rejection of Raman silica bands from the fibers. These probes are useful with clear liquids, opaque liquids, solids, and gases, and do not require insertion into a sample but can operate from a distance. However, they are more expensive, are sensitive to alignment of the optical components (and thus temperature), and are somewhat lower in Raman scattering efficiency for clear liquids.

The first filtered probe design was developed by Carraba and Rauh[9] and is shown in Figure 3. The probe employs two fibers, one for transmitting and one for collecting. When the laser light enters the probe it is collimated and filtered with a laser bandpass filter to remove the silica Raman scattering generated in

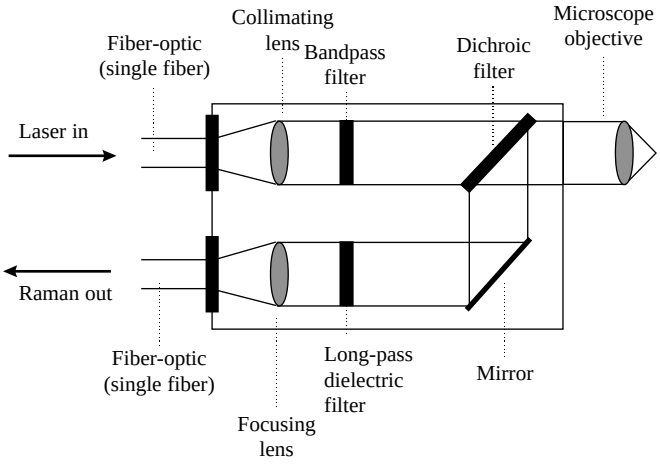

Figure 3 Diagram of filtered probe design. (Adapted from Carrabba and Rauh.[9])

the fiber length. The light strikes the sample and the scattered radiation is collimated, filtered to remove the Rayleigh line (which would excite Raman scattering in the collection fiber), and refocused into the collection fiber. This style of probe provides complete overlap of the excitation and collection fields of view, and this helps to compensate for transmission losses from the numerous optical components. Dilor developed a very similar probe geometry which was also designed to be noncontacting. The lens used at the tip of the probe to focus and collect radiation was of a sufficiently long focal length to be some distance (up to several inches) from the sample. Kaiser Optical has a probe with similar geometry as well. Currently, filtered probes such as these are the most common choice for industrial process applications.

2.2 Lasers and Optical Fibers

Three types of lasers are typically used for on-line fiber optic Raman systems. These lasers are small, rugged, air-cooled, and operate at 115 V. Except for the helium–neon laser, the cost per milliwatt for these portable lasers is quite high relative to large-frame ion lasers. However, large-frame lasers operating at 220 V, 30 A, and 3 gal min^{-1} of cooling water are impractical for on-line instruments. These portable laser systems all have relative strengths and weaknesses as summarized in Table 1.

2.3 Spectrographs

The three main types of spectrographs used in process Raman analyzers are axial transmissive, Czerny–Turner, and Echelle. The strengths and weaknesses of these spectrograph designs are summarized in Table 2. The holographic filter technology, which is part of the axial transmissive design from Kaiser Optical, is the industry

Table 1 Comparison of laser characteristics

Laser system	Advantages	Disadvantages
532 nm (frequency-doubled diode-pumped YAG)[a]	good QE for CCD detector (532–676 nm), roughly 20–33%	fluorescence can be significant
	relatively high Raman scattering efficiency due to shorter wavelength	some attenuation through long lengths of optical fibers
	high-power lasers available	high cost per milliwatt
	good long-term reliability	
633 nm (helium–neon)	good QE for CCD detector (633–848 nm), roughly 40–50%	some fluorescence still observed at this wavelength
	very long lifetime, very reliable	only low-power lasers available
	little attenuation through optical fibers	
	low cost per milliwatt	
785 nm diode laser	low fluorescence	relatively low Raman scattering efficiency
		short lifetimes, 4–6 months of continuous operation
		QE of CCD detector varies significantly over range, 40–4%
		high cost per milliwatt
		mode stabilization critical to prevent mode-hopping

[a] YAG, yttrium aluminum garnate; QE, quantum efficiency.

Table 2 Comparison of spectrograph designs

Spectrograph design	Advantages	Disadvantages
Axial transmissive (Kaiser)	high throughput, up to 80% transmission at center of grating range	transmission at edges of grating range falls to roughly 20%
	multiplexing possible as spectrum from single fiber is only about 10 pixels high	
	only two regions must be spliced together to achieve the full spectrum	
	relatively high resolution (about 5 cm^{-1}) for 5000 cm^{-1} coverage	
Czerny–Turner (ISA/Dilor)	fairly flat response across grating range	lower transmission efficiency, roughly 40% of transmission
	allows multiplexing as spectrum from single fiber is only about 10 pixels high	lower resolution
	low-density grating allows full spectral coverage with no splicing	
Echelle (EIC Laboratories)	use multiple orders to achieve full 4000 cm^{-1} coverage	multiplexing not possible as full CCD is used for single probe
	can achieve very high resolution	spectrum must be pasted together from hundreds of pixels on CCD

standard and is used by most of the vendors of Raman spectrometers for process applications.

2.4 Detectors

Array detectors have become an important component in Raman spectroscopy. A review article on array detectors for spectroscopy[10] discusses the characteristics of various detectors for spectroscopic measurements. Virtually all commercially available Raman spectrometers employ a silicon CCD detector. The most common type of CCD detector for Raman spectroscopy is the uncoated, front-illuminated CCD detector peaking at roughly 40% quantum efficiency (QE) in the range 650–750 nm. The detectors drop to roughly 1% QE at 1050 nm and 400 nm. Thus, these detectors are quite good for excitation with 500–650 nm laser sources.

One approach to enhancing QE across the visible and into the ultraviolet and near-IR is the use of thinned, back-illuminated CCD detectors. These detectors can range as high as 80% QE at their peak response and can extend the detection range to 200–1100 nm. However, these detectors are still quite expensive and may not be as rugged as the thicker, front-illuminated CCD detectors.

3 RAMAN AS A PROCESS ANALYTICAL TOOL

3.1 Strengths

As a process tool, Raman will compete most frequently against mid-IR and near-IR techniques. The mid-IR generally provides a high level of spectral information, but no good choices for fiber optic probes. The near-IR has good choices for fiber optics, but the spectral information is often limited. Raman combines the best feature of each IR technique by providing a high level of spectral information with good choices for fiber optic probes. An additional advantage of Raman probes is the backscattering collection geometry in the probe tip. There is no gap that the sample must flow through as in mid- and near-IR transmission probes, which is very important for the analysis of viscous polymers and solid slurries. Filtered probes do not need to contact the sample, but may operate from a distance through a sight glass – a very important advantage with corrosive streams. As the information content of Raman spectra is high with individual bands often being baseline resolved (or nearly so), it is possible in many instances to avoid complicated chemometric data handling approaches in favor of much more direct band height or area ratios. These direct methods require much less time and effort to develop and maintain, and they are generally more robust to changes in process conditions which may have occurred since a chemometric calibration was developed.

Modern Raman spectrometers provide sufficient sensitivity for many applications to allow relatively rapid sampling rates, in the seconds to minutes time regime, which is the speed often necessary for meaningful process control. These instruments can be compact and portable, require no special utilities, and are certainly rugged for short-term use. Some on-going applications will provide insight into long-term performance and reliability. Part of the ruggedness of Raman spectrometers is due to the wavenumber stability which results from the presence of only a single moving part – the camera shutter in front of the detector. The other optics in these spectrometers are fixed in position so that they do not move with time, further contributing to wavenumber stability. Another beneficial feature is the ability to multiplex several probes onto a single spectrometer with no moving parts. Most CCD detectors are large enough to be partitioned into separate areas, each of which is large enough to hold an entire spectrum. The laser beam can also be divided into several beams with each beam directed into a separate probe. The Raman scattering from each probe can then be focused onto a portion of the CCD dedicated to that probe. This allows continuous and simultaneous monitoring from several probes with the camera shutter remaining the only moveable part in the system.

3.2 Weaknesses

Compared to mid-IR and near-IR techniques, there are a number of disadvantages as well. Samples that fluoresce are always troublesome. A common approach with fluorescence is to move to longer wavelength lasers, but this sacrifices sensitivity. From a safety standpoint it must always be verified beforehand that a sample will not degrade or burn under the highly focused laser beam. Many of the lasers in use today do not have long lifetimes from a process analysis point of view, so provision must be made for replacing them in a timely fashion. Some important functional groups are weak Raman scatterers, such as OH and NH. Sensitivity among Raman, mid-IR, and near-IR is an issue which must be addressed on a case-by-case basis, as Raman and IR often have complementary spectral features. There is a general perception within the spectroscopic community that Raman is less sensitive than either IR technique. Our feeling is that this perception must be qualified. One reason to avoid dogmatic statements in this area is due to the variety of lasers, detectors, and optical configurations available in Raman spectrometers. This leads to a wide range of sensitivities which can be achieved, and the field is in a rapid state of change with significant new options being introduced each year. This contrasts sharply with mid- and near-IR instruments where components are relatively standardized within each technique and sensitivity among different instruments within each technique is generally comparable. Our experience at this time is that Raman can provide sensitivity which is basically comparable to the mid- and near-IR, provided that instrument components which maximize Raman sensitivity are selected. For good cases, sensitivities in the 100 ppm range are feasible, and 0.1–1% is commonly achieved. Many of the advances being made in Raman instrumentation may swing the sensitivity comparison in favor of Raman in the near future.

4 APPLICATIONS

There is a large body of literature dealing with the remote monitoring capabilities of Raman spectroscopy. However, many of these are evaluations done in laboratory settings and the number of process applications in real industrial environments is much smaller. Some of these have been selected for discussion, with the intention of conveying the breadth of process analysis problems to which Raman spectroscopy can be applied.

4.1 Distillation Streams

Industrial distillation processes are used to purify a wide variety of materials. These separations are often very

energy consuming, and precise control of the distillation conditions is needed to enhance the purity of the final material along with minimizing the energy costs. Small improvements in the performance of these columns can have large monetary impacts. Thus, it is not surprising that some of the earliest efforts at process control by Raman spectroscopy were directed at distillation columns.[11–13] These initial efforts focused on three separations: acetic acid and solvent; isopropanol and water; and cumene and cumene hydroperoxide. The first two applications involved a high concentration of water, and hence were difficult by near-IR, whereas the third involved detection of a peroxide group which was basically IR inactive. An insertable probe was used for these measurements, and liquid samples were measured from particular points in the distillation streams. This work demonstrated that Raman could provide sufficient sensitivity and time resolution for meaningful process control of these separations.

Another application of on-line Raman spectroscopy is for measuring the composition of methylchlorosilane distillation streams.[14] These streams consist of various mixtures of methyl- and hydrogen-containing chlorosilane monomers which are being separated by distillation for later use in the production of silicone polymers. Near-IR measurement of these streams was not feasible because several of the species of interest would have been weak or transparent. Raman monitoring of these streams used the Dow Chemical insertable probe[6] which could withstand the corrosive environment caused by residual HCl. The use of Raman in this example takes advantage of the high information content of the spectra along with the capability for faster time resolution relative to current monitoring by gas chromatography. The nine species of interest all have strong and distinct Raman bands from SiCl and SiH moieties, and the data was collected at 5 min intervals. Figure 4 shows the concentration profile with time for a two-component stream where large changes in stream composition are followed in very great detail.

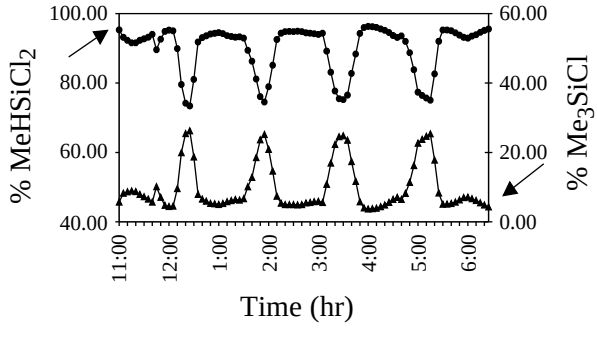

Figure 4 Concentrations in a two-component methylchlorosilane stream showing how the two species fluctuate with time.

4.2 PCl_3 Production

An on-line Raman spectrometer has been used to monitor various species during PCl_3 production[15] where P_4 and Cl_2 gas are mixed in a closed reactor. PCl_5 is an additional by-product. This problem could not be addressed by near-IR because all the species of interest are basically transparent in that region. Raman monitoring of this corrosive and hazardous mixture occurred with a probe operating from a distance through a viewing window, thus taking advantage of the ability of Raman spectroscopy to sample noninvasively. Accurately measuring the reactor contents has two important roles in the synthesis of PCl_3. First, ensuring that excess P_4 is present allows efficient control over the reaction rate as this rate can then be controlled by the Cl_2 feed rate. Second, PCl_5 levels can be monitored. A build-up of this material is to be avoided as it can react with P_4 and interfere with the PCl_3 production. This process had been previously monitored[16] with a Fourier transform Raman spectrometer, but was subsequently changed to a visible laser–CCD system to achieve both higher sensitivity and lower operating costs.

4.3 Petrochemicals

Raman monitoring of petrochemicals and fuels has received much attention because Raman has a generally high sensitivity for aromatic species, and the spectra of related species such as xylene isomers often show distinctive differences. A chromatographic process to separate *p*-xylene from other aromatic and nonaromatic species has been monitored on-line at four different points simultaneously.[17] One spectrometer and four fiber optic probes were used for this application. The laser beam was divided into four parts using a beam splitter, and each of the four beams was then directed into a separate fiber optic probe. The CCD detector was electronically partitioned into four separate areas, and the Raman scattered light from each probe was directed onto one of these four areas. The combined use of a beamsplitter and partitioning of the CCD allowed one instrument to monitor four different points simultaneously with no moving parts, thus minimizing instrumental drift and producing a very robust instrument for process environments. Raman monitoring of this process was capable of controlling the process by following several chemical species over wide concentration changes.

A number of other chemicals or properties of petroleum fuels have been studied by Raman spectroscopy for potential on-line measurements. Some of these include benzene, octane number, density, and volatility,[18] and a series of papers dealing with oxygen content from ethers and alcohols, benzene and other aromatics, octane number, and Reid vapor pressure.[19–22] Some of these studies were comparisons of Raman with

mid- and near-IR measurements. Raman spectroscopy was found to be a suitable tool for each of these applications, as were mid- and near-IR when they were also studied. Preference for Raman as a process analytical tool in these applications would probably stem from capabilities for noninvasive sampling, multiplexing several streams to one spectrometer, and the higher degree of spectral distinction among some species.

4.4 Syndiotactic Polystyrene Production

An especially difficult on-line application for which Raman spectroscopy has proved successful is for monitoring the production of syndiotactic polystyrene.[23] During the polymerization reaction process, liquid styrene monomer is converted to a free-flowing polystyrene powder which becomes airborne in mechanically agitated reactors. The outlet powder contains significant amounts of unreacted monomer absorbed on or in the particles, and control of this outlet monomer concentration is critical for achieving reliable and efficient performance of the reactor. Spectral measurement of solids suspended in air is a difficult problem because quantitation of results is highly dependent upon the particle size. This problem is more severe with IR transmission measurements than for Raman as a scattering measurement does not require the light to pass completely through the particles. The problem is further exacerbated for transmission measurements when the transparent medium is air, due to the large refractive index mismatch between the particles and air. Consequently, near-IR transmission was not a realistic possibility for monitoring this suspension of solid particles.

In the analysis of styrene to syndiotactic polystyrene on suspended particles, there is a particle size range which allows Raman spectroscopy to be used quantitatively to determine the concentration of liquid styrene dispersed on the surface of the polymer particles. For this particle size range, the Raman signal from the liquid monomer coating relative to that from the underlying solid polymer particles is linear over the operating range of the production reactor. The analytical results are based on measurement of the C=C stretching band at roughly 1640 cm^{-1}, which is present only in the styrene monomer, relative to the 1000 cm^{-1} band due to the aromatic ring breathing mode, which is present in both the monomer and polymer spectra. Representative spectra from this on-line application are shown in Figure 5. These bands are sufficiently resolved to allow a rugged calibration based on peak areas of the bands of interest. In addition to the styrene monomer and polymer concentrations, concentrations of comonomers and copolymers can be determined as well. This process Raman analyzer was successful in achieving the desired results, providing fast analysis (less than 5 min) and good precision (±0.2% relative concentration) for the monomer. Moreover the probe exhibited no fouling in a period of over 1 year of operation and no analyzer drift was observed. The on-line analysis allowed the plant to operate with reduced contamination and labor costs than experienced with grab sampling.

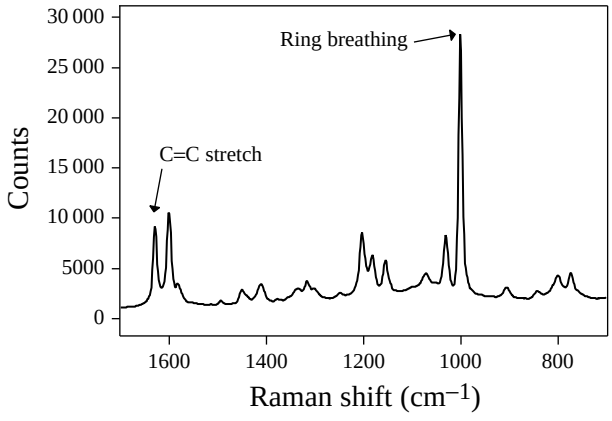

Figure 5 Raman spectrum measured during synthesis of syndiotactic polystyrene. The ratio of the C=C stretch and the ring breathing mode is used to monitor the course of the reaction.

5 SUMMARY

This article demonstrates the type of applications to which fiber optic Raman spectroscopy has been applied for process control. The number of potential applications for Raman spectroscopy in process control applications is really boundless. Because of the backscatter geometry employed in Raman spectroscopy, virtually any type of sample can be analyzed. Thus applications such as polymer fiber orientation, film morphology measurements, polymer crystallinity determinations, and pharmaceutical polymorph concentrations can all be performed on-line just as the more conventional chemical determinations of copolymer and comonomer concentrations. If the measurement or determination has been performed using Raman spectroscopy in the laboratory, the measurement can potentially be performed on-line as well. Many of these types of applications have been developed but cannot be reported for reasons of business propriety.

The technological developments of recent years have transformed Raman spectroscopy from an underutilized technique of last resort to a technique of comparable value and versatility to its more universally accepted counterpart, IR spectroscopy. The instrumentation is significantly more rugged than in the early 1990s and is still changing rapidly. Raman analyzers in standard class I Division II instrument housings are now standard

off-the-shelf units. Multiplexing several probes to the same instrument is now routinely offered as an option and rugged air-cooled solid-state lasers that operate efficiently at relatively low current ($\ll$15 A) are standard and relatively long-lived. The large step-changes that have occurred since the late 1980s in laser technology, CCD technology, and spectrograph/filter technology have allowed this type of analyzer to be built. The advances which need to be realized now are the development of manufacturing practices which result in lower-cost instruments and components.

Unlike earlier revolutions in Raman spectroscopy, this current revolution in fiber optic dispersive Raman spectroscopy has demonstrated significant impact, well beyond the analytical laboratory and within the manufacturing process itself. The technology is still improving and it is likely that the desirability of Raman spectroscopy as an on-line analytical tool will continue to increase in the foreseeable future.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
IR	Infrared
QE	Quantum Efficiency
YAG	Yttrium Aluminum Garnate

RELATED ARTICLES

Pharmaceuticals and Drugs **(Volume 8)**
Vibrational Spectroscopy in Drug Discovery, Development and Production

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Raman Scattering, Fundamentals

REFERENCES

1. M. Pelletier (ed.), *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford, 1999.
2. F. Adar, R. Geiger, J. Noonan, 'Raman Spectroscopy for Process/Quality Control', *Appl. Spectrosc. Rev.*, **32**(1,2), 45–101 (1997).
3. A.L. Lyon, C.D. Keating, A.P. Fox, B.E. Baker, L. He, S.R. Nicewarner, S.P. Mulvaney, M.J. Natan, 'Raman Spectroscopy', *Anal. Chem.*, **70**(12), 341R–361R (1998).
4. I.R. Lewis, P.R. Griffiths, 'Raman Spectrometry with Fiber-optic Sampling', *Appl. Spec.*, **50**(10), 12A–30A (1996).
5. T. Hirschfeld, 'Remote Fiber Fuorimetric Analysis', Dept. of Defense, Defense Programs, 17–21 (date uncertain).
6. R.D. McLachlan, G.L. Jewett, J.C. Evans, 'Fiber-optic Probe for Sensitive Raman Analysis', US patent 4 573 761, 1986.
7. S.D. Schwab, R.L. McCreery, 'Versatile, Efficient Raman Sampling with Fiber Optics', *Anal. Chem.*, **56**(12), 2199–2204 (1984).
8. P.E. O'Rourke, R.R. Livingston, 'Fiber Optic Probe Having Fibers with Endfaces Formed for Improved Coupling Efficiency and Method Using Same', US patent 5 402 508, 1995.
9. M.M. Carrabba, R.D. Rauh, 'Apparatus for Measuring Raman Spectra Over Optical Fibers', US patent 5 112 127, 1992.
10. J.M. Harnly, R.E. Fields, 'Solid-state Array Detectors for Analytical Spectrometry', *Appl. Spectrosc.*, **51**(9), 334A–351A (1997).
11. M.J. Roberts, A.A. Garrison, S.W. Kercel, E.C. Muly, 'Raman Spectroscopy for On-line, Real-time, Multipoint Industrial Chemical Analysis', *Process Control and Quality*, **1**, 281–291 (1991).
12. A.A. Garrison, D.S. Trimble, E.C. Muly, M.J. Roberts, S.W. Kercel, 'On-line Chemical Analysis Applied to Distillation Control: A Status Report', *Am. Lab.*, **22**(3), 24–27 (1990).
13. A.A. Garrison, E.C. Muly, M.J. Roberts, D.S. Trimble, C.F. Moore, 'Raman Spectroscopy for On-line Distillation Process Control', *Adv. Instrum. Control*, **44**(1), 357–363 (1989).
14. E.D. Lipp, R.L. Grosse, 'On-line Monitoring of Chlorosilane Streams by Raman Spectroscopy', *Appl. Spectrosc.*, **52**(1), 42–46 (1998).
15. G.J. Gervasio, M.J. Pelletier, 'On-line Raman Analysis of PCl_3 Reactor Material', *At-Process*, **3**(1,2), 7–11 (1997).
16. J.J. Freeman, D.O. Fisher, G.J. Gervasio, 'FT-Raman On-line Analysis of Phosphorus Trichloride Reactor Material', *Appl. Spectrosc.*, **47**(8), 1115–1122 (1993).
17. G. Marteau, G. Hotier, N. Zanier-Szydlowski, A. Aoufi, F. Cansell, 'Advanced Control of C_8 Aromatics Separation Process with Real-time Multipoint Online Raman Spectroscopy', *Process Control Qual.*, **6**, 133–140 (1994).
18. C.J. de Bakker, P.M. Fredericks, 'Determination of Petroleum Properties by Fiber-optic Fourier Transform Raman Spectrometry and Partial Least-squares Analysis', *Appl. Spectrosc.*, **49**(12), 1766–1771 (1995).
19. J.B. Cooper, K.L. Wise, W.T. Welch, R.R. Bledsoe, M.B. Sumner, 'Determination of Weight Percent Oxygen

in Commercial Gasoline: A Comparison between FT-Raman, FT-IR, and Dispersive Near-IR Spectroscopies', *Appl. Spec.*, **50**(7), 917–921 (1996).

20. J.B. Cooper, K.L. Wise, W.T. Welch, M.B. Sumner, B.K. Wilt, R.R. Bledsoe, 'Comparison of Near-IR, Raman, and Mid-IR Spectroscopies for the Determination of BTEX in Petroleum Fuels', *Appl. Spec.*, **51**(11), 1613–1620 (1997).
21. J.B. Cooper, P.E. Flecher, T.M. Vess, W.T. Welch, 'Remote Fiber-optic Raman Analysis of Xylene Isomers in Mock Petroleum Fuels Using a Low-cost Dispersive Instrument and Partial Least-squares Regression Analysis', *Appl. Spec.*, **49**(5), 586–592 (1995).
22. J.B. Cooper, K.L. Wise, J. Groves, W.T. Welch, 'Determination of Octane Numbers and Reid Vapor Pressure of Commercial Petroleum Fuels Using FT-Raman Spectroscopy and Partial Least-squares Regression Analysis', *Anal. Chem.*, **67**(22), 4096–4100 (1995).
23. A. Leugers, D. Arriola, D. Beaudoin, J. Weston, 'On-line Measurement of Conversion in Syndiotactic Polystyrene Production Using Raman Spectroscopy', 24th FACSS Conference, Providence, RI, 1997.

Sampling and Sample Preparation in Process Analysis

Kenneth E. Creasy and Troy W. Francisco
Honeywell International, Morristown, TN, USA

1 Introduction **1**
1.1 Objectives 1
1.2 Scope of Article 1
1.3 Limits of Article 2
2 Sampling Theory and Statistics **2**
2.1 Process Variability 2
2.2 Sample Size Determination 3
2.3 Sampling Frequency 3
2.4 Sources of Error 4
3 General Considerations **5**
3.1 Sampling System Criteria 5
3.2 Transport Time Requirement 5
3.3 Materials of Construction 6
4 Sampling of Fluids **7**
4.1 Design Considerations 7
4.2 Sampling Modes 9
4.3 Sampling Point Determination 11
4.4 Sample Transfer 11
4.5 Sample Conditioning 12
5 Sampling Mixed Phases **13**
5.1 Classifications 13
5.2 Design Considerations 14
5.3 Sampling Techniques 14
5.4 Sampling Point Determination 15
5.5 Sample Transfer 15
5.6 Sample Conditioning 16
6 Sampling Solids **16**
6.1 Design Considerations 16
6.2 Sampling Techniques 17
6.3 Sample Point Determination 17
6.4 Sample Transfer 18
6.5 Sample Conditioning 18
7 Calibration and Validation of Sampling Systems **18**
7.1 Frequency of Calibration 19
7.2 Long-term Performance 19
8 Maintenance **20**
8.1 Design Criteria 20
8.2 Documenting Maintenance 21
9 Safety **21**
9.1 First and Last Thought 21

Acknowledgments **21**
List of Terms **21**
Abbreviations and Acronyms **22**
Related Articles **22**
References **23**

The purpose of this article is to introduce the reader to the important design characteristics of sampling and sample preparation systems for process analytical measurements, as well as to discuss several of the most commonly used configurations for gas- and liquid-phase process streams. Design criteria essential to the success of any process analyzer installation are discussed at length, and the pitfalls associated with neglecting these items are highlighted. This article provides an overview of what must be considered when building the interface between the analytical instrument and the process stream. It is not intended to be a comprehensive dissertation on the subject or a tutorial. The reader is referred to literature references and textbooks for detailed explanations and relevant design formulae. In addition, we also provide several sections regarding calibration procedures, maintenance concerns, and safety considerations.

1 INTRODUCTION

1.1 Objectives

It is our intention to give the reader an introduction to sampling theory and application for process analysis. Industrial process control depends more on analytical measurements than ever before, and, in order to obtain a maximum value for any analytical effort, the entire analyzer package must give an accurate picture of what is happening in the process. Although there are many well-built, robust analyzers on the market today, the interface between the process and the analytical instrumentation is often neglected and gives rise to most of the dissatisfaction felt by users. This interface, comprising the sampling and conditioning systems, requires attention throughout the conception, design, implementation, and maintenance phases of the process analyzer project. Generally, the requirements for a process that operates near steady state (continuously) are similar to those for a single-run (batch) process, but we will endeavor to differentiate specific needs where appropriate.

1.2 Scope of Article

This article focuses on the theory behind sampling, as well as the major sampling and sample conditioning

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

or preparation techniques used in process analytical chemistry applications. Whenever possible, the reader will be referred to more complete discussions on sampling topics, or to specialized sampling topics outside the scope of this work. For a more comprehensive treatment of the subject, the reader is referred to Carr-Brion and Clarke[1] or to Houser.[2] Additional information may be found in a number of sources specific to process analyzers.[3–7]

1.3 Limits of Article

It is impossible to give the reader a comprehensive understanding of sampling systems and sample preparation in an article of this size. This article is intended to be more of an overview of the current state of the practiced art rather than a treatise on how to design, build, and maintain complex sampling systems. For detailed explanations or analyzer-specific information, please refer to the cited references. Although much effort has been expended in developing in vivo sampling systems,[8–10] these are very specialized systems and as such will not be discussed here.

2 SAMPLING THEORY AND STATISTICS

In order to understand the purpose and function of sampling systems, it is important to review a few statistical concepts. A statistical representation or sampling of the process means that the aliquot of the process stream withdrawn from the bulk matrix will have component species in proportions matching the process stream. In the case of truly representative samples, replicate withdrawals of material from the process stream will all yield the same analytical results (within the statistical variation of the sampling system and instrumentation). As will be pointed out many times in this article, the overall variation in the analytical results is the sum of the variances from the process, sample withdrawal, sample conditioning system, and the process analyzer, collectively known as the measurement system. A block diagram of the measurement system is depicted in Figure 1.

2.1 Process Variability

The primary goal of any sampling system is to capture a representative portion of the process stream for presentation to an analyzer. Frequently, this goal is at odds with conditioning this portion of the process stream and placing it in a suitable form for analysis. Ideally, proportions of species in the captured sample should be identical statistically to the entire process stream at the sampling point. From a practical standpoint this may be neither necessary nor desired – it may be important to exclude portions of the process stream, such as particles

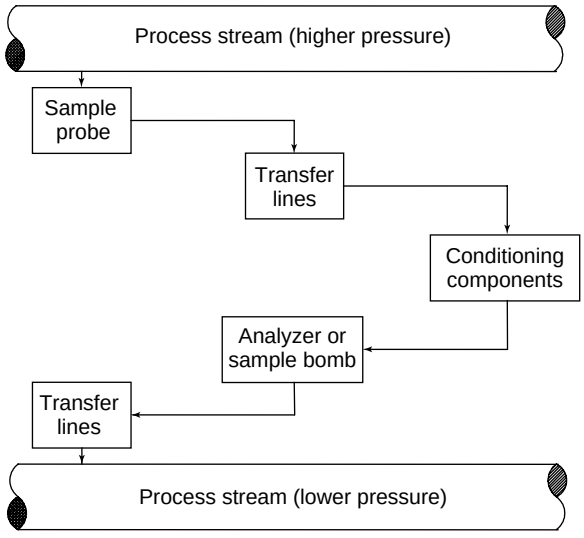

Figure 1 Major components of a process analysis measurement system.

or entrained air, prior to introducing the sample into an analyzer.

By conventional definition, a process that is in statistical control is within three standard deviations of the mean indicating variables. In normally distributed processes, 99.7% of all values will fall within these limits.[11] For process analyzers this means that the measured species or properties in the process stream do not vary by more than three standard deviations within that phase of the process. A more detailed description of statistical control is given in section 7.

The process variability is typically the very reason for having a sampling system. Great care and attention must be paid to the range of this variation and the heterogeneity of the process itself. There could be particle size, solvent, pressure, temperature, or flow variations, which may give rise to heterogeneity in the process stream. Heavier particles, for example, tend to flow along the bottom portions of process piping. Filtration can remove solid particles and maintain the integrity of the gaseous or liquid phases in the process stream. However, if the solid portion needs to be included in the analysis (or worse yet, is the key part of the analysis), such as in detecting particle size or in measuring emissions from smokestacks, disregarding this information would be critical.

The ultimate goal of any sampling system is to provide a statistically representative sample to the process analyzer without introducing any variation from the sampling or conditioning processes. In this manner, the analyzer will respond strictly to the process variation.[12]

Many types of process streams can be sampled. The more heterogeneous the stream, the greater the effort that must be placed in the design and operation

of a sample capturing and conditioning system. The American Society for Testing and Materials (ASTM) has published a standard describing the practical and statistical considerations necessary for sampling industrial chemicals.[13] As already mentioned, streams containing multiple phases may be sampled for process monitoring purposes. Nearly any stream can be sampled, provided that some motive force is available or can be added to extract an aliquot of the bulk material. Additionally, it must be possible to transfer this sample without altering the distribution of species in the sample or changing any of its key properties. This is sometimes quite difficult because the sample frequently continues to react after the sample is processed, and there is no way of quenching such a reaction without altering the components of the sample. In cases such as these, where true sampling is not possible, it is typically better to correlate the analytical results from the treated sample to the true conditions of the process stream.

Before determination of the sampling and analyzing methodologies, it is very important to consider how fast the process changes, and whether there is a pattern to this variation. For example, consider a wastewater stream flowing into a bioremediation vessel.[14,15] Periodically, this wastewater may contain a few litres of highly concentrated solvent that, in a flowing stream, passes a probe for a total of 30 s. A sampling and analysis cycle of 2 min would have a low chance to detect such an upset. If the amount of solvent reaching the wastewater treatment is not sufficient to cause any problems, then concerns are negligible. However, if this amount of solvent causes a major kill of the microorganisms in the pond, the analyzer system would have failed to perform its basic function – to protect the bioremediation pond.

If, on the other hand, sampling occurs far more frequently than the process could ever change, the sampling system is said to be inefficient. Oversampling is not necessarily a poor practice – it becomes so when the effort required to maintain such a system is too great to justify the sampling rate. Sampling should generally occur at a rate sufficient to detect normal rhythmic changes as well as significant upsets in the process stream. From this point of view the sampling must occur at a rate sufficient to detect representative changes in the process stream.

In all cases, it is important to learn as much as possible about the random process variation, as well as the characteristic timescales and magnitudes of process changes that you wish to detect. It is then possible to match the analysis cycle to the process.

2.2 Sample Size Determination

In determining the appropriate sample size, a designer of such systems should always ask two questions: what is the minimum sample that will be representative of the process stream? and what is the minimum sample required by the analyzer portion of the system?

If the process stream, when partitioned to any degree, yields a sample with the exact proportions of the parent stream, the stream content is homogeneous. Any impure material cannot be perfectly homogeneous; that is, if a single molecule could be sampled in such a system, it would not be representative of the whole. Therefore, care must be taken to ensure that the sample is large enough and repeatable enough to represent the entire process stream (on a statistical basis). Many systems are heterogeneous and multiple samples of material from these systems reveal proportional nonrepresentation of the process stream. In order to test whether a stream is homogeneous or heterogeneous, statistical analysis of repetitive samples is typically required.[11,12,16] If the spread of analysis is either nonrandomly distributed or is greater than the expected level within the measurement system (analyzer and operator), we suspect heterogeneity.

Capturing the process variation may be accomplished by taking multiple samples separated temporally or spatially. These may then be analyzed as replicates or as a composite sample. In either case, a more representative picture of the process stream may be developed. Caution should be exercised when drawing conclusions from composite samples, because the true process variation may be averaged out in such samples.

Operator, analyzer, and sampling system constitute the measurement system. The instrument system performs best when the sample stream is effectively homogeneous (the degree of heterogeneity does not contribute a statistically significant error to the measurement system), or when the sampling system provides a true, statistically-valid sample for the analyzer.

If done properly, sample conditioning (filtration, phase separation, heating, cooling, etc.) will not affect the required analysis. Any significant changes in the sample effected by the conditioning system must be scrutinized closely to ensure that the important characteristics of the process sample remain unchanged.

2.3 Sampling Frequency

How often a sample is withdrawn from a process stream defines the sampling frequency. This concept, although straightforward, is often overlooked in the design phase of the instrument system. Commonly, designers will spend a great deal of time and money to place a system in operation that will produce analyses every few seconds, only to discover later that the process is slow to change and the designer would have been better served to maximize the accuracy and precision.

It is always a trade-off between how often we must sample and how much time and effort we are willing to put forth. The bottom line is that we must sample at the frequency appropriate to capture the process stream variation, no matter what the cost, otherwise the analyzer will not provide useful data and will not be successful.

As indicated earlier, one must consider the rate of process change before recommending a measurement system. In order to control any process, feedback must occur on a timescale that is faster than the rate of change within the process stream. This is intuitive; process upsets must be captured within sufficient time to effect a change. Ideally, one must be able to know that an upset has or is about to occur – plant operators have differing opinions on exactly what constitutes knowledge in these cases. With today's analyzers, it is usually possible to sample, condition, and analyze fast enough to give the operator ample warning of the imminence or the beginning of a process upset.

In order to control any process, the analysis must give feedback within a shorter timescale than that in which significant changes are likely to occur. Many operators will not take action until two indications are received to suggest that a process is out of control, or at least progressing toward an out-of-control condition. The response required for batch operations is typically different than that for continuous processes. In the continuous process, species or properties being measured could be related to the final product, to a key control parameter such as pH, or to an impurity in the system. In a batch process, the moiety of interest is commonly that item whose target value specifies the product, or the intermediate material prior to the next production step.

Most sampling systems are continuous in nature; that is, the sample stream or some portion of the process stream is constantly passing through the plumbing that comprises the sampling system. It therefore follows that the sampling system is typically not the slowest portion of the measurement system. If designed correctly, the sampling portion of the instrument system will present to the analyzer a sample with minimal age, or the freshest sample possible. After a portion of the process stream has been removed and placed within the sampling system, changes in temperature, flow, materials of construction, and time itself can cause changes in a representative sample. This is the reason why we wish to minimize the time lags between extracting, conditioning, and analyzing a sample. When a single analyzer is measuring multiple streams, this consideration becomes even more important. Not only must a multiple-point sampling system deliver fresh sample, but it also must prevent cross-contamination between the streams.

When installing a single analyzer for multiple process stream monitoring, one must consider (in addition to the cross-contamination possibilities) the effects that multiplexing will have on the response time. Even when the sample transport system is carefully designed and extremely responsive, overloading a single analytical instrument can have catastrophic effects on its performance.

2.4 Sources of Error

There are many potential sources of error in the sampling portion of any measurement system. Errors may originate from the transfer of sample from the process stream to the container or line outside of the stream, from the transport of the sample to the laboratory or process analyzer, or from the transfer from the container or sample line into the analyzer.[17,18]

When capturing a process sample, the first question to ask is: what are the requirements of the sample? If a sample is to be wholly representative of the process at the sampling point, it must include all the species present in proportions matching the process. If one is interested in only gaseous or liquid components, not only must all other phases be excluded but the sample transfer must not be impeded by the presence of other phases. The sample probes – those devices that accommodate the transfer from the process pipe – must also be designed so that any undesired discrimination does not occur.

When transporting the sample in either a detachable container or through a sample line, leaks into or out of the vessel may occur. One may accept a small amount of leakage if neither partitioning of the sample nor safety or environmental concerns arise. It is rare that both conditions would be met.

Sampling systems fail for many reasons, including hardware failure, improper design, or incompatibility with either the process stream or the analyzer. Components such as filters, valves, and connectors may plug. This may be true especially if there are problems in maintaining the pressure, temperature or flow of the process stream. Every effort will be made to point out where such faults may occur in the specific sections that follow.

Even in cases where the sampling process has been minimized, there may still be problems. In the case of surface-sensitive techniques (e.g. attenuated total reflectance spectroscopy), the physical characteristics of the process fluid, such as miscibility or surface tension, may cause segregation of the sample at the probe. Also, key process variables, such as temperature and pressure, must be addressed to ensure that changing them will not adversely affect the sample. If possible, maintain the sample condition as closely as possible to the process conditions.

3 GENERAL CONSIDERATIONS

3.1 Sampling System Criteria

Along with being representative of the process parameter being investigated, the criteria of the sampling system are the deliverable commodities of the sampling system.[4] When designing a sampling system, these features should always be considered.

3.1.1 Robust

A robust system withstands not only the rigors of the process stream but also the rigors of the environment in which the system is installed. The system must function as intended over a reasonable period of time without the need for excessive maintenance. Both the ''reasonable period of time'' and the ''excessive maintenance'' are user-defined performances. Users will often invest a great deal of effort into maintaining a sampling system and instrument system that provides crucial process information. Of course, the ancillary consideration is also true; if an instrument system provides unimportant or inaccurate insight into the process, very little, if any, maintenance would seem reasonable.

3.1.2 Reproducible

Ideally, a reproducible sampling system is one that, when presented with identical samples, will yield identical results. Practically, such a system will persistently indicate the properties or species of interest under normally varying conditions. The reproducibility is a measure of the overall system precision, not that strictly of the analysis portion of the scheme. One would like the sample transport and conditioning to have little, if any, effect on the variation in the analytical results.

3.1.3 Timely

Any sample pulled from a process stream is a measure of the stream at that point in time. It is a snapshot of what has occurred, not of what is occurring. This realization is very important because users need to understand that, except for direct-responding in situ probes, all samples have an age associated with them. It should be the goal of the sampling system designer to make the capture and analysis of a process sample fast enough to detect deleterious events before problems arise. In the case of sampling for laboratory analysis, not only does the sampling system need to reflect accurately the process stream at that point in time, but also the sampling plan needs to consider how fast the process stream changes.[13]

3.1.4 Economical

Sampling systems often cost as much or more than the analyzers they provide streams for. The service they provide should not be minimized, but wherever possible this service should be economized. The elegant approach is usually the best – the designer should construct a system that is easy to follow and maintain. Additionally, thought needs to be given to whether a sampling system may be shared by an analyzer, or an analyzer shared by multiple sample streams, without harming the performance of any one unit.

3.1.5 Safe

The sampling system resulting from the design effort should be as safe as the process piping to which the stream connects. Caution should be exercised when the materials transported by the system are under pressure, contain corrosive or erosive materials, are flammable, explosive, or reactive, or may cause an unsafe condition to exist when removed from the process piping.

3.1.6 Maintainable

Any components of the sampling system that require routine maintenance or repair should be designed in such a way that there are alternate flow paths and proper clean-out facilities. Having to completely shut down a process to replace a filter is not a desirable design characteristic of any sampling system.

3.2 Transport Time Requirement

When moving a sample from a process stream, there are a number of objectives that need to be addressed. The first is the time lag that the user can afford; this depends largely upon how fast the process changes.[19,20] Next, remembering that the analyzer cycle time is additive to the sampling system lag time, what is the total delay time in the system? Can we live with this total cycle time? The motive force for moving the sample throughout the system then needs to be considered. What equipment such as pumps and gravity feeders exists? Finally, one should consider the phase of the sample. What is the phase in the process? What phase must the sample be in for the analyzer?

A number of strategies exist for moving samples through a system. These include using an in-line probe for direct analysis within the process stream, vaporizing samples from a high-pressure tap for vapor transfer to an analyzer, using added or existing pumps to make a slipstream, and using the pressure differential between the suction and discharge sides of the pump.

If the measurement technique is well suited for direct measurement within the process stream, and if the process stream conditions are not too rigorous for the probes associated with these techniques, in-line measurement may provide the most timely and

representative analyses.[5,21] The two main methods for installing in situ probes are: a single-point investigation of the process; or transmitting energy through the process stream from an excitation probe to a collector probe.[21–26] Optical probes for visible and near-infrared spectroscopic measurements may be mounted in the process stream, with the excitation and response signals transmitted back to a spectrometer using fiber-optic cables. Infrared spectroscopy has been performed in a similar manner with the light being transmitted through an inert gas-filled light pipe. Optical probes may be used in either single-point- or transmission-type modes. Potentiometric probes, such as those for measuring pH or ionic species, are single-point measurements that will measure directly in the process stream – if favorable conditions exist.[3–5] The drawback for measuring process streams in line is that very few process streams will not foul or coat these probes. Additionally, these measurement techniques may not be appropriate for the analyses required by the user. Finally, either any probe mounted directly into the process stream must be removable for replacement or maintenance, or the entire process must be stopped and the stream declared safe for maintenance to occur. Alternate flow paths allow for maintenance and repair without requiring a complete shutdown of the process. If possible, redundant instruments or analyzers will provide backup capability in case the primary unit goes down for any reason.

3.2.1 Vaporization

Direct vaporization at a sample tap may be the best approach if the process monitor requires a gaseous sample for analysis.[2,27,28] In most cases, the pressure is dropped though a heated regulator that volatilizes any liquid materials. The heated regulator is also used to prevent droplet formation in vapor sample streams due to the Joule–Thompson effect. This technique is commonly used when pumps or compressors are not readily available and when the gaseous sample does not need to be returned to the process stream.

3.2.2 Slipstream

The slipstream technique is probably the most common sampling technique for process analyzers and laboratory analysis. Figure 2 illustrates a ''fast-loop'' configuration of the slipstream technique. In this method, a process pump at the point of interest is used to move the sample to and from the process monitor. A small stream is taken off the pressure side of the pump, transported through the sampling system to the process analyzer or sample container, and returned back to the process stream through a small inlet on the suction side of the same pump. The pressure differential between the suction and discharge sides of the pump provides a strong force to move the sample through the system.[1–5,29,30] The amount of material being recycled back to the process is designed to be sufficiently small to be negligible. As shown in the diagram, two three-way valves may be used to block the sampling system from the process stream, as well as to purge or clean the sample lines.

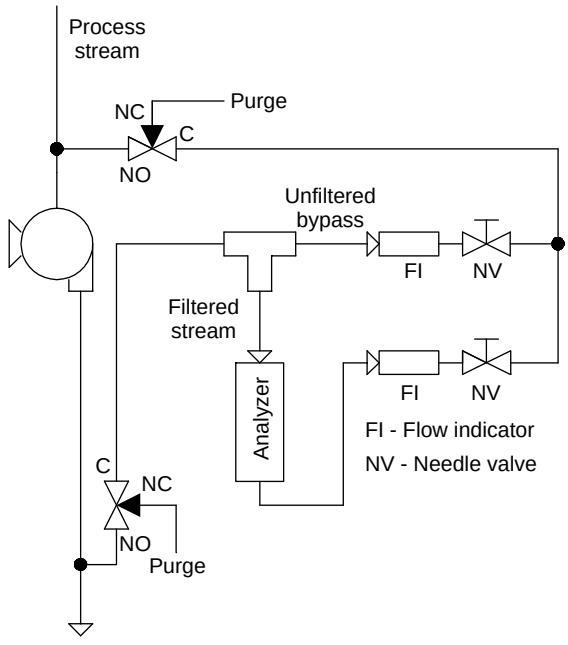

Figure 2 Fast-loop sampling for process analysis.

3.3 Materials of Construction

After function, the next most important criterion when building or designing a sampling system should be the materials of construction. The sampling system itself must be compatible with the process stream constituents as well as the temperatures, flows, and pressures of that stream. A sampling system designer must look at all the components that comprise the measurement system to see whether any opportunity exists for failure due to material incompatibility. Flowmeters may contain O-rings that swell in the presence of certain organic solvents. Valves may contain soft plastic seals that wear quickly in a stream containing hard particulates. Every effort must be made to investigate all potential failure modes of components to ensure the greatest reliability possible of sampling systems. In addition to *Perry's Chemical Engineer's Handbook*,[31] there are a number of good sources for defining and investigating material compatibility issues.[32,33]

3.3.1 Corrosive Systems

Some process streams are obviously corrosive, such as those in plants that produce highly corrosive materials such as chlorine or hydrofluoric acid. Other processes may use corrosive materials such as sulfuric acid, and others may produce corrosive materials as byproducts. Some streams, such as phosgene, are not corrosive under dry conditions but as little as 20 ppm of moisture will produce highly corrosive species, such as hydrochloric acid, which may also cause serious safety concerns.

3.3.2 Organic Systems

Whenever organic systems are monitored using either a sampling device or an entire sampling system, not only must one consider possible sources of corrosion but also possible sources of dissolution, swelling, and other interactions due to the solvating power of the process stream constituents. A practical approach would be to list all components that may contain elastomeric materials, such as valves, regulators, and pumps. Next, review the process stream constituents along with the extremes of temperature and pressure. Finally, make certain that all wetted parts are compatible with the process stream mixture or specify alternative materials that are compatible. Identify any unknown material interactions as initial points to check during early maintenance checks.

3.3.3 Inertness

A common misconception is that certain materials such as Teflon® or stainless steel are inert. For many analyses these materials are effectively inert, i.e. these materials do not interfere with the analysis being performed to a measurable degree. Teflon® is known to convey gases such as oxygen and to entrap certain materials such as methylene chloride and hydrofluoric acid. If percentage-level measurements are being performed, the interaction level is probably negligible. If, however, one is performing parts-per-million-level measurements, these interactions may be the cause of great uncertainty in the measurement system. Water will similarly adhere to metal surfaces. This is not ordinarily a problem, but when measuring very low levels in a water-sensitive process the measurement system will be slow to react to changes because of adsorption and desorption phenomena.

4 SAMPLING OF FLUIDS

Many of the concerns of designing sampling systems for liquids are the same as those used for gases. Sampling systems for liquids use many of the same components as those for gases, and the performance criteria are generally the same.

4.1 Design Considerations

Simplifying the measurement system should always be a primary objective of any designer of sampling systems. When moving fluids through a sampling train, one must realize that every added component affects the response time of the measurement system. Filters, valves, and even fittings can cause restrictions that slow the flow of fluid through the sample transport system, which, in turn, slows the overall response time. The sampling point also is often overlooked by sampling system designers. As pipe diameter decreases, laminar flow becomes more of a concern. Laminar flow approximates to plug flow, which means that poor mixing occurs in the stream. When mixing is poor, there is little opportunity to capture a sample that is representative of the entire process.[1,2,27,28,34]

The main criteria associated with all sampling schemes apply especially to fluids. Any captured sample must be representative, i.e. changes in any process moieties must have corresponding changes that are directly related in the portion extracted from the stream. Note that the concept of accuracy is not raised here. It is often far more important to detect changes in the process stream than to know exactly what the concentrations of those species are. To be certain, accuracy is an important function of any measurement system. However, depending on the nature of the measurement, and the process point being measured, a degree of accuracy may be sacrificed in favor of repeatability.

Repeatability refers to a measurement system's tendency to yield the same value when presented with the same sample. If the measurement system has a high degree of repeatability, changes in the process are easy to detect. If a system has good accuracy but poor reproducibility, control limits will be fairly large, meaning that small but important process changes are more difficult to detect.

4.1.1 Analyzer Requirements

There are many different types of analyzers that may be mated to process streams via sampling systems. These analyzers, which range from chromatographs to spectrometers, from pH sensors to moisture analyzers, and from turbidimeters to refractometers, may have different requirements for accepting a flowing stream. In some cases, the sampling system needs to drop the temperature, pressure, or flow of the stream before it is introduced to the analyzer. In other cases, extensive filtering or coalescing may be required within the sampling train. In still other cases, the phase of the sample may need

to be altered, as in vaporization for a gas chromatograph, or for certain moisture analyzers.

4.1.2 Phase Considerations

The simplest type of sampling occurs when the process stream is at near-ambient conditions of pressure and temperature and contains a safe, dry, clean stream. Unfortunately, few, if any, streams approach the ideal case. To further complicate matters, many streams contain multiple phases. One can envisage a cross-section of a pipe in a corrosive, hot process where the process stream is liquid, but at the bottom of the pipe are solid rust particles and at the top of the pipe are gas bubbles (perhaps dissolved air that after a pressure drop or temperature increase was no longer soluble in the liquid). In order to picture what is happening in the process stream, it is advantageous to exclude these nonparticipating species, unless we are specifically interested in the degree of corrosion or oxygen or moisture content. When faced with such a situation, there are a number of design considerations to explore. Is this the best sampling point? Can these other materials easily be excluded from the sample train without complicating the measurement system? How does the decision made affect the repeatability and representative nature of the sampling system?

4.1.3 Maintenance

The maintenance of the sampling system is often overlooked when designing the installation of a process analyzer system. Frequently, however, the components that comprise the sampling system often require more maintaining effort than the analyzer portion. It is here that the line sizes, filter sizes, valves, regulators, and flow indicators must be considered carefully. Often there is a trade-off between ease of design, ease of installation, cost, and maintainability when designing and constructing a sampling system. Functionality is not a negotiable factor and, hence, is the design foundation of any sampling system. High regard should be placed on the maintainability of the system – if the system cannot easily be maintained, it will not be made use of unless the measurement afforded by the system is of extreme value.

4.1.4 Multistream Sampling

The plumbing system used to monitor multiple streams is known as either a multipoint or multistream sampling system. It is common practice to install a single analyzer to monitor multiple streams in processing plants for a number of reasons.[35] The primary reason has been cost, where a number of very similar streams that require process analysis are in close proximity to one another. Under these conditions it may be possible to leverage the cost of a process analyzer by providing an analysis of all of these streams. Another important justification for using multistream sampling systems is the availability of space – a plant may not have the ground space to accommodate multiple gas chromatographs, for instance. Other considerations include safety, maintenance of a single analyzer, and modifications to existing systems. When designing a multipoint sampling system, there are additional requirements that need to be considered:

- response time from individual streams;
- manifold size and function;
- contamination from previous sample during the analysis (often called cross-talk);

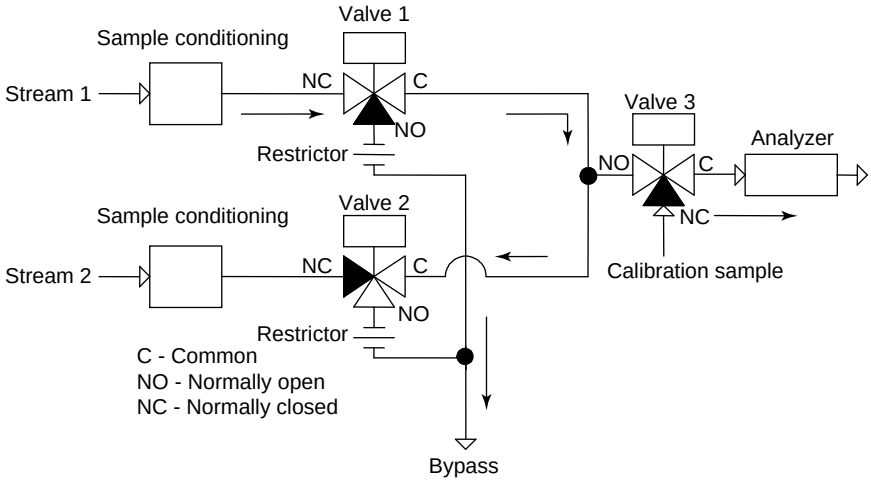

Figure 3 A version of a block and bleed set-up for multistream sampling. Closed ports are shown as blacked-out paths. Flow proceeds through valves 1 and 3 into the analyzer; excess pressure is vented through valve 2 and a restrictor.

- contamination from one stream into another; and
- synchronization of analysis events.

The most common valve configuration for multipoint sampling systems is the block and bleed approach shown in Figure 3. Although only two sampling points are illustrated here, the points may be duplicated many times over. This technique stops the flow of all nonselected streams from entering the analyzer and vents any sample between the blocked valve and the instrument to atmosphere or to the low-pressure return line. There is always one stream active but the remaining streams are blocked using this method. The valves are synchronized by connecting the pairs to the same pneumatic or electric actuators – this also prevents cross-talk to a large extent.

A better approach is the double block and bleed method, which has additional valve components that trap components from a previous stream behind two valves. The flow diagram of this technique is shown in Figure 4. The use of the second three-way valve allows the fluid remaining in the tubing to be cut off fully from the analysis train. Both valves are normally vented to atmosphere or an otherwise low-pressure stream, preventing pressure leakage past the valve.

The two other major techniques for multipoint sampling of fluids are backflushing and bypassing. In the backflush technique, the active stream, when selected, not only passes into the analyzer but also backward through the nonselected stream to an orifice just prior to the selection valve. The orifice connects to a low-pressure vent, such that the selected stream will flush out residual sample from the previous stream. In the bypassing technique, the nonselected streams flow continuously through the normally open port of a three-way valve. This prevents pressure from building at the valve itself. With the bypass principle, the pressure of the selected stream should always be much greater than the bypassed streams, preventing contamination. The bypass configuration also ensures fresh sample from the process, because the nonselected streams always flow from the sample tap. Neither the backflush nor bypass sampling system is generally used widely by itself. The double block and bleed system is commonly configured with bypass lines, especially when the discarded sample is recycled back to the process stream. All these methods may be used with a sample bomb or capturing device in place of the analyzer. The sample stream sweeps through the bomb, forcing out old sample or air with fresh sample.

4.2 Sampling Modes

There are basically two ways to make measurements for process streams: in the process stream itself or by extracting a sample from the stream. Laboratory analyses are of course extractive by nature in that a portion of the process stream is removed from the process and brought to an analyzer housed away from the process. A true in-line analyzer or probe is placed in the process kettle or flowing stream. On-line analyses can either be in-line or extractive. Some techniques obviously require one mode, e.g. all chromatographs use extractive sampling. In other cases, such as monitoring spectroscopic properties, one can use an in-line measurement or could extract sample in order to condition the stream for the analysis.

4.2.1 In-line

These types of analyses either are performed directly within the process stream or use a probe that looks directly into the process. Direct analysis probes such as those for pH, oxygen, and water are generally self-contained, except for power and signal umbilicals that bring in electricity and transmit readings. These probes are usually

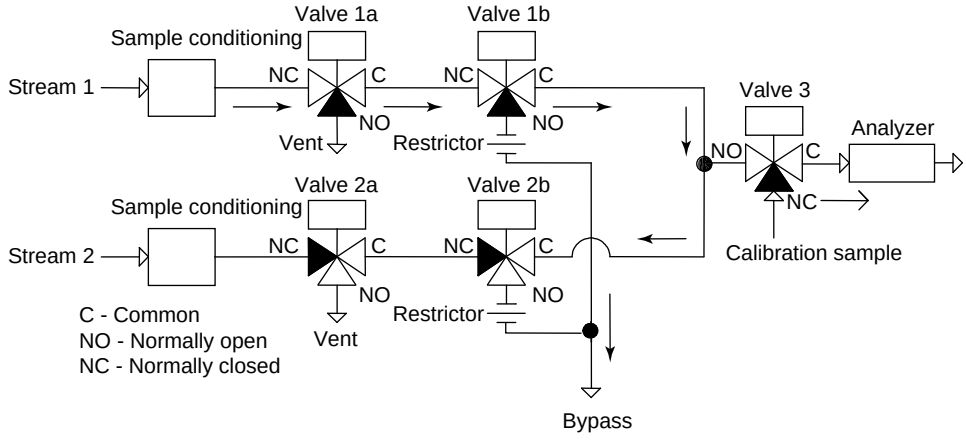

Figure 4 A version of a double block and bleed set-up for multistream sampling. Closed ports are shown as blacked-out paths. Flow proceeds through valves 1a, 1b, and 3 into the analyzer; excess pressure is vented through valve 2b and a restrictor.

mounted through a tee mated to the process line. Probes for spectroscopic analyzers such as refractometers or infrared photometers either may be mated to the process stream via a tee or will be built around the process pipe such that energy transmits across the stream.

The requirements for tee-mounted devices are generally the same regardless of the analytical technique employed. Single-point devices require that fresh sample comes in contact with the probe surface. The designer therefore needs to be certain that these probes are inserted into portions of the stream with turbulent flow, avoiding areas where plugging or dead space may occur. Because some spectroscopic instruments transmit energy through the bulk stream during the measurement phase, these instruments require that the portion of the process stream in the light path be representative of the process.

In either case, the most troublesome problem for in-line devices is fouling or coating of the active surface. With electrochemical probes, such as those for moisture or pH, fouling results in inaccurate results by prohibiting a representative portion of the process mixture from coming in contact with the sensing surface of the probe. For many transmission instruments, fouling may cause absorption of the transmitted energy, causing baseline shifts, inaccurate results and heavy maintenance. A number of methods have been used to mitigate or eliminate fouling. Occasionally, simply by choice of materials, fouling may be avoided. In other cases, a water, steam, or solvent flush is programmed into the analytical cycle to clean the sensor surface periodically. Additionally, one may raise the temperature or stream flow rate in order to prevent materials from adhering to the active surface of the probe. Some of the spectroscopic probes utilize ultrasonic cleaners for continual refreshing of the optical surfaces. Finally, maintenance can be performed preventatively, before the fouling is significant.

4.2.2 Open-path or Ambient

Direct measurement of chemical species in the atmosphere is a specialized case of in situ monitoring. Open-path monitoring involves measurement of gas-phase compounds by absorption spectroscopy realized between a light source and a detector spaced far apart.[36] This technique is gaining in popularity for fugitive and stack gas emission detection. The most common open-path techniques make use of the power of Fourier transform infrared (FTIR) spectroscopy.[37] The special efforts to be taken when sampling ambient air are thoroughly described in an ASTM standard.[38]

4.2.3 Extractive to Waste

As mentioned previously, process sampling for laboratory analysis requires the extraction of a sample that will rarely, if ever, be returned to the process stream. There are many devices used to capture a representative sample for subsequent analysis by a laboratory instrument. The major considerations for process analysis by a laboratory instrument are:

- timescale of analysis;
- response time required by plant operations;
- stability of the sample; and
- laboratory analyzer's requirements of the condition of the sample.

A sampling system designed for capturing a representative sample for transport to a laboratory has more demands placed on it than does the supply system for a similar on-line analyzer. By virtue of transporting a sample to a laboratory, these sampling systems must have a removable container that integrates into both the process and laboratory sides of the measurement system.

Some process analyzer systems also send all captive sample to a waste (or vent) stream. It is common in these systems to run a fast bypass loop near the analyzer while removing a portion for passage through the analyzer to a waste stream. This keeps the response time low while minimizing the process stream waste. This may be necessitated by the type of analysis, such as a gas chromatograph requiring that a liquid stream be vaporized continuously before filling a sample loop. More commonly, some processes, such as those in the food and pharmaceutical industry, require that no contamination occurs, and this technique is absolutely essential.

4.2.4 Extractive to Return

This is the most common sampling methodology for process analyzers to utilize. In this technique, a portion of the process stream is brought out of the main flow path through an analyzer and returned back to the process. This smaller, recycling, flow loop is usually referred to as a slipstream. Typically, the designer of the sampling system uses a pressure differential already existing within the plant as the motive force to move the portion of the process stream through the analyzer. For example, the inlet to the sampling system can be from the discharge (high-pressure) side of a process pump, and the sample is then returned to the suction (low-pressure) side of the pump. The sampling system designer first considers the pressure across the pump, then determines the flow needed for a satisfactory response time, and finally calculates the tubing and associated pressure drop of the sampling system. If the process does not contain sufficient motive force to supply the sampling system, additional hardware such as pumps or aspirators might need to be added to the sampling system. A complete discussion of

the calculation of pressure drop in a sampling system is given by Sherman and Rhodes.[4]

4.3 Sampling Point Determination

Determining the best sampling point for process analysis requires the designer to consider not only the analysis but also the process stream. The sampling system needs to be compatible with the process stream at the sample tap. The sampling point must also be meaningful from a control viewpoint. Finally, the sampling point and related hardware must also be maintainable. Often, the sampling point chosen for process analysis is a compromise among all considerations.

One other consideration is the integration of an extractive sampling point for laboratory analysis. Samples will have to be pulled occasionally for validation of the on-line analyzer. Building in a laboratory sampling point at the same location in the process makes sense in terms of direct comparability of the results.

Ideally, the best sampling point will yield the most information about the process stream. It will give a representative view of the production in a timely fashion, allowing experienced personnel to adjust key parameters before any upsets occur. This is a critical consideration – one must have the control requirements of the analysis closely aligned with the sampling point for the measurement system. Failure to marry these two concepts could lead to a failure of the installation to provide useful information. The sampling system designer must choose the sampling point carefully – he or she must select the tap and allow for transport and analysis time when calculating the response time of the measurement system.

Often the sample tap that would yield the best response time is not the best point for producing either a maintenance-free system or a representative sample. In these instances, a compromise must be made between what is expected and what is practical. Often, after a pressure drop in a gaseous flow the stream is more homogeneous due to turbulence at the restrictor. Although the high-pressure side would produce faster flow in a slipstream, the better choice may be to tap into the low-pressure side to catch a more representative sample.[30,39–41] This would be especially true if there still were enough pressure to move the sample quickly to the analyzer, or if an auxiliary device such as a pump could be added easily to the sampling system to augment the flow.

Sample stability in the transfer line leading to the analyzer or sample vessel is of great importance. Changes in flow, pressure, and temperature could lead to partitioning of the sample matrix, which causes the captured portion to lose its representative nature. A hot gas, for instance, could adsorb to the walls of cooler tubing; this adsorption leads to a memory effect. Adsorption of materials will cause species not to be detected as quickly as the flow would indicate, whereas desorption of these species into an otherwise ''clean'' sample stream would cause the analyzer to register values after the process stream was clear.

4.4 Sample Transfer

The response time of any process measurement system relies on the time it takes to transfer a sample from the process stream to the analysis device. In some cases, the measurement time is fairly long, such as in chromatography, where 10-min cycle times are typical. In other cases the instrument response time is negligible, of the order of seconds. It is in the latter case where sample transfer is typically the limiting portion of system response time. The two main design considerations of sample transfer are the linear velocity (speed) of the sample and the amount of plumbing (distance) to the analyzer through which the sample must travel.

Speed of response for process measurement systems can be achieved by either placing the analyzer close to the sampling point or allowing the sample to flow quickly from the tap to the analyzer. The electrical classification of the area near the sampling point as well as the available space often limit how close the analyzer may be placed to the process. In those cases where sample flow is not a problem, it is often best to locate the analyzer in a safe, convenient location and use a fast-loop bypass to waste or to a plant return. Obviously, if there are pollution or cost issues, it is best to return the sample back to a lower-pressure point in the process stream. The extractive-to-return system can be configured with a bypass that is always flowing at a high flow rate while a small portion of the sampling stream is sent to the analyzer or sample-capturing portion of the sampling system. This design ensures small lag times and can accommodate small-volume analyzers. It is important to note that the return points in fast-loop systems must be considered carefully. Fast-loop sampling relies on differential pressure within a process and should never be placed across control valves for fear of adversely affecting control of the process.

The amount of sample to be transferred to a process analyzer is determined by two factors. First, what dimension of piping needs to be used to keep the pressure drop at an acceptable level? Second, what portion of the process stream must be transferred to the analyzer for accurate representation of the process? Typically, a designer will aim for at least three volume flushes of the sample transfer lines for either laboratory or discontinuous analyzer sampling. This rule of thumb is employed to be certain that the captured sample is representative of the process by washing out sample

remaining in the lines from a previous analysis. For continuous analyzers, the constant flow through the pipe leads to a defined lag time for the entire measurement system.

Transfer lines and their associated hardware must be compatible with both the analyzer and the process stream and must be designed with the system requirements in mind. All components must be sized correctly to allow for the desired flow. The lines should be run in a linear fashion as much as is practical; bends and elbows slow response time due to axial mixing. Turns could also cause problems such as plugging or the formation of dead volumes. Wherever bends in the sample transfer lines are necessary, provisions should be made for cleaning and maintenance. Whenever possible, lines should be sloped to allow for proper drainage of condensable gases and to prevent particulates from settling.

The hardware also must be constructed of appropriate materials, and during the design process consideration of the process piping is used to determine these materials. Because of the smaller dimensions and analytical nature of these components, greater care needs to be taken. Corrosive and erosive processes are more severe with thin-walled sampling system tubing than they are with thick-walled process piping. In addition, adsorption is not a primary concern in process piping, but it would be in the stream leading to an analyzer. Finally, the valves and other components associated with the process stream are more resistant to plugging than their finer counterparts located in the sample transfer lines. An ASTM standard practice on flow and temperature control in process measurement systems discusses these concerns more completely.[42]

4.5 Sample Conditioning

A primary function of any sampling system is to condition the sample for analysis. The sampling (sample conditioning) system is responsible for making the transition from what is in the process stream to what the analyzer is capable of accepting without changing the sample composition in a meaningful way. Sample conditioning can be as simple as adjusting temperature or pressure, or as complex as quenching a reaction or vaporization of a liquid stream. The degree of conditioning may depend upon the physical state extremes of the process, or upon the chemical nature of the sample itself. The amount of conditioning is also dependent upon the constraints of the analyzer system itself, such as the size and amount of particulates that may enter analyzer components. Finally, conditioning is dependent upon the performance criteria of the sampling system itself, such as easy maintenance, and avoiding corrosion and plugging of the sample lines.

4.5.1 Representative Nature

In a strict sense, sample conditioning tends to make fundamental changes to the sample. By removing particulates using a filter, constituents of the process stream have been preferentially removed, and the sample is no longer representative of what was in the process stream. This is acceptable if there is an understanding that the removed components are either not part of the process or can be accounted for via calibration or some other technique.

4.5.2 Liquid Vaporization

Sample conditioning includes vaporization of liquids to gases. This is done for one of three reasons: to reduce lag times; to keep portions of a gas stream from condensing; or to present a gas sample to an analyzer requiring such. If either a liquid or gas could be analyzed as part of a measurement system, the gaseous form is often preferred because less sample may be used and higher flow rates may be achieved. Often, sample lines may be heat traced above the dew point of any component in the process stream in order to prevent condensation, which leads to disproportionate moiety concentrations in the captured sample.[43] Gas chromatographs may require vapor samples, depending upon the interface to the sampling system. By vaporizing samples as close as possible to the sampling tap, a fast-flowing representative sample is presented to the analyzer. Means used to vaporize samples include heating or insulating sample lines, steam-heating vaporization lines, and vaporizing regulators with electrical or steam-heating sources.

Another way to vaporize a portion of the sample stream is by stripping or sparging. In this technique, an inert gas is bubbled through a trapped portion of a liquid stream. The most volatile components travel with the inert gas to a gas analyzer whereas the spent liquid stream is sent to waste or back to the process stream. A similar technique is headspace analysis, where a portion of the liquid process stream is sent through a chamber. The liquid does not fill the chamber, instead volatile species travel to the headspace above the liquid and are subsequently sent to a gas analyzer using an inert gas. These techniques are good choices when the liquid streams are difficult to clean or contain high concentrations of dissolved solids. Both of these techniques require good control of flow rate and temperature to achieve reproducible analyses.

When considering process measurement systems, one needs to determine what phase the sample captured or presented to an analyzer will be in. Although in many cases either a liquid or gas may be used, one phase may be preferred over another. Gas chromatographs, because of maintenance associated with liquid injection valves, usually favor gas samples. Spectrometers, on the other hand, because of pathlength limitations, may favor

liquid samples. Other considerations include the ease of conditioning the sample further, the additional hardware required, and, most importantly, which phase maintains the sample integrity.

4.5.3 Temperature and Pressure

Both the sample temperature and pressure must be controlled in order to control the phase of the sample. It may be necessary to cool or heat the sample or to maintain or reduce its pressure as it travels to the analyzer portion of the measurement system. Bubble formation and disproportionation of the sample may be avoided by maintaining pressure or by cooling the sampling stream. Cooling may also be necessary if the process stream contains reactive species. The most common reasons to reduce pressure in a sampling system are to provide the analyzer portion with a low-pressure sample and to provide an easily manipulated sample stream.

An added benefit of reducing the sample pressure for analysis is safety. In the case where the sample is brought into an analyzer shelter or room, high-pressure sample lines are not desirable because of the possibility of oxygen displacement in the event of a leak.

4.5.4 Filtering and Separation

Because the goal of any sample conditioning system is to present the analyzer with a representative portion of the process stream ready for analysis, removal of particulates or other harmful species needs to be considered. As mentioned previously, removing entrained species is a preferred practice if these species are not of interest or cannot be accommodated. There are many separation techniques used in sampling systems; the method employed depends on the specific problem to be solved. Separation may be effected at the inlet to the sample transfer lines, within the lines, or at the outlet to the analyzer portion.

Particulates are removed from dry gases using filtering. Filter materials may be sintered metal or wire mesh made from appropriate materials. Filter elements may also be membranes, metal disk arrays or fiber tubes. Other materials such as ceramics and plastics have been used extensively; again, one must consider material compatibility with both the physical and chemical property extremes of the process. Self-cleaning filters, such as bypass filters, may be used in parallel to make a long-lasting, easily maintainable system.[34,44] Placing the filters in parallel also allows for maintenance without shutting the system down.

Cyclone filters are also used to remove particles from gas and liquid streams. In such a device, the incoming sample fluid forms a vortex; centrifugal action forces the separation of particles and droplets. Gas bubbles caught in liquid streams are also readily removed using the cyclonic action. Although the smallest of particles are not removed using this equipment, pressure drops are often significantly less than those experienced in pass-through filters. Cyclonic filters are also easily incorporated into fast-loop sampling systems. The main drawbacks of these devices are that flow rates of sample and bypass streams need to be determined experimentally and then controlled.

Another device used to separate liquids from gases is a coalescing filter. Liquids are removed from gases by being trapped on a porous bed, accumulating and dropping out of the flowing stream or by being sent to the bypass stream. Bubbles may be separated from liquids using an inverted coalescing filter, allowing the vapor to travel up to the top of the device.

5 SAMPLING MIXED PHASES

Phase heterogeneity occurs in numerous production processes. Here heterogeneity is defined as having more than one phase present in the flow stream being monitored, rather than differences within very small samples being withdrawn from the stream. Heterogeneity exists either by design or by accident, such as when temperature or pressure upsets occur within a plant. Sampling such systems requires some process knowledge as to whether all phases are to be sampled and analyzed, or whether the sampling system must effect some type of segregation. In this section, emphasis is placed on incorporation of all process variability into the measurement system, i.e. sampling the bulk and obtaining sufficient representation of the entire process stream.

5.1 Classifications

Regardless of the nature of the nonuniformity of the sample, the sampling system designer must attempt to create random heterogeneity in the sample that is captured. The portion of the process stream captured must reflect the heterogeneity of the process as a whole.[45–48] If the sample captured is biased in any fashion, the system must either be corrected or calibrated to account for the bias. In order to capture a representative sample, portions withdrawn from the process stream tend to be larger than those removed from homogeneous systems.

5.1.1 Gas–Liquid

Process streams consisting of a mixture of gases and liquids are typically not sampled for process control, often because this condition is not a normal mode of operation. Exceptions to this rule of thumb usually occur

due to temperature or pressure cycling, typically in batch vessels or occasionally in distillation systems.

5.1.2 Liquid–Liquid

Immiscible liquids, which in a static system would normally separate into two phases, can form emulsions in process streams. Emulsions are tiny droplets of a liquid (disperse phase) suspended in another liquid (dispersion medium). Some industrial processes are designed to use emulsions (such as block polymer additions to latex particles), whereas in other cases emulsions form due to process upsets (an organic liquid exceeding its saturation point in an aqueous mixture). When sampling such mixtures, it is typically important to maintain the emulsion as it is in the process stream, for fear of disproportionation occurring.

5.1.3 Solid–Fluid

Solids may be dispersed in either gas or liquid process streams. Smokestacks contain particulate matter suspended in flowing hot gases. Other processes may use fluidized beds or pneumatic transfer of solids. In each of these cases, sampling may need to capture both the solid and gas phases. In sampling such systems, one needs to be mindful of erosion and abrasion processes degrading the sample transfer lines and components. Process liquid streams may also contain solids by design or by lack of process control. Solubility loss can lead to solid particles precipitating from solution (which may be planned or unplanned). In other cases pastes, suspensions, or slurries may be formed to achieve process objectives.

5.2 Design Considerations

In all mixed-phase systems, the designer considers the rheological properties of the blended substances. Understanding how these materials flow often leads to how they are sampled. Solids in slurries, for instance may not settle in a static system, but when pumped they may exhibit segregation due to flow.[49,50] Pastes also may flow and maintain representation of the process under certain conditions but, when heated or overly compressed, could lose this character.

5.2.1 Criteria

The concerns for adequately sampling mixed phases are more complex than those for homogeneous streams. Every component placed in the sampling train could prejudice the measurement system. Not only must the designer worry about the captured sample representing the process stream, but he or she must also be concerned with material-handling issues such as blockage, abrasion, erosion, and phase separation. The entire system must be monitored for correct operation, not only upon start-up, but also after the system has been placed in operation.

In general, there are many considerations to be made in the design of these systems, including: the mass flow rate of the material in the process line; the physical state (size and shape of particles, density, degree of wetting and phase ratios) of the materials; and the actual chemical composition of the sample. Other important factors to consider are whether an off-line method is to be used, what equipment is available to help move the sample, and whether the process is to be sampled continuously or at regular intervals.

The first decision to be made when sampling a heterogeneous stream is whether both phases need to be captured. Correct process operation may require the amount of an unreacted starting material which exists only in the fluid phase. In such an example, it is necessary only to extract the fluid – although both phases could be sampled if convenient. In cases where there is solubility of a reactant or product in two phases and the total needs to be known, both phases obviously need to be representatively captured.

5.2.2 Sample Stability

Another important consideration is whether the sample changes after it is removed from the process, and what can be done to prevent these changes. The sampling process itself could cause some speciation; if the probe design will not allow all species to enter the sample train with equal probability, segregation occurs. If the process stream has a significantly different flow rate than the sample train, solids may settle or emulsions may separate into different phases. As with samples from homogeneous streams, concern must also be given to the possibility of the sample reacting or decomposing once removed from the process stream.

5.3 Sampling Techniques

Sampling heterogeneous systems occurs either at a single point or across an entire section of process plumbing. Point sampling makes use of a single, well-defined sample point connected to the process stream. It is the preferred method for homogeneous systems that do not suffer from segregation effects. Although simple in design, this method often fails to yield reliable information about heterogeneous systems. Point sampling is prone to plugging, and relies on the heterogeneous system being very well mixed at its inlet. This technique is appropriate for fine dispersions of a denser material in a less dense medium.

Isokinetic sampling. If both phases are needed to yield the analysis appropriate for process control, it is important

that neither is excluded at the sample tap.[50–53] For point sampling, it is important to have isokinetic sampling.[54] In this methodology the velocity of the fluid flowing into the sampling inlet should be the same as the velocity in the process stream. In addition, the orifice into the sampling system should be aligned isoaxially with the process stream flow, i.e. the process stream should flow directly into the sample tube. Discrimination occurs under nonisokinetic conditions because of the tendency of smaller particles to be pulled into the sampling orifice when its velocity is greater than the process flow rate, and the tendency of these same particles to move around the orifice when its velocity is slower than the process stream.

5.3.1 Cross-sectional Sampling

Cross-sectional sampling can be accomplished in one of three fashions: the entire contents of a cross-sectional volume could be moved from the process line to the sampling system; a slotted, sample-capturing device may be mounted across a section of the process plumbing; or multiple single-point devices can be installed to different depths within the process pipe. In the latter two cases, careful attention must be placed on the positioning of the probes in order to be certain that the sampling is not biased. In all cases, there is a better chance of capturing a representative sample if the tap occurs at a point where the process stream is well mixed.

5.4 Sampling Point Determination

Determination of the sampling point ideally relies more on production requirements for process feedback than it does on the analytical methodology. If the process property to be measured is available only within a small region of the process line or much of the process line has limited accessibility, the choice of sampling locations is limited. In addition, there are the more frequent requirements of lag time and required sample conditioning, which limit the choice of sampling location. Other points to consider are the ability to withdraw sample without interfering with the safe and effective operation of the process, as well as the installation of associated equipment needed to move the sample to an analyzer or sample vessel.

In order to sample a heterogeneous stream for laboratory or process analysis, one must have the ability to move the sample away from the process line to a vessel or analyzer and then move the stream to waste or back to the process. For less viscous samples, including many suspensions, emulsions, and slurries, the same flow considerations as for homogeneous fluid streams are made. For more viscous samples and streams with heavy solid loadings, care must be taken to include devices such as screw feeders to move a portion of the process stream into and through the sampling system.

Once a sample has been placed within the sample train, its representative nature must be maintained. Sample flow, temperature, and pressure may need to be controlled to achieve this goal. Additionally, equipment may need to be added to the sampling system to prevent blockage of lines. Finally, the entire sampling train needs to be compatible with the sample and the measurement objectives. Pumps, extruders, feeders, and other flow devices must be appropriate for the materials being monitored.

Maintaining these complex sampling systems requires a good deal of effort. Even in well-designed sampling schemes, systematic error or bias exists.[48] Measurement system bias is often largely due to the inability to capture a representative sample. Systems for heterogeneous processes, because of their complex nature, are more prone to these effects. Bias may occur because of design defects, temporary plant conditions, or aging of the sampling system itself. It is important to understand the bias of any system upon commissioning and to monitor changes in the bias over time. Because repeatability of the process analysis is very important for monitoring process changes, a slow change in the measurement system bias needs to be followed.

5.5 Sample Transfer

Transfer of heterogeneous samples can occur by many mechanisms. All of the devices used for homogeneous fluids, if appropriate, may be used, as well as devices specific for certain applications. Examples of other transfer devices are hoppers, conveyors, pneumatic devices, and metered feeders. Many problems are encountered in transferring heterogeneous materials, such as pluggage, abrasion, erosion, corrosion, and sample lag.[47]

The flow characteristics of the sampled materials often dictate how the materials will be transferred to the sample vessel or analytical instrument. The densities of the components along with the velocity and viscosity of the fluid determine the likelihood of separation. Care must be taken to keep the velocity of the material in the sample transfer lines above rates where deposition or phase separation would occur. Insulation or heat tracing may also be required, depending upon the physical and chemical properties of the stream constituents.

The amount of sample to place in transfer lines depends on a number of factors. First, the amount of sample required to accurately represent the process stream must be considered. Next, one must ask whether the sampling process is continuous or interval sampling is used. In addition, one needs to consider the flow rates to the vessel or analyzer required to maintain the representative

nature of the sample. Finally, one needs to consider the pressure drop in the sampling lines and the flow and pipe dimensions required for a timely sample.

Successfully transferring and conditioning heterogeneous streams require more effort than performing the same functions for homogeneous streams. In general, line blockage is more likely to occur. Additionally, there are typically more components in a heterogeneous sampling system. The process streams also tend to be more demanding because of erosion and related processes.

5.6 Sample Conditioning

Sample conditioning requires that the sample be prepared for analysis or storage. The ultimate goal of the sampling system is to condition the sample without changing the important parameters of the sample. There are many modes by which a portion of the process stream may be changed once the sample has been withdrawn from the production pipe. Solids may be lost from flowing streams if the path is too tortuous. Liquid droplets may drop from a gas stream if the temperature drops or if the materials of construction are not appropriate. Line blockage may occur if the solubility of the material changes due to temperature or pressure effects. All these modes must be considered when determining the condition methodology.

5.6.1 Phase

When both liquid and gas are present in a process stream it is usually good practice to convert to a single phase. If possible, it is usually better to convert all species to the gas phase because of the ease with which the sample may be transferred and filtered. For many reasons, such as analytical methodology, liquid boiling point, decomposition, and reactivity, this is not always practical. Likewise, it is also not always easy to force gases back into solution because of pressure and safety concerns.

When other mixed phases occur simultaneously in a process stream, the primary goal is to be certain that the representativeness is not lost in the sample-conditioning process. If solids can be dissolved into liquids by raising the temperature (without losing description of the process), then an acceptable approach has been found. Such cases rarely, if ever, occur. More likely, effort should be placed toward ensuring that the sample stream remains well mixed and nonreactive.

A notable exception to this rule of thumb is when the analytical system requires dilution. Diluents may be added to the sample stream outside of the process pipe if there are no safety, control, or analytical concerns. Dilution is beneficial in many cases because either it may cause a single phase to result or it may allow for easier transport of the sample stream. Caution should be exercised, however, in the generation of large waste streams that cannot be recycled back to a process stream. Flow injection techniques have recently addressed a number of concerns regarding dilution or reaction of process streams prior to analysis.[55–57]

5.6.2 Temperature and Pressure

As far as they affect the phase and therefore the representative nature of the extracted process sample, temperature and pressure are vital parts of sample conditioning. As with homogeneous systems, both stream properties often need to be controlled to be certain that samples can be withdrawn from and returned to the process. Control of these parameters also allows the analyzer to be presented with a recognizable sample, as well as keeping the sample train maintainable. In other cases, temperature extremes may be employed to quench reactions or to trap samples.[56,59]

5.6.3 Filtering

Filtering heterogeneous streams while maintaining representativeness is a difficult task, especially in those systems containing solids. In some cases, one of the phases needs to be excluded for the analysis; in other cases, coarse filtering is preferred in order to digest one of the two phases. In still other cases, obvious portions of a sample stream, such as metal filings, rust, or zeolite drying column particles, are to be discriminated against. Where possible, one can aim to filter these materials from the bulk process stream components. In a select number of cases, filter elements may be used as sample-collection devices for subsequent analyses from heterogeneous streams.[60–62]

6 SAMPLING SOLIDS

Solid process streams present many of the same sampling challenges as heterogeneous streams. Indeed, many solid process streams have a high degree of heterogeneity – solid particles are generally not of uniform size or composition. Because of these specific problems, sampling of solids needs to be considered as a separate topic. More detail for solid systems may be found in ASTM Standard E 877-93, which describes sampling of iron ores and related materials.[63]

6.1 Design Considerations

6.1.1 Particle Size

In the production of solid materials, particles are distinguished by their size. The existing particle size

range in a process pipe determines the sampling system methodologies that could be employed successfully. It stands to reason that all minimal dimensions within the sampling train are dependent on the maximum particle size existing in the process stream. Also important are the particle shapes existing in the process line; it is generally easier to remove spherical particles from a flowing stream than elongated particles. Thus, the particle geometry as well as size will determine the orifice dimensions of the sampling port.

6.1.2 Mixed-phase Streams

When describing measurement systems for solid streams, often the details indicate that two phases are actually present. Fluidized beds usually contain solids and gases, whereas recrystallization vessels contain solids and liquids. In other cases, the solids may be wet from process liquids, or may otherwise contain entrapped fluids with the particle matrix. Additionally, solids are frequently suspended in liquid or gas streams to help move the particles down the process pipeline. In a great number of these cases one is interested in some characteristic of the solid, and the fluid phase is inconsequential. Typically, the sampling system designer must be aware of density differences and potential sources of density changes in order to build a reliable sampling system.[64,65] Segregation frequently occurs in solid streams due to pathway differences arising from density and particle size variations.

6.1.3 Analyzer Requirements

Another key consideration in the design of these sampling systems is how the sample is analyzed. If the measurement is of particle size and distribution, great care must be taken to make certain that the analyzer receives a sample that is representative of the process. On the other hand, if the solid stream is to be pulverized, ground, dissolved, or digested, particle size integrity may not need to be maintained.

6.1.4 Erosion/Corrosion

The hardness of solid particles is also a point of concern. If the particles are relatively soft, poor uniformity of flow within and blockage of sampling lines might be a problem. With these types of materials it is very important to choose the motive components, such as pumps and conveyors, carefully. If the particles are relatively hard, erosion of sample transfer lines and associated components could become a serious problem. Once erosion does occur, corrosion usually becomes more likely because the active surface area exposed is likely to increase. Here, it is important to choose the materials of construction of the sampling train wisely.

6.2 Sampling Techniques

The sampling techniques used for heterogeneous systems are similar to those used for solid sampling. Either point sampling or cross-sectional sampling may be used in these systems, but one must understand the role that the transporting fluid, if any, plays in the conveyance process.

6.2.1 In-line

In-line analyses in solid process systems primarily focus on physical property measurements rather than on chemical composition. These systems also usually rely on a fluid phase for dispersion of the solid or as a part of the system. Examples of these analyses include turbidity and particle size measurements. Notable exceptions to this rule are some spectroscopic instruments such as diffuse-reflectance near-infrared spectroscopy, where the sample passes on a conveyor under a source and detector, or in microwave spectrometers, where the solid stream passes through a specially designed microwave chamber.[66]

6.2.2 Extractive

Although some solids flow freely, extractive systems used to measure solids typically either have a motive force device, such as a screw feeder, or use a trap and purge method for moving the sample from the process stream to the sampling train. In addition to screw-based devices, gravity, suction, and sliding mechanisms are all used to partition a small fragment of the process stream contents into the sampling system. In other cases, a portion of the process stream is dropped or forced into a chamber, typically by an actuated sampler, and then transferred pneumatically to the sample collection vessel or process analyzer.

6.3 Sample Point Determination

Determination of the sampling point in these systems is similar to determining the sampling point in heterogeneous systems. The space, representation, and lag time concerns are the same for solids as for heterogeneous systems. The greater likelihood of solids segregating in the process line often makes the sampling process somewhat more difficult. Because the measurement system will never yield a better picture of the process than the sample that enters the sampling lines, the greatest concern in determining the sampling point is often how to represent adequately the bulk process in the capturing phase.

One must understand how the solids move within the process lines in order to choose an appropriate sampling point as well as appropriate techniques. If segregation does tend to occur in the process line, then it is typically beneficial to capture a cross-sectional sample of the horizontal pipe, or as an alternative to capture a sample on a vertical run of pipe as the sample falls through that section. A final alternative would be to use multipoint sampling across a range of the process pipe.

6.4 Sample Transfer

As with the transfer of heterogeneous materials, moving solids through sampling lines requires the use of appropriate devices for specific applications. When designing the sample transfer portion of the overall measurement system, one needs to be concerned with blockages, erosion, contamination between streams, and increased lag times, as well as safety.

6.4.1 Gravity

Because gravity is omnipresent in production processes, it is often used as an inexpensive motive force for transferring samples through sampling lines. A simple tee may be fitted into a horizontal process pipe. Within this fitting, a door leading to a chute could be connected that would allow samples to drop to a collection device or analyzer inlet. In a less ideal situation, a hopper could be inserted into this vertical line to help control feed rates through the sampling system. The problem with hoppers both within the process and in sampling lines is that there is a residence time range associated with these devices that causes not only greater overall lag times but also an increased time uncertainty related to process changes.

6.4.2 Devices

As mentioned previously, high-pressure inert gas may be used to move solids to an analyzer or sample-capturing device. These systems work well for fairly hard, small (<1 mm) particles, but may suffer from segregation and increased lag time due to dead spots and from increased erosion rates due to the increased kinetic energy of the particles. Conveyor belts, either alone or in conjunction with feeders, are also used to move solids along sampling trains. These other devices must be sized appropriately and designed with the analyzer conditioning requirements in mind.

6.5 Sample Conditioning

Preparing solid samples for analysis may require anything from trying to maintain particles as they existed in the process pipe to total dissolution in a strongly acidic solution. In some cases the solid materials are relatively inert to their surroundings, whereas in others the solids are reactive. Conditioning requirements may necessitate careful temperature or flow control. Temperature needs to be controlled in many cases if the sample is ground as part of the conditioning process, or if significant thermal changes cause changes in particle size or hardness.

6.5.1 Grinding

Many measurement systems require comminution, or size reduction. In processes where the measurement occurs after digestion of process particles, such as analyzing ore minerals by atomic emission in an inductively coupled plasma, comminution is desirable, if not necessary. Other analytical techniques might perform better if the solid particles are of a smaller, more uniform size distribution. Chopping, grinding, crushing, scraping, and cutting devices may all be used to reduce the size of process particles.[63] Often, multiple devices are placed in a series configuration such that comminution takes place in controlled stages. The two major concerns of such devices are their ability to produce representative process samples and their need for maintenance.

6.5.2 Dissolution

Although much dissolution takes place in preparation for laboratory analyses, some systems make use of on-line dissolution. This is done for a variety of reasons: on-line analysis yields faster, more frequent results; the dissolution may be accomplished easily, with readily available solvents or reagents; and sample reactivity or safety concerns may lead to an on-line dissolution approach.

7 CALIBRATION AND VALIDATION OF SAMPLING SYSTEMS

As with any measurement plan, calibration and validation of sampling systems are vital to the value of the equipment. Calibration describes all the efforts made to determine the errors of the measurement system and the efforts placed into being certain that the instrument will yield a response proportional to the concentration of the species of interest. Validation is the quantitative performance of the measurement system and is typically accomplished by comparing the measurement system response to known or matched values.[67,68]

The sampling train and sampling conditioning devices are part of the entire measurement system. As such, they must be considered when evaluating the analysis of a process sample.[69] Because most analytical measurements are comparisons of responses to standards rather than

absolute indications, selection of calibration samples is of critical importance. These standards not only must reflect the variability of the process stream but also must represent the physical properties of the stream contents. This is an important point to note: different sampling systems and analyzers may respond differently to samples with the same constituent values but different physical characteristics. Typically, the best practice is to introduce calibration samples into the measurement system such that the sample undergoes the same conditioning as the portion of the process stream normally measured.

7.1 Frequency of Calibration

All sampling systems need to be calibrated before any process sample enters the architecture. This serves to reassure designers and users that the system will yield realistic results. Calibration prior to operation also allows various safety and other performance tests to be conducted before an actual sample is placed within the system. Once the error of the system is understood, one can make better judgements on the variability of the process as determined by the measurement system.

Recalibration, or at least verifying a calibration, should be performed periodically, with the periodicity determined by a number of factors. Systems prone to drifting or fouling should have calibration checks at least twice as frequently as these mishaps occur. If unusual results are produced by the measurement system, it is good practice to analyze a process sample by an independent analyzer; if disagreement exists, the process instrument system should have a calibration check. Some instrument manufacturers suggest calibration schedules as a part of preventative maintenance, whereas others have incorporated the means to manually or automatically send a calibration sample to the analyzer.

Instruments may be calibrated by the introduction of an external standard or may make use of internal standards added to a portion of the process stream within the sampling system.[70] Although generally agreed to be a less satisfactory approach, the most common practice for recalibrating the measurement system is to draw a sample from the process stream, analyze the sample using a laboratory analyzer, and adjust the process instrumentation to match the laboratory measurement. When using this technique, one must understand the magnitude of the errors associated with both instrument systems in order to decide whether the results are significantly different. If any calibration verification shows the measurement system to be statistically the same, no adjustments should be made.

An important use of calibration is for validation. The instrument needs to be examined periodically in conjunction with the validation process. Once calibrated,

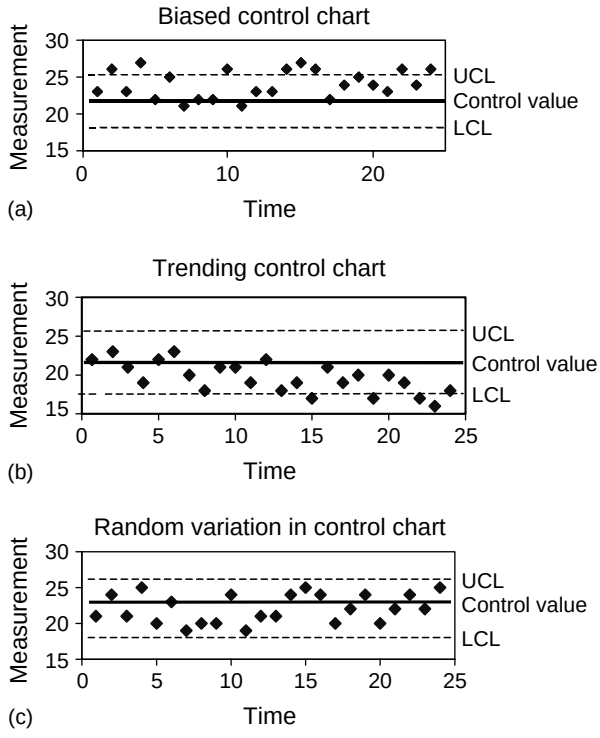

Figure 5 Commonly experienced control charts for tracking process analyzer (measurement system) performance.

the process analysis response can be tracked with control charts. This is truly the best method for determining drift in the measurement system. The calibration thus yields the starting point; once the starting point has been determined, the measurement system is judged on its bias and its repeatability.

An important practice in monitoring the performance of process analyzer systems, and hence sampling systems, is control charting. Control charting can be accomplished by measuring a standard sample periodically and plotting the results temporally. Figure 5 shows a series of control charts. Figure 5(a) shows a system exhibiting bias, where all (or the vast majority) of the responses obtained are greater than the control value. Figure 5(b) shows a system drifting over time. Although some drift may be acceptable, exceeding certain limits indicates the need for recalibration. Figure 5(c) shows a system with random error, or no net error. If the error values are of the same order as the precision of the instrument, this is a good indication that the system is performing well. If the errors in a control chart are more than three standard deviations of the noise, the system is said to be out of control, and some remedial action is required.

7.2 Long-term Performance

Tracking long-term performance of a measurement system requires documenting its calibration history. It

is often difficult to separate the performance of the sampling operation from that of the analyzer it supplies. When disparities exist between process and laboratory measurements, it is often the sampling system, not the analyzer, that has failed. Occasionally, the sampling system must be separated from the analytical instrument to determine the cause of failure. If the analyzer fails to provide an accurate measurement after being divorced from the sampling apparatus, the sampling system may still have failed in one of its missions, namely to provide the analyzer with a compatible sample. This must be determined and corrected, if necessary. If the analyzer yields adequate results after the sampling system is removed, the sampling system must be repaired or changed.

Determining the reliability of sampling systems requires a good deal of effort after the measurement system is initially started. The same is true whenever the sampling system has undergone changes. It is good practice to monitor continually the measurement system performance using control charts. This may be done using the control sample technique described earlier, or by comparing the process measurement against routine laboratory measurements by noting the differences between the two methods over time. Figure 6 shows a typical laboratory–process difference control chart where a drift is evident. As can be seen, the drift was eventually corrected through either recalibration or hardware maintenance. These types of control charts allow comparison over long periods of time and may indicate seasonal or process-related biases.[71,72] In addition, when the maintenance schedules and records are included, the overall performance is indicated.

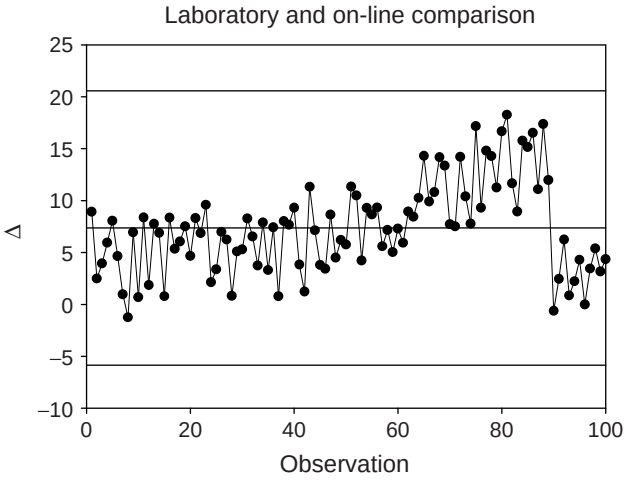

Figure 6 An example of a laboratory–process difference control chart. The discrepancy between the on- and off-line methods is represented by Δ.

8 MAINTENANCE

Good long-term performance of sampling systems is a function of understanding the sample capture, transfer, and conditioning of the process stream as it travels through the train. Understanding these processes, their failure modes, and corrective actions is vital to the measurement system performance. A good measurement system will be operational for more than 99% of the time and be available to make process measurements well over 95% of the time. Reliable measurement systems will also have a relatively long mean time between failures.[4,5]

8.1 Design Criteria

Sampling systems must be designed not only with the analyzer in mind but also with the end-user in mind. Components must be arranged so that the unit is maintainable by a skilled professional without the need for very specialized tools. Items commonly requiring attention, such as filters and valves, should be easily identifiable and accessible by maintenance personnel. The installation must meet all the necessary safety requirements of the production environment and be located in a convenient location. Recently, more attention has focused on the design of sampling systems for ease of maintenance by investigating component selection and modular construction.[73,74]

Designers must specify the most appropriate components and materials before constructing the sampling system. Maintenance concerns are lessened if the designer conducts a thorough analysis of material compatibility issues between the process stream and the sampling system. In addition, common sense must be exercised when choosing the components. If a multifunctional device can replace multiple devices, it is generally good practice to incorporate the higher-functioning device. Examples are a flowmeter with a needle valve incorporated within the stem and a vaporizing regulator.

Ideally, alterations should not need to be performed after the sampling system is installed. Occasionally, changes will need to be made in order to accommodate changes in plant design or the addition of a sampling point, or for more reliable performance. All changes must be well documented, and the associated measurement system must be calibrated and validated as if the system were new.

Components of sampling systems should be as common and readily available as possible. The parts must be of robust construction, with the ability to withstand the rigors of the process stream and plant environment. In critical applications, there should be redundancy within the sampling system, or all the parts necessary to build a complete system should be on location.

8.2 Documenting Maintenance

Documenting the maintenance of the sampling system as well as the associated analyzer allows for faster determination of sources of error within the measurement system. Preventative, routine maintenance should be performed on a regular basis and documented. This allows for adjusting the frequency of maintenance to appropriate intervals.

9 SAFETY

9.1 First and Last Thought

The safe operation of any measurement system should always be the first and last thought of any sampling system designer. The sampling system must be designed appropriately for the process and environment it is to be installed in.[75] Because its size relative to a process stream is small, sampling systems are more fragile. Safety devices incorporated within the process line, such as pressure-relief valves, must be included within the sampling system.[76] Starting the sampling system should not cause hazardous conditions, nor should the shutdown cause any harmful events. If the potential for hazardous conditions cannot be avoided, every effort must be taken to minimize the risk for personnel.[77] Release of chemicals should always be diverted away from locations where people may be found.

Only trained professionals must be allowed to work on sampling systems. These personnel must be well schooled in the safe operation of these systems. Tapping a process line while the process is operating through that line (hot tapping) should be a last-resort measure.

ACKNOWLEDGMENTS

The authors would like to acknowledge Eric Day for his help in preparing this article, especially the diagrams. We would also like to thank Sunita Saluja for her help in performing literature searches. Finally, we would like to thank AlliedSignal for its support of our efforts in process analytical chemistry.

LIST OF TERMS

At-line analysis	Samples are withdrawn from the process stream and analyzed in close proximity to the sampling point.
Backflushing	A multistream sampling technique where the selected process stream is used to flush out the nonselected sample lines.
Block and bleed	A multistream sampling technique where the nonselected streams are blocked by an isolation valve and the tubing between this valve and the process analyzer is vented or returned to the process.
Bubble point	The pressure at which volatile species in a process stream will form gas bubbles.
Bypassing	A multistream sampling technique where the pressure differential between streams provides the isolation.
Cross-sectional sampling	Extracting samples from varying positions in the process stream.
Cross-talk	Contamination between process streams in a multistream sampling system because of poor isolation of the separate streams.
Dew point	The temperature at which condensable gases in a process stream will form liquid droplets.
Double block and bleed	Similar to the block and bleed design, but the incorporation of a second isolation valve provides for even greater isolation between the nonselected process streams and the process analyzer. The tubing between isolation valves and between the second isolation valve and the process analyzer is vented or returned to the process.
Fast-loop sampling	A slip stream technique where a process pump is used to rapidly move the sample to a point near the process analyzer, and a fraction of this material is extracted for use in the analysis.
In-line analysis	The analytical probe is placed directly in the process stream with no need for extractive sampling.

In situ analysis	See *In-line analysis.*
Isokinetic sampling	In-line sampling where the velocity of the process stream is the same as the velocity of the material removed for analysis.
Joule–Thompson effect	A change in gas temperature when the gas pressure is reduced by expansion through an orifice. Many gases exhibit a temperature decrease when this occurs.
Laminar flow	Flow through piping or tubing with little or no mixing; the velocity of the material in the pipe or tube is in the direction of flow with very little perpendicular component.
Multistream sampling	The use of parallel sample flow paths and isolation valving to provide analytical results for more than one process stream using a single analyzer.
On-line analysis	A sample loop or other extractive method is used to transport a portion of the process stream to the process analyzer; or a general term used to denote an analytical method coupled directly to the process stream.
Point sampling	Extracting a single aliquot of the process material.
Process characteristic time	A measure of the frequency of process changes one wishes to detect. This time determines the minimum cycle time allowable for a process analyzer.
Response time	The total time required by the sampling system and process analyzer to provide an analytical result for control purposes.
Sample probe	A device that provides for the extraction of an aliquot of the process stream for analysis; or an in-line device that provides a direct reading of the process stream.
Sampling frequency	The rate at which samples are withdrawn from the process stream; or the rate at which a process analyzer provides an analytical result.
Sidestream	See *Slip stream.*
Slip stream	An extractive sampling line used to provide process material for an on-line analysis. The motive force used to move the sample is typically a pressure differential or a sample pump.
Sparging	An inert gas is bubbled through a sample in order to entrain volatile species in the gas. This gas is then sent to the process analyzer for analysis.
Stripping	See *Sparging*.
Transport time	The time required for the sampling system to move an aliquot of the process stream to the process analyzer.
Turbulent flow	Flow through piping characterized by directional changes in velocity and good mixing.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
FTIR	Fourier Transform Infrared

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring

Environment: Water and Waste **(Volume 4)**
Sampling Considerations for Biomonitoring

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

REFERENCES

1. K.G. Carr-Brion, J.R.P. Clarke, *Sampling Systems for Process Analysers*, Butterworth–Heinemann, Oxford, 1996.
2. E.A. Houser, *Principles of Sample Handling and Sampling Systems Design for Process Analysis*, Instrument Society of America, Pittsburgh, 1972.
3. G.D. Nichols, *On-line Process Analyzers*, John Wiley & Sons, New York, 1988.
4. R.E. Sherman, L. Rhodes, *Analytical Instrumentation*, Instrument Society of America, Raleigh, 1996.
5. K.J. Clevett, *Process Analyzer Technology*, John Wiley & Sons, New York, 1986.
6. G.K. Tyson, R.S. Wilson, 'Successful On-line Chemical Analysis', *Chem. Eng.*, **97**(12), 66–77 (1990).
7. J.J. Gunnell, 'Sampling Systems [for Analyzers]', in *Process Analytical Chemistry*, eds. F. McLennan, B.R. Kowalski, Blackie, Glasgow, UK, 14–38, 1995.
8. E. Dempsey, D. Diamond, M.R. Smyth, G. Urban, G. Jobst, I. Moser, E.M.J. Verpoorte, A. Manz, H.M. Widmer, K. Rabenstein, R. Freaney, 'Design and Development of a Miniaturized Total Chemical Analysis System for On-line Lactate and Glucose Monitoring in Biological Samples', *Anal. Chim. Acta*, **346**(3), 341–349 (1997).
9. J.J. vanderPol, M. Machnik, M. Biselli, T. PortelaKlein, C.D. deGooijer, J. Tramper, C. Wandrey, 'On-line Immunoanalysis of Monoclonal Antibodies during a Continuous Culture of Hybridoma Cells', *Cytotechnology*, **24**(1), 19–30 (1997).
10. E. Csoregi, T. Laurell, I. Katakis, A. Heller, L. Gorton, 'Online Glucose Monitoring by using Microdialysis and Amperometric Detection based on Wired Glucose Oxidase in Carbon Paste', *Mikrochim. Acta*, **121**(1–4), 31–40 (1995).
11. R.D. Moen, T.W. Nolan, L.P. Provost, *Improving Quality Through Planned Experimentation*, McGraw-Hill, Boston, 1991.
12. G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters*, John Wiley & Sons, New York, 1978.
13. American Society for Testing and Materials, *Standard Practice for Sampling Industrial Chemicals*, ASTM E 300-92, ASTM, Philadelphia, PA, 1992.
14. M.W. Sweeney, 'Instrumentation and Automation [in Wastewater Treatment]', *Water Environ. Res.*, **65**(4), 374–377 (1993).
15. P. Shanahan, 'Mixing and Transport [of Water Pollutants]', *Water Environ. Res.*, **64**(4), 514–523 (1992).
16. American Society for Testing and Materials, *Standard Practice for Acceptance of Evidence Based on the Results of Probability Sampling*, ASTM E 141-91, ASTM, Philadelphia, PA, 1991.
17. P.M. Girling, 'Resolution of Differences Between Laboratory and Process Analysers', *ISA Anal. Instrum.*, **17**, 16–19 (1979).
18. P.M. Gy, 'The Sampling of Particulate Materials – A General Theory', *Int. J. Miner. Process*, **3**, 289–312 (1976).
19. R.J. Akers, J.I.T. Stenhouse, 'A Theoretical and Experimental Study of Particle Sampling from Hydraulic Systems', *Proc. Inst. Mech. Eng.*, **72/96**, 45–54 (1976).
20. H.A. Kraftson, E.L. Weiss, 'Doublt Filtration Gas Sampling Systems for Advanced Applications', *Anal. Instrum.*, **7**, 226 (1969).
21. G.W. Ewing, *Analytical Instrumentation Handbook*, Marcel Dekker, New York, 1997.
22. H. Telle, 'In-situ Assaying of Materials using Laser Analytical Spectroscopy', *Proc. SPIE–Int. Soc. Opt. Eng.*, **2730**, 160–170 (1996).
23. T.M. Niemczyk, M. Delgado-Lopez, C.D. Newman, 'Multichannel Raman Spectroscopy Tackles Industrial Problems', *Laser Focus World*, **29**(3), 85–98 (1993).
24. Z. Mao, J. Demirgian, A. Mathew, R. Hyre, 'Use of Fourier Transform Infrared Spectrometry as a Continuous Emission Monitor', *Waste Manage.*, **15**(8), 567–577 (1995).
25. K.L. Gallaher, 'Reformulated Gasoline Analysis using FT-NIR Spectroscopy', *Proc. 1997 Air Waste Management Assoc.*, **90**, 608–613, 1997.
26. J. Crandall, 'Sample Systems in Process Near-infrared Spectroscopy I', *J. NIR News*, **5**(4), 5–8 (1994).
27. L.K. Barnes, 'Sampling Systems for Gas Analysers', *Instrum. Technol.*, **19**(1), 53–59 (1972).
28. L.K. Barnes, 'How to Design Gas Sampling Systems', *Hydrocarb. Process. Petrol. Refin.*, **42**, 157–162 (1964).
29. J.R.P. Clarke, 'Sampling for On-line Analysis', *Meas. Control*, **3**(8), 241–244 (1970).
30. J.R.P. Clarke, 'Some Tricks of the Sampling Trade', *Process Control Qual.*, **4**, 1–11 (1992).
31. R.H. Perry, D.W. Green, *Perry's Chemical Engineer's Handbook*, McGraw-Hill, New York, 1997.
32. B.G. Liptak, K. Venczel, *Instrument Engineers' Handbook*, Chilton, New York, 1969.
33. D.M. Considine, *Process/Industrial Instruments and Controls Handbook*, McGraw-Hill, New York, 1993.
34. E.A. Houser, R. Villalobos, 'Analysis of Catalytic Regenerator Flue Gas', *Anal. Instrum.*, **8**(Session III), paper 5 (1970).
35. E.A. Houser, 'Switching Process Samples', *Instrum. Technol.*, **16**(5), 51–54 (1969).
36. W.B. Grant, R.H. Kagann, W.A. McClenny, 'Optical Remote Measurement of Toxic Gases', *J. Air Waste Manage. Assoc.*, **42**(1), 18–30 (1992).
37. J.D. Tate, P. Chauvel, K. Taylor, 'Monitoring Ambient Air with Optical Sensing Systems', *Chem. Eng.*, **104**, 110–114 (1997).
38. American Society for Testing and Materials, *Standard Practice Planning the Sampling of the Ambient Atmosphere*, ASTM D 1357-95, ASTM, Philadelphia, PA, 1995.
39. S. Hiraoka, 'Current Status and Task of Gas Analysis Technology for Atmospheric Contamination Control', *Nensho Kenkyu*, **114**, 19–24 (1998).

40. L. Driscoll, 'Coping with High Pressure Letdown', *Chem. Eng.*, **83**, 113–118 (1976).
41. P. Staahlberg, P. Oesch, E. Leppaemaeki, A. Moilanen, M. Nieminen, J. Korhonen, 'Development of Sampling Systems and Special Analyses for Pressurized Gasification Processes', *LIEKKI 2 – Annual Review*, eds. J. Matinlinna, M. Hupa, Abo Akademi University Press, Arsbroken, 631–651, 1996.
42. American Society for Testing and Materials, *Standard Practice for Flow Control and Temperature Control for On-line Water Sampling and Analysis*, ASTM D 5540-94a, ASTM, Philadelphia, PA, 1994.
43. C.O. Bertram, V.J. Desai, E. Interess, 'Designing Steam Tracing', *Chem. Eng.*, **3 Apr.**, 74–80 (1972).
44. J.W. Sugar, J.H. Brubaker, 'Sampling for Wastewater Analysers Part 1, a Systematic Approach', *Instrum. Technol.*, **21**(6), 27–32 (1973).
45. Z.-l. Zhi, A. Rios, M. Valcarcel, 'Direct Processing and Analysis of Solid and Other Complex Samples with Automatic Flow Injection Systems', *Crit. Rev. Anal. Chem.*, **26**(4), 239–260 (1996).
46. P. Massacci, 'Minerals and Particles Image Analysis for Mineral Characterization and Process Modeling', in *Proc. Conf. on Mineral Processing: Recent Advances and Future Trends*, eds. S.P. Mehrotra, R. Shekhar, Allied Publishers, New Delhi, India, 254–261, 1995.
47. K.G. Carr-Brion, 'Sampling Systems for On-line Analysis of Materials that are Difficult to Handle', *Anal. Chim. Acta*, **238**, 11–16 (1990).
48. B.J. Hulley, 'Sampling and Sample Conditioning in On-line Fertilizer Analysis', *Chem. Eng.*, **77**, 410–413 (1970).
49. B.G. Liptak, 'Process Instrumentation for Slurries and Viscous Materials', *Chem. Eng.*, **74**, 176–179 (1967).
50. H. Nasr-El-Din, C.A. Shook, M.M. Esmail, 'Isokinetic Probe Sampling from Slurry Pipelines', *Can. J. Chem. Eng.*, **63**, 179–185 (1984).
51. N.A. Esmen, 'Nozzle Sampling Bias', *Am. Ind. Hyg.*, **36**(6), 362–364 (1976).
52. L.L. Sloss, 'Emissions of Air Toxics from Coal Combustion: Measurement and Control', *Proc. Int. Tech. Conf. Coal Util. Fuel Syst.*, **20**, 35–44 (1995).
53. G.C. Vezzoli, 'Systems for Analysis of Organic Substances Contained in Vehicle Emissions and in the Atmosphere', *Tecnol. Chim.*, **16**(8), 63–70 (1996).
54. B.B. Ferguson, R.F. Weston, 'Stationary Source Sampling', in *Industrial Environmental Control*, 2nd edition, ed. A.M. Springer, TAPPI, Atlanta, GA, 621–643, 1993.
55. K.H. Karstensen, 'Flow-injection Analysis (FIA) in Industrial Process Control', *Sent. Ind.*, **51**(8), 20–21, 23–24 (1991).
56. R.D. Hull, R.E. Malick, J.G. Dorsey, 'Dispersion Phenomena in Flow-injection Systems', *Anal. Chim. Acta*, **267**(1), 1–24 (1992).
57. K.E. Neu, 'Continuous Flow On-line Sampling Systems: On-line Process Monitoring for Soluble Constituent Analysis in WWTP's Begins with an Effective Filtrate Sample Production System™', *69th Annual Conference and Exhibition by the Water Environment Federation*, 655–666, Vol. 4, 1996.
58. G.J. Lynch, 'Use of On-line Ion Chromatography for High-temperature and High-pressure Reaction Studies', *Process Control Qual.*, **7**(1), 51–61 (1995).
59. B. Baaske, U. Telgheder, 'Automated Sampling System for the Direct Determination of Trace Amounts of Heavy Metals in Gaseous Hydrogen Chloride by Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **10**(12), 1077–1080 (1995).
60. J.G. Watson, J.C. Chow, 'Particle and Gas Measurements on Filters', in *Environmental Sampling: Trace Analysis*, ed. B. Market, VCH, Weinheim, Germany, 125–161, 1994.
61. K.W. Baughman, D.H. Love, 'Industrial Hygiene Chemistry. The Method Development Approach to Air Analysis', *Anal. Chem.*, **65**(10), 480A–484A, 486A–487A (1993).
62. N.C. Vandemerbel, H. Lingeman, U.A.T. Brinkman, 'Sampling and Analytical Strategies in Online Bioprocess Monitoring and Control', *J. Chromatogr. Sect. A*, **725**(1), 13–27 (1996).
63. American Society for Testing and Materials, *Standard Practice for Sampling and Sample Preparation of Iron Ores and Related Materials*, ASTM E 877-93, ASTM, Philadelphia, PA, 1993.
64. Y. Xiao, L. Halappa, 'Thermodynamic Properties of Chromium Bearing Slags and Minerals', *Lab. Mater. Process. Powder Metall.*, **B118**, 1–52 (1996).
65. Y. Yamashita, M. Kuwashima, T. Nonaka, M. Suzuki, 'Online Measurement of Cell Size Distribution and Concentration of Yeast by Image Processing', *J. Chem. Eng. Jpn.*, **26**(6), 615–619 (1993).
66. R.C. Woodward, 'On-line Bulk Material Analysis: The Alternative to Mechanical Sampling', *Bulk Solids Handling*, **11**(2), 495–497 (1991).
67. M.H. Ramsey, 'Sampling as a Source of Measurement Uncertainty: Techniques for Quantification and Comparison with Analytical Sources', *J. Anal. At. Spectrom.*, **13**(2), 97–104 (1998).
68. T. Trantow, R. Herzog, 'Causes of Systematic Errors in Stability Testing', *Pharm. Ind.*, **57**(2), 170–173 (1995).
69. J.G. Converse, 'Real Time Validation of On-line Measurements Utilizing Internal Standard Introduction', *Proc. 41st Annu. ISA Anal. Div. Symp.*, **29**, 1–10 (1996).
70. L.F. Peveto, J.G. Converse, 'Decisions to Change Analyzer Calibration Based on Statistical Quality Control Charts', *ISA Trans.*, **30**(1), 79–91 (1991).
71. American Society for Testing and Materials, *Standard Practice for Validation of Process Stream Analyzers*, ASTM D 3764-92, ASTM, Philadelphia, PA, 1992.
72. American Society for Testing and Materials, *Standard Practice for Applying Statistical Quality Assurance*

Techniques to Evaluate Analytical Measurement System Performance, ASTM D 6299-99, ASTM, Philadelphia, PA, 1999.
73. D.E. Podkulski, 'How do New Process Analyzers Measure up?', *Chem Eng. Prog.*, **93**, 33–46 (1997).
74. D.P. Mayeaux, 'Modular Sample Conditioning System', *US Patent 5 841 036*, 1998.
75. American Society of Mechanical Engineers (ASME), *Boiler and Pressure Vessel Code: Part 2, Materials Specification; Part 5, Non-Destructive Testing; Part 8, Pressure Vessels Div. 1. and Div. 2 – Alternative Rules*, ASME, New York, 1974.
76. P.T. Bodurtha, *Industrial Explosion Prevention and Protection*, McGraw-Hill, New York, 1980.
77. F.P. Lees, *Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control*, Butterworths, London, 1983.

Titration Techniques for Process Analysis

Tore K. Fossum
Tor Systems, Friendswood, USA

Duane K. Wolcott
Global FIA, Inc., Gig Harbor, USA

1 Introduction 1
2 Theory 1
3 Instrumentation 2
3.1 Overview 2
3.2 Sample Pretreatment 2
3.3 Sample Handling 2
3.4 Titrant Delivery 6
3.5 Titration Control 6
3.6 Titration Curve Shapes 7
3.7 Sensors 7
3.8 Cleanout 9
3.9 Communication 10
3.10 Waste Generation 10
3.11 Maintenance 10
3.12 Isolation and Explosion Proofing 10
3.13 Commercially Available Titrators 10
4 Chemistry 10
4.1 Reaction Kinetics of Titrant–Analyte Reaction 10
4.2 Solvent 10
4.3 Sensor Kinetics 11
4.4 Indicators 11
4.5 Materials Compatibility 11
4.6 Environmental Impact 11
5 Applications 11
5.1 Aqueous Acid–Base Titrations 11
5.2 Nonaqueous Acid–Base Titrations 12
5.3 Redox Titrations 12
5.4 Argentometry 13
5.5 Complexometric Titrations 13
Acknowledgments 14
Abbreviations and Acronyms 15
Related Articles 15
References 15

Titration remains a method of choice for quantitation and assay because the techniques are simple and the repeatability and accuracy can be better than with other analytical methods. Titration has moved from the laboratory bench to the process in order to obtain more timely information for the control of the process. This has become possible with the development of sampling techniques for getting clean samples from a pipe or tank, sensors suitable for remote measurements, highly precise automated titrant delivery pumps, and microprocessor-based controllers adapting the time-honored methodologies of the bench chemist.

Topics discussed in this paper include theory, sampling techniques, titration control, detection and applications using pH, argentometric, redox, and photometric sensors.

1 INTRODUCTION

Many chemical plant operators desire on-line titrimetric analysis, because the analysis of samples directly in the process stream makes more timely data available for adjusting the process. A secondary consideration may be to improve overall efficiency, by automating manually supervised or manually performed analyses. In some cases a third consideration is that some samples deteriorate with time or exposure, and thus a shortened time to analysis improves the quality of the data. A fourth consideration could be to prevent contact between personnel and highly hazardous samples or reagents.

Titration is often the most accurate method of quantification. It may be the only practical method, or it may be the easiest method, or even the fastest of the available analytical methods.

Many analyses done titrimetically in the laboratory can be done on line. Ideal conditions are homogeneous single-phase fluid samples at atmospheric pressure and room temperature, which are not excessively aggressive, and which react stoichiometrically with a titrant. Particles in the stream, high temperatures and pressures, and aggressive or corrosive characteristics of the sample complicate the sample pretreatment. Solid samples are not good candidates for process titration. The difficulty is with the repeatable measurement of a quantity of sample, because solid samples can normally not be measured by volume and must be weighed on a balance. Assays of solid samples are better done individually in the laboratory with manual or with automated titration techniques. Many attempts have been made to automate the weighing process using robotic systems, but these are so maintenance intensive and complex as to limit them to supervised laboratory environments.

2 THEORY

Titration is a batch process. The basic principle of titrimetric analysis is that the portion of interest of the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

sample will react quantitatively with a titrant of known concentration. The course of the titration is tracked with a sensor that detects the equivalence point, at which sample is fully reacted. There must be a mechanism for measuring the sample quantity and delivering it to the titration vessel. There must also be a means of measuring the amount of titrant added. When the analysis is finished, the vessel is flushed to prepare for the next measurement cycle. The calculation assumes that the equivalent amount of sample is equal to the equivalent amount of titrant (Equations 1 and 2):

$$N_S \times V_S = N_T \times V_T \tag{1}$$

or

$$\text{milliequivalents of sample} = N_T \times V_T \tag{2}$$

where N is the normality or equivalents per liter, and V is the volume in milliliters. A typical expression of concentration may then be written (Equation 3) as

$$\text{content}(\%) = \frac{N_T \times V_T \times M}{10 \times Z \times (V_S \text{ or } W_S)} \tag{3}$$

where M is the molecular weight and Z is the valency change, V_S represents the sample volume in milliliters (in which case one calculates percent by volume), and W_S represents the weight in grams (in which case weight percent is calculated). Introducing the sample density δ in grams per cubic centimeter, one can calculate weight percent using a volumetric sample (Equation 4):

$$\text{content}(\%) = \frac{N_T \times V_T \times M}{10 \times Z \times V_S \times \delta} \tag{4}$$

3 INSTRUMENTATION

3.1 Overview

The titration system must perform the processes of sample treatment, sample measurement, addition of solvents and reagents, the titration consisting of two interacting processes, sensor measurement and titrant addition, and cleanout. A control computer or built-in microprocessor controls these processes and provides an interface to the operator and to the process computer (Figure 1).

3.2 Sample Pretreatment

Precise analytical titration requires known amounts of homogeneous sample. In the real world of pipes and flowing streams, samples often require some pretreatment to obtain a measurable homogeneous sample. This may include filtration, separation, pressure reduction, degassing, and temperature adjustment. A flowing stream may have solid particles that can destroy valve surfaces, it could have dissolved salts that will precipitate out as the temperature drops, it may be under pressure and when that pressure is released degassing could occur, and it may have a hydrocarbon phase and an aqueous phase – these may need to be separated and analysis done on one or the other.

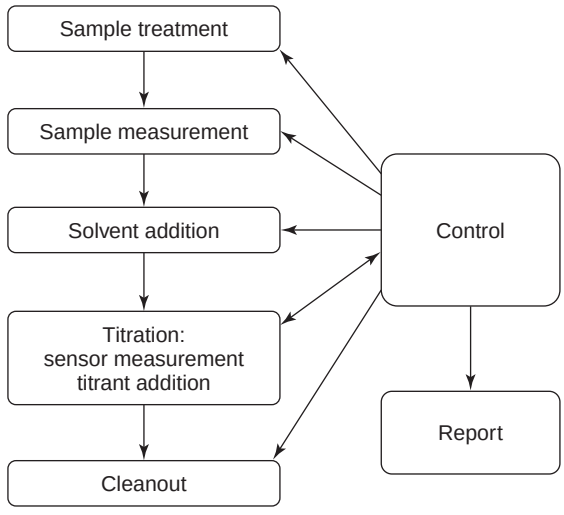

Figure 1 Functions of an on-line titrator.

Extracting a representative sample from a process stream for subsequent analysis involves some compromises. It is rare that a sample is sufficiently pristine that it can be fed directly to the inlet of the desired instrumentation. On-line titration is no different in this regard. To derive meaningful information about the process the sample must often be purified without corrupting the analyte, or by changing it in such fashion that the degree of difference from 'true process conditions' is known.

3.3 Sample Handling

3.3.1 Extraction

The first step in the chain of operations to obtaining meaningful on-line measurements is determining how to gain access to the process stream. It is desirable to design a sample-handling system that uses the minimum number of additional devices. There are many options.

Process Pump Probably the most desirable option is to tap across an existing process pump. This approach takes advantage of the fact that the capital expenditure (and maintenance responsibility) for the pump have been catered for. In such an arrangement, a sample port is added to exit piping, routed to the desired analyzer location, and back to an inlet port on the upstream side of the pump (Figure 2). Samples obtained in this manner are well mixed by virtue of their passage through the

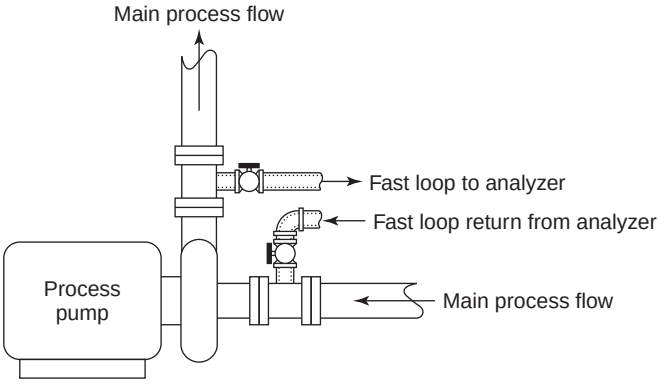

Figure 2 Sampling of tapping across an existing process pump.

pump. The main disadvantage of this approach is that large process changes are recirculated through the fast loop and can lead to an increased analysis lag time. If such lag times are a problem, the recycled sample must be reintroduced into the process downstream of the take-off point at a low-pressure point in the process.

Process Pressure If no convenient pump is available, but the process is under sufficient pressure, it may be possible to introduce an in-line connection at a high-pressure point. The existing process pressure provides the drive energy to force the process sample to the desired analyzer location, and back to a lower-pressure process point. It is tempting to make use of existing process connection points (blind flanges) to provide the connection points. However, extreme care must be taken in such cases to avoid various types of sampling problems. Blind side-arms on piping runs typically are mounted at right angles to existing piping. If the piping run to which connection is desired runs horizontally, such right-angle mounting may place the side-arm on top of the pipe, on the side of the pipe, or on the bottom of the pipe (Figure 3).

Any or all of these may present potential sampling problems. If connection is to the top of the pipe, gas bubbles will be preferentially routed into the sample line. If connection is to the bottom of the pipe, suspended

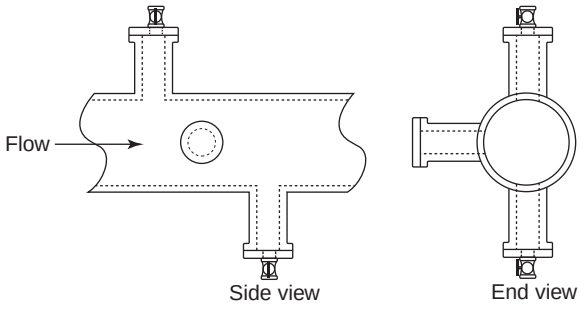

Figure 3 Side-arm sampling.

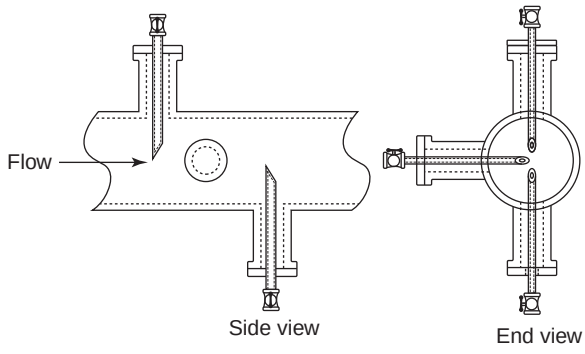

Figure 4 Stinger sampling.

solids will be routed preferentially into the sample line. The preferred connection is to a side-arm mounted on the side of the pipe. Even then, there is the probability that the volume of the side-arm will introduce significant time lags, as it represents a large, stagnant volume of process fluid that must be diluted out by the sample flow. With the addition of a properly designed stinger, any of the aforementioned side-arm piping orientations becomes a viable sample point (Figure 4). Note the shape of the end of the stinger, and its orientation with respect to process flow. This arrangement allows some extra discrimination against bubbles and/or particles flowing down the pipe. The stingers illustrated are all for the take-off point of the sample flow – the return point is not shown, being located further down the process or at some other low-pressure point.

Dedicated Pump The use of a dedicated pump to provide flow through the fast loop to the analyzer is usually not a popular option. It represents both an additional capital cost, and additional maintenance costs. Also, when such a pump fails, its replacement will always be a low-priority item compared to other maintenance emergencies, with increased analyzer downtime and decreased analyzer reliability as a result. However, a dedicated pump does provide several advantages in that it can be much smaller and less expensive, and does not require multiple taps into the process for extraction and return points (Figure 5).

The arrangement shown in Figure 5 shows the use of a flat plate style cross-washed inertial filter. Systems of this sort are usually plumbed up with piping (and dedicated pump) sufficiently large to allow the largest solids in the process to completely traverse the fast loop piping. This usually requires that the fast loop be three quarters to one inch (16 to 25 mm) pipe. With the addition of some automated valving and a different stinger style, this arrangement becomes a fairly sophisticated means of sample prescreening for larger particles – thereby allowing the use of a smaller cross-washed filter and

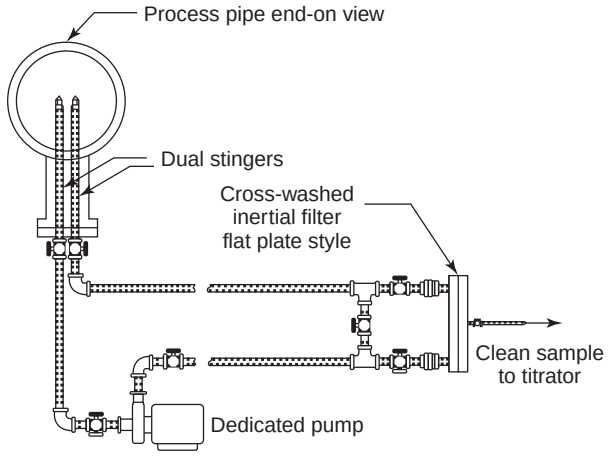

Figure 5 Dual stinger–dedicated pump sampling.

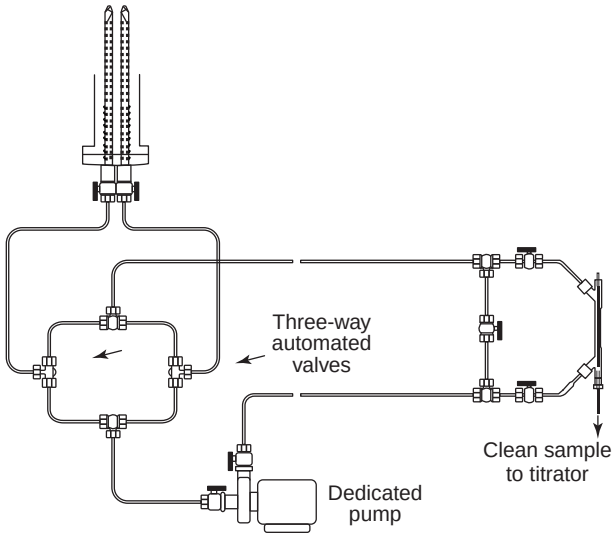

Figure 6 Dedicated pump with backwash loop.

stingers (Figure 6) and significantly smaller piping or tubing in the fast-loop plumbing.

The arrangement in Figure 6 provides automatic backwashing of the stingers with the pumped flow. In use, each stinger alternately serves as sample take-off and/or return, depending on the position of the three-way valves. However, flow through the rest of the fast-loop tubing remains monodirectional to and through the smaller, tube-in-tube cross-washed inertial filter.

The dual stinger sampling point with a dedicated pump and cross-washed filter will work well for a tank, a basin, or for surface water. The return stinger should be downstream of the sampling stinger.

3.3.2 Transport

Extraction of the sample from the process pipe is only the first step in the chain of events that must occur before quantitative information can be derived from the on-line titrimetric analysis. Once withdrawn from the process, the sample must be transported to the physical location of the instrument. Whether this transport is done with a process pump, process pressure, or dedicated analyzer pump, several factors must be observed.

Sample Integrity It is critical that the samples do not change significantly, or that any change is not detrimental to the quality of the analytical results or to the reliability of the instrument. In the sample transport step, the most likely parameter to change drastically is temperature. Many processes operate at elevated temperatures, and the extraction and transport processes can result in fairly drastic drops in temperature. As many process streams of interest in on-line titration may be fairly concentrated solutions, such temperature drops can result in solid material precipitating out of solution, with catastrophic effect on the sample system. Temperature drops can also result in shifts in chemical equilibrium, which can lead to a measurement result that differs from the true concentration of the target species in the process. Sometimes the reverse situation occurs – a grab sample is taken from the process and analyzed, yielding a substantially different result from the on-line instrument. In many cases, the error arises from a chemical change that has occurred in the grab sample. Finally, depending on the climatic conditions, cooling in transfer lines may be sufficient to allow sample freezing to occur. Depending on the specific process and situation, it may be necessary to insulate or provide heat to the transfer lines.

Time Lag The introduction of fast-loop piping (and other sample processing devices) introduces various time lags in the analytical results. An exhaustive treatment of the possible sources of lag time is beyond the scope of this article, but such time lags must be a consideration when designing and using the sample handling system. The closer the analyzer to the sampling point, the shorter the sample travel and the smaller the time lag, the better.

3.3.3 Physical Conditioning

Successful transport of the sample from the process is only the second step in the chain of events that must occur before quantitative information can be derived from the on-line titrimetric instrument. For most samples, further conditioning is required.

Filtration The presence of suspended solids is possible in any process system. Even high-purity water systems may have particulate matter resulting from piping or pump wear, so the addition of a filtration step ranges

from being highly desirable to absolutely essential. If the sample stream contains very low levels of particulate matter, it is possible to use in-line cartridge filters and simply change the filter element based on a time interval or by measuring pressure drop across the filter. However, if the levels of suspended particulate are higher, then cartridge-type filters will clog up too quickly, resulting in unacceptable maintenance requirements. The criteria for effective filtration must include not only how effective the particular filter is for removing particles above a certain size, but how long the time interval is between any sort of maintenance (changeout or backwash). The most useful type of filter is a cross-washed inertial filter (Figures 5 and 6), as these provide greatest reliability and minimize maintenance.

Gas Removal Many sample streams either contain bubbles, or contain dissolved species that form gas bubbles on removal from the process (usually due to the drop in pressure from process pressure towards ambient pressure. If the initial quantity of sample to be analyzed is determined volumetrically (as for instance by using a sample-loop injection valve), the presence of bubbles can contribute to significant analytical errors. There are two approaches to removing bubbles – bubble separation based on gravity and/or wettability, and membrane extraction. Gravity-based methods can handle very high quantities of formed gas bubbles, but are relatively ineffective at removing dissolved gases. The inverse is true for membrane extraction methods. Probably the best approach is to combine the two methods in series, with a gravity-separation device to remove the vast majority of formed bubbles, followed by a membrane device to remove any remaining dissolved gas. These two approaches are less suitable for cases where foam is present. Foams can sometimes be broken by forcing the foam and liquid through a porous element, and doing a gravity separation downstream of the unit.

Dissolution The sample stream may contain solid material whose presence contributes to the measured variable (an example is an alkaline stream containing sodium bicarbonate, where the desired titration measurement is total alkalinity). Sampling these types of streams is particularly difficult. What is usually done is to use some sort of valve to isolate a relatively large volume of the sample containing both liquid and solids for transport to the titration vessel. A standpipe sampler is shown in Figure 7.

In operation, the sample flows from the process into the inner cylinder of the device, filling it and overflowing to the drain. At some desired time interval, the three-way valve is actuated, blocking flow in from the process, and dropping an isolated aliquot of sample into the titration vessel.

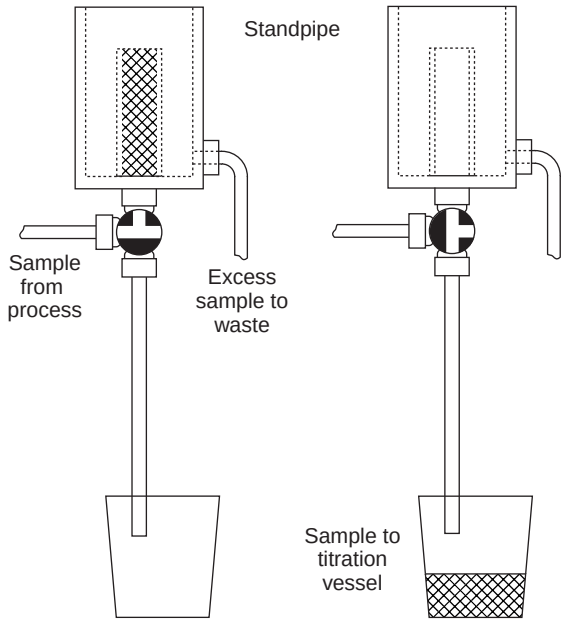

Figure 7 Standpipe sampler.

3.3.4 Sample Measurement

One simple way to measure the sample volume is to fill the titration vessel with sample, then aspirate down to a fixed volume (Figure 8). However, this is less precise than measuring with a valve-operated pipette, using a valved

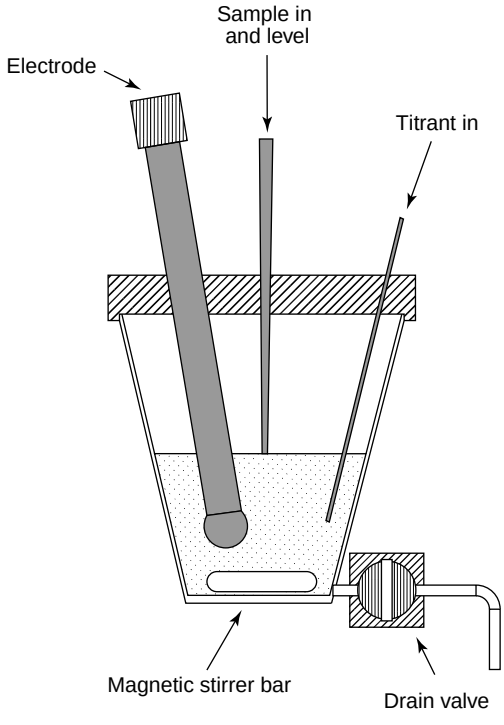

Figure 8 Titration vessel with aspirated sample.

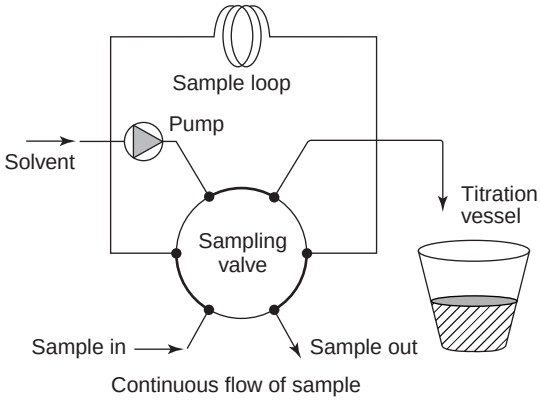

Figure 9 Volumetric sample loop with two-position valve.

sample loop or with a motor-driven piston burette. Using a piston burette has the drawback that, as it has some dead volume, numerous complete strokes are required to completely purge the syringe. The piston burette is subject to fouling and the sample must be free of particulates and precipitatable components. It has the advantages that the sample is measured quite precisely and that the volume of sample taken can be varied to accommodate various sample strengths. A valve-operated pipette is used for fixed-volume samples of moderate to large volume (5–50 mL). It is similar in operation to the Standpipe Sampler (see Figure 7). A two-position six-port injection valve can be used to set up a sample loop with the volume dependent on the length and internal diameter of the tube used to make the loop (Figure 9). In the first position, flow comes from the process through the loop and back to the process. To deliver the volume of the loop, the valve is moved to the second position and water or solvent pumped through to flush the contents into the titration vessel. Very small volumes can be sampled accurately with this method, typically from 0.1 mL up to 5 mL.

3.4 Titrant Delivery

Two types of pumps are commonly used in titrators. The glass-barrelled burette with a Teflon® piston can deliver very precise volumes at a controlled speed. The precision of delivery is typically better than 0.05% relative standard deviation (RSD) on a full burette volume. A valve is required to direct the flow from the reservoir and to the delivery tip. The other type of pump has a ceramic sleeve with a rotating piston. The piston has a cut-away face that exposes, alternately as it travels in or out, the outlet port and inlet port. The precision achieved by this pump is not as good as with the glass-barrelled syringe-type pump, but it will purge thoroughly with less volume, has fewer moving parts and can be run with one motor. The glass-barrelled syringe-type of pump requires two motors, one to drive the piston with a precision spiral-cut gear, and the other to turn the valve.

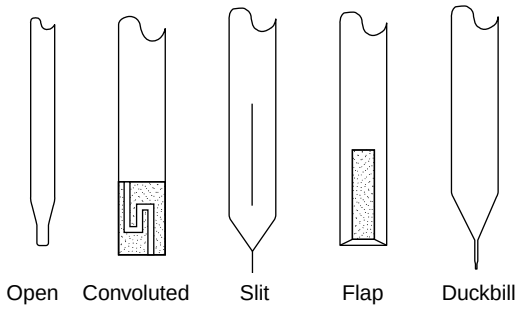

Figure 10 Titrant dispenser tip types.

The ultimate in precision is only achieved when there is a controlled release at the dispensing tip. An open tube type of tip, even of a very small diameter, will allow exchange of titrant with the solvent of the titration vessel. Various methods have been used to control this exchange (Figure 10). The tip path can be run in a convoluted fashion, reversing direction of a small-diameter tube; alternatively, an insert can be put in the tip with a flapper type of valve. Another approach is to have a straight Teflon® tube with a razor-cut longitudinal slit which is normally closed but opens under the pressure of the pumped titrant. Another way is to squeeze shut the heated end of a Teflon® tube so that the tube closes but does not seal. The two flaps appear a bit like the bill of a duck, and open under pressure.

3.5 Titration Control

The earliest automatic titrators simply delivered titrant until a preset potential measured on an electrode was reached. The volume was read and the result calculated. Somewhat later, volume delivery was coordinated with a chart recorder and the electrode reading drawn on a moving chart. The chart was evaluated for an inflection point on the curve. In state-of-the-art titration systems, computers control the titrant delivery and evaluate the data. A set point evaluation takes the volume at a particular pH or potential read by a sensor, either an electrode or a photometric detector. Typically there is some control, to deliver rapidly when far from the set point and to slow down as it is approached. This can be achieved by adding consistently small increments of titrant and allowing longer time between increments as the set point is approached. Another way is to vary the size of increments in a dynamic fashion, based on the slope of the curve. When the slope is flat, the increments are large, and when the slope increases, the increments decrease in volume. In this fashion, titrant is delivered rapidly when far from the region of interest and slowly near the

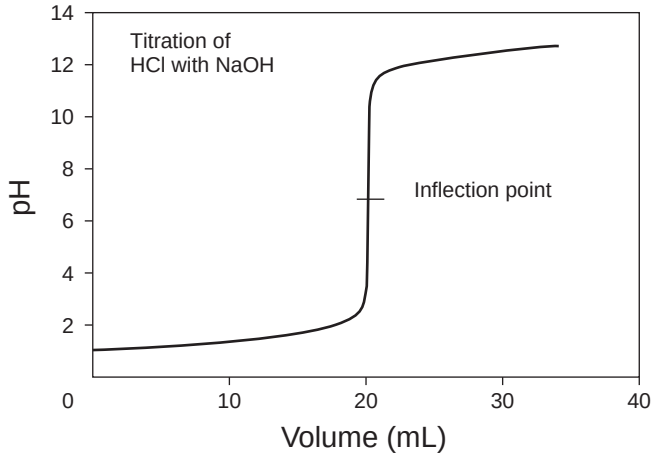

Figure 11 Titration with inflection point marked.

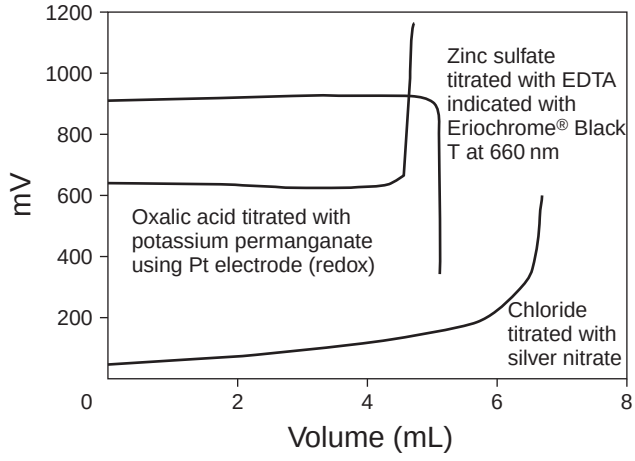

Figure 12 Various titration curves.

equivalence point. The end-point evaluation is based on being able to consistently have a set point or to find an inflection at or near the equivalence point, that point where the titrant and the analyte are in equal valence proportions. The timing of the titrant delivery may be strictly on a time basis, proportionally to approaching a set point, or based on the stability of the measured signal (equilibrium control). The more sophisticated titrators allow a combination of these timing schemes, such as the signal attaining a set equilibrium within a time window. The next increment is delivered either when the electrode signal is stable or at the end of the time window.

Inflections may be evaluated in several ways. The most common evaluation for potentiometrically measured titrations and many photometric titrations is to the center of an inflection. That is the point where the curvature changes direction (Figure 11). In some photometric titrations, the point where either the color begins to change or ceases to change is the desired point. The evaluation is for the intersection of two lines, the steepest part of the curve and the straight line where there is no color change before or after the color change, respectively. This type of evaluation can also be applied in conductivity titrations, when the rate of conductivity changes relative to titrant delivery changes because an absorbing element in the sample is saturated. For certain types of turbidimetrically detected surfactant titrations, surfactant activity can be detected by the maximum or minimum of turbidity generated by an immiscible two-phase mixture.

3.6 Titration Curve Shapes

The shape of the titration curve affects how well the inflection point can be anticipated. Well-behaved acid–base titrations have a predictable curve and the equivalence point can be estimated from a few points at the beginning of the titration. The titration can be conducted rapidly at first, then slowed proportionately as the inflection point approaches. If time is of no consequence, one can titrate a reaction with slow kinetics over a long period of time. However, in industrial analysis, time is often of the essence, and a rapid titration is pressed with some sacrifice of accuracy. In many oxidation–reduction titrations indicated with a platinum combination electrode, only a small shift in the electrode potential heralds the coming inflection point, which usually has a large jump in potential. The titration thus must be done with smaller increments than might be used on a comparable acid–base titration. This is often even more pronounced with dye indicators (Figure 12). Indicators that develop an absorbance might at first be diluted with the addition of titrant, thus increasing the transmittance. Then as color begins to change, the transmittance decreases abruptly, thus producing a titration curve with practically no possibility of anticipating the end-point. Therefore, the titration must be done with small increments in order to get an accurate result.

3.7 Sensors

Sensors can be potentiometric electrodes, conductivity detectors, polarizing electrodes, or photodetectors. The most commonly used potentiometric electrode is the glass electrode pair for detection of hydrogen ion activity, or pH. Electrodes are always in pairs; there must be a reference electrode and a sensing electrode. Today combination electrodes are made which have the reference electrode built into the same measuring rod that holds the glass sensor, but there are still two electrode half cells. The hydrogen electrode is defined to have a potential of zero, against which all other electrodes are referenced. However, it is not easy to use and is usually only used for certain types of research.

3.7.1 Reference Electrodes

The reference electrodes in common use are the silver–silver chloride half cell, the calomel half cell, and the Orion Ross™ reference electrode (Orion Research, Beverly, MA, USA). The calomel mercury cells are not much used because of the hazards associated with mercury. Most reference half cells are based on the reversible silver–silver chloride reaction on a silver wire suspended in an electrolyte solution. A low-flow ceramic frit or ground glass junction allows electrical charge transport via ionic travel between the electrolyte solution inside and the measured solution outside of the electrode. A clear flow path is essential as a clogged flow cuts the flow of ions and thus, charge, affecting the measured signal. Various modifications of the standard ceramic frit plug can improve the flow. These include an annular fritted ring, both fixed or movable ground glass joints, and a spring-tensioned movable glass joint. The standard electrolyte is potassium chloride saturated with silver chloride for the silver–silver chloride reference electrode, and potassium chloride for the Orion Ross™ reference electrode or calomel half-cell. If other electrolytes are required for compatibility with some part of the measured solution, a double-junction reference may be used where the upper chamber has the silver chloride-coated silver reference wire and potassium chloride saturated with silver chloride. A ceramic frit junction connects to the lower chamber, which may be filled with an appropriate electrolyte such as potassium chloride, lithium chloride in alcohol, potassium nitrate, or potassium perchloride.

The most common causes of reference electrode failure are clogging of the flow path by a solid-forming reactant such as silver ion precipitating with chloride, by clogging with oils or polymers, and by contamination with ammonia which dissolves the silver chloride on the silver wire. A silver–silver chloride reference electrode filled with a potassium chloride solution must have the electrolyte saturated with silver chloride; otherwise, the potassium chloride solution will dissolve the silver chloride. This electrolyte must be avoided for the Orion Ross™ reference electrode or the calomel reference electrode as the silver will poison these reference electrodes.

3.7.2 Glass Electrodes

Glass electrodes (pH electrodes) are sensitive to the presence of hydronium ions. The pH-sensitive glass is composed of silicon oxide doped with various alkaline earth metals. The addition of lithium or lanthanum to the silicon improves the selectivity and response to the hydrogen ion. When this glass is placed in aqueous media, the surface swells to form a hydrated layer. Proton transfer occurs at the phase boundary between the glass and the hydrated layer. A potential develops which is proportional to the hydrogen ion activity in the sample. Because this potential is a very weak signal, the measurement amplifier must have a very high input impedance so that it will draw very little current, so as not to corrupt the signal.

Care must be taken to ensure proper hydration of the glass. Protein, oil or polymers, which may coat the glass, must be removed, either mechanically by wiping or through the use of chemical solvents. If the electrode is used in media that dry the hydrated layer, such as the petroleum acid number solvent, which is composed of 49.5% toluene, 49.5% isopropyl alcohol, and 1% water, the hydration must be restored by soaking the electrode in water for 3–5 min between samples. Proper maintenance of the pH sensor will ensure accuracy and promote longevity of the electrode.

3.7.3 Temperature Correction

For accurate pH measurements when the sample temperature is different from the calibration temperature of the electrode, a Nernst correction must be applied to correct for the temperature dependency of the slope of the electrode. However, most acid–base titrations are indicated by the detection of an inflection point or by a set potential titration at a steep part of the titration response, and thus an exact pH is not necessary for an accurate titration. More important for an assay could be the temperature-related expansion or contraction of the titrant relative to its standardization temperature. This is approximately 0.1% at 25 °C for a 5 °C deviation from the standardization temperature with an aqueous titrant, and about four times that for an organic solvent-based titrant.

3.7.4 Redox Sensors

A platinum or gold electrode will detect the oxidation potential of ions in a solution. The potential swings are typically quite large, and the small deflection of the titration curve before the inflection is less pronounced and the jump, when it comes, more sudden than with aqueous pH titrations. Usually somewhat smaller titration increments and a more sensitive anticipation detection system must be used if a variable-increment titration is performed. The reference electrode most commonly used is silver–silver chloride. Iodometric or bromometric titrations may also be detected with a polarized conductance detector consisting of two platinum electrode elements and either a constant current (voltametric) or constant voltage (amperometric) current source. The conductance measured is dependent on the presence of both iodine and iodide, or bromine and

bromide, at the electrodes to support an oxidation at one electrode and a reduction at the other.

3.7.5 Argentometric Sensors

Titrations using silver nitrate to quantitate halides such as chloride, bromide, or iodide, reactive sulfur compounds such as hydrogen sulfide or mercaptans, or cyanide, can utilize a metallic silver electrode. In the case of halides, the electrode follows the silver ion concentration, which is necessarily low when halides are present as they combine to form an insoluble precipitate, and high when the halides have all been combined with silver and a slight excess of silver added. The reference typically used is a silver–silver chloride half cell filled with 10% potassium nitrate solution. For sulfur compounds, the electrode must be sensitized to sulfide by depositing a layer of silver sulfide. The potential follows the reactive sulfur concentration, which drops as silver ion is added. The same reference may be used as for the halide titrations, but the reference electrode may be subject to clogging by silver mercaptide, silver sulfide, or oils. For petroleum crude oils and for distillate fuels, a glass half cell may be used as the reference. For clean petroleum distillates, a sulfide-specific electrode may be used, which is based on a crystalline matrix containing silver sulfide. Cyanide titrations follow the same pattern as the halide titrations, except that the tightly bound silver cyanide complex is soluble and no precipitate is formed.

3.7.6 Specific Ion Electrodes

Electrodes have been developed with sensitivity to various species. In general, if an electrode will give a rapid and stable response, it can be used for titration. Some commercially available electrodes are shown in Table 1.

Specific ion electrodes may be used for direct measurement of ionic activity, for a Gran's plot–standard addition method[2] or for conventional titration. The standard addition method works by adding aliquots of known concentration of the ion in question and observing the potential change measured on the electrode. The original concentration can be calculated from the potential change relative to the amount of added ion.

Table 1 Commercially available specific ion electrodes[1]

ammonia, NH_3	lithium, Li^+
ammonium, NH_4^+	nitrate, NO_3^-
cadmium, Cd^{2+}	nitrogen oxide, NO_x
calcium, Ca^{2+}	perchlorate, ClO_4^-
carbonate, CO_3^{2-}	potassium, K^+
chloride, Cl^-	silver/sulfide, Ag^+/S^{2-}
copper, Cu^{2+}	sodium, Na^+
cyanide, CN^-	sulfur dioxide, SO_2
fluoroborate, BF_4^-	surfactant, X^+, X^-
iodide, I^-	thiocyanate, SCN^-
lead, Pb^{2+}	water hardness, Ca^{2+}, Mg^{2+}

3.7.7 Photodetectors

Photodetectors have several advantages over potentiometric electrodes, including faster response, better long-term reliability, and usability in low-polarity nonaqueous solvents. However, a stable indicating dye must be available which will indicate the reaction under investigation. The sample must be optically clear and free of suspended solids, it must not generate gases or precipitates, and it must not leave deposits on the detector windows.

Three types of photodetectors are available for titrimetric analyses. All use optical fibers to carry a light source to the detection point, where there is a light path through the solution. The light path can be direct or reflected from a mirror, with a return optical fiber light path to the detector. In one system a wide-spectrum incandescent light is reduced to a relatively narrow band with an interference filter in order to focus on one particular absorption window. In another system the light source is a light emitting diode (LED) which emits light within approximately a 15 nm band. The beam is chopped by switching the LED on and off, and the measured signal is compared to the background light. Another design uses two LED sources at different wavelengths in order to null out differences in optical density of the sample solution as the titration proceeds. In some cases, a dual-wavelength detector can be set with one wavelength where the indicator will decrease absorption, and the other wavelength set where the indicator will increase absorption, and thus improve the detected signal.

Optical detectors can be used to follow a titration and for color development analysis, such as the analysis for phosphate using barbituric acid. The photometric titration curve is typically different from potentiometric titrations in its curvature and also in that, in many cases, it is the point where the color begins or ceases changing, rather than the inflection point, which is sought.

3.8 Cleanout

At the conclusion of a titration, the vessel must be drained and rinsed prior to the next sample coming in. This is typically achieved by either a dedicated drain port at the bottom of the titration vessel (see Figure 9) or a small tube which goes to the bottom of the titration vessel through which the sample is aspirated. Water or solvent can be piped from a pressurized reservoir, and the flow controlled by a solenoid-operated valve, or it can be propelled by a pump.

3.9 Communication

The results are typically computed in the titrator. Depending on the set-up, they may be transmitted to a printer or to a process control computer via a serial output, such as current loop or RS232 or RS488, via a local area network, or via an analog output such as 4–20 mA. The systems may be under control of the process computer, with a start titration signal sent by a digital signal line or by serial port, or the titrator may run autonomously on its own timing.

3.10 Waste Generation

When hazardous solvents or samples are involved in the analysis, it becomes advantageous to design the system to run smaller samples in order to minimize the waste generated.

3.11 Maintenance

If a titrator in a process situation is essential for the operation of the process, there should be trained maintenance personel with parts readily available. The electrode fluid level and titrant and reagent levels should be checked on a regular basis, possibly daily or every 3 or 5 days.

3.12 Isolation and Explosion Proofing

In locations with flammable fumes or products it may be desirable to eliminate spark sources. A titrator should meet all local requirements for isolation and explosion proofing. For example, instead of using electrically driven pumps and valves, one might use pneumatically driven pumps and valves. The electronics, including the burette motors, may be enclosed and purged with an inert gas.

3.13 Commercially Available Titrators

At the time of this writing (1999) there are several manufacturers of on-line process titration instruments. A partial listing of these is in the acknowledgement section of this article.

4 CHEMISTRY

4.1 Reaction Kinetics of Titrant–Analyte Reaction

Several factors are important to getting a good analytical titration. The electrode must deliver a signal dependent upon the concentration of either the analyte or the titrant. This signal must be rapid relative to the speed of the overall titration, and it must accurately depict the progress of the titration. The speed of the reaction is important, as well as that the reaction must go well to completion. A fast reaction such as the reaction of acidic protons with alkaline hydroxyl ions in aqueous solution or the reaction of chloride with silver(I) ion can be followed readily. A reaction of hydroxyl ion with saponifiable epoxide groups will not run quickly to completion, so the usual procedure is to add a known excess of the reactant, then allow the reaction some time, perhaps under an elevated temperature, and then, when the reaction is complete, the excess hydroxyl ion is titrated with an acid. A similar situation may occur when attempting to titrate sterically hindered amines with perchloric acid. The reaction is too slow to be easily titrated within a reasonable time, so an excess of perchloric acid is added, followed by a reaction time and back-titration of the excess perchloric acid with sodium acetate.

4.2 Solvent

The kinetics of the electrode response and the stability of that response are strongly influenced by the polarity of the solvent in which the titration is carried out. A polar solvent such as water ionizes the reference electrolyte and allows rapid ionic travel, and thus, efficient charge travel between the measuring electrode and the reference electrode. A nonpolar solvent such as hexane will not permit enough ionic movement to carry an electrode signal. Alcohols provide usable polarity. Acetic acid is used as a solvent for the nonaqueous titration of bases using perchloric acid. The acetic acid forms the acetate with many bases, and in a sample containing mixed bases most will titrate under a single inflection point. The perchloric acid can titrate weak bases that are difficult to measure by other methods. The preferred reference electrode fill solution is potassium perchlorate in acetic acid, although lithium chloride in ethanol works quite well. Pyridine may be used as a nonaqueous solvent for titrating acids using tetraethyl ammonium hydroxide. With a mixture of carboxylic acids with different acid dissociation constants, one inflection is usually obtained.

The polarity of the solvent affects how well the inflection points of mixtures of analytes can be separated. When the polarity is high, as with water, acetic acid, or pyridine, the inflection points of mixtures tend to form either one single inflection or a poorly separated series of weak inflections due to dissociation in the solvent. If the titration can be carried out in a solvent with lower polarity, the sample is less dissociated and the various constituents will form better-defined inflection points. Thus, a titration of citric acid in water shows only one inflection point, but in *n*-propanol three inflection points are detectable. Sulfuric acid in water shows one inflection point, but in a mixture of toluene and isopropyl alcohol, the inflections due to the two different protons are clearly

separated. The drawback in this last titration is that the speed of charge travel is restricted in the low-polarity solvent and one must titrate very slowly in order to allow the electrode sufficient time to equilibrate.

4.3 Sensor Kinetics

The speed of the sensor response can influence the titration. The titration speed must be such that the sensor gives an accurate portrayal of analyte status. A slow-responding electrode, such as an ammonia electrode which measures pH changes brought about by gaseous ammonia permeating a Teflon® membrane, requires a commensurately slow titration. A fast-responding electrode, such as the silver electrode in a silver nitrate–halide determination, makes possible a rapid titration. If the electrode response is slow and the titration is done too quickly, the end-point or inflection point will be overshot, thus yielding a high result. In nonaqueous acid–base titrations a glass bulb electrode will usually respond slower than in an aqueous environment. The ionic charge travels slower in a less-polar medium. Likewise, measuring an absolute pH in low-ionic-strength water such as rainwater requires more time for stabilization than when working with high-ionic-strength seawater or mineral acids or bases. Indicator dyes in photometric titrations are usually selected to have a fast reaction, and the speed of response becomes dependent on the efficiency of the stirring.

4.4 Indicators

Indicators in some way point to the equivalence point. Visual indicators are dyes, for instance litmus or phenolphthalein, which change their color absorption at various levels of the target species, such as hydrogen ion concentration. Permanganate, by virtue of its deep purple color, can be self-indicating. A visual observation of an iodine titration end-point can be sharpened by the addition of starch, which forms a blue-black complex with iodine. Mercuric EDTA (ethylenediaminetetraacetic acid disodium salt) [139-33-3] can be used to indicate a calcium or hardness titration using a mercury-sensitive electrode.

4.5 Materials Compatibility

When planning to measure samples on line by means of titration, it is essential to consider the compatibility of the sample with the wetted materials of the titrator and sample handling system. Teflon® valves are used where possible, and stainless steel is avoided when analyzing strong mineral acids. Glass electrodes and components must be avoided when analyzing acidic fluoride. Nonaqueous solvents can swell or alter polymeric components. Glass is dissolved, albeit slowly, by strong aqueous bases.

4.6 Environmental Impact

Safe and environmentally compatible disposal of waste is an important concern with any chemical operation. Titration of cleanwater samples with sulfuric acid for carbonates generates a small quantity of nontoxic salts. This can safely be disposed of in a public sewer. However, if the sample contains heavy metals, or if the titrant contains silver or mercury, it may be poisonous to biological water treatment systems and must be treated as a hazardous waste. Titrations in nonaqueous solvents generate waste which may vaporize and be a hazard to personnel on-site if not disposed of safely. If the titration can be done effectively on a smaller scale, less hazardous waste is generated per sample with correspondingly reduced disposal costs. This approach is not effective for nonhomogeneous samples, where a large sample size is necessary to obtain a representative sample. Small sample size must also be weighed against the need for precision, where the repeatability of measuring an aliquot is better with larger sample sizes.

Consideration must be made for any vapors which may issue forth from the solvent, sample, or titrant. Vapors may be flammable, explosive, or poisonous. The hazard involved should be assessed and provision made for dissipation or entrapment of the vapors. Chemical absorbing filter systems work well and the filters are readily disposed of.

5 APPLICATIONS

5.1 Aqueous Acid–Base Titrations

In water, electrode-indicated titrations are predictable. There is usually enough ionic activity to transmit ionic charge in the solution and the glass sensing electrode will rapidly and repeatably indicate the hydrogen ion level (pH). A base such as sodium hydroxide may be titrated with an acid such as sulfuric acid with repeatability better than 0.05% RSD (Figure 13). The exception to this is with low-ionic-strength samples such as rainwater, which must be titrated slower and possibly with a Gran's calculation to determine the equivalence point.[3] Industrial samples typically have adequate ionic strength for titrimetric analysis. If they do not, a small amount of potassium chloride may be added to increase the ionic strength.

The titrants typically are stable for long periods, but periodic standardization is good practice for best accuracy. Bases may be standardized with potassium acid phthalate (KHP). Acids are typically standardized with tris(hydroxymethyl)aminomethane (THAM). Alkaline titrants should be protected from atmospheric carbon dioxide by blanketing under nitrogen, storing in a sealed, collapsible bag-type vessel, or by filtering with

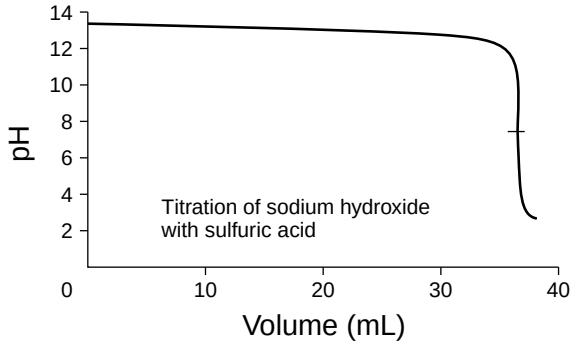

Figure 13 Titration of sodium hydroxide with sulfuric acid.

an alkaline filter made of ascarite or moisture protected sodium hydroxide pellets. In practice there is not much difference between using sulfuric or hydrochloric acids. Hydrochloric acid will evaporate, sulfuric acid will not. It is possible to standardize the titrant of an on-line titrator by making provision for volumetric introduction of a known standard, or one can take a grab sample and compare the lab analysis using standardized titrants with the on-line analysis. Table 2 shows some examples of aqueous acid–base titrations.

5.2 Nonaqueous Acid–Base Titrations

There are a great variety of nonaqueous acid–base titrations, many of which are not suitable at all for process titration. The reasons usually relate to the electrodes not being stable. A glass pH electrode must have the outer surface hydrated in order to be stable and have an optimum response. Good candidates for process titration are easy titrations, such as acidity in alcohols, or acid or alkaline strength of samples which can be readily mixed with low-molecular-weight alcohols such as methanol, ethanol, or propanol. Analysis of amines in ethanol typically works quite well. Unlike aqueous solutions, titrations to a set pH do not work well in nonaqueous media. The absolute potentials are not stable. However, the titration curve usually has a repeatable shape. Therefore, the end-point of the titration can repeatably be found by seeking the inflection point. If needed, the electrode can be hydrated with a water cycle between samples. Petroleum titrations, such as acid numbers, base numbers, and neutralization numbers, can normally only be done in the laboratory because the electrodes require frequent attention. The same case exists with the analysis of sulfides and mercaptans in petroleum samples by titration with silver nitrate. The electrodes are too finicky for process analysis.

Bases are typically standardized with KHP and acids with THAM. The exception is that perchloric acid in glacial acetic acid is typically standardized with KHP because it is acidic enough to titrate the phthalate as a base. Nonaqueous titrations are always evaluated to an inflection because absolute potentials are not as stable as in aqueous solutions. Table 3 shows some examples of nonaqueous acid–base titrations.

Table 2 Examples of aqueous acid–base titrations

Sample	Titrant	Evaluation
Sulfuric acid	sodium hydroxide	inflection or to pH 7.5
Sodium hydroxide	hydrochloric acid	inflection or to pH 7.0
Sodium carbonate	hydrochloric acid	inflection, generates CO_2 gas
Water alkalinity	sulfuric acid	pH 4.5 and 8.2, ppm as $CaCO_3$
Paper mill liquor, ABC test	sulfuric acid	three inflections, determines NaOH, H_2S, and CO_2
Boric acid in cooling water	potassium hydroxide	add mannitol, titrate to pH 8.5

5.3 Redox Titrations

Oxidation–reduction (redox) reactions can be used for a wide range of sample analyses. The sensing system used

Table 3 Nonaqueous acid–base titrations

Sample	Titrant	Solvent
Acidity in methanol	potassium hydroxide in methanol	neat
Amines	hydrochloric acid in isopropanol	isopropanol
THAM	perchloric acid in glacial acetic acid, indicated with crystal violet, detected with photodetector at 550 nm	glacial acetic acid
Amines	perchloric acid in glacial acetic acid	glacial acetic acid

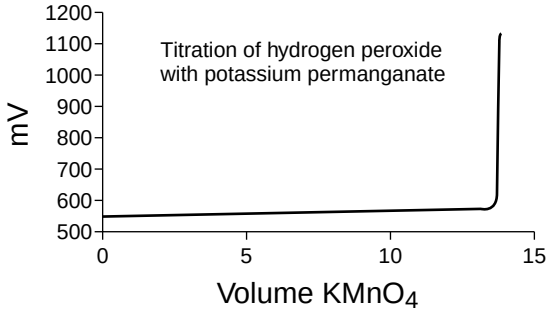

Figure 14 Redox titration: hydrogen peroxide with potassium permanganate.

Table 4 Redox titrations

Sample	Titrant	Indication	Solvent
Hydrazine	potassium iodate	platinum combination electrode	HCl and water, 1 : 1
Hydrogen peroxide	potassium permanganate	platinum combination electrode	1 M H_2SO_4
Copper(II) sulfate	sodium thiosulfate	platinum combination electrode	acidic, KI
Sulfurous acid in wine as SO_2	iodine	platinum combination electrode	neat sample
Hydrogen sulfide	iodine	platinum combination electrode	
Ascorbic acid in fruit juice	2,6-dichlorophenolindophenol sodium salt	photometric detection at 550 nm	water, adjust to pH 3.5
Iron(II)	Potassium permanganate	platinum combination electrode or photodetector at 550 nm	adjust to pH 1–2

is usually a platinum combination electrode. Analyses can be direct, such as the case of analyzing hydrogen peroxide with potassium permanganate (Figure 14), or indirect, such as the reaction of chlorine with iodide, which forms iodine and is subsequently titrated with potassium thiosulfate. Typically the solution is made strongly acidic. Table 4 shows some examples of redox titrations.

5.4 Argentometry

The most common argentometric titration is the analysis of chloride. Chloride content is often correlated to salt content in chemicals, food, or waters. The argentometric titration employs a silver combination electrode that measures the level of silver ion added as silver nitrate in a neutral to acidic solution. Although there is free ionic chloride present, it combines with the silver ion forming silver chloride, which precipitates, thus taking the silver out of solution. As the chloride ion concentration diminishes, the silver ion level begins to rise. This behavior can be expressed as the solubility product, K_{sp}, Equations (5) and (6):

$$Ag^+ + Cl^- \longrightarrow AgCl \quad (5)$$

$$K_{sp} = \frac{AgCl}{Ag^+} \times Cl^- \quad (6)$$

When the chloride is all combined with the silver, the silver ion concentration rises and the silver electrode measures a steep rise, which forms an inflection point (Figure 15). The amount of silver added thus corresponds to the amount of chloride added. This analysis is not specific to chloride, as bromide, iodide, and cyanide will also react with silver. In a mixture of iodide, bromide, and chloride the titration curve will have three inflection points, in that order. However, each inflection point does not necessarily reflect a stochiometric reaction of silver with only one halide species. Chloride will also react in the region associated with bromide, making this an inexact analysis. The pH of the solution is important – it must be neutral to acidic; the analysis will not work under alkaline

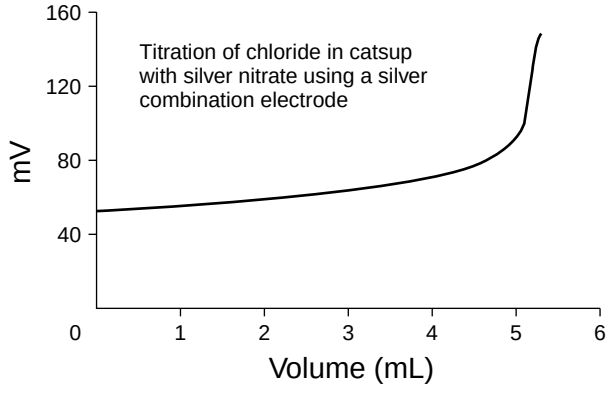

Figure 15 Titration of chloride with silver nitrate.

conditions. Usually an acid such as sulfuric, nitric, or acetic is added to bring the pH to the range 6–2. If the acid is too strong, it dissolves the silver in the indicating electrode. This dissolved silver will react with halides, thus lowering the apparent halide content. Cleaning the precipitated silver chloride may be a problem as it clings to electrodes and stirrers unless a small amount of nonionic surfactant such as Triton® X-100 is added to break up the precipitate. Table 5 shows some examples of argentometric titrations.

5.5 Complexometric Titrations

Complexometric titrations using EDTA may be used for a variety of divalent cations, such as zinc, copper, cadmium, nickel, and iron, but the most common analysis is the colorimetric determination of hardness. Hardness is determined in water treatment plants. The sample is buffered to pH 10 and Eriochrome® Black T indicator added. The EDTA chelates the calcium and magnesium in the sample. At the end-point, the EDTA concentration rises and causes the indicator to change from blue to purple. A colorimetric detector senses the color change. Magnesium can be masked by adjusting the pH to 12 with sodium hydroxide, and then titrating only the calcium. Other divalent cations may

Table 5 Examples of argentometric titrations

Sample	Titrant	Indication	Solvent
Silver(I)	thioacetamide at pH 5	silver combination electrode	2 M NaOH in water
Silver(I)	sodium chloride	silver combination electrode	water, neutral to acid
Chloride	silver nitrate	silver combination electrode	water, neutral to acid
Potassium cyanide	silver nitrate	silver combination electrode	water, must be alkaline; ammonia reduces interference from chloride
Mercaptan, aqueous or neat	silver nitrate	silver combination electrode, treated with Ag_2S	water, may be alkaline or acidic
Mercaptan, in petroleum distillate	silver nitrate	silver combination electrode, treated with Ag_2S or sulfide-specific electrode	isopropanol with 1% ammonia

Table 6 Complexometric titrations with EDTA[4]

Analyte	Indicator	pH	Absorption maxima (nm)	Remarks[a]
Magnesium	calmagite, 2-naphthol-4-sulfonic acid	10	630, 510	indicator 0.05% in water or 0.1% in 10% in alcohol
	calcon, 1-(2-hydroxy-1-naphthylazo)-2-naphthol-4-sulfonic acid	10	640, 520	indicator 0.5% in alcohol
	Eriochrome® Black T, 1-(1-Hydroxy-2-naphthylazo)-6-nitro-2-naphthol-4-sulfonic acid	10	670, 520	indicator 0.05 to 0.5% in ethanol
Calcium	calcon	12	640, 460	buffer with NaOH solution; a small amount of magnesium sharpens the color change
	murexide, 5,5′-nitrilodibarbituric acid ammonium salt	12	580	calcium only may be titrated with EGTA; magnesium may subsequently be determined with EDTA using Eriochrome® Black T
Zinc	Eriochrome® Black T	10	670, 520	
	alizarin fluorine blue	4.3	500	indicator 0.5% in 1 M ammonium acetate
	3-[di(carboxymethyl)aminomethyl]-1,2-dihydroxyanthraquinone alizarincomplexone			
Iron(III)	sulfosalicylic acid	1.5	520	
Nickel	murexide	10	580, 490	
Copper	alizarin fluorine blue	4.3	500	
	murexide	10	580, 490	

[a] Samples must be buffered to the pH indicated using an appropriate buffer: for buffer pH 4.5 use 60 mL glacial acetic acid, 50.0 g sodium acetate anhydrous, dilute to one liter; for buffer pH 10 use 570 mL NH_3 concentrated, 70 g NH_4Cl dilute with water to one liter; for buffer pH 12 use 1 M NaOH.
EGTA, ethylenebis(oxyethylenenitrilo)tetraacetic acid.

be titrated with a different adjustment of pH and appropriate indicators. Table 6 lists some buffer, indicator and cation combinations used in complexometric determinations.

ACKNOWLEDGMENTS

The manufacturers of titrators who supplied literature on instrumentation and applications are gratefully acknowledged:

Mettler-Toledo AG, Schwerzenbach, Switzerland
Applikon, Schiedam, The Netherlands
Tytronics Incorporated, Bedford, Massachusetts, USA
Maselli Misure, Parma, Italy
Metrohm Instrumente AG, Herisau, Switzerland.

Thanks are due for helpful suggestions on this article from Sam Lambros and Peter Boyle of pHoenix Electrode Co., Rob Hoogendijk of Applikon, Rolf Rohner of Mettler Toledo, and James L. Marshall of the University of North Texas.

ABBREVIATIONS AND ACRONYMS

EDTA	Ethylenediaminetetraacetic Acid Disodium Salt
EGTA	Ethylenebis(oxyethylenenitrilo)tetraacetic Acid
KHP	Potassium Acid Phthalate
LED	Light Emitting Diode
RSD	Relative Standard Deviation
THAM	Tris(hydroxymethyl)aminomethane

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Inorganic Environmental Analysis by Electrochemical Methods

Food **(Volume 5)**
Food Analysis Techniques: Introduction

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction

Pharmaceuticals and Drugs **(Volume 8)**
Robotics and Laboratory Automation in Pharmaceuticals Analysis

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Ion-selective Electrodes: Fundamentals

REFERENCES

1. 'Ion Selective Electrodes', Commercial Brochure, pHoenix Electrode Co., Houston, Texas, USA, 1995.
2. M. Mascini, 'Uses of Known Addition, Gran's Plots and the Related Methods with Ion-selective Electrodes', *Ion-select. Elect. Rev.*, **2**(1), 17–71 (1980).
3. G. Gran, 'Determination of the Equivalence Point in Potentiometric Titrations. Part II', International Congress on Analytical Chemistry, 1952, 77, 661; and G. Gran, *Acta Chem. Scand.*, **4**, 559 (1950).
4. K. Mooibroek, *Methodenführer Phototitration*, Mettler Toledo, Schwerzenbach, Switzerland, 1978.

Ultraviolet/Visible Spectroscopy in Process Analyses

Robert S. Saltzman
Bob Saltzman Associates, Inc., Wilmington, USA

1 Introduction 1
2 Brief History 1
3 Process Versus Laboratory Instruments 2
4 Types of Process Ultraviolet/Visible Analyzers 2
4.1 Filter Photometric Analyzers 2
4.2 Process Ultraviolet/Visible Spectrophotometers 4
5 Applications of Process Ultraviolet/Visible Analyzers 5
6 Calibration and Safety Issues 6
Abbreviations and Acronyms 6
Related Articles 6
References 7

Process analyzers based on selective ultraviolet (UV) and/or visible (VIS) absorption provide highly reliable on-line composition monitoring and closed-loop control throughout the process industries. Design criteria for ultraviolet/visible (UV/VIS) analyzers are different to those for laboratory instruments since the process instrument must meet the area electrical classification and often withstand harsh plant environments such as vibration, wide temperature swings, corrosive atmospheres and ''dirty'' samples.

The designs of various types of UV/VIS process analyzers are reviewed, and applications of these analyzers in major process industries are described.

1 INTRODUCTION

On-line composition analyses based on selective UV and/or VIS absorption have been used for process monitoring and control throughout the process industries for over 60 years.

New and improved UV and VIS light sources, detectors, interference filters and optical materials continue to be developed, expanding the potential for UV/VIS analyzer applications. Process analyzers using a linear photodiode-array detector (LPDAD) and holographic gratings have been developed to provide multicomponent and selective analyses not previously feasible.[1]

2 BRIEF HISTORY

During the past half century of design evolution, photometric analyzers based on UV/VIS absorption have found increasingly wide applicability throughout the process industries. UV/VIS analyzers were among the first plant stream analyzers used for direct monitoring of process stream composition and for closed-loop control.

Before 1930, Kurt Müller, in Germany, developed one of the first UV analyzers to monitor elemental mercury in air, using the mercury 253.7-nm atomic resonance line as the measuring wavelength.[2] In effect, it was an atomic absorption measurement. In 1939, Woodson et al. of General Electric (GE) also described a sensitive mercury analyzer using a GE G4T4 low-pressure discharge lamp as the source for the 253.7-nm line.[3]

Until the late 1950s, most UV analyzers were developed within petroleum and chemical companies for proprietary use, and a few were designed at university research facilities. Many of these early instruments either used the 253.7-nm mercury resonance line from a germicidal lamp as the measuring wavelength or adapted the Beckman DU laboratory spectrophotometer.

Some of the early UV/VIS analyzers performed adequately, were well applied and filled needs resulting in substantial savings. However, others were poorly designed for continuous plant service and were often misapplied. As a result of some of the early failures, it took almost a generation for some process companies to accept UV/VIS analyzers (and plant stream analyzers in general). The split-beam photometric analyzer was developed by DuPont in the mid-1950s initially to meet the need for accurate monitoring and control of chlorine gas concentration in the ''chloride'' process for producing titanium dioxide.[4] In 1962, about 200 of these instruments were applied within the DuPont Company when the commercial version was introduced as the first product of the newly formed Instrument Products Division. Other UV/VIS photometric analyzers, particularly the Analytic Systems (now Teledyne) photometric analyzer and the Halikainen UV analyzer, were commercially available and used by several companies at that time. However, the introduction of the split-beam analyzer, with its large application base within the DuPont Company, greatly expanded the use of photometric analyzers throughout the process industries.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

3 PROCESS VERSUS LABORATORY INSTRUMENTS

In contrast to a laboratory UV/VIS spectrophotometer, the process analyzer is designed to be used in a plant environment for the continuous or semicontinuous analysis usually for a specific component critical for process or environmental control. Whereas most laboratory UV analyses are on liquid phase samples, the bulk of process UV analyses are in the vapor phase. The measuring wavelengths used in process UV analyzers are most often provided by isolating an emission line from a metallic discharge lamp. Since the emission line is at a fixed wavelength and extremely narrow, the analyzer output can be highly linear with concentration based on Beer's law even if the measuring wavelength is away from an absorption peak. If the UV/VIS analyzer is used in an automatic control loop, high reproducibility with a minimum of dead zone may be essential, and the instrument usually must respond rapidly to changes in concentration. Hence, delays through long sampling lines must be avoided.

The problems associated with continuous sampling in a plant environment are usually formidable, and design criteria for process stream analyzers go well beyond those for the laboratory. Criteria used in designing UV/VIS process stream analyzers, developed from plant experience, have been described previously[1] and include rugged construction, serviceability, the accuracy, stability and reproducibility required for the application, linear and rapid response, and simplified sampling requirements.

Most process UV/VIS analyzers have separate light source, sample, and photometer housings. The three housings of the analyzer are easily separated, and the analyzer is readily adapted for monitoring process gas and liquid streams in pipeline cells installed directly in the process line. This modular construction permits the use of special cells such as those used with high- or low-temperature gas and liquid samples.

In recent years, the elements used in UV analyzers have been greatly improved. Advances in thin-film technology have resulted in more stable interference filters with higher resistance to temperature and improved transmittance with narrower band-passes. ''Quieter'' discharge lamps with longer lives are now available, and silicon photodiodes have been improved for use in the UV region with high signal-to-noise ratios and temperature resistance equivalent to that of vacuum phototubes. Highly pure quartz with high transmittance throughout the UV/VIS region is now available for windows, lenses and fiber optics, and other optical glasses are being developed.

4 TYPES OF PROCESS ULTRAVIOLET/ VISIBLE ANALYZERS

Several types of process analyzers based on UV/VIS absorption are commercially available. These instruments range from very simple to highly sophisticated designs. The types of analyzers described in this section are not the only commercial UV/VIS absorption analyzers, but are the most prevalent units used in industrial applications. Process UV/VIS analyzers may be divided into two basic types: (1) filter photometric analyzers and (2) spectrophotometers. Included in the filter photometer category are single beam, split-beam and ''filter wheel'' analyzers. Scanning spectrophotometers (with fiber optics) and process diode-array analyzers are the significant types of spectrophotometers used in process applications. Process UV analyzers based on étalon technology were introduced; however, since none are now commercially available, they will not be discussed.

4.1 Filter Photometric Analyzers

Filter photometric analyzers are the most common kind of process UV/VIS analyzers, and the simplest is the single-beam type; a functional diagram is shown in Figure 1. The output of this type of instrument will be affected by fluctuations and drift of the light source, dirt or bubbles in the sample cell, and any drift in the detector or detector circuit. As a result, single-beam instruments must operate on relatively low sensitivity (high absorbance) levels to provide reasonably stable analyses. Stabilized light sources and improved detector circuits, in recent years, have improved stability to some extent, and these simplified units are being used in applications such as the detector for semicontinuous colorimetric analyses where frequent automatic 'zero' adjustment is provided.

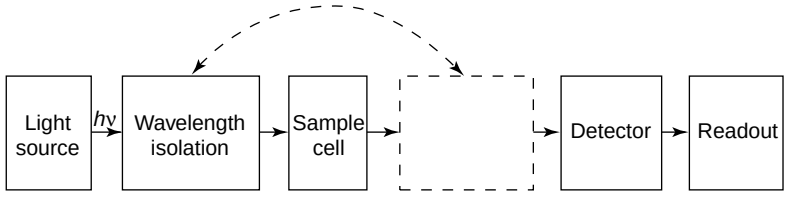

Figure 1 Functional diagram of single-beam analyzer.

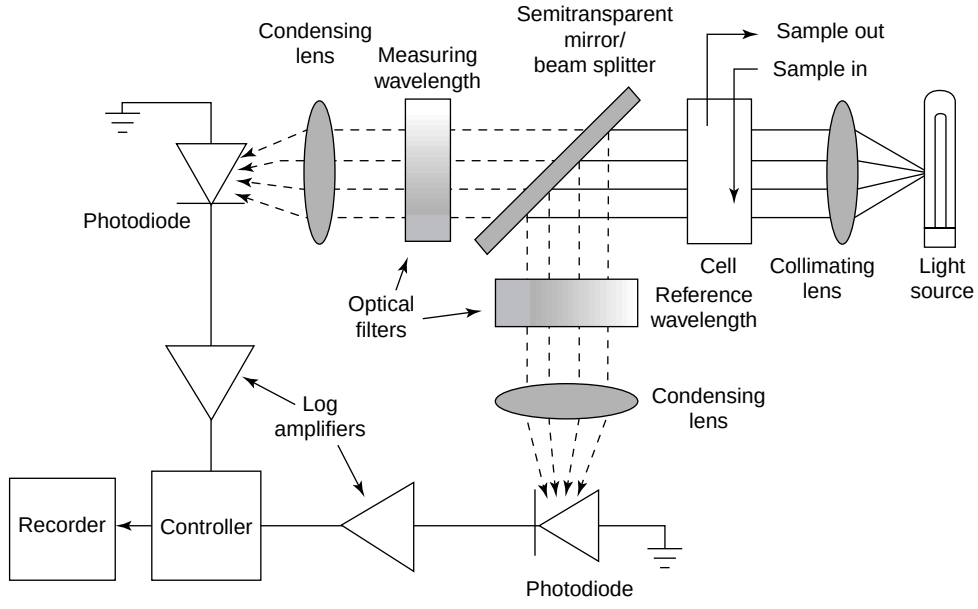

Figure 2 Functional diagram of split-beam UV/VIS analyzer. (Reproduced by permission of Ametek Process and Analytical Division.)

A functional diagram of a typical split-beam analyzer is shown in Figure 2. This type of design is based on simultaneous differential absorption measurements at two wavelengths and has several distinct advantages, including compensation for dirt and bubbles in the sample cell.[4] Radiation from the source, generally a metallic discharge lamp, is partially absorbed on passing through the sample. The radiation leaving the sample is divided into two beams by a semitransparent mirror. Each beam passes through an optical filter to a photodiode. The filter removes radiation at all wavelengths except the one to be measured. Radiation at the analytical wavelength is absorbed strongly by the measured component. Radiation at the reference wavelength is absorbed weakly or not at all by the measured component. Each photodiode develops a current directly proportional to the intensity of radiation striking the photodiode.

Each photodiode current is converted into a DC voltage proportional to the negative logarithm of the photodiode current, providing signals linear with concentration based on Beer's law. The output voltages are fed to a microprocessor which provides digital and analog outputs calibrated in concentration units.

Figure 3 is a functional diagram of a filter-wheel chopper photometric analyzer. The filter wheel contains two or more interference filters (at different selected wavelengths) continuously rotated by the chopper motor sequentially into the optical path. The source light is collimated by a series of lenses and directed through the filters and the sample cell. A detector lens focuses the transmitted light on to the detector. The detector

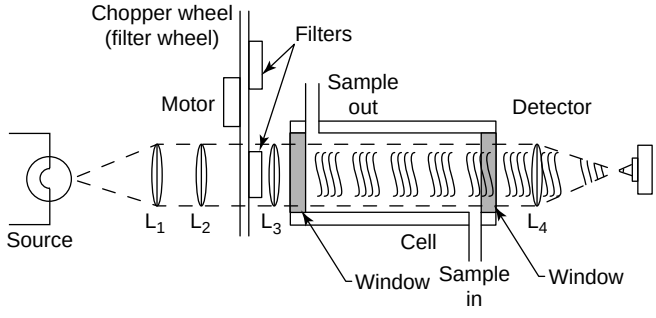

Figure 3 Functional diagram of a filter-wheel chopper UV/VIS analyzer. (Reproduced by permission of ABB Process Analytics.)

provides electrical signals proportional to the intensities at each wavelength. As in the split-beam analyzer, two wavelengths are used for a single concentration measurement; however, if more filters at selected wavelengths are used, multiple concentrations may be measured. The analyzer microprocessor is programmed to compensate for overlapping absorbances.

Using the new improved UV elements, several manufacturers of infrared (IR) filter photometric analyzers have been able to convert the basic optical bench of the IR analyzer for use in the UV/VIS region. The functional diagram in Figure 3 is typical of ABB Process Analytics, Ametek–Western Research and Anarad analyzers and several others.

The Ametek–Western Research analyzer uses up to three separate hollow-cathode sources with a system of

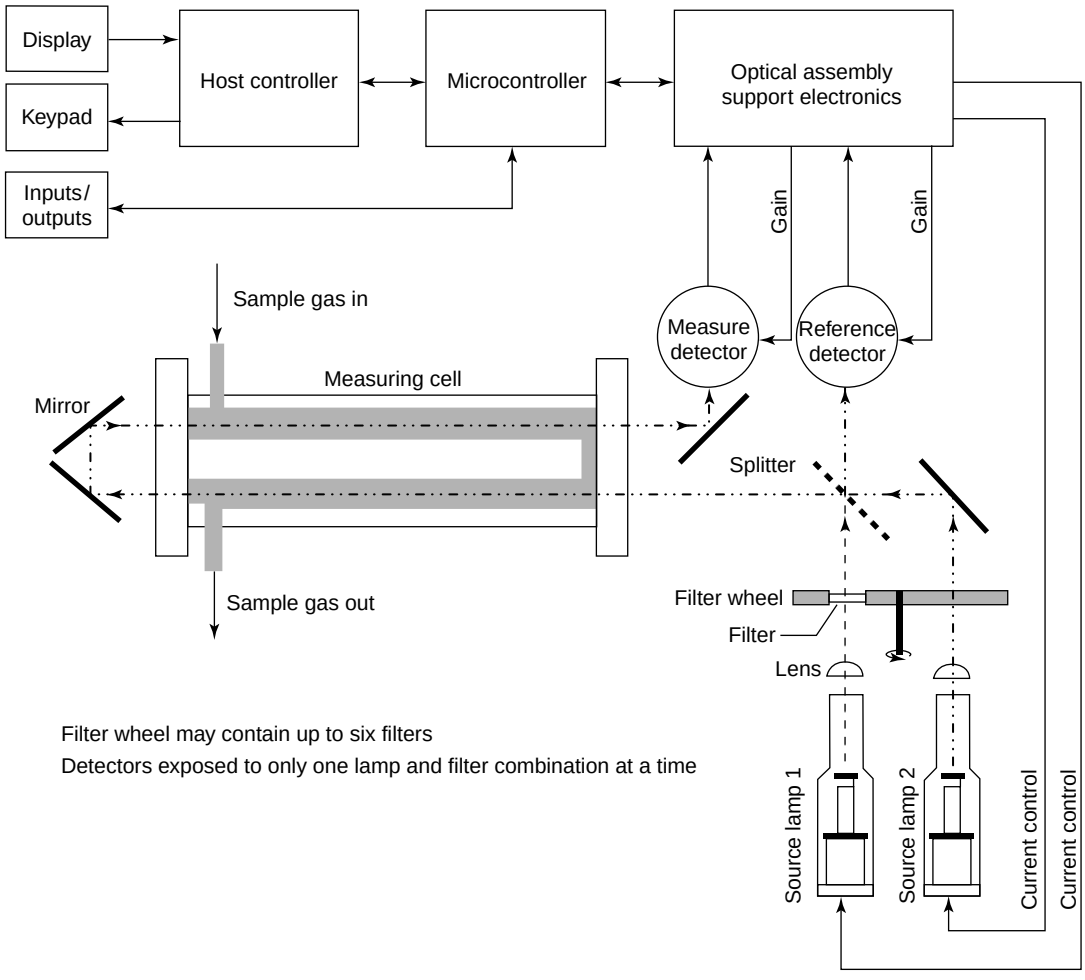

Figure 4 Functional diagram of the Ametek–Western Research multiwavelength analyzer. (Reproduced by permission of Ametek Process and Analytical Division.)

mirrors to provide several discharge lines from these lamps as measuring wavelengths. Figure 4 is a functional diagram of the Ametek–Western Research Series 900 analyzer.

4.2 Process Ultraviolet/Visible Spectrophotometers

It was realized early in the development of photometric analyzers that a laboratory scanning spectrophotometer could not be installed in a process plant environment with any acceptable degree of reliability. The optical bench of a spectrophotometer with an LPDAD is much less complex than that of a scanning spectrophotometer, where an optical element must be moved to provide spectrum intensities to a detector.

Process scanning UV/VIS spectrophotometers have optical benches functionally the same as those used in the laboratory but use fiber optics to transmit radiation through the sample cell from the monochromator to the photomultiplier detector. In general, the monochromator and the detector are installed in a clean, electrically classified ''general-purpose'' area with the sample cell in the plant area. Practically all of the uses for the process UV/VIS scanning spectrophotometers are for applications in the VIS and the near-UV region since the quartz in the fiber optics may have limited transmittance and can be ''solarized'' with time in the far-UV region. A computer, usually with chemometric software, is used with the spectrophotometer to provide the required output signals.

Process photodiode-array analyzers have been developed and applied for process control and emission monitoring. The optical assembly of the LPDAD spectrophotometer used in a process analyzer has no moving parts (except for a shutter in one design), and its detector assembly is compact and particularly rugged with superior signal-to-noise ratio performance and overall quality.

The process diode-array spectrophotometer shown in Figure 5 uses a deuterium lamp source, operating

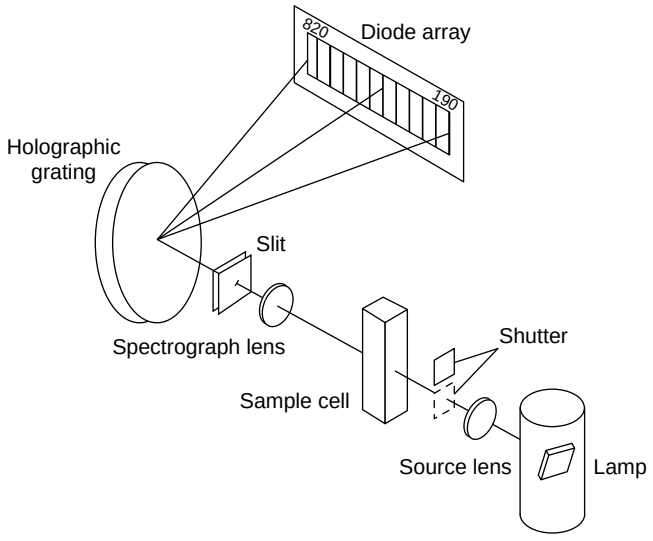

Figure 5 Functional diagram of process diode-array analyzer. (Reproduced by permission of Ametek Process and Analytical Division.)

at a reduced current and emitting a broad spectrum of radiation (light) from the vacuum-UV to the very near-IR region. The light is collimated and directed through the sample cell to the detector assembly where the light is focused on to a slit, striking a holographic grating which disperses the light on to the LPDAD. The grating is precisely positioned with respect to the diode array such that each diode sees light a fixed number (or fraction) of nanometers apart. High-resolution versions extending from 190 nm are used for source monitoring.

5 APPLICATIONS OF PROCESS ULTRAVIOLET/VISIBLE ANALYZERS

Thousands of process UV/VIS analyzers are being used throughout the process industries for environmental and process control. In some applications where the UV analyzer is used in a feedback control loop, the process may be inoperable without the analyzer.

Process UV/VIS analyzers are used extensively for monitoring of source emissions, particularly for sulfur dioxide (SO_2) and nitrogen oxides (NO_x) concentrations.[5] Other pollutants monitored by UV/VIS analyzers include ammonia, elemental mercury, aromatic compounds and elemental halogens.

In petroleum refining and in gas production, process UV/VIS analyzers are used extensively for control of the Claus sulfur recovery process.[6,7] Rapidly responding analyzer systems monitoring both the SO_2 and the hydrogen sulfide (H_2S) concentrations in the tail gas provide an output for feedback control of the air to the Claus furnace to maintain the optimum 2 : 1 ratio of [H_2S] to [SO_2]. Process UV analyzers are also used to monitor the H_2S concentrations in the Claus plant feed and in the ''lean'' amine in the sulfur removal process.[8]

The transmitted color grade is a primary quality measurement in petroleum refining, and process photometric analyzers are used extensively to monitor color grades from the very dark tars (Tag–Robinson color scale) through the refined oils (ASTM D-1500 color scale) to the hydrocarbons (Saybolt color scale).[9]

Process UV/VIS analyzers are used in a large variety of applications in the petrochemical and chemical industries. For example, elemental halogens, particularly chlorine, with ranges of 0–50 ppm to 100% are monitored in gaseous and liquid phases. Chlorine has a broad absorption over a wavelength span covered by few other materials. Materials of construction in contact with a halogen sample must often be carefully selected. Teflon®, Hastelloy C-176, nickel, etc. have been used in these applications. Synthetic sapphire windows are used where fluorine or hydrogen fluoride are present.

Precise control of many chlorination processes, such as in dichloroethane production, in phosgene manufacture, and in the titanium dioxide chloride process, have been made possible through the accurate measurement and control of chlorine concentration.

In general, oxidizing agents have strong absorption bands. Materials such as sodium hypochlorite (NaOCl), chlorine dioxide (ClO_2), hydrogen peroxide (H_2O_2), ozone (O_3), and potassium permanganate ($KMnO_4$) are examples of oxidizing materials whose concentrations are monitored with process UV/VIS analyzers. ''Bleaching'' steps can often be controlled by the monitoring of the oxidizing agent concentration.

Concentrations of aromatics in aliphatic hydrocarbons are monitored in very low concentrations, and unit processes for removal of the aromatics are often controlled with this sensitive measurement. In the manufacture of polycarbonates, bisphenol-A [2,2-bis(4-hydroxyphenol)-propane] is continuously monitored.[10]

Liquid color grading scales other than those used in petroleum refining include APHA (American Public Health Association), Rosin, Hazen and ICUMSA (International Commission for Uniform Methods of Sugar Analysis). Process photometric analyzers have been calibrated for each of these (and other) color scales and are successfully and accurately monitoring the color of a variety of liquid products on the specified color scale.[9] Several of the streams are monitored at very high temperatures($>200\,^{\circ}C$) and pressures(>150 bar).

Salts of transition metals such as iron, nickel, manganese, vanadium, and copper are usually strong absorbers and are monitored for process control.

In several chemical processes, sulfur-containing compounds such as SO_2, H_2S, carbonyl sulfide (COS) and carbon disulfide (CS_2) are monitored and controlled for safety and process efficiency.

In pharmaceutical processing, antibiotics generally have strong UV absorption and are being monitored with process UV analyzers. When the valuable product is collected after separation in a large process liquid chromatographic column, a process UV analyzer may be used to provide the concentration information required to determine when to start and stop the collection of the product.

The largest application of photometric analyzers in the production of wood pulp for paper is the monitoring and control of ClO_2 in bleaching operations. ClO_2 is highly corrosive, and the only satisfactory materials of construction in contact with the sample have been titanium, fluorocarbon resins and ceramics (quartz and glass). Vapor-phase ClO_2 in percentage concentrations is highly explosive and should not be heated. It is also readily photochemically decomposed when exposed to wavelengths below 420 nm, and it is advisable to install a titanium flame arrester between the analyzer system and the process.

UV photometric analyzers are used to monitor lignin concentration in brown stock and for control of pulp digesters.

6 CALIBRATION AND SAFETY ISSUES

In practically all gas-phase applications of UV photometric analyzers, Beer's law is precisely followed up to 1.5 absorbance (and often higher) if the measuring wavelength is a narrow line from a metallic discharge lamp. Hence an accurate initial analyzer calibration may be provided with only one or two prepared concentrations of the gas to be analyzed. It should be noted, however, that the calibration gas should be at the same regulated temperature and pressure as the sample will be in the field. Without absolute pressure control, the accuracy of a gas analysis may be limited to $\pm 2\%$ (because of barometric pressure changes). With a dual-wavelength, single-component analysis, an optical glass with a differential absorbance equal to that of an upscale concentration of the measured component may be inserted into light path (with ''zero'' gas in the sample cell) to give a quick but accurate check of the analyzer span.

For multicomponent analyses with overlapping absorptions, initial calibration can be highly time-consuming since an extensive calibration set of concentrations of the components must be fed to the analyzer to establish the correlation analysis software and to confirm the validity of the software package.

Often, process streams with UV/VIS-absorbing compounds are toxic and/or explosive, and special safety precautions must be taken when calibrating or working on the analyzers. It should also be noted that UV sources, such as a low-pressure mercury discharge lamp, may produce low levels of VIS light but intense UV radiation highly dangerous to human eyes. Safety glasses will block the most dangerous low-wavelength UV light and should always be worn when working with UV analyzers. However, even with safety glasses on, one should avoid exposure to the source radiation, which can also be damaging to the skin.

ABBREVIATIONS AND ACRONYMS

APHA	American Public Health Association
ICUMSA	International Commission for Uniform Methods of Sugar Analysis
IR	Infrared
LPDAD	Linear Photodiode-array Detector
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Lube Products. Molecular Characterization of Base Oils

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction • Chemometric Methods in Process Analysis • Sampling and Sample Preparation in Process Analysis

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data

Electroanalytical Methods **(Volume 11)**
Ultraviolet/Visible Spectroelectrochemistry

Electronic Absorption and Luminescence **(Volume 12)**
Detectors, Absorption and Luminescence

REFERENCES

1. R.S. Saltzman, J.R. Small, J. Steichen, 'New Process UV/VIS Diode Array Analyzer Development', *Anal. Instrum. (Research Triangle Park, NC)*, **23**, 69–83 (1988).
2. K. Müller, 'A New Method for Determining Mercury Content in Air', *Z. Phys.*, **65**, 738–758 (1930).
3. T.T. Woodson, 'A New Mercury Vapor Detector', *Rev. Sci. Inst.*, **10**, 308–311 (1939).
4. L.G. Glasser, R.S. Kansler, D.J. Troy, 'A Simple Photoelectric Process-stream Analyzer', *Rev. Sci. Instrum.*, **33**, 1062–1066 (1962).
5. R.S. Saltzman, J.A. Williamson, 'Monitoring Stationary Source Emissions for Air Pollutants with Photometric Analyzer Systems', *Anal. Instrum. (Research Triangle Park, NC)*, **9**(D-9), 1–6 (1971).
6. R.S. Saltzman, E.B. Hunt, Jr, 'A Photometric Analyzer System for Monitoring and Control of the $[H_2S]/[SO_2]$ Ratio in Sulfur Recovery Plants', *Anal. Instrum. (Research Triangle Park, NC)*, **10**, 55–59 (1972).
7. R.S. Saltzman, M. Mathews, L.M. Leonard, 'Improved Sulfur Recovery Plant Control Using New Process Analysis Systems', *Sulfur*, **91**, 89–100 (1991).
8. R.S. Saltzman, 'A Monitoring System for Low levels of Hydrogen Sulfide in Alkanol Amine Streams', *Anal. Chim. Acta*, **190**, 227–233 (1986).
9. R.S. Saltzman, J.S. Harrison, 'Continuous Process Stream Monitoring of Liquid Color Grades', *Anal. Instrum. (Research Triangle Park, NC)*, **3**, 209–217 (1965).
10. J.R. Small, 'A Continuous Analyzer Combining Liquid–Liquid Extraction and UV Photometry', *Anal. Instrum. (Research Triangle Park, NC)*, **3**, 41–47 (1965).

Pulp and Paper Matrices Analysis: Introduction

B. Bruce Sitholé
Pulp and Paper Research Institute of Canada, Pointe Claire, Canada

1 Wood Chemistry 1
2 Wood Pulping 4
2.1 Mechanical Pulps 4
2.2 Chemical Pulps 4
2.3 Recycled Pulps 4
3 Bleaching 5
4 Papermaking 5
5 Quality Control 6
6 Protection of Consumers 7
7 Odors 7
8 Environmental Protection 7
9 Estrogen Mimics 8
Abbreviations and Acronyms 8
Related Articles 8
References 8

The pulp and paper industry is a major player in the economies of many developed countries. As illustrated in Figure 1(a) and (b), the developed world accounts for more than 60% of the world output in pulp, paper and paperboard capacity. In one particular case, namely Canada, the industry contributes over 15% to the trade balance (see Figure 2). In order to ensure continual survival of the industry, it is necessary to continually improve and/or optimize the utilization of raw materials, the pulping processes, the papermaking process, the product quality properties and the protection of the environment. Analytical chemistry has an important role to play in the improvement and/or optimization of these parameters. This article summarizes how and where analytical chemistry is used, or can be used, to achieve the desired goals. Applications of analytical chemistry in the pulp and industry are reviewed in detail in the rest of the articles in this section on pulp and paper.

1 WOOD CHEMISTRY

Wood is the main raw material for the manufacture of pulp and paper. Its components can be divided into two groups, viz. macromolecular and low molecular weight substances. Macromolecular substances are the major constituents and they include cellulose, hemicelluloses and lignin. Cellulose **(1)** determines the character of wood fibers and enables them to be used for papermaking. It is a polysaccharide of glucose units with the formula $(C_6H_{10}O_5)_n$ where n is the number of repeating sugar units or the degree of polymerization. Hemicelluloses are polysaccharides of five different sugars from hexoses (glucose, mannose, galactose) and pentoses (xylose and arabinose). The hemicelluloses are susceptible to more degradation and dissolution than cellulose during chemical pulping. Lignin is an amorphous, highly polymerized substance that cements fibers together. It is a very complex polymer consisting primarily of *p*-hydroxyphenyl, guaiacyl and syringyl groups that are linked together in a three-dimensional structure and whose concentrations differ in softwoods and in hardwoods. A simplified structure of a partial softwood lignin molecule is shown here **(2)**.

The low molecular weight components are minor constituents that include pectins, extractives and inorganic matter. Although minor, these materials nonetheless do have major impacts on the pulp and papermaking processes. For example, lipophilic extractives (fatty acids, resin acids, glycerides, sterols, waxes, etc.) account for about 2–5% of the weight of wood but cause or contribute to process upsets (pitch deposition problems) that cost the North American industry upwards of $200 000 000 annually.

Analytical chemistry plays a major role in understanding the chemistry of the wood. For example, one of the major chemical differences between softwoods and hardwoods is the presence of resin acids (diterpene acids) in softwoods and not in hardwoods. In general lipophilic extractives are comprised of fatty acids, resin acids (softwoods only), fatty acid ester glycerides, waxes, sterols, steryl esters and alcohols, as shown in the example structures **(3–9)**. Analysis of the extractives has shown that they are important in facilitating or hindering the pulping and papermaking process. For example, the resin acids in softwoods are easily saponifiable during chemical pulping and consequently are easily removed from fibers by washing with water. This is one of the reasons why softwoods are the main source of fibers for chemical pulping. Hardwoods lack resin acids and, consequently, are difficult to process by chemical pulping. Normally pulping of hardwoods is done after storage of the wood for some time. The storage results in a decrease in the amount of wood resin in the wood. In addition, the chemical pulping of hardwoods is facilitated by addition of the wood resin (tall oil) collected from mills that pulp softwoods.[1] Here analytical chemistry plays a role by monitoring the chemistry of the wood resin with storage time. For example,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

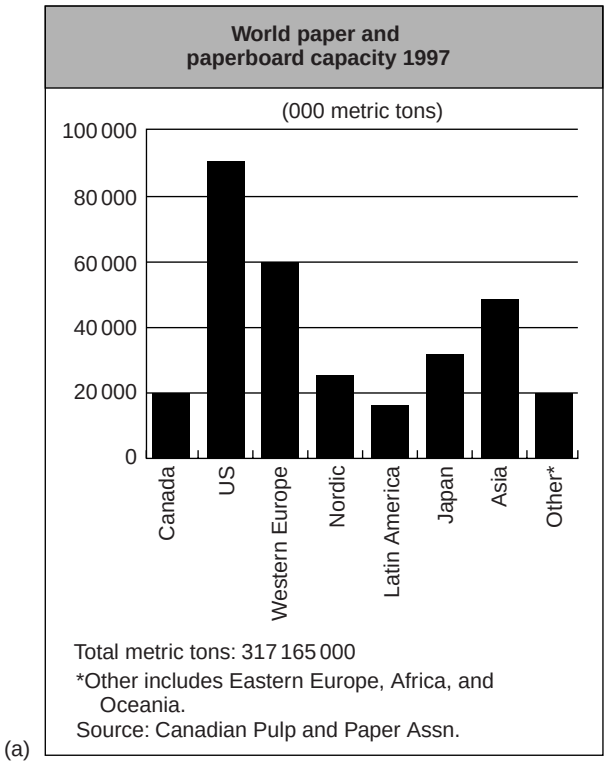

Figure 1 World production of pulp and paper.

compounds that cause the most pitch deposition problems are reduced by storing the wood prior to pulping.

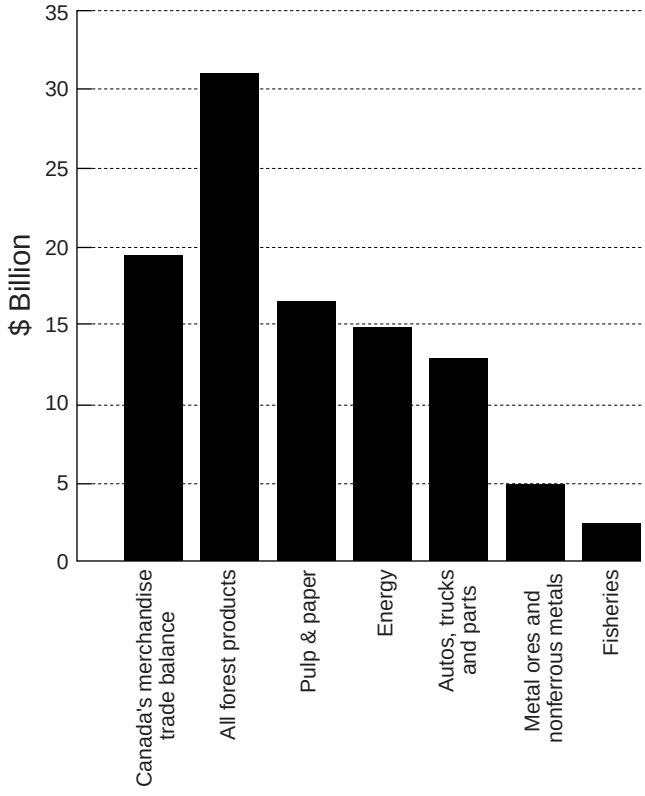

Figure 2 Contribution of the forest products industry to the economy of Canada (1998).

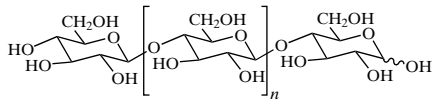

(1)

In addition, analyses of wood and bark show that bark contains 5–6 times more wood resin than the wood itself. This is one of the reasons why good bark removal is a prerequisite in pulp and paper manufacture.[2] Wood resin is known to contribute to the toxicity of mill effluents.[3] Hence mills are required to monitor for their presence in effluents to ensure that their effluent treatment systems remove them adequately. In analyzing wood extractives, the main analytical tools are gas chromatographic techniques (see **Pyrolysis in the Pulp and Paper Industry**).

Microscopic techniques are used to characterize the morphology of wood fibers. Thus, softwoods can be differentiated from hardwoods by the size and dimensions of the vessel elements. Fibers from softwood are longer than fibers from hardwoods, and this is why they are the main wood species for making printing, writing and packaging papers. Wet chemical methods are used to study the nature of wood, for example, to differentiate sapwood from heartwood by staining techniques.

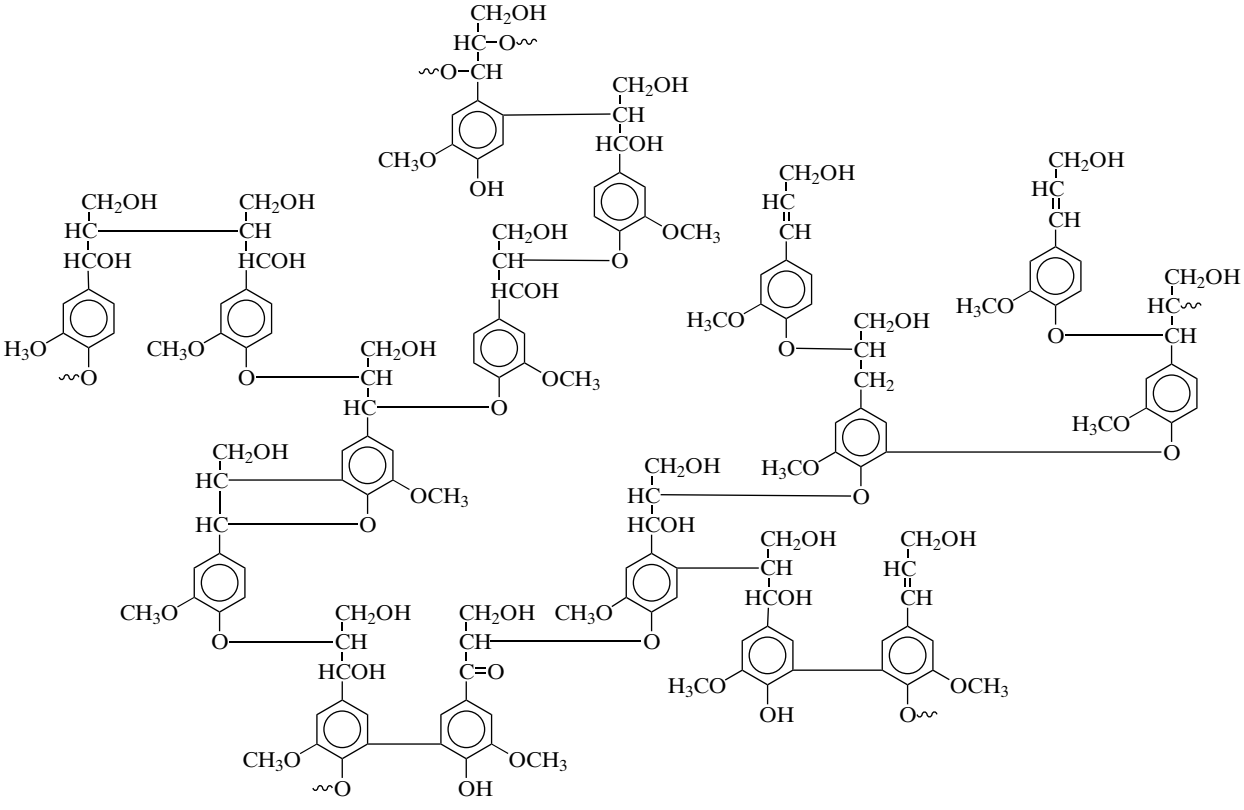

(2)

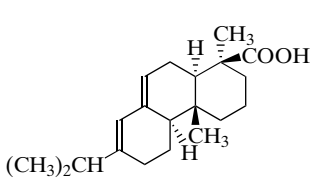

Abietic acid

(3)

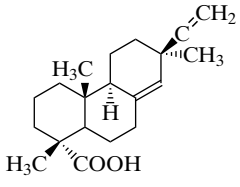

Pimaric acid

(4)

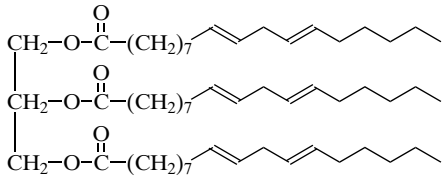

Triglyceride of linoleic acid

(8)

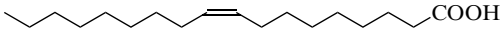

Oleic acid

(5)

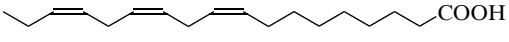

Linolenic acid

(6)

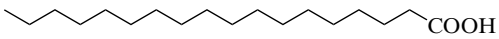

Stearic acid

(7)

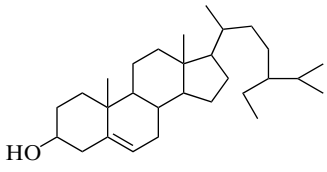

Sitosterol

(9)

The chemistry of cellulose and hemicelluloses in wood and pulps has been studied by wet chemical methods and instrumental methods such as spectroscopy and chromatography. In mechanical pulp furnishes, the pulping processes release substances from the wood that are cationic and subsequently compete with papermaking additives that such retention aids. Controlling or minimizing the cationic demand is an important concern in papermaking. Studies by Finnish scientists have shown that pectin polysaccharides such as glucomannans are major contributors to the cationic demand in mechanical pulp furnishes.[4] Analysis by gel permeation

chromatography revealed that these pectic substances have an average degree of polymerization of 130. Their identification led to the development of a pectinase enzyme aimed at destroying the compounds in pulps.[5]

2 WOOD PULPING

Wood is debarked and then cut into wood chips before pulping. The main pulping processes use mechanical and chemical processes to yield mechanical and chemical pulps.

2.1 Mechanical Pulps

Mechanical pulps are produced by mechanical grinding action on wood chips or wood, in the presence of water, to produce pulp. The pulp yields are very high (>95%) because very little wood components are lost in the aqueous matrix. Mechanical pulps are used to produce lower grade products such as newsprint, catalogs, paperbacks, wrapping papers and paperboards. These products do not require paper with high-strength physical properties. Mechanical pulps contain lignin and this material causes the paper to yellow in the presence of light (see **Mechanical Pulps, Ultraviolet/Visible Spectroscopy of Chromophores in**). Controlling the yellowing process can significantly increase the value of mechanical pulps and this is where analytical chemistry comes into the picture. The complex structure of lignin has challenged researchers to develop suitable methods to characterize the material.[6] Methods that have been used encompass classical and spectroscopic analytical techniques. For example, chemical degradation of lignin with alkaline nitrobenzene yields large amounts of aromatic aldehydes, namely *p*-hydroxybenzene, vanillin and syringaldehyde. This degradation method is extensively used to classify lignins from different wood species according to the amounts of the aldehydes generated. Spectroscopic methods that have been used to study the structure of lignin include ultraviolet (UV), infrared (IR) and nuclear magnetic resonance (NMR) techniques. UV spectroscopy has been used to monitor UV-absorbing functional groups (e.g. aromatics), IR for estimating the ratio of guaiacyl to syringyl units in lignin and NMR for the quantitative determination of functional groups such as aliphatic hydroxyl, phenolic hydroxyl and benzyl alcohol groups. The lignin has to be extracted from wood or pulps before analysis and several procedures have been developed. One typical extraction scheme for obtaining lignin from pulp is shown in Figure 3.

2.2 Chemical Pulps

Chemical pulps are made by cooking wood chips in chemicals such as caustic soda with sodium sulfide. The caustic soda dissolves most of the lignin from the wood matrix; consequently, the pulp yield from such a chemical pulping process is about 50%. In addition, some cellulose and hemicellulose compounds are dissolved during the cooking process and this is an undesirable occurrence because it results in pulps with lower than optimum papermaking properties. To optimize the cooking process, analytical chemistry is used to measure and evaluate parameters such as alkalinity, total titratable alkali and permanganate number (extent of lignin removal). The integrity of the pulps is determined by measuring their cellulose and hemicellulose contents. Cellulose is measured by chemical degradation methods or by gel permeation chromatography. Hemicelluloses can be measured by their depolymerization (enzymatic or otherwise) followed by analysis of the depolymerized monosaccharides by high-performance liquid chromatography (HPLC), anion-exchange chromatography or gas chromatography (GC). In recent years capillary ion electrophoresis has been increasingly used to analyze for carbohydrates in pulp and paper matrices (see **Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis**).

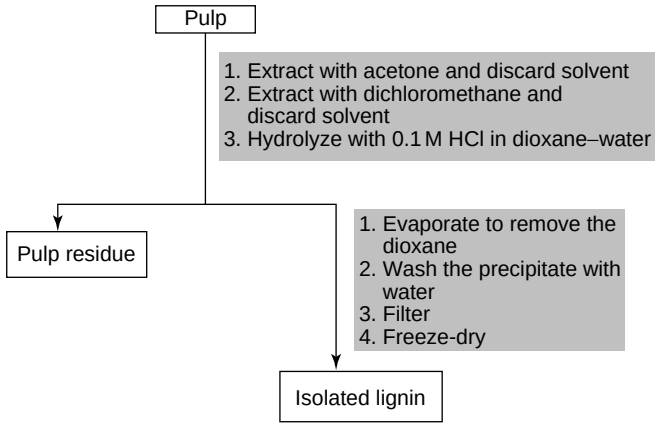

Figure 3 An extraction scheme for obtaining lignin from pulp.

After cooking, the cooking chemicals are removed and the pulp is washed free of the chemicals with water. This washing is an important step that is carefully monitored by measuring the concentrations of the cooking chemicals in the wash water. The cooking chemicals are sent to the recovery furnace where they are regenerated for reuse in the pulping process.

2.3 Recycled Pulps

Recycled or recovered paper is now a major source of fibers in papermaking. The recovered paper is reslushed and the printed ink removed, typically by a flotation process. A major problem with recycled paper is the presence of nonfiber components such as plastics, glues,

adhesives and pressure-sensitive labels. These materials tend to be tacky and stick to process equipment and the paper products, causing tears and holes in the products. Consequently, mill productivity is lowered by these contaminants because the papermaking process is shut down to remove and clean the contaminants from mill equipment. Analysis of these contaminants has shown that they include a diverse list of compounds (see **Pyrolysis in the Pulp and Paper Industry**). To minimize their impact on the pulp and papermaking process, it is important to ascertain their concentrations in the incoming furnish and their removal efficiencies in mill process cleaning unit operations.

Methods that have been used to analyze the contaminants in recycled fibers include Fourier transform infrared (FTIR) spectroscopy, solvent extraction, gel permeation chromatography, NMR, GC, pyrolysis/gas chromatography (Py/GC) and pyrolysis/gas chromatography/mass spectrometry (Py/GC/MS).

3 BLEACHING

Chemical pulps require bleaching to remove residual lignin that was not removed in the cooking process. The bleaching process results in a white pulp that is suitable for printing and writing papers such as copier grades, legal and government documents and high-quality books. The bleaching can be done with elemental chlorine-free reagents such as chlorine dioxide or totally chlorine-free reagents such as hydrogen peroxide, oxygen and ozone. In earlier years, chlorine was the main bleaching chemical until, with the help of analytical chemistry, it was ascertained that the chlorination led to the formation of chlorinated dioxins and furans that were detected in pulp and paper products and in mill effluents. Switching to a nonelemental chlorine bleaching process eliminated the presence of chlorinated compounds in pulp and paper matrices.

Since mechanical pulps are used in the production of lower grades of products, there is no need to remove the lignin present in the pulps. In any case, attempting to remove them by bleaching would be a very expensive process. Instead of bleaching, some grades may be brightened by partial removal of lignin, addition of tinting dyes and addition of fluorescent agents that will make the paper products appear brighter. Such brightened pulps are used mainly in making paper for printing books.

4 PAPERMAKING

After pulping and bleaching or brightening, the fibers are now ready for conversion into paper. To achieve this, a number of additives are added before, during and after the papermaking process. A partial listing of additives and their purpose is shown in Table 1.

To ensure efficient and optimal addition of additives, it is important to monitor for their concentrations in the paper or water matrix. In some cases, the additives

Table 1 Partial listing of additives that are used in pulp and papermaking processes

Class of additive	Addition point in the papermaking process	Purpose of additive	Example of additive
Deresination agent	Before	Improve removal of lipophilic extractives in chemical pulping	Tall oil (a mixture of rosin acid compounds)
Acids and bases	Before and during	Control pH	Sulfuric acid and sodium hydroxide
Wet strength	During	Impart wet strength to toweling and wrapping papers	Epichlorohydrin derivatives
Internal size	During	Impart hydrophobicity (water repellency)	Alkyl ketene dimer
Biocide	During	Control microbiological growths	Methylene bisthiocyanate
Filler	During	Improve paper brightness, reduce fiber content or induce permanence in the paper	Calcium carbonate
Coloring material	During	Impart desired color of the paper	Direct yellow
Defoamer	During	Control or eliminate foam	Silicone defoamer
Dispersant	During	Disperse wood resin particles	Alcohol ethoxylate
Retention aid	During	Retain fibers to form a sheet	Polyacrylamide
Optical brighteners	During	Improve apparent brightness	Diaminostilbene
Pitch and deposit control agents	During and after	Prevent accumulation of deposits	Talcum powder
Surface size	After	Impart hydrophobicity and reduce porosity of paper	Styrene–maleic anhydride
Coating	After	Reduce paper porosity and impart special printing properties	Styrene–butadiene rubber

may be deactivated before reaction with the paper. For example, alkyl ketene dimer sizing agent can hydrolyze to yield products that are very tacky and can cause process upsets by causing the paper to tear. The dissolved and colloidal substances in process waters are negatively charged. Consequently, they compete with fibers for cationic additives (e.g. retention aids) added in the papermaking process. Knowing the types and concentrations of these negatively charged compounds would be of help in optimizing the addition of additives. These negatively charged compounds have preoccupied papermaking chemists for a long time and recently analytical schemes have been devised to characterize them in process waters. For example, one method entails centrifugation of a pulp suspension to remove fibers, collection of the supernatant, liquid–liquid extraction with a nonpolar solvent, derivatization of the extracted compounds and analysis of the derivatized extractives by GC techniques.[7]

Another way to determine the effectiveness of additives is to install deposition plates in the process stream and then wait for deposits to accumulate on the plates. The plates can be periodically monitored visually for cleanliness, giving the operators an idea of what is happening in the process. The analytical chemist will remove the deposits to conduct chemical analyses that will determine the chemistry of the material that will have accumulated on the plates.

5 QUALITY CONTROL

A persistent problem in pulp and papermaking is the presence of dirt spots in the products. Pulp and paper products that are contaminated with the dirt spots have to be culled or sold at below premium prices. A wide variety of compounds can lead to the formation of these unwanted contaminants; they include pieces of bark, fly ash, lipophilic extractives, metal soaps, defoamer components, sizing agents and fillers. A good chemical analysis of these dirt spots is a prerequisite to determining the origin or cause of the problem.[8] Analytical methods that can be used to ascertain the nature of these contaminants include: spot tests (e.g. fizz test with acid to show the presence of calcium carbonate); ash content to determine if the contaminants are organic or inorganic; FTIR spectroscopy to determine the major functional groups present; solvent extraction schemes to separate components such as defoamer oils, wood resin, metal soaps, hydrolyzed size; and chromatographic techniques to understand the chemistry of the separated components. If the contaminants are present in large quantities, then a variety of analytical methods can be used to obtain information about their composition. In some instances, the contaminants may occur as dirt spots that are visible to the naked eye but too small for adequate analysis by techniques that require solvent extraction. In such instances, microanalytical techniques are used. For example, an FTIR instrument with a microscope attachment can be used to analyze very small samples of dirt spots or paper with very little sample preparation required (see **Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry**). Another technique that is useful for such small quantities is Py/GC where the samples are pyrolyzed to yield pyrolysis products that are subsequently analyzed by GC techniques (see **Pyrolysis in the Pulp and Paper Industry**). For semiquantitative determination of inorganic matter, microscopic techniques, specifically energy-dispersive spectroscopy, can be used.

Microscopy is a powerful complementary technique in the analysis of pulp and paper samples, as illustrated in the following example. Our laboratory was asked to ascertain the cause of poor print quality of a newsprint sheet made by one mill. Printing a solid black image on the sheet resulted in tiny white spots, all over the image, that measured about 2 mm across. Obviously the mill customer was not happy about this and sent back the shipment of paper. A visual examination of the sheet showed that there were what appeared to be fiber bundles in the nonprinted areas of the sheet. Examination of a nonprinted sheet also showed the presence of the fiber bundles. The fiber bundles were collected from the sheet using tweezers with the aid of a magnifying glass. Several of the spots were analyzed by FTIR spectroscopy and their spectra were compared with each other and with the spectrum of the good part of the sheet. The data showed that all the spectra were identical and indicated the presence of cellulosic matter. The spots were then analyzed by Py/GC/MS and, again, their pyrograms were identical with that of the good part of the sheet, showing that there was no difference in their chemical makeup. Finally, analysis of the fiber bundles by scanning electron microscopy (SEM) showed that the bundles were actually sclereids (see Figure 4). Sclereids are thick-walled, heavily lignified cells that are found in the inner bark of trees and are the same colour as the wood. Unlike normal wood fibers, the sclereids are difficult to pulp and therefore remain as fiber bundles that cannot absorb ink, hence the presence of white spots in the printed areas of the sheet. If SEM had not been available, we would not have been able to identify the cause of the problem.

Other analytical techniques that are useful in quality control include surface analytical techniques such as secondary ion mass spectrometry (SIMS), electron spectroscopy for chemical analysis (ESCA), and X-ray photoelectron spectroscopy (XPS). Further details on

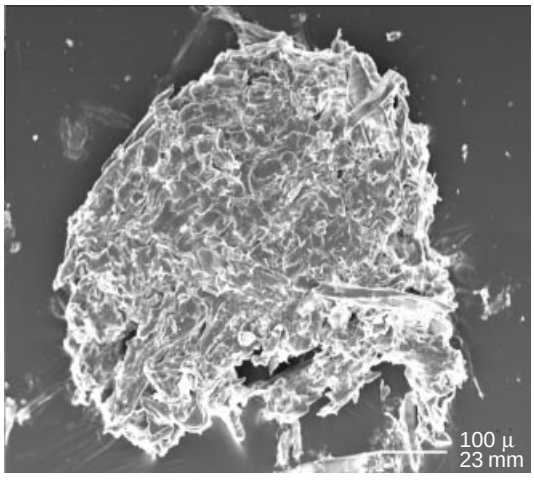

Figure 4 SEM image of a sclereid that caused printing problems on a newsprint sheet.

these techniques can be found in the article on surface analysis by XPS (see **X-ray Photoelectron Spectroscopy, Paper Surface Analysis by**).

6 PROTECTION OF CONSUMERS

Since paper products are used by consumers, it is important to ensure that the products will be safe for the consumer to use. This is especially important for products that will be used in food packaging. The types of compounds that should be monitored in food packaging materials include:[9] (a) compounds that may induce toxicity or allergic reactions, typical examples being halogenated compounds such as dioxins, epichlorohydrin, chloropropanols, solvents and certain dyes; (b) organoleptic substances that may affect the odor and taste of foodstuffs, examples being aldehydes and ketones; and (c) microbiological matter such as Gram-negative bacteria that produce endotoxins.

Regulations of the US Food and Drug Administration (FDA) specify that containers and package components are food additives if they ''may reasonably be expected to become a component, or to affect the characteristics, directly or indirectly, of food packaged in the container''. Methods have been developed to determine the migration of components from food packaging papers by use of a specially designed test cell and using solvents (e.g. water, hexane, chloroform and alcohol) that simulate different foods that may come into contact with the papers. The FDA requires that food-packaging papers made from recycled papers be free of optical brighteners.[10]

Table 2 Odors that may emanate from pulp and paper products

Material	Odor
Paper products	Sulfide, sour, chlorine
Coated paper	Volatile additives
Recycled pulp	Musty
Inks	Oxidizing oils, solvents
Adhesives	Solvents, added perfumes
Wax	Oil, residual solvents

7 ODORS

Odors in paper and board samples can be a nuisance and their analysis can be difficult. Examples of odors from some common paper products and materials associated with paper are shown in Table 2.

GC techniques are useful in characterizing odors from the paper products. For example, paper can be suspended in water in a jar with a gas-tight lid and then heated at 70 °C for 16 h to concentrate the volatiles in the headspace. The headspace gas is then sampled and analyzed by a suitable GC technique. Odor problems in food packaging from a Norwegian mill were attributable to oxidation of lipophilic extractives and degradation of paper additives.[11]

8 ENVIRONMENTAL PROTECTION

Process waters used in the pulp and papermaking process should be purified before discharge into the environment. In a few cases, mills are fully closed, meaning that no effluent waters are discharged. Instead, the process waters are purified by a variety of means and then reused in the pulp and papermaking process. However, the majority of mills do discharge waters into the environment. Such discharges require prepurification by a combination of chemical and biological treatment systems. Environmental regulations mandate that the waters be monitored for a variety of parameters such as chemical oxygen demand (COD), biological oxygen demand (BOD), total suspended solids (TSS), adsorbable organic halogen (AOX) and resin acids (because of their toxicity to fish). A variety of standard test procedures are available to monitor these parameters (see **Pulp and Paper Matrices**). Monitoring the amount of resin acids in mill effluents is important because of their toxicity. Typically, the resin acids are extracted by liquid–liquid extraction from effluents, but the requirements for pH adjustment before extraction are still under debate. Some studies report that optimum recoveries are obtained when

extracting samples adjusted to high pH,[12] whereas others report the opposite.[13]

9 ESTROGEN MIMICS

There is an increased interest in the analysis of phytosterols in pulp and paper mill effluents. Phytosterols are estrogen-like compounds that are present in the unsaponifiable fraction of wood resin. The most prevalent are campestanol, stigmasterol, stigmastanol, β-sitosterol, compesterol, and lupeol. These compounds are suspected to be estrogen mimics because their basic chemical structure is similar to that of estrogen hormones. The estrogen mimics are thought to induce low fertility in the male population. For example, Denton et al.[14] and Hunsinger et al.[15] showed that female mosquitofish were masculinized after exposure to plant sterols, whereas Knutson et al.[16] showed that sterols can affect fish reproduction. If they are indeed estrogen mimics, then it is important to ensure that they are absent in mill effluents discharged into receiving waters or in products used by consumers. However, literature data indicate that they are present in mill effluents and pulps. For example, Holmbom and Lehtinen[17] identified lupeol, cycloartenol and methylenecycloartenol in effluent from a Finnish pulp and paper mill; Oikari and Holmbom[18] identified β-sitosterol and betulinol in the final effluent from a European hardwood and softwood bleached kraft mill; and Jansson et al.[19] identified sitosterol, betulin, lupeol, stigmastanol, campesterol, campestanol, methylenecycloartenol, and cycloartenol in elemental chlorine-free and totally chlorine-free acetone extracts of oxygen bleached pulp. A study of the presence of these estrogen mimics in US mills showed that mill effluent treatment systems removed most of the compounds except for stigmasterol, which appeared to increase in an aerated stabilization basin treatment system.[20] This apparent increase was attributed to other sources of plant sterols such as algae.

ABBREVIATIONS AND ACRONYMS

AOX	Adsorbable Organic Halogen
BOD	Biological Oxygen Demand
COD	Chemical Oxygen Demand
ESCA	Electron Spectroscopy for Chemical Analysis
FDA	Food and Drug Administration
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
IR	Infrared
NMR	Nuclear Magnetic Resonance
Py/GC	Pyrolysis/Gas Chromatography
Py/GC/MS	Pyrolysis/Gas Chromatography/ Mass Spectrometry
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectrometry
TSS	Total Suspended Solids
UV	Ultraviolet
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Starch Analysis in Food

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis

Environment: Water and Waste **(Volume 4)**
Sampling Considerations for Biomonitoring

Food **(Volume 5)**
Lipid Analyses in Food

Pulp and Paper **(Volume 9)**
Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis • Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry • Mechanical Pulps, Ultraviolet/Visible Spectroscopy of Chromophores in • Pulp and Paper Matrices

Pulp and Paper cont'd **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry • X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

REFERENCES

1. L.H. Allen, 'Pitch Control During Production of Aspen Kraft Pulp', *Pulp Pap. Can.*, **89**(10), 87–91 (1988).
2. L.H. Allen, B.B. Sitholé, J.M. MacLeod, C.L. Lapointe, F.J. McPhee, 'The Importance of Seasoning and Barking in the Kraft Pulping of Aspen', *J. Pulp Pap. Sci.*, **17**(3), J85–J91 (1991).
3. J.M. Leach, A.N. Thakore, 'Isolation and Identification of Constituents Toxic to Juvenile Rainbow Trout, *Salmo gairdneri*, in Caustic Extraction Effluents from Kraft Mill Bleach Plants', *J. Fish. Res. Bd. Can.*, **32**(8), 1249–1257 (1975).

4. K. Sundberg, J. Thornton, B. Holmbom, R. Ekman, 'Effects of Wood Polysaccharides on the Stability of Colloidal Wood Resin', *J. Pulp Pap. Sci.*, **22**(7), J226–J230 (1996).
5. A. Kantelinen, O. Jokinen, M.-L. Sarkki, C. Pettersson, K. Sundberg, C. Eckerman, R. Ekman, B. Holmbom, 'Effects of Enzymes on the Stability of Colloidal Pitch', in *Proceedings, 8th International Symposium on Wood and Pulping Chemistry, June 6–9, 1995*, Gummerus Kirjapaino, Helsinki, 605–612, 1995.
6. S.Y. Lin, C.W. Dence, *Methods in Lignin Chemistry*, Springer, Berlin, 1992.
7. F. Orsa, B. Holmbom, 'A Convenient Method for the Determination of Wood Extractives in Papermaking Process Waters and Effluents', *J. Pulp Pap. Sci.*, **20**(10), J361–J366 (1995).
8. B.B. Sitholé, 'Analysis of Pitch Deposits', in *Pitch Control, Wood Resin and Deresination*, eds. E. Back, L.H. Allen, TAPPI Press, Atlanta, GA, Chapter 12, 2000.
9. A. Donetzhuber, B. Johansson, K. Johansson, M. Lovgren, E. Sarin, 'Analytical Characterization of the Gas Phases in Paper and Board Products', *Nord. Pulp Pap. Res. J.*, **14**(1), 48–60 (1999).
10. 'Code of Federal Regulation 21 CFR 176 260, US Food, Drug and Cosmetic Act', Food and Drug Administration, Washington, DC, 1990.
11. K. Wiik, M. Rødbotten, T. Bakken, T. Helle, 'Odor Analysis of Paper. Part II. A Case Study of the Odor Analysis and Abatement', in *Proceedings, Process and Product Quality Conference and Trade Fair*, TAPPI Press, Atlanta, GA, 11–17, 1997.
12. R. Voss, A. Rapsomatiotis, 'An Improved Solvent-extraction Based Procedure for the Gas Chromatographic Analysis of Resin and Fatty Acids in Pulp Mill Effluents', *J. Chromatogr.*, **346**, 205–214 (1985).
13. A.G. Casado, E.J.A. Hernandez, P. Espinosa, J.L. Vilchez, 'Determination of Total Fatty Acids (C_8–C_{22}) in Sludges by Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **826**, 49–56 (1998).
14. T.E. Denton, W.M. Howell, J.J. Allison, J. McCollum, B. Marks, 'Masculinization of Female Mosquitofish by Exposure to Plant Sterols and *Mycobacterium smegmatis*', *Bull. Environ. Contam. Toxicol.*, **35**(5), 627–632 (1985).
15. R. Hunsinger, B. Byram, W. Howell, 'Unchanged Gonadal Morphology of Mosquitofish Masculinized by Exposure to Degraded Plant Sterols', *J. Fish Biol.*, **32**, 795–796 (1988).
16. S. Knutson, K. Solomon, G. Van Der Kraak, P. Hodson, J. Parrott, 'Effect of Sitosterol on Fish Reproduction', in *Abstracts, 2nd SETAC World Congress*, Society of Environmental Toxicology and Chemistry, Pensacola, FL, 183, 1995.
17. B. Holmbom, K-L. Lehtinen, 'Acute Toxicity to Fish of Kraft Pulp Mill Wastewaters', *Pap. Puu*, **62**(11), 673–684 (1980).
18. A. Oikari, B. Holmbom, 'Ecotoxicological Effects of Process Changes Implemented in a Pulp and Paper Mill: A Nordic Case Study', in *Environmental Fate and Effects of Pulp and Paper Mill Effluents*, eds. M.R. Servos, K.R. Munkittrick, J.H. Carey, G.J. Van der Kraak, St. Lucie Press, Delray Beach, FL, 613–625, 1996.
19. M.B. Jansson, P. Wormald, O. Dahlman, 'Reactions of Wood Extractives During ECF and TCF Bleaching of Kraft Pulp', *Pulp Pap. Can.*, **96**(4), 42–45 (1995).
20. D. Cook, 'Development and Evaluation of a Method for the Determination of Phytosterols in Pulp and Paper Mill Effluents', National Council of the Paper Industry for Air and Stream Improvement (NCASI), Research Triangle Park, NC, Technical Bulletin No. 746, 1997.

Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis

Matti Ristolainen and Raimo Alén
University of Jyväskylä, Finland

1 Introduction 1
2 Chemical Composition of Wood 2
3 Delignification Processes 2
3.1 General Aspects 2
3.2 Kraft and Sulfite Pulping 4
3.3 Oxygen–Alkali Delignification and Bleaching 5
4 Sample Preparation 7
4.1 Neutral and Acidic Monosaccharides 7
4.2 Oligosaccharides and Polysaccharides 7
5 Methods 7
5.1 Detection Methods and Limits 7
5.2 Electrolyte Systems 7
5.3 Sample Injection Modes 9
5.4 Interfering Substances 9
6 Example Applications 9
6.1 Neutral and Acidic Monosaccharides 9
6.2 Oligosaccharides and Polysaccharides 10
7 Quality Control 11
7.1 Data Quality Objectives 11
7.2 Laboratory Operations 12
7.3 Instrument Performance 12
8 Comparison with Other Separation Methods 13
Abbreviations and Acronyms 13
Related Articles 14
References 14

Capillary zone electrophoresis (CZE) is one of the novel separation techniques, being applicable, for example, to virtually all important biomass-derived monosaccharides. However, in spite of this fact the utilization of this technique in the pulp and paper industry is currently rather limited. In this chapter practical guidance and some application examples are given to show how carbohydrate material (polysaccharides, oligosaccharides, and neutral and acidic monosaccharides) in pulps and spent liquors can be analyzed by CZE. Emphasis is placed on sample preparation and separation conditions including comparison with other separation methods, such as gas chromatography (GC) and high-performance liquid chromatography (HPLC), generally used for similar purposes. In addition, a short description of other significant degradation products (e.g. aliphatic carboxylic acids and lignin monomers) formed during delignification (e.g. cooking and bleaching) is given. It is concluded that these low-molecular-mass (LMM) degradation products could also be analyzed effectively by CZE.

1 INTRODUCTION

Several techniques have been used to characterize the carbohydrates of chemical pulps and carbohydrate-derived degradation products formed during cooking (present in spent liquors) and bleaching (present in bleaching effluents). Nearly all these carbohydrates originate from cellulose and the main hemicellulose constituents of wood. In addition, minor carbohydrate-containing residues such as arabinogalactans, pectins, and starch can also be regarded as carbohydrate sources.[1]

Pulp properties are mainly caused by different polysaccharides in pulp. However, carbohydrate-based degradation products (aliphatic carboxylic acids and oligo- and polysaccharide fragments) dissolved in black liquor during kraft delignification have an influence, for example, on the combustion properties of black liquor when it is burnt in the mill's recovery furnace. Therefore, detailed data on all kinds of carbohydrate-based substances, including their molecular mass distribution, are of great industrial importance.

The wood-based carbohydrates have conventionally been analyzed by various chromatographic systems: GC, HPLC, ion chromatography (IC), or gel permeation chromatography (GPC).[2–6] Capillary electrophoretic systems such as CZE, micellary electrokinetic chromatography (MEKC), capillary gel electrophoresis (CGE), and capillary electrochromatography (CEC) offer the most promising novel techniques for this purpose.[7–9] In particular, many CZE systems have been developed for the analysis of plant-derived carbohydrates. This is primaily due to the possibility of directly applying these systems, which have high separation efficiency and a short analysis time, to aqueous samples with a relatively wide molecular mass range.

The possibility of analyzing high-molecular-mass (HMM) carbohydrates has so far not been widely used for wood-processing applications. In contrast, most studies related to these applications have concentrated on neutral and acidic monosaccharides and only a few, for example, on oligosaccharides.[10,11] Thus, most

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

analyses of oligo- and polysaccharides have applied to other feedstocks (e.g. cotton[12] and food[13]) than wood and pulp. However, several substances, such as carboxylic acids,[14–27] phenols,[25,28–30] and inorganic compounds,[14,15,17,18,20–27] also typical of pulp production, can be analyzed, for example, by CZE. Such compounds are discussed in this article only as interfering substances of carbohydrate analyses, with the main emphasis being paid to neutral and acidic carbohydrates.

2 CHEMICAL COMPOSITION OF WOOD

The major chemical constituents of all wood species are the structural substances cellulose, hemicelluloses, and lignin. Other polymeric constituents, present in lesser and often varying quantities, are pectin, starch, and proteins. In addition to these macromolecular components, various nonstructural and mostly LMM compounds (extractives, some water-soluble organics, and inorganics) can be found in small quantities in both softwoods and hardwoods.

Softwoods and hardwoods differ typically from each other with regard to their chemical composition.[1,31–33] In both cases, the cellulose (degree of polymerization (DP) = 10 000–15 000, average molecular mass ($\overline{M}_w$) 1.6–2.4 × 10^6 Da) content is more or less the same (40–45% of the wood dry solids), but softwoods usually contain less hemicelluloses (DP = 100–200, $\overline{M}_w$ = 15 000–20 000 Da) and more lignin ($\overline{M}_w$ = 15 000–20 000 Da). The typical content of hemicelluloses in temperate-zone softwoods and hardwoods is, respectively, 25–30% and 30–35% of the wood dry solids. The lignin content of softwoods is typically in the range 25–30% of the wood dry solids, whereas the common lignin content of hardwoods varies between 20 and 25% of the wood dry solids. The other compounds (mainly extractives) in woods from temperate zones usually make up about 5% of the wood dry solids, but tropical species often exceed this value. Thus, in woods from temperate zones, the macromolecular substances building up the cell walls account for about 95% of the wood material. In contrast, for tropical woods this value may decrease to an average value of 90%.

It is typical that the structure of hemicelluloses is different between softwoods and hardwoods, whereas cellulose is a uniform component of all woods. Scheme 1 shows the partial chemical structure of cellulose[34,35] and the main hemicellulose constituents.[36,37] In softwoods the primary hemicellulose component is the acetylated galactoglucomannan (glucomannan, 15–20% of the wood dry solids), the nonacetylated arabinoglucuronoxylan (xylan) being a minor constituents (5–10% of the wood dry solids). In contrast, in hardwoods the primary hemicellulose component is the acetylated xylan (20–30% of the wood dry solids) and clearly less nonacetylated glucomannan (less than 5% of the wood dry solids) is present. Arabinogalactan occurs significantly (10–20%) in the heartwood of larches (*Larix sibirica/L. decidua*), whereas its content in other softwoods is generally less than 1%.[1,32] In addition, different galactans (acidic galactans) are present especially in reaction wood (i.e. compression wood in softwood and tension wood in hardwood).[1,32,36,38]

The building units of wood carbohydrates are pentoses (D-xylose, L-arabinose, and D-arabinose), hexoses (D-glucose, D-mannose, and D-galactose), and deoxyhexoses (L-rhamnose or 6-deoxy-L-mannose). Small amounts of certain uronic acids (4-*O*-methyl-D-glucuronic acid, D-galacturonic acid, and D-glucuronic acid) are also present. These units exist mainly as six-membered (pyranose) structures either in the α or β forms (Scheme 2).

Lignin is an amorphous polymer with a chemical structure that differs distinctly from the other macromolecular constituents of wood.[39–41] It is also characteristic that, unlike wood carbohydrates, the chemical structure of lignin is irregular in the sense that different structural elements (phenylpropane units) are not linked to each other in any systematic order. However, the close association between lignin and carbohydrate components in wood strongly suggests the existence of a chemical linkage between these constituents forming a lignin–polysaccharide complex (LPC) or a lignin–carbohydrate complex (LCC).[1,32]

The final organic constituent group – extractives – comprises an extraordinarily large number of diverse substances (i.e. several thousands of individual compounds) mainly with low molecular masses and which by a broad definition are either soluble in neutral organic solvents or water.[42–44] These substances belong to both lipophilic and hydrophilic types and are regarded as nonstructural wood constituents. The ash contents of commercial softwoods and hardwoods are generally in the range 0.3–1.5% of the wood dry solids.

3 DELIGNIFICATION PROCESSES

3.1 General Aspects

Pulping refers to different processes by which wood is converted into a product mass with liberated fibers predominantly used for papermaking.[1,31,45–48] These thermal conversions can be accomplished either chemically or mechanically, or by a combination of these two types of treatment. Thus, with regard to pulping

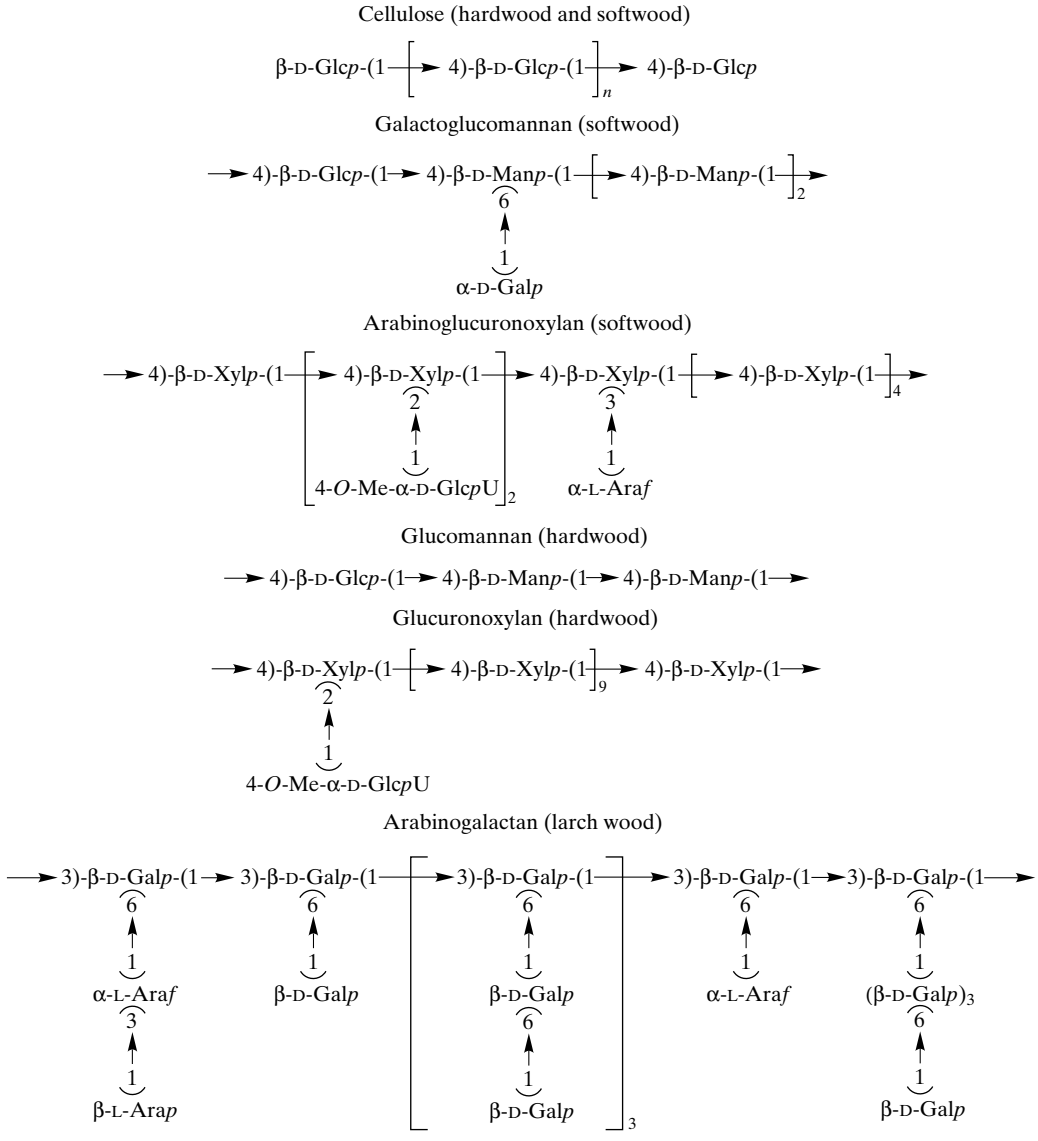

Scheme 1 Partial chemical structure of cellulose and the main hemicelluloses in softwood and hardwood.

processes, the term ''pulp'' is used collectively for chemical (yield 45–55% of wood feedstock), semichemical (yield 65–85% of wood feedstock), chemimechanical (yield 80–90% of wood feedstock), and mechanical (yield 91–98% of wood feedstock) pulps. Chemical pulping accounts for about 70% of the total worldwide production, and nowadays about 80% of chemical pulp is produced by the dominant kraft (sulfate) process. The importance of sulfite pulping has clearly decreased during recent decades. All the main chemical constituents (cellulose, hemicelluloses, lignin, and extractives) of wood feedstocks behave differently during wood processing, resulting in differences in the chemical composition of pulp products and waste liquors.

A general goal in cooking prior to bleaching is to obtain as low a kappa number (i.e. the residual lignin content of pulp) as possible without deactivating the residual lignin and with no degradation of product quality. Another current trend is the closing of process water circulation, to decrease the wastewater load. Thus, a totally effluent-free (TEF) mill represents the ultimate objective in pulp production. To avoid the corrosion problems caused by chlorine-containing compounds, oxygen-based chemicals (oxygen, ozone, hydrogen peroxide, and peracids) are used in the totally chlorine-free (TCF) bleaching process. The proportion of elemental chlorine-free (ECF; i.e. the use of chlorine dioxide and hypochlorite, but not chlorine) bleached pulps is gradually increasing worldwide, being about 50% today, whereas the corresponding TCF production is about 6%.

The aim of this section is to give an overview of delignification-based compounds that could be analyzed

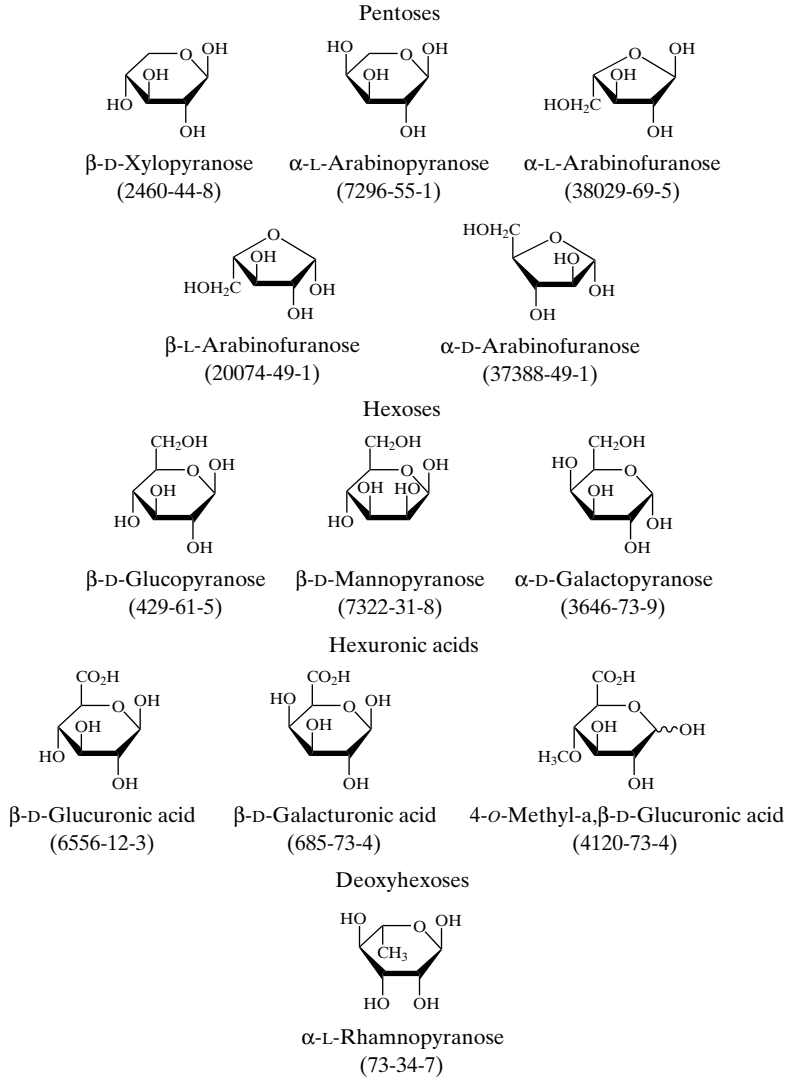

Scheme 2 The sugar moieties of wood hemicelluloses.

by different CZE techniques. However, only a few such applications have been reported.

3.2 Kraft and Sulfite Pulping

During kraft pulping,[1,32,46,47,49] or generally during the conditions prevailing at alkaline pulping, roughly half of the wood substance is degraded and dissolved into the cooking liquor, i.e. the black liquor. The organic matter in the black liquor is composed of the degradation products of lignin and polysaccharides in addition to a minor fraction comprising extractives (Table 1).[1,45,50,51] A substantial part of the hemicelluloses is converted to hydroxy carboxylic acids (Scheme 3). In addition, small amounts of dissolved polysaccharides are not completely degraded and can be found in the final black liquor. Dissolution and degradation of lignin during kraft pulping also produces a complex mixture of breakdown products with wide molecular mass distributions, ranging from simple LMM phenols (Scheme 3) to large macromolecules. Of the charged alkali, 70–75% is required to neutralize the aliphatic carboxylic acids and about 20% is consumed to neutralize the degradation products of lignin.

Figure 1 illustrates the typical material balance of wood organics for the kraft process (including the subsequent oxygen–alkali delignification) and the bleaching sequence overall, which produces kraft pulp with high brightness. As can be seen, approximately 90% of the lignin, 60% of the hemicelluloses, and 15% of the cellulose is dissolved during the cook.

There are several modifications of the sulfite method, designated according to the pH of the cooking liquor and producing pulps. These range from dissolving pulps for chemical end-uses to high-yield neutral

Table 1 Typical composition of the dry matter of scots pine (*Pinus sylvestris*) and silver birch (*Betula pendula*) kraft black liquors (percentage of the total dry matter)[31]

Component	Pine	Birch
Lignin	31	25
HMM (>500 Da) fraction	28	22
LMM (<500 Da) fraction	3	3
Aliphatic carboxylic acids	29	29
Formic acid	6	4
Acetic acid	4	8
Glycolic acid	2	2
Lactic acid	3	3
2-Hydroxybutanoic acid	1	5
3,4-Dideoxypentonic acid	2	1
3-Deoxypentonic acid	1	1
Xyloisosaccharinic acid	1	3
Glucoisosaccharinic acid	6	3
Others	3	3
Other organics	7	9
Extractives	4	3
Carbohydrates[a]	2	5
Miscellaneous	1	1
Inorganics[b]	33	33
Sodium bound to organics	11	11
Inorganic compounds	22	22

[a] Mainly hemicelluloses-derived fragments.
[b] Due to the presence of dead-load inorganics, this mass proportion may be higher in some cases.

sulfite semichemical (NSSC) grades.[1,32,52] The AQ (anthraquinone) alkaline sulfite methods lead generally to kraft-type pulps. All these cooking modifications result in differences in the chemical composition of pulp and spent liquor.

Typical sulfite spent liquors differ from alkaline pulping liquors in many respects, the majority of the organic material originating from lignosulfonates and hemicelluloses (Table 2). In principle, a variety of useful products can be produced from these liquors, but today most of their organic solids are burned with the generation of energy and recovery of the cooking chemicals.

3.3 Oxygen–Alkali Delignification and Bleaching

Oxygen–alkali delignification or simple oxygen delignification is a common process for delignifying unbleached pulps because by this process water pollution can be substantially reduced.[1,32,53] One of the most important barriers to extending oxygen delignification is the severe degradation of carbohydrates at higher degrees of delignification leading to inferior mechanical properties of fibers and low pulp yields. Due to this nonselectivity, delignification degree of the kraft pulp from conventional coking is normally in the range 35–50%, if a suitable inhibitor (i.e. a protector against carbohydrate degradation) is added. However, this technique offers the most prominent alternative to chlorination and a suitable way of decreasing the use of chlorine dioxide as well.

The overall composition of the spent liquors from delignification can be characterized by dividing the material into common groups of constituents, because only a minor fraction of the total material constitutes distinctly definable compounds. Table 3 illustrates the

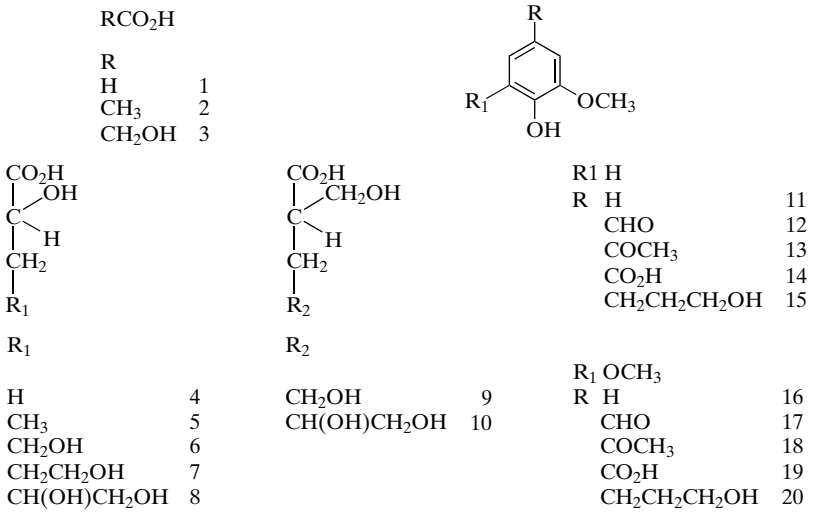

Scheme 3 The main LMM organic compounds in softwood and hardwood black liquors. Aliphatic carboxylic acids: formic acid (1), acetic acid (2), glycolic acid (3), lactic acid (4), 2-hydroxybutanoic acid (5), 3-deoxytetronic acid (6), 3,4-dideoxypentonic acid (7), 3-deoxypentonic acid (8), xyloisosaccharinic acid (9), and glucoisosaccharinic acid (10). Lignin monomers in softwood black liquors: guaiacol (11), vanillin (12), acetovanillone (13), vanillic acid (14), and dihydroconiferyl alcohol (15). Lignin monomers in hardwood black liquors: (11–15), syringol (16), syringaldehyde (17), acetosyringone (18), syringic acid (19), and dihydrosinapyl alcohol (20).

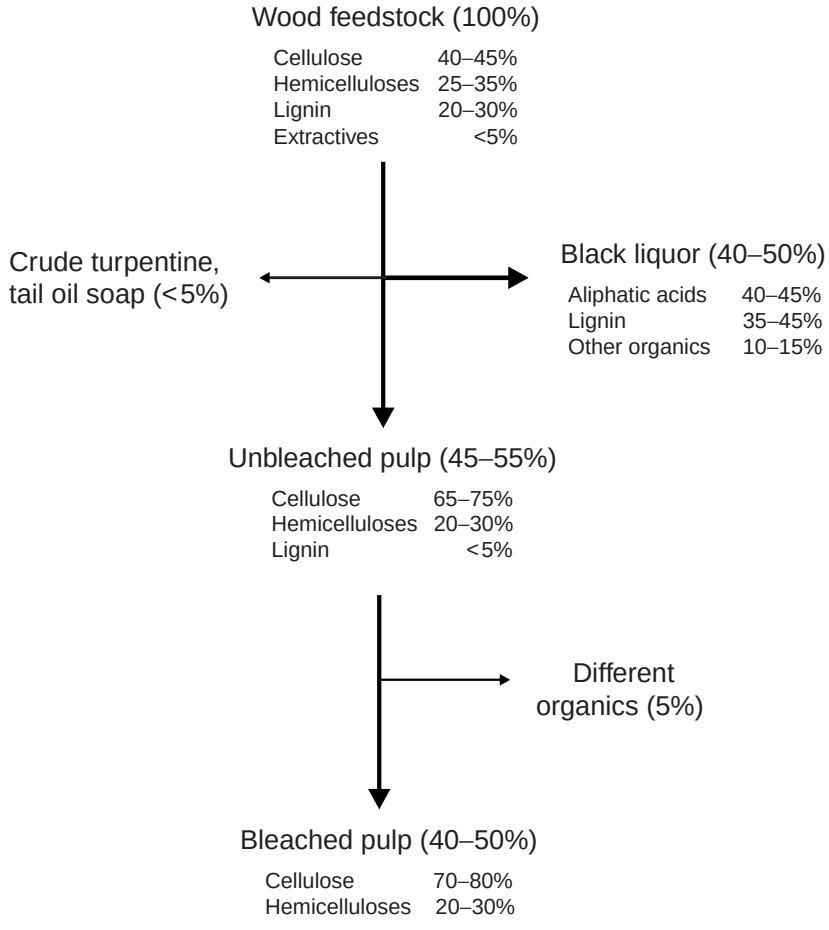

Figure 1 Typical material balance of wood organics for the kraft process for producing bleached pulp.

Table 2 Typical composition of Norway spruce (*Picea abies*) and silver birch (*Betula pendula*) spent acid sulfite liquors (kilograms per ton of pulp)[45]

Component	Spruce	Birch
Lignosulfonates	510	435
Carbohydrates	270	380
Monosaccharides	215	305
Arabinose	10	5
Xylose	45	240
Galactose	30	5
Glucose	25	10
Mannose	105	45
Oligo- and polysaccharides	55	75
Carboxylic acids	70	130
Acetic acid	30	75
Aldonic acids	40	55
Extractives	40	40
Others	30	55

typical composition of the organic fraction from kraft pulping, oxygen delignification, and TCF bleaching. These effluents differ typically from each other indicating

Table 3 Typical composition of dissolved organics in different spent liquors (percentage of the total amount)[45]

Component	Kraft pulping	Oxygen–alkali delignification	TCF bleaching
Lignin	45	50	10
Aliphatic carboxylic acids	45	25	40
Hemicelluloses	<5	15	45
Extractives and other organics	5	10	<5

clearly the nature of each process. It should be noted that, although the soluble extractives remain almost unchanged during oxygen delignification, they may be somewhat modified after bleaching.

In general, bleaching can be defined as a multistage chemical process applied to chemical and mechanical pulps to increase their brightness.[1,32,53,54] The bleaching reactions of pulp components are highly complex and can result in a wide variety of LMM degradation products. It can be concluded that CZE techniques are useful for analyzing many of these degradation products.

4 SAMPLE PREPARATION

4.1 Neutral and Acidic Monosaccharides

Depending on the samples and the CZE procedures used, different sample pretreatments may be required. In particular, removal of lignin-derived fragments, extractives, carboxylic acids, and inorganic substances may be nessesary if, for example, the black liquor is analyzed directly without acid treatment (i.e. by hydrolysis of hemicelluloses). These noncarbohydrate substances can interfere with derivatization, detection, separation of carbohydrates or electrokinetic sample introduction.

Purification of carbohydrate samples can be performed in many different ways.[55] The removal of inorganic substances and carboxylic acids can be accomplished by GPC, or a weakly basic anion exchange resin. Most of the lignin and extractives can be removed by extraction with organic solvents or by solid phase extraction (SPE; C18 or NH_2 phase).

4.2 Oligosaccharides and Polysaccharides

Oligo- and polysaccharides are analyzed either directly or after hydrolysis. The same sample purification systems can be also utilized in this case. The three methods that have been reported are sulfuric acid,[11] trifluoroacetic acid (TFA),[10,56–60] and enzymatic[10] hydrolysis. Of these methods, enzymatic hydrolysis is the most gentle and specific, although it is the most time-consuming technique. Furthermore, it allows the liberation of different uronic acid side groups of hemicelluloses. Hydrolyzates may also contain different oligomers after enzymatic peeling due to incomplete hydrolysis, and their detailed indentification can be difficult because of a lack of commercial model compounds.

The acid hydrolysis methods are usually utilized instead of enzymatic peeling because they are less time-consuming and the reagents are easily available. However, a significant drawback of acid hydrolysis is the loss of uronic acids during treatment.[55] The TFA can be removed after hydrolysis by evaporation, but the sulfuric acid should be neutralized either by an anion exchange resin or by calcium carbonate. In the latter case, however, some epimerization of monosaccharides may occur. An insoluble noncarbohydrate residue formed comprises mainly lignin and can be filtered off after acid treatment.[55]

5 METHODS

5.1 Detection Methods and Limits

The main detection systems in commercial capillary electrophoresis (CE) devices are ultraviolet (UV) or laser-induced fluorescence (LIF) detectors. In addition, a refractive index (RI) detector, a pulse amperometric detector (PAD), and an amperometric detector with constant potential (ADCP) have been used. The use of a mass-sensitive detector (MSD) for carbohydrate analysis is limited because it is difficult to find a buffer solution suitable for both separation and detection.

The main difficulties arise from the poor detection of carbohydrates by UV or LIF, as these compounds do not contain chromophores or fluorophores. In spite of this, carbohydrates have generally been detected directly by UV at 195 nm.[61] The sensitivity of this method is not very good and the detectability of carbohydrates has been improved by an indirect UV[62] or an indirect fluorescence detection method.[63] However, most migrated compounds respond to indirect UV and fluorescence detection due to displacement of the background electrolyte. Therefore, noncarbohydrate-based compounds can interfere with carbohydrate detection in these methods. Table 4 shows the different detection methods with the mass detection limits.

More uncommon detection methods such as PAD,[64] ADCP,[66] and RI[64] have also been used. These are nondiscriminatory techniques and the positive responses originate from, for example, carboxylic acids and aromatic compounds can confuse the peak assignment.[64] It is also possible that strongly alkaline conditions, which are used especially with the amperometric detectors, can lead to epimerization and chemical degradation of carbohydrates.[64] Noncompatibility of the gradient elution techniques is a further drawback of an RI detector. The sensitivity of this detection can be increased by extended light path capillaries or z-type detection cells,[68] but the most common method for improving detectability of carbohydrates is derivatization. This is normally based on reductive amination or condensation reactions.[64,69] Some common derivatization reagents are listed in Table 4.

Schwaiger et al.[67] have studied optimization of precolumn derivatization by 4-ABN. The high sensitivity in this case was achieved by derivatization of 6% 4-ABN solution for aldoses, ketoses, and uronic acids. A derivatization time of 15 min was sufficient at 90 °C. This study also showed that derivatization improved both sensitivity and resolution, mainly because the derivatized carbohydrates were present only in their open-chain forms.

5.2 Electrolyte Systems

The common buffer systems are based on sodium hydroxide, borate (made from boric acid or borax) or phosphate with the different background electrolytes and organic modifiers. Sodium hydroxide electrolytes have been used especially with ADCP or PAD.[66] The

Table 4 Common detection methods for carbohydrates

Detection method	Mass detection limit	Derivatization reagent or background electrolyte
ADCP[66] or PAD[64]	<50 fmol (ADCP)	
Direct fluorescence		
325/515 nm[a],[13] 325/514 nm,[8] 257/520 nm[8]	Few attomoles (325/515 nm)	ANTS
Direct UV		
195 nm[61]	Nanogram range	
214 nm[13,64]	15 fmol	ANTS
240 nm[12,57]	10 pmol[b]	2-AP
245 nm[65]	50 fmol	PMP
245 nm[10,12]	Few femtomoles	6-AQ
280 nm,[56] 285 nm,[67] 286 nm[11]	<10 fmol	4-ABN
280 nm,[56] 305 nm[58]	7 fmol (305 nm), 30 fmol (280 nm)	Et-4-AB
285 nm[59,60]	15–30 fmol	4-ABA
Indirect fluorescence		
442 nm[63]	2 fmol	Coumarin 343
Indirect UV		
227 nm[62]	0.05 mM[c]	Benzoic acid
256 nm[62]	0.1 mM[c]	Sorbic acid
267 nm[62]	0.1 mM[c]	Riboflavin
305 nm[62]	0.1 mM[c]	4-Aminosalicylic acid
RI[64]	0.1 mM[c]	

[a] Excitation/detection wavelength.
[b] Limit of detection.
[c] Concentration detection limit value.
ANTS, 8-aminonaphthalene-1,3,6-trisulfonic acid; 2-AP, 2-aminopyridine; PMP, 1-phenyl-3-methyl-5-pyrazolone; 6-AQ, 6-aminoquinoline; 4-ABN, 4-aminobenzonitrile; Et-4-AB, ethyl 4-aminobenzoate; 4-ABA, 4-aminobenzoic acid.

advantage is that derivatization can be avoided, because carbohydrates are dissociated under highly alkaline conditions. In addition, detection can be improved to the picomole range by indirect UV, for example with sorbic acid.[62] However, strong alkaline conditions can also induce epimerization and degradation. Borate and diethylamine (DEA) have been used to improve the separation effiency.[62]

Borate is normally considered as a buffer and sodium hydroxide as a chemical agent for adjusting pH. In the case of the borate buffer, which is the most common electrolyte used in the separation of monosaccharides, the pH is kept at a lower level than in the case of the sodium hydroxide electrolyte. The separation power of the former case is based on the negatively charged carbohydrate–borate complexes and their stability, which depends on the hydroxyl group positions in the carbohydrate structure. The orientation of the vicinal hydroxyl groups at C4/C5 and C3/C4 for arabinose, mannose, and glucose interferes with the mutual separation of these monosaccharides due to their almost equal electrophoretic mobilities.[56] The separation selectivity has been optimized, especially with a high borate concentration (200–500 mM), or with organic electroosmotic modifiers such as hexadimethrine bromide (HDB), acetone, acetonitrile, ethylene glycol, or aliphatic alcohols.[56] In addition, pH adjustment (pH 9.0–10.5) has been used to facilitate separation.[11,56,58,70]

Little attention has been given to other means of separation, mainly because of the drawbacks of the above-mentioned methods. A high borate concentration increases the Joule heating due to the high electric current and there is also a risk of precipitation. The stability of well-prepared 500 mM borate buffer has shown to be good under prolonged storage at the room temperature.[11] The formation of borate crystals can begin if the dissolution of borate into water is reduced, as for example during storage in the refrigerator. In some cases the organic modifiers can affect the buffer stability. However, organic modifiers have been shown to be very useful for changing the electroosmotic properties of borate buffer systems and to significantly shorten the analysis time.[56] Other approaches, such as the use of coated capillaries and voltage or buffer programming, have not reported in conjucation with the analysis of wood-based monosaccharides. Polyacrylamide[8] or polyether[12,71] coatings have been used but no data on voltage or buffer programming have been given for oligo- and polysaccharide separation systems.

Carbohydrates can also be separated under acidic (pH 2.0–2.5) conditions with a 30–200 mM phosphate buffer.[13] In this case the charge should be introduced

into the carbohydrate molecule because phosphate is not able to form complexes with carbohydrates and carbohydrates are not dissociated under acidic conditions. Reagents such as ANTS, 2-AP, and 6-AQ are used to introduce the ionogenic function.[12,13] The separation selectivity of phosphate buffer has been improved by tetrabutylammonium bromide.[12] These buffer systems are mainly used for the separations of oligo- and polysaccharides.

5.3 Sample Injection Modes

Samples can be introduced into the capillary by hydrodynamic or electrokinetic methods. In most cases, sample introduction is done in the hydrodynamic mode by applying vacuum[58–60,67] or pressure[11] for 1–10 s, but the siphoning[10,56,57,62,66] method has also been used. Electrokinetic sample introduction[66] is instrumentally the simplest alternative, but has been used only in limited cases for carbohydrate analysis. Its main problem is sampling bias, which leads to a poor reproducibility if, for example, sample matrices differ from each other or the sample contains multiple components as pulp-derived samples usually do.[72] The use of internal standards is a convenient way of improving the reproducibility and accuracy of the results.

5.4 Interfering Substances

Most samples from chemical pulping contain some interfering substances (Table 5). These substances have been reported to complicate carbohydrate analysis only in a few cases. Adsorption of enzymes (proteins) on to the silica capillary surface is the commonly known problem[72] that affects the electroosmotic flow. Carboxylic acids have been shown to make the peak assignment of carbohydrates more difficult in the nondiscrimated detection systems with sodium hydroxide electrolyte.[64]

Different methods can be used to avoid interference by aromatic compounds and carboxylic acids.[55] This is especially important when the carbohydrate composition is determined from samples with difficult matrices such as black liquors or bleaching effluents without acid treatment. In acid hydrolysis and especially in sulfuric acid hydrolysis, lignin material is precipitated and can be removed by filtration or centrifugation. Lignin fragments can also be removed, for example, by SPE, GPC, or by extraction with different organic solvents. Volatile carboxylic acids are evaporated during hydrolysis or sample preparation, but different extraction methods can also be applied. The potential interference of carboxylic acids can be reduced by selecting a suitable derivatization method and detection wavelength.

6 EXAMPLE APPLICATIONS

6.1 Neutral and Acidic Monosaccharides

Several CZE methods that have been developed for separating a wide range of neutral and acidic monosaccharides are also suitable for separating pulp-derived monosaccharides (Table 6). Examples of separated analytes are shown with their migration order and mass detection limits in Table 7. In most cases a borate buffer with different concentrations has been used, but sodium hydroxide solution with or without organic additives has also been applied. Only two of these systems have been directly used for analysis of pulp-derived carbohydrates. Rydlund and Dahlman[10] investigated spruce xylan of TMP after TFA and enzymatic hydrolysis. Neutral and acidic 6-AQ-derivatized monosaccharides were separated within less than 16 min by using 420 mM borate buffer. Ristolainen[11] has analyzed the carbohydrate composition of TCF bleaching effluents by using 500 mM borate buffer (Figure 2). In this investigation, CZE and GC separations were also compared to each other. The results indicated that the relative standard deviations for monosaccharide concentrations of the separately hydrolyzed samples were 1.5–8.7% in the CZE systems, whereas in the GC systems these values varied between 2.9 and 5.6%.

The baseline separation of arabinose, glucose, and mannose has generally been a difficult task in many

Table 5 Interferences of CZE analysis by chemical pulping-related carbohydrates

Source	Negative effects	Solution proposal
Aromatic materials, carboxylic acids, hydroxy carboxylic acids, or extractives	Detection, peak assignment, or electrokinetic sample introduction	Preliminary separations; acid hydrolysis; derivatization; other detection methods or by wavelength
Enzymes from enzymatic hydrolysis	Adsorption on to the silica surface[72]	Preliminary separations; use of high or low pH buffers; washing the capillary between runs; adding modifiers to the buffer; modification of the capillary wall
Inorganic salts	Electrokinetic sample introduction	Hydrodynamic sample introduction; preliminary separations

Table 6 Summary of CZE methods developed for the separation of derivatized carbohydrates from different sources

Method	Sample	Capillary			Buffer, pH	Detection (nm, derivate)	Hydrolysis
		Length (cm) × diameter (μm)	Temperature (°C)	Applied potential (kV)			
1	Arabic gum[57]	65 × 50		15	200 mM borate, 10.5	240, 2-AP	2 M TFA
2	Model substances[61]	94 × 75	60	20	50 mM borate, 9.3	195	
3	*Flos matricariae*[58]	72 × 50	30	25	200 mM borate, 10.5	305, Et-4-AB	6 M TFA
4	Model substances[66]	73 × 50		11	100 mM NaOH	ADCP (Cu)	
5	*Radix althaeae*[59]	72 × 50	30	28	150 mM borate, 10.0	285, 4-ABA	4 M TFA
6	Model substances[67]	72 × 75	30	25	175 mM borate, 10.5	285, 4-ABN	
7	Isolated birch xylans[60]	72 × 50	30	28	150 mM borate, 10.0	285, 4-ABA	2 M TFA
8	Spruce TMP xylan[10]	43 × 30		1.2 W	420 mM borate, 9.0	245, 6-AQ	2 M TFA
9	Sugar cane bagasse[56]	57.5 × 50	26	−18	350 mM borate[a], 10.5	280, 4-ABN	2 M TFA
10	Model substances[62]	90 × 50		17	5% DEA, NaOH, 12.6	256, Sorbic acid	
11	Bleaching effluent[11]	48.5 × 50	20	16.0	500 mM borate, 9.5	286, 4-ABN	H_2SO_4

[a] Buffer contained also 5% methanol, 5% acetonitrile, and 0.001% HDB to modify the electroosmotic flow. TMP, thermomechanical pulp.

Table 7 List of pulp-derived carbohydrates analyzed by several CZE methods and their respective mass detection limits

Method[a]	Neutral monosaccharides[b]						Acidic monosaccharides[b]			Oligosaccharides[c]			Mass detection limit
	Arabinose	Galactose	Glucose	Mannose	Rhamnose	Xylose	4-*O*-Methyl-glucuronic	Galacturonic	Glucuronic	Cellooligosaccharides	Mannooligosaccharides	Xylooligosaccharides	
1	**D**	**E**	C		**A**	B		G	F				10–100 mM
2		B	C	A		D		E	F	DP 2			Nanogram range
3	**D**	**E**	**C**		**A**	B		**G**	**F**	DP 2			7 fmol[d]
4		A	B	D	C	E				DP 2			<50 fmol[e]
5	**C**	**D**	**B**	B		**A**		**F**	**E**	DP 2			15 fmol[d]
6	D	C	C		A	B		F	E	DP 2			1 fmol[d]
7	**C**	**D**	**B**			**A**		F	E				15–30 fmol
8	**E**	**F**	**C**	**D**	A	**B**	**G**	I	H			DP 2–6	Few femtomol[f]
9	**D**	**C**	**F**	E	H	**G**		**A**	B	DP 2			30 fmol
10	C	A	B	F	D	E							0.1 mM
11	**E**	**F**	**C**	**D**	A	**B**	G	I	H	DP 2–6	DP 2–5	DP 2–5	<10 fmol[g]

[a] Numbering, see Table 6.
[b] Order of migration as indicated in alphabetical order. The bold letter indicates that analyte was found in the sample (see Table 5).
[c] Range of oligosaccharides as indicated by the DP.
[d] Measured for glucose.
[e] Measured for fructose.
[f] Measured for glucuronic acid.
[g] Measured for neutral monosaccharides.

investigations. Nguyen et al.[56] have expected that a complete separation could not be achieved with borate buffer, even if the concentration was increased to a level of 500 mM (Figure 3). This separation problem was avoided by addition of organic modifiers such as 5% methanol, 5% acetonitrile, and 0.001% HDB into the 350 mM borate buffer. The additives changed the direction of the electroosmotic flow and the separation was made by the so-called co-electroosmotic techniques. This kind of procedure markably shortened the total analysis time. Ristolainen[11] has recently demonstrated the complete baseline separation with 500 mM borate buffer without organic modifiers. In this case, however, the analysis time was much longer and especially the peaks of acidic monosaccharides were much wider than those detected in the case of the co-electroosmotic system. Voltage programming was not used; the migration time of galactose and acidic monosaccharides could be significantly shortened using this approach.[68]

6.2 Oligosaccharides and Polysaccharides

The development of CZE separation systems for wood-based oligo- and polysaccharides has been surprisingly

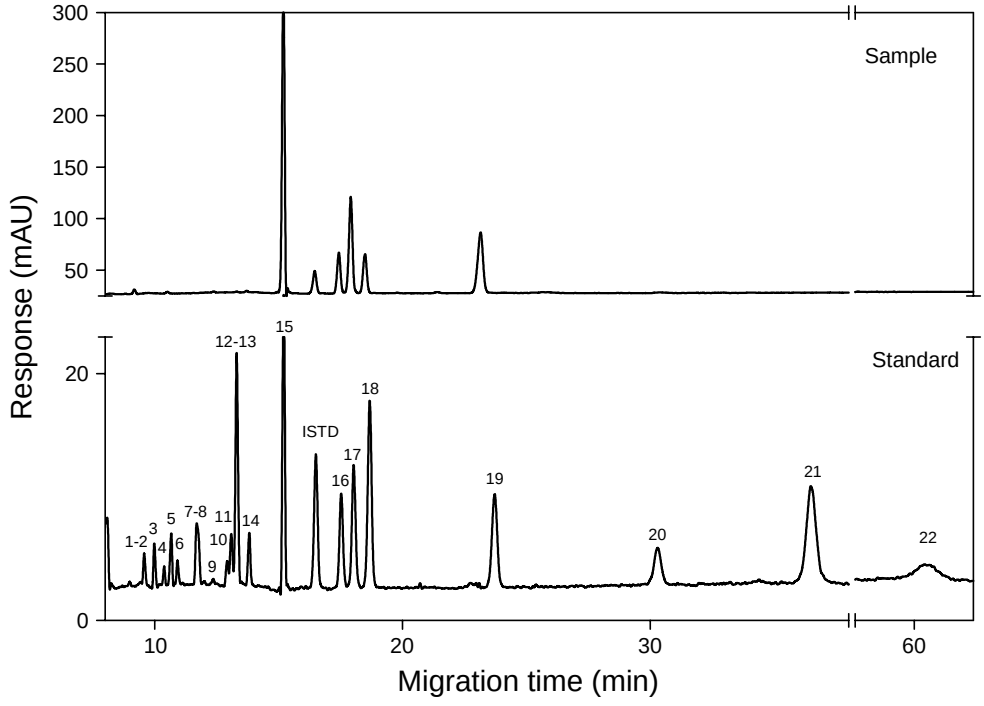

Figure 2 Separation of 4-ABN derivatives of carbohydrates in a standard carbohydrate solution and the hydrolyzate (H_2SO_4) from the bleaching effluent.[11] Peak assignment: xylopentaose (1), cellohexaose (2), xylotetraose (3), cellopentaose (4), xylotriose (5), cellotetraose (6), xylobiose (7), cellotriose (8), mannopentaose (9), mannotetraose (10), mannotriose (11), rhamnose (12), mannobiose (13), cellobiose (14), xylose (15), glucose (16), mannose (17), arabinose (18), galactose (19), 4-*O*-methylglucuronic acid (20), glucuronic acid (21), and galacturonic acid (22). (Reprinted from M. Ristolainen, 'Characterization of Totally Chlorine-free Effluents from Kraft Pulp Bleaching. II. Analysis of Carbohydrate-derived Constituents after Acid Hydrolysis by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **832**, 203–209, Copyright (1999), with permission of Elsevier Science.)

slow, even though it is evident that this technique, as well as the other electrophoretic separation techniques, has great potential to separate these compounds. Rydlund and Dahlman[10] have studied xylooligosaccharides up to xylohexaose in addition to some of their acid forms. Ristolainen[11] has studied wood-based oligosaccharides but only with model compounds (Figure 2). A relatively good separation was achieved for cellooligosaccharides (DP 2–6), mannooligosaccharides (DP 2–5), and xylooligosaccharides (DP 2–5). It is also possible to analyze simultaneously monosaccharides by this system, thus representing a suitable way of monitoring hydrolysis. The amount of cellobiose has been determined by several other CZE systems.[56,58,59,61,66,67]

The CZE analysis of polysaccharides is utilized in areas other than those related to wood processing and its utility in separating maltopolysaccharides up to 30 DP units has been demonstrated.[13] Higher separation efficiencies have been achieved by other electrophoretic separation techniques such as the use of CGE for 3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde-derived hydrolyzates of poly(galacturonic acid).[9]

7 QUALITY CONTROL

Good laboratory practice is essential in order to obtain reliable results. This section highlights a few important guidelines and recommendations.

7.1 Data Quality Objectives

The correct peak assignments should be confirmed by model and reference compounds, but the detection limits and the linear response range should also be determined. The use of reference compounds in CZE separations is more critical than in many other separation systems to ensure the correct peak interpretation. In addition, the purity of peaks in the actual sample should be monitored, for example by UV or mass spectroscopy spectra; small changes in the separation conditions may give hints in many cases if there exist possible separation difficulties. It is also a good practice to collect temperature, voltage, current, and pressure profiles during runs to ensure a proper repeatability of the results. Changes in these parameters normally indicates possible difficulties, with

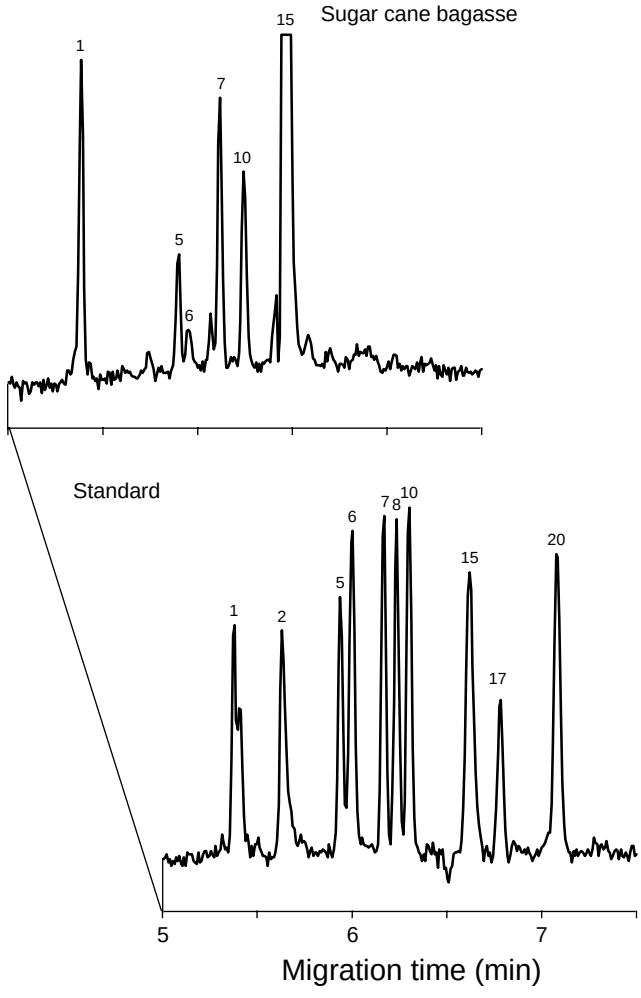

Figure 3 Co-electroosmotic separation of the 4-ABN derivatives of carbohydrates from a sugar cane bagasse hydrolyzate (TFA).[56] Peak assignment: galacturonic acid (1), glucuronic acid (2), galactose (5), fucose (6), arabinose (7), mannose (8), glucose (10), xylose (15), cellobiose (17), and rhamnose (20). (Reproduced from Nguyen et al.[56] by permission of Friedr. Vieweg & Sohn Verlagsgesellschaft.)

the separation facilitating the rejection of nonvalid results.

7.2 Laboratory Operations

Chemically pure reagents and well-prepared, filtered, and properly stored electrolyte or buffer solutions are important for a good separation. In addition, degassing of the buffer is recommended to avoid the formation of bubbles in the capillary and especially in the z-type detection cells. The solution level of the buffer and sample vials should be equal and they must be filtered. Otherwise, the reproducibility is affected by siphoning or the presence undissolved particles.

7.3 Instrument Performance

Some CE devices are equipped with an option that automatically replaces and adjusts the equal solution level in the buffer resevoirs after each run. This replenishment option has been shown to improve the quality of the separation and reproducibility.[72] Unfortunately, there is a high risk that the strong borate buffer will block this system due to crystallization. Therefore, rinsing with water is recommended after each sequence.

A new fused-silica capillary is rinsed with dilute sodium hydroxide solution followed by water and buffer. The same procedure is also used to regenerate the capillary after each run. It is debatable whether this should be done with the running buffer, especially in the case of an acidic buffer. It is therefore good practice to test which method gives the best reproducibility.

Monitoring of the current is advisable because it indicates the condition of buffer, capillary, and electrodes. In particular, it is a good indicator of problems in the CZE systems, where highly concentrated borate buffer

or sodium hydroxide solutions can crystallize between capillary and electrode, forming a current leak. Changes in the current may also be due to the formation of bubbles or obstructions within the capillary, or to temperature and voltage fluctuations and, in some cases, adsorption of compounds on to the capillary wall. Sodium hydroxide also has a tendency to dissolve the fragile fused-silica capillary. However, many of these problems can be avoided if the capillary is replaced frequently enough, well-prepared solutions are used, and the CZE device is serviced weekly.

8 COMPARISON WITH OTHER SEPARATION METHODS

Conventional methods for determining carbohydrates utilize GC and HPLC. Although GC provides high sensitivity and separation efficiency for monosaccharides, it is not a good approach for determining oligo- or polysaccharides. In general, it has two major drawbacks: first, the presence of multiple peaks, due to the existing equilibrium between different ring forms of monosaccharides including the simultaneous formation of α- and β-anomers; and second, the low volatility of monosaccharides.[55] However, the volatility of monosaccharides can be increased by derivatization using direct silylation or acetylation after reduction to alditols. The problem of multiple peaks can also be avoided by the latter method. A flame ionization detector (FID) is suitable for the most purposes with a detection limit of about 5 ng.[6] However, an electron capture detector (ECD) with halogenated derivatives can be used if a more sensitive detection is needed. In this case, the sensitivity is 100–1000-fold compared to that of FID.

In HPLC or IC techniques as well in CZE systems the isomeric nature of monosaccharides does not present a major problem because mutarotation is so fast that the different forms are able to elute as a single, although somewhat wider, peak. The sensitivity of HPLC or IC is better than that of CZE but, in the former approaches, the limiting factor is a moderate separation efficiency.[72] However, third-generation latex-agglomerated anion exchangers have shown the ability to resolve wood-based monosaccharides without derivatization.[73] The limited stability of these anion exchangers to organic solvents is one drawback that prevents effective removal of organic contaminants. Therefore, the difficult matrices of wood-based carbohydrates may affect the performance of the anion exchange column without sample clean-up. The CZE systems have a high separation efficiency and are capable of simultaneously resolving mono- and oligosaccharides. In addition, the amount of mobile phase needed in CE analysis is much less than that needed in the IC or HPLC techniques. One major drawback of the CZE approach is the need to perform a sample clean-up and to derivatize the analytes, both of which are time-consuming and may affect the quantitative recovery of preparations.

ABBREVIATIONS AND ACRONYMS

ADCP	Amperometric Detector with Constant Potential
ANTS	8-Aminonaphthalene-1,3,6-trisulfonic acid
AQ	Anthraquinone
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CGE	Capillary Gel Electrophoresis
CZE	Capillary Zone Electrophoresis
DEA	Diethylamine
DP	Degree of Polymerization
ECD	Electron Capture Detector
ECF	Elemental Chlorine-free
Et-4-AB	Ethyl 4-Aminobenzoate
FID	Flame Ionization Detector
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HDB	Hexadimethrine Bromide
HMM	High-molecular-mass
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
LCC	Lignin–Carbohydrate Complex
LIF	Laser-induced Fluorescence
LMM	Low-molecular-mass
LPC	Lignin–Polysaccharide Complex
MEKC	Micellary Electrokinetic Chromatography
MSD	Mass-sensitive Detector
NSSC	Neutral Sulfite Semichemical
PAD	Pulse Amperometric Detector
PMP	1-Phenyl-3-methyl-5-pyrazolone
RI	Refractive Index
SPE	Solid Phase Extraction
TCF	Totally Chlorine-free
TEF	Totally Effluent-free
TFA	Trifluoroacetic Acid
TMP	Thermomechanical Pulp
UV	Ultraviolet
2-AP	2-Aminopyridine
4-ABA	4-Aminobenzoic Acid

4-ABN	4-Aminobenzonitrile
6-AQ	6-Aminoquinoline

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis • Monosaccharides and Sugar Alcohol Analysis

Environment: Water and Waste **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction • Pulp and Paper Matrices

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods

REFERENCES

1. E. Sjöström, *Wood Chemistry – Fundamentals and Applications*, Academic Press, San Diego, USA, 21–108 and 114–248, 1993.
2. T. Vuorinen, R. Alén, 'Carbohydrates', in *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*, eds. E. Sjöström, R. Alén, Springer, Berlin, Germany, 37–75, 1999.
3. S.C. Churms, 'Recent Progress in Carbohydrate Separation by High-performance Liquid Chromatography based on Size Exclusion', *J. Chromatogr. A*, **720**(1–2), 151–166 (1996).
4. S. Honda, 'High-performance Liquid Chromatography of Mono- and Oligosaccharides', *Anal. Biochem.*, **140**(1), 1–47 (1984).
5. S.C. Churms, 'Recent Developments in the Chromatographic Analysis of Carbohydrates', *J. Chromatogr.*, **500**(1), 555–583 (1990).
6. C.J. Biermann, 'Introduction to Analysis of Carbohydrates by Gas-Liquid Chromatography (GLC)', in *Analysis of Carbohydrates by GLC and MS*, eds. C.J. Biermann, G.D. McGinnis, CRC Press, Boca Raton, Florida, USA, 1–17, 1989.
7. S. Honda, 'Separation of Neutral Carbohydrates by Capillary Electrophoresis', *J. Chromatogr. A*, **720**(1–2), 337–371 (1996).
8. M. Stefansson, M. Novotny, 'Resolution of the Branched Forms of Oligosaccharides by High-performance Capillary Electrophoresis', *Carbohyd. Res.*, **258**(1–2), 1–9 (1994).
9. J. Liu, O. Shirota, M.V. Novotny, 'Sensitive, Laser-assisted Determination of Complex Oligosaccharide Mixtures Separated by Capillary Gel Electrophoresis at High Resolution', *Anal. Chem.*, **64**(8), 973–975 (1992).
10. A. Rydlund, O. Dahlman, 'Efficient Capillary Zone Electrophoretic Separation of Wood-derived Neutral and Acidic Mono- and Oligosaccharides', *J. Chromatogr. A*, **738**(1), 129–140 (1996).
11. M. Ristolainen, 'Characterization of Totally Chlorine-free Effluents from Kraft Pulp Bleaching. II. Analysis of Carbohydrate-derived Constituents after Acid Hydrolysis by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **832**(2), 203–209 (1999).
12. W. Nashabeh, Z. El Rassi, 'Capillary Zone Electrophoresis of Linear and Branched Oligosaccharides', *J. Chromatogr.*, **600**(2), 279–287 (1992).
13. C. Chiesa, C. Horváth, 'Capillary Zone Electrophoresis of Malto-oligosaccharides Derivatized with 8-Aminonaphthalene-1,3,6-trisulfonic acid', *J. Chromatogr.*, **645**(2), 337–352 (1993).
14. X. Xiong, S.F.Y. Li, 'Selection and Optimization of Background Electrolytes for Simultaneous Detection of Small Cations and Organic Acids by Capillary Electrophoresis with Indirect Photometry', *J. Chromatogr. A*, **822**(1), 125–136 (1998).
15. J. Chen, B.P. Preston, M.J. Zimmerman, 'Analysis of Organic Acids in Industrial Samples. Comparison of Capillary Electrophoresis and Ion Chromatography', *J. Chromatogr. A*, **781**(1–2), 205–213 (1997).
16. D. Volgger, A.J. Zemann, G.K. Bonn, M.J. Antal, Jr, 'High-speed Separation of Carboxylic Acids by Co-electroosmotic Capillary Electrophoresis with Direct and Indirect UV Detection', *J. Chromatogr.*, **758**(2), 263–276 (1997).
17. D.R. Salomon, 'Applications of Capillary Ion Electrophoresis in the Pulp and Paper Industry', *J. Chromatogr.*, **602**(1–2), 219–225 (1992).
18. M.T. Ackermans, F.M. Everaerts, J.L. Beckers, 'Quantitative Analysis in Capillary Zone Electrophoresis with Conductivity and Indirect UV Detection', *J. Chromatogr.*, **549**(1–2), 345–355 (1991).
19. M. Shirao, R. Furuta, S. Suzuki, H. Nakazawa, S. Fujita, T. Maruyama, 'Determination of Organic Acids in Urine

by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **680**(1), 247–251 (1994).
20. E. Dabek-Zlotorzynska, J.F. Dlouhy, 'Application of Capillary Electrophoresis in Atmospheric Aerosols Analysis: Determination of Inorganic and Organic Anions', *J. Chromatogr. A*, **671**(1–2), 389–395 (1994).
21. P.E. Jackson, 'Analysis of Oxalate in Bayer liquors: a Comparison of Ion Chromatography and Capillary Electrophoresis', *J. Chromatogr.*, **693**(1), 155–161 (1995).
22. N. Avdalovic, C.A. Pohl, R.D. Rocklin, J.R. Stillian, 'Determination of Cations and Anions by Capillary Electrophoresis Combined with Suppressed Conductivity Detection', *Anal. Chem.*, **65**(10), 1470–1475 (1993).
23. P. Kuban, B. Karlberg, 'On-Line Dialysis Coupled to a Capillary Electrophoresis System for Determination of Small Anions', *Anal. Chem.*, **69**(6), 1169–1173 (1997).
24. A. Röder, K. Bächmann, 'Simultaneous Determination of Organic and Inorganic Anions in the Sub-µmol/L Range in Rain Water by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **689**(2), 305–311 (1995).
25. D. Volgger, A. Zemann, G. Bonn, 'Determination of Phenols, Inorganic Anions, and Carboxylic Acids in Kraft Black Liquors by Capillary Electrophoresis', *J. High Resol. Chromatogr.*, **21**(1), 3–10 (1998).
26. S.M. Masselter, A.J. Zemann, G.K. Bonn, 'Determination of Inorganic Anions in Kraft Pulping Liquors by Capillary Electrophoresis', *J. High Resol. Chromatogr.*, **19**(3), 131–136 (1996).
27. D.R. Salomon, J. Romano, 'Applications of Capillary Ion Electrophoresis in the Pulp and Paper Industry', *J. Chromatogr.*, **602**(1), 219–225 (1992).
28. A. Hilmi, J.H.T. Luong, A.-L. Nguyen, 'Capillary Electrophoretic Separation of Chlorophenols using Amperometric Detection', *J. Chromatogr. A*, **761**(1–2), 259–268 (1997).
29. O. Maman, F. Marseille, B. Guillet, J.-R. Disnar, P. Morin, 'Separation of Phenolic Aldehydes, Ketones and Acids from Lignin Degradation by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **755**(1), 89–97 (1996).
30. S. Masselter, A. Zemann, O. Bobleter, 'Analysis of Lignin Degradation Products by Capillary Electrophoresis', *Chromatographia*, **40**(1–2), 51–57 (1995).
31. R. Alén, 'Structure and Chemical Composition of Wood', in *Papermaking Science and Technology – Forest Products Chemistry*, ed. P. Stenius, Fapet, Finland, Vol. 3, Chapter 1, 2000.
32. D. Fengel, G. Wegener, *Wood–Chemistry, Ultrastructure, Reactions*, Walter de Gruyter, Berlin, Germany, 26–239 and 268–560, 1989.
33. S. Saka, 'Chemical Composition and Distribution', in *Wood and Cellulosic Chemistry*, eds. D.N.-S. Hon, N. Shiraishi, Marcel Dekker, New York, USA, 59–88, 1991.
34. A.D. French, N.R. Bertoniere, O.A. Battista, J.A. Cuculo, D.G. Gray, 'Cellulose', in *Kirk-Othmer–Encyclopedia of Chemical Technology*, 4th edition, eds. J.I. Kroschwitz, M. Howe-Grant, John Wiley & Sons, New York, USA, 476–496, Vol. 5, 1993.
35. K. Okamura, 'Structure of Cellulose', in *Wood and Cellulosic Chemistry*, eds. D.N.-S. Hon, N. Shiraishi, Marcel Dekker, New York, USA, 89–112, 1991.
36. K. Shimizu, 'Chemistry of Hemicelluloses', in *Wood and Cellulosic Chemistry*, eds. D.N.-S. Hon, N. Shiraishi, Marcel Dekker, New York, USA, 177–214, 1991.
37. N.S. Thompson, 'Hemicellulose', in *Kirk-Othmer–Encyclopedia of Chemical Technology*, 4th edition, eds. J.I. Kroschwitz, M. Howe-Grant, John Wiley & Sons, New York, USA, 54–72, Vol. 13, 1995.
38. T.E. Timell, *Compression Wood in Gymnosperms*, Springer, Heidelberg, Germany, Vols. 1–3, 1986.
39. W. Glasser, S. Sarkanen (eds.), *Lignin–Properties and Materials*, ACS Symposium Series 397, American Chemical Society, Washington, DC, USA, 1989.
40. S.Y. Lin, S.E. Lebo, Jr, 'Lignin', in *Kirk-Othmer–Encyclopedia of Chemical Technology*, 4th edition, eds. J.I. Kroschwitz, M. Howe-Grant, John Wiley & Sons, New York, USA, 268–289, Vol. 15, 1995.
41. A. Sakakibara, 'Chemistry of Lignin', in *Wood and Cellulosic Chemistry*, eds. D.N.-S. Hon, N. Shiraishi, Marcel Dekker, New York, USA, 113–176, 1991.
42. W.E. Hillis (ed.), *Wood Extractives and Their Significance to the Pulp and Paper Industries*, Academic Press, New York, USA, 1962.
43. J.W. Rowe (ed.), *Natural Products of Woody Plants: Chemicals Extraneous to the Lignocellulosic Cell Wall*, Springer, Heidelberg, Germany, Vols. 1 and 2, 1989.
44. Y. Kai, 'Chemistry of Extractives', in *Wood and Cellulosic Chemistry*, eds. D.N.-S. Hon, N. Shiraishi, Marcel Dekker, New York, USA, 215–256, 1991.
45. R. Alén, 'Basic Chemistry of Wood Delignification', in *Papermaking Science and Technology – Forest Products Chemistry*, ed. P. Stenius, Fapet, Finland, Vol. 3, Chapter 2, 2000.
46. T.M. Grace, E.W. Malcolm (eds.), *Pulp and Paper Manufacture – Alkaline Pulping*, TAPPI & CPPA, Montreal, Canada, Vol. 5, 1989.
47. S.A. Rydholm, *Pulping Processes*, John Wiley & Sons, New York, USA, 1965.
48. G.A. Smook, *Handbook for Pulp & Paper Technologists*, 2nd edition, Academic Press, San Diego, USA, 114–248, 1993.
49. A. Mimms, M.J. Kocurek, J.A. Pyatte, E.E. Wright (eds.), 'Kraft Pulping–A Compilation of Notes', 2nd printing, TAPPI Press, Atlanta, USA, 1993.
50. K. Niemelä, R. Alén, 'Characterization of Pulping Liquors', in *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*, eds. E. Sjöström, R. Alén, Springer, Heidelberg, Germany, 193–226, 1999.

51. J. Gierer, 'Chemistry of Delignification, Part I: General Concept and Reactions During Pulping', *Wood Sci. Technol.*, **19**, 289–312 (1985).
52. O.V. Ingruber, M.J. Kocurek, A. Wong (eds.), *Pulp and Paper Manufacture–Sulfite Science & Technology*, TAPPI & CPPA, Montreal, Canada, Vol. 2, 1985.
53. C.W. Dence, D.W. Reeve (eds.), *Pulp Bleaching–Principles and Practice*, Tappi Press, Atlanta, USA, 1996.
54. O. Dahlman, R. Mörck, J. Knuutinen, 'Analysis of Bleach Liquors', in *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*, eds. E. Sjöström, R. Alén, Springer, Heidelberg, Germany, 233–266, 1999.
55. M. Ristolainen, *Analysis of the Organic Material Dissolved during Totally Chlorine-Free Bleaching*, PhD Thesis, University of Jyväskylä, Finland, 17–25, 1999.
56. D.T. Nguyen, H. Lerch, A. Zemann, G. Bonn, 'Separation of Derivatized Carbohydrates by Co-electroosmotic Capillary Electrophoresis', *Chromatographia*, **46**(3–4), 113–121 (1997).
57. S. Honda, S. Iwase, A. Makino, S. Fujiwara, 'Simultaneous Determination of Reducing Monosaccharides by Capillary Zone Electrophoresis as the Borate Complexes of *N*-2-Pyridylglycamines', *Anal. Biochem.*, **176**(1), 72–77 (1989).
58. A.E. Vorndran, E. Grill, C. Huber, P.J. Oefner, G.K. Bonn, 'Capillary Zone Electrophoresis of Aldoses, Ketoses and Uronic Acids Derivatized with Ethyl *p*-Aminobenzoate', *Chromatographia*, **34**(3–4), 109–114 (1992).
59. E. Grill, C. Huber, P. Oefner, A. Vorndran, G. Bonn, 'Capillary Zone Electrophoresis of *p*-Amonobenzoic Acid Derivatives of Aldoses, Ketoses and Uronic acids', *Electrophoresis*, **14**(10), 1004–1010 (1993).
60. C. Huber, E. Grill, P. Oefner, O. Bobleter, 'Capillary Electrophoretic Determination of the Component Monosaccharides in Hemicelluloses', *Fresenius J. Anal. Chem.*, **348**(12), 825–831 (1994).
61. S. Hoffstetter-Kuhn, A. Paulus, E. Gassmann, H.M. Widmer, 'Influence of Borate Complexation on the Electrophoretic Behavior of Carbohydrates in Capillary Electrophoresis', *Anal. Chem.*, **63**(15), 1541–1547 (1991).
62. Y. Liu, C. Shu, J.D. Lamb, 'High-performance Capillary Electrophoretic Separation of Carbohydrates with Indirect UV Detection using Diethylamine and Borate as Electrolyte Additives', *J. Cap. Elec.*, **4**(3), 97–103 (1997).
63. T.W. Garner, E.S. Yeung, 'Indirect Fluorescence Detection of Sugars Separated by Capillary Zone Electrophoresis with Visible Laser Excitation', *J. Chromatogr.*, **515**(1), 639–644 (1990).
64. A. Paulus, A. Klockow, 'Detection of Carbohydrates in Capillary Electrophoresis', *J. Chromatogr. A*, **720**(1–2), 353–376 (1996).
65. S. Honda, K. Yamamoto, S. Suzuki, M. Ueda, K. Kakehi, 'High-performance Capillary Zone Electrophoresis of Carbohydrates in the Presence of Alkaline Earth Metal Ions', *J. Chromatogr.*, **588**(1–2), 327–333 (1991).
66. L.A. Colón, R. Dadoo, R.N. Zare, 'Determination of Carbohydrates by Capillary Zone Electrophoresis with Amperometric Detection at a Copper Microelectrode', *Anal. Chem.*, **65**(4), 476–481 (1993).
67. H. Schwaiger, P.J. Oefner, C. Huber, E. Grill, G.K. Bonn, 'Capillary Zone Electrophoresis and Micellar Electrokinetic Chromatography of 4-Aminobenzonitrile Carbohydrate Derivatives', *Electrophoresis*, **15**(7), 941–952 (1994).
68. S.F.Y. Li, *Capillary Electrophoresis–Principles, Practice and Applications*, Elsevier, Amsterdam, The Netherlands, 55–375, 1993.
69. S. Hase, 'Precolumn Derivatization for Chromatographic and Electrophoretic Analyses of Carbohydrates', *J. Chromatogr. A*, **720**(1–2), 173–182 (1996).
70. J. Vindevogel, P. Sandra, 'Simultaneous pH and Ionic Strength Effects and Buffer Selection in Capillary Electrophoretic Techniques', *J. Chromatogr.*, **541**(1–2), 483–488 (1991).
71. S. Honda, A. Makino, S. Suzuki, K. Kakehi, 'Analysis of the Oligosaccharides in Ovalbumin by High-performance Capillary Electrophoresis', *Anal. Biochem.*, **191**(2), 228–234 (1990).
72. D.R. Baker, *Capillary Electrophoresis*, John Wiley & Sons, New York, USA, 9–227, 1995.
73. J. Weiss, *Ion Chromatography*, VCH, Weinheim, Germany, 43–55 and 438–442, 1995.

Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Denys F. Leclerc
Pulp and Paper Research Institute of Canada, Vancouver, Canada

1 Introduction **2**
1.1 Definition of Fourier Transform Infrared Spectroscopy and Related Acronyms 2
1.2 Summary of Advantages and Limitations 2
1.3 Definition and Use of the Mid- and Near-infrared Regions 3
1.4 Topical Overview of Article 3

2 History **3**
2.1 Traditional Pulp and Paper Applications of Infrared Spectroscopy 3
2.2 Modern Development of Fourier Transform Infrared Spectroscopy 3
2.3 Early Fourier Transform Infrared Spectroscopy Applications to Pulp and Paper Problems 4
2.4 Current Scope of Pulp and Paper Applications of Fourier Transform Infrared Spectroscopy 4

3 Sampling Methods and Calibration Techniques **4**
3.1 Sampling Methods 4
3.2 Calibration Methods 8

4 Laboratory Applications **11**
4.1 Lignocellulosic Material Characterization 11
4.2 Extractives Analysis 12
4.3 Contaminant Analysis 13
4.4 Characterization of Papermaking Additives 15

5 On-line Applications: Analysis of Process Liquors **17**
5.1 Kraft Pulping 17
5.2 Alkaline Sulfite Anthraquinone Methanol Pulping 18
5.3 Other Applications 19

6 Quality Control and Troubleshooting **19**
6.1 Instrument Performance Issues 19
6.2 Data Quality Issues 20

7 Development of Quantitative Methods **20**
7.1 Univariate Calibration 20
7.2 Multivariate Calibration 21

8 Comparison with Other Spectroscopic Methods **22**
8.1 Comparison of Technology 22
8.2 Business Issues Relating to Process Control Instrumentation 24

Acknowledgments **25**

Abbreviations and Acronyms **25**

Related Articles **25**

References **25**

Fourier transform infrared spectroscopy (FTIRS) is the combined use of infrared (IR) light, for illuminating a light-absorbing sample, with an interferometer, for producing an interference pattern from which an absorption spectrum characteristic of the sample is recovered through Fourier transformation. The majority of current IR pulp and paper applications involve the routine characterization of samples that contain variable amounts of water, a very strong IR absorber. Consequently, qualitative IR analysis of pulp and paper samples is best performed with a Fourier transform infrared (FTIR) spectrometer because of the larger optical throughput offered by an interferometer over a dispersive spectrometer. Also, an interferometer simultaneously records all IR frequencies present in the spectrum. These two advantages enable one to acquire high-quality spectra in a few minutes or less. Additionally, since an interferometer accurately determines IR frequencies with the output of a visible-light laser, quantitative determinations which use multivariate calibration methods such as partial least squares (PLS) are much less prone to instrument drift, and hence more accurate. Liquid and solid samples can be analyzed either neat or suitably diluted in a nonabsorbing matrix, through a variety of sampling techniques such as transmission, diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS), attenuated total reflectance (ATR), IR microscopy, and photo-acoustic spectrometry (PAS). Detection limits for quantitative methods vary from 0.01 to 1 percent, depending on the sampling method and the complexity of a sample matrix. However, FTIR spectrometry cannot resolve more than a few components at a time in liquid and solid samples because of the presence of strongly overlapping bands. Therefore, FTIR spectroscopy must often be used in combination with separation techniques such as gas chromatography (GC) or solid-phase extraction (SPE), especially when a complete analysis of the sample is required.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Griffiths and de Haseth[1] describe in detail the theory, practice and historical development of FTIRS. Early industrial applications were presented in a series of monographs edited by Ferraro and Basile.[2–4] Practical aspects of FTIRS are discussed in Ferraro and Krishnan.[5] Jones[6] gives an overview of the subject.

1.1 Definition of Fourier Transform Infrared Spectroscopy and Related Acronyms

FTIRS is a technique that measures the frequency dependence of the interaction of radiation with matter. FTIRS is the combined use of IR light, for illuminating a light-absorbing sample, with an interferometer, for producing an interference pattern from which an absorption spectrum characteristic of the sample is recovered through Fourier transformation. In FTIRS, the selection rules governing band assignments are the same as those found for dispersive IR spectroscopy. However, the greater energy throughput of a FTIR spectrometer, as well as the multiple-scan capability and higher resolution which can be achieved with computing technology, have made possible the use of IR spectroscopy in much lower energy sampling situations than was hitherto possible with a dispersive instrument. The majority of IR spectrometers now use a Michelson interferometer (Figure 1) to produce the interference pattern. Referring to Figure 1, a collimated cylindrical beam emanating from an IR light source is split in half by a beamsplitter. One half of the beam is reflected from a fixed mirror, and the other half from a moving mirror. Both halves recombine and give rise to the interference pattern. Unlike a dispersive instrument where the light interacting with the sample has a narrowly defined frequency, the interference pattern contains light having a wide range of frequencies covering either the near-infrared (NIR) or the mid-infrared (Mid-IR) region of the electromagnetic spectrum. Since the sample absorbs only at certain frequencies, the shape of the interference pattern, or interferogram, changes upon interacting with the sample, thereby producing a spectrum characteristic of the sample. The signal is transformed into an electrical signal by either a deuterated triglycine sulfate (DTGS) or a mercury cadmium telluride (MCT) [29870-72-2] detector. An array processor then manipulates the data. The spectrum is then obtained by performing a fast Fourier transform (FFT) with the Cooley–Tukey algorithm.[1] A computer controls the interferometer, accumulates, stores and manipulates interferogram data, performs the Fourier transformation, as well as post-spectral manipulation such as plotting, calibration and correlation. Manipulation of interferogram data consists of co-adding successive scans, apodization and phase correction. The first operation improves the signal-to-noise ratio of a spectrum, whereas the latter two remove instrumental artifacts.

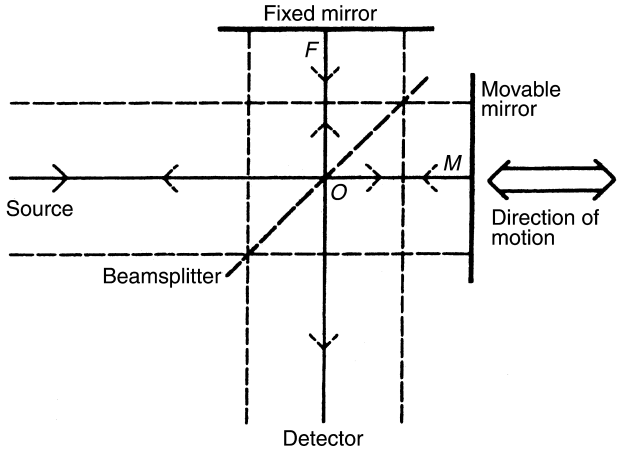

Figure 1 Schematic diagram of a FTIR spectrometer based on the Michelson interferometer. The median ray is shown as a solid line, and the extremes of the collimated beam are shown as broken lines. (Reproduced by permission of John Wiley & Sons from P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, Copyright © 1986, John Wiley & Sons, New York.)

1.2 Summary of Advantages and Limitations

IR analysis of pulp and paper samples is best performed with a FTIR spectrometer because of the two major advantages offered by an interferometer over a dispersive spectrometer. These advantages are quite distinct from each other, and are discussed in detail elsewhere.[1–4,6] The first advantage is called the throughput or Jacquinot's advantage, and results in an optical throughput which is about 200 times larger in an interferometer than that found in a dispersive spectrometer. Thus, a much larger amount of radiation is admitted thought the aperture of an interferometer than through the narrow slit of a dispersive-instrument monochromator.[6]

The throughput advantage is reinforced by the interferometer's ability to use the entirety of the light signal by simultaneously observing all available frequencies in the spectral region being observed, thereby improving its signal-to-noise ratio by a factor proportional to the square root of the number of available frequencies. The light signal is then resolved into individual frequencies by using the Fourier transformation. This second advantage is called the multiplex or Fellgett's advantage, and enables one to acquire a high-quality spectrum in a few minutes or less by co-adding a large number of scans. In the Mid-IR, the multiplex advantage varies from a factor of about 16 for a resolution of 16 cm^{-1} to a value of 64 for 1 cm^{-1}.[1,2] The combination of the multiplex and throughput advantages produces a large linear dynamic range, thereby making FTIRS very suitable for quantitative work.[1]

Finally, since an interferometer accurately determines IR frequencies with the output of a visible-light laser, quantitative determinations which use multivariate calibration methods such as PLS are much less prone to instrument drift, and hence more accurate. The increased accuracy is referred to as the laser-reference or Connes' advantage, and is important when one wants to avoid problems of peak broadening and/or peak shifting during signal averaging. Subtle spectral features are therefore preserved and the level of spectral noise is lower, thereby improving the quality of calibrations. Although this little-known advantage is not as important as the first two, the frequency accuracy which is routinely achievable with FTIR spectrometers is much higher in practice than that available with dispersive instrumentation, especially when using short-wavelength radiation in the NIR region of the spectrum.

The main practical advantage of FTIRS is that liquid and solid samples can be analyzed with little preparation through a variety of sampling techniques such as transmission, DRIFTS, ATR, IR microscopy, and PAS. ATR is also known as multiple internal reflectance (MIR). Detection limits for FTIR quantitative methods vary from 0.01 to 1 percent, depending on the sampling method and the complexity of a sample matrix. The main disadvantage of FTIRS is that it cannot resolve more than a few components at a time in liquid and solid samples because of the presence of strongly overlapping bands. Therefore, FTIRS must often be used in combination with separation techniques such as GC or SPE, especially when a complete analysis of the sample is required.

1.3 Definition and Use of the Mid- and Near-infrared Regions

The frequency range use for routine pulp and paper applications normally extends from 400 to 14 000 cm^{-1}, and covers both the Mid-IR and NIR ranges. The presence of very strongly absorbing fundamental absorption bands in the Mid-IR region (400 to 4000 cm^{-1}) dictates the use of sampling techniques with a very small effective path length which minimize the extent of these absorption bands, such as ATR and DRIFTS. These sampling techniques are particularly effective when water is present in a sample. On the other hand, in the NIR region (4000 to 14 000 cm^{-1}) the majority of studies have been performed with cheaper dispersive instrumentation, which have sufficient throughput for most routine applications.

1.4 Topical Overview of Article

A brief historical survey of the development of early dispersive IR spectroscopy applications in the pulp and paper industry will first be presented, followed by an overview of the most useful FTIR sampling methods and calibration techniques. Representative laboratory and on-line applications will be demonstrated. Data quality problems frequently encountered with pulp and paper samples will then be diagnosed, and remedies will be suggested. Next, selected examples relating to the development of quantitative methods will be given. Finally, FTIRS will be compared to other spectroscopic techniques, such as Raman and NIR spectroscopy.

2 HISTORY

2.1 Traditional Pulp and Paper Applications of Infrared Spectroscopy

Chemists have been using IR spectroscopy for nearly fifty years to characterize lignocellulosic samples. Prior to the 1980s, these studies were performed with dispersive instrumentation, which used either a grating or a prism. Data acquisition was slow and cumbersome, and relied on analog technology to record data. Nevertheless, the scope of solid-phase applications that were then pioneered is just as broad as that of current research.

For example, lignin [9005-53-2] variations between different softwood species were characterized.[7] The groundwood content of paper was measured by calculating the ratio of the lignin peak to that of cellulose [9004-34-6].[8] Similarly, the kappa number of pulps was determined by using MIR and a cellulose peak as an internal standard.[9] The photo-degradation of cellulose was reported to be a surface effect[10] and to be correlated to an increase in the amount of carbonyl or carboxyl groups, as evidenced by the appearance of a band at 1740 cm^{-1}. Lipids in beech bark were investigated.[11] The total hydroxyl content of cellulose esters was determined by NIR.[12] Kraft pulping process gases were analyzed by IR spectroscopy.[13] It was concluded that, because of the cumbersome sample preparation and transport problems then associated with IR analysis, coulometric titration or GC was better suited for mill use.

2.2 Modern Development of Fourier Transform Infrared Spectroscopy

Griffiths and de Haseth have briefly discussed the historical development of FTIRS.[1] IR spectroscopy underwent a period of partial eclipse during the 1960s and early 1970s because of the rapid growth of other structural determination techniques such as nuclear magnetic resonance (NMR) and mass spectrometry (MS), as well as the development of separation techniques such as gas and liquid chromatography. MS and gas and liquid chromatography thus replaced IR spectroscopy as the

method of choice in performing quantitative analysis. Also, the lack of versatility of sampling methods other than transmission, tedious preparation protocols and the need to scrupulously dry samples precluded the use of IR spectroscopy in many industrial applications.

Even though physicists developed interferometers at the turn of the century, these instruments were of little use to chemists until gas lasers and minicomputers became available in the late 1960s. The minicomputer allowed spectral computation to be done in real time. The helium–neon (He–Ne) laser made possible the monitoring of the travel of the moving mirror, and permitted the interferogram to be digitized at precisely equal intervals, thereby creating an internal wavelength standard.[1] In the late 1970s, the rapid commercialization of FTIR spectrometers spurred the development of new IR sampling techniques such as DRIFTS, ATR, FTIR microscopy and PAS, all of which took advantage of the greater throughput available with interferometers. Computational software packages, which calculated the FFT immediately after acquiring the data, also became available. Additionally, these packages performed post-spectral processing of the mass of digital data now available. FTIR instrumentation thus became more user-friendly, and hence more attractive for pulp and paper scientists. In the mid-eighties, FTIR spectrometers began to be used in the pulp and paper industry. By the mid-nineties, a wide range of applications had been developed, most notably in the fields of contaminant analysis and pulp manufacture process control. FTIRS is now used for routine analysis in most pulp and paper laboratories around the world and, increasingly, for process control applications in mill environments.

2.3 Early Fourier Transform Infrared Spectroscopy Applications to Pulp and Paper Problems

In the late 1980s, the first pulp and paper applications of FTIRS appeared in the literature. For example, a method was developed with the use of DRIFTS[14] for estimating lignin in unbleached pulps. The method is based on a linear correlation between the area of the 1510 cm^{-1} IR band and kappa number. The relationship holds for a range of hardwood and softwood pulps having kappa numbers ranging from 10 to 120. In another application, an ATR method for controlling the cooking degree of pulp during pulp digestion was described.[15] The method analyzes the composition of the cooking liquors before and during the cook, and uses correlations of either peak heights or peak integrals between 1400 and 1550 cm^{-1} with kappa numbers. Finally, pulp and paper samples were examined by IR microscopy.[16] The utility of the technique for identifying deposits, laminates and surface coatings was demonstrated.

2.4 Current Scope of Pulp and Paper Applications of Fourier Transform Infrared Spectroscopy

FTIRS is now used in various ways in the pulp and paper industry, especially for laboratory bulk analysis of condensed-phase samples that are difficult to analyze by any other method. Important applications include the characterization of lignocellulosic materials such as pulp and lignin samples, papermaking additives, extractive and contaminant analysis, and process control. The technique is particularly useful for the in situ analysis of contaminants that are difficult to extract from their matrix. FTIRS is also best applied whenever high sensitivity and/or speed is needed, such as in the analysis of pulping liquors, demanding quantitative determinations, as well as process and quality control applications.

3 SAMPLING METHODS AND CALIBRATION TECHNIQUES

The purpose of this section is to briefly review the most important sampling methods and calibration techniques that are currently available to a person interested in using IR spectroscopy to pursue the determination and/or characterization of pulp and paper-related samples. This review is not meant to be exhaustive: the relative emphasis given to each sampling method reflects the degree of use of that method in the pulp and paper industry. For example, gas analysis, which has seen little activity in recent years, will not be discussed. The reader is referred for more details to the reviews on sampling methods by Krishnan and Ferraro,[17] Friese and Banerjee,[18] Faix,[19] and to the monograph by Griffiths and de Haseth.[1] Haaland[20] has extensively reviewed univariate and multivariate calibration methods.

3.1 Sampling Methods

A variety of methods is available for sample determination by IR spectroscopy. The nature of the sample, the goals of the analysis and the time available dictate the choice. The vast majority of IR pulp and paper applications used to be done with transmission cells. Elimination of all traces of water is critical in obtaining good-quality spectra. In the transmission method, Beer's law (Equation 1) expresses the intensity of IR radiation traversing a sample:

$$I = I_o \times 10^{-acl} \tag{1}$$

where I_o is the intensity of the source, I is the intensity of the transmitted radiation, a is the absorption coefficient, c is the concentration of the absorbing species and l is the optical path length of the radiation in the sample. Spectra

are customarily presented either as absorbance, A, or transmittance, T, which, by rearranging Equation (1), are defined by Equation (2):

$$T = \frac{I}{I_o} = 10^{-acl} A = -\log_{10} T = acl \qquad (2)$$

Thus the absorbance is linearly proportional to the concentration of the absorbing component. The linear relationship between A and c expressed by Equation (2) is the conceptual basis for the calibration methods used in IR spectrometry, such as PLS. Other than the well-known transmission method, the most widely used IR sampling methods for pulp and paper applications are DRIFTS, ATR, IR microscopy and, to a much lesser extent, PAS. These methods will now be briefly reviewed. The advantages and drawbacks of each method will also be discussed in a pulp and paper industry context.

3.1.1 Diffuse Reflectance Infrared Fourier Transform Spectroscopy

The technique of diffuse reflectance predates the advent of FTIR spectrometers, but did not gain wide acceptance until an accessory with efficient collection optics was designed in the late 1970s.[21] Referring to Figure 2, the light from the source is collected by a flat-mirror system (3), reflected onto a large hemiellipsoidal mirror (2), which then focuses it on the sample (1). The sample usually consists of a small amount of powdered sample which has been diluted in a highly reflective, nonabsorbing matrix such as powdered potassium chloride [7447-40-7]

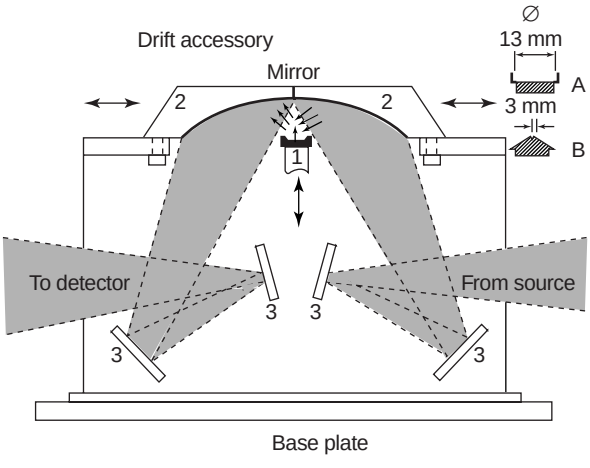

Figure 2 Diffuse reflectance accessory (Spectra-Tech Inc.). 1 Movable stage with device for removing specular reflectance; 2 Hemiellipsoidal mirrors; 3 Mirror system; A Macro sample holder; B Micro sample holder. (Reproduced by permission of Springer-Verlag from O. Faix, in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 83–109 © 1992.)

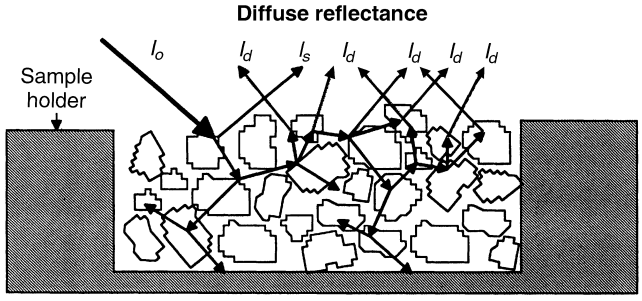

Figure 3 A diagram of the path taken by the radiation during DRIFTS sampling. (Reproduced by permission from M.A. Friese, S. Banerjee, in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, 119–141 Copyright CRC Press, Boca Raton, Florida © 1995.)

(KCl). As shown in Figure 3, IR light having an intensity I_o is partly reflected (I_s) through specular reflectance from the sample surface and partly transmitted to a depth of a few micrometers, where it is either completely absorbed by the sample or scattered (I_d) back to the surface. This process is repeated a few times until some of the light is re-emitted (I_d) at the surface in all directions. The re-emitted light is thus diffuse, and contains spectral features characteristic of the sample. This light is then collected by a second hemiellipsoidal mirror and sent to the detector by a second series of flat mirrors similar to the first.

The reflectance data can be converted to a transmission-like spectrum by using the Kubelka–Munk model.[17] This model relates the reflectance to the absorption coefficient, a, found in Beer's law according to Equation (3):

$$\frac{2.303\,ac}{s} = \frac{(1 - R_\infty)^2}{2R_\infty} \qquad (3)$$

where c is the concentration of the absorbing species, s is the scattering coefficient of the sample, and R_∞ is the reflectance of an infinitely thick sample, as determined relative to a nonabsorbing standard. Since dilution renders the scattering coefficient, s, nearly independent of sample absorption, the right-hand side of Equation (3) is now proportional to ac, and the spectrum of a diluted sample will resemble that of a transmission spectrum.

The main advantage of using DRIFTS is that minimal sample preparation is needed. DRIFTS is an easy means of obtaining the IR spectra of solids such as lignin and cellulose which are difficult to study with the transmittance method. It is very difficult to grind such materials into fine, uniform powders, but DRIFTS can be performed with coarsely ground materials. For example, samples as small as 0.25 mm can be analyzed as received. With larger samples, one must scrape a very small amount of sample off the surface of the object to be analyzed, grind one part of sample into a powder with 20 to 50 parts of KCl, and put the mixture into a sample cup. Even liquid extracts of

either dissolved of suspended solids can be analyzed with the use of a solvent elimination technique:[22] one simply presses freshly ground KCl into the sample cup, put a drop of the extract onto the KCl, and let the volatile solvent evaporates. Transmission-like spectra are obtained. All components present in sufficient quantities will be detected: strong IR absorbers such as esters and acids can be detected in sub-microgram quantities. The DRIFTS method is particularly useful when it is used in combination with a user-developed spectral library customized for the types of samples and contaminants commonly found at a particular mill. One can identify unknown samples by searching the spectral library.

Specular reflectance is produced by mirror-like surfaces, and distorts DRIFTS spectra. Even though DRIFTS accessories are now designed to remove most of the specular reflectance, the presence of inverted peaks produced by residual specular reflectance is the main disadvantage of this technique. Specular reflectance can be partially removed by increasing the dilution of the sample. However, if the major component of the sample is affected strongly by specular reflectance, further dilution of the sample will wash out the spectral features of the minor components, which defeats the purpose of the analysis. DRIFTS is mainly a qualitative tool and, even though it is possible to develop quantitative methods with multivariate calibration, is less amenable to quantitative determination than the transmission method. Accessories are available from a wide variety of suppliers.

3.1.2 Attenuated Total Reflectance

ATR, which was introduced in the 1960s, successfully addressed the problem posed by highly absorbing samples.[23] ATR is a technique for analyzing films, powders, pastes and liquids with the aid of an internal reflection accessory, and is ideal for studying the surface composition of films or coatings, and for determining the concentration of dissolved components in aqueous solutions. As an alternative to the transmission method, this technique is especially useful when components such as water with high extinction coefficients absorb so strongly that their absorbance is no longer linear. Referring to Figure 4, IR radiation from the source is made to propagate at an angle through a reflecting rod. This rod is made of an IR-transparent material with a high index of refraction, η_1, such as zinc selenide [1315-09-9], zinc sulfide [1314-98-3] or germanium [7440-56-4], which is surrounded by a solid or liquid sample with a lower index of refraction η_2. When this radiation strikes the internal surface of the high-index material, the incident beam splits into a refracted part and a reflected part. Let θ be the angle of incidence as defined from a line perpendicular to the internal surface of the reflecting material. If the angle of incidence is greater than a critical

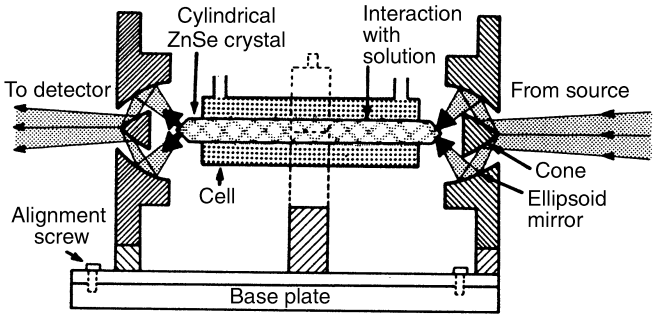

Figure 4 Liquid sampling device (The CIRCLE®, Spectra-Tech Inc.) according to the principle of ATR. (Reproduced by permission of Springer-Verlag from O. Faix, in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 233–241 © 1992.)

value θ_c, no refraction occurs and the beam is totally reflected back into the high-index material. The reflected ray will have nevertheless propagated for a very short distance called the penetration depth, d_p, into the sample and returned to the reflecting material, after having been slightly absorbed by the surrounding sample. The reflected ray is thus attenuated, and provides structural and quantitative information about the chemical components present in the sample. This process is repeated a number of times, N_r, along the reflecting medium before the radiation reaches the detector. The effective path length, l, is thus N_r times the penetration depth, d_p, and is given by Equation (4):[23]

$$l = N_r d_p = \frac{N_r \lambda}{2\pi(\eta_1^2 \sin^2\theta - \eta_2^2)^{0.5}} \qquad (4)$$

where λ is the wavelength of the IR radiation in air. Since the penetration depth usually falls between 0.05 λ and 0.2 λ, the effective path length l can vary from 0.25 to 4 μm, depending on the wavelength and reflecting material.[17] Thus, l is very small in comparison with other FTIR sampling techniques, even though it increases ten times from 4000 to 400 cm^{-1}. For archival purposes, an ATR spectrum can be converted to a transmission-like spectrum by using Equation (3). The converted spectrum will then exhibit undistorted features.

The main advantage of this sampling technique is the ability to study bulk samples that contain strongly absorbing species such as water. The absorbance of dissolved species can therefore be measured against that of water without saturating the detector. Another advantage is the capacity to perform surface analysis. Finally, this technique requires minimal sample preparation.

A major drawback of this sampling technique is that the surface of many reflecting rods is fragile and easily attacked by harsh chemicals. Many reflecting-rod surfaces exhibit a tendency to foul or etch over time, even if

one tries to minimize long-term exposure to process liquors. Calcium scale deposits are especially hard to clean, and particular care must be taken by the analyst to avoid damaging the surface: a mildly acidic solution is recommended. Consequently, reflecting rods must be re-polished periodically; as a result, there is a slight increase in total effective path length after each re-polishing operation. For laboratory applications, this represents a small problem: calibrations used in the laboratory can be adapted so as to account for this gradual change. In process applications, however, the choice of optical-element material must be carefully considered so that the sampling system can be designed so as to minimize lengthy exposure of the material to the process environment, thereby maximizing its lifetime. A slip-stream sampling configuration is an excellent way of achieving this objective. Reflecting rods and accessories are available from a wide range of vendors in either rectangular cross-section configurations for laboratory liquid and solid sampling, or in cylindrical cross-section configurations for process monitoring of liquid streams.

3.1.3 Fourier Transform Infrared Microscopy

When the sample to be analyzed is smaller than 0.25 mm, the use of a FTIR microscope is required. A schematic of a FTIR microscope is shown in Figure 5. This specially designed apparatus must focus the IR beam from the spectrometer onto the sample, collect the IR radiation that has interacted with the sample, deliver it to a detector, and allow the user to view and define the area to be analyzed. The microscope thus must have identical geometric configurations for the IR and visible beams. This requirement is accomplished through the use of Cassegrainian optics. Visual inspection of the sample is done with the microscope, a spot is selected, and the IR beam is then focused on the sample by a Cassegrain mirror condenser. The transmitted/reflected IR radiation is then collected with a Cassegrain mirror objective and sent to a liquid nitrogen-cooled MCT detector. The spatial resolution is approximately between 4 and 30 μm in the Mid-IR range. The practical range of sample sizes is from 20 to 30 μm. FTIR microspectroscopy and its application to paper samples are described in detail by Sommer et al.[24] Practical aspects of FTIR microscopy are also discussed in Humecki.[25]

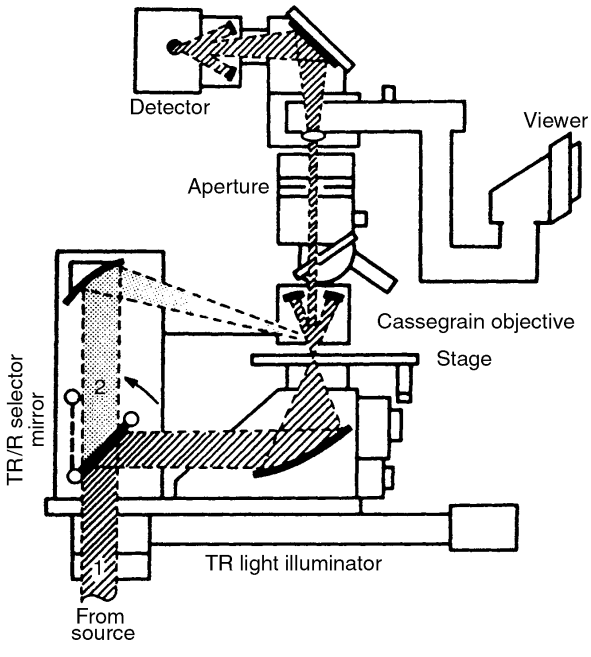

Figure 5 FTIR microscope UMA 300A (Bio-Rad Inc., Digilab Division) operating in either the transmission (1) or in the reflectance (2) mode. (Reproduced by permission of Springer-Verlag from O. Faix, in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 83–109 © 1992.)

The main advantage of this technique is its ability to analyze very small samples, down to one pulp fiber diameter. However, unlike visible microscopy, this technique cannot resolve lignocellulosic structures found within the fiber wall, which are of the same size as the incident radiation and below the diffraction limit (10 μm). Another drawback is the need to maintain the detector at liquid nitrogen temperatures for greater sensitivity, if it is not being thermo-electrically cooled. Finally, the cost of a FTIR microscope is roughly the same as that of a stand-alone FTIR spectrometer, thereby doubling the effective cost of FTIR microscopy, compared to other sampling methods.

3.1.4 Photo-acoustic Spectrometry

Fourier transform PAS is a method whereby the sample is irradiated in a closed cell with light whose intensity is modulated by a rapid-scanning Fourier transform spectrometer. The modulation frequency is proportional to the IR frequency, and goes from 125 to 1250 Hz for a DTGS-equipped FTIR spectrometer with a mirror velocity of 0.16 cm s^{-1}. The intensity of the light penetrating into the sample falls off exponentially with depth, with a decay constant of $1/(ac)$, where ac is the absorptivity coefficient from Beer's law. As shown in Figure 6, the absorption of radiation by the sample causes the temperature of the sample to rise. The modulated heat diffuses to the surface and is then transferred from the sample surface to an IR-transparent gas with good thermal properties such as helium [7440-59-7]. This produces a modulated heating and pressurizing of the gas inside the cell, which is then picked up as sound waves by a microphone. The strength of the signal is proportional to the value of ac and inversely proportional to the square

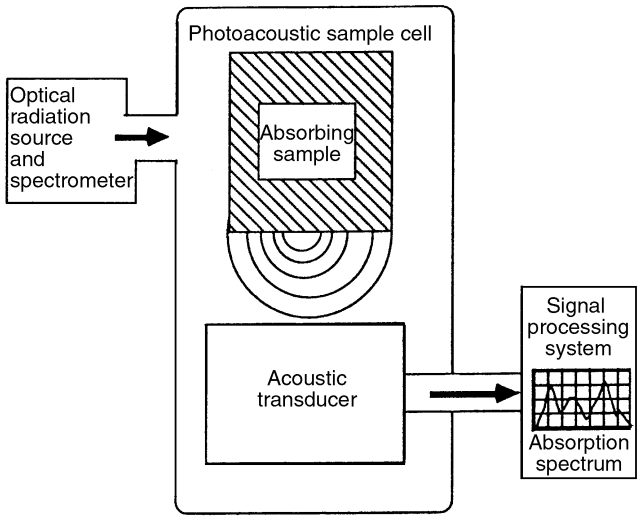

Figure 6 Schematic diagram of a photo-acoustic spectrometer. (Reproduced by permission from J.F. McClelland, *Anal. Chem.*, **55**(1), 89A–105A © 1983 American Chemical Society.)

root of the modulation frequency. The PAS signal is thus about three times weaker at 4000 cm^{-1} than that at 400 cm^{-1}. Even so, the resulting spectrum is very similar to an absorbance spectrum. Penetration depths vary from a few micrometers to a few tens of micrometers, and can be adjusted by changing the modulation frequency. Since there is no wavelength range, one detector can be used from the ultraviolet (UV) to the far IR.

The main advantage of PAS is that one can obtain spectra on any type of opaque solid or semi-solid material. In addition, the PAS signal is produced only at those frequencies where absorption by the sample occurs, with little or no interference from scattered radiation. It is especially useful for minerals that exhibit strong specular reflectance such as talc [14807-96-6], alum [7784-24-9] and calcium carbonate [471-34-1]. PAS is thus another technique that has benefited from the advent of FTIR spectrometers.

Unfortunately, after an initial surge in interest regarding the study of wood samples,[26] this technique has been recently eclipsed by DRIFTS, mainly because one can now acquire DRIFTS spectra of wood samples which are virtually free of specular reflectance with the accessories now available on the market. Also, the absolute PAS signal cannot be used directly for quantitative analysis: one must use internal standards and peak ratios. Finally, commercial PAS accessories are not as widely available as they were a few years ago.

Table 1 summarizes the advantages and drawbacks of each sampling method, the kind of samples that can be analyzed with the method, along with an estimate of the concentration range over which the method gives optimal results.

3.2 Calibration Methods

In the recent past, one had to isolate a spectral band in order to perform IR quantitative analysis. Nowadays, the ability to use the entire spectral range has greatly extended the range of samples that can now be quantitatively analyzed by IR spectroscopy. Indeed, the IR spectrum of samples can also be used for the determination of the chemical and physical properties of individual chemical components in the sample.[20] An illustration of the power of multivariate calibrations over univariate determinations is given in Figure 7. Figure 7(a) shows how an impurity may affect the spectrum of the species being determined. In Figure 7(b) we see that, in the presence of that impurity, the real absorbance reading, A_t, at a frequency ν_1 is overstated by the error $(A_m - A_t)$, but nevertheless falls on the calibration curve. Thus, it is impossible to tell whether c_m or c_t constitutes the real value. On the other hand, if a second frequency ν_2 is used, the impurity will be detected Figure 7(c) as an outlier to be treated with suspicion.

The main advantages of multivariate methods over univariate methods are that:

Table 1 Summary of FTIR sampling methods

Method	Matrices	Advantages	Drawbacks
Transmission (>10 ppm)	Liquids, thin sheets, KBr disks	Bulk analysis, good spectral quality	Tedious preparation with solid samples
DRIFTS (0.1–100%)	Ground KCl, Wood, pulp, paper	Near-surface analysis, minimal preparation, solvent elimination	Specular reflectance with talc, clay, alum, wood, pulp, paper; specks >0.25 mm
ATR (0.05–100%)	Pulp, paper, liquors	Solids: surface; liquids: bulk	Film formation on IRE; chemical attack
FTIR microscopy (1–100%)	Paper, thin sheets, laminates	Can analyze very small contaminants	Cost, maintenance, limited resolution
PAS (0.1–100%)	Solids, pastes, oils	Bulk analysis; no specular reflectance	Signal depends on modulation frequency

IRE, internal reflection element.

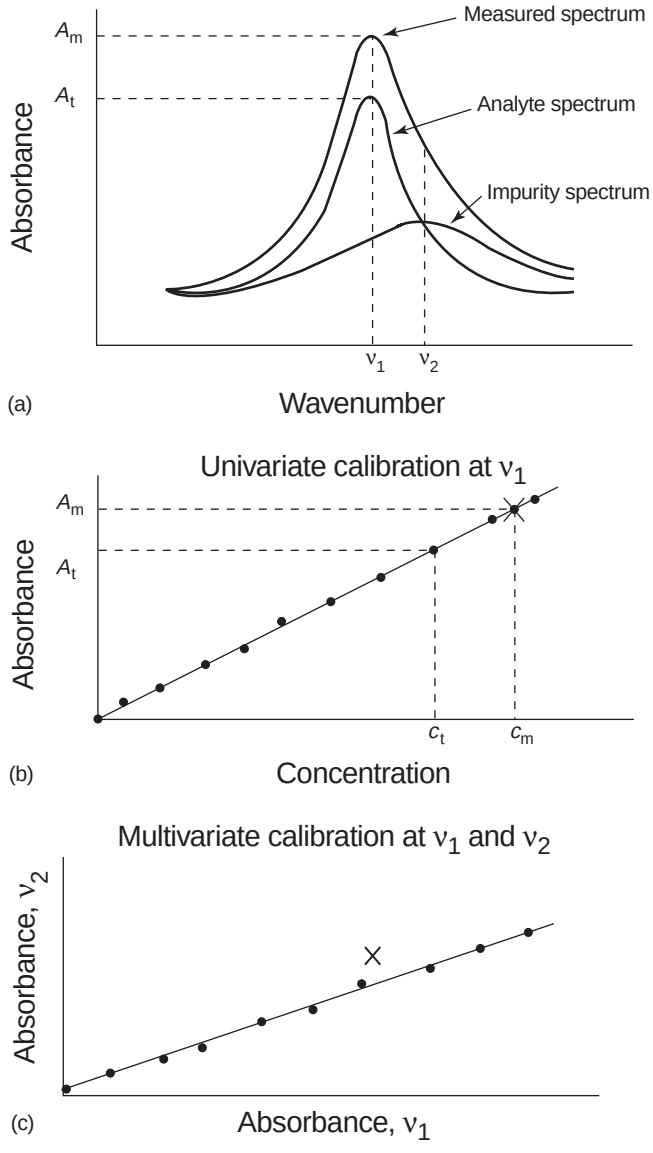

Figure 7 Advantage of multivariate calibrations over univariate calibrations. The 'X' in (b) and (c) indicates data obtained from a sample containing both the analyte and the impurity. (Reproduced by permission of Academic Press and of the author from D.M. Haaland, in *Practical Fourier Transform Infrared Spectroscopy: Industrial and Laboratory Chemical Analysis*, eds. J.R. Ferraro, K. Krishnan, Academic Press, San Diego, CA, 395–468 © 1990.)

1. redundant spectral data provide increased precision;
2. baseline variations can be accounted for; and
3. outliers are readily detected.

In this section, we give a brief description of two different, but related, multivariate calibration methods: classical least squares (CLS) and PLS. We also provide a summary of their respective advantages and drawbacks. For a description of calibration methods such as inverse least squares (ILS), principal component analysis (PCA) and principal component regression (PCR), the reader is referred to the review written by Haaland[20] and to a detailed discussion by Haaland and Thomas.[27]

3.2.1 Classical Least Squares Calibration

Since the basis of this method is the linear form of Beer's Law as given by Equation (2), this method is easier to grasp than the other multivariate methods, the majority of which are variants of CLS utilizing different models with various assumptions. The general calibration procedure used in CLS is common to all multivariate calibration methods, and consists of two main steps: a calibration step and a prediction step. First, one acquires i absorbance readings from the IR spectra of a series of n sample mixtures of all m components known to be present in the unknown samples. That series must span the range of variation of all the factors affecting the spectra of unknown samples. The calibration model is then calculated with the use of matrix algebra, which enables one to solve for each of the i spectral frequencies n simultaneous linear equations in m unknowns. From Equation (2), one sets $k = al$ for a constant path length and obtains Equation (5), for a series of n mixtures, assuming that the absorbance values are additive:

$$A_{n,i} = \sum_{j=1}^{m} c_{n,j} k_{j,i} + e_{n,i} \tag{5}$$

where $e_{n,i}$ are the residual errors between the least squares fit line and the actual absorbance values. Since these are calibration samples, the concentrations are known and the absorbance values have been measured. Therefore, one obtains m plots of $k_{m,i}$ for all i data points, thereby reconstructing the pure-component spectrum for each of the m components, as shown in the right-hand side of Figure 8. These pure-component spectra can then be used in the prediction phase of the analysis for determining the concentrations of the m components in the unknown sample. A least-squares procedure calculates the best linear combination of pure-component spectra that minimizes the sum of squared differences calculated from the $e_{n,i}$ values found between the actual and calculated spectra.[20] A better way of solving simultaneous equations is by using linear algebra. Therefore, rewriting Equation (5) in matrix form and expressing the spectra presented in the absorbance matrix as rows rather than columns, we obtain Equation (6):

$$\mathbf{A} = \mathbf{CK} + \mathbf{E_a} \tag{6}$$

where $\mathbf{A}$ is the $n \times i$ matrix containing the absorbance values of the n samples at i frequencies, $\mathbf{C}$ is the $n \times m$ matrix of the m component concentrations in

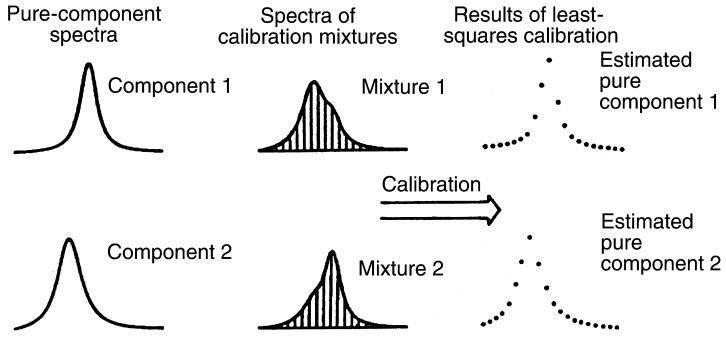

Figure 8 Details of the CLS calibration. (Reproduced by permission of Academic Press and of the author from D.M. Haaland, in *Practical Fourier Transform Infrared Spectroscopy: Industrial and Laboratory Chemical Analysis*, eds. J.R. Ferraro, K. Krishnan, Academic Press, San Diego CA, 395–468 © 1990.)

the n samples, and **K** constitutes the $m \times i$ matrix of pure-component spectra. $\mathbf{E_a}$ is the $n \times i$ matrix showing the spectral noise present in the spectra. The least squares solution for the **K** matrix in Equation (6) is then Equation (7):

$$\langle \mathbf{K} \rangle = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathbf{T}}\mathbf{A} \tag{7}$$

where $\langle \mathbf{K} \rangle$ is the least-squares estimate for the **K** matrix coefficients. Prediction for samples of unknown concentration is straightforward since **K** can be obtained from Equation (7) and **A** is calculated from the spectral data. The vector equivalent of Equation (6) is then solved for the concentration.

The advantages of CLS are that it can be used for moderately complex mixtures and that the full spectrum can be used, thereby averaging out spectral noise. However, one must know the concentration of all components. Also, the method assumes that components do not interact. Finally, the calibration is extremely sensitive to baseline and collinearity effects.

3.2.2 Partial Least Squares Calibration

Some of the limitations of the CLS calibration can be overcome by using factor analysis methods such as PLS. The latter is a factor analysis method that creates a new coordinate system by generating mixed-component spectra instead of pure-component spectra. This method simultaneously decomposes both the calibration spectra and the concentration data into two basis sets of loading vectors and two sets of scores: one set for the spectral data, and the other for the constituent concentrations. Thus, we have Equations (8) and (9):

$$\mathbf{A} = \mathbf{TB} + \mathbf{E}_{\mathrm{a}} \tag{8}$$

$$\Updownarrow$$

$$\mathbf{C} = \mathbf{TV} + \mathbf{E}_{\mathrm{c}} \tag{9}$$

where **A** is the $n \times i$ matrix of spectral data (n spectra with i frequencies) expressed as absorbance values. **T** is the $n \times h$ score (intensity) matrix in the new coordinate system of the h PLS loading vectors for the n sample spectra, whereas **B** is the $h \times i$ loading-vector matrix of coefficients relating the scores (intensities) of **T** to the absorbance values. Referring to Equation (9), **C** is the $n \times m$ matrix of component concentration data (n spectra with m components). **V** is the $h \times m$ loading-vector matrix of coefficients relating the scores (intensities) of **T** to the concentrations, whereas $\mathbf{E}_c$ is the $n \times m$ matrix of concentration errors not fitted by the model. The rank h is the optimal number ($h < n$; $h < i$) of loading vectors for modeling the calibration spectra in **A**. Equation (9) is similar to that used in the ILS method, whereby concentrations are modeled as a function of absorbance for a number of selected frequencies. Indeed, in a noise-free system, the set of full-spectrum simultaneous linear equations is over-determined and only a small number of frequencies are needed to fully characterize the calibration spectra in **A**. Therefore, the true spectral variation will be concentrated in the first few loading vectors. These loading vectors are linear combinations of the original calibration spectra. The two sets of scores are forcibly related to each other by swapping their contents before the contribution of each factor is removed from the raw data. The vector equivalent of Equation (9) is thus solved one component at a time for each value of m. The use of the concentration information during the decomposition process causes spectra containing high component concentrations to be weighted more heavily than low-concentration spectra. PLS thus takes advantage of the correlation that already exists between the spectral data and the component concentrations. Instead of obtaining pure-component spectra as in CLS, one obtains a set of loading vectors that not only model the spectral contribution of each component, but also Beer's law nonlinearities, chemical interactions, and baseline fluctuations. By contrast, in the CLS calibration model, the factors are constrained to follow Beer's Law, with one factor per analyte present. The intensities are the

Table 2 Summary of calibration methods

Method	Applications	Advantages	Drawbacks
Univariate/peak height measurements	Simple mixtures; no peak overlap	Simple calibration procedure	Spectral noise; matrix interferences
CLS	Simple matrices; some peak overlap	Less spectral noise; full-spectrum method	Baseline effects; collinearity problems; must know active species
ILS	One component of interest in mixtures	Few frequencies; less complex calibration	More samples needed for calibration
PLS	Many components; complex mixtures	Same as CLS; correct for baseline effects; overcomes collinearity problems	Sensitive to design of experiments

component concentrations, whereas the loading vectors are the pure-component spectra.

The main advantage of PLS is that it can be used for very complex mixtures since only the concentration of the constituents of interest is required, as opposed to CLS where the concentration of all constituents is required. Another advantage of the PLS calibration method is its ability to overcome collinearity effects. These effects are often seen in spectral data whereby the presence of inter-correlated variables leads to unstable calibrations. PLS is also a full-spectrum method, unlike ILS, so that the spectral noise can be spread out over many frequencies. This type of calibration can sometimes predict samples in the presence of contaminants not present in the original calibration. Thus, the method is particularly useful for pulp and paper applications since samples show considerable band overlap and significant matrix interferences. However, a large number of samples are required for generating an accurate calibration. Table 2 summarizes the advantages and drawbacks of various types of calibration.

4 LABORATORY APPLICATIONS

For a good overview on applications of FTIRS to pulp and paper problems, the reader is referred to the excellent reviews by Banerjee[18] and Faix.[19] In this section, we will cover qualitative applications such as wood species characterization, lignin analysis, determination of extractives and contaminants, as well as the analysis of papermaking additives. A discussion of the development of quantitative methods will be provided in a later section, along with some specific examples.

4.1 Lignocellulosic Material Characterization

4.1.1 Wood Species Classification

Subtle interspecies differences in cellulose, hemicellulose [9034-32-6], lignin, lignocellulose [11132-73-3] and extractive content can be detected through by careful spectroscopic analysis. DRIFTS can be used for differentiating between species of coniferous wood.[28] A small number of frequencies are selected by performing a combination of PCA and correlation. The approach works best for species with high levels of extractives.

4.1.2 Cellulose and Lignin Characterization

Deconvolution can be used for measuring the increase in the degree of crystallinity[29] observed for cellulose I in samples originating from ramie, cotton, sulfate pulp and bacterial cellulose (Figure 9a). The same trend can be observed (Figure 9b) through an increase in the band at 3440 cm^{-1}, with the sequence going from viscose fibers (amorphous) to cellulose crystals. By using DRIFTS and potassium bromide [7758-02-3] (KBr) pellets, one can perform[30] a structural analysis of lignin that relates IR spectral information to several structural features in lignin samples as determined by conventional methods. As shown in Figure 10, good correlations are obtained for the hydrolysis ratio (a), the number of phenolic hydroxyl (b) and methoxy groups (c) per phenyl propane unit. Significantly better results are obtained with the DRIFTS method than with the KBr pellet method because of the latter's sensitivity to moisture content. One can also monitor the changes that occur in wood during de-lignification with various pulping agents such as cold sodium chlorite [7758-19-2], caustic soda [1310-73-2], soda-anthraquinone [84-65-1] and neutral sodium sulfite [7757-83-7] with KBr pellets. Changes in fiber oxidation and in the amount of lignin, as well as the degradation of hemicellulose, can easily be detected.[31]

4.1.3 Color/Brightness Reversion Studies

The brightening of mechanical pulp with hydrogen peroxide [7722-84-1] can be studied with the use of PAS.[32] One must select three internal-standard peaks (3400, 1510 and 1056 cm^{-1}). A decrease in peak intensities in the region from 1750–1600 cm^{-1} is found to correspond

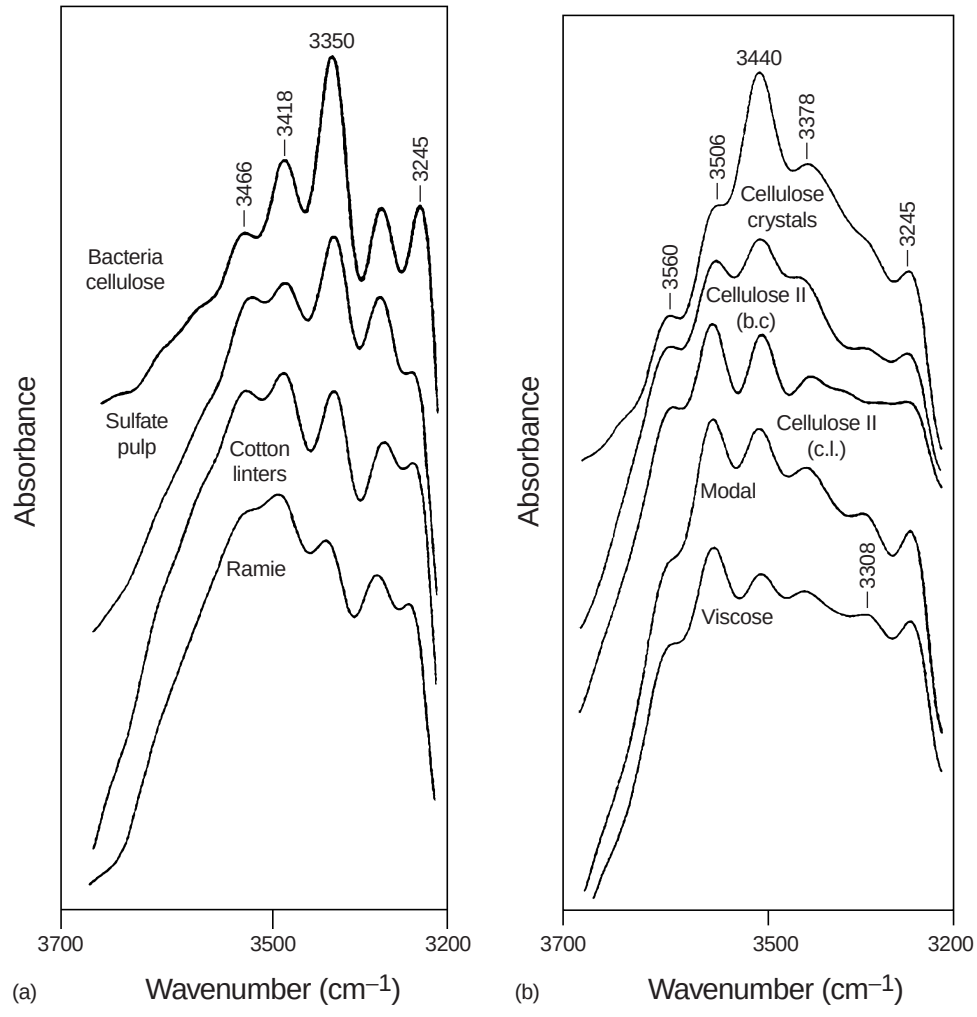

Figure 9 (a) Deconvoluted FTIR spectra (3200–3700 cm^{-1}) of various samples of cellulose I; (b) deconvoluted FTIR spectra (3200–3700 cm^{-1}) of various samples of cellulose II; c.l. = cotton linters, b.c. = bacterial cellulose. (Reproduced by permission of the author from D. Fengel, *Holzforschung*, **46**(4), 283–288 © 1992.)

to alkaline deacetylation (1740 cm^{-1}) and to the net removal of conjugated carbonyl structures (1650 cm^{-1}). The effect of irradiation wavelength on the color reversion behavior of unbleached and bleached mechanical pulp can be studied.[33] As shown in Figure 11, a decrease in the guaiacyl peak for lignin (1507 cm^{-1}) is observed at shorter irradiation wavelengths, accompanied by a simultaneous increase in carbonyl content (1719 cm^{-1}). An increase in the peak for conjugated carbonyl groups is also observed at 1678 cm^{-1} for irradiation frequencies of 390 and 500 nanometers.

4.2 Extractives Analysis

Sitholé[34] reviewed analytical methods for the analysis of extractives. The use of FTIRS and NIR spectroscopy were briefly discussed.

4.2.1 Resin and Fatty Acids

The composition of crude tall oil [8002-26-4] can be elucidated with the use of DRIFTS.[35] As shown in Figure 12, one obtains a relatively featureless spectrum, except for the carboxyl and acid bands. Nevertheless, characteristic absorption peaks appear at 1720, 1460 and 1278 cm^{-1}. When the analyst is faced with a complex matrix, a sequential extraction scheme must be used in conjunction with FTIRS so as to maximize the benefits of the technology.

4.2.2 Triglycerides

Triglycerides deposited on cotton fabrics[36] can be studied with the aid of the ATR sampling technique. The triglycerides originate from artificial body soil, which constitutes a representative mixture of free fatty

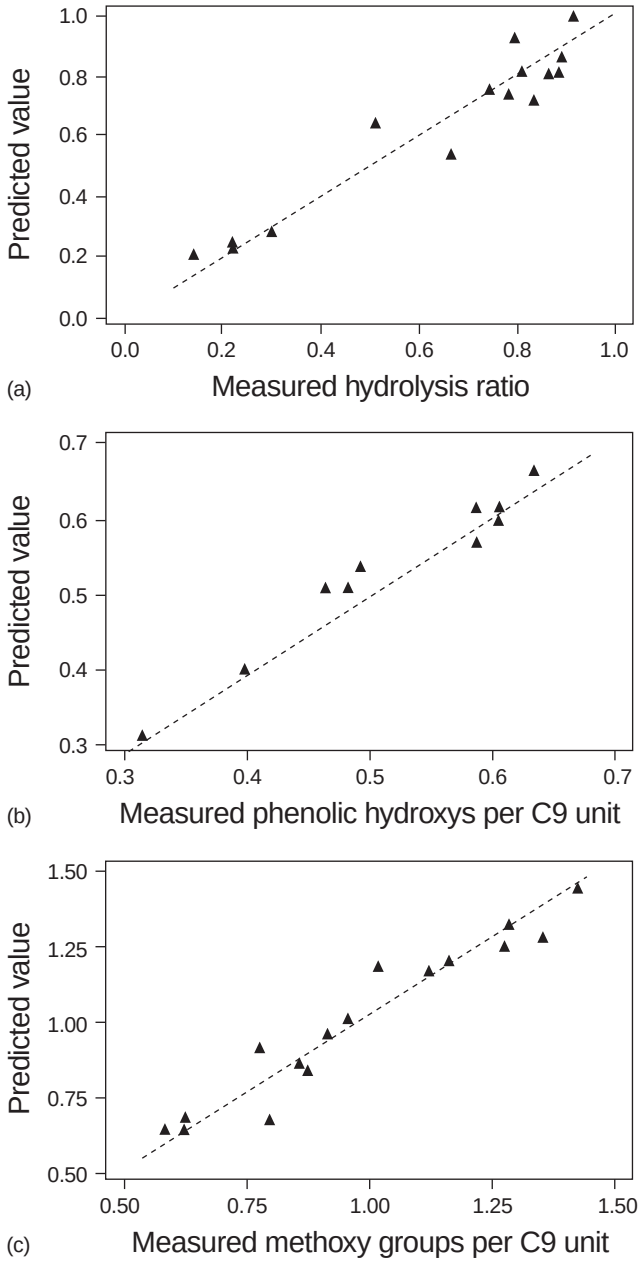

Figure 10 (a) Hydrolysis ratio measured by the conventional procedure versus the value predicted by DRIFTS; (b) Phenolic hydroxy concentration versus the value predicted by the DRIFTS method; (c) Methoxy group concentration measured by the conventional procedures versus the DRIFTS-predicted value. (Reproduced by permission of the authors from T.P. Schultz, W.G. Glasser, *Holzforschung*, **40**(Suppl.), 37–44 © 1986.)

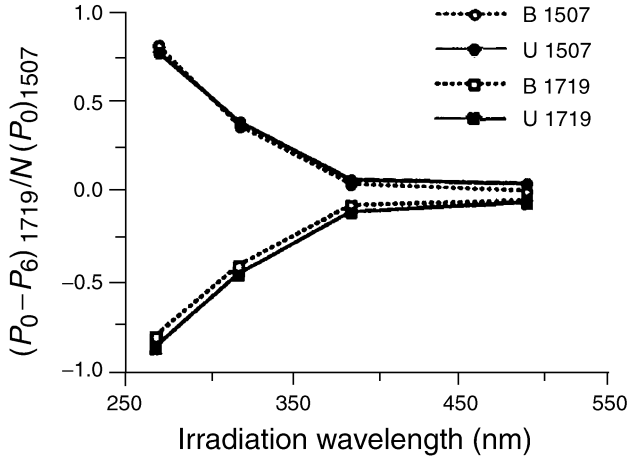

Figure 11 Increase in the 1719 cm^{-1} band due to irradiation, relative to the initial peak height of 1507 cm^{-1} band, $(P_0 - P_6)_{1719}/N(P_0)_{1507}$ (lower curves). The normalized decrease, D_N, of the 1507 cm^{-1} band is shown for comparison (upper curves). (Reproduced by permission of the authors from F. Kimura, T. Kimura, D.G. Gray, *Holzforschung*, **48**(4), 343–348 © 1994.)

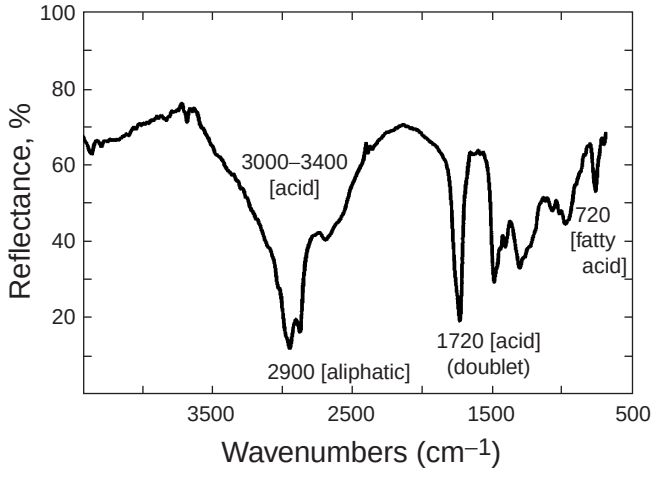

Figure 12 DRIFTS spectrum of crude tall oil dissolved in methyl *tert*-butyl ether (MTBE). (Reproduced by permission of the Technical Section, CPPA from D.F. Leclerc, M.D. Ouchi, *J. Pulp Pap. Sci.*, **22**(3), J112–J117 © 1996.)

acids [61790-12-3], triglycerides, mono- and diglycerides, hydrocarbons and cholesterol [57-88-5]. As shown in Figure 13, pressing the cotton fabric several times against the ATR element results in a steady increase in the triglyceride ester peak at 1750 cm^{-1}. This technique is directly applicable to the study of sterol esters deposited on pulp.

4.3 Contaminant Analysis

DRIFTS has proven to be very useful as a method for analyzing plastic, pitch and dirt specks found in the kraft process and on market pulp sheets because of the minimal preparation required. Samples can even be analyzed as received. Here are a few examples.

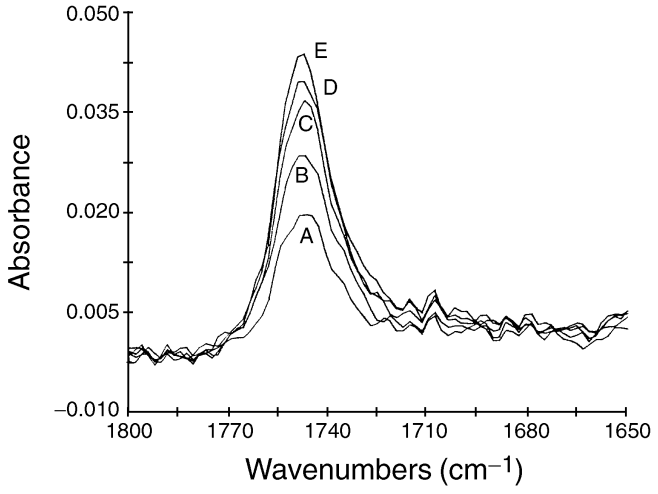

Figure 13 IR spectra of artificial body soil transferred to the ATR crystal after successive presses from first press (A) to fifth press (E). (Reproduced by permission of the Society for Applied Spectroscopy from R.S. Shreedhara Murthy, D.E. Leyden, J.K. Grime, *Appl. Spectrosc.*, **39**(6), 1047–1050 © 1985.)

4.3.1 Stickies/Deposits and Plastics

Recycled fiber being used in pulp and paper mills contains contaminants such as adhesives, hot melts, coating binders, ink residues, de-inking chemicals and wet-strength resins. These contaminants are referred to as 'stickies'. Since these contaminants often contain more than one component, IR identification is often difficult without some form of fractionation. Stickies and deposits can then be analyzed by either DRIFTS or transmission, after sequential extraction with organic solvents such as chloroform [67-66-3], methanol [67-56-1] or acetone [67-64-1].[35] Detailed descriptions of analytical procedures for sequential extraction are given by Sitholé et al.[37] and Guo and Douek.[38] The extract is then applied directly on an alkali halide salt, and the solvent is made to evaporate prior to analysis. FTIR analysis can even distinguish between fatty acid and resin acid soaps of aluminum and calcium:[37] calcium fatty acid soaps have a characteristic doublet at 1590 and 1540 cm^{-1}, whereas calcium resin acid soaps have absorption peaks at 1549 and 1405 cm^{-1}. Aluminum fatty acid soaps absorb strongly at 3690, 1590, 1472 and 985 cm^{-1}, whereas their resin-acid counterparts only absorb strongly at 1585 and 1435 cm^{-1}.

Indeed, some of these characteristic peaks may sometimes be masked in the spectrum of a deposit sample, but not always. For example, a DRIFTS spectrum of a typical solid soap deposit is shown in Figure 14.[35] The deposit is simply mixed with potassium chloride (KCl) and ground to a powder. The peak characteristic for soap deposits appears at 1550 cm^{-1}, whereas the peak characteristic of an acid (1720 cm^{-1}) is absent. Small amounts of defoamer (3295 cm^{-1}) and talc (3665 cm^{-1} and 1015 cm^{-1}) are also present. On the other hand, a Uhle box deposit will contain contaminants from various sources, and is best analyzed following sequential extraction firstly with ethanol, and secondly with chloroform. As shown in Figure 15,[38] FTIR analysis successfully identified the major component of the deposit as a SBR polymer. Polymeric contaminants in market kraft pulp can also be analyzed.[39] Results indicate that polyethylene [9002-88-4], polypropylene [25085-53-4] and latexes are prevalent in the contaminated pulp.

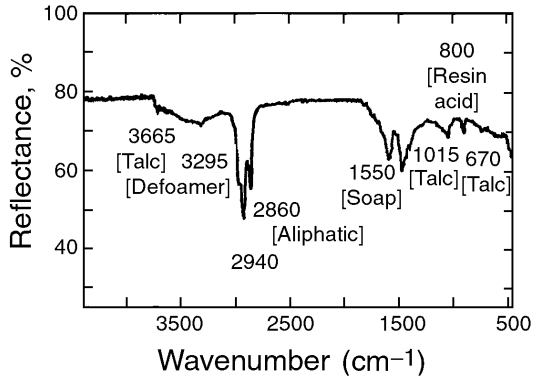

Figure 14 DRIFTS spectrum of a solid deposit collected from a pulp dryer screen. (Reproduced by permission of the Technical Section, CPPA from D.F. Leclerc, M.D. Ouchi, *J. Pulp Pap. Sci.*, **22**(3), J112–J117 © 1996.)

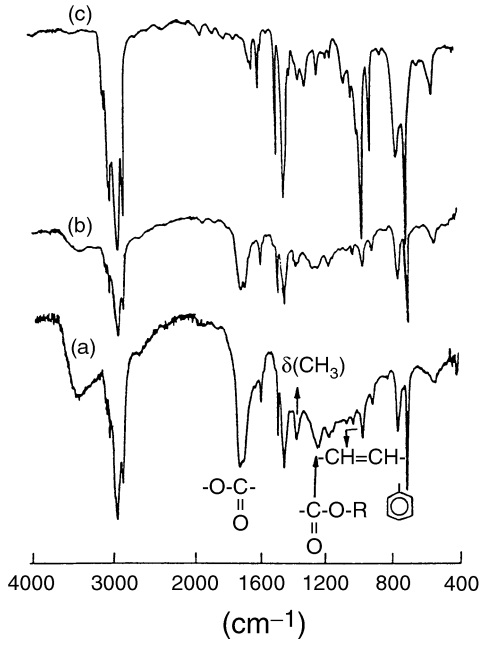

Figure 15 FTIR spectra of (a) acetone extract and (b) chloroform extract of Uhle box deposit; (c) styrene-butadiene rubber (SBR) copolymer. (Reproduced by permission of the Technical Section, CPPA from X.-Y. Guo, M. Douek, *J. Pulp Pap. Sci.*, **22**(11), J431–J439 © 1996.)

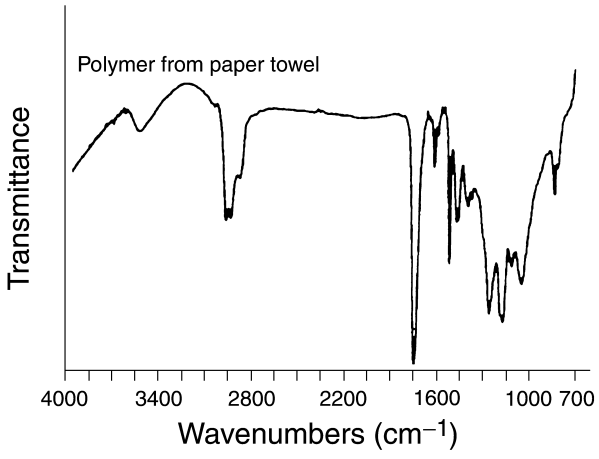

Figure 16 IR spectrum of a polymer removed from a paper towel (ethyl acrylate/styrene copolymer). (Reproduced by permission of Marcel Dekker from A.J. Sommer, P.L. Lang, B.S. Miller, J.E. Katon, in *Infrared Microspectroscopy: Theory and Applications*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, New York, 245–258 © 1988.)

4.3.2 Micro-sized Contaminants

Very small contaminants are more homogeneous, and thus easily analyzed by IR microscopy.[24] A particularly interesting example is the study of particulate contaminants in recycled paper products. Figure 16 shows the IR spectrum of a polymeric contaminant taken from a paper towel. The spectrum matches that of an ethyl acrylate/styrene copolymer [25066-97-1]. Examination of dust particles from an office environment shows that paper products are a major source of dust.[40]

IR microscopy can also be employed to solve contaminant problems associated with raw stock supplies, pulping operations, sheet formation and the use of recycled pulp. For example, contaminants such as poly(vinyl alcohol) (PVA) [9002-89-5], chitin, defoamers and silicone oils can be identified as well.[41] The analysis of pulp and paper spots and defects, deposits, fibers, and laminates is very easy with FTIR microscopy. For example, the photomicrograph shown in Figure 17(a) represents a cross-section of a laminate, in this case, a plastic carton.[16] Four layers are visible in the micrograph. The top and center tie layers are about 100 μm thick and are cellulosic in nature (spectrum: Figure 17c), whereas the two pigmented layers are made of polyethylene (spectrum: Figure 17b).

4.4 Characterization of Papermaking Additives

The ever-increasing use of recycled pulp fibers and the continued trend to lower the grammage of paper sheets have made the judicious use of papermaking additives even more important in maintaining wet- and dry-end sheet strength during the papermaking process. The

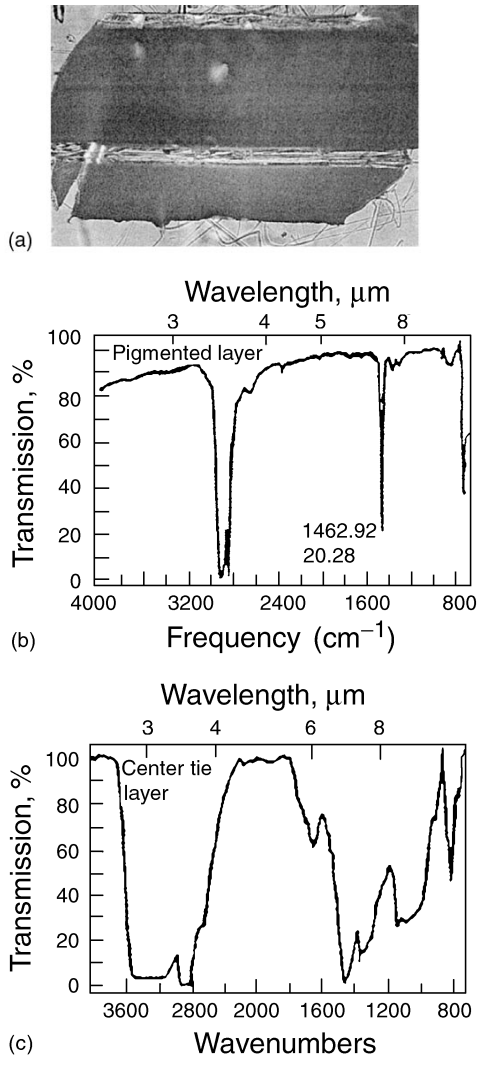

Figure 17 (a) Photomicrograph (100×) of a plastic carton and IR spectra of (b) pigmented layers and (c) the center tie layer and surface coating. (Reproduced by permission from K.M. Sweeney, *Tappi J.*, **72**(2), 171–174 Copyright © 1989 TAPPI.)

following examples show that FTIR can play an important role in elucidating the mechanism behind the effect of novel additives on paper strength.

4.4.1 Wet-end Additives

Cross-linking agents such as 1,2,3,4-butanetetracarboxylic acid [1703-58-8] (BTCA) or poly(maleic acid) [26099-09-2] (PMA) improve the wet strength of paper by producing ester cross-linking bonds between cellulose molecules. DRIFTS analysis of BTCA-treated paper[42] shows that there is a perfect correlation ($r^2 = 0.999$) between the ester carbonyl band absorbance of the treated paper and its wet-strength retention. As shown in

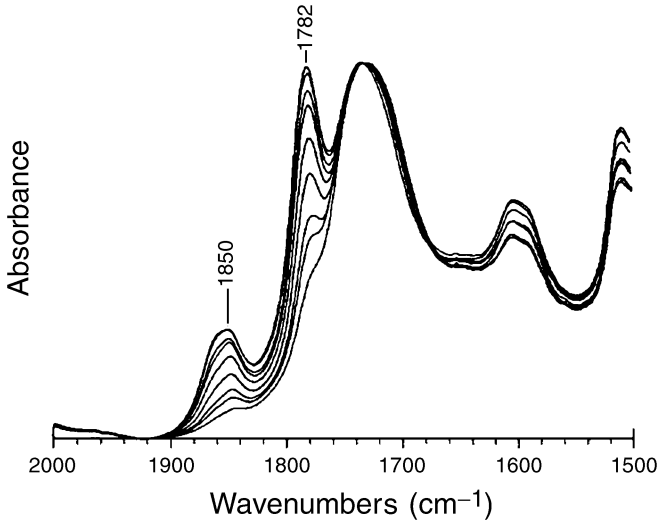

Figure 18 IR spectra of the sheets treated with 4% PMA and cured at different temperatures for two minutes. The curing temperatures are 120, 130, 140, 150, 160, 170, 180 and 190 °C (from bottom to top). (Reproduced by permission from C.Q. Yang, Y. Xu, D. Wang, *Ind. Eng. Chem. Res.*, **35**(11), 4037–4042 © 1996 American Chemical Society.)

Figure 18, a five-membered cyclic anhydride is formed as an intermediate because the asymmetric stretching band at 1782 cm^{-1} is stronger than the symmetric stretching band found at 1850 cm^{-1}, as expected for a cyclic anhydride.

4.4.2 Dry-end Additives and Coatings

Synthetic dry-strength resins such as polyacrylamide [9003-05-8] (PAM) resins have unique properties that make them especially effective as dry-strength additives. Tensile strength generally increases with the amount of additive retained by the sheet and the degree of beating. The extent of PAM retention on handsheets has been studied[43] with the use of the ATR sampling technique. As shown at the top of Figure 19, the strength of the amide absorption band at 1670 cm^{-1} seems to increase nonlinearly with beating degree and the retention of PAM, a spurious effect that is in reality due to surface roughness. By using the cellulose band at 1315 cm^{-1} as an internal standard, the nitrogen content of the sheet as measured by the strength of the amide band now linearly increases with PAM retention, irrespective of beating degree.

Coated sheets contain a variety of additives and fillers like hydroxymethylcellulose [37353-59-6] (HMC), poly(vinyl acetate) [9003-20-7] (PVAc), clay (a mixture of silicates) and calcium carbonate. A DRIFTS spectrum of a coated sheet[18] is shown in Figure 20. The alkyl ether and carbonyl peaks characteristic of PVAc can be clearly

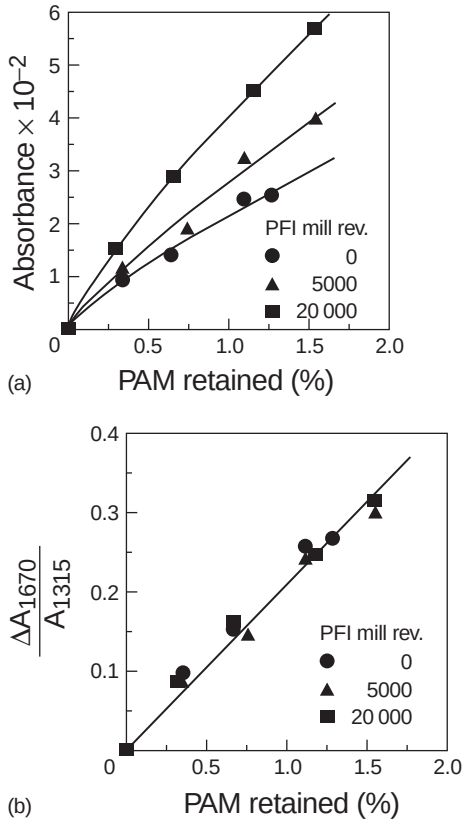

Figure 19 (a) Relationship between the absorbance of the band at 1670 cm^{-1} in difference ATR spectra and the amount of PAM retained in the sheets; (b) relationship between the corrected absorbance value at 1670 cm^{-1} in difference ATR spectra and the amount of PAM retained in the sheets. (Reproduced by permission of the authors from D. Tatsumi, T. Yamauchi, K. Murakami, *Nord. Pulp Pap. Res. J.*, **10**(2), 94–97 © 1995.)

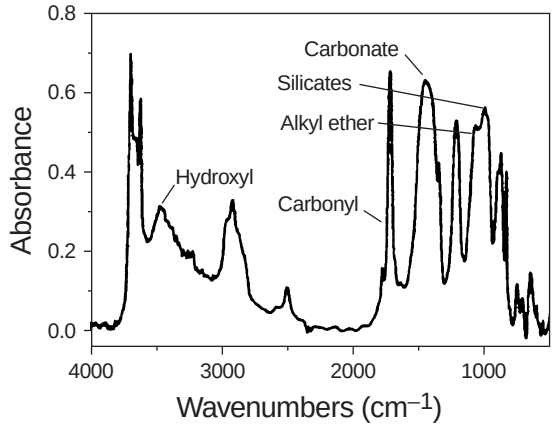

Figure 20 DRIFTS spectrum of a base sheet coated with a mixture of clay, calcium carbonate, PVAc, and HMC. (Reproduced by permission from M.A. Friese, S. Banerjee, in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, 119–141 Copyright CRC Press, Boca Raton, Florida © 1995.)

seen, along with the hydroxyl peak for HMC and those for carbonate and clay.

5 ON-LINE APPLICATIONS: ANALYSIS OF PROCESS LIQUORS

FTIR spectrometers are well suited for on-line applications because of their ability to rapidly determine the concentration of several components at once. The short analysis time being provided by FTIR spectrometry is a prerequisite for process control. Also, because of the need for small path lengths, ATR is the sampling method of choice for aqueous streams, especially in pilot-plant environments and if one carefully selects the right IRE for the application. With the use of multivariate calibration techniques, the chemical components present in liquors (lignin, carbohydrate, and inorganic compounds) can be determined from the FTIR spectra. Individual calibrations can be developed for each type of liquor being monitored. However, because of the limited throughput of chalcogenide fibers, Mid-IR applications cannot use remote sampling locations, and must rely on sampling systems to bring samples to the analyzer, thereby limiting sample throughput. A schematic of a typical on-line installation is given in Figure 21.[44]

On the other hand, in the NIR fiber optics can be used for sensing remote locations. The superior throughput and wavelength accuracy of FTIR spectrometers are essential for using long path lengths, for sampling multiple aqueous streams, and for accurately determining individual components. The use of fiber optics not only increases sample throughput, but is an extremely cost-effective way of interfacing one single instrument to many individual applications which, a few years ago, would have needed their own dedicated instrument. The main drawback is that, since spectra in the NIR region of the spectrum consist of wide overtone and combination bands, a calibration is more difficult to implement, and one must design the calibration experiment carefully to avoid problems such as collinearity. A thorough discussion of the opportunities and challenges presented by process applications of FTIRS is given by Hassell and Bowman.[44]

5.1 Kraft Pulping

A rapid ATR method for determining the effective alkali and inorganic content of kraft process liquors was developed.[45] An ATR spectrum of kraft black liquor is shown in Figure 22. The peak assignments are given in Table 3. Requiring no reagents, the method is faster and more reliable than current laboratory methods. This method has since been commercialized with a mill-hardened spectrometer. Good results are reported when the method is combined with a PLS calibration method:[46] results from the new on-line method closely follow those obtained with a nearby existing on-line titrator. A report has also been issued on the mill performance of this technology.[47] The general assessment is that the method will follow process trends over the long term.

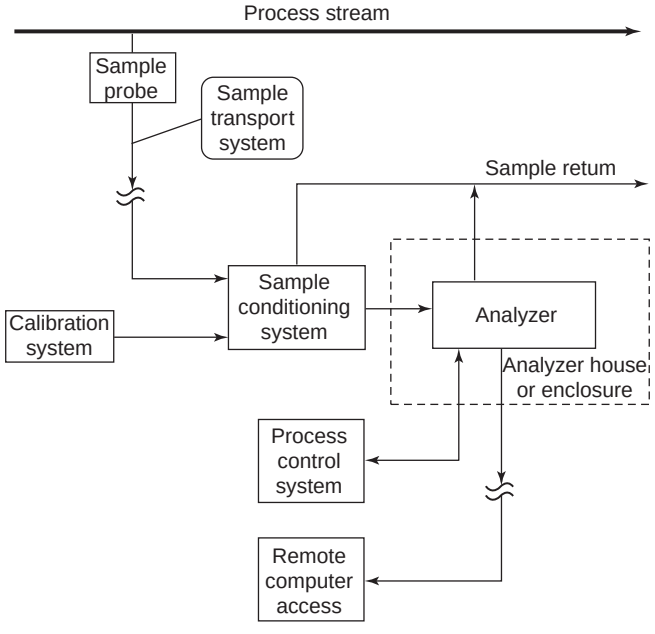

Figure 21 Generalized schematic diagram of a process analyzer. (Reproduced by permission of the Society for Applied Spectroscopy from D.C. Hassell, E.M. Bowman, 'Process Analytical Chemistry for Spectroscopists', *Appl. Spectrosc.*, **52**(1), 18A–29A © 1998.)

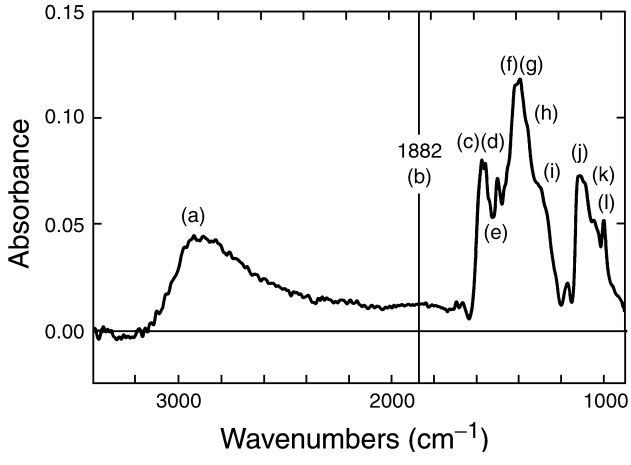

Figure 22 ATR absorbance spectrum of a typical kraft black liquor. Peak assignments identified by lower case letters are given in Table 3. (Reproduced by permission of the Technical Section, CPPA from D.F. Leclerc, R.M. Hogikyan, *J. Pulp Pap. Sci.*, **21**(7), J231–J237 © 1995.)

Table 3 Peak assignments for kraft black-liquor spectra

Reference to Figure 22	Frequency (cm^{-1})	Source	Assignment
(a)	2957	Sodium carbonate [497-19-8]	C−O stretch, $2\nu_3$
(b)	1882	Sodium hydroxide [1310-73-2]	O−H stretch
(c)	1595	Syringyl lignin [9005-53-2]	Aromatic C=C stretch
(d)	1550	Carboxylate ion	Antisymmetric COO^- stretch
(e)	1491	Guaiacyl lignin	Aromatic C=C stretch
(f)	1405	Carboxylate ion	Symmetric COO^- stretch
(g)	1386	Carbonate ion	C−O stretch, ν_3
(h)	1355	Hemicellulose [9034-32-6]	C−O stretch
(i)	1295	Guaiacyl lignin	Aromatic C=C bending
(j)	1104	Sodium sulfate [7757-82-6]	S−O stretch, ν_3
(k)	1035	Lignin	C−O stretch
(l)	1001	Sodium thiosulfate [7772-98-7]	S−O stretch, ν_3

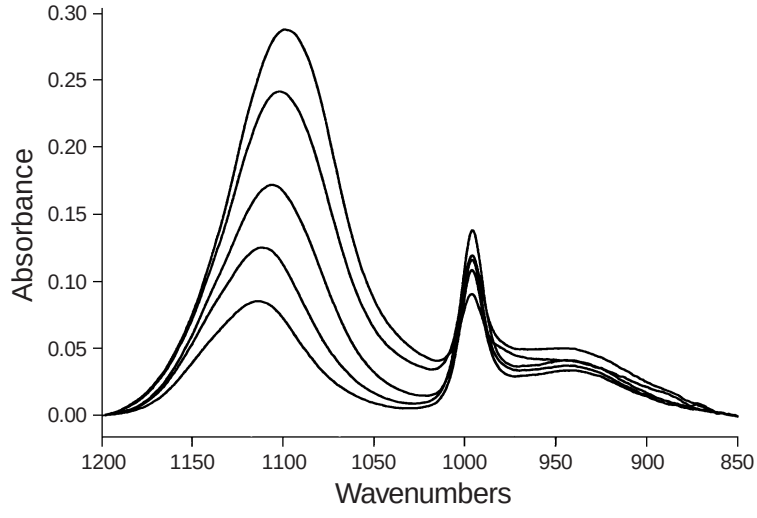

Figure 23 FTIR spectra of white liquor solutions at increasing times (from bottom to top). (Reproduced by permission of the Society for Applied Spectroscopy from J. Bandekar, R. Sethna, M. Kirschner, *Appl. Spectrosc.*, **49**(11), 1577–1582 © 1995.)

However, an acid wash is required from time to time, and the optical element must be replaced regularly.

The course of the oxidation reaction of white liquor by molecular oxygen has been studied in an ATR sample cell.[48] An example of an oxidation run is shown in Figure 23, whereby increasing amounts of sodium sulfate [7757-82-6] (1104 cm^{-1}) and sodium thiosulfate 7772-98-7 (995 cm^{-1}) are formed as the reaction proceeds.

An univariate ATR method for measuring kappa numbers in laboratory cooks has been proposed.[49] The method measures the absorbance produced by dissolved lignin in black liquor at 1118 cm^{-1}, and works well for laboratory cooks for which the chip moisture content and the liquor-to-wood ratio is known precisely. However, it would be difficult to implement in the mill where these values are usually not determined with great precision.

Recent advances in FTIR instrumentation and software have made possible the use of the NIR region of the spectrum for determining aqueous components such as dissolved electrolytes. Advantages over previous methods include: no sample preparation, short measurement times, long path lengths and the capacity to use fiber-optic technology for real-time, in situ analysis. A NIR kraft-liquor analysis method has been reported,[50] whereby a PLS calibration is carried out for measuring the concentration of sodium hydroxide 1310-73-2, sodium hydrosulfide 16721-80-5 and sodium carbonate 497-19-8. Particularly good results were obtained for effective alkali. A similar method has also been demonstrated for hydrosulfide.[51]

5.2 Alkaline Sulfite Anthraquinone Methanol Pulping

Another practical use of liquid-state FTIR spectrometry concerns the analysis of alkaline sulfite anthraquinone methanol (ASAM) liquors, as well as control of the ASAM pulping process. A method for controlling the cooking degree of pulp is described.[15] The method

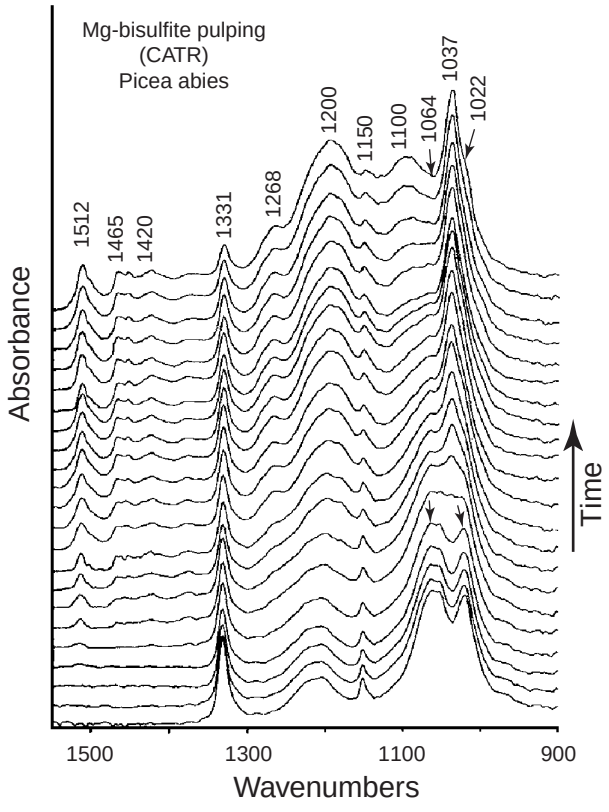

Figure 24 Absorption spectra of spent liquors of a magnesium bisulfite pulping sequence for spruce. Peak assignments are: free SO_2 (1331 and 1150 cm^{-1}), HSO_3^- (1200, 1064 and 1022 cm^{-1}), lignosulfonates (1037 cm^{-1}), lignin (1512 cm^{-1}). (Reproduced by permission of Springer-Verlag from O. Faix, in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 233–241 © 1992.)

monitors the amount of dissolved lignin present in the liquor during the cook. The method uses correlations of either peak heights or peak integrals between 1400 and 1550 cm^{-1} with kappa numbers. Correlations with the level of inorganic compounds dissolved in the liquor such as sulfite also gave positive results.[52] An example is shown in Figure 24, which shows the changes in the FTIR spectrum of magnesium-based bisulfite [7757-88-2] (HSO_3^-) liquor with cooking time. A decrease in the levels of both free sulfur dioxide [7446-09-5] (SO_2) (1331 cm^{-1}) and bisulfite (1022 and 1064 cm^{-1}) is accompanied by an increase in the concentrations of lignin (1512 cm^{-1}) and lignosulfonates [8061-54-9] (1037 cm^{-1}).

5.3 Other Applications

The degree of nitration of cellulose nitrate [9004-70-0] was determined[53] with the combination of a gel permeation chromatography (GPC) column and an on-line transmission flow cell. The solvent was tetrahydrofuran [109-99-9] (THF). The antisymmetric nitrate stretching region around 1650 cm^{-1} was used for estimating the degree of nitration.

Table 4 summarizes the scope of pulp and paper applications of FTIRS with regard to analyte, choice of sampling method, matrices, and key references.

6 QUALITY CONTROL AND TROUBLESHOOTING

In this section, we will briefly discuss problems that affect data quality such as deviations from Beer's law, apodization, low resolution, and spectral noise caused by excessive vibration. Griffiths and de Haseth[1] have discussed these problems and their effects on quantitative analysis in more detail.

6.1 Instrument Performance Issues

6.1.1 Laboratory Performance

In most cases, the FTIR spectrometer will be located in the technical department, which is very often located next to the mill. Low- and high-frequency mechanical vibrations are the norm, and care must be taken to insulate the instrument from vibrations. If not, then a stronger apodization must be used so that lowering the resolution preserves the integrity of the data. Most pulp and paper applications deal with bands that are relatively broad, and so the choice of the apodization function

Table 4 Summary of pulp and paper applications of FTIR spectrometry

Analyte	Sampling method	Matrices	Comments	Refs.
Wood species	DRIFTS	Wood	Extractives	28
Lignin/cellulose	DRIFTS/ATR/PAS	Wood/pulp/paper	Specular reflectance	29, 33
Extractives	DRIFTS/ATR	Pulp/paper	Sequential analysis	34, 35
Stickies/deposits	DRIFTS/microscopy	Pulp/paper		24, 38
Fillers/coatings	DRIFTS/ATR	Paper		18
Additives	DRIFTS/ATR	Paper		42, 43
Inorganic ions	ATR/NIR	Process liquors	ZnS/SiO_2 best materials	45–52
Cellulose nitrate	Transmission	THF		53

will not affect band shapes significantly. However, some spectra will exhibit a combination of sharp peaks, i.e. lignin, and broad peaks such as carbonate, sulfate, etc., and so the choice of apodization will depend on the application. In general, a stronger apodization will distort sharp peaks very significantly. Also, sharp peaks with an absorbance greater than unity will exhibit deviations from Beer's law even if the extent of apodization is small. Therefore, it is best to tailor the application so that none of the peaks will have an absorbance greater than 0.7. For instance, in an ATR application, one would select an optical element with a shorter effective path length.

6.1.2 Performance at Mill Sites

Instrument air which is available at mill sites is usually used for operating pneumatic valves, not for purging optical instrumentation, and its moisture content often fluctuates, especially in the summer. Since this could impair long-term performance of the optical system, it is best to use an instrument that either has a permanently purged optical compartment, or one that is continually dried by a desiccant which is replaced on a regular basis.

6.2 Data Quality Issues

6.2.1 Reststrahlen (Inversion) Bands

Samples with microscopic crystalline domains such as minerals often display spectral artifacts called reststrahlen or inversion bands, which are produced by specular reflection when analyzed by DRIFTS even at very low concentrations, whereas amorphous samples like lignin do not. Similar anomalous dispersion effects with pulp fibers have been reported.[54,55] An example of these effects is shown in Figure 25: spectral distortions can still

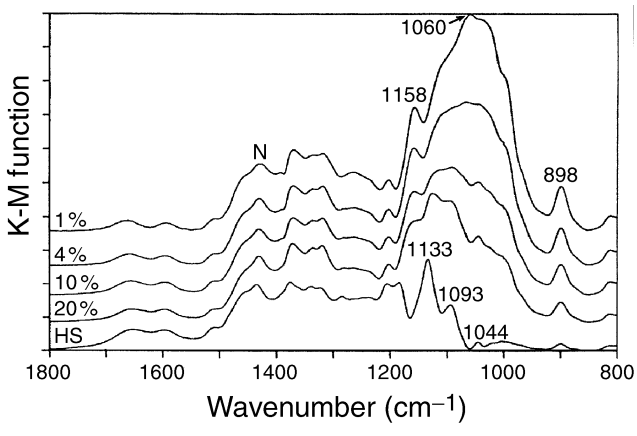

Figure 25 DRIFTS spectra of unbleached kraft pulp at different percent concentrations of samples in KBr. HS denotes the handsheet spectrum. (Reproduced by permission of Marcel Dekker from J. Bouchard, M. Douek, J. Wood, *Chem. Technol.*, **13**(4), 481–499 © 1993.)

be seen at a concentration as low as one percent. On the other hand, no specular reflection is observed with extracted lignin samples. This effect has been attributed to the structural features present in the cellulose portion of pulp fibers.[55] Therefore, DRIFTS analysis of pulp fibers must be done at the maximum dilution ratio dictated by the application, but not so much that any useful spectral features are lost.

6.2.2 Optimization of Attenuated Total Reflectance Sensitivity

Some ATR applications call for enhanced sensitivities. One can either increase the penetration depth, or the number of reflections. Of the two, the latter is the most effective since increasing the penetration depth only further distorts the fingerprint part of the spectrum. This can be done quite easily by selecting an optical element with a conical angle of 30° instead of 45°, or by buying a flow cell that accommodates a longer optical element. A conical angle of 30° effectively doubles the path length, which should satisfy most requirements.

7 DEVELOPMENT OF QUANTITATIVE METHODS

In this section, we will show a few typical examples of quantitative method development. We will also talk about the need for sound experimental planning during the design and calibration phases of a typical multivariate calibration method.

7.1 Univariate Calibration

When performing an univariate calibration, it is necessary to locate an isolated band so that the analyte can be quantified without interferences. Lignin exhibits some sharp absorption bands that do not significantly overlap with others. In particular, the band at 1510 cm^{-1} is amenable to peak-area measurements, especially for dry pulp. However, the water band at 1640 cm^{-1} is very wide and, even when present in small quantities, interferes significantly with any measurement which depends on a subjective choice of baseline. Nevertheless, a DRIFTS method, which estimates the amount of lignin in air-dried unbleached pulps, has been developed.[14] Lignin-free cotton linters were used as the reference material. A lignin spectrum is thus obtained after spectral subtraction of the cellulose contribution. As shown in Figure 26, a linear relationship is found between the area of the band at 1510 cm^{-1} and kappa number for a wide variety of species. However, the data show a significant amount of scatter, especially for low kappa numbers, probably

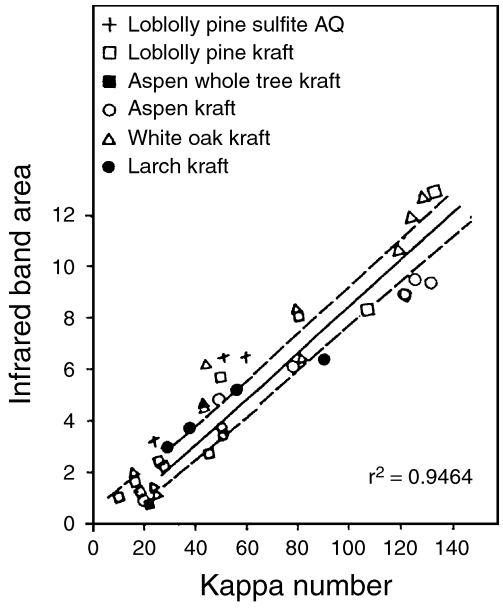

Figure 26 Regression line of the band area at 1510 cm^{-1} obtained after spectral subtraction versus kappa number for a number of wood species. (95% confidence limits are given by dashed lines.) (Reproduced by permission from S.A. Berben, J.P. Rademacher, L.O. Sell, D.B. Easty, *Tappi J.*, **70**(11), 129–133 Copyright © 1987 TAPPI.)

because of differences in the amount of specular reflection emitted by each pulp sample. In fact, it is reported that the results are no better than those obtained with traditional analysis. Therefore, univariate calibration must be used only when one is reasonably sure that a single physical phenomenon such as Beer's law or internal reflection is responsible for the spectral changes being monitored. Since most pulp and paper applications of FTIR deal with complex matrices, multivariate calibration must be used so that optimal results are obtained.

7.2 Multivariate Calibration

The selection of independent calibration samples will be discussed in the context of modern experimental design principles, and with regard to the analysis at hand. We will then present an example of a typical experimental design, followed by a brief discussion of the respective merits of the CLS and PLS calibrations, in a pulp and paper context.

The selection of calibration samples is essential to the success of any method, particularly when dealing with a large number of components per sample. Two-level factorial designs[56] enable the experimenter to detect any number of interactions between components, and permit the construction of robust models. Component concentrations are varied in a systematic and orthogonal fashion, so that the greatest amount of information is gathered from a small number of samples. These types of design allow the detection of interaction and nonlinear effects. When one is reasonably sure that no significant interactions occur, a half-factorial design is usually sufficient and more effective, since only half the number of samples is needed. Table 5 shows such a half-factorial calibration design for a four-component hypothetical green liquor for which the concentration range (in g L^{-1} as Na_2O) is as follows: EA = 5(−) to 45(+); NaHS = 5(−) to 35(+); Na_2CO_3 = 30(−) to 80(+); Na_2SO_4 = 2(−) to 12(+). The target concentration usually constitutes the center point of the range. The extremes of the range are calculated so that they lie outside the possible range of value found in the mill. Another type of design is the constrained mixture design[27], which is most suitable for modeling the spectral behavior of mixed-species pulps such as spruce, pine and fir. Depending on the application, both types of design can be employed for selecting the composition of samples.

7.2.1 Classical Least Squares Calibrations

Pulp and paper samples have complex matrices and usually contain minor constituents that are not of interest to the researcher. More often than not, these minor constituents exhibit a significant spectra response. Also, the spectra of these samples exhibit spurious effects such as baseline shifts that are hard to characterize, let alone quantify. Unfortunately, a CLS calibration requires that

Table 5 Example of a two-level half-factorial design with four factors for a hypothetical green liquor

Run/sample	Effective alkali (gL^{-1}) (as Na_2O)	Sodium sulfide (gL^{-1}) (as Na_2O)	Sodium carbonate (gL^{-1}) (as Na_2O)	Sodium sulfate (gL^{-1}) (as Na_2O)
1	5	5	30	2
2	45	5	30	12
3	5	35	30	12
4	45	35	30	2
5	5	5	80	12
6	45	5	80	2
7	5	35	80	2
8	45	35	80	12

the concentration of every constituent be known and assumes that the response is due entirely to constituents present in the calibration. Therefore, CLS calibrations have very little utility for pulp and paper users of FTIRS, and will not be discussed further.

7.2.2 Partial Least Squares Calibrations

The PLS calibration technique is tailor-made for pulp and paper applications, since only knowledge of the constituent of interest is required for the calibration. In addition PLS can sometimes be used for predicting unknown samples, even in the presence of contaminants not originally present in the original calibration samples. The PLS naturally lends itself to the accurate determination of pulp constituents. For example, the proportion of phenolic hydroxyl in milled wood lignins (MWL) was predicted versus that determined by aminolysis.[57] Results are shown in Figure 27. Best results are achieved when the spectral range from 800 to 2000 cm^{-1} is used. The PLS calibration gave absolute standard errors of 0.06% and prediction errors of 0.11%, a result which is six times better than that obtained by a simple linear regression based on the absorbance of the acetoxy bands at 1744 and 1766 cm^{-1}. Another method uses the DRIFTS technique[58] to rapidly determine lignin and carbohydrates such as arabinose [5328-37-0], galactose [59-23-4], glucose [50-99-7], mannose [3458-28-4] and xylose [25990-60-7] in pulp, as well as assess process properties such as yield. Table 6 provides a checklist for choosing the best calibration method.

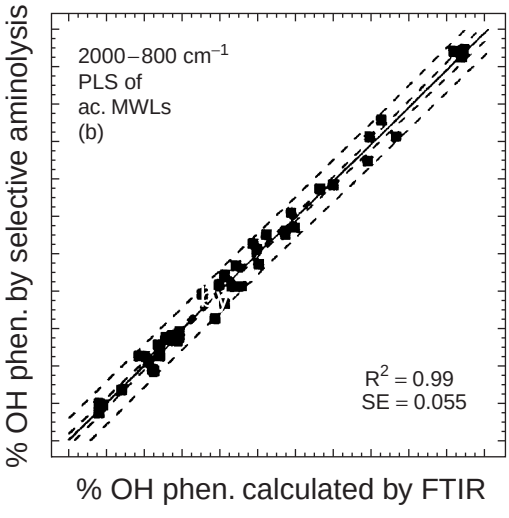

Figure 27 FTIRS prediction of the proportion of phenolic hydroxyl in acetylated MWL versus that determined by aminolysis. Inner dotted lines: 95% confidence limits: outer dotted lines: 95% prediction limits. (Reproduced by permission of the authors from O. Faix, J.H. Böttcher, *Holzforschung*, **47**(1), 45–49 © 1993.)

8 COMPARISON WITH OTHER SPECTROSCOPIC METHODS

In this section, we will first compare FTIRS with other spectroscopic methods such as NIR dispersive spectrometry, Raman spectroscopy and MS, and then briefly examine business issues relating to the purchase and maintenance of process instrumentation.

8.1 Comparison of Technology

8.1.1 Near-infrared Dispersive Spectroscopy

In the NIR region (4000 to 14 000 cm^{-1}) the weakness of overtone bands results in higher detection limits, which can be overcome by using thicker or undiluted samples. This enables the sample to be analyzed with little or no preparation. Much longer path lengths (from 0.5 mm to 1 cm) are also available for aqueous samples. NIR absorption bands are sensitive to their local environment, and one can distinguish between polar, steric and hydrogen bonding effects. In the reflectance mode, the penetration depth of IR radiation is about 10 to 100 μm, versus a few millimeters for NIR radiation. Detection limits are 100 parts per million for NIR whereas for the Mid-IR these are about 1 part per million for the transmission mode.[59]

A comparison between the use of NIR and FTIR for the analysis of pine lignin in pulped wood fibers was performed with the use of the transmission method.[60] Even though the two methods yielded similar calibration results, sample preparation and analysis for the NIR technique were much faster and simpler. On the other hand, FTIR gave the best standard error of prediction. Therefore, the advantages of using a FTIR spectrometer

Table 6 Checklist for choosing a quantitative method

Example	Method/region	Interactions (?)	Design	Refs.
Lignin/inorganics	Univariate/Mid-IR	No	Peak heights	14, 15, 45, 52, 53
Kappa no.	ILS/Mid-IR	No	Half-factorial	15, 52
Pulp	PLS/Mid-IR/NIR	Yes	Full-factorial	27, 57, 58, 60
Liquor analysis	PLS/Mid-IR/NIR	Yes	Full-factorial	46, 50, 51

are not so obvious, especially for qualitative analysis, since cheaper dispersive instruments have sufficient throughput for most laboratory applications.

The latest trend in NIR spectroscopy is to use an acoustic-optic tunable filter (AOTF) as a monochromator. This technology has been reviewed elsewhere.[6] An AOTF device has a number of advantages. First, a NIR AOTF spectrometer is cheaper than a FTIR spectrometer. Second, it can scan very rapidly and more than one wavelength can be recorded simultaneously. Third, it has a large aperture. However, resolution is not constant with frequency and is modest at best (between 30 and 50 cm^{-1} at 10 000 cm^{-1} in the noncollinear configuration) and the throughput, although greater than that of a dispersive instrument, is not sufficient for multiple-point fiber-optic applications. Finally, AOTF instruments do not have the frequency accuracy provided by the laser referencing of FTIR spectrometers, and need to be re-calibrated after servicing, thereby generating significant calibration maintenance costs.

The advantages of using a FTIR spectrometer for demanding NIR applications are thus readily apparent, especially when one desires to use a single instrument for developing quantitative methods to be applied at multiple sampling locations throughout the mill site without the need for re-calibration.

8.1.2 Classical Raman and Fourier Transform Raman Spectroscopy

Since both IR spectroscopy and Raman spectroscopy deal with molecular vibrations, the two techniques can be compared. The kind of information one can gather from a Raman experiment is similar but not identical to that obtained with IR spectroscopy. A molecular vibration is Raman-active if a change in polarizability occurs during the transition, whereas it is IR-active if a change in dipole moment is produced by the vibration. Thus, strongly covalent bonds have intense Raman bands, whereas molecules with strongly polar bonds produce intense IR bands. For example, water is a strong IR absorber, but a weak Raman scatterer. Water can therefore be present in samples analyzed by Raman spectroscopy. Since the selection rules are different for the two types of spectral transition, the information which one can gather from a Raman spectrum complements that which is obtained from an IR spectrum. Raman technology has been reviewed by Christensen et al.[61] Pulp and paper applications of Raman spectroscopy are discussed in a review.[62]

Raman scattering is a very inefficient process, and so the intensity of bands is very small, unless the Raman process can be enhanced by resonance. Traditionally, most pulp and paper samples could not be analyzed with classical Raman spectroscopy because of the presence of laser-induced fluorescence (LIF) produced by lignin and its degradation products. LIF completely swamps the Raman signal, unless it is minimized when using NIR excitation, in combination with a FTIR interferometer. However, a small residual fluorescence background still remains for samples that are high in lignin content. Nevertheless, the advent of Fourier transform Raman spectroscopy (FTRS) has enabled a number of previously inaccessible applications such as elucidating the mechanism of color reversion[63] to be carried out. FTRS also benefits from the throughput, multiplex and laser-reference advantages. Thanks to the last advantage, the sharp spectral features found in a Raman spectrum are better preserved than with a dispersive instrument since there is no peak broadening during signal averaging. However, since the Raman signal varies as the inverse fourth power of the wavelength, peaks generated in FTRS with NIR excitation are weak and, for samples with high lignin content such as wood,[64] difficult to quantify against the background produced by residual fluorescence. Also, data acquisition is still slow because of the poor noise performance of NIR detectors, as compared to the performance of visible-range detectors used for classical Raman spectroscopy. High-powered lasers are used for circumventing this problem, but these generate a thermal background that is just as high as a fluorescence background.[61] Finally, the cost of FTRS is double that of FTIR instruments, not to mention that excitation lasers need to be replaced every two or three years.

In general, for low-fluorescence samples such as chemimechanical pulp, semi-bleached pulp and paper surfaces, Raman spectroscopy is more advantageous than IR spectroscopy since the amount of moisture in the sample does not interfere with the measurement. On the other hand, FTIRS is the preferred method for the analysis of wood and pulp samples having high levels of lignin and fluorescent chromophores.

8.1.3 Mass Spectrometry

Secondary ion mass spectrometry (SIMS) is the most widely used mass-spectrometry sampling technique for pulp and paper applications.[65] SIMS is a solid-surface analysis technique that destructively samples the topmost layer of the surface by focusing an energetic beam of ions onto a solid-phase sample and secondary ions are sputtered from the surface and analyzed by a mass spectrometer. Even though the penetration depth of these ions is much smaller than 1 μm, the IR-spectrometry counterpart for this technique is the ATR sampling technique. SIMS mainly provides information about the spatial distribution and nature of trace fragments from heavier inorganic species such as the calcium in calcium carbonate and simple organic molecules, whereas ATR provides nondestructive bulk analysis of the uppermost 1 μm of the surface. Ideally, one should

perform a preliminary ATR characterization followed by the more sensitive SIMS. An example is that of dry-strength additive analysis.[65] Whereas ATR can detect strongly absorbing functional groups such as the amide absorption band in PAM, it cannot distinguish between additives having the same amide functional group but different types of aliphatic chains. On the other hand, these aliphatic groups can be readily identified and differentiated by the SIMS technique, thanks to differences in their fragmentation patterns. The technique is particularly useful for measuring the effectiveness of sizing agents. Detter-Hoskin and Busch[65] have discussed pulp and paper applications of the SIMS technique.

Table 7 summarizes some of the advantages and disadvantages of different spectrometric methods being used in pulp and paper applications.

8.2 Business Issues Relating to Process Control Instrumentation

Improvements in IR analyzer technology, combined with advances in digital control and process-control systems, have led to increased use of FTIR spectrometers in closed-loop automated control applications. In the following paragraphs, we will briefly examine some of the most important issues involved in the use of IR analyzers for process control applications.

Key requirements for using an IR analyzer for closed-loop control are repeatability, accuracy, reliability and suitable response time. Because of the need for recording a reference spectrum from time to time, most present-day installations use the slip-stream configuration and a sample conditioning system shown in Figure 21. This enables a single analyzer to sample different types of process streams. However, this introduces a significant dead time, during which the analyzer is not available. This dead time should be no more than one-sixth of the time constant of the process. For pulp and paper operations such as the digester and recausticizing units, this time constant is often measured in hours. Therefore, an effective measurement time of five to ten minutes is sufficient. Nevertheless, the increased availability of multiple in situ fiber-optic probes monitored with a single instrument is making the use of IR analyzers much more attractive for process-control applications, since the dead time is effectively reduced to zero. San Giovanni[66] has reviewed in detail the technical requirements for process-control instrumentation as well as the economic aspects of those requirements.

Table 7 Comparison of FTIR with other spectrometric techniques

Method	Advantages	Drawbacks
FTIR	Nondestructive, on-line, some structural information	Moisture sensitive (solids), matrix interferences
NIR	On-line applications	Overlapping peaks
FTRS	No water interferences	Fluorescence, expensive equipment
MS	Detailed structural information	Destructive, expensive equipment

8.2.1 Purchasing Considerations and Maintenance Requirements

No instrument, whether used for laboratory analysis or for process-control measurements, is free of maintenance requirements. The cost of purchasing a laboratory instrument such as an IR spectrometer is generally measured against the need for increased analytical capability. Such capability will not be generally used every day, but will be available when needed. Purchasing costs are therefore amortized over the projected life of the instrument, whereas maintenance costs are absorbed into operating budgets. Since there is no direct link to benefits, no consideration is given to the amount of time needed to achieve complete return on investment (ROI).

The situation is radically different for a process-control IR analyzer. Compared to other types of instrumentation for process control, an analyzer is relatively expensive, more complex and sensitive to its environment, and requires more regular maintenance by trained personnel. Therefore, capital expenditures leading to an installation must be justified by the benefits generated through continuous use. An example of a cost-benefit analysis is given below.

8.2.2 Benefits and Costs: An Example

The investment cost that is necessary for the purchase of an on-line IR analyzer is discretionary in nature, i.e, directly related to improved profitability. Benefits in pulp and paper mills are usually linked with increased production and/or reduced process and product-quality variability. For example, assuming that a typical mill produces 1000 tons of pulp per day, implementation of a closed-loop digester control system will generate annual savings of between US$500 000 to US$1 000 000, mainly through reduced use of chemicals, more consistent liquor quality, increased production, and lower downgrading costs. If the cost of the analyzer is US$250 000, the ROI is therefore between three to six months. Typical assumptions on improvements to be realized are:

1. reduced white liquor flow (5 to 10%);
2. lowered K number standard deviation (30%), reduced bleach-plant chemical consumption ($1 per K number deviation per ton);

3. drop in off-grade production (25%), penalty for off-grade pulp (US$30 per ton);
4. increased production (1%), marginal profit on additional tonnage (US$100 per ton).

Therefore, even though savings are mill-specific and, quite often proprietary, it is nevertheless possible to give a range of probable savings by making general assumptions on the process improvements made after implementation of a closed-loop control application. Van Fleet[67] has given a very detailed discussion of benefits and costs associated with closed-loop control applications.

ACKNOWLEDGMENTS

The author wishes to thank Ms. Kelly Wadham for her help in scanning and organizing the drawings, and Dr D.W. Francis for his advice on calibration methods.

ABBREVIATIONS AND ACRONYMS

AOTF	Acoustic-optic Tunable Filter
ASAM	Alkaline Sulfite Anthraquinone Methanol
ATR	Attenuated Total Reflectance
BTCA	1,2,3,4-Butanetetracarboxylic Acid
CLS	Classical Least Squares
DRIFTS	Diffuse Reflectance Infrared Fourier Transform Spectroscopy
DTGS	Deuterated Triglycine Sulfate
FFT	Fast Fourier Transform
FTIR	Fourier Transform Infrared
FTIRS	Fourier Transform Infrared Spectroscopy
FTRS	Fourier Transform Raman Spectroscopy
GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HMC	Hydroxymethylcellulose
ILS	Inverse Least Squares
IR	Infrared
IRE	Internal Reflection Element
LIF	Laser-induced Fluorescence
MCT	Mercury Cadmium Telluride
Mid-IR	Mid-infrared
MIR	Multiple Internal Reflectance
MS	Mass Spectrometry
MTBE	Methyl *tert*-Butyl Ether
MWL	Milled Wood Lignins
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
PAM	Polyacrylamide
PAS	Photo-acoustic Spectrometry
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PMA	Poly(Maleic Acid)
PVA	Poly(Vinyl Alcohol)
PVAc	Poly(Vinyl Acetate)
ROI	Return On Investment
SBR	Styrene-butadiene Rubber
SIMS	Secondary Ion Mass Spectrometry
SPE	Solid-phase Extraction
THF	Tetrahydrofuran
UV	Ultraviolet

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction • Pulp and Paper Matrices

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Chemometrics **(Volume 11)**
Multivariate Calibration of Analytical Data • Signal Processing in Analytical Chemistry

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Databases, Infrared • Theory of Infrared Spectroscopy • Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

REFERENCES

1. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, New York, 1986.

2. J.R. Ferraro, L.J. Basile (eds.), *Fourier Transform Infrared Spectroscopy: Applications to Chemical Systems*, Academic Press, New York, Vol. 1, 1978.
3. J.R. Ferraro, L.J. Basile (eds.), *Fourier Transform Infrared Spectroscopy: Applications to Chemical Systems*, Academic Press, New York, Vol. 2, 1979.
4. J.R. Ferraro, L.J. Basile (eds.), *Fourier Transform Infrared Spectroscopy: Applications to Chemical Systems*, Academic Press, New York, Vol. 3, 1982.
5. J.R. Ferraro, K. Krishnan (eds.), *Practical Fourier Transform Infrared Spectroscopy: Industrial and Laboratory Chemical Analysis*, Academic Press, San Diego, CA, 1990.
6. R.W. Jones, 'Infrared Technology', in *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edition, John Wiley & Sons, New York, 379–416, Vol. 14, 1995.
7. K.V. Sarkanen, H.M. Chang, G.G. Allan, 'Species Variation in Lignins: II. Conifers Lignins', *Tappi J.*, **50**(12), 583–590 (1967).
8. C.D. Smith, J.K. Wise, 'Infrared Spectrometric Examination of Paper. I. Techniques and Fiber Composition', *Anal. Chem.*, **39**(14), 1698–1701 (1967).
9. J. Marton, H.E. Sparks, 'Determination of Lignin in Pulp and Paper by Infrared Multiple Internal Reflectance', *Tappi J.*, **50**(7), 363–368 (1967).
10. R.L. Desai, 'Photodegradation of Cellulose – A Surface Effect', *Pulp Paper Can.*, **71**(14), 51–52 (1970).
11. M. Streibl, K. Konečný, M. Mahdalík, 'Chemical Composition of Beech Bark: I Investigation of Lipids.', *Cell. Chem. Technol.*, **3**(6), 653–656 (1969).
12. R.L. Jackson, 'Determination of the Total Hydroxyl Content of Cellulose Esters by Near-infrared Spectroscopy', *Tappi J.*, **51**(12), 560–563 (1968).
13. G.N. Thoen, D.C. Nicholson, 'Infrared Analysis of Kraft Pulping Process Gases', *Tappi J.*, **53**(2), 224–226 (1970).
14. S.A. Berben, J.P. Rademacher, L.O. Sell, D.B. Easty, 'Estimation of Lignin in Wood Pulp by Diffuse Reflectance Fourier Transform Infrared Spectrometry', *Tappi J.*, **70**(11), 129–133 (1987).
15. O. Faix, U. Welkener, R. Patt, 'Method for Controlling the Digestion of Pulp by IR Spectroscopy', US Patent No. 4,743,339, May 10th, 1988.
16. K.M. Sweeney, 'FTIR Microscopy of Pulp and Paper Samples', *Tappi J.*, **72**(2), 171–174 (1989).
17. K. Krishnan, J.R. Ferraro, '5. Techniques Used in Fourier Transform Infrared Spectroscopy', in *Fourier Transform Infrared Spectroscopy: Applications to Chemical Systems*, eds. J.R. Ferraro, L.J. Basile, Academic Press, New York, 149–193, Vol. 3, 1982.
18. M.A. Friese, S. Banerjee, '6. FTIR Spectroscopy', in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, CRC Press, Boca Raton, 119–141, 1995.
19. O. Faix, '4. Characterization in Solid State: 4.1 Fourier Transform Infrared Spectroscopy'; '5. Characterization in Solution: Spectroscopic Methods – 5.2 Fourier Transform Infrared Spectroscopy', in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 89–109, 233–241, 1992.
20. D.M. Haaland, '8. Multivariate Calibration Methods Applied to Quantitative FTIR Analyses', in *Practical Fourier Transform Infrared Spectroscopy: Industrial and Laboratory Chemical Analysis*, eds. J.R. Ferraro, K. Krishnan, Academic Press, San Diego, CA, 395–468, 1990.
21. M.P. Fuller, P.R. Griffiths, 'Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry', *Anal. Chem.*, **50**(13), 1906–1910 (1978).
22. D.T. Kuehl, P.R. Griffiths, 'Microcomputer-controlled Interface between a High Performance Liquid Chromatograph and a Diffuse Reflectance Infrared Fourier Transform Spectrometer', *Anal. Chem.*, **52**(9), 1394–1399 (1980).
23. N.J. Harrick, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, Ossining, NY, 1987.
24. A.J. Sommer, P.L. Lang, B.S. Miller, J.E. Katon, '17. Applications of Molecular Microspectroscopy to Paper Chemistry', in *Infrared Microspectroscopy: Theory and Applications*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, New York, 245–258, 1988.
25. H.J. Humecki (ed.), *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York, 1995.
26. M.-L. Kuo, J.F. McClelland, S. Luo, P.-L. Chien, R.D. Walker, C.-Y. Hse, 'Applications of Infrared Photoacoustic Spectroscopy for Wood Samples', *Wood Fiber Sci.*, **20**(1), 132–145 (1988).
27. D.M. Haaland, E.V. Thomas, 'Partial Least Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information', *Anal. Chem.*, **60**(11), 1193–1202 (1988).
28. J.R. Nault, J.F. Manville, 'Differentiation of Some Canadian Coniferous Woods by Combined Diffuse and Specular Reflectance Fourier Transform Infrared Spectrometry', *Wood Fiber Sci.*, **24**(4), 424–431 (1992).
29. D. Fengel, 'Characterization of Cellulose by Deconvoluting the OH Valency Range in FTIR Spectra', *Holzforschung*, **46**(4), 283–288 (1992).
30. T.P. Schultz, W.G. Glasser, 'Quantitative Structural Analysis of Lignin by Diffuse Reflectance Fourier Transform Infrared Spectrometry', *Holzforschung*, **40**(Suppl.), 37–44 (1986).
31. A.J. Michell, 'Usefulness of Fourier Transform Infrared Difference Spectroscopy for Studying the Reactions of Wood During Pulping', *Cell. Chem. Technol.*, **22**(2), 105–113 (1988).
32. F.G.T. St-Germain, D.G. Gray, 'Photoacoustic Fourier Transform Infrared Spectroscopic Study of Mechanical Pulp Brightening', *J. Wood Chem. Technol.*, **7**(1), 33–50 (1987).
33. F. Kimura, T. Kimura, D.G. Gray, 'FTIR Study of the Effect of Irradiation Wavelength on the Color Reversion of Thermomechanical Pulps', *Holzforschung*, **48**(4), 343–348 (1994).

34. B.B. Sitholé, 'Modern Methods for the Analysis of Extractives from Wood and Pulp: A Review', *Appita*, **45**(4), 260–264 (1992).
35. D.F. Leclerc, M.D. Ouchi, 'Rapid Characterization of Dirt Specks in Kraft Pulp by Infrared Spectrometry', *J. Pulp Pap. Sci.*, **22**(3), J112–J117 (1996).
36. R.S. Shreedhara Murthy, D.E. Leyden, J.K. Grime, 'Determination of Triglycerides on Cotton Fabrics Using Fourier Transform Attenuated Total Reflection Infrared Spectroscopy', *Appl. Spectrosc.*, **39**(6), 1047–1050 (1985).
37. B. Bruce Sitholé, T.N. Tran, L.H. Allen, 'Quantitative Determination of Aluminum Soaps in Pitch Deposits', *Nord. Pulp Pap. Res. J.*, **11**(2), 64–69 (1996).
38. X.-Y. Guo, M. Douek, 'Analysis of Deposits/Stickies from Newsprint Mills Using Recycled Fiber', *J. Pulp Pap. Sci.*, **22**(11), J431–J439 (1996).
39. D.F. Leclerc, R.W. Gooding, J.T. Wearing, 'Off-line Characterization of Polymeric Contaminants in Pulp by Diffuse Reflectance Infrared Spectrometry', *J. Pulp Pap. Sci.*, **18**(6), J202–J206 (1992).
40. P.L. Lang, J.E. Katon, A.S. Bonanno, 'Identification of Dust Particles by Molecular Microspectroscopy', *Appl. Spectrosc.*, **42**(2), 313–317 (1988).
41. A.J. Sommer, J.E. Katon, 'Infrared Microspectroscopy: A High Information Content Microanalytical Technique for the Analysis Contaminants in the Pulp and Paper Making Industry', Proc. 1993 TAPPI Pulping Conf. (Atlanta GA), Book 2, TAPPI Press, Atlanta GA, 543–552, 1993.
42. C.Q. Yang, Y. Xu, D. Wang, 'FTIR Spectroscopic Study of the Polycarboxylic Acids Used for Paper Wet Strength Improvement', *Ind. Eng. Chem. Res.*, **35**(11), 4037–4042 (1996).
43. D. Tatsumi, T. Yamauchi, K. Murakami, 'FT-IR spectroscopy in the evaluation of dry handsheets with an acrylamide-based dry-strength resin', *Nord. Pulp Pap. Res. J.*, **10**(2), 94–97 (1995).
44. D.C. Hassell, E.M. Bowman, 'Process Analytical Chemistry for Spectroscopists', *Appl. Spectrosc.*, **52**(1), 18A–29A (1998).
45. D.F. Leclerc, R.M. Hogikyan, 'Rapid Determination of Effective Alkali and Dead-load Concentrations in Kraft Liquors by Attenuated Total Reflectance Infrared Spectrometry', *J. Pulp Pap. Sci.*, **21**(7), J231–J237 (1995).
46. T.P. Trung, D.F. Leclerc, C. Roberts, D. Osmond, J.A. McIlwain, 'On-line Infrared Technology for Digester and Causticizer Liquor Analysis: Part 1 – Pilot-plant Analyzer Calibration and Performance at a Mill-site Digester', Proc. 1996 CPPA Technical Section Pacific/Western Branches Joint Conference, Jasper, Alberta, Session 3, Paper 5, May 15–18, 1996.
47. R.L. Yeager, 'Better Analysis of Pulping Liquors Aids Operation at Cariboo Pulp Mill', *Pulp Pap.*, **70**(6), 111–112, 115 (1996).
48. J. Bandekar, R. Sethna, M. Kirschner, 'Quantitative Determination of Sulfur Oxide Species in White Liquor by FT-IR', *Appl. Spectrosc.*, **49**(11), 1577–1582 (1995).
49. A.J. Michell, 'Kappa Number Determination in Kraft Pulping by FTIR Spectroscopic Measurements on Spent Liquors', *Tappi J.*, **73**(4), 235–236 (1990).
50. S. Vanchinathan, S. Ananth. J.M. Jarvis, G.A. Krishnagopalan, 'Kraft-liquor Alkali Analysis Using Near-infrared Spectroscopy', *Tappi J.*, **79**(10), 187–191 (1996).
51. D.F. Leclerc, 'Determination of Sodium Sulfide and Sulfidity in Green Liquors and Smelt Solutions' US Patent No. 5,616,214, April 1st, 1997.
52. O. Faix, H.-L. Schubert, R. Patt, 'Continuous Process Control of Pulping by FT-IR Spectroscopy', Proc. 1989 International Wood and Pulping Chemistry Symposium, Raleigh NC, TAPPI Press, Atlanta GA, 1–8, 1989.
53. C.W. Saunders, L.T. Taylor, 'Determination of the Degree of Nitration of Cellulose Nitrates via GPC/FT-IR Using an On-line Flow Cell', *Appl. Spectrosc.*, **45**(5), 900–905 (1991).
54. A.J. Michell, 'An Anomalous Effect in the DRIFTS Spectra of Wood and Paper', *J. Wood Chem. Technol.*, **11**(1), 33–40 (1991).
55. J. Bouchard, M. Douek, 'Structural and Concentration Effects on the Diffuse Reflectance FT-IR Spectra of Cellulose, Lignin and Pulp', *J. Wood Chem. Technol.*, **13**(4), 481–499 (1993).
56. G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, New York, 1978.
57. O. Faix, J.H. Böttcher, 'Determination of Phenolic Hydroxyl Group Contents in Milled Wood Lignins by FT-IR Spectroscopy Applying Partial Least Squares (PLS) and Principal Component Regression (PCR)', *Holzforschung*, **47**(1), 45–49 (1993).
58. S. Backa, A. Brolin, 'Determination of Pulp Characteristics by Diffuse Reflectance FT-IR', *Tappi J.*, **74**(5), 218–226 (1991).
59. J.M. Pope, '7. Near-infrared Spectroscopy of Wood Products', in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, CRC Press, Boca Raton, 142–151, 1995.
60. T.P. Schultz, D.A. Burns, 'Rapid Secondary Analysis of Lignocellulose: Comparison of Near-infrared (NIR) and Fourier Transform Infrared (FT-IR)', *Tappi J.*, **73**(5), 209–212 (1990).
61. K.A. Christensen, E.A. Todd, M.D. Morris, 'Raman Spectroscopy', in *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edition, John Wiley & Sons, New York, 416–430, Vol. 14, 1995.
62. U.P. Agarwal, R.H. Atalla, '8. Raman Spectroscopy', in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, CRC Press, Boca Raton, 152–181, 1995.
63. U.P. Agarwal, 'Assignment of the Photoyellowing Related 1675 cm^{-1} Raman/IR Band and its Implication to the Mechanism of Color Reversion in Mechanical Pulps', Proc. 1997 International Wood and Pulping Chemistry Symposium, Montréal QC, Technical Section, CPPA, Montréal QC, K4: 1–4, 1997.

64. I.R. Lewis, N.W. Daniel, N.C. Chaffin, P.R. Griffiths, 'Raman Spectrometry and Neural Networks for the Classification of Wood Types – 1', *Spectrochim. Acta*, **50A**(11), 1943–1958 (1994).
65. L.D. Detter-Hoskin, K.L. Busch, '10. SIMS: Secondary Ion Mass Spectrometry', in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, CRC Press, Boca Raton, 206–234, 1995.
66. J.P. San Giovanni, 'Process Control', in *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edition, John Wiley & Sons, New York, 129–174, Vol. 20, 1995.
67. S. Van Fleet, 'Digester Optimization and Control', in *Pulp and Paper Manufacture*, eds. M.J. Kocurek, T.M. Grace, B. Leopold, E.W. Malcolm, 3rd edition, Alkaline Pulping, TAPPI/CPPA, Atlanta/Montreal, 181–240, Vol. 5, 1989.

Mechanical Pulps, Ultraviolet/ Visible Spectroscopy of Chromophores in

John A. Schmidt
Paprican, Quebec, Canada

1 Introduction **1**
2 Origin of Ultraviolet/Visible Absorption in Mechanical Pulps **2**
2.1 The Nature of Lignin 2
2.2 Absorption of Substituted Benzenes Related to Lignin 3
2.3 Absorption and Determination of Soluble Lignins 4
3 Ultraviolet/Visible Diffuse Reflectance Spectroscopy **7**
3.1 Kubelka–Munk Theory 7
3.2 Equations for Determination of Absorption and Scattering Coefficients 8
3.3 Brightness and Its Relation to Reflectance Spectra 9
3.4 Instrumentation, Sample Preparation and Calculations 11
3.5 Major Chromophores in Mechanical Pulps 12
3.6 Sample Spectra 14
4 Other Techniques **17**
4.1 Derivative Spectroscopy 17
4.2 Transient Spectroscopy 17
Abbreviations and Acronyms **17**
Related Articles **18**
References **18**

Mechanical pulps are those where the starting wood is rendered into a suspension of fibers by the application of mechanical energy. Most of the lignin component is retained in the fibers by this process. The yellowish coloration of wood fiber is a consequence of the presence of chromophores in the lignin that absorb both visible and ultraviolet (UV) light. In most cases, the fibers must be bleached before they are acceptable for their final use. Lignin-containing fibers are also susceptible to yellowing when exposed to UV light. To study the chemistry of these processes in pulp fibers, a technique capable of detecting chromophores in materials that both absorb and scatter radiation is needed. Diffuse reflectance ultraviolet/visible (UV/VIS) spectroscopy is such a technique. For nonscattering materials, the Beer–Lambert law provides a simple linear relationship between absorbance and the concentration of absorbers. This has been used for the analysis of soluble lignins and pulping liquors. For scattering materials, there is no similar simple linear relationship between reflectance and concentration of absorbers. A single reflectance measurement from an opaque sample can be related to the ratio of absorption to scattering by the Kubelka–Munk remission function, $F(R_\infty) = k/s = (1 - R_\infty)^2/2R_\infty$, *where* R_∞ *is the reflectance of an opaque sample,* k *is the specific absorption coefficient and* s *is the specific scattering coefficient. The specific absorption coefficient* k *is linearly related to absorber concentration. If the scattering coefficient remains constant, the remission function is proportional to chromophore concentration. However, for many samples of interest, such as pulps prepared by different mechanical processes, scattering is not constant. Here, it would be uncertain whether a difference in remission function should be attributed to a change in absorption or a change in scattering. Also, in samples where the chromophores are not homogeneously distributed throughout the sample thickness, such as sheets yellowed by light, the linearity of the remission function no longer holds. In such cases,* k *can be calculated explicitly by taking additional reflectance measurements from thin transmitting samples that are suspended over backgrounds of known reflectance. While spectra of* k *versus* λ *(wavelength) are usually broad and featureless, difference spectra can reveal a surprising amount of detail about the changes in chromophore content.*

1 INTRODUCTION

Higher plants contain three types of biological polymers: celluloses, hemicelluloses, and lignins. Celluloses are linear polymers of the glucose dimer cellobiose, while hemicelluloses are complex polysaccharides containing a variety of carbohydrate building blocks. Lignins are polymers of methoxy-substituted phenylpropane. Papermaking fiber is obtained from plants by disintegrating the woody tissue into an aqueous suspension of its constituent fibers. The fibers are bleached to an appropriate whiteness and are then ''reassembled'' at the paper machine into sheets. The majority of paper produced in the world today originates from wood fiber, with small contributions from nonwood sources such as bamboo, kenaf, sugar-cane bagasse and straw.

There are two major ways to disintegrate plant tissue into fibers. In chemical pulping, lignin is removed by hydrolysis. The dominant process used today is kraft

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

pulping, alkaline hydrolysis catalyzed by sodium sulfide. This process and the subsequent bleaching stages remove not only lignin, but also the hemicellulose component of the fibers. Kraft pulping produces fibers that form a strong sheet and that can be bleached to a high and stable whiteness, at about 45% yield based on the starting wood.

In mechanical pulping (also known as high-yield or ultra high-yield pulping), mechanical energy is applied to disintegrate wood into fibers.[1] The mechanical energy is applied either by grinding or by refining; feeding wood chips into the center of two grooved rotating discs and collecting fibers at the periphery. Mild chemical pretreatments, such as steaming at up to 130 °C, or short digestion with bisulfite or sulfite, may be used to alter the material properties of the wood and assist disintegration into intact fibers. These treatments are not severe enough to dissolve much of the lignin, and the yield of fiber from wood is at least 80%, more commonly 90–95%. Almost all new mechanical pulp installations built since 1975 use refiner pulping techniques.

Mechanical pulps have traditionally been restricted to the manufacture of short-life, low-whiteness products such as newsprint. However, projected fiber shortages and the high capital cost of new kraft mills have made the use of bleached mechanical pulps in higher value paper grades increasingly attractive. Softwood thermomechanical pulp (TMP) can be bleached economically to ISO (International Standardization Organization) brightness of about 80% with alkaline hydrogen peroxide, while 87% is achievable for chemithermomechanical pulp (CTMP) from hardwoods. While acceptable for many uses, this is still less than the 92% ISO brightness possible for fully bleached kraft pulp. The most serious barrier to wider use of mechanical pulps is their tendency to yellow rapidly under UV irradiation. Both of these limitations arise from the presence in the lignin of small concentrations of chromophores, groups that absorb strongly near-UV (300–400 nm) or visible (400–700 nm) light.

Changes in chromophore content that result in perceptible, even severe, changes in color usually involve only a small number of lignin groups. Often, they cannot easily be detected by techniques such as NMR (nuclear magnetic resonance), or even vibrational spectroscopies, especially in the solid state. Thus, diffuse reflectance UV/VIS spectroscopy is often the best method for studying chromophore changes in wood fibers.

2 ORIGIN OF ULTRAVIOLET/VISIBLE ABSORPTION IN MECHANICAL PULPS

Mechanical pulps absorb strongly in the UV region, from 200 nm to greater than 360 nm. These absorptions tail into the visible region and contribute to the characteristic colors of various woods. Chromophores with maxima in the visible region (e.g. quinones, flavones) also exist. Cellulose and hemicellulose contain largely saturated bonds to carbon and therefore contribute very little to this absorption. Lignin, due to its benzenoid structure, is responsible for most of the UV/VIS absorption of mechanical pulps, although there is some contribution from low-molecular mass extractives. This is true even for highly colored woods such as western hemlock, red cedar, etc.

2.1 The Nature of Lignin

The biosynthesis of lignin occurs by radical polymerization of one or more of the propenyl phenol precursors shown in Figure 1. Lignin units are classified according to their degree of methoxy substitution. Those without methoxy substitution are *p*-coumaryl units, those with one *o*-methoxy group are guaiacyl units, and those with two *o*-methoxy groups are syringyl units. Coniferous woods contain almost exclusively guaiacyl units and 1–5% *p*-coumaryl units, while deciduous species contain approximately 60% syringyl units and 40% guaiacyl units. Grasses can contain all three types of lignin unit.

The biosynthesis of lignin is a radical reaction, most often described as a random process leading to a statistical distribution of the possible bonding patterns.[2] However, Davin et al.[3] recently isolated a so-called dirigent protein that couples lignin precursors stereospecifically. This is the first clear evidence that lignin biosynthesis may be more precisely controlled than previously thought.

Unlike most biological and synthetic polymers, lignins contain a variety of interunit bonding patterns. Lignin structure is usually specified by the frequency of the various bonding patterns and functional groups, expressed as a percentage of the total number of phenyl propane, or C_9, units. Some of the more important structures are shown in Figure 2. Side-chain carbons are labeled, moving away from the benzene ring, as α, β and γ. Benzene carbons are numbered, with the carbon bearing the side chain as carbon 1.

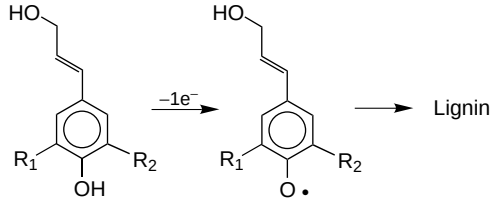

Figure 1 Biosynthesis of lignin occurs by radical polymerization of propenyl phenol precursors. $R_1 = R_2 = H$ are *p*-coumaryl units, $R_1 = H$ and $R_2 = OMe$ are guaiacyl units, and $R_1 = R_2 = OMe$ are syringyl units.

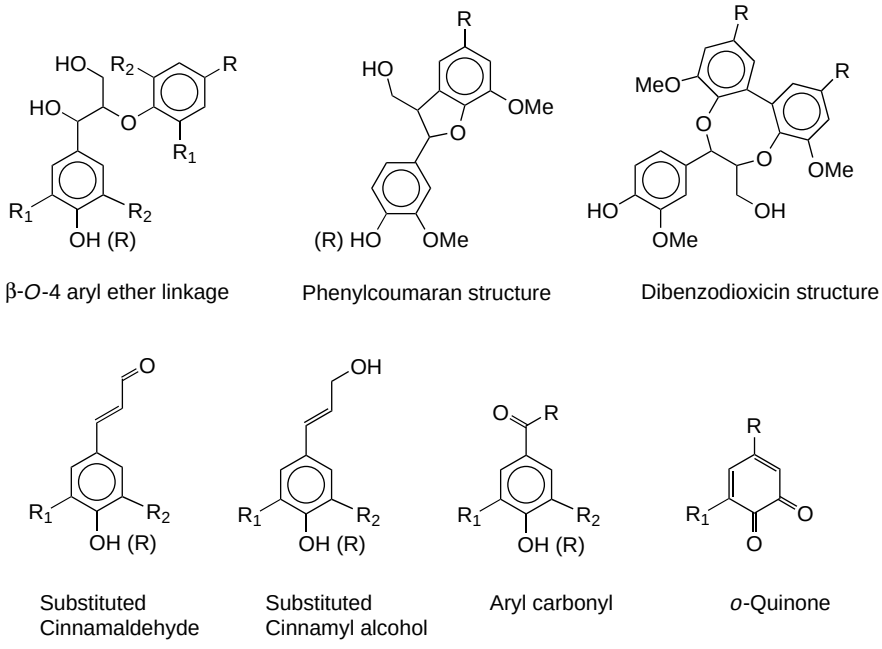

Figure 2 Structures of some important lignin functional groups and interunit linkages.

Approximately 10–15% of lignin units have a free (unetherified) phenolic group, which is one of the most reactive functional groups. The most frequent interunit linkage is the β-*O*-4 structure, where the phenolic oxygen of one unit is bonded to the β carbon of the next. In guaiacyl lignins, the unmethoxylated *ortho* position can bond at a β carbon, giving the fused ring phenylcoumaran structure. Guaiacyl units can also form C–C bonds at the unmethoxylated *ortho* position, yielding biphenyls. Karhunen et al.[4–6] have proposed that biphenyl units are bonded in guaiacyl lignins as dibenzodioxicin structures. Substituted cinnamaldehyde and cinnamyl alcohol structures are end units of the polymer. Cinnamaldehydes have an intense UV absorption at 350–360 nm that contributes to the color of wood. Finally, *o*-quinones, although present in very small amounts (about 0.7%), are also considered as one of the major contributors to the color of wood and mechanical pulps.

2.2 Absorption of Substituted Benzenes Related to Lignin

The maxima, extinction coefficients and designations of the three absorption bands for benzene in solution are shown in Table 1.[7] The band designations ^1B, ^1L_a, and ^1L_b arise from the Platt system of nomenclature for polycyclic aromatic hydrocarbons,[7] of which benzene is the simplest example. The two longest wavelength bands (256 nm, 203 nm) are symmetry forbidden, but have a finite intensity because they are vibronically coupled to the fully allowed transition at 180 nm.

Table 1 Absorption bands of benzene

λ (nm)[a]	ε (L mol^{-1} cm^{-1})[a]	Designation in Platt notation
180	46 000	^1B
203	7 400	^1L_a
256	220	^1L_b

[a] Values from Jaffé and Orchin[7].

Substitution on the benzene ring shifts the absorption maxima of the ^1L_a and ^1L_b transitions by roughly equal amounts and increases their intensity, largely as a consequence of resonance effects. The wavelength shifts are always to the red, regardless of whether the substituent is electron donating (*ortho*, *para* directing) or electron withdrawing (*meta* directing). The ^1B band is also red shifted, often to wavelengths longer than 200 nm.

In disubstituted compounds two types of shifts are observed.[8] If both substituents are of the same type (*o*, *p*-directing or *m*-directing), there is no additive effect due to the second substituent; the band displacements are those observed for the most displaced monosubstituted compound. Thus the ^1L_a band (203 nm in benzene) is shifted by 26.5 nm in benzoic acid, 65 nm in nitrobenzene, and 61 nm in *p*-nitrobenzoic acid.[7]

If one of the substituents is electron donating and the other is electron withdrawing, further wavelength shifts are observed when the second substituent is added,[7–10] as shown in Table 2. For example, in the

Table 2 Wavelength shifts of the ^1L_a and ^1L_b bands in some disubstituted benzenes

	^1L_a		^1L_b		Ref.
	λ_{max} (nm)	$\varepsilon \times 10^{-3}$ (L mol^{-1} cm^{-1})	λ_{max} (nm)	$\varepsilon \times 10^{-3}$ (L mol^{-1} cm^{-1})	
Benzene	203	7.40	256	0.22	7
Benzaldehyde	249	1.14	279[a]	–	8
Phenol	210	6.24	270	1.45	8
p-Hydroxybenzaldehyde	283	16.0	–	–	8
o-Hydroxybenzaldehyde	256	12.6	324	3.4	9
m-Hydroxybenzaldehyde	254	10.1	314	2.58	9

[a] Doub and Vandenbelt[8] did not observe this weak band, probably because of the broadening of the ^1L_a band in hydroxylic solvents. This band is observed in hydrocarbon solvents such as cyclohexane.[10]

Table 3 Absorption maxima of some substituted benzaldehydes in cyclohexane[10]

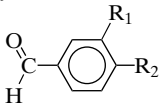

	^1L_a Band λ_{max} (nm)	^1L_b Band λ_{max} (nm)
$R_1 = R_2 = H$	241	279
$R_1 = H$, $R_2 = OMe$	267	281
$R_1 = OMe$, $R_2 = H$	248	306
$R_1 = R_2 = OMe$	272	301

hydroxybenzaldehydes, the ^1L_a band is shifted by 53 nm in the *o*-isomer and 51 nm in the *m*-isomer, almost the sum of the shifts for phenol (7 nm) and benzaldehyde (46 nm). The wavelength shift in *p*-hydroxybenzaldehyde is greater than additive, 80 nm, and is large enough to obscure the ^1L_b band. In the *o*- and *m*-disubstituted compounds, it is the ^1L_b band that is red shifted by an amount greater than the sum of the individual substituent shifts.

The spectra of trisubstituted benzenes with one electron-withdrawing and two electron-donating substituents contain the most red-shifted bands found in the constituent disubstituted compounds.[11] This is illustrated in Table 3 for 3,4-dimethoxybenzaldehyde. The ^1L_a band is the most red-shifted in *p*-methoxybenzaldehyde (267 nm), while the ^1L_b band is the most red shifted in *m*-methoxybenzaldehyde (306 nm). Both of these red-shifted bands appear in 3,4-dimethoxybenzaldehyde.

2.3 Absorption and Determination of Soluble Lignins

Soluble lignins are obtained as byproducts of chemical pulping. They can also be obtained from wood or lignin-containing pulp by acidolysis or enzymic treatment, or by milling followed by solvent extraction (so-called milled-wood lignin). The use of UV/VIS absorbance spectroscopy to quantify such lignins, and as an aid to study their chemistry, will be described briefly in this section.

Figure 3 shows the spectra of two milled-wood lignin samples, one obtained from TMP made from a mixture of eastern North American black spruce and balsam fir, and the other obtained from aspen wood. These spectra are typical of most soluble lignins. Above 300 nm, the absorbance declines continuously, without any clear features. A discrete characteristic maximum occurs at about 280 nm, and a shallow minimum appears at around 260 nm. Below 260 nm, the absorbance increases sharply with decreasing wavelength and the spectrum depends somewhat on the solvent and the specific lignin preparation. In some cases, another maximum is observed at around 230 nm as in Figure 3. In others, this second absorption maximum is blue shifted to about 205 nm.

The absorbance at 280 nm (less often at 205 nm) is the basis of several techniques for the quantitative

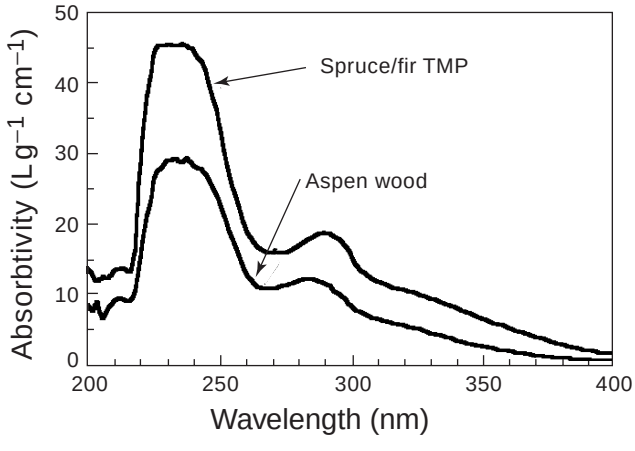

Figure 3 Absorbance spectrum of milled-wood lignin obtained from spruce/balsam fir TMP (0.025 g L^{-1}) and aspen wood (0.05 g L^{-1}) dissolved in 80 : 20 2-methoxyethanol–water.

determination of lignin, according to the Beer–Lambert law, Equation (1).

$$A = DlC \tag{1}$$

A is the absorbance, l the path length in cm, D the absorptivity, and C the concentration. The absorptivity D is analogous to the more familiar extinction coefficient ε. Concentration of lignin solutions is usually expressed in $g\,L^{-1}$, so that the units of D are $L\,g^{-1}\,cm^{-1}$. The absorptivity depends on species and solvent, and varies from 10 to about 26. Summaries of absorptivities are available in reviews by Lin,[12] and Fengel and Wegener.[13]

2.3.1 Determination of Soluble Lignins

Although lignin model compounds are soluble in a wide range of solvents, the choice of solvent for isolated lignins is more limited. Nonetheless, solvent selection is important for quantitative work and care should be taken to assure complete dissolution of the material.

Lignosulfonates are usually soluble in water or water–ethanol mixtures. Lignin from alkaline pulping processes is soluble in water at alkaline pH, but precipitates at neutral or lower pH. A widely used solvent for this lignin and milled-wood lignin is a 2:8 mixture of water with either 2-methoxyethanol or 1,4-dioxane. Dimethylformamide or hexafluoropropanol have also been used for intractable samples.

The following procedure for determination of a soluble lignin sample is adapted from Lin:[12]

1. Weigh 0.5 g of lignin into a 1 L volumetric flask and make up to the mark with solvent. For aqueous solutions, the pH is adjusted to 5 with acetic acid prior to full dilution.
2. Avoid heating or exposure to UV light. If possible, remove oxygen by bubbling a gentle stream of nitrogen (e.g. from a syringe tip) through the solution for 5–10 min. This is particularly important for solutions prepared at alkaline pH, since the phenolate ion reacts readily with oxygen.
3. Dilute the stock solution 10-fold to make the final solution. Determine the absorbance using pure solvent in a matching cuvette as the reference. For quantitative work, the absorbance at 280 nm is usually used. If necessary, the dilution should be adjusted so the absorbance is between 0.2 and 0.8, the most highly linear range for the Beer–Lambert law. A spectrum can be acquired, if desired.
4. Determine the concentration of lignin from the absorbance using the Beer–Lambert law and an appropriate value of the absorptivity. The preferred procedure, of course, would be to construct a calibration graph of absorbance versus concentration, using a known pure sample of the lignin under study.

2.3.2 Determination of Lignin in Wood or Pulp

In addition to determining soluble lignins, UV absorbance has also been used to determine the lignin content of solid plant tissue or pulp. The solid sample is solubilized by an appropriate chemical treatment, usually acid hydrolysis, and the lignin content is determined from the absorbance at 280 nm. Dence[14] has summarized various techniques reported in the literature. The acetyl bromide method has gained general acceptance because of its simplicity, precision and lack of interference from nonlignin materials. A procedure according to Dence[14] is described here:

1. Wood samples are ground in a Wiley mill and the 40–60 mesh fraction is Soxhlet-extracted with 1:2 ethanol–benzene. Pulp samples do not require milling, but are solvent exchanged through acetone and dried at 60 °C and 20 mmHg before the ethanol–benzene extraction.
2. Sample size is generally 5 mg for woods and 10–15 mg for pulp. The samples are placed in 15 mL glass vessels with 5 mL 25% acetyl bromide in acetic acid and 0.2 mL of 70% perchloric acid. The bottles are sealed with PTFE (polytetrafluoroethylene)-coated silicon caps to avoid moisture contamination.
3. The samples are heated in an oven at 70 °C, with swirling at 10-min intervals.
4. The digested samples are transferred to 100-mL flasks containing 10 mL 2 M NaOH and 25 mL acetic acid. After rinsing the bottles, the solution is made up to the mark with acetic acid.
5. The absorbance is measured at 280 nm and the lignin content is determined using Equation (2)

$$\text{Lignin}\% = \frac{100(A_s - A_b)V}{DW} - B \tag{2}$$

A_s is absorbance of the sample, A_b is the absorbance of a blank sample carried through the digestion procedure, V is the solution volume in liters, D is the absorptivity of a lignin standard, and W is the mass of the sample in grams. B is a correction factor for pulp samples; $B = 1.70$ for kraft pulps and 1.38 for sulfite pulps. As in the determination of soluble lignins, it is important to use the appropriate value of the absorptivity. Dence has compiled representative values from the literature.[14] The best procedure, when possible, is to construct a calibration graph for the lignin in question where the lignin content has been determined by some standard procedure such as Klason lignin analysis.

2.3.3 Absorption Difference Spectra: $\Delta\varepsilon$ Methods

Difference spectroscopy has been enormously useful in interpreting the UV/VIS spectra of lignins. In this technique, the absorption spectrum measured before a given chemical treatment is subtracted from that recorded after the treatment (or vice versa). In the simplest case, a compound, *C* (concentration of C), is transformed to a product *P* (concentration of P). At any time during the reaction, these quantities are related by mass balance, Equation (3), where C_0 is the initial concentration of *C*

$$P(t) + C(t) = C_0 \tag{3}$$

The difference in absorbance ΔA is given by Equation (4),

$$\Delta A = l[\varepsilon_P P(t) + \varepsilon_C C(t) - \varepsilon_C C_0] \tag{4}$$

where *l* is the path length through the sample cell, and ε_P and ε_C are the extinction coefficients for *P* and *C*, respectively. Substituting the mass balance for C_0 yields Equation (5).

$$\Delta A = l(\varepsilon_P - \varepsilon_C)P(t) = l\Delta\varepsilon P(t) \tag{5}$$

Note that absorptions by groups that are unaffected by the chemical treatment will cancel in the subtraction.

Lin[12] has described procedures for the measurement of ionization difference spectra. Absorption difference spectra have been used to estimate the content of phenolic groups, carbonyl groups,[15,16] ethylenic double bonds,[16] noncondensed phenolic groups[17] and phenylcoumarans[18,19] in soluble lignins, and to study the photochemistry of lignin solutions.[20,21] Two brief examples are given below to illustrate the technique.

The absorption of a phenolate ion is shifted by at least 10 nm to longer wavelength relative to the corresponding phenol. If electron-withdrawing groups are conjugated with the benzene ring, especially when *para* to the phenol function, the shifts are considerably larger. For

Table 4 Functional groups in spruce milled-wood lignin estimated from absorption difference spectra

Structure	Estimated by alkaline borohydride reduction[15]		Estimated by catalytic hydrogenation[16]	
	λ_{max} (nm)	Proportion of total C_9 units	λ_{max} (nm)	Proportion of total C_9 units
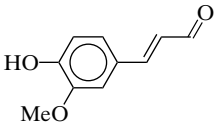	400	<1	–	–
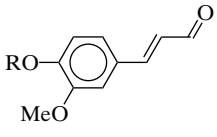	340	3	250 343	3–4
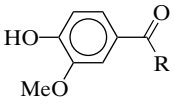	350	<1	–	–
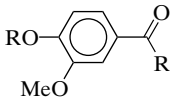	310	5–6	270 280	5
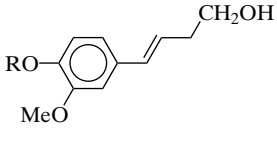	–	–	262–269 295–298	3

example, 3,4-dimethoxyacetophenone and 3-methoxy-4-hydroxyacetophenone both have an absorption peak at 310 nm in neutral solution. In alkaline solution, this peak is unchanged in dimethoxyacetophenone, but is shifted to 350 nm for 3-methoxy-4-hydroxyacetophenone.

Adler and Marton[15] estimated the numbers of phenolic and etherified lignin units containing carbonyl groups in either the α or γ positions of spruce milled-wood lignin, by determining such ionization difference spectra before and after reduction with $NaBH_4$. The positions of the absorption maxima for the various lignin structures and their estimated content in spruce milled-wood lignin are shown in Table 4. The appropriate values of $\Delta\varepsilon$ to use in Equation (5) were determined from experiments on appropriate model compounds. Note that, even though the difference spectra for etherified γ-carbonyls and phenolic α-carbonyls both have maxima in the alkaline-borohydride difference spectrum at 340–350 nm, they could be easily distinguished because the phenolic carbonyls are reduced much more slowly.

In another study, Marton and Adler[16] estimated the content of $\alpha-\beta$ double bonds in spruce milled-wood lignin by comparing the ionization difference spectra before and after catalytic hydrogenation. The results are summarized in Table 4. This experiment also yielded estimates of the amounts of etherified conjugated α and γ carbonyl groups that agree well with those obtained by the borohydride reduction experiment.

3 ULTRAVIOLET/VISIBLE DIFFUSE REFLECTANCE SPECTROSCOPY

Figure 4 shows the diffuse reflectance spectra of unbleached and peroxide-bleached TMP from black spruce. Reflectance is the ratio of the intensity of reflected to incident light, and is thus mathematically analogous to the transmittance of the Beer–Lambert law. Unlike transmittance, however, reflectance is not easily rendered to a quantity proportional to chromophore concentration.

Kubelka and Munk developed the basic theory of diffuse reflectance in the 1930s,[22] and Van den Akker[23] extended it to paper shortly afterwards. However, the arduous calculations required to generate spectra severely limited its use. The development of microprocessor-controlled spectrophotometers and inexpensive computers in the 1990s has made diffuse reflectance measurements much more accessible.

3.1 Kubelka–Munk Theory

Figure 5 shows the essential elements of the Kubelka–Munk model for absorption and scattering of light. By definition, *diffuse* reflectance concerns light that penetrates the sample surface and can be absorbed by chromophores or scattered by internal surfaces. Specular or mirror-like reflectance must be excluded from the measurement. The sample is a composite of layers capable of both diffusely scattering and absorbing radiation, with the absorbers or chromophores *uniformly* distributed throughout the layers. The passage of light through the sample is described in terms of two fluxes: the incident light flux, i, which is normal to the sample surface, and the reflected light flux, j, which travels in the opposite direction. Reflectance is the ratio of the reflected flux to the incident flux, j/i, at the sample surface.

This model leads to two coupled differential equations, which can be solved for various boundary conditions. Details of this derivation are available in Kortüm[24] or Kubelka.[22] Robinson[25] has published a convenient summary of the Kubelka–Munk relations useful for the characterization of paper. For opaque samples (no

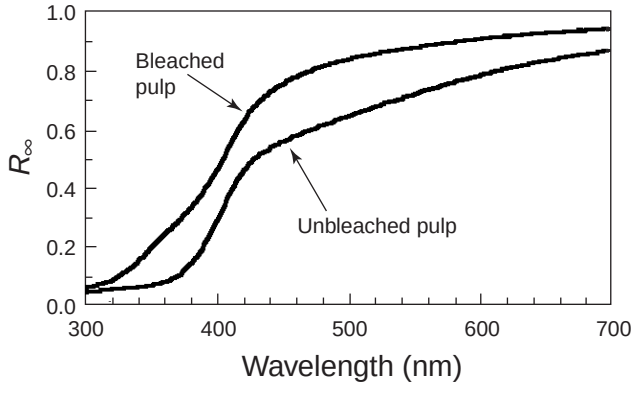

Figure 4 Diffuse reflectance spectra of unbleached and peroxide-bleached TMP from black spruce. R_∞ is the reflectance of an infinitely thick sample.

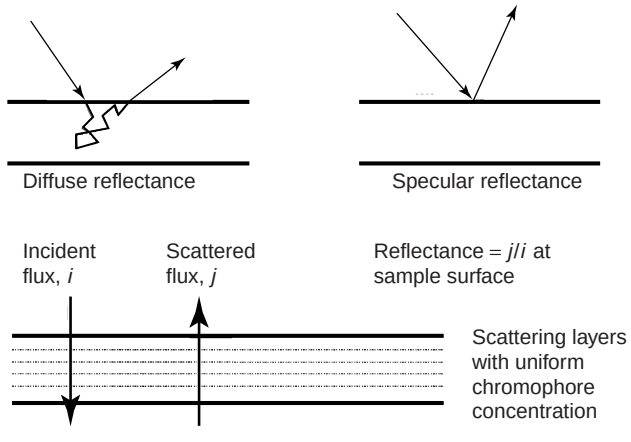

Figure 5 Basic elements of the Kubelka–Munk model for absorption and scattering of light of highly scattering materials.

incident radiation passes through), the solution is the Kubelka–Munk remission function $F(R_\infty)$, Equation (6). R_∞ is the reflectance of an opaque sample, k is the specific absorption coefficient and s is the specific scattering coefficient

$$F(R_\infty) = \frac{k}{s} = \frac{(1 - R_\infty)^2}{2R_\infty} \tag{6}$$

The specific absorption coefficient k is given by Equation (7)

$$k(\lambda) = 2\sum_i \varepsilon_i(\lambda) C_i \tag{7}$$

This quantity is analogous to the absorption coefficient of the Beer–Lambert law, that is the product of extinction coefficient and absorber concentration (the Kubelka–Munk relations reduce to the Beer–Lambert law in the limit of zero scattering). The linearity of Equation (7) is strictly valid in the limit of low absorption, just as the Beer–Lambert absorbance of solutions is linear with absorber concentration only at low absorbance.[26] Koukoulas and Jordan[27] have shown that k is linear with absorber concentration over a range of about $60\,m^2\,kg^{-1}$ (see section 3.2 for comments on the units of k). Polcin and Rapson determined that mechanical pulps obey Equation (7) well in the visible region.[28,29] In the UV region, the strong absorption of lignin causes departure from strict linearity, but useful information regarding relative absorption changes and the position of absorption maxima can still be obtained.

It is clear from Equations (6) and (7) that chromophore concentration is not linear with respect to the observable quantity, R_∞. R_∞ is analogous to transmittance of transparent solutions and is sometimes plotted as $\log(1/R_\infty)$ by analogy with the relationship between absorbance and transmittance in the Beer–Lambert law.[26,30] Unlike the situation in solution, $\log(1/R_\infty)$ is not linear with absorber concentration.

Equation (6) can be rearranged to express R_∞ as a function of k. This is plotted in Figure 6, where s has been set to $30\,m^2\,kg^{-1}$, a typical value of the scattering coefficient for high-yield pulps. The graphical presentation highlights an important consequence of the nonlinearity of the Kubelka–Munk equation: the effect of a change in absorption coefficient on reflectance depends strongly on the initial reflectance. A small change in absorption coefficient will affect a highly reflective sample substantially, but may have a minimal effect on a sample of low reflectance. As a later example will show, this can cause startling differences in the appearance of difference spectra, depending on how the raw data are presented.

The remission function, k/s, is linear with respect to chromophore concentration if the scattering coefficient, s, is constant. If this constraint is met, then differences in the remission function will be proportional to changes in the absorption coefficient and, hence, the chromophore concentration. However, in some samples of interest, e.g. pulps made from different refining processes, the scattering coefficients cannot be assumed to be constant. In this case, the absorption and scattering coefficients must be determined explicitly.

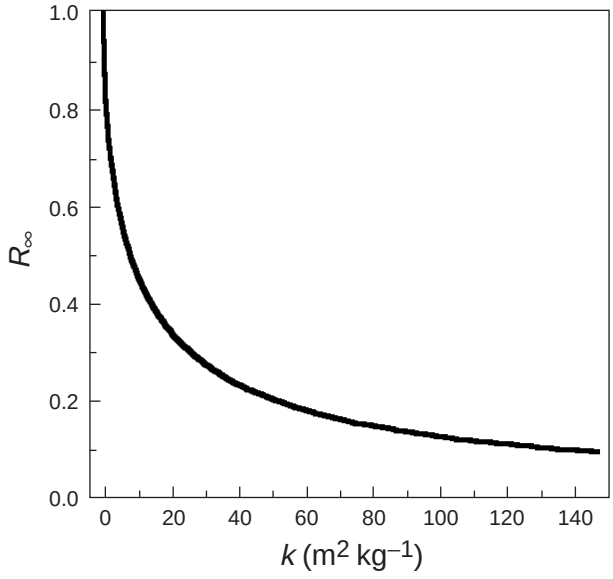

Figure 6 R_∞ as a function of absorption coefficient k, where $s = 30\,m^2\,kg^{-1}$.

3.2 Equations for Determination of Absorption and Scattering Coefficients

While measurements of R_∞ can only be used to calculate the ratio k/s, separate values of k and s can be obtained with an additional reflectance measurement and the use of Equation (8) (another of the equations derived by Kubelka and Munk).

$$s = \frac{1}{b}\left(\frac{R_\infty}{1 - R_\infty^2} \quad \ln \quad R_\infty\left(\frac{1 - R_\infty R_0}{R_\infty - R_0}\right.\right. \tag{8}$$

R_0 is the reflectance of a thin transmitting sample suspended over a black cavity of zero reflectance, b is the basis weight (mass per unit area) of the transmitting sheet. Once the scattering coefficient s has been calculated from Equation (8), the absorption coefficient can be calculated from Equation (6).

A note on units is required here. For most highly scattering materials, the sample thickness is used as the quantity b in Equation (8), which gives m^{-1} as the SI (System International) unit for absorption coefficient. However, paper samples are easily compressed and the sample thickness is not easily controlled. A more precisely defined property is the basis weight, or mass

per unit area, usually specified in $kg\,m^{-2}$. Van den Akker[23] showed that basis weight can be used in the derivation of the Kubelka–Munk equations in place of path length with no loss in mathematical rigor. The SI unit for the absorption coefficient that results is $m^2\,kg^{-1}$. Formally, values calculated this way are called specific absorption coefficients and specific scattering coefficients and lower case letters k and s are used in equations.

A problem of great interest in the chemistry of mechanical pulps is their rapid yellowing on exposure to UV light. This occurs only in the surface layers of a thick paper sheet, so that, after light exposure, the homogeneous distribution of chromophores needed for strict applicability of the Kubelka–Munk equations no longer holds. Absorption coefficients determined using Equations (6) and (8) under these conditions would not be linear with chromophore concentration.

Rigorous values for changes in k under such conditions can still be obtained using samples that are sufficiently thin so that a homogeneous chromophore concentration is maintained. This condition can be tested experimentally by determining that the reflectance is the same on both sides of the sample. In practice, samples with a basis weight in the range of $10–20\,g\,m^{-2}$ are sufficient. A procedure for preparation of such sheets is given in section 3.4.2. The suitability of these low-basis weight sheets is demonstrated visually by the scanning electron micrographs shown in Figure 7. These micrographs show cross-sections of a standard brightness sheet, basis weight $200\,g\,m^{-2}$ and a sample of $10\,g\,m^{-2}$ basis weight. The $10\,g\,m^{-2}$ sheet is, on average, no more than two fiber widths thick. Light from a modestly powered source will not be significantly attenuated on passing through such a sample.

Determination of k requires measurement of both R_0 and R_{wb}, the reflectance of the sample when it is suspended over a highly reflective white backing of reflectance R_w. R_∞ can then be calculated from another of the Kubelka–Munk relations, Equation (9).

$$R_\infty + \frac{1}{R_\infty} = \frac{R_0 - R_{wb} + R_w}{R_w R_0} + R_{wb} \qquad (9)$$

Equation (9) is quadratic in R_∞, but one of the roots of the equation can be eliminated on physical grounds since $0 \le R_\infty \le 1$. The scattering coefficient s is then calculated from Equation (8), and k is calculated from Equation (6).

3.3 Brightness and Its Relation to Reflectance Spectra

Brightness, or diffuse blue reflectance factor, is a standard industrial measurement of the whiteness of white or nearly white paper. It is, in essence, the reflectance of a fairly broad band of blue light from an opaque sample, measured under tightly prescribed conditions. Methods for the measurement of brightness and specifications for brightness measurement devices are published by several standard setting bodies: the Technical Association of the Pulp and Paper Industry (TAPPI) in the United States (T 452), the Pulp and Paper Technical Association of Canada (E.1), the Scandinavian Pulp and Paper Technical Association (SCAN C11:75), and the ISO (ISO 2470). Jordan[31] has given a comprehensive description of brightness measurement to which the reader is referred for detailed information.

Brightness is often referred to loosely as the reflectance measured at 457 nm and in research work such measurements from a scanning spectrophotometer are sometimes reported as equivalent to brightness. This procedure can lead to erroneous results, since the bandwidth of UV/VIS spectrophotometers is much narrower than the standard brightness filter.

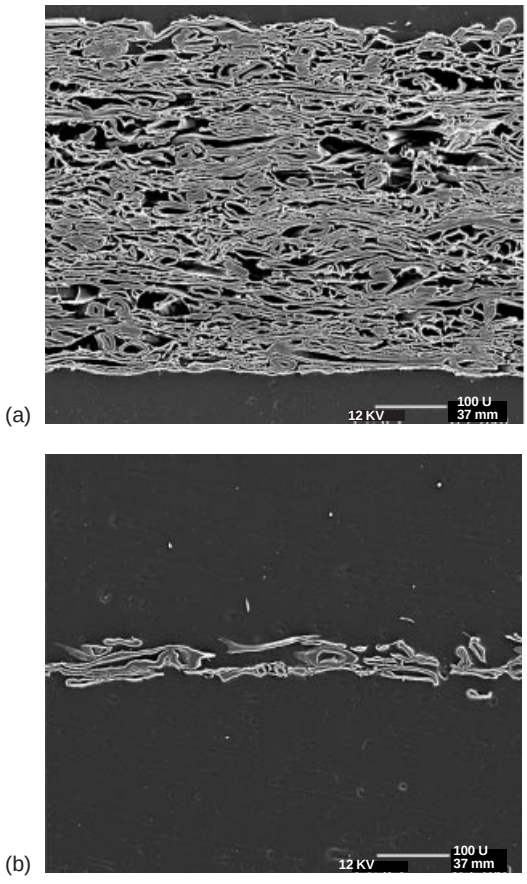

Figure 7 Scanning electron micrographs (170 × magnification) of sheets made from peroxide-bleached TMP. (a) Standard ISO brightness handsheet ($200\,g\,m^{-2}$). (b) Thin sheet of $10\,g\,m^{-2}$ basis weight. The thin sheet is barely two fibers thick, and will attenuate only slightly incident UV light. Photochemical yellowing thus occurs throughout its thickness.

An approximation of ISO brightness can be obtained from a reflectance spectrum by applying Equation (10),

$$\text{ISO brightness} = \sum_{\lambda} f(\lambda) R_{\infty}(\lambda) \qquad (10)$$

where $f(\lambda)$ is the spectral distribution function of a reflectometer equipped for measuring ISO brightness. Table 5 lists values of $f(\lambda)$ specified in ISO standard 2470 for the measurement of ISO brightness. These values include a normalization factor, so that brightness will be expressed as a percentage when R_{∞} is expressed as a fractional value. The response function should be chosen to correspond to the data interval of the spectral measurement.

Values obtained in this way cannot be called ISO brightness for contract verification or other commercial purposes and will not match exactly measurements from standard brightness measuring devices such as those available from Datacolor or Technidyne. This is because the exact requirements of integrating sphere geometry, illuminant and so on are not met. It is, however, an improvement over the procedure of single-wavelength measurements and will be useful to researchers who wish to relate their spectral results to this recognizable commercial scale.

Table 5 Spectral distribution functions, $f(\lambda)$, of a reflectometer equipped to measure brightness according to ISO standard 2470

Wavelength (nm)	Data intervals		
	5 nm	10 nm	20 nm
400	0.107	0.213	0.425
405	0.309		
410	0.715	1.430	
415	1.291		
420	1.942	3.885	7.728
425	2.752		
430	3.680	7.364	
435	4.790		
440	6.145	12.295	24.259
445	7.467		
450	8.801	17.609	
455	10.038		
460	10.668	21.345	42.463
465	10.593		
470	9.462	18.933	
475	7.734		
480	5.665	11.334	22.548
485	3.627		
490	2.166	4.333	
495	1.184		
500	0.597	1.195	2.378
505	0.235		
510	0.032		

As explained below, further caution is required when measuring samples that are known or suspected to contain fluorescent brighteners. Such materials absorb long-wavelength UV light (usually 340–400 nm) and re-emit it as blue fluorescence at $400\,\text{nm} < \lambda < 500\,\text{nm}$. Fluorescent brighteners are often added to paper to enhance whiteness. If samples are illuminated with a light source that is too rich in UV content the apparent brightness will be too high, owing to detection of fluorescent light as well as reflected light.

Standard brightness reflectometers use broad-band illumination of the sample. Reflected light is detected after it passes through a standard filter with the response function of Table 5. Alternatively, instruments of the abridged spectrophotometer type use a monochromator for detection and brightness is calculated by a microprocessor that applies Equation (10). To avoid excess excitation of fluorescence, ISO standard 2470 specifies that the light source in such instruments must be filtered to give a spectral distribution corresponding to the CIE (Commission Internationale d'Eclairage) C standard illuminant.

Scanning spectrophotometers use the opposite configuration of optical components, monochromatic illumination with broad-band detection. The above protocol for limiting UV content cannot be replicated with this type of instrument, so that ISO brightness calculated by applying Equation (10) to reflectance spectra will differ from standard measurements more for fluorescent samples than for nonfluorescent samples. It also follows from the above that reflectance measurements in the UV region could be substantially in error, since reflected and fluorescent light cannot be distinguished. The effect of fluorescence could be controlled with an instrument that uses both monochromatic detection and monochromatic illumination. An absorption spectrometer with this configuration (Bispectral Fluorescence Colorimeter, Labsphere, North Sutton, NH) has recently become commercially available.

3.3.1 Postcolor Number

Giertz[32] first proposed postcolor number, Equation (11), as a way to monitor the chromophores contributing to pulp brightness

$$\text{postcolor } (PC) \text{ number} = [F(R_{\infty,1}) - F(R_{\infty,2})] \times 100 \qquad (11)$$

$F(R_{\infty,i})$ is the remission function of Equation (6), calculated using standard brightness values in place of R_{∞}. It has been used in many studies of chromophore changes that occur during bleaching or reversion of mechanical pulps.

Many standard pulp and paper laboratories do not have easy access to a spectrophotometer, but are equipped with a brightness-measuring device. In such cases, *PC* number is an appealing measurement. It is quick to do, and gives reasonably accurate estimates of the changes in absorption coefficient for samples where the process under study has not seriously affected the scattering coefficient or the homogenous distribution of chromophores throughout the thickness of the sheet.

A difficulty with *PC* number is that it reflects only absorption changes that occur in the region between 400–500 nm (see Table 5). However, many of the chromophores that are responsible for the yellowish cast of mechanical pulps have their strongest absorption in the near-UV region, 300–400 nm. This drawback to Equation (11) is easily remedied if, instead of using a brightness measurement for R_∞, an appropriate spectral measurement can be done.[28] Many modern spectrophotometers have as part of the standard software package the ability to convert reflectance spectra to the corresponding remission functions.

Even when spectra are measured, the problems of using the remission function to compare samples where chromophores are not homogeneously distributed (e.g. light-induced yellowing), or where the samples have different scattering coefficients (e.g. pulps produced by different pulping processes) remain. In such cases, chemical changes can most confidently be studied by explicitly calculating k and s using the techniques described in section 3.2.

3.4 Instrumentation, Sample Preparation and Calculations

3.4.1 Instrumentation

A scanning UV/VIS spectrophotometer with an integrating sphere attachment is required to measure reflectance spectra. Baseline stability is more difficult to maintain when doing reflectance measurements than with transmission measurements, so a double-beam instrument is strongly recommended.

Integrating sphere attachments have been available for many years from the manufacturer of Varian Cary and Perkin-Elmer instruments, and from Labsphere (North Sutton, NH), a specialist manufacturer of reflectance equipment. They are generally available in two size ranges: 6–7.5 cm in diameter and approximately 15 cm in diameter. Efficient functioning of the sphere requires that the sample port occupy only a small percentage of the internal surface area. For paper samples and particularly for the determination of k and s using thin sheets, larger spheres are preferred to lessen the effects of formation inhomogeneity.

As pointed out in section 3.1, specular reflectance must be excluded from measurements of diffuse reflectance. The area of the integrating sphere wall where specular reflectance from the sample falls is fitted with a removable attachment appropriate to the measurement being made. A glossy white plaque is used for measurements of total reflectance, while a black gloss trap is used for diffuse reflectance.

A computer is required to calculate k and s, to average spectra, and to generate difference spectra. Modern spectrophotometers are equipped with a computer interface and have either a native language to manipulate spectra, or can export data to spreadsheets or widely usable formats (e.g. ASCII).

3.4.2 Sample Preparation

Thin sheets that transmit some incident radiation to a background are required for the calculation in Equations (8) and (9). Because of the strong absorption of lignin, situations may arise through normal measurement error where R_0 equals or exceeds R_∞. Either of these conditions will lead to a singularity in Equation (8). Since the absorption of lignin becomes stronger as the wavelength is decreased from 400–300 nm, the thickness of the sample will determine the lower wavelength limit to which spectra can be measured. With practice, sheets of a basis weight as low as $10\,g\,m^{-2}$ can be made with the procedure shown below. Sheets of this basis weight allow reliable measurements to about 300 nm. Absorption coefficients can actually be obtained below 300 nm for sheets of this basis weight, however, the data tend to be noisy and should be interpreted with care.

Sheets of a basis weight as low as $10\,g\,m^{-2}$ can be prepared by the following procedure. Place a 74 μm Teflon™ mesh over the wire of a static sheet-forming machine, taking care that no air bubbles are trapped between the mesh and the wire. Air bubbles will leave holes in the final sheet. Dilute an appropriate amount of pulp to obtain the desired basis weight to the full volume of the sheet machine and then filter onto the Teflon™ mesh. Place the mesh and the sheet on a kraft blotter, mesh side down, and place an 18.5 cm diameter sheet of filter paper on top. Press the sheet manually with three or four passes of a couch roll. The Teflon™ mesh can now be carefully peeled away, leaving the formed sheet on the filter paper. Place the sheet and filter paper sheet-side down on a stainless steel plate and press for 5 min at 0.34 MPa. After air drying, carefully separate the formed sheet from the filter paper with the aid of a spatula. This step is the most difficult and may require some practice. Mechanical pulps bleached with peroxide are easiest to work with, since several washing steps will have removed most of the sticky extractive material, and the exposure

to alkali and peroxide creates anionic groups that aid fiber–fiber bonding. Unbleached pulps require the most care. Cut $5 \times 5\,cm^2$ samples using a scalpel, being careful to avoid any holes that may have formed in the sheet. A rigid plastic template is useful here to obtain sheets of a known uniform area. Weigh the sheets and determine the basis weight.

A potential problem of using sheets of $10–20\,m^2\,kg^{-1}$ basis weight is that formation of fibers into a sheet may be nonuniform over the area probed by a spectrophotometer beam. Averaging the results of three to six samples can minimize the error associated with this nonuniformity. In the author's experience, standard deviations on the absorption coefficients vary from a maximum of 5–10% at wavelengths below 300 nm to 1–2% at wavelengths above 300 nm.

3.4.3 Calculations

Some spectrophotometers report reflectance measurements as percentage values. In Equations (6), (8) and (9) the reflectance values must be fractional. Because of the nonlinear nature of these equations, factors of one hundred will not cancel simply.

Once the raw data (R_0, R_{wb}, R_w, and b) has been collected, the calculation of Equations (9), (8) and (6) (in order) are done to determine the absorption and scattering coefficients. Some spectroscopic software packages have a native programming language that will allow this type of manipulation, as well as averaging and subtraction of spectra. Otherwise, the data can be exported to a spreadsheet program or transmitted as ASCII characters and manipulated by user-written software.

Lignin contains many overlapping UV absorptions and plots of k versus λ will usually show steadily increasing absorption on going from 400–300 nm, with no discrete absorption maxima. Subtraction of the appropriate spectra to generate difference spectra will usually reveal maxima that are characteristic of the chromophore changes occurring (see section 3.6).

3.5 Major Chromophores in Mechanical Pulps

3.5.1 Coniferaldehyde

Coniferaldehyde end groups (second entry in Table 4) are responsible for some of the yellowish tint of softwoods.[33] The longest wavelength maximum for etherified coniferaldehyde model compounds occurs at around 340 nm. Although this is below the visible region, the band is very broad and there is nonzero absorption at wavelengths as high as 430 nm. This is sufficient for this group to contribute to ISO brightness, since the spectral distribution function (see Table 5) has finite values to wavelengths as low as 400 nm.

3.5.2 Biphenyls

Soluble lignins and biphenyls have much shallower minima at 250–270 nm than the monomeric lignin-like phenols. Pew and Connors[34,35] prepared a number of *o*-*o*′-dihydroxybiphenyl derivatives by dehydrogenative coupling of simple guaiacyl phenols, and concluded that much of the difference between the spectra of lignins and monomeric phenols could be accounted for by the presence of a substantial number of biphenyl structures in lignin. The work by Karhunen et al.,[4–6] which reports substantial amounts of biphenyl groups bonded in dibenzodioxicin structures (Figure 2), would appear to be consistent with this view.

3.5.3 Quinones

Quinones have long been considered to be contributors to the color of wood and high-yield pulps.[36–38] *o*-Quinones, in particular, are invoked in many reaction schemes purporting to explain pulp bleaching reactions, as well as reactions proposed to explain darkening under the influence of heat or light.[39]

Generally, *p*- and *o*-quinones each show three characteristic bands, although the weaker bands are sometimes obscured under the stronger ones.[40,41] Table 6 shows the general characteristics of *p*- and *o*-quinones and the band assignments, while Table 7 gives specific data on some quinones of relevance to lignin. The differences between *o*- and *p*-quinones are readily apparent to the eye; *p*-quinones are yellow, while *o*-quinones tend to be dark red.

Imsgard et al.[44] determined that spruce milled-wood lignin contained 0.7% *o*-quinone structures by reducing the quinones to the corresponding catechols and then measuring the visible spectrum of the complex formed between the catechols and ferric ion. This agrees well with other values in more recent literature. Zakis[45] found 0.8% *o*-quinones in spruce milled-wood lignin, based on the ability of quinones to oxidize hydrazine and release a stoichiometric amount of nitrogen. Argyropoulos and

Table 6 Absorption bands of simple benzoquinones[41]

	λ_{max} (nm)	log ε	Assignment
p-Benzoquinones	240–300	3.8–4.5	$\pi \rightarrow \pi^*$
	285–485	2.4–3.5	$\pi \rightarrow \pi^*$
	420–460	1.2–2.3	$n \rightarrow \pi^*$
o-Benzoquinones	250–300	2.6–4.2	$\pi \rightarrow \pi^*$
	370–470	2.8–3.5	$\pi \rightarrow \pi^*$
	500–580	1.4–1.8	$n \rightarrow \pi^*$

Table 7 UV/VIS absorption data for some quinones relevant to lignin (sh = shoulder)

Compound	λ_{max} (nm)	Log ε	Solvent	Refs.
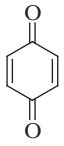	246 288 439	4.42 2.50 1.35	$CHCl_3$	40
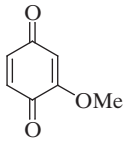	254 357	4.26 3.21	$CHCl_3$	40
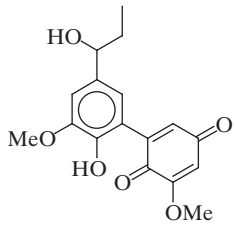	286 328 400 (sh)	4.4×10^3 3.6×10^3 1.4×10^3	95% EtOH	33, 35
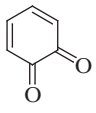	254 368 587	3.04 3.28 1.35	Diethyl ether	40
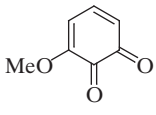	269 465 545 (sh) 575 (sh)	2.66 3.26 1.78 1.56	CH_2Cl_2	40
	370 480 (sh)	3.31 2.8	Water	42
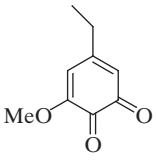	470 570 (sh)	3.19 2.60	$CHCl_3$	43
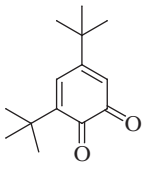	256 402 550	3.5 3.28 1.77	CH_3OH	40

Heitner[46] found 0.7% *o*-quinone groups in black spruce TMP, based on ^{31}P NMR spectra of the trimethyl phosphite–quinone adducts.

The spectra of *o*-quinones with an adjacent methoxy group exhibit an unusual solvent effect. In nonaqueous solvents, these compounds show a prominent maximum at approximately 480 nm. In an aqueous solvent, the absorption at 480 nm is diminished, and a second more intense maximum appears at about 370 nm. Adler et al.[42] attribute this behavior to an equilibrium between the *o*-quinone and a monohydrate adduct.

3.5.4 Metal Complexes

Spruce milled-wood lignin contains, in addition to 0.7% quinone groups, 1% catechol groups.[44] Lignin catechols form complexes with ferric ions with a broad absorption from 470–700 nm, λ_{max} around 560–800 nm, and ε in the

range $1-2 \times 10^3\,L\,mol^{-1}\,cm^{-1}$.[44,47] Ferric and ferrous ions also form complexes with other lignin structures such as phenols and biphenyls. However, the extinction coefficients for these complexes are 10–100 times smaller than those of the catechol complexes. Other common metal ions (Cu^+, Mn^{2+}, Zn^{2+}) have little effect on the color of high-yield pulp fibers.

3.5.5 Hydrogen-bonded Complexes

There are sporadic reports suggesting complex formation between different lignin structures as a source of color. The color of kraft lignin has been attributed both to quinhydrones,[48] donor–acceptor complexes between a quinone and a hydroquinone, and to quinone–phenol complexes.[49,50] Such complexes are held together by hydrogen bonds and can be particularly strong in the solid state. They often have an intense visible absorption due to a charge-transfer transition.[51,52]

The spectra of simple *o*-and *p*-quinones adsorbed on high-yield pulps have been recorded.[53,54] Table 8 compares these spectra with the corresponding spectra in solution. For the *o*-quinones in particular, the solution spectra are significantly different from those recorded when the compound was adsorbed on high-yield pulp. These few observations might be consistent with the formation of complexes between the added quinones and phenol, catechol or other donor groups in the lignin. However, much more systematic work is required before concluding that these observations cannot be explained in terms of the red shifting and band broadening that are usually expected for electronic spectra in the solid state.

3.6 Sample Spectra

The key to effective and efficient use of reflectance spectroscopy when studying chromophore chemistry is to understand the limitations of each of the ways of obtaining and presenting data, and to match this to information that is required from the experiment. Two examples are presented here; others are summarized elsewhere.[55,56]

3.6.1 Difference Spectra for Peroxide Bleaching

Alkaline hydrogen peroxide is an effective bleaching agent for mechanical pulps and much effort has been expended to understand the nature of the chromophores involved in the bleaching reactions.[38] Figure 4 shows the reflectance spectra obtained for opaque samples of black spruce TMP before and after peroxide bleaching. The spectra show clearly that the bleached sample has a significantly higher reflectance at all wavelengths and that for both samples reflectance drops rapidly for $\lambda < 400$ nm. No information with respect to absorption maxima is evident for spectra presented in this way.

Figure 8(a) shows the differences in R_∞ for the two pulps, $\Delta R_\infty = R_{\infty,\text{bleached}} - R_{\infty,\text{unbleached}}$. Two maxima, each corresponding to a reflectance increase of approximately 0.2, appear at 380 nm and 480 nm. This differs sharply from the spectrum in Figure 8(b), the difference in absorption coefficients, $\Delta k = k_{\text{bleached}} - k_{\text{unbleached}}$, for the same two pulps. A strong bleaching with $\lambda_{max} = 360$ nm occurs in the UV. There is a general decrease in the absorption coefficient in the visible region, but there is no clear absorption maximum. While the changes in R_∞

Table 8 Comparison of absorption spectra of quinones in solution and adsorbed on lignin-containing pulp

Compound	λ_{max} (nm) in solution (log ε)	Ref.	λ_{max} adsorbed on lignin-containing pulp	Ref.
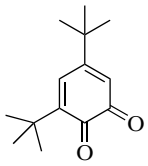	256 (3.50) 402 (3.28) 550 (1.77)	40	440	53
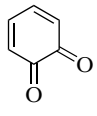	254 (3.04) 368 (3.28) 587 (1.35)	40	400 540	52
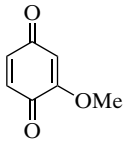	254 (4.26) 357 (3.21)	40	385	53

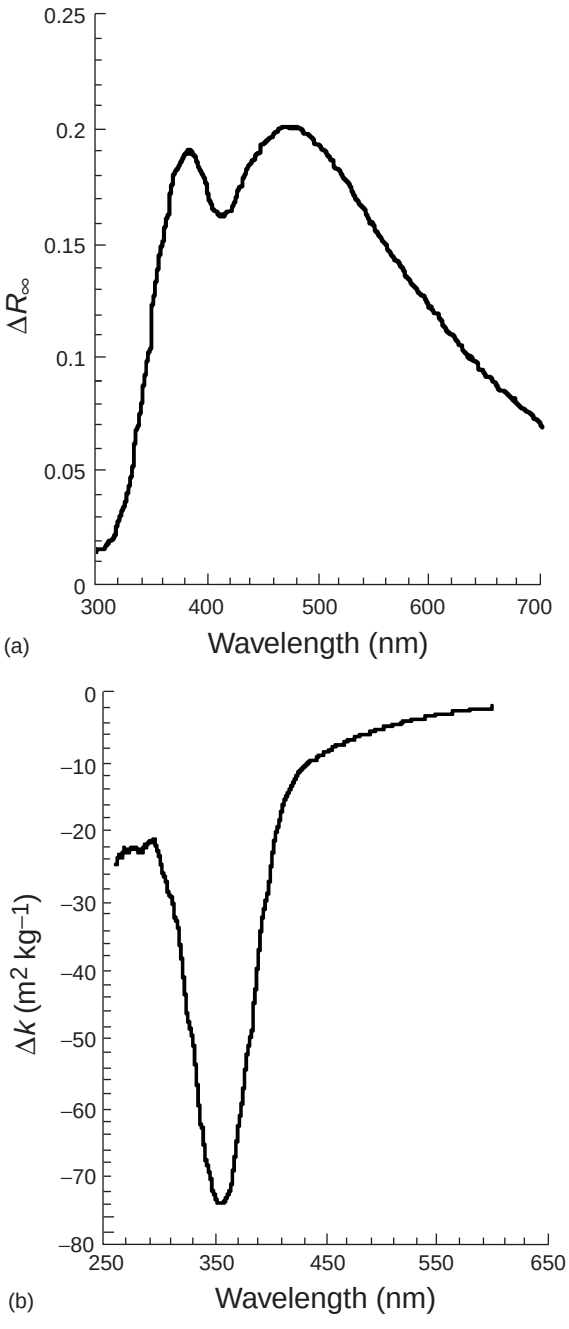

Figure 8 Difference spectra for peroxide bleaching of black spruce TMP: (a) $\Delta R_\infty = R_{\infty,\text{bleached}} - R_{\infty,\text{unbleached}}$; (b) $\Delta k = k_{\text{bleached}} - k_{\text{unbleached}}$.

at 480 nm and 380 nm are almost equal, the changes in k at these two wavelengths differ by about a factor of eight.

The reason for the startling differences between spectra of ΔR_∞ and Δk is the nonlinearity of R_∞ with respect to k, shown earlier in Figure 6. Recall that Figure 6 was computed assuming a typical value of s for mechanical pulps, of 30 m^2 kg^{-1}. At 480 nm, the reflectance of the unbleached pulp is about 0.55, and a decrease in k of 10 m^2 kg^{-1} is sufficient to cause an increase in reflectance of 0.2. At 380 nm, where the initial reflectance is only about 0.12, k must decrease by about 80 m^2 kg^{-1} to achieve a reflectance increase of 0.2.

The two major chromophores removed in peroxide bleaching are coniferaldehyde ($\lambda_{max} = 360$ nm) and *o*-quinones.[38] *p*-Quinones may also be removed, although the evidence for their presence in native lignin is less clear. *o*-Quinones have very broad absorption bands with λ_{max} in the region of 480 nm (see Table 7), although the intensity of this band may be diminished somewhat in the presence of water (section 3.5.3).

The absorption due to coniferaldehyde is quite strong and tails well into the visible region. This probably accounts for the inability to see a discrete peak due to *o*-quinones in the spectrum of Δk; the broad *o*-quinone absorption is probably merged with the tailing coniferaldehyde absorption. Note that the absorption of *o*-quinones, already broad in solution, is even broader where the *o*-quinones have been generated in high-yield pulp. This can be achieved by the reaction of Fremy's salt, potassium nitrosodisulfonate, with lignin phenolic groups.[43,53,57]

The above suggests that spectra of ΔR_∞ more effectively reveal the appearance or disappearance of weak peaks in the visible region. However, one should also note that the position of λ_{max} can differ substantially between these two presentations, at least in the UV region. Thus, while the spectrum of ΔR_∞ shows a maximum at 380 nm, the spectrum of Δk shows a maximum at 360 nm. The absorption maximum observed for Δk is red shifted by about 20 nm from the value observed for etherified coniferaldehyde in solution (Table 4), a value that can reasonably be explained in terms of red shifting in the solid state. Thus, it is more likely to be the correct absorption maximum than that observed for ΔR_∞.

3.6.2 Difference Spectra for Light-induced Yellowing

A second example is taken from work on the light-induced yellowing of mechanical pulps. It illustrates the profound effect that residual amounts of a chromophore can have on absorption difference spectra.

Figure 9 shows the absorption difference spectra observed for UV irradiation of two samples of peroxide-bleached black spruce TMP.[58] The sample in Figure 9(a) was bleached with a single, 3% charge of hydrogen peroxide. This amount of peroxide is typical in the industrial production of mechanical pulps with an ISO brightness in excess of 70%.

Alkaline hydrogen peroxide reacts quantitatively with coniferaldehyde model compounds within minutes.[59] However, several studies have shown that the reaction

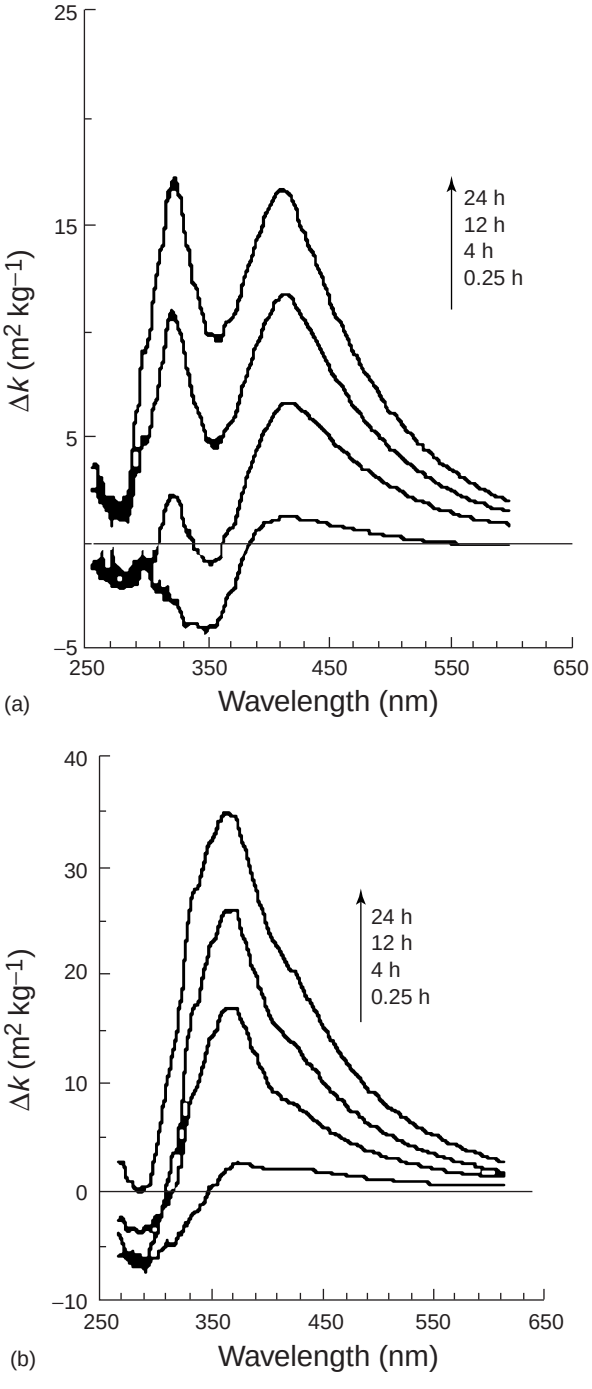

Figure 9 Difference spectra for light-induced yellowing of peroxide-bleached black spruce TMP: (a) TMP bleached with a single 3% charge of peroxide on pulp; (b) TMP bleached with two consecutive charges of 4% peroxide on pulp. The legend indicates the number of hours of irradiation.

in mechanical pulp is slower and less efficient[60,61] and that this amount of peroxide will leave residual coniferaldehyde groups in the pulp. Two consecutive bleaches using 4% peroxide are sufficient to remove all coniferaldehyde groups.[58] The difference spectra for irradiation of black spruce TMP bleached in this way are shown in Figure 9(b).

The marked difference between the two sets of difference spectra in Figure 9 occurs because coniferaldehyde is bleached by UV light.[58] The spectra in Figure 9(b) show a dominant absorption increase at 360 nm, with a shoulder at 416 nm. The spectra in Figure 9(a) show instead absorption maxima at 330 nm and 416 nm, and a minimum at 360 nm. Photobleaching of coniferaldehyde has slowed the overall absorption increase at 360 nm sufficiently to show that there are smaller independent increases in absorption, with λ_{max} at about 330 nm and 416 nm.

These spectra indicate that the irradiation of mechanical pulps produces at least three different chromophores. The peak in the visible region with λ_{max} at about 416 nm is usually attributed to photochemical formation of *o*-quinones. Clear evidence for this proposition is the observation of ^{31}P NMR signals for *o*-quinone–trimethylphosphite adducts.[46] Difference spectra for irradiation of pulps reduced with sodium borohydride indicate that the 330 nm arises from oxidation of benzyl alcohol groups in the pulp to the corresponding aldehydes or ketones.[58] There is less clear evidence for the identity of the chromophore responsible for the strong absorption increase at 360 nm. Recent phloroglucinol staining experiments suggest that *p*-quinones are likely candidates.[62]

3.6.3 Color Tests

Nakano and Meshitsuka have described a number of color-forming reactions for lignin and specific lignin structures.[63] Of these, the phloroglucinol reaction (Wiesner reaction) is the most widely used and will be described briefly here.

Coniferaldehyde residues react with acidic alcoholic solutions of phloroglucinol (1,3,5-trihydroxybenzene) according to the reaction in Figure 10 to give an intensely red complex with $\lambda_{max} = 550$ nm. The reagent is prepared very simply by mixing two volumes of a 2% ethanolic solution of phloroglucinol with one volume of concentrated hydrochloric acid. Since the reagent is not very stable, the phloroglucinol solution is stored in the dark and is mixed with the acid immediately before use. The color reaction is instantaneous upon placing a drop of the solution onto a sample. Because of its sensitivity and ease of use, this reagent is often used by papermakers to test for the presence of mechanical fiber in paper.

Davidson et al.[64] reported the difference spectra of spruce wood sections and bleached softwood mechanical pulp that had been irradiated and then stained with phloroglucinol. The irradiated and stained samples showed an absorption increase with $\lambda_{max} = 550$ nm, which they attributed to photochemical formation of coniferaldehyde. However, in a subsequent publication,[62] they

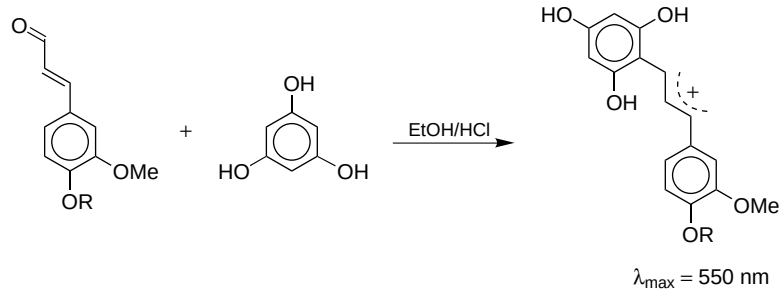

Figure 10 Wiesner reaction of coniferaldehyde residues with phloroglucinol.

found that phloroglucinol also gives a similar color reaction with methoxy-*p*-quinone, which would seem more consistent with the behavior of coniferaldehyde reported in section 3.6.2.

4 OTHER TECHNIQUES

4.1 Derivative Spectroscopy

Derivative spectroscopy is used to resolve more of the fine structure in a spectrum known or suspected to consist of several overlapping bands. The absorbance or reflectance is plotted as the second derivative with respect to wavelength. Spectral features accentuated as higher derivatives are taken, at the expense of increased noise in the spectrum. Lin applied the technique to soluble lignins[65] and found that the absorption due to added phenol in a solution of lignosulfonate can be accurately resolved and quantified. However, a significant limitation of derivative spectroscopy when applied to soluble lignins is that it cannot resolve absorptions in the important region between 300 and 400 nm.

Derivative spectroscopy can also be applied to diffuse reflectance spectra, although only one report has appeared in the literature. Michell et al.[66] were able to resolve some spectral features in the 300–400 nm region in a study of bleaching and yellowing of mechanical pulps from *Eucalyptus regnans*.

4.2 Transient Spectroscopy

Transient spectroscopy is the extension of difference spectroscopy to very short timescales, so that short-lived chemical species can be monitored. The absorbance or reflectance of the sample is measured before and after a brief, intense light pulse strikes the sample, generating a significant population of excited states, and/or reactive intermediates. The excitation source is usually a pulsed laser. In absorbance mode, the data are handled exactly as described in Equations (3–5) of section 2.3.3. The technique has only recently been extended to measure reflectance from opaque samples. More detailed information on transient absorbance measurements is available in a review by Hadel,[67] and information on transient reflectance measurements is available in a review by Wilkinson and Kelly.[68]

The transient absorbance experiment has been applied both to studies of lignin model compounds[69] and to soluble lignin preparations.[70] Guaiacyl-like aromatic ketones (e.g. 3,4-dimethoxyacetophenone) show unusual photophysical behavior. In most acetophenones, the initially formed first excited singlet state is converted to the lowest triplet state with 100% efficiency and all photochemical reactions occur from the triplet. In contrast, triplet-quenching studies indicate that chemistry can occur from both the singlet and triplet states in lignin ketones.[71]

Diffuse reflectance transient spectroscopy has been used to study the reactive intermediates in the photoyellowing of mechanical pulp.[72,73] A transient absorption was observed with λ_{max} at about 450 nm and was identified as a carbonyl triplet state based on two experimental observations, quenching by oxygen and an increase in absorption intensity when phenolic groups were etherified. Both oxygen and phenols quench carbonyl triplet states.

ABBREVIATIONS AND ACRONYMS

CIE	Commission Internationale d'Eclairage
CTMP	Chemithermomechanical Pulp
ISO	International Standardization Organization
NMR	Nuclear Magnetic Resonance
PTFE	Polytetrafluoroethylene
SI	System International
TAPPI	Technical Association of the Pulp and Paper Industry
TMP	Thermomechanical Pulp
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Luminescence in Environmental Analysis

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction

Surfaces **(Volume 10)**
Differential Reflectance Spectroscopy in Analysis of Surfaces

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

REFERENCES

1. R.A. Leask, *Mechanical Pulping*, Pulp and Paper Manufacture, ed. M.J. Kocurek, Joint Textbook Committee of the Paper Industry, Montreal, Vol. 2, 1987.
2. E. Sjöström, *Wood Chemistry: Fundamentals and Applications*, Academic Press, New York, 1993.
3. L.B. Davin, H.-B. Wang, A.L. Crowell, D.L. Bedgar, D.M. Martin, S. Sarkanen, N.G. Lewis, 'Stereoselective Bimolecular Phenoxy Radical Coupling by an Auxiliary (Dirigent) Protein Without an Active Center', *Nature*, **275**, 362–366 (1997).
4. P. Karhunen, P. Rummakko, J. Sipilä, G. Brunow, I. Kilpeläinen, 'The Formation of Dibenzodioxocin Structures by Oxidative Coupling. A Model Reaction for Lignin Biosynthesis', *Tetrahedron Lett.*, **36**(25), 4501–4504 (1995).
5. P. Karhunen, P. Rummakko, J. Sipilä, G. Brunow, I. Kilpeläinen, 'Dibenzodioxocins; A Novel Type of Linkage in Softwood Lignins', *Tetrahedron Lett.*, **36**(1), 169–170 (1995).
6. P. Karhunen, P. Rummakko, A. Pajunen, G. Brunow, 'Synthesis and Crystal Structure Determination of Model Compounds for the Dibenzodioxocine Structure Occurring in Wood Lignins', *J. Chem. Soc., Perkin Trans. 1*, 2303–2308 (1996).
7. H.H. Jaffé, M. Orchin, *Theory and Applications of Ultraviolet Spectroscopy*, John Wiley & Sons, New York, 295–303, 1962.
8. L. Doub, J.M. Vandenbelt, 'The Ultraviolet Absorption Spectra of Simple Unsaturated Compounds. I. Mono- and *p*-Disubstituted Benzene Derivatives', *J. Am. Chem. Soc.*, **69**, 2714–2723 (1947).
9. L. Doub, J.M. Vandenbelt, 'The Ultraviolet Absorption Spectra of Simple Unsaturated Compounds. II. *m*- and *o*-Disubstituted Benzene Derivatives', *J. Am. Chem. Soc.*, **71**, 2414–2420 (1949).
10. R.G.E. Morales, M. Toporowicz, 'Assignation Spectrale des Bandes Électroniques du 3-méthoxy, du 4-méthoxy et du 3,4-diméthoxybenzaldehyde', *Spectrochim. Acta, A*, **37**, 11–15 (1981).
11. L. Doub, J.M. Vandenbelt, 'The Continuity of the Ultraviolet Bands of Benzene with Those of its Derivatives. Application to Certain Trisubstituted Derivatives', *J. Am. Chem. Soc.*, **77**, 4535–4540 (1955).
12. S.Y. Lin, 'Ultraviolet Spectrophotometry', in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 217–232, 1992.
13. D. Fengel, G. Wegener, *Wood: Chemistry, Ultrastructure, Reactions*, Walter de Gruyter, Berlin, 157–161, 1984.
14. C.W. Dence, 'The Determination of Lignin', in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 33–61, 1992.
15. E. Adler, J. Marton, 'Zur Kenntnis der CarbonylGruppen im Lignin. I', *Acta. Chem. Scand.*, **13**, 75–96 (1959).
16. J. Marton, E. Adler, 'Carbonyl Groups in Lignin III. Mild Catalytic Hydrogenation of Björkman Lignin', *Acta Chem. Scand.*, **15**, 370–383 (1961).
17. E. Adler, K. Lundquist, 'Estimation of Uncondensed Phenolic Units in Spruce Lignin', *Acta Chem. Scand.*, **15**(1), 223–224 (1961).
18. E. Adler, K. Lundquist, 'Spectrochemical Estimation of Phenylcoumaran Elements in Lignin', *Acta Chem. Scand.*, **17**, 13–26 (1963).
19. S.I. Falkehag, J. Marton, E. Adler, 'Chromophores in Kraft Lignin', *Advances in Chemistry: Lignin Structure and Reactions*, ACS, Washington, 75–89, Vol. 59, 1966.
20. S.Y. Lin, K.P. Kringstad, 'Photosensitive Groups in Lignin and Lignin Model Compounds', *Tappi*, **53**(4), 658–663 (1970).
21. S.Y. Lin, K.P. Kringstad, 'Some Reactions in the Photoinduced Discoloration of Lignin', *Norsk Skogindustri*, **9**, 252–256 (1971).
22. P. Kubelka, 'New Contributions to the Optics of Intensely Light-Scattering Materials. Part I', *J. Opt. Soc. Am.*, **38**, 448–457 (1948).
23. J.A. Van den Akker, 'Scattering and Absorption of Light in Paper and Other Diffusing Media', *Tappi*, **32**(11), 498–457 (1949).
24. G. Kortüm, *Reflectance Spectroscopy*, Springer-Verlag, Berlin, 107, 1969.
25. J.V. Robinson, 'A Summary of Reflectance Equations for Application of the Kubelka–Munk Theory to Optical Properties of Paper', *Tappi*, **58**(10), 152–153 (1975).
26. R.W. Frei, M.M. Frodyma, V.T. Lieu, 'Diffuse Reflectance Spectroscopy', in *Comprehensive Analytical Chemistry*, ed. G. Svehla, Elsevier, Amsterdam, 263–354, 1975.

27. A.A. Koukoulas, B.D. Jordan, 'Effect of Strong Absorption on the Kubelka–Munk Scattering Coefficient', *J. Pulp Paper Sci.*, **23**(5), J224–J232 (1997).
28. J. Polcin, W.H. Rapson, 'Spectrophotometric Study of Wood Chromophores in situ. I. The Method of Differential $(K/S)_\lambda$ Curves Related to Bleaching and Color Reversion', *Tappi*, **52**(10), 1960–1965 (1969).
29. J. Polcin, W.H. Rapson, 'Effects of Bleaching Agents on the Absorption Spectra of Lignin in Groundwood Pulps: Part II. Oxidative Bleaching', *Pulp Paper Mag. Can.*, **72**(3), 80–91 (1971).
30. H. Hirashima, M. Sumimoto, 'Fundamental Properties of Mechanical Pulp Lignins. I. Visible Light Spectrophotometry and Nitrobenzene Oxidation', *Mokuzai Gakkaishi*, **32**(9), 705–712 (1986).
31. B. Jordan, 'Brightness: Basic Principles and Measurement', in *Pulp Bleaching – Principles and Practice*, eds. C.W. Dence, D.W. Reeve, TAPPI Press, Atlanta, 695–716, 1996.
32. H.W. Giertz, 'Om Massans Eftergulning', *Svensk Popperstidn.*, **48**(13), 317 (1945).
33. J.C. Pew, W.J. Connors, 'Color of Coniferous Lignin', *Tappi*, **54**(2), 245–251 (1971).
34. J.C. Pew, 'Evidence of a Biphenyl Group in Lignin', *J. Org. Chem.*, **28**, 1048–1054 (1963).
35. J.C. Pew, W.J. Connors, 'New Structures from the Enzymic Dehydrogenation of Lignin Model *p*-Hydroxy-α-carbinols', *J. Org. Chem.*, **34**, 580–584 (1969).
36. J. Gierer, 'Basic Principles of Bleaching. Part 2: Anionic Processes', *Holzforschung*, **44**(6), 395–400 (1990).
37. J. Abbot, 'Reactions of *ortho*-Quinones and the Model Compound 4-*t*-Butyl-1,2-benzoquinone in Alkaline Solution', *Res. Chem. Intermed.*, **21**(3–5), 535–562 (1995).
38. C.W. Dence, 'Chemistry of Mechanical Pulp Bleaching', in *Pulp Bleaching – Principles and Practice*, eds. C.W. Dence, D.W. Reeve, TAPPI Press, Atlanta, 161–181, 1996.
39. C. Heitner, 'Light-induced Yellowing of Wood-containing Papers. An Evolution of the Mechanism', in *Photochemistry of Lignocellulosic Materials*, eds. C. Heitner, J.C. Scaiano, American Chemical Society, Washington, DC, 2–25, 1993.
40. S. Berger, A. Rieker, 'Identification and Determination of Quinones', in *The Chemistry of the Quinoid Compounds*, ed. S. Patai, John Wiley & Sons, London, 163–229, 1974.
41. S. Berger, P. Hertl, A. Rieker, 'Physical and Chemical Analysis of Quinones', in *The Chemistry of Quinoid Compounds*, eds. S. Patai, Z. Rappoport, John Wiley & Sons, London, 29–86, Vol. II, 1988.
42. E. Adler, R. Magnusson, B. Berggren, H. Thomelius, 'Periodate Oxidation of Phenols II. Oxidation of 2,6-Dimethoxyphenol and 3-Methoxycatechol. Formation of a 1,2-Naphthoquinone', *Acta Chem. Scand.*, **14**(3), 515–528 (1960).
43. H.-J. Teuber, G. Staiger, 'Reaktionen mit Nitrosodisulfonat, VIII. Mitteil. *ortho*-Benzochinone und Phenazine', *Chem. Ber.*, **89**, 802–827 (1955).
44. F. Imsgard, S.I. Falkehag, K.P. Kringstad, 'On Possible Chromophoric Structures in Spruce Wood', *Tappi*, **54**(10), 1680–1684 (1971).
45. G.F. Zakis, A.A. Melkis, B.Y. Nieberte, 'Methods for Determination of Quinone Carbonyl Groups in Lignins', *Khim. Drev.*, (3), 82–86 (1987).
46. D.S. Argyropoulos, C. Heitner, '^{31}P NMR Spectroscopy in Wood Chemistry. Part IV. Solid State ^{31}P NMR of Trimethyl Phosphite Derivatives of Chromophores and Carboxylic Acids Present in Mechanical Pulps; a Method for the Quantitative Determination of *ortho*-Quinones', *Holzforschung*, **48** Suppl., 112–116 (1994).
47. Y. Ni, Z. Li, A.R.P. van Heiningen, 'Minimization of the Brightness Loss Due to Metal Ions in Process Water for Bleached Mechanical Pulps', *Pulp Paper Can.*, **98**(10), T396–T399 (1997).
48. C. Steelink, T. Reid, G. Tollin, 'On the Nature of the Free-Radical Moiety in Lignin', *J. Am. Chem. Soc.*, **85**, 4048–4049 (1963).
49. G.S. Furman, W.F.W. Lonsky, 'Charge-Transfer Complexes in Kraft Lignin Part 1, Occurrence', *J. Wood Chem. Technol.*, **8**(2), 165–189 (1988).
50. G.S. Furman, W.F.W. Lonsky, 'Charge-Transfer Complexes in Kraft Lignin Part 2: Contribution to Color', *J. Wood Chem. Technol.*, **8**(2), 191–208 (1988).
51. M.C. Depew, J.K.S. Wan, 'Quinhydrones and Semiquinones', in *The Chemistry of Quinoid Compounds*, eds. S. Patai, Z. Rappoport, John Wiley & Sons, London, 963–1018, 1988.
52. R. Foster, M.I. Foreman, 'Quinone Complexes', in *The Chemistry of the Quinoid Compounds*, ed. S. Patai, John Wiley and Sons, London, 257–333, 1974.
53. W. Giust, F. McLellan, P. Whiting, 'Alkali Darkening and Its Similarity to Thermal Reversion', *J. Pulp Paper Sci.*, **17**(3), J73–J79 (1991).
54. G. Leary, D. Giampaolo, 'The Darkening Reactions of TMP and BTMP During Alkaline Peroxide Bleaching', *J. Pulp Paper Sci.*, **25**(4), 141–147 (1999).
55. J. Schmidt, C. Heitner, 'Use of Diffuse Reflectance UV-Visible Spectroscopy to Characterize Chromophores in Wood Fibers', in *Progress in Lignocellulosics Characterization*, ed. D. Argyropoulos, TAPPI Press, Atlanta, 179–199, 1999.
56. J.A. Schmidt, C. Heitner, 'The Use of UV-Visible Diffuse Reflectance Spectroscopy for Chromophore Research on Wood Fibers: A Review', *Tappi*, **76**(2), 117–123 (1993).
57. J.A. Schmidt, F. Kimura, D.G. Gray, 'IR and UV Spectroscopic Study of Borohydride Reduced Mechanical Pulp During Monochromatic and Wide Band Irradiation', *Res. Chem. Intermed.*, **21**(3–5), 287–301 (1995).
58. J.A. Schmidt, C. Heitner, 'Light-induced Yellowing of Mechanical and Ultra-high Yield Pulps. Part 3.

Comparison of Softwood TMP, Softwood CTMP and Aspen CTMP', *J. Wood Chem. Technol.*, **15**(2), 223–245 (1995).
59. G. Gellerstedt, R. Agnemo, 'The Reactions of Lignin with Alkaline Hydrogen Peroxide. Part III. The Oxidation of Conjugated Carbonyl Structures', *Acta Chem. Scand. B*, **34**, 275–280 (1980).
60. G. Gellerstedt, L. Zhang, 'Formation and Reactions of Leucochromophoric Structures in High Yield Pulping', *J. Wood Chem. Technol.*, **12**(4), 387–412 (1992).
61. H. Hirashima, M. Sumimoto, 'Fundamental Properties of Mechanical Pulp Lignins. II. Behaviors of Coniferyl Aldehyde Type Structures in Pulp Lignin', *Mokuzai Gakkaishi*, **33**(1), 31–41 (1987).
62. S. Origgi, V. Trichet, A. Castellan, R.S. Davidson, 'The Action of Phloroglucinol in the Presence of Hydrochloric Acid on Photoaged, Totally Chlorine-free, Bleached Chemical Pulps', *J. Photochem. Photobiol. A: Chem.*, **103**, 159–162 (1997).
63. J. Nakano, G. Meshitsuka, 'The Detection of Lignin', in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer-Verlag, Berlin, 23–32, 1992.
64. R.S. Davidson, H. Choudhury, S. Origgi, A. Castellan, V. Trichet, G. Capretti, 'The Reaction of Phloroglucinol in the Presence of Acid with Lignin-containing Materials', *J. Photochem. Photobiol. A: Chem.*, **91**, 87–93 (1995).
65. S.Y. Lin, 'Derivative Ultraviolet Spectroscopy of Lignin and Lignin Model Compounds – A New Analytical Technique', *Svensk. Papperstidn.*, **85**, R162–R171 (1982).
66. A. Michell, P.J. Nelson, C.W.J. Chin, 'Diffuse Reflectance Spectroscopic Studies of the Bleaching and Yellowing of *Eucalyptus regnans* Cold Soda Pulp', *Appita*, **42**(6), 443–448 (1989).
67. L.M. Hadel, 'Laser Flash Photolysis', in *Handbook of Organic Photochemistry*, ed. J.C. Scaiano, CRC Press, Boca Raton, 279–292, 1989.
68. F. Wilkinson, G. Kelly, 'Diffuse Reflectance Flash Photolysis', in *Handbook of Organic Photochemistry*, ed. J.C. Scaiano, CRC Press, Boca Raton, 293–314, 1989.
69. A.B. Berinstain, M.K. Whittlesley, J.C. Scaiano, 'Laser Techniques in the Study of the Photochemistry of Carbonyl Compounds Containing Ligninlike Moieties', in *Photochemistry of Lignocellulosic Materials*, eds. C. Heitner, J.C. Scaiano, ACS, Washington, DC, 111–121, 1993.
70. M.G. Neumann, A.E.H. Machado, 'The Role of Oxygen in the Photodegradation of Lignin in Solution', *J. Photochem. Photobiol. B: Biology*, **3**, 473–481 (1989).
71. J.A. Schmidt, E. Goldszmidt, C. Heitner, J.C. Scaiano, A.B. Berinstain, L.J. Johnston, 'Photodegradation of α-Guaiacoxyacetoveratrone: Triplet-state Reactivity Induced by Protic Solvents', in *Photochemistry of Lignocellulosic Materials*, eds. C. Heitner, J.C. Scaiano, American Chemical Society, Washington, 122–128, 1993.
72. F. Wilkinson, A. Goodwin, D.R. Worrall, 'Diffuse Reflectance Laser Flash Photolysis of Thermomechanical Pulp', in *Photochemistry of Lignocellulosic Materials*, eds. C. Heitner, J.C. Scaiano, ACS, Washington, DC, 86–98, 1993.
73. J.A. Schmidt, C. Heitner, G.P. Kelly, F. Wilkinson, 'Diffuse-reflectance Laser-flash Photolysis of Mechanical Pulp. Part 1. Detection and Identification of Transient Species in the Photolysis of Thermomechanical Pulp', *J. Pulp Pap. Sci.*, **16**(4), J111–J117 (1990).

Pulp and Paper Matrices

Maurice Douek
Pulp and Paper Research Institute of Canada, Pointe Claire, Canada

1 Introduction 1
2 Applications of Chemical Analysis in the Pulp and Paper Industry 1
3 Test Methods 2
4 Principal Components of Wood 2
4.1 Cellulose 2
4.2 Hemicelluloses 3
4.3 Lignin 3
4.4 Extractives 3
4.5 Inorganics 3
5 Chemical Characterization of Wood Pulp 3
5.1 Determination of Carbohydrates 3
5.2 Determination of Lignin 6
5.3 Determination of Extractives 7
5.4 Determination of Inorganics 11
6 Analysis of Process Liquids 12
6.1 Analysis of Process Liquors 12
6.2 Analysis of Process Waters and Effluents 17
7 Analysis of Paper 19
8 Analysis of Deposits and Contaminants 21
9 Conclusion 24
Abbreviations and Acronyms 25
Related Articles 26
References 26

Chemical analysis plays an important role in pulp and paper research and in the day-to-day operation of pulp and paper mills. It is used in practically every aspect of the papermaking process, including the characterization of raw materials and products, optimization of process parameters, addressing environmental concerns, and diagnosing operation and product quality problems. This article provides an overview of the major types of chemical analyses conducted on pulp and paper matrices. The emphasis is mainly on the principle, significance, and applications of the various analyses. Major areas of application include the analysis of wood pulp, process liquids, paper, and deposits and contaminants. Analytical techniques and instrumentation range from the traditional or standard wet chemical and instrumental methods designed primarily for routine testing in mills to highly sophisticated techniques used mainly for research purposes.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Chemical analysis plays an important role in pulp and paper research and in the day-to-day operation of pulp and paper mills. This article provides an overview of the major types of chemical analyses performed on pulp and paper matrices. The emphasis is mainly on the principle, significance, and applications of the various analyses. Details of analytical techniques and instrumentation can be found in specialized textbooks on the subject[1–3] and in the references provided in each section. It is assumed that the reader is somewhat familiar with the basic terminology and processes involved in pulp and paper manufacture. If this is not the case, the reader is advised to consult some of the many excellent references[4–8] on the subject.

2 APPLICATIONS OF CHEMICAL ANALYSIS IN THE PULP AND PAPER INDUSTRY

Practically every aspect of the pulp and paper-making process relies on the use of certain types of chemical analyses. The majority of these analyses fall into one of the following areas of applications:

- To characterize raw materials and products. The principal raw material is wood. Depending on the particular end-use, several factors such as lignin content, type and amount of carbohydrates, extractives content and composition may have to be determined to assess their effect on pulping conditions and on the quality and properties of the final product. Products may be broadly classified into wood pulp, paper, and board. Chemical analyses are commonly performed on pulp, either bleached or unbleached. These must often be characterized for a number of parameters, including lignin content, viscosity, carbohydrate level, extractives content and composition, and metal profile. Many of the tests performed on pulp are also applicable to paper and board products. Both of these end products, however, are also tested for the level and composition of various additives such as fillers, coatings, and sizing agents.
- To establish and optimize process parameters. During pulp manufacture, analyses of pulping liquors are carried out frequently to monitor the concentration of active components in these liquors and adjust the pulping conditions accordingly. Unbleached pulp

and also bleaching liquors must be characterized in order to control the charge of bleaching chemicals. In the kraft recovery process, pulping and spent liquors are analyzed on a regular basis to determine the level of makeup chemicals that must be used in the system.

- To address environmental concerns. A wide range of analyses are performed regularly on effluents and receiving waters from pulp and paper mills to monitor various parameters such as organic content, level of specific pollutants, color, and toxicity, and to ensure that environmental regulations are being met.
- To diagnose operation and product quality problems. Chemical analysis often provides a valuable tool to diagnose various mill problems. Typical operation problems may include incomplete combustion of black liquors, or black-outs, in kraft recovery boilers, pitch deposits, scaling, or corrosion of process equipment, to mention only a few. Problems involving product quality often involve contamination of the finished product with dirt specks which can downgrade its quality and cause a loss of production. In many cases, application of selected chemical analyses can help to establish the cause of the problem and in devising suitable means of resolving them.

3 TEST METHODS

Based on the foregoing, it is evident that a variety of chemical analyses are required to characterize wood, pulp, paper, and process liquors and to determine their properties for a diversity of end-uses. Accordingly, a large number of test methods have been developed over the years for mill applications and for research purposes. Many of these analyses, particularly those that are commonly used in mills, have been adopted as standards or official test methods by a number of organizations worldwide. A list of major pulp and paper standards organizations is shown in Table 1.

Table 1 Pulp and paper standards organizations

Name of organization	Abbreviation	Country
Technical Association of the Pulp and Paper Industry	TAPPI	USA
Pulp and Paper Technical Association of Canada	PAPTAC	Canada
International Standardization Organization (Technical Committee 6: Paper, Board and Pulps)	ISO	International
Scandinavian Pulp, Paper and Board Testing Committee	SCAN	Denmark, Finland, Norway, Sweden
Deutsche Institut für Normen	DIN	Germany
American Society for Testing and Materials	ASTM	USA
British Paper and Board Makers Association	BPBMA	UK
Australian Pulp and Paper Industry Technical Association	APPITA	Australia and New Zealand

4 PRINCIPAL COMPONENTS OF WOOD

In order to appreciate better the significance of chemical tests conducted on pulp and paper matrices, it is appropriate first to review the chemical composition of the starting raw material, namely wood.

As shown in Table 2, wood is made up of several components, including cellulose and hemicelluloses, which together comprise the hollocellulose or carbohydrate fraction of wood, lignin, extractives, and inorganics.

Table 2 Chemical composition of wood

Component	Approximate concentration range (%, based on dry weight of wood)
Cellulose	40–50
Hemicelluloses	15–20 (softwoods)
	20–35 (hardwoods)
Lignin	23–33 (softwoods)
	16–25 (hardwoods)
Extractives	2–10
Inorganics	0.2–3

4.1 Cellulose

This is the main component of wood, making up 40–50% of its weight. Cellulose consists of a straight or linear polysaccharide chain made up of anhydro glucose units connected with each other by $1\rightarrow4$-β-glucosidic linkages to form a continuous linear chain (Figure 1). Cellulose has a partially crystalline structure with a degree of polymerization (DP) ranging from 7000 to 10 000, depending on its source and method of isolation. One of its important chemical properties is its resistance to alkalis, most acids, and organic solvents. It is, however, readily degraded by concentrated sulfuric acid, and this property is commonly used to separate it from lignin.

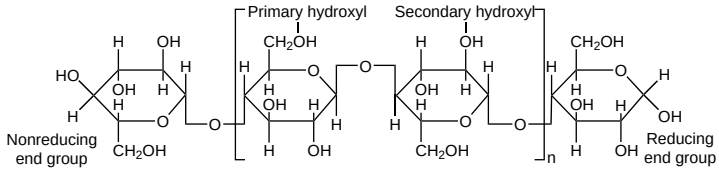

Figure 1 Cellulose structure.

4.2 Hemicelluloses

These are polysaccharides of considerably lower molecular weight than cellulose, making up 15–20% of the weight of coniferous wood (softwoods) and 20–35% of deciduous wood (hardwoods). Unlike cellulose, which is linear, hemicelluloses are made up of branched chains consisting of glucose, xylose, mannose, arabinose, and galactose units. In softwoods, the major hemicellulose sugars are mannose and xylose, whereas in hardwoods, xylose predominates. The DP of hemicelluloses is considerably lower than that of cellulose, typically ranging between 50 and 300. They are largely soluble in aqueous alkali solutions, and are readily degraded by acids.

4.3 Lignin

The content of lignin in softwoods varies between 23 to 33% and in hardwoods between 16 and 25%. Lignin has a highly aromatic branched chain structure, consisting mainly of phenylpropane units with aryl ether and alkyl linkages. Its molecular weight varies greatly, depending on the method of isolation, ranging from 2000 to 10^6. One of the important properties of lignin is that it is essentially insoluble in concentrated sulfuric acid. In its native form, it is almost insoluble in organic solvents.

4.4 Extractives

These are relatively low molecular weight substances generally making up 2–10% of the weight of wood. As their name implies, extractives are mostly soluble in a variety of organic solvents; the most commonly used ones are ethanol–benzene (1:2), dichloromethane, acetone, chloroform, and diethyl ether. The major classes of wood extractives include the terpenes, polyphenols, including flavanones and lignans, and resins such as fatty and resin acids, glycerides, waxes, and sterols. Certain types of extractives, such as tannins, gums, sugars, and starches, are water soluble.

4.5 Inorganics

Inorganics usually make up 0.2–3% of wood, depending on species. Most North American wood species have an ash content below 1%. The major inorganic components of wood are calcium, potassium, and magnesium, present mainly as silicates, phosphates, or sulfates. Several minor elements are also present including, iron, manganese, sodium, chlorine, and aluminum.

Depending on the type and extent of treatment to which the wood is subjected during pulping, bleaching, refining, or other paper-making processes, there is a certain degree of removal, degradation, or chemical modification of some of these wood components. Accordingly, many of the chemical tests performed on wood pulp are aimed at determining, either directly or indirectly, the extent of these changes by measuring the amount and chemical nature of these various components.

The next section examines the various methods and techniques commonly used for the characterization of wood pulp. Subsequent sections will deal with the analysis of process liquids, paper, and deposits and contaminants. Standard PAPTAC and TAPPI test methods are mentioned in the text, as appropriate. Related methods from other standard organizations (e.g. ISO, SCAN, APPITA, and ASTM) can be found in the corresponding PAPTAC or TAPPI methods.

5 CHEMICAL CHARACTERIZATION OF WOOD PULP

The major types of wood pulp manufactured by the pulp and paper industry are listed in Table 3. These can be broadly classified as mechanical, chemimechanical, semichemical, or chemical. As indicated in Table 3, each type can be further subdivided into several classes to reflect more specifically the process used in the mill. Certain tests are only applicable to chemical pulps, while others are intended for all types of pulp.

5.1 Determination of Carbohydrates

Several tests are commonly performed in pulp mills to monitor the degree of cellulose and hemicelluloses removal or degradation during pulping and bleaching. These changes can affect the quality of the pulp, particularly its strength properties. Minimum loss of these components is desirable in order to preserve pulp yield and to achieve optimum papermaking properties. For example, a high retention of hemicelluloses during

Table 3 Major types of wood pulp

Type of pulp	Pulping process
Mechanical	SGW
	RMP
	TMP
Chemimechanical	CMP
	CTMP
Semichemical	NSSC
Chemical	Kraft
	Soda-AQ (NaOH–anthraquinone)
	Polysulfide
	Sulfite (acid, bisulfite, neutral, alkaline)

SGW, stone groundwood; RMP, refiner mechanical pulp; TMP, thermomechanical pulp; CMP, chemimechanical pulp; CTMP, chemithermomechanical pulp; NSSC, neutral sulfite semichemical.

pulping is associated with improved beatability of the pulp, which in turn results in higher tensile strength.

In process control applications, the most widely used methods for evaluating the level or extent of degradation of carbohydrates are based on the determination of pulp viscosity, alkali solubility and pentosan content. These will be considered first. For more detailed carbohydrate profiles and molecular weight distributions, both of which are often required in pulping and bleaching research, several instrumental methods have been developed and will be subsequently described.

5.1.1 Pulp Viscosity

The viscosity of a pulp is related to the DP of the cellulose, and therefore, indicates the extent of its degradation. The more commonly used standard test for measuring pulp viscosity is applicable only to bleached or delignified pulp with a lignin content below 4% (PAPTAC G.24; TAPPI T230). The test involves dissolving a pulp specimen in cupriethylenediamine (CED) – a solvent specific for cellulose – and measuring the viscosity of the resulting solution with a capillary viscometer. If the lignin content is higher than 4%, the sample must first be delignified with a sodium chlorite–acetic acid solution prior to viscosity measurement (PAPTAC G.10U). The average DP of cellulose can be estimated from calculated intrinsic viscosity values (SCAN C15). Comparison between several methods for determining pulp viscosity and conversion between DP and intrinsic viscosity have been reported by Sihtola et al.[9]

5.1.2 Alkali Solubility

Several tests are used to determine the solubility of pulp in various alkali solutions, as a measure of cellulose and hemicellulose removal or degradation. Extraction of wood or pulp in hot 1% sodium hydroxide removes hemicelluloses and degraded cellulose (PAPTAC G.6 and G.7; TAPPI T212). Degraded cellulose represents the portion of the original cellulose which has been broken down into lower molecular weight fragments during pulping or bleaching. This test is applicable to all types of samples, including wood and chemical or mechanical pulp. The following tests are generally applicable only to bleached or delignified pulp.

10 and 18% alkali solubility (PAPTAC G.26; TAPPI T235): in this test, a pulp sample is extracted with a 10 or 18% sodium hydroxide solution at 25 °C and filtered. The dissolved organic matter in the filtrate is determined by oxidation with potassium dichromate, followed by titration of the excess dichromate with ferrous ammonium sulfate. Both degraded cellulose and hemicelluloses are removed in 10% sodium hydroxide; at this alkali concentration, swelling of cellulose fibres is at a maximum. At 18% sodium hydroxide concentration, limited amounts of degraded cellulose are removed, and the dissolved material consists mainly of hemicelluloses. It must be emphasized that the solubility in these alkali solutions does not represent the total hemicelluloses or cellulose content of a pulp; it is only a measure of hemicelluloses or degraded cellulose that can be dissolved under the conditions of this test.

α-, β-, and γ-cellulose (PAPTAC G.29; TAPPI T203): this is somewhat analogous to the alkali solubility test, and is used to determine three carbohydrate fractions of pulp. A pulp sample is treated consecutively with 17.5 and 9.45% sodium hydroxide and filtered. The insoluble fraction after this treatment is called α-cellulose and is a measure of the portion of undegraded, higher molecular weight cellulose in the sample. The precipitate produced after acidification of the filtrate is referred to as β-cellulose and represents the lower molecular weight, degraded cellulose portion. The material remaining soluble after acidification is γ-cellulose and consists mainly of hemicelluloses.

5.1.3 Pentosans

This test provides an indication of hemicellulose retention in pulp during pulping and bleaching. The standard test for determining pentosans is applicable to wood, and to all types of pulps, and consists of hydrolysis of the pentosans in the sample to furfural using boiling hydrochloric acid, followed by distillation and colorimetric determination of furfural with orcinol–ferric chloride (PAPTAC G.12; TAPPI T223). The test, however, provides a better indication of hemicellulose content in hardwood pulp, since softwood hemicelluloses consist of a mixture of pentose and hexose units, whereas hardwood hemicelluloses are mainly made up of pentoses. The use of orcinol to determine total carbohydrates in pulp leachates was also

reported by Willis et al.[10] Standard sugar solutions were used to calibrate the method. Appropriate correction factors were applied to account for the absorbance due to lignin. Provided that independent calibration curves are obtained for each type of sample, this technique could be used as a rapid method for determining total carbohydrates in pulp filtrates and in process liquids.

5.1.4 Detailed Carbohydrate Profiles

The test methods described so far are particularly useful for estimating the proportions of various carbohydrate fractions in pulp. However, a detailed profile of the type and amounts of individual monosaccharides can also be required. A large number of publications has dealt with the determination of monosaccharides in wood and pulp. This is generally accomplished by hydrolysis of the sample, followed by separation of monosaccharides using one of several chromatographic techniques.

The hydrolysis is generally carried out by a two-step procedure with sulfuric acid (TAPPI T249). Primary hydrolysis is performed at 30 °C for 1 h. This is followed by dilution of the sample and a secondary hydrolysis step at 120 °C. The samples are then neutralized before analysis. An alternative hydrolysis procedure involves the use of trifluoroacetic acid (TFA).[11] The advantages of using TFA are shorter reaction times and no need for neutralization of the sample since TFA can be removed by evaporation. The drawback of using TFA is that, for samples with higher lignin contents, there may be incomplete hydrolysis. If longer hydrolysis times are used, sugar degradation may occur and therefore correction values must be applied to the results.

After hydrolysis, the sugars can be analyzed by various chromatographic procedures. Paper chromatography was used for many years,[12] but has now been largely replaced by gas or liquid chromatography. In gas chromatography (GC), the monosaccharides must be derivatized prior to analysis. The more common derivatives are alditol acetates, which are prepared using sodium borohydride as reducing agent and acetic anhydride for acetylation.[13] This lengthy derivatization procedure has been considerably shortened with the use of 1-methylimidazole as a catalyst during acetylation.[14] Monosaccharides can also be converted to trimethylsilyl (TMS) derivatives.[15] Although this is a faster procedure than alditol acetate derivatization, the chromatograms are considerably more complex, since several peaks are produced for stable isomers of each neutral sugar.

High-performance liquid chromatography (HPLC) is also widely used for determination of monosaccharides in wood and pulp hydrolyzates.[16–18] In HPLC analysis, derivatization is not required, and consequently the analysis time is considerably shorter than with GC analysis. Although most of the earlier HPLC analyses were based on refractive index (RI) detection, more recently, the use of high-performance anion-exchange chromatography (HPAEC) columns coupled with pulsed amperometric detection (PAD) has been reported by several workers.[19–21] The major advantages of PAD are its linearity over a much wider range of sugar concentration and higher sensitivity compared with RI detection. Figure 2 shows a typical HPAEC/PAD trace for an unbleached kraft pulp hydrolyzate, obtained using a Carbopac ion-exchange column. The high selectivity achieved with this type of column is clearly evident. In spite of the high glucose level present in the sample, the adjacent galactose peak is well resolved.

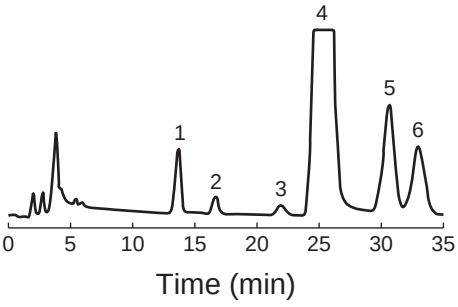

Figure 2 HPAEC/PAD of an unbleached kraft pulp hydrolyzate.[20] Internal standard, 2-deoxy-D-glucose, present at 5 mg L^{-1}. Peaks: 2-deoxy-D-glucose (1), arabinose (2), galactose (3), glucose (4), xylose (5), mannose (6). (Reproduced from J. Sullivan, M. Douek, *J. Chromatogr. A.*, **671**, 339–350 (1994), Copyright 1994, with permission from Elsevier Science.)

The chromatographic methods described above allow the determination of the five major monosaccharides originating from the hollocellulose (cellulose and hemicelluloses) fraction of wood and pulp. In order to determine the amount of cellulose and hemicelluloses separately, it is necessary to estimate the glucose contribution to the hemicelluloses. This contribution can be calculated from a ratio of mannose to glucose of 3 : 1 for softwoods[22] and 2 : 1 for hardwoods.[23] The cellulose fraction can then be calculated by difference between total glucose and that associated with hemicelluloses.

One of the drawbacks of sulfuric acid hydrolysis is that only neutral sugars can be determined by this approach. Sugar acids, such as galacturonic acid, which are formed during kraft pulping, are usually degraded during this treatment. This problem can be overcome through the use of enzymatic hydrolysis.[24] Using a commercial enzyme mixture for hydrolysis, followed by HPAEC/PAD, monosaccharides and sugar acids, including 4-*O*-methylglucuronic and hexenuronic acids, were determined in several types of laboratory and mill kraft pulp. Another method which was found suitable for quantifying sugar acids involves the use of acid

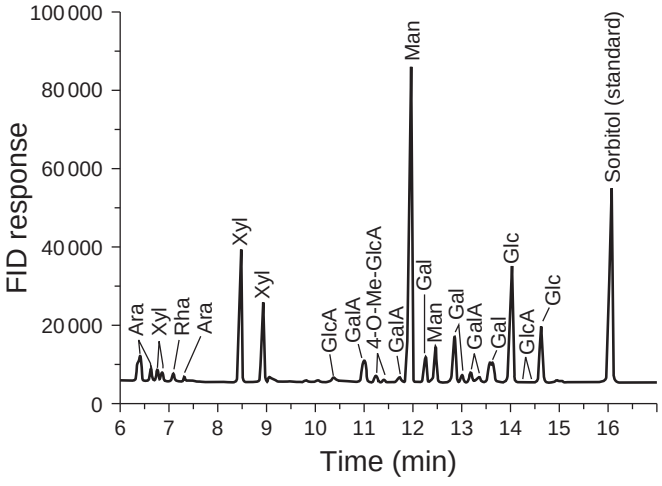

Figure 3 Gas chromatogram of silylated sugar units obtained by acid methanolysis of TMP fibers.[25] Ara = arabinose, Xyl = xylose, Gal = galactose, Glc = glucose, Man = mannose, Rha = rhamnose, GlcA = glucuronic acid, 4-*O*-MeGlcA = 4-*O*-methylglucuronic acid, GalA = galacturonic acid. (Reproduced by permission of Mentor Communications AB from A. Sundberg, K. Sundberg, C. Lollandr, B. Holmbom, *Nordic Pulp Paper Res. J.*, **11**(4), 216–220 (1996).)

methanolysis.[25] Hydrolysis was carried out in a mixture of hydrochloric acid in anhydrous methanol. This was followed by silylation and analysis of sugars by GC. Figure 3 shows a gas chromatogram of silylated sugar units obtained after acid methanolysis of TMP fibers. All sugar components, including uronic acids, are well resolved. The peaks were identified by gas chromatography/mass spectrometry (GC/MS) and by analyzing acid methanolysis products of analytical-grade sugars. One of the advantages of this approach is that cellulose is not degraded significantly during methanolysis. Hence only the noncellulosic glucose, originating from hemicelluloses, can be measured.

More recently, capillary electrophoresis (CE) was used for the separation and quantitation of neutral and acidic mono- and oligosaccharides in wood-derived hemicelluloses.[26] Mono- and oligosaccharides were derivatized by reductive amination with 6-aminoquinoline, and separated as their borate complexes by capillary zone electrophoresis (CZE). High sensitivity was achieved with on-column ultraviolet (UV) detection at 245 nm.

5.1.5 Molecular Weight Distribution of Carbohydrates

One of the common approaches for determining molecular weight distribution of pulp holocellulose is by size-exclusion chromatography (SEC) of their tricarbanilate derivatives.[27,28] Holocelluloses are derivatized with phenyl isocyanate in the presence of pyridine, as a solvent and catalyst, to form the cellulose tricarbanilate (CTC). The CTC is recovered from the pyridine solution by precipitation in a nonsolvent. SEC is then performed with tetrahydrofuran (THF) as eluent using SEC columns of different pore size connected in series. The advantage of this derivatization procedure is that it leads to complete trisubstitution of cellulose with minimum degradation. It was shown, however, that the loss of the low molecular weight fraction could occur during the reprecipitation step,[29] and must be accounted for by performing the SEC analysis on both the soluble and insoluble fractions.

A relatively new technique, matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) mass spectrometry (MS) has also been applied to the determination of the molecular weight distribution of carbohydrates. The principle of this technique was recently reviewed by Harvey.[30] Samples are diluted with a suitable matrix and ionized by irradiation with a laser. Desorbed ions from the vaporized sample are then detected by MS. The technique was used by Dahlman et al.[31] to determine the molecular weight distribution of oligosaccharides in hardwood kraft pulp. The main advantages of this technique are its high sensitivity, wide mass range, simple sample preparation, and no fragmentation of the molecules analyzed.

5.2 Determination of Lignin

This is undoubtedly one of the more important tests on chemical pulp, since the main goal of chemical pulping is to remove the lignin from the wood with minimum degradation of the carbohydrates. Consequently, one of the major objectives of measuring lignin is to assess the effect of a particular pulping or bleaching process on the degree of delignification. Another important purpose is to determine bleaching requirements. The higher the level of residual lignin in unbleached pulp, the greater the amount of bleaching chemicals that must be applied in order to achieve a target brightness.

Comprehensive textbooks and reviews have been written on lignin chemistry and on methods of lignin determination.[1,32,33] Lignin content can be determined by both direct and indirect methods. By far the most widely used direct method is the so-called Klason lignin determination, in which the lignin is measured gravimetrically (PAPTAC G.8 and G.9; TAPPI T222). In this method, a sample of 1–2 g of finely ground and pre-extracted wood or pulp is treated with 72% sulfuric acid at room temperature for 2 h, diluted with water, and then boiled for 4 h to hydrolyze and solubilize the carbohydrates completely. The residue, after filtration, is weighed and is referred to as acid-insoluble or Klason lignin. A modified procedure for smaller samples (200 mg) was reported by Effland.[34] A small portion of lignin

is dissolved during acid hydrolysis. This acid-soluble lignin is determined spectrophotometrically, from the UV absorbance at 205 nm of the filtrate from the Klason lignin determination (PAPTAC G.8 and G.9; TAPPI UM250). The amount of acid-soluble lignin varies from less than 1% in softwoods and kraft pulps to 1–4% in most hardwoods. The acid-soluble lignin content in groundwood and high-yield pulp is similar to that of the original wood. The total lignin content is determined from the sum of the acid-insoluble and acid-soluble lignin in the sample.

Among the most widely used indirect methods are those based on oxidation of lignin with potassium permanganate (PAPTAC G.18; TAPPI T236), or chlorination with sodium hypochlorite (PAPTAC G.32; TAPPI T253). In the former procedure, a pulp sample is treated with an acidified potassium permanganate solution. The excess permanganate is determined by titration with sodium thiosulfate after addition of potassium iodide. The permanganate consumed in the reaction is related linearly to the lignin content and expressed as ''kappa number''.[35] This test is mostly applicable to chemical and semichemical pulps with a yield below 70%. It is not applicable to mechanical or high-yield pulps because the lignin in such pulp is not completely oxidized by the permanganate. For conventional kraft and sulfite pulps with yields below 70%, the relationship between kappa number and Klason lignin is essentially linear and given by Equation (1):[36]

$$\text{kappa number} \times 0.15 = \text{Klason lignin} \qquad (1)$$

A similar principle is used in pulp chlorination, except that in this case the sample is treated with a sodium hypochlorite solution, and the result is expressed as the ''hypo number''.[37] Unlike the kappa number, the hypo number test is applicable to all types and grades of pulp, including high-yield pulps. The relationship with lignin was also found to be linear.[38] Because of their rapidity and simplicity, both tests are particularly well-suited for control use in pulp mills. However, only the kappa number test has gained widespread applicability, primarily in the kraft pulp industry. On-line optical kappa number analyzers, such as the STFI OPTI-kappa, have also been developed.[39] The methods are based on the UV light absorption or reflectance by the sample, and the results obtained with these analyzers were generally found to be in good agreement with the conventional kappa number test.

Another approach for determining lignin in wood and pulp which has received some attention involves dissolution in a mixture of acetyl bromide and acetic acid, followed by measuring the absorbance of the solution at 280 nm.[40] This method was later improved by incorporating a perchloric acid digestion step to ensure faster and complete dissolution of all types of samples.[41] Excellent agreement with the total lignin content of wood and pulps, determined as the sum of the Klason and acid-soluble lignin, was obtained using this modified procedure.

Infrared (IR) spectroscopy has also been used for estimating lignin and carbohydrate content in wood and pulp,[42–44] and for studying the chemical changes in wood during pulping.[45] The near-infrared (NIR) region is particularly suitable for lignin measurements, and several applications of this technique to the determination of lignin and cellulose have been reported.[46–48] The advantage of the NIR technique is that it is considerably faster than wet chemical methods; however, it requires extensive calibrations against other established techniques for each type of wood or pulp species in order to be used reliably. When separate calibrations were used for hardwood and softwood pulps, good agreement was found between lignin determined by NIR spectroscopy and Klason lignin.[47]

5.2.1 Functional Groups in Lignin

A comprehensive review of various methods for the determination of functional groups in lignin has recently been published.[49] Major functional groups in lignin include phenolic hydroxyl, carbonyl, methoxyl, sulfonate, and carboxyl, all of which can have a significant effect in papermaking and on pulp properties. For example, acidic groups, including sulfonates and carboxylates, in combination with metal ions, can influence fiber swelling and various papermaking properties such as beating, drainage, and the retention of cationic chemicals. Both sulfonate and carboxylate groups can be conveniently determined by conductometric titration with sodium hydroxide after conversion of the acid groups to their hydrogen form.[50] Brightness reversion and yellowing of paper are partly due to light-induced oxidation of phenolic structures to quinones. One of the simpler methods for determining phenolic hydroxyl groups is based on the difference in UV absorbance (250 or 300 nm) of phenolic units in neutral and alkaline solutions.[51] The increase in absorption on increasing the pH from 6 to 12 is related to the concentration of phenolic groups in the lignin. The measurements, however, must be calibrated against lignin model compounds.

5.3 Determination of Extractives

So far, we have examined the analysis of the major components of wood and pulp: cellulose, hemicelluloses, and lignin. Extractives are considered to be a minor component, making up 2–10%, and generally less than 5%, of the weight of wood. As indicated previously,

they consist of many different classes of compounds, ranging from water-soluble materials such as tannins, gums, sugars, and starches, to those soluble in various organic solvents. The chemistry and significance of wood extractives have been described in detail,[52] and methods for their analysis have been reviewed by Sitholé.[53] From a pulp and paper-making point of view, the class of extractives that has received the most attention is known as wood resins. By definition, wood resins are compounds soluble in neutral, organic solvents and consist of nonpolar, lipophilic compounds, including fatty and resin acids, sterols, steryl esters, terpenes, and glycerides. A relatively high level of wood resin in pulp is undesirable because it increases the risk of pitch deposition. Certain wood species, such as birch and aspen, contain high levels of neutral wood resin compounds, mainly steryl esters, which are particularly difficult to remove during pulping and can be a serious source of deposition on paper-making equipment. In addition, during pulp bleaching with chlorine compounds, wood resin compounds such as fatty and resin acids can be chlorinated and contribute to the total organic halogen content in the effluents.

In order to assess the propensity of wood or pulp to give rise to these problems, it is important to determine both the level and the types of extractives present. Several solvents are specified in various standard methods (PAPTAC G.13 and G.20; TAPPI T204). The use of 1:2 ethanol–benzene mixture has now been practically eliminated because of the carcinogenic hazard associated with the use of benzene. Halogenated solvents, such as dichloromethane and chloroform, are also being discouraged, because of health hazards. For these reasons, the most recent versions of the PAPTAC standards specify the use of acetone for solvent extraction (PAPTAC G.13 and G.20). In general, extraction with dichloromethane gives lower amounts of extractives in both wood and pulp compared with either acetone or ethanol–benzene. On the other hand, the acetone extractives content in wood is somewhat lower than that obtained with ethanol–benzene. In the case of pulp, however, the acetone extractives content is in most cases comparable to that determined with ethanol–benzene because the compounds originally present in wood that are soluble in ethanol–benzene but not in acetone, such as polyphenols and low molecular weight carbohydrates, are largely removed during the pulping and washing processes.

A Soxhlet-type extraction apparatus is traditionally used for wood and pulp; however, more recently, a rapid extraction unit, the Soxtec System HT® (Tecator, Fisher Scientific) was evaluated and the results compared to those obtained with the traditional Soxhlet apparatus.[54] A schematic diagram of the Soxtec extraction system is shown in Figure 4(a–c). Extraction is carried out with boiling solvent instead of condensed solvent as is the case with Soxhlet extraction. The main advantages of the Soxtec system are much shorter extraction times and recovery of the major portion of the solvent. Efficient rinsing of the sample after extraction must be carried out either by double extraction or extended rinsing. Good agreement was observed between the Soxhlet and double-rinsed Soxtec extraction with acetone on various samples of wood and pulp.[54] The relative amounts of various components were also essentially the same by both methods for the acetone extraction of aspen wood.

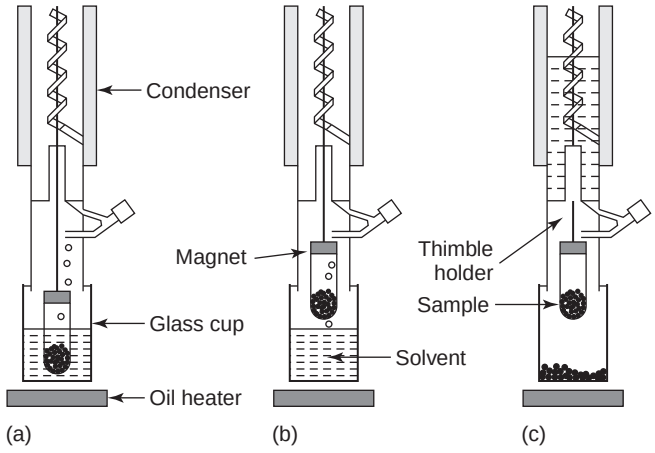

Figure 4 Soxtec extraction system,[54] showing (a) extraction with boiling solvent, (b) sample rinsing with condensed solvent, and (c) solvent recovery. (Reproduced by permission of TAPPI Press from B.B. Sitholé, P. Vollstaedt, L.H. Allen, *TAPPI J.*, **74**(11), 187–191 (1991).)

More rapid and effective means of extraction than those described above are also being explored. Supercritical fluid extraction (SFE) is a relatively new technique which involves the use of a fluid above its critical temperature and pressure. Under these conditions, the fluid behaves like a gas in terms of its extracting ability. As reported by Sitholé in his review of methods for analysis of extracts in wood and pulp,[53] there have been a limited number of publications dealing with SFE of wood and pulps, and percent recoveries are considerably different from those obtained by Soxhlet extraction. More recently, Sequeira and Taylor[55] reported on the use of supercritical fluid carbon dioxide (SF-CO_2) and SF-CO_2 modified with 2% methanol for the extraction of a CTMP and a bleached kraft pulp. The SFE results were comparable to those obtained by Soxhlet extraction with ethyl acetate. Although SFE is faster than conventional liquid extraction, with typical average extraction time of 30 min to 2 h, compared with 1–4 h for Soxtec and 4–48 h for Soxhlet, suitable conditions must still be established before the technique can be widely used on pulp and paper samples.

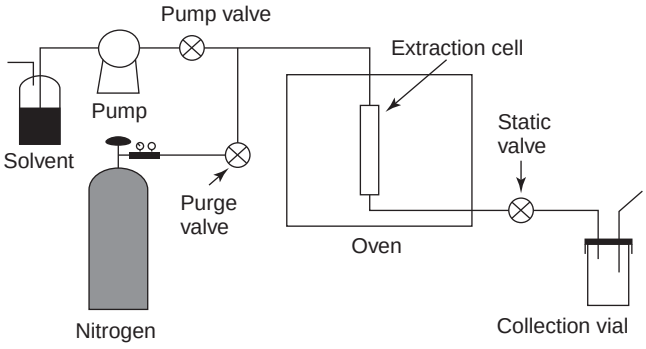

Figure 5 Schematic of ASE™ system.[56] (Reprinted with permission from B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, *Anal. Chem.*, **68**(6), 1033–1039 (1996), Copyright 1996 American Chemical Society.)

Another potentially attractive extraction technique is referred to as accelerated solvent extraction (ASE™). It is based on the use of organic solvents at temperatures and pressures above their boiling point.[56] The extraction rate is considerably increased as a result of increased solubility of extracts, improved mass transfer of the solvent, and disruption of surface equilibria which occur at high temperatures and pressures. A schematic of the ASE™ system is shown in Figure 5. The solvent is pumped into a stainless-steel extraction cell containing the sample. The cell is heated in an oven at elevated temperature (50–200 °C) and pressure (500–3000 psi). The sample extract is then purged with nitrogen into a sealed collection vial and weighed. Typical extraction times are 12–18 min per sample, and solvent consumption is also considerably reduced to 15–40 mL per sample, compared with 50–100 mL for Soxtec and 200–500 mL for Soxhlet. Recoveries of various compounds from reference materials,[56] soil samples,[57] and pesticides and herbicides[58] with ASE™ were in excellent agreement to those obtained by Soxhlet extraction. The application of ASE™ to the determination of dioxins and furans in environmental samples has also been described.[59] The technique is being evaluated by several laboratories for potential applications to wood, pulp, and other samples from the pulp and paper industry.

5.3.1 Detailed Analysis of Extractives

Analysis of extractives in wood and pulp involves preliminary solvent extraction of the sample, followed by fractionation and chromatographic analysis of the extracts. Capillary GC is commonly used for these analyses. A comprehensive scheme for the separation and analysis of extractives in pulp was proposed by Ekman and Holmbom (Figure 6).[60] Lipophilic and polar extractives were successively extracted with dichloromethane and acetone–water (9 : 1). A portion of the lipophilic fraction

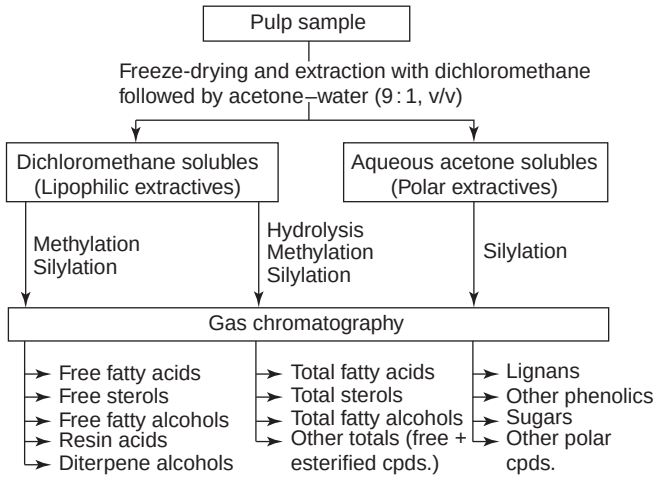

Figure 6 Scheme for the separation and analysis of extractives in pulp.[60] (Reproduced by permission of Mentor Communications AB from R. Ekman, B. Holmbom, *Nordic Pulp Paper Res. J.*, **1**, 16–24 (1989).)

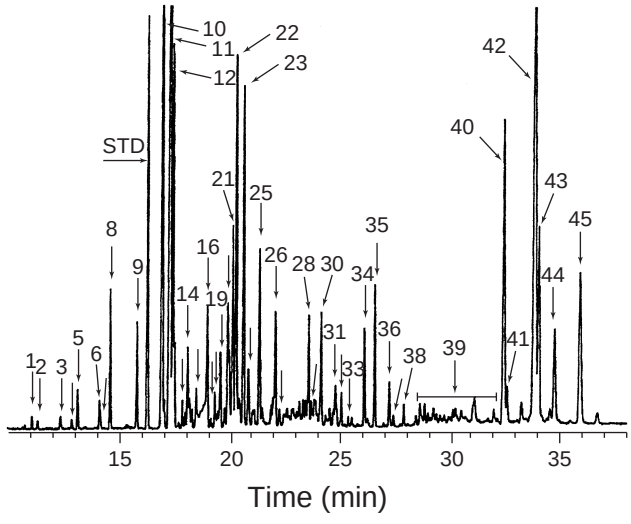

Figure 7 Total ion chromatogram (TIC) from MS detection of the lipophilic extractives in a pulp sample.[60] Components are identified in Table 4. GC conditions: capillary column, HP-1 (12 m × 0.20 mm i.d., film thickness 0.33 μm); column oven, heated from 120 °C at 6 °C min^{-1} to 290 °C; carrier gas, He (0.8 mL min^{-1}, splitting ratio 1 : 15). (Reproduced by permission of Mentor Communications AB from R. Ekman, B. Holmbom, *Nordic Pulp Paper Res. J.*, **1**, 16–24 (1989).)

was hydrolyzed to determine total (free and esterified) extractives. Both hydrolyzed and unhydrolyzed fractions were then derivatized by methylation and silylation. Analysis was carried out by capillary GC combined with MS. Figure 7 shows a TIC of lipophilic extractives in a pulp sample. Identification of the various compounds is shown in Table 4.

Table 4 Identification of compounds corresponding to the peaks in Figure 7.[a(60)] (Reproduced by permission of Mentor Communications AB from R. Ekman, B. Holmbom, *Nordic Pulp Pap. Res. J.*, **1**, 16–24 (1989))

Peak no.	Compound	Peak no.	Compound
1	Myristic acid (14:0)	34	Tetracosanoic acid (24:0)
2	*cis*-Ferulic acid[b]	35	Tetracosanol
3	*anteiso*-Pentadecanoic acid (15:0 ai)	36	Squalene
4	Pentadecanoic acid (15:0)	37	Docosanedioic acid
5	*trans*-Ferulic acid[b]	38	22-Hydroxydocosanoic acid
6	*anteiso*-Hexadecanoic acid (16:0 ai)	39	Lignans, partially methylated
7	Hexadecenoic acid (16:1)	40	Campesterol
8	Palmitic acid (16:0)	41	Campestanol
9	*anteiso*-Heptadecanoic acid (17:0 ai)	42	Sitosterol
10	Pinolenic acid (5, 9, 12–18:3)	43	Sitostanol
11	Linoleic acid (9, 12–18:2)	44	Cycloartenol
12	Oleic acid (9–18:1) + (11–18:1)	45	Methylenecycloartenol
13	Stearic acid (18:0)	46	Monosaccharides
14	Thunbergol	47	Glucose anomers
15	*anteiso*-Nonadecenoic acid (19:1 ai)	48	Resin acid TMSi esters
16	*cis*-Abienol	49	Oxidized resin acids and phenolics
17	Palustrol	50	Resveratrol
18	Pimaric acid (Pi)	51	Sucrose
19	Sandaracopimaric acid (Sa)	52	Isorhapontigenin
20	Eicosatrienoic acid (20:3)	53	Astringenin
21	Isopimaric acid (Ip)	54	Isolariciresinol
22	Levopimaric (Levo) + palustric acid (Pal)	55	Secoisolariciresinol
23	Dehydroabietic acid (DeAb)	56	Liovil
24	Eicosanoic acid (20:0)	57	Conidendric acid
25	Abietic acid (Ab)	58	Picearesinol
26	Neoabietic acid (Neo)	59	Matairesinol
27	Neoabietol	60	Allohydroxymatairesinol
28	Docosanoic acid (22:0)	61	Hydroxymatairesinol
29	15-Hydroxydehydroabietic acid	62	α-Conidendrin
30	Docosanol	63	Lariciresinol
31	Tricosanoic acid	64	Oxomatairesinol
32	*anteiso*-Tricosanol	65	Lignan, MW of TMSi derivative = 590
33	Tricosanol	66	Pinoresinol

[a] Compounds were detected by MS of the Me esters/TMSi ethers (abbreviations in parentheses). Several compounds (e.g. peaks 46–66) listed in this table are not present in Figure 7, but were detected on other types of samples or using different conditions.
[b] Me ester/Me ether internal standard heptadecanoic acid.

Another scheme involves the use of ion-exchange chromatography to fractionate acetone extracts into weak acids, strong acids and neutrals.[61] A weak ion-exchange resin bed was used for this purpose, and elution was carried out successively with diethyl ether, diethyl ether saturated with CO_2, and formic acid to recover neutrals (hydrocarbons, alcohols, esters, etc.), weak acids (fatty and resin acids), and strong acids, respectively. Samples were derivatized with diazomethane and analyzed with capillary GC. A novel feature of this approach is the use of short (2 m) fused-silica capillary columns to separate the main classes of extractives into fatty acids, fatty alcohols, sterols and triterpene alcohols, waxes, diglycerides, steryl esters, and triglycerides. A typical chromatogram of a standard mixture is shown in Figure 8, and the peaks are identified in Table 5. This technique can be combined with fast temperature programming to elute the compounds of interest in about 15 min.

Other approaches have also been developed for the separation of extractives into individual classes. The use of reversed-phase HPLC with a mass detector has been reported by Suckling et al.[62] for the analysis of fatty acids, resin acids and triglycerides in softwood extractives. More recently, thin-layer chromatography (TLC) was used to separate acids and neutral compounds, including sterols, triglycerides, and steryl esters, in extracts from bleached CTMP and kraft pulps.[63] Extracts were applied to silica-coated plates and neutrals were eluted with heptane–acetone–ammonia and fatty and resin acids with dichloromethane–methanol–ammonia. The eluted compounds were quantified using a scanning UV detector.

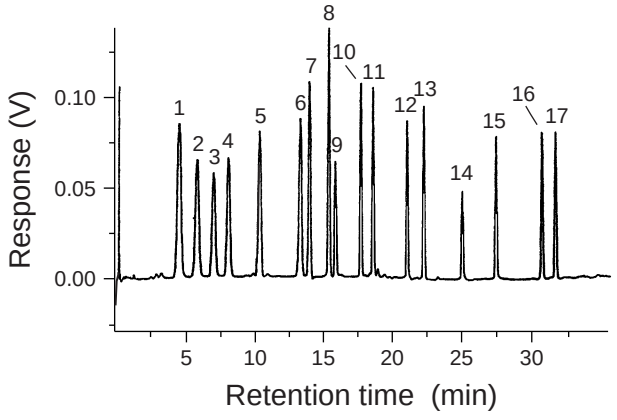

Figure 8 Typical gas chromatogram of a standard mixture of extractives as determined by the short column procedure.[61] Peaks are identified in Table 5. (Reproduced by permission of Walter de Gruyter GmbH & Co. KG from B.B. Sitholé, J.L. Sullivan, L.H. Allen, *Holzforschung.*, **46**(5), 409–416 (1992).)

Table 5 Identification of compounds corresponding to the peaks in Figure 8.[61] (ref = reference standards). (Reproduced by permission of Walter de Gruyter GmbH & Co. KG from B.B. Sitholé, J.L. Sullivan, L.H. Allen, *Holzforschung.*, **46**(5), 409–416 (1992))

Peak no.	Compound	Concentration ($mg\,mL^{-1}$)	Retention time (min)
1	Manool	1.017	4.43
2	Heptadecanoic acid	1.041	5.72
3	Linoleic acid	0.955	6.87
4	Stearic acid	0.921	7.92
5	Dehydroabietic acid	1.008	10.13
6	Heneicosanoic acid (ref.)	0.993	13.07
7	Behenyl alcohol	1.026	13.72
8	Tricosanoic acid	1.071	15.10
9	Monostearin	1.001	15.57
10	Cholesterol (ref.)	1.008	17.38
11	Stigmasterol	0.926	18.25
12	Tricaprin	1.025	20.67
13	Octadecyl stearate	0.950	21.85
14	Distearin	1.002	24.55
15	Cholesterol stearate	1.006	26.88
16	Tristearin	1.039	30.10
17	Trinonadecanoin	1.010	31.03

Analysis of individual fatty and resin acids is best carried out on medium-length (20 m) fused-silica capillary columns such as SE-30 or Carbowax columns, as reported by Dorris et al.[64] In this work, the use of equivalent chain length for expressing the retention of a compound, and also the effects of several variables, including temperature, splitting ratio, and method of injection, on the precision and accuracy of the analysis were critically examined.

Fatty and resin acids must be converted into volatile derivatives prior to GC analysis. Derivatization procedures were recently described in a review by Sitholé on methods of extractives analysis.[53] They include conventional methylation with diazomethane, trimethylsilylation, ethylation, and pentafluorobenzylation, and also rapid in situ methylation with trimethylammonium hydroxide. Because of its speed and simplicity, in situ methylation is often used for many applications.

A rapid spectrophotometric method for determining total fatty and resin acids in pulp and paper samples was reported by Sitholé.[65] After solvent extraction, the resin and fatty acids (RFA) were complexed with cupric ions, extracted, and their concentrations determined colorimetrically. The results were in good agreement with those of the GC method.

5.4 Determination of Inorganics

The next class of minor components of wood and pulp are the inorganics, which together generally make up less than 1% of the weight of wood. Potassium, calcium, and magnesium account for 70–80% of the total ash content. In spite of their presence at such low levels, inorganics can have a significant negative impact on a mill operation and on pulp properties. The content and composition of inorganics in wood and pulp must be determined for a variety of reasons: to conduct mass balance studies in order to establish the distribution of particular elements in the system,[66,67] to assess the impact of certain elements, such as calcium, on scaling and pitch deposition,[68] to determine the effect of various inorganic constituents in kraft liquors on the pulping and recovery processes,[69] or to carry out fundamental studies of the distribution of elements at the morphological level,[70] to list only a few. The presence in pulp of certain transition metals, particularly iron, manganese, and copper, can have several deleterious effects on pulp properties and on the consumption of bleaching chemicals. In particular, metal ions are often responsible for brightness reversion and a decrease in pulp brightness through the formation of colored complexes with lignin. They can also induce, through catalytic and free radical reactions, the degradation of cellulose during oxygen and ozone bleaching. In addition, metal ions cause the decomposition of hydrogen peroxide used in pulp bleaching. These problems can be considerably aggravated if high levels of bark are present in the wood furnish, owing to the much higher level of metals in bark.[71] The type and levels of metals must therefore be determined to control such problems.

In finished paper and board products, the determination of inorganics provides information on the level and type of fillers, coatings, and other additives, such as clay, calcium carbonate, titanium dioxide, alum, and talc.

The total level of inorganics is generally determined by simply destroying the organic matter in a furnace at

a temperature of 525–575 °C (PAPTAC G.10; TAPPI T211) or 900 °C (PAPTAC G.11; TAPPI T413). The portion of ash insoluble in hydrochloric acid is usually taken as an approximate measure of the silicates content, and referred to as acid-insoluble ash (PAPTAC G.33; TAPPI T244).

Determination of specific elements in wood and pulp requires either dry ashing or wet digestion of the sample prior to analysis. Dry ashing is generally performed at 525–575 °C. The use of higher temperatures is not recommended because it can lead to substantial losses of metals, particularly sodium, calcium, copper, and iron.[72] This problem can be circumvented by the use of wet ashing techniques. The latter involve treating a wood or pulp sample with oxidative chemicals to destroy the organic matter and liberate the free metals in solution. Various mixtures can be used for digestion, including sulfuric acid and hydrogen peroxide,[72] or more conveniently perchloric and nitric acid (PAPTAC G.30). However, the use of perchloric acid is especially hazardous because of the risk of formation of explosive perchlorate compounds. Specially designed fumehoods and precautions are required for perchloric acid digestions.[73] Microwave digestion in pressurized vessels has become increasingly popular, partly because of safety considerations, since no perchloric acid is used, and because it is considerably faster than conventional wet digestion methods. It has recently been applied to the digestion, using only nitric acid, of a wide range of kraft mill samples.[74] It has also been used for digestion of tree foliage and several standard materials, including pine needles, using sequential addition of nitric acid, hydrogen peroxide, and hydrochloric acid.[75]

Following dry ashing or wet digestion, analysis of elements is generally carried out by atomic absorption spectrometry (AAS) or by inductively coupled plasma atomic emission spectrometry (ICPAES). Although flame AAS is commonly used,[70,76,77] graphite furnace (electrothermal) atomization has also been applied to the determination of trace metals in pulp.[78] ICPAES is finding increasing use in pulp and paper applications because of its ability to determine sequentially or simultaneously a wide range of elements, including aluminum, silicon, phosphorus, and sulfur.[74]

X-ray methods can also be applied to the determination of elements in pulp and paper samples. Kocman and Bruno[79] used X-ray fluorescence (XRF) spectrometry for determining a wide range of fillers, including clay, titanium dioxide, calcium carbonate, and talc, in fine papers, newsprint, and recycled stock. Paper samples were first pulverized in a laboratory blender, pressed into a pellet using a hydraulic press, and then analyzed with a wavelength-dispersive X-ray instrument. Keitaanniemi and Virkola[80] also used XRF spectrometry for the determination of inorganic components in solid samples, including wood, bark, pulp, furnace smelt, dust, lime, lime mud, grits, and dregs. The advantage of this technique is that the sample can be analyzed directly without prior ashing or digestion. However, the XRF analysis is subject to matrix effects and requires careful calibration to yield reliable results.

Another powerful nondestructive method for elemental analysis is neutron activation analysis (NAA). Samples are irradiated with a neutron source and, based on the energy and level of emitted γ-rays, the type and concentration of a particular element can be determined. This technique was applied to the measurements of elements in black spruce wood,[81] and the results were generally in good agreement with those obtained by ICPAES.

Energy-dispersive spectroscopy (EDS) combined with scanning electron microscopy (SEM) is an effective and rapid tool for mapping the distribution of elements in a sample. Ormerod et al.[82] described the application of this technique to the determination of fillers, coating additives, and contaminants throughout a paper sheet. Hillis and de Silva applied a similar technique to the analysis of deposits in morphological regions of wood.[83] This approach, however, is generally used only for qualitative or semiquantitative analysis.

Elemental distributions at the morphological level can also be obtained through the use of secondary ion mass spectrometry (SIMS). Bailey and Reeve[81] described the application of this technique for determining the spatial distribution of metals in several morphological features of black spruce wood.

6 ANALYSIS OF PROCESS LIQUIDS

For the purpose of this discussion, process liquids have been divided into two parts: process liquors which include pulping, bleaching, and spent liquors, and process waters and effluents.

6.1 Analysis of Process Liquors

Routine analyses are performed on pulping liquors to determine the concentration of active chemicals, to adjust digester charge, and to follow the progress of a cook. In the latter application, samples can be withdrawn at various stages during pulping and analyzed to determine the consumption of various chemicals. The major types of pulping liquors are those used in the sulfite and kraft processes.

6.1.1 Sulfite Liquors

In the sulfite process, pulping liquors are produced by absorbing sulfur dioxide gas in a base, such as calcium,

Table 6 Typical sulfite cook specifications.[84] (Reproduced by permission of PAPTAC from O.V. Ingruber, M.J. Kocurek, A. Wong, *Pulp and Paper Manufacture*, 3rd edition, Joint Textbook Committee of the Paper Industry, Montreal, Vol. 4, 1985)

Chemical	Acid	Bisulfite	Neutral	Alkaline
Total SO_2 charge, % on o.d. wood	20	20	20	20
Free SO_2, % of total SO_2 (true free SO_2)	80	0	0	0
Bound SO_2, % of total SO_2 (true combined SO_2)	20	100	100	100
Liquor to wood ratio (o.d. wood)	4:1	4:1	4:1	4:1
Buffer: Na_2CO_3, Na_2O equivalents ratio	–	–	0.20	–
Alkali: NaOH, Na_2O equivalents ratio	–	–	–	0.10–0.50
Additive(s), AQ, % on o.d. wood	–	–	–	0.02–0.20
Starting pH	1–2	3–5	7–9	10–13
Palmrose liquor analysis (g SO_2 per 100 mL liquid)				
Total SO_2	5	5	5	5
Free SO_2	4	2.5	–	–
Combined SO_2	1	2.5	–	–
Actual SO_2 concentration (g SO_2 per 100 mL liquid)				
SO_2 as excess 'H_2SO_3'	3	0	0	0
SO_2 as $MHSO_3$[a]	2	5	2.5	0
SO_2 as M_2SO_3[a]	0	0	2.5	5

[a] M = monovalent cation (Na^+, NH_4^+, $\frac{1}{2}Ca^{2+}$, $\frac{1}{2}Mg^{2+}$).

magnesium, sodium, which are used in their hydroxide or other suitable form, or ammonia. Sulfur dioxide can be present in sulfite liquors in various forms: as combined SO_2 in the form of sulfite or bisulfite, or as free SO_2. The sum of these two gives the total SO_2 in the liquor. Standard methods for analyzing sulfite liquors involve measuring the free and total SO_2 by titration (PAPTAC J.1, J.13; TAPPI T604). The combined SO_2 is determined by difference. Total SO_2 is traditionally determined by titration with potassium iodate and free SO_2 by titration with sodium hydroxide. Sulfite cooking liquors cover a wide range of specifications depending on the pH of the liquor and the ratio of SO_2 to alkali. Table 6[84] shows the approximate composition of the four major sulfite processes over a starting pH ranging from 1–2 for the acid process to 10–13 for the alkaline process. As indicated in Table 6 and discussed in detail by Ingruber,[84] the calculation of both true free and combined SO_2 depends on the pH level, and may differ from free and combined SO_2 determined by the standard titration method.

Determination of the various forms of SO_2 is important because of their significant impact on the pulping rate and pulp properties. The level of combined SO_2 affects chip penetration, pulping rate, and strength properties. Increasing the level of combined SO_2 results in increased brightness and tensile strength of the pulp. On the other hand, too high a level of combined SO_2 can result in a decreased pulping rate. Free SO_2 affects mainly the pulping rate, and total SO_2 has an impact on several factors, including chip penetration, rejects level, brightness, and bleach demand. A particularly important parameter in sulfite cooking is pH, which influences pulp yield, pulping rate, brightness, and strength properties. It is especially critical in bisulfite pulping where a small variation in pH can have a significant effect on these factors.

6.1.2 Kraft Liquors

In the kraft process, pulping and spent liquors are continuously generated in a cyclic liquor system (Figure 9).[85] The pulping or cooking liquor, also known as white liquor, is combined with the wood in a digester. After cooking, the spent liquor, referred to as black liquor, is removed during washing, concentrated in several stages of evaporation and then burned in a recovery furnace. The resulting smelt is dissolved in water to produce the so-called green liquor. The latter is converted back to white liquor in a causticizer. This section examines the more common types of analyses conducted on kraft white, green, and black liquors.

The active components in a kraft white liquor are sodium hydroxide and sodium sulfide. The main parameters that are determined routinely in white liquors are the effective alkali (EA) charge, the initial alkali concentration, and the sulfidity, all of which have a profound effect on the pulping process and on pulp quality. The EA is defined as $NaOH + \frac{1}{2}Na_2S$, expressed as Na_2O, and sulfidity as $Na_2S/(NaOH + Na_2S)$, expressed as Na_2O.

Increasing the EA reduces pulp yield, but an excess EA at the end of the cook also results in improved pulp brightness. The alkali concentration charged at the start of the cook affects mainly chip penetration and the rate of delignification. Both the rate of delignification and pulp strength properties are improved with increased sulfidity.

The traditional method for analyzing white liquors is known as the ABC titration, which involves a three-step titration with hydrochloric acid, with successive addition of barium chloride and formaldehyde.[86] The end-points are used to calculate the various alkali charges, and also the concentration of sodium carbonate. The latter is present as a result of incomplete conversion of sodium carbonate to sodium hydroxide in the causticizing step of the kraft recovery process. In addition to carbonate,

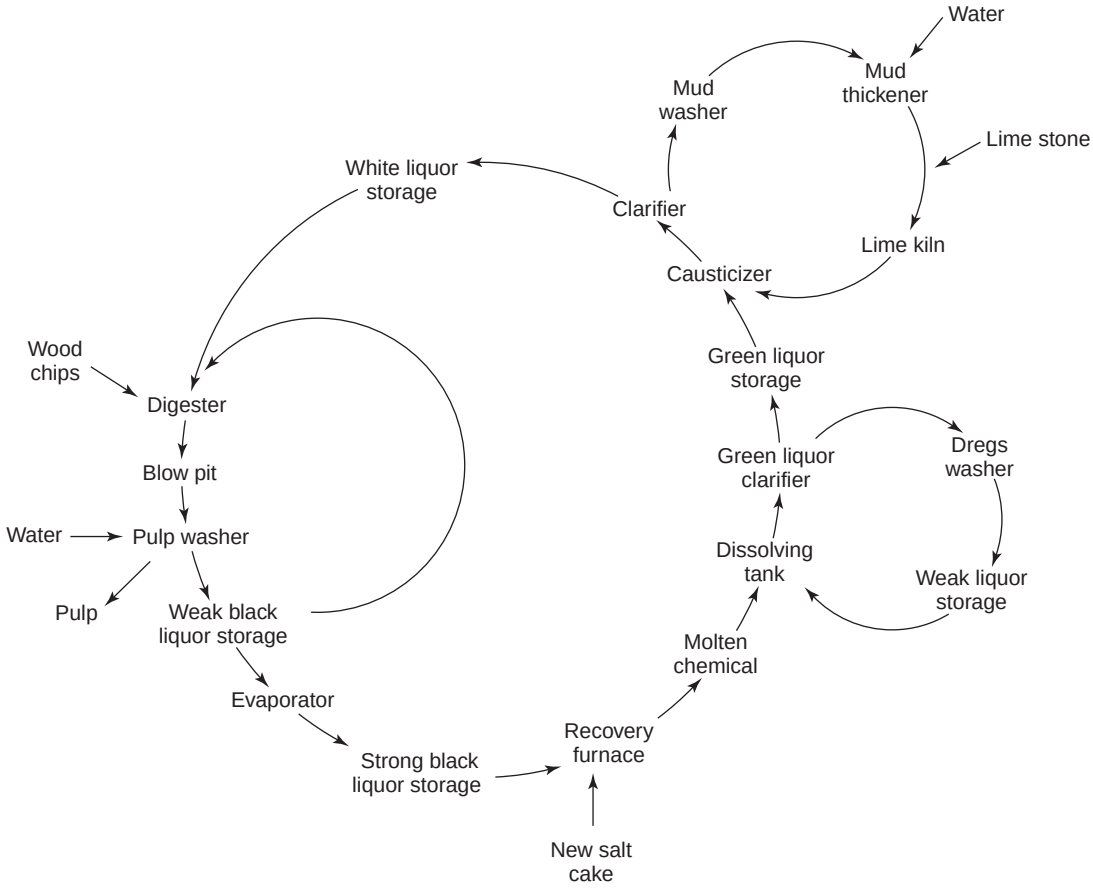

Figure 9 Diagram showing cyclic nature of the kraft recovery process.[85] (Reproduced by permission of PAPTAC from T.M. Grace, B. Leopold, E.W. Malcolm, M.J. Kocurek, *Pulp and Paper Manufacture*, 3rd edition, Joint Textbook Committee of the Paper Industry, Montreal, Vol. 5, 1989.)

several other nonprocess components are present in white liquors and are referred to collectively as 'dead load'. They include a range of inorganic sodium salts, such as sodium sulfate, thiosulfate and chloride, and metal ions including potassium, calcium, magnesium and iron, and silicates. These originate from the wood, chemicals added in the recovery process, or from incomplete conversion of compounds in the recovery process. A high dead load is undesirable because it reduces the efficiency of lignin removal, and can give rise to various operating problems such as corrosion and deposition on process equipment.

The ABC titration can also be used to analyze green liquors for sulfide and carbonate. However, in addition to sodium salts, green liquors also contain a number of impurities, consisting mainly of metallic compounds, such as silicates, aluminates, and carbon, which must be largely removed by sedimentation in a clarifier before the causticizing process. Therefore, total solids and inorganics content must be known because they can affect the efficiency of the clarification process. The ratio of lime to calcium carbonate must also be determined because it affects the causticizing efficiency. A ratio close to 1 is desirable for optimum efficiency.

Black liquors are also analyzed for a variety of parameters. Residual alkali provides an indication of alkali consumption during the cook and is most accurately determined by potentiometric autotitration to a weak inflection point.[87] An excess alkali at the end of the cook is usually desirable because it ensures maximum solubilization of the resin, which, in turn, reduces the risk of pitch deposition. Analyses of sodium and dissolved solids content in black liquors are also performed routinely as a measure of washing efficiency and to determine the amount of makeup chemicals. A rapid method for determining kraft black liquor solids was proposed by McDonald.[88] Other black liquor tests include viscosity, boiling point, specific heat, and soaps and inorganics content. All of these parameters affect the efficiency of evaporation and subsequent burning in the recovery boiler.

In addition to several standard or official methods for the analysis of kraft liquors (PAPTAC J.12, J.15; TAPPI T624 and T625), a variety of instrumental techniques have

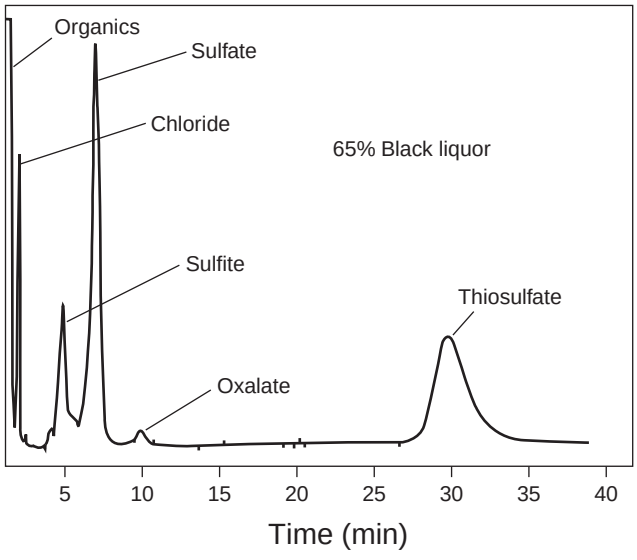

Figure 10 Ion chromatogram of a kraft black liquor.[86] (Reproduced by permission of PAPTAC from T.M. Grace, B. Leopold, E.W. Malcolm, M.J. Kocurek, *Pulp and Paper Manufacture*, 3rd edition, Joint Textbook Committee of the Paper Industry, Montreal, Vol. 5, 1989.)

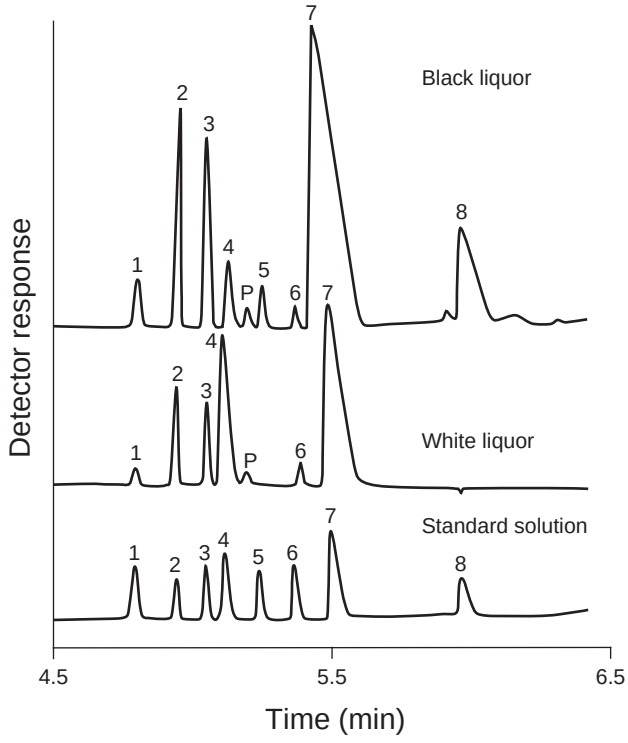

Figure 11 Separation of anions in a standard solution, white liquor and black liquor by CE.[99] Electrolyte, 10 mM chromate, 2 mM tetradecyltrimethyl-ammonium hydroxide (TTAOH) (pH 11.4); separation current, −22 μA. Peaks: chloride (1), thiosulfate (2), sulfate (3), sulfide (4), polysulfide (P), oxalate (5), sulfite (6), carbonate (7), formate (8). The standard solution contained 5 mg L^{-1} of each ion. The white and black liquors were diluted 1 : 1000 and 1 : 250, respectively, in He-sparged water. (Reproduced by permission of TAPPI Press from J. Sullivan, M. Douek, *Proceedings of 1999 TAPPI Pulping Conference*, TAPPI Press, Atlanta, GA, 1999.)

also been devised for determining the active chemicals and dead-load components in these liquors. Potentiometric titration with mercuric chloride using a sulfide ion-selective electrode has been widely used for the analysis of various sulfur species.[89] Pulse polarography was also shown to be an effective technique for determining sulfide, sulfite, and thiosulfate.[90] Ion chromatography (IC) is a highly versatile and proven technique for measuring a wide range of inorganic and organic anions in kraft and other types of pulp and paper liquors.[91–96] Standard methods for the analysis of pulping liquors, based on IC, have also been introduced (PAPTAC J.20; TAPPI T699). One of the major drawbacks of the technique is that high levels of dilution are required, which can result in measurement errors, particularly for sulfide, sulfite, and thiosulfate owing to shifts in equilibrium concentration between sulfur species during sample dilution. Many of these problems can be largely circumvented, however, by analyzing the samples immediately after dilution.[96] A typical ion chromatogram of a kraft black liquor is shown in Figure 10.[86] More recently, CE was used for the determination of various sulfur species, and also several other anions, including hydroxide, chloride, oxalate, and carbonate, in kraft pulping liquors.[97–99] Examples of electropherograms of white and black liquors and of a standard solution are shown in Figure 11.[99] The main advantages of the CE approach compared with IC are its speed, its ability to determine simultaneously a wide number of ions in a sample, and the use of inexpensive capillary columns.

Considerable progress has also been achieved in the area of on-line measurement of kraft liquors. Alkali and sulfide consumptions during kraft cooking were measured using a conductivity detector and on-line IC.[100] Another approach involves the use of attenuated total reflectance IR spectroscopy. EA, carbonate, sulfate, and thiosulfate could be determined by this technique in laboratory experiments on kraft liquors.[101] It has also been applied to on-line measurements of EA and carbonate in these liquors.[102]

In a modified version of kraft pulping, referred to as polysulfide pulping, significant amounts of sodium polysulfide (Na_2S_x, where $x = 2–5$) are generated in the white liquor, typically by oxidation of sodium sulfide present in the liquor. The use of polysulfide in kraft pulping is known to produce an increase in pulp yield. This beneficial effect is attributed to the oxidation by polysulfide of the reducing end groups of polysaccharides, thereby stabilizing them against the so-called peeling or degradation reaction

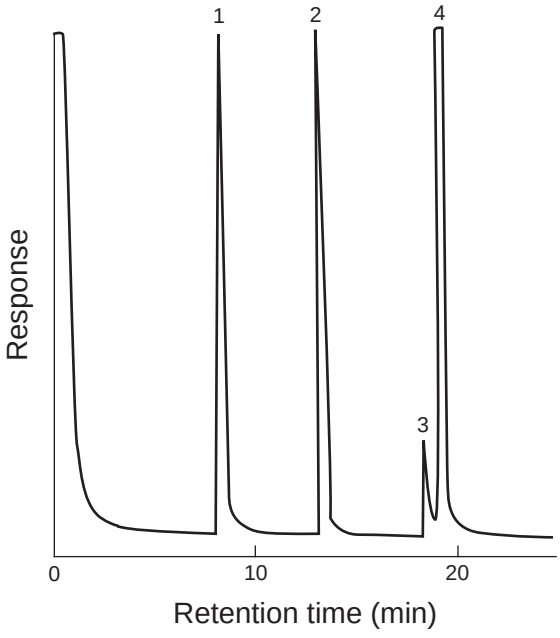

Figure 12 Gas chromatogram obtained in the determination of sulfur, from the decomposition of polysulfide, by the TPP method.[104] Peaks: 1 = internal standard; 2 = TPP; 3 = triphenylphosphine oxide (TPPO); 4 = TPPS. TPPO is formed by the action of TPP with dissolved oxygen in the sample. (Reprinted from L.G. Borchardt, D.B. Easty, *J. Chromatogr.*, **299**, 471–476 (1984), Copyright 1984, with permission from Elsevier Science.)

which takes place during conventional kraft pulping. Several methods have been devised for measuring the concentration of polysulfide in these liquors. The potentiometric method with mercuric chloride and a sulfide ion-selective electrode described by Papp[89] can also be used for determining polysulfide. The polysulfide is determined by measuring thiosulfate before and after addition of sulfite to the sample. A spectrophotometric method based on the measurement of absorbance at 285 nm has also been reported by Teder.[103] Another method involves reduction of polysulfide with sodium amalgam and measuring the sulfide content before and after the reduction step by potentiometric titration with mercuric chloride (TAPPI T694). More recently, a direct method was reported involving decomposition of polysulfide to elemental sulfur. The sulfur was then derivatized with triphenylphosphine (TPP), and the resulting triphenylphosphine sulfide (TPPS) was determined by GC with flame ionization detection (FID).[104] Figure 12 shows a typical gas chromatogram obtained by this method. Good agreement was reported with the sodium amalgam method on commercial polysulfide liquors. A modified approach to this method involves gravimetric determination of elemental sulfur precipitated by acidification of polysulfide solutions.[105] The same authors recently reported using a new end-point in the ABC titration of polysulfide liquors to obtain a measure of polysulfide along with other sulfoxy anions.[106]

6.1.3 Bleaching Liquors

Bleaching liquors represent another type of process liquors that require frequent analysis. The subject of pulp bleaching is covered in detail in several excellent textbooks and monographs.[107–109] The major bleaching agents used in the pulp and paper industry are listed in Table 7. Kraft bleaching liquors consist mainly of chlorine dioxide, sodium hydroxide, sodium hypochlorite, and hydrogen peroxide. The use of elemental chlorine in pulp bleaching has decreased dramatically since its role in the formation of chlorinated dioxins and furans was recognized. Oxygen and ozone are used to a limited extent at the moment, mainly in totally chlorine-free (TCF) bleaching installations. Mechanical pulp are commonly bleached with hydrogen peroxide or sodium hydrosulfite. Several standard methods, based primarily on titrimetry, are available for analysis of generator and spent bleaching liquors (PAPTAC J.2, J.14, J.16, J.17 and J.22; TAPPI T611, T613 and T622). A more recent TAPPI method (TAPPI T700) is based on the use of IC for determining the major chlorine and oxychlorine species. Several problems with the use of IC for the determination of these species, including incomplete recovery of chlorine dioxide during analysis and interferences with organic ions, were recently identified and resolved.[110] Based on analytical conditions developed in this work, excellent recovery of spiked bleaching effluent samples was obtained. Figure 13(a) and (b) shows typical ion chromatograms of a sample of a chlorine dioxide effluent. The use of both conductivity and UV modes of detection allowed better separation of certain ions to be achieved. For example, the response of chlorite with a UV detector was significantly higher than that of organic acids, providing a more reliable means of quantifying chlorite in the presence of high concentrations of formate.

Table 7 Major bleaching agents used in the pulp and paper industry

Bleaching agent	Chemical formula
Chlorine	Cl_2
Chlorine dioxide	ClO_2
Sodium hydroxide	NaOH
Hydrogen peroxide	H_2O_2
Sodium hypochlorite	NaOCl
Sodium hydrosulfite	$Na_2S_2O_4$
Oxygen	O_2
Ozone	O_3

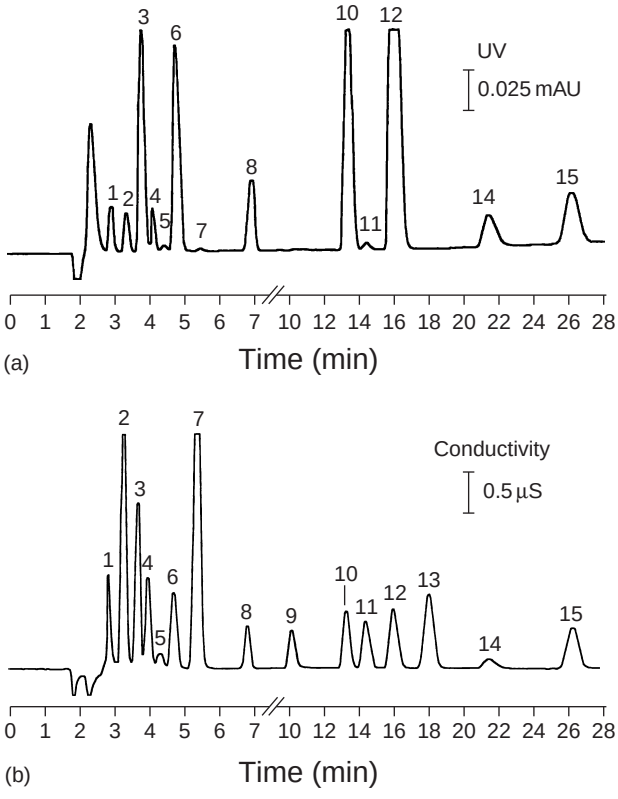

Figure 13 Separation of anions on IonPac AG12A/AS12A columns using a borate–tetraborate eluent with coupled (a) UV and (b) suppressed conductivity detection.[110] Each solute present at 2 mg L^{-1} (except carbonate, which arises from dissolved CO_2 present in the dilution water). Peaks: acetate (1), formate (2), chlorite (3), chloroacetate (4), carbonate (5), bromate (6), chloride (7), dichloroacetate (8), orthophosphate (9), bromide (10), chlorate (11), nitrate (12), sulfate (13), trichloroacetate (14), oxalate (15). (Reprinted from J. Sullivan, M. Douek, *J. Chromatogr. A.*, **804**, 113–121 (1998), Copyright 1998, with permission from Elsevier Science.)

6.2 Analysis of Process Waters and Effluents

Mill effluents consist of process waters and filtrates that are discharged from various unit operations of the pulp and paper-making process. Effluents are characterized for a wide range of parameters and for various purposes, including assessing the level of priority pollutants, conducting mass balance studies to establish the distribution of various components in a system, and determining the effect of various operating conditions, such as washing efficiency, white water closure, and water treatment processes on the removal or buildup of various components in the system. With the trend toward increased system closure in mills to limit the discharge of effluents into receiving waters, the levels of dissolved, colloidal, and suspended material could also increase significantly. This could, in turn, impact on paper machine runnability and product quality.

Table 8 Common tests performed on process waters and effluents

pH
Conductivity
Solids
Turbidity
Color
BOD
COD
TOC
AOX
Hardness
Silica
Metals: potassium, calcium, magnesium, iron, copper, manganese, aluminum
Inorganic anions: fluoride, chloride, chlorate, nitrate, phosphate, sulfate

BOD, biochemical oxygen demand; COD, chemical oxygen demand; TOC, total organic carbon; AOX, adsorbable organic halogens.

Some of the common tests conducted on process waters and effluents are listed in Table 8. Most of these tests are based on well-established industry methods and standards (PAPTAC H.4; TAPPI T620).[111] Several of the instrumental techniques for determining some of these parameters, including IC for anion analysis and AAS or ICPAES for metals, have already been described in previous sections. Other tests for determining basic properties such as solids, turbidity, color, and pH, are straightforward and require no further explanation. Subsequent discussion will therefore be limited to the more elaborate tests and to those based on the use of new analytical schemes and instrumentation.

The BOD, COD, and TOC tests provide a measure of the organic matter in the sample. The BOD is a measure of the amount of oxygen consumed by decomposition of organic matter by microorganisms added to the sample. In the standard BOD test (PAPTAC H.2), the oxygen content in the sample is measured before and after incubation with microorganisms over a 5-day period. Several sources of error have been identified to account for the high variability of the test results,[112–114] including water quality, viability of the seed source, use of nitrification inhibitor, and presence of toxicants in the sample. However, the BOD test continues to be widely used as an indicator of the level of organic pollutants in wastewaters.

One of the major drawbacks of the standard BOD test is that it takes 5 days to complete, which makes it impractical for process control. A number of faster alternative approaches have been proposed. One of them is simply based on correlating 24-h BOD results with the 5-day conventional test.[115] A good correlation was obtained between the two tests with untreated and

biologically treated paper mill effluent. Another report showed good correlation between BOD 5-day test results on spent sulfite liquor and carbohydrate content.[116]

Instrumental techniques, based on respirometry, for measuring oxygen uptake on a continuous basis, have received considerable attention. Several types of commercial instruments are available, including manometric, volumetric, electrolytic, and direct-input respirators, depending on the method used for determining oxygen consumption.[117] In all cases, carbon dioxide produced during oxygen uptake is removed from the air space. More recently, the use of microbial oxygen sensors with very fast response time of the order of 30 s have been reported. These sensors consist of cells immobilized on an oxygen electrode. In amperometric-type sensors,[118,119] the acceleration of respiration rate is measured by the change in electric current. An optical fiber sensor for BOD measurement has also been reported.[120] In this case, the sensing membrane contains an oxygen-sensitive fluorescent material. Typical response times are 5–10 min. Another advantage of optical type sensors is that oxygen is not consumed in the process. BOD results with both amperometric and optical sensors were found to correlate well with the 5-day BOD test on various municipal wastewaters and effluents. Although BOD sensors have the potential for considerably reducing measurement time, they have not yet been widely adopted for commercial pulp and paper mill applications.

The COD test is a measure of oxidizable material in the sample under specified conditions. In one of the standard tests (PAPTAC H.3), denoted the ''open reflux'' method, the sample is refluxed with potassium dichromate under strongly acidic conditions. Excess dichromate is determined at the end of the reflux period by titration with ferrous ammonium sulfate. In general, COD test results are much higher than those obtained with the BOD test, because the conditions of oxidation are much more severe. The determination of COD can also be carried out more rapidly and conveniently by a ''closed reflux'' colorimetric method.[121] In one version of this approach, the sample is placed in a small digestion tube and heated in an oven at 150 °C for 2 h using a mixture of chromic and sulfuric acid. The time can be reduced further by using microwave digestion.[122] After cooling, the absorbance is measured at 600 nm. Absorbance at that wavelength is related to the increase in Cr^{3+} resulting from the oxidation of chromium from the hexavalent ($Cr_2O_7^{2-}$) to the trivalent state. Complete systems for the closed reflux method are available commercially (e.g. from Hach). The method has been shown to give comparable results to the standard titrimetric procedure.

In the TOC test, all the organic carbon originating from both oxidizable and nonoxidizable matter is determined in the sample. Organic carbon is converted into CO_2 using either UV radiation, chemical oxidants, or high temperature, and the liberated CO_2 is then measured by IR spectroscopy or coulometric titration.[123]

Another group parameter for evaluating the level of halogenated organic material is AOX. This test measures the level of halogens, primarily present as chlorine, bound with organic compounds. In mills using chlorine compounds for bleaching, AOX levels must be below certain levels, as determined by environmental regulations. In standard AOX tests[124] (PAPTAC H.6; SCAN W9), the sample is adsorbed on granular activated carbon (GAC). Inorganic halides are removed by washing. The adsorption can be carried out in GAC-packed columns (column method) or by shaking the sample with GAC for a specified time (shaker method), as illustrated in Figure 14(a) and (b).[125] The GAC with adsorbed organic material is pyrolyzed in a combustion furnace and the organic halogens are determined by microcoulometric titration. Several potential sources of variability in the AOX test, including discrepancies between shaker and column results, and sample storage conditions, as well as measures to avoid these problems, have been reported by Sullivan and Douek.[126]

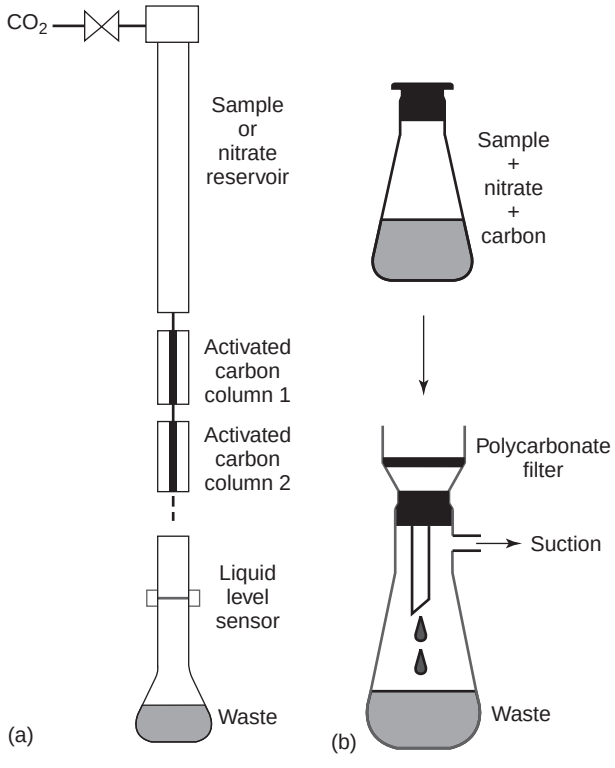

Figure 14 Schematic diagram illustrating the adsorption of samples on activated carbon for AOX determinations, using (a) the column or (b) the shaker method.[125] (Reproduced by permission of Southam from S.M. Odendahl, K.M. Weishar, D.W. Reeve, *Pulp Paper Can.*, **91**(4), T136–T140 (1990).)

In addition to the various tests described so far, detailed analyses of the dissolved and colloidal material released into process waters and effluents can provide valuable information on the nature and level of specific components present in the system. For example, certain lipophilic extractives, such as fatty and resin acids, sterols, and glycerides, can produce an increase in effluent toxicity, give rise to deposition problems on process equipment, and affect paper quality.

Several schemes, using a wide range of analytical techniques, for separating this material into classes or individual components have been devised. A simple scheme for separating both the lipophilic and polar extractives in a water sample from groundwood pulping of spruce was devised by Ekman and Holmbom.[60] The scheme is similar to that shown in Figure 6 for pulp samples, except that in the initial step the lipophilic material was separated from polar compounds by extraction with diethyl ether. A portion of the ether extract was saponified by alkaline hydrolysis to liberate the free fatty acids, alcohols, and sterols bound as esters in the original sample. Both the extracts and residue after extraction were then derivatized and analyzed by GC or GC/MS.

The use of methyl *tert*-butyl ether (MTBE) for extracting RFA in pulp mill effluents has also been reported.[127] Extraction of the sample at pH 9 with this solvent appears to overcome several problems, such as emulsion formation and lignin precipitation, which occur with other solvent extraction techniques. Örsa and Holmbom[128] have also demonstrated the use of MTBE for extraction of free and esterified wood resin, and also phenolic low molecular weight components, such as lignans, from TMP process waters, followed by GC. They determined the dissolved lignin from the UV absorption of the extracted water sample at 280 nm. The ability to distinguish between low molecular weight lignans and high molecular weight lignin is important, because lignans do not consume cationic polymers used in paper-making, whereas lignin does. On the other hand, lignans may reduce paper brightness.

Another convenient method for isolating lipophilic extractives from pulp mill process waters involves the use of solid-phase extraction. Backa et al.[129] have shown that fatty and resin acids and triterpenoic extractives were efficiently retained on a C_{18} hydrocarbon-substituted solid-phase disposable column. Undesirable material retained on the column was removed with an aqueous alkali solution containing methanol. The extractives were then eluted by sequential addition of ethanol and dichloromethane and analyzed by GC. The method was primarily developed for highly contaminated samples such as black liquors, but could also be applied to other types of process liquids. A similar technique, using silica Sep-Paks or C_{18} extraction cartridges, was applied by Sweeney[130] for the separation of several organic fractions in various types of pulp and paper samples, including mill process streams, and for black liquor cleanup. Compared with conventional solvent extraction techniques, solid-phase extraction methods are relatively simpler and faster, and require much less sample and solvent.

The environmental impact of mill effluents is also dependent on the molecular weight distribution of the organic material in a sample. For example, some of the low molecular weight organic halogen compounds produced during pulp bleaching accumulate in aquatic organisms. On the other hand, high molecular weight material, which constitutes the major portion of organic halogen compounds, is considered to be biologically inert. A large body of work has been published by several workers on the determination of molecular weight distribution and characterization of various fractions in bleached kraft mill effluents. Jokela and Salkinoja-Salonen[131] described the use of aqueous and nonaqueous SEC and ultrafiltration to determine the molecular weight distribution of pulp bleaching organic halogen compounds. Their work showed that both dilution of sample and solvent type can significantly affect the apparent molecular weight distribution, due to micelle formation in concentrated samples, and intermolecular association in various solvents. Detailed structural analysis of the high molecular weight fractions of both alkaline and acidic kraft mill spent bleach liquors was reported by Lindström and Österberg.[132,133] The liquors were subjected to several steps of ultrafiltration and fractions were characterized by a variety of techniques including infra-red and ^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectroscopy for structural identification. GC, MS, and microwave multiple plasma emission spectroscopy for elemental analysis were used on low molecular weight fractions and on degradation products. Similar detailed studies were conducted by Dahlman et al.[134,135] on the high molecular weight fractions from elemental chlorine-free (ECF) and TCF bleaching effuents. Determination of dissolved carbohydrates from unbleached and bleached TMP pulp has also recently received considerable attention, because of their potential impact on the paper-making process.[136] In particular, polygalacturonic acids, released during peroxide bleaching of mechanical pulp, were shown to have a detrimental effect by consuming significant amounts of cationic retention aids through complex formation.[137,138]

7 ANALYSIS OF PAPER

As indicated at the beginning of this article, a large proportion of chemical tests conducted in the pulp and

Table 9 Major types of additives used in paper-making

Type of additive	Examples
Pitch control agents	Alum, talc, nonionic and ionic surfactants
Defoamers	Fatty acid diamide, hydrophobic silica (in oil, oil/water or water emulsions)
Retention aids	Alum, polyacrylamide, poly(ethylene oxide), polyethyleneimine
Wet-strength resins	Urea–formaldehyde, melamine–formaldehyde, polyamine–epichlorohydrin
Internal sizing agents	Rosin, AKD, ASA
Pigments and fillers	Calcium carbonate, clay, titanium dioxide
Surface sizing agents	Starch, poly(vinyl alcohol), SMA, carboxymethylcellulose
Dry-strength adhesives	Starches, gum
Coatings	Pigments: clay, calcium carbonate, titanium dioxide
	Binders: starches, styrene–butadiene latex, PVA
	Additives: plasticizers, dispersants

AKD, alkylketene dimers; ASA, alkenylsuccinic anhydride; SMA, styrene–maleic anhydride; PVA, poly(vinyl acetate).

paper industry are performed on wood pulp, process liquors and effluents. There are relatively few quantitative chemical analyses that are routinely performed on paper products. Some of the methods described for wood pulp, for determination of basic parameters such as ash, silicates, metals, and extractives content, are also generally applicable to paper products. Most of the tests performed on paper are designed at determining various additives used in paper-making for a wide range of applications. A list of the major types of additives, used in paper-making, is given in Table 9. For control purposes, simple qualitative analyses, involving the use of spot tests with specific dyes, chemical reactions with selected reagents, solubility tests, observation of physical characteristics, and microscopic examination, are used to establish the presence or absence of a particular additive. A comprehensive description of some of the traditional detection and control tests applied to paper products is given by Browning,[139] and in the various industry standard methods. Only the more recent developments in the analysis of paper additives will be covered in the remaining part of this section.

One of the most widely used paper-making additives is starch. It is used for improving the dry strength of paper, in surface sizing, as binder in coating formulations, and as adhesive in paper and board converting processes. Standard methods for determining starch in paper (PAPTAC G.23; TAPPI T419) are based on extraction of the starch, addition of iodine to the filtrate, and detection of the blue starch–iodine complex. For quantitative measurements, this is done by measuring the absorbance at about 580 nm. However, the iodine test is not considered reliable because it is dependent on the ratio of amylose to amylopectin, the two polymeric forms present in starch, and on the molecular weight of the starch, and because both of these factors could change during processing and sample preparation.[140] Another approach which circumvents these problems has recently been developed. It involves hydrolysis of the starch to glucose using a combination of two types of enzymes specific to starch, α-amylase and glucoamylase. After hydrolysis, the glucose is determined by HPLC[141] or spectrophotometrically.[140] This enzymatic hydrolysis method has now been adopted by several laboratories. However, since some types of modified starches, such as cationic starches, will produce a lower yield of glucose after hydrolysis, it is always preferable to compare the results with a calibration curve obtained with the same type of starch as that used in the paper sample.

Pyrolysis/gas chromatography (py/GC) has been widely applied to the determination of various paper and board additives, including wet-strength resins, such as polyamide–epichlorohydrin,[142] neutral sizing agents such as AKD and ASA,[143] surface sizing agents, including SMA and styrene acrylate,[144] and a variety of other polymers in paper.[145] The advantages of this technique are its speed and simplicity, since practically no sample preparation is required. However, pyrograms can be very complex, particularly if several polymers are present in the sample. In order to remove the interference from the paper itself, a low-temperature prepyrolysis technique was described by Crockett et al.[145] After the paper had been decomposed and its products vented, subsequent pyrolysis was carried at higher temperatures. This was followed by mass spectral analysis of the pyrolysis products to identify various polymers present in the sample.

A method for determining free and bound AKD in paper samples was also reported by Sitholé et al.[146] It involves extraction of the paper with chloroform before and after digestion with hydrochloric acid or sodium carbonate, followed by GC. For better results, free AKD was converted by hydrolysis to stearone prior to GC analysis. A simpler and faster method for determining AKD was also reported by Nyarku and Sitholé.[147] The method is based on potentiometric titration and the results were in good agreement with the chromatographic method.

Gel permeation chromatography (GPC) coupled with IR spectroscopy has also been applied to the identification of PVA and styrene–butadiene-based latex binders in paper coatings.[148] The paper samples were first extracted with acetone, and THF was used in the GPC system as

the mobile phase. A nebulizer was used to vaporize the effluent from the GPC column and deposit it on a collection disk on a rotating stage. After evaporation of the solvent, the nonvolatile extractables on the collection disk were analyzed by reflectance IR spectroscopy. From a knowledge of the disk rotation speed and the distance between a spot on the disk and the start point, the spot could be correlated to a time-fraction on the GPC chromatogram. The GPC/IR approach could potentially be applied to other extractable polymers. One of its main advantages is that it allows various polymers in paper extracts to be separated with little or no interference from other components. It is also superior to sequential solvent extraction, which is more time-consuming and not entirely selective.

Surface spectroscopic techniques have also been applied to the characterization of paper additives. Electron spectroscopy for chemical analysis (ESCA) was used to identify pigments (clay and calcium carbonate) and latex binders on coated papers.[149] This technique, together with SIMS, was also used to characterize internal and surface sizing agents in paper.[150] The difference between the two techniques is that ESCA can only detect the presence of such additives, but provides little structural information. On the other hand, with SIMS, the chemical structure of the additive can be identified. Both techniques, however, can provide information on the spatial distribution of sizing agents on the paper surface.

8 ANALYSIS OF DEPOSITS AND CONTAMINANTS

In pulp and paper production, a large number of chemical components are released, in soluble, colloidal or insoluble form, into process waters. Some of them are transferred from pulp fibers into the surrounding liquid phase, whereas others are added to the system for specific purposes. Major components present in the paper-making process include carbohydrates, lignin, wood resin, soaps, defoamers, sizing agents, and inorganic materials such as clay, talc, and calcium carbonate. Under certain conditions of pH, temperature, and degree of mixing, or at certain concentration levels, some of these components can precipitate or agglomerate, resulting, in some cases, in major operational and quality-related problems. Common symptoms include deposition on paper-machine equipment, forming fabrics, and press felts. The consequences of these so-called pitch deposits range from premature replacement of fabrics and felts, to sheet breaks and holes in the paper. Small particles can also be carried through to the finished product, causing dirt spots or streaks which can downgrade the quality of the paper and create further problems during printing and converting processes. These problems can be considerably aggravated when recycled fibers are used in the papermaking furnish. Contaminants from wastepaper, commonly referred to as stickies, originate from a variety of sources, including pressure-sensitive adhesives, hot melts, coating binders, ink, wood resin, rosin size, and wet-strength resin.[151,152] In addition, deinking chemicals used in the flotation process can contribute to the deposits.

Analysis of a particular deposit on process equipment or contaminant in paper is often the first step toward finding an appropriate solution to the problem. By tracing the source of deposition or contamination to specific components in the system, more effective control measures can be devised. Most deposits consist of a complex mixture of compounds present in varying proportions, and consequently elaborate analytical schemes are required for their identification. A review of the various analytical procedures and techniques for analysis of pitch deposits was published recently by Sitholé.[153]

A general scheme for the analysis of paper machine deposits has been described by Sjöström and Holmbom.[154] It involves the separation and fractionation of various components in a deposit, followed by characterization of each fraction using various techniques. This general scheme was applied by Sjöström et al. to several other deposits and spots from a newsprint mill using deinked pulp.[155] As outlined in Figure 15, samples were first examined by microscopy to determine a number of

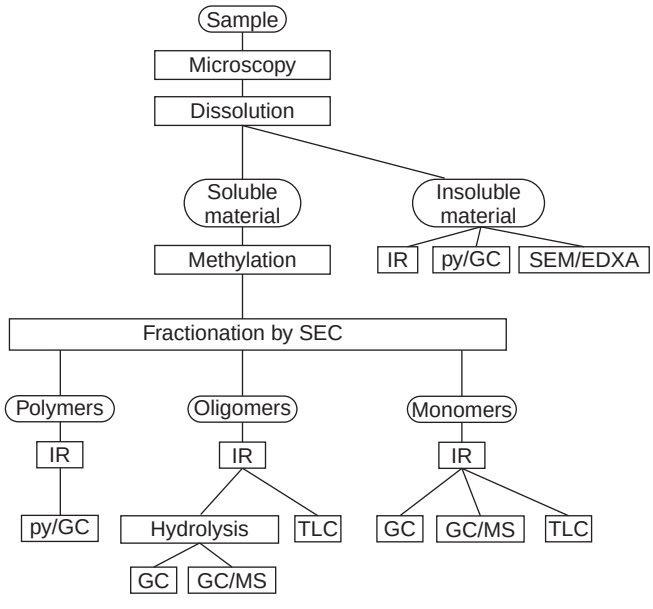

Figure 15 Scheme for analysis of paper machine deposits.[154] (Reproduced by permission of the Finnish Paper and Timber Journal Publishing Co. from J. Sjöström, B. Holmbom, *Pap. Timber*, **70**(2), 151–156 (1988).)

features such as homogeneity, color, stickiness, and presence of fibers and particles. They were then dissolved in THF and the soluble material was further fractionated by SEC into polymers, oligomers, and monomers. Details of the SEC technique were described by the same authors in another publication.[156] IR analysis was performed on each SEC fraction for general identification of major functional groups. py/GC was used for identifying polymers, such as styrene–butadiene copolymer, in the high molecular weight fraction. The second fraction consisting mainly of triglycerides and sterol esters was hydrolyzed to liberate free fatty and resin acids which could then be determined by GC and GC/MS. The same techniques were used to characterize the third fraction consisting mainly of fatty and resin acids and other low molecular weight components such as plasticizers used in adhesive formulations. TLC was also used on the second and third fractions to establish the polarity and confirm the identity of the various components. The use of TLC for identifying pitch and paper additives, such as retention aids and defoamers, in mill deposits was reported previously.[157]

A method specifically designed for analysis of deposits from newsprint mills using recycled fiber was proposed by Guo and Douek.[151] As illustrated in Figure 16, it consists of sequential solvent extraction of a deposit with ethanol and chloroform. The ethanol extracts were redissolved in hexane and the hexane-soluble fraction was fractionated using solid-phase extraction as described by Sweeney.[130] Several techniques, including Fourier transform infrared (FTIR) spectrometry, GC, NMR, and thermogravimetric analysis (TGA), were used to characterize both soluble and insoluble fractions. Metal soaps were also determined on a separate fraction of the deposit as described by Sitholé et al.[158] The advantage of this scheme is that it simplifies the identification of the major classes of compounds, including wood resins, hydrocarbon oil, and extractable and nonextractable polymeric material, by separating them into individual fractions on the basis of differences in solubilities in different solvents. However, certain types of compounds, particularly polymers with wide molecular weight distributions, could be present in more than one fraction. One particular fraction obtained from the ethanol extract of a paper-machine dryer fabric deposit was readily identified as PVA by comparing its FTIR spectrum with that of a standard PVA sample (Figure 17a and b).[159]

A complete scheme for separating dissolved and colloidal substances, including stickies, from recycled waste paper was recently described by Holmbom.[160] After centrifuging to remove fibers and other particles, one portion of the sample was analyzed for carbohydrates, uronic acids, and pectins, using methanolysis, followed by GC of silylated derivatives. A second portion was extracted with MTBE. The MTBE phase was analyzed by GC for various lipophilic compounds and lignans. For

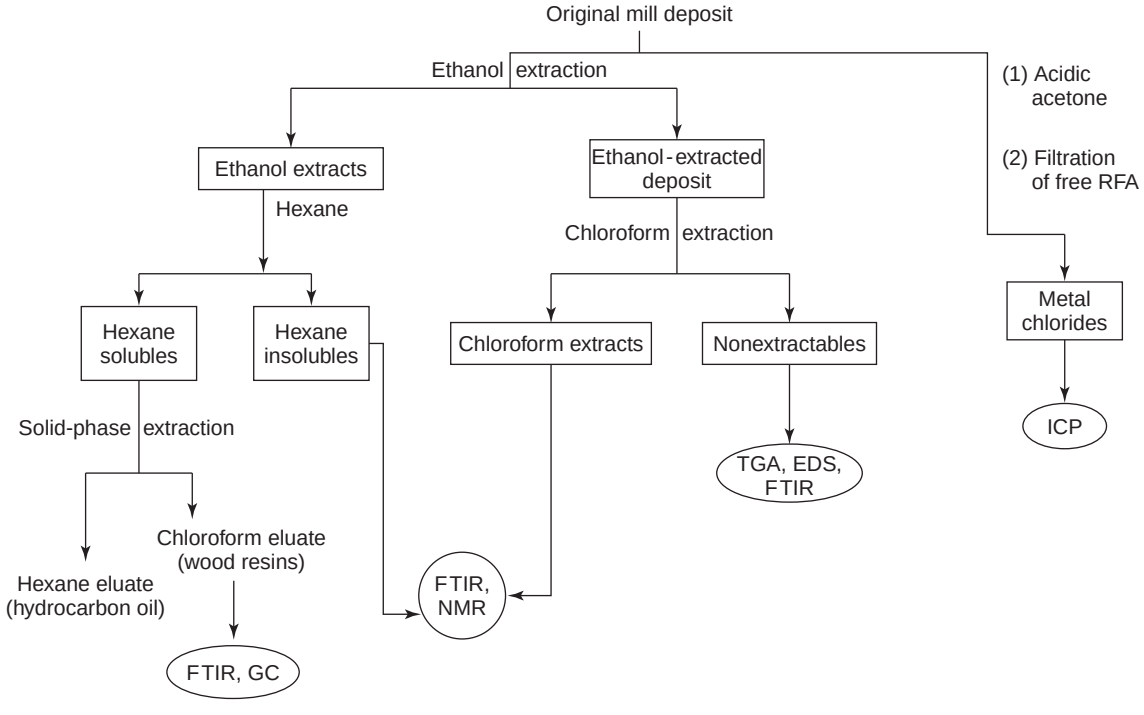

Figure 16 General scheme for analysis of deposits from mills using recycled fiber.[151] (Reproduced by permission of PAPTAC/TAPPI from X.-Y. Guo, M. Douek, *J. Pulp Paper Sci.*, **22**(11), J431–J439 (1996).)

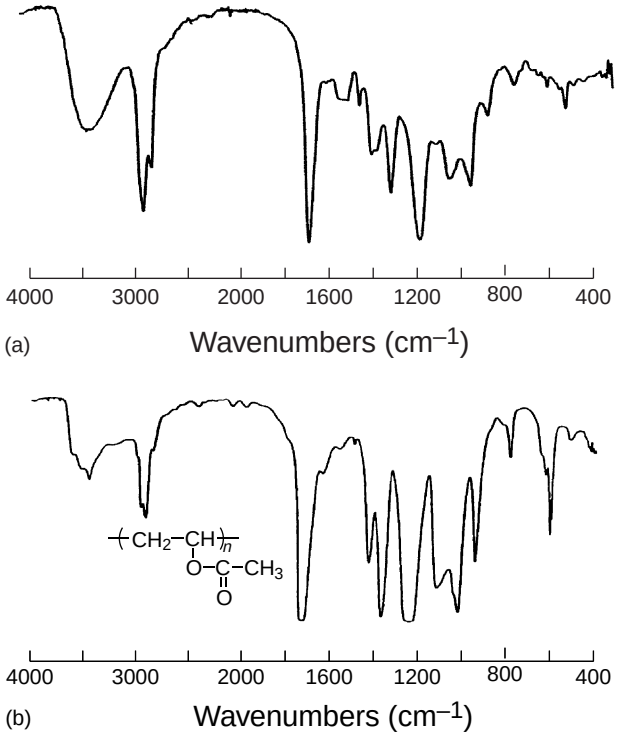

Figure 17 FTIR spectra of one of the fractions from the ethanol extract of a paper-machine dryer fabric (a) and of PVA (b).[159] (Reproduced by permission of TAPPI Press from M. Douek, X.-Y. Guo, *Proceedings of TAPPI 1997 Recycling Symposium*, TAPPI Press, Atlanta, GA, 313–330, 1997.)

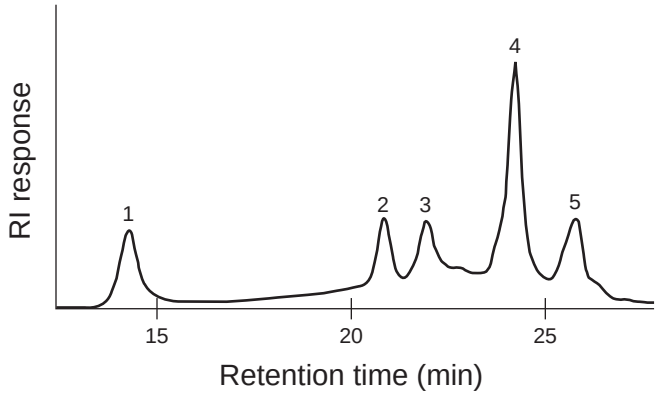

Figure 18 HPLC/SEC of THF extractives of a deinked pulp.[160] Main components of peaks: synthetic polymers (1), triglycerides (2), steryl esters (3), fatty acids (4), resin acids (5). (Reproduced by permission of PIRA International from B. Holmbom, *Proceedings of 1997 PIRA Wet End Chemistry Conference and COST Workshop*, 1–12.)

the analysis of stickies, a pulp, paper, or process liquid sample was freeze-dried and refluxed in THF, filtered and the extract was then fractionated by HPLC/SEC. A typical chromatogram of THF extracts of a deinked pulp sample is shown in Figure 18. Each SEC fraction was then analyzed by py/GC/MS. Pyrolysis components in the polymeric fraction are shown in Table 10.

Analytical schemes have also been devised for the analysis of metal soaps[158,161] and defoamers[162] in pitch deposits. Calcium and aluminum soaps of fatty acids, because of their highly sticky nature, can enhance the deposition of other components in pitch deposits, and are therefore considered to play an important role in the deposition process. The scheme developed by Dorris et al.[161] for analysis of calcium soaps in kraft mill brownstock pitch deposits involves extracting the deposit with acetone to remove free acids and neutrals. The acetone-extracted deposits were then Soxhlet extracted with chloroform to remove calcium soaps. The chloroform extracts were subsequently filtered to separate the soaps from material insoluble in cold chloroform. The types and amounts of acids bound to calcium ions were also determined by hydrolysis of calcium soaps followed by analysis of the liberated acids by capillary GC. This scheme was modified for aluminum soaps because the latter are partially soluble in hot acetone and chloroform.[158] In this case, the initial extraction with acetone was conducted at room temperature to remove wood resin and hydrocarbon oil. This was followed by refluxing with acidic acetone to yield metal-bound acids and aluminum chloride precipitate. The aluminum was then determined by AAS, and the acids were analyzed by GC and GC/MS. The complete scheme, shown in Figure 19, was applied to the determination of aluminum soaps in newsprint mill deposits. For the analysis of amide defoamers,[162] another potential contributor to pitch deposits, the approach taken was to extract the deposit with chloroform, followed by isolation of the fatty acid diamides, the active ingredient of amide defoamers, by recrystallization and filtration. The mineral oil, one of the main ingredients of defoamers, was separated by adsorption chromatography on silica gel. The fractionation scheme for isolating mineral oil and diamides from pitch deposits is illustrated in Figure 20.

The use of ^{13}C-NMR for the analysis of pitch extracts has been reported by Suckling and Ede.[163] The method can be used to determine quantitatively the amounts of triglycerides, fatty and resin acids, and esters. The technique is nondestructive and allows one to assess whether other organic components such as defoamers and retention aids are present. However, it requires more than 1 g of sample for analysis, and only material completely soluble in the solvent used for NMR spectroscopy can be analyzed. Proton NMR spectroscopy has also been used for the characterization of synthetic polymers in a pitch deposit[164] and, in combination with GC and SEC, for the analysis of so-called white pitch deposits, originating from coated papers.[165]

Table 10 Pyrolysis components (and fragment ions shown in brackets) used for identification and quantification of polymers by py/GC/MS.[160] (Reproduced by permission of PIRA International from B. Holmbom, *Proceedings of 1997 PIRA Wet End Chemistry Conference and COST Workshop*, 1–12)

Polymer	Primary pyrolysis component	Secondary component
EVA[a]	1-Decene (55, 140)	1-Undecene (55, 154)
SB[b]-latex, SIS[c], SBS[d]	Styrene (104)	4-Phenylcyclohexene (104, 158)
Butyl acrylate latex	Butyl acrylate (55, 56)	
PVA, fresh	Naphthalene (128)	Benzaldehyde (106)
PVA, cross-linked	2-Butenal (70)	2,4-Hexadienal (81, 96)

[a] EVA, ethylene vinyl acetate.
[b] SB, styrene butadiene.
[c] SIS, styrene isoprene styrene.
[d] SBS, styrene butadiene styrene.

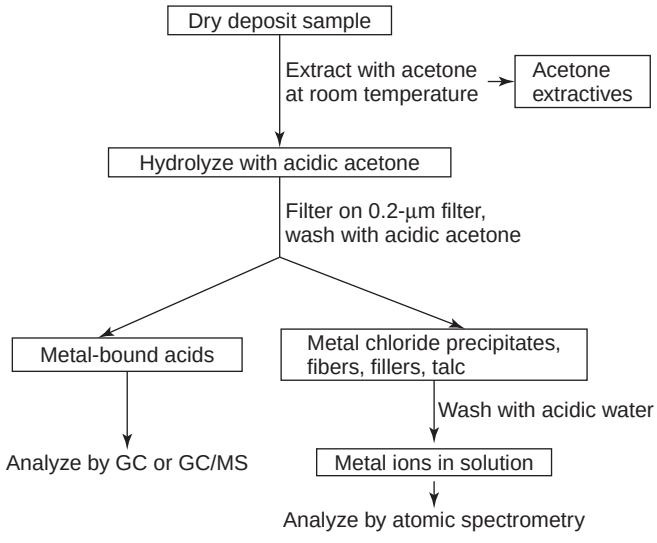

Figure 19 Schematic diagram of the analytical scheme for determination of metal soaps in fresh pitch deposits.[158] (Reproduced by permission of Mentor Communications AB from B.B. Sitholé, T.N. Tran, L.H. Allen, *Nordic Pulp Pap. Res., J.*, **11**(2), 64–69 (1996).)

Another relatively new hyphenated technique for identifying the source of contaminants in mill samples is thermogravimetry (TG)/IR/MS.[166] The loss of weight by TG allows one to determine the proportion of organics in the sample, and analysis of the volatile material can be achieved by IR spectroscopy, followed by MS. This combination of techniques was used to determine a variety of contaminants in mill deposits, including common sizing agents, AKD and ASA, and a residual surfactant, a dimethyl-substituted amine, which was interfering with the sizing process. The technique is relatively rapid and requires only a few milligrams of sample with minimal preparation prior to analysis.

py/GC is a particularly well-suited technique for identification of synthetic polymers in complex mixtures.[167,168] It requires as little as 1 mg of sample for analysis and no sample preparation. A fingerprint of a particular polymer can often be obtained from its fragmentation pattern. This is illustrated in Figure 21(a) and (b), which shows pyrolysis chromatograms of a contaminant in the dryer section of a tissue mill and of a PVA standard.[168] The similarity between the two chromatograms is evident, and suggests that the contaminant originates from PVA which was introduced with the recycled wastepaper. Even complex mixtures can be analyzed by careful manipulation of the pyrolysis and chromatographic conditions. However, in such cases, identification of the various components is best carried out by combining MS with py/GC, as described by Hardell[169] and del Rio et al.[170]

Analysis of spots and specks in pulp and paper can also be conveniently carried out by IR spectroscopy. The technique is rapid, and can be performed on sub-millimeter specks with little or no sample preparation. It is especially useful for identifying relatively pure contaminants or those consisting of a few major components. Leclerc et al.[171] showed the application of diffuse reflectance infrared Fourier transform (DRIFT) spectrometry to the characterization of plastic and other contaminants, including polyethylene, polypropylene, and latexes in pulp, and more recently to the identification of wood resins, metal soaps, talc, clay, and defoamers in pitch deposits and dirt specks from market kraft pulp.[172] The use of a miniature diamond anvil cell is particularly convenient for FTIR analysis, in the transmission mode, of dirt specks and spots, as described by Douek and Guo.[159]

9 CONCLUSION

This overview of the major types of chemical analyses conducted on pulp and paper matrices is by no means exhaustive. Owing to space limitations, only a relatively few publications, representative of each area of pulp

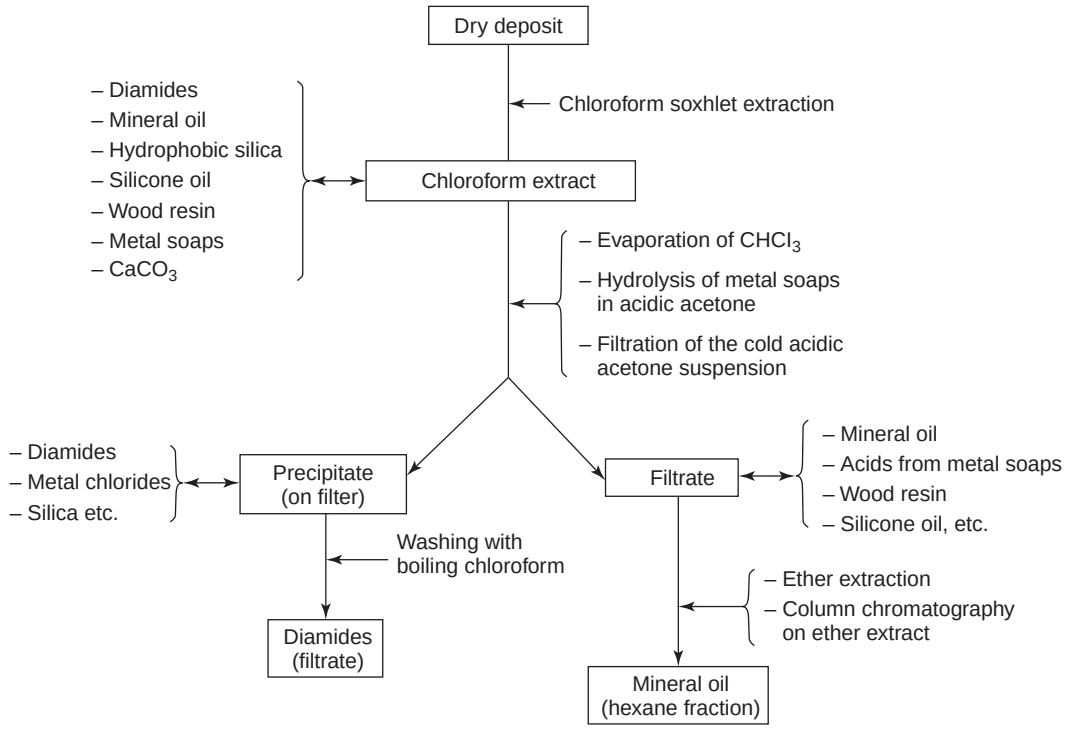

Figure 20 Fractionation scheme for isolating mineral oil and diamides from a pitch deposit.[162] (Reproduced by permission of PAPTAC from G.M. Dorris, M. Douek, L.H. Allen, *J. Pulp Pap. Sci.*, **11**(5), J149–J154 (1985).)

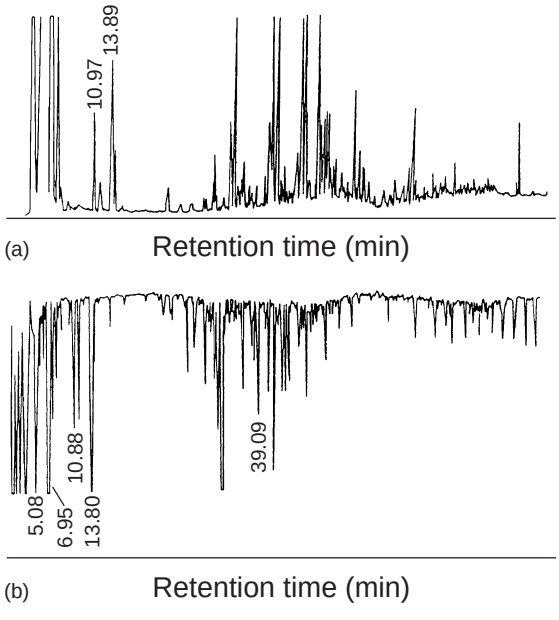

Figure 21 Pyrolysis chromatograms showing the presence of PVA in streak marks on paper from a tissue mill: (a) sample of contaminating material; (b) PVA standard.[168] (Reproduced by permission of PAPTAC/TAPPI from B.B. Sitholé, L.H. Allen, *J. Pulp Pap. Sci.*, **20**(6), J168–J172 (1994).)

and paper analysis, could be cited. The intention was to provide the reader with a sense of the broad scope of analytical techniques and instrumentation used in the pulp and paper field. The analyses described in this article range, in fact, from the traditional wet-chemical and instrumental methods designed primarily for routine testing in mills to highly sophisticated techniques used mainly for research purposes. As new analytical developments and measurement tools become available, some of these will no doubt continue to find their way into pulp and paper mills and research laboratories.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AKD	Alkylketene Dimers
AOX	Adsorbable Organic Halogens
APPITA	Australian Pulp and Paper Industry Technical Association
ASA	Alkenylsuccinic Anhydride
ASE™	Accelerated Solvent Extraction
ASTM	American Society for Testing and Materials
BOD	Biochemical Oxygen Demand
BPBMA	British Paper and Board Makers Association
CE	Capillary Electrophoresis

CED	Cupriethylenediamine
CMP	Chemimechanical Pulp
COD	Chemical Oxygen Demand
CTC	Cellulose Tricarbanilate
CTMP	Chemithermomechanical Pulp
CZE	Capillary Zone Electrophoresis
DIN	Deutsche Institut Für Normen
DP	Degree of Polymerization
DRIFT	Diffuse Reflectance Infrared Fourier Transform
EA	Effective Alkali
ECF	Elemental Chlorine-free
EDS	Energy-dispersive Spectroscopy
ESCA	Electron Spectroscopy for Chemical Analysis
EVA	Ethylene Vinyl Acetate
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GAC	Granular Activated Carbon
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GPC	Gel Permeation Chromatography
HPAEC	High-performance Anion-exchange Chromatography
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
IR	Infrared
ISO	International Standardization Organization
MALDI/TOF	Matrix-assisted Laser Desorption/ Ionization Time-of-flight
MS	Mass Spectrometry
MTBE	Methyl *tert*-Butyl Ether
NAA	Neutron Activation Analysis
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
NSSC	Neutral Sulfite Semichemical
PAD	Pulsed Amperometric Detection
PAPTAC	Pulp and Paper Technical Association of Canada
PVA	Poly(vinyl acetate)
py/GC	Pyrolysis/Gas Chromatography
RFA	Resin and Fatty Acids
RI	Refractive Index
RMP	Refiner Mechanical Pulp
SB	Styrene Butadiene
SBS	Styrene Butadiene Styrene
SCAN	Scandinavian Pulp, Paper and Board Testing Committee
SEC	Size-exclusion Chromatography
SEM	Scanning Electron Microscopy
SFE	Supercritical Fluid Extraction
SGW	Stone Groundwood
SIMS	Secondary Ion Mass Spectrometry
SIS	Styrene Isoprene Styrene
SMA	Styrene–Maleic Anhydride
TAPPI	Technical Association of the Pulp and Paper Industry
TCF	Totally Chlorine-free
TFA	Trifluoroacetic Acid
TG	Thermogravimetry
TGA	Thermogravimetric Analysis
THF	Tetrahydrofuran
TIC	Total Ion Chromatogram
TLC	Thin-layer Chromatography
TMP	Thermomechanical Pulp
TMS	Trimethylsilyl
TOC	Total Organic Carbon
TPP	Triphenylphosphine
TPPO	Triphenylphosphine Oxide
TPPS	Triphenylphosphine Sulfide
TTAOH	Tetradecyltrimethyl-ammonium Hydroxide
UV	Ultraviolet
XRF	X-ray Fluorescence

RELATED ARTICLES

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction • Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis • Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Pulp and Paper cont'd **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry • X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

REFERENCES

1. B.L. Browning, *Methods of Wood Chemistry*, John Wiley & Sons, New York, Vols I and II, 1967.
2. E. Sjöström, R. Alén (eds.), *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*, Springer Series in Wood Science, Springer, Berlin, 1998.
3. D.S. Argyropoulos (ed.), *Advances in Lignocellulosics Characterization*, TAPPI Press, Atlanta, GA, 1999.
4. E. Hägglund, *Chemistry of Wood*, Academic Press, New York, 1951.
5. S.A. Rydholm, *Pulping Processes*, John Wiley & Sons, London, 1965.

6. J.P. Casey (ed.), *Pulp and Paper Chemistry and Chemical Technology*, 3rd edition, Wiley-Interscience, New York, Vols I–IV, 1980.
7. M.J. Kocurek, F. Stevens (eds.), *Pulp and Paper Manufacture*, Joint Textbook Committee of the Paper Industry, Montreal, Vols 1–10, 1983.
8. G.A. Smook (ed.), *Handbook for Pulp and Paper Technologists*, 2nd edition, Angus Wilde Publications Inc., Vancouver, 1992.
9. H. Sihtola, B. Kyrklund, L. Laamanen, I. Palenius, 'Comparison and Conversion of Viscosity and DP-Values Determined by Different Methods', *Pap. Timber*, **45**(4a), 225–232 (1963).
10. J.M. Willis, F. Commodari, D.A.I. Goring, 'The Spectrophotometric Determination of Carbohydrate in Leachates from Ultra High Yield Sulphite Pulp', *J. Pulp Pap. Sci.*, **11**(3), J63–J68 (1985).
11. D. Fengel, G. Wegener, 'Hydrolysis of Polysaccharides with Trifluoroacetic Acid and its Application to Rapid Wood and Pulp Analysis', *Adv. Chem. Ser.*, **181**, 145–158 (1979).
12. J.F. Saeman, W.E. Moore, R.L. Mitchell, M.A. Milletett, 'Techniques for the Determination of Pulp Constituents by Quantitative Paper Chromatography', *TAPPI*, **37**(8), 336–343 (1954).
13. L.G. Borchardt, C.V. Piper, 'A Gas Chromatographic Method for Carbohydrates as Alditol Acetates', *TAPPI*, **53**(2), 257–260 (1970).
14. A.B. Blakeney, P.J. Harris, R.J. Henry, B.A. Stone, 'A Simple and Rapid Preparation of Alditol Acetates for Monosaccharide Analysis', *Carbohydr. Res.*, **113**, 291–299 (1983).
15. J.M. MacLeod, J.V. Benko, 'Sugars and Sugar Acids in Lignosulphonate Products', *J. Wood Chem. Technol.*, **2**(2), 207–219 (1982).
16. F.E. Wentz, A.D. Marcy, M.J. Gray, 'Analysis of Wood Sugars in Pulp and Paper Industry Samples by HPLC', *J. Chromatogr. Sci.*, **20**, 349–352 (1982).
17. G. Bonn, R. Pecina, E. Burtscher, O. Bobleter, 'Separation of Wood Degradation Products by High-performance Liquid Chromatography', *J. Chromatogr.*, **287**, 215–221 (1984).
18. W.E. Kaar, L.G. Cool, M.M. Merriman, D.L. Brink, 'The Complete Analysis of Wood Polysaccharides Using HPLC', *J. Wood Chem. Technol.*, **11**(4), 447–463 (1991).
19. R.C. Petterson, V.H. Schwandt, 'Wood Sugar Analysis by Anion Chromatography', *J. Wood Chem. Technol.*, **11**(4), 495–501 (1991).
20. J. Sullivan, M. Douek, 'Determination of Carbohydrates in Wood, Pulp, and Process Liquor Samples by High-performance Anion-exchange Chromatography with Pulsed Amperometric Detection', *J. Chromatogr. A*, **671**, 339–350 (1994).
21. P.J. Wright, A.F.A. Wallis, 'Rapid Determination of Carbohydrates in Hardwoods by High Performance Anion Exchange Chromatography', *Holzforschung*, **50**, 518–524 (1996).
22. A.R. Mills, T.E. Timell, 'Constitution of Three Hemicelluloses from the Wood of Engelmann Spruce (*Picea Engelmanni*)', *Can. J. Chem.*, **41**, 1389–1395 (1963).
23. G.D. McGinnis, F. Shafizadeh, 'Cellulose and Hemicellulose', in *Pulp and Paper Chemistry and Chemical Technology*, 3rd edition, ed. J.P. Casey, Wiley-Interscience, New York, 33, Vol. I, 1980.
24. M. Tenkanen, T. Hausalo, M. Siika-aho, J. Buchert, 'Use of Enzymes in Combination with Anion Exchange Chromatography in the Analysis of Carbohydrate Composition of Kraft Pulps', in *Proceedings of the 8th International Symposium on Wood and Pulping Chemistry, Helsinki, Finland*, Gummerus Kirjapaino Oy, Jyväskylä, 189–194, Vol. 3, 1995.
25. A. Sundberg, K. Sundberg, C. Lollandr, B. Holmbom, 'Determination of Hemicelluloses and Pectins in Wood and Pulp Fibers by Acid Methanolysis and Gas Chromatography', *Nordic Pulp Pap. Res. J.*, **11**(4), 216–220 (1996).
26. A. Rydlund, O. Dahlman, 'Efficient Capillary Zone Electrophoretic Separation of Wood-derived Neutral and Acidic Mono- and Oligosaccharides', *J. Chromatogr. A*, **738**, 129–140 (1996).
27. L. Valtasaari, K. Saarela, 'Determination of Chain Length Distribution of Cellulose by Gel Permeation Chromatography Using the Tricarbanilate Derivative', *Pap. Timber*, **57**(1), 5–10 (1975).
28. L.R. Schroeder, F.C. Haigh, 'Cellulose and Wood Pulp Polysaccharides, Gel Permeation Chromatographic Analysis', *TAPPI*, **62**(10), 103–105 (1979).
29. B.F. Wood, C.G. Hill, 'The Effect of Precipitation on the Molecular Weight Distribution of Cellulose Tricarbanilate', *J. Appl. Polym. Sci.*, **32**, 3703–3712 (1986).
30. D.J. Harvey, 'Matrix-assisted Laser Desorption/Ionisation Mass Spectrometry of Oligosaccharides and Glycoconjugates', *J. Chromatogr. A*, **720**, 429–446 (1996).
31. O. Dahlman, A. Rydlund, A. Lindquist, 'Characterization of Carbohydrates from Chemical Pulps Using Capillary Electrophoresis and MALDI/TOF/MS', in *Proceedings of the 9th International Symposium on Wood and Pulping Chemistry, Montreal, Canada*, Technical Section, CPPA, L5-1–L5-4, 1997.
32. K.V. Sarkanen, C.H. Ludwig (eds.), *Lignins, Occurrence, Formation, Structure and Reactions*, John Wiley & Sons, New York, 1971.
33. S.Y. Lin, C.W. Dence (eds.), *Methods in Lignin Chemistry*, Springer, Berlin, 1992.
34. M.J. Effland, 'Modified Procedure to Determine Acid-insoluble Lignin in Wood and Pulp', *TAPPI*, **60**(10), 143–144 (1977).
35. V. Berzins, J.E. Tasman, 'The Permanganate Consumption of Pulp Materials, II. The Kappa Number', *Pulp Pap. Mag. Can.*, **58**, 149–153 (1957).

36. J.E. Tasman, 'The Permanganate Consumption of Pulp Materials, IV. The Kappa Correction Coefficient', *Pulp Pap. Mag. Can.*, **60**(Convention Issue), 231–235 (1959).
37. J.D. McLean, 'Hypo Number', *Pulp Pap. Mag. Can.*, **66**, T103–T106 (1965).
38. B. Kyrklund, G. Strandell, 'Applicability of the Chlorine Number for Evaluation of the Lignin Content of Pulp', *Pap. Timber*, **51**(4a), 299–305 (1969).
39. E. Kubulnieks, S.-O. Lundqvist, T. Petterson, 'The STFI OPTI-Kappa Analyzer – Applications and Accuracy', *TAPPI J.*, **70**, 38–42 (1987).
40. J. Marton, 'Determination of Lignin in Small Pulp and Paper Samples Using the Acetyl Bromide Method', *TAPPI*, **50**(7), 335–337 (1967).
41. K. Iiyama, A.F.A. Wallis, 'An Improved Acetyl Bromide Procedure for Determining Lignin in Woods and Wood Pulps', *J. Wood Sci. Technol.*, **22**, 271–280 (1988).
42. O. Faix, R. Patt, O. Beinhoff, 'Basics of Fourier-transform Infrared Spectroscopy (FTIR) and its Use in Preparation and Analysis of Chemical Pulps', *Papier*, **41**(12), 657–663 (1987).
43. S. Backa, A. Brolin, 'Determination of Pulp Characteristics by Diffuse Reflectance FTIR', *TAPPI J.*, **74**, 218–226 (1991).
44. S. Banerjee, 'Interpreting Multicomponent Infrared Spectra', *TAPPI J.*, **75**(8), 147–149 (1992).
45. A.J. Michell, 'Usefulness of Fourier Transform–Infrared Difference Spectroscopy for Studying the Reactions of Wood During Pulping', *Cellulose Chem. Technol.*, **22**, 105–113 (1988).
46. M.D. Birkett, M.J.T. Gambino, 'Estimation of Pulp Kappa Number with Near-infrared Spectroscopy', *TAPPI J.*, **72**(9), 193–197 (1989).
47. D.B. Easty, S.A. Berben, F.A. DeThomas, P.J. Brimmer, 'Near-infrared Spectroscopy for the Analysis of Wood Pulp: Quantifying Hardwood–Softwood Mixtures and Estimating Lignin Content', *TAPPI J.*, **73**(10), 257–261 (1990).
48. L. Wallbäcks, U. Edlund, B. Nordén, 'Multivariate Characterization of Pulp Using Solid-state ^{13}C NMR, FTIR, and NIR', *TAPPI J.*, **74**(10), 201–206 (1991).
49. C.-L. Chen, Y.-Z. Lai, C.W. Dence, R.P. Beatson, 'Functional Group Analysis', in *Methods in Lignin Chemistry*, eds. S.Y. Lin, C.W. Dence, Springer, Berlin, Chapter 7, 1992.
50. S. Katz, R.P. Beatson, A.M. Scallan, 'The Determination of Strong and Weak Acidic Groups in Sulfite Pulps', *Sven. Papperstidn.*, **87**, R48–R53 (1984).
51. A.S. Wexler, 'Characterization of Lignosulfonates by Ultraviolet Spectrometry', *Anal. Chem.*, **36**(1), 213–221 (1964).
52. W.E. Hillis (ed.), *Wood Extractives and Their Significance to the Pulp and Paper Industries*, Academic Press, London, 1962.
53. B.B. Sitholé, 'Modern Methods for the Analysis of Extractives from Wood and Pulp: A Review', *Appita*, **45**(4), 260–264 (1992).
54. B.B. Sitholé, P. Vollstaedt, L.H. Allen, 'Comparison of Soxtec and Soxhlet Systems for Determining Extractives Content', *TAPPI J.*, **74**(11), 187–191 (1991).
55. A.J. Sequeira, L.T. Taylor, 'Supercritical Fluid Extraction of Wood Pulps', *J. Chromatogr. Sci.*, **30**, 405–408 (1992).
56. B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: A Technique for Sample Preparation', *Anal. Chem.*, **68**(6), 1033–1039 (1996).
57. B.E. Richter, J.L. Ezzell, D. Felix, K.A. Roberts, D.W. Later, 'An Accelerated Solvent Extraction System for the Rapid Preparation of Environmental Organic Compounds in Soil', *Am. Lab.*, **27**(4), 24–28 (1995).
58. J.L. Ezzell, B.E. Richter, W.D. Felix, S.R. Black, J.E. Meikle, 'A Comparison of Accelerated Solvent Extraction with Conventional Solvent Extraction for Organophosphorus Pesticides and Herbicides', *LC/GC*, **13**(5), 390–398 (1995).
59. 'Extraction of Polychlorinated Dibenzo-*p*-dioxins and Polychlorinated Dibenzofurans from Environmental Samples Using Accelerated Solvent Extraction (ASE™)', Dionex Application Note 323, Dionex, Sunnyvale, CA, 1996.
60. R. Ekman, B. Holmbom, 'Analysis by Gas Chromatography of the Wood Extractives in Pulp and Water Samples from Mechanical Pulping of Spruce', *Nordic Pulp Pap. Res. J.*, **7**(1), 16–24 (1989).
61. B.B. Sitholé, J.L. Sullivan, L.H. Allen, 'Identification and Quantitation of Acetone Extractives of Wood and Bark by Ion Exchange and Capillary GC with a Spreadsheet Program', *Holzforschung*, **46**(5), 409–416 (1992).
62. I.D. Suckling, S.S. Gallagher, R.M. Ede, 'A New Method for Softwood Extractives Analysis Using High Performance Liquid Chromatography', *Holzforschung*, **44**(5), 339–345 (1990).
63. M. Sandström, M.A. Norborg, A. Ericsson, 'Applications of Thin-layer Chromatography to Process Control in the Pulp and Paper Field', *J. Chromatogr. A*, **730**, 373–379 (1996).
64. G.M. Dorris, M. Douek, L.H. Allen, 'Operating Variables in the Analysis of Tall Oil Acids by Capillary Chromatography', *J. Am. Oil Chem. Soc.*, **59**(11), 494–500 (1982).
65. B.B. Sitholé, 'A Rapid Spectrophotometric Procedure for the Determination of Total Resin and Fatty Acids in Pulp and Paper Matrices', *TAPPI J.*, **76**(10), 123–127 (1993).
66. M. Douek, L.H. Allen, 'The Distribution of Calcium in Kraft Mill Brownstock Systems', *Sven. Papperstidn.*, **83**(15), 425–429 (1980).

67. V.A. D'Souza, V.C. Hand, R.L. Schaefer, 'Concentration of Metals in Recycled Papers', Preprints of 1998 TAPPI Recycling Symposium, New Orleans, LA, 163–175, 1998.
68. L.D. Markham, J.R.G. Bryce, 'Formation of Calcium Carbonate Scale in a Continuous Digester', *TAPPI*, **63**(4), 125–127 (1980).
69. T.M. Grace, D.G. Sachs, H.J. Grady, 'Determination of the Inorganic Composition of Alkaline Black Liquors', *TAPPI*, **60**(4), 122–125 (1977).
70. T. Fossum, N. Hartler, J. Libert, 'The Inorganic Content of Wood', *Sven. Papperstidn.*, **75**(8), 305–309 (1972).
71. A. Wong, 'Problems of Wood Inorganics in the Kraft Pulping of Chips with Bark', Preprints of 1981 Annual Meeting CPPA Technical Section, Montreal, A203–A216, 1981.
72. L.H. Phifer, J.B. Maginnis, 'Dry Ashing of Pulp and Factors which Influence it', *TAPPI*, **43**(1), 38–44 (1960).
73. 'Perchloric Acid ($HClO_4$)', National Safety Council Data Sheet 311, Revision A, National Safety Council, Chicago, 1971.
74. K. Sirén, 'Analytical Methods for Determination of Metals, Silicon and Phosphorus in Kraft Mill Samples', in *Proceedings of 1998 TAPPI International Chemical Recovery Conference*, TAPPI Press, Atlanta, GA, 1113–1119, 1998.
75. Y.P. Kalra, D.G. Maynard, F.G. Radford, 'Microwave Digestion of Tree Foliage for Multi-element Analysis', *Can. J. For. Res.*, **19**, 981–985 (1989).
76. P.O. Bethge, R. Rådeström, 'Determination of Trace Metals in Pulp by Atomic Absorption Spectrophotometry', *Sven. Papperstidn.*, **69**(22), 772–776 (1966).
77. L.L. Landucci, N. Sanyer, 'Influence of Transition Metals in Oxygen Pulping', *TAPPI*, **58**(2), 60–63 (1975).
78. N.M. Morris, V.W. Tripp, 'Determination of Metals in Wood Pulps with the Graphite Furnace', *TAPPI*, **59**(4), 146–147 (1976).
79. V. Kocman, P. Bruno, 'Accurate Determination of Six Common Fillers in Fine Papers, Newsprint, and Recycled Stock by X-ray Fluorescence Spectrometry', *TAPPI J.*, **78**(4), 129–134 (1995).
80. O. Keitaanniemi, N.-E. Virkola, 'Amounts and Behavior of Certain Chemical Elements in Kraft Pulp Manufacture: Results of a Mill Scale Study', *Pap. Timber*, **60**(9), 507–522 (1978).
81. J.H.E. Bailey, D.W. Reeve, 'Spatial Distribution of Trace Elements in Black Spruce by Imaging Microprobe Secondary Ion Mass Spectrometry', *J. Pulp Pap. Sci.*, **20**(3), J83–J86 (1994).
82. D.L. Ormerod, G.C. Simon, D.L. Gibbon, 'Paper Mill Additive and Contaminant Distribution Mapping Using Energy Dispersive X-ray Analysis', *TAPPI J.*, **71**(4), 211–214 (1988).
83. W.E. Hillis, D. de Silva, 'Inorganic Extraneous Constituents of Wood', *Holzforschung*, **33**, 47–53 (1979).
84. O.V. Ingruber, 'Sulfite Pulping Cooking Liquor and the Four Bases', in *Pulp and Paper Manufacture, Part I, Sulfite Science and Technology*, 3rd edition, eds. O.V. Ingruber, M.J. Kocurek, A. Wong, Joint Textbook Committee of the Paper Industry, Montreal, 3–23, Vol. 4, 1985.
85. T.M. Grace, B. Leopold, E.W. Malcolm, M.J. Kocurek (eds.), *Pulp and Paper Manufacture, Part I, Alkaline Pulping*, 3rd edition, Joint Textbook Committee of the Paper Industry, Montreal, Vol. 5, 1989.
86. T.M. Grace, B. Leopold, E.W. Malcolm, M.J. Kocurek (eds.), *Pulp and Paper Manufacture, Part II, Kraft Liquors*, 3rd edition, Joint Textbook Committee of the Paper Industry, Montreal, Vol. 5, 1989.
87. E. Milanova, G.M. Dorris, 'Determination of Residual Alkali in Black Liquors', *Nordic Pulp Pap. Res. J.*, **9**(1), 4–15 (1994).
88. K.L. McDonald, 'Rapid Determination of Kraft Black Liquor Solids', *TAPPI*, **62**(1), 80–81 (1979).
89. J. Papp, 'Determination of Sulphur Compounds in Alkaline Pulping Liquors with a Sulphide Ion-selective Electrode', *Sven. Papperstidn.*, **74**(10), 310–315 (1971).
90. D.L. Noel, 'Sulfur Compounds in Kraft Pulping Liquor – Determination by Pulse Polarography', *TAPPI*, **61**(5), 73–76 (1978).
91. G.O. Franklin, A.W. Fitchett, 'Fast Chemical Characterization of Pulping and Bleaching Process Liquors by Ion Chromatography', *Pulp Pap. Can.*, **83**(10), T271–T275 (1982).
92. D.B. Easty, M.L. Borchardt, A.A. Webb, 'Analysis of Pulping Liquors by Ion Chromatography: Evaluation and Validation', *Pap. Timber*, **67**(9), 501–506 (1985).
93. J. Krishnagopalan, M. Hill, A.L. Fricke, 'Chromatographic Analysis of Kraft Liquor Anions', *TAPPI J.*, **68**(9), 108–112 (1985).
94. D.B. Easty, J.E. Johnson, 'Recent Progress in Ion Chromatographic Analysis of Pulping Liquors: Determination of Sulfide and Sulfate', *TAPPI J.*, **70**(3), 109–111 (1987).
95. S.K. Murarka, N.S. Fairchild, 'Ion Chromatography: Is It the 'Now' Technique for Monitoring Process Streams in a Paper Mill? Part II: Selected Applications', *Pulp Pap. Can.*, **88**(12), T441–T446 (1987).
96. M. Douek, J. Sullivan, J. Ing, 'Effect of Sample-related Sources of Error on the Analysis of Pulp and Paper Mill Process Liquids by Ion Chromatography: Part I – Determination of Sulphur Species', *J. Wood Chem. Technol.*, **13**(3), 439–462 (1993).
97. S.M. Masselter, A.J. Zemann, G.K. Bonn, 'Determination of Inorganic Anions in Kraft Pulping Liquors by Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **19**, 131–136 (1996).
98. D. Volgger, A. Zemann, G. Bonn, 'Determination of Phenols, Inorganic Anions, and Carboxylic Acids in Kraft Black Liquors by Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **21**, 3–10 (1998).

99. J. Sullivan, M. Douek, 'Rapid Analysis of Inorganic and Organic Anions in Kraft Pulping Liquors by Capillary Zone Electrophoresis', in *Proceedings of the 1999 TAPPI Pulping Conference*, TAPPI Press, Atlanta, GA, 15, 1999.
100. S.-J. Liaw, G.A. Krishnagopalan, 'On-line Measurement of Sulfide and Alkali Concentrations During Kraft Pulping', *TAPPI J.*, **75**(9), 219–224 (1992).
101. D.F. Leclerc, R.M. Hogikyan, 'Rapid Determination of Effective Alkali and Dead-load Concentrations in Kraft Liquors by Attenuated Total Reflectance Infrared Spectrometry', *J. Pulp Pap. Sci.*, **21**(7), J231–J237 (1995).
102. T.P. Trung, D.F. Leclerc, C. Roberts, D. Osmond, J.A. McIlwain, 'Infrared Technology to Analyse Kraft Processes Used During Pulping and Caustification: Part I – Calibration of the Sensor in the Pilot Plant and its Performance in the Mill with a Continuous Digester', in *Proceedings of the 1996 CPPA Conference Technologique Estivale, May 1996*, Technical Section, CPPA, Montreal, 35–42, 1996.
103. A. Teder, 'Spectrophotometric Determination of Polysulfide Excess Sulfur in Aqueous Solutions', *Sven. Papperstidn.*, **70**(6), 197–200 (1967).
104. L.G. Borchardt, D.B. Easty, 'Gas Chromatographic Determination of Elemental and Polysulphide Sulfur in Kraft Liquors', *J. Chromatogr.*, **299**, 471–476 (1984).
105. G.M. Dorris, V.C. Uloth, 'Analysis of Oxidized White Liquors: Determination of Polysulfides by Gravimetry', *J. Pulp Pap. Sci.*, **20**(8), J211–214 (1994).
106. G.M. Dorris, V.C. Uloth, 'Analysis of Oxidized White Liquors: Potentiometric Titrations for the Determination of Polysulfides and Sulfoxy Anions', *J. Pulp Pap. Sci.*, **20**(9), J242–J248 (1994).
107. R.P. Singh (ed.), 'The Bleaching of Pulp', 3rd edition, TAPPI Press, Atlanta, GA, 1979.
108. V. Lorås, 'Bleaching', in *Pulp and Paper Chemistry and Chemical Technology*, 3rd edition, ed. J.P. Casey, Wiley-Interscience, New York, Chapter 5, Vol. I, 1980.
109. C.W. Dence, D.W. Reeve (eds.), *Pulp Bleaching – Principles and Practice*, TAPPI Press, Atlanta, GA, 1996.
110. J. Sullivan, M. Douek, 'Determination of Inorganic Chlorine Species in Kraft Mill Bleach Effluents by Ion Chromatography', *J. Chromatogr. A*, **804**, 113–121 (1998).
111. *Standard Methods for the Analysis of Water and Wastewater*, 20th edition, American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, 1998.
112. W.E. Razzell, 'Pitfalls and Procedures in BOD Tests', *Pulp Pap. Mag. Can.*, **73**(7), T186–T189 (1972).
113. P.J. Leblanc, 'A Review of the Biochemical Oxygen Demand (BOD_5) Test', *Pulp Pap. Mag. Can.*, **75**(3), T85–T90 (1974).
114. V.T. Stack, 'Critique of the Standard Biochemical Oxygen Demand Method', *J. Test. Eval.*, **12**(4), 246–250 (1984).
115. R. Stalford, '24 Hour BOD Analysis', in *TAPPI Proceedings, 1993 Environmental Conference, Boston, MA*, TAPPI Press, Atlanta, GA, 757–759, 1993.
116. J.J. Garceau, S.N. Lo, 'A Rapid Estimation of BOD_5 of Spent Sulfite Liquor', *TAPPI J.*, **61**(12), 85–86 (1978).
117. 'Biochemical Oxygen Demand – Respirometric Method', in *Standard Methods for the Analysis of Water and Wastewater*, 20th edition, American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, Method 5210D, 1998.
118. K. Riedel, K.-P. Lange, H.-J. Stein, M. Kühn, P. Ott, F. Scheller, 'A Microbial Sensor for BOD', *Water Res.*, **24**(7), 883–887 (1990).
119. T.C. Tan, F. Li, K.G. Neoh, Y.K. Lee, 'Microbial Membrane-modified Dissolved Oxygen Probe for Rapid Biochemical Oxygen Demand Measurement', *Sens. Actuators B*, **8**, 167–172 (1992).
120. C. Preininger, I. Klimant, O.S. Wolfbeis, 'Optical Fiber Sensor for Biological Oxygen Demand', *Anal. Chem.*, **66**, 1841–1846 (1994).
121. 'Chemical Oxygen Demand – Closed Reflux, Colorimetric Method', in *Standard Methods for the Analysis of Water and Wastewater*, 20th edition, American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, Method 5220D, 1998.
122. W.F. Jardim, J.J.R. Rohwedder, 'Chemical Oxygen Demand (COD) Using Microwave Digestion', *Water Res.*, **23**(8), 1069–1071 (1989).
123. 'Total Organic Carbon (TOC)', in *Standard Methods for the Analysis of Water and Wastewater*, 20th edition, American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, Method 5310, 1998.
124. 'Dissolved Organic Halogen', in *Standard Methods for the Analysis of Water and Wastewater*, 20th edition, American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, Method 5320, 1998.
125. S.M. Odendahl, K.M. Weishar, D.W. Reeve, 'Chlorinated Organic Matter in Bleached Chemical Pulp Production. Part II: A Review of Measurement Techniques for Effluents', *Pulp Pap. Can.*, **91**(4), T136–T140 (1990).
126. J. Sullivan, M. Douek, 'Method and Sample-related Problems in the Determination of AOX in Effluents', *TAPPI J.*, **79**(8), 145–154 (1996).
127. R.H. Voss, A. Rapsomatiotis, 'An Improved Solvent-extraction Based Procedure for the Gas Chromatographic Analysis of Resin and Fatty Acids in Pulp Mill Effluents', *J. Chromatogr.*, **346**, 205–214 (1985).
128. F. Örsa, B. Holmbom, 'A Convenient Method for the Determination of Wood Extractives in Papermaking Process Waters and Effluents', *J. Pulp Pap. Sci.*, **20**(12), J361–J366 (1994).

129. S. Backa, A. Brolin, N.-O. Nilvebrant, 'Analyzing Wood Extractives from Process Streams by Using Alkaline Reversed-phase Extraction Followed by In Situ Ion Pairing', *TAPPI J.*, **72**(8), 139–144 (1989).
130. K.M. Sweeney, 'Solid-phase Extraction Techniques in the Pulp and Paper Industry', *TAPPI J.*, **71**(1), 137–140 (1988).
131. J.K. Jokela, M. Salkinoja-Salonen, 'Molecular Weight Distributions of Organic Halogens in Bleached Kraft Pulp Mill Effluents', *Environ. Sci. Technol.*, **26**, 1190–1197 (1992).
132. K. Lindström, F. Österberg, 'Characterization of the High Molecular Mass Chlorinated Matter in Spent Bleach Liquors (SBL). Part I. Alkaline SBL', *Holzforschung*, **38**, 201–212 (1984).
133. F. Österberg, K. Lindström, 'Characterization of the High Molecular Mass Chlorinated Matter in Spent Bleach Liquors (SBL). Part II. Acidic SBL', *Holzforschung*, **39**, 149–158 (1985).
134. O. Dahlman, I. Haglind, R. Mörck, F. de Sousa, L. Strömberg, 'Chemical Composition of Effluents from Chlorine Dioxide Bleaching of Kraft Pulps Before and After Secondary Effluent Treatment', in *Proceedings of the 1993 International Environmental Symposium*, EUCEPA, Vienna, 193–215, 1993.
135. O. Dahlman, A. Reimann, L. Strömberg, R. Mörck, 'On the Nature of High Molecular Weight Effluent Materials from Modern ECF- and TCF-bleaching', in *Proceedings of the 1994 International Pulp Bleaching Conference*, Technical Section, CPPA, Montreal, 123–132, 1994.
136. J. Thornton, R. Ekman, B. Holmbom, F. Örså, 'Polysaccharides Dissolved from Norway Spruce in Thermomechanical Pulping and Peroxide Bleaching', *J. Wood Chem. Technol.*, **14**(2), 159–175 (1994).
137. J. Thornton, R. Ekman, B. Holmbom, C. Eckerman, 'Release of Potential ''Anionic Trash'' in Peroxide Bleaching of Mechanical Pulp', *Pap. Timber*, **75**(6), 426–431 (1993).
138. J.W. Thornton, 'Enzymatic Degradation of Polygalacturonic Acids Released from Mechanical Pulp During Peroxide Bleaching', *TAPPI J.*, **77**(3), 161–167 (1994).
139. B.L. Browning, *Analysis of Paper*, 2nd edition, Marcel Dekker, New York, 1977.
140. S. Malton, K. Kuys, I. Parker, N. Vanderhoek, 'Measurement of Starch in Paper, and Application to Investigations of Cationic Starch Adsorption', in *Proceedings of the 49th Appita Annual General Conference*, Victoria, Australia, 67–74, 1995.
141. N. Birosel-Boettcher, 'Measurement of Starch in Paper by High-pressure Liquid Chromatography', *TAPPI J.*, **76**(3), 207–208 (1993).
142. T. Yano, H. Ohtani, S. Tsuge, T. Obokata, 'Determination of Wet-strength Resin in Paper by Pyrolysis/Gas Chromatography', *TAPPI J.*, **74**, 197–201 (1991).
143. T. Yano, H. Ohtani, S. Tsuge, T. Obokata, 'Determination of Neutral Sizing Agents in Paper by Pyrolysis/Gas Chromatography', *Analyst*, **117**, 849–852 (1992).
144. B. Sitholé, B. Ambayec, 'Identification and Determination of Surface Sizing Agents on Sized Papers', Preprints, 1999 PAPTAC 85th Annual Meeting, A95–A99, 1999.
145. T.D. Crockett, A.A. Webb, L.G. Borchardt, D.B. Easty, 'Identification of Polymers in Paper by Pyrolysis/Gas Chromatography', *J. Chromatogr.*, **407**, 330–339 (1987).
146. B.B. Sitholé, S. Nyarku, L.H. Allen, 'Chromatographic Methods for the Determination of Alkylketene Dimer in Pulp and Paper Matrices', *Analyst*, **120**, 1163–1169 (1995).
147. N. Nyarku, B.B. Sitholé, 'Analysis of Alkyl Ketene Dimer (AKD) by Potentiometric Titration', *Can. J. Chem.*, **72**(2), 274–278 (1994).
148. N.S. Adams, 'Analysis of Synthetic Materials in Paper and Paper Coatings Using GPC/IR', in *Proceedings of 1995 TAPPI Process and Product Quality Conference*, TAPPI Press, Atlanta, GA, 109–120, 1995.
149. G. Ström, G. Carlsson, A. Schulz, 'Chemical Composition of Coated Paper Surfaces Determined by Means of ESCA', *Nordic Pulp Pap. Res. J.*, **1**, 105–112 (1993).
150. J.S. Brinen, 'The Observation and Distribution of Organic Additives on Paper Surfaces Using Surface Spectroscopic Techniques', *Nordic Pulp Pap. Res. J.*, **1**, 123–129 (1993).
151. X.-Y. Guo, M. Douek, 'Analysis of Deposits/Stickies from Newsprint Mills Using Recycled Fiber', *J. Pulp Pap. Sci.*, **22**(11), J431–J439 (1996).
152. B.B. Sitholé, W. Kamutzki, 'Contaminants in Recycled Papers', in *Wood Resin, Deresination and Pitch Control*, eds. E.L. Back, L.H. Allen, TAPPI Press, Atlanta, GA, 1998.
153. B.B. Sitholé, 'Analysis of Pitch Deposits', in *Wood Resin, Deresination and Pitch Control*, eds. E.L. Back, L.H. Allen, TAPPI Press, Atlanta, GA, 1998.
154. J. Sjöström, B. Holmbom, 'A Scheme for Chemical Characterization of Deposits in Pulp and Paper Production', *Pap. Timber*, **70**(2), 151–156 (1988).
155. J. Sjöström, B. Holmbom, L. Wiklund, 'Chemical Characteristics of Paper Machine Deposits from Impurities in Deinked Pulp', *Nordic Pulp Pap. Res. J.*, **6**(4), 123–126 (1987).
156. J. Sjöström, B. Holmbom, 'Size Exclusion Chromatography of Deposits in Pulp and Paper Mills', *J. Chromatogr.*, **411**, 363–370 (1987).
157. D.G. Greer, 'Analysis of Pulp and Paper Mill Deposits by Thin-layer Chromatography', *TAPPI*, **55**(11), 1619 (1972).
158. B.B. Sitholé, T.N. Tran, L.H. Allen, 'Quantitative Determination of Aluminum Soaps in Pitch Deposits', *Nordic Pulp Pap. Res. J.*, **11**(2), 64–69 (1996).
159. M. Douek, X.-Y. Guo, 'An Overview of the Chemical Nature of Deposits/Stickies in Mills Using Recycled

Fiber', in *Proceedings of TAPPI 1997 Recycling Symposium*, TAPPI Press, Atlanta, GA, 313–330, 1997.
160. B. Holmbom, 'Analysis of Dissolved and Colloidal Substances Generated in Deinking', in *Proceedings of 1997 Pira International Wet End Chemistry Conference and COST Workshop*, Pira International, Leatherhead, UK, 1–12, 1997.
161. G.M. Dorris, M. Douek, L.H. Allen, 'Analysis of Metal Soaps in Kraft Mill Brownstock Pitch Deposits', *Pulp Pap. Can.*, **84**(3), T15–T20 (1983).
162. G.M. Dorris, M. Douek, L.H. Allen, 'Analysis of Amide Defoamers in Kraft Mill Pitch Deposits', *J. Pulp Pap. Sci.*, **11**(5), J149–J154 (1985).
163. I.D. Suckling, R.M. Ede, 'A Quantitative ^{13}C Nuclear Magnetic Resonance Method for the Analysis of Wood Extractives and Pitch Samples', *Appita*, **43**(1), 77–80 (1990).
164. C.J. Biermann, 'The Use of Proton NMR for Pitch Analysis – A Case Study', *TAPPI J.*, **71**(8), 135–136 (1988).
165. C.J. Biermann, M.-K. Lee, 'Analytical Techniques for Analyzing White Pitch Deposits', *TAPPI J.*, **73**(1), 127–131 (1990).
166. J.M. McGuire, C.C. Lynch, 'Characterization of Non-microbiological Paper Mill Deposits by Simultaneous TG/IR/MS', *Anal. Chem.*, **68**(15), 2459–2463 (1996).
167. N. Dunlop-Jones, L.H. Allen, 'A Rapid Method for the Qualitative Analysis of Plastic and ''Sticky'' Contaminants by Pyrolysis/Gas Chromatography', in *Recycling Paper: From Fiber to Finished Product*, ed. M.J. Coleman, TAPPI Press, Atlanta, GA, 463–467, Vol. 2, 1990.
168. B.B. Sitholé, L.H. Allen, 'Quantitative Analysis of Synthetic Polymers in Pulp and Paper Mill Pitch Deposits by Pyrolysis Gas Chromatography', *J. Pulp Pap. Sci.*, **20**(6), J168–J172 (1994).
169. H.L. Hardell, 'Characterization of Impurities in Pulp and Paper Products Using Pyrolysis/Gas Chromatography/Mass Spectrometry, Including Direct Methylation', *J. Anal. Appl. Pyrol.*, **27**, 73–85 (1993).
170. J.C. del Rio, A. Gutiérrez, F.J. González-Vila, F. Martin, 'Application of Pyrolysis/Gas Chromatography/Mass Spectrometry to the Analysis of Pitch Deposits and Synthetic Polymers in Pulp and Paper Mills', *J. Anal. Appl. Pyrol.*, **49**, 165–177 (1999).
171. D.F. Leclerc, R.W. Gooding, J.T. Wearing, 'Off-line Characterization of Polymeric Contaminants in Pulp by Diffuse Reflectance Infrared Spectrometry', *J. Pulp Pap. Sci.*, **18**(6), 202–206 (1992).
172. D.F. Leclerc, M.D. Ouchi, 'Rapid Characterization of Dirt Specks in Kraft Pulp by Infrared Spectrometry', *J. Pulp Pap. Sci.*, **22**(3), J112–J117 (1996).

Pyrolysis in the Pulp and Paper Industry

B. Bruce Sitholé
Pulp and Paper Research Institute of Canada, Pointe Claire, Canada

1 Introduction 1
2 Instrumentation 2
2.1 Isothermal Furnace Pyrolyzers 3
2.2 Curie Point Pyrolyzers 4
2.3 Resistively Heated Pyrolyzers 5
2.4 Thermal Extraction Pyrolyzers 6
2.5 Off-line Pyrolysis/Gas Chromatography 7
2.6 Pyrolysis/Mass Spectrometry 7
2.7 Pyrolysis/Fourier Transform Infrared Spectroscopy 7
2.8 Automated Analytical Pyrolysis 8
3 Data Analysis 8
3.1 Fingerprinting of Samples 9
3.2 Pyrolysis/Gas Chromatography/Mass Spectrometry Library 10
4 Pyrolysis Alkylation 12
5 Applications of Analytical Pyrolysis 16
5.1 Analysis of Natural Polymers in Wood and Pulps 16
5.2 Analysis of Deposits and Impurities by Normal Pyrolysis 22
5.3 Analysis of Deposits by Pyrolysis Alkylation 23
5.4 Analysis of Deposition Problems in Coated Papers 24
5.5 Analysis of Rosin Sizing Agents 24
5.6 Analysis of Alkyl Ketene Dimer and Alkenyl Succinic Anhydride Sizing Agents 24
5.7 Analysis of Wet-strength Resin 26
5.8 Analysis of Sulfur Compounds 27
5.9 Analysis of Ink 28
5.10 Analysis of Photocopier Toners 28
6 Pretreatment of Samples Before Analytical Pyrolysis 29
7 Slow Pyrolysis of Wood Samples 29
8 Simultaneous Pyrolysis Alkylation of Synthetic Polymers 31
8.1 Alkyd Resins 31
8.2 Unsaturated Polyester Resins 31
8.3 Polyester Fibers 32
8.4 Phenolic Polymers 32
8.5 Poly(vinyl acetate) 32
8.6 Methacrylic Acid Copolymers 33
8.7 Cellulose Acetate Butyrate 33
8.8 Polycyanoacrylates 33
Abbreviations and Acronyms 34
Related Articles 34
Appendix 1: Pyrolysis Nomenclature 34
References 36

Analytical pyrolysis is a powerful technique that is seldom used in the pulp and paper industry. This article illustrates the utility of the technique to study the chemistry of wood and pulps, mechanistic aspects of pulping, troubleshooting process upsets in pulp and papermaking, and evaluating the quality of finished products. The technique has advantages over other analytical techniques in that little sample preparation is required and enough data can be obtained from very small amounts such as dirt spots on paper. In many cases good data can be obtained by analyzing the pyrolysis products by gas chromatography (GC) with flame ionization detection (FID). Detection with mass spectrometry (MS) offers the advantage of absolute qualitative and quantitative determination of the chemistry of the analytes. In many instances the chromatography can be improved by on-line derivatization of the pyrolysis products with an alkylating agent.

This article takes the reader through a survey of the different types of analytical pyrolysis systems, through data analysis and processing, and examples of applications of analytical pyrolysis in the pulp and paper industry. Ample illustrations are used to highlight the usefulness of the technique.

1 INTRODUCTION

Pyrolysis techniques involve the application of thermal energy to induce the transformation or degradation of compounds. The techniques can be divided into two types, applied pyrolysis and analytical pyrolysis. Applied pyrolysis is concerned with the production of chemicals, e.g. the pyrolysis of coal to generate oil. Analytical pyrolysis deals with the structural identification and quantitation of pyrolysis products with the ultimate aim of establishing the identity of the original material and the mechanisms of its thermal decomposition. This article is devoted to analytical pyrolysis.

Many organic compounds can be analyzed by GC, provided that they are volatile or can be derivatized into compounds that are amenable to analysis by GC. Compounds that cannot be analyzed by GC can instead

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

be analyzed by other chromatographic techniques such as liquid chromatography and capillary electrophoresis. The analysis of many polymers by any of these analytical techniques tends to be difficult mainly because they require that the analytes be brought into solution, and this is usually difficult to achieve. Such compounds are candidates for analytical pyrolysis, which entails applying thermal energy to break them into products that are volatile enough to be analyzed by GC. If the pyrolysis is done under controlled conditions and inert atmospheres, the resultant pyrolysis fragments give characteristic patterns that are "fingerprints" of the original polymers. Further information on the chemistry of the fragments can be obtained by detecting them by MS and, in some instances, by Fourier transform infrared (FTIR) spectroscopy. The pyrolysis can be done by rapid heating (flash pyrolysis) or temperature-programmed heating (slow pyrolysis) of the polymers. Slow pyrolysis products can be sent to a mass spectrometer or the pyrolysis can be done directly in a mass spectrometer.

Although there are a number of books and reviews on analytical pyrolysis techniques, e.g. an introduction to pyrolysis,[1] a bibliography,[2] an applications monograph,[3] a handbook,[4] and an extensive bibliography of 594 references on synthetic polymers,[5] there do not seem to be any that are specifically targeted for pulp and paper matrices (e.g. none of the 11 chapters in the handbook edited by Wampler[4] specifically address pulp and paper analysis). This article addresses this shortcoming by reviewing how pyrolysis techniques are aptly suited to the analysis of samples from the pulp and paper industry. The pyrolysis techniques are useful in understanding mechanistic aspects of natural polymers present in wood, and how they affect pulp and papermaking processes. They can also be used to optimize pulp and papermaking processes and to troubleshoot production and product quality problems. This review will include applications of pyrolysis/mass spectrometry (Py/MS), pyrolysis/Fourier transform infrared (Py/FTIR) spectroscopy, pyrolysis/gas chromatography (Py/GC), and pyrolysis/gas chromatography/mass spectrometry (Py/GC/MS), but the main emphasis will be on the last two techniques. Theoretical aspects of pyrolysis are beyond the scope of the article but can be found in the references listed above.

Since the pyrolysis unit, the pyrolyzer, is the heart of analytical pyrolysis, it is important to understand how pyrolysis is achieved. The next section reviews the different pyrolysis modes that are currently available.

2 INSTRUMENTATION

Analytical pyrolysis can be looked upon as occurring in two steps, namely fragmentation of the sample to be analyzed and detection of the pyrolysis products by a suitable detector. In general, lower pyrolysis temperatures cause less fragmentation, resulting in smaller amounts of pyrolysis products. Very low pyrolysis temperatures may not cause fragmentation of samples but may remove volatile matter that may interfere in the interpretation of the pyrograms. For example, poly(vinyl chloride) contains residual softening agents and antioxidants that can be removed by thermal desorption. Higher temperatures can result in extensive fragmentation of samples resulting in larger amounts of pyrolysis products and more complex pyrograms that may be difficult to interpret by the detector used. This can be solved by introducing an extra step in the analytical pyrolysis procedure, namely separation of the pyrolysis products by GC. After pyrolysis, the fragments are swept on to a chromatographic column. Typically, an inert gas, helium, is used to flush the pyrolyzer at high flow rates ($50\,mL\,min^{-1}$ and above), thereby avoiding recombination of the fragments. The chromatographic process separates the fragments according to their volatility. Usually, the column is maintained at a low temperature at the beginning to allow condensation of the pyrolysates into a narrow plug at the head of the column. The highest sensitivity is achieved by direct flow without splitting, but normally an appropriate split ratio is used to avoid overloading the column and to induce a reasonable flow rate during the chromatographic process. The separated pyrolysis products can then be detected by a suitable detector such as a FTIR spectrometer, flame ionization detector, or mass spectrometer. The last two are the most common types of detection and will be the focus of this article. Analytical pyrolysis requires that the sample be heated to pyrolysis temperatures in a reproducible manner. The reproducibility of fragmentation patterns is determined by parameters such as mode of heating, sample size, sample preparation, and flow rate. For example, large samples may experience temperature gradients during pyrolysis and this will result in the generation of irreproducible pyrograms. Also, pyrolysis of large samples may result in contamination of analytical columns, especially the inlet side. Back rinsing of the column with solvents or cutting off a small portion of the column ($\sim 30\,cm$) can restore the performance of capillary columns. Other techniques that have been employed to avoid contamination of columns include installation of disposable precolumns (similar to what is done in high-performance liquid chromatography),[6] design of a splitter system to reduce column contamination without affecting the performance of the chromatographic column,[7] and modification of the Py/GC inlet system for capillary columns to allow proper identification of the pyrolysis products, including those with lower volatilities that remain in the pyrolysis chamber.[8] Analysis of standards at regular intervals is a

good way of ensuring that a Py/GC system is generating reproducible pyrograms.

Analytical pyrolysis has traditionally been carried out using laboratory-made instruments and this has made it difficult to develop standardized instrumentation. Also, the technique has been practised in so many different ways that it has resulted in a bewildering array of terms and terminologies. An attempt was therefore made to develop a common nomenclature.[9] The nomenclature is listed in Appendix 1. Later, the International Union of Pure and Applied Chemistry (IUPAC) published recommendations entitled "Nomenclature and Terminology for Analytical Pyrolysis".[10]

A good pyrolysis unit should meet certain specifications, such as a very low dead volume and a short and heat-insulated connection between the pyrolyzer and the detector or GC column to prevent condensation of high-boiling pyrolysis products, rapid heating of the sample, and accommodation of different sample forms. The variety of instrumentations used can be categorized into three kinds, namely isothermal furnace pyrolyzers, Curie point (inductively heated filament) pyrolyzers, and resistively heated pyrolyzers. Features and operations of each type of pyrolyzer will be described below. The different modes of sample heating in the three modes of pyrolysis designs imply that the resultant pyrograms may be different from one instrument to another. Such discrepancies therefore require that care should be taken when comparing pyrograms from different pyrolysis systems.

2.1 Isothermal Furnace Pyrolyzers

In this class of pyrolyzer, the furnace is held isothermally at a specific temperature, typically the pyrolysis temperature, and the sample is introduced into the hot furnace. A carrier gas at a high flow rate is used to sweep the pyrolysis products onto a chromatographic column for analysis. Figure 1 is a schematic of the apparatus.

In most cases, a liquid sample in a syringe is injected into the hot zone where it vaporizes before undergoing pyrolysis. The inlet carrier gas rapidly sweeps the pyrolysis products on to the chromatographic column. Typically, high inlet gas flows, e.g. $50\,mL\,min^{-1}$, are used. One therefore has to be careful that the sample in the syringe is not swept out of the needle housing during the injection process. Solid samples can be introduced by prior dissolution of the samples in suitable solvents or the use of syringes designed for injection of solid samples. There are two designs of such syringes. In the first design, one needle resides inside another needle. After the solid sample has been placed on a groove or indentation of the inner needle, the syringe is then injected into the pyrolyzer and the plunger depressed to expose the contents of the inner needle to the hot

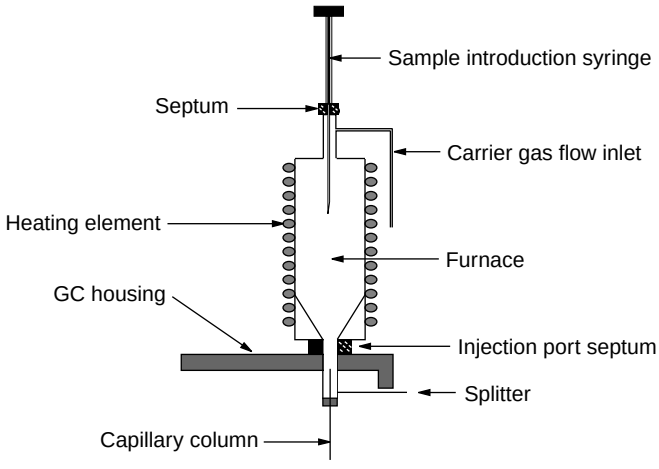

Figure 1 Schematic of a furnace pyrolyzer.

furnace. In the second design, the injector is supplied with two interchangeable plungers: a spiral tip plunger for injecting solids and a flat-tipped plunger for solid samples in the form of granules, fibers, or powders that can be packed into the bore of the needle with the plunger retracted. A pelletizer can be used for producing micropellets that can be transferred into the bore of the solids injector. One manufacturer (Shimadzu) has a unique sample dropping method that is said to shorten the time of heating the sample to the decomposition temperature and to minimize the loss of the low-boiling components. Consequently, all of the pyrolysis products load on to the GC column at a faster rate.

A heating coil wrapped around the furnace is used to heat the furnace to a specified temperature and a temperature sensor sends a signal to a controller that adjusts the conditions to maintain a constant temperature. The wall thickness and insulation of the furnace affect the time required to reach a set temperature equilibrium. The pyrolysis temperature experienced by the sample may not be the same as that of the furnace because of temperature gradients induced by the wall thickness of the sample holder, the flow rate and the size of the sample. The simple nature of isothermal pyrolyzers implies that they are easy to construct in-house as they do not require sophisticated controls. The pyrolyzers are ideal for situations that do not require frequent changes of pyrolysis temperature or temperature gradients.

Furnace pyrolyzers tend to be large and cumbersome and require inordinately long times to reach equilibrium temperatures. Their pyrolysis tubes are also large and thus it is not easy to ascertain the actual pyrolysis temperatures of the analyzed samples. The pyrolyzers require large amounts of samples and are inconvenient for cases where the sample size is limited, e.g. dirt spots on paper. A splitter inside the GC housing reduces the flow rate to the

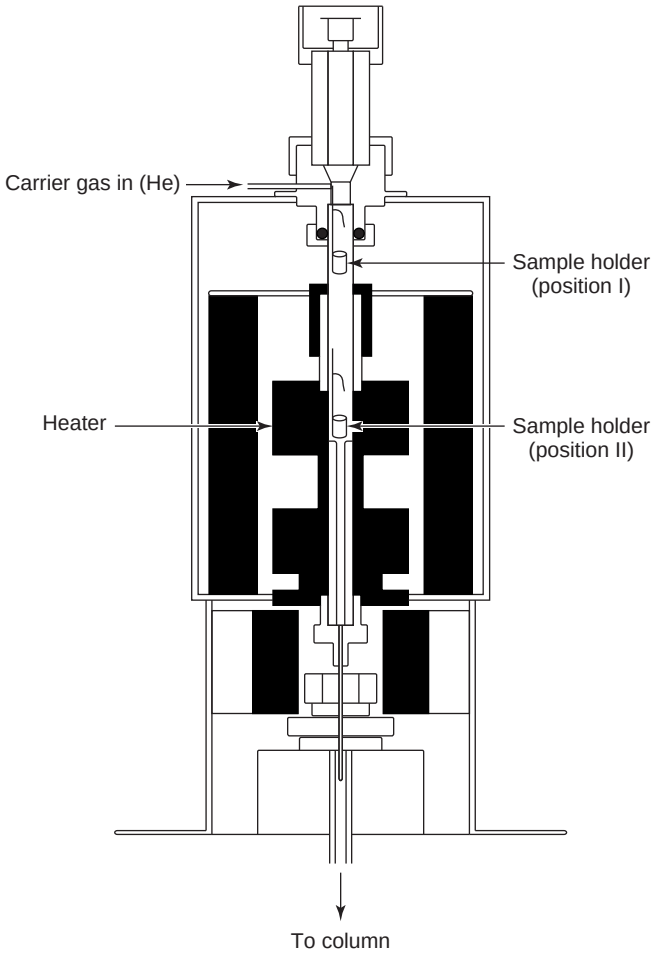

Figure 2 Schematic of a modified microfurnace pyrolyzer.

column to normal analytical flow rates, and is maintained at the highest temperature that will be experienced by the chromatographic column.

A more recent design of the pyrolyzer is a microfurnace unit that can analyze small samples of about 100 µg. In the unit, shown schematically in Figure 2, the sample for analysis is placed in a platinum cup that is held above the pyrolyzer at room temperature (position I). Lowering of the sample to position II allows for pyrolysis of the sample. This arrangement is claimed to be suitable for labile compounds because the sample is not exposed to high temperatures before the pyrolysis step.[11]

2.2 Curie Point Pyrolyzers

Figure 3 shows a schematic of a Curie point pyrolyzer. Here the active heating element is a ferromagnetic metal wire that is placed in a radiofrequency field to induce eddy currents that cause a rise in temperature. The temperature rises until it reaches the Curie point, where the ferromagnetic wire becomes paramagnetic and ceases

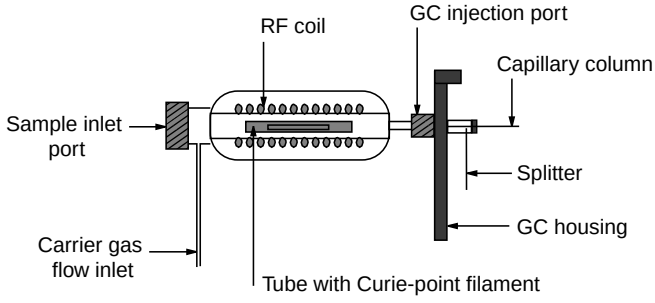

Figure 3 Schematic of a Curie point pyrolyzer.

induction of eddy currents. The Curie point temperature reached depends on the type of metal used: the values for cobalt, nickel, and iron are 1128, 359, and 770 °C, respectively. These values are specific for the metals in question and can only be altered by making alloys of the metals. For example, a nickel–iron alloy wire common in many Curie point systems reaches a maximum at 510 °C. A series of ferromagnetic alloys are available with temperature ranges from 160 to 1040 °C. Ballistic heating enables the pyrolysis temperatures to be reached within 20–30 ms. The sample, typically 10–20 µg, is coated on the wire, and is then dried while rotating the wire. Application of the sample as a melt which then solidifies on the wire can be used for coating the wire. Reproducible coating of samples on the wire is essential to ensure the production of pyrograms that are fingerprints of the samples to be analyzed. Samples that are difficult to coat on the wire can be introduced by bending, flattening, or crimping the wire, or by making indentations on the wire that will retain the samples.

Deterioration of the sample wires with usage will eventually affect the reproducibility of the pyrograms generated by the system. Cleaning of all components that are in contact with the sample and pyrolysis products is essential for long-term reproducibility. The cleaning can be done by scrubbing or sanding or by cleaning with acid and/or solvent. Heating at elevated temperatures such as in a flame for a few seconds is ideal for removing deposits that are difficult to remove by other cleaning methods. However, pyrolyzer components that are susceptible to oxidation should not be heated in a flame.

Curie point pyrolyzers can be automated with multichannel autosamplers that allow for automatic and unattended analysis of up to 24 samples. For example, DyChrom (Santa Clara, CA, USA) markets a unit that can be used with 20 different ferromagnetic alloy wires to cover temperature range from 160 to 1040 °C.

Curie point pyrolyzers are considered inferior to microfurnace pyrolyzers because they cannot maintain accurate and high oven temperatures at their outlets ends attached to gas chromatographs. Consequently,

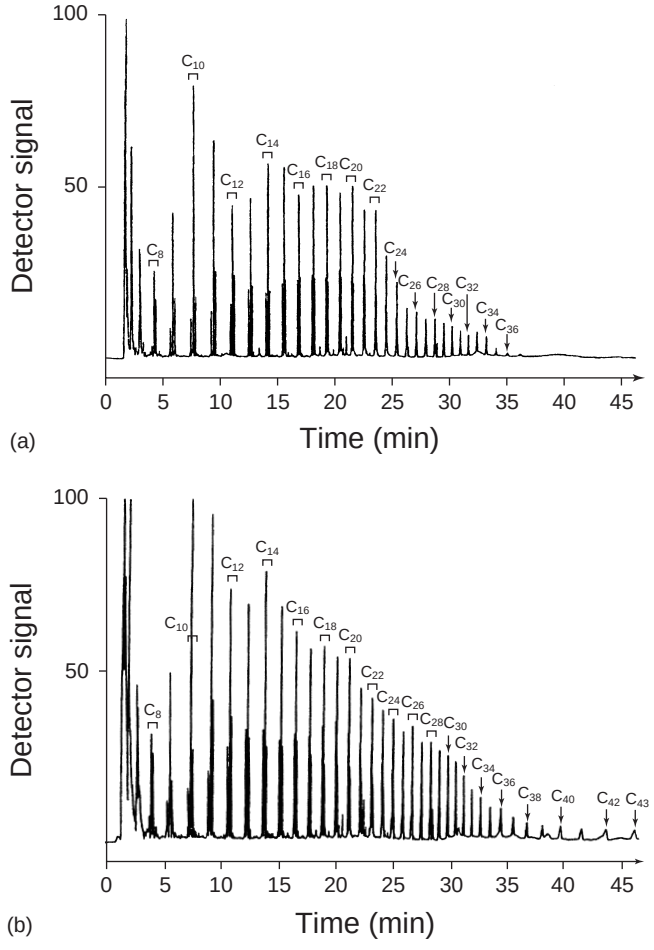

Figure 4 Comparison of pyrograms generated with (a) normal and (b) modified Curie point pyrolyzers.

high-molecular-weight fragments such as styrene trimers are lost and not readily detectable by pyrolysis using a Curie point analyzer.[1] These shortcomings were apparently overcome by the development of a Curie point pyrolyzer with a modified maintenance oven that allows adequate transfer of high-molecular-weight fragments into the GC analytical column.[12] Figure 4(a) and (b) illustrate the advantage of such a modified pyrolyzer; it is evident that the modified pyrolyzer facilitates the detection of pyrolysis fragments of polypropylene with chain lengths exceeding C_{38}. Chromatography of styrene trimers, from pyrolysis of polystyrene, was significantly improved using this analyzer.[12]

2.3 Resistively Heated Pyrolyzers

The principle of resistively heated pyrolyzers (or filament probes) entails rapid heating of a sample from ambient to pyrolysis temperature in a matter of milliseconds and this is achieved by applying an electrical current to a filament that holds either the sample or a tube that contains

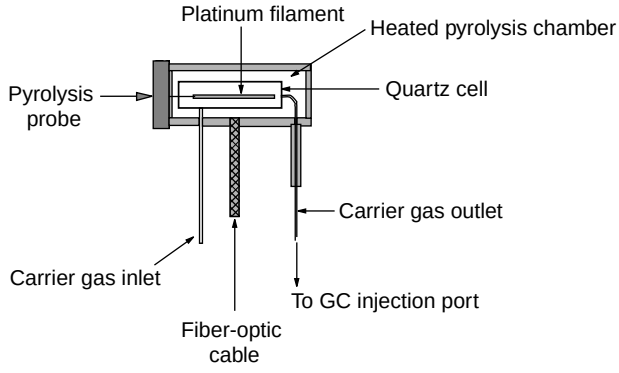

Figure 5 Schematic of a resistively heated pyrolyzer.

the sample. An early design of such a system featured a ribbon or coil that was an element in a Wheatstone bridge. The bridge was set such that when balanced at the pyrolysis set-point temperature a capacitor was discharged to rapidly heat the filament to a temperature of 1000 °C at rates of 20 °C ms^{-1}. Modern designs can now heat at rates that vary from 0.01 °C min^{-1} to 30 °C ms^{-1}. The filaments used exhibit high electrical resistance and wide operating temperatures and they can be made of iron, platinum, and Nichrome.[13] They can be configured as either a ribbon for direct sample placement or as a coil for holding samples in reusable quartz tubes or boats.

The absence of a barrier between the sample and a ribbon filament ensures good contact between the sample and filament, thus enabling rapid heating of the sample. In one design, shown schematically in Figure 5, two current pulses of variable time and amplitude are applied to the filament during pyrolysis: the first heats the filament to the set pyrolysis temperature and the second compensates for cooling effects to enable the sample to experience the same temperature during the entire pyrolysis time.[14] The cooling effects are induced by a fast gas flow that sweeps the pyrolysis products on to the chromatographic column. Thus the current pulses control the length and shape of the computer-generated temperature and time profiles, thereby guaranteeing accuracy and reproducibility. This is considered critical as temperature affects the generation of pyrolysis degradation products. The rapid heating minimizes decomposition of the sample during the temperature rise time (TRT). The pyrolysis time can be changed according to the size of the sample to prevent overheating of the filament after sample pyrolysis: such overheating may lead to unwanted secondary decomposition of the pyrolysis products. The pyrolysis temperature can be monitored in two ways: electrical resistance of the filament and filament light emission of the filament/sample contact area. Both can be displayed and compared on a screen or downloaded to a printer. The comparison is done to confirm the reproducibility

of the pyrolysis profile. Such a pyrolyzer is believed to be well suited for quantitative analysis because of the rapid heating under very reproducible conditions.[14,15] However, the filaments used are very thin and fragile and they break if current is applied to them in the absence of a gas flow. Residual pyrolysis products on the filaments need to be removed by heating over a flame. Each time they are replaced, the system has to be calibrated to ensure that the correct pyrolysis profiles will be attained. In some designs, the filaments come in precalibrated cases, and an optical output will read the correct temperatures.

Coiled filaments are more rugged than ribbon filaments. Since they are not in direct contact with analytes, they are consequently not contaminated by pyrolysis products or nonpyrolyzed matter and therefore do not require cleaning. These filaments can last for very long periods before replacement.

The filament probes are housed in heated chambers to minimize condensation of the pyrolysis products. The high pyrolysis temperatures in resistively heated pyrolyzers can result in deposition of pyrolysis products on the walls of the chambers that are at lower temperatures. To prevent such contamination, the chambers can be lined with quartz or the filament can be housed in reusable quartz tubes. The use of quartz also minimizes catalytic and secondary reactions that occur on hot metal surfaces.

The filament probes were originally interfaced to GC injection ports by attaching a heated interface to the septum retainer of the gas chromatograph. Carrier gas was then passed through the heated chamber into the injection port via a needle assembly. When the probe was removed and reloaded for each run, air was introduced into the column with undesirable consequences especially for MS detectors. This problem has now been overcome by valved interfaces that isolate the GC injection port from the pyrolysis probe so that the probe can be removed or inserted without interrupting the carrier flow in the gas chromatograph.

Filament probes are ideally suited for sequential pyrolysis where a sample is heated at one temperature to pyrolyze interfering compounds and at a higher temperature to pyrolyze the analytes of interest. Examples of applications of this technique will be discussed later in section 5. An example of the reproducibility of a resistively heated pyrolyzer is shown in Table 1 for the pyrolysis of polystyrene. Polystyrene is often used to check the reproducibility of a pyrolyzer because its pyrolysis products vary from low- to high-boiling components and the main products are styrene monomer, dimer, and trimer, as illustrated in Figure 6. The data in Table 1 show that the relative standard deviation is within 2%.[16]

Table 1 Reproducibility of a resistively heated pyrolyzer (pyrolysis of polystyrene at 700 °C for 2 s)

	Area of styrene	Area of dimer	Area of trimer	Total area
First replicate	274 330	6622	13 639	294 591
Second replicate	274 743	6642	13 931	295 316
Third replicate	267 800	6490	14 008	288 298
Mean	272 291	6585	13 859	296 475
RSD (%)	1.4	1.3	1.4	2.2

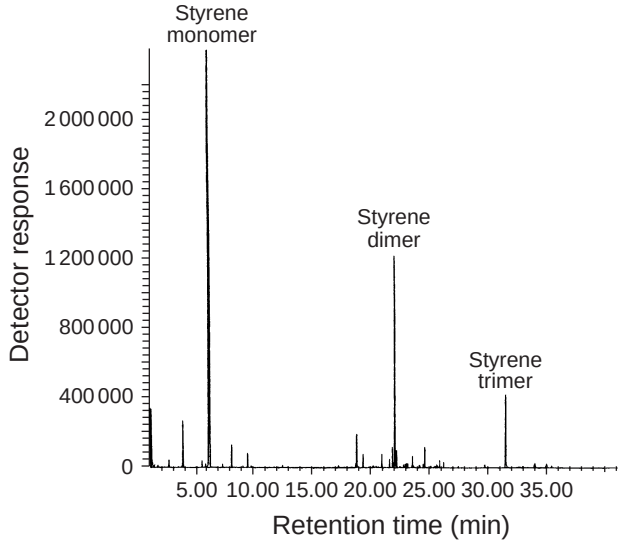

Figure 6 Polystyrene is often used as a pyrolysis standard; its pyrolysis products range from low- to high-boiling fragments.

2.4 Thermal Extraction Pyrolyzers

Humble Instruments (Humble, TX, USA), markets a flow-programmable thermal extraction/pyrolysis injector that is said to offer true, in-sample temperature measurement for the accurate study of the decomposition of complex samples. The unit can be used for vaporization (thermal extraction) and nonisothermal (constant heating rate) pyrolysis of solid or liquid samples. Thermal extraction GC is a solvent-free technique for fingerprinting the volatile components present in solid or liquid samples. Volatile compounds are thermally volatilized from the sample matrix into a GC oven where they are cryogenically focused on-column resulting in well-resolved peaks (this is equivalent to syringe injection of the volatile sample). This technique should be applicable to the analysis of volatile and odorous materials in pulp and paper matrices. Thermal extraction at high temperatures, usually from 340 to 650 °C, can be used for the analysis of high-molecular-weight compounds such as polymers. The pyrolyzer can also be used to automatically acquire multiple heating rate cycles with true temperature acquisition for kinetic determinations. Pyrolysis profiles consisting of time, true temperature acquisition, and rate (FID

response) are processed using one of the kinetic models in Lawrence National Laboratory's Kinetics® program. However, the pyrograms generated may be different from those obtained with other pyrolyzers because of the non-isothermal pyrolysis conditions.

The unit can be equipped with a microscale sealed vessel for the generation of closed-system pyrolysis products. The vessel is a 40 µL glass tube that may be sealed after sample and internal standard loading. The tubes are heated under tightly controlled conditions. Products generated by this pyrolysis technique do not contain the intermediate reactants formed during open-system, nonisothermal, or flash pyrolysis.

2.5 Off-line Pyrolysis/Gas Chromatography

In off-line Py/GC, the pyrolysis is done in the absence of a chromatograph and the pyrolysis products are trapped for subsequent analysis by GC or gas chromatography/mass spectrometry (GC/MS). A typical experimental set-up is shown in Figure 7. The apparatus is flushed with an inert gas (helium) before pyrolysis of the sample that is placed between silanized quartz wool packing in a quartz tube. The quartz wool allows the pyrolysis products to pass through but traps nonpyrolyzed compounds that may be swept by the gas flow. The pyrolysis products condense in the quartz tube that is cooled with liquid nitrogen. After pyrolysis, the pyrolysis products condensed in the quartz tube are rinsed with a suitable solvent and analyzed quickly to minimize time-dependent secondary reactions.

2.6 Pyrolysis/Mass Spectrometry

In Py/MS, a mass spectrometer detects the pyrolyzed compounds. The pyrolysis probe is inserted directly in the mass spectrometer. This offers advantages such as collection of repetitive scans during the pyrolysis, detection of different decomposition steps of the analyte, detection of thermally labile compounds because of the vacuum conditions in the mass spectrometer, and detection of high-molecular-weight compounds. High-molecular-weight compounds can be detected because there is no interface between the pyrolyzer and the mass spectrometer (in Py/GC/MS high-molecular-weight materials are lost by adsorption on the walls of the interface surfaces). Disadvantages of Py/MS are that the technique detects mixtures of pyrolysis products and further analysis of the mass spectra is required to identify each pyrolysis product. Also, quantitative analysis of the pyrolyzed components is difficult.

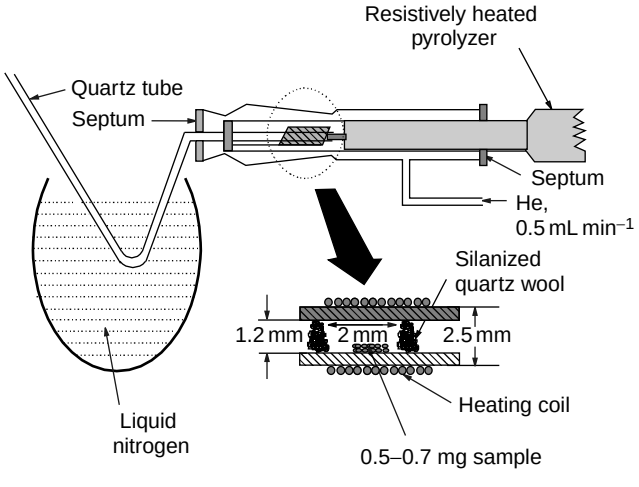

Figure 7 Schematic of an off-line pyrolysis system.

2.7 Pyrolysis/Fourier Transform Infrared Spectroscopy

In Py/FTIR spectroscopy, the pyrolysis products are directed to an infrared (IR) cell. In early designs, pyrolysis was done in a test tube and the pyrolysis products were condensed or deposited on an IR-transparent plate or crystal.[17] This, however, resulted in recombination of the pyrolysis products and consequently did not give an accurate view of the pyrolysis process. This limitation was partially solved by use of a heated transfer line to transport the volatiles into the IR beam.[18] The procedure resulted in dilution of the pyrolyzed products that were therefore difficult to detect. Direct-pyrolysis FTIR spectroscopy overcomes these problems because the sample is located <3 mm below the IR beam via a specially designed interface.[19] The interface, a cylindrical cell with windows on each end, houses the pyrolysis probe. Upon pyrolysis, the volatiles diffuse immediately into the IR beam where they are detected. The sensitivity is very good with this technique because of minimal sample dilution. Samples may also be heated at slow rates (e.g. 0.01 °C min^{-1}) to carry out time-evolved studies and simulated distillations.

Several variables can affect the accuracy of the results obtained by direct Py/FTIR spectroscopy. They include pyrolysis temperature, cell temperature, and the flow rate of the gas through the pyrolyzer. Cell temperature is important to prevent condensation of the pyrolysis products on the cell windows. The gas flow rate can be optimized to provide adequate ventilation without losing sensitivity. The optimized parameters for one commercial Py/FTIR instrument are a cell temperature of 200 °C and a cell flow rate of 40 mL min^{-1}.[20]

Py/FTIR spectroscopy is ideal for clean samples or mixtures whose components pyrolyze at distinct temperatures. The technique is more rapid than Py/GC because the chromatographic step is eliminated. Thus in cases where one is concerned with quality control of materials, such as paper coating formulations, Py/FTIR would be the method of choice over Py/GC.

2.8 Automated Analytical Pyrolysis

In the majority of cases, pyrolyzers are marketed by companies that do not manufacture gas chromatographs. This sometimes poses a problem in ensuring that a pyrolyzer will be able to communicate properly with a GC column to which it is attached. For example, triggering a GC run should simultaneously trigger the pyrolyzer. This can be accomplished by modification of the electronics of the GC system, provided that this will not negate the warranty of either instrument. Another potential problem may be conflict in the software of the two systems: GC vendors normally warranty only their software running on the GC systems that they sell. Pyrolysis products fouling mass spectrometers are considered the fault of the operator even though the pyrolysis system in use may have been faulty. Understandably, the vendor of the pyrolysis unit will not take responsibility for the mass spectrometer. These problems could be avoided if companies made complete dedicated pyrolysis GC/MS systems. Dutch researchers developed the first fully integrated pyrolysis mass spectrometer in the 1970s by integrating a Riber quadrupole mass analyzer with a Curie point pyrolyzer.[21] Further development led to the production of a fully automated system that utilized high-speed ion counting and computerized data-processing techniques.[22] This advance resulted in the development of two commercial instruments, the Extranuclear 5000 from Extranuclear Laboratories (USA) and the Pyromass 8-80 from VG Gas Analysis (UK). The two instruments were based on conventional MS and, consequently, were very expensive for most laboratories. A low-cost dedicated instrument was developed in the mid 1980s.[23] The instrument came with a microcomputer that was capable of performing on-line pattern recognition analysis. Currently, there is a dedicated Py/GC/MS system that has found wide use in the microbiological field. The instrument, marketed by Horizon Instruments (East Sussex, UK), is an automated and dedicated Py/GC/MS system that can process 150 samples unattended. It also uses a Curie point pyrolyzer and a quadrupole mass spectrometer with a mass range of 12–400 Da.[24] Another dedicated system is available from Shimadzu Scientific Instruments (Maryland, USA). It is a furnace pyrolyzer attached to a gas chromatograph with a quadrupole mass spectrometer. The system, however, does not have autosampling capabilities.

Automation in resistively heated pyrolyzers is difficult, if not impossible, in systems that use ribbon filaments. This is because the filament has to be cleaned after each pyrolysis run to remove pyrolysis products of the previous run. In cases where the filament is a coil, automation is possible because a fresh tube is brought into the coil before each pyrolysis run. For example, CDS Analytical (Pennsylvania, USA) has an automated pyrolysis unit that can run 36 samples. The unit is equipped with software that enables control of the pyrolysis probe and programming of the autosampler. The samples, placed in 2×40 mm quartz tubes, can be loaded on a carousel that feeds the samples on to the pyrolyzer by gravity. The pyrolyzer, equipped with a platinum filament coil, has five sensors for monitoring the home and carousel number locations, the tray used, presence of sample tube, and discharge of the tube from the pyrolyzer. Cleaning of the pyrolyzer can be done by valving to the gas chromatograph for blank runs or out to vent to remove contaminants.

3 DATA ANALYSIS

Pyrolysis under controlled conditions results in the production of pyrolysis products that are often used to fingerprint the nature of the original sample. Thus an unknown pyrogram can be identified by comparison with standards run under the same conditions. In cases where only a few pyrolysis products are generated the comparison can be made visually, e.g. by overlaying of pyrograms. However, visual comparison tends to be complicated in situations where pyrolysis results in large numbers of pyrolysis products. In such cases, multivariate data analysis may yield more information than visual comparison. Multivariate data analysis is considered useful in evaluating data that contains correlated variables as in the pyrolysis of pulp and paper matrices where such pyrolysis results in degradation products that can be correlated with constituents in the pyrolyzed matrices. Theoretical treatises on multivariate data analysis have been reported.[25] The analysis reduces a large amount of complex data down to a form where it can be readily understood, e.g. graphical output in two or three dimensions. Two statistical methods that can manipulate data to show correlations are principal component analysis (PCA) and partial least squares (PLS).

PCA is used to identify correlations within large data sets and to reduce the number of dimensions required to display the variation within the data. A new set of axes, principal components, is constructed, each of which accounts for the maximum variation not accounted for by previous principal components. With pyrolysis mass spectra, PCA is used essentially as a data reduction technique prior to performing PLS analysis. The information obtained from principal component plots can be used to identify samples that deviate from the norm and/or outliers within the data and as a test for reproducibility.

The PLS method finds correlations between the principal components and generates new vectors based on two principles: a good approximation of the separation

between the variables and the highest correlation possible between the variables.

Both methods generate their models with an internal validation of significance called cross-validation: parts of the data are kept out of the model development, and then predicted by the created model and compared with the actual values.

Discrimination of similar pyrograms can be done by comparison of individual peaks in the pyrograms. For example, Meier et al.,[26] used a software package, called MatchFinder®, that is designed for analysis of complex pyrograms. The program analyses retention times and responses (peak areas or heights) of each peak in the pyrogram. Initially, trial assignments are made to determine if there are any peaks in the pyrograms that correspond to each other. An annealing algorithm is then used to avoid suboptimal solutions. The matching process is rapid, typically 3–4 min. An overall measure of the similarity between pyrograms is indicated by the ''match quality''. Analysis of the Py/GC data by computer base discrimination methods showed that the origins of teak, mahogany, and eucalyptus could be discerned but it was not possible to determine the origin of pine wood samples.

3.1 Fingerprinting of Samples

When heated to pyrolysis temperatures, polymers dissociate at certain bonds, resulting in degradation products, usually monomers, that may rearrange to form dimers, trimers, and other derivatives of the monomer. The amounts and ratios of these degradation products are a function of the pyrolysis temperature. For example, pyrolysis of polystyrene at three different temperatures results in the generation of mainly monomers, dimers, and trimers whose ratios vary with the pyrolysis temperature.[27] This is illustrated in Figure 8(a–c) for pyrolysis of the polymer at 600, 700 and 800 °C: the amount of trimer generated decreased with increase in temperature and very little is produced at 800 °C. The patterns of the pyrograms shown will be reproducible if the pyrolysis conditions can be controlled to strict tolerances. These resultant pyrograms can then be used to ''fingerprint'' the pyrolyzed polymer. The fingerprinting is different from that generated by IR data in that the pyrolysis conditions influence the pattern of the pyrograms. Although other parameters such as the instrument used, sample size, and geometry may also influence the pyrogram pattern, Py/GC has found widespread use in fingerprinting a wide variety of polymers.[4,28,29]

Fingerprinting is usually easy when dealing with pure polymer samples where there are no interferences from sample matrices. Complications arise when analyzing complex multicomponent samples that may be partly organic materials having different melting points.

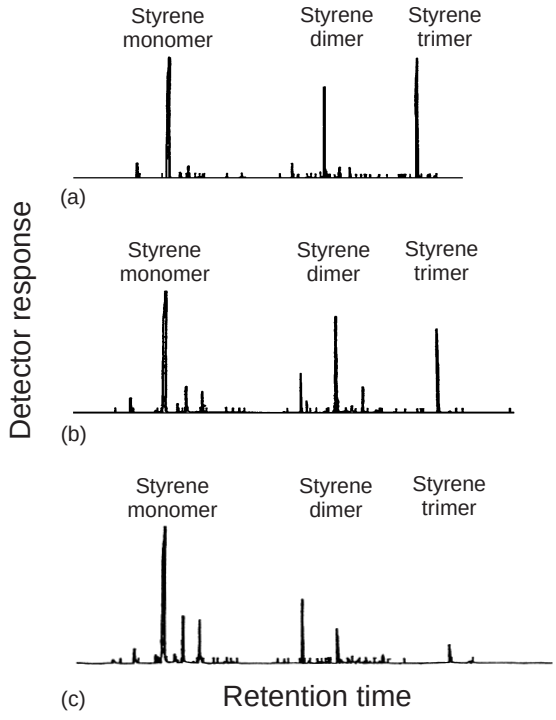

Figure 8 Effect of temperature on the pyrolysis patterns of polystyrene: (a) 600, (b) 700, and (c) 800 °C.

Pyrograms from such matrices are so complex that it is difficult to discern the nature of the components present. However, this can be overcome by multistep pyrolysis that permits selective volatilization of individual components from the sample or division of the sample into temperature-dependent fractions. For example, interferences from cellulosic products in the pyrolysis of paper samples can be overcome by a two-step procedure: at one temperature to volatilize the cellulosic products and at a higher temperature to pyrolyze the polymer(s) of interest.[30,31] Simultaneous pyrolysis derivatization techniques can also be used to minimize matrix interferences, as will be discussed later in this article.

Determining the origin of wood species is an important and desirable thing to do because different wood species have different characteristics such as fiber length, fiber wall thickness, and wood resin content. These characteristics will affect the pulp and papermaking properties of the woods. Conventional methods for differentiating wood species involve macroscopic and microscopic techniques. However, the techniques are not capable of detecting small differences between same species that may be induced by the environmental conditions where the trees grew. Analytical pyrolysis may be used to fingerprint different wood species. For example, it can be used to distinguish softwoods from hardwoods, in that the pyrograms of their wood extractives are different: softwood extractives

will show the presence of resin acids and fatty acids whereas hardwoods will contain fatty acids sterols and steryl esters but no resin acids.[32] In another example, pyrolysis of acetone extracts of *Eucalyptus* species shows the presence of fatty acids, ß-sitosterol, sitostanol, fucosterol, saturated and unsaturated fatty acids, and α-tocopherol.[33]

A comprehensive study was undertaken to differentiate wood species from different areas by using Py/GC, Py/MS, near-infrared (NIR) spectroscopy, and multivariate data analysis.[26] The wood species studied and their origins are shown in Table 2. For Py/GC work a resistively heated probe was used, whereas a Curie point pyrolyzer was used for Py/MS studies. Pyrolysis of the samples gave complex data that were evaluated using a dynamic chromatographic profile matching algorithm for Py/GC studies and chemometric methods for data reduction in the case of Py/MS studies.

The MS raw data and results for multivariate analysis, PCA, and canonical variates analysis of the MS data obtained by Py/MS showed that there was good reproducibility within a group, as demonstrated by the closeness of the points in each group. This therefore demonstrated that Py/MS can be used to differentiate wood species harvested in different areas.

Table 2 Wood samples obtained for fingerprinting studies

Botanical name	Origin	Common name
Eucalyptus saligna SM	Uruguay	Eucalipto colorado
	Australia	Sydney blue gum
	South Africa	Sydney blue gum
Eucalyptus viminalis Labill.	Yugoslavia	Eucalyptus
	South Africa	Ribbon gum, white gum
	Australia	Ribbon gum, white gum
Swietenia macrophylla King	Brazil	Mahogany
	Burma	Mahogany
	Honduras	Caboa
	Indonesia	Grootbladige mahagony
	Cuba	Caboa de Honduras
Pinus radiata D. Don	Spain	Pino insigne
	New Zealand	Radiata pine
	Chile	Pino radiata
	California, USA	Monterey pine
	Sudan	–
	Uruguay	Pino insignis
	South Africa	Radiata pine
Tectona grandis L.	India	Teak
	Cuba	Teca
	Burma	Teak
	Papua New Guinea	Teak
	Indonesia	Teak
	Honduras	Teca
	Nigeria	Tiek
	Costa Rica	Teca

The pyrograms obtained by Py/GC were processed via a software package, MatchFinder®, that compares pyrograms by evaluating the retention times and responses of each peak in the pyrograms. Figure 9(a–h) show the match qualities of the wood samples analyzed, in duplicate, by Py/GC. It is clear that the reproducibility is not very good for the duplicates, except for the pine wood samples. In addition, the discrimination of samples is not as clear as with Py/MS analysis.

3.2 Pyrolysis/Gas Chromatography/Mass Spectrometry Library

The need for a library of pyrolysis mass spectra has been recognized since the early days of analytical pyrolysis, but progress has been slow owing to the complex pyrograms generated. In 1985, Italian researchers described a library of mass spectra of polymers containing 200 entries.[34] The spectra were obtained by direct Py/MS (electron impact) that entailed heating the polymers in the direct insertion probe of a mass spectrometer, at a constant heating rate and recording the repetitive mass spectra during the temperature rise time (TRT). The library could be used both as a database for library searches and as a training tool for pattern recognition analysis. Two pattern recognition methods used on the library produced identical results. A major limitation of the library is that the spectra were generated on fragments produced by heating in a direct insertion probe. The absence of chromatographic separation of the pyrolysis fragments implies that only pure polymers could be analyzed and identified by the library. Analysis of unknown mixtures that produce complex spectra would be difficult to identify using the library.

Matheson et al.[35] are creating a computer library for the identification of polymers by Py/GC/MS, the aim being eventually to encompass most known polymers. Since Py/GC/MS total ion chromatograms are complex, it is obviously time-consuming to look at individual peaks in a pyrogram for identification purposes. The library under creation uses a scheme that selects peaks that are crucial to the identification of unknown compounds; these are then compared with known spectra, thus simplifying the analysis of the compounds.

It has been mentioned earlier that pyrolysis conditions will affect the resultant pyrograms and that different pyrolyzers may yield differing pyrolysis products. Therefore, a first step in the creation of the library was the selection of Py/GC/MS conditions. Using a pulse pyrolyzer, pyrograms were generated at three temperatures, 500, 600, and 850 °C, for ethyl methacrylate, and it became apparent that the highest temperature was too high because only small molecules remained and structural detail was lost. Pyrolysis at the lowest temperature

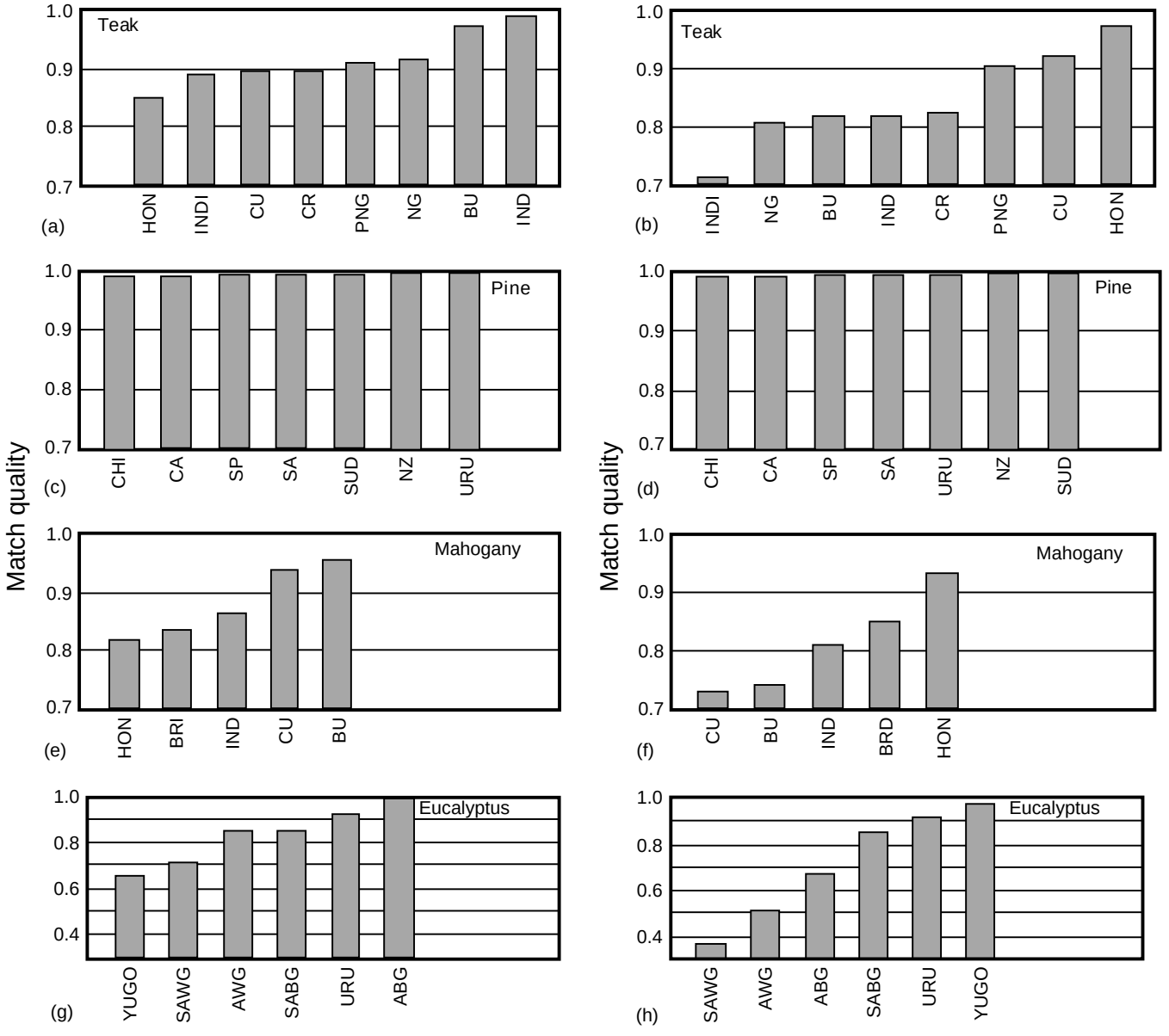

Figure 9 Use of Py/GC to differentiate wood species from different areas.

resulted in a few pyrolysis products that did not reveal much structural information. The more complex pyrograms were obtained from pyrolysis at 650 °C; both large and small fragments were obtained that were necessary for identification. The same process was repeated on other types of polymers and, for most of them, 650 °C gave good pyrograms. Library entries for polymers that do not pyrolyze well at this temperature will be done after pyrolysis at higher temperatures. The basic GC conditions required for the separation of polymers have been in use for some time. It is desirable to chromatograph to as high a temperature as possible. Temperature programming at 6 °C min^{-1} seems to be ideal for a 30-m capillary column. Faster programming results in the overlap of too many peaks, while a slower program results in very long analysis times.

The development of a Py/GC/MS library is a noble idea. The pyrolysis products of many polymers are unusual compounds that are not found in many commonly used mass spectral libraries and this makes it difficult to identify peaks even if the expected pyrolysis products are known. Creation of a library of polymer pyrolysis products by inclusion of peaks that are identified in the polymers should simplify the analysis of the polymers. However, inclusion of a select number of peaks in the library may hinder polymer identification as the library

becomes more complex. The authors do not want to include more peaks for the sake of simplifying the library. They are confident that the majority of polymers will have unique combinations of a limited number of peaks. Hence identification will not require determining the mass spectra of more than a few peaks.

The work done by Matheson et al.[35] has resulted in the production of a commercial Py/GC/MS library that, at present, is limited only to systems that use Hewlett-Packard (now Agilent Technologies Inc.) mass spectrometer equipment. Development of the library was based on the premise that ''the pattern of peaks in a chromatogram is more difficult to reproduce and recognize than the pattern in a mass spectrum, since the spacing along the *x* axis (time) is variable with column dimensions and pneumatic parameters. In addition, a reconstructed ion chromatogram consists of thousands of individual spectra, making it a very large data file, so storage of pyrograms for comparison rapidly uses disk space''.[36] To overcome these limitations, creation of the library takes advantage of the mass spectrometer data analysis software that permits the averaging of spectra under a peak to produce a single spectrum representative of a section of the chromatogram. Since there is no limit on the amount of time that may be incorporated into the averaging, an entire pyrogram may be averaged and stored as a single spectrum, requiring only about 0.02% of the disk space needed to store the original chromatogram. The averaged spectrum for the pyrogram may be added to the library of the mass spectrometer, or a new library of pyrolyzed polymer spectra created. The computer of the mass spectrometer is used to search and match these spectra to identify unknown polymers just as it does for the mass spectra of individual compounds. The averaged mass spectra for the pyrograms become a simulated pyrolysis mass spectrum, independent of the chromatographic conditions. It is believed that the combination of the averaged mass spectra and the pyrogram total ion chromatogram provides a very powerful two-pronged technique for the analysis of polymeric systems, since it permits analysis of the individual peaks in the pyrogram using standard searching libraries while providing the chromatographic pattern characteristic of the polymer, and it also permits the existing computer searching techniques to be applied to the polymer as a whole.[36]

A limitation of such a library is that the search capability is restricted to a polymer index in the library and users must literally carry the electronic files or hard-copy formats of the library so as to have access to it. A report by Wang[37] states that the creation and maintenance of an electronic database requires certain issues to be addressed, such as inputting data from different sources, maintenance (adding, deleting, and modifying entries), access and distribution, compatibility and integration with other databases, access from different computer platforms, and the cost of the software. The report shows how a hypertext markup language (HTML) approach to creating and maintaining a Py/GC database can address these issues. HTML is a standardized language used on the Worldwide Web. It is widely used and accepted because it is simple to program and easy to understand, is very efficient in linking files and images, and is independent of computer platform (IBM-PC, Macintosh, or any type of UNIX workstation). An unlimited number of users can view the database at the same time.

The mass spectra of complex pyrograms can be interpreted with the help of expert systems. For example, expert systems have been developed for the interpretation of pyrolysis mass spectra of condensation polymers such as polyamides and polyesters.[38–40] To develop the interpretation methods, the pyrolysis products were first analyzed and correlated with the degradation pathways and the structural characteristics of the polymers. An expert system was then developed to facilitate and automate the interpretation of the mass spectral characteristics. The expert system is based on rules that were empirically extracted but includes various algorithms for processing the information on the unknown mass spectra and sorting out the results of the search. The most significant information that is extracted from the unknown spectra is the repeating unit of the polymer. Determination of the repeating unit is done by using empirical algorithms and an autocorrelation algorithm.

4 PYROLYSIS ALKYLATION

A large number of compounds found in pulp and paper matrices contain carboxylic acid and phenolic groups. They include natural compounds such as fatty acids, resin acids, esters, soaps of fatty and resin acids, lignins, and hemicellulose adducts, and synthetic polymers such as polyesters, phenolic resins, and polymer additives. The synthetic polymers originate from the erosion of plastic equipment in the pulp and papermaking process and from the recycling of paper that contains plastics, glues, adhesives, coatings, and other polymer additives. Inadvertent introduction of plastic materials into the process, such as styrofoam cups, chewing gum, hard hats, pens, etc., is another source of synthetic polymers in pulp and paper mills. Once introduced into the process, these polymeric materials can be pulverized into small particles that affect productivity and product quality. A large number of these unwanted contaminants stick to process equipment to cause process upsets at the manufacturing site and/or at the customer site in the printing or conversion

processes. They are commonly referred to as ''stickies'', owing to their tacky characteristics. Mills that supply pulps to customers who are sensitive to plastic contaminants take extra measures to minimize the introduction of plastic materials in the process. These include banning the use of plastic cups and plastic pens by mill personnel and visitors alike, and requiring that logs brought into the mill have labels/tags that are not made of plastic.

Mills that make recycled papers take extra measures to try to reduce the amount of these contaminants in the final product. These include using high-quality furnish (low in synthetic polymeric content) and installing efficient cleaning and screening unit operations. If production problems occur due to deposition or accumulation of these contaminants, it is necessary to analyze them to determine their composition and then work backwards to ascertain their origin. Analytical pyrolysis is a valuable tool in the characterization of the synthetic polymers. The diverse synthetic polymers that may be found are listed in Table 3.

Pyrolysis of these compounds yields a wide range of compounds that vary from nonpolar (e.g. alkanes and alkenes) to highly polar (e.g. alcohols and carboxylic acids). If carboxylic acids are the starting compounds they can undergo decarboxylation during pyrolysis.[41] The polar compounds can give useful information about the structure of the polymer.[29] Examples include the pyrolysis of polyesters to give polybasic acids and alkenes,[42] epoxies to phenolic compounds,[43] and polyamides to amines and carboxylic acids.[44] The polar compounds are normally difficult to analyze by GC because they are prone to adsorption on wall surfaces in the pyrolyzer, the injection system, and the analytical column. Polar pyrolyzates often show peak-tailing characteristics, poor reproducibility, long chromatographic times, and in some cases loss of peaks.[29] In conventional GC, these problems can be overcome by derivatizing with an alkylating reagent, such as tetramethylammonium hydroxide (TMAH), prior to analysis by GC. The derivatization can be done externally or by in situ derivatization in the injection port of a gas chromatograph.[44,45] This saves time and is less likely to lead to sample loss during manipulation. These procedures, however, cannot be used for on-line Py/GC.

Instead, the pyrolysis products can be derivatized prior to GC analysis. An early example of this procedure was flash derivatization, with a butylating agent, of a copolymer of vinyl acetate, crotonic acid, and a complex vinyl ester. The reaction resulted in the formation of butyl acetate and a mixture of higher isomeric butyl esters.[46] Analysis of these compounds by normal Py/GC does not give results that are in agreement with wet chemical methods or spectroscopic techniques because they undergo decarboxylation during pyrolysis.[41] This problem can be overcome by pyrolytic alkylation. This technique has been studied and used in numerous applications.[44,45] It has been called by a number of terminologies including reactive pyrolysis, on-line pyrolysis, thermally assisted hydrolysis, pyrolysis methylation, and simultaneous pyrolysis methylation (mainly because the most widely used alkylating agent contains methyl groups). The mechanism of

Table 3 Unwanted synthetic polymers that are found in pulp and paper matrices

Contaminant	Definition	Sources of contaminant
Latex	Latex rubber used as an adhesive or as a lining or coating	Magazines, books, envelopes, sealing tapes, laminated boards
Latexes	Non-water-soluble latexes present in coating mixtures	Books, stationery, bottle tops, wallpaper
Non-water-soluble adhesives	Thermoplastic adhesives extensively used for bookbinding	Books, magazines, board
Fly pasting adhesives	Non-water-soluble adhesives used for joining reels on high-speed printing processes	Newspapers, magazines
Waxed laminates	Wax used to laminate two or more webs of paper or board	Books, cartons, bottle tops, laminated papers, paper tubes
Asphalt	Used for coating or laminating paper and board	Multiwall sacks, cartons, corrugated cases, laminated papers, packaging products
Parchment laminates	Vegetable parchment laminated to or within other grades of paper or board	Cartons, boxes, laminated boards
Wet-strength resins	Urea–formaldehyde resins used to impart wet strength	Books, photographic papers, envelopes, greeting cards, paper cups, paper towels, paper sacks and bags
Synthetic fiber	Synthetic fibers bonded together into a fabric	Books, stationery, greeting cards
Polyethylene, PVA, Saran™, cellulosic materials	Synthetic materials used as film formers, coatings or laminates in paper and board	Books, milk cartons, paper sacks and bags, paper cups, greeting cards
Inks	Inks containing synthetic resins, commonly used in rotogravure printing	Stationery, magazines, die-stamps, wallpaper

PVA, poly(vinyl acetate).

the reaction is shown in Equation (1):

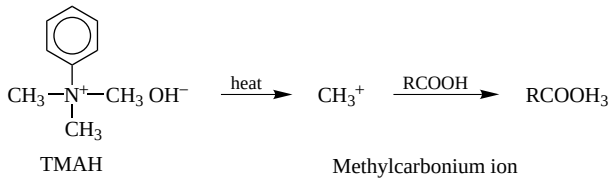

(1)

The methylcarbonium ion will react with compounds that have reactive amino, hydroxyl, and carboxyl functional groups.

The technique has been adapted for Py/GC analysis. Pyrolysis of carboxylic acids, carboxylic acid salts, esters, polyhydric alcohols, and phenolics in the presence of an alkylating reagent such as TMAH avoids decarboxylation and degradation of the compounds resulting in their alkylation.[29,47,48] The alkylation leads to improved chromatographic analysis of the derivatized compounds. For example, pyrolysis of polyesters involves scission of the RCOO–R bond resulting in the formation of the respective carboxylic acid and alkene.[42] Pyrolysis with a methylating agent results in the formation of the methyl esters of the carboxylic acids and methyl ethers of the alcohol. An example of the powerful effect of the derivatization is illustrated in Figure 10(a) and (b). It is evident that pyrolysis alkylation results in a clean and sharp pyrogram owing to methylation of the pyrolyzed rosin size fragments.

Analytical pyrolysis in the presence of a methylating agent but at a lower pyrolysis temperature has been termed thermally assisted hydrolysis or methylation (THM).[49] Analytical pyrolysis requires the absence of oxygen to avoid oxidation of the sample. The THM technique is especially sensitive to the presence of traces of oxygen because the strong alkaline medium facilitates autoxidation reactions.[50]

A typical pyrolysis alkylation procedure entails placing 2 μL of an alkylating agent such as TMAH (25% w/w aqueous solution) or tetrabutylammonium hydroxide (40% w/w aqueous solution) on the sample (~5 μg) placed in the pyrolyzer. The sample is then pyrolyzed and the pyrolysis products are analyzed by GC. Pyrolysis methylation is inappropriate for the derivatization of low-molecular-weight pyrolysis products that have very short retention times or co-elute with other pyrolysis products. For such compounds pyrolysis butylation results in the formation of butyl compounds that exhibit good chromatographic properties. The procedure has been used successfully for vinyl acetate, methacrylic acid, and cyanoacrylate homopolymers and copolymers.[29] Examples illustrating the use of pyrolysis alkylation in the analysis of synthetic polymers are given in section 8.

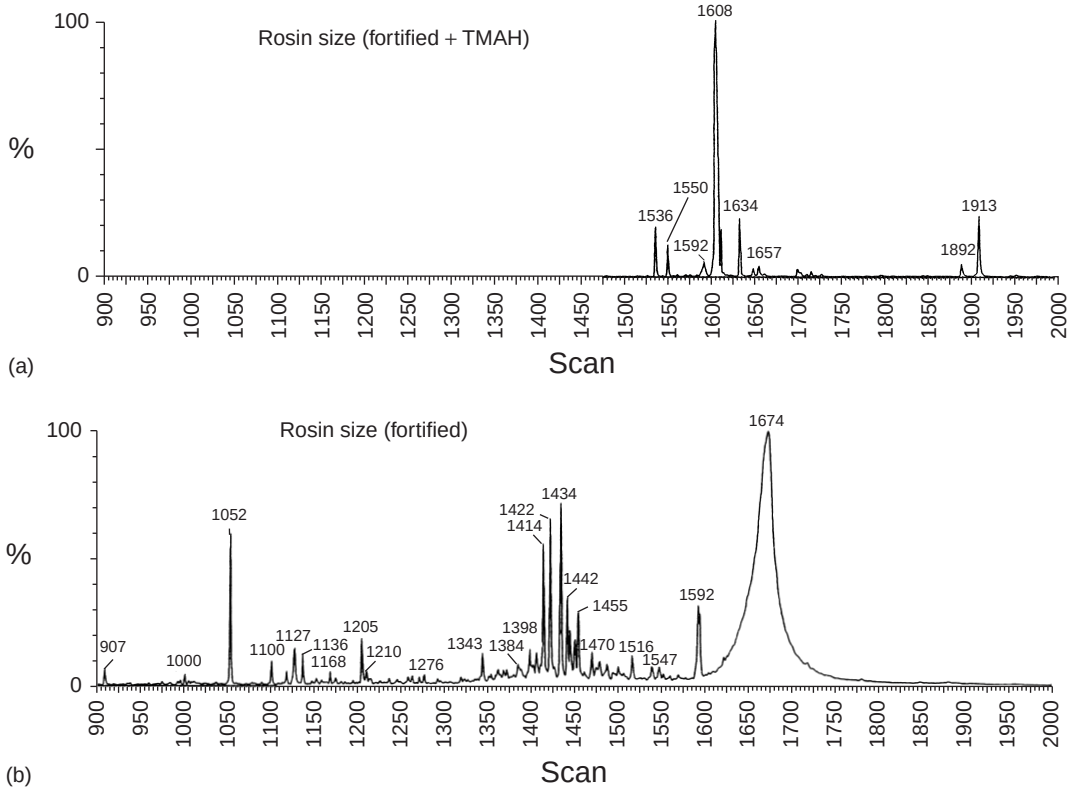

Figure 10 Effect of in situ alkylation on the pyrolysis of rosin size.

Although pyrolysis methylation has found widespread use in various pulp and paper matrices,[32,33] it has a limitation in that it is unable to distinguish between the methoxyl groups originally present in a polymer and the free hydroxyl groups that become methylated after pyrolysis. Such a distinction is needed in lignin chemistry where lignins are characterized by the ratio of phenolic groups to hydroxyl and methoxyl groups. A butylating agent does not have the limitation of its methyl counterpart. For example, Martin et al.[51] have shown that pyrolysis in the presence of tetrabutylammonium hydroxide results in derivatization of the free hydroxyl groups present in lignin before pyrolysis (to form *O*-butyl ether methoxyl groups). The mechanism is thought to be thermally assisted hydrolysis followed by alkylation, and not true pyrolysis followed by in situ alkylation of the products.

A potential problem with TMAH is that it can result in the production of multiple peaks in chromatograms owing to isomerization of the compound. Another problem is that the methylation can result in the production of benzenecarboxylic acids from unoxidized lignin.[52] Therefore, caution is required when interpreting the presence of benzenecarboxylic acids as pre-existing structural units of lignin.

Still another problem with pyrolysis methylation is that the process does not distinguish fatty acids from fatty acid esters. For example, analysis of wood extractives containing free fatty acids and glycerides of fatty acids will yield a pyrogram with peaks corresponding to methylated fatty acids with no indication of their origin: the free fatty acid will be methylated whereas the glycerides will be hydrolyzed first to release glycerol and fatty acids that will then also be methylated. A report by Hardell and Nilvebrant[53] shows that pyrolysis using a methylating agent TMAH and an acetylating agent (tetramethylammonium acetate (TMAAc)) can distinguish free and esterified acids according to Equations (2) and (3):

$$\underset{\text{free acid}}{\text{RCOOH}} + \underset{\text{triglyceride}}{\text{R}^1\text{COOR}^2} \xrightarrow{\text{TMAH}} \underset{\text{methylated acids}}{\text{RCOOMe} + \text{R}^1\text{COOMe}} \quad (+\ \text{R}^2\text{OMe}) \qquad (2)$$

$$\underset{\text{free acid}}{\text{RCOOH}} + \underset{\text{triglyceride}}{\text{R}^1\text{COOR}^2} \xrightarrow{\text{TMAAc}} \underset{\text{methylated acid}}{\text{RCOOMe}} + \underset{\text{triglyceride}}{\text{R}^1\text{COOR}^2} \qquad (3)$$

TMAAc has less alkalinity than TMAH and consequently does not degrade esterified acids.

Figure 11(a) and (b) illustrate the selectivity of the approach: reaction of TMAH with tripalmitin yields two

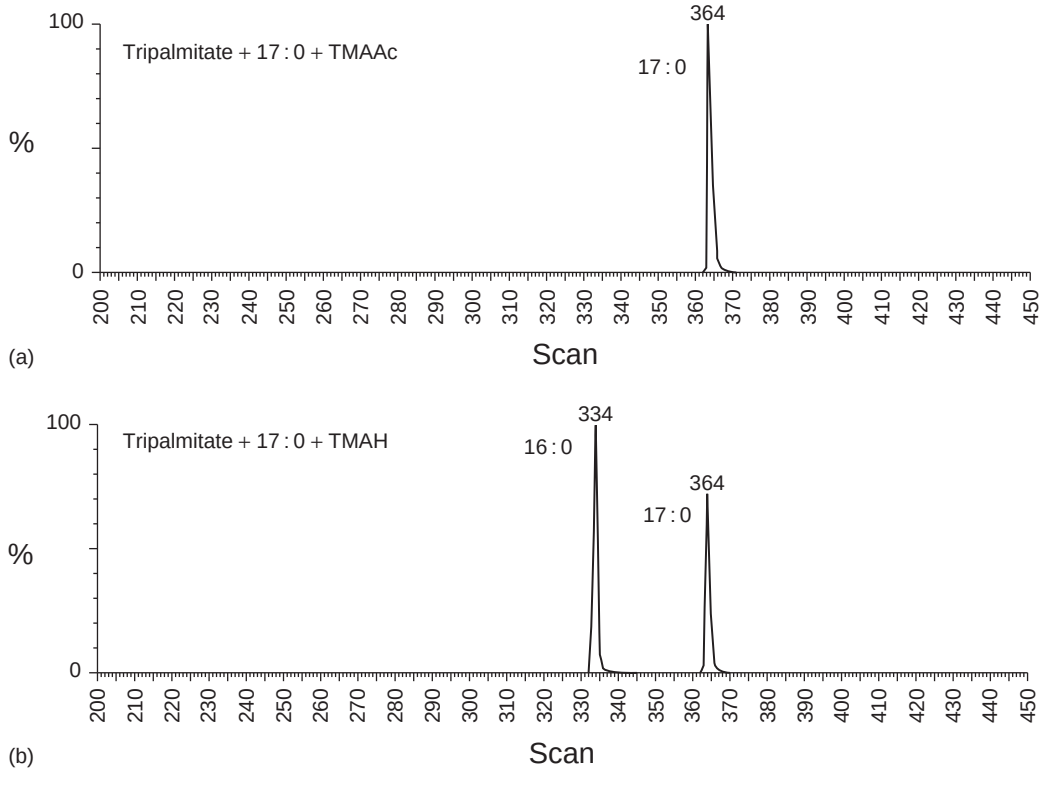

Figure 11 Effect of (a) TMAAc and (b) TMAH on the alkylation of triglycerides.

peaks corresponding to methylated palmitic acid and unreacted tripalmitin. Reaction of the same compound with TMAAc yields only one peak corresponding to tripalmitin. No other peak is detected, indicating that TMAAc does not degrade the tripalmitin.

Pyrolytic alkylation is different from fusion GC or derivatization prior to Py/GC analysis. It uses the alkylation agent to hydrolyze/derivatize polymers in the injection port or in the pyrolysis chamber. In fusion GC a large excess of the reagent is mixed with the polymer and the mixture is heated to above the melt temperature, resulting in hydrolysis or cleavage of the polymers and generating low-molecular weight compounds that are amenable to GC analysis.[54] Derivatization prior to Py/GC analysis entails reaction of functional groups in the polymer prior to analysis. This requires that the polymer be stable enough to resist attack from the derivatizing agent below the temperature at which pyrolysis takes place. The technique can be used to study the degradation mechanisms of polymers. For example, the degradation mechanisms of a poly(vinylcarboxylic acid) and its alkyl ester analog were studied after derivatization of the acid by reaction with TMAH for 2 h at 100 °C in an air-cooled reflux flask.[55]

5 APPLICATIONS OF ANALYTICAL PYROLYSIS

5.1 Analysis of Natural Polymers in Wood and Pulps

The major composition products of wood are cellulose, hemicelluloses, lignin, and extractives. The average composition of these compounds in softwood and hardwoods is illustrated in Figure 12.

Cellulose determines the character of wood fibers and enables them to be used for papermaking. It is a polysaccharide of glucose units with the formula $(C_6H_{10}O_5)_n$ where n is the number of repeating sugar units or the degree of polymerization. Decreasing the molecular weight of cellulose molecules below a certain level will cause deterioration in strength of the paper products. Hemicelluloses are polysaccharides of five different sugars from hexoses (glucose, mannose, galactose) and pentoses (xylose, arabinose). These sugars combine with uronic acids to form various polymeric structures that are associated with the cellulose portion of wood or with lignin. The hemicelluloses are susceptible to more degradation and dissolution than cellulose during chemical pulping. Lignin is an amorphous, highly polymerized substance that cements fibers together. It is a very complex polymer consisting primarily of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) groups that are linked together in a three-dimensional structure and whose concentrations differ in softwoods and in hardwoods. In general, hardwoods contain traces of hydroxyphenyl groups, 29–35% guaiacyl groups, and 40–60% syringyl groups, whereas softwoods contain traces of hydroxyphenyl groups, >95% guaiacyl groups, and 1% syringyl groups.[56] Usually, lignins are classified as G, G/S or H/G/S lignin depending on the presence of the different groups. The lignin is dissolved during chemical pulping to free the cellulosic fibers. Extractives are a diverse list of compounds that are soluble in water or lipophilic solvents: they include resin acids, fatty acids, fats, terpenoids, phenols, and alcohols.

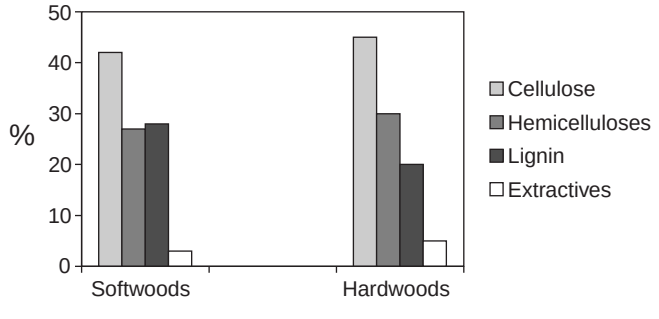

Figure 12 Concentrations of major wood components.

The main objective of chemical (kraft) pulping is to dissolve lignin and free the cellulosic fibers. The idea is to remove as much lignin as possible without doing too much damage to the cellulose and hemicellulose matter. It is therefore desirable to analyze for these compounds to monitor the pulping process. Traditional methods for measuring carbohydrates in kraft pulps are time-consuming; they entail acid hydrolysis, reduction to form alditols, acetylation of the alditols, and GC analysis of the acetates (see the article **Pulp and Paper Matrices**). The hydrolysis step can lead to erratic results because the polysaccharides in the pulps have different glycosidic linkages that do not hydrolyze at the same rate. For example, the hydrolysis of furanosides in xylans proceeds 10–1000 times faster than that of pyranosides.[57] Lignin is measured by gravimetry after its precipitation with acid or by spectrophotometry.

Py/GC can be used for the better characterization and analysis of native polymers, as will be shown in the next few subsections.

5.1.1 Cellulose

An early study on analysis of microcrystalline cellulose with a Curie point pyrolyzer showed that furfural and levoglucosenone were the major pyrolysis products.[58] Small amounts of 1,6-anhydro-3-deoxy-β-D-glucopyranosen were detected. Pyrolysis of glucose and oligosaccharide standards with up to five glucose units showed that very

small amounts of the cellulose pyrolysis products were obtained. From this, two pyrolysis mechanisms were proposed.

Later, Py/GC/MS with a Curie point pyrolyzer was used to analyze microcrystalline cellulose.[59] High-resolution MS identified 96 compounds that were classified into several groups, namely carbonyl compounds, acids and methyl esters, furans, pyrans, anhydro sugars, and hydrocarbons. Levoglucosan was the major product detected. The data enabled the authors to propose several mechanisms for the pyrolysis of the cellulose. Later work by Wada et al.,[60] using the same kind of pyrolyzer, showed that the degree of substitution of the constituent glucose unit influenced the amount of levoglucosan generated.

5.1.2 Carbohydrates

Py/GC studies of carbohydrates have shown that the method can successfully identify the saccharide composition of a wide range of heteropolysaccharides by the ability of the pyrolysis technique to produce individual anhydro sugar fragments which are further separated and identified by GC/MS.[61,62] However, quantitative analysis of polysaccharides has been difficult when noncarbohydrate materials and metals are present, as is the situation with wood and pulp samples. A study by Kelly et al.[63] showed that removing metal ions, e.g. by acid washing, overcomes the difficulty, resulting in the generation of reproducible yields of anhydro sugar products.

The technique is applicable to both mechanical and chemical pulps and the resultant pyrograms are similar to that of pure microcrystalline cellulose and a xylan polymer. The pyrograms reveal differences in the cellulose compositions of the pulps: high-yield mechanical pulps retain their original saccharide content, although some hemicellulose is lost during pulping; lower yield chemical pulps have much higher concentrations of cellulose and much lower concentrations of hemicellulose that is dissolved during cooking. Table 4 lists the identities of the pyrolysis products and shows that almost all the peaks are pyrolysis products of pentose or hexose units. This study showed that acid washing apparently reduces the concentrations of the unknown peaks.

Table 4 Compounds identified upon pyrolysis of wood pulps. (Adapted from Kelly et al.[63])

Peak no.	Molecular weight	Compound name
1	86	Unknown
2	74	1-Hydroxy-2-propanone
3	84	2(3*H*)-Furanone
4	96	2-Furaldehyde
5	110	5-Methyl-2-furaldehyde
6	126	Levoglucosenone
7	114	4-Hydroxy-5,6-dihydro-(2*H*)-pyran-2-one
8	142	Unknown
9	126	5-Hydroxymethyl-2-furaldehyde
10	144	1,5-Anhydro-4-deoxy-D-*glycero*-hex-1-en-3-ulose
11	132	1,4-Anhydroxylopyranose
12	162	1,6-Anhydrogalactopyranose
13	162	1,6-Anhydromannopyranose
14	162	1,4-Anhydrogalactopyranose
15	162	1,6-Anhydroglucopyranose
16	162	1,6-Anhydrogalactofuranose
17	162	1,6-Anhydroglucofuranose

5.1.3 Lignins

Lignins can be characterized by a number of methods that entail degradation of the compounds, the most common and simplest of which is oxidation with nitrobenzene.[64] However, these methods, including the simplest one, are time-consuming and tedious. Py/GC is ideal for characterizing lignin groups, as has been shown by a number of researchers.[65–68] For example, Faix et al.[65] studied lignins in several wood species by Py/GC/MS and off-line Py/GC. This extensive study addressed issues such as determining the best method for wood and lignin samples, the overall yields of pyrolysis products identified by GC, the reproducibility of the pyrolysis experiments, whether or not it is necessary to isolate lignin from lignocellulosics, and compared Py/GC/MS and off-line Py/GC with other degradation methods for lignin. The pyrolysis was done with a resistively heated pyrolyzer. Four lignocellulosic samples from beech, spruce, bamboo and teak wood, and lignins isolated from the samples were used in the study. Py/GC/MS analysis allowed the assignment of 67 mass spectra to the most prominent lignin-derived degradation products, most of which were phenols containing methoxy groups and C_1-C_3 saturated and unsaturated alkyl substituents. Fifty of these compounds could be identified by off-line Py/GC using authentic standards. A comparison of the off-line Py/GC (with a flame ionization detector) and Py/GC/MS pyrograms is shown in Figure 13(a) and (b). A major characteristic of the on-line pyrogram is the presence of a broad peak due to levoglucosan. It is evident that there are differences in the intensities of peaks in the two pyrograms and this is related to the peculiarities of the methods used and the specific sensitivities of their detectors. Conclusions from the study were that off-line Py/GC is a suitable method for lignin classification of lignicellulosic materials with no prior isolation of the lignin; pyrolysis allows the quantitative determination of H : G : S ratios of lignins and the results are similar to those obtained by nitrobenzene

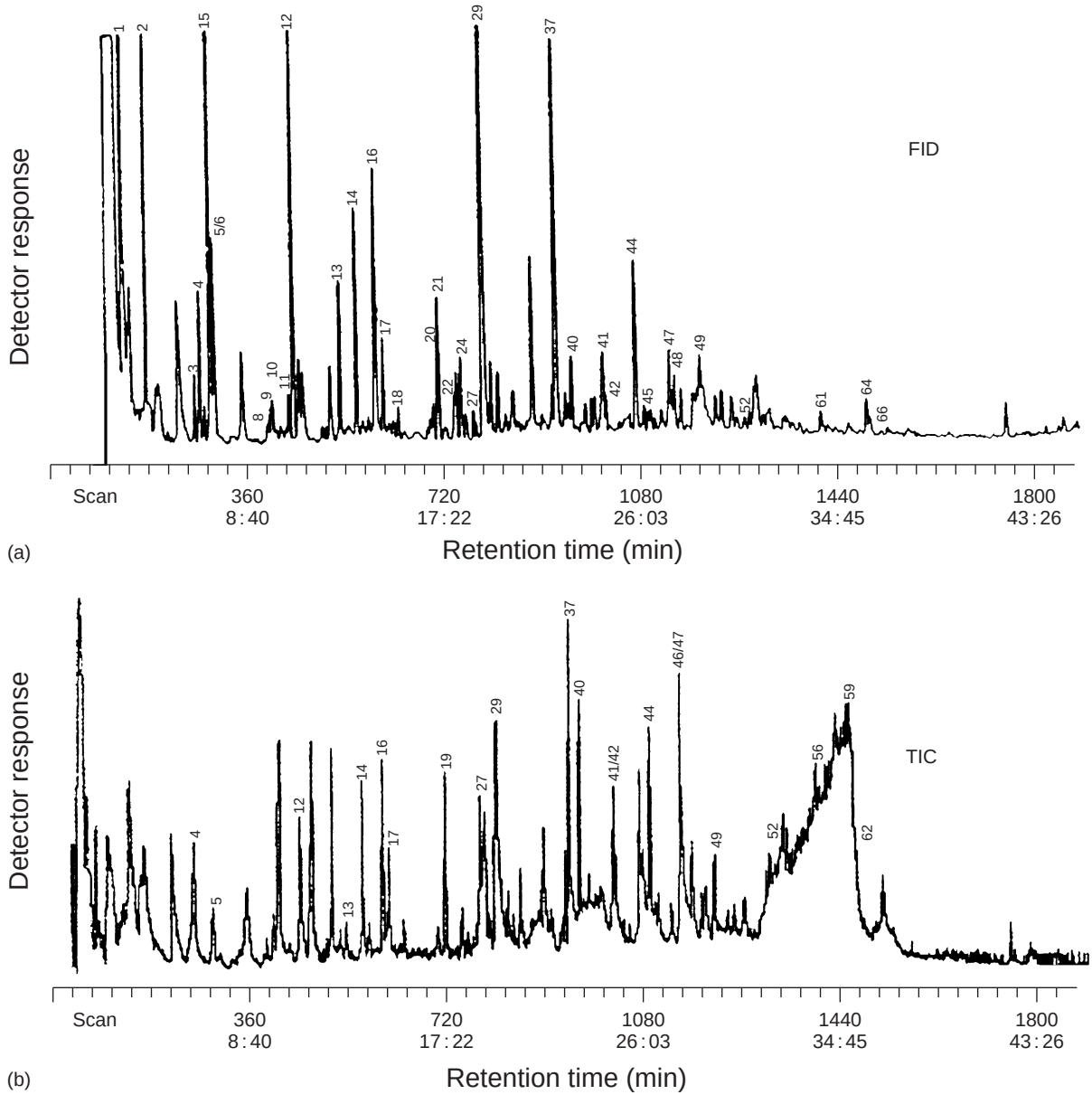

Figure 13 Comparison of (a) Py/GC and (b) Py/GC/MS of a lignin sample. (Reprinted from ref. 65, with permission from Elsevier Science.)

oxidation and FTIR analyses; Py/GC/MS facilitates the interpretation of complex lignocellulosic pyrograms, for example, by excluding ions of carbohydrate-derived degradation products.

Japanese researchers used Py/GC to evaluate the uniformity of delignification of beech wood chips by monitoring residual lignin in alkali pulps.[69] Analysis of the samples identified 10 pyrolysis products whose structures were related to vanillin, isoeugenol, syringols, and guaiacols. The calculated S : G ratio correlated with the Klason lignin contents of kraft and soda–anthraquinone beech pulps. Analysis of this ratio in samples of chip slices during a cook can be used to evaluate the uniformity of delignification.

The same researchers also used Py/GC to analyze lignin in different paper furnishes including printing and writing paper, wood-containing paper, wood-free paper, and wood-containing coated paper.[70] Their results showed that papers containing hardwood mechanical pulps could be clearly distinguished from the other papers, owing to the presence of syringyl-type pyrolysis products. An advantage of Py/GC over other traditional methods for lignin measurement such as the Klason technique is that it is not affected by the presence of fillers in the papers.

Other workers have used Py/GC to differentiate paperboards made from mechanical and biomechanical wheat-straw pulps and various proportions of wastepaper.[71] Their results showed that yields of methoxyphenols with an oxidized *n*-chain correlated with wheat-straw content, whereas the yields of cinnamyl and syringyl phenolics correlated with wastepaper content. In addition, the yields of six pyrolysis products, of which anhydro sugars were the major components, correlated with the pulping method.

Py/GC/MS has been used to determine the amount of lignin present in paper mill wastewaters.[72] The residual chemical oxygen demand (COD) in treated effluent was attributed to lignins that originated from conifers. The pyrolysis yielded guaiacol and 4-methylguiacol fragments that were used to quantify the lignin in the effluent. The former was the best compound for the quantification.

Detection of high-molecular-weight products provides information on building blocks of lignin. However, the compounds are not volatile enough to be detected by normal Py/GC but can be detected when the pyrolysis is performed in the presence of alkylating agents that increase their volatility (on-line pyrolysis alkylation). For example, Izumi et al.[67] used the technique to characterize the lignin of more than 30 hardwood species. The major products generated were syringaldehyde and guaiacol. The data obtained were compared with those given by the traditional nitrobenzene oxidation method that yields syringaldehyde and vanillin as the major products and the results showed that there was good linear correlation between the two ($R > 0.85$). The authors concluded from this that Py/GC is applicable to the analysis of lignocellulosic materials with a wide range of syringaldehyde to guaiacol ratios.

Hardell and Nilvebrant[53] used on-line pyrolysis methylation for the structural analysis of isolated and native lignins in wood and mechanical pulp. The samples studied were milled wood lignin (MWL) prepared from Scandinavian spruce (MWL and acetone-extracted lignin), and chemithermomechanical pulp (CTMP) bleached with hydrogen peroxide. The authors observed that normal pyrolysis resulted in low pyrolysis yields of the lignins due to loss or adsorption of the pyrolysis products on the GC column. When the pyrolysis was performed at lower temperatures in the presence of a methylating agent, the dominant lignin products retained more of their structural information, and previously nonvolatile carboxylic acids were methylated and made volatile. The compounds identified using on-line pyrolysis methylation are listed in Table 5.

Pyrolysis at 360 °C was sufficient for MWLs in the presence of a methylating agent. No pyrolysis products were obtained when the lignin was pyrolyzed in the absence of the methylating agent at this temperature. The pyrolysis products listed in Table 5 show that all the products were based on the expected monomeric lignin veratryl units. The main products contained side chains with methylated and carboxylic acid groups together with small amounts of keto and aldehyde groups. Fractionated pyrolysis of the lignin in the presence of the methylating agent, that is, initial pyrolysis at 360 °C followed by another pyrolysis of the same sample at a higher temperature (650 °C), showed that the pyrolysis products generated at the lower pyrolysis temperature accounted for 95% of the total pyrolysis products of both pyrolyses.

The results obtained by Hardell and Nilvebrant did not show any pyrolysis products that contained phenolic groups. This is contrary to literature data that showed the presence of phenolic compounds. Studies have shown that incomplete methylation is possible at times.[73,74] The authors speculate that the incomplete methylation may be due to limited access of the sample to the methylating agent, or to nonintimate contact between the sample and the pyrolysis filament. They conclude that the pyrolysis of MWL in the presence of a methylating agent can be interpreted as thermally assisted hydrolysis followed by in situ methylation rather than a traditional pyrolysis. This pyrolysis technique yields fragments that are in better agreement with the accepted structure of lignins than the fragments obtained after conventional pyrolysis.

Table 5 Compounds identified after on-line pyrolysis methylation of spruce MWL. (Adapted from Hardell and Nilvebrant[50] with permission from Elsevier Science.)

Compound[a]	Molecular weight	Relative amount
V−CH=CH_2	164	Minor
V−CH=CHMe (*cis*)	178	Trace
V−CH_2OMe	182	Major
V−CHO	166	Minor
V−CH_2CHO	180	Trace
V−CH=CHMe (*trans*)	178	Trace
V−CHOMe	180	Trace
V−COOMe	196	Major
V−CH=COMeMe	208	Trace
V−CH=CHOMe (*cis*)	194	Major
V−$CH_2CH_2CH_2$OMe	210	Major
V−CH=CHOMe (*trans*)	194	Major
V−CHOMeCH_2OMe	226	Major
V−COMe=CHMe (*cis*)	208	Minor
V−COMe=CHMe (*trans*)	208	Minor
V−CH=COMeCH_2OMe	238	Minor
V−CHOMeCH_2CH_2OMe	240	Minor
V−CH_2CHOMeCH_2OMe	240	Trace
V−CHOMeCOOMe	240	Major
V−CH(CH_3)OMeCOOMe	254	Major
V−CHOMeCHOMeCH_2OMe	270	Major
V−CHOMeCHOMeCH_2OMe	270	Major
V−COMe=$CHCH_2$OMe (*cis*)	238	Trace
V−COMe=$CHCH_2$OMe (*trans*)	238	Minor
V−CH=CHCOOMe	222	Trace

[a] V = veratryl (3,4-dimethoxyphenyl).

On-line pyrolysis methylation of lignin at a higher temperature does yield pyrolysis products that contain phenolic compounds. For example, data generated by Kuroda and Izumi[75] showed that pyrolysis at 500 °C released high-molecular-weight compounds that were volatile enough to be detected by GC. Pyrolysis of five lignin dimer model compounds provided little information on the β-O-4 lignin building blocks. However, the technique was found to be ideal for the analysis of the β-β, β-5, and β-1 building blocks in lignin.

Py/GC studies on lignins seldom involve absolute quantitation of the lignin in the analyzed samples. The few that have been used are based on off-line Py/GC that entails trapping of the pyrolysis products for subsequent analysis. Obviously this methodology defeats some of the advantages of analytical pyrolysis, such as on-line analysis and minimal sample preparation. A report by Bocchini et al.[76] shows that absolute quantitation of lignin by Py/GC/MS is possible using an internal standard. Tests on linearity and reproducibility of response of several standards showed that 1,3,5-tri-*tert*-butylbenzene was the best. The compound vaporized in the hot pyrolysis interface during the 3-min equilibration and condensed as a plug on the GC column, thereby avoiding its loss by thermal fragmentation. Correction factors for the main lignin pyrolysis fragments were calculated and the technique was then used to analyze a wheat straw sample. The results for pyrolysis fragments of the lignin are shown in Table 6.

The data show that the most abundant compounds were phenolics, namely 4-vinylguaiacol, 4-vinylphenol, 2,6-dimethylphenol, vanillin, and 4-vinyl-2,6-dimethoxyphenol. These compounds are characteristic of the pyrolysis of lignocellulosics and are products of thermal cleavage at different sites of the phenylpropanoid structure of lignin. According to the authors, the total amount of lignin listed in Table 6 is consistent with bibliographic data obtained using classical analytical techniques such as neutral-detergent fiber analysis and Klason lignin.

Studies have shown that lignin and carbohydrates can be analyzed simultaneously in both mechanical and chemical pulps by Py/GC/MS.[15,77] The studies also showed that, for chemical pulps, there was a relatively good correlation between the sugar components (arabinose, xylose, mannose, galactose, and glucose) determined by Py/GC/MS and those obtained by traditional methods. Changes in the chemical composition of softwood pulp fibers induced by kraft pulping were monitored by measuring the amounts of lignin and carbohydrates in kraft pulps using analytical pyrolysis and multivariate data analysis, namely PCA and PLS analysis. To quantify the lignin and carbohydrates in the pulp, a multivariate calibration model was calculated by correlating the Py/GC/MS results with those obtained from the traditional acid hydrolysis/derivatization GC procedure. The ability of the calibration model to predict the lignin and carbohydrate content was studied using a number of test samples whose Klason lignin content varied from 85 to 25 and pulp yields from 59.7 to 43.1%. The lignin and the main carbohydrate components (glucose, xylose, and mannose) in the test samples could be readily predicted from Py/GC/MS analysis. For the minor sugar components (arabinose and galactose), the difference between the predicted values from Py/GC/MS and those obtained by acid hydrolysis was larger owing to the poor reproducibility of their analysis by the wet-chemical method. The reproducibility of the quantitative analysis of lignin and the five main sugars in kraft pulp was better in Py/GC/MS than in the traditional method using acid hydrolysis/derivatization GC. This was proof that improved quantification of lignin and the main sugar components can be achieved by using Py/GC/MS.

Table 6 Py/GC/MS of wheat straw lignin using an internal standard. (Adapted from Bocchini et al.[76] with permission from Elsevier Science.)

Peak	Compound	Concentration ($g\,kg^{-1}$)
1	Phenol	0.75
2	2-Methylphenol	0.15
3	4-Methylphenol	0.30
4	Guaiacol	1.34
5	4-Ethylphenol	0.11
6	4-Methylguaiacol	1.33
7	4-Vinylphenol	2.36
8	4-Ethylguaiacol	0.61
9	4-Vinylguaiacol	5.60
10	2,6-Dimethoxyphenol	1.76
11	Eugenol	0.06
12	Vanillin	1.42
13	*cis*-Isoeugenol	0.08
14	2,6-Dimethoxy-4-methylphenol + *trans*-isoeugenol	1.48
15	Homovanillin	0.54
16	1-(4-Hydroxy-3-methoxy-phenyl)propyne	trace (<0.005)
17	Acetovanillone	0.02
18	4-Ethyl-2,6-dimethoxyphenol	0.04
19	Guaiacylacetone	0.37
20	4-Vinyl-2,6-dimethoxyphenol	1.81
21	4-Allyl-2,6-dimethoxyphenol	0.33
22	4-Propyl-2,6-dimethoxyphenol	0.10
23	*cis*-4-Propenyl-2,6-dimethoxyphenol	0.19
24	Syringaldehyde	1.14
25	*trans*-4-Propenyl-2,6-dimethoxyphenol	0.93
26	Acetosyringone	0.18
27	*trans*-Coniferyl alcohol	0.33
28	Syringylacetone	0.12
	Total	23.5
	RSD (total) (%)	8.0

Further studies have confirmed that Py/GC/MS can be used for the simultaneous analysis of polysaccharides and lignins. For example, Kleen and Lindstrom[78] used the technique to analyze the compounds in newsprint process waters. The pyrolysis products were compared with those obtained from pyrolysis of CTMP, its isolated fractions, and the newsprint paper itself, to ascertain the characteristic features of the dissolved polysaccharides and lignins in the process waters. In addition, the lignins and polysaccharides in fines were compared with those present on the long fiber fractions. The results showed that the main compounds in the dissolved substances were galactomannan, arabinogalactan and polyarabinan that originated from hemicelluloses and pectic compounds. Only small amounts of xylan were found. Lignin or lignin-like compounds with a chemical structure different from that of the lignin in the fines were found to be closely associated with the polysaccharides dissolved in the waters. The authors conclude that the lignin in CTMP, especially the fines fraction, is enriched with coniferyl alcohol units.

Py/GC/MS studies by Hardell and Nilvebrant[50,53] have also shown that the technique can be used for simultaneous analysis of lignin and carbohydrates. Their studies on a lignocellulosic sample by THM with pyrolysis at 360 °C showed that lignin fragments were the main products but small amounts of fragments derived from carbohydrates were also observed.[50] Analysis by conventional analytical pyrolysis had previously led to misidentification of the carbohydrates as products derived from syringyl lignin units.[79] Pyrolysis of a methoxyhydroquinone standard using the THM technique confirmed the presence of the carbohydrate, leading the authors to interpret this as evidence of the presence of carbohydrate–lignin linkages.

Analytical pyrolysis has been used to characterize chlorinated organic compounds in pulp and paper samples. For example, Van Loon and Boon[80] used the techniques for qualitative analysis of chlorolignins and lignosulfonates in pulp mill effluents discharged into the Rhine river. Three different kinds of compounds were isolated from the effluents: high-molecular-weight macromolecules that were isolated by ultrafiltration and concentrated by freeze-drying; volatile compounds that were isolated by a closed-loop stripping analysis system before trapping on to carbon and desorption with carbon disulfide; and apolar and medium-polarity compounds that were isolated by adsorption on XAD-4 resins followed by desorption with dichloromethane and then concentration by evaporation to small volumes. The high-molecular-weight compounds were studied by Py/MS using a resistively heated pyrolyzer and the results showed small peaks due to guaiacyl or syringyl derivatives and intense peaks from small sulfur- or chlorine-containing products, namely H_2S, CH_3SH, CH_3SCH_3, SO_2, CS_2, CH_3Cl, and HCl. The authors explained that this was due to the highly sulfonated, chlorinated, and oxidized nature of the chlorolignins and lignosulfonates that consequently have relatively low aromatic contents and high aliphatic group contents. The guaiacyl derivatives were detected in effluents from a mill pulping softwood, whereas the syringyl derivatives were found in hardwood mill effluents. This illustrates that analysis of mill effluents by Py/MS can be used to determine the wood furnish used in the mill. Further study of the generation of SO_2 led the authors to conclude that the technique can also be used for specific quantitative determination of lignosulfonates. The high-molecular-weight material was also studied by Py/GC/MS using a Curie point pyrolyzer and the results showed that about 150 products were obtained that varied in molecular weight from 44 (carbon dioxide) to 272 (1,2-diguaiacylethene). Chloroguaiacols, structurally very specific pyrolysis products of aromatic chlorolignins, were also observed and were proposed as chlorolignin markers for quantitative determination of chlorolignins in process waters.

Later studies by Koda et al.[81] described a Py/GC/MS procedure to determine the fate of high-molecular-weight chloroorganics produced by chlorine bleaching. The compounds were extracted from chlorinated pulps with aqueous dioxane or aqueous sodium hydroxide and the extracts were then subjected to ultrafiltration (cut-off 10 000 Da) to obtain the high-molecular-weight compounds. The compounds were analyzed by Py/GC/MS, using a microfurnace pyrolyzer, and the pyrolysis products were compared with those obtained after pyrolysis of MWL and residual lignins isolated from oxygen- and ozone-bleached pulps. The data showed that the major pyrolysis products in all the samples were alkane derivatives but pyrolyzates containing aromatic structures were also present in the chlorine-bleached lignins. This is the first study showing the presence of alkanes after pyrolysis of lignins – the authors postulated that the alkanes are generated by in situ hydrogenation reaction of some common precursors during pyrolysis. The most significant difference between chlorine-oxidized lignins and other oxidized lignins was the presence of small amounts of chlorinated phenols in the chlorine-bleached lignins.

Flodin et al.[82] used Py/GC and Py/GC/MS techniques to search systematically for halogenated structures in natural organic matter. Results for analytical pyrolysis with on-line methylation on aquatic fulvic acid (FA) and two samples of high-molecular-weight organic matter isolated from bleached kraft mill effluents (BKME) strongly indicated the presence of chlorinated 4-hydroxyphenyl structures in the original FA samples, whereas the BKME samples showed that the main chlorinated

pyrolysis products were guaiacyl- and syringyl-derived compounds.

5.2 Analysis of Deposits and Impurities by Normal Pyrolysis

Pulp and papermaking processes are sometimes plagued with deposition problems that cause breaks and/or off-quality products in the form of dirt and specks in the product. Such products are culled or sold at below premium prices. Rapid characterization of the deposits is essential in order to find solutions to the problems. Analytical methods have been developed to determine the composition of the deposits; they range from simple spot tests[83] to sophisticated techniques that include solvent extraction and instrumental analysis by FTIR spectroscopy, GC, GC/MS, and scanning electron microscopy.[84] In many instances, the concentrations of impurities in pulp and paper products are too low for analysis by solvent extraction followed by GC analysis of the extracts. FTIR spectroscopy with a diamond cell or a microscope can be used to characterize the dirt spots but it may not give definite answers on the composition of the spots. Analytical pyrolysis offers an alternative rapid technique to characterize these deposits. No (or minimal) sample preparation is required and very small amounts are required. The following examples illustrate applications of Py/GC in troubleshooting deposition problems in pulp and paper mills.

5.2.1 Deposit Sample from a Coated Paper Mill

A coated paper mill employing a mixture of groundwood and kraft pulps was experiencing deposition problems on the paper machine foil blades. The deposit was grey in color and had a rubber-like texture. Py/GC analysis of a sample (see Figure 14a and b) showed the presence of latex, styrene–butadiene copolymer, in the deposit.[85] The copolymer is a major component in paper coating formulations. In this case, repulping of off-quality coated paper was the cause of the deposition problem. The amount of latex in the deposit was determined to be 10.3% by comparing the peak area at 7.37 min with peak areas of copolymer standards at the same retention time. The larger peak eluting at 4.90 min was not used for quantitation because its total area was not reproducible, and is probably due to other components in the latex formulation such as stabilizers, preservatives, or emulsifiers.

5.2.2 Sticky Deposit from a Fine Paper Mill

A fine paper mill had a deposit that seemed to have a texture usually associated with sticky deposits found in mills that use or make recycled papers. This was a puzzle

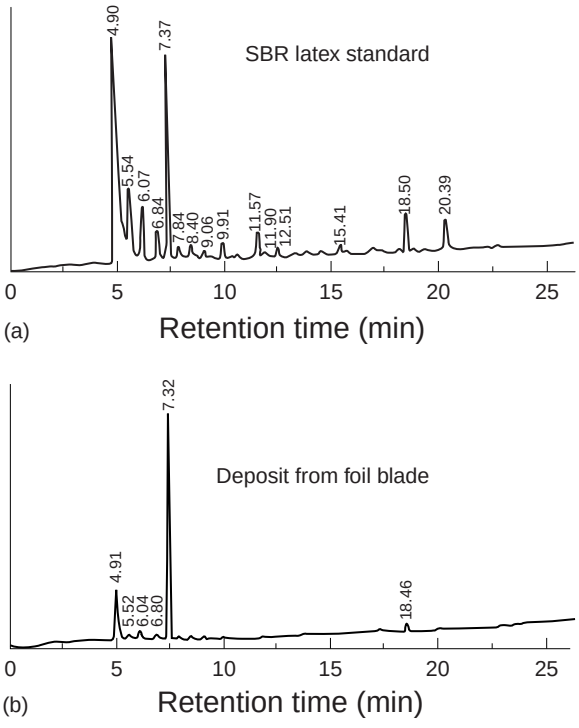

Figure 14 Analysis of a mill deposit by Py/GC. (First published in ref. 85 and reproduced here by kind permission of the Journal.)

since the mill was using virgin fiber with no recycled content. The texture of the deposit was very similar to that of chewing gum, which led to speculation that chewing gum could have been inadvertently introduced into the mill process. Py/GC analysis of the deposit sample and a sample of masticated chewing gum indicated that most of the peaks in the pyrogram of the deposit sample matched those of the chewing gum sample, as shown in Figure 15(a) and (b). Quantitation using the peaks areas at 5.65 min showed that there was 80% gum in the deposit.

5.2.3 Streaks on Tissue Paper from a Tissue Mill

A tissue paper mill experienced production problems due to plastic-like streaks that were stuck on to the paper in the dryer section of the mill. Py/GC indicated that the contaminant was due to PVA, as shown in Figure 16(a) and (b). Quantitation of the PVA in the sample showed that the sample was 100% PVA. The polymer had been introduced into the mill system with recycled paper and was melting to cause streaking in the dryer section.

5.2.4 Impurities in a Totally Chlorine-free Bleached Pulp

Py/GC/MS was one of the techniques that was used for the analysis of impurities that occurred during the

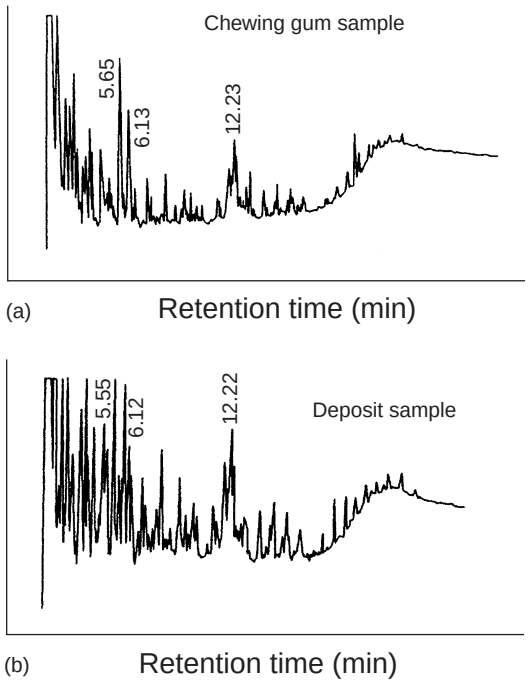

Figure 15 Analysis of a mill deposit by Py/GC. (First published in ref. 85 and reproduced here by kind permission of the Journal.)

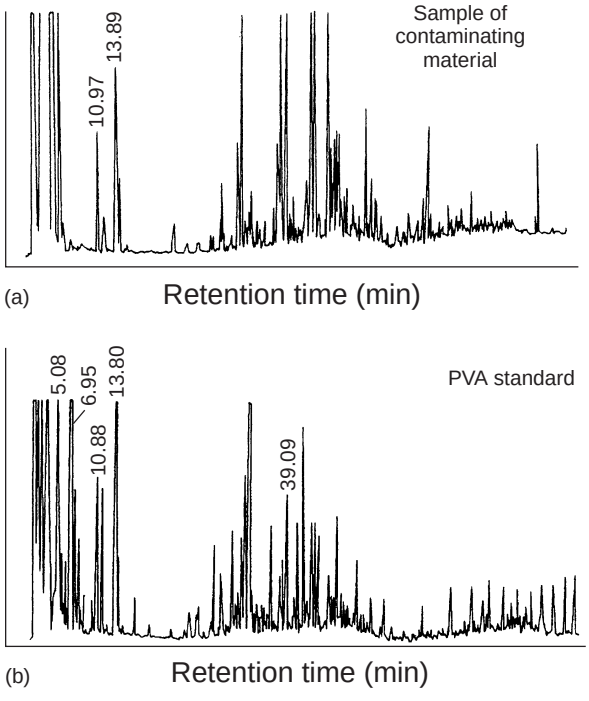

Figure 16 Analysis of a mill deposit problem by Py/GC. (First published in ref. 85 and reproduced here by kind permission of the Journal.)

manufacture of kraft pulp in a mill that practised totally chlorine-free bleaching on Iberian Eucalyptus.[86]

Analysis of the material left after solvent extraction showed that it was composed of small particles of isoprene rubber whereas the lipophilic material was constituted of wood extractives.

5.3 Analysis of Deposits by Pyrolysis Alkylation

Many of the compounds present in deposits and impurities contain polar groups such as carboxylic acid and alcohols. Their pyrolysis products are difficult to analyze on a GC column, but the analysis can be improved by pyrolysis alkylation of the spots. Ideally the impurities should be extracted from the fiber matrix using a pair of tweezers. Wetting the paper and observation under a stereo microscope is helpful in doing the extraction. Studies by Hardell[32] showed that a resistively heated pyrolyzer can be used for the analysis of the impurities. The sample is placed in the middle of the platinum filament and ~0.5 μL of TMAH (10% w/w aqueous solution) is added to the sample using a syringe. The mixture is dried before inserting the probe into the pyrolysis chamber. The sample is pyrolyzed for 2 s at 600 °C. The pyrolysis products are analyzed by GC or GC/MS. The following are examples of this analytical technique.

5.3.1 Specks in Paperboard

A multilayer paperboard product was contaminated with specks that occurred between the two layers of board. The paper was wetted and the specks were removed with as few fibers as possible. Analysis by pyrolysis alkylation showed that the main products were resin acid and maleo-resin acid methyl esters.[32] Fortified sizes were produced by treatment of rosin with maleic acid (or maleic anhydride) to form maleo-resin acid adducts. The presence of these compounds in the specks was therefore evidence that fortified rosin size was the cause of the dirt problem.

5.3.2 Cratering in "Blown Vinyl" Wallpaper

A wallpaper product, consisting of a paper sheet coated with a layer of poly(vinyl chloride), showed defects in the form of craters that measured about 1 mm. In some cases a colored spot was present underneath the crater. Analysis of material from one of the spots in the paper by normal Py/GC showed one main product, a di-ester of *o*-phthalic acid. The compound is widely used as a plasticizer. Pyrolysis alkylation of the spot revealed that there was no wood resin in the spot and the single phthalate ester peak had been transformed into three peaks corresponding to partial and total alkylation of the original plasticizer.[32] Figure 17(a) and (b) show pyrograms of the spot with and without alkylation.

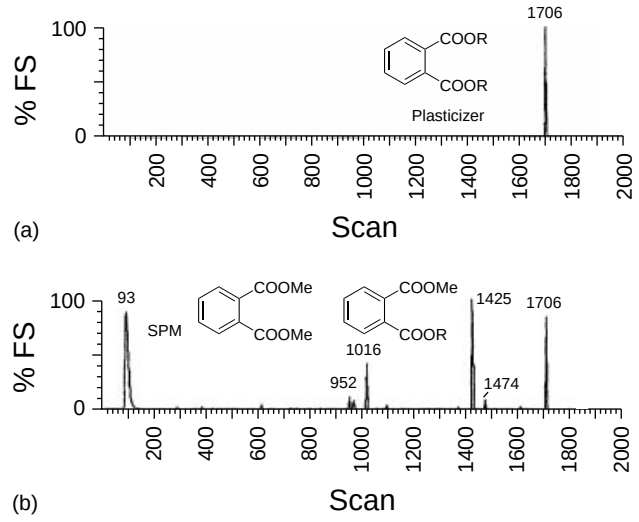

Figure 17 Use of pyrolysis alkylation to determine the identity of a spot found on paper. (Reproduced from ref. 32, with permission from Elsevier Science.)

Py/GC/MS is very useful in tracking down the origin of deposits. Deposits induced by hardwood resin will contain, among other compounds, saturated and unsaturated fatty acids, those induced by softwood resin will contain the same components plus resin acids, those from rosin size will contain only resin acids, and those from alkyl ketene dimer (AKD) will contain derivatives of saturated fatty acids. Py/GC can differentiate among these components.[32] This is very useful for mills that sell market pulps to customers who blend the pulps with others from different suppliers. If the customers experience deposition problems, Py/GC can help in determining which pulp or process caused the deposition problem.

5.4 Analysis of Deposition Problems in Coated Papers

Coated papers are made by coating with formulations that contain latexes. If the finished product does not meet specifications, it is reslushed and added to the papermaking furnish. The use of this reslushed furnish, termed coated broke, can cause sticky deposit problems because of the latex components that contain tacky materials. Such deposition problems are called white pitch to differentiate them from the deposits caused by wood resin that tend to be black and/or brown. The major component of latex is a tacky styrene butadiene rubber. The white pitch problems are often solved by gradual introduction of the broke into the main furnish. Knowing the amount of latex in the broke is useful in determining how much of the broke should be added. Py/GC has been used to determine the amount of latex in coated broke. Py/GC of a commercial latex sample showed that the main pyrolysis products were butylene, methylmethacrylic acid, and styrene. Quantitation of latex in the broke was calculated based on the peak intensity of the styrene peak. Analysis of a mill sample showed that 68% of the latex was in the fines fraction of the broke, 28% was in the coarse fiber fraction, and 3% was associated with dissolved and colloidal substances.[87]

5.5 Analysis of Rosin Sizing Agents

The sizing of paper entails attaching or anchoring of sizing agents on to cellulosic fibers. Determination of the amount of size present on sized papers is difficult to do by solvent extraction because the solvents do not remove the bound sizes. Py/GC with on-line methylation using TMAH has been used to measure rosin sizing agents in paper. The pyrolysis products, methyl esters of the rosin adducts, were used to measure the rosin size. The rosins retained in the paper were determined with a standard deviation of about 1%.[88]

5.6 Analysis of Alkyl Ketene Dimer and Alkenyl Succinic Anhydride Sizing Agents

AKD and alkenyl succinic anhydride (ASA) are neutral sizes added to paper and board to impart hydrophobicity. These sizes are considered to be reactive sizes that react with cellulose fibers thereby imparting the hydrophobicity.[89] Because these sizes can hydrolyze during storage or in the papermaking process before they react with fibers, it is important to ascertain how much of the compounds are present in sized sheets in order to optimize the sizing process. It is also important to measure the hydrolyzed components as they are sticky and can cause production problems. Methods have been developed for measuring these sizes in paper and deposit samples, but they are long and complex.[90,91]

Py/GC offers an alternative and rapid method for the direct measurement of the sizes in paper samples. A report by Yano et al.[92] showed that Py/GC could be used to identify AKD and ASA and the former was shown to correlate with the hydrophobicity of the paper. Vrbanac and Dixon[93] also used the technique for quantitative analysis of the sizes in papers. Resistively heated pyrolyzers connected to a GC or GC/MS system were used. Figure 18(a) and (b) show that the pyrogram of an AKD standard matches very well that of a paper sample sized with AKD. Pyrolysis of AKD yields three main peaks corresponding to 12/14, 14/14, and 14/16 carbon side-chain lengths on the AKD. Natural polymers in the paper do not interfere because they elute much earlier than the AKD peaks. By spiking known amounts of AKD in toluene on to unsized paper, a linear calibration plot was developed using areas of the three main peaks. The calibration was used to show that the amount of size in the paper was 6.82 lb t^{-1}. However, a

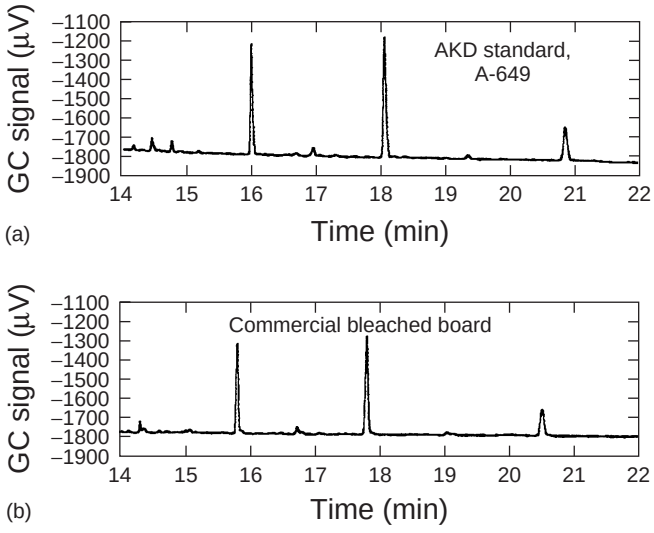

Figure 18 Analysis of AKD sizing agent by Py/GC.

benchmark of three commercial liquid packaging board samples showed that AKD content as measured by Py/GC does not necessarily correlate with the off-machine sizing test that is used to evaluate the sizing or hydrophobicity of the sized papers. While the sizing tests showed significant differences between the three samples, the AKD contents of two of the papers were not statistically different.

Pyrolysis of ASA (structure shown in Figure 19a) results in a complex pyrogram as shown in Figure 20. The pattern of the pyrogram is the same irrespective of whether or not pyrolysis alkylation is performed. This complexity indicates that ASA has a large number of homologs or there is chain breakage during pyrolysis, yielding numerous pyrolysis products. A mass spectral library search of the pyrolysis products on a resistively heated pyroprobe linked to a GC column and an ion trap quadrupole mass spectrometer shows octadecenyl-succinic acid, dodecen-1-yl-(−)-succinic anhydride and a pentamethylheptadecatetraenecarboxylic acid as some of the major degradation products.[94] Data from Vrbanac and Dixon[93] show that the pyrolysis calibration curve

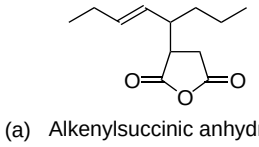

(a) Alkenylsuccinic anhydride (ASA)

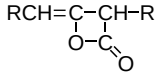

(b) Alkyl ketene dimer (AKD)
R varies from C_{14} to C_{20}

Figure 19 Structures of AKD and ASA internal sizing agents.

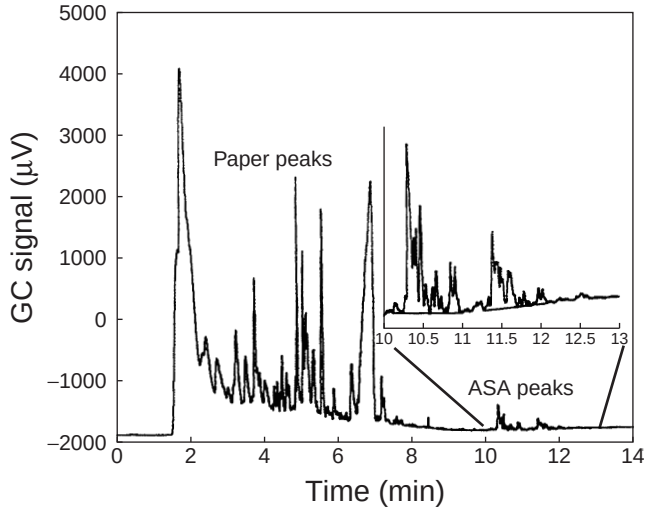

Figure 20 Pyrolysis of ASA yields a complex pyrogram (ASA in toluene spiked on paper, $1\,g\,L^{-1}$ concentration, $1\,\mu L$ of solution).

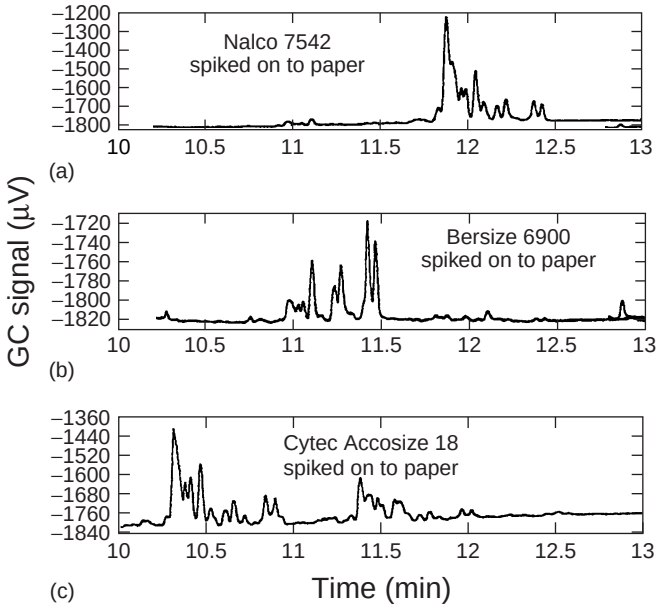

Figure 21 ASA samples from different manufacturers yield different pyrograms.

for ASA is not linear but can still be used to calculate the amount of size in paper. Analysis of ASA products from different suppliers shows that they have different pyrograms as shown in Figure 21(a–c). Each ASA has a different peak pattern, confirming that the isomeric olefin stocks for making ASA are different. These peak patterns can be used for qualitative identification of the sizes. An unfortunate aspect of Py/GC of ASA is that hydrolyzed ASA cannot be distinguished from unhydrolyzed ASA because their pyrograms are similar. This means that the

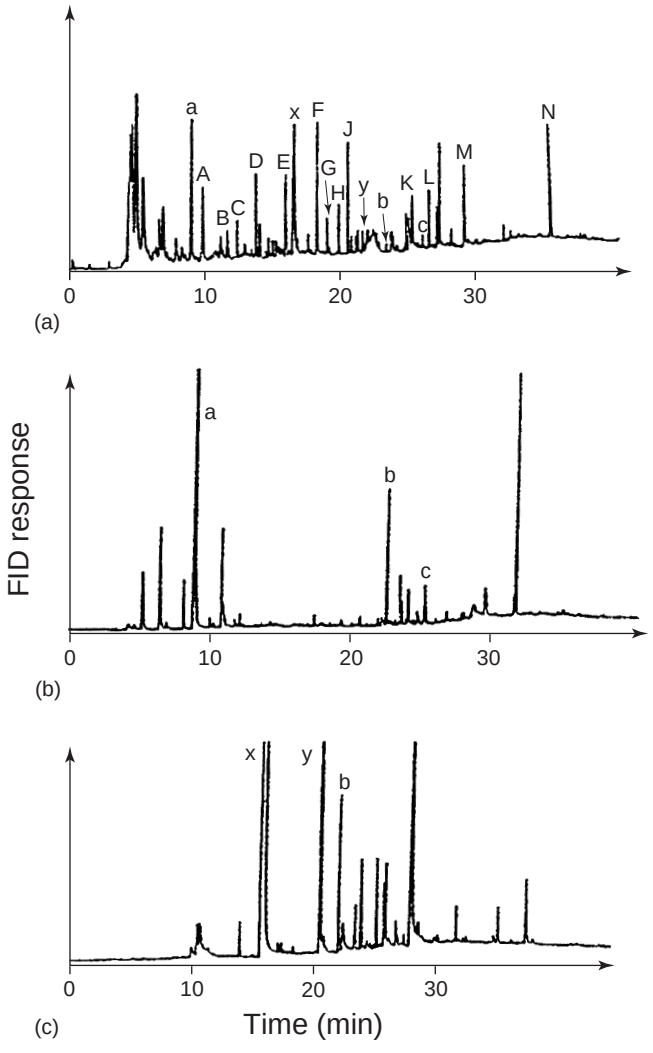

Figure 22 Pyrolysis of (a) black ink on paper and (b) black ink at 650 °C and (c) headspace of analysis of ink at 280 °C. Peaks A–N are pyrolysis fragments of paper alone; a–c result from pyrolysis of ink alone; x and y are volatiles from ink. (Copyright *LC–GC Magazine*. Reprinted with permission.)

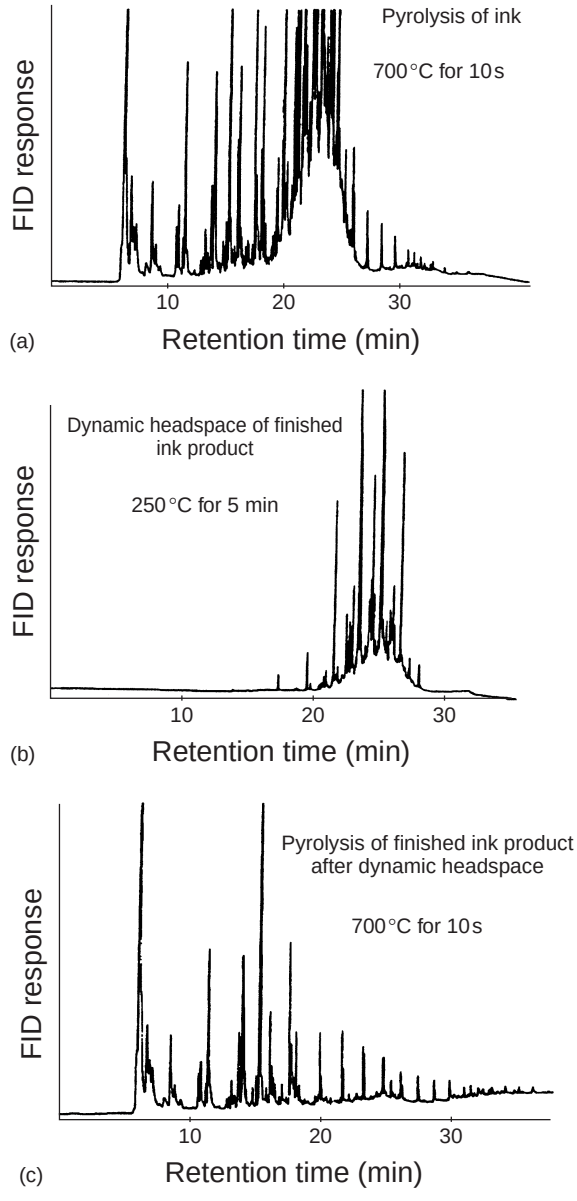

Figure 23 Analysis of ink by Py/GC and dynamic headspace sampling. (Copyright CDS Analytical Inc. Reprinted with permission.)

technique measures total ASA in the sheet with no indication of the amount of ASA that has reacted to impart sizing on fibers. Indeed, the use of Py/GC to diagnose the cause of poor sizing on a commercial paper production revealed that ASA content does not correlate with the hydrophobicity of the paper. Analysis of a paper with normal hydrophobicity and another with low hydrophobicity showed that the problem paper contained 26% more ASA than the normal paper. The cause of the lower hydrophobicity was thought to be either hydrolysis of the ASA or poor distribution and orientation of ASA on the fibers.[93]

Py/GC has been used to study the sizing mechanism of AKD.[95] Results from the study showed that the addition of a cationic polymer to the pulp suspension was effective at promoting sizing and there was a good correlation between the amount of AKD in the paper and its degree of hydrophobicity.

5.7 Analysis of Wet-strength Resin

Ascertaining the structure and mechanism of polyamide–epichlorohydrin (PAE) resins for use as wet-strength additives in paper is important but is difficult to do reliably because the additives are present in papers in very small amounts. Py/GC has been used to measure reliably the compounds in paper. Cyclopentanone derived

from the adipic acid component of PAE was the key product for measuring the polymer in paper. Analyses were reported to be reproducible within a 4.8% relative standard deviation.[96]

5.8 Analysis of Sulfur Compounds

Although the kraft pulping process that entails cooking wood chips in sulfide liquors is the dominant technology in the pulp and paper industry, very little is known about the fate of the sulfur component in the process (e.g. the chemical nature of sulfur in lignocellulosics). This is probably due to a lack of suitable analytical techniques for sulfur compounds. The analysis of sulfur compounds can be problematic. For example, analysis of polysulfide is hampered by the rapid deterioration of the sulfur in the samples to be analyzed. A new method for characterization of sulfur in wood and chemical pulps has been developed that employs fractionated pyrolysis.[97] In fractionated pyrolysis, a sample is heated at progressively increasing temperatures for different times to enable study of pyrolysis products of the sample at the different temperatures. The influence of different pyrolysis conditions and sample size was studied on sulfur-treated wood residues and sulfur-containing model compounds. The pyrolysis was performed using a resistively heated pyrolyzer linked to a gas chromatograph with a flame photometric detector. The fractionated pyrolysis was conducted in two ways. The first was sequential pyrolysis as follows: three times at 300 °C, three times at 400 °C, twice at 600 °C, and once at 1350 °C, each time using a pyrolysis time of 2 s. The second was pyrolysis

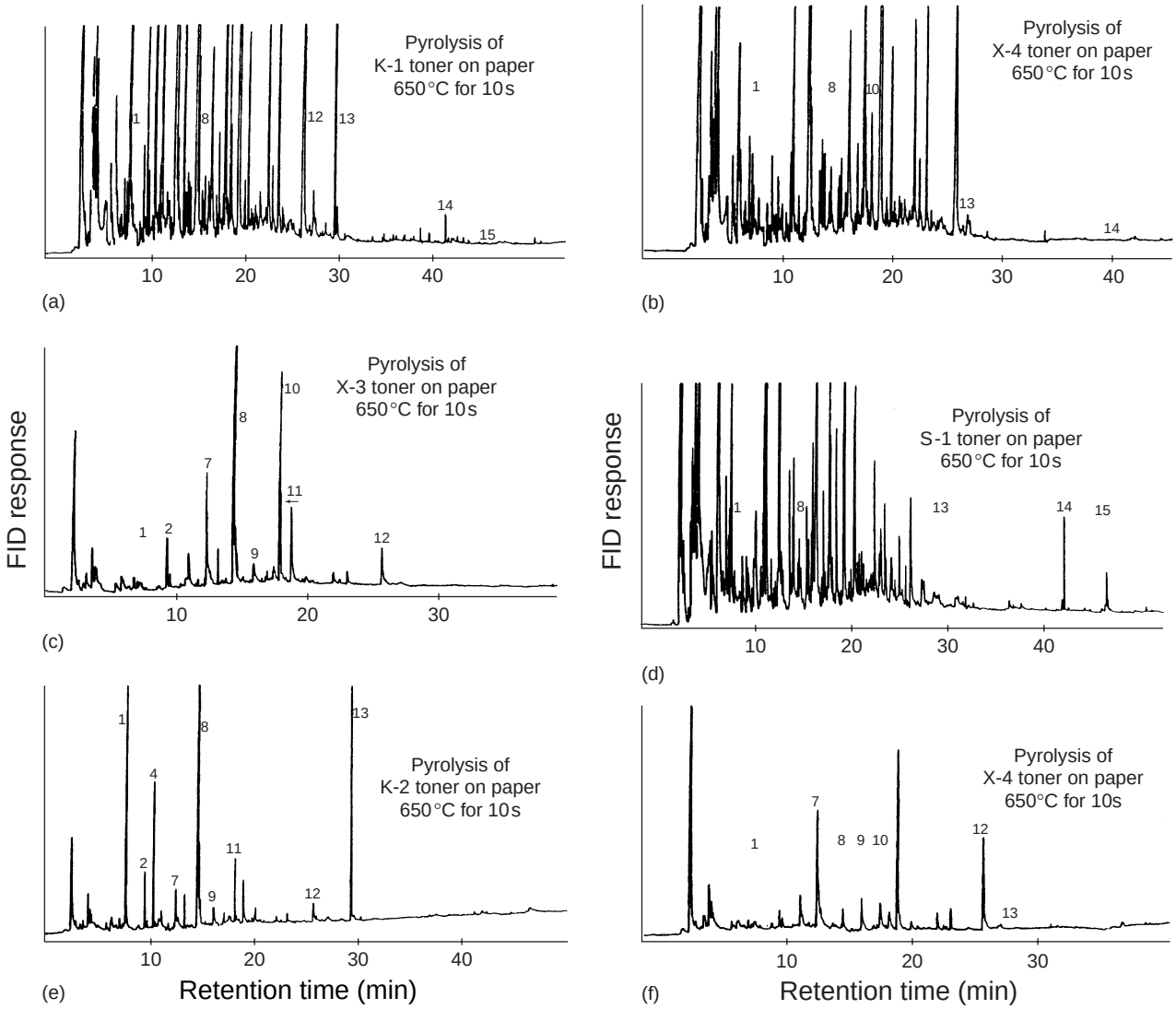

Figure 24 Py/GC can be used to differentiate toners used in different makes of photocopier machines. (Copyright ASTM. Reprinted with permission.)

once at each temperature for 2 s. The main pyrolysis products that were identified were H_2S, COS, SO_2, CH_3SH, and CS_2, that is, they were inorganic except for methanethiol (CH_3SH). The study seems to indicate that the majority of the sulfur in wood is inorganic, and is not associated with wood protein as had been reported previously.[98]

5.9 Analysis of Ink

Printing inks are difficult to analyze because they are complex formulations of solid, volatile, and semi-volatile materials such as pigments, petroleum resins, waxes, and oils. Analytical pyrolysis may be used to characterize the solid components but the presence of volatile materials may interfere. Figure 22(a–c) show pyrograms obtained from analysis of black ink on paper, the ink by itself, and volatile components present in the ink formulation.[99] The pyrograms show that both volatile and non-volatile components of the ink are present in the ink on the paper sample.

Wampler and Levy[99] developed a two-step procedure to simplify the analysis of ink formulations as illustrated in Figure 23(a–c). Figure 23(a) shows analysis of an ink formulation by Py/GC. It is evident that the pyrogram is complex and has a mass of unresolved peaks that elute between 20 and 26 min. If the sample is first heated at 250 °C for 5 min in the probe of a resistively heated pyrolyzer, volatile hydrocarbon oils, which make up a large portion of the ink, are removed, leaving behind solid resins and waxes. The removed volatiles can then be chromatographed to give a pyrogram as shown in Figure 23(b). This pyrogram is akin to dynamic headspace analysis of the ink.

Pyrolysis of the remaining solids at 700 °C for 10 s and chromatography of the pyrolysis products result in a pyrogram, shown in Figure 23(c), that has near baseline separation of major peaks and also has specific peaks that may be identified and associated with individual solid constituents of the ink formulation. For example, peaks eluting between 12 and 19 min are due to hydrocarbon resins, whereas those eluting between 20 and 25 min are due to normal alkanes and alkenes degraded from the wax.

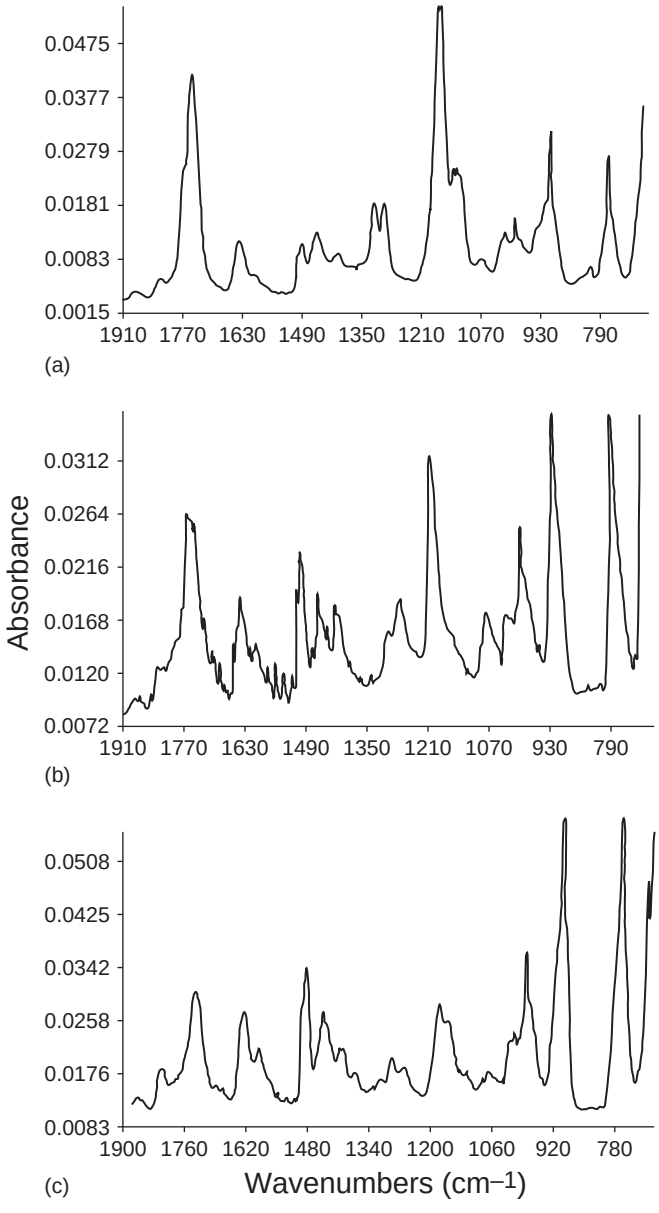

Figure 25 IR spectroscopy can be used for detection in pyrolysis. (a) Xerox 9500; (b) Canon NP/15; (c) Mita DC 131. (Copyright CDS Analytical Inc. Reprinted with permission.)

5.10 Analysis of Photocopier Toners

Analytical pyrolysis has also been used to characterize toner materials from photocopiers. Toners consist of a colorant such as carbon black or iron oxide embedded in an organic polymeric matrix. The polymer formulations used vary from one photocopy manufacturer to another and these may be easily differentiated using Py/GC and Py/GC/MS.[100] Analysis of nine photocopy samples representing four manufacturers revealed that, within a manufacturer, there was little difference seen in the toner formulations used for different models. The pyrograms ranged from complex to relatively simple, as illustrated in Figure 24(a–f). Identification of the major peaks indicate that materials used in the photocopying process include methyl methacrylate, toluene, glycidol, furfural, styrene, butyl methacrylate, dodecyl methacrylate and tetradecyl methacrylate.

Photocopier toners have also been characterized by Py/FTIR spectroscopy. The colorant materials that absorb IR radiation remain when the sample is pyrolyzed.

The volatilized products are easily detected by FTIR spectroscopy. For example, analysis of toners from three manufacturers shows that, although they are all based on polystyrene, the FTIR spectra of their pyrolysis products are different.[20] As can be seen in Figure 25(a–c), the three toners can be easily distinguished based on the ratios of the absorbance at absorption bands that match the pattern formed by styrene (major peaks at 774 and 910 cm^{-1}) and those of butyl methacrylate (major peak at 1165 cm^{-1}).

6 PRETREATMENT OF SAMPLES BEFORE ANALYTICAL PYROLYSIS

The pretreatment of samples before analysis can affect the resulting chromatograms and can lead to the incorrect identification of the analyzed polymers. For example, pyrolysis of PVA samples yields different pyrogram patterns depending on whether or not the PVA emulsion was dried at high temperature before analysis.[85] The pyrograms shown in Figure 26 illustrate this. In this case a match with the deposit sample was obtained after the emulsion had been dried before pyrolysis. This implies that standards should be subjected to temperatures similar to those experienced by the contaminants in the deposit samples in order to obtain meaningful chromatographic patterns in the pyrograms.

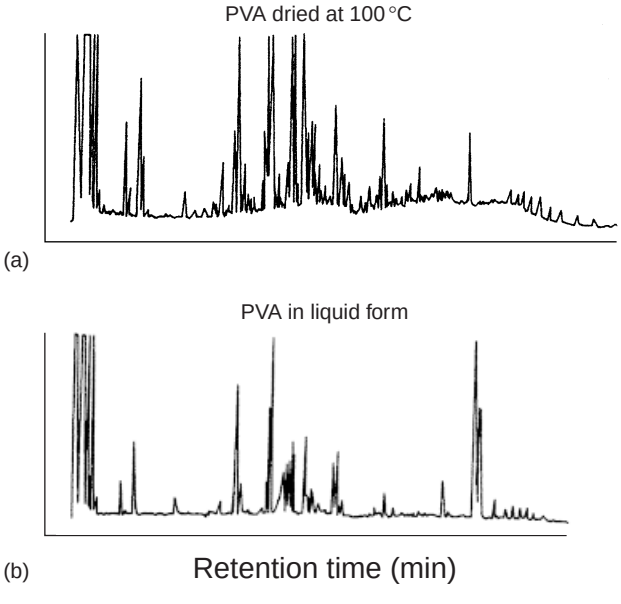

Figure 26 The way in which a sample is handled before pyrolysis can affect the pattern of the pyrolysis pyrogram. (First published in ref. 85 and reproduced by kind permission of the Journal.)

Representative sampling can be a problem when analyzing samples in the microgram range, such as dirt spots on paper. Consider a hypothetical case of two plastic particles in a dirt spot on a sheet of paper. On sampling a few micrograms of the deposit it is very likely that these particles will be missed since there are many more nonplastic components in the deposit (such as wood resin, calcium carbonate, talc, and clay) than the plastic components. Analysis of the whole sample is one way to overcome this dilemma. However, this is not always feasible because the sample may be too large for analysis (in analytical pyrolysis sample sizes of a few micrograms are desirable). Another way is to ensure that the sample is homogeneous before sampling. A sample homogenization procedure has been developed that entails refluxing the sample in a suitable solvent.[85] The solvent dissolves and disperses the polymer particles plus any other organic matter and distributes them evenly throughout the sample matrix. On evaporating the solvent, a film surface remains on top of the sample, portions of which can then be sampled for analysis.

7 SLOW PYROLYSIS OF WOOD SAMPLES

As mentioned in the Introduction, analytical pyrolysis is normally done under rapid heating conditions to prevent condensation of the pyrolysis products. We have seen that such pyrolysis can induce reproducible degradation of wood extractives, lignins, and carbohydrate-type compounds. In many cases the degradation patterns can be used to fingerprint the parent compounds. It is interesting, however, to see how slow pyrolysis affects the degradation of lignocellulosic materials. A completely different kind of behavior is observed. For example, DeGroot et al.[101] studied slow pyrolysis products of wood. The wood was heated at 250 °C on a thermal balance and the evolved products were monitored by FTIR spectroscopy, which showed that about 60% of the weight loss was due to the formation of water, carbon dioxide, methanol, acetic acid, and formic acid. The heated wood was analyzed for glycoses and uronic acids and the data led the authors to determine the events that occur when wood is pyrolyzed. They found that the first chemical events that occur in the pyrolysis of wood include ready decomposition of uronic acids in the hemicelluloses and pectic substances to form carbon dioxide, water, char (or char precursors), and some methanol. The decomposition may lead to further pyrolysis of the xylose units to which the uronic acids are attached in the hemicelluloses. Acetyl ester groups in the hemicelluloses are much more resistant to pyrolysis, but are released slowly as acetic acid. Formic acid

is released at a slow but steady rate by unknown mechanisms and is probably derived from degradation of hemicellulose.

A report by Ingemarsson et al.[102] discusses slow pyrolysis of wood samples from several species of European spruce and pine trees. The samples were subjected to slow pyrolysis in a reactor and the products analyzed by GC/MS, GC/FTIR/FID, and by trapping on solid-phase extraction cartridges for later analysis by GC/MS. A schematic of the pyrolyzer, a variant of a furnace pyrolyzer, is shown in Figure 27. The pyrolyzer can be heated up to 900 °C but in this case it was maintained at 550 °C with a nitrogen gas flow rate of 40 mL min^{-1}.

It is interesting to note, from the data, that the pyrolysis of pine and spruce samples produced substantial amounts of carbon monoxide and carbon dioxide even though the pyrolysis was done in the absence of oxygen. Also, substantial amounts of gaseous oxygenated compounds were produced, such as methanol, acetaldehyde, acrolein, propanal, acetone, acetic acid, butyric acid, ethyl formate, and phenol. The nongaseous compounds detected by GC/MS for the different wood species studied are shown in Table 7.

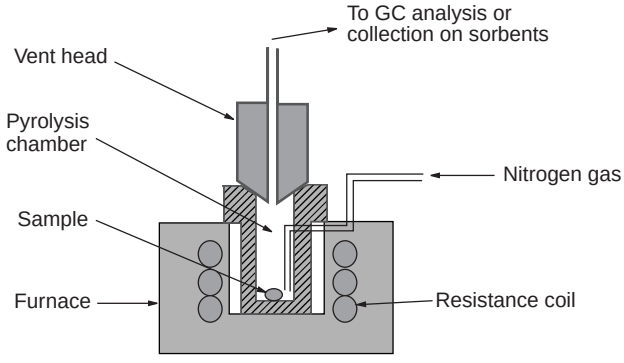

Figure 27 Schematic of a pyrolyzer for conducting slow pyrolysis studies.

The data show that the emissions of organic compounds from pyrolysis of spruce and pine wood samples are very complex but there are significant differences between spruce and pine species. Unfortunately, the patterns of

Table 7 Compounds evolved during slow pyrolysis of four different pine and four different spruce wood species. (Adapted from Ingemarsson et al[102])

Compound	Molecular formula	Molecular weight	Pines[a]				Spruces[b]			
			S	P	L	A	N	W	S	B
Acetaldehyde	C_2H_4O	44	✓	✓	✓	✓	✓	✓	✓	✓
2-Propanone	C_3H_6O	58	✓	✓	✓	✓	✓	✓	✓	✓
Furan	C_4H_4O	68			✓		✓			✓
2,3-Butanedione	$C_5H_{10}O$	86	✓	✓	✓	✓	✓	✓		✓
2-Butanone	C_4H_8O	72	✓		✓		✓		✓	
2-Methylfuran	C_5H_6O	82	✓	✓	✓	✓	✓		✓	✓
Ethyl formate	$C_4H_8O_2$	88	✓	✓		✓		✓		
Benzene	C_6H_6	78			✓					✓
1-Hydroxy-2-propanone	$C_3H_6O_2$	74	✓	✓		✓		✓		
1,1-Diethoxyethane	$C_6H_{14}O_2$	118	✓	✓		✓				
Toluene	C_7H_8	92		✓	✓	✓	✓	✓	✓	✓
2-Furancarboxaldehyde	$C_5H_4O_2$	96		✓		✓		✓		
Ethylbenzene	C_8H_{10}	106	✓	✓		✓				
β-Pinene	$C_{10}H_{16}$	136		✓	✓			✓		
1,5,5,6-Tetramethyl-1,3-cyclohexadiene	$C_{10}H_{16}$	136				✓	✓	✓		✓
β-Myrcene	$C_{10}H_{16}$	136	✓	✓	✓	✓		✓	✓	
δ-3-Carene	$C_{10}H_{16}$	136	✓	✓	✓		✓	✓		✓
1-Methyl-2-(1-Methylethyl)benzene	$C_{10}H_{16}$	134	✓			✓				
DL-Limonene	$C_{10}H_{16}$	136	✓	✓	✓	✓	✓	✓	✓	✓
2-Methoxyphenol	$C_7H_8O_2$	124		✓	✓	✓		✓	✓	
Alloocimene	$C_{10}H_{16}$	136				✓		✓		
2-Methoxy-4-methylphenol	$C_8H_{10}O_2$	138	✓	✓		✓		✓	✓	✓
4-Ethyl-2-methoxyphenol	$C_9H_{12}O_2$	152		✓		✓		✓	✓	
2-Methoxy-6-vinylphenol	$C_9H_{10}O_2$	150	✓	✓	✓	✓	✓	✓	✓	✓
2-Methoxy-4-(propenyl)phenol[c]	$C_{10}H_{12}O_2$	164				✓	✓	✓	✓	✓
2-Methoxy-4-(propenyl)phenol[c]	$C_{10}H_{12}O_2$	164				✓			✓	✓
2-Methoxy-4-(propenyl)phenol[c]	$C_{10}H_{12}O_2$	164	✓			✓				✓
2-Methoxy-4-(propenyl)phenol[c]	$C_{10}H_{12}O_2$	164	✓	✓	✓	✓	✓	✓		✓

[a] Pines: S = Scotch pine; A = Austrian pine; L = Lodgepole pine; P = Ponderosa pine.
[b] Spruces: N = Norway spruce; W = White spruce; B = Black spruce; S = Sitka spruce.
[c] Isomers similar to 2-methoxy-4-(propenyl)phenol. A more complete identification could not be made using the MS Library.

the pyrolysis products cannot be used to fingerprint the species. The authors conclude that the emission of large amounts of toxic aldehydes and ketones illustrate that incomplete combustion of wood biomass can lead to environmental health problems.

8 SIMULTANEOUS PYROLYSIS ALKYLATION OF SYNTHETIC POLYMERS

The following are examples of how the analysis of synthetic polymers can be characterized by analytical pyrolysis using simultaneous alkylation procedures. This information is useful in understanding the sources of contaminants and deposit problems that occur in mills that make and/or use recycled paper furnishes.

8.1 Alkyd Resins

Alkyd resins are made from polyfunctional acids, polyhydric alcohols, and long-chain fatty acids from vegetable oils. Conventional Py/GC/MS of the compounds shows aldehydes, alkanes, alkenes, and phthalic anhydride.[29] These compounds only reflect the composition of the polymer indirectly because the aldehydes, alkenes, and alkanes are low-molecular-weight fragments that have little structural similarity to the long-chain fatty acids and polyols. Pyrolysis alkylation, however, results in hydrolysis and methylation of the polar components, yielding methyl esters of the polybasic acid and long-chain fatty acids and methyl ethers of the polyhydric alcohol as shown in Figure 28(a) and (b).[29] The formation of dimethyl *o*-phthalate, methyl ethers of pentaerythritol, and the fatty acid methyl esters shows that the resin is a pentaerythritol *o*-phthalic alkyd with a fatty acid composition similar to that of soya-bean oil.

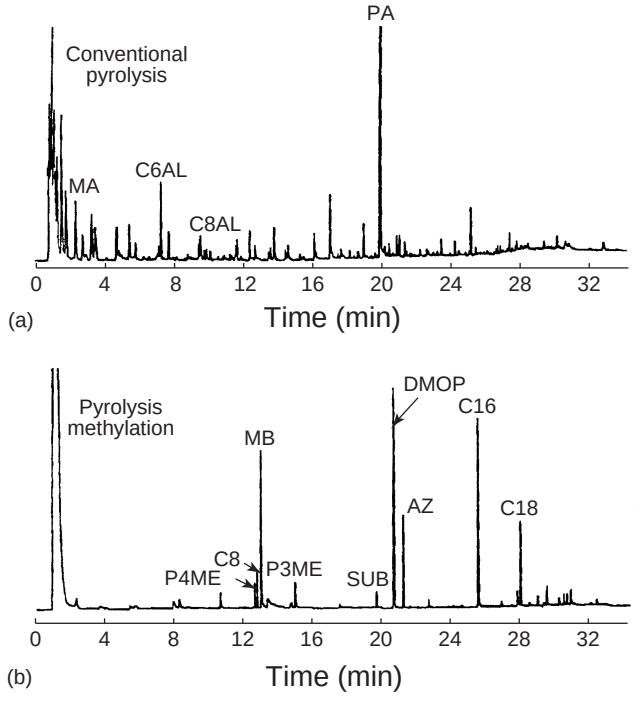

Figure 28 Pyrolysis alkylation of an alkyd resin polymer. (Reproduced from ref. 29, with permission from Elsevier Science.)

8.2 Unsaturated Polyester Resins

Styrene unsaturated polyester resins are produced by esterification of *o*- or isophthalic acids, maleic anhydride, and polyhydric alcohols (propylene glycol and neopentyl glycol), and can be coreacted with modifying agents such as adipic acid. The styrene is added in the formulation to facilitate cross-linking.

The main products of conventional pyrolysis of one such polymer, a diethylene glycol–isophthalic–adipic acid-modified polyester, include toluene, styrene, styrene oligomers, and cyclopentanone. Pyrolysis methylation of the same polymer gives dimethyl isophthalate, methyl ethers of diethylene glycol, dimethyl adipate, and methyl benzoate (Figure 29a and b). The polybasic acid in the polymer is indicated by the dimethyl phthalate isomer,

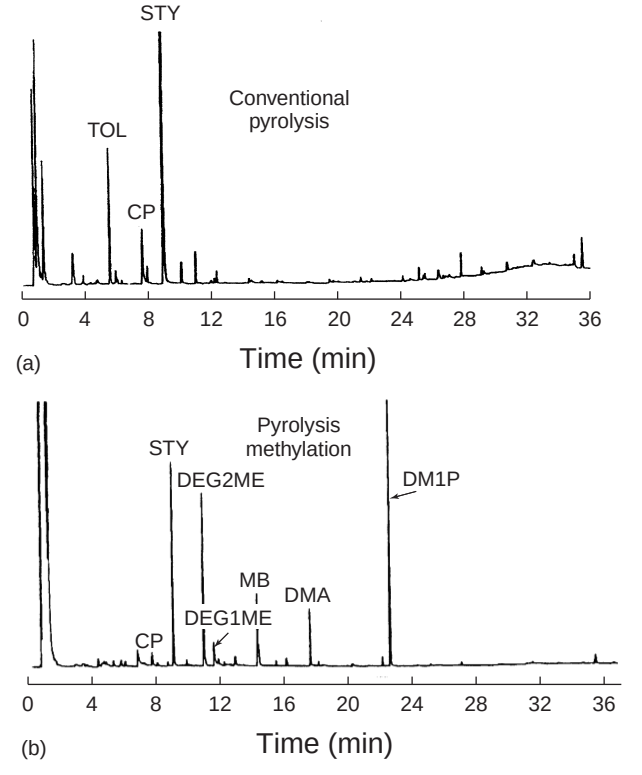

Figure 29 Pyrolysis alkylation of an unsaturated polyester resin polymer. (Reproduced from ref. 29, with permission from Elsevier Science.)

the polyol by the corresponding methyl ethers, and the adipic acid by its dimethyl ester.[29]

8.3 Polyester Fibers

Polyester fibers are made mainly from poly(ethylene terephthalate) (PET) and used for making fabrics used in pulp and papermaking processes. Sugimura and Tsuge[42] have shown that their main conventional pyrolysis products include benzene, benzoic acid, and biphenyl. Pyrolysis methylation causes esterification of the terephthalic acid to dimethyl terephthalate and benzoic acid to methyl benzoate. Figure 30(a) and (b) illustrate the enhanced sensitivity of the methylation procedure. It is evident that pyrolysis methylation of a 1-mm fiber yields much sharper pyrograms than conventional pyrolysis of a 40-mm length of fiber.

8.4 Phenolic Polymers

Phenolic polymers that are commonly found include epoxy, polycarbonate, and phenol–formaldehyde resins. Conventional pyrolysis of epoxy resins derived from phenol A and epichlorohydrin results in mainly substituted phenolic compounds,[43] those from polycarbonates produce similar phenolic products and diphenyl carbonate,[103] whereas phenol–formaldehyde resin yields substituted phenols whose structure is dependent on the macromolecular structure of the resin.[104] Pyrolysis methylation results in the formation of methyl ethers of phenol and substituted phenolic products and gives more detailed information on the composition of the polymers, as illustrated in Figure 31(a) and (b) for an epoxy resin.[29]

8.5 Poly(vinyl acetate)

Vinyl acetate is usually copolymerized with other vinyl monomers to make surface coatings and adhesives. Conventional pyrolysis of PVA forms aromatic compounds

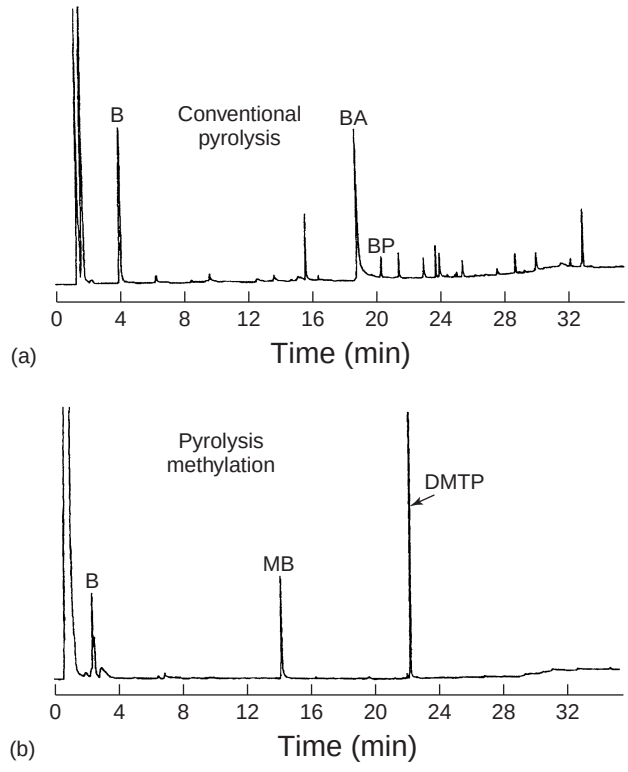

Figure 30 Pyrolysis alkylation of polyester fibers. (Adapted from Challinor.[29] Reproduced from ref. 29, with permission from Elsevier Science.)

and acetic acid and the latter is used as a diagnostic marker for the presence of vinyl acetate in the polymer. However, the polar nature of acetic acid results in very poor chromatography of the compound. Pyrolysis alkylation with a butylating agent improves the chromatography and sensitivity of polymers that contain vinyl acetate. A typical pyrogram for a PVA adhesive pyrolyzed in the presence of TBAH is shown in Figure 32(a–c). A disadvantage of pyrolysis butylation with TBAH is the presence of by-products, *n*-butanol and tributylamine,

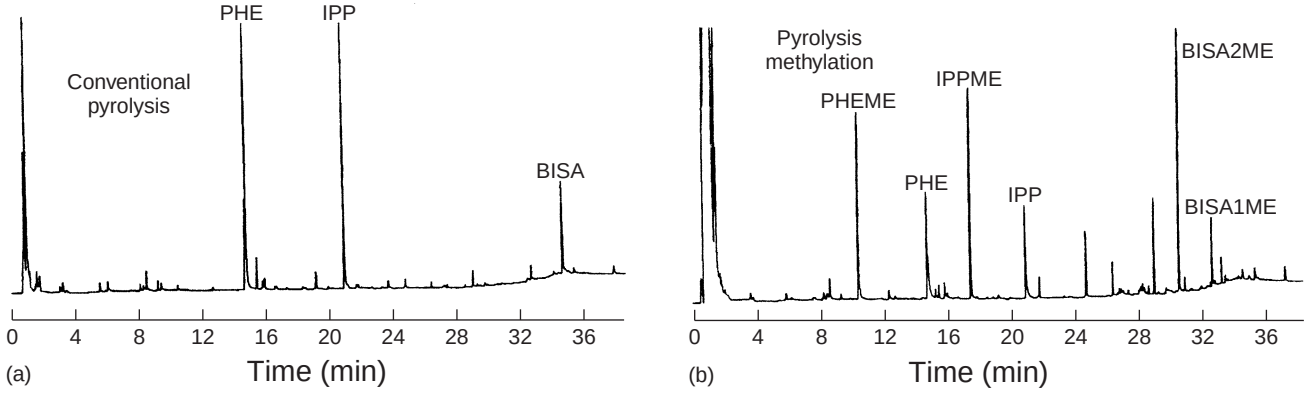

Figure 31 Pyrolysis alkylation of phenolic polymers. (Reproduced from ref. 29, with permission from Elsevier Science.)

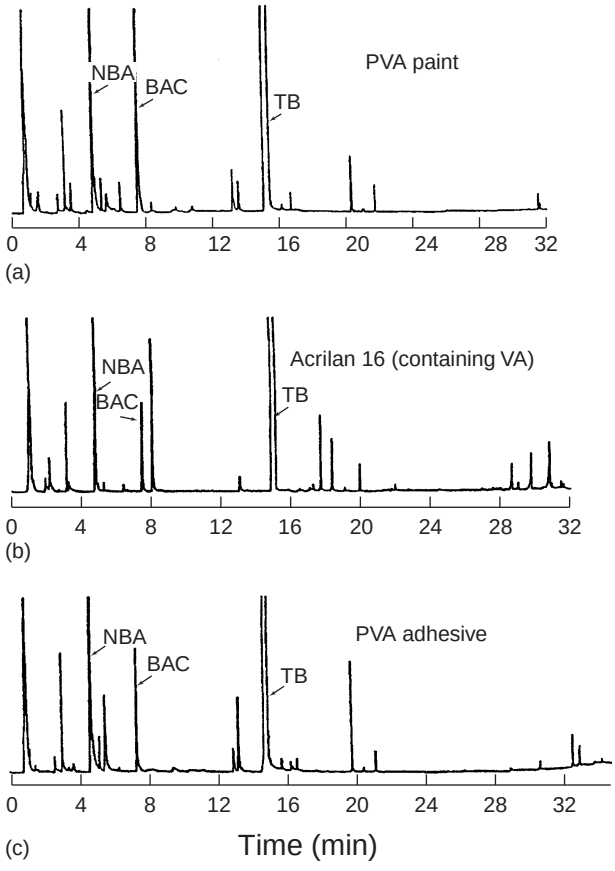

Figure 32 Pyrolysis alkylation of PVA. (Reproduced from ref. 29, with permission from Elsevier Science.)

in the pyrogram.[29] Fortunately, the compounds generally do not interfere with the interpretation of the pyrogram.

8.6 Methacrylic Acid Copolymers

Methacrylic acid is copolymerized with acrylic acid in high-quality coatings and adhesives. The acid, produced in conventional pyrolysis of this copolymer, exhibits poor chromatographic properties that are significantly improved by pyrolysis butylation of the polymer due to the formation of butyl methacrylate.[29] The formation of the methacrylate confirms the presence of the acid in the polymer, as illustrated in Figure 33.

8.7 Cellulose Acetate Butyrate

Pyrolysis products of cellulose acetate butyrate, a flow-promoter additive in acrylic lacquer formulations, are not detected by conventional Py/GC. However, pyrolysis butylation results in the detection of butyl esters of acetic and butyric acids, as shown in Figure 34.[29]

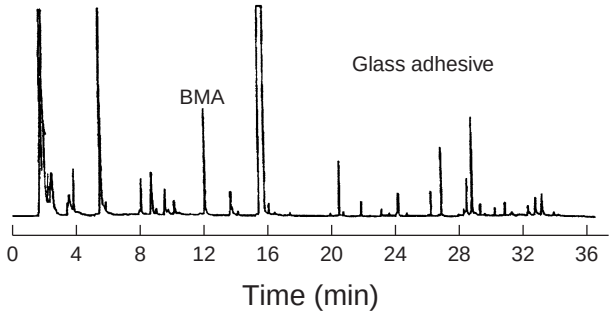

Figure 33 Pyrolysis alkylation of methacrylic acid copolymers. (Reproduced from ref. 29, with permission from Elsevier Science.)

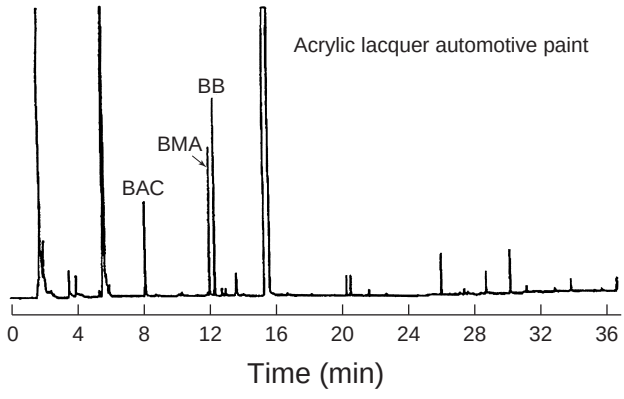

Figure 34 Pyrolysis alkylation of cellulose acetate butyrate. (Reproduced from ref. 29, with permission from Elsevier Science.)

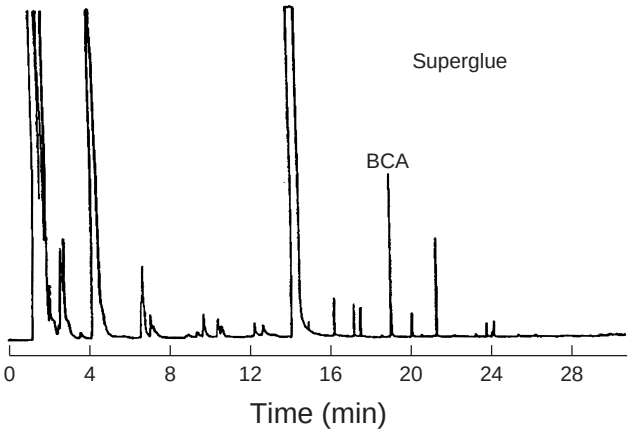

Figure 35 Pyrolysis alkylation of polycyanoacrylates. (Reproduced from ref. 29, with permission from Elsevier Science.)

8.8 Polycyanoacrylates

Polycyanoacrylates are used to make high-performance adhesives. Their pyrolysis products are polar and therefore difficult to separate by chromatography.

Pyrolysis butylation yields a number of compounds, the major one being butyl cyanoacrylate (Figure 35).

ABBREVIATIONS AND ACRONYMS

AKD	Alkyl Ketene Dimer
ASA	Alkenyl Succinic Anhydride
BKME	Bleached Kraft Mill Effluents
COD	Chemical Oxygen Demand
CTMP	Chemithermomechanical Pulp
FA	Fulvic Acid
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/MS	Gas Chromatography/ Mass Spectrometry
HTML	Hypertext Markup Language
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
MS	Mass Spectrometry
MWL	Milled Wood Lignin
NIR	Near-infrared
PAE	Polyamide–Epichlorohydrin
PCA	Principal Component Analysis
PET	Poly(ethylene terephthalate)
PLS	Partial Least Squares
PVA	Poly(vinyl Acetate)
Py/FTIR	Pyrolysis/Fourier Transform Infrared
Py/GC	Pyrolysis/Gas Chromatography
Py/GC/IR	Pyrolysis/Gas Chromatography/ Infrared Spectroscopy
Py/GC/MS	Pyrolysis/Gas Chromatography/ Mass Spectrometry
Py/IR	Pyrolysis/Infrared Spectroscopy
Py/MS	Pyrolysis/Mass Spectrometry
THM	Thermally Assisted Hydrolysis or Methylation
THT	Total Heating Time
TMAAc	Tetramethylammonium Acetate
TMAH	Tetramethylammonium Hydroxide
TRT	Temperature Rise Time
TTP	Temperature–Time Profile
TVA	Thermal Volatilization Analysis

RELATED ARTICLES

Food **(Volume 5)**
Lipid Analyses in Food

Polymers and Rubbers **(Volume 9)**
Pyrolysis Techniques in the Analysis of Polymers and Rubbers

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices

Chemometrics **(Volume 11)**
Clustering and Classification of Analytical Data

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses

APPENDIX 1: PYROLYSIS NOMENCLATURE

Pyrolysis	A chemical degradation reaction that is induced by thermal energy (alone). (The term ''pyrolysis'' generally refers to an inert environment.)
Oxidative pyrolysis	A pyrolysis that occurs (only) in the presence of an oxidative atmosphere.
Catalytic pyrolysis	A pyrolysis that is influenced by the addition of a catalyst.
Pyrolyzer	A device for performing pyrolysis.
Pyrolysis reactor	That portion of the pyrolyzer in which the pyrolysis takes place.
Pyrolyzate	The products of pyrolysis.
Volatile pyrolyzate	That portion of the pyrolyzate that has appreciable vapor pressure under the conditions of the pyrolysis.
Pyrolysis residue	That portion of the pyrolyzate that does not leave the reactor.
Char	A solid carbonaceous residue.
Tar	A liquid pyrolysis residue.
Analytical pyrolysis	The characterization of a material or a chemical process by the instrumental analysis of its pyrolyzate.

Applied pyrolysis	The production of commercially useful materials by means of pyrolysis.
Isothermal pyrolysis	A pyrolysis during which the temperature is essentially constant.
Nonisothermal pyrolysis	A pyrolysis during which the temperature varies significantly.
Temperature–time profile (TTP)	A graphical representation of temperature versus time for a particular pyrolysis experiment or pyrolyzer.
Total heating time (THT)	The time between the onset and conclusion of sample heating in a pyrolysis experiment.
Final temperature (T_f)	The final (steady-state) temperature which is attained by a pyrolyzer. (The terms "equilibrium temperature" and "pyrolysis temperature" may be used when referring to an isothermal pyrolysis; they are not recommended for use with a nonisothermal pyrolysis, however.)
Maximum temperature (T_a)	The highest temperature in a TTP.
TRT	The time required for a pyrolyzer to go from its initial temperature to the final temperature.
Flash pyrolysis	A pyrolysis that is carried out with a fast TRT (to a constant final temperature).
Continuous mode (furnace) pyrolyzer	A pyrolyzer in which the sample is introduced into a pyrolyzer preheated to the final temperature.
Pulse mode pyrolyzer	A pyrolyzer in which the sample is introduced into a cold pyrolyzer which is then heated rapidly.
Filament (ribbon) pyrolyzer	A pyrolyzer in which the sample is placed on a metal filament (ribbon) that is resistively or inductively heated to induce pyrolysis.
Coil pyrolyzer	A pyrolyzer in which the sample (sometimes placed in a tubular vessel) is placed in a metal coil that is heated to induce pyrolysis.
Curie point pyrolyzer	A pyrolyzer in which a ferromagnetic sample carrier is inductively heated to its Curie point.
Temperature-programmed pyrolysis	A pyrolysis in which the sample is heated at a controlled rate over a range of temperatures during which pyrolysis occurs.
Stepwise pyrolysis	A pyrolysis in which the sample temperature is raised stepwise, i.e. pyrolysis is carried out at a selected temperature for a selected period of time, then the pyrolyzer is stepped rapidly to a higher temperature.
Sequential pyrolysis	A pyrolysis in which the same sample is repetitively pyrolyzed under identical conditions (T_f, TRT, and THT) until no further products are detected.
Fractionated pyrolysis	A pyrolysis in which the sample is pyrolyzed at different temperatures for different times in order to study special fractions of the sample.
Pyrogram	A chromatogram or spectrum of a pyrolyzate.
Py/GC	A pyrolysis technique in which the volatile pyrolyzates are separated and detected by GC.
Py/MS	A pyrolysis technique in which the volatile pyrolyzates are analyzed by MS.
Py/GC/MS	A pyrolysis technique in which the volatile pyrolyzates are analyzed by on-line GC/MS.
Pyrolysis/infrared spectroscopy (Py/IR)	A pyrolysis technique in which the pyrolyzates are analyzed by IR spectroscopy.
Pyrolysis/gas chromatography/infrared spectroscopy (Py/GC/IR)	A pyrolysis technique in which the volatile pyrolyzates are analyzed by on-line GC/IR spectroscopy.
Thermal volatilization analysis (TVA)	A pyrolysis technique in which the pressure of the volatile pyrolyzates is recorded as the sample is heated.

REFERENCES

1. W.I. Irwin, *Analytical Pyrolysis: A Comprehensive Guide*, Marcel Dekker, Inc., New York, 1982.
2. T.P. Wampler, 'A Selected Bibliography of Analytical Applications 1980–1989', *J. Anal. Appl. Pyrol.*, **16**(4), 291–322 (1989).
3. S.A. Liebman, E.J. Levy, *Pyrolysis and GC in Polymer Analysis*, Marcel Dekker, Inc., New York, 1985.
4. T.P. Wampler (ed.), *Applied Pyrolysis Handbook*, Marcel Dekker, Inc., New York, 1995.
5. J.K. Haken, 'Pyrolysis Gas Chromatography of Synthetic Polymers – A Bibliography', *J. Chromatogr. A*, **825**, 171–187 (1998).
6. A. Mitchell, M. Needleman, 'Technique for the Prevention of Column Contamination in Pyrolysis Gas–Liquid Chromatography', *Anal. Chem.*, **50**(4), 668–669 (1978).
7. Y. Sugimura, S. Tsuge, 'Fundamental Splitting Conditions for Pyrogram Measurements with Glass Capillary Gas Chromatography', *Anal. Chem.*, **50**(14), 1968–1972 (1978).
8. J.M. Halkett, H.R. Schulten, 'Simple Pyrolysis Chamber Modification for Capillary Column Curie Point Pyrolysis Gas Chromatography–Mass Spectrometry of Complex Biomaterials', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **9**, 596–597 (1986).
9. I. Ericsson, R.P. Lattimer, 'Pyrolysis Nomenclature', *J. Anal. Appl. Pyrol.*, **14**(2), 219–221 (1989).
10. P.C. Uden, 'Nomenclature and Terminology for Analytical Pyrolysis', *Pure Appl. Chem.*, **65**(11), 2405–2409 (1993).
11. H. Yokoi, Y. Ishida, H. Ohtani, S. Tsuge, T. Sonoda, T. Ona, 'Characterization of Within-tree Variation in *Eucalyptus camaldulensis* by Pyrolysis–Gas Chromatography', *The Analyst*, **124**(5), 669–674.
12. N. Oguri, A. Onishi, 'Curie Point Pyrolysis with a Modified Maintenance Oven for Capillary Column GC', *LC/GC*, **12**(7), 541–545 (1994).
13. E.M. Anderson, I. Ericsson, 'Thermal Degradation of Organic Polymers Using Different Metals as the Pyrolysis Filament', *J. Anal. Appl. Pyrol.*, **3**(1), 35–47 (1981).
14. I. Tydén-Ericsson, 'A New Pyrolyzer with Improved Control of Pyrolysis Conditions', *Chromatographia*, **6**(8–9), 353–358 (1973).
15. M. Kleen, *Characterization of Wood and Pulp Using Analytical Pyrolysis and Multivariate Data Analysis*, Thesis, Royal Institute of Technology, Stockholm, Sweden, 1993.
16. Bulletin #397, Extend your GC's capabilities. Analyze insoluble or high molecular weight materials with Pyrola® 2000, Alltech Associates, Deerfield, IL, 1998.
17. D.L. Harms, 'Identification of Complex Materials', *Anal. Chem.*, **25**, 1140–1155 (1953).
18. J.O. Lephardt, 'Fourier Transform Infrared Evolved Gas Analysis (FT/IR/EGA)', *Appl. Spectrosc. Rev.*, **18**(2), 265–303 (1982–1983).
19. J.W. Washall, T.P. Wampler, 'Direct Pyrolysis FT-IR Spectroscopy for Polymer Analysis', *Spectrosc.*, **6**(4), 38–42 (1991).
20. M.J. Matheson, *The Analysis of Photocopier Toners by Pyrolysis/FT-IR*, CDSolutions, CDS Analytical Inc., Oxford, PA, USA, Pamphlet, #493.
21. H.L.C. Meuzelaar, P.G. Kistemaker, 'Technique for Fast and Reproducible Fingerprinting of Bacteria by Pyrolysis Mass Spectrometry', *Anal. Chem.*, **45**(3), 587–590 (1973).
22. H.L.C. Meuzelaar, P.G. Kistemaker, W. Eshuis, A.J.H. Boerboom, in *Advances in Mass Spectrometry*, ed. N.R. Daley, Heyden, London, 1452–1456, 1978.
23. R.E. Aries, C.S. Gutteridge, T.W. Ottley, 'Evaluation of a Low-cost Automated Pyrolysis–Mass Spectrometer', *J. Anal. Appl. Pyrol.*, **9**(2), 81–98 (1986).
24. J. Maddock, T.W. Ottley, 'Pyrolysis Mass Spectrometry: Instrumentation, Techniques, and Applications', in *Applied Pyrolysis Handbook*, ed. T.P. Wampler, Marcel Dekker, New York, Chapter 3, 1993.
25. G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York, 1992.
26. D. Meier, J. Moltran, O. Faix, 'Discrimination of Wood Species from Various Provenances Using Pyrolysis–Gas Chromatography, Pyrolysis–Mass Spectrometry, Near Infrared Spectroscopy and Multivariate Data Analysis', in *Proceedings, 8th International Symposium on Wood Pulp Chemistry*, Gummerus Oy, Helsinki, 283–288, 1995.
27. C. David, 'Thermal Degradation of Polymers', in *Comprehensive Chemical Kinetics*, eds. C.H. Bamford, C.F.H. Tipper, Elsevier, New York, Vol. 14, 1975.
28. N. Dunlop-Jones, L.H. Allen, 'Rapid Method for the Qualitative Analysis of Plastic and Sticky Contaminants by Pyrolysis Gas Chromatography', *Tappi J.*, **71**(2), 109–113 (1988).
29. J.M. Challinor, 'A Pyrolysis–Derivatisation–Gas Chromatography Technique for the Structural Elucidation of some Synthetic Polymers', *J. Anal. Appl. Pyrol.*, **16**(4), 323–333 (1989).
30. T.D. Crockett, A.A. Webb, L.G. Borchardt, D.B. Easty, 'Identification of Polymers in Paper by Pyrolysis Gas Chromatography. Reduction of Paper Interference by Pre-pyrolysis', *J. Chromatogr.*, **407**, 330–339 (1987).
31. B.B. Sitholé, B. Ambayec, 'Identification and Determination of Surface Sizing Agents on Sized Papers', *Proceedings, 85th Annual Meeting, Pulp and Paper Technical Association of Canada, Montreal*, Vol. A, A95–A99, 1999.
32. H.-L. Hardell, 'Characterization of Impurities in Pulp and Paper Products Using Pyrolysis–Gas Chromatography–Mass Spectrometry', *J. Anal. Appl. Pyrol.*, **27**(1), 73–85 (1993).

33. F.J. González-Vila, A. Gutiérrez, F. Martin, T. Verdejo, 'Application of Analytical Pyrolysis to the Characterization of Eucalyptus Extractives and Pitch Deposits from a Pulp Mill', *J. Anal. Appl. Pyrol.*, **40–41**, 501–510 (1997).
34. D. Garozzo, G. Montaudo, 'Identification of Polymers by Library Search of Pyrolysis Mass Spectra and Pattern Recognition Analysis', *J. Anal. Appl. Pyrol.*, **9**(1), 1–17 (1985).
35. M.J. Matheson, T.P. Wampler, L. Johnson, L. Atherly, L. Smucker, 'A Pyrolysis–GC/MS Library for Polymers', *Am. Lab.*, **5**, 24C–24F (1997).
36. *Polymer Pyrolysis Library*, CDS Analytical, Oxford, PA, 1998.
37. F.C.-Y. Wang, 'An HTML Approach to Creating and Maintaining a Chromatography Database', *Anal. Chem.*, **69**(19), 667A–671A (1997).
38. M.C. Statheropoulos, C.G. Georgakopoulos, G. Montaudo, 'An Expert System for the Interpretation of Pyrolysis Mass Spectra of Condensation Polymers', *J. Anal. Pyrol.*, **23**(1), 15–25 (1992).
39. C.G. Georgakopoulos, M. Statheropoulos, G. Parissakis, G. Montaudo, 'A Method for the Interpretation of Pyrolysis–Mass Spectra of Polyesters', *J. Anal. Pyrol.*, **34**(2), 127–142 (1995).
40. C.G. Georgakopoulos, M.C. Statheropoulos, G. Montaudo, 'An Expert System for the Interpretation of Pyrolysis Mass Spectra of Condensation Polymers', *Anal. Chim. Acta*, **359**(1–2), 213–225 (1998).
41. F. Martin, J.C. González-Vila, J.C. del Rio, T. Verdejo, 'Pyrolysis Derivatization of Humic Substances. 1. Pyrolysis of Fulvic Acids in the Presence of Tetramethylammonium Hydroxide', *J. Anal. Appl. Pyrol.*, **28**(1), 71–80 (1994).
42. Y. Sugimura, S. Tsuge, 'Studies on Thermal Degradation of Aromatic Polyesters by Pyrolysis–Gas Chromatography', *J. Chromatogr. Sci.*, **17**(5), 269–272 (1979).
43. K. Peltonen, 'Gas Chromatographic–Mass Spectrometric Determination of Phenols and Heterocyclic Nitrogen Compounds in the Thermal Degradation Products of Epoxy Powder Paint', *J. Anal. Appl. Pyrol.*, **10**(1), 51–57 (1986).
44. W.C. Kossa, J. MacGee, S. Ramachandran, A.J. Webber, 'Pyrolytic Methylation–Gas Chromatography: A Short Review', *J. Chromatogr. Sci.*, **17**(4), 177–187 (1979).
45. S.J. Abraham, W.J. Criddle, 'Quantitative Studies on the Pyrolytic Methylation of Simple Carboxylic Acids and Phenols', *J. Anal. Appl. Pyrol.*, **9**(1), 53–64 (1985).
46. W. Goetz, P. Lasserre, G. Kaba, 'Characterization and Identification of Synthetic High Polymers Using Pyrolysis–Gas Chromatography', in *Cosmetic Analysis: Selective Methods and Techniques*, ed. P. Boré, Marcel Dekker, New York, 1985.
47. J.W. de Leeuw, M. Baas, 'The Behavior of Esters in the Presence of Tetramethylammonium Salts at Elevated Temperatures; Flash Pyrolysis or Flash Chemolysis?', *J. Anal. Appl. Pyrol.*, **26**(3), 175–184 (1993).
48. F.J. González-Vila, J.C. Del Rio, F. Martin, T. Verdejo, 'Pyrolytic Alkylation–Gas Chromatography–Mass Spectrometry of Model Polymers. Further Insights into the Mechanism and Scope of the Technique', *J. Chromatogr., A*, **750**(1+2), 155–160 (1996).
49. J.M. Challinor, 'Pyrolysis Capillary Gas Chromatography of Alkyd Paints in Forensic Science', *J. Forensic Sci. Soc.*, **24**(3), 451–459 (1984).
50. H.-L. Hardell, N.-O. Nilvebrant, 'Analytical Pyrolysis of Spruce Milled Wood Lignins in the Presence of Tetramethylammonium Hydroxide', *Nord. Pulp Pap. Res. J.*, **11**(2), 121–126 (1996).
51. M.J. Martin, J.C. del Rio, F.J. González-Vila, T. Verdejo, 'Thermally Assisted Hydrolysis and Alkylation of Lignins in the Presence of Tetra-alkylammonium Hydroxides', *J. Anal. Appl. Pyrol.*, **35**(1), 1–13 (1995).
52. P.G. Hatcher, M.A. Nanny, R.D. Minard, S.D. Dible, D.M. Carson, 'Comparison of Two Thermochemolytic Methods for the Analysis of Lignin in Decomposing Gymnosperm Wood: The CuO Oxidation Method and the Method of Thermochemolysis with Tetramethylammonium Hydroxide (TMAH)', *Org. Geochem.*, **23**(10), 881–888 (1996).
53. H.-L. Hardell, N.-O. Nilvebrant, 'A Rapid Method to Distinguish Between Free and Esterified Fatty Acids by Pyrolysis Methylation Using Tetramethylammonium Acetate or Hydroxide', *J. Anal. Appl. Pyrol.*, **52**(1), 1–14 (1999).
54. P.A. Vimalasiri, J.K. Haken, R.P. Burford, 'Gas Chromatographic Analysis of Polyether-based Polyurethanes after Acid Fusion', *J. Chromatogr.*, **355**(1), 141–147 (1986).
55. F.C.-Y. Wang, 'Polymer Degradation Mechanism Reselection through Derivatization for Qualitative Pyrolysis Gas Chromatography Analysis', *Anal. Chem.*, **70**(17), 3642–3648 (1998).
56. L. LaFleur, 'Sources of Chemicals in Effluents', in *Environmental Fate and Effects of Pulp and Paper Mill Effluents*, ed. M.R. Servos, K.R. Munkittrick, J.H. Carey, G.J. Van der Kraak, St. Lucie Press, Delray Beach, FL, 22, 1996.
57. G.O. Aspinall (ed.), *The Polysaccharides*, Academic Press, New York, Vol. 1, 1982.
58. A. Ohnishi, A. Kato, E. Takagi, 'Curie Point Pyrolysis of Cellulose', *Polymer J.*, **7**(4), 431–437 (1975).
59. A.D. Pouwels, A. Tom, G.B. Eijkel, J.J. Boon, 'Characterization of Beech Wood and its Holocellulose and Xylan Fractions by Pyrolysis–Gas Chromatography–Mass Spectrometry', *J. Anal. Appl. Pyrol.*, **11**, 417–436 (1987).
60. M. Wada, S. Fujishige, S. Uchino, N. Oguri, 'Pyrolysis of Flame-retardant Cellulose Fibers Using a Curie-point Pyrolyzer', *J. Soc. Fiber Sci. Technol.*, **52**(10), 558–561 (1996).

61. R.J. Helleur, D.R. Budgell, E.R. Hayes, 'Identification of the Composition of Oligosaccharides and Polysaccharides by Pyrolysis–Capillary Gas Chromatography', *Anal. Chim. Acta*, **192**(2), 367–372 (1987).
62. R.J. Helleur, 'Characterization of the Saccharide Composition of Heteropolysaccharides by Pyrolysis–Capillary Gas Chromatography–Mass Spectrometry', *J. Anal. Appl. Pyrol.*, **11**, 297–311 (1987).
63. J. Kelly, M. Mackey, R.J. Helleur, 'Quantitative Analysis of Saccharides in Wood Pulps by Quartz-tube Pyrolysis–Polar Phase Gas Chromatography', *J. Anal. Appl. Pyrol.*, **19**, 105–117 (1991).
64. H.-M. Chang, G.G. Allan, 'Oxidation', in *Lignins – Occurrence, Formation, Structure and Reactions*, eds. K.V. Sarkanen, C.H. Ludwig, Wiley-Interscience, New York, 1971.
65. O. Faix, D. Meier, I. Grobe, 'Studies on Isolated Lignins and Lignins in Woody Materials by Pyrolysis–Gas Chromatography–Mass Spectrometry and Off-line Pyrolysis–Gas Chromatography with Flame Ionization Detection', *J. Anal. Appl. Pyrol.*, **11**, 403–416 (1987).
66. A. Izumi, K. Kuroda, H. Ohi, A. Yamaguchi, 'Structural Analysis of Lignin by Pyrolysis–Gas Chromatography; Comparative Studies of Pyrolysis–Gas Chromatography and Nitrobenzene Oxidation for the Determination Method of Lignin Composition in Hardwood', *Jpn. Tappi J.*, **49**(9), 1339–1346 (1995).
67. A. Izumi, H. Ohi, Y. Yasuta, and K. Kuroda, 'Analytical Pyrolysis of Hardwood Type Dehydrogenation Polymers and Lignins', in *Proceedings, 9th International Symposium on Wood Pulp Chemistry*, CPPA, Montreal, 41-1–41-4, 1997.
68. J. Ralph, R.D. Hatfield, 'Pyrolysis–GC/MS Characterization of Forage Materials', *J. Agric. Food Chem.*, **39**(8), 1426–1437 (1991).
69. S. Suzuki, A. Izumi, H. Ohi, K. Kuroda, A. Yamaguchi, 'Structural Analysis of Alkaline Cooked Lignin by Pyrolysis–Gas Chromatography. Evaluation of Uniformity in Delignification of Beech Wood Chips', *Jpn. Tappi J.*, **48**(9), 94–102 (1994).
70. S. Suzuki, H. Ohi, K. Kuroda, A. Yamaguchi, 'Analysis of Lignin in Various Kinds of Paper by Pyrolysis–Gas Chromatography', in *Proceedings, 62nd Pulp and Paper Research Conf.*, Japan TAPPI, Tokyo, 86–91, 1995.
71. G. Galletti, P. Bocchini, M. Guadalix, G. Almendos, S. Camarero, A. Martinez, 'Pyrolysis Products as Markers in the Chemical Characterization of Paperboards from Wastepaper and Wheat Straw Pulps', *Bioresourc. Technol.*, **60**(1), 51–58 (1997).
72. C. Schiegel, B. Helmreich, P.A. Wilderer, 'Method of Quantification of Lignin in Paper Mill Wastewaters by Pyrolysis GC/MS', in *Wom Wasser 88*, eds. K. Haberer, B. Hamburger, M. Jekel, P. Koppe, S. Schmidt, Wiley-VCH, Weiheim, Germany, 137–147, 1997.
73. P.G. Hatcher, D.J. Clifford, 'Flash Pyrolysis and In Situ Methylation of Humic Acids from Soil', *Org. Geochem.*, **21**(10/11), 1081–1092 (1994).
74. J.M. Challinor, 'Characterization of Wood by Pyrolysis Derivatization–Gas Chromatography–Mass Spectrometry', *J. Anal. Appl. Pyrol.*, **35**(1), 93–107 (1995).
75. K. Kuroda, A. Izumi, 'Analysis of Lignin by Pyrolysis–Methylation: Pyrolysis Products Derived from β-O-4, β-β, β-5, and β-1 Substructures', *J. Jpn. Wood Res. Soc.*, **43**(1), 112–114 (1997).
76. P. Bocchini, G.C. Galletti, S. Camarero, A.T. Martinez, 'Absolute Quantitation of Lignin Pyrolysis Products Using an Internal Standard', *J. Chromatogr., A*, **773**(1 + 2), 227–232 (1997).
77. M. Kleen, G. Lindblad, S. Backa, 'Quantitation of Lignin and Carbohydrates in Kraft Pulps Using Analytical Pyrolysis and Multivariate Data Analysis', *J. Anal. Appl. Pyrol.*, **25**, 209–227 (1993).
78. M. Kleen, K. Lindstrom, 'Polysaccharides and Lignins in Newsprint Process Waters Studied by Pyrolysis–Gas Chromatography–Mass Spectrometry', *Nord. Pulp Pap. Res. J.*, **9**(2), 111–119 (1994).
79. D.J. Clifford, D.M. Carson, D.E. McKinney, J.M. Bortiatynski, P.G. Hatcher, 'A New Rapid Technique for the Characterization of Lignin in Vascular Plants: Thermochemolysis with Tetramethylammonium Hydroxide (TMAH)', *Org. Geochem.*, **23**(2), 169–175 (1995).
80. W.M.G.M. van Loon, J.J. Boon, B. de Groot, 'Qualitative Analysis of Chlorolignins and Lignosulfonates in Pulp Mill Effluents entering the River Rhine Using Pyrolysis–Mass Spectrometry and Pyrolysis–Gas Chromatography–Mass Spectrometry', *J. Anal. Appl. Pyrol.*, **20**(2), 275–302 (1991).
81. K. Koda, H. Shintani, Y. Matsumoto, G. Meshitsuka, 'Environmental Fate of High Molecular Weight Chloroorganics Produced by Chlorine Bleaching (1). Development of a New Analytical Method by the Use of Pyrolysis Gas Chromatography', in *Proceedings, 1998 International Pulp Bleaching Conference*, Helsinki, Gummerus Oy, Jyvaskyla, 609–612, 1998.
82. C. Flodin, M. Ekelund, H. Boren, A. Grimvall, 'Pyrolysis–GC/AED and Pyrolysis–GC/MS Analysis of Chlorinated Structures in Aquatic Fulvic Acids and Chlorolignins', *Chemosphere*, **34**(11), 2319–2328 (1997).
83. I.H. Isenberg, *Pulp and Paper Microscopy*, 3rd edition, Institute of Paper Chemistry, Wisconsin, USA, 1967.
84. B.B. Sitholé, 'Analysis of Pitch Deposits', in *Pitch Control, Wood Resin and Deresination*, eds. E. Back, L.H. Allen, TAPPI Press, 2000.
85. B.B. Sitholé, L.H. Allen, 'Quantitative Analysis of Synthetic Polymers in Pulp and Paper Mill Pitch Deposits by Pyrolysis Gas Chromatography', *J. Pulp Pap. Sci.*, **20**(6), 168–172 (1994).
86. J.C. del Rio, A. Gutiérrez, J. Gonzalez-Villa, 'Analysis of Impurities Occurring in a Totally Chlorine-free

Bleached Kraft Pulp', *J. Chromatogr. A*, **830**, 227–232 (1999).
87. H. Ono, T. Miyanishi, 'Studies on Wet-end Optimization (2) – Quantitative Analysis of SB Latex in Coated Broke Determined by Pyrolysis Gas Chromatography', in *Proceedings, 62nd Pulp & Paper Research Conf.*, Japan TAPPI, Tokyo, paper 17, 82–85, 1995.
88. Y. Ishida, H. Ohtani, T. Kato, S. Tsuge, T. Yano, 'Determination of Rosin Sizing Agents in Paper by Pyrolysis–Gas Chromatography–Mass Spectrometry', *Tappi J.*, **77**(3), 177–183 (1994).
89. B. Evans, 'Alkyl Ketene Dimer Sizes', in *The Sizing of Paper*, ed. W.F. Reynolds, TAPPI Press, Atlanta, 33–50, 1989.
90. B.B. Sitholé, S. Nyarku, L.H. Allen, 'Chromatographic Methods and Procedures for the Determination of Alkyl Ketene Dimer and its Hydrolysis Product in Pulp and Paper Matrices', *Analyst*, **120**(4), 1163–1169 (1995).
91. P.J. Dart, D.V. McCalley, 'Determination of Alkylketene Dimer Sizing-agent Products in Paper by Capillary Gas Chromatography', *Analyst*, **115**(1), 13–16 (1990).
92. T. Yano, H. Ohtani, S. Tsuge, T. Obokata, 'Determination of Wet-strength Resin in Paper by Pyrolysis–Gas Chromatography', *Analyst*, **117**(5), 849–852 (1992).
93. M. Vrbanac, D. Dixon, 'Quantitative Analysis of AKD and ASA in Paper Using Pyrolysis Gas Chromatography', in *Proceedings, Scientific and Technical Advances in the Internal Sizing of Paper and Board*, London, UK, paper 11, Pira International, Leatherhead, UK, 1997.
94. B.B. Sitholé, B. Ambayec, Unpublished data, 1998.
95. A. Isogai, 'AKD Sizing, its Retention and Sizing Mechanisms', in *Proceedings, Scientific and Technical Advances in the Internal Sizing of Paper and Board*, London, UK, paper 10, Pira International, Leatherhead, UK, 1997.
96. T. Yano, H. Ohtani, S. Tsuge, T. Obokata, 'Determination of Wet-strength Resin in Paper by Pyrolysis–Gas Chromatography', *Tappi J.*, **74**(2), 197–201 (1991).
97. P. Selsbo, I. Ericsson, M. Kleen, 'Characterization of Sulfur in Wood Pulps Using Pyrolysis–Gas Chromatography with Sulfur-selective Detection', *J. Anal. Appl. Pyrol.*, **43**(1), 1–14 (1997).
98. W.L. Banwart, J.M. Bremner, 'Evolution of Volatile Sulfur Compounds from Soil Treated with Sulfur-containing Organic Materials', *Soil Biol. Biochem.*, **8**(5), 439–443 (1976).
99. T. Wampler, E. Levy, 'Pyrolysis GC in the Analysis of Inks and Papers', *LC GC*, **4**(11), 1112–1116 (1986).
100. E. Levy, T. Wampler, 'Applications of Pyrolysis Gas Chromatography–Mass Spectrometry to Toner Materials from Photocopiers', *J. Forensic Sci.*, **31**(1), 258–271 (1986).
101. W.F. DeGroot, W.-P. Pan, M.D. Rahman, G.N. Richards, 'First Chemical Events in Pyrolysis of Wood', *J. Anal. Appl. Pyrol.*, **13**(2), 221–231 (1988).
102. Å. Ingemarsson, U. Nilsson, M. Nilsson, J.R. Pedersen, J.O. Olsson, 'Slow Pyrolysis of Spruce and Pine Samples with GC/MS and GC/FTIR/FID', *Chemosphere*, **36**(14), 2879–2889 (1998).
103. S. Tsuge, T. Okumoto, Y. Sugimura, T. Takeuchi, 'Pyrolysis–Gas Chromatographic Investigation of Fractionated Polycarbonates', *J. Chromatogr. Sci.*, **7**(4), 253–256 (1969).
104. M. Blazsó, T. Tóth, 'Thermal Decomposition of Methylene Bridges and Methyl Groups at Aromatic Rings in Phenol–Formaldehyde Polycondensates', *J. Anal. Appl. Pyrol.*, **10**(1), 41–50 (1986).

X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Krishna M. Pamidimukkala
Champion International Corporation, West Nyack, USA

1 Introduction **1**
2 X-Ray Photoelectron Spectroscopy **2**
2.1 Surface Charging of Paper Samples 2
2.2 Quantitative Analysis 2
2.3 Depth Profile 3
2.4 Sample Handling 3
2.5 X-ray Photoelectron Spectroscopy Spectrum 3
2.6 Carbon Chemistry 3
3 Applications **4**
3.1 Pulp Characterization 4
3.2 Coated Papers 7
3.3 Nonimpact Printing Papers 10
3.4 Adhesion 12
3.5 Paper Additives 13
Acknowledgments **15**
Abbreviations and Acronyms **15**
Related Articles **15**
References **15**

X-ray photoelectron spectroscopy (XPS) also popularly known as electron spectroscopy for chemical analysis (ESCA) involves the irradiation of a solid surface in ultrahigh vacuum with monoenergetic soft X-rays. The X-rays cause electrons of an atom to be ejected and to be ionized as photoelectrons. The number of electrons and their kinetic energies are then measured. The energy of the emitted electron is very characteristic of the atom from which the electron originated and its environment. The XPS spectrum is a plot of the number of emitted electrons per energy interval versus their binding energies. Quantitation can be achieved either by fundamental parameter calculation or comparison with appropriate standards. In the pulp and paper industry, XPS has successfully been used to study migration of binder in coated paper, the surface chemistry of pulps and its effect on the properties of the final products; the adhesion between paper and polymer, toner or ink; and the chemical additives on the paper surface.

As with most analytical instrumentation, XPS has its limitations and disadvantages and is often used in conjunction with other, complementary techniques such as SIMS (secondary ion mass spectrometry) and FTIR (Fourier transform infrared) spectroscopy for the identification of organic compounds on paper surfaces. XPS is surface sensitive which renders it useless if the underlying cause of the problem might be other than surface related. XPS provides only semiquantitative information. A more precise quantification of the elements is possible, but requires exhaustive calibration.

1 INTRODUCTION

The surface composition and the surface chemistry of uncoated and coated papers strongly affect their performance. Since paper frequently is used for some form of printing application, the paper surface and its interaction with the printing medium (inks) become very critical for good print quality. With the advent of new and ever changing printing technologies, great demand is placed on paper manufacturers to change the commodity paper technologies to meet the challenges of new paper end-user requirements. The ability of the paper manufacturer to come up with paper substrates to accommodate new printing technologies expanded the markets for both print providers and paper vendors. For example, paper is now marketed which was specifically developed for ink jet printers, laser printers and color copiers. The surface chemistry and surface energetics of the papers play a vital role in the development of new paper grades and optimizing the paper performance.

A number of analytical techniques are now available to provide surface chemistry and surface compositional information without significant interference from the bulk of the sample. XPS, also known as ESCA, is ideally suited to study the paper surfaces. Although the name XPS is scientifically more correct, the commonly used name is ESCA, the title given by Siegbahn who developed this technique in 1967. Since then, several commercial XPS instruments have been developed. XPS, which has now matured into one of the most powerful and valuable methods for surface analysis, gained its distinction due to:

1. ease of interpretation of the data;
2. ability to detect all elements (except hydrogen);
3. wealth of chemical bonding information and;
4. its application to a large number of areas including, organics, semiconductors, fibers, films, powders, metals and biomaterials.

The application of this technique for the analysis of cellulosic materials was pioneered by Dorris and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Gray.[1–3] Since then there have been several research papers published involving the use of XPS for paper, pulp and wood surface analyses. Paper is a composite material consisting of wood/cellulose fibers, filler and other additives. Because of its heterogeneous nature, the surface chemical composition can vary significantly and hence so too can its performance. Thus, it is only natural to extend the application of XPS for surface analysis of paper.

This article is not meant to be a rigorous treatment of the theory underlying the XPS or of instrumentation, but a concise review of the literature with more emphasis placed on the application of XPS for paper surface analysis. However, basic principles, advantages, and disadvantages of XPS are discussed briefly. The reader will find suitable references in the literature for detailed descriptions and theoretical aspects of this and other surface analysis instruments.[4–6] Istone[6] gave a good account of the theory and a few applications of XPS for paper surface analysis.

2 X-RAY PHOTOELECTRON SPECTROSCOPY

XPS method is based on Einstein's classic photoelectric effect. When a clean solid surface is bombarded by X-ray photons it emits electrons. The emitted electron, before leaving the surface, has to overcome the binding energy (BE) of the orbital in which it was located in the atom. The amount of remaining incident radiation energy results in the velocity of the ejected electron, that is its kinetic energy (KE). From the energy of the incident radiation and the KE one can calculate the BE of the electron. The BE identifies the elements present and the oxidation state of that element. The energy of the emitted electron, which is characteristic of the atom from which the electron originated and its environment, is measured by a hemispherical electron analyzer in most commercial XPS instruments. The KE of an electron leaving the solid surface is given by Equation (1):

$$\mathrm{KE} = h\nu - \mathrm{BE} - \phi_s \tag{1}$$

where $h\nu$ = energy of the photon, ϕ_s = spectrometer work function.

Since the mean free path of the electrons is very small, only the electrons that are generated from the top few atomic layers of the surface can reach the detector. Hence XPS is deemed as a surface sensitive technique. The most commonly used X-ray sources are Mg Kα, Al Kα, and monochromatic Al Kα. The monochromatic source is normally used for delicate organic surfaces including paper for which high energy resolution and minimal sample damage are necessary. The disadvantage of a monochromatic source is charge buildup on the sample surface. This is generally not a major problem with nonmonochromatic sources because white radiation neutralizes the charge on the sample surface.

2.1 Surface Charging of Paper Samples

For electrically insulating samples such as papers and polymers, a positive charge on the sample surface will build up as photoelectrons are ejected. Without a method for charge neutralization, photoelectron peaks tend to change their shape and shift to higher binding energies, sometimes by as much as several hundred electron volts (eV). The current standard method for neutralization is a source of low energy electrons, usually an electron flood gun. Even with good charge neutralization, minor shifting of peaks could occur and the spectrum should be corrected for charging. It is ordinarily done by choosing a peak for which the correct BE is known. The entire spectrum is then shifted so that the reference peak is in the correct position. In the case of paper or polymer samples, the hydrocarbon component of the C (1s) line at 285.00 eV is used as a reference line. If the sample is heterogeneous (as is the case with paper) particles of different materials can charge to different extents. This phenomenon, known as differential charging, makes the interpretation of the spectrum more complicated.

2.2 Quantitative Analysis

A rigorous treatment of the theory underlying quantitative XPS analysis is beyond the scope of this article. In practice, the peak areas for the elements are converted into fractional atomic concentrations using the relative atomic sensitivity factors as shown in Equation (2):

$$C_a = \frac{I_a/S_a}{\sum_n I_n/S_n} \tag{2}$$

where C_a = fractional atomic concentration of element a, I = peak intensity, S = relative atomic sensitivity factor, and the sum is performed over all elements detected in the sample.

Because of the fairly crude assumptions, the concentrations measured using this approach are probably associated with an error of about 10%. Values of S_n may be calculated from theory or derived empirically by recording spectra from standard materials. XPS instruments are generally provided with the atomic sensitivity factors and appropriate algorithms as a part of a software package for calculating the concentrations.

2.3 Depth Profile

The maximum sampling depth obtainable, even under optimum conditions, is only about 10 nm which corresponds to about 16 carbohydrate chain layers, because the thickness of a single glucose unit is approximately 0.6 nm.[7] In most cases, the information about the surface composition within this analysis depth is extremely valuable and further analysis may not be necessary. However, XPS can also be used to investigate the variation of the composition with depth into the sample thereby producing a composition–depth profile of the surface of interest. The inert gas ion bombardment sputtering method is ordinarily used to remove sample surface beyond 10 nm. But for a composition–depth profile, in the range 1–10 nm, the angle-dependent technique is normally adapted. This is accomplished by selecting a narrow acceptance angle and varying the take off angle.

2.4 Sample Handling

For the majority of XPS applications, sample preparation and mounting are not difficult. The paper (or other samples) is used as received. The samples are mounted or mechanically held to the sample stage which is then introduced into the chamber for analysis after initial pump down. For pulp samples, laboratory hand sheets are usually made without any additives. But for characteristic bulk composition analysis, pulp or other samples are ground to powder which can be introduced by one of several methods; carefully dusting the powder on a adhesive tape, pressing powder into an indium foil, or pressing the powder into pellets. Since the XPS instrument is surface sensitive, extreme protection must be taken to avoid any contamination of the sample during and after the sample preparation.

2.5 X-ray Photoelectron Spectroscopy Spectrum

For a typical ESCA investigation, a broad scan survey spectrum (low resolution) is usually obtained first, especially if the sample is an unknown. Once the elements are identified from the survey spectrum, detailed (high resolution) scans of selected elements of interest can be obtained for chemical state identification and quantitative analysis. Survey scans ordinarily range from 1000 to 0 eV BE which is sufficient for the identification of detectable elements. A higher range of 1100–0 eV is recommended for paper samples containing Na or Mg because the strongest peaks for these elements occur above 1000 eV.

Typically the survey scans are obtained at high-pass energies and for short periods of time. The high resolution scans, on the other hand, are obtained using a narrow BE range, low-pass energy and for longer periods of time. Quantitative analysis and deconvolution manipulations of the data are done on the high-resolution spectra. The survey spectrum is displayed as a plot of BE versus intensity (counts per second per unit energy interval). Several peaks are observed in ESCA spectra but the most intense peaks are photoelectron lines. Other peaks often observed are: Auger lines, shake-up lines, valence band lines and so on which are not discussed here.

In general, interpretation of the ESCA spectrum is most readily accomplished by first identifying the lines that are almost always present, specifically those of carbon and oxygen, then identifying the other major and minor lines. Most instrument manufacturers have software programs to identify the peaks using the BE of each peak. Also, standard BE tables are available in many standard text books for manual identification of the peaks.[6] An ESCA survey spectrum of a coated paper is shown in Figure 1. From this spectrum it can be inferred that the coating consisted of clay (Al and Si) and calcium carbonate (Ca) as pigments and probably a latex binder.

2.6 Carbon Chemistry

XPS spectra not only characterize the elements present on a surface but also provide the chemical bonding information for the elements. It has been observed that the energy of the electrons emitted from a given element may be altered (or shifted) depending on the type of chemical bond formed by the element in question. The magnitude of this *chemical shift* indicates the types of chemical bond present on the surface. For example, the energy of the electrons originating from carbon bonded to other carbons (C–C, 285.0 eV) is different from the energy of the electrons originating from carbon bonded to oxygen (C–O, 286.5 eV or C=O, 288.5 eV). Binding energies and chemical shifts for carbon (1s) and other atoms in organic compounds have been measured and are available in the literature.

Dorris and Gray[2] and Gray[3] showed from high-resolution ESCA data for mechanically refined softwood pulp that the carbon (1s) spectra consisted of three overlapping but distinctly different peaks. Based on the binding energies reported in the literature for carbon–oxygen bonds, they attributed the low-BE peak (C1) to lignin, (C–C bonds); the middle peak (C2) to cellulose (C–OH bonds); and the high-BE peak (C3) to cellulose (O–C–O bonds) where two oxygen atoms are bonded to carbon at the reducing end. Simple deconvolution methods, normally provided by the instrument software, are used to estimate the relative abundance of each class of carbon atom. The C4 carbon atoms, representing carbon bonded to a carbonyl and noncarbonyl oxygen atoms (O–C=O) are

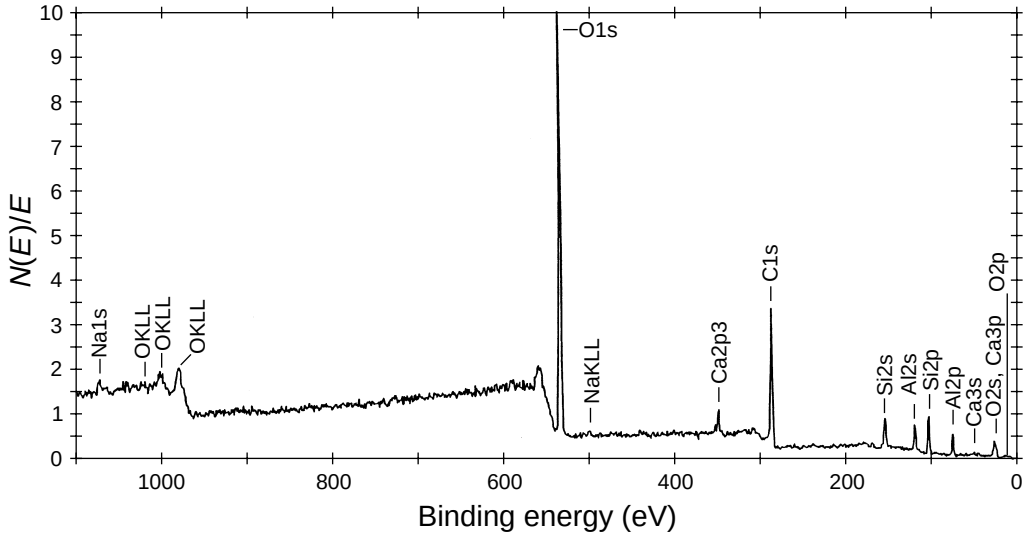

Figure 1 XPS survey spectrum of a coated paper.

Table 1 Classification of C (1s) and O (1s) XPS lines for cellulosic materials[14,15]

Group	Chemical shift (eV)	Carbon or oxygen bond to	Examples
Carbon			
C1	0.0 ± 0.4	C−C, C−H	Hydrocarbons, lignin, resins
C2	1.5 ± 0.4	C−O, C−O−C	Cellulose, alcohols, ether
C3	3.0 ± 0.4	C=O, O−C−O	Cellulose, aldehyde,
C4	4.5 ± 0.4	O−C=O (or CO_3)	Organic acids, carbonate
Oxygen[a]			
O1	-1.5 ± 0.4	O−C=**O**, Ph−O−Ph	Lignin, acrylate, phenyl ether
O2	0.0 ± 0.4	C=O, C−O, C−O−C, **O**−C=O	Cellulose, acrylate
O3	1.5 ± 0.4	Ph−OH, Ph−O−C	Lignin

[a] The oxygen atom of interest is underlined. The reference lines for C1 = 285.0 eV and for O2 = 533.2 eV.

generally difficult to detect owing to low concentrations of carboxylic acid groups present on the fiber surface. The C4 peak, in the case of coated papers where the concentrations are greater, is clearly seen corresponding to carbon bonded to three oxygen atoms as in calcium carbonate which is used as a pigment in coating formulations.

Because carbon and oxygen are the primary constituents of paper, the BE shifts for these two elements are shown in Table 1. It should be noted that the observed chemical shifts and the absolute binding energies of the elements to some extent depend on the instrument, surface neutralization efficiency, sample type and surface roughness.

Surface damage of cellulosic materials can occur as result of exposure to high-energy X-ray radiation which might cause some physical and chemical changes without changing their appearance. Ahmed et al.[8] studied the chemical degradation of Whatman filter paper when exposed to a high-energy X-ray source. With increasing exposure time, the O/C ratio and C2 component decreased, while the C1 and C3 components increased. They explained the results by a double dehydroxylation mechanism in which the cellulose was converted into an enone. Chemical changes of the cellulose surface structure have also been reported by Carlsson and Strom.[9] The effect of these chemical changes should be taken into account in the analysis of ESCA data.

3 APPLICATIONS

3.1 Pulp Characterization

The surface chemical composition of paper-grade fibers has great influence on fiber–fiber bonding, fiber–paper additive interactions, and ultimately on the final properties of the paper. The fibers undergo complex changes

Table 2 O/C ratios and C (1s) components of various pulps

Pulp	O/C Ratio	C1	C2	C3	C4	Refs.
Cellulose (theoretical)	0.83	0	83	17	0	1, 10
Filter paper	0.75	7.8	71.8	17.4	3.0	9
Lignin (theoretical)	0.31–0.40	49.2	48.8	2.0		1
Extractives	0.12	93	5		2	12
Aspen, CTMP	0.47	34.2	59.4	6.4		10, 11
Aspen, CMP	0.41	43.1	50.8	6.1		10, 11
Aspen, explosion	0.51	32.5	60.7	6.8		10, 11
Bleached softwood kraft	0.76	10	71	18	1	13
Unbleached softwood kraft	0.67	17	67	15	1	13
Bleached softwood	0.75	14.8	65.1	19.6	0.6	[a]
Bleached hardwood	0.73	15.0	69	16.3	0.8	[a]
Unbleached groundwood	0.58	48.0	44	6.6	1.3	[a]
Eucalyptus pulp	0.75	14	67	17	1.2	[a]
BCTMP (50 : 50 hard/soft wood)	0.58	29	59	11	1	[a]

CTMP, chemithermomechanical pulp; CMP, chemimechanical pulp; BCTMP, bleached chemithermomechanical pulp.

[a] All XPS spectra and results presented in this article were generated by the author on the Perkin-Elmer PHI-5500 Multi-Technique System at Champion International Corporation in West Nyack, NY. All experiments were performed using monochromatic aluminum X-rays ($K_\alpha = 1486.7$ eV) at 400 W (15 keV). The survey spectra were acquired at a pass energy of 93.9 eV whereas the multiplex spectra were acquired at a pass energy of 58.7 eV. The analysis area was about 800 μm in diameter. High-resolution carbon spectra were obtained at pass energies of either 58.7 eV or 5.87 eV. Curve fitting and quantitation were performed using routines provided by the PHI instrument software. In all cases the experiments were performed on three different sample locations and an average was calculated.

during the pulp manufacturing process, creating new fiber surfaces at each stage of pulping, bleaching, drying, and any other chemical treatment. Here, a few examples of how XPS has been used to characterize fiber surface at different stages of the pulp manufacturing process are discussed. The XPS survey spectrum of pulps shows only peaks due to oxygen and carbon and perhaps chlorine from the bleaching chemical. But the high-resolution carbon spectra and oxygen-to-carbon ratios of pulps reveal a great deal information regarding their surface structure.

Dorris and Gray[1,2] studied the surface chemical composition of three mechanical pulps (stone ground, refiner mechanical and thermomechanical pulps) by XPS and showed that these fibers were distinctly different from one another with respect to the oxygen–carbon ratio and much different from the cotton fibers or the filter paper. From the molecular formula for pure cellulose, $C_6H_{10}O_5$, the theoretical O/C ratio can be calculated to be 0.83 (five oxygens/six carbons). Almost all pulp fibers exhibit ratios much less than 0.83 due to the presence of carbon-rich lignin and or other extractives on the surface of the fibers. Practically it is difficult, in the manufacturing process, to remove the lignin completely without severely damaging the fiber surface. Lignin, a large complex molecule, exhibited a O/C ratio of about 0.3. In most cases, the O/C ratio of fibers falls between 0.83 and 0.3.

Extractive components, such as abietic acid which has an O/C ratio of 0.1, also reduce the O/C ratio of pulps. A list of O/C ratios for various pulps is shown in Table 2. The purpose of this table is only to demonstrate the differences in the surface chemical composition of various pulps. The actual concentrations of these components in a given situation might vary significantly depending upon the instrument, extent of bleaching, contamination, etc.

High-resolution carbon spectra of pulps have also been used to characterize the pulp surfaces. Theoretically, there is 0% C1, 83% C2 and 17% C3 for pure cellulose; that is all the carbons in cellulose are linked to at least one oxygen. In practice, some of the C1 on fiber surfaces is always seen due to the presence of surface lignin, extractives and other additives which contribute to the C1 content. Using Freudenberg's empirical model, Dorris and Gray[1] came up with the ratio of carbon contents for lignin; $C1 : C2 : C3 = 49.2 : 48.8 : 2.0\%$. Other empirical formulas and lignin extracted from other sources might give different ratios for these three classes of carbon. Abietic acid, on the other hand, has 95% C1 component.

High-resolution carbon spectra for bleached kraft hardwood and unbleached groundwood pulps are shown in Figure 2. High levels of C1 in groundwood pulp is a strong indication of high levels of lignin and other extractives compared with bleached hardwood pulp.

Very little attention is usually paid to the oxygen spectra of cellulose fibers because of its complex shift behavior compared to C (1s). The complexity arises from the electronegativity and polarizability of substituent groups interacting directly or indirectly with oxygen. However,

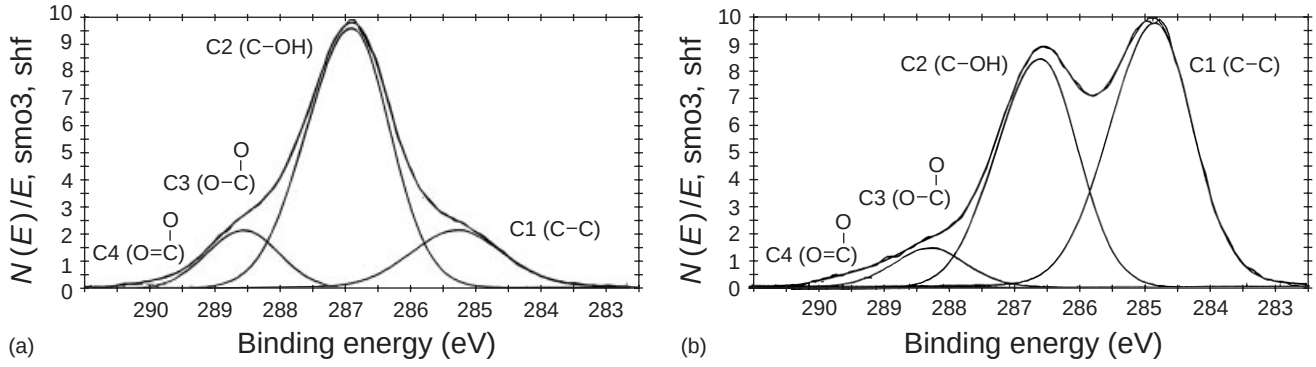

Figure 2 Curve resolved XPS C (1s) spectra of (a) bleached hardwood and (b) unbleached groundwood pulp samples. See Table 1 for an explanation of C1, C2, C3 and C4. The raw data was smoothed (smo3) and shifted (shf) to the carbon reference peak at 285.0 eV.

the works of Koubaa et al.,[14] Kamdem et al.,[15] and Hua et al.[10,11] shed some light on the shift behavior of oxygen in cellulose fibers. The general classification of oxygen peak components is shown in Table 1.

The effect of a steam explosion pulping process on fiber composition was examined by Hua et al.[10,11] The explosion pulps showed higher O/C ratios than CMP and CTMP, indicating that the explosion pulping process had removed more extractives and lignin from the fiber surfaces. Because this facilitates hydrogen bonding between fibers, the paper made from explosion pulps provides much greater strength than the paper made from conventional pulps.

The effect of digestion parameters on the surface chemical composition of fibers was investigated by Laine et al.[12] using XPS. A series of unbleached pulp samples were prepared by digestion in the presence of three different levels of effective alkali such that the kappa numbers are in the range 18–57. The kappa number is the volume (in milliliters) of 0.1 N $KMNO_4$ consumed by 1 g of moisture free pulp in 0.5 N sulfuric acid after a 10-min reaction time at 25 °C under conditions such that one-half of the permanganate remains unreacted. The 50% residual permanganate is titrated to determine the exact consumption. The kappa number test is used to monitor the degree of cook (extent of delignification during pulping) and to control the process between bleaching stages. Bleached pulps always exhibit low kappa numbers and unbleached pulps give high numbers. From the XPS data for these and organic-solvent-extracted pulps Laine et al. concluded that:

1. There was strong enrichment of lignin and extractives (relative amount of alkyl C1) on the fiber surfaces.
2. Increase in digestion time (i.e. decreasing kappa number) resulted in a decreased amount of C1 for all the pulps.
3. The alkali dosage had little effect on the surface lignin content but strongly affected the surface concentration of extractives.
4. The carboxylic group content in the surface decreased with increasing kappa number.

The effects of bleaching chemicals and the bleaching sequence on the surface chemical composition were also determined using XPS by Laine et al.[13] Here, unbleached softwood kraft pulps were treated with different conventional bleaching chemicals (oxygen, peroxide, ozone, and chlorine dioxide) in multiple sequences. Both the bulk and surface concentrations of lignin and extractives were investigated after each bleaching sequence. While oxygen and peroxide were very ineffective in removing lignin, the strong oxidant ozone removed a significant amount of lignin from the fiber surfaces. The effectiveness of chlorine dioxide in removing lignin strongly depended on the number of treatments. The surface content of lignin, calculated from the XPS data, showed a linear relationship to the surface brightness of these pulps; a small reduction in lignin showed a significant increase in brightness.

Koubaa et al.[14] studied the effect of drying temperature on the surface composition of birch and spruce CTMP pulps by XPS. At temperatures much higher than the glass transition temperatures of lignin, fluid lignin covered the surface of the fibers as evidenced by the decrease in O/C ratios and in the fractional areas of C2 for both birch and spruce pulps.

Buchert et al.[7] applied ESCA to characterizing enzyme-treated pulps and to determine the location of various components on the fiber surfaces. Owing to the specificity of the enzymes (xylanase and mannase) they were able to peel off the surface sugars (xylan and glucomannan) selectively without destroying the fiber structure. ESCA analysis of the treated

and untreated birch kraft pulps has shown that the concentration of extractives was reduced when treated with xylanase as compared with treatment with mannase, indicating that the extractives are preferentially associated with xylan. While there was no effect on surface lignin content with birch pulps, the lignin content increased in enzyme-treated pine pulp. They suggested that the lignin preferentially located under xylan which, when removed by enzyme treatment, exposed lignin to the surface.

3.1.1 Modification of Cellulose Fibers

Cellulose fibers are generally modified to improve adhesion, wettability and bonding properties. Surface modification is usually done by flame, plasma, corona, and chemical treatments. ESCA is ideally suited and primarily used to confirm the surface chemical compositional changes with respect to each of these treatments. Several examples of plasma treatment of cellulose fibers to enhance adhesion properties can be found in the latter part of this article.

Felix et al.[16] reported significant compositional changes including the formation of new bonds, C−N, C=N, N−C=O, on the fiber surface when treated with ammonia plasma. Belgacem et al.[17] studied the corona treatment of cellulose fibers by XPS. The corona treated fibers showed high levels of surface oxygen compared with untreated fibers, indicating corona-induced surface oxidation of impurities making the surface more hydrophilic. They also observed an increase in the carboxylic content (C4) of treated fibers at all corona current levels.

Apart from the more traditional uses of wood pulps, there has been a growing interest in the use of wood pulps in composite materials because of their excellent mechanical properties, low density, availability and cost. However, the problem of improving the compatibility between the hydrophilic cellulose fibers and hydrophobic synthetic polymers has been a major issue. As coupling agent, a bifunctional molecule is often used to overcome incompatibility between these two phases. Zadorecki and Flodin[18] studied the grafting of trichloro-*s*-triazine-based coupling agents onto cellulose fibers to improve adhesion between fibers of cellulose and polyester. The amounts of oxygen and nitrogen were used for calculating the concentrations of cellulose and coupling agent, respectively. XPS data strongly suggested that the coupling agent concentrated on the treated fiber surface.

Kazayawoko et al.[19] studied the reaction between bleached kraft cellulose and lignin and maleated polypropylene, a coupling agent. High-resolution carbon spectra of the treated pulps showed high levels of C1, compared with untreated pulp, indicating surface enrichment of aliphatic carbon due to a polypropylene backbone. The appearance of C4 (O−C=O) in the treated pulps indicated an esterification reaction between the hydroxy groups of cellulose fibers and anhydride groups of maleated polypropylene.

XPS has also been used to study the weathering of wood. Wood surfaces when exposed to natural environmental conditions, such as moisture, ultraviolet radiation and fungi, undergo discoloration and erosion. Lange et al.[20] measured the O/C ratios of European pine wood weathered at variable times after treating with CrO_3 and NaOCl solutions. The increase in O/C ratio, which signifies weathering, was found to be much slower for CrO_3-treated wood, compared with the increase in untreated wood. Kaldas et al.[21] used the C1/C2 ratio to determine the progress of the oxidation of wood treated with chromated copper arsenate (CCA) solution. The variation in this ratio suggested that oxidation of C2 (C−OH, cellulose) predominated on initial contact with CCA solution but the oxidation of C1 (C−C, lignin) became more important with extended exposure to the treating solution.

XPS has been proven to be a valuable tool for the characterization of chemical and mechanical pulps and pulps which are chemically modified to enhance certain properties. XPS could also be used as a problem solving tool. For example, first the fiber strength and performance issues could be addressed by measuring the surface lignin, extractives or other contaminants which interfere with the hydrogen bonding of fibers; and, second, washing and bleaching efficiencies of the manufacturing process could be found by measuring the enrichment of chemical contaminants on the fiber surface.

3.2 Coated Papers

Pigment coatings are applied to paper to improve printing characteristics. The coating mixture, which is commonly referred to as coating color, is applied as an aqueous slurry of three basic components: pigment, binder, and minor amounts of other additives such as lubricants and viscosity modifiers. Pigments which represent the largest portion of a dried coating, (75–90% by weight of the total solids) are naturally occurring clay and calcium carbonate. Binders are those materials which cause coating pigments to be bonded to each other and to adhere to the paper base. All coating binders are natural or synthetic organic polymers such as starch, styrene–butadiene rubber latex (SBR) or acrylic lattices. TiO_2 is often used to meet the optical properties of lightweight coated (LWC) papers. If the sample is an unknown, a survey spectrum is useful to qualify the elements in the target sample (see Figure 1). Most of the literature published has been focused on the application of XPS to measuring the concentrations of pigment and binder (P/B) in coated papers. The P/B ratios vary significantly depending upon the grade, and

also heavily influence the final properties (print, optical and physical) of the coated papers.

Fujiwara and Kline[22] described a method for determining the surface concentrations of binders and pigments in multicomponent coated papers. A series of model papers was prepared with known concentrations of SBR latex–starch binders and clay/$CaCO_3$ pigments, and analyzed by XPS. Using the atom-percent concentrations from the XPS data for calcium, silicon and high-resolution carbon, they have developed a series of numerical equations for calculating the coating composition. In their calculations they assumed that the total C1 and C2 signals were primarily due to SBR latex and starch, respectively. The method works extremely well for simple coating systems. But the P/B ratio calculations for unknowns would become complicated if the samples contained plastic pigment or if the paper is only partially covered, exposing fibers on the surface. Our experience of coating composition calculations using the method of Fujiwara and Kline has been successful in most cases but the concentrations of latex and starch have always been found to be higher than what was in the original coating color. This might be due to the presence of other coating components such as thickeners and dispersing agents and uncoated fibers which might contribute to the total C1 and C2 signals.

The studies of Tomimasu et al.[23,24] of coated papers represent a comprehensive investigation of the effects of pigment composition, particle size, surface roughness, drying temperature on the surface binder concentration, and z-directional distribution of binder. The signals from starch and latex were differentiated using fluorine-labeled latex. The binder concentrations were represented as C1/Si (or Ca) ratio. From the XPS data, they concluded that the binder/pigment ratio (a) was not affected by supercalendering, (b) was extensively affected by pigment particle size and (c) that SBR latex selectively migrated to the surface rather than starch.

Strom and Carlsson[25] showed that the elemental composition was independent of the sampling depth indicating that the coating surface is homogeneous in the z-direction within that sampling depth. The dispersion agents in the pigment slurries were expected to form a thin overlayer, but the angle-dependent work showed otherwise. The apparent area fraction or the concentration of binder decreased with increased pigment particle size. The effects that binder type and pigment level have on the surface chemical compositional changes of laboratory and pilot-coated samples were also investigated. Paper gloss increased with an increase in surface pigment (clay) concentration.

Gron and Strom[26] studied the effect of binder type on the surface chemical composition of coated papers using cationic and anionic starch. From the XPS data for nitrogen and clay (Al and Si) they concluded that cationic starch is more on the surface and covered the pigment particles more efficiently than anionic starch. Also, the C2 content (assumed to be due to starch) was twice as high on the surface coated with cationic starch than that of the surface coated with anionic starch.

Brungard and Cleland[27] reported the use of XPS to estimate the concentrations of clay, TiO_2, and $CaCO_3$ pigments in coated papers. They showed that more accurate concentrations of these pigments in coated papers can be obtained only when empirical calibration data rather than calculated peak sensitivity factors were used. Empirical calibration curves were generated from the XPS data using various but known concentrations of three pigments and binder systems. The XPS pigment-to-matrix ratio data were always lower than the theoretical ratio calculated from the component weight percents. They showed that calendering, coat weight, and the type of binder did not significantly affect the pigment concentration measurements. TiO_2 could not be quantified above 1% due to overlay of Al_2O_3 which is normally deposited on TiO_2 particles surfaces to improve dispersion.

3.2.1 Binder Migration and Print Mottle

Print mottle problems have often been attributed to selective migration of binder (starch or latex) to the surface or to the base. The process of migration is very complicated and is influenced by several factors, including coating composition, percent solids, binder type and level, base stock, and drying rate. A detailed discussion of various parameters and their effects on binder migration is beyond the scope of this article. Print mottle is often attributed to the nonuniform absorption of ink due to defects in the coated surface, resulting in areas of high and low print gloss. Areas of higher surface binder content may receive less ink than those of lower binder content resulting in print mottle. Other forms of print mottle, such as back trap mottle and water-interference mottle, can also occur as a result of transfer of ink from one blanket to the other printing blanket in a multicolor printing process. A few examples of how XPS technique was used to determine the migration behavior and its effect on print mottle are discussed. XPS has successfully been used, in some cases, to prove or disprove the migration behavior of binders in coated papers.

Engstrom, Strom and Norrdahl[28] studied the effect of coating drying process on the binder migration and on print mottle. The atomic ratio N_C/N_{Si}, where N_C represents the binder (latex + starch) and N_{Si} represents the clay concentration, was used for quantifying the surface concentration of binder. The greater the ratio, the more binder is present in the coating surface. The print mottle, as measured by the Croda print mottle test, decreased with an increase in the evaporation rate, but

the ESCA experiments showed no significant difference in N_C/N_{Si} ratio, implying that binder migration was not influenced by evaporation rate. In other experiments, in which differing coat weight papers samples were examined, mottle increased with increased N_C/N_{Si} ratio. It appears from these results that evaporation rate is less important when compared with the available water phase for evaporation for migration to occur. The binder migration increased when the time between the coating application blade and the dryer was increased. Application of fluid coating colors onto the surface of a porous paper base and their exposure to elevated temperatures in drying systems results in migration of and redistribution of the aqueous binder as long as there is sufficient mobility in the freshly applied coating color.

Arai et al.[29] established a strong correlation between binder-to-pigment ratio and print mottle. By selecting an effective XPS measurement width close to the size of the mottling pattern, XPS spectra were obtained by scanning the coated surface across the entire mottling area. The binder distribution, which was evaluated as the ratio of binder to pigment, was plotted against the scanning length. The coefficient of variation in the data was used as an indication of mottling tendency of the papers. The print mottle was found to correlate well with the binder-to-pigment ratio and was independent of z-directional distribution of coating components.

Yamazaki et al.[30] studied the effects of binder migration and latex film formability on print mottle using laboratory and pilot-coated paper samples. The time interval between the coating application and drying was varied to determine the migration tendencies of starch and latex. By comparing the surface concentration with the concentration in the coating color they showed that latex and starch migrated differently. The surface latex concentration was unaffected by the time interval. The starch, on the other hand, migrated either to the surface or to the base sheet depending upon whether the time interval was low or high, and was more sensitive to coat weight. Although they could not establish an obvious relationship between the mottling tendency and surface starch, the samples in which the starch migrated towards the coated surface showed severe mottle. Since starch migration was found to be much more sensitive to drying time interval and coat weight, they concluded that starch has more influence on print mottle.

Zang and Aspler[31] measured the print density on different model papers and showed how surface binder distribution differences affected the amount of ink transferred to the coated papers. Model samples were prepared by applying the coating colors to a commercial LWC paper base and to a nonabsorbent Mylar® film and were allowed to air dry. The idea here was that the direction of binder migration depends on the relative rates at which water is drained into the basestock and at which water evaporates from the coating surface. In the case of LWC-coated samples, water flows towards the basestock causing the binder to migrate towards the basestock. On the other hand, in the case of Mylar® film, water evaporates mainly from the surface causing the binder to migrate towards the surface. XPS results were used to support their binder migration mechanism theory. By assuming that the carbon content is a measure of latex content, they showed that the surface latex content at a given latex addition level was higher for the coatings on Mylar® than for those on LWC paper. For coatings prepared under identical formulation, the amount of ink transferred to the coatings was less on nonabsorbent Mylar® than to those on LWC paper due to differences in surface binder content.

Kim et al.[32] studied the effects of binder type, drying temperature, and solids content on back-trap mottle. XPS data was used to help quantify the degree of binder migration of several pilot coated samples under different experimental conditions. The XPS data taken at 1 mm increments provided the confirming support to the coated paper which was most resistant to mottle. The distribution of the atomic concentrations for mottle-resistant coated paper was more consistent (less variation) than other coated papers which exhibited large variations in the atomic concentration data. The high C/Ca (binder/pigment) ratios suggested high binder content for two coatings when harsh late-drying temperatures were used. These observations suggested that the binder type and drying temperatures have strong influence on the binder migration behavior and print mottle.

Sanders and Bashey[33] reported the application of XPS for solving a few common web offset printing problems, dusting, milking or glazing and scum deposits, which are directly related to paper/press interactions. Milking or glazing are typically white deposits on the printing blanket while the scum is a deposit along edge of the printing area. ESCA analyses of the clean blanket, glaze, scum, ink and paper showed that the major chemical surface constituents of the glazed area were phosphorus and calcium. Phosphorus, derived from the fountain solution, reacted with the calcium from the alkaline paper under acid conditions forming an insoluble calcium hydrogen phosphate which was deposited on the blanket. The scum (edge deposit), rich in organic nitrogen, originated from the ink. Similarly, image area deposits and the loss of print density problems of the coated papers have also been resolved using XPS.

The binder migration and print mottle of coated papers are influenced by several factors: base stock, choice of pigments, choice of binders, the drying process and percent solids. Each one of these factors has a small

or large influence, depending on the situation, on the migration behavior. However, severity of the migration and mottle behaviors are often the result of the severity of the experimental conditions and associated with the combined effect of various factors involved.

3.3 Nonimpact Printing Papers

3.3.1 Laser Printing Papers

Various functional properties of laser, xerographic, and electrophotographic papers have been discussed in the literature.[34,35] The paper designed for laser printing should be resistant to heat, pressure, and curl, and also should possess other properties such as smoothness, electrical resistivity, percent moisture for its optimal performance. One of the major attributes of electrophotographic paper is toner adhesion. Toner adhesion or fusing fix is described as how well an image is fused to the paper and its resistance towards scratching.

Unlike other conventional printing processes, the molten toner in the laser printing process does not enter into the fiber walls nor does it penetrate deep into the paper. Because the interaction of the toner is limited to the surface of the paper, the surface chemistry and the surface energetics of the paper play a major role in the print quality and toner adhesion of laser papers. XPS studies of electrophotographic papers have been used for studying toner adhesion problems and for identifying contaminants on paper surfaces.[35] Contaminants such as waxes or adhesives are detrimental to the performance of the printer.

Farrow et al.[36] studied by XPS the surface chemistry of several papers with varying pulp composition, internal size, and external size and correlated the surface hydrocarbon component (C1) with toner adhesion. Increased levels of C1 component reduced toner adhesion. Sanders et al.[37] studied the surface composition of 25 model papers, which had different levels of internal size, external size, and fillers. The model papers were printed on three different high speed copiers and the degree of fusing fix was correlated with the surface energies of the papers. The changes in the surface energies of these papers were explained in terms of the C1 concentration. Borch[38] had shown that there is considerable variation in toner adhesion depending on the additives used for sizing papers. The variation in the toner strength was attributed to the changes in the surface energetics.

XPS analyses of 20 uncoated commercial laser papers were performed in our laboratory. They exhibited a wide range of surface elemental compositions depending upon the type of sheet: acid, alkaline, sized or unsized. From these data a typical range of chemical composition for laser papers can be drawn (Table 3).

Table 3 Surface elemental composition of laser printing papers

Element	Concentration range (%)
C	59–69
O	29–39
Ca	0.2–1.3
Si	0.3–1.3
Al	0.3–1.6
Na	0.0–0.6
O/C	0.42–0.66

Calcium, silicon and aluminum originated from filler materials. Sodium might be due to common salt and/or from the surface sizing additive. It is a common practice to add sodium chloride to xerographic papers to control the electrical properties of the paper.

3.3.1.1 X-ray Photoelectron Spectroscopy Analysis of Toner Films Toner adhesion not only depends on the substrate (paper) but on the toner characteristics as well. It is therefore desirable to understand the composition and the chemistry of toners.

XPS spectra of freshly prepared molten toner films from three different manufacturers were obtained. The XPS data are shown in Table 4. From these results, it is evident that all three toners are distinctly different with respect to oxygen, carbon and silicon contents. The percentages listed in Table 4 are the surface concentrations of various elements and are not to be confused with the bulk composition which could be quite different. However, in all cases, the observed carbon was totally in the C1 form, indicating that the toner surfaces are very hydrophobic.

Most xerographic toners are low surface energy materials primarily composed of either styrene/acrylic-based, polyester-based, or pure acrylic-based polymers. High-resolution spectra of silicon exhibited a broad peak with a shoulder indicating two forms of silicon. Based on the binding energies of the silicon peaks (104.2 eV and 102.5 eV) it is assumed that the two peaks represent inorganic (SiO_2) and organic silicon. It is known that organic silicon compounds are used in toners to control the fluid properties of the toner. Inorganic SiO_2 particles are used to coat toner particles to

Table 4 Surface elemental composition of toner films

Toner	Elemental composition (%)				Ratio
	O	C	Si	Al	O/C
Toner 1	4	95		1.0	0.04
Toner 2	15	76	9		0.19
Toner 3	30	52	18		0.58

prevent toner particle–particle agglomeration. The toner chemistry is very complex and ever changing. It is not the intention of this paper to provide a detailed description but to make the reader aware of the importance of differences in toners when evaluating the papers for toner adhesion.

The surface energy of the paper should be higher than the surface energy of the toner for it to be wetted by the molten toner. Tirtaamadja et al.[39] reported surface energies of several electrophotograpic papers which ranged between 25 and 34 dyn cm^{-1}, which are well above or close to 26 dyn cm^{-1} surface energy reported for styrenated molten toner. Internal sizing agents such as alkyl ketene dimer (AKD) and alkyl succinic anhydride (ASA) have been known to reduce the surface energy of the paper substrate resulting in poor toner adhesion. Surface sizing agents such as styrene malaeic anhydride (SMA) have been reported to improve the toner adhesion due to compatibility between the toner resin and SMA. It should be noted that toner adhesion and print quality are complex phenomena which on balance are the result of reaching a desirable compromise of many paper properties including the surface chemistry. A preliminary investigation of toner adhesion tests for about 11 commercial papers showed a correlation with O/C ratio. As the O/C ratio is decreased, toner adhesion was found to be increased.

3.3.2 Ink Jet Printing Papers

The requirements for ink jet papers are radically different from laser papers. In ink jet printing, uniformly shaped droplets of aqueous dye solutions are jetted on to the paper to form the image. The paper structure and surface chemistry requirements for good print quality are so unconventional that new paper grades specifically designed for ink jet printing are being marketed. The three most important requirements for coated or uncoated paper for ink jet applications are:

1. rapid but controlled absorption of liquid ink to prevent smearing and feathering;
2. shallow penetration of liquid ink (lateral spread) into the base sheet to prevent show through;
3. the dye in the solutions should permanently bond to and stay on the surface.

The dye retention and the absorption of the water to a large extent are surface phenomenon whereas the lateral spread of the water depends on the bulk composition of the coating and the base stock.

The performance of various specially coated and surface-sized papers was discussed by Lyne and Aspler.[40] To obtain optimum color quality and permanence, most paper manufacturers developed new coatings and sizing additives specifically for ink jet applications. Most of these additives are patented or kept as trade secrets. The most common ink jet coatings consist of silica, calcium carbonate or clay or a combination thereof as pigments and poly(vinyl alcohol) (PVOH) or starch as binder. Poly(vinyl pyrrolidone) (PVP) is often used to improve ink receptivity.[41] In very specialized coating applications, a cationic polymer, such as quaternary ammonium (QA) salts, is used for dye fixation and improved print quality.[42] We have performed XPS analysis of several ink jet papers and the results are shown in Table 5. All paper samples analyzed were either commercially available or given to us as free samples at trade shows.

Table 5 XPS data for coated ink jet printing papers

Sample	Elemental composition (%)					Ratio	
	O	C	Si	N(T)	N(Q)	O/C	C2/C1
1	41	41	16.1	1.2	0.91	1.01	0.48
2[a]	30	63	0.3	3.9		0.47	0.83
3	37	48	11.5	2.5	1.68	0.77	0.31
4	57	20	21.8	0.8	0.18	2.79	0.66
5[a]	38	46	11.8	3.4	0.28	0.82	0.73
6	45	39	11.7	2.6	1.1	1.16	0.76
7[a]	21	71	0.2	6.1		0.30	0.64
8	31	67		0.6	0.3	0.47	0.85
9	48	31	18.5	1.7	1.3	1.55	1.09
10	46	41	12.1	1.3	0.8	1.11	1.92
11	42	41	15.0	1.8	1.5	1.03	0.55
12	56	21	22.6	0.9		2.73	0.97
13	58	16	24.1	1.2		3.55	0.68
14	25	66	7.8	1.3		0.38	0.20
15	33	59	5.8	2.3		0.57	0.72
16	56	21	22.2	0.8		2.64	0.95
17	51	25	21.5	2.2		2.03	1.21
18	42	39	16.9	1.6		1.07	0.69
19[a]	47	35	15.8	1.5		1.34	1.13

[a] Also contained low levels of calcium. Numbers might not add to 100% due to rounding of numbers and omission of elements of low concentrations. N(T) = total nitrogen, N(Q) = contribution from QA salt to the total nitrogen.

Most of the uncoated papers, both acid or alkali, have been claimed for multiuse, i.e. for copy, laser and ink jet printing. About 23 such paper samples were analyzed and the composition was found to be in the ranges of O, 32–39%; C, 58–68%; and minute amounts of filler materials, either clay or calcium carbonate. The O/C and C2/C1 ratios of uncoated samples were in the ranges of 0.4–0.6 and 0.7–2.0, respectively.

The coated papers or papers with special treatment specifically designed for ink jet applications exhibited complex surface chemistry. Although each sample is distinctly different with respect to the concentrations of each element, all of them contained high levels of silicon and nitrogen. Silica pigment, owing to its high-surface area and high-internal pore volume is commonly used

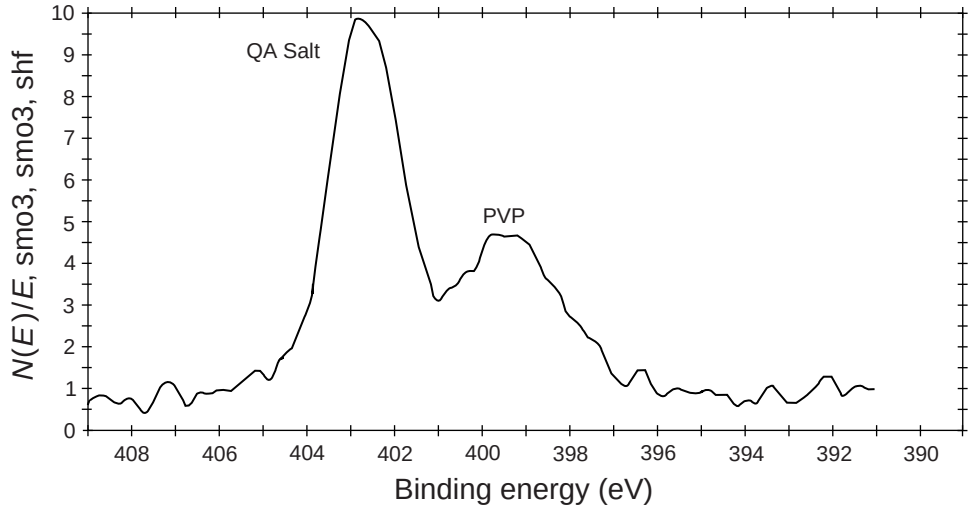

Figure 3 XPS multiplex spectrum of nitrogen (1s) from a coated ink jet printing paper. The raw data was smoothed (smo3) and shifted (shf) to the carbon reference peak at 285.0 eV.

as a hydrosink for the absorption of water in the ink. Some coated papers exhibited two peaks for nitrogen, 399.5 eV and 402.5 eV as shown in Figure 3. Comparing the binding energies for standard reference materials with the binding energies for the samples, we theorized that the peaks for nitrogen at 399.5 eV and 402.5 eV may have been respectively due to nitrogen in PVP and QA salt. The XPS spectrum alone can not identify these compounds. Based on this assumption the respective amounts for the total nitrogen and QA salt were calculated.

The performance of ink jet paper depends on several factors including the type of silica, types of PVP, PVOH and any other ingredient that might be present in the coating formulation. New coating formulations are constantly being developed to improve and/or add certain attributes to the paper quality. The purpose of this analysis is only to demonstrate how XPS techniques can be used for analyzing the surface chemistry of ink jet papers. It should be pointed out that knowledge of the surface composition, which could be different from the bulk composition, would not alone be sufficient to develop paper coating with superior performance.

3.4 Adhesion

XPS has been proven to be the most valuable tool for addressing adhesion-related problems. The XPS method provides distinctive signatures of any contaminants on the surface which might interfere with adhesion. In the food packaging industry, good adhesion between the board and the plastic is important and critical for the functioning of the package. Milk, juice cartons, and paper coffee cups are made of polyethylene extruded board. In order for the plastic to adhere, the surface energy of the board should be higher than that of the plastic. Wet strength resins and internal sizing agents are added to the basestock to impart wet strength to the food board. Like the plastics, these are low surface energy materials which reduce the surface energy of the board if present excessively on the surface. This reduced surface energy results in poor bonding between the board and polyethylene. The weakly adhered polyethylene would ultimately be delaminated due to mechanical forces during carton manufacturing, printing or when exposed to liquids. Critical reviews of other factors affecting the adhesion of polymers to the surface are available in the literature.

The capability of XPS to deal with adhesion problems was demonstrated by Istone.[35] He showed that the C1 component of the problem board was significantly more than the good board causing the surface energy of the problem board to be lower. Wet strength was ruled out as the source of the contamination since the resin used was a polyamide and no nitrogen was observed in the XPS spectrum. AKD, which was used as an internal sizing agent, has very low surface energy (18 dyn cm^{-1}) and was found to be the source of C1 contamination of the surface.

Westerlind, Larsson and Rigdahl[43] evaluated the degree of adhesion between greaseproof paper and low-density polyethylene (LDPE) using XPS. The paper and LDPE samples were plasma treated in the presence of ammonia and oxygen. The XPS analysis of the untreated and plasma-treated samples showed a dramatic increase of oxygen amounts both on the greaseproof paper and LDPE. The significant improvement in the adhesion between the plasma-treated paper surface and LDPE was explained by the incorporation of polar oxygen and nitrogen-containing groups. These polar groups increase

the surface energy and wetting properties of the paper surface causing the polyethylene melt to penetrate into the fiber network.

Nussbaum[44] reported the use of XPS to evaluate the glue adhesion property of flame-treated wood surfaces. The O/C ratio was found to increase from 0.24 for untreated pine wood to 0.57 for flame-treated pine wood. The high-resolution C (1s) spectra showed a dramatic increase in the C2: hydroxy or ether and C3: carbonyl groups. Although flame treatment led to a remarkable increase in the wettability and surface oxidation level of the flame-treated wood, no distinguishable improvement was observed with the glue, indicating that some other adhesion mechanism was operative.

Carlsson and Strom[9] measured the surface composition of hydrogen and oxygen plasma-treated filter paper and greaseproof paper by XPS. They showed that surface chemistry was very important for good adhesion between these papers and polyethylene. The changes in the adhesive strength between modified and unmodified papers and polyethylene were discussed in terms of the changes in the O/C ratios of the papers. A small decrease in the O/C ratio produced a large decrease in adhesion.

XPS can also be applied to the study of adhesives. Often the adhesion failure between the paper and adhesive might be due to the incompatibility and/or poisoning of adhesive resins. Brewis et al.[45] studied the loss of adhesive strength of adhesives treated with zirconium solutions which are added in the recycling industry to make stickies less tacky. XPS analysis of adhesives treated with zirconium acetate and ammonium zirconium carbonate solutions showed that zirconium was absorbed in all adhesives tested. Ammonium zirconium carbonate has been used in the paper industry for several years as a sizing agent in certain applications. XPS analysis of envelopes in our laboratory with poor and good adhesion showed the presence of zirconium on the surface of the envelopes with poor adhesion. Zirconium has a strong affinity for oxygenated species and readily interacts with the functional groups of the adhesive. It may be that the zirconium on the problem paper surface readily attached to the adhesive reducing its adhesive strength by blocking off the functional groups which otherwise bond to the fibers on the surface.

Hemminger[46] used ESCA to study the extent of wax bloom to the surface of latex-coated papers. Wax is used in top coat latex formulations to improve antiblocking and slip properties. Podhajny[47] suggested a strategy which involved the use of ESCA when FTIR failed to identify the cause of poor heat seal problem with paper/poly/foil/polyethylene base stock.

3.5 Paper Additives

3.5.1 Surface Size Analysis

One of the important properties of paper is its resistance to wetting. Paper, if untreated, absorbs water because of its high porosity and hydrophilic nature due to the presence of hydroxy groups of cellulose and hemicellulose. To impart water resistance, most commodity papers are manufactured with internal sizing agents, hydrophobic low-surface energy materials which are added to the papermaking stock before the sheet is formed. These internal sizing agents, which must attach to the fibers through the use of an appropriate functional group, present a hydrophobic surface on the fiber. The major internal sizing agents used commercially are fortified rosin, ASA and AKD. XPS has been used to study the migration of these sizes to the surface and their adverse effects on adhesion.[35]

Surface sizing agents, on the other hand, are applied as a solution to the surface of the paper after it has been formed. Papers that are to be surface sized often already contain internal size to regulate penetration of the surface-sizing materials into the sheet structure. The primary purpose of the surface size is to improve the print quality although several other benefits can be achieved. Surface sizes can either be hydrophilic (such as starch, carboxymethyl cellulose, PVOH) or hydrophobic (such as SMA, polyurethane or styrene acrylic polymer). The hydrophilic sizes improve water resistance by filling the voids and sealing the sheet which reduces the absorption of water by capillary action. However, the nonpolar synthetic polymers form a film on the surface of the paper. The film minimizes water absorption, improves ink holdout and improves surface picking resistance.

XPS can be readily used for detecting low levels of surface size on paper. Brinen et al.[48] reported quantitative determination of *N*,*N*,*O*-tris(octadecylcarbamoyl)hydroxylamine (HABI), an alkaline size, on paper surfaces by XPS. As the amount of HABI applied increased, the amount of C1 present on the surface was increased. A linear relationship was established between the C1/O and N/O ratios and the amount of HABI applied. More recently, Brinen[49] reported a similar increase of the C1 component of papers sized with internal and external sizing agents.

Several other reports exist in the literature showing the distribution of sizing agents by time-of-flight secondary ion mass spectrometry (TOFSIMS).[50] In our laboratory, we have also studied the XPS analysis of papers sized with three different synthetic polymers and the results are shown in Table 6. Because of the application of these sizing additives at very low levels, complete coverage of the paper surface is not expected. As a result, the sized papers exhibited XPS survey spectra very similar to the

Table 6 Surface elemental composition of surface-sized papers and reference sizing materials[a]

Paper sized with	Elemental composition (%)				Carbon chemistry (%)				Ratio	
	O	N	Ca	C	C1	C2	C3	C4	O/C	C1(f)[b]
Starch	42	0.4	0.7	57	17	65	17		0.73	9.7
SAP	37		0.7	62	32	52	16	0.3	0.60	19.8
AAL	36	0.4	1.1	63	39	47	14	0.3	0.57	24.6
SMA	34		1.1	65	44	44	12	0.6	0.52	28.6
SAP[b]	14			86	32	52	16	0.3	0.16	27.5
AAL[b]	15			85	39	47	14	0.3	0.18	33.1
SMA[b]	24			76	44	44	12	0.6	0.31	33.5

SAP, styrene acrylic polymer; AAL, alkyl acrylic latex.

[a] All XPS spectra and results presented in this chapter were generated by the author on the Perkin-Elmer PHI-5500 Multi-Technique System at Champion International Corporation in West Nyack, NY. All experiments were performed using monochromatic aluminum X-rays ($K_\alpha = 1486.7$ eV) at 400 W (15 keV). The survey spectra were acquired at a pass energy of 93.9 eV whereas the multiplex spectra were acquired at a pass energy of 58.7 eV. The analysis area was about 800 μm in diameter. High-resolution carbon spectra were obtained at pass energies of either 58.7 eV or 5.87 eV. Curve fitting and quantitation were performed using routines provided by the PHI instrument software. In all cases the experiments were performed on three different sample locations and an average was calculated.

[b] Reference material, applied as thin film on the paper and dried at 60 °C. C1(f) = Total C × (C1/100).

survey spectra of the unsized papers. Complete coverage of the paper should give a spectrum and O/C ratios similar to the reference standard.

The differences observed in the O/C ratios of papers sized with synthetic sizes were only minimal and did not directly correlate with the O/C ratios of the synthetic polymer reference standards. Since the synthetic sizing agents are generally carbon-rich, the C1 component of the surface-sized paper is increased relative to the starch-sized papers as shown in Table 6. The C1 fraction (i.e. percent of total carbon present as C1) increased with the increase in the C1 fraction of the standard reference materials.

3.5.2 Coefficient of Friction

The application of XPS for studying the frictional properties of papers is rather limited in the literature. Frictional properties of paper, measured as static or dynamic coefficient of friction (COF), play a major role in the use and handling of paper. The frictional properties for each grade of paper should be optimized; if not, the paper either slips or cause excessive wear on the machines in high-speed converting and/or printing operations. A good account of various factors affecting the COF of paper and application of XPS for studying COF is given by Gurnagul et al.[51] They found a linear relationship between the surface O/C ratio and static COF for newsprint samples. Increased levels of carbon (low O/C ratio) decreased the static friction between two papers. The differences in the amounts of surface carbon were attributed to the extractives present on the surface of the newsprint. The report also showed that corona discharge treatment of the newsprint samples showed a 30% increase in the COF due to increase in the surface oxidation of extractives on the surface. The O/C ratio increased from 0.39 for untreated to 0.47 for treated samples.

XPS and COF data were obtained in our laboratory for several commercial xerographic papers. All the samples were analyzed as received and consisted of both acid, alkaline, US and European sheets. The O/C ratio for each sheet was calculated and plotted against the static (paper against paper) COF in Figure 4. Even though there is a lot of scatter in the data, the trend, i.e. decrease in COF with the decrease in O/C ratio, is readily apparent. A linear relationship is not expected since several factors influence the COF; fiber, surface additives, surface extractives, pigments, filler type, and filler levels. Other factors which influence the COF are internal sizing agents such as ASA and AKD.

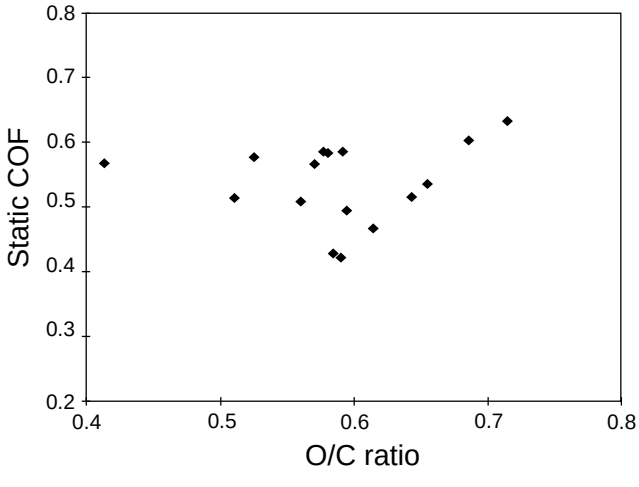

Figure 4 Plot of static COF as a function of O/C atomic ratio.

XPS will be a valuable tool to determine and optimize the variables affecting COF. For example, XPS can be used to determine a threshold level for an internal sizing agent, or wet strength resin, for a given grade of paper, filler and fiber, before COF would become unacceptable. However, surface chemistry and chemical composition cannot alone define the frictional properties of paper. For example, supercalendering of coated papers changes the frictional properties without adversely changing the surface chemistry.

ACKNOWLEDGMENTS

The author wishes to express his deep sense of gratitude and appreciation to the senior management (Dr G. Closset) of the Champion International Corporation, a leading manufacturer of paper and forest products, and to the management (Drs F. Antonucci, F. Renk and R. Thorman) of Applied Technologies (Technology arm of the Champion International) for time and permission to publish this article. The author also expresses his sincere thanks to Dr W. Istone (also of Champion International) for his valuable advice and encouragement over the years. Thanks are also due to Mr R. Stewart for reviewing the documents. The author would like to thank his wife, Anjula, for her love and patience.

ABBREVIATIONS AND ACRONYMS

AKD	Alkyl Ketene Dimer
ASA	Alkyl Succinic Anhydride
BCTMP	Bleached Chemithermomechanical Pulp
BE	Binding Energy
CCA	Chromated Copper Arsenate
CMP	Chemimechanical Pulp
COF	Coefficient of Friction
CTMP	Chemithermomechanical Pulp
ESCA	Electron Spectroscopy for Chemical Analysis
FTIR	Fourier Transform Infrared
HABI	*N*,*N*,*O*-tris(octadecylcarbamoyl)-hydroxylamine
KE	Kinetic Energy
LDPE	Low-density Polyethylene
LWC	Lightweight Coated
P/B	Pigment and Binder
PVOH	Poly(vinyl alcohol)
PVP	Poly(vinyl pyrrolidone)
QA	Quaternary Ammonium
SBR	Styrene–Butadiene Rubber Latex
SIMS	Secondary Ion Mass Spectrometry
SMA	Styrene Malaeic Anhydride
TOFSIMS	Time-of-flight Secondary Ion Mass Spectrometry
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Surfaces **(Volume 10)**
Surfaces: Introduction • Auger Electron Spectroscopy in Analysis of Surfaces • Ion Scattering Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

REFERENCES

1. G.M. Dorris, D.G. Gray, 'The Surface Analysis of Paper and Wood Fibers By ESCA-I', *Cellul. Chem. Technol.*, **12**, 9–23 (1978).
2. G.M. Dorris, D.G. Gray, 'The Surface Analysis of Paper and Wood Fibers by ESCA, II Surface Composition of Mechanical Pulps', *Cellul. Chem. Technol.*, **12**, 721–734 (1978).
3. D.G. Gray, 'The Surface Analysis of Paper and Wood Fibers, III. Interpretation of Carbon (1s) Peak Shape', *Cellul. Chem. Technol.*, **12**, 735–743 (1978).
4. L.A. Casper, C.J. Powell (eds.), *Industrial Applications of Surface Analysis*, ACS Symposium Series 199, American Chemical Society, Washington, DC, 1982.
5. C.R. Brundle, C.A. Evans, Jr, S. Wilson (eds.), *Encyclopedia of Material Characterization*, Butterworth-Heinemann, Boston, 1992.
6. W.K. Istone, 'X-Ray Photoelectron Spectroscopy', in *Surface Analysis of Paper*, eds. T.E. Conners, S. Banerjee, CRC Press, 235–268, 1995.
7. J. Buchert, G. Carlsson, L. Viikari, G. Strom, 'Surface Characterization of Unbleached Kraft Pulps by Enzymatic Peeling and ESCA', *Holzforschung*, **50**(1), 69–74 (1996).
8. A. Ahmed, A. Adnot, J.L. Grandmaison, S. Kaliaguine, J. Doucet, 'ESCA Analysis of Cellulosic Materials', *Cellul. Chem. Technol.*, **21**, 483–492 (1987).
9. M.G. Carlsson, G. Strom, 'Adhesion between Plasma Treated Cellulosic Materials and Polyethylene', *Surf. Interface Anal.*, **17**, 511–515 (1991).
10. X. Hua, S. Kaliaguine, B.V. Kokta, A. Adnot, 'Surface Analysis of Explosion Pulps by ESCA. Part 1. Carbon

(1s) Spectra and Oxygen-to-Carbon Ratios', *Wood Sci. Technol.*, **27**, 449–459 (1993).
11. X. Hua, S. Kaliaguine, B.V. Kokta, A. Adnot, 'Surface Analysis of Explosion Pulps by ESCA. Part 2. Oxygen (1s) and Sulfur (2p) Spectra', *Wood Sci.Technol.*, **28**, 1–8 (1993).
12. J. Laine, P. Stenius, G. Carlsson, G. Strom, 'Surface Characterization of Unbleached Kraft pulps by means of ESCA', *Cellulose*, **1**, 145–160 (1994).
13. J. Laine, P. Stenius, 'The Effect of ECF and TCF Bleaching on the Surface Chemical Composition of Kraft Pulp as Determined by ESCA', *Nordic Pulp Pap. Res. J.*, **11**(3), 201–210 (1996).
14. A. Koubaa, B. Riedl, Z. Koran, 'Surface Analysis of Press Dried-CTMP Paper Samples by ESCA', *J. Appl. Polym. Sci.*, **61**, 545–552 (1996).
15. D.P. Kamdem, B. Riedl, A. Adnot, S. Kaliaguine, 'ESCA Spectroscopy of Poly(methyl methacrylate) Grafted onto Wood Fibers', *J. Appl. Polym. Sci.*, **43**, 1901–1912 (1991).
16. J. Felix, P. Gatenholm, H.P. Schreiber, 'Plasma Modification of Cellulose Fibers: Effects on Some Polymer Composite Properties', *J. Appl. Polym. Sci.*, **51**, 285–295 (1994).
17. M.N. Belgacem, G. Czeremuszkin, S. Sapieha, A. Gandini, 'Surface Characterization of Cellulose Fibers by XPS and Inverse Gas Chromatography', *Cellulose*, **2**, 145–157 (1995).
18. P. Zadorecki, P. Flodin, 'Surface Modification of Cellulose Fibers. I. Spectroscopic Characterization of Surface Modified Cellulose Fibers and Their Copolymerization with Styrene', *J. Appl. Polym. Sci.*, **30**, 2419–2429 (1985).
19. M. Kazayawoko, J.J. Balatinecz, R.T. Woodhams, R.N. Sodhi, 'XPS of Lignocellulosic Materials Treated with Maleated Polypropylenes', *J. Wood Chem. Technol.*, **18**(1), 1–26 (1998).
20. P.J. de Lange, A.K. de Kreek, A. van Linden, N.J. Coenjaarts, 'Weathering of Wood and Protection by Chromium Studied by XPS', *Surf. Interface Anal.*, **19**(1–12), 397–402 (1992).
21. M.L. Kaldas, P.A. Cooper, R. Sodhi, 'Oxidation of Wood Components during Chromated Copper Arsenate (CCA-C) Fixation', *J. Wood Chem. Technol.*, **18**(1), 53–67 (1988).
22. H. Fujiwara, J.E. Kline, 'ESCA Analysis of Pigment Paper', in *International Process and Materials Quality Evaluation Conference Proceedings*, TAPPI Press, Atlanta, GA, 157–167, 1986.
23. H. Tomimasu, S. Ogawa, Y. Sakai, T. Yamasaki, T. Ogura, 'ESCA to Analyze Surface Binder Concentration of Coated Paper', in *TAPPI Coating Conference Proceedings*, TAPPI Press, Atlanta, GA, 35–43, 1986.
24. T. Arai, H. Tomimasu, 'Measuring Method of Chemical Distribution in Coated Layer by ESCA', *Jpn. TAPPI*, **43**(2), 1–16 (1989).
25. G. Strom, G. Carlsson, 'Chemical Composition of Coated Paper Surfaces Determined by means of ESCA', *Nordic Pulp Pap. Res. J.*, **8**(1), 105–112 (1993).
26. J. Gron, G. Strom, 'Influence of Starch Modification on the Chemical Composition and Structure of Coated Layers', *Nordic Pulp Pap. Res. J.*, **13**(2), 119–123 (1998).
27. N.L. Brungard, A.J. Cleland, 'Using XPS to Estimate Relative Pigment Concentrations on Coated Paper', in *TAPPI Coating Conference Proceedings*, TAPPI Press, Atlanta, GA, 13–22, 1994.
28. G. Engstrom, G. Strom, P. Norrdahl, 'Studies Of The Drying And Its Effect On Binder Migration And Offset Printing', *TAPPI Coating Conference Proceedings*, 35–43, 1987 and *Tappi J.*, December, 45–49, 1987.
29. T. Arai, T. Yamasaki, K. Suzuki, T. Ogura, Y. Sakai, 'The Relationship Between Print Mottle and Coating Structure', *Tappi J.*, May, 47–52 (1988).
30. K. Yamazaki, Y. Hattori, T. Nishioka, K. Fujita, 'Study on Print Mottle Effect of Binder Migration and Latex Film Formability During Consolidation Process', in *TAPPI Coating Conference Proceedings*, TAPPI Press, Atlanta, GA, 17–25, 1993.
31. Y.-H. Zang, J.S. Aspler, 'The Effect of Surface Binder Content on Print Density and Ink Receptivity of Coated Paper', in *Advanced Coating Fundamentals Symposium Proceedings*, TAPPI Press, Atlanta, GA, 165–170, 1997.
32. L. Kim, E. Wittbrodt, D. Smith, M. Rolf, B. Langolf, M. Pollock, J. Ropr, J. Stolarz, T.J. Green, 'Reduction of Black-trap Mottle Through Optimization of the Drying Process for Paper Coatings, Part I', in *TAPPI Coating Conference Proceedings*, TAPPI Press, Atlanta, GA, 111–125, 1997.
33. N.D. Sanders, A.R. Bashey, 'Application of UHV Surface Analysis to Understanding Paper/Press Interaction Problems', in *International Printing and Graphic Arts Conference Proceedings*, Canadian Pulp and Paper Association, Montreal, 57–65, 1994.
34. I.L. Seldin, 'Throughput Requirements For Xerography', in *First International Symposium on Competing in the Business Papers Market of the Future Proceedings, IV, 1–11 May*, Miller Freeman, New York, 1985.
35. W.K. Istone, 'SIMS and XPS for the Characterization of Surface Defects in Paper Products', *J.Vac. Sci. Technol. A*, **12**(4), 2515–2522 (1994).
36. M.M. Farrow, A.G. Miller, A.M. Walsh, 'Surface Chemistry of Business Papers: ESCA Studies', *Colloids and Surfaces in Reprographic Technology*, ACS Symposium Series 200, 455–474, American Chemical Society, Washington, D.C., 1982.
37. D.J. Sanders, D.F. Rutland, W.K. Istone, 'The Effect of Paper Properties on Fusing Fix', *J. Imaging Sci. Technol.*, **40**(2), 175–179 (1996).
38. J. Borch, 'Sizing and its Effect on Paper–Polymer Adhesion', *Pap. Technol. Ind.*, 388–395 (1985–1986).
39. Tirtaatmadja, I.H. Parker, K. Kuys, 'Surface Energetics of Copy Paper', *Appita*, 419–426 (1995).

40. M.B. Lyne, J.S. Aspler, 'Paper for Ink Jet Printing', *Tappi J.*, 106–110 (1985).
41. D.R. Titterington, C.W. Jaeger, H.P. Le, 'Development of Transparent Receiving Sheet for Ink Jet Printing', *J. Imaging Sci.*, **30**(1), 4–9 (1985).
42. D.W. Donigian, P.C. Wernett, M.G. McFadden, J.J. McKay, 'Ink Jet Dye Fixation and Coating Pigments', in *Coating/Papermakers Conference Proceedings*, TAPPI Press, Atlanta, GA, 393–409, 1998.
43. B. Westerlind, A. Larsson, M. Rigdahl, 'Determination of the Degree of Adhesion in Plasma-Treated Polyethylene/Paper Laminates', *Int. J. Adhesion Adhesives*, **7**(3), 141–146 (1987).
44. M. Nussbaum, 'Oxidative Activation of Wood Surfaces by Flame Treatment', *Wood Sci. Technol.*, **27**, 183–193 (1993).
45. D.M. Brewis, J. Comyn, J.L. Tegg, R.H. Duncan, I. McAlpine, P.J. Moles, B. Smith, J. Treverton, 'The Reduction in Contact Adhesion of Some Adhesives for Paper by Treatment with Zirconium Chemicals', *J. Adhesion*, **26**, 215–235 (1988).
46. C. Hemminger, 'ESCA Analysis as a Tool to Study Surfaces of Treated Paper', in *TAPPI Coating Conference Proceedings*, TAPPI Press, Atlanta, GA, 55–61, 1985.
47. R.M. Podhajny, 'How Can I Identify What is Causing Poor Heat-seal?', *Converting Magazine*, **14**(10), 20 (1996).
48. J.S. Brinen, J.C. Calbick, R.D. Cody, 'Quantitative Determination of Alkaline Sizes on Paper Surfaces by ESCA. Part I. HABI', *Surf. Interface Anal.*, **14**, 245–249 (1989).
49. J.S. Brinen, 'The Observations and Distribution of Organic Additives on Paper Surfaces using Surface Spectroscopic Techniques', *Nordic Pulp Pap. Res. J.*, **8**(1), 123–129 (1993).
50. J. Ozaki, 'Surface Characterization of Rosin Sizing Agent in Paper by Means of EPMA, ESCA and TOFSIMS', *Nordic Pulp Pap. Res. J.*, **12**(4), 260–266 (1997).
51. N. Gurnagul, J.T. Wearing, M.D. Ouchi, D.G. Sparkes, N. Dunlop-Jones, 'Factors Affecting the Coefficient of Friction of Paper', Pulp Paper Research Institute Canada, Pulp and Paper Report, PPR 871, 1991.

Remote Sensing: Introduction

Paul M. Mather
The University of Nottingham, Nottingham, UK

Analytical chemistry is concerned with instrumental and laboratory measurements that are made in the context of particular problems or tests. One of the procedures employed in analytical chemistry is spectroscopy, which is used to measure the degree of absorption and scattering of electromagnetic radiation impinging on an object, or the characteristics of electromagnetic radiation emitted by an object, in order to identify and infer the properties of that object. This same procedure forms the basis of remote sensing of the Earth's surface from aircraft and satellites. Furthermore, imaging procedures such as scanning electron microscopy (SEM) are widely used by analytical chemists; the link with imaging of the Earth's surface by remote sensing is even clearer. Thus, the principles of interpretation of remotely sensed images are, consciously or otherwise, similar to those used by analytical chemists.

Digital images of the Earth's surface taken from space have been available for more than 30 years. Early civilian applications of this space imagery were in the fields of weather forecasting and meteorology. Photographs taken by astronauts during the Apollo 9 mission in 1968 demonstrated that a wealth of cartographic, ecological and geological information could be derived from images of the Earth taken from orbital altitudes. Until this time, the main source of spatial information used by Earth scientists, cartographers and others was the air photograph. Space photographs from Apollo also showed the benefits of synoptic or broad-scale views of the Earth available from altitudes of several hundred kilometers.

The National Aeronautics and Space Administration (NASA), with the support of the US Department of Agriculture and the US Geological Survey, designed and built the first automatic Earth satellite system for terrestrial observations during the late 1960s. Called ERTS-1 (Earth Resources Technology Satellite-1), this satellite was launched successfully in 1972. It carried a scanning instrument called the MultiSpectral Scanner (MSS), which collected digital images of the Earth's surface in four spectral wavebands (green, red, and two near infrared (NIR) bands) for an area 92.5 km on either side of the sub-satellite track. Its orbit was such that ERTS-1 could survey all of the Earth's land and sea surfaces between latitudes 82 °N and 82 °S over a period of 18 days. In order to maintain consistency from one observation of a given point to the next, ERTS-1 had a near-circular, sun-synchronous orbit with a daytime equatorial crossing time of 0945 local time. On the launch of the second ERTS in 1972 the programme was renamed ''Landsat'' and, to date, seven Landsats have been launched. Landsats 1–3 were similar in design to ERTS-1, but Landsat-4 and Landsat-5 carried a more advanced sensor called the Thematic Mapper (TM). This instrument collects images in seven wavebands, from the blue-green end of the optical spectrum to the middle infrared. The spatial resolution (as measured by pixel size on the ground) of the TM instrument is 30 m compared with the 80-m resolution of the MSS. Landsat-6 failed to reach orbit, but Landsat-7 was launched successfully in 1999. It carries the Enhanced TM Plus (ETM+) instrument, which includes a panchromatic channel with a ground resolution of 15 m.

Landsat MSS images were intended primarily for use in agricultural crop monitoring and classification. However, it soon became clear that information relevant to a range of applications could be derived from MSS imagery. For example, Figure 1 shows a Landsat MSS image of the area around Phnom Penh, Cambodia, taken on 3 January 1973. This image is a color composite, in which ground-reflected solar energy in the spectral waveband 0.8–1.1 μm (NIR) is shown in red, ground reflection in the 0.6–0.7 μm (red) waveband is shown in green, and reflection in the 0.5–0.6 μm (green) waveband is shown in blue. Water bodies have low reflectance in the NIR band, with progressively higher reflectance in the red and green bands, depending upon the amount of sediment present in the water. So, for example, the turquoise band running from the top to the bottom down the centre of the image is a turbid river. The black linear feature running across the left side of the image is a river (or canal) of relatively clear water, and the separation between the clear and turbid water is clearly apparent in the lower left corner, where the two water features coalesce. Red areas are vegetated, as the degree of vegetation vigor is shown by the magnitude of the reflected energy in the NIR waveband.

Figure 2 shows the area around the city of Las Vegas, Nevada, in the same color composite format as Figure 1. In comparison with the Phnom Penh image, there is relatively little vegetation (red color) or water (black or turquoise color), though the layout of the city is clear, and the topographic and geological structures in the area on the right of the image centre are clear. Alluvial fans – composed of sedimentary debris laid down by ephemeral streams – are visible on the left side of the image.

Advances in sensor design over the past 30 years have produced significant improvements in the quality of the images returned to Earth from orbiting satellites.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

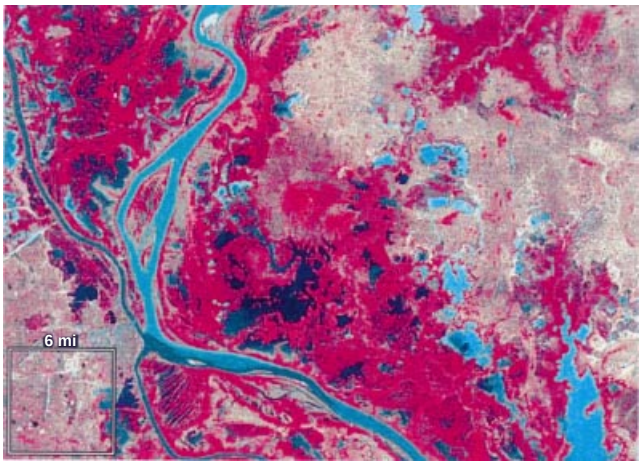

Figure 1 ERTS-1 MSS image of Phnom Penh, Cambodia, taken on 3 January 1973. Black areas are clear water. Red areas are vigorous vegetation. Sediment-laden streams appear as turquoise. This image is a composite, showing reflection in the 0.8–1.1 μm waveband in red, reflection in the 0.6–0.7 μm waveband in green and reflection in the 0.5–0.6 μm waveband in blue. (Source: http://edcwww.cr.usgs.gov/17dhf/L7MMO/gallery.htm)

Figure 2 Landsat MSS image of the Las Vegas area, acquired 13 September 1972. (Source: http://edc.usgs.gov/earthshots/slow/LasVegas/LasVegas)

Perhaps the most immediately impressive development has been in the spatial resolution of imaging sensors. Figure 2 shows the city of Las Vegas, Nevada, as seen by Landsat's MSS. The first commercial satellite, IKONOS, was launched in September 1999 by the American Space Imaging Corporation. Figure 3 shows a portion of the first image received from IKONOS, covering the area around the Jefferson Memorial in Washington, DC. The spatial resolution of this image is 1 m. The IKONOS satellite also collects multispectral imagery with a resolution of 4 m. While the Landsat

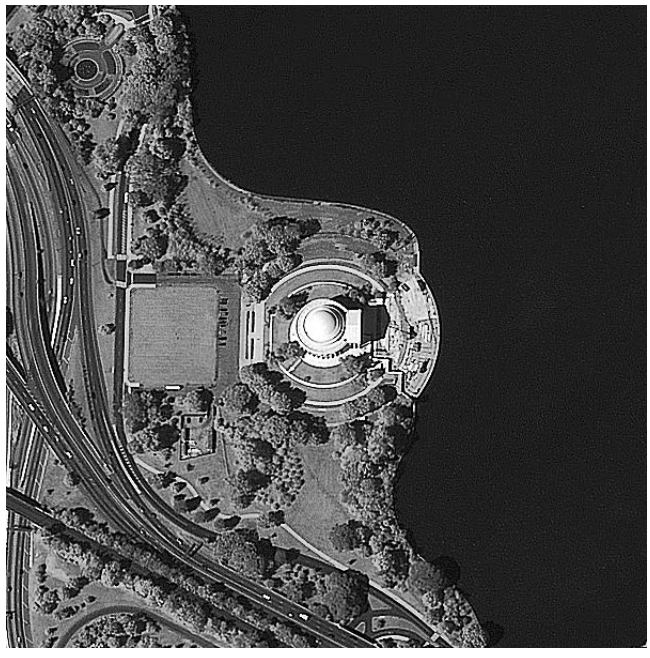

Figure 3 The Jefferson Memorial, Washington, DC. Image collected by Space Imaging's Ikonos satellite on 30 September 1999. This panchromatic image has a spatial resolution of 1 m. The satellite's orbital height is 680 km. (For more details, see http://www.spaceimaging.com/products/ikonos.html)

MSS image of Las Vegas gives us a broad picture of the city and its region, providing information on both the physical landscape and the density of the vegetation cover, the IKONOS image provides a level of spatial detail that was previously associated with aerial photography. The images of Death Valley shown in Figure 6 also demonstrate the effects of spatial resolution on image interpretation and use. The spatial resolution of the images shown in Figure 6(a–c) are 15 m, 30 m, and 90 m, respectively.

Quality is not measured solely in terms of spatial resolution. The term more properly refers to the appropriateness to the intended application of the geographical scale, plus the spectral, radiometric and temporal properties of the source of imagery. Spectral resolution refers to the number and location of the spectral wavebands in which data are collected by the instrument. The precision of the measurements (radiometric resolution) is reflected in the number of levels used in quantizing the signal (Landsat MSS had 64 levels, while the newer TM instrument measures reflectance on a 0–255 scale). Temporal resolution refers to the time between repeat coverage of a given area. The repeat cycle of an orbiting satellite is related primarily to the characteristics of the orbit, which – for most Earth observing satellites – is near-circular and polar (the exceptions being

the geostationary meteorological satellites such as GOES and Meteosat). However, some instruments such as the high resolution visible (HRV) sensor carried by the French/Belgian/Swedish SPOT satellites can be moved laterally through ±30°. This capability allows the instrument to be pointed off-nadir and thus collect information for a specified target at a higher temporal frequency than would be the case for a nadir-pointing instrument. The SPOT HRV's pointing capability enables the system to collect stereoscopic imagery, because the target can be observed from first one side and then the other on adjacent orbits.[1] Further details can be found in the article **Satellite and Sensor Systems for Environmental Monitoring**.

The articles in this section deal with applications that cover a range of spatial scales, from global to local. Each of these applications requires data at a specific spatial scale, with measurements in a particular spectral band or bands, and with an appropriate temporal frequency. A number of operational and experimental Earth satellite systems are available to meet these different requirements. For example, the article **Biological Oceanography by Remote Sensing** considers the way in which information from satellite-borne scanners is used to assess the abundance of phytoplankton and the concentration of dissolved solids and sediment in the upper layers of the ocean. The biological productivity in the oceans, the optical properties of ocean water, the interaction of winds and currents with ocean biology, and the way in which human activities influence the oceans can be investigated using data from these scanners. An example is the Coastal Zone Color Scanner (CZCS) operated between October 1978 and June 1986, on board the Nimbus 7 spacecraft. The CZCS instrument was specifically designed to measure the main absorption bands of phytoplankton and algae, as shown in Figure 4. The use of satellite remote sensing to study and monitor sea ice is a second important practical application in the field of oceanography. It is covered in **Sea Ice Monitoring by Remote Sensing**.

As noted above, the properties of remote sensing instruments such as the Landsat TM or the SPOT HRV sensors can be summarized in terms of their spatial resolution (the resolving power of the instrument), spectral resolution (the number and width of wavebands in which data are collected), radiometric resolution (the signal-to-noise ratio) and the temporal resolution. Each instrument is designed for one or several related applications, and the contributions in this section of the Encyclopaedia survey specific application areas. These contributions can be divided into the following groups:

- cartography
- oceanography

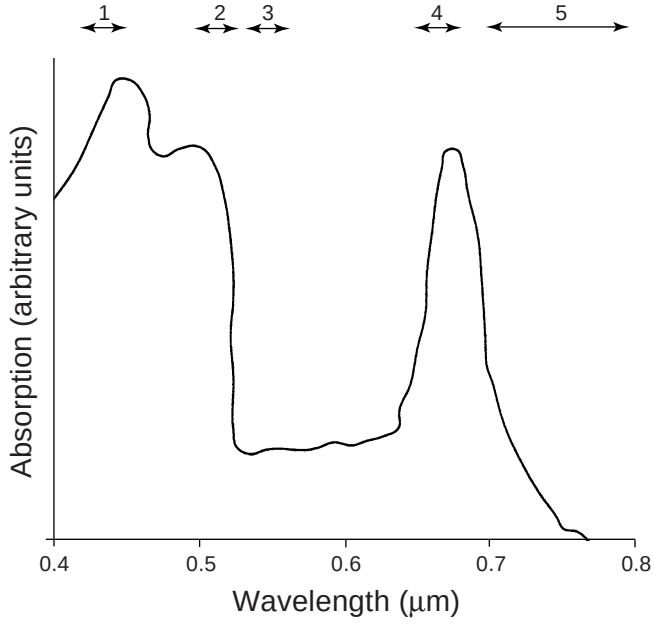

Figure 4 Idealized absorption curve for green algae, showing location of the Nimbus-7 CZCS wavebands 1–5. (Based in part on T. Alföldi,[2] Figure 6. Reproduced by permission of the *Journal of Soil and Water Conservation*, Ankeny, USA.)

- geology and mineral exploration
- land cover monitoring
- environmental monitoring
- techniques and analytical methods.

The contributions in the section that focuses on oceanography are considered above. Cartographic applications use satellite-derived data to map the two and three-dimensional shape of the Earth's terrestrial surfaces. The use of two-dimensional satellite imagery, particularly that derived from the SPOT HRV, to collect planimetric (X, Y) information is well established. More recent developments now permit satellite-borne instruments to derive Earth surface elevations to decimeter accuracy or better. This technique is known as elevation modeling, and the product – a three-dimensional representation of the Earth's surface – is termed a digital elevation model, or DEM. This information is of great importance to Earth scientists engaged in explaining and predicting the behavior of terrestrial systems. Elevation modeling using spaceborne data is discussed in articles **Elevation Modeling and Displacement Mapping using Radar Interferometry** and **Elevation Modeling from Satellite Data**. While the SPOT HRV instrument has been widely and successfully used to generate DEMs, recent research has involved the use of interferometric synthetic aperture radars (InSAR) which are capable both of measuring surface elevations and (in differential mode) of mapping ground displacements following, for

example, earthquakes. Optical mapping instruments such as those carried by SPOT and IKONOS provide stereoscopic images similar, in principle, to those acquired by aircraft. The SPOT satellite provides stereoscopic imagery with a ground resolution of 10 m, while the IKONOS spacecraft's camera is capable of producing imagery that has a locational accuracy of 2 m horizontally and 3 m vertically.

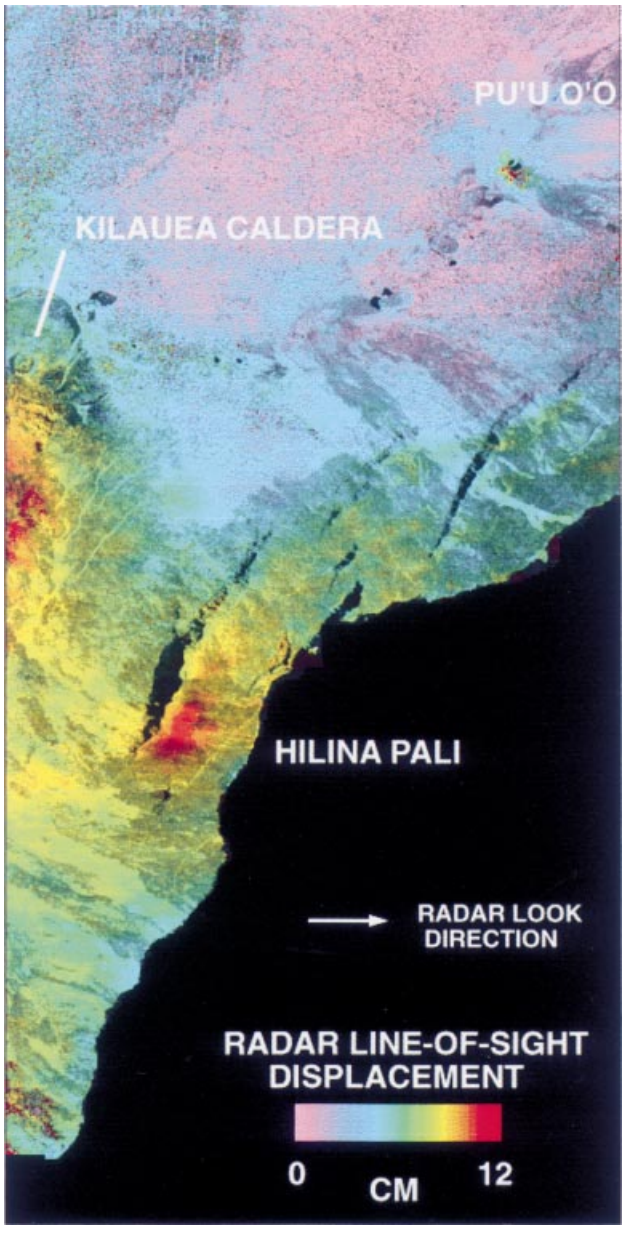

Figure 5 Kilauea, Hawaii, showing ground displacement between April and October, 1994. The data were collected by the SIR-C/X-SAR instrument carried by the Space Shuttle. See text for further elaboration. (Image courtesy of NASA Jet Propulsion Laboratory, Pasadena, California. http://www.jpl.nasa.gov/radar/sircxsar/kilauea2.html)

Figure 5 shows the ground displacement that occurred between April and October 1994, in the region of Kilauea volcano, Hawaii. The image area is approximately 40 km by 80 km. The colors indicate the displacement of the surface in the direction that the radar was pointing. The radar data used to generate this image were acquired from the SIR-C instrument, developed by the American, German, and Italian space agencies. The ground surface shape was derived from two sets of synthetic aperture radar (SAR) images, collected during Shuttle flights in April and October, 1994. Each set consists of two radar images collected on different orbits. Height information for each set was determined using interferometric methods, and the difference (displacement) is calculated from analysis of the differences between the two sets of height measurements. This type of information is of value in delimiting the extent of earthquake activity, and in monitoring surface movements in tectonically active areas of the world.

Geological mapping and mineral resources exploration is a major area of terrestrial remote sensing applications, and these are described in two articles entitled **Hyperspectral Remote Sensing: Data Collection and Exploitation** and **Imaging Spectrometry for Geological Applications**. Instruments capable of collecting data for large areas of the Earth's land surface in a large number of narrow spectral bands and at high spectral resolution and hyperspectral instruments is providing new opportunities for detailed geological mapping, especially in arid and semi-arid areas. The article on **Stellar Spectroscopy** shows that techniques used in terrestrial applications are equally relevant to the exploration of extra-terrestrial objects.

Figure 6(a) and (b) shows two examples of images collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) developed by the NASA Jet Propulsion Laboratory, Pasadena, California. This instruments collects information in 224 contiguous spectral channels between 400 and 2500 nm. It is flown on a NASA ER-2 aircraft (a modified U2) at an altitude of approximately 20 km. Image (a) is a composite, made up of data from three channels in the visible and NIR range of the AVIRIS. The bright red areas in this image show vegetation. Image (b) is also made up of three AVIRIS bands, located in the short-wave infrared region of the spectrum. The turquoise regions in the upper left corner delineate ground covered by limestone fragments. Image (c) was collected by the thermal infrared MSS and, in this image, surfaces with a quartz content are shown in red. The three images also show the impact of spatial resolution: image (a) has a spatial resolution of 15 m, while (b) has a resolution of 30 m. The thermal infared image (c) has a resolution of 90 m.

(a) (b) (c)

Figure 6 Death Valley: three views using AVIRIS imaging spectrometer data (a and b) and thermal infrared MSS data (c). (a) Visible and NIR image (spatial resolution is 15 m); (b) short-wave infrared (spatial resolution is 30 m); (c) thermal infrared (spatial resolution is 90 m). The size of the area shown on these images is 12 × 50 km. (Image courtesy of NASA Jet Propulsion Laboratory, Pasadena, California: http://asterweb.jpl.nasa.gov/application/geology/death.htm)

Land cover change may result from a variety of factors. Human activity, in the form of agricultural practices, is one major driver of change. Economic activities, such as logging and forest burning, are also widely recognized as processes that have a global impact. The effects of global environmental changes also influence the nature of land cover. Land cover changes may also promote environmental change – this positive feedback loop is an important component of global climate models. Our knowledge of the distribution, status and dynamics of the Earth's terrestrial vegetation is restricted and uncertain, and monitoring from space is perhaps the only feasible way of maintaining our awareness of these changes. **Global Land Databases for Environmental Analyses** considers ways in which data derived from remote sensing sources can be stored, managed and accessed, while **Land Cover Assessment and Monitoring; Temperate Forest Resource Assessment by Remote Sensing** and **Tropical Forest Resource Assessment by Remote Sensing** are concerned with different aspects of monitoring various types of land cover. As pressure on marginal lands builds, as a result of population increase, as exploration for minerals and hydrocarbons extends more widely, and as the effects of atmospheric pollution spread globally, so the capability to monitor and assess fragile ecosystems grows in importance. **Semiarid Land Assessment: Monitoring Dry Ecosystems with Remote Sensing** and **Polar Environments Assessment by Remote Sensing** consider the impact of remote sensing technology on human ability to measure and assess change in two such fragile regions of the planet.

No survey of the applications of Earth observation by remote sensing would be complete without a survey of the methods used in processing the images acquired by aircraft and satellites. A wide range of processing techniques are required if uses of such images are to be valid and reliable. Correction of the raw data for flaws and deficiencies, calibration in order to recover radiance data from quantized counts, and geometric registration of the images with a conventional map coordinate system are needed if remotely sensed data are to be used with other geographical data in the solution of problems. **Processing and Classification of Satellite Images** summarizes the use of advanced pattern recognition techniques, such as artificial neural networks, in the generation of thematic maps from multispectral imagery. This type of information, derived from satellite and aircraft imagery, is now widely used in geographical information systems.

As population pressures increase, the future of the planet depends on the timely acquisition of data concerning the state of our environment, and the use of these data in monitoring and modeling change. Earth observation by remote sensing offers an effective and affordable way of collecting such data. The articles in this section provide an overview of the methods used in the application of remotely sensed data to a variety of environmental problems.

ABBREVIATIONS AND ACRONYMS

AVIRIS	Airborne Visible Infrared Imaging Spectrometer
CZCS	Coastal Zone Color Scanner
DEM	Digital Elevation Model
ERTS-1	Earth Resources Technology Satellite-1
HRV	High Resolution Visible

InSAR	Interferometric Synthetic Aperture Radar
MSS	MultiSpectral Scanner
NASA	National Aeronautics and Space Administration
NIR	Near Infrared
SAR	Synthetic Aperture Radar
SEM	Scanning Electron Microscopy
TM	Thematic Mapper

REFERENCES

1. P.M. Mather, *Computer Processing of Remotely-sensed Images*, John Wiley and Sons, Ltd., Chichester, England, 1999.
2. T.T. Alföldi, 'Remote Sensing for Water Quality Monitoring', in *Remote Sensing for Resource Management*, eds. C.J. Johanssen, J.L. Saunders, Ankeny, Iowa: Soil Conservation Society of America, 317–328, 1982.

Biological Oceanography by Remote Sensing

M.A. Srokosz
Southampton Oceanography Centre, Southampton, UK

1 Introduction **1**
1.1 A Brief History of Ocean Color Measurements from Space 2
2 Light in the Ocean **3**
2.1 Some Definitions 3
2.2 In-water Constituents and Bio-optics 4
2.3 Measurements from Space and the Effect of the Atmosphere 5
2.4 Case 1 and Case 2 Waters 6
3 Satellites and Sensors **6**
3.1 Missions and Sensors Characteristics 6
3.2 Algorithms, Including Atmospheric Correction 9
3.3 Calibration and Validation 10
4 Applications **11**
4.1 Measurements of Phytoplankton 11
4.2 The CO_2 Problem 15
4.3 Biophysical Interactions 17
4.4 Assimilation of Data into Models 20
4.5 Dimethyl Sulfide, Climate, and Gaia 22
4.6 Commercial Application – Fisheries 22
4.7 Possible Future Applications 22
4.8 Afterword 23
Acknowledgments **23**
Abbreviations and Acronyms **23**
Related Articles **24**
References **24**

Biological oceanography may be studied from space using sensors on satellites that determine the color of the ocean. The presence of phytoplankton (microscopic algae) in the upper layers of the ocean changes the color of the water as seen from above. In simplified terms, this is due to the selective absorption of blue light by the phytoplankton pigments (primarily chlorophyll) which changes the appearance of the water from blue to green. These changes in color can be observed using a satellite-borne spectroradiometer that measures the water-leaving radiance in a number of bands in the visible part of the electromagnetic spectrum. The limitations of the technique are first, that only information on the phytoplankton in the upper layers of the ocean can be obtained (light does not penetrate very far into the ocean). Second, most of the signal measured by the satellite sensor originates in the atmosphere (due to the molecular and aerosol scattering of photons there), so careful correction for atmospheric effects is necessary if good ocean data are to be obtained. Of course, in the presence of clouds the sensor will not "see" the ocean surface at all, and no data will be obtained. Third, only one component of the ocean ecosystem, namely the phytoplankton, can be studied by this means. Despite these limitations, satellite observations of ocean color have given new insights into biological oceanography on a global scale that could not have been obtained by any other means of observation. Observations of ocean color have contributed to a better understanding of the biophysical interactions that determine the phytoplankton productivity, the seasonal and interannual variations of the phytoplankton biomass on global scales, and the role of phytoplankton in the climate system. They have also contributed to the improved modeling of biogeochemical processes in the ocean.

1 INTRODUCTION

The purpose of this article is to describe the application of satellite remote sensing techniques to the study of biological oceanography. In one sense it seems strange to think that a sensor (instrument) flying on a satellite several hundreds of kilometers above the ocean surface can tell us anything at all about the biology of the ocean. This initial reaction is to a large extent correct, in that measurements from space can only tell us something directly about one very specific component of the ocean biology. That component is the phytoplankton, microscopic plants that live in the near-surface waters of the ocean. The reason that this component can be observed from space is that the phytoplankton contain pigments that are necessary for photosynthesis (primarily chlorophyll-a) and their presence in the water changes the color of the water, as seen from above the sea surface, usually from blue to green. Thus a satellite-borne sensor that makes measurements in the visible part of the electromagnetic spectrum can be used to measure the change in color, and so provide information on the phytoplankton.

Given that only a single component of the ocean biological system can be measured from space, one might ask: why bother? Plankton are the most abundant life form in the world's oceans, both in terms of weight and of numbers. Phytoplankton are microscopic plants, while zooplankton are the microscopic and small animals that feed on the phytoplankton. A cubic meter of seawater

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

will contain millions of these small plants. Phytoplankton are the oceanic equivalent of terrestrial plants, forming the basic element of the oceanic food chain. The total phytoplankton biomass is greater than that of all the marine animals taken together (zooplankton, fish, and so on). In addition to their role in the food chain, they have a significant role in the world's climate system. Their presence in the water causes light to be scattered and absorbed, which warms the upper layers of the ocean. They produce chemical compounds that escape into the atmosphere and have a role in the formation of clouds. More fundamentally, in growing they use carbon dioxide from the atmosphere that has been absorbed into the ocean. When they die, some proportion of the plankton fall out of the upper layers of the ocean and become part of the seabed sediments, thus removing carbon from the system. Therefore, phytoplankton have a major role in the global carbon cycle and may be important in either ameliorating or accelerating the effects of anthropogenic emissions of CO_2 into the atmosphere. At the present time it is not known how the phytoplankton will respond to the warming occurring due to the increase of greenhouses gases.[1]

The preceding paragraph shows the importance of phytoplankton and why measuring them and their behavior is necessary, but why do it from space? As noted above, phytoplankton are ubiquitous in the world's oceans, but the traditional ship-based methods of observation are unable to give a truly global view of the phytoplankton in the ocean. Thus observations from space are the only means of obtaining a global view. In order to understand the oceanic ecosystem, of which phytoplankton are just one component, ship-based and other types of measurements are still necessary. However, the ability to measure one component of the system from space has brought many new insights into the biology of the ocean on the global scale. These will be described later in this article.

There are, of course, drawbacks to measuring phytoplankton by satellite remote sensing methods. Since the measurement of ocean color is made using the visible part of the electromagnetic spectrum, the major problem that arises is the presence of clouds, which prevents the sensor from seeing the sea surface and thus making measurements. (Other satellite sensors that measure in a different part of the electromagnetic spectrum, particularly the microwave part, can see through clouds, but do not provide information on ocean biology directly).[2–4] As some areas of the world's oceans are more prone to cloud cover than others this could lead to bias in the measurements. A second problem is that of the depth of penetration of light into the ocean. The satellite sensor measures light exiting from the sea surface representing the end-result of complex interactions (absorption and scattering) of the light entering the ocean with the constituents (such as phytoplankton) present in the water, and the water itself. Depending on the constituents present, the exiting light represents information about the constituents over some depth (the details of this will be considered further below). This depth varies from place to place, so interpretation of the measurements may not be straightforward. A particular example of this problem is that of the so-called deep chlorophyll maximum.[5] This occurs when the surface waters are depleted of the nutrients necessary for the phytoplankton to grow. Here the balance between the phytoplankton's need for light (available from above) and nutrients (available from below) to grow, means that the bulk of the phytoplankton growth takes place well below the surface and may not be visible to the satellite sensor. Thus any estimate of phytoplankton activity, particularly primary production, based on ocean color measured from space will need to account for this type of situation.

Despite the drawbacks mentioned in the previous paragraph, the ability to measure ocean color from space has brought many new insights into ocean biology and these will be described later in this article. To begin, a brief history of the measurement of ocean color from space will be given.

1.1 A Brief History of Ocean Color Measurements from Space

The first true ocean color sensor was the Coastal Zone Color Scanner (CZCS), which was launched by NASA (National Aeronautical and Space Administration) on the Nimbus-7 satellite in late 1978 and operated until mid-1986.[6] This followed on from the work of Clarke et al.[7] who showed that chlorophyll concentration in the ocean surface waters could be estimated from airborne measurements of the light leaving the sea surface. CZCS made measurements in four channels in the visible part of the spectrum, one channel in the near-infrared, and one in the infrared (IR) (see Table 2 for details of the sensor). The latter channel allowed simultaneous measurement of the sea surface temperature (SST), but failed early in the mission. In addition, the sensor showed degradation over the life of the mission, which meant that the data had to be carefully processed to take this into account.[8] A final drawback was that data were not acquired continuously globally during the mission, owing to the limits of power and on-board data recording. Nevertheless, data from most parts of the globe were acquired and the first truly global picture of phytoplankton activity in the ocean was obtained by averaging the data over time.

The next sensor due to be launched was the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) flown on SeaStar, but various delays meant the Ocean Color and

Temperature Sensor (OCTS) was launched first by the National Space Development Agency of Japan (NASDA) on ADEOS (Advanced Earth Observation Satellite). OCTS was operational from August 1996 until June 1997, when the ADEOS suffered a catastrophic failure (Tables 3 and 4). In addition to measuring ocean color, OCTS measured SST using channels in the IR. SeaWiFS, a collaborative venture between NASA and the Orbital Sciences Corporation (OSC), was launched in August 1997, shortly after the failure of ADEOS. It continues to operate well and provide data globally (Table 5). Both SeaWiFS and OCTS have more channels in the visible part of the spectrum than CZCS, and this allows for better retrieval of biological information from the ocean (see section 3).

Another sensor that is flying in space, on the Indian IRS-P3 satellite, is an experimental one developed by DLR (German Space Agency) in Germany, the Modular Optoelectronic Scanner (MOS). It was launched in March 1996 and does not provide global data. It does have similar channels in the visible part of the spectrum to OCTS and SeaWiFS, but a much narrower swath (200 km). Two further sensors capable of measuring ocean color are due to be launched. The Moderate Resolution Imaging Spectroradiometer (MODIS) is due to be launched by NASA on the first Earth Observing System (EOS) platform in late 1999.[9] The Medium Resolution Imaging Spectrometer (MERIS) is due for launch in early 2001 on the European Space Agency's (ESA) satellite Envisat.[10] Both sensors have more channels than OCTS and SeaWiFS (Tables 6, 7 and 8). MODIS and MERIS are not just designed for ocean color measurements but will also provide data on the atmosphere, and on terrestrial vegetation. Finally, the experimental Ocean Color Imager (OCI) is due to be launched on the Taiwanese satellite ROCSAT-1 in 1999.

The above discussion has given details of the sensors that have specific ocean color capability, but it is also worth noting that other sensors that measure in the visible part of the spectrum have been used occasionally.[2] In general, these sensors are not sufficiently sensitive for ocean color measurements, an exception being the detection of coccolithophore blooms using the visible band of AVHRR (advanced very high resolution radiometer); an instrument designed for measuring SST. Owing to their high reflectivity, coccolithophores can be seen by less-sensitive sensors such as AVHRR (see section 4.1.3 below). In this article the focus will be on those sensors specifically designed for ocean color measurements.

In order to understand how it is possible to obtain information about biological activity from satellite ocean color sensors, it is necessary to consider first the behavior of light in the ocean. This is the subject of the next section (section 2). Following this a more detailed description of the sensors and algorithms used to retrieve ocean color and biological information is given (section 3). Finally (section 4) the application of that information to study ocean biology will be discussed.

2 LIGHT IN THE OCEAN

The subject of light in the ocean is a vast one, as evidenced by the more than one thousand references given in a standard text by Kirk.[11] It is not possible in a brief article to do justice to all these aspects, so the focus will be on those most relevant to the remote sensing of ocean color. In this context light will be taken to mean electromagnetic radiation of wavelengths ca. 400–700 nm, to which the human eye responds, and which plants, including phytoplankton, can use for photosynthesis.[11,12] In terms of color, blue light has wavelengths of ca. 450 nm, green light ca. 520 nm and red light ca. 650 nm. The light of wavelengths 400–700 nm is usually referred to as photosynthetically active (or available) radiation (PAR).[11,13]

2.1 Some Definitions

Light (photons) entering the ocean is subject to two physical processes, absorption and scattering (for more details of the definitions in this section see Mobley[14] and Kirk[11]). The description of the optical properties associated with these processes is usually divided into inherent and apparent optical properties (IOPs and AOPs). IOPs depend only on the medium and can be specified for light of any given wavelength λ by the absorption and scattering coefficients, $a(\lambda)$ and $b(\lambda)$, and the volume scattering function, which describes the directional properties of the scattering. The AOPs depend both on the medium (thus on the IOPs) and on the geometric structure of the ambient light field. Examples of AOPs are the diffuse, or vertical, attenuation coefficients $K(\lambda, z)$, which describe the approximately exponential decrease with depth z of various properties of the light field. The IOPs and the AOPs depend on both the water and on what is in the water, for example, phytoplankton.

In terms of the light properties of interest in remote sensing, the important quantities are radiance, irradiance, upwelling (upward) and downwelling (downward) irradiance, diffuse attenuation coefficient, and the optical depth. These are all considered to depend on the frequency or wavelength of the light, as ocean color sensors measure light in discrete bands of the electromagnetic spectrum. Therefore, in terms of wavelength, they are expressed in units as per nanometer (nm^{-1}), that is per unit spectral bandwidth. Equivalent quantities integrated

over part or the whole of the visible spectrum are defined without nm^{-1}. Radiance $L(\lambda)$ is the optical property appropriate to light energy leaving an extended source or incident on a surface, such as the ocean, and has units of $W\,m^{-2}\,sr^{-1}\,nm^{-1}$ (the last but one factor is per steradian, a measure of solid angle). Irradiance $E(\lambda)$ is the radiant flux per unit area of a surface and has units of $W\,m^{-2}\,nm^{-1}$. Downwelling irradiance $E_d(\lambda)$ and upwelling irradiance $E_u(\lambda)$, are the values of the flux passing down, or up, through a horizontal surface. E_d and E_u are obtained from L, the radiance incident on the surface, by integrating with respect to the solid angle w for the upper and lower hemispheres, respectively, see Equations (1) and (2). Thus

$$E_d = \int_{\text{upper hemisphere}} L \cos v \, dw \quad (1)$$

$$E_u = -\int_{\text{lower hemisphere}} L \cos v \, dw \quad (2)$$

where v is the zenith angle.[11] The diffuse or vertical attenuation coefficient at a depth z is given by Equation (3)

$$K(\lambda, z) = \frac{-1}{E(\lambda, z)} \frac{dE}{dz} \quad (3)$$

and has units of m^{-1}. If K_d does not vary much with depth, an optical depth can be defined as $1/K_d$, where K_d is the diffuse attenuation coefficient for the downwelling irradiance E_d. It can be shown that the information on chlorophyll concentration obtained from an ocean color sensor is that from approximately one optical depth, which can be regarded as the depth that the sensor sees into the ocean.[15]

The key quantity of interest in measuring ocean color from space is the water-leaving radiance L_w. This represents the light leaving the sea surface resulting from the absorption and scattering by the water itself and by in-water constituents, such as phytoplankton, of light incident on the sea surface. It is this light that contains information about what is in the water. However, the light (radiance) that the satellite sensor measures L_s originates from a number of sources (see, for example, Robinson[2] or Kirk[11]). Even in relatively clear atmospheric conditions L_w is only 10–20% of L_s.[11] This means that the effect of the atmosphere must be accounted for in deriving L_w from L_s. This problem is discussed in section 2.3 below. For various technical reasons that will not be discussed here, two other related quantities are sometimes used in the measurement of ocean color rather than L_w. These are the normalized water-leaving radiance; that is, approximately the radiance that would exit the ocean in the absence of the atmosphere and with the sun at the zenith.[16] Another alternative to L_w is the reflectance ρ_w; that is, L_w normalized with respect to the extraterrestrial solar irradiance.[16] The important point to note is that whichever quantity is used it can be related to the presence of phytoplankton in the water, which will be discussed in the next subsection.

The final definition to be given in this section is that of the euphotic zone. Kirk[11] states that a useful rule of thumb in aquatic biology is that significant phytoplankton photosynthesis takes place down to a depth z_e at which the downwelling irradiance of PAR falls to 1% of its value just below the sea surface. The layer in which $E_d(\text{PAR}, z)$ is greater than or equal to 1% of $E_d(\text{PAR}, 0)$ is known as the euphotic zone. It can be shown that the depth that the ocean color sensor sees into the ocean is approximately $z_e/4.6$ (one optical depth). In waters with a low concentration of chlorophyll ($0.1\,mg\,m^{-3}$) this depth is about 25 m, whereas for waters with a higher concentration ($10\,mg\,m^{-3}$) it is about 5 m.[17] High concentrations of chlorophyll reduce the penetration of light into the ocean, owing to absorption and scattering.

2.2 In-water Constituents and Bio-optics

In simple terms the presence of phytoplankton in the water changes the color of the water as seen by the color sensor. Waters low in phytoplankton pigments reflect more blue light than green, whereas waters high in pigments reflect more green light as a result of the selective absorption of blue light by the pigments. This means that the shape of the light spectrum changes and by measuring in different bands of the spectrum it is possible to quantify the concentration of pigment present in the water.[2,11] The pigments that absorb the light are chlorophyll-a and phaeopigments, so the estimate obtained from the color data is the pigment concentration. However, the phaeopigments are a small fraction of the total (ca. 10%),[18] so the pigment concentration can be regarded as the chlorophyll-a concentration in the case of CZCS, owing to the inherent error in the algorithms used to recover this information (section 3.2). In the section on applications, the terms chlorophyll concentration or pigment concentration will both be used when discussing CZCS data. In order to extract the pigment concentration information from the ocean color measurements it is necessary to develop bio-optical algorithms and the theoretical basis of these is considered next.

It can be shown theoretically that the water-leaving radiance $L_w(\lambda)$ is related to the reflectance $R(\lambda) = E_u(\lambda)/E_d(\lambda)$ evaluated just below the sea surface.[2,15] Theoretical modeling suggests that $R(\lambda)$ in turn depends on the absorption $a(\lambda)$ and backscattering $b_b(\lambda)$ coefficients (note that the scattering coefficient $b(\lambda) = b_b(\lambda) + b_f(\lambda)$, the scattering in the backward and

forward directions[11]). For b_b/a small, $R = 0.33 b_b/a$;[2,19] more generally R may be regarded as a function of $[b_b/(a + b_b)]$.[2,20] In addition, the absorption and scattering coefficients can be decomposed into the contributions due to the water itself and to each of the constituents in the water, and so related to the chlorophyll concentration. On this basis it is possible to develop what are known as semianalytical models of ocean color, where the behavior of $a(\lambda)$ and $b_b(\lambda)$ is established through a combination of modeling and measurements.[20] For CZCS the standard algorithms were in fact based on an empirical approach, where the ratio of the water-leaving radiance in two bands was compared to values of the pigment concentration measured in situ (section 3.2).[15] This is a simple measurement that compares the water-leaving radiance in the blue to that in the green part of the visible spectrum, the ratio decreasing with increasing concentration of pigments.[20] For future missions both empirical and semianalytical approaches to estimating pigment concentration will be used (see, for example, Esaias et al.[9]). In principle, the semianalytical approach should provide improved information as it should reduce some of the uncertainties associated with the empirical approach.[20] In either case accurate measurements of the water-leaving radiance in the sensor bands need to be obtained and this requires that the data be corrected for atmospheric effects, which are discussed next.

2.3 Measurements from Space and the Effect of the Atmosphere

It is well known that at least 80–90% of the signal received by an ocean color sensor at the top of the atmosphere originates from the atmosphere rather than the ocean, even in relative clear atmospheric conditions.[11] In the presence of clouds no signal from the sea surface is obtained at all, of course. Therefore the removal of the atmospheric effects from the signal received by the sensor is crucial if accurate measurements of the water-leaving radiance in the various sensor bands are to be obtained. The signal received by the satellite sensor may be written as Equation (4)[9]

$$L_s(\lambda) = L_r(\lambda) + L_a(\lambda) + L_{ra}(\lambda) + T(\lambda)L_{sg}(\lambda) + t(\lambda)L_{wc}(\lambda) + t(\lambda)L_w(\lambda) \quad (4)$$

where L_r is the scattering of the photons due to air molecules (known as Rayleigh scattering), L_a is the scattering due to atmospheric aerosols (dust, water droplets, salt, and so on), L_{ra} represents interactions between the two previous effects, L_{sg} is the sunglint contribution (the direct reflection of sunlight from the sea surface), L_{wc} is the whitecap contribution (due to the presence of breaking waves), and L_w is the desired water-leaving radiance. T and t are the direct and diffuse transmittance of the atmosphere, that is the effects of the atmosphere on the signal from the sea surface. Sunglint cannot be corrected for, so data contaminated by sunglint are usually discarded.[16] Considerable effort has been devoted to obtaining accurate corrections with regard to the other terms. Gordon[16] gives a comprehensive review.

As the applications discussed later are based on data from the CZCS instrument, a brief explanation of the atmospheric correction technique developed for it will be given here. The ability to correct for atmospheric effects is partly dependent on the measurements made by the sensor. If measurements are obtained in a sufficient number of independent bands in the electromagnetic spectrum, then the effects of the atmosphere can be accounted for (to a lesser or greater degree depending on what is in the atmosphere). As CZCS had only four bands in the visible, the possibilities for atmospheric correction were limited, so simplifying assumptions were made. In particular, the Rayleigh scattering was assumed to be single scattering (photons have only one scattering encounter with air molecules) and this can be calculated theoretically, as can the diffuse transmittance of the atmosphere. The interaction term between the aerosol scattering and Rayleigh scattering term was ignored (being a secondary effect). The effect of whitecaps on the sea surface was also ignored.

The aerosol scattering term was treated in a simplified manner (see the appendix of Gordon et al.[20]). If the water-leaving radiance is known in two bands at one position then the ratio $S(\lambda_1, \lambda_2) = L_a(\lambda_1)/L_a(\lambda_2)$ may be found (L_w known, L_s measured, t and L_r calculated, all other terms ignored). Making the assumptions that the aerosol scattering phase function is approximately independent of wavelength, it can be shown under the single scattering approximation that $S(\lambda_1, \lambda_2) = \varepsilon(\lambda_1, \lambda_2)\ F(\lambda_1)/F(\lambda_2)$, where the Fs represent the ratio of the extraterrestrial solar irradiance for the two wavebands, corrected for absorption by atmospheric ozone, and $\varepsilon(\lambda_1, \lambda_2)$ is the ratio of the aerosol optical thicknesses at λ_2 and λ_1. ε depends only on the type of aerosol present in the atmosphere and not its concentration. For the ocean a reasonable assumption is that the aerosol type is the same across an image and only the concentration varies. Thus if ε can be determined for the pairs of CZCS bands somewhere in an image, then the water-leaving radiances can be calculated for the whole image. This was done based on the clear water radiance concept which gives $L_w(520)$, $L_w(550)$ and $L_w(670)$ when the pigment concentration is less than $0.25\,\mathrm{mg\,m^{-3}}$. This allows $\varepsilon(520,670)$ and $\varepsilon(550,670)$ to be determined for pixels satisfying the clear water condition, and $\varepsilon(443, 670)$ is extrapolated from these values. The further assumption

that $L_w(670) = 0$ then allows the water-leaving radiance in the other three bands to be determined. This is based on the fact that water absorbs light strongly in the red/near-IR region of the spectrum,[11] which is a reasonable assumption for clear oceanic waters. There are problems with this approach if there is no clear water in the image or if the aerosol type varies across the image. For the first problem, an iterative approach is possible by starting with an initial guess for $L_w(670)$.[20,21] The second problem, together with the problems due to the simplifying assumptions made in the CZCS approach, can only be dealt with by a more sophisticated pixel by pixel atmospheric correction procedure, which can incorporate the information available from the extra channels of the new sensors. These are being developed and applied.[16,22]

The CZCS atmospheric correction procedure worked reasonably well for the open ocean, but less well in coastal areas. It is not valid to make assumptions about the constancy of aerosol type near land, and about the water-leaving radiance in the red/near-IR region when there are sediments in the water.[11,21] For reasons like these it has proved useful to distinguish between two different water types when using the data from color sensors, as explained in the next section.

2.4 Case 1 and Case 2 Waters

Before proceeding further in the discussion of ocean color measurements it is necessary to define the terms Case 1 and Case 2 waters, which are often used in the remote sensing of ocean color. These terms were originally used by Morel and Prieur.[19] Case 1 waters are ones where the optical signature is due to the presence of phytoplankton and their by-products. Case 2 waters are ones where the optical signature may also be influenced by the presence of suspended sediments, dissolved organic matter and terrigenous particles from rivers and glaciers. In general, Case 1 waters are those of the open ocean, while Case 2 waters are those of the coastal seas. The derivation of geophysical parameters from Case 2 waters is much more complex, as the presence of the various particulates in the water, in addition to the phytoplankton, affects the color signal measured by the satellite sensor.[2,23] This also makes atmospheric correction of the data more difficult (see section 2.3). Given these complications and that the focus of this article is on biological oceanography, further discussion will be restricted to Case 1 waters. It is worth noting that the new generation of satellite sensors (SeaWiFS, MODIS, MERIS; see next section) which possess more bands in the visible part of the spectrum than did CZCS, will allow better discrimination between the various in-water constituents in Case 2 waters.[24] Almost all the applications of ocean color data that will be discussed below (section 4) will be from open ocean Case 1 waters.

3 SATELLITES AND SENSORS

From the previous section, it is clear that it is necessary to measure the spectral properties of the light leaving the ocean, and to correct for the effects of the intervening atmosphere on the measurements, in order to obtain useful biological measurements from spaceborne sensors. In this section, the characteristics of various ocean color sensors that have been, or will be, flown in space are described. In addition, an overview of the algorithms used to derive biological information will be presented together with an explanation of how these have been calibrated and validated.

3.1 Missions and Sensors Characteristics

As noted in the historical introduction (section 1.1) a number of ocean color sensors have been or are to be flown in space. The characteristics of each sensor will be given in the following subsections. Here a summary of the instruments and satellites is provided in Table 1. (The experimental sensors MOS and OCI mentioned in the historical survey have been excluded from what follows.) Details of earlier visible band sensors that were not specifically designed for ocean color studies but have been used for such may be found in Robinson[2] and Stewart.[3]

Optical sensors are essentially of two types, pushbroom or scanning. Pushbroom sensors have optics that enable a line of data, in an across-track direction, to be acquired at the same time. Scanning sensors have optics that scan in the across-track direction, thus acquiring a line of data. The swath width of the sensor is determined by the height above the surface and the maximum angle of view across-track to either side of nadir. The instantaneous field of view (IFOV) of the sensor fixes the size of the pixels that make up each line of data. The angular resolution of the sensor's optical system and the height of the satellite orbit determine the IFOV. Owing to the viewing geometry, for a fixed angular resolution, this means that pixels at nadir are somewhat smaller than those out towards the edge of the instrument swath. An image is built up of successive

Table 1 Ocean color sensors and satellites

Sensor	Satellite	Agency	Dates
CZCS	Nimbus-7	NASA	10/78–6/86
OCTS	ADEOS	NASDA	8/96–6/97[a]
SeaWiFS	SeaStar	OSC/NASA	launched 8/97
MODIS	EOS-AM1	NASA	due for launch 1999
	EOS-PM1		due for launch 2000
MERIS	Envisat	ESA	due for launch 2001

[a] Mission ended prematurely owing to unfortunate catastrophic failure of ADEOS satellite.

lines of data as the satellite moves along its orbit. Some optical sensors flown in space also have the capability to tilt, thus being able to look ahead or behind the satellite, rather than just down at nadir. This capability is used to avoid sunglint problems over certain parts of the satellite's orbit, by tilting the sensor away from the sunglint (the direct reflection of the sun in the sea surface). Clearly the degree of tilt will affect both the swath width and the pixel size.

In the following, details of the individual sensors will be given, including the type of sensor (pushbroom or scanning, tilting or nontilting), the orbit height, the swath width, the spatial resolution (pixel size) and, most important, the bands in the electromagnetic spectrum in which each sensor measures. At the typical orbit altitudes and swath widths of the satellites and sensors, respectively, global coverage of the oceans is acquired over about three days (on-board recording capability permitting).

3.1.1 Coastal Zone Color Scanner

CZCS was a scanning sensor able to scan ±39.34° each side of nadir. The satellite altitude of 955 km meant that the resulting swath width was 1659 km when the sensor was in nadir-looking mode. The sampling of the scan was such as to give 1968 pixels across the swath. The sensor could tilt ±20°, in steps of 2° to avoid sunglint. The IFOV (ca. 0.05°) gave pixels that varied from 825 m at nadir to 1653 m at the edge of the swath (again in nadir-looking mode). The swath width and pixel size varied with the degree of tilt. The sensor had six channels (Table 2), four in the visible part of the spectrum, one in the near-IR, and one in the thermal IR.[21,25]

Owing to limitation of power on board the Nimbus-7 satellite, CZCS was limited to only two hours of operation each day. Furthermore, limited data-recording capability meant that much of the data had to be acquired when the satellite was within reception range of a ground recording station. When combined with the problems of cloud cover this meant that, over its lifetime 1978–1986, CZCS data coverage was somewhat patchy. Despite this, it provided the first global view of the biology of the oceans and subsequent work on the data has produced a calibrated and consistent data set for that period.[6]

Table 2 CZCS bands

Band no.	Band center (nm)	Band width (nm)	S/N
1	443	20	350
2	520	20	342
3	550	20	280
4	670	20	209
5	750	100	50
6	11 500	2000	Thermal IR band

S/N, Signal-to-noise ratio.

Over its period of operation the visible band detectors were found to be suffering from decreasing and variable sensitivity,[8,15] and careful correction for this effect was necessary in order to make full use of the data.[8] The broad near-IR band (band 5; see Table 2) was only useful for land–sea–cloud discrimination. In addition, the thermal IR band detector failed shortly after launch so that SST data were not available. Most of the processing effort has therefore focused on data from the four visible bands (bands 1 to 4).[21]

3.1.2 Ocean Color and Temperature Sensor

After a gap of 10 years from the demise of CZCS, OCTS was launched in August 1996. Unfortunately the mission was short lived as the ADEOS satellite suffered a catastrophic failure in June 1997, thus acquiring only nine months of data. Although launched prior to SeaWiFS (which was to have been launched about 1994), OCTS was designed in the light of SeaWiFS and was therefore in some respects a very similar instrument. It had six visible bands and two near-IR bands, which corresponded very closely to the SeaWiFS bands (compare Tables 3 and 5). In addition, OCTS had four bands in the thermal IR (see Table 4) which allowed the retrieval of SST, thus providing contemporaneous information on ocean biology (from color) and physics (from SST). The spatial resolution at nadir was approximately 700 m. The nominal orbit altitude was 700 km, giving a swath width of about 1400 km. The sensor was of the scanning type, with the ability to tilt ±20° forward and backwards to avoid sunglint.[26]

3.1.3 Sea-viewing Wide Field-of-view Sensor

SeaWiFS was designed as the successor to CZCS, but for a variety of reasons its launch was delayed until

Table 3 OCTS visible bands

Band no.	Band center (nm)	Bandwidth (nm)	Comments
1	412	20	
2	443	20	
3	490	20	
4	520	20	
5	565	20	
6	665	20	
7	765	40	Atmospheric correction
8	865	40	Atmospheric correction

Table 4 OCTS thermal bands (μm)

3.55–3.88	8.25–8.80	10.3–11.4	11.4–12.5

August 1997, over 10 years after the demise of CZCS. It is now in orbit and providing very high quality ocean color data. Unlike CZCS, it has no bands in the IR, so provides no information on SST (its successor MODIS will have IR bands, see section 3.1.4). A comparison of the capabilities of CZCS and SeaWiFS is given by Hooker et al.[25] SeaWiFS flies on the SeaStar satellite operated by OSC, with NASA buying the data for distribution to the scientific community, while OSC sell the data to commercial organizations for other applications (see section 4.6).

SeaWiFS has six bands in the visible part of the electromagnetic spectrum and two bands in the near-IR for atmospheric correction purposes.[27] These bands are very similar to those of OCTS (compare Tables 3 and 5). The increased number of visible bands, plus the bands in the near-IR, and the improved sensitivity (see S/N in Tables 2 and 5) will allow better atmospheric correction of the data and improved retrieval of biological information, compared with CZCS.[24] SeaWiFS is a scanning instrument and has the capability to tilt $\pm 20^\circ$ forwards or backwards to avoid sunglint. It is operated in two spatial-resolution modes, local area coverage (LAC) with a 1.1 km resolution at nadir, and global area coverage (GAC) with a 4.5 km resolution at nadir. At the nominal orbit altitude of 705 km, the LAC has a swath width of 2801 km, while the GAC has a swath width of 1502 km. The on-board recording system has only limited capacity for LAC data, so this capability is primarily for use when the satellite is within range of a ground receiving station.

3.1.4 Moderate Resolution Imaging Spectroradiometer

MODIS is due for launch on the EOS-AM1 satellite in late 1999. Details of the instrument may be found in Barnes et al.[28] This instrument has 36 bands in the visible and IR part of the electromagnetic spectrum and will be used for atmospheric and terrestrial studies, as well as for biological oceanography. The bands and their uses are listed in Tables 6 and 7. At nadir, bands 1 and 2 have a spatial resolution of 250 m, bands 3–7 of 500 m and bands 8–36 of 1 km. Thus the spatial resolution of the bands being used for biological oceanography (8–16) is similar to that of CZCS, OCTS, and SeaWiFS. The bands are narrower and the sensitivity greater than for the earlier instruments (see Tables 2, 3 and 5). The instrument is of the scanning type and has a swath width of 2330 km at its nominal operational altitude of 705 km, with 1354 pixels across the swath. In order to obtain the best

Table 5 SeaWiFS bands

Band no.	Band center (nm)	Bandwidth (nm)	S/N	Comments
1	412	20	499	
2	443	20	674	
3	490	20	667	
4	510	20	616	
5	555	20	581	
6	670	20	447	
7	765	40	455	Atmospheric correction
8	865	40	467	Atmospheric correction

Table 6 MODIS visible bands

Primary use	Band no.	Bandwidth (nm)	Required S/N
Land/cloud boundaries	1	620–670	128
	2	841–876	201
Land/cloud properties	3	459–479	243
	4	545–565	228
	5	1230–1250	74
	6	1628–1652	275
	7	2105–2155	110
Ocean color/ phytoplankton/ biogeochemistry	8	405–420	880
	9	438–448	838
	10	483–493	802
	11	526–536	754
	12	546–556	750
	13	662–672	910
	14	673–683	1087
	15	743–753	586
	16	862–877	516
Atmospheric water vapor	17	890–920	167
	18	931–941	57
	19	915–965	250

Table 7 MODIS IR bands

Primary use	Band no.	Bandwidth (μm)	Required NEΔT(K)[a]
Surface/cloud temperature	20	3.660–3.840	0.05
	21	3.929–3.989	2.00
	22	3.929–3.989	0.07
	23	4.020–4.080	0.07
Atmospheric temperature	24	4.433–4.498	0.25
	25	4.482–4.549	0.25
Cirrus clouds water vapor	26	1.360–1.390	150 (S/N)
	27	6.535–6.895	0.25
	28	7.175–7.475	0.25
	29	8.400–8.700	0.05
Ozone	30	9.580–9.880	0.25
Surface/cloud temperature	31	10.780–11.280	0.05
	32	11.770–12.270	0.05
Cloud top altitude	33	13.185–13.485	0.25
	34	13.485–13.785	0.25
	35	13.785–14.085	0.25
	36	14.085–14.385	0.35

[a] NEΔT, Noise equivalent temperature difference.

possible measurements the instrument has four on-board calibration systems (see Barnes et al.[28] for details).

Although it was conceived as a two-instrument system (MODIS-T, with tilt capability, and MODIS-N, nadir looking), owing to cost constraints the tilting capability had to be foregone and a single instrument was designed. In order to compensate for the loss of oceanographic data due to sunglint (which could have been reduced using the tilt mechanism), a second MODIS instrument will be flown on the EOS-PM1 platform. It turns out that it is cheaper to design a single instrument and fly a duplicate, than to design and fly the original two-instrument system. The two instruments together will provide approximately the same global coverage as a single tilting instrument.[9]

The MODIS capabilities for ocean observations are described by Esaias et al.,[9] so only a brief summary is given here. It should be noted that MODIS's ability to measure SST (using the IR bands), in addition to ocean color, will allow it to provide data on biophysical interactions. The end-result of the improved instrument performance is the ability to obtain more information on ocean biology. Basic information derived will concern the water-leaving radiance in the various bands. From these data information on pigment concentrations, chlorophyll-a, coccolithophores (see section 4.1.3), phycoerythrin (a specific algal pigment), ocean primary production (see section 4.1.2), and solar-stimulated chlorophyll fluorescence will be obtained.[9] MODIS will be the first satellite instrument to provide data on phycoerythrin and solar-stimulated chlorophyll fluorescence (see section 4.7.2).

3.1.5 Medium Resolution Imaging Spectrometer

MERIS[10] is the ESA equivalent of NASA's MODIS instrument described in the previous section, but without the bands in the IR part of the electromagnetic spectrum. It too will provide data for atmospheric and terrestrial, as well as oceanographic, studies (see Table 8). MERIS is a 15-band programmable instrument, so that the band positions and widths can in principle be changed during the mission. Only 15 of the 16 preliminary bands, listed in Table 8, will be acquired and transmitted when MERIS is in orbit. It will fly on ESA's Envisat satellite at a nominal altitude of 800 km, giving it a swath width of about 1150 km, with 740 pixels across the swath. MERIS is a nontilting pushbroom type of instrument, so it will suffer from sunglint problems similar to MODIS.

MERIS had been designed to be useful for both global and regional studies. For this reason it has two spatial resolution modes. The so-called full-resolution mode has a resolution of 300 m at nadir, while the reduced-resolution mode has a resolution of 1200 m at nadir, the latter being similar to CZCS, OCTS, SeaWiFS, and MODIS. The full-resolution mode is intended for regional studies only, as the on-board data recording capacity is insufficient to capture all the data at this resolution on a global scale. Table 8 gives some indication of the types of information that will be derived from the various MERIS bands (see also Rast[10]). As with MODIS, particular care has been taken with the calibration systems for MERIS to ensure high quality of data.

Table 8 MERIS bands

Band	Band center (nm)	Bandwidth (nm)	Comments
1	412.5	10	Yellow substance, turbidity
2	442.5	10	Chlorophyll-absorption maximum
3	490	10	Chlorophyll, other pigments
4	510	10	Turbidity, suspended sediments, red tides
5	560	10	Chlorophyll reference, suspended sediments
6	620	10	Suspended sediment
7	665	10	Chlorophyll-absorption
8	681.25	7.5	Chlorophyll fluorescence
9	705	10	Atmospheric correction, red edge
10	753.75	7.5	Oxygen absorption reference
11	760	2.5	Oxygen absorption
12	765	5	Oxygen absorption
13	775	15	Aerosols, vegetation
14	865	20	Aerosols correction over the ocean
15	890	10	Water vapor absorption reference
16	900	10	Water vapor absorption

3.2 Algorithms, Including Atmospheric Correction

As noted earlier (section 2), the critical measurement made by an ocean color sensor is the value of the water-leaving radiance L_w in each band in which the sensor measures. These measurements can then be related to the in-water biological constituents that are of interest to the biological oceanographer. The standard algorithms used to produce the global CZCS chlorophyll (pigment) concentration data set that is now widely available[6] are given by McClain et al.[29] They are empirical algorithms that take into account the changes in the chlorophyll concentration. Thus the satellite-determined chlorophyll concentration C_{sat} is given by Equations (5) and (6)

$$C_{sat} = 1.13 \left(\frac{L_w(443)}{L_w(550)} \right)^{-1.71} \quad \text{for } 0 < C_{sat} < 1.5\,\text{mg}\,\text{m}^{-3} \tag{5}$$

and

$$C_{\text{sat}} = 3.33 \left(\frac{L_{\text{w}}(520)}{L_{\text{w}}(550)} \right)^{-2.44} \quad \text{for } 1.5\,\text{mg}\,\text{m}^{-3} < C_{\text{sat}}. \tag{6}$$

The change in algorithm at $C_{\text{sat}} = 1.5\,\text{mg}\,\text{m}^{-3}$ is necessary as the value of $L_{\text{w}}(443)$ becomes too small to quantify accurately at greater concentrations owing to the digitization and signal-to-noise characteristics of CZCS. The estimation of the near-surface pigment concentration using these algorithms is accurate to within 35% for Case 1 waters.[6]

The above algorithms are based on comparisons of the satellite data with in situ measurements and are therefore empirical in nature. More sophisticated algorithms have been developed for CZCS data that rely on modeling the dependence of the water-leaving radiance on the phytoplankton pigment concentration (for example, the so-called semianalytic model of Gordon et al.[20]). In addition, regional algorithms have been developed to improve the retrieval of chlorophyll concentration in specific areas (for example, the Southern Ocean).[30] These types of algorithm have not been applied routinely to CZCS data. The new generation of ocean color sensors, with more bands and improved digitization and S/N, will allow more sophisticated algorithms to be employed and more, and more accurate, biogeochemical information to be recovered from the data.[9,24]

In order to obtain the water-leaving radiance values a simple atmospheric correction algorithm has been applied to the CZCS data set. A standard atmospheric aerosol type was assumed in order to correct for the presence of aerosols in the atmosphere.[6] This correction introduces errors in regions where other types of atmospheric aerosol are present, such as off the north-west African coast, which is affected by Saharan dust. Other atmospheric correction algorithms have been developed and used for specific circumstances, some details of which have been given by Barale and Schlittenhardt.[21] More sophisticated algorithms have been designed for the new generation of ocean color sensors (see, for example, Gordon and Wang[22] for SeaWiFS and Gordon[16] for MODIS). All the algorithms used for ocean color data have to screen the data to eliminate sunglint effects, and some of the more sophisticated algorithms take into account other effects that contribute to the radiances measures by the sensor, such as the presence of whitecaps on the sea surface. It is important to note that as the new generation of ocean color sensors have more radiometric sensitivity than CZCS, this in turn requires a better atmospheric correction algorithm if more accurate measurements of water-leaving radiances are to be made. The extra bands that the new sensors have allow for the atmospheric correction procedure to be much improved over that used for CZCS.[16]

The spatial and temporal coverage of the oceans provided by CZCS was patchy. The data have been averaged to provide so-called higher level products, such as weekly and monthly composites. These can be used more easily to study such phenomena as seasonal variations (section 4.1). Thus at Level 1 there are the individual CZCS images, with calibrated radiances, having a spatial resolution of 1 km. At Level 2 there are the derived geophysical parameters for each CZCS images, at a 4 km spatial resolution, the derived parameters being the phytoplankton pigment concentration, the diffuse attenuation coefficient, normalized water-leaving radiances at 440, 520, and 550 nm, and the aerosol radiance at 670 nm. At Level 3, the data have been binned onto an Earth-grid with about an 18.5-km resolution at the equator. The Level 3 data are available as daily, weekly (5 days) and monthly averages. Full details of these data are given by Feldman et al.[6] SeaWiFS and OCTS data are now becoming available in similar formats over the Internet.

3.3 Calibration and Validation

The calibration and validation of spaceborne ocean color sensors is vital if the data obtained are to be used in any quantitative manner, rather than just as images of the sea. Considerable effort has been and will continue to be devoted to the calibration and validation of the data from ocean color sensors. Particular emphasis on the quality of the atmospheric correction applied to the data is necessary, as this has such a large impact on the retrieved water-leaving radiance values that provide the basic input into all the bio-optical algorithms.[8] As noted in the previous subsection, if the improved radiometric sensitivity of the new sensors means that a better atmospheric correction be applied to the data, this in turn will need to be validated.[31]

The details of the calibration and validation process are rather involved and so will not be discussed in detail here. There are on-board calibration systems that monitor the performance and stability of the sensor (see, for example, Evans and Gordon[8] for CZCS and Barnes et al.[28] for MODIS). More generally, what are required are accurate measurements of the optical properties of the water and of the atmosphere (particularly with regard to atmospheric aerosols), and of the phytoplankton and associated pigments found in the water. Instruments that can be deployed from both ships and buoys have been developed to make such measurements (see, for example, Clark et al.[31]). In addition, measurements of other contributions to the radiance seen by the satellite sensor, such as that due to whitecaps, need to be made. Even if such measurement techniques are available, care

must be taken to make measurements across a range of conditions for the calibration and validation process to be successful and useful. A further consideration is the temporal and spatial sampling of the in situ measurements compared with those made by the satellite. Clearly a ship can survey only a small part of an area that the satellite can see instantaneously, and in the time it takes to carry out the survey conditions may have changed (for example, owing to the currents advecting the phytoplankton around). Similarly how representative are data measured by a buoy at a single point, compared with the typical 1-km square pixel measurement obtained by the satellite, given the spatial variability of the phytoplankton?

The end-result of the calibration and validation efforts for the CZCS global data set was that the accuracy of the pigment concentration retrieval was shown to be 35% in Case 1 waters, and within a factor of two otherwise.[6] The aim for SeaWiFS is to obtain water-leaving radiances to within 5% and chlorophyll-a concentration to within 35% across the range $0.05-50\,\mathrm{mg\,m^{-3}}$.[25] To achieve this a comprehensive calibration and validation plan has been adopted.[32] Similar procedures are being adopted for the other ocean color missions[31] and should provide well-calibrated data for use in scientific studies.

4 APPLICATIONS

The previous sections have given some indication of the complexity of obtaining biological information from remotely sensed ocean color measurements from space. In this section the focus will be on how such measurements may be used to improve our understanding of biological oceanography. All the examples that will be discussed rely on the use of CZCS data. Although OCTS and SeaWiFS have provided and are providing new ocean color data (see section 2), little has yet appeared in the open literature on the application of these data to the study of ocean biology. In addition to scientific applications of the data, which will be the main focus of this section, the commercial applications of the data will also be briefly discussed, as will potential future scientific applications of data from the new generation of ocean color sensors (MODIS, MERIS).

Before proceeding it is useful to define and discuss a number of terms that will be used in this applications sections.[5,13,17] The growth of phytoplankton in the upper layers of the ocean is controlled by the availability of sunlight and nutrients (such as nitrate, silicate, phosphate, and iron) and by predation (the algae being eaten by zooplankton). In general, the phytoplankton may be regarded as passively advected by the turbulent flow in the ocean surface layer, the so-called mixed layer. A phytoplankton bloom occurs when the factors affecting growth are such that rapid growth can occur. For example, blooms may be caused by the injection of nutrients into the mixed layer due to the presence of a cyclonic eddy, which causes local upwelling of nutrient-rich water.[33] The spring bloom occurs in certain parts of the ocean (for example the North Atlantic Ocean) when the mixed layer, deepened by the effects of winter storms, begins to shallow as the ocean begins to heat up in the spring (a process known as restratification). The mixing down of the layer in winter has entrained fresh nutrients into the layer. It has also reduced phytoplankton growth as the phytoplankton have been mixed down by the turbulence in the layer away from the euphotic zone. The shallowing of the layer in spring means that the phytoplankton spend more time in the euphotic zone, allowing them to grow rapidly (abundant sunlight and nutrients). At this stage predation is low because zooplankton numbers are low, so phytoplankton growth outstrips zooplankton grazing and a bloom occurs. As the zooplankton begin to grow rapidly and grazing increases, and the phytoplankton use up the nutrients in the mixed layer, growth begins to slow and then the numbers decay owing to mortality and predation. The zooplankton numbers then decrease (owing to lack of food) and by late summer the bloom has finished. This cycle is repeated each year. Various variations of this cycle are possible, but will not be discussed here (see, for example, Mann and Lazier[5] and Longhurst[17]). Regions of the ocean that have low concentrations of the nutrients required for phytoplankton growth are called oligotrophic, while those that have high nutrient concentrations are called eutrophic. Many of the nutrients necessary for phytoplankton growth (such as CO_2) are available in the ocean in sufficient quantities not to limit growth. The key nutrients necessary for phytoplankton growth that may only be available in low concentrations, and therefore limit growth, are nitrate, silicate, and phosphate (sometimes called macronutrients). In addition, small quantities of so-called micronutrients (trace elements such as iron) are also necessary for the growth of some species of phytoplankton. Ocean areas where phytoplankton growth is limited by lack of iron or some other process (such as grazing by zooplankton), but that have an abundance of the macronutrients are called high nitrate (or nutrient) low chlorophyll (HNLC) areas.

4.1 Measurements of Phytoplankton

As noted in the introduction, the primary component of the biology about which information is obtained from measurements of ocean color is phytoplankton. In this subsection three particular applications of ocean color relating to phytoplankton directly will be considered: measurement of chlorophyll and related pigments, of

primary production, and of coccolithophores (a particular type of phytoplankton). The examples discussed are illustrative rather than comprehensive; other information may be found in the reviews of Abbott and Chelton[34] and Aiken et al.[24]

4.1.1 Chlorophyll and Other Pigments

As discussed earlier (section 2) ocean color measurements give information about the phytoplankton present in the near-surface layers of the ocean due to the presence of chlorophyll-a and other pigments necessary for photosynthesis in the phytoplankton and to associated colored degradation products from the phytoplankton. Thus the basic information provided from the ocean color sensor is a measure of the surface concentration of chlorophyll-a and associated pigments. Relating this concentration to the phytoplankton is a nontrivial exercise,[11] but the satellite chlorophyll concentration measurements in themselves have provided a unique insight into the global biological behavior of the oceans.

Perhaps the simplest observation that has been made has arisen from producing seasonal (spring, summer, autumn, winter) pictures of the global chlorophyll concentrations derived from CZCS data (see, for example, McClain et al.[29] and Figure 1). These pictures show: (1) the occurrence of the spring bloom in the North Atlantic, (2) the relatively constant biological behavior of the Southern Ocean, (3) the low-concentration (desert-like) regions of the subtropical gyres, (4) the high concentrations in the Arabian Sea during the summer monsoon, and (5) the higher concentrations in the northern and tropical Atlantic compared with equivalent regions of the Pacific Ocean. The reasons for these phenomena are related to the ocean physics and will be discussed below (section 4.3). The point to note here is that, while these phenomena had been observed previously from scattered ship-based observations, CZCS data on chlorophyll concentrations provided the first global and spatially coherent view of them.

Yoder et al.[35] have used the monthly CZCS chlorophyll concentration values to look at the biological seasonal cycle in the oceans on a global scale. They averaged the data spatial into latitude bands, defining an equatorial band and northern and southern hemisphere subtropical and subpolar bands. Despite some problems associated with the coverage available from CZCS data, fewer data having been acquired in the southern hemisphere than in the northern, they were able to compare the seasonal changes in the chlorophyll concentration for the Atlantic, Pacific, and Indian Oceans in the appropriate latitude bands. They found that the coverage of the equatorial Atlantic Ocean was poor and concluded that the annual cycle, with a maximum in December, may not be representative. For the equatorial Pacific Ocean they found no seasonal cycle, while for the equatorial Indian Ocean they concluded that the maximum in August/September is related to the subtropical monsoon cycle there. For the subtropical Atlantic and Pacific Oceans, they found that the seasonal cycles are similar, with the winter chlorophyll concentrations approximately double those of the summer. This pattern was attributed to the higher nutrient flux into the mixed layer in winter and the relatively high solar irradiance during winter (compared with higher latitudes). The subtropical northern Indian Ocean was found to be anomalous, with the highest chlorophyll concentrations in the summer months. This was explained by the upwelling of nutrient-rich waters during the summer monsoon. For the subpolar waters of the North Pacific and North Atlantic Oceans they found the existence of the spring bloom, which was more pronounced in the North Atlantic Ocean. This bloom is the result of the increase in solar irradiance in the spring, the shallowing of the mixed layer (due to solar heating and restratification) and the corresponding growth in phytoplankton. This growth outstrips the zooplankton grazing rate initially, but by late spring or summer the phytoplankton losses are greater than their growth and the bloom declines. The subpolar waters of the southern hemisphere were found not to exhibit this pattern. This hemispherical asymmetry may be due to differences in micronutrient (iron) availability, solar irradiance, vertical mixing, or zooplankton grazing. Overall, the seasonal patterns found by Yoder et al.[35] are consistent with predictions based on simple models of predator–prey (zooplankton–phytoplankton) interactions with implicit assumptions about growth limitation by nutrients and solar irradiance. In some respects this agreement is surprising as there is considerable spatial and temporal variability in the distribution of phytoplankton (section 4.3). Therefore the averaging procedure used[35] might have suppressed or distorted any seasonal signal, which it has not. Instead, their results confirm on the large scale a picture of the seasonal behavior of the phytoplankton that was arrived at originally from a more limited set of in situ observations. Banse and English[36] have given a complementary view of seasonal cycles which focuses more on specific areas and considers interannual variability. Their results are in broad agreement with those of Yoder et al.,[35] but show more regional detail as they did not use latitudinal averaging.

Moving from the global to a more regional scale, Sullivan et al.[30] have studied the distribution of phytoplankton blooms in the Southern Ocean (south of 30 °S) using CZCS data, a region poorly sampled by traditional ship-based measurements. They derived a regional pigment retrieval algorithm to improve the estimation of chlorophyll concentration from the CZCS data.

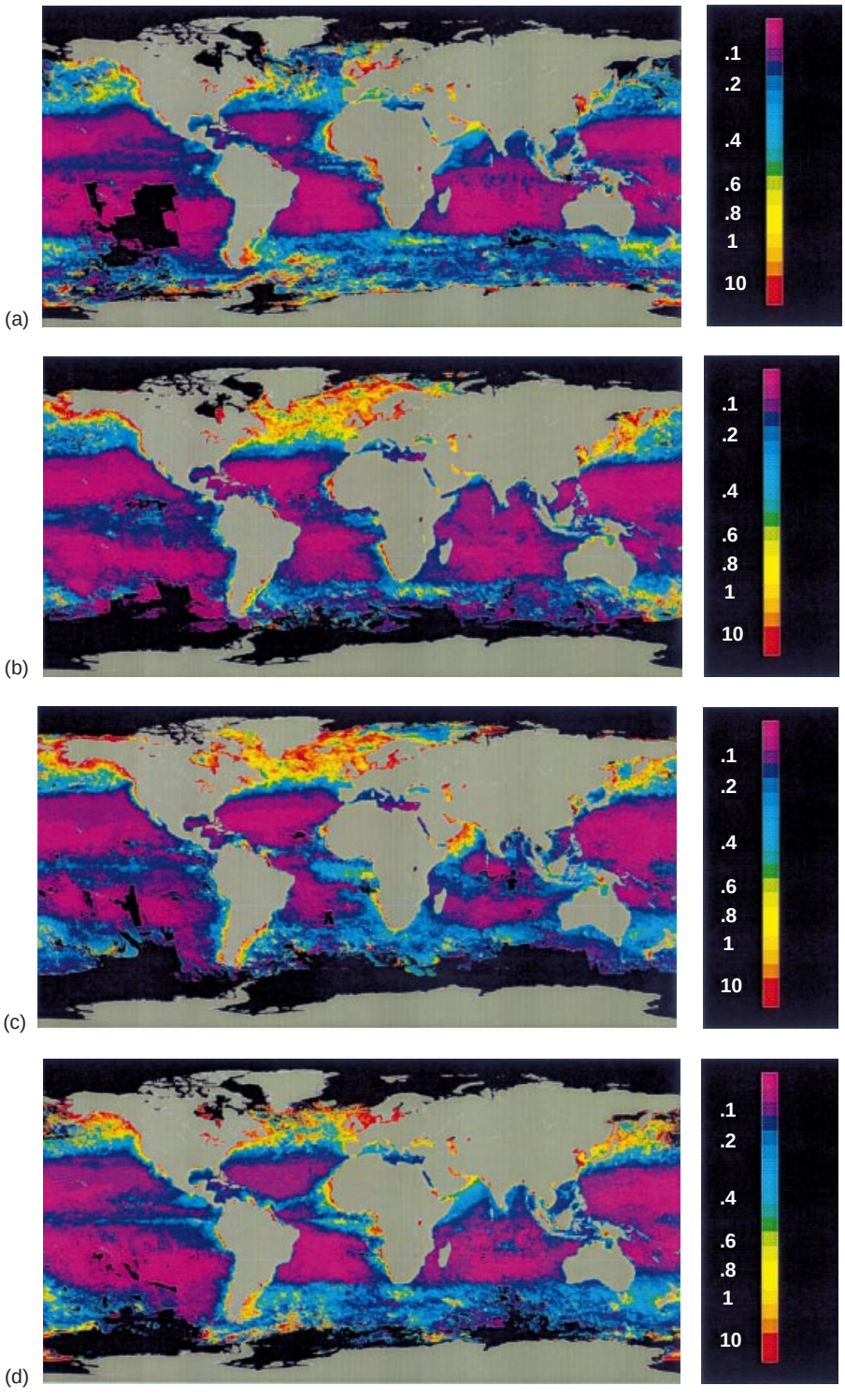

Figure 1 (a) CZCS seasonal sea surface chlorophyll distribution for the northern hemisphere winter months, using data from 1979 to 1986. Note that the color scale is logarithmic in chlorophyll (phytoplankton pigment) concentration (units of mg chlorophyll m^{-3}). Data provided by NASA/Goddard Space Flight Center. (b) As for (a) but for northern hemisphere spring. (c) As for (a) but for northern hemisphere summer. (d) As for (a) but for northern hemisphere autumn. Note the following features of the data shown in the figure: (i) the bloom of phytoplankton (yellow/orange/red) in the North Atlantic Ocean during the northern hemisphere spring, (ii) the bloom of phytoplankton (yellow/orange/red) in the Arabian Sea during the summer monsoon, (iii) the ''desert'' regions, low phytoplankton (purple), in the subtropical gyres, (iv) black indicates missing data. Even using data from 1979 to 1986 gaps exist in the CZCS data record (see section 3.1.1), (v) the data are illustrative of results discussed in sections 4.1.1 and 4.3.1.

They found that blooms were localized to three regions: (1) in shallower waters (near continental margins, islands, and over shoals), (2) in coastal polynyas of the Antarctic sea ice zone, and (3) downstream of the continents (South America, Africa, Australia plus New Zealand) that interrupt the flow of the major circumpolar currents. In relation to the blooms downstream of the continents, they conclude that transport of iron (thought to be the limiting micronutrient in the Southern Ocean) from the adjacent continental shelves stimulates and sustains these blooms. They provide evidence of latitudinal banding of the chlorophyll concentration around the Antarctic continent and link this to the physical processes that occur there. The results obtained[30] are now being used in conjunction with other observations to study the ecology of the Southern Ocean.[37] This shows that the CZCS data are both of intrinsic interest and of value in gaining a better understanding of the ecology of the oceans, in combination with other data. This points the way forward for the use of the new ocean color data that are being and that will be acquired.

Many other examples of the use of chlorophyll concentration data from CZCS could be given; the ones given have been chosen to illustrate the unique value of remotely sensed ocean color data and their potential for understanding the biology of the oceans.

4.1.2 Primary Production

Phytoplankton are the main primary producers of the upper ocean, in that they convert inorganic compounds (nutrients, such as nitrate and silicate) into organic compounds through photosynthesis. This is the beginning of the oceanic food chain and the amount of organic material (biomass) produced is known as the primary production.[13] The rate at which biomass is produced is known as the primary productivity. The net primary production takes into account losses due to respiration and is the amount of photosynthetically fixed carbon available to the next level in the food chain.[38] The primary production may further be subdivided into new and regenerated components. New production is that based on new nutrients that have entered the euphotic zone, while regenerated production is that which occurs due to the recycling of nutrients there (through processes such as microbial breakdown of dead organic matter and fecal pellets). The ratio of new to total primary production is known as the f-ratio.[39] Understanding the oceanic primary production is important for the carbon cycle and the CO_2 problem (see section 4.2). It is also important in terms of assessing the sustainability of the global fisheries,[40] an increasingly vital issue given the increasing world population's requirements for food.

In order to estimate primary production it is necessary to have not only the surface chlorophyll information from ocean color data, but also information about the photosynthesis–light relationship and possibly the structure of the chlorophyll distribution in the vertical. Here two recent attempts to estimate the primary production of the oceans from CZCS data are described, those of Longhurst et al.[39] and Field et al.[38] References to earlier attempts may be found in these papers.

Longhurst et al.[39] use chlorophyll estimates from CZCS and an approach developed by Platt and Sathyendranath[41] to calculate primary production. They divide the ocean up into a number of biogeochemical provinces, based on in situ and satellite data, in order to specify the spatial and temporal variability of the parameters that are needed by their algorithm for primary production. The provinces are an attempt to characterize the biological, chemical, and physical variability of the oceans. As well as the surface chlorophyll value from CZCS, their algorithm requires information about the depth of the chlorophyll maximum, the standard deviation around the peak value, and the ratio of the chlorophyll peak at its maximum to the total peak biomass. The latter information is compiled from an extensive database of ship-based observations. From this information a vertical chlorophyll profile is constructed at each point on a global 1° grid on a quarterly basis (centered on the 15th day of January, April, July, and October). The calculations were restricted to a quarterly basis owing to the sparsity of in situ measurements available. Surface radiation was computed from the sun angle and climatological information from cloud cover. This was combined with experimentally derived information on the photosynthesis–light relationship, representing polar, westerlies, trade-wind, and coastal domains, to calculate the total primary production. The calculation makes no allowance for the presence of suspended sediments in coastal waters for which the CZCS chlorophyll algorithm is inadequate and which will affect the estimates obtained. Taking this and other uncertainties into account and integrating over the year, Longhurst et al.[39] estimate the annual primary production as 44.7–50.24 Gt C per year (1 Gt C is one Gigatonne of carbon = 1 Pg C, one Petagram of carbon = 10^{15} g of carbon). They found that this figure is in reasonable agreement with extrapolations based on a few good in situ measurements.

Field et al.[38] calculated the net primary production for both oceanic and terrestrial biospheres. For the ocean they used a depth integrated model, which requires information about the surface chlorophyll concentration, the depth of the euphotic zone, PAR, and a temperature-dependent maximum chlorophyll-specific carbon fixation rate (the amount of biomass measured in carbon per unit chlorophyll per day). The temperature field is also obtained from satellite data (from AVHRR), as is the solar irradiance (from Bishop and Rossow[42]).

This model is based on the work of Behrenfeld and Falkowski,[43] who suggest that the improvement gained by using a vertically resolved model, such as that of Longhurst et al.,[39] is negligible compared with using a depth-integrated model. Their resulting estimate of global net primary production is 48.5 Gt C per year, which lies in the middle of the range calculated by Longhurst et al.[39] It is interesting to note that while the production is of the order of 50 Gt C per year, the actual phytoplankton biomass is only ca. 1 Gt C. This implies that the phytoplankton biomass turns over approximately once per week on average.[44] This is consistent with the fact that the phytoplankton lifecycle is relatively short, of the order of a day.

Prior to the availability of ocean color data, estimates of global primary production were more difficult to obtain.[39,44] Both of the approaches discussed above rely on more than just the satellite data to obtain these estimates. Neither method is able to distinguish between new and regenerated production, which it is necessary to do in studying carbon fluxes into the ocean (section 4.2). However, Sathyendranath et al.[45] have proposed a method for doing this using satellite data (see section 4.2 below). Algorithms for estimating primary production are being developed for MODIS.[9]

4.1.3 Coccolithophores

Coccolithophores are phytoplankton that form external calcium carbonate $CaCO_3$ scales called coccoliths, which are a few micrometers in diameter and ca. 250–750 nm in thickness. These can form multiple layers, which eventually detach and sink to the sea floor. Coccolithophores are also one of the principal producers of dimethyl sulfide (DMS).[46] Their importance is probably greatest during a bloom, where their concentrations can reach up to 115 million cells per liter.[47] The most abundant of the species is *Emiliania huxleyi*, which can be found throughout most of the world's oceans, with the exception of the polar oceans.[48] They can be detected in satellite imagery because the presence of the coccoliths leads to high reflectance in the surface waters due to their intense scattering of light.[47] Essentially they act like small mirrors suspended in the water and cause a significant portion of the incoming light to be reflected back out from the water.

Whereas phytoplankton pigments change the water-leaving radiance differential across the spectrum owing to absorption, coccolithophore blooms tend to increase the radiance uniformly owing to scattering.[9] The resulting appearance of the ocean can be milky, which is how these blooms were first observed by eye. A consequence of this is that coccolithophore blooms can be detected using the visible channels of the AVHRR, which are not sensitive enough for studying other changes in ocean color.[48] Another consequence is that, in the case of ocean color sensors, a different algorithm needs to be used to estimate their abundance from satellite data.[47] It also means that it is possible to monitor one specific species of phytoplankton, whereas the standard chlorophyll measurements obtained from ocean color do not allow discrimination between different species of phytoplankton. Two effects of the presence of large numbers of coccoliths in the water are first, an increase in the ocean's albedo, and, second, a shading effect that reduces the light level in deeper water (while the scattering of photons increases the light level in the surface waters). These effects have not been studied using satellite data to date (1999).

The most comprehensive study of coccolithophore blooms in the global ocean using ocean color data, from CZCS, is that of Brown and Yoder.[47] They mapped the distribution of blooms using CZCS five-day composite normalized water-leaving radiance data from 1978 to 1986. The data used have a spatial resolution of 20 km. An automatic spectral classification scheme was used to detect the blooms, based on the spectral characteristics that have been obtained from in situ measurements. Monthly and annual composites were calculated from the five-day composite analyses. They found that coccolithophore blooms annually covered an average area of 1.4×10^6 km^2 of the global ocean from 1979 to 1985. This represents ca. 0.5% of the ocean surface. Blooms were most extensive in the subarctic North Atlantic Ocean, annually covering an area of 10^5 km^2 (approximately equivalent to the size of England). As the scattering of light is due to the presence of coccoliths in the water, rather than of the cells themselves, the results are biased towards the declining stage of the blooms when the proportion of coccoliths to cells is greatest. Based on these results they were able to make estimates of $CaCO_3$ and DMS production, using further assumptions regarding the depth of the mixed layer and the concentrations of cells and related chemicals within it. They concluded that on a regional scale the blooms are a significant source of $CaCO_3$ and DMS. In contrast, on a global scale, the blooms detected in CZCS imagery play only a minor role in the production of $CaCO_3$ and DMS and their flux from the mixed layer to deeper waters and to the atmosphere, respectively.

4.2 The CO_2 Problem

Phytoplankton live in the surface sunlit waters of the ocean. When they grow and reproduce, they absorb carbon dioxide and other chemicals (nutrients) from the water. When phytoplankton blooms occur, the surface waters become depleted in CO_2. This induces an

imbalance between the atmosphere and ocean that allows the ocean to absorb CO_2 from the atmosphere. Most of the CO_2 removal is not permanent. Approximately 90% of the phytoplankton die and are decomposed in the surface waters rereleasing their CO_2. The remaining fraction (ca. 10%) sinks down into the deeper ocean where it remains out of contact with the atmosphere for long periods of time (as much as 1000 years). A very small fraction ($\ll$1%) reaches the sea floor where it becomes part of the sea floor sediments and is therefore totally removed from the system (on timescales shorter than geological ones, that is millions of years).[49] Ocean color measurements have been used to look at two parts of this process, the transfer of gases between atmosphere and ocean, and the so-called biological pump that removes the CO_2 to the deep ocean and sediments.

4.2.1 Air–Sea Gas Transfer

The air–sea transfer of gases is dependent on many complex processes including wave breaking, the production of droplets and spray, the entrainment of air bubbles into the water and their subsequent behavior, and the presence or absence of surfactants. Clearly this complexity is too difficult to model in all its aspects. Therefore the net air–sea flux of CO_2 is usually parameterized in terms of a wind speed-dependent gas exchange coefficient multiplied by the difference in the partial pressure of CO_2 (pCO_2) in the air and in the water.[50] The presence of phytoplankton in the surface waters of the ocean, which use the CO_2 for photosynthesis, decreases the pCO_2 of the surface waters allowing greater uptake of CO_2 from the atmosphere. At the present time the potential feedback (whether positive or negative) between changes in climate (increasing atmospheric CO_2) and the phytoplankton are not understood.[1]

In situ measurements in the North Atlantic Ocean[51] have shown that the spatial variability of the oceanic pCO_2 is correlated with the spatial variability of the SST and chlorophyll concentration. These results were obtained in the springtime, during the bloom period. Watson et al.[51] suggest that using satellite-derived SST and chlorophyll values may allow the determination of the pCO_2 value. Subsequent modeling studies by Antoine and Morel[52] (see also Antoine and Morel[53]) have shown that the relationship between pCO_2, SST, and chlorophyll varies spatially and temporally, but may be sufficiently stable on a seasonal basis to make a pCO_2 estimate based on satellite data. To date (1999) this does not appear to have been done.

However, a related study has been carried out by Erickson and Eaton,[54] who studied the flux of carbonyl sulfide from ocean to atmosphere, using CZCS data and an ocean general circulation model. The CZCS chlorophyll data are related empirically to the maximum potential carbonyl sulfide concentration in the surface ocean. The assumption is that the chlorophyll data are representative of the maximum supply of organosulfur compounds that are available for photooxidation. The maximum potential concentration is then related to the actual concentration taking account of the surface radiation field of the ocean. Using an appropriate gas transfer coefficient (see above and Liss and Merlivat[55]) and information on surface radiation, the wind field, and SST, Erickson and Eaton[54] calculate the gas flux for a five-year period, with a 2.8° spatial resolution, and a 24-hour temporal resolution. Computed values of surface concentrations of carbonyl sulfide are said to agree with experimental data on a regional basis to within the uncertainties of the calculation. They found two orders of magnitude variation in the spatial and temporal gas flux, with the comment that the technique is potentially extendible to other biogeochemically important gases, such as CO_2.

To conclude this section, it is worth noting that the gas transfer coefficient itself may be estimated from satellite scatterometer or passive microwave radiometer data on the oceanic wind field. Etcheto et al.[56] have used passive microwave radiometer wind speed data and the Liss and Merlivat[55] formulation of the transfer coefficient's dependence on wind speed to do this. They show that the seasonal variations are large and need to accounted for in calculating the flux of CO_2. Together with satellite-based SST and chlorophyll measurements, this provides the basis for global calculation of the air–sea flux of CO_2 and other biogeochemically important gases.

4.2.2 Biological Pump

The biological pump refers to the process by which part of the of the primary production is removed from the surface layers of the ocean to the ocean interior, as sinking organic particles and as dissolved organic matter. Thus the biological pump reduces the carbon in the surface layers of the ocean (and consequently the CO_2 in the atmosphere) and increases the carbon content of the ocean interior, a small fraction of which becomes sediment on the ocean floor.[1] If the processes controlling the primary production were in a steady state, the export production (that part of the primary production lost from the surface layers) would be equal to the new production due to nutrients entering the upper layers of the ocean (section 4.1.2). Thus knowledge of the f-ratio (new-to-total production) should give some information about the behavior of the biological pump.[57] The estimation of the f-ratio, or equivalently the new and total production, from satellite data has been carried out by Sathyendranath et al.[45] for the Georges Bank area. Using AVHRR SST data, they derived the f-ratio using relationships

between temperature, nitrate, and the f-ratio based on in situ measurements. Combining this with estimates of primary production from CZCS data, they were able to estimate new production. Such an approach may provide information on the biological pump, but the assumption that the processes controlling primary production may be regarded as in a steady state has been questioned.[44] Storms, eddies, El Niño, and other transient physical processes affect the delivery of nutrients to the ocean's surface layers (see section 4.3). On short timescales (up to interannual ones) a steady state cannot be assumed. Over longer timescales an approximate steady-state situation may exist. The processes involved are too complex to be understood using only limited observations (such as SST and ocean color from satellites and some in situ data) and the best way to study them is probably through combining the observations with biophysical models (see section 4.4). Empirical algorithms for estimating new, export, and primary production are being developed for use with MODIS.[9]

4.3 Biophysical Interactions

It is well known that the biology in the oceans is influenced by the physical and chemical processes occurring there.[5,17] These processes occur across the whole range of time- and space-scales present in the ocean, therefore biological distributions in the sea are patchy on a variety of spatial and temporal scales.[58] This patchiness results from interactions between physical and biological factors with the dominant forcing functions changing with scale. Physical circulation dominates basin scale variability and sets the overall context within which biological distributions are correlated to water masses. At the mesoscale, variability is dominated by fronts and eddies[59,60] and at scales below this, variability becomes a function of biological behavior interacting with physical processes such as turbulence and mixing.[61,62] Clearly it is not possible to study the microscale processes, involving turbulence and mixing, with a satellite sensor that has a spatial resolution of ca. 1 km. However, it has proved possible to gain insight into the large scale (gyre scale, O(1000 km)) and mesoscale processes (O(10–200 km)) and these are discussed in the next two subsections.

4.3.1 Large Scale

The links between the large-scale O(1000 km) physical structure of the ocean and its biology has been understood for some time on the basis of in situ measurements made from ships and by using simple models (see, for example, Sverdrup[63] and Mann and Lazier[5]). However, as noted by Longhurst,[17] the distribution of phytoplankton in the ocean was only known in broad terms, while CZCS ocean color data revealed novel information about the global seasonal distribution (see section 4.1.1), despite the fact that CZCS could not directly detect the presence of deep chlorophyll maxima. The basic latitudinal variation in phytoplankton production is due to the decrease in light available for photosynthesis from the equator to the poles and the (inverse) increase in wind mixing which brings nutrients into the surface layers. Both these effects vary seasonally. The ability of the wind mixing to bring nutrients into the surface waters is affected by the degree of stratification (variation of the density gradient with depth) of the ocean. Superimposed on this rather simple picture are the effects of the large-scale ocean currents and gyre circulations, which modify the basic latitudinal variations (Figure 1; see Mann and Lazier[5] and Longhurst[17] for more detail).

A description of the physical effects leading to the spring bloom in the North Atlantic (north of ca. 40 °N) has been given above (see the beginning of the section), and this bloom has been observed in CZCS data.[17,29] It has proved possible to relate the northward progression of the bloom during the spring and summer, as observed in CZCS data, to the latitudinal decrease in the mixed layer depth and increase in light levels over that part of the year.[21] The North Atlantic Ocean is anomalous in having such a strong spring bloom feature.[17] To the south, the subtropical gyre of the North Atlantic Ocean is oligotrophic, with low levels of phytoplankton, owing to the basic downwelling nature of the flow in the center of the gyre which prevents the supply of new nutrients to the surface layers. This is true of the central gyres of all the oceans, as can be seen from the paper by Falkowski et al.[44] (see also Barale and Schlittenhardt[21]), who show the seasonal variation of upper ocean chlorophyll concentrations, from CZCS, alongside the large-scale flow derived from satellite altimeter data. These are the so-called ''desert'' regions in terms of phytoplankton productivity. It can be seen from CZCS data that there are other regions of the ocean that are more productive, but not as much as might be expected. These are the subarctic eastern Pacific Ocean, the eastern equatorial Pacific Ocean and the Southern Ocean. Despite physical processes that supply sufficient macronutrients for high levels of primary production to occur (as is the case in the North Atlantic Ocean), they appear to be limited by the lack of iron (or grazing by zooplankton; see Mann and Lazier[5] and Falkowski et al.[44]). These are the HNLC areas of the ocean.

A specific example of the effect of the large-scale ocean physics on the biology is that of the El Niño in the equatorial Pacific.[64] This has been studied using CZCS data and a model by Halpern and Feldman.[65] For non-El Niño conditions they found results consistent with the HNLC scenario. They were able to observe a

reduction in phytoplankton pigment concentration during the 1982–1983 El Niño due to the smaller nitrate flux into the mixed layer, this being the result of the reduction of the upwelling of nutrient-rich waters when El Niño occurs. It was shown that the amplitude of the annual cycle of pigment concentration was small (almost undetectable), whereas the impact of the El Niño (an interannual variation) was substantial. Owing to the sparsity of CZCS data for the region in 1983–1984 they were unable to examine the effect on the biology of the La Niña event that followed the El Niño. The recent El Niño (1997–1998), one of the strongest on record, has been observed by SeaWiFS and this will no doubt in time provide new insights into the impact of the event on the ocean biology.

The growth in computing power since 1990 has meant that more realistic modeling of the physics of the ocean has been possible and also the development of basin-scale coupled biophysical models. Using such models it is possible to predict the overall seasonal changes in the phytoplankton in the ocean. CZCS data have provided the means to check whether these predictions are valid on temporal and spatial scales that would not be possible using traditional ship-based measurements. One example is the work of Sarmiento et al.[57] who modeled the North Atlantic (20 °S to 60 °N) and found that their predictions of the seasonal variations in the surface chlorophyll concentration matched well overall except for some specific regions. For example, the model predicted higher concentrations of chlorophyll along the equator because the model physics leads to too much nutrient being supplied, allowing more phytoplankton growth than is observed. The Sarmiento et al.[57] model did not have the spatial resolution to account for the effects of eddies, which more recent modeling has been able to do.[66] Mesoscale features, such as eddies, can play an important role in the development of the biology and these will be considered next.

4.3.2 Mesoscale

For the purposes of this discussion the mesoscale is defined as spatial scales in the range O(10–200 km). This scale includes such phenomena as eddies, rings, and fronts in the ocean.[5] One effect of the physics on the biology is through horizontal advection, where the currents move the phytoplankton around, and in some situations trap the phytoplankton in discrete patches of water such as a ring or eddy. Another effect is that of the vertical velocities associated with eddies, rings, and fronts. Upward velocities can bring more nutrients into the euphotic zone, allowing the phytoplankton to grow.[33] Downward velocities can move the phytoplankton out of the euphotic zone thus slowing their growth. A secondary effect may be through the influence of the horizontal and vertical currents on the swimming zooplankton that graze the phytoplankton. As numerous studies of mesoscale effects exist (see the review by Abbott and Chelton),[34] only a few examples will be discussed here.

The first example to be considered is that of warm core rings off the east coast of the USA (see Figure 2). These rings form when Gulf Stream meanders are pinched off and trap the warmer Gulf Stream waters in a ring (or eddy). These rings are typically of O(100 km) in diameter. Brown et al.[18] describe observations of one such ring using CZCS data and SST data from AVHRR. These observations showed that the ring and the Gulf Stream waters, which were warmer than the surrounding waters, had lower concentrations of pigment during a period when a phytoplankton bloom occurred. Garcia-Moliner and Yoder[67] have carried out a more detailed study of pigment concentration in warm core rings in the same region, again using CZCS and AVHRR data. The correspondence between the SST from the AVHRR and the pigment concentration can be seen in the images in their paper. Examining four rings in detail they found that there is decreasing pigment concentration with increasing temperature. They examined a number of hypotheses about the physical processes occurring that affect the biological structure of the ring and conclude that advective entrainment of the surrounding waters at the periphery of the ring is the main factor leading to variability in the pigment concentration. Considerable effort has been devoted to the study of ring processes (see references in Garcia-Moliner and Yoder[67] and Olson[68]) and ocean color data have contributed to these. Weeks and Shillington[69] looked at CZCS imagery around South Africa and showed an image where the Agulhas current and the rings that spin off from it have lower chlorophyll concentration than the surrounding waters. They found an area of high chlorophyll concentration along the east coast of South Africa, in the Benguela upwelling, which brings nutrient-rich waters up to the surface allowing phytoplankton growth to occur.

An area that has been studied in great detail using CZCS data is the California Current system off the west coast of the USA. Denman and Abbott[70] have studied SST from AVHRR and ocean color from CZCS for the area. They used cross-spectrum analysis to look at the timescales of the evolution of the mesoscale features (particularly upwelling jets and the dynamically less active regions between). They concluded that the patterns of pigment concentration, which they observed in the CZCS data, are controlled by the two-dimensional horizontal mesoscale current field of the upper ocean. They restricted their analysis to horizontal length scales greater than 25 km, and noted that for shorter length scales, biological processes (phytoplankton growth, mortality, sinking, and predation by zooplankton) may be important

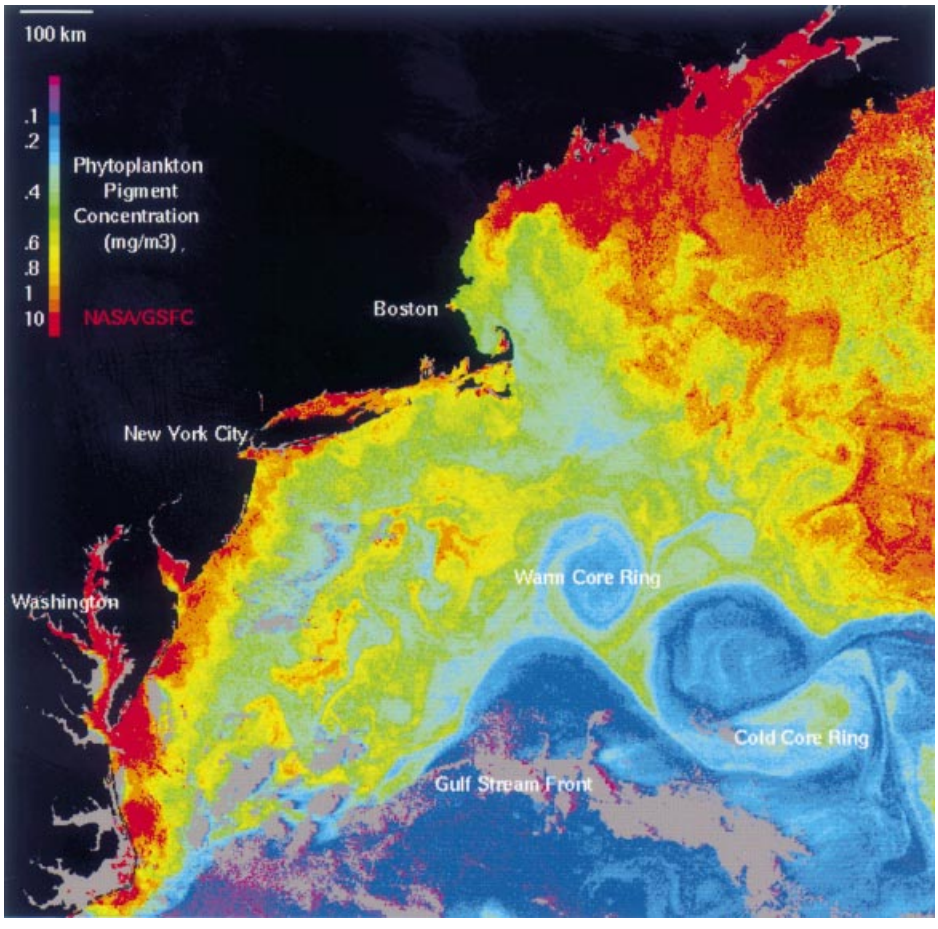

Figure 2 CZCS sea surface chlorophyll (units of mg chlorophyll m^{-3}) image of the east coast of the USA. Clearly visible is the lower productivity water (blue) associated with the warmer waters of the Gulf Stream and with warm core rings shed by the Gulf Stream. The cooler nutrient-rich shelf waters, inshore of the Gulf Stream, show higher productivity (yellow/orange/red). Grey areas denote missing data (primarily due to cloud cover). A variety of mesoscale features – rings, eddies, fronts – can be seen in the image, showing the importance of biophysical interactions for biological productivity. This image is illustrative of the results discussed in section 4.3.2. Data provided by NASA/Goddard Space Flight Center.

for the patterns of patchiness found in the pigment concentrations.

Another example of mesoscale interactions are the filament structures observed in the Portuguese upwelling region by Sousa and Bricaud.[71] The upwelling is driven by local winds and occurs in the period late June to October. CZCS images for the region showed no significant patterns in winter and spring, but during the upwelling period filaments of high chlorophyll concentration were found to extend up to 200 km off the Portuguese coast. These filaments are related to topographic features, which affect the currents flowing in the region of the Portuguese coast. Sousa and Bricaud[71] also examined the relationship between SST and chlorophyll concentration and found variations consistent with the hypothesis that this relationship depended on the stage of the local upwelling event. The scenario was that initially the upwelled cold nutrient-rich water would have low chlorophyll concentration. This water would be advected offshore while the phytoplankton were growing and SST was increasing. The chlorophyll concentration would then decline as the nutrients were exhausted. Similar behavior had previously been observed off the Californian coast.[72] Upwelling regions are known to be highly productive, owing to the supply of nutrients from the deeper waters and are therefore of considerable biological interest.

A secondary application of ocean color data is that of using the biology to infer dynamical information. This has proved useful in a number of situations, a particular example being the case where, because solar heating of the near surface layer leads to uniform ocean surface temperatures, the SST signal of ocean features has not been visible in the IR satellite data. As the color signal (if it exists) is unaffected by such an effect then dynamical structures, such as eddies, can still be detected despite

their lack of a SST signature. Thompson and Gower[73] give an example of an eddy near Vancouver Island that was detected by its CZCS ocean color signal, but could not be seen in the IR AVHRR image. Another approach to obtaining dynamical information from ocean color images is that of feature tracking using successive images (see, for example, Garcia and Robinson[74]). The implicit assumption behind this approach is that the quantity being measured that defines the feature, in this case ocean color, acts as a passive tracer that is simply advected by the ocean currents. As the ocean color is a function of the phytoplankton in the surface waters, a quantity that can be changing rapidly, this assumption is only true on short timescales. Therefore this means of obtaining dynamical information has not been used very much.

4.3.3 Influence of Biology on the Physics

In the previous two subsections the influence of ocean physics on the ocean biology has been considered in detail as this is the primary biophysical interaction. However, a secondary interaction is the effect of the presence of phytoplankton on the solar heating of the surface layers of the ocean.[75,76] The presence of phytoplankton can affect the absorption of solar irradiance by the ocean in different way, depending on their concentration and distribution with depth. Ocean color data from CZCS have been used to study this effect for the equatorial Pacific Ocean.[77] Lewis et al.[77] conclude that some of the discrepancies between observed and modeled SSTs for the equatorial Pacific Ocean may be due to the neglect of penetrating solar irradiance in the models. In order to allow for this effect it is necessary to include biology as well as physics in the models. Some recent coupled biophysical models (for example, that of Oschlies and Garcon[66] for the North Atlantic Ocean) take account of this effect.

4.4 Assimilation of Data into Models

For many and various reasons, perhaps most importantly understanding the role of the oceans in the carbon cycle and the influence of anthropogenic CO_2 emissions into the atmosphere on that cycle,[50] considerable interest exists in developing robust oceanic ecosystem models.[78] By coupling these biological models to physical models of the ocean circulation, the aim is to provide a description, and possibly a prediction, of the contribution of the ocean to global biogeochemical cycles, such as the carbon cycle.[79,80] Another application of the models would be to understand the behavior of the oceanic food chain better, including fisheries.[81]

In attempting to carry this out two problems become apparent. First, in contrast to the modeling of the atmosphere or ocean, where a basic description of the physics is provided by the Navier–Stokes equations of fluid dynamics,[82] there is no basic set of equations that describe the ocean ecosystem. Therefore a heuristic approach is generally taken, where sufficient biological components are modeled to describe the problem of interest.[78] For example, a simple three-component (also called a three-compartment) oceanic ecosystem model might include phytoplankton, zooplankton, and nutrients and has been used with a degree of success for some studies.[83] The second problem is related to the first, in that such heuristic models have many parameters that are known to varying degrees of accuracy (for example, the seven-compartment model of Fasham et al.[84] has 27 parameters). This limits the potential applicability of the models, unless the parameters can be better determined. This provides an ideal application for data assimilation in its guise of fitting models to data.[85]

Another purpose of data assimilation is that of improving the predictive power of models. This is a familiar application in the context of meteorological forecasting. It is also beginning to be used more widely in the area of physical ocean modeling, but is in its infancy with regard to assimilation of data into biological models. Many techniques have been developed for the assimilation of data into models of atmosphere and ocean physics.[86,87] The applicability of these techniques to oceanic ecosystem models is by no mean automatic and much work remains to be done in this area.

As has been made clear thus far, data assimilation is useful for improving both descriptive and predictive powers of models. However, for this to be possible it is necessary to have good measurements of the biology to assimilate into the ecosystem model. Such measurements exist at a limited number of sites (such as Bermuda),[88] or for particular experiments (such as the North Atlantic Bloom Experiment – NABE),[89] but these may not be representative of what is happening in the ocean basin or on the global scale. The availability of ocean color data from OCTS and from SeaWiFS provides measurements that might help in studying these larger scales. The limitation is that only one component (phytoplankton) of the ecosystem is measured, but in combination with a number of in situ measurements that might be sufficient to improve the models, through data assimilation.

The first person to assimilate data into a coupled biophysical model appears to have been Ishizaka,[90] for the southeastern USA continental shelf area. The data assimilated came from CZCS and the assimilation was of the predictive type. The model was a four-compartment ecosystem model (nutrient, phytoplankton, zooplankton, detritus), vertically integrated with horizontal advection and eddy diffusion, but with a semiempirical upwelling/downwelling term added.[91] The coupling to the physics is through the advective velocities, which

Ishizaka[92] obtained from optimally interpolated circulation fields. Ishizaka's assimilation procedure is that of direct insertion, where the phytoplankton values in the model were simply replaced by those estimated from the CZCS ocean color data. Three types of adjustment were applied to the other compartments (nutrient, zooplankton, detritus) to allow for conservation of biomass, and this seemed to make little difference to the results obtained. The model was run forward from the assimilation time and the results compared with CZCS data at later times (comparisons made over a period of a few days). Two cases were considered, one in which the biological interaction and upwelling/downwelling terms were switched off, so that the resulting biological distributions were due purely to the physical advective and diffusive processes, and one where the biological interactions were included. The results suggested that advection dominated in this case, with biological processes acting as a secondary factor. However, the impact of the data assimilation lasted only a few days, indicating the need to assimilate data frequently (every 1–2 days) to keep the model updated and the errors in prediction small. Assimilation of the data had a positive effect overall, but did degrade some aspects of the model (estimates of nutrient fluxes as compared with in situ data). Ishizaka[90] also tested the impact of only assimilating data over part of the region modeled, thus simulating the effect of partially cloudy data. This too seemed to give improved results, though it did cause discontinuities at the boundaries between the areas where data were available for assimilation and where there were none. Direct insertion of data is known to cause problems in the physical model, as the model adjusts by radiation of waves,[86] but this did not appear to cause problems for the advective–diffusive biological model used by Ishizaka.[90] One observation to be made about the study is that the use of a vertically integrated model avoids the issue of how to relate the satellite surface observation of chlorophyll to the subsurface structure in the biology (there being none in this model, as there is no depth dependence). Subsequently, Ishizaka[93] has reported some preliminary results on the use of direct insertion with simpler ecosystem models. He concludes that, ''. . .assimilation of phytoplankton data into a simple time-dependent model is not straightforward. Differences in error sources, methods of assimilation, and the timing of the assimilation all result in different model solutions. . . Furthermore, much *more* complex ecosystem structure is preferable for the real simulation and *this* increases uncertainty of the effects of the data assimilation.''

Sarmiento et al.[57] compared their results for the seasonal chlorophyll distribution of the Atlantic Ocean, obtained by embedding the Fasham et al.[84] ecosystem model in a ocean general circulation model, with data from CZCS (of the type shown in Figure 1). They found reasonable agreement overall, but some specific discrepancies. Armstrong et al.[94] followed this up by attempting to assimilate CZCS data into the same model at frequencies of 1 and 5 day^{-1}. Comparisons improved overall compared with the unforced case of Sarmiento et al.,[57] except at high latitudes, with the more frequent forcing giving better results. Sarmiento et al.[51] attributed the discrepancy at high latitudes as being related to the modeling of the zooplankton grazing of the phytoplankton (see their paper for details). They therefore incorporated multiple grazing chains in the model, giving it a total of 13 compartments (nitrate, ammonium, bacteria and dissolved organic nitrogen, plus three compartments each for detritus, phytoplankton, and zooplankton).[94] Assimilation of CZCS data at the same frequencies as before led to much better agreement. The assimilation technique used was ''nudging'',[86] applied to bring the model chlorophyll values in the upper layer of the model (top 10 m) towards the surface observations from CZCS. This procedure requires the model to adjust the subsurface biological distributions and violates the conservation of biomass, which is either added or removed from the model depending on whether the satellite–model difference in chlorophyll is positive or negative. Simplistically one might think that forcing the model values towards the observations would eventually result in complete agreement, but this is not the case owing to deficiencies in the model physics and biology. This is a case of using data assimilation to fit a model to data and thus to improve the model.

The previous two models discussed were two-dimensional (horizontal) and three-dimensional in space. A one-dimensional (vertical) model has been developed by Prunet et al.[95] This is a 10-compartment ecosystem model coupled to a one-dimensional mixed layer[95] and the data assimilated are from Station Papa (Station P; 50 °N, 145 °W). The difference from the other studies is that Prunet et al.[96] assimilated data into both the biological and the physical components of their model. A variational assimilation technique was used. In their first paper[95] they considered assimilating only surface chlorophyll data into the model and found that this can only partially constrain the parameters of the model and does not constrain the vertical structure of the chlorophyll. In their second paper[96] they used a simpler four-compartment ecosystem model (nitrate, phytoplankton, zooplankton, detritus) coupled to the same mixed layer model as in their first paper.[95] They found that surface chlorophyll assimilation is not sufficient to reproduce the seasonal cycle of surface chlorophyll, temperature, and nitrate in a robust manner. By additionally assimilating surface nitrate and temperature the model is

improved. Comparison between the two models suggests that choice of model structure can affect the results obtained. Although satellite ocean color data have not been used in these studies, they are important as they show that satellite-derived surface chlorophyll values on their own may not provide sufficient information to constrain models through data assimilation. Despite this, it is undoubtedly the case that as new ocean color data become available from the next generation of sensors, they will be used in assimilation studies (along with other data) to improve the biophysical modeling of the oceans.

4.5 Dimethyl Sulfide, Climate, and Gaia

It is known that various species of phytoplankton produce dimethyl sulfoniopropionate (DMSP), a precursor of DMS.[1,97] Some time ago Charlson et al.[98] proposed a link between the DMS originating in the ocean due to phytoplankton and the formation of cloud condensation nuclei (CCN) in the atmosphere. The DMS is oxidized in the atmosphere to form sulfate particles which act as CCN. Changing the CCN concentration affects the formation of clouds, which in turn influences cloud albedo and hence the climate. Changes in the climate in turn affect the phytoplankton in the ocean, thus providing a feedback mechanism. Under the Gaia hypothesis of the Earth as a self-regulating system,[97,99] the increased albedo of the earth would reduce the temperature and light beneath the clouds, and so reduce photosynthesis and the growth of DMS-producing phytoplankton. Less DMS would in turn mean fewer CCN, fewer clouds, and an increase in light allowing more phytoplankton to grow. These feedbacks would allow the climate system to remain in balance.

In order to understand this process better, attempts have been made to estimate the flux of DMS from the ocean to the atmosphere. One of these has made use of CZCS ocean color data.[100] Using an empirical relationship between the CZCS chlorophyll data and DMS, together with a gas transfer coefficient estimated using climatological winds (see section 4.2.1), Thompson et al.[100] are able to estimate the flux of DMS from atmosphere to ocean. They note that the complexity of the phytoplankton production of DMS means that a valid relationship between the two will only exist in certain regions and at certain seasons. This is due to the fact that the production of the DMS precursor DMSP is highly species-specific.[1] There are also considerable (factor 2) uncertainties associated with the DMS gas transfer coefficient.[101] Nevertheless Thompson et al.[100] were able to obtain estimates that are consistent, for a specific region and season, with in situ measurements and an estimate based on photochemical model and atmospheric measurements. They note that future ocean color sensors with increased spectral resolution may enable a more robust determination of the relationship between phytoplankton and DMS to be established, by providing quantitative information on species abundance or on the productivity of those phytoplankton groups that are strong producers of DMS. It should be noted that anthropogenic emissions of sulfur far exceed those from phytoplankton. Therefore, their impact on the climate system, particularly in the industrialized northern hemisphere, is greater.[1,5]

4.6 Commercial Application – Fisheries

Although the primary focus of this article has been on the use of remotely sensed ocean color data for research into biological oceanography, it is worthwhile mentioning briefly that the data have also been successfully used for commercial purposes. The importance of commercial applications of the data is evidenced by the fact that SeaWiFS was launched and is operated by a commercial company OSC. The primary commercial application of ocean color data is in fisheries.[102] It has been used, often in conjunction with SST data, to guide fishermen and fishing vessels to areas where fish might be found. This use is based on the food chain principle that an abundance of phytoplankton leads to an abundance of zooplankton that feed on them. This in turn leads to an abundance of fish that eat the zooplankton. Thus the presence of high levels of phytoplankton, as measured by changes in ocean color, is taken to indicate the potential presence of fish. More sophisticated approaches that take into account the preferences of certain fish species for particular temperature conditions (hence the use of SST in conjunction with ocean color), can be used to determine where to fish.[103] Even in a situation where fish catches are limited by legislation (as is the cases for European Community countries), the time and fuel saved in searching for fish may improve the profitability of fishing, a tangible benefit to the fisherman or fishing fleet operator. SeaWiFS data are being used to provide such services for fishermen, as CZCS data were in the past.

4.7 Possible Future Applications

In this final section a number of possible future applications of ocean color measurements from space will be briefly discussed. These are not exhaustive, but are given here to show the potential for the use of ocean color data. Despite the many existing applications that have been discussed above, which were based primarily on the use of CZCS data, the data from the new improved ocean color sensors (OCTS, SeaWiFS, MODIS, MERIS) will doubtless lead to the development of many new applications.

4.7.1 Discrimination of Different Phytoplankton Species

As discussed in section 4.1.3 it has already proved possible to gain information from ocean color measurements on one particular species of phytoplankton, namely coccolithophores. One paper[104] has shown that it may be possible to detect blooms of the cyanobacterium *Trichodesmium* (a species of phytoplankton), using CZCS data. With the increased spectral discrimination and more bands that will be available from future sensors (MODIS, MERIS) it may be possible to discriminate between different species on the basis of their photosynthetic pigments. Different combinations and amounts of pigments in different species will produce varying responses in the light spectrum, as measured by the color sensors. The standard algorithms being proposed for MODIS include one for the detection of coccolithophores.[9] No doubt other algorithms will be developed as the data become available, as was the case for CZCS.

4.7.2 Measurement of Other Biochemical Constituents

In the earlier discussion of applications (sections 4.1, 4.2, and 4.5) it was shown that even the rather limited CZCS data were able to provide some information on a variety of biochemical constituents apart from chlorophyll-a. The new generation of sensors should allow further studies of DMS, carbonyl sulfide, pCO_2, and other chemical constituents related to the phytoplankton. MODIS algorithms have been developed to retrieve information on the algal pigment phycoerythrin.[9] In addition, MODIS and MERIS are capable of measuring the phytoplankton fluorescence peak at around 685 nm (see Tables 6 and 8, and Esaias et al.[9]). Of the light energy absorbed by phytoplankton chlorophyll, a small amount O(1%) is re-emitted as fluorescence.[11] This provides an alternative means of measuring chlorophyll-a, as demonstrated by Neville and Gower[105] using airborne measurements. Again as data become available, more new algorithms will be developed.

4.7.3 Use of Data with Models and in situ Measurements

One area that has been hampered by the lack of ocean color data is that of the use of data with models and in situ measurements. This is a developing area of study, both in terms of the use of data for model validation, and for the improvement of models by data assimilation (see section 4.4). Computer power has increased in such a way that high-resolution coupled biophysical models of entire ocean basins are now possible. Ocean color data can contribute to the validation of such models,[57] to the study of specific processes,[66,106] and to the improvement of the models through data assimilation.[94]Interestingly all these examples are of the North Atlantic Ocean, one of the most studied ocean basins. The improved quality of the data from the new generation of sensors will add impetus to such studies. However, there will be a continuing need to make in situ observations to complement those from ocean color sensors, in order to understand and model other components of the ecosystem and their changes with depth.

4.7.4 Monitoring of Harmful Algal Blooms

The environmental and economic impact of harmful algal blooms (HABs) have increased as human activities influence the coastal ecosystem. Various types of phytoplankton produce toxins that can poison fish, thus affecting fisheries and commercial fish farming, or through shellfish that can enter the food chain and cause the poisoning of human beings (for a recent review see Richardson[107]). The early detection of such blooms could help in preventing or alleviating some of the consequences that follow. In some literature HABs are referred to as ''red tides'' indicating that the presence of these harmful phytoplankton changes the water color to red or brown. The resulting color change can in principle be detected using ocean color measurements, though it may not be possible to distinguish changes in color due to harmful and nonharmful phytoplankton.[108] The new generation of ocean color sensors provides the opportunity to detect HABs from space and to monitor their development and impact.

4.8 Afterword

The preceding review of the use of remotely sensed ocean color data to study biological oceanography makes no claims to be comprehensive or exhaustive (at least a book would be required to do the job properly). Instead the aim has been to give an overview of what has been achieved using the somewhat limited CZCS data set that has been available. The potential of data that are becoming, or will become, available from the new generation of satellite ocean color sensors (OCTS, SeaWiFS, MODIS, MERIS) has also been indicated. The availability of these data mean that an exciting period in the development of satellite-based biological oceanography is just beginning.

ACKNOWLEDGMENTS

I am grateful to Paolo Cipollini for helpful comments on the first draft of this article.

ABBREVIATIONS AND ACRONYMS

ADEOS	Advanced Earth Observation Satellite

AOP	Apparent Optical Properties
AVHRR	Advanced Very High Resolution Radiometer
CCN	Cloud Condensation Nuclei
CZCS	Coastal Zone Color Scanner
DLR	German Space Agency
DMS	Dimethyl Sulfide
DMSP	Dimethyl Sulfoniopropionate
EOS	Earth Observing System
ESA	European Space Agency
GAC	Global Area Coverage
HAB	Harmful Algal Bloom
HNLC	High Nitrate (or Nutrient) Low Chlorophyll
IFOV	Instantaneous Field of View
IOP	Inherent Optical Properties
IR	Infrared
LAC	Local Area Coverage
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MOS	Modular Optoelectronic Scanner
NASA	National Aeronautical and Space Administration
NASDA	National Space Development Agency of Japan
NEΔT	Noise Equivalent Temperature Difference
OCI	Ocean Color Imager
OCTS	Ocean Color and Temperature Sensor
OSC	Orbital Sciences Corporation
PAR	Photosynthetically Active (or Available) Radiation
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
S/N	Signal-to-noise Ratio
SST	Sea Surface Temperature

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors • High-performance Liquid Chromatography of Biological Macromolecules

Environment: Trace Gas Monitoring **(Volume 3)**
Airborne Instrumentation for Aerosol Measurements

Field-portable Instrumentation **(Volume 4)**
Chemical-sensing Networks: Satellite-based

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Global Land Databases for Environmental Analyses • Hyperspectral Remote Sensing: Data Collection and Exploitation • Polar Environments Assessment by Remote Sensing • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy • Fluorescence Lifetime Measurements, Applications of

REFERENCES

1. K. Denman, E. Hofmann, H. Marchant, 'Marine Biotic Responses to Environmental Change and Feedbacks to Climate', in *Climate Change 1995: the Science of Climate Change*, eds. J.T. Houghton, L.G. Meira Fihlo, B.A. Callander, N. Harris, A. Kattenberg, K. Maskell, Cambridge University Press, Cambridge, 483–516, 1996.
2. I.S. Robinson, *Satellite Oceanography: An Introduction for Oceanographers and Remote-sensing Scientists*, Ellis Horwood, Chichester, 1985.
3. R.H. Stewart, *Methods of Satellite Oceanography*, University of California Press, Berkeley, 1985.
4. C.S. Nilsson, P.C. Tildesley, 'Imaging of Oceanic Features by ERS 1 Synthetic Aperture Radar', *J.Geophys. Res.*, **100**, 953–967 (1995).
5. K.H. Mann, J.R.N. Lazier, *Dynamics of Marine Ecosystems: Biological–Physical Interactions in the Oceans*, 2nd edition, Blackwell Science, Oxford, 1996.
6. G. Feldman, N. Kuring, C. Ng, W. Esaias, C. McClain, J. Elrod, N. Maynard, D. Endres, R. Evans, J. Brown, S. Walsh, M. Carle, G. Podesta, 'Ocean Color: Availability of the Global Data Set', *Eos, Trans. Am. Geophys, Union*, **70**, 634–635, 640–641 (1989).
7. G.L. Clarke, G.C. Ewing, C.J. Lorenzen, 'Spectra of Backscattered Light from the Sea Obtained from Aircraft as a Measurement of Chlorophyll Concentration', *Science*, **167**, 1119–1121 (1970).
8. R.H. Evans, H.R. Gordon, 'Coastal Zone Color Scanner, ''System Calibration'': A Retrospective Examination', *J. Geophys. Res.*, **99**, 7293–7307 (1994).
9. W.E. Esaias et al., 'An Overview of MODIS Capabilities for Ocean Science Observations', *IEEE Trans. Geosci. Remote Sensing*, **36**, 1250–1265 (1998).
10. M.A. Rast (ed.), *MERIS: The Medium Resolution Imaging Spectrometer*, ESA, Noordwijk, 1996.
11. J.T.O. Kirk, *Light and Photosynthesis in Aquatic Ecosystems*, 2nd edition, Cambridge University Press, Cambridge, 1996.
12. P.G. Falkowski, J.A. Raven, *Aquatic Photosynthesis*, Blackwell Science, Abingdon, 1997.
13. C.M. Lalli, T.R. Parsons, *Biological Oceanography: An Introduction*, 2nd edition, Butterworth-Heinemann, Oxford, 1997.
14. C.D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*, Academic Press, San Diego, 1994.

15. H.R. Gordon, R.W. Austin, D.K. Clark, W.A. Hovis, C.S. Yentsch, 'Ocean Color Measurements', *Adv. Geophys.*, **27**, 297–333 (1985).
16. H.R. Gordon, 'Atmospheric Correction of Ocean Color Imagery in the Earth Observing System Era', *J. Geophys. Res.*, **102**, 17081–17106 (1997).
17. A. Longhurst, *Ecological Geography of the Sea*, Academic Press, San Diego, 1998.
18. O.B. Brown, R.H. Evans, J.W. Brown, H.R. Gordon, R.C. Smith, K.S. Baker, 'Phytoplankton Blooming off the US East Coast: A Satellite Description', *Science*, **229**, 163–167 (1985).
19. A. Morel, L. Prieur, 'Analysis of Variations in Ocean Color', *Limnol. Oceanogr.*, **22**, 709–722 (1977).
20. H.R. Gordon, O.B. Brown, R.H. Evans, J.W. Brown, R.C. Smith, K.S. Baker, D.K. Clark, 'A Semi-analytical Radiance Model of Ocean Color', *J. Geophys. Res.*, **93**, 10909–10924 (1988).
21. V. Barale, P.M. Schlittenhardt (eds.), *Ocean Color: Theory and Applications in a Decade of CZCS Experience*, Kluwer Academic, Dordrecht, 1993.
22. H.R. Gordon, M. Wang, 'Retrieval of Water-leaving Radiances and Aerosol Optical Thickness over the Oceans with SeaWiFS: A Preliminary Algorithm', *Appl. Optics.*, **33**, 443–452 (1994).
23. P.M. Holligan, T. Aarup, S.B. Groom, 'The North Sea: Satellite Color Atlas', *Cont. Shelf Res.*, **9**, 667–765 (1989).
24. J. Aiken, G.F. Moore, P.M. Holligan, 'Remote Sensing of Oceanic Biology in Relation to Global Climate Change', *J. Phycol.*, **28**, 579–590 (1992).
25. S.B. Hooker, C.R. McClain, A. Holmes, 'Ocean Color Imaging: CZCS to SeaWiFS', *Mar. Tech. Soc. J.*, **27**, 3–15 (1993).
26. S. Saitoh, 'OCTS on ADEOS', in *Oceanographic Applications of Remote Sensing*, eds. K. Ikeda, F. Dobson, CRC Press, Boca Raton, FL, 473–480, 1995.
27. S.B. Hooker, W.E. Esaias, 'An Overview of the SeaWiFS Project', *Eos, Trans. Am. Geophys. Union*, **74**, 241, 245–246 (1993).
28. W.L. Barnes, T.S. Pagano, V.V. Salomonson, 'Prelaunch Characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1', *IEEE Trans. Geosci. Remote Sensing*, **36**, 1088–1100 (1998).
29. C.R. McClain, G. Feldman, W. Esaias, 'Oceanic Biological Productivity', in *Atlas of Satellite Observations Related to Global Change*, eds. R.J. Gurney, J.L. Foster, C.L. Parkinson, Cambridge University Press, Cambridge, 251–263, 1993.
30. C.W. Sullivan, K.R. Arrigo, C.R. McClain, J.C. Cosimo, J. Firestone, 'Distributions of Phytoplankton Blooms in the Southern Ocean', *Science*, **262**, 1832–1837 (1993).
31. D.K. Clark, H.R. Gordon, K.J. Voss, Y. Ge, W. Broenkow, C. Trees, 'Validation of Atmospheric Correction over the Oceans', *J. Geophys. Res.*, **102**, 17209–17217 (1997).
32. C.R. McClain, W.E. Esaias, W. Barnes, B. Guenther, D. Endres, S.B. Hooker, G. Mitchell, R. Barnes, *Calibration and Validation Plan for SeaWiFS, NASA Technical Memorandum 104566* Vol. 3, NASA Goddard Space Flight Center, Maryland, 1992.
33. D. McGillicuddy, A.R. Robinson, D.A. Siegel, H.W. Jannasch, R. Johnson, T.D. Dickey, J. McNeil, A.F. Michaels, A.H. Knap, 'Influence of Mesoscale Eddies on New Production in the Sargasso Sea', *Nature*, **394**, 263–266 (1998).
34. M.R. Abbott, D.B. Chelton, 'Advances in Passive Remote Sensing of the Ocean', *Rev. Geophys.*, Suppl., 571–589 (1991).
35. J.A. Yoder, C.R. McClain, G.C. Feldman, W.E. Esaias, 'Annual Cycles of Phytoplankton Chlorophyll Concentrations in the Global Ocean: A Satellite View', *Global Biogeochem. Cyc.*, **7**, 181–193 (1993).
36. K. Banse, D.C. English, 'Seasonality of Coastal Zone Color Scanner Phytoplankton Pigment in the Offshore Oceans', *J. Geophys. Res.*, **99**, 7323–7345 (1994).
37. C.T. Tynan, 'Ecological Importance of the Southern Boundary of the Antarctic Circumpolar Current', *Nature*, **392**, 708–710 (1998).
38. C.B. Field, M.J. Behrenfeld, J.T. Randerson, P. Falkowski, 'Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components', *Science*, **281**, 237–240 (1998).
39. A. Longhurst, S. Sathyendranath, T. Platt, C. Caverhill, 'An Estimate of Global Primary Production in the Ocean from Satellite Radiometer Data', *J. Plankton Res.*, **17**, 1245–1271 (1995).
40. D. Pauly, V. Christensen, 'Primary Production Required to Sustain Global Fisheries', *Nature*, **374**, 255–257 (1995).
41. T. Platt, S. Sathyendranath, 'Oceanic Primary Production: Estimation by Remote Sensing at Local and Regional Scales', *Science*, **241**, 1613–1620 (1988).
42. J.K. Bishop, W.B. Rossow, 'Spatial and Temporal Variability of Surface Solar Irradiance', *J. Geophys. Res.*, **96**, 16839–16858 (1991).
43. M.J. Berhenfeld, P.G. Falkowski, 'A Consumer's Guide to Phytoplankton Primary Productivity Models', *Limnol. Oceanogr.*, **42**, 1479–1491 (1997).
44. P.G. Falkowski, R.T. Barber, V. Smetacek, 'Biogeochemical Controls and Feedbacks on Ocean Primary Production', *Science*, **281**, 200–206 (1998).
45. S. Sathyendranath, T. Platt, E.P.W. Horne, W.G. Harrison, O. Ulloa, R. Outerbridge, N. Hoepffner, 'Estimation of New Production in the Ocean by Compound Remote Sensing', *Nature*, **353**, 129–133 (1991).
46. M.D. Keller, W.K. Bellows, R.R.L. Guillard, *Dimethyl Sulfide Production in Marine Phytoplankton: Review*, American Chemical Society Symposium Series, 167–182, Vol. 393, 1989.
47. C.W. Brown, J.A. Yoder, 'Coccolithophorid Blooms in the Global Ocean', *J. Geophys. Res.*, **99**, 7467–7482 (1994).

48. P.M. Holligan, et al., 'A Biogeochemical Study of the Coccolithophore *Emiliania huxleyi* in the North Atlantic', *Global Biogeochem. Cyc.*, **7**, 879–900 (1993).
49. W.H. Schlesinger, *Biogeochemistry: An Analysis of Global Change*, 2nd edition, Academic Press, San Diego, 1997.
50. U. Siegenthaler, J.L. Sarmiento, 'Atmospheric Carbon Dioxide and the Ocean', *Nature*, **365**, 119–125 (1993).
51. A.J. Watson, C. Robinson, J.E. Robinson, P.J. le B. Williams, M.J.R. Fasham, 'Spatial Variability in the Sink for Atmospheric Carbon Dioxide in the North Atlantic', *Nature*, **350**, 50–53 (1991).
52. D. Antoine, A. Morel, 'Modelling the Seasonal Course of the Upper Ocean pCO_2(II). Validation of the Model and Sensitivity Studies', *Tellus*, **47B**, 122–144 (1995).
53. D. Antoine, A. Morel, 'Modelling the Seasonal Course of the Upper Ocean pCO_2(I). Development of A One-dimensional Model', *Tellus*, **47B**, 103–121 (1995).
54. D.J. Erickson, B.E. Eaton, 'Global Biogeochemical Cycling Estimates with CZCS Satellite Data and General Circulation Models', *Geophys. Res. Lett.*, **20**, 683–686 (1993).
55. P.S. Liss, L. Merlivat, 'Air–sea Gas Exchange Rates: Introduction and Synthesis', in *The Role of Air–Sea Gas Exchange in Geochemical Cycling*, ed. P. Baut-Menard, D. Reidel, Norwood, 113–128, 1986.
56. J. Etcheto, J. Boutin, L. Merlivat, 'Seasonal Variations of the CO_2 Exchange Coefficient over the Global Ocean Using Satellite Wind Speed Measurements', *Tellus*, **43B**, 247–255 (1991).
57. J.L. Sarmiento, R.D. Slater, M.J.R. Fasham, H.W. Ducklow, J.R. Toggweiler, G.T. Evans, 'A Seasonal Three-dimensional Ecosystem Model of Nitrogen Cycling in the North Atlantic Euphotic Zone', *Global Biogeochem. Cyc.*, **7**, 417–450 (1993).
58. L.R. Haury, J.A. McGowan, P.H. Wiebe, 'Patterns and Processes in the Timescale of Plankton', in *Spatial Patterns in Plankton Communities*, ed. J. Steele, Plenum Press, New York, 277–327 1978.
59. J.D. Woods, 'Scale Upwelling and Primary Production', in *Towards a Theory on Biophysical Interactions in the World Ocean*, ed. B.J. Rothschild, Kluwer Academic, Dordrecht, 7–83, 1988.
60. V.H. Strass, 'Chlorophyll Patchiness Caused by Mesoscale Upwelling at Fronts', *Deep-Sea Res.*, **36**, 75–96 (1992).
61. C.S. Davis, G.R. Flierl, P.H. Wiebe, P.J.S. Franks, 'Micropatchiness, Turbulence and Recruitment in Plankton', *J. Mar Res.*, **49**, 109–151 (1991).
62. K.L. Denman, A.E. Gargett, 'Biological–Physical Interactions in the Upper Ocean: The Role of Vertical and Small Scale Transport Processes', *Annu. Rev. Fluid Mech.*, **27**, 225–255 (1995).
63. H.U. Sverdrup, 'On Conditions for the Vernal Blooming of Phytoplankton', *J. Cons. Per. Int. Exp. Mer.*, **18**, 287–295 (1953).
64. S.G.H. Philander, *El Nino, La Nina, and the Southern Oscillation*, Academic Press, San Diego, 1990.
65. D. Halpern, G. Feldman, 'Annual and Interannual Variations of Phytoplankton Pigment Concentration and Upwelling Along the Pacific Equator', *J. Geophys. Res.*, **99**, 7347–7354 (1994).
66. A. Oschlies, V. Garcon, 'Eddy-induced Enhancement of Primary Production in a Model of the North Atlantic Ocean', *Nature*, **394**, 266–269 (1998).
67. G. Garcia-Moliner, J.A. Yoder, 'Variability in Pigment Concentration in Warm-core Rings as Determined by Coastal Zone Color Scanner Satellite Imagery from the Mid-Atlantic Bight', *J. Geophys. Res.*, **99**, 14277–14290 (1994).
68. D.B. Olson, 'Rings in the Ocean', *Ann. Rev. Earth Planet. Sci.*, **19**, 283–311 (1991).
69. S.J. Weeks, F.A. Shillington, 'Interannual Scales of Variation of Pigment Concentrations from Coastal Zone Color Scanner Data in the Benguela Upwelling System and the Subtropical Convergence Zone South of Africa', *J. Geophys. Res.*, **99**, 7385–7399 (1994).
70. K.L. Denman, M.R. Abbott, 'Time Scales of Pattern Evolution from Cross-spectrum Analysis of Advanced Very High Resolution Radiometer and Coastal Zone Color Scanner Imagery', *J. Geophys. Res.*, **99**, 7433–7442 (1994).
71. F.M. Sousa, A. Bricaud, 'Satellite-derived Phytoplankton Pigment Structures in the Portuguese Upwelling Area', *J. Geophys. Res.*, **97**, 11343–11356 (1992).
72. M.R. Abbott, P.M. Zion, 'Satellite Observations of Phytoplankton Variability during an Upwelling Event', *Cont. Shelf Res.*, **6**, 661–680 (1985).
73. R.E. Thompson, J.F.R. Gower, 'A Wind-induced Mesoscale Eddy over the Vancouver Island Continental Slope', *J. Geophys. Res.*, **90**, 8981–8993 (1985).
74. C.A.E. Garcia, I.S. Robinson, 'Sea Surface Velocities in Shallow Seas Extracted from Sequential CZCS Satellite Data', *J. Geophys. Res.*, **94**, 12681–12691 (1989).
75. M.R. Lewis, J.J. Cullen, T. Platt, 'Phytoplankton and Thermal Structure in the Upper Ocean: Consequences of Nonuniformity in the Chlorophyll Profile', *J. Geophys. Res.*, **88**, 2565–2570 (1983).
76. J.T.O. Kirk, 'Solar Heating of Water Bodies as Influenced by their Inherent Optical Properties', *J. Geophys. Res.*, **93**, 10897–10908 (1988).
77. M.R. Lewis, M.E. Carr, G.C. Feldman, W. Esaias, C. McClain, 'Influence of Penetrating Solar Radiation on the Heat Budget of the Equatorial Pacific', *Nature*, **347**, 543–545 (1990).
78. M.J.R. Fasham, 'Modelling the Marine Biota', in *The Global Carbon Cycle*, ed. M. Heimann, Springer-Verlag, New York, 457–504, 1993.
79. M.R. Abbott, 'Report of the US JGOFS Workshop on Modeling and Data Assimilation', US JGOFS Planning Report 14, 1992.

80. J.L. Sarmiento, R.A. Armstrong, 'US JGOFS Synthesis and Modeling Project Implementation Plan', US JGOFS Report, 1997.
81. E. Hofmann, T. Osborne, T. Powell, J. Price, B. Rothschild, J. Roughgarden, 'Theory and Modeling in GLOBEC: A First Step', US GLOBEC Report, 1991.
82. A.E. Gill, *Atmosphere–Ocean Dynamics*, Academic Press, San Diego, 1982.
83. G.T. Evans, J.S. Parslow, 'A Model of Annual Plankton Cycles', *Biol. Oceanogr.*, **3**, 327–347 (1985).
84. M.J.R. Fasham, H.W. Ducklow, S.M. McKelvie, 'A Nitrogen Based Model of Plankton Dynamics in the Oceanic Mixed Layer', *J. Mar. Res.*, **48**, 591–639 (1990).
85. W.C. Thacker, R.B. Long, 'Fitting Dynamics to Data', *J. Geophys. Res.*, **93**, 1227–1240 (1988).
86. M. Ghil, P. Malanotte-Rizzoli, 'Data Assimilation in Meteorology and Oceanography', *Adv. Geophys.*, **33**, 141–266 (1991).
87. P. Malanotte-Rizzoli (ed.), *Modern Approaches to Data Assimilation in Ocean Modeling*, Elsevier, New York, 1996.
88. A.F. Michaels, A.H. Knap, R.L. Dow, K. Gundersen, R.J. Johnson, J. Sorensen, A. Close, 'Seasonal Patterns of Ocean Biogeochemistry at the US JGOFS Bermuda Atlantic Time-series Study Site', *Deep-sea Res. I*, **41**, 1013–1038 (1994).
89. H.W. Ducklow, R.P. Harris, 'Introduction To the JGOFS North Atlantic Bloom Experiment', *Deep-sea Res.*, **40**, 1–8 (1993).
90. J. Ishizaka, 'Coupling of CZCS Data to a Physical–Biological Model of the Southeastern US Continental Shelf Ecosystem, 3. Nutrient and Phytoplankton Fluxes and CZCS Data Assimilation', *J. Geophys. Res.*, **95**, 20201–20212 (1990).
91. J. Ishizaka, 'Coupling of CZCS Data to a Physical–Biological Model of the Southeastern US Continental Shelf Ecosystem, 2. An Eulerian Model', *J. Geophys. Res.*, **95**, 20183–20199 (1990).
92. J. Ishizaka, 'Coupling of CZCS Data to a Physical–Biological Model of the Southeastern US Continental Shelf Ecosystem, 1. CZCS Data Description and Lagrangian Particle Tracer Experiments', *J. Geophys. Res.*, **95**, 20176–20181 (1990).
93. J. Ishizaka, 'Data Assimilation for Biogeochemical Models', in *Towards a Model of Ocean Biogeochemical Processes*, eds. G.T. Evans, M.J.R. Fasham, Springer-Verlag, New York, 295–316, 1993.
94. R.A. Armstrong, J.L. Sarmiento, R.D. Slater, 'Monitoring Ocean Productivity by Assimilating Satellite Chlorophyll into Ecosystem Models', in *Ecological Time Series*, eds. T.M. Powell, J.H. Steele, Chapman and Hall, New York, 379–390, 1995.
95. P. Prunet, J.-F. Minster, D. Ruiz-Pino, I. Dadou, 'Assimilation of Surface Data in a One dimensional Physical–Biogeochemical Model of the Surface Ocean 1. Method and Preliminary Results', *Global Biogeochem. Cyc.*, **10**, 111–138 (1996).
96. P. Prunet, J.-F. Minster, V. Echevin, I. Dadou, 'Assimilation of Surface Data in a One dimensional Physical–Biogeochemical Model of the Surface Ocean 2. Adjusting a Simple Trophic Model to Chlorophyll, Temperature, Nitrate, and pCO_2 Data', *Global Biogeochem. Cyc.*, **10**, 139–158 (1996).
97. T.M. Lenton, 'Gaia and Natural Selection', *Nature*, **394**, 439–447 (1998).
98. R.J. Charlson, J.E. Lovelock, M.O. Andreae, S.G. Warren, 'Oceanic Phytoplankton, Atmospheric Sulfur, Cloud Albedo and Climate', *Nature*, **326**, 655–661 (1987).
99. J. Lovelock, *Gaia: A New Look at Life on Earth*, Oxford University Press, Oxford, 1995.
100. A.M. Thompson, W.E. Esaias, R.L. Iverson, 'Two Approaches to Determining the Sea-to-air Flux of Dimethyl sulfide: Satellite Ocean Color and a Photochemical Model with Atmospheric Measurements', *J. Geophys. Res.*, **95**, 20551–20558 (1990).
101. M.O. Andreae, P.J. Crutzen, 'Atmospheric Aerosols: Biogeochemical Sources and Role in Atmospheric Chemistry', *Science*, **276**, 1052–1058 (1997).
102. P.C. Fielder, G.B. Smith, R.M. Laurs, 'Fisheries Applications of Satellite Data in the Eastern North Pacific', *Mar Fish. Rev.*, **46**, 1–13 (1984).
103. R.M. Laurs, P.C. Fielder, D.R. Montgomery, 'Albacore Tuna Catch Distributions Relative to Environmental Features Observed from Satellites', *Deep-sea Res.*, **31**, 1085–1099 (1984).
104. A. Subramaniam, E.J. Carpenter, 'An Empirically Derived Protocol for the Detection of Blooms of the Marine Cyanobacterium *Trichodesmium* using CZCS Imagery', *Int. J. Remote Sensing*, **15**, 1559–1569 (1994).
105. R.A. Neville, J.F.R. Gower, 'Passive Remote Sensing of Phytoplankton via Chlorophyll α-Fluorescence', *J. Geophys. Res.*, **82**, 3487–3493 (1977).
106. R.G. Williams, M.J. Follows, 'The Ekman Transfer of Nutrients and Maintenance of New Production over the North Atlantic', *Deep-sea Res. I*, **45**, 461–489 (1998).
107. K. Richardson, 'Harmful or Exceptional Phytoplankton Blooms in the Marine Ecosystem', *Adv. Mar. Biol.*, **31**, 302–385 (1997).
108. J.J. Cullen, A.M. Ciotti, R.F. Davis, M.R. Lewis 'Optical Detection and Assessment of Algal Blooms', *Limnol. Oceanogr.*, **42**, 1223–1239 (1997).

Elevation Modeling and Displacement Mapping using Radar Interferometry

Didier Massonnet
Centre National d'Etudes Spatiales, Toulouse, France

1 Introduction 1
2 History 2
3 Radar Imagery and the Principles of Interferometry 2
3.1 Physical and Geometrical Properties of Radar Images 2
3.2 Interference between Radar Images 3
3.3 Limits Caused by Changes in the Geometry and the Surface 4
3.4 Elimination of Systematic Contributions 4
4 Application to Topographic Calculations 4
4.1 Set of Equations and Accuracy 4
4.2 Suppression of the Measurement Ambiguity 5
4.3 Dedicated Systems 5
5 Application to the Measurement of Displacements 5
5.1 Set of Equations and Accuracy 5
5.2 Examples of Use 6
6 Ambiguity of Interpretation 7
6.1 Signature from Various Contributions 7
6.2 Methods for Improving Interpretation 7
7 Availability of Data and Systems 7
7.1 Availability of Archived Radar Data 7
7.2 Availability of Processing Tools 7
7.3 Principle of Data Selection 8
8 Comparison with Other Methods and Prospects 8
8.1 Comparison with Stereoscopy for Topographic Measurement 8
8.2 Comparison with Global Positioning System for Measurement of Shifts 9
8.3 Prospects for Dedicated Space Missions 9
Abbreviations and Acronyms 10
Related Articles 10
References 10

Topographic modeling consists of producing a digital description of the relief of a given area. This description is generally given on a regular grid in two coordinates. The product consists of the value of the altitude of the terrain for each position of the grid, with respect to a reference shape such as an ellipsoid (approximation of the shape of the Earth) or the geoid (shape of the potential of gravity at sea level). The grid itself is defined in a cartographic or a geographic coordinate system. Common systems include the Universal Transverse Mercator (UTM), Lambert or, alternatively, coordinates in latitude/longitude. Topographic data are used in a variety of applications, from the correction of geometric distortion in remote sensing images to the prediction of the volume of materials to be removed in public works. Their applications include military uses, such as collision avoidance for low-flying aircraft or missiles, or civilian ones, such as optimal positioning of transmitters to cover an area for mobile communications. Radar interferometry is a promising technique for producing topographic maps on a large scale. It makes use of the very peculiar features of the images produced by synthetic aperture radar (SAR) instruments placed on board satellites. Used in differential mode, this technique can also produce a unique product: the map of the displacements which occur in a given area between two passes by a radar satellite. Again, the data can be organized as a digital file placed on a grid similar to the ones used for topographic mapping. However, in the case of displacement measurement, the useful piece of data placed on each node is the displacement experienced by the node, expressed in appropriate units of length. As for each new technique, there are a number of points to take care of before interpreting the data. The purpose of this article is to describe the radar data, both historically and technically, and to detail their two main applications in topography and displacement mapping. The main issues of the interpretation and the availability of the data are covered. Finally, we compare the performance of radar interferometry with other techniques, in both domains, and we try to outline its future.

1 INTRODUCTION

Radar imagery is often thought of as a substitute for the optical image, when the latter is not available because of cloud cover. This way of seeing things is doubly false

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

because on the one hand, radar imagery has unique applications, which cannot be matched by optical imagery, and on the other hand because radar imagery is not completely insensitive to cloud cover, which is one of its main limits in the applications that we will discuss in this article.

After a brief history of radar imagery which will be followed by a description of the main technical characteristics of radar images, we will discuss one of the most spectacular techniques made possible with this type of data: interferometry.

This technique mainly produces two kinds of information, usually mixed together on the same image: topographic information and information on ground shifts. There are also undesirable contributions due to changes in the atmosphere during imaging. In order to avoid interpretation errors, one must know how to detect these contributions.

We will discuss the principles of the interferometric comparison of radar images, the limits of this technique and the systematic contributions that must be removed. Then we will go into the details of topographic computation using interferometry, which can produce, in the best case use of satellite data, altimetric accuracy of 1 or 2 m on cells with 30 m sides. We will then discuss the details of the computation of shifts on the Earth's surface, for which accuracy of the order of 2 mm has already been observed.

We will also address the problem of ambiguous interpretation and how to deal with it, the availability of space- or airborne systems, archives of radar data and software processing tools. Finally we will compare interferometry with competing techniques, by pointing out its advantages and disadvantages, and finish with the prospects for space missions dedicated to this technique.

2 HISTORY

Radar imagery is almost as old as radar itself. Radar instruments were flown on aircraft at the beginning of the 1940s, with detection goals similar to those of radar devices on the ground. The first processing intended to improve radar capabilities in terms of ground imagery date from the beginning of the 1950s, with ''beam sharpening'' techniques. The ultimate evolution of this processing, which allowed radar images to compete with optical images in terms of resolution, was formalized in the beginning of the 1960s. Wiley played a crucial role in these changes.[1] Nevertheless, the computation requirements for the proposed processing led their designers to use analog methods based on optical computing with laser. The use of the phase of the radar signal by comparing two images taken simultaneously dates from the beginning of the 1970s.[2] This was extended to two images taken at different moments from space at the beginning of the 1980s.[3] In the mid-1980s, a suggestion was made that interferometry could be used to measure ground shifts. Because so little new space data was available, it took several years before the technique, with various capacities, was demonstrated.[4,5]

The first steps of radar processing have only historic value. Current data can be processed by everyday workstations without compromising quality, in a very reasonable time that can only decrease with the regular increase in computing power. The reproducibility of digital processing makes it easy to use computed interference.

3 RADAR IMAGERY AND THE PRINCIPLES OF INTERFEROMETRY

3.1 Physical and Geometrical Properties of Radar Images

The main difficulty with radar imagery is the large wavelength used. Some radio waves have the advantage of passing through clouds, but their wavelength is larger than optical ones by a factor of one hundred thousand. To obtain the same resolution as with an optical system, a radar system must therefore have an instrument aperture one hundred thousand times larger. To imitate an optical observation satellite such as Système Probataire Pour l'Observation de la Terre (SPOT), whose telescope has a diameter of more than 30 cm, a radar antenna of thirty kilometers would therefore be needed. It is then obvious that radar imagery cannot rely on the separating power of the instrument. Two ideas are used to replace it: firstly, range imagery, which allows different points on the ground to be separated by the time that the radar wave takes to complete the round-trip distance. This wave is sampled at tens of megahertz, which corresponds to separations of the order of tens of meters on the ground. Each sample includes the wave amplitude and phase. Secondly, there is Doppler imagery, which separates the points along the direction of the velocity of the radar carrier by detailed analysis of the phases of the successive radar signals. In order for these techniques to work together to obtain a well-resolved image in the two main directions, the instrument must look to the side (Figure 1). These principles and their consequences are discussed in detail in several papers.[6] They allow the data provided by the instrument (''raw'' data) to be converted into image data (''synthesized'' data). We will simply mention the main characteristics here. The

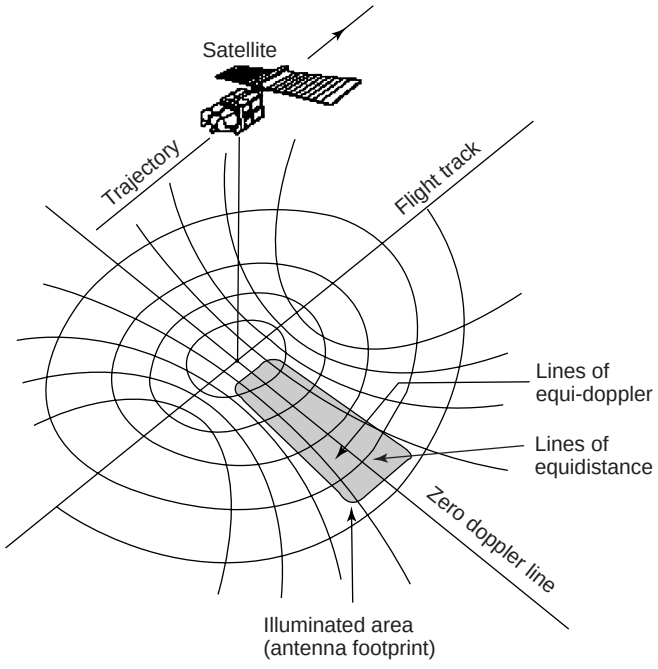

Figure 1 Viewing geometry of radar. The instrument looks to the side to take advantage of the possibility of sorting echo samples by distance (closed lines). Synthetic aperture processing permits points to be distinguished by their Doppler effect (open lines). Both principles work together to cut the antenna print into high resolution pixels.

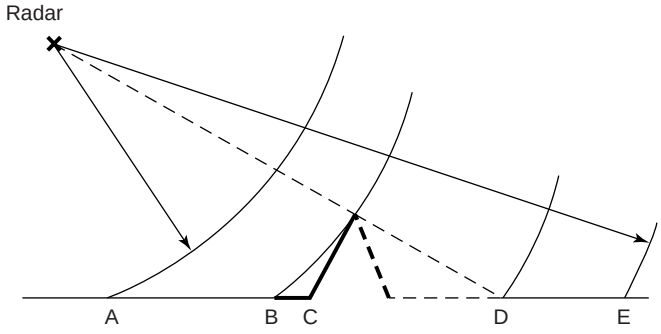

Figure 2 Difficulties caused by the principle of distance imagery. Distance ordering is compatible with cartographic coordinates between points A and B. Between points B and C the slope of the mountain is mixed with the contributions of points in the valley which share the same distance, a phenomenon called layover. In particular, the top of the mountain is seen, in the distance, before C, even if C is found before the top on a map. A more conventional effect is found between the distances of C and D, where shadowing prevents any signal from returning to the radar. Unlike layover, shadowing may occur in optical images.

principle of range imaging creates specific artifacts, which have no equivalent in optics. Depending on the incidence angle of the radar wave, two points (for example the peak of a mountain and the bottom of a valley) can be at the same range from the radar and their contributions will be irremediably mixed together in the same point of the radar image. This is the ''overlay'' phenomenon (Figure 2).

The principle of Doppler imagery, also called synthetic aperture, requires large quantities of calculations, a feature which is fortunately attenuated by current computer systems. It leads to a geometry of the image tied to the position of the radar carrier and to its velocity vector. The geometry of the image thus does not depend on the orientation of the radar, which makes locating it much more precise than for the equivalent optical imagery. On the other hand, the radar image has an unusual appearance because it functions in amplitude and phase. In short, a radar image is formed from complex numbers. An element of the final image (pixel) is thus the sum of complex numbers which are representative of the different elementary targets to be found in the pixel. The amplitude of two identical targets present in the same pixel may thus be reinforced if their phases are identical, or cancel each other out if the phases are opposite. In the general case, the amplitude of the ''synthesized'' radar image is characterized by scintillation noise (''speckle''), which makes reading it more difficult than for an optical image.

The phase information at the pixel level is the ultimate information on range, as the fractional part of the radar range when the latter is expressed in units of the wavelength. The phase of each pixel in the image, however, results from the interaction of multiple elementary targets located at various ranges within the pixel, and varies randomly between neighboring pixels in a same image. But if the configuration of the elementary targets does not change, the ''random'' contribution in the phase between two radar images can be eliminated by subtracting their phases point by point. The difference of phase will just reveal the change of range for the pixel globally. This is the principle of radar interferometry.

3.2 Interference between Radar Images

The phase of a radar image used in comparison with that of a second radar image, records all the differences in radar range between the two images. The two images that we merge to form the image of the point-by-point difference of their phases, called the interferogram, have, in general, different perspectives, since they were not acquired from exactly the same spot, and a time difference, since they were not acquired at the same moment. These two differences almost always occur together in an interferogram. The difference in time is eliminated in systems with two radar antennas that can create two images simultaneously. The difference in

perspective can disappear exceptionally, when a satellite passes over almost exactly the same point to acquire the second image.

These two differences are behind the two types of information provided by interferometry. The difference in perspective creates the topographic information in the interferogram. The difference in time creates information on displacement. Each of these differences is also responsible for difficulties when implementing the technique.

3.3 Limits Caused by Changes in the Geometry and the Surface

The difference in perspective between two images should be limited. Let us assume we have two distinct targets, located at two ends of a pixel (one close to the radar, the other far from it). The targets will always be mixed together on the same pixel. It can thus be understood that if the variations in phase undergone by the two targets during the change from one image to the other are approximately identical, the variation in phase will be significant at the scale of the pixel. In the opposite case, the phase difference between the two images will, again, result in random contributions, because they are variable within a pixel. The difference in phase cannot then be exploited.

In the limiting case where the targets are located, in range, at the two extremities of a pixel, the stability of the phase variation will be guaranteed by the stability of the incidence angle over the two images. If it changes, the pathlength between the two targets on one image will be different from the pathlength on the other image. For example, a pixel of 10 m observed with 30° incidence assumes a round-trip pathlength of 10 m between two extreme targets. In order to limit the variation of this pathlength to a fraction of the wavelength, for example 1 cm, the incidence angle in the second image must be between 29.967° and 30.033°! A more concrete way of showing this is to express the maximum acceptable difference in position between the points from which the images are acquired. For a satellite in orbit at an altitude of about 1000 km, this difference is of the order of a kilometer (Figure 3). One can therefore only combine images separated by a whole number of orbital cycles of the satellite, when it is supposed to be in exactly the same place, and actually is found within 1 km in general (section 7.1).

3.4 Elimination of Systematic Contributions

Even if the previous condition is fulfilled, the slight difference in the incidence angle between the two images will create a progressive shift from one image to the next. For example, even if this difference only creates a tenth of a phase cycle on a pixel, it will create a full cycle every ten pixels, and several hundreds of cycles over the width of an image. Interference fringes of this type, called ''orbital fringes'', can be eliminated by computations based on the trajectories. Because of uncertainties in the orbit, some fringes may remain after this correction. These fringes are used to refine the trajectories in the same way that the alignment of mirrors is adjusted in an optical interferometer.

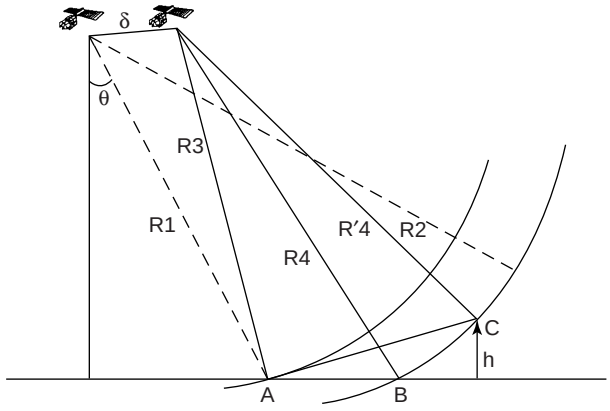

Figure 3 Geometric limitations of radar interferometry. The distance between R_1 and R_2 represents the size of a pixel in the distance (point A to B on the ground) in the first image. To create a proper interferometric effect, the corresponding difference of distance R_4-R_3 in the second image should not differ from the size of a pixel by more than half the wavelength (or a full wavelength in round-trip distance), which typically restricts δ to less than one kilometer in the usual conditions of a spaceborne radar. If we consider a change of altitude h in the pixel, now representing AC on the ground, the distance to the second image will not be the expected R_4, but R'_4. The difference is observed as an unexpected phase change in the interferogram, corresponding to the fractional part of R'_4-R_4 in units of half the wavelength. h is the altitude of ambiguity if R'_4-R_4 equals half the wavelength. In round-trip distance, it represents one full phase cycle on the interferogram. To remain within the limits of interferometry, a change in altitude larger than or equal to the altitude of ambiguity cannot occur across a single pixel.

4 APPLICATION TO TOPOGRAPHIC CALCULATIONS

4.1 Set of Equations and Accuracy

The elimination of the orbital fringes leaves some ''relief fringes''. By looking again at Figure 3, we can see that if the terrain changes elevation in a pixel, that is between the ranges R_1 and R_2, which do not change, the distance R_4 will be changed to R'_4. This alteration will be expressed by an additional pathlength revealing the relief. The pathlength will be observed as a phase difference between two adjacent pixels.

To avoid manipulating complicated geometric equations, it is practical to combine all the parameters involved in the altitude measurement, such as the orbital separation, the wavelength and the incidence angle, in one single significant value, the altitude of ambiguity. This is defined by the variation in altitude that produces a relief fringe. It does not generally vary much on one interferogram and is thus a simple and practical way to characterize one. Assuming a horizontal difference between image acquisition points of δ, an incidence angle of θ, an observation range of R and a wavelength λ, the altitude of ambiguity h is given approximately by Equation (1):

$$h = \frac{R\lambda \tan(\theta)}{2\delta} \quad (1)$$

Figure 4 shows an example of a topographic interferogram.

4.2 Suppression of the Measurement Ambiguity

In mountainous terrain, and after elimination of orbital fringes, an interferogram looks like a family of interference fringes analogous to contour lines, separated in elevation by the local value of the altitude of ambiguity (Figure 4). The main difference between this interferogram and a topographic map is that the ''contour lines'' present as fringes are not numbered. In fact, the radar signal does not allow an order to be attributed to the fringes since each one corresponds to geometric differences between images that are usually less than one hundredth the size of the pixel, which is beyond the possibilities of correlation measurements. The only method for renumbering the fringes, an operation commonly called ''phase unwrapping'', is to analyze the continuity of the interferogram. This is what we do when we analyze a topographic map. The noisy areas on the interferogram, as well as the places where the topographic fringes are very close together, pose significant problems to algorithms for automatic unwrapping of fringes, for which many solutions have been proposed.[7,8]

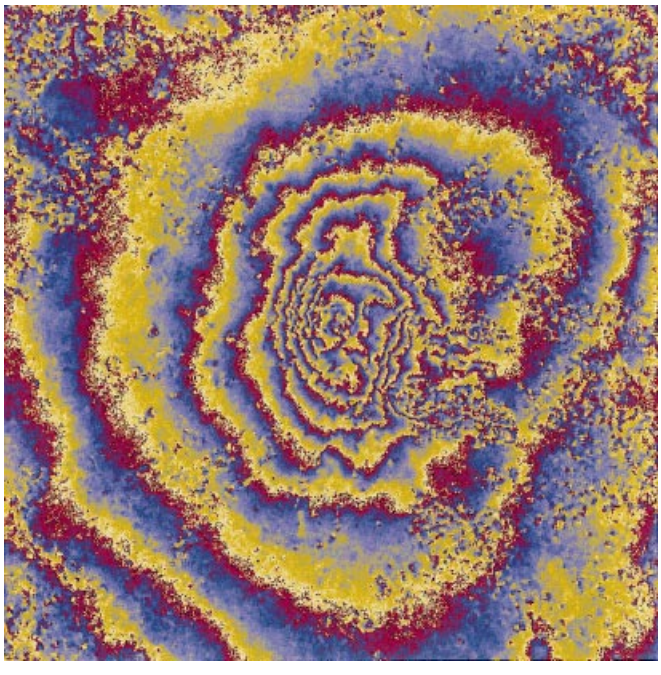

Figure 4 Example of ''topographic'' interferogram on Mount Etna, Sicily. Each fringe corresponds to a change in elevation of 250 m.

4.3 Dedicated Systems

Interferometry provoked a strong interest in the computation of topography. Given the sensitivity of interferograms to other factors, such as the variation in atmospheric thickness discussed in section 6.1, systems with simultaneous imaging, for which these contributions can be eliminated between two images, were preferred.

Even though space radar systems with two antennas can be imagined (section 8.3), this concept has mostly been used for airborne systems, which offer economic means of computing the topography with interesting operational characteristics. For aircraft, which are much closer to their target than a satellite, the distance between the antennas is at most several meters, and not a kilometer (section 3.3). It is therefore possible to install the two antennas on the same plane, with either a horizontal or a vertical separation. The distance between the two antennas is then well known. The main problem is to properly take into account the roll that affects the aircraft. This can make the observational base ''wobble''.

5 APPLICATION TO THE MEASUREMENT OF DISPLACEMENTS

5.1 Set of Equations and Accuracy

Assuming that orbital and topographic effects have been compensated for, an interferogram can reveal the variations in range linked to the time elapsed between images. Even though these variations may include atmospheric effects (section 6.1), they are mainly useful for measuring groundshifts. If one part of the landscape observed by an image has moved, this displacement will be seen as fringes on the interferogram. More precisely, if the displacement at point A is represented

by a vector D and the position of the satellite is S, in any common reference frame, this displacement will be expressed (Equation 2) by a number of fringes n equal to:

$$n = \frac{2D \cdot (S - A)}{\lambda} \qquad (2)$$

where the factor 2 takes into account the round-trip path and where $\cdot$ represents the dot product. The accuracy of this kind of measurement therefore depends on the direction of the shift and the way in which a fraction of a phase cycle can remain significant. Shifts directed towards the satellite are observed the best. On the other hand, for a shift along the ground track of the satellite, where D is perpendicular to $(S - A)$, the interferogram is not modified. The vector $(S - A)$, after being normalized and expressed in a geographical reference frame (for example: North, East and Up), gives a precise view of the sensitivity of the measurement. For example, a typical value for a satellite such as the European radar satellite (ERS) launched in 1991, ERS-1, is, for mid-latitudes, (−0.08, 0.35, 0.93). The significant fraction of the phase cycle depends on the signal-to-noise ratio on the interferogram, which is a result of the power of the radar signal but even more of the way in which the layout of the elementary targets has been maintained in the time elapsed between two images. From this point of view, a desert, rocky or urban landscape would be very favorable, an aquatic landscape (ocean, river or swamp) would not yield any result and a landscape covered with vegetation would give an average result depending on the elapsed time and the way in which the radar waves pass through the vegetation to reach ''solid'' targets on the ground. Long wavelengths manage this more easily. The signal-to-noise ratio of an interferogram is therefore essentially experimental data. Accuracy of the order of 2 mm have nevertheless been confirmed in favorable cases. Accuracy also depends on the way the other components of interferometric information have been removed. For example, if the interferogram used for measuring shifts has an altitude of ambiguity of 100 m, and the terrain model used to remove the topographic contribution may have errors of 20 m, one-fifth of a fringe of topographic origin may remain. If the wavelength is 5 cm, a ''noise'' corresponding to 0.5 cm (1 cm for the round-trip path) will remain in the interferogram.

5.2 Examples of Use

Measuring shifts with interferometry is more interesting than topographic measurement, because there is almost no competition (section 8). The limitations of the technique are threefold:

1. the ambiguous character of the measurement, which must be unwrapped;
2. geometric sensitivity which varies depending on the direction of the displacement;
3. applicability which depends on the type of terrain observed.

They have not prevented spectacular developments in the measurement of deformations caused by earthquakes,[5] glaciers,[9,10] volcanoes[11] and even phenomena with smaller amplitudes like post-seismic motion[12,13] or continental drift.[14] In the field of hydrology (Figure 5), interferometry has proved to be capable of measuring the quantity of water pumped by a geothermal plant.[15] Irrigation also has a direct effect on the phase,[4] which may come from geometric causes (swelling) or physical ones (change in electrical conductivity).

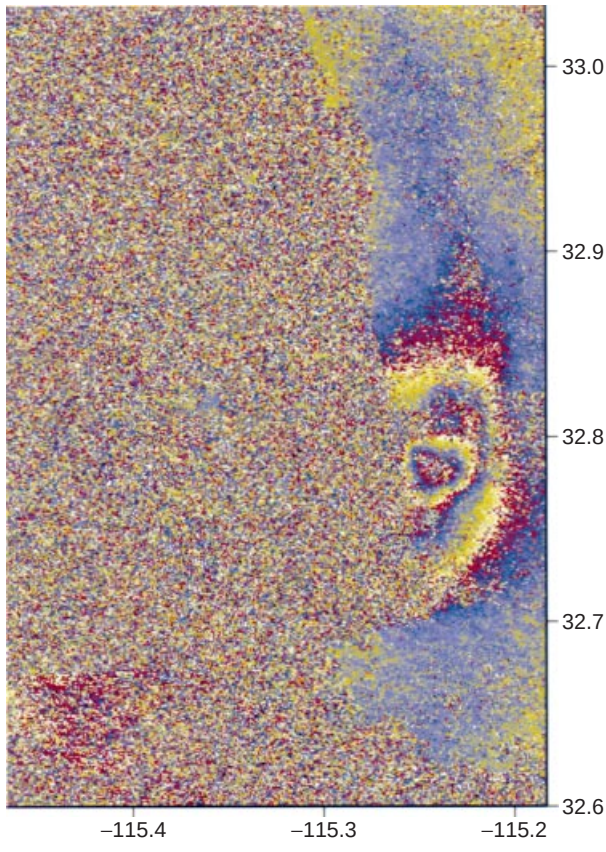

Figure 5 Example of a differential interferogram. In California, the East Mesa Geothermal plant created a subsidence bowl by pumping water for two years. Each fringe represents about 30 mm of deepening. The ellipse-shaped bowl is 18 km by 10 km. The volume of the bowl stands at 4 million cubic meters where the loss of water recorded by the plant over the same period stands at 5 million cubic meters. Blurred areas to the west are used for agriculture, so that the soil did not keep the stability required for interferometry over the 2 years which elapsed between image acquisitions.

6 AMBIGUITY OF INTERPRETATION

6.1 Signature from Various Contributions

A system for classifying phase structures[16] should be used to enable correct interpretation, if several interferograms are available. The following in particular can be distinguished:

- Structures tied to the orbital difference: this involves phase structures viewed in several interferometric combinations of images, with an amplitude proportional to the topographic sensitivity of the interferometric pairs, which itself is tied to the distance between their orbits. This behavior characterizes the non-compensated or poorly corrected topographic residual.
- Atmospheric artifacts: this term includes structures that appear, with the same amplitude, in all the interferometric combinations of a given image. This behavior characterizes heterogeneity in the conditions of atmospheric propagation at the exact moment of image acquisition. Because the responsible image can be identified, the sign of the artefact is known. Artifacts that lengthen the trajectory are attributed to pressure waves or to turbulence in the troposphere and those that shorten it to local neutralizations in the ionosphere.
- Structures that depend on a date: this involves phase structures viewed with the same amplitude in several interferometric combinations of images, whenever the acquisition dates of the images that form them include a particular date. This behavior characterizes geophysical events like earthquakes.

This list can be expanded, for example with phase structures viewed in interferograms with an amplitude proportional to the time elapsed between the two images that form them. This behavior characterizes geophysical deformations that are continuous in time.

6.2 Methods for Improving Interpretation

Various simple methods can improve the appearance or content of interferograms. Numerous filters have been proposed. The simplest one recombines the phases of an interferogram with the corresponding amplitude from the source radar images, then averages the complex numbers formed this way in a neighborhood. The phase of the averaged complex number formed this way will be much less noisy than the initial phases, with some deterioration in the spatial resolution of the image. This method assumes that the interferogram has already had its densest fringes removed, in particular the orbital fringes, otherwise an average might blur the results. One can also use the systematic characterization of the contents of the interferograms described in section 6.1 to attenuate or amplify this or that contribution. For example, if an interferogram has, because of the orbital gap of the images that form it, a topographic sensitivity of one fringe for 100 m of difference in elevation, while another interferogram has only fifty meters, the combination of twice the first interferogram minus the second will not contain any topographic effects. If the combination of different interferograms can be done with whole multiplying coefficients, the combination is possible before any fringe unwrapping.

7 AVAILABILITY OF DATA AND SYSTEMS

7.1 Availability of Archived Radar Data

Space radar images that can be combined must come from the same satellite, or from two identical satellites that are in almost the same position (typically within a kilometer, section 3.3) at different times. Satellites repeat their position at the end of their orbital cycle. Among the usable satellites is ERS-1, which went through two distinct 3-day orbital cycles, now completed, and a 35-day cycle which it is still in and which has taken up the largest part of its mission. The ERS-2, satellite, launched in 1995, has a radar identical to that of ERS-1, and is therefore compatible. It is on the same 35-day cycle, but with a one day lag (''tandem'' concept). By combining the ERS-1 and ERS-2 data, one can thus obtain gaps of 35, 34 or 1 day. The data are in C-band (wavelength 5.6 cm). Still in C-band, but incompatible with ERS because of orbital and instrumental differences, are the data from RADARSAT (Radar Satellite of the Canadian Space Agency), a Canadian satellite launched in 1995, with a 24-day cycle and which features the possibility of imaging with various incidence angles. In L-band (wavelength 23 cm), there are data from the Japanese Earth Resource Satellite (J-ERS), which is on an orbit with a 44-day repeat cycle. Intrinsically less accurate than C-band, L-band can obtain better results on terrain covered with vegetation, since the longer wavelength can penetrate the vegetation and reach the ground. Short experiments have also taken place with radar on board the space shuttle (Shuttle imaging radar (Sir), Sir-A, Sir-B and Sir-C missions).

7.2 Availability of Processing Tools

The usefulness of radar images for computing topography and groundshifts involves a much larger public than the community of specialists working with this type of image. It is therefore necessary that the various processing steps be conducted by non-specialist users.

The processing steps include

1. the synthesis of radar images from raw radar data (section 3.1). Images that are already synthesized are available on the market;
2. the point-by-point registration of the two images to be combined by resampling;
3. the computation of the phase difference;
4. finishing work which includes fine-orbit adjustment and aligning the product in the desired cartographic geometry.

This processing is accessible to non-specialists who use scientific or off-the-shelf software. Commercial software uses standard workstations. When processing radar data one should begin with raw data. The inconvenience of this is the large quantity of computations to be performed, but though these computations are long, they are controlled by simple parameters and do not present any problem. On the other hand, the raw data are easier to concatenate than the processed data (when one site extends over more than one image in length). Optimal implementation of interferometry requires that the processing parameters for scenes to be combined be identical rather than optimized for each scene, which would be the case for scenes processed in a conventional fashion. The use of raw data with interferometric goals can impose a common value on these processing parameters, a compromise between recommended values for each scene. Finally, ordering scenes in raw form may be less expensive.

7.3 Principle of Data Selection

The choice of radar images to be combined with interferometry depends on the aim of the measurement and the nature of the terrain to be observed. For topographic computation, short time differences are preferable (the ideal being simultaneous images) as are altitudes of ambiguity compatible with the type of relief observed. For example, if altitude differences of 1 km are in the scene, it is preferable to choose an altitude of ambiguity of 100 m rather than an altitude of ambiguity of 10 m, although it may mean restarting with a lower altitude of ambiguity after a first pair has roughly calculated the topography. Finally, it is always necessary to check the elevation map with several sets of data, in order to tag atmospheric artifacts. For measurements of drifts that are constant in time, one should wait until the drift reaches a measurable amplitude. For instantaneous motion (earthquake), the images must simply be taken before and after the event. For measuring motions, data with a high altitude of ambiguity are favored, to make it easier to eliminate topographic contributions. In all cases, climatic considerations are involved. In Iceland, one can only work with images acquired in summer because of the snow.[14] Paradoxically, it is sometimes better to work with images taken at an interval of one year rather than six months, because of seasonal cycles. One can take advantage of the capability of radar to work at night. Since it is an active instrument, it is its own source of light and at night the atmospheric conditions and states of vegetation are more stable.[11] When there is dense vegetation, it is better to use L-band data. The choice of optimal data is thus largely a question of opportunity and depends on the know-how of the specialist of the object of the measurement and on his knowledge of the sites.

8 COMPARISON WITH OTHER METHODS AND PROSPECTS

8.1 Comparison with Stereoscopy for Topographic Measurement

In optical stereoscopy, the difference in perspective between the images creates a deformation that depends on the relief, and this deformation is used to compute the relief. The critical parameters of the method are the pixel size and the ''base-to-height'' ratio. The smaller the pixel the better the measurement of the deformation. Depending on the measurement methods used, it is possible to detect deformations with an accuracy equal to a more or less large fraction of the pixel size. The ''base-to-height'' ratio affects the extent of the deformation, for a given relief. In optical stereoscopy, one can obtain a ratio of one. That is to say, that the distance between the points from which the images are taken is equal to the altitude of the imaging instrument. In these conditions, an elevation difference in the observed landscape is expressed by a deformation of the same size between the images. The accuracy in altitude is then equal to a fraction of the pixel size, or typically a few meters.

In radar interferometry, the base-to-height ratio is obviously very low, as we saw in section 3.3, and can hardly exceed one thousandth. Even though the equations of geometric sensitivity to the difference in altitude are different from those in optics, one can conclude that over a given relief the deformation between two radar images which can be combined by interferometry is typically one thousand times less than that of two stereoscopic images. On the other hand, the measurement is based on the phase difference, which allows, depending on the wavelength and the power of the signal, an accuracy of the order of a centimeter to be obtained. This accuracy compensates for the small base and helps explain why the results of the two methods are quantitatively comparable. They both lead to accuracy of a few meters

for elevation measurements. They are, however, very different qualitatively.

Interferometry can function all-weather and thus offers the total certainty of a result. It is nevertheless necessary to check this result with several trials in order to detect the contribution of atmospheric effects. Furthermore, some surfaces, like forests, only yield mediocre results. In comparison, stereoscopy with an optical satellite with lateral repositioning will have a low probability of success in case of dense cloud cover. For example, if a territory is free of clouds only 10% of the time, the probability of obtaining an image pair without clouds is only 1%. If the satellite can acquire the two images by looking ahead, then behind (along-track stereoscopy), the probability stays close to 10% but no space stereoscopic system of this type currently exists.

Interferometry is superior over flat and poorly contrasted terrain, for which the correlation between optical images ''does not work''. It is also better in background quality. For weak altimetric signals, interferometry produces a rather smooth result while the correlation residuals are noisier. On the other hand, optical stereoscopy is better for terrain that is mountainous, contrasted by types of surface and differences in lighting. In these areas, the geometric problems of radar such as overlay and shortening have major impact. Besides, it is difficult to count the numerous fringes in these areas, and they may even be interrupted.

The structure of the error generated by each technique is very different: the atmosphere is ''all or nothing'' in optics. If there are clouds, no measurement is possible. In radar interferometry, the atmosphere introduces an undetectable measurement bias, except by comparing several results over the same site. The error in radar is thus more concealed.

8.2 Comparison with Global Positioning System for Measurement of Shifts

There are no existing techniques which compete with interferometry for measuring groundshifts from space. While it is possible to measure these shifts by comparing optical images taken before and after the events, one is unlikely to obtain measurements whose accuracy would be much better than a meter given current civil optical satellites. On the other hand, using specific devices on the ground, it is possible to obtain local measurements of very high accuracy, in particular with the differential global positioning system (GPS) technique.

The advantages of this technique are

1. its relative insensitivity to the nature of the ground over which the receiver is installed, provided that the ground is stable and the site is obstacle free;
2. its ability to measure in three dimensions, even though the measurement of the vertical component is less accurate than the others.

Depending on the conditions and the integration times, the measurement accuracy can be less than a centimeter.

The advantages of interferometry are

1. lower cost (excluding from the computation the cost of the satellites in both cases, GPS and radar, since the missions of the space systems go beyond geodesy);
2. a measurement with very high spatial density, at least one point per hectare in general, which allows better understanding of phenomena with complex morphology.

8.3 Prospects for Dedicated Space Missions

Interferometry has produced very spectacular results[17] by using data from radar satellites that were not designed for this goal. Nevertheless, the technique is so efficient that all future radar missions will take its particular requirements into account. In order to consider one or more space missions dedicated to interferometry, they must promise decisive advantages. We will distinguish between topographic applications and dynamic applications.

For topographic applications, the main difficulty comes from the atmospheric contribution. To suppress it, simultaneous image acquisition would be sufficient. This can be done either with a satellite with two antennas, or with a system of two very close satellites. In the first case, the system is very specific and its precision is limited by the difficulty in constructing a long beam to separate the two antennas sufficiently. This idea is, however, at the base of National Aeronautics and Space Administration (NASA)'s Shuttle Radar Topography Mission. In the second case, the system is less specific but there must be two copies of the satellite. To eliminate atmospheric effects, the two satellites must fly at an interval of a few seconds so that the atmosphere has not had time to change. The European Space Agency (ESA) experiment which consisted of flying the ERS-1 and ERS-2 satellites with a one-day interval did not succeed from this point of view. One can also imagine that one of the two satellites is passive and just receives the signal, which would lower the cost, but make the mission more specific.

For applications related to ground motions, the added value of a dedicated system mainly resides in its capacity to reuse old radar archives, in order to highlight slow shifts on image pairs taken at very different times. Other technical elements are important, but less specific: precision in the repeat cycle of the trajectory

which minimizes the topographic contribution, auxiliary instruments that provide simultaneous information on the state of the atmosphere, the capability of pointing with different angles and the capability for using varied wavelengths. Being able to reuse an archive will require a satellite with very precise technical characteristics, which must cover the same orbit as its predecessor, use the same wavelength and have the same orientation. Several mission concepts have been proposed. Their success will depend on their capability to distinguish themselves from conventional radar satellites of the future, which will all have interferometric capability.

ABBREVIATIONS AND ACRONYMS

ERS	European Radar Satellite
ESA	European Space Agency
GPS	Global Positioning System
J-ERS	Japanese Earth Resource Satellite
NASA	National Aeronautics and Space Administration
RADARSAT	Radar Satellite of the Canadian Space Agency
SAR	Synthetic Aperture Radar
Sir	Shuttle Imaging Radar
SPOT	Système Probataire Pour l'Observation de la Terre
UTM	Universal Transverse Mercator

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Elevation Modeling from Satellite Data • Polar Environments Assessment by Remote Sensing • Processing and Classification of Satellite Images

REFERENCES

1. Carl A. Wiley. 'Synthetic Aperture Radars: A Paradigm for Technology Evolution'. *IEEE Trans. Aerospace Electron. Syst.*, **21**(3), 440–443 (1985).
2. L.C. Graham, 'Synthetic Interferometer Radar for Topographic Mapping', *Proc. IEEE, 62*, **6**, 763–768 (1974).
3. H. Zebker, R. Goldstein, 'Topographic Mapping from Interferometric SAR Observations', *J. Geophys. Res.*, **91**, 4993–5001 (1986).
4. A.K. Gabriel, R.M. Goldstein, H.A. Zebker, 'Mapping Small Elevation Changes over Large Areas: Differential Radar Interferometry', *J. Geophys. Res.*, **94**, 9183–9191 (1989).
5. D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, Th. Rabaute, 'The Displacement Field of the Landers Earthquake Mapped by Radar Interferometry'. *Nature*, **364**, 138–142 (1993).
6. J.C. Curlander, R.N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, John Wiley & Sons, New York, 1991.
7. R.M. Goldstein, H.A. Zebker, C.L. Werner, 'Satellite Radar Interferometry: Two Dimensional Phase Unwrapping', *Radio Science*, **23**(4), 713–720 (1988).
8. D. Ghiglia, L. Romero, 'Robust 2D Weighted Unwrapping that Uses Fast Transforms and Iterative Methods', *J. Opt. Soc. Am.*, **11**, 107 (1994).
9. R.M. Goldstein, H. Engelhardt, B. Kamb, R.M. Frolich, 'Satellite Radar Interferometry for Monitoring Ice Sheet Motion: Application to an Antarctic Ice Stream', *Science*, **262**, 1525–1530 (1993).
10. J.J. Mohr, N. Reeh, S.N. Madsen, 'Three Dimensional Glacial Flow and Surface Elevation Measured', *Nature*, **391**, 273 (1998).
11. D. Massonnet, P. Briole, A. Arnaud, 'Deflation of Mount Etna Monitored by Spaceborne Radar Interferometry'. *Nature*, **375**, 567–570 (1995).
12. D. Massonnet, W. Thatcher, H. Vadon, 'Detection of Postseismic Fault Zone Collapse following the Landers Earthquake', *Nature*, **382**, 612–616 (1996).
13. G. Peltzer, P. Rosen, F. Rogez, K. Hudnut, 'Postseismic Rebound in Fault Step-overs Caused by Pore Fluid Flow', *Science*, **273**, 1202–1204 (1996).
14. H. Vadon, F. Sigmundsson, 'Crustal Deformation from 1992 to 1995 at the Mid-Atlantic Ridge, Southwest Iceland, Mapped by Satellite Radar Interferometry', *Science*, **275**, 193–197 (1997).
15. D. Massonnet, T. Holzer, H. Vadon, 'Land Subsidence Caused by the East Mesa Geothermal Field, California, Observed Using SAR Interferometry', *Geophys. Res. Lett.*, **24**(8), 901–904 (1997).
16. D. Massonnet, K.L. Feigl, 'Discrimination of Geophysical Phenomena in Satellite Radar Interferograms', *Geophys. Res. Lett.*, **22**(12), 1537–1540 (1995).
17. D. Massonnet, K.L. Feigl, 'Radar Interferometry and its Application to Changes in the Earth's Surface'. *Review Geophys.*, **36**(4), 441–500 (1998).

Elevation Modeling from Satellite Data

Thierry Toutin
Canada Centre for Remote Sensing, 588 Booth Street, Ottawa, Canada, K1A 0Y7

1 Introduction 1
2 Historical Background and Basic Concepts 2
2.1 Constructing the Third Dimension 2
2.2 Vision and Perception 2
2.3 Depth Perception with Remote Sensing Data 3
3 Shadow and Shade for Clinometry 3
3.1 Basic Concepts 3
3.2 Application with Visible and Infrared Images 4
3.3 Applications with Synthetic Aperture Radar Images 5
4 Binocular Disparity for Stereoscopy 7
4.1 Basic Concepts 7
4.2 Visible and Infrared Sensors 7
4.3 Synthetic Aperture Radar Sensors 11
4.4 Mixed Sensors 14
4.5 Processing, Methods and Errors 15
5 Other Methods 18
5.1 Interferometry 18
5.2 Polarimetry 19
5.3 Altimetry 20
6 Concluding Remarks 21
Acknowledgments 23
Abbreviations and Acronyms 23
Related Articles 23
References 23

Most geoscientific applications using georeferenced cartographic data need a good knowledge and visualization of the topography of the earth's surface. For example, mapping of geomorphological features is hardly feasible from a single image; three-dimensional (3-D) information has to be generated or to be added for a better interpretation of the two-dimensional (2-D) data.

Since the early emergence of earth observation satellites, researchers have investigated different methods of extracting 3-D information using satellite data. Apart from a few early stereoimages acquired with hand-held cameras during the Gemini and Apollo missions, the first experiments to extract 3-D data using stereoviewing from space began with the Earth Terrain Camera (ETC) flown on board SkyLab in 1973–74.

Since these early experiments, various analog or digital sensors in the visible or in the microwave spectrum have been flown to provide researchers and geoscientists with spatial data for extracting and interpreting 3-D information on the earth's surface. Although the shape-from-shading technique can be applied to optical sensor (OPS) images, stereoviewing using space camera or digital scanner images was, and still is, the most common method used by the mapping, photogrammetry and remote sensing communities.

However, side-looking synthetic aperture radar (SAR) data also give the opportunity to extract 3-D information using image-processing techniques appropriate to the nature of the data. With SAR data, three main methods have been developed: radargrammetry, clinometry and interferometry. Radargrammetry (similar to the stereoviewing of optical data) uses two images acquired from different viewpoints to generate a stereopair and stereoviewing. Clinometry takes advantage of the SAR shading and shadowing in the image, and interferometry uses mainly the SAR signal data instead of the image. Other methods (polarimetry and altimetry) are also used for 3-D information extraction.

This article reviews the different methods and sensors used to extract absolute or relative elevation and assesses their performance using the results from various research and commercial organizations. It also discusses the respective advantages, difficulties and constraints of the sensors, the methods and the technologies used to take into account the strength of each. It further assesses how they perform as complementary sources and systems for extracting elevation data in an operational context.

1 INTRODUCTION

At one time, a hilltop provided the best vantage point from which to observe nature's workings, but now discoveries in optics, photography and flight allow us to see the earth as never before. Advanced methods in computing and signal-processing technologies have enabled us to increase our ability to visualize, perceive and extract information from the earth's surface. Today earth observation satellites orbit our planet collecting data to produce images, which allow us to monitor, understand and plan the use of our world's resources.

Remote sensing has evolved into an important supplement to ground observations and aerial photographs

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © Copyright Property of the Government of Canada

in the study of terrain features, such as the ground elevation. With the advent of instruments that produce images from electromagnetic radiation beyond which the human eye and cameras are responsive, human ''vision and perception'' have been greatly extended.[1,2]

Why is it important that the third dimension be conveyed? Because humans are naturally able to see in three dimensions. The ''naturalness'' of a 3-D representation of reality enhances our ability to interpret 2-D imagery. Cartographers, engineers, geologists, hydrologists and other geoscientists use different 3-D viewing methods to perceive the ground elevation in order to understand better the earth's surface. For example, representation of the third dimension supplies important information about the relationship between land shape and structure, slopes and waterways, surface material and vegetative growth.

A digital elevation model (DEM), which is a digital representation of the earth's relief, is now one of the most important data structures used for geospatial analysis. Unfortunately, DEMs of usable details are still not available for much of the Earth, and when they are available they frequently lack sufficient accuracy. The digital format of a DEM made it easier to derive additional information for various applications, so that elevation modeling has become an important part of the international research and development (R&D) programs related to geospatial data.

Owing to the high spatial resolution of recent satellite sensors (Landsat–Thematic Mapper (TM), Système pour l'Observation de la Terre (SPOT)–High Resolution in the Visible (HRV), Indian Remote Sensing Satellite (IRS)–Linear Imaging Self-scanned Sensor (LISS), European Remote Sensing Satellite (ERS)–SAR, RADARSAT–SAR, etc.), a large number of researchers around the world have investigated elevation modeling and the production of DEMs. There is plenty of literature describing the methods, algorithms and accuracy assessment of DEMs,[2–6] including review articles.[7–11] They have addressed different, but generally not all, aspects of DEM generation from satellite data. A completely comprehensive and up-to-date review is not available.

Furthermore, recent research into modeling computer vision on human vision has led to the advent of new alternatives applied to satellite imagery. Current research in computer vision assumes that if a computer program can be made to ''see'' things as a human would, the algorithm must have some basis in human vision. Consequently, to develop better an understanding of the different methods used to derive elevation from satellite images, the relationship between depth vision and perception and terrain elevation representation has first to be addressed. Only the basic concepts and the historical background of natural depth perception relevant for remote sensing applications are presented. The different methods (clinometry, stereoscopy, interferometry, polarimetry and altimetry) are then presented and their applicability to the variety of data is reviewed (space photographs, digital sensor in the visible and infrared (VIR) spectrum and SAR). Finally, some concluding remarks on these methods and the future prospects of the next generation of satellites are presented.

2 HISTORICAL BACKGROUND AND BASIC CONCEPTS

2.1 Constructing the Third Dimension

Throughout history, humans have tried to represent what they saw and understood through images. Everything from cave walls, to canvasses, to computer screens have been used to express perception of our surroundings. Maps have provided one means of showing the relationship between humans and their environment. Towns, roads, rivers, mountains, valleys and where the land meets the sea have been drawn in an organized fashion for centuries. Mapmakers have always sought ways in which to represent both the location and the 3-D shape of land.

Mapmakers and other illustrators have traditionally used rendering techniques such as shading, overlapping and perspective views to create a 3-D effect. Leonardo da Vinci (1452–1519) demonstrated in 1492 the principles of optical projection. His German contemporary, Albrecht Dürer (1471–1528), produced an outline of the laws of perspective, and in 1525 he constructed samples of mechanical devices with which he made true perspective drawings of nature scenes. His devices included an apparatus for producing stereoscopic drawing.[3] In the last 100 years, many advances in representing three dimensions have been made. Stereomodels, anaglyphs or polarized images, chromostereoscopic images and holograms can provide 3-D information about our planet whereas 2-D flat images cannot.

2.2 Vision and Perception

For humans, the information provided by the eyes undoubtedly plays the dominant role in our interpretation of the environment. However, the power to integrate the viewed image, to recognize its contour, its color and its relationship with other objects indicates that the process of vision does not merely consist of ''seeing'' but also of ''perceiving'' and understanding through the central nervous system. The eye, considered as part of the brain, is fundamentally an organizer. The eye/brain, starting with the activity of the retina, is actively building a world of objects: our mental model in psychology. This

suggests that a priori knowledge is useful for a better interpretation and understanding of the image: to have a clear idea of what to look for, where to look and how to look.[12]

Perception, or perceiving, refers to the process whereby sensory stimulation is translated into organized experience. That experience, or precept, is the joint product of the stimulation and the process itself. In the ''depth'' context, the visual system (the process) creates the 3-D world (the precept) we experience from the 2-D pattern projected on to the retinas (the stimulation). But the fact that we can see depth fairly well with one eye closed, or in a photograph or painting, indicates that two eyes are not necessary for a satisfying sense of depth. This dichotomy suggests an intimate relationship between what might be called ''object recognition'' and perception of three-dimensionality. Unfortunately, at this point, we know little about how the brain identifies objects, so a large portion of ''depth perception'' is not understood.[13]

However, in modern psychology, it is accepted that depth perception is based upon four physiological cues (accommodation, convergence, binocular disparity and motion parallax) and six psychological cues (image size, linear perspective, areal perspective, overlapping, shade and shadow, texture gradient).[14] These cues are treated as additional pieces of information which, when added to a flat picture on the back of the eye, make depth perception possible. The brain combines these cues in our mental model with the 2-D picture to produce judgements about the relationship of objects in space.

2.3 Depth Perception with Remote Sensing Data

Within the field of remote sensing, it is generally recognized that psychological factors, such as perception, play a major role, but researchers devote little time or no time to studying the psychological aspects of the remote sensing processes.[12] In fact, it has been shown that the interpretation of cartographic information can be facilitated by using 3-D or perspective representations when compared with a flat 2-D display.[15] Since terrain relief modeling is based on the principal concepts related to human depth perception, what are the main cues that play a role in depth perception of remotely sensed data?

Perspective is the most popular and widely used with remote sensing data. It combines different cues such as linear perspective, overlapping and texture gradient. It also takes advantage of the viewer's conceptual knowledge of the perspective phenomena. This psychological cue is thus only used for a representation and visualization of the terrain topography combined with remote sensing images and not for terrain modeling.

Shade and shadows are familiar phenomena, which can help one to judge the size and shape of objects by providing profile representations. It is particularly helpful if the objects are very small or lack tonal contrast with their surroundings. For example, large look angle SAR images, which approximate low sun angle aerial photography in order to accentuate minute surface irregularities, are becoming important in geological investigations.

Shading is sometimes confused with shadowing. Shading is the variation of brightness exhibited in the image. It arises primarily because some parts of a surface are oriented so as to reflect more of the incident illumination towards the sensor.[16] Since shading provides cues all over the surface not just along special contours, this principle is used with the shape-from-shading technique to derive terrain slope and elevation. Shadow on a surface results when another surface intercepts the illumination from the source. It only provides localized cues (along special contours) to shape, although the shadow of a curved surface cast on another curved surface is very difficult to interpret. This principle is used to derive the elevation of specific targets such as buildings, trees, etc.

Binocular disparity and convergence are the two physiological cues when viewing imagery in stereoscopy. Binocular disparity predominates with optical images because it reproduces the natural process of human binocular vision. It is important when viewing radar images, but the shade and shadow cues also have a strong and cumulative effect on stereoradar imagery. As an example on a quasiflat terrain, the psychological cues overcome the binocular disparity when looking at the radar stereopair in pseudoscopy (apparent reversal of natural relief when inverting the viewing position of the two images).[17] Owing to the specific geometric and radiometric aspects of SAR images, it may take our brain time to assimilate this non-natural stereoviewing, mainly when both geometric and radiometric disparities are large.[18] However, since depth perception is an active process (brain and eye) and relies on an intimate relationship with object recognition, with experience radar images can be viewed in stereo as easily as VIR satellite images.[19] This principle is used in satellite photogrammetry and radargrammetry by computing the terrain elevation from the measured parallaxes (related to the binocular disparity) between the two images.

3 SHADOW AND SHADE FOR CLINOMETRY

3.1 Basic Concepts

Shadow has been used for a long time in astronomy. In 1610, Galileo observed Moon spots. The first ambiguity was to determine if these spots were shadows or low

reflectivity surfaces. Looking at their evolution as a function of the sun illumination, he concluded them to be shadows from the moon relief.[6] Later, the height of craters was determined using the lengths of shadows of the crater edges.[20]

One of the first applications of shape-from-shading was used in robot vision to detect the 3-D shape of industrial objects with a diffuse reflecting surface. Using the principle that an image of a smooth object known to have a uniform surface will exhibit gradations of brightness, or shading, the shape can be determined to map the height of this surface. Because there are two degrees of freedom to surface orientation, the reflectivity does not uniquely determine the local normal but a set of possible normal directions. These directions describe a cone, whose axis is the illumination direction, and the half-angle the incidence angle. Consequently, local operation on the brightness alone cannot be used to determine the shape of the surface and its orientations. Additional constraint must therefore be added: generally the surface is assumed to be continuous and smooth, so that the surface orientations of neighboring surface elements are not independent. If the reflectivity function and the position of the illumination source are known, the shape can thus be obtained from the shading.

The application of this concept to remote sensing data is less evident owing to the sensitivity of shading to reflective properties of the earth's surface. Even if this reflectivity function has been described with data from many experimental observations,[21–23] a general Lambertian model is often chosen for simplification when a small range of incidence occurs.[24] The local slope is then computed from the pixel reflectivity value and transformed into relative elevation by integration pixel by pixel. In other words, shape-from-shading makes uses of the sensitivity of microtopography, but it cannot provide absolute location. Some reference elevation information is needed to derive the absolute elevation. More details of VIR or SAR sensors are given in the following sections. In summary, the accuracy of this technique with remote sensing data is limited by intrinsic radiometric and geometric ambiguities:

1. The reflectivity is not only dependent on the local incidence angle, but also on the albedo related to land cover, rugosity, humidity, etc., as a function of the sensor. Miscalibration and SAR speckle are also sources of error.
2. The determined incidence angle yields a set of potential orientations whose normal directions describe a cone. Furthermore, this method only determines slopes, reference elevations have to be known and the accuracy is limited by the height error propagation.

On the other hand, cast shadow and occluded areas can also be used to extract relative heights[25,26] or to determine ground control points (GCPs).[27,28] The shadow areas occur when the ground surface is not illuminated by the source (''backslope'' related to the illumination source), and the occluded areas occur when the ground surface is not visible from the sensor (''backslope'' related to the sensor). Since the illumination source is the sensor with SAR images, the effects of these two concepts are mixed, but they are different with VIR images since the illumination source, the sun, and the sensor are different. While occluded areas are completely without information, shadow areas in VIR images have some information because the terrain receives some diffuse sun illumination partially reflected by the terrain. The impacts of shadow/occluded areas for VIR and SAR images are addressed in the next two sections.

3.2 Application with Visible and Infrared Images

For VIR images, shadow and occluded areas are different since the illumination source is the sun. Satellite VIR images are acquired from the descending path of a sun-synchronous orbit, and the local solar time is generally before or around noon (e.g. for SPOT, the local solar time of the descending node is around 10.30 a.m.). Consequently, west-looking images will have shadow and occluded areas in the same direction. Care must be taken to separate these two effects. Since the sun elevation angle around noon will generate shadow with steep slopes, it can be consistently measured only from vertical structures such as buildings or trees (in a row or isolated), or very rough terrain. However, with a low sun elevation angle (in wintertime), the relief perception of a rugged terrain is inverted.[29] Occluded areas could only be used to extract elevation information with off-nadir viewing images, but this has been never addressed.

This method using shadow length measurements is largely used with aerial photographs in which the pixel resolution is much better than the object height.[30] To our knowledge, no attempt has been realized with space photographs. Knowing the sun and sensor geometry, the same method can be applied to VIR images, such as panchromatic SPOT images, even if the resolution is coarser.[24] Using a simple trigonometric solution, elevation was computed with 3.7-m accuracy over a sample of 42 well-defined buildings. A correction for the known terrain slopes was also introduced. Since the shadow length is manually measured at the pixel unit (10 m for panchromatic SPOT images), these good results can be accounted for by the size of the building (up to 60 m with a mean of 30 m), and the large shadow cast (up to 18 pixels with a mean of 8 pixels). Hartl and Cheng[31] computerized the method and applied it over a complete

city. Only 30% of over 78 800 buildings were extracted, with calculated heights less than 20 m for 90% of them. Seventy-seven buildings were randomly selected to check their height error. The root mean square error (RMSE) was about 6 m, with only 11 buildings having errors larger than 10 m (SPOT resolution). The high building density and the overlapping of gray value were the main factors leading to the larger error.

Since shadow boundary is a key point in the process, different tools have been developed to determine it more accurately. Meng and Davenport[32] created an edge-image template using the point-spread function of the sensor. After a manual rough location of the edge, a correlation process between the template and the actual image determines the best location of the shadow edge within 1/100th of a pixel. Unfortunately, no ground truth data have been provided for checking the accuracy.

On the other hand, Shettigara and Sumerling[33] developed a four-step process using the spatial information of a panchromatic SPOT image and the spectral information of the infrared band of a SPOT image. First, an appropriate threshold to delimit shadows in the images is selected. Shadows cast by rows of trees are used to estimate the mean heights of trees. Calibration curves are then constructed to relate the actual mean heights of trees to the estimated heights. Finally, heights of industrial buildings are computed using their shadow lengths and the calibration curves, without any correction for the terrain slopes. A 3-m accuracy has been measured with only three 12-m tall buildings. Although the shadow determination is more sophisticated, the results are similar to those with the first method.[26] The advantage of this method, when compared with the first one, is that the shadow boundaries are located with sub-pixel accuracy using an optimum threshold that enables smaller building heights to be estimated. The disadvantages are that two SPOT images are used, and some ground data for the tree-row heights are necessary to determine the calibration curves. Conversely, the first method does not use ground data for height estimation. No attempt to verify the method over a complete city in a real environment has been attempted, as with the first method.[31]

Shape-from-shading can be applied to VIR imagery since information concerning the terrain is contained in multiscanner data. With homogeneous surfaces where variations in reflectance may only refer to topographic surface differences, rather than to land cover effect, a simple reflectance model can be used to derive the topographic information. Lodwick and Paine[34] used two Landsat-1 and -2 images (July and October) over an ice cap on Baffin Island, Canada, to obtain high and low sun angles and maximum difference in sun azimuth. They allowed for the reflectivity with a simplified Lambertian model[21] and two empirical models to resolve the radiometric ambiguity. The slopes being in the sun-azimuth direction resolved the conic ambiguity. They first demonstrated with a training sample that the reflectance conditions of an ice cap are non-Lambertian for a large range of incidence angles, as did Smith et al.[24] for pine-forest cover types. The first empirical model used a second-order polynomial computed over training samples. Difficulties in obtaining representative training samples were the main source of errors even if a high correlation was obtained. The second empirical model applied a simple linear model between typical maximum slopes in the ''sun-facing and away-facing'' directions and the reflectance values at the 1% level of the gray value histogram. It was the best solution to generate height differences, with broad agreement with values observed on the map.

Finally, a weighted third-order surface adjustment was carried out with nine control points to transform the 50-m posting slopes in the sun-azimuth direction into elevations. The results for the basic shape of the ice cap surface compared well with the base map. Some variations could be partly explained by the surges (melting and refreezing) of the ice-cap surface. No quantitative accuracy results were given owing to the lack of precise and digital topographic information.

These qualitative results should have generated some interest in the scientific community to expand on this work. However, no other results with different study site and data sets have been presented to date. It seems that most of the research effort on the applicability of the method has been directed towards SAR data.

3.3 Applications with Synthetic Aperture Radar Images

For SAR images, shadow and occluded areas are mixed, since the illumination source is the sensor itself. Therefore, they are completely without information and the boundaries of a cast shadow are easier to determine than with VIR images. Depending on the SAR look angles, only the steepest slopes produce shadow/occluded areas. Height determinations can then be made for the same type of vertical structures (buildings, rows of trees, etc.) for which such measurements have been made on VIR imagery. La Prade and Leonardo[25] used simple trigonometric models and knowledge of the SAR geometry to translate measured shadows into heights. However, layover lengths have to be added to the shadow lengths to take into account the good positioning of the base of the vertical structure.

Cast shadows provide only localized cues to shape. On the other hand, shade provides cues all over the surface, not only along special contours. This radiometric information (the radar backscatter) of each image pixel is used in radar clinometry to determine the local

orientation of the terrain and then the elevation by the integration of slopes.

Radar clinometry, as an adaptation of photoclinometry developed by Horn,[35] has been further developed by Wildey for the mathematical equations and later for its operational feasibility in anticipation of the Magellan mission to map Venus.[36] Radar clinometry capabilities and limitations are well known, even if the research studies have been limited.[38–41] At first, the principle appears simple: essentially the inversion of a mathematical expression of the radar backscatter in terms of the albedo and the local incidence angle. As mentioned previously, there are more severe limitations due to intrinsic radiometric and geometric ambiguities when the method is applied to SAR images of general terrain surfaces, and not only to homogeneous surfaces such as in the previous experiment with VIR images.[34]

The first radiometric ambiguity is related to the inversion of the model since it depends on two parameters. The SAR backscatter of the surface is altered if the surface properties vary from place to place. In this case assuming uniform reflecting properties (constant albedo) will recover a shape (incidence angle) that is different from the actual one. This approximation, as an extension of photoclinometry, was also used with SAR by Wildey[36] and Frankot and Chellapa.[35] More sophisticated models,[22,23,42] which take into account the SAR and surface interaction (surface geometry, vegetation, soil properties, geographic conditions, etc.) have been developed. They should now permit one to establish a more realistic backscattering model of the intensity[43] than the traditional Lambertian model used for a homogeneous surface.[44] No attempt to use them has been made due to a relative decline of the method in the scientific community during the last decade. Other radiometric problems that are not completely controlled and fully resolved are specific to SAR sensors,[43,45] namely speckle, miscalibration and ''discretized'' sampling.

The conic ambiguity is related to the definition of the incidence angle. Even when accurately determined, it does not uniquely define the orientation of the surface but a set of possible orientations. Their normal directions describe a cone with the axis being the illumination direction. Since there are two degrees of freedom to surface orientation, it takes two numbers to specify the direction of a unit vector perpendicular to the surface.[35] One brightness measurement at each picture cell only gives one equation for two unknowns at every cell. Additional constraint or assumption has to be made to resolve this conic ambiguity. One assumption implemented by Wildey[36,37] is the hypothesis of local cylindricity. It enforces a local continuity between adjacent pixels to define a local cylinder. Since there is no iteration in the solution process, the local-cylindricity method is sensitive to integration approximations and image noise. It then tends to accumulate these effects along the full DEM reconstruction leading to ''pseudo systematic'' errors.[4]

Other assumptions or constraints to resolve the conic ambiguity implemented by Frankot and Chellapa[38] are the notions of integrability and regularization. The first states that heights can be integrated along any path since these values are independent of the integration path. This constraint acts as a smoothing process. The second limits the amount of allowable oscillation in the computed terrain surface. Furthermore, they used an iterative approach starting from an approximate existing DEM. Differences between gray values of the real image and the SAR synthetic image, predicted from the latest estimated DEM, are used to improve the terrain's slopes and heights and to converge to the final DEM. In conjunction with the two previous integrability and regularization constraints, this approach tends to spread out the speckle noise errors instead of propagating them.

Thomas and Kober[46] expanded this iterative approach to multiple images. Some spot heights derived from stereoscopy or other sources can also be added as supplementary constraints of the estimated heights. Using multiple image algorithms gives better stability and robustness with noisy images during the iteration procedure, as well as a faster convergence. However, it does not fully resolve the two basic ambiguities.

For the conic ambiguity, Guindon[41] quantitatively demonstrated that the SAR image gray level is not an effective indicator of local incidence angle, and hence is not an accurate measure of the overall local terrain surface normal direction. It is only a strong indicator of the range component of the terrain slope. It can therefore only be used to derive elevation profiles for individual image range lines. Since no significant detectable information is available about azimuthal slope, an additional source of ''azimuthal control'' data is required to tie the adjacent range line elevation profiles to a common and absolute origin. Paquerault and Maître[43] therefore developed a two-step strategy to compute these two components of the incidence angle. First, they computed the range component from the backscatter gray-level pixel, and integrated it along a range line. They then applied a contextual Markovian strategy to modify successively, in a random order, each pixel orientation of the full image. This second step enabled them (i) to take into account the azimuth component of the incidence angle, (ii) to link together the adjacent range lines elevation profiles and (iii) to reduce the noise error propagation.

Despite the developments in the mid-1990s, SAR shape-from-shading remains a marginal technique, applied mainly in difficult situations such as tropical land cover or extraterrestrial sites without ground truth.

This is mainly due to the fact that the radiometric ambiguity between the terrain albedo, the radar backscattering cross-section and the incidence angle is rarely solved, except on a homogeneous terrain surface with a Lambertian model. However, a large part of the earth, without cartography, approximates to these homogeneous surfaces.

4 BINOCULAR DISPARITY FOR STEREOSCOPY

4.1 Basic Concepts

In about 1600, the German astronomer Johannes Kepler (1571–1630) gave the first precise definition of stereoscopy, and a Florentine painter, Jacopo Chimenti, produced one of the first hand-drawn stereopicture pairs (Wicar Museum, Lille, France, ca. 1600).

In modern photogrammetry, ''stereoscopy is the science and art that deals with the use of images to produce a 3-D visual model with characteristics analogous to that of actual features viewed using true binocular vision''.[47]

In stereoscopic space perception, two major cues are used: the convergence and the binocular disparity. Convergence is the ability to focus the optical axes of the two eyes on to a single object. The sensing of the amount of muscular tension in the eyes resulting from different convergence angles provides a cue to the absolute distance to the viewed point. The binocular disparity (or parallax) is the disparity or the ''difference'' between the images of an object projected on to each retina. The degree of disparity between the two projected images depends on the convergence angle. The binocular disparity is considered the most important perception cue over medium distance, and is the only one used in stereophoto- or radargrammetry for quantitative elevation extraction.

The three main applications of stereoscopy are

1. as an interpretation aid in qualitatively recognizing the 3-D form of an object;
2. the quantitative estimation of slopes and relative heights; and
3. the quantitative and precise measurement of 3-D coordinates of planimetric and altimetric features.

In the last 50 years, first optico-mechanical and later analytical and digital 3-D photogrammetric systems capitalizing on the binocular parallax and convergence cues have been developed for aerial photographs to meet the needs of these three applications of stereoscopy (especially the last). It was Helava who developed the main concepts behind these analytical and digital systems in 1957. Most stereoworkstations have now been adapted to process stereodata from the same satellite sensors (space photo, VIR or SAR), but only a few can simultaneously process mixed sensor stereopairs. Stereoimages displayed on the screen are separated either spatially, radiometrically or temporally. Photogrammetric principles for space photographs (colinearity and coplanarity conditions) and their equivalent for remote sensing data mathematically solve the relationship between 2-D image coordinates and 3-D ground coordinates. The hardware and software to derive information from the 2-D digital imagery have thus allowed the mapping process to become more automated,[48] but not completely, with occasional unmatched expectations.[49]

Among all the new developments of stereoworkstations, DEM generation is an important R&D topic. In fact, any satellite data can be used to generate a stereoscopic pair and simulate the natural depth perception, as soon as the terrain is imaged from two different viewpoints. Since the stereoscopic methods to extract elevation, based on the binocular disparity and parallax, are ''more or less'' the same, stereoscopic capabilities of different sensors are first analyzed. The processing, the methods and the error propagation are then addressed.

4.2 Visible and Infrared Sensors

Two main categories of VIR sensors have to be considered: space cameras and digital scanners.

4.2.1 Space Cameras

For a long time, space camera technology remained in the military domain. Since the techniques and technologies of space photographs are derived from classical aerial photographs, photogrammetrists have postulated that space cameras would be the next logical step for topographic mapping.

The first significant satellite photogrammetry experiment was done using imagery on the Apollo 15, 16 and 17 missions to the moon. A lunar control net with a 30-m relative accuracy in the three coordinate axes was generated, and 1 : 25 000 topographic ortho-photo maps were produced.[50] After a few early stereo hand-held photographs acquired during the Gemini and Apollo missions, the ETC experiment on board Skylab-D in 1974 produced the first along-track stereoviewing images from space. One of the first attempts to measure heights from space images was made by Mott.[51] He reported an RMSE of 120 m for a strip of four Skylab models and concluded that the minimum contour intervals to be plotted should be approximately 250 m.

The ETC was followed by the German Democratic Republic's MKF-6M multiband camera flown on the Soviet spacecraft Soyuz 22 and Salyut in 1976. None of these were metric cameras (MCs) capable of producing

acceptable accuracy, even with an image pixel size of 17–30 m. Furthermore, wide spaces between ground tracks hindered the mapping of large areas.[9,52]

In 1978, the USSR flew the KATE140 MC on Salyut, acquiring panchromatic images with 60-m resolution. Later, the USSR developed RESURS, a series of remote spacecraft based on the recoverable Vostok capsule. They carried different multiband MC (KATE-200, KFA-1000, LK-1000 and MK4) with retrievable film on missions lasting between 2 and 4 weeks. The ground resolution varied from less than 5 to 30 m. At the same time, the German Zeiss MC initiated by the European Space Agency (ESA) took panchromatic and near-infrared images during the Space Shuttle STS-9 mission between November 28 and December 7, 1983. Later, the ITEK Large Format Camera (LFC) initiated by the National Aeronautics and Space Administration (NASA) was flown on the Space Shuttle STS-41-G on October 5–10, 1989. The LFC has a Forward Motion Compensation (FMC) system to produce a better image quality.

Since these different MCs have along-track stereo-capabilities, elevation can be derived. However, most of the research work has been on the estimation of the stereoacuity,[9,50,51] or on the evaluation of planimetric and altimetric accuracies over a limited number of points.[53] Other results were reported by various authors at the MC Workshop held in Oberpfaffenhofen, Germany, in February 1985, and at the ACSM–ASPRS Annual Meeting held in Washington, DC, in March 1986. They mainly used an analytical stereoplotter for which earth curvature correction has been added to the normal photogrammetric bundle adjustment, but not for varying atmospheric conditions.[54]

Two experiments generated contour lines, one with MC data[55] and the other with LFC data,[56] both on analytical stereoplotters. In the first experiment, 50-m contour lines were digitally plotted and compared with 1 : 25 000 topographic maps. It shows a standard error of about 30 m with larger errors in the steepest areas. In the second experiment, difficulty in extracting 20-m contour lines was reported. The height accuracy of the extracted contour lines was only computed from 30 points, and was 15 m on average for stereopairs with a base-to-height (B/H) ratio of more than 0.6. Although less significant, these LFC stereopair results are better owing to the FMC system. The accuracy obtained was of the same order as the predicted accuracy for generating contour lines at 20–30-m intervals with LFC data,[50] but it does not completely meet the requirements of cartography, particularly in mountainous areas.[52]

The main reasons why these data have not been used for operational DEM production are as follows:[55]

- the limited distribution of the data relative to the amount of acquired images;
- the experimental nature of the data and system, and the lack of decision to make it fully operational and repetitive; and
- the relatively poor quality of the data.

Consequently, the stereocapabilities of the MC and LFC have mainly become a source of planimetric feature content for mapping using traditional photogrammetric techniques with an analytical stereoplotter.[57]

4.2.2 Digital Scanners

To obtain stereoscopy with images from satellite scanners, two solutions are possible:

1. along-track stereoscopy from the same orbit using fore and aft images; and
2. across-track stereoscopy from two different orbits.

The latter solution was the most used since 1980: first with Landsat from two adjacent orbits, then with SPOT using across-track steering capabilities and finally with IRS 1C/D by ''rolling'' the satellite. In the last few years, the first solution, as applied to space frame cameras, gained renewed popularity. Examples are, the Japan Earth Resources Satellite (JERS)-1 OPS, the German Modular Opto-electronic Multispectral Stereo Scanner (MOMS), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), the IRS P5 and most of the high-resolution satellites such as Orb-View1 and Quick-Bird. Only Ikonos will have simultaneously along- and across-track capability.

4.2.2.1 ''Adjacent Orbit'' Stereo In the case of Landsat (either Multi Spectral Scanner (MSS) or TM), stereoscopic acquisition is only possible from two adjacent orbits since the satellite acquires nadir viewing images, and the tracking orbit ensures a repeat path consistent within a few kilometers. In fact, since the mean B/H ratio with Landsat MSS is around 0.1, one needs relief of about 4000 m to generate a parallax of five Landsat MSS pixels (80-m resolution). Owing to its quasipolar orbit, the coverage overlap grows from about 10% with a B/H ratio of 1.8 at the Equator to about 85% with a B/H ratio of 0.03 at 80° latitude. From 50° north and south the coverage overlap (45%) with a B/H ratio less than 0.12 permits quasioperational experiments for elevation extraction. Welch and Lo[58] extracted the elevation of 10 control points from different color-photograph stereopairs acquired from Landsat-1. They designed a precise parallax-bar instrument with various viewing magnifications (10–30×), and obtained an RMSE for the elevation between 300 and 500 m. They noticed a

large error in the parallax difference measurements on the analog photographs. Digital processing should thus allow a better parallax measurement accuracy.

Simard[59] and Simard and Slaney[60] then used digital Landsat MSS and Landsat TM stereopairs, respectively with a B/H ratio of 0.11 over the mountains (2000-m elevation variation) in British Columbia, Canada. Ehlers and Welch[61] also applied the method using Landsat TM data with a larger B/H ratio (0.17) over a low relief (500-m elevation variation). For the three experiments, the images are first corrected for the geometric distortions related to the platform, sensor and look geometry. The residual misregistration (or parallax) between the images of the stereopair thus reflects the relief effect. Cooper et al.[62] also suggested correcting for the earth's curvature, if it is not done. Since the east–west component accounts for almost 98% of the total parallax, a simplified one-dimensional model to compute the elevation from the measured parallax can be used (Equation 1):[59]

$$\mathrm{d}h(x, y) = \mathrm{d}p(x, y)\frac{H}{B} \quad (1)$$

where dh and dp are the relative height and parallax at each image point (x,y), respectively.

This equation can be modified for the different orientation of the scan lines, but the variation in B/H is less than 0.004.[61] These models are an approximation of the stereogeometry, which is only good because of the coarse satellite image resolution (30–80 m), the poor B/H (0.1–0.2) and the final expected accuracy (50–100 m).

The parallax for each pixel is measured using a hierarchical cross-correlation technique with variable reference window size.[59] The window size for the search window can also be adjusted according to image content and signal-to-noise ratio.[61] More details on the method are given in section 4.5.

Qualitative and visual evaluation of the resulting DEM or the derived contour lines show generally good agreement when overlaid on the ortho-rectified Landsat MSS imagery[59] or with the existing map contour lines.[61] Quantitative evaluation gives an RMSE of about 45 m when compared with independent check points (ICPs),[60,61] and of 60–70 m with a low-precision 1 : 250 000 map-derived DEM.[62] The resulting variations of this last study can be accounted for by the low-precision map DEM and from the correlation process using edge matching instead of gray-level matching. More details on the correlation results and performances are given in section 4.5.

The stereoscopic capabilities with Landsat data still remain limited because:

- it can be applied for large areas only in latitudes higher than 45–50° north and south;
- it generates a small B/H ratio leading to elevation errors of more than 50 m; and
- only medium to high relief areas are suitable for generating sufficient vertical parallax.

4.2.2.2 "Across-track" Stereo To obtain good geometry for better stereoplotting, the intersection angle should be large in order to increase the stereoexaggeration factor, or equivalently the observed parallax, which is used to determine the terrain elevation. According to Light et al.,[63] B/H ratios of 0.6–1.2 are a typical value to meet the requirements of topographic mapping. The SPOT system with its across-track steering capabilities (±26°) can generate such B/H ratios. In conjunction with a finer pixel size (10 m for a panchromatic image), a more precise model has to be used to transform the parallax extracted from the raw SPOT images into an elevation value. Since the perspective of the SPOT push-broom scanner is a conico-cylindrical perspective (conical for imaging a line and cylindrical for the displacement of the satellite), new geometric and stereoscopic models, equivalent to colinearity and coplanarity equations in photogrammetry, have to be developed for the generation of a precise DEM. To transform image coordinates or parallax into map coordinates, the parametric model has to take into account:

- the distortions relative to the platform (position, velocity, orientation);
- the distortions relative to the sensors (orientation angles, instantaneous field of view, detector signal integration time);
- the distortions relative to the earth (geoid–ellipsoid including relief); and
- the deformations relative to the cartographic projection (ellipsoid–cartographic plane).

Some of the first studies were undertaken at the Institut Géographique National (IGN), France, from raw-type simulated stereo SPOT data generated by the Centre National d'Études Spatiales (CNES), the French Space Agency.[64–66] These three studies reported a 3-m accuracy in both planimetry and altimetry with the simulated stereopair (B/H ratio of about 1.1). Furthermore, 20-m contour lines were generated using automatic correlation, and qualitatively compared with contour lines generated from aerial photographs with an analytical stereoplotter.[65] Quantitative results have also been presented using the Matra Traster analytical stereoplotter with the same simulated SPOT stereopair[67] and showed an elevation error of 5 m with 80% confidence.

Simulation work on georeferenced-type SPOT data was also done in Canada and a DEM was generated with an RMSE of 5.7 m from geo-referenced stereoimages corrected for systematic distortions (satellite, sensor,

earth's curvature and rotation) with a B/H ratio of 0.5. Other studies with simulated SPOT images were later conducted around the world (US, UK, Australia, Sweden, etc.) using analytical stereoplotter or automatic correlation methods.[62,68]

After the launch of SPOT-1 in February 1986, CNES sponsored the SPOT Preliminary Evaluation Program (PEPS) to assess SPOT capabilities for thematic and topographic mapping. In preparation for the launch and early PEPS data, considerable research was carried out to develop robust and rigorous mathematical models describing the specific acquisition geometry of the SPOT–HRV sensors.[64,69–76]

Most of these researchers used the photogrammetric solution (colinearity conditions for the conic perspective of a single image line) and took into account the displacement of the satellite (cylindrical perspective) to link the equations. Since the parameters of neighboring lines are highly correlated, and satellite positions and attitude can be computed from on-board recording systems, the mathematical equations can be reduced to a minimum of 8–10 unknowns depending on the development and implementation of the solution. Only some of them have been adapted to process stereodata (coplanarity condition).

Most of the results with real data were presented at the SPOT-1 Image Utilization, Assessment, Results Symposium held in Paris, in November 1987.[77] Academic research results rather than operational systems or projects dominated the Symposium. In general, an accuracy of 10 m or less for the planimetry and the elevation was achieved. The differences are mainly dependent on the accuracy of the SPOT geometric modeling and its implementation in the workstation since good cartographic data were generally used. For DEM generation two main processing methods have been presented:

- using an analytical stereoplotter or
- using a digital image analysis system.

The first method uses a stereoanalysis system with SPOT data on transparency photographs. Following the Traster System developed by Matra in conjunction with IGN, France, for the software aspects,[67] different universities or mapping agencies around the world developed solutions in collaboration with photogrammetric instrument manufacturers: the Kern DRS-1,[78] the Zeis Planicomp,[79,80] the Wild Aviolyt[81] and the Canadian NRC Anaplot-1.[82] Contour lines can be interactively stereoplotted to further generate DEM. Petrie[83,84] provided useful information on the progress of analytical stereoworkstations and their processing capabilities. Later, Hottier and Albattah[85] described a method by resampling raw SPOT stereoimages to generate a pair of quasi-epipolar images that is suitable for stereoplotting on an analog stereoplotter. The processed stereoimage pair was plotted on a Wild AG1 without either excessive Y-parallax or significant loss of information.

When digital photogrammetric workstations became more available, the different analytical solutions and software were ported into these fully digital systems. Some of them also took advantage of low-cost personal computers.[86–88] Dowman et al.,[89] Heipke[90] and Walker and Petrie[91] are useful sources of information on the progress of digital stereoworkstations and their processing capabilities.

The second method uses fully digital images and processing without any stereoviewing capabilities most of the time. The DEM is automatically derived from the digital SPOT images using correlation techniques and a geometric SPOT model.[92–96]

The IRS 1C and 1D also have across-track stereocapability. This is achieved by rolling the satellite rather than steering the instrument. In anticipation of the planned stereo IRS 1C data, Malleswara et al.[97] used the "adjacent orbit" stereotechnique with the IRS 1A linear imaging self-scanning sensor data, which does not have across-track stereocapability. Using the same methodology with Landsat data (geo-referenced data, least-squares matching and approximated elevation modeling) they generated a DEM over three study sites with stereoimages displaying various overlaps (16–27%) and B/H ratios (0.12–0.14).

The DEMs were then checked with 30 ICPs, and showed an error with 90% confidence of about 35 m with a slight correlation between the error and the B/H. Jacobsen[98] carried on the investigation with three IRS 1C images (two off-nadir, $B/H = 0.8$, and one nadir) over Hannover, Germany. When compared with over 80 ICPs he obtained an accuracy of ± 1.1 pixels (6.5 m) in planimetry and elevation. No DEM was extracted. The results are worse than those generally obtained with SPOT or with other IRS 1C data because:

- he used a nonparametric solution instead of using a rigorous photogrammetric solution adapted to the specific geometry and characteristics of the LISS sensors;[99,100] and
- the attitude data are not always consistent and accurate.[99]

Few results on DEM extraction have been published, owing to the limited availability of stereo IRS images. Cheng et al.,[101] generated a DEM (least-squares matching, rigorous photogrammetric modeling) from raw IRS 1C LISS stereoimages ($B/H = 0.52$) over a mountainous area in Arizona, USA (elevation variation of 2100 m). They reported an elevation accuracy of about 10 m when compared with both ICPs and digital DEM of the United

States Geological Survey (USGS). It is still worse (1.7 corresponding pixels) than results on the same type of relief with SPOT (about one corresponding pixel or better). This is most likely due to the inconsistent attitude data. Further work with IRS 1D could provide a better answer if stereodata were more available to researchers.

4.2.2.3 "Along-track" Stereo Launched in 1992, JERS had the capability to acquire along-track stereoimages by the use of forward and nadir linear array sensors, named OPS. The 15° forward-looking image and the nadir-looking image (18-m ground resolution) generate a stereopair with a B/H ratio of 0.3. The simultaneous along-track stereodata acquisition gives a strong advantage in terms of radiometric variations versus the multidate across-track stereodata acquisition. This was confirmed by the very high correlation success rate (82.6%),[102] which can compensate for the weaker stereogeometry.

Few results on DEM extraction from JERS data[102,103] have been presented after the first Japanese experiment.[104] All experiments have generated DEMs with the correlation method and photogrammetric solutions. Although the methods used are approximately the same, Westin[103] obtained results (20 m) twice as good as Maruyama et al.[104] or Raggam and Almer.[102] This 20-m accuracy corresponds to one pixel spacing, which needed an automatic parallax measurement accuracy of better than one-third of a pixel with a 0.3 B/H ratio. Even when the GCPs were separated from over a 200-km distance on the same image strip, the interpolated distance and the distribution of the control data did not affect the elevation accuracy.

The German MOMS is another push-broom scanner with along-track stereocapability. This development started with MOMS 1 in 1979, with the first experiment flown mainly as a technical verification of the instrument line. In a second step, experimental MOMS 2 data (ground resolution of 13.5 m) were acquired during the German space laboratory mission in 1993 for testing the map generation potential. Since the system has fore-and-aft scanners (±21°), a B/H ratio of 0.8 can be obtained. Both methods, with an analytical plotter[105] and with digital correlation,[106] have been used over an Australian test site to produce 10- and 5-m intermediate contour lines and DEMs, respectively. Checked only with ICPs, a DEM error of 16 m was measured. Qualitative evaluation of the contour lines achieved very good consistency (even for the 5-m contour lines) with the ground truth. It thus enables scale-ups to 1 : 25 000 to be mapped. These better results are accounted for by the superiority of human depth perception when compared with automatic correlation techniques with this data set.[105] Owing to the bad quality of the control data in the Australian data set, they expected to improve consistently the height accuracy to 5 m with the third MOMS 2P/PRIRODA mission to be flown on the Russian space station MIR. However, the first experiment with these third mission data (18-m resolution and 0.8 B/H ratio) showed a degradation of the DEM accuracy to 25–30 m[107] while a second experiment achieved a 10-m accuracy.[108] The large discrepancy between these last two experiments can be accounted for by the different type of relief and/or by the different geometric modeling of the 3-D-array scanner. Future studies will confirm the potential accuracy of this VIR scanner.

4.3 Synthetic Aperture Radar Sensors

In the 1960s, stereoscopic methods were first applied to radar images to derive ground elevation leading to the development of radargrammetry.[109] It was shown that some specific SAR stereoconfigurations would produce the same elevation parallaxes as those produced by air photographs. Consequently, elevation could be derived on traditional stereoplotters. They can only be used to measure target elevations. Furthermore, Carlson[110] developed a single-path technique to generate radar stereoimages, which made it easier to view and to measure parallax for elevation computation than the traditional two-path technique. However, the lack of radar stereopairs led mainly to theoretical studies[111–114] or simulated data processing experiments.[115,116]

During the 1980s, improvements in SAR systems, with parallel investigations into the theory, allowed the demonstration of stereoradar with same-side or opposite-side viewing. These theoretical studies and practical experiments[114] confirmed that the opposite-side stereoconfiguration is superior to the same-side stereoconfiguration. The difficulty in using these geometrically superior configurations stems from the illumination differences that are too pronounced to be stereoscopically viewed and the difficulty of finding corresponding points and features. Figure 1 illustrates the intersection geometry with the radar parallax due to elevation for different stereoconfigurations (same versus opposite side; steep versus shallow look angles).

To obtain a good geometry for better stereoplotting, the intersection angle (Figure 1) should be large in order to increase the stereoexaggeration factor, or equivalently the observed parallax, which is used to determine the terrain elevation. Conversely, to have good stereoviewing, the interpreter (or the image matching) prefers a stereopair as nearly identical as possible, implying a small intersection angle. Consequently, large geometric and radiometric disparities hinder both stereoviewing and precise stereoplotting. Since the reduction of one disparity could compensate for the other disparity, a compromise

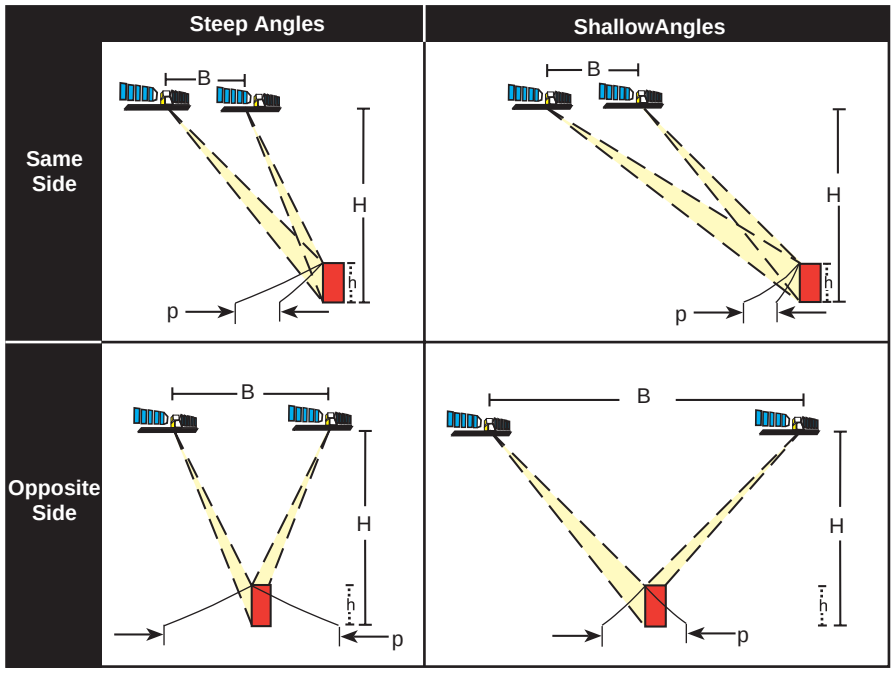

Figure 1 The intersection geometry with the radar parallax (p) due to the terrain elevation (h) for different stereo SAR configurations (same side versus opposite side; step versus shallow look angles). (Courtesy of the Canada Centre for Remote Sensing, Natural Resources Canada.)

has to be reached between better stereoviewing (small radiometric differences) and a stronger geometry and plotting (large parallax).

The common compromise for any type of relief is to use a same-side stereopair, thus reducing both disparities. Unfortunately, this does not optimize the full potential of stereoradar for all terrain topography. A compromise to reduce the radiometric difference of an opposite-side stereopair is to invert the radiometry of one image.[118,119] When processing digital images, Fullerton et al.,[119] added a local brightness change to exclude some image features from the radiometric inversion. A low frequency or a sparse DEM can also be used to reduce the geometric disparity, as applied with success to iterative hierarchical SAR image matching.[120] Another potential compromise is to use an opposite-side stereopair over rolling topography.[121] The rolling topography reduces the parallax difference and also the radiometric disparities (no layover, shadow and little foreshortening), making possible stereoviewing and good stereoplotting.

However, with spaceborne platforms, parallel flights (from the opposite or same side) are very rare. Even sun-synchronous satellite orbits are only parallel near the Equator. Elsewhere, crossing flight lines or convergent stereoconfigurations must be considered. No differences exist between computations for parallel flight lines and those for crossing flight lines if rigorous intersection geometry is applied. This was confirmed with the Spaceborne Imaging Radar (SIR) A and B shuttle missions of 1981 and 1984.[117,119,120] The last two studies processed radar photographs on an analytical stereoplotter, the Kern DSR-1, which was adapted to process stereo SAR images.[124] The first study used a fully digital method with iterative hierarchical matching. The results achieved for the DEM were of the order of 60–100 m, due mainly to the poor SIR A resolution or radiometric and geometric image quality. Furthermore, with the SIR B SAR system, stereopairs with intersection angles ranging from 5° to 23° can be created.[125]

Since the launch at the beginning of the 1990s of different satellite sensors (Almaz, ERS, JERS, etc.) radargrammetry again became a ''hot'' R&D topic. First, the Russian Almaz-1 SAR system could have acquired images with different angles to obtain stereoimages in the latitude range 0–72°. Yelizavetin[126] digitally processed two images with 38° and 59° look angles acquired over a mountainous area of Nevada, USA. No quantitative results were given. Stereoscopy with ERS–SAR data is achieved using an image with its normal look angle (23°) and a second image with the roll–tilt mode angle (35°) to generate a same-side stereopair.[127,128] It can also be done with two normal look angle (23°) images from ascending and descending orbits to generate an opposite-side stereopair.[121,129] Comparison of these research results, 20 versus 40 m,[130] confirmed

the superiority of the opposite-side stereopair. With the JERS–SAR, stereoscopy has been achieved with adjacent orbits generating a small overlap with a small intersection angle.[101] The digital correlation method was used to generate a 75-m accurate DEM.

Results obtained with simulated and spaceborne SAR data can be summarized as follows:

1. Kaupp et al.[114] found that the optimum intersection angles are about 40–45°;
2. Domik[115] showed that the best subjective stereoimpressions were obtained with shallow look angles (50–70°), and at an intersection angle of 20°;
3. Leberl et al.[122,124] showed that the highest accuracy is not necessarily achieved with the largest intersection angles;
4. Fullerton et al.[118] noted that higher ground resolution does not necessarily lead to higher height accuracy; and
5. better accuracy is more consistently achieved with an opposite side stereopair.[119,121,130]

These reported results are inconsistent and practical experiments do not clearly support theoretical expectations: for example, larger intersection angles and higher spatial resolution do not translate into higher accuracy. In various experiments, accuracy trends even reverse, especially for rough topography. Only in the extreme case of low relief does accuracy approach theoretical expectations.

By analogy with photogrammetry, theoretical error analyses were first developed by Rosenfield[111] and La Prade.[113] They related an error of an exterior orientation element in the left and right images to the resulting error in the stereomodel. These first analyzes were mainly limited to absolute errors and to comparing same-side with opposite-side stereo. Leberl[114] applied a more general approach to error propagation, identifying both relative and absolute errors regardless of the stereoconfiguration. As a summary, an estimation of the error in the cross-track and elevation coordinates, E_x and E_h, respectively, due to an error in range, E_r, for the measurement of a target in the stereomodel is given by Equations (2) and (3):

$$E_x = \frac{(\cos^2\theta_L + \cos^2\theta_R)^{1/2}}{\sin\Delta\theta} E_r \tag{2}$$

$$E_h = \frac{(\sin^2\theta_L + \sin^2\theta_R)^{1/2}}{\sin\Delta\theta} E_r \tag{3}$$

where θ_L and θ_R are the look angles of the left and right images, respectively, and $\Delta\theta$ is the intersection angle, i.e. the difference between the two look angles.

As shown in Equations (2) and (3), the error modeling accounts only for SAR geometric aspects (look and intersection angles, range error) and completely neglects the radiometric aspects (SAR backscatter) of the stereopair and of the relief. This error propagation modeling can then only be applied when the radiometry has a minor role and impact with respect to the geometry, such as during the stereomodel set-up with GCPs, which are radiometrically well-defined points.[131,132] The residuals error of the least-squares adjustment of the stereomodel are thus correlated with the intersection angle.

Since SAR backscatter and consequently the image radiometry are much more sensitive to the incidence angle than the VIR reflectance, especially at low incidence angles,[11] the theoretical error propagation has a major limitation as a tool for predicting accuracy and selecting appropriate stereoimages for DEM generation. Care must therefore be taken in attempting to extrapolate VIR stereoconcepts to SAR.

Before RADARSAT, Canada's first earth observation satellite, was launched in November 1995, it was difficult to acquire different stereoconfigurations to address this point precisely, namely the impact of radiometry in the error propagation. RADARSAT (Figure 2) with its

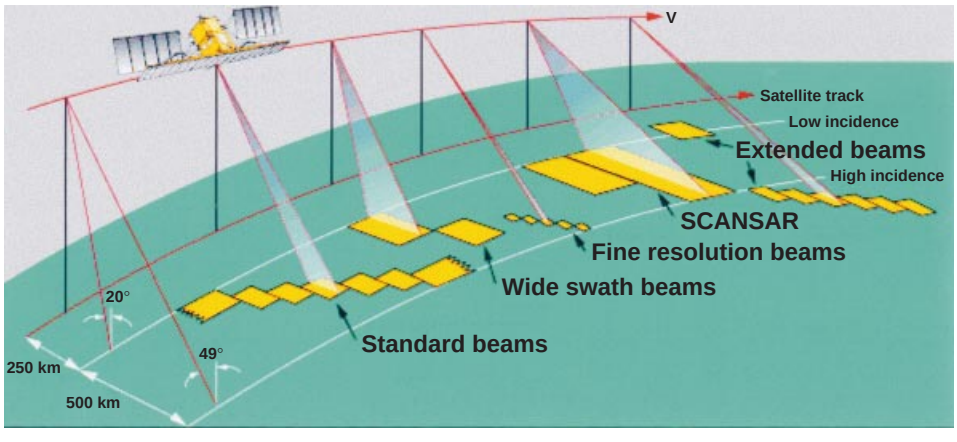

Figure 2 Operating modes of RADARSAT SAR. (Courtesy of the Canadian Space Agency.)

various operating modes, imagery from a broad range of look directions, beam positions and modes at different resolutions[132] fills this gap. Under the Applications Development and Research Opportunity (ADRO) program sponsored by the Canadian Space Agency (CSA), researchers around the world have undertaken studies on the stereoscopic capabilities by varying the geometric parameters (look and intersection angles, resolution, etc.). Most of the results were presented at the final RADARSAT ADRO Symposium held in Montreal, Canada in 1998.[124] There was general consensus in the results of the DEM extraction accuracy: slightly more than one resolution for the fine-mode (12 m), and slightly better for the standard mode (20 m), whatever the method used (digital stereoplotter or image correlation). Relative elevation extraction was also addressed from a fine-mode RADARSAT stereopair for the measurement of canopy heights in the tropical forest of Brazil.[17]

However, there was no significant correlation between the DEM accuracy and the intersection angle or the vertical parallax ratio.[135] In fact, most of the results showed that the principal parameter that has a significant impact on the precision of the DEM is the type of the relief (and its slopes).[136] The greater the variation between two look angles (e.g. 23° and 47°), the more the quality of the stereoscopic fusion deteriorated. This cancels out the advantage obtained from the stronger stereogeometry. On the other hand, although a higher resolution (fine mode) produces a better quality image, it does not change the stereoacuity for a given stereoconfiguration (e.g. intersection angle) and it does not significantly improve the DEM accuracy. Furthermore, although the speckle creates some confusion in the stereoplotting, it does not degrade the DEM accuracy because the correlation method or the human stereoviewing ''behaves like a filter''. Preprocessing the images with an adaptive speckle filtering does not improve the DEM accuracy;[137] it can slightly reduce the image contrast and smoothes the relief (especially the low relief).[135]

Since the type of relief is an important parameter in the DEM accuracy, it is strongly recommended that the DEM accuracy be ascribed values that reflect the different areas of the relief. Furthermore, in the choice of a stereoscopic pair for DEM generation, both the geometric and radiometric characteristics must be jointly evaluated taking into account the SAR and surface interaction (surface geometry, vegetation, soil properties, geographic conditions, etc.). The advantages of one can compensate for the deficits of the other.

4.4 Mixed Sensors

Owing to the increasing amount of image data, it is very common to have data from different sensors over the same terrain area. The traditional stereoscopic technique can thus be applied. By perceiving the different radiometry in the brain, the stereoscopic fusion of mixed sensors can also provide a virtual 3-D model of the terrain surface. Few results have been published on the use of mixed stereosensors to generate DEMs. Welch et al.[138] used a 23° viewing angle SPOT image (band 3) and a Landsat TM image (band 4) with the automatic stereocorrelation capability of the Desktop Mapping System.[85] Comparison of profiles for the stereoextracted DEM with the existing 1 : 50 000 topographic maps indicated an RMSE of about ±100 m. This large error is mainly due to the polynomial coregistration process instead of a rigorous parametric geometric model. In fact, Raggam and Almer[139] generated a 50-m accurate DEM from a 23° viewing angle SPOT image (band 1) and a Landsat TM image (band 2). A proper relative registration process was used to generate the epipolar images for the measurement of corresponding image points with an automatic stereocorrelation process. They reported a 65% success rate in the correlation step due to the radiometric difference between the two images and to the homogeneous nature of some areas (snow fields, glacier or shadow). Human interaction is still required to reject blunders or to fill the mismatched areas in order to optimize the DEM results. This requires a digital stereoworkstation, not only with automatic matching, but also with full stereoscopic capabilities (GCP and tie points stereoplotting, 3-D DEM editing, 3-D cartographic feature extraction, etc.).[140]

In fact, the brain can generate the perception of depth combining, for example, the spatial information from a SPOT panchromatic image and the spectral information from a Landsat TM image for stereoplotting when image matching fails. Toutin[140] reported an altimetric accuracy of 37 m for the elevation data extracted from a raw 26° viewing angle SPOT P and Landsat TM (band 1) stereopair. The 10-m resolution of the SPOT P image, and the fact that elevation data are extracted directly from the raw image (no polynomial coregistration or epipolar image resampling), account for the better results. More difficulties have been reported by Akeno[141] when trying to generate a DEM from a National Oceanic and Atmospheric Administration (NOAA)/Advanced Very High Resolution Radiometer (AVHRR) and Landsat MSS stereopair owing to the large resolution difference (1 km versus 80 m). He registered the two images using image-to-image correlation and degraded the Landsat MSS image to the AVHRR resolution. He reported 320-m accuracy over the good matched DEM points. The main difficulty was to obtain the sub-pixel accuracy in the correlation process, applied in the NOAA/AVHRR image rectification and the parallax measurement.

When two optical images are not available, a stereopair can be generated and viewed by combining optical and SAR images. Moore[142] first addressed the principle theoretically. He used simultaneously acquired infrared line-scanner and SAR images. In neither case was the visual stereoeffect perfect except near a 45° viewing angle. Various scaling factors were also applied to different areas of the stereopair to obtain the proper stereoeffect for the height determination. No quantitative measurement was achieved owing to the lack of an ''adapted'' stereoplotter.

Further evaluation was done with SIR B and Landsat TM images.[143] Only moderate results over 27 extracted points were reported, mainly due to pixel offset error in the registration of the images and the approximated angular values used in the simplified elevation computation equation. Using a better parametric solution, Raggam et al.[144] extracted a DEM from a multiband SPOT and airborne SAR stereopair. Since no meaningful results can be obtained from automatic image correlation, they interactively measured 500 corresponding image points and computed the elevation off-line. Results of the comparison with a reference DEM showed a standard deviation of 60 m with a 42-m bias and minimum/maximum errors of about ±250 m. More recently, Toutin[145] further investigated the mapping feasibility of mixed sensor stereopairs with parametric geometric solutions ported into a fully digital stereoworkstation adapted to process on-line VIR and SAR stereopairs. From the raw images (no epipolar resampling), the data are interactively stereoextracted and then directly compared with a checked DEM. An accuracy of 20 m with no bias and minimum/maximum errors of less than ±100 m has been achieved from two different SPOT-P and ERS–SAR stereopairs, one being an opposite-side stereopair and the other a same-side stereopair. The full on-line stereocapabilities in the GCPs plotting and elevation measurements account for the good results. Comparisons of the two stereopair results showed that the elevation parallax, which contributes to the determination of the elevation, is mainly dominated by the SAR geometry with its high sensitivity to the terrain relief. Conversely, the radiometry of the SPOT-P images mainly contributes to the determination of the features with the quality of the image content.

4.5 Processing, Methods and Errors

The different processing steps to produce DEMs using stereoimages can be described in broad terms as follows:

1. to acquire the stereoimage data with supplementary information such as ephemeris and attitude data if available;
2. to collect GCPs to compute or refine the stereomodel geometry;
3. to extract the elevation parallax;
4. to compute the 3-D cartographic coordinates using 3-D stereointersection; and
5. to create and post-process the DEM (smoothing, filtering, 3-D editing, etc.).

Steps 2 and 4 involve mainly geometric issues and step 3 involves radiometric issues while steps 1 and 5 involve both geometric and radiometric issues. Since the stereomodel geometry computed from the GCPs and step 4 are related and dependent on the type of images, they are addressed in step 1.

4.5.1 Acquiring Stereoimage Data

With VIR images two types of data can be used: raw images with only normalization and calibration of the detectors (e.g. level 1A for SPOT), or geo-referenced images (e.g. level 1B for SPOT) corrected for the systematic distortions due to the sensor, the platform and the earth's rotation and curvature.

Raw 1A imagery is preferred by photogrammetrists for use in analytical or digital stereoworkstations. As mentioned previously, the geometric modeling solution employs the well-known colinearity and coplanarity equations. They have been adapted to suit the geometry of scanner imagery, but also have benefited from theoretical work in celestial mechanics to determine better the satellite's osculatory orbit and parameters.[146,147] More details on the development of the solutions and their implementation in the workstations can be found in the different referenced papers.

Since they have been systematically geo-referenced, the ''level 1B'' images just retain the elevation parallax. To compute the cartographic 3-D coordinates (step 4), the 3-D stereointersection modeling is then reduced with a simpler 2-D polynomial-based solution for the planimetry, and separately with a simple parallax equation solution for the elevation (Equation 1). This method was mainly applied in the first experiments with Landsat[59–62] since the approximation generated by the method is smaller than the final expected accuracy. However, with SPOT stereoimages (better resolution and larger B/H ratio), the approximation is no longer valid and generates poorer results than with ''raw'' stereoimages.[78,148] The solution to overcome this approximation when using 1B stereoimages is to convert the 1B images back into 1A images using the reverse transformation,[149] or to ''reshape and resize'' the 1B images to the raw imagery format.[150] This 1B geometric modeling can be mathematically combined with the normal 1A geometric modeling to avoid multiple image resampling. Although this mathematical procedure used for 1B stereoimages works better than the simple parallax approximation, it

is still recommended that raw stereoimages with the rigorous parametric solution (colinearity and coplanarity equations) be used.

The SAR images are standard products in slant or ground range presentations. They are generated digitally during postprocessing from the raw signal SAR data (Doppler frequency, time delay). The ground range presentation is the most popular product since the pixel spacing on the ground is roughly the same for the different look angle images. It then facilitates the stereoviewing and matching. The geometric modeling solution to compute the stereomodel and 3-D intersection starts generally either from the traditional Doppler and range equations,[128,131] from the equations of radargrammetry[4] or from generalized equations.[129] Their mathematical developments are different, and also depend on the method used (e.g. analytical or digital stereoworkstation, digital image or visual matching).

4.5.2 Collecting Ground Control Points

Whatever the VIR and/or SAR geometric modeling used for the stereomodel and 3-D intersection, some GCPs have to be acquired to refine the stereomodel with a least-squares adjustment process in order to obtain a cartographic standard accuracy. Since the polynomial modeling does not reflect the geometry of viewing, it requires many GCPs (20 and more) spread over the full stereopair. Each image modeling is computed separately, which does not set up a relative orientation between the images. Furthermore, the elevation is computed from an approximated solution. Consequently, this modeling cannot be used to provide the high cartographic accuracy required with the latest generation of satellites such as SPOT, IRS, MOMS, ERS and RADARSAT.

With parametric modeling such as those defined previously, few GCPs (1–6) are required. In an operational environment their number will vary as a function of their accuracy. They should preferably be spread at the border of the stereopair to avoid extrapolation in planimetry. It is also preferable to cover the full elevation range of the terrain. Different types of GCPs can be used:

- full control points with known XYZ coordinates;
- altimetric points with known Z coordinate; and
- tie points with unknown cartographic coordinates.

The two last types are useful to reinforce the stereogeometry and fill in gaps where there is no XYZ GCP. Furthermore, GCPs displayed only on one image in or outside the stereopair can also be acquired as complementary points to the ''stereo'' GCPs. Combined with tie points they can also help to avoid extrapolation in planimetry in areas where there is no ''stereo'' GCP.

The final accuracy of the stereogeometry is mainly dependent on the GCP's cartographic and image coordinates. The first can be obtained from a global positioning system (GPS), air photographic surveys, map digitizing, etc. The image coordinates are plotted interactively on the plotter or the screen. Since some workstations do not have full stereoscopic capabilities, the image coordinates are obtained simultaneously in ''double monoscopy''. This plotting will then create artificial X- and Y-parallaxes (few pixels) between the images, and the parallax errors will propagate through the stereomodel (relative and absolute orientations), the stereointersection and finally the DEM. The error propagation is much larger with a SAR stereopair than with a VIR stereopair where the plotting accuracy is about 1/3 pixel and $B/H \approx 1$. Owing to the same-side geometry with small intersection angles (8–20°) of SAR stereopairs, this error propagation due to the ''double monoscopic'' plotting increases with shallower look angles and smaller intersection angles.[132] Consequently, the DEM accuracy can decrease with a 20–40% ratio, depending on the stereopair geometry.[135] True stereoscopic plotting using human depth perception permits a better relative correspondence of the GCP between the images (SAR but also VIR) and a better absolute positioning on the ground. It is also a requisite that the two images are computed together, and not separately, to obtain a relative orientation between them.

4.5.3 Extracting Elevation Parallax

Two main methods can be used to extract the elevation parallax using image matching: the computer-assisted (visual) and automatic methods. These two methods can, of course, be integrated to take into account the strength of each one.

The computer-assisted visual matching is an extension of the traditional photogrammetric method to extract elevation data (contour lines) on a stereoplotter. It requires full stereoscopic capabilities to generate the on-line 3-D reconstruction of the stereomodel and the capture in real time of 3-D planimetric and elevation features. For elevation, the contour lines or an irregular grid DEM can be generated. The stereoscopic viewing is completed on the computer screen using a system of optics. The stereoimages are separated spatially, radiometrically or temporarily. Spatial separation is achieved using two monitors or a split screen and an optical system using mirror and/or convex lenses. Radiometric separation is achieved by anaglyphic or polarization techniques with colored or polarized lenses, respectively. Temporal separation is achieved by an alternate display of the two

images and special synchronized lenses. Petrie,[84] Dowman et al.,[89] Audet and Lapierre,[151] Heipke[90] and Walker and Petrie[91] presented the latest developments in analytical and digital stereoworkstations over the last 20 years. Furthermore, Makarovic[152] gave a comprehensive comparison between analytical and digital techniques and systems.

To retain real 3-D performance in a stereoworkstation, the images are resampled into an epipolar or quasi-epipolar geometry, in which only the X-parallax related to the elevation is retained.[69,153] Another solution to control the image positioning from the raw imagery is to follow automatically the dynamic change by cancelling the Y-parallax using the previously computed stereomodel.[87,129] In the same way as with a conventional stereoplotter, the operator cancels the X-parallax by fusing the two floating marks (one per image) on the ground. The system then measures the 2-D parallax between the images for each point and computes the XYZ cartographic coordinates using the 3-D intersection. The visual matching then combines in the brain a geometric aspect (fusing the floating marks together) and a radiometric aspect (fusing the floating marks on the corresponding image point). Some automatic tasks (displacement of the images or cursors, prediction of the corresponding image point position, etc.) are added.

However, computer-assisted visual matching, principally used with paper-format images and analytical stereoworkstations, is a long and expensive process to derive a DEM. When using digital images automated image matching can therefore be used. Since image matching has been a lively research topic for the last 20 years, an enormous body of research work and literature exists on stereomatching of different VIR and SAR sensors.

Most of the research studies on satellite image matching are based on Marr's research at the Massachusetts Institute of Technology (MIT) into the modeling of human vision.[154] If a computer program can be realized to see things as a human would, then the algorithm must have some basis in human visual processing. The stereodisparity is based on ''correct'' assumptions about the real world:[155] (i) a point of the surface has a unique position in space at any one time and (ii) matter is cohesive. The first generation of image matching based on these assumptions is gray-level image matching. Gray-level matching between the two images really implies that the radiometric intensity data from one image, representing a particular element of the real world, must be matched to intensity data from the second image, representing the same real-world element.

Although a satellite image of the real world represented by gray-levels is not like a random-dot stereogram (easily matchable), gray-level matching has been widely studied and applied to remote sensing data. Most of the matching systems operate on reference and search windows. For each position in the search window, a match-value is computed from gray-level values in the reference window. The local maximum of all the match-values computed in the search window is the good spatial position of the searched point. The match-value can be computed with the normalized cross-correlation coefficient,[59] the sum of mean normalized absolute difference,[156] the stochastic sign change or the outer minimal number estimator, etc. The first approach is considered to be the most accurate[157] and is widely used with remote sensing images. It was also noticed that matching errors were smaller with SPOT images and digitized aerial photographs than with SAR images. The last two match-value computation methods have rarely or never been used by the remote sensing community.

Another solution to the problem of matching, introduced by Förstner,[158] is the least-squares approach, minimizing the squares of the image gray-level differences in an iterative process. This method makes possible the use of well-known mathematical tools and the estimation of the error. Rosenholm[159] found that the more complicated least-squares method applied on simulated SPOT images did not give any significant improvement when compared with the cross-correlation coefficient. However, this least-squares method seems to be more accurate with real SPOT data.[7] No attempts have been made with SAR images.

The notion of least-squares matching in the object domain (ground) rather than in the image domain was later introduced by Helava.[160] Predicted image densities, corresponding to each ground element ''groundel'', are mathematically computed with known geometric and radiometric image parameters, and matched to the original one. The uncertainty in the parameters of a particular groundel is resolved by least-squares calculation. An advantage of this approach is the use of more than two images from the same or different sensors to make the least-squares solution meaningful, and a disadvantage is the ability to model correctly the groundel attributes for each image. It is mainly used with air photographs since more than two images overlap the same ground area and their geometry and radiometry are better controlled.

Since one of the constraints was either missing or incorrectly implemented in gray-level matching, Marr developed a second generation of image matching: feature-based matching.[161] The same element of the real world may look considerably different in remote sensing images acquired at different times and with different geometries between the sensor, the illumination and the terrain. Instead, edges in the images reflect the true structures.[62] Feature-based matching has not been very

popular in the remote sensing community, but successful applications have been achieved with simulated SPOT and Landsat TM.[63] The DEM results were not as good as those obtained by Simard and Slaney[60] with Landsat TM stereopairs using gray-level matching. Hähn and Förstner[162] also found that least-squares matching is more accurate than feature-based matching, conversely to Marr's theoretical prediction.

Hybrid approaches can be thus be used to achieve better and faster results by combining gray-level matching, feature-based matching with a hierarchical multiscale algorithm and also computer-assisted visual matching. The feature-based approach may produce good results for identified features, but no elevation at intermediate points. These results can then be used as seed points for the gray-level matching. Another hybrid approach is to generate gradient amplitude images as a first step with gray-level values derived from the original stereoimages instead of gradient images with only binary edge values. In a second step, any gray-level matching technique can be used on these preprocessed images.[163] The linear gradient operator can be designed to be optimal to remove noise (such as the SAR speckle) and to enhance edges. The first preliminary results with SAR stereoimages showed a 10–15% improvement in the DEM reconstruction, not always significant or consistent, but at least with fewer blunders, owing to the noise removal.

Although the computer-assisted visual matching is a long process, it has been proved to be more accurate with photographs or VIR data[105,144] and with SAR data.[134,144,156] It can therefore be used either to eliminate the blunders or to fill the mismatched areas or in areas where the automated image matching gives errors larger than one pixel (about 10% for SPOT, 15% for digitized photographs and 40–50% for SAR images[157]). It can also be used to correct the lake elevations or to generate seed points for automatic matching.

Other developments tested principally for airborne images, but rarely with satellite images, include the global approach, scale space algorithms, relational matching, consideration of breaklines, multiple image primitives, etc. Some other research studies used the recognition of corresponding structures[164] or of uniform regions,[165,166] a moment-based approach with fine-invariant features[167] or a wavelet transform approach.[168] They were only used to extract well-defined GCPs for image registration between different spaceborne VIR images.

More development could be done to integrate these solutions for generating seed points to gray-level matching. Some apparent contradictions should also be the issue of future research studies, such as:

- the theoretical prediction of Marr[154] that feature-based matching is better than gray-level matching versus better experimental results with gray-level matching than with feature-based matching;
- the theoretical automated image matching error (much better than one pixel) versus the experimental results (one and more depending on the data); and
- the ''so-called'' superiority of computer matching over the visual matching versus the experimental results.

These overall comments confirm our first statement that image matching has been a lively research topic for the past 20 years and it may be for the next 20 years!

5 OTHER METHODS

5.1 Interferometry

Radar interferometry is an alternative to the conventional stereoscopic method for extracting relative or absolute elevation information. It uses the advantages of SAR systems and of digital image processing: all-weather, night and day capabilities and automated or semiautomated processing. Imaging radar interferometry combines complex images recorded either with two antennas at different locations or with the same antenna at two different times. The phase difference information between the SAR images is used to measure changes in the range, on the subwavelength scale, for corresponding points in an image pair. By analyzing the phase variations, these distances can be translated into elevation or displacement on the ground. Since the interferometric techniques are largely detailed in another article (see **Elevation Modeling and Displacement Mapping using Radar Interferometry**) only the basic aspects are given here for understanding and comparing them with previously developed methods.

The first proposal and experiments in radar interferometry were in the field of planetary mapping in the context of radar astronomy.[169] They used an earth-based range-Doppler system to map Venus, and the interferometric information was used only to resolve the ambiguity in the mapping of southern and northern hemispheres. The first airborne application was done by Graham[170] with two antennas carried on an aircraft. ''Classified'' until 1980, the technique was later extended to SIR B data acquired from two separate passes acquired over several days.[171] After the launch of ERS 1 in 1991, numerous multipass satellite interferometric studies have been accomplished,[172,173] subsequently with Almaz-1[174] and with RADARSAT.[175] With existing satellite SAR data, only the repeat-pass system (one satellite antenna and two passes) can consistently generate interferometric data through the combination of complex images since there

is at present no satellite system with two antennas. Dual-antenna system data are only available from the 10-day Shuttle Radar Topography Mission (SRTM) realized in February 2000.

The images are registered and the phase difference is computed for each pixel. The main product is the interferogram (phase difference), and the secondary product is a coherence image, which indicates the correlation between the two SAR images. The phase difference, which is related to the two-way path difference of the radar echoes, is only known with a 2π ambiguity. It is then necessary to ''unwrap'' the phase differences for the determination of the absolute interferometric phase to within a constant.[176] The terrain elevations can thus be derived using an accurate SAR imaging model. However, there are geometric and radiometric limitations in the computation of the terrain elevation since the phase difference contains several components: topographic, atmospheric, displacement and noise.

To resolve these ambiguities and to address these different components, different developments in SAR interferometric processing took place, first with airborne SAR data in the 1980s and later with spaceborne. These include:

- the impact of the interferometric baseline on the elevation accuracy;[177]
- the evaluation of different atmospheric phenomena and the way to characterize them;[178]
- the differential interferometry combining multiple interferograms to estimate subcentimeter surface displacement;[179]
- the time interval between the image acquisitions.[180]

Current developments also include the use of satellite radar interferometry to study dynamic phenomena and their relative elevation displacement. Combining a SAR interferogram generated from two ERS 1 SAR data acquired before and after an earthquake with a DEM, the topographic component is removed from the interferogram, and the displacement field of an earthquake is mapped.[181] Topographic and displacement components can also be separated by combining three radar images to generate two interferograms.[173] The estimation of the displacement field using radar data alone, without any terrain information, is then possible. Similarly, using repeat-track interferometry with a very small cross-track baseline, which generates an interferogram with little sensitivity to topography (small topographic component), Goldstein et al.[182] measured and estimated ice sheet motion. This technique is currently applied to RADARSAT data from the Antarctica mapping mission to measure ice motions[183] and to analyze glacier flow dynamics.[184]

So far, the atmospheric component and the image coherence are the main limitations of the interferometric method for operational DEM generation. The coherence images can also be used as SAR interferometric signatures for land use classification with ERS 1 SAR repeat-pass data.[185] The interferometric correlation over forested areas was found to be significantly lower than over open canopies, small vegetation, bare soils and urban areas. The results strongly support deforestation studies, forest mapping and monitoring since it was possible not only to distinguish coniferous, deciduous and mixed forest stands, but also regrowth and clear-cut areas.

Although these results indicated that the scene coherence over forested areas was low, the interferometric technique has been used to estimate the topography and tree heights in specific conditions: a boreal forest in wintertime where the coherence varies from 0.2 to 0.5.[186] The interferometric phase information used to estimate the tree height relative to an open field and compared with in situ measurements demonstrated that the scattering center at the C-band is close to the top of the trees if the forest is dense. The good coherence obtained is also a result of the stiffness of the ''frozen'' branches on the top of the boreal forest during the wintertime. An increased sensitivity of the degree of coherence to other environmental parameters (temperature, precipitation, snowfall and soil moisture change) was also noticed.

5.2 Polarimetry

SAR polarimetry has been used with success for thematic classification studies involving natural scenes and manufactured targets. A recently developed application of the SAR polarimetry involves both a direct measure of terrain azimuthal slopes and a derived estimate of the terrain elevations.[187] The method is mainly based on empirical comparisons, supported by preliminary theoretical analysis, between the terrain local slope and the copolarized signature maximum shift. This has been validated over different geographical areas and different types of natural targets using different DEMs as reference. Although it was only tested with airborne P- and L-band SAR platforms, it is worth mentioning it, since future satellite missions (ENVISAT, RADARSAT-2) will generate dual-polarimetric or full polarimetric SAR data.

Polarimetric SAR measures the amplitude and phase terms of the complex scattering matrix. Based on a theoretical scattering model[188] for tilted, slightly rough dielectric surfaces, the azimuthal surface slope angles and signature-peak orientation displacements produced by such slopes are proportional over a range of azimuthal slopes. Empirical studies showed that an azimuthal angle

of a open-field terrain caused a proportional shift of the copolarized polarimetric signature maximum from its flat position by an angle almost equal to the terrain slope.[187]

Since forest scattering is more complex than open-field terrain scattering, radiative transfer models or discrete scatter formulations[189] of forest backscatter from a sloping terrain have to be used to modify the open-terrain algorithm. Schuler et al.[187] undertook experiments with airborne polarimetric P-band SAR data (6.6 m in range by 12.1 m in azimuth with four looks) over forested terrain with slopes up to 30° in the Black Forest, Germany. They obtained low RMSEs (2–3°) and high correlation values for the measured slopes and the derived elevation profiles when compared with an accurate DEM. Attempts to use shorter wavelength radar (C- or L-bands) yielded profiles with larger errors for forested terrain, mainly for the C-band. The larger slope estimation indicates that canopy and/or branch scattering dominates over the terrain relief scattering.

The technique was later applied with L-band SAR data over non-forested areas.[190] A simplified closed-form approximation to the relationship between the copolarized maximum shift and the measured covariance matrix elements is first established. Covariance matrices generated from experimental or modeling data can then be used as input parameters to derive the link with the terrain azimuthal slope. Azimuthal direction slopes can then be computed from the polarimetric SAR data without any prior knowledge of the terrain. By integrating the slope profiles in the azimuthal direction, the relative terrain elevation can be derived. To obtain the absolute elevation, one elevation point must be known along each slope profile.

To obtain 2-D topographic elevation and slope maps, sets of elevation profiles spaced throughout the range direction have to be available. Two orthogonal-pass SAR data provide a solution for generating an elevation surface with only one elevation point.[190] The shape-from-shading technique, which generates slopes in the across-track direction, could also be another solution.

To apply this technique with satellite SAR data (mainly C- and X-bands), future studies should be directed towards an analysis based on volume scattering to take into account the more complicated situation of the SAR backscattering with forested or agricultural areas. With such scattering models, quantitative slope and elevation values could be derived from the relationship between radiation frequency, incidence angle and type of scattering. However, the main drawbacks of this emergent technique are the volume scattering models but also the limited availability of polarimetric data to evaluate the robustness of the technique with different topographic and land cover situations.

5.3 Altimetry

There are two kinds of altimeter: the laser and the radar. The laser scanner represents an advanced method for topographic mapping. It could replace aerial photography in a short-term period, since the laser represents an emerging technology that is making its transition from the proof-of-concept, prototype stage to a readily available and reliable commercial survey instrument.[191] Automatic DEM generation over large areas is thus produced in a short delivery time. Its use is mainly restricted to airborne platforms, although a proposal was made, without success, for a satellite laser altimeter of high resolution to map the polar ice sheets.[192] The instrument would have also provided useful data on cloud-top heights and the ocean surface.

The satellite radar altimeter measures the height of a reflecting facet scanned by the passage of the instrument overhead. It uses the echo delays from within the pulse-limited footprint to estimate the minimum radar range.[193] Outside the pulse, a limited footprint for a flat surface is determined by the pulse length.[194] As an example, the radar altimeter of GEOSAT has a pulse-limited footprint about 2 km in diameter, which increases to many kilometers as large-scale surface roughness increases. Other system limitations (inaccuracies in the timing, tropospheric and ionospheric delays of the radar wave propagation and orbit determination errors) introduce errors in the range precision, which indirectly affect the spatial resolution. However, a dual-frequency altimeter should correct for the residual ionospheric uncertainty.[195] Since orbit errors are usually regarded as long-wavelength phenomena (40 000 km), least-squares adjustment using trend and bias parameters is another method to reduce errors at cross-points between ascending and descending orbit passes.[196]

The two main disadvantages of conventional radar altimeters are the large footprint and its dilation with rougher terrain. A third disadvantage is that most of the radiated power falls outside the footprint and cannot be used for height estimations. Although the radar altimeter can measure distances within 10 cm, these drawbacks have reduced its applications to flat surfaces such as the ocean surface and the ice sheets. The delay/Doppler radar altimeter[197] should overcome some of these difficulties:

1. More equivalent looks are accumulated to achieve a relatively small along-track footprint size (of the order of 250 m for a Ku-band altimeter). This minimizes the unwanted terrain dependence of the footprint size and position.
2. The entire along-track signal history contributes to height measurements rather than only the small pulse-limited area. It uses much more of the

instrument's radiated energy, and thus increases the efficiency of the system.

Measuring the distance between the satellite and the ground is only the first step of a longer procedure to convert it into an elevation relative to the geoid.[198] The geoid is the equipotential surface at mean sea level, with regard to the earth's gravitation only. The other gravitational attractions (such as those of the sun, the moon, etc.) are classified as perturbations, which give rise to the height of the sea surface relative to the geoid.

Furthermore, the geoid is in general approximated by an ellipsoid, which deviates in height from the geoid by distances of about 50 m owing to uneven mass distribution, and the satellite orbit is determined relative to the reference ellipsoid. Therefore, both measurements (height above the sea surface and satellite orbit relative to ellipsoid) have to be obtained with the same degree of accuracy. In fact, errors in the final sea-surface height (above the geoid) calculation arises from inaccuracies in the modeling of the satellite orbit.[198] To reduce this error the satellite should have as many ground-tracking stations as possible with globally dispersed coverage. Now the use of GPS provides sufficiently accurate means for determining the position of the altimeter satellite relative to the ellipsoid.

To resolve ocean-surface topography finally, the geoid shape relative to the reference ellipsoid should be known within a 10 cm accuracy over the ocean length scales being studied. This is not possible at this time. In fact it is the reverse. Geodesists used satellite altimetry to map the geoid over the ocean with an accuracy of about 1 m. Consequently, the best geoid estimate by geodesists is not dependent on the sea-surface height that oceanographers want to extract.[199] For oceanographers, the time-varying aspects of ocean dynamics can thus be addressed with repeat orbits such as (i) the ocean currents or geostrophic balance[200] (ii) ocean-circulation determination and tidal studies[201] and (iii) ocean bathymetry.[202]

With GEOSAT, the large number of repeat orbits (up to 64) can be processed with a third-degree polynomial to adjust the residuals between repeated altimeter profiles and a mean height profile. This averaging procedure improves the determination of the mean topography of the sea surface with an effective precision determined to be 1.5 cm.[196] This can then be used to characterize gravity anomalies. Furthermore, short-wavelength orbit errors are also reduced. The method can also be applied on other data sources with many collinear tracks such as ERS 13- and 35-day repeat cycle data and Topex/Poseidon. The use of multiple satellite data sets yields improved track coverage and increases the feasibility of recovering the height and consequently the geopotential field in the across-track direction.

Since the 1990s, the ERS 1 and Topex/Poseidon altimeters have offered improved data quality and global data coverage in comparison with the previous altimeters from GEOS-3, Seasat and GEOSAT. In combination with its other sensors, the ERS altimeter has been contributing to scientific development in the following areas of ocean dynamic research:[203]

- improved description of 2-D surface waves and wave heights for wave forecasting;
- ocean topography from the mesoscale to the global scale; and
- ocean-surface wind field measurement at high spatial resolution.

6 CONCLUDING REMARKS

Elevation modeling from satellite data has been a vibrant R&D topic for the last 30 years since the launch of the first civilian remote sensing satellite. Different data (space photographs, VIR scanner, SAR, altimeter) in different formats (analog, digital) can be processed by different methods (shadowing/shading, stereoscopy, interferometry, polarimetry) taking advantage of the different sensor and image characteristics (geometric, radiometric, phase) using different types of technology (analog, analytical, digital) and processing (interactive, automatic).

Most of the techniques were proposed and tested in the early years. However, the limited availability of data and associated technologies has restricted their evolution in comparison with traditional photogrammetry. Their respective evolution is also a function of the research effort in terms of physical parameter modeling and data processing.

The shadowing method, providing localized cues along special contours, is principally used to derive the relative elevation of a specific target. Despite good results, the method remains limited to specific applications. Conversely to shadowing, shading provides cues all over the studied surface, but can be applied successfully only on homogeneous surfaces. Combined with the empirical approach to resolve the different ambiguities, the method also remains marginal whatever the potential accuracy. Both methods have generated limited interest in the scientific community. On the other hand, stereoscopy is the most preferred and used method by the mapping, photogrammetry and remote sensing communities, most likely owing to the heritage of well-developed stereophotogrammetry. The latest advances in computer vision to model human vision have led to the advent of new automatic image processing approaches applied to satellite stereoscopy. Thus, the mapping

process has become more automated, but not completely with occasional unmatched expectations. Inversely, radar altimetry and polarimetry have remained more at the level of scientific interest in the physical parametric modeling without much effort made in data or image processing development.

Whether certain methods evolve depends on the trends of scientific interest. SAR stereoscopy was popular around the 1980s with the development of radargrammetry equations and the first interesting and promising results with SIR B. However, SAR image processing and related technologies to extract elevation data (such as image matching) were not mature enough and led to a temporary decline. In the same way, most of the R&D in the shape-from-shading method was done in the 1980s starting with the Venus radar mapper. At the same time, there was great interest in SPOT data with enormous research around the world in both physical parametric modeling and image processing, taking advantage also of the R&D in digital photogrammetry.

When ERS 1 was launched, scientists became enthusiastic over interferometric techniques using previously developed parametric modeling. Most research efforts were then focused in the first years on image processing (coherence image, phase unwrapping) and very few on the newly identified and poorly quantified physical artefacts (atmospheric conditions, sensor calibration). With the launch of RADARSAT in 1995, there was renewed interest in radargrammetry because researchers could take advantages of the R&D in image matching realized for SPOT at the end of the 1980s, and in the new computer technologies.

It seems obvious that the R&D of the new millennium will be focused on the use of the high-resolution satellites (VIR and SAR) and the development of their associated technologies. Already some research studies have looked at the stereoscopic potential of ASTER[204,205] and of the US high-resolution satellites[206–208] and the interferometric potential of SRTM.[209] Already same data acquired from these types of new satellites (ASTER, KOMSAT, IKONOS, SRTM, etc.) are used to demonstrate their real potential for topographic mapping.

Since any sensor, system or method has its own advantages and disadvantages, solutions to be developed in the future for operational DEM generation should exploit the complementarity between the different sensors, methods and processing. It has already been optimized in stereoscopy combining VIR and SAR data where the radiometric content of the VIR image is combined with the high sensitivity of SAR to the terrain relief and its ''all-weather'' capability to obtain the second image of the stereopair.

With the same method, some complementary aspects can be applied:

- two SAR stereoscopic pairs from ascending and descending orbit paths partially to complement the backslopes of each stereopair;
- two interferometric pairs, one with a small baseline (to help the phase unwrapping), and the other with a larger baseline (to increase the accuracy); and
- two polarimetric images from ascending and descending orbit paths to resolve the across-track ambiguity and reduce the required number of known elevation points.

The complementarity of methods has already been tried with SAR where stereoscopy is used to generate seed points needed for the clinometry or to generate an approximate DEM to help the phase unwrapping in interferometry. The loss of coherence in interferometry in forested areas can be completed with clinometry, which is well suited to the homogeneity of the forest cover. Clinometry and polarimetry could be combined since the former gives elevation information in the illumination direction and the latter in the azimuthal direction. Only one polarimetric SAR image would therefore be necessary. The shadowing method to extract building or tree heights could also be used to reduce a stereoscopic or interferometric digital surface model (DSM) in a DEM, or inversely. DSMs of cities can be used in various applications related to geographic information systems.

The complementarity can also apply at the processing level: (i) using the visual matching to seed points to the automatic matching or to postprocess and edit raw DEMs (occlusion, shadow or mismatched areas); or (ii) using stereomeasurements of geomorphological features (thalweg and crest lines, lake surfaces, etc.) to increase the mapping consistency of the DEM. Furthermore, it has been proved in most of the previous experiments that the user has to make judgements and decisions at different stages of the processing, regardless of the level of automatic processing to obtain the final DEM product: the ''know-how'' of the users could favorably complement the computer capability in different processing steps.

In the past, high-quality DEMs have been generated with traditional photogrammetry in such a way that they were used for many purposes. At present, DEMs are considered the most permanent and reusable geo-related data set over time. Although the need, requirements and specifications of DEM products are difficult to determine owing to their multiple uses by different user communities, a global DEM generation is realized in February 2000 with the US–German Space Radar Laboratory embarked on a US shuttle mission, the SRTM. It uses two-frequency single-pass interferometry with a second receiving antenna to generate DEMs over all land surfaces between -56° and $+60^\circ$ latitudes.[209,210]

The accuracy of the released DEM generated by the US C-band radar interferometry should be of the order of digital terrain elevation data (DTED) level-1 accuracy. The accuracy of the DEM generated by German X-band radar interferometry will be slightly better, but with only a partial coverage of the landmass.

Is, then, a global DEM with a unique specification not a ''wager''? Are these output statistics describing the global DEM helpful enough? Can we accept DEMs, which cannot produce ''good-looking contour lines'' in a mapping sense? Are we satisfied nowadays in producing ''throwaway DEMs'' at a particular scale that are maybe only good for a specific application? Will it fulfil the requirements of all DEM users? The other satellite data resellers hope not, since many new satellites with high-resolution VIR or SAR images with along- or across-track stereocapability are launched in the same time frame by US, Canadian, European, Indian, Russian, Japanese, etc., private or governmental organizations.

ACKNOWLEDGMENTS

The author would like to thank his CCRS friends Drs Brian Brisco, Bert Guindon, Laurence Gray, Karim Mattar and Ridha Touzi for the time they spent reviewing and improving this article.

ABBREVIATIONS AND ACRONYMS

ADRO	Applications Development and Research Opportunity
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CNES	Centre National d'Études Spatiales
CSA	Canadian Space Agency
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTED	Digital Terrain Elevation Data
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ETC	Earth Terrain Camera
FMC	Forward Motion Compensation
GCP	Ground Control Point
GPS	Global Positioning System
HRV	High Resolution in the Visible
ICP	Independent Check Point
IGN	Institut Géographique National
IRS	Indian Remote Sensing Satellite
JERS	Japan Earth Resources Satellite
LFC	Large Format Camera
LISS	Linear Imaging Self-scanned Sensor
MC	Metric Camera
MOMS	Modular Opto-electronic Multispectral Stereo Scanner
MSS	Multi Spectral Scanner
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OPS	Optical Sensor
PEPS	Preliminary Evaluation Program
R&D	Research and Development
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SIR	Spaceborne Imaging Radar
SPOT	Système pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
TM	Thematic Mapper
USGS	United States Geological Survey
VIR	Visible and Infrared
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Elevation Modeling and Displacement Mapping using Radar Interferometry • Processing and Classification of Satellite Images

REFERENCES

1. R.J. Lyon, 'Remote Sensing: Visions Beyond Sight', *Stanford Today*, Series, **1**(17), 2–7 (1966).
2. *Manual of Remote Sensing*, 3rd edition, American Society of Photogrammetry and Remote Sensing, Falls Church, VA, 1998.
3. *Manual of Photogrammetry*, 4th edition, American Society of Photogrammetry, Falls Church, VA, 1980.
4. F. Leberl, *Radargrammetric Image Processing*, Artech House, Norwood, MA, 1990.
5. J. Williams, *Geographic Information from Space*, John Wiley & Sons, Chichester, 1995.
6. L. Polidori, *Cartographie Radar*, Gordon and Breach Science Publishers, Amsterdam, 1996.
7. T. Day, J.-P.A. Muller, 'Quality Assessment of Digital Elevation Models Produced by Automatic Stereomatchers from SPOT Image Pairs', *Photogrammetric Rec.*, **12**(72), 797–808 (1988).
8. M. Lemmens, 'A Survey on Stereo-image Matching Techniques', *Int. Arch. Photogrammetry Remote Sensing*, **27**(B8), V11–V23 (1988).

9. M. Buchroithner, 'Stereoviewing from Space', *Adv. Space Res.*, **19**(1), 29–40 (1989).
10. H. Maître, F. Turpin, J.-M. Nicolas, 'Cartographie Automatique Radar: l'Apport du Traitement d'Images', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **148**, 6–14 (1997).
11. L. Polidori, Th. Toutin, 'Cartographie du Relief par Imagerie Radar: l'État de l'Art', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **152**(4), 12–23 (1998).
12. R.R. Hoffman, 'Remote Perceiving: a Step Towards a Unified Science of Remote Sensing', *Geocarto Int.*, **5**(2), 3–13 (1990).
13. R.M. Friedhoff, W. Benzon, *The Second Computer Revolution: Visualization*, Freeman, New York, 1991.
14. T. Okoshi, *Three-dimensional Imaging Techniques*, Academic Press, New York, 1976.
15. S.V. Bemis, J.L. Leeds, E.A. Winer, 'Operator Performance as a Function of Type of Display: Conventional Versus Perspective', *Hum. Factors*, **30**, 162–169 (1988).
16. B. Horn, M. Brooks, (eds.) *Shape from Shading*, MIT Press, Cambridge, MA, 1989.
17. Th. Toutin, S. Amaral, 'Stereo RADARSAT Data for Canopy Height in Brazilian Forests', *Can. J. Remote Sensing*, **26**(3), 189–199 (2000).
18. Th. Toutin, 'Depth Perception with Remote Sensing Data', in *Proceedings of the Seventeenth EARSeL Symposium on Future Trends in Remote Sensing, Lingby, Denmark, 17–19 June 1997*, 401–409, 1998.
19. Th. Toutin, C. Vester, 'Radar and Stereoscopy Tutorial', CCRS Web, http://www.ccrs.nrcan.gc.ca/cers/eduref/sradar/srintrue.html.
20. T. Rindfleisch, 'Photometric Method for Lunar Topography', *Photogrammetric Eng.*, **32**(2), 262–277 (1966).
21. P. Teillet, B. Guindon, D. Goodenough, 'On the Slope/Aspect Correction of Multispectral Scanner Data', *Can. J. Remote Sensing*, **8**(2), 84–106 (1982).
22. F.T. Ulaby, M.C. Dobson, *Handbook of Radar Backscattering Statistics for Terrain*, Artech House, Norwood, MA, 1988.
23. G. Domik, F. Leberl, J. Cimino, 'Dependence of Image Grey Values on Topography in SIR-B Images', *Int. J. Remote Sensing*, **19**(5), 1013–1022 (1988).
24. J.A. Smith, T.L. Lin, K.H. Rawson, 'The Lambertian Assumption and Landsat Data', *Photogrammetric Eng. Remote Sensing*, **46**(9), 1183–1189 (1980).
25. G. La Prade, E. Leonardo, 'Elevations from Radar Imagery', *Photogrammetric Eng.*, **35**(4), 366–371 (1969).
26. F. Cheng, K.-H. Thiel, 'Delimiting the Building Heights in a City from the Shadow in a Panchromatic Image. Part 1: Test of 42 Buildings', *Int. J. Remote Sensing*, **16**(3), 409–415 (1995).
27. P.A. Brivio, A. Della Ventura, A. Rampini, R. Schettini, 'Automatic Selection of Control Points from Shadow Structures', *Int. J. Remote Sensing*, **13**(10), 1853–1860 (1992).
28. Th. Toutin, D. Hoja, E. Höppner, A. Rémond, C. King, 'GCPs Selection from Multi Source Data Over Mountainous Topography', in *Proceedings of IGARSS'98, Seattle, July 6–10*, 2339–2341, Vol. 5, 1998.
29. A.K. Saraf, J.D. Das, B. Agarwal, R.M. Sundaram, 'False Topography Perception Phenomena and its Correction', *Int. J. Remote Sensing*, **17**(18), 3725–3733 (1996).
30. A. Huertas, R. Nevatia, 'Detecting Building in Aerial Images', *Comput. Vision Graphics Image Process.*, **41**, 131–152 (1988).
31. Ph. Hartl, F. Cheng, 'Delimiting the Building Heights in a City from the Shadow on a Panchromatic SPOT Image: Part 2: Test of a Complete City', *Int. J. Remote Sensing*, **16**(15), 2829–2842 (1995).
32. J. Meng, M. Davenport, 'Building Assessment with Subpixel Accuracy Using Satellite Imagery', *Proceedings SPIE*, **2656**, 65–76 (1996).
33. V.K. Shettigara, G.M. Sumerling, 'Height Determination of Extended Objects Using Shadows in SPOT Images', *Photogrammetric Eng. Remote Sensing*, **64**(1), 35–44 (1998).
34. G.H. Lodwick, S.H. Paine, 'A Digital Elevation Model of the Barnes Ice-cap Derived from Landsat-MSS Data', *Photogrammetric Eng. Remote Sensing*, **51**(12), 1937–1944 (1985).
35. B. Horn, 'Obtaining Shape from Shading Information', *The Psychology of Computer Vision*, McGraw-Hill, New York, Chapter 4, 115–155, 1975.
36. R.L. Wildey, 'Topography from Single Radar Images', *Science*, **224**, 153–156 (1984).
37. R.L. Wildey, 'Radarclinometry for the Venus Radar Mapper', *Photogrammetric Eng. Remote Sensing*, **52**(1), 41–50 (1986).
38. T.R. Frankot, R. Chellapa, 'A Method for Enforcing Integrability in Shape from Shading', *IEEE Trans. Pattern Anal. Machine Intell.*, **10**(4), 439–451 (1988).
39. T.R. Frankot, R. Chellapa, 'Estimation of Surface Topography from SAR Imagery Using Shape from Shading Techniques', *Artif. Intell.*, **43**, 271–310 (1990).
40. J. Thomas, W. Kober, F. Leberl, 'Multiple-image SAR Shape from Shading', in *Proceedings of IGARSS'89, Vancouver, July 10–14*, 592–596, 1989.
41. B. Guindon, 'Development of a Shape-from-shading Technique for the Extraction of Topographic Models from Individual Spaceborne SAR Images', *IEEE Trans. Geosci. Remote Sensing*, **28**(4), 654–661 (1990).
42. J.B. Boisvert, T. Pultz, R.J. Brown, B. Brisco, 'Potential of Synthetic Aperture Radar for Large Scale Soil Moisture Monitoring: a Review', *Can. J. Remote Sensing*, **22**(1), 2–13 (1995).
43. S. Paquerault, H. Maître, 'La Radarclinométrie', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **148**, 20–29 (1997).
44. W. Keidel, 'Application and Experimental Verification of an Empirical Backscattering Cross-section Model

for the Earth's Surface', *IEEE Trans. Geosci. Remote Sensing*, **20**(1), 67–71 (1982).
45. L. Polidori, 'Digital Terrain Models from Radar Images: a Review', in *Proceedings of the International Symposium on Radars and Lidars in Earth and Planetary Sciences, Cannes, September 2–4*, ESA SP-328, 141–146, 1991.
46. J. Thomas, W. Kober, 'Radarclinometry – Shape from Shading: Generalized *N*-Image Algorithm', in *Radargrammetric Image Processing*, ed. F. Leberl, Artech House, Norwood, MA, sections 15.4–15.7, 435–551 (1990).
47. G. La Prade, S.J. Briggs, R.J. Farrell, E.S. Leonardo, 'Stereoscopy', in *Manual of Photogrammetry*, 4th edition, American Society of Photogrammetry, Falls Church, VA, Chapter X, 519–544, 1980.
48. U.V. Helava, 'On System Concepts for Digital Automation', *Int. Arch. Photogrammetry Remote Sensing*, **27**(B2), 171–190 (1988).
49. A. Grün, 'Digital Photogrammetric Stations: a Short List of Unmatched Expectations', *Geo Info Mag.*, **11**(1), 20–23 (1997).
50. F. Doyle, 'A Large Format Camera for Shuttle', *Photogrammetric Eng. Remote Sensing*, **45**(1), 73–78 (1979).
51. R.K. Mott, 'The Use of Satellite Imagery for Very Small Scale Mapping', *Photogrammetric Rec.*, **8**(46), 458–475 (1975).
52. R. Kostka, 'Anwendungsbezogene Untersuchungen an Weltraum Photos von Gebirgsregionen', *Mitt. Forstl. Bundes-Versuchsanst. Wien*, **157**, 139–146 (1986).
53. A.A.C. Meneguette, 'Evaluation of Metric Camera Photography for Mapping and Coordinates Determination', *Photogrammetric Rec.*, **11**(66), 699–709 (1985).
54. K. Jacobsen, W. Müller, 'Evaluation of Space Photographs', *Int. J. Remote Sensing*, **9**(10–11), 1715–1721 (1988).
55. G. Ducher, 'Metric Camera Data: Assessment and Potential Use at l'Institut Géographique National', *Photogrammetric Rec.*, **11**(66), 671–689 (1985).
56. S. Murai, 'Cartographic Accuracy of Stereospace Photographs Taken by Large Format Camera: a Case Study in Japan', *Proceedings of ISPRS Symposium Mapping from Modern Imagery, Edinburgh, September 8–12*, **26**(B4), 732–737 (1986).
57. I.F.G. Whittington, 'An Assessment of the Capability of Stereo Shuttle Photographs as a Source of Feature Content for Mapping', *Int. J. Remote Sensing*, **10**(6), 965–987 (1989).
58. R. Welch, C.P. Lo, 'Height Measurements from Satellite Images', *Photogrammetric Eng. Remote Sensing*, **43**(10), 1233–1241 (1977).
59. R. Simard, 'Digital Stereo-enhancement of Landsat-MSS Data', in *Proceedings of the Seventeenth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, May 9–13*, 1275–1281, 1983.
60. R. Simard, R. Slaney, 'Digital Terrain Model and Image Integration for Geologic Interpretation', in *Proceedings of the Fifth Thematic Conference on Remote Sensing for Exploration Geology, Reno, Nevada, September 29–October 2*, 49–60, 1986.
61. M. Ehlers, R. Welch, 'Stereocorrelation of Landsat-TM Images', *Photogrammetric Eng. Remote Sensing*, **53**(9), 1231–1237 (1987).
62. P.R. Cooper, D.E. Friedman, S.A. Wood, 'The Automatic Generation of Digital Terrain Models from Satellite Images by Stereo', *Acta Astronaut.*, **15**(3), 171–180 (1987).
63. D.L. Light, D. Brown, A. Colvocoresses, F. Doyle, M. Davies, A. Ellasal, J. Junkins, J. Manent, A. McKenney, R. Undrejka, G. Wood, 'Satellite Photogrammetry', in *Manual of Photogrammetry*, 4th edition, American Society of Photogrammetry, Falls Church, VA, Chapter XVII, 883–977, 1980.
64. H. Guichard, 'Étude Théorique de la Précision dans l'Exploitation Cartographique d'un Satellite à Défilement: Application à SPOT', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **90**, 15–26 (1983).
65. G. de Masson d'Autumne, 'Corrélation Numérique d'Images Quelconques Selon les Lignes Quasi-épipolaires par Approximations Successives', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **95**, 23–32 (1984).
66. Th. Toutin, 'Analyse Mathématique des Possibilités Cartographiques du Système SPOT', Thèse de Docteur-Ingénieur, École Nationale des Sciences Géodésiques, St-Mandé, 1985.
67. C. Vigneron, P. Denis, 'Stéréorestitution d'Images SPOT à l'Aide du Logiciel Traster IGN/Matra', in *Proceedings of the Eighteenth International Symposium on Remote Sensing of Environment, Paris, October 1–5*, 455–464, 1984.
68. K.V. Vincent, P.K. Pleitner, D.H. Coupland, H. Schultz, E.R.B. Oshel, 'New Digital Elevation Mapping Software Applied to SPOT Simulation Data', in *Proceedings of the 1984 SPOT Symposium, Scottsdale, AZ, May 20–23*, 92–97, 1984.
69. G. de Masson d'Autumne, 'Le Traitement Géométrique des Images de Télédétection', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **73–74**, 5–16 (1979).
70. V.I. Khizhnichenko, 'Coordinates Transformation when Geometrically Correcting Earth Space Scanner Images', *Earth Explor. Space*, **5**, 96–103 (1982) (in Russian).
71. Th. Toutin, 'Analyse Mathématique des Possibilités Cartographique du Satellite SPOT', *Mémoire de DEA*, École Nationale des Sciences Géodésiques, St-Mandé, 1983.
72. G. Konecny, E. Kruck, P. Lohmann, 'Ein universeller Ansatz für die geometrische Auswertung von CCD-Zeilenabtasteraufnahmen', *Bildmessung Luftbildwesen*, **54**(4), 139–146 (1986).
73. D.J. Gugan, 'Practical Aspects of Topographic Mapping from SPOT Imagery', *Photogrammetric Rec.*, **12**(69), 349–355 (1987).

74. W. Kratky, 'Rigorous Stereo Photogrammetric Treatment of SPOT Images', in *Proceedings of the SPOT-1 Image Utilization, Assessment, Results, Paris, November*, 1281–1288, 1987.
75. F.C. Paderes, E.M. Mikhail, J.A. Fagerman, 'Batch and On-line Evaluation of Stereo SPOT Imagery', in *Proceedings of the ASPRS–ACSM Convention, Baltimore, MD*, 31–40, Vol. 3, 1989.
76. T. Westin, 'Precision Rectification of SPOT Imagery', *Photogrammetric Eng. Remote Sensing*, **56**(2), 247–253 (1990).
77. CNES, 'SPOT-1: Utilisation des Images, Bilan, Résultats', *Proceedings of the SPOT-1 Symposium, Paris*, Centre National d'Études Spatiales, 1987.
78. D.J. Gugan, I.J. Dowman, 'Accuracy and Completeness of Topographic Mapping from SPOT Imagery', *Photogrammetric Rec.*, **12**(72), 787–796 (1988).
79. R. Priebbenow, E. Clerici, 'Cartographic Applications of SPOT Images', in *Proceedings of the SPOT-1 Image Utilization, Assessment, Results, Paris, November*, 1189–1194, 1987.
80. G. Konecny, P. Lohmann, H. Engel, E. Kruck, 'Evaluation of SPOT Imagery on Analytical Photogrammetric Instruments', *Photogrammetric Eng. Remote Sensing*, **53**(9), 1223–1230 (1987).
81. J.C. Trinder, B.E. Donnelly, K.L. Keong, 'SPOT Mapping Software for Wild Aviolyt BC2 Analytical Plotter', in *Proceedings of IGARSS'88, Edinburgh, 12–16, September*, 1801–1806, 1988.
82. W. Kratky, 'On-line Aspects of Stereophotogrammetric Processing of SPOT Images', *Photogrammetric Eng. Remote Sensing*, **55**(3), 311–316 (1989).
83. G. Petrie, 'Developments in Analytical Instrumentation', *ISPRS J. Photogrammetry Remote Sensing*, **45**, 61–89 (1990).
84. G. Petrie, 'Trends in Analytical Instrumentation', *ITC J.*, 364–383 (1992).
85. Ph. Hottier, M. Albattah, 'Transformation d'un Couple de Scènes SPOT en un Couple d'Images Épipolaires Analogiquement ou Analytiquement Restituable sans Déformation du Relief', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **123**, 13–27 (1991).
86. R. Welch, 'Desktop Mapping with Personal Computers', *Photogrammetric Eng. Remote Sensing*, **55**(11), 1651–1662 (1989).
87. Th. Toutin, C. Nolette, Y. Charbonneau, P.-A. Gagnon, 'Stéreo-restitution Interactive des Données SPOT: Description d'un Nouveau Système', *J. Can. Télédétection*, **29**(2), 146–151 (1993).
88. Th. Toutin, M. Beaudoin, 'Real-time Extraction of Planimetric and Altimetric Features from Digital Stereo SPOT Data Using a Digital Video Plotter', *Photogrammetric Eng. Remote Sensing*, **61**(1), 63–68 (1995).
89. I.J. Dowman, H. Ebner, C. Heipke, 'Overview of European Developments in Digital Photogrammetric Workstations', *Photogrammetric Eng. Remote Sensing*, **58**(1), 51–56 (1992).
90. C. Heipke, 'State-of-the-art Digital Photogrammetric Workstations for Topographic Applications', *Photogrammetric Eng. Remote Sensing*, **61**(1), 49–56 (1995).
91. A.S. Walker, G. Petrie, 'Digital Photogrammetric Workstation', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B2), 384–395 (1996).
92. D.J. Gugan, I.J. Dowman, 'Design and Implementation of a Digital Photogrammetric System', *Int. Arch. Photogrammetry Remote Sensing*, **26**(B2), 100–109 (1986).
93. P. Denis, 'Génération de Modèle Numérique de Terrain à Partir de Données SPOT', *Int. Arch. Photogrammetry Remote Sensing*, **26**(B3/2), 176–185 (1986).
94. H. Guichard, G. Ruckebusch, E. Sueur, S. Wormser, 'Stéréorestitution Numérique SPOT: une Approche Originale de Formation des Images Épipolaires et de Mise en Correspondance', in *Proceedings of the SPOT-1 Image Utilization, Assessment, Results, Paris, November*, 1371–1392, 1987.
95. L. Renouard, 'Création Automatique de MNT à Partir de Couples d'Images SPOT', in *Proceedings of the SPOT-1 Image Utilization, Assessment, Results, Paris, November*, 1347–1355, 1987.
96. R. Simard, Th. Toutin, A. Leclerc, S.R. Haja, M. Allam, R. Boudreau, R. Slaney, 'Digital Terrain Modelling with SPOT Data and Geological Applications', in *Proceedings of the SPOT-1 Image Utilization, Assessment, Results, Paris, November*, 1205–1212, 1987.
97. T. Malleswara Rao, K. Venugopala Rao, A. Ravi Kumar, R.P. Rao, B.L. Deekshatula, 'Digital Terrain Model from Indian Remote Sensing Satellite Data from the Overlap of Two Adjacent Paths Using Digital Photogrammetric Techniques', *Photogrammetric Eng. Remote Sensing*, **62**(6), 726–731 (1996).
98. K. Jacobsen, 'Calibration of IRS-1C Pan Camera', in *ISPRS Workshop on Sensors and Mapping from Space, Hannover, September 29 to October 2*, 163–170, 1997.
99. Th. Toutin, Ph. Cheng, K. Seidel, 'Indian Remote Sensing Satellite: Geocoding and DEM Extraction', in *Proceedings of the 20th Canadian Symposium on Remote Sensing, Calgary, Alberta, May 10–13*, 29–32, 1998.
100. Ph. Cheng, Th. Toutin, 'Unlocking the Potential for IRS-1C Data', *Earth Obs. Mag.*, **7**(3), 24–26 (1998).
101. Ph. Cheng, Th. Toutin, C. Stohr, 'Automated DEM Extraction from Air Photos or Optical Satellite Images', in *Proceedings of the 13th International Conference Applied Geologic Remote Sensing, Vancouver, BC*, March 1–3, Vol. 1, 56–63, 1999.
102. J. Raggam, A. Almer, 'Assessment of the Potential of JERS-1 for Relief Mapping Using Optical and SAR Data', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B4), 671–676 (1996).
103. T. Westin, 'Photogrammetric Potential of JERS-1 OPS', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B4), 937–942 (1996).

104. H. Maruyama, R. Kojiroi, T. Ohtsuka, Y. Shimoyama, S. Hara, H. Masaharu, 'Three Dimensional Measurement by JERS-1 OPS Stereo Data', *Int. Arch. Photogrammetry Remote Sensing*, **30**(B4), 210–215 (1994).
105. E. Dorrer, W. Maier, V. Uffenkamp, 'Stereocompilation of MOMS-02 Scenes on the Analytical Plotter', in *Proceedings of the MOMS-02 Symposium, Köln, 5–7 July*, 95–110, 1995.
106. F. Ackerman, D. Fritsch, M. Hahn, F. Schneider, V. Tsingas, 'Automatic Generation of Digital Terrain Models with MOMS-02/D2 Data', in *Proceedings of the MOMS-02 Symposium, Köln, Germany, 5–7 July*, 79–86, 1995.
107. H. Raggam, K. Gutjahr, A. Almer, 'MOMS-2P und RADARSAT: Neue Sensoren zur stereometrischen Geländemodellerstellung', *Vermessung Geoinf.*, **4/97**, 267–280 (1997).
108. W. Kornus, M. Lehner, H. Ebner, H. Froba, T. Ohlhof, 'Photogrammetric Point Determination and DEM Generation Using MOMS-2P/PRIRODA Three-line Imagery', *Int. Arch. Photogrammetry Remote Sensing*, **32**(B4), 321–328 (1998).
109. G. La Prade, 'An Analytical and Experimental Study of Stereo for Radar', *Photogrammetric Eng.*, **29**(2), 294–300 (1963).
110. G.E. Carlson, 'An Improved Single Flight Technique for Radar Stereo', *IEEE Trans. Geosci. Electron.*, **GE-11**, 199–204 (1973).
111. G.H. Rosenfield, 'Stereoradar Techniques', *Photogrammetric Eng.*, **34**, 586–594 (1968).
112. G. Gracie, J.W. Bricker, R.K. Brewer, R.A. Johnson, 'Stereoradar Analysis', Report FTR -1339-1, US Engineering Topography Laboratory, Fort Belvoir, VA, 1970.
113. G. La Prade, 'Subjective Considerations for Stereo Radar', in *Proceedings of the Thirty-sixth Annual Meeting of the American Society of Photogrammetry, Washington, DC, March 1–6*, 640–651, 1970.
114. F. Leberl, 'Accuracy Analysis of Stereo Side Looking Radar', *Photogrammetric Eng. Remote Sensing*, **45**(8), 1083–1096 (1979).
115. V. Kaupp, L. Bridges, M. Pisaruk, H. MacDonald, W. Waite, 'Simulation of Spaceborne Stereoradar Imagery: Experimental Results', *IEEE Trans. Geosci. Remote Sensing*, **21**(2), 400–405 (1983).
116. G. Domik, 'Evaluation of Radar Stereoviewability by Means of Simulation Techniques', in *Proceedings of IGARSS'84, Paris*, ESA-SP-215, 623–646, 1983.
117. F. Leberl, 'Imaging Radar Applications to Mapping and Charting', *Photogrammetria*, **32**(3), 75–100 (1976).
118. K. Yoritomo, 'Methods and Instruments for the Restitution of Radar Pictures', Invited Paper at the Twelfth Congress of the International Society for Photogrammetry, Ottawa, July 24–August 4, 1972.
119. J.K. Fullerton, F. Leberl, R.E. Marque, 'Opposite Side SAR Image Processing for Stereoviewing', *Photogrammetric Eng. Remote Sensing*, **52**(9), 1487–1498 (1986).
120. R. Simard, F. Plourde, Th. Toutin, 'Digital Elevation Modelling with Stereo SIR-B Image Data', in *Proceedings of the International Symposium on Remote Sensing for Resources Development and Environmental Management, ISPRS Commission VII, Enschede, The Netherlands*, 161–166, 1986.
121. Th. Toutin, 'Opposite-side ERS-1 SAR Stereo Mapping Over Rolling Topography', *IEEE Trans. Geosci. Remote Sensing*, **34**(2), 543–549 (1996).
122. M. Kobrick, F. Leberl, J. Raggam, 'Radar Stereo Mapping with Crossing Flight Lines', *Can. J. Remote Sensing*, **12**(9), 132–148 (1986).
123. F. Leberl, G. Domik, J. Raggam, M. Kobrick, 'Radar Stereo-mapping Techniques and Applications to SIR-B Images of Mount Shasta', *IEEE Trans. Geosci. Remote Sensing*, **24**(4), 473–481 (1986).
124. J. Raggam, F. Leberl, 'SMART – a Program for Radar Stereomapping on the Kern DSR-1', in *Proceedings of the ASP/ACSM Annual Convention, Washington, DC, March 11–16*, American Society of Photogrammetry, Falls Church, VA, 765–773, 1984.
125. F. Leberl, G. Domik, J. Raggam, J. Cimino, M. Kobrick, 'Multiple Incidence Angle SIR-B Experiment Over Argentina: Stereoradargrammetric Analysis', *IEEE Trans. Geosci. Remote Sensing*, **24**(4), 482–491 (1986).
126. I.V. Yelizavetin, 'Digital Terrain Modeling from Radar Image Stereopairs', *Mapping Sci. Remote Sensing*, **30**(2), 151–160 (1993).
127. J. Raggam, A. Almer, W. Hummelbrunner, D. Strobl, 'Investigation of the Stereoscopic Potential of ERS-1 SAR Data', in *Proceedings of the Fourth International Workshop on Image Rectification of Spaceborne Synthetic Aperture Radar, Loipesdorf, 26–28 May*, 81–87, 1993.
128. Z.-G. Twu, I. Dowman, 'Automatic Height Extraction from ERS-1 SAR Imagery', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B2), 380–383 (1996).
129. Th. Toutin, 'Generating DEM from Stereoimages with a Photogrammetric Approach: Example with VIR and SAR Data', *EARSeL J. Adv. Remote Sensing*, **4**(2), 110–117 (1995).
130. L. Marinelli, Th. Toutin, I. Downan, 'Génération de MNT par Radargrammétrie: état de l'Art et Perspectives', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **148**, 88–96 (1997).
131. S. Sylvander, D. Cousson, P. Gigord, 'Étude des Performances Géométriques des Images RADARSAT', *Bull. Soc. Fr. Photogrammétrie Télédétection*, **148**, 57–65 (1997).
132. Th. Toutin, 'Evaluation de la Précision Géométrique des Images de RADARSAT', *Can. J. Remote Sensing*, **24**(1), 80–88 (1998).
133. S. Parashar, E. Langham, J. McNally, S. Ahmed, 'RADARSAT Mission Requirements and Concepts', *Can. J. Remote Sensing*, **19**(4), 280–288 (1993).

134. Canadian Space Agency, 'Bringing Radar Application Down to Earth', *Proceedings of the RADARSAT ADRO Symposium, Montreal, October 13–15*, CD-ROM, Canadian Space Agency, 1998.
135. Th. Toutin, 'Error Tracking of Radargrammetric DEM from RADARSAT Images', *IEEE Trans. Geosci. Remote Sensing*, **37**(5), 2227–2238 (1999).
136. Th. Toutin, 'Evaluation of Radargrammetric DEM from RADARSAT Images in High Relief Areas', *IEEE Trans. Geosci. Remote Sensing*, **38**(2), 782–789 (2000).
137. I.J. Dowman, Z.-G. Twu, P.H. Chen, 'DEM Generation from Stereoscopic SAR Data', in *Proceedings of ISPRS Joint Workshop on Sensors and Mapping from Space, Hannover, September 29–October 2*, 113–122, 1997.
138. R. Welch, R.T. Jordan, J.C. Luvall, 'Geocoding and Stereodisplay of Tropical Forest Multi-sensor Datasets', *Photogrammetric Eng. Remote Sensing*, **56**(10), 1389–1392 (1990).
139. J. Raggam, A. Almer, 'A Multi-sensor Stereomapping Experiment', in *Proceedings of the ACSM/ASPRS/Auto Carto Annual Convention Baltimore, MD*, 173–182, Vol. 4, 1991.
140. Th. Toutin, 'SPOT and Landsat Stereo Fusion for Data Extraction Over Mountainous Areas', *Photogrammetric Eng. Remote Sensing*, **64**(2), 109–113 (1998).
141. K. Akeno, 'DEM Generation from Multi-sensor Stereopairs AVHRR and MSS', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B4), 36–40 (1996).
142. R.K. Moore, 'Heights from Simultaneous Radar and Infrared', *Photogrammetric Eng.*, **5**(7), 649–651 (1969).
143. A.L. Bloom, E.J. Fielding, X.-Y. Fu, 'A Demonstration of Stereophotogrammetry with Combined SIR-B and Landsat-TM Images', *Int. J. Remote Sensing*, **9**(5), 1023–1038 (1988).
144. J. Raggam, A. Almer, D. Strobl, 'A Combination of SAR and Optical Line Scanner Imager for Stereoscopic Extraction of 3-D Data', *ISPRS J. Photogrammetry Remote Sensing*, **49**(4), 11–21 (1994).
145. Th. Toutin, 'Stereomapping with SPOT-P and ERS-1 SAR Images', *Int. J. Remote Sensing*, **21**(8), 1657–1674 (2000).
146. P.R. Escobal, *Methods of Orbit Determination*, Krieger, Malabar, FL, 1965.
147. CNES, 'Le Mouvement du Véhicule Spatial en Orbite', *Proceedings of Symposium, Toulouse*, Centre National d'Études Spatiales, 1980.
148. N. Al-Roussan, G. Petrie, 'System Calibration, Geometric Accuracy Testing and Validation of DEM and Ortho-image Data Extracted from SPOT Stereopairs Using Commercially Available Image Processing Systems', *Int. Arch. Photogrammetry Remote Sensing*, **32**(4), 8–15 (1998).
149. N. Al-Roussan, P. Cheng, G. Petrie, Th. Toutin, M.J. Valadan Zoej, 'Automated DEM Extraction and Ortho-image Generation from SPOT Level-1B Imagery', *Photogrammetric Eng. Remote Sensing*, **63**(8), 965–974 (1997).
150. M.J. Valadan Zoej, G. Petrie, 'Mathematical Modelling and Accuracy Testing of SPOT Level-1B Stereopairs', *Photogrammetric Rec.*, **16**(91), 67–82 (1998).
151. H. Audet, A. Lapierre, 'Revue des Systèmes et Techniques de Visualisation Stéréoscopique et de leur Application en Géomatique', in *Comptes-Rendus du Seizième Symposium Canadien sur la Télédétection, Sherbrooke, Québec, 7–10 juin*, 449–454, 1993.
152. B. Makarovic, 'Analytical Versus Digital Photogrammetric Techniques and Systems', *Int. Arch. Photogrammetry Remote Sensing*, London, UK, February 13–14, **28**(2), 304–321 (1990).
153. H.H. Baker, T.O. Binford, 'Depth from Edge and Intensity Based Stereo', in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence, Vancouver, BC*, 631–636, 1981.
154. D. Marr, *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman, San Francisco, 1982.
155. D. Marr, T. Poggio, 'A Computation of Stereodisparity', *Proceedings R. Soc. London*, **B194**, 283–287 (1977).
156. H. Ramapriyan, J. Strong, Y. Hung, C. Murray, 'Automated Matching Pairs of SIR-B Images for Elevation Mapping', *IEEE Trans. Geosci. Remote Sensing*, **24**(4), 462–472 (1986).
157. F. Leberl, K. Maurice, J.K. Thomas, M. Millot, 'Automated Radar Image Matching Experiment', *ISPRS J. Photogrammetry Remote Sensing*, **49**(3), 19–33 (1994).
158. W. Förstner, 'On the Geometric Precision of Digital Correlation', *Int. Arch. Photogrammetry*, **24**(B3), 176–189 (1982).
159. D. Rosenholm, 'Numerical Accuracy of Automatic Parallax Measurement of Simulated SPOT Images', *Can. J. Remote Sensing*, **12**(2), 103–113 (1986).
160. U.V. Helava, 'Object Space Least-squares Correlation', *Proceedings of the ACSM–ASPRS Annual Convention, St. Louis, Missouri*, 46–55, Vol. 3, 1988.
161. D. Marr, E. Hildreth, 'A Theory of Edge Detection', *Proceedings R. Soc. London*, **B207**, 187–217 (1980).
162. M. Hähn, W. Förstner, 'The Applicability of Feature-based and a Least-squares Matching Algorithm for DEM Acquisition', *Int. Arch. Photogrammetry Remote Sensing*, Kyoto, Japan, July 1–10, **27**(B9), III137–III150 (1988).
163. Ph. Paillou, M. Gelautz, 'Relief Reconstruction from SAR Stereo Pairs: The ''Optimal Gradient'' Matching Method', *IEEE Trans. Geosci. Remote Sensing*, **37**(4), 2099–2107 (1999).
164. A. Della Ventura, A. Rampini, R. Schettini, 'Image Registration by Recognition of Corresponding Structures', *IEEE Trans. Geosci. Remote Sensing*, **28**(3), 305–314 (1990).
165. J. Petit-Frère, 'Prise en Compte des Différences Photométriques Entre Images dans les Techniques de

Stéréorestitution', *Int. Arch. Photogrammetry Remote Sensing*, **29**(B4), 392–398 (1992).
166. M. Abbasi-Dezfouli, T.G. Freeman, 'Stereoimages Registration Based on Uniform Patches', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B2), 101–106 (1996).
167. J. Flusser, T. Suk, 'A Moment-based Approach to Registration of Images with Affine Geometric Distortions', *IEEE Trans. Geosci. Remote Sensing*, **32**(2), 382–387 (1994).
168. J.-P Djamdji, A. Bijaoui, 'Disparity Analysis: A Wavelet Transform Approach', *IEEE Trans. Geosci. Remote Sensing*, **33**(1), 67–76 (1995).
169. A.E.E. Rodgers, R.P. Ingalls, 'Venus Mapping: the Surface Reflectivity by Radar Interferometry', *Science*, **165**, 797–799 (1969).
170. L.C. Graham, 'Synthetic Interferometer Radar for Topographic Mapping', *Proceedings IEEE*, **62**(6), 763–768 (1974).
171. A.K. Gabriel, R.M. Goldstein, 'Crossed-orbit Interferometry: Theory and Experimental Results from SIR-B', *Int. J. Remote Sensing*, **9**(5), 857–872 (1988).
172. D. Massonnet, T. Rabaute, 'Radar Inteferometry: Limits and Potential', *IEEE Trans. Geosci. Remote Sensing*, **31**(2), 455–464 (1993).
173. H.A. Zebker, C. Werner, P.A. Rosen, S. Hensley, 'Accuracy of Topographic Maps Derived from ERS-1 Interferometric Radar', *IEEE Trans. Geosci. Remote Sensing*, **32**(4), 823–836 (1994).
174. I.V. Yelizavetin, Ye.A. Ksenofontov, 'Precision Terrain Measurement by SAR Interferometry', *Mapping Sci. Remote Sensing*, **33**(1), 1–19 (1996).
175. D. Geudtner, P. Vachon, K.E. Mattar, A.L. Gray, 'RADARSAT Repeat-pass SAR Interferometry: Results Over an Arctic Test Site', in *Proceedings of the GER'97 Symposium: Geomatics in the Era of RADARSAT, Ottawa, May 25–30*, CD-ROM, 1997.
176. R.M. Goldstein, H. Zebker, C. Werner, 'Satellite Radar Interferometry: Two-dimensional Phase Unwrapping', *Radio Sci.*, **23**(4), 713–720 (1988).
177. F. Li, R.M. Goldstein, 'Studies of Multi-baseline Spaceborne Interferometric Synthetic Aperture Radars', *IEEE Trans. Geosci. Remote Sensing*, **28**(1), 88–97 (1990).
178. D. Massonnet, K. Feigl, 'Discrimination of Geophysical Phenomena in Satellite Interferograms', *Geophys. Res. Lett.*, **22**, 1537–1540 (1995).
179. A.K. Gabriel, R. Goldstein, H. Zebker, 'Mapping Small Elevation Changes Over Large Areas: Differential Radar Interferometry', *J. Geophys. Res.*, **94**(B7), 9183–9191 (1989).
180. P. Vachon, D. Geudtner, L. Gray, R. Touzi, 'ERS-1 Synthetic Aperture Radar Repeat-pass Interferometry Studies: Implications for RADARSAT', *Can. J. Remote Sensing*, **21**(4), 441–454 (1995).
181. D. Massonnet, M. Rossi, C. Carmona, F. Adragana, G. Peltzer, K. Feigl, T. Rabaute, 'The Displacement Field of the Landers Earthquake Mapped by Radar Interferometry', *Nature (London)*, **364**, 138–142 (1993).
182. R.M. Goldstein, H. Engelhardt, B. Kamb, R.M. Frolich, 'Satellite Radar Interferometry for Monitoring Ice Sheet Motion: Application to an Antartic Ice Stream', *Science*, **262**, 1525–1530 (1993).
183. A.L. Gray, K.E. Mattar, P.W. Vachon, R. Bindschadler, K.C. Jezek, R. Forster, 'InSAR Results from the RADARSAT Antarctic Mapping Mission Data: Estimation of Glacier Motion Using a Simple Registration Procedure', in *Proceedings of IGARSS'98, Seattle, July 6–10*, 1638–1640, Vol. III, 1998.
184. R.R. Forster, K.C. Jezek, H. Gyoo Sohn, A.L. Gray, K.E. Mattar, 'Analysis of Glacier Flow Dynamics from Preliminary RADARSAT InSAR Data of the Antarctic Mapping Mission', in *Proceedings of IGARSS'98, Seattle, July 6–10*, 2225–2227, 1998.
185. U. Wegmüller, Ch. Werner, 'SAR Interferometric Signature of Forest', *IEEE Trans. Geosci. Remote Sensing*, **33**(2), 1153–1161 (1995).
186. J.O. Hagberg, L.M.H. Ulander, J. Askne, 'Repeat-pass Interferometry Over Forested Terrain', *IEEE Trans. Geosci. Remote Sensing*, **33**(2), 331–340 (1995).
187. D.L. Schuler, J.S. Lee, G. De Grandi, 'Measurement of Topography Using Polarimetric SAR Images', *IEEE Trans. Geosci. Remote Sensing*, **34**(5), 1266–1277 (1996).
188. G.R. Valenzuela, 'Scattering of Electromagnetic Waves From a Tilted Slightly-rough Surface', *Radio Sci.*, **11**(3), 1057–1066 (1968).
189. S.L. Durden, J.J. van Zyl, H.A. Zebker, 'Modelling and Observation of Radar Polarization Signature of Forested Areas', *IEEE Trans. Geosci. Remote Sensing*, **27**(3), 290–301 (1989).
190. G.F. De Grandi, D.L. Schuler, T.L. Ainsworth, J.S. Lee, 'Topography Sensing by Polarimetric SAR: Theoretical Basis and Application Using Orthogonal-pass AIRSAR Data', in *Proceedings of IGARSS'97, Singapore, August 11–15*, 1371–1375, 1997.
191. M. Flood, B. Gutelius, 'Commercial Implications of Topographic Terrain Mapping using Scanning Airborne Laser Radar', *Photogrammetric Eng. Remote Sensing*, **63**(4), 327–329, 363–366 (1997).
192. J.L. Bufton, J.E. Robinson, M.D. Femiano, F.S. Flatow, 'Satellite Laser Altimeter for Measurement of Ice Sheet Topography', *IEEE Trans. Geosci. Remote Sensing*, **GE-20**(4), 544–549 (1982).
193. R.K. Moore, C.S. Williams, 'Radar Terrain Return at Near-vertical Incidence', *Proceedings IRE*, **45**, 228–238 (1997).
194. G.S. Brown, 'The Average Impulse Response of a Rough Surface and its Application', *IEEE Trans. Antennas Propag.*, **AP-25**(1), 1169–1170 (1998).
195. J. Goldhirsh, J.R. Rowland, 'A Tutorial Assessment of Atmospheric Height Uncertainties for High-precision Satellite Altimeter Missions to Monitor Ocean Currents',

IEEE Trans. Geosci. Remote Sensing, **GE-20**(4), 418–433 (1982).
196. J.P. Noréus, 'Improved Resolution of Geosat Altimetry Using Dense Sampling and Polynomial Adjusted Averaging', *Int. J. Remote Sensing*, **16**(15), 2843–2862 (1995).
197. R.K. Raney, 'The Delay/Doppler Radar Altimeter', *IEEE Trans. Geosci. Remote Sensing*, **36**(5), 1578–1588 (1998).
198. I.S. Robinson, *Satellite Oceanography: An Introduction for Oceanographers and Remote Sensing Scientists*, Ellis Horwood, Chichester, 1985.
199. R.G. Brown, W.D. Kahn, D.C. McAdoo, W.E. Himwich, 'Roughness of the Marine Geoid From Seasat Altimetry', *J. Geophys. Res.*, **88C**, 1531–1540 (1983).
200. S. Pond, G.L. Pickard, *Introductory Dynamical Oceanography*, 2nd edition, Pergamon Press, Oxford, 1983.
201. D.E. Cartwright, 'Detection of Large-scale Ocean Circulation and Tides', *Philos. Trans. R. Soc. London*, **A309**, 361–370 (1983).
202. J.G. Marsh, T.V. Martin, 'The Seasat Altimeter Mean Surface Model', *J. Geophys. Res.*, **87C**, 3269–3280 (1982).
203. ESA, 'New Views of the Earth: Scientific Achievements of ERS-1', ESA SP1176/I Report, European Space Agency, Paris, 1995.
204. M. Tokunaga, S. Hara, 'Overview of DEM Product Generated by Using ASTER Data', *Int. Arch. Photogrammetry Remote Sensing*, **31**(B4), 874–878 (1996).
205. R. Welch, T. Jordan, H. Lang, H. Murakani, 'ASTER as a Source for Topographic Data in the Late 1990's', *IEEE Trans. Geosci. Remote Sensing*, **36**(4), 1282–1289 (1998).
206. H.M. Ridley, P.M. Atkinson, P. Aplin, J.-P. Muller, I. Dowman, 'Evaluation of the Potential of the Forthcoming US High Resolution Satellite Sensor Imagery at the Ordnance Survey', *Photogrammetric Eng. Remote Sensing*, **63**(8), 997–1005 (1997).
207. V. Kaufmann, W. Sulzer, 'Über die Nutzungsmöglichkeit hochauflösender amerikanischer Spionage-Satellitenbilder (1960–1972)', *Vermessung Geoinf.*, **3/97**, 166–173 (1997).
208. R. Li, 'Potential of High Resolution Satellite Imagery for National Mapping Products', *Photogrammetric Eng. Remote Sensing*, **64**(12), 1165–1170 (1998).
209. M. Werner, 'Shuttle Radar Topography Mission', in *Proceedings of ISPRS Workshop: Sensors and Mapping from Space, Hannover, September 29–October 2*, 9–11, 1997.
210. R.L. Jordan, B.L. Huneycutt, M. Werner, 'The SIR-C/X-SAR Synthetic Aperture Radar', *IEEE Trans. Geosci. Remote Sensing*, **33**(4), 829–839 (1995).

Global Land Databases for Environmental Analyses

Thomas R. Loveland
US Geological Survey EROS Data Center, Sioux Falls, USA

1 Introduction 1
2 Global Land Databases 1
2.1 Land Cover Legends 4
2.2 Other Global Land Data 4
3 Examples of Global Land Data Applications 5
3.1 Climate and Meteorology 6
3.2 Atmospheric Chemistry 6
3.3 Ecological Applications 6
3.4 Hydrological Modeling 7
3.5 Resource Management 7
4 Future Global Land Data Directions 8
Abbreviations and Acronyms 9
Related Articles 9
References 9

Global land databases present the area, spatial extent, location, and patterns of land characteristics. Land cover, the natural and artificial materials on the land surface, is the most widely available and used land data theme, but soils and elevation global databases are also found. These land data sets are key inputs to many global and regional environmental models and provide a means to parameterize key landscape processes used in studies of climate and meteorology, atmospheric chemistry, ecological dynamics, hydrology, and resource management.

1 INTRODUCTION

Global land databases present the area, spatial extent, location, and patterns of important land characteristics. While virtually any landscape theme (i.e. soils, ecosystems, terrain) can be represented in spatial databases, the most widely available and commonly used land databases in the environmental sciences involve land cover. Land cover can be defined as the natural and artificial materials on the land surface at a given point in time (e.g. cropland, forest land, wetlands, water, snow and ice, and built-up structures).

Many environmental applications require data that provide descriptions of the types of land cover or vegetation and their characteristics (i.e. canopy density and height, leaf area, albedo). Land cover affects available energy and energy partitioning between sensible and latent heat fluxes through changes in albedo, roughness, and stomatal control of water losses.[1] The specific land cover types represented in a database correlate with the biological, thermodynamic, or chemical pathways corresponding to different vegetation associations. The types and definitions of land cover found in different global databases are variable and depend on the requirements of specific applications.

Global land databases can be either tabular or spatial. Tabular databases provide statistics of land cover area. Spatial databases are maps that provide information on the location, patterns, and extent of land cover. Spatial databases are especially important inputs to grid-based environmental models. Each grid cell has a size (e.g. 1 km^2, 1° latitude by 1° longitude), and an earth location that is commonly expressed in latitude and longitude or map projection coordinates.

2 GLOBAL LAND DATABASES

There is significant variability in the characteristics of global land cover databases (Table 1). In the 1980s, several coarse-resolution databases (1° latitude by 1° longitude) were produced for use in general circulation models (GCMs)[2,3] and global carbon studies.[4]

The Olson and Watts database, generally referred to as the Olson Global Ecosystems map, originally consisted of 49 categories.[3] These categories represent natural, anthropogenic or natural/modified vegetation mosaics of land cover and are optimized for use in global carbon cycle research. The Olson database has a spatial resolution of 0.5° latitude by 0.5° longitude. As such, it was the highest resolution global database of this era. It was compiled using hundreds of published maps. The accuracy of this database is unknown, as no systematic accuracy assessment has been conducted. The Olson database has undergone continual updating and improvement. The current Olson Global Ecosystems legend now consists of 94 land cover classes. However, Olson did not produce global map based on the more detailed legend.

The Matthews global database consists of two data layers.[2] The primary layer consists of a 32-category map of potential natural vegetation and has a grid resolution of 1° latitude by 1° longitude. A secondary layer includes

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Characteristics of global land cover databases

Characteristic	Matthews (1983)[2]	Olson and Watts (1982)[3]	Wilson and Henderson-Sellers (1985)[4]	DeFries and Townshend (1994)[5]	Loveland et al. (1999)[6]
Grid cell resolution	1° latitude by 1° longitude	0.5° latitude by 0.5° longitude	1° latitude by 1° longitude	1° latitude by 1° longitude	1 km by 1 km
Thematic detail and attributes	32 classes; percentage cropland	49 classes	53 classes	14 classes	7 land cover maps with between 6 and 75 classes
Source materials	Existing maps	Existing maps	Existing maps	AVHRR satellite imagery	AVHRR satellite imagery
Dates of sources	Pre-1983	Pre-1983	Pre-1985	1987	1992–93
Mapping approach	Compilation	Compilation	Compilation	Supervised classification	Unsupervised classification
Intended application	Climate modeling	Carbon cycle studies	Climate modeling	Climate modeling	Multipurpose
Validation approach	None	None	None	Training field	Unbiased sampling

AVHRR, advanced very high-resolution radiometer.

estimates for each grid cell of the percentage of cultivated land. The two layers can be combined to produce a map that approximates actual land cover. The Matthews database was developed for use in a GCM. Like the Olson map, the Matthews map was compiled from an exhaustive review of hundreds of source materials, including maps, texts, atlases, and statistical summaries. A strength of the Matthews product is that the source materials used were meticulously documented. The accuracy of this database has not been established.

A third global database is the Wilson and Henderson-Sellers global land cover database.[4] Their database has a spatial resolution of 1° latitude by 1° longitude and contains 53 categories of land cover. Each grid cell includes a primary and secondary land cover attribute. The land cover categories include physiognomic elements, including canopy density, and have inferences to land use activities. Like the Matthews and Olson databases, the Wilson and Henderson-Sellers product was compiled from previously published materials and was designed for use in GCMs. As with the other two coarse-resolution land cover databases, the accuracy of the Wilson and Henderson-Sellers product is unknown.

Although the quality of the previously described data sets is unknown, studies suggest that improved spatial resolution and map accuracy are needed.[5,7] For example, Leemans et al.[7] made a comparison of the natural vegetation categories of the three databases (representing 75% of the global land surface). Kappa statistics were computed and used to rank similarities and differences between complex spatial patterns. The kappa results indicate only poor to fair agreement between all maps. This is also indicated by the low correspondence between all spatial patterns. Leemans et al. found that up to 60% of the cells were classified differently.[7] This difference was most notable between Olson and Watts[3] and Matthews.[2] Based on the comparison, the authors concluded that these were only an approximate representation of actual land cover. Their conclusion was further strengthened by the observation that even in the tabular databases developed from country census, large differences occur between the data.

Defries and Townshend[5] compared the Olson, Wilson and Henderson-Sellers, and Matthews classifications to determine their relative consistency. They generalized the categories from the three data sets into 15 categories, and then compared the spatial extent of each category in each data set. Their conclusions reinforced those made by Leemans et al.[7] They found that even when the global area of a given land cover class was similar in all three data sets, the area of spatial agreement could be much smaller because the distributions were not geographically coincident. Overall, the land area where the data sets agree, for the 15 classes, was only 26% of the total surface area.

The Olson and Watts, Matthews, and Wilson and Henderson-Sellers databases were pioneering efforts. Although imperfect, they provided the first attempts to produce science-quality land databases. The differences between these global compilations of land cover exist because (1) they rely on compilation from numerous separate sources, since direct ground observations using an agreed system of classification has never been directly attempted at a global scale, and (2) widely varying criteria are used in classification.[8]

In the mid-1990s, newer databases of higher resolution started to be developed from satellite remotely sensed data.[6,9] The new data sets are based on remotely sensed data from the AVHRR on-board US National

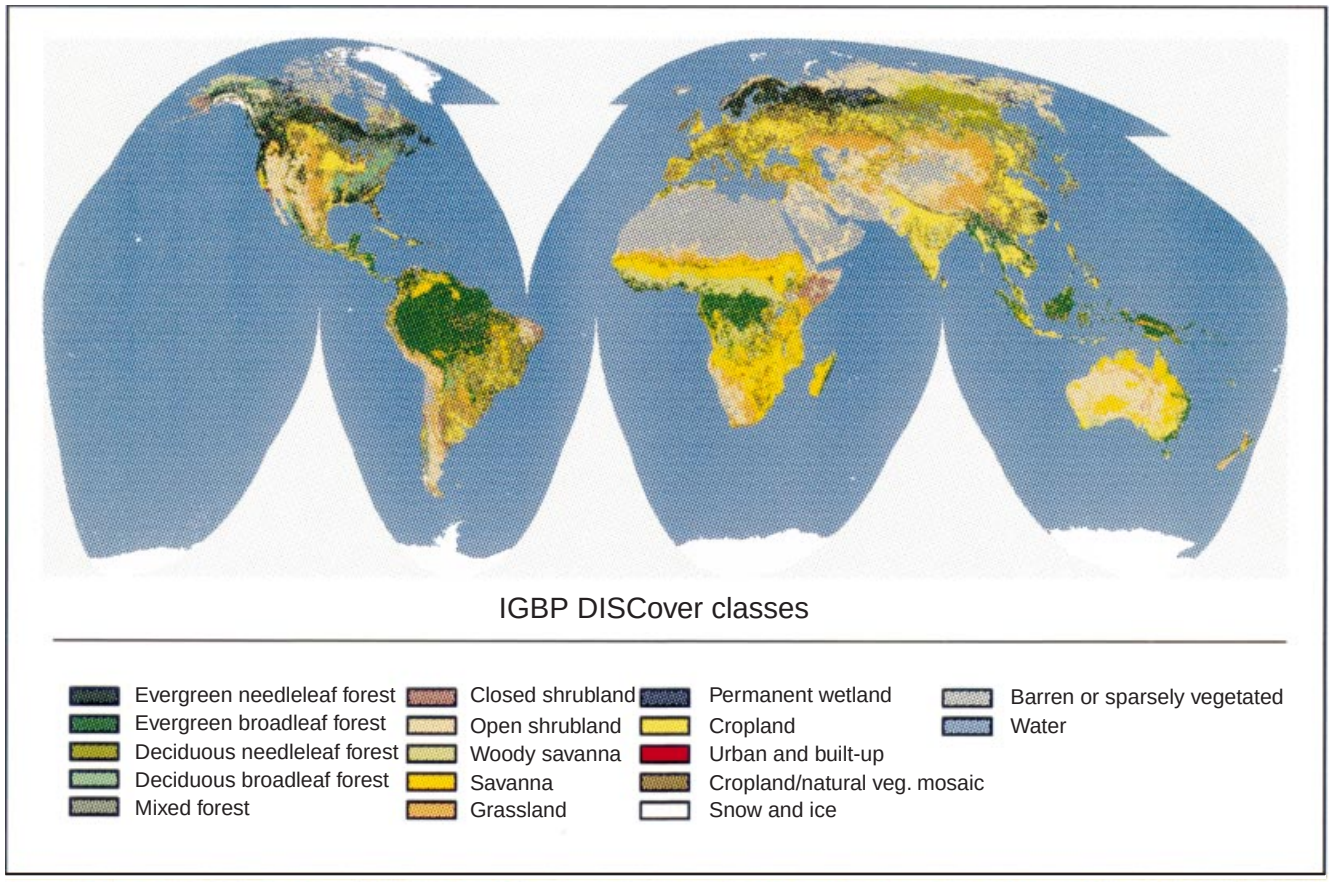

Figure 1 Global land cover based on the IGBP DISCover land cover legend. (Reproduced from Loveland et al.[9] © US Government.)

Oceanic and Atmospheric Administration (NOAA) polar-orbiting meteorological satellites.

Defries et al.[10] assembled a coarse-resolution land cover product from AVHRR data. This data set has 1° grid cells but was developed from a 1987 AVHRR data set averaged to 1° resolution. It was developed using supervised classification of a global 1° latitude by 1° longitude AVHRR normalized difference vegetation index data set prepared by Los et al.[11] This data set was designed specifically for use in land atmosphere interactions modeling, and specifically the Simple Biosphere Model (SiB), and has 14 categories corresponding to SiB model requirements.

In 1997, the US Geological Survey (USGS) completed a 1-km^2 spatial resolution global land cover characteristics database interpreted from 1992–93 AVHRR data.[6] The USGS database differs in that it is not based on a single land cover legend, but a database of land cover characteristics that can be interpreted on a case-by-case basis to create the data sets needed for specific applications. This database provides land cover based on seven different land cover legends commonly used for global-scale environmental research: (1) Olson Global Ecosystems for carbon cycle studies; (2) SiB for land–atmosphere interaction modeling; (3) Simple Biosphere Model Two (SiB2), and update of the original land–atmosphere interaction scheme; (4) Biosphere Atmosphere Transfer scheme, another land–atmosphere interaction model legend; (5) Running Global Land Cover, for biogeochemical modeling; (6) USGS Land Use and Land Cover, for general applications; and (7) International Geosphere Biosphere Programme (IGBP) Data and Information System Land Cover (DISCover) legend, for general environmental modeling. Figure 1 presents global land cover based on the DISCover legend. The DISCover data set is the first global land data set in which a formal accuracy assessment was completed to establish data quality. An independent accuracy assessment conducted by IGBP-affiliated researchers determined that the overall accuracy is 73.5%.[12]

In addition to spatial databases, statistical summaries of specific land cover types are compiled by the United Nations Food and Agriculture Organization (UNFAO). UNFAO agriculture and forestry statistical databases are

based on inventory data provided by member states, or estimated using statistical sampling.[13,14] In addition, the World Resources Institute (WRI) publishes annual statistical abstracts that include a variety of land cover themes.[15] The WRI statistics are compiled from data collected by other organizations.

2.1 Land Cover Legends

It often appears that there are as many land cover legends as there are applications of land cover data. Because of the variety of land cover legends used in global studies, this discussion will be limited to the most common.

Although there are several general-purpose land cover legends, there are also many special-purpose schemes developed for specific applications. The SiB, SiB2, and Biosphere–Atmosphere Transfer Scheme (BATS) land–atmosphere interaction models are examples of single-purpose land cover legend. The classes defined for these systems represent the unique parameterization requirements of the affiliate model.

The land cover legends developed and refined by Olson and Watts (Olson Global Ecosystems) is noteworthy because it was designed specifically for global applications.[3] Although not specifically developed for use with remotely sensed data, Olson has refined class definitions to permit straightforward use with satellite imagery. An added strength of Olson is that he has developed attributes describing land use, seasonality, and climate for each category. As previously stated, this legend now has 94 classes. Olson continues to add new classes and refine and improve definitions and attributes as new applications for the scheme arise.

The Matthews legend, consisting of 35 categories, is important because of its longevity. Like Olson, the Matthews database has been used in many applications. As a result, it is pervasive and accepted by many modelers. Matthews is based on the well-known United Nations Educational, Scientific, and Cultural Organization (UNESCO) vegetation classification system, and stresses potential rather than actual vegetation types.[16]

The land cover classification scheme developed by Anderson et al., although not designed for global-scale studies, is important because it is one of the few designed specifically for use with remotely sensed data.[17] The Anderson system is based on several important considerations. First, it is hierarchical with subsequent levels of the hierarchy defined to be mapped by increasingly larger scales of remotely sensed data. Second, Anderson classes were developed to yield consistent accuracies of at least 85% at all levels of the hierarchy. Third, the land cover classes were defined to serve as surrogates for land use. The principles from which the legend was developed have been widely accepted. As a result, it is not uncommon to see elements of the Anderson system in other legends.

The Running et al. global land cover system is an example of a legend designed both for use with remotely sensed data and for use at the global scale.[18] The Running strategy is based on definitions of three canopy components: vegetation structure (termed above-ground biomass by Running), leaf longevity, and leaf type. Vegetation structure defines whether the vegetation retains perennial or annual above-ground biomass, an issue for seasonal climate and carbon-balance modeling. It is also a determinant of the surface roughness length parameter that climate models require for energy and momentum transfer equations. Leaf longevity (evergreen versus deciduous canopy) is a critical variable in carbon cycle dynamics of vegetation, and affects seasonal albedo and energy transfer characteristics of the land surface. Leaf longevity indicates whether a plant annually must completely regrow its canopy, or a portion of it, with inferred consequences to carbon partitioning, leaf litterfall dynamics, and soil carbon. Leaf type (needleleaf, broadleaf, and grass) affects gas exchange characteristics.

The IGBP developed a land cover legend for use in global change research.[19] The legend consists of 17 general cover types selected based on the requirements of the IGBP Core Projects. It was designed specifically for use with 1-km remotely sensed data from the AVHRR. The categories embrace the philosophy presented by Running et al. but with modifications to (1) be compatible with classifications systems currently used for environmental modeling (e.g. SiB, and BATS), (2) provide, where possible, land-use implications, and (3) represent landscape mixtures and mosaics.[18] The IGBP legend retains key elements of Running, including climate-independent class definitions and reliance on ancillary remotely sensed measures, such as vegetation greenness indices, as relative indicators of temporal dynamics of biophysical properties.

2.2 Other Global Land Data

Global environmental studies also require data on other themes including spatial databases on soil types and characteristics and elevation. Digital elevation model (DEM) data provide estimates of the average elevation within grid cells. DEM data can be used to derive important terrain characteristics such as slope, aspect, hydrologic basins, and stream networks. The USGS, for example, distributes a 1-km^2 resolution global DEM.[20] A 1° latitude by 1° longitude resolution global soils database has been produced from the UNFAO Soils Map of the World.[21] This data set has 1° latitude/longitude

Table 2 World Wide Web sources for global land data sets

Data set	Internet URL
Matthews Global Land Cover	www.giss.nasa.gov
Olson and Watts Global Ecosystems	www.ngdc.noaa.gov
Wilson and Henderson-Sellers Land Cover	www.ngdc.noaa.gov
Defries (University of Maryland) Land Cover	glcf.umiacs.umd.edu/glcf_data.html
USGS Global Land Cover Database	edcwww.cr.usgs.gov/landdaac/glcc/glcc.html
USGS DEM	edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html
FAO/Zobler Global Soils Database	edcwww.cr.usgs.gov/glis/hyper/guide/world_soil
UNFAO Cropland and Forestry Statistics	apps.fao.org
WRI Global Statistics	www.wri.org

resolution, and includes attributes on soil types, soil texture, soil slope, and soil proprieties.

Table 2 provides World Wide Web access information for all of the data sets discussed.

3 EXAMPLES OF GLOBAL LAND DATA APPLICATIONS

Global land data sets are used in a wide range of applications, including climate modeling, biogeochemical cycle analyses, ecosystem dynamics, carbon cycle studies, hydrological processes, and resource management. The specific data requirements vary according to the geographic extent, spatial resolution, and thematic attribute requirements of each study. It is important to note that the applications of land cover data vary both within and between disciplines and applications (Table 3). The following sections provide brief summaries of selected applications of global land cover data.

Global environmental applications have two basic components. First, numerical models accounting for interactions and feedbacks between different earth systems are needed. Second, a large amount of geographically referenced, validated data will be needed to parameterize these models adequately. Land cover data are used to partition the landscape into geographic units corresponding to a broad suite of environmental parameters, such as surface roughness, albedo, latent and sensible heat flux, and associated biogeochemical processes and cycles. Changes in the distribution of land cover alter the regional and possibly global balance of these fluxes. Changes in land cover distribution include both the often-cited conversions of one land cover type to another (e.g. forest to cropland, grassland to urban), and also the seasonal changes in land cover parameters. The seasonal development of foliage,

Table 3 Land cover characteristics input requirements and spatial scale for selected modeling applications and models

Type	Model	Classification scheme	Spatial scale	Associated attributes
GCMs	NASA/GSFC	SiB	$4 \times 5^{\circ}$	SiB, NDVI
	University of Maryland–COLA	Simplified SiB	$4.5 \times 7.8^{\circ}$, $1.8 \times 2.8^{\circ}$	SiB, NDVI
	NCAR–CCM	BATS	$2 \times 4^{\circ}$	BATS, NDVI
Mesoscale meteorological models	CSU–RAMS	LEAF	Nested grids of 1, 10, 40 km	LEAF, NDVI
	PSU/NCAR–MM4	BATS	Nested grids of 4, 12, 36 km	BATS set and NDVI derivatives
Hydrological models	Watershed precipitation/runoff	Basic classes	2.5, 5, 10 km	Model specific
	Agricultural chemicals runoff	Anderson level II	Country level or 1 km	Model specific
Ecosystem models	RHESSys	Basic biomes	1–50 km	RHESSys, NDVI
	CENTURY	Anderson level II	1–50 km	NDVI derivatives
	Biogenic emissions	Key species (oak, hickory, etc.)	20 km	NDVI derivatives

COLA, Center for Ocean–Land–Atmosphere; CSU–RAMS, Colorado State University, Regional Atmospheric Modeling System; GSFC, Goddard Space Flight Center; LEAF, Land–Ecosystem–Atmosphere Feedback; NCAR–CCM, National Center for Atmospheric Research, Climate Community Model; PSU/NCAR-MM4, Penn State University, National Center for Atmospheric Research, Mesoscale Meteorological Model 4; RHESSys, Regional Hydrological Ecosystem Simulation System.

for example, over a period of less than a few weeks can affect local weather.[22]

3.1 Climate and Meteorology

Land cover data are increasingly important in GCMs used to simulate global climate and in mesoscale meteorological models that provide regional weather forecasts. Several land surface parameterization models have been developed for the purpose of estimating suites of parameters that affect the interaction between land cover and the atmosphere.[22] Land cover databases provide boundary conditions from which surface fluxes of energy, moisture, and momentum are estimated. The SiB model by Sellers et al.[23] and the BATS model by Dickinson et al.[24] were among the earliest land surface parameterization models to use land cover data to realistically represent landscape processes in atmospheric models.

There is considerable experimental evidence that modeled atmospheres are sensitive to landscape characteristics, and that changes in water availability at the earth's surface, or changes in albedo, produce large changes in the numerically simulated climates.[25] Numerical experiments using global circulation models have demonstrated that landscape characteristics affect not only the regional atmosphere, but also modify global climate. Land cover parameters important in affecting mesoscale circulation include leaf area index as a function of height, stomatal resistance, albedo, aerodynamic roughness, displacement height, percentage coverage, and photometric properties. Pielke and Avissar conclude that global data sets need higher spatial resolution than the 1980s 1° cells so that GCMs can provide information on local circulation.[25]

The impact of land cover on climate is illustrated in the Copeland et al. study comparing the United States climate generated using both current and presettlement land cover conditions.[26] In this simulation study, they found that differences in vegetation parameters, resulting from the change of land cover from natural to modified conditions, affected most climate variables. For example, they found that changes in roughness length, leaf area index, and albedo changed significantly when natural vegetation was replaced by anthropogenic land uses. Overall, Copeland et al. found that the mean daily temperature increased by 0.05 K because of land cover change.[26] Locally, temperature changes were in excess of 1–2 K warmer or cooler, though warming dominated. The temperature changes were the result of a change in the surface energy balance. Albedo changes led to changes in net radiation available at the surface, but changes in the partitioning of that between latent and sensible heat flux due to leaf area index and fractional coverage changes determined whether a region heated or cooled. Precipitation rates generally increased throughout the USA, by about 5% (0.09 mm day^{-1}). Local changes in precipitation rates were typically 5–10% above historical land cover scenarios. Changes tended to occur along the ecotones, or boundaries between vegetation types. The overall result of this modeling suggests that current land cover has caused summertime surface conditions to be warmer and wetter than natural landscapes would indicate.

Modeling exercises using global land databases continue to probe the sensitivity of climate system to surface processes. These efforts can only be successful if global-scale data sets are available for initialization and validation.

3.2 Atmospheric Chemistry

Investigations of atmospheric chemistry use land cover to estimate the amount and mixing of biogenic emissions in the atmosphere. Emission inventories of biogenic volatile organic compounds (BVOC), such as isoprenes and monoterpenes from vegetation, are important inputs to models that simulate photochemical smog.[27] The impacts of BVOC can be important in understanding ambient air quality, since BVOC emissions from vegetation can be greater than those from anthropogenic sources. BVOC emission potentials for different land cover types vary by more than three orders of magnitude.[28]

The chemistry of the atmosphere is strongly influenced by ecological processes that control the emission of water and trace gases from plants. Vegetation-based emissions of BVOC could have a significant impact on tropospheric chemistry by influencing the processes that control the formation of atmospheric haze. Isoprene and monoterpenes are typically regarded as the predominant BVOC emitted by plants. The land cover elements that control natural BVOC emissions are (1) landscape average emission potential, (2) foliar density, and (3) emission fluxes that are based on interaction with light and temperature conditions.[28] The accuracy of regional BVOC emission models is limited by a lack of emission rate measurements and appropriate land cover databases. Most regional emission models classify landscapes into three or four woodland categories and several nonwoodland landscapes. A significant improvement should be expected if landscapes are classified according to species composition.

3.3 Ecological Applications

Biogeochemical models address a variety of ecological processes and have contributed to significant advances in the understanding of the global cycles responsible for the compositions of the atmosphere, oceans, and sediments on the surface of the earth.[29] Land cover attributes describing community composition or vegetation types

are used in biogeochemical models (e.g. CENTURY by Parton et al.[30] and BIOME-BGC by Hunt et al.[31]) to parameterize assimilation rates, carbon allocation, and nutrient use efficiency, which influence CO_2 uptake during photosynthesis.[22]

There are numerous studies illustrating the use of global land databases for ecological assessments. For example, Burke et al.[32] highlight the importance of land cover in simulating the temporal dynamics of soil organic matter and plant production in grazed grasslands and agroecosystems using the CENTURY model. Land cover, along with other important driving variables including surface-soil physical properties, monthly precipitation and temperature, and plant nitrogen and lignin contents were inputs to an assessment of the impacts of a CO_2 doubling scenario on soil carbon. Burke et al.[32] found that 35–50% of soil organic matter in the Great Plains was lost as a result of 50 years of cultivation in the region. The simulation suggested that all croplands had significant losses of soil organic matter but that rangelands did not. Intermediate losses occurred in irrigated corn because of the relatively high tillage intensity, and lowest simulated losses occurred in continuous wheat and dryland corn. Across all cultivated areas, an average of 44% of soil organic matter in the top 20 cm was lost; across the entire region (i.e. all land use types), 21.5% of surface carbon was estimated to be lost over 50 years.

One of the most modeled ecological parameters, global net primary productivity (NPP), is complicated because the models used differ markedly in approach, complexity, and land cover inputs.[33] NPP models use a wide range of spatial data, including vegetation species, climate, soils, plant characteristics, disturbance regimes, and a number of other natural and anthropogenic factors. Past global databases lacked the land cover attributes needed for NPP modeling. Estimation of NPP is complicated by the fact that the plant characteristics that affect this parameter vary both within and among species. Improvements in global data are a critical foundation for accurate NPP estimates for a wide array of global change scenarios.

Global data sets have also been used to assess large-area patterns of environmental diversity.[34] Land cover data can be used to approximate large-area biodiversity patterns. Land cover serves as a surrogate for habitat and can thus be used to identify potential distributions of flora and fauna. In a study of US land cover, Wickham et al.[34] observed that the general pattern of increasing diversity from east to west is in contrast with latitudinal gradients in species diversity at continental global scales. This north–south terrestrial species gradient appears to result from a similar gradient in available energy.

A general conclusion regarding global land cover data for ecological applications is that the ecological community stressed the need to develop land data sets of the highest possible resolution.[35] For ecological assessments, grid cell sizes may sometimes need to be aggregated, but more frequently fine resolution is essential for ecological models.

3.4 Hydrological Modeling

Modeling in hydrology appears to have lagged behind atmospheric and ecological processes modeling.[36] Hydrological modeling requires land cover data for modeling the physical processes of the hydrological cycle. However, land cover is typically treated as uniform through a watershed. Vegetation interacts with the atmosphere by its controls on the return of water back to the atmosphere, as evaporation and transpiration. Therefore, land cover has a significant impact on the hydrological cycle and consequently on other components of the climatic system. Land cover influences the exchange of energy, water, carbon, and other substances at the land–surface–atmosphere interface. Its form affects the absorption of solar radiation. Plant leaves and branches increase the surface area for evaporation, intercept part of precipitation, and by increasing friction for air movement increase the thickness of the boundary layer. The IGBP biological aspects of the hydrologic cycle (BAHC) core project aims to serve as a catalyst for the improvement of future hydrological models. BAHC plans stress the need explicitly to assess and represent the relevant topographic, vegetation, soil, and geological parameters that control soil moisture and surface and subsurface hydrological conditions in order to improve the results of hydrological simulations.[37]

3.5 Resource Management

There are many regional applications of large-area land cover databases in activities associated with resource management. For example, US Department of Agriculture policy makers used land cover data to assess drought and flood conditions for crop land areas.[38] Land cover data provided a means to focus flood and drought analysis on key agricultural areas in the midwestern USA.

In a unique application, the US Air Force uses coarse-resolution land cover to model the potential for bird collisions with low-flying aircraft.[39] The Air Force uses a bird avoidance model to provide localized data on bird distributions and abundance throughout the continental USA.

4 FUTURE GLOBAL LAND DATA DIRECTIONS

Perhaps the clearest direction in global land data set characteristics is the movement toward higher resolution (i.e. 1 km). Although there are currently no global environmental modeling applications using 1-km data, or for that matter resolutions higher than 50 km^2, there are a number of significant reasons why a 1-km global land cover database is important for global environmental studies. These include:

- The 1-km global land cover database will have improved historical relevance since it can serve as a time-specific baseline of regional to global land cover patterns and areal extents.
- A wider range of applications can be met with a higher resolution land cover data set. As illustrated in Table 3, the grid cell dimensions vary both within and between types of environmental modeling applications. A 1-km global land cover database has the spatial resolution that meets the minimum spatial resolution requirements of all current large-area environmental models.
- A variety of fundamental research opportunities are possible, particularly those dealing with scaling and landscape heterogeneity issues.
- Regional modeling can be conducted at higher resolutions. For example, mesoscale meteorological models are now being run at resolutions from 1- to 40-km grid cells.
- Flexible sub-grid cell parameterization requires incorporation or analysis of landscape heterogeneity. Higher resolution data provides a means either to apply flexible aggregation strategies to represent the internal heterogeneity of grid cells, or an ability to incorporate higher resolution parameterization models to calculate grid cell biophysical parameters. Thus, 1-km data can be aggregated to other grid dimensions (e.g. 10-, 20-, 50-, 100-km, etc.) using intelligent methods for determining the landscape types most important to the problem being addressed.
- Coupled models can be run with differing resolution land cover inputs, but still based on consistent land cover data. For example, a GCM with a 4° latitude by 5° longitude horizontal grid can be used to estimate future climate trends. The climate results of the GCM can be input to a mesoscale meteorological model operating at a 40-km^2 resolution to simulate regional hydrological variables. This can then be input to agriculture production models operating at a 1-km resolution. The 1-km source data from the global land cover database can be aggregated to the grid cell resolutions used by all models, allowing use of a single, consistent land cover source.

Several additional trends will likely affect future global land cover databases. One challenge is the need for increasingly more complex descriptions and attributes describing land cover and associated characteristics. A key problem is associated with the fact that land cover data have traditionally been represented using discrete scales of measurement. However, mathematical models typically use continuous data.[40] As a result, thematic data, such as the categories used with traditional land cover and soils maps, must be converted to parameter values (e.g. stomatal resistance, surface roughness, leaf area index) which can be used in numerical calculations. An example of this is the effort to quantify forest canopy characteristics such as percentage canopy cover and percentage leaf type.[41] The trend toward continuous variables means that researchers must use measurements from limited field or laboratory studies and judgment to select the values that are associated with particular land cover classes, and modify and retest models based on the new types of inputs.

A related challenge is that environmental processes operate on many different time and space scales, and there may be scale thresholds at which critical processes change. To accommodate increasingly dynamic models, data sets will need to be multiscale and multitemporal, while still being consistent over large areas.

New global land data sets will have accuracy assessments that quantify data quality levels. This trend will provide the science quality data needed for new generations of earth systems models. While the recent IGBP DISCover database has documented accuracy, there is still considerable work needed to develop sound, yet practical, validation standards.

Finally, a new generation of satellite sensors will become operational in the early years of the twenty-first century that will provide imagery that are particularly well suited for developing global land data. The French satellite Système Probatoire pour l'Observation de la Terre (SPOT) vegetation sensor and the US Earth Observing System Moderate Resolution Imaging Spectrometer (MODIS), for example, will provide global data with appropriate resolution and spectral characteristics for generating a wide range of new land data sets. There are already plans to establish specialized, operational quantitative databases using MODIS data. MODIS processing plans call for the routine development of global data sets on albedo, land temperature, burned areas, land cover, and land cover change.[42]

Environmental models are rapidly increasing in sophistication owing to (1) improvements in computing hardware performance, (2) increasing use of geographic

information systems and object-oriented databases, and (3) scientific advancements in earth systems science.[40] Data inputs to these models must be based not only on today's immediate needs, but also the anticipated requirements of the next generations of models. A 1-km global land cover database should meet the needs of the majority of the modeling community for the next several years.

ABBREVIATIONS AND ACRONYMS

AVHRR	Advanced Very High-resolution Radiometer
BAHC	Biological Aspects of the Hydrologic Cycle
BATS	Biosphere–Atmosphere Transfer Scheme
BVOC	Biogenic Volatile Organic Compounds
DEM	Digital Elevation Model
DISCover	Data and Information System Land Cover
GCM	General Circulation Model
IGBP	International Geosphere Biosphere Programme
MODIS	Moderate Resolution Imaging Spectrometer
NOAA	National Oceanic and Atmospheric Administration
NPP	Net Primary Productivity
SiB	Simple Biosphere Model
SiB2	Simple Biosphere Model Two
SPOT	Système Probatoire pour l'Observation de la Terre
UNESCO	United Nations Educational, Scientific, and Cultural Organization
UNFAO	United Nations Food and Agriculture Organization
USGS	US Geological Survey
WRI	World Resources Institute

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Land Cover Assessment and Monitoring • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring • Semiarid Land Assessment: Monitoring Dry Ecosystems with Remote Sensing • Temperate Forest Resource Assessment by Remote Sensing • Tropical Forest Resource Assessment by Remote Sensing

REFERENCES

1. R.R. Nemani, S.W. Running, 'Satellite Monitoring of Global Land Cover Changes and Their Impact on Climate', *Clim. Change*, **31**, 395–413 (1995).
2. E. Matthews, 'Global Vegetation and Land Use: New High Resolution Data Bases for Limited Studies', *J. Climatol. Appl. Meteorol.*, **22**, 474–487 (1983).
3. J.S. Olson, J.A. Watts, *Major World Ecosystem Complex Map*, Oak Ridge National Laboratory, Oak Ridge, TN, 1982.
4. M.F. Wilson, A. Henderson-Sellers, 'A Global Archive of Land Cover and Soils Data for Use in General Circulation Models', *J. Climatol.*, **5**, 119–143 (1985).
5. R.S. Defries, J.R.G. Townshend, 'NDVI-derived Land Cover Classification at a Global Scale', *Int. J. Remote Sens.*, **15**, 3567–3586 (1994).
6. T.R. Loveland, B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Yang, J.W. Merchant, 'Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data', *Int. J. Remote Sens.*, in press.
7. R. Leemans, W. Cramer, J.G. van Minnen, 'Prediction of Global Biome Distribution Using Bioclimatic Equilibrium Models', in *Global Change: Effects on Coniferous Forests and Grasslands*, eds. A. Breymeyer, D.O. Hall, J.M. Melillo, John Wiley and Sons, New York, 1997.
8. J.R.G. Townshend, C. Justice, W. Li, C. Gurney, J. McManus, 'Global Land Cover Classification by Remote Sensing: Present Capabilities and Future Possibilities', *Remote Sens. Environ.*, **35**(2), 243–255 (1991).
9. R.S. Defries, J.R.G. Townshend, 'Global Land Cover: Comparison of Ground-based Data Sets to Classifications with AVHRR Data', in *Environmental Remote Sensing from Regional to Global Scales*, eds. G.M. Foody, P.J. Curran, John Wiley and Sons, New York, 84–110, 1995.
10. R. Defries, M. Hansen, J.G.R. Townshend, 'Global Discrimination of Land Cover Types from Metrics Derived from AVHRR Pathfinder Data', *Remote Sens. Environ.*, **54**, 209–222 (1995).
11. S.O. Los, C.O. Justice, C.J. Tucker, 'A Global 1° by 1° NDVI Data Set for Climate Studies Derived from the GIMMS Continental NDVI Data', *Int. J. Remote Sens.*, **15**(17), 3493–3518 (1994).
12. J. Scepan, 'Thematic Validation of High Resolution Global Land Cover Data Sets', *Photogramm. Eng. Remote Sens.*, **65**(9), 1051–1060 (1999).

13. Food and Agriculture Organization of the United Nations, *Agrostat-PC*, FAO, Rome, 1993 (diskette).
14. Food and Agriculture Organization of the United Nations, 'Forest Resources Assessment 1990: Global Synthesis', FAO Forestry Paper 124, FAO, Rome, 1995.
15. World Resources Institute, *World Resources 1998–99: a Guide to the Global Environment*, Oxford University Press, New York, 1998.
16. UNESCO, *International Classification and Mapping of Vegetation*, UNESCO, Paris, 1973.
17. J.R. Anderson, E.E. Hardy, J.T.Roach, R.E. Witmer, 'A Land Use and Land Cover Classification System for Use with Remote Sensor Data', US Geological Survey Professional Paper 964, US Geological Survey, Reston, VA, 1976.
18. S.W. Running, T.R. Loveland, L.L. Pierce, 'A Vegetation Classification Logic Based on Remote Sensing for Use in Global Biogeochemical Models', *Ambio*, **23**(1), 77–81 (1994).
19. T.R. Loveland, A.S. Belward, 'The IGBP–DIS Global Land Cover Data Set (DISCover): First Results', *Int. J. Remote Sens.*, **18**(15), 3289–3295 (1997).
20. D.B. Gesch, K.S. Larson, 'Techniques for Development of Global 1-Kilometer Digital Elevation Models', in *Pecora 13 Proceedings Human Interactions with the Environment: Perspectives from Space*, American Society of Photogrammetry and Remote Sensing, Bethesda, MD, 568–572, 1998.
21. L. Zobler, 'A World Soil File for Global Climate Modeling', Technical Memorandum 87802, National Aeronautics and Space Administration, Washington, DC, 1986.
22. G.B. Bonan, 'Land–Atmosphere Interactions for Climate System Models: Coupling Biophysical, Biogeochemical, and Ecosystem Dynamical Processes', *Remote Sens. Environ.*, **51**(1), 57–73 (1995).
23. P.J. Sellers, Y. Mintz, Y.C. Sud, A. Dalcher, 'A Simple Biosphere Model (SiB) for Use Within General Circulation Models', *J. Atmos. Sci.*, **43**, 505–531 (1986).
24. R.E. Dickinson, A. Henderson-Sellers, P.J. Kennedy, M.F. Wilson, 'Biosphere–Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model', NCAR Technical Note NCAR/TN–275+STR, NCAR, Boulder, CO, 1986.
25. R.A. Pielke, R. Avissar, 'Influence of Landscape Structure on Local and Regional Climate', *Landscape Ecol.*, **4**(2/3), 135–155 (1990).
26. J.H. Copeland, R.A. Pielke, T.G.F. Kittel, 'Potential Climatic Impacts of Vegetation Change: a Regional Modeling Study', *J. Geophys. Res.*, **101**(ND3), 7409–7418 (1996).
27. E. Kinnee, C. Geron, T. Pierce, 'United States Land Use Inventory for Estimating Biogenic Precursor Emissions', *Ecol. Appl.*, **7**(1), 46–58 (1997).
28. A. Guenther, 'Seasonal and Spatial Variations in Natural Volatile Organic Compound Emissions', *Ecol. Appl.*, **7**(1), 34–45 (1997).
29. D.S. Schimel, T.G. Kittel, W.J. Parton, 'Terrestrial Biogeochemical Cycles: Global Interactions with the Atmosphere and Hydrology', *Tellus*, **43AB**, 188–203 (1991).
30. W.J. Parton, J.W.B. Stewart, C.V. Cole, 'Dynamics of C, N, P, and S in Grassland Soils: a Model', *Biogeochemistry*, **5**, 109–131 (1988).
31. E.R. Hunt, Jr, S.C. Piper, R. Nemani, C.D. Keeling, R.D. Otto, S.W. Running, 'Global Net Carbon Exchange and Intra-annual Atmospheric CO_2 Concentrations Predicted by an Ecosystems Process Model and Three-dimensional Atmospheric Transport Model', *Global Biogeochem. Cycles*, **10**(3), 431–456 (1996).
32. I.C. Burke, W.K. Lauenroth, W.J. Parton, C.V. Cole, 'Interactions of Landuse and Ecosystem Structure and Function: a Case Study in the Central Great Plains', in *Integrated Regional Models: Interactions Between Humans and Environment*, eds. P. Groffman, G. Likens, Chapman and Hall, New York, 79–95, 1994.
33. C.B. Field, J.T. Randerson, C.M. Malmstrom, 'Global Net Primary Production: Combining Ecology and Remote Sensing', *Remote Sens. Environ.*, **51**(1), 74–88 (1995).
34. J.D. Wickham, T.G. Wade, K.B. Jones, K.H. Riitters, R.V. O'Neill, 'Diversity of Ecological Communities of the United States', *Vegetatio*, **119**, 91–100 (1995).
35. S.L. Ustin, C.A. Wessman, B. Curtiss, E. Kasischke, J. Way, V.C. Vanderbilt, 'Opportunities for Using the EOS Imaging Spectrometers and Synthetic Aperature Radar in Ecological Models', *Ecology*, **72**(6), 1934–1945 (1991).
36. L.T. Steyaert, T.R. Loveland, J.F. Brown, B.C. Reed, 'Integration of Environmental Simulation Models with Satellite Remote Sensing and Geographic Information Systems Technologies: Case Studies', in *Proceedings: Pecora 12 Symposium on Land Information from Space-based Systems*, American Society of Photogrammetry and Remote Sensing, Bethesda, MD, 407–417, 1994.
37. 'International Geosphere–Biosphere Program, Biospheric Aspects of the Hydrological Cycle (BAHC): the Operational Plan', IGBP Global Change Report No. 27, International Geosphere–Biosphere Program, Stockholm, 1993.
38. P. Doraiswamy, G. Hart, M. Craig, P. Cook, 'The Anomalous '93 Growing Season – How USDA Used AVHRR Data', in *Proceedings, ASPRS Annual Convention*, American Society of Photogrammetry, Reno, NV, 144–151, Vol. 1, 1994.
39. R.P. DeFusco, 'Modeling Bird Hazards to Aircraft: a GIS Application Study', *Photogramm. Eng. Remote Sens.*, **59**(10), 1481–1487 (1993).
40. K.K. Kemp, 'Spatial Models for Environmental Modeling With GIS', in *Proceedings: 5th International Symposium on Spatial Data Handling*, International Geographical Union, Charleston, SC, 524–533, Vol. 2, 1992.
41. R.S. Defries, C.B. Field, I. Fung, C.O. Justice, P.A. Matson, E. Matthews, H.A. Mooney, C.S. Potter, K. Prentice, P.J. Sellers, J. Townshend, C.J. Tucker, S.L. Ustin, P.M.

Vitousek, 'Mapping the Land Surface for Global Atmosphere–Biosphere Models: Toward Continuous Distributions of Vegetation's Functional Properties', *J. Geophys. Res.*, **100**(D10), 20 867–20 882 (1995).

42. S.W. Running, C.O. Justice, V. Salomonson, D. Hall, J. Barker, Y. Kaufmann, A. Strahler, A. Huete, J.-P. Muller, V. Vanderbilt, Z. Wan, P. Teillet, D. Carneggie, 'Terrestrial Remote Sensing Science and Algorithms Planned for EOS/MODIS', *Int. J. Remote Sens.*, **17**(15), 3587–3620 (1994).

Hyperspectral Remote Sensing: Data Collection and Exploitation

Stephanie Sandor-Leahy and John Shepanski
TRW Inc., Redondo Beach, USA

1 Introduction **1**
1.1 Hyperspectral Imaging 1
1.2 Information Resources 2
2 Spectroscopic Considerations **3**
2.1 Interaction of Light with the Atmosphere 3
2.2 Spectral Properties of Selected Materials 4
3 Applications Overview **6**
4 Data Collection **7**
4.1 Early Instruments 7
4.2 Contemporary Instruments 7
4.3 TRW Imaging Spectrometer III: A Dispersive Hyperspectral Imager 8
4.4 Livermore Imaging Fourier Transform Infrared Spectrometer: An Imaging Fourier Transform Spectrometer 10
5 Data Correction **10**
5.1 Types of Correction 10
5.2 Computing a Reflectance Data Cube 12
6 Spectral Scene Analysis **13**
6.1 Spectral Matching 14
6.2 Spectral Unmixing 14
7 Future Developments **15**
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **17**

Hyperspectral imaging (HSI) spectrometers are remote sensing instruments that acquire images in a large number, typically hundreds, of contiguous spectral channels throughout the visible to long-wave infrared (IR) portions of the spectrum from 0.4 to 14 μm.[1] These systems are usually flown on aircraft platforms and use either platform motion or mirror mechanisms to scan a region of the earth's surface. The high-resolution spectral features represented in the data cube allow for discrimination of materials in a scene. Because of the large number of spectral channels that are acquired by these instruments, they are termed hyperspectral, in contrast to multispectral instruments that obtain relatively few spectral bands. Like all remote sensing technologies, HSI must contend with data perturbations caused by atmospheric effects. In addition, data analysis must account for spectral mixing of multiple constituent materials within each pixel's field of view (FOV).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Hyperspectral Imaging

HSI is a form of spectroscopy, the study of how electromagnetic radiation is reflected, scattered, absorbed by, or emitted from the surface of a material. Hyperspectral sensors typically operate throughout the visible to long-wave IR portions of the spectrum from 0.4 to 14 μm[1] (Figure 1) and measure an effectively continuous spectrum for each pixel in a scene, generating an image data cube with two spatial dimensions and one spectral (Figures 2 and 3). Solids, liquids, and gases interact with photons differently at different wavelengths. The wavelength dependence of absorption processes is used to derive information about the chemistry of a material and to identify a material from its spectral signature. For example, some of the major spectral absorption features in vegetation can be attributed to plant pigments, such as chlorophylls, which produce a highly characteristic vegetation reflectance signature (Figure 4).

Hyperspectral systems can be contrasted with traditional multispectral systems (see section 4.1) which produce radiance measurements corresponding to an integrated response over broad spectral channels. A typical multispectral instrument obtains one measurement over a wide spectral band whereas the hyperspectral system obtains approximately 100 measurements covering the same spectral region (see Figure 5). Multispectral systems differentiate amongst general classes of materials, while hyperspectral instruments support detailed identification and quantification of materials.

Hyperspectral remote sensing combines the analytical power of spectroscopy with the established data collection and processing techniques of remote sensing but encounters new challenges. For example, chemical analysis via spectroscopy has not typically been complicated by data corruption due to inhomogeneous scene illumination and atmospheric absorption and scattering. The analysis is further complicated if pixel-to-pixel quantitative comparisons of surface features are desired and there exist substantial illumination and water vapor variations over the FOV. Analysis algorithms robust to these spatial inhomogeneities are required.

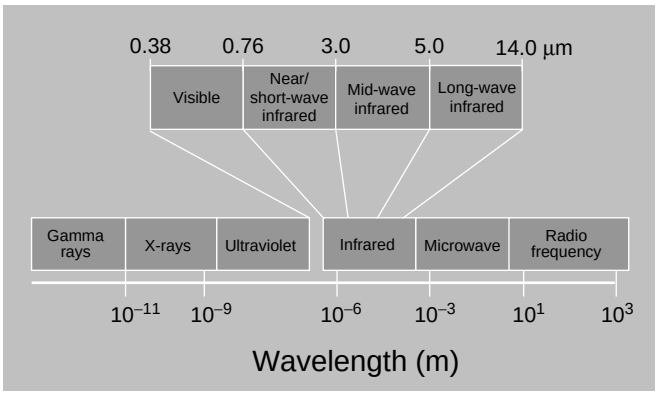

Figure 1 Electromagnetic spectrum.

1.2 Information Resources

Hyperspectral remote sensing is a young field; not many texts are available that specifically deal with its capabilities and challenges. In regard to multispectral remote sensing, which shares many challenges with HSI, Muller provides a good introduction and historical perspective.[2] Landgrebe[3] gives an excellent discussion of multispectral and hyperspectral data analysis and information extraction. Good general overviews of spectral reflectance signatures and remote sensing technology and analysis are available.[4–6] Wolfe[7] has provided a most useful work on the physics of imaging spectrometers. Various government agencies and research institutes have published technical reviews and reports.[1,8] Proceedings from the AVIRIS (airborne visible/infrared imaging spectrometer) workshop, which has been held by the Jet Propulsion Laboratory (JPL) since 1989, provide an excellent source of papers on HSI applications.

Numerous articles on hyperspectral applications and analysis techniques have appeared in remote sensing journals such as the *IEEE Transactions on Geoscience*

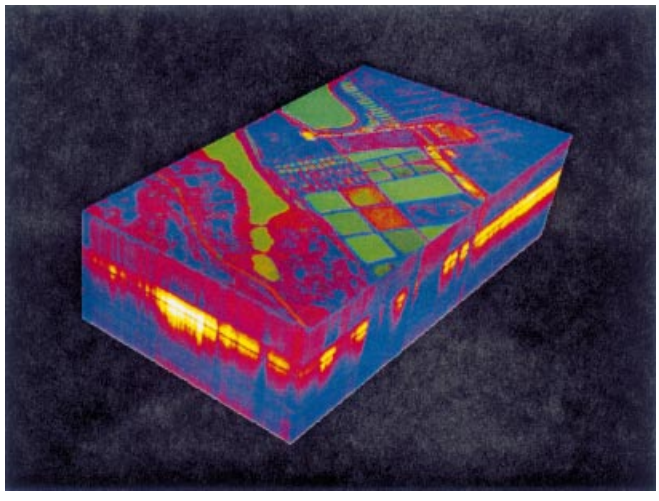

Figure 3 A hyperspectral data cube.

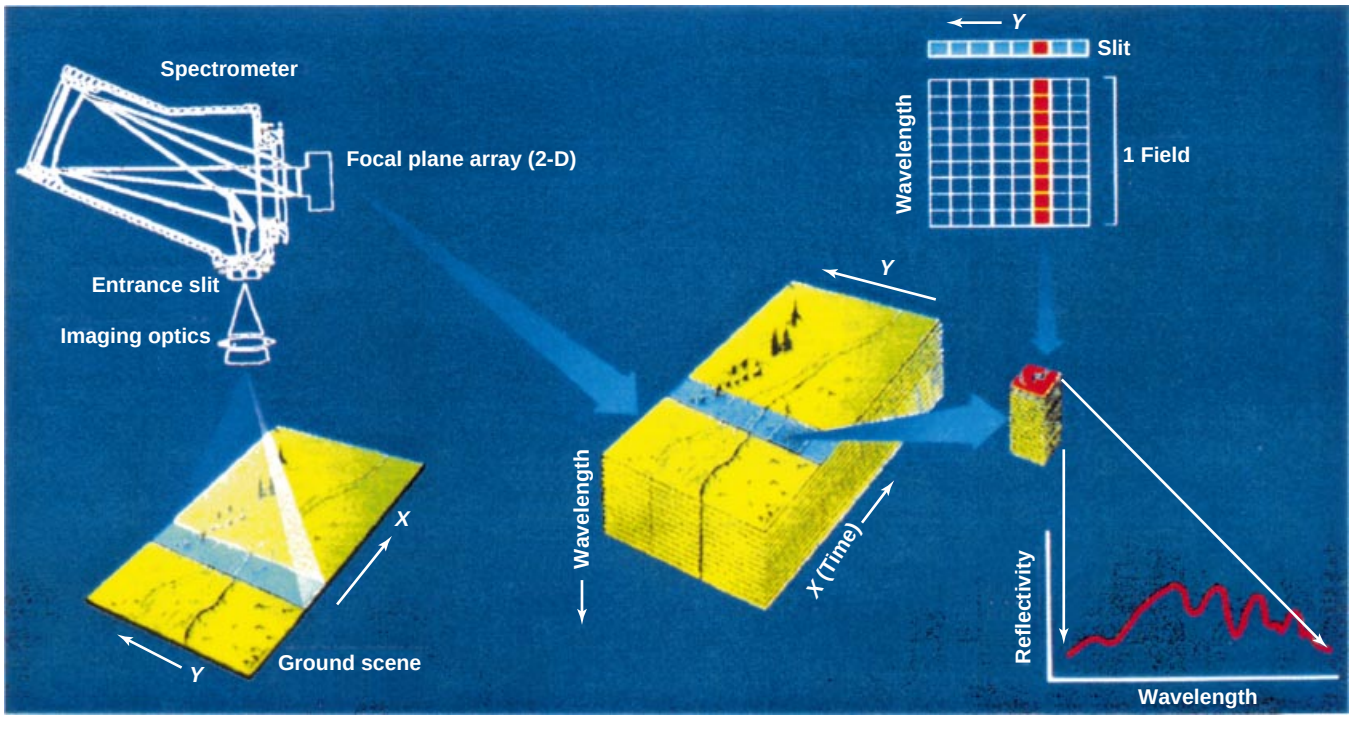

Figure 2 Generation of a hyperspectral image for a pushbroom imaging system with a two-dimensional focal plane array.

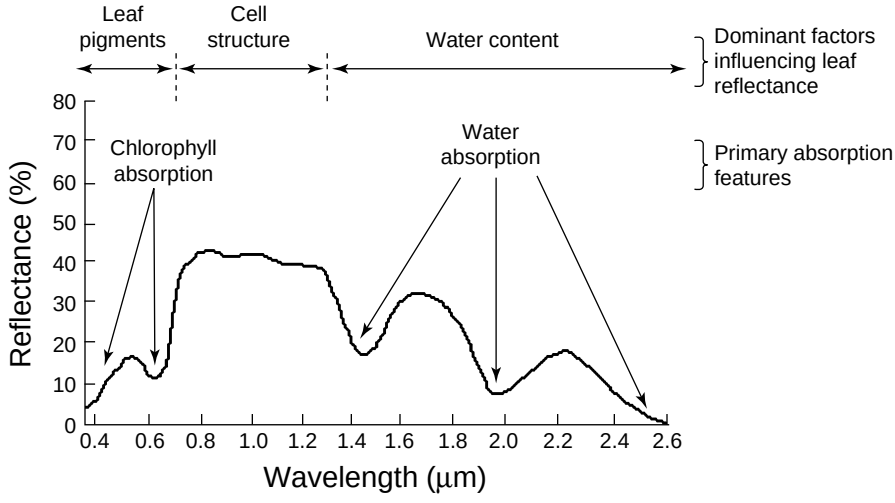

Figure 4 Dominant features of a typical vegetation spectrum.

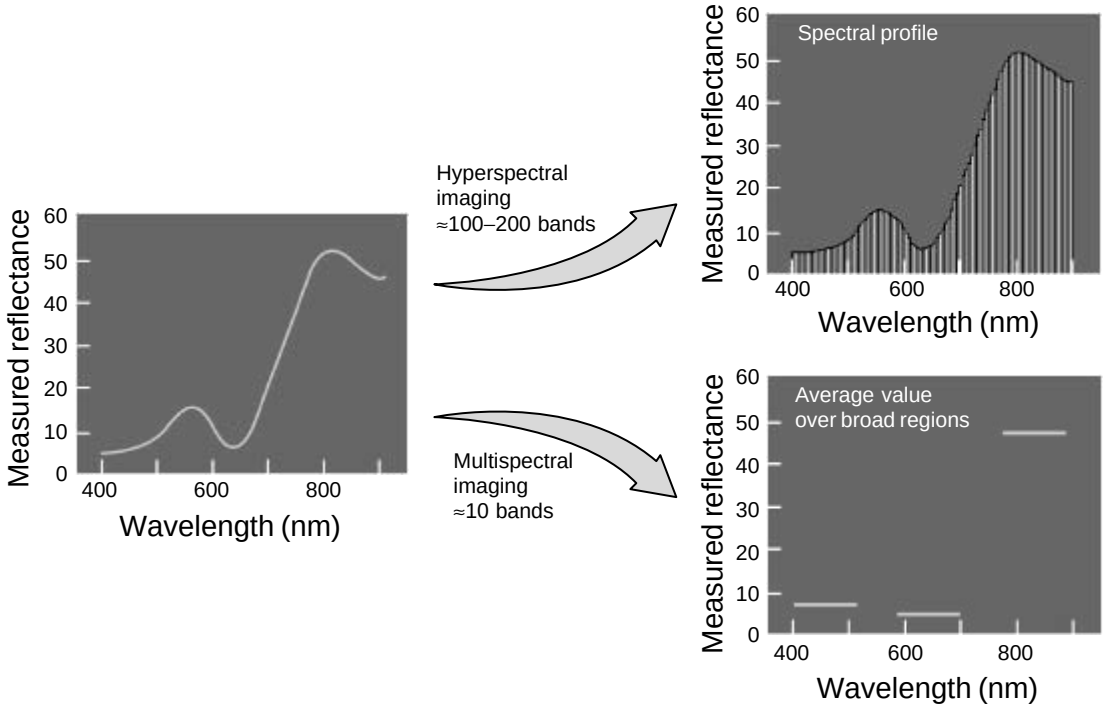

Figure 5 Number of spectral measurements for hyperspectral vs multispectral imaging systems.

and Remote Sensing, *Remote Sensing of Environment*, *International Journal of Remote Sensing*, and *Photogrammetric Engineering and Remote Sensing*. Over the last several years, most remote sensing and optical instrumentation conferences have held sessions devoted to HSI and its applications. These conferences include Applied Geologic Remote Sensing, Remote Sensing for Marine and Coastal Environments, International Airborne Remote Sensing Conference, the Annual SPIE Meeting (Optical Science, Engineering, and Instrumentation): Imaging Spectrometery Conference and others. The proceedings from these meetings are an important resource of information on HSI.

2 SPECTROSCOPIC CONSIDERATIONS

2.1 Interaction of Light with the Atmosphere

When radiant energy is incident on a material's surface, a portion of that energy will be reflected, absorbed,

or transmitted at various wavelengths. From these surface interactions one can infer the characteristics of the material being observed. However, all radiation recorded by a remote hyperspectral imager passes through some distance, or path length, of atmosphere. Each time radiant energy passes through the atmosphere, it undergoes absorption and scattering, significantly altering the intensity and spectral composition of the light that is available for detection by the spectrometer. Therefore, the atmosphere is a primary determinant of the operational range of the instrument.

Atmospheric absorption processes convert incoming radiation to heat through an exchange of energy between a molecule and the incident electromagnetic energy. The most efficient absorbers of solar radiation are water vapor, carbon dioxide, ozone, and oxygen. These atmospheric constituents attenuate radiation very strongly in certain wavelength bands and it is the wavelength ranges outside these main absorption bands that are used for HSI. The wavelength regions that are useful for imaging are known as atmospheric transmission windows (see Figure 6). Although the earth cannot be observed in regions of high atmospheric absorption, these wavelength ranges can be used for atmospheric sounding and are the basis of current atmospheric retrieval and correction algorithms (see section 5.2).

In the atmospheric transmission windows, scattering by atmospheric particles is the dominant mechanism affecting the spectrometer's ability to correctly measure surface radiance. Scattering removes energy from the incident electromagnetic wave and re-emits some portion of it at the same wavelength but in different directions. When atmospheric particles are much smaller than the wavelength of incoming radiation, Rayleigh scattering is the dominant scattering mechanism. Rayleigh scattering is proportional to $(1/\lambda)^4$, where λ is the wavelength of incoming radiation; therefore, this mechanism preferentially attenuates shorter wavelengths. Rayleigh scattering accounts for the blue color of the sky. Since the shorter wavelengths (blue to ultraviolet regions of the spectrum) are scattered more strongly than the longer (red) wavelengths, an observer will see blue light when looking at the sky in any direction.

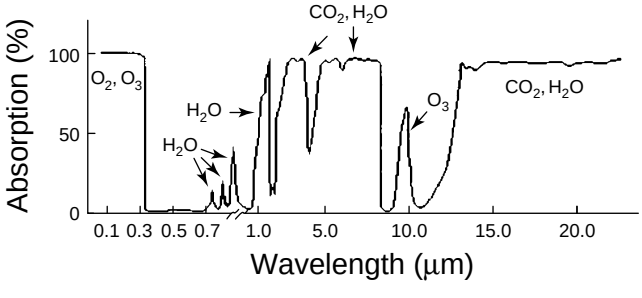

Figure 6 The absorption of radiation by gases in the atmosphere. The wavelength regions where absorption is low are known as atmospheric transmission windows.

When the sizes of the atmospheric particles are approximately equal to the wavelength of the incoming radiation, Mie scattering occurs. This mechanism is wavelength dependent but influences longer wavelengths than Rayleigh scattering. Generally, atmospheric Mie scattering becomes significant when dust, haze, fog or smoke are present.

Nonselective scatter, where radiation is scattered equally in all directions, occurs when the atmospheric particles are much larger than the wavelength of the incoming radiation. For example, water droplets with diameters on the order of 5 to 10 μm cause nonselective scattering. In the visible wavelengths, thick clouds or fog appear white or bluish-white due to isotropic scattering by these larger particles.

2.2 Spectral Properties of Selected Materials

In the visible/near-infrared (VNIR, 0.4–1.1 μm) to shortwave infrared (SWIR, 1.1–2.5 μm) wavelength range, hyperspectral instruments measure reflected solar radiation after it passes through some path length of atmosphere. The shapes of reflected spectra depend on the atmosphere and on a variety of surface properties (e.g. pigmentation, moisture content and leaf morphology, mineral and moisture contents of soils, and level of sedimentation in water). Figure 7 shows hyperspectral images with corresponding spectral radiance curves (in the 0.4 μm to 2.5 μm range) for a variety of land cover types. Some of the characteristic features of these materials are visible in the spectral curves.

The major absorption features in vegetation spectra can be attributed to plant pigments (chlorophylls, zanthophyll, and carotenoids) and water. The dip in the spectrum at 0.65 μm is caused by the strong absorption of incoming radiation by chlorophyll a and b in the leaves. The other two dips at 1.45 to 1.55 μm and 1.90 to 1.95 μm are caused by water absorption in the leaves. The absorption features seen in vegetation spectra are related to organic compounds common to the majority of plant species.[9,10] Other minor features are attributable to components including cellulose, lignin, proteins, starches, and sugars (see Table 1). In the 0.7 to 1.0 μm range, leaves are highly reflective due to the cellular structure of leaf tissue. In this wavelength range, vegetation spectra typically exhibit a strong plateau, the height of which decreases as moisture stress increases.

Bodies of water have a low spectral reflectance (less than 10%) in the visible, blue-green region. At wavelengths longer than 0.75 μm, water absorbs almost all incoming solar radiation. Therefore, most applications of

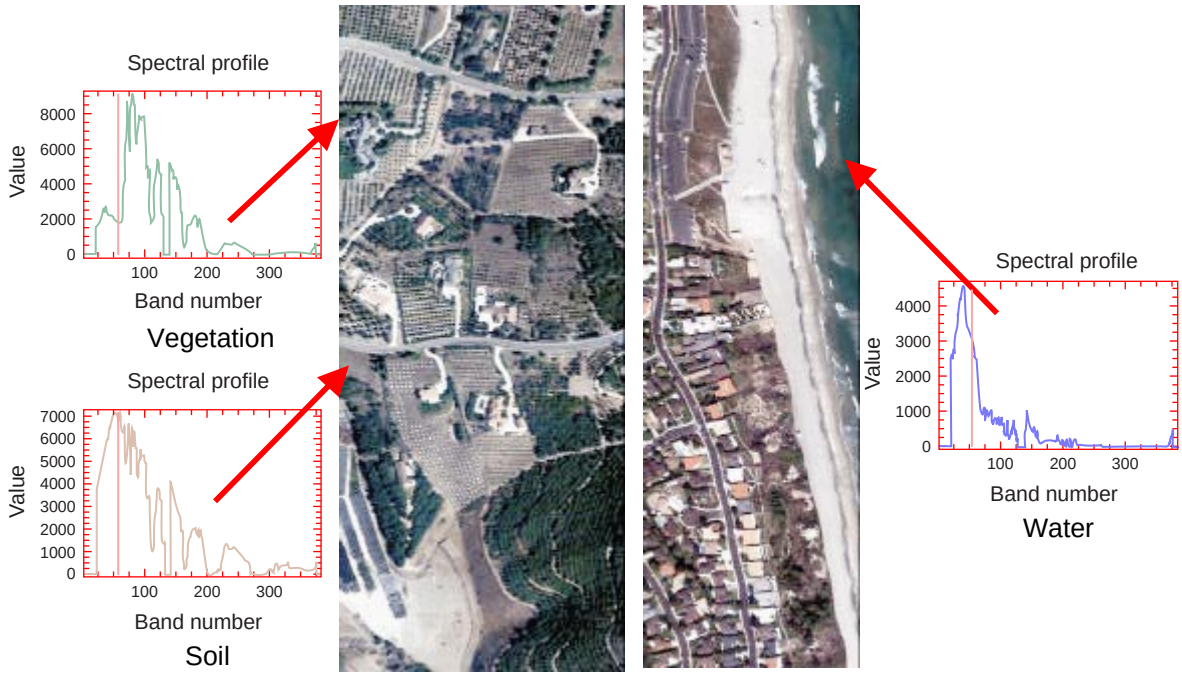

Figure 7 Hyperspectral imagery of different ground cover types, acquired by the TRWIS (TRW Imaging Spectrometer) III instrument. Each pixel of the image corresponds to a 384-band spectral measurement of a ground sample. The gap in the spectral curves (starting at band number 128) corresponds to the overlap region between the SWIR and VNIR spectrometers. The images have been calibrated to radiance units of 0.01 W $(m^2\ sr\ \mu m)^{-1}$.

Table 1 Wavelength ranges that contain features of interest for vegetation spectra

Wavelength range (μm)	Vegetation phenomenology
0.45–0.67	Approximate chlorophyll absorption bands
0.7–0.9	Differentiation between deciduous and coniferous trees
0.7–1.3	Reflectance resulting from internal leaf structure, plant stresses
Beyond 1.3	Leaf reflectance approximately inversely related to total water content
1.0–2.4	Cellulose, sugar, starch, etc., contents in leaves

HSI to water environments use the wavelength range of approximately 0.4 to 0.6 μm, where valuable information about water condition can be gleaned (see section 3). In this range, however, Rayleigh and aerosol scattering contribute strongly to the reflected energy measured by a hyperspectral sensor, and an accurate estimate of this atmospheric contribution is necessary.[11] Water reflectance can also change dramatically as a result of the presence of suspended organic or inorganic materials, and shallow, clear water bodies can exhibit an increased reflectance due to reflections from bottom materials. Taken together, all of these phenomena make applications of remote sensing to water particularly difficult.

HSI reveals a wealth of information about minerals, and geology is perhaps the most mature application of imaging spectrometry to earth observation. Solar radiation induces electronic transitions, charge transfers, and molecular vibrations, which generate a variety of distinctive spectral features useful for mineral identification. In the 0.4 to 2.5 μm range, wavelength dependent absorption bands in mineral spectra are due to electronic transitions. For example, water and hydroxyl produce particularly characteristic absorption features at approximately 1.4, 1.9, 2.2, and 2.7 μm. The strong hydroxyl features have proven to be particularly diagnostic of clay mineralogy.[12] Carbonates produce a number of absorption features at 1.9, 2.0, 2.16, 2.35, and 2.55 μm.

In the long-wave infrared region (LWIR, 8–14 μm), rather than measure reflected solar radiation, hyperspectral instruments detect the energy being emitted from the earth's surface or from objects on the ground.[1] In the thermal region a material's emissivity is the property of interest, where emissivity is a wavelength dependent measure of a material's ability to emit energy in comparison to a blackbody. Real objects are not perfect radiators and emit only a fraction of the energy of a blackbody at the same temperature. Spectral features of minerals in the LWIR are due to molecular vibrations, and similarly in the VNIR/SWIR range, distinctive features in the emissivity spectrum can be used for mineral identification.

For example, reststrahlen bands are characteristic features of silicate minerals in the wavelength range beyond 8 μm. The reststrahlen features are emissivity minima that move from shorter to longer wavelengths as silicate content decreases. In general, signatures of materials tend to be flatter and exhibit less peak-to-valley spectral variation in the LWIR than in the reflected VNIR/SWIR ranges.

3 APPLICATIONS OVERVIEW

Because HSI yields an abundance of information about materials covering the earth's surface, a wide range of problems can be addressed using this technology. This section will briefly describe some of these applications. More details can be found in the references and in the other sources mentioned in section 1.2.

As previously noted, geology and mineralogy are perhaps the most mature applications of HSI. Multispectral imagery has long been applied to geologic exploration. Imagery from the Landsat Thematic Mapper and Multispectral Scanner has been used to distinguish rock types, create mineral maps, and delineate major folds and fracture systems. However, the spectral resolution of Landsat and other multispectral scanners is not sufficient to uniquely determine the mineralogic composition of rocks. For example, a feature at 2.2 μm can be caused by sericite or kaolinite. With measurements of sufficient spectral resolution the minerals can be differentiated.[8] An absorption feature due to pyrophyllite is centered at 2.17 μm. The feature is broadened when other hydroxyl-bearing minerals are present. For a multispectral system, the broadened feature is indistinguishable from the characteristic spectrum of alunite, and only measurements of sufficiently high spectral resolution can detect that alunite is not present. Figure 8 shows an example where high spectral resolution in particular wavelength bands is necessary to differentiate between minerals.

HSI allows not only mineralogic identification and classification but also quantitative estimates of surface mineral abundance. Early work with the airborne AVIRIS sensor (see section 4.2) demonstrated that hyperspectral data could be used to derive both detailed mineral maps of large areas and mineral abundance estimates at the individual pixel level.[13–15] These remotely derived image products were verified through ground studies of the surveyed areas. In recent years, as the signal-to-noise performance of detectors has improved and processing algorithms have been refined, even finer spectral features have been measured, leading to extremely detailed mineral maps and accurate abundance estimates at the sub-pixel level.[16] The region of Cuprite, Nevada has

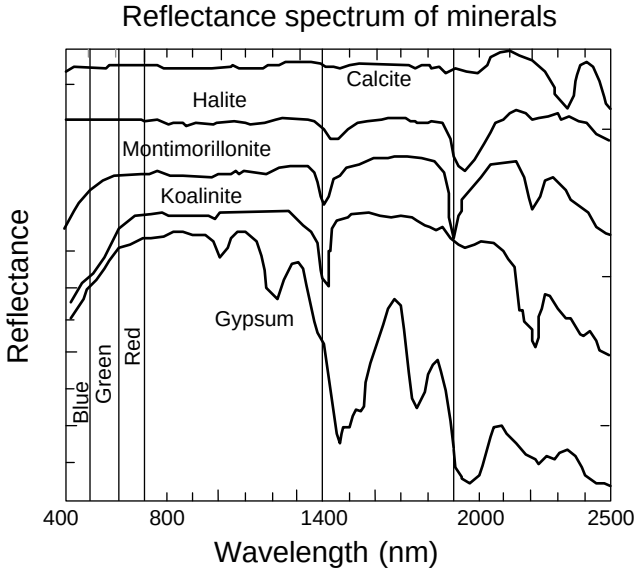

Figure 8 Spectral reflectance curves for a variety of minerals.

been used as testing ground for new scanners and mineral classification algorithms. Over a number of years, this area has been characterized by ground, airborne, and satellite measurements, and the detailed maps that have been generated from hyperspectral data are an indication of the advancements in mineralogic identification that are possible with HSI.

Hyperspectral sensors have been used for oceanographic studies and to image bodies of water, including oceans, coastal regions, inland lakes and rivers. The high spectral resolution of these instruments has contributed to the understanding of complex water environments, such as coastal regions, where suspended sediments, bottom reflectance, and dissolved organic matter make the interpretation of optical remote sensing data particularly difficult.[17] In water bodies where there is sufficient penetration of solar energy, bathymetry (depth measurement) has been performed. Hyperspectral image data have been used to address an important component of marine remote sensing: the detection of phytoplankton to derive estimates of biological activity.[18] In certain water bodies, multispectral sensors have been used to detect pigments, such as chlorophyll *a*, a photosynthetic pigment common to all algal groups. However, hyperspectral sensors allow measurement of specific absorption features that are unique to pigments other than chlorophyll *a*. The detection of these pigments is applicable to the understanding and characterization of algal blooms, which can be harmful to humans and the ecosystems of bays and rivers.[19]

Another major application area for hyperspectral remote sensing is the observation of vegetation. Traditionally, remote sensing techniques, from aerial photography to multispectral imaging, have been used to

map the area and extent of vegetative regions. But HSI has been used to produce accurate maps of vegetation varieties for a wide range of landscapes, from national parks to agricultural fields. This technology also generates a wealth of information about the biophysical and biochemical attributes of a living and changing vegetation ecosystem. Several studies have investigated the relationship between remotely sensed hyperspectral data and the foliar chemistry of forests, grasslands, and wetland canopies.[9,20–23] These data have been used to generate maps of foliar nitrogen, which is strongly related to the rate of photosynthesis and can be used to predict forest growth rates.[24] Hyperspectral data have also been used to produce estimates of canopy water content, which are applied to risk assessment and mitigation for fire-prone regions of urbanized landscapes.[25] The ability to measure the spectral changes in vegetation during its growth cycle also has implications for forestry and agriculture, particularly in the area of precision farming.[26] HSI can be used to perform crop and forest inventories, crop health assessments and pre-visual stress detection without extensive field work.

Along with remote sensing of the earth's environment, HSI has demonstrated utility in the area of target discrimination and identification. Because man-made materials can be accurately identified, and even quantified, by remote measurement of their distinct spectral signatures, hyperspectral data are used to accomplish a number of target detection missions. These applications include distinguishing vehicles, target panels, and fabrics in a variety of natural and urban backgrounds, anomaly detection, classifying target signatures in real time, and detecting subpixel-sized targets.[27–29] HSI is also relevant to the problem of detecting buried land mines.[30,31] Many well-known hyperspectral image analysis techniques (see section 6) are directly applicable to this type of data, but a primary concern of data exploitation in this regime is to minimize the false alarm rate of the detection process. By reducing the number of incorrectly identified targets, hyperspectral image analysis becomes well suited to the problem of real-time target detection.

4 DATA COLLECTION

A number of approaches are available for multispectral imaging and HSI,[1,7,32] but most fall in three categories: those using bandpass optical filters, dispersive elements (prisms or diffraction gratings), and interferometers. Devices based on optical filters provide the easiest technical solution for multispectral imaging and were the basis of the first multispectral sensors. However, these suffer from a practical limitation in the number of spectral bands they can monitor. In the early 1980s the fabrication of large focal plane arrays facilitated the development of dispersive and interferometric hyperspectral imagers capable of simultaneously measuring hundreds of spectral bands.

Dispersive HSI sensors (see section 4.3), take instantaneous spatial/spectral frames, concatenating many frames over time to furnish the second spatial dimension of a data cube. On the other hand, an imaging Fourier transform spectrometer (IFTS, see section 4.4), takes instantaneous two-dimensional spatial frames of spectral interference amplitudes, concatenating many frames and applying a fast Fourier transform (FFT), to produce the spectral dimension of a data cube.

4.1 Early Instruments

The first multispectral scanner was developed by the Environmental Research Institute of Michigan in 1963 as an airborne instrument.[1] The age of orbital multispectral imagery began in 1972 with the launch of Landsat-1. These and other early instruments typically employed 6 to 16 spectral filters to select wavelength bands that facilitate the characterization of terrain features. The development of two-dimensional focal plane arrays made HSI possible. An IFTS was first described in 1972 by Potter,[33,34] who also helped produce a simple instrument in 1980.[35] The first deployed HSI, the Airborne Imaging Spectrometer (AIS), was built by NASA (National Aeronautics and Space Administration) JPL in 1983, and used a 64×64 element focal plane. AIS was a near-IR and SWIR imager that employed dispersive optics.[1,8] The development of true IFTSs was delayed until the early 1990s, pending the availability of high speed microcomputer boards capable of rapid FFTs.[32]

4.2 Contemporary Instruments

The number of multispectral and hyperspectral instruments has grown steadily in the 1980s and 1990s. At the time of writing, orbital collection remains the domain of multispectral imagers like Landsat and SPOT (Système Pour l'Observation de la Terre),[1] but this situation will change soon with the launch of NASA's EO-1 satellite that carries the Hyperion HSI sensor. A variety of airborne hyperspectral imagers (AHIs) are in service. Some are IFTSs (e.g. SMIFTS (Spatially Modulated Imaging Fourier Transform Spectrometer), LIFTIRS (Livermore Imaging Fourier Transform Infrared Spectrometer), HRIS (High-resolution Imaging Spectrometer)), but most use dispersive optics (e.g. AVIRIS, HYDICE (Hyperspectral Digital Imagery Collection Experiment), TRWIS III, SEBASS (Spatially Enhanced Broadbard Array Spectrograph System), AHI).[1,32,36,37] The latter group use slit camera designs that capture

images by scanning scenes. Some are ''pushbroom'' scanners (e.g. TRWIS III), that is, the scan motion is provided by the platform's movement and is in the direction of flight. Others are ''whiskbroom'' scanners (e.g. AVIRIS), that sweep the imaging slit back and forth perpendicular to the direction of flight.

Beside construction differences, hyperspectral imagers can be grouped according to spectral domain. Fourier transform spectrometers are capable of resolving finely separated spectral lines and have proven most useful in the medium-wave infrared (MWIR, 3–5.5 μm) and LWIR regimes where we observe the vibrational and rotational structures of gas absorption and emission.[1] Dispersion-based instruments tend to perform better for broader spectral features, such as those found at VNIR and SWIR wavelengths. Dispersive instruments are also useful in the MWIR and LWIR regimes when high spectral resolution is not needed, but precise radiometric calibration is required.

Whereas multispectral devices often cover bands throughout the visible and IR spectrum (e.g. AHS, DAIS, MVIS), present hyperspectral sensors tend to concentrate on one or two ranges, such as VNIR (e.g. CHRISS (Compact High-resolution Imaging Spectrographic System), TRWIS B), VNIR/SWIR (e.g. HYDICE, AVIRIS, TRWIS III), SWIR/MWIR (e.g. SMIFTS), or MWIR/LWIR (e.g. LIFTIRS, AHI).[1,32,36,37] Focal plane materials and operating temperatures vary markedly for these spectral regimes. Ambient temperature silicon charge-coupled device (CCD) arrays, like those used in video cameras, work well for VNIR sensors. In the SWIR regime, cooled InSb or HgCdTe photodiode arrays are frequently used; HgCdTe, PtSi and Si : In arrays have been used in the MWIR; HgCdTe, Si : As and Si : Ga in the LWIR.[1] Materials used for MWIR and LWIR detectors require extreme, stable cooling to function reliably. LWIR devices have a further complication: uncooled collection optics emit radiation at LWIR wavelengths and this contributes background signal at the focal plane.

In the next two sections we present detailed descriptions of dispersive and Fourier transform HSI devices.

4.3 TRW Imaging Spectrometer III: A Dispersive Hyperspectral Imager

The TRWIS III, shown in Figure 9, is a dual hyperspectral imager. A VNIR instrument covers the 0.4 to 1.0 μm spectral regime, and a co-aligned SWIR imager covers 0.9 to 2.5 μm.[36] Table 2 summarizes the instrument parameters. The TRWIS III measures a 13.1 degree swath in 384 spectral bands with 256 pixels of spatial resolution. It is a slit camera, meaning its optical FOV must scan over a scene to render an image. Typical TRWIS III

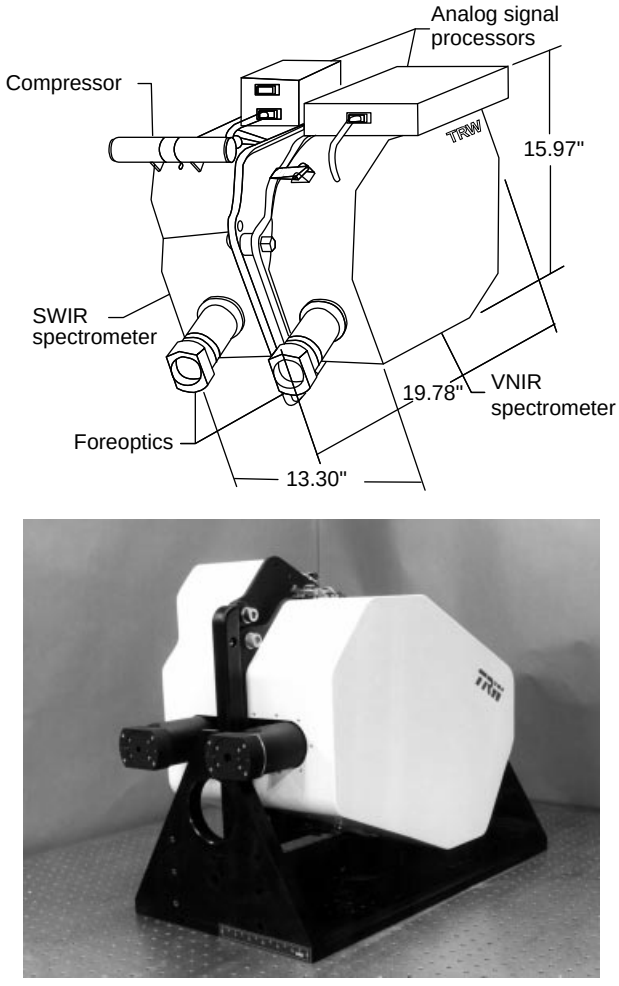

Figure 9 TRWIS III hyperspectral imager.

images comprise 1500 lines and are rendered in about 30 s. The information content of one hyperspectral image is huge: a $384 \times 256 \times 1500$ pixel data cube comprises nearly 300 Mbytes (12 bit precision, stored as 16 bit integers).

Figure 10 offers a simplified diagram illustrating the basic operation of a grating-based HSI like TRWIS III. Foreoptics focus the input scene on a narrow slit. A bandpass filter limits the wavelength range of accepted light. The image component passing through the slit is collimated and directed to a diffraction grating. Here the spectral components of the image are dispersed at different angles such that the axis of dispersion is perpendicular to the slit. Light diffracted off the grating is focused on the VNIR or SWIR focal plane. The VNIR device uses a silicon CCD array and the SWIR instrument uses a HgCdTe photodiode array maintained at 115 K by a cryogenic cooler.

A single scan frame contains both spatial and spectral data. The frame acquisition rate typically ranges between

Table 2 TRWIS III performance parameters

Parameter	Measured performance
Weight	62 kg (sensor head), 103 kg (electronics), 60 kg (data recorder)
Power at 28 V DC	<740 W peak
Instantaneous FOV	0.9 mrad
Full FOV	13.1 (256 crosstrack pixels)
Number of spectral channels	384
Wavelength range	380–2450 nm
Spectral bandwidth	5.25 nm (VNIR), 6.25 nm (TBD) (SWIR)
Spectral band purity (VNIR and SWIR)	<21% adjacent channels, <2% two channels away, <0.8% three channels away
MTF at 0.56 cycles mrad^{-1} (Nyquist)	0.24–0.4 (VNIR)
Crosstrack spectral error	<±1.2 nm (VNIR), <±0.61 nm (SWIR)
Spatial co-registration of spectral channels	<±0.18 (VNIR and SWIR)
Frame rate	15, 30, or 60 Hz
GPS position knowledge	<5 m when differential signal available
Pointing knowledge	<100 μrad roll, pitch, and yaw
Spectral calibration accuracy	<1.0 nm
Radiometric calibration accuracy	<5% (1 σ)
Quantization	12 bits
Data storage capacity	165 Gbytes per data tape

15 and 240 Hz and determines how quickly the camera slit can be scanned over the scene to produce square spatial pixels. Scanning faster than this critical rate does not overlook data, but stretches the pixel footprint – the ground sample distance (GSD) – along the scan direction. Scanning slower over-samples the scene spatially along the scan direction. Even at the critical scan rate, where the pixel footprint is square, there are some data mixing along the spatial scan direction because light is collected continuously at the focal plane. Approximately 25% of the signal in a pixel comes from landscape immediately before and after the current ground sample area (see Figure 11).

Samples of TRWIS III data cubes are shown in Figure 7. These images are partial representations of the full data cube – three spectral bands corresponding to typical red, green and blue wavelengths are merged to form a composite color picture. The spectral profiles of some representative spatial pixels are shown. An alternative presentation of HSI data involves combining bands in specific ratios to extract hidden details or highlight differences between spectral profiles. Samples of this exploitation technique and the analysis methods that produce them are presented in section 6.

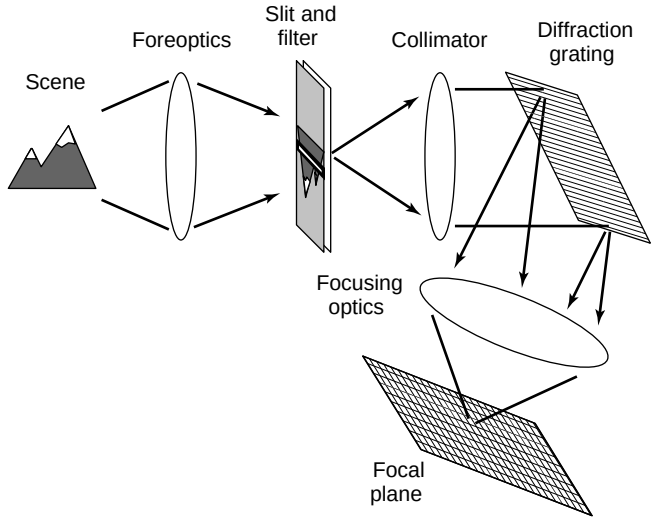

Figure 10 Schematic outline of a grating-based HSI. A scene is focused on the sensor's entrance slit. A filter passes a selected range of wavelengths, which are collimated and incident on a diffraction grating. The axis of diffraction is selected perpendicular to the orientation of the entrance slit. The diffracted slit image is focused on a focal plane: one axis corresponds to spatial position on the slit, the other axis is spectral. As the slit is scanned across the scene, multiple image frames are stacked to form a data cube.

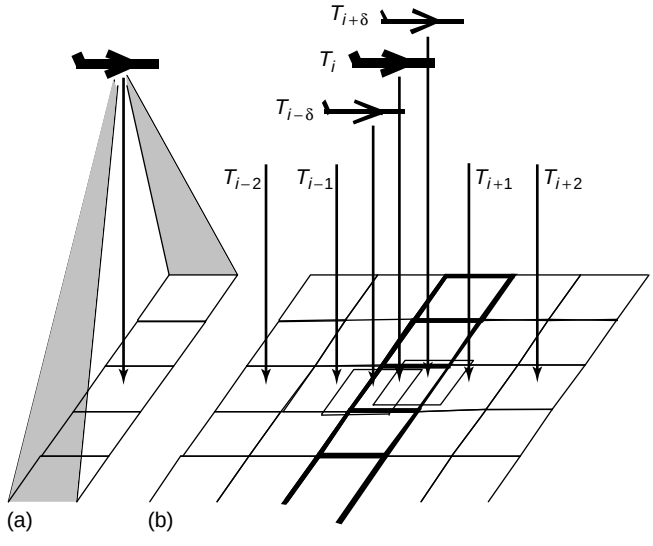

Figure 11 "Pushbroom" HSI data collection. The instantaneous FOV of a nadir-oriented dispersive HSI collects spectral profiles for a spatial strip of pixels (a). The light collected over the time interval $T_{i-1/2}$ to $T_{i+1/2}$ constitutes a single frame of hyperspectral data. Multiple successive frames are concatenated to form a data cube (b). Note the position of the aircraft at three closely spaced times: observe that a single data frame contains photons from ground sample areas immediately behind ($T_{i-\delta}$) and in front of ($T_{i+\delta}$) the frame.

4.4 Livermore Imaging Fourier Transform Infrared Spectrometer: An Imaging Fourier Transform Spectrometer

The LIFTIRS is based on a Michelson interferometer fitted with an IR focal plane.[38] A simplified diagram of this type of IFTS is presented in Figure 12. IR light from the observed scene is collimated and split into two paths. One path is equipped with a moving mirror so that the length of its optical path, x, can be smoothly varied. Light from this path is recombined with light from the fixed-length path, forming an optical interference pattern. The pattern is a function of x, and is focused on a GaSi image plane. The focal plane image an array of instantaneous samples of interferograms – one for each pixel – that cannot be interpreted without additional data frames. The optical path length, x, is changed while successive frames are collected. At each spatial pixel the ordered set of interference amplitudes comprises the Fourier transform of the pixel's spectral profile. An FFT applied to the data set produces a spectrum for each pixel. The performance parameters of LIFTIRS are listed in Table 3.

The distance traveled by the moving mirror determines the spectral resolution. A bandpass optical filter limits the measured spectral range, and the shortest wavelength passed by this filter determines how many frames need to be collected. Unlike filter- or dispersion-based HSI, an IFTS has adjustable spectral resolution. Scanning the mirror over a large span resolves fine spectral lines, which is desirable for detecting and discriminating the molecular vibrational and rotational transitions of gases. This is of particular use in the LWIR regime where many gaseous industrial by-products have spectral activity. In practice, sample noise increases with smaller resolved bandwidths, so the best signal-to-noise ratios are achieved when the IFTS bandwidth matches the line widths of gases being classified and discriminated.[34]

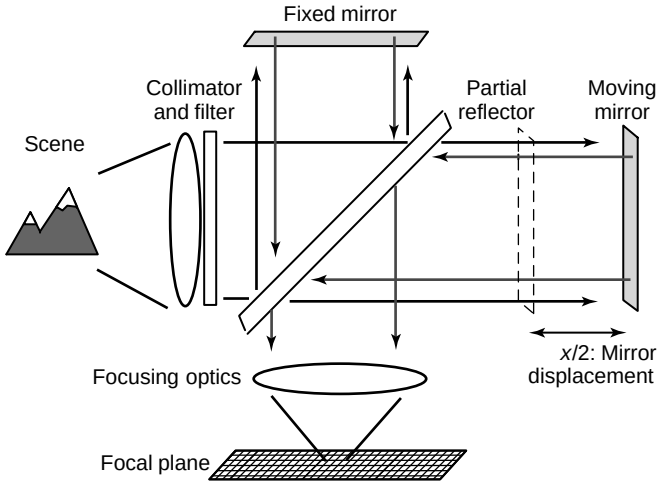

Figure 12 Schematic outline of an IFTS. This instrument is based on a Michelson interferometer. Light rays from the scene are collimated and filtered before being split onto two paths by the partial reflector (heavy arrows). The light on both paths are reflected by mirrors, recombined by the partial reflector – producing optical interference – and focused on the focal plane (light arrows). The mirror on one path is movable in order to change the optical path length, which alters the optical interference pattern. As mirror displacement changes, successive frames build up the Fourier transform of each pixel's spectral profile.

Table 3 LIFTIRS operating parameters. (Reproduced by permission of SPIE from M.R. Carter et al., *Proc. SPIE*, Vol. 2480, 380–386, 1995.[38])

Parameter	LWIR	MWIR
Detector	Ga : Si	InSb
Spectral range (μm)	8–12.5	3.3–4.9
Image format	128 × 128	256 × 256
Pixel size (μm)	50	30
Instantaneous FOV (mrad)	0.55	0.35
Quantum efficiency (%)	4	92
Beam splitter	KBr/Ge	KBr/Ge
Frame rate (maximum) (Hz)	1000	125
Spectral resolution	Variable to 0.25 cm^{-1}	Variable to 0.25 cm^{-1}

A weakness of the IFTS is its sensitivity to random or periodic temporal fluctuations of pixel spectra.[32] If these fluctuations fall within the bandwidth of the instrument, false spectral components are introduced while true components are destroyed. As an additional problem, an airborne IFTS requires very agile mounting systems to compensate for roll and pitch motion of the aircraft. Yaw motions (i.e. direction shifts) present a more difficult problem, introducing rotation within the image as data are collected. If successive image pixels do not correspond to the same point on the ground, spectral corruption is unavoidable.

Figure 13 illustrates the potential of the IFTS for gaseous pollution monitoring. A mixture of SF_6 and NH_3 was released from a stack at a test facility.[34] The presence of both gases is detected using their characteristic absorption frequencies. Later, after the stack is purged for 15 min, the flow of SF_6 is terminated, but to the surprise of the observers, substantial amounts of residual NH_3 continue to be detected in the stack plume.

5 DATA CORRECTION

5.1 Types of Correction

Several data correction steps are required to properly capture the spectral content of a scene. Some of these corrections compensate for solar and atmospheric

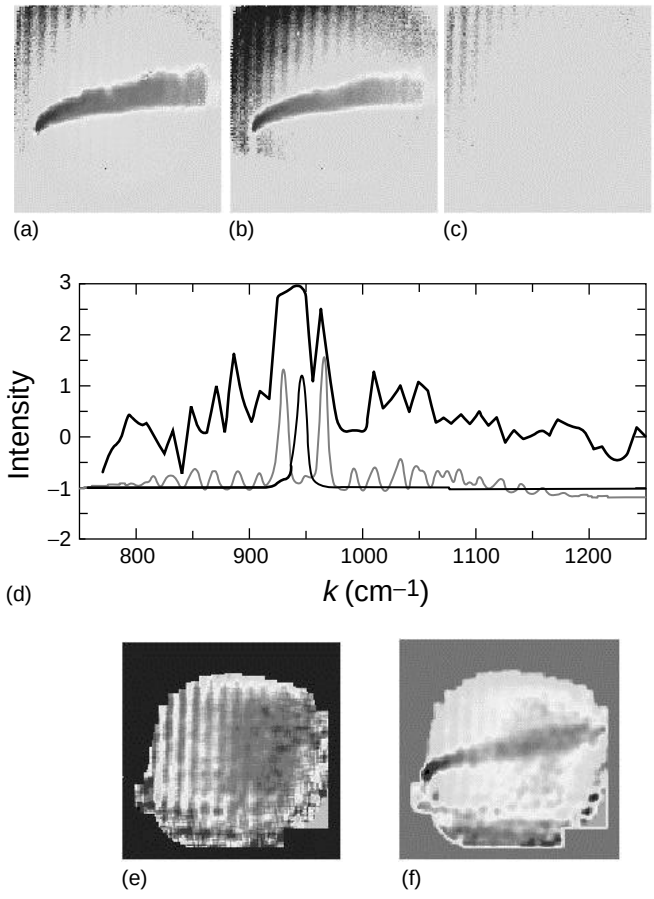

Figure 13 LIFTIRS ground level observation of a gas plume. Two gases, (a) SF_6 and (b) NH_3, are released from a stack and easily detected at their characteristic absorption frequencies (946 and 966 cm^{-1}, respectively). The 980 cm^{-1} sky background (c) is relatively uniform. (d) The measured spectral profile of the plume is the dark tracing, while reference spectra of SF_6 (946 cm^{-1} peak) and NH_3 (928 and 966 cm^{-1} peaks) are shown as lighter tracings. After a 15-min stack purge, no trace of SF_6 is detected (e), but the LIFTIRS team was surprised to find a persisting signature of NH_3 (f). This unexpected observation of residual NH_3 shows the power of HSI for monitoring gaseous pollutants. (Reproduced by permission of SPIE from C.L. Bennett et al., *Proc. SPIE*, Vol. 2480, 335–444, 1995.[34])

effects while others handle the optical and electrical characteristics of an HSI sensor.

Let us follow a solar photon on its way from the sun to its information impact in an HSI data cube (Figure 14). The solar spectrum is essentially that of a blackbody at 5800 K with embedded absorption lines. Passing into the atmosphere, solar radiation is both absorbed and scattered. Gases such as oxygen, ozone and carbon dioxide are strong absorbers, but water vapor is by far the most significant gas affecting solar illumination at ground level. Light interacting with air molecules causes Rayleigh scattering, especially towards the blue end of the spectrum. Aerosols and particulate matter in the atmosphere produce Mie scattering, which acts over a wider spectral range. Clouds attenuate or completely block sunlight from illuminating a patch of ground. In addition to direct sunlight, skylight (i.e. scattered and refracted sunlight) and light reflected off nearby landscape also contribute illumination at ground level. At mid-IR wavelengths, hot objects such as running engines or motor exhaust, contribute photons to the scene. At long-IR wavelengths thermal emissions from ambient temperature objects and landscape also contribute.[1]

At ground level the elements in the scene selectively absorb or reflect incident photons. Many natural materials have a substantial Lambertian component (Lambertian surfaces have no preferred direction of reflection), and have low to mid-range reflection coefficients (0 to 20%). Some materials such as snow, ice and water, as well as many man-made materials like metal and glass, possess substantial specular reflection properties; that is, they behave somewhat like mirrors. At certain angles light glints from strong specular reflectors can saturate a sensor designed for viewing Lambertian materials.

The reflected spectra of surface features pass through a span of atmosphere to the HSI sensor. Along the way, atmospheric absorption and scattering again contribute their effects. Unwanted light is scattered into the HSI's optical path from the sun, skylight and light reflected from adjacent landscape features. Finally, thermal inhomogeneities in the atmosphere causes subtle lensing perturbations that disrupt the wavefront of reflected light.

Light reaching the sensor is absorbed, reflected or scattered to some extent by optical elements. These elements also introduce geometric and chromatic aberrations, and possibly image ghosts. The diffraction limit of the optical device constrains how sharply the image can be focused on the photo-active plane. Further, shot noise is produced in the detector due to statistical fluctuations in photon flux.

The focal plane has opto-electrical characteristics that affect data collection: frequency-dependent quantum conversion efficiency, inactive ''dead'' pixels, ''hot'' or fluctuating pixels, individual pixel gains and offsets, electrical echoes, readout noise, limited dynamic range, quantization noise, $1/f$ noise (i.e. drift), and dark current. For LWIR sensors, uncooled optical elements of the sensor and enclosure components emit radiation, contributing to the measured signal. A thorough pre-collection calibration regimen is essential to characterize and minimize the effects of imperfect optics and electronics. During data collection itself periodic measures of the focal plane's radiometric gains and offsets are needed to handle the inescapable drift of sensor responsivity and dark current. Further calibration is needed to check the HSI's optical path relative to the platform's pointing

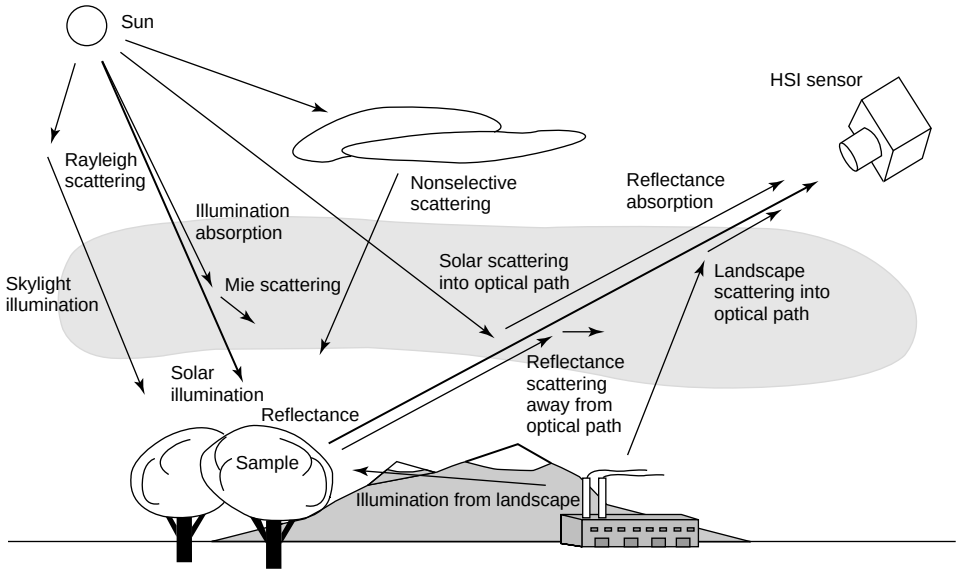

Figure 14 Atmospheric and landscape reflection effects on hyperspectral remote sensing.

direction and to properly assign spectral bands to specific focal plane rows. Airborne instruments without fully stabilized mounts are subject to unavoidable roll, pitch and yaw motions that cause image distortion. In regard to data fusion tasks, if the sensor is not pointing straight down (i.e. in the ''nadir'' viewing direction), the oblique viewing angle will introduce geometric distortions and image re-sampling will be required to correct the image.

Corrections for these effects can be accomplished with varying degrees of success. The most difficult tasks in remote sensing are characterizing the incident light at the surface, and properly correcting for the atmospheric effects of absorption and scattering. While these can have huge effects on spectral analysis algorithms, we are becoming increasingly adept at handling these issues, and are able to calculate the corrections at an acceptable rate using present day workstations.[39,40]

5.2 Computing a Reflectance Data Cube

Spectral reflectance, the ratio of reflected to illuminating radiance, is an intrinsic property of materials, which makes it an attractive feature on which to base classification and quantification algorithms. Most natural objects have a slightly glossy surface. The specular characteristics of their surfaces are apparent for a narrow range of viewing angles relative to the illuminating source. For the purposes of remote sensing, the specular aspects of most materials can be set aside because observing conditions are usually chosen to minimize their contribution.

Calculating surface reflectance by remote observation is a challenging problem. Figure 15 shows the process of generating a reflectance data cube from raw data. The difficulties arise primarily from two sources: sensor drift and atmospheric effects. For sensors the ''offset'' (also called ''dark current'' or ''background count'') tends to drift over time as $1/f$ noise. Generally, the longer the wavelength being monitored, the more this drift affects the measurement. In order to calculate a radiometrically correct data cube, a ''dark field'' (i.e. sensor frames recorded with the shutter closed) is recorded immediately before or after data collection.

Atmospheric effects are the most difficult aspect of calculating reflectance data cubes because many required parameters are difficult to obtain. Figure 14 illustrates the complicated phenomenology that is involved. Water vapor is a strong atmospheric absorber and varies in concentration laterally over tens of meters (GSD). Ozone concentration also varies substantially in the lower atmosphere near cities. Sophisticated atmospheric modeling programs such as MODTRAN can be used for calculating atmospheric optical behavior once key parameters are available. Extracting these parameters directly from spectral data is a goal of current algorithmic development. For example, ATREM (Atmospheric Removal Program)[40] uses the absorption ratios of water bands to estimate atmospheric humidity on a pixel by pixel basis. Other approaches also include scattering phenomena and use recursive data smoothing and eigenvalue analysis.[39]

Calculating a reflectance data cube begins by measuring the sensor gain profile for each pixel in the detector array (or at least verifying the profile obtained in laboratory tests). An on-board calibration source is essential. Both a dark field (D in Figure 15) and calibration ''white field'' (C) are collected. The radiometric power of the calibration source, P_{cal}, is accurately known from

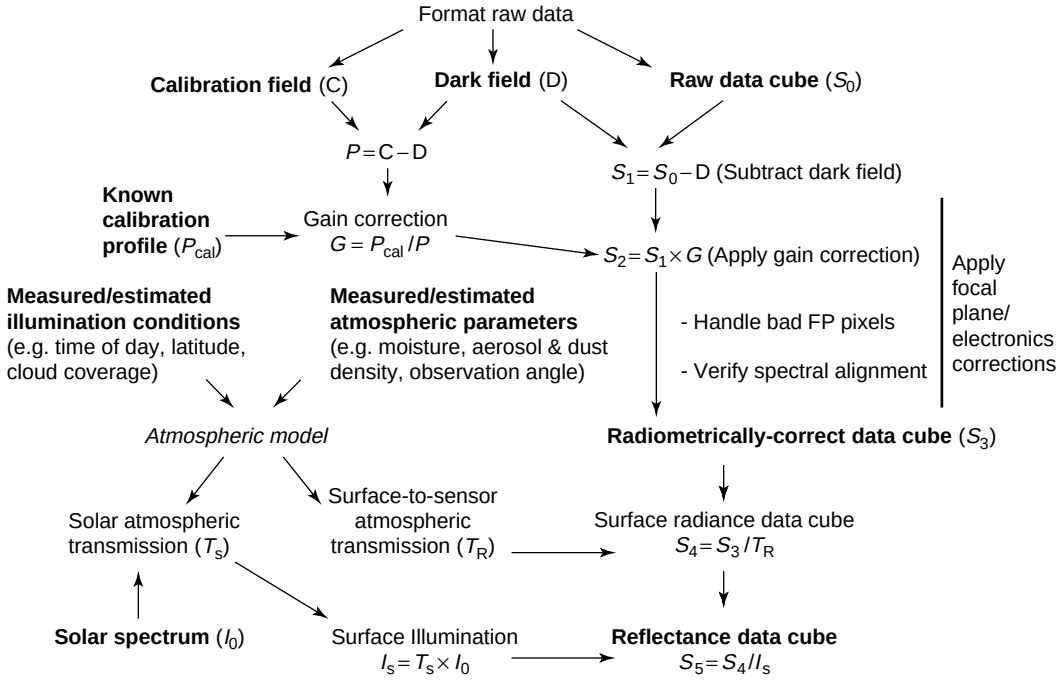

Figure 15 Process of calculating a reflectance data cube from raw data.

laboratory measurements. The gain coefficient for each focal plane pixel, G – the conversion factor from focal plane electron counts to radiometric quantification – is calculated using Equation (1) (see Figure 15):

$$G = \frac{P_{\text{cal}}}{\text{C} - \text{D}} \tag{1}$$

This value of G for a given pixel will vary with wavelength, reflecting the frequency dependence of the detector's quantum efficiency. Other factors affect G, such as inhomogeneities in the focal plane's fabrication and its readout electronics. Unlike dark current, G should not change rapidly over time.

Converting a raw data cube to a reflectance data cube proceeds as outlined in Figure 15. The associated dark field offset, D, is subtracted from the raw data cube, S_0, and gain correction is applied. Bad focal plane pixels are noted, and the spectral alignment is confirmed by noting the pixel positions associated with key atmospheric absorption lines. This produces a radiometrically correct data cube.

The next step is to remove atmospheric perturbations. An atmospheric model is used to calculate the transmission profile of the optical path between the surface and the HSI sensor, T_R. Similarly, the ground-level solar illumination spectrum, $(T_S I_0)$, is computed based on our best knowledge of atmospheric parameters. With these transmission profiles in hand, the reflectance data cube, S_5, is calculated using Equation (2):

$$S_5 = \frac{(S_0 - \text{D})G}{T_R T_S I_0} \tag{2}$$

6 SPECTRAL SCENE ANALYSIS

Hyperspectral image analysis is often concentrated on the problem of classification, i.e. detection and identification of a particular target or surface cover type. However, HSI allows for more than simple classification of image pixels. Analysis techniques have been developed that produce quantitative estimates of material abundance in a hyperspectral scene. Because such a wide variety of applications are addressed with this technology, hyperspectral analysis methods also vary widely and include techniques for spectral matching and feature fitting, linear and nonlinear unmixing, and structural modeling.

In general, a hyperspectral data cube is an image of some portion of the earth's surface, in which a spectral reflectance curve is associated with each pixel in the image. In some instances, a pixel is homogeneous in the sense that the area of ground represented by that pixel contains only one type of material. However, most natural surfaces do not consist of a single uniform material, and a pixel can cover a spatial area containing different materials, each with its own individual spectral signature. In this case, the pixel will not contain a pure spectrum but a mixture of individual component

spectra, and the pixel is termed mixed. The spectrally unique signatures that combine to produce the statistically significant variations and mixtures in a hyperspectral image are called endmembers. Many hyperspectral image processing algorithms address the problems of accurate image endmember determination, separation of endmembers from background spectra, and estimation of endmember abundance.

6.1 Spectral Matching

A spectral angle mapper (SAM) is used to correlate measured spectra with a library of endmember spectra. The method can be used to detect targets in known or unknown backgrounds and to classify materials in an image. SAM treats spectra as vectors in N-dimensional space, where N is the number of spectral bands per pixel, and matches pixels to reference spectra by computing angles in N-dimensional space. Let $\mathbf{u}$ and $\mathbf{r}$ be vectors of length N representing an unknown and a reference spectrum, respectively. SAM determines the similarity of $\mathbf{u}$ to $\mathbf{r}$ through computation of the angle a between them, as shown in Equation (3):

$$a = \cos^{-1}\left(\frac{\mathbf{u}^T\mathbf{r}}{\|\mathbf{u}\|\|\mathbf{r}\|}\right) \tag{3}$$

If the two spectra are similar, the angle a between them will be small.

Using SAM, an image is classified by first calculating spectral angles between each image pixel and each reference spectrum, then assigning the pixel to the material class whose reference spectrum produces the best match, i.e. the smallest spectral angle. The reference spectra can come from a library of materials assumed to be in the scene or from the image itself. Figure 16 shows a hyperspectral image of an agricultural region near a river bend that has been classified using the SAM algorithm. In this case, material classes correspond to seven different land cover classes. The reference spectra were chosen from the image itself, and those pixels that did not generate small enough angles with the reference spectra were left unclassified (the black regions of the classified image). For this image, the SAM technique produced an accurate classification of river water depth as well as land cover types.

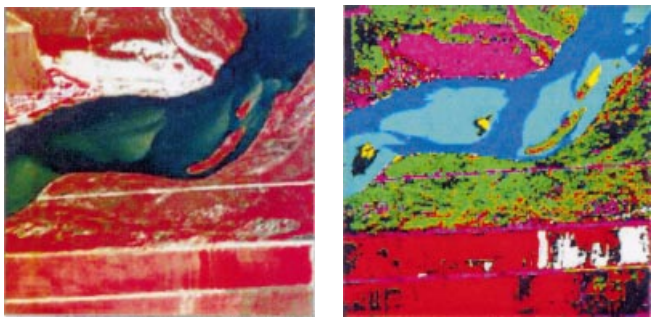

Figure 16 A hyperspectral image classified using the SAM algorithm. (Color key: red, white – vegetation types 1 and 2; green – scrub; purple – bare soil; yellow – river sand; light blue – shallow river water; dark blue – deeper river water; black – unclassified pixels.)

Because SAM depends on angles between spectral vectors and not the lengths of these vectors, the algorithm is insensitive to variations in illumination across a scene. This technique is useful for classification of general land cover types and even lends itself well to applications that require image analysis in real-time. However, more sophisticated processing is necessary to estimate material abundance and to deal with issues such as mixed pixels.

Another technique that is used to correlate a signature of interest with measured data is spectral matched filtering. This method is particularly useful for target detection in hyperspectral images and is based on well-known techniques in signal detection.[41] Let $\mathbf{d}$ be a vector representing the target signature of interest and $\mathbf{p}$ an image pixel. The matched filter output, which is the vector dot product $\mathbf{d}^T\mathbf{p}$, gives a measure of how well the image pixel matches the desired signature. This technique has been extended to deal with the problem of detecting small targets in highly-structured backgrounds.[42] In this case, an optimum matched filter is developed that takes into account the spatial as well as spectral variability of the image data and maximizes the probability of detecting a target at a particular false alarm rate. Adaptive matched filtering algorithms can be shown to achieve near-optimum performance for target detection without precise knowledge of target or background spectra.[43,44]

6.2 Spectral Unmixing

Spectral unmixing techniques provide quantitative abundance estimates of individual component spectra in a hyperspectral scene.[45–50] Unmixing is also used for detection of targets of interest, particularly at the sub-pixel level, and for removal of undesired signature components from hyperspectral measurements.[51,52] The simplest model of a mixed pixel is a linear one. Research into mixing scales has indicated that linear mixing occurs when each individual photon encounters only one type of material before being reflected.[53] Nonlinear mixing occurs when a single incident photon interacts with more than one material before being reflected.[54] For example, the microscopic (or intimate) mixtures that are typical of many geologic situations engender nonlinearly mixed pixels. Even for the case of a nonlinear mixture model, a change of variables can linearize the unmixing problem, yielding a first-order approximation of endmember abundance.

The linear mixture model is described as follows. Let $\mathbf{p}_i$ be an $l \times 1$ vector denoting the ith pixel in a hyperspectral image, where l is the number of spectral bands, and suppose there are k endmembers of interest in the scene. That is, there are k different classes of materials represented in the hyperspectral image. We will let $\mathbf{M}$ denote the $l \times k$ matrix whose columns contain the endmember signatures of interest, and $\mathbf{a}_i = (a_{i1} a_{i2} \cdot a_{ik})^T$ be a $k \times 1$ abundance vector, where a_{ij} denotes the fraction of the jth signature in the pixel $\mathbf{p}_i$. The mixed pixel is then modeled as shown in Equation (4):

$$\mathbf{p}_i = \mathbf{M}\mathbf{a}_i + \mathbf{n} \tag{4}$$

where $\mathbf{n}$ is an $l \times 1$ vector representing noise which we will assume here to be Gaussian with zero mean and covariance $\sigma^2 I$ (I is the $l \times l$ identity matrix).

The unmixing problem is to estimate the spectral abundance vector $\mathbf{a}_i$ given the observation $\mathbf{p}_i$. For the model described above, the maximum likelihood estimate is also the least squares estimate of $\mathbf{a}_i$, and is given by Equation (5):

$$\hat{\mathbf{a}}_i = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{p}_i \tag{5}$$

provided there are no constraints placed on the components of $\hat{\mathbf{a}}_i$. Therefore, $\hat{\mathbf{a}}_i$ is an unconstrained estimate of the fractional abundance of each endmember in pixel $\mathbf{p}_i$. Figure 17 shows an example of linear spectral unmixing for a hyperspectral image of the Cuprite, Nevada region. The colors correspond to the fractional abundance of six endmembers for each individual pixel.

The accuracy of an abundance estimate derived by spectral unmixing depends directly on the chosen matrix of endmembers, which come from the image itself, from libraries of laboratory or field measurements, or from estimates of target and background signatures. A number of automatic and manual endmember selection techniques have been developed including construction of endmembers from the dominant eigenvectors of a principal component analysis of the image, or formulation of the endmember selection process as a problem in convex geometry.[47,55,56] Automated endmember selection methods yield rapid, repeatable processing of large volumes of data and have been implemented for real-time analysis of hyperspectral imagery.[57] Manual techniques are useful for interactive exploration of the high-dimensional data space spanned by the spectra in a hyperspectral cube.[58] Each technique for endmember determination has its own advantages and drawbacks, and the choice of method depends strongly on the imaging application.

The analysis techniques described above are representative of the core methods that have been developed for hyperspectral image classification and material identification. Since the applications for hyperspectral technology are wide ranging and this field of study has become increasingly active in recent years, there are of course many aspects of this work that have not been described here. For example, there have been significant advances in the areas of vegetation structural modeling, foliar chemistry analysis from hyperspectral images, hyperspectral data compression and real-time processing. The reader is encouraged to explore the references for details of this research.

Figure 17 Derived endmember abundance map of Cuprite, Nevada. Unmixing was performed using six endmembers selected by the automated N-FINDR technique. (This figure was provided by Ed Winter of Technical Research Associates, Inc.)

7 FUTURE DEVELOPMENTS

The field of HSI, including instrumentation and analysis, has grown significantly in the last decade. Recent developments in focal plane array technology, high-speed data acquisition and processing hardware, and increasingly accurate analysis techniques have enabled the growth of HSI as a quantitative remote sensing tool. Although the technology has advanced greatly, future developments could result in increased utility of HSI. For example, larger focal plane arrays and faster data acquisition hardware could allow for even wider area coverage with high spatial, as well

as spectral, resolution. Another aspect of the technology that will soon see advancements is HSI of the earth from space. At the time of writing, NASA is preparing a hyperspectral instrument payload for launch within 1 year. This first hyperspectral look at the earth from space will generate exciting new developments in this quickly evolving and expanding branch of remote sensing.

ACKNOWLEDGMENTS

The authors wish to thank Lane Darnton, Mark Folkman, John Brock and Richard Leahy for their careful review of this paper and their helpful suggestions for its improvement.

ABBREVIATIONS AND ACRONYMS

AHI	Airborne Hyperspectral Imager
AIS	Airborne Imaging Spectrometer
ATREM	Atmospheric Removal Program
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
CCD	Charge-coupled Device
CHRISS	Compact High-resolution Imaging Spectrographic System
FFT	Fast Fourier Transform
FOV	Field of View
GSD	Ground Sample Distance
HRIS	High-resolution Imaging Spectrometer
HSI	Hyperspectral Imaging
HYDICE	Hyperspectral Digital Imagery Collection Experiment
IFTS	Imaging Fourier Transform Spectrometer
IR	Infrared
JPL	Jet Propulsion Laboratory
LIFTIRS	Livermore Imaging Fourier Transform Infrared Spectrometer
LWIR	Long-wave Infrared Region
MWIR	Medium-wave Infrared
NASA	National Aeronautics and Space Administration
SAM	Spectral Angle Mapper
SEBASS	Spatially Enhanced Broadbard Array Spectrograph System
SMIFTS	Spatially Modulated Imaging Fourier Transform Spectrometer
SPOT	Système Pour l'Observation de la Terre
SWIR	Short-wave Infrared
TRWIS	TRW Imaging Spectrometer
VNIR	Visible/Near-infrared

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Automotive Emissions Analysis with Spectroscopic Techniques • Differential Optical Absorption Spectroscopy, Air Monitoring by • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Matrix Isolation Spectroscopy in Atmospheric Chemistry

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Dyes, Environmental Analysis of • Explosives Analysis in the Environment • Industrial Waste Dumps, Sampling and Analysis • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Luminescence in Environmental Analysis • Soil Sampling for the Characterization of Hazardous Waste Sites • Water Analysis: Organic Carbon Determinations

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction • Aircraft-based Flux Sampling Strategies • Chemical-sensing Networks: Satellite-based • Field-based Analysis of Organic Vapors in Air

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Biological Oceanography by Remote Sensing • Imaging Spectrometry for Geological Applications • Land Cover Assessment and Monitoring • Polar Environments Assessment by Remote Sensing • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring • Sea Ice Monitoring by Remote Sensing • Semiarid Land Assessment: Monitoring Dry Ecosystems with Remote Sensing • Temperate Forest Resource Assessment by Remote Sensing • Tropical Forest Resource Assessment by Remote Sensing

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared • Spectral Databases, Infrared • Theory of Infrared Spectroscopy • Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

General Articles **(Volume 15)**
Multivariate Image Analysis • Quantitative Spectroscopic Calibration

REFERENCES

1. R. Anderson, W. Malila, R. Maxwell, L. Reed, *Military Utility of Multispectral and Hyperspectral Sensors*, Infrared Information Analysis Center, Environmental Research Institute of Michigan, Ann Arbor, 1994.
2. J.-P. Muller (ed.), *Digital Image Processing in Remote Sensing*, Taylor & Francis, London, 1988.
3. D. Landgrebe, 'Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data', in *Information Processing for Remote Sensing*, ed. C.H. Chen, World Scientific Publishing Co., NJ, 1999.
4. T.M. Lillesand, R.R. Kiefer, *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc., New York, 1994.
5. J.A. Richards, *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, 1993.
6. R.M. Hoffer, 'Biological and Physical Considerations in Applying Computer-aided Analysis Techniques to Remote Sensor Data', in *Remote Sensing: The Quantitative Approach*, eds. P.H. Swain, S.M. Davis, McGraw-Hill, New York, 1978.
7. W.L. Wolfe, *Introduction to Imaging Spectrometers*, SPIE Optical Engineering Press, Bellingham, WA, 1997.
8. 'HIRIS, High-resolution Imaging Spectrometer: Science Opportunities for the 1990s', Earth Observing System, Instrument Panel Report, NASA, Vol. IIc, 1987.
9. D.L. Peterson, J.D. Aber, P.A. Matson, D.H. Card, N. Swanberg, C.A. Wessman, M. Spanner, 'Remote Sensing of Forest Canopy and Leaf Biochemical Contents', *Rem. Sens. Environ.*, **24**, 85–108 (1988).
10. B.-C. Gao, A.F.H. Goetz, 'A Linear Spectral Matching Technique for Retrieving Equivalent Water Thickness and Biochemical Constituents of Green Vegetation', in *Proceedings of the Third Airborne Annual JPL Geoscience Workshop (AVIRIS, TIMS, and AIRSAR)*, Jet Propulsion Laboratory, Pasadena, CA, 35–37, 1992.
11. Z. Lee, K.L. Carder, S.K. Hawes, R.G. Steward, T.G. Peacock, C.O. Davis, 'Model for the Interpretation of Hyperspectral Remote-sensing Reflectance', *Appl. Opt.*, **33**, 5721–5732 (1994).
12. R.N. Clark, *Reflectance Spectra*, AGU Handbook of Physical Constants, American Geophysical Union, Washington, DC, 178–188, 1995.
13. R.N. Clark, G.A. Swayze, A. Gallagher, 'Mapping Minerals with Imaging Spectroscopy', U.S. Geological Survey, Office of Mineral Resources Bulletin 2039, 1993.
14. R.N. Clark, G.A. Swayze, A. Gallagher, N. Gorelick, F. Kruse, 'Mapping with Imaging Spectrometer Data using the Complete Band Shape Least-squares Algorithm Simultaneously Fit to Multiple Spectral Features from Multiple Materials', in *Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, JPL Publication 91-28, Jet Propulsion Laboratory, Pasadena, CA, 2–3, 1991.
15. J.W. Boardman, A.F.H. Goetz, 'Sedimentary Facies Analysis using AVIRIS Data: A Geophysical Inverse Problem', in *Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, JPL Publication 91-28, Jet Propulsion Laboratory, Pasadena, CA, 4–13, 1991.
16. J.W. Boardman, J.F. Huntington, 'Mineralogic and Geochemical Mapping at Virginia City, Nevada using 1995 AVIRIS Data', in *Proceedings of the Twelfth International Conference on Applied Geologic Remote Sensing*, ERIM International Inc., I-191–I-198, 1997.
17. K.L. Carder, R.G. Steward, R.F. Chen, S. Hawes, Z. Lee, C.O. Davis, 'AVIRIS Calibration and Application in Coastal Oceanic Environments: Tracers of Soluble and Particulate Constituents of the Tampa Bay Coastal Plume', *Photogramm. Eng. Remote Sens.*, **59**, 339–344 (1993).
18. L.L. Richardson, V.G. Ambrosia, 'Remote Sensing of Algal Pigments to Determine Coastal Phytoplankton Dynamics in Florida Bay', in *Proceedings of the Fourth International Conference on Remote Sensing for Marine and Coastal Environments*, ERIM, Ann Arbor, MI, I-75–I-81, 1997.
19. T.J. Malthus, L. Grieve, M.D. Harwar, 'Spectral Modeling for the Identification and Quantification of Algal Blooms: A Test of Approach', in *Proceedings of the Fourth International Conference on Remote Sensing for Marine and Coastal Environments*, ERIM, Ann Arbor, MI, I-223–I-232, 1997.
20. M.E. Martin, J.D. Aber, 'Measurements of Canopy Chemistry with 1992 AVIRIS Data at Blackhawk Island and Harvard Forest', in *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, Jet Propulsion Laboratory, Pasadena, CA, 113–116, 1993.
21. C.A. Wessman, J.D. Aber, D.L. Peterson, J.M. Melillo, 'Remote Sensing of Canopy Chemistry and Nitrogen Cycling in Temperate Forest Ecosystems', *Nature*, **335**, 154–156 (1988).
22. G.P. Asner, 'Biophysical and Biochemical Sources of Variability in Canopy Reflectance', *Remote Sens. Environ.*, **64**, 234–253 (1998).

23. V.C. LaCapra, J.M. Melack, M. Gastil, D. Valeriano, 'Remote Sensing of Foliar Chemistry of Inundated Rice with Imaging Spectrometry', *Remote Sens. Environ.*, **55**, 50–58 (1996).
24. M.E. Martin, J.D. Aber, 'Estimation of Forest Canopy Lignin and Nitrogen Concentration and Ecosystem Processes by High Spectral Resolution Remote Sensing', *Ecol. Appl.*, **7**, 431–443 (1997).
25. S.L. Ustin, D.A. Roberts, J. Pinzon, S. Jacquemoud, M. Gardner, G. Scheer, C.M. Castaneda, A. Palacios, 'Estimating Canopy Water Content of Chaparral Shrubs using Optical Methods', *Remote Sens. Environ.*, **65**, 280–291 (1998).
26. C.J. Johannsen, 'Farming by the Inch', in *The World and I*, The Washington Times, Washington, DC, 148–155, 1996.
27. R.C. Olsen, S. Bergman, R.G. Resmini, 'Target Detection in a Forest Environment using Spectral Imagery', *Proc. SPIE*, **3118**, 46–56 (1997).
28. J. Bowles, J. Antoniades, M. Baumback, J. Grossmann, D. Haas, P. Palmadesso, J. Stracka, 'Real Time Analysis of Hyperspectral Data Sets using NRL's ORASIS Algorithm', *Proc. SPIE*, **3118**, 38–45 (1997).
29. R.D. McGregor, M.H. Nguyen, I.S. Robinson, A. Weisberg, J.X. Tull, S. Augustin, R. Crow, 'Target Detection using HSI Systems: An Update of TRW Results', *Proc. SPIE*, **3438**, 52–65 (1998).
30. A.T. DePersia, A.P. Bowman, A.L. Giles, E.M. Winter, F. Badik, M. Schlangen, P. Lucey, T. Williams, J. Johnson, J. Hinrichs, K. Horton, G. Allen, A. Stocker, A. Oshagan, B. Schaff, B. Kendall, M.R. Carter, C.L. Bennett, D.J. Fields, 'ARPA's Hyperspectral Mine Detection Program', in *Proceedings of the International Symposium on Spectral Sensing Research*, Melbourne, 1995.
31. P.G. Lucey, T. Williams, M. Mignard, J. Julian, D. Kokobun, G. Allen, D. Hampton, W. Schaff, M.J. Schlangen, E.M. Winter, W.B. Kendall, A.D. Stocker, K.A. Horton, A.P. Bowman, 'AHI, an Airborne Long-wave Hyperspectral Imager', *Proc. SPIE*, **3431**, 36–43 (1998).
32. M.R. Carter, C.L. Bennett, D.J. Fields, J. Hernandez, 'Gaseous Effluent Monitoring and Identification using an Imaging Fourier Transform Spectrometer', *Proc. SPIE*, **2092**, 16–26 (1993).
33. A.E. Potter, Jr, 'Multispectral Imaging System', US Patent No. 3702735, Nov. 14, 1972.
34. C.L. Bennett, M.R. Carter, D.J. Fields, F.D. Lee, 'Infrared Hyperspectral Imaging Results from Vapor Plume Experiments', *Proc. SPIE*, **2480**, 435–444 (1995).
35. C.W. Wells, A.E. Potter, T.H. Morgan, 'Near-infrared Spectral Imaging Michelson Interferometer for Astronomical Applications', *Proc. SPIE*, **226** (1980).
36. S. Sandor-Leahy, D. Beiso, M. Figueroa, M. Folkman, D. Gleichauf, T. Hedman, P. Jarecke, S. Thordarson, 'The TRWIS III Hyperspectral Imager: Instrument Performance and Remote Sensing Applications', *Proc. SPIE*, **3438**, 13–22 (1998).
37. P.G. Lucey, T. Williams, M. Mignard, J. Julian, D. Kokobun, G. Allen, D. Hampton, W. Schaff, M.J. Schlangen, E.M. Winter, W.B. Kendall, A.D. Stocker, K.A. Horton, A.P. Bowman, 'AHI, an Airborne Long-wave Hyperspectral Imager', *Proc. SPIE*, **3431**, 36–43 (1998).
38. M.R. Carter, C.L. Bennett, D.J. Fields, D. Lee, 'Livermore Imaging Fourier Transform Infrared Spectrometer (LIFTIRS)', *Proc. SPIE*, **2480**, 380–386 (1995).
39. A.F.H. Goetz, J.W. Boardman, B. Kindel, K.B. Heidebrecht, 'Atmospheric Corrections: On Deriving Surface Reflectance from Hyperspectral Imagers', *Proc. SPIE*, **3118**, 14–22 (1997).
40. B.-C. Gao, C.O. Davis, 'Development of a Line-by-line-based Atmosphere Removal Algorithm for Airborne and Spaceborne Imaging Spectrometers', *Proc. SPIE*, **3118**, 132–141 (1997).
41. H.L. Van Trees, *Detection, Estimation and Modulation Theory*, Wiley & Sons, New York, 1968.
42. J.Y. Chen, I.S. Reed, 'A Detection Algorithm for Optical Targets in Clutter', *IEEE Trans. Aerospace Electronic Systems*, **AES-23**, 46–59 (1987).
43. A.D. Stocker, I.S. Reed, X. Yu, 'Multi-dimensional Signal Processing for Electro Optical Target Detection', *Proc. SPIE*, **1305**, 218–231 (1990).
44. X. Yu, I.S. Reed, A.D. Stocker, 'Comparative Performance Analysis of Adaptive Multispectral Detectors', *IEEE Trans. Signal Processing*, **41**, 2639–2656 (1993).
45. F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H. Goetz, 'The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data', *Remote Sens. Environ.*, **44**, 145–163 (1993).
46. J.W. Boardman, 'Inversion of Imaging Spectrometry Data using Singular Value Decomposition', in *12th Canadian Symposium on Remote Sensing*, IEEE, Canada, Vol. 4, 2069–2072, 1989.
47. J.W. Boardman, 'Automated Spectral Unmixing of AVIRIS Data using Convex Geometry Concepts', in *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop*, Jet Propulsion Laboratory, Pasadena, CA, 11–14, 1993.
48. F.A. Kruse, A.B. Lefkoff, J.B. Dietz, 'Expert System-based Mineral Mapping in Northern Death Valley, California/Nevada, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sens. Envrion.*, **44**, 309–336 (1993).
49. J.W. Boardman, F.A. Kruse, 'Automated Spectral Analysis: A Geologic Example Using AVIRIS Data, North Grapevine Mountains, Nevada', in *Proceedings of the Tenth Thematic Conference on Geologic Remote Sensing*, Environmental Research Institute of Michigan, Ann Arbor, MI, I-407–I-418, 1994.
50. J.W. Boardman, F.A. Kruse, R.O. Green, 'Mapping Target Signatures via Partial Unmixing of AVIRIS Data', in *Summaries of the Fifth Annual JPL Airborne Earth*

Science Workshop, Jet Propulsion Laboratory, Pasadena, CA, 23–26, 1995.

51. J. Harsanyi, C.-I. Chang, 'Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach', *IEEE Trans. Geosci. Remote Sensing*, **32**, 779–784 (1994).
52. C.-I. Chang, X. Zhao, M.L.G. Althouse, J.J. Pan, 'Least Squares Subspace Projection Approach to Mixed Pixel Classification for Hyperspectral Images', *IEEE Trans. Geosci. Remote Sensing*, **36**, 898–912 (1998).
53. R.B. Singer, T.B. McCord, 'Mars: Large Scale Mixing of Bright and Dark Surface Materials and Implications for Analysis of Spectral Reflectance', in *Proceedings of the Tenth Lunar and Planetary Science Conference*, 1835–1848, 1979.
54. R.B. Singer, 'Near-infrared Spectral Reflectance of Mineral Mixtures: Systematic Combinations of Pyroxenes, Olivine, and Iron Oxides', *J. Geophys. Res.*, **86**, 7967–7982 (1981).
55. M.O. Smith, P.E. Johnson, J.B. Adams, 'Quantitative Determination of Mineral Types from Reflectance Spectra using Principal Components Analysis', in *Proceedings of the Fifteenth Lunar and Planetary Science Conference*, 1985.
56. M.E. Winter, 'Fast Autonomous Spectral Endmember Determination in Hyperspectral Data', in *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, ERIM, Vancouver, BC, Canada, Vol. II, 337–344, 1999.
57. J. Bowles, M. Daniel, J. Grossmann, J. Antoniades, M. Baumback, P. Palmadesso, 'Comparison of Output from ORASIS and Pixel Purity Calculations', *Proc. SPIE*, **3438**, 148–156 (1998).
58. A. Bateson, B. Curtiss, 'A method for Manual Endmember Selection and Spectral Unmixing', *Remote Sens. Environ.*, **55**, 229–243 (1996).

Imaging Spectrometry for Geological Applications

Freek D. van der Meer
International Institute for Aerospace Surveys and Earth Sciences (ITC), Enschede, The Netherlands

1 Introduction and Historical Perspective 1
1.1 Objectives of Imaging Spectrometry 1
1.2 History of Imaging Spectrometry 3
2 Physics of Reflectance and Emittance Spectroscopy 3
2.1 Physics of Radiation 3
2.2 Reflectance Spectroscopy 4
3 Airborne Imaging Spectrometer Systems 7
3.1 Overview 7
3.2 Airborne Visible/Infrared Imaging Spectrometer 7
3.3 Digital Airborne Imaging Spectrometer (7915) 8
4 Future Planned Spaceborne Imaging Spectrometer Systems 8
5 Data Acquisition and Preprocessing of Imaging Spectrometer Data 9
5.1 Overview 9
5.2 Spectral Preprocessing Chain 10
5.3 Spatial Preprocessing Chain 12
5.4 Noise Adjustment 13
6 Radiance to Reflectance Correction 14
6.1 Overview 14
6.2 Relative Reflectance 15
6.3 Absolute Reflectance 16
6.4 In-flight Calibration 16
7 Thematic Analysis Techniques for Absorption Feature Extraction 17
7.1 Overview 17
7.2 Binary Encoding 17
7.3 Waveform Characterization 17
7.4 Spectral Feature Fitting 18
7.5 Spectral Angle Mapping 18
7.6 Spectral Unmixing 19
7.7 Constrained Energy Minimization 21
7.8 Classification 22
7.9 Cross-correlogram Spectral Matching 23
8 Extracting Emissivity from Thermal Infrared Data 23
9 Applications of Imaging Spectrometry 25
9.1 Overview 25
9.2 Mineral Mapping: Surface Mineralogy 25
9.3 Mineral Mapping: Exploration 25
9.4 Mineral Mapping: Lithology 27
9.5 Vegetation Stress and Geobotany 27
9.6 Environmental Geology 27
9.7 Petroleum-related Studies 28
9.8 Atmospheric Effects Resulting from Geological Processes 28
9.9 Volcanic Hazards 28
9.10 Thermal Infrared Studies 28
Acknowledgments 30
Abbreviations and Acronyms 30
Related Articles 31
References 31

Reflectance and emittance spectra have been used for many years to obtain compositional information about the Earth's surface. Electronic transition and charge transfer processes associated with transition-metal ions such as Fe, Ti, Cr, and so on produce diagnostic absorption features in the visible and near-infrared (VNIR) wavelength region of the spectra of minerals and rocks, while vibrational processes in H_2O and OH^- produce fundamental overtone absorptions in the shortwave infrared (SWIR) portion of the spectrum. Field and laboratory studies have demonstrated that reflectance spectroscopy can be used for mineral identification and further geological studies. The translation of these measurements to imaging data led to the development of the field of imaging spectrometry in remote sensing. Imaging spectrometers acquire images in a large number of narrow contiguous spectral bands to enable the extraction of reflectance spectra at a pixel scale that can be directly compared with similar spectra measured either in the field or in a laboratory. In this chapter, the use of imaging spectrometer data for geological applications is discussed by describing the processing chain of data analysis from the raw data acquisition to prospective applications. Several topics are addressed.

An appraisal of future spaceborne imaging spectrometer missions and potential prospective applications provides insight into this vastly growing field of research.

1 INTRODUCTION AND HISTORICAL PERSPECTIVE

1.1 Objectives of Imaging Spectrometry

When light interacts with a mineral or rock, light of certain wavelengths is preferentially absorbed while light of other

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

wavelengths is transmitted in the substance. Reflectance, defined as the ratio of the intensity of light reflected from a sample to the intensity of the light incident on it, is measured by reflection spectrophotometers which are composed of a light source and a prism to separate light into different wavelengths. The light beam interacts with the sample and the intensity of reflected light at various wavelengths relative to a reference standard of known reflectance is measured by a detector. Thus a continuous reflectance spectrum of the sample is obtained in the wavelength region measured.

Reflectance spectra have been used for many years to obtain compositional information about the Earth's surface. Similarly, it has been shown that spectral reflectance in the VNIR offers a rapid and inexpensive technique for determining the mineralogy of samples and obtaining information about chemical composition. Electronic-transition and charge-transfer processes (e.g. changes in the energy states of electrons bound to atoms or molecules) associated with transition-metal ions such as Fe, Ti, Cr, and so on, largely determine the position of diagnostic absorption features in the VNIR wavelength region of the spectra of minerals.[1–3] In addition, vibrational processes (e.g. small displacements of the atoms about their resting positions) in H_2O and OH^- produce fundamental overtone absorptions.[4,5] Electronic transitions produce broad absorption features that require higher energy levels than do vibrational processes and therefore take place at shorter wavelengths.[4,6] The position, shape, depth, and width of these absorption features are controlled by the particular crystal structure in which the absorbing species is contained and by the chemical structure of the mineral. Thus, variables characterizing absorption features can be directly related to the mineralogy of the sample.

Remote sensing (e.g. the observation of a target by a device separated from it by some distance and thus without physical contact) of the surface of the Earth from aircraft and from spacecraft provides information not easily acquired by surface observations. Until recently, the main limitation of remote sensing was that surface information lacked detail owing to the broad bandwidth of sensors available. Work on high-spectral-resolution radiometry has shown that earth surface mineralogy can be identified using spectral information from sensor data.[6] Conventional sensors (e.g. Landsat multispectral scanner (MSS) and thematic mapper (TM), and Système Probatoire de l'Observation de la Terre (SPOT)) acquire information in a few separate spectral bands of various widths (typically in the order of 0.1–0.2 μm), thus smoothing to a large extent the reflectance characteristics of the surface.[7] Most terrestrial materials are characterized by spectral absorption features typically 0.02–0.04 μm in width.[8] High-spectral-resolution remotely sensed images are acquired to produce reflectance or radiance spectra for each pixel in the scene. Imaging spectrometers acquire images in a large number (typically over 40) of narrow (typically 0.01–0.02 μm in width) contiguous (i.e. adjacent and not overlapping) spectral bands to enable the extraction of reflectance spectra at a pixel scale that can be directly compared with similar spectra measured either in the field or in a laboratory. Different names have been coined for this field of remote sensing including ''imaging spectrometry'', ''imaging spectroscopy'' and ''hyperspectral remote sensing''. Although they have a different meaning in the sense of a direct translation of the term (i.e. spectrometry for ''measuring'', spectroscopy for ''seeing'', hyperspectral for ''too many bands''), the significance and perception to the remote sensing community is the same: ''the acquisition of images in hundreds of registered, contiguous spectral bands such that for each picture element of an image it is possible to derive a complete reflectance spectrum''[9] (Figure 1).

The objective of imaging spectrometry is to measure quantitatively the components of the Earth system from calibrated spectra acquired as images for scientific research and applications. Thus it is of interest to measure physical quantities at the Earth surface such as upwelling radiance, emissivity, temperature, and reflectance. Based upon the molecular absorptions and constituent scattering characteristics expressed in the spectrum we seek to:

- detect and identify the surface and atmospheric constituents present;
- assess and measure the expressed constituent concentrations;

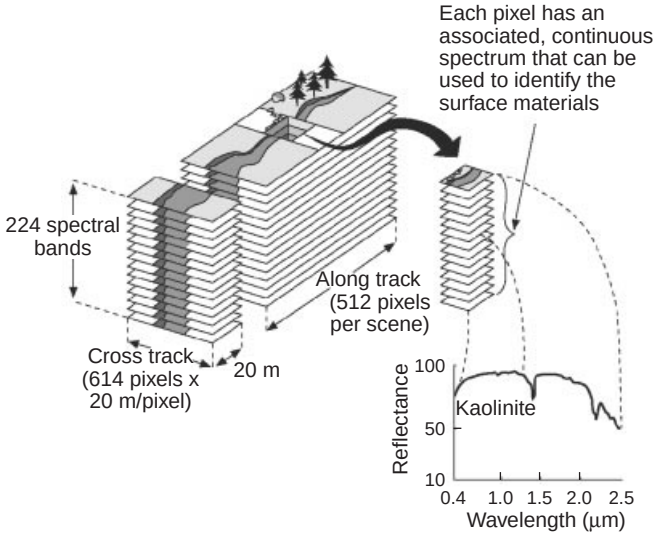

Figure 1 Concept of imaging spectrometry. (Reproduced by permission of ITC from S.M. de Jong, *ITC Journal*, **98**(1), 1–10 (1998).)

- assign proportions to constituents in mixed spatial elements;
- delineate spatial distribution of the constituents;
- monitor changes in constituents through periodic data acquisitions;
- simulate, calibrate, and intercompare sensors;
- validate, constrain, and improve models.

Through measurement of the solar reflected spectrum, a wide range of scientific research and applications are being pursued using signatures of energy, molecules, and scatterers in the spectra measured by imaging spectrometers. These fields include:

- atmosphere: water vapor, cloud properties, aerosols, absorbing gases;
- plant ecology: including chlorophyll, leaf water, lignin, cellulose, pigments, structure, vegetation species and community maps, nonphotosynthetic constituents;
- geology and soils: including mineralogy, soil type;
- coastal and inland waters: including chlorophyll, plankton, dissolved organics, sediments, bottom composition, bathymetry;
- snow and ice hydrology: including snow cover fraction, grainsize, impurities, melting;
- biomass burning: including subpixel temperatures and extent, smoke, combustion products;
- environmental hazards: including contaminants directly and indirectly, geological substrate;
- calibration: including aircraft and satellite sensors, sensor simulation, standard validation;
- modeling: including radiative transfer (RT) model validation and constraint;
- commercial: including mineral exploration, agriculture and forest status;
- algorithms: including autonomous atmospheric correction, advanced spectra derivation;
- other: human infrastructure, mine detection, and so on.

1.2 History of Imaging Spectrometry

The first civilian airborne spectrometer data were collected in 1981 using a one-dimensional profile spectrometer developed by the Geophysical Environmental Research (GER) Company which acquired data in 576 channels covering the 0.4–2.5 μm wavelength range[10] followed by the Shuttle Multispectral Infrared Radiometer (SMIRR) in 1981. The first imaging device was the Fluorescence Line Imager (FLI; also known as the Programmable Line Imager, PLI) developed by Canada's Department of Fisheries and Oceans (in 1981) followed by the Airborne Imaging Spectrometer (AIS), developed at the NASA (National Aeronautics and Space Administration) Jet Propulsion Laboratory which was operational from 1983 onward. This instrument acquired data in 128 spectral bands in the range of 1.2–2.4 μm with a field-of-view (FOV) of 3.7 degrees resulting in images of 32-pixels width.[11] A later version of the instrument, AIS-2, covered the 0.8–2.4 μm region acquiring images 64 pixels wide.[12] In 1987 NASA began operating the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS[13]). AVIRIS was developed as a facility that would routinely supply well-calibrated data for many different purposes. The AVIRIS scanner makes possible the simultaneous collection of images in 224 contiguous bands resulting in a complete reflectance spectrum for each 20×20 m pixel in the 0.4–2.5 μm region with a sampling interval of 10 nm.[11,14,15] The FOV of the AVIRIS scanner is 30 degrees resulting in a ground FOV of 10.5 km. Private companies now recognize the potential of imaging spectrometry resulting in several sensors becoming available for specific applications. Examples are the GER Company imaging spectrometer that became operational in 1986 and the Compact Airborne Spectrographic Imager (CASI) that became operational in 1989. Currently many space agencies and private companies in developed and developing countries operate their own instruments.

In this article, the use of imaging-spectrometer data for geologic applications is discussed by describing the processing chain of data analysis from the raw data acquisition all the way through to the prospective applications. The following topics are addressed:

- physics of reflectance and emittance spectroscopy;
- airborne and spaceborne instruments;
- data acquisition and (spectral, spatial, noise adjustment) preprocessing;
- radiance to (absolute and relative) reflectance correction;
- data-analysis techniques; and
- geological applications (mineral and lithological mapping, geobotanical studies, environmental studies, petroleum-related studies, atmospheric geology, volcanology and thermography).

2 PHYSICS OF REFLECTANCE AND EMITTANCE SPECTROSCOPY

2.1 Physics of Radiation

Thermal radiation is emitted by all objects at temperatures above absolute zero. Consider an area dA and radiation arriving from a direction θ to the normal of dA, but in the range of directions forming a solid angle of $d\Omega$ steradians. Radiance, L in $W\,m^{-2}\,sr^{-1}$, is defined by

Equation (1) as

$$d\Phi = L\, dA\, d\Omega \cos\theta \tag{1}$$

where $d\Phi$ is the total power incident on dA. The total power falling on dA from all directions is given by the integration over 2π steradians for the case of no absorption or scattering by Equation (2)

$$\Phi = dA \int_{2\pi} L \cos\theta\, d\Omega = E\, dA \tag{2}$$

where E is the irradiance measured in $W\,m^{-2}$. In the reverse case when E measures the total radiance leaving dA we refer to exitance, M, and the total power emitted by the source is the radiant intensity, I, defined in Equation (3) as

$$\Phi_{\text{total}} = I\, d\Omega \tag{3}$$

We can also express these quantities spectrally by introducing an interval of wavelength λ and integrating over $d\lambda$. Spectral radiance according to Planck's relationship is given in Equation (4) by

$$L_\lambda = \frac{2hc^2}{\lambda^5}(e^{hc/\lambda kT} - 1)^{-1} \tag{4}$$

where h is Planck's constant, k is Boltzmann's constant ($1.38 \times 10^{-23}\,J\,K^{-1}$), c is the speed of light (in a vacuum) and T is the temperature. The total outgoing radiance of a blackbody of temperature T is given in Equation (5) by

$$L = \int_0^\infty L_\lambda\, d\lambda = \frac{2\pi^4 k^4}{15c^2h^3} T^4 \tag{5}$$

where $\sigma = 2\pi^4k^4/(15c^2h^3)$ is known as Stefan–Boltzmann's constant ($5.67 \times 10^{-8}\,W\,m^{-2}\,K^{-4}$). Wien's displacement law gives the relation between the wavelength at which the maximum radiation is reached, $\lambda_{\max}$, and the temperature of the blackbody as, Equation (6)

$$\lambda_{\max} = \frac{c_w}{T} \tag{6}$$

where c_w is a constant (Wien's constant) of $2.898 \times 10^{-3}\,K\,m$.

When radiation interacts with a surface, it is partly absorbed into the substance and partly scattered or reflected by the object. Consider a collimated beam of radiation incident on a surface at an incidence angle θ_0. The irradiance E is given by $F\cos\theta_0$, where F is the flux density, and scattered into a solid angle $d\Omega$ in a direction θ_1. The outgoing radiance of the surface as a result of this illumination is L_1 in the direction (θ_1, ϕ_1) where ϕ_1 is the azimuthal angle. The bidirectional reflectance distribution function (BRDF) R (in sr^{-1}) is defined in Equation (7) as

$$R = \frac{L_1}{E} \tag{7}$$

where R is a function of the incident and scattered directions and can thus be noted as $R(\theta_0, \phi_0, \theta_1, \phi_1)$. The reflectivity of the surface, r (also known as albedo), is the ratio of the total power scattered to the total power incident given in Equation (8) as

$$r(\Theta_0, \Phi_0) = \int_{\Theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} R \cos\Theta_1 \sin\Theta_1\, d\Theta_1\, d\phi_1 \tag{8}$$

Two extreme cases of scattering surface can be defined: the perfectly smooth surface (specular surface) and the perfectly rough surface (the Lambertian surface). A perfect Lambertian surface will scatter all the radiation incident upon it so that the radiant exitance M is equal to the irradiance E and the albedo is unity. A measure of roughness of a surface is given by the Rayleigh criterion. For a surface to be smooth according to the Rayleigh criterion it should satisfy Equation (9)

$$\frac{\Delta h \cos\Theta_0}{\lambda} < \frac{1}{8} \tag{9}$$

where Δh is the surface irregularity of height and λ the wavelength considered.

2.2 Reflectance Spectroscopy

Radiation incident on a material is preferentially absorbed by molecules forming the structure of the substance at wavelengths predetermined by quantum-mechanical principles. The total energy of a molecule W_t is the sum of the electronic energy W_e, the vibrational energy W_v and the rotational energy W_r given in Equation (10) as

$$W_t = W_e + W_v + W_r \tag{10}$$

Changes in energy states of molecules due to changes in rotational energy levels do not occur in solids and will not be treated here further. This discussion is restricted to changes in electronic states and vibrational processes. Harmonic vibration is described by Equation (11)

$$w_v = \overline{\frac{f(m_1 + m_2)}{m_1 m_2}} \tag{11}$$

where f is the restoring force constant (spring constant), v the vibrational states $v = 0, 1, 2, \ldots$ and m_1 and m_2 are the masses of the molecule. The possible energy states of the harmonic vibration are given by quantum mechanics

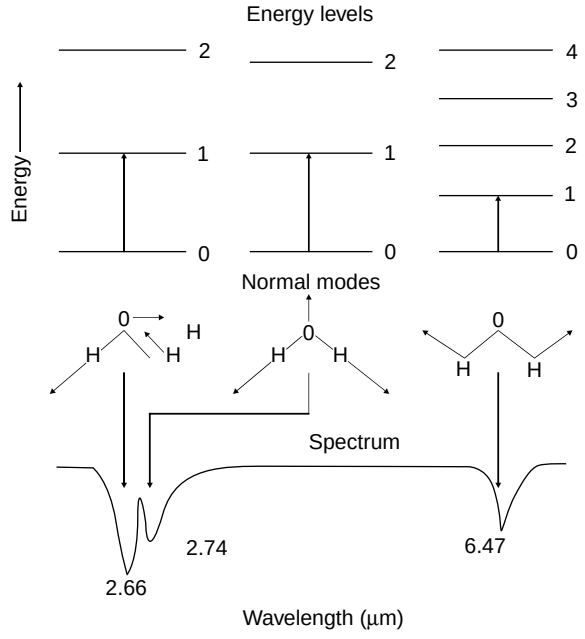

Figure 2 Vibrational energy level diagram for water vapor.

in Equation (12) as

$$W_v = \left(v + \tfrac{1}{2}\right) hw_v \tag{12}$$

where h is Planck's constant equal to $6.6 \times 10^{-34}\,\mathrm{W\,s^2}$. These define the spectral regions where absorption can occur as a result of vibrational processes (Figure 2).

Electrons in solid materials can occupy a discrete number of energy states $q = 0, 1, 2, \ldots$. When photons are incident on a material they interact with the electrons, and the absorption of a photon of a proper energy hv may cause the transition of the electron to a higher state. Absorption of energy into the medium results in absorption features in reflectance spectra. The velocity of an electron is given by Equation (13)

$$v = \frac{qh}{2mL} \tag{13}$$

where m is the mass of the electron and $L = Nl$ (l is the internuclear bond length and N the number of electrons). Thus the total energy of the electron is the sum of the kinetic energy $mv^2/2$ and the potential energy (often set to zero). Now the energy states of the electron can be described by the following formula, Equation (14)

$$W_\mathrm{q} = \frac{mv^2}{2} = \frac{q^2h^2}{8mL^2} \tag{14}$$

Each molecular orbital starting from the one with the lowest energy can accommodate only two electrons with antiparallel spin orientations (i.e. the Pauli exclusion principle). Therefore N electrons can maximally occupy $q = N/2$ states in the lowest energy state representing an energy of $W_{N/2}$ and the lowest empty one has $W_{(N/2)+1}$. We can now calculate that absorption of a photon of proper energy hv causes the transition of an electron from orbital $q = 0$ to $q = 1$ corresponding to an energy difference given in Equation (15) of

$$\Delta W = W_{(N/2)=1} - W_{N/2} = \frac{h^2}{8mL^2}(N+1) \tag{15}$$

Acknowledging that $hv = hc/\lambda$ where c is the speed of light (in a vacuum) and λ is the wavelength at which the absorption occurs, we can find the longest wavelength λ_max which can be absorbed by the molecule owing to the first transition as, Equation (16)

$$\lambda_\mathrm{max} = \frac{8mc}{h}\frac{N^2l^2}{N+1} \tag{16}$$

To be able to interpret or predict the wavelengths at which absorptions due to electronic transitions occur it is useful to know the energy level schemes of the molecules involved in the transitions.

Reflectance spectra of minerals are well known[4,5,16–24] and several studies have been conducted to determine reflectance spectra of rocks.[25–30] The reflectance characteristics of rocks can be simulated accurately by studying the compound effect of reflectance of minerals in a spectral mixture forming the rock. Reflectance spectra of mineral mixtures are a systematic combination of the reflectance of the pure mineral components or end-members in the mixtures. Several investigators have examined the systematics of mixing theoretically and empirically, and found that if the scale of mixing is large, or macroscopic, the spectral systematics of mixing different materials are linear while for microscopic or intimate mixtures the systematics are nonlinear.[31,32] Mathematical models for the analysis of reflectance spectra of mineral mixtures and rocks have been presented by many workers. Examples of minerals showing absorption features in the near-infrared (NIR) due to electronic transition processes are shown in Figure 3 while Figure 4 shows some minerals with characteristic absorption features in the SWIR due to vibrational processes. Figure 5 shows reflectance curves of rocks in the thermal electromagnetic (EM) region.

Reflectance spectra of minerals measured by different spectroradiometers with different spectral resolution are stored in spectral libraries that are available in digital format (e.g. Grove et al.,[24] Clark et al.[21]). A large variety of spectroradiometers now exist that measure the visible (VIS)/SWIR part of the spectrum (e.g. Smith,[33] Curtiss and Goetz,[34] Pearson et al.,[35] and Analytical Spectral Devices at http://www.asdi.com/) and/or the middle infrared/thermal infrared (MIR/TIR) wavelength range

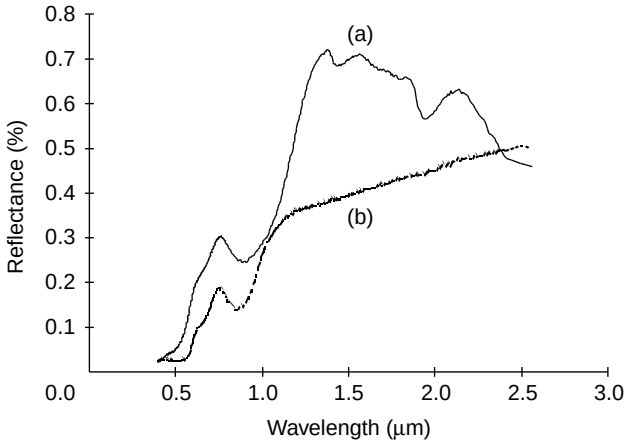

Figure 3 Spectra of (a) goethite and (b) hematite showing diagnostic absorption features short of 1.0 μm due to electronic transitions.

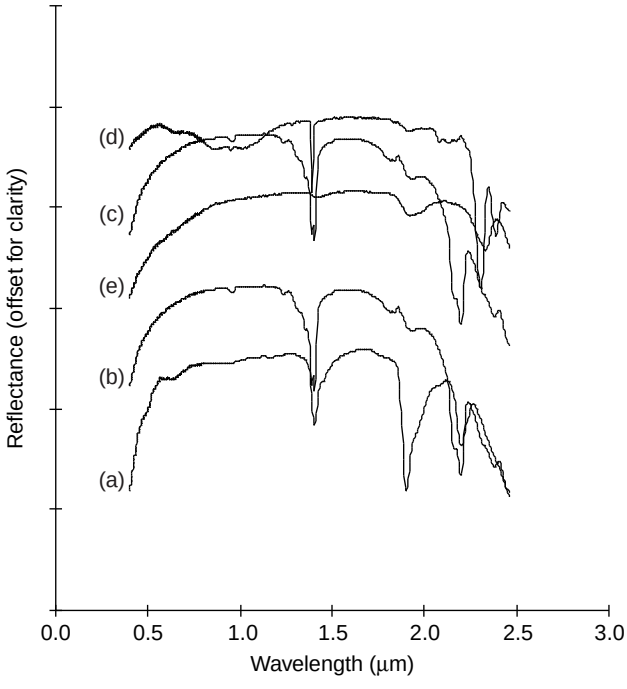

Figure 4 Spectra of minerals characterized by diagnostic absorption features due to vibrational processes: (a) montmorillonite (1.4 and 1.9 μm due to H_2O), (b) kaolinite (1.4 μm due to OH), (c) alunite (2.2 μm due to Al–OH), (d) talc (2.3 μm due to Mg–OH), and (e) calcite (2.34 μm due to CO_3).

(e.g. Hoover and Kahle).[36] Similar spectral libraries are available for vegetation (e.g. Elvidge[37]) and soils (e.g. Baumgardner et al.,[38] Condit[39]). Emissivity of terrestrial materials (i.e. minerals, rocks, soils, vegetation) in the thermal atmospheric window has been recorded and stored in a data base by Salisbury and D'Aria[40] partly based on earlier studies by Buettner and Kern,[41] Elvidge,[42] Lyon,[43] Pinkley and Williams,[44] Salisbury and D'Aria,[45] Salisbury and Milton,[46] Salisbury and Walter,[47] Schaaf and Williams,[48] Vincent et al.,[49] Walter and Salisbury,[50] and Wong and Blevin.[51] Further studies on this topic can be found in Salisbury[52] and Henderson et al.[53]

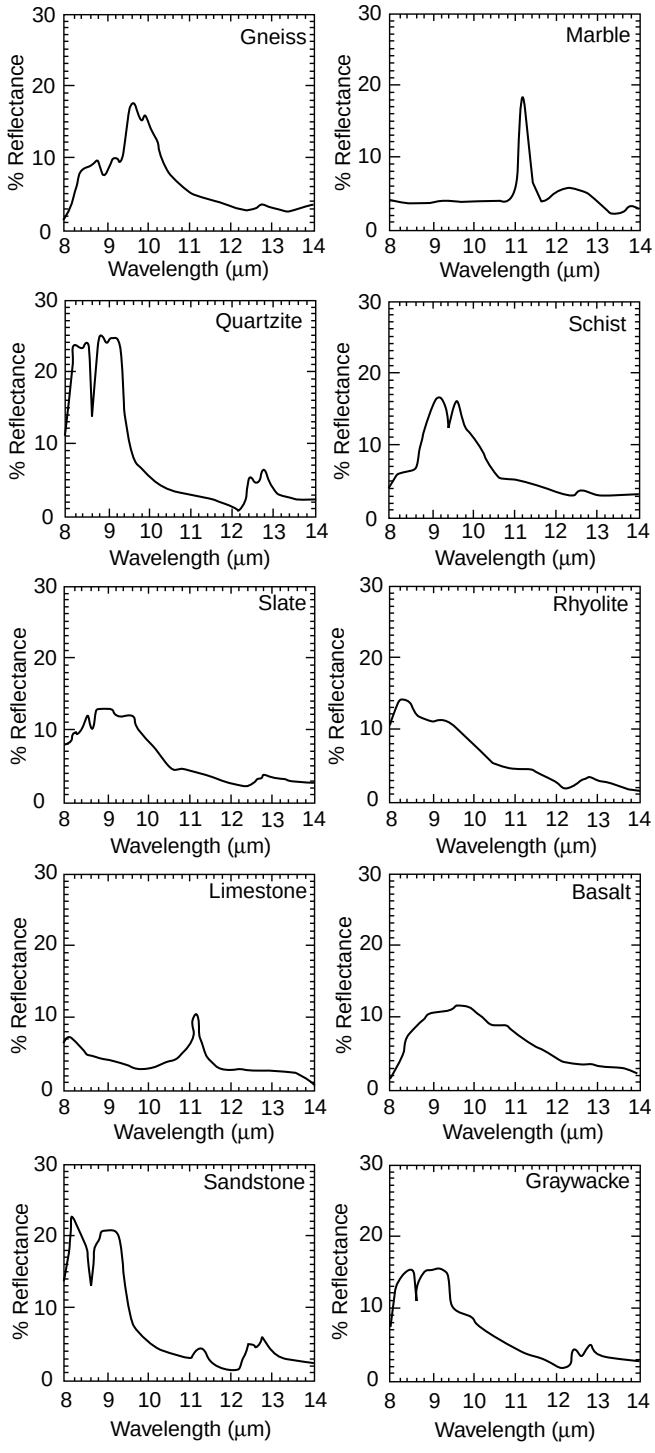

Figure 5 Directional hemispherical reflectance spectra of fresh rock.

3 AIRBORNE IMAGING SPECTROMETER SYSTEMS

3.1 Overview

It is impossible to describe all currently operational AIS systems in detail. An attempt to provide and maintain an overview of sensors resulted in ''Michael Schaepman's comprehensive list of imaging spectrometers'' at http:/www.geo.unizh.chl/~schaep/research/apex/IS_list.html. This contains a list of 45 (on 22 July 1998) airborne and spaceborne imaging spectrometer systems including a rudimentary description of their characteristics. A similar list is provided at http://www.eol.ists.ca/documents/IS-team-canada/ and in Staenz.[54] In the section on the ''History of imaging spectrometry'' an overview of development that led to some of the currently operational airborne systems is given. I provide a basic description here of two systems, AVIRIS, and Digital Airborne Imaging Spectrometers (DAIS) 7915, in order to give examples of the type and amount of data that is provided by these systems. In Table 1, a selection of sensors that are often used for geologic imaging spectrometry studies are listed.

3.2 Airborne Visible/Infrared Imaging Spectrometer

AVIRIS delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels with a wavelength-sensitive range of 10 nm nominal width in the 0.4–2.5 μm wavelength region. The instrument is operated aboard NASA's ER-2 airplane (a modified U2) at approximately 20 km above sea level with an average airspeed of about 730 km hr^{-1}. AVIRIS uses a scanning mirror to sweep back and forth (whiskbroom scanning), producing 614 pixels for the 224 detectors for each scan. Each pixel produced by the instrument

Table 1 Selection of AIS systems currently operational for geological remote sensing[a]

Sensor	Number of bands	Spectral coverage (nm)	Bandwidth (nm)	GIFOV (mrad)
AHS	48	440–12700	20–1500	2.5
AMSS (Geoscan)	32	490–1090	20–71	2.1 × 3.0
	8	2020–2370	60	
	6	8500–12000	550–590	
ARES (Lockheed)	75	2000–6300	25–70	1.17
AVIRIS (NASA)	224	400–2450	9.4	1
DAIS 7915 (DLR)	32	498–1010	16	3.3, 2.2,
	8	1000–1800	100	1.1
	32	70–2450	15	
	1	3000–5000	2000	
	6	8700–12300	600	
GERIS (GER)	24	400–1000	25.4	2.5, 3.3,
	7	1000–2000	120	4.5
	32	2000–2500	16.5	
HYDICE (Naval Research Lab.)	206	400–2500	7.6–14.9	0.5
HyMAP (Integrated Spectronics, Australia)	128	420–2481	16–20	variable
ISM (DES)	64	800–1700	12.5	3.3, 11.7
	64	1500–3000	25	
MAIS (China)	32	450–1100	20	3
	32	140–2500	30	4.5
	7	8200–12200	400–800	3
MIVIS (Italy)	20	433–833	20	2.0
	8	1150–1550	50	
	64	2000–2500	8	
	10	8200–12700	400–500	
SFSI (CCRS)	120	1200–2400	10	0.33
TRWIS-III (TRW)	396	400–2500	6.9–13.8	0.9

[a] After Staenz.[57] For a more detailed update see websites listed in section 3.1. AHS, airborne hyperspectral scanner; AMSS airborne multispectral scanner; GERIS, Geophysical Environmental Research Imaging Spectrometer; GIFOV, Ground Instantaneous Field-of-view; HYDICE, Hyperspectral Digital Imagery Collection Experiment; HyMAP, hyperspectral mapping; ISM, imaging spectroscopic mapper; MAIS, Modular Airborne Imaging Spectrometer; MIVIS, Multispectral Infrared and Visible Imaging Spectrometer; SFSI, SWIR full spectrographic imager; TRWIS, TRW imaging spectrometer.

covers an approximately 20 m^2 area on the ground (with some overlap between pixels), thus yielding a ground swath about 11 km wide. The ground data is recorded on board the instrument along with navigation and engineering data and the readings from the AVIRIS on-board calibrator.

Some details of the instrument are as follows:

- mass: 340 kg;
- encoding: 10 bit data encoding through 1994, 12 bit from 1995;
- spectrometers: four grating spectrometers (A, B, C, D);
- detectors: silicon (Si) detectors for the visible range; indium–antimonide (InSb) detectors for the NIR;
- scanning: whiskbroom scanning with 12 Hz scanning rate;
- bandwidth: 10 nm nominal, calibrated to within 1 nm;
- FOV (field-of-view): 30 degrees total FOV (full 614 samples);
- IFOV (instantaneous field-of-view): 1 mrad IFOV, calibrated to within 0.1 mrad.

More details on AVIRIS can be found in Vane et al.[13] and Green et al.[55]

3.3 Digital Airborne Imaging Spectrometer (7915)

Funded by the European Union, DLR (Deutsche Forschungsanstalt führ Luft-und Raumfahrt e.V.) operates the 79-channel DAIS 7915[56] built by the GER Company. This sensor, which is the successor of the 63-channel GERIS,[57] covers the spectral range from the visible to the thermal infrared (TIR) wavelengths at variable spatial resolution from 3 to 20 m depending on the carrier aircraft flight altitude. Six spectral channels in the 8–12 μm region could be used for the retrieval of temperature and emissivity of land surface objects. These and 72 narrow-band channels in the atmospheric windows (e.g. those wavebands that pass relatively undiminished through the atmosphere) between 0.450 and 2.45 μm allow investigation of land surface processes with a special emphasis on vegetation/soil interactions. The instrument has four spectrometers covering the 4–12.6 μm wavelength range as follows:

- spectrometer 1: 0.4–1.0 μm 32 bands bandwidth = 15–30 nm detector: Si;
- spectrometer 2: 1.5–1.8 μm 8 bands bandwidth = 45 nm detector: InSb;
- spectrometer 3: 2.0–2.5 μm 32 bands bandwidth = 20 nm detector: InSb;
- spectrometer 3: 3.0–5.0 μm 1 band bandwidth = 200 nm detector: InSb;
- spectrometer 4: 8.0–12.6 μm 6 bands bandwidth = 900 nm detector: mercury cadmium, telluride infrared (IR) detectors (MCT).

The main radiometric and geometric parameters include:

- dynamic range: 15 bit;
- sensitivity VNIR: NER (noise equivalent radiance) $<0.025\,mW\,cm^{-2}\,sr^{-1}\,\mu m^{-1}$
  SWIR: NER $<0.025\,mW\,cm^{-2}\,sr^{-1}\,\mu m^{-1}$
  MIR/TIR: NET (noise equivalent temperature) <0.1 K;
- FOV: 0.894 rad (26 degrees) on DO 228, depending on aircraft max. 39;
- IFOV: 3.3 mrad (0.189 degrees);
- GIFOV: depending on aircraft altitude 5–20 m;
- scan frequency: adjustable according to aircraft altitude between 6–24 Hz;
- image pixels per line: 512.

More details can be found in Strobl et al.[58]

4 FUTURE PLANNED SPACEBORNE IMAGING SPECTROMETER SYSTEMS

A number of orbital imaging spectrometer missions are planned or being prepared currently. HIRIS, the High-resolution Imaging Spectrometer (NASA)[59,60] is designed to acquire images in 192 spectral bands simultaneously in the 0.4–2.5-μm wavelength region with a ground resolution of 30 m and a swath width of 30 km. This scanner has a spectral resolution of 9.4 nm in the 0.4–1.0-μm wavelength region and a 11.7-nm spectral resolution in the 1.0–2.5-μm wavelength range. Though an interesting design, there are no plans for the HIRIS to fly. NASA is now supporting instead the Japanese Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER),[61,62] a high-spectral-resolution imaging spectrometer planned for launch in 1999 on the EOS AM1 platform. The instrument is designed with three bands in the VNIR spectral range with a 15-m spatial resolution, six bands in the SWIR with a 30-m spatial resolution, and five bands in the TIR with a 90-m spatial resolution. The IR/VNIR and SWIR bands have a spectral resolution in the order of 10 nm. Simultaneously, a single band in the NIR will be provided along track for stereo capability. The swath width of an image will be 60 km with 136 km crosstrack and a temporal resolution of <16 days. Also on the EOS-AM1, the Moderate Resolution Imaging Spectroradiometer (MODIS) is planned as a land remote sensing instrument with high revisting

time. MODIS is mainly designed for global change research.[63] The European Space Agency (ESA) are developing two spaceborne imaging spectrometers:[64] The Medium-resolution Imaging Spectrometer (MERIS) and the High-resolution Imaging Spectrometer (ESA) (HRIS); now renamed PRISM, Process Research by an Imaging Space Mission.[65] MERIS, currently planned as payload for the satellite Envisat-1 (see also http://www.esoc.esa.de/external/mso/envisat.html/) to be launched in 1999 by the Ariane 5, is designed mainly for oceanographic application and covers the 0.39–1.04-μm wavelength region with 1.25 nm bands at a spatial resolution of 300 m or 1200 m.[66] PRISM, currently planned for Envisat-2 to be launched around the year 2000, is a spectrometer similar to HIRIS which will cover the 0.4–2.4-μm wavelength range with a 10 nm contiguous sampling interval at a 32 m ground resolution. The first satellite imaging spectrometer to be tested was the LEWIS Hyperspectral Imager (HSI) from the TRW company which was launched in 1997 but failed. It covered the 0.4–1.0-μm range with 128 bands and the 0.9–2.5-μm range with 256 bands of 5-nm and 6.25-nm bandwidth respectively. Several other instruments are in the design stage. The Chinese Imaging Spectrometer (CIS) will have 36 bands in the visible to SWIR wavelength region and the National Space Development Agency of Japan (NASDA) global imager (GLI). The Commonwealth Scientific and Industrial Research Organization (CSIRO) ARIES (Australian Resource Information and Environment Satellite) with 32 bands from 0.5 to 1.0 μm and 32 bands from 2.0 to 2.5 μm is in the planning stage. Many private companies are exploring the possibility of developing and launching tailor-made instruments. An overview on these instruments, as well as any other review of currently available sensors, would be outdated and obsolete before publication.

5 DATA ACQUISITION AND PREPROCESSING OF IMAGING SPECTROMETER DATA

5.1 Overview

The chain of analysis steps carried out during preprocessing of imaging spectrometer data is described in this section. This generally is conducted by the manufacturers, institute/agency, or commercial company that maintains and operates the instrument. In most cases the user is provided with at-sensor radiance data and the necessary instrument characteristics (e.g. spectral response functions (SRFs), band passes, etc.) needed for further analysis. The on-ground calibration of a visible to thermal infrared (VIS/TIR) imaging spectrometer consists of a sequence of procedures providing the calibration data files and ensuring the radiometric, spectrometric, and geometric stability of the instrument:

- measurement of noise characteristics of the sensor channels;
- measurement of the dark current of the channels;
- measurement of the relative SRF of the channels;
- derivation of the effective spectral bandwidth of the channels;
- definition of the spectral separation of the channels;
- measurement of the absolute radiometric calibration coefficients (the ''transfer functions'' between absolute spectral radiance at the entrance to the aperture of the sensor and the measured radiation-dependent part of the output signal of each channel);
- definition of the NER and the noise equivalent temperature difference (NEΔT) of the channels;
- measurement of the IFOV, the spectral resolution and the deviations in the band-to-band registration.

A typical laboratory calibration facility for VIS/TIR wide-angle video spectrometric airborne sensors as found at DLR-Oberpfaffenhofen consists of the following four main parts:

1. spectrometric/geometric calibration part (SCP);
2. relative radiometric calibration part (RRCP);
3. thermal absolute calibration part (TACP);
4. absolute radiometric calibration part (ARCP).

Part 1 (SCP) comprises a mirror collimator with a flat folding mirror at the parallel-light output of the collimator. The mirror collimator has a focal length of 150 cm and a circular aperture of 30-cm diameter. At the focus position of the mirror collimator the following can be installed:

- the output slit of a monochromator for the 0.4–2.5 μm spectral region;
- the interference filter arrangement;
- different illuminated point or slit targets (for geometric calibration).

The flat folding mirror at the collimator output allows us to realize different illumination directions of the parallel light from the collimator output to the sensor entrance.

Part 2 (RRCP) is mobile and included because it is impossible to check the absolute radiometric calibration of the installed sensor by means of an integrating sphere (IS). The IS consists of an optical module and power supply unit. The optical module contains four halogen lamps of 100 W and several dissipative panels made of ground glass. The dimensions of the illuminated surface are 40 cm by 55 cm.

Part 3 (TACP) is mobile and consists of two water-cooled blackbodies, each with dimensions of $100 \times 100\,cm^2$, one at ambient or up to 5 °C below ambient temperature, the other at 25–30 °C above ambient temperature, and the supply module. The TACP is used for the radiometric calibration of the mid-IR and the thermal channels of the sensor in the laboratory and in the hangar.

Part 4 (ARCP) is used for the absolute on-ground calibration of the sensor in the visible to shortwave infrared (VIS/SWIR) spectral region (0.4–2.5 μm) by means of an IS (IS; Figure 6). The interior of the IS (1.65 m in diameter) is coated with barium sulfate. The eighteen 200-W lamps are mounted internally. Each lamp has its own individual power supply, so that each may be lit independently, providing 18 equal steps of radiance. The maximum radiance of the IS is $0.08\,mW\,sr^{-1}cm^{-2}\,nm^{-1}$ at a wavelength of 1.0 μm. The homogeneity of the radiance at the 40 cm × 55 cm rectangular output part of the sphere is better than 1%, if at least four lamps are lit. The IS has to be calibrated from time to time by establishing the ratio between the output of the sphere and a well-known radiance standard, the so-called absolute diffuse source (ADS).

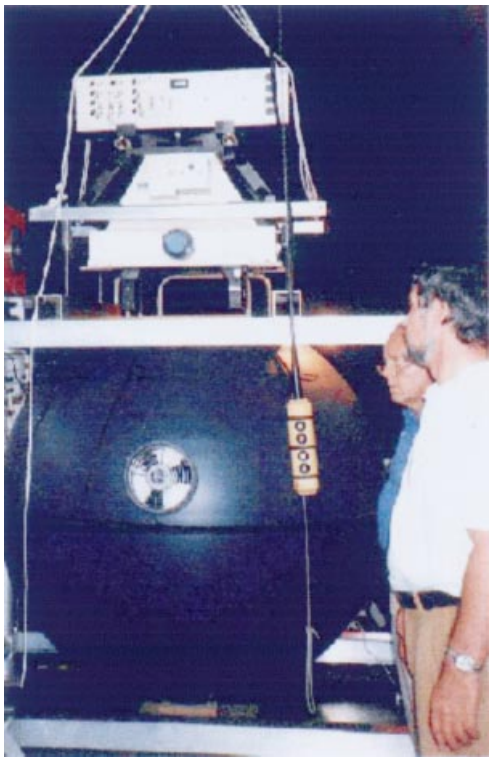

Figure 6 DAIS 7915 imaging spectrometer operated by DLR from Oberpfaffenhofen mounted on the IS for spectral calibration measurements. (Reproduced by permission of DLR, D. Oertel.)

The following preprocessing steps are described in detail below:

- spectral calibration
- spatial calibration
- noise adjustment.

The following sections are derived from discussions with engineers and scientists involved in the calibration of AISs, a subject poorly covered in literature. Those seeking more details on these topics are referred to the following articles and book chapters: data acquisition (Goetz[67]) spectral calibration (Strobl,[68,69] Green,[70] Green et al.,[71] Oertel et al.,[72]; Strobl et al.,[73]), geometric calibration and geocoding (Meyer[74]), and signal-to-noise calculation (Gao,[75] Curran and Dungan[76]).

5.2 Spectral Preprocessing Chain

Data acquisition by imaging spectrometers can be done using the whiskbroom, pushbroom, or staring principle. Whiskbroom imagers are electromechanical scanners. On-axis optics or telescopes with scan mirrors sweep from one edge of the swath to the other. The FOV of the scanner can be detected by a single detector or a single-line-detector. Simultaneously the movement of the satellite or airplane guarantees the sweeping scan over the earth. This means that the dwell time for each ground cell must be very short at a given IFOV because each scan line consists of multiple ground cells which will be detected. In whiskbroom scanning each pixel is viewed separately which allows a wide FOV and only one detector in each spectral band needs to be calibrated. The disadvantage is the short dwell time (i.e. the time that the instrument ''sees'' each pixel), which limits the spatial and spectral resolution as well as the ratio of signal-to-noise. Furthermore, the rotating mirror often causes resonance that may be observed in the data as striping.

Pushbroom scanners are electronic scanners that use a line of detectors to scan over a two-dimensional scene. The number of pixels is equal to the number of ground cells for a given swath. The motion of the aircraft or the satellite provides the scan in the along-track direction, and thus the inverse of the line frequency is equal to the pixel dwell time. By using a two-dimensional detector, one dimension can represent the swath width (spatial dimension, y) and the other the spectral range. These imaging spectrometers can be subdivided into wide-field imagers (MERIS, ROSIS (Relative Optics System Imaging Spectrometer)) and narrow-field imagers (HSI, PRISM).

Staring imagers are also electronic scanners. They detect a two-dimensional FOV instantly. The IFOV along and cross track corresponds to the two dimensions of the detector area array. Two subgroups of staring imagers

are the wedge-imaging spectrometer (WIS) and thermal-delay integration imagers (TDIs). If the incoming light passes a linear wedge filter, each row of the ground segment is seized by the detector row for a determined wavelength. For very high ground resolution and low-sensitivity applications, *xn* rows of the ground can be traced by using a TDI. The light from the scene will be separated by a linear filter for spectral band definition. On the two-dimensional detector the signal for this line can be read out from multiple rows caused by the forward movement of the sensor. Therefore the sensitivity of a TDI with n rows is n times that of an imager using the pushbroom principle.

Incoming light or radiation from the surface is split by beam splitters into the wavelength ranges specified by the instrument. By means of a set of lenses, the incoming photons are projected onto the array containing light-sensitive elements on which the charge accumulating is proportional to the integrated light intensity. For signal read-out, the charges accumulated on the detector elements are passed through an amplifier and digitizer for analog to digital conversion. This results in a digital signal, referred to here as raw radiance, digitized in 8, 12 or 16 bit. Detector materials currently used are silicon for the 0.4–1.0-μm, lead sulfide for the 0.8–2.5-μm, and indium antimonide for the 0.8–5.0-μm range. Spectral separation into the channels needed can be done through a dispersion element (grating/prism) or a filter-based system. Dispersion elements collect spectral images by using a grating or a prism. The incoming EM radiation will be separated into distinct angles. The spectrum of a single ground pixel will be dispersed and focused at different locations of one dimension of the detector array. This technique is used for both whiskbroom and pushbroom image-acquisition modes. HSIs use mainly gratings as the dispersive element. A narrow band of a spectrum can be selected by applying optical bandpass filters (tunable filters, discrete filters, and linear wedge filters). A linear wedge filter functions by transmitting light at a center wavelength that depends on the spatial position of the illumination in the spectral dimension. The detector behind the device receives light from different wavelengths of the scene.

Spectral response is not homogenous when measured over the area covered by a pixel. The point spread function (PSF) describes the decline of the measured signal and is found using a monochromator covering the sensor designated wavelength coverage and different pinhole targets.

For each spectral channel that the sensor acquires, the radiance is variable. Rather than a channel sensing only photons of one particular wavelength, the channel measures radiance in a wavelength range that stretches from a few nanometers lower to a few nanometers higher

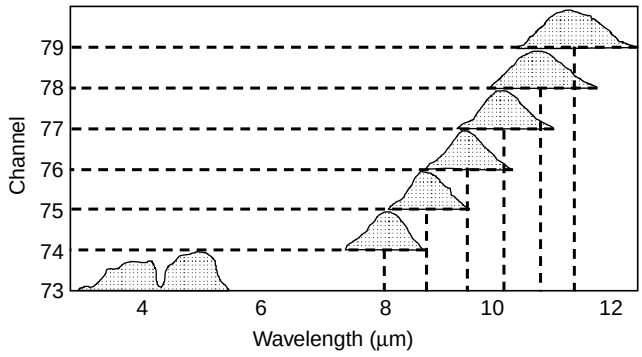

Figure 7 SRFs for the MIR/TIR channels of the DAIS 7915. (Reproduced by permission of DLR, P. Strobl.)

wavelength than the center wavelength of the channel. The curve describing the decline of the radiance levels around the central channel wavelength for each channel is known as the SRF (Figure 7) which can again be deduced from monochromator measurements.

Imaging spectrometers take indirect measurements of physical parameters in the sense that the digitized signal recorded is directly proportional to the incoming photon energy but not measured in any physically meaningful unit. The relation between the raw digitized signal and a meaningful physical parameter is established after radiometric correction yielding spectral radiance measured as the photon flux power per unit solid angle per wavelength interval. During radiometric calibration the radiometric response function is derived from the relation between the signal, caused on the detectors by the bombardment by photoelectrons, and the incoming radiance. This function translates raw radiance into spectral radiance. The function is measured by mounting the sensor onto a so-called IS, which is a half sphere reference surface coated with highly reflective barium sulfate and isolated from daylight. A set of lamps produce light of known spectral radiance in the IS which can be compared with the measurements of the sensor. These are cross-calibrated using a field spectroradiometer with a known standard. The radiometric response function corrected for the spectral response and geometric response as well for the temporal response (not further elaborated here) gives the at-sensor spectral radiance. Since the radiometric response function is linear it can be represented by two coefficients, the gain and offset, c_1 and c_0, respectively for each channel. Calculation of the at-sensor radiance, L, from raw recorded digital signal, DN, using given gain and offset values is done by Equation (17)

$$L = c_0 + c_1 \times xDN = \mathrm{mW\,cm^{-2}\,sr^{-1}\,\mu m^{-1}} \qquad (17)$$

Often thermal sensors are calibratable using two on-board blackbodies (bb), one with a low temperature, T_{bb1}

and a second with a high temperature T_{bb2}. The at-sensor radiance L_{bb} can now be found from Equation (18) as

$$L_{bb} = L_1(T_{bb1}) + \frac{L_2(T_{bb2}) - L_1(T_{bb1})}{DN_{bb2} - DN_{bb1}}(DN - DN_{bb1}) \tag{18}$$

where DN_{bb1} and DN_{bb2} are the digital numbers (DNs) for blackbody 1 and 2 and $L_1(T_{bb1})$ and $L_2(T_{bb2})$ are the corresponding spectral radiances for blackbody 1 and 2.

Most instruments measure physical parameters indirectly by generating and recording a signal, that is a DN, which is related to this physical parameter, that is radiance. The empirical relationship between the raw signal and the desired physical parameter is done through instrument calibration. The radiometric response function defines the relation between the signal caused by $N_{photons}$ and the incoming spectral radiance, L_λ. Thus each image channel needs to be calibrated to derive its radiometric response function which translates raw signal into (at-sensor) spectral radiance through a linear relationship. $N_{photons}$ is not only a function of the incident spectral radiance but is also related to sensor characteristics of which the SRF and the PSF are the two most important. Image channels are generally defined with an upper and lower wavelength defining the portion of the EM spectrum for which they are sensitive. However, what actually is sensed in terms of the wavelength range of the EM spectrum and its contribution to the total signal is represented by the SRF, recording the relative contribution to the signal of a channel of each wavelength portion of the EM spectrum. Similarly the PSF defines the aerial sensitivity of the instrument. Recall that spectral radiance is an integration over area.

Given the effects of atmospheric attenuation and sensor characteristics, remotely sensed data need to be properly calibrated to derive radiance or reflectance at sensor or at the surface. Absolute radiance or reflectance calibration can be achieved through the use of RT models that model an ideal atmosphere. Relative measurements can be derived by applying calibration coefficients provided by the instrument manufacturers. These coefficients allow translation of the given raw DNs to at-sensor radiance and reflectance and are usually the result of an experiment in an IS, that is a pure reflecting environment in which sets of lamps allow radiation to be created with known radiance levels which are in turn measured by the sensor. The relation between the radiance levels and the sensor registered digital values provides an empirical coefficient for the correction. Conversion from the raw digital values to at-sensor spectral radiances is accomplished with Equation (19)

$$L_\lambda = \text{Gain} \times DN + \text{Bias} \tag{19}$$

5.3 Spatial Preprocessing Chain

Aircraft data suffer from spatial distortions related to the carrier movements and the ruggedness of the terrain. Modern aircraft remote sensing campaigns are flown with on-board global positioning systems (GPSs) for absolute location of the aircraft at acquisition time and with on-board gyros that record tilt of the aircraft in terms of roll, pitch, and yaw. In geometric correction, for each pixel the original observation geometry is reconstructed based on the flight line, aircraft altitude, surface topography, and aircraft navigational information. The result of the correction is geocoded at sensor radiance data. Geometric distortion of AIS data resulting from data recording can be characterized through four effects (Figure 8; Rothfuss[77]):

1. Panoramic effect: owing to the scanning with constant angular scanning speed, pixels become larger from nadir to the left- and right-hand sides of the scan line.
2. Over- and undersampling: owing to nonperfect synchrony between air-speed, altitude, and scan speed in the flight direction, redundant information is scanned or data holes occur.
3. Geometric distortions due to projection: owing to movements of the aircraft, roll, pitch, and/or yaw distortions may occur.
4. Drift effect: owing to side wind or other effects the airplane may become off course and as a result distortions occur in the flight path.

In addition, topographic effects (1) result in shift of pixel locations compared with the true position and (2) affect the pixel size. In order to allow a geocoding of AIS data, on-board gyroscopic measurements record the attitude of the aircraft in terms of roll, pitch, and yaw and differential GPS measurements record the flight path in x, y, and z-absolute coordinates. This information can be used for geocoding of the image data. Three approaches can be applied:

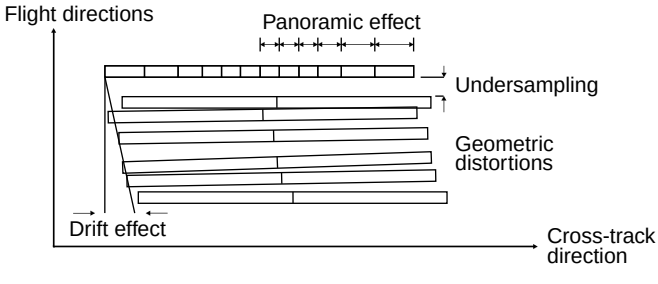

Figure 8 Common geometric distortions found in AIS data. (Reproduced by permission of DLR from H. Rothfuss, *VDI Verlag* (1994).[77])

1. geocoding using control points and registration to a map base;
2. geocoding using pixel transformations through gyroscope data;
3. parametric geocoding using gyroscopic data, flight line information, and a digital terrain model.

The first approach is commonly used in remote sensing and assumes planar correlation between a number of ground control points that are characterized in the image as well as on a map to which the image data has to be transformed. Geocoding using registration to a map base using classical control point registration methods with *n*-term polynomial or rubber sheeting image warping techniques has not been applied very successfully. With satellite data such transformations are commonly used with success. However in the case of airborne data with complex nonplanar distortions, results are typically less accurate. An example of geocoding using gyroscope data is found in Clark et al.[78] Here pixel transformations using on-board navigation and engineering data are used to obtain geocoded imaging spectrometer data. These authors demonstrate that for AVIRIS absolute geocoding to accuracies less than 100 m nominal can be achieved. The most advanced geocoding is parametric geocoding[74] for which airplane position and attitude (i.e. navigation data, x, y, z coordinate from a differential global positioning system (DGPS) and engineering data, roll, pitch, and yaw or true heading) as well as a digital elevation model (DEM) (with same coordinate system as the airplane, accuracy about the pixel size of the image data) are needed. Geocoding proceeds in four steps:

1. calculating the current observation geometry, i.e. the effective and ideal pixel-location vector from airplane position and attitude;
2. finding the intersection point to the surface, i.e. creating an oversampled DEM with accuracy better than the nominal pixel size and calculating test-vectors on the DEM to obtain the corresponding pixel vector;
3. obtaining the four corner points of each pixel, i.e. calculating effective vectors for each corner, and obtaining the DEM coordinates;
4. resampling to the original DEM geometry.

This approach, dependent on data characteristics and accuracy/resolution of the DEM, allows positional accuracies in the order of tens of meters to be achieved for AVIRIS data.

5.4 Noise Adjustment

Signal in imaging spectrometry is considered to be the quantity measured by an imaging spectrometer sensor, whereas noise describes the random variability of the signal. The quantification of the noise level alone is not a very useful measure of the quality of the imaging spectrometer data set, since the effect is more severe when the signal is low. Therefore, in most studies the signal-to-noise ratio is used as an estimate of the ratio of the signal's mean to its standard deviation. Imaging spectrometer data sets contain both periodic (coherent) sensor noise that can be removed and random noise that cannot. The signal-to-noise calculation is done on data sets with periodic noise removed. The remaining random noise can be additive noise, which is independent of the signal, and multiplicative noise which is proportional to the signal. The major part of the noise in imaging spectrometer data sets is additive and decreases sharply with both an increase in wavelength and atmospheric absorption. This random noise component consists of random sensor noise (which is image-independent), intrapixel variability (resulting from spatially heterogeneous pixel contents), and interpixel variability.

An increase in signal-to-noise ratio can be obtained by reducing the noise and retaining the signal. A method for doing this is the maximum noise fraction (MNF) transform developed by Green et al.[79] The MNF algorithm is a method for ordering data cubes into components of image quality using a cascaded principal components transform that selects new components in order of decreasing signal-to-noise ratio. In this approach, the gray levels are considered to be linear combinations of a uncorrelated signal component and a correlated noise component in which the gray level covariance is the sum of the signal and the noise covariance matrices. The difficulty in applying the MNF technique lies in finding these covariance matrices. The gray level covariance matrix can be readily derived as the sample covariance matrix of the data, but the noise covariance matrix is more complex to assess. Green et al.[79] showed that estimating the noise covariance matrix is unnecessary when noise occurs in one band only. These authors developed a procedure known as minimum/maximum autocorrelation factors (MAF) to estimate the noise covariance matrix for more complex cases. MAF exploits the fact that in most remote sensing data the signal at any pixel is strongly correlated with the signal at neighboring pixels, while the noise is not. The original data are transformed into linear orthogonal combinations in order of increasing spatial correlation which can be subsequently treated. The common approach is to apply low-pass filtering to the low-order MAF factor images which contain most of the low spatially correlated noise and a degraded signal component and removal of the high-MAF factor images that contain almost only noise. After filtering, the cleaned MAF factor images are backtransformed to the original data space. In Figure 9, the first, tenth, twentieth, and

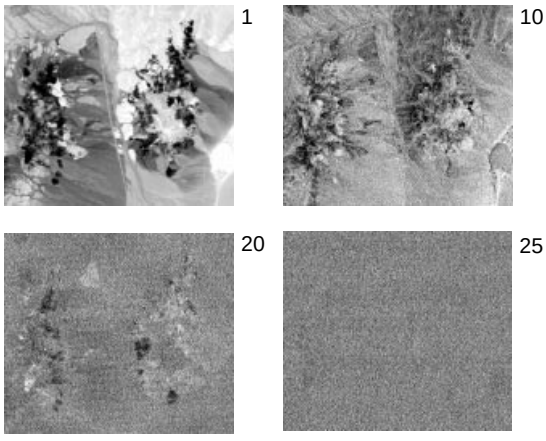

Figure 9 MNF bands 1, 10, 20, and 25 from a AVIRIS data set acquired in 1995 over the NASA test-site at Cuprite, Nevada.

twenty-fifth MNF bands of an AVIRIS data set over the NASA test-site at Cuprite, Nevada acquired in 1995 are shown.

6 RADIANCE TO REFLECTANCE CORRECTION

6.1 Overview

Raw imaging spectrometer data have the general appearance of the solar irradiance curve, with radiance decreasing towards longer wavelengths, and exhibit several absorption bands due to scattering and absorption by gases in the atmosphere. The major atmospheric water-vapor bands (H_2O) are centered approximately at 0.94 μm, 1.14 μm, 1.38 μm, and 1.88 μm, the oxygen (O_2) band at 0.76 μm, and the carbon dioxide (CO_2) bands near 2.01 μm and 2.08 μm. Additionally, other gases including ozone (O_3), carbon monoxide (CO), nitrous oxide (N_2O), and methane (CH_4), produce noticeable absorption features in the 0.4–2.5-μm wavelength region (Figure 10). The effect of atmospheric calibration algorithms is to rescale the raw radiance data provided by imaging spectrometers to reflectance by correcting for atmospheric influence thus shifting all spectra to nearly the same albedo. The result is a data set in which each pixel can be represented by a reflectance spectrum that can be directly compared to reflectance spectra of rocks and minerals acquired either in the field or in the laboratory. Reflectance data obtained can be absolute radiant energy or apparent reflectance relative to a certain standard in the scene. Comparative analysis of several methods for radiance to reflectance correction of imaging spectrometer data can be found in Rast et al.,[80] Roberts et al.[81] and Van der Meer.[82] Calibration to reflectance can be

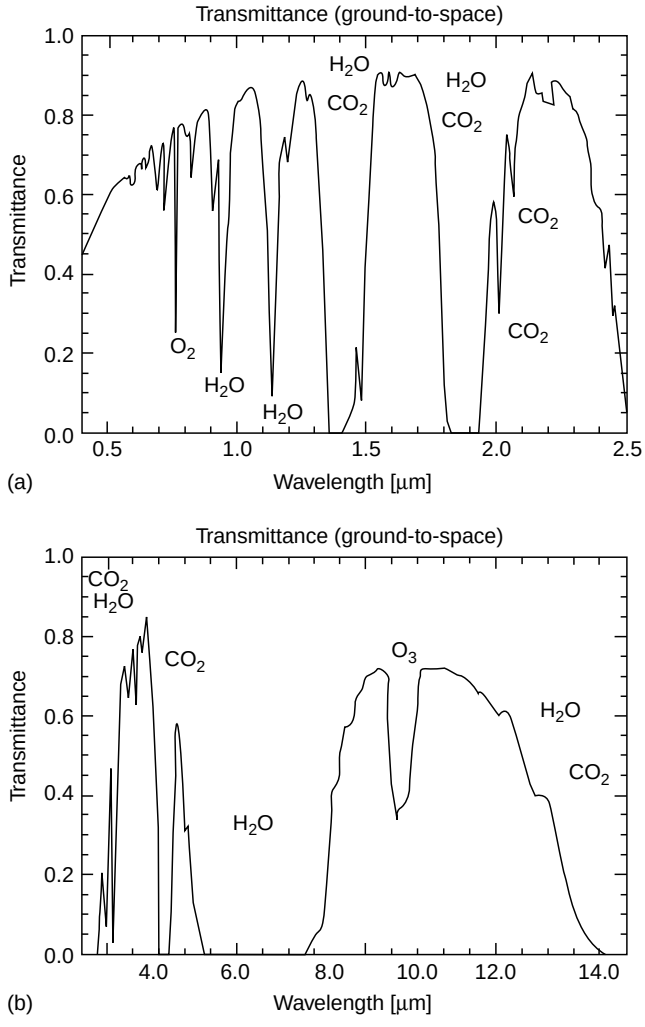

Figure 10 Atmospheric transmittance spectra (ground to space) for the VIS/SWIR (a) and MIR/TIR (b) wavelength regions with main absorbing gases indicated.

conducted to result in absolute or relative reflectance data. An example of a bright and dark target spectrum before and after atmospheric correction is shown in Figure 11.

The optical properties of the atmosphere have been described in section 5.2 as well as the underlying basic assumptions of the physics of remote sensing. Interaction of photons with constituents of the atmosphere take the form of scattering by particles (aerosol, dust, etc.) or molecules (gases) and absorption. The combined effect of absorption and scattering is known as extinction. Radiation reaching the sensor can be split into four components: path radiance, reflected diffuse radiance, reflected direct radiance and reflected radiance from the neighborhood. RT codes model the atmosphere's optical behavior given user-defined boundary conditions. The inverse problem of atmospheric correction of imaging

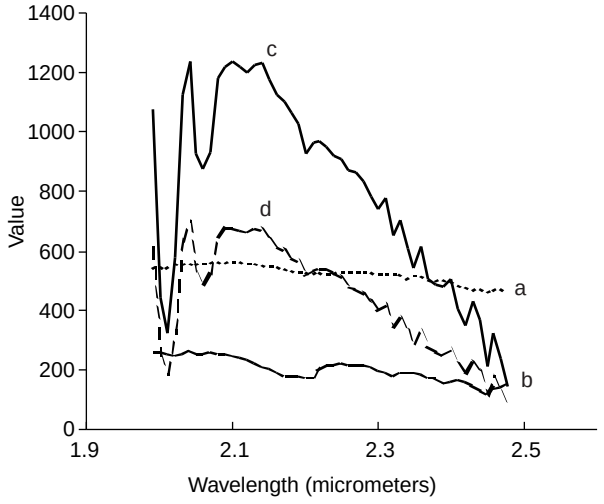

Figure 11 Bright and dark target single pixel spectrum from an AVIRIS scene before and after atmospheric correction. (a) Bright-after; (b) dark-after; (c) bright-before; (d) dark-before.

spectrometer data with the aim of obtaining radiance and/or reflectance at the ground surface can be achieved in three ways:

1. empirical correction methods to obtain apparent surface reflectance;
2. use of RT codes to obtain absolute reflectance;
3. in-flight calibration of airborne optical sensors.

It should be noted that empirical approaches such as those listed above only approximate the highly variable processes in time and space in the atmosphere controlling the transfer of radiance to and from the Earth's surface to the sensor.

6.2 Relative Reflectance

In relative reflectance data, reflectivity is measured relative to a standard target from the scene. Correction methods currently available for this purpose include:

- flat-field correction
- internal average relative reflectance (IARR) correction, and
- empirical line (EL) correction.

The purpose of the flat-field correction is to reduce the atmospheric influence in the raw imaging spectrometer data and eliminate the solar irradiance drop-off, as well as any residual instrument effects. This is achieved by dividing the whole data set by the mean value of an area within the scene which is spectrally and morphologically flat and spectrally homogeneous. The flat-field chosen should have a high albedo to avoid decreasing the signal-to-noise ratio. This can also be achieved by increasing the number of pixel spectra used to produce the flat-field spectrum. In order to select a flat-field target area properly, ground truth data are necessary to ensure that the calibration target is indeed spectrally flat. In this case, the flat-field method removes the solar irradiance curve and major gaseous absorption features as well as system-induced defects.

The IARR correction method,[83–85] allows the calibration of raw imaging spectrometer data to reflectance data when no calibration information is available. This procedure uses an "average reference spectrum" (ARS) calculated as the average pixel spectrum of the entire scene. This spectrum is divided into each image radiance spectrum to produce a relative reflectance spectrum for each pixel. Care should be taken when cover types with strong absorption features are present in the scene. In such a case the IARR correction method may cause artefacts which may be wrongly interpreted as being spectral features.

Conversion of raw imaging spectrometer data to reflectance data using the EL method[83,84,86] requires the selection and spectral characterization of two calibration targets,[81], thus assuming a priori knowledge of each site. This empirical correction uses a constant gain and offset for each band to force a best fit between sets of field spectra and image spectra characterizing the same ground areas, thus removing atmospheric effects, residual instrument artefacts, and viewing geometry effects. The correction requires four basic steps:[86]

1. The first step is the choice of two ground target regions with a wide albedo range (e.g. a dark and a bright target) and acquiring field spectra characterizing these targets.
2. Next, multiple pixels are selected that are associated with each of the two ground targets.
3. Then, an over-determined system of linear equations may be constructed for each band in which the number of unknowns is two, the gain and offset values, and the number of knowns is equal to the total number of image pixels chosen. Solving these equations by means of least-squares fitting provides gain and offset spectra and a standard error for these values for each parameter at each wavelength. The gain spectrum is an inverse solar irradiance curve, whereas the offset spectrum is a negative correction factor that increases with wavelength.
4. The final step, the actual calibration of the data, is a multiplication of the instrument DN values by the proper gain factor adding a corresponding offset value.

6.3 Absolute Reflectance

Absolute reflectance data without a priori knowledge of surface characteristics can be obtained using atmosphere models. These models correct for scattering and absorption in the atmosphere due to water vapor and mixed gases as well as for topographic effects and different illumination conditions. The 0.94-μm and 1.1-μm water absorption bands are used to calculate water vapor in the atmosphere while transmission spectra of the mixed gases in the 0.4–2.5-μm wavelength region are simulated on the basis of the water vapor values found and the solar and observational geometry. Scattering effects in the atmosphere are modeled using RT codes. A typical atmospheric correction algorithm using RT codes typically models the atmosphere's behavior under incident radiation through deriving transmission spectra of the main atmospheric gases and water vapor which are integrated with the effects of atmospheric scattering from aerosols and molecules. User input on the atmospheric condition required is the date, time, and location of data take, ozone depth, aerosol type, visibility, and elevation. This is derived from radiosonde data or meteorological stations in the area to be imaged. RT codes often used are low-resolution transmittance (LOWTRAN) code,[87] moderate-resolution transmittance (MODTRAN) code,[88,89] 5S (simulation of the satellite signal in the solar spectrum) and 6S (second simulation of the satellite in the solar spectrum).[90] An example of an atmospheric correction algorithm using RT codes to derive surface reflectance data for DAIS and AVIRIS can be found in Richter[91] and Gao et al.,[92] respectively. For a recent discussion on RT codes consult Richter[91,93] or Zagolski and Gastellu-Etchegorry.[94]

In general, in atmospheric correction the reflective (VIS/SWIR) and thermal spectral regions are treated separately since the influence of the sun dominates the solar reflective region while it can be neglected in the thermal region. The basic relation defining spectral radiance in the VIS/SWIR region is given by Equation (20)

$$L_\lambda = L_0(\lambda) + \frac{E_g(\lambda)}{\pi}[\tau_{\mathrm{dir}}(\lambda) + \tau_{\mathrm{dif}}(\lambda)]\rho(\lambda) \qquad (20)$$

where $L_0(\lambda)$ is the path radiance for a blackbody ($\rho = 0$), $E_g(\lambda)$ is the global irradiance on the ground, $\tau_{\mathrm{dir}}(\lambda)$ is the direct atmospheric transmittance (ground to sensor), $\tau_{\mathrm{dif}}(\lambda)$ is the diffuse atmospheric transmittance (ground to sensor), and $\rho(\lambda)$ is the reflectance of a Lambert surface. The basic relation defining spectral radiance in the thermal region is given by Equation (21)

$$L_\lambda = [\varepsilon_\lambda L_{\mathrm{bb}\lambda}(T) + (1 - \varepsilon_\lambda)L_{\mathrm{sky}\lambda}]\tau_\lambda + L_{\mathrm{atm}\lambda} \qquad (21)$$

where L_λ is the spectral radiance for the wavelength λ, ε_λ is the surface emissivity at wavelength λ, $L_{\mathrm{bb}\lambda}(T)$ is spectral radiance from a blackbody at surface temperature T, $L_{\mathrm{sky}\lambda}$ is spectral radiance incident upon the surface from the atmosphere, τ_λ is spectral atmospheric transmission, and $L_{\mathrm{atm}\lambda}$ is the spectral radiance from atmospheric emission and scattering that reaches the sensor. Atmospheric values in these equations are typically simulated using RT codes in conjunction with radiosonde data to derive wavelength-dependent surface emissivity (TIR) or reflectance (VIS/SWIR).

For a spaceborne sensor, the following relation can be demonstrated to relate ground reflectance, ρ, to the DN value,[95] Equation (22)

$$\rho = \frac{1}{a_1} \quad \frac{\pi(c_0 + c_1)DN}{E_{\mathrm{s}} \cos \Theta_{\mathrm{s}}} - a_0 \qquad (22)$$

where E_{s} is the extraterrestrial solar irradiance, Θ_{s} is the solar zenith angle, and a_0 and a_1 are atmospheric functions relating planetary albedo, ρ_{p} to ground albedo, ρ, in Equation (23) as

$$\rho_{\mathrm{p}} = a_0 + a_1 x \rho \qquad (23)$$

For airborne sensors the term $E_{\mathrm{s}} \cos \Theta_{\mathrm{s}}$ has to be replaced by E_{g} the global downwelling flux at the sensor altitude. The adjacency effect is approximately taken into account by calculating an $N \times N$ pixel low-pass filter image $\bar{\rho}^{(1)}$ of the reflectance data $\rho^{(1)}$ and by weighting the difference of $\rho^{(1)} - \bar{\rho}^{(1)}$ for each pixel with a correction function q to obtain the corrected reflectance values $\bar{\rho}^{(1)}$ in Equation (24) as

$$\rho^{(2)} = \rho^{(1)} + q(\rho^{(1)} - \bar{\rho}^{(1)}) \qquad (24)$$

where q is a function of the strength of the adjacency effect depending on the diffuse transmittance (ground to sensor) which results in the scattering of radiation of neighboring fields onto the FOV (see also Richter[93]).

6.4 In-flight Calibration

In-flight calibration of imaging spectrometers has a dual purpose: validating the laboratory calibration of the instrument and absolute calibration of the instrument. As laboratory calibration yields coefficients to convert raw data (DNs) to a physical quantity (at-sensor radiance) it cannot be assumed that these calibration coefficients also apply to the airborne environment. Therefore, in-flight calibration is an additional method of obtaining calibration coefficients and monitoring the instrument's radiometric stability over a flight season. For most instruments, in-flight calibration is performed at the beginning and end of the flight season. A homogenous bright area on the ground is selected as target. AVIRIS

is often calibrated over a playa near Rogers Dry Lake, CA, USA, and DAIS is calibrated at the La Grau test site[96] in France which is also used for calibration of the SPOT satellite. Ancillary information measured during the calibration flight is generally surface reflectance, atmospheric optical depths, and atmospheric water vapor, aerosols (size distribution function, spectral optical thickness), atmospheric gas transmission from passive optical measurements at ground level, and radiosonde data. These in situ measurements are used to constrain a RT code predicting spectral radiance incident onto the instruments aperture resampled to its SRF. Predicted and measured radiance values can be evaluated to derive radiometric calibration coefficients for each channel as well as their radiometric accuracy. Based on the calibration coefficients and the specified atmospheric data, reflectance data are obtained. Details on in-flight calibration are found in Green[70] and Slater et al.[97]

7 THEMATIC ANALYSIS TECHNIQUES FOR ABSORPTION FEATURE EXTRACTION

7.1 Overview

Once reflectance-like imaging spectrometer data is obtained, the logical next step is to map absorption features to determine surface composition. ''Conventional'' remote sensing techniques such as band depth mapping using ratios or differences or visual inspection of image spectra can be applied. These often provide useful information and visualizations of the data, but need experience in interpretation. Algorithms used here are restricted to those developed specifically to make use of or cope with the high spectral dimensionality typical for imaging spectrometer data. Such techniques include:

- binary encoding
- waveform characterization
- spectral feature fitting (SFF)
- spectral angle mapping (SAM)
- spectral unmixing
- constrained energy minimization (CEM)
- classification, and
- cross-correlogram spectral matching (CCSM).

7.2 Binary Encoding

A simple binary code for a reflectance spectrum can be described by Equation (25) as[9]

$$\begin{aligned} h(\mathbf{n}) &= 0 \quad \text{if } x(\mathbf{n}) \le T \\ &= 1 \quad \text{if } T < x(\mathbf{n}) \end{aligned} \tag{25}$$

where $x(\mathbf{n})$ is the brightness value of a pixel in the nth channel, T is the user-specified threshold which often equals the average brightness value of the spectrum, and $h(\mathbf{n})$ is the resulting binary code for the pixel in the nth band.

A variation of this simple encoding is to break the spectral range into a number of subregions and to code separately within these subregions following the same procedure as described above. A modification of the simple encoding has been proposed by Jia and Richards[98] who exploit multiple thresholds. Their method consists of determining the mean brightness of a pixel vector and then additionally setting upper and lower thresholds. The binary code in two-bit format can now take four different values, Equation (26)

$$\begin{aligned} h(\mathbf{n}) &= 00 \quad \text{if } x(\mathbf{n}) \le T_1 \\ &= 01 \quad \text{if } T_1 < x(\mathbf{n}) \le T_2 \\ &= 10 \quad \text{if } T_2 < x(\mathbf{n}) \le T_3 \\ &= 11 \quad \text{if } T_3 < x(\mathbf{n}) \end{aligned} \tag{26}$$

where T_1 is the lower threshold, T_2 is the mean brightness of the spectrum, and T_3 is the higher threshold.

Binary encoding provides a simple mean of analyzing data sets for the presence of absorption features. However, with the simple 0–1 coding much depends on the threshold chosen. Comparison of 0–1 coded pixel and laboratory spectra yields a qualitative indication of the presence or absence of absorption features where the depth, and thus the significance of the absorption feature, is not considered.

7.3 Waveform Characterization

In waveform characterization,[99] first the imaging spectrometer data are normalized to produce reflectance-type images. The upper convex hull is calculated as an enveloping curve on the pixel spectra which have no absorption features. Next, the hull-quotient reflectance spectrum is derived by taking the ratio between the pixel reflectance spectrum and the enveloping upper convex hull.[100] These hull quotient spectra are used to characterize absorption features, known to be attributed to a certain mineral of interest, in terms of their position, depth, width, asymmetry, and slope of the upper convex hull.

The absorption band position, λ, is defined as the band having the minimum reflectance value over the wavelength range of the absorption feature. The relative depth, D, of the absorption feature is defined as the reflectance value at the shoulders minus the reflectance value at the absorption band minimum. The width of the

absorption feature, W, is given by Equation (27)

$$W = \frac{A_{\text{all}}}{2D} \tag{27}$$

where A_{all} is the sum of the area left of the absorption band minimum, A_{left}, and the area right of the absorption band minimum, A_{right}, forming the total area under the convex hull enclosing the absorption feature. The symmetry factor, S, of the absorption feature is defined by Equation (28) as

$$S = 2\left(\frac{A_{\text{left}}}{A_{\text{all}}} - 1\right. \tag{28}$$

where A_{left} is again the area of the absorption from starting point to maximum point. Values for S range from -1.0 to 1.0 where S equals 0 for a symmetric absorption feature. The slope of the upper hull, ϕ, characterizes the slope of the upper hull over the absorption feature and is defined by Equation (29) as

$$\phi = \tan^{-1}\left(\frac{R_{\text{e}} - R_{\text{s}}}{\lambda_{\text{e}} - \lambda_{\text{s}}}\right. \tag{29}$$

where R_{e} and R_{s} are the reflectances at the ending and starting point of the absorption feature, respectively, and λ_{e} and λ_{s} are the wavelengths at the ending and starting point of the absorption feature, respectively.

As a result of the waveform characterization, five images (e.g. a position, depth, width, symmetry, and slope of upper hull image for each absorption feature selected) can be generated defining the similarity between a pixel spectrum and a laboratory spectrum of a mineral of interest based on the presence of characteristic absorption features in both spectra. However, Okada and Iwashita[99] give no solution to the mineral mapping from these images. Techniques need to be developed to estimate the probability of a mineral occurrence quantitatively from the waveform characteristics.

Data normalization using a single internally derived reference spectrum can result in spectral curves that are distorted from those observed in true reflectance spectra. To avoid this problem, Crowley et al.[101] developed a method of mineral mapping from an imaging spectrometer using relative absorption band-depth images (RBD) generated directly from radiance data. In essence, RBD images provide a local continuum correction[102] removing any small channel-to-channel radiometric offsets, as well as variable atmospheric absorption and solar irradiance drop-off for each pixel in the data set. To produce a RBD image, several data channels from both absorption band shoulders are summed and then divided by the sum of several channels from the absorption band minimum. The resulting absorption band-depth image gives the depth of an absorption feature relative to the local continuum which can be used to identify pixels having stronger absorption bands indicating that these may represent a certain mineral.

7.4 Spectral Feature Fitting

SFF, also known as the TRICORDER algorithm,[101,103,104] builds on waveform characterization in that it is a absorption feature-based method for matching image to laboratory spectra. Again continuum removed pixel spectra are used and compared to continuum reference spectra of known mineralogy (possibly derived from a spectral library). A least-squares fit is calculated band by band between each reference end member and the unknown pixel spectra. The root mean square (RMS) error of this fit is used to create an RMS image that shows pixels that are more and pixels that are less similar to the selected reference end member (see Figure 12).

7.5 Spectral Angle Mapping

SAM[105] calculates the spectral similarity between a test reflectance spectrum and a reference reflectance spectrum

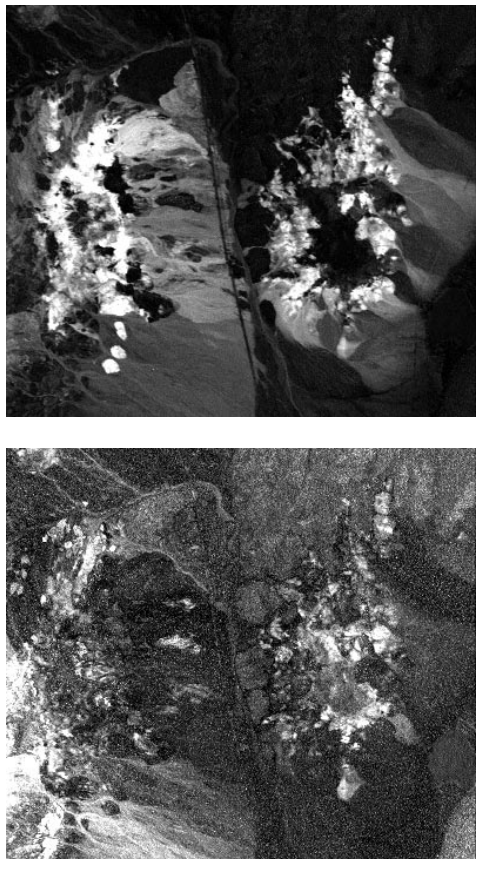

Figure 12 Spectral similarity image (a) and RMS image (b) from SFF for the mineral alunite using 1995 AVIRIS data from Cuprite, Nevada.

assuming that the data are correctly calibrated to apparent reflectance with dark current and path radiance removed. The spectral similarity between the test (or pixel) spectrum, t, and the reference (or laboratory) spectrum, r, is expressed in terms of the average angle, Θ, between the two spectra as calculated for each channel, i, in Equation (30) as

$$\Theta = \cos^{-1} \frac{\sum_{i=1}^{n} t_i r_i}{\sqrt{\sum_{i=1}^{n} t_i^2 \sum_{i=1}^{n} r_i^2}} \qquad (30)$$

In this approach, the spectra are treated as vectors in a space with dimensionality equal to the number of bands, n. The outcome of the SAM for each pixel is an angular difference measured in radians ranging from zero to $\pi/2$ which gives a qualitative estimate of the presence of absorption features that can be related to mineralogy (Figure 13).

7.6 Spectral Unmixing

The most widely used method for extracting surface information from remotely sensed images is image classification. With this technique, despite the stochastic concept of the method, each pixel is assigned to one of several known categories or classes through a statistical separation approach. Thus an image is decomposed into an image containing only thematic information of the classes previously selected as the expected image elements. In general, a training sample set of pixels is defined by the user to train the classifier. The spectral characteristics of each training set are defined through a statistical or probabilistic process from feature spaces, and unknown pixels to be classified are statistically ''compared'' with the known classes and assigned to the class to which they mostly resemble. In this way thematic information is obtained disregarding the mostly compositional nature of surface materials. Reflected radiation from a pixel as observed in remote sensing imagery has rarely interacted with a volume composed of a single homogenous material because natural surfaces composed of a single uniform material do not exist in nature. Most often the EM radiation observed as pixel reflectance values results from the spectral mixture of a number of ground spectral classes present at the surface sensed. Singer and McCord[106] showed that if the mixing scale is macroscopic such that photons interact with one material rather than with several materials, the mixing can be considered

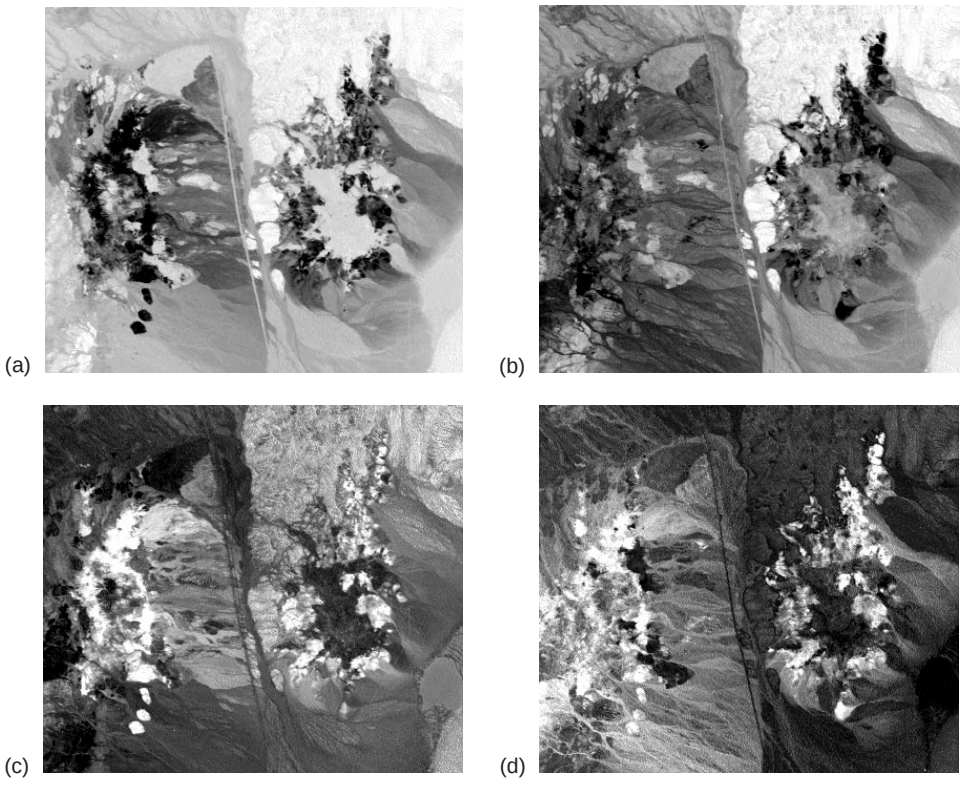

Figure 13 Spectral angle maps for (a) alunite, (b) kaolinite, (c) calcite and (d) silica using 1995 AVIRIS data from Cuprite, Nevada. Gray levels indicated the spectral angle in radians (note that brighter pixels are less similar).

linear and the resulting pixel reflectance spectrum is the linear summation of the individual material reflectance functions multiplied by the surface fraction they constitute. Various sources contribute to spectral mixing: (1) optical imaging systems integrate reflected light from each pixel, (2) all materials present in the FOV contribute to the mixed reflectance sensed at a pixel, and (3) variable illumination conditions due to topographic effects result in spectrally mixed signals. Mixing can be considered a linear process if: (1) no interaction between materials occurs, each photon seeing only one material, (2) if the scale of mixing is very large as opposed to the size of the materials, and (3) if multiple scattering does not occur.

Rather than aiming to represent the landscape in terms of a number of fixed classes, mixture modeling and spectral unmixing[107] acknowledge the compositional nature of natural surfaces and strive to find the relative or absolute fractions (or abundance) of a number of spectral components or end-members that together contribute to the observed reflectance of the image. Therefore the outcome of such analysis is a new set of images that for each selected end-member portray the fraction of this class within the volume bound by the pixel. Mixture modeling is the forward process of deriving mixed signals from pure end-member spectra, while spectral unmixing aims to do the reverse, deriving the fractions of the pure end-members from the mixed pixel signal. A review of mixture modeling and spectral unmixing approaches can be found in Ichoku and Karnieli.[108]

Mixture modeling aims to find the mixed reflectance from a set of pure end-member spectra. This is based on work by Hapke[109] and Johnson et al.[110] on the analysis of RT in particulate media at different albedos and reflectances by converting reflectance spectra to single-scattering albedos. Acknowledging the fact that in remote sensing spectra are measured in bidirectional reflectance, Mustard and Pieters[111] derived the following expression relating bidirectional reflectance $R(i, e)$ defined as the radiant power received per unit area per solid angle viewed from a specific direction $e = \cos^{-1}\mu$ and a surface illuminated from a certain direction $i = \cos^{-1}\mu_o$ by collimated light to the mean single-scattering albedo W in Equation (31) as

$$R(i, e) = \left\{ \frac{W}{4(\mu + \mu_o)} \right\} \{H(\mu)H(\mu_o)\} \tag{31}$$

where e is the viewing angle, $\mu_o = \cos(i)$, $\mu = \sin(e)$ and $H(\mu)$ is a function describing multiple scattering between particles that can be approximated by Equation (32)

$$H(\mu) = \frac{1 + 2\mu}{1 + 2\mu(1 - W)^{0.5}} \tag{32}$$

The mean single-scattering albedo of a mixture is a linear combination of the single-scattering albedos of the end-member components weighted by their relative fraction in Equation (33) as

$$W(\lambda) = \sum_{j=1}^{n} W_j(\lambda) F_j \tag{33}$$

where W is the mean single-scattering albedo, λ is the spectral waveband, n is the number of end-members and F_j is the relative fraction of component j. F_j is a function of the mass fraction M_j, the density, ρ_j and the diameter d_j of the end-member j in Equation (34) as

$$F_j = \frac{M_j/\rho_j d_j}{\sum_{j=1}^{n} M_j/\rho_j d_j} \tag{34}$$

Thus these equations allow us to model reflectance properties and mean single-scattering albedos of mixtures and form the basis of mixture-modeling studies. Spectral unmixing is a deconvolution technique that aims to estimate the surface fractions of a number of spectral components (or end-members) together causing the observed mixed spectral signature of the pixel. A linear combination of spectral end-members is chosen to decompose the mixed reflectance spectrum of each pixel, R_i, into fractions f_j of its end-members, Re_{ij}, by Equation (35)

$$R_i = \sum_{j=1}^{n} f_j \, Re_{ij} + \varepsilon_i \quad \text{and} \quad 0 \le \sum_{j=1}^{n} f_j \le 1 \tag{35}$$

where R_i is the reflectance of the mixed spectrum in image band i for each pixel, f_j is the fraction of each end-member j calculated band by band, Re_{ij}, is the reflectance of the end-member spectrum j in band i, i is the band number, j is each of the n image end-members and ε_i is the residual error or the difference between the measured and modeled DN in band i. A unique solution is found from this equation by minimizing the residual error, ε_i, in a least-squares solution. This residual error, ε_i, is the difference between the measured and modeled DN in each band and should in theory be equal to the instrument noise in the case where only the selected end-members are present in a pixel. Residuals over all bands for each pixel in the image can be averaged to give a RMS error, portrayed as an image, which is calculated from the difference of the modeled (R_{jk}) and measured (R'_{jk}) pixel spectrum in Equation (36) as

$$RMS = \sum_{k=1}^{m} \frac{\sqrt{\sum_{j=1}^{n} \frac{(R_{jk} - R'_{jk})^2}{n}}}{m} \tag{36}$$

where n is the number of spectral bands and m the number of pixels within the image. The solution to the unmixing is found through standard matrix inversion such as the gaussian elimination method. Boardman,[112] however, suggested the use of singular value decomposition of the end-member matrix which has the advantage that singular values can be used to evaluate the orthogonality of the selected end-members. If all end-members are spectrally unique, all singular values are equal. For a degenerate set of end-members all but one singular value will be zero.

A solution to the mixing as found by a process of matrix inversion is give by Settle and Drake.[113] These authors define a vector of expected pixel signals $\boldsymbol{\mu}_i = \{\mu_{i1}, \mu_{i2}, \ldots \mu_{in}\}^T$ for the n ground-cover classes giving an expected mixed pixel signal under strictly linear conditions in Equation (37) as

$$f_1\mu_1 + f_2\mu_2 + \cdots + f_c\mu_c = Mf \tag{37}$$

The columns of the matrix M are the vectors $\boldsymbol{\mu}_i$ which are the end-member spectra. The observed signal of pure pixels will exhibit statistical fluctuations due to sensor noise characterized by a noise variance–covariance matrix N_i. Therefore pixels with the mixture f will exhibit fluctuations around their mean value Mf characterized by the noise covariance matrix $N(f)$ given by Equation (38)

$$N(f) = f_1N_1 + f_2N_2 + \cdots + f_cN_c \tag{38}$$

If the $N(f)$ is independent of f, thus when the noise components are uncorrelated, the linear model can be defined according to Settle and Drake[113] in Equation (39) as

$$x = Mf + \boldsymbol{e} \tag{39}$$

where $\boldsymbol{e}$ is the vector of errors satisfying Equation (40)

$$E(\boldsymbol{e}) = 0 \quad \text{and} \quad E(\boldsymbol{e}\boldsymbol{e}^T) = N \tag{40}$$

This signifies that the expectation of $\boldsymbol{e}$ is zero and that the expectation of the noise component is close to the sensor noise component.

The spectral unmixing model requires a number of spectrally pure end-members to be defined which cannot exceed the number of image bands minus one in order to allow a unique solution to be found that minimizes the noise or error of the model. Furthermore, constraining factors arise where fractions sum to unity and fractions for the individual end-members vary from 0 to 100%. Constrained unmixing makes the assumptions stated above. Other cases of unmixing not taking into account these assumptions are referred to as unconstrained unmixing.

Two versions of spectral unmixing are generally implemented: constrained and unconstrained unmixing. Constraining assumptions that can be implemented separately are (1) that the fractions are non-negative and (2) that the fractions sum to unity. These constraints are only meaningful when considering the scientific field of applications (e.g. the earth science perspective). From a statistical viewpoint, or even from an image interpretation viewpoint, it is meaningless to force mixture models to be constrained to the data. The set of end-members should describe all spectral variability for all pixels, produce unique results, and be of significance to the underlying science objectives. Selection of end-members can be achieved in two ways:

1. selecting end-members from a spectral (field or laboratory) library; and
2. deriving end-members from the purest pixels in the image.

The second method has the advantage that selected end-members were collected under similar atmospheric conditions. The end-members resulting from application of the first option are generally denoted as ''known'' end-members while the second option results in ''derived'' end-members. Identification of the purest pixels in the scene is done through compression of the data using principal component analysis (PCA) following a method developed by Smith et al.[114] Spectrally pure pixels are found at the vertices of the polygon that bounds the data space of the principal components. Selection of spectral end-members often is an iterative process where additional end-members are selected on the basis of clearly visible spatial patterns in the RMS error image until the RMS error obtained does not show any obvious systematic spatial patterns of error distribution. Boardman et al.[115] introduced an algorithm for finding the most spectrally pure pixels using the pixel purity index (PPI). The PPI is computed by an iterative process of projecting n-dimensional scatterplots (where n is the number of bands) onto a random unit vector and recording the number of times that each pixel is marked as extreme. An example of the results obtained through spectral unmixing of AVIRIS data from Cuprite are shown in Figure 14.

7.7 Constrained Energy Minimization

CEM developed by Farrand and Harsanyi[116] is an extension to spectral unmixing. CEM maximizes on a pixel-by-pixel basis the response of a target signature and suppresses the response of undesired background signatures. It is assumed that foreground and background signatures are mixed linearly, as is the case when each photon only interacts with one material. The CEM strives at finding a vector that suppresses the unknown background signature while enhancing the

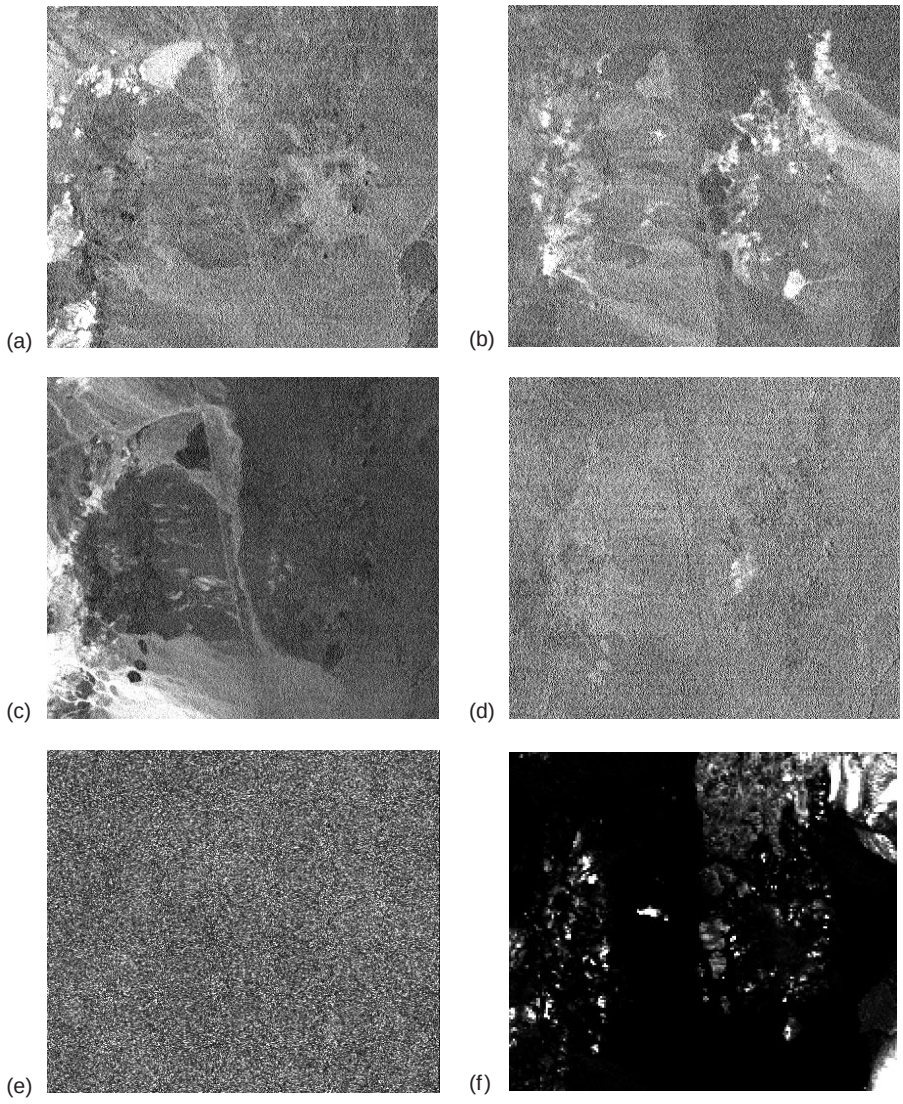

Figure 14 Abundance (fraction) images for (a) calcite, (b) kaolinite, (c) illite/muscovite, and (d) buddingtonite for 1995 AVIRIS data from Cuprite, Nevada. The resulting RMS image is shown (e) together with the PPI (f) image used for end-member selection.

target signature. This is achieved by minimizing the total output energy of all pixels and by assuming the energy of an individual pixel summed across the wavelength range to be one when applied to the target pixel spectrum. The result of CEM is a vector component image that is comparable to fraction abundance images typically obtained through unmixing.

7.8 Classification

Classification of remotely sensed imagery into groups of pixels having similar spectral reflectance characteristics is often an integral part of digital image analysis. On the basis of the spectral reflectance that training pixels (e.g. pixels known to represent a ground class of interest) exhibit, pixels for which the ground cover type is unknown are classified. Classification routines aim to compare the observed spectral reflectance of pixels with unknown composition with that of training pixels, and assign the unknown pixel to that group which most resembles their spectral reflectance characteristics. Techniques making use of training data sets are referred to as supervised classification algorithms as opposed to unsupervised classification techniques in which no foreknowledge of the existence of ground classes is required. Supervised and unsupervised image classification techniques have been widely used in the analysis of conventional remote sensing data type (e.g. Landsat MSS and TM, and SPOT) and several studies have been undertaken to develop algorithms based on classification

for hyperspectral data types. Cetin et al.[117] and Cetin and Levandowski[118] use n-dimensional probability density functions (PDFs) based on the principle of the maximum likelihood classifier to analyze AVIRIS, Thermal Infrared Multispectral Scanner (TIMS), and Landsat TM images. Lee and Landgrebe[119] give algorithms for the minimum distance classifier in high-spectral-resolution imagery. An approach using neural networks is given in Benediktsson.[120] An alternative classification algorithm for imaging spectrometer data based on indicator kriging is given in Van der Meer.[121,122]

7.9 Cross-correlogram Spectral Matching

CCSM[123,124] is a new approach toward mineral mapping from imaging spectrometer data using the cross-correlogram of pixel and reference spectra. A cross-correlogram is constructed by calculating the cross-correlation at different match positions, m, between a test spectrum (i.e. a pixel spectrum) and a reference spectrum (i.e. a laboratory mineral spectrum or a pixel spectrum known to represent a mineral of interest) by shifting the reference spectrum over subsequent channel positions by Equation (41):

$$r_m = \frac{n\sum \lambda_r\lambda_t - \sum \lambda_r \sum \lambda_t}{\sqrt{n\sum \lambda_r^2 - \left(\sum \lambda_r\right)^2 \; n\sum \lambda_t^2 - \left(\sum \lambda_t\right)^2}} \tag{41}$$

where r_m is the cross-correlation at match position m, λ_t is the test spectrum, λ_r is the reference spectrum, n is the number of overlapping positions (spectral bands), and m the match position. The statistical significance of the cross-correlation coefficient can be assessed by a Student's t-test and the skewness can be calculated as an estimation of the goodness-of-fit. The cross-correlogram for a perfectly matching reference and test spectrum is a parabola around the central matching number ($m = 0$) with a peak correlation of one. Deviations from this shape indicate a different surface mineralogy. Mineral mapping on a pixel-by-pixel basis is achieved by extracting three parameters from the cross-correlograms and combining these into a statistical estimate of the goodness-of-fit of the two spectra compared: the correlation coefficient at match position zero, the moment of skewness (based on the correlation differences between match numbers of equal but reversed signs, e.g. $m = 4$ and $m = -4$), and the significance (based on a Student's t-test testing the validity of the correlation coefficient at $m = 0$). In order to evaluate the surface mineralogy maps a RMS error assessment procedure is proposed in Van der Meer and Bakker[125] in which the error is calculated from the difference between the calculated pixel cross-correlogram and the ideal cross-correlogram calculated for the reference in Equation (42) as (Figure 15):

$$RMS = \sqrt{\frac{\sum_0^M (R_M - R'_M)}{N}} \tag{42}$$

where R_M is the pixel cross-correlation at match position m, R'_M is the reference cross-correlation at match position m, N is the number of match positions, M is the match number.

8 EXTRACTING EMISSIVITY FROM THERMAL INFRARED DATA

Thermal radiance measured from the Earth's surface in the TIR region consists of temperature and emissivity information. Variations in emissivity relate to the chemistry and texture of materials at the surface. However, these are blurred by the effects of temperature variations which are an order of magnitude larger that the emissivity effects. Planck's function (defined earlier) relates surface temperature of a perfect blackbody to spectral radiance. Most surface materials do not emit radiance like a perfect blackbody, their spectral emissivity (ε_{ij}) is defined as the ratio of the radiance of the material to that of a blackbody at the same temperature given in Equations (43) and (44) as

$$\varepsilon_{ij} = \frac{L_{ij}}{L_{ij}^{bb}} \tag{43}$$

or

$$L_{ij} = \varepsilon_{ij} L_{ij}^{bb} \tag{44}$$

where L_{ij} is the radiance of the material and L_{ij}^{bb} the radiance of a blackbody of similar temperature. From this it can be seen that, if radiance is measured from the surface in n channels, there will always be $n + 1$ unknowns, n emissivities (one per channel) and an unknown surface temperature. Assumptions are needed to solve these equations, and several methods were proposed to find emissivities and temperature. A common approach is the thermal log residuals (see Hook et al.[126]) approach that utilizes Wien's approximation of Planck's law, Equation (45)

$$L_{ij}^{BB} = \frac{C_1}{\lambda_j^5 \pi (e^{C_2/\lambda_j \tau_i})} \tag{45}$$

where C_1 and C_2 are the first and second radiation constants. Taking natural logs of the surface radiance

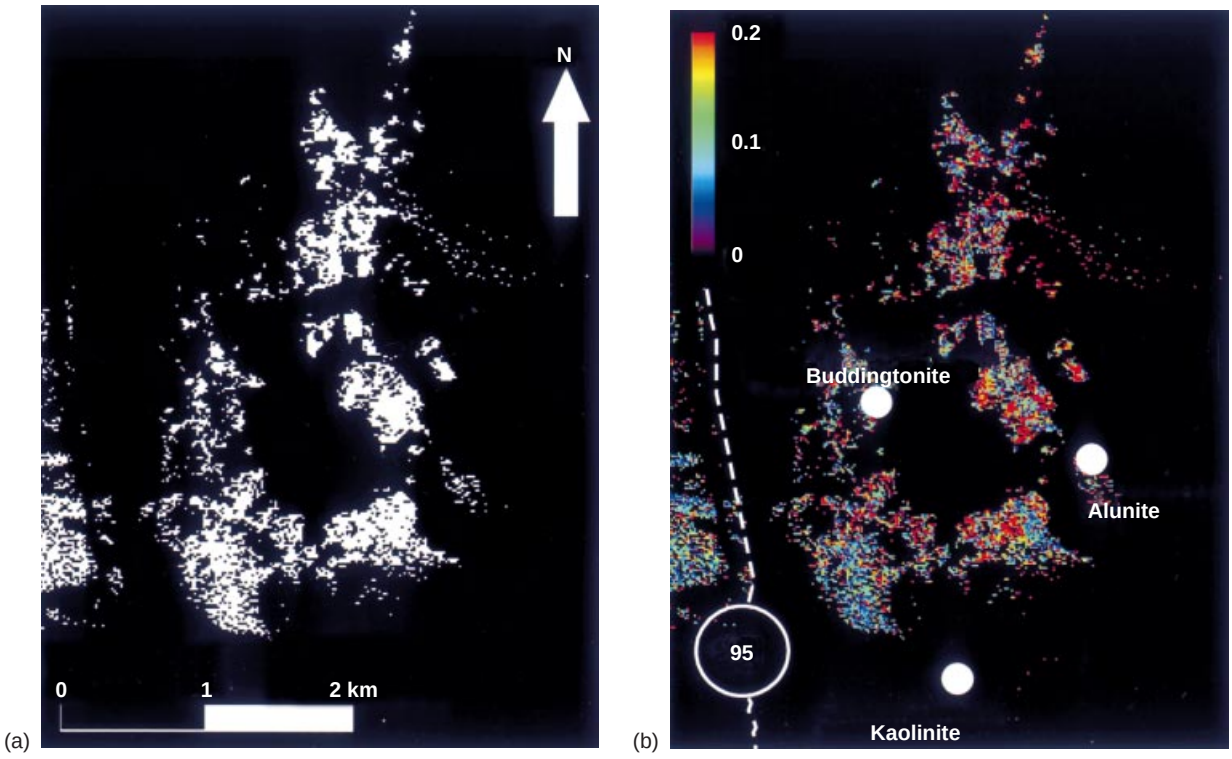

Figure 15 Surface mineralogy map for alunite (a) at Cuprite, Nevada and corresponding RMS (b) from 1995 AVIRIS data.

using Wien's approximation we obtain Equation (46)

$$\ln L_{ij} = \ln \varepsilon_{ij} + \ln C_1 - 5 \ln \lambda_j - \ln \pi - \frac{C_2}{\lambda_j T_i} \tag{46}$$

This relation is manipulated with a set of means derived from the data in order to obtain the thermal log residual value consisting only of the emissivity term in Equation (47) by

$$Y_{ij} = X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet} \tag{47}$$

where X_{ij} is the wavelength-weighted log of the Wien radiance at pixel i in channel j in Equation (48) equal to

$$X_{ij} = \lambda_j \ln \varepsilon_{ij} + \lambda_j \ln C_1 - 5\lambda_j \ln \lambda_j - \lambda_j \ln \pi - \frac{C_2}{T_i} \tag{48}$$

$X_{i\bullet}$ is the geometric mean over all channels for pixel j in Equation (49) equal to

$$X_{i\bullet} = \frac{1}{n} \sum_{j=1}^{n} \lambda_j \ln \varepsilon_{ij} + \frac{\ln C_1}{n} \sum_{j=1}^{n} \lambda_j - \frac{n-1}{n} \sum_{j=1}^{n} \lambda_j \ln \lambda_j - \frac{\ln \pi}{n} \sum_{j=1}^{n} \lambda_j - \frac{C_2}{T_1} \tag{49}$$

with n as the number of channels.

$X_{\bullet j}$ is the geometric mean mean over all pixels in the scene for channel i in Equation (50)

$$X_{\bullet j} = \frac{\lambda_j}{N} \sum_{i=1}^{N} \ln \varepsilon_{ij} + \lambda_j \ln C_1 - 5\lambda_j \ln \lambda_j - \lambda_j \ln \pi - \frac{C_2}{N} \sum_{i=1}^{N} \frac{1}{T_i} \tag{50}$$

where N is the number of pixels in a channel.

$X_{\bullet\bullet}$ is the geometric mean over all pixels in all channels given by Equation (51)

$$= \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_j \ln \varepsilon_{ij} + \frac{\ln C_1}{n} \sum_{j=1}^{n} \lambda_j - \frac{n-1}{n} \sum_{j=1}^{n} \lambda_j \ln \lambda_j - \frac{\ln \pi}{n} \sum_{j=1}^{n} \lambda_j - \frac{C_2}{N} \sum_{i=1}^{N} \frac{1}{T_i} \tag{51}$$

Now the residual value is found in Equation (52) as

$$Y_{ij} = \lambda_j \ln \varepsilon_{ij} - \frac{1}{n} \sum_{j=1}^{n} \lambda_j \ln \varepsilon_{ij} - \frac{\lambda_j}{N} \sum_{i=1}^{N} \ln \varepsilon_{ij} + \frac{1}{nN} \sum_{j=1}^{n} \sum_{i=1}^{N} \lambda_j \ln \varepsilon_{ij} \tag{52}$$

If $l_j = \lambda_j/\lambda_\bullet$ and $e_{ij} = (\varepsilon_{ij})^{l_j}$ where $\lambda_\bullet$ is the mean of the central wavelengths of all channels then Equation (53) follows

$$\frac{Y_{ij}}{\lambda_\bullet} = \ln e_{ij} - \frac{1}{n}\sum_{j=1}^{n} \ln e_{ij} - \frac{1}{N}\sum_{i=1}^{N} \ln e_{ij} + \frac{1}{nN}\sum_{j=1}^{n}\sum_{i=1}^{N} \ln e_{ij} \quad (53)$$

Thus the thermal log residuals represent variations in a combination of emissivities for a given channel rather than variations of temperature and emissivity.

9 APPLICATIONS OF IMAGING SPECTROMETRY

9.1 Overview

In this section, applications of imaging spectrometry are discussed. Technological studies on sensor performance, calibration and validation of sensor characteristics as well as studies that evaluate new processing strategies are not elaborated. Earth-science-related applications, especially geological studies, are emphasized. Most of the studies are discussed in a rudimentary fashion; the interested reader is referred to the references for further details.

9.2 Mineral Mapping: Surface Mineralogy

Goetz et al.[14] demonstrated that spectroscopy may aid in surface mineralogy mapping. This early work was the start of a large number of similar types of work. Murphy[127] for the first time succeeded in mapping jasperoid in the Cedar mountains. An expert system approach without a priori knowledge of the area was successfully implemented by Dwyer et al.[128] where they applied AVIRIS data to surface mineralogy mapping in the Drum mountains of Utah (USA). Similarly, Gaddis et al.[129] presented techniques that use only information contained within a raw AVIRIS data set to estimate and remove additive components (atmospheric scattering and instrument dark current), normalize multiplicative components (instrument gain, atmospheric transmission), and enhance, extract, and map surface composition and mineralogy. Kruse et al.[130] presented mineral mapping using an expert-systems approach. Ben-Dor et al.[131] and Ben-Dor and Kruse[132] demonstrated that mineral maps (derived using spectral unmixing techniques) from GER 63-channel imaging spectrometer data over the Makhtesh Ramon area in Israel provide additional geological details not yet revealed on existing geological maps. This work builds on earlier studies by Kaufman et al.[133] and Crowley et al.[101] using AIS data (AVIRIS) of the Ruby mountains, Montana for mineral discrimination using RBD images. Crowley[134] also presented a study on mapping of evaporite minerals with AVIRIS in Death Valley. A study by Farrand and Seelos[135] showed the possibility of mapping faults by evaluating mineral maps for linear or curvilinear features. These authors were able to delineate some previously unmapped faults in the Summitville area from mineral end-member images obtained using the CEM technique. The use of classification methods including neural networks for mapping surface mineralogy from AVIRIS data of a volcanic terrain on Iceland is described in Benediktsson[120] Baugh et al.[136] showed the potential of AVIRIS for mapping ammonium-type minerals.

9.3 Mineral Mapping: Exploration

Numerous studies on remote sensing to aid in mineral exploration have been conducted at Cuprite and Goldfield, Nevada (USA), two of NASA's test sites for instrument calibration and geological mapping. Most of these studies center on mapping of hydrothermal alteration to aid in mineral prospecting (particularly for gold). Calibration and testing of new sensors as well as validating new processing strategies is typically conducted at Cuprite. Cuprite and Goldfield are areas of extensive hydrothermal alteration within a sequence of rhyolitic welded ash flow and air fall tuffs[137,138] which can be subdivided into three mappable units: silicified rocks, opalized rocks, and argillized rocks. Silicified rocks form a large irregular patch extending from the middle to the south end of the area. The silicified core represents the most intensely altered rocks at Cuprite containing quartz, calcite, and minor alunite and kaolinite. Opalized rocks contain abundant opal and as much as 30% alunite and kaolinite. Locally, an interval of soft poorly exposed material mapped as argillized rock separates fresh rock from opalized rock. In the argillized rocks, plagioclase is altered to kaolinite, and glass is altered to opal and varying amounts of montmorillonite and kaolinite. The distribution of these alteration assemblages is characteristic of a fossilized hot-spring deposit, which often contains gold.

Volcanism at Cuprite began in Oligocene with eruption of rhyolite and quartz latite flows. Hydrothermal alteration is related to a second, early Miocene, period during which dacite and andesite flows were extruded and hot acidic brines began circulating through the volcanics. The general geology of the Cuprite mining district is treated in more detail in Abrams et al.[139] Mineral zones can be mapped as rings around a central vent formed by altered silica. The alteration zones are composed of kaolinite, alunite, buddingtonite, and various ferrigenous minerals

characterizing the host rocks. Some of the sedimentary units show mappable amounts of calcite and dolomite. The mineral zonation is indicative of a sulfuric acid-charged system with hydrothermal fluids emitted in a hot spring-type of environment. The circular distribution of the mineral zones demonstrates that alteration occurred along a central vent with the lateral mineral zoning controlled by a decrease of acidity and temperature. It should be noted that alunite is a pathfinder mineral for gold; in fact for many years gold has been explored at mines near the town of Goldfield.

Some examples of recent studies at Cuprite (Figure 16) and Goldfield (data and (in some cases) the analytical technique given in parentheses) include those of Resmini et al.[140] (HYDICE, CEM), Okada and Iwashita[99] (waveform characterization), Abrams and Hook[141] (simulated ASTER data, decorrelation stretching), Nedeljkovic and Pendock[150] (AVIRIS, PDFs), Carrere and Abrams[143] (AVIRIS), Kruse et al.[86] (GER 63-channel imaging spectrometer), Hook et al.[144] (AVIRIS, Geological Scanner Geoscan), Van der Meer and Bakker[124] (AVIRIS, CCSM).

Other examples of gold mineral exploration through mapping of hydrothermal alteration zones are described by Crosta et al.[145] who use Geoscan AMSS MK-II data in an area in the Rio Itapicuru greenstone belt in the northeastern part of Brazil, Ferrier and Wadge[146] who use AVIRIS data from southern Spain, Hutsinpiller[147] who used AIS-I data from Virginia City (Nevada), Feldman and Taranik[148] who used AIS-I data to map alteration minerals in the Tybo mining district (Nevada), and Crosta et al.[149] who use AVIRIS data for alteration mapping at Bodie (California).

Other alteration processes that have been studied in imaging spectrometer data are the serpentinization of ultramafic rocks which relates to major asbestos deposits. Based on the laboratory studies of Hunt and Evarts[150] it has been shown that the degree of serpentinization of ultramafic rocks can be estimated through spectral analysis of imaging spectrometer data.[151] Beratan et al.[152]

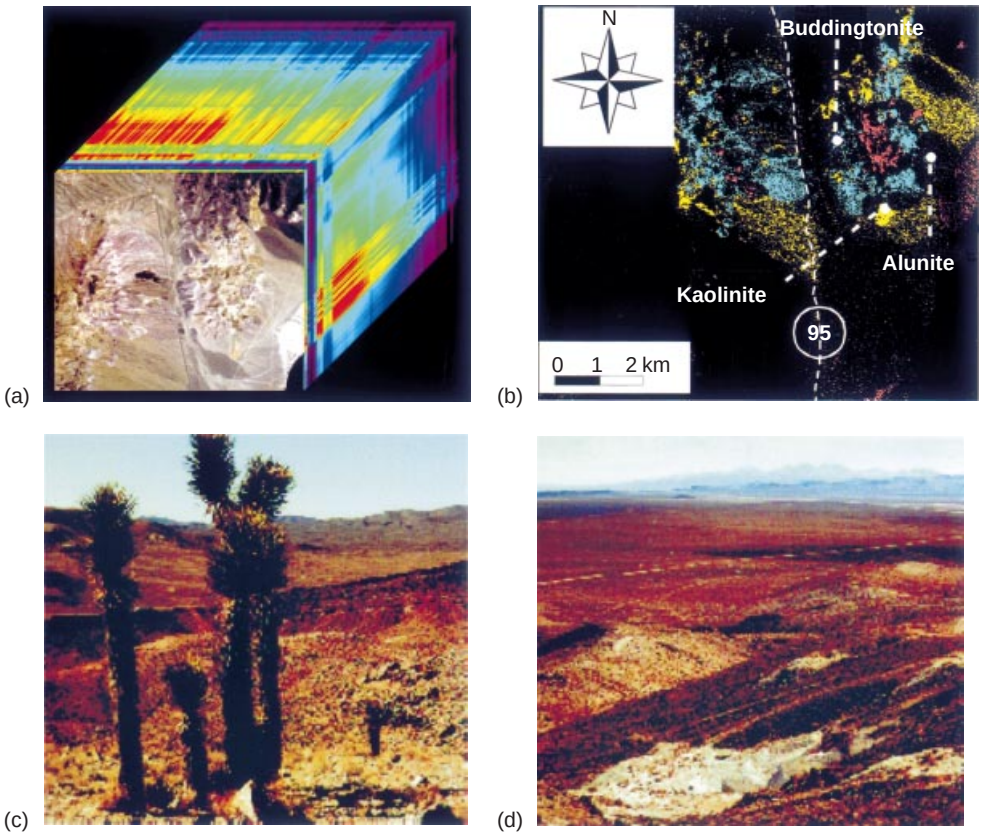

Figure 16 Cuprite, Nevada as seen from AVIRIS. (a) Image cube (face is a red, green, blue image with $R = 2.2$, $G = 2.3$, $B = 2.4\,\mu m$) where the sides of the cube are formed by spectral slices through the (in this example) 40 image channels that form the SWIR part of the spectrum. Color coding is in terms of reflectance (red is high, black is low reflectance), each line represents one channel. (b) Surface mineralogy showing kaolinite in yellow, alunite in cyan, silica in red, and buddingtonite in dark blue. (c) Field photographs of the area showing Joshua trees with rhyolite in the background. (d) Field photograph showing large deposits of white alunite in the front and hills consisting of kaolinite in the middle.

used AVIRIS to map potassium metasomatism, a type of alteration marked by addition of large amounts of potassium at the expense of sodium. These authors mapped hematite as a proxy for potassium since potassium itself has no distinct absorption features to be evaluated. Van der Meer et al.[153] used data from the Chinese MAIS to map surface mineralogy in the Jinchuan ultramafic intrusion. This work contributed to prospecting for nickel–copper sulfide mineralizations in the Gansu province of China. Another case of hydrothermal alteration mapping in rhyolites and acid volcanics is presented in Loercher et al.[154] who studied the Torfajokull volcanic complex of Iceland using AVIRIS and map zonation patterns related to active geothermal fields. Mapping of mineral-bound ammonium (buddingtonite) in hydrothermally altered rocks to permit the determination of fluid chemistry in fossil hot springs was demonstrated by Baugh and Kruse[155] for the southern Cedar mountains using AVIRIS data.

9.4 Mineral Mapping: Lithology

An example of remote mineralogic and lithologic mapping using AVIRIS is presented for the Ice River Alkaline complex of British Columbia (Canada) by Bowers and Rowan.[156] These authors use spectral unmixing techniques to map mineralogy related to shales, slates and limestones incorporating vegetation and weathering products into the analysis. Integration of optical, thermal and radio detection and ranging (RADAR) data from AVIRIS, TIMS, and airborne synthetic aperture RADAR (AIRSAR) by Keirein-Young[157] shows the synergy of such an approach while mapping surface geology and morphology in alluvial fans in the Death Valley area (California, USA). Alternatively, geological imaging spectrometry may be used to assist metamorphic facies mapping for regional geological investigations (e.g. Rowan et al.,[158] Van der Meer[159]). Through the analysis of the position of the absorption band in the 2.3 μm wavelength range it has been shown that mapping calcite versus dolomite is possible from imaging spectrometer data thus allowing us to map dolomitization patterns from space.[160,161] This has potential for petroleum exploration since porosity increase due to dolomitization makes these rocks potential reservoir rocks. Mustard[162] used the relationship between soil, grass, and bedrock to map in a serpentinite melange using AVIRIS data. A last example comes from Lang and Cabral-Canoz[163] who use AVIRIS data for tectonostratigraphic mapping in the Northern Guerrero state of southern Mexico.

9.5 Vegetation Stress and Geobotany

Indirect detection of mineral deposits using imaging spectrometer data has been attempted through analysis of spectra of ''stressed vegetation''. Collins et al.[164] were the first to report the shift of the red edge (e.g. the point of maximum slope on the reflectance spectrum of vegetation between red and NIR wavelengths) toward the blue end of the spectrum as a result of stress due to copper in the subsurface. These authors claimed that the position and shape of the red edge could be used as a guide in mineral prospecting. Their work has led to a still ongoing debate on the potential use of the red-edge shift. Although the amount and direction of shift as a result of geochemical stress on vegetation is uncertain, most workers agree that the red edge can be used to asses the vitality of plant communities (see, for example, Boochs et al.[165] for a discussion on this topic). Work on the use of vegetation stress indicators derived from AVIRIS data can be found in Clark et al.[166] and Lelong et al.[167] An overview of techniques for delineating the red edge can be found in Dawson and Curran.[168]

9.6 Environmental Geology

High-spectral-resolution data of the Geoscan Mk II were used to delineate mercury-contaminated mill tailings within an Environmental Protection Agency (EPA) Superfund site in north central Nevada where surface geochemical data could be used for validation of the results.[169] Similarly, Lehmann et al.[170] investigated GER spectra of a vegetation-covered mine waste deposit. Farrand and Harsanyi[116] map the distribution of mine tailings in the Coeur d'Alene River valley, Idaho, through their CEM Technique. Farrand[171] used AVIRIS data to map pernicious trace metals, ferric oxides, and oxyhydroxide minerals in an acid water environment. The trace metals are released into the environment as a result of mining operations. Traditionally, such phases are mapped in the field using geochemical sampling and laboratory analysis. Bianchi et al.[172] used data from the MIVIS to map oil spills over land by relating the parts per million of oil with the oil hyperspectral information gathered by the imaging sensor. Their study focuses on the oil spilled during the blow-out from an Agenzia Generale Italiana Petroli (AGIP) rig located within the Ticino Regional park in Italy. Mapping of abundance images of goethite, jarosite, hematite, kaolinite, muscovite, montmorillonite along with vegetation and soil from AVIRIS over the Leadville area of Colorado aided in an environmental hazard assessment while muscovite/illite mostly corresponded to mine tailings and iron minerals often were associated with dark smelter slag materials.[173] The combination of airborne Daedalus-ATM (Airborne Thematic Mapper) scanner data and imaging spectrometer (reflective and thermal) data from DAIS was used to map the area of a former uranium mining site in eastern Germany.[174] Mapping of temperature anomalies and

differentiation of surface materials allowed assessment of the hydrologic conditions at the site and provided information relevant for recultivation. In a paper by Clark et al.[175] a US Geological Survey (USGS) initiative is outlined for monitoring of superficial geology, vegetation communities, and environmental materials in the US National Parks to aid in ecosystems environmental mapping.

9.7 Petroleum-related Studies

Bammel and Birnie[176] studied the spectral reflectance response of big sagebrush to hydrocarbon-induced vegetation stress in the Bighorn basin of Wyoming and found a significant blue shift (to shorter wavelengths) of the green reflectance peak at 0.56 μm and a red trough (0.67 μm) as stress indicators. However, these authors concluded that the correlation found between geobotanical anomalies in relation to surface or subsurface hydrocarbons could not be demonstrated with statistical significance. Yang et al.[177] surveyed two profiles in an oil-rich basin in the delta of the Yellow river in China. Field spectral measurements of monocultural winter wheat over areas known to have light hydrocarbon gas seepages constrained from subsurface gas measurements showed an anomalous red shift (contrary to the result of Bammel and Birnie[176]) of the red edge of the vegetation spectrum. Although a visual relationship between gas measurements and spectral anomalies could be seen, no statistical significant correlation could be demonstrated. Mapping of dolomitization patterns through mapping dolomite versus calcite concentration is a goal of remote sensing since dolomites have a 12% greater porosity than limestones containing calcite, thus making them interesting host rocks for oil accumulation.

Windeler and Lyon[160] demonstrated the possibility of separating calcite from dolomite in Geoscan airborne data although their study centered on mineral prospecting in skarn-type deposits. In a series of publications, Van der Meer[161,178] demonstrated that imaging the shift of the carbonate absorption would allow mapping of compositional variations in dolomite and calcite percentages. Hydrocarbon soil geochemistry and airborne spectrometer data were integrated with Landsat TM imagery for a portion of the Sao Francisco basin of Central Brazil by De Oliveira and Crosta.[179] This approach allowed these authors to map both soil and vegetation anomalies related to hydrocarbon microseepages.

9.8 Atmospheric Effects Resulting from Geological Processes

Most studies use imaging spectrometer data calibrated to reflectance. As discussed previously, raw imaging spectrometer radiance data show absorption features that can be attributed to atmospheric gases. In contrast, these absorption features can be used to map quantities and differences in these gases in the atmospheric column. Using ratios of absorption band depth estimates of water bands at 0.95 μm and 1.15 μm, it has been demonstrated that water vapor total column abundance can be mapped from imaging spectrometer data.[180–183] Similarly, De Jong and Chrien[184] and De Jong[185] detected abnormally high abundances of carbon dioxide and methane in AVIRIS data from the Mammoth Mountain area which they attributed to renewed volcanic activity.

9.9 Volcanic Hazards

Hydrothermally altered rocks on stratovolcanoes are closely linked to edifice failures and the generation of destructive debris flows while altered rocks form zones of weakness along fractures and contain hydrous clay minerals that modify the physical properties of derbis flows.[186] Mapping alteration minerals on the slopes of Mount Rainier using AVIRIS data allowed Crowley and Zimbelman[187] to delineate hazardous sectors and develop a method also applicable to other areas.

9.10 Thermal Infrared Studies

Geological studies using TIR multispectral data date back to the early 1980s when Kahle and Goetz[188] and Kahle[189] derived emissivity information and thermal inertia from the first TIMS data and used it to map mineralogy and some lithologies. Following on from these studies were the first examples that showed that the age of lava flows could be constrained from thermal data.[190] In 1992, Hook et al.[126] presented a comparison of techniques for extracting emissivity information from TIR data for geological studies. Abrams and Hook[141] use the TIMS data to simulate ASTER data and evaluate the use of this data for geological mapping purposes (Figure 17). A comparison of TIR emissivity spectra measured in situ and in the laboratory, and similar spectra derived from NASA's airborne TIMS in Cuprite (Nevada) using spectral matching techniques, have demonstrated the complementary nature of the thermal region and the reflective region of the EM spectrum.[191] Many silicate minerals that have no diagnostic absorption features in the reflective part of the spectrum and which can therefore be difficult to map using VIS/SWIR spectroscopy show strong features in the thermal part of the spectrum, thus allowing more accurate mapping with thermal spectrometers such as TIMS. TIR spectra of TIMS acquired over sedimentary rocks in the Tarim basin allowed Bihong and Xiaowei[192] to map sandstones, siltstones, argillites, and carbonates accurately using decorrelation stretching. In a decorrelation-stretched color-composite TIMS image, temperature

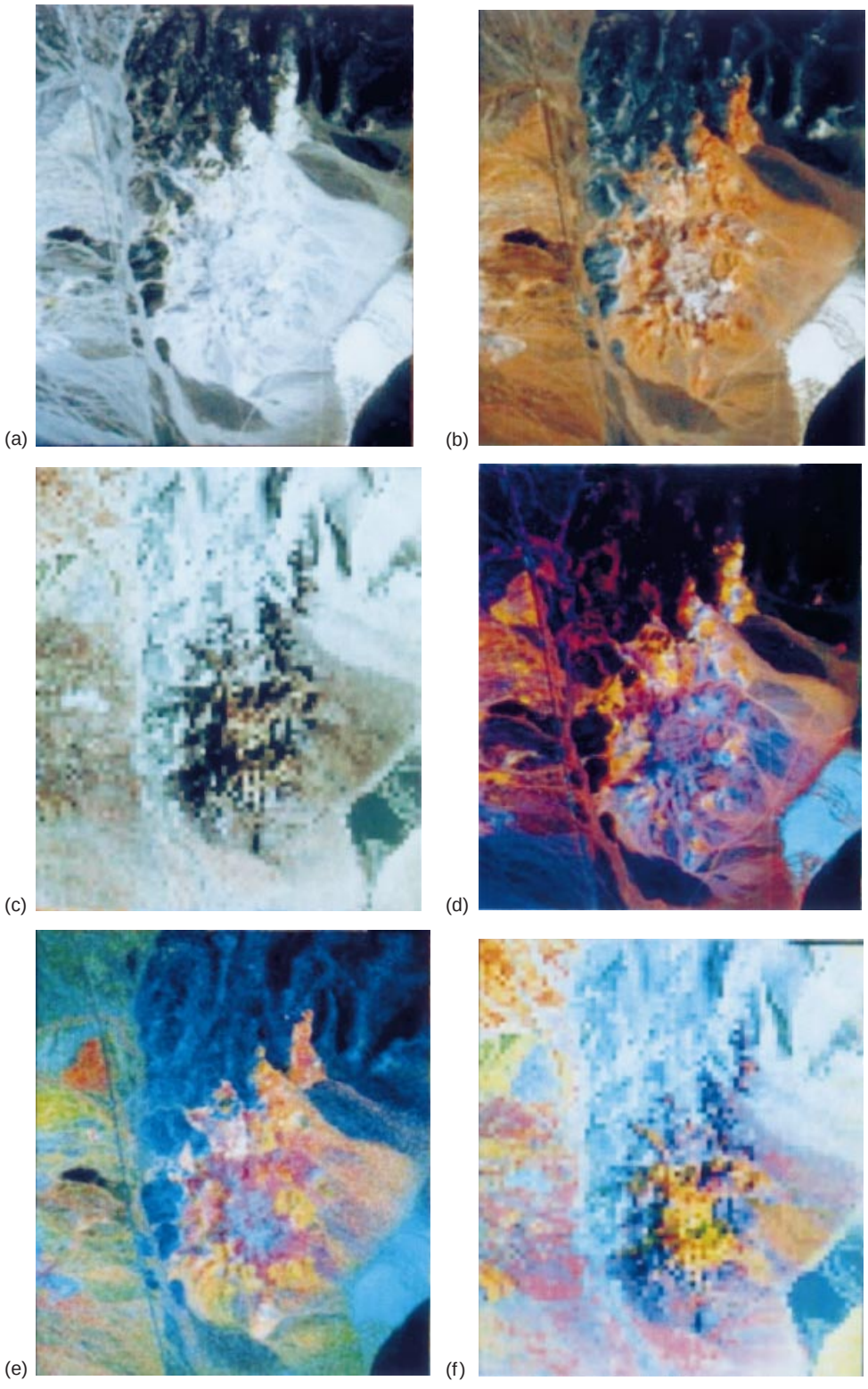

Figure 17 Linearly stretched (a–c) and decorrelation stretched (d–f) color composites of simulated ASTER data. (a) and (d) display bands 3, 2, and 1 in RGB; (b) and (e) display bands 4, 5, and 9 in RGB; and (c) and (f) display bands 13, 12, and 10 in RGB. The ASTER bands are 1, 0.52–0.60 μm; 2, 0.63–0.69 μm; 3, 0.76–0.86 μm; 4, 1.6–1.7 μm; 5, 2.145–2.185 μm; 9, 2.36–2.43 μm; 10, 8.125–8.475 μm; 12, 8.925–9.275 μm; 13, 10.25–10.95 μm. (Reproduced by permission of IEEE (© 1995, IEEE) from Abrams and Hook, *IEEE Trans. Geosci. Remote Sensing*.[141])

variations show as intensity differences while emissivity differences show as color variations in the image.[193]

ACKNOWLEDGMENTS

The author thanks Dieter Oertel, Peter Strobl and Harald Rothfuss from the Deutsche Forschungsanstalt führ Luft- und Raumfahrt e.V. (DLR), Steven de Jong from the University of Utrecht and IEEE for allowing the use of some of their view graphs for illustration. Peter Strobl, Dieter Oertel (DLR), Paul Curran (University of Southampton), Steven de Jong (University of Utrecht) and Harold Lang (NASA Jet Propulsion Laboratory) are thanked for reviewing the manuscript.

ABBREVIATIONS AND ACRONYMS

ADS	Absolute Diffuse Source
AGIP	Agenzia Generale Italiana Petroli
AHS	Airborne Hyperspectral Scanner
AIRSAR	Airborne Synthetic Aperture RADAR
AIS	Airborne Imaging Spectrometer
AMSS	Airborne Multispectral Scanner
ARCP	Absolute Radiometric Calibration Part
ARIES	Australian Resource Information and Environment Satellite
ARS	Average Reference Spectrum
ASTER	Advanced Spaceborne Thermal Emission and Reflectance Radiometer
ATM	Airborne Thematic Mapper
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BRDF	Bidirectional Reflectance Distribution Function
CASI	Compact Airborne Spectrographic Imager
CCSM	Cross-correlogram Spectral Matching
CEM	Constrained Energy Minimization
CIS	Chinese Imaging Spectrometer
CSIRO	Commonwealth Scientific and Industrial Research Organization
DAIS	Digital Airborne Imaging Spectrometers
DEM	Digital Elevation Model
DGPS	Differential Global Positioning System
DLR	Deutsche Forschungsanstalt führ Luft-und Raumfahrt e.V.
DN	Digital Number
EL	Empirical Line
EM	Electromagnetic
EPA	Environmental Protection Agency
ESA	European Space Agency
FLI	Fluorescence Line Imager
FOV	Field-of-view
Geoscan	Geological Scanner
GER	Geophysical Environmental Research
GERIS	Geophysical Environmental Research Imaging Spectrometer
GIFOV	Ground Instantaneous Field-of-view
GLI	Global Imager
GPS	Global Positioning System
HIRIS	High-resolution Imaging Spectrometer (NASA)
HRIS	High-resolution Imaging Spectrometer (ESA)
HSI	Hyperspectral Imager
HYDICE	Hyperspectral Digital Imagery Collection Experiment
HyMAP	Hyperspectral Mapping
IARR	Internal Average Relative Reflectance
IFOV	Instantaneous Field-of-view
IR	Infrared
IS	Integrating Sphere
ISM	Imaging Spectroscopic Mapper
LOWTRAN	Low-resolution Transmittance
MAF	Minimum/Maximum Autocorrelation Factors
MAIS	Modular Airborne Imaging Spectrometer
MCT	Mercury Cadmium, Telluride Infrared Detectors
MERIS	Medium-resolution Imaging Spectrometer
MIR/TIR	Middle Infrared/Thermal Infrared
MIVIS	Multispectral Infrared and Visible Imaging Spectrometer
MNF	Maximum Noise Fraction
MODIS	Moderate Resolution Imaging Spectroradiometer
MODTRAN	Moderate-resolution Transmittance
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency of Japan
NER	Noise Equivalent Radiance
NET	Noise Equivalent Temperature
NEΔT	Noise Equivalent Temperature Difference
NIR	Near-infrared
PCA	Principal Component Analysis
PDF	Probability Density Function
PLI	Programmable Line Imager
PPI	Pixel Purity Index

PRISM	Process Research by an Imaging Space Mission
PSF	Point Spread Function
RADAR	Radio Detection and Ranging
RBD	Relative Absorption Band-depth
RMS	Root Mean Square
ROSIS	Relative Optics System Imaging Spectrometer
RRCP	Relative Radiometric Calibration Part
RT	Radiative Transfer
SAM	Spectral Angle Mapping
SCP	Spectrometric/geometric Calibration Part
SFF	Spectral Feature Fitting
SFSI	SWIR Full Spectrographic Imager
SMIRR	Shuttle Multispectral Infrared Radiometer
SPOT	Système Probatoire de l'Observation de la Terre
SRF	Spectral Response Function
SWIR	Shortwave Infrared
TACP	Thermal Absolute Calibration Part
TDI	Thermal-delay Integration Imager
TIMS	Thermal Infrared Multispectral Scanner
TIR	Thermal Infrared
TM	Thematic Mapper
TRWIS	TRW Imaging Spectrometer
USGS	US Geological Survey
VIS/SWIR	Visible to Shortwave Infrared
VIS/TIR	Visible to Thermal Infrared
VNIR	Visible and Near-infrared
WIS	Wedge-imaging Spectrometer

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Field-portable Instrumentation **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

Remote Sensing **(Volume 10)**
Hyperspectral Remote Sensing: Data Collection and Exploitation • Satellite and Sensor Systems for Environmental Monitoring

Infrared Spectroscopy **(Volume 12)**
Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

REFERENCES

1. R.G. Burns, *Mineralogical Application to Crystal Field Theory*, Cambridge University Press, Cambridge, 1970.
2. J.B. Adams, 'Visible and Near-infrared Diffuse Reflectance: Spectra of Pyroxenes as Applied to Remote Sensing of Solid Objects in the Solar System', *J. Geophys. Res.*, **79**, 4829–4836 (1974).
3. J.B. Adams, 'Interpretation of Visible and Near-infrared Diffuse Reflectance Spectra of Pyroxenes and other Rock Forming Minerals', in *Infrared and Raman Spectroscopy of Lunar and Terrestrial Materials*, ed. C. Karr, Academic Press, New York, 91–116, 1975.
4. G.R. Hunt, 'Spectral Signatures of Particulate Minerals in the Visible and Near-infrared', *Geophysics*, **42**, 501–513 (1977).
5. G.R. Hunt, J.W. Salisbury, 'Visible and Near-infrared Spectra of Minerals and Rocks, I. Silicate Minerals', *Modern Geology*, **1**, 283–300 (1970).
6. A.F.H. Goetz, 'Imaging Spectrometry for Studying Earth, Air, Fire and Water', *EARSel Adv. Remote Sensing*, **1**, 3–15 (1991).
7. A.F.H. Goetz, L.C. Rowan, 'Geologic Remote Sensing', *Science*, **211**, 781–791 (1981).
8. G.R. Hunt, 'Electromagnetic Radiation: the Communication Link in Remote Sensing', in *Remote Sensing in Geology*, eds. B. Siegal, A. Gillespie, Wiley, New York, 1980.
9. A.F.H. Goetz, G. Vane, J.E. Solomon, B.N. Rock, 'Imaging Spectrometry for Earth Remote Sensing', *Science*, **228**, 1147–1153 (1985).
10. H.Y. Chiu, W. Collins, 'A Spectroradiometer for Airborne Remote Sensing', *Photogramm. Eng. Remote Sensing*, **44**, 507–517 (1978).
11. G. Vane, A.F.H. Goetz, 'Terrestrial Imaging Spectroscopy', *Remote Sensing Environ.*, **24**, 1–29 (1988).
12. C. LaBaw, 'Airborne Imaging Spectrometer 2: The Optical Design', in *Proceedings of the Society of Photo-optical Instrumentation Engineers*, eds. G. Vane, A.F.H. Goetz, D.D. Norris, SPIE, Vol. 834 (1987).
13. G. Vane, R.O. Green, T.G. Chrien, H.T. Enmark, E.G. Hansen, W.M. Porter, 'The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sensing Environ.*, **44**, 127–143 (1993).
14. A.F.H. Goetz, L.C. Rowan, M.J. Kingston, 'Mineral Identification from Orbit: Initial Results from the Shuttle Multispectral Infrared Radiometer', *Science*, **218**, 1020–1031 (1982).
15. G. Vane, A.F.H. Goetz, 'Terrestrial Imaging Spectrometry: Current Status, Future Trends', *Remote Sensing Environ.*, **44**, 117–126 (1993).
16. G.R. Hunt, J.W. Salisbury, 'Visible and Near-infrared Spectra of Minerals and Rocks, II. Carbonates', *Modern Geology*, **2**, 23–30 (1971).
17. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, III.

Oxides and Hydroxides', *Modern Geology*, **2**, 195–205 (1971).
18. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, IV. Sulfides and Sulfates', *Modern Geology*, **3**, 1–14 (1971).
19. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, V. Halides, Arsenates, Vanadates, and Borates', *Modern Geology*, **3**, 121–132 (1972).
20. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, VI. Additional Silicates', *Modern Geology*, **4**, 85–106 (1973).
21. R.N. Clark, T.V.V. King, M. Kleijwa, G.A. Swayze, N. Vergo, 'High Spectral Resolution Reflectance Spectroscopy of Minerals', *J. Geophys. Res.*, **95**, 12653–12680 (1990).
22. J.K. Crowley, 'Visible and Near-infrared (0.4–2.5 μm) Reflectance Spectra of Playa Evaporite Minerals', *J. Geophys. Res.*, **96**, 16231–16240 (1991).
23. S.J. Gaffey, L.A. McFadden, D. Nash, C.M. Pieters, 'Ultraviolet, Visible, and Near-infrared Reflectance Spectroscopy: Laboratory Spectra of Geologic Materials', in *Remote Geochemical Analysis: Elemental and Mineralogical Composition*. eds. C.M. Pieters, P.A.J. Englert, Cambridge University Press, Cambridge, 43–78, 1993.
24. C.I. Grove, S.J. Hook II, E.D. Paylor, 'Laboratory Reflectance Spectra of 160 Minerals, 0.4 to 2.5 Micrometers', NASA-JPL Publication, 92–2, Pasadena, 300, 1992.
25. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, VII. Acidic Igneous Rocks', *Modern Geology*, **4**, 217–224 (1973).
26. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, VIII. Intermediate Igneous Rocks', *Modern Geology*, **4**, 237–244 (1973).
27. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, IX. Basic and Ultrabasic Igneous Rocks', *Modern Geology*, **5**, 15–22 (1974).
28. G.R. Hunt, J.W. Salisbury, C.J. Lenhoff, 'Visible and Near-infrared Spectra of Minerals and Rocks, X. Stoney Meteorites', *Modern Geology*, **5**, 115–126 (1975).
29. G.R. Hunt, J.W. Salisbury, 'Visible and Near-infrared Spectra of Minerals and Rocks: XI. Sedimentary Rocks', *Modern Geology*, **5**, 211–217 (1976).
30. G.R. Hunt, J.W. Salisbury, 'Visible and Near-infrared Spectra of Minerals and Rocks: XII. Metamorphic Rocks', *Modern Geology*, **5**, 219–228 (1976).
31. D.B. Nash, J.E. Conel, 'Spectral Reflectance Systematics for Mixtures of Powdered Hypersthene, Labradorite, and Ilmenite', *J. Geophys. Res.*, **79**, 1615–1621 (1974).
32. R.B. Singer, 'Near-infrared Spectral Reflectance of Mineral Mixtures: Systematic Combinations of Pyroxenes, Olivine, and Iron Oxides', *J. Geophys. Res.*, **86**, 7967–7982 (1981).
33. M.W. Smith, 'Design and Initial Performance Evaluation of a Portable Short Wave Infrared Spectrometer', *Proceedings SPIE*, San Diego, Ca., Vol. 1762-14, 1992.
34. B. Curtiss, A.F.H. Goetz, 'Field Spectrometry: Techniques and Instrumentation', *International Symposium on Spectral Sensing Research*, July, 1994.
35. R.L. Pearson, L.D. Miller, C.J. Tucker, 'Hand-held Spectral Radiometer to Estimate Gramineous Biomass', *Appl. Opt.*, **15**, 416–418 (1976).
36. G. Hoover, A.B. Kahle, 'A Thermal Emission Spectrometer for Field Use', *Photogramm. Eng. Remote Sensing*, **53**, 627–632 (1987).
37. C.D. Elvidge, 'Visible and Near-infrared Reflectance Characteristics of Dry Plant Material', *Int. J. Remote Sensing*, **11**(10), 1775–1795 (1990).
38. M.F. Baumgardner, E.R. Stoner, L.F. Silva, L.L. Biehl, 'Reflectance Properties of Soils', in *Advances of Agronomy*, ed. N. Brady, Academic Press, New York, 1–44, 1985.
39. H.R. Condit, 'The Spectral Reflectance of American Soils', *Photogramm. Eng. Remote Sensing*, **36**, 955–966 (1970).
40. J.W. Salisbury, D.M. D'Aria, 'Emissivity of Terrestrial Materials in the 8–14 μm Atmospheric Window', *Remote Sensing Environ.*, **42**, 83–106 (1992).
41. J.K. Buettner, C.D. Kern, 'The Determination of Infrared Emissivities of Terrestrial Surfaces', *J. Geophys. Res.*, **70**, 1329–1337 (1965).
42. C.D. Elvidge, 'Thermal Infrared Reflectance of Dry Plant Materials', *Remote Sensing Environ.*, **26**, 265–285 (1988).
43. R.J.P. Lyon, 'Analysis of Rocks by Spectral Infrared Emission (8 to 25 microns)', *Econ. Geol.*, **60**, 715–736 (1965).
44. L.W. Pinkley, D. Williams, 'Optical Properties of Seawater in the Infrared', *J. Opt. Soc. Am.*, **66**, 554–558 (1976).
45. J.W. Salisbury, D.M. D'Aria, 'Infrared (8–14 μm) Remote Sensing of Soil Particle Size', *Remote Sensing Environ.*, **42**, 157–165 (1992).
46. J.W. Salisbury, N.M. Milton, 'Thermal Infrared (2.5 to 13.5 μm) Directional Hemispherical Reflectance of Leaves', *Photogramm. Eng. Remote Sensing*, **54**, 1301–1304 (1988).
47. J.W. Salisbury, L.S. Walter, 'Thermal Infrared (2.5 to 13.5 μm) Spectroscopic Remote Sensing of Igneous Rock types on Particulate Planetary Surfaces', *J. Geophys. Res.*, **94**, 9192–9202 (1989).
48. J.W. Schaaf, D. Williams, 'Optical Constants of Ice in the Infrared', *J. Opt. Soc. Am.*, **63**, 726–732 (1973).
49. R.K. Vincent, L.C. Rowan, R.E. Gillespie, C. Knapp, 'Thermal-infrared Spectra and Chemical Analyses of Twenty-six Igneous Rock Samples', *Remote Sensing Environ.*, **4**, 199–209 (1975).

50. L.S. Walter, J.W. Salisbury, 'Spectral Characterization of Igneous Rocks in the 8 to 12 μm Region', *J. Geophys. Res.*, **94**, 9203–9213 (1989).
51. C.L. Wong, W.R. Blevin, 'Infrared Reflectances of Plant Leaves', *Aust. J. Biol. Sci.*, **20**, 501–527 (1967).
52. J.W. Salisbury, 'Mid-infrared Spectroscopy: Laboratory Data'. in *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, eds. C.M. Pieters, P.A.J. Englert, Cambridge University Press, Cambridge, 79–98, 1993.
53. B.G. Henderson, P.G. Lucey, B.M. Jakosky, 'New Laboratory Measurements of Mid-IR Emission Spectra of Simulated Planetary Surfaces', *J. Geophys. Res.*, **101**, 14969–14975 (1996).
54. K. Staenz, 'A Decade of Imaging Spectrometry in Canada', *Can. J. Remote Sensing*, **18**(4), 187–197 (1992).
55. R.O. Green, M.L. Eastwood, C.M. Sarture, T.G. Chrien, M. Aronsson, B.J. Chippendale, J.A. Faust, B.E. Pavri, C.J. Chovit, M. Solis, M.R. Olah, O. Williams, 'Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sensing Environ.*, **65**(3), 227–248 (1998).
56. S. Chang, M.J. Westfield, F. Lehmann, D. Oertel, R. Richter, 'A 79-Channel Airborne Imaging Spectrometer', *Proceedings of SPIE*, Vol. 1937, 164–172 (1993).
57. W. Collins, S. Chang, 'The Geophysical Environmental Research Corp. 63 Channel Airborne Imaging Spectrometer and 12 Band Thermal Scanner', *Proceedings of SPIE*, Vol. 1298, 62–71 (1990).
58. P. Strobl, A. Nielsen, F. Lehmann, R. Richter, A. Mueller, 'DAIS System Performance, First results from the 1995 Evaluation Campaigns', In: Proceedings 2nd International Airborne Remote Sensing Conference and Exhibition, San Francisco, Vol. II, 325–334, 1996.
59. A.H. Goetz, M. Herring, 'The High Resolution Imaging Spectrometer for EOS', *IEEE Trans. Geosci. Remote Sensing*, **27**(2), 136–144 (1989).
60. J.P. Kerekes, D.A. Landgrebe, 'Parameter Trade-offs for Imaging Spectroscopy Systems', *IEEE Trans. Geosci. Remote Sensing*, **29**(1), 57–65 (1991).
61. A.B. Kahle, F.D. Palluconi, S.J. Hook, V.J. Realmuto, G. Bothwell, 'The Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER)', *Int. J. Imag. Syst. Technol.*, **3**, 144–156 (1991).
62. H. Fujisada, F. Sakuma, A. Ono, M. Kudoch, 'Design and Preflight Performance of ASTER Instrument Protoflight Model', *IEEE Trans. Geosci. Remote Sensing*, **36**(4), 1152–1160 (1998).
63. C.O. Justice, E. Vermote, J.R.G. Townshend, R. Defries, D.P. Roy, D.K. Hall, V.V. Salomonson, J.L. Privette, G. Riggs, A. Strahler, W. Lucht, R.B. Myneni, Y. Knyazikhin, S.W. Running, R.R. Nemani, Z. Wan, A.R. Huete, W. van Leeuwen, R.E. Wolfe, L. Giglio, J.P. Muller, P. Lewis, M.J. Barnsley, 'The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research', *IEEE Trans. Geosci. Remote Sensing*, **36**(4), 1228–1249 (1998).
64. M. Rast, 'ESA's Activities in the Field of Imaging Spectroscopy', in *Imaging Spectroscopy: Fundamentals and Prospective Applications*. eds. F. Toselli, J. Bodechtel, Kluwer Academic Publishers, Dordrecht, 167–191, 1992.
65. W. Posselt, P. Kunkel, E. Schmidt, U. Del Bello, R. Meynart, 'Process Research by an Imaging Space Mission', In: Proceedings of the Advanced and Next-generation Satellites II, eds. H. Fuijisada, G. Calamai, M. Sweeting. SPIE, Toarmina, September 1996, Vol. 2957, 1996.
66. M. Rast, J.L. Bézy, 'The ESA Medium Resolution Imaging Spectrometer (MERIS): Requirements to its Mission and Performance of its System', In: Proceedings of the RSS'95 Remote Sensing in Action, eds. P.J. Curran, C. Robertson, 11–14 September 1995, Southampton, UK, 125–132, 1995.
67. A.F.H. Goetz, 'Principles of Narrow Band Spectrometry in the Visible and IR: Instruments and Data Analysis', in *Imaging Spectroscopy: Fundamentals and Prospective Applications*, eds. F. Toselli, J. Bodechtel, Kluwer Academic Publishers, Dordrecht, 21–32, 1992.
68. P. Strobl, 'DAIS 7915 Laboratory Calibration', Presented at the First DAIS 7915 LSF Workshop, Oberpfaffenhofen, 10–12 October 1996.
69. P. Strobl, 'DAIS 7915 Data facility, Pre-Processing and Archiving', Presented at the First DAIS 7915 LSF Workshop, Oberpfaffenhofen, 10–12 October 1996.
70. R.O. Green, 'Determination of the In-flight Spectral and Radiometric Characteristics of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', in *Imaging Spectrometry: Fundamentals and Prospective Applications*, eds. F. Toselli, J. Bodechtel, Kluwer Academic Publishers, Dordrecht, 103–123, 1992.
71. R.O. Green, G. Vane, J.E. Conel, 'Determination of In-flight AVIRIS Spectral, Radiometric, Spectral and Signal-to-noise Characteristics using Atmospheric and Surface Measurements from the Vicinity of the Rare-earth-bearing Carbonatite at Mountain Pass, California', In: Proceedings of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Performance Evaluation Workshop, ed. G. Vane, Pasadena, USA, NASA-JPL Publication 88-38, 162–184, 1988.
72. D. Oertel, F. Lehmann, S.H. Chang, 'A New 79-Channel Airborne Imaging Spectrometer: Its Functioning, Main Characteristics and Calibration', Presented at the ISPRS Workshop 'From Pixels to Sequences' Zurich, March 22–24, 1995.
73. P. Strobl, A. Müller, D. Schläpfer, M. Schaepman, 'Laboratory Calibration and Inflight Validation of the Digital Airborne Imaging Spectrometer DAIS 7915 for the 1996 Flight Season', in *Algorithms for Multispectral and Hyperspectral Imagery III*, SPIE Vol. 3071, Orlando, 20–25 April 1997.
74. P. Meyer, 'A Parametric Approach for the Geocoding of Airborne Visible/Infrared Imaging Spectrometer

(AVIRIS) Data in Rugged Terrain', *Remote Sensing Environ.*, **49**(2), 118–130 (1994).
75. B.C. Gao, 'An Operational Method for Estimating Signal-to-Noise Ratios from Data Acquired with Imaging Spectrometers', *Remote Sensing Environ.*, **43**, 23–33 (1993).
76. P.J. Curran, J.L. Dungan, 'Estimation of Signal-to-Noise: A New Procedure Applied to AVIRIS Data', *IEEE Trans. Geosci. Remote Sensing*, **27**(5), 620–628 (1989).
77. H. Rothfuss, 'Verarbeitung und einsatzt abbildender spektometerdaten (GER) mit unterstutzenden bodenmessungen zur erkundung einer landwirtschaftlich genutzten altlastflache' VDI Verlag, Dusseldorf, 1994.
78. R.N. Clark, E.C. Livo, R.F. Kokaly, 'Geometric Correction of AVIRIS Imagery Using On-Board Navigation and Engineering Data' In: Summaries of the 7th Annual JPL Airborne Earth Science Workshop, ed. R.O. Green, Jan 12–14, Pasadena, CA, 1998.
79. A.A. Green, M. Berman, P. Switzer, M.D. Graig, 'A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal', *IEEE Trans. Geosci. Remote Sensing*, **26**, 65–74 (1988).
80. M. Rast, S.J. Hook, C.D. Elvidge, R.E. Alley, 'An Evaluation of Techniques for the Extraction of Mineral Absorption Features from High Spectral Resolution Remote Sensing Data', *Photogramm. Eng. Remote Sensing*, **57**(10), 1303–1309 (1991).
81. D.A. Roberts, Y. Yamaguchi, R.J.P. Lyon, 'Calibration of Airborne Imaging Spectrometer Data to Percent Reflectance using Field Spectral Measurements', In: Proceedings of the 19th International Symposium on Remote Sensing of Environment, ERIM, 21–25 October, Ann Arbor, 679–688, 1985.
82. F. van der Meer, 'Calibration of Airborne Visible/Infrared Imaging Spectrometer Data (AVIRIS) to Reflectance and Mineral Mapping in Hydrothermal Alteration Zones: an example from the 'Cuprite Mining District'', *Geocarto Int.*, **9**(3), 23–37 (1994).
83. F.A. Kruse, G.L. Raines, K. Watson, 'Analytical Techniques for Extracting Geologic Information from Multichannel Airborne Spectroradiometer and Airborne Imaging Spectrometer Data', In: Proceedings of the Fourth Thematic Conference on Remote Sensing for Exploration Geology, ERIM, 1–4 April, San Francisco, 309–324, 1985.
84. F.A. Kruse, 'Use of Airborne Imaging Spectrometer Data to Map Minerals Associated with Hydrothermally Altered Rocks in the Northern Grapevine Mountains, Nevada, and California', *Remote Sensing Environ.*, **24**, 31–51 (1988).
85. E. Ben-Dor, F.A. Kruse, 'The Relationship between the Size of Spatial Subsets of GER 63 Channel Scanner Data and the Quality of the Internal Average Relative Reflectance (IARR) Atmospheric Correction Technique', *Int. J. Remote Sensing*, **15**, 683–690 (1994).
86. F.A. Kruse, K.S. Keirein-Young, J.W. Boardman, 'Mineral Mapping at Cuprite, Nevada with a 63-channel Imaging Spectrometer', *Photogramm. Eng. Remote Sensing*, **56**, 83–92 (1990).
87. R.G. Isaacs, W.C. Wang, S. Goldenberg, 'Multiple Scattering LOWTRAN and FASCODE models', *Appl. Opt.*, **26**, 1272–1281 (1987).
88. A. Berk, L.S. Bernstein, D.C. Robertson, 'MODTRAN: A Moderate Resolution Model for LOWTRAN 7', GL-TR-89-0122, *Geophys. Lab.*, Beford, MA, 1989.
89. A. Berk, L.S. Bernstein, G.P. Anderson, P.K. Acharya, D.C. Robertson, J.H. Chetwynd, S.M. Adler-Golden, 'MODTRAN Cloud and Multiple Scattering Upgrades with Application to AVIRIS', *Remote Sensing Environ.*, **65**(3), 367–375 (1998).
90. D. Tanre, 'Description of a Computer Code to Simulate the Satellite Signal in the Solar Spectrum: the 5S code', *Int. J. Remote Sensing*, **11**, 659–668 (1990).
91. R. Richter, 'Atmospheric Correction of DAIS Hyperspectral Data', *Comp. Geosci.*, **22**, 785–793 (1996).
92. B.C. Gao, K.B. Heidebrecht, A.F.H. Goetz, 'Derivation of Scaled Surface Reflectances from AVIRIS Data', *Remote Sensing Environ.*, **44**, 165–178 (1993).
93. R. Richter, 'A Spatially Adaptive Fast Atmospheric Correction Algorithm', *Int. J. Remote Sensing*, **17**, 1201–1214 (1996).
94. F. Zagolski, J.P. Gastellu-Etchegorry, 'Atmospheric Correction of AVIRIS Images with a Procedure based on the Inversion of the 5S Model', *Int. J. Remote Sensing*, **16**, 3115–3146 (1995).
95. Y.J. Kaufman, 'Atmospheric Effect on Spatial Resolution of Surface Imagery', *Appl. Opt.*, **23**, 3400–3408 (1984).
96. G. Rondeaux, M.D. Steven, J.A. Clark, G. Mackay, 'La Grau: a European Test Site for Remote Sensing Validation', *Int. J. Remote Sensing*, **19**(14), 2775–2788 (1998).
97. P.N. Slater, S.F. Biggar, R.G. Holm, R.D. Jackson, Y. Mao, M.S. Moran, J.M. Palmer, B. Yuan, 'Reflectance and Radiance-based Methods for the In-flight Absolute Calibration of Multispectral Sensor', *Remote Sensing Environ.*, **22**, 11–37 (1987).
98. X. Jia, J.A. Richards, 'Binary Coding of Imaging Spectrometer Data for Fast Spectral Matching and Classification', *Remote Sensing Environ.*, **43**, 47–53 (1993).
99. K. Okada, A. Iwashita, 'Hyper-multispectral Image Analysis based on Waveform Characteristics of Spectral Curve', *Adv. Space Res.*, **12**(7), 433–442 (1992).
100. A.A. Green, M.D. Graig, 'Analysis of Aircraft Spectrometer Data with Logarithmic Residuals', In: Proceedings of the Airborne Imaging Spectrometer Data Analysis Workshop, eds. G. Vane, A.F.H. Goetz, Pasadena, USA, NASA-JPL Publication 85-41, 111–119 (1985).
101. J.K. Crowley, D.W. Brickey, L.C. Rowan, 'Airborne Imaging Spectrometer Data of the Ruby Mountains, Montana: Mineral Discrimination using Relative

Absorption Band-depth Images', *Remote Sensing Environ.*, **29**, 121–134 (1989).
102. R.N. Clark, T.L. Roush, 'Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications', *J. Geophys. Res.*, **89**, 6329–6340 (1984).
103. R.N. Clark, A. Gallagher, G.A. Swayze, 'Material Absorption Band Depth Mapping of Imaging Spectrometer Data using a Complete Band Shape Least-squares Fit with Library Reference Spectra', In: Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, ed. R.O. Green, Pasadena, NASA-JPL Publication 90-54, 176–186, 1990.
104. J.K. Crowley, G.A. Swayze, 'Mapping Minerals, Amorphous Materials, Environmental Materials, Vegetation, Water, Ice, and other Materials: The USGS Tricorder Algorithm', In: Summaries of the Fifth Annual JPL Airborne Earth Science Workshop, JPL Publication 95-1, 39–40, 1995.
105. F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H. Goetz, 'The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data', *Remote Sensing Environ.*, **44**, 145–163 (1993).
106. R.B. Singer, T.B. McCord, 'Mars: Large Scale Mixing of Bright and Dark Surface Materials and Implications for Analysis of Spectral Reflectance', In: Proceedings of the 10th Lunar and Planetary Science Conference, Houston, 19-23 March 1979, 1835–1848 (1979).
107. J.B. Adams, M.O. Smith, P.E. Johnston, 'Spectral Mixture Modelling: a New Analysis of Rock and Soil Types at the Viking Lander 1 Site', *J. Geophys. Res.*, **91**, 8098–8112 (1985).
108. C. Ichoku, A. Karnieli, 'A Review of Mixture Modelling Techniques for Sub-pixel Land Cover Estimation', *Remote Sensing Rev.*, **13**, 161–186 (1996).
109. B. Hapke, 'Bidirectional Reflectance Spectroscopy, 1, Theory', *J. Geophys. Res.*, **86**, 3039–3054 (1981).
110. P.E. Johnson, M.O. Smith, S. Taylor-George, J.B. Adams, 'A Semiempirical Method for Analysis of the Reflectance Spectra of Binary Mineral Mixtures', *J. Geophys. Res.*, **88**, 3557–3561 (1983).
111. J.F. Mustard, C.M. Pieters, 'Quantitative Abundance Estimates from Bidirectional Reflectance Measurements', *J. Geophys. Res.*, **92**, 617–626 (1987).
112. J.W. Boardman, 'Inversion of Imaging Spectrometry Data using Singular Value Decomposition', In: Proceedings 12th Canadian Symposium on Remote Sensing IGARSS'89, Vol. 4, 2069–2072, 1989.
113. J.J. Settle, N.A. Drake, 'Linear Mixing and the Estimation of Ground Cover Proportions', *Int. J. Remote Sensing*, **14**, 1159–1177 (1993).
114. M.O. Smith, P.E. Johnston, J.B. Adams, 'Quantitative Determination of Mineral types and Abundances from Reflectance Spectra using Principal Component Analysis', *J. Geophys. Res.*, **90**, 797–804 (1985).
115. J.W. Boardman, F.A. Kruse, R.O. Green, 'Mapping Target Signatures via Partial Unmixing of AVIRIS Data', In: Proceedings of the Fifth JPL Airborne Earth Science Workshop, JPL Publication 95-1, Vol. 1, 23–26, 1995.
116. W.H. Farrand, J.C. Harsanyi, 'Mapping the Distribution of Mine Tailings in the Coeur d'Alene River Valley, Idaho, through the use of a Constrained Energy Minimization Technique', *Remote Sensing Environ.*, **59**, 64–76 (1997).
117. H. Cetin, T.A. Warner, D.W. Levandowski, 'Data Classification, Visualization, and Enhancement Using n-Dimensional Probability Density Functions (nPDF): AVIRIS, TIMS, TM, and Geophysical Applications', *Photogramm. Eng. Remote Sensing*, **59**(12), 1755–1764 (1993).
118. H. Cetin, D.W. Levandowski, 'Interactive Classification and Mapping of Multi-dimensional Remotely Sensed Data using n-dimensional Probability Density Function (nPDF)', *Photogramm. Eng. Remote Sensing*, **57**(12), 1579–1587 (1991).
119. C. Lee, D.A. Landgrebe, 'Analyzing High-dimensional Multispectral Data', *IEEE Trans. Geosci. Remote Sensing*, **31**(4), 792–800 (1993).
120. J.A. Benediktsson, 'Classification and Feature Extraction from AVIRIS Data', *IEEE Trans. Geosci. Remote Sensing*, **33**(5), 1194–1205 (1995).
121. F. van der Meer, 'Extraction of Mineral Absorption Features from High-spectral Resolution Data using Non-parametric Geostatistical Techniques', *Int. J. Remote Sensing*, **15**(11), 2193–2214 (1994).
122. F. van der Meer, 'Performance of the Indicator Classifier on Simulated Image Data', *Int. J. Remote Sensing*, **17**(3), 621–627 (1996).
123. F. van der Meer, W. Bakker, 'CCSM: Cross Correlogram Spectral Matching', *Int. J. Remote Sensing*, **18**(5), 1197–1201 (1997).
124. F. van der Meer, W. Bakker, 'Cross Correlogram Spectral Matching (CCSM): Application to Surface Mineralogical Mapping using AVIRIS Data from Cuprite, Nevada', *Remote Sensing Environ.*, **61**(3), 371–382 (1997).
125. F. van der Meer, W. Bakker, 'Validated Surface Mineralogy from High-spectral Resolution Remote Sensing: a Review and a Novel Approach applied to Gold Exploration using AVIRIS Data', *Terra Nova*, **10**(2), 112–119 (1998).
126. S.J. Hook, A.R. Gabell, A.A. Green, P.S. Kealy, 'A Comparison of Techniques for Extracting Emissivity Information from Thermal Infrared Data for Geologic Studies', *Remote Sensing Environ.*, **42**, 123–135 (1992).
127. R.J. Murphy, 'Mapping of Jasperoid in the Cedar Mountains, Utah, U.S.A., using Imaging Spectrometer Data', *Int. J. Remote Sensing*, **16**(6), 1021–1042 (1995).
128. J.L. Dwyer, F.A. Kruse, A.B. Lefkoff, 'Effects of Empirical versus Model-based Reflectance Calibration on Automated Analysis of Imaging Spectrometer Data:

a Case Study from the Drum Mountains, Utah', *Photogramm. Eng. Remote Sensing*, **61**(10), 1247–1254 (1995).
129. L.R. Gaddis, L.A. Soderblom, H.H. Kieffer, K.J. Becker, J. Torson, K. Mullins, 'Decomposition of AVIRIS Spectra: Extraction of Surface-reflectance, Atmospheric and Instrumental Components', *IEEE Trans. Geosci. Remote Sensing*, **34**(1), 163–178 (1996).
130. F.A. Kruse, A.B. Lefkoff, J.B. Dietz, 'Expert System-based Mineral Mapping in Northern Death Valley, California/Nevada, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sensing Environ.*, **44**, 309–336 (1993).
131. E. Ben-Dor, F.A. Kruse, A.B. Lefkoff, A. Banin, 'Comparison of Three Calibration Techniques for Utilization of GER 63-channel Aircraft Scanner Data of Makhtesh Ramon, Negev, Israel', *Photogramm. Eng. Remote Sensing*, **60**(11), 1339–1354 (1994).
132. E. Ben-Dor, F.A. Kruse, 'Mineral Mapping of Makhtesh Ramon, Negev, Israel using GER 63-channel Scanner Data and Linear Unmixing Techniques', In: Proceedings Tenth Thematic Conference and Workshop on Applied Geologic Remote Sensing, San Antonio, Texas, 9–12 May, Vol. I, 215–226, 1994.
133. H. Kaufman, W. Weisbrich, M. Beyth, Y. Bartov, A. Itamar, S. Ronen, U. Kafri, 'Mineral Identification using GER-II Data Acquired from Makhtesh Ramon/Negev, Israel', *EARSeL Advances in Remote Sensing*, **1**(1), 82–92 (1991).
134. J.K. Crowley, 'Mapping Playa Evaporite Minerals with AVIRIS data: a First Report from Death Valley, California', *Remote Sensing Environ.* **44**, 337–356 (1993).
135. W.H. Farrand, A. Seelos, 'Using Mineral Maps Generated from Imaging Spectrometer Data to Map Faults: an Example from Summitville, Colorado', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. II, 222–230, 1996.
136. W.M. Baugh, F.A. Kruse, W.W. Atkinson, 'Quantitative Geochemical Mapping of Ammonium Minerals in the Southern Cedar mountains, Nevada, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sensing Environ.*, **65**(3), 292–308 (1998).
137. J.P. Albers, J.H. Stewart, 'Geology and Mineral Deposits of Esmeralda County, Nevada', *Nevada Bureau of Mines and Geology Bulletin*, **78**, 1–80 (1972).
138. R.P. Ashley, 'Geologic Map and Alteration Map, Cuprite Mining District, Nevada (unpublished)', 1971.
139. M.J. Abrams, R.P. Ashley, L.C. Rowan, A.F.H. Goetz, A.B. Kahle, 'Mapping of Hydrothermal Alteration in the Cuprite Mining District, Nevada using Aircraft Scanner Images for the Spectral Region 0.46–2.36 μm', *Geology*, **5**, 713–718 (1977).
140. R.G. Resmini, M.E. Kappus, W.S. Aldrich, J.C. Harsanyi, M. Anderson, 'Mineral Mapping with Hyperspectral Digital Imagery Collection Experiment (HYDICE) Sensor Data at Cuprite, Nevada, U.S.A.', *Int. J. Remote Sensing*, **18**(7), 1553–1570 (1997).
141. M.J. Abrams, S.J. Hook, 'Simulated Aster Data for Geologic Studies', *IEEE Trans. Geosci. Remote Sensing*, **33**(3), 692–699 (1995).
142. V. Nedeljkovic, N. Pendock, 'Finding Spectral Anomalies in Hyperspectral Imagery using Nonparametric Probability Density Estimation Techniques', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. I, 197–204, 1996.
143. V. Carrere, M.J. Abrams, 'An Assessment of AVIRIS Data for Hydrothermal Alteration Mapping in the Goldfield Mining District', In: G. Vane (ed.), Proceedings of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Performance Evaluation Workshop, Pasadena, USA, NASA-JPL Publication 88-38, 134–154, 1988.
144. S.J. Hook, C.D. Elvidge, M. Rast, H. Watanabe, 'An Evaluation of Short-wave-infrared (SWIR) Data from the AVIRIS and GEOSCAN Instruments for Mineralogical Mapping at Cuprite, Nevada', *Geophysics*, **56**(9), 1432–1440 (1991).
145. A.P. Crosta, I.D.M. do Prado, M. Obara, 'The Use of Geoscan AMSS Data for Gold Exploration in the Rio Itapicuru Greenstone Belt (BA), Brazil', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. II, 205–214, 1996.
146. G. Ferrier, G. Wadge, 'The Application of Imaging Spectrometry Data to Mapping Alteration Zones associated with Gold Mineralization in Southern Spain', *Int. J. Remote Sensing*, **17**(2), 331–350 (1996).
147. A. Hutsinpiller, 'Discrimination of Hydrothermal Alteration Mineral Assemblages at Virginia City, Nevada, Using the Airborne Imaging Spectrometer', *Remote Sensing Environ.*, **24**, 53–66 (1988).
148. S.C. Feldman, J.V. Taranik, 'Comparison of Techniques for Discriminating Hydrothermal Alteration Minerals with Airborne Imaging Spectrometer Data', *Remote Sensing Environ.*, **24**, 67–83 (1988).
149. A.P. Crosta, C. Sabine, J.V. Taranik, 'Hydrothermal Alteration Mapping at Bodie, California, using AVIRIS Hyperspectral Data', *Remote Sensing Environ.*, **65**(3), 309–319 (1998).
150. G.R. Hunt, R.C. Evarts, 'The Use of Near-infrared Spectroscopy to Determine the Degree of Serpentinization of Ultramafic Rocks', *Geophysics*, **46**, 316–321 (1981).
151. F. Van der Meer, 'Estimating and Simulating the Degree of Serpentinization of Peridotites using Hyperspectral Remotely Sensed Imagery', *Nonrenewable Resources*, **4**(1), 84–98 (1995).

152. K.K. Beratan, B. Peer, N.W. Dunbar, R. Blom, 'A Remote Sensing Approach to Alteration Mapping: AVIRIS Data and Extension-related Potassium Metasomatism, Socorro, New Mexico', *Int. J. Remote Sensing*, **18**(17), 3595–3609 (1997).
153. F. van der Meer, L. Fan, J. Bodechtel, 'MAIS Imaging Spectrometer Data Analysis for Ni–Cu Prospecting in Ultramafic Rocks of the Jinchuan Group, China', *Int. J. Remote Sensing*, **18**(13), 2743–2761 (1997).
154. G. Loercher, S. Endres, S. Sommer, 'Mapping Hydrothermal Alterations in the Landmannalaugar Area (Iceland) using MAC-Europe 1991 AVIRIS data', In: Proceedings Tenth Thematic Conference and Workshop on Applied Geologic Remote Sensing, San Antonio, Texas, 9-12 May, Vol. I, 333–340, 1994.
155. W.M. Baugh, F.A. Kruse, 'Quantitative Geochemical Mapping of Ammonium Minerals using Field and Airborne Spectrometers, Cedar Mountains, Esmeralda County, Nevada', In: Proceedings Tenth Thematic Conference and Workshop on Applied Geologic Remote Sensing, San Antonio, Texas, 9–12 May, Vol. II, 304–315, 1994.
156. T.L. Bowers, L.C. Rowan, 'Remote Mineralogic and Lithologic Mapping of the Ice River Complex, British Columbia, Canada, using AVIRIS Data', *Photogramm. Eng. Remote Sensing*, **62**(12), 1379–1385 (1996).
157. K.S. Keirein-Young, 'The Integration of Optical and Radar Data to Characterize Mineralogy and Morphology of Surfaces in Death Valley, California, U.S.A.', *Int. J. Remote Sensing*, **18**(7), 1517–1541 (1997).
158. L.C. Rowan, C. Anton-Pancheco, D.W. Brickey, M.J. Kingston, A. Payas, N. Vergo, J.K. Crowley, 'Digital Classification of Contact Metamorphic Rocks in Extremadura, Spain using Landsat Thematic Mapper Data', *Geophysics*, **52**, 885–897 (1987).
159. F. van der Meer, 'Metamorphic Facies Zonation in the Ronda Peridotites Mapped from GER Imaging Spectrometer Data', *Int. J. Remote Sensing*, **17**(9), 1633–1657 (1996).
160. D.S. Windeler, R.J.P. Lyon, 'Discriminating Dolomitization of Marble in the Ludwig Skarn near Yerington, Nevada Using High-resolution Airborne Infrared Imagery', *Photogramm. Eng. Remote Sensing*, **57**, 1171–1177 (1991).
161. F. van der Meer, 'Mapping Dolomitization through a Coregionalization of Simulated Field and Image-derived Reflectance Spectra; a Proof-of-concept Study', *Int. J. Remote Sensing*, **19**(8), 1615–1620 (1998).
162. J.F. Mustard, 'Relationship of Soil, Grass, and Bedrock over the Kaweah Serpentinite Melange through Spectral Mixture Analysis of AVIRIS Data', *Remote Sensing Environ.*, **44**, 293–308 (1993).
163. H.R. Lang, E. Cabral-Canoz, 'Preliminary Analysis of AVIRIS Data for Tectonostratigraphic Assessment of Northern Guerrero State, Southern Mexico', Presented at the Geoscience workshop, 1996, Pasadena, CA, 1998.
164. W. Collins, S.H. Chang, G. Raines, F. Canney, R. Ashley, 'Airborne Biogeochemical Mapping of Hidden Mineral Deposits', *Econ. Geol.*, **78**, 737–743 (1983).
165. F. Boochs, G. Kupfer, K. Dockter, W. Kuhbauch, 'Shape of the Red Edge as Vitality Indicator for Plants', *Int. J. Remote Sensing*, **11**, 1741–1753 (1990).
166. R.N. Clark, T.V.V. King, C. Ager, G.A. Swayze, 'Vegetation Species and Stress Indicator Mapping in the San Luis Valley, Colorado using Imaging Spectrometer Data', *Remote Sensing Environ.*, submitted (1997).
167. C.C.D. Lelong, P.C. Pinet, H. Poilve, 'Hyperspectral Imaging and Stress Mapping in Agriculture: a Case Study on Wheat in Beauce (France)', *Remote Sensing Environ.*, **66**(2), 179–191 (1998).
168. T.P. Dawson, P.J. Curran, 'A New Technique for Interpolating the Reflectance Red Edge Position', *Int. J. Remote Sensing*, **19**(11), 2133–2139 (1998).
169. L.K. Fenstermaker, J.R. Miller, 'Identification of Fluvially Redistributed Mill Tailings using High Spectral Resolution Aircraft Data', *Photogramm. Eng. Remote Sensing*, **60**(8), 989–995 (1994).
170. F. Lehmann, H. Rothfuss, R. Richter, 'Evaluation of Imaging Spectrometer Data (GER) for the Spectral Analysis of an old Vegetation Covered Waste Deposit', In: Proceedings of the 10th International Geoscience and Remote Sensing Symposium, IGARSS '90, May 20–24, Washington, DC, 1613–1616, 1990.
171. W.H. Farrand, 'Identification and Mapping of Ferric Oxide and Oxyhydroxide Minerals in Imaging Spectrometer Data of Summitville, Colorado, U.S.A., and the Surrounding San Juan Mountains', *Int. J. Remote Sensing*, **18**(7), 1543–1552 (1997).
172. R. Bianchi, R.M. Cavalli, C.M. Marino, S. Pignatto, M. Poscolieri, 'Evaluation of the Spatial Distribution in Percent Coverage of the Oil Spilled during the Trecate Blow-out Based on the Analysis of Airborne Hyperspectral MIVIS Data', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. I, 66–75, 1996.
173. F.A. Kruse, 'Mineral Mapping for Environmental Hazards Assessment using AVIRIS Data, Leadville, Colorado, USA', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. II, 526–533, 1996.
174. A. Mueller, F. Lehmann, H. Rothfuss, 'The Potential of Imaging Spectrometry (DAIS 7915) for the Monitoring of Recultivation Activities in Mining Areas', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. II, 350–357, 1996.
175. R.N. Clark, G.A. Swayze, L. Rowan, E. Live, K. Watson, 'Mapping Surficial Geology, Vegetation Communities, and Environmental Materials in Our National Parks: The USGS Imaging Spectroscopy, Integrated Geology,

Ecosystems, and Environmental Mapping Project', Presented at the Airborne Geoscience Workshop 1996, Pasadena, CA, 1996.
176. B.H. Bammel, R.W. Birnie, 'Spectral Reflectance Response of Big Sagebrush to Hydrocarbon-induced Stress in the Bighorn Basin, Wyoming', *Photogramm. Eng. Remote Sensing*, **60**(1), 87–96 (1994).
177. H. Yang, J. Zhang, F. van der Meer, S.B. Kroonenberg, 'Spectral Characteristics of Wheat Associated with Hydrocarbon Microseepages', *Int. J. Remote Sensing*, **20**(4), 807–813 (1999).
178. F. van der Meer, 'Sequential Indicator Conditional Simulation and Indicator Kriging Applied to Discrimination of Dolomitization in GER 63-channel Imaging Spectrometer Data', *Nonrenewable Resources*, **3**(2), 146–164 (1994).
179. W.J. De Oliveira, A.P. Crosta, 'Detection of Hydrocarbon Seepage in the Sao Francisco basin, Brazil, through Landsat TM, Soil Geochemistry and Airborne/Field Spectrometry Data Integration', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. I, 155–165, 1996.
180. V. Carrere, J.E. Conel, 'Recovery of Atmospheric Water Vapor Total Column Abundance from Imaging Spectrometer Data around 940 nm – Sensitivity Analysis and Application to Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Data', *Remote Sensing Environ.*, **44**, 179–204 (1993).
181. B.C. Gao, A.F.H. Goetz, 'Column Atmospheric Water Vapor and Vegetation Liquid Water Retrievals from Airborne Imaging Spectrometer Data', *J. Geophys. Res.*, **95**, 3549–3564 (1990).
182. B.C. Gao, A.F.H. Goetz, 'Retrieval of Equivalent Water Thickness and Information Related to Biochemical Components of Vegetation Canopies from AVIRIS Data', *Remote Sensing Environ.*, **52**, 155–162 (1995).
183. R. Frouin, P.Y. Deschamps, P. Lecomte, 'Determination from Space of Atmospheric Total Water Amounts by Differential Absorption near 940 nm: Theory and Airborne Verification', *J. Appl. Meteorol.*, **29**, 448–460 (1990).
184. S.M. De Jong, T.G. Chrien, 'Mapping Volcanic Gas Emissions in the Mammoth Mountain Area using AVIRIS', Presented at the Sixth Earth Science Workshop, March 4–8, 1996, Jet Propulsion Laboratory, Pasadena, 1996.
185. S.M. De Jong 'Imaging Spectrometry for Monitoring Tree Damage caused by Volcanic Activity in the Long Valley Caldera, California', *ITC J.*, **1998**(1), 1–10 (1998).
186. J.K. Crowley, D.R. Zimbelman, 'Mapping Hydrothermally Altered Rocks on Mount Rainier, Washington: Application of Airborne Imaging Spectrometry to Volcanic Hazards Assessment', In: Proceedings of the Eleventh Thematic Conference and Workshop on Applied Geologic Remote Sensing, Las Vegas, Nevada, 27–29 February, Vol. II, 460–461, 1996.
187. J.K. Crowley, D.R. Zimbelman, 'Mapping Hydrothermally Altered Rocks on Mount Rainier, Washington, with Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data', *Geology*, **25**(6), 559–562 (1997).
188. A.B. Kahle, A.F.H. Goetz, 'Mineralogic Information from a New Airborne Thermal Infrared Multispectral Scanner', *Science*, **222**, 24–27 (1983).
189. A.B. Kahle, 'Surface Emittance, Temperature, and Thermal Inertia Derived from Thermal Infrared Multispectral Scanner (TIMS) Data for Death Valley, California', *Geophysics*, **52**(7), 858–874 (1987).
190. M.J. Abrams, E. Abbott, A. Kahle, 'Combined Use of Visible, Reflected Infrared and Thermal Infrared Images for Mapping Hawaiian Lava Flows', *J. Geophys. Res.*, **96**(B1), 475–484 (1991).
191. Y. Ninomiya, T. Matsunaga, Y. Yamaguchi, K. Ogawa, S. Rokugawa, K. Uchida, H. Muraoka, M. Kaku, 'A Comparison of Thermal Infrared Emissivity Spectra Measured in situ, in the Laboratory and Similar Spectral Derived from Thermal Infrared Multispectral Scanner (TIMS) Data in Cuprite, Nevada, U.S.A.', *Int. J. Remote Sensing*, **18**(7), 1571–1581 (1997).
192. F. Bihong, C. Xiaowei, 'Thermal Infrared Spectra and TIMS Imagery Features of Sedimentary Rocks in the Kalpin Uplift, Tarim Basin, China', *Geocarto Int.*, **13**(1), 69–75 (1998).
193. A.R. Gillespie, 'Enhancement of Multispectral Thermal Infrared Images: Decorrelation Contrast Stretching', *Remote Sensing Environ.*, **42**, 147–155 (1992).

Land Cover Assessment and Monitoring

Eric Lambin
Université catholique de Louvain, Louvain-la-neuve, Belgium

1 Introduction **1**
1.1 Land Cover and Global Change 1
1.2 Major Land Cover Change Processes 2
1.3 Definitions 2
1.4 Remote Sensing Systems Used 3
2 Land Cover Attributes measured by Remote Sensing **4**
2.1 In the Visible Range of the Spectrum 4
2.2 In the Thermal Range of the Spectrum 4
2.3 Relationships between Vegetation Indices and Surface Temperature 5
2.4 In the Microwave Range of the Spectrum 7
2.5 In the Spatial Domain 8
3 Discrete versus Continuous Representations of Land Cover **9**
4 Land Cover Mapping **9**
4.1 Multispectral Approaches 9
4.2 Multitemporal Approaches 10
4.3 Categorical and Ordinal Legends for Land Cover Maps 10
4.4 Spatial Unmixing of Land Cover Information 11
4.5 Land Cover Maps and Interannual Climatic Variability 12
5 Land Cover Monitoring **12**
5.1 Reference State for Land Cover 12
5.2 Change Detection Methods 13
5.3 Mapping Ecosystem Disturbances 14
5.4 Land Cover Monitoring Systems 14
6 Perspectives **15**
6.1 Model Inversion 15
6.2 New Sensors and Multisensor Approaches 15
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Land-use changes are cumulatively transforming land cover at an accelerating pace, mainly in the tropics. These changes in terrestrial ecosystems are closely linked with the issue of the sustainability of socioeconomic development since they affect essential parts of our natural capital such as climate, soils, vegetation, water resources and biodiversity. Changes in terrestrial ecosystems brought about by human activity are driven by land cover conversion, land degradation, land-use intensification or other forms of land cover modification. The term land cover *refers to the attributes of a part of the Earth's land surface and immediate subsurface, including biota, soil, topography, surface and groundwater, and human structures. The term* land use *refers to the purposes for which humans exploit the land cover. One generally distinguishes between land cover* conversion*, the complete replacement of one cover type by another, and land cover* modification*, more subtle changes that affect the character of the land cover without changing its overall classification. Land cover modifications are generally more prevalent than land cover conversions. Two types of satellite sensors are used in land cover applications: fine spatial resolution systems and coarse spatial resolution/high temporal frequency systems. Data are collected in the visible, thermal and microwave ranges of the spectrum. Information in the spatial domain also provides valuable information on land surface attributes. These data are analyzed either for land cover mapping or for land cover monitoring. Current research is moving toward the direct estimation of biophysical surface parameters by model inversion methods. Multisensor approaches are also increasingly being developed.*

1 INTRODUCTION

1.1 Land Cover and Global Change

Land-use changes are cumulatively transforming land cover at an accelerating pace, mainly in the tropics.[1,2] These changes in terrestrial ecosystems are closely linked with the issue of the sustainability of socioeconomic development since they affect essential parts of our natural capital such as climate, soils, vegetation, water resources and biodiversity.[3] Actually, changes in land cover have important implications for a range of issues such as biosphere–atmosphere interactions, species and genetic diversity associated with endangered habitats, soil conditions, water and sediment flows and sustainable use of natural resources in the development process of human societies.[4] However, quantitative data on where, when and why such changes take place globally are still incomplete and sometimes inaccurate. Our knowledge and understanding of land

cover change processes are too fragmented for assessing the broader impact of these changes on natural and human systems. While our knowledge about interactions between land surface processes and climate change has achieved recent advances,[5–7] there are still a number of unknowns.

Changes in land cover are possibly driven by five categories of causes: long-term natural changes in climatic conditions; geomorphologic and ecological processes (e.g. soil erosion and natural vegetation dynamics); human-induced alterations of vegetation cover and landscapes (e.g. deforestation and land degradation); interannual climatic variability (e.g. recurrent droughts and floods); and human-induced climatic changes. These processes have different impacts on ecosystems and human societies, occur at different rates and scales, and are characterized by varying degrees of reversibility. Because of considerable interactions between these processes, the explanation and prediction of land cover changes is a complex task. The study of land cover change processes requires a three-level approach:[8]

1. direct measurements of the rate, location, spatial pattern and temporal characteristics of land cover changes;
2. case studies and field investigations to gain insights into local-scale dynamics of land cover changes;
3. identification of the broad scale factors that drive land cover changes and allow projection of future trends.

Thus, there is a need both for comparative analyses of the main processes of land cover changes and for advanced methods to monitor and model land cover changes at regional scales. An important research priority is the accurate measurement of land cover change magnitudes and the categorization of change processes. Suitable measurements will allow evaluation of the major driving forces of changes, their potential effects and, eventually, prediction of future trends.

1.2 Major Land Cover Change Processes

Changes in terrestrial ecosystems brought about by human activity are driven by land cover conversion, land degradation, land-use intensification or other forms of land cover modification. The most widespread contemporary example of *land cover conversion* is tropical deforestation. It is estimated that forest-cover conversion reached an average of 15.5 million hectares per year for the period 1981 through 1990, which translates to an annual deforestation rate of 0.8%.[9] These figures have been critically examined by Grainger.[10] Deforestation is, in most cases, the result of complex causality chains which originate beyond the forestry sector.[11] Forest conversion in the humid tropics accounts for a significant part of the carbon flux from terrestrial ecosystems[12,13] and contributes to a large share of the loss of biodiversity.[14] Another economically and demographically important process of conversion is urbanization. However, less than 2% of the Earth's surface can be considered ''urban'' and only 0.2% is densely built-up.[15]

Land degradation is most severe and widespread in semi-arid regions. It encompasses processes such as soil erosion or soil salinization, which are sometimes described by the controversial concept of ''desertification''.[16] Land degradation implies a decline in the usable natural resource base and, thus, directly affects the food supply. It is usually associated with sociopolitical mechanisms leading to a ''pressure of production on resources''.[17] A synthesis of the most recent global assessments of human-induced land degradation estimated that 69.5% of the world drylands are affected by various forms of land degradation.[18] However, these assessments were not based on systematic measurement and do not allow derivation of spatially disaggregated rates of desertification. There are strong interactions between dryland degradation, the primary productivity of vegetation and climate.[19,20]

Land-use intensification may be associated with agricultural, agroforestry or grazing systems. Intensified management of land can be based on such techniques as irrigation, fertilizer use or the integration of different production activities.[21] It may be driven by population pressure, market demand or political economy factors.

1.3 Definitions

1.3.1 Land Cover versus Land Use

The term *land cover* refers to the attributes of a part of the Earth's land surface and immediate subsurface, including biota, soil, topography, surface and groundwater, and human structures. Examples of some broad categories of land cover are: boreal forest, tropical savanna, temperate grasslands, croplands, wetlands and settlements.[22] The term land use refers to the purposes for which humans exploit the land cover. Common land uses include agriculture, grazing, forestry, mineral extraction and recreation. Forest is a land cover dominated by woody species, which may be exploited for land uses as varied as recreation, timber production or wildlife conservation. Changes in human land use are frequent causes of land cover conversion and modification. Monitoring land surface processes by remote sensing is directed to changes in biophysical attributes of the surface, i.e. the land cover.

1.3.2 Land Cover Conversion versus Land Cover Modification

One generally distinguishes between land cover conversion, the complete replacement of one cover type by another, and land cover modification, more subtle changes that affect the character of the land cover without changing its overall classification. Land cover modifications may result in degraded ecosystems. Land cover modifications are generally more prevalent than land cover conversions. In principle, the monitoring of land cover conversions (e.g. agricultural expansion or deforestation) can be performed by a simple comparison of successive land cover maps (e.g. derived by classification of remote sensing data or by field surveying). However, the comparison of land cover classifications for different dates does not allow the detection of subtle changes within land cover classes: even if some of the attributes of one class have changed, the magnitude of these changes will not always be large enough to justify a shift from one land cover category to another, unless the vegetation classification identifies a very large number of narrowly defined categories. Therefore, monitoring land cover changes can only be achieved through repetitive measurements of biophysical attributes which characterize the land cover. In this article, remote sensing approaches to measure indicators of the state of land cover and its change are reviewed. The main emphasis of this review is on methods applicable at broad spatial scales.

1.4 Remote Sensing Systems Used

1.4.1 Medium to Fine Spatial Resolution Systems

A number of sensors have acquired multispectral data at a spatial resolution of dozens of meters over several decades. The most widely used among these sensors are: Landsat Multispectral Scanner (MSS) (nominal spatial resolution of 80 by 80 m), Landsat Thematic Mapper (TM) (nominal spatial resolution of 30 by 30 m), and SPOT (Système Probatoire Pour l'Observation de la Terre) High-resolution Visible (HRV) Imaging Instrument (nominal spatial resolution of 20 by 20 m for the multispectral product ''XS'' and nominal spatial resolution of 10 by 10 m for the panchromatic product). As an example, the spectral bands of the Landsat TM sensor are: 0.45–0.52 μm, 0.52–0.60 μm, 0.63–0.69 μm, 0.76–0.90 μm, 1.55–1.75 μm, 2.08–2.35 μm and 10.40–12.50 μm. Time series of data from these fine spatial resolution sensors can be created to monitor changes in landscapes at a local scale.

A new generation of fine spatial resolution sensors is due to be launched around the turn of the century. These sensors are capable of generating imagery with spatial resolutions as fine as 1 m in panchromatic mode and 4 m in multispectral mode. These systems are also characterized by high geometric precision, short revisit intervals and rapid data supply.[23] However, they have narrow swath widths, which limits data acquisition to relatively small areas. They also have few, broad spectral bands. Such imagery will provide greater spatial detail on the land surface.

1.4.2 Coarse Spatial Resolution/High Temporal Frequency Systems

For some applications, measures of land cover change must be represented at a level of spatial aggregation which matches the spatial resolution of ecosystem models. The potential of coarse spatial resolution (1 km or coarser) remote sensing data with a high frequency of observation (daily) for monitoring changes in the attributes of the Earth's surface is well documented.[24,25] Radiance data measured from the Advanced Very High-resolution Radiometer (AVHRR) on the National Oceanic and Atmospheric Administration (NOAA) series of orbiting platforms have been widely used. These data cover the period from June 1981 to present. The daily NOAA AVHRR Local Area Coverage (LAC) data have a spatial resolution of 1.1 km at nadir, which is well suited to monitor seasonal vegetation dynamics at the scale of the landscape. Continental scale, daily Global Area Coverage (GAC) data have a minimum resolution-cell size of 3.3 by 5.5 km and are produced by resampling the 1.1 by 1.1 km spatial resolution data.[26] The spectral bands of the NOAA AVHRR sensor are: 0.58–0.68 μm, 0.72–1.10 μm, 3.55–3.93 μm, 10.30–11.30 μm and 11.50–12.50 μm.

The Along Track Scanning Radiometer (ATSR)-1 sensor embarked on the Earth Resource Satellite (ERS)-1 satellite produced high-quality 1-km spatial resolution data.[27] The sensor collects data in the short-wave infrared (SWIR) part of the spectrum (i.e. reflected energy) and in two thermal infrared channels (i.e. emitted energy). (ATSR-1's middle infrared channel failed early in the mission.) These data are collected in two 500 km swath looks, one at nadir and the other one some 600 km in front of the platform, at an angle of 55°. Repeat coverage at the equator is between 6 and 9 days, although this will alter depending on the ERS-1 operational cycle. The SWIR channel (ATSR's channel 1b) overlaps with the Landsat TM channel 5. The thermal channels (ATSR's channels 2 and 3) overlap with the NOAA AVHRR channels 4 and 5. The ATSR-2 sensor has the same specifications, but with two additional bands in the red and near-infrared (NIR) parts of the spectrum. It has been providing data with two look angles for several years now. The forthcoming ENVISAT (environmental satellite) Mission of European Space Agency (ESA) will include an Advanced Along Track Scanning Radiometer (AATSR) instrument.

A new promising sensor for land cover monitoring is the Moderate Resolution Imaging Spectroradiometer (MODIS), as part of National Aeronautics and Space Administration (NASA)'s Earth Observing System (EOS).[28] This instrument is similar to the NOAA AVHRR sensor in that it provides frequent temporal coverage of the Earth, acquiring images on a two-day repeat. Unlike NOAA AVHRR, the instrument images in 36 spectral bands from 0.4 to 14 μm selected for studies of atmosphere, ocean and land. The land bands sample regions of the spectrum that are largely similar to those sampled by the Landsat TM, except that bandwidths are much narrower. Spatial resolution of these bands is 500 m, while resolution of two bands in the red and NIR is 250 m. The MODIS instrument will be much better calibrated than AVHRR, and because atmospheric data are acquired simultaneously with land data, it should be possible to routinely correct MODIS land observations for atmospheric effects. A number of MODIS standard land products will be generated, to improve our capability for land surface monitoring at a global scale.[29] MODIS is now scheduled for flight on both morning pass and afternoon pass platforms of EOS, due for launch at the turn of the century.

A sister instrument of MODIS is the Medium Resolution Imaging Spectrometer (MERIS), as part of the forthcoming ENVISAT Mission of ESA. It will include 15 spectral bands in the visible and NIR range, at a 250 m spatial resolution. Both the position and the bandwidths of the spectral bands will be fully programmable by ground command. The data will be fully calibrated on board and corrected for atmospheric effects.

Finally, several active microwave sensors are available: the Synthetic Aperture Radar (SAR) on the ERS-1 and Japanese Earth Resource Satellite (JERS)-1 satellites, and the forthcoming Advanced Synthetic Aperture Radar (ASAR) sensor as part of the forthcoming ENVISAT mission. For all these sensors, a radar beam illuminates the ground and an image of the ground is build up from the return signals, which are dependent on the scattering properties of the individual targets. ERS-1/JERS-1 SAR is a version of side-looking radar, operating in imaging mode, in which high spatial resolution (30 by 30 m for ERS-1, 18 by 18 m for JERS-1) is obtained. The frequency of the radar signal is 5.3 Ghz (C-Band) for ERS-1 and 1.25 Ghz (L-Band) for JERS-1. Both systems have a single polarization (VV for ERS-1 and HH for JERS-1). The incidence angle of the system in normal operation is 23°. This can be varied up to 35°. ENVISAT ASAR is an advanced active microwave instrument, employing a number of new technological developments.

Note that this enumeration of sensors used for land cover applications is not exhaustive. Only the most commonly used instruments for land applications have been mentioned here.

2 LAND COVER ATTRIBUTES MEASURED BY REMOTE SENSING

2.1 In the Visible Range of the Spectrum

Many authors have characterized the surface using vegetation indices, which are arithmetic combinations of spectral bands. The most widely used of these indices is the normalized difference vegetation index (NDVI). The NDVI is calculated (Equation 1) from reflectances in the red and NIR parts of the electromagnetic spectrum as:

$$\frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \tag{1}$$

Vegetation indices derived from remote sensing data are related to several variables quantifying vegetation activity and state. Green leaf area index (LAI) is inversely related to red reflectance and positively related to NIR reflectance.[30] Several studies showed that ratio vegetation indices such as the NDVI were related to the fractional vegetation cover of a surface.[31,32] Empirical studies and simulations with radiative transfer models support the interpretation of the NDVI in terms of fraction of photosynthetically active radiation (PAR) absorbed by the vegetation canopy, canopy attributes (e.g. green biomass or green LAI), state of the vegetation (i.e. vegetation vigor or stress) and instantaneous rates associated with the activity of the vegetation.[33–35] Seasonal variations of vegetation indices are related to vegetation phenology and biome seasonality.[36] Many other indices have been defined. They are variants of the NDVI and display different levels of insensitivity to perturbing factors such as soil color changes or atmospheric effects.[37]

2.2 In the Thermal Range of the Spectrum

Land surface skin brightness temperature (T_s) can be derived (Equation 2) from the thermal channels of the NOAA AVHRR using the split-window technique:[38]

$$T_s \text{ (in degree Celsius)} = T_4 + 3.33(T_4 - T_5) - 273 \tag{2}$$

where T_4 and T_5 are true brightness temperatures in NOAA AVHRR channels 4 and 5 (in degree Kelvin). This formula assumes a constant surface emissivity. The correspondence between the thermal channels of the AVHRR and ATSR sensors means that the split-window algorithm to estimate land surface temperatures (T_s) from NOAA AVHRR data can be adapted for use with ATSR data.[39] The ATSR however has a higher

radiometric resolution (12 bit) than the AVHRR (10 bit) for the thermal channels. The potential error magnitude on satellite-derived T_s is 2–3 °C.

T_s is related, through the surface energy balance equation, to surface moisture availability and evapotranspiration, as a function of latent heat flux.[40] Empirical studies have demonstrated the usefulness of thermal data to locate dense forest boundaries.[41] In tropical regions, surface temperature is largely a function of evaporation, when water availability is not limited. Energy is transferred to the atmosphere in the form of latent heat through bare soils evaporation, plant transpiration and direct evaporation of water intercepted by plant canopies. This heat loss determines to a large extent, through the surface energy balance equation, the daily mean temperature of the surface.[42–44] Main determinants of surface evaporation are radiation, the moisture availability at the surface, the presence of active vegetation, wind speed, surface roughness and atmospheric demand. In the absence of sensible heat advection, net radiation sets an upper limit to the latent heat flux transfer, because of the balance between incoming and outgoing energy fluxes at the Earth's surface. Evaporative losses of vegetated surfaces also depend on factors such as rooting depth of plants, stomatal resistance, canopy height and structure, soil hydraulic properties, topography and the physiographic and geomorphic characteristics of hydrologic systems. Surface resistance (both aerodynamic diffusion resistance and canopy resistance) is thus influenced by land cover factors such as the amount and type of vegetation and soil characteristics.

2.3 Relationships between Vegetation Indices and Surface Temperature

2.3.1 Biophysical Processes Controlling Surface Temperature and Vegetation Index

Published studies demonstrated the advantage of combining vegetation index (VI) with T_s data for land cover and land cover change analysis.[45–47] A negative relationship between remotely sensed vegetation indices and surface brightness temperature has been widely observed.[48–55] This relationship may be driven by several mechanisms:

1. An increase in green biomass is often associated with a reduction in surface resistance to evapotranspiration, greater transpiration and a larger latent heat transfer. The presence of actively transpiring vegetation thus increases the share of latent heat flux with regard to total available energy. In general, soil water extraction by plant roots occurs more rapidly and at much greater depth than water diffusion to the soil surface. This relationship also depends on the synoptic state of the atmosphere (air temperature and vapor pressure deficit), soil moisture availability and the physiological activity of the vegetation.
2. Radiant surface temperature variations over a vegetated surface result from variations in the proportion of bare soil visible to the radiometer. During daylight hours, plant leaves are cooler than exposed bare soil because the heat capacity of plant leaves is much lower than the heat capacity of soil.[56] Variations in the proportions of radiances emitted from shaded or unshaded bare soil and vegetation lead to coarse scale variations in apparent aggregated brightness temperature at the level of ground scene elements. The proportions of these scene components within the sensor instantaneous field of view is a function of fractional vegetation cover, canopy three-dimensional architecture, solar elevation angle and viewing geometry. The importance of this effect was demonstrated by Friedl and Davis[54] over a well-watered tallgrass prairie.
3. The thermal inertia of scene components also influences the T_s–VI relationship. Thermal inertia measures the thermal response of a material to temperature changes. It is a function of thermal conductivity, thermal capacity and density. It is affected by both permanent surface characteristics (i.e. soil, landform and geological setting) and the transients such as soil moisture, ventilation, vegetative cover or albedo. The thermal inertia of vegetation canopies is lower than the thermal inertia of soils. Mineral constituents of soil have large densities and thermal conductivities compared to foliage,[56,57] and both mineral constituents and the organic matter in soil have much higher heat capacities than foliage. Moreover, soil moisture has a pronounced effect on soil thermal properties. High soil moisture further increases the thermal inertia of the substrate relative to vegetation canopies. Note that stomatal control of transpiration also influences the diurnal temperature range. Transpiration by plants causes a lower daytime temperature of vegetated surfaces relative to bare soils, and thus has a conflicting effect with the lower thermal inertia of vegetation canopies. The more vegetated a pixel, the more evapotranspiration suppresses the maximum temperature and the smaller the day–night temperature difference. As documented by Whitehead et al.,[58] this leads to a negative relationship between a VI and the diurnal temperature range, and thus to a negative relationship between a VI and daytime surface temperature.
4. Increasing vegetation height increases surface roughness, until a canopy reaches complete closure. An increase in the roughness of canopies increases turbulent exchange of water above the canopy and the

efficiency of energy dissipation, and thus decreases surface temperature. Even the presence of non-transpiring vegetation could be associated (if other conditions are met) with a lower surface temperature than for bare soil, owing to higher aerodynamic surface roughness. This increase in surface roughness and the resultant decrease in aerodynamic resistance occur to a slight extent for agricultural crops but to a much larger degree for trees, scattered bushes and brush. However, model simulations by Smith and Choudhury[59] showed that vegetation height did not significantly affect the T_s–NDVI relationship.

5. Over bright soils, a decrease in green vegetation increases surface albedo, which results in reduced surface heating. However, Goward et al.[48] showed that this effect is small compared to latent heat exchanges associated with evapotranspiration.

The dominance of one of these processes over the others depends on the land cover type, the climatic and surface moisture conditions and the specific location (e.g. soil type, landform, local climate, spatial heterogeneity in surface attributes, latitude). On surfaces with a low moisture content, latent heat exchanges associated with evapotranspiration seem to be the dominant mechanism accounting for variations in surface temperature with increases in vegetation cover. On well-watered surfaces, it is rather the difference between soil and vegetation thermal properties associated with changes in fractional vegetation cover that drives the T_s–NDVI inverse relationship. Spatial variations in surface attributes will cause a negative T_s–VI relationship in space. The relationship between T_s and VI brought by seasonal variations in surface attributes depends on environmental conditions. When energy is the main limiting factor to vegetation – e.g. in high-latitude regions – changes in VI are positively correlated in time to changes in T_s. This comes from the senescence of vegetation in fall with temperature decreases, followed in spring by vegetation growth with temperature increases. In this case, variations in T_s control variations in VI over the phenological cycle. When water is the main limiting factor to vegetation – as is the case in the warm regions of Africa – changes in VI are expected to be negatively correlated in time to changes in T_s, for the reasons discussed above.

2.3.2 *The Slope of the Surface Temperature and Vegetation Index Relationship*

Several authors have attempted to interpret the partial derivative – or the sensitivity – of the T_s–NDVI relationship in biophysical terms. Goward et al.[48] suggested that the rate of change in surface temperature with changes in the amount of vegetation could be diagnostic of surface resistance to moisture flux. Simulations of water and energy flows in the soil–plant–atmosphere system, using a canopy reflectance model, confirmed that an increase in vegetation amount, as indicated by the VI, leads to a decrease in the area-averaged minimum canopy resistance and to an increase in latent heat flux. Nemani and Running[50] related, over conifer forests, the slope of the T_s–NDVI relationship to regional surface resistance due to soil water deficit – i.e. the landscape ''dryness''. Smith and Choudhury[59] considered a greater variety of land cover types and noted that two separate effects influence the T_s–NDVI slope: soil evaporation – the dominant effect for areas with a low NDVI – and leaf stomatal resistance – the dominant effect for areas with a high NDVI. Later, on the basis of broad scale data, Nemani et al.[60] concluded that the slope of the T_s–NDVI relationship is primarily controlled by fractional vegetation cover, surface moisture status and local meteorological conditions. Over the well-watered surface studied by Friedl and Davis,[54] soil background temperature (which is affected by soil moisture) more than surface partitioning of energy into latent versus sensible heat was driving the T_s–NDVI relationship. This complicates and reduces the generality of previous interpretations of the T_s/NDVI relationship, for instance in terms of surface Bowen ratio conditions.[51] Finally, Hope and McDowell[61] and Friedl and Davis[54] observed that the T_s–NDVI relationship depends on land cover type.

In summary, these studies have tended to interpret the partial derivative of the T_s–VI relationship only on the basis of one or two of the processes driving it. Simulation studies which would integrate all five processes discussed previously still have to be conducted. Note that the studies reviewed above were performed at a variety of spatial scales: from fine resolution data that can resolve local variations in canopy density, to coarse resolution data that resolve landscape units. In the latter case, the spatially aggregated T_s and VI data are mainly sensitive to continental scale patterns of land use and land cover.

2.3.3 *The Surface Temperature and Vegetation Index Space*

Previous findings by Price,[52] Goward and Hope[51] and Nemani et al.[60] on the T_s–VI feature space are summarized on Figure 1. Over bare soil, variations in radiant surface temperature are highly correlated with variations in surface water content.[62,63] Thus, points A and B on Figure 1 represent, respectively, dry bare soil (low VI, high T_s) and moist bare soil (low VI, low T_s). As the fractional vegetation cover increases, surface temperature decreases (as seen above, this relationship is driven by several biophysical mechanisms). Point C on Figure 1 corresponds to continuous vegetation canopies

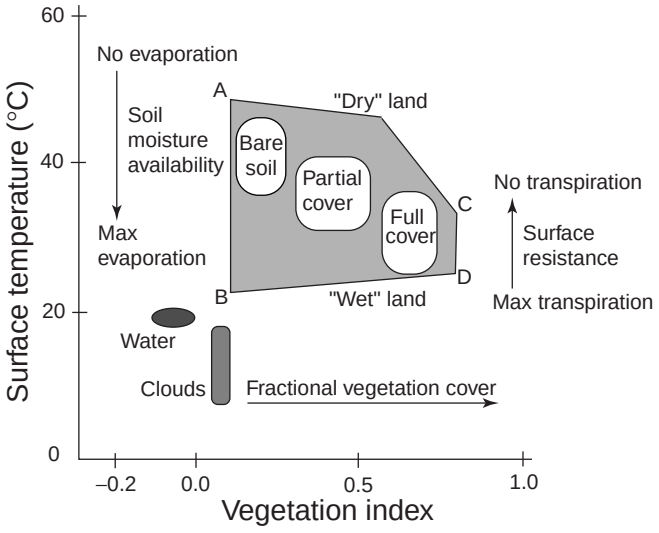

Figure 1 Surface temperature (T_s) and VI space. (Reproduced by permission of Taylor & Francis from E.F. Lambin, D. Ehrlich, *Int. J. Remote Sens.*, **17**, 463–487 (1996).)

with a high resistance to evapotranspiration (high VI, relatively high T_s) e.g. due to a low soil water availability. Point D corresponds to continuous vegetation canopies with low resistance to evapotranspiration (high VI, low T_s) e.g. on well-watered surfaces. The upper envelope of observations in that space, A–C, represents the low-evapotranspiration line (i.e. dry conditions). The lower envelope B–D represents the line of potential evapotranspiration (wet conditions). Recently, Moran et al.[64] defined a similar ''VI/temperature trapezoid'' in a space defined by fractional vegetation cover and surface minus air temperatures.

Data values from any scene will be distributed, during the growing season, between these four extreme points of the T_s–VI space. Outside the growing season, surfaces in regions which are affected by low air temperatures, i.e. in high latitudes, will have part of their time trajectories outside the area defined by A–B–C–D, with T_s and VI approaching zero in winter. Open bodies of water (and clouds) are characterized by lower VI and T_s.

The studies reviewed above led to a good understanding of the biophysical relationships of the T_s–VI space. This knowledge can be used to improve current strategies for land cover mapping and change detection, by defining measurements in this feature space. The main issue of land cover mapping is how to partition this space. The main issue of land cover change analysis is how to characterize time trajectories in that space.

2.4 In the Microwave Range of the Spectrum

Information on surface processes is also acquired by sensors that operate in the microwave portion of the electromagnetic spectrum. The cloud-penetrating capability of microwaves makes them an ideal tool to work in areas with a permanent cloud cover (e.g. equatorial regions). Note that microwave reflections and emissions from the land surface are controlled by totally different processes from those that control reflection and emissions in the visible or thermal part of the spectrum. Both active and passive microwave sensing are used. Active systems supply their own source of energy or illumination while passive systems sense energy that is naturally emitted or reflected from the Earth surface, at extremely low levels however. As of today, there have been fewer successful applications of microwave remote sensing for land cover monitoring or mapping than those using optical sensing. There is however a large potential provided that the interpretation of microwave data is approached through a physical modeling approach.

In active systems, the intensity of microwave returning from a given surface depends, on one hand, on the wavelength and polarization of the signal and, on the other hand, on the geometric characteristics of the surface (i.e. shape and orientation of objects, as well as their surface roughness) and its electrical characteristics (i.e. reflectivity and conductivity of materials, as measured by their complex dielectric constant). Note that changes in microwave signal intensity are often related to changes in moisture content more closely than they are to changes in surface materials. In general, green vegetation is a good reflector of microwave energy, because of its high moisture content and large surface area. Also, soil moisture and surface wetness conditions are well detected, particularly with microwave data at long wavelengths. On forest landscapes, the influence of soil moisture variations on the backscattering coefficient in the C-band appears to be the prime cause of dynamic behavior in temporal series of SAR imagery, even when relatively low levels of standing biomass are present.[65]

The C-band signal tends to saturate with increases in biomass at quite low levels of biomass. Therefore, C-band data are not good for discriminating forests. Even low vegetation canopies can give rise to returns very similar to those from a full forest canopy. The L-band provided by JERS-1 has improved discriminating power of forests as radiation penetrates further into the canopy.[65] Other authors also found that the L-band was more sensitive to stem volume in a boreal forest than the C-band.[66,67] The potential of the L-band from JERS-1 to retrieve above-ground biomass density in an equatorial forest was also evaluated as being promising. However, speckle and image texture set a biomass limit at 31 tones per hectare for retrieval purpose.[68] In general, the C-band is more dependent on weather. Thus, the L-band seems more attractive for biomass estimation.

A recent project has produced a subcontinental scale mosaic of Central Africa based on microwave data, recorded by the ERS-1 SAR sensor.[69] Identification of vegetation types and detection of changes in vegetation cover is most favorable with microwave data on flat to undulating terrain. For example, major success has been reported on the mapping of flooded forest based on ERS-1 and JERS-1 radar data.[70] Mountainous or dissected terrain provides significant difficulties. Combined imagery from the ERS-1 and JERS-1 SAR systems was tested by Kellndorfer et al.[71] for regional-to-global scale vegetation classification. This required high-resolution digital elevation model data for terrain correction of geometric and radiometric distortions. An unsupervised clustering was then applied to the two-dimensional SAR feature space. The combination of orbital L- and C-band SAR data is quite powerful for structural vegetation characterization.[71]

Seasonal contrasts can be exploited on microwave data to discriminate between different cover types (e.g. evergreen versus deciduous formations), given the frequent coverage of the data. Change classes are defined from multitemporal SAR data based on their unique signature of radar backscattering time series. Change classes can then be interpreted in terms of thematic classes, e.g. rice cultivation systems.[72] However, not all time variations in the backscattered microwave signal are of phenological origin.

Image texture analysis on microwave data is also often used for canopy characterization and landscape degradation mapping. The derivation of land surface parameters from microwave data using a model-inversion approach is the subject of continuing research. The problems associated with the extraction of thematic information in complex topography warrant particular attention as, for example, a lot of the tropical forests of the world are located on mountainous terrain.[73] The development of algorithms for exploiting the synergy between optical and microwave data sources is also the focus of current research. Clearly, the all-weather capacity of microwave remote sensing combined with the fine spatial resolution of the data opens promising avenues for land cover assessment and monitoring.

2.5 In the Spatial Domain

A major attribute of a landscape is its spatial pattern, i.e. the arrangement in space of its different elements. The concept of landscape spatial pattern covers, for example, the patch size distribution of residual forests, the location of agricultural plots in relation to natural vegetation, the shapes of fields or the number, types and configuration of landscape elements (i.e. their spatial heterogeneity). Landscape spatial pattern is seldom static because of both natural changes in vegetation and human intervention. The spatial dynamics of landscapes interact with ecological processes which have important spatial components,[74] e.g. flows of energy and matter between landscape components, biological productivity, biodiversity, the spread of disturbances.

Remote sensing offers the possibility to analyze changes in spatial structure at the scale of landscapes.[75–77] Textural information of high spatial resolution remote sensing images is related to biophysical parameters (crown diameter, distance between trees, tree density, tree positioning, LAI, tree height, etc.) and acquisition parameters (spatial resolution, spectral domain and viewing and illumination configurations).[78] Texture can be measured using a variety of indices such as variance, variograms, mathematical morphology or fractals.[79]

Indicators of the degradation of the vegetation cover can be derived from such measures.[80,81] For example, Pickup and Foran[82] developed a method to monitor arid landscapes used for pastoralism based on the spatial variability of the vegetation. The spatial autocorrelation function and mean-variance plots of a spectral indicator were found to be successful in discriminating between the cover responses typical of good and poor rainfall years. For drought conditions, the decrease in spatial autocorrelation with increasing spatial lag was rapid since the ground surface is bare and most of the vegetation signal comes from scattered areas of trees and shrubs. A low decay rate of the autocorrelation function indicated a greater spatial uniformity of the landscape, e.g. during wet periods when more ground cover is present, reducing the contrast between the bare soil signal and that produced by trees and shrubs.[82] Similar observations were made by Lambin[83] over the seasonal and interannual cycle of three West African landscapes. Vogt[84] also analyzed the seasonal changes in spatial structure of a West African landscape, showing that there is a marked seasonal cycle in the spatial structure of a VI, surface temperature and albedo, and that zones of ecological transition have an identifiable seasonal dynamic in spatial structure. However, the monitoring of these spatial variability measures only provide a qualitative description of the cover state. Hudak and Wessman[85] related a textural index to woody plant densities and woody plant encroachment across savanna landscapes. Actually, heterogeneity in vegetation structure has a significant influence on local pixel variance in high spatial resolution images. These authors found that, in a savanna landscape, woody stem count correlated best with image texture.

Lambin and Strahler[86] showed that the detection by remote sensing of land cover change processes is improved when using both spectral (e.g. VI and surface temperature) and spatial indicators of surface condition. Their study suggested that the detection of interannual

changes in landscape spatial structure is more likely to reveal long-term and long-lasting land cover changes, while spectral indicators are more sensitive to fluctuations in primary productivity associated with the interannual variability in climatic conditions. The long-term monitoring of landscape spatial pattern in addition to other biophysical variables might thus lead to the detection of a greater range of processes of landscape modification.[86]

The spatial pattern of a changing landscape also provides some information on the processes of land cover change. Certain categories of changes in human land use tend to fragment the landscape (e.g. expansion of extensive agricultural exploitation, land cover changes driven by small-scale logging, overgrazing or desertification around deep wells). Other land use changes increase landscape homogeneity (e.g. intensive cultivation or ranching over large areas). Spatial processes of gap expansion in a forest cover have been modeled to predict the total gap area and gap size distribution.[87] One can make the hypothesis that landscapes with a very low or very high level of disturbance are characterized by a low spatial heterogeneity while landscapes with a medium level of disturbance are very heterogeneous. Recent remote sensing observations grossly supported this hypothesis, e.g. in a study of forest-cover disturbances in Papua New Guinea[88] and another study of forest fragmentation in New England.[89]

Current research investigates the effect of changes in spatial resolution on the image variance, image texture and spatial patterns. The variance of a remotely sensed image is determined by the interaction of scene properties with the spatial characteristics of the sensor, mainly its spatial resolution.[90] Image variance is related to image information content. Texture is caused by the variability or uniformity of image tone or color. The higher the texture, the greater the degree of spatial heterogeneity in the data. Image texture varies with the scale of the data. When texture analysis is applied to images of the same scene with different spatial resolutions, it is argued that the highest texture indicates the resolution level at which most land surface processes operates and, therefore, at which a landscape should be analyzed.

3 DISCRETE VERSUS CONTINUOUS REPRESENTATIONS OF LAND COVER

The land surface can be represented as a set of spatial units which are each associated with an attribute. These attributes are either a single land cover category (i.e. leading to a discrete representation of land cover) or a set of values for continuous biophysical variables (i.e. leading to a continuous representation of land cover).[91] The correspondence between these two representations can be established through a table which associates to each land cover category the average range of values for the biophysical variables.

In the continuous representation of land cover, the biophysical variables vary continuously not only in space but also in time, at the seasonal and interannual scale. By contrast, in the discrete representation of land cover, each spatial unit is represented by a single categorical value which is stable over a season. Interannual changes in the values of the biophysical attributes of the surface are described, in the discrete representation of land cover, as land cover conversions only if the changes exceed the range of values which is characteristic of the land cover category (i.e. if the magnitude of the change is such that the values of all biophysical attributes falls within the range of another land cover class during the entire seasonal cycle). By contrast, a land cover modification, which is not detectable with the discrete representation of land cover, implies that variations in the values of the biophysical attributes remain within the range of values which is characteristic of the land cover category. Processes which lead to changes in the seasonal dynamics or in the fine scale spatial variability of biophysical attributes would also be described as land cover modifications. Changes that would only affect the values of some of the biophysical attributes, for just part of the seasonal cycle would probably also enter in that category.

This suggests that monitoring land cover changes by remote sensing requires the measurement of a set of indicators of the biophysical attributes of the surface, the seasonality of these attributes and their fine scale spatial pattern. These information requirements correspond to the three major information sources provided by remote sensing.

4 LAND COVER MAPPING

4.1 Multispectral Approaches

Classification is a method by which labels are attached to pixels in view of their character.[92] This character is generally their spectral response in different spectral ranges. It may also include their spatial attributes (i.e. texture) or temporal signature (see below). This labeling is implemented through pattern recognition procedures, the patterns being pixel vectors. The most commonly used classification methodology is based on the maximum likelihood, a probabilistic classification method which assumes that spectral classes can be described by a normal probability distribution in multispectral space.[92] More advanced classification techniques have been developed recently. Contextual classifiers label a pixel in the context

of its neighbors in space. It exploits spatial information by quantifying the neighborhood relationships of an image. Decision tree classifiers based on a large set of metrics have been used for global land cover mapping.[93] Neural networks are classifier networks, having a decision tree structure, where a collection of simple classifiers is brought together to solve a complex problem.[94] Neural network classifiers have also been proposed for global land cover mapping.[95] A number of studies have demonstrated that neural network-based classification can produce more accurate results than conventional approaches. The reasons include:[96]

1. neural network classifiers are distribution-free and can exploit nonlinear data patterns;
2. neural network classification algorithms can easily accommodate ancillary data;
3. neural network architectures are quite flexible and can be easily modified to optimize performance;
4. neural networks are able to handle multiple subcategories per class.

4.2 Multitemporal Approaches

The International Geosphere-Biosphere Programme (IGBP) has endorsed the use of coarse spatial resolution, high temporal frequency imagery for deriving land cover parameters for global change studies.[97] A number of studies have produced broad scale land cover maps from NOAA's AVHRR.[98–103] These studies were based on NDVI data time series of one year or less. Rather than analyzing isolated dates, the multitemporal classification approach is based on a comparison of the temporal development curve, or time-trajectory, for different spatial sampling units of indicators derived from NOAA AVHRR data, such as vegetation indices, surface temperature or spatial structure. Several researchers have shown that a high temporal frequency of measurement, which is a characteristic of the AVHRR, ATSR or MODIS sensors, allows the characterization of vegetation types by analysis of the seasonal or phenological variations of these spectral indicators.[104,105]

The production of land cover maps by classification of VI or surface temperature temporal development curves, on a pixel-by-pixel basis, requires the analysis of time-trajectory shapes for all pixels. This should take place after filtering the multitemporal data for noise introduced by atmospheric effects, sensor calibration and variations in viewing geometry. Various attempts have been made to develop mathematical expressions that fit vegetation development curves, using logistic or exponential expressions,[106] a combination of those for different phases of the radiometric evolution of a vegetative cover,[107] or by jointly using the results of the theories of generalized shift and of projection pursuit.[108] Other studies have used a few simple parameters to characterize the evolution in time of vegetation indices: the onset of greenness, the peak of greenness, the length of growing season, the integrated VI, the amplitude of the variation, etc.[100,101,109,110] Samson[111] proposes a skew index and a range index to characterize the shape of the seasonal profile of vegetation indices. Tucker et al.[98] used the information from the first two principal components of annual NDVI time series as measures of mean annual NDVI value and seasonality to classify the time-profiles of land cover types on a continental scale in Africa. For our objective of comparing successive temporal profiles to detect subtle land cover changes, a method of analysis which captures the maximum detail contained in the time-trajectory of the indicator is needed.

The seasonal dynamic of a remotely sensed indicator can also be represented by a point in a multidimensional space, with the number of dimensions of this space corresponding to the number of observations.[45] These observations will correspond either to the maximum values of the indicator for pre-specified compositing periods, or to the results of a sampling through time of a continuous function that describes the shape of the time-trajectory of the indicator. If we assume monthly-composed data or a monthly rate of sampling, twelve successive images will be available. The value taken by the indicator under consideration can be represented, for each pixel, by a point in the twelve-dimensional temporal space defined by the vector (Equation 3):

$$\mathbf{p}(i, y) = \begin{matrix} \mathrm{I}(t_1) \\ \mathrm{I}(t_2) \\ \cdots \\ \cdots \\ \mathrm{I}(t_n) \end{matrix} \tag{3}$$

where $\mathbf{p}(i, y)$ is the multitemporal vector for pixel i and the year y, I are the values of the indicator under consideration for pixel i at the time periods t_1 to t_n, n being the number of time dimensions. The magnitude of this vector, $|\mathbf{p}|$, measures the accumulated value of the indicator through the year. The direction of this vector, measured by the direction angles of $\mathbf{p}$, is a synthetic quantifier of the seasonal pattern of the indicator, i.e. the shape of the curve (Lambin and Strahler[45]).

4.3 Categorical and Ordinal Legends for Land Cover Maps

Major applications of land cover mapping are the monitoring of land cover changes and the modeling of land cover change processes as well as their climatic, ecological and socio-economic impacts. Land cover mapping is also performed to support land management

and land-use planning in order to achieve sustainable development. Each of these applications has different requirements in terms of legend, scale and accuracy of the land cover map. For example, for land-use planning, the characteristics of the map should be adapted to the specific needs of the land managers and planners, whether these are rural communities, states or regional institutions. These needs will depend on the main environmental constraints of a specific region and on the socio-economic objectives of the land managers.

Legends of land cover maps are typically at the nominal level: each item of the legend is a specific category, unrelated to the other items. This is the lowest form of measurement, in which the classes differ only in kind, not by degree.[112] For land management, it is useful to upgrade the legend of land cover maps into ordinal levels. In this case, each item of the legend is related to the other items according to an ordinal scale which can be defined on the basis of a variable of interest for land management. For ordinal-level classifications, the classes differ by degree and may be placed in rank order and therefore represents a higher level of measurement than a nominal-level classification. It thus provides a higher level of information to land-use planners and natural-resources managers.

In a few recent studies, noncategorical mapping of land cover were generated by retrieving sub-pixel information on, for example, forest cover percentage using a linear, spectral mixture model,[113] a fuzzy membership function[112] or an empirical calibration function between fine and coarse spatial resolution data.[114] These previous studies derived a continuous variable on forest cover percentage from spectral or spatial information, with no or little indication on forest cover conditions (e.g. degree of disturbance, degradation or heterogeneity of the cover).

4.4 Spatial Unmixing of Land Cover Information

The production of broad scale land cover maps generally relies on remote sensing data at a coarse spatial resolution. Thus, complex landscape mosaics are represented by a single value or a single land cover category. As a result, the estimation of land cover proportions from broad scale maps is associated with a systematic bias due to spatial aggregation effects. The magnitude of this error depends on the spatial resolution of the map, the initial proportions of the landscape in the different land covers and the spatial arrangement of the land covers at the fine resolution.[114]

Backward scaling of land cover proportions, which means retrieving the fine scale proportions of land cover classes from the coarse scale data, can be performed through a ''double sampling with regression estimator approach''. This method is based on a regression between an ''auxiliary'' variable, which can be measured over the entire population from coarse resolution data, and the ''target'' variable (i.e. land cover proportions at a fine resolution), which is measured over a sample of calibration sites. The parameters of this regression are estimated over the sample of fine and coarse resolution data, and then applied over the entire population of coarse resolution data, in order to predict the ''true'' land cover proportions (i.e. at a fine resolution). This statistical calibration estimator is known as inverse calibration model.[115] It has proved to be more reliable than classical calibration to correct misclassifications in remote sensing.[116,117]

The double sampling approach for backward scaling can be applied in two ways:

1. by modeling spectral mixtures at the scale of coarse resolution pixels[118–125] or
2. by modeling the influence of the spatial pattern of land cover classes on the proportional error for blocks of coarse resolution pixels.[114,126,127]

In the first approach, referred to as the mixed pixel estimator, a direct relationship between land cover proportions at a fine resolution and the brightness values at a coarse resolution is calibrated over a sample of sites. It does not require classification of the data. The assumptions are that:

- brightness values are stable across space;
- the brightness value at a coarse resolution is mainly determined by the cover proportion of the class of interest;
- the deviations between cover proportions and brightness data are consistent throughout the range of values.

This approach can be implemented using a variety of methods, e.g. linear mixture models, fuzzy classifiers or Bayesian methods.

In the second approach, the land cover proportions at a fine resolution are related to the land cover proportions measured on coarse resolution classifications and the influence of the landscape spatial pattern on this relationship is modeled. Spatial information can either be introduced explicitly as a continuous, exogenous variable in the model[114] or indirectly, through a spatial stratification of the study area based on landscape attributes.[121] When the spatial pattern of the landscape is not included in a correction model, the parameters of the function are specific to a landscape type and not applicable over large regions. This approach thus consists in post-classification correction procedures.

4.5 Land Cover Maps and Interannual Climatic Variability

Several authors indicated that NOAA AVHRR-based land cover maps could be improved by classifying multiyear NDVI time series rather than single year NDVI data. Tucker et al.[98] for example, in their pioneer land cover classification of Africa, noted that errors in biome delimitation were produced because their analysis relied only on a single year of data. This was due to "... the interannual variability in vegetation response to climatic conditions"[105] and the high sensitivity of NDVI to the year-to-year variations in rainfall and soil moisture.[128]

The production of land cover maps from the classification of multiyear datasets raises the question of the intrinsic timescale of the concept of land cover. Clearly, daily, monthly or seasonal variations in biophysical attributes of the surface do not qualify as land cover changes. Some biophysical attributes of ecosystems also fluctuate at the interannual timescale, owing to climatic variability. Such variations are often independent of long-term climate trends. Do these fluctuations qualify as land cover changes? The title of the study by Tucker et al.[129] – i.e. "Expansion and contraction of the Sahara desert..." – illustrates the uncertainty of this issue. Intuitively, it is not the boundaries of the Sahara as a geographic region, and of the desert as a distinct land cover class, which fluctuates in space from one year to the other by about 100 km. The fluctuation rather affects one of the many biophysical attributes defining a land cover type (i.e. the green biomass productivity) which responds to interannual variations in rainfall.[130]

These comments suggest that the interannual climatic variability leads to changes in ecological conditions which have limited meaning in terms of land cover and that these fluctuations hide more fundamental characteristics of land cover types. This could be explained either because one year of observation lacks statistical representativity to produce a robust land cover classification, or because the remotely sensed variable used in previous land cover classifications is oversensitive to fluctuating attributes of the land surface. Indeed, NDVI is largely driven by short-term climatic characteristics more than by long-term, more intrinsic land cover characteristics. Accordingly, there are two ways to improve current NOAA AVHRR-based land cover classifications: either one averages a climate-driven, but easily accessible variable (such as the NDVI) over several years, or one measures a new variable which is sensitive to land cover quite independently from short-term climatic variations.

5 LAND COVER MONITORING

5.1 Reference State for Land Cover

Prior to detecting changes, one has to define a reference state for land cover. This can be performed in two ways: (i) on the basis of a bioclimatic reconstruction of potential climax or predisturbance vegetation or (ii) on the basis of actual observations of the state of land cover in every location at a given time. These two approaches lead to fundamentally different assessments of changes. Studies investigating the impact of land cover changes on the global climate relied on crude approximations of changes over a very long period, taking as a reference land cover prior to human disturbances.[131–134] The approach is only valid if the reference state is defined through retro-projections by a model integrating all the factors constraining vegetation conditions, i.e. not only temperature and rainfall (as in climate-correlative methods) but also soil characteristics and natural disturbances such as fires. Moreover, the bioclimatic approach does not treat the temporal dimension of land cover changes explicitly: the timing, cyclicity or reversibility of changes are unknown.

By contrast, the definition of a reference state of land cover from direct observations has the advantage of clearly defining the temporal dimension of the changes detected. A baseline inventory identifies the land cover distribution as it actually occurred in a recent past. However, owing to a lack of historical data on land cover, this approach is restricted to a few decades. For the recent period (i.e. last 30 years), remote sensing techniques allow land cover mapping at broad spatial scales.[135,136]

The definition of an appropriate reference state as a standard against which to measure deviations occurring in any particular year stems from a simple conceptual model of land cover, where (Equation 4):

$$\begin{aligned}(\text{Land cover indicator})_{\text{year } i} &= (\text{Land cover indicator})_{\text{optimum conditions}} \\ &\quad - f\,(\text{Stress factors}) \end{aligned} \tag{4}$$

The land cover indicator is a continuous variable measuring biophysical properties of the surface. In the above formulation, the state of land cover for a given year is defined by reducing the optimum conditions that could be found in a site by some function of unusual stress factors (i.e. anomalous or exceptional constraints). Since most vegetation covers are in a constant state of flux, a measure of normal or average conditions would be meaningless. It is replaced here by the values of surface attributes corresponding to the optimum conditions for a certain observation period. It is assumed that these

values occur at a time when the stresses are at a minimum. Examples of natural or anthropogenic stresses that reduce these optimum conditions are droughts, exceptional fires, floods, vegetation clearing or land degradation. These values of land cover attributes corresponding to optimum conditions are indeed specific to the area for which they have been observed. They should therefore be estimated separately for every sampling unit. They can be reached occasionally (in case of fluctuating land cover conditions) or only once, either early in the observation period (in case of a continuous decrease in vegetation cover) or at the end of the observation period (in case of reforestation, vegetation regrowth, etc.). Land cover change detection measures the physical changes brought by stresses affecting the surface. This is computed from Equation (5):

$$(\text{Land cover indicator})_{\text{optimum conditions}} - (\text{Land cover indicator})_{\text{year } i} \quad (5)$$

5.2 Change Detection Methods

5.2.1 Detection Based on a Few Isolated Observations

Classic land cover change detection techniques are based on the comparison of sequential land cover maps derived from remote sensing data or other sources for the same area. For every sampling unit of the maps, the land cover categories at the two dates are compared. The comparison of successive maps does not allow the detection of subtle changes within broad land cover classes. Since most coarse spatial resolution maps usually represent only a few broad categories, only drastic land cover conversions are identified while land cover modifications remain undetected. Moreover, change maps only exhibit an accuracy similar to the product of the accuracy of each individual map.

A land cover category is an aggregate concept which is defined by a set of biophysical attributes of the surface. Rather than detecting changes on the basis of land cover categories, change detection is better performed on the basis of the continuous variables defining these categories, whether these are reflectance values measured by a satellite sensor or biophysical attributes derived by model inversion. Singh[137] and Coppin and Bauer[138] review techniques used for this comparison, such as image differencing, image ratioing, multispectral or multitemporal change vector analysis, image regression or multitemporal linear data transformation. Empirical studies demonstrated that there is not a single optimal change detection technique but that different techniques are best suited for different change patterns.

Recent studies have detected changes in land cover between successive dates by combining two techniques[139–141] in a multistage approach:

1. postclassification comparison, that is, overlaying and comparing the two successive land cover classifications;
2. band differencing, that is, subtracting reflectances measured at the successive dates.

The postclassification comparison leads to a categorical map which indicates, for every pixel, the land cover classes at the two successive observation years.[142] Its ability to characterize ''from'' and ''to'' identifiers is essential for the definition of land cover change trajectories. The disadvantage of this method is to be dependent on the classification accuracy of the maps which are compared. The second change detection technique used in combination with the postclassification comparison further increases the reliability of the final land cover change map. The band differencing method leads to a measure of change along a continuum of change intensity. It indicates the degree to which the change has modified the vegetation cover, using the surface reflectance as a proxy. While a direct evaluation of change areas is not possible with this latter method, it allows for the detection of land cover modifications as well as of land cover conversions. Several recent studies have demonstrated the good performances obtained with the band differencing method for land cover change detection.[138,141,143,144]

When the two methods are combined, the unchanged areas identified based on the image differencing map are labeled based on the first-date land cover map. The changed areas identified based on image differencing are labeled using the result of the post-classification comparison method. This allows reduction of the misclassifications on the final land cover change map by reducing the area for which the results of the postclassification method are retained. In other words, the results of the postclassification method are only applied to the areas which have been identified as being affected by change based on the image differencing method, which is more reliable for change detection but fails to characterize the ''from'' and ''to'' land cover classes.

5.2.2 Detection Based on Continuous Time Series

Land cover changes take place at a variety of temporal scales, for example short events with detectable effects only for a few months, modifications in seasonal trajectories of ecosystem attributes, processes that affect the land cover through several seasonal cycles and long-term, permanent changes. Land cover changes may affect, and therefore be indicated by, the phenology of the vegetation cover. The analysis of the temporal trajectories of vegetation indices based on high temporal frequency remote sensing data allows monitoring of vegetation phenology and biome seasonality.[145]

Processes such as a shortening of the growing season, a dephasing of the phenology of different vegetation layers or modifications of the cover due to disturbances such as fires can only be detected if interannual changes in the seasonal trajectories of vegetation covers are analyzed. For any landscape with a strong seasonal signal, the detection of interannual changes needs to explicitly take into account the fine scale temporal variations. If data from only one or a few dates a year are used to measure interannual changes, the undersampling of the temporal series hinders the change detection accuracy and might lead to the detection of spurious changes.[83]

A quantitative evaluation of differences in seasonal development curves of remotely sensed data was applied for land cover change analysis by some authors.[146] A measure of the deviation in seasonal trajectories can be computed between successive years. Subtle processes of land cover change can be detected this way. Results of these two studies in Africa suggested that land cover changes mostly involve erratic variations in land cover conditions due to droughts, temporary modifications in seasonality, shifts in the timing of rains and episodic events to which most ecosystems display a high resilience. In a study that opens a new line of research, Goward and Prince[147] provided empirical data, measured by remote sensing, that indicate some persistence or lag between vegetation activity and climate dynamics. In some ecosystems, the response time of vegetation to short-term climatic fluctuations might provide useful information on key characteristics of the ecosystem, such as its level of degradation.

In addition to the monitoring of remotely sensed indicators of land cover changes, one can integrate a monitoring of its proximate causes or of other surface processes closely associated with the state of the vegetation cover. The best example of such process for tropical ecosystems is biomass burning. Recent research has improved the ability to monitor active fires and burnt areas by remote sensing.[148,149] On one hand, land cover changes may result from repeated biomass burning or be caused by a change in the fire regime. On the other hand, ecosystem degradation is likely to lead to a modification of the seasonal distribution and spatial patterning of fires as the diffusion of fires through the landscape will be altered.

5.3 Mapping Ecosystem Disturbances

The ecological variable of interest is often the degree of ecosystem disturbance, including both natural and anthropogenic, as well as recent and past disturbances. Disturbance refers to ''any relatively discrete event in time that disrupts ecosystem, community or population structure, and changes resources, substrate availability or the physical environment''.[150] In forest ecosystems, particular attention is paid to events that create an opening in the canopy.[151] Any vegetation cover or landscape mosaic can be conceptualized as following a development trajectory until the impact of some disturbance disrupts this development. The cover type is then forced into another trajectory, more or less remote from the initial one, depending on the magnitude, type, duration and return-interval of the disturbance. Terms like forest degradation and forest fragmentation are commonly found in the literature to describe such processes. The first expression, forest degradation, refers to changes to the forest which negatively affect the state of the forest ecosystem and, in particular, lower its biological productivity. A forest cover is termed fragmented when it is ''... broken down into disjunct parts by conversion of parts of the cover to non-forest land cover types''.[152]

The repetitive or episodic impact of disturbances on landscapes is a highly dynamic process. It is highly variable since it depends on the forcing agent (e.g. a typhoon or a fire). The detection and modeling of ecosystem disturbance therefore requires successive observations of surface conditions with a high time frequency. Repetitive observations at a scale of an entire country or a region is generally not feasible because:

1. time series of broad scale remote sensing observations (e.g. from the NOAA AVHRR sensor) date back less than two decades;
2. some areas are affected by a nearly permanent cloud cover which makes satellite sensor observation in the optical domain most difficult.

An alternative, more feasible, approach consists in defining monotemporal indicators of ecosystem disturbances. Candidate variables that could be detected by remote sensing are, for example, tree density, the spatial heterogeneity of the vegetation cover and surface moisture status (i.e. water stress at the level of the canopy). To infer the impact of disturbances on vegetation cover through a monotemporal approach requires some assumptions on the initial conditions of the landscape. Since vegetation covers are in a constant state of flux, such assumptions are easier to formulate for long-term studies, e.g. by defining, as the reference state, the vegetation cover before any human impact rather than the vegetation cover as it was supposed to be a few decades ago.

5.4 Land Cover Monitoring Systems

Some global-scale systems for monitoring tropical deforestation and dryland degradation have been put in place. Quantitative estimates of rates of change over the recent period (i.e. the last few decades) have been collected systematically for forest ecosystems by three international projects: ''Forest Resource

Assessment'',[153] ''Landsat Pathfinder Tropical Deforestation''[135] and ''TREES''.[136] These projects, as well as other tropical deforestation assessments over countries such as Brazil, all integrate remote sensing data, but with different strategies in terms of spatial and temporal sampling of observations.[154] Measuring rates of dryland degradation is a much more complex challenge given the strong interaction between erratic fluctuations in rainfall due to climatic variability and anthropogenic changes in vegetation cover.[129,130] Hulme and Kelly[155] have attempted to separate the two effects by comparing time series of a remotely sensed VI and rainfall data. They concluded that longer time series of observations would be needed before reaching any firm conclusion. Only local-scale studies backed up by field work and by time series of aerial photographs and satellite images produce convincing evidence of localized dryland degradation.[156,157] To date, there is no attempt to apply such systematic measurements at a global scale.

6 PERSPECTIVES

6.1 Model Inversion

Physically-based approaches, relying on model inversion, have been developed.[158–161] In these methods, a scene model is constructed to describe the form and nature of the energy and matter within the scene and their spatial and temporal order.[162] This scene model is coupled with an atmospheric and a sensor model. This coupled remote sensing model is then inverted to infer, from the remotely sensed measurements, some of the properties or variables of the scene which were unknown.[162] Concerning complex landscapes, these methods are not yet sufficiently robust to be applied routinely for a large-scale monitoring of land cover changes, because it is difficult to represent adequately in a model the natural variability in landscape structure and composition. For example, Franklin and Strahler[163] applied a model inversion approach to generate regional estimates of tree size and density in West Africa, using Landsat TM data. They used the Li-Strahler canopy reflectance model.[164] A woodland stand is modeled geometrically as a group of objects casting shadow on the background. The reflectance of a pixel is modeled as a linear combination of the signatures of the major scene components (i.e. illuminated tree crown, illuminated background, shadowed tree and shadowed background), weighted by their relative areas. Inversion results over a sample of observations were satisfactory for the estimation of tree density and cover but poor for tree size. The application of this type of approach for monitoring woodland degradation would probably require further developments for a more robust approach.

In the future, radiative transfer models should allow the scientific community to derive, using a physical approach, surface parameters such as: hemispherical albedo, land surface temperature and emissivity, soil moisture, snow cover, LAI, net primary production, total biomass, evapotranspiration, incident short-wave radiation, outgoing long-wave radiation, PAR and fraction of absorbed photosynthetically active radiation (FPAR). These parameters are important input variables in ecosystem models such as Soil-Vegetation-Atmosphere-Transfer (SVAT) models, energy and water models, biogeochemical models, or resource management models.

The dependence of land surface reflectance on sun and sensor viewing geometry is governed by the bidirectional reflectance distribution function (BRDF). Multiple-view-angle data of the Earth's surface (which will be provided by future satellite sensors) will permit the exploitation of the BRDF of surfaces to extract a greater number of surface attributes by remote sensing.[165,166] Actually, as the reflectance distribution of vegetation is strongly anisotropic, multiple-view-angle observations of terrestrial ecosystems contain additional and unique information beyond that acquired with nadir or single-angle spectral measurements alone.[167] Variations in the BRDF of vegetation results primarily from differences in canopy- and landscape-level structural characteristics (such as ground coverage and spatial distribution of vegetation, crown geometry, LAI, leaf angle distribution, foliage clumping), along with leaf biogeochemical and soil textural attributes.[167] Multi-view angle measurements of the vegetation BRDF will allow improved retrieval of important canopy structural characteristics (e.g. LAI) and biophysical variables (e.g. fraction of PAR absorbed by plant canopies).[168] This opens promising avenues for a more quantitative characterization of the terrestrial biosphere from space.

6.2 New Sensors and Multisensor Approaches

The coming years will see an increase and qualitative improvement in the Earth observation systems on orbit. These new sensors will include both very fine spatial resolution systems (i.e. a few meters) with a low spatial coverage and high temporal frequency, coarse spatial resolution systems, with an increase in spectral bands, sensor calibration and better georeferencing of the data. The potential for multisensor analysis (i.e. combining data from different sensors) will increase greatly. Three types of combinations will be interesting:

1. fine and coarse spatial resolution data, to combine a detailed view of the land surface over a few representative locations with an exhaustive view of the surface at a coarser spatial resolution;

2. data at approximately the same spatial resolution (fine or coarse) but from different sensors to increase the temporal frequency of coverage of a given area;
3. data at approximately the same spatial resolution but in different spectral ranges (visible, thermal, microwave) to describe more comprehensively surface processes.

Multisource data-fusion techniques are being developed for vegetation classification. Fusion of complementary data leads to a more consistent recognition of vegetation patterns. Depending on where the combining operations take place in the information extraction process, data fusion is categorized at four different levels: signal, pixel, feature and decision level.[169] There are four types of fusion methods: statistical, fuzzy logic, Dempster–Shafer evidence theory and neural network.[170] The neural network approach has the inherent ability to incorporate multisource data and also the capability to identify mixed pixels. Data-fusion techniques have led to a number of operational applications, e.g. to delineate land cover at appropriate thematic and spatial resolutions to support a quantitative inventory of forest carbon stocks in North America.[171]

ABBREVIATIONS AND ACRONYMS

AATSR	Advanced Along Track Scanning Radiometer
ASAR	Advanced Synthetic Aperture Radar
ATSR	Along Track Scanning Radiometer
AVHRR	Advanced Very High-resolution Radiometer
BRDF	Bidirectional Reflectance Distribution Function
EOS	Earth Observing System
ERS	Earth Resource Satellite
ESA	European Space Agency
FPAR	Fraction of Absorbed Photosynthetically Active Radiation
GAC	Global Area Coverage
HRV	High-resolution Visible
IGBP	International Geosphere-Biosphere Programme
JERS	Japanese Earth Resource Satellite
LAC	Local Area Coverage
LAI	Leaf Area Index
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NOAA	National Oceanic and Atmospheric Administration
PAR	Photosynthetically Active Radiation
SAR	Synthetic Aperture Radar
SPOT	Système Probatoire Pour l'Observation de la Terre
SVAT	Soil-Vegetation-Atmosphere-Transfer
SWIR	Short-wave Infrared
TM	Thematic Mapper
VI	Vegetation Index

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Global Land Databases for Environmental Analyses • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring • Semiarid Land Assessment: Monitoring Dry Ecosystems with Remote Sensing • Tropical Forest Resource Assessment by Remote Sensing

REFERENCES

1. B.L. Turner II, W.B. Meyer, D.L. Skole, 'Global Land-use/Land Cover Change: Towards an Integrated Study', *Ambio*, **23**, 91–95 (1994).
2. R.A. Houghton, 'The Worldwide Extent of Land-use Change', *Bioscience*, **44**, 305–313 (1994).
3. J.R. Mather, G.V. Sdasyuk, *Global Change: Geographical Approaches*, University of Arizona Press, Tucson, 1991.
4. W.B. Meyer, B.L. Turner II, *Changes in Land Use and Land Cover: A Global Perspective*, Cambridge University Press, Cambridge, UK, 1994.
5. W.N. Adger, K. Brown, *Land Use and the Causes of Global Warming*, John Wiley & Sons, Chichester, 1994.
6. A. Henderson-Sellers, 'Land-use Change and Climate', *Land Degrad. Rehabil.*, **5**, 107–126 (1994).
7. R.E. Dickinson, 'Land-atmosphere Interaction', *Rev. Geoph. Suppl.*, 917–922 (1995).
8. D.L. Skole, W.A. Chomentowski, W.A. Sallas, A.D. Nobre, 'Physical and Human Dimensions of Deforestation in Amazonia', *Bioscience*, **44**, 314–322 (1994).
9. Food and Agriculture Organization of the United Nations (FAO), 'Forest Resource Assessment 1990, Global Synthesis', FAO Forestry Paper, 124, FAO, Rome, 1995.

10. A. Grainger, 'An Evaluation of the FAO Tropical Forest Resource Assessment, 1990', *The Geog. J.*, **162**, 73–79 (1996).
11. E.F. Lambin, 'Modeling Deforestation Processes: a Review', TREES Publications Series B: Research Report 1, EUR 15744 EN, European Commission, Luxembourg, 1994.
12. R.K. Dixon, S. Brown, R.A. Houghton, A.M. Solomon, M.C. Trexler, J. Wisniewski, 'Carbon Pools and Flux of Global Forest Ecosystems', *Science*, **263**, 185–190 (1994).
13. K. McGuffie, A. Henderson-Sellers, H. Zhang, T.B. Durbidge, A.J. Pitman, 'Global Climate Sensitivity to Tropical Deforestation', *Global Planet. Ch.*, **10**, 97–128 (1995).
14. S.T. Pimm, G.J. Russell, J.L. Gittleman, T.M. Brooks, 'The Future of Biodiversity', *Science*, **259**, 347–350 (1995).
15. W.B. Meyer, B.L. Turner II, 'Human Population Growth and Global Land-use/cover Change', *An. Rev. Ecol. System.*, **23**, 39–61 (1992).
16. T. Binns, 'Is Desertification a Myth?', *Geography*, **75**, 106–113 (1990).
17. P. Blaikie, H.C. Brookfield, *Land Degradation and Society*, Methuen, London, 1987.
18. H. Dregne, M. Kassas, B. Rozanov, 'A New Assessment of the World Status of Desertification', *Desertif. Control Bull.*, **20**, 6–19 (1991).
19. J. Charney, P.H. Stone, 'Drought in the Sahara: A Biogeophysical Feedback Mechanism', *Science*, **187**, 434–435 (1975).
20. W.H. Schlesinger, J.F. Reynolds, G.L. Cunningham, L.F. Huenneke, W.M. Jarrell, R.A. Virginia, W.G. Whitford, 'Biological Feedbacks in Global Desertification', *Science*, **247**, 1043–1048 (1990).
21. R. McC Netting, *Smallholders, Householders, Farm Families and the Ecology of Intensive, Sustainable Agriculture*, Stanford University Press, Stanford, 1993.
22. B.L. Turner II, D.L. Skole, S. Sanderson, G. Fischer, L. Fresco, R. Leemans, 'Land-use and Land-cover Change: Science/Research Plan', IGBP Report, 35, HDP Report, 7, The Royal Swedish Academy of Sciences, Stockholm, 1995.
23. P. Aplin, P.M. Atkinson, P.J. Curran, 'Fine Spatial Resolution Satellite Sensors for the Next Decade', *Int. J. Remote Sens.*, **18**(18), 3873–3881 (1997).
24. J.P. Malingreau, 'Global Vegetation Dynamics: Satellite Observations over Asia', *Int. J. Remote Sens.*, **7**, 1121–1146 (1986).
25. S.W. Running, T.R. Loveland, L.L. Pierce, 'A Vegetation Classification Logic Based on Remote Sensing for Use in Global Biogeochemical Models', *Ambio*, **23**, 77–81 (1984).
26. K.B. Kidwell, 'NOAA Polar Orbiter Data User's Guide', NOAA NESDIS, National Climate Data Center, Washington, DC, 1986.
27. A.J. Prata, R.P. Cechet, I.J. Barton, D.T. Llewellyn-Jones, 'Along Track Scanning Radiometer for ERS-1 – Scan Geometry and Data Simulation', *IEEE Trans. Geosci. Remote Sens.*, **28**(1), 3–13 (1990).
28. V.V. Salmonson, W.L. Barnes, P.W. Maymon, H.E. Montgomery, H. Ostrow, 'MODIS: Advanced Facility Instrument for Studies of the Earth as a System', *IEEE Trans. Geosci. Remote Sens.*, **27**, 145–153 (1989).
29. C.O. Justice, E. Vermote, J.R.G. Townshend, R. Defries, D.P. Roy, D.K. Hall, V.V. Salomonson, J.L. Privette, G. Riggs, A.H. Strahler, W. Lucht, R.B. Myneni, Y. Knyazikhin, S.W. Running, R.R. Nemani, Z. Wan, A.R. Huete, W. van Leeuwen, R.E. Wolfe, L. Giglio, J-P. Muller, P. Lewis, M.J. Barnsley, 'The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research', *IEEE Trans. Geosci. Remote Sens.*, **36**, 1228–1248 (1998).
30. C.J. Tucker, P.J. Sellers, 'Satellite Remote Sensing of Primary Production', *Int. J. Remote Sens.*, **7**, 1395–1416 (1986).
31. A.R. Huete, R.D. Jackson, D.F. Post, 'Spectral Response of a Plant Canopy with Different Soil Backgrounds', *Remote Sens. Environ.*, **17**, 37–53 (1985).
32. J.P. Ormsby, B.J. Choudary, M. Owe, 'Vegetation Spatial Variability and its Effect on Vegetation Indices', *Int. J. Remote Sens.*, **8**, 1301–1306 (1987).
33. M. Kumar, J.L. Montheit, *Remote Sensing of Plant Growth, In Plants and the Daylight Spectrum*, Academic Press, London, 1981.
34. G. Asrar, R.B. Myneni, B.J. Choudhury, 'Spatial Heterogeneity in Vegetation Canopies and Remote Sensing of Absorbed Photosynthetically Active Radiation: A Modeling Study', *Remote Sens. Environ.*, **41**, 85–103 (1992).
35. R.B. Myneni, S. Maggion, J. Iaquinta, J.L. Privette, N. Gobron, B. Pinty, D. Kimes, M. Verstraete, D. Williams, 'Optical Remote Sensing of Vegetation: Modeling, Caveats and Algorithms', *Remote Sens. Environ.*, **51**, 169–188 (1995).
36. C.O. Justice, J.R. Townshend, B.N. Holben, C.J. Tucker, 'Analysis of the Phenology of Global Vegetation Using Meteorological Satellite Data', *Int. J. Remote Sens.*, **6**, 1271–1318 (1985).
37. M.M. Verstraete, B. Pinty, 'Designing Optimal Spectral Indexes for Remote Sensing Applications', *IEEE Trans. Geosci. Remote Sens.*, **GE-34**, 1254–1265 (1996).
38. J.C. Price, 'Land Surface Temperature Measurements from the Split-window Channels of the NOAA 7 Advanced Very High Resolution Radiometer', *J. Geophysical Res.*, **85**(D5), 7231–7237 (1984).
39. A.J. Prata, 'Land Surface Temperatures Derived from the Advanced Very High Resolution Radiometer and the Along-Track Scanning Radiometer 1.Theory', *J. Geophysical Res.*, **98**, 16689–16702 (1993).
40. T.N. Carlson, E.M. Perry, T.J. Schmugge, 'Remote Sensing of Soil Moisture Availability and Fractional

Vegetation Cover for Agricultural Fields', *Agricul. Forest Meteo.*, **52**, 45–69 (1990).
41. C.J. Tucker, B.N. Holben, T.E. Goff, 'Intensive Forest Clearing in Rondonia, Brazil, as Detected by Satellite Remote Sensing', *Remote Sens. Environ.*, **15**, 255–261 (1984).
42. J.L. Montheit, *Principles of Environmental Physics*, Edward Arnold, London, 1973.
43. W. Brutsaert, *Evaporation into the Atmosphere*, Kluwer Academic, Dordrecht, 1982.
44. N.J. Rosenberg, B.L. Blad, S.B. Verma, *Microclimate: The Biological Environment*, John Wiley & Sons, New York, 1983.
45. E.F. Lambin, A.H. Strahler, 'Change-vector Analysis: a Tool to Detect and Categorize Land-cover Change Processes Using High Temporal-resolution Satellite Data', *Remote Sens. Environ.*, **48**, 231–244 (1994).
46. E.F. Lambin, D. Ehrlich, 'Combining Vegetation Indices and Surface Temperature for Land-cover Mapping at Broad Spatial Scales', *Int. J. Remote Sens.*, **16**(3), 573–579 (1995).
47. R.R. Nemani, S.W. Running, 'Land Cover Characterization Using Multi-temporal Red, Near-infrared and Thermal-infrared Data From NOAA/AVHRR', *Ecol. Appl.*, **7**, 79–90 (1996).
48. S.N. Goward, G.D. Cruickshanks, A.S. Hope, 'Observed Relation Between Thermal Emission and Reflected Spectral Radiance of a Complex Vegetated Landscape', *Remote Sens. Environ.*, **18**, 137–146 (1985).
49. A.S. Hope, D.E. Petzold, S.N. Goward, R.M. Ragan, 'Simulated Relationships Between Spectral Reflectance, Thermal Emissions, and Evapotranspiration of a Soybean Canopy', *Water Res. Bull.*, **22**, 1011–1019 (1986).
50. R.R. Nemani, S.W. Running, 'Estimation of Regional Resistance to Evapotranspiration from NDVI and Thermal-IR AVHRR Data', *J. Appl. Meteo.*, **28**, 276–284 (1989).
51. S.N. Goward, A.S. Hope, 'Evaporation from Combined Reflected Solar and Emitted Terrestrial Radiation: Preliminary FIFE Results from AVHRR Data', *Adv. Space Res.*, **9**, 239–249 (1989).
52. J.C. Price, 'Using Spatial Context in Satellite Data to Infer Regional Scale Evapotranspiration', *IEEE Trans. Geosci. Remote Sens.*, **28**, 940–948 (1990).
53. F. Achard, F. Blasco, 'Analysis of Vegetation Seasonal Evolution and Mapping of Forest Cover in West Africa with the Use of NOAA AVHRR HRPT Data', *Photogram. Eng. Remote Sens.*, **56**, 1359–1365 (1990).
54. M.A. Friedl, F.W. Davis, 'Sources of Variation in Radiometric Surface Temperature over a Tallgrass Prairie', *Remote Sens. Environ.*, **48**, 1–17 (1994).
55. S.N. Goward, R.H. Waring, D.G. Dye, J. Yang, 'Ecological Remote Sensing at OTTER: Satellite Macroscale Observations', *Ecol. Applic.*, **4**, 322–343 (1994).
56. D.M. Gates, *Biophysical Ecology*, Springer-Verlag, New York, 1980.
57. D. Hillel, *Introduction to Soil Physics*, Academic Press, New York, 1982.
58. V.S. Whitehead, W.R. Johnson, J.A. Boatright, 'Vegetation Assessment Using a Combination of Visible, Near-IR and Thermal-IR AVHRR Data', *IEEE Trans. Geosci. Remote Sens.*, **24**, 107–112 (1986).
59. R.C.G. Smith, B.J. Choudhury, 'On the Correlation of Indices of Vegetation and Surface Temperature over South-eastern Australia', *Int. J. Remote Sens.*, **11**, 2113–2120 (1990).
60. R.R. Nemani, L. Pierce, S.W. Running, S. Goward, 'Developing Satellite-derived Estimates of Surface Moisture Status', *J. Appl. Meteo.*, **32**, 548–557 (1993).
61. A.S. Hope, T.P. McDowell, 'The Relationship Between Surface Temperature and a Spectral Vegetation Index of a Tallgrass Prairie: Effects of Burning and Other Landscape Controls', *Int. J. Remote Sens.*, **13**, 2849–2863 (1992).
62. R.D. Jackson, R.J. Reginato, S.B. Idso, 'Wheat Canopy Temperature: a Practical Tool for Evaluating Water Requirements', *Water Resources Management*, **13**, 651–656 (1977).
63. T.J. Schmugge, 'Remote Sensing of Surface Soil Moisture', *J. Appl. Meteo.*, **17**, 1549–1557 (1978).
64. M.S. Moran, T.R. Clarke, Y. Inoue, A. Vidal, 'Estimating Crop Water Deficit Using the Relation Between Surface-air Temperature and Spectral Vegetation Index', *Remote Sens. Environ.*, **49**, 246–263 (1994).
65. K. Grover, S. Quegan, C. da Costa Freitas, 'Quantitative Estimation of Tropical Forest Cover by SAR', *IEEE Trans. Geosci. Remote Sensing*, **GE-37**, 479–489 (1999).
66. L. Kurvonen, J. Pulliainen, M. Hallikainen, 'Retrieval of Biomass in Boreal Forests from Multitemporal ERS-1 and JERS-1 SAR Images', *IEEE Trans. Geosci. Remote Sensing*, **GE-37**, 198–204 (1999).
67. J.E.S. Fransson, H. Israelsson, 'Estimation of Stem Volume in Boreal Forests Using ERS-1 C- and JERS-1 L-band SAR Data', *Int. J. Remote Sens.*, **20**(1), 123–137 (1999).
68. A. Luckman, J. Baker, M. Honzak, R. Lucas, 'Tropical Forest Biomass Density Estimation Using JERS-1 SAR: Seasonal Variation, Confidence Limits and Application to Image Mosaics', *Remote Sens. Environ.*, **63**, 126–139 (1998).
69. J.P. Malingreau, F. De Grandi, M. Leysen, 'TREES/ERS-1 Study: Significant Results over Central and West Africa', *Earth Observation Quat.*, **48**, 6–11 (1995).
70. G.F. De Grandi, P. Mayaux, A. Rosenqvist, S. Saatchi, M. Simard, M. Leysen, 'Flooded Forest Mapping at Regional Scale in the Central Africa Congo River Basin: First Thematic Results Derived by ERS-1 and JERS-1 Radar Mosaics', Proceedings of the Workshop 'Retrieval of Bio- and Geo-physical Parameters from SAR Data for Land Applications', ESTEC/ESA, 21–23, October 1998.

71. J.M. Kellndorfer, L.E. Pierce, M.C. Dobson, F.T. Ulaby, 'Toward Consistent Regional-to-global-scale Vegetation Characterization Using Orbital SAR Systems', *IEEE Trans. Geosci. Remote Sensing*, **GE-36**, 1396–1411 (1998).
72. S.C. Liew, S-P. Kam, T-P. Tuong, P. Chen, V.Q. Minh, H. Lim, 'Application of Multitemporal ERS-2 Synthetic Aperture Radar in Delineating Rice Cropping Systems in the Mekong River Delta, Vietnam', *IEEE Trans. Geosci. Remote Sensing*, **GE-36**, 1412–1420 (1998).
73. J.P. Malingreau, G. Duchossois, 'The TREES/ERS-1 SAR '94 project', *Earth Observation Quat.*, **48**, 1–5 (1995).
74. M.G. Turner, 'Landscape Ecology: The Effect of Pattern on Process', *Annu. Rev. Ecol. Syst.*, **20**, 171–197 (1989).
75. C.P. Dunn, D.M. Sharpe, G.R. Guntenspergen, F. Stearns, Z. Yang, 'Methods of Analyzing Temporal Changes in Landscape Pattern', in *Quantitative Methods in Landscape Ecology. The Analysis and Interpretation of Landscape Heterogeneity*, eds. M.G. Turner, R.H. Gardner, Springer-Verlag, New York, 1991.
76. J.M. Briggs, J.M. Nellis, 'Seasonal Variation of Heterogeneity in the Tallgrass Prairie: a Quantitative Measure Using Remote Sensing', *Photogram. Eng. Remote Sens.*, **57**, 407–411 (1991).
77. M.G. Turner, R.H. Gardner, *Quantitative Methods in Landscape Ecology. The Analysis and Interpretation of Landscape Heterogeneity*, Springer-Verlag, New York, 1991.
78. V. Bruniquel-Pinel, J.P. Gastellu-Etchegorry, 'Sensitivity of Texture of High Resolution Images of Forest to Biophysical and Acquisition Parameters', *Remote Sens. Environ.*, **65**, 61–85 (1998).
79. C.W. Emersonn, N.S-N. Lam, D.A. Quattrochi, 'Multiscale Fractal Analysis of Image Texture and Pattern', *Photogram. Eng. Remote Sens.*, **65**, 51–61 (1999).
80. D.L.B. Jupp, J. Walker, L.K. Pendridge, 'Interpretation of Vegetation Structure in Landsat MSS Imagery: A Case Study in Disturbed Semi-arid Eucalypt Woodland. Part 2. Model-based Analysis', *J. Environ. Management*, **23**, 35–57 (1986).
81. D.E. De Pietri, 'The Spatial Configuration of Vegetation as an Indicator of Landscape Degradation due to Livestock Enterprises in Argentina', *J. Appl. Ecology*, **32**, 857–865 (1995).
82. G. Pickup, B.D. Foran, 'The Use of Spectral and Spatial Variability to Monitor Cover Change on Inert Landscapes', *Remote Sens. Environ.*, **23**, 351–363 (1987).
83. E.F. Lambin, 'Change Detection at Multiple Temporal Scales: Seasonal and Annual Variations in Landscape Variables', *Photogram. Eng. Remote Sens.*, **62**, 931–938 (1996).
84. J. Vogt, 'Characterizing the Spatio-temporal Variability of Surface Parameters from NOAA-AVHRR Data', Report EUR 14637 EN, Agriculture Series, Joint Research Centre, Institute for Remote Sensing Applications, 266, 1992.
85. A.T. Hudak, C.A. Wessman, 'Textural Analysis of Historical Aerial Photography to Characterize Woody Plant Encroachment in South African Savanna', *Remote Sens.*, **66**, 317–330 (1998).
86. E.F. Lambin, A.H. Strahler, 'Indicators of Land Cover Change for Change-vector Analysis in Multitemporal Space at Coarse Spatial Scales', *Int. J. Remote Sens.*, **15**, 2099–2119 (1994).
87. T. Kubo, Y. Iwasa, N. Furumoto, 'Forest Spatial Dynamics with Gap Expansion: Total Gap Area and Gap Size Distribution', *J. Theo. Biol.*, **180**, 229–246 (1996).
88. C. Estreguil, E. Lambin, 'Mapping Forest Disturbances in Papua New Guinea with AVHRR Data', *J. Biogeog.*, **23**, 757–773 (1996).
89. J.E. Vogelmann, 'Assessment of Forest Fragmentation in Southern New England Using Remote Sensing and Geographic Information Systems Technology', *Conserv. Biol.*, **9**, 439–449 (1995).
90. J.B. Collins, C.E. Woodcock, 'Geostatistical Estimation of Resolution-dependent Variance in Remotely-sensed Images', *Photogram. Eng. Remote Sens.*, **65**, 41–50 (1999).
91. R.S. DeFries, C.B. Field, I. Fung, C.O. Justice, S. Los, P.A. Matson, E. Matthews, H.A. Mooney, C.S. Potter, K. Prentice, P.J. Sellers, J.R.G. Townshend, C.J. Tucker, S.L. Ustin, P.M. Vitousek, 'Mapping the Land Surface for Global Atmosphere-biosphere Models: Toward Continuous Distributions of Vegetation's Functional Properties', *J. Geophysical Res.*, **100**(D10), 20 867–20 882 (1995).
92. J.A. Richards, *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin, 1993.
93. M. Hansen, R. Dubayah, R. DeFries, 'Classification Trees: an Alternative to Traditional Land Cover Classifiers', *Int. J. Remote Sens.*, **17**, 1075–1081 (1996).
94. P.M. Atkinson, A.R.L. Tatnall, 'Introduction: Neural Networks in Remote Sensing', *Int. J. Remote Sens.*, **18**(4), 699–709 (1997).
95. G.A. Carpenter, M.N. Gjaja, S. Gopal, C.E. Woodcock, 'ART Neural Networks for Remote Sensing: Vegetation Classification from Landsat TM and Terrain Data', *IEEE Trans. Geosci. Remote Sens.*, **35**(2), 308–325 (1997).
96. S. Gopal, C.E. Woodcock, A.H. Strahler, 'Fuzzy Neural Network Classification of Global Land Cover from a 1° AVHRR Data Set', *Remote Sens. Environ.*, **67**, 230–243 (1999).
97. J.R.G. Townshend (ed.), 'Improved Global Data for Land Applications', Report no. 20, International Geosphere-Biosphere Programme, Stockholm, 1992.
98. C.J. Tucker, J.R. Townshend, T.E. Goff, 'African Land Cover Classification Using Satellite Data', *Science*, **227**, 369–375 (1985).
99. J.R. Townshend, C.O. Justice, V. Kalb, 'Characterization and Classification of South American Land Cover

Types Using Satellite Data', *Int. J. Remote Sens.*, **8**, 1189–1207 (1987).
100. D. Lloyd, 'A Phenological Classification of Terrestrial Vegetation Cover Using Shortwave Vegetation Index Imagery', *Int. J. Remote Sens.*, **11**, 2269–2279 (1990).
101. T.R. Loveland, J.W. Merchant, D.O. Ohlen, J. Brown, 'Development of a Land-cover Database for the Conterminous US', *Photogram. Eng. Remote Sens.*, **57**, 1453–1463 (1991).
102. A.C. Millington, R.W. Critchley, T.D. Douglas, *Estimating Woody Biomass in Sub-Saharan Africa*, The World Bank, Washington, DC, 1994.
103. T.A. Stone, P. Schlesinger, R.A. Houghton, G.M. Woodwell, 'A Map of the Vegetation of South America Based on Satellite Imagery', *Photogram. Eng. Remote Sens.*, **60**, 541–551 (1994).
104. G. Thomas, A. Henderson-Sellers, 'Evaluation of Satellite Derived Land Cover Characteristics for Global Climate Modeling', *Climate Ch.*, **11**, 313–348 (1987).
105. J.R.G. Townshend, C. Justice, W. Li, C. Gurney, J. McManus, 'Global Land-cover Classification by Remote Sensing: Present Capabilities and Future Possibilities', *Remote Sens. Environ.*, **35**, 243–255 (1991).
106. G.D. Badhwar, H.K. Henderson, 'Application of Thematic Mapper Data to Corn and Soy-bean Development Stage Estimation', *Remote Sens. Environ.*, **17**, 197–201 (1985).
107. F. Baret, G. Guyot, 'Suivi de la Maturation de Couverts de blé par radiométrie dans les domaines Visible et Proche-infrarouge', *Agronomie*, **6**, 27–36 (1986).
108. M.Ya. Antonovski, V.M. Buchstaber, L.S. Veksler, 'Application of Multivariate Statistical Analysis for the Detection of Structural Changes in the Series of Monitoring Data', Working Paper WP-91-37, International Institute for Applied Systems Analysis, Laxenburg, Austria, 86, 1991.
109. J.B. Odenweller, K.I. Johnson, 'Crop Identification Using Landsat Temporal-spectral Profiles', *Remote Sens. Environ.*, **14**, 39–54 (1984).
110. J.P. Malingreau, 'The Vegetation Index and the Study of Vegetation Dynamics', in *Applications of Remote Sensing to Agrometeorology*, ed. F. Toselli, Commission of European Community & European Space Agency, Brussels, 285–303, 1989.
111. S.A. Samson, 'Two Indices to Characterize Temporal Patterns in the Spectral Response of Vegetation', *Photogram. Eng. Remote Sens.*, **59**, 511–517 (1993).
112. G.M. Foody, 'Ordinal-level Classification of Sub-pixel Tropical Forest Cover', *Photogram. Eng. Remote Sens.*, **60**, 61–65 (1994).
113. A.M. Cross, J.J. Settle, N.A. Drake, R.T. Paivinen, 'Subpixel Measurement of Tropical Forest Cover Using AVHRR Data', *Int. J. Remote Sens.*, **12**, 1119–1129 (1991).
114. P. Mayaux, E.F. Lambin, 'Estimation of Tropical Forest Area from Coarse Spatial Resolution Data: A Two-step Correction Function for Proportional Errors due to Spatial Aggregation', *Remote Sens. Environ.*, **53**, 1–15 (1995).
115. P.J. Brown, 'Multivariate Calibration', *J. Roy. Stat. Soc.*, **B44**, 287–321 (1982).
116. R.L. Czaplewski, G.P. Catts, 'Calibration of Remotely-sensed Proportion or Area Estimates for Misclassification Error', *Remote Sens. Environ.*, **39**, 29–43 (1992).
117. T.A. Walsh, T.E. Burk, 'Calibration of Satellite Classifications of Land Area', *Remote Sens. Environ.*, **46**, 281–290 (1993).
118. R.P. Pech, A.W. Davis, R.R. Lamacraft, R.D. Graetz, 'Calibration of Landsat Data for Sparsely Vegetated Semi-arid Rangelands', *Int. J. Remote Sens.*, **7**(12), 1729–1750 (1986).
119. R. Nelson, 'Regression and Ratio Estimators to Integrate AVHRR and MSS Data', *Remote Sens. Environ.*, **30**, 201–216 (1989).
120. L.R. Iverson, E.A. Cook, R.L. Graham, 'A Technique for Extrapolating and Validating Forest Cover across Large Regions: Calibrating AVHRR Data with TM Data', *Int. J. Remote Sens.*, **10**(11), 1805–1812 (1989).
121. Z. Zhu, D.L. Evans, 'US Forest Types and Predicted Percent Forest Cover from AVHRR Data', *Photogram. Eng. Remote Sens.*, **60**(5), 525–531 (1994).
122. W.J. Ripple, 'Determining Coniferous Forest Cover and Forest Fragmentation with NOAA-9 Advanced Very High Resolution Radiometer data', *Photogram. Eng. Remote Sens.*, **60**, 533–540 (1994).
123. P. Puyou-Lascassies, G. Flouzat, M. Gay, C. Vignolles, 'Validation of the Use of Multiple Linear Regression as a Tool for Unmixing Coarse Spatial Resolution Images', *Remote Sens. Environ.*, **49**, 155–166 (1994).
124. E.J. Van Kootwijk, H. Van Der Voet, J.J.M. Berdowski, 'Estimation of Ground Cover Composition per pixel after Matching Image and Ground Data with Subpixel Accuracy', *Int. J. Remote Sens.*, **16**(1), 97–111 (1995).
125. C.A. Hlavka, M.A. Spanner, 'Unmixing AVHRR Imagery to Assess Clearcuts and Forest Regrowth in Oregon', *IEEE Trans. Geosci. Remote Sens.*, **33**(3), 788–795 (1995).
126. F.A. Roesch, P.C. Van Deusden, Z. Zhu, 'A comparison of Various Estimators for Updating Forest Area Coverage Using AVHRR and Forest Inventory Data', *Photogram. Eng. Remote Sens.*, **61**, 307–311 (1995).
127. A. Moody, 'Using Landscape Spatial Relationships to Improve Estimates of Land Cover Area from Coarse Spatial Resolution Remote Sensing', *Remote Sens. Environ.*, **64**, 202–220 (1998).
128. A.C. Millington, P.J. Styles, R.W. Critchley, 'Mapping Forests and Savannas in Sub-Saharan Africa from Advanced Very High Resolution Radiometer (AVHRR) imagery', in *Forest-Savanna Boundaries*, ed. P.A. Furley,

J. Proctor, J.A. Ratter, Chapman & Hall, London, 1992.
129. C.J. Tucker, H.E. Dregne, W.W. Newcomb, 'Expansion and Contraction of the Sahara Desert from 1980 to 1990', *Science*, **253**, 299–301 (1991).
130. U. Hellden, 'Desertification – Time for an Assessment?', *Ambio*, **20**, 372–383 (1991).
131. C. Sagan, B.T. Owen, J.B. Pollack, 'Anthropogenic Albedo Changes and the Earth's climate', *Science*, **206**, 1363–1368 (1979).
132. A. Henderson-Sellers, V. Gornitz, 'Possible Climatic Impacts of Land Cover Transformations, with Particular Emphasis on Tropical Deforestation', *Climatic Ch.*, **6**, 231–257 (1984).
133. R.A. Houghton, J.D. Unruh, P.A. Lefebvre, 'Current Land Cover in the Tropics and its Potential for Sequestering Carbon', *Global Biogeochem. Cycles*, **7**, 305–320 (1993).
134. R.R. Nemani, S.W. Running, R.A. Pielke, T.N. Chase, 'Global Vegetation Cover Changes from Coarse Resolution Satellite Data', *J. Geophysical Res.*, **101**, D3, 7157–7162 (1996).
135. D. Skole, C. Tucker, 'Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988', *Science*, **260**, 1905–1910 (1993).
136. J.P. Malingreau, F. Achard, G. D'Souza, H.J. Stibig, J. D'Souza, C. Estreguil, H. Eva, 'AVHRR for Global Tropical Forest Monitoring: The Lessons of the TREES Project', *Remote Sens. Rev.*, **12**, 29–40 (1995).
137. A. Singh, 'Digital Change Detection Techniques Using Remotely-sensed Data', *Int. J. Remote Sens.*, **10**(6), 989–1003 (1989).
138. P.R. Coppin, M.E. Bauer, 'Digital Change Detection in Forest Ecosystems with Remote Sensing Imagery', *Remote Sensing Rev.*, **13**, 207–234 (1996).
139. P.G. Pilon, P.J. Howarth, R.A. Bullock, P.O. Adeniyi, 'An Enhanced Classification Approach to Change Detection in Semi-arid Environments', *Photogram. Eng. Remote Sens.*, **54**(12), 1709–1716 (1988).
140. S.A. Sader, 'Spatial Characteristics of Forest Clearing and Vegetation Regrowth as Detected by Landsat Thematic Mapper Imagery', *Photogram. Eng. Remote Sens.*, **61**(9), 1145–1151 (1995).
141. R.D. Macleod, R.G. Congalton, 'A Quantitative Comparison of Change-detection Algorithms for Monitoring Eelgrass from Remotely Sensed Data', *Photogram. Eng. Remote Sens.*, **64**(3), 207–216 (1998).
142. M.E. Jakubauskas, K.P. Lulla, P.W. Mausel, 'Assessment of Vegetation Change in a Fire-altered Forest Landscape', *Photogram. Eng. Remote Sens.*, **56**(3), 371–377 (1990).
143. D.M. Muchoney, B.N. Haack, 'Change Detection for Monitoring Forest Defoliation', *Photogram. Eng. Remote Sens.*, **60**(10), 1243–1251 (1994).
144. W.K. Michener, P.F. Houhoulis, 'Detection of Vegetation Changes Accuracy with Extensive Flooding in a Forested Ecosystem', *Photogram. Eng. Remote Sens.*, **63**(12), 1363–1374 (1997).
145. C.O. Justice, J.R. Townshend, B.N. Holben, C.J. Tucker, 'Analysis of the Phenology of Global Vegetation Using Meteorological Satellite Data', *Int. J. Remote Sens.*, **6**, 1271–1318 (1985).
146. E.F. Lambin, D. Ehrlich, 'Land-cover Changes in Sub-Saharan Africa (1982–1991): Application of a Change Index Based on Remotely-sensed Surface Temperature and Vegetation Indices at a Continental Scale', *Remote Sens. Environ.*, **61**, 181–200 (1997).
147. S.N. Goward, S.D. Prince, 'Transient Effects of Climate on Vegetation Dynamics: Satellite Observations', *J. Biogeog.*, **22**, 549–563 (1995).
148. C.O. Justice, J. Kendall, P. Dowty, R.J. Scholes, 'Satellite Remote Sensing of Fires during the SAFARI Campaign Using NOAA Advanced Very High Resolution Radiometer Data', *J. Geophys. Research*, **101**(D19), 23,851–23,863 (1996).
149. H. Eva, E.F. Lambin, 'Remote Sensing of Biomass Burning in Tropical Regions: Sampling Issues and Multisensor Approach', *Remote Sens. Environ.*, **64**, 292–315 (1988).
150. S.T.A. Pickett, P.S. White, *The Ecology of Natural Disturbance and Patch Dynamics*, Academic Press, New York, 1985.
151. C. Uhl, D. Nepstad, R. Buschbacher, K. Clark, B. Kauffman, S. Subler, 'Studies of Ecosystem Response to Natural and Anthropogenic Disturbances', in *Alternatives to Deforestation: Steps toward Sustainable Use of the Amazon Rainforest*, ed. A. Anderson, Columbia University Press, New York, 24–42, 1990.
152. United Nations Food and Agriculture Organization, 'Forest Resources Assessment 1990: Tropical Countries', FAO Forestry Paper, 112, FAO, Rome, 1993.
153. United Nations Food and Agriculture Organization, 'Forest Resources Assessment 1990, Global Synthesis', FAO Forestry Paper, 124, FAO, Rome, 1995.
154. M.W. Downton, 'Measuring Tropical Deforestation: Development of the methods', *Environ. Cons.*, **22**, 229–240 (1995).
155. M. Hulme, M. Kelly, 'Exploring the Links Between Desertification and Climate Change', *Environment*, **35**, 4–11, 39–45 (1993).
156. S. Lindqvist, A. Tengberg, 'New Evidence of Desertification from Case Studies in Northern Burkina Faso', *Geografiska Annaler*, **75A**, 127–135 (1993).
157. F. Sefe, S. Ringrose, W. Matheson, 'Desertification in North-central Botswana: Causes, Processes, and Impacts', *J. Soil Water Cons.*, **51**, 241–248 (1996).
158. N.S. Goel, D.E. Strebel, 'Inversion of Vegetation Canopy Reflectance Models for Estimating Agronomic Variables. i. Problem Definition and Initial Results Using Suits Model', *Remote Sens. Environ.*, **13**, 487–507 (1983).

159. X. Li, A.H. Strahler, 'Geometric-optical Bidirectional Reflectance Modeling of a Conifer Forest Canopy', *IEEE Trans. Geosci. Remote Sensing*, **GE-24**, 906–919 (1986).
160. B. Pinty, M.M. Verstraete, 'On the Design and Validation of Surface Bidirectional Reflectance and Albedo Models', *Remote Sens. Environ.*, **41**, 155–167 (1992).
161. N. Gobron, B. Pinty, M.M. Verstraete, Y. Govaerts, 'A Semi-discrete Model for the Scattering of Light by Vegetation', *J. Geophysical Res.*, **102**, 9431–9446 (1997).
162. A.H. Strahler, C.E. Woodcock, J.A. Smith, 'On the Nature of Models in Remote Sensing', *Remote Sens. Environ.*, **20**, 121–139 (1986).
163. J. Franklin, A.H. Strahler, 'Invertible Canopy Reflectance Modeling of Vegetation Structure in Semiarid Woodland', *IEEE Trans. Geosci. Remote Sensing*, **GE-26**, 809–825 (1988).
164. X. Li, A.H. Strahler, 'Geometric-optical Bidirectional Reflectance Modeling of a Conifer Forest Canopy', *IEEE Trans. Geosci. Remote Sensing*, **GE-24**, 906–919 (1986).
165. W. Wanner, X. Li, A.H. Strahler, 'On the Derivation of Kernels for Kernel-driven Models of Bidirectional Reflectance', *J. Geophys. Res.*, **100**, 21077–21090 (1995).
166. M.J. Barnsley, D. Allison, P. Lewis, 'On the Information Content of Multiple-view-angle (MVA) Images', *Int. J. Remote Sens.*, **18**(9), 1937–1960 (1997).
167. G.P. Asner, B.H. Braswell, D.S. Schimel, C.A. Wessman, 'Ecological Research Needs from Multiangle Remote Sensing Data', *Remote Sens. Environ.*, **63**, 155–165 (1998).
168. P. Bicheron, M. Leroy, 'A Method of Biophysical Parameter Retrieval at Global Scale by Inversion of a Vegetation Reflectance Model', *Remote Sens. Environ.*, **67**, 251–266 (1999).
169. C. Pohl, J.L. Van Genderen, 'Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications', *Int. J. Remote Sens.*, **19**, 823–854 (1998).
170. X. Dai, S. Khorram, 'A Hierarchical Methodology Framework for Multisource Data Fusion in Vegetation Classification', *Int. J. Remote Sens.*, **15**, 3697–3701 (1998).
171. R.S. Lunetta, J.G. Lyon, B. Guindon, C.D. Elvidge, 'North American Landscape Characterization Data Set Development and Data Fusion Issues', *Photogram. Eng. Remote Sens.*, **64**(8), 821–829 (1998).

Polar Environments Assessment by Remote Sensing

Shusun Li
University of Alaska Fairbanks, Fairbanks, USA

1 Introduction – Polar Environment, Characteristics and Importance 1
2 Remote Sensing of Snow and Ice Surface Properties and Snow and Ice-covered Areas 2
2.1 Optical Properties of Ice 3
2.2 Cloud Mask – Use of Spectral Difference at Reflective Infrared Wavelength Regions 4
2.3 Measurement of Snow Properties by Remote Sensing 4
2.4 Mapping Snow Depth Distribution Using Sequences of Snowmelt Images and Energy Balance Model 5
2.5 Use of Synthetic Aperture Radar Interferometry for Investigation of Redistribution of Snow 6
2.6 Use of Multi-frequency Polarimetric Synthetic Aperture Radar 6
3 Remote Sensing of Freeze and Thaw Status of the Surface 7
3.1 Use of Passive Microwave Brightness Temperature and Radar Backscatter to Detect Surface Freeze and Thaw Status 7
3.2 Use of Synthetic Aperture Radar Interferometry to Study Frost Heaving and Thaw Settlement 7
4 Remote Sensing of Glacier, Lake Ice and River Ice 8
4.1 Use of Satellite Synthetic Aperture Radar Images for Mapping Polar Ice Sheets 8
4.2 Use of Altimetry to Measure Ice Sheet Topography and Mass 9
4.3 Use of Microwave Brightness to Estimate Surface Accumulation of Ice Sheets 9
4.4 Use of Synthetic Aperture Radar Backscatter and Interferometry to Track Glacier Ice Motion 10
4.5 Use of Backscatter to Map Lake Ice and Lake Depth 11
4.6 Use of Visible Images and Synthetic Aperture Radar Interferometry to Map River Ice 11
5 Remote Sensing of Soil Moisture 13
5.1 Use of Microwave Brightness Temperature for Soil Moisture 13
5.2 Use of Single Band Synthetic Aperture Radar Backscatter 13
5.3 Use of Dual-polarization Imaging Radars 14
6 Remote Sensing of Arctic Vegetation and Forest Fire 14
Acknowledgments 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Polar regions are those surrounding the North and South Poles and situated within the Arctic and Antarctic Circles of the Earth. The polar regions are characterized by environments harsh to human and other organisms: low mean annual temperature and long nights and extensive snow and ice cover in winter. Because of the harsh conditions and poor accessibility, assessment of different aspects of the polar environment, such as light, temperature, snow and ice cover, soil and vegetation, largely relies on information acquired by remote sensors, i.e. airborne- and satellite-borne electromagnetic instruments. This article briefly describes how and to what extent polar environmental information is currently derived through remote sensing methods.

1 INTRODUCTION – POLAR ENVIRONMENT, CHARACTERISTICS AND IMPORTANCE

Polar environments include both polar oceans and polar land environments. The current article will be devoted only to polar land environments in both Arctic and Antarctic regions. Polar land environments are characterized by the cold climate, the extreme seasonal variation in incoming solar radiation and the consequent drastic seasonal changes in air and surface temperatures. As a result, polar regions are covered by snow and other various forms of ice at least during their long winter. In summer snow and ice may melt if the incoming radiation is sufficiently large and the environment

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

becomes sufficiently warm. In places where the ground surface eventually becomes snow and ice free, the low surface albedo of bare soil further accelerates snowmelt in the surrounding areas. With abundant solar illumination and soil moisture, vegetation flourishes on the snow-free surface during a very short period. Because water is frozen either seasonally or permanently, the polar land environments are the main land component of the terrestrial cryosphere. During the summer melt season, snow cover and glaciers release large amounts of fresh water. Seasonal and permanent frost in soils reduces the amount of water that can be stored in soils. By reducing infiltration, frozen soils can dramatically increase the runoff generated from melting snow. Assessment of the polar environment by remote sensing should focus on the investigation of surface changes related to these seasonal phenomena. With global climate change being a crucial issue in environmental studies, assessment should also be focused on inter-annual changes of the main seasonal phenomena in polar regions. Because global climate modeling studies indicate that the effects of climate changes may be amplified at higher latitudes owing to strong positive feedback of snow and ice to the surface energy balance, the assessment of long-term changes in polar regions becomes extremely important. These studies will improve our understanding of the surface processes in the polar regions and the impact of the interaction among the cryosphere, the weather and climate system, the geosphere and the biosphere.

The main issues to be discussed in this article include: (1) remote sensing of surface albedo and surface temperature in relation to snow distribution; (2) remote sensing of surface soil thaw and freeze status; (3) remote sensing of other forms of ice existing on land surfaces, including glaciers, ice sheet, lake ice and river ice; (4) remote sensing of soil moisture in tundra; and (5) remote sensing of vegetation and forest fire.

2 REMOTE SENSING OF SNOW AND ICE SURFACE PROPERTIES AND SNOW AND ICE-COVERED AREAS

Among various snow and ice surface properties, albedo and surface temperatures are of the most significant importance in meteorological and climate research. Albedo is defined as the ratio of the upwelling shortwave irradiance at the surface to the incident shortwave irradiance. Surface albedo is an important geophysical parameter because it determines the shortwave energy exchange at the Earth's surface. For weather forecasting and global climate change modeling, accurate albedo estimates are required by a general circulation model (GCM).[1–3] In the polar regions, the albedo of snow is much greater than that of any other surface cover. Because of its broad geographic distribution, snow has an important influence on the global radiation budget and global climate. The possible importance of snow and ice albedo feedback to climate change was first recognized by Croll.[4] This positive snow and ice albedo feedback is easily understood. The onset of surface melting leads to a reduction of albedo. The latter increases the absorption of shortwave energy at the surface, and results in further enhancement of the melting rate. This mechanism has proven very important in numerous simulations of global warming.[5–7] Snow and ice albedo is especially important in warm seasons when the solar illumination is one of the dominant factors of surface energy exchange. The transition from snow-covered to snow-free conditions can be very rapid. For example, tundra and prairie snow in Alaska can melt completely in 2 weeks on average. Monitoring changes of snow cover over large areas in polar environments during the melt season when access to these remote areas is extremely difficult is almost impossible without remote sensing techniques.

Temperature is another parameter of vital meteorological and climatological importance because it is a dominant factor in determining the outgoing longwave radiation from the surface and the sensible and latent heat exchange between the surface and the atmosphere. In the cold winter, the surface temperature in the polar regions can be extremely low. During the melt season, the snow and ice surface is close to isothermal conditions although there are still certain diurnal changes. After snow melt, the surface temperature of bare soil can increase rapidly owing to the low albedo of bare soils.

Since albedo is the wavelength-integrated hemispheric reflectance in solar spectra, it can be remotely sensed at visible and reflective infrared (IR) wavelengths under clear sky conditions. Surface temperature can be estimated from its brightness temperature if the surface emissivity is known. Because the emission from the terrestrial surface peaks at thermal infrared (TIR) wavelengths and surface emissivity is close to unity at those wavelengths, surface temperature can be estimated with remote sensors operating at those wavelengths. Because of the atmospheric effects, remote sensors often operate in narrow spectral windows where the atmospheric effects are minimal. These sensor response windows are always much narrower than the solar spectral region or the TIR region. Derivation of surface albedo and temperature from spectral reflectance and emission obtained from remote sensing not only requires atmospheric correction of the remotely sensed data, but also needs basic knowledge of the spectral characteristics of snow- and ice-covered surfaces. In addition, cloud cover can be a problem in remote sensing of snow and ice surfaces because clouds are similar

to snow and ice surfaces in many respects: they are highly reflective in solar spectra and can appear as cold objects when sensed in TIR wavelength bands. Therefore, separation of snow and ice surface from clouds is necessary for remote sensing of snow and ice surface properties under cloudy conditions. Since strati are common in winter and cumuli are dominant in summer in polar regions, remote sensing in solar reflective and TIR regions reveals its inherent weakness when frequent monitoring of snow and ice is needed. With much stronger penetration capability, microwave techniques can be used to provide all-weather and day-and-night information of the polar regions.

2.1 Optical Properties of Ice

Detailed study of the wavelength dependence of spectral reflectance and emissivity on snow and ice properties would lay a firm foundation of remote sensing of snow and ice properties that are of vital interest. These properties depends on the optical properties of ice.

Snow is a collection of ice grains and air and when at 0 °C it also has a significant fraction of liquid water. Snow also includes particulate impurities – dust, soot, pollen and other plant material – and chemical impurities – small amounts of major cations and anions. Thus the optical properties of snow depend on the bulk optical properties and geometry of the ice grains, the liquid water inclusions and the solid and soluble impurities. In the visible and near-infrared (NIR) wavelengths, the bulk optical properties of ice and water are very similar. The reflectance and transmittance of the snow pack in this wavelength region depend on the wavelength variation of the refractive index of ice, the grain size distribution of the snow, the depth and density of the snow pack and the size and amount of the impurities.

In the microwave region, water is much more absorptive than ice, so at those wavelengths, minute amounts of water dramatically change the electromagnetic signature of the snow pack.

The most important optical property of ice, which causes spectral variations in the reflectance of snow in the visible and NIR wavelengths, is that the absorption coefficient (i.e. the imaginary part of the refractive index) varies by seven orders of magnitude at wavelengths from 0.4 to 2.5 μm. Normally, the refractive index is expressed as a complex number, $n + ik$. Figure 1(a) and (b) shows the real and imaginary parts, respectively, of the refractive index for ice and water. The importance properties to note are (1) the spectral variation in real part n is small and the difference between ice and water is not significant; (2) the absorption coefficients k of ice and water are very similar, except for the region between 1.55 and 1.75 μm, where ice is slightly more absorptive; (3) in the visible region, k is small and ice is transparent; and (4) in the NIR region, ice is moderately absorptive, and the absorption increases with increase in wavelength.[8]

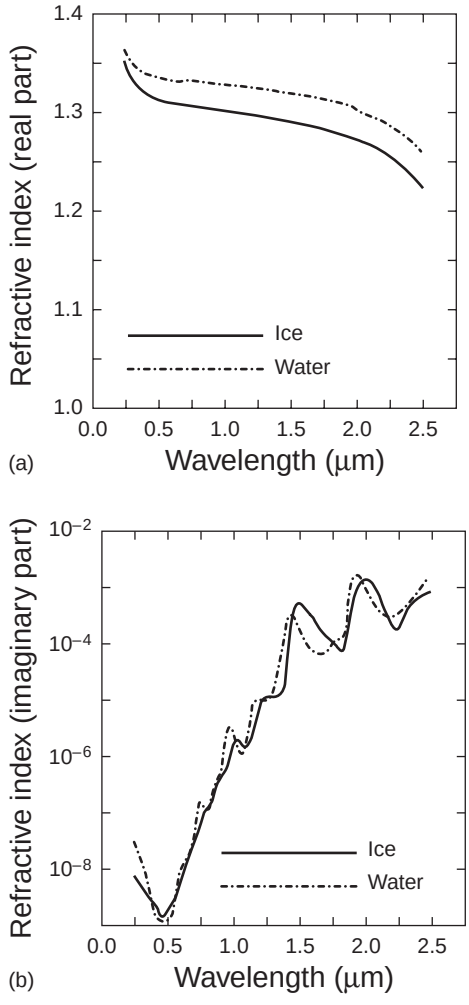

Figure 1 Complex refractive index ($n + ik$) of ice. (a) Real part of refractive index (n); (b) imaginary part of refractive index (k). (Reproduced by permission from Dozier.[8])

Using a radiative transfer-based snow model,[9,10] the spectral reflectance of snow can be calculated. Figure 2 shows the spectral reflectance of pure deep snow for visible and NIR wavelengths, for snow grain radii from 50 to 1000 μm. This snow grain radius range represents a range for new snow to spring snow, although the grain clusters in coarse spring snow can exceed 0.5 cm in radius. Because ice is so transparent at visible wavelengths, increasing the grain size does not appreciably affect the reflectance. In the NIR region, however, ice is moderately absorptive. Therefore, the reflectance is sensitive to grain size, and the sensitivity is greatest at 1.0–1.3 and at 1.55–1.75 μm. Because the ice grains are strongly forward scattering at NIR wavelengths, the reflectance increases with increase in illumination angle (Figure 3), especially for larger grains.

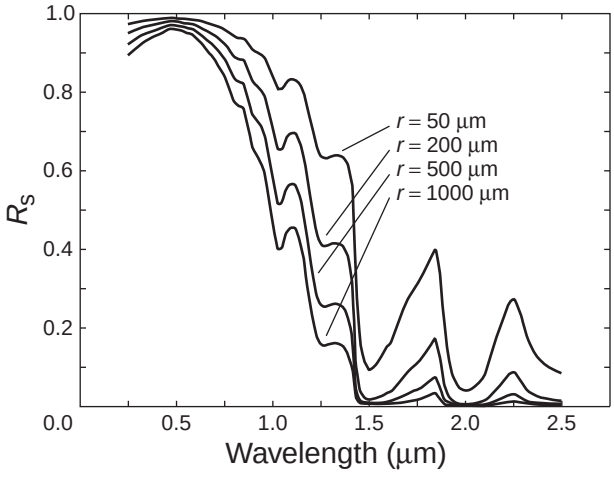

Figure 2 Directional–hemispherical reflectance R_s of snow at an illumination angle $\theta_0 = 60°$ for wavelengths from 0.4 to 2.5 μm and grain size r from 50 to 1000 μm. (Reproduced by permission from Dozier.[8])

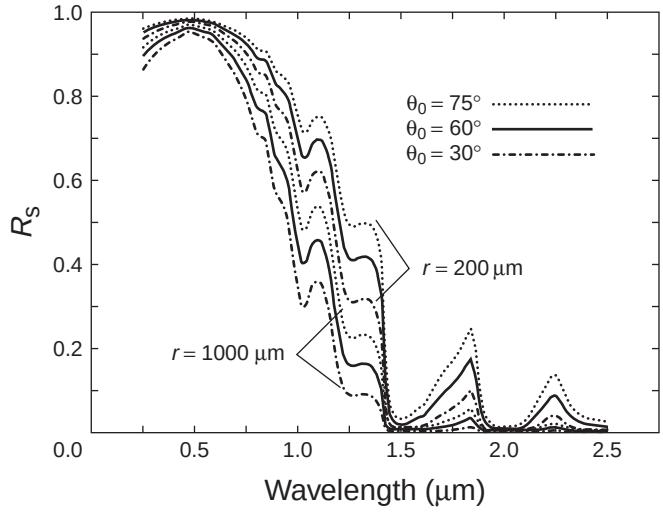

Figure 3 Directional–hemispherical reflectance R_s of snow at illumination angle $\theta_0 = 30°$, 60° and 75°, for wavelengths from 0.4 to 2.5 μm and grain size $r = 200$ and 1000 μm. (Reproduced by permission from Dozier.[8])

According to the radiative transfer model, the presence of liquid water in the snow does not by itself affect the reflectance. The observed changes in reflectance during snow melt result from the increased crystal sizes and from an effective size increase caused by the two-to-four grain clusters that form in wet unsaturated snow.[11,12]

At visible wavelengths, reflectance is insensitive to snow grain size, but is affected by two variables, finite depth and the presence of absorbing impurities. For pure snow with grain radius $r = 1000$ μm, the reflectance is perceptibly reduced when the snow depth is reduced to 100 mm snow water equivalence. Warren and Wiscombe[10] also showed that minute amounts of absorbing impurities reduces snow reflectance at visible wavelengths, where pure ice is highly transparent. Their calculation shows that soot concentrations as low as 0.1 ppmw (parts per million by weight) are sufficient to reduce reflectance. The effect of the absorbing impurities is apparently enhanced when they are inside the snow grains because refraction focuses the light on the absorbers.[13,14]

2.2 Cloud Mask – Use of Spectral Difference at Reflective Infrared Wavelength Regions

Both snow and clouds have high reflectance in visible wavelength. This is understandable because snow pack is composed of snow grains and clouds consist of either water droplets or ice crystals. Because of their spectral similarity, discrimination of snow and clouds can be a challenging task in remote sensing for snow and ice applications. Fortunately, snow and cloud discrimination is possible using the spectral information in the region between 1.55 and 1.75 μm, where ice is more absorptive and reflectance is most sensitive to ice grain sizes. In this wavelength region, water clouds have higher reflectance than snow, because water droplets are less absorptive. Also, ice clouds (such as cirrus) are more reflective than snow because the former is composed of ice crystals (a few micrometers at most) that are much smaller than the snow grains in regular snow pack (50–1000 μm). Therefore, the bands of the Landsat Thematic Mapper (TM) allow snow–cloud discrimination.[15,16] Using color images composed of TM bands 2, 5 and 7, clouds can clearly discriminated from snow by spectral features. A new example given by Riggs et al.[17] uses wavelength bands at 0.547, 0.869 and 1.609 μm to form a color composite which clearly shows the difference among sea ice, open water and clouds.

2.3 Measurement of Snow Properties by Remote Sensing

The advanced very high-resolution radiometer (AVHRR) on the National Oceanic and Atmospheric Administration (NOAA) operational polar orbiting satellites scans across-track up to 56° off nadir and therefore views the surface at widely varying view angles. The AVHRR sensor consists of one visible, one NIR and two TIR channels, and also one mid-infrared (MIR) channel that is sensitive to a combination of thermal and reflected solar fluxes. The visible and NIR radiance measurements of the AVHRR sensor have been used for mapping snow-covered areas.[18]

In the late 1980s, the National Aeronautics and Space Administration (NASA) started an ambitious remote

sensing program, the Earth Observing System (EOS). EOS is designed to provide long-term observations in the area of climate and terrestrial and marine ecosystems and the supporting information system necessary to develop a comprehensive understanding of how the Earth functions as a unified system.[19] After a continuous, decade-long effort, NASA is finally about to launch Terra, the EOS "flagship". Terra will begin collecting a new global data set on which to base future scientific investigations on our complex home planet. The moderate resolution imaging spectroradiometer (MODIS) is one of the key instruments on Terra and her successors. MODIS will acquire reflection and emission of the Earth system daily in 36 bands at spatial resolutions of 250 m (bands 1–2), 500 m (bands 3–7) and 1000 m (bands 8–36).[20,21] The multi-angle imaging spectroradiometer (MISR) is another key instrument on Terra. MISR will image the Earth simultaneously at nine different angles in each of four color bands. With MISR, it is feasible to study every aspect of the Earth's climate system that scatters sunlight differently at different angles.[22]

The NASA science team will use MODIS data to develop a global snow map as a routine product.[23,24] The global snow cover will be mapped daily at 500-m resolution using MODIS bands 1, 2, 4 and 6. A pixel will be designated as snow covered if it is approximately 50% or more snow covered. The expected accuracy of the MODIS-derived snow map will vary with land cover type. According to field, aircraft and satellite measurements, the accuracy is 15% in forest, 10% in mixed agriculture and forest areas and 5% in prairie areas, barren and sparsely vegetated areas, wetlands, tundra and permanent snow and ice fields.[24]

The AVHRR images are also used for the derivation of surface spectral albedo in polar regions under clear sky conditions with the atmospheric and surface anisotropic effects corrected, and the resulted narrow-band reflectances are further converted into a broad-band albedo (0.4–3.0 μm).[25] Comparison of the AVHRR-derived surface albedo over the Greenland Ice Sheet with field measurements indicates that before the onset of snow melt, agreement is good (<5%) between the satellite-derived and field measured surface albedo. After melt begins, it is difficult to compare satellite values with the field data owing to melt water ponding on the ice sheet surface. Because AVHRR has a nominal spatial resolution of 1 km, it could not map the details of surface albedo in mountainous areas where variations in solar illumination make the derivation of snow surface albedo more difficult, or during snowmelt when the surface albedo shows greater spatial variations. Sensors with finer spatial resolution are needed for such cases. Shi et al.[26] used images from the Landsat TM to derive surface albedo with the help of a digital elevation model (DEM).

The accuracy of the satellite-derived surface albedo depends in part on the accuracy of the primary atmospheric input variables used in the atmospheric correction. Most of the current space-borne sensors, such as AVHRR and Landsat, themselves, are not able to provide sufficient information on atmospheric conditions (especially the amounts of aerosols) necessary for accurate albedo derivation. The accuracy of the satellite-derived surface albedo also depends on the appropriateness of surface anisotropic correction. In the previously mentioned example of derivation of surface albedo using AVHRR images, Stroeve et al.[25] used a radiative transfer model to simulate the anisotropy of the snow surface reflectance.

In the EOS era, because measurements will be made in a large number of spectral bands through MODIS and over a wide range of viewing directions through both MODIS and MISR, the atmospheric conditions and surface anisotropy can be accurately derived from those sensors. Consequently, surface albedo can be more accurately estimated. In fact, the bidirectional reflectance distribution function (BRDF)/albedo product is listed as one of the high-level products to be derived. The BRDF provides a quantitative representation of surface reflectance for any type of surface, including Lambertian and anisotropic surfaces.

MODIS BRDF/albedo is a general-purpose product to be derived from Terra satellite instruments. Because of the large number of spectral channels available on the MODIS sensor and the various sensor viewing angles associated with MISR, the combination of MODIS and MISR is superior to any existing satellite sensors for correction of the atmospheric and surface anisotropic effects.[27] Two sub-types of albedo products will be produced: (1) simple nadir broad-band surface albedo for global climate models that require shortwave energy partitioning; and (2) multiangle spectral albedos for other advanced MODIS products.[27] Meanwhile, the BRDF will be generated. The BRDF is a key element in the derivation of surface albedo because it specifies the behavior of surface scattering as a function of illumination and view angles at a particular wavelength.

2.4 Mapping Snow Depth Distribution Using Sequences of Snowmelt Images and Energy Balance Model

To assess the hydrological or climatic impact of tundra and prairie snow, the most fundamental information needed is the snow depth distribution. The depth distribution evolves and changes throughout the winter, but typically reaches a maximum near the end of winter. This is the initial distribution for modeling the spring run-off, assessing winter frost penetration under the snow and describing the evolution of the albedo as the snow melts.

The end-of-winter snow depth distribution in tundra, steppe and prairie can be recovered from a sequence of images of the snow-covered area acquired during the snowmelt if an energy balance model is used to calculate the melt rate.[28] Images acquired at visible and NIR wavelengths are the most straightforward means for mapping these snow-covered areas because the snow has a much higher albedo than the surrounding darker ground. However, frequent occurrences of clouds over tundra areas such as Alaska's Arctic Slope reduce the probability of obtaining a suitable suite of images during the rapid melt, which often takes less than 10 days. Gaps in the image sequence can only be bridged by images acquired by microwave sensors that are capable of mapping the Earth's surface even under overcast conditions and during periods of dark.

Using Cline's idea, we collected a snowmelt sequence at a study area at Ivotuk, Alaska, during the 1999 melt season, using a combination of air photographs, Système Probatoire Pour l'Observation de la Terre (SPOT) and AVHRR visible images acquired under clear sky conditions, and Earth Resources Satellite (ERS) synthetic aperture radar (SAR) and RADARSAT SAR images acquired in both clear and cloudy conditions.[29] The combined image sequence depicts the details of a 10-day snowmelt process. The results will be used with a melt model to recover the end-of-winter snow depth distribution in the study area. The study also shed light on the relation between radar backscatter and snow pack conditions.

2.5 Use of Synthetic Aperture Radar Interferometry for Investigation of Redistribution of Snow

Recently, satellite SAR interferometry, available via the European Space Agency (ESA) First ERS-1, has provided all-weather and day-and-night capabilities for detecting subtle surface changes.[30] Using satellite SAR interferometry, we produced the basic pattern of redistribution of wind-drifted snow in the Alaskan Arctic region.[31] The coherence image of an interferogram measures the resemblance of radar phases between two acquisitions a few days apart in the vicinity of each pixel location. Formation and evolution of sastrugi, or snow dune, on a new snow slab with a certain thickness can significantly reduce the phase coherence. As a result, areas and times with and without significant snow accumulation and drifting are revealed very clearly on the coherence images of the interferograms. After a new snow fall and a following strong west wind event in February 1994, patterns of relatively high phase coherence on the windward slopes and low coherence on the leeward slopes and vast tundra form the most striking features on the interferograms which cover the middle of February through the middle of March 1994. These patterns show less accumulation of new snow on the windward slopes due to wind acceleration and an opposite process on the leeward slopes. This finding not only agrees with previous field investigations, in the sense that small-scale terrain relief of the order of meters actively affects the redistribution of wind-drift snow on the tundra, but also leads to an improved knowledge that large-scale topography with surface relief of the order of tens to hundreds of meters plays a key role in determining the larger scale redistribution patterns of new snow fallen on the North Slope of Alaska. The interferograms also identify an event with bands in wind, i.e. bands a few kilometers wide and hundreds of kilometers long, possibly caused by bands of different concentrations of ice particles suspended in the air. This is a phenomenon that observers noticed in previous field investigations but were never able to have a big picture of it.

2.6 Use of Multi-frequency Polarimetric Synthetic Aperture Radar

As mentioned in the previous section, the application of SAR images with single wavelength and single polarization in environmental studies requires terrain correction of the SAR images to remove the geometric distortions inherent in the slant range mapping of radar sensors, and to normalize radar backscatter on a sloping surface.[32,33] Unlike single polarization SAR, polarimetric SAR provides information in full polarization, including horizontal transmit, horizontal receive (HH), vertical transmit, vertical receive (VV), horizontal transmit, vertical receive (HV) and vertical transmit, horizontal receive (VH). The polarimetric data allow pixel-by-pixel derivation of the complete scattering matrix and thereby detailed information about the geometric structure and dielectric property of a target.[34] The concept of application of a combination of multifrequency and full polarization to map snow is investigated in the SIR-C/X-SAR (shuttle imaging radar-C and X-band synthetic aperture radar) experiment. SIR-C/X-SAR is a joint project between NASA, the German Space Agency (DARA) and the Italian Space Agency (ASI). SIR-C provides radar polarimetric images simultaneously at the L-band (24 cm) and C-band (5.6 cm). The X-band (3 cm) SAR with VV polarization provides information in the third frequency. Because NASA space shuttles cover only the mid-low latitude, no high-latitude examples are available.

Using SIR-C/X-SAR, Shi and Dozier[35] developed a procedure to map wet snow without using a DEM. Although the experiment was made over site at a mid-latitude site at Mammoth Mountain, California, the procedure is potentially applicable to polar regions.

The basic ground cover types during the experiment include coniferous forest, short vegetation such as brush and grass, bare soil, snow and lake. The directly measured polarimetric values at each pixel are used to derive 20 polarimetric quantities at the L- and C-bands and 10 backscatter ratios between different frequencies. Because the derived polarimetric quantities and backscatter ratios are various forms of ratios among the basic polarimetric measurements on a pixel basis, the terrain effects are cancelled. Consequently, the derived values can be used for snow mapping without using a DEM. Using those derived quantities, a decision tree classifier (DTC) was implemented based on training sets. On April 17 1994, when the snow was wet, the SAR-derived snow map was about 77% as accurate as the TM binary classification. The example demonstrates the potential of multifrequency SAR polarimetry in snow mapping.

3 REMOTE SENSING OF FREEZE AND THAW STATUS OF THE SURFACE

3.1 Use of Passive Microwave Brightness Temperature and Radar Backscatter to Detect Surface Freeze and Thaw Status

Measurements of the Earth's surface in the microwave regions is largely insensitive to weather conditions and solar illumination. Microwave sensors appear ideal to measure the properties of the cryosphere because the microwave signal is sensitive to the dielectric constant of surface materials, which in turn is highly sensitive to the phase of water, ice or liquid.

Passive microwave sensors have demonstrated sensitivity to snow and surface freeze–thaw transitions. Ground-based measurements indicate that the surface brightness temperature increases substantially when wet soil is freezing, owing to a significant increase in emissivity even though the soil temperature is decreasing.[36,37] When multiple-frequency microwave brightness temperature data are available, such as those acquired by the Nimbus 7 scanning multi-channel microwave radiometer (SMMR), the spectral gradient of brightness temperature provides further information of the soil frozen and thaw status. The 10.7–37 GHz brightness temperature spectral gradient is especially useful because wet soils have high brightness temperature spectral gradient values whereas frozen soils have low values.[38–40] These facts make it feasible to classify four types of soils, frozen, hot (and dry), wet (and cool) and mixed soils, using spaceborne microwave radiometers.[40] The major problem with passive microwave radiometers is that their spatial resolution is too coarse (>10 km) to obtain information adequate for the analysis of hydrological processes at relatively small scales.

Also having all-weather and day-and-night capabilities, satellite SAR maps the surface at spatial resolutions much finer than passive microwave radiometers. The radar backscatter is very sensitive to the freeze and thaw status of moisture in soil and vegetation because freezing results in a dramatic decrease in the dielectric constant due to the cessation of the rotation of the polar water molecules. Ground-based scatterometer measurements show a 3–6 dB drop in surface backscatter coefficients when bare soil freezes.[36,37] A similar drop in backscatter was observed with freezing at a forest site in interior Alaska using C-band, VV polarization SAR imagery acquired from ERS-1.[41] Radar backscatter was observed to rise by 4 dB in Siberia during the spring thawing process using time series of the ERS scatterometer and SAR images.[42,43]

3.2 Use of Synthetic Aperture Radar Interferometry to Study Frost Heaving and Thaw Settlement

SAR interferometry is also used to detect frost heaving of the active layer over permafrost.[44] Permafrost plays an important role in polar and global climatology, primarily owing to its unique characteristic in the ground thermal regime, inhibition of the groundwater recharge and movement, enhancement of runoff,[45] restriction to plant growth, and sensitive response to moisture and trace gas exchanges (e.g. CO_2 and CH_4) between the atmosphere and polar soils. Recent studies show that global warming can turn the permafrost region from a net carbon sink to a source, thus acting as a positive feedback to global climate change.[46] Monitoring changes in the timing and duration of the active layer freeze–thaw cycle and thickness of the active layer provides an extremely sensitive way to study climatic change.[47] However, it is difficult to monitor the changes over large areas through in situ measurements under the Arctic weather conditions. A study site was chosen near the Toolik Lake within the Imnivait watershed at the foothills of the Brooks Range, Alaska. The site is located in a region of continuous permafrost. The low boundary of the permafrost is between 250 and 300 m below the surface. Although the topography is gentle and the elevation range is small, the thickness of the active layer over permafrost varies significantly as a function of hill slope position. A deformation pattern was derived from a pair of ERS-1 SAR images obtained on October 4 and August 31, 1995. Terrain effects on fringes in both interferograms were removed using a third interferogram made from a pair of ERS-1 and -2 SAR images acquired on October 4 and 5, 1995. During the period time, the active layer began to freeze, and hence the surface was

Figure 4 Color composite of interferograms near Toolik Lake, Alaska. The green component of the composite is made of a differential interferogram derived from a pair of SAR images acquired on August 30 and October 4, 1995. The blue component is made of the coherence image of the same interferogram. The red component is made of a differential interferogram derived from a pair of SAR images acquired on August 30, 1995 and August 15, 1996. Terrain effects on fringes in both interferograms have been removed using a third interferogram made of a tandem pair of ERS-1 and -2 SAR images acquired on October 4 and 5, 1995, respectively. The green tone at the top of the hills and along the river channels represents a frost heave 1–1.5 cm higher than its surrounding red toned area for the period between August 30 and October 4, 1995.

uplifting. In the resulting interferogram (Figure 4), at the top of hills and along main river channels, during the early freezing season between later August and early October, frost heave is 1–1.5 cm higher than at the valley bottom. This pattern matches well the spatial variations of the thickness of the active layer near the Toolik site. According to Walker et al.,[48] the active layer is around 0.7–1.0 m thick near the top of the hills, and reduces to 0.3–0.4 m at the valley bottom. The active layer along the river channels is also thick because of the thermal effects of stream water. The derived pattern of differential frost heaving can be explained by freezing of about half of the active layer with a water content up to 60% by volume and a thickness difference between 30 and 70 cm.

4 REMOTE SENSING OF GLACIER, LAKE ICE AND RIVER ICE

4.1 Use of Satellite Synthetic Aperture Radar Images for Mapping Polar Ice Sheets

Satellite SAR images have been used to map ice sheets in Greenland and in Antarctica. ERS-1 SAR images have been used to form a mosaic of the ice sheet in Greenland.[49] The mosaic reveals the basic zonation of the ice sheet. The zones have backscatter signatures related to the structure of the snowpack, which varies with the balance of accumulation and melt at various elevations. From the center of the ice sheet towards its edge, there exist four distinct zones (or facies). The central part is the dry snow facies. The dark signal in radar backscatter of this zone results from the low-volume scattering of fine, dry snow grains in the cold, deep snow pack of the dry snow facies. Outside the dry snow facies is the percolation facies. In the percolation facies, where vertical ice fingers or pipes and horizontal ice lenses develop, a bright radar signature is caused by strong double bounce of the radar beams between the ice pipes and ice lens, and surface scattering from the ice surfaces. Further outside, the dark appearance of glaciers with moist snow cover is the consequence of the strong microwave absorption within the moisture-rich snow pack. The bare ice in the outmost part of the ice sheet reveals a moderate backscatter because of lack of the microwave-absorptive moist snow layer. The boundaries of zones can be accurately located on the mosaic. The SAR images also reveal a large flow feature in northeast Greenland that is similar to ice streams in Antarctica and may play a major role in the discharge of ice from the ice sheet.

In normal operation, all of the current and previous polar orbiting satellite SARs have been oriented to the right side of the orbital plane to maximize coverage of the Arctic regions. This normal operation leaves a relatively small area without coverage at the North Pole and a larger hole at the South Pole. In order to obtain coverage of Antarctica, a special in-orbit 180° yaw maneuver of the Radarsat was performed to re-orient the satellite to the left-looking configuration. The new orientation is referred to as the Antarctic Mode and it covers the entire Antarctica. Figure 5 shows the coverage of Antarctica during the Antarctic Mode.

The Antarctic Mode was implemented during the Radarsat Antarctic Mapping Project (RAMP) in September and October 1997. Approximately 4000 SAR images were collected during that time period. The image data were combined with a recently created DEM for removing geometric distortion and making radiometric normalization. The results form a terrain-corrected SAR mosaic of the entire ice sheet of Antarctica. The mosaic reveals

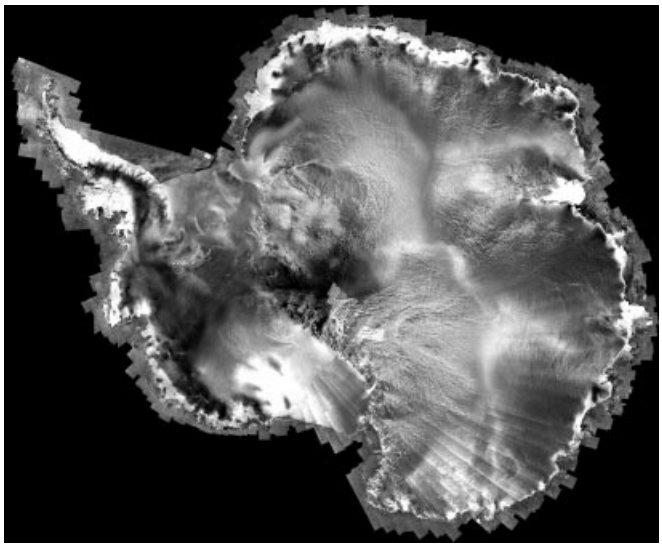

Figure 5 SAR mosaic of Antarctica generated by Jezek's group at Ohio State University using images acquired by Canadian RADARSAT SAR sensor. (Reproduced by permission from ftp://iceberg.mps.ohio-state.edu/pub/JEZEK/mos-5km.gif.)

extraordinary details about the Antarctic ice sheet. It provides a new view of Antarctica and permits quantitative analysis of surface properties over all of East and West Antarctica (Figure 5).

In addition, Radarsat SCANSAR captured the break-off of a large tabular iceberg from the Ronne Ice Shelf, Antarctica, that occurred over the period from October 20 to November 1, 1998 (at http://www.asf.alaska.edu/user_serv/feature.html). The iceberg, identified as A-38, measures approximately 150×50 km (larger than the area of the state of Delaware) and is seen to break up into smaller pieces as it drifts northwards. The resulting pieces were identified as A38A, -B, and -C. The northernmost section of the iceberg, A38B, is seen to drift 50 km over the 10-day period involved in this animation. Note the top left section of ice shelf that breaks off after being struck by the main iceberg. The image showing the ice shelf prior to the iceberg breaking off was acquired approximately 1 year earlier.

4.2 Use of Altimetry to Measure Ice Sheet Topography and Mass

Altimeters are nadir-viewing instruments which transmit short-duration electromagnetic pulses with known power in a pencil beam towards the Earth's surface, then measure the reflected energy in a number of time gates. The signal amplitudes in successive gates create a waveform. The time delay until receipt of the reflected signal, when coupled with a knowledge of the velocity of propagation through the ionosphere and wet troposphere, can be converted to an accurate measurement of the elevation of surface topography, given that the orbit ephemeris is accurately determined.

Satellite radar altimetry has a history of more than two decades. The first such instrument was on board GEOS 3 in 1975. Because it only covered areas between 60° north and south latitude, it missed almost all ice sheets. Seasat, launched in 1978, extended coverage to 72° north and south latitude. However, it only lasted for 98 days. Geosat came next, with a coverage similar to that of Seasat, and successfully operated from 1985 to 1992. The ERS series began in 1991 and extended coverage to 81.5° north and south latitude.

Because the antenna size of these altimeters range from 1 to 2 m in diameter, they have a large beam footprint on the surface of the Earth. Altimeter data over ice sheet should be corrected to remove slope-induced errors. Further, measurements are adjusted at crossovers, where successive satellite paths intersect.[50] ERS-1 altimeter data have been processed with this method.[51] The resulting DEMs with a 10-km footprint covers the areas of the Antarctica to the north of 81.5 °S. Further south, the hole was filled using an airborne-based data set. This data set is an important component of a recently created DEM of Antarctica, which gains improvement in a Geographic Information System (GIS) environment through integrating the best available topographic data from a variety of sources.[52] The available Antarctica DEM serves as an input to ice sheet flow models, and helps determine the location of features such as grounding lines and correct geometric radiometric distortions inherent in SAR images.

Measurement of elevation change by space- or airborne altimeters offers a method of determining changes in ice volume and therefore mass balance. The distribution of crossover differences in an area tends to resolve regional changes in elevation.[50] The technique has been used to estimate decadal elevation changes of ice sheets in south Greenland[53] and East Antarctica.[54] Considering the large spatial (-15 to $+18\,\mathrm{cm\,yr^{-1}}$) and temporal variations ($\pm 15\,\mathrm{cm\,yr^{-1}}$) in the former case, the derived average growth rate of $1.5 \pm 0.5\,\mathrm{cm\,yr^{-1}}$ is too small to determine if the Greenland ice sheet is undergoing a long-term change due to a warmer polar climate. In the latter case, inconsistency of the adjusted Seasat to Geosat vs Seasat to ERS-1 results was found. Both cases indicate the need for further improvements in satellite altimetry using laser techniques and more accurate satellite orbit tracking methods.

4.3 Use of Microwave Brightness to Estimate Surface Accumulation of Ice Sheets

Accurate estimation of the annual rate of net mass accumulation at the surface of the Antarctic and Greenland

ice sheets is a key element in global climate modeling. In addition to estimation of the accumulation rate by measurement of ice sheet surface elevation changes through altimetric techniques, the rate can also be estimated from passive microwave brightness temperatures.

The temperature at the surface of ice sheets in Antarctica and interior Greenland is always well below freezing point. For dry snow, the emissivity of a snowpack depends on the grain size, which in turn depends on accumulation rate. With time, snow grains metamorphose to ever larger grains until they eventually form pure ice. Hence low-accumulation regions have relatively older, and therefore relatively larger, grains nearer the surface than do high-accumulation regions. Based on radiative transfer theory, snowpack in low-accumulation regions ought to be more reflective at microwave wavelengths. According to energy conservation, such regions tend to be less absorptive at microwave wavelengths. Kirchhoff's law further indicates that a less absorptive surface must be less emissive under thermodynamic equilibrium. Consequently, the snowpack in low-accumulation regions has lower emissivity than that in the high-accumulation regions of the ice sheet, as indicated by theoretical analysis and microwave measurements.[55,56]

The microwave emissivity of snowpack is estimated by the mean annual passive microwave brightness temperature divided by the mean annual surface temperature, based on Nimbus-5 electronically scanning microwave radiometer (ESMR) and Nimbus-7 TIR data. Meanwhile, two coefficients are determined for each region or each ice sheet to fit best the field-measured accumulation rates with a theoretical-based hyperbolic function. Once the coefficients have been determined, the microwave emissivity and mean annual surface temperature are used to estimate the net mass balance. The method was used to derive accumulation values in Greenland and Antarctica. The results appears to be reliable for the areas of dry-snow facies. Vaughan et al.[57] refined the method for Antarctica through improvements in basin delineation using automated techniques on an improved DEM, digital calculation of surface areas and basin-wide integrations and interpolation of sparse data using independent background field. Their results indicate that the total net surface mass balance for conterminous grounded ice sheet is 1811 Gton yr^{-1} and for the entire ice sheet including ice shelves and embedded ice rises 2280 Gton yr^{-1}. The uncertainty is at least ±5%.

4.4 Use of Synthetic Aperture Radar Backscatter and Interferometry to Track Glacier Ice Motion

SAR backscatter images have been used to track ice motion in glaciers and ice sheets. In essence, this technique is no different from the determination of ice motion using visible images using cross-correlation techniques.[58,59] The cross-correlation technique using SAR backscatter has been applied to measure the surface ice velocities on Greenland ice sheet,[49] Pine Island Glacier in West Antarctica[60] and Malaspina Glacier in Alaska.[61]

The surface ice flow in northeast Greenland was derived by computer-determined displacement of features in two images acquired 105 days apart.[49] Results for the middle reach of Lakobshavns Isbrae show that the ice flow in this area is extremely rapid; the velocity is as high as 2 km yr^{-1}. The accuracy of the velocity derived is about ±80 m yr^{-1} based on a one-pixel misregistration between the two images; this error is proportional to the inverse of the time interval.

Pine Island Glacier is one of the major ice streams draining West Antarctica. Lucchitta et al.[60] calculated average velocities for both its grounded and floating parts by tracking crevasses and other patterns moving with ice on two sequential images acquired in February and December 1992 by ERS-1 SAR. Velocities in the fast-moving central parts of the glacier ranged from about 1.3 km yr^{-1} on the grounded part to 2.6 km yr^{-1} on the floating part. The velocity increases rapidly just below the grounding line. For the floating part, velocities are approximately 0.3 km yr^{-1} faster than previously established from Landsat images. The new observation suggests that the discharge is larger and, therefore, the mass balance less positive than previously thought.

Malaspina Glacier is a large piedmont lobe on the coast of southcentral Alaska fed by glaciers from the St. Elias Mountains. It is known for its size and the immense folded moraines on its surfaces. The surface moraines move with the ice, forming patterns clearly recognizable in satellite imagery. From an exact repeat orbit, ERS-1 SAR images with a pixel spacing of 12.5 m were acquired on June 29, 1992 and June 14, 1993.[61] The US Geological Survey (USGS) DEM of a 3 arcsec resolution of the same area was re-sampled to produce a DEM of 30 m spacing through an optimum interpolation method such as kriging.[62] The resulting DEM was used for removing the geometric and radiometric distortions on each SAR image. Then the cross-correlation method was applied to the entire image area on the 20 × 20 pixel grid to map surface displacement (Figure 6). The maximum velocity derived was 399 m yr^{-1}. The velocity vectors close to the eastern edge of Malaspina Glacier are mostly in the range 80–300 m yr^{-1}. Because terrain correction is performed and a long time interval of 1 year is involved, an accuracy of 30 m yr^{-1} is expected. A comparison with earlier measurements implies that the current results are consistent with those obtained using the more traditional

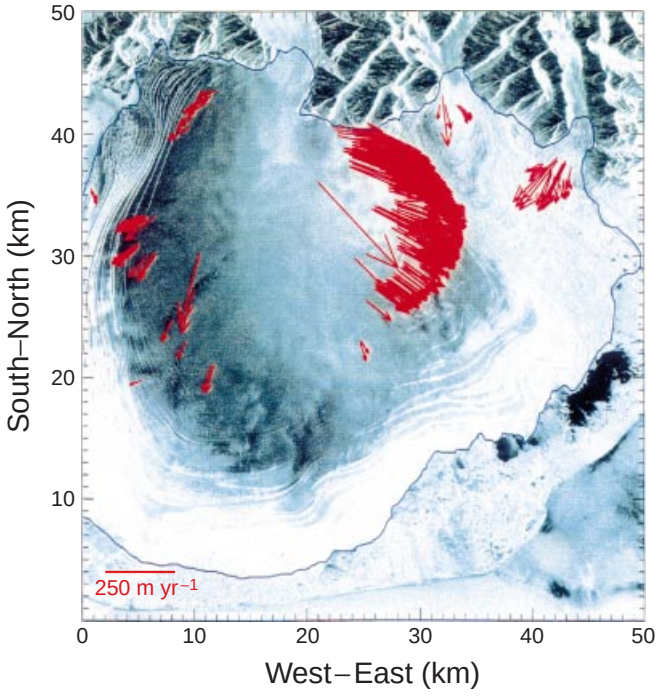

Figure 6 Surface ice velocities on Malaspina Glacier, derived from cross-correlation of SAR images acquired on June 29, 1992 and June 14, 1993. The direction of each vector is in the direction of ice movement; vector length is proportional to the magnitude of the velocity, in accordance with the scale at the lower left. The background is the SAR image of June 29, 1992. (Reproduced by permission from Lingle and Voronina.[61])

"point picking" method, indicating that this new approach is a practical and cost-effective way to measure glacier surface velocities.

Interferometry provides a methodology that greatly extends the accuracy of ice motion tracking. Massonnet and Feigl[30] provided a special summary of the application of satellite radar interferometry (SRI) for tracking ice motion of glaciers and ice sheets. Interested readers can also refer to the article in this encyclopedia by Massonnet **Elevation Modeling and Displacement Mapping using Radar Interferometry**.

4.5 Use of Backscatter to Map Lake Ice and Lake Depth

SAR backscatter images can also be used to study shallow tundra lakes.[63] The North Slope of Alaska is a relatively flat and featureless region of tundra with thousands of lakes, located between the Brooks Range and the Arctic Ocean. Most of the lakes are shallow, with few exceeding 2–4 m deep, and apart from a short open-water season from June to August they are ice covered. During the summer the lakes provide important habitats for plants and animals. In winter those lakes that do not freeze completely to the bottom can continue to support aquatic life and serve as sources of fresh water, a material that is in short supply in an otherwise frozen landscape. Ice core analysis shows that ice on all lakes there includes a clear layer overlying a layer with tubular bubbles oriented parallel to the direction of ice growth. The clear ice may also be overlain by a discontinuous layer of bubbly snow ice. Sequential ERS-1 SAR images acquired in NW Alaska, winter 1991–92, show that backscatter from lakes is low (−16 to −22 dB) at the time of initial ice formation because the specular nature of the upper and lower ice surfaces cause the radar pulse to be reflected away from the radar. As the ice thickens during the autumn, backscatter rises rapidly. The radar backscatter of floating ice can quickly reach the maximum values of the order of −6 to −7 dB. Once the ice freezes to the lake bottom, low backscatter values of −17 to −18 dB are observed. Model analysis[63] indicates that the strong forward scattering caused by tubular bubbles and the reflection at the ice–water interface jointly result in a double-bounce mechanism, which is responsible for bright appearance of floating lake ice on radar images. Using a sequential SAR images to track when and where the lake ice becomes grounded, Kozlenko and Jeffries[64] were able to depict the bathymetry of the tundra lakes. The results are of vital importance to studies of tundra lake ecosystems and monitoring of the impact of climate changes in the region.

4.6 Use of Visible Images and Synthetic Aperture Radar Interferometry to Map River Ice

Aufeis deposits (also called overflow ice, stream icings or naleds) develop on many rivers in high latitudes and in alpine regions. They form when the water pressure in streams is great enough to force water through cracks in the ice cover. The resulting overflows refreeze into additional layers of ice on the surface. The increased hydrostatic pressure generally results from restricted water flow caused by ice growth in stream channels. Refreezing of the overflowing water occurs during periods of subfreezing air temperature; such conditions almost always exist in polar regions during winter. Calcium carbonate deposits on some icings indicate that sometimes water has flowed through calcareous bedrock.[65] Climate, hydrology, geology, permafrost and topography influence the occurrence of aufeis and fluctuations in its activity.[66] Deposits of aufeis form every year on many rivers on the North Slope of Alaska, and, through repeated episodes of overflow and freezing, they may reach thicknesses of more than 3 m. The presence of aufeis in a drainage basin tends to stabilize river discharge in the same way as a glacier does, by providing melt water during hot, dry periods. For example, the Sagavanirktok River has no glaciers in its drainage area, yet its flow remains strong all summer,

partly due to the melting of aufeis deposits which range in area between 10^2 and 10^3 km^2. Indeed, the melting aufeis contributes to the difference in hydrographs between the Sagavanirktok River, which has extensive aufeis deposits in its head water area, and the Kuparuk River, which does not; both rivers enter the Arctic Ocean within 20 miles of each other near Prudhoe Bay. In addition, areas susceptible to icing may present engineering problems for construction of buildings, highways and other structures. Therefore, it is important to know the distribution and volume of aufeis in a drainage basin.

It is possible to map aufeis using Landsat Multispectral Scanner (MSS) imagery.[67–70] The visible wavelength (MSS band 5) data can be used to map residual ice after the spring thaw and the NIR wavelength range (MSS band 7) is useful for mapping late-winter and spring overflows. In an extensive investigation of Alaskan aufeis deposits, using Landsat imagery, Dean[69] found that most icings on the North Slope occur east of the Sagavanirktok River, where many icings coincide with known springs.[71]

Landsat imagery has proved useful in locating aufeis deposits at the end of winter but, because of the lack of solar illumination, aufeis processes cannot be mapped with visible and NIR imagery during the winter when the accumulation is most active. On the other hand, after the aufeis deposits have formed in the fall and continue to grow during the winter, the surface changes are subtle.[72] This makes the use of remote sensing for investigating winter aufeis processes even more difficult. An effort was made to use SRI to map the extent of aufeis near the junction of the Ivishak and Echooka Rivers in the northern Brooks Range during the winter of 1994. Both are important tributaries of the Sagavanirktok River.

Li et al.[73] have used interferometric coherence to map aufeis. The interferometric coherence between two radar acquisitions was used to map the distribution and development of aufeis in the study area. Coherence is a measure of the degree of resemblance of radar phases between two SAR images acquired a few days apart. If accretion occurs in the time between passes, the coherence between phases from two acquisitions over the aufeis is low relative to the surroundings. A time series of 12 interferograms were produced to cover the time period between January and March 1994.

On all interferograms, lakes and aufeis deposits along the rivers all show significant irregular changes in radar returns, as indicated by low coherence values. In the northern part of the coherence map (Figure 7), many lakes, ranging in size from a few hundred meters to 1.5 km, all appear dark. To the south of the junction of the Ivishak and the Echooka rivers, a narrow band along the Ivishak River is identified as an area of active aufeis development based on its proximity to the river channels

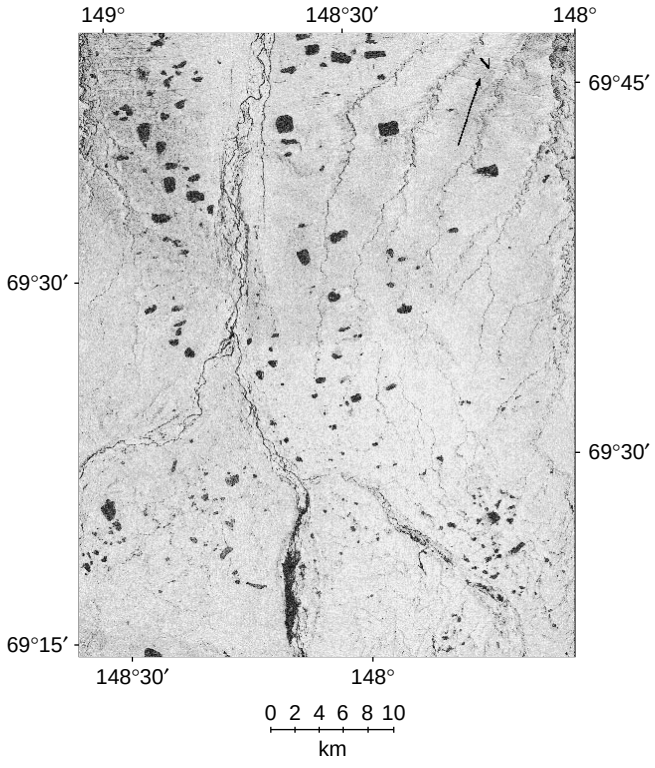

Figure 7 Coherence map of the interferogram generated from the ASF complex SAR images acquired on January 8 and 11, 1994. River ice and lake ice are characterized by low coherence (dark appearance) on the image. (Adapted by permission from Li, Benson, Shapiro and Dean.[73])

and its dark appearance on the coherence map. Similar patterns are also seen on all other interferograms in the time series.

Since the time series of interferograms started in January, the lakes were clearly frozen at that time. It is also certain that these lakes were not affected by overflows. The low coherence value of the lakes is caused by the lowering of the ice–water interface as the ice thickens. This boundary is the major scattering interface, so lowering it changes the scattering characteristics with respect to radar propagation. This increases the randomness in the phase relations between the two passes and leads to a low coherence value. For the aufeis areas, the accretion of new ice layers on top of the ice cover changes both the surface and the internal scattering pattern in a random manner, resulting in irregular changes of the phases over the extent of aufeis deposits. Consequently, the coherence between phases from the two passes over the aufeis is significantly lower than for all the other objects. Although both lakes and aufeis deposits appear dark in the coherence map, it is possible to distinguish one from the other, because (1) the locations of the lakes are known, (2) aufeis deposits always occur along river channels and (3) the backscatter

values of active aufeis always change, whereas the most significant lake ice backscatter changes occur primarily when the lake ice freezes to the bottom.[63] Owing to a lack of significant backscatter changes, most lakes disappear in the background, leaving aufeis as the only prominent object in the backscatter change maps, which are one of the by-products of SAR interferometric processing.

5 REMOTE SENSING OF SOIL MOISTURE

5.1 Use of Microwave Brightness Temperature for Soil Moisture

The microwave emission from soil is strongly dependent on the near-surface soil moisture content. Water has a dielectric constant of about 80 at the L-band compared with 3.5 for dry soil, and it may occupy 0–59% of the total soil volume, depending on external conditions and soil type. Consequently, the effective bare-soil emissivity may range about 0.6 for wet soil to 0.9 for dry soil. Competing effects from variations in soil temperature, surface roughness, vegetation cover and soil texture reduce the overall sensitivity of microwave brightness temperature (radiobrightness) to soil moisture. Moisture remains a dominant effect, however, as long as the vegetation cover is moderate.

The application of microwave brightness temperature in other regions provides useful information to researchers on polar regions. Using data sets acquired by an airborne electronically scanned thinned array radiometer (ESTAR) at Washita watershed in southwest Oklahoma in June 1992, Jackson et al.[74] derived soil moisture over a large area. The watershed was saturated with a great deal of standing water at the outset of the study. During the experiment of nine contiguous days there was no rainfall and the surface soil moisture exhibited a dry-down pattern over the period. Significant variations in the level and rate of change in surface soil moisture were noted over areas dominated by different soil textures. ESTAR images acquired on eight days over the nine-day experiment period show spatial patterns that are clearly associated with soil textures and temporal patterns of drainage and evaporation processes. Because ESTAR is a synthetic aperture passive microwave radiometer and it was flown at an altitude of 2200 m or lower, the resulting data sets have a grid size of 200 m and a swath of 3.2 km. The images were formed into a mosaic to cover larger areas, and the finer grid cells of the original images were integrated into larger cells of 9×9 km, which is close to what satellite passive microwave radiometers can achieve. The results show that a strong relation is retained between the passive microwave brightness temperature and spatially averaged soil moisture.

While the soil moisture in the top 2–5 cm layer can be directly estimated through a single snapshot by a passive microwave radiometer, Galantowics et al.[75] have demonstrated theoretically that sequential data assimilation through Kalman filter optimal estimation can retrieve the profile of soil moisture and temperature from periodic passive microwave observations. In their analysis, a 1-m soil state profile was recovered over an eight-day period from daily L-band observations following an intentionally poor initial state estimate.

In addition, multiple frequency passive microwave measurements in the 6–18 GHz range have been used to retrieve multiple parameters of land surface, including soil moisture, vegetation water content and surface roughness.[76] This approach has been tested using data from the Nimbus-7 SMMR for the years 1982–85 over African Sahel. The approach discriminates well between soil moisture, vegetation and temperature variations.

5.2 Use of Single Band Synthetic Aperture Radar Backscatter

SAR backscatter is used to estimate soil moisture. Because the radar return is strongly related to soil moisture content,[77–79] ERS-1 SAR can be used to map soil moisture conditions. Li et al.[80] generated a SAR mosaic for the entire state of Alaska using ERS-1 SAR images. The mosaic is particularly useful for this application. On the mosaic, three gigantic bands with each having backscatter values distinctively different from those in the other bands are seen between the Brooks Range and the Arctic coast (Figure 8).

The first band of high backscatter values along the Arctic coast stretches about 500 km between the delta of the Canning River in the east and the Icy Cape in the west. This band has a width ranging from 50 to more than

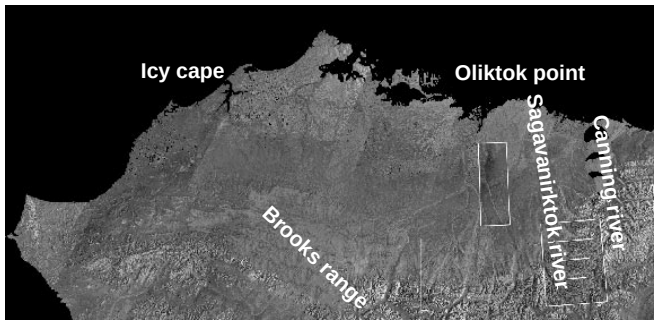

Figure 8 SAR mosaic of the North Slope of Alaska. The non-terrain corrected areas are enclosed in the thin-framed white boxes. On the mosaic, three gigantic bands with each having backscatter values distinctively different from those in the other bands are seen between the Brooks Range and the Arctic coast. (Reproduced by permission from Li, Guritz, Logan, Shindle, Grooves, Carsey, MacMahon and Olmsted.[80])

100 km and is characterized by numerous arctic lakes and flat terrain. Geographically, the band coincides with the Arctic coastal plain. The existence of numerous lakes exerts a strong influence on the high backscatter level of the band. On the mosaic, most lakes look very bright. This agrees with other investigators' observations[81] that arctic lakes usually have high backscatter during short summer. The high radar backscatter from lakes in summer is caused by roughening of the lake surface due to strong winds. In fact, weather records have documented that surface wind speeds along the Arctic coast are persistent and strong compared with those in more interior regions. It is of particular interest that the areas between the lakes also contribute significantly to the high backscatter values of this band. For example, near the Oliktok Point, to the west of the Sagavanirktok River, the coastal plain appears bright in the mosaic. Close examination indicates that land between lakes has a higher backscatter value than lakes in the area and the overall high backscatter level in the vicinity is mainly caused by backscatter from land. A previous geological survey of the area[82] reported that the ground cover of this area is wet tundra where the peaty soil has a shallow, active layer and is saturated throughout the summer.

To the south of the first band lies the second band of distinct radar backscatter values. This band of relatively low backscatter values is about 50 km wide on average. This band is the Arctic foothills characterized by rolling uplands of moist tundra.

Further south is the third band of the Brooks Range, with distinct mountainous radar backscatter characteristics, i.e. very bright foreslopes and dark backslopes.

The bright appearance of wet tundra and the relatively dark signature of moist tundra agree with the model and laboratory experimental results regarding the proportional relation between soil moisture and radar backscatter values for bare soils.[77] The low biomass in tundra makes the application of the bare soil moisture model relatively straightforward.

5.3 Use of Dual-polarization Imaging Radars

In the previous section, it was discussed that radar backscatter increases with increase in soil moisture. However, soil surface roughness also plays an important role in determining radar backscatter. For such a problem with more than one unknown, multiple polarized backscatter would provide a better solution. In practice, dual copolarized backscatter (HH and VV), instead of full polarimetric SAR, is used for the estimation of soil moisture. Backscatter obtained by imaging radars with longer wavelength, such as the L-band, is particularly useful in this application because of the deeper penetration into the soil. Omitting the usually weaker cross-polarized (HV or VH polarization) returns makes the estimation less sensitive to system noise and the presence of vegetation, and simplifies the calibration process. Empirical algorithms have been developed for inverting soil moisture content and surface root mean square (RMS) height from imaging radar images. Based on scatterometer data, Dubois et al.[83] derived empirical equations for L-band dual copolarized backscatters as functions of two unknowns: RMS and dielectric constant. Because the two unknowns can be solved from measurements of copolarized backscatters from two independent equations, and the dielectric constant of soil can lead to soil moisture, estimation of soil moisture is feasible. The algorithm is tested with data sets acquired with the AIRSAR system and used to derived soil moisture images using spaceborne SIR-C SAR data. A comparison with in situ data acquired at Chickasha, OK, shows that the algorithm infers soil moisture with an accuracy of 4.2%. The best results are achieved when the surface roughness is around 10 cm RMS for the L-band, the incidence angle is $>30^{\circ}$ and the volumetric soil moisture content is $<35\%$. Shi et al.[84] developed an algorithm based on a fit of a more theoretical model, the single scattering integral equation method. The algorithm is applied to basically the same L-band copolarized SAR image data. The RMS error of soil moisture was found to be 3.4%.

6 REMOTE SENSING OF ARCTIC VEGETATION AND FOREST FIRE

The world's boreal forest region is an active component of the Arctic and sub-Arctic regions. The total boreal forest covers a little less than 10% of the total land surface. Recent surveys indicate that 714 Gt of carbon are stored in the vegetation and ground layers of the ecosystem found in the boreal forest region, which represents >37% of the total terrestrial carbon pool. Boreal forest can be divided into two basic classes: those dominated by black spruce and those dominated by white spruce, based on their relation with microclimate and climate change. In Alaska, the white spruce forests are typically found on sites with warmer and better drained soil conditions, while black spruce forests are found on sites with cooler and more poorly drained soils. These local conditions are the primary factors behind the patterns of above- and below-ground carbon accumulation. In mature white spruce forests, the carbon levels are 70–90 t ha^{-1} in the living and dead vegetation and 20–40 t ha^{-1} in the ground layer. The cooler growing conditions in mature black spruce forest leads to lower carbon accumulation in the vegetation layer (20–30 t ha^{-1}), but higher levels of carbon in the ground layer because of lower rates of decomposition

(100–150 t ha^{-1}). Thus, on average, there is more carbon stored in black spruce than in white spruce forest. In a global warming scenario, a shift from black spruce to white spruce is to be expected.

It is well known that the boreal forest in the Upper Yukon and Koyukuk regions is especially prone to destruction by fire. Many trees have branches to the ground and commonly are covered with lichens and mosses. Shrub undergrowth abounds. Relatively low precipitation, long hours of sunshine and very high air temperatures accompanied by low relative humidity in summer greatly increase the fire hazard in an already flammable region where mosses and lichens are extremely dry. Once caused by either lightning or human accident, fire spreads readily in such an environment.[85] Because the forest fires in the region are widespread over large inaccessible areas, it is essential to map the extent of forest fires, especially the extent of the most recent ones, using satellite imagery. Recent studies[86] have shown that old fire boundaries (up to 40–50 years old) can be mapped in the boreal forest zone using Landsat TM data. This is possible because of the sensitivity of the shortwave IR bands of TM to patterns of regrowth in these forest. SAR imagery is also a useful tool for the identification of forest fire scars in the Arctic region. On the Alaska SAR mosaic,[80] several conspicuous bright patches have been found in the Upper Yukon and Koyukuk regions (Figure 9). These patches range from 10 000 to 500 000 acres in size. A comparison of the mosaic with the forest fire records collected at the US Bureau of Land Management shows that all those bright patches are recent forest fire scars. Table 1 lists the five largest bright patches in the region. Of these five patches, four are scars of 1991 forest fires and the remaining one is a 1990 fire scar. Also, the sizes of the fire scars estimated from the SAR mosaic compare with the fire records fairly well. Considering that the component SAR images were acquired in 1992, the ability of the mosaic to map the recent boreal forest fire scars is very encouraging.

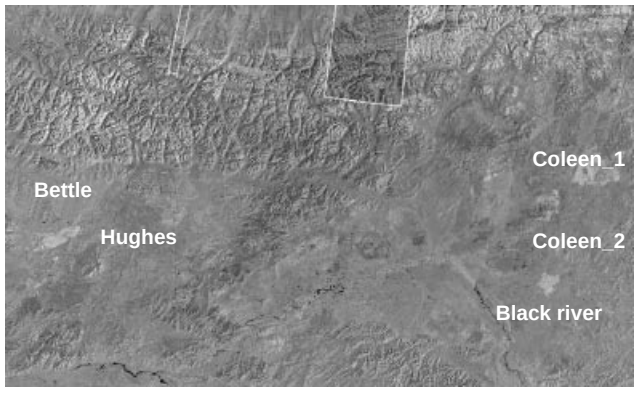

Figure 9 SAR mosaic of the Upper Yukon Region of Alaska. Forest fire scars can be easily identified because of their distinct radar signatures. (Reproduced by permission from Li, Guritz, Logan, Shindle, Grooves, Carsey, MacMahon and Olmsted.[80])

According to Kasischke et al.[87] and French et al.,[88] recent boreal forest fire scars can be identified on ERS-1 SAR images based on the difference in their radar backscatter from the surrounding areas. The difference in image intensities between burned and unburned areas is mainly caused by exposure of soil in the burned areas. Many boreal forest scars have brighter SAR signatures because severe forest fires burn the forest canopy completely and expose bare soil to radar illumination. Furthermore, severe fires consume the surface peat layer which insulates the ground and helps lower the average ground temperature. Elimination of this insulating layer causes a 5–7 °C increase in average ground temperature for at least several years after the fire, which in turn causes a melting of the underlying permafrost and an increase in soil moisture. As mentioned in section 5, a higher moisture content in bare soil leads to stronger radar

Table 1 Comparison of forest fire scars identified in Alaska SAR mosaic and recorded in forest fire files

Name of map quadrangle	Coleen	Coleen	Black river	Hughes	Bettle
Location	Rabbit Mtn–Lake Creek–Spike Mtn	Graphite Lake–Coleen Mtn	NE of Little Black River, W of Black River	Zane Hills–Koyukuk River	Koyukuk River–Alatna Hills
Center: lat./long.	67°30′N/142°W	67°08′N/143°15′W	66°25′N/143°W	66°15′N/154°50′W	66°40′N/152°15′W
Estimated size	35 × 40 km	25 × 30 km	20 × 25 km	20 × 30 km	25 × 25 km
SAR signature	Very bright	Assorted dark and bright	Bright	Assorted dark and bright	Bright
Fire record No.	A143(1990)	A144 (1988) B562 (1991)	B460 (1991)	B541 (1991)	B569 (1991)
Fire duration	6/29–9/14	6/28–10/1 6/30–9/27	6/26–9/16	6/29–9/30	6/30–9/30
Fire size in record (acres)	464 545	144 000 123 930	80 900	184 520	249 784

backscatter in general. This explains why many recent forest scars appear brighter on SAR images. On the other hand, the burned areas may have low soil moisture due to better drainage, resulting in lower SAR backscatter values. Also, regrowth of vegetation on fire-disturbed areas can either increase or decrease the image intensities by masking the effect of soil moisture on radar backscatter.

It is interesting that variation of SAR backscatter within individual patches may even give a clue to local forest fire history or the variation in damage caused by the fire. For example, the patch Coleen_2 between Graphite Lake and the Coleen Mountain, centered at 67°08′N and 143°15′W, has an assorted dark and bright appearance in image intensities. Further examination of the fire records shows that the dark middle part of the patch also experienced an earlier fire between June 28 and October 1, 1988. To the north and south of this part, the areas with only the damage of recent fire (1991) show a much brighter radar signature. Another example is the patch Coleen_1 between Rabbit Mountain and Spike Mountain, with its central location at 67°30′N and 142°W. This patch experienced fire between June 24 and September 14, 1990. Detailed fire records show that this fire scar is actually related to five smaller fires in the vicinity. In other words, this area was not uniformly burned. This explains the variation of SAR signatures within this large patch.

ACKNOWLEDGMENTS

Preparation of this article was partially supported by NASA grant NAG5-8614 and CRREL grant DACA89-99-K-0001. The author is very grateful to Dr J. Dozier of the University of Calfornia, Santa Barbara, Dr Kenneth Jezek of Ohio State University and his colleague, Dr Craig Lingle, for allowing to use of their figures in this article.

ABBREVIATIONS AND ACRONYMS

ASI	Italian Space Agency
AVHRR	Advanced Very High-resolution Radiometer
BRDF	Bidirectional Reflectance Distribution Function
DARA	German Space Agency
DEM	Digital Elevation Model
DTC	Decision Tree Classifier
EOS	Earth Observing System
ERS	Earth Resources Satellite
ESA	European Space Agency
ESMR	Electronically Scanning Microwave Radiometer
ESTAR	Electronically Scanned Thinned Array Radiometer
GCM	General Circulation Model
HH	Horizontal Transmit, Horizontal Receive
HV	Horizontal Transmit, Vertical Receive
IR	Infrared
MIR	Mid-infrared
MISR	Multi-angle Imaging Spectroradiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Landsat Multispectral Scanner
NASA	National Aeronautics and Space Administration
NIR	Near-infrared
NOAA	National Oceanic and Atmospheric Administration
RAMP	Radarsat Antarctic Mapping Project
RMS	Root Mean Square
SAR	Synthetic Aperture Radar
SIR-C/X-SAR	Shuttle Imaging Radar-C and X-band Synthetic Aperture Radar
SMMR	Scanning Multi-channel Microwave Radiometer
SPOT	Système Probatoire Pour l'Observation de la Terre
SRI	Satellite Radar Interferometry
TIR	Thermal Infrared
TM	Thematic Mapper
USGS	US Geological Survey
VH	Vertical Transmit, Horizontal Receive
VV	Vertical Transmit, Vertical Receive

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Elevation Modeling and Displacement Mapping using Radar Interferometry • Satellite and Sensor Systems for Environmental Monitoring • Sea Ice Monitoring by Remote Sensing

REFERENCES

1. W.D. Sellers, 'A Global Climate Feedback Based on the Energy Balance of the Earth–Atmosphere System', *J. Appl. Meteorol.*, **8**, 392–400 (1969).
2. K.P. Shine, A. Henderson-Sellers, 'The Sensitivity of Thermodynamic Sea Ice Model to Changes in Surface

Albedo Parameterization', *J. Geophys. Res.*, **90**(D1), 2243–2250 (1985).
3. P.J. Sellers, 'Remote Sensing of the Land Surface for Studies of Global Change', International Satellite Land Surface Climatology Project Report, Columbia, MD, 1993.
4. J. Croll, *Climate and Time in Geologic Relations, a Theory of Secular Change of the Earth's Climate*, Isbister, London, 1875.
5. W.M. Washington, G.A. Meehl, 'General Circulation Model CO_2 Sensitivity Experiments: Snow–Sea Ice Albedo Parameterizations and Globally Averaged Surface Air Temperature', *Clim. Change*, **8**, 231–241 (1986).
6. R.E. Dickinson, G.A. Meehl, W.M. Washington, 'Ice Albedo Feedback in a CO_2-doubling Simulation', *Clim. Change*, **10**, 241–248 (1987).
7. D. Rind, R. Healy, C. Parkinson, D. Martinson, 'The Role of Sea Ice in $2 \times CO_2$ Climate Model Sensitivity, I: The Total Influence of Sea Ice Thickness and Extent', *J. Clim.*, **8**, 449–463 (1995).
8. J. Dozier, *Remote Sensing of Snow in Visible and Near-infrared Wavelengths, Theory and Applications of Optical Remote Sensing*, John Wiley & Sons, New York, 527–547, 1989.
9. W.J. Wiscombe, S.G. Warren, 'A Model for the Spectral Albedo of Snow, I, Pure Snow', *J. Atmos. Sci.*, **37**, 2712–2733 (1980).
10. S.G. Warren, W.J. Wiscombe, 'A Model for the Spectral Albedo of Snow, II, Snow Containing Atmospheric Aerosols', *J. Atmos. Sci.*, **37**, 2734–2745 (1980).
11. S.C. Colbeck, 'Grain Clusters in Wet Snow', *J. Colloid Interface Sci.*, **72**, 371–384 (1979).
12. S.C. Colbeck, 'Classification of Seasonal Snow Cover Crystals', *Water Resour. Res.*, **22**, 59S–70S (1986).
13. T.C. Grenfell, D.K. Perovich, J.A. Ogren, 'Spectral Albedos of an Alpine Snowpack', *Cold Reg. Sci. Technol.*, **4**, 121–127 (1981).
14. C.F. Bohren, 'Applicability of Effective-medium Theories to Problems of Scattering and Absorption by Nonhomogeneous Atmospheric Particles', *J. Atmos. Sci.*, **43**, 468–475 (1986).
15. J. Dozier, 'Snow Reflectance from Landsat-4 Thematic Mapper', *IEEE Trans. Geosci. Remote Sens.*, **GE-22**, 323–328 (1984).
16. J. Dozier, D. Marks, 'Snow Mapping and Classification from Landsat Thematic Mapper Data', *Ann. Glaciol.*, **9**, 97–103 (1987).
17. G.A. Riggs, D.K. Hall, S.A. Ackerman, 'Sea Ice Extent and Classification Mapping with the Moderate Resolution Imaging Spectroradiometer Airborne Simulator', *Remote Sens. Environ.*, **68**, 152–163 (1999).
18. M. Matson, 'NOAA Satellite Snow Cover Data', *Palaeogeogr. Palaeoecol.*, **90**, 213–218 (1991).
19. R. Kahn, D. Wenkert, 'The Earth Observing System', in *Encyclopedia of Planetary Sciences*, eds. J.H. Shirley, R.W. Fairbridge, Chapman and Hall, London, 211–215, 1997.
20. V.V. Salomonson, W.L. Barnes, P.W. Maymon, H.E. Montgomery, H. Ostrow, 'MODIS: Advanced Facility Instrument for Studies of the Earth as a System', *IEEE Trans. Geosci. Remote Sens.*, **27**(2), 145–153 (1989).
21. C. Justice, D. Hall, V. Salomonson, J. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Knjazihhin, S. Running, R. Nemani, E. Vermote, J. Townshend, R. Defries, D. Roy, Z. Wan, A. Huete, W. van Leeuwen, R. Wolfe, L. Giglio, J.-P. Muller, P. Lewis, M. Barnsley, 'The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research', *IEEE Trans. Geosci. Remote Sens.*, **36**(4), 1228–1249 (1998).
22. D.J. Dinner, J.C. Beckert, T.H. Reilly, C.J. Bruegge, J.E. Conel, R. Kahn, J.V. Martonchik, T.P. Ackerman, R. Davies, S.A.W. Gerstl, H.R. Gordon, J.-P. Muller, R. Myneni, R.J. Sellers, B. Pinty, M.M. Verstraete, 'Multiangle Imaging Spectroradiometer (MISR) Description and Experiment Overview', *IEEE Trans. Geosci. Remote Sens.*, **36**, 1072–1087 (1998).
23. D.K. Hall, G.A. Riggs, V.V. Salomonson, 'Development of Methods for Mapping Global Snow Cover Using Moderate Resolution Imaging Spectrometer Data', *Remote Sens. Environ.*, **54**, 127–140 (1995).
24. D.K. Hall, J.L. Foster, V.V. Salomonson, A.G. Klain, J.Y.L. Chien, 'Error Analysis for Global Snow-cover Mapping in the EOS Era', Presented at IGARSS'98, 1998.
25. J. Stroeve, A. Nolin, K. Steffen, 'Comparison of AVHRR-derived and In Situ Surface Albedo over the Greenland Ice Sheet', *Remote Sens. Environ.*, **62**(3), 262–276 (1997).
26. J.C. Shi, T. Painter, J. Dozier, 'Estimation of Snow Surface Albedo Using Landsat Thematic Mapper', in *Proceedings of IGARSS'96*, IEEE No. 96CH35875, IEEE, Lincoln, NE, 2249–2251, 1996
27. A.H. Strahler, W. Lucht, C.B. Schaaf, T. Tsang, F. Gao, X. Li, J.-P. Muller, P. Lewis, M.J. Barnsley, 'Algorithm Theoretical Basis Document, MODIS Product ID: MOD43 (MODIS BRDF/albedo), Version 5.0', NASA MODIS Team, Washington, DC, April, 1999.
28. D. Cline, R. Bales, J. Dozier, 'Estimating the Spatial Distribution of Snow in Mountain Basins Using Remote Sensing and Energy Balance Modeling', *Water Resour. Res.*, **34**(5), 1275–1285 (1998).
29. S. Li, M. Sturm, S. Kyriazis, J. Chen, T. Logan, 'A Multisensor Approach for Mapping the Change of Snow Covered Area at Ivotuk, Alaska, During the 1999 Melt Season', Presented at the 50th Arctic Science Conference, 19–22 September 1999, Denali National Park.
30. D. Massonnet, K.L. Feigl, 'Radar Interferometry and Its Application to Changes in the Earth's Surface', *Rev. Geophys.*, **36**(4), 441–500 (1998).
31. S. Li, M. Sturm, C. Benson, 'Pattern of Redistribution of Wind-drifted Snow at Imanvait Creek on the North Slope,

Alaska, During a Storm in February, 1994, Revealed by ERS-1 SAR Interferometry', in *PIERS96 on Polar Remote Sensing*, ed. J.A. Kong, Institüt für Meteorologie und Geophysik, Innsbruck, Austria, July 8–12, 50, 1996.
32. H. Rott, C. Matzler, D. Strohl, S. Bruzzi, K.B. Lenhart, 'Study of SAR Land Applications for Snow and Glacier Monitoring', Technical Report 6618/85/F/FL(SC), European Space Agency, Paris, 1988.
33. J. Shi, J. Dozier, 'Measurement of Snow- and Glacier-covered Areas with Single Polarization SAR', *Ann. Glaciol.*, **17**, 72–76 (1993).
34. J.J. Van Zyl, H.A. Zebker, C. Elachi, 'Image Radar Polarization Signature: Theory and Observations', *Radio Sci.*, **22**(4), 529–543 (1987).
35. J. Shi, J. Dozier, 'Mapping Seasonal Snow with SIR-C/X-SAR in Mountainous Areas', *Remote Sens. Environ.*, **59**, 294–307 (1997).
36. U. Wegmuller, C. Matzler, E. Schanda, 'Microwave Signatures of Bare Soil', *Adv. Space Res.*, **9**(1), 307–316 (1989).
37. U. Wegmuller, 'The Effect of Freezing and Thawing on the Microwave Signatures of Bare Soil', *Remote Sens. Environ.*, **33**, 123–135 (1990).
38. B. Zuerndorfer, A.W. England, F.T. Ulaby, 'An Optional Approach to Mapping Freezing Terrain with SMMR Data', in *Remote Sensing Science for the Nineties*, ed. R. Mills, IGARSS '90, the University of Maryland, College Park, Maryland, May 20–24, 1990, IEEE, Washington, DC, 1153–1156, 1990.
39. C. Dobson, F.T. Ulaby, B. Zuerndorfer, A.W. England, 'Mesoscale Monitoring of the Soil Microwave Radiometry', Final Report to Orbiting Satellite Project, NASA/Goddard Space Flight Center, NASA, Grant NAG5-852, 30, 1990.
40. B. Zuerndorfer, A.W. England, 'Radiobrightness Decision Criteria for Freeze and Thaw Boundaries', *IEEE Trans. Geosci. Remote Sens.*, **30**(1), 89–102 (1992).
41. E. Rignot, J.B. Way, 'Monitoring Freezing–Thaw Cycles along North–South Alaskan Transects Using ERS-1 SAR', *Remote Sens. Environ.*, **49**, 131–137 (1994).
42. K. Boehnke, V. Wismann, 'Thawing of Soils in Siberia Observed by the ERS-1 Scatterometer Between 1992 and 1995', in *Proceedings of IGARSS*, IEEE, Lincoln, NE, 2264–2266, 1996.
43. K. Boehnke, V. Wismann, 'Thawing Processes During Siberian Spring Observed by ERS Scatterometer and SAR', in *Proceedings of IGARSS*, ed. T.I. Stein, IEEE, Singapore, 1826–1828, 1997.
44. Z. Wang, S. Li, 'Detection of Winter Frost Heaving of the Active Layer of Arctic Permafrost Using SAR Differential Interferograms', in *Proceedings of IEEE 1999 International Geoscience and Remote Sensing Symposium 1999 (IGARSS'99)*, ed. T.I. Stein, IEEE No. 99CH36293, 1946–1948, IEEE, Hamburg, 1999.
45. P.J. Williams, M.W. Smith, *The Frozen Earth*, Cambridge University Press, Cambridge, Chapter 5, 1989.
46. W.C. Oechel, S.J. Hastings, G. Vourlitis, M. Jenkins, G. Riechers, N. Grulke, 'Recent Change of Arctic Tundra Ecosystems from a Net Carbon Sink to a Source', *Nature (London)*, **361**, 520–523 (1993).
47. F.E. Nelson, A.H. Lachenbruch, M.K. Woo, E.A. Koster, T.E. Osterkamp, M.K. Gavrilova, G.D. Cheng, 'Permafrost and Change Climate', in *Proceedings of the Sixth International Conference on Permafrost*, South China University of Technology Press, Wushan, 987–1005, Vol. II, 1993.
48. M.D. Walker, D.A. Walker, K.R. Everett, 'Wetland Soils and Vegetation, Arctic Foothills, Alaska', *US Fish Wildl. Serv. Biol. Rep.*, **89**(7), (1989).
49. M. Fahnestock, R. Bindschadler, R. Kwok, K. Jezek, 'Greenland Ice Sheet Surface Properties and Ice Dynamics from ERS-1 SAR Imagery', *Science*, **262**, 1525–1530 (1993).
50. H.J. Zwally, A.C. Brenner, J.A. Major, R.A. Bindschadler, J.G. March, 'Growth of Greenland Ice Sheet: Measurement', *Science*, **246**, 1587–1589 (1989).
51. J.L. Bamber, 'A Digital Elevation Model of the Antarctic Ice Sheet Derived from ERS-1 Altimeter Data and Comparison with Terrestrial Measurements', *Ann. Glaciol.*, **20**, 48–64 (1994).
52. K.F. Noltimier, K.C. Jezek, H.G. Sohn, B. Li, H. Liu, F. Baumgartner, V. Kaupp, J.C. Curlander, B. Wilson, R. Onstott, 'RADARSAT Antarctic Mapping Project – Mosaic Construction', in *Proceedings of IEEE 1999, International Geoscience and Remote Sensing Symposium 1999 (IGARSS'99)*, ed. T.I. Stein, IEEE No. 99CH36293, IEEE, Hamburg, 2349–2351, 1999.
53. C.H. Davis, C.A. Kluever, B.J. Haines, 'Elevation Changes of the Southern Greenland Ice Sheet', *Science*, **279**, 2086–2088 (1998).
54. C.S. Lingle, D.N. Covey, 'Elevation Changes on the East Antarctic Ice Sheet, 1978–93, from Satellite Radar Altimetry: A Preliminary Assessment', *Ann. Glaciol.*, **27**, 7–18 (1998).
55. H.J. Zwally, 'Microwave Emissivity and Accumulation Rate of Polar Firn', *J. Glaciol.*, **18**, 195–215 (1977).
56. H.J. Zwally, M.B. Giovinetto, 'Accumulation in Antarctica and Greenland Derived from Passive Microwave Data: A Comparison with Contoured Compilations', *Ann. Glaciol.*, **21**, 123–130 (1995).
57. D.G. Vaughan, J.L. Bamber, M. Giovinetto, J. Russell, A.P.R. Cooper, 'Reassessment of Net Surface Mass Balance in Antarctica', *J. Clim.*, **12**, 933–946 (1999).
58. T.A. Scambos, M.J. Dutkiewwicz, J.C. Wilson, R.A. Bindschadler, 'Application of Images Cross-correlation to the Measurement of Glacier Velocity Using Satellite Image', *Remote Sens. Environ.*, **42**, 177–186 (1992).
59. R.A. Bindschadler, P.L. Vornberger, D.D. Blankenship, T.A. Scambos, R. Jacobs, 'Surface Velocity and Mass Balance of Ice Streams D and E, West Antarctica', *J. Glaciol.*, **42**, 461–475 (1996).

60. B.K. Lucchitta, C.F. Rosanova, K.F. Mullins, 'Velocity of Pine Island Glacier, West Antarctica, from ERS-1 SAR Images', *Ann. Glaciol.*, **21**, 277–283 (1995).
61. C.S. Lingle, V.A. Voronina, 'Ice Velocity and the Origin of Folded Medial Moraines on Melaspina Galcier, Alaska, from Satellite Synthetic Aperture Radar Imagery', Project Report to National Park Service, Wrangle–St. Elias National Park, National Natural Landmark (NNL) Program, 1997.
62. E.H. Isaaks, R.M. Srivastava, *Applied Geostatistics*, Oxford University Press, New York, 1989.
63. M.O. Jeffries, K. Morris, W.F. Weeks, H. Wakabayashi, 'Structural and Stratigraphic Features and ERS-1 Synthetic Aperture Radar Backscatter Characteristics of Ice Growing on Shallow Lakes in NW Alaska, Winter 1991–1992', *J. Geophys. Res.*, **99**(C11), 22 459–22 471 (1994).
64. N. Kozlenko, M.O. Jeffries, 'Bathymetric Mapping of Shallow Water in Thaw Lakes on the North Slope of Alaska with Spaceborne Imaging Radar', *Arctic*, in press.
65. D.K. Hall, 'Mineral Precipitation in North Slope River Icings', *Arctic*, **33**(2), 343–348 (1980).
66. K.L. Carey, 'Icing Developed from Surface and Groundwater', US Army Cold Regions Research and Engineering Laboratory Monograph, No.111-D3, US Army Corps of Engineerings, Hanover, NH, 1–67, 1973.
67. C.E. Sloan, C. Zenone, L.R. Mayo, 'Icing Along the Trans-Alaska Pipeline Route', *US Geol. Surv. Prof. Pap.*, No. 979 (1976).
68. D.K. Hall, C. Roswell, 'The Origin of Water Feeding Icings on the Eastern North Slope of Alaska', *Polar Rec.*, **20**(128), 433–438 (1981).
69. K.G. Dean, 'Stream-icing Zones in Alaska', Report of Investigations, State of Alaska, Department of Natural Resources, Division of Geological and Geophysics Surveys, No 84–16, State of Alaska, Dept. of Natural Resources, Fairbanks, AK, 1–20, 1984.
70. R.S. Williams, Jr, 'Glaciers and Glacial Landforms', *Geomorphology from Space, a Global Overview of Regional Landforms*, eds. N. Short, R.W. Blair, Jr, National Aeronautics and Space Administration, Washington, DC, Special Publication 486, 521–596, 1986.
71. J.M. Childers, C.E. Sloan, J.P. Mecke, J.W. Nauman, 'Hydrologic Reconnaissance of the Eastern North Slope, Alaska', US Geological Survey Open-file Report, No. 77–492, US Geological Survey, Denver, CO, 1–65, 1977.
72. E. de K. Leffingwell, 'The Canning River Region Northern Alaska', *US Geol. Surv. Prof. Pap.*, No. 109, US Government Printing Office, Washington, DC, 1919.
73. S. Li, C. Benson, L. Shapiro, K. Dean, 'Aufeis in the Ivishak River, Alaska, Mapped From Satellite Radar Interferometry', *Remote Sens. Environ.*, **60**, 131–139 (1997).
74. T.J. Jackson, D.M. Le Vine, C.T. Swift, T.J. Schmuggege, F.R. Schiebe, 'Large Area Mapping of Soil Moisture Using the Ester Passive Microwave Radiometer in Washita'92', *Remote Sens. Environ.*, **53**, 27–37 (1995).
75. J.F. Galantowics, D. Entekhabi, E.G. Njoku, 'Test of Sequential Data Assimilation for Retrieving Profile Soil Moisture and Temperature from Observed L-band Radiobrightness', *IEEE Trans. Geosci. Remote Sens.*, **37**(4), 1860–1870 (1999).
76. E.G. Njoku, L. Li, 'Retrieval of Land Surface Parameters Using Passive Microwave Measurements at 6–18 GHz', *IEEE Trans. Geosci. Remote Sens.*, **37**(1), 79–93 (1999).
77. M.C. Dobson, F.T. Ulaby, 'Active Microwave Soil Moisture Research', *IEEE Trans. Geosci. Remote Sens.*, **GE-24**, 23–36 (1986).
78. A. Beaudoin, T.L. Toan, Q.H.J. Gwyn, 'SAR Observations and Modeling of the C-band Backscatter Variability Due to Multiscale Geometry and Soil Moisture', *IEEE Trans. Geosci. Remote Sens.*, **28**, 886–894 (1990).
79. O. Yisok, K. Sarabandi, F.T. Ulaby, 'An Empirical Model and an Inversion Technique for Radar Scattering from Bare Soil Surfaces', *IEEE Trans. Geosci. Remote Sens.*, **31**, 370–381 (1992).
80. S. Li, R. Guritz, T. Logan, M. Shindle, J. Grooves, F. Carsey, J. MacMahon, C. Olmsted, 'Summer Environmental Mapping Potential of a Large-scale ERS-1 SAR Mosaic of the State of Alaska', *Int. J. Remote Sens.*, **20**(2), 387–401 (1999).
81. K. Morris, M.O. Jeffries, W.F. Weeks, 'Ice Processes and Growth History on Arctic and Sub-arctic Lakes Using ERS-1 SAR Data', *Polar Rec.*, **31**(177), 115–128 (1995).
82. J. Brown, R.A. Kreig, (eds.), *Guidebook to Permafrost and Related Features Along the Elliott and Dalton Highways, Fox to Prudhoe Bay, Alaska, Fourth International Conference on Permafrost, July 18–22, 1983*, University of Alaska, Fairbanks, Alaska, Division of Geological and Geophysical Surveys, Department of Natural Resources, State of Alaska, Fairbanks, 1–230, 1983.
83. P.C. Dubois, J. van Zyl, T. Engman, 'Measuring Soil Moisture with Imaging Radars', *IEEE Trans. Geosci. Remote Sens.*, **33**(4), 915–926 (1995).
84. J. Shi, J. Wang, A.Y. Hsu, P.E. O'Neill, E.T. Engman, 'Estimation of Bare Surface Soil Moisture and Surface Roughness Parameter Using L-band SAR Image Data', *IEEE Trans. Geosci. Remote Sens.*, **35**(5), 1254–1266 (1997).
85. H.J. Lutz, 'Fire as an Ecological Factor in the Boreal Forest of Alaska', *J. For.*, **58**, 454–460 (1960).

86. H. Epp, R. Lanoville, 'Satellite Data and Geographic Information Systems for Fire and Resource Management in the Canadian Arctic', *Geocarta Int.*, **11**, 97–103 (1996).
87. E.S. Kasischke, L.L. Bourgeau-Cahvez, N.H.F. French, 'Observations of Variations in ERS-1 SAR Image Intensity Associated with Forest Fires in Alaska', *IEEE Trans. Geosci. Remote Sens.*, **32**(1), 206–210 (1994).
88. N.H.F. French, E.S. Kasischeke, L.L. Bourgeau-Chavez, P.A. Harrell, 'Sensitivity of ERS-1 SAR to Variations in Soil Water in Fire-disturbed Boreal Forest Ecosystems', *Int. J. Remote Sens.*, **17**(15), 3037–3053 (1996).

Processing and Classification of Satellite Images

Graeme G. Wilkinson
Kingston University, Kingston Upon Thames, UK

1 Introduction **2**
1.1 Remote Sensing, Major Applications, and Need for Image Processing 2
1.2 Spectral Characteristics of Natural Surfaces 2
2 Recent Satellites **2**
2.1 Optical/Infrared Systems 2
2.2 Radar Satellites 3
3 Image Capture Systems and Trends **5**
3.1 Imaging Geometries 5
3.2 Spectral and Spatial Characteristics of Data and Data Volumes 6
3.3 Orbits and Temporal Repeat Cycles 6
3.4 Imaging Spectrometers 7
3.5 Very High Spatial Resolution Imaging 7
4 Early Stages of Image Processing and Enhancement **7**
4.1 Geometric Correction and Geocoding 7
4.2 Noise Removal 7
4.3 Spectral Feature Enhancement 8
4.4 Principal Components Analysis and Spectral Data Reduction Techniques 8
4.5 Normalized Vegetation Indices 8
5 Image Classification **9**
5.1 Classification as a Signal Inversion Problem 9
5.2 Choice of Spectral Bands, Multitemporal Analysis, and Data Fusion 9
5.3 Per-pixel Approaches and Segmentation 9
5.4 Unsupervised Data Clustering 9
5.5 Self-organizing Feature Maps for Data Clustering 10
5.6 Supervised Classification 10
5.7 Nonparametric Methods 11
5.8 Parametric Approaches 11
5.9 Statistical Classification and the Maximum Likelihood Method 11
5.10 Neural Network Classification: Multilayer Perceptrons 11
5.11 Combined Classification Approaches and Soft Classification 12
5.12 Alternatives to Classification: Spectral Mixture Modeling 12
6 Classification Refinement **13**
6.1 Spatial Smoothing and Rule-based Postprocessing 13
6.2 Structural Scene Models and Use of Textural Analysis 13
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

Earth-observing satellites primarily gather information about the environment as digital images acquired by multispectral imaging sensors. These images consist of large arrays of picture elements (pixels). Images are usually recorded in 3–7 spectral bands, though new generation sensors providing 18–36 spectral bands are in development. The spectral bands are located in the visible, infrared, millimeter, and microwave parts of the electromagnetic spectrum. The spatial resolutions of images at the subsatellite point at ground level (dictated by the sensor's instantaneous field of view, IFOV) are principally in the range 10 m–1 km depending on the satellite. Very-high-resolution satellite-borne sensors providing imagery with resolution as high as 1 m will soon become routinely available. Images are primarily used for mapping Earth resources and monitoring the state of the environment (terrestrial, atmospheric, and oceanic). Pixels consist of digitized radiances which are usually quantized on a scale of 0–255. Interpretation of surface features requires the images to be processed and classified by computer. The processing involves several stages, typically: (1) geometrical rectification and adjustment to a suitable cartographic coordinate system, (2) correction and calibration for atmospheric or system effects on detected radiances, (3) classification or product generation. Classification involves the transformation of the detected multispectral radiances in the image into meaningful descriptors of the surface – usually as thematic classes which can be displayed as a map. The mathematical conversion process is usually based on training an appropriate classifier algorithm by use of example spectral radiances from known surface objects (ground truth data).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Classifier algorithms may either be nonparametric or parametric (e.g. based on a statistical model). Neural network algorithms have found useful application as classifiers for satellite image data. Products may also be generated by computing environmentally meaningful indices from spectral radiances such as the normalized difference vegetation index (NDVI).

1 INTRODUCTION

1.1 Remote Sensing, Major Applications, and Need for Image Processing

The science of satellite remote sensing, also frequently known as Earth observation, began in 1960 with the launch of the first civilian (nonmilitary) satellites designed for meteorological observation. The TIROS series (Television and Infrared Observation Satellite) became the first satellites to acquire images of synoptic weather patterns and in so doing provided an expedient tool in meteorological forecasting and climate modeling. This remained the principal application of remote sensing until 1972 when the first satellite was launched dedicated to the monitoring of the Earth's land surface using a multispectral imaging system. This satellite was known as the Earth Resources Technology Satellite (ERTS)-1 but was later renamed Landsat-1. The Landsat series has since proved to be one of the main sources of global environmental information ever conceived and still continues to provide valuable data with the satellites making a repeat coverage of the entire planet every 16–18 days. Program continuity over several decades has been of inestimable value in monitoring gradual or long-term changes in the biosphere and natural ecosystems (such as progressive soil degradation, desertification, deforestation, increasing urbanization, and loss of biodiversity).

The main advantage of satellite remote sensing over alternative forms of environmental data gathering is that large global surface areas can be monitored without the need for labour-intensive ground level surveys. In addition, satellite observations are also less costly than aerial surveys per unit of Earth surface area in the context of long-term/large-area mapping and monitoring. It is estimated that by the year 2000 the total volume of data gathered by Earth observing satellites will be of the order of 10 Tbytes per week, mostly consisting of digital images. These will be derived from a variety of sensors operating in different parts of the electromagnetic spectrum on-board a collection of satellites built and launched by many different countries, governmental agencies, and commercial enterprises.[1] One of the main reasons for the expansion of data gathering capability has been the growing concern over environmental changes. In the United States this resulted in the development of the Earth Observing System (EOS) which is the cornerstone of the Earth Science Enterprise (ESE) program of the National Aeronautics and Space Administration (NASA). The EOS encompasses a number of satellites each housing several different sensing systems each aimed at assessing specific hydrologic, biogeochemical, atmospheric, ecological, and geophysical processes.[2]

Images from remote sensing satellites need to be processed by computer in order to gain maximum utility from them. Images record the radiance, reflectance, or backscatter of electromagnetic radiation from terrestrial surfaces. Image processing is thus principally concerned with the identification of particular terrestrial features or measurement of environmental variables through analysis of the spectral observations encoded in the digital multispectral images recorded by the satellite sensor.

1.2 Spectral Characteristics of Natural Surfaces

The reflective or emissive properties of terrestrial surfaces differ significantly. This provides the key to satellite observation of the Earth. Figure 1 shows typical reflectance spectra of common terrestrial surfaces through the visible and near-infrared part of the spectrum. Such reflectance curves should be taken as being indicative only, since precise spectra of particular surfaces observed may vary significantly depending on the condition or composition. Water has a very low reflectance throughout the visible and infrared region. Wet soils therefore generally have lower reflectance than dry soils. Vegetation has high reflectance in the 800–1200 nm near-infrared region. The reflectance of diseased or damaged vegetation is usually reduced markedly in this region. The spectral bands of satellite sensors are strategically located at the sensor design stage to facilitate easy identification of surface types by analysis of the signature detected across all the bands. The bands must also be located at wavelengths at which the Earth's atmosphere is highly transparent. The atmosphere is relatively highly transparent at visible wavelengths, though absorption bands of water vapor, carbon dioxide, oxygen, ozone, and trace gases become significant in the infrared region. The presence of dust and aerosols can affect the transmission of electromagnetic radiation throughout the visible region.

2 RECENT SATELLITES

2.1 Optical/Infrared Systems

Table 1 lists the principal remote sensing satellites which have been operational during the 1990s and also includes those which are due to be launched in the late 1990s for

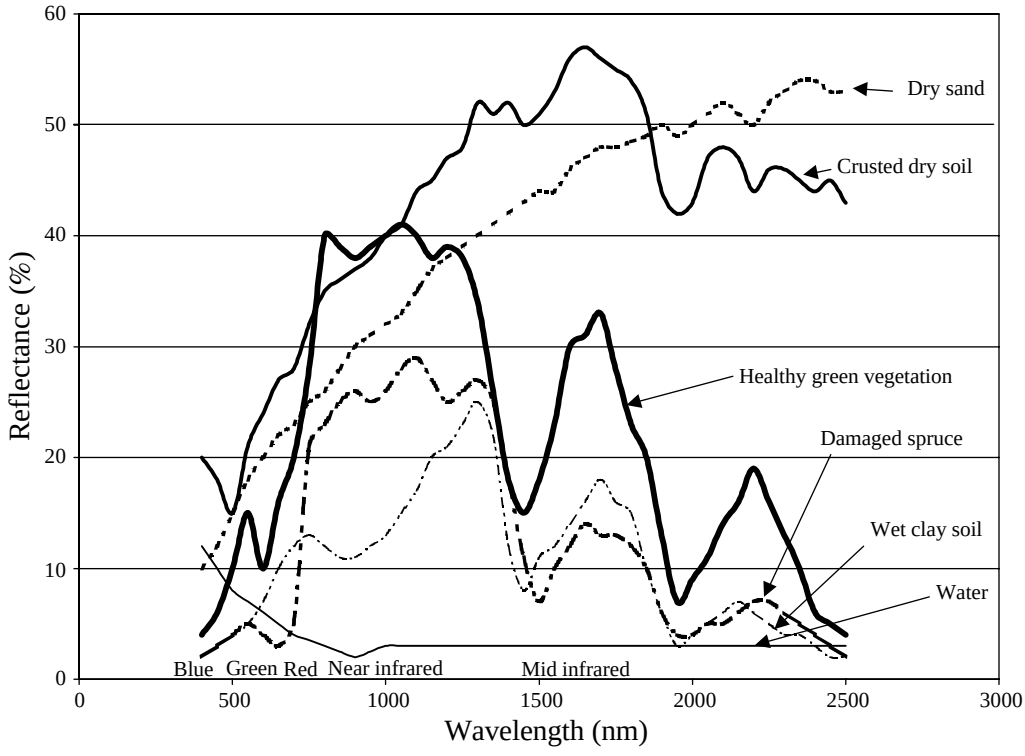

Figure 1 Approximate spectral reflectance curves of common natural surfaces.

service beyond the year 2000. More detailed information about Earth observation satellites is given by Kramer.[1]

The main satellites providing multispectral optical and infrared imagery can be divided into three groups:

- those providing relatively low-to-medium spatial resolution imagery (primarily for meteorological applications or global environmental change monitoring, such as the National Oceanic and Atmospheric Administration (NOAA) series, and the more recent ENVISAT and TERRA satellites, all giving imagery with spatial resolutions in the 250 m–1.1 km range);
- those providing high spatial resolution imagery (primarily for mapping and monitoring applications, such as the Systeme Probatoire de l'observation de la Terre (SPOT) and Landsat series with spatial resolutions in the 10–30 m range);
- those providing very high spatial resolution (primarily for detailed mapping applications, such as the later additions to the Indian remote sensing satellite (IRS) series and new generation satellites such as QuickBird with spatial resolutions in the 1–5 m range).

In addition, the SPOT-4 satellite is notable in combining sensors with high and low spatial resolution. The more advanced low-to-medium resolution sensors such as the medium resolution imaging spectrometer (MERIS) on the ENVISAT satellite and the moderate resolution imaging spectrometer (MODIS) on the TERRA satellite have much higher numbers of spectral bands in comparison to earlier sensors.

In the optical and infrared region of the spectrum, all of the satellites mentioned above rely on the use of passive radiometers which measure either reflected solar radiation or blackbody emission from terrestrial surfaces. Currently, however, ''active'' laser-based systems are also under development for measurement of vegetation canopy height. The first remote sensing satellite carrying an operational sensor of this kind will be the Vegetation Canopy Lidar (VCL) mission which forms part of NASA's Earth System Science Pathfinder Project (ESSP).

2.2 Radar Satellites

Table 1 also lists the principal remote sensing satellites which provide radar imagery in the C and L bands. These include ERS-1, ERS-2, JERS-1, Radarsat and ENVISAT. These all provide high-resolution ground images at microwave (centimeter) wavelengths by using the signal integration technique of ''aperture synthesization''. This technique enables the sensor to generate very-high-resolution radar images by effectively simulating the existence of a very large baseline antenna on the satellite.

Table 1 Principal remote sensing satellites[a]

Satellite	Main imaging sensor	Spectral bands of sensor (wavelength) (nm)	Pixel ground resolution (m)	Date of launch/ expected launch
SPOT-3	HRV	*Multispectral*:		1993
		500–590	20	
		610–680	20	
		790–890	20	
		Panchromatic:		
		510–730	10	
SPOT-4	HRVIR	*Multispectral*:		1998
		500–590	20	
		610–680	20	
		790–890	20	
		1580–1750	20	
	VEGETATION	430–470	1150	
		610–680	1150	
		780–890	1150	
		1580–1750	1150	
Landsat-5	TM	*Multispectral*:		1984
		450–520	30	
		520–600	30	
		630–690	30	
		760–900	30	
		1550–1750	30	
		2080–2350	30	
		10 400–12 500	120	
NOAA-14	AVHRR	580–680	1100	1994
		725–1000	1100	
		1580–1640	1100	
		3550–3930	1100	
		10 300–11 300	1100	
		11 500–12 500	1100	
ERS-1	SAR	Radar C-band 5.66 cm	30	1991
ERS-2	SAR	Radar C-band 5.66 cm	30	1995
JERS-1	SAR	Radar L-band: 23.5 cm	18	1992
Radarsat	SAR	Radar C-band: 5.6 cm	9 (high resolution mode)	1996
IRS-1C	LISS	*Multispectral*:	23.5	1995
		520–590	23.5	
		620–680	23.5	
		770–860	23.5	
		1550–1750	70.8	
		Panchromatic:		
		500–750	5	
Quickbird	QBM	*Multispectral*:	4	4th quarter 1998
		450–520	4	
		530–590	4	
		630–690	4	
		770–900	4	
	QBP	*Panchromatic*:		
		450–900	1	

Table 1 (*continued*)

Satellite	Main imaging sensor	Spectral bands of sensor (wavelength) (nm)	Pixel ground resolution (m)	Date of launch/ expected launch
Landsat 7	ETM+	*Multispectral*:		1998
		450–520	30	
		520–600	30	
		630–690	30	
		760–900	30	
		1550–1750	30	
		2080–2350	30	
		10 400–12 500	60	
		Panchromatic:		
		500–900	15	
ENVISAT	MERIS	15 bands – programmable in range 400–1050	300	1999
	ASAR	Radar C-band with 5 possible polarization modes	30	
TERRA	MODIS	36 in total:	Various:	1998
		21 bands in range 400–3000	250 500	
		15 bands in range 3000–14500	1000	
VCL	VCL	Pulsed lasers operating at 1064	25	2000

[a] In most cases only the main imaging sensor is indicated although some of these satellites carry multiple sensing systems. This table is not exhaustive. HRV, high resolution visible; HRVIR, high resolution visible and infrared; TM, thematic mapper; AVHRR, advanced very high resolution radiometer; LISS, linear imaging self-scanning sensor; QBM, QuickBird Multispectral; QBP, QuickBird Panchromatic; ETM+, enhanced thematic mapper+.

The synthetic aperture radar (SAR) images so obtained are extremely valuable as an adjunct to HRV or infrared imagery obtained from other satellites. A comprehensive treatise on the SAR technique is provided by Curlander and McDonough.[3] The advanced SAR (ASAR) system under development for the ENVISAT satellite provides radar images with multiple polarizations.

3 IMAGE CAPTURE SYSTEMS AND TRENDS

3.1 Imaging Geometries

Remote sensing satellites use a combination of their orbital motion and a mechanical or electronic system to build up images covering the ground beneath the satellite as it moves in orbit. Some satellites also have the capability to view locations off-nadir (and can be programmed to do so according to ''customer'' requirements). The most common type of imaging geometry now used is the ''push broom'' geometry (Figure 2) in which the satellite captures complete lines of pixels simultaneously as it flies. This procedure enables the satellite to gather a sequence of one-dimensional pixel arrays which, when stacked as rows, form a two-dimensional image.

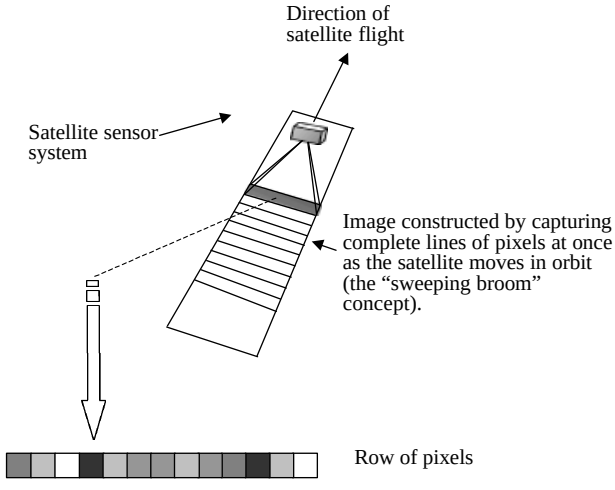

Figure 2 Satellite imaging scanning geometry: the ''push broom'' mode.

Normally the sensor system will gather information simultaneously in each of its spectral channels using appropriate optical devices (e.g. beam splitters) so that each pixel will consist of a set of reflectance measurements representing the spectral characteristics of each area viewed.

3.2 Spectral and Spatial Characteristics of Data and Data Volumes

As indicated in Table 1, a wide variety of data of differing spectral and spatial characteristics are gathered from Earth observing satellites. Sensors are generally intended for particular applications and are designed to have the appropriate combination of spectral and spatial characteristics. These characteristics include:

- spatial resolution or pixel size (which may vary across the captured image)
- swath width of the Earth ''scene'' which is imaged
- number of spectral channels
- location and bandwidth of spectral channels
- quantization accuracy of spectral measurements.

Although it would be ideal to maximize spatial resolution, scene width, number of channels, and quantization accuracy, this can lead to an unmanageable quantity of image data for practical purposes. Excessive data volumes lead to problems in data transmission, storage, and analysis. Sensor designers therefore decide on an appropriate trade-off between the imaging characteristics according to the primary intended application. Gathered data volume is also affected by the repeat cycle of the imaging system.

The VEGETATION sensor flown on board the SPOT-4 satellite is a notable example of an instrument designed specifically for one primary application: the monitoring of global vegetation. It has a spatial resolution of 1.15 km, a swath width of 2200 km, four spectral channels, and high precision 10-bit quantization (i.e. 1024 signal levels in each channel) leading to a sensor data rate of 500 kbit s^{-1}. The locations of the spectral channels are optimized for the detection of chlorophyll and to remove atmospheric influences on the detected ground radiances.

3.3 Orbits and Temporal Repeat Cycles

To investigate dynamic environmental phenomena, remote sensing satellites must be capable of generating images of the same location on Earth on a regular repeat basis. For meteorological applications there needs to be a repeat cycle of the order of a few hours at most. For many land applications, repeat cycles of a few days or even weeks suffice.

Most remote sensing satellites fly in one of two main types of orbit:

- geostationary orbit in which the satellite constantly stays over the same point on the Earth's surface (this type of orbit is used for some meteorological satellites such as the Geostationary Operational Environmental Satellite (GOES) series – the orbital altitude must be approximately 36 000 km)
- sun synchronous polar orbits in which the satellite orbits approximately from one pole to the other in a constant ''vertical'' plane whilst the Earth rotates beneath it (typical orbital altitudes of the order of 800–1500 km are common).

Satellites in geostationary orbit view only one side of the globe from a constant (and relatively distant position). They are thus able to generate repeat images by simply rescanning the globe as fast as the optical and electronic systems permit. Satellites in lower polar orbits view only narrow strips of the Earth as they fly (usually in push broom mode) and can only generate repeat imagery when their orbital track brings them back over the same point

Table 2 Orbital altitudes and repeat cycles of some remote sensing satellites[a]

Satellite	Orbital type	Orbital altitude (km)	Repeat cycle
GOES	Geostationary	35 770	30 min
SPOT-4	Sun-synchronous, polar	830	26 days
ERS-2	Sun-synchronous, polar	780	3 days (reference orbit), [35 days for full SAR imaging coverage]
Radarsat	Sun-synchronous, circular	798	24 days
Landsat-7	Sun-synchronous, polar	705	16 days
ENVISAT	Sun-synchronous, circular	800	35 days [3 days for full MERIS mapping repeat coverage]

[a] Repeat cycles can be reduced for satellites which permit programmed off-nadir viewing, e.g. the SPOT series.

on the Earth's surface. In most cases the orbit parameters are designed so that a fixed repeat cycle of a few days or weeks is established. Examples of satellite orbital characteristics and temporal repeat cycles are given in Table 2.

3.4 Imaging Spectrometers

Imaging spectrometers are designed to capture hyperspectral imagery, that is images with a large number of usually narrow closely spaced spectral channels. Such sensors usually provide from at least 15 up to approximately 250 spectral channels. The first satellite-borne imaging spectrometers will be the MODIS sensor to be carried on TERRA and the MERIS sensor to be carried on ENVISAT (Table 1). Imaging spectrometer systems have also been developed for airborne remote sensing such as the GER-63 (Geophysical Environmental Research Corporation 63 channel spectrometer) and the 224 channel AVIRIS (Airborne Visible and InfraRed Imaging Spectrometer) developed by the NASA JPL (Jet Propulsion Laboratory). Imaging spectrometers are of most interest for applications where precise discrimination of surface features is required, such as in ecological or geological studies. Hyperspectral imagery can be regarded as forming a data cube since the multiple spectral measurements effectively add a third dimension to an image pixel array. Imaging spectrometers have the potential to generate very high data volumes. In order to keep data rates within currently acceptable bounds, satellite-borne systems such as MODIS or MERIS will operate only at medium spatial resolution rather than at high spatial resolution.

3.5 Very High Spatial Resolution Imaging

Very high spatial resolution imaging usually refers to satellite images gathered with pixel sizes of 10×10 m or less. A number of civilian satellites capable of acquiring such images have been developed during the late 1990s, representing a new trend in satellite remote sensing. The origins of this technology lie with earlier military surveillance satellites. Very-high-resolution images are ideal for large-scale mapping (e.g. 1 : 10 000 scale) and for detection and analysis of artificial structures on the landscape.

Several very-high-resolution remote sensing satellite systems are in development by commercial organizations because of the economic potential of detailed land surface observations. Such systems will not gather data continuously because of the very high data rates that would be produced. Instead most very-high-resolution systems are designed to be programmable so that the sensor views a particular chosen ground location as it passes.

4 EARLY STAGES OF IMAGE PROCESSING AND ENHANCEMENT

4.1 Geometric Correction and Geocoding

Raw digital satellite images captured by satellites usually need to be corrected for geometrical distortions caused by sensor aberrations and orbital instability. They also usually need to be transformed to an appropriate map projection and cartographic grid (geocoded) so that they can be integrated with environmental data derived from alternative sources (such as digital maps). This involves transforming an image I with pixel array coordinates (x, y) into a new image I' with pixel coordinates (v, w). The transformation is achieved with a deformation mapping (M), Equation (1)

$$I'(v, w) = M[I(x, y)] \tag{1}$$

The transformed image must be generated at regular integer grid points (v, w). The transformation may involve a combination of scaling, rotation, and translation. The deformation mapping M can be computed exactly if the image distorting function is known. However in most practical cases it is determined by identifying ground control points (GCPs) in the image and in the reference grid and deducing a polynomial function which maps from the image to the reference grid and vice versa. Easily identifiable spatial features such as road intersections, or points on coastlines are usually used as GCPs.

4.2 Noise Removal

Satellite images usually contain noise as a result of the sensing and data transmission processes. This may manifest itself in images as pixel ''drop outs'' (where the radiance value of an individual pixel has been seriously corrupted). This can be corrected by modifying pixel values which anomalously differ from their neighborhood average. Noise is also manifested as image striping which can be corrected using special purpose algorithms. For some purposes, such as classification, noise can be removed by applying a low-pass filter to an image, which results in local smoothing of pixel radiance values. Filtering can be performed by any of a large number of neighborhood operators, such as the 3×3 Sobel Operator. Further information on such operators can be obtained from standard image processing and computer vision texts.[4,5] In some cases, it is desirable to use filters which preserve abrupt changes in image intensity due to edges of real spatial features on the landscape, such as the Nagao–Matsuyama filter for edge preserving smoothing.[6]

SAR images are subject to a special form of random multiplicative speckle noise arising from the coherent

radar imaging process. Individual pixel values by themselves cannot be taken to be indicative of the radar backscatter coefficient of a particular location on the Earth. Speckle filtering must be carried out to estimate the true radar backscatter coefficient of each pixel. A number of algorithms have been developed to remove speckle noise such as that described by Nezry and De Groot.[7] See also Curlander and McDonough[3] for a wider discussion of the SAR speckle phenomenon.

4.3 Spectral Feature Enhancement

Spectral measurements in satellite images are often confined to a narrow part of the overall dynamic range of the sensor's response. It is therefore often useful to expand the measurements to a wider dynamic range as a means of improving the discrimination of surface features. Simple computational techniques can be used to perform this, such as the linear contrast stretch (Figure 3) or the nonlinear histogram equalization technique.[5]

4.4 Principal Components Analysis and Spectral Data Reduction Techniques

It is often difficult for the information content of large numbers of spectral bands in satellite images to be fully comprehended by the user. Techniques have therefore been adopted to concentrate a large proportion of the spectral information into a smaller number of synthetic channels created by applying a geometrical transformation to the initial multidimensioned spectral space. One of the most common approaches is the Principal Components Transform (PCT) which creates a new set of orthogonal axes in spectral space with maximum variance. The first principal component image thus contains more information about the variability of the viewed landscape than any individual spectral channel alone. In very-high-dimensional spectral spaces, the technique of projection pursuit[8] may be used as a search procedure to determine an optimum transformation of the data. This has been applied to hyperspectral imagery.[9] Besides computing principal components, other linear combinations of spectral measurements may be defined to emphasize certain aspects of ground features. A good example is the tasseled cap transform. This can be used to generate synthetic spectral channels to emphasize brightness, greenness, and wetness of land surface features by applying linear weighting factors to the original spectral channels, for example of the Landsat TM sensor of Crist and Cicone.[10]

4.5 Normalized Vegetation Indices

Satellite images are frequently used to study vegetation and estimate surface biomass. The relative amount of green vegetation in a pixel can be approximately deduced by deriving a vegetation index from a combination of different spectral measurements. A commonly used index is the NDVI which is calculated in Equation (2)

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - R}{\mathrm{NIR} + R} \tag{2}$$

where NIR denotes reflected radiance in the near-infrared spectral channel and R denotes the reflected radiance in the red visible spectral channel. The computation of NDVI images can be used as a data reduction technique and a method of emphasizing differences in vegetation cover. The NDVI parameter emphasizes the difference between the near-infrared and red reflectance which has been shown to be a good indicator of live green vegetation with high leaf chlorophyll content. Healthy green vegetation has high near-infrared reflectance (reflection by chlorophyll) and low red reflectance (absorption by chlorophyll). NDVI computation is particularly useful in agricultural remote sensing. Empirical relationships between NDVI and crop yield can be used, for example, to predict end of season harvests.[11]

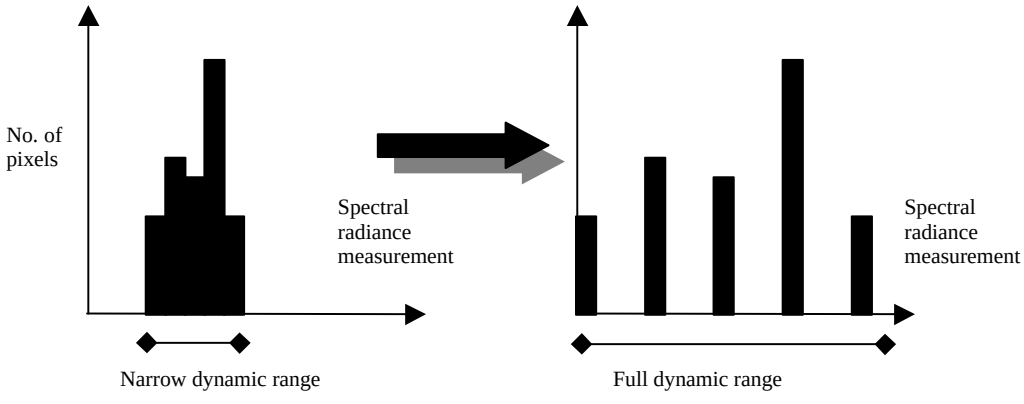

Figure 3 Simple spectral feature enhancement by contrast stretch.

5 IMAGE CLASSIFICATION

5.1 Classification as a Signal Inversion Problem

The classification of multispectral satellite images is one of the most commonly applied analysis techniques. Classification involves performing a transformation from the numerical spectral measurements into meaningful classes or labels which can describe a landscape. Classification effects a transformation from a physical measurement into a cartographic or thematic description of the surface, for example into terms such as forest, urban area, water surface, apple orchards, and so on. As such, classification can be viewed as a signal inversion process. Many different techniques exist for classification of satellite images, most based on statistical techniques or learning algorithms. These ''learn'' an association between spectral measurements and class labels from observations of known ground features, so-called ''ground truth'' information.

5.2 Choice of Spectral Bands, Multitemporal Analysis, and Data Fusion

Ideally, classification of individual pixels or groups of pixels should be performed using as much information as possible from as many sources as are available. In practice this normally means using all of the spectral channel measurements available from a given satellite sensor. However, classification can be improved significantly by using images taken on different dates (the multitemporal approach). Multitemporal images can particularly help to differentiate vegetation classes by virtue of the fact that different species have differences in their phenological cycles which are manifested in varying spectral response curves. For example, deciduous forests can be easily separated from coniferous forests by combined use of summer and winter imagery.

Besides using multitemporal imagery, classification can be improved by using images acquired in different parts of the electromagnetic spectrum by sensors on additional satellites. For example, the combination of optical and infrared imagery acquired by one satellite with radar imagery acquired by another can make a significant overall improvement in the accuracy of a landscape classification. This is an example of data fusion. The process of data fusion can also include the use of ancillary geographical data sets acquired from conventional ground level surveys. Usually the data fusion process is nontrivial because the multiple images or ancillary data sets may have been sampled at different spatial resolutions and stored in different cartographic projections, or as image frames in different orientations. This may necessitate a considerable level of additional image preprocessing. A good review of data fusion in remote sensing is provided by Pohl and Van Genderen.[12]

5.3 Per-pixel Approaches and Segmentation

The process of classification and the generation of class labels may be performed either for individual pixels (''per-pixel'' approaches) or for contiguous groups of pixels that constitute a meaningful spatial object to which a single class label can be applied. In order to do the latter, an image has to be initially divided into the separate spatial groupings of pixels, a process known as ''segmentation''.[13] Segmentation can be performed by detecting discontinuities in radiance in an image which delineate the borders of homogeneous land parcels (the edge detection approach). Alternatively it can be performed by finding groups of adjacent pixels which are spectrally similar and locally merging additional pixels to these groups, if they satisfy a spectral similarity constraint (the region growing approach). Both approaches are aimed at dividing an image into a mosaic of landscape parcels which are spectrally similar and therefore likely to define single objects or landscape types which can be classified and labelled. In addition, several hybrid approaches have been developed which integrate and take advantage of both the edge and the region methods.[14] The spectral radiance of an entire parcel is obtained by averaging the spectral measurements of each pixel from which it is composed. This averaging process removes the spectral variability of each individual pixel, thus in most cases making the classification of the parcel more accurate.

5.4 Unsupervised Data Clustering

The different spectral radiance values recorded by a typical multispectral satellite sensor can be regarded as independent measurements of properties of the Earth's surface. As such they create a multidimensional measurement or ''feature'' space in which all the pixels are scattered. The radiances of pixels corresponding to similar land cover types can normally be expected to be grouped in similar regions of this measurement space (Figure 4). It is therefore often useful to apply clustering algorithms to the data to determine the locations of the most important pixel groupings in the measurement space and so determine the dominant land cover classes in the image. This is usually done by an iterative procedure such as the k-nearest neighbor algorithm.[15] Once the dominant data clusters have been found, it is necessary to assign class labels to them in order to attach some physical meaning. In practice this may not be easy as the spectral groupings may not correspond to physical land cover classes in an obvious way. Unsupervised data clustering and subsequent assignment of labels are therefore rarely used as the sole means of classifying satellite image data. Normally ground truth information is used to guide the classification. This is the supervised

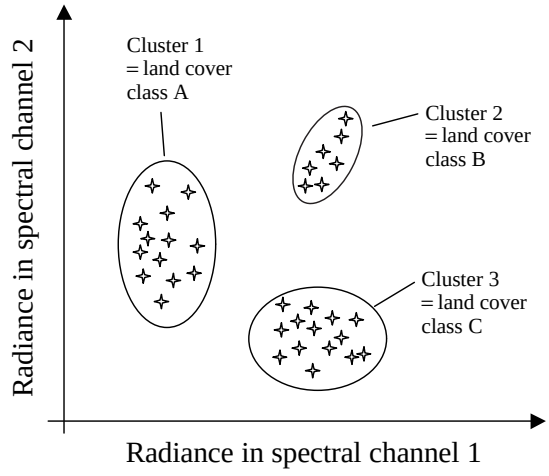

Figure 4 Pixel clusters in a two-dimensional spectral measurement space.

rather than the unsupervised approach. However the results of an initial unsupervised classification can be used to guide the selection of representative pixels for a subsequent supervised classification.

5.5 Self-organizing Feature Maps for Data Clustering

Recently, artificial ''neural network'' algorithms have also been used for unsupervised data clustering in remote sensing. The self-organizing map neural network algorithm developed by Kohonen[16] models data using a multidimensional array of competing ''neurons'' which each learn to represent a prototype cluster from a given data set. The array of neurons effectively becomes a map of the natural relationships between the patterns (spectral measurements) given to the network. Self-organizing networks have been found to be powerful tools for complex pattern recognition problems. In remote sensing they have been found to reveal interesting land surface features in complex data sets.[17] However their usefulness is not universally agreed upon as it has also been found that they demand excessive computation time in comparison with other methods for data clustering in the remote sensing context.[18]

5.6 Supervised Classification

Supervised classification is the approach most commonly adopted in remote sensing and one which is often central to the image analysis process since it concerns the direct transformation from measured radiances to thematic map. It is a vital step in the generation of a geographical product from the physical measurements encoded in the satellite images. A good description of supervised classification, together with some excellent practical examples, is provided by Thomas et al.[19]

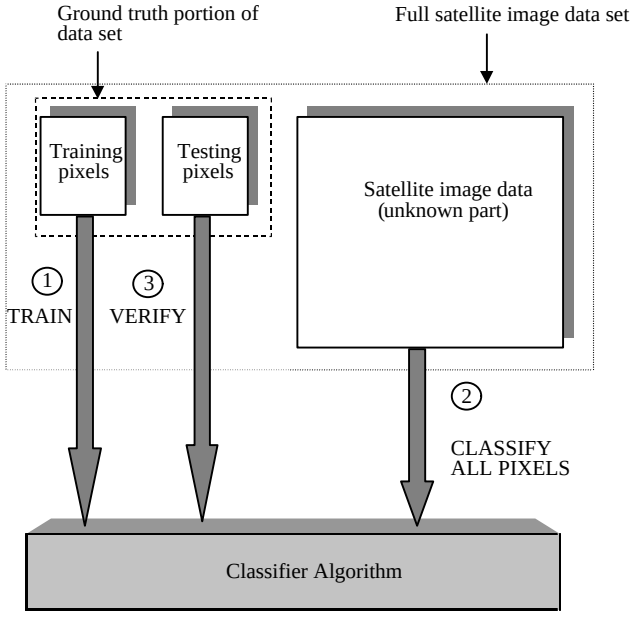

Figure 5 Principle of supervised classification.

The key element of supervised classification of satellite images is to make use of ground truth obtained from a prior field survey to create a reduced set of pixels selected from the image whose thematic classes are known. This reduced set of pixels can then be used to train a classifier algorithm. The objective of supervised classification is then to assign all the pixels in an image to thematic map classes (such as forest, urban area, etc.) by making use of this trained classifier algorithm (Figure 5). Ideally this should happen with a high degree of precision if an accurate map of the landscape is to be produced.

The accuracy of a classification is determined partly by the quality of the ground truth data and partly by how well the restricted set of ground truth pixels is representative of the full image (the ''generalization capability'' of the classifier). It is also determined by the spectral ''separability'' of the classes (it is often found that some classes are very similar, for example certain types of vegetation may have very close spectral response curves). In order to measure accuracy, it is common to use only part of the ground truth data for training the classifier algorithm and to use the remainder to test it, that is to see if it produces the classes it is supposed to with the test pixels provided. The results of such tests are normally described in the form of a two-dimensional confusion matrix[20] which shows how testing pixels from the different classes are distributed by the classifier between all the classes in use. Ideally testing pixels from class N should all be allocated by the classifier to that same class. In practice, it is rare for more than 80% of the pixels to be allocated correctly

because of the generalization and class separability problems.

5.7 Nonparametric Methods

There are many types of classifier algorithm in use in remote sensing and most commercial software packages used by satellite image analysts offer several alternatives. The objective of training a classifier is to define ''discrimination surfaces'' which divide the multidimensional spectral space of the image into hyper-regions corresponding to different thematic classes.

The simplest forms of classifier are those which rely on nonparametric methods in which there is no attempt to create a parameterized model of how the pixels of different classes are distributed. In the ''parallelepiped'' or ''box'' classifier, simple linear or planar discrimination surfaces are defined in spectral space to define the classes (Figure 6). In the ''minimum distance classifier'', class means are first identified in spectral space and then a pixel is simply allocated to the nearest class on the basis of Euclidean distance. This is equivalent to dividing the spectral space by use of a Voronoi tessellation[21] (Figure 7). Sometimes threshold distances from means are used in order to avoid classifying pixels which are spectrally too dissimilar from any of the available classes.

5.8 Parametric Approaches

As seen in Figures 6 and 7, nonparametric methods usually rely on simple planar discrimination surfaces to define a classifier's behavior. This is a relatively crude approach, especially when the pixels corresponding to a class are distributed in an irregular or nonsymmetric manner. Parametric approaches to classification make use of parameterized models of the classes in the spectral feature space. These are generally much more powerful than nonparametric methods and lead to higher overall classification accuracy.

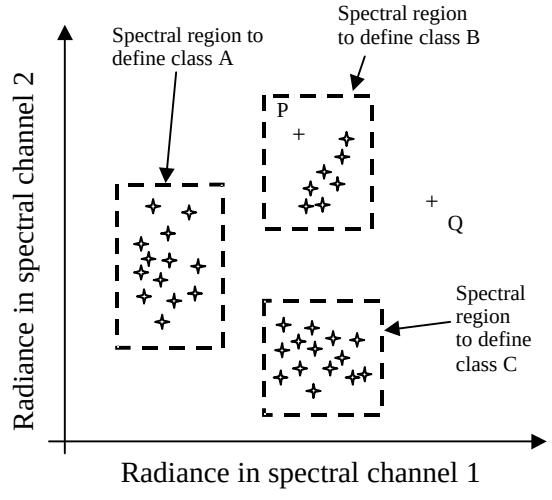

Figure 6 Parallelepiped or box classifier in a two-dimensional spectral space. Pixel P would be assigned to class B; pixel Q could not be assigned to a defined class.

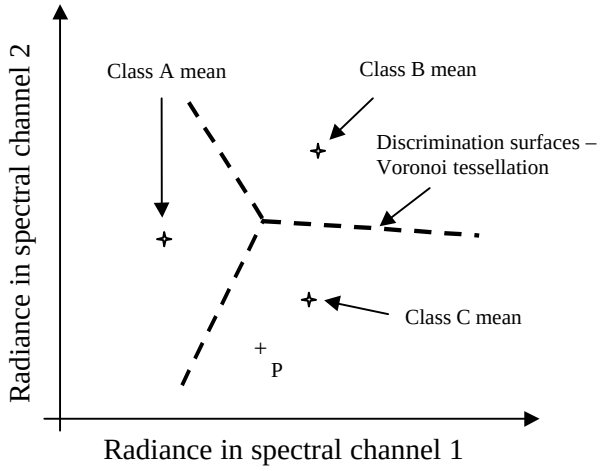

Figure 7 Minimum distance classifier in a two-dimensional spectral space. The dashed discrimination surfaces divide the space into regions which define the nearest class mean; pixel P would thus be assigned to class C.

5.9 Statistical Classification and the Maximum Likelihood Method

The most common parametric approach is to use a statistical method and to model classes according to the distributions of the training pixels. Most often classes are modeled by using the multivariate form of the normal probability density function. Pixels are then classified by assigning them to the class to which they have the highest statistical likelihood of belonging. This approach effectively results in classes being discriminated on the basis of ellipsoidal equiprobability surfaces in the spectral feature space (Figure 8).

5.10 Neural Network Classification: Multilayer Perceptrons

Since the late 1980s, supervised classification of satellite image data has also been performed by artificial neural network algorithms[22] which have rapidly gained in popularity for satellite image analysis.[23] In supervised classification, their layers of interconnected processing elements are trained to perform a mathematical transformation between spectral radiances and thematic map classes.

The most common type of neural network adopted in remote sensing for supervised classification is the multilayer perceptron (Figure 9) which is trained by

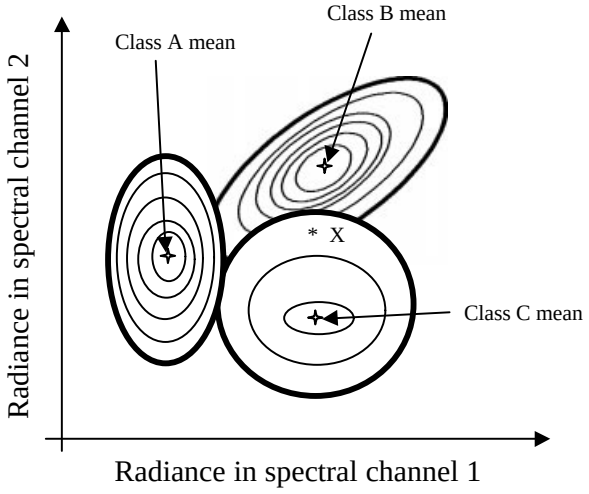

Figure 8 Maximum likelihood classification. The classifier models each class by a normal distribution which can be represented as concentric equiprobability contours in the spectral feature space. Pixel X has a higher probability of being in class C than class B even though in spectral distance terms it is closer to the mean of class B.

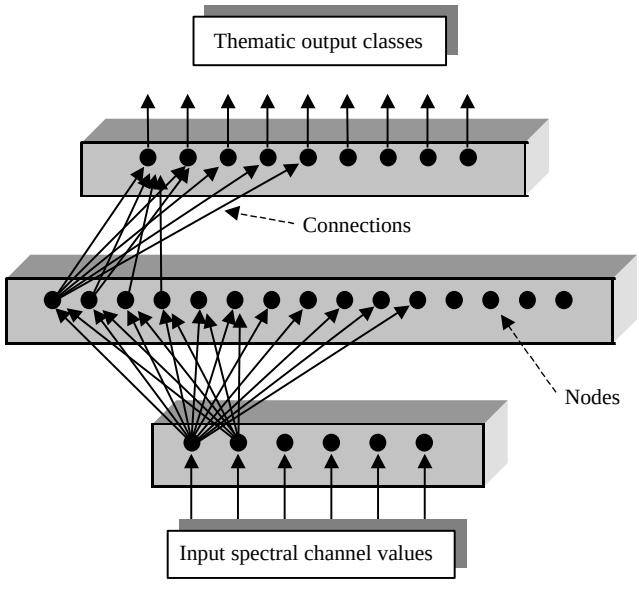

Figure 9 Multilayer perceptron neural network classifier. Such a network can be trained to encode a mathematical transformation between spectral values and their thematic class associations. The nodes in the network are linked by connections which carry weighting factors. Each node performs a nonlinear mathematical operation on its inputs to derive its own output which it passes on to other nodes. The network is shown partially connected for clarity. Usually such a network has connections between all nodes in successive layers.

the backpropagation algorithm. A good review of this technique can be found in Paola and Schowengerdt.[24] They have often been found to perform as well as statistical classifiers in terms of overall classification accuracy. Their popularity as a classification technique in remote sensing can be partly attributed to the fact that they do not make any initial assumptions about the form of the distribution of the satellite data and appear well suited to fused data sets comprising images from multiple sources. The main disadvantage of the technique is that training can be a long procedure sometimes requiring several hours of computer processor time, depending on the number of pixels, spectral channels, and classes, apart from the size of network selected.

5.11 Combined Classification Approaches and Soft Classification

Since statistical and neural network classification approaches have different strengths and weaknesses, some attempts have been made to integrate them, for example by Wilkinson et al.[25] The essence of this integrated approach is to use both statistical and neural classifiers in parallel and then to reclassify any pixels for which they disagree. The reclassification is performed by a second neural network which is trained specifically to deal with ''difficult'' pixels (usually outlier pixels on the edge of the main class clusters).

In some cases, it may be difficult to assign pixels in satellite images to single thematic classes because they contain a mixture of land cover types. They are hence regarded as mixed pixels. Various methodologies have been developed for ''soft'' classification of such mixed pixels.[26] These are mainly concerned with fuzzy set approaches in which pixels may have partial membership of more than one class.

5.12 Alternatives to Classification: Spectral Mixture Modeling

Spectral mixture modeling is an alternative method of describing mixed pixels. Instead of being interpreted as a mixture of thematic land cover classes, pixels are regarded as a mixture of pure spectral signatures of certain types of reflecting terrestrial surface (e.g. bare soil, bright green vegetation, dry senescent vegetation). Normally these signatures are chosen to be at the extremities of the spectral measurement space and are known as end-member spectra. Each pixel can then be given a labeling which specifies its relative abundance of each end-member spectral component. For this technique to provide unique labelings, the number of end-member spectra used cannot exceed the number of independent spectral channels plus one. Bastin[27] compares spectral mixture modeling to both maximum likelihood classification and fuzzy classification and concludes that fuzzy classification gives the best prediction of subpixel land cover areas. However the

extraction of subpixel information is currently the subject of continued research.[28]

6 CLASSIFICATION REFINEMENT

6.1 Spatial Smoothing and Rule-based Postprocessing

Classification by itself generates thematic image products which usually contain some degree of error. This is often manifested as scattered pixels which are labeled with quite a different class to their immediate neighbors. This kind of error can be removed to a large extent by applying filtering or spatial smoothing operations to classification products. Such postclassification refinement can be performed using, for example, spatial context information or rules about expected class occurrences stored in a knowledge-based system.[29]

6.2 Structural Scene Models and Use of Textural Analysis

As spatial resolutions of satellite images have improved, there has been growing interest in the detection and classification of visible structural features on the Earth's surface such as geological lineaments and artificial objects. The classification of such features can be of enormous benefit in applications such as mineral prospecting and topographic mapping in remote areas. This generally requires the use of computational models to describe such objects and ''intelligent'' algorithms to detect them. A good state of the art survey of recent results in this field is provided by Gruen et al.[30]

Classification of images may also be improved by using the image ''texture'' in the vicinity of a pixel, determined by the local spatial variability of the spectral radiances. Local texture parameters may be computed for each pixel in an image and used as additional spectral feature channels in the classification process. The texture measures described by Haralick et al.[31] are commonly used in satellite remote sensing for this purpose.

ABBREVIATIONS AND ACRONYMS

ASAR	Advanced SAR
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Airborne Visible and InfraRed Imaging Spectrometer
EOS	Earth Observing System
ERTS	Earth Resources Technology Satellite
ESE	Earth Science Enterprise
ESSP	Earth System Science Pathfinder Project
ETM	Enhanced Thematic Mapper
GCPs	Ground Control Points
GER	Geophysical Environmental Research
GOES	Geostationary Operational Environmental Satellite
HRV	High Resolution Visible
HRVIR	High Resolution Visible and Infrared
IFOV	Instantaneous Field of View
IRS	Indian Remote Sensing Satellite
JPL	Jet Propulsion Laboratory
LISS	Linear Imaging Self-Scanning Sensor
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectrometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
PCT	Principal Components Transform
QBM	QuickBird Multispectral
QBP	QuickBird Panchromatic
SAR	Synthetic Aperture Radar
SPOT	Systeme Probatoire de l'Observation de la Terre
TIROS	Television and Infrared Observation Satellite
TM	Thematic Mapper
VCL	Vegetation Canopy Lidar

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Global Land Databases for Environmental Analyses • Hyperspectral Remote Sensing: Data Collection and Exploitation • Land Cover Assessment and Monitoring • Satellite and Sensor Systems for Environmental Monitoring

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. H.J. Kramer, *Observation of the Earth and Its Environment – Survey of Missions and Sensors*, 3rd edition, Springer-Verlag, Berlin, 1996.
2. G. Asrar, R. Greenstone (eds.), *1995 MTPE EOS Reference Handbook*, EOS Project Science Office, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 1995.

3. J.C. Curlander, R.N. McDonough, *Synthetic Aperture Radar*, John Wiley & Sons, New York, 1991.
4. J.C. Russ, *The Image Processing Handbook*, 2nd edition, CRC Press, Boca Raton, 1994.
5. D.H. Ballard, C.M. Brown, *Computer Vision*, Prentice Hall, Englewood Cliffs, NJ, 1982.
6. M. Nagao, T. Matsuyama, 'Edge Preserving Smoothing', *Computer Graphics and Image Processing*, **9**, 394–407 (1979).
7. E. Nezry, H. De Groof, 'The Refined Gamma–Gamma Maximum *a posteriori* Adaptive Speckle Filter for SAR Images', Commission of the European Communities, Joint Research Centre, Ispra, Italy, Technical Report: EUR-16087-EN, 1994.
8. P.J. Huber, 'Projection Pursuit', *The Annals of Statistics*, **13**(2), 435–475 (1985).
9. L. Jimenez, D. Landgrebe, 'High Dimensional Feature Reduction via Projection Pursuit', Proc. International Geoscience and Remote Sensing Symposium (IGARSS 94), held at Pasadena, 8–12 August, 1994, IEEE Press, Piscataway, NJ, Vol. 2, 1145–1147, 1994.
10. E.P. Crist, R.C. Cicone, 'A Physically-based Transformation of Thematic Mapper Data – The TM Tasseled Cap', *IEEE Trans. Geosci. Remote Sensing*, **GE-22**(3), 256–263 (1984).
11. Eurostat (Statistical Office of the European Communities), 'The Impact of Remote Sensing on The European Statistical Information System', Office for Official Publications of the European Communities, Luxembourg, 1998.
12. C. Pohl, J.L. Van Genderen, 'Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications', *Int. J. Remote Sensing*, **19**(5), 823–854 (1998).
13. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.
14. J. Le Moigne, J.C. Tilton, 'Refining Image Segmentation by Integration of Edge and Region Data', *IEEE Trans. Geosci. Remote Sensing*, **33**(3), 605–615 (1995).
15. M. James, *Classification Algorithms*, Collins, London, 1985.
16. T. Kohonen, *Self-organization and Associative Memory*, Springer-Verlag, New York, 1989.
17. M. Schaale, R. Furrer, 'Land Surface Classification by Neural Networks', *Int. J. Remote Sensing*, **16**(16), 3003–3031 (1995).
18. P. Boekaerts, E. Nyssen, J. Cornelis, 'A Comparative Study of Topological Feature Maps Versus Conventional Clustering in Multi-spectral Scene Identification in METEOSAT Imagery', in *Neurocomputation in Remote Sensing Data Analysis*, eds. I. Kanellopoulos et al., Springer-Verlag, Berlin, 232–241, 1997.
19. I.L. Thomas, V.M. Benning, N.P. Ching, 'Classification of Remotely Sensed Images', Adam Hilger, Bristol, 1987.
20. R.G. Congalton, 'A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data', *Remote Sensing of Environ.*, **37**, 35–46 (1991).
21. M.F. Worboys, *GIS A Computing Perspective*, Taylor and Francis, London, 1995.
22. Y.-H. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA, 1989.
23. I. Kanellopoulos, G.G. Wilkinson, F. Roli, J. Austin (eds.), *Neurocomputation in Remote Sensing Data Analysis*, Springer-Verlag, Berlin, 1997.
24. J.D. Paola, R.A. Schowengerdt, 'A Review and Analysis of Backpropagation Neural Networks for Classification of Remotely-sensed Multi-spectral Imagery', *Int. J. Remote Sensing*, **16**(16), 3033–3058 (1995).
25. G.G. Wilkinson, F. Fierens, I. Kanellopoulos, 'Integration of Neural and Statistical Approaches in Spatial Data Classification', *Geographical Systems*, **2**, 1–20 (1995).
26. E. Binaghi, P.A. Brivio, A. Rampini (eds.), *Soft Computing in Remote Sensing Data Analysis*, World Scientific, Singapore, 1996.
27. L. Bastin, 'Comparison of Fuzzy C-Means Classification, Linear Mixture Modeling and MLC Probabilities as Tools for Unmixing Coarse Pixels', *Int. J. Remote Sensing*, **18**(17), 3629–3648 (1997).
28. A.P. Cracknell, 'Synergy in Remote Sensing – What's in a Pixel ?', *Int. J. Remote Sensing*, **19**(11), 2025–2047 (1998).
29. C.C. Kontoes, D. Rokos, 'The Integration of Spatial Context Information in an Experimental Knowledge-based System and the Supervised Relaxation Algorithm – Two Successful Approaches to Improving SPOT-XS Classification', *Int. J. Remote Sensing*, **17**(16), 3093–3106 (1996).
30. A. Gruen, E.P. Baltsavias, O. Henricsson (eds.), *Automatic Extraction of Man-made Objects from Aerial and Space Images (II)*, Birkhäuser Verlag, Basel, 1997.
31. R.M. Haralick, K. Shanmugam, I. Dinstein, 'Textural Features for Image Classification', *IEEE Trans. on Systems, Man and Cybernetics*, **SMC-3**, 610–621 (1973).

Satellite and Sensor Systems for Environmental Monitoring

Wim Bakker
ITC, Enschede, The Netherlands

1 Introduction **1**
1.1 The Remote Sensing System 2
2 Historical Overview **4**
3 Physical Principles of Remote Sensing **6**
3.1 The Electromagnetic Spectrum 6
3.2 Orbits 11
4 Remote Sensors **18**
4.1 Passive Radiometry 19
4.2 Active Radiometry 24
5 Platforms for Remote Sensing **27**
5.1 Aircraft 27
5.2 Drones 27
5.3 (Stratospheric) Balloons 28
5.4 Sounding Rockets 28
5.5 Vessels 28
6 Earth Observation Satellites and Sensors **29**
6.1 Low-resolution Systems 30
6.2 Medium-resolution Systems 32
6.3 High-resolution Systems 37
6.4 Imaging Spectrometry Systems 40
6.5 Imaging Radio Detection and Ranging Systems 41
6.6 Future Systems 44
7 Remote Sensing Images **45**
7.1 What Type of Satellite Image? 48
7.2 Which Image? 48
7.3 Image Processing and Image Products 49
8 Conclusion **51**
Abbreviations and Acronyms **51**
Related Articles **52**
Further Reading **52**
References **53**

More satellites monitoring some aspect of the Earth are launched each day. This article gives an overview of the science and practice of remote sensing (RS). It goes into the technical details of the physics behind RS, lists some typical contemporary and future Earth observing (EO) satellites, and explains the factors that determine the quality of images from space. This relatively short article cannot cover the entire field of RS, but the references and further reading list can be used for further study.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

One could argue that RS dates back to the invention of Aristotle's camera obscura. However, practical RS started after the invention of a stable photographic process by Nièpce and Daguerre, who published their results in 1839. Soon enough, cameras were carried on-board virtually everything that could fly, from hot-air balloons to kites and rockets. The images thus returned by cameras offered an unprecedented view of the Earth and our environment. Large patches of land could be imaged by just a few photographs.

However, this unique viewing capability also poses a challenge in interpreting these images. For one thing, vertical viewing is completely different from normal viewing, providing a very different perspective than what we are familiar with. Usually the sense of depth is lost when viewing a single photograph. The combination of an unfamiliar perspective with a very different scale and lack of recognizable detail can make the most familiar object unrecognizable in an image. Another problem is that our eyes are limited in the spectral sense, being sensitive to only a very small wavelength range, between roughly 400 and 700 nm, the so-called visible part of the spectrum, and the representation of wavelengths outside this window is more difficult for us to comprehend. The interpretation of radio detection and ranging (radar) images, again, are more difficult to interpret, because the way in which radar waves interact with materials requires rethinking the process of how we see objects with radar. All this, the sensors, the platforms, and the interpretation, justify the following definition:

RS is the science of acquiring, processing, and interpreting images that record the interaction between electromagnetic energy and matter.[1]

RS has also been called an ''art'',[2] obviously because it requires skills and experience to interpret RS images.

Major uses of RS images include:

- weather observation and prediction
- topographic mapping
- coastal zone management
- oceanography
- minerals, oil and gas exploration
- agricultural monitoring
- environmental monitoring
- education

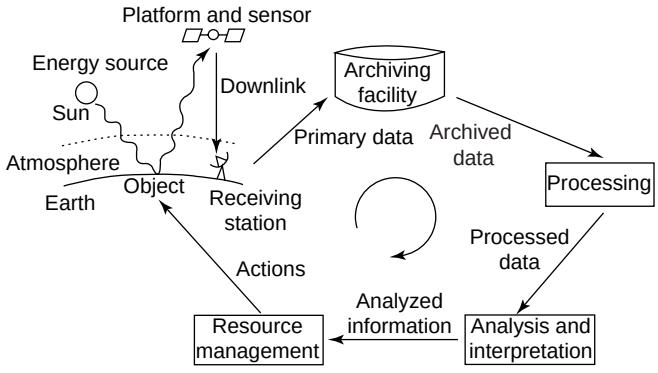

Figure 1 The RS system.

- disaster relief management
- intelligence gathering.

1.1 The Remote Sensing System

The term ''RS activities'' means the operation of RS space systems, primary data collection and storage stations, and activities in processing, interpreting and disseminating the processed data.[3] The last definition comes close to defining RS as a *closed system*, consisting of the following components (see Figure 1).

1.1.1 The Scene

Any object on Earth. Objects are roughly classified as land, ocean, and atmosphere objects, because measurement devices typically are tuned for the measurement task at hand. Usually these three different classes require different types of instruments, although that does not mean that the data from one such type of sensor cannot be used for other purposes.

If the images cover a large part of the Earth, knowledge of the shape and movement of the Earth is also required for the interpretation of the images.

1.1.2 The Energy Source

In principle, the whole spectrum of electromagnetic radiation is available for acquiring information about the object, but in practice only limited parts of the spectrum are used.

In fact, the RS of other sources of energy and fields is possible, such as the measurement of radioactive decay and the magnetic and gravitational field. However, the discussion of these types of measurements is outside the scope of this article.

1.1.3 The Atmosphere

Although good for breathing, the atmosphere is a source of unwanted effects in RS images. Energy is added to, or subtracted from, the signal coming from the object, making atmospheric corrections necessary if absolute measurements are required. Furthermore, various gases of the atmosphere are the cause of its opacity in some of the frequency bands in the electromagnetic spectrum. The frequency bands that are transmitted are the so-called atmospheric windows. Most of the spectral bands used by remote sensors fall entirely within these atmospheric windows, unless, for instance, the absorption itself is an indication of the abundance of certain constituents of the atmosphere. A minor effect of the atmosphere is the displacement caused by refraction.

1.1.4 The Remote Sensing Instrument

The RS instrument can be any type of sensor. Treated in this article are the so-called imaging sensors, sensors that yield two-dimensional data from the Earth. Typically, the remote sensors operate in the ultraviolet (UV), the visible, the infrared (IR), the thermal IR, and the microwave parts of the spectrum. Note that the sensors operating in the optical part of the spectrum, i.e. that part of the spectrum where the energy can be manipulated by optics (extending from roughly 0.3 to 14 μm), can be divided into the optical system, and the detectors themselves. Both subsystems can give rise to their own type of errors. Also note that by the term ''instrument'' we usually mean the entire device, whereas ''sensor'' means only the detector and electronics part. However, these terms are commonly loosely interchanged and used for indicating the same thing: the imaging device.

1.1.5 The Platform

As mentioned before, remote sensors can be flown by any type of platform. This article focuses on airborne RS, but mostly RS by spacecraft. These artificial satellites are launched by rockets and brought into orbit with the help of complex celestial mechanics. Users of RS data usually need not be bothered with how this is actually done and, except for some considerations about minor orbit corrections and attitude control, nothing will be said about this here. The most important factor for the spaceborne platform is its final orbit, as each type of orbit has its own advantages and disadvantages. The height of the orbit clearly determines the details that can be discriminated by the sensor. Its inclination determines the percentage of the Earth's surface that can be imaged, and various parameters together determine the repeat cycle of the orbit and this in turn will determine how often a sensor will be able to take an image of a certain location on Earth. This and more will be discussed in one of the following sections.

1.1.6 The Command and Receiving Station(s)

Left to itself, the satellite will quickly drift from its predetermined orbit, lose its attitude, and may finally be unrecoverable altogether. The satellite has to be taken care of, its status must be checked on a daily basis, and commands must be sent to the satellite to take corrective actions. Furthermore, the sensors must be programmed as to where the next images will be taken. The satellite can transmit the data it records in real time to a ground station, but then a station must be within reach of the satellite. With practical downlink technology, this means that the satellite must be within sight (in the ''line of sight'') of the ground station. Clearly, this requires having ground stations all around the globe to obtain global coverage. There are two ways to overcome the need for having many ground stations. One way is to carry a mass-memory device on-board the platform, so that images can be recorded and played back off-line later when the satellite is over a receiving station. Another way is to relay the data from the RS satellite via communication satellites to the ground. All possibilities are currently being used and have their own specific advantages and disadvantages.

1.1.7 The Archiving Facility

The bytes keep raining to the ground by the megabyte, and most of the information is stored somewhere for possible later use. Some archiving centers already have hundreds of terabytes in their archives, and it is growing daily. What do we do with all these data? These archives are a great source of data for global change detection. But to what lengths should we go to maintain these archives? Already today complete archives of images recorded on microfilm are abandoned, simply because it is too expensive to keep them! Modern systems may even suffer from the fact that there is no archive of previously recorded data from the same sensor. The new images may be incompatible with older sensors, making a comparison very hard, if not impossible.

1.1.8 The Processing and Interpretation

The processing and interpretation of RS data is an essential part of the RS activities, without which most of the data would be virtually worthless. The purpose of the analysis is to transform the data to higher levels of information in order to make the result useful for the application at hand. The processing itself can be divided into three categories, although there is not really a sharp distinction between them. Preprocessing corrects for systematic and random errors. Image processing is a set of techniques for manipulating digital images, the result always being an image. In image analysis, the result of the processing may no longer be image data, but some form of information.

1.1.9 The Dissemination of the Data

The information is then disseminated directly or via resellers to the end-users. Currently there is a blooming market of value-added resellers, who buy and process data for third parties.

1.1.10 The Management of the Earth

The ultimate goal of all these information-gathering activities, of course, is to use the knowledge thus acquired to our advantage for managing the Earth. Many of the RS systems are currently used for monitoring, just watching for changes, not for direct action. Most of these management processes are slow processes, and the fact that RS is part of a closed system is not always as obvious as, for instance, using a computer vision system in a brewing company for removing bad bottles from a conveyor belt. However, some tasks do require a quick response, such as fire detection or any other form of disaster monitoring. Such tasks require near real-time dissemination of the data and, unfortunately, this is, as yet, not the case for most of the traditional land RS systems. There is one prominent exception to this rule, namely meteorology. Meteorologists have a long history of global distribution of all kinds of data in real time across global networks, including satellite imagery. Meteorologists were the first to understand that satellite-based observations ''see between the cracks'' of ground-based observations, providing the earliest clues to the onset of the potential development of severe weather.

In the RS system as defined above, the data flow changes from low-level data to high-level information in a number of steps. Usually the following types of data are distinguished.

1.1.11 The Primary Data or Raw Data

The term ''primary data'' means the raw data that are acquired by remote sensors borne by a space object and that are transmitted or delivered to the ground from space by telemetry in the form of electromagnetic signals, by photographic film, magnetic tape, or any other means.[3]

1.1.12 The Processed Data

The term ''processed data'' means the products resulting from the processing of the primary data, needed to make such data usable.[3] With processed data we simply mean the satellite images that we mostly encounter.

1.1.13 Analyzed Information

The term ''analyzed information'' means the information resulting from the processing and interpretation of data, inputs of data and knowledge from other sources, where a distinction can be made between processing by a computer, and interpretation by a human operator. In practice, this distinction is not a clear one as computers are heavily used in RS and in most applications the touch of a human operator is still needed to drive the process of retrieving information. Analyzed information comes in many forms. One important form is digital data that can be ingested by information systems, usually consisting of a database and a set of tools for manipulating the elements in the database for further analysis.

Tools for the analysis and presentation of the information acquired by RS have traditionally been in the separate fields of photogrammetry, image processing, geoinformation systems (GIS), and cartography, but recently these fields have merged and many software packages have emerged that have strong capabilities in all fields. Such software enables the user to process the data from beginning to end, from raw data to high-quality output maps and information.

The discussion of GIS and resource management information systems is outside the scope of this article, but details can be readily obtained from standard works such as those by Aronoff[4] and McCloy.[5]

The task of producing a useful information product from the primary data is not a trivial one and requires many processing steps and knowledge about the object, the Earth, the source of energy, the atmosphere, the remote sensor, and the platform. Most of these themes will be discussed in the following sections.

2 HISTORICAL OVERVIEW

At the end of World War II the Germans used their V2 rockets to bomb cities in Belgium and the UK. Very few people would have guessed that these terrible weapons would be used shortly after the war for a much nobler goal: to gain knowledge of the upper atmosphere and solar radiation. About 100 confiscated V2s (together with German scientists and engineers such as Dr Wernher von Braun) were transported to the USA and launched successively during the following 5 years. Instead of explosives the nose of the rocket now contained scientific instruments for measuring the chemical composition of the atmosphere, for measuring wind speed, for measuring electric and magnetic fields, and for solar and cosmic radiation. Sometimes these rockets would also carry photographic cameras, for taking pictures of the Earth and the solar spectrum. It was difficult to retrieve the cameras in one piece, so soon the images were transmitted back to the Earth during the flight. Around 1950 the stock of V2s became exhausted and the USA started to build its own rockets. In the former USSR similar developments were going on. During the ''Cold War'' this took the form of a rocket-building competition. Of course, this way of monitoring the Earth and the Sun has disadvantages. The time for one observation using a rocket is short; the time above 80 km was typically less than 2 min. The cost of the launch with respect to the observation time was very high. It was nowhere near a continuous observation of interesting phenomena. The idea was launched of developing a ''long-play'' rocket, one that would be launched to a certain height and circle the Earth at such a speed that it would ''fall'' around the Earth in a circular orbit. The required speed at the surface of the Earth would be $7.9\,\mathrm{km\,s^{-1}}$; at 500 km high the speed would be $7.6\,\mathrm{km\,s^{-1}}$. To launch a satellite in such a circular orbit is a very complex problem, requiring a profound knowledge of orbital mechanics. Then the satellite needs to have instruments that will be able to operate in the harsh space environment, and the signals of the instruments must somehow be received on Earth in order to obtain the data from the instruments. Another problem is the atmosphere of the Earth. The satellite should fly high enough in order not to experience the drag of the atmosphere. However, even at a 300-km height the satellite will experience some drag. It will slow down over time and eventually, after months or even years, at a 90-km height it will be heated by friction with the air and totally burn before it hits the Earth's surface.

The first practical realization of this idea of artificial satellites occurred on 4 October 1957, when the Russians launched their Sputnik I. A second, much larger, Russian satellite followed on 3 November 1957. The first American satellite, Explorer I, was launched on 31 January 1958.

The Space Age probably started with the theoretical basis for rockets developed by Hermann Oberth and Robert H. Goddard directly after World War I, and reached its apotheosis on 20 July 1969 when Apollo 11 landed on the Moon. Neil Armstrong was the first human being to set foot on the Moon. As he placed his left foot on the dusty surface, a breathless world could just distinguish his words: ''That's one small step for a man, one giant step for mankind''.

The main disappointment of the Apollo program was that, after Apollo 17, humankind seemed as far as ever from understanding the origin of the Moon. The first Moon landings, the technical mastery of space travel, and the final scientific expeditions to the lunar surface must nevertheless rank as the most significant step so far in humanity's development.

In the meantime, space was already being used for communication and observation. Systematic satellite

observations of the Earth began in 1960 with the launch of Television and Infrared Observation Satellite (TIROS) 1, the first meteorological satellite, using television tubes as a low-resolution imaging system.

Beginning in the mid-1960s, a large number of studies of the application of color IR and multispectral (MS) photography were undertaken under the sponsorship of the National Aeronautics and Space Administration (NASA), leading to the launch of a MS imager on the Earth Resources Technology Satellite (ERTS) 1. ERTS-1 was later renamed Landsat 1; it was to be followed by four more Landsat satellites in the 1970s and 1980s.

In 1981, the Space Shuttle provided a new platform for remote sensors. The second Shuttle flight carried a package of Earth-oriented sensors including imaging radar, a MS radiometer, and a visible ocean color scanner.

Radar work expanded dramatically during World War II. Today, the diversity of applications of radar is truly startling. It can be used for mapping the topography of the ocean floor, for atmospheric phenomena, for surface structures, for subsurface structures, and for measuring height and wind speed. Radar sensors can be configured in many different modes, including altimetry, scatterometry, synthetic aperture radar (SAR), polarimetry, and interferometry.

A more recently introduced RS instrument is the light amplification by stimulated emission of radiation (laser), which was first developed in 1960. It is mainly being used for atmospheric studies, altimetry, and surface studies using fluorescence.

One more technology that is waiting to become spaceborne is hyperspectral imaging. Hyperspectral imaging or imaging spectrometry takes images in many narrow contiguous spectral bands, as opposed to the MS systems that take only images in a few broad bands. Unfortunately, the first attempt to launch a civil imaging spectrometer failed in 1997, when NASA's satellite Lewis went out of control and burned in the atmosphere.

RS has also expanded dramatically using planetary spacecraft. Images were acquired of the surfaces of the Moon, Mercury, and Mars and the atmospheres of Venus, Jupiter, Saturn, and Uranus. Currently, other objects are also under investigation such as the Sun, asteroids, and comets. Further, there is renewed interest in the Moon, Mars, and the Jovian moons. Although the RS satellites and the planetary probes have many things in common, there are also many differences. The probes will not be discussed here.

In the next millennium large platforms, such as the International Space Station (ISS), may be used as the host for an array of sophisticated sensors that can be replaced periodically by more advanced types when they become available. Major advantages of such a platform are that it will be permanently inhabited, which means that instruments can be serviced whenever necessary, and sensors can be pointed manually at, for instance, disaster areas.

Throughout the Cold War, the USA and the former Soviet Union viewed control of satellite RS paramount to holding an intelligence ''high ground''. The race to space was on. The end of the Cold War produced major changes in government policies restricting the development of new RS capabilities.

At the time of writing, RS is at the brink of a new era. One revolution is the commercialization of high-resolution RS satellites. In 1994, the US government decided to allow the civil commercial development of RS satellite systems providing higher spatial resolution, with up to 1-m resolution. They were afraid that the presumed US leadership in civil satellite RS would erode further if other countries were to develop the 1-m data market first. Several US companies are engaged in a new race to space, trying to be the first in space with a high-resolution sensor.

In general, one can say that the trend is towards ''high resolution'': high spatial resolution, high spectral resolution, high radiometric resolution, and high temporal ''resolution''. The last item, the capability to revisit areas frequently, can also be achieved by RS satellite constellations. The current miniaturization trend for satellites makes that possible. Already at this time 10 microsatellites can be built and launched for the price of one traditional satellite. Smaller satellites can be launched riding piggyback with other microsatellites on a larger satellite. Problems such as shielding, attitude control, and controlling a constellation, instead of a number of individual satellites, will eventually be overcome and pave the way for another boom in space exploitation.

Table 1 lists some highlights of the first Earth images from space.

Table 1 Some highlights of the first Earth images from space

Launch date	Spacecraft	Details
07/08/59	Explorer 6	TV pictures from space
01/04/60	TIROS 1	Weather satellite
06/08/61	Vostok 2	First global picture of the Earth
26/07/63	Syncom 2	Geo-stationary satellite
01/07/67	Dodge 1	1st full-faced color picture of the Earth
23/07/72	Landsat 1	1st Earth resources satellite, ERTS-1
27/06/78	Seasat 1	Sea resources satellite, radar
12/11/81	SIR-A	Shuttle Imaging radar
22/02/86	SPOT 1	10-meter resolution
21/04/95	ERS 1 and 2	First INSAR tandem mission
27/11/97	TRMM	First spaceborne rainfall radar

SIR, Shuttle Imaging Radio Detection and Ranging; SPOT, Système Pour l'Observation de la Terre; ERS, Earth Remote Sensing Satellite; INSAR, Synthetic aperture radio detection and ranging interferometry; TRMM, Tropical Rainfall Measuring Mission.

3 PHYSICAL PRINCIPLES OF REMOTE SENSING

3.1 The Electromagnetic Spectrum

From the previously mentioned definition of RS, it is clear that the knowledge of the electromagnetic field is important for the interpretation of RS images. This section deals with some of the properties relevant for RS. The physical nature will be discussed, in addition to the interaction of electromagnetic radiation with various materials and surfaces. All this is needed to understand what an object will look like in a RS image.

3.1.1 Electromagnetic Energy

Electromagnetic radiation is a form of radiation in which energy is carried by oscillating electric and magnetic fields, which move together and travel through space at the speed of light. Electromagnetic radiation is generated by the acceleration of an electric charge, for example, by the movement of electrons in an antenna. The electric and magnetic fields can be thought of as oscillating at right angles to one another as they move together through space. In the 19th century, James Maxwell developed a set of equations that describe the behavior, with the changing electric field producing the changing magnetic field. The equations include a constant c, which is the speed at which the waves move through empty space. It turned out that this constant is precisely the speed of light, proving that light is a form of electromagnetic radiation. Electromagnetic radiation can also travel through some media, at a slightly slower speed than c.

The entire spectrum, from γ-rays to longwave radio waves, is explained by Maxwell's equations, with the appropriate wavelength for the radiation. Quantum theory also describes the radiation in terms of a stream of particles, called photons; each quantum of radiation with a particular frequency f carries hf units of energy, where h is Planck's constant, 6.63×10^{-34}. γ-Rays are the most energetic and radio waves the least energetic.

In terms of the quantum description, photons behave like particles obeying Bose–Einstein statistics (bosons). Our everyday experience makes it hard to understand how something can behave as both waves and particles, yet many experiments have shown that this is just the way the quantum world works, and physics uses whichever description is convenient. Blackbody radiation, a term explained below, can be described perfectly in terms of photons as a ''gas'' of bosons.[6]

It should be noted that electromagnetic radiation is sometimes expressed as having a frequency, sometimes as having a wavelength λ, and another time as having a particular energy. Frequency f, wavelength λ and energy E are directly related to each other by the simple Equations (1) and (2):

$$c = f\lambda \tag{1}$$

$$E = hf \tag{2}$$

Two beams of light (or any other form of electromagnetic radiation) are coherent when the phase difference between their waves is constant. This can only occur if both beams are of exactly the same wavelength. They are noncoherent if there is a random phase relation. In active systems such as radar and laser, the term coherence means that in addition to the amplitude (intensity) of the reflected beam, the phase can also be measured. The phase can be used for measuring small differences in height. The phase difference $\Delta\theta$ as a result of a difference in height Δh can be expressed by Equation (3):

$$\Delta\theta = \frac{2\Delta h}{\lambda} \tag{3}$$

Height differences of less than one wavelength λ can be measured. By combining the phase information from different observations the topography of the terrain can be reconstructed, and even small displacements of the order of a few centimeters can be measured. The technique of combining the phases of two or more observations is called interferometry. INSAR (also referred to as IFSAR) has been used to visualize depressions or bulges in the terrain caused by earthquakes and active volcanoes. In the future, INSAR may be used as an additional tool for predicting earthquakes or volcanic eruptions by monitoring changes in height.

The Doppler effect is the observed increase or decrease in frequency when the source moves towards or away from, respectively, the observer. Everybody knows the Doppler effect from the lowering of the pitch of the siren when a fire engine passes at high speed. In imaging SAR the Doppler effect, caused by the fast movement of the satellite along the objects, is used to determine whether an object is before, next to, or after the satellite. By cleverly combining all these signals, a better spatial resolution can be achieved, even if the radar antenna is short.

Polarization defines the orientation of the electric field of an electromagnetic wave. Horizontal (H) and vertical (V) polarization refer to the electric field vector being parallel and normal, respectively, to the surface of the medium upon which the wave is incident.

A radar beam can be transmitted horizontally polarized, and the beam can be received horizontally polarized (HH); the beam can be transmitted horizontally and received vertically (HV); the beam can be transmitted vertically and received vertically (VV); or the beam can be transmitted vertically and received horizontally (VH).

HH and VV are called the like-polarization modes and HV and VH are called the cross-polarization modes. If a beam were transmitted horizontally polarized, one would not expect a material to return the signal vertically polarized. Obviously, if this happens, it tells us something about the material. Polarization offers yet another possible way for identifying materials on the surface of the Earth.

The fact that reflected radiation from smooth surfaces is highly polarized can be used to reduce the amount of glint in, for instance, wide-field oceanographic images.

Table 2 Microwave band designation codes

Microwave band	Central wavelength (cm)
Ka	0.85
K	1.2
Ku	2
X	3
C	5
S	10
L	20
P	70

3.1.2 The Spectrum

The most familiar example of a spectrum is the rainbow spectrum of visible light, which can be displayed using a prism or seen in a rainbow itself. White light is made up of a mixture of wavelengths. The spectrum of colors seen by the human eye covers a range from red (with the longest wavelength of about 700 nm) through orange, yellow, green, blue, and indigo to violet (with the shortest wavelength of about 400 nm). The brightness of each color in the spectrum shows how strongly that component of white light contributes to the overall brightness of the source. The properties of the spectrum can be determined accurately either photographically or using electronic detectors. The spectrum extends well beyond the visible range (VIS) at both ends, into the UV and beyond and into the IR and beyond.[6] Usually the IR part of the spectrum is divided into the near-infrared (NIR) (around 1 μm wavelength), the short-wave infrared (SWIR) (around 2 μm), the mid-wave infrared (MWIR) (around 4 μm), and the thermal infrared (TIR) (around 10 μm). Well beyond the TIR region is the microwave range, at wavelengths between roughly 1 mm and 1 m. Most of the Earth observation satellites have detectors that operate in the above-mentioned wavelength ranges (UV, VIS, NIR, SWIR, TIR, and microwave). The combined VIS and NIR range is sometimes designated the visible and near-infrared (VNIR) range. The far-infrared (FIR) range, located between the TIR and the microwave regions, is hardly ever used for RS, simply because the atmosphere is opaque for these frequencies (see section 3.1.5). The microwave range has its own codes for designating the bands (according to the IEEE 521-1984 standard) (Table 2). However, it must be said that these are old codes used in the communications industry, and microwaves are best expressed in frequency or wavelength.

3.1.3 Blackbody Radiation

An object that absorbed all the electromagnetic radiation that fell on it would be a perfect blackbody. Blackbody radiation is the radiation that would be produced by such a hypothetical object when it was warmed. Although a blackbody is an idealized concept, the radiation emitted from objects can be approximately described in terms of equivalent blackbody radiation. All matter at temperatures above absolute zero (0 K) continuously emits electromagnetic radiation. The Sun, for example, radiates rather like a blackbody with a temperature of about 6000 K.[6] Most objects on the Earth have a temperature around room temperature. Hence the radiation emitted by these objects can be approximated by a blackbody of 300 K temperature.

The total amount of energy radiated by a blackbody is expressed by the Stefan–Boltzmann law, which states that the energy is proportional to T^4. Therefore, the energy increases very rapidly with increase in temperature. An object at 6000 K radiates 160 000 times more energy than the same object at 300 K. The shape of the curve, with respect to wavelength, is given by Planck's law (see Figure 2). Two things are obvious: First, the height of the curve increases with increase in temperature and the energy radiated is proportional to the area under the curve; and second, the top of the curve shifts to shorter wavelengths and more energy is radiated at the shorter

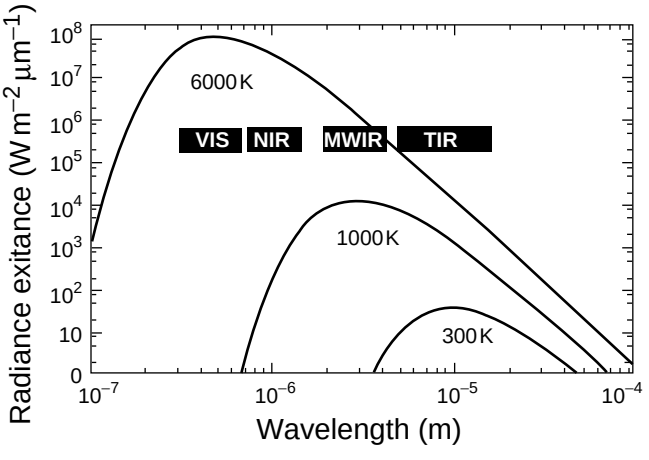

Figure 2 The blackbody radiation curves for 300, 1000, and 6000 K. Indicated are the locations of the VIS, NIR, MWIR, and TIR bands.

wavelengths. This is why a warm lump of iron radiates invisible radiation, a slightly hotter lump of iron glows red, an even hotter lump glows white hot, and eventually it will glow bluish white. The top of the curve first shifts into the red part of the spectrum and moves towards the blue part with increasing temperature. This shift is determined by Wien's displacement law, which states that the wavelength at which the blackbody radiation curve reaches its maximum is inversely proportional to the temperature (Equation 4):

$$\lambda_{max} = \frac{A}{T} \tag{4}$$

where λ_{max} = wavelength of maximum spectral radiant exitance in micrometers, $A = 2900\,\mu m\,K$ and T = temperature in kelvin.

The peak of the 6000 K curve of the Sun is located in the middle of the visible part of the spectrum. This may not come as a surprise, because the evolutionary adaptation of the detectors in our eyes will have tuned these in such a way that they are most sensitive to the energy that is abundantly available, sunlight.

The peak of the 300 K curve of objects at room temperature, the approximate ambient temperature of the Earth, is located around 10 μm. As this type of radiation is strongly related to the temperature of objects, this part of the spectrum is called the TIR.

As passive remote sensors measure reflected radiation from the Sun, or measure energy emitted by the object itself, most of the sensors operate in the VIS and TIR parts of the spectrum. However, there are many interesting absorption, reflection, and emission features present in the NIR, the SWIR and the MWIR regions, which explains the many sensors operating in this part of the spectrum. Examples of these features are the high reflection of vegetation in the NIR region, the absorption dips of certain rock minerals in the MWIR region, and fire detection: a fire usually has a temperature of about 1000 K, and has peak radiation around 3–4 μm, in the MWIR region. This means that areas with fires stand out as bright spots on the satellite images taken in the MWIR region, and can be easily detected.

3.1.4 Absorption, Emission, Reflection, Transmission, and Refraction

Although the concept of a blackbody is a way to describe ideal radiation principles, real materials do not behave as blackbodies. Instead, all materials emit less than the energy emitted from a blackbody at the equivalent temperature. The ratio of the real energy radiated by the material over the energy that would be radiated by a blackbody at the same temperature is referred to as the material's emissivity, ε. By definition, ε will have values between 0 and 1. The emissivity, just like reflection, may have different values for different wavelengths λ and viewing angle θ. Furthermore, depending on the material, emissivity can vary with temperature. If a material has an emissivity that is constant for all wavelengths, it is called a graybody.

Water comes close to being a blackbody in the 6–14 μm TIR range, with an emissivity of 0.98–0.99. We have seen that a blackbody is both a perfect absorber and a perfect radiator. This is true in general, and is expressed by Kirchhoff's law, which states that the spectral emissivity of an object equals its spectral absorptance (Equation 5):

$$\varepsilon(\lambda) = \alpha(\lambda) \tag{5}$$

This means that good absorbers are good emitters. The emissivity factor must be taken into account if one wants to determine the temperature of objects. In the TIR range the emitted radiance is usually expressed as a temperature. As objects radiate less energy than blackbodies, their apparent temperature, or brightness temperature, is less than the real temperature, or kinetic temperature, of the object. Bearing in mind that the energy is proportional to T^4, it can be shown that Equation (6) holds:

$$T^4_{measured} = \varepsilon T^4_{real} \tag{6}$$

To understand the reflective properties of surfaces, we have to make a distinction between a perfect specular reflector and a perfect diffuser. Specular reflection is the mirror-like reflection of a smooth surface, where the angle of incidence equals the angle of reflection. Just as we learn little from the mirror by looking at the mirror – we see ourselves – RS usually is not interested in the specular type of reflection. Mostly this type of reflection gives rise to unwanted effects, such as "hot spots" and glint in images, and precautions must be taken not to see this. For instance, wide-angle oceanographic sensors must be pointed away from the specular reflection of the Sun on the water surface, in order to avoid overload of the sensor and sea glint in the images.

Diffuse reflection is the matte reflection of a rough surface. A perfect diffusing surface backscatters the incident radiance in all directions according to Lambert's cosine law. Lambert's law states that the radiant energy of a diffuse surface varies with the viewing angle (Equation 7):

$$I = I_0 \cos\theta \tag{7}$$

where I_0 is the intensity along the normal to the surface and I the intensity along the line at an angle θ with the normal.

As there is no sharp boundary between a smooth and a rough surface, real materials exhibit a combination of the

specular and diffuse reflection properties. To make things worse, real materials are not perfect diffusers. In fact, the reflecting properties of materials can only be measured in the field, and are a function of the angle of incidence of the radiation and the viewing angle. This function is called the bi-directional reflectance distribution function (BRDF), and usually shows a smooth transition between the specular and diffuse reflection properties of a material. The BRDF depends on wavelength and is determined by the structural and optical properties of the surface, such as shadow casting, multiple scattering, mutual shadowing, transmission, reflection, absorption and emission by surface elements, facet orientation distribution, and facet density.

In addition to absorption, reflection, and emission, materials may also transmit energy, passing part of the radiation unharmed. The ratio between the energy hitting the surface and the energy passed by the material is called the transmission, τ. For preservation of the total amount of energy, Equation (8):[2]

$$\alpha(\lambda) + \rho(\lambda) + \tau(\lambda) = 1 \tag{8}$$

states that the absorptance, reflection, and transmission add up to 1.

For thermal radiation remember that the absorptance and emissivity may be freely interchanged. Furthermore, most objects in RS are assumed to be opaque to thermal radiation (Equation 9):

$$\tau(\lambda) = 0 \tag{9}$$

such that we can write (Equation 10)

$$\varepsilon(\lambda) + \rho(\lambda) = 1 \tag{10}$$

This shows that the lower an object's reflectance, the higher is its emissivity, and vice versa, the higher an object's reflectance, the lower is its emissivity. For instance, water has a nearly negligible reflectance in the thermal spectrum, and as a result it has an emissivity of almost 1.

Furthermore, the material changes the direction of the transmitted energy beam. Snell's refraction law expresses the relationship between the incident geometry and the transmitted geometry, and depends on the speed of light in both media (Equation 11):

$$\frac{\sin\theta_i}{\sin\theta_r} = \frac{c_0}{c} \tag{11}$$

where θ_i is the angle of incidence, θ_r is the angle of refraction, c is the speed of light outside the material, and c_0 is the speed of light in the material.

The ratio c_0/c is also known as the refractive index, n.

For certain materials, the speed of light c_0 is a function of the wavelength λ, causing the angle of reflection to differ slightly for different wavelengths (colors) in the VIS. This causes the incident beam of mixed wavelengths to be broken up into a fan of split wavelengths. This effect is called dispersion, and can be used to unravel a mixed beam of light, composed of many wavelengths, into its components. An example of such a device is, of course, the prism, which is capable of splitting white light into a rainbow of pure colors.

Other dispersive elements, but based on a different principle, are gratings. Gratings are polished surfaces with a large number of equidistant, parallel grooves. Dispersive elements are used at the heart of imaging spectrometers.

3.1.5 Atmospheric Effects

Like any other material, the constituents of the atmosphere absorb, transmit, reflect, and refract the energy passing through it. Also, the atmosphere, because it has a temperature above 0 K, emits energy by itself. Carbon dioxide, water vapor, and ozone dominate the interaction of the atmosphere with electromagnetic radiation.[7]

3.1.5.1 Absorption Carbon dioxide (CO_2), absorbs energy near the wavelengths 2.01, 2.08, 4.3, and 15 μm. CO_2 plays an important role in the energy budget of the Earth and its atmosphere, and is known to be a ''greenhouse'' gas.

Water vapour (H_2O), is particularly important because of the strong absorption lines in the IR and microwave regions. Water vapor concentration is highly variable in space and time, and may vary from $10^{-2}\,g\,m^{-3}$ in a very cold and dry climate up to $30\,g\,m^{-3}$ in hot and humid regions.[7] However, even within one satellite image the water vapor concentration may vary from location to location, which makes it an extremely difficult factor to correct for. The absorption bands in the NIR region are known as the water vapor absorption bands, in the VNIR range centered at approximately 0.94, 1.14, 1.38, and 1.88 μm.

Ozone (O_3), strongly absorbs UV radiation and causes a shortwave cutoff of the atmospheric transmission below 0.3 μm. Ozone is mostly concentrated at altitudes between 20 and 50 km and its distribution is also highly variable. The total column ozone may vary from 100 to 500 Dobson units with season and type of weather. The amount of ozone is expressed in Dobson units, where 100 Dobson units is equivalent to a 1 mm layer of ozone at sea-level pressure at a temperature of 0 °C.

Oxygen (O_2), which absorbs at 0.76 μm, and a number of other minor constituents, ''trace gases'', such as methane (CH_4) and nitrous oxide (N_2O), also have minor effects on the absorptive properties of the atmosphere.

Obviously, it does not make much sense to build an Earth observation sensor for wavelengths that are

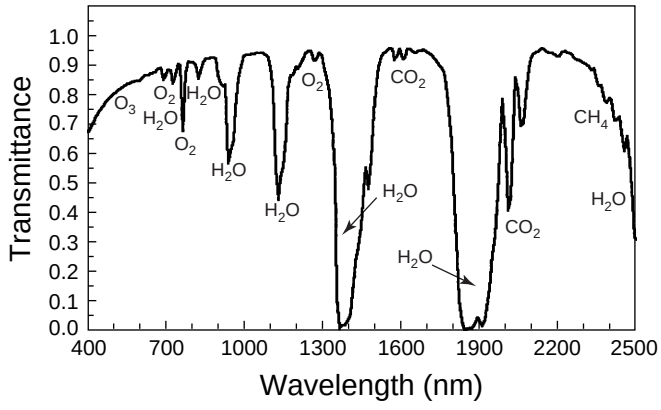

Figure 3 The atmospheric transmittance. Wavelengths with high transmission are called the ''atmospheric windows''. The dips in the curve are caused by atmospheric absorption, while the decrease in transmittance towards 400 nm wavelength is caused by atmospheric scattering.

absorbed by the atmosphere. Most sensors operate in the spectral regions where the atmospheric transmission is high. These regions are called atmospheric windows. Luckily, for humans, the visible part of the spectrum falls within such an atmospheric window. We would not see very much otherwise. There are a number of windows in the NIR and MWIR regions, and the TIR region falls within one window. Then there is darkness until we enter the microwave range, where there are a number of windows again. The atmospheric windows clearly determine the spectral band placing of the Earth observation satellites (see Figure 3).

Major atmospheric windows, where atmospheric absorption is low, are as follows:

- 0.3–1.3 μm VNIR
- 1.5–1.8 μm SWIR
- 2.0–2.5 μm MWIR
- 3.5–4.1 μm MWIR
- 7.0–15.0 μm TIR
- 1 mm–1 m (microwave).

Some satellites measuring properties of the atmosphere may look *at* the absorption bands of the atmosphere rather than through the atmospheric windows. For instance, the spectral range around 6–7 μm is sometimes called the water vapor band (WV) as it gives a measure of the amount of water vapor in the atmosphere. It should be noted, however, that there are many water vapor absorption bands, especially in the VNIR range.

The amount of absorption by the atmosphere can be expressed by an exponential relationship with the optical depth (Equation 12):

$$\alpha \sim e^{-\text{depth}} \tag{12}$$

The optical depth is the product of the extinction coefficient, the density of the medium, and the length of the optical path through the medium (Lambert–Beer law).[8]

3.1.5.2 Particulate Scattering Particulates are atmospheric dust particles, called aerosols, with diameters between 0.1 and 10 μm. Particles tend to scatter radiation in all directions, depending on the size of the particle and the wavelength of the radiation. Three types of scatter models are used for particles of sizes smaller than the wavelength, equal to the wavelength, and larger than the wavelength.

Rayleigh scattering is the scattering by particles small in size compared with the wavelength being scattered. It is dominant at elevations of 9–10 km above the surface. Roughly, the scattering is inversely proportional to λ^4. For the visible part of the spectrum, this means that the blue light is scattered roughly 10 times as much as the red light. A blue sky is a manifestation of Rayleigh scattering. In the absence of scattering, the sky would be black.

Another type of scatter is Mie scatter, which exists when the size of atmospheric particles more or less equals the wavelength. It is a low atmosphere scattering, up to 5-km height, caused by dust, pollen, smoke, and small water droplets. The effects are approximately inversely proportional to the wavelength λ, and affect electromagnetic radiation mostly in the visible region.

The last type is the nonselective scattering in the lower atmosphere by particles much larger than the incident wavelength. It gets its name because the effects are roughly independent of the wavelength. Nonselective scattering is the primary cause of haze. Water droplets, for example, cause such scatter. They commonly have a diameter in the range 5–100 μm and scatter all VIS, NIR and MWIR wavelengths about equally. In the VIS, equal amounts of all colors are scattered, hence fog and clouds appear white.

Scattering causes haze in imagery. It is visible as a blue haze on photographs of distant objects. In black and white photography this blue haze can be filtered by using a blue-blocking filter, which improves the haze penetration (but reduces sensitivity).

Atmospheric scattering causes the decrease in transmission in Figure 3 towards the smaller wavelengths (blue).

3.1.5.3 Atmospheric Emission The atmosphere has a temperature above 0 K and therefore radiates energy. This means that sensors looking at the thermal radiation of Earth objects also receive an unwanted dose of atmospheric radiation.

On the other hand, the emissivity together with the absorptance can be used for building a temperature profile of the atmosphere. The atmosphere has an absorptance that varies with wavelength from about 0% at 10 μm to

about 100% at 16 μm. An IR sounder with, for instance, five channels in the region from 11 to 16 μm probes the radiation of the atmosphere to different depths. Since the absorption is high at 16 μm, only the top of the atmosphere is sensed by the 16-μm channel. The 11-μm channel, however, senses the radiation from all attitudes. The other channels are somewhere in between.

3.1.5.4 Atmospheric Refraction The atmosphere also causes radiation to be bent towards the Earth, causing a minor virtual displacement of the object. For incidence angles less than about 45°, Equation (13) is applicable to satellite observations:[9]

$$\Delta x \approx 2.5 \sin\theta (\cos\theta)^{-4.5} \tag{13}$$

where Δx is the observed displacement in meters and θ is the look angle.

The displacement is about 0.5 m at a 10° look angle, 1 m at 20°, 2.5 m at 30°, 5 m at 40°, and about 8 m at 45°. This means that it was never really an important factor in the traditional RS instruments with a resolution greater than 10 m and look angles less than 30°. However, for modern high-resolution systems with extreme pointing capabilities, the displacement is of the order of a few pixels and will become apparent.

The Earth's ionosphere and the water vapor content of the atmosphere may have a considerable effect on microwaves. At lower frequencies (150 MHz), daytime ionospheric biases can easily reach several kilometers in range.[8]

The atmospheric effects are difficult to model. The best model currently available is probably the Second Simulation of the Satellite Signal in the Solar Spectrum (6S),[10] which is the basis of many models and programs for the correction of atmospheric effects. However, a very simple model for the energy reaching the sensor sufficing for most applications is given by Equation (14):

$$L_{\text{sensor}} = \rho\tau^2 L_{\text{sun}} + L_{\text{skylight}} \tag{14}$$

where L_{sun} is the energy from the Sun, τ is the transmittance of the atmosphere (which is traversed twice), and L_{skylight} is the ambient light caused by scattering. The reflectance ρ of the object is the parameter in which we are interested. This can be approximated by Equation (15):

$$L_{\text{sensor}} = a\rho + b \tag{15}$$

The factors a and b can be determined from known objects present in the image. The approximation gives us a direct relationship between the values measured by the RS instrument and the physical phenomena on the ground.

3.1.6 Fluorescence

Some materials consist of molecules that absorb photons of high energy and re-emit the absorbed energy in one or more photons of lower energy. The light is transformed from a short wavelength (high energy) to a long wavelength (low energy). The term fluorescence is especially used when materials that are subjected to UV radiation start to glow with various colors. However, plants also re-emit a portion of the absorbed visible light into the red and NIR region between 0.65 and 0.75 μm. Fluorescence can be used for the identification of materials. In RS, fluorescence is applied by flashing a laser and recording the spectrum of the re-emitted light.[8]

3.2 Orbits

The way in which an artificial satellite "flies" is explained by the laws of Kepler. Once launched into orbit, the satellite will remain, theoretically at least, in a continuous free fall around the Earth, just like the Moon orbits the Earth, and just like the Earth itself orbits the Sun. It requires neither a pilot to fly it nor fuel, except for making minor orbit corrections. The parameters of the orbit determine to a large extent the monitoring capabilities of the sensor.

Orbit selection and sensor characteristics are closely related to the strategy required to achieve the desired results. Different types of orbits are required to achieve continuous monitoring, repetitive coverage of different periodicities, global mapping, or selective imaging. Most Earth-orbiting RS satellites use circular orbits. Planetary orbiters usually have elliptical orbits, which is less taxing on the spacecraft orbital propulsion system and combines some of the benefits of high and low orbits, thus allowing a broader flexibility to achieve multiple scientific objectives.[7]

Although Newton's universal law of gravitation and Einstein's field theory of gravity apply to satellite orbits, they can best be visualized and explained by using Kepler's laws of planetary motion. Kepler's laws can be applied directly to satellites orbiting the Earth, and explain the properties of the typical orbits used in RS.

3.2.1 Kepler's Laws

Kepler's laws of planetary motion were derived by the German astronomer Johannes Kepler, whose analysis of the observations of the 16th century Danish astronomer Tycho Brahe enabled him to announce his first two laws in 1609 in his publication "Astronomia Nova" ("New Astronomy") and a third law a decade later, in 1619, in his "Harmonice Mundi" ("Harmony of the World").

Brahe's observations were remarkable because of their accuracy. He performed his minute observations without

the help of telescopes, as these were only to be invented in the early 17th century by Galileo! Kepler went on to provide the beginning of a theory of the telescope in his *Dioptrice*, published in 1611.

It is interesting that Kepler derived all three of these laws by dead reckoning: he did not have calculus available to assist him (it had not been invented then).

He began to think of planets as moving in ellipses when his data from the orbit of Mars showed that it did not follow the path of a circle, which is what everyone assumed at the time.

It is interesting that, should Kepler have been studying another planet, he would not have noticed any discrepancy between its true and predicted (circular) orbit. Mars is the only planet at the right distance from Earth and with an eccentricity large enough so that its orbit could be distinguished from a circle with the equipment available at the time.

Kepler's three empirical laws of planetary motion can be stated as follows:

1. Each planet moves in an ellipse with the Sun at one focus point (orbits are ellipses) (see Figure 4).
2. A line connecting the planet to the Sun sweeps out equal areas in equal times (law of equal areas) (see Figure 5).
3. The squares of the periods T of the planets are directly proportional to the cube of their mean distances A from the Sun; T^2/A^3 is a constant (the harmonic law).

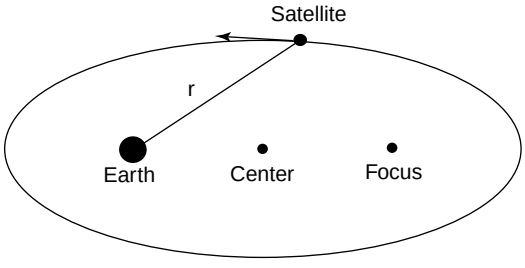

Figure 4 Kepler's first law applied to satellites orbiting Earth; the satellite moves in an ellipse that has the Earth in one of the focus points.

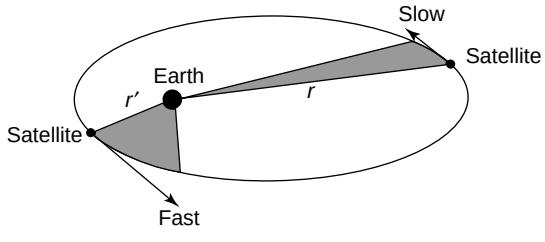

Figure 5 Kepler's second law applied to satellites orbiting Earth; the satellite has a higher velocity near to Earth.

Knowledge of these laws, especially the second (the law of equal areas), proved crucial to Isaac Newton in 1684–85. He then formulated his famous law of universal gravitation between the Earth and the Moon, and between the Sun and the planets, postulated by him to have validity for all objects anywhere in the Universe. Newton showed that the motion of bodies subject to central gravitational force need not always follow the elliptical orbits specified by the first law of Kepler, but can take paths defined by other, open conic curves. The motions can also be in parabolic or hyperbolic orbits, depending on the total energy of the body. Thus, an object of sufficient energy, e.g. a comet, can enter the solar system and leave again without returning.

The third law can be recognized in Equation (16) for calculating the orbital period of a certain satellite as function of its height:[7]

$$T = 2\pi \sqrt{\frac{r^3}{gR^2}} \tag{16}$$

where $g = 9.81\,\mathrm{m\,s^{-2}}$, $R = 6380\,\mathrm{km}$, and r is the orbit radius.

It can be noted that Newton used his new vector calculus and laws of motion and gravitation to show that Kepler was right. One day in 1689 he came up to his friend, Edmund Halley, and casually mentioned to him that he had proved that, with a $1/r^2$ force law like gravity, planets orbit the Sun in the shapes of conic sections. This undoubtedly took Halley aback, as Newton had just revealed to him the nature of the Universe (at least the Universe as it was known then). Halley then pressed Newton to publish his findings, but Newton realised that he had forgotten the proof. After struggling to remember how he had proved the theorem, he published his work and it later appeared in full form in his *Principia*. Later, Halley examined the orbits of comets and, armed with Newton's new understanding of the effects of gravity, computed the orbit of Halley's comet. He predicted that it orbits the Sun once every 76 years, and named the date of the next passage. Unfortunately, he died before that year, but astronomers saw the comet come just when he said it would. For that, the comet received his name: Halley's comet.

The usefulness of Kepler's laws extends to the motions of natural and artificial satellites in addition to unpowered spacecraft in the orbit of planets.

As formulated by Kepler, the laws do not take into account gravitational interactions (as perturbing effects) of the various planets with each other, or of the Moon and the Sun with artificial satellites. The general problem of accurately predicting the motions of more than two bodies under their mutual attractions is complicated; analytical solutions of the three-body problem are unobtainable except for some special cases.

The physical principles of orbits are treated extensively by Rees.[9]

3.2.2 Orbital Models

Basically, orbital models are used for two things:

1. To be able to calculate the exact location of the satellite with respect to the Earth. The position of the satellite is used for determining the exact location of each scanned pixel on Earth.
2. To know when the satellite will pass overhead and will take images of the area of interest.

Of course, point 1 is the most difficult, because it needs high accuracy. The results are used to perform geometric corrections on the satellite images. There are many programs available that are useful for predicting when the satellite will pass overhead.

Many orbital models exist, even though they all use the same theory of orbital motion. The differences are in the integration techniques used and the perturbations of the orbits that are accounted for. Each model requires its own particular set of orbital elements. Simple conversions between element sets, even though they may look similar, will lead to inaccuracy! Most orbital element sets will have a close resemblance to the Keplerian elements discussed below.

Images acquired by RS are subject to different distortions due to the Earth's curvature and rotation, the spacecraft's speed, altitude and attitude, the scan skew, and the projection of a spherical surface on a flat image. These distortions, if not properly accounted for, will prevent meaningful comparison among images, particularly in sequential image analysis or multisatellite data studies. Therefore, the need for image referencing, i.e. identifying the geographic coordinates corresponding to an image pixel or locating a pixel corresponding to given geographic coordinates, is inevitable in many satellite data studies.

In order to be able to establish the exact location of a pixel of a digital satellite image, a number of things must be known. First, of course, we must know what we are looking at; we must have a model of the Earth itself. Second, the characteristics of the imaging sensor must be known. Last, we must know the exact position, velocity, and attitude of the satellite at the moment of imaging.

The precision obtained depends on, amongst other things, the particular model of the Earth used. Most of the time this will be a nonspherical geoid model. The Earth models are beyond the scope of this article and will not be discussed here.

The geometric characteristics of commonly used sensors will be discussed elsewhere. This section will tackle the problem of the position of the satellite.

3.2.3 Often Used Orbital Models

This section discusses the most commonly used orbital models, the SGP, SGP4, SDP4, SGP8, and SDP8 models, as they were adapted by the North American Aerospace Defense Command (NORAD), which is the NASA-related organization responsible for tracking all satellites. The NORAD regularly issues updated orbital element sets. These sets contain the so-called two-line elements (TLE) for each satellite (see the next section).

The models are analytical, which means that the result can be calculated without having to perform numerical integration. This means that the calculations for determining the position of a satellite can be performed within a limited amount of time.

Five mathematical models for the prediction of satellite position and velocity are available. The first of these, SGP, developed in 1966, is used for near-Earth satellites. This model uses a simple gravitational model and it takes the drag effect on mean motion as linear with time.

The second model, SGP4, was developed in 1970 and is used for near-Earth satellites. This model was obtained by simplification of the more extensive analytical theory developed in 1969, which uses a more advanced gravitational model and a power density function for its atmospheric model.

The next model, SDP4, is an extension of SGP4 to be used for deep-space satellites. The deep-space equations were developed in 1979 and model the gravitational effects of the Moon and the Sun in addition to certain sectoral and tesseral Earth harmonics which are of particular importance for half-day and one-day period orbits.

The SGP8 model of 1980 is used for near-Earth satellites and is obtained by simplification of an extensive analytical theory, which uses the same atmospheric models as SGP4, but integrates the differential equations in a very different way.

Finally, SDP8 model is an extension of SGP8 to be used for deep-space satellites. The deep-space effects are modeled in SDP8 with the same equations as used in SDP4.

NORAD classifies all space objects as near-Earth (period less than 225 min) or deep-space (period greater than 225 min). The value of the mean motion can be used to determine whether the satellite is near-Earth or in deep space.

All this does not mean that there are no other models in use. Different models may be used for different requirements of the final result. If the result does not require high accuracy, then a simpler model than the SGP4 model could be used. If high accuracy is required for just one specific satellite, then a specific model might be used for that orbit. For instance, for the National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites the following models are known to be used (D'Souza et al.[11] 153–193).

Table 3 Example of a NORAD TLE set, with its meaning listed below

NOAA 12							
1 21263U	91032A	94321.74937578	0.00000127	00000-0	76078-4	0 2713	
2 21263	98.6030	346.0459	0013195	147.3105	212.8891	14.22467535182330	

Satellite	NOAA 12, one of the polar orbiting satellites
Epoch	Year 1994
	Day of year 321
	Fraction of day 0.74937578
Inclination	98.6030°
RAAN	346.0459°
Eccentricity	0.0013195
Argument of perigee	147.3105°
Mean anomaly	212.8891°
Mean motion	14.2246753 revolutions per day
Drag	0.00000127

1. circular model
2. elliptical model
3. Brouwer–Lyddane model
4. SGP4 model
5. Kloster model.

Furthermore, the NOAA uses its own specific orbital element set for the NOAA polar orbiting satellites, the so-called TBUS part IV elements.

3.2.4 Orbital Elements Sets

Satellite orbital elements are numbers that tell us the orbit of each satellite. The orbital elements are the parameters that drive the orbital model.

3.2.4.1 The Keplerian Elements Seven different numbers are needed to describe accurately an orbit in space and time. Although many different sets of numbers will do (including position and velocity at a given time), there is a standard set called the Keplerian elements that are most used and most useful.

Epoch. The time when the elements were made.

Inclination. The angle between the orbital plane and the equatorial plane.

Right Ascension of Ascending Node. The right ascension of ascending node (RAAN) is the angle between the ascending node, where the orbit crosses the equator going south to north, and the point of Ares, the reference point for the right ascension.

Argument of Perigee. The angle between the perigee and ascending node, or between the major axis and the line of nodes.

Eccentricity. A parameter that determines how squashed an ellipse is. Also, the ratio between the center-to-focus distance and the semimajor axis.

Mean Motion. The average rotation rate of the satellite.

Mean Anomaly. The position of the satellite in the orbit at the epoch, measured in an angle of an equivalent circular orbit moving at constant rate.

An example of a NORAD TLE is given in Table 3.

3.2.5 Perturbations

If orbital motion were that of an ideal two-body system, then the orbital elements would remain constant in time except for the mean anomaly, which would vary linearly in time. In reality, a number of perturbations slowly cause the orbit to change:

1. The Earth is not a perfect sphere, and the gravitational field of the Earth is irregular owing to inhomogeneities.
2. The additional pull of the Moon, the Sun, and the planets.
3. The solar wind.
4. The atmospheric drag and lift.
5. The Earth's magnetic field.

The mean orbital elements of a near-Earth satellite exhibit the following effects:

1. Semimajor axis and eccentricity vary approximately linearly with time owing to atmospheric drag.
2. Inclination remains constant.
3. Argument of perigee and RAAN vary linearly with time owing to the Earth's oblateness, and they have a small quadratic variation with time owing to the indirect effect of the changing semimajor axis.
4. Mean anomaly has an additional linear variation with time due to the Earth's oblateness and an

additional small quadratic variation due to the changing semimajor axis.

Although it may seem that all these effects are highly undesirable, one is actually used for creating a special type of orbit that most RS satellites use. The linear change with time of the RAAN means that the orbit is slowly rotating around the Earth's axis. This rotation is called precession, and is comparable to the slow wobble of a spinning top. Where the gravitation tries to overturn the top, because of the spinning, it actually starts to precess. More or less the same happens with satellite orbits. Because of its oblateness, the Earth has more mass on the equators and gravity tries to decrease the inclination, but, because the satellite is spinning in its orbit, the orbit precesses instead. If the parameters of the orbit are carefully chosen, the orbit can be designed in such a way that the period of precession exactly matches the 1-year orbital period of the Earth around the Sun. This means that the orbit can be chosen in such a way that the satellite always records images at the same local solar time, all year long. Such an orbit, called a sun-synchronous orbit, would not be possible without precession. The height of the orbit and its inclination are the most important factors to determine the precession period.

3.2.6 Atmospheric Drag

Usually the drag of the atmosphere will be the most important factor for changes in the satellite's orbit. Most of the error, however, will be an error in the along-track direction of the orbit. This means that the errors in the satellite position will mostly look like a timing error; it will not pass overhead at the exact predicted time.

Bear in mind, however, that a timing error of 1 s does not look so dramatic, but in fact the satellite travels through space at a velocity of about $7\,\mathrm{km\,s^{-1}}$. A timing error of 1 s means an error in location of 7 km! Compared with the pixel size of the Landsat Thematic Mapper (TM) sensor of 30 m, this already means a location error of 250 pixels on the ground!

Most models add the following ''element'' to the Keplerian elements.

Drag. The drag orbital element simply tells us the rate at which mean motion is changing owing to drag or other related effects. Precisely, drag is half of the first time derivative of mean motion.

3.2.7 Typical Satellite Orbits

As may be clear from the previous section, although all orbits are ellipses, many choices remain for the orbital parameters, making many different orbits possible. Still, most satellites are placed in orbits with specific characteristics. Each orbit has its own particular advantages and disadvantages. The most commonly used orbits will be discussed here.

3.2.7.1 Circular Orbit Orbits are ellipses, and yet most satellites are in orbits that are almost a perfect circle. EO satellites are designed to fulfill one specific task, and a change in altitude would cause an undesired change in scale of the images.

On planetary missions often a highly elliptical orbit is used, because the solitary spacecraft has to fulfill a large diversity of scientific objectives. The lowest point in the orbit yields detailed images of the surface, while the highest point offers a vantage point on the planet. Only by using an elliptical orbit can a three-dimensional view of the magnetosphere be constructed.

Advantage of a circular orbit.

- Stable, no change of scale in the images.

Advantages of an elliptical orbit.

- Offers detail of the surface and a synoptic view on the planet.
- Needed for studying the magnetosphere.

3.2.7.2 Low Earth Orbit A low Earth orbit (LEO) is an orbit near the top of the atmosphere (150–300 km high) with a period of about 90 min.

Advantages.

- These orbits are the easiest to get satellites into. All other orbits require more energy and rocket fuel and thus larger and more expensive rockets.
- These orbits are ''closest'' to the Earth, so the sensor does not have to be very sophisticated in order to achieve high resolution.
- For active systems like SAR, a LEO requires the least power.

Disadvantages.

- These orbits are closest to the atmosphere and will suffer most from drag. This means that these satellites must often correct the orbit, otherwise they will spiral towards the Earth fairly rapidly and burn in the atmosphere. This is especially true in the years with high solar activity when the atmosphere swells owing to the solar wind and increased solar radiation.

- In such a low orbit the satellite views only that portion of the Earth's surface above the horizon, which is about 3% of the total surface.

These orbits are perfect for short-time missions of the order of weeks and months. Most spy satellites during the Cold War used such low orbits. More recently, the Space Shuttle is designed specifically for using low orbits.

3.2.7.3 Near-earth Orbit Orbits between 300 km high and roughly twice the Earth's radius, 6400 km high. Most satellites are in a near-Earth orbit (NEO), high enough to experience little drag by the atmosphere, but low enough to view much detail of the Earth's surface.

Advantage.

- Less drag.

Disadvantage.

- The sensor has to be more advanced in order to achieve the same resolution as in a LEO.

3.2.7.4 Prograde Orbits having an inclination angle between 0 and 90° are the eastward-launched orbits. It is cheaper to launch a rocket in an eastern direction, because it already has the speed of the rotational speed of the Earth, which can be as much as 0.5 km s^{-1} on the equator. A major disadvantage of this type of orbit is that the precession rotates in the wrong direction: instead of rotating with the Earth, the satellite precesses against the Earth, making a Sun-synchronous orbit impossible.

Advantage.

- Requires less fuel than a retrograde orbit.

Disadvantage.

- Cannot be used for a Sun-synchronous orbit.

3.2.7.5 Retrograde Orbits with angles between 90 and 180° are the westward-launched orbits. The latter are the so-called retrograde orbits.

Advantage.

- Makes Sun-synchronous orbit possible. All Sun-synchronous orbits are retrograde orbits.

Disadvantage.

- Retrograde orbits require more fuel as the rocket has to compensate for the Earth's rotation, which may be as much as 0.5 km s^{-1}, which is considerable compared with the final orbiting speed of about 7.9 km s^{-1}.

3.2.7.6 Polar or Near-polar If a satellite were to be launched with an inclination angle of 0° then it would orbit the Earth over the equator. In a low orbit this would mean that it could only see parts of the Earth along the equator. In order to see more of the Earth, the orbit would have to have a certain inclination. If the inclination were 60° then the satellite would fly over the Earth between latitude 60° south and 60° north; it would never see parts of the Earth at latitudes higher than 60°!

Ideally, for a satellite to see the entire Earth, it would have an inclination between 80 and 100°. Such an orbit, over the poles, is called polar or near-polar. A satellite in this orbit is sometimes called a polar orbiter.

Advantage.

- Eventually covers the entire Earth, also offering a good view of the poles.

3.2.7.7 Sun-synchronous An orbit chosen in such a way that the satellite always passes overhead at the same local solar time is called Sun-synchronous. Most orbits cross the equator, going from north to south (descending), at mid-morning, so images are taken when the Sun angle is low and the resultant shadows reveal terrain relief. Note that night-time images can be recorded during the ascending phase of the orbit.

Advantage.

- The Sun's position and lighting conditions for all the images are always the same (within certain limits). This makes it very easy to compare images from different dates.

Disadvantages.

- Varying Sun angles might help geologists to reveal subtle structural details that are not always visible at constant solar angles.
- It requires a retrograde orbit.
- In order to keep it Sun-synchronous, the subtle relationship between the orbital parameters must be preserved, which makes it more difficult to maintain.

Understanding the dynamics of the Sun-synchronous orbit is not so trivial. A constant orbit could never be Sun-synchronous. An orbit can only be Sun-synchronous if it precesses around the Earth at the same rotational speed (in degrees) as the Earth orbits around the Sun. In other words, the orbital plane rotates around the Earth at an angular rate of about 1° per day, in order to match exactly the angular rotation of the Earth around the Sun of 360° during the 365 days of 1 year (see Figure 6).

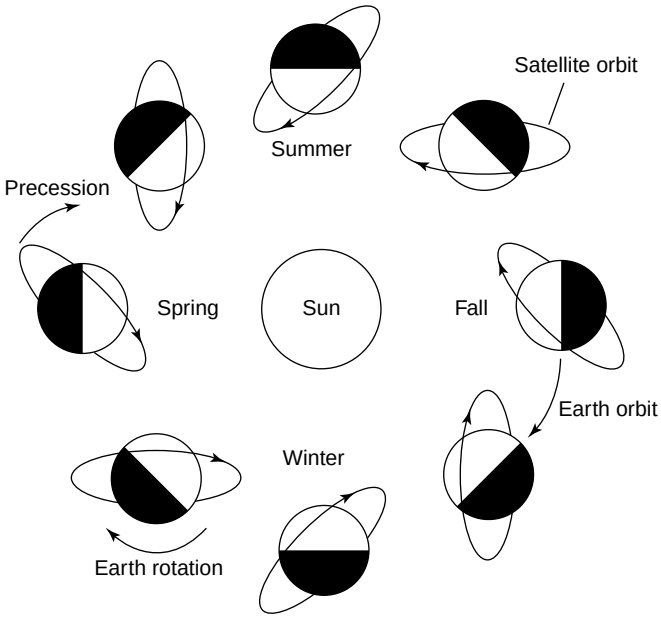

Figure 6 Precession of the satellite orbit. Sun-synchronous orbits have a precession period of exactly 1 year.

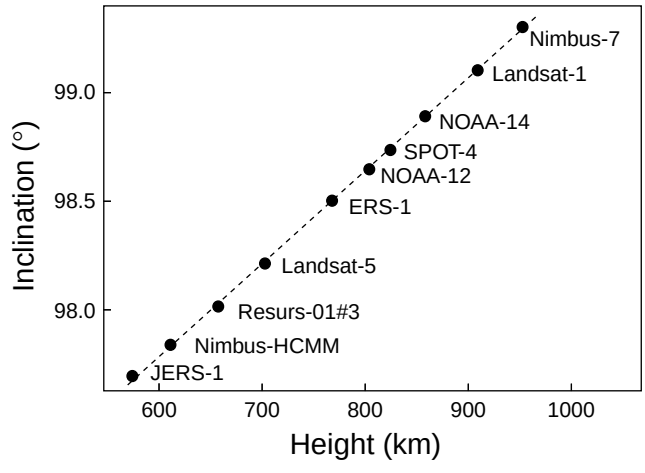

Figure 7 Inclination plotted against height for some Sun-synchronous satellites. Note that Sun-synchronous orbits are only possible on the dotted line.

All Sun-synchronous orbits are retrograde. The most important parameters for achieving a Sun-synchronous orbit are the height and the inclination. Most satellites are at 800 km height and 99° inclination. Figure 7 plots the relationship between the height and the inclination of Sun-synchronous orbits.

Most modern high-resolution Earth resources satellites are in a near-circular, near-polar, retrograde, Sun-synchronous orbit.

3.2.7.8 Geo-synchronous A geo-synchronous orbit has an orbital period that is an integer multiple or submultiple of the Earth's rotation rate, resulting in a repeating ground track. Very often the term geo-synchronous is confused with the term geo-stationary (see the next section). Every geo-stationary orbit (GEO) is geo-synchronous, but not every geo-synchronous orbit is geo-stationary. In fact, most RS satellites have a repeating ground track, but are not geo-stationary. The footprint of the satellite's orbit on Earth, its ''path'', is kept fixed (within certain limits), and maps can be obtained outlining these paths on the Earth's surface. Images of an entire path are chopped up at regular intervals, the ''rows''. Thus each scene or ''picture'' can be indicated by the ''path and row'' of its ''scene center''. For instance, the path and row grid of the Landsat satellites is called the World Reference System (WRS). The regular return of the satellite on a certain path is called the repeat cycle. The repeat cycle together with the pointing capability of the sensor determine how much time will pass before a sensor can generate another image of the same location. This is called the revisit time. Landsat 5 has a repeat cycle of 16 days and its sensor is not steerable, which means it has a revisit time of 16 days. The SPOT 4 satellite has a repeat cycle of 26 days, but its sensors can be directed to the left and the right, and thus achieve a revisit time of 4–6 days depending on the latitude of the scene.

Owing to perturbations, a satellite may slowly drift from its fixed path pattern. Small variations may be permitted before corrective action is taken by the mission control center.

Most RS satellites take daytime images during the descending phase of the orbit, which means that nighttime images can be recorded during the ascending phase of the orbit. Landsat, for instance, records thermal IR TM (a sensor on-board Landsat) images in the ascending phase. Radar systems record their images on only one side of the track. The possibility of also taking images at night (which for an active system does not really matter!) means that objects can be viewed from both sides. This opens up the possibility of doing stereoscopy with radar images. The science of measuring height and positions in stereo pair radar images is referred to as radargrammetry.

Advantage.

- The satellite follows a fixed pattern on the Earth, which makes reference of scenes by a path and row grid easy.

Disadvantage.

- As satellites slowly drift from their orbit, a geo-synchronous satellite must be corrected occasionally to keep it on the track of its predetermined paths.

3.2.7.9 Geo-stationary Orbit Arthur C. Clarke was the first to conceive the idea of stationary satellites, in

1945.[12] From their orbits 35 680 km above the equator, GEO satellites match the Earth's rotational period of about 24 h and thus remain stationary in relation to a selected point on the Earth's surface. They view about 45% of the globe, compared with the 3% of a spacecraft in a 320-km orbit. The view from GEO is ''Olympian'' compared with the ''worm's-eye'' view of ground-based observers.

Advantages.

- Continuous observations. A picture of a hemisphere's weather can be obtained every 15–30 min on an operational basis, and even every minute whenever necessary, e.g. in severe weather.
- From its vantage point in the sky, it views almost half of the globe at the same time.
- The satellite is always in view, which allows a real-time downlink for data.
- The position of the satellite, in fact, is more or less fixed (within 0.5° or so) in the sky, so it is very easy to point a fixed dish at it. Data transmission is easy, because the dish does not have to follow the satellite.

Disadvantages.

- Because of its height it requires a heavy rocket, which makes it the most expensive orbit to get into.
- Because of the distance to the Earth, the resolution of RS instruments from this orbit is low.
- Because of its position in the equatorial plane, it offers only an oblique view on the poles, which degrades the resolution at higher latitudes.

Modern meteorological weather satellite systems use a combined system of geo-stationary satellites and polar orbiters. The geo-stationary satellites offer a continuous view, whereas the polar orbiters offer a better resolution (see Figure 8).

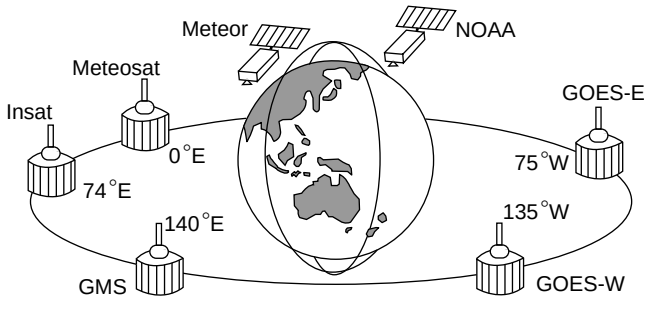

Figure 8 The global system of meteorological satellites; a combination of polar orbiters and geo-stationary satellites. (selection of systems only).

3.2.8 Other Orbits

As pointed out, the laws of Kepler are fine for a two-body problem, Sun–planet or Earth–satellite, but are useless for three or more bodies. For a satellite circling two massive bodies, it appears that there are five points where the pull of gravity of these two bodies is at equilibrium. These points are called Lagrangian or libration points L1–L5. At these points a satellite can be positioned at zero velocity with respect to both bodies. Three of these points are on the line passing through the centers of mass of the two massive bodies. L3 is beyond the most massive body, L2 beyond the least massive body, and L1 between the bodies. The other points L4 and L5 are located at equal distances, in an equilateral triangle, from the two main components in the orbit of the least massive object.

L4 and L5 are stable points, and a satellite positioned at one of these points, even after small perturbations, will always remain at this point. A satellite located at L1, L2, or L3 is unstable, and will eventually wander off if slightly disturbed. A thruster engine must be provided for small orbit corrections.

The L4 and L5 points of the Sun–Jupiter system are actually occupied by a number of asteroids referred to as the Trojans. The first Trojan, Achilles, was discovered in 1906 and now more than 50 are known.

The L1 point has already been used by a number of solar observation satellites, studying the Sun and the solar wind outside the Earth's magnetic field, for instance the Solar and Heliospheric Observatory (SOHO). In 2000 the Triana satellite will be placed at the Earth–Sun L1 point, about 1.5×10^6 km from the Earth, where it can maintain its fixed position between the Sun and the Earth and continuously see the full sunlit disk of the Earth. In the far future the L4 or L5 point of the Earth and Moon may be used for a space station.

4 REMOTE SENSORS

The terms instrument, sensor, and detector are not strictly defined, and are usually freely interchanged, although by detector is usually meant the energy-measuring element. By sensor we generally mean the detector plus optics. The term instrument mostly refers to the entire module, containing the sensor and electronics, as it is attached to the satellite platform. The platform, in turn, contains all sensor instruments plus additional instruments for maintaining orbit and attitude, communication, telemetry and command link, and more.

Remote sensors can be classified in many ways. The classification followed here first distinguishes passive sensors from active sensors, then classifies them by increasing complexity.

MS sensors use devices that scan the Earth and take measurements point by point rather than taking one picture in one shot, as in aerial photography. Either the point measurements are collected in an electromechanical way (with a moving mirror) or fully electronically (with no moving parts). This has the obvious disadvantage that the measurements are taken at slightly different times and may be difficult to relate to each other. However, MS sensors have some major advantages over photography:

- Photographic systems are limited to the spectral range 0.3–0.9 μm. MS systems have an extended range from 0.3 to roughly 14 μm. This range is usually referred to as the optical domain, which is the spectral range that can be manipulated by using optics (lenses). Some sensors may even operate in the microwave region of the electromagnetic spectrum, which roughly spans the wavelengths between 1 cm and 1 m; these waves can only be handled with electromagnetic fields.
- MS systems can record images in many very narrow spectral bands, which is referred to as imaging spectrometry.
- Photographic data are difficult to calibrate because of the photochemical processing involved. MS data are as stable as the electronics involved, which are much more stable than the photochemical processing. Calibration tries to reverse the measurement values back to the physical units of the observed phenomena.
- Electronic detectors also allow a much larger dynamic range, which is the operational range between the minimum and the maximum input that the detector can handle.
- Photographic systems require transportation of film to and from the platform. MS data allow electronic transmission to ground receiving stations, and can even offer near real-time viewing of the ground scene.
- MS data are readily available in digital form and do not need an additional scanning step for further electronic processing.

4.1 Passive Radiometry

Most RS systems use the indirect solar energy reflected by the object as their prime source. The system does not actively probe the object with some form of energy, hence the name ''passive sensors''. Another source of energy may be radiation emitted by the objects themselves. All objects having a temperature above 0 K emit energy. At 300 K, the top of the blackbody curve is located in the middle of the TIR region of the spectrum. Objects may also emit other forms of energy. Well known are the images of city lights from space taken by the Defense Meteorological Satellite Program (DMSP) satellites. The same satellite has also taken images from the Aurora Polaris.

A detector is a device that transduces incident energy into a signal. Typically, a detector needs a certain amount of time to build up a signal. This time is called the integration time, or dwell time. The signal from the detector will be proportional to the aperture of the detector and the integration time. Preferably, there is a linear relationship between the energy falling on and the signal coming from the detector.

Detectors for the IR range must be cooled in order to reduce stray radiation from the sensor itself to be detected.

4.1.1 Aerial Photography

Photography has long been the only means of recording images by RS. It is still being used intensively on-board aircraft, and there are a number of spaceborne cameras. For instance, the Russian KVR-1000 camera flown on-board a Cosmos satellite delivers images for Space Information – 2 m resolution (SPIN-2). The images are available on the Internet.[13]

Only a few aspects of photography will be treated here. Other sources such as the *Manual of Remote Sensing*[14] and Lillesand[2] are excellent sources for further reading.

The following elements can be distinguished in a camera:

- the camera body and film magazine
- the optical system, comprised of a number of lenses
- the film
- the diaphragm
- the shutter
- the filter
- the exposure control system
- an auxiliary data system.

A camera projects the scene on to a film in a so-called central projection. A displacement in height of the object will cause a displacement in the image plane. Although this is bad for topographic mapping, because objects are not imaged at the expected location in the image, these small displacements can actually be used to measure the height of the object. Using two or more images taken from different positions does this, and is called stereoscopy. Photogrammetry is the science of measuring positions and heights in aerial photographs.[15] Distortions in the optical system (aberrations) and stretch of the film determine the geometric accuracy of the image.

The grain size of the film, the quality of the optics, the aperture, and other factors determine the spatial resolution of a camera system. The optical quality of a photographic system is usually determined by constructing a modulation transfer function (MTF). The

MTF plots the relative modulation depth of the response of a square-wave test pattern offered to the camera. However, although the MTF gives a good indication of the resolution of the system, one should remember that the final resolution of a photograph depends on many factors.

The spectral characteristics of a photographic system are determined by the sensitivity of the film and the filters used. Normal emulsion is sensitive only to blue light. The film can be sensitized to have an extended spectral sensitivity into the green (orthochromatic emulsion), red (panchromatic (PAN) emulsion) and IR (IR-sensitive emulsion) regions. Colored-glass or coated-glass filters can be used to limit the sensitivity to a certain spectral range.

Advantages.

- large image format
- high information density (one aerial photograph has an equivalent information content of about 600 Mb)
- high geometric integrity.

Disadvantages.

- transmission of film
- radiometry is difficult to control
- limited spectral coverage in the VNIR range
- analog output.

4.1.2 Cathode-ray Tubes

Although the early systems carried the equivalent of television tubes (TIROS-1, Landsat 1), these are hardly used now because of the advantages of solid-state detector systems. Sometimes modern video equipment is carried on-board aircraft for recording the state of the environment.

4.1.3 Solid-state Detectors

Electro-optical systems as radiation detectors are transducers that transform electromagnetic radiation into electrons or other detectable signals to convey the measurement information of the viewed scene. The description of types of detector materials, such as photoconductors, photodiodes, and charge-coupled devices (CCD) can be found, for instance, in the *Manual of Remote Sensing*.[14] Here, we will mention only a selection of general properties of detectors.

The whole range of input intensity values, between a maximum and a minimum signal that a detector can handle, is called the dynamic range. The ratio between a change in output value as a result of the changing input value is referred to as sensitivity, gain factor, or gamma. An ideal detector has the same sensitivity over the dynamic range, i.e. the gain factor is independent of the input level. Such a detector is said to be linear. The output of a digital detector is quantized into a limited number of levels by an analog-to-digital converter (ADC). The higher the number of levels in the dynamic range, the more intensity values in the input can be distinguished. The smallest input level that can still be distinguished in input is called the radiometric resolution; more levels improve the radiometric resolution. Usually, the smallest level that can be resolved in input is determined by the noise in the system. The more noise in the detector system, the more the radiometric levels will become blurred and the less the radiometric resolution will be.

Each detector has a useful sensitivity in its own spectral range. The graph that plots the sensitivity of the sensor for each wavelength is called the spectral response function (SRF). A filter can be applied for narrowing the spectral range of a sensor. The position of the central wavelength of the SRF determines the spectral band of the sensor. The width of the SRF determines the spectral bandwidth. The bandwidth is usually determined by the difference of the two wavelengths where the SRF is half as high as its maximum peak value. This is called the full width at half-maximum (fwhm) value.

Preferably all the characteristics of the sensor are known, and they can be determined by calibration. Calibration is the characterization of a sensor in the spatial, spectral, and radiometric domains. The process of calibration involves standard or invariant objects, and can be done either before or after the sensor is taken into operation. Ideally, calibration makes the measurements independent of the instrument. This means that when a particular physical entity is to be measured at different times and places or with different instruments, the results should always be the same. The ultimate measurements should carry only object attributes, not sensor attributes. Unfortunately, the nature of the interactions of electromagnetic waves with the object, the atmosphere, and the sensor is highly complex, which makes it almost impossible to obtain the ideal calibration for all practical situations.[8]

Apart from that, sensors appear to change over time, which makes frequent recalibration essential. Several internal or external sources can be used for in-flight calibration. If the sensor is sufficiently linear by nature, a dark and a bright (or cold and hot) source are sufficient to construct a new calibration curve. Still, the calibration method itself may degrade over time. A constant external source may not be so constant after all. A diffuser, used to soften the light of the Sun directed on to the sensor for calibration purposes, may degrade over time, necessitating a second, less frequently used diffuser. For instance, the Medium Resolution Imaging Spectrometer (MERIS) sensor on-board the environmental satellite (Envisat) will apply the latter technique. Of course, this

calibration of the calibrator can be carried out with endless recursion. The fact is that the calibration and validation of remote sensors are attracting more and more attention and some organizations execute regular underflights for comparison of airborne and spaceborne data.

The calibration data can then be used to determine the physical properties of the object under study (radiance, reflectance, temperature, etc.) from the digital number (DN) values.

4.1.4 The Sounder

The simplest sensor is composed of one straight-downward looking detector. Such a detector would take measurements of small ground areas of the Earth. During its orbit the satellite can measure the energy from a number of points along its path on the Earth. Each measurement is taken at discrete distances on the Earth, and is called a sample. The distance between the samples is called the sample distance. The radius of the area covered by one such measurement is called the resolution. The DN expresses the value of the measurement and is called the DN value, the radiometric value, or simply the ''value''.

The sample distance and the resolution together determine how much spatial detail can be observed in the recorded measurements. The sensor could be instructed to take more than one sample within the resolution of the detector. However, this would not make any sense, as the detector would measure most of the ground area of the previous sample! In this case, the detail that one can observe is limited by the resolution. On the other hand, a sensor built with a resolution much smaller than the sample distance would not make much sense either, as the sample distance limits the amount of detail in this case. Besides, the sensor would not even measure much of the total area at all! This is exactly the reason why most sensors are designed in such a way that the sample distance more or less equals the resolution. Hence sample distance and resolution are often freely interchanged. The problem with this is that this leads to thorough confusion (even among RS experts!) about the very different natures of sampling and resolution. Sampling, when and where the next measurement is taken, is determined by the electronics and the orbit, whereas resolution, the extent of one measurement, is determined by the angular aperture of the detector and the quality of the optics. The angular aperture is called the instantaneous field of view (IFOV).

In order to build up an image of the Earth, many measurements, or samples, must be recorded. By taking samples along the path of the satellite and combining these with sample of adjacent paths, a two-dimensional matrix of measurements can be built up. Thus, the satellite image is born.

Note: suppose the speed of a (hypothetical) satellite's footprint on the Earth is $6\,\mathrm{km\,s^{-1}}$ and that we want to take samples every 1 km. In order to achieve the sample distance of 1 km along the track, the sensor must take six samples every second. In other words, the detector has one-sixth of a second to take one measurement. In order to achieve the sample distance of 1 km across the track, the pattern of paths of the satellite's successive polar orbits must be spaced at 1 km at the equator. The circumference of the Earth is about 40 000 km, which means that the satellite completes about 13 orbits per day (one day contains 86 400 s; during one day the satellite travels $6\,\mathrm{km\,s^{-1}} \times 86\,400\,\mathrm{s}$, which is 518 400 km or 40 000 km per orbit around the Earth.) It takes the satellite 6670 s (40 000 km/$6\,\mathrm{km\,s^{-1}}$) to complete one orbit. During the time the satellite rotates around the Earth, the Earth itself rotates around its axis. The rotational speed of the Earth is about $1670\,\mathrm{km\,h^{-1}}$ (40 000 km/24 h) on the equator, which means that the Earth has moved 3086 km at the next equator crossing of the satellite. This in turn means that the next orbit is not adjacent to the previous orbit! We will have to wait longer for the satellite to visit adjacent paths. However, this poses a problem: as each orbit lasts more than 1 h (6670 s = 1.9 h), things might change on Earth. The Sun-angle changes, the cloud cover (CC) may change, or the phenomena under investigation may change. This is unacceptable. Another problem is that it would take 40 000 orbits (to achieve a 1-km sample distance on the equator) to acquire a global image. This would need more than 8 years! (40 000 orbits/13 orbits/day for 365 days/year = 8.4 years).

Although a sounder has the advantage that it always looks straight down, which means the measurements are independent of the height of the object, it is not a practical solution for taking high-resolution images of the Earth. An obvious solution would be to take multiple measurements across the track of the satellite. Both the scanner and the pushbroom sensor do exactly that. They can record an image in a short time, and the pattern of their paths on the Earth does not have to be so dense.

4.1.4.1 Microwave Sounders Microwave sounders are used for measurements to map surface temperatures, subsurface structures, or atmospheric profiles of temperature, composition, and pressure as a function of altitude.[8] Sounders use the emission of objects themselves as a source of energy. The microwave region is in the low-energy tail of the 300 K blackbody radiation curve (see paragraph 3.1.3). In this spectral region, all objects in the natural environment emit microwave radiation, albeit faintly. Because of the extremely weak magnitude, the signal is very noisy compared with that which cameras, visible or IR scanners, or radar provide. Common to all radiometer designs is the trade-off between antenna beam width and system sensitivity. Because of

the very low levels of radiation available to be passively sensed in the microwave region, a comparatively large antenna beam width is required to collect enough energy to yield a detectable signal. Consequently, passive microwave radiometers are characterized by low spatial resolution.(2)

Because microwave sounders operate on the same energy curve as TIR sensors, images from sounders bear a strong resemblance to TIR images, although transmission of the top layer may unveil microwave emission from subsurface structures.

4.1.5 The Scanner

A combination of a single detector plus a rotating mirror can be arranged in such a way that the detector ''beam'' will sweep a straight line of the Earth across the track of the satellite at each rotation of the mirror. The forward motion of the satellite takes care of positioning the satellite above successive lines (see Figure 9). Thus the scanner is born. Because of the sweeping motion, the across-track scanner is sometimes referred to as a whiskbroom.

The aperture or opening angle determines the tiny area that the detector sees at one glance of the Earth (the ground cell), and is called the IFOV. The IFOV is constant over the entire line, and this means that the resolution degrades at increasing scan angles left and right of the satellite track. The along-track resolution decreases with $1/\cos\theta$ and the across-track resolution decreases with $1/\cos^2\theta$, where θ is the scan angle. Clearly, the fact that the scanner measures ground cells at constant angles causes geometric distortions. The curvature of the Earth causes additional distortions, especially for a large field of view (FOV).

An extreme example of this is the advanced very-high resolution radiometer (AVHRR) sensor, which has a FOV of 110°, where the resolution degrades from approximately 1 by 1 km at nadir to 6 by 3 km at the limb of the sensor.

The swath pattern and orbital path are designed in such a way that images of adjacent paths have a degree of lateral overlap. The amount of overlap is minimal at the equator, but the side overlap increases towards the poles.

Advantages.

- simple detector
- simple optics
- wide FOV capability
- extended spectral sensitivity (compared with a CCD).

Disadvantages.

- short dwell time
- moving parts
- lower geometric stability.

Note: consider the same (hypothetical) satellite as mentioned above, but now equipped with a scanner instead of a sounder. Again, we want to have a sample spacing of 1 km. Furthermore, suppose we want to record 2000 samples per line, which gives the sensor a swath on Earth of 2000 km (not considering distortions). The speed of the satellite of 6 km s^{-1} has not changed, because it is in the same orbit. This means that in order to achieve a line spacing of 1 km the sensor has to record six lines per second, which in turn means the detector has to record 2000 samples per second! This shows that the dwell time of the scanner has decreased dramatically by a factor of 2000 compared with the dwell time of the sounder (not considering the dead time of the rotating mirror). Obviously, there is a trade-off between the number of samples per line and the time the detector has to take one sample. A major advantage is that global coverage can be achieved within 20 orbits (40 000 km Earth radius/2000 km swath), which is less than 2 days.

The dwell time or integration time is the time needed by the detector to build up one measurement. At the expense of a much shorter dwell time, the sensor can record large swaths at the same time, and global coverage can be achieved within a reasonable time. If the dwell time becomes too short, the noise levels of the sensor increase. In order to give the detector longer dwell times, multiple detectors can work in parallel. For instance, the TM sensor of the Landsat satellite uses 16 detectors in parallel per spectral band. In addition, the TM uses an oscillating mirror rather than a rotating mirror, and sweeps the Earth in both left–right and right–left motions, which reduces the dead time of the mirror almost to zero. By these measures the TM achieves an acceptable dwell time for its detectors.

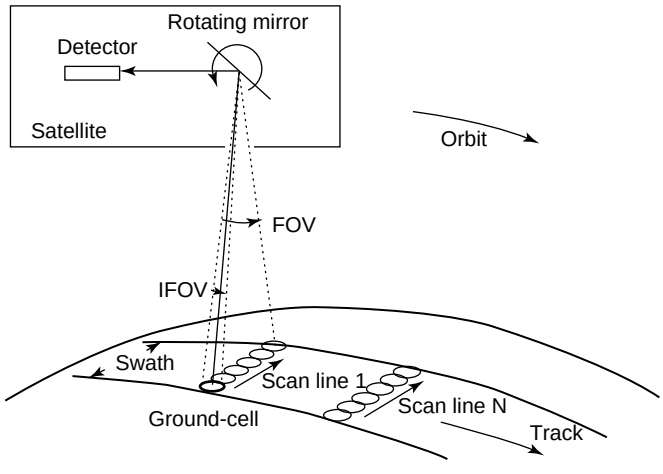

Figure 9 Principle of the whiskbroom scanner.

The use of parallel detectors comes at a cost. Instead of having to calibrate one detector, many detectors must be checked. Differences between the detectors are visible in the recorded images. TM images may suffer from horizontal banding caused by differences in sensitivity of its 16 detectors per band. The banding pattern will repeat itself after 16 lines (see also Figure 14). This banding is called ''striping''. Also, a single detector sensor may exhibit striping, as the characteristics of the detector may be influenced by the received energy (e.g. overload). If a detector saturates on a very bright object, it may give a faulty response on the next few pixels and lines.

The ultimate step in parallelism would be to record all samples across the scan line with an individual detector. This would also remove the need for a moving mirror! Such an arrangement of an array of detectors recording all samples of a line in parallel is referred to as a pushbroom.

4.1.6 The Conical Scanner

The conical scanner also uses a rotating mirror, but in a different geometrical arrangement. The detector sweeps the ground in circles (see Figure 10). The advantage of this is that the angle of incidence is the same for all objects, and also the path through the atmosphere is a constant. A disadvantage is the extreme geometrical distortion caused by the cycloid scanning pattern. In fact, the same pattern can also be achieved by a rotating satellite, which requires no moving mirror. Many passive microwave atmospheric sounders use conical scanners. One MS system to use this principle is the MSU-SK sensor carried on-board the Russian Resurs-O1 satellites.

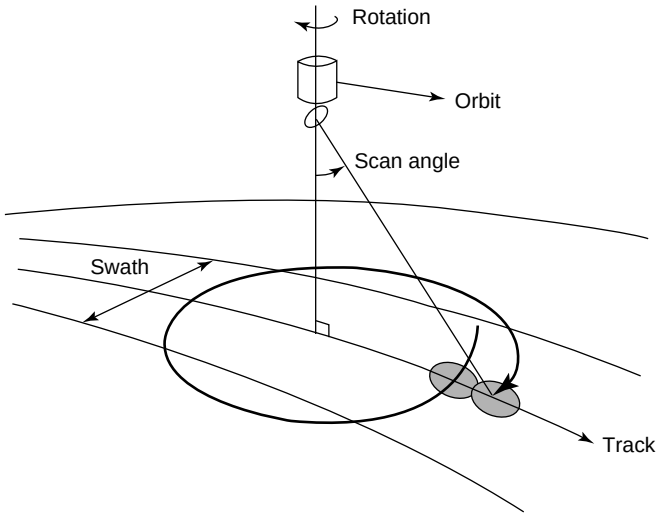

Figure 10 Principle of the conical scanner (after Lillesand and Kiefer.[2])

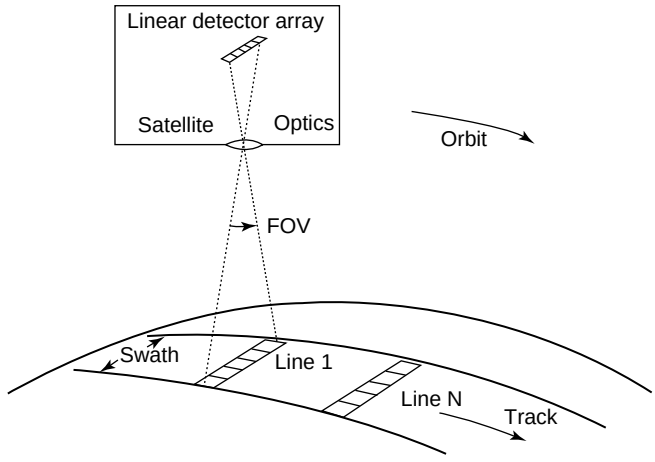

Figure 11 Principle of the pushbroom sensor. A linear array of detectors is used to image an entire line simultaneously.

4.1.7 The Pushbroom

Modern pushbroom sensors use linear array CCDs (see Figure 11), which are composed of thousands of equally spaced detectors. They act as a linear camera, recording one entire line at the time. The pushbroom requires no moving parts, as opposed to the scanner, and offers better signal levels because of the increased dwell time.

Also, the geometric distortion of a pushbroom is less than that of a scanner, because the central projection of each line causes no distortions in vertical images. If the terrain is flat there is a one-to-one correspondence with the position of the object and the position of the detector. Actually, the opening angle of each detector varies with its position in the array, and is not a constant like the IFOV of the scanner. Therefore, it is incorrect to speak of the IFOV of a pushbroom.

Note: again, considering the hypothetical satellite mentioned above, by using a pushbroom, each individual detector can take as much one-sixth of a second to form one measurement of one ground cell.

Advantages.

- The pushbroom has no moving parts, which gives it a longer life expectancy.
- All elements of the pushbroom are at fixed distances from each other. This means that recorded data will not suffer from geometric distortions caused by irregular movements of the scan mirror.
- No IFOV.
- Long dwell time.

Disadvantage.

- State-of-the-art CCDs have a limited spectral sensitivity up to about 2.5 μm.
- Needs wide FOV optics.

Also with the pushbroom the parallelism comes at a cost. Although one might expect all the detectors of the linear CCD to have similar characteristics, there is a limit to the control in the production process of CCDs. Each element will have its own properties, and instead of having to calibrate one, or just a few, detectors, one may have to calibrate thousands of detectors of the linear CCD. Differences between detectors are visible in the recorded images. As opposed to the horizontal striping in scanner images, pushbroom images may suffer from vertical banding.

4.1.8 The Two-dimensional Charge-coupled Device

The two-dimensional version of the linear CCD chip can be used as a replacement of the film in a camera. The advantage of using an electronic detector over film is that the data are immediately available in digital form. The process of taking snapshots of the Earth along the orbit of the satellite is referred to as step staring. A geo-stationary satellite can apply the two-dimensional CCD as a video camera pointed at the Earth; this is referred to as fully staring. Some heat-sensitive geo-stationary missile detectors seem to apply this technique. The TRMM satellite carries a lightning imaging sensor (LIS), an optical staring imaging system to detect the rate, position, and radiant energy of lightning flashes.

4.1.9 The Imaging Spectrometer

A modern and popular design of the imaging spectrometer is the pushbroom using a dispersive element together with a two-dimensional CCD chip. One line of the Earth is projected on to a dispersive element (grating or prism), which separates the line image into a series of lines, each corresponding to a spectral band. In this way each column of the chip records the spectrum of one ground cell. Each row images each ground cell in one particular spectral band. The whole process is repeated as the platform moves one line along the track. Imaging spectrometers (or hyperspectral imagers) allow the simultaneous acquisition of images in many, contiguous, narrow bands. Traditional remote sensors have few and broad bands, whereas imaging spectrometers may have hundreds of narrow bands. They allow the recording of spectra of each ground cell much like using a spectrometer in a laboratory. One major problem of using an imaging spectrometer is the high volume of the data. Some existing airborne systems allow averaging of rows or columns on the two-dimensional CCD chip, thereby offering the possibility of exchanging spatial resolution for spectral resolution and vice versa (for instance the CASI instrument).

Some instruments are imaging spectrometers by design, but deliver only a limited number of bands from the whole range of possibilities. The advantage of this is that the band position and width can be reprogrammed at a later stage. For instance, the MERIS instrument of the Envisat satellite delivers only 15 bands, while it is actually a fully fledged imaging spectrometer. MERIS's spectral resolution is actually very high, and it allows the bands to be reprogrammed at wavelength steps of 1.25 nm!

4.2 Active Radiometry

Unlike in airborne radar observation, where relatively simple side-looking airborne radio detection and ranging (SLAR) can be carried, spaceborne RS requires a different technique. Spaceborne radar would require an unacceptably long antenna in order to achieve the same resolutions as airborne radar. Instead, a long antenna in space is synthesized by combining signals from different positions along the orbit of the satellite. Such a radar system is called synthetic aperture radar (SAR). Building such radars demands the utmost of electronics, signal processing, electromagnetic fields, and antenna design. Also, the interpretation of radar images is difficult, because the return signal from the object can be influenced by many factors. For instance, the reflected intensity from trees is affected by the leaves, the branches, the trunk, the shape and structure of the branches, the surface below the tree, the surface roughness of the soil below, the humidity of the soil, and subsurface structures under the soil. Radar can be used for determining the amount of vegetation on the ground, but then the other parameters, such as surface roughness and soil humidity, must be known.

Imaging radars have been flown on a number of spacecraft, including Seasat, the Japanese Earth Resources Satellite (JERS) 1, ERS 1 and 2, Radarsat 1, and the Space Shuttle. Shuttle Imaging Radar C (SIR-C) is currently the most advanced imaging SAR, capable of shooting radar images at three different microwave frequencies and multipolarization mode. With this system, up to nine different radar images can be obtained from an object simultaneously, offering an unprecedented view of the object's composition and structure. Future plans for the Shuttle include flying a second radar system connected with a 60-m boom to the first in order to acquire, for the first time, simultaneous images for interferometric SAR.

4.2.1 Altimetry

Radars are also used in a nonimaging mode as altimeters. Altimeters use the ranging capability of radar sensors to measure the surface topographic profile, by simply emitting a pulse and measuring the time between emission

and reception. The height h bears a linear relationship to the time lapse t (Equation 17):[7]

$$h = \frac{ct}{2} \tag{17}$$

Of course, the pulse travels at speed c through space. Altimeters have been flown on a number of spacecraft, including Seasat, ERS 1 and 2, and TOPEX/Poseidon. The water height of the oceans is influenced by a number of factors. Seasat is known for the first images of the topography of the ocean floor. A dip in the ocean floor causes a dip in the water surface, because there is less gravitational pull from the Earth's crust. The water level can vary by roughly 150 m, caused by the ocean floor topography. Knowing the mean water level of the oceans, minor variations on top of the mean level can be measured. TOPEX/Poseidon is known for measuring the ocean temperature by these changes. There appears to be a direct relationship between temperature and water height. The sea surface temperature reflects the temperature in the top few centimeters of water, and the temperature can change dramatically with depth in some cases. The sea surface measured by altimetry is related to the temperature at *all* depths, in addition to other parameters, such as the water salinity and ocean currents. Note that, in general, the movement of currents has a greater effect on sea level (± 1 m) than heating and/or winds (± 12 cm).[16]

All spaceborne radar altimeters to date have been wide-beam systems, limited in accuracy by their pulse duration. Such altimeters are useful for smooth surfaces (oceans), but are ineffective over relatively high relief continental terrain. A fundamental problem in narrow-beam spaceborne radar altimetry is the physical constraint of antenna size. Again, large antennas are required for small radar footprints.[8]

4.2.2 Scatterometers/Spectrometers

Scatterometers are radar sensors that provide the backscattering cross-section of the surface area illuminated by the sensor area. They are particularly useful in measuring the ocean backscatter as a means of deriving the near-surface wind vector. The strength of the radar backscatter is proportional to the amplitude of the water surface capillary and small gravity waves (Bragg scattering), which in turn is related to the wind speed and direction near the surface.[7]

4.2.3 Imaging Synthetic Aperture Radar

Imaging radar data are being used in a variety of applications, including geological mapping, ocean surface observation, polar ice tracking, and vegetation monitoring. Most of these applications are still in the development stage, because our knowledge is still limited on how to extract information from radar image data and how to use the various possibilities of radar parameters, such as frequency, polarization, incidence angle, resolution, and swath width. Qualitatively, radar images can be interpreted in very much the same way as optical images. In general, one could say that radar reveals more of the structure of an object, and optical images show more of the composition.

Man-made objects usually stand out very bright in radar images, because many objects act as corner reflectors for the radar signal. Houses, bridges, and other constructions can easily be found.

4.2.3.1 Speckle The first impression of radar images is that they have a very noisy character. This noisy appearance results from the backscatter contributions of the many scattering points of the surface area, even within one ground-cell. Even though its properties do not change, the same ground-cell may look completely different from different positions (an image from exactly the same position will give exactly the same measurement). This type of ''noise'' is related to location rather than time. The noise is caused by the coherent scatterers and not by bad signals by the radar system, and is called speckle. The way to improve the noisy appearance of radar imagery is to take a number of images and average them. This multilook technique will reduce the speckle by the square root of the number of ''looks''.

4.2.3.2 Pulse-time Measurement Actual imaging SARs use a series of pulses instead of a continuous signal. Each transmitted wave front hits the target surface at near range and sweeps across the swath to the far range. The return pulses are timed and recorded. The signals of each swath are said to be in the slant range; the objects are in ground range. The transformation of signals in the slant range to the ground range is not a trivial one and involves the incidence angle and the height of the object, and sometimes cannot be solved at all because of overlapping signals (see below), i.e. signals from different objects arriving at the same time. This problem of signals arriving at the same time is also the reason why all imaging radars are side-looking. This removes the left–right ambiguity of signals from objects at the right and the left side of the radar.

The pulse bandwidth B determines the cross-track resolution (Equation 18):

$$C = \frac{c}{2B \sin\theta} \tag{18}$$

where c = speed of light, B = pulse bandwidth, and θ = incidence angle).[8] The presence of the incidence

angle θ in the equation is the reason for varying resolution and scale across the slant-range of a radar image.

The along-track resolution A of a real-aperture radar is inversely proportional to the antenna length L (Equation 19):

$$A = \frac{h\lambda}{L\cos\theta} \tag{19}$$

where h = height above ground, λ = wavelength, L = antenna length, and θ = incidence angle.[8]

However, interestingly enough, theoretically the along-track resolution A of a SAR is proportional to the antenna length L (Equation 20):

$$A = \frac{L}{2} \tag{20}$$

In theory, the resolution for a 1-m antenna would be 0.5 m, independent of the height! However, since short antennas have stability problems, radar design is replete with tradeoffs among operating range, resolution, wavelength, antenna size, and overall system complexity.[2]

4.2.3.3 Synthetic Aperture Radar Polarimetry The roughness and composition of materials may cause a change in the polarization angle of the return signal with respect to the incidence angle. This change in polarization thus offers yet another possibility to characterize materials on the surface of the Earth. Current operational systems typically only record the like-polarization modes, either transmitting horizontally polarized signals and receiving horizontally polarized signals (HH polarization), or transmitting vertically polarized signals and receiving vertically polarized signals (VV polarization). However, experimental systems have already proven the usefulness of the other polarization modes. Future radar satellites will also be capable of recording the cross-polarization modes, horizontal transmission–vertical reception (HV), and vertical transmission–horizontal reception (VH).

4.2.3.4 Synthetic Aperture Radar Interferometry Recently radar RS has proven very valuable with the advent of interferometry. The use of SAR as an interferometer is rather new owing to the stringent requirements as regards the stability of the satellite orbit. INSAR offers the possibility to map the Earth's land and ice topography and to measure small displacements over large temporal and spatial scales with subcentimeter accuracy. INSAR makes use mainly of the phase measurements in two or more SAR signals of the same scene, acquired at two different moments or at two slightly different locations. By interference of the two images, very small slant-range changes can be related to topography and surface deformations. INSAR thus has the potential for mapping centimeter-scale ground displacements over a region of many tens of kilometers in size at a resolution of a few meters, making it one of the most promising space geodetic techniques for monitoring Earth's surface deformations.

4.2.4 Precipitation Radar

Radar can also be applied to measure the amount of rainfall. The TRMM provides the first detailed monitoring of tropical rainfall. The objective of the TRMM is to obtain three-dimensional rainfall distribution over land and oceans in the tropical belt between latitude 35° north and 35° south. Two-thirds of global rain falls within this tropical belt, and knowledge of climate processes here is important for understanding climate processes and changes on a global scale.

In addition to the precipitation radar, TRMM carries instruments for measuring radiance from clouds and a lightning sensor.

4.2.5 Light Detection and Ranging

Light detection and ranging (lidar) refers to laser-based RS. Lidar uses principles very similar to those used in radar. The main difference is the difference in wavelength: radar uses microwaves whereas lidar uses visible light. In fact, the term lidar is a generic term used for a variety of sensors operating on different concepts. Going into detail for each of them would be beyond the scope of this article. For more information we refer to, for instance, Kramer.[8] The main applications are measuring distance (height), movement, and chemical composition. By measuring the intensity, polarization, and spectral properties of the return signal as a function of time, one can obtain information on the properties of the atmosphere.

Light Detection and Ranging Applications[8]

- Atmospheric research: meteorology, climatology, boundary layer physics, pollution, visibility, radiative budget, atmospheric dynamics and chemistry (composition).
- Solid-Earth research: crustal movement, gravity field, Earth kinematics and terrain mapping [digital terrain/elevation model (DTM/DEM)].
- Precise orbit determination.

Lidars are active instruments and can operate day and night. A disadvantage is that lidars cannot penetrate through optically thick layers (clouds).

5 PLATFORMS FOR REMOTE SENSING

Soon after the invention of photography, cameras were taken into the sky and aerial photography was born. Even today cameras and sensors are carried on anything that can observe the Earth at altitudes ranging from 1 m to about 30 km. Airborne RS is still intensively used, and will probably never disappear altogether, even though one might be inclined to think so in this modern age of space flight and technology. The reason for this is that airborne RS and satellite RS complement each other rather than compete with each other. Also, although this article deals mainly with satellite RS, it will be interesting to list a number of properties of both airborne and satellite RS next to each other.

5.1 Aircraft

Above 300 m we enter the realm of the aircraft. Virtually every type of existing aeroplane has been used, at some time or other, for the purpose of Earth observation. There is no such thing as the universal plane for RS; each plane has its own specific advantages and disadvantages. In the early days, the main problem was to fly high enough in order to cover large enough areas. Satellite images became available at the beginning of the 1970s, and a new niche was found: the highly specialized pinpointed analysis of areas within satellite coverage. The demand widened to require a new generation of slow-flying aircraft at low altitudes providing specialized cover for individual customers, and often operated by the users themselves.

In recent years, the ''minimum'' aircraft, or micro-light, has entered the field as a slow-flying camera platform. These new aircraft continue to surprise us, and several have been designed recently with work rather than sport in mind. They may yet gain our esteem in this respect.

Small, single-engined light aircraft have long been the workhorses in oblique and vertical aerial photography. The altitude of these planes is usually below 3000 m, and limited to 4500 m, flying for single projects.

The next step up is the light twin-engined aircraft, operating at altitudes ranging from 900 to 6000 m for the piston-engined twins to 7500 m for the turbo-charged piston-engined aircraft. It offers larger carrying capabilities for payloads, faster speeds, better attitude control, offering a stable platform, better climbing performance, and better safety because of the second engine. Although there is some overlap with the earlier mentioned aircraft, this comes at the expense of higher investment and maintenance costs.

Pressurized turbo-prop aircraft are operated at altitudes from 4500 to around 10 000 m. Although high in capital cost, the scarcity or expense of aviation gasoline in many countries makes the turbo-prop, which uses kerosene as fuel, attractive as a survey aircraft. Airports everywhere are now geared up to jet aircraft and keep large quantities of fuel on hand for large airliners, making life easier for the turbo-prop than for the piston-engined aircraft.[17]

Above 9000 m, the pure jet aircraft is the only platform available. The maximum altitude of these aircraft is in the region of 16 000 m.[17] The top of the league is held by special reconnaissance aircraft, such as the beautifully sinister Lockheed SR71 Blackbird, capable of flying at altitudes in excess of 26 km!

Next in line, although not really an aircraft in the strict sense, is the Space Shuttle, operating at heights between 200 and 300 km. Several cameras, electronic scanners, and radar systems have flown on-board the Space Shuttle.

The main emphasis in the past has been aerial photography. These days, however, it is more often on electronic sensing than on photography.

5.2 Drones

These aircraft, which look and act like outsized model aeroplanes, have been around for some time. Drones, also called remotely piloted vehicles or aircraft or unmanned aerial vehicles (UAV), are currently very much under development.

Drones have several advantages over ordinary aeroplanes:

- They are easy to deploy, control and recover.
- They are becoming cheaper to build as the state of the art improves and off-the-shelf technology becomes the norm.
- They are small and lightweight, and can be easily transported to any location.
- They require no pilot on-board and can fly, for instance, over enemy territory without fear of loss of life, or they can fly very high without the need for a pressurized cabin.
- They can fly autonomously, and be programmed to fly predetermined flight paths, using satellite navigation systems.
- They can fly nonstop, sometimes even for days.

NASA is has its experimental Pathfinder and Centurion planes in its Environmental Research Aircraft and Sensor Technology (ERAST) program. The planes are powered by solar energy. NASA wants those planes to go ''slower, higher, and longer''. Eventually these planes will fly as slowly as 24 km h^{-1}, at altitudes of 30 km, for missions lasting as long as 4 days, carrying an instrument payload of about 90 kg.

The military also have their UAV programs for airborne reconnaissance using planes with catchy names such as Hunter, Predator, Global Hawk, and Dark Star.

Already the communications world has shown interest in this type of plane. Such a plane could circle a large city, serving as a relay station for communication channels. As such it would be a compromise between an Earth-bound network of transmitters and a satellite communication system.

The drones, together with small-satellite technology, have a great future for the communications and Earth observation businesses.

5.3 (Stratospheric) Balloons

Balloons have been used for decades to conduct scientific studies. While the basics of ballooning have not changed, the balloon size has increased and their reliability has improved greatly. Stratospheric balloons are ideal for making observations from the ''edge of the atmosphere''. They go as high as 40 km, where it is too high for any plane and too low for any satellite. Balloons are made of thin 20-μm polyethylene material, about the same thickness as ordinary sandwich wrap. The balloon system includes a balloon, a parachute, and the payload that contains the instruments to conduct the experiment. Balloon payloads provide us with information on the atmosphere, the universe, the Sun and the near-Earth and space environment.

A balloon flight mission is relatively simple. The balloon is partially filled with helium and launched with the payload suspended beneath it. As the balloon rises, the helium expands and fills out the balloon until it reaches its peak altitude 2–3 h after launch. A major disadvantage of balloons is that they cannot be steered, the path being determined by wind direction only.

Unique Features of Balloons[18]

- Balloons are inexpensive.
- Balloons offer a quick response method for performing scientific investigations.
- They can be readied for flight in as little as 6 months.
- Balloons are mobile, meaning that they can be launched where the scientist needs to conduct the experiment.
- A balloon can carry a payload weighing as much as 4000 kg.
- They can fly to an altitude of 42 km.
- Flights last an average of 12–24 h.
- Some special-purpose, long-duration balloon flights have lasted more than 2 weeks.

5.4 Sounding Rockets

The space age for RS received its first boost in 1891, when a patent was granted to Ludwig Rahrmann of Germany for a ''New or Improved Apparatus for Obtaining Bird's Eye Photographic Views''. The apparatus was a rocket-propelled camera system that was recovered by parachute. Sounding rockets are still being used to ''fill the gap'' between balloons and satellites. Typically they will be launched up to 250 km height and make a fall back to Earth by parachute, in the meantime taking measurements. The overall time in space is brief, typically 5–20 min. Sounding rockets are mainly used for studying the upper atmosphere chemistry. They provide the only platform with which direct in situ measurements can be made in these regions.

Furthermore, suborbital flights play a crucial role in the development and testing of instruments that ultimately lead to experiments on space missions.

Unique Features of Sounding Rockets[19]

- quick, low-cost access to high altitudes where observations can be made of radiation at wavelengths absorbed by the Earth's atmosphere;
- direct access to the mesosphere and lower thermosphere (40–150 km);
- low cost;
- rapid response times for examining new phenomena;
- flight of payloads especially crafted to study specific phenomena;
- relatively large payload masses can be carried;
- provision of several minutes of micro-gravity;
- ability to use the Earth's limb as an occulting disk to permit observations close to the Sun;
- ability to tailor payloads to study specific targets;
- access to remote sites;
- long dwell times at apogee;
- slow vehicle speed with respect to the ambient medium;
- collection of vertical parameter profiles;
- ability to recover and refly instruments.

5.5 Vessels

Bathymetry is the measurement of ocean floor topography. Generally, a satellite cannot perform bathymetry directly, simply because most of the electromagnetic energy is absorbed by the water. There are two exceptions. One is the bathymetry of shallow coastal zones where some of the energy does reach the bottom, and reaches the top surface again. The other technique is based on the fact that the top surface reflects, to a certain extent, the height of the ocean floor; the topography can then be measured by using altimetry. First maps of the

ocean floors were generated around 1978 using data from the Seasat satellite.

Bathymetry can best be performed using sonar (ultrasound) on-board a vessel. Unfortunately, this is a time-consuming technique, and large portions of the Earth's ocean floor still remain unexplored.

6 EARTH OBSERVATION SATELLITES AND SENSORS

There have been many Earth observation satellites since the 1960s, too many to mention them all. Just a few typical examples will be chosen and described here.

The satellites will be categorized according to the spatial resolution of their sensors. Some satellites will carry a number of different sensors, in which case the satellite is listed under the category of what is regarded as its most important or most interesting sensor for Earth observation. Next to the low, medium, and high spatial resolution categories, there are two special categories, imaging spectrometry and radar satellites.

In addition to RS instruments, satellites carry a host of other instruments (see Figure 12). Important for RS are instruments for orbital and attitude measurement and orbital and attitude control. The orbital position and velocity may be measured by means of

- ground tracking stations
- on-board reception of a satellite navigation system.

A satellite may measure its own attitude (orientation in space) in the roll, pitch, and yaw planes by means of

- inertial reference units, such as gyroscopes
- tracking the position of known stars
- Earth-horizon scanning sensors.

Once the orbital position and satellite attitude have been determined, ground-based computers at the mission control center calculate the necessary changes required for returning the satellite to its optimum configuration. Command signals may be sent to the satellite with instructions to use its thruster rockets to make the appropriate corrections.

A few typical examples of sensors in the following categories will be described:

- low-resolution sensors, with a spatial resolution larger than 100 m
- medium-resolution sensors, with a spatial resolution between 10 and 100 m

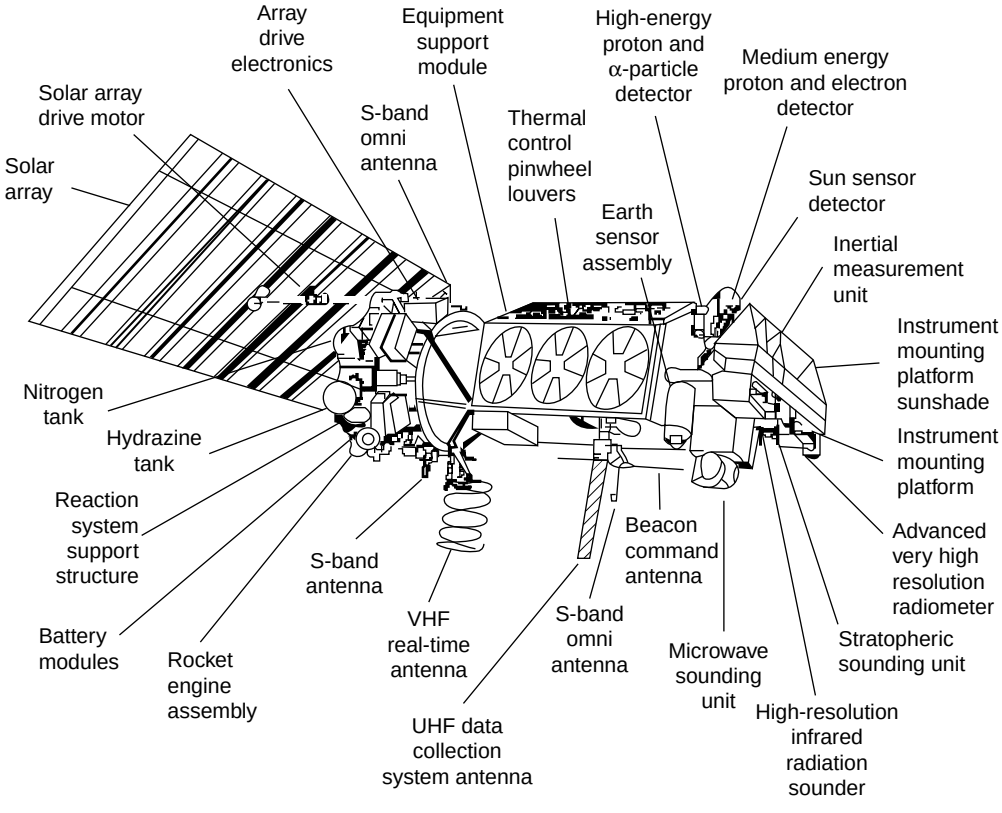

Figure 12 Schematic picture of the TIROS-N spacecraft carrying many different modules and instruments. Copyright NOAA.

- high-resolution sensors, with a spatial resolution better than 10 m
- imaging spectrometry sensors, with a high spectral resolution
- radar sensors, with an active microwave system.

It should be noted that the distinction between the low-, medium-, and high-resolution sensors is arbitrary and other groupings are possible. Sometimes the term very-high resolution is used for satellite sensors having a resolution better than a few meters. An explanation of some parameters is given in Table 4.

Good sources for the description of satellites and sensors are Kramer[8] and Jane's.[20]

Table 4 Explanation of the parameters of satellite technical data[a]

Parameter	Explanation
System	Name of the system, sequence number, launch dates
Orbit	Type of orbit
Sensor	Name and type of sensor
Spectral bands	Band positions/ranges expressed in nm, μm, GHz
Bandwidth	Width of the spectral bands (imaging spectrometry)
Number of bands	Total number of bands (imaging spectrometry)
Polarization modes	HH, HV, VV, VH polarization (radar systems)
Spatial resolution (and IFOV)	Ground resolution, expressed in m, km. In some cases the IFOV is also given, expressed in degrees
Swath width (and FOV)	Width of the scene covered by the satellite, expressed in km. In some cases this is also given as FOV, which relates to area observed, expressed in degrees
Incidence angle	Or look angle; expresses the off-nadir pointing capability of the satellite/sensor in degrees
Revisit time	Theoretical revisit time based on sensor-platform characteristics (disregarding cloud coverage)
Downlink	Capacity for downlinking the data, expressed in megabits per second (Mbit s^{-1})
Recorder	On-board recording capability, expressed in gigabits (Gbit)

[a] When a certain capability is not listed for a particular satellite and sensor, that does not necessarily mean that it does not have this capability; it just means that it is not listed. The list here is not meant to be complete; it is given as a comparative list. For completeness the reader is urged to consult one of the reference books mentioned above, or one of the many on-line resources.

6.1 Low-resolution Systems

Typically, these sensors have a spatial resolution larger than 100 m. They trade much detail in the image for large coverages, and frequent revisits. A global system of geo-stationary and polar orbiting satellites is used for the observation of global weather. Other satellites are used for oceanography and mapping phenomena on a continental or even global scale. For instance, images of the NOAA satellites are used for the food early warning system (FEWS) in Africa.

One major difference between the weather satellite group and the higher resolution groups is that these programs are the most operationally robust. Planning of replacement satellites typically is 10 years in advance, and if an operational satellite breaks down it can quickly be replaced.

Many of the medium- and high-resolution satellites also carry a wide-field, low-resolution sensor. With different sensors at different resolutions on one platform, the same phenomenon can easily be studied at different scales.

Characteristics of Low-resolution Sensors

- spatial resolution >100 m
- 3–7 spectral bands
- swath >500 km
- daily revisit capability.

Some low-resolution sensors are listed in Table 5.

6.1.1 Meteosat

Meteosat 1 was the first European meteorological geo-stationary satellite. Meteosat 5 is the currently the primary satellite, with Meteosat 6 as standby. Meteosat is controlled by Eumetsat, an international organization representing 17 European states.

Meteosat itself (Table 6) is part of a global network of geo-stationary meteorological satellites distributed around the equator. The other satellites are owned by the USA (GOES-E and GOES-W), Russia (GOMS), Japan (GMS), India (INSAT), and China (FENGYUN). The Meteosat satellite covers Europe, Africa, and the Atlantic Ocean. It observes the Earth in three spectral bands, the visible band, the WV band, and the TIR band at a low resolution of 2.5 km (5 km TIR). An image is provided every 30 min.

Meteosat Second Generation (MSG) will appear in 2000, together with the first polar orbiting Metop satellite.

6.1.2 National Oceanic and Atmospheric Administration Satellite

The workhorses of meteorology have been the NOAA satellites. The program has evolved over several

Table 5 Some low-resolution sensors

Platform	Orbit	Sensor	Spatial resolution	Spectral bands	Swath width (km)	Revisit time
Meteosat	GEO	VISSR	2.5 km	3	Earth disk	30 min
NOAA	Polar	AVHRR	1 km	7	3000	Daily
Resurs-O1	Sun-sync.	MSU-SK1	200–300 m	4	760	3–5 days
SeaStar	Sun-sync.	SeaWiFS[a]	1.1 km	8	2800	Daily

[a] WiFS, wide-field sensor.

Table 6 Meteosat

System	Meteosat 5, since 2 March 1991 Meteosat 6 (backup satellite), since 20 November 1993
Orbit	Geo-stationary, 0° longitude (Meteosat 6 on 9°W)
Sensor	VISSR[a]. Platform spin rate 100 rpm
Spectral bands (μm)	0.5–0.9 (VIS) 5.7–7.1 (WV) 10.5–12.5 (TIR)
Spatial resolution (km)	2.5 (VIS and WV) 5 TIR
Swath width	Full Earth disk (FOV = 18°)
Revisit time (min)	30

[a] VISSR, visible and infrared spin scan radiometer.

generations of satellites (TIROS, ESSA, TIROS-M, and TIROS-N, to the NOAA-KLM series), starting with TIROS-1 and culminating in the most recent NOAA-15 operational satellite. They have been the host for a shipload of different instruments for measuring the atmosphere's temperature and humidity profiles, the Earth's radiation budget, space environment, instruments for distress signal detection (search and rescue), instruments for relaying data from ground-based and airborne stations, and more (see Figure 12).

For Earth observation, the most interesting instrument is the AVHRR scanner. The AVHRR scans the Earth in five spectral bands: band 1 in the visible red around 0.6 μm, band 2 in the NIR around 0.9 μm, band 3 in the MWIR around 3.7 μm, and bands 4 and 5 in the TIR around 11 and 12 μm, respectively. This combination of bands makes the AVHRR suitable for a wide range of applications, from measurement of cloud coverage, to sea surface temperature, vegetation, land and sea ice. The disadvantage of the AVHRR is its coarse resolution of about 1 km at the nadir point. However, the major benefit of the AVHRR lies in its high temporal frequency of coverage. The NOAA satellites are operated in a two-satellite system. Both satellites are in a Sun-synchronous orbit; one satellite will always pass around noon and midnight and the other always passes in the morning and

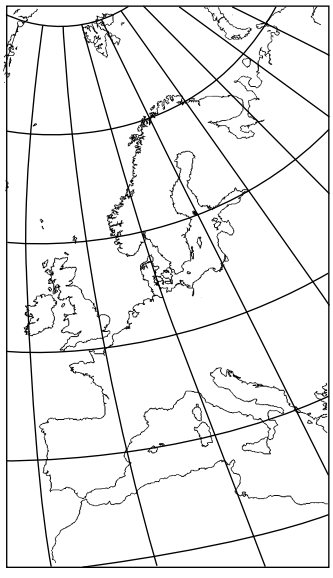

Figure 13 Example of the swath of the AVHRR. A large part of Europe is imaged in a single image!

in the evening. The AVHRR sensors have an extreme FOV of 110°, and together they give a global coverage each day! Every spot on Earth is imaged at least twice each day, depending on latitude. It is the instrument for the observation of phenomena on a global scale, although, owing to its frequent revisit time, it is being used for many monitoring projects on a regional scale. See Figure 13 for an impression of the coverage over Europe of a single AVHRR scene.

The imagery of the AVHRR is also known by other names. High-resolution Picture Transmission (HRPT) is the digital real-time reception of the imagery by a ground station. There are over 500 HRPT receiving stations registered by the World Meteorological Organization (WMO) worldwide. The satellite can also be programmed to record a number of images. Such images, although having the same characteristics as HRPT, are called Local Area Coverage (LAC). Next to the 1-km resolution LAC, the satellite can resample the data on the fly to 4-km resolution Global Area Coverage (GAC). Finally, two bands of 4-km resolution imagery are transmitted by an analog weather fax signal from the satellite, which can

be received by relatively simple and low-cost equipment. This is called Automatic Picture Transmission (APT).

US President Clinton in May 1994 approved a long-mooted plan to combine NOAA's polar orbiter with the military, halving the number of operational satellites to two, thus achieving a considerable reduction in operational costs. In about 2003, the American program will even merge with the European plan for polar orbiters, when the NOAA-N and Metop-1 satellites will join forces. Metop-2 and NOAA-N′ are scheduled for 2007.

Much more can be said about the characteristics, and use and abuse of the AVHRR imagery. Two excellent sources are Cracknell[21] and Belward et al.[11]

Note: NOAA-15 carries an improved sensor, the AVHRR/3. It has an extra spectral channel called band 3A at 1.6 μm SWIR. The old band 3 is now called band 3B. The AVHRR images will still have five bands of information, because band 3A is switched on during the day and band 3B during the night. Furthermore, bands 1, 2, and 3A have increased sensitivity at low light and energy levels. The prime reason for this is to improve ice, snow and aerosol products produced from the data.

A description of NOAA systems is given in Table 7.

6.1.3 Resurs-O1

Maybe it is not altogether fair to list the Russian Resurs in the low-resolution category as it is actually equivalent to the US Landsat. However, the satellite is best known for its relatively cheap, large coverage images of the MSU-SK conical scanner. Receiving stations are located only in Russia and in Sweden. The Swedish Space Corporation Satellitbild also processes and distributes the images. Taking data from Resurs' recorder allows it to reach a growing global market. With a swath of 760 km and a resolution of about 250 m, Resurs fills the gap between the 1-km resolution NOAA images and the 30-m resolution Landsat images.

Technical data are given in Table 8.

Table 7 NOAA

Operational system	NOAA-14 (since 30 December 1994) NOAA-15 (since 13 May 1998)
Orbit	850 km, 98.9°, Sun-synchronous (afternoon or morning)
Sensor	AVHRR/2, electromechanical rotating mirror scanner
Spatial resolution and sampling	1 km at nadir 6 km at limb of sensor IFOV = 1.4 mrad Sample step 0.96 mrad
Spectral bands	Band 1 580–680 nm Band 2 725–1100 nm Band 3 3.55–3.93 μm Band 4 10.3–11.3 μm Band 5 11.4–12.4 μm
Swath width (km)	~2800 (FOV = 55.4° both sides)
Revisit time	2–14 times per day, depending on latitude

Table 8 Resurs-O1

Operational system	Resurs-O1 No. 4, since 10 July 1998
Orbit	835 km, 98.75°, Sun-synchronous
Sensors	MSU-E pushbroom CCD MSU-SK conical scanner
Spatial resolution (m)	30 (MSU-E) 200–300 (MSU-SK)
Spectral bands (μm)	Band 1 0.5–0.6 Band 2 0.6–0.7 Band 3 0.7–0.8 Band 4 0.8–1.1 Band 5 10.4–12.6
Swath width (km)	760
Revisit time (days)	3–5

6.1.4 OrbView-2 or SeaStar/SeaWiFS

Launched on 1 August 1997, SeaStar will deliver multi-spectral ocean-color data to NASA for 5 years. This is the first time that the US Government has purchased global environmental data from a privately designed and operated RS satellite. SeaStar carries the SeaWiFS sea-viewing WiFS, which is a next generation of the Nimbus 7's Coastal Zone Color Scanner (CZCS). SeaWiFS measures ocean surface-level productivity of phytoplankton and chlorophyll. However, SeaStar was originally designed for ocean color but was later changed to make it able to measure the higher radiances from land. Hence it provides a more environmentally stable vegetation index than that derived from NOAA's AVHRR, which is inaccurate under hazy atmospheric conditions because of its single VIS and NIR channels.

Band 1 looks at gelbstoffe (yellow substance), bands 2 and 4 at chlorophyll, band 3 at pigment, and band 5 at suspended sediments. Bands 6, 7 and 8 look at atmospheric aerosols, and are provided for atmospheric corrections.

Technical data are given in Table 9.

6.2 Medium-resolution Systems

These sensors all have medium area coverage, a medium spatial resolution, a moderate revisit capability, and MS band characteristics comparable to those of the current Landsat and SPOT satellites.

The scale of the images of these satellites makes them especially suited for land management and land-use planning for extended areas (regions, countries,

Table 9 OrbView 2

System	SeaStar, since 1 August 1997
Orbit	705 km, 98.2°, Sun-synchronous, equator crossing at noon
Resolution (km)	1.1
Swath width (km)	2800
Revisit time (days)	1
Incidence angles (°)	+20, 0, −20, to avoid specular reflection (glint) from the Sun
Spectral bands (nm)	402–422, 433–453, 480–500, 500–520, 545–565, 660–680, 745–785, 845–885

continents). This explains why most of the satellites in this group are being funded and operated by governments. Unfortunately, the usefulness of these satellites is not accepted by the general public as it is for the weather satellites. This is the reason why these programs are not as operationally robust as the weather satellite programs. The failure of Landsat 6 was a major setback, and it took the USA another 6 years to replace the 15-year-old Landsat 5.

Most of these medium-resolution satellites are in a Sun-synchronous orbit.

Characteristics of Medium-resolution Satellites

- spatial resolution between 10 and 100 m
- 3–7 spectral bands
- swath between 50 and 200 km
- incidence angles from 0 to 30 degrees
- revisit 3 days and more.

Some medium-resolution satellites are listed in Table 10.

6.2.1 Landsat

The workhorses of Earth observation truly are the Landsat satellites, having been in operation since 1972. Landsat 5 has even been in operation for 15 years! The Landsat satellites were developed in the 1960s. Landsat 1–3 were enhanced versions of the Nimbus weather and research satellites, and were originally known as the ERTS. The moment Landsat 1 became operational, its vivid flow of revealing images was regarded as sensational by the early investigators. The quality of the images led to the information being put to immediate practical use. It became clear that they were directly relevant to the management of the world's food, energy, and environment.

In 1981, as it became clear that the Landsat program would represent a total US investment of $1 billion, the US Reagan administration decided that the RS program must either die or be taken over by the private sector. Since that decision the Landsats have had a confusing history of ownerships. The latest state is that the Eros Data Center sells all multispectral scanner (MSS) and TM data older than 10 years, and Space Imaging EOSAT sells the new TM images. Landsat 7 data will be available at the low cost of reproduction only.

The Landsat satellites have flown the following sensors:

- The Return Beam Vidicon (RBV) Camera
- MSS
- TM.

The RBV basically consisted of three television cameras, each operating at their own spectral band. The ground resolution of the RBV was roughly 80 m, similar to that of the MSS. Even though the RBV had better geometric characteristics, the MSS was the sensor mostly used. One reason for this was that the RBV cameras exhibited many technical problems, but the main reason was that the MSS data were available in digital form (the signal from the RBV was an analog video signal). This made possible the direct further processing of the data, without the need for an intermediate conversion step. Applications of the MSS data have been widespread in many fields. Also, even though MSS images are not recorded any longer, the MSS data are still popular, first, because of the archive going back to the 1970s, MSS data can be used for monitoring changes, and second, because

Table 10 Some medium-resolution satellites[a]

Platform	Sensor	Spatial resolution (m)	Spectral bands	Swath width (km)	Pointing capability (°)	Revisit time (days)
Landsat 4 and 5	TM	30	7	185	No	16
Landsat 7	ETM+	15 (PAN)	8	185	No	16
SPOT 1–3	HRV	10 (PAN)	3	60	±27	4–6
SPOT 4	HRVIR	10 (PAN)	4	60	±27	4–6
Resource21	M10	10 (PAN)	6	205	±30	3–4
CBERS	HRCC + IRMSS	20	9	120	±32	3
Terra EOS AM-1	ASTER	15	14	60	±24	5

[a] EOS, Earth observing system; PAN, panchromatic; HRV, high-resolution visible; HRVIR, high-resolution visible/infrared; CBERS, China/Brazil Earth Resources Satellite; ASTER, Advanced Spaceborne Thermal Emission and Reflectance Radiometer.

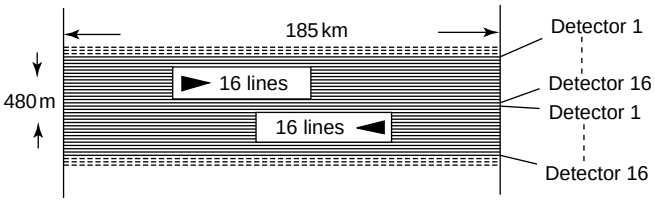

Figure 14 Scanning pattern of the Landsat TM instrument: 16 lines are simultaneously imaged moving first to the left and then to the right in an oscillating way. Line spacing is 30 m.

the MSS data were put in the public domain and can be obtained for $200 per scene.

The TM, carried by Landsat 4 and 5, is the third sensor. Several major differences exist between the basic design of the MSS and the design of the TM. To improve the dwell time, i.e. the time a single detector has to take one measurement of one ground-cell, it uses an oscillating mirror and acquires data during both the forward and reverse sweeps of its mirror (see Figure 14). Furthermore, it has 16 detectors per band, such that 16 lines can be acquired at the same sweep of the mirror (four for the thermal band 6). At any one instant, surface radiation is sensed by a total of 100 TM detectors $[(6 \times 16) + 4]$. Because of the longer dwell time, the TM offers improved radiometric sensitivity over the MSS. TM data are quantized to eight bits (256 possible radiance levels), whereas MSS data are quantized to six bits (64 levels). The TM has four times more levels over the same range than the MSS. Hence the radiometric resolution, the capability to notice small brightness differences between objects, has improved.

The first three Landsat satellites relied upon on-board recorders to store data until the satellite passed within range of a ground station. These recorders were often the first component to fail. Landsat 4 and 5 used a new communication system, called the Tracking and Data Relay Satellite (TDRS) system, which eliminates the need for on-board recorders. The TDRS system consists of satellites in a GEO and a centralized ground station. The first TDRS satellite was launched in 1983. The TDRS system made it possible to receive data in real time from the Landsat satellites all over the world, except for a small strip over Asia, India, and the Indian Ocean. Currently, because of problems with a transmitter on-board, the TDRS links have been lost, and Landsat 5 data can only be received by ground stations.

Landsat 4 was launched on 16 July 1982, but a failing power system prompted the launch of Landsat 5 2 years later on 1 March 1984.

The loss of Landsat 6 in October 1993 was a severe blow to the system and the USA must continue to rely on the aging Landsat 5. Both Landsat 4 and 5 have suffered from degrading subsystems and sensors and are expected to fail at any moment. The construction of Landsat 7 was set back considerably when two power supply units for its ETM+ sensor failed a test and had to be replaced. Landsat 7 was successfully launched on 15 April 1999.

Details of Landsat are given in Table 11.

Table 11 Landsat

Operational system	Landsat 5, since 1 March 1984
Orbit	705 km, 98.2°, Sun-synchronous, 09.45 a.m. local time equator crossing
Repeat cycle (days)	16
Sensor	TM, electromechanical oscillating mirror scanner
Spatial resolution TM (m)	Bands 1–5 and 7 30 Band 6 120
Spectral bands TM (μm)	Band 1 0.45–0.52 Band 2 0.52–0.60 Band 3 0.63–0.69 Band 4 0.76–0.90 Band 5 1.55–1.75 Band 6 10.4–12.50 Band 7 2.08–2.35
Swath width (km)	185 (FOV = 15°)
Downlink (Mbit s^{-1})	X-band 84.9

6.2.2 Landsat 7

The Landsat 7 Data Policy Plan makes the Landsat 7 data available to all users at the cost of fulfilling requests. In fact, Landsat 7 was added to NASA's EOS program, and there will be no distinct Landsat 8 as instruments will be carried within EOS after demonstration within the New Millennium Program (NMP). The NMP Earth Observer 1, NMP/EO-1, will carry a prototype Advanced Land Imager (ALI) for returning 10-m resolution PAN and 30-m resolution MS imagery. ALI has a pushbroom design, which is five times lighter, and which requires five times less power than the current Landsat sensors. NMP/EO-1 will fly less than 1 min ahead of Landsat 7, allowing accurate pair-wise comparison of the two technologies. Eventually, ALI will fly on the EOS AM-2 mission in 2004 to follow on from Landsat 7.

Details of Landsat 7 are given in Table 12.

6.2.3 Système Pour l'Observation de la Terre 1–3

First named Système Probatoire d'Observation de la Terre (Test System for Earth Observation), but later renamed to Système Pour l'Observation de la Terre (System for Earth Observation), the SPOT system has been in operation since 1986. The SPOT satellites each carry two identical HRV sensors, consisting of CCD linear arrays. Essentially, all the points of one line

Table 12 Landsat 7

Operational system	Landsat 7, launched 15 April 1999
Orbit	705 km, 98.2°, Sun-synchronous, 10 a.m. local time equator crossing
Repeat cycle (days)	16
Sensor	ETM+, electromechanical oscillating mirror scanner; imagery can be collected in low- or high-gain modes; high gain doubles the sensitivity
Spatial resolution (m)	30 PAN 15 Band 6 60
Spectral bands (μm)	Band 1 0.45–0.52 Band 2 0.52–0.60 Band 3 0.63–0.69 Band 4 0.76–0.90 Band 5 1.55–1.75 Band 6 10.4–12.50 Band 7 2.08–2.35 Band 8 PAN 0.50–0.90
Swath width (km)	185 (FOV = 15°)
Downlink	X-band 2 × 150 Mbit s^{-1}, 300 Mbit s^{-1} playback
On-board recorder	375 Gbit SSR for about 100 ETM+ scenes

SSR, solid-state recorder.

are imaged at the same time by the many detectors of the linear array. SPOT 1 was the first satellite to use this pushbroom technology. Pushbroom sensors avoid mechanical problems associated with moving scan mirrors. An additional advantage is that each detector has more time to perform a measurement of one ground-cell, which means it has a much longer dwell time, which in turn means that a better signal can be achieved.

A disadvantage of having so many detectors is that the sensitivities of all these detectors may be slightly different, requiring a calibration of thousands of detectors! A bad calibration of a pushbroom is visible as a pattern of vertical darker and brighter lines in the image, called vertical striping.

The SPOT sensors can be tilted from the normal downward viewing mode by ±27°, which offers the possibility of viewing objects from two different sides (see Figure 15). These stereo images can be used to determine the height of objects, or even the height of the terrain. The height model thus determined is referred to as a Digital Terrain Model (DTM).

SPOT 1 was already put in standby mode, but it has been activated again after SPOT 3 failed on 14 November 1996. SPOT 3 ran out of power when a wrong series of commands was sent to the satellite, and could not be recovered.

Details of SPOT sensors are given in Table 13.

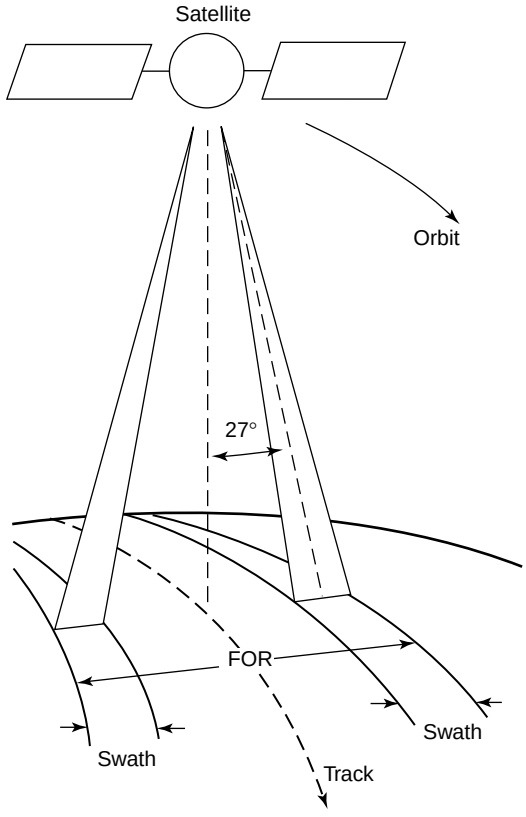

Figure 15 Pointing capabilities of the SPOT sensors, up to 27 degrees side-looking. The field of regard (FOR) is more than 950 km. Swath width is 60–80 km (after Kramer.[8])

Table 13 SPOT sensors

Operational system	SPOT 1, since 22 February 1986 SPOT 2, since 22 January 1990
Orbit	832 km, 98.7°, Sun-synchronous, 10.30 a.m. local time equator crossing
Repeat cycle (days)	26
Sensor	2 × HRV, pushbroom linear CCD array
Spatial resolution (m)	MS mode 20, PAN 10
Spectral bands MS (nm)	Band 1 500–590 Band 2 610–680 Band 3 790–890 PAN 510–730
Swath width (km)	60
Look angle (°)	±27, left and right from nadir
Revisit time (days)	4–6
Downlink (Mbit s^{-1})	X-band 50

6.2.4 Système Pour l'Observation de la Terre 4

SPOT 4 is considered a second-generation satellite, and carries a number of improvements over the older SPOTs:

1. An extra spectral band was added in the MWIR.
2. The PAN band has a narrower bandwidth.
3. Increased on-board storage capacity.
4. It carries a new wide FOV sensor for vegetation.
5. Its design life has been extended.
6. It carries an experimental laser communication device.

HRVIR instruments have superseded the HRV optical instruments on SPOT 1, 2, and 3. The HRVIR has three bands in the VIS and NIR spectrum and, in addition to the HRV channels, carries an additional band in the MWIR spectrum.

Another important addition to SPOT 4 is the VEGETATION instrument. This instrument is the first to use the HRVIR's linear-array technology to produce high-quality imagery at lower resolution with greatly reduced distortion, compared with electromechanical sensors, such as the AVHRR and WiFS.

With a resolution of 1.15 km at nadir and a swath width of 2250 km, the VEGETATION instrument will cover almost all of the globe's landmasses while orbiting the Earth 14 times a day. Only a few zones near the equator will be covered every other day. Areas above 35° latitude will be seen at least once daily. The HRVIR and VEGETATION have three identical spectral bands. The first band of VEGETATION is shifted towards the blue absorption band of chlorophyll (0.43–0.47 μm).

The laser inter-satellite communication device PASTEL is an experiment. The objective is to transmit imaging data from SPOT 4 to the ground over a high-rate laser link via the future Artemis geostationary data relay satellite of the European Space Agency (ESA).

With SPOT 5 scheduled to come on-stream around 2002, the SPOT family of Earth observation satellites still has many long years of service ahead of it. SPOT 5's two high-resolution geometric (HRG) instruments will further enhance its performance. Spatial resolution will improve to 5 m. Further, SPOT Image plans to offer a 2.5-m sample distance product by combining two HRG images.

Details of SPOT 4 are given in Table 14.

Table 14 SPOT 4

Operational system	SPOT 4, since 24 March 1998
Orbit	Same as SPOT 1, 2, 3
Sensors	2 × HRVIR, pushbroom linear CCD array, VEGETATION
Spatial resolution	MS mode 20 m PAN 10 m VEGETATION 1.15 km
Spectral bands MS (nm)	Band 1 500–590 Band 2 610–680 Band 3 790–890 Band 4 1.58–1.75 μm
PAN mode (nm)	610–680 (same as MS band 2)
Swath width (km)	60
VEGETATION bands (μm)	Band 1 0.43 Bands 2–4 same as HRVIR
VEGETATION swath width (km)	2250

Table 15 Resource21

Operational system	4 satellites to be launched in 2000
Orbit	740 km, 98.6°, Sun-synchronous, 10.30 a.m. crossing at ascending node
Sensor	M10 (MS 10-m resolution), pushbroom
Spectral bands (μm)	0.45–0.52 blue, 0.53–0.59 green, 0.63–0.69 red, 0.76–0.90 NIR, 1.55–1.68 SWIR, 1.23–1.53 cirrus clouds
Spatial resolution (m)	10 VNIR 20 SWIR 100+ cirrus
Swath width (km)	205
Look angle (°)	About ±30 cross-track
Revisit	3 or 4 days with 4 satellites

6.2.5 Resource21

Resource21 is the name of a commercial RS information services company. Resource21 will combine satellite and aircraft RS to provide twice-weekly information products within hours of data collection, based on 10-m resolution MS imagery. The complete system will consist of four satellites. Every area on Earth will be visited by one of the satellites every three or four days, resulting in a revisit twice a week. In other words, the constellation will give a global coverage every three or four days at 10-m resolution. Details are given in Table 15.

The main application areas of Resource21 are agriculture (''precision farming'') and environment and natural resource monitoring.

At the time of writing it is not clear whether there is enough funding for continuation of the Resource21 program.

6.2.6 China/Brazil Earth Resources Satellite (CBERS)

The CBERS is also known as Ziyuan-1. The development and launch were realized by a trade agreement between China and Brazil. Its main imaging system has SPOT-like characteristics, using linear CCDs. Also a WFI will be provided. Plans provide for a second CBERS-2 satellite,

Table 16 CBERS[a]

Operational system	CBERS, to be launched in 1999
Orbit	778 km, 98.5°, Sun-synchronous, repeat cycle 26 days
Sensors	Pushbroom HRCC, IRMSS, WFI
Spatial resolution (m)	20 and 260 VIS 80 IR 160 TIR
Spectral bands MS (μm)	Band 1 0.51–0.73 Band 2 0.45–0.52 Band 3 0.52–0.59 Band 4 0.63–0.69 Band 5 0.77–0.89
IR/TIR (μm)	0.50–1.10, 1.55–1.75, 2.08–2.35, 10.4–12.5
Swath width (km)	120
Incidence angle (°)	±32 side-looking,
Revisit time (days)	3
Downlink (Mbit s^{-1})	2 × 53
WFI bands (μm)	0.63–0.69, 0.77–0.89
WFI swath width (km)	900

[a] HRCC, high-resolution charge-coupled device; IRMSS, Infrared Multispectral Scanner; WFI, Wide-field Imager.

Table 17 EOS AM-1

Operational system	Terra EOS AM-1, to be launched December 1999
Orbit	705 km, 98.2°, Sun-synchronous, 10.30 a.m. equator crossing, repeat cycle 16 days
Sensors	ASTER
ASTER carries three telescopes (μm)	VNIR 0.56, 0.66, 0.81 SWIR 1.65, 2.17, 2.21, 2.26, 2.33, 2.40 TIR 8.3, 8.65, 9.10, 10.6, 11.30
Spatial resolution (m)	VNIR 15 SWIR 30 TIR 90
Pointing capabilities (°)	±8.5 SWIR/TIR, ±24 VNIR
Swath width (km)	60
Revisit time (days)	5
Downlink (Mbit s^{-1})	150 via TDRS
Recorder	140 Gbit SSR

and CBERS-3 and -4 are projected to have a higher resolution. Details are given in Table 16.

6.2.7 Terra/Earth Observing System AM-1

EOS is the centerpiece of NASA's Earth Science mission. The EOS AM-1 satellite, later renamed Terra, is the flagship of the fleet and will be launched in mid-1999. It carries 5 RS instruments including the moderate-resolution Imaging Spectroradiometer (MODIS) and the ASTER. ASTER is a high spectral resolution imaging spectrometer. The instrument is designed with three bands in the VIS and NIR spectral range with a 15-m resolution, six bands in the SWIR range with a 30-m resolution, and five bands in the TIR range with a 90-m spatial resolution. The VNIR and SWIR bands have a spectral bandwidth of the order of 10 nm. Simultaneously, a single band in the NIR range will be provided for along-track stereo viewing capability. The swath width of the image will be 60 km with a revisit time of about 5 days.

Also on EOS AM-1 (Table 17) is MODIS, which is planned as a land RS instrument with a high revisit capability. MODIS is mainly designed for global change research.

6.3 High-resolution Systems

This was once the exclusive domain of spy satellites in terms of resolution. Already in the 1960s spy satellites existed that had a resolution of better than 10 m. Civil high-resolution satellites had to wait until the very end of the 20th century. The major breakthrough was one of policy rather than technology. The US Land RS Act of 1992 concluded that a robust commercial satellite RS industry was important to the welfare of the USA and created a process for licensing private companies to develop, own, operate, and sell high-resolution data from Earth-observing satellites. Two years later, four licenses for 1-m systems were granted, and currently the companies involved are struggling to get their satellite first in space. This revolution promises to set off an explosion in the amount and use of image data. Applications that are not possible with today's space or aircraft systems may well become routine.

Much is expected in the field of GIS technology. GIS databases will be constructed using the 1-m images, reducing reliance on out-of-date paper maps. Highly accurate topographic maps, called DTMs, can also be developed from the images and added to the databases. Because they cover large areas, high-resolution satellite images are expected to replace aerial photographs for certain types of detailed mapping such as gas pipeline routing, urban planning, and real estate.

Forestry and agriculture are also expected to benefit greatly from the increased resolution. The health and growing stage of crops can be monitored to areas within the field boundaries. This is known as "precision farming" and enables the farmer to apply water, fertilizer, and pesticides to specific areas in the field. In forestry, individual trees could be identified and mapped over enormous areas such as the Amazon. Tree species can be isolated by determining their signature in the electromagnetic spectrum.

The high-resolution imaging capability requires a change in instrument design to a pushbroom with a large telescope technique, resulting in a new spacecraft design.

In contrast to the medium-resolution satellites, this group has limited MS coverage, or even just PAN capabilities. They do have extreme pointing capabilities to extend to potential imaging coverage. The pointing capability can also be used for last-minute reprogramming of the satellite if CC makes this necessary.

The private sector has shown an almost exclusive interest in high-resolution systems. Obviously, it is believed that these systems represent the space capability needed to create commercially valuable information products. On the other hand, the pure commercial nature with no government funding implies a high risk. This was demonstrated by the loss of the EarlyBird satellite and the resulting reorganization at the EarthWatch Company. Reliability and consistent service provision in any business depend very much on back-up solutions in case of need. Most companies in the high-resolution business have such a back-up satellite in store so as to be able to launch a replacement satellite at short notice. However, still, the loss of one satellite means a loss of millions of dollars, which may be considerable for a business just starting in this field.

Characteristics of High-resolution Satellites

- Spatial resolution better than 10 m
- 1–4 spectral bands
- Swath width less than 100 km
- Extreme pointing capabilities
- Revisit 3 days and more.

Some high-resolution satellites are listed in Table 18.

6.3.1 Indian Remote Sensing Satellite 1C and 1D

Having been the seventh nation to achieve orbital capability in July 1980, India is pressing ahead with an impressive national program aimed at developing launchers in addition to nationally produced communications, meteorological, and Earth resources satellites. Currently, India has more than a half a dozen RS satellites under development, intended for launch in the next decade. The entire fleet will be a mix of niche-role and multi-role satellites, which will join India's existing constellation of polar orbiters. The US Company Space Imaging EOSAT plans to market imagery and data from all satellites that India expects to fly by 2005. The IRS-1C and -1D offer improved spatial and spectral resolution, on-board recording, stereo viewing capability and more frequent revisits. They carry three separate imaging sensors, the WiFS, the Linear Imaging Self-Scanning Sensor (LISS), and the high-resolution PAN sensor.

The WiFS provides regional imagery acquiring data with 800-km swaths at a coarse 188-m resolution in two spectral bands, VIS (620–680 nm) and NIR (770–860 nm), and will mainly be used for vegetation index mapping. The WiFS offers a rapid revisit time of 3 days.

LISS-3 serves the needs of MS imagery clients, possibly the largest of all current data user groups. LISS-3 acquires four bands (520–590, 620–680, 770–860, and 1550–1750 nm) with a 23.7-m spatial resolution, which makes it an ideal complement to data from the aging Landsat 5 TM sensor.

The most interesting of the three sensors is the PAN sensor with a resolution of 5.8 m, giving the IRS-1C and -1D the highest resolution of any civilian RS satellite currently in orbit. With its 5.8-m resolution, the IRS-1C and -1D can cover applications that require spatial detail and scene size between the 10-m SPOT satellites and the upcoming 1-m systems. The PAN sensor can be pointed off-nadir up to $\pm 26^{\circ}$ and thus offers stereo capabilities and a possible frequent revisit of about 5 days, depending on the latitude. Working together, the IRS-1C and -1D (Table 19) will also cater for users who need a rapid revisiting rate.

In late 1999, India will initiate a high-resolution mapping program with the launch of the IRS-P5, which has been dubbed Cartosat-1. It will acquire 2.5-m resolution PAN imagery. There seem to be plans to improve future Cartosat-2 to 1-m resolution.

Table 18 Some high-resolution satellites

Platform[a]	Sensor	Spatial resolution (m)	MS	Swath width (km)	Pointing capability (°)	Revisit time (days)
IRS-1C and D	PAN	5.8	4 bands	70	±26	5
Cosmos	KVR-1000	~2	No	160	No	N/A
OrbView-3	PAN	1	4 bands	8	±45	3
Ikonos 1	OSA	1	4 bands	11	±30	1–3
QuickBird	QBP	1	4 bands	27	±30	1–3
Eros A+	CCD	1.8	No	12.5		

[a] IRS, Indian Remote Sensing Satellite; OSA, Optical Sensor Assembly.

Table 19 IRS-1C and -1D

Operational system	IRS-1C, launched 28 December 1995 IRS-1D, launched 28 September 1997
Orbit	817 km, 98.6°, Sun-synchronous, 10.30 a.m. equator crossing, 24-day repeat cycle
Resolution (m)	5.8 PAN
Spectral band (μm)	0.50–0.75
Swath width (km)	70
Incidence angle (°)	±26, steerable
Revisit time (days)	5
Recorder	OBTR, for 24 min of data

OBTR, on-board tape recorder.

Table 20 SPIN-2

Operational system	Cosmos, last launch 17 February 1998
Orbit	LEO, 212 × 276 km, 70.4°
Sensor	KVR-1000 camera
Resolution (m)	~2
Spectral band (nm)	510–760
"Swath"/coverage (km)	40 × 160 scenes

6.3.2 Space Information – 2 m Resolution/KVR-1000

Data from the Russian KVR-1000 camera, flown on a Russian Cosmos satellite, is marketed under the name of SPIN-2 (Table 20). It provides high-resolution photography of the USA in accordance with a Russian–American contract. Currently SPIN-2 offers the worlds' highest resolution, commercially available satellite imagery. SPIN-2 PAN imagery has a resolution of about 2 m. The data are single band with a spectral range between 510 and 760 nm. Individual scenes cover a large area of 40 by 180 km. Typically, the satellite is launched and takes images for 45 days, before it runs out of fresh film. The last mission was in February–March 1998. The archive goes back to 1980, and is available on-line via the Internet.

6.3.3 OrbView-3

OrbView-3 (Table 21) will produce 1-m resolution PAN and 4-m resolution MS imagery.

Table 21 OrbView-3

Operational system	OrbView-3, to be launched 19 November 1999
Orbit	470 km, Sun-synchronous, 10.30 a.m. equator crossing
Spatial resolution (m)	1
Swath width (km)	8
Look angle (°)	Up to 45
Revisit (days)	3

Table 22 Ikonos

Operational system	Ikonos-1, launch failed 27 April 1999 Ikonos-2, is to be launched late 1999
Orbit	680 km, 98.2°, Sun-synchronous
Repeat cycle (days)	14
Sensor	OSA pushbroom telescope
Spectral band PAN (μm)	0.45–0.90
Spectral band MS (μm)	0.45–0.52, 0.52–0.60, 0.63–0.69, 0.76–0.90
Spatial resolution (m)	1 PAN 4 MS
Swath width (km)	11
Look angle (°)	±30 into any direction
Revisit time (days)	1–3

6.3.4 Ikonos

The Ikonos satellite system (Table 22) was initiated as the Commercial Remote Sensing System (CRSS). The satellite will routinely collect 1-m PAN and 4-m MS imagery. Mapping North America's largest 100 cities is an early priority. Unfortunately, the launch of the first Ikonos failed on 27 April 1999, because the protecting rocket nose fairing refused to come off, giving the satellite too much mass to reach orbit. Ikonos-2 will probably be launched later in 1999.

The OSA features a telescope with a 10-m focal length (folded optics design) and pushbroom detector technology. Simultaneous imaging in the PAN and MS modes is provided. A body-pointing technique for the entire spacecraft permits a pointing capability of ±30° in any direction.

Tasking commands can be modified up to 10 min before satellite contact, accommodating last-minute priorities or changing weather conditions.

6.3.5 QuickBird

QuickBird is (Table 23) the next-generation satellite of the EarlyBird satellite. Unfortunately, EarlyBird-1 was lost shortly after launch in December 1997. After its launch in late 1999, QuickBird-1 will offer 1-m resolution.

The sensor can be pointed off-axis ±30° into any direction, offering unprecedented stereo viewing capabilities.

6.3.6 Eros

Eros (Table 24) is the result of a joint venture between the USA and Israel. The Eros A+ satellite will have a resolution of about 1.8 m. The follow-up satellite Eros B will have a sub-meter resolution of about 80 cm.

Table 23 QuickBird

Operational system	QuickBird-1, to be launched late 1999
Orbit	600 km, 52° inclination
Sensor	QBP QBM
Spectral band QBP (μm)	0.45–0.90
Spectral band QBM (μm)	0.45–0.52, 0.53–0.59, 0.63–0.69, 0.77–0.90
Spatial resolution (m)	1 QBP 4 QBM at nadir
Swath width (km)	27
Look angle (°)	±30 in any direction
Revisit (days)	0.5–2.5 in a two-satellite configuration
Downlink (Mbit s^{-1})	X-band, 300

QBP, QuickBird Panchromatic; QBM, QuickBird Multispectral.

Table 24 Eros

Operational system	Eros A+, to be launched in 2000
Orbit	480 km, 50°
Sensor	CCD
Resolution (m)	1.8
Swath width (km)	12.5

6.4 Imaging Spectrometry Systems

These satellites will use a near-continuous spectrum to capture all of the information in the reflected radiance. This capability promises to make possible entirely new applications and to improve the accuracy of current MS analysis techniques. The technique is called imaging spectrometry, also referred to as imaging spectroscopy and hyperspectral imaging.

The push for imaging spectrometers has a long history in the geophysical field. Aircraft-based experiments have shown that measurements of the continuous spectrum allow greatly improved mineral identification. Clay minerals that cannot be distinguished by a human observer at a 1-m distance can be identified in hyperspectral images taken from a 20-km height! Recent airborne campaigns have also indicated significant improvements in the accuracy of vegetation identification and stress measurement.

The term ''hyperspectral'' is used for sensors capable of recording between 100 and 1000 bands. Future sensors, capable of recording more than 1000 spectral bands, are referred to as ''ultraspectral'' sensors.

Calibration and processing of hyperspectral data are difficult because of the large number of channels and the sheer amount of data. Algorithms developed specifically to make use of or cope with the high spectral dimensionality of hyperspectral data include:

- binary encoding[22]
- waveform characterization[23]
- wavelets[24]
- artificial neural networks[25]
- derivative analysis[26]
- model inversion[27]
- spectral feature fitting[28,29]
- spectral angle mapping[30]
- spectral unmixing[31]
- constrained energy minimization[32]
- classification[33]
- cross-correlation spectral matching.[34]

Characteristics of Imaging Spectroscopy Satellites

- spatial resolution about 10–30 m
- more than 100 spectral bands
- swath width 5–15 km
- a revisit time of 3–7 days.

Some imaging spectrometry satellites are listed in Table 25.

6.4.1 OrbView-4

OrbView-4 (Table 26) will be the successor of the OrbView-3 high-resolution satellite. As with OrbView-3, OrbView-4's high-resolution camera will acquire 1-m resolution PAN and 4-m resolution MS imagery. In addition, OrbView-4 will acquire hyperspectral imagery.

Table 25 Some imaging spectrometry satellites

Platform	Sensor	Spatial resolution (m)	Spectral bands	Spectral range (μm)	Swath width (km)	Revisit time (days)
OrbView 4		8	200	0.45–2.5	5	3
NMP/EO-1	Hyperion	30	220	0.4–2.5	7.5	
	LAC	250	256	0.9–1.6	185	
Aries-1		30	64	0.4–1.1	15	7
			64	2.0–2.5		

Table 26 OrbView-4

Operational system	OrbView-4, to be launched in 2000
Orbit	Same as OrbView-3
Spectral range (nm)	450–2500
Number of bands	200
Spatial resolution (m)	8
Swath width (km)	5
Spectral resolution (nm)	10
Spectral coverage	Continuous
Revisit time (days)	<3

6.4.2 Earth Observing Satellite

Earth Observing (EO) satellite 1 is the New Millennium Program Earth Observing (NMP/EO-1) satellite. It is an experimental satellite carrying three advanced instruments as a technology demonstration. It carries the ALI, which will be used in conjunction with the ETM+ sensor (see Landsat 7 above) for comparison of the two sensors. In addition to the MS instrument it carries two hyperspectral instruments, the Hyperion and the Linear Etalon Imaging Spectrometer Atmospheric Corrector (LEISAC).

The focus of the Hyperion instrument is to provide high-quality calibrated data that can support the evaluation of hyperspectral technology for spaceborne EO missions. It provides hyperspectral imagery in the 0.4–2.5-μm region at continuous 10-nm intervals. The spatial resolution will be 30 m.

The LEISAC is intended to correct mainly for water vapor variations in the atmosphere using the information in the 890–1600-nm region at 2–6-nm intervals. In addition to atmospheric monitoring, LEISAC will also image the Earth at a spatial resolution of 250 m. The imaging data will be cross-referenced to the Hyperion data where the footprints overlap.

EO-1 (Table 27) is scheduled for launch at December 1999. The launch of EO-1 was delayed for 4 months in order to add a hyperspectral instrument to replace the one lost in the failure of the Lewis satellite. A flawed attitude control system and inadequate spacecraft monitoring by ground controllers doomed NASA's Lewis imaging spectrometry satellite.

Table 27 EO-1

Operational system	EO-1, to be launched December 1999	
Orbit	Same as Landsat 7, one minute earlier	
Instrument	Hyperion	LAC
Spectral range (μm)	0.4–2.5	0.9–1.6
Spectral resolution (nm)	10	2–6
Spectral coverage	Continuous	Continuous
Number of bands	220	256
Spatial resolution (m)	30	250
Swath width (km)	7.5	185

Table 28 Aries-1

Operational system	Aries-1
Orbit	500 km, Sun-synchronous
Sensors	VNIR and SWIR hyperspectral, PAN
Spectral resolution (μm)	0.4–1.1 and 2.0–2.5 in 128 bands
Spatial resolution (m)	10 PAN 30 hyperspectral
Swath width (km)	15
Look angle (°)	±30 across track
Revisit time (days)	7

6.4.3 Aries-1

Aries-1 (Table 28) is a purely Australian initiative to build a hyperspectral satellite, mainly targeted at geological applications for the (Australian) mining business. Aries-1 will become operational in 2001.

6.5 Imaging Radio Detection and Ranging Systems

Radar's cloud-penetrating, all-weather capability already makes it the instrument of necessity for many observational tasks. For areas usually covered by clouds (e.g. tropical areas), it may be the only instrument available for mapping tasks. However, radar has also proven invaluable in assessing floods and other disasters where weather conditions rule out optical images. Its sensitivity to soil moisture makes it useful for agriculture. Surface geometry is clearly visible in radar images, which is useful for identifying fault lines and other structures. Its penetration capabilities allow imaging topography under forested areas and mapping of subsurface structures.

The most important drawback of radar images is that the reflections of radar signals are very complex functions of the physical and structural properties, and also the water content of the target objects. This means that the interpretation of radar images is completely different to the interpretation of optical images. The complexity of radar images is responsible for the long development period faced by radar systems.

Characteristics of Radar Satellites

- spatial resolution between 10 and 100 m
- single or multiple (up to three) radar frequencies
- single or multiple polarization modes (HH, VV, HV, VH)
- swath 15–500 km
- revisit time 5–35 days.

Some radar satellites are listed in Table 29.

Table 29 Some radar satellites

Platform	Sensor[a]	Spatial resolution (m)	Frequency band	Polarization	Swath width (km)	Incidence angle (°)
ERS-1 and -2	AMI/SAR	30	C-band	VV	100	23 (up to 35)
Radarsat-1	SAR	8–100	C-band	HH	50–500	10–60
Envisat-1	ASAR	30 150	C-band	HH + VV, HH/HV, VV/VH	56–100 405	17–45
Space Shuttle (2 × in 1994)	SIR-C X-SAR	25	C- + L- bands X-band	HH, VV, HV, VH VV	15–60 15–45	20–55
JERS-1	SAR	18	L-band	HH	75	35

[a] AMI, Active Microwave Instrument; ASAR, Advanced Synthetic Aperture Radar; Envisat, Environmental Satellite.

6.5.1 European Remote Sensing Satellite

The ERS-1 satellite's three primary all-weather instruments provide systematic, repetitive, global coverage of oceans, coastal zones, and polar ice caps, monitoring wave height and wavelengths, wind speed and direction, precise altitude, ice parameters, sea-surface temperature, cloud top temperature, CC, and atmospheric water vapor content.

1. The AMI can operate as a wind scatterometer or SAR.
2. The Along-track Scanning Radiometer and Microwave Sounder (ATSR-M) provides the most accurate sea-surface temperature to date.
3. The Radio Detection and Ranging Altimeter (RA) measures large-scale ocean and ice topography and wave heights.

The follow-on ERS-2 provides continuity until Envisat-1 appears. ERS-1 and -2 (Table 30) operated simultaneously from 15 August to May 1996, the first time that two identical civil SARs worked in tandem. The orbits were carefully phased to provide 1-day revisits, allowing the collection of interferometric SAR image pairs and improving temporal sampling. Although still working perfectly, lack of funding required it to be put on standby from May 1996. Extending ERS-2 operations beyond 2000 is under discussion.

In addition to the instruments mentioned above, ERS-2 also carried a Global Ozone Monitoring Experiment (GOME) to determine ozone and trace gases in the troposphere and stratosphere.

The combination of images taken during the descending and ascending parts of the orbit can be used for stereoscopic viewing. The technique for determining height from stereo radar images is called radargrammetry.

Table 30 ERS-1 and -2

Operational system	ERS-1, since 17 July 1991 ERS-2, since 21 April 1995
Orbit	Within 771 × 797 km, 98.55°, Sun-synchronous, 35-day repeat cycle
Sensors	AMI, RA, ATSR-M
Spatial resolution (m)	30
Frequency (GHz)	C-band, 5.3
Polarization	VV
Swath width (km)	100
Incidence angle (°)	23 (up to 35 with "roll–tilt" attitude)
Downlink (Mbit s^{-1})	105, real-time only

6.5.2 Radarsat

Designed, constructed, launched, and operated by the Canadian Space Agency (CSA), Radarsat (Table 31) is the world's most powerful commercial radar RS satellite totally dedicated to operational applications. Its C-band SAR has seven different modes of 10–100-m resolution and 50–500-km swath widths combined with 25 beam positions ranging from 10 to 60° incidence angles. Hence a wide variety of products can be offered to the customer. The Radarsat system is designed to operate with no backlog, so that all imagery received within a 24-h period can be processed and distributed within 24 h. For the customer this means that once the satellite has acquired the data, it can be on the desk within 3 days.

Radarsat-2 will have a resolution of 3 m.

Table 31 Radarsat-1

Operational system	Radarsat-1, since 4 November 1995
Orbit	797 km, 98.6°, Sun-synchronous, 24-day repeat cycle
Sensors	SAR, antenna 1.5 by 15 m
Spatial resolution (m)	8–100
Frequency (GHz)	C-band, 5.3
Polarization	HH
Swath width (km)	50–500
Incidence angle (°)	10–60, selectable
Downlink	2 X-band at 85–100 Mbit s^{-1}
Recorders	2 × 10 min of data

Table 32 Envisat-1

Operational system	Envisat-1, to be launched in May 2000
Orbit	800 km, 98.55°, crossing equator at 10 a.m., 35-day repeat cycle
Sensors	ASAR, and many others
Frequency (GHz)	C-band, 5.331
Polarizations	HH and VV, or HH/HV, or VV/VH modes
Spatial resolution (m)	30 or 150
Swath width (km)	56–100 (7 different swaths)
Incidence angle (°)	17–45

6.5.3 Environmental Satellite

In May 2000, the ESA will launch Envisat-1 (Table 32), an advanced polar-orbiting Earth observation satellite, which will provide measurements of the atmosphere, ocean, land, and ice. Envisat-1 is a large satellite carrying a large number of sensors. The most important sensors for land applications are the ASAR, the MERIS, and the Advanced Along-track Scanning Radiometer (AATSR).

ASAR, operating at the C-band, ensures continuity with the image mode SAR and the wave mode of the ERS-1/2 AMI. It features enhanced capability in terms of coverage, a range of incidence angles, polarization, and modes of operation. The ERS-1 and -2 could only be switched on for 10 min each orbit, whereas the ASAR can take up to 30 min of high-resolution imagery each revolution.

In the normal image mode, the ASAR will generate high spatial resolution products (30 m) similar to the ERS SAR. It will image one of the seven swaths located over a range of incidence angles spanning from 15 to 45° in HH or VV polarization. In other modes it is capable of recording the cross-polarization modes (HV and VH). In addition to the 30-m resolution, it offers wide swath modes, providing images of a wider strip (405 km) with medium resolution (150 m).

The ESA is developing two spaceborne imaging spectrometers:

- MERIS to fly on Envisat-1
- Process Research by an Imaging Spectrometer (PRISM) on-board Envisat-2.

Although MERIS delivers only 15 bands, it is a fully fledged imaging spectrometer by design; the 15 bands are fully programmable in central wavelength and bandwidth.

PRISM will cover the 0.4–2.4 μm wavelength range with a 10-nm contiguous sampling interval at a 32-m ground resolution. PRISM can be pointed in all directions, permitting measurement of the BRDF dependence of the wavelength.

Table 33 SIR-C/X-SAR

Operational system	Space Shuttle, Missions STS-59 and STS-68, 9–20 April 1994 and 30 September–11 October 1994, respectively
Orbit	LEO, 225 km, 57°
Sensors	SIR-C and X-SAR
Spatial resolution (m)	25
Frequencies (GHz)	X, C, and L-bands, 9.6, 5.3, and 1.25 GHz, respectively
Polarizations	HH, VV, HV, VH; X-SAR only VV
Swath width (km)	15–60 (15–45 X-SAR)
Incidence angle (°)	20–55

6.5.4 Shuttle Imaging Radar C/X SAR

Together the Shuttle Imaging Radar C (SIR-C, built by NASA's jet propulsion laboratory) and the X-band SAR (X-SAR, jointly built by Germany and Italy) (Table 33) comprise the world's most advanced radar system ever flown. The system is capable of recording radar images at three different wavelengths (X, C, and L-bands) and in four different polarization modes (HH, VV, HV, and VH). The first mission was flown on-board the Space Shuttle flight STS-59 in April 1994 and the second mission was on-board STS-68 in September–October 1994. Roughly 20% of the Earth was imaged at up to 10-m resolution.

NASA plans a Shuttle Radio Detection and Ranging Topography Mission (SRTM) in September 1999 to create a global digital elevation map with 16-m height accuracy at 30-m horizontal intervals. SRTM will use the SIR-C antenna working interferometrically with an additional antenna on a 60-m long deployable mast. Approximately 80% of the Earth's landmass (everything between 60° north and 56° south latitude) will be imaged in 1000 data takes.

6.5.5 Japanese Earth Resources Satellite 1

JERS-1 (Table 34) is the Japanese Earth Resources Satellite. In addition to the SAR system it carries the Optical Sensor (OPS), which has seven downward-looking bands and one forward-viewing band for stereo viewing. The

Table 34 JERS-1

Operational system	JERS-1, 11 February 1992 to 11 October 1998
Orbit	568 km, 97.7°, Sun-synchronous
Sensors	SAR
Frequency (GHz)	L-band, 1.275
Polarization	HH
Resolution (m)	18
Swath width (km)	75
Incidence angle (°)	35

original design lifetime was 2 years, but JERS-1 operated until October 1998, when a short-circuit in its solar panels immobilized it. The next Japanese radar system will be on-board the Advanced Land Observing Satellite (ALOS) to be launched in 2002. The new SAR will have multiple resolution, 10-m high-resolution and 100-m ScanSAR mode, multiple polarization modes (HH and VV), and a variable off-nadir pointing angle capability.

6.6 Future Systems

There is a push for ever-improving resolution: spatial, spectral, radiometric, and temporal resolution. The current trend of building constellations of smaller and smarter microsatellites will certainly be an important factor in this respect.

6.6.1 New Techniques

New techniques for measuring aspects of the Earth's environment will emerge. NASA, for instance, has many innovative sensors and satellites under development. The Space Shuttles will continue to carry experimental sensors that will eventually be carried by operational systems in its Earth Science System. Three EO missions are defined carrying technologies that could revolutionize space-based Earth observations.

EO-1 is scheduled for launch in December 1999. It will demonstrate an ALI system and hyperspectral imaging technologies that may eventually replace the current measurement approach used by Landsat satellites.

EO-2 will fly an IR laser in the cargo bay of the Space Shuttle to demonstrate the capabilities of a space-based lidar to measure accurately atmospheric winds from the Earth's surface to a height of about 15 km. This flight is scheduled for launch in early 2001.

EO-3 will be a geo-stationary system with new technologies such as

- a high-resolution thermal imaging sensor
- a synthetic aperture sounder
- a Fourier transform imaging spectrometer
- a tropospheric trace-gas imager.

Not all these sensors are new, but the idea of flying them in a GEO is new.

Future developments include the following:

- Earth observation will profit from the current boom in communication.
- Miniaturization of satellites to the size of a suitcase weighing 50 kg.
- Application of solid-state sensors and recorders.
- The application of constellations of EO satellites.
- New propulsion methods, e.g. the application of ion motors.
- New attitude control methods, e.g. magnetic torque.
- The application of laser/lidar from space.
- The use of P-band radar from space (currently not allowed by international law).
- The installation of high-resolution sensors in a GEO.

Table 35 Satellite launches in 1999

Spacecraft	Launch date	Country	Remarks	Ground resolution
Landsat 7	15 April 1999	USA	NOAA	15 m
IKONOS-1	27 April 1999	USA	Launch failed	1 m
Fengyun-1C	10 May 99	China	Weather	1.1 km
Oceansat/IRS-P4	25 May 1999	India	Ocean color	350 m
QuikSCAT	19 June 1999	USA	NASA, wind speed	Scatterometer
GOES-L	July 1999	USA	NASA/NOAA, weather	1 km
CBERS 1	Unknown 1999	China/Brazil		20 m
Ikonos-2	Indefinite	USA	Commercial	1 m
EOS AM-1/Terra	August 1999	USA	NASA	15 m
DMSP	August 1999	USA	Defense, weather	500 m
OrbView 3	November 1999	USA	Commercial	1 m
NMP/EO-1	December 1999	USA	NASA	10 m, IS 315 bands
QuickBird	Late 1999	USA	Commercial	0.8 m
NOAA-L	December 1999	USA	Weather	1 km
Eros A	December 1999	USA/Israel	Based on the Ofeq 3	1.5 m
Meteor 3M#1	1999	Russia	Microwave	1.5 km
RESURS-O1#5	September 1999	Russia		30 m
MT-Sat	1999	Japan	GEO	1 km
Eros B	1999	USA/Israel		1 m
Electro-GOMS	1999	Russia	GEO	1.25 km
IRS P5	End 1999	India	a.k.a. Cartosat-1	2.5 m
SAC-C	December 1999	Argentina		35 m

Table 36 Satellite launches beyond 1999

Spacecraft	Launch date	Country	Remarks	Ground resolution
ADEOS-2	2000	Japan		250 m
Envisat-1	May 2000	Europe		30 m
IRS P6	June 2000	India	Resourcesat	10 m PAN, 23 m
Resource21	2000	USA	Commercial	10 m
Aries-1	2000	Australia	Imaging spectrometer, 100 bands	30 m
OrbView 4	2000	USA	Commercial, hyperspectral, 280 bands	1 m
Meteor 3M-2	2000	Russia	GEO	
Shuttle	May 2000	USA	Radar topography mission	25 m
CBERS 2	2000	China/Brazil		20 m
EOS PM-1	December 2000	USA	NASA	250 m
Meteosat 8	2000	Europe	GEO	1 km
Radarsat 2	2000	Canada	Radar	3 m
NOAA-M	April 2001	USA	Weather	1 km
CBERS 3	2002	China/Brazil		3 m
CBERS 4	2002	China/Brazil		3 m
ALOS	2002	Japan	Including radar	2.5 m
SPOT 5	2002	France		5 m
IRS P7	2003	India	Oceansat and Metsat	
IRS 2A	2003	India		
IRS P8	2003	India	Climatsat	

- The use of ''new'' orbits, e.g. the Lagrangian point L1 of the Sun–Earth system or the L4 point of the Earth–Moon system.
- The application of high-speed laser communication.
- The use of satellite-to-satellite communication.
- Earth observation from the ISS (see below).

6.6.2 The International Space Station

The ISS will offer unique observation capabilities. Its orbit covers 75% of the Earth's surface, containing 95% of the Earth's population. It will offer a long-term window on the world and, because astronauts always inhabit it, it offers unrivalled observational capabilities on disaster events. Telescopes can easily be pointed at disaster zones. Currently, the US Laboratory Window Observational Research Facility (WORF) is in the requirements, definition, and conceptual design phase. Many aspects of the Space Shuttle-based EO program will be considered in the design and operation of the WORF.

6.6.3 Future Missions

Tables 35 and 36 are intended to give the reader an impression of the missions that are scheduled for launch in the coming two years (last revision 22 June 1999).

7 REMOTE SENSING IMAGES

A remotely sensed image is a two-dimensional collection of samples. Each of its elements is a pixel (picture element) whose intensity is a DN value. The number of possible values is usually limited to 256 (8 bits) providing 256 brightness values called gray levels in the image.

The value of a pixel is literally the most visible characteristic of a sample. However, in addition to the value, each pixel carries a number of implicit characteristics. Each sample is a measurement taken at a certain location, at a certain time, measuring a certain property of the target object. The measured property is usually the amount of radiation received in a certain spectral range, and is determined by the design of the sensor. The value and the time of acquisition are recorded at the time of imaging. But what about the location of a sample? It appears that the Earth location of a pixel can be determined from its location in the image matrix, but this requires knowledge of the sensor design, of the satellite's position and attitude, and of the shape and topography of the Earth itself. The process of determining the Earth location of pixels is referred to as geo-referencing.

The information content of digital images can be summarized as follows:

- Where? Spatial information
- What? Spectral (and other) information
- How much? Radiometric information
- When? Temporal information.

Both the sampling and the resolution in the spatial, spectral, radiometric, and temporal domains determine the amount of detail that can be observed from a digital image. Although there is a strong mathematical analogy in the concepts of resolution and sampling in these domains, many different factors play a role in this.

For well-balanced systems, the resolution will be of the same order of magnitude as the sampling intervals, which leads to confusion of these terms. For instance, often the term ''temporal resolution'' is used to describe the revisit capability of a remote sensor, when actually ''temporal sampling'' is meant.

Sampling is defined as the interval between successive measurements, and is determined by the following.

- Spatial sampling is determined by the sampling speed of the sensor and the spacing of the elements of the detector array (pushbroom).
- Spectral sampling is determined by the position of the center of the spectral bands. In MS sensors band placing is determined by the sensitivity of the detector and transmission function of whatever filter was applied. Only in hyperspectral instruments could one speak of spectral sampling, and it is then determined by the spacing of the elements of the detector array.
- Radiometric sampling is determined by the dynamic range of the sensor and the number of levels (quantizations) that the ADC can handle.
- Temporal sampling is determined by the orbit repeat cycle of the satellite, together with the off-axis pointing capability of the sensor.

Resolution is defined as the minimum separation between two objects at which they can still be discriminated, and is determined by the following factors.

- Spatial resolution is determined by the aperture (or photon catching area) of the detector and the quality of the optical system used. It can to a certain extent be expressed as a point spread function (PSF) or a MTF. The MTF is actually the Fourier transform of the PSF. The PSF reveals how much an image of an ideal point source is blurred by the sensor. It should be noted that the concept and use of a PSF only has meaning for linear systems. Practical systems hardly ever are such ideal linear systems. The ''size'' of both the PSF and SRF (see next point) functions is sometimes defined by taking the width at half of the maximum value of the function fwhm. The spatial resolution of a scanner system is referred to as the instantaneous field of view (IFOV).
- Spectral resolution is related to bandwidth and is determined by the SRF. The SRF plots the sensitivity of the sensor for all wavelengths. A wide skirt in this function is an indication of bad out-of-band rejection (see spectral accuracy below).
- Radiometric resolution is determined by the ability of the ADC to resolve different levels of input. In practice this is limited by the amount of noise present in the electronics. Noise levels can be expressed as the commonly used signal-to-noise ratio (SNR), which expresses noise level with respect to signal level. A more useful measure is to express the noise level as an equivalent input level to the sensor. Thus the noise can be expressed as the noise equivalent radiance (NER) or noise equivalent temperature difference (NEΔT). Basically the NER determines the smallest difference in input levels that can be discriminated by the sensor.
- Temporal resolution is determined by the dwell time of the sensor. If there is a considerable movement during the dwell time, this will be visible as a motion blur in the result. As the movement of satellites is smooth, very simple measures can be taken to reduce the amount of motion blur caused by the dwell time, e.g. by simply pointing the detector backwards during a scan.

The accuracy is defined as the exactness of the measurement, and is determined by the following factors.

- Spatial accuracy is determined by the mechanical stability of the sensor (scan mirror jitter, thermal stress, film stretch), the quality of the optical system (lens aberration), and pointing stability of the platform (attitude control).
- Spectral accuracy is determined by the band placing (accuracy of the central wavelength), and out-of-band rejection (spectral separation).
- Radiometric accuracy is determined by dark current or offset, nonlinearities in the response function, overload, cross-talk, etc. Furthermore, the response of a sensor may vary with the position in the image. A sensor may exhibit light falloff very much like in a photographic camera. Sensors may degrade with time (aging), and regular calibration is necessary in order to preserve a certain standard in absolute accuracy of the sensor.
- Temporal accuracy is not an issue in RS as normally the interval between the images is of the order of hours and days.

Resolution, sampling, and accuracy to a large extent determine the visible quality of images. However, there are additional external image quality factors, as follows.

- Owing to transmission problems or recording problems there may be missing or incomplete data (dropouts, dropped lines).
- There may be additional noise from various external sources. Spacecraft and sensors can be confused by what is called a single-event upset (SEU). SEUs

occur in space when a single high-energy particle hits the spacecraft electronics, disrupting the spacecraft's normal operations.

- There may be unfavorable atmospheric conditions such as haze and CC.
- There may be differences in down-welling irradiance. For instance, the distance between the Earth and the Sun varies by approximately 1%, which causes a small difference in energy available to remote sensors over the year. Reflective properties can only be determined if the exact amount of down-welling energy is known. This is particularly important for airborne sensors where part of the energy is already lost by the amount of atmosphere overhead. Nowadays, many airborne systems also measure the down-welling energy in addition to the images. It should be noted that characterization of the down-welling irradiance is just as important, and in some respects more difficult, than the characterization of the up-welling measurement part of the sensor.
- There may be glint or hot-spots visible in the image, due to an unfavorable solar angle.

On satellite systems the bandwidth of the downlink enforces a practical limit on the amount of detail. The amount of data is proportional to

- the number of spatial samples (number of pixels, and lines)
- the number of spectral bands
- the number of radiometric levels (number of bits per sample)
- the revisit frequency
- the amount of auxiliary data in the telemetry.

This means that there is a simple tradeoff between these numbers. A satellite with a large coverage has only a limited pixel size. An imaging spectrometer cannot offer a high spatial resolution. High-resolution satellites offer only a limited number of bands, and so on. There is one exception to the rule, namely spaceborne photography. For instance, the KVR-1000 camera has a high resolution but at the same time it has a large coverage. Once in a while it drops a canister of exposed film, the equivalent of an enormous electronic transmission rate. Of course, this comes at a cost, too. Once all the film has been used, it is the end of the mission. The tradeoff between swath width, number of bands, and spatial resolution is immediately visible when we plot the spatial resolution of a number of satellites against the swath width. Figure 16 clearly shows the trade-off between resolution and swath width. High-resolution satellites have a small swath width; medium-resolution Landsat-like satellites have large swath widths. The SPIN-2, using a KVR-1000 camera, is also visible. For imaging spectrometry satellites the limiting factor will be the number of bands.

Future high-speed satellite-to-satellite communications links may increase the downlink data-rate capabilities, relieving some of the constraints on the data volume.

The tradeoff between resolutions also implies that the ideal RS system does not exist! If a good sensor is good

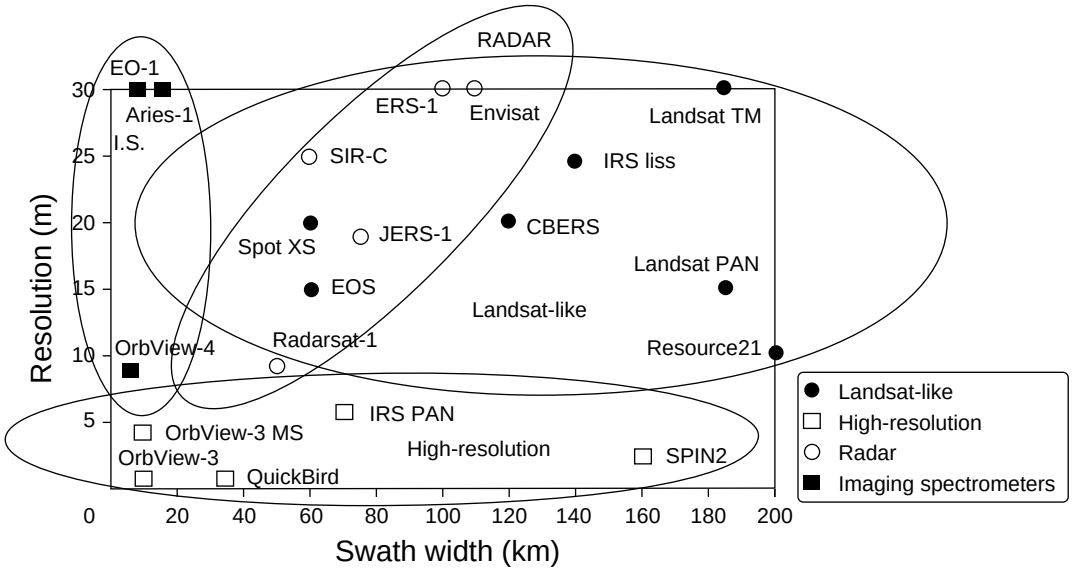

Figure 16 Spatial resolution versus swath width for a number of satellites. The downlink capacity of a satellite imposes a practical limit on the data transfer rate, which means there is a trade-off between spatial resolution, swath width, and number of bands. Indicated by ellipses are four typical classes of satellites. The low-resolution systems are not included in this figure.

for one application this does not mean that it is suitable for another. That is the reason why there are so many systems; they complement each other rather than being competitors in the same market. Even if different sensors have similar characteristics, this is good for the user, because it offers additional opportunities for acquisition. Apart from that, simple economics states that prices go down where there is competition.

7.1 What Type of Satellite Image?

Selecting satellite images for specific projects is not an easy task. Many factors determine whether a satellite image will meet the requirements of a project within the boundary conditions. As pointed out before, trade-offs exist between spatial resolution, spatial coverage, number of spectral bands, radiometric resolution, temporal coverage, and revisit time. The Netherlands, for instance, is 90% covered by two full-scene 30-m resolution Landsat TM images. If the project were to require the 20-m resolution of MS/SPOT, then one would require about 30 SPOT images, costing 10 times more than the Landsat images. Also, the required computing power for processing the images (storage and CPU) would increase by a factor of 10. All this is at the loss of a few spectral bands.

However, there are more factors. For instance, The Netherlands is one of the places where the chance of obtaining one cloud-free image during 10 days is less than 10%. In order to cover the entire country of The Netherlands one might end up having to buy as many as 20 full-frame and quarter-scene Landsat TM images. Clearly, these images can never be acquired in a single day, or even in a single season, and maybe not even in a single year, as the satellite only passes once every 16 days. Obviously, this will result in a number of images having very different lighting conditions, because of the different solar angles in winter, spring, summer, and fall. Also, there will be a large variation in the growing stages of the crops and natural vegetation. All this means that, if one wants to produce an image map of The Netherlands from satellite images, a considerable amount of time will have to be spent on radiometric processing and carefully stitching the images together. The stitching must be done in such a way that the differences between the individual images are not too obvious in the final result.

Of course it *is* possible to acquire a satellite image of the entire country of The Netherlands in a single day by selecting NOAA imagery. One can even have two images each day, but this is at the expense of the resolution of only 1 km.

7.2 Which Image?

Once it is settled what type of imagery one requires for a project, it must be decided *which* images meet the requirements of the project. Fortunately, many data providers have on-line catalogs on the Internet. The user can enter search queries and the system will return information about the availability of imagery. Each entry will list the meta-data (information about the image data) and very often a quick-look (the satellite image in reduced format). The query form has the following entries:

- Acquisition date. The required acquisition date can be entered as a single date or a range of dates. Sometimes a seasonal range can be entered enabling queries such as ''show all April images for the years 1980–1990''.
- Geographic coverage. The required location can be entered as a single point, a box (by entering two points), or a free polygon (by entering many points). Sometimes the scene center can be entered as a path–row pair or an orbit–frame pair.
- Cloud cover (CC). Of course, unless one is a meteorologist, a fully clouded image is worthless, but sometimes one has to accept some clouds simply because cloud-free images are not available. CC can be listed for the entire image, or even for each quarter of the image. Values usually range from 0% (cloud-free) to 100% (fully clouded). Some systems have special codes for cloud (and snow) cover; SPOT designates cloud-free images with an A, codes B, C, and D are used for partly clouded images, and a fully clouded image is coded E. Further, remember that 10% CC is only good if that single cloud is *not* over the study-area! The quick-look image is an almost mandatory item of the image database.

The catalog system returns a list of images matching the query. In addition to the acquisition date, geographic coverage, and CC, the following items are worth noting:

- Technical quality. This is actually the figure for the overall performance of the sensor, platform, transmission, and preprocessing. If this figure is low, one has to be careful; using such an image may yield surprises during later processing and interpretation. However, this does not mean that an image with a low technical quality should not be used at all. This figure should simply be regarded as a sort of warranty statement by the data supplier.
- Quick-look image. This is probably the most interesting part of the image database. The quick-look image is a degraded version of the original image. It is poor enough to prevent its misuse for analysis, but good enough for recognizing features in the area of interest and for checking for clouds and haze.

7.3 Image Processing and Image Products

Effective management by the use of RS data relies on accurate and timely information. The need to manage our activities and their influence on our environment drives an ever-increasing demand for high-quality information.

In order to achieve the highest quality, the data received from the satellite has to undergo a number of preprocessing and processing steps. Preprocessing corrects for known errors and distortions. The result of the preprocessing is image data that can be directly used for further distribution, analysis, and interpretation. However, in order to arrive at project-specific information, further processing may be needed.

Preprocessing steps involve the following.

- Radiometric correction. During the radiometric correction the DN values are transformed into physical units. Noise and systematic errors are removed as much as possible. Radiometric correction relies on the calibration data of the sensor.
- Geometric correction. Although the Earth location of each image point is implicitly known by its relative orientation in the image matrix, most images are transformed in such a way that the image coordinates carry a direct relationship with a map projection. The process of determining the transformation is referred to as geo-referencing; the transformed image is then said to be geo-coded.
- Atmospheric correction. Actually, the correction for atmospheric effects is also some form of radiometric correction, as it operates on the values only. However, the parameters and models used for the atmospheric corrections are so different and so complex that it is justified to separate these corrections from the radiometric correction. An accurate atmospheric correction requires knowledge of the state of the atmosphere at the time of acquisition; parameters such as aerosol, water vapor, and ozone content should be recorded together with the image data. Unfortunately, it is not yet common practice to supply these atmospheric data with the image data. The bare minimum atmospheric correction takes care of haze and skylight.

These corrections are not necessarily done at one and the same location. From satellite to end-user the data may go through one or more of the following facilities:

- The Ground Station or Data Acquisition Facility acquires the raw (telemetry) data transmitted by the satellite, and records it directly on disks for later backup on digital media.
- The Mobile Ground Station. The coverage of the globe by ground receiving stations is not complete. The mobile station is an attractive solution to fill the gaps, but also an efficient means to perform acquisition in a remote location for a long period of time when a lot of images are needed for a particular project. Its use is also of great value to reduce the time between data reception and its operational use.
- The Data Processing Facility processes raw image data in order to generate standard digital image products, generates quick-look data, stores data and meta-data in a catalog, allows for catalog consultation, services user orders, and disseminates the data in a processed form on exchange media such as computer-compatible tape (CCT), CD-ROM, or 8-mm digital video tape.
- The Archiving Center. Sometimes the archiving and dissemination functions are separated from the processing facility into an archiving center.
- The Value Added Facility or Value Added Resellers. These facilities or companies process the standard digital images to value-added digital image products such as ortho-images, mosaics of geocoded images, and digital elevation models (DEMs).
- The User Support Services help the user with catalog lookups, with advice on images and image products, and with satellite programming requests if the user wants the satellite to record new data. User services enable the users to order products, and offer after-sales service.
- The Clearinghouse or Data Warehouse is a one-stop counter for all sorts of spatial data.

The above-mentioned facilities are not always strictly separated; combinations of functions are possible such as the European processing and archiving facilities (PAFs).

The resulting image data are available in a number of processed image products. So-called processing levels designate these products. In most standard imagery the following levels can be found:

- Level 1A, the ''raw'' image data; only the bare minimum of radiometric corrections is performed.
- Level 1B, the standard image data; additional corrections to level 1A are performed, such as corrections for Earth rotation (skew) and curvature. Sufficient auxiliary data are added to enable the user to perform the geo-coding themselves.

These standard products are available as full scene or, to reduce cost, as a quarter scene, or floating scene, and sometimes any subset may be selected from the image. However, one has to realize that many still regard the

''subset'' as a value-added product, which, consequently, should cost a ''value-added'' price. This may change in the near future as flexibility of the archiving centers and the software increases.

The next levels 2A to 2C are the geo-coded products. A higher level means a higher geometric accuracy. These products usually cost a factor of two more, so if the end-user has the equipment (and the time) usually they will perform these steps at their own facilities.

- Level 2A, the systematically geometric corrected product. Only the location and attitude of the satellite are used for the geo-coding. Accuracy is of the order of a few pixels.
- Level 2B, the precision geometric corrected product. To improve accuracy points in the image are related to map locations [ground control point (GCP)]. Accuracy is of the order of 1 pixel on flat terrain.
- Level 2C, the ortho-image. Level 2B does not yet take into account the displacement caused by height in the terrain. Level 2C uses GCP and a DEM or DTM to reach an overall accuracy of less than 1 pixel.

It should be noted that although levels 1A and 1B are often used as a de facto standard, there is no such definition for the higher levels. Levels 2A to 2C are often used, but many other designations and terms are used for various other products, e.g.

- Fused products. For instance, PAN and MS data can be integrated to obtain data that have both a high spatial resolution and a high spectral information content.
- Stereo products. A pair of images taken from different incident angles to permit enable stereoscopy and height measurement.
- Multitemporal product. Images from the same location but from different dates.
- Many levels exist for radar imagery, such as single-look complex (SLC), multi-look complex (MLC), and precision image (PRI).
- Mosaic products, which are composed of many different images stitched together. Studies spanning large areas may require such products. These products require the utmost radiometric ''matching'' of the scenes, the greatest accuracy of geometric correction, and the highest storage capacity.
- Image maps. These are satellite image data presented in a map-like way; they contain a coordinate grid, a legend, meta-data, etc.

7.3.1 Committee on Earth Observation Satellites

The Committee on Earth Observation Satellites (CEOS)[35] was created in 1984 as a result of the international Economic Summit of Industrialized Nations and serves as the focal point for international coordination of space-related Earth observation activities. Policy and technical issues of common interest related to the whole spectrum of Earth observation satellite missions and data received from such missions are addressed. CEOS encourages complementarity and compatibility among space-borne EO systems through coordination in mission planning, promotion of full and nondiscriminatory data access, setting of data product standards, and development of compatible data products, services, and applications.

7.3.2 Committee on Earth Observation Satellites International Directory Network

The CEOS International Directory Network (IDN) is an international effort to assist researchers in locating information on available data sets. The IDN provides open, on-line access to information on worldwide scientific data including Earth sciences (geoscience, hydrospheric, biospheric, satellite RS, atmospheric sciences), space physics, solar physics, planetary science, and astronomy/astrophysics. The IDN describes data held by university departments, government agencies, and other organizations. The search for data sets can be done on-line via one of the local nodes of the IDN, e.g. on NASA's Global Change Master Directory.[36]

For the end-user, CEOS offers two obvious advantages:

- Data query. Descriptions of data sets can be queried through the IDN.
- Data format. Most satellite data comply with the CEOS ''standard''.

However, the last point does not mean that all the data come in the same form! The CEOS format defines a structure for the data and a standard for a description of the meta-data that comes together with the satellite data, but it does not say anything about the meaning of all the entries in the meta-data and the way in which the meta-data should be used. The meta-data information can be anything related to the (image) data, including:

- the data acquisition facility
- the processing facility
- software version
- date and time of acquisition
- size; number of bands, lines, and pixels per line
- calibration data

- sensor-specific data
- Earth location data
- map projection
- orbital elements
- Sun elevation and azimuth angle
- annotation text
- data histograms.

8 CONCLUSION

Pictures of the Earth from space are marvellous images. These images are not just beautiful to look at; they can also be used for a diversity of applications in daily life. With these images floods can be predicted or icebergs detected. With these images a geologist can discriminate details which lead to underground water stores and new deposits of mineral resources. With these images, maps can be made of the vegetation in large areas. Many more applications are possible.

Examining the Earth from space is a wonderful means and a powerful tool. Most scientists are enthusiastic. Some scientists air their criticisms and refuse to see the potential that these images can offer. The best way is to see for yourself and judge whether these images offer anything else besides just beauty, and to try to recognize and understand the many details these images show, possibly aided by proven image processing and GIS techniques.

Earth observation is still in its infancy and the big ''boom'' of RS is yet to come. Possibly, in the near future, the Earth will be circled by constellations of small and low-cost RS satellites that meticulously monitor all aspects of the Earth's surface and atmosphere. The deluge of data that these systems will deliver will hopefully help us to understand the processes that rule this tiny spaceship called Earth.

ABBREVIATIONS AND ACRONYMS

AATSR	Advanced Along-track Scanning Radiometer
ADC	Analog-to-digital Converter
ALI	Advanced Land Imager
ALOS	Advanced Land Observing Satellite
AMI	Active Microwave Instrument
APT	Automatic Picture Transmission
ASAR	Advanced Synthetic Aperture Radio Detection and Ranging
ASTER	Advanced Spaceborne Thermal Emission and Reflectance Radiometer
ATSR-M	Along-track Scanning Radiometer and Microwave Sounder
AVHRR	Advanced Very-high Resolution Radiometer
BRDF	Bi-directional Reflectance Distribution Function
CBERS	China/Brazil Earth Resources Satellite
CC	Cloud Cover
CCD	Charge-coupled Device
CCT	Computer-compatible Tape
CEOS	Committee on Earth Observation Satellites
CRSS	Commercial Remote Sensing System
CSA	Canadian Space Agency
CZCS	Coastal Zone Color Scanner
DEM	Digital Elevation Model
DMSP	Defense Meteorological Satellite Program
DN	Digital Number
DTM	Digital Terrain Model
Envisat	Environmental Satellite
EO	Earth Observing
EOS	Earth Observing System
ERAST	Environmental Research Aircraft and Sensor Technology
ERS	Earth Remote Sensing Satellite
ERTS	Earth Resources Technology Satellite
ESA	European Space Agency
FEWS	Food Early Warning System
FIR	Far-infrared
FOR	Field of Regard
FOV	Field of View
fwhm	Full Width at Half-maximum
GAC	Global Area Coverage
GCP	Ground Control Point
GEO	Geo-stationary Orbit
GIS	Geoinformation System
GOME	Global Ozone Monitoring Experiment
HRCC	High-resolution Charge-coupled Device
HRG	High-resolution Geometric
HRPT	High-resolution Picture Transmission
HRV	High-resolution Visible
HRVIR	High-resolution Visible/Infrared
IDN	International Directory Network
IFOV	Instantaneous Field of View
INSAR	Synthetic Aperture Radio Detection and Ranging Interferometry
IR	Infrared
IRMSS	Infrared Multispectral Scanner
IRS	Indian Remote Sensing Satellite
ISS	International Space Station
JERS	Japanese Earth Resources Satellite
LAC	Local Area Coverage

laser	Light Amplification by Stimulated Emission of Radiation
LEISAC	Linear Etalon Imaging Spectrometer Atmospheric Corrector
LEO	Low Earth Orbit
lidar	Light Detection and Ranging
LIS	Lightning Imaging Sensor
LISS	Linear Imaging Self-Scanning Sensor
MERIS	Medium Resolution Imaging Spectrometer
MLC	Multi-look Complex
MODIS	Moderate-resolution Imaging Spectroradiometer
MS	Multispectral
MSG	Meteosat Second Generation
MSS	Multispectral Scanner
MTF	Modulation Transfer Function
MWIR	Mid-wave Infrared
NASA	National Aeronautics and Space Administration
NEO	Near-Earth Orbit
NER	Noise Equivalent Radiance
NEΔT	Noise Equivalent Temperature Difference
NIR	Near-infrared
NMP	New Millennium Program
NOAA	National Oceanic and Atmospheric Administration
NORAD	North American Aerospace Defense Command
OBTR	On-board Tape Recorder
OPS	Optical Sensor
OSA	Optical Sensor Assembly
PAF	Processing and Archiving Facility
PAN	Panchromatic
PRI	Precision Image
PRISM	Process Research by an Imaging Spectrometer
PSF	Point Spread Function
QBM	QuickBird Multispectral
QBP	QuickBird Panchromatic
RA	Radio Detection and Ranging Altimeter
RAAN	Right Ascension of Ascending Node
radar	Radio Detection and Ranging
RBV	Return Beam Vidicon
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SEU	Single-event Upset
SIR-C	Shuttle Imaging Radar C
SLAR	Side-looking Airborne Radio Detection and Ranging
SLC	Single-look Complex
SNR	Signal-to-Noise Ratio
SOHO	Solar and Heliospheric Observatory
SPIN-2	Space Information – 2 m Resolution
SPOT	Système Pour l'Observation de la Terre
SRF	Spectral Response Function
SRTM	Shuttle Radio Detection and Ranging Topography Mission
SSR	Solid-state Recorder
SWIR	Short-wave Infrared
TDRS	Tracking and Data Relay Satellite
TIR	Thermal Infrared
TIROS	Television and Infrared Observation Satellite
TLE	Two-line Elements
TM	Thematic Mapper
TRMM	Tropical Rainfall Measuring Mission
UAV	Unmanned Aerial Vehicle
UV	Ultraviolet
VIS	Visible Range
VISSR	Visible and Infrared Spin Scan Radiometer
VNIR	Visible and Near-infrared
WFI	Wide-field Imager
WiFS	Wide-field Sensor
WMO	World Meteorological Organization
WORF	Laboratory Window Observational Research Facility
WRS	World Reference System
WV	Water Vapor Band
6S	Second Simulation of the Satellite Signal in the Solar Spectrum

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Airborne Instrumentation for Aerosol Measurements

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Hyperspectral Remote Sensing: Data Collection and Exploitation • Imaging Spectrometry for Geological Applications • Land Cover Assessment and Monitoring • Processing and Classification of Satellite Images

FURTHER READING

H.J. Buiten, J.G.P.W. Clevers (eds.), *Land Observation by Remote Sensing: Theory and Applications*, eds. H.J.

Buiten, J.G.P.W. Clevers, Gordon and Breach, Zurich, 1993.
R.N. Colwell (ed.), *Manual of Remote Sensing*, Sheridan Press, Falls Church, Vols. I and II, 1983.
A.P. Cracknell, L.W.B. Hayes, *Introduction to Remote Sensing*, Taylor & Francis, London, 1991.
R. Harris, *Satellite Remote Sensing – An Introduction*, Routledge & Kegan Paul, London, 1987.
E. Hyatt, *Keyguide to Information Sources in Remote Sensing*, Mansell Publishing, London, 1988.
K.B. Kidwell, *NOAA Polar Orbiter Data Users Guide*, NOAA/NESDIS/NCDC/SDSD, Washington, DC, 1991.
P.M. Mather, *Computer Processing of Remotely Sensed Images – An Introduction*, Wiley, Chichester, 1987.
J.P. Muller (ed.), *Digital Image Processing in Remote Sensing*, Taylor & Francis, London, 1988.
J.A. Richards, *Remote Sensing Digital Image Analysis – An Introduction*, Springer, Berlin, 1993.
R.A. Schowengerdt, *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, New York, 1983.
N.H. Short, *The Landsat Tutorial Workbook – Basics of Satellite Remote Sensing*, NASA Reference Publication 1078, NASA, Washington, DC, 1982.
B.S. Siegal, A.R. Gillespie (eds.), *Remote Sensing in Geology*, Wiley, New York, 1980.
P.H. Swain, S.M. Davis (eds.), *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York, 1978.
R. Turnill (ed.), *Jane's Spaceflight Directory*, Jane's Publishing Company, London, 1986.

REFERENCES

1. F.F. Sabins, *Remote Sensing: Principles and Interpretation*, 3rd edition, W.H. Freeman, New York, 1996.
2. T.M. Lillesand, R.W. Kiefer, *Remote Sensing and Image Interpretation*, 3rd edition, Wiley, New York, 1994.
3. Principles Relating to Remote Sensing of the Earth from Space, United Nations Resolution 41/65, adopted in 1986, http://www.un.or.at/OOSA/treat/rs/rs.html.
4. S. Aronoff, *Geographic Information Systems: A Management Perspective*, WDL Publications, Ottawa, 1989.
5. K.R. McCloy, *Resource Management Information Systems: Process and Practice*, Taylor & Francis, London, 1995.
6. J. Gribbin, *Companion to the Cosmos*, Weidenfeld & Nicolson, London, 1996.
7. C. Elachi, *Introduction to the Physics and Techniques of Remote Sensing*, Wiley, New York, 1987.
8. H.J. Kramer, *Observation of the Earth and its Environment: Survey of Missions and Sensors*, 3rd edition, Springer, Berlin, 1996.
9. W.G. Rees, *Physical Principles of Remote Sensing*, Cambridge University Press, Cambridge, 1990.
10. E. Vermote, D. Tanré, J.L. Deuzé, M. Herman, J.J. Mocrette, 'Second Simulation of the Satellite Signal and the Solar Spectrum, 6S User Guide', ftp://loaser.univ-lille1.fr/6S.
11. G. D'Souza, A.S. Belward, J.P. Malingreau (eds.), *Advances in the Use of NOAA AVHRR Data for Land Applications*, Kluwer, Dordrecht, 1996.
12. A.C. Clarke, 'Extra-terrestrial Relays', *Wireless World*, 305–308 (1945).
13. TerraServer, http://www.terraserver.com.
14. *Manual of Remote Sensing. Vol. I. Theory, Instruments and Techniques*, eds. F.J. Janza, H.M. Blue, J.E. Johnston, *Vol. II. Interpretation and Applications*, eds. R.G. Reeves, A. Anson, D. Landen, American Society of Photogrammetry, Falls Church, 1975.
15. C.C. Slama, C. Theurer, S.W. Henriksen (eds.), *Manual of Photogrammetry*, 4th edition, American Society of Photogrammetry, Falls Church, 1980.
16. TOPEX/Poseidon, http://topex-www.jpl.nasa.gov.
17. R. Graham, R.E. Read, *Manual of Aerial Photography*, Focal Press, London, 1986.
18. NASA Scientific Balloons, http://pao.gsfc.nasa.gov/gsfc/lithos/balloon/balloon.htm.
19. NASA Sounding Rocket Program, http://lepjas.gsfc.nasa.gov/rocket/RocketProgram.html.
20. P. Clark (ed.), *Jane's Space Directory 1997–98*, 13th edition, Jane's Information Group, Alexandria, 1997.
21. A.P. Cracknell, *The Advanced Very High Resolution Radiometer AVHRR*, Taylor & Francis, London, 1997.
22. X. Jia, J.A. Richards, 'Binary Coding of Imaging Spectrometer Data for Fast Spectral Matching and Classification', *Remote Sens. Environ.*, **43**, 47–53 (1993).
23. K. Okada, A. Iwashita, 'Hyper-multispectral Image Analysis Based on Waveform Characteristics of Spectral Curves', *Adv. Space Res.*, **12**(7), 433–442 (1992).
24. A. Graps, 'An Introduction to Wavelets', *IEEE Comput. Sci. Eng.*, **2**(2), 50–61 (1995).
25. P.M. Atkinson, A.R.L. Tatnall, 'Neural Networks in Remote Sensing', *Int. J. Remote Sensing*, **18**(4), 699–709 (1997).
26. E.M. Rollin, E.J. Milton, 'Processing of High Spectral Resolution Reflectance Data for the Retrieval of Canopy Water Content Information', *Remote Sens. Environ.*, **65**, 86–92 (1998).
27. S. Jacquemoud, F. Baret, B. Andrieu, F.M. Danson, K. Jaggard, 'Extraction of Vegetation Biophysical Parameters by Inversion of the PROSPECT + SAIL Models on Sugar Beet Canopy Reflectance Data. Application to TM and AVIRIS Sensors', *Remote Sens. of Environ.*, **52**(3), 163–172 (1995).
28. J.K. Crowley, D.W. Brickey, L.C. Rowan, 'Airborne Imaging Spectrometer Data of the Ruby Mountains, Montana: Mineral Discrimination Using Relative Absorption Band-depth Images', *Remote Sens. Environ.*, **29**, 121–134 (1989).

29. R.N. Clark, A. Gallagher, G.A. Swayze, 'Material Absorption Band Depth Mapping of Imaging Spectrometer Data Using a Complete Band Shape Least-squares Fit with Library Reference Spectra', in *Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, Pasadena, USA*, ed. R.O. Green, NASA–JPL Publication 90–54, NASA–JPL, Pasadena, CA, USA, 176–186, 1990.
30. F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H. Goetz, 'The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data', *Remote Sens. Environ.*, **44**, 145–163 (1993).
31. J.J. Settle, N.A. Drake, 'Linear Mixing and the Estimation of Ground Cover Proportions', *Int. J. Remote Sens.*, **14**, 1159–1177 (1993).
32. W.H. Farrand, J.C. Harsanyi, 'Mapping the Distribution of Mine Tailings in the Coeur d'Alene River Valley, Idaho, Through the Use of a Constrained Energy Minimization Technique', *Remote Sens. Environ.*, **59**, 64–76 (1997).
33. H. Cetin, D.W. Levandowski, 'Interactive Classification and Mapping of Multi-dimensional Remotely Sensed Data Using *n*-Dimensional Probability Density Function (*n*PDF)', *Photogramm. Eng. Remote Sens.*, **57**(12), 1579–1587 (1991).
34. F. Van der Meer, W. Bakker, 'Cross Correlogram Spectral Matching (CCSM): Application to Surface Mineralogical Mapping Using AVIRIS Data from Cuprite, Nevada', *Remote Sens. Environ.*, **61**(3), 371–382 (1997).
35. CEOS, http://www2.ncdc.noaa.gov/CEOS.
36. GCMD, http://GCMD.gsfc.nasa.gov.

Sea Ice Monitoring by Remote Sensing

Stein Sandven, Ola M. Johannessen, and Kjell Kloster
Nansen Environmental and Remote Sensing Centre, Bergen, Norway

1 Introduction **1**
1.1 The Role of Sea Ice in the Climate and Weather System 1
1.2 Sea Ice as a Barrier for Ship Traffic, Fisheries and Offshore Operations 3
2 Brief History of Sea Ice Monitoring **4**
3 Physical Properties of Sea Ice **5**
3.1 Large-scale Ice Parameters Defined by the World Meteorological Organization 5
3.2 Small-scale Ice Structure and Growth of Ice 6
4 Overview of Sea Ice Remote Sensing Methods **8**
4.1 Principal Methods, Instruments and Surface Characteristics 8
4.2 Electromagnetic Properties of the Macroscopic Components 9
4.3 Basic Remote Sensing Parameters 10
5 Imaging Radar Systems **12**
5.1 Basic Properties of Imaging Radar Systems 12
5.2 Some Important Parameters and Properties 12
5.3 The Radar Equation 13
6 Case Study of Passive Microwave Remote Sensing of Sea Ice **15**
6.1 Observation of Brightness Temperature 15
6.2 Algorithms for Retrieval of Ice Parameters 16
6.3 Example of Ice Analysis from Special Sensor Microwave Imager Data 17
7 Case Study of Active Microwave Remote Sensing of Sea Ice **17**
7.1 General Interaction Between Synthetic Aperture Radar and Sea Ice 17
7.2 Synthetic Aperture Radar Data Characteristics 19
7.3 Joint Analysis of Synthetic Aperture Radar Images and In Situ Measurements 19
7.4 Summer Conditions 20
7.5 Conclusion from the Synthetic Aperture Radar Ice Classification Study 21
8 Future Perspectives on Sea Ice Remote Sensing **21**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Sea ice *is defined as ice which is formed as a result of freezing of seawater. Sea ice occurs at the surface of the ocean in areas where the surface temperature is cooled to the freezing point, which is about* $-1.8^{\circ}C$ *for sea water with a salinity of about 35 parts per thousand. Ice formed in lakes and rivers and icebergs coming from glaciers and ice sheets are not defined as sea ice.*

Monitoring by remote sensing *is defined as any measurement technique which can be used to observe sea ice repeatedly by instruments on board earth observation satellites. Monitoring also includes the process of making observational data available for users a short time after the observations have been obtained (i.e. a few hours). Remote sensing of ice can also be done with aircraft, submarines and other vehicles, but these techniques are not discussed in this article.*

1 INTRODUCTION

1.1 The Role of Sea Ice in the Climate and Weather System

Sea ice is a part of the cryosphere which interacts continuously with the underlying oceans and the overlaying atmosphere. The growth and decay of sea ice occur on a seasonal cycle at the surface of the ocean at high latitudes. As much as 3×10^7 km^2 of the earth's surface can be covered by sea ice. In the Northern Hemisphere, the extent of sea ice (area enclosed by the ice boundary) fluctuates each year from a minimum in September, when most of the ice is confined to the central Arctic Ocean, Greenland Sea and Canadian Archipelago, to a maximum in March, when the ice covers almost the entire Arctic Ocean and many adjacent seas. In the Southern Hemisphere, the annual fluctuation is even greater, from a minimum in February to a maximum in September when the ice surrounds the Antarctic continent and extends equatorward

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

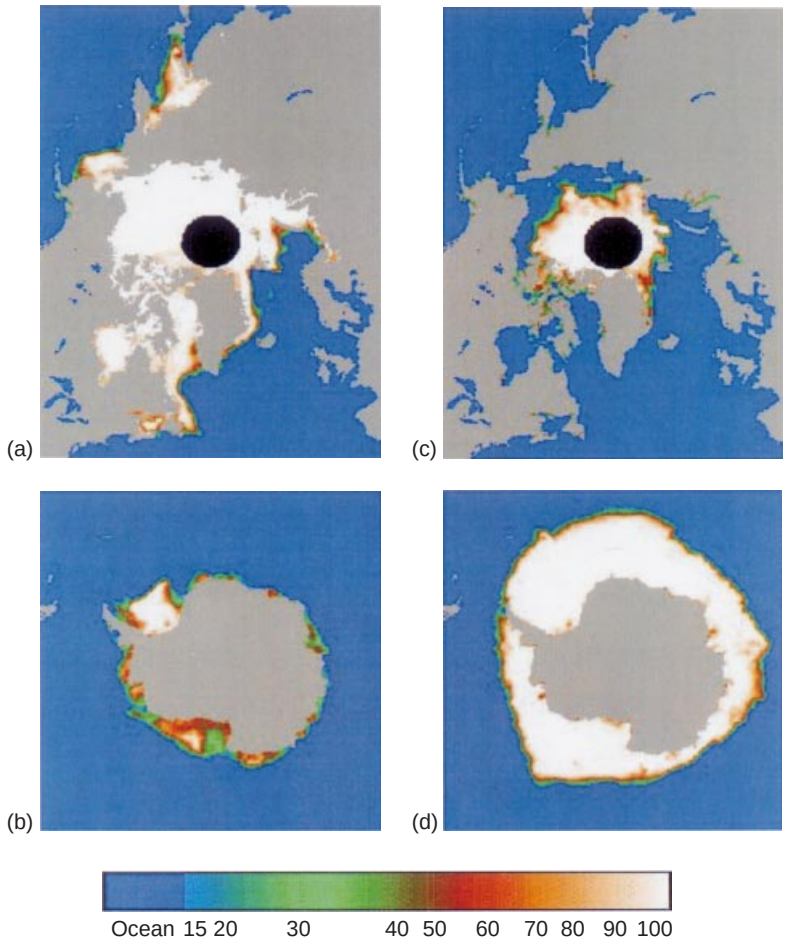

Figure 1 Maps of maximum and minimum sea ice extent and concentration in the Arctic and Antarctic obtained from passive microwave (PMW) satellite data from the Special Sensor Microwave Imager (SSM/I). The maps are from March 1993 [(a) and (b)] and September 1993 [(c) and (d)]. Ice concentrations, indicated by the color bar, were computed from SSM/I data using the Norwegian Remote Sensing Experiment (NORSEX) algorithm.[2]

to 55–65 °S.[1] Figure 1(a–d) shows an example of the maximum and minimum ice extent observed by PMW satellite data.

The largest volume of sea ice is found in the northern hemisphere in March, $0.05 \times 10^6\,km^3$, which is nearly twice the maximum ice volume in the southern hemisphere. The reason for this is the mean thickness of the Arctic sea ice, which is about 3 m, whereas the mean thickness of the Antarctic sea ice is 1–1.5 m.

Sea ice research and monitoring are important issues for many countries at high latitudes, including those who operate in Antarctica. Sea ice imposes severe restrictions on ship traffic in Arctic countries, it is a sensitive climate indicator and it plays an important role in exploration and exploitation of marine resources.

Sea ice has many roles in the global climate system. For one, it serves as an effective insulator between the ocean and the atmosphere, restricting the exchange of heat, mass, momentum and chemical constituents. During winter when there is a large temperature difference between the cold atmosphere and the relatively warm ocean surface, ocean-to-atmosphere heat transfer is essentially limited to areas of open water and thin ice within the pack. The winter flux of oceanic heat to the atmosphere from open water can be two orders of magnitude larger than the heat flux through an adjacent thick ice cover. As a result, the distribution of open water and thin ice is particularly important to the regional heat balance.

Another important role of sea ice in the global climate system is that it affects surface albedo. Ice-free ocean generally has albedos below 10–15%, whereas snow-covered sea ice albedos average about 80%. A fresh snow cover on the ice can increase the surface albedo to values as high as 98%, whereas melt ponds can decrease the ice albedo to as low

as 20%. Because the albedo of snow-covered sea ice is high, relative to that of open water, the presence of sea ice considerably reduces the amount of solar radiation absorbed at the earth's surface. This is most significant in summer, when the insolation, or solar heating, is high.

Sea ice processes also affect oceanic circulation directly by the rejection of salt to the underlying ocean during ice growth. This increases the density of the water directly under the ice, thereby inducing convection that tends to deepen the mixed layer. This convection contributes to driving the thermohaline circulation of the ocean and, in regions with density structures that were initially weak or unstable, can lead to overturning and deep water formation. Much of the world oceans' deep and bottom water is believed to be formed in polar latitudes by these mechanisms. Conversely, the input of relatively fresh water to the ocean during ice melt periods tends to increase the stability of the upper layer of the ocean, inhibiting convection. Furthermore, the net equatorward transport of ice in each hemisphere produces a positive freshwater transport and a negative heat transport.

On a hemispheric scale, the seasonal variability of ice extent and ice edge location is controlled by atmospheric and oceanic forcing, which include ocean temperature and salinity and atmospheric temperature and winds. The location of the ice edge will in turn feed back on several atmospheric and oceanic processes, which have an effect on the regional weather, such as the generation of polar lows. On a regional scale surface roughness of the ice and the drag coefficient depend upon ridging and rafting, both of which can be produced by wind- or wave-induced ice convergence. Ice observation is therefore important for weather forecasts at high latitudes.

Data from polar orbiting satellites are used extensively in research and in monitoring of sea ice extent and other ice parameters.[3,4] Ice thickness, which is another important ice parameter, cannot be measured directly by any spaceborne instrument today.[5] The key objective in sea ice science during the next decade is to achieve the capability for synoptically measuring sea ice thickness in both hemispheres. Data on ice thickness are very sparse, especially in Antarctica. Our estimates of sea ice volume, which are mainly based on model results owing to a lack of data, can have errors of ±50%. In the Arctic there are a few synoptic surveys of ice thickness obtained by submarine sonar and point measurements from expeditions and drifting ice stations. The data sets available provide some information about regional and seasonal variability, but they are too sparse to provide any information about temporal changes.

General circulation models predict enhanced climatic warming in polar areas, and this should be reflected in a decreased mean sea ice thickness and possible changes in the intensity of pressure ridging. Without a technique to map sea ice thickness routinely, however, such changes cannot be detected. Only a satellite-borne method can achieve the required coverage in time and space without prohibitive cost, and this will have to be a technique which measures topography directly, since imaging methods (active microwave (AMW) PMW, visual, thermal infrared (TIR)) do not yield thickness but only an identification of ice type.

1.2 Sea Ice as a Barrier for Ship Traffic, Fisheries and Offshore Operations

The presence of sea ice represents a major limitation for ships and offshore operations in high latitudes in both hemispheres. The sea ice, which is on average 2–3 m thick, can only be penetrated by ice-strengthened vessels or icebreakers with sufficient ice class. Most ships and fishing vessels are not ice-strengthened and must therefore avoid all ice areas. In many cases, when the ice concentration is 100% and the ice pressure is high, even the most powerful icebreakers can have problems moving forward through the icepack. Offshore platforms for ice-covered areas must have much stronger construction than that required in ice-free waters. Also, harbors and loading terminals on the coast require stronger construction in areas of sea ice. In countries and regions where sea ice occurs it is therefore of primary importance to monitor the sea ice regularly and produce ice forecasts to assist ship traffic, fisheries and other marine operations.[6] In the Northern Sea Route, the longest ice navigation route in the world, Russia has built up an extensive ice service to support the ship traffic.[7] The service includes ice monitoring and forecasting as well as icebreaker escorting of cargo ships (Figure 2).

Figure 2 Picture of Russian nuclear-powered icebreaker *Sibiria* leading a convoy of cargo vessels sailing through 2 m thick first-year (FY) ice in the Northern Sea Route.[6]

2 BRIEF HISTORY OF SEA ICE MONITORING

Sea ice observation from coastal stations and ships has a history of more than 100 years. Regular sea ice charting, however, using aircraft and satellites, has developed mostly after the Second World War. Aircraft surveys has been the main observation method, but the use of satellite data has developed gradually over the last three decades as a consequence of the instruments and observation techniques available from satellites.

The first satellite sensors providing views of the large-scale structure and motion of sea ice utilized visible and TIR channels such as those on board the early Nimbus, Tiros and Earth Resources and Technology Satellite (ERTS, later renamed Landsat). By the late 1960s, it was apparent that the sequential synoptic observations needed for sea ice and climate studies could not be acquired by satellite-borne visible and infrared (IR) sensors, which are limited to cloud-free and well-illuminated conditions. Sea ice exists in regions which are dark for several months and are frequently cloudy in the remaining months.[1]

Therefore, it has been necessary to develop observation methods using microwaves which are able to penetrate clouds and are not dependent on light conditions. The first PMW remote sensing systems for satellites were launched on the Russian Cosmos 243 and Cosmos 384 in 1968 and 1970, respectively. In the USA, PMW technology was first used in remote sensing of sea ice during the early 1970s, when the Electrically Scanning Microwave Radiometer (ESMR) was flown on Nimbus-5 over the Arctic.[8]

The period since 1970 has been one of great advancement in remote sensing of sea ice. After the ESMR period 1973–76, a more advanced satellite instrument, the scanning multichannel microwave radiometer (SMMR) was operated on Nimbus-7 for 9 years, from 1978 to 1987. This instrument provided the longest and first regular time series of global sea ice data. PMW observations have fairly coarse resolution (typically 30 km) and are more suitable for large-scale or global monitoring than for regional and local observations. AMW systems, such as side-looking radar (SLR) and synthetic aperture radar (SAR) were developed for aircraft surveillance and used in ice monitoring to provide detailed maps of the ice conditions, especially in areas of heavy ship traffic.

In 1978, Seasat was the first satellite which provided high-resolution SAR images of sea ice, but the satellite operated for only about 3 months. With spaceborne SAR data, which combines high spatial resolution with independence of cloud cover and light conditions, it is possible to observe sea ice with much better accuracy than visible, TIR and PMW methods. European Remote Sensing Satellite (ERS-1) represented a major milestone in satellite SAR remote sensing from 1991, because the satellite delivered tens of thousands of SAR images of sea ice from most ice covered regions in the world. The satellite SAR technology is being improved and SAR systems offered by the Canadian RADARSAT and the European Environmental Satellite (ENVISAT) satellites have wide swath, multimode and dual polarization. Other microwave systems such as scatterometer and SLR data have also shown promising results in sea ice observations.

In order for SAR and other satellite data to become central elements in ice observation, products useful for a wide range of users, interpretation techniques and algorithms must be improved and streamlined. The data products must be developed based on requirements from established users.

The status of sea ice monitoring can be summarized as follows: it is

- of national importance in countries whose coasts are ice-infested during the winter;
- of considerable economic importance in sea transportation;
- one of the most important and appropriate applications of satellite data with potential to increased use as new satellite data becomes available;
- an established and organized activity in many countries where sea ice has an impact on sea transportation;
- important for weather forecasting;
- important for climate change detection.

Use of satellite data in sea ice monitoring is developing steadily. The main aspects of this development are as follows:

- satellite methods using visible/IR data and PMW data are used operationally in many countries;
- the introduction of spaceborne SAR data represents a major milestone and improvement of ice monitoring quality;
- synergy with other data (aircraft, in situ, etc.) is essential for optimum extraction of ice information from satellite data;
- in many regions only satellites can provide data because aircraft and ships cannot cover the whole Arctic or Antarctic ice area;
- the market for sea ice data and related services is expected to grow because of increased activities and more focus on environmental issues in polar regions;
- considerable improvements of ice monitoring products and services are needed, which require more use of SAR data from satellites.

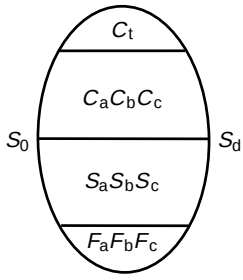

C_t	Total concentration of ice in area, reported in tenths
$C_aC_bC_c$	Partial concentration in tenths of thickest (C_a), second thickest (C_b) and third thickest (C_c) ice types with C_a, C_b and C_c 1/10 or more. If only one thickness type is present it equals C and the second level is left blank
$S_aS_bS_c$	Stage of development (age) of ice concentration reported by C_a,C_b and C_c
$F_aF_bF_c$	Predominant form of ice (floe size) corresponding to S_a,S_b and S_c, respectively
S_aS_b	Development stage (age) of remaining ice types. S_0 if reported is a trace of ice type thicker/older than S_a.S_d is a thinner ice type which is reported when there are four or more ice thickness types

$F_aF_bF_c$	Form of ice	Width
0	Pancake	
1	Brash	
2	Ice cakes	<20 m
3	Small floe	20–100 m
4	Medium floe	100–500 m
5	Big floe	500–2000 m
6	Vast floe	2–10 km
7	Giant floe	>10 km
8	Fast ice	
9	Icebergs	
X	No form	
C	Ice in strips in which concentration is C	

$S_aS_bS_c$	Stage of development	Thickness
1	New	<10 cm
2	Nilas	<10 cm
3	Young	10–30 cm
4	Gray	10–15 cm
5	Gray–white	15–30 cm
6	Firstyear	>30 cm
7	Thin firstyear/white	30–70 cm
1	Medium firstyear	70–120 cm
4	Thick firstyear	>120 cm
7	Old	
8	Secondyear	No defined ranges
9	Multiyear	
Δ	Icebergs	

Figure 3 The WMO's standard sea ice nomenclature and its graphical presentation in the Egg code.

3 PHYSICAL PROPERTIES OF SEA ICE

3.1 Large-scale Ice Parameters Defined by the World Meteorological Organization

The definition of sea ice terminology has been standardized by the Sea Ice Working Group of the World Meteorological Organization (WMO),[9] which established a nomenclature used by all institutions producing ice maps today. The main ice parameters, with examples of their values, are shown in Figure 3. The parameters are shown with symbols, such as the Egg code, in ice charts produced by operational ice centers.

The WMO nomenclature does not include the ice characteristics on smaller scale, i.e. on scales from centimeters to millimeters, which are important for several remote sensing methods. Limitations of the conventional ice codes from an ice navigation point of view are according to Lensu et al.[10]

1. The classification of ice types is unnecessarily detailed for younger and thinner ice types which exist only for shorter periods of time in the freeze-up and early winter season. On the other hand, thicker, deformed and multiyear (MY) ice types are not characterized in as much detail as is needed. The classification should be proportional to the decrease in ship speed and increase in damage probability.
2. There is no quantitative reference to deformed ice types such as rafted ice, in spite of the fact that deformed ice can be predominant and several times thicker than level ice.
3. There is no quantitative reference to ice ridges or to their size and frequency of occurrence.
4. There is no reference to lead size, frequencies or orientations.
5. The relation of regional ice characteristics to what is experienced by an ice-going vessel is uncertain.

6. The codes cannot optimally use the information that is available from SAR images.
7. The terminology has no clear connection to geophysical ice models used in forecasting.

The main reason for these shortcomings is that the WMO nomenclature was defined in 1970, when

- no operative ice models existed;
- no high-resolution satellite data were available; and
- very few data on ice thickness, floe size and ridge distribution existed.

The code is clearly aimed to display visual and mainly qualitative observations, for example those made onboard a vessel. On the other hand, it is not feasible that a single ice code would satisfy all possible requirements. According to Lensu et al.,[10] a possible scenario for the future development of the ice codes could therefore be threefold:

- a geophysical code with theoretically sound concepts which can be related to satellite imagery and ice modeling;
- a navigational code which can be related to ship speed and damage probability; and
- an observational code as a further development of the Egg code.

3.2 Small-scale Ice Structure and Growth of Ice

A complete description of ice freezing and melting, the main physical processes responsible for the large seasonal variability in sea ice extent and volume, should start with a discussion of the structure of the H_2O molecule as it changes in the phase transition between solid and liquid. The fact that solid water (ice) has a lower density than liquid water, which is caused by this change in molecular structure, has the important implication that sea ice floats on top of the ocean. In this section, we shall describe only the most important physical properties very briefly, focusing on those which are important for remote sensing. The basic physical parameters of sea ice are temperature, salinity, a crystal structure which incorporates brine and air, surface roughness, snow cover and the presence of liquid water on top of the ice, which frequently occur in the summer time when ice and snow melts at the ice surface.

Since sea ice is formed by freezing of salt water, there are important effects of the salt during the freezing process: some salt is released from the ice which is formed, and some ice is trapped in brine pockets of varying size. Ice formed during a winter season (FY ice) contains typically from 6 to 10‰ salt if the ice is formed from normal seawater with salinity of 35‰. The brine gradually drains from the upper part of the ice, causing MY ice to have a salinity of less than 1‰ in the surface layer, which is practically fresh water.

In the first stage of freezing, small ice crystals are formed at the surface, called frazil ice. As freezing continues and more ice crystals are formed, the crystals coagulate to form a soupy layer called grease ice. Frazil ice and grease ice dampen the short gravity waves at the sea surface, which has a significant impact on radar remote sensing of open ocean. An example of this dampening is shown in the photograph in Figure 4(a). If the freezing is allowed to continue without disturbance from surface waves, a weakly consolidated layer of elastic ice is formed. When this layer is less than 10 cm thick, it is called nilas. As it grows thicker and becomes less elastic, it is called gray ice (10–15 cm) and gray–white ice (15–30 cm). The process of brine drainage gradually replaces the brine pockets with voids of air which change the ice's visual appearance from almost black nilas to bright grey and grey–white ice, which is significant in visual remote sensing of ice.

When wind and waves act on the ice freezing process, which is the normal situation in the marginal ice zone, the grease ice forms pancakes, which are small, rounded floes of elastic ice with a size from 30 cm to 3 m in diameter. The wave action causes continuous collisions between these floes, which generate raised rims which are typically 5–20 cm high (Figure 4b). The rims of the pancakes are

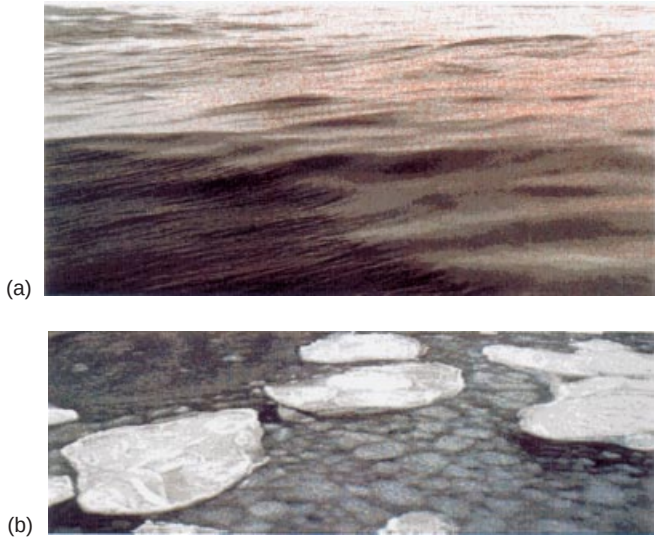

Figure 4 (a) Grease ice which dampens the short surface waves observable in the ice-free area on the left side of the image. (b) Pancake ice of different age. The smaller floes are 30–50 cm in diameter and 1–2 days old. The larger floes, 1–2 m in diameter, are several days old and composed of clusters of smaller pancakes. The photographs were taken by the Nansen Center during the Seasonal Ice Zone Experiment (SIZEX)'92 experiment.

efficient scatterers of radar waves used in remote sensing of sea ice.

In addition to the thermodynamic processes (freezing and melting) and the effects of the salt contents, there are important mechanical forces on the ice from winds and ocean currents, which cause the ice cover to converge or diverge. Convergence causes the formation of ridges, rafted ice and underwater keels, while divergence generates openings in the ice cover, called leads and polynyas. The leads and polynyas normally have open water or thin ice (nilas or grey ice) at the surface, which is detectable by most remote sensing methods if the resolution is good enough. The variability in surface roughness generated by ridging, rafting or other processes, such as pancake formation, is most readily observed by radar remote sensing data. The radar backscatter attains higher values in areas of high surface roughness than in areas of smooth ice surface.

In most areas, the sea ice is covered with layers of snow which have impact on the remote sensing signal, primarily the visual methods because the snow cover determines the surface albedo. Microwave remote sensing waves can penetrate snow unless the snow is wet or there are several layers of snow separated by thin ice. In regions with heavy snowfall, such as the East Greenland marginal zone, the ice can be loaded with wet heavy snow, which causes the whole ice floe to be flooded by seawater. When this occurs, the microwave remote sensing signals from the ice change. At the snow–ice interface, the snow is quickly recrystallized, which also has an effect on the radar return signal. The snow cover undergoes a characteristic seasonal cycle which affects the microwave remote sensing observations of sea ice (illustrated in Figure 5a–d). In the winter season, lasting 6–8 months from late autumn to the spring, defined by temperatures well below 0 °C for both snow and ice, the snow cover is dry and practically transparent to microwaves. In late spring and early summer, when the temperature approaches 0 °C, the snow becomes humid and eventually moist, which creates melt pools at the interface between the snow and ice. Continued warming during the summer period due to air temperatures above 0 °C can melt away much of the snow and the top layer of the ice. In periods of most intense melting, a large part of the surface can be covered by melt pools (Figure 6a and b). For microwave remote sensing the amount of water at the surface is the most critical factor, whereas for optical remote sensing the variability of the surface albedo is most important. Typical values of the albedo are 0.8 for snow, 0.6 for dry ice and 0.2 for water-covered ice.

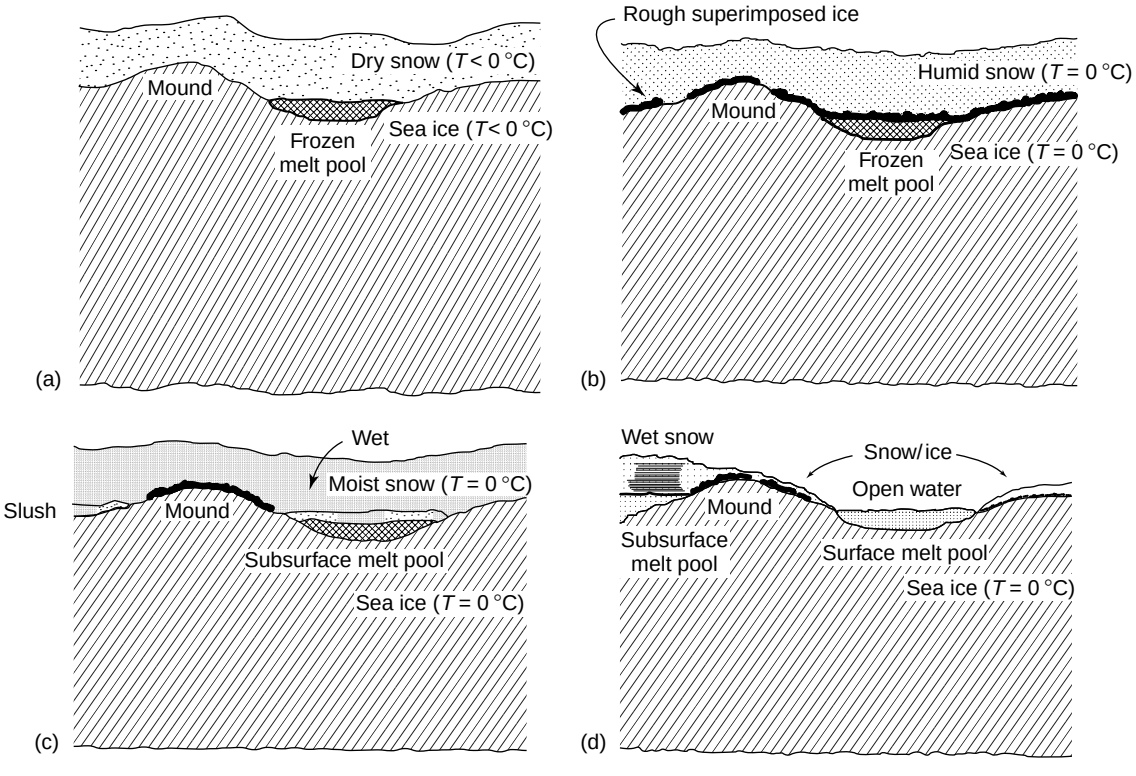

Figure 5 Snow and ice conditions on a MY floe which have impact on microwave remote sensing of sea ice. The four panels show characteristic conditions during (a) winter and early spring, (b) late spring, (c) early/midsummer, and (d) late summer.[11]

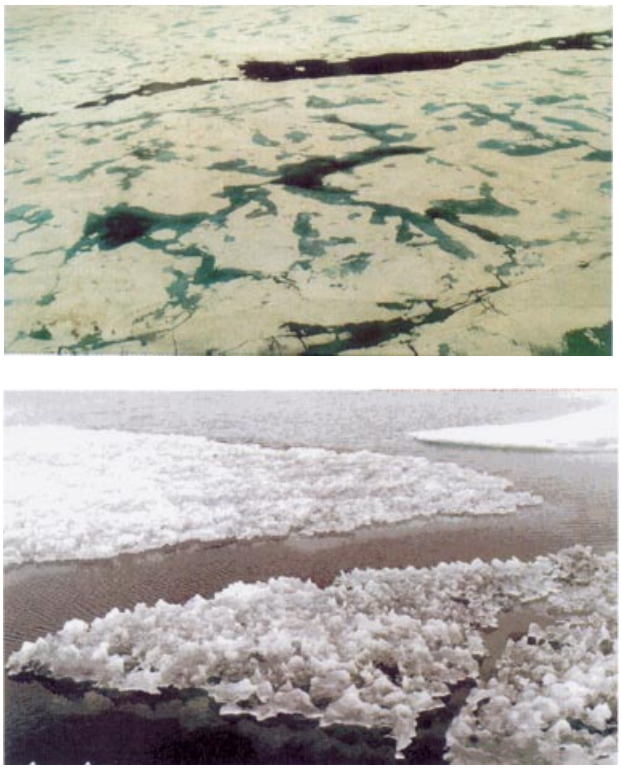

Figure 6 Photographs of melting ice taken in the area north of Svalbard in August 1991 by the Nansen Center. (a) Aircraft picture of water ponds on top of the ice during the melt season. Distance across the image center is about 3 km. (b) Magnification of the ice surface which is subject to intense melting. Distance across the image center is about 2 m.

The albedo is a key parameter controlling the incoming radiation, which determines the surface melt rate of snow and ice.

At the beginning of the cold season, all ice which survived the summer season becomes MY ice. The top layer (first meter) of a MY floe is harder, has lower salinity and is more optically transparent than a FY floe. The ridges of a MY floe are older, with more rounded and smoothed forms compared with the ridges of a FY floe. The most smooth areas are caused by refrozen pools covered with snow. The low salinity of the upper layer of MY ice makes it possible to discriminate FY ice from MY ice by microwave remote sensing during winter. This discrimination is not possible in the melt season, with water and wet snow on top of the ice. Another phenomenon of importance for remote sensing during winter time is the formation of frost-flowers on top of thin, high-saline ice which occur in cold weather. These crystals, which have dimensions of a few centimeters, cause very high backscatter of the AMW radar signal.

4 OVERVIEW OF SEA ICE REMOTE SENSING METHODS

4.1 Principal Methods, Instruments and Surface Characteristics

All remote sensing techniques for observation of the earth's surface use electromagnetic (EM) radiation in the visible/near-infrared (VNIR), TIR and microwave bands of the EM spectrum. There are three principal measurement techniques applicable to ocean and sea ice observations:

1. measurement of the part of the incoming solar radiation that is reflected at the surface of the earth (VNIR remote sensing);
2. measurement of the thermal radiation from the surface (TIR and PMW remote sensing); and
3. measurement of the return signal from active source, especially microwave radar methods using several types of instruments which measure backscattered radiation from the surface.

The VNIR, TIR and the microwave channels utilize intervals of the EM spectrum with high atmospheric transmission, such as in the optical bands at 0.4–2.5, 3.5–4.0 and 10–13 μm. The EM waves in these intervals do not generally penetrate clouds, so remote sensing observations of the earth's surface in these bands can only be done during cloud-free conditions. This is a severe limitation because sea ice occurs in regions where clouds are present most of the time, which inhibits regular observations of sea ice by satellite sensors operating in these bands. In the microwave area, at wavelengths above 0.3 cm, EM waves generally penetrate clouds, which makes it feasible to obtain regular, daily observations of sea ice (Figure 7a and b).

The most common satellite instrument used for measuring VNIR and TIR is the advanced very high-resolution radiometer (AVHRR), which provides images with a resolution of about 1 km in 4–5 different frequency channels. PMW radiometers, such as SSM/I, observe in a similar manner the emitted radiation in the microwave area in 4–5 channels and at different polarizations. The ground resolution varies in the range 15–150 km, depending on wavelength. AMW observations can be obtained by several types of instruments. The most common are SAR, SLR and scatterometers. The SAR instruments provide high-resolution images, with pixel size down to about 10 m, whereas the SLR provides medium-resolution images with a resolution of 1–2 km. Scatterometer data have coarser resolution, varying from 50 to about 10 km. The most common wavelengths for AMW instruments are 2.3 cm (X-band), 5.6 cm (C-band) and 23 cm (L-band).

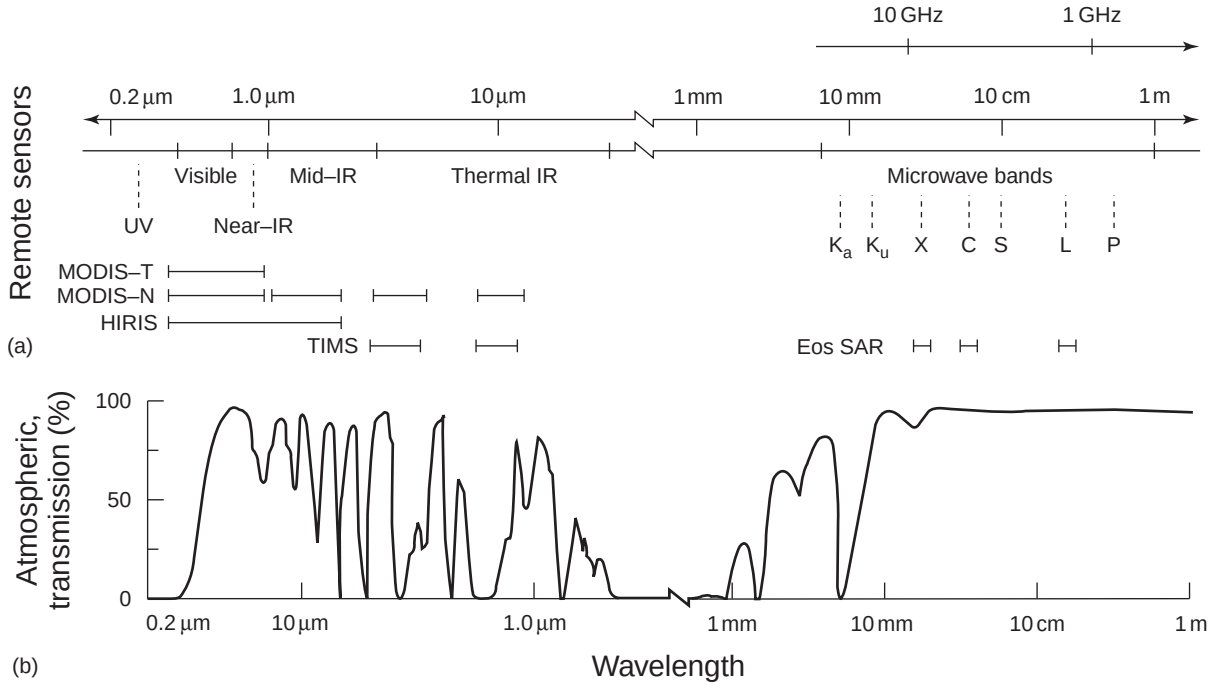

Figure 7 (a) EM spectrum showing the bands used in remote sensing, together with the operating area for some sensors. (b) Atmospheric transmission of the EM spectrum.

The basic parameter observed by VNIR is the albedo, or alternatively the reflection coefficient. For TIR and PMW it is emitted energy as a function of surface temperature and emissivity, and for AMW it is a backscatter coefficient (also called normalized radar cross-section). These three basic parameters reflect many physical and chemical variables describing the ice surface within the field-of-view of the satellite instrument. The sea ice surface to be analyzed based on the measured albedo, energy or backscatter coefficient from a satellite sensor has three macroscopic components:

- open water outside the ice edge, in leads/polynyas and on top of the ice in summer;
- ice with varying amount of salt water intrusions (salt pockets);
- snow on top of the ice.

The following properties of the macroscopic components have an impact on the remote sensing measurements:

- the percentage and distribution of the three components;
- the temperature of the components;
- the salinity of the components and the distribution of salt intrusions in the ice;
- the crystal structure of the ice and the snow;
- the occurrence of layers (e.g. snow–ice) and rough surfaces (e.g. floes, ridges, frost-flowers).

In addition, there are variables defined by the remote sensing instrument which also have significant impact on the measurements:

- the frequency (wavelength) of the radiation;
- the angle of incidence for the radiation;
- the polarization of the radiation if different from vertical.

4.2 Electromagnetic Properties of the Macroscopic Components

The most important EM properties of any surface are *reflection* and *emission*. The latter, expressed by a dimensionless coefficient called *emissivity*, can be computed from the complex dielectric constant (or the relative permittivity), $e = e' - ie''$, which characterizes the electrical properties of the media; e' is referred to as the dielectric constant and e'' as the dielectric loss factor. Alternatively, e can be estimated from the complex index of refraction, n, where $n^2 = e$. The reflection coefficient r is defined according to Equation (1):

$$r = \frac{n-1}{n+1}^2 \tag{1}$$

The dielectric constant, together with the surface roughness of the target, determine the emissivity. At microwave frequencies, the emitted energy, called brightness temperature, T_b, is a product of surface temperature, T_s, and the emissivity.

4.2.1 Open Water

Observation of open water is necessary as a reference for identification of sea ice, especially for the detection of the ice edge and areas of new ice formation. Within sea ice, open water can occur in leads and polynyas, and it is important to know if there is open water or thin ice in these areas. Examples of reflection coefficients for calm, pure water at two different wavelengths and normal incidence angle, using Equation (1) are as follows:

for visible radiation at 0.55 μm, $n = 1.33$, which gives $r = 0.02$;
for TIR radiation at 11 μm, $n = 1.20$, which gives $r = 0.008$.

The low reflection coefficients show that water is dark at visible and almost a blackbody at TIR wavelengths. For microwave remote sensing, water temperature and salinity are also important. Equations for computing reflection/emission coefficients from the dielectric constant can be found in Strogyn.[12] Examples of reflection and emission coefficients, at normal incidence, 273 K and 35‰ salinity are as follows:

for microwave radiation at 10 GHz, $e = 38 - 41i$, $r = 0.61$ and $e = 0.39$;
for microwave radiation at 37 GHz, $e = 9 - 18i$, $r = 0.47$ and $e = 0.53$.

Seawater is therefore a graybody at microwave frequencies, with emissivity as a function of frequency.

4.2.2 Microwave Electrical Properties of Ice

In the microwave wavelength region and for the physical temperatures encountered, the Rayleigh–Jeans approximation of the Planck radiation law pertains. The radiated power, usually expressed as brightness temperature, T_b, is therefore proportional to the physical temperature, T_s, at the surface. Most real objects emit only a fraction of the radiation that a perfect emitter would do at the same physical temperature. This fraction defines the emissivity of the object as T_b/T_s. Spatial variations in T_b observed over the surface of the earth are due primarily to variations in the emissivity of the surface material and secondarily to variations in temperature.

The value of the dielectric constant e' for freshwater ice has been measured as 3.17 at 10 GHz and varies little with frequency.[13] For saline FY ice, e' is higher and strongly dependent on both temperature and salinity of the ice. Only a small amount of brine will alter the dielectric constant of the ice, owing to the high value of e for salt water. A brine volume of 4‰ was found to give apparent dielectric constant of about 4.5. Lower temperatures will reduce the dielectric constant.

For sea ice, the dielectric constant e' is relatively constant with frequency above 1 MHz, but e'' is not. There is a minimum in e'' at 3–8 GHz, with higher values for lower and higher frequencies. For FY ice at 283 K and 8‰ salinity, the minimum e'' is approximately 0.3. As the temperature decreases, e'' will increase because precipitated salt will go back into solution. Furthermore, e'' will decrease with decreasing salt content. MY ice has a lower e'' than FY ice and its temperature dependence is weaker. Hence microwave radiation penetrates deeper into MY ice than FY ice.

Empirical relations have been established for the dielectric constant and thus the emissivity for sea ice at different microwave frequencies (Figure 8a and b). For example, the emissivity of seawater at 19 GHz and vertical incidence is about 0.44 compared with 0.92 for FY ice and 0.84 for MY ice. In contrast to water, it is not only the temperature, salinity, incidence angle and polarization which determine the emissivity of sea ice; the structure of the crystals and of the brine pockets – different for various ice types and ice ages – are also important.

4.2.3 Microwave Electrical Properties of Snow

There is a large difference in electrical properties between dry snow and wet snow with high liquid water fraction. Snow without salt can be considered dry in temperatures below −1 °C. Wet snow has a high e'', with correspondingly high absorption and emission. In summer, therefore, microwaves will only sense the top layer of the snow cover on the ice.

In winter, dry snow is almost transparent, with very small e''. However, the bottom snow layer on FY ice can contain salt, and with 4‰ salt the effect is to double the value of e'' from the value of 0.3 for pure ice. The dielectric constant e', which is independent of frequency, is only weakly dependent on temperature, and increases with increase in snow density. For a density of 0.2 g cm^{-3}, e' has a value of approximately 1.3.

4.3 Basic Remote Sensing Parameters

For a radiometer in the TIR or the microwave region, electrical and thermal properties of the surface will largely determine the basic parameter measured by the sensor. For a visual radiometer and in particular for a radar, the geometrical structure of the surface is often the

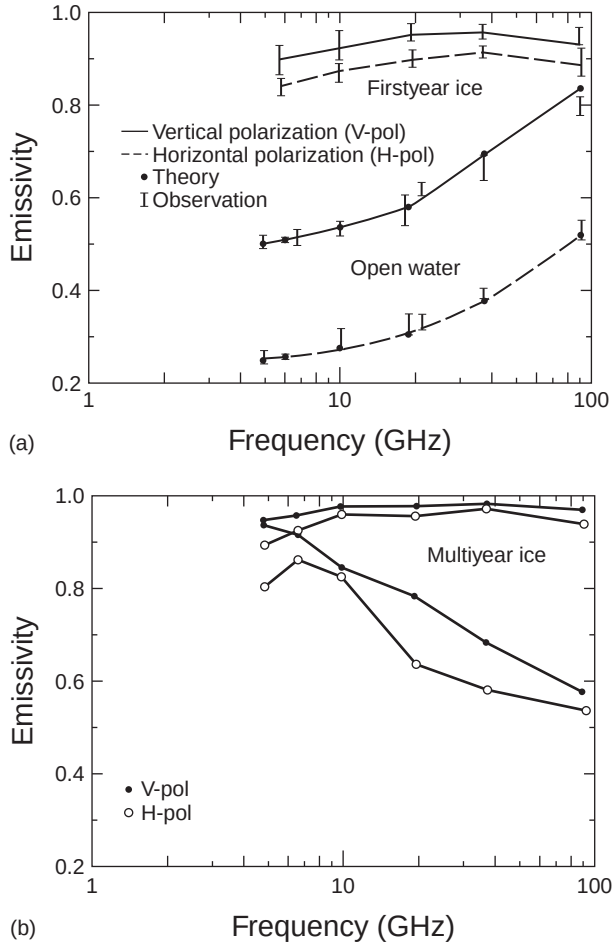

Figure 8 (a) Theoretical and observed emissivities for FY ice and open water both for vertical and horizontal polarization at 50° incidence angle as function of frequency. (b) Observed emissivities for MY ice as function of frequency. Note the high variability in emissivity of MY ice due to temperature change; the two upper graphs are for summer conditions and the two lower graphs for winter conditions–properties.

most important factor. Some relations between surface parameters and the basic remote-sensed parameters are discussed in this section.

4.3.1 Albedo

In the visual part of the spectrum, the albedo (or the reflection coefficient) is measured as a value between 0 (total absorption of the sunlight) and 1 (total diffuse reflection of the sunlight). Of the macroscopic sea ice components, new snow has the highest values of about 0.8. As the snow ages, the albedo decreases to values equal to dry old ice of 0.6 or lower. For very old and wet surfaces of ice and snow, the value can be as low as 0.2. Open water has an even lower albedo, varying somewhat with incidence angles and wind. The vertical reflection coefficient of calm pure water is only about 0.02.

For snow and ice, the visual albedo decreases slowly with increasing wavelength. In the near-IR region the decrease is more rapid for most snow and ice types, reaching very low values below 0.1 for wavelengths over 1.5–2.0 μm. This low albedo can be utilized in the discrimination of snow/ice from clouds. This is a major problem in the visual part of the spectrum where snow and clouds often show a similar signal.

4.3.2 Surface Temperature

In the TIR part of the spectrum, the surface signal expressed as the radiance obtained by remote sensing can be used as input to the Planck equation of radiation to find the surface temperature T_s if the emissivity e of the surface is known. For water, the value of e in the most used thermal spectral band of 10–12 μm is very high and stable, about 0.99, and accurate measurement (to 0.1–0.2 K) of the water surface skin temperature can be made. Snow and ice also have high emissivity in this region (values not well known) and approximate temperature measurements (to 1–2 K) of the upper layer can be made.

4.3.3 Brightness Temperature

As discussed above, the brightness temperature T_b provides the measurement of the PMW emission from the surface. The brightness temperature is defined by the real surface temperature T_s and the emissivity e by the relation $T_b = T_s e$. Depending on the application, various methods can be applied to separate the two variables e and T_s. For the most frequently used frequencies between about 6 and 90 GHz, the emissivity of both ice and water show large variations. Whereas e for calm water can be calculated fairly accurately from the electric properties, the value and variation of e for the various forms of ice and snow are less accurately known, and therefore often have to be measured empirically. PMW instruments typically have a very low-resolution cell in which many different components of snow, ice and water are present. Multifrequency measurements are widely applied in algorithms for determining the relative coverage (concentration) of the sea ice components and possibly also of other parameters.

4.3.4 Radar Backscatter Coefficient

For AMW systems, the most important parameter is the backscatter coefficient (σ_0), which expresses a measure of the energy scattered towards the source of the radiation, at a scattering angle of 180°. The backscatter coefficient and other related parameters are discussed further in section 5 which deals with imaging radar systems.

5 IMAGING RADAR SYSTEMS

5.1 Basic Properties of Imaging Radar Systems

Imaging radars are used for the detection and ranging of objects and features at the earth's surface. Imaging radars provide their own source of illumination in the microwave portion of the EM spectrum, at wavelengths of the order of 10 000 times longer than those in the visible part of the spectrum. Because of this, radars can operate independently of solar illumination and can penetrate clouds and precipitation. An imaging radar sends out energy in the form of an EM wave which strikes an object, causing it to be reflected back. The radar system senses the intensity of the reflected wave and at the same time measures the time taken for a round trip. Radar systems are different from more conventional visible and TIR systems in the following ways:

- radar is extremely sensitive to the surface roughness of the area being imaged;
- radar does not detect the visible color of the surface, but detects the moisture (or lack of) and electrical properties of the surface;
- radar systems record the phase and polarization characteristics of the reflected microwave pulse;
- radar produces images with speckle due to the coherent nature of the system;
- radar produces images with certain geometrical distortions, such as slant range geometry, image layover and shadowing; and
- radar is sensitive to the motion of objects in the imaged area.

A further description of an imaging radar system requires the definition of a set of variables that will lead to the remote sensing parameters used in the observation of sea ice and other surface features by spaceborne radar systems. Extensive descriptions of radar remote sensing principles have been published by Ulaby et al.[14,15] The following review is based on the radar remote sensing manual for the polar oceans by Shuchman et al.[16]

5.2 Some Important Parameters and Properties

The overall geometry of the ERS SAR systems is shown in Figure 9. The ERS SAR looks to the right of the orbit, covering the area between 244 and 344 km away from the *nadir*, which is the ground projection of the satellite orbit. The two main orientations are called *azimuth* (along the satellite orbit) and *range* (the look-direction of the SAR, perpendicular to the orbit). Range is also defined more precisely as a distance. The measurements of travel time of the reflected microwaves are readily converted to distance, which in radar usage is always called *range*, defined according to Equation (2):

$$R = c\frac{\tau}{2} \tag{2}$$

where c is the speed of light and τ is the travel time of the radar pulse. Range is also used to define the location of an object in the cross-track direction, which is perpendicular to the satellite flight direction (along track), called azimuth. In imaging theory, *resolution* of a radar system is defined as the radar's ability to discriminate between two point targets, observable as bright spots in a particular image. Two different principles determine the resolution of a radar system: azimuth resolution is defined by the beamwidth β of the radar, which is the angle by which the radar beam expands, and is a function of the antenna size and range. The location of a reflecting object is somewhere within the cross-hatched area in Figure 10, which is the area on the surface illuminated by a radar pulse. The along-track dimension, or azimuth resolution r_a, is defined according to Equation (3):

$$r_a = R\beta \tag{3}$$

The beamwidth of the radar is a function of the antenna length, with a larger antenna producing a narrower beam. For real aperture radars, such as SLRs, the only way to obtain fine azimuth resolution is to have a very short range or a very large antenna. SARs overcome this problem by utilizing the along-track motion of the satellite (or aircraft) to produce a very large antenna.

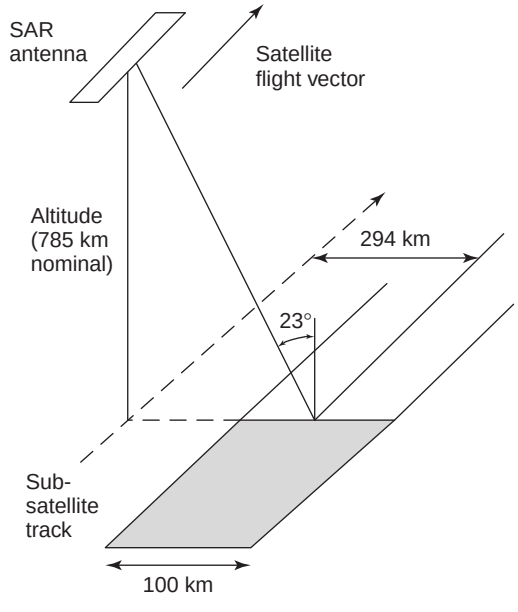

Figure 9 Geometry of the SAR system of the ERS-1/2 satellites.

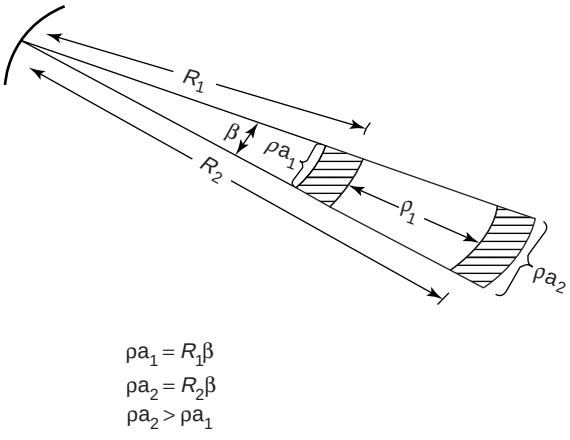

Figure 10 Illustration of the dependence of azimuth resolution ($r_a = \rho\alpha$) on range (R) and antenna beamwidth (β). R_1 and R_2 are the near-range and far-range, respectively, of the area illuminated by a radar pulse.[16]

The range resolution r_r of a radar system is given by Equation (4):

$$r_r = c\frac{t_p}{2} \tag{4}$$

where t_p is the duration of the transmitted radar pulse and c is speed of light. Range resolution is the distance that two objects must be separated to be detected at two unique ranges.

Imaging radars operate at microwave frequencies between 400 MHz and 35.2 GHz, corresponding to wavelengths between 62 and 0.85 cm. Radar frequencies are identified by letter designations, the most commonly used being the K-band (30 GHz, 1 cm), X-band (9.4 GHz, 3.2 cm), C-band (5.3 GHz, 5.7 cm), L-band (1.25 GHz, 23.5 cm) and P-band (450 MHz, 62 cm). At these wavelengths the EM waves are not appreciably attenuated by clouds, precipitation or the earth's atmosphere. Therefore, good-quality radar images can be obtained during all kinds of weather and light conditions. Another characteristic of imaging radar systems is polarization, which is defined as the direction of the electric vector is perpendicular to an EM wave. The electric vector is the plane in which the EM energy is transmitted. Visible light is unpolarized because the direction of the electric vector is randomly distributed. For radar systems it is common to emit pulses which are either horizontally (H) or vertically (V) polarized. The radar antenna can receive either the horizontally or vertically oriented return signals, or both.

In contrast to visible and IR remote sensing, radar remote sensing is extremely sensitive to the surface roughness. For smooth surfaces, there is near specular reflection of the radar waves, and little energy is reflected back to the radar. As the surface becomes rougher, the radar wavelengths interact with the surface in a way which causes diffuse reflection (Figure 11a–c). The amount of roughness controls the distribution of the reflected energy, and thus the intensity in the radar image. The wavelength and incidence angle of the radar wave determine the extent to which a surface is rough or not. There is a criterion for this [Equation 5]:

$$\Delta h \cos q_1 > \frac{\lambda}{8} \tag{5}$$

where Δh is the vertical scale of the surface roughness, q_1 is the incidence angle and λ is wavelength.

5.3 The Radar Equation

All radar measurements can be described by a basic equation which relates power, distance, reflectivity and antenna characteristics, which can be formulated as Equation (6):

$$P_R = \frac{P_t G \sigma A}{(4\pi)^2 R^4} \tag{6}$$

where P_R is power received, P_t is power transmitted, G is the gain of the antenna, σ is the radar cross-section and A is antenna area. A can also be expressed as $\lambda^2 G/4\pi$,

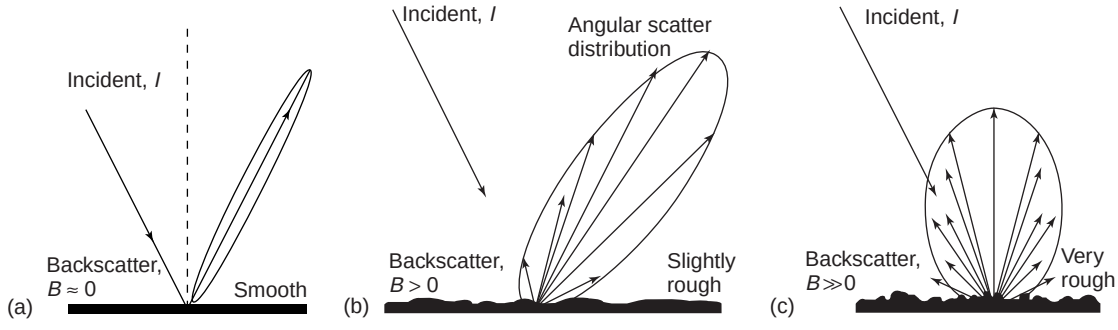

Figure 11 Qualitative illustration of how radar backscatter (B) at the ERS SAR incidence angle (23°) increases with surface roughness.

which is used to rewrite Equation (6) as a product of four terms [Equation 7]:

$$P_R = \underset{(1)}{\frac{P_t}{4\pi R^2}} \underset{(2)}{G} \underset{(3)}{\frac{\sigma}{4\pi R^2}} \underset{(4)}{A} \quad (7)$$

The energy of the outward propagating wave, which is spherically expanding, is given by term (1). This spherically expanding wave is focused down to an angular beamwidth by the antenna so that the fluxes become higher by a factor of G, term (2), over that of a spherically expanding wave. The focused energy impinges on an object which has a radar cross-section σ, which is defined as the equivalent of a perfectly reflecting object of a given area which reflects isotropically (spherically), as shown by term (3). Finally, the antenna area, A, term (4), intercepts a portion of the reflected wave so that this portion of the flux produces the power received by the antenna.

The basic radar return equation is general; it can be applied to any object of any shape or composition. For imaging over areas of terrain or ocean, a reflection coefficient is defined, σ_0, which is the radar cross-section, σ, per unit area. The radar Equation (6) can then be expressed as Equation (8):

$$P_R = \frac{\lambda^2}{(4\pi)^3} \int \frac{P_t G^2 \sigma_0}{R^4} \, dA \quad (8)$$

The averaged received power for a radar can then be determined by examining the integral radar equation for distributed targets. The radar scattering coefficient, σ_0, also called the backscatter coefficient, is important because it is an absolute measure of scattering behavior. It is a function of frequency, incidence angle, polarization and the scattering characteristics of the illuminated area. It is defined to be unity for a surface of small metal balls which exhibit ideal diffuse scattering without impacts of varying incidence angles. The backscatter from various

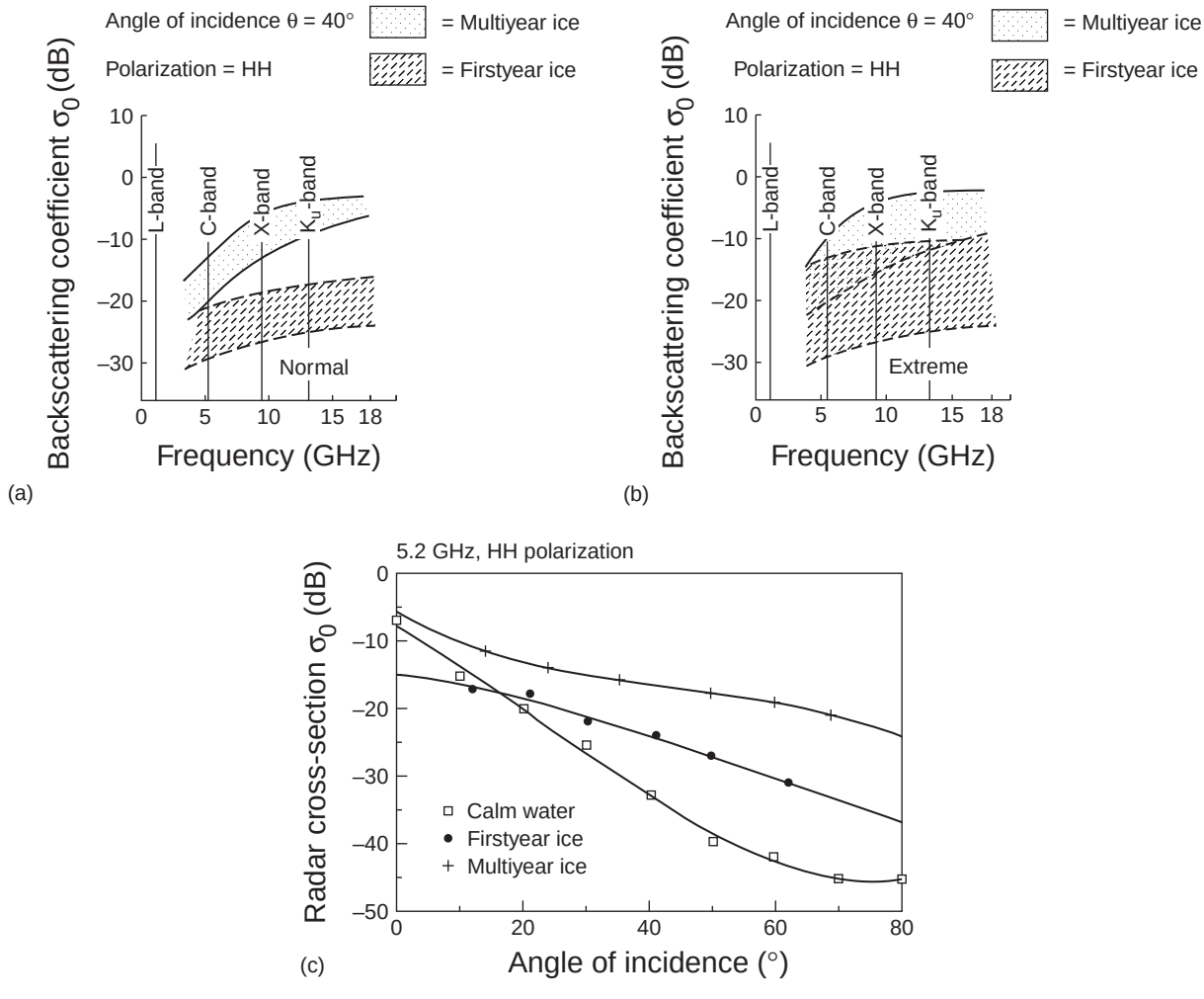

Figure 12 Backscattering coefficient σ_0 as a function of (a) and (b) frequency and (c) incidence angle for two major ice types.[17]

sea ice surfaces is generally much smaller than unity at the most used incidence angles (20–50°). Examples of the dependence of the backscatter coefficient on frequency and incidence angles for different ice types are shown in Figure 12(a–c). In section 7, examples of measurements of σ_0 by SAR for different ice types and surface conditions will be discussed.

6 CASE STUDY OF PASSIVE MICROWAVE REMOTE SENSING OF SEA ICE

6.1 Observation of Brightness Temperature

As discussed in section 4.2.1, the microwave brightness temperature T_b of the earth's surface depends on the electrical properties of the surface, embodied in its emissivity e and the physical temperature of the radiating portion of the surface T_s. This may be expressed by the following relation in terms of the wavelength λ and polarization p:

$$T_b[\lambda, p] = e[\lambda, p]T_s \tag{9}$$

This relationship is true only for e and T_s independent of depth, a typical assumption for sea ice.[18] The radiative transfer equation is the basis for the development of algorithms that convert the satellite radiance data into geophysical parameters. The microwave radiances received by the satellite are composed of various contributions from the earth, atmosphere and space and are illustrated schematically in Figure 13. The radiation received by the satellite, which is a function of wavelength and polarization as shown in Equation (9), can be expressed by Equation (10):

$$T_b = eT_s e^{-\tau} + T_{up} + (1-e)T_{down}e^{-\tau} + (1-e)T_{sp}e^{-2\tau} \tag{10}$$

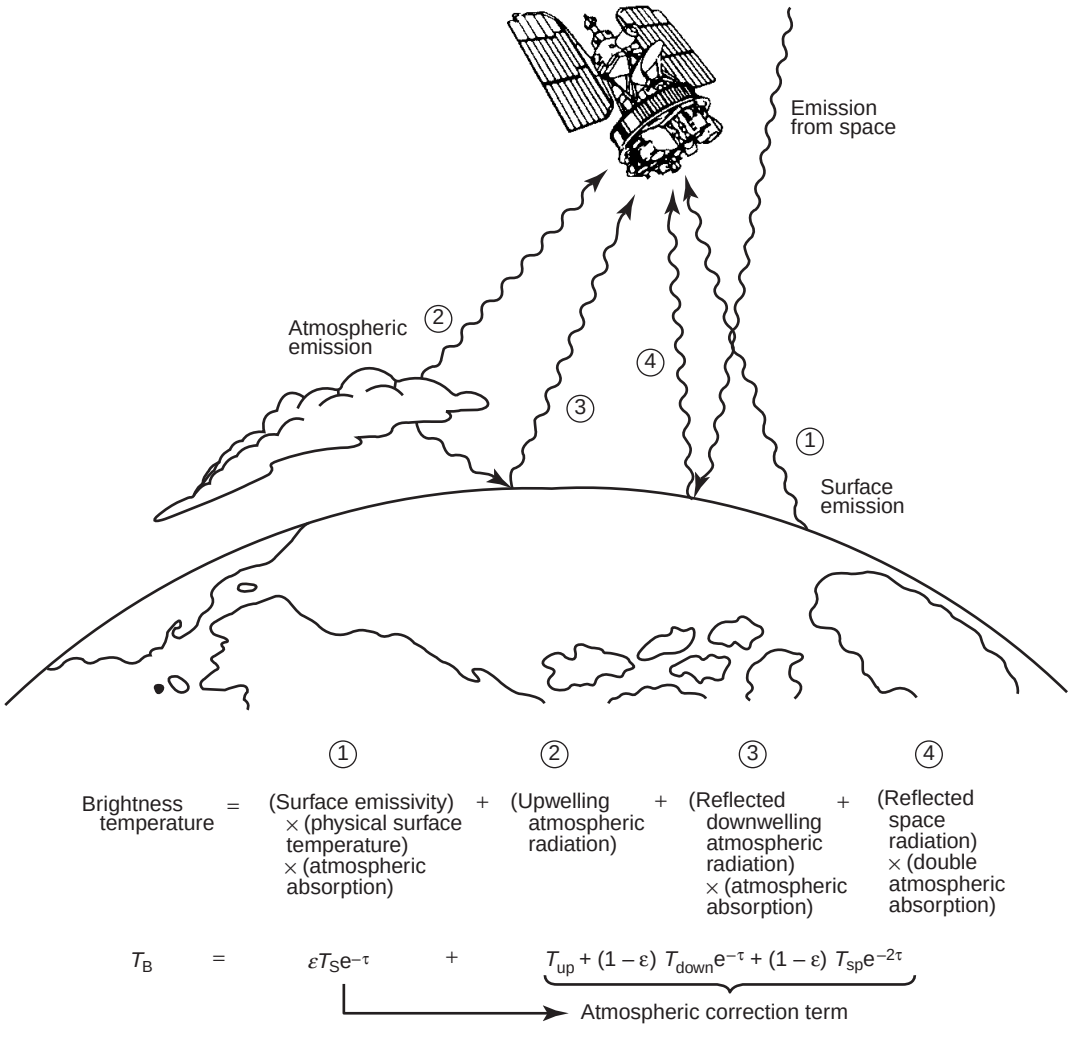

Figure 13 The main components of the brightness temperatures, observed by PMW radiometer in a satellite, are expressed by four terms in the radiative transfer equation.[18]

where T_b, T_s and e are as before, $e^{-\tau}$ represents atmospheric absorption, T_{up} is the atmospheric upwelling radiation, T_{down} is the atmospheric downwelling component and T_{sp} is the cosmic background component, as illustrated in Figure 13.[19]

6.2 Algorithms for Retrieval of Ice Parameters

There are several algorithms for estimation of sea ice concentration from brightness temperature observed in several channels and both polarizations.[18] An example of such an algorithm is NORSEX-85H, which is an extension of the NORSEX algorithm developed by Svendsen et al.[2] based on the SMMR data available from 1978 to 1987; 85H means that the extended version takes advantage of the improved spatial resolution of the 85-GHz channels provided by the SSM/I system which is currently in operation.

The basic algorithm was developed after the NORSEX marginal ice zone experiment near Svalbard in 1979 conducted by the NORSEX Group.[20] The algorithm computes area concentration of total ice and two ice types: MY ice and FY ice. Two channels are used: 18 GHz vertical polarization and 37 GHz vertical polarization. Only vertical polarization is used, since it was found to be significantly less sensitive to surface structure (snow layering, snow density, roughness) than horizontal polarization.

The ice emissivities and the atmospheric opacities used in the algorithm were basically measured during the NORSEX'79 experiment. Ocean water emissivities are theoretical values.[12] The brightness temperature of the surface is modeled as the linear combination of the area fraction of water, FY ice and MY ice in the field-of-view. A simple atmospheric model with surface air temperature as a parameter is used to convert satellite-observed brightness temperature to surface-observed brightness temperature.

The NORSEX algorithm requires that input brightness values are adjusted by a few degrees K to give 100% water over calm, open polar water. These adjustments are made based on a few suitable scenes where the state for the atmosphere and the surface is known, and thereafter kept largely unchanged. Adjustments are different for SMMR and SSM/I.

A rough estimate of the average surface air temperature over 100% pack-ice from climatological data is used as input for each scene. The NORSEX algorithm output is % water, % FY ice and % MY ice with a spatial resolution of approximately 60 km for SSM/I; 60 km is the resolution of the 19 GHz channel.

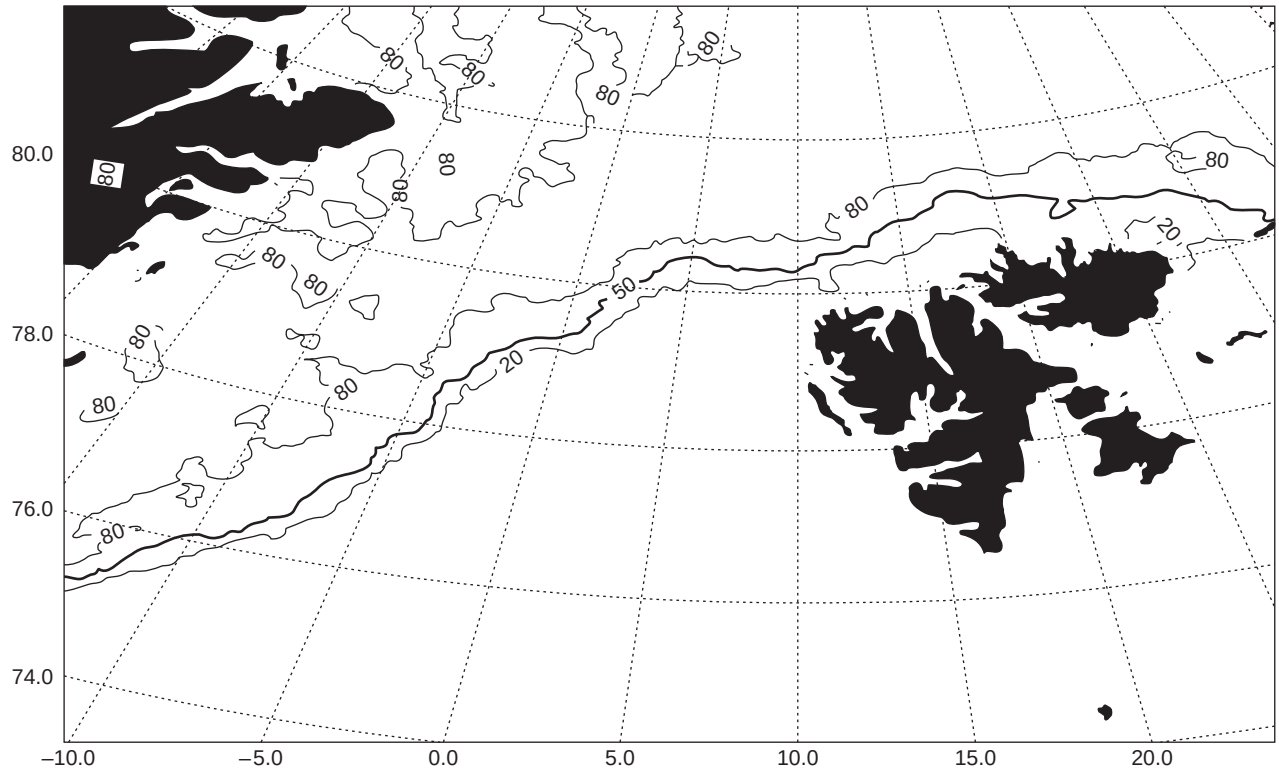

Figure 14 Example of ice concentration map derived from SSM/I data in the Svalbard area. Isolines for 20, 50 and 80% ice concentration are drawn. Open water is areas with less than 20% ice.

The 85 GHz channel, H polarization (85H), is very sensitive to the difference between water and FY ice and to the atmosphere. This channel is used to sharpen the ice–water boundary in the following way.

The 85H is spatially lowpass filtered to the 18V/37V algorithm resolution of 60 km. Concentration values are then multiplied with the ratio of the 85H signal to the lowpass-filtered 85H signal. This results in an image with a spatial resolution of 15–20 km, given by the 85H signal, and with ice concentration values given primarily by the basic NORSEX algorithm.

The NORSEX-85H algorithm was developed for use under the SIZEX'92 experiments in the Greenland and Barents Sea.[21] The ice edge defined by the 30–50% contour lines compares favorably with the ice–water boundary seen in near-simultaneous SAR images both under SIZEX'92 and later field experiments.

6.3 Example of Ice Analysis from Special Sensor Microwave Imager Data

The most common ice analysis using SSM/I data is to calculate total ice concentration by an algorithm as described in the previous section. An example of ice concentration map is shown in Figure 14. The data were received on file in the form of geolocated brightness temperatures, and the required channels were resampled on to a raster with 6-km spacing. On this raster, ice was computed using the NORSEX algorithm and thereafter sharpened by using the channel 85H signal as explained above. Then a contouring algorithm was applied to plot the 20%, 50% and 80% isolines for concentration. The map was produced on the same day as the satellite measurements were obtained, 27 September 1997, by transferring data from the SSM/I data distribution service in the USA. The SSM/I data are used for daily ice monitoring all over the world.

The most direct, quantitative means to study the global sea ice cover are satellite PMW remote sensors. Sea ice time series derived from multichannel PMW data are among the longest continuous satellite-derived geophysical records, extending over two decades. The Nimbus-7 SMMR provided data from 1978 to 1987, and the follow-up SSM/I onboard Defense Meteorological Satellite Program (DMSP) satellites have provided data since 1987. The 25×25 km gridded data sets are issued by the National Snow and Ice Data Center (USA). The brightness temperature data are used to calculate total ice concentration (the percentage of ice-covered ocean within an image pixel), from which total ice area (the area of ice-covered ocean) and total ice extent (the area within the ice–ocean margin) are derived.

Analysis of SMMR and SSM/I records taken separately revealed a greater reduction in Arctic sea ice area and extent during the SSM/I period. The decreases from 1987 to 1994 were ca. 4% per decade compared with ca. 2.5% per decade from 1978 to 1987,[4] with no significant trends found in the Antarctic. Since then, merged SMMR–SSM/I time series have been produced and analyzed, establishing the trends more firmly. Bjørgo et al.[22] established the trend in Arctic ice area and extent (1978–95) to be about -0.3×10^6 km^2 per decade (Figure 15a and b), corresponding to ca. 3% per decade, with no significant change in the Antarctic. The 3% per decade decrease in the Arctic ice extent (1978–97) was subsequently corroborated in a separate analysis[23] that also confirmed the hemispheric asymmetry seen earlier.[4,22] Cavalieri et al.[23] found a slight (ca. 1.5%) increase in the Antarctic, which may be considered significant. The hemispheric ice covers fluctuate quasi-periodically, with predominant periods between 3 and 5 years, although their variability is apparently not correlated.[23]

The capability to monitor interannual variations in MY ice area from SMMR and SSM/I data has recently been exploited using winter data, when FY and MY ice signatures permit their distinction. The analysis revealed a relatively large (ca. 7% per decade) reduction in the MY ice area from 1978 to 1998 (Figure 15c), compared with a ca. 2% per decade decrease in the total ice area in winter. This finding is supported by an SMMR–SSM/I data analysis that found an 8% increase (5.3 days) in the length of the sea ice melt season in the Arctic from 1978 to 1996.[24] It is also corroborated by spatially and temporally fragmentary observations (from submarine sonar transects) of ice thickness decreases, and also oceanographic data that have revealed changes in Arctic water masses since the 1970s that are reasoned to stem from a substantial (ca. 2 m) melting of perennial MY ice. If this trend were to continue, it could eventually lead to a markedly different sea ice regime in the Arctic, altering heat and mass exchanges and also ocean stratification.

7 CASE STUDY OF ACTIVE MICROWAVE REMOTE SENSING OF SEA ICE

The most important AMW remote sensing instrument for ice monitoring is the SAR, which provides high-resolution images of the ice surface. In this section we present some results of the SIZEX experiment, which was an ERS-1 SAR ice validation experiment in the Barents Sea in March 1992.[21]

7.1 General Interaction Between Synthetic Aperture Radar and Sea Ice

The radar backscatter from sea ice is highly dependent on physical ice properties such as salinity content,

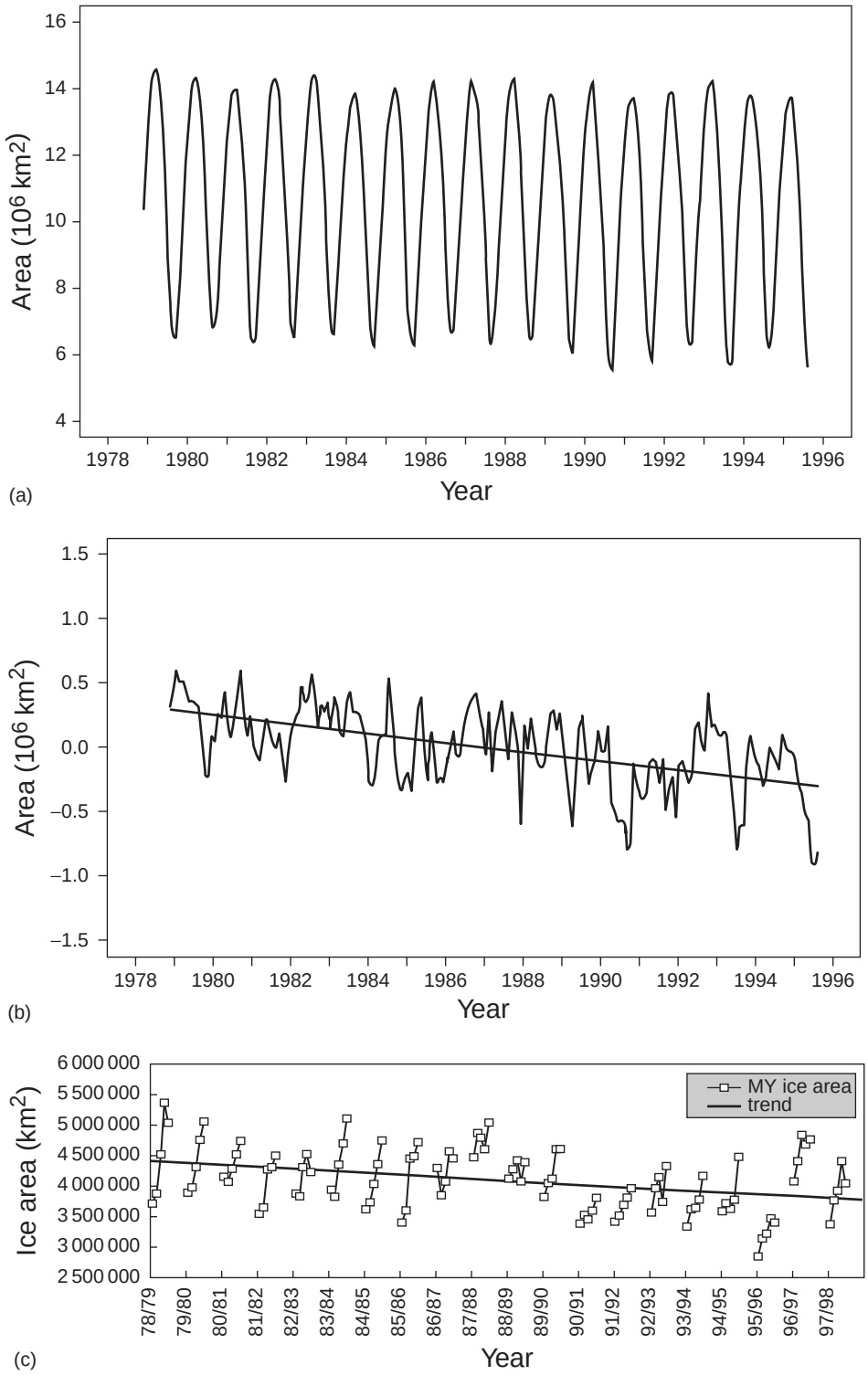

Figure 15 Time series of Arctic ice area derived from SMMR and SSM/I satellite PMW data: (a) monthly mean and (b) anomalies in ice area in 1978–95 where the linear regression indicates a ca. 31 000 km^2 yr^{-1} decrease, corresponding to ca. 3% per decade.[22] (c) Fraction of MY (i.e. having survived the summer melt) sea ice area in winter (November–March), 1978–98, where the linear regression indicates a ca. 30 000 km^2 yr^{-1} decrease, corresponding to ca. 7% per decade.

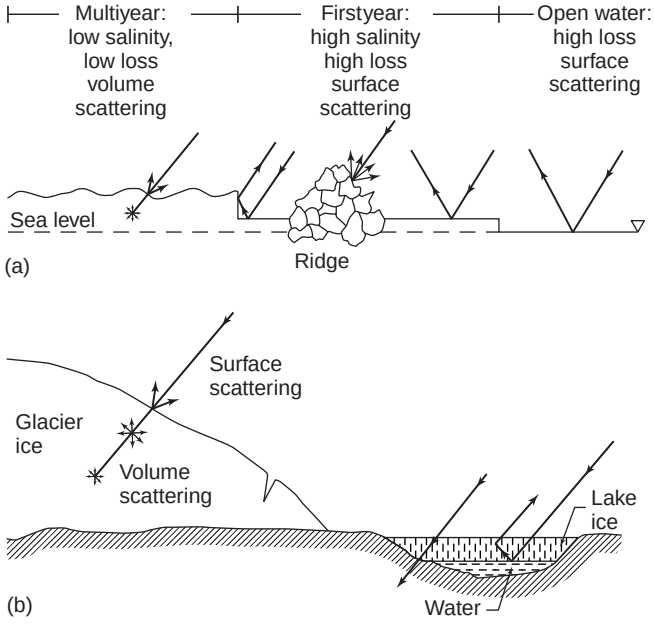

Figure 16 Schematic illustration of the interaction between radar microwaves and ice types: (a) sea ice and (b) glaciers, ice sheets and lake ice.

temperature, surface roughness, snow layers and presence of water. Figure 16(a) and (b) is a conceptual illustration of the interaction between AMW and different types of ice such as sea ice, including MY, FY (smooth and rough), open water, glaciers and lake ice. MY ice is characterized by low salinity in the surface layer (<2.0‰), which allows the penetration of the microwaves into the ice (volume scattering). FY ice, on the other hand, has higher salinity in the surface layer (5–7‰), which causes mainly surface scattering. The volume scattering from the MY ice tends to give higher backscatter than the surface scattering from smooth FY ice. However, FY ice can also give higher backscatter due to rough surface caused by ridges, small floes with many irregular edges and snow crystals. Open ocean gives very low backscatter in the case of no wind, but usually there is some wind which causes short surface waves and high backscatter values.[25] In the SIZEX'92 experiment some of these problems were studied in more detail.

7.2 Synthetic Aperture Radar Data Characteristics

During SIZEX'92, C-band ERS-1 SAR images with VV polarization were downlinked and processed at Tromsø Satellite Station and distributed to the Nansen Center for further analysis in addition to real-time use during the field experiment. SAR backscatter values from several ice types and open water conditions were analyzed during the experiment. Each image was averaged from 16×20 m to 100×100 m pixels, thereby removing much of the speckle noise. The images were then normalized to an incidence angle of 23° (in the center of the swath) by applying antenna pattern range spread and incidence angle corrections as described by the European Space Agency (ESA).[26] Finally, the pixels in the images were calibrated to σ_0 backscatter values by the relation $\sigma_0 = 20 \log V - 46.9$, where V is the digital value of the images analyzed at the Nansen Center. The lowest backscatter observed in the images had values of $V = 14$ (grease ice), and this was assumed to correspond to the noise floor of −24 dB. The accuracy of this calibration is estimated to be 1–2 dB.

7.3 Joint Analysis of Synthetic Aperture Radar Images and In Situ Measurements

The primary in situ measurements were profiles of snow and ice parameters taken at several sites from ships and helicopter. A total of 21 sites were investigated during the SIZEX experiment from March 2 to 11, 1992. At each site, several ice cores were drilled to obtain some spatial statistics of the ice parameters. The ice cores were analyzed for temperature, salinity, density and brine volume. Ice type was identified, ice thickness was measured and surface roughness was characterized. Snow thickness, temperature, grain type, grain size, density and layer description were also obtained at each site. An example of a SAR scene extending from open water to the interior of the ice pack is shown in Figure 17(a). From the SAR images and the in situ observations, the ice conditions in this part of the Barents Sea could be divided into three zones, reflecting different physical processes. In the following discussion, each of these zones, which are denoted A, B and C, are described in terms of their SAR signatures. Results of classification of the SAR images into six classes are shown in Figure 17(b).

7.3.1 Zone A: the Interior of the Ice Pack

In the upper left part of the SAR image in Figure 17(a) many MY floes were found about 100 km into the ice pack. The backscatter value of these floes were typically -9.0 ± 1.5 dB. Between the MY floes, areas of undeformed FY ice typically 2–3 m thick were observed. This ice had backscatter values from −10 to −13 dB. Refrozen leads with smooth thin ice had the lowest backscatter values, between −19 and −13 dB. Ridges and leads with open water were found to have variable and higher backscatter, above −8 dB. In the interior of the ice pack the ERS-1 SAR demonstrated its ability to discriminate between (a) young ice in refrozen leads, (b) MY ice in large floes, (c) rubble fields/ridges and (d) smooth FY ice, using simple thresholding of backscatter values.

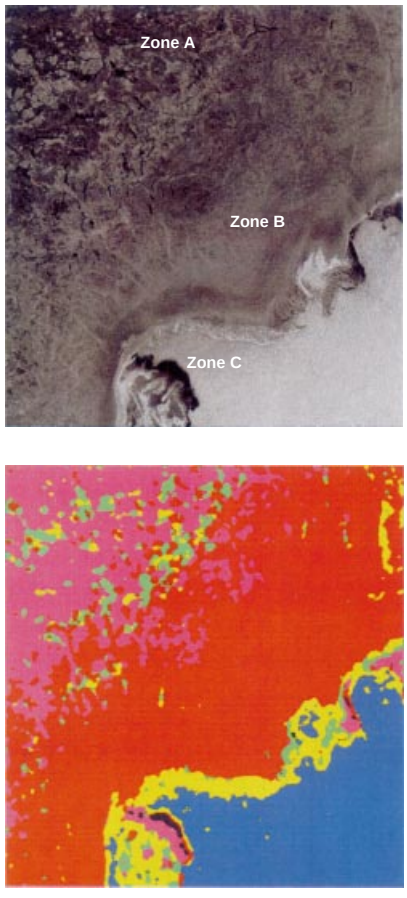

Figure 17 (a) An ERS-1 SAR image from the SIZEX experiment, covering 100 × 100 km of the marginal ice zone with characteristic zones denotes A, B and C. (b) The result of a classification of ice types and open water from the SAR image in (a). Color code for the classification: open water, blue; grease ice, gray; pancake ice, yellow; deformed FY ice, red; MY ice, green; and undeformed FY ice, violet.

7.3.2 Zone B: the Small Floe Area

This zone is characterized by small floes, typically 10–100 m in size, which have been broken up by surface waves penetrating from the open ocean into the ice pack. The width of the zone depends on the intensity of the wave field in the preceding days. During the SIZEX experiment the zone was 20–30 km wide and was clearly identified in the SAR images as a more uniform zone between open ocean and the interior of the ice pack (Figure 17a). The dominant ice type is rough FY ice 2–3 m thick with backscatter values ranging from −10 to −6.5 dB. These values are significantly higher than for the undeformed FY ice in zone A, owing to the increased surface roughness. This surface roughness is caused by the edges and ridges of the numerous small floes which cannot be identified in the SAR images. During the experiment the ice concentration was generally high (>95%). The open water or thin ice areas between the floes was in the range 1–10 m, which could not be observed in the SAR images.

MY floes, which drift southwards with the East Spitzbergen Current, could also occur in this zone. Similarly to the FY floes, the MY floes tended to break up, owing to the wave field. The SAR signature of small MY floes was similar to that of surrounding FY floes. Drifting icebergs with a horizontal scale of 100 m and a draft of 5–8 m were frequently observed in the zone, but their SAR signature even in full-resolution images was diffuse. Reliable observations of these icebergs could not be made by the ERS-1 SAR.

7.3.3 Zone C: the Area of Ice Formation

Most of the ice in the Barents Sea is formed locally as the ice edge advances southwards during the freezing season. In the experiment, the areas of ice formation were mapped by SAR images and documented by in situ observations. The first stage in ice formation is grease ice, which dampens out the short surface waves and causes low radar return signals. This ice had the lowest backscatter of all ice types, ranging from −24 to −14 dB. After 1–2 days of freezing, the grease ice starts to form pancake ice with characteristic edges which causes a high radar return, typically above −7.0 dB. As the pancake ice grows thicker during the winter it forms 1–2 m thick FY ice. The ice in this zone is constantly exposed to surface waves and is therefore characterized by a rough surface. Grease ice and calm water (wind speed below 3 m s^{-1}) can have overlapping backscatter values which make the interpretation of the SAR images ambiguous; however, no wind speeds below 3 m s^{-1} were observed in the experiment. Open water in the SAR images had higher backscatter than any of the observed ice types, from −4.5 to −3.0 dB. The backscatter values for the most important ice types in the Barents Sea are summarized in Figure 18.

7.4 Summer Conditions

The classification into three zones suggested for winter conditions would not be valid in the summer, because the interaction between radar microwaves and the ice surface is completely different from winter conditions owing to the different temperature. Zone C with ice formation would be absent. Zones A and B would be detectable, but the discrimination between different ice types becomes less significant. Owing to wet snow and melt water on top of the ice, the SAR backscatter values from MY floes is reduced compared with the winter situation and becomes similar to the signature for FY ice. It is therefore difficult to separate the two ice types based solely on backscatter levels.[3] Thin ice types such as grease ice, pancake ice and young ice are usually not found in the summer.

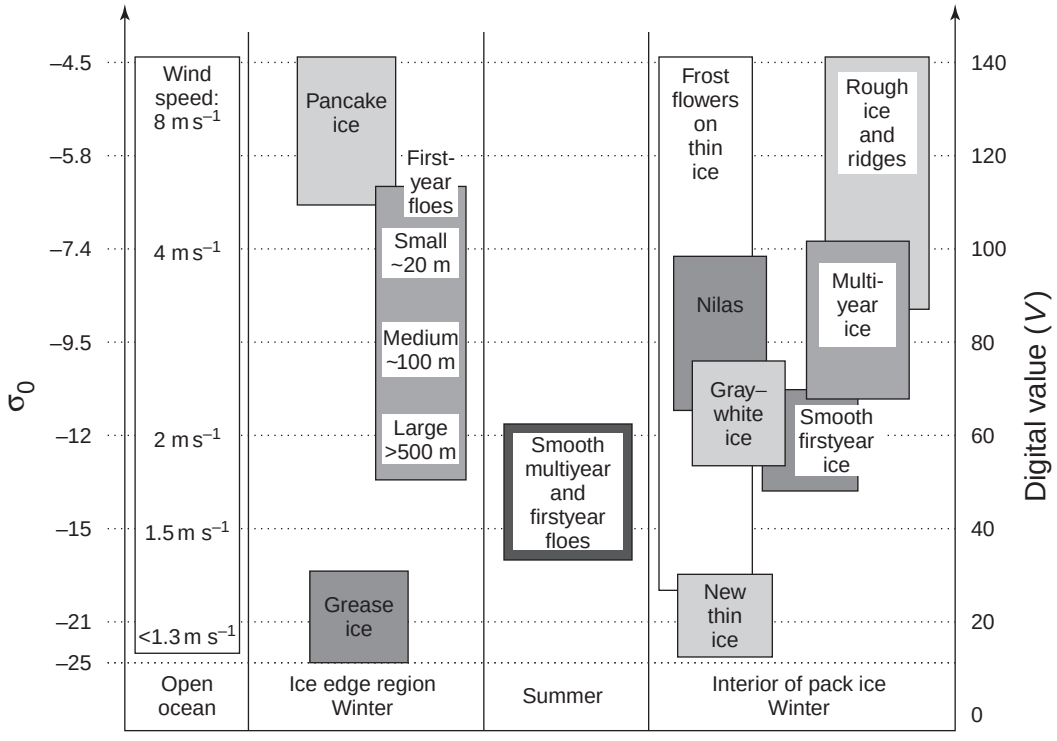

Figure 18 Backscatter coefficient for ERS SAR (C-band, VV polarization and 23° incidence angle), expressed in σ_0 (left) and digital value (right), for different regions and ice types as observed during the SIZEX experiment and other SAR ice validation experiments. Note that ERS SAR backscatter for different wind speeds over open ocean is included for comparison because the SAR backscatter for open water can often be similar to that of some of the ice types.[7]

7.5 Conclusion from the Synthetic Aperture Radar Ice Classification Study

In the SIZEX'92 experiment, ERS-1 SAR scenes of all the major ice types in the Barents Sea were analyzed for winter conditions. These included:

- 3–4 m thick MY ice floes which originate from the Arctic Ocean north of Svalbard;
- refrozen leads with smooth thin ice;
- consolidated FY ice 2–3 m thick which is formed between MY floes in the interior of the ice pack during the present winter season;
- rough ice in leads, ridges and rubble fields;
- FY ice 2–3 m thick in a 20–30 km wide zone inside the ice edge broken up in typically 10–50 m large floes due to wave action;
- new ice formed outside the ice edge (grease ice and pancake ice).

The most difficult factor which influences the SAR ice classification and ice concentration algorithms is the variable backscatter from open water due to the wind speed. Ice classification algorithms have been applied to the SAR images from the SIZEX experiment, and the results from use of the Wackerman algorithm are shown in Figure 17(b).[27] Such classifications will normally result in ambiguities, because different ice types can have similar SAR signatures. Also, open water can have a similar signature to some of the ice types, in particular pancake ice and ice with rough surface. In order to improve the algorithm for ice type determination and thus also for the ice edge detection and ice concentration estimation, it is necessary to know the open water signature as a function of wind speed. The wind speed can thus be used as an input parameter in the algorithms.

The main result of the SIZEX program and other SAR ice investigations[3] is an improved understanding of ice processes and their signatures in SAR images. All the previous SAR ice studies have contributed to the establishment of look-up tables, such as in Figure 18, for many ice types at different seasons.[17] Although there are many phenomena and features observable in SAR images which remain to be explained, the general knowledge of SAR ice signatures is fairly well established.

8 FUTURE PERSPECTIVES ON SEA ICE REMOTE SENSING

Sea ice remote sensing by use of satellites is expected to become increasingly important both for climate

monitoring[4,22] and for process studies investigating the interaction between ocean and atmosphere.[28,29] After the SIZEX experiment, there have been developments towards operational use of satellite SAR data in ice monitoring in many countries.[6] The ESAs ERS program has been very instrumental in developing SAR ice monitoring methods, including algorithm development, validation and user demonstrations. From 1996, the Canadian RADARSAT has provided SAR images covering up to 500 km wide swaths which are very useful in operational ice monitoring. The ice centers in the USA and Canada use RADARSAT SAR data as the main source of information,[30] while, in Europe, Finland, Sweden and Denmark have introduced these data in regular ice monitoring service in the Baltic Sea and Greenland waters.[6] Several new satellites, carrying SAR, scatterometer, PMW radiometers and other relevant instruments, will have an increasingly important role in ice monitoring. The next satellite of interest will be ENVISAT, scheduled for launch in 2001, which will have Advanced Synthetic Aperture Radar (ASAR), a new SAR with the possibility of obtaining data in several polarizations and at different resolutions and incidence angles. Several other SAR satellites, which will have ice monitoring as one of their tasks, are planned for the next decade.

ACKNOWLEDGMENTS

We thank many of our colleagues, mentioned in the reference list, for valuable contributions to this article.

ABBREVIATIONS AND ACRONYMS

AMW	Active Microwave
ASAR	Advanced Synthetic Aperture Radar
AVHRR	Advanced Very High-resolution Radiometer
DMSP	Defense Meteorological Satellite Program
EM	Electromagnetic
ENVISAT	Environmental Satellite
ERS	European Remote Sensing Satellite
ERTS	Earth Resources and Technology Satellite
ESA	European Space Agency
ESMR	Electrically Scanning Microwave Radiometer
FY	First-year
IR	Infrared
MY	Multiyear
NORSEX	Norwegian Remote Sensing Experiment
PMW	Passive Microwave
SAR	Synthetic Aperture Radar
SIZEX	Seasonal Ice Zone Experiment
SLR	Side-looking Radar
SMMR	Scanning Multichannel Microwave Radiometer
SSM/I	Special Sensor Microwave Imager
TIR	Thermal Infrared
VNIR	Visible/Near-infrared
WMO	World Meteorological Organization

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Polar Environments Assessment by Remote Sensing • Satellite and Sensor Systems for Environmental Monitoring

REFERENCES

1. P. Gloersen, W.J. Campbell, D.J. Cavalieri, J.C. Comiso, C.L. Parkinson, H.J. Zwally, 'Arctic and Antarctic Sea Ice, 1978–1987. Satellite Passive–Microwave Observations and Analysis', NASA SP-511, National Aeronautics and Space Administration, Washington, DC, USA, 1992.
2. E.A. Svendsen, K. Kloster, B.A. Farrelly, O.M. Johannessen, J.A. Johannessen, W.J. Campbell, P. Gloersen, D. Cavalieri, C. Matzler, 'Norwegian Remote Sensing Experiment: Evaluation of the Nimbus 7 Scanning Multichannel Microwave Radiometer for Sea Ice Research', *J. Geophys. Res.*, **88**(C5), 2781–2791 (1983).
3. O.M. Johannessen, W.J. Campbell, R. Shuchman, S. Sandven, P. Gloersen, J.A. Johannessen, E.G. Josberger, P.M. Haugan, 'Microwave Study Programs of Air–Ice–Ocean Interactive Processes in the Seasonal Ice Zone of the Greenland and Barents Seas', in *Microwave Remote Sensing of Sea Ice*, ed. F. Carsey, AGU Geophysical Monograph No. 68, Chapter 13, 261–289, American Geophysical Union, Washington DC 20009, USA, 1992.
4. O.M. Johannessen, M. Miles, E. Bjørgo, 'The Arctic's Shrinking Sea Ice', *Nature (London)*, **376**, 126–127 (1995).
5. P. Wadhams, 'Sea Ice Thickness Changes and Their Relation to Climate Change', in Nansen Centennial Volume, American Geophysical Union Monograph No. 85, 337–362, American Geophysical Union, Washington DC 20009, USA, 1994.
6. S. Sandven, H. Gronvall, A. Seina, H.H. Valeur, M. Nizovsky, H. Steen Andersen, V.E.J. Haugen, 'Operational Sea Ice Monitoring by Satellites in Europe. OSIMS Final Report', NERSC Technical Report No. 148, 1998, Nansen Center, Bergen, Norway.

7. O.M. Johannessen, A.M. Volkov, V.D. Grischenko, L.P. Bobylev, S. Sandven, K. Kloster, T. Hamre, V. Asmus, V.G. Smirnov, V.V. Melentyev, L. Zaitsev, 'ICEWATCH – Ice SAR Monitoring of the Northern Sea Route', in *Operational Oceanography. The Challenge for European Co-operation*, ed. J.H. Stel, Proceedings of the First International Conference on EuroGOOS, 7–11 October 1996, The Hague, The Netherlands, Elsevier Oceanography Series No. 62, Elsevier, Amsterdam, 224–233, 1997.
8. W.J. Campbell, 'NASA Remote Sensing of Sea Ice AIDJEX', in *Proceedings of the World Meteorological Organization Technical Conference, Tokyo, Japan*, World Meteorological Organization, Tokyo, Japan, No. 350, 55–56, 1973.
9. 'WMO Sea Ice Nomenclature', WMO Report No. 259, TP 145, 1970, World Meteorological Organization, Geneva.
10. M. Lensu, S. Heale, K. Riska, P. Kujula, 'Ice Environment and Ship Hull Loading Along the Northern Sea Route', INSROP Working Paper No. 66, 1996.
11. R.G. Onstott, T.C. Grenfell, C. Maetzler, C.A. Luther, E.A. Svendsen, 'Evolution of Microwave Sea Ice Signatures During Early and Midsummer in the Marginal Ice Zone', *J. Geophys. Res.*, **92**, 6825–6835 (1987).
12. A. Strogyn, 'Equations for Calculating Dielectric Constant of Saline Water', *IEEE Trans. Microwave Theory Tech.*, **MTT-19**, 733–736 (1971).
13. M. Hallikainen, D.P. Winebrenner, 'The Physical Basis for Sea Ice Remote Sensing', in *Microwave Remote Sensing of Sea Ice*, ed. F. Carsey, AGU Geophysical Monograph No. 68, Chapter 3, 29–46, American Geophysical Union, Washington DC 20009, USA, 1992.
14. F.T. Ulaby, R.K. Moore, A.K. Fung, 'Real Aperture Side-looking Airborne Radars', in *Microwave Remote Sensing – Active and Passive: Volume II – Radar Remote Sensing and Surface Scattering and Emission Theory*, Addison-Wesley, Reading, MA, 562–629, 1982.
15. F.T. Ulaby, R.K. Moore, A.K. Fung, 'Synthetic Aperture Side-looking Airborne Radar Systems', in *Microwave Remote Sensing – Active and Passive: Volume II – Radar Remote Sensing and Surface Scattering and Emission Theory*, Addison-Wesley, Reading, MA, 630–745, 1982.
16. R.A. Shuchman, C.C. Wackerman, L.L. Sutherland (eds.), *The Use of Synthetic Aperture Radar to Map the Polar Oceans. A Remote Sensing Manual*, ERIM, Ann Arbor, MI, 1991.
17. R.G. Onstott, 'SAR and Scatterometer Signatures of Sea Ice', in *Microwave Remote Sensing of Sea Ice*, ed. F. Carsey, AGU Geophysical Monograph No. 68, Chapter 5, 73–104, American Geophysical Union, Washington DC 20009, USA, 1992.
18. K. Steffen, J. Key, D.J. Cavalieri, J. Comiso, P. Gloersen, K.St. Germain, I. Rubinstein, 'The Estimation of Geophysical Parameters Using Passive Microwave Algorithms', in *Microwave Remote Sensing of Sea Ice*, ed. F. Carsey, AGU Geophysical Monograph No. 68, Chapter 10, 201–231, American Geophysical Union, Washington DC 20009, USA, 1992.
19. C.T. Swift, D.J. Cavalieri, 'Passive Microwave Remote Sensing for Sea Ice Research', *EOS Trans. Am. Geophys. Union*, **66**, 1210–1212 (1985).
20. NORSEX Group (O.M. Johannessen, W.J. Campbell, coordinators), B.A. Farrelly, J.A. Johannessen, O.M. Johannessen, E. Svendsen, K. Kloster, I. Horjen, C. Matzler, W.J. Campbell, J. Crawford, R. Harrington, L. Jones, C. Swift, V.E. Delnore, D. Cavalieri, P. Gloersen, S.V. Hsiao, O.H. Shemdin, T.W. Thomson, R.O. Ramseier, 'Norwegian Remote Sensing Experiment in a Marginal Ice Zone', *Science*, **220**(4599), 781–787 (1983).
21. S. Sandven, O.M. Johannessen, M.W. Miles, L.H. Pettersson, K. Kloster, 'Barents Sea Seasonal Ice Zone Features and Processes from ERS-1 SAR: SIZEX92', *J. Geophys. Res.*, **104**(C7), 15 843–15 857, July (1999).
22. E. Bjørgo, O.M. Johannessen, M.W. Miles, 'Analysis of Merged SMMR/SSMI Time Series of Arctic and Antarctic Sea Ice Parameters', *Geophys. Res. Lett.*, **24**, 413–416 (1997).
23. D.J. Cavalieri, P. Gloersen, C.E. Parkinson, H.J. Zwally, J.C. Comiso, 'Observed Hemispheric Asymmetry in Global Sea Ice Changes', *Science*, **278**, 1104–1106 (1997).
24. D.M. Smith, 'Recent Increase in the Length of the Melt Season of Perennial Arctic Sea Ice', *Geophys. Res. Lett.*, **25**, 655–658 (1998).
25. R.A. Shuchman, R.G. Onstott, 'Remote Sensing of Polar Oceans', in *Polar Oceanography, Part A: Physical Science*, ed. W.O. Smith, Jr, 123–169, Academic Press, San Diego, 1990.
26. N. Caneva, 'ERS-1 SAR Calibration', *J. Geophys. Res.*, **104**(C7), 15 843–15 857 (1999).
27. C.C. Wackerman, D. Miller, 'An Automated Algorithm for Sea Ice Classification in the Marginal Ice Zone Using ERS-1 Synthetic Aperture Radar', ERIM Technical Report, Ann Arbor, MI, 1996.
28. S. Sandven, O.M. Johannessen, 'Seasonal Ice Zone Studies', in *The Sea: Ocean Engineering Science*, eds. B. Le Méhauté, O.M. Hanes, John Wiley, New York, 567–591, Vol. 9, 1990.
29. O.M. Johannessen, S. Sandven, W.P. Budgell, J.A. Johannessen, R. Shuchman, 'Observation and Simulation of Ice Tongues and Vortex Pairs in the Marginal Ice Zone', in *Nansen Centennial Volume, American Geophysical Union Monograph* No. **85**, 109–136, American Geophysical Union, Washington DC 20009, USA, 1994.
30. B. Ramsey, M. Manore, L. Weir, K. Wilson, D. Bradley, 'Utilization of RADARSAT Data in the Canadian Ice Service', paper presented at the International Symposium Geomatics in the Era of RADARSAT, Ottawa, Canada, May 25–30, 1997.

Semiarid Land Assessment: Monitoring Dry Ecosystems with Remote Sensing

Joachim Hill
University of Trier, Trier, Germany

1 Introduction **1**
2 Dryland Ecosystems **2**
2.1 Climate and Vegetation 3
2.2 Soils 4
2.3 Land Use 4
3 Remote Sensing and the Assessment of Environmental Conditions in Semiarid Regions **5**
3.1 The Importance of Assessing Surface Properties 5
3.2 Remote Sensing Systems 6
4 Remote Sensing Data Analysis **7**
4.1 Data Preprocessing 8
4.2 Scene Models: Fundamental Considerations, Indicators and Interpretation Concepts 9
4.3 Improved Scene Models: Detecting Surface Properties Based on Spectral Decomposition Modeling 12
5 Remote Sensing of Semiarid Lands: From Resource Assessments to Regular Monitoring of Land Degradation Processes **13**
5.1 Vegetation Assessments and Monitoring in Dryland Systems 14
5.2 Examples for the Assessment of Soil Characteristics in Dry Ecosystems 15
5.3 Monitoring Arid Ecosystems: the Desertification Issue 19
6 Concluding Remarks **19**
Abbreviations and Acronyms **20**
Related Articles **21**
References **21**

Semiarid lands belong to the world's dryland environments, which are permanently, seasonally or temporarily subjected to a significant deficit in moisture availability. Drylands, or dry ecological systems, include the proper deserts as well as their semiarid and subhumid dry margins. The areas referred to range from the cold polar deserts (which, because of their highly specific nature, are not considered in this article) to warm tropical deserts surrounded by tropical semiarid and subhumid drylands, including subtropical Mediterranean latitudes.

Remote sensing is the scientific approach to extracting information from an object without coming into physical contact with it. Broad definitions would need to include computer vision, astronomy, medical and biological imaging, sonar, observation of terrestrial and marine ecosystems and many other fields. Within the restricted context of this article we shall limit our perspective to collecting information about the state of terrestrial ecosystems through air- and space-borne sensor systems. This process is based on the interaction of electromagnetic radiation with the surface constituents, and increasing knowledge about these interaction mechanisms makes it possible to interpret remotely sensed data with regard to specific environmental conditions and their change in time.

1 INTRODUCTION

Dryland areas cover approximately one-third of the continental surface of the earth. In particular the semiarid ecosystems provide important land resources for adapted agricultural production and grazing systems. While often considered areas at risk in the context of global climatic change and worldwide desertification dynamics, the importance of thoroughly monitoring the state of the environment in these areas has long been recognized. However, with regard to the spatial extension but, at the same time, limited accessibility of dryland systems, it becomes obvious that terrestrial observation alone is hardly able to cope with this task. Remote sensing with air- or space-borne sensor systems provides a comprehensive spatial coverage, is intrinsically synoptic and collects objective, repetitive data, and is thus ideally suited for monitoring environmentally sensitive areas. The major problem associated with its use is to interpret quantitatively a measured signal that has interacted with remote objects in terms of the properties of these objects.

Since most semiarid lands are characterized by frequently cloud-free atmospheres it is obvious that remote sensing systems operating in the optical range of the wavelength spectrum are and have been used predominantly. Besides airborne systems for individual surveys on local to regional scales, several geostationary and polar-orbiting satellites [e.g. METEOSAT/GOES, National Oceanographic and Atmospheric Administration (NOAA) (of the USA) advanced very high resolution radiometer

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

(AVHRR), land observing satellite (Landsat) [of the National Aeronautics and Space Administration (NASA) of the USA], système pour l'observation de la terre (SPOT) (French spacecraft)] are available which operate in the reflective and emissive domain and can be used for regional to global assessments. As continuity for most operational remote sensing systems (which are already operating for 10–20 years) seems guaranteed, they can be used for continuous environmental monitoring but also for retrospective studies on environmental change that has occurred in the past. For the future, one can expect that remote sensing systems with increased spectral resolution (imaging spectrometers) as well as microwave sensors may further increase application perspectives. As most of the planned spaceborne hyperspectral systems, however, will collect data with a relatively moderate spatial resolution in order to keep data transmission rates within realistic limits, their use will demand the use of powerful algorithms for detecting surface characteristics at subpixel scales; spectral mixture analysis (SMA) is such a technique which, as it is based on the paradigm of linear additive mixtures of the reflectance from distinct material components, appears particularly suited to the assessment of dryland ecosystems with their low canopy structures and predominantly woody (i.e. nontransparent) plant components. However, it must be accepted that future low spatial/high temporal resolution satellite sensors will be facing specific detection limits as opposed to already operational remote sensing systems with high spatial resolution that provide spectral measurements in the most important wavelength regions needed for detecting plant- and soil-related characteristics.

2 DRYLAND ECOSYSTEMS

Arid and semiarid lands cover approximately one-third of the continental surface of the earth.[1] Their delineation is not uncomplicated since the main characteristics of dry ecosystems, i.e. the shortage of rainfall and resulting drought, exhibit substantial regional variations which must be related to variable temperature regimes. About two-thirds of them are located in warm climates (between 15° and 35° in both hemispheres), and thus belong to the deserts and semideserts of tropical and subtropical regions, respectively. In the northern hemisphere, cold drylands extend directly from tropical and subtropical regions until 55° latitude, and cover large parts of Central Asia and North America (Great Basin, Interior Plains); compared with that, we find only very limited areas of this type in the southern hemisphere (eastern Patagonia, southern New Zealand).

Besides referring to general characteristics such as limited water resources, highly irregular rainfall patterns, incomplete plant cover with mostly xeromorphic species, and specific risks for agricultural use, the definition of drylands is a matter of distinct parameters. In 1977, UNESCO[2] published its synthetic map of the distribution of drylands (Figure 1) where the degree of aridity is defined by the bioclimatic index P/PET [ratio between total annual precipitation (P) and potential average evapotranspiration (PET) (i.e. quantity of water lost by evaporation directly from the soil and by transpiration of a plant cover) from measurements at 1600 meteorological stations]. Four grades of aridity are distinguished (see also Mainguet[1]):

- *Hyperarid*: P/PET < 0.03, the average rainfall is 10–50 mm a^{-1} and the interannual coefficient of variability exceeds 40% (a zone corresponding to extreme deserts without vegetation, except for some ephemerals and xerophytic bushes in the beds of wadis).
- *Arid*: 0.03 < P/PET < 0.20, the average rainfall is 50–100 mm a^{-1} and the interannual coefficient of variability is 30–40% (barren areas or areas covered by sparse vegetation of perennial and annual plants; pastoral nomadism is possible, but no rain-fed agriculture).
- *Semiarid*: 0.20 < P/PET < 0.50, the average rainfall is 150–500 mm a^{-1} and the interannual coefficient of variability is 30–40% (covered by open vegetation cover where perennial plants dominate; extensive livestock breeding is possible, also rain-fed agriculture which, however, remains hazardous).
- *Subhumid–dry*: 0.50 < P/PET < 0.75, the average annual precipitation is 500–800 mm, spread over a rainy season of 6 months. Aridity is a recurring seasonal constraint, and the interannual coefficient of variability may still reach 20% (permanent rain-fed agriculture is practised with cultures adapted to seasonal drought).

It is seen that all continents, including even Europe with its Mediterranean fringe, possess drylands where aridity and drought are the two most apparent climatic facts (*aridity* implies a permanent pluviometric deficit; *droughts* result from a temporary pluviometric deficit in relation to the normal precipitation). A more detailed review of arid ecosystem characteristics is provided by Goodall and Perry,[4] special emphasis on arid soils is, for example, given by Dregne,[5] and specific management issues are thoroughly discussed by Beaumont.[6] In the following section, some of those issues which are of specific importance when the use of remote sensing for an

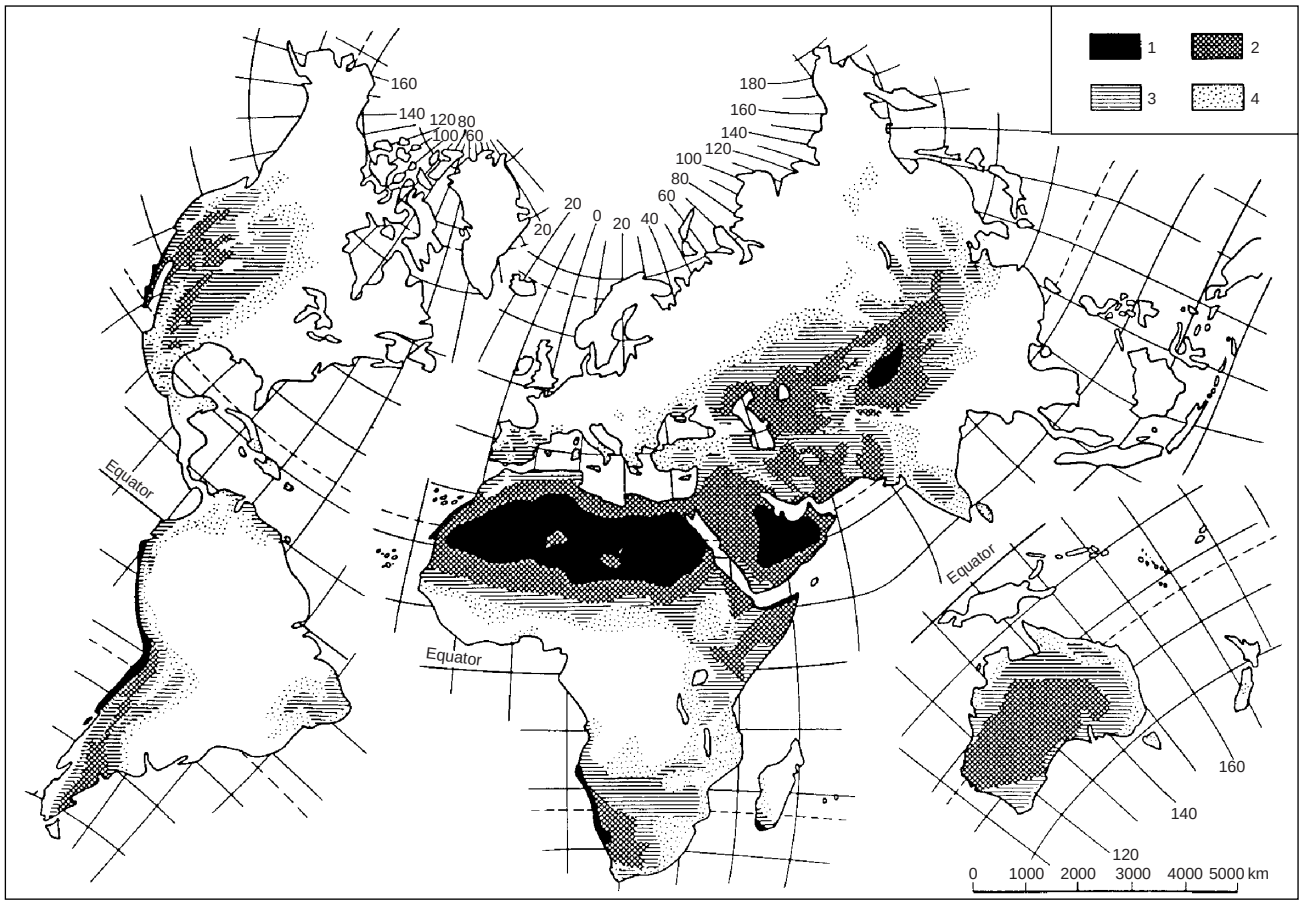

Figure 1 World distribution of drylands.[2] The different types are: 1, hyperarid; 2, arid; 3, semiarid; 4, subhumid–dry. [Reproduced by permission from Mainguet.[3]]

assessment of arid lands is taken into consideration are reviewed in more detail.

2.1 Climate and Vegetation

The average annual rainfall in arid to semiarid ecosystems roughly varies between 0 and 500 mm a^{-1} and, in general, remains below the potential evapotranspiration rate. The distribution of rainfall over time is highly irregular, such that dry spells may also occur during so-called ''wet'' periods. Since most dry ecosystems are characterized by frequently cloud-free atmospheres, global irradiance is higher than in most other ecosystems and, as ground albedo amounts to higher values than in more humid regions, radiation balance in arid to semiarid regions is only slightly positive, or even negative. Rock, soil and plant surfaces become intensely heated during the day whereas, owing to the low moisture content of the atmosphere and soils, temperatures decrease rapidly during night-time.

Plants of arid and semiarid ecosystems have to adapt to frequently occurring periods of heat and water stress by developing xeromorphic characteristics (e.g. reducing the size but increasing the number of leaves, increasing the number of stomata per leaf area, etc.). Typically, drylands possess an open vegetation cover; in moderately dry areas the vegetation is distributed uniformly where proportional cover appears directly related to the amount of rainfall or, more precisely, plant-available water. With severely increased aridity, i.e. under desert conditions, this uniform distribution is replaced by a concentration of plants in channels of the drainage network; this ''contracted'' vegetation rarely covers more than 10% of the surface.[7] The sensitive nature of the vegetation cover in drylands is a result of their weak persistence and stability. This reflects both their dependence on precipitation (especially their variability) and their resilience, i.e. their capacity for rehabilitation when favorable conditions are re-established.[1] It is a direct consequence that plant density or proportional cover is one of the most important parameters for assessing the resource potential or state of the environment in semiarid

regions, often more important than species composition. Differences in proportional cover may indicate climatic or hydrological gradients, or respond to effects of edaphic aridity (as described in the following section); vegetation changes over time may relate to climatic fluctuations and/or provide evidence for human pressure upon the environment. More important even, proportional plant cover as one of the most important environmental indicators is directly tangible through remote sensing systems,[8] provided that specific precautions are taken with regard to the methodology of data analysis.

2.2 Soils

Apart from the soils of steppe areas (e.g. phaeozems, chernozems) with their particularly high amount of organic carbon, soil formation tends to decrease rapidly with increasing aridity since, owing to the almost continuous shortage of water, chemical weathering becomes less important than in humid regions. Soils in semiarid regions often have very little organic material and therefore consist almost exclusively of rock and mineral particles. Also, the importance of leaching processes (i.e. transport of soluble salts, carbonates, Fe and Al oxides, fulvic acids and clay minerals) is reduced and, under increasing aridity, is replaced by the movement of free carbonates and salts with ascending water (capillary rise). Arid soils (xerosols) are thus characterized by relatively high concentrations of carbonates, soluble salts and, depending on the parent substrate, SiO_2-rich components in the upper soil horizons. Frequently, these form hard crusts (i.e. calcrete, silcrete) close to the surface which considerably influence infiltration capacity and the soil water budget, but also the spectral properties of the surface. Microphytic communities tend to produce coatings on rock and soil surfaces in arid and semiarid areas.[9] The mineral grains of soils are then held together by a fibrous mat of cyanobacteria and other biological elements;[10–12] these biological crusts may considerably modify runoff and infiltration characteristics of the parent substrate[13] and contribute to the stabilization of surfaces.[14]

Although already mentioned implicitly, it must be emphasized that soil texture appears to be the most important factor for the water balance of arid soils as it controls infiltration and the capillary rise of water. In fine-textured (i.e. loamy) soils rainfall can hardly infiltrate and is therefore subject to evaporative losses. More coarsely textured soils (e.g. sandy or stony substrates) permit water to infiltrate rapidly to greater depths where it is well protected against evaporation after the topsoil layer dries out; additionally, almost all infiltrated water is readily available to plants, and salt concentrations in the soil remain low owing to the higher leaching efficiency. This implies that care must be taken as the postulated positive relation between increasing average rainfall and leaching intensity is to be modified with regard to local surface properties: water availability and therefore the rate of water infiltration and leaching are positively related to the rock/soil ratio (i.e. the proportion of runoff-generating rocky surfaces compared with the water-receiving soil surfaces per unit area). Relationships between climate and environmental conditions which are valid on a global scale therefore become problematic when local factors such as topography, lithology and soils determine the redistribution of water.[15] It is therefore mandatory that remote sensing needs to focus on the precise characterization of surface properties if it is supposed to contribute to environmental assessments of arid lands, a point which needs to be discussed again in the context of the ongoing debate on desertification and climate change.

2.3 Land Use

Although drylands belong to the less densely populated regions of the world, extensive rain-fed agriculture and/or grazing are common features of the dry mid-latitudes or the subtropical regions. The frontier between these land use systems is drawn by a rainfall threshold up to which rain-fed agriculture is still possible; depending on temperature regimes, soil characteristics (i.e. edaphic aridity) and specific cultivation techniques (dry farming) this rainfall limit is reached between 250 and 350 mm a^{-1}.[16]

The main agricultural products of rain-fed agricultural systems in drylands are cereal crops (wheat, barley) which provide more than 50% of the world production. As the yield per unit area is relatively low, the farms tend to compensate for this by increasingly large production acreages and a high degree of mechanization. However, a sharp distinction needs to be made between developed and third-world countries: in the latter, the pressure of an enormous population growth severely limits the availability of land resources and imposes the need for more intense production systems (i.e. at the expense of higher risks of failing).

Extensive grazing in semiarid ecosystems is conducted in the form of either ranching or nomadism. Ranching is a modern organizational scheme for extensive grazing and was mainly developed in the USA and Australia, while grazing in the drylands of Africa and Central Asia is still dominated by nomads. One of the most important environmental risks results from exceeding the carrying capacity of the land by increasing the number of cattle per unit area. This does not only occur in well-developed economies but many cases are known where the abandonment of nomadism has led to severe land degradation processes which are usually summarized under the term ''desertification''.[3,17] Evidently, remote

sensing has to play an important role in monitoring the effects of such degradation processes where, once more, the focus has to be on the detection of surface properties related to vegetation and soil resources. On the other hand, remote sensing not only has to interact in the field of disaster assessment and mitigation but, even more important, in analyzing the resource availability for sustainable agricultural developments in drylands.

In many dryland ecosystems with accessible water resources, irrigated agriculture has been introduced. Of course, superb yields are guaranteed (often at the expense of excessive fertilizer and pesticide input into the system) as long as the water resources (frequently fossil groundwater) are not exhausted. Environmental problems usually result from high salt and Na concentrations in the soil, in particular when the supply of irrigation waters is not adequate to provide sufficient leaching. Remote sensing may primarily contribute to the resource management in irrigation areas by mapping the spatial extension of specific cultures (for which water consumption is known), or by detecting indications of salt accumulation in topsoil substrates. Efforts have also been devoted to the use of space techniques for deriving energy fluxes for water management.[18–21]

3 REMOTE SENSING AND THE ASSESSMENT OF ENVIRONMENTAL CONDITIONS IN SEMIARID REGIONS

Remote sensing is commonly introduced as the science of collecting information about objects without coming into physical contact with them; in earth observation the most important medium for transmitting this information is electromagnetic radiation in the optical and microwave region. There are numerous excellent textbooks available which give a thorough description of the physical background,[22–24] basic principles of image interpretation,[25,26] digital image processing,[27–29] and specific sensors and fields of application.[30–35] Most of these books, however, tend to emphasize either the optics and physics of remote sensing or digital image processing and sensor aspects.

Here, we shall attempt to treat remote sensing assessments of arid and semiarid ecosystems as a continuous process which includes aspects of energy – matter interaction, radiation propagation, sensor characteristics and image processing – an approach which has recently been devised as the image interpretation chain.[36] It is in the consequence of this approach that one has to go beyond emphasizing the general advantages of remote sensing such as providing a synoptic, repetitive and consistent perspective over large areas when discussing its suitability for environmental assessments. Primary importance should be attributed to analyzing more closely which relationships exist between parameters of interest and corresponding electromagnetic signals from the earth.

3.1 The Importance of Assessing Surface Properties

It is widely agreed that environmental change in arid, semiarid and dry subhumid ecosystems is not necessarily driven by climatological variables but frequently triggered by processes which result from adverse human impact on these fragile ecosystems.[3,37] Also, the productivity of dryland systems largely depends on surface properties which, as they control water availability and the spontaneous emergence and development of new plants and dust production during wind storms, might dominate climatic variables. It follows that surface properties are to be considered more indicative than long-term climatic conditions. The ability to draw concise conclusions with respect to land resources and environmental change will thus depend on the capability to assess specific surface characteristics (i.e. vegetation cover and composition, parent material and soils, including mineralogical and biological crusting) from remote sensing data, and to analyze multiannual time series of images. Here, retrospective studies might be as important as continuous monitoring of environmental change by remote sensing.[38]

In addition to the presence or absence of, and the temporal changes in, the abundance of photosynthetic and dry (i.e. woody) plant components, important surface properties for the assessment of semiarid ecosystems are the type of parent material (lithology), the proportion of rock and soil surfaces per unit area (as the rock/soil ratio carries important implications for water concentration and increased leaching capacities above climatic averages), specific soil properties (such as the content of organic carbon or iron oxides which assist in identifying whether a soil profile has been truncated by erosion processes, or whether it is to located in a sediment and nutrient sink), the presence of sand sheets and dunes (with their associated capability to protect infiltrated water against evaporation) and biological crusts (with their positive feedback on sand stabilization and runoff concentration). All these can play a decisive role not only in the assessment of land resources but also in the context of assessing desertification dynamics.[17] However, one has to abolish too simplistic views such as to expect that remote sensing assessments may automatically flag ''desertified'' regions by, for example, indicating a change in the proportion between potential and real evapotranspiration (notwithstanding the existing methodological problems). What remote sensing assessments can provide is to reveal physical changes in surface properties; however, whether the intrusion of a sand sheet, for example, is to be understood

as an indicator of desert encroachment (i.e. degradation) or as providing improved conditions for the infiltration of rainfall and its protection against evaporative losses (i.e. a resource upgrade), is to be decided by the user/expert in the local and regional context, respectively. Remote sensing, if handled with appropriate methodological and conceptual knowledge, provides quantitative and qualitative indicators but not directly their interpretation. Additionally, it must be recognized that the instantaneous state of the surface may fluctuate seasonally and/or annually in response to rainfall, illustrating the dynamic nature of the system.[39] Repeated observations over longer time periods are therefore indispensable for assessing significant changes.

3.2 Remote Sensing Systems

When compared with the early 1970s when the first Landsat system was placed in a space orbit, remote sensing systems exhibit an enormous diversity in terms of spectral, spatial and temporal parameters,[35] and which system is to be used depends on the user requirements (Figure 2). If a frequent repetitive coverage with relatively low spatial resolution is desired (e.g. for meteorology), one would certainly be inclined to base the approach on the AVHRR system available from the polar-orbiting satellites of the NOAA series, or even on data from geostationary satellites such as METEOSAT or GOES.[40] Alternatively, one might look for the highest spatial and spectral resolution available, even at the expense of relatively low repetition rates, and would thus choose one of the available earth observation satellite systems (Landsat, SPOT, IRS).

The modern era of satellite remote sensing started with the multispectral scanner system (MSS) in 1972, which provided for the first time high-resolution earth images in digital form [four spectral bands in the visible (VIS) and near-infrared (NIR) region, 80 m spatial resolution, 180 × 180 km frame size]. Two identical follow-up systems (Landsat-2 and -3) were launched during the 1970s before, in 1982, the second Landsat generation with the Thematic Mapper (TM) [30 m spatial resolution (the thermal band of the Landsat TM has a spatial resolution of 120 m), seven spectral bands in the VIS, NIR and short-wave infrared (SWIR), and thermal infrared (IR)] and, in 1986, 1990 and 1993 the SPOT high-resolution visible instrument (HRV) with 20 m spatial resolution in three spectral bands, and 10 m resolution in a panchromatic

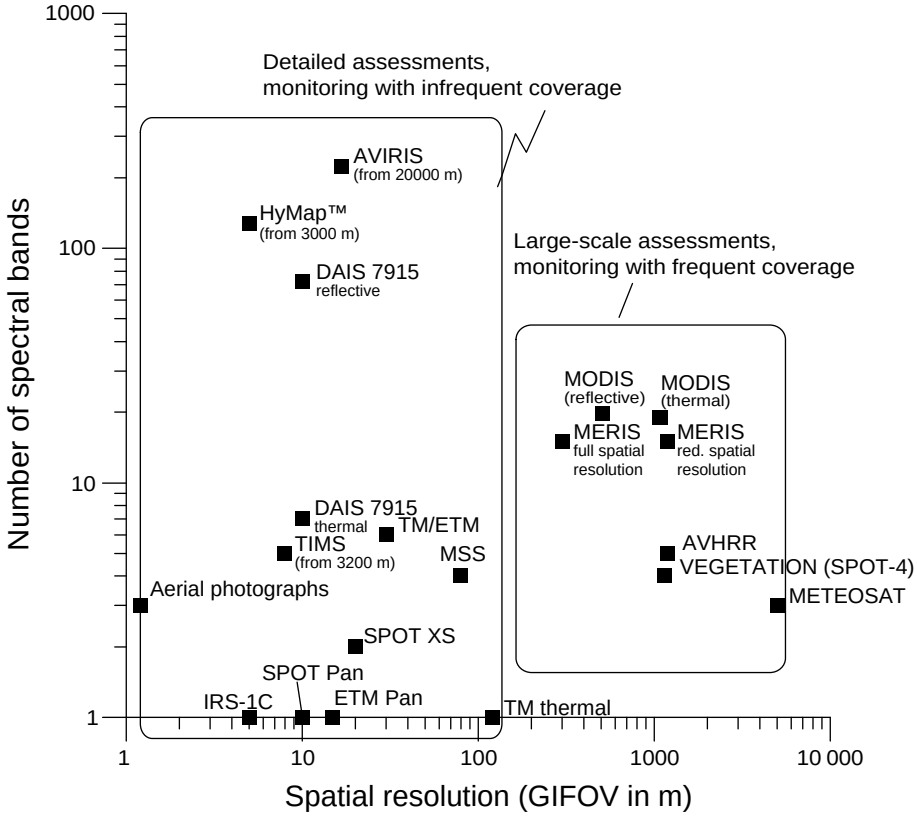

Figure 2 Important operational, planned, and experimental remote sensing systems in a two-dimensional parameter space determined by spectral and spatial resolution. DAIS, Digital Airborne Imaging Spectrometer.

channel, became available. Recently, the importance of SWIR channels for operational satellite systems has been underpinned by the layout of the HRV and the vegetation monitoring system (VEGETATION) on SPOT-4. Also, the new concept of imaging spectrometry, i.e. the acquisition of reflectance signatures in numerous narrow and contiguously spaced spectral channels,[41] has meanwhile provided the user community with a range of powerful, yet largely experimental airborne sensor systems such as the airborne visible and infrared imaging spectrometer (AVIRIS) and HyMap™.[34,42] Considerable efforts have been made to construct hyperspectral imaging systems which are able to observe the earth from space orbits; however, to find a suitable cost–benefit compromise the planned systems, such as the moderate resolution imaging spectrometer (MODIS) (to be provided by NASA) and the medium resolution imaging spectrometer (MERIS) on board the future environmental research satellite (ENVISAT) platform [European Space Agency (ESA)], will collect their high-spectral resolution data with rather moderate spatial detail (250–1200 m).

Which systems are to be considered the most useful ones for arid ecosystem observation is largely to be seen in relation to the basic considerations in the previous section. As most information on the composition of surface elements (vegetation, soils, rocks, biological crusts) is provided by passive remote sensing systems, we shall limit our discussion to sensors that measure electromagnetic radiation either emitted directly or reflected by the objects. Although spectacular images from Shuttle-based radar experiments have revealed the existence of palaeodrainage systems in the eastern Sahara,[43] the use of active remote sensing systems (i.e. synthetic aperture radar) is, in arid ecosystem research, to a large extent still experimental, and the discussion on microwave remote sensing therefore is beyond the scope of this application article. Also, no emphasis is given to thermal sensors since a clear strategy to employ these systems in arid ecosystem research is not recognizable at the moment. With regard to the importance of surface properties in arid regions, there is of course a connection with geological applications of thermal remote sensing,[44] but a direct link to nongeological resource assessments has not yet been established, apart from the use of thermal data for estimating fluxes in irrigated areas. The description of the energy exchanges at the earth's surface represents a major sub-discipline of meteorology, where in particular the evapotranspiration term is of great interest to agriculture and water-use specialists.[21,45] Although considerable progress in processing thermal IR images has been achieved,[46,47] a major problem results from the fact that, in conjunction with the remote sensing data, explicit values of some meteorological variables are needed which are not readily found over large areas. Also, thermal remote sensing provides instantaneous measurements, and these can only be extrapolated to daily values under relatively stable atmospheric conditions; it is therefore not surprising that this type of approach has not yet achieved operational status.

The focus is therefore on air- and space-borne multispectral systems, including high spectral resolution sensors (i.e. imaging spectrometers), but also on color aerial photographs which still provide a cost-attractive yet efficient tool for surveying important ecosystem parameters in arid and semiarid environments. The traditional techniques of air-photo interpretation are well established and routinely used,[26] and will not be considered in much detail. It is only to be mentioned that digital image processing has also opened up new avenues for analyzing aerial photographs.[48] Emphasis will be primarily placed on the processing chain, i.e. on digital image processing, radiometric analysis and thematic concepts for assessing target properties.

4 REMOTE SENSING DATA ANALYSIS

Optical remote sensing systems measure the spectral properties of surfaces; they cannot measure specific resources or land degradation directly. The major problem associated with their use is to interpret quantitatively a measured signal that has interacted with remote objects in terms of the properties of these objects. This requires a thorough understanding of the radiation interactions with those elements that characterize the surface conditions in which we are finally interested.

However, before discussing appropriate scene models, one needs to understand the initial part of the processing chain (i.e. data preprocessing) which deals with the geometric rectification of digital imagery and with turning uncalibrated image gray values into physical quantities. Engineering data about the detector sensitivity (i.e. calibration coefficients) permit the multispectral scanner to reconvert an encoded digital number (DN) into measured radiance, and radiative transfer calculations can be used to correct for atmospheric effects, such that the surface-reflected radiance is restored from the satellite-measured signal. Dividing this term by the downwelling solar irradiance provides us with an important primary parameter which is termed bidirectional reflectance, ρ (Figure 3); albedo and surface temperature are other primary parameters which, as a result of similar processing chains, can be derived from optical remote sensing systems.

The processing chain further involves appropriate scene models which can be used to convert primary parameters into thematic information. Here the problem

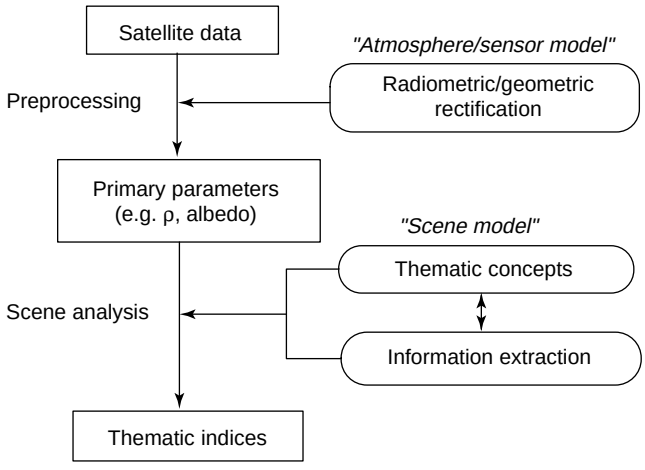

Figure 3 The conversion of satellite raw data into standardized thematic information layers.

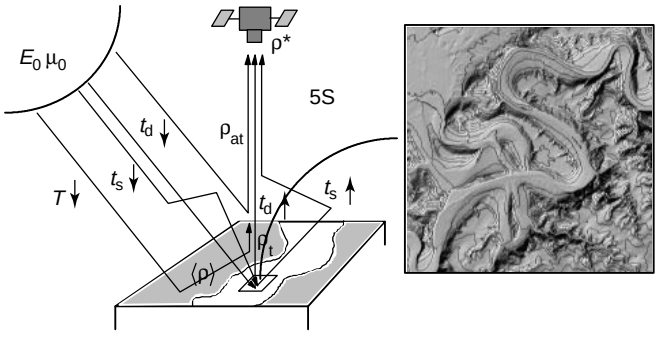

Figure 4 Schematic representation of the modified 5S code for atmospheric correction of multispectral data in the reflective domain of the optical spectrum. If digital elevation data with an appropriate spatial resolution are available, it is also possible to correct the terrain-induced illumination differences. [After Hill et al.[54]]

is not so much to identify a particular surface type, but to characterize it. A variety of methods have been proposed which range from empirical spectral indices to the design and inversion of physically based models. While the applicability of the various approaches depends on the nature and accuracy of the desired information and the availability of resources (i.e. sensor characteristics), an important prerequisite for their operational use is that they must satisfy specific requirements in terms of standardization and portability. Therefore, an important issue in satellite remote sensing is to conceptualize and streamline the data analysis in such a way that consistent indicators for environmental assessments can be obtained.

4.1 Data Preprocessing

In the context of this application article, not too much emphasis will be placed on the geometric rectification of digital images since the techniques and methods to be used are well described and considered largely operational.[27,29]

More important are the methods which are used to compute primary (i.e. physical) parameters from gray-level images. Today, we are in the favorable situation that various radiative transfer codes are available which, if constrained with the same atmospheric parameters, provide consistent results.[49] Frequently, atmospheric correction methods are employed which are based on the formulation of radiative transfer as developed by Tanré et al.[50] Modifications have been applied to this concept in order, for example, to account for atmospheric extinction processes as a function of sensor and terrain altitude[51,52] and to compensate efficiently terrain-induced illumination effects by incorporating digital elevation data into the correction process.[53]

This modified 5S code (Figure 4) makes extensive use of analytical expressions and preselected atmospheric models, resulting in a short execution time. It provides corrections for atmospheric absorption, scattering and pixel adjacency effects, where diffusion and absorption processes are assumed to be independent. Upward and downward transmission coefficients are derived by introducing the auxiliary quantity of optical thickness τ which measures the total extinction of a light beam due to molecular and aerosol scattering when passing through an airmass. Multiple scattering is accounted for, and the absorbing atmospheric gases (H_2O, O_3, CO_2, O_2) are assumed to condense at the top of the atmosphere and at the top of the layer between the earth's surface and the sensor altitude. With this or comparable transfer codes, reflectance factor retrievals from the Landsat TM bands have been achieved with an accuracy of ± 0.005 to 0.02,[55–57] and it is important to note that this level of precision can also be accomplished when the scattering optical depth is estimated from scene data itself.[51] The major problem in retrieving reflectance factors from operational satellites lies in the sensitivity to the absolute radiance calibration of the sensors. However, there is an increasing awareness in the scientific user community towards this problem. In-flight calibrations for Landsat TM and SPOT data are conducted at various high-reflectance sites which is useful for determining at least an absolute radiometric calibration gain and monitoring the sensor degradation with time.[58,59] However, it must be also noted that the successful use of radiative transfer calculations for high spectral resolution imagers depends on the availability of key parameters that adequately characterize atmospheric conditions during the flight (i.e. absorption optical depth, in particular that of water vapor).

Alternatively, corrections of radiometric scene-to-scene variations for retrospective and monitoring studies

can also be obtained through empirical approaches which are based on the analysis of radiometrically invariant surfaces.[36,60] However, it needs to be recalled that these techniques only produce relative adjustments of variable scene radiometry as long as the reference scene is not atmospherically corrected. The use of scene models that are based on absolute reflectance might therefore be seriously restricted or impossible. Also, radiometric normalization is not able to account for nonlinear effects such as the variation of atmospheric conditions with terrain altitude, the reflectance contribution from the environment of a pixel and terrain-induced variations of illumination, and, therefore, reflectance.

4.2 Scene Models: Fundamental Considerations, Indicators and Interpretation Concepts

In order to assess land resources or land degradation processes, it is necessary to define diagnostic indicators. These may be primary indicators (e.g. high salt content in salinized soils) or secondary indicators which are produced by the problem (e.g. reduced vigor of vegetation). The most useful indicators are those which have distinct spectral signatures and are unique to a particular issue.[61] As part of the scientific background to this application article, we therefore need to discuss some issues in more detail, such as principles of soil and vegetation reflectance, limitations of conventional vegetation indices, detection thresholds for photosynthetic vegetation, the influence of specific surface constituents in arid regions (i.e. biotic crusts, desert varnish) and spectral characteristics related to nonphotosynthetic vegetation and variable surface conditions. This is essential for evaluating the use of remote sensing for resource inventory and monitoring, and for selecting appropriate indicators and sensor systems. Besides, it should not be forgotten that the development of suitable indices derived from remote sensing measurements and their interpretation for resource assessments and the surveillance of degradation processes (desertification) also require a conceptual framework which allows one to draw concise conclusions with regard to land surface conditions. Although these underlying concepts might vary as a function of regional ecosystem characteristics (i.e. in relation to physiographic conditions such as parent material, aridity, etc.), it should be ensured that results from different regions can be consistently evaluated on a higher hierarchical level.

4.2.1 Soil Reflectance Characteristics

The spectral reflectance of soils is a cumulative property which derives from the inherent spectral behavior of heterogeneous combinations of minerals, organic matter and soil water.[62] The most comprehensive studies to derive a typology of soil reflectance classes were conducted by Condit[63] and Baumgardner et al.;[64] the latter measured the reflectance of 564 soil samples from soil series in the USA and Brazil in the wavelength range between 0.52 and 2.32 μm, and they were able to define five major soil reflectance types which are identified by curve shape and the presence or absence of spectral absorption bands caused by organic matter content, iron oxides and soil minerals (Figure 5). It is believed that any observed soil spectrum resembles one of these spectral curves.

Reflectance spectra of moist soils exhibit strong absorption bands centered at 1.4 and 1.9 μm (Figure 5) which, together with weaker absorption bands at 0.97, 1.20, and 1.77 μm, are attributable to overtones and combinations of the fundamental vibration frequencies of water molecules in the soil. The bands at 1.4 and 1.9 μm are broad, indicating an unordered arrangement of water molecules at various sites in the soil.[64] The relevance of individual mineral spectra to soil reflectance was reviewed by Mulders.[65] Although absorption features in the 2.0–2.5 μm range are of limited use for the identification of soil characteristics (because soil materials are complex mixtures of primary and secondary minerals both of which can already be part of the parent material), spectral absorption in the VIS to NIR range conveys important information about the presence of iron oxides. Even trace amounts of hematite and goethite can cause absorption bands and, in spite of the difficulties outlined by Geerken,[66] quantitative relationships can be established when only a limited range of soils, for example tropical soil types, are considered.[67] Since soil color is largely determined by spectral absorption

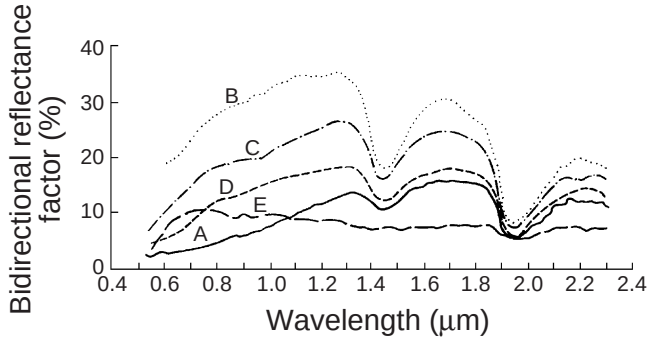

Figure 5 Characteristic soil bidirectional reflectance spectra. (A) Developed, fine textured soils with high (>2%) organic matter content; (B) undeveloped soils with low (<2%) organic matter and low (<1%) iron oxide content; (C) developed soil with low (<2%) organic matter and medium (1–4%) iron oxide content; (D) moderately coarse textured soils with high (>2%) organic matter content and low (<1%) iron oxide content; (E) fine textured soils with high (>4%) iron oxide content. [Reproduced by permission from Baumgardner et al.[64]]

features, attempts have been made to relate spectral measurements to standard color description used for soil description, such as the Munsell Soil Color Charts. It was shown that soil spectra can be simulated from Munsell color data, and that soil color can be remotely sensed when a limited number of reflectance measurements from the VIS spectrum are available.[68,69]

The essential question is how spectral characteristics, including soil color indices, can be used to assess pedogenic properties or the state of soil resources in arid, semiarid or subhumid ecosystems. It is difficult to provide a general answer on this issue. An approach which, without any doubt, is useful for remote sensing studies refers to concepts which consider soil development to be either progressive or regressive with time.[70] Under progressive development, soils become better differentiated by horizons, and horizon contrasts become stronger. Pedogenetic processes involve the formation of clay-sized particles by weathering of larger grains, the alteration of clay minerals to other clay-mineral species, and the release and accumulation of iron by weathering. Some solids (silt, clay and $CaCO_3$) and ions (Ca^{2+}, Na^{+}, etc.) dissolved in rainwater are added from the atmosphere, and topsoil organic matter contents increase with the decomposition of plant and animal residues. Transfers within the soil profiles result in the accumulation of silt and clay, Fe, Al, $CaCO_3$, gypsum or halite in the B horizon, or, owing to bioturbation processes, to the soil surface. In contrast, regressive pedogenesis refers to the addition of material to the surface at a rate that suppresses soil formation (i.e. eolian dunes, sand encroachment, glacial moraines, distal fans, etc.), or suppression of pedogenesis by surface erosion. Both progressive and regressive pedogenesis cause alterations of the soil surface that, to a certain extent, are spectrally detectable.[53,71,72] Under more humid conditions (i.e. more than 300 mm annual rainfall) this may be related to the intensity of soil brunification and rubification; another indicator may be the organic matter content of the topsoil material.[73] However, it must be kept in mind that, in well-developed soil profiles, such models may not be applicable because extensive erosion may also occur without exposing the subsoil or zones with significant changes in soil mineralogy. In more arid ecosystems additional effects such as the soil/rock ratio, sand encroachment into rocky or loessic environments or the presence of biotic crusts may become very important.[15]

As this concept seems to provide a widely applicable, general framework for the successful use of remote sensing techniques, the objective of differentiating between more or less favorable areas merely condenses on classical detection problems, i.e. to identify specific tracer substances or surface expressions, and then to define optimized methods for maximizing the detection capabilities based on specific sensor systems. The methods to apply cover a wide range, including semiempirical relations between specific spectral variables and soil parameters, spectral mixing models and color indices derived from two or more spectral variables in the VIS range of the optical spectrum. Some examples will be reviewed more closely in section 5. Although the potential for mapping soils on a compositional basis seems to increase through the advent of hyperspectral imaging systems, a number of problems remain: the effects of vegetative cover must be separated from the mineralogical response, and the influence of the moisture content on the signal (which is intrinsically variable) must be efficiently compensated. It should also not be forgotten that erosion mapping using the vegetation response has been successfully adopted as an alternative approach.[74] Vegetation in dryland systems, however, exhibits specific characteristics which are discussed in the following section.

4.2.2 Reflectance Characteristics of Vegetation in Dryland Systems

Although sparse vegetation communities, typical of deserts and semiarid regions, make only a small contribution to the global carbon pool, they are an important indicator for environmental conditions, including disturbances introduced by human actions. The high spectral contrast of vegetation relative to soils has provided many successful studies which use optical remote sensing systems for observing vegetation communities.[75,76] Since plants have a distinct spectral signature with low reflectance in the VIS part of the spectrum and high reflectance in the NIR region, attempts have been made to use this spectral feature for estimating green vegetation cover and biomass through various, but functionally equivalent, ''vegetation indices''.[77] The most widely used indices are the simple ratio (SR) and the normalized difference vegetation index (NDVI) defined as $SR = \rho_{NIR}/\rho_{RED}$ and $NDVI = (\rho_{NIR} - \rho_{RED})/(\rho_{NIR} + \rho_{RED})$, where ρ_{RED} and ρ_{NIR} is the measured reflectance in the red and in the NIR spectral regions, respectively. These indices have frequently been used with the AVHRR data from the NOAA satellites, but also with high spatial resolution instruments, such as SPOT or Landsat.

The plant cover of arid regions usually consists of perennial plants (which are neither abundant nor very dense) and some short-lived (i.e. annual) herbaceous vegetation (which only appears for some weeks after the rainy periods). However, owing to their morphological and structural adaptations, the spectral response of the perennial plants in dry ecosystems is different from the typical reflectance of green plants. Because water

and chlorophyll pigment contents of the green leaves are low and the dominant parts of perennial desert plants are woody components, one finds that important spectral features of photosynthetic plants (i.e. chlorophyll absorption around 450 and 670 nm, the steep spectral transition of the red edge between 650 and 750 nm, and liquid water absorption around 1450 and 1950 nm) are much less developed than in plants from more humid regions. [This, of course, is not the case for irrigated crops with sufficient water supply (Figure 6).] The spectral response of vegetation in arid areas thus depends mainly on the proportions of green to dry and woody parts. An additional complication is introduced through the fact that dry and woody plant material exhibits not a unique but a highly variable spectral response, depending mainly on the level of plant decomposition.[78]

A commonly used approach for retrieving vegetation parameters from spectral vegetation indices involves principally three steps:[79]

- for a given date of satellite overpass, the derivation of a relationship between a spectral vegetation index and one vegetation parameter, based on the statistical analysis of satellite-derived index values obtained from a sample of contrasted reference sites with known values of fractional cover of green vegetation (F_g) or leaf area index (LAI) (from simultaneous ground measurements);
- inversion of the relationship and application to each pixel of the image encompassing the reference sites;
- spatial and temporal extrapolation of the results.

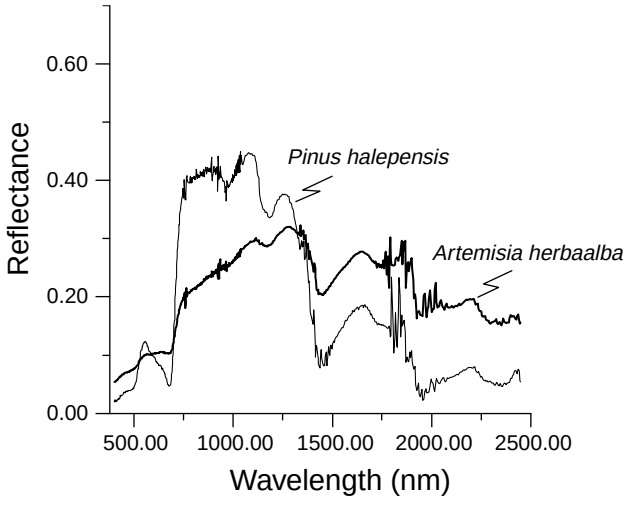

Figure 6 Spectral reflectance of photosynthetic plants (*Pinus halepensis*) compared with a spectrum of desert shrub (*Artemisia herba alba*) which includes a large proportion of dead stems and litter.

The last point is the more problematic, because empirical relationships may be valid only within the limited range of vegetation and soil conditions occurring within the studied area and therefore not transferable to other situations.[8] Nevertheless, some empirical relationships between NDVI and fractional cover (F) or fractional cover of green vegetation (F_g) have been proposed[80,81] but, as NDVI is influenced by soil albedo variations,[82] it is suggested that one should use ''soil-adjusted'' or orthogonal indices, especially in the case of sparse canopies. Other limitations in deriving regression-type relationships between spectral vegetation indices and ground-measured structural parameters are due to measurement errors in the field. The number of direct measurements is generally low because they are time-consuming and labor-intensive, and indirect measurements may not provide consistently accurate estimates of vegetation parameters. Similar restrictions apply to estimates of the above-ground biomass,[79] and it must be realized that the uncertainties of empirical relationships will easily increase owing to scale problems as a function of the sensor spatial resolution.[83]

Some concern has also been expressed about the sensitivity of spectral vegetation indices to the state of the atmosphere, illumination and observation geometry, and, in particular, the background reflectance of soils and rocks.[84,85] While some of these disadvantages can be compensated by the use of atmospheric corrections and/or by compositing a number of days of data, Pinty and Verstraete[76] have proposed an alternative index [global environment monitoring index (GEMI)] which is assumed to compensate for most atmospheric and illumination conditions. However, this new index also is not able to overcome the major drawback of two-band indices, which is the contamination of vegetation abundance estimates due to the background reflectance of soil and rock surfaces. Recently, it has once more been shown by Price[46] that bare soil spectra alone may occupy already the lower 30% of the usable NDVI value range between 0 and 1 (Figure 7). In view of these undesired properties, we would of course prefer to use an approach which subtracts, or corrects for, the contribution of soil-reflected radiation instead of accepting the resulting measurement uncertainty.

Several modifications have been suggested[84,86–88] which reduce primarily the influence of moisture-induced soil brightness differences but not the influence of wavelength-dependent variations in soil–rock spectral response. However, they do not provide a generic solution for compensating the side effects from spatially variable substrate types. To summarize, a breakthrough based on the use of band ratioing techniques, i.e. within the concepts of two-band indices, could only then be expected when soil spectra would exhibit no slope between 0.6 and

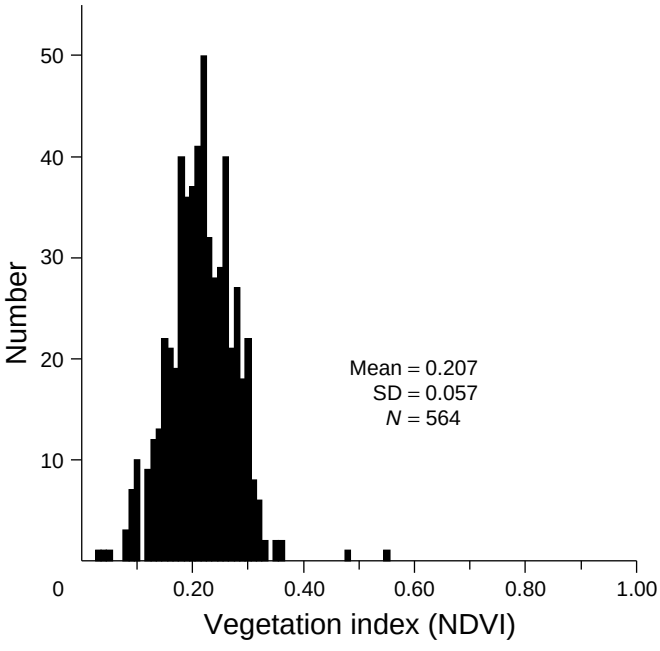

Figure 7 Histogram illustrating the number of soil spectra having given values of NDVI, at intervals of 0.01. The spectral data are from 564 spectra contained in the Purdue University soil spectral database (see also Figure 5). [Modified after Price.[46]]

0.8 μm, which is in contradiction to our knowledge of soil spectral reflectance (Figure 5). A more specific limitation of conventional vegetation indices is determined through the fact that they are only consistent with green vegetation that contains active chlorophyll and water in the foliage. Since vegetation in arid regions is not always alive, we have indeed to pose the question of how ratios and indices such as the NDVI function under conditions when most of the plant biomass is dead stems and litter.

4.2.3 Reflectance Properties of Biotic Crusts and Surface Coatings

Rocks and soils are exposed to various processes that strongly affect their surface characteristics. A major effect of chemical weathering is an enrichment of Fe-bearing minerals at the rock surface. Good solubility of iron in connection with capillary force and precipitation supported by bacteria results in limonitic crusts that, owing to Fe charge-transfer bands induced by the iron-rich crust, cause strong absorptions in the VIS range between 0.4 and 0.6 μm. The dominant spectral effect in the NIR and SWIR region is a uniform decrease in overall reflectance, but the increase in absorption in the VIS range might simulate a substrate richer in iron than actually present.[66,89,90] Also microphytes form coatings on rock and soil surfaces in arid and semiarid areas,[9] where the kind of populating organisms (mosses, lichens, algae, fungi, bacteria and cyanobacteria) often depend on the orientation of rock surfaces. In the Negev highlands of Israel, for example, south-facing surfaces of limestones are populated by endolithic cyanobacteria carving pinhead holes, whereas north-facing surfaces are covered by epilithic lichens forming smooth-faced coatings.[91] The effects of various lichens on the spectral behavior of rock surfaces can hardly be predicted, and in addition to overall intensity changes (albedo) we also find a dependence of spectral response on the type of microphytes.[89]

On sandy substrates one frequently finds biogenic crusts, where the mineral grains are held together by a fibrous mat of cyanobacteria and other biological elements.[15] The influence of the crust on runoff and infiltration capacity was, for example, examined by Yair,[13] and their spectral properties have been studied to some extent by Karnieli and Tsoar[92] and Hill et al.[48] The photosynthetic activity of microphytes, algae and mosses causes spectral absorption due to chlorophyll around 670 nm, which is expected to influence abundance estimates of vascular plants derived from multispectral remote sensing data. Hill et al.[54] had noticed similar effects when measuring green foliage cover with high spectral resolution data over lichen- and algae-covered limestone surfaces in France: sparse vegetation was overestimated when it occurred on limestone/lichen-covered background which provided a photosynthetic component on its own. However, they could also demonstrate that improved estimates were obtained when the corresponding background spectrum was introduced in the analysis.[53]

4.3 Improved Scene Models: Detecting Surface Properties Based on Spectral Decomposition Modeling

With regard to the approaches discussed so far, one needs to draw attention to the fact that uncertainties in measuring vegetation abundance with optical sensors can be minimized by accounting for the reflective properties of sometimes highly variable background materials.[53,93] Conversely, natural vegetation can significantly mask and alter the spectral response of the ground. Therefore, instead of attempting to develop separate soil- or vegetation-related indices which are based on spectral reflectance measurements in single or multiple bands, it seems more appropriate to use dedicated spectral decomposition techniques.

One of the most promising approaches for computing the proportional abundance of materials which occur within a specific surface area (i.e. pixel) is based on computationally decomposing multispectral measurements with reference to a finite number of pure spectral components, i.e. endmembers (Figure 8). The method has

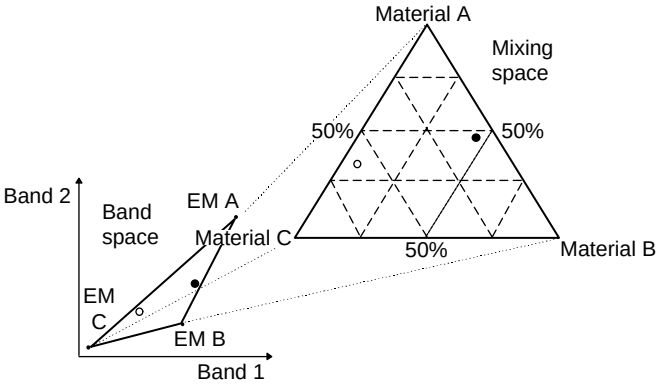

Figure 8 The concept of spectral unmixing, illustrated using an example of a sensor system with two spectral bands and three endmember materials.

become known as SMA,[94,95] and it assumes that most of the spectral variation in multispectral images is due to mixtures of a limited number of surface materials, and that these mixtures can, to a first approximation, be described as a result of additive (linear) spectral mixing (i.e. where each photon contacts only type of surface material). Either multiple scattering would need to be accounted for with nonlinear models or the variables of the linear mixing equation would need to be delinearized by using, for example, single-scattering albedo instead of reflectance.[96] The computation of proportional abundances can then always be solved with a system of linear equations. In matrix notation one can write Equation (1)

$$A \cdot X = B \tag{1}$$

where A is the $m \cdot n$ matrix of spectral endmember signatures (usually derived from so-called spectral libraries, i.e. a suitable collection of laboratory and field measurements), with each column containing one of the endmember spectral vectors. X denotes the $n \cdot 1$ unknown vector of abundances and B the $m \cdot 1$ observed data vector (i.e. measured reflectance of one pixel). The unknown vector of abundances is calculated with Equation (2)

$$X = A^{-1} \cdot B \tag{2}$$

i.e. essentially by inverting the endmember matrix. Evidently, a unique solution is possible as long as the number of spectral endmembers corresponds to the number of spectral bands. Also, if the problem is underdetermined, i.e. the number of unknown fraction components (i.e. abundances) exceeds the number of useful spectral bands by one, a solution can still be obtained by assuming that the set of endmembers is exhaustive (i.e. the sum of the computed endmember fractions is equal to one). This assumption, however, is not unproblematic since it is difficult to be sure that a sufficient number of spectral endmembers have been defined for a given data set. However, the typical case encountered in spectral unmixing (in particular when hyperspectral data are used) is that the number of bands is greater than the number of endmember spectra. The linear mixing model then becomes overdetermined such that the endmember matrix cannot be inverted. In this case a solution can be obtained through the pseudoinverse Equation (3)[97]

$$X = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} \cdot B + \varepsilon \tag{3}$$

which minimizes the mean-square error in fitting the abundance estimates to the data, and renders the computation of abundance estimates equivalent to a rotational transform of the image.[29] The added term ε represents the residual error of the model.

As the specification of endmembers is one of the major problems in spectral unmixing, it is an important feature of the method that the validity of mixing models can be analyzed by calculating error terms, such as the average root mean square (RMS) or spectral band residuals (i.e. ε). The average RMS error for the image provides a measure for how much of the spectral variability was explained by the selected endmembers, and the image of the RMS error for each pixel indicates such objects which could not be adequately modeled. In particular, the latter provides valuable indications for identifying which spectral components are not represented in the model. The spectral band residuals provide additional indicators of modeling errors. They are calculated by subtracting the modeled reflectance in each spectral band from the original sensor-measured reflectance. Positive residuals occur when the measured spectrum has higher reflectance at a specific wavelength than the modeled spectrum, indicating that the modeled spectrum contains absorption features which were lacking in the measured reflectance signature. Negative band residuals indicate the presence of absorption features in the measured spectrum which are absent or less pronounced in the modeled spectrum. Both the overall RMS error and the band residuals provide important diagnostics for handling uncertainties of the mixture model and for its optimization.

5 REMOTE SENSING OF SEMIARID LANDS: FROM RESOURCE ASSESSMENTS TO REGULAR MONITORING OF LAND DEGRADATION PROCESSES

Depending on the scientific perspectives, a variety of approaches and strategies have been proposed for assessing environmental characteristics of arid, semiarid and

subhumid ecosystems based on remote sensing.[98] The different perspectives expressed therein not only emphasize methodological preferences but, more fundamentally, reflect important paradigms of scientific disciplines. It has already been argued earlier why this application article concentrates on remote sensing approaches which are focused on the characterization of surface properties and not on climatological variables. The reason is not so much that general circulation models have not yet been able to suggest unequivocally whether global warming will result in a decrease or increase in rainfall in the subtropical belts. It is of even greater importance that they are not able to account for regional or local scale variations in surface properties (i.e. lithology, topography, soils and biological variables) which, in dry ecosystems, exert a strong influence on the environment and control water availability.[15]

After the International Convention on Desertification of the United Nations (i.e. a framework for programs on national, subnational and regional levels) entered into force, the need to measure land degradation processes increased. Whereas standard methods for undertaking such measurements are imperfect or expensive,[99] satellite-based remote sensing offers considerable potential. However, the relatively low spatial resolution provided by remote sensing satellites implies that it is impossible to detect small features such as rills or gullies, or to identify the mix of plant species or to detect species change over time. Instead, one needs to exploit the spectral contrast between soil and vegetation and their intermediate states to develop surrogate measurements of degradation. Operational techniques, however, have been implemented in only a few locations; some dedicated research projects are, for a considerable part, funded by the European Union (e.g. DeMon-I and DeMon-II; see also Lacaze et al.[79] and Hill et al.[100]), and are expected to provide guidelines for the implementation of more standardized approaches in the near future.

5.1 Vegetation Assessments and Monitoring in Dryland Systems

As already mentioned, one of the most important surface characteristics of arid regions to be assessed by remote sensing is vegetation cover. Appropriate management of arid lands requires information on cover classes and vegetation types on relatively bare lands, and the extent of changes due to management as opposed to changes from natural causes.[61] Arid lands are particularly suited to remote monitoring,[101] and Graetz,[102] for example, considers the use of satellites as the only way in which arid rangelands in Australia could be adequately monitored.

Although the differentiation between photosynthetic and dry vegetation is largely dependent on the available spectral range (it is obvious that sensor systems involving spectral bands in the SWIR region provide better data to separate green and dry vegetation components; this is, among others, one reason for which the new HRV system on board SPOT-4 has been equipped with one spectral channel around 1.65 μm), Pickup et al.[103] in Australia have already been able to produce reasonable estimates of vegetation cover involving green and dry biomass even with Landsat MSS data (which only provide spectral bands in the VIS and NIR part of the spectrum). A specific problem in using optical remote sensing systems for mapping and quantifying vegetation in arid lands is the question of what amount of green foliage is detectable with air- or space-borne instruments. Siegal and Goetz[93] had already demonstrated that significant changes in the reflectance characteristics of bare substrates need a vegetation cover of more than 10%, and that above 30% vegetation cover the vegetation-related signal component becomes more important than the soil reflectance. This by no means implies that vegetation cover below 30% is not detectable by remote sensing systems. Here we must emphasize again that ratio-based vegetation indices provide the worst estimates. Only vegetation indices which attempt to measure the abundance of green foliage with respect to a red versus NIR baseline (e.g. the ''perpendicular vegetation index'' proposed by Richardson and Wiegand[86]) tend to perform somewhat better, but appear still worse than conceptually different approaches developed under the spectral mixing paradigm.

SMA has been successfully used for a wide range of applications in geology, ecology, land use and land degradation studies and even coastal water mapping.[54,104] Owing to its direct response to preselected component spectra (i.e. typical desert shrubs and appropriate substrates) and increased robustness against soil color differences in the absence of dense vegetation, the method appears suited to overcome the conceptual limitations of two-band vegetation indices, such that we can expect improved estimates of vegetation abundance in areas with a sparse cover of mainly woody plants. Graetz and Gentle[105] and Pech et al.[106] have already used mixture models to estimate cover in arid Australian landscapes, and the technique has been successfully applied to Landsat TM images for mapping vegetation cover in semiarid parts of California (Figure 9). Although the scatter in Figure 9 is substantial, its orientation along the 1 : 1 relation between field-measured vegetation cover (in percent) and the vegetation fraction indicates an improvement in comparison with empirical relations between proportional cover and various vegetation indices. Also, not all of the variability in the estimated vegetation fraction should be attributed to the remote sensing estimates.

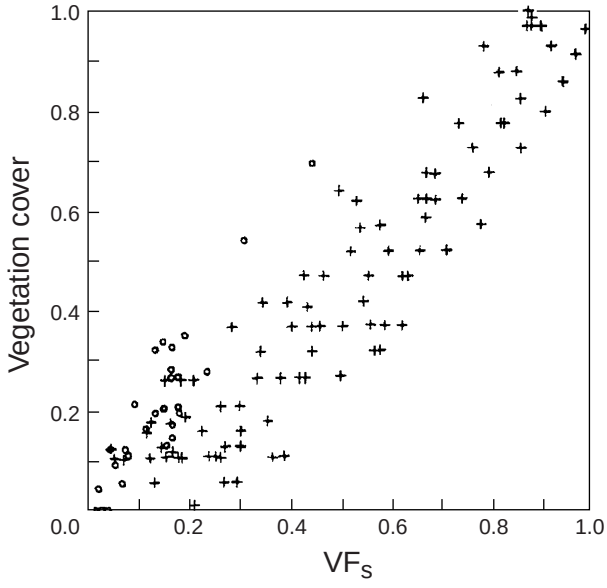

Figure 9 Comparison between vegetation cover measured in the field and calculated through the shade-normalized vegetation fraction (VF_S) derived from SMA. The symbols refer to Landsat TM observations of the bajadas (○) and the valley floor (+) of the semiarid Owens Valley, CA, USA. [Reproduced by permission from Smith et al.[107]]

Some of it reflects methodological difficulties in assessing precisely the proportional vegetation cover, a typical problem with field observations (for low shrub communities of the bajadas, a consistent overestimation of the projected area of green leaves and woody biomass in the field records has been reported). However, part of the scatter may still be due to soil color variations or other background materials (rocks, woody components or litter), or to the nonlinear scattering of light which was known to play an important role for dense plant canopies with high LAI (i.e. irrigated areas).[108] More recently, Ray and Murray[108] demonstrated that nonlinear mixing is also important in the observation of plants in arid environments.

Since it is not advisable to compensate for background-related errors in vegetation abundance estimates by simply adding additional endmember spectra to the mixing model, Hill et al.[53] proposed a strategy which aims at minimizing the number of endmember spectra while optimizing the selection of endmembers which describe the mixture of materials present in the background for each individual pixel. Spectral band residuals and, to a lesser extent, the RMS error of the model seem to provide valid criteria for controlling this selection. Their approach is based on the assumption that the mixed reflectance signature of a pixel is conditioned by three primary components which represent so-called ''foreground'' and ''background'' materials, and shade. While a green vegetation spectrum constantly represents the foreground material (which may provide partial or complete cover of the ground), the background component is allowed to vary between various soils, parent rock types, senescent vegetation or leaf litter.[54] This approach improved the vegetation abundance estimates for low coverage but also helped to identify the presence of dry plant matter, which inevitably provokes erroneous abundance estimates when using vegetation indices, no matter whether classical or improved indices are applied. It is believed that such developments in spectral unmixing hold some potential further to improve and standardize vegetation estimates in dryland systems. Moreover, it was demonstrated that spectral unmixing can also provide valuable information on soil composition when the spectral combination of components is conceptualized in the sense of intimate mixtures (as opposed to spatial-spectral mixing as determined by the sensor spatial response).

Besides the ongoing development of high spectral resolution airborne scanner systems, it has almost been neglected that aerial photographs with their extremely high spatial resolution (i.e. in the range of centimeters, depending on scale) can not only be used in the traditional way based on classical photointerpretation and/or photogrammetry approaches. The availability of high-quality scanning devices provides the possibility of generating multispectral images from aerial photographs which can then be analyzed by using classical image processing techniques. If spectral reference measurements are available it is even possible to calibrate color photographs to reflectance. As there are not many references available in the literature, the approach will be briefly illustrated by using an example from our own ongoing research on the disturbance and regeneration of biotic crusts in a sandy arid ecosystem of the Negev Desert (Israel). In this framework, digitized aerial photographs are used successfully to differentiate the proportional cover by photosynthetic and woody vegetation components (Figure 10). The methodological approach is based on SMA, which, after stratifying the aerial photographs into vegetated and nonvegetated spatial domains, provides the possibility of using reflectance spectra of specific components measured in the field.

5.2 Examples for the Assessment of Soil Characteristics in Dry Ecosystems

Not all studies of dryland ecosystems have relied solely on vegetation cover as an indicator of land suitability or degradation. Escadafal et al.[110] and Hill et al.,[54] for example, have been able to use parameters indicative of soil conditions and hence to apply concepts taken from pedology and geomorphology to obtain qualitative

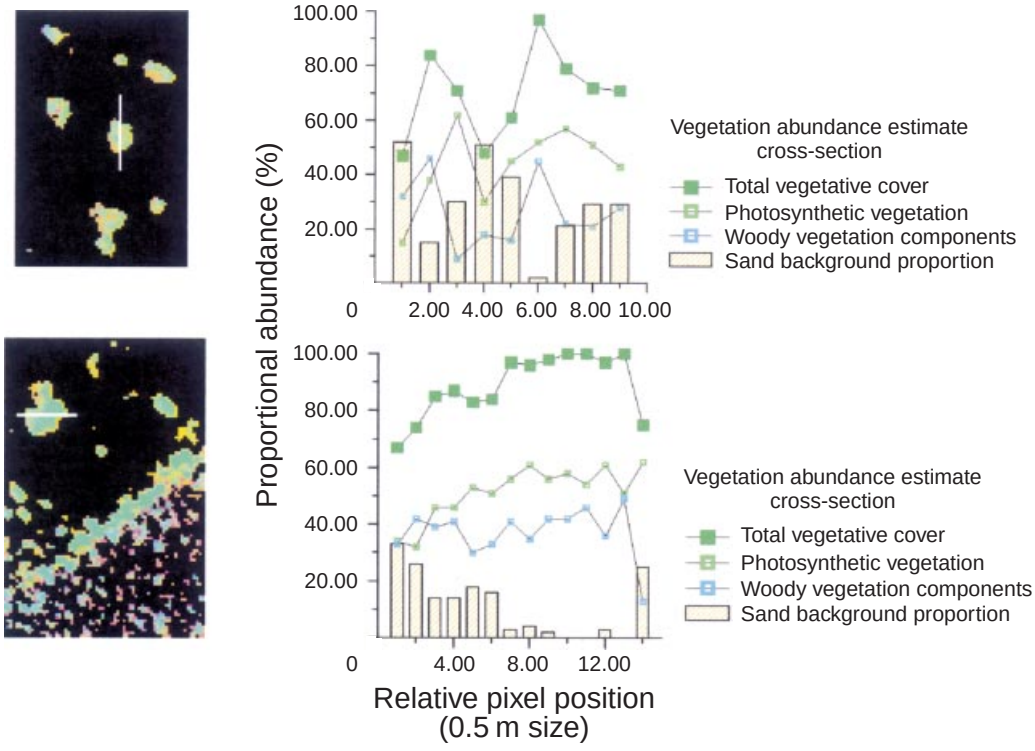

Figure 10 Cross-section through individual desert shrubs in a sandy arid ecosystem of the Negev Desert (Nizzana, Israel). The diagrams which are based on the spectral unmixing of true-color aerial photographs[48] show the proportional cover of photosynthetic and woody vegetation against the sandy background.

indications of soil degradation or erosion processes. SMA allowed the estimation of relative amounts of rock fragments and soil particles on the surface. Since, within a specific context, soil erosion leads to an increase in rock cover in source areas and the accumulation of soil material as colluvium elsewhere, rock content can be used as an indicator of degradation. This corresponds to defining the erosional state of soils as a function of the mixing ratio between developed soil substrates and parent material components which, of course, need to be spectrally distinct from each other. While Fischer[111] has been able to adopt these principles for mapping soils of different ages in an area of young glacial deposits, Hill et al.[53] successfully used this approach to map areas affected from soil degradation and erosion under sub-humid conditions in Mediterranean France.

They used a five-endmember mixing model which was applied to spectral field and laboratory measurements of soils, and later to multi- and hyperspectral images. According to taxonomic criteria, soil depth, and surface constituents, these soil spectra had been assigned to four soil condition levels (I, fair; II, degraded; III, severely eroded; and IV, parent rock). It was assumed that the surface reflectance characteristics of well-developed soils predominantly respond to the spectrum of a corresponding prototype spectrum, whereas those of degraded (i.e. eroded) or undeveloped soils produce increasing fractions of parent rock material (in this case limestone and marls) owing to the increased content of stone fragments and chemically unaltered material. Applying this approach to process real earth observation satellite images, of course, requires that the images were corrected for atmospheric effects (see section 4.1). Normalizing the estimated mixing proportions with regard to the shade and vegetation abundance allows one to retain only that part of the signal which relates to the soil reflectance. A soil condition map can then be obtained through a combination of cluster analysis and automatic classification of the corresponding images where user-defined thresholds (Figure 11) have to be observed. In both cases, the overall mapping accuracy for the above-mentioned four soil condition levels exceeded the 75% level.[53]

Similarly, based on digitized and calibrated high spatial resolution true-color aerial photographs, Hill et al.[48] could also demonstrate that spectral unmixing of reflectance signatures from the nonvegetated domain allowed the successful identification and spatial differentiation of substrates and biological crusts in a sandy arid ecosystems (Figure 12). Based on hydrological concepts, Pickup and Chewings[112] in Australia developed the erosion cell approach whereby the landscape is divided into

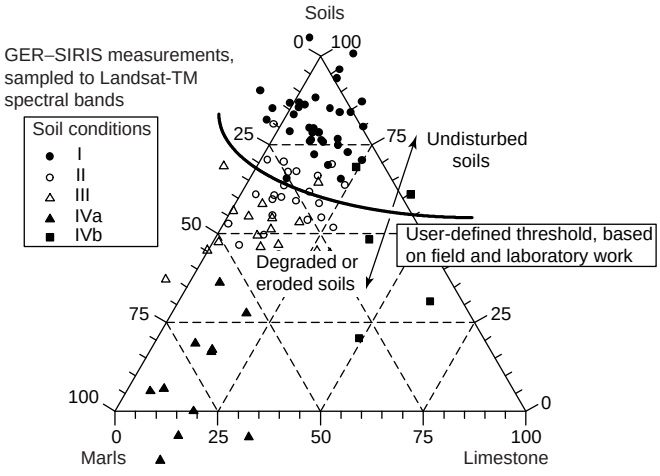

Figure 11 The separation between undisturbed soils and degraded/eroded substrates, displayed in a ternary diagram of endmember abundances resulting from the unmixing of reflectance spectra with a three endmember model (one soil and two parent material components). [Modified after Hill et al.[53]]

the production zone, where there is a net soil loss, a transfer zone with intermittent erosion and deposition and a sink where accumulation occurs.

Both approaches can be combined with quantitative estimates of soil properties such as organic carbon, thereby leading to increasingly differentiated possibilities to analyze environmental conditions in dryland ecosystems. Relating proxy data from a semiarid ecosystem in SE Spain (average rainfall 300–350 mm a^{-1}) to the erosion cell approach, Hill and Schütt,[73] for example, suggest that organic carbon in xeromorphic soils can be understood as a tracer substance which highlights areas of accumulation and relative stability (sediment sinks) where soil conditions are favorable because of higher infiltration and water retention capacity, better aggregation and increased nutrient availability.[113] Based on this premise, the objective of differentiating more favorable sink areas from active erosion and transport zones on the basis of optical remote sensing images merely condenses on a detection problem, which is to identify the organic matter content of dryland soils based on spectral indicators. Although the general relationship between organic matter and soil brightness and color have long been recognized, quantitative relationships are difficult to obtain because albedo is influenced by additional properties such as soil moisture, content of iron or mafic minerals. However, summarizing the results of a review paper by Schulze et al.,[114] one can state that such relationships can be developed within and even among soil landscapes if soil textures do not vary too much. Soil texture also controls whether relationships between soil organic matter (SOM) and soil color are linear (in silty and loamy soils) or curvilinear (sandy-textured soils). Consequently, the relationship between organic matter and soil color (here to be understood as a surrogate of spectral reflectance in the VIS region) may only then become unpredictable if the soil texture varies too widely.[114]

While more advanced procedures such as multiple-scattering models applicable to soil reflectance are still in

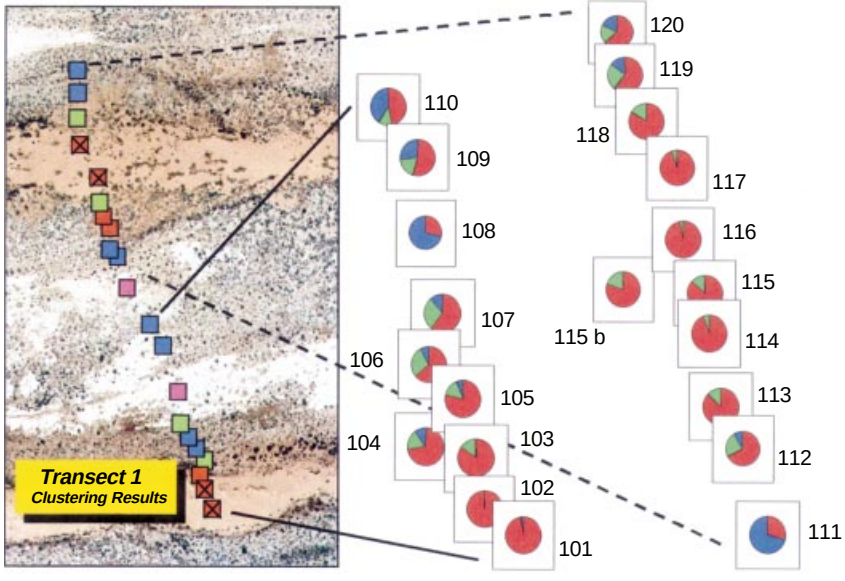

Figure 12 The sequence of diagrams on the right relates to estimates of the relative proportion of sand (dark gray), silt/clay (black) and biological crust material (light gray) in the surface material along a sampling transect at the Nizzana Research Site (Negev, Israel).

their infancy as they do not yet fully incorporate fundamental soil properties such as moisture content, organic matter content and structure,[62] empirical relations still play an important role. Some of the classical relations between organic carbon content and spectral variables were based on the reflectance in the VIS range. Traditionally one has tried to analyze the relationships to soil color and soil brightness in the VIS range of the solar spectrum. In continuation of this work, major efforts concentrated on correlating specific spectral bands with SOM. More recently, Ben-Dor and Banin[115] used the full reflective spectrum to predict SOM and other mineralogical and chemical soil properties based on multivariate statistics. Although they extended the approach to the Landsat TM bands,[116] no application to real imagery was demonstrated. As this missing link is common to many studies, Hill and Schütt[73] proposed a method which can be applied to high-resolution spectra in addition to the spectral resolution of the Landsat TM, and which can be applied to real images, provided that the data have been corrected for atmospheric effects (i.e. are available as primary parameter of the reflectance ρ). It is based on the fact that soil brightness is influenced by more parameters than organic carbon. Also, the specific influence of SOM on spectral reflectance is not expressed in narrow absorption bands, as is the case, for example, for iron oxides or carbonates, but it determines the overall shape of the reflectance curve in the VIS, NIR and mid-IR range of the spectrum (Figure 5). For this reason, they used the coefficients of a quadratic function (which was fitted to the original soil spectrum) to parameterize the most important characteristics of the original spectrum influenced by SOM, i.e. slope and curvature. Organic matter content (in weight percent) could then be estimated through a multiple linear regression between SOM and the coefficients of the parabolic curve fit; also, it was demonstrated that this model could be used to map successfully the organic carbon content in their study site based on atmospherically corrected Landsat TM images (Figure 13).

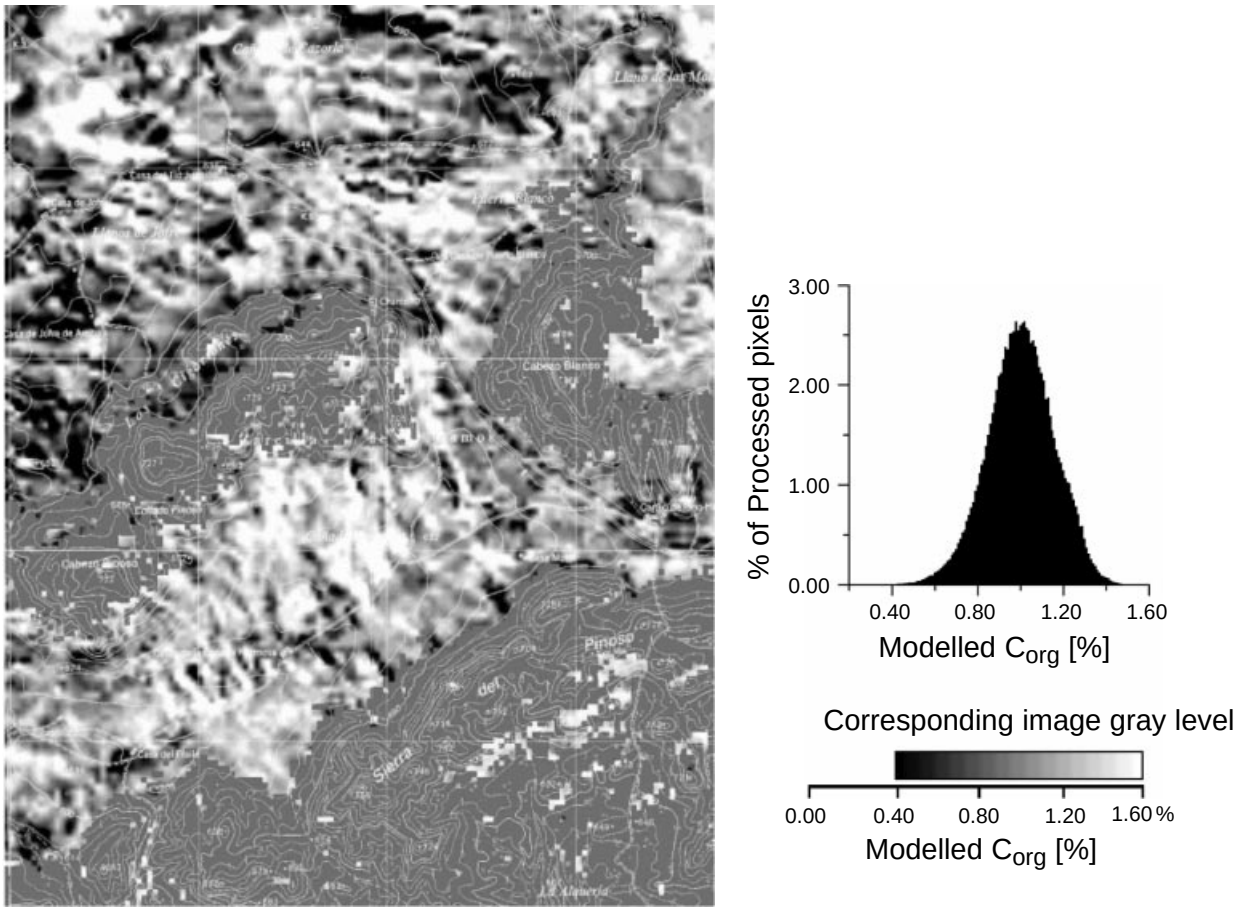

Figure 13 Landsat TM-based map of soil organic carbon in the semiarid Guadalentin area in SE Spain, based on the relation presented in Hill and Schütt.[73] In the image, gray levels correspond to SOM contents between 0.2 and 1.6%; areas of seminatural vegetation are masked. [Reproduced by permission from Hill and Schütt.[73]]

5.3 Monitoring Arid Ecosystems: the Desertification Issue

Resource assessments and continuous monitoring of environmental parameters are complementary issues to be observed for a sustainable management of dryland ecosystems. Therefore, an essential requirement for the applicability of remote sensing approaches is that suitable concepts and processing chains can be transferred and applied to locations with a wider range of climatic and lithologic conditions without changing too many of the processing parameters. Another important issue in monitoring is standardization, mainly because any improvements in the analysis may easily be incorporated into existing monitoring schemes if this implies only the reprocessing of already available primary parameter products or image-derived information layers.

Assessing and monitoring desertification by conventional means, in particular, have traditionally been lacking standardization because of the range of criteria and indicators.[17] The various data sources available through remote sensing offer the possibility of gaining environmental data over both large areas and relatively long time periods. In this context, much attention has been given to using the daily global observations provided by the AVHRR on board the NOAA weather satellites. While it was initially believed that the disadvantages of the low spatial resolution (the pixel dimension varies between 1 and approximately 5 km at off-nadir positions) would be compensated by the benefits of spatial extensiveness and high temporal resolution, it is now well understood that the compilation of calibrated time series involves highly demanding processing chains[117,118] and very careful consideration of calibration issues.[119,120] The result is that suitable time series are almost exclusively provided by huge research organizations [e.g. NASA, Joint Research Center (European Commission) (JRC), Food and Agriculture Organization (FAO)][121,122] and, if data are desired at 1 km resolution, do not cover very long periods of time. Additionally, it had meanwhile to be accepted that, owing to its limited spectral and spatial resolution, the information content of the AVHRR data is not very suited to the application of high-level quantitative analysis procedures, and the well-known NDVI product suffers from the shortcomings described in section 4.

NOAA data have nevertheless been used to compare seasonal and annual biomass fluctuations in the Sahel and to study the complex spatial and temporal patterns at the Sahara desert front.[123] Helldén[39] could show that previously developed concepts of an advancing desert front in the Sudan were incorrect, and that changes in biomass were related to the variability of rainfall, giving evidence of a fast recovery of vegetation in wet years. With regard to the available earth observation satellite data archives (regular Landsat MSS data acquisition started in 1972 and Landsat TM coverage in 1983), we wish particularly to emphasize the importance of retrospective studies based on earth observation satellite images from Landsat MSS, TM and SPOT, which may provide better spatial scales and spectral information for understanding the present situation, but also for devising approaches for regular monitoring on local to regional scales. Long-term studies for degraded sites in semiarid to subhumid Mediterranean regions covering a time span of more than 25 years with earth observation satellites were conducted, for example, by Hill et al.[100] Also their results revealed a largely differentiated evolution of mountainous ecosystems which in total had been considered as severely damaged by overgrazing. It seems obvious that the identification of degraded areas in the sense of an environmental inventory provides the fundamental basis for understanding better the processes of land degradation from their spatial context.

6 CONCLUDING REMARKS

Important characteristics of dryland ecosystems have been outlined to illustrate options for assessing and monitoring land resources and environmental change with remote sensing systems. Although it is unrealistic that remote sensing will replace traditional sources of data for inventory and monitoring, there is, without any doubt, an obvious role for it to play in assessing and monitoring the state of the environment in arid ecosystems. Unlike meteorological approaches which concentrate on the exchange of matter, momentum and radiation between the earth's surface and the atmosphere, the importance of defining indicators and thematic interpretation pathways for the characterization of actual land surface conditions was emphasized. We have shown that both inventory and mapping of dryland systems are required to define the current status of resources and provide a baseline for monitoring, and that surveys must be repeatable and comparable, and therefore demand a standardized methodological framework.

The advantages of remote sensing result from its synoptic nature, comprehensive spatial information and objective, repetitive coverage. While remote sensing has initially been used primarily for resource mapping and inventory, it turns out that monitoring and predictive modeling are becoming more important and successful. However, no matter whether one intends to map surface properties or the objective is to estimate fluxes based on remote sensing, data require that the primary parameters, such as the spectral surface reflectance or surface temperature, are first retrieved with adequate

precision. This raises immediately the issue of adequate radiometric corrections. While atmospheric effects have been a significant factor in the failure of scene models (in particular for monitoring concepts), much progress has been achieved in the past years, and we are now in the position that large parts of the radiometric preprocessing are considered a routine operation, similar to geometric rectification. At present, the remaining problems in retrieving surface reflectance from satellite data appear related more to the absolute radiance calibration of the sensor systems than to methodological drawbacks. This ensures that more advanced scene models, such as the spectral mixing paradigm or invertible physically based analytical models, can be used to derive quantitative estimates and improved indicators for land resources and degradation processes.

Although several examples have been presented which appear suited for a standardized data analysis within operational schemes, additional efforts are required, in particular for the definition of suitable sampling schemes (i.e. the selection of monitoring sites which are representative of large eco-regions), for extending the thematic concepts for the interpretation of remotely sensed primary parameters as a function of ecosystem characteristics, for the incorporation of ancillary information (climatic records, lithology, topographic information), and for improving specific processing modules with respect to the physiographic variability of arid ecosystems. Finally, we wish to emphasize the importance of a retrospective analysis of earth observation satellite images which have been available since the 1970s. The integrated interpretation of the satellite-derived information layers, available climatic records and results from detailed field studies may in fact provide a new perspective to understanding environmental change in arid ecosystems.

The growth of Geographic Information Systems (GIS) has already greatly enhanced the integration of remote sensing with other types of data, opening up more and more pathways for the application of spatialized simulation models (concerning biomass production, runoff generation, soil erosion, etc.) which relate dynamic landscape processes to changes in spectral–spatial characteristics captured by remote sensing. Additionally, the satellite-based Global Positioning System provides high-precision geometric reference data for areas which were traditionally badly documented by maps. This is an additional element in the success story of GIS which are increasingly used to complement and improve the use of remote sensing data. The integration of remote sensing and GIS is expected to contribute substantially to assessing land capability for sustained production, to facilitate the identification of environmental indicators and interactions and to improve predictive capabilities. It thus forms the basis for drafting and implementing efficient land management plans which are needed to avoid land degradation under inadequate management practices.

Also, we are facing intriguing perspectives with regard to suitable sensor systems available in the future. Landsat 7, successfully launched in 1999, carries the Enhanced Thematic Mapper (ETM) with seven spectral bands in the VIS and IR portion of the electromagnetic spectrum, thermal IR capabilities and a panchromatic band with 15 m spatial resolution. The SPOT earth observation satellites, together with the series of Indian remote sensing satellites, provide a strong backup for operational applications, and there are indications that very high spatial resolution satellite systems, such as QuickBird and Ikonos (commercial 1-m panchromatic resolution satellites with 4–5 m resolved multispectral data from the VIS to NIR range) will become available in the future. Fusing data from high spectral and spatial resolution systems may significantly increase our observational capabilities. Besides the ongoing development of multi- and hyperspectral airborne scanner systems which, with the commercially available HyMap™ system developed by Integrated Spectronics Pty Ltd (Sydney, Australia), is approaching a phase of true operationality, it has almost been neglected that aerial photographs with their extremely high spatial resolution (i.e. in the range of centimeters, depending on scale) can not only be used in the traditional way based on classical photointerpretation and/or photogrammetry approaches. The availability of high-quality scanning devices provides the possibility of generating multispectral images from aerial photographs which can then be analyzed by using classical image processing techniques. If spectral reference measurements are available it is even possible to calibrate color photographs to reflectance and apply high-level image processing techniques, thereby providing a low-cost alternative.

Moreover, polar platforms will be launched by the USA and the ESA where advanced sensor systems for environmental monitoring will be placed, including hyperspectral imaging systems. It is also believed that new forms of synergism between low spatial resolution satellites with more frequent coverage (e.g. the NOAA AVHRR, the VEGETATION instrument on board SPOT-4, MERIS, etc.) and high spatial and spectral resolution systems should be devised; active microwave systems [e.g. European remote sensing satellite (ERS) 1 and 2 (of ESA), RADARSAT] are also expected to become more and more incorporated.

ABBREVIATIONS AND ACRONYMS

AVHRR	Advanced Very High Resolution Radiometer

AVIRIS	Airborne Visible and Infrared Imaging Spectrometer
DAIS	Digital Airborne Imaging Spectrometer
DN	Digital Number
ENVISAT	Environmental Research Satellite
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ETM	Enhanced Thematic Mapper
FAO	Food and Agriculture Organization
GEMI	Global Environment Monitoring Index
GIS	Geographic Information Systems
HRV	High-resolution Visible Instrument
IR	Infrared
JRC	Joint Research Center (European Commission)
LAI	Leaf Area Index
Landsat	Land Observing Satellite
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectrometer
MSS	Multispectral Scanner System
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NOAA	National Oceanographic and Atmospheric Administration
P/PET	Ratio between Total Annual Precipitation (P) and Potential Average Evapotranspiration (PET)
RMS	Root Mean Square
SMA	Spectral Mixture Analysis
SOM	Soil Organic Matter
SPOT	Système pour l'Observation de la Terre
SR	Simple Ratio
SWIR	Short-wave Infrared
TM	Thematic Mapper
VEGETATION	Vegetation Monitoring System
VIS	Visible

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Global Land Databases for Environmental Analyses • Hyperspectral Remote Sensing: Data Collection and Exploitation • Imaging Spectrometry for Geological Applications • Land Cover Assessment and Monitoring • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring

Infrared Spectroscopy **(Volume 12)**
Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Spectral Databases, Infrared

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. M. Mainguet, *Aridity, Droughts and Human Development*, Springer-Verlag, Berlin, 1999.
2. UNESCO, 'Map of the World Distribution of Arid Regions', *MAB Technical Notes*, No. 7, UNESCO, Paris, 1977.
3. M. Mainguet, *Desertification, Natural Background and Human Mismanagement*, 2nd edition, Springer-Verlag, Berlin, 1994.
4. D.W. Goodall, R.A. Perry (eds.), *Arid-land Ecosystems: Structure, Functioning and Management*, Cambridge University Press, Cambridge, 2 vols., IBP 16 and 17, 1979/81.
5. H.E. Dregne, *Soils of Arid Regions*, Elsevier Scientific, Amsterdam, 1976.
6. P. Beaumont, *Drylands, Environmental Management and Development*, London, 1989.
7. H. Walter, S.W. Breckle, *Ökologie der Erde, Bd. 1.: Ökologische Grundlagen in globaler Sicht*, Gustav Fischer Verlag, Stuttgart, 1991.
8. R.D. Graetz, 'Remote Sensing of Terrestrial Ecosystem Structure: An Ecologist's Pragmatic View', in *Remote Sensing of Biosphere Functioning*, eds. R.J. Hobbs, H.A. Mooney, Springer-Verlag, New York, 5–30, 1990.
9. N.E. West, 'Structure and Function of Microphytic Crusts in Wildland Ecosystems of Arid to Semi-arid Regions', *Adv. Ecol. Res.*, **20**, 179–223 (1990).
10. J. Belnap, J.S. Gardner, 'Soil Microstructure in Soils of the Colorado Plateau: The Role of the Cyanobacterium *Microcoleus Vaginatus*', *Great Basin Nat.*, **53**, 40–47 (1993).
11. J. Belnap, D.A. Gillette, 'Vulnerability of Desert Biological Crusts to Wind Erosion: The Influences of Crust Development, Soil Texture, and Disturbance', *J. Arid Environ.*, **39**, 133–142 (1998).
12. O.L. Lange, G.J. Kidron, B. Büdel, A. Meyer, E. Kilian, A. Abeliovitch, 'Taxonomic Composition and Photosynthetic Characteristics of the ''Biological Soil Crusts'' Covering Sand Dunes in the Western Negev Desert', *Funct. Ecol.*, **6**, 519–527 (1992).

13. A. Yair, 'Runoff Generation in a Sandy Area. The Nizzana Sands, Western Negev, Israel', *Earth Surf. Processes Landforms*, **15**, 597–609 (1990).
14. A. Danin, Y. Bar-Or, I. Dor, T. Yisraeli, 'The Role of Cyanobacteria in Stabilisation of Sand Dunes in Southern Israel', *Ecologia Mediterranea*, **15**(1/2), 55–64 (1989).
15. A. Yair, 'The Ambiguous Impact of Climate Change at the Desert Fringe: Northern Negev, Israel', in *Environmental Change in Drylands: Biogeographical and Geomorphological Perspectives*, eds. A.C. Millington, K. Pye, John Wiley & Sons, Chichester, 199–227, 1994.
16. J. Schultz, *Die Ökozonen der Erde*, 2nd edition, Verlag Eugen Ulmer, Stuttgart, 1995.
17. D.S.G. Thomas, N.J. Middleton, *Desertification. Exploding the Myth*, John Wiley & Sons, Chichester, 1994.
18. G.J.A. Nieuwenhuis, E.H. Smidt, H.A.M. Thunissen, 'Estimation of Regional Evapotranspiration of Arable Crops from Thermal Infrared Images', *Int. J. Remote Sensing*, **6**, 1319–1334 (1985).
19. M. Menenti, 'Understanding Land Surface Evapotranspiration with Satellite Multispectral Measurements', *Adv. Space Res.*, **13**, 89–100 (1993).
20. W.G. Bastiaanssen, H. Pelgrum, T. van der Wal, R.A. Roebeling, 'Aggregation of Evaporative Fraction by Remote Sensing from Micro to Macro Scale', in *The Use of Remote Sensing for Land Degradation and Desertification Monitoring in the Mediterranean Basin. State of the Art and Future Research*, eds. J. Hill, D. Peter, EUR 16732 EN, Office for Official Publications of the European Communities, Luxembourg, 115–132, 1996.
21. G.J. Roerink, W.G.M. Bastiaanssen, J. Chambouleyron, M. Menenti, 'Relating Crop Water Consumption to Irrigation Water Supply by Remote Sensing', *Water Resources Manage.*, **11**, 445–465 (1997).
22. R.N. Colwell (ed.), *Manual of Remote Sensing*, American Society for Photogrammetry and Remote Sensing, Falls Church, VA, Vol. I, 1983.
23. P.N. Slater, *Remote Sensing – Optics and Optical Systems*, Addison-Wesley, Reading, MA, 1980.
24. C. Elachi, *Introduction to the Physics and Techniques of Remote Sensing*, John Wiley & Sons, New York, 1987.
25. P.H. Swain, S.M. Davis (eds.), *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York, 1978.
26. T.M. Lillesand, R.W. Kiefer, *Remote Sensing and Image Interpretation*, John Wiley & Sons, New York, 1987.
27. J.A. Richards, *Remote Sensing Digital Image Analysis – An Introduction*, Springer-Verlag, Berlin, 1993.
28. J.R. Jensen, *Introductory Digital Image Processing – A Remote Sensing Perspective*, Prentice Hall, Upper Saddle River, NJ, 1996.
29. R.A. Schowengerdt, *Remote Sensing. Models and Methods for Image Processing*, Academic Press, San Diego, CA, 1997.
30. B. Siegal, A. Gillespie, *Remote Sensing in Geology*, John Wiley & Sons, New York, 1980.
31. C. Elachi, *Spaceborne Radar Remote Sensing: Applications and Techniques*, IEEE Press, New York, 1988.
32. G. Asrar (ed.), *Theory and Applications of Optical Remote Sensing*, Wiley-Interscience, New York, 1989.
33. J.C. Curlander, R.N. McDonnough, *Synthetic Aperture Radar – Systems and Signal Processing*, John Wiley & Sons, New York, 1991.
34. J. Hill, J. Mégier (eds.), *Imaging Spectrometry – A Tool for Environmental Observations*, Kluwer Academic Publishers, Dordrecht, 1994.
35. H.J. Kramer, *Observation of the Earth and its Environment: Survey of Missions and Sensors*, Springer-Verlag, Berlin, 1996.
36. J.R. Schott, *Remote Sensing: The Image Chain Approach*, Oxford University Press, New York, 1996.
37. F. Perez-Trejo, *Desertification and Land Degradation in the European Mediterranean*, EUR 14850 EN, Office for Official Publications of the European Communities, Luxembourg, 1994.
38. R.D. Graetz, 'Empirical and Practical Approaches to Land Surface Characterisation and Change Detection', in *The Use of Remote Sensing for Land Degradation and Desertification Monitoring in the Mediterranean Basin. State of the Art and Future Research*, eds. J. Hill, D. Peter, EUR 16732 EN, Office for Official Publications of the European Communities, Luxembourg, 9–21, 1994.
39. U. Helldén, 'Desertification – Time for an Assessment', *Ambio*, **20**(8), 372–383 (1991).
40. S.Q. Kidder, T.H. Vonder Haar, *Satellite Meteorology, an Introduction*, Academic Press, San Diego, CA, 1995.
41. A.F.H. Goetz, G. Vane, J.E. Solomon, B.N. Rock, 'Imaging Spectrometry for Earth Remote Sensing', *Science*, **228**, 1147–1153 (1985).
42. T. Cocks, R. Jenssen, A. Stewart, I. Wilson, T. Shields, 'The HyMap™ Airborne Hyperspectral Sensor: The System, Calibration and Performance', in *Proceedings of the 1st EARSeL Workshop on Imaging Spectrometry, RSL, University of Zurich, Switzerland, 6–8 October 1998*, eds. M. Schaepman, D. Schläpfer, K. Itten, EARSeL, Paris, 37–42, 1998.
43. G.G. Schaber, J.F. McCauley, C.S. Breed, 'The Use of Multifrequency and Polarimetric SIR-C/X-SAR Data in Geologic Studies of Bir Safsaf, Egypt', *Remote Sensing Environ.*, **59**, 337–363 (1997).
44. A.B. Kahle, 'Surface Emittance, Temperature and Thermal Inertia Derived from Thermal Infrared Multispectral Scanner (TIMS) data for Death Valley, California', *Geophysics*, **52**(7), 858–874 (1987).
45. W.P. Kustas, R.D. Jackson, G. Asrar, 'Estimating Surface Energy-balance Components from Remotely Sensed Data', in *Theory and Applications of Optical Remote Sensing*, ed. G. Asrar, Wiley-Interscience, New York, 604–627, 1989.
46. J.C. Price, 'Land Surface Temperature Measurements from the Split Window Channels of the NOAA 7

Advanced Very High Resolution Radiometer', *J. Geophys. Res.*, **89**, 7231–7237 (1984).
47. C. Coll, V. Caselles, J.A. Sobrino, E. Valor, 'On the Atmospheric Dependence of the Split-window Equation for Land Surface Temperature', *Int. J. Remote Sensing*, **15**, 105–122 (1994).
48. J. Hill, Th. Udelhoven, B. Schütt, A. Yair, 'Differentiating Biological Soil Crusts in a Sandy Arid Ecosystem Based on Hyperspectral Data Acquired with DAIS-7915', in *Proceedings of the 1st EARSeL Workshop on Imaging Spectrometry, RSL, University of Zurich, Switzerland, 6–8 October 1998*, eds. M. Schaepman, D. Schläpfer, K. Itten, EARSeL, Paris, 427–436, 1998.
49. J.E. Conel, R.O. Green, R.E. Alley, C.J. Bruegge, V. Carrere, J.S. Margolis, G. Vane, T.G. Chrien, P.N. Slater, S.F. Biggar, P.M. Teillet, R.D. Jackson, M.S. Moran, 'In-flight Radiometric Calibration of the Airborne Visible/Near-infrared Spectrometer (AVIRIS)', *Proc. SPIE*, **924**, 179–195 (1988).
50. D. Tanré, C. Deroo, P. Duhaut, M. Herman, J.J. Morcrette, J. Perbos, P.Y. Deschamps, 'Description of a Computer Code to Simulate the Signal in the Solar Spectrum: The 5S Code', *Int. J. Remote Sensing*, **11**(4), 659–668 (1990).
51. J. Hill, B. Sturm, 'Radiometric Correction of Multitemporal Thematic Mapper Data for Use in Agricultural Land Cover Classification and Vegetation Monitoring', *Int. J. Remote Sensing*, **12**, 1471–1491 (1991).
52. J. Hill, *High Precision Land Cover Mapping and Inventory with Multi-temporal Earth Observation Data. The Ardeche Experiment*, EUR 15271, Office for Official Publications of the European Communities, Luxembourg, 1993.
53. J. Hill, J. Mégier, W. Mehl, 'Land Degradation, Soil Erosion and Desertification Monitoring in Mediterranean Ecosystems', *Remote Sensing Rev.*, **12**, 107–130 (1994).
54. J. Hill, W. Mehl, V. Radeloff, 'Improved Forest Mapping by Combining Corrections of Atmospheric and Topographic Effects, in *Sensors and Environmental Applications of Remote Sensing, Proceedings 14th EARSeL Symposium, Göteborg, Sweden, 6–8 June 1994*, eds. J. Askne, A.A. Balkema, Rotterdam, 143–151, 1995.
55. R.G. Holm, M.S. Moran, R.D. Jackson, P.N. Slater, B. Yuan, S.F. Biggar, 'Surface Reflectance Factor Retrieval from Thematic Mapper Data', *Remote Sensing Environ.*, **27**, 47–57 (1989).
56. R.C. Wrigley, M.A. Spanner, R.E. Slye, R.F. Pueschel, H.R. Aggarwal, 'Atmospheric Correction of Remotely Sensed Image Data by a Simplified Model', *J. Geophys. Res.*, **97**, 18797–18814 (1992).
57. B.L. Markham, R.N. Halthore, S.J. Goetz, 'Surface Reflectance Retrieval from Satellite and Aircraft Sensors: Results of Sensor and Algorithm Comparisons During FIFE', *J. Geophys. Res.*, **97**, 18785–18795 (1992).
58. P.N. Slater, S.F. Biggar, R.G. Holm, R.D. Jackson, Y. Mao, M.S. Moran, M. Palmer, B. Yuan, 'Reflectance- and Radiance-based Methods for the In-flight Absolute Calibration of Multispectral Sensors', *Remote Sensing Environ.*, **22**, 11–37 (1987).
59. K.J. Thome, D.I. Gellmann, R.J. Parada, S.F. Biggar, P.N. Slater, M.S. Moran, 'In-flight Radiometric Calibration of Landsat-5 Thematic Mapper from 1984 to Present', in *Recent Advances in Sensors, Radiometric Calibration, and Processing of Remotely Sensed Data*, eds. P.S. Charez, R.A. Schowengerdt, SPIE – International Society for Optical Engineering, Bellingham, WA, 126–130, 1993.
60. V. Caselles, M.J. López, 'An Alternative Simple Approach to Estimate Atmospheric Correction in Multitemporal Studies', *Int. J. Remote Sensing*, **10**, 1127–1134 (1989).
61. R.M. Johnston, M.M. Barson, 'An Assessment of the Use of Remote Sensing Techniques in Land Degradation Studies', *Bureau Rural Resources Bull.*, No. 5, Australian Government Printing Services, Canberra, 1990.
62. J.R. Irons, R. Weismiller, G.W. Petersen, 'Soil Reflectance', in *Theory and Applications of Optical Remote Sensing*, ed. G. Asrar, Wiley-Interscience, New York, 66–106, 1989.
63. H.R. Condit, 'The Spectral Reflectance of American Soils', *Photogrammetric Eng.*, **36**, 955–966 (1970).
64. M.F. Baumgardner, E.R. Stoner, L.F. Silva, L.L. Biehl, 'Reflective Properties of Soils', *Adv. Agron.*, **38**, 1–44 (1985).
65. M.A. Mulders, *Remote Sensing in Soil Science*, Elsevier, Amsterdam, 1987.
66. R. Geerken, 'Informationspotential von Spektral Hochauflösenden Fernerkundungsdaten für die Identifizierung von Mineralen und Gesteinen. Laborversuche und Anwendungsbeispiele in der Geologie', PhD Dissertation, University of Karlsruhe, 1991.
67. J. Madeira, 'Etude Quantitative des Relations Constituants Minéralogiques–Réflectance Diffuse des Latosols Brésiliens. Application à l'Utilisation Pédologique des Données Satellitaires TM (Région de Brasilia)', Thèse de Doctorat, Université Pierre et Marie Curie, Paris, 1991.
68. R. Escadafal, M.C. Girard, D. Courault, 'Munsell Soil Color and Soil Reflectance in the Visible Spectral Bands of Landsat Data (MSS and TM)', *Remote Sensing Environ.*, **27**, 37–46 (1989).
69. R. Escadafal, 'Remote Sensing of Soil Color: Principles and Applications', *Remote Sensing Rev.*, **7**, 261–279 (1993).
70. P.W. Birkeland, 'Soil-geomorphic Research – A Selective Overview', *Geomorphology*, **3**, 207–224 (1990).
71. R.A. Weismiller, G.E. Van Scoyoc, S.E. Pazar, K. Latz, M.F. Baumgardner, 'Use of Soil Spectral Properties for Monitoring Soil Erosion', in *Soil Erosion and Conservation*, eds. S.A. El-Swaify, W.C. Moldenhauer, A. Lo, Soil Conservation Society of America, Ankeny, IA, 1984.

72. G. Pickup, D.J. Nelson, 'Use of Landsat Radiance Parameters to Distinguish Soil Erosion, Stability, and Deposition in Arid Central Australia', *Remote Sensing Environ.*, **16**, 195–209 (1984).
73. J. Hill, B. Schütt, 'Die Verwendung von Fernerkundungsdaten zur Erfassung komplexer Gefüge von Erosions- und Stabilitätsbereichen in Mediterranen Trockengebieten', *Trierer Geogr. Studien*, **16**, 185–201 (1997).
74. G. Pickup, 'The Erosion Cell – A Geomorphic Approach to Landscape Classification in Range Assessment', *Aust. Rangelands J.*, **7**, 114–121 (1985).
75. C.J. Tucker, 'Red and Photographic Infrared Linear Combinations for Monitoring Vegetation', *Remote Sensing Environ.*, **8**, 127–150 (1979).
76. B. Pinty, M.M. Verstraete, 'GEMI: A Non-linear Index to Monitor Global Vegetation from Satellites', *Vegetatio*, **101**, 15–20 (1992).
77. C.R. Perry, L.F. Lautenschlager, 'Functional Equivalence of Spectral Vegetation Indices', *Remote Sensing Environ.*, **14**, 169–182 (1984).
78. J.B. Adams, M.O. Smith, 'Changes in Vegetated Landscapes: A Process-model Approach to Multispectral Remote Sensing', in *The Use of Remote Sensing for Land Degradation and Desertification Monitoring in the Mediterranean Basin. State of the Art and Future Research*, eds. J. Hill, D. Peter, EUR 16732 EN, Office for Official Publications of the European Communities, Luxembourg, 169–180, 1996.
79. B. Lacaze, 'Spectral Characterisation of Vegetation Communities and Practical Approaches to Vegetation Cover Changes Monitoring', in *The Use of Remote Sensing for Land Degradation and Desertification Monitoring in the Mediterranean Basin. State of the Art and Future Research*, eds. J. Hill, D. Peter, EUR 16732 EN, Office for Official Publications of the European Communities, Luxembourg, 149–166, 1996.
80. P. Warnant, L. Francois, D. Strivay, F. Robinet, 'Forcing of a Global Model of Plant Productivity with Climatic and Remote Sensing Data, in *Vegetation, Modeling and Climatic Change Effects*, eds. F. Veroustraete, R.J.M. Ceulemans, I.I.P. Impens, J.B.H.F. Van Rensbergen, SPB Academic Publishing, The Hague, 179–186, 1994.
81. J.M.C. Pereira, T.M. Oliveira, J.P.C. Paúl, 'Fuel Mapping in a Mediterranean Shrubland Using Landsat TM Imagery', in *Proceedings of International Workshop on Satellite Technology and GIS for Mediterranean Forest Mapping and Fire Management*, eds. P.J. Kennedy, M. Karteris, Office for Official Publications of the European Communities, Luxembourg, 97–106, 1994.
82. M.F. Jasinski, 'Sensitivity of the Normalized Difference Vegetation Index to Subpixel Canopy Cover, Soil Albedo, and Pixel Scale', *Remote Sensing Environ.*, **32**, 169–187 (1990).
83. S.D. Prince, 'Satellite Remote Sensing of Primary Production: Comparison of Results from Sahelian Grasslands', *Int. J. Remote Sensing*, **12**, 1301–1312 (1991).
84. A.R. Huete, R.D. Jackson, D.F. Post, 'Spectral Response of a Plant Canopy with Different Soil Backgrounds', *Remote Sensing Environ.*, **17**, 37–53 (1987).
85. M.B. Satterwhite, P.J. Henley, 'Spectral Characteristics of Selected Soils and Vegetation in Northern Nevada and Their Discrimination Using Band Ratioing Techniques', *Remote Sensing Environ.*, **23**, 155–175 (1987).
86. A.J. Richardson, C.L. Wiegand, 'Distinguishing Vegetation from Soil Background Vegetation', *Photogrammetric Eng. Remote Sensing*, **43**, 1541–1552 (1977).
87. J.P.G.W. Clevers, 'The Application of a Weighted Infrared Vegetation Index for Estimating Leaf Area Index by Correcting for Soil Moisture', *Remote Sensing Environ.*, **29**, 25–37 (1989).
88. F. Baret, G. Guyot, D. Major, 'TSAVI: A Vegetation Index which Minimises Soil Brightness Effects on LAI and APAR Estimation', in *Proceedings of the IGARSS '89*, Vancouver, Canada, July 10–14, 1355–1358, 1989.
89. R. Geerken, H. Kaufmann, 'Spectral Effects of Surface Characters and Rock Accessories. A Critical Contribution to the Applicability of High Spectral Resolution Sensor Data', in *Proceedings of the 9th EARSeL Symposium, 27 June–1 July 1989, Espoo, Finland*, EUR 12827 EN, Office for Official Publications of the European Communities, Luxembourg, 501–511, 1990.
90. J.B. Adams, F. Palmer, J.T. Staley, 'Rock Weathering in Deserts: Mobilization and Concentration of Ferric Iron by Microorganisms', *Geomicrobiol. J.*, **10**, 99–114 (1992).
91. A. Danin, J. Garty, 'Distribution of Cyanobacteria and Lichens on Hillsides of the Negev Highlands and Their Impact on Biogenic Weathering', *Z. Geomorphol., N.F.*, **27**(4), 423–444 (1983).
92. A. Karnieli, H. Tsoar, 'Spectral Reflectance of Biogenic Crust Developed on Desert Dune Sand Along the Israel–Egypt Border', *Int. J. Remote Sensing*, **16**(2), 369–374 (1995).
93. B.S. Siegal, A.F.H. Goetz, 'Effect of Vegetation on Rock and Soil Type Discrimination', *Photogrammetric Eng. Remote Sensing*, **43**, 191–196 (1977).
94. J.B. Adams, M.O. Smith, A.R. Gillespie, 'Simple Models for Complex Natural Surfaces: A Strategy for the Hyperspectral Era of Remote Sensing', in *Proceedings of the IGARSS '89 Symposium*, Vancouver, Canada, July 10–14, 16–21, 1989.
95. M.D. Craig, 'Minimum-volume Transforms for Remotely Sensed Data', *IEEE Trans. Geosci. Remote Sensing*, **32**, 542–552 (1994).
96. P.E. Johnson, M.O. Smith, S. Taylor-George, J.B. Adams, 'A Semi-empirical Method for Analysis of the Reflectance Spectra for Binary Mineral Mixtures', *J. Geophys. Res.*, **88**, 3557–3561 (1983).

97. G.H. Golub, C.F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1983.
98. J. Hill, D. Peter (eds.), *The Use of Remote Sensing for Land Degradation and Desertification Monitoring in the Mediterranean Basin. State of the Art and Future Research*, EUR 16732 EN, Office for Official Publications of the European Communities, Luxembourg, 1996.
99. G. Pickup, 'New Land Degradation Survey Techniques for Arid Australia – Problems and Prospects', *Aust. Rangeland J.*, **11**, 74–82 (1989).
100. J. Hill, P. Hostert, G. Tsiourlis, P. Kasapidis, Th. Udelhoven, 'Monitoring 20 Years of Intense Grazing Impact on the Greek Island of Crete with Earth Observation Satellites', *J. Arid Environ.*, **39**, 165–178 (1998).
101. P.T. Tueller, 'Remote Sensing Science Application in Arid Environments', *Remote Sensing Environ.*, **23**, 143–154 (1987).
102. R.D. Graetz, 'Satellite Remote Sensing of Australian Rangelands', *Remote Sensing Environ.*, **23**, 313–331 (1987).
103. G. Pickup, V.H. Chewings, D.J. Nelson, 'Estimating Changes in Vegetation Cover Over Time in Arid Rangelands Using Landsat MSS Data', *Remote Sensing Environ.*, **43**, 243–263 (1993).
104. P.N. Bierwirth, T.J. Lee, R.V. Burne, 'Shallow Seafloor Reflectance and Water Depth by Unmixing Multispectral Imagery', *Photogr. Eng. Remote Sensing*, **59**(3), 331–338 (1993).
105. R.D. Graetz, M.R. Gentle, 'The Relationships Between Reflectance in the Landsat Wavebands and the Composition of an Australian Semi-arid Shrub Rangeland', *Photogrammetric Eng. Remote Sensing*, **48**, 1721–1730 (1982).
106. R.P. Pech, A.W. Davis, R.D. Graetz, 'Reflectance Modelling and the Derivation of Vegetation Indices for an Australian Semi-arid Shrubland', *Int. J. Remote Sensing*, **7**, 389 (1986).
107. M.O. Smith, S.L. Ustin, J.B. Adams, A.R. Gillespie, 'Vegetation in Deserts I: A Regional Measure of Abundance from Multispectral Images', *Remote Sensing Environ.*, **31**, 1–26 (1990).
108. D.A. Roberts, M.O. Smith, J.B. Adams, 'Green Vegetation, Nonphotosynthetic Vegetation and Soils in AVIRIS Data', *Remote Sensing Environ.*, **44**, 255–269 (1993).
109. T.W. Ray, B.C. Murray, 'Nonlinear Spectral Mixing in Desert Mixing', *Remote Sensing Environ.*, **55**, 59–64 (1996).
110. R. Escadafal, A. Belgith, A. Ben Moussa, 'Indices Spectraux Pour la Télédétéction de la Dégradation des Milieux Naturels en Tunisie Aride', in *Proceedings of the 6th International Symposium on Physical Measurements and Signatures in Remote Sensing, 17–24 January 1994, Val d'Isère, France*, CNES, Paris, 253–259, 1994.
111. A.W. Fischer, 'Mapping and Correlating Desert Soils and Surfaces with Imaging Spectroscopy', in *Proceedings of the 3rd Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, JPL Publication 91-28, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 23–32, 1991.
112. G. Pickup, V.H. Chewings, 'Forecasting Patterns of Soil Erosion in Arid Lands from Landsat MSS Data', *Int. J. Remote Sensing*, **9**(1), 69–84 (1988).
113. A.C. Imeson, F. Perez-Trejo, L.H. Cammeraat, 'The Response of Landscape Units to Desertification', in *Mediterranean Desertification and Land Use*, eds. C.J. Brandt, J.B. Thornes, John Wiley & Sons, New York, 447–469, 1996.
114. D.G. Schulze, J.L. Nagel, G.E. van Scoyoc, T.L. Henderson, M.F. Baumgardner, D.E. Stott, 'Significance of Organic Matter in Determining Soil Colors', in *Soil Color*, eds. J.M. Bigham, E.J. Ciolkosz, Soil Science Society of America, Madison, WI, 71–90, 1993.
115. E. Ben-Dor, A. Banin, 'Visible and Near Infrared (0.4–1.1 μm) Analysis of Arid and Semiarid Soils', *Remote Sensing Environ.*, **4**, 110–124 (1994).
116. E. Ben-Dor, A. Banin, 'Evaluation of Several Soil Properties using Convolved TM Spectra', in *Monitoring Soils in the Environment with Remote Sensing and GIS*, eds. R. Escadafal, M.A. Mulders, L. Thiombiano, ORSTOM, Paris, 135–149, 1996.
117. F. Achard, G. D'Souza, 'Collection and Pre-processing of NOAA-AVHRR 1 km Resolution Data for Tropical Forest Resource Assessment', TREES Series A: Techn. Doc. No. 2, EUR 16055 EN, Office for Official Publications of the European Communities, Luxembourg, 1994.
118. D. Koslowsky, 'Mehrjährige Validierte und homogenisierte Reihen des Reflexionsgrades und des Vegetationsindexes von Landoberflächen aus Täglichen AVHHR-Daten hoher Auflösung', *Meterologische Abhandlungen*, Neue Folge, Bd. 9, Verlag Dietrich Reimer, Berlin, 1996.
119. B.N. Holben, Y.J. Kaufmann, J.D. Kendall, 'NOAA-11 AVHRR Visible and Near-IR In-flight Calibration', *Int. J. Remote Sensing*, **11**, 1511–1519 (1990).
120. Y.J. Kaufmann, B.N. Holben, 'Calibration of the AVHRR Visible and Near-infrared Bands by Atmospheric Scattering, Ocean Glint and Desert Reflection', *Int. J. Remote Sensing*, **14**, 21–52 (1993).
121. M.E. James, S.N.V. Kalluri, 'The Pathfinder AVHRR Land Data Set: An Improved Coarse Resolution Data Set for Terrestrial Monitoring', *Int. J. Remote Sensing*, **15**, 3347–3363 (1994).
122. M. Maiden, S. Greco, 'NASA's Pathfinder Data Set Program: Land Surface Parameters', *Int. J. Remote Sensing*, **15**, 3333–3345 (1994).
123. C.J. Tucker, B.J. Choudhury, 'Satellite Remote Sensing of Drought Conditions', *Remote Sensing Environ.*, **23**, 243–251 (1987).

Stellar Spectroscopy

John M. Porter and Iain A. Steele
Astrophysics Research Institute, Liverpool John Moores University, Birkenhead, UK

1 Introduction 1
2 Stellar Atmospheres – What Do We Expect to See? 3
2.1 Hydrostatic Equilibrium 4
2.2 Physical State of the Gas 4
2.3 Energy Transport in the Atmosphere 4
2.4 Outer Atmospheres – Chromospheres and Coronae 5
2.5 Examples of Atmospheres 5
2.6 Radiation 6
2.7 Gray Atmospheres 6
2.8 Formation of Continuum Emission 6
2.9 Formation of Spectral Lines 8
2.10 A Return to Real Spectra 8
3 Spectral Lines 9
3.1 Natural Width 9
3.2 Thermal Broadening 9
3.3 The Voigt Line Profile 10
3.4 Pressure Broadening 10
3.5 Zeeman Splitting 11
3.6 Velocity Fields 11
4 Practical Techniques 13
4.1 Instrumentation 13
4.2 Data Acquisition 13
4.3 Line Identification 14
5 Basic Characterization of Stars 14
5.1 Spectral Classification 14
5.2 Abundance Analysis 15
5.3 Radial Velocity Analysis 16
6 Wavelength Regimes 17
6.1 Radio Spectroscopy (Wavelength 200 μm–100 m) 17
6.2 Far-infrared Spectroscopy (Wavelength 20–200 μm) 18
6.3 Near- and Mid-infrared Spectroscopy (Wavelength 1–20 μm) 18
6.4 Ultraviolet Spectroscopy (Wavelength 100–3000 Å) 18
6.5 High-energy Spectroscopy (Wavelength < 100 Å) 19
7 Conclusions – What Have We Learnt? 20
Acknowledgments 20
Abbreviations and Acronyms 20
Related Articles 20
Further Reading 20

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Spectroscopy started with the examination of the Sun's light, making stellar spectroscopy the oldest branch of the subject. We describe modern stellar spectroscopy, and highlight some of the major successes in understanding the light output from stars and what we can deduce about the underlying objects. We present a description of stellar atmospheres considering the mechanical structure and the transport of energy. This leads on to a discussion of how the continuum and spectral lines of stellar spectra are formed and an explanation of the main features observed. The spectral lines observed in stellar spectra are then discussed and the information which can be obtained from examining their line shapes is examined. These include atmospheric pressure and temperature and both coherent and turbulent velocity fields (from stellar rotation and atmospheric convection currents). Magnetic fields, if large enough, may also be detected from careful examination of spectral lines.

Following a discussion of some practical aspects of observing stars, the main observational stellar classification scheme is presented, along with its interpretation regarding the underlying mass, radius, temperatures and evolutionary phase of the star. The effect of the chemical composition on line strengths is discussed, along with observational techniques of measuring it. The information which is obtained from different wavelength regimes – from radio waves to γ-rays – is presented, along with possible difficulties associated with observing in each regime.

Finally, we discuss what we have learnt from examination of stellar spectra and the current state of theory which is applied to them.

1 INTRODUCTION

The oldest branch of spectroscopy involves examining the light from stars – the field of spectroscopy started with experiments by Newton on the nature of light, using the Sun as his source. Helium was discovered by Lockyer in 1868 when examining emission lines in the solar spectrum (these lines actually originate from the solar corona). Since then, myriad applications of the technique have arisen, producing some significant advances in the study of matter. Spectroscopy has been a vital tool in the study of stars and has provided us with an excellent observational test bed on which to confront our theories. Here, we provide a brief introduction to some aspects of modern stellar spectroscopy.

Astronomy is a difficult science mainly because essentially all the information we receive from stars and galaxies is light. We cannot perform any experiments on the stars, we can only observe them – making this possibly the most extreme branch of remote spectroscopy as well as the oldest. This means that in order to understand anything about the universe, we must acquire as much information as we can from the photons we receive.

What, then, are our aims for stellar spectroscopy? The goals are to determine the physical conditions (pressure, temperature, density, chemical composition and ionization state) of the gas in a star's atmosphere – from this we can describe the conditions of the emitting object as we see it. Then we can attempt to combine many such observations to constrain where that star is in its evolution by comparing them with theories of stellar formation and evolution. We also find information on rotation of the star and any other velocity fields in the atmosphere (turbulence and convection), and we can measure strong magnetic fields. Examining the same star at different wavelengths also gives us information about any circumstellar material, stellar winds and any natal gas left over from the star formation process.

There are several reasons why we should desire to understand stars: first of all for the stars themselves – they are large, powerful, nuclear-driven machines, with a wide range of temperatures and pressures in the same body. Therefore, it is obvious that our curiosity is aroused regarding the detailed formation and evolution processes. There are many questions in stellar astronomy still left to be answered, e.g. how exactly do stars form?, what does rotation do to a star?, what exactly happens in the center of a star?, how does the energy produced there propagate through the star?, what processes occur in the death throes of a star? These questions need a variety of physics to understand them – from fluid dynamics through atomic, nuclear and particle physics to optics. From this, we can use stellar evolution to study other physics; for example, we can provide tests of theories of general relativity, galaxy formation and evolution, cosmology and particle physics. Therefore, this is a considerable challenge!

Let us concentrate on stars, then. The light we observe originates from the very outer regions of the star; we can only see the atmosphere. Much of the information we can derive about the star is indirect and comes from comparing models of stars to our observations. From this technique, we can *infer* masses, radii, temperatures, internal structures, energy generation and transport methods of the stars, even though only perhaps several hundred stars have accurate masses and radii *measured*. The typical mass range of stars is from 0.1 solar masses ($M_\odot = 2 \times 10^{30}$ kg) up to perhaps 100 solar masses. Radii range from 0.01 solar radii ($R_\odot = 6.96 \times 10^8$ m) up to a few thousand solar radii. Atmospheric temperatures of stars lie in the range from 1000 up to 50 000 K (the Sun's atmospheric temperature is 5780 K). The temperature which characterizes the emission is termed the 'effective' temperature, $T_{\rm eff}$ – this is defined to be the temperature of a star with the same radius emitting a blackbody spectrum if it produced the same total emission. The total flux from the stars is $F = \sigma T_{\rm eff}^4$, where σ is Stefan's constant. The total energy emitted by the star is this flux, multiplied by the total surface area of the star $4\pi R^2$, or $L = 4\pi R^2 \sigma T_{\rm eff}^4$.

A star's ingredients are hydrogen, making up around 70–75% by mass of the star, helium (around 25–30% by mass) and the rest (0.1–2%) the other elements in the periodic table – termed 'metals' in astronomy. A star is essentially a self-gravitating ball of gas which is prevented from collapsing under its own gravitational attraction by nuclear reactions in the central regions. The reactions mainly fuse hydrogen to helium, but in later stages of a star's life fusion produces other elements up to iron. The energy liberated in these reactions diffuses outward from the star, and eventually escapes into space, producing the photons we observe.

If we observe a sample of stars (e.g. several hundred of the Sun's close neighbors), then we find that if we plot

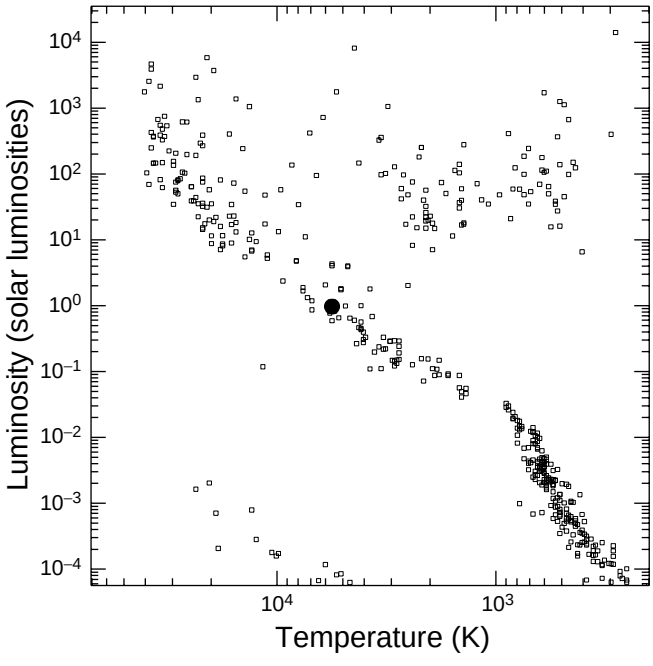

Figure 1 Plot of the effective temperature versus total output (luminosity) for the 300 nearest and 200 brightest stars (note temperature increases from right to left). The luminosity of the stars has been scaled to that of the Sun, which is represented by a filled circle (temperature ~ 5800 K). The stars fall into three main groupings: a large band of stars from the bottom right-hand side to the top left-hand side (the main sequence stars), an area in the top right-hand side (the giant stars) and some stars in the bottom left-hand side (the white dwarfs).

the stars' total output (luminosity) versus the effective temperature there are several distinct groups of stars (see Figure 1). The biggest group of stars (around 90% by number) are those which have temperatures from a few 10^3 to a few 10^4 K and radii from $0.1R_\odot$ to tens of $R_\odot$. These stars are termed ''main sequence stars'' and lie on a diagonal locus from top left to bottom right in Figure 1. Another group has high output and low temperature corresponding to the stars having large radii – the ''giant'' stars. White dwarfs are the low-luminosity, hot stars in the lower left-hand portion of Figure 1. As the main-sequence stars are the most numerous, we shall concentrate mainly on these.

The lifetimes of stars vary widely – the most massive stars (those which live the shortest time) have lifetimes of less than 10^6 years, whereas the low-mass stars (0.1–1 solar mass) live for $>10^{10}$ years. The life of a star may be broken into three stages: the first is where the star forms and there are no significant nuclear reactions in the center (termed the ''pre-main sequence'' phase). Once on the main sequence, the stars spend most of their life (~90%) fusing hydrogen into helium. When the hydrogen fuel has become exhausted in the center then the star evolves off the main sequence and becomes a giant star. These ''post-main sequence'' stars lie on the upper right-hand side of Figure 1, and some of them eventually produce the hot white dwarf stars as remnants of their evolution.

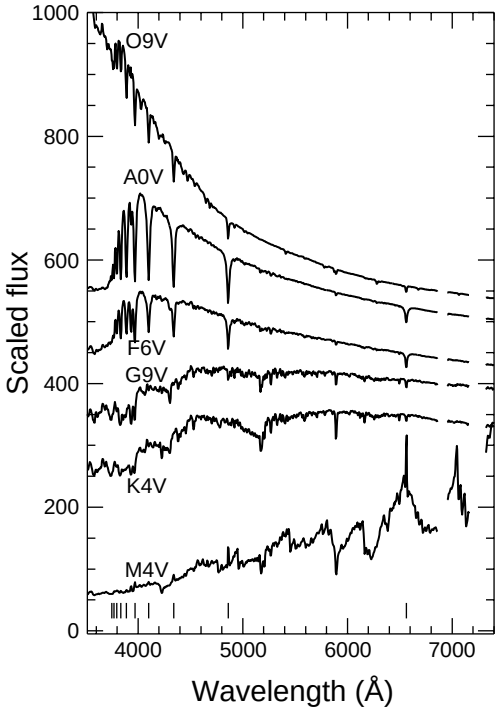

Figure 2 Examples of stellar spectra in the visible wavelength region. The spectra have been scaled and shifted to show the typical features for each temperature. The stellar parameters for these stars are given in Table 1. The positions of the hydrogen lines are marked with the bars at the bottom of the figure.

Table 1 Parameters for the selection of stars whose spectra are presented in Figure 2

Mass ($M_\odot$)	Radius ($R_\odot$)	Temperature (K)	Spectral type
20.0	8.0	30 000	O9V
3.0	2.5	9900	A0V
1.4	1.3	6500	F6V
0.8	0.8	5000	G9V
0.7	0.7	4200	K4V
0.25	0.3	2900	M4V

Masses and radii are given in units of solar mass and radius. For an explanation of the spectral type, see section 5.1.

What do we see, then? In Figure 2, the visible spectra of six stars spanning the main sequence are presented (see also Table 1). There are many changes from the hottest star (the most massive) at the top of the figure to the coolest (the least massive) at the bottom. In this article we shall describe the emission from the stars and present the framework by which we can understand them.

2 STELLAR ATMOSPHERES – WHAT DO WE EXPECT TO SEE?

The physical conditions within a star vary enormously from the center to the surface. In the central regions (the core – comprising around 10% of the mass of the star), thermonuclear reactions provide energy to support the star against gravitational collapse, fusing hydrogen into helium. Temperatures in the cores of stars range from a few 10^6 to a few 10^7 K, depending upon the mass of the star. The energy created there travels outward through the star and is finally released in the atmosphere. As the body of the star is opaque to photons, then we cannot directly see the core and the only information we receive is from the outer layers (the atmosphere) of the star (in the case of neutrino astronomy, we can directly probe the core of the star, as the neutrinos do not interact with the rest of the star).

Although the atmosphere is rather small in its physical extent (for the Sun, the atmosphere is only ~300 km thick, compared with the solar radius of $\sim 7 \times 10^5$ km), it is very important that we understand both the structure and the way energy is transported in the atmosphere – after all, we obtain essentially all our information about the state of the star from this region.

A zero-order approximation to the radiative emission of any star is that it is a blackbody at a characteristic temperature T_{eff}. Although this approximation is in fact

surprisingly good for most stars, in order to interpret the spectrum of a star correctly we must understand the region from which the photons we see come – the atmosphere.

2.1 Hydrostatic Equilibrium

Gas in the atmosphere is attracted towards the star due to gravity; it is prevented from collapsing by the pressure in the gas. The equilibrium attained when these two oppositely directed forces balance is called *hydrostatic equilibrium*. This leads to a relation between the radial pressure gradient of the gas $\mathrm{d}P_r/\mathrm{d}r$ at radius r, the mass within this radius M_r and the density ρ_r (Equation 1):

$$\frac{\mathrm{d}P_r}{\mathrm{d}r} = -\frac{GM_r\rho_r}{r^2} \tag{1}$$

where $G = 6.676 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$ is the gravitational constant.

2.2 Physical State of the Gas

The gas in the atmosphere can be well approximated by an ideal gas and so the pressure is given by Equation (2):

$$P_r = \frac{k\rho_r T_r}{\mu m_{\mathrm{H}}} \tag{2}$$

where k is Boltzmann's constant, T_r is the temperature of the gas at radius r and m_{H} is the mass of a hydrogen atom. The new quantity here, μ, is the *mean molecular weight*. This is defined as the average mass per particle, measured in hydrogen atom masses, e.g. for ionized hydrogen $\mu = \frac{1}{2}$ whereas for neutral hydrogen $\mu = 1$. In very hot stars, the pressure due to radiation has a significant effect and it must be taken into account. This radiation pressure is only a function of temperature, $P_{\mathrm{rad}} = \frac{1}{3}aT_r^4$, where a is the radiation constant.

From the outer edge of the atmosphere traveling inwards, the pressure increases, as do the density and temperature. The state of the elements in the atmosphere must also be taken into account when calculating its structure and emission properties – consideration must be taken into account of the ionization structure. This is determined via the Saha equation describing a collisionally dominated gas at temperature T (Equation 3):

$$\frac{n_1}{n_0}P_{\mathrm{e}} = \frac{(2\pi m_{\mathrm{e}})^{3/2}(kT)^{5/2}}{h^3}\frac{2u_1(T)}{u_0(T)}\exp\ -\frac{\eta}{kT} \tag{3}$$

where n_1/n_0 is the ratio of ions to neutrals, P_{e} is the pressure exerted by the electrons which have mass m_{e}, h is Planck's constant, η is the ionization potential and $u_1(T)$ and $u_0(T)$ are the partition functions of the ionized and neutral species, respectively, at temperature T. Finally, the excitation of each elemental species also needs to be calculated in order to discuss the flow of energy through the star (see below). The fraction of atoms (or ions) which are in the ith energy level is proportional to the Boltzmann factor and the statistical weight g_i (Equation 4):

$$N_i \propto g_i \exp\ -\frac{\chi_i}{kT} \tag{4}$$

where the statistical weight is $g_i = 2J + 1$, where J is the inner quantum number (e.g. for hydrogen $g_i = 2i^2$), and χ_i is the excitation potential (the difference in energy between the bound state i and the ground state).

2.3 Energy Transport in the Atmosphere

The energy created in the stellar core by nuclear fusion travels out toward the edge of the star and in the atmosphere it is finally able to escape, and so the star emits photons. The transport of the energy may occur by two different mechanisms: radiative and convective transport (in white dwarfs where the star is supported against gravity by electron degeneracy pressure, heat conduction by electrons may become important, but for the current argument, conduction is neglected as it is very small for most stars). Radiative transport is *always* present, but convective energy transport may only be present in certain regions of the star at certain times in the star's evolution.

All stars have regions where the energy is transported radiatively (i.e. by photons) and convectively (i.e. by bulk motions of gas). Indeed, the energy transport mechanism has some effect on the evolution of a star, e.g. convection homogenizes the chemical composition of the star by mixing, whereas radiative transport does not. In physical terms, if a small parcel of gas is moved adiabatically in pressure equilibrium with its surroundings and it is less dense in its new position than its surroundings, then it will continue to rise – this region will therefore be unstable to convective motions. Convectively unstable regions obey the Schwarzschild instability criterion, relating the pressure and temperature gradients to those for an adiabatically stratified gas (Equation 5):

$$\frac{\mathrm{d}\log T}{\mathrm{d}\log P} > \frac{\gamma - 1}{\gamma} + \frac{\mathrm{d}\log\mu}{\mathrm{d}\log P} \tag{5}$$

where γ is the ratio of specific heats. In convective regions the combined temperature and pressure gradients are large enough that it is energetically favorable to move the gas physically (i.e. convect) than energy transport via radiation. Note that the composition and ionization state of the gas enters into the instability criterion (via μ) as this effects the equation of state (Equation 2). In the examples of atmospheres in Figure 3(a–d), convective regions are denoted by thick lines.

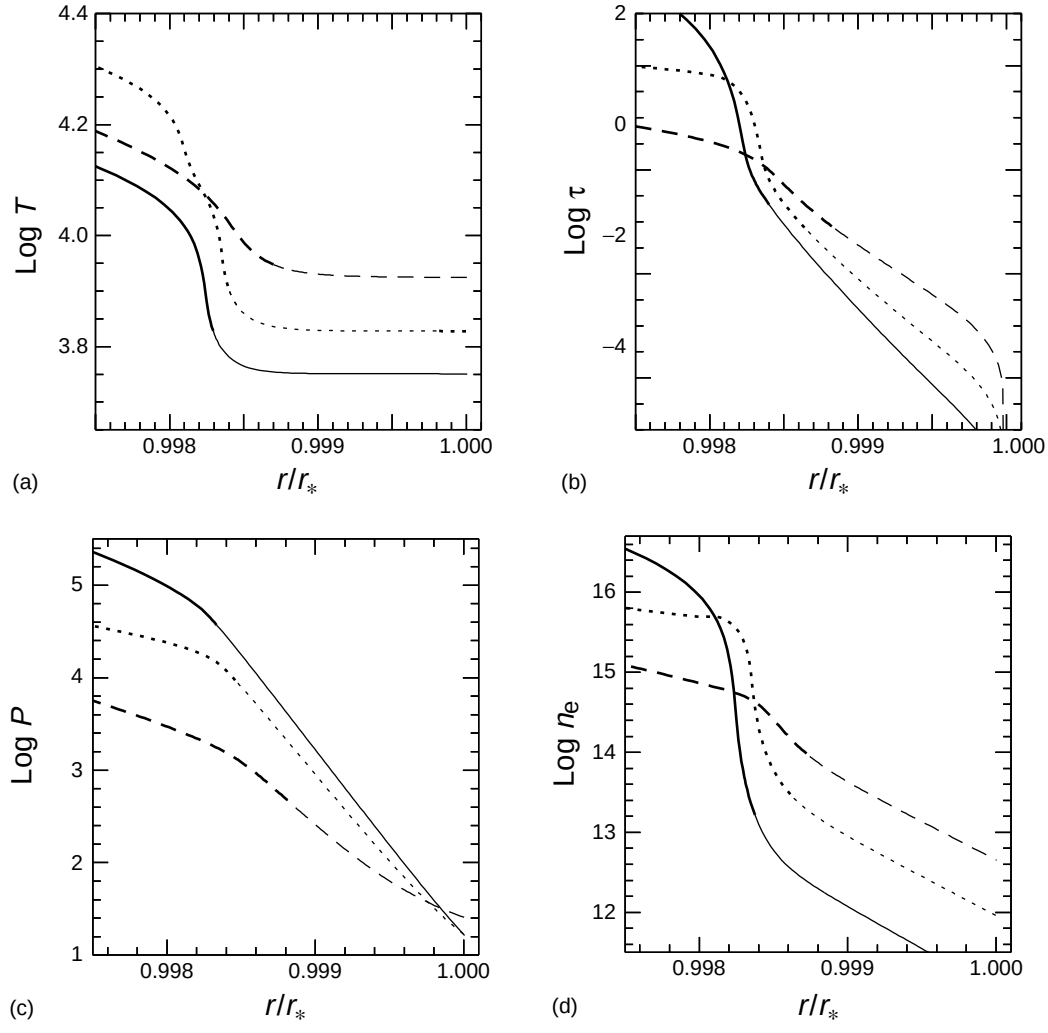

Figure 3 Model atmospheres for three stars with stellar parameters of 1.3 $M_{\odot}$, 1.5 $R_{\odot}$, 6700 K (solid line), 1.7 $M_{\odot}$, 1.7 $R_{\odot}$, 8000 K (dotted line) and 2.4 $M_{\odot}$, 2.0 $R_{\odot}$, 10 000 K (dashed line). Plotted as functions of radius (scaled to the stars' radii) are the logarithms of (a) temperature, (b) gray optical depth, (c) pressure and (d) number density of electrons. The convective regions are shown as thick lines.

2.4 Outer Atmospheres – Chromospheres and Coronae

We have described the main ingredients from which the mechanical structure of a stellar atmosphere can be determined. Although these ingredients apply to all stars, there are extra physical effects which modify the outer regions of some stars' atmospheres. In low–intermediate mass stars (masses up to about the solar mass), magnetic fields have a significant effect. When magnetic field lines twist owing to turbulence in the atmosphere, they may realign themselves to gain a lower energy state by breaking and reconnecting. The extra energy liberated in the reconnection process heats the gas at the outer regions of the atmosphere. For a solar-type star, the decrease in temperature with increasing radius is reversed owing to the magnetic reconnection heating and several hot layers are created – the chromosphere (having temperatures up to 25 000 K) and the coronae (temperatures up to 2×10^6 K). In fact, the thermal energy in the corona of the Sun is so great that the gas there becomes unbound and forms the solar wind.

2.5 Examples of Atmospheres

In order to calculate the full mechanical structure of the atmosphere, we must solve the equation for hydrostatic equilibrium (Equation 1) along with the equation of state (Equation 2) whilst at the same time solving for the ionization and excitation structure (Equations 3 and 4). Clearly this is a difficult process and is analytically intractable. Therefore, the equations are solved numerically. In Figure 3(a–d) the calculated atmospheres of some main sequence stars are shown. As can be clearly seen, the temperature, pressure and

electron density increase as we go deeper into the atmosphere and star. There is a significant change in all three models at ~0.9985 of the stars' radii, which corresponds to the point where energy can escape from the star in the form of photons. To understand this characteristic point further, we must examine the flow of radiation through a star.

2.6 Radiation

The star loses its energy to the rest of the universe by photons escaping from the atmosphere. When considering radiative energy transport, the primary quantity to consider is the *intensity* of the photons I_ν at a frequency ν: this is the amount of energy with frequency ν travelling in a given direction per unit area, per unit solid angle. As they pass through the atmosphere, photons may be absorbed by atoms or ions, scattered in some direction, or created by an excited atom/ion or an ion recombining with an electron. The number of photons absorbed and scattered clearly depends on the number of photons present, so for a gas with absorption and scattering alone, the decrease in the intensity with distance is proportional to the intensity itself. Any extra photon emission from the gas depends only on the composition and physical conditions of the gas (this ignores stimulated emission, which is an excellent approximation). Therefore, we can write an expression for the intensity I_ν: the change in intensity $\mathrm{d}I_\nu$ in a distance $\mathrm{d}s$ is given by Equation (6):

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\kappa_\nu I_\nu + \varepsilon_\nu \tag{6}$$

where κ_ν is the *opacity* of the gas describing the scattering and absorption processes and ε_ν is the number of photons emitted each second per unit volume (both of which are functions of frequency ν).

By dividing by the opacity, this equation (the *equation of radiative transfer*) becomes Equation (7):

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = I_\nu - S_\nu \tag{7}$$

where we have defined two new quantities. S_ν is the *source function*. If the star is in complete thermodynamic equilibrium, or if the thermodynamic properties do not vary enormously over the mean free path of a photon (termed *local* thermodynamic equilibrium), then S_ν is equal to the Planck blackbody function. Local thermodynamic equilibrium is usually a reasonably good approximation in stellar atmospheres, although many modern calculations do not use it – these solve for the excitation and ionization state exactly using a detailed calculation of the collisional and radiative processes. The other new quantity turns out to be extremely useful: $\mathrm{d}\tau_\nu = -\kappa_\nu \,\mathrm{d}s$ is the *optical depth* and characterizes how far a photon will travel before it interacts: at an optical depth of unity, 63% of photons have been absorbed or scattered. Optical depths for stellar atmospheres are always measured from infinity inwards (Equation 8):

$$\tau_\nu = -\int_\infty^r \kappa_\nu \,\mathrm{d}r \tag{8}$$

We find that at an optical depth of $\tau_\nu = \frac{2}{3}$, the flux from the star is equal to the source function. Also, most of the radiation we receive originates at optical depths of less than $\frac{2}{3}$, and so this value is conventionally used to describe how far we can 'see' into a star's atmosphere. The surface at $\tau_\nu = \frac{2}{3}$ is called the *photosphere*. Note that the physical location of the photosphere is dependent upon frequency ν through the opacity. Hence at different frequencies, we see to different physical depths in the atmosphere.

2.7 Gray Atmospheres

There are three sources of opacity present in atmospheres: bound–free absorption, free–free absorption and bound–bound absorption. The first two of these have free electrons as their final state, whereas the last involves an electron jumping between two bound states in an atom or ion. Bound–bound transitions therefore give rise to spectral lines, and involve very specific energies (and hence wavelengths) of photons, whereas the first two can involve many wavelengths, and so provide a continuous absorption coefficient.

Although the opacity is a function of frequency ν, it is often averaged over all wavelengths when calculating model atmospheres – called the *gray* opacity. Examples of the gray optical depth (i.e. that from the gray opacity) are shown in Figure 3(b) for the model atmosphere examples. It is worthy of note that where the optical depth becomes less than a few, there is a severe change in the thermodynamic quantities – clearly radiation (and hence energy) may escape out of these regions. Also note that the atmosphere may be optically thin $\tau < \frac{2}{3}$ in regions which are convecting for some stars. This allows us to ''see'' the convective motions in the atmosphere.

2.8 Formation of Continuum Emission

In the visible–radio wavelength regime, the continuous absorption coefficient tends to be dominated by hydrogen for all but the hottest stars (where helium free–free and bound–free opacity is important) and the coolest stars where molecules form in the atmosphere. At shorter wavelengths [into the ultraviolet (UV)], Rayleigh scattering off metallic ions can become important.

As hydrogen contributes most to opacity, let us use it as an example – many other elements show similar features in their continuous absorption coefficient. For

temperatures higher than ~8000 K, the bound–free opacity dominates, whereas below this level, it is the H^- ion which provides most of the opacity (below 16 000 Å via bound–free transitions, above via free–free transitions). Bound–free absorption introduces "edges" into the opacity corresponding to wavelengths where the atom is ionized from a particular level. For example, there are absorption edges at 912, 3647 and 8206 Å as the atoms with electrons in principle quantum number $n = 1$ (Lyman), $n = 2$ (Balmer) and $n = 3$ (Paschen), respectively, are ionized. Of course, the free–free opacity does not show any edges.

The total opacity is the sum of all the individual opacity weighted by the number distribution of hydrogen atoms in each excited or ionized state governed by the Boltzmann equation. We present three examples of the total continuous absorption coefficient due to hydrogen in Figure 4, where we have broken the opacity down into its constituent parts.

What effect does the variation of the opacity with wavelength have on observations? Assuming local thermodynamic equilibrium, the source function S_ν is the blackbody function. As the radiation flows out of the atmosphere, S_ν has the features of the opacity imposed on it. Therefore, edges may be introduced to the spectrum corresponding to the sharp changes in the continuous absorption coefficient – in the visible regime, this produces the Balmer discontinuity at 3647 Å for hotter stars. These features are illustrated in Figure 5, where two examples of the spectrum produced from model atmosphere are shown: the dashed lines are the blackbody

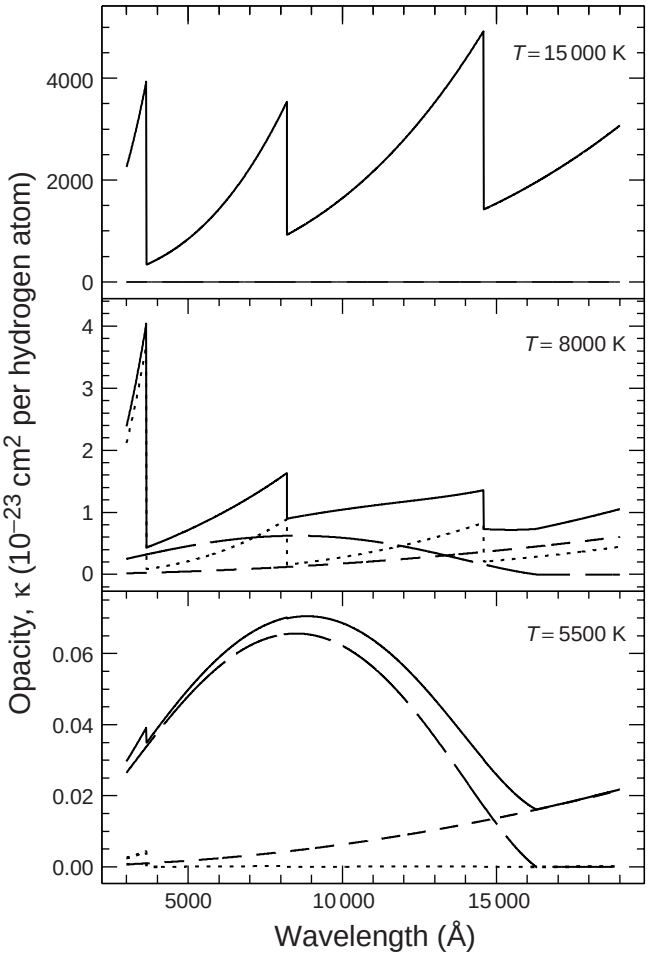

Figure 4 The continuous opacity coefficient for three temperatures. The dotted line corresponds to atomic hydrogen bound–free transitions, the short dashed line to H^- free–free absorption and the long dashed line to H^- bound–free absorptions. The solid line is the total of the three absorption coefficients.

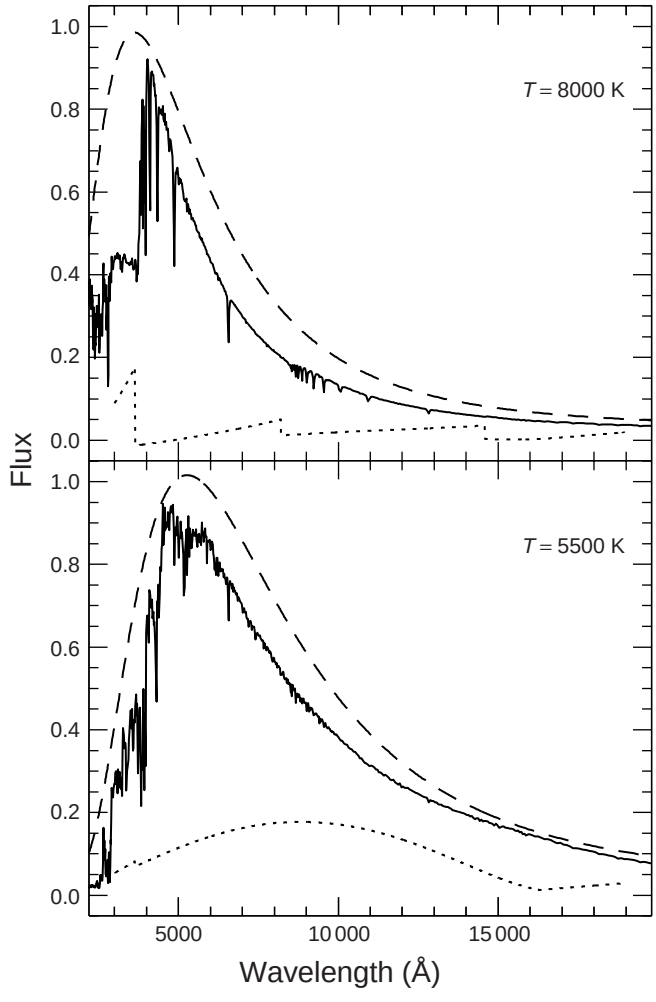

Figure 5 Influence of the opacity on the resultant spectrum: the dotted line is the absorption coefficient of Figure 4 for reference and the dashed line is the blackbody emission at the temperature of the star scaled by ~1.1 for clarity. The solid lines are spectra calculated from model atmospheres. Note the effect of the different total opacity on the resultant spectrum, e.g. the obvious Balmer discontinuity at 3647 Å for the 8000 K star is not present for the 5500 K star.

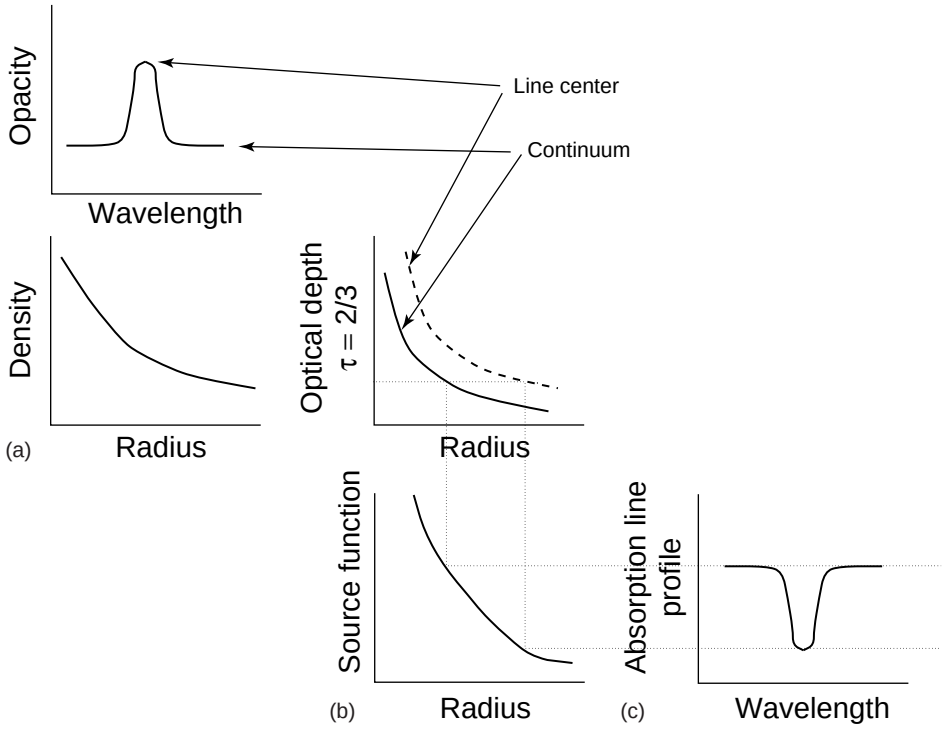

Figure 6 Schematic representation of why spectral lines forming at the photosphere are absorption lines: the opacity and density yield the optical depth. Most of the photons escape at optical depths of less than $\tau = \frac{2}{3}$. As the opacity is higher in the center of the line than in the continuum, the optical depth for the line center (dashed line) is higher than for the continuum (solid line). Therefore, the source function at an optical depth of $\frac{2}{3}$ is less in the line center than in the continuum, producing an absorption line.

spectra for the same temperatures as the models – note that the large Balmer discontinuity for the hotter star is negligible for the cooler star, which is a direct by-product of the opacity (denoted by the dotted lines in Figure 5).

In fact, the size of the Balmer discontinuity can be used to measure the temperature of the star. At temperatures greater than about 9000 K, the ratio of the opacity on each side of the discontinuity becomes proportional to the number of hydrogen atoms in the second and third quantum levels. This is governed by the Boltzmann equation, which is only a function of temperature. For other ranges of temperatures, the size of the discontinuity depends on both the electron density and temperature.

2.9 Formation of Spectral Lines

The continuous spectrum is useful for determining several parameters about the star, most notably the temperature. More detailed information can be obtained from examining features caused by bound–bound transitions in the atmosphere – the spectral lines.

The lines correspond to significant increases in the opacity over a very limited frequency range. As the photosphere is always defined at an optical depth of $\tau_\nu = \frac{2}{3}$, then the photosphere for the lines is physically higher up in the atmosphere, and so the source function S_ν is less. An increase in opacity and a decrease in source function lead to a net decrease in intensity, and so the lines formed in the photosphere are always seen as *absorption* lines (i.e. the line intensity is less than the continuum on either side). This is illustrated schematically in Figure 6(a–c). In cooler stars the generation of the chromosphere and corona due to magnetic field lines reconnecting lead to an emission line spectrum containing lines from (possibly highly) ionized ions. This is because at high temperatures the line source function is large (although the continuum source function is small – in this case the chromospheric and coronal gas is definitely not in thermal equilibrium) and hence there is a net increase in intensity, thereby producing an emission line.

2.10 A Return to Real Spectra

Let us now return to the spectra shown in Figure 2 and point out some of the features which have been described. For the hot stars, the spectrum rises toward low wavelengths, until a maximum is reached. The hydrogen lines are particularly obvious for these stars. Neutral helium lines appear at 5015, 5875 and 6678 Å. The NaI (neutral sodium) line at 5889 Å is particularly visible for the cooler stars, as is a magnesium line at 5183 Å. Neutral calcium lines appear at 4426, 5588 and 6162 Å. Cooler

stars have many more lines, giving the spectra a more ''ragged'' appearance. For the lowest-temperature stars the hydrogen lines finally come into emission – a sign of the active chromosphere region.

Owing to their temperatures, the cooler stars may form molecules in their atmospheres. The spectra of the cooler stars readily display evidence for molecules – wide absorption bands created by TiO are clearly visible.

3 SPECTRAL LINES

Aside from being useful in determining what elements are present in the star's atmosphere and what ionization states they are in, spectral lines offer a practical way of determining other parameters of the star through their line shape. At a first glance the lines should be infinitely thin as they are produced by electrons jumping from one set of quantum numbers to another – a transition which has a precise value of energy. However, there are several ways in which the line is broadened, or possibly split into several components. These include the natural width and also Stark and Zeeman splitting, although for most stars the line shapes are dominated by the Doppler effect. We must be aware, however, that although the shapes of spectral lines can in principle produce a lot of information, there is also the intrinsic broadening of the spectrograph to consider. This instrumental broadening is *always* present in measured spectra, and must be taken into account in interpreting line shapes – any intrinsic line shape is convolved with the instrumental line shape.

3.1 Natural Width

Classically, the electron–nucleus system is seen as an oscillator. When the electron emits radiation the oscillator is damped, and hence there is a small spread of energies with which the photon can emerge. This leads to the Lorentzian (or damping) profile (Equation 9):

$$\phi_{\mathrm{L}} \propto \frac{\gamma/4\pi}{(\nu-\nu_0)^2+(\gamma/4\pi)^2} \tag{9}$$

where ν_0 is the central wavelength, $\gamma = 8\pi^2 e^2 \nu_0^2/3m_{\mathrm{e}}c^3$ is the damping constant, e is the electronic charge and c is the speed of light.

A quantum description of the natural width uses Heisenberg's uncertainty principle, $\Delta E \Delta t \approx \hbar$, where Δt is the typical time an electron is in a given excited state and ΔE is the energy uncertainty. If a line has a large transition probability, then Δt is small, leading to a large ΔE, and the line is broadened. This quantum approach, of course, leads the the same line shape from the classical approach, and the profile is shown as a dashed line in Figure 7. The half-width in frequency characterizing the profile is $\delta\nu = \gamma/4\pi$. More usefully, this corresponds to a *wavelength* natural half-width (which is independent of wavelength) of $\delta\lambda_{\mathrm{N}} = 1.2 \times 10^{-4}$ Å.

3.2 Thermal Broadening

The atoms in the atmosphere are moving owing to their thermal energy – their number distribution with velocity is Maxwellian. The typical thermal velocity for an element with mass number A is $v_0 = \sqrt{2RT/A}$, where R is the gas constant and T is the temperature. Some of the emitting atoms/ions will be moving toward the observer and some away with a velocity v, leading to a shift in frequency due to the Doppler effect of $\Delta\nu/\nu_0 = \pm v/c$ (the positive/negative sign corresponds to atoms moving toward/away from the observer).

When the Maxwellian distribution of velocities is considered, an infinitely thin line is broadened into the ''Doppler'' profile (Equation 10):

$$\phi_{\mathrm{D}} \propto \frac{1}{\Delta\nu_{\mathrm{D}}} \exp\ -\left(\frac{\nu-\nu_0}{\Delta\nu_{\mathrm{D}}}\right)^2 \tag{10}$$

where $\Delta\nu_{\mathrm{D}} = v_0\nu/c$ is the typical thermal width (the combined effect of natural and thermal broadening are discussed below). This profile is shown as a dotted line in Figure 7.

As the line is formed at different geometric depths on the photosphere, the temperature of the atoms contributing to the line also varies. This mimics an additional random velocity component to the line and is called microturbulence (note that this is strictly a misnomer, as this effect is not really caused by turbulence). Also, it is possible that the optical depth of the line is less than $\tau_\nu = \frac{2}{3}$ in convective regions (see Figure 3a–d). In this case the photosphere may be formed in a convective region, and assuming that the motions have a

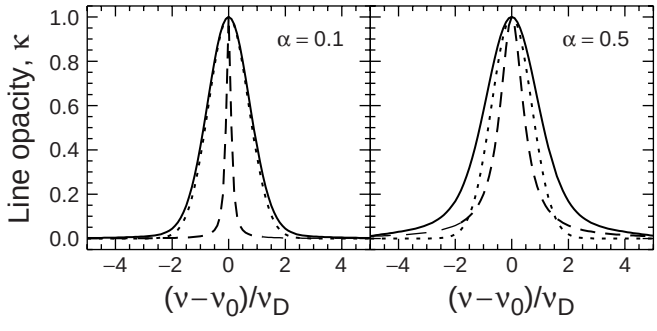

Figure 7 Doppler (dotted line) and Lorentzian or damping (dashed line) line profiles. The solid lines are the Voigt line profile formed from a convolution of the Doppler and Lorentzian profiles for values of the ratio of half-widths of the Lorentzian to Doppler profiles of $\alpha = 0.1$ and $\alpha = 0.5$.

velocity distribution similar to the distribution of thermal velocities then they may be treated in exactly the same way as thermal broadening. These random convective motions are called macroturbulence. They can be included in the line profile in a particularly simple fashion – the typical velocity dispersion v_0^2 in the line profile is increased by summing the component velocity dispersions, $v_0^2 = v_{\text{th}}^2 + v_{\text{micro}}^2 + v_{\text{macro}}^2$, where v_{micro} and v_{macro} are the velocities characterizing the micro- and macroturbulence, respectively and v_{th} is the thermal width.

As an estimate of the typical Doppler widths of lines, consider a gas at a temperature of 6000 K. The thermal velocity of hydrogen is about $10\,\text{km}\,\text{s}^{-1}$. At the first hydrogen Balmer line at a wavelength of 6563 Å (denoted as Hα), this velocity corresponds to a wavelength Doppler half-width of $\delta\lambda_{\text{D}} = v_0\lambda/c = 0.22$ Å, which clearly dominates over the natural width. Both macro- and microturbulent velocities are found to be of the order of a few kilometers per second, making the thermal motions most important in line broadening for hydrogen. However, for higher-mass elements such as iron ($A = 56$), then the thermal speeds are $1.6\,\text{km}\,\text{s}^{-1}$ and consequently the macroturbulence may be dominant.

3.3 The Voigt Line Profile

Both the natural width and the Doppler broadening should contribute to a typical line profile. In the middle regions of the line where the energy is only slightly different from the central energy, the damping of the oscillator is small and the line is dominated by the Doppler profile. Toward the edges of the line (the wings), the damping may become dominant over the Doppler broadening and so the line follows the Lorentzian profile. The final profile is then a convolution of the two profiles and is called the Voigt profile ϕ_{V}. Two examples are given as solid lines in Figure 7 corresponding to ratios of half-widths of 0.1 and 0.5.

3.4 Pressure Broadening

A line may be broadened by the effects of pressure. The two effects which are important here are collisional broadening and Stark broadening.

3.4.1 Collisional Broadening

An atom in an excited state can be influenced by another passing atom. If the timescale for a ''collision'' is small compared with the natural lifetime of the excited atom, then it can de-excite sooner than it would have done in isolation. This reduces Δt in the quantum picture leading to a larger ΔE. Therefore, this collisional de-excitation will broaden the shape of the line in exactly the same way as natural broadening does – the damping constant γ of Equation (9) is increased. The amount of broadening is dependent on the natural lifetime (given by the inverse of the Einstein A_{ji} coefficient for a given line, where i and j are the two levels, j the higher), and the collisional de-excitation rate $R_{ji} = 8.63 \times 10^{-12} n_{\text{e}} \Omega_{ji} / T_{\text{e}}^{1/2} g_j\,\text{s}^{-1}$, where n_{e} and T_{e} are the electron density and temperature respectively, g_j is the statistical weight of level j and Ω_{ji} is the collision strength which is a constant for a given transition. If there are Γ transitions per second caused by collisions, then the effective damping constant $\gamma_{\text{eff}} = \gamma + \Gamma$. However, this form of collisional broadening is often masked by the Doppler effect, described above.

3.4.2 Stark Broadening

When an atom or ion is subjected to an electric field, small changes are induced in the excitation levels of the atom or ion. This is the Stark effect, which can lead to extra broadening of the lines. When the atmospheric pressure is large, the weak electric fields around ions slightly change neighboring atoms'/ions' emission frequencies. In the quantum picture, the change in excitation energies shortens the lifetime of electrons in these levels and leads naturally to a larger ΔE owing to the uncertainty principle. The effect depends upon proximity of ions and is described as the ''molecular'' Stark effect (again this is a misnomer, as molecules are not necessarily involved: the name differentiates it from the Stark effect in a large-scale electric field). The line broadening generated by the molecular Stark effect is exactly the same in form to the Lorentzian profile, and produces its own constant γ_{S} characterizing the broadening.

The change in energy induced by a close perturber on the transition between two levels is $\Delta E \propto R^{-m}$, where the integer m describes the type of interaction and R is the distance to the perturber. If the energy change is linearly dependent on the applied electric field $\mathcal{E}$, then $\Delta E \propto \mathcal{E} \propto e/r^2$, and hence $m = 2$ corresponds to *linear* Stark effect. The perturber is most likely to be a proton or an electron for $m = 2$ (this has most effect on hydrogen lines), electrons for $m = 4$ (the quadratic Stark effect, which may be important for hot stars) and neutral hydrogen for $m = 6$ (Van der Waals effect, which is more important for cool stars). The energy change corresponds to a shift in frequency of $\Delta\nu = C_m R^{-m}$, where C_m is a constant dependent on the type of effect.

Calculations and measurements for the values of C_m are difficult and indeed well known for only a few simple atoms. For the linear Stark effect in the case of large field strengths, an analytical expression for the opacity as a function of wavelength in the line wings may be derived (Equation 11):

$$\phi_{\text{S}} = \frac{C\mathcal{E}_0^{3/2}}{(\lambda - \lambda_0)^{5/2}} \tag{11}$$

where $\mathcal{E}_0$ is the typical field strength defined as the electronic charge divided by the square of the mean interparticle separation $\bar{r}$ (Equation 12):

$$\mathcal{E}_0 = \frac{e}{\bar{r}^2} = e\left(\frac{4\pi n}{3}\right)^{2/3} \tag{12}$$

where n is the number density of particles. The values of C in this expression vary for each line – for the hydrogen Balmer series, Hα 6563 Å, $C = 3.13 \times 10^{-16}$; Hβ 4861 Å, $C = 0.89 \times 10^{-16}$; Hγ 4340 Å, $C = 0.44 \times 10^{-16}$; Hδ 4101 Å, $C = 0.31 \times 10^{-16}$. This broadening of lines can be seen in Figure 8, where the Balmer series of hydrogen lines for three hot stars (of similar temperature at ~25 000 K) are plotted. The continuum variation has been divided out of the spectra so that a direct comparison of the lines can be made. The bottom line corresponds to a main-sequence star which has a high pressure in the atmosphere, the middle line to a giant star and the top line to a supergiant star which has low pressure in the atmosphere.

Another effect of an electric field is the lifting of the energy degeneracy of the levels. This should produce several broadened lines for each transition, which blend into each other. For hydrogen the degeneracy increases with principal quantum number n, although the energy differences decrease as $1/n^3$. Therefore, the effects should be more noticeable in higher transitions (i.e. it will affect Hγ more than Hβ and Hβ more than Hα). At some n the broadening of the lines becomes larger than the differences in the lines themselves and the hydrogen lines blend into each other (this can be easily seen in Figure 8). This provides a method of measuring electron pressure – the upper level n_u where the lines blend is related to the electron density n_e via $\log n_e = 23.26 - 7.5 \log n_u$.

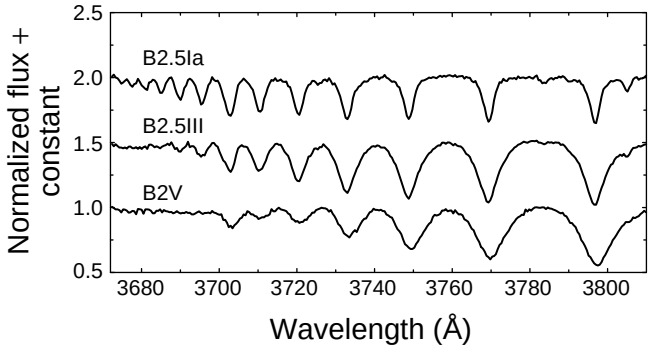

Figure 8 The Balmer hydrogen line series for three stars of temperature ~ 25 000 K. The continuum has been divided into these stars to illustrate the effect of pressure broadening on the lines. The bottom line corresponds to the star which has the highest pressure in the atmosphere and the top line to the star with the lowest atmospheric pressure.

3.5 Zeeman Splitting

The Zeeman effect occurs when the emitting gas is in a magnetic field. The spin of the electron and the magnetic moment of the orbital interact with the magnetic field, and the energy degeneracy of the quantum state is lifted. Therefore, there exists a splitting of the previously degenerate energy levels and instead of a single component, three components are seen (corresponding to different magnetic quantum numbers). The central component, at a wavelength unaffected by the magnetic field, is denoted as the π component, whilst the two other components are denoted σ^+ and σ^- for the higher and lower energy (corresponding to longer and shorter wavelengths, respectively). The degree of separation between the π and σ components is given by Equation (13):

$$\Delta\lambda_Z = 4.7 \times 10^{-9}\lambda^2 B \text{ Å} \tag{13}$$

where B is the magnetic field strength measured in tesla and λ is measured in ångstroms.

Magnetic fields vary widely in stars – in main-sequence stars B may be as large as 1 T for the most magnetically active stars, whereas for white dwarf stars fields of up to 10^4 T have been found [although the majority of white dwarfs (~96%) have not been observed to have any magnetic effects]. Therefore, Zeeman splitting will be negligible in most stars, and it is only clearly detectable in some white dwarf stars. We should note that magnetic fields may produce other features in the radio spectrum (see section 6.1). Figure 9 shows an example of Zeeman splitting of the hydrogen lines in the star PG 2329+267, which has a total magnetic field strength of ~230 T (these observations can only measure the line-of-sight magnetic field, which is around 160 T for this star). Note also in Figure 9 the very wide Stark broadened wings of the hydrogen lines.

3.6 Velocity Fields

The velocity fields we have dealt with so far are coherent only over a small part of the star, and hence we have assessed their broadening effect by assuming that motions are uncorrelated and are distributed in a Gaussian fashion about a mean of zero. However, there are velocity fields which are coherent over scale lengths similar to the physical size of the star which have to be taken into account. The first of these is stellar rotation. All stars rotate at some level, and rotation velocities up to the breakup velocity (where the centripetal effects of rotation and gravity balance and so the gas becomes unbound, and the star breaks up) have been observed.

Rotation affects the line profile by Doppler shifting parts of the star – some are approaching the observer

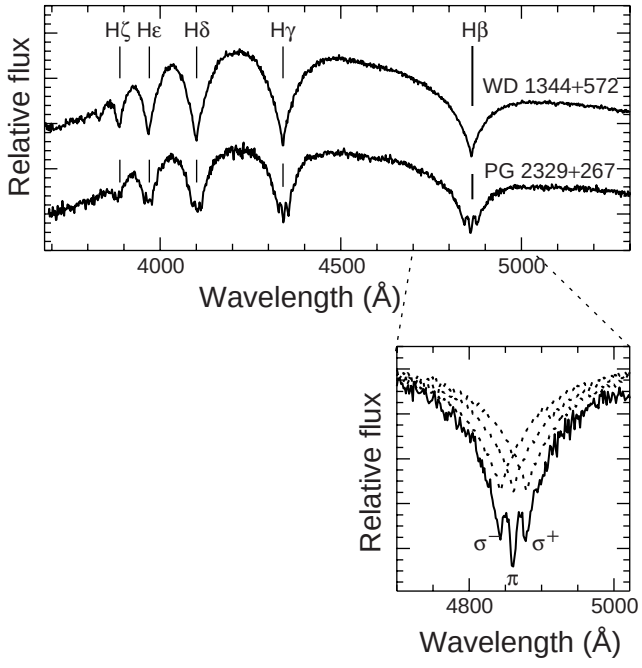

Figure 9 Example of Zeeman splitting of hydrogen lines. The top spectrum is a non-magnetic white dwarf star (WD 1344 + 572). The lower spectrum is for the magnetic white dwarf star PG 2329 + 267 which has a magnetic field of 230 Tesla. The region around Hβ 4681 Å has been expanded to display the Zeeman splitting, showing the line profile for PG 2329 + 267 (solid line) with the reference line profile of WD 1344 + 572 (dotted lines) shifted as if it had the same magnetic field to illustrate how the Zeeman split line profile is built up.

and some are receding, and hence the effect is symmetric about the line center. Because of the coherent nature of the rotation, we cannot simply add another velocity component to the Doppler profile, but must convolve the Doppler/Voigt profile with the profile characteristic of rotation.

The rotation profile as a function of wavelength is given by Equation (14):

$$\mathcal{R} = \frac{(2-\varepsilon)[1-(\lambda-\lambda_0)^2/\Delta\lambda^2]^{1/2} + \dfrac{\pi\varepsilon}{2}[1-(\lambda-\lambda_0)^2/\Delta\lambda^2]}{\pi\Delta\lambda(1-\varepsilon/3)} \qquad (14)$$

where ε is a coefficient of order unity describing the variation of angle of incidence of the observer to the atmosphere across the observed hemispherical surface of the star. $\Delta\lambda$ is the wavelength shift corresponding to the projected rotational velocity at the equator ($v \sin i$, where v is the equatorial velocity and i is the angle from the polar axis to the line of sight). This profile is shown with examples of observed lines in Figure 10. Often the rotation velocity of a star is large enough to dominate over most of the line profile.

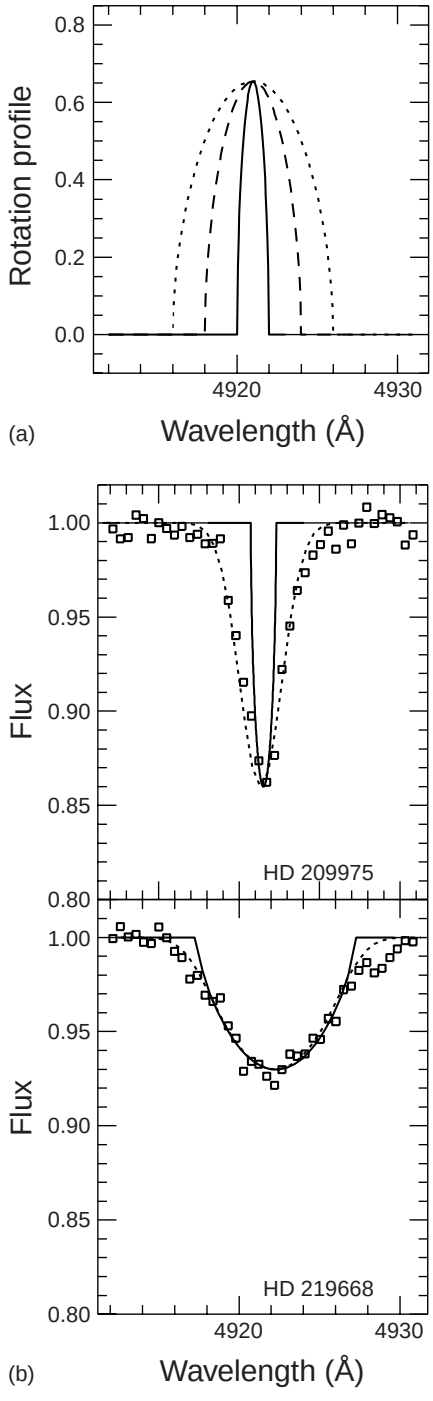

Figure 10 Rotational profile and examples for the HeI 4921 Å line. In (a), three profiles have been calculated. These have $\Delta\lambda = 5$ Å (dotted line), 3 Å (dashed line) and 1 Å (solid line), which correspond to projected rotational velocities $v \sin i$ of 300, 180 and 60 km s^{-1}, respectively. In (b), the He 4921 Å line profile for two stars is presented (squares): HD 209 975 has a projected rotational velocity of ∼30 km s^{-1} and HD 219 668 has $v \sin i \approx 300$ km s^{-1}. The solid lines are the rotational profiles given in Equation (15) and the dotted profiles are the rotational profile convolved with the instrumental line broadening of ∼1.5 Å.

Other velocity fields may be present around stars owing to the accelerating winds that they may possess. These winds have no effect on the photospheric lines as in general they are diffuse flows. However, some lines may be particularly affected by the flow – for example, in hot stars some carbon, nitrogen and silicon lines in the UV region can display characteristic profiles of winds (see section 6.4).

4 PRACTICAL TECHNIQUES

4.1 Instrumentation

To obtain a spectrum of an astronomical source requires three basic components: a telescope, a spectrograph and a detector. The telescope is responsible for collecting photons from the source and bringing them to a focus. An aperture placed at this focus admits light into the spectrograph, which disperses the light. The resulting spectrum is then imaged by a detector, usually some form of array such as a charged-coupled device (CCD). If a long slit is used as the entrance aperture and the spectrograph and detector array are correctly oriented, then the resulting spectrum will be dispersed along one detector axis in wavelength. The other axis will reflect the spatial distribution of the incoming light along the slit.

The fundamental parameter defining an astronomical spectrograph is the resolving power R, defined according to Equation (15):

$$R = \frac{\lambda}{\delta\lambda} \tag{15}$$

where λ is the central wavelength and $\delta\lambda$ the resolution (roughly defined as the minimum resolvable separation of two spectral lines). As a general rule, increasing R will decrease the sensitivity of the instrument as the light will be dispersed over a greater physical area.

We conventionally divide spectrographs into three resolution ranges, low, medium and high. Low-resolution spectrographs have $R < 1000$ – they are designed for high efficiency, often using all transmission optics (i.e. no reflection) and are therefore most useful for observing very faint objects. Medium-resolution spectrographs have $R \approx 1000–10\,000$ and typically use reflecting optics. Their resolution is sufficient to allow spectral classification using the Morgan and Keenan (MK) system (see section 5.1) and radial velocity determinations for binary systems. High-resolution spectrographs, with $R > 10\,000–100\,000$, often use a cross-dispersed, échelle design in order to obtain a wide wavelength coverage at high resolution. However, because of their complicated optics and high dispersion, their sensitivity limits them to relatively bright objects.

4.2 Data Acquisition

The exact observing procedure followed to obtain a fully calibrated spectrum of an astronomical source naturally varies from instrument to instrument, and also depends on the reason for obtaining the spectrum. For instance, for some applications (e.g. radial velocity determinations) a precise wavelength calibration is vital, whereas for others (e.g. spectral classification) it is not. In this section we therefore simply give a generic overview of the procedure, assuming our aim is to obtain a wavelength and flux (intensity) calibrated spectrum of a star.

The first task in taking a stellar spectrum is, of course, to acquire the object on to the spectrograph slit (which has a typical projected width on the sky of $\sim$1 arcsec). For bright objects a slit-viewing television camera can be used to see the reflection off the slit jaws of the telescope image, and the telescope pointing adjusted to place the object of interest on the slit. For faint objects which cannot be directly viewed, a blind offset procedure must be used, i.e. the slit-viewing camera is used to align the telescope on a nearby bright star, and then a precise offset to the telescope coordinates is applied. When observing objects close to the horizon (i.e. through a thick layer of the Earth's atmosphere), we must be aware of the effect of differential atmospheric refraction. The difference in refractive indices for red and blue light in air can cause a significant shift in the apparent position of the object in the sky between these colors which can be enough to move the object out of the slit at one end of the spectrum. In order to counter this effect, many astronomical spectrographs have the ability to rotate the slit so that it can be aligned so that it runs parallel to the direction of refraction (i.e. perpendicular to the horizon) at any given sky position.

Once we have taken an exposure of the spectrum, it needs to be corrected for several aberrations due to the detector itself, the spectrometer and the effect of the Earth's atmosphere. First the spectrum must be calibrated in wavelength so that we know exactly which wavelength corresponds to which pixels of the CCD image. This uses the traditional spectroscopic approach of using arc lines to provide wavelength references, typically CuAr or CuNe. Because of the extreme faintness of many stellar spectra, it is not feasible to expose the arc lines and source simultaneously on the detector. Therefore, an arc calibration image is usually obtained both before and after the source exposure (the two arc exposures allowing calibration of any flexing of the spectrometer due to the changing gravity vector as the telescope tracks the source across the sky.)

Next we consider removing the signature of the detector from the data. The first effect we must consider is the removal of small-scale, pixel-to-pixel variations in

detector sensitivity. We do this by constructing a so called ''lamp-flat'' – effectively a map of these variations which we can divide into our image. This is obtained by observing a tungsten lamp (which has a smooth, blackbody-like spectrum) mounted to shine on to the entrance slit. However, because the tungsten lamp light path is not the same as that followed by light coming through the telescope from the star, the lamp-flat does not accurately reflect the large-scale spatial sensitivity variations across the detector array. To map these, we must observe a ''sky-flat'' at twilight by placing the slit over a blank area of sky that is relatively bright. This sky-flat can then be collapsed in the wavelength direction to make an overall map of the spatial variation. In turn this is then divided into the lamp-flats to produce a master flat-field which can be divided into an image to remove the spatial and small-scale wavelength sensitivity variations.

Finally, we must consider removing the large-scale wavelength sensitivity variation from the spectrum. In the visible wavelength regime this is caused mainly by the dependence of the quantum efficiency of the detector on wavelength, i.e. the responsivity of the detector is not constant over the wavelength range over which it is used. However, in the near-infrared (NIR) region ($\lambda = 1–20\,\mu m$) it is dominated by the strongly wavelength-dependent water vapor absorption of the Earth's atmosphere. The variation can be eliminated by observing a ''standard star'' which has a well known spectrum. Dividing the observed standard star spectrum by its actual spectrum yields a pixel-to-pixel correction curve.

This may then be divided into the observed spectrum of the source in which we are interested to produce its ''real'' spectrum, which is wavelength calibrated and free of aberrations produced by the spectrometer and detector. However, we should note that we are not only observing the star, but also the night sky around the star. The spectrum will therefore contain contaminating telluric ''sky'' emission lines (mainly caused by OH airglow in the upper atmosphere), which can be removed by extracting a sky region (i.e. one devoid of object flux) from the image of the same size as the object region and subtracting the resulting sky spectrum from the object spectrum. Finally, we are left with a spectrum of the object star itself free from any aberration (although residual errors introduced by the photon noise in the object and sky and the noise characteristics of the detector will still remain in the spectrum).

4.3 Line Identification

Spectral line identification may be carried out using either laboratory or stellar line lists. These traditionally group species in terms of multiplets, with the relative strengths of the components given. This means that if a line is suspected to be due to a certain species, the identification can be confirmed by searching for the other components of the multiplet in the spectrum and checking that they have the appropriate strengths. We note here that astronomers have a ''nonstandard'' method of referring to the ionization state of atoms and ions. In this method, neutral species (i.e. atoms) are indicated by the appropriate chemical symbol, followed by the small roman numeral I. Therefore, neutral hydrogen is referred to as HI, and neutral oxygen by OI. Ionized species are referred to by a roman numeral suffix with the value of one more than the ionization state. Therefore, singly ionized helium (He^+) is referred to in the astronomical literature as HeII and triply ionized nitrogen (N^{3+}) as NIV.

5 BASIC CHARACTERIZATION OF STARS

5.1 Spectral Classification

Spectral classification is a fast and efficient tool for deriving the basic parameters (luminosity and surface temperature) of stars. Classically, it is based on a ''by eye'' comparison of the strength and width of spectral features between the object of unknown spectral type and spectra of stars of known spectral type (''spectral standards''). Alternatively, the comparison can be carried out by computer using minimization techniques. Also, the spectral type of a star may be found by taking broad-band observations – the most common scheme for doing this uses three filters covering the visible spectrum: U (UV) from 3000 to 4000 Å, B (blue) from 3800 to 5200 Å and V (visible, yellow–green) from 4800 to 6200 Å. Comparison of the flux ratios in these filters to spectral standards provides the spectral type – essentially this is *very* crude spectroscopy.

The primary reference source for comparison stars is the MK list of standards. The MK classification scheme classifies spectra on the relative strengths of lines to yield a ''spectral class'' and on the widths of those lines to yield a ''luminosity class''. The core MK spectral classes are O, B, A, F, G, K and M. Each of these is nominally divided into 10 subclasses which are indicated by arabic numerals, e.g. a B2 star is closer to an O star than it is to a B7 star. The main observational characteristics of each spectral class are given in Table 2. From the table, it is clear that species with high ionization and excitation potentials dominate the so called early (O–B) spectral classes, indicating that the temperature in their photosphere is large. In contrast the late (K–M) spectral classes show lines of neutral atoms and molecules, indicating a much lower temperature (see Figure 2).

In section 3.4 we demonstrated that the width of photospheric absorption lines is affected by pressure broadening, with broader lines indicating higher pressure

Table 2 MK spectral classes and their main line features (see section 5.1)

Spectral class	Spectral features
O	SiIV, NIII and HeII strongest; H, HeI lines weak
B	SiII, SiIII, HeI strongest, H stronger
A	H strongest, CaII weak
F	CaII stronger, H weaker, numerous singly ionized metals
G	CaII strongest, H weak, weak neutral metals
K	CaII weaker, H very weak, strong neutral metals, weak molecules
M	Strong molecular bands (TiO, VO)

The O type stars are the hottest and the M type the coolest (also see Table 4).

and hence surface gravity. As we have seen in section 1, there is a linear relationship between the luminosity of a star L and its surface area $A = 4\pi R^2$ (where R is the star's radius), i.e. $L = 4\pi R^2 \sigma T_{\text{eff}}^4$. Also, there is an inverse proportionality between the acceleration due to gravity at its surface (surface gravity) and radius, $g = GM/R^2$, where M is the star's mass. Therefore, it is apparent that a more luminous star of a given effective temperature will have a lower surface gravity and hence pressure. Consequently, the photospheric lines in a more luminous star will be affected less by pressure broadening than a less luminous star, as illustrated in Figure 8. We can therefore use the width of the photospheric lines in a spectrum to derive the ''luminosity class'' of a star which is denoted by a roman numeral suffix to the spectral class (see Table 3).

The combination of spectral class and luminosity class is often called a ''spectral type'', and gives a fairly complete description of the nature of a star (Table 4). For example, by knowing a star to be of spectral type G2V we can say its effective temperature is ~5800 K and its luminosity is 1 $L_\odot$ (i.e. one solar luminosity $L_\odot = 3.8 \times 10^{26}$ W). Similarly, the star Betelgeuse in the constellation of Orion has a spectral class of M2I. It therefore has $T_{\text{eff}} \approx 3300$ K and $L \approx 7 \times 10^4\, L_\odot$. Additional suffixes may be added to the spectral type to give further information about the spectrum, often indicating nonstandard features (see Table 5). For example, the B0 dwarf star γ Cas shows strong emission lines of hydrogen in its spectrum, and therefore has the spectral type B0Ve. Also for close binary systems a composite spectrum of both components may be recorded, in which case the two spectral types are given joined by a + symbol (e.g. B5V + F7III).

Table 3 MK luminosity classes

Luminosity class	Star type
I	Supergiants
II	Bright giants
III	Giants
IV	Subgiants
V	Main sequence (dwarfs)
VI	Subdwarfs

Table 4 Variation of temperature and luminosity with spectral type

Spectral class	T_{eff} (K)	Luminosity class		
		Class V ($L/L_\odot$)	Class III ($L/L_\odot$)	Class I ($L/L_\odot$)
O5	40 000	8×10^6	–	–
B0	28 000	4×10^4	9×10^4	3×10^5
B5	15 500	870	2900	1×10^5
A0	9900	70	160	5×10^4
A5	8500	17	70	3×10^4
F0	7400	7	25	2×10^4
F5	6600	3.5	22	1×10^4
G0	6000	1.5	28	1×10^4
G5	5500	0.8	50	1×10^4
K0	4900	0.4	70	1×10^4
K5	4100	0.17	200	2×10^4
M0	3500	0.06	420	5×10^4
M5	3000	0.01	1250	1×10^5

The spectral class fixes the temperature and the luminosity class fixes the luminosity.

Table 5 Spectral type suffixes

Suffix	Meaning
a, ab, b	Subdivisions of luminosity class (a brightest)
e	Emission lines present
m	Strong metallic lines
n	Very broad (nebulous) lines
p	Peculiar spectrum
s	Sharp lines
v	Variable spectrum

5.2 Abundance Analysis

Spectral classification gives us a good idea of the physical conditions in the photosphere of the star. However, if we wish to determine the chemical abundances of different species in the photosphere, we must use more quantitative techniques. The basic method employed for this determination is the ''curve of growth'' method. For very high precision work this can be supplemented by ''spectral synthesis'', where computed model atmosphere grids with differing abundances and photospheric conditions can be used to generate artificial spectra for comparison with the observed spectrum. However, in many cases the curve of growth method gives a sufficiently good determination and is much quicker to carry out.

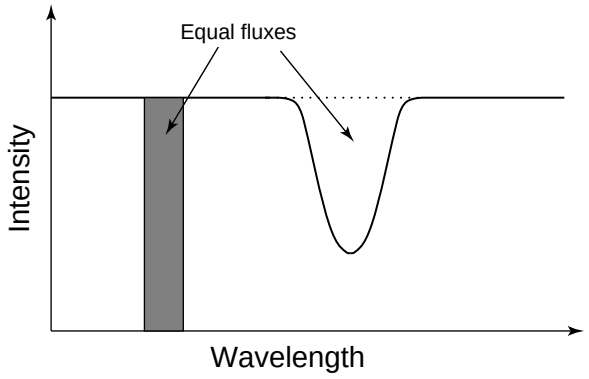

Figure 11 The equivalent width (W) of an absorption line is the width of the continuum containing the equivalent flux (i.e. equal area) to the line.

To apply any quantitative method we must first have a technique for measuring the strength of absorption lines. The standard measure of the strength of a spectral line is the ''equivalent width'' (W). As the name suggests, this defines the strength of a line as equal to the equivalent width of nearby continuum that contains the same total flux as the line (see Figure 11). W therefore has the dimensions of wavelength, and is typically measured in Å or milliångstroms (mÅ). W is independent of spectral resolution and absolute flux calibration, and therefore provides a simple, instrument-independent measure of line strengths.

We must now ask how W varies in proportion to the line absorption coefficient (κ_L), which for a given species will be proportional to the abundance (N, the number density of the species). First we consider the case where $\kappa_L \ll \kappa_C$ (the continuum absorption coefficient). The line width will be governed by the core of the Voigt profile, i.e. by the thermal properties of the atmosphere (and therefore will be independent of N), although the line depth will be proportional to κ_L, and so W will be proportional to $\kappa_L \propto N$. Therefore, for these ''weak'' lines there is a linear relationship between N and W.

Now consider the case where $\kappa_L \approx \kappa_C$. The line width will still be governed by the core of the Voigt profile (i.e. is effectively constant), but many of the photons in the line are being scattered or absorbed, and so the line will appear dark, i.e. there are no photons there – this is termed saturation. Therefore, the line depth is constant and so increasing N has no effect on W and we cannot determine abundances for such lines.

Finally, we consider the case of very strong lines where $\kappa_L \gg \kappa_C$. In this case, while the line depth is still saturated, the damping wings of the Voigt profile start to become significant. It can be shown that the strength of the wings is proportional to $\sqrt{N}$, and therefore W is proportional to $\sqrt{N}$.

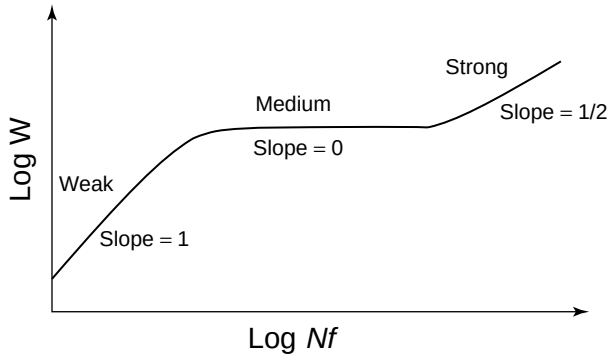

Figure 12 Curve of growth for spectral lines (see section 5.2).

Combining these three cases together, we can produce a ''curve of growth'', showing how W varies with Nf (where f is the line oscillator strength) (see Figure 12). To use the curve of growth to determine the abundance of a species, we plot the observed values of $\log W$ for the various transitions of the species versus $\log f$ for each transition. The resulting observed curve of growth can then be compared with the theoretical curve by sliding it along the $\log Nf$ axis. When the two curves overlap, $\log N$ can be determined by the offset between the two graphs.

We note here that it is traditional in astrophysics to measure abundances relative to the solar hydrogen abundance, which is set at $\log N = 12.00$. For example, on this scale the solar Li abundance is $\log N \approx 0.7$ by number, i.e. a factor of $10^{11.3}$ smaller. In Table 6 the abundances of the elements are given for the Sun.

5.3 Radial Velocity Analysis

Additional information about the nature of a star can often be obtained by measuring the central wavelengths of the lines. Until now we have considered that the line profiles are symmetric about the central wavelength, although we must examine the case when the star is moving relatively with respect to us. In this case we can measure the radial velocity of the star (in a few cases the tangential velocities of stars can be measured directly). For a many-lined spectrum this is best achieved by using a cross-correlation technique to compare with radial velocity standards. The offsets of the lines from the rest wavelengths that are measured can be converted to velocities using the Doppler equation (Equation 16):

$$\frac{v}{c} = \frac{\Delta\lambda}{\lambda_0} \tag{16}$$

where c is the speed of light, $\Delta\lambda$ the Doppler shift and λ_0 the rest wavelength of the line.

These radial velocities can often tell us much about a star. In combination with the object's position in the sky,

Table 6 Abundances of the elements up to iron for the Sun

Atomic number	Element		Log (abundance)	
			Number	Mass
1	Hydrogen	H	12.00	12.00
2	Helium	He	10.93	11.53
3	Lithium	Li	0.7	1.6
4	Beryllium	Be	1.1	2.0
5	Boron	B	<3	<4
6	Carbon	C	8.52	9.60
7	Nitrogen	N	7.96	9.11
8	Oxygen	O	8.82	10.02
9	Fluorine	F	4.6	5.9
10	Neon	Ne	7.92	9.22
11	Sodium	Na	6.25	7.61
12	Magnesium	Mg	7.42	8.81
13	Aluminum	Al	6.39	7.78
14	Silicon	Si	7.52	8.97
15	Phosphorus	P	5.52	7.01
16	Sulfur	S	7.20	8.71
17	Chlorine	Cl	5.6	7.2
18	Argon	Ar	6.8	8.4
19	Potassium	K	4.95	6.54
20	Calcium	Ca	6.30	7.90
21	Scandium	Sc	3.22	4.87
22	Titanium	Ti	5.13	6.81
23	Vanadium	V	4.40	6.11
24	Chromium	Cr	5.85	7.57
25	Manganese	Mn	5.40	7.14
26	Iron	Fe	7.60	9.35

Abundances are expressed as logarithm of both number and mass abundance on a scale for which hydrogen is 12.00. (The data is from C.W. Allen, *Astrophysical Quantities* (see Further Reading list).)

they allow us to make an estimate of the distance to the object (by comparison with the rotational distribution of stars in the galaxy) . This also allows estimates to be made of the age and chemical composition of the object, by comparing it with stars in a similar physical position. Many stars show variable radial velocities. This can indicate that the star is in a binary system, in which case monitoring the velocity variations over an orbit allows an estimate of the ratio of masses of the two stars in the binary. Very precise radial velocity variation measurements (accuracy $\sim 1\,\mathrm{m\,s^{-1}}$) of nearby stars can in fact be used to detect extra-solar planets around some stars. Variable radial velocities may also indicate stellar pulsations in isolated stars as the atmosphere rises and falls – the analysis of the modes of these pulsations (astroseismology) can yield much important information about the interior structure of stars.

6 WAVELENGTH REGIMES

Our discussion has so far concentrated on optical spectroscopy (taken to imply a wavelength range of $\lambda \approx 3000$–$10\,000$ Å). However, it is possible to carry out spectroscopy at wavelengths ranging from the radio to γ-rays. The astrophysical processes occurring at these different wavelengths are often very different from those we have described as being responsible for the optical spectrum of the source, and therefore allow us to probe different regions of the stellar environment. Each wavelength regime also requires different detector technologies and observing techniques. Here, the Earth's atmosphere has been a driving force; for example, at high frequencies it is completely opaque, and so space-borne telescopes have been necessary. This section provides a brief summary of both the techniques used and the physical processes involved for the wavelength regime.

6.1 Radio Spectroscopy (Wavelength 200 μm–100 m)

At wavelengths longer than around 1 cm the Earth's atmosphere is transparent to radio waves. These can be detected using radio telescopes (essentially a parabolic reflecting antenna which focuses incoming radio waves to a small pick-up antenna). The radio waves induce an electrical signal in the pick-up antenna that can be tuned using a resonant circuit and down-converted to a lower (intermediate) frequency (longer wavelength) using a heterodyne mixing technique. The intermediate frequency can then be demodulated and the resulting frequency spectrum digitized by computer. The smallest channel width available is usually ~128 kHz, giving a spectral resolution at 5 GHz of $R \approx 20\,000$.

Below wavelengths of ~2 cm, the Earth's atmosphere starts to become opaque, mainly owing to water vapor, but several ''transparent'' bands are available at wavelengths stretching down to ~350 μm. At wavelengths below a few millimeters, traditional radio engineering techniques are no longer applicable owing to the extremely small size of the microwave circuitry required. Instead, bolometer arrays are employed that detect the heating effect of the incoming radiation. The use of filters allows a rough spectral discrimination, but only with a resolution R of <10.

An important mechanism responsible for radio continuum emission is synchrotron radiation, i.e. that produced by accelerated electrons in a magnetic field. Accelerated electrons are produced in stellar sources such as some interacting binary stars and rapidly rotating remnant cores of supernovae – neutron stars (pulsars). The energy spectrum of the electrons is often characterized by a power law as is the emitted radio continuum spectrum.

Maser action can produce line emission in radio spectra. For low-density regions such as those found in pre-main sequence objects and the outer atmospheres of cool supergiants, the population inversion necessary for maser emission can be created. In these regions the density is sufficiently low that a metastable level may become populated, the pumping mechanism being either

radiation, collisions or even chemical. Molecules that can be excited in this way include OH, water vapor and CO.

Another process responsible for line features in radio spectra is line transitions between closely adjacent energy states. For example, neutral hydrogen (HI) demonstrates the famous 21.1 cm line due to a forbidden hyperfine transition involving the alignment of the electron and proton spins. Such transitions cannot occur in the relatively dense environment near a star owing to collisional de-excitation, but they can be superimposed on a stellar spectrum in the intervening interstellar medium between us and the star. The dynamics of the lines can be measured using their Doppler profiles and shifts, and so we can gain information on the diffuse interstellar gas in the galaxy.

6.2 Far-infrared Spectroscopy (Wavelength 20–200 μm)

Owing to the opacity and temperature of the Earth's atmosphere in this region, satellite-based telescopes are required to explore the far-infrared (FIR) region. The basic optical design of FIR telescopes and instruments is similar to that in optical spectroscopy and detectors are typically based on exotic semiconductor hybrids, such as Ge : Ga and Si : B. However, the effect of the thermal environment of the optics is much more important than that for an optical telescope. Applying Wien's displacement law, we see that blackbody emission peaking at 100 μm corresponds to a temperature of 20–30 K. Therefore, we must cool both the telescope and instrument to temperatures lower than this to prevent emission from the telescope completely dominating over the astronomical source. The recent Infrared Space Observatory (ISO) satellite has been the only infrared (IR) satellite capable of relatively high-resolution spectroscopy, giving resolutions between a few hundred and a few tens of thousands using two grating spectrometers.

The FIR region allows us to see relatively cool material such as dust and cool gas which may be in the process of forming stars, or may be part of a star's ejecta as it dies. Often molecular and atomic features (e.g. CI 157 μm) can be observed in star-forming regions and premain sequence objects.

6.3 Near- and Mid-infrared Spectroscopy (Wavelength 1–20 μm)

Unlike FIR spectroscopy, NIR and mid-infrared (MIR) spectroscopy can be carried out from the ground, although it is best carried out from high, dry sites such as Mauna Kea, Hawaii, or Antarctica. Even at the best sites, however, continuous coverage of the region 1–20 μm is not possible from the ground. Rather, we can observe through a number of spectral windows, centered at 1.2, 1.6, 2.2, 3.5, 5, 10 and 20 μm where the atmosphere is transparent. Telescopes for IR use are designed in order to reduce their thermal emissivity, as a major contributor to the background at wavelengths > 2.5 μm is the thermal emission from the telescope.

Modern spectrographs for IR use are based on the use of cooled optics and gratings, and can be designed for spectral resolutions from a few hundreds to tens of thousands. Detection generally uses array detectors based on hybrid semiconductors such as InSb that must be cooled to around 35 K and have a rapid readout as the fast buildup of thermal charge within the arrays and the very bright sky background mean that exposure times must be kept short (typically <10 s at 2.2 μm, and only tens of milliseconds at 10 μm). Because the sky background can be much brighter in the IR region than the astronomical source, very accurate sky subtraction is vital. This is sometimes achieved by having a telescope secondary mirror that can be chopped on and off source at rates between 0.1 and 10 Hz depending on the observing wavelength. The physical processes giving rise to IR spectral features in stars are similar to those we have discussed for optical spectroscopy, although the dominant opacity sources are often very different. For instance, the optical spectra of M stars are dominated by TiO bands (see Figure 2) whereas in the region 1–2.5 μm the dominant opacity source is water vapor. An important feature in many IR spectra is a ~10 μm silicate feature that is seen in very cool objects and is a signature of dust formation in their atmospheres at temperatures of less than around 2000 K. In some cases an important advantage of NIR spectroscopy over optical is that IR photons suffer much lower interstellar absorption by dust grains than optical photons, and therefore using IR observations we can penetrate into dusty regions such as the center of our galaxy and star forming regions such as those in Taurus and Orion.

6.4 Ultraviolet Spectroscopy (Wavelength 100–3000 Å)

UV photons with wavelengths less than around 3000 Å are blocked by the Earth's atmosphere and are therefore accessible only from satellites. At wavelengths above around 1000 Å, instrument and detector technologies similar to those employed for optical spectroscopy can be used. At wavelengths shorter than around 1000 Å, grazing incidence techniques such as those employed for X-ray telescopes (see below) must be employed.

In addition to normal photospheric lines, a key feature in many UV stellar spectra for stars of up to a few solar masses is strong, coronal emission lines (see section 2.4). These lines are typically of species with very high excitation energies (e.g. CIV 1545, NV 1238 and SiIV 1393 Å). Such lines are indicative of very high

temperatures up to 100 000 K, and the spectral energy distribution of the continuum emission indicates that the coronal temperature reaches over 10^6 K. For higher-mass stars exactly the same lines are present, indicative of ionization states higher than expected from the photospheric temperatures. These lines display very large Doppler shifts ($>10^3$ km s^{-1}) and have been successfully explained as a ''wind'' flowing from the star. The wind is driven by the radiation pressure from the large flux of high-energy UV photons being absorbed by line transitions in the gas.

6.5 High-energy Spectroscopy (Wavelength < 100 Å)

High-energy photons are usually labelled with their energy as opposed to their wavelength. The photon energy in electronvolts (eV) is related to its wavelength in meters by Equation (17):

$$E = \frac{1.25 \times 10^{-6}}{\lambda} \tag{17}$$

There are two methods of differentiating an X-ray from a γ-ray: the first merely considers the energy (E) of the photons associated with the radiation. A simple definition might therefore be that photons of energy 0.1–100 keV are X-rays and higher-energy photons are γ-rays. This is somewhat arbitrary, however, and an alternative is to consider the *origin* of the radiation. By this definition, only nuclear processes give rise to γ-rays, whereas X-rays originate from radiation processes involving electrons.

X- and γ-radiation is blocked by the Earth's atmosphere and, as for UV radiation, is accessible only via satellite experiments. A number of different technologies can be used to detect high-energy radiation. Owing to their penetrating nature, X-rays are naturally hard to focus, and grazing incidence techniques must be used to construct a telescope. The most commonly used detector for X-rays is the X-ray CCD, although position-sensitive gas proportional counters are also employed. Owing to the low count rates and large energy per photon (generating many photoelectrons per photon), X-ray CCDs can be read out continuously, allowing them to tag each arriving photon with an energy. It is therefore possible without a dispersive element to obtain some spectral resolution, typically $R \approx 10$–100. To obtain higher resolutions a reflection grating spectrometer must be employed. For γ-rays, detectors may be based on the photoelectric effect, Compton scattering or pair production. Typical detectors achieve $R \approx 10$ using a similar photon tagging mechanism to X-ray CCDs. Solid-state detectors based on high-purity germanium can increase the resolution to $R \approx 300$. The principal processes that give rise to high-energy emission from stellar sources are (a) thermal emission from stellar coronae (see section 2.4), (b) accretion of stellar material on to a compact remnant of stellar evolution and (c) nuclear processes in supernovae (exploding massive stars). Stellar coronae have already been discussed and so we shall give a brief discussion of processes (b) and (c).

Accretion models are now accepted to explain the general properties of most strong stellar X-ray sources. These are considered to be binary stars (i.e. two objects orbiting around their common center of mass) in which one of the objects is a compact object produced at the end of stellar evolution (i.e. a white dwarf, a neutron star or a black hole) on to which mass is being transferred from the ''normal'' star. Accretion is a very efficient process at releasing gravitational potential energy. (Neutron star accretion will typically produce $\sim 10^{16}$ J kg^{-1} of accreted material, whereas nuclear fusion of 1 kg of hydrogen to helium yields only $\sim 5 \times 10^{14}$ J.) For neutron stars we expect photon energies in the range 1 keV–50 MeV, and observations show the resulting spectral energy distribution is generally well characterized as bremsstrahlung.

Spectral lines in accreting sources are generally due to one of two mechanisms. Lines such as the iron Kα line, which is often visible at around 6.7 keV, are produced by fluorescence as harder X-rays illuminate the in-falling material. Analysis of the shape and wavelength of such lines for relativistic distortions allows a determination of the geometry of the system and the nature of the central object. Second, cyclotron lines may be visible, indicating the presence of a strong magnetic field. This allows a determination of the magnetic field strength from the cyclotron line frequency (gyrofrequency) f_c (Equation 18):

$$f_c = \frac{eB}{2\pi\gamma m_e} \tag{18}$$

where e is the electron charge, B the magnetic field density, m_e the electron mass and γ the relativistic gamma factor.

At the end of their lives, stars with masses larger than about 10 solar masses explode in a supernova, typically releasing about 10^{42} J in a couple of hundred days. We might expect them to be strong emitters of X- and γ-radiation owing to their extremely high luminosities and temperatures. However, supernovae are optically thick to high-energy emissions and the high-energy radiation is degraded to lower energies before the expansion of the supernova shell allows them to escape. The principal source of high-energy emission from young supernovae is from the radioactive decay from elements produced in the explosion. For example, 847 keV flux was detected from the supernova 1987A which was caused by the decay of Co^{56}, itself produced by the decay of Ni^{56}. The intensity and shape of the γ-ray line can be used to determine the mass of Ni^{56} synthesized in the explosion.

7 CONCLUSIONS – WHAT HAVE WE LEARNT?

We have found that stellar spectroscopy can be used to calculate the physical conditions of the atmospheric regions of the star. However, we have also seen that interpretation of spectra must be done in the framework of stellar structure and evolution. We have shown that stellar spectroscopy can provide more than temperatures and pressures – large- and small-scale velocity fields, ages of stars, circumstellar material and aspects of stellar structure may be detected and measured. There are some significant problems in obtaining spectra, from the intrinsic limiting resolution of the spectrograph to the problems of using satellite observatories for some wavelength regimes. As we cannot perform experiments on any star, then the study of stars in all wavebands is vital.

Using spectra, the theory of stellar evolution has been developed during the twentieth century and, although there are still unsolved problems, has largely been confirmed. It is fitting that the oldest branch of spectroscopy has been so successful as a tool for studying stars.

ACKNOWLEDGMENTS

The data plotted in Figure 1 comes from the Gliese catalogue (the nearest stars) and the Yale Bright Star Catalogue (the brightest stars). Dr Andy Newsam and Dr Tom Marsh are thanked for help in producing Figure 9. The model atmospheres spectra in Figure 5 were calculated by Dr R. Kurucz.

Dr Phil James is thanked for reading and making comments on the draft of this article.

ABBREVIATIONS AND ACRONYMS

CCD	Charged-coupled Device
FIR	Far-infrared
IR	Infrared
ISO	Infrared Space Observatory
MIR	Mid-infrared
MK	Morgan and Keenan
NIR	Near-infrared
UV	Ultraviolet

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Hyperspectral Remote Sensing: Data Collection and Exploitation • Imaging Spectrometry for Geological Applications

Atomic Spectroscopy **(Volume 11)**
Glow Discharge Optical Spectroscopy and Mass Spectrometry

Electronic Absorption and Luminescence **(Volume 12)**
Near-infrared Absorption/Luminescence Measurements

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview

FURTHER READING

There are many texts regarding the emission and analysis of stellar spectra. Here we list a few particularly relevant texts for the interested student.

1. E. Böhm-Vitense, *Introduction to Stellar Astrophysics*, Cambridge University press, Cambridge, Vols. 1–3, 1989. These three books discuss different aspects of stellar astronomy – Vol. 2 (*Stellar Atmospheres*) is a good introduction, Vol. 1 concerns *Basic Stellar Observations and Data* and Vol. 3 discusses *Stellar Structure and Evolution*.
2. D.F. Gray, *The Observation and Analysis of Stellar Photospheres*, Cambridge University Press, Cambridge, 1992. A more advanced text on atmospheres.
3. C.R. Kitchen, *Astrophysical Techniques*, Adam Hilger, Bristol, 1992. An excellent introduction to observing techniques and instruments.
4. C.W. Allen, *Astrophysical Quantities*, Athlone Press, London, 1976. A handbook containing a vast amount of data relevant to astronomy, including stellar parameters and typical emissions, oscillator strengths for lines and collision strengths.

Temperate Forest Resource Assessment by Remote Sensing

F. Mark Danson
Telford Institute of Environmental Systems, University of Salford, Manchester, UK

1 Introduction 1
2 Background 2
3 Mapping Forest Resources 3
4 Optical Remote Sensing of Forest Biophysical Properties 3
4.1 Estimating Forest Biophysical Properties 3
4.2 Estimating Forest Biochemical Properties 4
4.3 Estimating Forest Stand Properties 6
5 Microwave Remote Sensing of Forest Biophysical Properties 6
6 Remote Sensing Inputs to Forest Ecosystem Models 7
7 Developments in Forest Resource Assessment by Remote Sensing 8
8 Conclusions 9
Abbreviations and Acronyms 9
Related Articles 10
References 10

Remote sensing from airborne and spaceborne platforms provides a means of mapping forest resources at a range of spatial scales. Local-scale mapping using aerial photography is a well established technique for determining stocking density, forest cover and forest health over relatively small areas. Satellite remote sensing can be used to map forest resources over larger areas, including regional to global scales. Techniques have been developed to map forest species and forest cover using spectral and spatial information in remotely sensed images. This work has been extended to map key biophysical and biochemical properties such as leaf area index (LAI) and leaf nitrogen content. These data have been coupled with other information such as meteorological data to drive process-based ecosystem simulation models which may be used to monitor the production of forests and the rates of exchanges of energy and matter with the atmosphere and hydrosphere. Forests play a key role as a source of wood products but they are increasingly also being seen as playing a key role in the global climate system. Remote sensing is the only technique available to monitor forest resources at local, regional and global scales, and to develop an understanding of their role in the global ecosystem.

1 INTRODUCTION

The world's temperate forests occur as a broad belt across the Northern hemisphere between approximately 40 and 70° north with no equivalent biome in the southern hemisphere. The region is dominated by the evergreen coniferous Boreal forest which extends northwards to the arctic tundra and southwards into temperate deciduous and mixed forest and, in drier areas, into steppe forest and the grasslands of central Eurasia and North America.[1] The structure and composition of the forests vary greatly with climate and site conditions but within the region, the moist forests of northwestern North America have the highest biomass productivity of any forest area on Earth.[2] From an economic point of view the conifer forests are important as a timber resource both for saw wood and for pulp wood. The mixed deciduous forests have historically also provided a valuable resource for building and for fuel, but their present distribution has been greatly affected by human activity, particularly in Western Europe and Asia. Continuous forest cover in these regions is generally restricted to areas which are not suitable for agriculture and, where forest cover is present, it is invariably secondary regrowth on areas which have previously been cleared.

Recent advances in modeling the fluxes of energy and nutrients in forest ecosystems have highlighted the importance of temperate forest ecosystems in the global climate system and in particular their role in the global carbon cycle. The Boreal forest covers some $15 \times 10^6\,km^2$, represents approximately 21% of the forested land surface of the Earth and accounts for 13% of the carbon stored in the biomass and 43% of the carbon stored in the soil.[3] These forests therefore represent an important terrestrial carbon "sink" and research has indicated that climatic warming may lead to significant changes in the carbon cycle and ecological functioning of the Boreal forest with feedbacks impacting on the global climate system.[4] There are still great uncertainties about the size of the energy and carbon fluxes between the biosphere and the atmosphere but the analysis of time-series satellite imagery is beginning to provide relevant, reliable and spatially comprehensive data.[5]

Information on temperate forest resources is required at a range of spatial and temporal scales from local

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

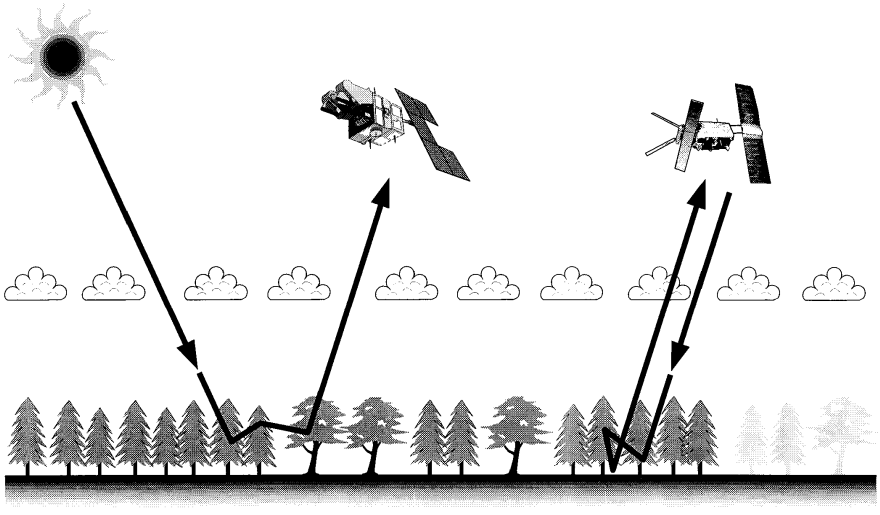

Sensor system:	Optical (passive)	SAR (active)
Wavelength:	400–3000 nm	1–100 cm
Scattering medium:	Leaf surface/interior	Branches, trunks, ground
Absorbing medium:	Chloroplasts, cell walls and cell water	Negligible absorption

Figure 1 Interaction mechanisms for forest canopies.

forest inventories used for resource management purposes and updated annually, to global data on carbon, water and energy fluxes required for modeling climate change over a number of decades. Remote sensing instruments flown on aircraft and on satellites play a crucial role in providing the data to derive relevant forest resource information at all these scales.

2 BACKGROUND

Like all applications of remote sensing, the measurement of forest resources relies on the interaction of electromagnetic radiation with the surface and interpretation of the reflected (or emitted) signal. Optical sensor systems measure reflected radiation in one or more discrete wavebands located in the spectral range 400–3000 nm, whereas active synthetic aperture radar (SAR) systems measure backscattered microwave radiation at wavelengths between 1 and 1000 cm. Optical wavelengths are several orders of magnitude smaller than the leaves, needles and branches that make up a forest canopy and consequently radiation may be both absorbed and scattered by these components. In the case of the longer microwave wavelengths, scattering from leaves, branches, trunks and the ground is the dominant mechanism (Figure 1). It follows that optical remote sensing systems may provide information on the amount of foliage and its biochemical properties whereas microwave systems provide information on woody biomass and forest structure. In addition to potential information in the spectral reflectance signal, remote sensing instruments are imaging devices which may provide additional spatial information related to the three-dimensional structure of the canopy and the spatial resolution of the imaging sensor.[6]

The problem of extracting information on forest resources from remotely sensed data may be reduced to one of relating spectral, spatial and temporal variation in the data to variation in the biophysical and structural properties of the forest canopies. There is no direct physical relationship between the remotely sensed signal and the biophysical characteristics of the forest canopy, but the physical processes of scattering and absorption are controlled by the canopy characteristics. Extracting information on forest resources therefore depends on developing techniques to infer this information from the remotely sensed data.[7] The inference process may involve qualitative interpretation of imagery by a trained operator, the development of statistically based calibration equations or the use of mathematical modeling techniques.

3 MAPPING FOREST RESOURCES

Local-scale mapping of forest resources using aerial photography and ground-based field surveys remains important today, even with the availability of 1-m spatial resolution data from space.[8] Manual air-photograph interpretation is required in applications which require the direct identification of tree species or an inventory based on surveying individual trees. Skilled air-photograph interpreters use qualitative properties of aerial photography, such as shape, pattern, tone and texture, to produce stock maps and forest inventory information, usually at scales of 1 : 25 000 to 1 : 5000.[9] This work is generally coupled with surveys at ground level to measure directly wood volume, basal area and tree heights. Forest resource assessment by satellite remote sensing is unlikely to replace air-photograph interpretation at scales above 1 : 25 000, but where applications require maps of larger areas, the use of satellite images may enable the production of a cheaper and more consistent product.[10]

The advantages of satellite sensor data for mapping forest areas have been recognized since the launch of the first civilian high-spatial resolution sensor, Landsat-1, in 1972. Thematic maps may be produced to show forest species, forest cover or forest condition. The procedure relies on the assumption that a given class (e.g. broadleaf forest) has a distinct and consistent spectral response or spectral signature. Most studies to date have used ''traditional'' supervised multispectral image classification techniques where areas of known class, or training areas, are defined a priori in an image and their spectral response quantified in terms of the reflectance mean and covariance between wavebands.[11] Each pixel in the image is then classified according to its statistical probability of membership to a given class. Many studies have shown that this approach to mapping forest areas can be successful at a range of spatial scales from local, to regional and global.[12–14] A weakness, however, is that the spatial information in the image data is not incorporated in the classification procedure and recent work has addressed this problem with the development of contextual classifiers which use both spectral and spatial characteristics of the imagery and evidential reasoning which allows data from different sources and at different scales to be incorporated.[15–17]

The application of an artificial neural network (ANN) to remotely sensed land cover classification has also been an important recent development.[18] An ANN involves the use of artificial intelligence techniques and provides a powerful tool to discriminate patterns in many different types of data. ANNs are particularly useful in classifying remotely sensed data because, unlike the traditional statistically based methods, they make no assumptions about the statistical distribution of the input data, can handle noisy data and can accommodate ancillary data from other sources.[19] A number of studies have shown that land cover classification accuracy based on a neural network approach is higher than that based on statistical classifiers, particularly when the training data sets are small.[20,21] Studies which have used ANNs to classify forest areas have reached similar conclusions.[16,22]

4 OPTICAL REMOTE SENSING OF FOREST BIOPHYSICAL PROPERTIES

Although image classification methods may produce useful information on the location and spatial extent of forests, and descriptive information on forest type or species, they do not provide quantitative information about forest biophysical or structural properties. In fact, the spectral properties of all vegetation canopies is a function of these properties and information on canopy biophysical properties should therefore be accessible from remotely sensed measurements. Variables such as LAI, the one-sided area of leaves per unit ground area, and leaf biochemistry affect forest function in terms of light interception and absorption, nutrient cycling and productivity.[23] They vary both spatially and temporally across forest areas and are difficult and expensive to measure. They are, however, the key spatial variables required to drive forest ecosystem simulation models at a range of spatial scales and so considerable effort has been expended in developing remote sensing techniques to map these variables over extensive forest areas.

4.1 Estimating Forest Biophysical Properties

Optical remote sensing has focused on the estimation of forest LAI or biomass and has built on earlier successful work to estimate these variables for agricultural crops and grassland canopies.[24,25] At visible wavelengths light absorption is the dominant process and in general as LAI or biomass increase, visible reflectance decreases so that a negative asymptotic relationship between red reflectance (R) (600–700 nm) and LAI is expected. At near-infrared (NIR) wavelengths (700–900 nm) leaf absorption is low and leaf reflectance and transmittance are high. A positive asymptotic relationship is therefore expected between NIR reflectance and LAI or biomass.[26] The spectral contrast between R and NIR reflectance has been used to develop ''vegetation indices'', which are linear or non-linear combinations of the reflectance in two or more

wavebands and a number of studies have attempted to correlate vegetation indices with forest LAI.

Early studies by Peterson et al.[27] and Spanner et al.[28] examined the relationships between the LAI and reflectance of forest sites across the western United States. They found a significant negative relationship with R and a significant positive relationship with the simple ratio (SR) vegetation index (NIR/R) or normalized difference vegetation index (NDVI), but no significant relationship with NIR reflectance. Other studies have also found only weak correlations between forest LAI and the NDVI[29,30] and it has been suggested that the spatially variable contribution of soil or understorey to the scene reflectance, coupled with variations in canopy structure and leaf optical properties, may confound the application of vegetation indices for estimating forest LAI.[31] For example, Badwhar et al.[32] found that the relationships between the LAI of aspen (*Populus tremula*) stands in Minnesota and R and NIR reflectance were seasonally variable and in August, when there was full canopy cover and a well-developed understorey vegetation layer, no significant relationships were found. Curran et al.[33] found a positive relationship between the LAI of slash pine (*Pinus elliottii*) stands in Florida and the NDVI, although again the strength of the relationship varied seasonally because of changes in understorey LAI. Chen and Cihlar[34] found that the relationship between the SR vegetation index derived from Landsat thematic mapper (TM) data and the LAI of a Boreal conifer forest was strongest in the late spring when the understorey vegetation was not fully developed. This work was followed up by a detailed study which examined the seasonal variation in the understorey reflectance of the stands and its impact on vegetation indices.[35] In contrast to the findings of earlier studies, Fassnacht et al.[36] found a strong positive relationship between LAI and NIR reflectance in Wisconsin. These studies together indicate that the strength of the relationships between forest LAI and vegetation indices such as the NDVI may be site, time and species specific and that above an LAI of about 5–6 the NDVI may not be sensitive to LAI variations. A number of authors have also suggested that the expected positive relationship between LAI and NIR reflectance may break down at higher levels of LAI because increased inter-canopy shadowing causes a decrease in NIR reflectance.[17,26]

Recent research has sought to improve the estimation of forest LAI by the application of new vegetation indices based on either broad wavebands or high spectral resolution data from field spectrometers or airborne imaging spectrometers (AIS). Nemani et al.[37] found that incorporating a middle-infrared (MIR) waveband with R and NIR normalized the effect of variable canopy cover on the relationship with LAI. The approach worked well at one test site but has yet to be evaluated for other forest environments. Foody et al.[38] explored a number of different vegetation indices for mapping tropical forest biophysical properties and highlighted the potential of data in the 3–5 μm region of the MIR band which exhibited stronger correlations with the stand variables than the NDVI. High spectral resolution data, where measurements of surface radiance are made in several hundred narrow wavebands, have also been used to estimate forest LAI. Several experiments have shown that calculation of first or second derivatives of the canopy reflectance may suppress the effects of variation in understorey reflectance and allow more accurate estimation of LAI.[39,40] High spectral resolution data may also be used to calculate the position of the ''red edge'', the point of inflection in the reflectance spectrum round 720 nm. Correlations between the red-edge position and forest LAI have been found to be stronger than those with single wavebands or vegetation indices based on broad wavebands.[30,41]

Significant advances have been made in estimating forest LAI from remotely sensed data and these have been partly facilitated by the use of new techniques and instruments for measuring the LAI of forest stands using light interception. This has allowed the traditional and expensive destructive sampling techniques to be complemented by techniques which allow rapid measurement of LAI at a large number of sites.[42] Furthermore, large multidisciplinary experiments such as Oregon Transect Ecosystem Research (OTTER) and the Boreal Ecosystem and Atmosphere Study (BOREAS) have provided comprehensive ground and remotely sensed data sets with which to develop and test LAI estimation techniques.[43,44] Further developments are likely to involve a greater emphasis on the application of radiative transfer models of forest canopy reflectance to investigate how differences in leaf optical properties and forest structure and composition affect the relationships between forest LAI and spectral reflectance.[45]

4.2 Estimating Forest Biochemical Properties

Chlorophyll and water are the main chemicals in vegetation responsible for absorbing incoming radiation from the Sun. The characteristic spectral response of vegetation canopies is dominated by the broad absorption features of these two chemicals and it follows that canopy reflectance is closely related to their concentrations in the canopy. A number of studies have used remote sensing to estimate the chlorophyll and water content of forest canopies in investigations of forest damage caused by environmental stress and for inputs to ecological models. Riggs and Running[46] attempted to detect water stress in conifer canopies using high spectral resolution data. However, they found that variation in canopy cover

between stands caused a larger variation in canopy reflectance than differences in water status. Dawson et al.[45] related the ratio of reflectance at 970 nm to that at 900 nm, termed the water index (WI),[47] to the water content of Boreal forest stands and found a weak negative correlation. They suggested that a priori knowledge of forest biophysical variables such as LAI may be necessary to determine accurately the canopy water content from remotely sensed data. Similar problems have been encountered in studies that have attempted to measure forest canopy chlorophyll content using remotely sensed data.[45,48] The effects of variation in canopy cover and/or LAI on the relationships between forest canopy chlorophyll and water content is a problematic area. However, Fourty and Baret[49] have shown that if the canopy cover is high or can be considered constant, retrieval of these biochemical properties should be possible.

Pioneering work in the late 1980s began to explore the use of remote sensing for determining the relationships between other forest canopy biochemicals and rates and patterns of biogeochemical cycling (BGC). Ecologists had established the importance of foliar nitrogen content of forest vegetation which in conjunction with LAI is closely related to photosynthetic capacity and nitrogen uptake.[50] In addition, the ratio of leaf lignin to leaf nitrogen has been shown to be related to rates of litter decomposition.[51] Measurement of these and other biochemicals over large areas of forest was therefore seen as a key step toward mapping the spatial characteristics of forest nutrient cycles.[52]

The techniques of laboratory-based NIR reflectance spectroscopy were developed by analytical chemists in the 1950s for measuring the crude protein, dry matter, lignin and other biochemicals in dried ground vegetation samples. However, the application of these techniques in ecological remote sensing became possible only with the development of a new generation of field spectroradiometers and imaging spectrometers which allowed the measurement of surface reflectance in a large number of narrow wavebands. The leaf biochemicals of interest have fundamental absorption features in the 2500–15 000 nm region but also have overtone and combination absorption features in the 1100–2500 nm shortwave infrared reflectance (SWIR) region measured by many remote sensing instruments. The challenge for ecological remote sensing was to extend the laboratory techniques from analysis of dried ground samples to fresh leaves, whole branches and complete canopies.

Experiments on dried ground samples confirmed that leaf nitrogen and lignin content could be estimated using laboratory instruments. Wessman et al.[53] collected samples of leaves and leaf litter from both deciduous and coniferous species. The samples were dried and ground and their reflectance measured using a laboratory spectrometer. Multiple linear regression was used to define calibration relationships between the first and second derivatives of reflectance and nitrogen and lignin content. The biochemicals were estimated with standard errors of 0.11 and 2.9%, respectively. Other studies provided similar results and also showed that the wavelengths selected by stepwise multiple regression techniques were close to known absorption features for a given leaf biochemical.[52,54]

Extension of this work to fresh foliage further indicated the potential of the technique for estimating the biochemistry of intact leaves.[52] Although the presence of water in the leaves may suppress the biochemical signal because of water absorption in the SWIR region, several studies have shown that leaf nitrogen, protein, cellulose, lignin, chlorophyll and starch content may be estimated from reflectance measurements.[55–57] It does appear, however, that the wavebands selected in stepwise regression for fresh leaves are not always clearly related to known absorption features for the biochemicals. Grossman et al.[58] showed that band selection depended on whether the chemical data were expressed in terms of leaf concentration (grams per gram) or as leaf content (grams per square meter) and was also sensitive to the set of calibration samples used. They also highlighted the problem of overfitting where high coefficients of multiple determination (r^2) were obtained for datasets produced by randomization of the measured data.

The results of the work on estimating leaf biochemistry stimulated experiments on estimating the biochemistry of complete forest canopies using data from imaging spectrometers. The first experiments were conducted using the AIS, which recorded a 32×32 pixel image with 128 wavebands of approximately 9 nm bandwidth over a range of 855–2036 nm (tree mode) or 1200–2400 nm (rock mode).[59] Wessman et al.[60] conducted an experiment using AIS data of a mixed species forest in Wisconsin. They established a regression equation, based on three wavelengths, between the first-difference reflectance spectra and the lignin content of 18 stands ($r^2 = 0.85$). They then used a previously established inverse relationship between nitrogen mineralization (kilograms per hectare per year) and lignin to produce the first map of annual nitrogen mineralization for a forest area.

Further work using imaging spectrometers has been conducted in two major experiments, the OTTER project[43] and the National Aeronautics and Space Administration (NASA)-funded Accelerated Canopy Chemistry Program (ACCP).[61] These involved the use of the airborne visible infrared imaging spectrometer (AVIRIS) which has 224 wavebands in the range 410–2450 nm and a higher signal-to-noise ratio (SNR)

than the earlier AIS. In the OTTER project, AVIRIS data were acquired in 1990 and 1991 over six study sites along a 250-km east–west transect representing forest vegetation zones from coastal rainforest to semiarid scrub.[40,62] Strong correlations between the AVIRIS first difference spectra and total nitrogen, chlorophyll and lignin measured both as canopy biochemical content (kilograms per hectare) and foliar concentration (milligrams per square centimeter). Correlations for starch were weaker for both concentration and content.

In the ACCP, AVIRIS data were acquired over a number of forest test sites in 1992 and 1993. Martin and Aber[63] used data for Blackhawk Island, Wisconsin and Harvard Forest, Massachusetts, to examine spatial and temporal variation in canopy lignin and nitrogen concentrations for a total of 40 stands. They found strong correlations between first difference spectra and biochemical concentrations for both sites separately and for data combined from both sites. However, when the calibration equation derived for one site was used to predict biochemical concentrations at the other, the relationships were not significant. Errors in atmospheric correction and low SNR were suggested as possible causes for the inconsistency. Curran et al.[64,65] also used AVIRIS data to estimate the biochemical composition of a slash pine canopy in Florida. They found very strong correlations between the remotely sensed data and chlorophyll, nitrogen, lignin and cellulose measured both as concentration and content.

Remote sensing is the only technique available for estimating forest biochemical properties over large areas. The research outlined above shows the great potential of the technique for providing biochemical information related to forest productivity and health and for input into process-based ecosystem models. Research to date has been limited by the availability of high spectral resolution image data with which to develop and test the techniques. The recent availability of data from the spaceborne moderate resolution imaging spectrometer (MODIS) and medium resolution imaging spectrometer (MERIS) and advances in radiative transfer modeling of leaf and canopy reflectance should ensure further progress.[64]

4.3 Estimating Forest Stand Properties

Much of the research to develop techniques for estimating the biophysical and biochemical properties of forest canopies has been driven by the needs of forest ecologists who require the data to drive ecosystem simulation models as discussed in section 6. However, commercial forest managers may require local data on variables such as stand age, average tree height, basal area or timber volume.[66,67] These variables may have no direct physical relationship with the remotely sensed signal, but they may be correlated with it through indirect relationships with LAI or canopy cover.[68] A large number of studies have sought relationships between forest stand variables and a range of remotely sensed data with mixed success. Strong negative relationships have been found between R and NIR reflectance satellite sensor data and wood volume or basal area in a range of forest environments[69–72] while other studies have found strong correlations with forest stand age and tree density.[68,73,74] Relationships established by correlation and regression techniques depend mainly on the existence of a consistent monotonic change in reflectance as the stand develops. Such relationships may exist up to a point where the forest stand reaches full canopy cover but not thereafter.[75] Furthermore, spatial variations in canopy cover or differences in terrain slope may reduce the strength of the empirical relations.[76]

5 MICROWAVE REMOTE SENSING OF FOREST BIOPHYSICAL PROPERTIES

SAR instruments operate at microwave wavelengths between about 1 cm and 1 m and provide image data which are sensitive to the macro structure of a forest canopy. Microwave signals, in contrast to optical wavelengths, are unaffected by atmospheric conditions, including clouds, because the wavelength of the radiation is several orders of magnitude larger than the atmospheric particles.[77,78] This means that SAR instruments have an all-weather day/night imaging capability, which is a major advantage for imaging the cloudy temperate regions of the Earth. The processing and analysis of SAR data are relatively complex with the magnitude of the backscattered radiation dependent on both wavelength and polarization. However, in general terms longer wavelengths penetrate deeper into the canopy, interact with the tree stems and ground and are sensitive to the total woody biomass whereas shorter wavelengths are scattered within the canopy and may be sensitive to canopy leaf area. SAR may transmit and receive radiation in either vertical (V) or horizontal (H) planes or polarizations and several studies have shown that the microwave response to forest biomass is also dependent on polarization.

Recent research on the microwave response of forest canopies has benefited from the availability of data from spaceborne instruments, in particular the C-band (3 cm) active microwave instrument (AMI) on ERS-1 and ERS-2, the L-band (23 cm) SAR on JERS-1 and more complex data from the Shuttle

Imaging Radar C (SIR-C) flown on the Space Shuttle Endeavour in 1994 (C and L band). In addition, sophisticated airborne SAR instruments such as the C-, L- and P-band (50 cm) airborne synthetic aperture radar (AirSAR) have provided an opportunity to test the application of multifrequency polarimetric data.

Many studies have found positive nonlinear relationships between microwave backscatter and forest above ground biomass, with saturation of the signal occurring at different levels depending on the wavelength measured.[79] For example, Ranson et al.[80] showed that SIR-C data could be used to map above ground woody biomass of Boreal forests to within 16 t ha^{-1} with sensitivity up to an upper limit of 150 t ha^{-1}. Similar wavelength-dependent upper limits to biomass estimation have been found by other authors, including Imhoff,[81] who found that the C band was sensitive to forest biomass up to 20 t ha^{-1}, L band up to 40 t ha^{-1} and P band up to 200 t ha^{-1}. Other studies using AirSAR data have shown that cross-polarized data are more sensitive to forest biomass than like-polarized data (Figure 2).[78,82]

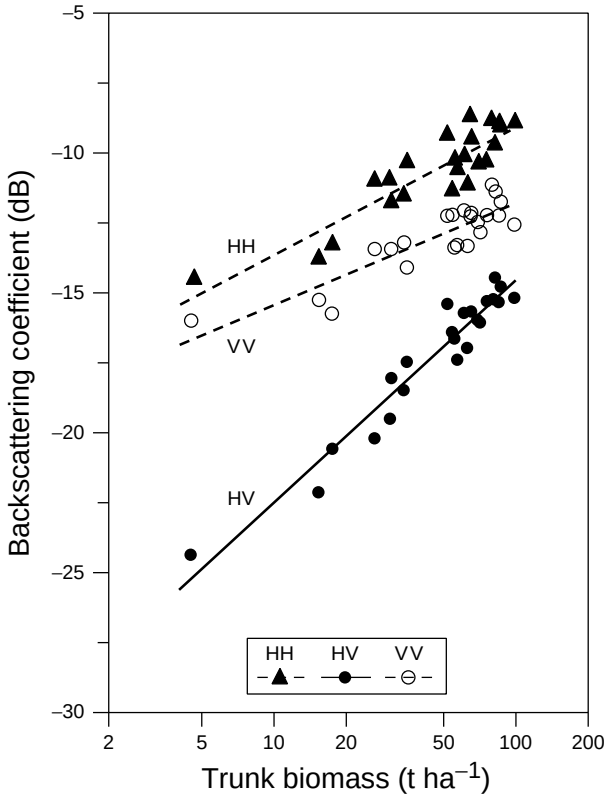

Figure 2 Relationships between SAR data and forest volume (P band). (Reproduced by permission from le Toan et al.[82] © 1992 IEEE.)

In addition to the work on estimating forest biomass, there have been many investigations of the application of SAR data for mapping forest areas. Unlike classification of data from optical remote sensing systems, SAR classification of forest cover types exploits differences in macro structure between stands of different species or different age or density.[83,84] Other advances have been made with nonimaging radar sensors such as the use of dual-frequency ranging scatterometers to estimate forest stand structural properties.[85,86]

6 REMOTE SENSING INPUTS TO FOREST ECOSYSTEM MODELS

Recent research using global climate models (GCMs) has suggested that forest ecosystems may be very sensitive to increases in atmospheric CO_2 concentrations through physiological and structural responses, with important feedbacks to global climate.[87] Climatic simulations and some recent measurements point to large temperature increases in the northern high latitudes where temperate forests are the main landcover.[88–90] Intimately related to these climatic effects is the nature of the global carbon cycle. Seasonal variations in atmospheric CO_2 have been related to size of the terrestrial carbon store and suggest that the northern forests represent a net ''sink'' for atmospheric carbon of 1–2 Gt yr^{-1} due to an increase in the length of the growing season at higher latitudes.[90,91] The spatial extent of the temperate forest biome, its relationship to climate and its importance as a carbon store therefore make it important to understand its function and to model the key ecosystem processes. This is a difficult task, however, because of the complex interactions between the forest and atmosphere (Figure 3) and the nature of the feedback mechanisms between the climate system and the biophysical, biogeochemical and structural characteristics of the forests.[44,92,93] Detailed point measurements of these processes have been made using instrumented towers at forest test sites,[94,95] but extrapolating these measurements up to regional scales (10–1000 km) requires the data on forest biophysical and biochemical properties derived from remotely sensed data using the techniques described earlier.[44]

Forest ecosystem models describe the processes of photosynthesis and evapotranspiration which control the exchanges of energy and matter with the vegetation and they may be used to estimate the rates of cycling of carbon, water and nitrogen through the ecosystem in addition to net primary productivity (NPP).[96,97] The Forest BGC model of Running and Coughlan[98] was designed specifically to use remotely sensed input

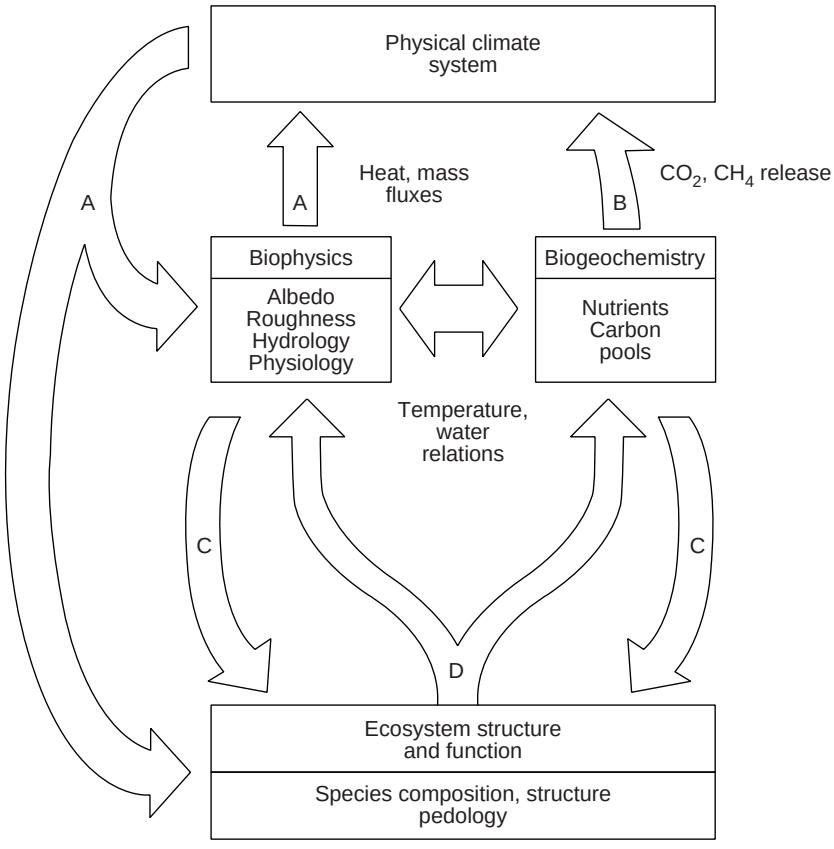

Figure 3 Interactions between the Boreal forest and atmosphere. (A) influence of changes in physical climate on biophysical processes; (B) influence of changes in nutrient cycling rates; (C) influence of changes in biogeochemical processes on forest species composition and structure; (D) influence of changes in forest structure on biophysical characteristics and biogeochemical process rates. (Reproduced from Sellers et al.[44] with permission from the *Journal of Geophysical Research*.)

data to determine LAI and a range of meteorological variables. At a local scale, Forest BGC was used to estimate the carbon accumulation in a Sitka spruce (*Picea sitchensis*) plantation in south Wales, UK.[99] High spectral resolution image data from the compact airborne imaging spectrometer (CASI) were used to relate the reflectance red edge to stand LAI and leaf nitrogen content and to drive the model. Estimates of stem carbon production from the model were comparable to estimates derived from tree cores (root mean square error = $0.34\,\mathrm{Mg\,C\,ha^{-1}\,yr^{-1}}$). Running et al.[96] used NDVI data from the National Oceanographic and Atmospheric Administration (NOAA) advanced very high resolution radiometer (AVHRR) to estimate LAI for a 1200-km^2 forested area in northwestern Montana and coupled this with daily meteorological data to estimate regional evapotranspiration and photosynthesis using Forest BGC. Estimates of aboveground net primary production and stem volume growth for the sites across the region were in the same range as those from earlier local-scale studies. Validation of the results for the 1.1-km pixel area of the AVHRR instrument was not possible, however, and the sensitivity of the model to the remotely sensed LAI estimates was highlighted as an area which required further investigation.

Liu et al.[100] used a land cover classification from TM data to stratify a Boreal forest area into different forest types before estimating LAI for input into a forest productivity model. A similar approach was used by the same group[101] to produce a Canada-wide NPP map using remotely sensed data, daily meteorological data and information on soil texture. All of this work shows the considerable potential for remote sensing inputs to forest ecosystem models; as techniques for estimating the model variables become more sophisticated and more accurate, and techniques for assimilating other data are developed, further advances are likely.

7 DEVELOPMENTS IN FOREST RESOURCE ASSESSMENT BY REMOTE SENSING

Assessment of forest resources by remote sensing depends on the existence of relationships between the

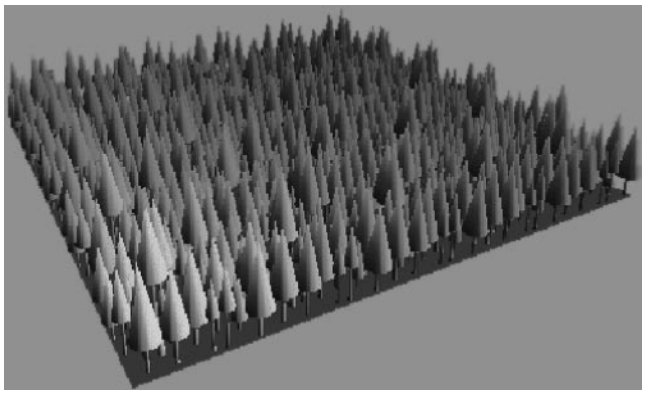

Figure 4 Simulation of forest canopy reflectance using the Flight model of North.[103]

forest spectral signature and the variable of interest, e.g. canopy nitrogen concentration. Much of the work done to date has sought statistical relationships between the spectral variables and the forest variables, often with some success. A problem arises, however, when no statistical relationship can be found. Recent work has started to address this problem by developing radiative transfer models that describe the interaction of electromagnetic radiation with the leaves, branches and canopies of forests.[102–104] These models can be used to simulate the reflectance of forest canopies by running them in the ''forward'' mode where data on the forest canopy variables are the inputs and the spectral signature is the output (Figure 4). They have also been used to estimate forest biophysical properties by applying them in the ''inverse'' mode where the spectral signature is the input and estimates of the forest biophysical variables are the outputs.[105–107] Although this work is in the early stages of development, the approach is potentially more robust and more accurate than the statistical techniques currently used.

Studies at a regional to global scale have been stimulated by the results of work that has suggested a close link between vegetation and climate. Myneni et al.[5] measured the change in global NDVI using AVHRR data from 1981 to 1991 and found evidence of increased vegetation activity over the period, particularly north of 45° north across the evergreen coniferous forest belt. The change in NDVI suggested an increase in the length of the active growing season in this region and was consistent with the observed increase in the amplitude of the seasonal cycle of atmospheric CO_2 concentration. Further advances in understanding the links between vegetation growth and climate are likely as the length of the satellite record increases and new spaceborne sensors for global environmental monitoring are launched.

At a local scale there have been recent developments in exploiting the spatial variation in remotely sensed images to complement techniques that use the spectral variation. A number of studies have shown that high spatial resolution data may be used to automatically determine structural variables in forest stands and to increase the accuracy of estimates of biophysical variables.[17,108,109] The availability of high spatial resolution data from space should provide the stimulus for further developments in this area.

8 CONCLUSIONS

Forest resource assessment by remote sensing began in the first part of the twentieth century with local-scale forest mapping from aerial photography. Since that time there have been significant advances in remote sensing technology which have allowed forest resources to be assessed over much larger areas. Advances in understanding remotely sensed data have led to exciting new developments in determining the biophysical, biochemical and structural properties of forest canopies and have provided the key input data to forest ecosystem models. In addition to their role as a source of timber products, forests are increasingly being seen as an ecological resource and as a key component of the global ecosystem. The combination of new remote sensing technology and new data analysis techniques with advances in remote sensing science and ecosystem modeling have assured a critical role for remote sensing in mapping, monitoring and managing temperate forest resources.

ABBREVIATIONS AND ACRONYMS

ACCP	Accelerated Canopy Chemistry Program
AirSAR	Airborne Synthetic Aperture Radar
AIS	Airborne Imaging Spectrometer
AMI	Active Microwave Instrument
ANN	Artificial Neural Network
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Airborne Visible Infrared Imaging Spectrometer
BGC	Biogeochemical Cycling
BOREAS	Boreal Ecosystem and Atmosphere Study
CASI	Compact Airborne Imaging Spectrometer
GCM	Global Climate Model
LAI	Leaf Area Index
MERIS	Medium Resolution Imaging Spectrometer
MIR	Middle-infrared
MODIS	Moderate Resolution Imaging Spectrometer

NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NOAA	National Oceanographic and Atmospheric Administration
NPP	Net Primary Productivity
OTTER	Oregon Transect Ecosystem Research
R	Red Reflectance
SAR	Synthetic Aperture Radar
SIR-C	Shuttle Imaging Radar C
SNR	Signal-to-noise Ratio
SR	Simple Ratio
SWIR	Shortwave Infrared Reflectance
TM	Thematic Mapper
WI	Water Index

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction • Global Land Databases for Environmental Analyses • Hyperspectral Remote Sensing: Data Collection and Exploitation • Land Cover Assessment and Monitoring • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring • Tropical Forest Resource Assessment by Remote Sensing

REFERENCES

1. O.W. Archibold, *Ecology of World Vegetation*, Chapman and Hall, London, 1995.
2. C.C. Grier, K.A. Vogt, M.R. Keys, R.L. Edmonds, 'Biomass Distribution and Above- and Below-ground Production in Young and Mature *Abies amabalis* Zone Ecosystems of the Washington Cascades', *Can. J. Forest Res.*, **11**, 155–167 (1991).
3. W.H. Schlesinger, *Biogeochemistry: An Analysis of Global Change*, Academic Press, San Diego, CA, 1991.
4. T.M. Smith, R. Leemans, H.H. Shugart, 'Sensitivity of Terrestrial Carbon Storage to CO_2-induced Climate Change: Comparison of Four Scenarios Based on General Circulation Models', *Climatic Change*, **21**, 367–384 (1992).
5. R.B. Myneni, C.D. Keeling, C.J. Tucker, G. Asrar, R.R. Nemani, 'Increased Plant Growth in the Northern High Latitudes from 1981–1991', *Nature (London)*, **386**, 698–702 (1997).
6. D. Marceau, P. Howarth, D. Gratton, 'Remote Sensing and the Measurement of Geographical Entities in a Forested Environment. 1. The Scale and Spatial Aggregation Problem', *Remote Sens. Environ.*, **49**, 93–104 (1994).
7. F.M. Danson, S.E. Plummer, S.A. Briggs, 'Remote Sensing and the Information Extraction Problem', in *Advances in Environmental Remote Sensing*, eds. F.M. Danson, S.E. Plummer, John Wiley & Sons, Chichester, 171–177, 1995.
8. P. Aplin, P.M. Atkinson, P.J. Curran, 'Fine Spatial Resolution Satellite Sensors for the Next Decade', *Int. J. Remote Sens.*, **18**, 3873–3881 (1997).
9. M.A. Wulder, 'Optical Remote Sensing Techniques for the Assessment of Forest Inventory and Biophysical Parameters', *Prog. Phys. Geogr.*, **22**, 449–476 (1998).
10. D. Leckie, 'Advances in Remote Sensing Technologies for Forest Survey and Management', *Can. J. Forest Res.*, **20**, 464–483 (1990).
11. P.M. Mather, *Computer Processing of Remotely Sensed Data*, 2nd edition, John Wiley & Sons, Chichester, 1998.
12. J. Cihlar, J. Beaubien, Q. Xiao, J.M. Chen, Z. Li, 'Land Cover of the BOREAS Region from AVHRR and Lansdat Data', *Can. J. Remote Sens.*, **23**, 163–175 (1997).
13. S.N. Goward, C.J. Tucker, D.G. Dye, 'North American Vegetation Patterns Observed with NOAA Advanced Very High Resolution Radiometer', *Vegetation*, **64**, 3–14 (1985).
14. T.R. Loveland, A.S. Belward, 'The IGBP-DIS Global 1 km Land Cover Data Set, DISCover: First Results', *Int. J. Remote Sens.*, **18**, 3291–3295 (1997).
15. K. Sharma, A. Sarker, 'A Modified Contextual Classification Technique for Remote Sensing Data', *Photogramm. Eng. Remote Sens.*, **64**, 273–280 (1998).
16. D. Peddle, G.M. Foody, A. Zhang, S.E. Franklin, E. LeDrew, 'Muti-source Image Classification II. An Empirical Comparison of Evidential Reasoning and Neural Network Approaches', *Can. J. Remote Sens.*, **20**, 396–407 (1994).
17. M.A. Wulder, E.F. LeDrew, S.E. Franklin, M.B. Lavigne, 'Aerial Image Texture Information in the Estimation of Northern Deciduous and Mixed Wood Forest Leaf Area Index (LAI)', *Remote Sens. Environ.*, **64**, 64–76 (1998).
18. H. Bischof, W. Schneider, A.J. Pinz, 'Multispectral Classification of Landsat Images Using Neural Networks', *IEEE Trans. Geosci. Remote Sens.*, **30**, 482–489 (1992).
19. G.M. Foody, M.K. Arora, 'An Evaluation of Some Factors Affecting the Accuracy of Classification by an Artificial Neural Network', *Int. J. Remote Sens.*, **18**, 799–810 (1997).
20. G.M. Foody, 'Using Prior Knowledge in Artificial Neural Network Classification with a Minimal Training Set', *Int. J. Remote Sens.*, **16**, 301–312 (1995).
21. G.F. Hepner, T. Logan, N. Ritter, N. Bryant, 'Artificial Neural Network Classification Using a Minimal Training

Set: Comparison to Conventional Supervised Classification', *Photogramm. Eng. Remote Sens.*, **56**, 469–473 (1990).
22. J. Ardö, P. Pilesjö, A. Skidmore, 'Neural Networks, Multitemporal Landsat Thematic Mapper Data and Topographic Data to Classify Forest Damages in the Czech Republic', *Can. J. Remote Sens.*, **23**, 217–229.
23. G.B. Bonan, 'Importance of Leaf Area Index and Forest Type When Estimating Photosynthesis in Boreal Forests', *Remote Sens. Environ.*, **43**, 303–314 (1993).
24. P.J. Curran, 'Multispectral Remote Sensing of Vegetation Amount', *Prog. Phys. Geogr.*, **4**, 315–341 (1980).
25. G. Asrar, E.T. Kanemasu, M. Yoshida, 'Estimates of Leaf Area Index from Spectral Reflectance of Wheat Under Different Agricultural Practices and Solar Angle', *Remote Sens. Environ.*, **17**, 1–11 (1985).
26. F.M. Danson, 'Developments in the Remote Sensing of Forest Canopy Structure', in *Advances in Environmental Remote Sensing*, eds. F.M. Danson, S.E. Plummer, John Wiley & Sons, Chichester, 52–69, 1995.
27. D.L. Peterson, M.A. Spanner, S.W. Running, K.B. Teuber, 'Relationship of Thematic Mapper Simulator Data to Leaf Area Index of Temperate Coniferous Forests', *Remote Sens. Environ.*, **22**, 323–341 (1987).
28. M.A. Spanner, L.L. Pierce, D.L. Peterson, S.W. Running, 'Remote Sensing of Temperate Coniferous Forest Leaf Area Index: The Influence of Canopy Closure, Understorey Vegetation and Background Reflectance', *Int. J. Remote Sens.*, **11**, 95–111 (1990).
29. F.G. Hall, Y.E. Shimabukuro, K.F. Huemmrich, 'Remote Sensing of Forest Biophysical Structure in Boreal Forest Stands of *Picea mariana* Using Mixture Decomposition and Geometric Reflectance Models', *Ecol. Appl.*, **5**, 993–1013 (1995).
30. F.M. Danson, S.E. Plummer, 'Red-edge Response to Forest Leaf Area Index', *Int. J. Remote Sens.*, **16**, 183–188 (1995).
31. F.G. Hall, D.R. Peddle, E.F. Drew, 'Remote Sensing of Biophysical Variables in Boreal Forest Stands of *Picea mariana*', *Int. J. Remote Sens.*, **17**, 3077–3081 (1996).
32. G.D. Badwhar, R.B. MacDonald, F.G. Hall, J.G. Carnes, 'Spectral Characterization of Biophysical Characteristics in a Boreal Forest: Relationship Between Thematic Mapper Band Reflectance and Leaf Area Index for Aspen', *IEEE Trans. Geosci. Remote Sens.*, **24**, 122–128 (1986).
33. P.J. Curran, J.L. Dungan, H.L. Gholz, 'Seasonal LAI in Slash Pine Estimated with Landsat TM', *Remote Sens. Environ.*, **39**, 3–13 (1992).
34. J.M. Chen, J. Cihlar, 'Retrieving Leaf Area Index of Boreal Conifer Forests Using Landsat TM Images', *Remote Sens. Environ.*, **55**, 153–162 (1996).
35. J.R. Miller, H.P. White, J.M. Chen, D.R. Peddle, G. McDermid, R.A. Fournier, P. Sheperd, I. Rubenstein, J. Freemantle, R. Soffer, E. LeDrew, 'Seasonal Changes in Understorey Reflectance of Boreal Forests and Influence on Canopy Vegetation Indices', *J. Geophys. Res.*, **102**, 29 475–29 482 (1997).
36. K.S. Fassnacht, S.T. Gower, M.D. MacKenzie, E.V. Nordheim, T.M. Lillesand, 'Estimating the Leaf Area Index of North Central Wisconsin Forests Using the Landsat TM', *Remote Sens. Environ.*, **61**, 229–245 (1997).
37. R. Nemani, L.L. Pierce, S.W. Running, L. Band, 'Forest Ecosystem Processes at the Watershed Scale: Sensitivity to Remotely Sensed Leaf Area Index Estimates', *Int. J. Remote Sens.*, **14**, 2519–2534 (1993).
38. G.M. Foody, D.S. Boyd, P.J. Curran, 'Relationships Between Tropical Forest Biophysical Properties and Data Acquired in AVHRR Channels 1–5', *Int. J. Remote Sens.*, **17**, 1341–1355 (1996).
39. P. Gong, R. Pu, J.R. Miller, 'Correlating Leaf Area Index of Ponederosa Pine with Hyperspectral CASI Data', *Can. J. Remote Sens.*, **18**, 275–282 (1992).
40. L.F. Johnson, C.A. Hlavka, D.L. Peterson, 'Multivariate Analysis of AVIRIS Data for Canopy Biochemical Estimation Along the Oregon Transect', *Remote Sens. Environ.*, **47**, 216–230 (1994).
41. D.T. Shaw, T.J. Malthus, J.A. Kupiec, 'High-spectral Resolution Data for Monitoring Scots Pine (*Pinus sylvestris*) Regeneration', *Int. J. Remote Sens.*, **13**, 2601–2608 (1998).
42. J.M. Chen, P.M. Rich, S.T. Gower, J.M. Norman, S. Plummer, 'Leaf Area Index of Boreal Forests: Theory, Techniques, and Measurements', *J. Geophys. Res.*, **102**, 29 429–29 443 (1997).
43. D.L. Peterson, R.H. Waring, 'Overview of the Oregon Transect Ecosystem Research Project', *Ecol. Appl.*, **4**, 211–225 (1994).
44. P.J. Sellers, F.G. Hall, R.D. Kelly, A. Black, D. Baldocchi, J. Berry, M. Ryan, K.J. Ranson, P.M. Crill, D.P. Lettenmaier, H. Margolis, J. Cihlar, J. Newcomer, D. Fitzjarrald, P.G. Jarvis, S.T. Gower, D. Halliwell, D. Williams, B. Goodison, D.E. Wickland, F.E. Guertin, 'BOREAS in 1997: Experiment Overview, Scientific Results and Future Directions', *J. Geophys. Res.*, **102**, 28 731–28 769 (1997).
45. T.P. Dawson, P.J. Curran, S.E. Plummer, P.R.J. North, 'Determining Causal Effects of Foliar Biochemical Concentrations on Forest Canopy Reflectance Using Physical Models and Airborne Spectrometry', in *Observations and Interactions*, Remote Sensing Society, Nottingham, 123–128, 1997.
46. G.A. Riggs, S.W. Running, 'Detection of Canopy Water Stress in Conifers Using the Airborne Imaging Spectrometer', *Remote Sens. Environ.*, **35**, 51–68 (1991).
47. J. Peñuelas, I. Filell, L. Serrano, R. Savé, 'Cell Wall Elasticity and Water Index (R 970 nm/R 900 nm) in Wheat Under Different Nitrogen Availabilities', *Int. J. Remote Sens.*, **17**, 373–382 (1996).
48. P.J. Curran, J.L. Dungan, H.L. Gholz, 'Exploring the Relationship Between Reflectance Red Edge and

Chlorophyll Content in Slash Pine', *Tree Physiol.*, **7**, 33–48 (1990).
49. T. Fourty, F. Baret, 'Vegetation Water and Dry Matter Contents Estimated From Top-of-the-atmosphere Reflectance Data: A Simulation Study', *Remote Sens. Environ.*, **61**, 43–45 (1997).
50. H.L. Gholz, 'Environmental Limits on Aboveground Net Primary Production, Leaf Area and Biomass in Vegetation Zones of the Pacific Northwest', *Ecology*, **63**, 469–481 (1982).
51. J.M. Melillo, J.D. Aber, J.M. Muratore, 'Nitrogen and Lignin Control of Hardwood Leaf Litter Decomposition Dynamics', *Ecology*, **63**, 621–626 .
52. D.L. Peterson, J.D. Aber, P.A. Matson, D.H. Card, N. Swanberg, C.A. Wessman, M.A. Spanner, 'Remote Sensing of Forest Canopy and Leaf Biochemical Contents', *Remote Sens. Environ.*, **22**, 323–341 (1988).
53. C.A. Wessman, J.D. Aber, D.L. Peterson, J.M. Melillo, 'Foliar Analysis Using Near Infrared Reflectance Spectroscopy', *Can. J. Forest Res.*, **18**, 6–11 (1988).
54. D.H. Card, D.L. Peterson, P.A. Matson, J.D. Aber, 'Prediction of Leaf Chemistry by Use of Visible and Near Infrared Reflectance Spectroscopy', *Remote Sens. Environ.*, **26**, 123–147 (1988).
55. S. Jacquemoud, G. Verdebout, G. Schmuck, G. Andreoli, B. Hosgood, 'Investigations of Leaf Biochemistry by Statistics', *Remote Sens. Environ.*, **54**, 180–188 (1995).
56. L.F. Johnson, C.R. Billow, 'Spectroscopic Estimation of Total Nitrogen Concentration in Douglas-fir Foliage', *Int. J. Remote Sens.*, **17**, 489–500 (1996).
57. T.M. McLellan, M.E. Martin, J.D. Aber, J.M. Melillo, K.J. Nadelhoffer, 'Comparison of Wet Chemistry and Near-infrared Reflectance Measurements of Carbon-fraction Chemistry and Nitrogen Concentration of Forest Foliage', *Can. J. Forest Res.*, **21**, 1689–1693 (1991).
58. Y.L. Grossman, S.L. Ustin, S. Jacquemoud, E.W. Sanderson, G. Schmuck, J. Verdebout, 'Critique of Stepwise Multiple Linear Regression for the Extraction of Leaf Biochemistry from Leaf Reflectance Data', *Remote Sens. Environ.*, **56**, 182–193 (1996).
59. G. Vane, A.F.H. Goetz, 'Terrestrial Imaging Spectrometry: Current Status, Future Trends', *Remote Sens. Environ.*, **44**, 117–126 (1993).
60. C.A. Wessman, J.D. Aber, D.L. Peterson, J.M. Melillo, 'Remote Sensing of Canopy Chemistry and Nitrogen Cycling in Temperate Forest Ecosystems', *Nature (London)*, **335**, 154–156 (1988).
61. J. Aber (ed.), 'Accelerated Canopy Chemistry Program', Final Report to NASA-EOS1 WG, NASA, Washington, DC, 1994.
62. P.A. Matson, L.F. Johnson, C.R. Billow, J.R. Miller, R. Pu, 'Seasonal Patterns and Remote Spectral Estimation of Canopy Chemistry Across the Oregon Transect', *Ecol. Appl.*, **4**, 280–298 (1994).
63. M.E. Martin, J.D. Aber, 'High-spectral Resolution Remote Sensing of Forest Canopy Lignin, Nitrogen and Ecosystem Processes', *Ecol. Appl.*, **7**, 431–443 (1997).
64. P.J. Curran, J.A. Kupiec, 'Imaging Spectrometry: A New Tool for Ecology', in *Advances in Environmental Remote Sensing*, eds. F.M. Danson, S.E. Plummer, John Wiley & Sons, Chichester, 71–78, 1995.
65. P.J. Curran, J.A. Kupiec, G.M. Smith, 'Remote Sensing the Biochemical Composition of a Slash Pine Canopy', *IEEE Trans. Geosci. Remote Sens.*, **35**, 415–420 (1997).
66. J. Ardö, 'Volume Quantification of Coniferous Forest Compartments Using Spectral Radiance Recorded by Landsat Thematic Mapper', *Int. J. Remote Sens.*, **13**, 1779–1786 (1992).
67. S. Poso, T. Häme, M. Simila, 'Forest Inventory by Compartments Using Satellite Imagery', *Silva Fenn.*, **21**, 69–94 (1987).
68. F.M. Danson, P.J. Curran, 'Factors Affecting the Remotely Sensed Response of Coniferous Forest Canopies', *Remote Sens. Environ.*, **43**, 55–65 (1993).
69. F.M. Danson, 'Preliminary Evaluation of the Relationship Between SPOT-1 HRV Data and Forest Stand Parameters', *Int. J. Remote Sens.*, **8**, 1571–1574 (1987).
70. W.J. Ripple, S. Wang, D.L. Isaacson, D.P. Paine, 'A Preliminary Comparison of Landsat TM and SPOT-1 HRV Multispectral Data for Estimating Coniferous Forest Volume', *Int. J. Remote Sens.*, **12**, 1971–1977 (1991).
71. W.B. Cohen, T.A. Spies, 'Estimating Structural Attributes of Douglas-Fir/Western Hemlock Forest Stands from Landsat and SPOT Imagery', *Remote Sens. Environ.*, **28**, 131–141 .
72. X. Baulies, X. Pons, 'Approach to Forest Inventory and Mapping by Means of Multispectral Airborne Data', *Int. J. Remote Sens.*, **16**, 61–80 (1995).
73. R.R. DeWulf, R.E. Goosens, B.P. DeRoover, F.C. Borry, 'Extraction of Forest Stand Parameters from Panchromatic and Multispectral SPOT-1 Data', *Int. J. Remote Sens.*, **11**, 1571–1588 (1990).
74. J.A. Brockhaus, S. Khorram, 'A Comparison of SPOT and Landsat-TM Data for Use in Conducting Inventories of Forest Resources', *Int. J. Remote Sens.*, **13**, 3035–3043 (1992).
75. T. Nilson, U. Peterson, 'Age Dependence of Forest Reflectance: Analysis of Main Driving Factors', *Remote Sens. Environ.*, **48**, 319–331 (1994).
76. F.M. Gemmell, 'Effects of Forest Cover, Terrain and Scale on Timber Volume Estimation with Thematic Mapper Data in a Rocky Mountain Site', *Remote Sens. Environ.*, **51**, 291–305 (1995).
77. S. Quegan, 'Recent Advances in Understanding SAR Imagery', in *Advances in Environmental Remote Sensing*, eds. F.M. Danson, S.E. Plummer, John Wiley & Sons, Chichester, 89–104, 1995.

78. J.R. Baker, P.L. Mitchell, R.A. Cordey, G.B. Groom, J.J. Settle, M.R. Stileman, 'Relationships Between Physical Characteristics and Polarimetric Radar Backscatter for Corsican Pine Stands in Thetford Forest, UK', *Int. J. Remote Sens.*, **15**, 2827–2894 (1994).
79. J.E.S. Fransson, H. Israelsson, 'Estimation of Stem Volume in Boreal Forests Using ERS-1 C-and JERS-1 L-Band SAR Data', *Int. J. Remote Sens.*, **20**, 123–137 (1999).
80. K.J. Ranson, G. Sun, R.H. Lang, N.S. Chauhan, R.J. Cacciola, O. Kilic, 'Mapping of Boreal Forest Biomass from Spaceborne Synthetic Aperture Radar', *J. Geophys. Res.*, **102**, 599–610 (1997).
81. M.L. Imhoff, 'Theoretical Analysis of the Effect of Forest Structure on Synthetic-aperture Radar Backscatter and the Remote Sensing of Biomass', *IEEE Trans. Geosci. Remote Sens.*, **33**, 341–352 (1997).
82. T. le Toan, A. Beaudoin, J. Riom, D. Guyon, 'Relating Forest Biomass to SAR Data', *IEEE Trans. Geosci. Remote Sens.*, **30**, 403–411 (1992).
83. K.J. Ranson, G. Sun, 'Northern Forest Classification Using Temporal Multifrequency and Multipolarimetric SAR Images', *Remote Sens. Environ.*, **47**, 142–153 (1994).
84. S.S. Saatchi, E. Rignot, 'Classification of Boreal Forest Cover Types Using SAR Images', *Remote Sens. Environ.*, **60**, 270–281 (1997).
85. J. Hyyppä, J. Pulliainen, M. Hallikainen, A. Soatsi, 'Radar Derived Standwise Forest Inventory', *IEEE Transa. Geosci. Remote Sens.*, **35**, 392–404 (1997).
86. E. Næsset, 'Estimating Timber Volume of Forest Stands Using Airborne Laser Scanner Data', *Remote Sens. Environ.*, **61**, 246–253 (1997).
87. R.A. Betts, P.M. Cox, S.E. Lee, F.I. Woodward, 'Contrasting Physiological and Structural Vegetation Feedback in Climate Change Simulations', *Nature (London)*, **387**, 796–799 (1997).
88. M.E. Schlesinger, J.F.B. Mitchell, 'Climate Model Calculations of the Equilibrium Climatic Response to Increased Carbon Dioxide', *Rev. Geophys.*, **25**, 760–798 (1997).
89. W.L. Chapman, J.W. Walsh, 'Recent Variations of Sea Ice and Air Temperature in High Latitudes', *Bull. Am. Meteorol. Soc.*, **74**, 33–47 (1993).
90. C.D. Keeling, J.F.S. Chin, T.P. Whorf, 'Increased Activity of Northern Vegetation Inferred from Atmospheric CO_2 Measurements', *Nature (London)*, **382**, 146–149 (1996).
91. A.S. Denning, I. Fung, D.A. Randall, 'Latitudinal Gradient of Atmospheric CO_2 Due to Seasonal Exchange with Land Biota', *Nature (London)*, **376**, 240–243 (1995).
92. G.B. Bonan, 'Atmosphere–Biosphere Exchange of Carbon Dioxide in Boreal Forests', *J. Geophys. Res.*, **96**, 7301–7312 (1991).
93. G.B. Bonan, S.L. Pollard, B.K. Welson, 'Effects of Boreal Forest Vegetation on Global Climate', *Nature (London)*, **359**, 716–718 (1992).
94. P.G. Jarvis, J.M. Massheder, S.E. Hale, J.G. Moncrieff, M. Rayment, S.L. Scott, 'Seasonal Variation of Carbon Dioxide, Water Vapour and Energy Exchanges of a Boreal Black Spruce Forest', *J. Geophys. Res.*, **102**, 2853–2866 (1997).
95. M.B. Lavigne, M.G. Ryan, 'Growth and Maintenance Respiration Rates of Aspen, Black Spruce and Jack Pine Stems at Northern and Southern BOREAS Sites', *Tree Physiol.*, **17**, 543–552 (1997).
96. S.W. Running, R.R. Nemani, D.L. Peterson, L.E. Band, D.F. Potts, L.L. Pierce, M.A. Spanner, 'Mapping Regional Forest Evapotranspiration and Photosynthesis by Coupling Satellite Data with Ecosystem Simulation', *Ecology*, **70**, 1090–1101 (1989).
97. W.P. Cropper, Jr., H.L. Gholz, 'Simulation of the Carbon Dynamics of a Florida Slash Pine Plantation', *Ecol. Model.*, **66**, 231–249 (1993).
98. S.W. Running, J.C. Coughlan, 'A General Model of Forest Ecosystem Processes for Regional Applications. I. Hydrologic Balance, Canopy Gas Exchange and Primary Production Processes', *Ecol. Model.*, **42**, 125–154 (1988).
99. N.S. Lucas, R. Green, P.J. Curran, S.E. Plummer, F.M. Danson, 'Driving a Forest Ecosystem Model With Rededge Data', *Int. J. Remote Sens.*, in press.
100. J. Liu, J.M. Chen, J. Cihlar, W.M. Park, 'A Process Based Boreal Ecosystem Productivity Simulator Using Remote Sensed Inputs', *Remote Sens. Environ.*, **62**, 158–175 (1997).
101. J.M. Chen, J. Liu, J. Cihlar, M.L. Goulden, 'Daily Canopy Photosynthesis Model Through Temporal and Spatial Scaling for Remote Sensing Applications', *Ecol. Model.*, submitted.
102. X. Li, A.H. Strahler, C.E. Woodcock, 'A Hybrid Geometric Optical–Radiative Transfer Approach for Modelling Albedo and Directional Reflectance of Discontinuous Canopies', *IEEE Trans. Geosci. Remote Sens.*, **33**, 466–480 (1995).
103. P.R.J. North, 'Three-dimensional Forest Light Interaction Model Using a Monte Carlo Method', *IEEE Trans. Geosci. Remote Sens.*, **34**, 946–956 (1996).
104. J. Chen, S.G. Leblanc, 'A Four-scale Bidirectional Reflectance Model Based on Canopy Architecture', *IEEE Trans. Geosci. Remote Sens.*, **35**, 1316–1337 (1997).
105. C.E. Woodcock, J.B. Collins, S. Gopal, V.D. Jakabhazy, X. Li, S. Macomber, S. Rygerd, V.J. Harwood, J. Levitan, Y. Wu, R. Warbington, 'Mapping Forest Vegetation Using Landsat TM Imagery and a Canopy Reflectance Model', *Remote Sens. Environ.*, **50**, 240–254 (1994).
106. P. Gong, D.X. Wang, S. Liang, 'Inverting a Canopy Reflectance Model Using a Neural Network', *Int. J. Remote Sens.*, **20**, 111–122 (1999).

107. F.M. Gemmell, J. Varjo, 'Utility of Reflectance Model Inversion Versus Two Spectral Indices for Estimation Biophysical Characteristics in a Boreal Forest Test Site', *Remote Sens. Environ.*, **68**, 95–111 (1999).
108. B.A. St-Onge, F. Cavayas, 'Estimating Forest Stand Structure from High Resolution Imagery Using the Directional Variogram', *Int. J. Remote Sens.*, **16**, 1999–2021 (1995).
109. B.A. St-Onge, F. Cavayas, 'Automated Forest Structure Mapping from High Resolution Imagery Based on Directional Semivariogram Estimates', *Remote Sens. Environ.*, **61**, 82–95 (1997).

Tropical Forest Resource Assessment by Remote Sensing

Doreen S. Boyd
Kingston University, Kingston upon Thames, UK

1 Introduction **1**
2 Tropical Forest Extent and Change Dynamics **3**
2.1 Delineating Forest Cover from Nonforest Cover 4
2.2 Fire Occurrence 10
3 Tropical Forest Types **10**
4 Tropical Forest Properties **14**
5 Future Capabilities **16**
Abbreviations and Acronyms **16**
Related Articles **17**
References **17**

The earth's tropical forests are a global resource, requiring action for their wise utilization, management and protection at spatial scales from the local through to the global. The acquisition of data pertaining to tropical forest resources presents many challenges,[1] many of which may be overcome through the use of remote sensing. Reasons for the adoption of remote sensing for the assessment of tropical forest resources are compelling. Remote sensing can be used to observe the dynamics of tropical forests at almost any level of detail. It allows the frequent measurement and monitoring of the earth's tropical forests on a continuous basis, from which accurate and informed judgments on their resources at any given time can be made. Accordingly, there has been progressive evolution of remote sensing approaches for the collection of tropical forest resource information. With the imminent launch of new spaceborne sensors which will be able to acquire more information on tropical forests, remote sensing will continue to be a crucial tool in the assessment of tropical forest resources.

1 INTRODUCTION

Tropical forest regions are often impenetrable in nature, they cover vast areas and are complex environments whose exploitation has ramifications globally as well as locally. The plethora of data on forest resources, collected using traditional techniques, often lack uniformity of coverage, quality and content. The information available is scattered and diverse, compilation is difficult and synthesis all too rare.[2] Many of the inherent difficulties of working within tropical forest environments, however, may be overcome through the use of remote sensing.

Remote sensing can be used to observe the dynamics of tropical forests at almost any level of detail.[3] It allows the frequent measurement and monitoring of the world's tropical forests on a continuous basis,[4] from which accurate and informed judgments on their resources at any given time can be made. Accordingly remote sensing approaches for the collection of tropical forest resource information have progressively evolved, with the focus shifting from remotely sensed data acquired by aerial cameras to spaceborne sensors.[5] Particular benefits of the adoption of remote sensing for the assessment of tropical forest resources are that (1) the data are spatially continuous and consistent in nature; (2) synoptic coverage is achieved with more intensive spatial sampling than the sparse point samples traditionally taken, allowing the spatial extension of ground estimates which would be impossible without extensive and costly ground surveys;[6] (3) data are collected in a digital format with access to them and the methods for their processing becoming increasingly easy.[7] Moreover, the digital format of remotely sensed data is conducive to computer-based analysis and integration with other data sets within a geographical information system (GIS);[8] (4) remote sensing provides a low cost per unit area means of resources data collection and (5) the frequency of data collection is high which permits the incorporation of temporal information into an analysis,[9] affording a long-term database for the examination of the dynamics of resource use. The latter point is also a particular advantage in view of the fact that many areas of tropical forests undergo gradual rather than dramatic change. Periods of drought, for example, have lead to the gradual degradation of large tracts of forest in several regions of southeast Asia.[10]

Any resource information which may be inferred from the remotely sensed data acquired by spaceborne sensors will always be expressed first as a radiative property (R) of the forest and any background in view, which is, after atmospheric correction and to a first approximation, a function (f) of the location (x), time (t), wavelength (λ) and viewing geometry (θ) of the given ground-resolution element:[11] (Equation 1)

$$R = f(x, t, \lambda, \theta) \quad (1)$$

These radiative properties are then processed and interpreted to extract resource information, of which there are three levels of detail. The first level refers to

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Sample of spaceborne sensors currently operating to provide remotely sensed data for the measurement of tropical forest resources

Platform and sensor	Swath width (km)	Spatial resolution (m)	Temporal resolution (days)	Wavelengths sensed						
				Visible 0.5–0.8 μm	NIR 0.8–1.5 μm	SWIR 1.5–3.0 μm	MIR 3.0–5.0 μm	TIR 8.0–15.0 μm	C-band microwave 5.6 cm	L-band microwave 23 cm
Landsat multispectral scanner	185	80	18	✓	✓					
Landsat TM	185	30 120[c]	16	✓	✓	✓		✓		
SPOT HRV sensor	60[a] and 80[b]	20	26	✓	✓					
MOMS multispectral electronic self-scanning radiometer	100	50	17	✓						
IRS linear imaging self-scanning sensor	70	40	24	✓	✓	✓				
NOAA AVHRR	2700	1100	Twice daily	✓	✓		✓	✓		
ERS along track scanning radiometer	500	1100	6	✓	✓	✓	✓	✓		
Seasat SAR	100	25	152							✓
Space shuttle shuttle imaging radar	100	40	Opportunity dependent							✓
ERS SAR	100	35	35						✓	
JERS SAR	75	18	44	✓	✓					✓
RADARSAT SAR	45	10	24						✓	

Based on Curran and Foody (1994).[111]
[a] Under nadir viewing.
[b] Under off-nadir viewing (maximum angle rotation 27°).
[c] Thermal channel at 120 m spatial resolution.
NIR, near-infrared; SWIR, shortwave-infrared; MIR, middle-infrared; TIR, thermal-infrared; TM, thematic mapper; SPOT, System Probatoire d'Observation de La Terre; HRV, high-resolution visible; MOMS, modular optoelectronic multispectral scanner; IRS, Indian Remote Sensing Satellite; NOAA, National Oceanic and Atmospheric Administration; AVHRR, Advanced Very High-resolution Radiometer; ERS, European Remote Sensing Satellite; SAR, Synthetic Aperture Radar; JERS, Japanese Earth Resources Satellite.

information on the spatial extent of tropical forest cover, which can be used to assess the rate of forest dynamics; the second level comprises information on the types of forest encompassing forested areas and the third level provides information on the biophysical properties of these forests. The assessment of such tropical forest information over time puts into effect the comprehensive monitoring of tropical forest resources.

Preprocessing requirements of the remotely sensed data prior to resource information extraction is dependent primarily on the subsequent use of the data. In general, the two processes of image rectification and restoration and image enhancement are performed. Image rectification and restoration involves the processing of raw digital data for geometric and radiometric corrections.[12] This enables the user of the data to locate themselves accurately in relation to the ground and utilize remotely sensed data acquired over time and space. The objective of image enhancement is to form new data from the original data in order to increase the amount of information that can be extracted.[12] Enhancements involving multiple spectral bands of imagery (e.g. vegetations index production, principal component analyses) have been readily employed. Of most relevance to the remote sensing of tropical forest resources is the production of vegetation indices. Vegetation indices are simple mathematical expressions that combine spectral measurements in two or more spectral channels to provide nondestructive estimates of the properties of vegetation. The most popular and widely used index is the normalized difference vegetation index (NDVI) which is a two-dimensional transformation of channels acquiring radiation in visible and NIR wavelengths, $((\mathrm{NIR} - \mathrm{visible})/(\mathrm{NIR} + \mathrm{visible}))$. The NDVI is an indicator of instantaneous biophysical rates of vegetation and has been found to be highly correlated with vegetation properties such as above-ground biomass and leaf-area index.[13]

There are a number of spaceborne sensors operating to provide remotely sensed data for the measurement of tropical forest resources (Table 1 provides a sample of these). The data from these sensors may also experience radiometric and atmospheric calibration and various levels of preprocessing or may be compiled to provide complete coverage of an area of interest. For instance, there are currently a number of data sets available from the data acquired by the NOAA AVHRR.[14] One such data set is available at 1 km spatial resolution, the large area coverage (LAC). This data set is routinely collected for several continents and data-processing stations are being set up around the world for complete AVHRR "1 km" coverage.[15] Another data set, the Global Area Coverage (GAC) sets, is displayed at a reduced spatial resolution (4 km) through the on-board resampling of the radiance recorded, thus providing data on truly continental and global scales.[16] The Global Rain Forest Mapping (GRFM) Project has acquired contiguous SAR data sets of the earth's major tropical forests using the JERS-1 satellite. This effort led by the Earth Observation Research Center of the National Space Development Agency of Japan (NASDA) is focused on the tropical forest regions of south and central America, central and western Africa, southeast Asia and northern Australia. Each region was observed at least once during a "single season" between September 1995 and January 1997. The final data set comprises 7500 processed SAR scenes and forms the only data set of its kind (NASDA, 1998; http://www.eors.nasda.go.jp/Sciences/Forest).

Unfortunately, there is no one spaceborne sensor currently in use that meets fully the requirements of a tropical forest resources assessment system, with the spatial resolution and frequency of coverage logistically competitive and exclusive of one another.[2] Moreover, the radiation from which tropical forest resources are inferred are sampled at suboptimal spectral resolutions. However, with consideration given to the nature of the desired resource product, an appropriate sensor or combination of sensors may be employed.

2 TROPICAL FOREST EXTENT AND CHANGE DYNAMICS

Techniques for measuring the extent of tropical forests and their change have evolved rapidly in the 1980s and 1990s, with the development of remote sensing techniques vastly improving the capability to do so.[17] A comprehensive evaluation of tropical forest resources requires information on how much tropical forest remains and where those areas are situated.[18] Remote sensing is particularly well placed to provide this, particularly at the regional to global scales.[19]

The remote sensing of the extent of tropical forests and their change dynamics falls naturally into two procedures. The first centers on the delineation of forest from nonforests and the calculation of the areal extent of forest cover. When this exercise is performed on two or more dates, an estimate of deforestation or regeneration within that time period may be made. The second procedure uses the occurrence of forest fires as an indicator of active areas of forest burning which can be used to estimate forest loss per se. Both procedures enlist a multitude of remote sensing methods, employing data from a host of sensors to perform their tasks. The optimal approach to acquiring information on tropical forest cover and their change dynamics is still under consideration.

2.1 Delineating Forest Cover from Nonforest Cover

Pioneering studies illustrated the potential of fine spatial resolution sensors[20,21] for the delineation of forest cover from nonforest cover, prompting the use of sensors such as those carried on the Landsat and SPOT series of satellites. These sensors have been employed to study forest area dynamics mainly at local and national scales.[22–24] The use of these data, particularly at regional and global scales, is, however, compromised by their expense (cost per image multiplied by number of images required to cover area of study), data load and processing requirements and the low frequency of data acquisition, compounded further by the cloud cover in tropical regions and frequent smoke from forest fires.[25] Furthermore, the classification of land cover requires good knowledge of the area under study, which is difficult to achieve beyond the national level.[18] Much consideration has been given to the use of remotely sensed data acquired by the NOAA AVHRR series of sensors. By virtue of its properties, that is its spatial resolution and temporal resolution (providing near-daily coverage), this sensor provides an invaluable data source for coarse-scale land cover studies and has been proposed as the foundation of a global monitoring system.[18,26] The use of SAR systems which operate in microwave wavelengths affords the certainty of cloud-free data. The ever increasing proliferation of spaceborne SAR sensors can build on the recommendation of early studies, which report them to be well suited to the mapping of forest cover.[27,28]

A number of approaches using remotely sensed data have been used to delineate forest cover from nonforest cover. Each approach has implicit advantages and disadvantages and are discussed in turn.

The visual interpretation of a remotely sensed image conducted by trained people with knowledge of the forested area under study has been used in a number of studies, particularly those focusing on national or subnational areas. The interpreter relies on spectral and spatial pattern recognition. The shape, size and pattern characteristics of pixels having similar spectral responses in a single channel or a combination of up to three channels may provide insight to the arrangement of forested areas and nonforested areas. Tucker et al.[29] inspected a NOAA-7 AVHRR image acquired on 9 July 1992 of Rondonia, Brazil and relied mainly on spectral differences between pixels. A series of linear features of higher spectral response in the SWIR channel, and to a lesser extent in the visible and NIR channels (combined within the NDVI) along highway BR-364, were detected and classified as nonforested areas. These were confirmed through field verification as forest clearings associated with a large colonization program. Other researchers have taken advantage of image texture. Texture is evident in fine spatial resolution imagery and refers to the variance of pixel values associated with a particular object. Texture is prevalent in images of tropical forests acquired by sensors such as Landsat TM and SPOT HRV and is probably a result of canopy complexity and the relative sizes of crowns and pixels. The texture associated with a forested area will be different from that of a nonforested area and therefore can be used in the delineation between these two cover types.[30]

Human interpretation of images allows the large variations in feature appearance that occur with different satellite viewing angles and differential bidirectional reflectance of different surfaces to be acknowledged and accounted for. However, the approach can be time-consuming, difficult and subjective rather than operational. Furthermore, it may be wasteful of information, since it is based on an image representation of, at most, three spectral bands at any one time. With the advent of more sophisticated methods, the visual interpretation approach is more often used in the preliminary inspection of imagery or in combination with other approaches.[31]

The binary division of remotely sensed data into a forest or nonforest class allows the application of a radiance threshold to remotely sensed data, whereby those areas having a spectral response on either side of the threshold are allocated to a respective class.[32,33] The success of this approach is enhanced if images are chosen at a time when spectral contrast is maximized. SWIR and MIR radiation provides a relatively large spectral contrast (compared with that in visible and NIR wavelengths) between mature forest and other land cover.[25,34,35] In the separation of dense forests from degraded less homogeneous vegetation, NIR radiation appears useful.[34] SAR data have also been used in thresholding, whereby the use of L-band microwaves provides better discrimination than C-band.[36]

Full utilization of the spectral information acquired by a sensor for the delineation of forests from nonforests can be achieved using digital unsupervised or supervised classifications. The overall objective of image classification procedures is to categorize automatically all pixels of the area of study into land cover classes, in this case to represent forest and nonforest, using the pixel-by-pixel spectral pattern present.[12] In an unsupervised classification the image data are classified by the automatic clustering of spectrally similar pixels. These clusters are then identified by comparing the classified image to ground reference data. The unsupervised approach can be favored over the supervised classification since it does not require a demanding definition of training areas, it is objective and faster to execute.[31] Supervised classifications involve three main stages, the initial training stage, the classification stage and the output stage. The training stage involves the characterization of pixels of known class membership,

in this case forest and nonforest, from which class spectral ''signatures'' are derived. Reference data for the training stage may be derived from reliable cartographic data or via field verification. In the classification stage, these training statistics are used to allocate pixels of unknown class membership to a class in accordance with some decision rule. After the entire data set has been classified, the results are presented in the output stage, which may consist of a thematic map or tables of area statistics and the like. There are many examples of successful unsupervised and supervised classifications of forest and nonforested areas.[37,38,40] The quality of the classification output is evaluated by comparison against some ground or other ancillary reference data from which quantitative measures of accuracy may be derived.[39]

By studying the temporal response of remotely sensed data, that is imagery acquired over the same area at different times of year, a considerable amount of useful information can be obtained for the delineation of forests from nonforests. The general change in spectral response from areas of forest and nonforest can be characterized and utilized.[41] However, it must be remembered that availability of imagery may be restricted by cloud cover contamination and costs. Moreover, preprocessing requirements are demanding since comparison of images requires the time-consuming task of radiometric and atmospheric correction of the remotely sensed data. Nonetheless, the attractiveness of this approach has been advocated. Conway[42] visually inspected a multitemporal sequence of ERS-1 SAR imagery over a study area located in Papua New Guinea. The intensity of the forest pixels was seen to be more temporally stable in comparison with that of the nonforest types. This inspection revealed that savanna pixels were best discriminated from forest pixels during the dry season. Other studies have employed NOAA AVHRR NDVI and surface temperature (derived from the TIR channels) temporal variations.[43] The temporal evolution curves of NDVI and surface temperature for a number of vegetation types of West Africa are shown in Figure 1. These curves illustrate that discriminating between the forest cover and nonforest cover is possible, particularly at certain time periods. The significance of this information is evident in Figure 2 which illustrates the enhanced separability achieved between forest and nonforest using an image acquired in the dry season (22 January and 1 February) as opposed to the wet season (17 May).

The combination of remotely sensed data from multiple sensors (synergy) has been demonstrated to provide improved delineation of forests from nonforests. Synergy

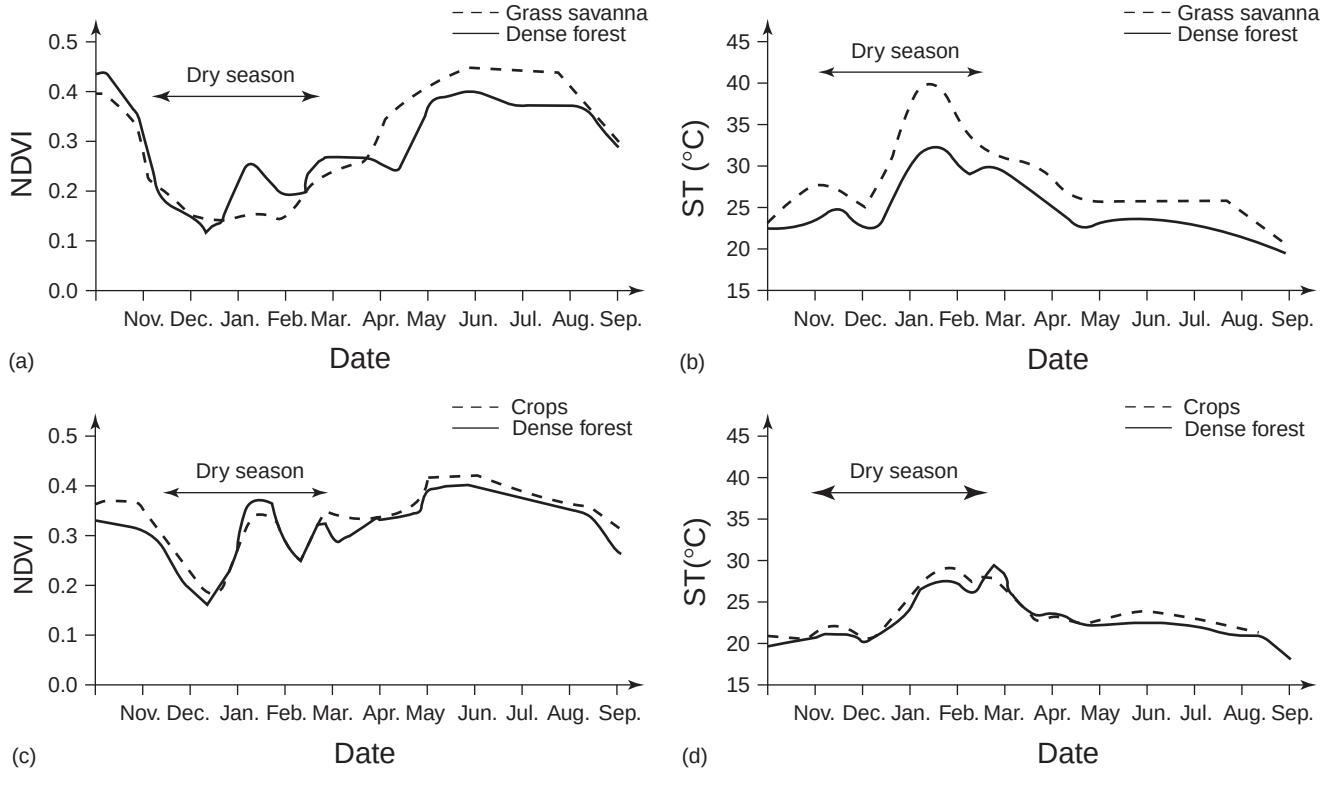

Figure 1 Temporal evolution curves of NOAA AVHRR NDVI and surface temperature for two study areas: (a) and (b) Marahque National Park and (c) and (d) Haut-Sassandra Forest, within the Guinean zone of tropical forests in West Africa. (Reproduced with permission of the American Society for Photogrammetry and Remote Sensing, Achard and Blasco, 1990.[43])

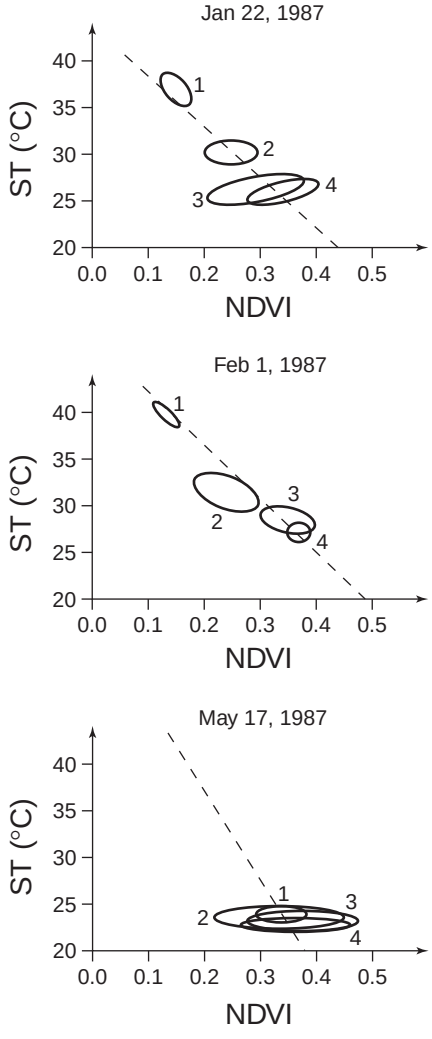

Figure 2 Spectral separability achieved between forest and nonforest using an image acquired in the dry season (22 January and 01 February) as opposed to the wet season (17 May). 1 = Grass savanna of the Marahque National Park; 2 = dense forest of Marahque National Park; 3 = crops of the Haut-Sassandra Forest and 4 = dense forest of the Haut-Sassandra Forest. (Reproduced with permission of the American Society for Photogrammetry and Remote Sensing, Achard and Blasco, 1990 (figure at 75%).[43])

within remote sensing refers to the utilization of two or more data sets together in order to extract more information from their combination than could be extracted from the respective data sources individually.[44] The use of multisensor data allows the combination, on the one hand of spectral information from different parts of the electromagnetic spectrum, and on the other of fine and coarse spatial resolution data.

The synergy of spectral data from different sensors allows utilization of the exclusive information about the forest and nonforest that is provided by specific parts of the spectrum. Particularly attractive is the use of data acquired by optical sensors with microwave sensors. Optical sensors provide information primarily about canopy characteristics whereas microwave sensors provide information on stand attributes, such as wood biomass and geometric structure, as well as soil characteristics (e.g. humidity). Furthermore, the cloud penetrating capabilities of microwave sensors allow areas that have missing data to be included in the analyses, particularly if multitemporal methods are being employed.[45]

Nezry et al.[46] used LHH band SAR data acquired by the space-shuttle mounted Spaceborne Imaging Radar (SIR)-B superimposed with the three channel SPOT-1 HRV XS (high-resolution visible multispectral) data to classify the vegetation of central Sumatra using the maximum likelihood method. A visual accuracy assessment revealed the classification to be more accurate than that obtained using the SIR-B imagery alone. Correlation coefficients between the three SPOT channels and the SIR-B channel of part of the study area revealed the nature of the information provided by these images to be globally very different.

The benefits of using multispatial sensor data has been demonstrated by many researchers.[47–49] The ability of a particular remote sensor to undertake the task of mapping forest and nonforests is dependent upon the sizes of the deforested areas under study, the spatial arrangement of these and the spectral contrast between the deforested areas and the original forest.[50] For example, large discrete clearances with distinct geometric boundaries, such as cattle ranches in eastern Amazonia will often be detected at the 1.1 km spatial resolution expressed in the AVHRR sensor data, even in cases where spectral contrast is minimal.[50,51] However, close to the margins of clearance and where a complex arrangement of forest and nonforest areas or small clearances exist (e.g. in Rondonia, Brazil), the estimation of the true spatial extent of forest clearance would require the use of data collected at a finer spatial resolution. However, the choice of sensor in a particular study may be determined by practicalities such as availability of funds, processing capabilities and time constraints, rather than theoretical knowledge. Coarse spatial resolution imagery is often used in large scale studies, which produce estimates of nonforest/forest which are inaccurate locally because of spatial aggregation errors, but acceptable over the entire population.[18] Whereas fine spatial resolution imagery used in the more local scale studies produces accurate estimates for that area, extrapolations over other areas where the imagery is unavailable may be inaccurate because of spatial variability.

The problems associated with using coarse spatial resolution data have been particularly well documented.[4,17] Generally, as the spatial resolution of an image decreases, so a greater proportion of pixels will have a partial forest

cover and the accuracy of a forest/nonforest classification will decrease. Pixels comprising imagery acquired by current coarse spatial resolution sensors typically represent a ground area of 1.21 km^2 and so the vast majority of pixels will contain two or more land cover classes.[52] These mixed pixels cannot be accommodated or appropriately represented in the methods described above, where each pixel may only be associated with one class, in this case forest or nonforest, leading to a classification error. Skole and Tucker[53] suggested that NOAA AVHRR data overestimated the deforested area in the Legal Amazon by as much as 50% in comparison to estimates derived using Landsat TM data. Cross et al.[54] also observed that at least 66% of the AVHRR ground pixel needed to be forested for inclusion in a forested class within the classified image. As a consequence the extent of forest cover may be underestimated which could result in an overestimation of deforestation rate.

It is possible to compensate for misclassification bias by focusing on the validation and correction of regional estimates of deforestation, although the spatial distribution of deforestation will still be erroneous. Forest/nonforest classifications based on coarse spatial resolution data are compared with classifications of a sample of coregistered finer spatial resolution data.[33] Corrections may take the form of a simple regression between the classification at the two spatial resolutions by obtaining the relationship between the two sets of data as an indication of the extent to which the coarse spatial resolution data represent the areally integrated spectral response of the ground surface at the pixel resolution of the sensors.[55] Others have stratified the coarse resolution classification according to the degree of forest fragmentation across the area of study prior to regression formulation.[56] Misclassifications occurring as a result of image quality, atmospheric variations, topographic and bidirectional reflectance effects, a particular problem with studies using image mosaics, may also be addressed using this approach.

Another approach aimed at improving the classification accuracy of coarse spatial resolution imagery is to unmix the land cover composition of each pixel. Thus rather than derive a conventional ''hard'' image classification whereby each pixel is classified as either forest or nonforest, estimates of the class composition of each pixel may be derived. These may be used to derive fraction images, which display the proportional coverage of a particular class (in this case either forest or nonforest) in each pixel.[57] Although such techniques fail to provide information on the exact location of the subpixel compositions, the accuracy of forest cover estimation is often improved. Techniques currently receiving attention are mixture modeling, artificial neural networks (ANNs) and subpixel calibration.

Linear mixture modeling is one approach used to produce the class composition of pixels, as a precursor to the production of fraction images. This approach assumes that the spectral response recorded for a pixel is simply an additive function of the spectral response of each class weighted by its proportional coverage of the area represented by the pixel. The use of this approach in a forest/nonforest classification requires knowledge of image end-members, which in this case would be the spectral response in each channel utilized from a pure forested pixel and a pure nonforest pixel. This information is used within the linear mixture model[58,59] and for each pixel in the image the proportion of each class computed. Cross et al.[54] implemented a linear-mixture model with channels 1 to 4 of AVHRR to derive forest cover estimates of Rondonia, Brazil and Ghana. The model produced image outputs in which the pixel intensities indicated the proportion of forest cover per square kilometer. Comparisons were drawn with a Landsat TM and an AVHRR classification performed using traditional methods, and improvements in the accuracy of forest cover extent noted. In general, mixture models applied to coarse resolution data enhanced the classification of forest/nonforest where discrete areas of forest are cleared but the models are less successful where a complex type of forest clearance exists (e.g. shifting cultivation) and in areas where only slight variations in the spectral response of forest canopies indicate forest disturbance.[60] Mixing from vegetated targets is often distinctly nonlinear in optical wavelengths owing to factors such as multiple scattering[61] and in thermal wavelengths owing to Planck's function. This restricts the available data to be used in the model, though success using MIR radiation corrected for thermal emission has been noted.[52] Obtaining the class spectral response or end-member spectra to represent the remotely sensed response that would be observed if the pixel represented a homogeneous region comprised of either forest or nonforest, is a further problem. As the vast majority of pixels in an image are likely to be mixed, it is therefore unlikely that the end-member spectra could be derived directly from the imagery. Moreover, nonforest classes may need to be represented by a variety of spectra since pastures of different ages are spectrally distinct,[62] as are plantations, roads and so on. It is possible to use library spectra or, as Adams et al.[63] demonstrated, laboratory or field spectra of materials contributing to the reflectance of each pixel over the study area. Here, reflectance spectra from 400–1000 nm of primary forest canopy were measured from towers, and samples of soil and vegetation were collected in the field and measured in the laboratory over the wavelength range 400–2500 nm. In this instance, these spectra were used in a spectral mixture analysis method applied to multitemporal Landsat TM imagery.

An ANN approach has also been used as a method of deriving the class composition of pixels. ANNs are a form of artificial intelligence that can be trained to learn by example and therefore may be used to recognize the forest/nonforest composition of pixels if trained using appropriate data. Foody et al.[57] used an ANN to map tropical forest cover over a 3000 km^2 region of the Mato Grosso, Brazil. The ANN was trained using coarse spatial resolution NOAA AVHRR data acquired in visible, NIR and MIR channels of which the land cover composition (forest, nonforest and water) had been determined from a coregistered Landsat TM image. The training data comprised data with a range of class compositions allowing the network to recognize the relationship between spectral response and forest cover, nonforest cover and water cover, respectively. Once trained, the ANN was used to allocate a class composition of forest, nonforest and water to the remaining pixels of the AVHRR image using their spectral response. The success of this method can be seen visually in Figure 3, whereby the fraction images produced via the ANN provides a far more realistic representation of the forest cover at the study site. Use of an ANN is attractive since it is able to use mixed pixels in the training stage and it does not assume linear mixing of the classes comprising the pixel's spectral response.

In the ANN approach outlined above, the network is essentially being employed as a regression model whereby the training data is used to determine the relationship between percentage of forest cover and spectral response. Indeed this approach can also be adopted using standard empirical regression modeling. Here a relationship between percentage forest cover and spectral response is determined through classification of finer spatial resolution imagery into forest and nonforest cover. The class proportion of a pixel in the coarse spatial resolution imagery is determined by the percentage of fine spatial resolution pixels classified as forests comprising the same area as the coarse spatial resolution pixel. This is conducted for a sample of coarse spatial resolution pixels and the resultant empirical relationship is applied to the remainder of the imagery and the percentage forest cover of the whole study area determined.[64]

The success in using remotely sensed data for the assessment of tropical forest extent and change dynamics has led to the implementation of major programs that are using remotely sensed data to map and monitor tropical forests, including the following.

2.1.1 United Nations Food and Agricultural Organization

The United Nations Food and Agricultural Organization (FAO) has undertaken periodic assessments of the

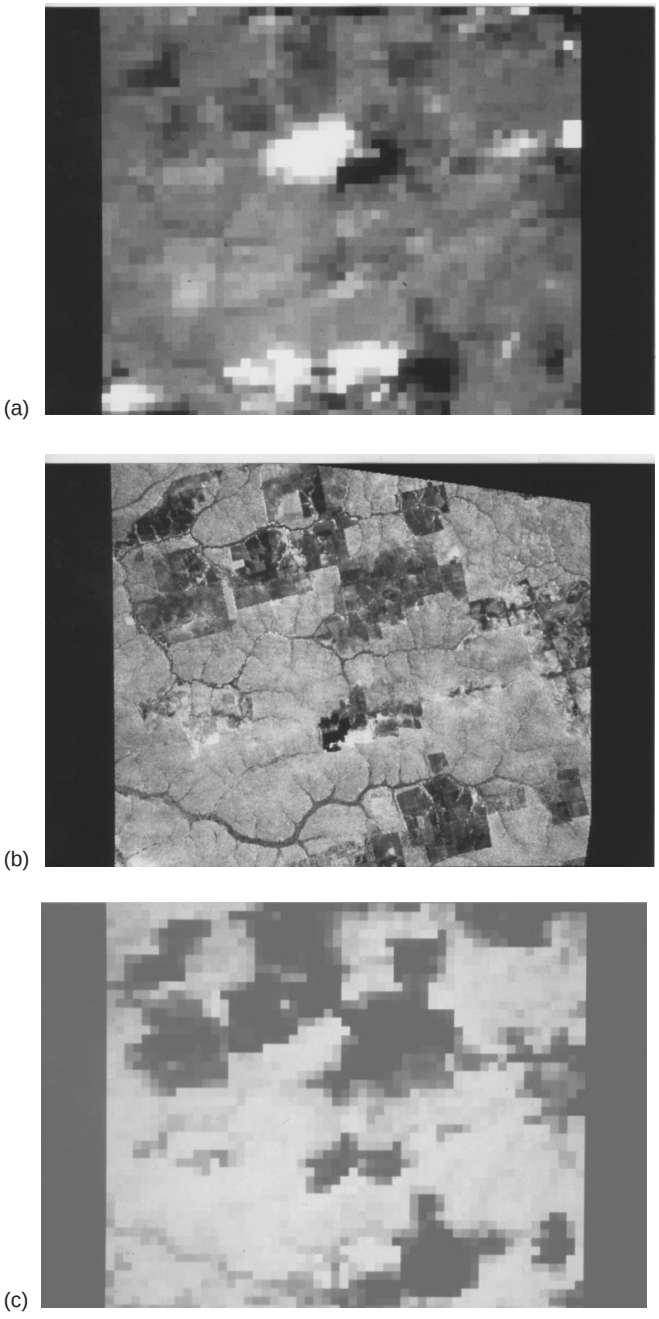

Figure 3 Forest cover delineated from nonforest cover using a supervised classification on (a) the NOAA AVHRR image (light tone area is cloud) and (b) the Landsat TM image. The forested area is readily identifiable on Landsat TM image (gray grainy tone) compared with the NOAA AVHRR image. (c) displays the fraction image illustrating forest cover derived by unmixing the AVHRR pixels with an ANN. The proportion of forest cover in a pixel is related to image tone (black = 100% forest cover and white = 0% forest cover). Note the improved delineation of forest cover from nonforest cover achieved on the fraction image over the hard classification of the NOAA AVHRR image. (Reproduced by kind permission of Kluwer Academic from Foody et al., *Plant Ecology*, 1997.[57])

earth's forest resources, including those of tropical regions, since 1946. In 1990 the Global Forest Resources Assessment was launched which had a remit to provide comprehensive information on the current state of global forests and recent changes including areal extent, growing stock, management and conservation and environmental functions of the forests. This assessment realized the contribution that remote sensing could make to the collection of such information. A remote sensing survey (FAO-RS) was included in the assessment and aimed to produce the resources information needed for the tropical forest zone. The survey employed 117 Landsat TM images, covering 10% of potential tropical forest land. Of these images, 47 covered African locations, 30 covered Asian and 40 Latin American. At each location satellite images of the best quality and appropriate season, separated by an approximate 10-year interval, were selected for interpretation. Local teams having good knowledge of the locations interpreted the multidate images visually to identify nine cover classes, four of which were ''forest classes''; closed forest (canopy cover >40%); open forest (canopy cover 10–40%); long fallow (forest affected by shifting cultivation) and fragmented forests (mosaic of forest and nonforest). Used within a GIS along with geo-referenced data such as ecological zones and population density, these land cover classes indicated tropical forest-cover change at global scales for the decade 1980 through 1990 and then the current (1990) state of tropical forest cover.[65]

2.1.2 *Tropical Ecosystem Environment Observations by Satellites*

The Tropical Ecosystem Environment Observations by Satellites (TREES) project of the European Commission's Joint Research Center (JRC) initiated in 1990 aimed to develop global forest cover assessment methods for the humid tropics and produced a set of tropical forest-cover maps and forest area statistics by country. The project makes use of NOAA AVHRR and ERS-1 SAR data, which are mostly selected in the dry season in order to obtain better contrast between forests and other more seasonal cover types. When seasonal forest is present, the use of a few dates at different stages of the dry season allows the distinction to be made between the evergreen and deciduous formations. Classification by unsupervised methods are applied to quality cloud-free parts of single-date images. Clusters derived are interpreted and labeled. Labeling is based on available field knowledge, ancillary information (existing national forest maps) and a visual analysis of spatial distribution pattern. Single date classifications are assembled into a simple mosaic map and overlapping classifications are chosen on the basis of a ''reliability'' indicator. The final product is a map of forest distribution at 1.1 km spatial resolution. Three main classes are distinguished relating to the proportion of forest cover in the AVHRR pixels: (1) dense humid forest (>70% forest cover in each pixel);(2) fragmented or secondary forests (10–70% cover);(3) nonforest (<10% cover) including savanna, agricultural land and urban areas. Maps are corrected using finer spatial resolution Landsat TM imagery.[18,34,56]

2.1.3 *The Landsat Pathfinder Tropical Forest Inventory Project*

This project was conducted by the National Aeronautics and Space Administration (NASA).[66] It aimed to provide maps of the rates of deforestation in the tropics through the intensive use of high-spatial-resolution satellite data from three epochs; early 1970s, mid-1980s and mid-1990s. Coverage was of three regions containing 75% of the world's tropical forests and the source of information was wall-to-wall coverage of high-spatial-resolution data (Landsat MSS (Multispectral Scanning System) and TM). Data processing combines automated classification and visual image interpretation. Results are manually edited and merged into a seamless database. The three epochs' interpretation are fed into a GIS and deforestation rates are computed.

2.1.4 *The Global Resource Information Database Approach*

This is the environmental data support component of the Global Environment Monitoring System (GEMS) at the United Nations Environment Programme (UNEP). The mission of Global Resource Information Database (GRID) is to provide a globally comprehensive data base of current tropical forest cover, derived from the analysis of NOAA AVHRR imagery. The proposal is to update the map at least every 3–5 years. The analysis of the imagery and the construction of the database is being undertaken in collaboration, both formally and informally, with 12 research/educational organizations in Brazil, Europe and the USA.[67]

2.1.5 *The Large-scale Biosphere Atmosphere Experiment*

The Large-scale Biosphere Atmosphere (LBA) in Amazonia is an international research initiative led by Brazil. The land use and land cover change component of this experiment is interested in the spatial pattern of forest cover and its change. Current remote-sensing algorithms and techniques on remotely sensed data acquired by a broad constellation of satellite sensors will be used to provide such information. Further development of algorithms to process microwave sensor images will also be conducted (LBA; http://daacl.esd.ornl.gov/lba_cptec).

2.2 Fire Occurrence

Remotely sensed data may be used to provide information on biomass burning which can indicate the magnitude and temporal dynamics of forest cover change through deforestation by burning events. The occurrence of biomass burning is indicated by the presence of active fires, burn scars and smoke plumes, all of which are detectable via remote sensing. There are many examples of the use of remote sensing to observe smoke plumes over tropical forests (e.g. Hellfert and Lulla).[68] However, in terms of measuring changes in forest cover it is important to be able to detect when and where and how much of an area of forest has been burnt. These variables are best estimated through active fire detection which provides the temporal and locational information and burn area estimation which allows the quantification of change in forest cover through burning.[69] Inferences of fire activity type may also be made from such information, which in turn is indicative of the mechanisms promoting forest change. For example, regular patterns of fire activity are indicative of organized clearance,[51] linear fire patterns may suggest the movement of the fire front or road construction[70,71] and multitemporal analysis detecting small, scattered and repeated fire occurrences in one particular area may indicate the rotation cycles for shifting cultivation.[72]

Theoretically, active fires may be detected by any remote sensor carrying an MIR or TIR channel. Active fires emit radiation vigorously in these wavelengths, providing a signal that appears widely divergent from its surroundings. The peak spectral emission associated with forest fires varies from between 8–12 μm for a cool forest fire (at 500 K), 2.9 μm for a forest fire (at 1000 K) to 1.6 μm for a maximum heat fire (1800 K), thus indicating that at these wavelengths the best contrast between a fire pixel and a nonfire pixel will be obtained.[73] Detectors using wavelengths between 1.5 and 3.0 μm, i.e. those carried on the Landsat TM sensor, face problems of solar contamination, for example from sunglint from cloud tops. Consequently, it has been the AVHRR sensor, and more recently the ERS-2 ATSR sensor, which measure radiation at 3.53 μm, that has been used for detecting active deforestation fires (e.g. Malingreau et al.,[74] Matson and Holben,[75] Pereira and Setzer,[76] Perrin and Milington,[77] Wooster et al.[78]). The general approach is to set a threshold in the 3.53 μm channel with additional thresholds set in other channels to avoid confusion with high-reflective areas. Because of changes in fire size and background temperature, the setting of these thresholds is best done by ecological region,[69] which can be aided by the use of a contextual classification approach.[79] The temporal resolution of the NOAA AVHRR satellite sensor also makes it an attractive proposition for measuring active fires and its daily revisit cycle increases the likelihood of fires in action being recorded. There are problems, however, with using AVHRR data. The 1.1 km spatial resolution and threshold temperature leads to the saturation of the pixel by small intense fires, medium to large fires of moderate intensity and by many sorts of fires if the background temperature is sufficiently high; this may lead to an imprecise location of burning activity within the ground pixel and little indication of the impact of fire on the forest environment. Small fires may be missed which also leads to incorrect estimation of remaining forest cover. Kaufman et al.[80] estimated that flaming fires with temperatures greater than 800 K can be detected on AVHRR imagery if they occupy an area greater than 1/10 000 of the pixel, while smoldering fires with temperatures above 500 K need to occupy more than 1/1000 of the pixel surface.

Burn areas provide a spectrally distinct response from the surrounding vegetation, which can be measured using remote sensing. Reflectance in the visible part of the electromagnetic spectrum is lower than that of the forest owing to the low albedo of burn scars, and NIR reflectance falls owing to the loss of photosynthetically active material in green vegetation. SWIR reflectance of fire scars is also much lower than that of the surroundings and this part of the spectrum has been used successfully for the purpose of fire scar detection in a number of studies.[69] This radiation is currently measured by sensors such as Landsat TM and ERS ATSR-2. MIR radiation may also be used. Here the radiation acquired at MIR wavelengths is corrected for thermal emission to leave the reflected MIR, which behaves in a similar way to SWIR reflectance, its spectral neighbor.[81]

The success of spatially delimiting burn scars as a prerequisite to estimating change in forest cover by fire events is highly dependent on the spatial resolution of the sensor used. The spatial resolution of fine resolution sensors (between 20 and 80 m) is generally sufficient to capture the spatial pattern of burn scars. However, at coarser spatial resolutions the burn scars are unlikely to be satisfactorily detected. There is the additional consideration, however, that the spectral distinctness of burn scars will diminish over time, with prompt vegetation growth and removal of the ash layer by wind. This has led to the suggestion that a multisensor approach should be adopted in which regional burnt area estimates from coarse spatial resolution data are calibrated on the basis of a sample of fine spatial resolution estimates of burnt areas.[69]

3 TROPICAL FOREST TYPES

The forest cover delineated from nonforest cover during the first level of resource information extraction will

Table 2 Review of the interrelated factors determining the spectral response from different ages of regenerating tropical forests

Spectral wavelengths	Regeneration age of 0–2 years	Regeneration age of 3–7 years	Regeneration age of 8+ years
Visible	Strong pigment absorption	Moderate pigment absorption	Asymptotic level of pigment absorption
NIR	High reflection from physiological components of leaves within the canopy	Shadows and multiple scattering within the canopy decreases NIR radiation	Shadows and multiple scattering within the canopy decreases NIR radiation
SWIR	Strong canopy water and soil water absorption	Strong canopy water absorption and sensitivity to canopy geometrical changes	Sensitivity to canopy geometrical changes and canopy water absorption
MIR	Mixture of strong canopy water and increasing canopy transpiration	Strong canopy water absorption and increasing canopy transpiration. Increasing sensitivity to canopy geometrical changes	Increasing canopy transpiration, coupled with increasing canopy shadows
TIR	Increasing canopy transpiration	Increasing canopy transpiration	Increasing canopy transpiration, coupled with increasing canopy shadows

comprise a number of different forest types. Remotely sensed data have been used to enhance resource detail and to provide a resource assessment which relies on the knowledge of forest type. Such information, based on, for example, regeneration age, structural composition and phenological characteristics of tropical forests, is important for the assessment of resource quantity. Different forest types lend themselves to different economic uses, e.g. fuelwood versus construction timber, as well as being of a differing conservational value. The accurate measurement of forest regeneration age is also pertinent for the development of regional and global carbon budgets[82] and the provision of information on tropical forest recovery after loss. Furthermore, since forest age may be used to infer biophysical properties such as biomass and basal area,[83,84] as well as other variables such as canopy structure, roughness and species composition,[85–87] its measurement is important for tropical forest resource quality assessment.

Studies exploring the remotely sensed response from tropical forests of different ages have demonstrated the capabilities of remote sensing to derive regeneration age class maps of tropical forests, at least up to the point where the primary and secondary forest spectral responses ''blend''.[88] Age class maps of tropical forests via remote sensing have been derived using one of three approaches. One has utilized the resolved relationships between remotely sensed response and age. Mausel et al.,[89] McMorrow,[90] Steininger[91] and Boyd et al.[92] have all shown that the combined effect of interrelated factors such as pigment absorption, canopy water content absorption, multiple scattering within the canopy, shadows within the canopy, canopy transpiration and understorey reflectance[93] will determine the radiation between visible and TIR wavelengths observed from the regenerating forest canopy. The influence of each of these factors will change in importance as the forest canopy ages, increasing in structural and geometrical complexity and developing a multilayered canopy.[89] Table 2 reviews the factors determining the spectral response from regenerating tropical forests of different ages.

On the basis of such information outlined in Table 2 and preliminary analyses, Lucas et al.[94] formulated a method for mapping of four stages of forest regeneration across the Legal Amazon using NOAA AVHRR imagery. The spectral response of each pixel in the visible, NIR and MIR channels was examined and a regeneration stage allocated on the basis of the following criteria. Stage one of regeneration was assigned to those pixels having a MIR response higher than the mean MIR response associated with mature forests. Stages II and III of regeneration were allocated to those pixels having a lower MIR response than the mean for mature forests and a maximum NIR response. The division between the two stages was based on the average visible response; pixels for which the visible response was less than and greater than the average were categorized as stage II and stage III, respectively. Stage IV was assigned to pixels having a NIR response below the mean for proximal mature forests. Accuracy assessment of the derived age class map was conducted using coregistered Landsat TM imagery at a number of sites across the Legal Amazon as reference data. The confusion matrices constructed confirmed a relatively high degree of correspondence for all sites.

Another approach to deriving age class maps has been to classify remotely sensed images of regenerating tropical forests into age classes via a supervised classification. Foody et al.[95] adopted the maximum likelihood method

to classify a Landsat TM image of an area near Manaus, Brazil into six regenerating forest age classes and achieved a high level of class separability. Of the classification errors observed, most were found to be associated with the youngest forest age classes. At the study site, two main successional pathways during regeneration were followed and spectral differences between areas of forest of the same age but on different pathways were most apparent with the younger forest. Splitting the regenerating forests by successional pathway was found to increase classification, with an accuracy of over 90% obtained for the resultant 11 class classifications.

The last approach has been to use a time series of coregistered, multispectral images (usually Landsat TM, SPOT HRV and/or Landsat MSS data).[96,97] The example of Lucas et al.[98] can be used to illustrate the approach. The forests comprising an area north of Manaus, Brazil were aged using subset and georegistered Landsat TM imagery acquired in 1985, 1988, 1989 and 1991. By using statistics from over 60 training sites representing known land cover, and applying a minimum distance classification to the red, NIR and SWIR channels to each of the Landsat TM subsets, the broad categories of primary forest, regenerating forest and pasture were classified for each date. A majority filter was then applied to each classified image to ensure that isolated pixels were reallocated to the surrounding class. Changes in land cover at each pixel location over this six-year period were then documented to provide information on the approximate age of regenerating forests. A model of changes in land cover at Manaus, based on the multitemporal analysis is shown in Figure 4. The seven categories of land cover (five categories of regenerating forests, plus pasture and primary forest) were mapped using a GIS to produce an age class map of the forests at the study site. Ideally, the multitemporal images are acquired yearly to map regenerating forest age accurately. The image sequences, however, often have gaps due to the difficulties of obtaining cloud-free scenes, and costs and temporal discontinuities between satellite sensors. It must be remembered that this missed information could introduce unknown errors into the mapping of forest age.[88]

A number of studies have also separated forest types defined on the basis of structural and bioclimatic attributes based largely on the degree of the forests' canopy closure, see Table 3 for examples. Roy et al.[99] and Saxena et al.[100] have explored the forest types to be identified using remotely sensed data by both visual and supervised classification techniques. An analysis of the cost effectiveness of these techniques revealed the latter to be a more economical proposition. However, consideration must be given to the availability of funds, facilities and accuracy of resource information required. Other studies have concentrated on purely digital classification

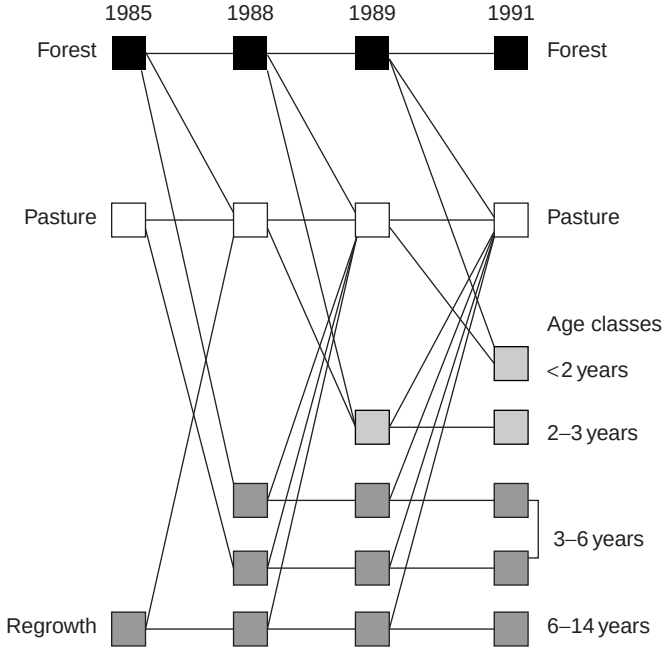

Figure 4 Age class model tracing the main changes in land cover for the study area between 1985 and 1991. (Reproduced by permission of Taylor & Francis from Lucas et al., 1993.[98])

techniques with refinements to improve the accuracy of classifications. For example, a stratified approach was adopted by Sudhaker et al.[103] Here forest vegetation was masked from plantations/social forest areas and classifications conducted in isolation and then composited. Classification accuracy also could be increased through the use of contextual and ancillary information, such as historical land-use information, site characteristics (i.e. topography) and geobotanical relationships (e.g. Paradella et al.,[104] Tuomisto et al.,[105] Brondizio et al.[102]).

The ability to map forest types using remotely sensed data could also be enhanced using interannual multitemporal data. An interannual multitemporal data set offers the potential for any temporal alteration in the spectral response from different forest types as a result of phenological change (e.g. leaf shedding, canopy greeness and senescence) to be exploited within a classification procedure. Forest types such as dense evergreen, dense seasonal and tropical deciduous have been distinguished in this way (e.g. Achard and Estreguil.[106]). At the continental and global scales the temporal variation in NOAA AVHRR NDVI has been used to map a range of vegetation types including a number in the tropical forest biome.[107,108]

The heterogeneity of the tropical forest cover types and the highly complex spectral response from them account for the limited number of forest types currently identifiable by remote sensing. An improvement in spatial, spectral and radiometric resolution of remote sensors will

Table 3 Examples of remote sensing studies with the aim of classifying different forest types

Study Location Satellite sensor	Roy et al.[99] Adaman Islands, India Landsat TM		Saxena et al.[100] Southwest India Landsat TM		Foody and Hill[101] Southeast Peru Landsat TM	Brondizio et al.[102] Pará, Brazil Landsat TM	Sudhakar et al.[103] Bengal, India IRS-1A LISS II
Methodology	Visual interpretation of false-color composite image based on tone/color, texture, pattern and association	Supervised maximum likelihood classification based on red, NIR and SWIR channels	Supervised maximum likelihood classification based on green, red, NIR and SWIR channels		Supervised maximum likelihood classification based on NIR and SWIR channels	Unsupervised and supervised classification using a hybrid maximum likelihood/texture algorithm based on green, red, NIR and SWIR channels informed by vegetation inventories and field-based studies documenting historical land use.	Supervised maximum likelihood classification on stratified imagery
Forest types identified	*Structural:* 1. Evergreen 2. Semievergreen 3. Moist deciduous 4. Mangrove (old) 5. Mangrove (young) 6. Plantation 7. Bamboo 8. Nonforest	*Structural:* 1. Evergreen 2. Semievergreen 3. Moist deciduous 4. Mangrove (old) 5. Mangrove (young) 6. Plantation 7. Bamboo 8. Degraded forest 9. Nonforest	*Detailed structural:* 1. Evergreen 2. Moist deciduous (sparse crown cover) 3. Moist deciduous (dense crown cover) 4. Open forest 5. Rubber plantation 6. Eucalyptus plantation 7. Teak plantation 8. Scrub (dense) 9. Scrub (open) 10. Cropped area 11. Fallowed area	*Bioclimatic:* 1. Evergreen 2. Moist deciduous 3. Dry deciduous 4. Open forest 5. Plantation 6. Grass/Scrub 7. Cropped area 8. Fallowed area	*Ecological groupings:* 1. Agriculture, bamboo, lower floodplain and premontane forests 2. Permanently and seasonally waterlogged swamp forest 3. Middle upper floodplain, terra firme forests, plateau forests	*Structural:* 1. Upland forest 2. Floodplain forest 3. Açaì Palm forest 4. Mangrove 5. Pasture 6. Pasture with palm 7. Initial secondary succession 8. Intermediate secondary succession 9. Advanced secondary succession 10. High savanna 11. Low savanna 12. Imperata savanna	*Bioclimatic:* 1. Alpine scrubs grassland 2. Subtropical broadleaved forest 3. Temperate coniferous forest 4. Mixed forest 5. Sal forest 6. Sal plantation 7. Tea gardens
Overall classification accuracy (%)	87.21	96.84	84.2	85.78	94.84	94.1	85.0

enhance the potential detail to be exploited. However, it must be recognized that while the total information content of remotely sensed data is determined by such prefixed sensor attributes, thematic details gained from the data depend on the interpretation techniques used for information extraction.[100]

4 TROPICAL FOREST PROPERTIES

The measurement of tropical forest biophysical properties (e.g. leaf and wood biomass) provides an indication of resource quality, as well as resource management strategy information. In regions of southeast Asia and Africa, shifting cultivation and selective logging have resulted in structural alteration of the forest rather than wholescale clearance (Green and Sussman;[24] Gilruth et al.[109]) and these would be best detected through biophysical property estimation. Plantation development, site productivity and the process of natural forest regeneration after deforestation, for example, can also be inferred from biophysical properties.

Two basic approaches can be employed to relate remotely sensed data to biophysical variables. In physical modeling, the canopy radiative transfer processes are simulated mathematically with valuable insights into the fundamental factors driving the relationships between remotely sensed data and vegetation biophysical properties gained.[110] However, their use for tropical forest resource assessment is presently hindered by factors such as the heterogeneity of the canopy, the dynamic characteristics of the canopy optical properties and external effects, such as atmospheric path radiance and absorption. An alternative and more operational approach is empirical modeling, whereby the quantitative relationship between remotely sensed data and various derivatives (e.g. vegetation indices) and ground-based biophysical property data is calibrated by interrelating known points of coincident observation of the two. Often statistical regression procedures are used. The use of radiometrically and atmospherically calibrated remotely sensed data to develop an empirical model affords its application at other spatial and temporal resolutions.

A fundamental problem often faced by researchers studying tropical forests is the lack of appropriate ground data which may be brought together with the remotely sensed data.[111] This may account for the limited number of studies that have been conducted which focus on the estimation of biophysical properties of tropical forests using remotely sensed data. For those few studies that have actual biophysical data, initial attention has been placed on the use of remotely sensed radiation acquired in the visible and NIR wavelengths. In visible wavelengths, absorption by leaf pigments is large and therefore there is little reflection, while in NIR wavelengths leaf scattering is large and, therefore, reflection from the canopy is also large. Shadow decreases canopy reflectance in both visible and NIR wavelengths. As a consequence, leaf biomass would be expected to exhibit a negative relationship with visible radiation and either no or a positive relationship with NIR radiation.[111] Improved relationship strengths are obtained when the visible and NIR radiation are combined in the NDVI, which has been shown to be directly related to the leaf area index of temperate forests,[112] as well as a number of other vegetation types. Generally, however, researchers have reported weak relationships between the NDVI and a range of tropical forest biophysical properties, including Net Primary Productivity (NPP),[113] stand structure (e.g. height, diameter of main stem) and biomass,[114] and tree density and basal area.[115] This weak relationship may be a function of a number of factors.[116] First, the asymptotic relationship, with the sensitivity of the NDVI to biophysical properties seen to decline with large amounts of vegetation. Second, there is attenuation by the large atmospheric water and aerosol load above tropical forests. Third, low reflectance in the red and NIR wavelengths is observed from tropical forest canopies and fourth is the ecological and physical complexity of tropical forest environments. Subsequent studies have therefore concentrated on the validity of using remotely sensed data acquired in other parts of the electromagnetic spectrum for the estimation of tropical forest biophysical properties, namely, the infrared and microwave wavelengths.

The suitability of using MIR radiation for the estimation of tropical forest biophysical properties has been advocated. MIR radiation acquired by the NOAA AVHRR sensor, either alone or within vegetation indices, provides stronger relationships with tropical forest biophysical properties than visible and NIR radiation and derivatives of these.[117] Moreover, within tropical environments, MIR radiation offers favorable atmospheric penetration capabilities. It is able to penetrate to a greater depth through anthropogenic or organic smoke particles than visible or NIR radiation.[118] Contemporary thinking assigns the interaction between the forests and MIR radiation to a number of interrelated factors, principally the water content, surface temperature and structure and roughness of the canopy.

An increase in forest amount, represented by biophysical properties such as biomass and tree density, will lead to a decrease in the MIR radiation from the forest, resulting in an inverse relationship between the two variables.[116] It has been suggested, however, that the use of MIR radiation at and beyond the regional scale may be unreliable. MIR radiation is composed of a mixture of reflected and emitted radiation and, at these scales, the emitted component may be more sensitive to external

confounding variables than to the intrinsic properties of the earth's surface. Emitted radiation, though related to forest canopy properties, may also be subject to varying soil moisture conditions, atmospheric conditions such as wind speed and air vapor conductance, as well as to site specific factors such as topography and aspect. Indeed, it is factors such as these that have been related more strongly to tropical forest canopy thermal response than to biophysical properties such as basal area and tree density.[119] It may be preferable, therefore, to correct for thermal emission and use only the reflected component of the MIR radiation.[118]

The reflected component of MIR radiation from a vegetation canopy is hypothesized to be principally a function of the water content of the canopy. An increase in vegetation amount will be accompanied by an increase in the amount of liquid water in the canopy, absorbing the incident radiant component of MIR radiation.[118] Canopy structure, generally becoming more complex as vegetation amount increases, is also believed to be an important factor determining the reflected component of MIR radiation from a vegetation canopy. The vertical structure of the canopy has an effect on radiation absorption, by trapping incoming photons thereby producing shadows which will decrease the intensity of the reflected component of MIR radiation. So although two vegetation types may have the same leaf reflectance in the MIR spectral region, their overall reflectance will be a function of the canopy structure which may differ between the two vegetation types.[118]

An exploratory study by Boyd et al.[120] has revealed that correcting MIR radiation for thermal emission, to derive MIR reflectance, increases the strength of the relationship between radiation acquired in MIR wavelengths and total forest biomass of West African tropical forests. The use of MIR reflectance, either alone or within the vegetation index VI3 ((NIR − MIR reflectance)/(NIR + MIR reflectance)), provided the strongest relationship with total forest biomass. This suggests that MIR reflectance may be more sensitive to changes in forest properties than the reflectance in visible and NIR wavelengths and should be used to estimate the biomass of tropical forests.

The use of SAR remote sensing to estimate biophysical properties of tropical forests has also shown promise. It has been shown both experimentally and by theoretical modeling that the penetration of microwaves into the forest is dependent primarily upon wavelength, with a few basic categories evident. Short-wavelength (e.g. X-band ≈ 3 cm) microwaves interact with the surface of the canopy promoting scattering on the leaves, twigs or branches; medium-wavelength (e.g. C-band ≈ 6 cm) microwaves interact with the volume of the canopy/trunk volume and longer-wavelength (e.g. L-band ≈ 22 cm) microwaves penetrate the canopy, interacting with the ground/trunk.[111,121] The depth of microwave penetration is further dependent on the angle of incidence of the sensor and canopy openness[122,123] and moreover on the polarization of the microwave.[124] Many SARs can both transmit and receive microwaves at two polarizations and this enhances the information provided, particularly on surface roughness and geometric regularities in the forest stand.[125] Through each of these interaction types, specific biophysical properties, most commonly above-ground biomass and canopy roughness, of forests may be estimated, at least up to certain thresholds. The values reported for broad-leaved evergreen forests are 20 t (tonnes) ha^{-1} for C-band and 40 t ha^{-1} for L-band under similar conditions,[126] though the use of cross-polarization microwaves and differing viewing geometries may extend that.

The suitability of SAR for tropical forest studies is determined by this saturation level, which is well below the biomass attained for mature tropical forests. Research has revealed that L-band microwaves can be used to discriminate between different levels of regenerating forest biomass (up to approximately 20 years old, 60 t ha^{-1}) and that cross-polarized backscatter is more sensitive to changes in biomass density than monopolarized backscatter. The use of C-band SAR imagery is limited to differentiating between very-low-biomass or clear-cut areas and those with some vegetation when it is dry.[127–129] Improvements in the strength of the relationship between microwave data and biomass have been attained using backscatter ratios (e.g. LHV/LHH) and, moreover, by stratifying the biomass data by forest type.[130] SAR data could also be suitable for estimating biophysical properties of plantation forests. Rosenqvist[121] has obtained strong relationships between LHH backscatter measured by the JERS-1 SAR sensor and the height of deciduous rubber trees, and furthermore between the CVV backscatter measured by the ERS-1 AMI (Active Microwave Instrumentation) sensor and LHH backscatter measured by JERS-1 SAR and the leaf area index of oil palm stands.

Once an empirical model has been developed between remotely sensed data and a tropical forest biophysical property, that model may be extrapolated across a larger desired area with the use of radiometrically and atmospherically corrected remotely sensed data to provide a map delimiting forests of classes of biophysical properties. This has been conducted for southern Cameroon[131] where a map of the carbon content of the forest was produced on the basis of basal area classes. Initial image processing used an unsupervised classification of NOAA AVHRR imagery in channels 1, 2 and 3, guided by field experience and available baseline data, to delineate the regenerating

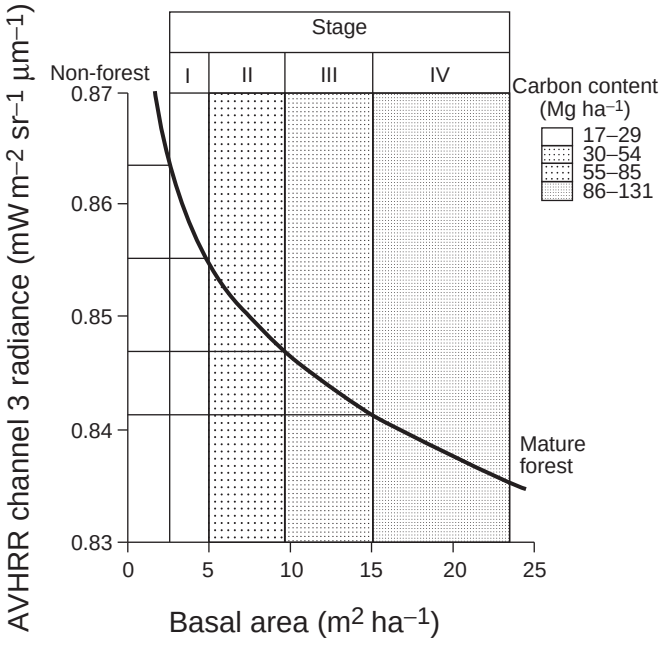

Figure 5 Regression model between basal area and NOAA AVHRR MIR radiance used to determine four main stages of tropical forest regeneration with their associated carbon content.

forests from mature forests. A significant negative logarithmic regression model derived between NOAA AVHRR channel 3 radiance and the biophysical property of basal area of tropical forests measured during a field campaign (Figure 5) was then applied to imagery of the area. From this exercise a map depicting the spatial extent of four classes of stages of regenerating forests (equating with four ranges of basal area estimated by the regression model), mature forest and nonforest was produced.

5 FUTURE CAPABILITIES

As tropical forest resources face increasing pressure so the need to acquire more accurate and timely information about their current state increases. This task will be much aided by the planned launch of new spaceborne sensors, such as Earth Observing System (EOS) AM-1 MODIS (Moderate Resolution Imaging Spectrometer), Landsat ETM (Enhanced Thematic Mapper), SPOT Vegetation, SPOT-Radar SAR and Polar Orbiting Environmental Mission (POEM) Envisat Advanced Along Track Scanning Radiometer (AATSR), which will allow the user to choose a sensor that will acquire the remotely sensed data that is best suited for the tropical forest resource information required. Through advances in technology, future sensors will have improved spectral, spatial, radiometric and temporal properties over current sensors. Moreover, more sensors will have multiple view angle capabilities. Access to remotely sensed data will also increase, with a multitude of data products which have undergone different levels of processing abounding. Different data products will enable users to choose products available at a particular processing level, enabling them to work at coarse spatial resolutions to study global resources or at finely detailed resolutions allowing the study of resources in targeted regions. With the launch of such spaceborne sensors remote sensing will continue to be a crucial tool in the assessment of tropical forest resources.

ABBREVIATIONS AND ACRONYMS

AATSR	Advanced Along Track Scanning Radiometer
AMI	Active Microwave Instrumentation
ANN	Artificial Neural Network
AVHRR	Advanced Very High-resolution Radiometer
EOS	Earth Observing System
ERS	European Remote Sensing Satellite
ETM	Enhanced Thematic Mapper
FAO	United Nations Food and Agricultural Organization
GAC	Global Area Coverage
GEMS	Global Environment Monitoring System
GIS	Geographical Information System
GRFM	Global Rain Forest Mapping Project
GRID	Global Resource Information Database
HRV	High-resolution Visible
HRV XS	High-resolution Visible Multispectral
IRS	Indian Remote Sensing Satellite
JERS	Japanese Earth Resources Satellite
JRC	Joint Research Center
LAC	Large Area Coverage
LBA	Large-scale Biosphere Atmosphere
MIR	Middle-infrared
MODIS	Moderate Resolution Imaging Spectrometer
MOMS	Modular Optoelectronic Multispectral Scanner
MSS	Multispectral Scanning System
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency of Japan
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NOAA	National Oceanic and Atmospheric Administration
NPP	Net Primary Productivity

POEM	Polar Orbiting Environmental Mission
SAR	Synthetic Aperture Radar
SIR	Spaceborne Imaging Radar
SPOT	System Probatoire d'Observation de La Terre
SWIR	Shortwave-infrared
TIR	Thermal-infrared
TM	Thematic Mapper
TREES	Tropical Ecosystem Environment Observations by Satellites
UNEP	United Nations Environment Programme

RELATED ARTICLES

Remote Sensing **(Volume 10)**
Global Land Databases for Environmental Analyses • Land Cover Assessment and Monitoring • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring • Temperate Forest Resource Assessment by Remote Sensing

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. N. Myers, *Conservation of Tropical Moist Forests*, National Academy of Sciences, Washington, DC, 1980.
2. J.P. Malingreau, M.M. Verstraete, F. Achard, 'Monitoring Global Deforestation: A Challenge for Remote Sensing', in *TERRA-1: Understanding the Terrestrial Environment*, ed. P. Mather, Taylor and Francis, London, 203–209, 1992.
3. R. Baltaxe, 'The Application of Remote Sensing to Tropical Forest Cover Monitoring, A Review of Practices and Possibilities', in *Remote Sensing Yearbook*, eds. A.P. Cracknell, L. Hayes, Taylor and Francis, London, 33–51, 1987.
4. A. Grainger, 'Rates of Deforestation in the Humid Tropics, Estimates and Measurements', *Geographical J.*, **159**, 33–44 (1993).
5. S.A. Sader, T.A. Stone, A.T. Joyce, 'Remote Sensing of Tropical Forests, an Overview of Research and Applications Using Non-photographic Sensors', *Photogrammetric Eng. Remote Sensing*, **56**, 1343–1351 (1990).
6. D.M. Stoms, J.E. Estes, 'A Remote Sensing Research Agenda for Mapping and Monitoring Biodiversity', *Int. J. Remote Sensing*, **14**, 1839–1860 (1993).
7. D.E. Wickland, 'Future Directions for Remote Sensing in Terrestrial Ecological Research', in *Theory and Applications of Optical Remote Sensing*, ed. A. Asrar, John Wiley & Sons, New York, 691–724, 1989.
8. J.S. Miguel-Ayanaz, G.S. Biging, 'An Iterative Classification Approach for Mapping Natural Resources from Satellite Imagery', *Int. J. Remote Sensing*, **17**, 957–981 (1996).
9. S.W. Running, C.O. Justice, V. Salomonson, D. Hall, J. Barker, Y.J. Kaufman, A.H. Strahler, A.R. Huete, J.-P. Muller, V. Vanderbilt, Z.M. Wan, P. Teillet, D. Carneggie, 'Terrestrial Remote Sensing Science and Algorithms Planned for EOS/MODIS', *Int. J. Remote Sensing*, **15**, 3587–3620 (1994).
10. J.P. Malingreau, 'Global Vegetation Dynamics: Satellite Observations over Asia', *Int. J. Remote Sensing*, **7**, 1121–1146 (1986).
11. M.M. Verstraete, B. Pinty, 'Extracting Surface Properties from Satellite Data in the Visible and Near-infrared Wavelengths', in *TERRA-1: Understanding the Terrestrial Environment*, ed. P.M. Mather, Taylor and Francis, London, 203–209, 1992.
12. T.M. Lillesand, R.W. Kiefer, *Remote Sensing and Image Interpretation*, 3rd edition, John Wiley & Sons, New York, 1994.
13. P.J. Curran, 'Multispectral Remote Sensing for the Estimation of Green Leaf Area Index', *Philos. Trans. R. Soc. London, Ser. A*, **309**, 257–270 (1983).
14. J.P. Malingreau, A.S. Belward, 'Scale Considerations in Vegetation Monitoring Using AVHRR Data', *Int. J. Remote Sensing*, **13**, 2289–2307 (1992).
15. IGBP, 'The International Geosphere-biosphere Programme, A Study of Global Change. The Initial Core Projects', International Geosphere-biosphere Programme, Stockholm, 1990.
16. K.B. Kidwell, 'NOAA Polar Orbiter Data TIROS-N, NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11 & NOAA-12. Users Guide', NASA, Washington, DC, 1991.
17. H.A. Downton, 'Measuring Tropical Deforestation: Development of Methods', *Environ. Conservation*, **22**, 229–240 (1995).
18. P. Mayaux, F. Achard, J.P. Malingreau, 'Global Tropical Forest Area Measurements Derived from Coarse Resolution Satellite Imagery: A Comparison with Other Approaches', *Environ. Conservation*, **25**, 37–52 (1998).
19. H.A. Mooney, R.J. Hobbs, 'Introduction', in *Remote Sensing of Biosphere Functioning*, eds. H.A. Mooney, R.J. Hobbs, Springer-Verlag, New York, 1–4, 1989.
20. J.P. Lanley, *Tropical Forest Resources, F.A.O. Forestry Paper 30*, United Nations, Food and Agriculture Programme Publications, Rome, 1992.
21. G.M. Woodwell, J.E. Hobbie, R.A. Houghton, J.M. Melillo, B. Moore, A.B. Park, B.J. Peterson, G.R. Shaver, 'Measurement of Changes in Vegetation of the Earth by Satellite Imagery', in *The Role of Vegetation in the Global Carbon Cycle, SCOPE 23*, ed. G.M. Woodwell, New York, John Wiley & Sons, 221–240, 1984.

22. G.M. Woodwell, R.A. Houghton, T.A. Stone, R.F. Nelson, W. Kovalick, 'Deforestation in the Tropics: New Measurements in the Amazon Basin Using Landsat and NOAA/AVHRR Imagery', *J. Geophys. Res.*, **92**, 2157–2163 (1987).
23. A.T. Tardin, R. Cunha, 'Evaluation of Deforestation in the Legal Amazon Using Landsat TM Images', INPE Publication 5015-RPE/609, INPE, São Paulo, 1990.
24. G.M. Green, R. Sussman, 'Deforestation History of the Eastern Rainforests of Madagascar from Satellite Images', *Science*, **248**, 212–215 (1990).
25. J.P. Malingreau, C.J. Tucker, 'Large Scale Deforestation in the Southeastern Amazon of Brazil', *Ambio*, **17**, 49–55 (1988).
26. J.C. Gervin, A.G. Kerber, R.G. Witt, Y.C. Lu, R. Sekhon, 'Comparison of Level 1 Land Cover Classification Accuracy for MSS and AVHRR Data', *Int. J. Remote Sensing*, **6**, 47–57 (1985).
27. T. Suzuki, M. Shimada, 'Japanese Earth Observation Satellite Program and Application of JERS-1 Sensors Data to Forest Monitoring', World Forest Watch Conference on Global Forest Monitoring, INPE, São Paulo, **47**, 1992.
28. H.D. de Groof, J.P. Malingreau, A. Sokeland, F. Achard, 'A First Look at ERS-1 Data over the Tropical Forest of West Africa', World Forest Watch Conference on Global Forest Monitoring, INPE, São Paulo, **19**, 1992.
29. C.J. Tucker, B.N. Holben, T.E. Goff, 'Intensive Forest Clearing in Rondonia, Brazil, as Detected by Satellite Remote Sensing', *Remote Sensing Environ.*, **15**, 255–261 (1984).
30. R. Riou, F. Seyler, 'Texture Analysis of Tropical Rain Forest Infrared Images', *Photogrammetric Eng. Remote Sensing*, **63**, 515–521 (1997).
31. G. D'Souza, J.P. Malingreau, H.D. Eva, 'Tropical Forest Cover of South and Central America as Derived from Analyses of NOAA AVHRR data', TREES Series B: Research Report no. 3, European Commission Publication, Brussels, 1995.
32. S. Amaral, 'Deforestation Estimates in NOAA/AVHRR and TM/Landsat Images for a Region in Central Brazil', *Int. Arch. Photogrammetry and Remote Sensing*, Commission VII, **29**, 764–767 (1992).
33. N. Laporte, C. Justice, J. Kendall, 'Mapping the Dense Humid Forest of Cameroon and Zaire Using AVHRR Satellite Data', *Int. J. Remote Sensing*, **16**, 1127–1145 (1995).
34. J.P. Malingreau, J.P. Tucker, N. Laporte, 'AVHRR for Monitoring Global Tropical Deforestation', *Int. J. Remote Sensing*, **10**, 855–867 (1989).
35. A. Cross, 'Tropical Forest Monitoring Using AVHRR, Towards an Automated System for Change Detection', Final Report to UNEP-GRID, Swindon: Natural Environmental Research Council, 1991.
36. K.D. Grover, S. Quegan, 'Studying Forestry in Brazilian Amazonia Using SAR', Proceedings of the 22nd Annual Conference of the Remote Sensing Society, Nottingham, Remote Sensing Society, 476–484, 1996.
37. S.S. Saatchi, J.V. Soares, D.S. Alves, 'Mapping Deforestation and Land Use in Amazon Rainforest by Using SIR-C Imagery', *Remote Sensing Environ.*, **59**, 191–202 (1997).
38. R. Paivinen, R. Witt, 'Application of NOAA/AVHRR for Tropical Forest for Forest Inventory and Monitoring', Department of Forest Mensuration and Research Notes, **21**, University of Helsinki, Helsinki, 163–170, 1988.
39. R.G. Congalton, 'A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data', *Remote Sensing Environ.*, **37**, 35–46 (1991).
40. R. Nelson, N. Horning, 'AVHRR-LAC Estimates of Forest Area in Madagascar, 1990', *Int. J. Remote Sensing*, **14**, 1463–1474 (1993).
41. J.P. Malingreau, F. Achard, G. D'Souza, H.J. Stibig, J. D'Souza, C. Estreguil, H. Eva, 'AVHRR for Global Tropical Forest Monitoring, The Lessons of the TREES Project', *Remote Sensing Rev.*, **12**, 29–40 (1995).
42. J. Conway, 'Evaluating ERS-1 SAR Data for the Discrimination of Tropical Forest from Other Tropical Vegetation Types in Papua New Guinea', *Int. J. Remote Sensing*, **18**, 2967–2984 (1997).
43. F. Achard, F. Blasco, 'Analysis of Seasonal Evolution and Mapping of Forest Cover in West Africa with use of NOAA AVHRR HRPT Data', *Photogrammetric Eng. Remote Sensing*, **56**, 1359–1365 (1990).
44. A.P. Cracknell, 'Synergy in Remote Sensing – What's in a Pixel', *Int. J. Remote Sensing*, **19**, 2015–2048 (1998).
45. E. Rignot, W. Salas, D. Skole, 'Mapping Deforestation and Secondary Growth in Rondonia, Brazil, Using Imaging Radar and Thematic Mapper Data', *Remote Sensing Environ.*, **59**, 167–179 (1997).
46. E. Nezry, E. Mougin, A. Lopes, J.P. Gastellu-Etchegorry, Y. Laumonier, 'Tropical Vegetation Mapping with Combined Visible and SAR Spaceborne Data', *Int. J. Remote Sensing*, **14**, 2165–2184 (1993).
47. R. Nelson, B. Holben, 'Identifying Deforestation in Brazil Using Multiresolution Satellite Data', *Int. J. Remote Sensing*, **7**, 429–448 (1986).
48. H. Jeanjean, A. Husson, 'Saemeo-France Project, Detailed Studies with SPOT Data and Global Perception with NOAA Data', Training in Remote Sensing Applications to Deforestation in Southeast Asia, Centre National D'Etudes Spatial CNES, Paris, 34–49 (1991).
49. H. Jeanjean, F. Achard, 'A New Approach for Tropical Forest Area Monitoring Using Multiple Spatial Resolution Satellite Sensor Imagery', *Int. J. Remote Sensing*, **18**, 2455–2461 (1997).
50. J.R.G. Towshend, C.O. Justice, 'Selecting the Spatial Resolution of Satellite Sensors Required for Global Monitoring of Land Transformations', *Int. J. Remote Sensing*, **9**, 187–236 (1988).

51. J.P. Malingreau, 'Remote Sensing for Tropical Forest Monitoring', in *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, eds. A.S. Belward, C.R. Valenzuela, Kluwer Academic, Dordrecht, 253–278, 1991.
52. B.N. Holben, Y.E. Shimabukuro, 'Linear Mixture Model Applied to Coarse Spatial Resolution Data from Multispectral Satellite Sensors', *Int. J. Remote Sensing*, **14**, 2231–2240 (1993).
53. D. Skole, C. Tucker, 'Tropical Deforestation and Habitat Fragmentation in the Amazon, Satellite Data from 1978 to 1988', *Science*, **260**, 1905–1910, (1993).
54. A.M. Cross, J.J. Settle, N.A. Drake, R.T.M. Pavinen, 'Subpixel Measurement of Tropical Forest Cover Using AVHRR Data', *Int. J. Remote Sensing*, **12**, 1119–1129 (1991).
55. A.P. Cracknell, *AVHRR The Advanced Very High-resolution Radiometer*, Taylor and Francis, London, 1997.
56. P. Mayaux, E.F. Lambin, 'Estimation of Tropical Forest Area from Coarse Spatial Resolution Data, A Two Step Correction Function for Proportional Errors due to Spatial Aggregation', *Remote Sensing Environ.*, **53**, 1–15 (1995).
57. G.M. Foody, R.M. Lucas, P.J. Curran, M. Honzak, 'Mapping Tropical Forest Fractional Cover from Coarse Spatial Resolution Remote Sensing Imagery', *Plant Ecology*, **131**, 143–154 (1997).
58. N.A. Quarmby, J.R.G. Towshend, J.J. Settle, M. Milnes, T.L. Hindle, N. Silleos, 'Linear Mixture Modeling Applied to AVHRR Data for Crop Area Estimation', *Int. J. Remote Sensing*, **13**, 415–425 (1992).
59. J.J. Settle, N.A. Drake, 'Linear Mixing and the Estimation of Ground Cover Proportions', *Int. J. Remote Sensing*, **14**, 1159–1177 (1993).
60. R.M. Lucas, 'Integration of NOAA AVHRR and Fine Spatial Resolution Imagery for Tropical Forest Monitoring', in *Advances in the Use of AVHRR Data for Land Applications*, eds. J.P. Malingreau, A. Belward, Kluwer Academic, Dordrecht, 371–394, 1996.
61. C.C. Borel, S.A.W. Gerstl, 'Nonlinear Spectral Mixing Models for Vegetative and Soil Surfaces', *Remote Sensing Environ.*, **47**, 403–416 (1994).
62. J.F.L. de Moraes, F. Seyler, C.C. Cerri, B. Volkoff, 'Land Cover Mapping and Carbon Pools Estimates in Rondonia, Brazil', *Int. J. Remote Sensing*, **19**, 921–934 (1998).
63. J.B. Adams, D. Sabol, V. Kapos, R.A. Filho, D.A. Roberts, M.O. Smith, A.R. Gillespie, 'Classification of Multispectral Images Based on Fractions of Endmembers, Application to Land-cover Change in the Brazilian Amazon', *Remote Sensing Environ.*, **52**, 137–154 (1995).
64. R. Defries, M. Hansen, M. Steininger, R. Dubayah, R. Sohlberg, J.R.G. Townshend, 'Subpixel Forest Cover in Central Africa from Multisensor, Multitemporal Data', *Remote Sensing Environ.*, **60**, 228–246 (1997).
65. FAO, 'Forest Resources Assessment 1990: Survey of Tropical Forest Cover and Study of Change Processes', FAO, FAO Forestry Paper 130, Rome, Italy, 1996.
66. W. Chomentowski, B. Salas, D. Skole, 'Landsat Pathfinder Project Advances Forest Mapping', GIS World, **7**, 34–38 (1994).
67. S. Jaakkola, 'Managing Data for the Monitoring of Tropical Forest Cover: The Global Resource Information Database Approach', *Photogrammetric Eng. Remote Sensing*, **56**, 1355–1357 (1990).
68. M.R. Helfert, K.P. Lulla, 'Mapping Continental-scale Biomass Burning and Smoke Palls over the Amazon Basin as Observed from the Space Shuttle', *Photogrammetric Eng. Remote Sensing*, **56**, 1367–1373 (1990).
69. H. Eva, E.F. Lambin, 'Remote Sensing of Biomass Burning in Tropical Regions: Sampling Issues and Multisensor Approach', *Remote Sensing Environ.*, **64**, 292–315 (1998).
70. T. Qingxi, L. Jiyuan, 'Monitoring Forest Cover Change Using Remote Sensing Techniques in China', Proceedings of International Space Year (IYS) World Forest Watch (WFW) Conference, Sao Jose dos Campos, Brazil, **71**, 1992.
71. J.P. Malingreau, 'A Tropical Deforestation Alarm System', Draft Report, Joint Research Centre, Ispra, 1990.
72. E. Lambin, D. Ehrlich, 'Identification of Tropical Deforestation Fronts at Broad Spatial Scales', *Int. J. Remote Sensing*, **18**, 3551–3568 (1997).
73. J.M. Robinson, 'Fire From Space: Global Fire Evaluation Using Infrared Remote Sensing', *Int. J. Remote Sensing*, **12**, 3–24 (1991).
74. J.P. Malingreau, G. Stephens, L. Fellows, 'Remote Sensing of Forest Fires: Kalimantan and North Borneo', *Ambio*, **17**, 49–55 (1985).
75. M. Matson, B. Holben, 'Satellite Detection of Tropical Burning in Brazil', *Int. J. Remote Sensing*, **8**, 509–516 (1987).
76. M.C. Pereira, A.W. Setzer, 'Spectral Characteristics of Fire Scars in Landsat-5 TM Images of Amazonia', *Int. J. Remote Sensing*, **14**, 2061–2078.
77. M. Perrin, A. Millington, 'ATRS-2 Data for Burned Area Detection', Proceedings of the 23rd Annual Conference of the Remote Sensing Society, Nottingham, Remote Sensing Society, 165–170 (1997).
78. M.J. Wooster, P. Ceccato, S.P. Flasse, 'Indonesian Fires Observed Using AVHRR', *Int. J. Remote Sensing*, **19**, 383–386 (1998).
79. H.D. Eva, S. Flasse, 'Contextual and Multiple-threshold Algorithms for Regional Active Fire Detection with AVHRR Data', *Remote Sensing Rev.*, **14**, 333–351 (1996).
80. Y.J. Kaufman, A.W. Setzer, C.O. Justice, C.J. Tucker, M.C. Pereira, I. Fung, 'Remote Sensing of Biomass Burning in the Tropics', in *Fire in the Tropical Biota*, ed. J.G. Goldammer, Springer-Verlag, Berlin, 400–417, 1990.

81. J.M.C. Pereira, 'A Comparative Evaluation of NOAA/AVHRR Vegetation Indices for Burned Surface Detection and Mapping', *IEEE Trans. Geosci. Remote Sensing*, **37**, 217–226 (1999).
82. S.A. Brown, C.A.S Hall, W. Knabe, J. Raich, M.C. Trexler, P. Woomer, 'Tropical Forests, Their Past Present and Potential Future Role in the Terrestrial Carbon Budget', *Water, Air Soil Pollut.*, **70**, 71–94 (1993).
83. J.G. Saldarriaga, D.C. West, C. Uhl, M.L. Tharp, 'Long-term Chronosequence of Forest Succession in the Upper Rio Negro of Colombia and Venezuela', *J. Ecology*, **76**, 938–958 (1998).
84. S.A. Brown, A.E. Lugo, 'Tropical Secondary Forests', *J. Tropical Ecology*, **6**, 1–32 (1990).
85. M.D. Swaine, J.B. Hall, 'Early Succession in Tropical Forest', *J. Ecology*, **71**, 601–627 (1983).
86. C. Uhl, C.F. Jordan, 'Succession and Nutrient Dynamics Following Forest Cutting and Burning in Amazonia', *Ecology*, **65**, 1476–1490 (1984).
87. C. Uhl, 'Factors Controlling Slash and Burn in Amazonia, *J. Ecology*, **75**, 371–407 (1987).
88. D.S. Kimes, R.F. Nelson, D.L. Skole, W.A. Salas, 'Accuracies in Mapping Secondary Tropical Forest Age from Sequential Satellite Imagery', *Remote Sensing Environ.*, **65**, 112–120 (1998).
89. P. Mausel, Y. Wu, L. Yinghong, E.F. Moran, E.S. Brondizio, 'Spectral Identification of Successional Stages Following Deforestation in the Amazon', *Geocarto Int.*, **4**, 61–71 (1993).
90. J.M. McMorrow, 'Relation of Oil Palm Spectral Response to Stand Age', *Int. J. Remote Sensing*, **16**, 3203–3209 (1995).
91. M.K. Steininger, 'Tropical Secondary Forest Regrowth in the Amazon, Age, Area and Change Estimation with Thematic Mapper Data', *Int. J. Remote Sensing*, **17**, 9–27 (1996).
92. D.S. Boyd, G.M. Foody, P.J. Curran, R.M. Lucas, M. Honzak, 'An Assessment of Landsat TM Middle and Thermal Infrared Wavebands for the Detection of Tropical Forest Regeneration', *Int. J. Remote Sensing*, **16**, 249–261 (1996).
93. A.L.C. McWilliam, J.M. Roberts, O.M.R. Cabral, M.V. B.R. Leitao, A.C.L. De Osta, G.T. Maitell, C.A.G.P. Lamparoni, 'Leaf Area Indices and Above-ground Biomass of *Terra Firme* Rain Forest and Adjacent Clearings in Amazonia', *Functional Ecology*, **71**, 310–317 (1993).
94. R.M. Lucas, M. Honzak, P.J. Curran, G.M. Foody, 'Regeneration of Tropical Forests Within the Legal Amazon', *Int. J. Remote Sensing*, (submitted).
95. G.M. Foody, G. Palubinskas, R.M. Lucas, P.J. Curran, M. Honzak, 'Identifying Terrestrial Carbon Sinks: Classification of Successional Stages in Regenerating Tropical Forest from Landsat TM Data', *Remote Sensing Environ.*, **55**, 205–216 (1996).
96. D.S. Alves, D.L. Skole, 'Characterizing Land Cover Dynamics Using Multitemporal Imagery', *Int. J. Remote Sensing*, **17**, 835–839 (1996).
97. R.C. Frohn, K.C. Mgwire, V.H. Dales, J.E. Estes, 'Using Satellite Remote Sensing Analysis to Evaluate a Socio-economic and Ecological Model of Deforestation in Rondônia, Brazil', *Int. J. Remote Sensing*, **17**, 3233–3255 (1996).
98. R.M. Lucas, M. Honzak, G.M. Foody, P.J. Curran, C. Corves, 'Characterizing Tropical Secondary Forests Using Multitemporal Landsat Sensor Imagery', *Int. J. Remote Sensing*, **14**, 3061–3067 (1993).
99. R.S. Roy, B.K. Ranganath, P.G. Diwakar, T.P.S. Vohra, S.K. Bhan, J.J. Singh, V.C. Pandian, 'Tropical Forest Type Mapping and Monitoring Using Remote Sensing', *Int. J. Remote Sensing*, **12**, 2205–2225 (1991).
100. K.G. Saxena, A.K. Tiwari, M.C. Porwal, A.R.R. Menon, 'Vegetation Maps, Mapping Needs and Scope of Digital Processing of Landsat Thematic Mapper Data in Tropical Region of South-west India', *Int. J. Remote Sensing*, **13**, 2017–2037 (1992).
101. G.M. Foody, R.A. Hill, 'Classification of Tropical Forest Classes from Landsat TM Data', *Int. J. Remote Sensing*, **12**, 2353–2367 (1996).
102. E. Brondizio, E. Moran, P. Mausel, Y. Wu, 1996. 'Land Cover in the Amazon Estuary: Linking of the Thematic Mapper with Botanical and Historical Data', *Photogrammetric Eng. Remote Sensing*, **62**, 921–929 (1996).
103. S. Sudhakar, S. Sengupta, I. Venkata Ramana, A.K. Raha, B.K. Bardham Roy, 'Forest Cover Mapping of West Bengal with Special Reference to North Bengal Using IRS-1B Satellite LISS II Data', *Int. J. Remote Sensing*, **17**, 29–42 (1996).
104. W.R. Paradella, M.F.F. Da Silva, N.De A. Rosa, C.A. Kushigbor, 'A Geobotanical Approach to the Tropical Rain Forest Environment of the Carajás Mineral Province (Amazon Region, Brazil), Based on Digital TM- Landsat and DEM Data', *Int. J. Remote Sensing*, **15**, 1633–1648 (1994).
105. H. Tuomisto, A. Linna, R. Kalloila, 'Use of Digitally Processed Satellite Images in Studies of Tropical Rain Forest Vegetation', *Int. J. Remote Sensing*, **15**, 1595–1610 (1994).
106. F. Achard, C. Estreguil, 'Forest Classification of South-east Asia Using NOAA AVHRR Data', *Remote Sensing Environ.*, **54**, 198–208 (1995).
107. J.C. Towshend, C.O. Justice, V. Kalb, 'Characterization and Classification of South American Land Cover Types Using Satellite Data', *Int. J. Remote Sensing*, **8**, 1189–1207 (1987).
108. T.A. Stone, P. Schlesinger, R.A. Houghton, G.M. Woodwell, 'A Map of the Vegetation of South America Based on Satellite Imagery', *Photogrammetric Eng. Remote Sensing*, **60**, 541–551 (1994).
109. P.T. Gilruth, C.F. Hutchinson, B. Barry, 'Assessing Deforestation in the Guinea Highlands of West Africa

Using Remote Sensing', *Photogrammetric Eng. Remote Sensing*, **56**, 1375–1382 (1990).
110. J.M. Chen, J. Cihlar, 'Retrieving Leaf Area Index of Boreal Conifer Forests Using Landsat TM Images', *Remote Sensing Environ.*, **55**, 153–162 (1996).
111. P.J. Curran, G.M. Foody, 'The Use of Remote Sensing to Characterize the Regenerative States of Tropical Forests', in *Environmental Remote Sensing from Regional to Global Scales*, eds. G.M. Foody, P.J. Curran, John Wiley & Sons, Chichester, 44–83, 1994.
112. D.L. Peterson, M.A. Spanner, R.W. Running, K.B. Teuber, 'Relationship of Thematic Mapper Simulator Data to Leaf Area Index of Temperate Coniferous Forests', *Remote Sensing Environ.*, **22**, 323–341 (1987).
113. E.O. Box, B.N. Holben, V. Kalb, 'Accuracy of the AVHRR Vegetation Index as a Predictor of Biomass, Primary Productivity and Net CO_2 Flux', *Vegetatio*, **80**, 71–89 (1989).
114. S.A. Sader, R.B. Waide, W.T. Lawrence, A.T. Joyce, 'Tropical Forest Biomass and Successional Age Class Relationships to a Vegetation Index Derived from Landsat TM Data', *Remote Sensing Environ.*, **28**, 143–156 (1989).
115. G.M. Foody, P.J. Curran, 'Estimation of Tropical Forest Extent and Regenerative Stage Using Remotely Sensed Data', *J. Biogeography*, **21**, 223–244 (1994).
116. D.S. Boyd, P.J. Curran, 'Using Remote Sensing to Reduce Uncertainties in the Global Carbon Budget: The Potential of Radiation Acquired in Middle Infrared Wavelengths', *Remote Sensing Rev.*, **16**, 293–327 (1998).
117. G.M. Foody, D.S. Boyd, P.J. Curran, 'Relationships Between Tropical Forest Biophysical Properties and Data Acquired in AVHRR Channels 1–5', *Int. J. Remote Sensing*, **17**, 1341–1355 (1996).
118. Y.J. Kaufman, L.A. Remer, 'Detection of Forests Using MID-IR Reflectance, An Application for Aerosol Studies', *IEEE Trans. Geosci. Remote Sensing*, **32**, 672–683 (1994).
119. J.E. Nichol, 'Monitoring Tropical Rain Forest Microclimate', *Photogrammetric Eng. Remote Sensing*, **61**, 1159–1165 (1995).
120. D.S. Boyd, G.M. Foody, P.J. Curran, 'The Relationship Between the Biomass of Cameroonian Tropical Forests and Radiation Reflected in Middle Infrared Wavelengths (3.0–5.0 μm)', *Int. J. Remote Sensing*, **20**, 1017–1024 (1999).
121. A. Rosenqvist, 'Evaluation of JERS-1, ERS-1 and Almaz SAR Backscatter for Rubber and Oil Palm Stands in West Malaysia', *Int. J. Remote Sensing*, **17**, 3219–3231.
122. F.T. Ulaby, R.K. Moore, A.K. Fung, 'Microwave Remote Sensing Active and Passive. Volume II Radar Remote Sensing and Surface Scattering and Emission Theory', Addison-Wesley, Reading, MA, 1982.
123. P.N. Churchill, A.J. Sieber, 'The Current Status of ERS-1 and the Role of Radar Remote Sensing for the Management of Natural Resources in Developing Countries', in *Remote Sensing and Geographical Information Systems for Resource Management in Developing Countries*, eds. A.S. Belward, C.R. Valenzuela, Kluwer Academic, Dordrecht, 111–143, 1991.
124. C. Elachi, *Spaceborne Radar Remote Sensing: Applications and Techniques*, IEEE Press, New York, 1988.
125. K. Carver, 'Synthetic Aperture Radar/Earth Observing System', Instrument Panel Report, National Aeronautics and Space Administration, Washington, DC, Vol. IIf, 1988.
126. M.L. Imhoff, 'Radar Backscatter and Biomass Saturation: Ramifications for Global Biomass Inventory', *IEEE Trans. Geosci. Remote Sensing*, **33**, 511–518 (1995).
127. C.da C.F. Yanasse, S.J.S. Sant'Anna, A.C. Frery, C.D. Rennó, J.V. Soares, A.J. Luckman, 'Exploratory Study of the Relationships Between Tropical Forest Regeneration Stages and SIR-C L and C Data', *Remote Sensing Environ.*, **59**, 180–190 (1997).
128. A.J. Luckman, J. Baker, T.M. Kuplich, C. da C.F. Yanasse, A.C. Frery, 'A Study of the Relationship Between Radar Backscatter and Regenerating Tropical Forest Biomass for Spaceborne SAR Instruments', *Remote Sensing Environ.*, **60**, 1–13 (1997).
129. A. Luckman, J. Baker, M. Honzak, R. Lucas, 'Tropical Forest Biomass Density Estimation Using JERS-1 SAR: Seasonal Variation, Confidence Limits, and Applications to Image Mosaics', *Remote Sensing Environ.*, **63**, 126–139 (1998).
130. G.M. Foody, R.M. Green, R.M. Lucas, P.J. Curran, M. Honzak, I. Do Amaral, 'Observations on the Relationships Between SIR-C Radar Backscatter and the Biomass of Regenerating Tropical Forests', *Int. J. Remote Sensing*, **18**, 687–694 (1997).
131. R.M. Lucas, M. Honzak, P.J. Curran, G.M. Foody, D.T. Nguele, 'Tropical Forest Regeneration in Cameroon: An Interpretation of NOAA AVHRR Data', *Int. J. Remote Sensing* (submitted).

Steel and Related Materials: Introduction*

Thomas R. Dulski
Carpenter Technology Corporation, Reading, USA

Metallurgy and chemistry have been intimately linked since long before either qualified as a modern science. The alchemist's pursuit of a Philosopher's Stone to transmute base metals into gold might be regarded as the origin of the *synthetic* phase of this symbiotic relationship. However, the analytical phase is much more ancient. Three thousand years ago fire assay techniques were being used to verify the purity of precious metals. These were, by any assessment, *analytical* procedures.

It is difficult from a modern perspective to appreciate fully how important the analysis of metals was to the emerging theory and practice of chemical science in the 18th and 19th centuries. Early texts such as *Metallurgic Chemistry*[1] (published in 1776 in English translation from the original German) demonstrate that even nascent, erroneous ideas (phlogiston theory, for example) were already being tested in practical application to commerce in ores and metals.

It could be argued that the Industrial Revolution was a revolution driven by metal production. The Bessemer process in 1856, the first open hearths in 1888, and the first electric arc furnaces in 1899 represent an upward curve of speed and quality in steel production. In 1886 the Hall/Heroult electrolytic process, in one leap, changed aluminum from a hard-won precious metal into a commodity. As material performance came to be scrutinized more closely, the concept of close compositional control in metal production emerged as a new imperative.

Long before this time it must have occurred to someone that a chemical analysis of the product *during* its production might be very useful, but now the idea took on a new urgency. In 1914 Johnson, who was Chief Chemist at Crucible Steel's Park Works in Pittsburgh, Pennsylvania, published *Rapid Methods for the Chemical Analysis of Special Steels, Steel-making Alloys, and Graphite*.[2] This is a volume that reflects the tremendous body of analytical work that had become associated with steel production by the early years of the 20th century.

Laboratories were designed and staffed to perform labor-intensive wet chemical procedures with great efficiency so that in-process hot metal could be adjusted to final compositional limits. Personnel, space, and laboratory equipment were committed to each elemental determination. Common tasks such as sample preparation, the weighing of test portions, and the preparation of stock reagent solutions were typically assigned as entry-level positions. Analysts specializing in carbon or silicon or manganese determinations might, with experience, advance to more involved procedures – perhaps chromium or nickel or molybdenum, especially at mills that produced the new ''rustless'' grades of steel.

Ores, ferroalloys, coke, and slag required their own separate crew of specialists. Various forms of physical testing of the finished metal were sometimes included in the chemistry laboratory's responsibilities as well. For a detailed picture of the routine operations in the chemistry laboratory of a large basic Steel Works in this period, see *Technical Analysis of Steel and Steel Works Materials* by Sisco (1923).[3] Similar situations prevailed in other metal industries. The 1930s and 1940s saw the publication of several important basic texts that served as references for many of these operations.[4–7]

However, the pace of developments in metal production as the century progressed soon made such manpower-intensive laboratories obsolete: the high-frequency vacuum induction furnace was introduced in 1927; the electroslag remelting furnace was developed in 1935; the vacuum arc remelting furnace was in use by 1945; the basic oxygen process was introduced in 1952 and improved a decade later; continuous casting began in 1963 and argon–oxygen decarburization in 1970; and large-scale ladle refining was introduced in 1980. These improvements have made possible the rapid production of new steels and alloys, some with unique properties and some of unprecedented purity.

Analytical developments paralleled these metallurgical advances: the modern ultraviolet/visible spectrophotometer was introduced in 1940; the first direct-reading arc/spark atomic emission spectrometer was marketed in 1950, the first X-ray fluorescence spectrometer in 1955, and the first atomic absorption spectrophotometer in 1960; the direct current plasma emission spectrometer was introduced in 1965 and the graphite furnace atomic absorption spectrophotometer in 1970; the inductively coupled plasma atomic emission spectrometer followed in 1975; and in 1985 the inductively coupled plasma mass spectrometer was introduced. All of these tools, and more, are still in use today. Each has contributed to the speed and efficiency of the metals analysis laboratory.

The production of nearly all commercial alloys has now become an extremely rapid process. Moreover, compositional specifications are commonly stricter than those of the past and often require assurance that deleterious impurities do not exceed a maximum at trace or ultratrace levels. The analytical control function has

* The ideas and statements contained in this introduction are those of the individual author and are not in any way endorsed by, and shall not be construed as the ideas of, Carpenter Technology Corporation.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

kept pace through the judicious application of high-speed instrumentation and automation and by the use of new techniques of extreme sensitivity.

Currently, except in isolated niches, classical wet analytical chemistry is absent from the workplace, replaced by physical instrumental methods. Solution-based techniques remain important in many laboratories, however, and so the science and lore of dissolution have been retained in some quarters – today, such methods are increasingly augmented by developments in closed-vessel microwave oven technology. Flame and electrothermal atomic absorption spectrophotometry and direct and inductively coupled plasma atomic emission spectrometry have become indispensable when time is available to use them (see the article **Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis** in this section).

However, time is often not available, as in process control applications where 100 tons of molten metal may be ''on hold'' while the laboratory completes its work. In these instances only work on solid metal samples will do. It is here that the power of certain physical methods and their associated mathematical treatments become evident. Often a pneumatic tube system is used to transport a chilled slug of the molten alloy from furnace-side to the laboratory. The sample is then cut, polished, and analyzed, and the results transmitted to the furnace control room in less than 5 min. There are only a few viable technologies for work at this pace: atomic emission using arc/spark excitation; X-ray fluorescence spectrometry; and thermal evolution techniques (see the article **Thermal Evolution Methods for Carbon, Sulfur, Oxygen, Nitrogen and Hydrogen in Iron and Steel Analysis** in this section). Both atomic emission polychromators and simultaneous X-ray fluorescence spectrometers are formidable analytical engines, each capable of generating great quantities of data. Sometimes as many as 25 elements are reported in the allotted 5-min interval, and for some complex alloys these may include sizable groups of major, minor, trace, and even ultratrace constituents.

Recently, robotics has been employed in this task with the development of the so-called ''container laboratory''. Here, programmed robots prepare the sample, present it to the instrument, and even restandardize the instrument with appropriate reference materials as required (see the article **Automation of Analytical Control in the Steel and Metals Industry** in this section). Advances in software are leading to heuristic expert systems that can deal with a wide range of alloy compositions by sensing count rates and then selecting among a variety of calibration curves stored in memory. Time-resolved exposures optimize each element channel for the required sensitivity and precision at the relevant concentration range.

In certain metal industries, such as basic steel and aluminum, the robotized laboratory is already becoming commonplace. But unattended automated analysis may not be practical for many types of metal commodities. In particular, it must be emphasized that all the speed and efficiency of this type of automated analysis are based upon the availability of appropriate, accurately certified reference materials, because all such systems are based on comparative methodology. Without reference materials of closely matching composition, none of this *tour de force* is possible.

Unfortunately, in recent years the availability of reference materials with reliable certified values has become endangered because the infrastructure to certify them has been seriously eroded. It is in those few remaining niches of classical wet analytical chemistry that the best hope for the future lies. It is certainly folly to believe that a certification program based on the use of comparative methodology using alloy reference materials can produce new reference materials with any more than a blurred image of the truth.

Most analysts are aware that accurate sampling is the *sine qua non* of analytical chemistry because all the conscientious labwork in the world is for naught if the test portion does not exactly reflect the test object. And so it is with metal analysis, where that 100 tons of molten metal may be represented by a 50-g slug, and no less for the furnace charge materials – highly heterogeneous scrap, notoriously segregated ferroalloys, ores, and fluxes – that are bought and sold for use in metal production in enormous lots. Proper sampling is a complex art that must be an important concern of the metal analyst (see the articles **Metal Analysis, Sampling and Sample Preparation in** and **Iron Ore, Sample Preparation and Analysis of** in this section).

In many cases the modern metal analysis laboratory is asked to do much more than analyze metal, as difficult a task as that often is. In addition to ore, slag, and refractory materials, typical nonmetal samples include process solutions, such as plating or pickling baths, contact cooling water, furnace gases, and impinger samples from stacks and workplace atmospheres. There also may be inorganic or organic coatings on metal surfaces to identify and quantify, and here specialized techniques often must be brought into play. Instrumental approaches in this area include Fourier transform infrared spectroscopy, Auger spectroscopy, secondary ion mass spectrometry, and others – each a highly specialized field.

It is ironic that, despite all the remarkable advances in analytical science, the metals analyst still remains poised between two worlds. There is no question that advances in physical instrumentation have provided the means to achieve unprecedented accuracy and precision with amazing speed, but at the same time the 3000-year-old

art of fire assay persists as the most reliable means of quantifying certain precious metals. Although silent robots provide the analytical control for vast quantities of steel and aluminum, their performance hinges critically on a diminished store of alloy reference materials, certified by methods developed almost 100 years ago.

The analysis of metals has a rich history, as we have seen, but its future may take leaps and turns that we can now only dimly imagine. High-temperature superconductors, amorphous alloys, shape-memory alloys, directionally solidified and single-crystal alloys, and metal–matrix composites are some of the high-technology material challenges for the analytical chemist today. No one can say which, if any, will be routine commodities in the 21st century. It *is*, however, reasonable to expect new alloys with imposing compositional specifications. Metallurgical processes are likely to continue to increase in speed as older, inefficient mills and equipment are phased out and replaced. The metals analyst can be expected to meet these new challenges with new analytical tools, such as inductively coupled plasma- and glow discharge mass spectrometry, which are already in use in many metals analysis laboratories.

In this section we have collected a series of articles that represent a picture of industrial steel analysis as it is currently practiced. We have also included some additional topics that reflect the way some of the same challenges are being met for other metal commodities, and with unique methodologies (see articles **Nickel Ore and Metals Analysis; Noble Metals, Analytical Chemistry of** and **Nuclear Magnetic Resonance in Metals Analysis**). The reader will find here both the innovative and the routine. It is a picture that conveys a sense of the complexity and depth of this field – an area as demanding and rewarding as any in analytical chemistry.

REFERENCES

1. C.E. Gellert, *Metallurgic Chemistry*, Adelphi Strand, London, 1776.
2. C.M. Johnson, *Rapid Methods for the Chemical Analysis of Special Steels, Steel-making Alloys, and Graphite*, 2nd edition, John Wiley & Sons, New York, 1914.
3. F.T. Sisco, *Technical Analysis of Steel and Steel Works Materials*, McGraw-Hill, New York, 1923.
4. G.E.F. Lundell, J.I. Hoffman, H.A. Bright, *Chemical Analysis of Iron and Steel*, John Wiley & Sons, New York, 1931.
5. H.V. Churchill, R.W. Bridges, *Chemical Analysis of Aluminum*, Aluminum Co. of America, New Kensington, PA, 1935.
6. US Steel Corp., *Sampling and Analysis of Carbon and Alloy Steels*, Reinhold, New York, 1938.
7. E. Gregory, W.W. Stevenson, *Chemical Analysis of Metals and Alloys*, 2nd edition, Blackie & Son, London, 1942.

Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis

Maria Grazia Del Monte
Centro Sviluppo Materiali, Rome, Italy

1 Introduction 1
2 History 2
3 Iron and Steel Application 4
3.1 Aluminum 5
3.2 Antimony 6
3.3 Arsenic 6
3.4 Barium 6
3.5 Bismuth 6
3.6 Boron 7
3.7 Cadmium 7
3.8 Calcium 7
3.9 Cerium 7
3.10 Cobalt 7
3.11 Chromium 7
3.12 Copper 7
3.13 Lanthanum 7
3.14 Lead 7
3.15 Magnesium 8
3.16 Manganese 8
3.17 Mercury 8
3.18 Molybdenum 8
3.19 Nickel 8
3.20 Niobium 8
3.21 Phosphorus 8
3.22 Praseodymium 8
3.23 Rare-earth Elements 9
3.24 Selenium 9
3.25 Silicon 9
3.26 Silver 9
3.27 Sulfur 9
3.28 Thallium 9
3.29 Tantalum 9
3.30 Tellurium 9
3.31 Tin 9
3.32 Titanium 9
3.33 Tungsten 10
3.34 Vanadium 10
3.35 Yttrium 10
3.36 Zinc 10
3.37 Zirconium 10
4 Sampling and Sample Preparation 10
5 Spectrometric Measurement Process 12
5.1 Interferences 12
5.2 Calibration 13
6 Quality Assurance Criteria 14
7 Standard Methods of Analysis 18
Abbreviations and Acronyms 18
Related Articles 18
References 19

The use of solution-based atomic absorption spectrometry (AAS) and atomic emission spectrometry (AES) in iron and steel analysis is presented. The article is arranged in different sections dealing with the history relevant to the introduction and the acceptance of these techniques in iron and steel laboratories and their application to iron and steel analysis. A special emphasis is given to the application studies described in the literature for determining the almost forty elements which can be present in steels.

Topics such as sampling and sample preparation, spectrometric measurement, interferences, calibration, and quality assurance criteria are discussed. Finally a brief overview of the standard methods of analysis is presented.

1 INTRODUCTION

Steels are iron-based alloys with a carbon content generally below 2%. In order to be technically competitive on the market with materials such as plastics, aluminum, copper, glass and even cement, steels have to satisfy many property requirements (hardness, tensile and impact strength, electrical and magnetic properties, wear resistance, soundness, fabricability, formability, surface finish, cleanness, corrosion resistance in a host of different environments, resistance to failure by hydrogen embrittlement, stress corrosion, fatigue, etc.).

Although the chemical composition is not a real requirement, nevertheless it is the vital starting point from which any desired property or group of properties can be consistently produced. By varying the chemical composition and applying various thermal treatments, different properties can be achieved and, therefore, different grades of steel produced. Plain carbon steel is converted into special steel when sufficient concentrations of other elements are introduced to alter the properties significantly (e.g. up to 3% nickel, chromium and/or molybdenum, vanadium, etc., to produce ''low-alloy steels'' for engineering purposes; up to 9% nickel or chromium to form ''medium-alloy steels'', etc.).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

"Stainless steels" are one of the best known special steels. They are made by the addition of at least about 11% chromium. This minimum confers tarnish- and corrosion-resistance in gaseous and liquid media that would attack plain carbon steels.

There are over 200 types (grades) of stainless steels, so that the American Iron and Steel Institute (AISI) uses a three-digit system to separate standard grades of wrought stainless steels into four general classes, based on composition: series 200 (chromium, nickel, and manganese), series 300 (chromium and nickel), series 400 (chromium) and series 500 (low chromium).

In iron and steel production it is mandatory to keep the content of different elements under rigorous control during all transformation steps from raw materials to finished products in order to be sure that the composition of the steel falls within an established range of acceptability.

Chemical composition control not only concerns steel during its production but also all the other materials related to the process. These include fuels, oils, raw materials such as iron ores, ferroalloys and steel scraps, waters, pickle liquor, lubricants, effluents, wastes.

In the early days of industrial steel-making, chemical analysis played a minor role in the production. However, fairly soon, it was obvious that knowledge about the chemical composition had to be known during the processing. For example, when the Bessemer process was introduced around 1860 it became necessary to have better control of the carbon content during production and the need for fast carbon determination was increased.

Chemical laboratories grew quickly in importance. All the analyzes were based on wet chemical methods such as gravimetry, potentiometry, titrimetry, ultraviolet and visible spectrophotometry. The complexity of the analysis became very high when production of stainless steels started in the early 1900s.

In the mid 1940s the first simultaneous emission spectrometers were able to determine a large number of elements in solid steel samples. Another breakthrough in steel analysis came in the 1950s when the X-ray fluorescence technique was introduced.

Since that time analytical chemistry in iron and steel plants followed two directions: high-speed spectroscopic methods were used for the production control, while classical wet methods were used for the control of the finished products, for the certification of reference materials necessary for the calibration of the spectrometric apparatus and for the analysis of all those materials not directly linked to the speed of the production process or not detectable (for size, quantity, or concentration) by instrumental methods.

The determination of element concentrations by wet methods changed drastically at the end of the 1960s when flame atomic absorption spectrometry (FAAS) was introduced in iron and steel laboratories in order to satisfy the pressing requirements of sensitivity, accuracy, speed and low cost.

Early in the history of steel-making, trace elements were totally neglected but in the 1970s it was discovered that some elements, present in low concentration in steels, like Pb, Bi, As, had detrimental effects on the mechanical properties such as hardenability, hot workability, creep resistance, etc. It was also found that some elements, such as Ca, Mg, Zr, Ce had positive effects on the properties even in very low concentrations. As the need for trace analysis increased, it became clear that the sensitivity of FAAS was not sufficient and extraction procedures employed prior to instrumental determination were necessary.

This obstacle was overcome with the adoption in the 1970s of electro thermal atomic absorption spectrometry (ETAAS).

At the end of the 1970s AES with plasma sources, owing to its linear calibration over wide concentration ranges, became the preferred methodology for major and minor element determination. A few laboratories are attempting to extend the detection limit (DL) of conventional plasma AES by employing preliminary solvent extractions, ion exchange, precipitation techniques, or linking ion chromatography to plasma AES to circumvent spectral interferences. Volatilization of the analyte compound followed by direct introduction of the volatile species into the plasma is another means to increase the sensitivity and decrease the interferences.

2 HISTORY

FAAS as an analytical tool for chemical analysis was introduced by Walsh[1–2] and Alkemade and Milatz[3] in the 1950s. Walsh suggested that measurement of the absorbance at the peak of the absorption line profile would give a linear relationship with concentration over a wide range of absorbance values.

Since then FAAS has become one of the best and most widely used analytical methods for the determination of major and minor constituents. FAAS concerns the determination, in liquid or dissolved samples, of concentrations of elements with atoms capable of absorbing radiation within the optical spectrum range.

The sample is converted into an atomic vapor by use of an air/acetylene or a nitrous oxide/acetylene flame. The number of atoms capable of absorbing any particular wavelength of transmitted light is proportional to the concentration of these atoms in the optical path in the flame. FAAS relies on measuring the peak absorbance of

a spectral line by using a spectral light source emitting a sharp line.

In its simplest form the radiant source employed is a hollow cathode lamp emitting radiation from excited atoms of the same element as that to be determined. This radiation, in the form of a line spectrum, is absorbed by the sample which has been atomized at 2000–3000 °C.

The resulting spectrum of the source radiation is one with lines of reduced intensity, the magnitude of the reduction being proportional to the concentration of the element in the atomized vapor. A monochromator is used to select a single wavelength for examination and changes in the intensity of this wavelength are related to different analyte concentrations.

FAAS gained popularity in iron and steel laboratories for the following reasons:

- capability of analyzing most elements
- good sensitivity
- ease of handling
- only a few interferences.

The replacement of the flame by an electrothermal sampling device improved the DL by one to three orders of magnitude. Among the electrothermal atomization techniques, graphite furnace atomic absorption spectrometry (GFAAS) is most popular.

This technique is ideal for ultratrace determination and requires only a few microliters or micrograms of a sample. Originated by Massman[4] and L'vov[5] the technique employs a graphite furnace to produce ground-state atoms by rapid volatilization of a small quantity of sample. The improved sensitivity is attributed partly to the fact that the residence time of the atoms in the light beam is much longer in the furnace than in the flame.

The graphite furnace can be used with almost all atomic absorption spectrometers. The energy required for atomization is supplied by applying a high electrical current through a graphite tube, where the sample has been placed. The furnace is located in the sample compartment so that light from the light source passes through the graphite tube. When the furnace is fired, the generated atomic vapor absorbs light from the source. The absorption signal is transient and a peak-shaped signal is produced as the atom concentration within the furnace rises, and then falls. Either peak height or peak area may be used for quantitation. As sample components can affect the rate of atom formation and hence the signal shape, integration over the peak area could result in better accuracy compared to peak height evaluation. A programmable power supply provides precise control of temperature programming for the drying and atomization steps. Micro amounts of solid steel samples can be analyzed directly without prior dissolution. Usually in GFAAS a measured volume of the test sample solution or a weighed mass of the test sample is introduced into the graphite atomizer. For this reason the atomization signal is proportional to the analyte mass in the test sample portion and not to its concentration.

The low cost of the GFAAS and the low limits of detection render this technique competitive for trace and ultratrace analysis in comparison with new techniques devoted to ultratrace analysis, such as inductively coupled plasma/mass spectrometry (ICP/MS).

Plasma AES, although introduced in 1964 by Greenfield[6] and Fassel[7], has found application in iron and steel laboratories only since 1980, after a certain number of papers showed its good performance in the analysis of steel.

The reasons for the introduction of this technique in iron and steel laboratories were:

- very large dynamic range
- very low noise signal
- most elements can be measured
- fast
- high sensitivity
- linear calibration over wide concentration ranges
- fundamental calibration from pure reagents
- ability to determine a programme of elements simultaneously.

In atomic spectrometry the absorption and emission of energy can take place with atoms and ions only in discrete quantities which are characteristic for each element. In emission spectrometry the atoms or ions are excited thermally and the number of photons emitted during relaxation is measured. The identification of the elements takes place by means of the wavelength of the radiation, while the concentration of the element is proportional to the intensity of the radiation. Since all the elements present are excited at the same time and therefore emit radiation simultaneously, it is possible to determine several elements at the same time.

In DCP (direct current plasma) and ICP (inductively coupled plasma) spectrometry the excitation takes place in a plasma.

ICP is generated in the following way: a gas, usually argon, flows through a system of three concentric tubes surrounded by a coil. A high-frequency alternating current is sent through the coil. Some charged particles are introduced into the gas by means of a spark discharge. They are accelerated into the electromagnetic field of the coil and can ionize other argon atoms by means of collisions. These ions are also accelerated and they can again ionize atoms or revert to atoms by recombination. This produces a chain reaction which leads within a fraction of a second to a state of equilibrium whereby the number of ionization and recombination reactions are

equal to each other. In the resulting plasma approximately 1% of the argon atoms are ionized and the gas has a temperature of about 8000–10 000 K.

If a solution is introduced into the plasma, the solvent will evaporate, after which the compounds present will melt, liquefy, evaporate and decompose into the constituent atoms. These atoms may be ionized. Both atoms and ions will be excited. The excited species will revert to lower energy state, whereby the absorbed energy will be released by the emission of photons. The wavelengths of the emitted radiation are characteristic for each element. In addition to this line spectrum there is a background which is caused primarily by recombination radiation and molecular emission bands. The spectrum can be analyzed by means of the dispersion of the radiation in a spectrometer. The intensity of radiation at a given wavelength is measured by means of a photomultiplier tube. A correction is made for background emission. The net intensity is proportional to the concentration of the analyte element over a range of typically up to six decades. The proportionality factor depends to a great extent on the analytical wavelength used and on the type and adjustment of the instrument which is used.

In the case of the DCP source the aerosol hits the surface of the plasma in the centre of three branches. Due to this the temperature available is lower than in the ICP and will be about 6000 K.

Plasma produces higher temperatures than chemical flames. These higher temperatures allow the dissociation of many types of compounds, allowing a wider range of determination than FAAS.

3 IRON AND STEEL APPLICATION

In steel and iron determination the analyst must cope with extreme differences in element concentrations, amounting to as much as six or seven orders of magnitude. From the analytical chemistry point of view, steel can be classified in three categories, in which the concentration ranges of elements are given in Table 1.

AAS (flame and electrothermal sources), as well as plasma AES, are exploited in iron and steel analysis for determining practically all the elements of the periodic table, excluding the inert gases, halogens, H, C, N, O.

The selection of AAS or plasma AES depends on the type of analysis required (the element, its concentration, the matrix composition, accuracy and uncertainty, cost of analysis) and the exigencies of the laboratory and how often the instrument is to be used on a routine basis. Plasma AES, for example, is used in routine methods when determining elements at high levels (>6%) with a precision (repeatability) of 0.5–0.6%.

When multiple elements are required on one sample, inductively coupled plasma/atomic emission spectroscopy (ICP/AES) is preferred, owing to the multielement capacity of AES.

For the determination of a single element, FAAS is quicker and more cost effective. Plasma AES is clearly superior to FAAS in the dynamic range covered and this allows a large range of elements be determined on a single solution preparation without any dilution. A few elements such as B, Ce, La, Nb, P, Pb, Pr, Sn, Ta, Ti, V, W and Zr have good DLs in the ICP but in FAAS sensitivities are poor. Usually, for high-precision trace analysis, electrothermal sources are required.

The concept of trace level in steel and iron has changed with time as the knowledge of the properties of elements increased. Today, trace analysis in metals can be defined as analysis involving element concentrations up to 50 $\mu g\,g^{-1}$ (ppm) of sample. This means that the real concentration in the solution will be less than 0.5 $\mu g\,ml^{-1}$, with the solution containing less than 1%(m/m) of solid sample.

A comparison of the current DL for pure elements in acid solution for the ICP/AES, FAAS and graphite furnace techniques is reported in Table 2.

The limit of quantitation (LOQ) in steel analysis is generally 10–100 DL.

In the literature, interferences from the matrix, chemical and ionization effects and background interferences were documented along with procedures for eliminating or compensating for them. The main strengths of these techniques today are the availability of proven methodologies and reference standards for almost all the elements of the periodic table.

One of the first papers concerning steel analysis was published in 1962 by Belcher and Bray[8] which described a procedure for magnesium determination. Butler, Fassel, and Kniseley[9] were the first to report on analytical ICP work for iron and steels and related alloys. With the financial support of the Commission of the European Communities many application studies have been carried out in the field of the analysis of steels and related materials in order to evaluate the capability of flameless AAS[10–14] and plasma spectrometry[15–18] as techniques to be used routinely in iron and steel laboratories. As far as plasma spectrometry is concerned, a line coincidence table for 40 elements that do not show any spectral interferences from the matrix up to a concentration of 5 $mg\,ml^{-1}$ of Fe has been worked out by Wittmann.[19] Reviews on the application of AAS to the analysis of iron and steel alloys were prepared by Ohls and Sommer,[20] Stafilov[21] and Kobayashi.[22] Bloedorn,[23–24] Ambrose,[25] and Fujimoto[26] presented a review of improvement of the reproducibility, sample pretreatment and calibration in ICP/AES. This section

Table 1 Steel and iron categories and concentration range of elements, % (m/m)

Element	Unalloyed steel	Alloy steel	Iron
Aluminum	0.001 to 0.10	0.001 to 1.50	0.005 to 0.50
Antimony	–	0.001 to 0.05	0.005 to 0.05
Arsenic	0.0005 to 0.10	0.0005 to 0.10	0.01 to 0.10
Barium	<0.001%		
Bismuth	0.001 to 0.10	0.001 to 0.10	0.001 to 0.05
Boron	0.0001 to 0.0008	0.0005 to 0.10	0.001 to 0.10
Carbon	0.0001 to 0.50	0.001 to 2.00	2.00 to 4.50
Calcium	0.0003 to 0.01	–	–
Cadmium	<0.00001	<0.00001	
Cerium	0.005 to 0.05	0.005 to 0.05	0.005 to 0.05
Chromium	0.005 to 0.30	0.01 to 35.0	0.01 to 30.0
Cobalt	0.001 to 0.10	0.001 to 17.0	0.001 to 5.00
Copper	0.005 to 0.40	0.01 to 5.0	0.01 to 8.0
Lanthanum	0.001 to 0.1	–	–
Lead	0.001 to 0.40	0.001 to 0.50	0.001 to 0.20
Mercury	<0.0001	<0.0001	<0.0001
Magnesium	–	–	0.001 to 0.10
Manganese	0.01 to 1.80	0.01 to 20.0	0.05 to 3.0
Molybdenum	0.001 to 0.08	0.01 to 10.0	0.01 to 5.0
Neodymium	0.0001 to 0.01	–	–
Nickel	0.005 to 0.30	0.01 to 30.0	0.01 to 40.0
Niobium	0.003 to 0.06	0.003 to 5.0	–
Nitrogen	0.0005 to 0.05	0.001 to 0.50	–
Phosphorus	0.001 to 0.05	0.001 to 0.50	0.01 to 1.00
Praseodymium	0.0001 to 0.01	–	–
Selenium	0.001 to 0.10	0.01 to 0.50	0.001 to 0.10
Silicon	0.001 to 0.50	0.01 to 5.00	0.10 to 6.00
Silver	<0.001%	–	–
Strontium	<0.001	–	–
Sulfur	0.0001 to 0.50	0.001 to 0.50	0.005 to 0.25
Thallium	<0.001	–	–
Tantalum	0.01 to 0.05	0.01 to 1.00	–
Tellurium	0.001 to 0.10	0.001 to 0.10	0.001 to 0.35
Tin	0.001 to 0.10	0.001 to 0.10	0.001 to 0.35
Titanium	0.001 to 0.05	0.01 to 5.00	0.001 to 0.20
Tungsten	0.005 to 0.10	0.005 to 25.0	0.001 to 0.20
Vanadium	0.005 to 0.10	0.005 to 4.00	0.005 to 0.50
Yttrium	<0.001	–	–
Zinc	–	–	0.005 to 0.20
Zirconium	0.005 to 0.05	0.005 to 0.20	–

provides a survey of a few of the most interesting application studies described in the literature for the iron and steel analysis by AAS and AES solution-based spectrometry with reference to each element of the periodic table.

3.1 Aluminum

Aluminum was determined in low-alloy steel by Klein[27] and Rains.[28] Specimen preparation, calibration standard preparation, analysis condition and instrument parameters were investigated.

The determination of aluminum was carried out by FAAS using a nitrous oxide–acetylene flame, with no interferences, by Headridge and Sowerbutts,[29] and Manning.[30] The iron concentration was matched. Brivot[31] determined aluminum by FAAS and by spectrophotocolorimetry.

Shaw and Ottaway,[32] Yu and Cui,[33] Atsuya,[34] and Zyukova[35] determined trace amounts of aluminum in steels by GFAAS. Persson[36] also determined aluminum in steel by GFAAS. The sample was dissolved in hydrochloric acid/nitric acid and ammonium sulfate was added to eliminate the influence of hydrochloric acid. The iron content in the sample reduced the sensitivity by about 20% and was therefore matched in the reference solution.

Collin[37] analyzed the various nonmetallic phases in steel which have considerable influence on the physical and mechanical properties. Kobayashi[38] and Naka[39] carried out the determination of acid-soluble and

Table 2 Elemental DL comparison for FAAS, GFAAS, ICP/AES

Element	FAAS mg L^{-1}	Graphite furnace mg L^{-1}	ICP mg L^{-1}
Aluminum	0.03	0.0001	0.02
Antimony	0.03	0.0001	0.1
Arsenic	0.02	0.0002	0.1
Barium	0.01	0.00004	0.005
Bismuth	0.02	0.0001	0.05
Boron	1.0	0.001	0.01
Calcium	0.001	0.0005	0.005
Cadmium	0.0005	0.000003	0.003
Cerium	–	–	0.03
Chromium	0.002	0.00001	0.005
Cobalt	0.006	0.00001	0.02
Copper	0.001	0.00002	0.0005
Lanthanum	–	–	0.02
Lead	0.01	0.00005	0.5
Mercury	0.17	0.02	0.0005
Magnesium	0.0001	0.000004	0.0005
Manganese	0.001	0.00001	0.001
Molybdenum	0.03	0.00002	0.01
Neodymium	–	–	0.02
Nickel	0.004	0.0002	0.01
Niobium	1.0	–	0.1
Phosphorus	50	0.03	1.0
Praseodymium	–	–	0.0001
Selenium	0.1	0.0005	0.05
Silicon	0.05	0.0001	0.01
Silver	0.001	0.0002	0.2
Strontium	–	–	0.00005
Thallium	0.001	0.0001	–
Tantalum	1.0	–	0.01
Tellurium	0.02	0.001	0.1
Tin	0.05	0.0005	0.1
Titanium	0.05	0.0005	0.0005
Tungsten	1.2	–	0.05
Vanadium	0.04	0.0002	0.005
Zinc	0.0001	0.00001	0.005
Zirconium	0.35	–	0.005

insoluble aluminum. The DL was 0.3 ppm acid soluble and 0.1 ppm of acid insoluble when 0.5 g of specimen was used.

1 g of steel dissolved in a mixture of HCl and HNO_3 and diluted to a concentration of 5 mg mL^{-1} of sample was analyzed for Al at 396.1 nm by ICP/AES by Butler[9] and Gomez Coedo[40–41] while λ 308.22 nm was used by Ward[42] and Hughes.[15]

3.2 Antimony

Traces of antimony were determined by extraction followed by FAAS determination by Hein.[43] The hydride technique was used by Fleming,[44] Vanloo,[45] and Bombach[46] while the graphite furnace was adopted by Barnett,[47] Frech,[48] Bettinelli[49] and by Del Monte Tamba.[50] They discovered that chromium enhances the signal, but above 200 mg L^{-1} the enhancement is constant, so that this quantity of chromium was always added. Flow injection hydride generation was applied by Weltz.[51]

λ 206.84 nm was used by Hughes[15] for determining antimony by ICP/AES.

Traces of antimony were determined by ICP/AES by Zhang[52] by solvent extraction in methyl isobuthyl ketone.

3.3 Arsenic

The hydride technique was applied by Fleming[44] and the graphite furnace by Barnet,[47] Del Monte Tamba[50] and Bettinelli.[49]

The method for the determination of arsenic in stainless steel by ICP/AES at the wavelength 228.81 nm was described by R. Diemiaszonek.[53] Ward,[42] Hughes[15] and Zhang[52] analyzed arsenic at λ 193.70 nm. Mendez Garcia[54] determined arsenic by ICP/AES by direct introduction of arsine into the plasma. Arsenic from the sample solution was extracted on-line as AsI_3 into a xylene solution. The xylene extract continuously reacting with sodium borohydride in dimethylformamide and acetic acid generated AsH_3 which was introduced into the plasma.

Arsenic in austenitic steel was determined by Nakayama.[55] He decomposed the sample in a mixture of HCl and HNO_3, filtered and separated the solution obtained in an anion exchange chromatographic column; the resulting eluent was treated with a mixture of $NaBH_4$ and HCl, and then subjected to ICP/AES. The determination of hydrides of As in steels was carried out by Ozaki[56] by coupling a polytetrafluoroethylene phase-separator layer to an ICP atomic emission spectrometer. Weltz[51] developed a method for the determination of arsenic in steel using flow injection hydride generation atomic absorption spectrometry (FI/HGAAS) and L-cysteine as a reducing and releasing agent.

3.4 Barium

This minor element in most iron oxides is usually lost in the refining process of low-alloy iron casting and steels. The determination of barium by FAAS was described by Rains,[28] Belcher[57] and David.[58] According to David[58] the nitrous oxide/acetylene flame has been shown to be very useful. In this flame, barium is highly ionized and an ionization suppressant must be added.

Barium was determined by ICP/AES by Hughes[15] at the wavelength 455.40 nm.

3.5 Bismuth

Fleming,[44] Weltz[59] and Vanloo[60] determined bismuth by applying the hydride technique, while Barnett[47]

used GFAAS while applying the analyte addition technique. GFAAS was used by Bettinelli[49] and also by Dulski[61] who used spiked samples for calibration. Del Monte Tamba,[12,50] Takada[62] and Frech[63] determined bismuth on solid samples. Cadore[64] developed a method for the determination of trace amounts of bismuth using flow injection and AAS with hydride generation. A comparison between flow injection hydride generation AAS and GFAAS was carried out by Bettinelli.[65]

The method for the determination of bismuth in stainless steel by ICP/AES at the wavelength 306.77 nm was described by R. Diemiaszonek.[53] Zhang[52] determined bismuth in iron and steel by solvent extraction and ICP/AES.

3.6 Boron

The method for the determination of boron in stainless steel by ICP/AES at the wavelength 249.68 nm was described by R. Diemiaszonek.[53] Boron was determined by Hughes[15] by ICP/AES using λ 249.77 nm and by Belliveau[66] using a multielement dc plasma emission spectrometer. Boron down to 1 ppm was determined by Baena Liberato[67–69] and by Chiba[70] by ICP/AES using a solvent extraction approach. Lopez Molinero[71] set up a method for determining acid soluble and acid insoluble boron in steels based on ICP/AES.

3.7 Cadmium

Trace of cadmium were directly determined by Liu[72] in stainless steels by platform GFAAS after dissolution of the sample in aqua regia, and redilution in $1+1$ HNO_3. The detection range is $1.10\,\mu g\,g^{-1}$ and the relative standard deviation is 9.3–14.3%. Barnett[47] determined cadmium by GFAAS, applying the analyte addition technique, while Del Monte Tamba,[50] and Lundberg[73] analyzed cadmium by GFAAS using solid samples.

Cadmium was determined by ICP/AES at wavelength 226.50 nm by Hughes.[15]

3.8 Calcium

Calcium in steels by AAS was determined by Reins,[28] Belcher,[74] and Tugane.[75] Headridge[76] determined trace amounts of calcium in stainless steel by solvent extraction followed by AAS. Traces of calcium were determined by carbon furnace by Glenc[77] and Alvarado.[78]

Hughes[15] determined calcium by ICP/AES at λ 315.89 nm. Metallic and chemically bound calcium in steels was determined by Golloch[79] by ICP/AES employing electrothermal atomization.

3.9 Cerium

Cerium was determined in steel by ICP/AES by Butler,[9] using the wavelength 456.2 nm, by Diemiaszonek[53] using λ 413.76 nm and by Hughes[15] at λ 395.25 nm.

3.10 Cobalt

Sprague,[80] Cobb,[81] Gregorczyk[82] and Wycislik[83] determined cobalt by FAAS. Cobalt in tool steel was analyzed by Knight,[84] while trace cobalt in steel was analyzed after chloroform extraction of its 1-nitroso-2-naphthol complex by Eskell.[85] Cobalt was analyzed by ICP/AES by Ward[42] and Hughes[15] at λ 228.62 nm.

3.11 Chromium

Chromium was analyzed by FAAS by Kinson,[86] Barnes,[87] Gomez Coedo,[88] Thomerson,[89] Ottaway,[90] Cobb,[91] Pandly,[92] Grimaldi,[93] Gregorcyznik,[94] Husler,[95] Sire.[96] Wycislik,[78] and Fogg[97] used the masking of iron with fluoride in extractive atomic absorption determination of chromium in steels. Chromium was analyzed by ICP/AES at λ 357.8 nm by Butler,[9] at λ 302.43 nm by Demiaszonek,[52] at λ 267.72 nm by Ward[42] and Hughes[15] and at λ 350.2 nm by Belliveau[66] and Gomez Coedo.[98] Souza[99] determined chromium in steels by plasma AES after electrodissolution in a flow injection. AAS and ICP/AES were used by Chin[100] for the determination of chromium in low-alloy steel armor.

3.12 Copper

Sprague,[80] Wycislik[78] and Kinson[101] determined copper by FAAS. They found that the determination of copper is influenced both by the acid content and the iron concentration, so they matched the reference solution correspondingly.

Copper was determined by ICP/AES at λ 324.7 nm by Butler,[9] Ward,[42] and Hughes.[15]

3.13 Lanthanum

The method for the determination of lanthanum in stainless steel by ICP/AES at the wavelength 398.8 nm was described by Butler.[9] R. Diemiaszonek[53] used λ 408.67 nm.

3.14 Lead

Lead occurs as a residual element in steel and can cause catastrophic failure in cast iron. Gluc[102] determined lead by AAS in the range 0.0005–0.04%. T.C Rains,[28] Barnett,[47] Shaw and Ottaway,[103] Dulski and Bixler,[61]

Bettinelli,[49] Frech,[104–106] and Mile[107] have described analytical methods for the determination of lead in steel by GFAAS. Special procedures were developed to overcome analyte losses due to chloride during the charring step. A few authors recommended a separation to remove the lead from the iron matrix prior to the determination. Fleming[44] determined lead by the hydride technique. The analysis of the solid sample by GFAAS was carried out by Del Monte Tamba,[12] Andrew[108] and Frech.[63]

Butler et al.[9] and Zhang[51] determined lead by ICP/AES using the wavelength 405.7 nm.

3.15 Magnesium

Magnesium is present in steel as a residual element. In 1962 Belcher[8] developed a procedure which is still frequently used today. Some 1 g of sample is dissolved in 30 mL of 1 + 1 hydrochloric acid and oxidized with 5 mL nitric acid. The solution is evaporated to dryness, mildly heated for a further 5 min, taken up in 10 mL hydrochloric acid and then diluted with strontium chloride solution.[2] The determination of magnesium is carried out by FAAS using either an air/acetylene flame or a nitrous oxide/acetylene flame. Belcher,[57] Rains,[28] and Sprague[80] determined Mg by graphite furnace. Hughes[15] determined magnesium by ICP/AES at λ 279.55 nm.

3.16 Manganese

Manganese was determined by FAAS by Klein,[27] Brivot,[31] Sprague,[80] Gregorczyk,[82] Knight,[84] Husler,[95] Sire.[96] Wycislik,[83] Belcher,[109] Atsuya,[110] Hubbold,[111] and by ICP/AES, at λ 403.0 nm by Butler[9] while Hughes[15] and Ward[42] used λ 257.61 nm, as did Belliveau[66] that determined manganese by dc AES. Souza[99] proposed for its determination a flow injection electrodissolution procedure. Chin[100] determined manganese both via AAS and ICP/AES.

3.17 Mercury

Mercury was determined by ICP/AES at λ 253.65 nm by Hughes.[15]

3.18 Molybdenum

David[112] showed that molybdenum absorption is very dependent on the air to acetylene ratio. A serious suppression of absorption happens when manganese, calcium and strontium are present. An acetylene/air flame was also used by Gregorczyk[82] for determining molybdenum in low- and medium-alloy steel, while Thomerson,[89] Kirkbright[113] and Klein[27] used a nitrous oxide/acetylene flame. FAAS was applied for molybdenum determination by Grimaldi,[93] Hussler,[95] Wycislik,[83] Endo,[114] Goto[115] and Wada.[116] Knight[84] determined molybdenum in tool steels. A preliminary extraction with benzoinoxine–methyl isobutyl ketone (MIBK) was carried out by Castillo.[117] Traces of molybdenum in cast iron have been determined by Gu[118] by FAAS extraction from 60%(v/v) HCl with ethyl-*N*-laurylthioacetate (ELTA) into butylacetate and back extraction with 1%(v/v) ammonium hydroxide. Bodrov and Nickolaev[119] determined molybdenum in steels using a L'vov furnace and found no interferences other than from a high concentration of titanium and niobium. Molybdenum was analyzed by ICP/AES at λ 203.03 nm by Ward,[42] at λ 287.15 nm by Hughes[15] and at λ 281.6 nm by Belliveau[66] and Chin.[100]

3.19 Nickel

Nickel was analyzed by FAAS by Klein,[27] Kinson,[120] Sprague,[80] Gregorczyk,[82] Knight,[84] Sire,[96] Wycislik,[83] Welz,[121] and by ICP/AES at λ 351.5 nm by Butler,[9] at λ 300.36 nm by Demiaszonek,[53] at λ 231.60 nm by Hughes[15] and Ward.[42]

Souza[99] determined nickel in ferritic stainless steels by ICP/AES.

3.20 Niobium

Schiller[122] determined niobium at concentrations below 0.3% in low-alloy steels by FAAS, while Thomerson[123] determined niobium down to 0.05%. Niobium was also analyzed by Martin using FAAS.[124]

Niobium was determined by ICP/AES at λ 405.8 nm by Butler,[9] at λ 313.08 nm by Ward,[42] and at λ 309.42 nm by Gomez Coedo.[41]

3.21 Phosphorus

Phosphorus was determined by GFAAS by Whiteside[125] and Welz.[126] Welz used the Zeeman background correction.

The method for the determination of phosphorus in stainless steel by ICP/AES at the wavelength 253.40 nm was described by R. Diemiaszonek.[53] Several emission lines were analyzed by Wallace.[127] λ 214.91 nm was used by Ward,[42] λ 178.28 nm by Hughes.[15] Phosphorus in steel was isolated by ion-exchange chromatography and analyzed by ICP/AES by Demidova.[128]

3.22 Praseodymium

Praseodymium was determined by ICP/AES at λ 417.9 nm by Butler,[9] while Hughes[15] used λ 180.73 nm.

3.23 Rare-earth Elements

There is only a limited amount of reported work on the rare earth elements, although many absorption lines have been observed. Generally, their determination is carried out by spectrophotometry.[129]

Total rare earths in steel were determined by Zhao[130] by an indirect method based on the reversal of the phosphate suppression of magnesium absorption intensity.

3.24 Selenium

Peterson[131] determined selenium free of interferences in stainless steels in the range 0.2–0.5% by dissolving the sample in hydrochloric acid/nitric acid and treating it with sulfuric acid/phosphoric acid.

Fleming[44] determined selenium by the hydride technique, while Dulski[61] and Bettinelli[49] adopted GFAAS. A spectral interference due to iron was reported. This interference could be eliminated by applying a Zeeman effect background correction.

3.25 Silicon

Knight[84] and McCarthy[132] determined silicon by FAAS. For the direct determination of silicon in cast iron 0.5 g sample is warmed with 25 mL of 1.5 M sulfuric acid until solution is complete. Then 10 mL 12% ammonium persulfate solution are added and the solution is heated to boiling for about one minute until is clear. After cooling the solution is diluted, filtered and nebulized into the flame directly. Zhuang[133] applied Pd salt as a chemical modifier for the determination of Si by GFAAS.

Silicon was determined by ICP/AES at λ 288.16 nm by Ward,[42] and at λ 251.61 nm by Hughes[15] and Dorado Lopez.[134] Belliveau[66] determined silicon by dc plasma emission spectrometry.

3.26 Silver

Silver is a deleterious element in steel and can be determined by FAAS with an air/acetylene flame as low as 0.0002%.[21] Hofton[135] determined silver after extraction of dithizonate in MIBK. Traces of silver were determined by GFAAS by Dulski.[61] Aziz Alrahman[136] determined silver on solid samples by GFAAS. Hughes[15] determined silver by ICP/AES at λ 328.07 nm.

3.27 Sulfur

Sulfur was determined by Mendez Garcia[137] in low-alloy steel down to 20 ppb by mixing an alkaline solution of the sample with 6 M hydrocloric acid and injecting the generated hydrogen sulfide directly into the plasma.

3.28 Thallium

Thallium was determined by GFAAS by Dulski.[61] He found that a direct determination by the graphite furnace technique brought a saving of time and provided satisfactory results. Calibration was carried out through spiked samples.

3.29 Tantalum

Tantalum was determined by FAAS by Obrusnik.[138]

A concentration of 0.0012% of tantalum was determined at the wavelength 240.06 nm by DCP/AES by Belliveau[66] and by ICP/AES by Ward[42] and Danzaki[139] at λ 240.06 nm. Hughes[15] and Diemiaszonek[53] determined tantalum at λ 295.19 nm.

3.30 Tellurium

Marcec[140] determined traces of tellurium by precipitating with tin (II) chloride and extracting the diethyldithiocarbamate complex with pentylacetate, while Barnett[141] described a procedure without prior extraction where the sample is dissolved in a mixture of HNO_3 and H_2SO_4. Background correction was used. Fleming[44] determined tellurium by the hydride technique while Dulski[61] and Bettinelli[49] by GFAAS.

Tellurium was determined by ICP/AES at λ 238.58 nm by Hughes.[15]

3.31 Tin

Fleming[44] and Gluc[102] determined tin in steels by the hydride technique while Barnett,[141] Ratcliffe,[142] Del Monte,[143] Xuan[144] and Bettinelli[49] applied GFAAS. Weltz[145] accurately determined tin in steels in the range 0.008–0.1% by flow injection hydride generation AAS with matrix-free standard solutions for calibration.

Tin was analyzed at wavelength 326.23 nm by DCP/AES[66] and by ICP/AES at λ 242.17 nm by Diemiaszonek.[53] λ 284.00 nm was used by Hughes.[15] Zhang[52] determined trace amounts of tin by ICP/AES after solvent extraction with MIBK.

3.32 Titanium

Headridge[146] determined titanium in the nitrous oxide/acetylene flame with no interferences. Only the iron concentration was matched.

2 g of steel were dissolved in a mixture of HCl and HNO_3. After the solution was cooled, potassium and lanthanum chloride were added and the solution obtained was diluted to the mark in a 100-mL volumetric flask. The determination of titanium in high-alloy steels and

iron sand was carried out by Negishi.[147] Titanium has been preconcentrated by extraction from 0.1 HCl/0.1 M HF with a solution of Amberlite LA2 in benzene followed by back extraction into the aqueous phase with 1.2 M HCl/0.5 M HF. Mo, V, and Zr interfere. Titanium in stainless steel by ICP/AES at the wavelength 249.68 nm was determined by R. Diemiaszonek[53] and at λ 334.94 nm by Ward,[42] Gomez Coedo[40–41] and Hughes.[15]

Okamoto[148] extended the DL of conventional ICP for titanium determination employing electrothermal atomization.

3.33 Tungsten

Tungsten was analyzed by FAAS by Husler[95] and Tindall.[149] They pulverized the samples and dissolved them slowly in concentrated hydrochloric acid in a water-bath at 60–70 °C. Musil[150] reduced tungsten with tin(II)chloride and extracted the thiocyanate complex into MIBK. The organic phase was nebulized directly into a nitrous oxide/acetylene flame.

Butler et al.[9] and Diemiaszonek[53] determined tungsten by ICP/AES using the wavelength 400.8 nm. Ward[42] used λ 207.91 nm.

3.34 Vanadium

Vanadium was determined by Klein[27] 2 g of steel were dissolved in a mixture of HCl and HNO_3. After the solution was cooled, potassium and lanthanum chloride were added and the solution obtained was diluted to the mark in a 100-mL volumetric flask. The determination of vanadium can be carried out in a nitrous oxide/acetylene flame with no interferences when the iron concentration is matched. Husler[95] used a nitrous oxide/acetylene flame. After the addition of iron to the reference solution and potassium as ionization buffer, the determination in this flame was free of interferences. FAAS was used also by Sire.[96] Knight[84] and Husler[95] determined vanadium in tool steels and Del Monte Tamba[50] determined vanadium by GFAAS. Vanadium by ICP/AES was determined by Hughes[15] using wavelength 311.07 nm. Ward[42] used λ 292.40 nm, while Gomez Coedo[40] used λ 309.31 nm. Okamoto[148] determined vanadium by ICP/AES coupled with an electrothermal source.

3.35 Yttrium

Yttrium was determined by ICP/AES by Karandashev[151] in pure iron and low-alloyed steel by means of extraction through columns of tetraphenylmethylendiphosphine oxide stationary phase on styrene-divinylbenzene macroporous copolymer support.

3.36 Zinc

Frech[104–105] and Takada[62] determined zinc on solid samples by GFAAS. Zinc was determined by ICP/AES at λ 213.86 nm by Hughes.[15]

3.37 Zirconium

Stelze[162] analyzed zirconium in low-alloy steel by FAAS. Most of the metals at 1000 mg L^{-1} or higher concentration level interfere with the zirconium absorbance signal. 0.0010% of zirconium were analyzed at the wavelength 349.62 nm by DCP/AES by Belliveau,[66] and at λ 349.6 nm and at λ 343.82 nm by ICP/AES. Butler,[9] Gomez Coedo,[40] Ward,[42] Hughes[15] used λ 339.20 nm.

4 SAMPLING AND SAMPLE PREPARATION

Iron and steel are sampled from the liquid for process control and as solids for specification checks. In this paper only the sampling of the steel product is considered. Sampling and preparation of the iron and steel samples for the determination of chemical composition are carried out in accordance with the International Standards Organization (ISO) 14284.[153] Samples must be sound, free from blow holes and representative of the mean chemical composition of the sample product. They must be free from surface coatings and from moisture, dirt or other forms of contamination. If necessary, the surface of the metal must be degreased by means of a suitable solvent, making sure that the degreasing process does not affect the correctness of analysis. Chips are obtained by machining (drilling, milling, turning or punching). Tools, machines and containers used during the preparation of the sample must be cleaned beforehand to prevent any contamination of the sample. Machining has to be carried out in a such a way as to avoid overheating. Chips must be thoroughly mixed before weighing the test portion, by rolling or gently tumbling the container. Where machining of the sample to obtain chips is impractical, it can be cut or broken into pieces. These pieces can be crushed using a percussion mortar or a vibratory grinding mill.

AAS and AES are predominantly techniques for the analysis of liquid or dissolved samples. Steel and iron samples, therefore, usually must be brought into solution.

However, procedures have been described and techniques are available for the direct analysis of solid samples.

Iron and steel are usually brought into solution using an acid digestion with hydrochloric acid, nitric acid or

with their mixture in varying proportions, sometimes followed by a perchloric, phosphoric or sulfuric acid fuming, or a fusion. The type of sample pretreatment is governed by the analytical task. For example, steels with an high content of tungsten always require the addition of phosphoric acid after the treatment with HCl/HNO_3 mixture.

In general, an acid digestion is preferred because it does not increase the total dissolved solids content in the test sample solution. However, a few test samples are not completely dissolved by an acid digestion because silicates and other refractory compounds such as oxides, carbides and nitrides may remain undissolved.

If the total content of all analyte is of interest a fusion with sodium carbonate, lithium metaborate is often required, usually followed by an acid treatment with nitric or hydrochloric acid.

In general, care should be taken that the total dissolved solids content in the test sample solution is not too high in order to avoid interferences by concomitants. The viscosity of solution can have a significant effect on nebulization efficiency. The maximum tolerable dissolved solids content depends on the type of matrix and on instrumental details, such as the nebulizer burner system and the means of solution introduction. For example, much higher total dissolved solids contents can be tolerated if the test sample solution is introduced via a carrier liquid (flow injection analysis) instead of by conventional aspiration.

In the literature many acid mixtures for steel dissolution are reported For example, Beyer[154] warmed about 500 mg of sample with 10 mL of 1 + 1 hydrochloric acid and added concentrated nitric acid drop by drop until all soluble metal had been dissolved. Kinson and Belcher[120] found that tungsten steel can be better dissolved in phosphoric acid/hydrofluoric acid. Husler[95] dissolved tool steels in nitric acid/hydrofluoric acid and Schiller[122] employed the same acid mixture for dissolving niobium in low-alloy steels.

When low-alloy steel has to be analyzed by FAAS it can be brought easily into solution also with 5% sulfuric acid and 0.5 g of ammonium persulfate. This type of acid has to be avoided when GFAAS is used. Dissolution in nitric acid is preferred. Hydrochloric acid must also be avoided owing to the problems which may be encountered from the loss of volatile chloride species. If chloride is present, excess nitric acid has to be added to the sample.

The analytical procedures reported in the literature for determinations by plasma AES do not differ drastically, since this technique requires approximately similar conditions (in particular the salt content of the solution and the amount of sample that can be fed into the plasma in a unit of time (minute)).

A simple procedure for dissolving plain carbon steel when silicon and zirconium have to be analyzed by plasma AES is as follows.

0.5 g of sample is gently warmed in a mixture of hydrochloric acid and nitric acid until the reaction ceases. The solution obtained is transferred into a 100 ml polyethylene volumetric flask, cooled and a small amount of hydrofluoric acid is added. The flask is immediately closed. After a few minutes the solution is brought up to the mark with distilled water. For such hydrofluoric acid-bearing solutions the plasma instrument has to be equipped, with an alumina torch.

While dissolution is facilitated by heating, care must be taken during the thermal treatment because some analytes may be lost as volatile species. Volatile loss during acid dissolution can be prevented or minimised by carrying out the digestion in a closed pressure vessel or by boiling under reflux. Also, in steps involving ignition and dry ashing, volatile elements and compounds are particularly prone to complete or partial escape. The analyte may also be lost by precipitation, coprecipitation, or sorption on container walls.

At the end of 1980s closed vessel systems designed to be rapidly heated by microwave energy had been adopted. Microwave heating is used with closed vessels because of increasing boiling temperature and hence increasing efficiency of the dissolution. For example, the performance of nitric acid is enhanced considerably in closed microwave vessels. The boiling point reaches 176 °C at about 75 psi, substantially increasing oxidizing power and decreasing reaction times. A study carried out with the financial aid of the European Commission set up steel dissolution procedures using microwave ovens.[155]

Most steels, with tungsten content below 5%, can be dissolved in such vessels with a very small amount of hydrochloric, nitric and hydrofluoric acid.

Steel with a tungsten content higher than 5% can be dissolved in a microwave oven by treating 0.25 g of sample with 5 mL of hydrochloric acid, 2 mL of nitric acid, 2 mL of phosphoric acid, 5 mL of water and 2 drops of hydrofluoric acid. Alternatively, instead of 2 mL of phosphoric acid, 1 mL of hydrofluoric acid can be used. In this case, when the reaction ceases and the solution is cold, 10 mL of boric acid must be added.

Analyte preconcentration and separation techniques may be used successfully for test sample solutions with low analyte and or high matrix concentrations. Solvent extraction procedures are a particularly attractive separation technique. A popular extraction system uses chelation by ammonium pyrrolidine dithiocarbamate followed by extraction into 4 methyl-pentan-2-one (methyl isobutyl ketone). However, sample solutions resulting from extraction often have a small volume and a high salt content. Procedures such as ion exchange

or coprecipitation may also be used for preconcentration. Off-line manual preconcentration is time-consuming and subject to contamination or analyte losses. On-line procedures using flow techniques work reliably, with a significant reduction of time and reagent consumption and with a greatly reduced risk of contamination. Of course, all analyte preconcentration and separation requires complete dissolution of the analyte in the test sample solution and its conversion into a well-defined oxidation state.

A special case of analyte separation is the conversion into a gaseous compound. The best-known example of this technique is hydride generation. Hydride generation can only be applied for selected elements such as antimony, arsenic, bismuth, selenium, tellurium and tin.

Mercury can be easily reduced to the metal which can be volatilized from the solution by a flow of air or inert gas (cold vapor technique).

Whenever an iron and steel sample is treated in order to obtain a test sample solution, there is a potential risk that analyte is introduced through the use of reagents, vessels and from the surrounding air. As for the grade of the reagents, the nature of the impurities may or may not be important, depending on how the reagent is to be used. If the acid is to be used to digest a sample for the determination of trace metals, the presence of metallic impurities in the acid may have a significant effect on the results. For this application, high-purity, metal-free acids are required. Significant contamination can be caused by elements that occur in high concentrations in the environment (e.g. aluminum, calcium, iron, silicon, sodium or zinc). Contamination by laboratory vessels, tools, reagents and loss of the analyte by sorption or volatilization, must be avoided during test sample solution preparation. When needed, ultrapure reagents should be used in conjunction with 18–25 MΩ cm deionized water. Both new and used laboratory vessels must be scrupolously cleaned before use, especially for work at trace levels.

In order to detect unintended analyte introduction and to minimize the resulting errors, it is essential to prepare blank solutions for each batch of samples.

The reagent blank solution must contain all reagents that were used to prepare the test sample solution, in the same concentration as in the test sample solution. If the test portion is subject to an acid digestion, a fusion or any other pretreatment, such as a preconcentration or a separation step, the reagent blank solution should be treated in the same way.

Random contamination usually manifests itself by very large relative fluctuations between replicate blanks.

Contamination may be minimized by decreasing handling through automation, using a separate set of labware for each type of test samples and each type of determination, and or by working in closed systems.

5 SPECTROMETRIC MEASUREMENT PROCESS

Ideally, one would like instruments to produce a noise-free response linearly related to the amount of determinand. In practice, optimization of the signal usually requires both instrumental adjustment and sample modification, either through classical methods of separation (i.e. the removal of interferent elements, and/or the preconcentration of analytes), the addition of matrix modifiers, or buffers, or releasing agents, or the use of matrix-matched reference solutions in order to eliminate or minimize interferences.

After having chosen the wavelength recommended by the instrument manufacturer or by the literature, a preliminary analysis has to be carried out in order to detect the presence of interferences and to evaluate how to avoid them.

The approach for avoiding interferences depends on the nature of the analyte and on the matrix (low-alloy steel, stainless steels, etc.). Optimization of the instrument will be carried out during the determination procedure according to the pertinent guides.[156–160]

5.1 Interferences

Interferences are generally classified into spectral and nonspectral interferences. Interferences can be specific for a particular analyte or be nonspecific. Their influence on the analytical signal can be additive (causing a parallel shift of the analytical function) or multiplicative (changing the slope of the analytical function).

Spectral interference is a general term used to cover various forms of overlapping of lines from the matrix and the concomitant elements with the analytical line of the analyte element. This overlapping can take the following forms:

- direct coincidence of lines
- overlapping of very close lying lines
- overlapping of broadened lines
- continuum radiation.

Nonspectral interferences are due to a change in the number of free atoms of the analyte element. They are classified according to the place or stage at which the particular interference occurs (transport, solute volatilization, vapor phase-dissociation, ionization-, spatial distribution). Nonspectral interferences may often be recognized by an underproportional reduction of the measured value upon dilution or by an incomplete or too high recovery of added analyte.

Every spectral interference can adversely influence both the precision and the accuracy of the determination.

Once the presence of spectral interferences has been recognized, an attempt should always be made to see whether the analyte element can be determined free of interferences at another line.

When a matrix effect is present, measures to be taken to remove the interference or to correct for it are a function of the technique used.

In FAAS one or more of the following measures should be taken:

- Use of a flame with a higher temperature and higher reducing power, such as the nitrous oxide/acetylene flame, in which molecular components are more completely dissociated.
- Addition of a spectrochemical buffer in order to enhance the formation of analyte atoms.
- Use of matrix-matching or the method of additions if the interference is nonspecific.
- Separation of the analyte element and interfering concomitants by extraction, precipitation, etc.

In GFAAS matrix effects can be minimized by the

- addition of an appropriate chemical modifier;
- optimization of the temperature program;
- atomization from the L'vov platform;
- using integrated absorbance for signal evaluation;
- use of a more powerful background correction system such as the Zeeman correction effect, instead of continuum source background correction;
- separation of the analyte element and interfering concomitants by extraction, precipitation, etc.;
- use of the method of additions when nonspectral interferences are present.

Loss of As, Se, Te during ashing of the sample can be reduced by the addition of nickel ions to form nickel arsenide and this type of compound.

Cd normally lost at 500 °C can be stabilized by 2% ammonium phosphate up to 1000 °C. Such modification not only reduces losses on ashing, but also permits the use of higher ashing temperature and thus may allow background and other matrix effects to be minimized.

In AES an appropriate choice of spectral lines is necessary in order to avoid line coincidence. Limitations arise from the spectral range of the instrument and from mechanical constraints such as the minimum distance between two exit slits. Compromises are required which often preclude the use of the best line. Typical line coincidence problems are encountered in steel analysis by AES with the following lines: B 249.68 W 249.66, Al 396.15 Mo 396.15, Ti 334.94 Cr 334.93, Mn 257.61 W 257.62, Cu 324.75 Mo 324.76, Pb 220.35 Ni 220.35 Co 220.34 W 220.38. Only one line of P (178.28 nm) is free from interferences in the presence of Fe.

An exact knowledge of the radiation background and of the emission lines of concomitants in the neighborhood of the analytical line is essential to achieve satisfactory correction and accurate results. A means to detect interferences is to compare the spectrum of the pure analyte in acid solution with the spectrum of the analyte in the sample solution and then scanning across a short region (0.1–0.2 nm) of the spectrum in the vicinity of a proposed analyte line. Efficient and accurate correction for these interferences normally requires a means of measurement of the background intensity adjacent to the analyte wavelength, preferably on both sides as well as a means of applying previously determined correction factors to compensate for line overlap interferences.

The internal standard technique is often used in steel analysis when simultaneous instruments are used. In this approach an element that does not occur in measurable quantities in the sample is added in precisely the same amount to all blanks, standard and sample solutions. In many cases yttrium or scandium will meet these requirements. The emission intensity of the internal standard element is measured at its own emission line at the same time and in the same way as that of the analyte. Each element channel's signal is divided by the signal of the reference element before further calculation takes place.

5.2 Calibration

Calibration can be carried out according to three techniques: the standard calibration technique with or without matrix-matching, the bracketing technique and the method of additions.

In the standard calibration technique, calibration solutions with equally spaced analyte concentrations that span the desired working range have to be prepared. The number of these solutions is linked to the technique and to the desired uncertainty. Generally in AAS five solutions must be prepared, where one is the blank value. In AES, due to its wide linearity range, calibration is possible even with a single sample, with a suitable blank.

In the bracketing technique, calibration solutions just below and just above the analyte concentration of the test sample solution are used.

In the method of additions the approximate concentration level of the test sample solution must be known in order to establish the analyte addition volumes that must be used. The concentration of the test sample solution with the highest addition must still fall within the linear range. The method of additions cannot be used when spectral interferences are present.

Generally, in iron and steel analysis, when the standard calibration or bracketing technique is used, the main constituents of the matrix (iron, in the case of nonalloyed

steel; iron, chromium, and nickel in alloyed steel) and the acid medium of the calibration solutions must mimic those of the test sample. If the test solution is obtained through preconcentration and separation procedures the blank calibration solution has to be prepared by the same pretreatment procedure.

When very low concentrations of analyte must be determined, or when it is not possible to completely match the test sample, the method of additions is preferred. Its application in AES is seldom necessary, however.

6 QUALITY ASSURANCE CRITERIA

In iron and steel analysis, as in many other fields, the use of validated methods as well as traceability calibration have become a subject of increasing and fundamental significance.

''Validated methods'' means that the required tolerance of all measurements undertaken within the method (volume, temperatures, masses), the forms of the determined measured, including speciation, and the effect of the interferences, have been widely investigated and quantified under realistic conditions, significant sources of error have been singled out and adequate means of controlling them have been identified.

Several approaches may be used to validate methodology. When reference samples are available that are similar in all respects to the test samples, the process is very simple. It consists of analyzing a sufficient number of reference sample and comparing the results to the expected or certified values. The use of appropriate reference materials can also enable analysts to demonstrate the accuracy of the results, calibrate equipment and methods, monitor laboratory performance and enable comparison of methods by their use as transfer standards.

When suitable reference material is not available, several other approaches can be used. For example it is possible to compare the results of the method to be validated against those obtained from the same sample using other validated methods. (When using validated methods it is necessary to have a complete understanding of them before working through them for the first time. This will ensure the greatest degree of reliability from them and will allow one to compare the results obtained from one method against those obtained by another method based on completely different physico-chemical principles.)

One possible tool of validation is the use of spiked samples. Spiking does not necessarily have to be restricted to the given analyte. Spiking may include anything added to the sample in order to gauge the effect of the addition. The sample could be spiked with varying amounts of a particular interferent, for example, with the aim of establishing at what concentration of interferent the determination of the analyte is adversely affected. The splitting of the sample gives an idea of the variation in the operational condition which can be expected during routine use of the method.

Validation can also be carried out through collaborative tests (interlaboratory study, proficiency testing schemes, certification exercises etc.). Collaborative tests can identify systematic errors.

Measurement on analogous reference materials provides a further validation tool. Even if an exact match between reference material and sample cannot be obtained, the use of the closest match available is better than nothing. It is important to take the reference material through the whole analytical procedure. This will provide a true reflection of the capability of the method and the equipment.

Many Certified Reference Materials (CRMs) are available for iron and steel analysis. Included among steel CRM producers are NIST (National Institute of Standards and Technology), BAM (Bundesanstalt fur Materialforschung und Prufung), IRSID (Institute de Recherche de la Siderurgie), BS (Brammer Standard Company), BAS (Bureau of Analyzed Samples), CTIF (Centre Technique des Industries de la Fonderie), JK (Jerkontoret), JSS (The Iron and Steel Institute of Japan) and others. BAM, BAS, CTIF, IRSID and JK are members of a producers' group founded for the production of CRM in the iron and steel industry, originally responsible to the European Community for Coal and Steel. Today this group is acting on behalf of the European Committee for Iron and Steel Standardization (ECISS). The finished samples are sold under the name EURONORM CRM.

As concerns traceability, much has been written and many standardizing bodies are working to more clearly define the scientific and statistical requirements for achieving measurement traceability. However, a clear definition of the scientific and statistical requirements for achieving measurement traceability in iron and steel analysis is missing too.

Therefore, in the framework of ISO (ISO TC 17 SC1) an ''ad hoc'' group has been set up in order to study the traceability of steel analysis from SI units to routine determinations. The group would investigate documents concerning existing CRM, compare certified values and those from SC1 tests, draft requirements for methods and CRMs and ultimately publish a guideline document.

Table 3 Summary of ISO TC17 AAS and ICP/AES methods

FAAS	
ISO 4940 Determination of nickel content	
Range of application	0.002% – 0.5%
Standard deviation of the most concentrated solution	<1.0% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	<0.30 μg mL^{-1} λ 352.5 nm <0.15 μg mL^{-1} λ 232.0 nm
Characteristic concentration	<0.50 μg mL^{-1} λ 352.5 nm <0.10 μg mL^{-1} λ 232.0 nm
Repeatability at 0.002%	0.0007
Reproducibility	0.010
Repeatability at 0.5%	0.0176
Reproducibility	0.0435
ISO 4943 Determination of copper content	
Range of application	0.004% – 0.5%
Standard deviation of the most concentrated solution	<1.0% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution 1.0%
Limit of detection	<0.15 μg ML^{-1}
Characteristic concentration	<0.10 μg ML^{-1}
Repeatability at 0.005%	0.0004
Reproducibility	0.0007
Repeatability at 0.5%	0.028
Reproducibility	0.046
ISO 9647 Determination of vanadium	
Range of application	0.005% – 1.0%
Standard deviation of the most concentrated solution	<1.0% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.3 μg ML^{-1}
Characteristic concentration	1.0 μg ML^{-1}
Repeatability at 0.005%	0.0005
Reproducibility	0.0016
Repeatability at 1.0%	0.0268
Reproducibility	0.0404
ISO 9658 Determination of aluminum	
Range of application	0.005% – 0.20%
Standard deviation of the most concentrated solution	1.5% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.1 μg ML^{-1}
Characteristic concentration	1.0 μg ML^{-1}
Repeatability 0.005%	Soluble 0.00073 Total 0.00111
Reproducibility	Soluble 0.00224 Total 0.00212
Repeatability 0.20%	Soluble 0.00472 Total 0.00416
Reproducibility	Soluble 0.0136 Total 0.00961

(*continued overleaf*)

Table 3 (*continued*)

FAAS	
ISO 10138 Determination of chromium	
Range of application	0.002% – 2.0%
Standard deviation of the most concentrated solution	<1.5% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.05 μg mL^{-1} λ 357.9 nm
	<0.15 μg mL^{-1} λ 425.4 nm
Characteristic concentration	<0.2 μg mL^{-1} λ 357.9 nm
	<0.5 μg mL^{-1} λ 425.4 nm
Repeatability 0.002%	0.00030
Reproducibility	0.00074
Repeatability 2.0%	0.0251
Reproducibility	0.00576
ISO 10697-1 Determination of calcium content. Part 1. Determination of acid soluble calcium content	
Range of application	0.0005% – 0.003%
Standard deviation of the most concentrated solution	<1.5% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.02 μg ML^{-1}
Characteristic concentration	0.05 μg ML^{-1}
Repeatability 0.0005%	0.00018
Reproducibility	0.00023
Repeatability 0.0030%	0.00028
Reproducibility	0.00052
ISO 10697-2 Determination of calcium content. Part 2 Determination of total calcium content	
Range of application	0.0005% – 0.005%
Standard deviation of the most concentrated solution	<1.5% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.02 μg mL^{-1} λ 422.7 nm
Characteristic concentration	0.05 μg mL^{-1}
Repeatability at 0.0005%	0.00017
Reproducibility	0.00030
Repeatability at 0.0050%	0.00051
Reproducibility	0.00078
ISO 10700 Determination of manganese content	
Range of application	0.002% – 2.0%
Standard deviation of the most concentrated solution	<1.5% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.02 μg mL^{-1} λ 279.5 nm
Characteristic concentration	0.1 μg ML^{-1}
Repeatability at 0.002%	0.00029
Reproducibility	0.00062
Repeatability at 2.0%	0.024
Reproducibility	0.061
ISO 11652 Determination of cobalt content	
Range of application	0.003% – 5.0%
Standard deviation of the most concentrated solution	<1.5% of the mean absorbance of the most concentrated solution

Table 3 (*continued*)

FAAS	
Standard deviation of the least concentrated solution	<0.5% of the mean absorbance of the most concentrated solution
Limit of detection	0.05 μg mL^{-1} λ 240.7 nm
Characteristic concentration	0.3 μg mL^{-1}
Repeatability at 0.003%	0.00019
Reproducibility	0.00058
Repeatability at 5.0%	0.081
Reproducibility	0.704
ETAAS	
ISO 10698 Determination of antimony content	
Range of application	0.0005% – 0.010%
Standard deviation of the most concentrated solution	<10% of the mean absorbance of the most concentrated solution
Standard deviation of the least concentrated solution	<4% of the mean absorbance of the most concentrated solution
Limit of detection	20 pg λ 217.6 nm
Characteristic mass	25 pg
Repeatability at 0.0005%	0.00020
Reproducibility	0.00035
Repeatability at 0.01%	0.00092
Reproducibility	0.0022
ICP	
ISO 10278 Determination of manganese content	
Range of application	0.02% – 1.5%
Standard deviation	<0.9%
DL	0.02 mg L^{-1} λ 257.61 nm
BEC	0.6 mg L^{-1}
Repeatability at 0.02%	0.00023
Reproducibility	0.00064
Repeatability at 1.5%	0.0266
Reproducibility	0.0535
ISO 13898-1 Nickel,Copper, Cobalt. Sample dissolution	
ISO 13898-2 Nickel, Copper, Cobalt. Determination of nickel content	
Range of application	0.001% – 0.30%
Standard deviation of the most concentrated solution	<0.4%
DL	0.04 mg L^{-1} λ 231.60 nm
BEC	1.5 mg L^{-1}
Repeatability at 0.001%	0.00039
Reproducibility	0.00076
Repeatability at 0.3%	0.0075
Reproducibility	0.023
ISO 13898-3 Nickel,Copper, Cobalt. Determination of copper content	
Range of application	0.001% – 0.40%
Standard deviation of the most concentrated solution	<0.4%
DL	0.02 mg L^{-1} λ 324.75 nm
BEC	0.8 mg L^{-1}
Repeatability at 0.001%	0.00016
Reproducibility	0.00023
Repeatability at 0.4%	0.0038
Reproducibility	0.014

(*continued overleaf*)

Table 3 (*continued*)

ICP	
ISO 13898-4 Nickel, Copper, Cobalt. Determination of cobalt content	
Range of application	0.001% – 0.10%
Standard deviation of the most concentrated solution	<0.4%
Limit of detection	0.04 mg L^{-1} λ 228.62 nm
BEC	0.8 mg L^{-1}
Repeatability at 0.001%	0.00018
Reproducibility	0.00051
Repeatability at 0.1%	0.0014
Reproducibility	0.0040

7 STANDARD METHODS OF ANALYSIS

Many standard reference methods for steel analysis are solution-based AAS and AES methods.

Such methods are studied and set up on a collaborative basis, in the framework of international, European and national committees for standardization, such as ISO, Comité Européen de Normalisation (CEN), American Society for Testing Materials (ASTM), Association Francais de normalisation (AFNOR), British Standards Institution (BSI), etc. The standards are optimized to give the highest possible accuracy (precision and trueness), which means that the procedures, in many cases, differ from those normally used in routine analysis.

Such standard methods are evaluated by means of experiments in which a number of laboratories take part. The standardization experiment is designed to reveal and solve any technical problems and to estimate the characteristics of the analytical method. After this stage a precision experiment follows. The aim of this exercise is to establish a concentration range, as well as repeatability and reproducibility values for the method.

Reference standard methods are used to establish element concentrations for CRM and to resolve disputes or doubts when methods, analysts, or laboratories produce disparate results.

Table 3 lists the standard methods published by ISO for the determination of elements in iron and steels.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
AFNOR	Association Francais de Normalisation
AISI	American Iron and Steel Institute
ASTM	American Society for Testing Materials
BAM	Bundesanstalt fur Materialforschung und Prufung
BAS	Bureau of Analyzed Samples
BS	Brammer Standard Company
BSI	British Standards Institution
CEN	Comité Européen de Normalisation
CRM	Certified Reference Material
CTIF	Centre Technique des Industries de la Fonderie
DCP	Direct Current Plasma
DL	Detection Limit
ECISS	European Committee for Iron and Steel Standardization
ELTA	Ethyl-*N*-laurylthioacetate
ETAAS	Electro Thermal Atomic Absorption Spectrometry
FAAS	Flame Atomic Absorption Spectrometry
FI/HGAAS	Flow Injection Hydride Generation Atomic Absorption Spectrometry
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
ICP	Inductively Coupled Plasma
ICP/AES	Inductively Coupled Plasma/Atomic Emission Spectroscopy
ICP/MS	Inductively Coupled Plasma/Mass Spectrometry
IRSID	Institute de Recherche de la Siderurgie
ISO	International Standards Organization
JK	Jerkontoret
JSS	The Iron and Steel Institute of Japan
LOQ	Limit of Quantitation
MIBK	Methyl Isobutyl Ketone
NIST	National Institute of Standards and Technology

RELATED ARTICLES

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction • Metal Analysis, Sampling and Sample Preparation in

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

General Articles **(Volume 15)**
Microwave Techniques • Traceability in Analytical Chemistry

REFERENCES

1. A. Walsh, Australian patent 23041, 1953.
2. A. Walsh, 'Application of Atomic Absorption Spectra to Chemical Analysis', *Spectrochim. Acta*, **7**, 108 (1955).
3. C.T.J. Alkemande, J.M.W. Milatz, 'Double Beam Method of Spectral Selection with Flames', *Appl. Sci. Res. Section B.*, **4**, 289 (1955).
4. H. Massman, 'Atomic Absorption Analysis with a Heated Graphite Cell', *Method. Phys. Anal.*, **4**, 193 (1968).
5. B.V. L'vov, 'Atomic Absorption Spectra by Complete Vaporisation of a Substance in a Graphite Cell', *Inzh. Fizh. Zh.*, **2**, 44 (1959).
6. S. Greenfield, I.L.L. Jones, C.T. Berry, 'High Pressure Plasmas as Spectroscopic Emission Sources', *Analyst*, **89**, 713 (1964).
7. V.A. Fassel, R.A. Knisely, 'Inductively Coupled Plasma Optical Emission Spectrometry', *Anal. Chem.*, **46**, 11110A (1974).
8. C.B. Belcher, H.M. Bray, 'Determination of Magnesium in Iron by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **26**, 322 (1962).
9. C.C. Butler, V.A. Fassel, R.N. Kniseley, 'Inductively Coupled Plasma Optical Emission Spectrometry. Application to the Determination of Alloying and Impurity Elements in Low and High Alloy Steels', *Anal. Chem.*, **47**, 825 (1975).
10. Centro Sperimentale Metallurgico Spa (I, 00129 Rome, Italy), 'Application of Flameless Atomic Absorption Spectroscopy in Iron Metallurgical Analysis', *Comm. Eur. Communities*, (Rep.) EUR 7886, 1983.
11. H. Hocquaux, 'Determination of Some Trace Elements or Elements Present in Low Amount in Steels, Ferrous Metallurgical Products and Industrial Waters by Flameless Atomic Absorption Spectrophotometry', *Comm. Eur. Communities* (Rep.) EUR 7885, 1982.
12. M.G. Del Monte Tamba, 'Determination of Trace Elements in Materials of Interest to Steel Making by Flameless AAS Directly on Compact Specimens', *Comm. Eur. Communities*, (Rep.) EUR 9893, 1985.
13. H. Hocquaux, 'Flameless Atomic Absorption Spectrophotometry Application to Metallurgical Analysis', *Met. Corros. Ind.*, **56**, 216 (1981).
14. H. Hocquaux, 'Atomic Absorption Spectrometry for Analyses of Alloyed Steel', *Colloq. Metall.*, **25**, 207–216 (1982).
15. H. Hughes, ECSC Final Report of the British Steel Corporation Project, Middlesbrough, 1981.
16. K. Ohls, ECSC Final Report of the Hoesch Project, Dortmund, 1980.
17. J. Hancart, ECSC Final Report of the Centre Recherche Metallurgique Project, Liege, 1981.
18. A. Wittman, J. Hancart, W. Hughes, K. Ohls, in ed., R.M. Barnes, *Development in Atomic Plasma Spectrochemical Analysis*, Heyden, Philadelphia, 530, 1981.
19. A. Wittman, ECSC Final Report of the Institute de Recherches de la Siderurgie Project, 1981.
20. K. Ohls, D. Sommer, 'Application of Atomic Absorption Spectrometry in Ferrous Metallurgy', *Anal. Spectroscop. Libr.*, **5**, 227–274 (1991).
21. T. Stafilov, 'Application of Atomic Absorption Spectrometry to the Analysis of Iron and Steel. Application of Electrothermal Atomic Absorption Spectrometry', *Metalurgija*, **29**, 39–44 (1990).
22. T. Kobayashi, 'A Study of Ultra Trace Analysis by Graphite Furnace Atomic Absorption Spectrometry', *Materia*, **33**, 313–318 (1994).
23. W. Bloedorn, T. Rippert, D. Thierig, H. Unger, 'Trace and Elemental Analysis of Steels and Ferroalloys. Emission Spectroscopy with Inductively Coupled Plasma', *Labor. Praxix*, **16**, 876, 880–882, 884, 886 (1992).
24. W. Bloedorn, T. Rippert, D. Thierig, H. Unger, 'Trace and Elemental Analysis of Steels and Ferroalloys. Emission with Inductively Coupled Plasma Excitation, Part 2', *Labor. Praxix*, **16**, 982–987 (1992).
25. A.D. Ambrose, R. Jowitt, 'A Collaborative Study of ICP Analysis', Comm. Eur. Communities, EUR 14113, *Prog. Anal. Chem. Iron and Steel Ind.*, 377–384, 1992.
26. K. Fujimoto, Y. Matsumura, 'Ultra Trace Analysis and Determination Limit by Inductively Coupled Plasma Atomic Emission Spectroscopy', *Materia*, **33**, 319–324 (1994).
27. A.A. Klein, 'Analysis of Low Alloy Steel Using a Sequential Atomic Absorption Spectrophotometer Equipped with an Autosampler', in *New Analytical Techniques for Trace Constituents of Metallic and Metal Bearing Ores*, ed., A. Javier-Son ASTM STP747, 29–42, 1980.
28. T.C. Rains, 'Determination of Aluminum, Barium, Calcium, Lead, Magnesium and Silver in Ferrous Alloy by Atomic Emission and Absorption Spectrometry', in *New*

Analytical Techniques for Trace Constituents of Metallic and Metal Bearing Ores, ed., A. Javier-Son ASTM STP747, 43–54, 1980.

29. B. Headridge, A. Sowerbutts, 'The Atomic Absorption Spectrophotometric Determination of Total Aluminum in Steel after its Dissolution in a Pressure Bomb', *Analyst*, **98**, 57–64 (1973).
30. D.C. Manning, 'The Determination of Aluminium by Atomic Absorption Spectroscopy', *At. Absorption Newslett.*, **6**, 24 (1964).
31. F. Brivot, I. Cohort, G. Legrand, J. Lavrier, I. Voinovitch, 'Comparative Determination by Atomic Absorption Spectrometry and by Spectrophotocolorimetry of Nickel, Copper, Chromium, Manganese, Aluminium, and Silicon in Certain Steels', *Analusis*, **2**, 570 (1973).
32. F. Shaw, J.M. Ottaway, 'The Determination of Trace Amounts of Aluminium and Other Elements in Iron and Steel by Atomic Absorption Spectrometry with Carbon Furnace Atomization', *Analyst*, **100**, 217 (1975).
33. J. Yu, Z. Cui, 'Determination of Trace Amounts of Aluminium in Pure Iron and Low Alloy Steel by Graphite Furnace Atomic Absorption Spectrometry', *Fenxi Huaxue*, **11**(8), 612–615 (1983).
34. I. Atsuya, N. Sugiura, 'Determination of Microamounts of Aluminium by Atomic Absorption Spectroscopy using a Heated Graphite Atomizer. Application to Analysis of Steel', *Bunseki Kagaku*, **23**, 1170 (1974).
35. N.D. Zyukova, N.A. Panarina, 'Determination of Metallic Aluminium in Steels by Atomic Absorption with Electrothermal Atomization', *Zavod. Lab.*, **52**(2), 36–37 (1986).
36. J.A. Persson, W. Frech, A. Cedergren, 'Determination of Aluminium in Low Alloy and Stainless Steels by Flameless Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **89**, 119 (1977).
37. J. Collin, J. Sire, J. Merklen, in *Determination of Major Constituents of Oxides Present in Steel*, Proceedings of the International Congress of Atomic Absorption Fluorescence, Paris, 1971, John Wiley & Sons, New York, 713, 1973.
38. T. Kobayashi, K. Ide, H. Okochi, Y. Shimada, 'Determination of Trace Amount of Aluminium in Iron and Steels by Graphite Furnace Atomic Absorption Spectrometry', *Nippon Kinzoku Gakkaishi*, **55**, 970–974 (1991).
39. H. Naka, H. Kurayasu, Y. Inokuma, 'Determination of Acid Soluble Aluminium in Steel', *Sumitomo Kinzoku*, **42**, 9–14 (1990).
40. A. Gomez Coedo, M.T. Dorado Lopez, J.L. Jimenez Seco, 'Application of Plasma Spectrometry to the Determination of Aluminium, Vanadium, Titanium, Zirconium and Niobium in Steels', *Rev. Metal. CENIM*, **15**, 97 (1979).
41. A. Gomez Coedo, J.L. Jimenez Seco, I. Gutierrez Cobo, 'Determination of Aluminium, Niobium and Titanium in Microalloyed Steels by Using Inductively Coupled Plasma (ICP) with Spark Preatomisation', *7th Cong. Nac. Cienc. Tecnol. Metal*, **3**, 243–250, 1990.
42. A.F. Ward, L.F. Marciello, 'Analysis of Metal Alloy by Inductively Coupled Argon Plasma Optical Emission Spectrometry', *Anal. Chem.*, **51**, 2264 (1979).
43. G. Hein, 'Determination of Trace Elements in Steel by Means of Flame Atomic Absorption Spectrometry', *Metalurgija*, **29**, 79–82 (1990).
44. H.D. Fleming, E.G. Ide, 'Determination of Volatile Hydride Forming Metals in Steels by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **83**, 67 (1976).
45. B. Vanloo, R. Dams, J. Hoste, 'Determination of Antimony in Steel by Hydride Generation and Electrothermal Zeeman Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **160**, 329 (1984).
46. H. Bombach, B. Luft, E. Weinhold, F. Mohr, 'Determination of Arsenic, Antimony, Bismuth, Tin, Selenium and Tellurium in Steels Using Hydride Atomic Absorption Spectrometry', *Neue Huette*, **29**, 233 (1984).
47. W.B. Barnett, E.A. McLaughlin, 'The Atomic Absorption Determination of Antimony, Arsenic, Bismuth, Cadmium, Lead and Tin in Iron, Copper and Zinc Alloys with the Graphite Furnace', *Anal. Chim. Acta*, **80**, 285 (1975).
48. W. Frech, 'Rapid Determination of Antimony in Steels by Flameless Atomic Absorption', *Talanta*, **21**, 565 (1974).
49. M. Bettinelli, U. Baroni, G. Bizzarri, 'Determination of Trace Elements in Steels and Nickel Based Alloy using the Stabilized Temperature Platform Furnace and Zeeman Effect Background Correction', *Comm. Eur. Communities* EUR 14113, *Prog. Anal. Chem Iron and Steel Ind.*, 566–575, 1992.
50. M.G. Del Monte Tamba, N. Luperi, 'Determination of Traces of As, Sb, Bi, V in Steels and Cast Iron by Graphite Furnace Atomic Absorption Spectroscopy', *Metall. Ital.*, **72**, 253 (1980).
51. B. Weltz, M. Sucmanova, 'L-cysteine as a Reducing and Releasing Agent for the Determination of Antimony and Arsenic Using Flow Injection Hydride Generation Atomic Absorption Spectrometry. Part 2 Interferences Studies and the Analysis of Copper and Steel', *Analyst*, **118**, 1425 (1993).
52. J. Zhang, Z. Li, H. Huang, Z. Hu, L. Zhu, 'Determination of Trace Lead, Arsenic, Antimony, Bismuth, and Tin in Iron and Steel by Solvent Extraction and Inductively Coupled Plasma Atomic Emission Spectrometry', *Fenxi Huaxue*, **19**, 677–679 (1991).
53. R. Diemiaszonek, J.L. Mouton, C. Trassy, 'Study and Adaption of a 40 MHz Inductively Coupled Plasma Source to the Determination of Some Elements in Steels', *Analusis*, **7**, 96 (1979).
54. A. Mendez Garcia, E. Sanchez Uria, A. Sanz Medel, 'Tandem On Line Continuous Separation and Determination of Arsenic by Inductively Coupled Plasma Atomic

Emission Spectrometry', *Anal. Chim. Acta*, **234**, 133–139 (1990).
55. M. Nakayama, 'Method for Determination of Arsenic in Steel by Inductively Coupled Plasma Atomic Emission Spectroscopy with Removal of Metal Ions', *Jpn. Kokai Tokkyo, Koho*, **JP 03**, 105, 249 (1991).
56. E.A. Ozaki, E. de Olivera, 'Simultaneous Determination of Arsenic, Bismuth, and Antimony in Steels and Nickel Alloys by Inductively Coupled Plasma Atomic Emission Spectrometry with Hydride Generation', *J. Anal. At. Spectrom.*, **8**, 367–370 (1993).
57. C.B. Belcher, in *CRC Critical Reviews in Analytical Chemistry*, 121–186, Vol. 7, 1978.
58. D.J. David, in *Flame Emission and Atomic Absorption Spectrometry*, eds., J.A. Dean and T.C. Rains, Marcel Dekker, New York, 33–64, Vol. 3, 1975.
59. B. Welz, M. Melcher, 'Determination of Antimony, Arsenic, Bismuth, Selenium, Tellurium and Tin in Metallurgical Samples Using the Hydride Atomic Absorption Technique. Analysis of Low Alloy Steels', *Spectrochim. Acta*, **36B**, 439 (1981).
60. B. Vanloo, R. Dams, J. Hoste, 'Determination of Lead and Bismuth in Steels and Cast Irons by Zeeman Atomic Absorption Spectroscopy and Hydride Generation', *Fonderie Belge*, **53**, 7–16, (1983).
61. T.R. Dulski, R.R. Bixler, 'Determination of Trace Impurities in Iron and Nickel Base Alloys by Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **91**, 199 (1977).
62. K. Takada, K. Hirokawa, 'Origin of Double Peak Signals for Trace Lead, Bismuth, Silver and Zinc in a Microamount of Steel in Atomic Absorption Spectroscopy with Direct Electrothermal Atomization of a Solid Sample in a Graphite-cup Cuvette', *Talanta*, **29**, 849 (1982).
63. W. Frech, E. Lundberg, M.M. Barbooti, 'Direct Determination of Trace Metals in Solid Samples by Atomic Absorption Spectrometry with Electrothermal Atomizer. Part 4. Interferences Effects in the Determination of Lead and Bismuth in Steels', *Anal. Chim. Acta*, **131**, 45 (1981).
64. Solange Cadore, Nivaldo Baccan, 'Continuous Hydride Generation System for the Determination of Trace Amounts of Bismuth in Metallurgical Materials by Atomic Absorption Spectrometry Using an On Line Stripping Type Generator/Gas–Liquid Separator', *J. Anal. At. Spectr.*, **12**, 637–642 (1997).
65. M. Bettinelli, S. Spezia, G. Bizzarri, 'Determination of Bismuth in Steels and Nickel Alloys; Comparison between FI-HGAAS and GFAAS', *At. Spectrosc.*, **13**, 75–80 (1992).
66. J. Belliveau, H. Griffin, A. Savolainen, 'Multielement DC Plasma Emission Spectrometric Analysis of Steels and Irons', in *New Analytical Techniques for Trace Constituents of Metallic and Metal Bearing Ores*, ed., A. Javier-Son ASTM STP747, 77–90, 1980.
67. J.M. Baena Liberato, 'Determination of Boron in Steels by Inductively Coupled Plasma Spectrometry', *Comm. Eur. Communities*, EUR 13123, 1990.
68. J.M. Baena Liberato, M.J. Guio, E. Hildago, 'Determination of Boron in Steels at Trace Level by Means of ICP Spectrometry in Organic Medium', *Comm. Eur Communities*, EUR 14113 *Prog. Anal. Chem. Iron and Steel Ind.*, 241–246, 1992.
69. J.M. Baena Liberato, M.J. Guio Bonany, E. Hildago Castro, 'Determination of Boron by Using Inductively Coupled Plasma (ICP)', *7th Congr. Nac. Cienc. Technol. Metal*, **3**, 213–222, 1990.
70. J. Ciba, B. Smolec, 'Determination of Boron in Steel by an ICP Emission Spectrometric Technique', *Fresenius' J. Anal. Chem.*, **348**, 215–217 (1994).
71. A. Lopez Molinero, A. Ferrer, J.R. Castillo, 'Volatilisation of Methyl Borate in Iron Matrix. Determination of Boron in Steel by ICP Atomic Emission Spectrometry', *Talanta*, **40**, 1397–1403 (1993).
72. W. Liu, J. Li Hwang, H. Yin, 'Determination of Cadmium in Stainless Steels by Platform Graphite Furnace AAS', *Guangpuxue Yu Guangpu Fenxi*, **12**, 109–110, 116 (1992).
73. E. Lundberg, W. Frech, 'Direct Determination of Trace Metals in Solid Samples by Atomic Absorption Spectrometry with Electrothermal Atomizer Part 2. Determination of Lead in Steel and Nickel Base Alloys', *Anal. Chim. Acta*, **104**, 75 (1979).
74. C.B. Belcher, M.L. Taylor, 'Determination of Calcium and Magnesium in Steel by Atomic Absorption Spectrophotometry', *Anal. Chim. Acta*, **45**, 219 (1969).
75. F. Tugane, M. Kamakura, F. Aoyama, 'Determination of Trace Amounts of Calcium in Steel by Atomic Absorption Analysis', *Tetsu to Hagane*, **57**, 1192, (1971).
76. J.B. Headridge, J. Richardson, 'The Determination of Trace Amounts of Calcium in Stainless Steel by Solvent Extraction Followed by Atomic Absorption Spectrophotometry', *Analyst*, **94**, 968 (1969).
77. T. Glenc, J. Jurczyk, 'Determination of Trace Amounts of Calcium in Low and High Alloy Steels and Nickel Based Alloy by Atomic Absorption Spectrometry with Electrothermal Atomisation', *Hutnik*, **48**, 147 (1986).
78. J. Alvarado, F. Campos, J.M. Ottaway, 'Determination of Trace Levels of Calcium in Steels by Carbon Furnace Atomic Absorption and Atomic Emission Spectrometry', *Talanta*, **33**(1), 61–65 (1986).
79. A. Golloch, W. Funk, E. Thiemann, 'Development of an Analytical Method for the Determination of Metallic and Chemically Bound Calcium in Steels', *Steel Research*, **61**, 117–120 (1990).
80. S. Sprague, W. Slavin, 'The Determination of Copper, Nickel, Cobalt, Manganese, Magnesium in Iron and Steels by Atomic Absorption Spectrophotometry', in *Development in Applied Spectroscopy*, ed., E.N. Davis, Plenum Press, New York, Vol. 4, 1965.

81. W.D. Cobb, W.W. Foster, T.S. Harrison, 'Determination of Cobalt in Steels by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **60**, 430 (1972).
82. S. Gregorczyk, A.W. Wycislik, 'Determination of Manganese, Chromium, Nickel, Molybdenum, Cobalt, Copper in Low Alloy and Medium Alloy Steel by Atomic Absorption Spectrometry with Acetylene Air Flame', *Chem. Anal.*, **24**, 529 (1972).
83. A. Wycislik, 'Use of Secondary Spectral Lines in the Analysis of Alloy Steels and Electrochemically Isolated Structural Constituents by AAS Method', *Comm. Eur. Communities*, EUR 14113, *Prog. Anal. Chem. Iron and Steel Ind.*, 542–546, 1992.
84. D.M. Knight, M.K. Przyna, 'Determination of Copper, Chromium, Cobalt, Manganese, Molybdenum, Nickel, Silicon, Tungsten, Vanadium in Tool Steel by Atomic Absorption Spectrometry', *At. Absorpt. Newslett.*, **8**, 129 (1969).
85. E.J. Eskell, M.E. Pick, 'Electrothermal Atomic Absorption Spectrometric Determination of Cobalt in Steel after Chloroform Extraction of its 1-nitroso-2-naphthol Complex', *Anal. Chim. Acta*, **117**, 275 (1980).
86. K. Kinson, R.J. Hodges, C.B. Belcher, 'The Determination of Chromium in Low Alloy Irons and Steels by Atomic Absorption Spectrophotometry', *Anal. Chim. Acta*, **29**(2), 134–138 (1963).
87. L. Barnes Jr, 'Determination of Chromium in Low Alloy Steel by Atomic Absorption Spectrometry', *Anal. Chem*, **38**, 1083 (1966).
88. A. Gomez Coedo, J.L. Jimenez Seco, 'Determination of Chromium in Steel by Atomic Absorption Spectrophotometry', *Rev. Metal.*, **4**, 316 (1968).
89. D.R. Thomerson, W.J. Price, 'Determination of Chromium and Molybdenum in a Complete Range of Steels by Atomic Absorption Spectrometry with a Nitrous Oxide Acetylene Flame', *Analyst*, **96**, 321 (1971).
90. J.M. Ottaway, N.K. Pradhan, 'Determination of Chromium in Steel by Atomic Absorption Spectrometry with an Air Acetylene Flame', *Talanta*, **20**, 927 (1973).
91. W.D. Cobb, W.W. Foster, T.S. Harrison, 'The Determination of Chromium in Plain Carbon Steel and Low Alloy Iron and Steel by Atomic Absorption Spectrometry', *Analyst*, **101**, 255 (1976).
92. L.P. Pandly, A. Ghose, P. Dasgupta, A.S. Rao, 'Determination of Chromium in Low Alloy Steel by Atomic Absorption Spectrophotometry', *Talanta*, **25**, 482 (1978).
93. R. Grimaldi, G. Pellegrino, M.G. Chiaroni, G. Randi, 'Analysis of Low, Medium and High Alloy Steel by Determination of Chromium, Molybdenum', *Chim. Ind.*, **63**, 801 (1981).
94. S. Gregorcyzik, J. Grolewska, J. Habon, 'Determination of High Content of Manganese, Chromium, Nickel, in High Alloy Steel by Atomic Absorption Spectrometry', *Chem. Anal.*, **18**, 1065 (1973).
95. J.W. Husler, 'Determination of Chromium, Manganese, Molybdenum, Tungsten, and Vanadium in Tool and Alloy Steel by Atomic Absorption Spectrometry', *At. Absorption Newslett.*, **10**, 60 (1971).
96. J. Sire, J. Collin, J.A. Volnovich, 'Physical and Chemical Condition for Determination of Vanadium in Steel by Atomic Absorption Spectrometry and Extension to Copper, Chromium, Nickel and Manganese', *Spectrochim. Acta*, **33B**, 31 (1978).
97. A.G. Fogg, S. Sollymanloo, D. Burns, 'Masking of Iron with Fluoride in Extractive Atomic Absorption Spectrometric Determination of Chromium in Steel', *Talanta*, **22**, 541 (1975).
98. A. Gomez Coedo, M.T. Dorado Lopez, 'Application of Inductively Coupled Plasma Spectrometry to the Determination of Alloy Elements in Steels', *Rev. Metal*, **16**, 157 (1980).
99. I.G. Souza, H. Filho Bergamin, F.J. Krug, B.F. Reis, P.V. Oliveira, J.A. Nobrega, 'Analysis of Steels by Plasma Atomic Emission Spectrometry after Electrodissolution in a Flow Injection System', *Quim. Nova*, **15**, 204–207 (1992).
100. W.K. Chin, 'Method for the Determination of Chromium, Manganese, Molybdenum, Nickel and Boron in Low Alloy Steel Armor Using Atomic Absorption Spectroscopy and Inductively Coupled Argon Plasma Emission Spectroscopy', Report MTL-TR-91-35 Order No AD A 241532, 1991. Abstract no. 206.
101. K. Kinson, C.B. Belcher, 'The Determination of Minor Amount of Copper in Iron and Steel by Atomic Absorption Spectrophotometry', *Anal. Chim. Acta*, **31**, 180 (1964).
102. T. Gluc, J. Jurczyk, A. Robosz Kabza, 'Determination of Tin and Lead in Transformer, Carbon and Alloy Steels in the Range from 0.0005 to 0.04% by Atomic Absorption Method with Electrothermal Atomization', *Chem. Anal.*, **25**, 513 (1980).
103. F. Shaw, J.M. Ottaway, 'Determination of Trace Amount of Lead in Steel and Cast Iron by Atomic Absorption Spectroscopy with the Use of Carbon Furnace Atomisation', *Analyst*, **99**, 184 (1974).
104. W. Frech, A. Cedergren, 'Investigations of Reactions Involved in Flameless Atomic Absorption Procedures. Part 3. Application of High Temperature Equilibrium Calculations to a Multicomponent System with Special Reference to the Interference from Chlorine in the Flameless Atomic Absorption Method for Lead in Steel', *Anal. Chim. Acta*, **82**, 83–92 (1976).
105. W. Frech, G. Lundgren, S.E. Lunner, 'Routine Use of Flameless Atomic Absorption Spectrometry for the Determination of Lead in High and Low Alloy Steels', *At. Absorption Newslett.*, **15**, 57 (1976).
106. W. Frech, A. Cedergren, 'Analytical Investigation of the Reaction Involved in Flameless Atomic Absorption Procedures. Part I. Application of High Temperature Equilibrium Calculations to a Multicomponent System

with Special Reference to the Interference from Chlorine in the Flameless Atomic Absorption Method for Lead in Steels', *Chim. Acta*, **82**, 93–102 (1976).
107. B. Mile, C. Rowlands, A.V. Jones, 'Single Calibration Method for the Determination of Lead in Nickel Alloys and Steels by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **7**, 1069–1077 (1992).
108. D.G. Andrew, A.M. Aziz Alrahrvan, J.G. Headridge, 'Determination of Lead in Irons and Steels by Atomic Absorption Spectrophotometry with the Introduction of Solid Samples into an Induction Furnace', *Analyst*, **103**, 909 (1978).
109. C.B. Belcher, K. Kinson, 'Determination of Manganese in Iron and Steel by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **30**, 483 (1964).
110. I. Atsuya, 'The Determination of Manganese in Iron, Steel Aluminium Alloys and Ferroalloys by Atomic Absorption Spectrometry', *Bunseki Kagaku*, **14**, 592 (1965).
111. D.P. Hubbold, H.H. Monks, 'The Determination of Manganese in a Wide Range of Steels and Cast Irons by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **4**, 197 (1969).
112. D.J. David, in *Flame Emission and Atomic Absorption Spectrophotometry*, eds., W.T. Elwell, J.A.F. Gidley, Pergamon Press, New York, 107, 1966.
113. G.F. Kirkbright, A.M. Smith, T.S. West, 'Rapid Determination of Molybdenum in Alloy Steels by Atomic Absorption Spectroscopy in a Nitrous Oxide–Acetylene Flame', *Analyst*, **91**, 700–704 (1966).
114. Y. Endo, T. Hata, T. Nakahara, 'Determination of Molybdenum in Steels by Atomic Absorption Spectrophotometry', *Bunseki Kagaku*, **18**, 878 (1969).
115. H. Goto, Y. Kakita, M. Namiki, 'Determination of Molybdenum in Steel and Iron by Flame Photometry', *Bunseki Kagaku*, **19**, 1467 (1970).
116. K. Wada, 'Determination of Molybdenum in Steel by Atomic Absorption Spectrometry', *Bunseki Kagaku*, **21**, 221 (1972).
117. J.R. Castillo, M.A. Belarra, J. Arnares, 'Determination of Molybdenum in Steel and Iron by Atomic Absorption Spectrometry after Separation by a Benzoinoxine MIBK Extraction', *At. Spectrosc.*, **3**, 518 (1982).
118. Z. Gu, Z. Lu, Y. Lu, J. Tong, A. Yang, 'Determination of Molybdenum in Cast Iron by Atomic Absorption Spectrometry after Extraction with ELTA and Back Extraction with Aqueous Ammonia', *Guangpuxue Yu Guangpu Fenxi*, **13**, 67–70 (1993).
119. N.V. Bodrov, G.I. Nikolaev, 'Characteristics of the Atomic Absorption Determination of Molybdenum in a Graphite Cell', *Analit. Khim.*, **24**, 1314 (1969).
120. K. Kinson, C.B. Belcher, 'The Determination of Nickel in Iron and Steel by Atomic Absorption Spectrophotometry', *Anal. Chim. Acta*, **30**, 64 (1964).
121. B. Welz, J.E. Sebestyen, Joint. Sympos. Accurate Meth. Anal. Maj. Const., London, 1970.
122. R. Schiller, 'The Determination of Niobium at Concentration below 0.3% in Low Alloys Steel by Atomic Absorption', *At. Absorpt. Newslett.*, **9**, 111 (1970).
123. D.R. Thomerson, 'Direct Determination of Niobium Down to 0.05% in Steels and Other Metallic Alloys by Atomic Absorption Spectrophotometry', *Spectrovision*, **26**, 13 (1971).
124. M.J. Martin, 'Determination of Niobium in Steels by Atomic Spectrophotometry', *Analyst*, **93**, 394 (1972).
125. P.J. Whiteside, W.J. Price, 'Determination of Phosphorus in Steel by Atomic Absorption Spectrophotometry with Electrothermal Atomisation', *Analyst*, **102**, 618–620 (1977).
126. B. Welz, U. Voellkopf, Z. Grobenski, 'Determination of Phosphorus in Steel with a Stabilized Temperature Graphite Furnace and Zeeman Corrected Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **136**, 201–214 (1982).
127. G.F. Wallace, A Comparison of the Capabilities of Atomic Absorption and Inductively Coupled Plasma Spectroscopy for the Analysis of Metallurgical Samples, in *New Analytical Techniques for Trace Constituents of Metallic Bearing Ores*, ASTM STP 747, 53–63, 1981.
128. A.L. Demidova, P.G. Volkova Danilova, N.E. Kuz'menko, 'Determination of Phosphorus in Steel by Spectral Analysis with Inductively Coupled Plasma', *Vestn. Mosh. Univ. Ser. 2 Khim*, **34**, 83–86 (1993).
129. V.A. Alekseevskii, 'Spectrophotometric Determination of Rare Earth Elements in Steels', *Zavod. Lab.*, **58**, 12–13 (1992).
130. G. Zhao, L. Yang, 'Indirect Determination of Total Rare Earth Content Using Air Liquefied Petroleum Gas Flame', *Lihua Jianyan, Huaxue Fence*, **28**, 348–351 (1992).
131. A. Peterson, 'Determination of Selenium in Stainless Steels by Atomic Absorption Spectrometry', *At. Absorption Newslett.*, **9**, 129 (1970).
132. J.P. McCarthy, E.B. Nunn, C. Kinard, 'Atomic Absorption Spectrophotometric Methods for the Determination of Phosphorus and Silicon in Steel', *Anal. Chem. Symp. Ser.*, **19**, 349 (1984).
133. Z. Zhuang, P. Yang, X. Wang, Z. Deng, B. Huang, 'Preliminary Study on the Use of Palladium as a Chemical Modifier for the Determination of Silicon by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **8**, 1109–1111 (1993).
134. M.T. Dorado Lopez, M.A. Palacios Vida, A. Gomez Coedo, 'Use of a Microwave Furnace as a System for Dissolution of Samples for Atomic Absorption and Inductively Coupled Plasma (ICP) Analysis. Application to Determination of Silicon in Steels', *7th Congr. Nac. Cienc. Tecnol. Metal*, **3**, 205–211, 1990.
135. M.E. Hofton, D.P. Hubbard, in *Proceedings of Third International Congress of Atomic Absorption Fluorescence*', Paris, 1971, John Wiley & Sons, New York, 743, 1973.

136. M. Aziz Alrahman, J.B. Headridge, 'Determination of Silver in Iron and Steels by Atomic Absorption Spectroscopy with an Induction Furnace; Direct Analysis of Solid Samples', *Talanta*, **25**, 413 (1978).
137. A. Mendez Garcia, C. Garcia Ortis, J. Sanchez Uria, A. Sanz Mendel, 'Sulfur Determination by Inductively Coupled Plasma Atomic Emission Spectrometry via Continuous Generation of Hydrogen Sulfide', *Quim. Anal.*, **9**, 33–42 (1990).
138. I. Obrusnik, S. Posta, 'Determination of Tantalum in Standard Steels by Instrumental Neutron Activation Analysis and Atomic Absorption Spectrometry', *Radiochem. Radioanal. Lett.*, **32**, 161 (1978).
139. Y. Danzaki, K. Takada, K. Wagatsuma, 'Cupferron Separation for the Determination of Tantalum in High Purity Iron and Low Alloy Steel by ICPAES', *Fresenius' J. Anal. Chem.*, **362**, 421–423 (1998).
140. M.V. Marcec, K. Kinson, C.B. Belcher, 'The Determination of Minor Amounts of Tellurium in Iron and Steels by Atomic Absorption Spectrophotometry', *Anal. Chim. Acta*, **41**, 447 (1968).
141. W.B. Barnett, H.L. Kahn, 'The Determination of Tellurium in Steels by Atomic Absorption Spectroscopy with the Deuterium Background Corrector', *At. Absorption Newslett.*, **8**, 21 (1969).
142. B. Ratcliffe, C.S. Byford, P.B. Osman, 'The Determination of Arsenic, Antimony and Tin in Steels by Flameless Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **75**, 457 (1975).
143. M.G. Del Monte Tamba, N. Luperi, 'Determination of Tin in Plain Carbon, Stainless and High Speed Steels by Atomic Absorption Spectrophotometry with Electrothermal Atomisation', *Analyst*, **102**, 489–494 (1977).
144. W. Xuan, B. Cai, 'Determination of Tin in Iron and Steel by Graphite Furnace Atomic Absorption Spectrometry', *Yejin Fenxi*, **12**, 23–25 (1992).
145. B. Weltz, M. Schubert-Jacobs, T. Guo, 'Investigation for the Determination of Tin by Flow Injection Hydride Generation Atomic Absorption Spectrometry', *Talanta*, **39**, 1097–1105 (1992).
146. J.B. Headridge, D.B. Hubbard, 'The Determination of Titanium in Alloys by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **37**, 151 (1967).
147. A. Negishi, C. Tsurumi, 'Determination of Titanium in High Alloy Steels and Iron Sand by Atomic Absorption Spectrometry after Solvent Extraction with Benzene Solution of Amberlite LA-2', *Nippon Kagaku Kaishi*, **6**, 683–686 (1992).
148. Y. Okamoto, H. Murata, M. Yamamoto, T. Kumamaru, 'Determination of Vanadium and Titanium in Steel by Inductively Coupled Atomic Emission Spectrometry with Modified Use of a Tungsten Boat Furnace Atomizer for Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **239**, 139–143 (1990).
149. F.M. Tindall, 'Hints on Chemical Analyses. Tungsten Determination by Atomic Absorption Spectrophotometry', *At. Absorption Newslett.*, **16**, 37 (1977).
150. J. Musih, J. Dolezal, 'Determination of Small Amounts of Tungsten by Atomic Absorption Spectrometry after Extraction Separation', *Anal. Chim. Acta*, **92**, 301 (1977).
151. V. Karandashev, A.N. Turanov, H.M. Kuss, L. Kumpmann, L.V. Zadnepruk, V.E. Baulin, 'Extraction-chromatographic Separation of Yttrium, Rare Earth Elements, Bismuth, Thorium and Uranium from the Matrix Suitable for the Determination in Pure Iron and Low Alloyed Steels by ICP.MS and ICP.AES', *Mikrochim. Acta*, **130**, 47–54 (1998).
152. W. Stelzer, G.F. Wallace, D.W. Hoult, 'Analyze Low Alloy Steel', *Ind. Res. Dev.*, **22**, 166 (1980).
153. ISO 14284, 1996 Steel and Iron Sampling and Preparation of Samples for the Determination of the Chemical Composition.
154. M. Beyer, 'Preparation of Metallic Air Pollutants for Atomic Absorption Analyses', *At. Absorption Newslett.*, **8**, 23 (1969).
155. M.G. Del Monte, A. Gomez Coedo, 'Applicazione del Forno a Microonde Come Sistema Oggettivo, Sicuro e Rapido per la Dissoluzione e il Trattamento dei Campioni Nell'Analisi Chimica Siderurgica,' *Comm. Eur. Communities*, Rep. EUR 15800 (1996).
156. ISO 12235, 1996, General Guidelines for Inductively Coupled Plasma Emission Spectrometry (ICP.AES).
157. ISO 13204, 1996, General Guidelines for Atomic Absorption Spectrometry – Flame Analysis (FAAS).
158. ISO 13812, 1996, General Guidelines for Atomic Absorption Spectrometry. Graphite Furnace Analysis (ETAAS).
159. ASTM E 663 Standard Practice for Flame Atomic Absorption Analysis. Section: Metals Test Method and Analytical Procedures. Vol: Chemical Analysis of Metals and Metal Bearing Ores.
160. ASTM E 1097 Standard Guide for Direct Current Plasma Emission Spectrometry Analysis. Section: Metals Test Method and Analytical Procedures. Vol: Chemical Analysis of Metals and Metal Bearing Ores.

Automation of Analytical Control in the Steel and Metals Industry

Jean Petin
Mondorf, Luxembourg

1 Introduction **1**
2 History of Automation **2**
2.1 The Cockerill Approach 2
2.2 The Cabin Laboratory Concept 3
2.3 Inferences from These Two Approaches 4
3 Hardware **5**
3.1 The Components 5
3.2 Sample Handling 7
4 Software **8**
4.1 Sample Tracking Software 8
4.2 Software for Instrument Calibration and Quality Assurance 9
5 Problems and Benefits **9**
5.1 Maintenance and Troubleshooting 9
5.2 Economic Aspects 9
6 Outlook **9**
Acknowledgments **10**
Abbreviations and Acronyms **10**
Related Articles **10**
References **10**

In this paper automation is defined as the replacement of operators by automatic devices. The genesis and the functions of automatic systems in the production control laboratories of the steel and metals industry are described.

1 INTRODUCTION

Analysis for production control in the steel and metals industry is performed on solid samples. The two main techniques used are optical emission spectrometry (OES) and X-ray fluorescence (XRF). In both techniques the surface of metallic samples is prepared by taking off the surface layer, and oxidic and mineral samples are prepared as beads either by compression or fusion of a powder. For ancillary techniques small pieces of metallic samples are cut off and weighed before analysis.

Each piece of equipment performing one of these preparation or analysis steps can be considered as a workstation performing its task in a self-contained manner. Such autonomous stations have been developed progressively along with the development of computer technology. Microprocessors became part of the instruments to perform calculations and to control sample introduction, for example via a turntable, whilst sample preparation equipment was equipped with programmable logic controllers (PLCs) to monitor the operation. These were the starting points of automation.

In this paper automation is defined as the replacement of operators by automatic devices. An automated laboratory consists of fully autonomous workstations, a sample transfer system between these stations and the necessary software to control the set-up.

Automation of tasks in analytical chemical laboratories started in the early 1980s using small, human-scale robots to perform repetitive analysis mainly in the pharmaceutical and biochemical industry.

The development started in the USA and was promoted by manufacturers proposing tools and concepts to the user community.[1] Such systems could be installed on benches where the workstations were arranged around a robotized sample transfer system.

The development of laboratory automation for production control in the steel industry started at the same time in Europe, but it was initiated by the user. In view of the size of the equipment and the nature of the environment, the tools used were different from those used in the above-mentioned fields.

As will be seen later these tools had to be, at least in part, developed in response to and in association with the first users.

To some extent equipment and concepts used in the cement industry could be adapted to analytical control in the steel industry. The cement industry was the first to introduce fully automatic laboratories integrating sample taking, preparation and analysis. Cement manufacturing is a continuous process with samples being collected at short intervals, for example every 15 min, averaged and prepared for analysis every 2, 4 or 8 h. An aliquot of the average sample is ground to a given particle size, pressed into a pellet and analysed by XRF. This procedure could be adapted to the analysis of slags and sinters in the steel industry. In a minority of cases the sample is a fused bead, that is, the sample is mixed with a flux, usually a borate, and melted in a platinum crucible to form a glass. The melt is solidified on a platinum dish producing a fused bead. As mineralogical and grain-size effects are suppressed in the fused bead, the analytical results obtained are more accurate, but the technology to be automated is far more complex.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The technology used for the analysis of powdered material in the cement industry was adapted for the analysis of oxidic materials such as ores, slags or sinter in the steel industry by adding where necessary an additional step to crush the sample and take out any metallic particles before the final grinding.

In controlling the steel-making process, however, the primary task is the rapid analysis of pig iron and steel samples taken from the melt. The main analytical technique used for this purpose is spark ablation OES.[2] This technique not only permits analysis for metallic constituents but can also be used for the determination of carbon, sulfur and phosphorus in steel. It has also been possible to detect nitrogen and even oxygen by OES since the early 1990s. Automation of OES is more difficult, as the operator must evaluate the quality of the sample before the analysis and also the quality of the burn produced by the spark discharge process, therefore some objective criteria have to replace the judgement of the operator on the quality of the sample.

As will be seen later, these problems had already been solved with the implementation of the first systems and this had largely contributed to the rapid acceptance of the technology. The nonferrous metallurgical industries such as those dealing with Al-, Cu-, or Zn-based materials have rapidly accepted the concepts developed for steel. The only change made was to the method of preparation of metallic samples. Steel samples are relatively hard and the abrasion techniques used for the preparation of the sample surface are not applicable to the softer nonferrous samples.

2 HISTORY OF AUTOMATION

2.1 The Cockerill Approach

Although some attempts to automate steps of the analysis in steel plants had been made before, the automation project undertaken in the Montignies steel plant of the Belgian steel company Cockerill-Sambre[3] can be considered to be the first successful project to integrate the automation of sample transport, sample preparation and analysis by OES and XRF spectrometry of steel, pig iron and converter slag samples. This was complemented by manually operated C, S and N analyzers.

Planning started in 1982 in response to pressure from the metallurgists asking for a quicker analytical output in order to cope with the requirements for the analytical control of the LD basic oxygen converter steelmaking process.

A new laboratory situated at ground level 8 m beneath the converter floor was installed in a two-storey portable cabin. At the ground floor of this cabin the automated area included the arrival point of the sample transport tubes, a preparation machine for metallic samples, a preparation machine for slags, an X-ray spectrometer, an optical emission spectrometer and an image analyzer for the inspection of steel samples. All this equipment was arranged around an industrial robot (Figure 1). In addition a carbon–sulfur analyzer, a nitrogen analyzer and a glow discharge optical emission spectrometer were installed in the manually operating part of the ground floor. On the first floor of the cabin, ancillary equipment such as the computers, the air conditioning equipment and facilities for the operators were installed.

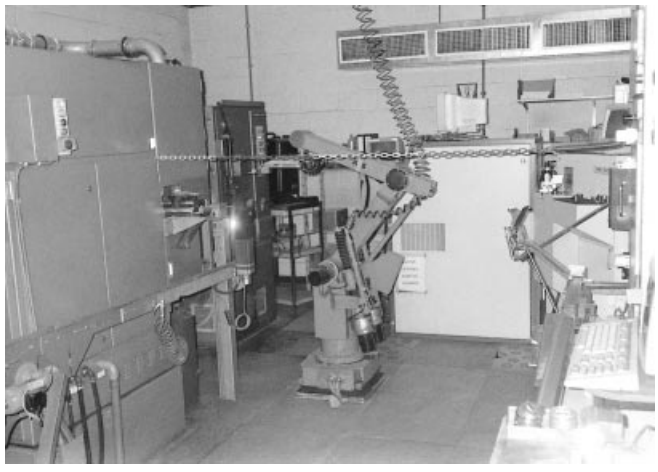

Figure 1 View of an automated laboratory. (Reproduced by permission of Cockerill-Sambre.)

Samples are identified on industrial computer terminals by the steelworks operatives before dispatching to the laboratory. Samples are transported either by gravity or by pneumatic tubes. Metallic samples which are transported pneumatically travel in a so-called open cartridge. The sample is gripped in a holder ahead of a pusher which strikes a ring at the arrival point causing the sample to be released and fall under gravity into a storage tube. Samples transported by gravity arrive directly in the storage tube. There is a separate arrival tube for each type of sample: steel samples from converters, ladle or continuous casting plant; iron samples from the mixer; desulfurization or blast furnace slags. Two computer controlled valves trigger the release one by one of the samples according to the fixed priorities.

The released sample falls onto a belt which is immersed in cooling water before it is automatically introduced into the preparation equipment. The sample is a so-called double medal lollipop sample (Figure 2). In the preparation machine the pin is first cut off by a disk and the thinner part is also removed for the determination of C, S and N. The surfaces of the main sample are then ground on an abrasive band in a manner such as to avoid

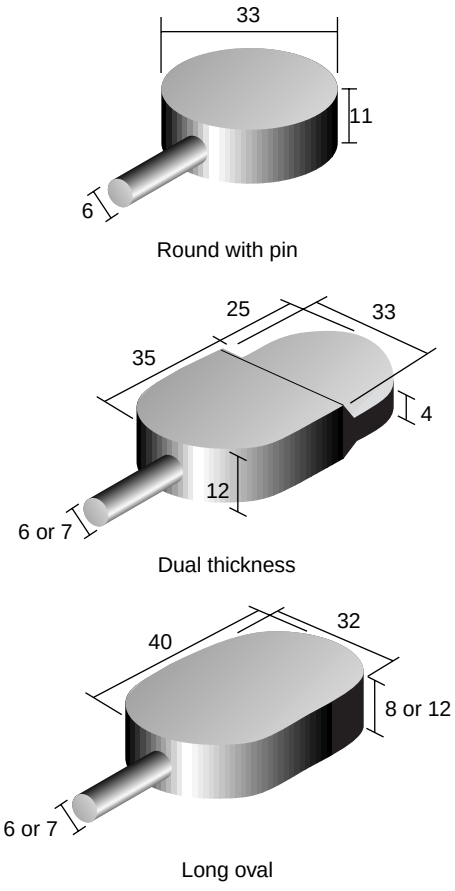

Figure 2 Samples used for liquid steel control. (Reproduced by permission of Soled Industries.)

excessive sample heating. After preparation the sample is taken up by the robot and transferred for inspection to the camera of the image analyzer. Defects are seen by the camera as different shades of gray and the computer determines the location suitable for a maximum of four good sparking areas.

The sample is then transferred by the robot to the spark stand of the OES spectrometer and positioned for spark discharge exposure on one of the predefined sound areas. A minimum of two spark discharges are performed and checked for agreement with the predefined deviation tolerance. If outside the limits, two more exposures can be performed in order to find an agreement. In case of failure a new sample is requested. After exposure the sample is stored by the robot.

Pig iron samples are prepared on the same equipment, but are analyzed by X-ray spectrometry.

Slag samples arrive by gravity in the laboratory and are manually inroduced into the preparation machine. This machine combines a crushing of the sample followed by a magnetic separation of metallic particles. The sample is then finely ground in a swing mill and pressed in a steel ring with a boric acid substrate. This ring is transported by the robot to the sample turntable of the X-ray spectrometer. After analysis the ring is put into a storing position.

Two operators perform ancillary work such as carbon, sulfur and nitrogen determination, thus completing the analysis of the steel samples. They also carry out background work on batch samples or initiate recalibration of the optical emission spectrometer and may also perform troubleshooting activities. In the case of major failure they will use the OES spectrometer to analyse manually steel and pig iron samples.

The Cockerill project was the origin of the development of the automation of production control laboratories in the steel industry and also led to the implementation of laboratory automation in the nonferrous metals industry.

Up to then only the cement industry was equipped with automatic laboratories using X-ray spectrometers for analysing samples in the form of pressed powders. The instrument suppliers were not ready for such automation projects in the steel industry; the major part of the necessary hardware and software had to be developed or adapted to the specific requirements in the steel industry.

The main developments[4] necessary for this project were:

- the open cartridge transport system;
- the preparation machine for steel and iron samples;
- the adaptation of an image analyzer used in metallography to the inspection of steel samples before analysis – a special illumination system for the sample and specific software was required;
- design of the gripping tool of the robot;
- in-house modification of the spark stand of the optical emission spectrometer for the automatic sample loading;
- addition of a metallic iron screening module to the slag preparation machine.

All the software was written in-house in macroassembler using a network of PDP 11 computers supplied by DEC®.

With the exception of the general study, the software and the coordination work performed in-house, the cost of the project was estimated at $1 million.

Thanks to the liberal attitude of the Cockerill staff, laboratory managers from all over the world were able to visit the project and convince themselves that automation of a laboratory for production control was feasible.

2.2 The Cabin Laboratory Concept

The project at Cockerill had shown that automation of spectral analysis of steel samples was possible. Since the

supplies market was not prepared for these developments, it meant that the user himself needed to become deeply involved. This was a costly exercise and at the same time it had been shown that automation did not seem to reduce the time of analysis. These considerations generated an alternative approach which took the form of a demonstration project under the aegis of the European Community for Coal and Steel (ECSC) research organization. The project involved collaboration between a steel company ARBED from Luxembourg and a German spectrometer manufacturer.[5] The idea was to bring the analytical facilities as close as possible to the steelmaking process and to install an emission spectrometer and a surfacing machine for the preparation of the steel samples in an air conditioned enclosure or cabin (Figure 3). The sample was introduced directly by the operator who sampled the melt, from the outside through a small window on to the table of an automatic horizontal surface grinding machine (Figure 4). The transfer between machine and spectrometer was performed by a small robot. This robot also had access to a magazine containing recalibration and control samples, making it possible to run the system in a fully automatic way without an operator being present. A daily maintenance of about 30 min was sufficient to keep the system in good condition. So it became possible to install such a system close to the vessels such as coverters, ladles or electric arc furnaces, where a rapid analytical response was required.

Instead of an image analyzer for the control of the quality of the sample, the intensity of an iron spectral line from a single spark discharge was monitored during the first 3 s of the presparking sequence. If the intensity was beyond a preset level, the discharge was interrupted,

Figure 3 View of the Spectrolux container laboratory. (Reproduced by permission of Spectro A.I.)

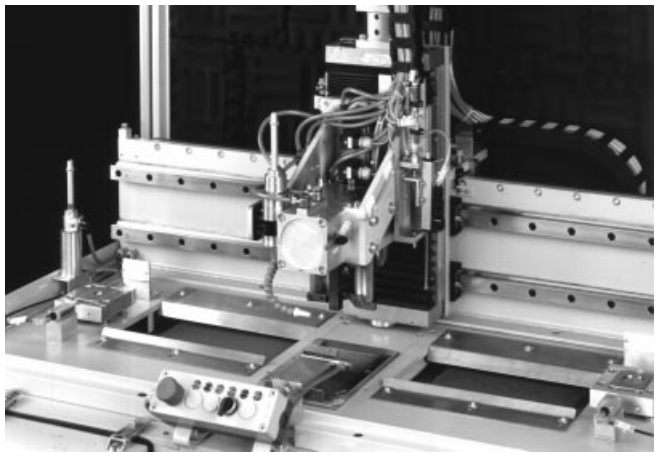

Figure 4 View of an automatic horizontal grinding machine. (Reproduced by permission of Spectro A.I.)

the sample was moved and a new discharge was initiated. This test is known as SEREPS (self-regulated prespark). Two values had to be within predefined limits before a result was released. Experience showed that the number of bad samples was reduced drastically by the fact that the operator who took the sample was aware of this checking process and was confronted 90 s later with the result.

Even during the testing of the system within ARBED and afterwards by Ovako Steel[6,7] and British Steel,[8] other spectrometer manufacturers took up the concept and began installing such systems.[9–11]

The first operational system, called Spectrolux at ARBED Esch-Schifflange works in Luxembourg, was located between two LD-AC converters and its cost was recovered in less than 6 months due merely to the increase in productivity at the steel plant.

British Steel, who had evaluated the prototype developed in the ECSC pilot project and had contributed by specifying the first industrial systems, was the second industrial user of the Spectrolux system.

2.3 Inferences from These Two Approaches

As already mentioned the Cockerill approach, which was a great success, had not only shown the feasibility of automation in this field, but also initiated the development of appropriate or modified equipment by the respective manufacturers. The first follow-up was the cabin concept developed in collaboration between a user and a spectrometer manufacturer. As part of this project a new sample surfacing machine was also developed by the spectrometer manufacturer.

The two main reasons for the development of the cabin system were the higher cost of the Cockerill system and the fact that such automation does not reduce significantly the analysis time.

It had been later shown that the automation of sample transport significantly increased the costs of the automation.[12] Setting up a cabin very close to the production vessels improved the response time compared to an automated laboratory where transport times were dependent on the distance between existing laboratories and the production site.

It is interesting to note that the two systems took almost equal shares of the market between 1985 and 1999.[13] The choice largely depends on the local conditions. Not only is the distance between production site and laboratory important, but the size of the steel-making plant must be considered also. Large plants where the decentralized cabin approach would multiply the installations generally use a centralized arrangement either in the existing premises or by creating new ones.[14–20] Sometimes the two schemes are combined by adding a central laboratory to a decentralized unit close to a production device such as a converter or a ladle furnace where the response time is especially critical.[21] In order to reduce costs, sample transport is not always included in the automation and an operator handles the introduction of the samples arriving by pneumatic tube into the analytical lines. This operator also takes care of any troubles that might arise.

Implementing a combined unit of a sample preparation machine and an optical emission spectrometer is fairly simple whether it is installed in a cabin or in a laboratory. Larger systems combining automatic sample transport, several lines using optical emission, and XRF and even automatic ancillary techniques used for carbon, sulfur or nitrogen analysis need careful planning. As manning levels in the steel industry have been significantly reduced, big projects are now entirely subcontracted. The project coordinator is either a separate engineering company or one of the suppliers. In-house developments are nowadays an exception. Nevertheless the customer is well advised to define his needs and organization very carefully and to ensure a close follow-up with the project coordinator, as equipment and software packages are delivered by different suppliers. A preliminary set-up and function test is highly recommended before delivery.

3 HARDWARE

3.1 The Components

3.1.1 Sample Transport to the Laboratory

In a centralized laboratory controlling the different vessels used in steel production, the samples arrive by pneumatic tube. The metal or slag sample taken from the production vessel is introduced into a carrier which is transported through the pneumatically driven transfer tube. The identity of the sample is registered on a computer terminal installed at the despatch station or deduced from the departure time and the despatch station location. At the automatic receiving station the sample is extracted from the carrier and transferred either by a robot or by a conveyor belt to the sample preparation machine.

In the case of a stage laboratory installed close to the production site there might be a mix of sample transport: by pneumatic carrier, manual transport or sample transport by gravity through a chute if the laboratory is at a lower level under a production unit.[17]

For cost reasons this step is not always automated and an operator dispatchs the samples and supervises the sample preparation equipment.

In the case of a cabin laboratory the sample is introduced into the system by an operator through a small opening in the cabin wall.

3.1.2 Metallic Sample Preparation

3.1.2.1 Steel Samples Steel samples are generally of the lollipop type (Figure 2). In order to perform a spectral analysis the sample must be prepared. This consists of removal of the pin at the end of the sample, the preparation of the sample surface by grinding and optionally removing a small sample for the separate determination of carbon and sulfur or nitrogen. Two different pieces of equipment have been developed for this task. One of these uses one or two narrow vertical grinding belts and may also include the tools to cut off the pin and punch out small pieces for the ancillary determinations. In this case the sample is a so-called double thickness sample with a spectral part having a thickness of 8–12 mm and a tongue having a thickness of only 4 mm. This tongue is punched in order to obtain small pieces each of about 1 g in weight which are used for the separate determination of carbon, sulfur or nitrogen. The second piece of equipment (Figure 4) uses a large horizontal grinding belt and includes a water cooling unit. Removal of the pin and the punching are performed on separate small units. It is important to take off sufficient material, normally a layer of 0.7–1.3 mm, in order to eliminate possible segregation within the sample and to obtain an even surface for the spectral analysis. The sample is positioned at the entry end of the equipment and from here on the preparation process is fully automatic. The same equipment is used to prepare the control and calibration samples of the emission spectrometer. These samples are introduced from the exit end as only a final grinding is necessary and these samples are then stored close to the spectrometer.

3.1.2.2 Nonferrous Metals Nonferrous metals such as copper alloys, zinc or aluminum are much softer than

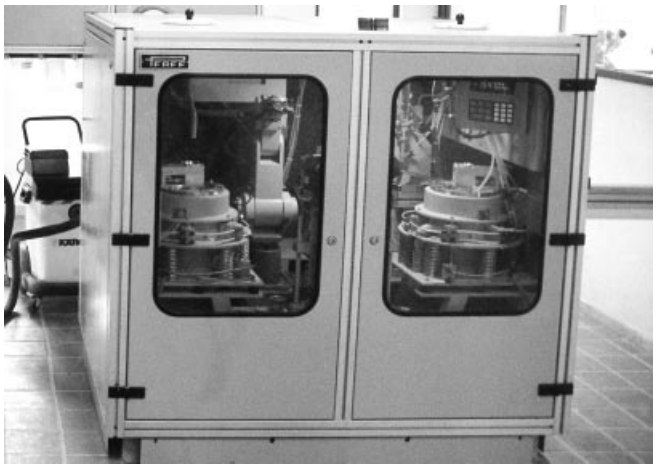

Figure 5 View of an automated slag preparation machine. (Reproduced by permission of Pfaff AQS.)

steel. In addition hard precipitates, for example dendrites, may be part of the sample. To prepare these soft samples and avoid a smearing effect of the precipitates a milling machine is used. Two different tools can be used for a preliminary coarse milling followed by a final fine milling. These pieces of equipment are also fully automatic and adapted to prepare both the control and calibration samples.

3.1.3 Oxide Sample Preparation

3.1.3.1 Pellet Preparation Oxide material is sampled from the process (sinter belt, blast furnace or steelmaking unit) and a small aliquot of about 100 g is sent to the laboratory. The automated preparation units are derived from those used in the cement industry which comprise a swing mill and an automatic press (Figure 5).

For their use in the steel industry an optional jaw crusher performs a coarse crushing and a magnetic separator eliminates any metallic particles included in the sample. Part of the automatic function is a cleaning cycle with compressed air and aspiration of the removed particles. Sometimes the vessel and the rings of the swing mill are cleaned with an organic solvent such as isopropyl alcohol. The final pellet may include a binder or be pressed into a steel ring to prevent partial crumbling during manipulation.

3.1.3.2 Bead Preparation In order to eliminate mineralogical and grain size effects, some users find it necessary to prepare a fused bead, that is, the sample is fused with a borate such as lithium tetraborate in order to obtain a glass in the form of a disk. This fusion is generally performed in a crucible made of a platinum–gold alloy (95–5%) heated to a temperature of about 1150 °C. In automatic systems heating is performed using either induction heating or gas burners. Such a system includes the automation of the weighing operations and of crucible cleaning (Figure 6).[22–24]

Figure 6 View of an automated bead preparation system. (Reproduced by permission of Soled Industries.)

3.1.4 Analysis by Optical Emission Spectrometry

As has already been shown in the section on the history of automation, steel and iron samples are normally analyzed by spark emission spectrometry. The hardware of the spectrometer used differs from a manual spectrometer only in respect of the addition of an automatic sample clamping unit to the spark stand as well as an automatic electrode cleaning device (Figure 7). Other differences are part of the software. A spectrometer control and recalibration strategy must be established and generally control samples are analysed at preset intervals. If the result is outside preset limits, a recalibration procedure is automatically initiated. In addition a so-called type calibration can be used for steel qualities with very narrow specifications. In such a case a type sample of similar composition to the unknown sample is analysed shortly before calibration and the two results are automatically matched. As we have seen the most difficult task to automate is the evaluation of the quality of the sample. Since the use of an image analyzer first developed by Cockerill is a rather expensive approach, alternative methods have been used. The simplest is the calculation of the standard deviation between several spark exposures on the same sample in comparison with preset limits. If the results of two exposures lie within these limits, the average is transmitted automatically to the production monitoring computer. If not, a maximum of four exposures are performed in order to find two consistent values,

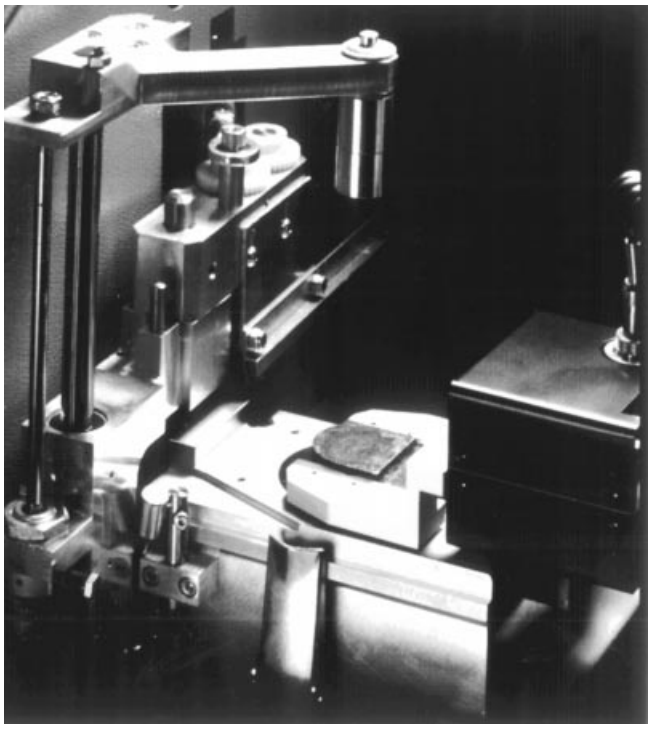

Figure 7 View of a spectrometer sample stand with automatic clamping and cleaning devices. (Reproduced by permission of ARL.)

otherwise the sample is rejected. Another test called SEREPS was developed as part of the cabin concept and combined with the previous method. As these latter tests are performed during the sparking process, samples with cracks or holes containing water introduced during the cooling operation are not detected and the water contaminates the spark stand. Human intervention for cleaning and drying therefore becomes necessary; weighing the sample before analysis reduces this risk.

3.1.5 Analysis by X-ray Fluorescence

Wavelength dispersive XRF is known to be a well established measurement technique giving very reliable results from the parts per million range to 100%. It is the preferred control method in any industry using minerals, and sample changers for this purpose had already been developed in the 1970s. An X-ray spectrometer with a sample changer is a self-contained workstation where only the software has to be adapted to include the station in an automated laboratory.

3.1.6 Ancillary Techniques

Ancillary techniques such as the determination of carbon and sulfur by combustion or nitrogen by carrier gas extraction in many laboratories are not completely automated, despite the fact that the analyzers can be considered as autonomous workstations. The reason for this situation is the fact that automation of central laboratories has reduced the manning levels but not completely dispensed with all the staff. Some examples of such automation, however, do exist.[25–27] Sample handling is performed by adding a human-scale robot to the equipment and bringing a 1-g sample portion to the working area via a small tube using compressed air for the transfer.

3.2 Sample Handling

3.2.1 Sample Transfer between Workstations

Although Cockerill used a large industrial robot for its automation project, conveyor belts or chains combined with manipulators or human-scale robots were widely used in some of the follow-up projects (Figure 8).[11,15,17,19] The main reason for this change was the cost of adapting the existing housing conditions. Transporting a sample with a belt or moving a sample on a slideway may result in inaccurate positioning or a sample may be blocked by going sideways.

As an automated system is only as good as its weakest link, it is of paramount importance to avoid such uncertainties. Whenever possible only ''pick and place'' movements should be used. The cabin approach previously described employed this concept, as did the Cockerill system except for the cooling step on arrival of the hot sample where a conveyor belt is used to transfer the metallic sample to the surfacing machine. Recently two companies have developed systems where a gripping device is mounted on a digitally indexed track. One uses a manipulator and the other uses a revolutionary double track based robotics system simultaneously operating two human-scale robots, one being placed upside down

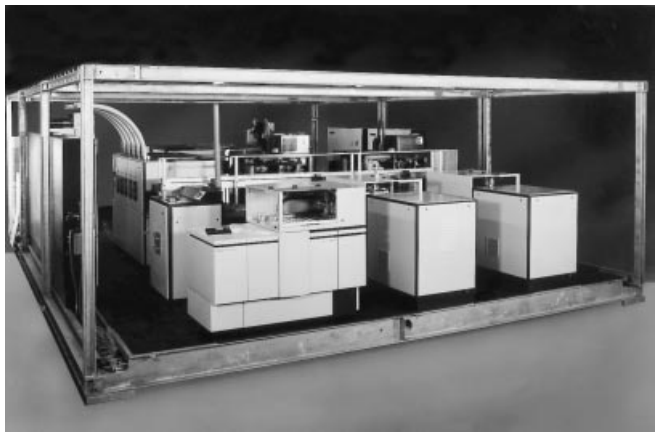

Figure 8 View of a conventional automated laboratory. (Reproduced by permission of Philips Analytical.)

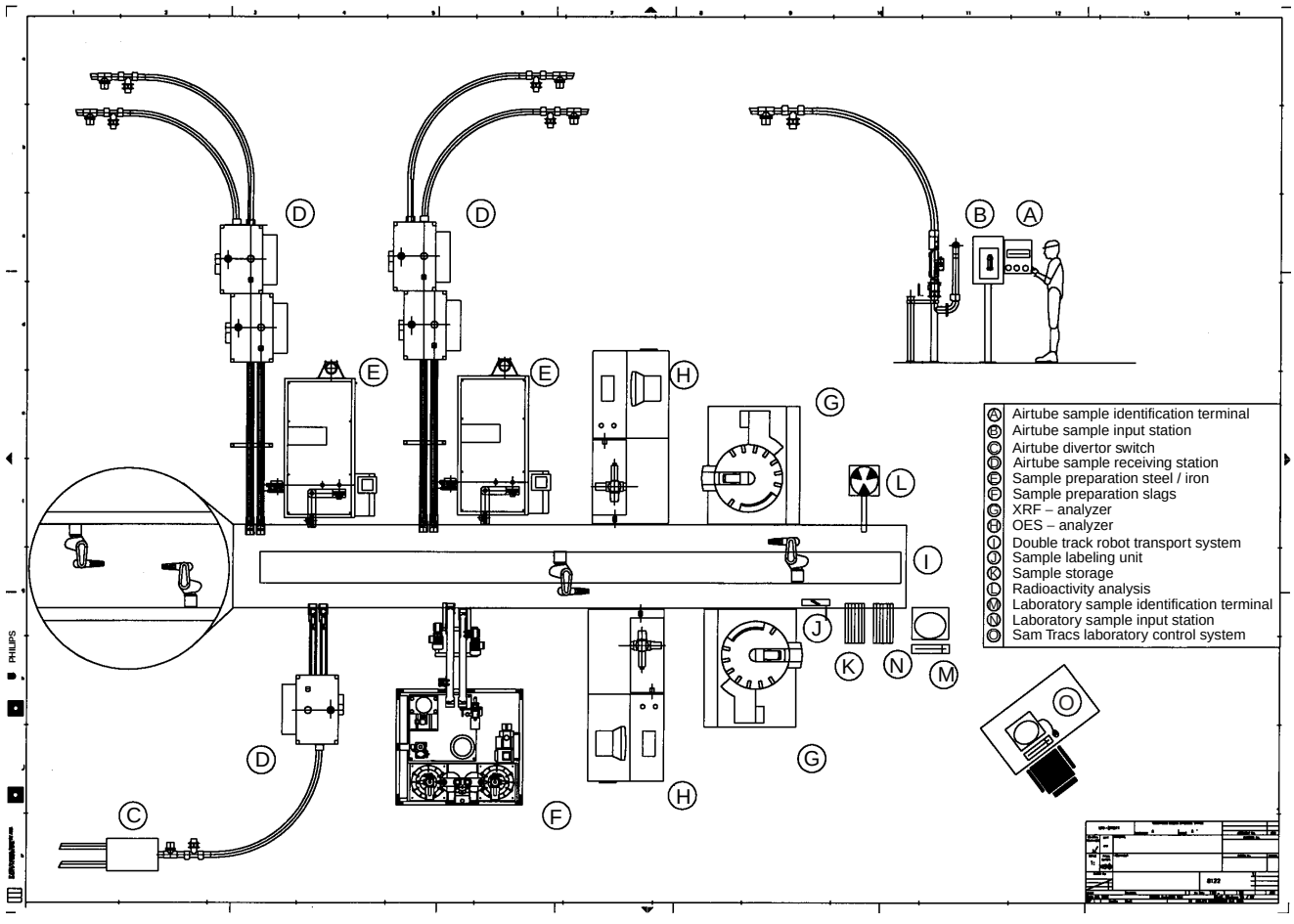

Figure 9 Layout of a double track robotic system. (Reproduced by permission of Philips Analytical.)

(Figure 9).[28] Anticollision software prevents the robots from colliding. The workstations are placed along the two sides of the track. For each transfer the nearest one is activated in order to optimize the speed of operation. This new arrangement is totally ''pick and place'' and has the same access facilities as a conveyor belt or chain.

3.2.2 Sample Identification

There are two aspects to sample identification. The sample must be identified in order for the system to choose the appropriate analytical program and the correct transmission pattern. In addition it may be necessary to physically identify the sample for storage and control purposes.

Sample identification may be done via identification terminals at the despatch stations of the automatic pneumatic system or on introduction into the cabin. For a cabin set-up the task of the operator may be reduced to simply identifying the vessel where the sample originated and the process computer of the steel plant then adds the details of the identification. It may also be deduced by the system from a specific despatch line and time.

If the sample transport is not automated, the operator at the arrival station has to identify the sample for the system and also has the option of putting an identification label on the sample. In automated systems physical identification has always been a problem. The early systems avoided automatic identification by manually labeling the samples after analysis based on their position on a storage rack. The first automatic labeling stations used ink-jet systems which were subject to troubles and malfunction. A much safer technique is the use of paper label dispensers. These labels provide an identification in clear print and by bar code.

4 SOFTWARE

4.1 Sample Tracking Software

The identity of each sample includes a code defining the sample preparation scheme and equipment, instruments used for analysis, the analytical scheme and criteria for result merging and transmission.

A supervising program coordinates the priorities and the different tasks to be performed. Due to historical reasons and the relatively small size of the market, the two principal suppliers of automatic sample transport systems are also suppliers of sample preparation equipment. They deliver with their package a sample tracking system covering the transport scheme, the sample preparation scheme and the sample priority and delivery scheme to the different analytical stations. Every workstation is either equipped with a PLC or a personal computer running the instrument specific software. Results coming from the different instruments are transmitted to a result managing program or to a laboratory information and management system (LIMS) which organizes the result merging, quality assurance and transmission to the final users. As a result of this complex arrangement supervising programs are at least partly customer specific and must be carefully tested before final installation. Most supervising programs include a synoptic chart showing in real-time the location and progress of the different samples. In the case of a cabin system containing only sample preparation equipment and an emission spectrometer the organizational functions are greatly simplified. Modern emission spectrometers are able to determine carbon and nitrogen with sufficient accuracy and precision to satisfy the requirements of an increasing number of process control situations in the steel industry.

4.2 Software for Instrument Calibration and Quality Assurance

The basis for the software is the package developed for manually operated spectrometers. These programs must be augmented by routines and algorithms permitting an automatic release of the results which satisfy predetermined conditions. For emission spectrometry this includes an evaluation of the quality of the burn, as has been explained previously in this text. Emission spectrometry and XRF are comparative methods. The instruments are delivered precalibrated or are calibrated on-site on delivery to suit the requirements of the user. Samples of known composition are analyzed and the calibration curves or algorithms as well as interelement or spectral interferences are stored in the computer asssociated with the instrument.

In order to check and maintain the calibration, control and recalibration samples are analyzed using a predefined strategy. If the results on control samples fall outside the predefined limits, a recalibration procedure is initiated. Generally a two-point calibration, checking the high point and the low point of the curve, is used. The results on these samples are used to calculate correction factors in order to revert to the original calibration. In a robotic system the calibration strategy has to be carefully planned and must take into account both continuous drifts and sudden changes which might occur especially with emission spectrometers. In order to record the quality of the instrument calibration and of the analytical results produced, all data on control and calibration samples are stored and statistically evaluated. It has been shown[29] that for an XRF spectrometer it is possible to implement a statistical process control routine by analyzing a control sample at regular intervals and displaying the results on a chart. Predefined criteria detect deviation trends and automatically initiate a recalibration taking into account not only deviations from the reference values but also trends.

5 PROBLEMS AND BENEFITS

5.1 Maintenance and Troubleshooting

In order to reduce downtimes a preventive maintenance scheme should be put in place. The impact of a failure can be reduced by doubling the number of instruments, maintaining older manual systems in operation or by implementing procedures for rapidly replacing failed components in the system. As the technology has advanced availability times of 99% or better are possible today.

Recently installed systems have a computer link to the manufacturer and distance diagnostics can reduce downtimes or even eliminate them by detecting possible problems before failure occurs.

5.2 Economic Aspects

Justification for automation was at first based on the prospect of reduced manning in the laboratories. Such considerations were often biased by the fact that during the prosperous years staff levels had not been optimized often for social reasons. Today it is recognized that the main justification comprises:

- the reduction of analysis time by locating the analytical facilities as close as possible to the production lines;
- the improvement of the quality of the results by eliminating the human factor;
- the social reform of replacing the monotonous and stressful operator tasks by a smaller number of jobs for higher qualified operators to supervise and maintain these high-tech installations.

6 OUTLOOK

Automation of the analytical equipment used for process control is today a well established procedure. Complete

systems as well as modular components are commercially obtainable and have been proven to be reliable with an availability close to 99%. Careful planning is necessary to define the needs of the process control function and to avoid backlogs of samples. This can be facilitated by the use of an appropriate software package.[30] For complex systems which integrate equipment and software from different manufacturers it is recommended that the services of a specialist company should be used and that a preinstallation test of the system should be included in the planning.

It has been shown recently[31] that the use of laser induced breakdown spectroscopy, where a high intensity laser beam vaporizes the material and excites the vapor to produce a plasma which is analyzed by an emission spectromer, permits the analysis of conductive and nonconductive material. Early versions of the system have been used to monitor pig iron production in blast furnaces using lollipop samples without the need to prepare the surface of the sample and also to determine the slag composition on vitreous samples obtained by controlled cooling of the liquid slag. Figure 10 shows such a system in its robotic form. For such a configuration sample preparation equipment is no longer necessary and equipment can be installed close to the furnace without the need for any specific operators. A furnace operator simply identifies the sample to the system and transfers it to the entry area where it is picked up by the robot.

Research on in situ analysis of liquid steel has up to now found only the occasional application in process control,[16] but much research effort has been expended in the last 25 years and recently initial trials in secondary steelmaking[32] appear to be quite promising.

Figure 10 View of an automated OES spectrometer with laser source. (Reproduced by permission of ARL.)

ACKNOWLEDGMENTS

I would have been unable to write this paper without the help and the updated information provided by my fellow steel chemists A. Daniels, Jean-Claude Hoët, Fernand Hoffert, Ray Jowitt, Patrick Kicq, Masao Saeki, Bernd-Joseph Schlothman and Gotthard Staats and by the manufacturers' representatives Pierre Brugnago, Volker Dreisbach, Colin Gamage, Reinhard Kamphoff and Huub Smit. The careful correction of the text by my friend Rhys Jenkins not only filled certain gaps in my English, but also included a critical assessment of the contents of the paper.

ABBREVIATIONS AND ACRONYMS

ECSC	European Community for Coal and Steel
LIMS	Laboratory Information and Management System
OES	Optical Emission Spectrometry
PLC	Programmable Logic Controller
SEREPS	Self-regulated Prespark
XRF	X-ray Fluorescence

RELATED ARTICLES

Process Instrumental Methods **(Volume 9)**
Sampling and Sample Preparation in Process Analysis

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction • Metal Analysis, Sampling and Sample Preparation in • Nickel Ore and Metals Analysis • X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Atomic Spectroscopy **(Volume 11)**
Laser-induced Breakdown Spectroscopy

X-ray Spectrometry **(Volume 15)**
Sample Preparation for X-ray Fluorescence Analysis • Wavelength-dispersive X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. *Advances in Laboratory Automation Robotics*, Zymark Center, Hopkinton, MA, 1984 and 1985.

2. K. Slickers, *Automatic Atomic Emission Spectroscopy*, Brühlsche Universitätsdruckerei, Giessen, 1993.
3. M. Lesir, 'A Fully Automated Steelplant Laboratory', in *Proceedings of the 38th Chemists' Conference*, British Steel, Middesbrough, 57–62, 1985.
4. A. Daniels, formerly of Cockerill-Sambre SA, personal communication, 1997.
5. J. Petin, A. Wagner, 'Spectrolux, a Different Concept for the Control of Liquid Steel', in *Proceedings of International Conference on Progress of Analytical Chemistry in the Iron and Steel Industry*, Laborlux SA, Luxembourg, 565–574, 1987.
6. G. Johansson, 'Experiences with the Laborlux Automatic Analysis System', in *Proceedings of the 41st Chemists' Conference*, British Steel, Middlesbrough, 115–120, 1988.
7. G. Johansson, J. Petin, *Steel Times*, **215**, 348 (1987).
8. R. Jowitt, B.D. Summerhill, 'Evaluation of the Spectrolux Automated Analysis Cabin', in *Proceedings of the 42nd Chemists' Conference*, British Steel, Middlesbrough, 107–114, 1989.
9. D. Lambie, 'Experiences with On-site Analysis', in *Proceedings of the 43rd Chemists' Conference*, British Steel, Middlesbrough, 17–20, 1990.
10. P.M. Rugg, 'On-plant Evaluation of the Hilger Analytical Ltd. Polyvac Automatic Laboratory', in *Proceedings of the 43rd Chemists' Conference*, British Steel, Middlesbrough, 21–24, 1990.
11. M.A. Bassin, 'Automatic Sample Mainipulation Systems for Optical Emission Spectrometers', in *Proceedings of International Conference on Progress of Analytical Chemistry in the Iron and Steel Industry*, Laborlux SA, Luxembourg, 591–602, 1987.
12. A. Pfeiffer, H.D. Reski, E. Thiemann, 'Die Emissionsspektrometrie im Bühnenlabor eines Stahlwerkes als integrierter Bestandteil eines Robotsystems', *Stahl Eisen*, **111**, 103–107 (1991).
13. Reference lists, Spectro Analytical Instruments, Germany and ARL, Switzerland, private communication, 1998.
14. K.H. Koch, K.D. Ohls, J. Flock, 'First Experience with an Automated System for Production and Quality Control Analysis', in *Proceedings of the 41st Chemists' Conference*, British Steel, Middlesbrough, 79–86, 1988.
15. W. Zimnik, C.-T. Chang, 'Automatisierung eines Spektrallaboratoriums in einem integrierten Hüttenwerk', *Stahl Eisen*, **111**, 71–75 (1991).
16. M. Saeki, 'Recent and Future Trends in OES for Analytical Control of Steel Production in Japan', in *Proceedings of the 42nd Chemists' Conference*, British Steel, Middlesbrough, 8–13, 1989.
17. P. Petesch, G. Böhm, 'Automatisierung eines Stahlwerkslaboratoriums', *Stahl Eisen*, **113**, 77–82 (1993).
18. K.A. Slickers, 'Die spektrometrische Stahlanalytik verläßt das Zentrallabor', *Stahl Eisen*, **113**, 103–104 (1993).
19. W. Marschinke, H.-J. von Wachtendonk, 'The New Automated Laboratory of Thyssen Stahl, Ruhrort Works', in *Proceedings of the Progress in Analytical Chemistry in the Steel and Metals Industry*, Office for Official Publications of the European Communities, Luxembourg, 264–273, 1994.
20. K.H. Koch, J. Flock, 'Rationalisierung der metallurgischem Prozeßanalytik durch Automation und Dezentralisation', *Stahl Eisen*, **116**, 71–74 (1996).
21. R. Sundvall, 'New Concept in the Steel Industry: Coupling an Automatic On-site Laboratory with an Automatic Main Laboratory', in *Proceedings of the 44th Chemists' Conference*, British Steel, Middlesbrough, 59–61, 1992.
22. J. Petin, A. Wagner, 'Automatic Preparation of Fused Beads for XRF Analysis', in *Advances in Laboratory Robotics*, Zymark Center, Hopkinton, MA, 367–377, Vol. 2, 1985.
23. H.A. Smit, A.F.P. van Eenbergen, H. Webster, J. Petin, 'IAX, Automatic Bead Machine for Automated XRF Applications', in *Proceedings of the 49th Chemists' Conference*, British Steel, Middlesbrough, 68–71, 1997.
24. S. Lüngen et al. 'The Automatic Melting Apparatus for Preparing Oxide Materials for the X-ray Analysis', in *Proceedings Analytical Chemistry in the Steel and Metals Industry*, Office for Official Publications of the European Communities, Luxembourg, 381–387, 1998.
25. J.R. Clinch, 'Automation of Combustion Analysis', in *Proceedings of the 41st Chemists' Conference*, British Steel, Middlesbrough, 43–44, 1988.
26. G.J. McManus, 'LTV Reaps the Rewards of Robotics', *Iron Age*, **Nov/Dec** (1992).
27. K.H. Koch, J. Flock, 'Die automatisierte Stickstoffbestimmung im Stahl als Teil der qualitätssichernden Prozeßanalytik', *Stahl Eisen*, **117**, 53–57 (1997).
28. Philips, Press Release, 'Hoogovens Installs Multi Phase Laboratory Automation', 1999.
29. J. Sabbioneda, F. Hoffert, J. van den Bosch, 'Using Robotics and On-line SPC for Analysis Control in Production', *Steel Times*, **221**, 12–14 (1993).
30. J.E. van den Bosch, 'Automated Laboratory Engineering, Times of Trial and Error Are Over', in *Proceedings Analytical Chemistry in the Steel and Metals Industry*, Office for Official Publications of the European Communities, Luxembourg, 628, 1998.
31. P. Eggiman, A. Lauritzen, E. Müller, 'Laser Spark Optical Emission Spectrometry: a Revolutionary Method for the Iron and Steel Industry', in *Proceedings Analytical Chemistry in the Steel and Metals Industry*, Office for Official Publications of the European Communities, Luxembourg, 321–328, 1998.
32. I.R.C. Whiteside, 'Application of In-situ Analysis in Secondary Steelmaking', in *Proceedings Analytical Chemistry in the Steel and Metals Industry*, Office for Official Publications of the European Communities, Luxembourg, 232–238, 1998.

Iron Ore, Sample Preparation and Analysis of

Om P. Bhargava
Burlington, Canada

1 Introduction 2
1.1 Sampling and Sample Preparation 2
1.2 Analytical Methods 2
2 Composition 2
2.1 Principal Ores 3
2.2 Concentration Ranges of Elements of Interest 3
3 Sampling and Sample Preparation 3
4 Physical Tests 3
4.1 Particle Size or Screen Analysis Test Methods 3
4.2 Tumbler Test Method 3
4.3 Crushing Strength Test Method 3
5 Analytical Competence and Safety Aspects 3
6 Elements of Interest and Choice of Methods 4
7 Common Instrumental Methods 4
8 Test Sample Weighing Procedure for Chemical Analysis 4
9 Determination of Hygroscopic Water 4
9.1 Moisture Determination 5
10 Determination of Loss on Ignition and Combined Water 5
10.1 Determination of Combined Water 5
11 Determination of Total Sulfur 5
12 Procedures for Test Sample Dissolution using Acid Dissolution or Alkali Fusion 5
13 Methods for the Determination of Total Iron (Redoximetric Titration) 5
13.1 Fusion/dissolution with Sodium Peroxide in a Zirconium Crucible 6
13.2 Silver Reductor Total Iron Redoximetric Titration Method 6
13.3 Titanium(III) Chloride Total Iron Reduction Method 6
14 Determination of Metallic Iron in Direct Reduced Iron by Bromine–Methanol Method 6
14.1 Dissolution of Metallic Iron 7
14.2 Determination of Metallic Iron by a Redoximetric Procedure 7
14.3 Determination of Metallic Iron by a Complexometric Procedure 7
15 Determination of Silicon by Gravimetry 7
16 Determination of Aluminum by Ethylene Diamine Tetra-acetic Acid-complexometric Titration 7
16.1 Procedure 8
17 Photometric Methods 8
17.1 Determination of Manganese by Periodate 9
17.2 Determination of Phosphorus by Phospho-molybdenum Blue 9
17.3 Determination of Silicon by Silico-molybdenum Blue 10
17.4 Determination of Titanium by Chromotropic Acid 11
17.5 Determination of Vanadium by *N*-benzoyl-phenylhydroxylamine 11
17.6 Determination of Arsenic by Silver Diethyldithiocarbamate–Ephedrine 12
18 Flame Atomic Absorption Spectrometric Methods 13
18.1 Determination of Calcium and Magnesium 13
18.2 Determination of Copper 15
18.3 Determination of Lead and/or Zinc 15
18.4 Determination of Nickel and/or Chromium 17
18.5 Determination of Sodium and Potassium 17
19 Inductively Coupled Plasma/Atomic Emission Spectrometry for Multi Elements 17
19.1 Procedure 17
20 X-ray Emission Spectrometry 18
20.1 Procedure 18
21 Standard/Certified Reference Materials, Quality Control and Quality Assurance 19
22 Statistics, Accuracy and Precision 19
Abbreviations and Acronyms 19
Related Articles 20
References 20

The iron ore industry is an important segment of the world economy. The main mineral oxides are magnetite and hematite. The composition of iron ore plays an important

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

role in both the iron (blast furnace) and steelmaking operations. Sampling is done both at the source and at the destination. Sampling protocols and preparation are described. The determination of iron, phosphorus, silica and manganese by methods such as gravimetric (Si) and volumetric (P and Mn) have been replaced by more rapid photometric or flame atomic absorption spectrometry (FAAS) methods. Instrumental methods such as combustion high-frequency furnace infrared (IR) (for C and S) FAAS and graphite furnace electrothermal methods, X-ray fluorescence (XRF) and inductively coupled plasma (ICP) spectrometric methods are used to determine analytes for routine process control. These methods are illustrated in the article. Rapid methods for chemical analysis and quality control, developed by the author, could also be cost effective.

1 INTRODUCTION

The iron ore industry is an important segment of the world economy. It includes natural lump ores and those ores which are processed to enhance the iron content of the products, which may be concentrates, pellets or direct reduced iron (DRI) for use in the iron and steel industry. In 1989 some 50 countries produced 982 million metric tons of iron ore. The world iron ore trade in 1989 amounted to about 423 million metric tons valued at about 11 billion US dollars. Brazil and Australia share the major portion of this trade. Canada, India and South Africa also significantly contribute to the export of iron ore.

The main mineral oxides are magnetite (Fe_3O_4 with Fe 72.4%), hematite (Fe_2O_3 with Fe 69.9%) and ilmenite ($FeTiO_3$ corresponding to Fe 36.8% and Ti 31.6%). Iron also occurs in the minerals such as carbonate (siderite) and sulfide (pyrite form FeS_2, with Fe 46.6% and S 53.4%) and various and sometimes complex iron silicates. The acceptability and the value of iron ore depend on the ore composition. The main constituent, iron, may vary from 46 to 72%, Si from 0.1 to 12%, Al from 0.1 to 4%. Constituents such as Mn may vary (0.01 to 8%), P from 0.005 to 1.4%, S from 0.001 to 0.9%, V from 0.001 to 0.4%. The composition of the iron ore plays an important role in both iron and steelmaking operations. Some components, from minor to trace elements, are environmentally sensitive. They include Ti, Ca, Mg, Na, K, Ba, V, F, Cu, Pb, Zn, Cr, Ni, Co, Sn, As, Sb, Cd, Bi, Mo, Fe(II), Cl, C, and hygroscopic moisture.

1.1 Sampling and Sample Preparation

Iron ore and iron-bearing feed stock is transported in bulk by ship (special carriers) or rail. The sampling is done both at the source and at the destination. Automated mechanical sampling stations with advanced technology are installed by some establishments, for example at Suldhana Bay (South Africa), South Wales (UK) and in Japan. The sampling techniques must provide a homogenous and representative sample of the consignment. After sample preparation (crushing, pulverizing, grinding, sieving etc.) the sample undergoes further steps for the preparation of an analytical sample according to established practice set out by International Standards Organization (ISO), American Society for Testing Materials (ASTM) and European and Japanese standards bodies.

1.2 Analytical Methods

In commercial analysis the determinations most commonly used are iron, phosphorus, silica and manganese. The older, tedious methods such as gravimetric (Si) and volumetric (P and Mn) have been replaced by more rapid photometric or FAAS methods. Reference and standard methods are used for contractual purposes to determine the value of the iron ore consignments including the characterization and composition of iron ore. The iron content as well as the concentration of the other constituents are important for blast furnace and steelmaking processes, such as basic oxygen furnace or electric furnace operations. These considerations necessitate various sampling protocols as well as method/s of analysis, both classical (standard chemical methods) and instrumental. A distinction must be made here in the choice of the method of analysis. Iron is a major element in the iron ore, concentrate, pellet, etc. (50–90% Fe). For contractual purposes this element is usually analyzed by very precise and accurate noninstrumental methods. Instrumental methods such as combustion high-frequency furnace IR (for C and S), FAAS and graphite furnace electrothermal methods, XRF and inductively coupled plasma/atomic emission spectrometry (ICP/AES) methods are used for determining analytes for routine process control. Gravimetric, volumetric, complexometric, photometric, FAAS, ICP/AES and XRF analysis methods are described. In this article rapid methods for chemical analysis, developed by the author, are illustrated that could also be cost effective to provide analytical quality control as well as reference analysis, worth consideration as standard methods, by organizations such as ISO and ASTM.

2 COMPOSITION

Iron ore encompasses rocks, minerals or aggregates of minerals, natural or processed such as concentrates,

pellets, or DRI, from which iron can be produced commercially.

2.1 Principal Ores

Principal ferriferous minerals occurring in iron ore are: red, brown and specular hematite, marlite and maghemite; magnetite; hydrated iron oxides, including goethite, and limonite; iron carbonates including siderite, chalybite, ankerite and other mixed carbonates; roasted iron pyrites or pyrite cinders; ferrites such as calcium ferrite occurring sometimes in natural ores, but mainly in fluxed pellets and sinters.

2.2 Concentration Ranges of Elements of Interest

Iron ore has diverse composition depending on the mineral characteristics of the ore-bearing materials. The main mineral oxides are magnetite (Fe_3O_4: Fe 72.4%), hematite (Fe_2O_3: Fe 69.9%), ilmenite ($FeTiO_3$: 36.8% Fe and Ti 31.6%). Iron also occurs in the minerals such as carbonate (siderite) and sulfide (pyrite form FeS_2: 46.6% Fe and 53.4% S). Concentration ranges of elements of interest are: Fe: 30–72%; Si: 0.1–10%; Ca: 0.01–10%; Al: 0.1–5%; Ti: 0.01–5%; Mg: 0.01–3%; Mn: 0.01–3%; P: 0.003–2%; S: 0.002–1%; Na and K: each 0.002–1%; V: 0.005–0.5%; Cu, Cr and Ni: each 0.003–0.1%; Pb and Zn: each 0.001–0.5%; Sn: 0.001–0.1%; water-soluble Cl: 0.005–0.1%; hygroscopic H_2O: 0.05–6% and loss on ignition: 0.25–10%; combined H_2O: 0.05–10%; Fe(II): 1– 25%.

3 SAMPLING AND SAMPLE PREPARATION

Importance of sampling is well recognized, as an analysis can only be as reliable as the sample on which it is made. Sampling is an important task which involves skill, judgement, reliability and experience. Sampling precedes sample preparation and involves both manual and automated systems for collecting samples (from the wagons as well as from the conveyer belts transferring the shipment from the cargo hold) in a well-designed system of sampling operations. The sampling and sample preparation steps are based on the objectives of the testing; a sampling plan specifies the minimum weights and number of increments required for each step in the procedure. Samples are then collected and a composite sample (subsample) is prepared for physical tests such as tumbler test, crushing strengths and chemical analysis. The subsample is dried, blended, divided, crushed, pulverized and ground, as required by the test method. This is also true for the manual sampling of iron ore, pellets etc.; using a scoop and combining the increments for sample preparation as stated earlier. Test samples for chemical analysis should be dry, weigh about 50 g, and have a maximum particle size to pass a No. 100 (150 μm) sieve.

4 PHYSICAL TESTS

Iron ore pellets, iron ore concentrates and sinter constitute a considerable portion of iron and steelmaking feed for iron and steel production. The physical tests, tumbler test and crushing strength are indicators of abrasion resistance and degradation by impact in the determination of quality problems associated with the production of pellets as well as performance in the blast furnace. Particle size or screen analysis is intended to be used for compliance with the compositional specification for particle size distribution.

4.1 Particle Size or Screen Analysis Test Methods

The sample is passed through a bank of standard sieves by agitation. The selected sieves are nested and a pan is fitted below the bottom sieve. The nested sieves are covered and mechanically agitated for a specified time. The cover and clamp are removed and contents of each sieve are transferred to a tared pan and weighed. The sieve fractions are then calculated.

4.2 Tumbler Test Method

The sample is placed in a tumbler drum (of specified material and design) which is rotated. The tumbled material is then removed and screened to determine the degradation.

4.3 Crushing Strength Test Method

A load is applied to a single pellet at a specified speed of the compressive platen until the pellet is broken. This procedure is repeated on all pellets of the test sample.

5 ANALYTICAL COMPETENCE AND SAFETY ASPECTS

It is expected that the analysis will be carried out in a properly equipped, environmentally clean laboratory, by qualified and trained analysts capable of performing common laboratory practices skillfully and safely. The work place hazardous materials information system should be in place. All chemicals used in the laboratory should have a material safety data sheet available and personnel acquainted with it. The laboratory shall have

in house practice of proper waste disposal (including appropriate devices for emissions control) procedures.

The safety aspects are crucial. Recently ISO has promulgated guidelines[1] for their inclusion in standards and that includes test methods. The guidelines particularly refer to harm, hazard, risk, protective measure and safety, defined as freedom from unacceptable risk. It is desirable to have a short description of safety aspects posted in the laboratory.

6 ELEMENTS OF INTEREST AND CHOICE OF METHODS

The choice of method depends on the chemical properties of the element to be determined. The precise iron content as well as the concentration of the other constituents of the iron ore and other iron-bearing feed stocks are crucial for blast furnace and steelmaking processes such as basic oxygen furnace or electric furnace operations. A distinction must be made here in the choice of the method of analysis. Iron is a major element (50 to 90% Fe) in the iron ore, concentrate, pellet, etc. For contractual purposes this element is usually analyzed by very precise and accurate standard (noninstrumental) methods. Standard methods are used to determine the value of the iron ore consignments including the characterization of iron ore, which of course includes the composition of the iron ore consignment. The reader is referred to articles by Lundell, Hoffman and Bright[2] and somewhat more recent by Harrison,[3] and standard methods for analysis of iron ore, from ISO, or national organizations such as ASTM, UK, Europe, Japan, Australia, etc.

Many standards methods have superseded the traditional methods, such as gravimetric and volumetric (used for Si, P and Mn). Faster photometric methods with improved precision and accuracy are now in use and amenable to automation. The latter are generally used for minor elements e.g. P, Mn, Si, Ti and V. Prior to the advent of FAAS the wet chemical methods for determination of Na and K, were very tedious with questionable accuracy. Elements such as Pb and Zn are now preferably determined by FAAS methods, for better sensitivity and relative freedom from interference. FAAS methods are also used to determine elements, such as Ca, Mg, Na, K, P, Cu, Ni, Cr, etc.

7 COMMON INSTRUMENTAL METHODS

The application of photometry and FAAS for analysis of a single element and multielements have already been alluded to. In industry, laboratory systems are common for sequential and simultaneous multi-elemental analysis. ICP/AES spectrometric methods are also being used for determining analytes for routine process control. However the samples need to be solubilized for both FAAS and ICP/AES. This presents constraints as acid dissolution is often not complete. The undissolved residue has to be separated, fused with a flux and combined with the main solution prior to aspiration and nebulization. The salt concentration poses a problem with salt build up and leads to interference. Alternatively, the sample is fused with borax (sodium tetraborate) in a platinum crucible and the melt is dissolved in acid. For elements such as Na and K, hydrofluoric acid treatment is essential to break down refractory silicates using Teflon™ or platinum ware for sample dissolution.

Carbon and sulfur are normally determined instrumentally by high-frequency combustion IR methods. The accuracy is similar to that of gravimetric (carbon) and iodometric (sulfur) methods (see section 11).

XRF methods are used in routine control analysis. The sample is briquetted with a binder or preferably fused. Common fluxes used in the preparation are borax and lithium carbonate. The fused hot melt is poured into a heated platinum dish to form a bead for subsequent measurement on an X-ray spectrometer equipped with suitable analyzing crystals and detectors.

Apparatus, laboratory ware, deionized (distilled) water and reagent purity should meet the specifications, outlined in national and international standards.

8 TEST SAMPLE WEIGHING PROCEDURE FOR CHEMICAL ANALYSIS

Iron ore is hygroscopic. The analysis is made on a dried sample basis and corrected if required on an as-received basis. The following procedure is recommended in all test methods where a weighed sample is used for chemical (solution) analysis. Transfer the analytical test sample to a weighing bottle. After drying in an electric oven at 100–105 °C for 1 h, cap the bottle and cool in a desiccator for 30 min. Momentarily release the cap to equalize pressure and weigh the capped bottle to the nearest 0.0002 g. Transfer the sample into a beaker or crucible for acid/fusion decomposition and reweigh the capped bottle to the nearest 0.0002 g. The difference of the two weights is the weight of the sample used in the test method.

9 DETERMINATION OF HYGROSCOPIC WATER

The importance of moisture content is twofold. Firstly the value is used in converting the iron content of the dried

sample on as-received basis of the cargo or consignment. Secondly samples may readily absorb moisture even waiting for analysis. It is customary to determine moisture in the sample preparation area in large drying ovens. The physical tests are also done on dry samples.

9.1 Moisture Determination

Transfer 2 g (accurately weighed to nearest 0.001 g) sample of the air-dried sample to a tared weighing bottle with stopper. Dry at 100–105 °C in an electric oven for 3–4 h. Stopper the bottle and transfer to a desiccator and cool for 30 min. Upstopper the bottle and weigh as before. Calculate percent moisture : loss in weight (g) divided by the original sample weight (g) and multiplied by 100.

10 DETERMINATION OF LOSS ON IGNITION AND COMBINED WATER

Normally loss in weight on ignition is due to combined water, organic matter and carbon dioxide from any carbonate present. With materials of high iron content (concentrates, prereduced iron) an overall weight gain is expected. It is prudent to determine both combined water and carbon dioxide separately.

10.1 Determination of Combined Water

Gravimetric and/or Karl Fischer titrimetric methods can be used. The combined water is determined in the test sample, usually a 2 g sample weight, by heating in a stream of dry air at a temperature of 1000 °C, in a combustion train with U-tubes for the preabsorption of sulfur on a silver spiral, followed by a pre-weighed U-tube containing magnesium perchlorate (gravimetric) or Karl Fischer reagent (titrimetrically), for determining the evolved water.

Carbon dioxide is determined by transferring the sample to a round-bottomed flask attached to a purification train to capture the sulfur gases and an absorption system containing a pre-weighed bulb with ascarite and magnesium perchlorate. Dilute hydrochloric acid is added via a tap funnel and heated to boiling. After absorption of the evolved carbon dioxide is complete the bulb is re-weighed for difference and percent CO_2 calculated.

11 DETERMINATION OF TOTAL SULFUR

Total sulfur can be determined by a few methods including the rarely used tedious barium sulfate gravimetric method. The most commonly used instrumental method is the combustion–iodate titrimetric method. In brief, the sulfur is converted to SO_2 by combustion in a stream of oxygen (in an induction furnace set-up using iron, tin and copper as accelerator) absorbed in an acidified starch–iodide solution and titrated continuously during the evolution with a standard solution of potassium iodate. The potassium iodate solution is standardized against a similar type of iron ore of known sulfur content, since the percentage sulfur (yield) evolved as SO_2, varies with different iron ore bodies.

12 PROCEDURES FOR TEST SAMPLE DISSOLUTION USING ACID DISSOLUTION OR ALKALI FUSION

The sample has to be dissolved for all methods requiring solution for traditional wet chemical procedures or by physico-chemical methods such as FAAS, direct current ICP/AES or by auto-analyzer procedures designed for such solutions. Sample decomposition procedure depends on the analyte of interest. Acid such as hydrochloric, singly or in combination with other acid/s, is used for sample dissolution. Acid dissolution invariably leaves an insoluble residue. The latter has to be separated by filtration, the residue ignited, silica removed by HF treatment (if required), and fused with a suitable flux and redissolved in the melt to obtain complete sample solution. If a fusion must be made, it might as well be done at the beginning to save time, provided the procedure can tolerate the flux used for fusion. Several test methods, adopted as ISO and ASTM Standards, and schemes,[4,5] have been developed by the author, which among others, include sodium peroxide sintering/fusion in a zirconium crucible to yield rapid and complete sample solution amenable for determining several analytes photometrically and by other physicochemical techniques. In an earlier study, a rapid method for decomposing iron ore in a vitreous carbon crucible by fusing with sodium peroxide was reported,[6] but this crucible is both heavily attacked and lasts only a few fusions, which makes it relatively expensive. A zirconium crucible is much more resistant to attack by sodium peroxide and lasts for dozens of fusions.

13 METHODS FOR THE DETERMINATION OF TOTAL IRON (REDOXIMETRIC TITRATION)

Iron is the most important and valuable element in iron ore. Several methods exist for its determination. The method involving hydrogen sulfide reduction and dichromate oxidation is very tedious. Acid dissolution of iron

ore is incomplete as well as tedious. Bhargava[6] described a rapid sample dissolution method for iron ore matrix and redoximetric titration of iron by conventional stannous chloride–mercuric chloride reduction and titration of ferrous iron with dichromate. Mercuric chloride has to be added to eliminate the interference of excess stannous chloride needed for the reduction of ferric iron. The ISO had been seeking a mercury pollution free method. Metallic silver (as powder) or generated by the reduction of silver nitrate solution with zinc rods, reduces ferric iron to the ferrous state at the acidity used in the dichromate redoximetric titration. An added feature of the silver reductor column is the ease of regeneration of silver after the latter has been exhausted, for repeated further reduction. Thus the silver reductor[7] method was developed to eliminate mercury pollution. The unique characteristic of using reduction by silver is the validation of its freedom from vanadium interference as is the case with other redoximetric methods for iron determination. The method is rapid, yields results with excellent precision and accuracy, avoiding the tedious manipulations of the acid dissolution procedure. The following procedure[7] uses fusion/dissolution with sodium peroxide in a zirconium crucible, providing complete sample solution and is applicable to the determination of several elements in iron ore.

13.1 Fusion/dissolution with Sodium Peroxide in a Zirconium Crucible

Accurately weigh sample (0.3 g), following the test sample weighing procedure in section 8, into a dry 50-mL zirconium crucible. Mix with 2 g of powdered sodium peroxide and fuse over a M'eker burner (low heat), swirling the crucible until the melt is cherry red and clear. Remove the crucible from heat and swirl until the melt solidifies on the crucible wall. Cool the crucible somewhat in air and place in a dry 250-mL beaker. Add 10 mL of water to the crucible and cover the beaker with a watch glass. After effervescence ceases, pour the contents of the crucible into the beaker and wash with 15–20 mL of water. Add 10 mL of concentrated HCl via the crucible as well as water rinsings into the beaker. Boil the solution for 3–4 min. Wash the watch glass and the walls of the beaker with water and continue boiling for about 30 s. Cool the solution and dilute to about 60 mL. Reserve this solution for the reduction of ferric iron via silver reductor and/or for other determinations.

13.2 Silver Reductor Total Iron Redoximetric Titration Method[7]

Prepare the silver reductor by placing silver powder 250 μm (125 g), over a lightly packed glass wool in a column 2 cm in diameter and 15 cm length fitted with a stop-cock and a 100-mL reservoir cup. Activate silver in the reductor by passing about 100 mL of 1 N HCl through the column at a rate of 35 mL min^{-1}. Place a 400-mL beaker containing 25 mL of mixed acids (prepared by carefully adding 150 mL of sulfuric and 150 mL of phosphoric acid to 600 mL of water, cooling and diluting to one litre with water) for collecting the eluate. Pass the reserved solution (section 13.1), through the reductor at a rate of 35 mL min^{-1} (until there is about 1 mL of acid left above the silver). Rinse the beaker 2–3 times with 1 N HCl and pass through the reservoir and the column as before. Add 5–6 drops of 0.3% sodium diphenylaminesulfonate indicator solution. Titrate immediately with standard potassium dichromate solution to a permanent purple end-point. Calculate the iron content by multiplying the titre (mL) by 5.585 and the normality of potassium dichromate, and divided by the sample weight in g. The silver reductor method has been adopted by ISO (ISO 9508)[8] and ASTM (E 1081).[9]

13.3 Titanium(III) Chloride Total Iron Reduction Method

Another method to eliminate mercury pollution involves the incorporation of titanium(III) chloride; however vanadium still interferes and has to be removed. This method also provides the choice of acid decomposition or decomposition by fusion. In the reduction step, most of the ferric iron is reduced in hot 6 M HCl medium with stannous chloride followed by a slight excess of titanium(III) solution. The excess titanium(III) is oxidized in the hot solution with perchloric acid. The solution is rapidly cooled and the reduced iron is titrated with a standard dichromate solution using sodium diphenylaminesulfonate as indicator. The titanium(III) chloride reduction method has also been adopted by ISO (ISO 9507)[10] and ASTM (E 1028).[11]

14 DETERMINATION OF METALLIC IRON IN DIRECT REDUCED IRON BY BROMINE–METHANOL METHOD

The term ''metallic iron'' means those forms of iron that are not bonded to oxygen nor present as pyrite. The method is particularly significant for metallic iron contents between 15% and 95% in reduced iron ores (DRI). The metallic iron is dissolved by treatment with bromine–methanol solution. The insoluble residue is separated by filtration. In the redoximetric method the iron is reduced to ferrous iron by stannous chloride and subsequently titrated with potassium dichromate solution. An alternative rapid complexometric titration method, developed by the author (unpublished

internal report), for routine control, is described in section 14.3.

14.1 Dissolution of Metallic Iron

Reflux test sample, accurately weighed (0.5 g) at room temperature in a dry 100-mL conical flask, with 50 mL of bromine–methanol solution (5 mL of bromine in 50 mL of methanol) using a magnetic stirrer, for 30 min. Filter immediately through a glass fiber filter under suction and collect the filtrate in a suction filter flask. Wash the filter 5–6 times with methanol (5 mL at a time) until the washings are colorless (free from bromine), then wash a further 3 times with water, to remove methanol. Transfer the solution to a 600-mL beaker, for the redoximetric determination (section 14.2) or to a 200-mL volumetric flask for the rapid complexometric titration method (section 14.3).

14.2 Determination of Metallic Iron by a Redoximetric Procedure

Add 20 mL of diluted (1 + 1) sulfuric acid to the filtrate, heat the solution cautiously until fumes of sulfuric acid appear, then continue heating to strong fuming. Remove from the hotplate, cool and add 10 mL of hydrogen peroxide (30% v/v). Evaporate until sulfuric acid fumes again evolve. Cool and repeat the oxidation with hydrogen peroxide to eliminate bromides and methanol. Add some water, rinsing the walls of the beaker and heat again to fuming. Cool, carefully add about 140 mL of water and 25 mL of concentrated HCl and boil until all salts are redissolved. Carry out the redoximetic determination[6] using tin(II) chloride, mercury(II) chloride, diphenylaminesulfonate indicator and potassium dichromate. Calculate the iron content by multiplying the titre (mL) by 5.585 and the normality of potassium dichromate, and dividing by the sample weight (g).

14.3 Determination of Metallic Iron by a Complexometric Procedure

Dilute the filtrate in the 200-mL volumetric flask to volume with water. Mix. Pipette 10 mL aliquot into a 400-mL beaker. Add about 50 mL of water and 0.5 g of ammonium persulfate. Boil for a few minutes to expel bromine. Dilute to 150 mL with warm water. Add about 0.1 g of salicylic acid. Titrate the warm solution (at about 40 °C) with 0.05 M ethylene diamine tetra-acetic acid (EDTA). The change at the equivalence point is from light reddish brown to distinct lemon yellow. Calculate the percentage of metallic iron by using the equivalence factor: 1 mL of 0.05 M EDTA is equivalent to 2.7925 mg Fe.

15 DETERMINATION OF SILICON BY GRAVIMETRY

The gravimetric method involves dehydration of silica. The sample is fused with sodium peroxide in a zirconium crucible. The melt is leached with water and dissolved in hydrochloric acid. Silica is separated by double dehydration with perchloric acid. The two precipitates after dehydration are combined, ignited in a weighed platinum crucible and reweighed after the ignition step. The difference in weight gives impure silica. Hydrofluoric acid and a little sulfuric acid are added to remove silicon as volatile silicon tetrafluoride. The ignition step is again repeated and the crucible reweighed. The difference in weight of the last two weighings gives pure silica. A blank determination is carried out simultaneously using the same amount of reagents and the same operations as described. Percent silica is calculated by subtracting the weight of blank from pure silica weight and dividing by sample weight, multiplying by 100. The gravimetric method is rarely used in routine analytical control. The double dehydration procedure is basically used for reference analysis and establishing silica content in reference standard materials. The preferred methods are photometric or XRF, which will be described in later sections.

16 DETERMINATION OF ALUMINUM BY ETHYLENE DIAMINE TETRA-ACETIC ACID-COMPLEXOMETRIC TITRATION

Some background information for the development of this technique is relevant. Older methods for determining aluminum such as the gravimetric aluminum phosphate method are both tedious and somewhat unreliable. Photometric methods, manual[12] and for auto-analyzer system[4,12] for aluminum determination have been reported. The sample is fused in a nickel crucible with sodium peroxide and solubilized with dilute HCl. After dilution, the photometry is carried out by treating an aliquot of the solution with ascorbic acid to reduce the ferric iron. The color complex of aluminum is developed with Chromazurol S in a solution buffered with sodium acetate at pH 5.3. Both synthetic and standard reference materials (SRMs) for iron ore can be used for calibration and subsequent calculation of result. However problems arose because of the unavailability of a consistent quality of Chromazurol S reagent. Earlier the author developed a direct method for the complexometric determination of aluminum in zinc alloy,[13] which became the basis of devising an alternative method for aluminum by EDTA-complexometric titration. This method[14] proved

suitable for reference analysis. The method was applied to a wide variety of iron ores with diverse composition giving excellent precision and accuracy. The results of protocol of international round robin tests and statistical evaluation of the developed procedure led to its adoption as an ISO standard (ISO 6830)[15] and by the ASTM (E 738).[16] In brief, the sample is fused in a zirconium crucible with a mixed flux of sodium carbonate and sodium peroxide. The fused mass is completely soluble in HCl. The R_2O_3 oxides are then precipitated with ammonia and redissolved in HCl. Elements such as iron, titanium and zirconium are separated from aluminum by solvent extraction with cupferron and chloroform. After removal of traces of organic matter from the aqueous phase, the solution is titrated with an excess of EDTA, which is back-titrated with standard zinc solution using xylenol orange as a metallochromic indicator. Addition of ammonium fluoride then releases equivalent aluminum EDTA (noteworthy), which is then titrated with zinc solution. Essential details (illustration of solvent extraction separation coupled with complexometric titration) as described in the procedure[14] are important.

16.1 Procedure

Transfer 0.50 g of sodium carbonate to a dry zirconium crucible. Add 0.10–0.30 g iron ore sample (containing 0.001–0.005 g of aluminum), weighed using the weighing procedure in section 8 to the crucible, followed by 2 g of sodium peroxide. Mix the contents with a dry stainless steel spatula. Follow the fusion and dissolution procedure (section 13.1). Bring the reserved solution to the boil. Add diluted (1 + 1) ammonia solution dropwise to raise the pH to 5. Boil for 1 min and immediately filter through a whatman No. 41 or equivalent filter paper, containing some paper pulp. Reserve the beaker. Wash the paper five times with hot ammonium chloride (1 g ammonium chloride per 100 mL water containing two drops of concentrated ammonia) wash solution. Discard the filtrate. Place the reserved beaker under the filter funnel. Dissolve the hydroxide precipitate by adding 5 mL of hot HCl dropwise. Wash the filter with nearly boiling diluted (1 + 5) HCl until the washings are free from iron. Adjust the volume of the filtrate to 50 mL with diluted (1 + 5) HCl. Cool to about 10 °C.

Transfer the solution to a cold (10 °C) 250-mL separating funnel. Use 25 mL of cold water (10 °C) for rinsing the beaker into the separating funnel. Add 20 mL of cold freshly prepared and filtered cupferron (6% in cold water at 10 °C) solution. Mix slightly. Add 20 mL of chloroform. Shake vigorously for 1 min. Let the layers separate. Draw off the lower organic layer. Add 5 mL of chloroform to the separating funnel to displace the cupferronates on the surface of the aqueous layer. Draw off the organic layer.

Table 1 Extraction guide, sample weight, volume (mL) of cupferron and chloroform

Sample (g)	Cupferron (mL)	Chloroform (mL)	Cupferron (mL)	Chloroform (mL)
0.3	20	20	15	20
0.2	15	20	–	–
0.1	10	20	–	–

Carry out further extraction/s with appropriate volumes of cupferron and chloroform (Table 1). Note that for 0.3 g sample weight an additional extraction with cupferron and chloroform is required as specified in Table 1.

Finally add two successive 20 mL portions of chloroform to the aqueous phase, shaking vigorously for 1 min. Let settle and separate. Draw off the organic layer. Wipe the stem of the separating funnel with a filter paper wick. Drain the aqueous phase into a 250-mL beaker, rinsing with 5 mL of diluted (1 + 5) HCl. Boil for a few minutes. Remove from the heat. Add 5 mL of nitric acid and 10 mL of perchloric acid. Cover with a ribbed cover glass. Evaporate nearly to dryness. Remove from heat and add 10 mL of diluted (1 + 1) HCl. Heat to dissolve the salts and then add 50 mL of water and bring to the boil. Filter through a whatman No. 41 paper and wash several times with hot water. Cool. Add an excess of 0.01 M EDTA (25 mL is sufficient) by pipette. Adjust the pH to 4 (using a pH meter) by dropwise addition first of 10% sodium hydroxide solution till the pH reaches 2.5 and then 1% sodium hydroxide solution. Dilute to 100 mL with water. Cover the beaker and bring to the boil. Keep boiling gently for 10 min to permit complete metal-complex formation. Add 15 mL of acetate buffer and seven drops of xylenol orange (0.1 g, made into a paste and diluted with water to 100 mL and stored in an amber-colored bottle) metallochromic indicator. Stir with the magnetic stirrer and titrate with standard 0.01 M zinc (prepared by dissolving 0.6538 g high purity zinc metal in nitric acid, adding about 500 mL of water, pH adjusted to 4, diluted to 1000 mL in a volumetric flask and mixed) solution. The color changes to a persistent (30 s) pink end-point. Add 10 mL of 10% ammonium fluoride solution (prepared freshly each day in a polyethylene beaker). Boil gently for 10 min, cool and titrate the liberated EDTA with standard 0.01 M zinc solution as before. Record the titre (mL). Percent Al is calculated by multiplying titre mL by 0.02698 and dividing by sample weight in g.

17 PHOTOMETRIC METHODS

The technique is basically based on Beer–Lambert's law where the absorbance of the analyte color complex of the

species formed, is proportional to analyte concentration. However in practice this is not as straightforward in all circumstances. Elaborate separation is sometimes necessary to isolate the analyte species prior to color formation. In this section recently developed photometric methods for As, Mn, P, Si, Ti and V will be described. Many of these methods have been adopted by ASTM and ISO as standards.

17.1 Determination of Manganese by Periodate

The range of the method is from 0.01 to 5.0%. The sample is decomposed by digestion with hydrochloric and nitric acids followed by fuming with perchloric acid to eliminate chlorides. The insoluble residue is removed by filtration, ignited and fused with sodium carbonate and the melt dissolved in the filtrate. The manganese is oxidized to permanganic acid with potassium periodate. After cooling, absorbance is measured at 545 nm. Calibrations are made using manganese standard solution to cover the range with treatment for photometry similar to the sample. For future development the author suggests that the sample could be decomposed by sodium peroxide fusion and the water-leached melt dissolved in perchloric acid; then proceed with periodate oxidation and photometry. This should be much faster than the tedious acid dissolution with added fusion step needed to recover the retained analyte in the insoluble residue.

17.2 Determination of Phosphorus by Phospho-molybdenum Blue

The molybdate/magnesia method of Lundell, Hoffman and Bright[2] has been the traditional method for reference analysis for determining phosphorus gravimetrically as magnesium pyrophosphate in minerals and iron ores. The method is quite tedious and cumbersome and suffers from titanium interference. The phosphovanadomolybdate photometric method also involves further extraction with methyl isobutyl ketone (MIBK) making it quite tedious and lengthy. Earlier the author developed a scheme[5] for universal dissolution of materials encountered in the steel industry and its application for the photometric determination of arsenic, phosphorus, titanium and vanadium. This will be described in section 17. The scheme uses fusion with sodium peroxide in a zirconium crucible for the above mentioned elements except for phosphorus. It was thought that zirconium may bind phosphorus as insoluble zirconium phosphate. Hence to avoid zirconium interference, a separate fusion of the sample in a vitreous carbon crucible with sodium peroxide and sodium carbonate was made. The melt was dissolved in perchloric acid, the medium used for developing the final molybdenum blue color for the photometric determination of phosphorus. In our investigations perchloric acid concentration was found to be critical and the usual practice of adding bismuth was found unsuccessful. Further, the use of perchloric acid is becoming rarer, particularly in North America, owing to its hazardous nature and strict safety control required. In a later study it was established that fusion with sodium peroxide in a zirconium crucible and dissolution of the melt in hydrochloric acid is amenable to the development of phosphomolybdenum blue color by addition of molybdate and hydrazine. However the concentration of hydrochloric acid was found important and needed investigation. The following developed method[17] was established and has been adapted as standard methods: (ISO 4687)[18] and (ASTM E 1070).[19]

17.2.1 Procedure

Accurately weigh sample (Table 2) following the test sample weighing procedure in section 8, into a dry 50-mL zirconium crucible, mix with 2 g of powdered sodium peroxide and fuse over a M'eker burner following the fusion dissolution procedure (section 13.1). Cool the crucible somewhat in air and place in a dry 250-mL beaker. Add 10 mL of water to the crucible and cover the beaker with a watch glass. After effervescence ceases, pour the contents of the crucible into the beaker and wash with 15–20 mL of water. Add 15 mL 12N HCl via the crucible as well as water rinsings into the beaker. Boil the solution for 2 min. Wash the watch glass and the walls of the beaker with water and continue boiling for about 30 s. Cool and transfer into a 50-mL volumetric flask, dilute the contents with water and mix.

17.2.1.1 Photometry Transfer (Table 2) an appropriate aliquot and the fusion blank (prepared by dissolving 4 g sodium peroxide in 40 mL of water, adding 30 mL 12N HCl, boiling for 2 min, cooling and diluting to 100 mL in a volumetric flask) to a 150-mL beaker. Add 15 mL of sodium sulfite ($100\,g\,L^{-1}$) solution, mix and bring to the boil. Add 20 mL of molybdate–hydrazine sulfite (section 17.2.1.2) solution, mix and bring to the boil then heat for 10 min in a boiling water-bath. Cool. Transfer to a 50-mL volumetric flask, dilute to the mark with water and mix. Within the next 2 h measure the absorbance of the solution and the calibration standards (section 17.2.1.4) at 725 nm with water as reference. Subtract the blank to obtain the net absorbance,

Table 2 Phosphorus photometric: sample weight, aliquot and fusion blank additions to aliquot

P content (%)	Weight of sample (g)	Sample aliquot (mL)	Fusion blank (mL)
0.005–0.15	0.3	10	none
0.15–0.50	0.1	10	none
0.50–1.0	0.1	5	5

and read off μg of phosphorus from the calibration graph. Calculation: percent P = μg of phosphorus × 0.005/[sample aliquot (mL) × sample weight (g)].

17.2.1.2 Ammonium Molybdate–Hydrazine Solution To 500 mL of water, add cautiously and slowly 300 mL sulfuric acid and cool. Add 20 g of ammonium heptamolybdate tetrahydrate $[(NH_4)_6Mo_7O_{24} \cdot 4H_2O]$. Stir to dissolve and dilute to one litre with water. Add 50 mL of this solution to 100 mL of water, followed by 20 mL of hydrazine sulfate (1.5 g L^{-1}) solution, and dilute to 200 mL.

17.2.1.3 Standard Phosphorus Solution Dry anhydrous disodium phosphate (Na_2HPO_4) at 105 °C, cool in a desiccator, then dissolve 0.2292 g in 200 mL of water. Transfer and dilute to volume in a 1-L volumetric flask with water and mix. This is the standard "A" 50 μg mL^{-1} phosphorus solution. Transfer 10.00 mL of this solution into a 50-mL volumetric flask, dilute to mark with water and mix. This solution "B" is the 10 μg mL^{-1} phosphorus solution for calibration.

17.2.1.4 Preparation of Calibration Curve Into a series of five 150-mL beakers transfer 10.00 mL portions of fusion blank solution and then 0, 0.50 and 2.50 mL of "B" and 1.00 and 2.00 of standard "A" 50 μg mL^{-1} phosphorus solution, corresponding to 0, 5, 25, 50 and 100 μg of phosphorus respectively. To each beaker add 15 mL of sodium sulfite solution, and continue from this point as described in photometry (section 17.2.1.1), for the test solution. Plot the net absorbance [obtained by subtracting the blank (0 μg P) absorbance] against μg of phosphorus in the colored solution.

17.3 Determination of Silicon by Silico-molybdenum Blue

Silicon may be determined by measuring the absorbance of the yellow or blue color complex formed by silicon with molybdate ions, the sensitivity being higher for the reduced molybdenum blue complex. Sanders and Cramer[20] studied various reducing agents such as tin(II) chloride, oxalic acid, iron(II) sulfate, sodium sulfite and 1-amino-2-naphthol-4-sulfonic acid. With iron(II), there is neither reduction of free molybdate nor interference from phosphorus or arsenic as there is with tin(II) chloride and it is not temperature-sensitive as is the case with sulfite. Finally, since iron is present in the iron samples anyway, it may as well be used as the reducing agent. In the method[21] described, the sample is fused with sodium peroxide in a zirconium crucible and the melt is dissolved in HCl. An aliquot of the solution is treated with molybdate and the silicomolybdate complex formed, is reduced to molybdenum with iron(II). Fluoride is added to provide a redox buffering system. The method is eminently useful for automation and applied to auto-analyzers[21] for universal photometric determination of silicon in multi-matrix such as iron ores, sinters, slags, iron and steel. Results are comparable to those obtained by reference methods as illustrated for various SRMs of iron ores and others of diverse composition. Only the manual photometric method is described in the following procedure.

17.3.1 Procedure

Accurately weigh a 0.2000 g sample, following the test sample weighing procedure in section 8, into a dry 50-mL zirconium crucible, mix with 2 g of powdered sodium peroxide and fuse over a M'eker burner following the fusion dissolution procedure (section 13.1). Add 10 mL of water to the crucible and cover the beaker with a watch glass. After effervescence ceases, pour the contents of the crucible into the beaker and wash with 15–20 mL of water. Add 30 mL of diluted (1 + 1) HCl via the crucible as well as water rinsings into the beaker. Boil the solution for 2 min. Wash the watch glass and the walls of the beaker with water and continue boiling for about 30 s and cool. Immediately transfer the solution to a 1-L volumetric flask and add about 500 mL of water and 40 mL of 8% v/v sulfuric acid with swirling and then dilute the contents with water and mix. Carry out a procedure blank exactly as described but omitting the sample.

17.3.1.1 Photometry Transfer 10.00 mL aliquot into a 50-mL volumetric flask containing 5 mL of 0.2% iron solution (prepared by dissolving 1.00 g of pure iron in 80 mL of 8% v/v sulfuric acid, adding 10 mL of 3% freshly prepared ammonium persulfate and boiling for 2 min, cooling and diluting to volume with water in a 500-mL volumetric flask). Add 10 mL of freshly prepared ammonium molybdate (1.6% w/v stored in a polyethylene bottle) solution and mix. After 3 min add 10 mL of sodium fluoride (2.4% w/v aqueous solution, stored in a polyethylene bottle). Mix well. Measure absorbance in a 10-mm cuvette with water as reference, at 665 nm. Subtract the procedure blank absorbance to obtain net absorbance to read off the percent Si in the sample.

17.3.1.2 Calibration Prepare calibration solutions using appropriate National Bureau of Standards (NBS), National Institute of Standards and Technology (NIST), British Chemical Standards (BCS) or local standard reference iron ore samples, and treat exactly as described under the procedure and photometry. Plot net absorbance against percent Si to obtain the calibration graph.

17.4 Determination of Titanium by Chromotropic Acid

Both the ISO method (ISO 4691)[22] and ASTM (E 878)[23] for determining titanium photometrically with diantipyrylmethane, use an acid dissolution followed by treatment and fusion of the insoluble residue. This makes the method unnecessarily time-consuming. The titanium chromotropic acid photometric method[5] coupled with universal technique for rapid dissolution, obviates this time-consuming step and yields rapid complete decomposition in a single step. In this study it was also found that pH control before formation of the titanium complex with chromotropic acid is important. The optimum pH is 4. Increasing the pH above 4 leads to a drop in absorbance. Relatively high concentrations of elements encountered in some iron ore, including phosphorus (e.g. BCS 303: 0.5% P) and vanadium (e.g. Philippine iron sand: 0.3% V), do not interfere. Zirconium or nickel leached during the fusion of the sample will not interfere. Identical results were obtained with zirconium or alternative nickel crucibles. The titanium complex with chromotropic acid is stable for 16 h under the conditions of the procedure. For standard reference samples the standard deviation was found to be 0.001% Ti at the 0.03% level and 0.008% Ti at the 0.5% level in the following procedure:[15]

17.4.1 Procedure

Accurately weigh 0.5000 g sample, following the test sample weighing procedure (section 8), into a dry 50-mL zirconium crucible, mix with 4 g of powdered sodium peroxide and fuse over a M'eker burner following the fusion dissolution procedure (section 13.1). Add 10 mL of water to the crucible and cover the beaker with a watch glass. After effervescence ceases, pour the contents of the crucible into the beaker and wash with 15–20 mL of water. Add 10 mL of concentrated HCl via the crucible as well as water rinsings into the beaker. Boil the solution for 2 min. Wash the watch glass and the walls of the beaker with water and continue boiling for about 30 s and cool. Transfer the solution to a 100-mL volumetric flask and dilute the contents with water to the mark and mix.

17.4.1.1 Photometry Transfer 10.00 mL aliquot into a 150-mL beaker. Add 10 mL of freshly prepared 5% ascorbic acid solution, mix, wait for 5 min, add 15 mL of freshly prepared 1% chromotropic acid solution and mix. Adjust (using a narrow-range pH paper) to pH 3.5–4.0 with acetate buffer (prepared by dissolving 220 g of sodium acetate trihydrate in water, adding 220 mL of glacial acetic acid and diluting the mixture to 1 L). Transfer to a 100-mL volumetric flask. Dilute to the mark with water. After 15 min measure absorbance at 470 nm in a 2-cm cuvette against water as reference. The color is stable for at least 2 h. Subtract the blank (0 calibration standard) to obtain net absorbance. Read off μg of titanium from the calibration graph. Calculate percent Ti by multiplying by 0.001 and dividing by sample weight in g.

17.4.1.2 Calibration Prepare fusion blank solution by weighing 0.300 g of pure iron powder into a zirconium crucible and carry out the procedure exactly. For the calibration, transfer 10.00 mL aliquot of fusion blank solution, into a series of 150-mL beakers. Add 0, 0.50, 1.0, 2.5, 5.0, 10.0 and 15.0 mL of 20 μg mL^{-1} titanium working solution (section 17.4.1.3). Carry out exactly as described in photometry (section 17.4.1.1). Plot net absorbance vs μg of titanium.

17.4.1.3 Standard Titanium Solution Fuse 0.1000 g of pure titanium (fine powder) with 2 g of potassium bisulfate in a platinum crucible at low heat. When clear melt is obtained, cool somewhat and dissolve the contents in 10 mL of diluted (1 + 1) H_2SO_4. Dilute to 1 L in a volumetric flask and mix. Dilute 20.00 mL of this solution to volume with water in a 100-mL volumetric flask to obtain the 20 μg mL^{-1} titanium working solution.

17.5 Determination of Vanadium by *N*-benzoyl-phenylhydroxylamine

Photometric determination of vanadium by BPHA, in steel including highly alloyed steel and high-temperature alloys, has been in use for a long time. It was thought desirable to investigate the application of BPHA for determining vanadium in iron ores using the rapid fusion/dissolution technique. After fusion of the sample and dissolution of the cooled melt in dilute sulfuric acid, vanadium was oxidized with permanganate and determined photometrically with BPHA. The critical point is ensuring oxidation of vanadium to the quinquevalent state before addition of HCl and subsequent extraction and color formation with BPHA. This is achieved in the described method.[15] Precision tests on two South African reference standard iron ores (10 replicates each) gave a standard deviation of 0.001% V at the 0.005% level and 0.002% V at the 0.05% level. Excellent agreement was found between the certified values Japan SRM 830-1: 0.31% V; 850-1: 0.05% V and South African SRM 11: 0.004% V; and SRM 12: 0.052% V; and those found with the procedure described as follows.

17.5.1 Procedure

Accurately weigh a 0.5000 g sample, following the test sample weighing procedure (section 8), into a dry 50-mL zirconium crucible, mix with 4 g of powdered sodium peroxide and fuse over a M'eker burner following the

fusion/dissolution procedure (section 13.1). Add 20 mL of water to the crucible and cover the beaker with a watch glass. After effervescence ceases, pour the contents of the crucible into the beaker and wash with 10 mL of water. Add 30 mL of diluted (1 + 4) sulfuric acid via the crucible as well as water rinsing into the beaker. Add 100 mL of diluted (1 + 4) sulfuric acid, then 1 or 2 drops of hydrogen peroxide until the color is clear yellow. Boil the solution for 2 min. Wash the watch glass and the walls of the beaker with water and continue boiling for about 30 s and cool. Transfer the solution to a 200-mL volumetric flask, dilute the contents with water to the mark and mix.

17.5.1.1 Photometry Transfer an appropriate aliquot into a 125-mL separating funnel (25 mL for 0.001–0.1% V, 10 mL for 0.1–0.2%, 5 mL for 0.2–0.5%, plus enough fusion blank solution (section 17.5.1.2) to give a total volume of 25 mL). Add 0.3% potassium permanganate (about 0.1 N) solution dropwise until a faint pink persists for 2 min, and then three drops more. Wait for 2 min. Add 20 mL of concentrated HCl and mix. Add 10 mL of freshly prepared BPHA (0.25% solution) in chloroform, and shake the funnel for 45 s. Let the layers separate; draw off the organic phase into a dry 50-mL beaker. Shake the aqueous phase in the funnel with 10 mL of chloroform for 30 s and draw off the organic layer in the beaker. Transfer the combined organic extracts into a dry 50-mL volumetric flask and make up the volume with chloroform. Measure the absorbance (within the next 2 h) in 1-cm cuvette at 530 nm against chloroform as reference. Subtract the blank (0-calibration) to obtain the net absorbance. Read off μg of vanadium from the calibration graph. Calculate percent V by multiplying μg of vanadium by 0.005 and dividing by sample weight (g) multiplied by sample aliquot (mL).

17.5.1.2 Calibration Prepare fusion blank solution by weighing 1.300 g of pure iron powder into a zirconium crucible, add 8 g of sodium peroxide and continue exactly as in the procedure. For the calibration curve, pipette 25 mL fusion blank solution into a series of 125-mL separating funnels. Add 0 (0 blank), 0.25, 0.5, 1.0, 2.0, 4.0, and 5.0 mL of 50 $\mu g\,mL^{-1}$ vanadium working solution (section 17.5.1.3) and enough water to make up the volume to 30 mL. Beginning from oxidation with potassium permanganate, continue exactly as described in photometry. Plot net absorbance against μg of vanadium.

17.5.1.3 Standard Vanadium Solution Weigh 0.230 g of ammonium metavanadate (NH_4VO_3) into a 150-mL beaker. Add 60 mL of water and gently simmer to dissolve. Cool, add 1 mL of concentrated sulfuric acid and dilute with water to the mark in a 100-mL volumetric flask. Pipette 5 mL of this solution into a 100-mL volumetric flask containing 50 mL of water and 1 mL of concentrated sulfuric acid and dilute with water to the mark in a 100-mL volumetric flask to obtain the working solution (V, 50 $\mu g\,L^{-1}$).

17.6 Determination of Arsenic by Silver Diethyldithiocarbamate–Ephedrine

In the past, reagents such as quercetin, morin and rutin have been employed for the photometric determination of arsenic as well as distillation separation as arsenic trichloride[24,25] and the subsequent formation of the arsenomolybdenum color complex. The present development is essentially based on the rapid photometric method[26] for the determination of arsenic in steel by absorbing the evolved arsine in silver diethyldithiocarbamate. It was established that the concentration of sulfuric acid and other parameters must be kept constant to avoid an erratic yield of arsenic and subsequent erratic results. In the universal fusion dissolution technique for iron ores, treating the fused melt with HCl and fuming with sulfuric acid leaves inconsistent amounts of the latter. It is more reliable to avoid fuming by dissolving the melt directly in dilute sulfuric acid for subsequent reduction with potassium iodide and stannous chloride in the photometry of the evolved arsine with silver diethyldithiocarbamate to obtain precise and accurate results. The procedure described,[5] also replaces the obnoxious smelling pyridine by substituting ephedrine in chloroform as solvent for silver diethyldithiocarbamate. The procedure is as follows.

17.6.1 Procedure

Run the test sample(s) and calibration standards simultaneously (taking identical sample weight and fusion in a zirconium crucible similar to calibration standards). Prepare calibration standards as follows: into each of six zirconium crucibles weigh 0.500 g of pure ferric oxide and 3 g of sodium peroxide for the range 0.0002–0.01% As or 0.05 g of ferric oxide and 1 g of sodium peroxide for the range 0.01–0.10% As. Mix thoroughly and carry out fusion for each crucible over a M'eker burner following the fusion/dissolution procedure (section 13.1). Cool the crucibles somewhat in air and place each in a separate dry 250-mL beaker. Add 10 mL of water into the crucible and cover the beaker with a watch glass. After effervescence ceases, pour the contents of the crucible into the beaker and rinse with 10 mL of water. Add 20 mL of diluted (1 + 1) sulfuric acid via the crucible as well as 10 mL of water. Transfer 0, 0.30, 0.50, 1.00, 2.00 and 5.00 of the 10 $\mu g\,mL^{-1}$ of arsenic working solution (section 17.6.1.1) into the beakers, avoiding any contamination with the simultaneously treated test sample(s). Gently boil the solutions for about 2 min. Cool. At this point the series of arsenic generation

and absorption apparatus described,[26] should be made ready by placing a small plug of lead acetate cotton (section 17.6.1.2) in position, connecting the capillary delivery tube removing the ''inner joints'' with the capillary tubes attached.

17.6.1.1 Arsenic Standard and Working Solution Dissolve 0.6601 g of arsenic trioxide (As_2O_3) in 20 mL of 20% w/v sodium hydroxide solution and dilute to 1 L with water in a volumetric flask. Dilute 10.00 mL of this solution ($As = 500\,\mu g\,mL^{-1}$) to the mark with water in a 500-mL volumetric flask. This working solution ($As = 10\,\mu g\,mL^{-1}$) is stable for at least a month.

17.6.1.2 Lead Acetate Cotton Dissolve 30 g of lead acetate trihydrate in 300 mL of water and make just acid with acetic acid. Soak absorbent cotton in this solution. Allow excess to drain out, then dry at 110 °C.

17.6.2 Photometry

Transfer the solutions to the respective generation flasks (test sample(s) and calibrations standards) with a minimum of water. To each flask add freshly prepared 15% potassium iodide solution, mix and let stand for 5 min. Add 40% stannous chloride (40 g $SnCl_2 \cdot 2H_2O$ in concentrated HCl and diluted to 100 mL with concentrated HCl) solution dropwise with swirling until the iron and the liberated iodine have been reduced (solution becomes colorless), then add five drops in excess. Let stand for 15 min. Meanwhile pipette 5 mL of arsine-absorption solution (section 17.6.2.1) into the absorption test-tubes. Quickly introduce 3.0 g of arsenic-free 20-mesh granulated zinc into the flask through a short-necked funnel and immediately fit the delivery and absorption system. Periodically swirl the flask. Let stand until no further gases are evolved (about 15–20 min). The funnel prevents any zinc from adhering to the ground glass joint (which would prevent a gas-tight seal). Immediately measure the absorbance in 1-cm cuvette at 540 nm against the arsine-absorption solution as reference. Subtract the blank (0-calibration) to obtain the net absorbance. Read off μg of arsenic from the calibration graph. Calculate percent arsenic by multiplying μg of arsenic by 10^{-4} and dividing by sample weight (g).

17.6.2.1 Arsine Absorption Solution Dissolve 0.41 g of ephedrine in 250 mL chloroform. Weigh 0.20 g of silver diethyldithiocarbamate into a 400-mL beaker, add a few milliliters of the ephedrine solution to make a paste, then add the remainder and stir to dissolve. Store in a glass-stoppered amber-colored bottle, away from light. Filter, if necessary, before use.

18 FLAME ATOMIC ABSORPTION SPECTROMETRIC METHODS

As indicated earlier, FAAS methods are used for elements such as Ca, Mg, Cu, Zn, Pb, Na, K, Ni and Cr. Prior to the advent of FAAS the wet chemical methods for determination of Na and K, particularly the Lawrence Smith method, were very tedious and of questionable accuracy. Basically, in FAAS technique, a solution containing a metal species is aspirated into a flame such as air/acetylene or nitrous oxide/acetylene. Some of the atoms may be thermally excited from the ground state for emission of characteristic radiation; however, most of the atomic population remains in the ground state capable of absorbing light energy of specific wavelengths corresponding to excitation to higher atomic energy levels. Hence if light from these wavelengths is passed through a flame containing atoms of the element (e.g. hollow cathode lamp), part of the energy will be absorbed and the absorption will be proportional to the concentration of these atoms in the flame. The instrument must meet the criteria for minimum sensitivity, curve linearity and minimum stability. Instrument manufacturer's safety instructions including igniting and extinguishing the nitrous oxide burner and recommended parameters for optimization etc. should be strictly followed. Acid dissolution sometimes leaves some analyte trapped in the insoluble residue and needing to be recovered by fusion dissolution. Examples of these procedures for Ca, Mg, will be described (section 18.1.1). Interference from some elements particularly the major element iron, is not uncommon; the latter needs to be separated by methods such as solvent extraction. Examples of two such methods, involving solvent extraction separation of iron, i.e. combined procedures for Pb/Zn (section 18.3.1) and Cr/Ni (section 18.4.1), will be described.

18.1 Determination of Calcium and Magnesium

The old tedious gravimetric methods for the determination of calcium (precipitation as calcium oxalate, ignited to CaO) and magnesium (precipitated as magnesium ammonium phosphate, ignited as magnesium pyrophosphate)[2] are hardly in use. A rapid scheme (unpublished) was devised for the anion exchange separation of iron and then determination of calcium and magnesium complexometrically. However these methods have now been superseded by FAAS, for both routine and reference analysis. The sample is dissolved in HCl and a small amount of HNO_3. After evaporation and dehydration, any insoluble matter is filtered, treated for the recovery of calcium and magnesium and added to the main solution. The solution is aspirated into the flame of the atomic absorption spectrometer.

A nitrous oxide/acetylene burner is used for calcium and an air/acetylene burner for magnesium. For calcium, a calcium-hollow cathode lamp and 422.7 nm (wavelength) are chosen. For magnesium, a magnesium-hollow cathode lamp and 285.2 nm (wavelength) are used. The following procedure details an example of acid dissolution in a FAAS method where the trapped insoluble analyte is recovered, but removal of iron is unnecessary.

18.1.1 Procedure

Weigh 1.000 g, sample as described in section 8 (test sample weighing procedure), and transfer to a 250-mL beaker. Carry a reagent blank through all steps of the procedure. Moisten the sample with a little water. Add 25 mL of HCl. Cover with a watch glass and then gently increase the heat to just below boiling until reaction ceases. Add 2 mL of HNO_3, digest for several minutes, evaporate to dryness and heat the salts for a few minutes. Add 5 mL of HCl, cover the beaker and warm for a few minutes. Add 50 mL of water and heat to dissolve the salts. Filter through a fine-texture filter paper collecting the filtrate in a 250-mL beaker. Wash the filter and residue alternately three times with hot diluted (1 + 9) HCl and hot water until the yellow color of ferric chloride is no longer apparent. Reserve the filtrate. Place the paper and residue in a platinum crucible and char it at 500 °C in a muffle furnace and then ignite at 700 °C for a few minutes. Cool the crucible. Add four drops of diluted (1 + 1) H_2SO_4 to moisten the residue in the crucible. Add 5 mL of HF. Cautiously and slowly heat to the absence of white SO_3 fumes. Cool, add 1 g of sodium carbonate and fuse over a M'eker burner until a clear melt is obtained. Cool the melt and dissolve it in the reserved filtrate. Remove and wash the crucible and cover. Cool. Transfer to a 200-mL volumetric flask, dilute to the mark with water, and mix. Transfer an appropriate aliquot (Table 3) of this solution to a 200-mL volumetric flask and add 40 mL of the lanthanum chloride solution and background solution and HCl (Table 3). Dilute to the mark with water, and mix. Proceed with the measurements described in section 18.1.2.

18.1.2 Calibrations, Reagents and Measurements for Flame Atomic Absorption Spectrometry

Calcium standard solution (1 mL = 25 μg Ca): dissolve 1.2487 g anhydrous calcium carbonate in 100 mL of diluted (1 + 3) HCl. When solution is complete, cool and dilute to volume with water in a 1-L volumetric flask. Transfer 10.00 mL of this solution to a 200-mL volumetric flask, dilute to the mark with water and mix.

Magnesium standard solution: (1 mL = 17.5 μg Mg): dissolve 0.3500 g of high-purity magnesium metal by slowly adding 75 mL of diluted (1 + 3) HCl. When solution is complete, cool and dilute to volume with water in a 1-L volumetric flask. Transfer 10.00 mL of this solution to a 200-mL volumetric flask, dilute with water to the mark and mix.

Background solution: dissolve 10.00 g of high-purity iron powder in concentrated HCl and oxidise with dropwise addition of HNO_3. Evaporate to a syrupy consistency, add 20 mL of HCl and dilute to 200 mL with water. Add 20 g of sodium chloride dissolved in 200 mL of water and dilute to volume with water in a 1-L volumetric flask.

Lanthanum chloride solution (50 g L^{-1}): dissolve 50 g of lanthanum chloride in 50 mL of HCl and 300 mL of hot water. Cool and dilute to volume with water in a 1-L volumetric flask.

Calcium and magnesium calibration standard solutions: transfer 5.00, 10.00, 20.00, 40.00, and 50.00 mL each of calcium and magnesium standard solutions to a series of 200-mL volumetric flasks. Dilute to 100 mL with water. Add 6 mL of HCl, 60 mL of background solution, 40 mL of lanthanum chloride solution, dilute to the mark with water and mix.

Zero calibration solution: transfer to a 200-mL volumetric flask 60 mL of background solution. Add 6 mL of HCl, 40 mL of lanthanum chloride solution, dilute to mark with water and mix.

Measurements: light the burner and aspirate water until the instrument comes under thermal equilibrium. Optimize the instrument response by adjusting the wavelength, fuel, (nitrous oxide/acetylene for Ca or air/acetylene for Mg) burner and nebulizer while

Table 3 Percent expected calcium and magnesium, aliquot, equivalent sample weight, background solution, and HCl addition

Element	Expected percent element in sample	Aliquot (mL)	Equivalent weight of sample (g)	Background solution to be added (mL)	HCl added (mL)
Calcium	0.05–0.5	50	0.25	45	4
Calcium	0.5–2.5	10	0.05	60	6
Calcium	2.5–10.0	2	0.01	60	6
Magnesium	0.05–0.25	40	0.2	50	4
Magnesium	0.25–1.25	10	0.05	60	6
Magnesium	1.25–5.0	2	0.01	60	6

aspirating the highest calcium or magnesium calibration solution to obtain the maximum absorbance. Aspirate water until a steady signal is obtained and adjust the instrument read out system to obtain zero absorbance. Aspirate the calcium or magnesium calibration standard solutions in the order of increasing absorbance starting with the zero calibration solution. Record the readings. Aspirate the test solutions at the proper points in the calibration series. Aspirate water between each calibration and test solution. Repeat all absorbance measurements at least two or more times. Obtain the net absorbance of each calibration solution by subtracting the average absorbance of the zero calibration solution. Similarly obtain the net absorbance of test solutions by subtracting the average absorbance of the reagent blank solution. Prepare a calibration curve for calcium and magnesium by plotting the net absorbance of the calibration solutions against micrograms of calcium or magnesium per mL. Convert the net absorbance values of the sample solutions to micrograms of calcium or magnesium per mL.

Calculation: calcium or magnesium percent $= (A \times 200)/(B \times 10\,000)$, where $A =$ micrograms of calcium and magnesium per mL and $B =$ sample represented in the 200-mL volume of the final test solution.

18.2 Determination of Copper

Range 0.003 to 1.0%: prior to the advent of FAAS, copper had been determined photometrically with reagents such as diethyldithiocarbamate and also by a specific reagent zinc-diethyldithiocarbamate,[27] and by 2,2′-biquinolyl. In the FAAS method, the sample is decomposed by HCl, HNO_3, HF and fumed after the addition of $HClO_4$. After dissolving the salts, filtration and dilution with water, copper is determined by FAAS using an air/acetylene flame using a copper-hollow cathode lamp at the resonance wavelength Cu 324.7 nm and steps similar to the calcium/magnesium procedure (section 18.1.1).

18.3 Determination of Lead and/or Zinc

Range 0.001 to 0.5% Pb and Zn. The lead molybdate gravimetric method for lead and the photometric dithizone methods for lead and zinc have questionable accuracy. In order to carry out the dithizone method, extensive separation of lead and zinc and other dithizone-complex forming elements need to be carried out. An anodic stripping voltammetric method[28] for the simultaneous determination of cadmium, copper, lead and zinc in iron ore, sinter etc. was devised for monitoring, for a research project. The FAAS method for lead and/or zinc by FAAS, though tedious, is preferable in terms of accuracy and precision. The sample is dissolved in HCl and HF, oxidized with HNO_3, filtered and the residue reserved. After evaporation the salts are dissolved in HCl, filtered and the filtrate evaporated to dryness and salts redissolved in HCl. Iron is then removed by MIBK solvent extraction. Lead and zinc are extracted into the aqueous phase which is treated with HNO_3 to destroy any organic matter and then evaporated to dryness. The reserved residue is ignited, fused with sodium carbonate, the melt dissolved in dilute HCl and then this solution is finally added to the evaporated and dried extracted lead and/or zinc. This solution is ready for aspiration into the flame of FAAS, using an air/acetylene burner, lead- and zinc-hollow cathode lamps and absorbance measurement at the respective resonance wavelength (Pb 283.3 nm) and (Zn 213.9 nm).

18.3.1 Procedure

Weigh 2.000 g as described in section 8 (test sample weighing procedure) and transfer to a 250-mL polytetrafluoroethylene (PTFE) beaker. Carry a reagent blank through all steps of the procedure. Moisten the sample with a little water. Add 40 mL of HCl and 10 mL of HF. Cover with a PTFE cover, heat gently and then increase the heat to 200 °C. Evaporate to dryness. Add 5 mL of HNO_3 and evaporate to dryness again. Dissolve the salts with 5 mL of HCl. Add 10 mL of water and filter through a fine-texture filter paper collecting the filtrate and washings into a 250-mL beaker. Wash the filter and residue with diluted (2 + 98) HCl until free from the yellow iron stain and then three times with hot water. Reserve the residue and combined filtrate and washings.

18.3.1.1 Treatment of Residue Place the paper and residue in a platinum crucible and char at 500 °C in a muffle furnace. Add 0.5 g of sodium carbonate and fuse over a M'eker burner until a clear melt is obtained. Cool the melt, add 5 mL of HCl, heat to dissolve and reserve the solution.

18.3.1.2 Treatment of Combined Filtrate and Washings and Removal of Iron Evaporate the filtrate and washings (section 18.3.1) just to dryness and add 20 mL of diluted (10 + 6) HCl and transfer to a 200-mL separating funnel. Rinse the beaker with 20 mL of diluted (10 + 6) HCl and combine with the solution in the separating funnel. Add 50 mL of MIBK and shake thoroughly for 1 min. Allow the layers to separate, then run the lower aqueous phase into a 250-mL beaker. Wash the organic phase by extracting with 10 mL of diluted (10 + 6) HCl and transfer the washings to the beaker. Heat the solution gently to expel almost all of the MIBK in the solution. Then add 5 mL of HNO_3 and evaporate to dryness. Dissolve the salts in 15 mL of diluted HCl (1 + 1). Combine this solution with the reserved solution from section 18.3.1.1. Transfer this solution to a 50-mL volumetric flask, dilute to the mark with water

Table 4 Dilution and reagent guide for test solution for lead using FAAS

Expected percent Pb in sample	Aliquot (mL)	Sodium carbonate to be added (g)	Diluted HCl (1 + 1) to be added (mL)
0.001–0.005	–	–	–
0.05–0.20	20	0.8	32
0.20–0.25	10	0.9	36
0.25–0.50	5	0.95	38

Table 5 Dilution and reagent guide for test solution for zinc using FAAS

Expected percent Zn in sample	Aliquot (mL)	Sodium carbonate to be added (g)	Dilute HCl (1 + 1) to be added (mL)
0.001–0.006	–	–	–
0.006–0.02	25	0.75	30
0.02–0.06	10	0.9	36
0.06–0.12	5	0.95	38
0.12–0.30	2	0.98	39
0.30–0.50	1	0.99	40

and mix. Depending on the concentration of lead and zinc (dilution and reagent guide Tables 4 and 5), use this solution for FAAS measurements either without dilution, or as specified. If dilution is required, transfer the appropriate aliquot to a 250-mL beaker. Add the amount of sodium carbonate and diluted HCl (1 + 1) as indicated for lead (Table 4) and for zinc (Table 5). Heat the solution to expel carbon dioxide. Cool and transfer the solution to a 100-mL volumetric flask, dilute to the mark with water and mix. Dilutions shown in Tables 4 and 5 will provide percent lead/zinc falling within the range of the calibration solutions in section 18.3.1.3. Transfer corresponding amounts of blank test solution to 250-mL beaker and add the same amount of sodium carbonate and diluted HCl (1 + 1) as used for the test solution. Heat the solution to expel carbon dioxide. Cool and transfer the solution to a 100-mL volumetric flask and dilute to the mark with water and mix. This solution is the diluted blank test solution.

18.3.1.3 Calibration Solutions, Reagents, and Measurements for Flame Atomic Absorption Spectrometry
Lead standard solutions: dissolve 0.5000 g of high-purity lead metal in 40 mL of diluted (1 + 1) HNO_3. Cool. Transfer to a 500-mL volumetric flask, dilute with water to volume and mix. This stock solution contains 1 mg of Pb per mL. Prepare standard solution ''A'' by transferring 10.00 mL of stock solution to a 100-mL volumetric flask, dilute to the mark with water and mix. Standard solution ''A'' contains 0.10 mg of Pb per mL. Prepare standard solution ''B'' by transferring 10.00 mL of standard solution ''A'' to a 100-mL volumetric flask, dilute to the mark with water and mix. Standard solution ''B'' contains 0.010 mg of Pb per mL.

Zinc standard solution: dissolve 0.5000 g of high-purity zinc metal in 40 mL of diluted (1 + 1) HCl. Cool. Transfer to a 500-mL volumetric flask, dilute with water to volume and mix. This stock solution contains 1 mg of Zn per mL. Prepare standard zinc solution by transferring 10.00 mL of stock solution to a 1000-mL volumetric flask, dilute to the mark with water and mix. This standard zinc solution contains 0.010 mg of Zn per mL.

Preparation of calibration solutions for lead: to a series of six 250-mL beakers transfer 1.0 g of sodium carbonate. Add 20 mL of HCl and transfer 0.0 (zero solution); 20.00 mL of standard solution ''B''; and 5.00, 10.00, 15.00 and 20.00 mL of standard solution ''A'' respectively. Heat the solution to expel carbon dioxide. Cool and transfer the respective solutions to a series of six 100-mL volumetric flasks, dilute to the mark with water and mix.

Preparation of calibration solutions for zinc: to a series of six 250-mL beakers transfer 1.0 g of sodium carbonate. Add 20 mL of HCl and transfer 0.0 (zero solution), 5.00, 10.00, 15.00, 20.00 and 25.00 mL of zinc standard solution respectively. Heat the solution to expel carbon dioxide. Cool and transfer the respective solutions to a series of six 100-mL volumetric flasks, dilute to the mark with water and mix.

Measurements: follow instrument set up and absorbance measurement exactly as in section 18.1.2, using appropriate wavelengths (Pb 283.3 nm; Zn 213.9 nm) and air/acetylene burner. Prepare calibration curves for lead or zinc by plotting the net absorbance of the calibration solutions against mg mL^{-1} of Pb or Zn. By referring to the calibration graph convert the net absorbance value of the test solution or diluted test solution to mg mL^{-1} of Pb or Zn. Calculate percent lead or zinc in the sample, commensurate with sample dilution.

18.4 Determination of Nickel and/or Chromium

As indicated in section 18, the FAAS method for the determination of nickel and/or chromium also involves the prior separation of iron by solvent extraction as described in detail in the procedures in section 18.3.1. For brevity, only the outline is described. The sample is dissolved in HCl and oxidized with HNO_3, and the filtered residue is reserved. After evaporation the salts are dissolved in HCl, filtered, the filtrate evaporated to dryness and salts redissolved in HCl. Iron is then removed by MIBK solvent extraction. Nickel and/or chromium are extracted into the aqueous phase which is then treated with HNO_3 to destroy any organic matter and evaporated to dryness. The reserved residue is ignited, fused with sodium carbonate–sodium tetraborate, and the melt dissolved in dilute HCl. This solution is finally added to the evaporated and dried extracted metal/s. This solution is ready for aspiration into the flame of FAAS, using an air/acetylene burner for nickel and a nitrous oxide/acetylene burner for chromium along with Ni- and Cr-hollow cathode lamps with absorbance measurement at resonance wavelengths Ni 232.0 nm and Cr 357.9 nm, respectively.

18.5 Determination of Sodium and Potassium

As indicated earlier, prior to the advent of FAAS, methods for the determination of these elements were tedious and somewhat unreliable. The FAAS method is quite straightforward. It is important to avoid contamination of sodium by glass apparatus. In brief, the sample is decomposed by HCl and HF in a PTFE beaker and evaporated to dryness. The treatment with HCl is repeated. After dissolution with HCl followed by appropriate dilution (in plastic volumetric flasks etc.), sodium and potassium are determined by FAAS using an air/acetylene flame using sodium- and potassium-hollow cathode lamps at the resonance wavelengths Na 589.0 nm and K 766.5 nm, respectively.

19 INDUCTIVELY COUPLED PLASMA/ATOMIC EMISSION SPECTROMETRY FOR MULTI ELEMENTS

With the ever increasing demand for rapid analytical control simultaneous determination of several elements is preferred. The ICP/AES technique uses solution and is generally used for minor elements. The iron ore sample is decomposed by fusion in a mixed sodium carbonate/sodium tetraborate flux and the melt dissolved in HCl. After dilution to volume the measurements are made on an ICP spectrometer for elements in iron ore, such as Al, Ca, Mg, Mn, P, Si, and T, and referred to calibration graphs using standard solutions. A procedure blank is carried out side by side for blank correction.

19.1 Procedure

Add 0.80 g of sodium carbonate to a 40-mL capacity platinum crucible or a suitable platinum-alloy crucible. Weigh 0.5000 g sample, as described in section 8 (test sample weighing procedure), transfer to the crucible and mix well with a platinum or stainless steel rod. Add 0.40 g of sodium tetraborate and repeat the mixing with the metal rod. Prefuse over a low dull red heat on a Bunsen burner with manual agitation. The mixture melts within 2–3 min. After this prefusion, place the crucible in a muffle furnace at 1020 °C for 15 min. Remove the crucible and gently swirl the melt as it solidifies. Allow to cool and the place a PTFE-coated stirring bar in the crucible and place the crucible in a 250-mL low-form beaker. Add 40 mL of diluted (1 + 1) HCl directly into the crucible, and 30 mL of water into the beaker. Cover and heat with stirring on a magnetic stirring hot plate until dissolution of the melt is complete. Remove and rinse the crucible and stirrer. Cool the solution and immediately transfer (to prevent reprecipitation) to a 200-mL volumetric flask. Dilute to volume with water and mix. Carry out a procedure blank side by side.

19.1.1 Calibration Solutions, Reagents and Measurements

Aluminum standard solution (1 mL = 1 mg Al): dissolve 1.000 g of high-purity aluminum in 20 mL of diluted (1 + 1) HCl in a covered beaker. Add about 10 drops of HNO_3. When dissolution is complete, add 20 mL of water and boil to expel oxides of nitrogen. Cool and transfer to a 1000-mL volumetric flask, dilute to the mark with water and mix.

Calcium standard solution (1 mL = 1 mg Ca): dissolve 2.4972 g of anhydrous calcium carbonate in 20 mL of diluted (1 + 1) HCl. When solution is complete, cool, dilute to volume with water in a 1-L volumetric flask and mix.

Magnesium standard solution (1 mL = 1 mg Mg): dissolve 1.000 g of high-purity magnesium metal by slowly adding 20 mL of diluted (1 + 1) HCl. When solution is complete, cool, dilute to volume with water in a 1-L volumetric flask and mix.

Manganese standard solution (1 mL = 1 mg Mn): dissolve 1.000 g of high-purity manganese metal by slowly adding 20 mL of diluted (1 + 1) HCl. When solution is complete, cool, dilute to volume with water in a 1-L volumetric flask and mix.

Phosphorus standard solution (1 mL = 1 mg P): dry potassium dihydrogen orthophosphate (KH_2PO_4) at 110 °C, cool it in a desiccator. Dissolve 4.3936 g in 200 mL of water. Transfer and dilute to volume in a 1-L volumetric flask with water and mix.

Silicon standard solution (1 mL = 1 mg Si): weigh 2.1393 g of pure silicon dioxide (finely ground, previously heated at 1000 °C for 45 min) into a platinum crucible. Mix with 5 g of sodium carbonate and fuse in a muffle furnace for 15 min at 1000 °C. Cool somewhat and dissolve the melt in 100 mL of water. Transfer to a 1-L volumetric flask, add about 400 mL of water and mix. Add 20 mL of diluted (1 + 1) HCl, dilute to volume with water and mix.

Titanium standard solution (1 mL = 1 mg Ti): dissolve 1.000 g of high-purity titanium metal in 100 mL of diluted (1 + 1) HCl. When solution is complete, cool, dilute to volume with water in a 1-L volumetric flask and mix.

Calibrations and reference solutions: for each of the elements Al, Ca, Mg, Mn, P, Si and Ti, prepare calibration solutions following the procedure in section 18.1, replacing the test sample with the equivalent amount of iron oxide (Fe_2O_3) of minimum purity 99.99%. Prior to dilution to 200 mL, the stock solution (element concerned) and diluted (1 + 1) HCl are added in sufficient amounts to retain the final acid concentration [40 mL of diluted (1 + 1) HCl]. In addition, to comply with the requirements of similarity, calibration solutions are prepared from reagents taken from the same containers, to minimize purity differences between batches.

Measurements: any conventional ICP spectrometer may be used. Set up according to manufacturer's recommendation and ensure compliance with the performance parameters (Table 6) to be carried out prior to measurements. Suggested analytical lines performance parameters are shown in Table 6. The correction method for spectral interferences using synthetic standard solutions is recommended. Instrument manufacturers may provide software for various calculations (statistics) and algorithms for correction for inter-elemental spectral interferences etc. After instrument set up, aspirate water until a steady signal is obtained and adjust the instrument read out system to obtain zero absorbance. Aspirate the calibration standard solutions in the order of increasing absorbance, starting with the zero calibration solution. Record the readings. Aspirate the test solutions at the proper points in the calibration series. Aspirate water between each calibration and test solution. Repeat all measurements at least two or more times. Obtain the net intensity of each calibration solution by subtracting the average intensity of the zero calibration solution. Similarly obtain the net intensities of test solutions by subtracting the average intensity of the reagent blank solution. Prepare a calibration curve for the element(s) by plotting the net intensity of the calibration solutions against micrograms of element per mL. By referring to the calibration graph, convert the net intensity value of the test solution to micrograms of element per mL for calculating percent element(s) in the sample.

Table 6 Suggested analytical lines (ICP/AES) and recommended performance parameters

Element	Wavelength (nm)	DL (μg/mL)	BEC (μg mL^{-1})	$RSDN_{min}$ (%)
Al	396.15 or 308.22	0.04	2.46	0.87
Ca	393.36 or 317.22	0.02	1.04	1.04
Mg	279.55 or 279.08	0.03	0.38	0.75
Mn	257.61	0.01	0.29	0.89
P	178.29*	0.07	2.15	1.04
Si	251.61 or 288.16	0.07	2.67	0.95
Ti	334.94 or 336.12	0.01	0.24	0.78

* Check and correct if necessary for Mn interference. DL, detection limit; BEC, background-equivalent concentration; $RSDN_{min}$, short-term precision.

20 X-RAY EMISSION SPECTROMETRY

Single as well as several elements are determined by sequential and/or simultaneous modes by this technique, used primarily for routine control. In brief, the sample is irradiated by an intense beam of X-ray photons, which may impinge on atom/s in the sample resulting in the characteristic fluorescence wavelengths spectra. The beam of fluorescent radiation, characteristic of some or all of the elements in the sample, is then resolved into its constituents through an analyzing crystal, e.g. lithium fluoride. The X-ray quanta of this characteristic monochromatic wavelength are collimated on a radiation detector such as a scintillation or proportional counter, which translates them into voltage pulses integrated over a fixed period of time. This electrical output is proportional to concentration of the element in the sample. The pressed powder technique using a binder for iron ore sample preparation, is hardly used, owing to inherent matrix problems. For quantitative analysis, the finely ground iron ore sample is fused with borax. The fused melt is poured into a platinum dish to form a flat bead ready for XRF measurements. Calibrations are prepared using iron ore certified reference materials (CRMs). Inter-elemental corrections and correlation coefficients are calculated by making fused beads of binary mixture (element plus pure ferric oxide).

20.1 Procedure

Add 8.00 g of borax (sodium tetraborate: prescreened stock) and 0.480 g sample into a glass vial. Cover the vial with a plastic snap and shake the vial to mix the borax

Table 7 Suggested XRF wavelengths and analyzing crystals for iron ore analysis

Element	Wavelength (nm)	Crystal	Element	Wavelength (nm)	Crystal
Ca	3.36	LiF	Mg	9.889	ADP
Si	7.126	PET	Al	8.339	PET
Fe	1.937	LiF	Mn	2.103	LiF
K	3.742	LiF	P	6.155	PET
Cr	2.291	LiF	V	2.505	LiF
Ti	2.76	LiF			

ADP, ammonium dihydrogen phosphate; PET, pentaerythritol.

and sample. Transfer the contents quantitatively to a platinum : gold (95 : 5) alloy crucible. Using a micropipette add 0.100 mL of sodium iodide solution (0.25 g mL^{-1}) and fuse at 1200 °C for 15 min. (Automated fusion apparatus is preferable.) Pour the melt on a heated platinum dish to form a flat and uniform bead, then air cool. Carry out XRF measurements using an appropriate analyzing crystal and detector.

20.1.1 Measurements

Table 7 lists suggested parameters for XRF analysis in conjunction with a rhodium target in the X-ray tube. Fused beads of binary mixtures (element plus pure ferric oxide) are prepared for the sought element, identically to the procedure) to compute inter-elemental correlation coefficients. Prepare standards for calibrations, using fused beads of iron ore SRMs/CRMs, exactly as in the procedure. Store the fused beads in a desiccator to avoid exposure to humidity. Plot percent element vs counts or intensity ratios and compute the concentration of element(s).

21 STANDARD/CERTIFIED REFERENCE MATERIALS, QUALITY CONTROL AND QUALITY ASSURANCE

It should be noted that SRMs and CRMs are used interchangeably in this article. Any use of a SRM depends on the ability to make valid inferences from the measurement results. This involves the tacit assumption or demonstrated evidence that the material is reliable and capable of challenging the measurement process. For quality assurance and quality control SRMs are used. Instrumental methods require SRMs for calibration. The latter are available from producers of SRMs such as NIST, British Standards, Euro Standards, Australian and Japanese Standards Institutions, etc. A useful resource is the NIST Special Publication 260-100.[29] It is important for the SRM users to be aware of the verification and accuracy of reference test methods used in the development of SRMs as well as the establishment of measurement traceability. Assurance of measurement compatibility is crucial for direct calibration of methods and instruments as well as internal and external quality assurance.

22 STATISTICS, ACCURACY AND PRECISION

These are an integral part of the establishment of analytical methods. Accuracy or trueness refers to the closeness of agreement between the arithmetic mean of large number of test results and the true or accepted reference value. Precision refers to the closeness of agreement between test results. Many factors, apart from variations between supposedly identical specimens, may contribute to the variability of results for an analytical method. These include, the operator; the equipment used; the calibration of the equipment; the environment (temperature, humidity, air pollution, etc.); the time elapsed between measurements. Statements for repeatability and reproducibility are essential indicators for the analysis as well as for quality control and quality assurance systems in the analytical laboratory. Standards on statistics, e.g. ISO 5725[30] and ASTM E 173[31] are useful for assessing these valuable statistics on analytical method performance, particularly the extensive round robin testing carried out for establishing standard methods.

ABBREVIATIONS AND ACRONYMS

ADP	Ammonium Dihydrogen Phosphate
ASTM	American Society for Testing Materials
BCS	British Chemical Standards
BEC	Background-equivalent Concentration
BPHA	*N*-benzoyl-phenylhydroxylamine
CRMs	Certified Reference Materials
DL	Detection Limit
DRI	Direct Reduced Iron
EDTA	Ethylene Diamine Tetra-acetic Acid
FAAS	Flame Atomic Absorption Spectrometry
ICP	Inductively Coupled Plasma
ICP/AES	Inductively Coupled Plasma/Atomic Emission Spectrometry
IR	Infrared
ISO	International Standards Organization
MIBK	Methyl Isobutyl Ketone

NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
PET	Pentaerythritol
PTFE	Polytetrafluoroethylene
RSDN	Short-term Precision
SRM	Standard Reference Material
XRF	X-ray Fluorescence

RELATED ARTICLES

Atomic Spectroscopy **(Volume 11)**
Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

X-ray Spectrometry **(Volume 15)**
Sample Preparation for X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Gravimetry • Titrimetry

REFERENCES

1. ISO/IEC Guide 51, Safety Aspects – Guidelines for Their Inclusion in Standards. International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland, 1997.
2. G.E.F. Lundell, J.R. Hoffman, H.A. Bright, *Chemical Analysis of Iron and Steel*, John Wiley & Sons, New York, 1931.
3. T.S. Harrison, *Handbook of Analytical Control of Iron and Steel Production*, Ellis Horwood, New York, 1979.
4. O.P. Bhargava, W.G. Hines, 'Automated Composite Analysis of Major Sinter Components', *Anal. Chem.*, **48**, 1701–1704 (1976).
5. O.P. Bhargava, M. Gimitro, W.G. Hines, 'Universal Technique for Rapid Dissolution of Materials Encountered in Steel Industry and its Applications on the Photometric Determination of Arsenic, Phosphorus, Titanium and Vanadium in Iron Ores', *Talanta*, **27**, 263–267 (1980).
6. O.P. Bhargava, 'Rapid Sample Dissolution and Determination of Total Iron in Iron Ore, Sinter, Concentrates and Agglomerates', *Analyst*, **101**, 125–127 (1976).
7. O.P. Bhargava, A. Alexiou, W.G. Hines, 'Mercury Pollution Free Rapid Method for Total Iron Determination in Iron Ores, Sinter and Related Materials', *Talanta*, **25**, 357–358 (1978).
8. ISO 9508 Determination of Total Iron Content – Silver Reduction Titrimetric Method, International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland.
9. ASTM E 1081 Test Method for Total Iron in Iron Ores by the Silver Reduction Titrimetric Method. Annual Books of Standards, Section 3, Volume 03.06, Analytical Chemistry for Metals, Ores, and Related Materials, ASTM 1916 Race Street, Philadelphia, PA 19103–1187 USA, 1995.
10. ISO 9507 Determination of Total Iron Content – Titanium(III) Chloride Reduction Titrimetric Method, International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland.
11. ASTM E 1028 Test Method for Total Iron in Iron Ores by the Pollution-free Titrimetric Method, Annual Books of Standards, Section 3, Volume 03.06, Analytical Chemistry for Metals, Ores, and Related Materials, ASTM 1916 Race Street, Philadelphia, PA 19103–1187 USA 1995.
12. O.P. Bhargava, W.G. Hines, 'Chromazurol S Spectrophotometric Determination of Alumina in Iron Ores, Sinters, and Open Hearth Slags', *Anal. Chem.*, **40**, 413–415 (1968).
13. O.P. Bhargava, 'Direct Complexometric Determination of Aluminum in Alzinoy', *Talanta*, **22**, 471–472 (1975).
14. O.P. Bhargava, 'Complexometric Determination of Aluminum in Iron Ores, Sinter, Concentrate and Agglomerates', *Talanta*, **26**, 146–148 (1979).
15. ISO 6830 Determination of Aluminum Content – EDTA Titrimetric Method, International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland.
16. ASTM E 738 Test Method for Aluminum in Iron Ores by Complexometric Titration, Annual Books of Standards, Section 3, Volume 03.06, Analytical Chemistry for Metals, Ores, and Related Materials, ASTM 1916 Race Street, Philadelphia, PA 19103–1187 USA, 1995.
17. O.P. Bhargava, M. Gimitro, 'Rapid Photometric Determination of Phosphorus as Phospho-molybdenum-blue in Iron Ores and Related Materials', *Talanta*, **31**, 301–303 (1984).
18. ISO 4687 Determination of Phosphorus Content. Part 1. Molybdenum Blue Spectrophotometric Method, International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland.
19. ASTM E 1070 Test method for Phosphorus in Iron Ores by the Phospho-Molybdenum-Blue Photometric Method. Annual Books of Standards, Section 3, Volume 03.06, Analytical Chemistry for Metals, Ores, and Related Materials, ASTM 1916 Race Street, Philadelphia, PA 19103–1187 USA, 1995.
20. W.F. Sanders, C.H. Cramer, 'Photometric Determination of Silicon in Steel' *Anal. Chem.*, **29**, 1139–42 (1957).
21. O.P. Bhargava, G.F. Pitt, W.G. Hines, 'Unified Automated Determination of Silicon in Iron Ores, Sinters and Slags, Iron and Steel', *Talanta*, **18**, 793–798 (1971).
22. ISO 4691 Determination of Titanium Content – Diantipyrylmethane Spectrophotometric Method, International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland.

23. ASTM E 878 Test Method for Titanium in Iron Ores by the Diantipyrylmethane Photometric Method. Annual Books of Standards, Section 3, Volume 03.06, Analytical Chemistry for Metals, Ores, and Related Materials, ASTM 1916 Race Street, Philadelphia, PA 19103–1187 USA, 1995.
24. P. Paklans, 'Spectrophotometric Determination of Traces of Arsenic by an Extraction Method', *Anal. Chem. Acta*, **47**, 225–227 (1969).
25. W.R. Nall, 'An Improved Method for Determination of Arsenic in Steel', *Analyst*, **96**, 398–400 (1971).
26. O.P. Bhargava, J.F. Donovan, W.G. Hines, 'Rapid Spectrophotometric Determination of Arsenic in Iron and Steel', *Anal. Chem.*, **44**, 2402–2404 (1972).
27. O.P. Bhargava, W.G. Hines, 'Rapid Spectrophotometric Determination of Copper in Steel with Zinc-diethyldithiocarbamate', *Talanta*, **16**, 743–745 (1969).
28. O.P. Bhargava, W.D. Lord, W.G. Hines, 'Anodic Stripping Voltammetric Determination of Lead and Zinc in Iron Ore, Sinter, Concentrates and Agglomerates', *American Laboratory*, **11**, 11–14 (1975).
29. NIST Special Publication 260-100 Standard Reference Materials (SRM), Handbook for SRM users, John K. Taylor, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.
30. ISO 5725 Accuracy (Trueness and Precision) of Measurement Methods and Results, International Organization for Standardization, ISO Central Secretariat 1, rue de Varembe, Geneva, Switzerland.
31. ASTM E 173 Conducting Interlaboratory Studies of Methods for Chemical Analysis of Metals, Annual Books of Standards, Section 3, Volume 03.05 Analytical Chemistry for Metals, Ores, and Related Materials, ASTM 1916 Race Street, Philadelphia, PA 19103–1187 USA, 1995.

Metal Analysis, Sampling and Sample Preparation in

Charles K. Deak
C.K. Deak Technical Services, Inc., Warren, USA

1 Introduction 1
2 Sampling 2
2.1 Molten Metals 2
2.2 Ingots, Billets, and Large Castings 4
2.3 Mill Products 4
2.4 Particulate Materials, Ores, and Minerals 5
2.5 Raw Materials 7
2.6 Finished Products 7
2.7 Scrap Metals and Metallic By-products 8
2.8 Special Considerations 11
3 Sample Preparation 12
3.1 Sample Comminution 12
3.2 Sample Melting 14
3.3 Mechanical Operations 16
3.4 Dissolution and Fusion 18
Acknowledgments 23
Abbreviations and Acronyms 23
Related Articles 23
References 23

This article deals with sampling and sample preparation for the chemical analysis of metals. It discusses the types of samples, their definitions, purposes, and requirements. Sampling of various types, shapes, and forms of metals ranging from liquid molten steel to scrap metals is described, along with ways of rendering a sample homogeneous. The various methods of making a sample suitable for the required analytical procedures are explained in the sample preparation section. This section discusses melting of samples in various types of laboratory melting furnaces and machining samples for use in all types of determinations using various techniques such as optical emission spectroscopy (OES), X-ray fluorescence spectroscopy (XRF), atomic spectroscopy, thermal analyses, wet chemistry and so on. Discussion also centers around dissolving and fusing samples to convert them into a solution which is required by wet chemical, atomic spectroscopy, and other procedures.

Sampling and analysis of fasteners (regulated in the United States), and instances in which the sample or the analysis is or may be involved in litigation are also included.

1 INTRODUCTION

When a metal is analyzed to establish its chemical composition, it is almost impossible and completely impractical to analyze the entire piece or lot of metal. Only a portion of the metal is analyzed, referred to as a sample. In fact, a distinction must be made between several forms of sample. According to the proper nomenclature[1] there are several kinds of sample:

- A *sample* is a portion of the material intended to be representative of the whole.
- A *gross sample* is a single large sample obtained by placing in a single container two or more increments taken from a lot.
- A *prepared sample* is a sample on which comminution, division, blending, or other procedures have been performed to make the sample ready for analysis.
- A *laboratory sample* is a subsample selected from the properly prepared gross sample of the lot for submission to one or more laboratories for chemical analysis where *select* is understood to mean: to mix and divide a sample (with or without comminution) to ensure that the subsample equally represents the original lot.
- A *test sample* is a subsample selected from the properly prepared laboratory sample which has a suitable sample weight or volume for one or more determinations by chemical analysis.

The laboratory sample that it is going to be analyzed has two basic requirements. First, the sample, regardless of whether it is a gross sample, prepared sample, laboratory sample, or test sample, has to be representative of the entire piece or lot of metal. In other words, the chemical composition of the sample must be identical to the metal it represents. This is a simple matter if the metal that the sample is to represent is completely homogeneous and any portion of the metal to be analyzed has exactly the same composition as any other portion. Then, any single portion, wherever it is taken from, would be a good representative sample. In the practical world, such perfect homogeneity does not normally exist and even if it does exist its existence cannot be assured until extensive tests are done. As a result, it is the sampling procedures and equipment that have to provide for the fact that

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the material is heterogeneous. Sampling procedures and equipment will be discussed in section 2.

The second requirement for a sample is that it should be suitable for analysis by the method chosen to perform the analysis or determination. For example, if the analysis of a metal is by atomic absorption spectroscopy (AAS) the sample has to be in the form of a solution. If the analysis is by OES using the point-to-plane technique, a solid surface of the correct size that has the proper surface preparation has to be available. Furthermore, if the gross sample or representative subsample is rendered homogeneous, it is much easier to obtain the final test sample. Procedures and equipment for improving homogeneity and rendering the sample suitable for particular methods of analyses will be discussed in section 3.

It should be remembered that the sampling and sample preparation procedures described and discussed in this chapter are intended only for the chemical analysis of metals and are not intended for the determination of physical properties such as hardness or tensile strength. The physical properties can change considerably during the discussed sampling and/or sample preparation procedures.

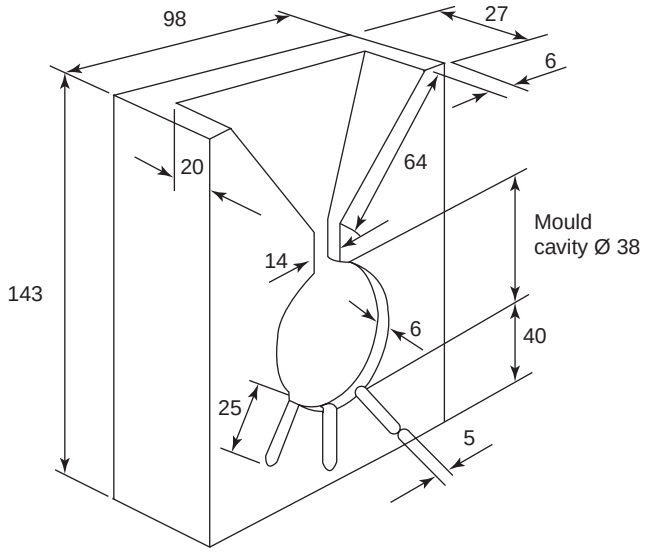

Figure 1 Combination-type vertical mold used for sampling liquid iron intended for cast iron production. Note that all dimensions are in millimeters and that the flat chill plate (not shown) has similar overall dimensions. (Reproduced by permission of ASTM (1997).[2])

2 SAMPLING

When sampling different metals or metal products the objective is always the same: to obtain the sample that is representative of the lot. However, the methods and tools used to obtain this objective obviously cannot be the same for liquid molten steel, or a 25-mm diameter titanium bar, or a 5000-kg lot of nickel base alloy scrap consisting of machine shop turnings. In other words, the type, form, and origin of the metal to be sampled have a strong influence on the method of sampling. The principal forms of metal that are customarily sampled will be discussed below.

2.1 Molten Metals

2.1.1 Iron and Steel

Several procedures for steel and iron are discussed in detail in American Society for Testing and Materials (ASTM) Standard Practice E 1806,[2] which discusses sampling procedures for liquid iron for steelmaking and pig iron production, liquid iron for cast iron production, and liquid steel for steel production. In pig or cast iron samples, it must be understood that the sample has to be chill-cast so the metallurgical structure will be suitable for spectroscopic analysis. Iron samples are generally cast into a coin or disk mold. Frequently these are combination molds in which one or more pin samples are attached that can be broken off later and used for thermal-type analyses such as carbon and sulfur determinations. A typical vertical mold is shown in Figure 1. A preheated steel spoon is filled with the molten iron by sampling from a stream or immersing in the melt. Any slag is removed and the liquid iron is poured into a cold mold without delay. It is important that the mold should be free from moisture.

Liquid iron and steel samples are frequently taken with probes. These fall into two categories, immersion sampling probes and stream sampling probes. Figure 2 shows typical immersion probes. Many of the probes are of the ''disk and pin'' variety, where the pin can easily be broken off and used for thermal analysis. These probes can be lowered into the melt manually or by mechanical means. Some of the probes also incorporate a thermocouple to take a temperature reading of the bath simultaneously. Figure 3 shows stream sampling and suction sampling probes. The probes frequently contain a deoxidant. There is one more type of probe, which is especially made for obtaining samples for hydrogen determinations. This type consists mostly of silica tubing mounted in cardboard tubing designed to produce pin or pencil samples 7–12 mm in diameter and up to 150 mm long.

2.1.2 Aluminum and Aluminum Alloys

Aluminum or aluminum alloys are cast most often into book molds, in which the sample disks are either in

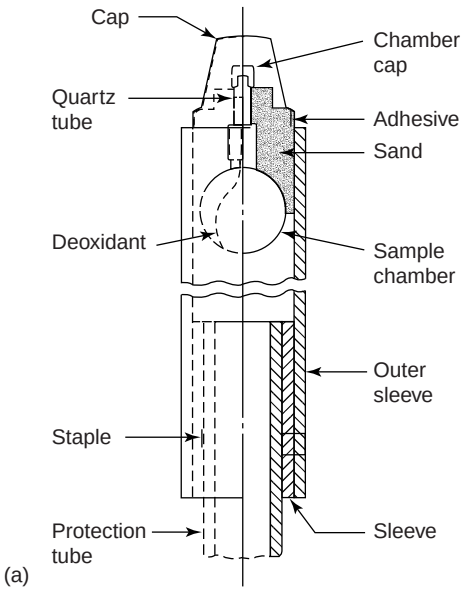

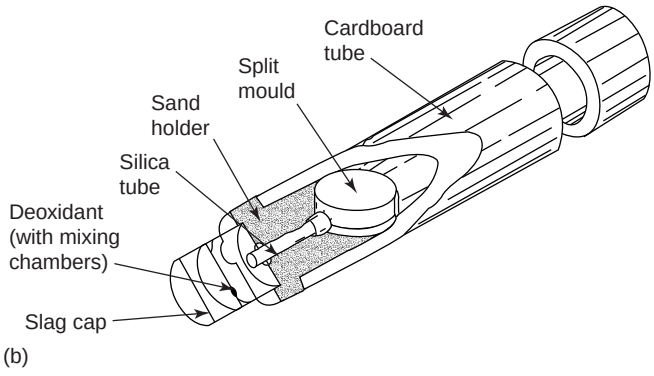

Figure 2 Examples of immersion probes filled by ferrostatic pressure: (a) with the deoxidant in the sample chamber and (b) with the deoxidant in a separate mixing chamber. (Reproduced by permission of ASTM (1997).[2])

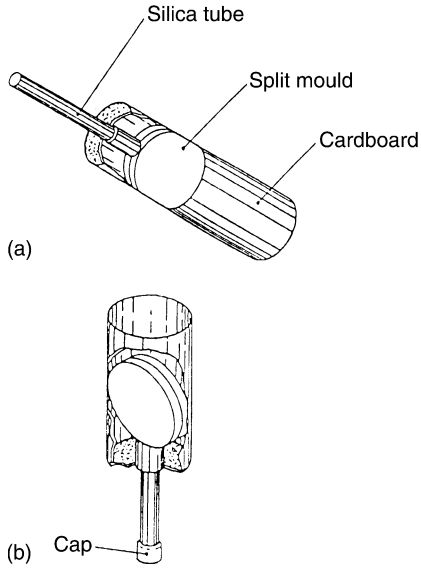

Figure 3 Examples of (a) stream sampling and (b) suction sampling probes. (Reproduced by permission from ASTM (1997).[2])

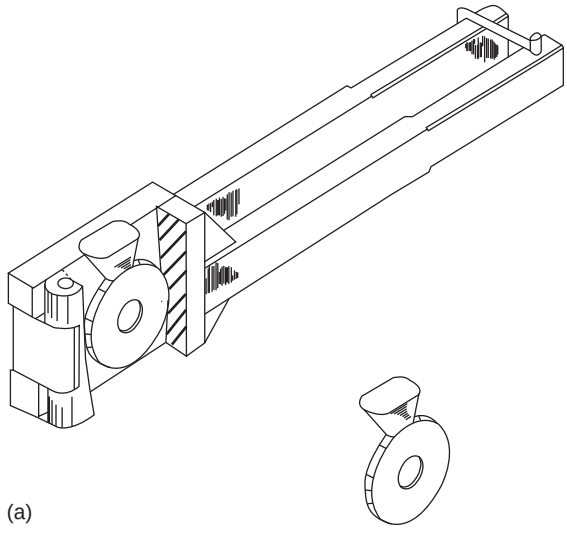

Figure 4 (a) Vertical book mold and specimen. (Reproduced by permission of ASTM (1997).[4]) (b) Vertical book mold. (Reproduced by permission of Angstrom Inc.)

a vertical (Figures 4a and b) or a horizontal position (Figure 5) when cast, depending on the mold type used. They both produce disks, usually 64 mm in diameter and 6–8 mm thick for the vertically cast disks and 6–13 mm thick for the horizontally cast disks. The advantage of the vertically cast disk is that a simpler and lighter mold is used and the 15–20 mm depression in the center makes machining easier and promotes a more uniform solidification. The disadvantage is that the area of excitation for the spectroscopic analysis is more limited. Generally, the two areas 90° from the sprue on the surface of the disk are to be used for the excitation. In the same way as the horizontally cast disk, the entire annular area can be used for excitation. The same size center depression in the horizontal mold also facilitates machining and promotes more uniform freezing. The disadvantage is that this mold is considerably heavier (3.5–4.5 kg). ASTM Standard Practice E 716[3] gives details of the sampling and sample

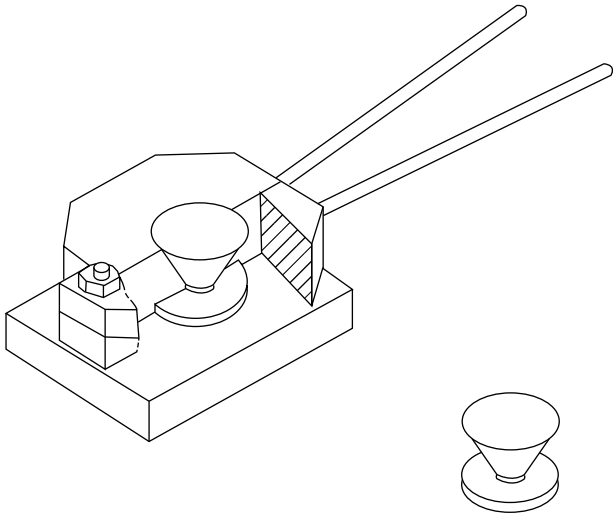

Figure 5 Horizontal mold and specimen. (Reproduced by permission of ASTM (1997).[4])

preparation of aluminum for spectrochemical analysis. For chemical analysis, samples can be obtained from the above described disks by taking drillings at areas that are also suitable for excitation.

2.1.3 Copper and Copper Alloys

Copper and copper alloys are generally sampled from the actual castings produced, or a disk sample is cast at the time the castings are made.

2.1.4 Zinc and Zinc Alloys

Zinc and zinc alloys are cast into the same types of mold as aluminum and aluminum alloys. ASTM E-634[4] specifies basically the same horizontal and vertical molds (see Figures 4a and b and Figure 5) as ASTM E 716.[3] The procedures are also identical (see section 2.1.2).

2.2 Ingots, Billets, and Large Castings

The main difficulty in sampling these items is their large size. Most of the sampling will have to be performed with hand or portable tools. The other problem is that owing to their size the cooling (solidification) of the metal is gradual from the outside of the piece of the metal where it is in contact with the mold towards the inside. As this is generally a relatively slow process, a certain amount of ''zone refining'' takes place. The impurities or even some of the alloying elements tend to migrate towards the center, where the metal stays liquid for the longest time. The general rule is to take the sample at half radius. In most instances the sample is obtained by taking drillings from the large solids. Usually 12–25-mm diameter spade drills or twist drills are used, with the larger ones being preferred. For harder metals they should be carbide tipped. If the ingot or billet is saw cut into sections, it is an easy matter to take the sample from the half radius. However, if it is torch cut, the heat affected area caused by torch cutting should be avoided. Either the drilling should be sited in another area or the drilling taken from the first 4–5 cm should be discarded. Another possible drilling area is at the bottom of the pieces at the half radius, but, again, the drillings taken from the first 2–3 cm should be discarded. If the drilling is done from the side, the drillings should be discarded until the half radius is reached, making sample-taking quite a tedious job. It is always a good idea to check with those responsible for production before taking a sample, so that sampling location does not cause any great loss of material.

There are instances where drillings do not provide the best samples and small solids are preferable. Sampling of reactive metals (titanium, zirconium, hafnium) for the determination of gases is a good example. The best way to obtain these samples is to trepan the ingot with a hollow core drill specially made for this purpose. The metals should be appropriately cooled during the trepanning operation to prevent additional pick-up of gases, or loss of some gases in other metals. The loss of hydrogen in steel is a good example and, in that case, packing the area with dry ice together with a slow rate of drilling would be a good way to minimize this loss. The pins can be removed by breaking them off with the help of a screwdriver or a cold chisel. Hollow core drills resulting in a 6 mm diameter pin are most commonly used. In many instances these pins can also be used for spectroscopic analysis by OES, although a special sample holder might be required.

Metal ingots produced by consumable arc melting, plasma arc melting, electron beam melting or electroslag melting are generally not in a molten state in their entirety at the same time. This generally results in minor variations in the chemical composition along the length of the ingot. Major variations can even occur depending on the feedstock used. Metals produced by these processes should be sampled and analyzed in at least three places along the length of the ingot at the top, center, and bottom.

2.3 Mill Products

Extrusions, (such as I-beams, angle irons, small bars, and wire), rolled products, (such as bar products and sheet metal products) and tubing, (seamless or welded) all fall into this category. They may have various finishes, for instance, as-rolled or as-extruded, pickled, ground, polished, and so on. The lot of material may

come from one heating, or an unknown number of heatings and this information may or may not be available to the sampler, depending on who does the sampling.

Most mill products can be sampled in one of two ways. Either a small piece can be cut off the product from the end or drillings can be taken from the end. ASTM Standard Practice E 1806[2] gives sampling positions for steel sections (see Figure 6). However, these locations can also be applied to other metals, since nonferrous sampling specifications generally are much less specific. A small slice cut off the end of a bar, extrusion, and so on, may be the preferred way, since in most cases, the slices lend themselves directly to spectroscopic analysis (particularly OES) after appropriate surface preparation. Special sample holders may be required. Many of the slices are also suitable for XRF analysis after appropriate machining is done to make them fit the sample holders. For slices that have small cross-sections, a longitudinal section will quite often be suitable.

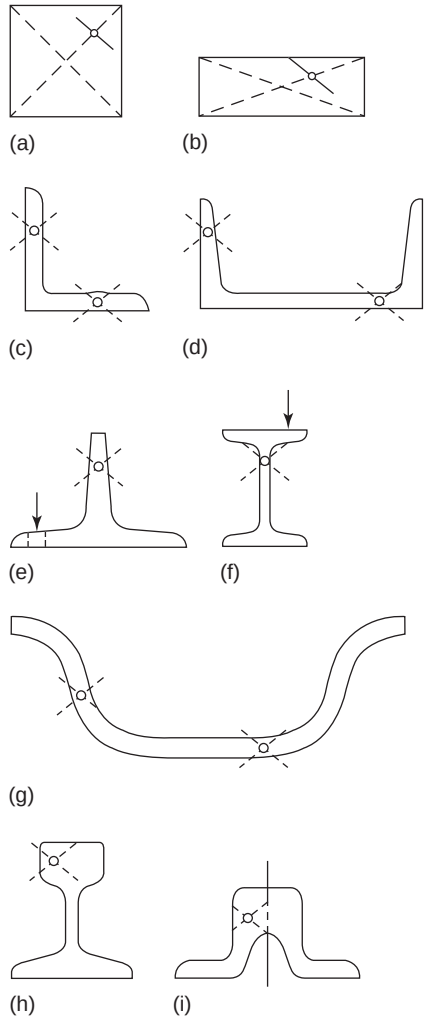

Figure 6 Sampling positions for steel sections: (a) billet or round, (b) slab, (c) angle, (d) channel, (e) tee, (f) and (g) beam, and (h) and (j) rail. (Reproduced by permission of ASTM (1997).[2])

2.4 Particulate Materials, Ores, and Minerals

Metal powders, shot, and uncompacted metal sponge, all fall into this category, as do friable metals, particularly ferroalloys that are highly friable and can readily be crushed (comminuted) to a small particle size. Ores, minerals, refractories, and slags may come in many sized chunks, but the large majority can readily be comminuted into small particles.

Some of these particulates may be fairly homogeneous while others, such as some ores and minerals, may be quite heterogeneous. The finer these particulates are crushed, the easier they are to mix into a homogeneous sample. However, this is only true to a certain limit. If the particle size is reduced below a certain point, the components separate. This sounds like an ideal situation as blending will then provide a homogeneous sample. However, when the constituents separate, particles of a different size, density, and shape are also obtained and these particles have a strong tendency to separate. For iron ores, ASTM Standard Practice E 877[5] gives great detail on calculating the minimum sample size based on the nominal size and the specific gravity of the ore.

For ferroalloys, ASTM Standard Practice E-32[6] deals with the sampling of ferroalloys and steel additives. Basically, these are available either in lumps over 50 mm in size or in the form of crushed alloy (below 50 mm in size). The method of sampling depends on whether the material is shipped in bulk or in containers and on the type of the ferroalloy. More expensive ferroalloys, such as ferrovanadium, ferromolybdenum, ferrotungsten, ferrocolumbium, ferrotitanium, ferrozirconium, and ferroboron are generally treated differently from the cheaper ones, such as ferrosilicon, ferromanganese, ferrophosphorus, silicomanganese, and 12–15% zirconium alloy.

For the lower-priced group, one lump is taken for each 270 kg sample of bulk material, generally, while loading or unloading materials or, if the alloy is in containers, one container in five is dumped and one lump is taken for each 55 kg of material that was dumped (equal to one lump for each 270 kg of the total shipment). This is the gross sample. The gross sample is then crushed in a jaw crusher (see Figure 7) to pass a 6.5 mm screen and riffled to not less than 9 kg. This sample is then reduced to about 2500 g and reduced in size on a roll crusher to pass a 2 mm screen. From this 2-mm sample approximately 170–220 g is then pulverized on a ring pulverizer (see Figure 8) or hardened steel mortar and pestle to pass a # 100 (150 μm) sieve. The sample is dried at 105 °C for an hour, and

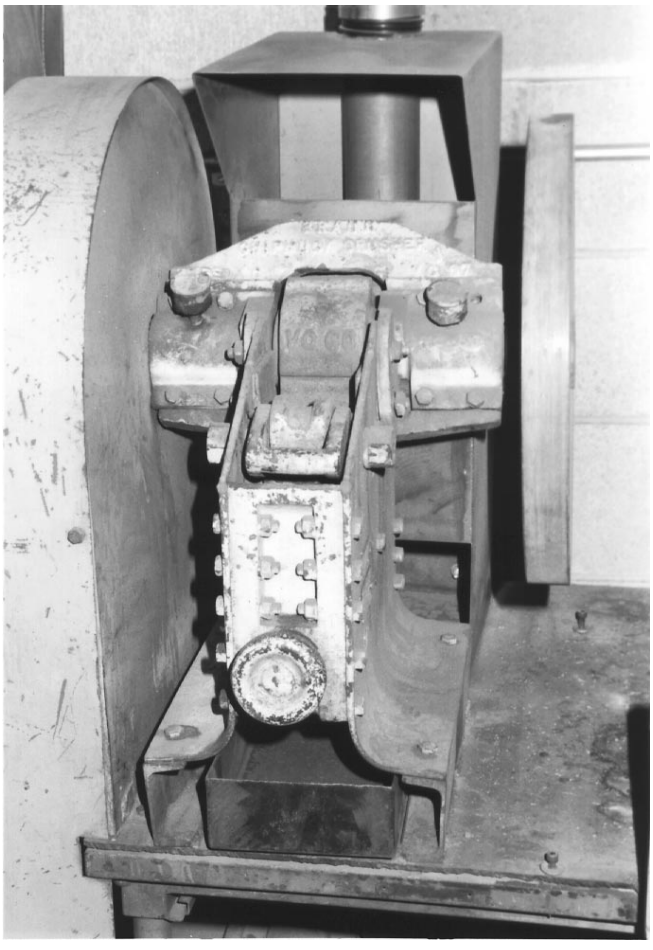

Figure 7 Jaw crusher.

Figure 8 Ring pulverizer.

divided by quartering if necessary. If the lot consists of crushed alloy, a 0.5% sample is taken during loading or unloading. This gross sample is crushed and pulverized as described above.

For the higher-priced ferroalloys, the procedure is basically the same except that all containers are emptied to form a cone-shaped pile. This pile is sampled by shoveling the pile of material. One shovelful of material from every three or four (depending on the size of the lot) is reserved as the gross sample. The sample is crushed as described above, except that larger portions are crushed during each step depending on the size of the lot. This is quite an elaborate procedure and is generally used for umpire analysis. For routine sampling, smaller samples are generally taken and not all containers are emptied. This procedure is a good example of a sampling process.

Reduction in sample size is mentioned a number of times. There are several simple ways to accomplish this task.

The first and simplest way may be by dumping the sample into a pile and transferring it into another pile by shoveling and placing every second, third, or fourth and so on shovelful into a sample container. The actual number will depend on the factor that the sample is to be reduced by. It should be noted that the more the sample is reduced by this method, the less accurate or representative it may become, depending on the heterogeneity of the material.

The second or ''quartering'' method is widely accepted. Here, again, the sample material is dumped into one pile and divided into four pie-shaped sections with a shovel (or a spatula, if the quantity is small and the quartering takes place on a benchtop). The four sections will be arbitrarily referred to as sections A, B, C, and D. Sections A and D are discarded (replaced in the original material) and sections B and C are blended together. This procedure is repeated until the blended sections B and C are of the size desired (see Figure 9).

The third method commonly used is the ''splitting'' or ''riffling'' method. This requires a sample splitter or riffle, a simple piece of apparatus (see Figure 10) also referred to as a Jones sampler, a Jones divider, a coal sampler, and so on. This unit basically consists of a large rectangular funnel that is divided into about a dozen narrow chutes. Each even chute conducts the material dumped on it to the left side of the splitter into a pan placed there, and each odd numbered chute dumps material to the right side of the splitter into a pan placed on that side of the chute. The chutes can normally be 10–50 mm wide, depending on the material that is being split. The sample that is to be reduced in size or split is dumped on the funnel with a special flat shovel that looks like a dustpan and is the full width of the sampler. Half of the sample will fall into each of the two sample pans. The contents of one of the two sample pans are discarded (replaced into the original material or reserved for potential resampling), while the contents of the other pan are dumped on the special flat shovel and split again. This is repeated until the desired sample size is reached.

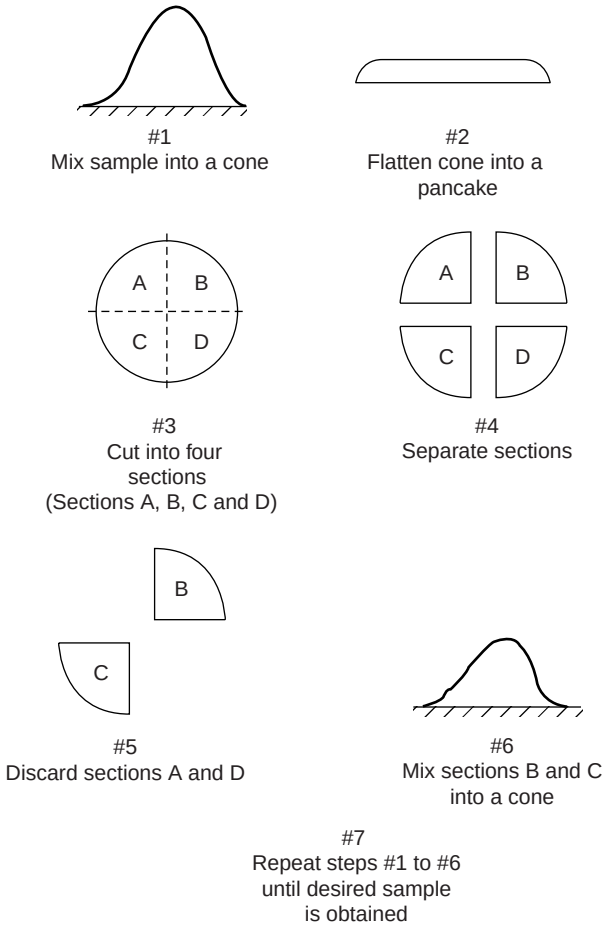

Figure 9 The quartering process.

Figure 10 Sample splitter.

2.5 Raw Materials

Raw materials can be defined as the material from which a product is made. In the metal industry the product is metal. The materials that go into the metal will depend on how integrated the particular producer is. A steel maker can start with iron ore, limestone, coal and so on and end up with pig iron. The pig iron can be converted into steel, in which case the additional raw materials may be alloying elements such as chromium, molybdenum, nickel (in elemental or ferroalloy form), and deoxidizers such as aluminum or titanium. Therefore, it can be seen that the raw materials of one producer may be the finished product of another producer. For this reason most raw materials used by the metal industry have already been discussed or will be discussed in other sections of this article. Ores, ferroalloys, and minerals are discussed in section 2.4. The alloying and deoxidizing metals, depending on what form they are used in, can be found in section 2.7. or if a ferroalloy is used, it is discussed in section 2.4. If new metal is used as an additive, it will probably fall into the category discussed in section 2.3 and sampling should be performed accordingly.

2.6 Finished Products

For the purpose of this discussion *finished products* will be considered to be metal parts products that have been produced for a specific application, such as a paper clip, screwdriver, chain, water faucet, turbine blade for a jet engine, or a fire extinguisher body. Generally these parts have to be analyzed either as a step in the final inspection (quality control), or as part of the incoming parts inspection program by the customer, or if problems arise with the part, in which case there may be litigation involved or at least contemplated. Fasteners, although they belong to this category, are regulated in the United States by federal law and have to be considered separately (see section 2.8).

For the purposes of the discussion in this chapter all nonmetallic parts, if any, will be disregarded. The sampling used has to depend on several factors:

- the reason for the inspection;
- the nature of the part, how critical the application is; how serious the consequences of failure of the part or product would be (with regard to loss of life or limb, property damage, or damage caused by downtime);
- the complexity of the specification;
- the past history.

If all these factors are considered it becomes evident that the reason for the inspection is a significant one, as the person performing a manufacturer's final outgoing inspection has considerably more knowledge about the product than the person who performs the incoming inspection at the purchaser's facility. The person performing the manufacturer's final inspection has or should have knowledge about factors such as first, was the entire

lot product made from the same coil of material or the same heating of metal? Second, was the incoming raw material (metal) checked, and did it meet all specifications? Could any other parts that look similar have been mixed into the lot? Third, did the material exhibit any difficulty or unusual behavior during the manufacturing? First-hand knowledge of all of these factors can influence the size of the sample taken. This gives the advantage to the sampler performing the manufacturer's final inspection. On the other hand, the sampler performing the customer's inspection does not have this information and must assume the worst. The only information that may be available is the past history of the shipment made by the producer of the parts and whether the supplier has been used long enough to build up a reliable history. Sampling that is performed owing to a problem such as product failure and where there is or could be litigation involved, will be addressed in section 2.8.

If the second factor, the product itself, is considered, it is evident that, for example, a turbine blade that goes into an engine of a jumbo jet requires tighter quality control than a paper clip or a screwdriver. This translates into the need for a larger and more representative sample of the turbine blades.

Most of the items that fall into the category of finished products discussed here can be sampled in the same way as those in the solid mill products category (see section 2.3).

2.7 Scrap Metals and Metallic By-products

Sampling scrap metals and metallic by-products can represent special problems. First of all, many if not most of the scrap metal products are unique in the sense that they are only available and marketed as scrap items. Good examples are metal turnings, millings, drillings, or broachings that are generated as a by-product of their respective machining operations. Scrap metal is sold mainly to melters to lower the cost of the metal produced. Generally, normal sampling procedures, rules, and customs do not necessarily apply to scrap metal. On the other hand, no generally accepted published sampling standards for scrap metal sampling are available. Most of the time, scrap metal sampling is not only a science but also an art. A good scrap metal sampler has to have many years of experience, must be familiar with the scrap metal to be sampled, its history and origin, and has to be familiar with the intended end-use of the metal produced from the particular scrap metal to be able to do a good job. Regretfully, little, if any reliable literature is available on this subject.

There are many types of scrap metal, turnings and millings, sheet metal scrap, small solids, mill scrap, fine residue type scrap, assemblies, baled and briquetted scrap, and "in-house" scrap. These types must be discussed separately.

2.7.1 Turnings and Millings

These originate from the machining operations of various metals. The turnings are drop-offs from a lathe or turning operation. The millings come from a milling operation. Most other machining operation scrap, such as boring from a boring mill operation, broachings from a broaching operation, drilling from a drill press, to name the more important ones, also fall into this category. Turnings represent probably the largest volume and can give the sampler the most problems. They are collected from the lathe operations in industrial plants and can represent any metal imaginable. Owing to their curly nature they are very tangled if purchased directly from the scrap generator (the manufacturing plant whose equipment generates the scrap) and do not lend themselves to normal sampling operations such as quartering, riffling, or just removing a handful. Turnings are generally a mixture of fine turnings from finishing operations. The larger heavier turnings are from first cut or roughing operations. There are two methods of sampling these turnings.

In the first method the entire lot of turnings can be crushed through a hammer mill (see Figure 11) or a ring crusher, and a sampling pipe is set in the stream coming from the output conveyor of the crusher, adjusted to collect about 0.1–0.2%. Turnings shipped by a scrap metal processor are generally already crushed.

The second method is faster and requires less equipment. However, it requires a more experienced sampler than the first method and is less accurate. The sampler has to estimate the proportion of the different types and sizes of turnings in the lot. He or she then hand picks a sample (possibly with the help of tin-snips) in approximately the same proportions as the lot, where the gross sample is

Figure 11 Hammer mill.

about 10–25 kg for a 5000 kg lot, passes this through a laboratory size hammer mill (run by a 3–4 kilowatt motor), and considers this to be the gross sample.

Warning: Some metal turnings such as magnesium, zirconium, hafnium, and titanium turnings, are highly flammable and can easily catch fire when crushed. Light magnesium turnings are not even sold as they are too flammable. Turnings are generally disposed of at the generation site by burning them under controlled conditions. However, the reactive metals can be crushed if proper precautions are taken:

- The crusher should be set up in an area well removed from all flammable materials, preferably out-of-doors.
- Crushing should always be performed under a constant and copious flow of water.
- The rate of crushing should be slow and even.
- Crushed turnings should not be left to accumulate at the output of the crusher, but removed continuously at frequent intervals.
- Zirconium and hafnium should only be crushed in small quantities in a small crusher.
- Appropriate fire-fighting equipment, in sufficient quantity, should always be available at the site of the operation.

There is one additional problem concerned with turnings and millings that has to be considered. When the turnings, millings, and so on, are generated, the machining generally is performed under a constant stream of cutting fluid to prevent overheating of the metal and/or to prolong the tool life. By the very nature of the cutting fluid, it adheres to the metals and much of it is retained in the shipment. Therefore, the bulk density of crushed turnings can be as low as 80–100 kg m^{-3}. This may not be a negligible item, since when the buyer negotiates the price he or she pays for the metal. In some instances, a light metal such as light-cut titanium turnings or millings can contain as much as 40–50% by weight of water and cutting oils. This can practically double the price of the metal, if the oil and moisture (O&M) is not taken into consideration. The O&M contents can rapidly change. They can decrease if left outside in the sun on a dry warm day, or increase when exposed to a heavy rain. To avoid this problem, a separate sample has to be taken for O&M determination, preferably at the time the weight of the incoming material is taken. The O&M content and its timely determination is a serious and recognized problem throughout the scrap metal industry, and no accepted standard or method is available for items that have over 5% O&M. A 2–5% by weight of allowance for O&M contents is often made for scrap metal turning shipments as a general, often unwritten, agreement unless processed (cleaned or degreased) crushed turnings (often referred to as chips) are traded. A frequently used procedure consists of dumping a few randomly selected drums on a large metal plate and quartering down the combined contents of the drums with a shovel to obtain a gross sample of 2–5 kg. This should be performed just after the lot is weighed. On this gross sample a combined O&M determination should be performed immediately by washing a large (approx. 100–500 g) test sample with an appropriate solvent or a detergent and drying the sample at a low temperature to a constant weight. If the lot is not crushed, the best way to handle the sampling is to crush the lot, weigh it immediately as it appears from under the crusher and simultaneously take a sample. An O&M determination is run on the sample immediately. The correction obtained is applied to the weight obtained after crushing, resulting in the weight of the dry metal. This method permits crushing with a flow of water and does not affect the weight of the dry metal.

2.7.2 Sheet Metal Scrap

Most of the sheet metal scrap consists of drop-off pieces from various manufacturing operations. They can be drop-offs from various shearing operations, or they can be the drop-offs from stamping operations. They can include rectangular cut pieces of various sizes, full size sheets, or ribbons of sheet metal with identical odd-shaped parts punched out of them and often referred to as skeleton scrap. Much of the sheet metal scrap that has been used in the aerospace industry (one of the largest generators of sheet metal scrap) is die marked with the grade of the metal, often stating the applicable ASTM or Aeronautical Material Specification (AMS) specification numbers. This can be a great help in evaluating the metal for uniformity.

The best way to take the actual sample is to remove small pieces from the sheet metal by shearing, nibbling, or clipping them with a pair of tin snips. If the lot to be sampled consists of several distinctive types, proportions should be estimated and the pieces sampled should be within those proportions. If the sample looks uniform to a person familiar with that type of scrap, a general sample should be taken from all areas of the lot. For best results the clippings, nibblings, or shearings should be melted into one solid homogeneous sample.

In some instances, the entire lot of sheet scrap is passed through a hammer mill and shredded into small, often postage stamp size, pieces. With this method much better sampling can be performed, particularly at the output conveyor of the hammer mill.

2.7.3 Small Solids

This can be either a very simple matter or a rather complex issue. Many small solids may consist of a lot of uniformly used jet engine turbine blades or small castings of identical parts with casting defects. These examples represent the simple cases. In these instances a simple plant scrap inspection method, as described in ASTM E 1916,[7] such as inspection with a metal sorting spectroscope, a thermoelectric comparator, or a chemical-spot check can verify the uniformity of the material, so that a random sampling of the pieces will result in a good sample.

If the pieces are heterogeneous, the situation can be much more complex. For example, the lot can consist of used drill bits. A visual inspection may indicate that all or at least most of them are made of high-speed tool steel based on the HS marking on most of the drills. However, preliminary tests may indicate that they are a mixture of T-1 and M-2 grade steel, and the size or type of the drills is no indication of the steel grade. This leaves the sampler with two alternatives. Either the entire lot is segregated by an experienced scrap sorter using a grinding wheel and a simple spark test, and then a random sample is taken or a large random sample is taken and that sample is evaluated. The second method is obviously faster and less costly. However, depending on the results of the test, the product may still have to be segregated to make it acceptable for the melter. Another example of a more complex sampling problem is when a lot of material consists of used miscellaneous turbine blades, known to consist of only nickel and/or cobalt base alloy or alloys. A preliminary examination of the blades indicates that the lot consists of two nickel base alloys and one cobalt base alloy containing considerable amounts of tungsten (wolfram). Although the cobalt alloy blades are slightly different in shape, it is not possible to estimate the percentage of each alloy or to get a representative sample. Therefore, there again are two alternatives. First, using a metal spectroscope, an experienced scrap sorter can segregate as many as 5 to 12 pieces per minute. It may be more economically advantageous to segregate the entire lot by 100% sorting, than to take a random sample of the segregated portions to be melted and completely analyzed. In the alternative case, only one large sample is taken and the lot evaluation is based on this sample. This will give an average composition that can be of use only if the material is used as a refinery base, a relatively low value material. In other words, an evaluation has to be made to determine if any upgrading of the material should be done prior to sampling.

2.7.4 Mill Scrap

Mill scrap are larger pieces of scrap generated by the mill or primary metal producer, such as bar and/or billet crop ends, hot tops and so on. Generally these pieces of metal can weigh from 5 kg to 200–300 kg. Usually this kind of scrap, unless its history is reliable and known completely, is evaluated piece by piece using scrap sorting techniques previously discussed. A final sample can be taken by compositing drillings or millings taken from each larger individual piece, together with a random sampling of the smaller pieces. If different alloys are found by the sorting, each one is kept separate.

2.7.5 Fine Residue Scrap

The materials that fall into this category are grinding residues, bag-house dust from melting operations, and residues from pickling tanks. Practically all of these materials are intended for recovery of one or more metals. Generally, they do not represent major sampling problems except that, frequently, they are mixed with cutting and grinding compounds, water, oils and so on. The sampler has to be alert, as these materials frequently contain hazardous substances such as lead. These occur most frequently in bag-house dust. Appropriate safety equipment should be used when sampling such materials.

2.7.6 Assemblies

These are new or used manufactured products or parts that consist of more than one piece. They are fastened together by welding, glue, the use of fasteners, or by other mechanical means. Some assemblies contain only metallic parts made from the same alloy. Some can be made from an assortment of alloys. In other instances there are also nonmetallic parts such as insulators, O-rings, hoses, or plastic parts that may be included in an assembly. Assemblies can range from small electronic gold-plated circuit boards with at least some of the electronic components attached, to complete (or nearly complete) 2–3 m tall jet aircraft engines. Seldom do such assemblies get much further in the metal recycling chain than a scrap processor who specializes in such assemblies.

Such assemblies, particularly the more complex ones, cannot truly be sampled. Purchases of such assemblies are made on either a past experience basis, or on a purchase gamble by making a very low bid. This is particularly true if the lot consists of several assemblies that are all different. To obtain a good evaluation, the entire assembly has to be disassembled and the obtained components identified. To get a true picture, the disassembled components that have a value have to be sampled and evaluated (tested) individually. Particular attention has to be paid to contaminants that cannot easily be removed, for example, a small hose fitting that is silver-soldered to a nickel base alloy part.

2.7.7 Baled and Briquetted Scrap

Because many forms of scrap metals are bent, twisted, and curled to such an extent that the bulk density is considerably reduced, shipping and handling are very expensive. Consequently, many scrap processors compress such products in large presses. Sheet metal scrap is frequently compressed under considerable pressure into $30 \times 30 \times 45$ cm bundles (the size can vary to a considerable extent), which are usually referred to as bundles, bales, or briquettes. Also, smaller often particulate-type scrap such as crushed turnings, millings and so on, is compressed into a shape resembling a large hockey puck. It can have a diameter normally ranging from 10–20 cm and can be about 7–10 cm thick. Such items are normally referred to as briquettes or sometimes as pucks.

These and sheet metal bundles in particular can represent a major sampling problem. The pucks made from turnings can generally be pried apart with a screwdriver and composites of samples taken from several pried-apart pucks resulting in a gross sample. However, compressed sheet metal bundles made of heavy gage sheet metal cannot be pried apart by ordinary means. Since the sheet metal is highly tangled in the bundles owing to the compression, it takes much more force to pull it apart than it took to make the bundles. The ideal solution is to take some random bundles, separately melt them and take a sample from the melt. This, however, in most facilities remains an unattainable utopia. Most of the time, taking clippings from the pieces that can be reached from the outside of the bundles is the best that can be done. Even this method is not simple if the sheet metal is heavy gage.

2.7.8 In-house Scrap

Scrap generated by the user is normally referred to as in-house scrap. If the user has a good housekeeping system and all scrap items are properly tagged and stored, many items only need some rapid alloy verification, if that, and do not need sampling. Other items that could become contaminated, such as machine shop turnings or broachings, can have a simpler sampling procedure performed during the pre-use processing. Results should be as expected, unless there is some breakdown in the system.

2.8 Special Considerations

There are several circumstances in which normal sampling practices do not apply. Some of these circumstances briefly discussed in this section are sampling and testing involved in shipments covered by buy–sell agreements or purchase orders, sampling and testing of certain fasteners, and cases involving litigation or potential litigation. These will be discussed here. However, laws and/or regulations change, so if any questions arise in these cases, for any other reason, the sampler and/or their company is strongly advised to seek competent legal advice.

2.8.1 Shipments Covered by Buy–Sell Agreements or Purchase Orders

When sampling is performed on a lot of materials that is covered by a buy–sell agreement or a purchase order, these documents sometimes specify testing and/or sampling procedures. In this case, the sampling and testing has to be performed as directed in the contract instead of in the customary way even if the customary way is an accepted sampling procedure. This may result in conflicts. For example, a lot of fasteners has to be sampled and tested in the United States according to a purchase order. The first thing to do is to establish if the fasteners involved are covered by the Fastener Quality Act (FQA) of 1990 (as modified),[8] or not. If the fasteners are covered by the Act, any provisions of the purchase order specifying sampling or testing (including who shall perform the sampling and testing) are superceded by the law. Competent legal advice should be sought since the contract generally does not override the law. Another confusing situation that can arise is that the standard or specification cited in the contract does not contain the information called for. This is another instance that has to be resolved through commercial negotiations or legal channels prior to performing the sampling and testing.

2.8.2 Fasteners

The sampling and testing of critical end-use fasteners produced, sold, or used in the United States are regulated by the FQA of 1990.[8] Basically, most metallic screws, nuts, bolts, and studs which have internal or external threads or load indicating washers with a nominal diameter of over 5 mm or 1/4 inch are covered under this law, if made to a standard specification which requires thorough hardening. In most cases, the inspection and testing, including the sampling on these fasteners has to be governed by the standards and specifications to which the manufacturer states that the fasteners have been manufactured. Testing laboratories have to be specially accredited by the National Institute of Standards and Technology (NIST), or NIST authorized accreditation bodies. The accreditations are based on International Organization for Standardization (ISO) Guide 25,[9] and involve auditing the laboratory requesting accreditation by the accrediting body.

2.8.3 Cases Involving Litigation or Potential Litigation

If sampling and testing is due to a litigation (or a potential litigation), resulting from a shipment that did not meet, or allegedly did not meet, certain requirements of the purchase order or other agreement, or product failure resulting in monetary damages, injury, or loss of life, all sampling, sample preparation and testing has to be performed with extra care. All samples, potential samples, lots to be sampled or potentially to be sampled shall be immediately secured, preserved, marked, and identified, and, if possible, sealed to prevent tampering. One person should be put in charge of the technical aspect of the case, who should work closely with the company lawyer handling the litigation. Any sampling, sample preparation, and testing should only be performed with this lawyer's approval, especially if the sampling, sample preparation, or testing is destructive in nature! Record keeping in particular should be discussed with the lawyer. No aspects of the case should be discussed with anyone else, including co-workers and superiors from the same company, except with the approval of the company lawyer. It should be understood that the opposing party may also be granted access to the sample(s) and possibly any or all records taken, including all notes and memos!

3 SAMPLE PREPARATION

Sample preparation deals with the procedures that result in a gross sample (or a representative portion of the gross sample) being rendered homogeneous and permitting a good laboratory sample to be taken which then can be prepared so that it is physically suitable for the analytical procedures that are to be performed.

For example, a lot of pure cobalt shot is to be analyzed in which each piece weighs about 2–5 g. The O_2, N_2, C, Al, Cr, Cu, Fe Mn, Ni, S, Bi, and Pb contents have to be determined. The best way, in this case, may be to determine the carbon and the sulfur on a carbon–sulfur determinator, the O_2 and N_2 on an oxygen–nitrogen determinator, the Al, Cr, Cu, Mn, Ni, and Fe on an OES and the Bi and Pb, owing to the fact that these two elements have to be determined on a parts per million or even fraction of a parts per million level, on a flameless atomic absorption spectroscopy (FAAS). Note that the purpose of this example is to show the different specimen requirements for different analytical techniques and is not intended as an endorsement of one method over another. For example, inductively coupled plasma spectroscopy (ICP) could be used instead of OES, or ICP mass spectroscopy (ICP/MS) instead of FAAS.

OES requires a solid piece of metal about 25–37 mm in diameter, with a flat surfaced bottom. The oxygen–nitrogen determinator requires about 0.5–1-g solid metal sample that fits into the small graphite crucible. The carbon–sulfur determinator requires a 1-g solid sample, although 1 g of chips may be used. FAAS requires a fairly concentrated solution of the metal. To prepare this solution, about 1 g of chips is required. If the gross sample has a weight of about 6–7 kg, it can probably be reduced by proper splitting techniques to one-half or even one-quarter of its weight, but below that, it could be questioned whether or not the test sample is representative of the lot. Furthermore, the sample would still be unsuitable for most analytical methods. The best choice is to melt a representative sample from the gross sample. The most suitable way of melting this material is in a vacuum induction furnace and because the Bi and Pb must also be determined at trace levels, the furnace has to be backfilled with an inert gas such as argon. The molten sample has to be cast into an appropriate mold. After cooling, the sample is removed from the furnace and the mold. It is then cut into slices with a slow-speed band saw under constant cooling to avoid heating the sample. One slice is surfaced on a belt sander and, will be the laboratory sample for OES. One thin slice is chopped up on a shear to produce small pieces for the determination of carbon, sulfur, oxygen, and nitrogen. Chips are then removed from a slice and carefully collected on a mill or drill press. These chips are dissolved in acid (see Table 1), diluted to volume, and used for FAAS.

The sample preparation can be a long and rather complicated procedure. It can involve much more time than the actual analysis itself and may require considerable capital investment. The steps will be discussed in the order that they are performed:

1. sample comminution
2. sample melting
3. mechanical operations
4. dissolution and fusion.

3.1 Sample Comminution

When discussing comminution of the sample, it is understood to mean the reduction in the size of the individual particles. The gross sample frequently consists of large pieces or particles. This makes it impossible to reduce the size of the gross sample to a laboratory size sample and still maintain it as a representative sample of the lot. It should be remembered that the laboratory sample is frequently smaller than a single particle of the gross sample. This is particularly true of ferro alloys (or master alloys), minerals, and ores. The answer is to reduce the size of the individual pieces or particles that make up the gross sample. Brittle materials, such as ferrotitanium or silicon metal are best reduced in size with the jaw crusher (see Figure 7). Jaw crushers are used

Table 1 Suggested acid dissolution of pure metal[a]

Element	Volume (mL)	Reagent and procedure
Aluminum	50	1 : 1 HCl : H_2O, plus 1 drop of Hg°
Antimony	18	HBr + 2 mL Br_2
Arsenic	20	1 : 1 HNO_3 : H_2O
Beryllium	50	1 : 1 HCl : H_2O
Bismuth	20	1 : 1 HNO_3 : H_2O
Cadmium	25	1 : 1 HCl : H_2O
Cobalt	20	1 : 1 HNO_3 : H_2O
Copper	20	1 : 1 HNO_3 : H_2O
Gallium	50	1 : 1 HCl : H_2O
Gold	30	3 : 1 HCl : HNO_3
Hafnium	15	HF
Indium	25	1 : 1 HCl : H_2O
Iron	20	1 : 1 HCl : H_2O
Lead	15	HNO_3
Manganese	20	HCl, dropwise HNO_3
Magnesium	30	1 : 10 HCl : H_2O
Mercury	20	1 : 1 HNO_3 : H_2O
Molybdenum	20	HCl, dropwise HNO_3
Nickel	30	1 : 1 HNO_3 : H_2O
Niobium	20	HF, then 5 mL HNO_3, dropwise
Palladium	30	3 : 1 HCl : HNO_3
Platinum	50	3 : 1 HCl : HNO_3
Rhenium	15	1 : 1 HNO_3 : H_2O in ice bath; no heat
Selenium	10	HNO_3
Silver	20	HNO_3
Tantalum	20	HF, then 5 mL HNO_3, dropwise
Tellurium	20	HCl, dropwise HNO_3
Thallium	20	1 : 1 HNO_3 : H_2O
Tin	50	1 : 1 HCl : H_2O
Titanium	20	H_2SO_4, 3 to 5 drops of HNO_3
Tungsten	10	HF, dropwise HNO_3
Zinc	20	HCl
Zirconium	15	HF

[a] One gram test portion is implied. Warm to complete reaction unless otherwise indicated. In most cases alternate dissolutions are possible. (Reproduced by permission of ASTM from Dulski (1996).[10])

extensively in the mineral and ore industry and by the master-alloy producers to reduce brittle lumps in size. They consist of a pair of heavy wear-resistant plates or jaws (see Figure 12) hammering at the material that is put between them until they are sufficiently reduced in size so they are able to pass between the two plates. The sample has to be passed through the jaw crusher several times. The distance between the jaws of the crusher is reduced each time the material is passed through the crusher. Jaw crushers can only be used with brittle materials and are not suitable for malleable materials such as aluminum, iron, and so on, or bulky materials such as machine shop turnings or light sheet metal scrap.

Comminution of malleable soft metals can pose a problem, especially if they are in a large solid form. Often there is no simple solution and the reduction of the gross sample size has to be accomplished without comminution of the gross sample, as was discussed in section 2.7.

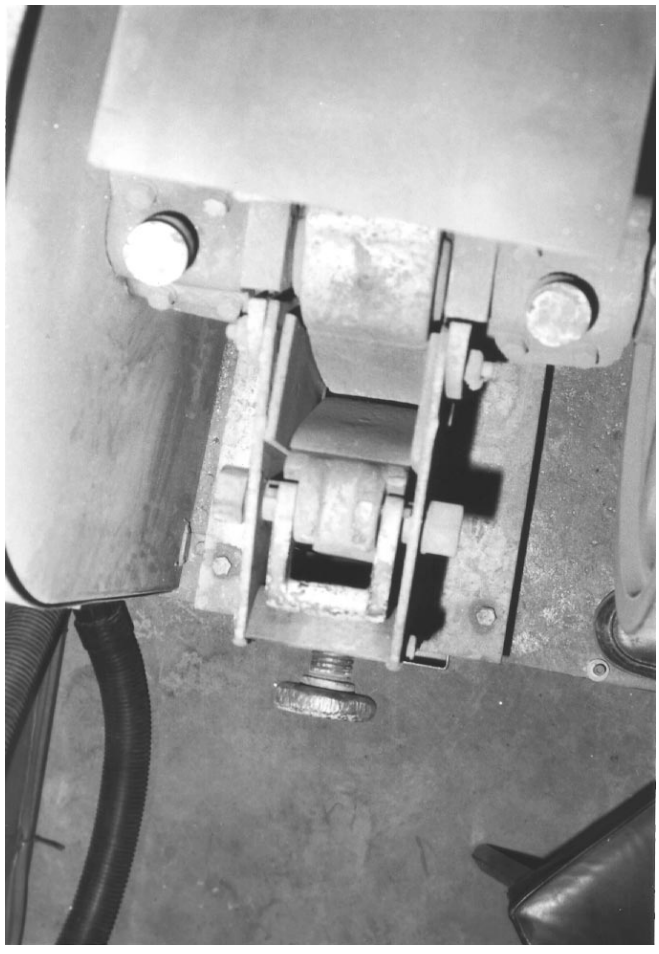

Figure 12 Jaws of jaw crusher (overhead view).

Metal turnings and light sheet metal scrap can present another problem. Owing to their nature, these items can become entangled thus preventing any proper reduction in the sample size by quartering or any other simple method. Normally the best solution is to use a hammer mill (see Figure 2) to comminute samples of this type. Once their particle size is reduced, they will no longer become entangled and the sample size can be reduced by quartering or riffling. A hammer mill is a rotary crusher in which a series of hammers, sometimes also referred to as clappers, crush the material against a set of grate bars until the material can pass through the grates. Such jaw crushers have been used for many decades by the scrap metal industry and are very efficient. Both jaw crushers and hammer mills are available in pilot plant or laboratory size. When crushing samples in a hammer mill that are subject to oxidation, the crushing should always be performed under a copious stream of water. The cooling effect of the water will help to avoid, or at least considerably reduce, oxidation and in flammable metals will reduce the chance of a fire. The crushing of

highly flammable metals such as magnesium should be avoided.

Another way of reducing the size of the sample pieces or particles is by shearing them to size with shears which can range from large alligator shears (frequently used by the scrap industry), to small hand nibblers. This is again a labor-intensive method.

It should be remembered that whenever a sample is passed through a piece of equipment such as a jaw crusher or a hammer mill, the possibility of contamination of the sample is always present. The contamination can generally arise from three sources described below.

The first most frequent source is contamination from materials previously passed through the equipment. After finishing the crushing of one sample, a considerable amount of the material can remain in inaccessible places in the crushers, particularly in hammer mills. It is not sufficient to use an airstream to blow out the contaminants. The equipment has to be opened up and completely cleaned out and inspected after *each* sample. Contamination from this source can be significant, particularly if the material processed in the equipment is of a different type than the next sample. When running more than one type of material through a piece of equipment, it may be advisable after type X material is finished and the equipment properly cleaned, to run a portion of the sample of type Y material through the equipment as a wash heat. Again properly clean the equipment and then process the actual sample of the type Y materials.

The second source of contamination can result from the crusher itself because of metal wear if proper materials are not used in the construction of the equipment or if the sample is too hard or abrasive. Although they normally are not a significant source of contamination, the critical construction parts of a crusher (the jaws of a jaw crusher and the hammers and grates or grate bars in a hammer mill) have to be made from the proper alloy.

Crushing operations generate a considerable amount of heat. Certain metals, such as reactive metals (titanium, zirconium and hafnium in particular) can pick up considerable amounts of gases owing to the heat. This third source of contamination, gas pickup, can be greatly reduced by running considerable amounts of water through the equipment during the crushing operation. In spite of these precautions, gas determinations should not be run on a sample that has been processed in a crusher unless the entire lot of material is similarly processed.

Warning: When crushing flammable metals such as reactive metals, magnesium and so on, a fire in the crusher is always a possibility even if water has been running through the crusher during the crushing operation. The equipment should be set up in an area where there are no flammable materials in the vicinity and where only small amounts of the flammable metal are close by, to minimize the size of the fire if it should occur. Operators should be trained in fire safety and fire-fighting procedures and there should be more than one person present at all times when operating and cleaning a crusher with flammable metals.

3.2 Sample Melting

The melting of a sample has a dual purpose. It renders the sample that is melted homogeneous and, at the same time, it puts it into a form that is suitable for many of the more commonly used analytical procedures such as OES and XRF. It also facilitates collection of the test sample for other analytical procedures. All melting procedures involve raising the temperature of the sample above its melting point, stirring the resulting liquid metal by some means to enhance its homogeneity, and finally cooling it in a suitable mold by a method in which the metal remains homogeneous.

Melting furnaces are classified by the way they generate the heat used to melt the metal. There are three types of furnaces in general use:

1. electric arc furnaces
2. induction furnaces
3. directly heated furnaces – these can be electrically heated or gas fired.

3.2.1 Arc Melting Furnaces

The laboratory-type arc melting furnaces (see Figure 13) consist of a water-cooled copper pot, which contains the sample to be melted. The copper pot and the metal sample which makes direct contact with it, act as the cathode, and a graphite or tungsten electrode is used as the anode. These components are enclosed in a tight enclosure that is provided with connections for vacuum and an inert gas, such as argon. A direct current electrical welder which has the capacity to supply 600 A acts as the power source for these furnaces. ASTM Standard Practices E 1010[11] and E 1306[12] discuss the apparatus and the melting procedures. In these furnaces, the melting pot also acts as the mold. These furnaces produce buttons up to about 36 mm in diameter that can weigh about 35–40 g for a titanium button or about 70–80 g for a steel or nickel alloy button.

These commercially produced laboratory arc melting furnaces are suitable for melting steel, iron, nickel alloys, cobalt alloys, reactive metals, or even smaller buttons of refractory metals. They are not suitable for melting aluminum and copper alloys, or generally low-melting-point alloys.

3.2.1.1 Advantages Arc melting furnaces are relatively low priced and economical to operate. They are the

Figure 13 Arc melting furnace (button furnace). (Reproduced by permission of Scianflone Scientific Instruments Corp.)

only laboratory furnaces that are able to melt titanium and other reactive metal alloys and to a certain degree, refractory metals. Buttons can be melted at a fairly fast rate (3–6 buttons per hour) and only simple maintenance is required. As the copper pots are easy to clean with a rotating wire brush, cross-contamination is kept at a very low level. Operators are relatively easy to train.

3.2.1.2 Disadvantages Owing to the high temperature of the electric arc, certain low-melting-point trace elements may be lost. Low-melting-point alloys, which include aluminum alloys, cannot be melted in these furnaces. The melts produced are relatively small and only what fits into the crucible can be melted as they do not have provision for making additions.

3.2.2 Induction Furnaces

An induction furnace consists of a refractory crucible surrounded by a water-cooled copper coil which induces heat in the metal charge of the crucible when the coil is connected to a high frequency power supply. The entire sample, when completely melted, is poured into a copper or iron ingot mold.

The entire assembly (crucible, coil, mold(s), pouring mechanism etc.) is enclosed in a chamber that can be evacuated and is used as a vacuum furnace or it can be backfilled with an inert gas such as argon and the melting can be done under a partial pressure of inert gas. Sample or alloying additions can be often made from one to several addition cups that are dumped onto a vibrating feeder that can feed the addition to the melt in the crucible at a controlled rate, and always under the control of the outside operator. Provisions for temperature measurement and bridge breakers for breaking up crusted tops can be incorporated as an option into these furnaces. Normally, laboratory-size vacuum induction furnaces (see Figure 14) are supplied with crucibles ranging from 5.5–13.5 kg steel capacity. Many furnaces are provided with a turntable within the chamber to accommodate several molds. The best molds are split molds machined from copper which produce a round ingot with a diameter of about 35 mm and a very large wall thickness in the range 25–35 mm to enhance rapid solidification of the sample.

The power supply frequency may vary depending on the manufacturer. However, it generally is in the 1000–10 000 Hz range for laboratory-size furnaces. The type of sample normally melted has to be considered, as lower frequencies have more penetrating power and are more suitable for melting larger chunks of metal, while higher frequencies have more of a ''skin effect'' and are more suitable for smaller thinner particles such as metal turnings. The coils used have to match the power supply.

Induction furnaces are generally suitable for melting steel, iron, nickel and nickel alloys, cobalt and cobalt

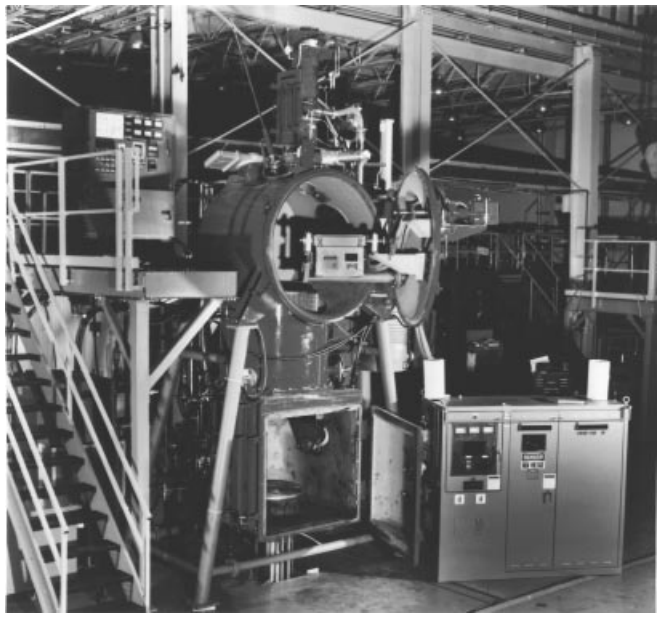

Figure 14 Laboratory vacuum-induction melting furnace. (Reproduced by permission of Consarc Corp.)

alloys, aluminum alloys, copper alloys, rare earth metals, and so on. They are not suitable for melting reactive metals, because molten reactive metals attack the refractories of the crucibles, or for refractory metals, such as tungsten or molybdenum, because their melting temperature is generally too high.

Often if one attempts to melt light low bulk density samples in an induction furnace, there will be a problem for the metal to ''couple'' to the coil. This problem can be overcome by compacting the sample in a press and making small briquettes from the low bulk density metal.

Air-melt induction furnaces are not generally used in laboratories for sample melting, except in foundries or by metal producers who want to simulate their production furnaces.

3.2.2.1 Advantages The biggest advantage of the laboratory vacuum-induction melting furnace is that it is able to melt relatively large samples, which can considerably enhance the quality of the sample. It is also suitable for melting a wide range of metals. If the furnace is under an inert gas, such as an argon atmosphere, the loss of trace level tramp elements in aerospace alloys can be prevented.

3.2.2.2 Disadvantages Reactive metal samples cannot be melted in an induction furnace. A laboratory-size vacuum-induction furnace is a fairly high-priced piece of equipment costing perhaps 10 to 20 times as much as a laboratory arc melting furnace. Melts are slow. It may take 1–2 h or more to produce a sample melt. Depending on the metal to be melted, a crucible may have a lifetime of 3 to 12 melts. Changing a crucible and drying it takes 18–24 h. It is considerably more difficult to clean a crucible properly after the melt in order to prevent cross-contamination from sample to sample. Ideally a separate furnace assembly (the coil, crucible, and the supporting box) should be available for each type of alloy melted to prevent cross-contamination. Changing these furnace assemblies between melts can also be time-consuming depending on the size of the crucible used.

The vacuum system requires extensive upkeep and maintenance to keep it in good operating order. An experienced operator is essential to produce proper uncontaminated samples and keep the equipment in proper condition.

3.2.3 Directly Heated Furnaces

Directly heated furnaces are the simplest of the three types of furnace discussed. They consist either of a crucible (carbon or refractory) heated by an electric heating element encased in a metal container and a lid, not unlike a small electric coffee pot, or of a crucible (carbon or refractory) supported in a refractory enclosure and heated by an air–gas flame. With the first type, which is also frequently used by jewelers, the molten metal is poured into the mold in a similar way as pouring from a coffee pot. With the second type the crucible is normally removed from the furnace with a pair of tongs, and with the help of the tongs, the melt is poured into the mold. These furnaces are used to melt precious metals (gold or silver), and generally low-melting alloys such as lead–tin alloys, aluminum, and so on.

3.2.3.1 Advantages These furnaces are very reasonably priced and require minimal maintenance. They are simple and fast to operate and do not require much special training to operate or maintain, with the exception of safety training. They are commercially available from stock.

3.2.3.2 Disadvantages They only operate as an ''air-melt'' furnace and in some instances fluxing is required to prevent oxidation. They only work with metal alloys that have a low melting point (actual top melting temperature depends on the unit).

3.3 Mechanical Operations

When a button or a small ingot is melted from the sample, or a slice is sawed off from a bar of steel, or a piece is cut off from a metal pipe, the pieces are not yet ready to analyze. They need additional preparation. Most of the preparation involves making sure that the sample piece is the right shape and the correct size and in many instances giving it an appropriate surface preparation. This can be achieved by a proper machine shop operation. An analytical laboratory, specializing in the analysis of metals will need a well equipped machine shop and should be equipped with a band saw, abrasive saw, bench top shear, mill, drill press, lathe, surface grinder (''swinging''-type preferred, see Figure 15), and a belt sander(s). Naturally the equipment depends on the work that the laboratory performs. A commercial testing laboratory specializing in metals in general may need all of the above equipment, while a laboratory serving only an aluminum foundry may only need a lathe and possibly a band saw for OES analysis, and perhaps a mill or a drill press if it also does wet chemical determinations.

It is not the object of this article to go into the operation details of all of the above-mentioned machinery. There are many excellent books written on the subject of machining various metals, such as one edited by Oberg.[13] A fully equipped machine shop should be operated by a duly qualified machinist. Even if only one or two machine tools are required, they should only be operated by personnel fully familiar with the operation of the equipment. This is not only important to maintain quality, but also very important from a safety point of view.

Figure 15 Swinging-type surface grinder.

The proper marking and remarking (should the original marking be destroyed in one of the mechanical operations) of all specimens cannot be emphasized enough. There are several good ways of marking specimens. They can be stamped with letter and number dies, engraved with a vibrating pen, engraved with an electric arc pen, and so on. Less desirable methods of marking samples are felt pens, as markings can be smudged by fingers (particularly oily fingers) or removed by solvents, labels can fall off and heavy specimens can also fall out of marked envelopes.

OES and XRF specimens have two basic requirements. First they must fit the sample holder of XRF spectrometers, or fit onto the arc or spark stand of the optical emission (OE) spectrometers, and second the surface on which the measurements are taken has to be flat and finished with an appropriate abrasive or machined in an appropriate way. If the specimens are melted in a laboratory furnace, the size is generally no problem. In button (arc) melting furnaces the size is normally not a problem, but one side of the button has to be machined flat. This is usually done on a swinging-type surface grinder, using an appropriate cup wheel, such as aluminum oxide. In the absence of this kind of grinder a lathe can be used, although the process is slower. A WC (tungsten carbide) tool bit is preferred. The final finish on the surface is done with a belt sander equipped with an appropriate belt.

When bars are cast in a laboratory furnace, the diameter of the bar is generally of the proper size and the bar only has to be sliced on a band saw. Sometimes an abrasive saw is required for harder metals. In most cases, high-speed tool steel blades are recommended for band saws. One of the surfaces again has to be prepared on a sander. Other round bar samples can be handled in the same way, but in some instances the diameter of the bar has to be reduced. This is generally best done on a lathe. With some solid samples, it may take a combination of machining steps to arrive at a suitable test specimen and the ingenuity of the machinist is put to a true test.

For the final surface preparation of XRF and OES analysis the usual way is to surface the specimens on a belt sander. The most common belts used for this purpose are alumina, silicon carbide, and zirconia. If the standard or procedure used specifies the grit size and abrasive used, it should be followed. If not, ASTM E 1257[14] can be used as a guide.

On disk specimens of softer (nonferrous) metals, such as aluminum, lead–tin alloys, and so on, the surface preparation has to be done with a lathe, as abrasive belts could cause smearing of the metal. When machining these surfaces, the automatic cross-feed on the lathe has to be engaged and very sharp tool bits must be used. It should be noted that the same feed rate has to be used each time on the cross-feed when machining specimens, otherwise the precision of the analyses will be affected. Precision also can be improved by the use of WC tool bits, because their sharpness is more constant and uniform than tools made from hand-ground tool steel.

Samples for wet chemical determinations are generally millings or drillings. They are obtained from solid specimens using a mill or a drill press. Sharp tools (mills or drill bits) have to be used, as dull tools do not cut, just burnish the metal and cause the sample to be contaminated. To prevent the drilling or milling from overheating, a coolant must be used. On vertically cast disk specimens, the drilling or milling should be done after the OES analysis is completed and the drilling should be taken from the same area. On horizontally cast specimens the drilling or millings can be taken at any time. If taken before the OES analysis is done, the final surface preparation should take place after the drillings or millings are taken. The drill press table, mill table, (including any vise, clamps, hold-downs), drill bits, or mills should be cleaned very well before and after each specimen is machined to prevent any

cross-contamination. The drill bits and mills should be inspected after each specimen for any chipping or breakage. A chip from a tool bit can seriously contaminate a specimen and has to be recovered before the drillings or millings are ready for their intended usage. The drillings and/or millings have to be degreased because of the cutting compounds in the coolants used during the machining operation. Normally the degreasing/cleaning is done in a beaker with acetone. The beaker should be marked. The best way to mark it is with a pencil on the spot provided by the manufacturer for that purpose. Marking pens should not be used, as they can run and become completely illegible if the marking comes into contact with the acetone or acetone vapors. The dried chips should be stored in marked glass vials or in marked envelopes.

Sheet metal specimens are often very difficult to drill or mill. Small clippings or nibblings are often used instead of drillings or millings.

Small solids for thermal determinations have to be cut from larger pieces by shearing or by saw. The use of an abrasive saw should be avoided because of excessive heat. For metals that can pick up gases readily, the best solution is to use a manual hacksaw at a slow rate to avoid the pickup of gases. The surfaces should be slowly abraded with a file to remove any oxide layers. Reactive metal specimens should be given a light pickling in 1 : 2 HF : H_2O. For carbon determinations in cast irons, particularly where the graphitic carbon is high, small solids are also recommended.

3.4 Dissolution and Fusion

Many analytical procedures call for a liquid solution sample. They can be instrumental analyses or determinations such as AAS (both flame or flameless), or plasma emission spectroscopy (PES, direct current or inductively coupled) to name those used most frequently now in 1999, and all of the wet chemical procedures. The wet chemical procedures are particularly important, as many of these classical procedures constitute the bulk of the definitive methods. Their significance lies in the fact that they are based on stochiometric chemical reactions and do not depend on the simultaneous use of and comparison to reference materials with known analysis. Even today, if there are no appropriate reference materials available, definitive methods have to be used to perform the required determinations.

As most analytical methods frequently referred to as standards include the sample dissolution and/or fusion part of the sample preparations, this article will only deal with general procedures with respect to the preparation of the solutions.

A solution of a particular metal can be prepared in two ways. It can be dissolved in an appropriate reagent at room temperature or at an elevated temperature. If the reagent is a liquid at room temperature, this operation is considered to be a dissolution and, as this reagent is nearly always an acid, it will be referred to as an *acid dissolution*. If the reagent is a solid at room temperature and has to be melted or fused at an elevated temperature to dissolve the metal, it will be referred to as a *fusion*. Naturally, the fusion itself must be dissolved in water or in a dilute acid to obtain a true aqueous solution. Therefore, the appropriate reagent for dissolution is generally an acid or a mixture of acids with possible additions. Fusions generally use molten salts or bases.

Warning: Before proceeding, it has to be noted that acid dissolution or fusion of metals deal most frequently with dangerous substances and elevated temperatures. Consequently, they should only be performed by people well versed in the technique or under the supervision of a person well versed in the technique, and only in a properly equipped laboratory. Appropriate personal safety equipment should be worn at all times even by experienced people. Particular care should be taken with hydrofluoric acid. Chemically this is a weak acid and contact with it does not cause immediate pain or any sensation. However, a throbbing pain may develop as much as an hour later by which time considerable nerve damage may have occurred.

3.4.1 Acid Dissolution of Metal Samples

The usual way to get metals into solution is by dissolving them in an inorganic acid or acid mixture. Most metals dissolve best in a nonoxidizing acid such as hydrochloric acid. Other metals, copper for example, dissolve best in an oxidizing acid, such as nitric acid. In many cases a mixture of acids is used to dissolve the sample.

Usually the entire sample needs to be dissolved and often special steps must be taken to get all components into solution, such as adding oxidizing agents or a catalyst to enhance the dissolution. A good example is the drop-by-drop addition of nitric acid to enhance the rate of dissolution when stainless steel is dissolved in 1 : 1 hydrochloric acid. A good illustration of catalytic enhancement is displayed when high-purity aluminum is dissolved in hydrochloric acid; the addition of a single drop of mercury will enhance dissolution of the aluminum.

There are procedures, however, that can take advantage of the insolubility of a certain element and enhance the insolubility to an extent that the sought element (or its compound) remains quantitatively insoluble. The determination of silicon in steel by the sulfuric acid or perchloric acid dehydration methods is a good example of this principle.

Naturally, when dissolutions are performed it is important to remember that when the entire sample has to be dissolved, elements can be lost during the

dissolution. Good examples would be (a) the loss of silicon and boron, when a sample is heated in the presence of hydrofluoric acid, (b) the loss of arsenic, when heated in hydrochloric acid without the presence of a considerable amount of nitric acid, and (c) the sudden volatilization of considerable amounts of chromium, when some hydrochloric acid is added to a fuming perchloric acid solution of a chromium-bearing alloy.

Another important factor is the rate of dissolution. Normally it is not practical to wait several days for a particular sample to dissolve. There are several factors that affect the rate of dissolution:

- the reagent mixture for the particular metal, considering not only the base metal (matrix) but also all constituents including possible contaminants;
- the sample shape and form. A particular concern is the surface-to-weight ratio. The larger the surface for a given weight of sample, the larger the surface on which the dissolution action (attack) can take place;
- the dissolution temperature; and
- any possible additions, such as a catalyst.

Let us look at these factors individually.

3.4.1.1 The reagent mixture This may be the most important factor. Most published standard procedures state the method of dissolution, including all of the reagents to be used. This should be followed exactly. It is a very important part of the procedure and if the method of dissolution is not accurately followed, a claim for following the procedure cannot be made, as the method of dissolution can also affect the subsequent steps is the procedure. For procedures that do not specify the method of dissolution Dulski[10] suggests the acid dissolution schemes for pure metals and metal alloys given in Tables 1 and 2.

3.4.1.1 The sample shape and form When acid is poured on a metal sample, the acid can attack the metal only on the surface of the metal. If, for example, a $0.25 \times 0.25 \times 0.25$ cm cube is cut from a larger block of steel, the cube will weigh about 1 g. The surface area, however, is only about 0.38 cm^2. This is a very small area and it will take a very long time to dissolve the sample. If, however, drillings or millings are taken from the same solid sample and a 1 g sample is weighed out from these drillings or millings, the surface area will probably be several square centimeters for the 1 g sample, the actual

Table 2 Suggested acid dissolution of metal alloys[a]

Alloy type	Volume (mL)	Regent and procedure
Aluminum alloys	20	1 : 1 HCl : H_2O, dropwise H_2O_2
Beryllium alloys	30	1 : 1 HCl : H_2O, dropwise HNO_3
Bismuth alloys	20	1 : 1 HNO_3 : H_2O + 5 g tartaric acid
Cobalt-base high-temp. alloys	50	HCl, dropwise HNO_3, low heat
Copper alloys	30	1 : 1 HNO_3 : H_2O
Die steels	50	1 : 1 HCl : H_2O, dropwise H_2O_2
Ferroboron	50	1 : 1 HCl : H_2O, dropwise HNO_3
Ferrochromium, low C	40	1 : 1 H_2SO_4 : H_2O
Ferromanganese	30	HNO_3
Ferromolybdenum	30	HNO_3 + 5 drops HF
Ferroniobium	40	1 : 1 HCl : HF, dropwise HNO_3
Ferrosilicon	15	HF, dropwise HNO_3
Ferrotitanium	30	1 : 1 H_2SO_4 : H_2O, 5 mL HF, dropwise HNO_3
Ferrotungsten	15	HF, then 5 mL HNO_3, dropwise
Ferrovanadium	30	HNO_3, in small portions
Gray iron	25	1 : 1 HNO_3 : H_2O; filter
Iron-base high-temp. alloys	50	HCl, dropwise HNO_3
Lead alloys	20	9 : 1 HBr : Br_2
Low-alloy steels	20	3 : 1 HCl : HNO_3
Manganese alloys	20	3 : 1 HCl : HNO_3
Nickel-base high-temp. alloys	50	HCl, dropwise HNO_3
Silicon steels	25	3 : 1 HCl : HNO_3 + 5 drops HF
Stainless steels	30	1 : 1 HCl : HNO_3
Tin alloys	30	10 : 1 HCl : Br_2
Titanium alloys	100	1 : 1 HCl : H_2O, 3–5 drops HNO_3
Zinc alloys	30	1 : 1 HCl : H_2O, dropwise HNO_3
Zirconium alloys	40	1 : 1 H_2SO_4 : H_2O; 2 mL HF dropwise

[a] One gram test portion is implied. Warm to complete reaction unless otherwise indicated. In most cases alternate dissolutions are possible. (Reproduced by permission of ASTM from Dulski (1996).[10])

surface area depending on how heavy the drilling or milling cuts are. This gives the acids 5 to 20 times the surface area to attack simultaneously, and speeds up the dissolution process considerably. Another advantage of using drillings or millings is the fact that they generally take much less work and are easier to obtain than cutting off a small solid piece.

3.4.1.2 Temperature As in most chemical reactions, the rate of dissolution is enhanced at higher temperatures. Some dissolutions are sufficiently exothermic not to require additional heating. In some instances, the acid actually has to be added in increments to keep the reaction under control. This is particularly true when the metal particles are very fine and/or the surface area is very large. In other instances, the reaction is rather slow and additional heating has to be provided to keep the reaction going. There is also an added benefit of the heating. It provides some stirring of the reagent. When dissolution of a metal takes place there is evolution of a gas, generally hydrogen, nitrous oxides, and so on. These gases evolve at the metal–reagent interface. If there is no stirring involved, the gas bubbles just adhere to the metal surface at the interface and the reaction stops, or at least slows down considerably. The movement generated by the heat transfer from the bottom of the beaker and the increased gas generation keeps the solution sufficiently stirred so that fresh reagent solution permanently contacts the metal. The heat source should always be adjustable and under control. A continuously adjustable electric hot plate is preferred by many chemists. The rate of heating should be just sufficient to keep the reaction going at a good rate, while keeping evaporation of the reagent to a minimum. A watch-glass cover also helps keep the evaporation rate to a minimum in addition to preventing contamination by possible falling debris. In a few instances, elevated heat is necessary to dissolve the sample. This can be accomplished in an autoclave-like dish sometimes referred to as a ''bomb''. It is usually a metallic (stainless steel) container that is lined with an inert plastic material such as Teflon™, which is suitable for the designed temperature, or a container made completely from a heat-resistant inert plastic.

The stainless steel containers are heated in a conventional laboratory oven at a prescribed temperature for an empirically determined time. The containers made entirely from plastic are generally made to be used in specially designed microwave ovens, where the dissolution time can be considerably reduced. Adhering to the manufacturer's given parameters is very important in the use of these methods. It is also very important to remember that the containers should never be opened before they reach room temperature. The pressure in these autoclaves also helps in the retention of volatile elements or compounds.

Figure 16 Laboratory microwave oven. (Reproduced by permission of CEM Corp.)

3.4.1.3 Microwave Ovens When discussing acid dissolution of samples, heating the samples was mentioned several times but only conventional hot plates and ovens were suggested as a source of heat. There is, however, a source of heat that has become popular since the early 1980s, namely laboratory microwave ovens (see Figure 16). These are actually very similar to the household units or those used by restaurants with some modifications. The modifications address the three problems of corrosion, safety, and convenience.

The problem in the corrosion area is obvious. Acids used to dissolve metals and metal-related substances are highly corrosive and so are the acid fumes generated during the dissolution process. Standard microwave ovens, with their electronic circuitry, are very sensitive to such a corrosive environment, particularly, since the electronic components and wiring are not hermetically sealed in the household and restaurant units. Modification consists in selecting the correct materials for construction and the appropriate methods of preventing the corrosive atmosphere from coming into contact with sensitive parts. The fumes can be eliminated or at least considerably reduced by dissolving the samples in sealed containers.

The question of safety is an important issue. Even if the corrosion problems are disregarded, much of the advantage of microwave ovens would be lost if the samples were placed in the ovens in an open or loosely covered vessel (beaker) as it is done on a hot plate. When a microwave oven is used, the samples are generally placed in sealed nonmetallic vessels that are capable of withstanding a limited amount of pressure, permitting a higher dissolution temperature without loss of volatile components. A microwave oven for closed vessels should at least have the capability of monitoring the pressure of each vessel as a safety measure. On many current units,

Table 3 Guide for selecting crucibles

Flux	Porcelain	Vycor[a]	Iron	Nickel	Zirconium	Platinum	Graphite
Na_2CO_3	X	X	X	X	X	E	X
K_2CO_3	X	X	X	X	X	E	X
Li_2CO_3	X	X	X	X	X	E	X
Na_2O_2	X	X	A	G	E	X	X
NaOH	X	X	A	G	E	X	A
KOH	X	X	A	G	E	X	A
$Na_2B_4O_7$	X	X	X	X	X	E	E
$Li_2B_4O_7$	X	X	X	X	X	E	E
H_3BO_3	X	X	X	X	X	E	E
$NaHSO_4$	A	E	X	X	X	E	X
$K_2S_2O_7$	A	E	X	X	X	E	X
MFI[b]	A	E	X	X	X	E	X

[a] Vycor or fused quartz.
[b] Muffle furnace ignitions.
E = Excellent G = Good A = Acceptable X = Do not use
(Reproduced by permission of ASTM from Dulski (1996).[10])

the pressure and temperature are monitored for each closed vessel and the unit will shut down automatically when a critical temperature or pressure is reached.

With respect to convenience, laboratory microwave ovens can be preprogramed on a time–temperature or time–pressure basis.

The most important advantage of microwave oven acid dissolution is that there is considerable saving of time. In addition, samples that are very hard to dissolve such as high-carbon ferrochromium or chromium metal can be dissolved in a reasonable time. This time saving is particularly significant when a large number of identical samples, or samples that require identical programming of the oven have to be dissolved.

3.4.1.4 Additives In rare instances, the dissolution rate of a metal is very slow and a catalyst must be added to accelerate it. A good example is high-purity aluminum. If we try to dissolve it in dilute HCl, the reaction rate is so slow, as to be barely perceptible. The addition of one drop of metallic mercury will accelerate the dissolution to a reasonable rate. The mercury can be recovered after the dissolution is complete. Generally any special additives are specified in the procedures.

3.4.1.5 Acid Insolubles When the acid dissolution is completed, the reaction stops, sometimes leaving a small amount of dark-colored residue on the bottom of the beaker. The residue may or may not contain any of the analytes. As this is generally not known, the insoluble residue must be recovered. If a particular standard analytical procedure is followed, the procedure will tell the analyst how to proceed. If no procedure for the recovery of insoluble residues is given, they are usually recovered on a filter paper, washed, and ignited in a platinum crucible. If the retention of silicon is of no importance, a drop of dilute H_2SO_4 and a drop or two of HF are added to volatilize the silicon by carefully evaporating it to dryness and igniting the crucible at about 550 °C. This is followed by fusing the residue, if any, in the platinum crucible with a small amount of flux (potassium pyrosulfate or sodium carbonate are frequently used). The cooled melt is then dissolved in a dilute acid and the solution is added to the filtrate. A guide for the selection of crucibles is given in Table 3.

3.4.1.6 Wet Ashing When there are organic materials or acid-insoluble particles present due to organic materials, there is a relatively simple method of removing the organic materials – the wet ashing procedure. The procedure basically consists of transferring the sample or the acid-insoluble residue that contains some organic material to a beaker and adding a quantity of concentrated nitric acid (not less than 30–50 mL) and 5–10 mL perchloric acid (70% $HClO_4$), then covering with a ribbed watch glass or a regular watch glass with glass hooks to separate the watch glass from the beaker. The beaker is placed on the hot plate and slowly heated. In the beginning, copious amounts of yellow nitric fumes will be emitted. When most of the nitric acid is evaporated, white fumes will start evolving. If the remaining solution starts to turn dark, it is a sign that there is still organic matter present and the beaker is immediately removed from the hot plate, left to cool slightly, more nitric acid is added and heating is resumed. When the solution no longer turns dark when the white fumes start to evolve, it is a sign that the organic matter has been decomposed and removed by oxidation. The heating is continued until the perchloric acid fumes are visible only in the upper part of the beaker. The beaker can be removed from the hot plate after 1 min of this fuming, cooled, and carefully diluted with water. This procedure must be carried out

Table 4 Suggested molten salt fusions for metals[a]

Material	Crucible and cover	Procedure
Chromium	Zirconium	10 g Na_2O_2 + 1 g Na_2CO_3 (line bottom with 2 g Na_2CO_3)
Ferroboron	Zirconium	10 g Na_2O_2
Ferrochromium (Hi C)	Zirconium	10 g Na_2O_2 + 1 g Na_2CO_3 (line bottom with 2 g Na_2CO_3)
Ferromanganese	Zirconium	10 g Na_2O_2 + 5 g Na_2CO_3 (**Use Extreme Caution!**)
Ferromolybdenum	Zirconium	10 g Na_2O_2
Ferroniobium	Zirconium	10 g Na_2O_2 + 2 g Na_2CO_3
Ferrosilicon	Zirconium	7 g Na_2O_2 + 7 g Na_2CO_3 (**Use Extreme Caution!**)
Ferrotantalum	Zirconium	10 g Na_2O_2 + 2 g Na_2CO_3
Ferrovanadium	Zirconium	10 g Na_2O_2
High-temp. alloys	Zirconium	10 g Na_2O_2 + 5 g Na_2CO_3 (line bottom with 2 g Na_2CO_3)
Niobium alloys	Fused silica	10 g $K_2S_2O_7$
Tantalum alloys	Fused silica	10 g $K_2S_2O_7$

[a] One gram test portion is implied. (Reproduced by permission of ASTM from Dulski (1996).[10])

carefully and certain precautions have to be observed. (a) Only glass beakers, flasks, hooks, watch glasses, or stirring rods should be used when fuming with perchloric acid. (b) Sufficient nitric acid should always be present. If the solution starts to turn dark it should be removed from the hot plate immediately, cooled, and more nitric acid added. (c) A beaker should never be left on the hot plate unattended if it contains perchloric acid. (d) The operation should be performed only in a clean fume hood that has been designed for use with perchloric acid and this fume hood should not be used with any organic matter. (e) Only experienced analysts should perform this procedure and they should wear appropriate safety equipment, such as safety glasses or goggles, or preferably face shields. Any other personnel working in or close to the same hood should also wear similar safety equipment.

For wet ashing in a microwave oven, in which nitric–perchloric acid or nitric–sulfuric acid systems are to be used, the manufacturer of the equipment should be consulted to avoid any safety-related problems.

3.4.2 Molten Salt Fusions

When the acid dissolution for a metal, ore, or related material is not successful, the next step is to fuse the material with flux in a crucible and then dissolve the melt in dilute acid. This procedure is referred to as a fusion or molten salt fusion. It should be pointed out that these salts in the salt fluxes used in these fusions are not always salts in the pure chemical sense, but also contain substances such as NaOH or Na_2O_2. However, to simplify, they will be referred to in this article as salts and salt fluxes. The procedure is basically simple. The finely divided sample is placed in an appropriate crucible containing some of the salt flux and covered with the balance of the flux. It is mixed with a clean smooth glass rod and heated first at low temperatures to expel any potential moisture present and then slowly to high temperatures. Either a gas burner (the ''Meker''-type is preferred), where the crucible is held on a tripod of appropriate size over the burner, or a muffle furnace is used. The crucible is covered with a lid of the same material. Frequently, the crucible containing the melt is swirled with a pair of tongs to facilitate completion of the fusion. When the fusion is completed, the melt is cooled to room temperature. The cooled melt is dissolved (leached) with an appropriate dilute acid or water that is later acidified. Again, appropriate safety precautions have to be taken and proper eye and face protection is essential.

If a standard procedure is followed, the crucible type and material are specified; if not refer to Table 3. Generally the sample size, flux, and flux quantity are also specified. These should be followed closely. If no procedure is found or no fluxes are suggested, fluxes suggested by Dulski[10] can be used as a good starting point (see Table 4).

A few words about crucibles. The most common crucibles are made from porcelain, fused silica (quartz), iron, nickel, platinum, and zirconium. Today, the platinum and zirconium and, in certain fusions, fused silica crucibles are most frequently used, although owing to the high price of platinum and zirconium crucibles, nickel or iron crucibles can be substituted, resulting in a shorter crucible life and potential for greater contamination by the crucible material. Normal crucible sizes used range from 25–40 mL.

3.4.2.1 Beads for X-ray Fluorescence Spectrometry Analysis A special application of molten salt fusions is the preparation of buttons, often referred to as beads, for direct analysis by XRF. The principal use of buttons is for samples where XRF analysis is desirable and there are difficulties in obtaining a suitable metallic specimen owing to various reasons such as the brittle nature of the metal, difficulties in remelting the sample without affecting the composition, and so on. A good example is ferrotitanium.

The samples are generally pulverized and a representative portion of the pulverized sample is fused with an appropriate salt or salt mixture. The fusion is most often performed in a heavy platinum or platinum–gold alloy crucible and the fusion is cast into an appropriate mold. The glass-like buttons or beads are then analyzed by XRF in the usual way, using analytical curves established with beads prepared from reference materials in the same way as the samples. Lithium metaborate, lithium tetraborate, sodium tetraborate (borax), and boric acid are among the most frequently used salts for this application.

ACKNOWLEDGMENTS

I would like to thank my wife Jenny for proof reading this article. I would also like to thank the American Society for Testing and Materials, Angstrom, Inc., CEM Corp., Cianflone Scientific Instruments Corp., and Consarc Corp. for supplying some of the pictures and illustrations used in this article.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AMS	Aeronautical Material Specification
ASTM	American Society for Testing and Materials
FAAS	Flameless Atomic Absorption Spectroscopy
FQA	Fastener Quality Act
ICP	Inductively Coupled Plasma Spectroscopy
ICP/MS	Inductively Coupled Plasma Mass Spectroscopy
ISO	International Organization for Standardization
NIST	National Institute of Standards and Technology
OE	Optical Emission
OES	Optical Emission Spectroscopy
O&M	Oil and Moisture
PES	Plasma Emission Spectroscopy
XRF	X-ray fluorescence spectroscopy

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction • Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis • Iron Ore, Sample Preparation and Analysis of • Nickel Ore and Metals Analysis • Noble Metals, Analytical Chemistry of • Nuclear Magnetic Resonance in Metals Analysis • Thermal Evolution Methods for Carbon, Sulfur, Oxygen, Nitrogen and Hydrogen in Iron and Steel Analysis • X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry • Microwave-induced Plasma Systems in Atomic Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Absorption Techniques in X-ray Spectrometry • Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Wavelength-dispersive X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Microwave Techniques • Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration • Traceability in Analytical Chemistry

REFERENCES

1. ASTM E 135 Standard Terminology Relating to Analytical Chemistry for Metals, Ores and Related Materials. ASTM Annual Book of Standards Volume 03.05, West Conshohocken, PA, USA, 1997.
2. ASTM E 1806 Standard Practice for Sampling Steel and Iron for Determination of Chemical Composition. ASTM Annual Book of Standards Volume 03.06, West Conshohocken, PA, USA, 1997.
3. ASTM E 716 Standard Practice for Sampling Aluminum and Aluminum Alloys for Spectrochemical Analysis. ASTM Annual Book of Standards Volume 03.06, West Conshohocken, PA, USA, 1997.
4. ASTM E 634 Standard Practice for Sampling Zinc and Zinc Alloys for Optical Emission Spectrometric Analysis. ASTM Annual Book of Standards Volume 03.06, West Conshohocken, PA, USA, 1997.
5. ASTM E 877 Standard Practice for Sample Preparation of Iron Ores and Related Materials. ASTM Annual Book of Standards Volume 03.06, West Conshohocken, PA, USA, 1997.
6. ASTM E 32 Standard Practice for Sampling Ferroalloys and Steel Additives for Determination of Chemical Composition. ASTM Annual Book of Standards Volume 03.05, West Conshohocken, PA, USA, 1997.

7. ASTM E 1916 Standard Guide for Identification and/or Segregation of Mixed Lots of Metals. ASTM Annual Book of Standards Volume 03.06, West Conshohocken, PA, USA, 1998.
8. (United States of America) Fastener Quality Act (PL 101-592, 1990) as amended by the National Technology Transfer and Advancement Act (PL 104-113, 1996) and its implementing regulations (1996 and 1998). Congress of the United States, Washington, DC, USA.
9. ISO Guide 25 General Requirements for the Competence of Calibration and Testing Laboratories, 3rd edition. International Organization for Standardization, Geneva, Switzerland, 1990.
10. T. Dulski *A Manual for the Chemical Analysis of Metals* ASTM Manual MNL 25, ASTM West Conshohocken, PA, 1996.
11. ASTM E 1010 Standard Practice for Preparation of Disk Specimens of Steel and Iron for Spectrochemical Analysis and Remelting. ASTM Annual Book of Standards, Volume 03.06, West Conshohocken, PA, USA, 1997.
12. ASTM E 1306 Standard Practice for Melting Metal and Alloy Samples by Electric Arc Remelting for Determination of Chemical Composition. ASTM Annual Book of Standards, Volume 03.06, West Conshohocken, PA, USA, 1997.
13. E. Oberg (ed.), *Machinery's Handbook*, Industrial Press, Brooklyn, New York, 1996.
14. ASTM E 1257 Standard Guide for Evaluating Grinding Materials Used for Surface Preparation in Spectrochemical Analysis. ASTM Annual Book of Standards, Volume 03.06, West Conshohocken, PA, USA, 1997.

Nickel Ore and Metals Analysis

J.R. Johnston
Formerly of Lakefield Research Limited, Lakefield, Canada

1 Introduction **2**
2 History **3**
3 Analysis of Nickel Ores **3**
3.1 General Sample Preparation and Preservation 3
3.2 Analysis of Nickel Sulfide Ores 4
3.3 Analysis of Nickel Laterite Ores 9
3.4 Quality Control for the Analysis of Nickel Ores 11
4 Analysis of Nickel Metal **12**
4.1 Common Forms of Refined Nickel Metal 12
4.2 Determination of Trace Impurities 13
4.3 Quality Control 15
5 Analysis of Nickel Oxides **15**
5.1 Sampling and Sample Preparation 15
5.2 Sample Dissolution and Analysis 16
5.3 Quality Control 16
6 Ferronickel **16**
6.1 Sampling and Sample Preparation 16
6.2 Analysis of Ferronickel 17
6.3 Summary 18
7 Analysis of Nickel Alloys **18**
7.1 Introduction 18
7.2 Sample Preparation 18
7.3 Determination of Major Constituents 19
7.4 Minor Elements in Nickel Alloys 20
7.5 Trace Elements in Nickel Alloys 21
7.6 Carbon and Sulfur in Nickel Alloys 22
7.7 Gases in Nickel Alloys 22
7.8 Recommended Methods for the Analysis of Nickel Alloys 22
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
List of Standards **23**
International Organization for Standardization Documents 23
National Standards 24
References **24**

This article examines contemporary routine and umpire quality methods for the analysis of nickel ores, nickel metal and nickel alloys. Sample preparation and sampling are also discussed, but the emphasis is on sample preparation and analysis, since sampling is all too often done by people not part of the analytical laboratory.

In order to analyze fully these materials, a laboratory must have at its disposal a wide range of instruments and methods, from the classical to the most modern, and frequently combinations of both. In keeping with the theme of this Encyclopedia, detailed methods are not given, and the reader should consult the references. The references themselves cover the period from roughly 1972 through 1997 but are not to be considered a definitive list. Unfortunately, most of the modern methods of analysis developed by chemists in nickel producers' laboratories have not been published, except as international or national standards. As an aid to the reader, a summary of International Organization for Standardization (ISO) and national standards is given immediately before the References. The author apologizes for any errors or omissions.

The section on nickel ores is divided into two broad classifications: sulfide and lateritic ores. Greater emphasis is given to the sample preparation of laterites since they are more difficult to handle and many laboratories do not have as much experience working with them as with sulfides. Each broad classification is broken down into the analysis of ''pay metals'', which are the reason the ore is mined, elements of environmental concern and other elements which are required for the metallurgical processing of the ore.

Products from refineries fall into three broad categories: nickel metal in various forms, nickel oxides and ferronickel. For nickel metal, the emphasis is on trace-element determinations; for nickel oxides and ferronickel, methods for the determination of nickel are also discussed.

Nickel alloys cover a wide range of materials, from stainless steels to high-nickel ''superalloys''. For these samples the analytical requirements range from the major components of the alloy to minor alloying constituents to trace-element determinations.

As an aid to the analytical chemist in the laboratory, tables are presented summarizing recommended methods of analysis, and in most instances an alternative method, in case the laboratory does not have the equipment necessary to select the first choice. These methods have proven themselves in the laboratories of nickel producers and commercial laboratories familiar with the analysis of nickel-bearing samples.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

For a general overview of the early development of the nickel industry and the analytical chemistry associated with it, the reader should consult *The Analysis of Nickel*,[1] *Analytical Chemistry of Nickel*[2] and two volumes in the series prepared by the Vernadskii Institute on the Analytical Chemistry of the Elements, *Analytical Chemistry of Nickel*[3] and *Analytical Chemistry of Cobalt*.[4] All of these books were written in the late 1960s.

Nickel is a transition-metal element and is closely associated with iron, cobalt and the platinum metals. The main use (about 65% of consumption) for nickel is as an alloying element in stainless steels, where the addition of nickel and chromium imparts greater corrosion resistance. High (35%) nickel, chromium and iron alloys are also used in petroleum refineries owing to their corrosion resistance and thermal resistance. Other important uses are in jet engine ''superalloys'', where the ability to withstand high temperatures is critical. In these alloys, nickel is the major constituent (>60%). Superalloy manufacturers demand very high-purity nickel and place severe restrictions on trace-element concentrations. The plating industry is also a major consumer of nickel, particularly for automobile components. A growing use for nickel is in batteries. Nickel–cadmium (NiCad) batteries have been in use for many years. A potential, but not yet commercial, use is for battery-powered vehicles. Many countries use nickel in coinage, often produced from nickel powder. Nickel powder is also used in powder metallurgical applications. Nickel is also used in catalysts.

Cobalt is used as an alloying element is high-temperature alloys and in high-strength magnets, e.g. cobalt–samarium magnets.

The platinum-group metals are most often used industrially as catalysts, and there is some use in jewelry.

The analysis of nickel ores and metals poses a number of unique problems for the analytical chemist. In ores, nickel is usually a minor component (1–10% by weight), and is often associated with other pay metals such as copper, cobalt, gold, silver and the platinum-group metals. In refined nickel metal, nickel itself is not determined directly. Rather, the trace impurities are determined, and the nickel content is estimated by difference. The trace impurities are of particular interest to superalloy manufacturers, since small amounts of some elements (Pb, Bi, Se, Te and Tl especially) can segregate to crystal boundaries, causing structural weaknesses in turbine blades.

In ferronickel, the two major components are iron and nickel. Nickel is determined, since the material is bought and sold on the basis of nickel content. Usually various minor and trace impurities are also determined. The principal market for ferronickel is the stainless-steel industry.

Some smelters produce nickel oxides as their final product. Some are fairly pure and are often destined for chemical manufacturers. Others are much less pure, almost an intermediate product containing considerable amounts of gangue material, and are sold either for further refining or to low-alloy steel producers. Nickel is determined in these oxides in addition to whatever other elements are deemed important for the end-use of the oxide.

In ores, several components besides the pay metals are also determined. Elements such as sulfur, silicon, magnesium, lead, arsenic, bismuth, selenium and zinc are of metallurgical and/or environmental interest.

Nickel ores occur as three broad types: sulfides, silicates and nickeliferous iron ores (laterites). There are other types, such as arsenides, but these are of relatively little importance. Sulfides, containing nickel mainly as pentlandite, a nickel–iron sulfide, account for about 30% of the world land-based nickel resources. Economically recoverable amounts of cobalt, copper, gold, silver and the platinum group metals are often associated with the nickel sulfide minerals. For the purposes of this article, silicates (saprolites) and nickeliferous iron ores (limonites) will be discussed under the broad term ''laterite'', since they can be analyzed in a similar fashion. The only additional by-product from laterites is cobalt, which may be present in sufficient quantities to justify its recovery. Of potential interest in the distant future are ''manganese nodules'' found on the ocean floor. As yet no way has been found to harvest these nodules in an environmentally safe and economic way. For more on nodules, see Haynes and Magyar.[5]

Nickel deposits are found widely throughout the world. Sulfides are found primarily in the Russian Federation (29%), Canada (26%), Australia (14%), China (11%) and Africa (13%). Nickel laterites (a weathered, clay-like material) are found mainly in the tropics. The laterite deposits are found near the surface in the Caribbean (Cuba, Dominican Republic, totaling 25%), New Caledonia (20%) Indonesia (16%), the Philippines (11%), Australia, (8%), Africa (8%) and Central and South America (8%). Minor amounts are found in the Balkans and in Greece.

In 1998, sulfides accounted for 53% of the worldwide production of about one million metric tonnes (10^6 t). This percentage is lower than in the 1960s when sulfides accounted for 75% of nickel produced. Most of the sulfide deposits are well below the surface, although there is a major near-surface deposit, yet to be developed, in the Voisey's Bay area in Labrador, Canada.

The production of nickel metal from ores involves several steps. For sulfides, the ore is crushed, then usually

put through concentration steps (froth flotation and magnetic separation) to separate metal-bearing minerals from as much of the gangue (usually silicates) as possible. The concentrate, upgraded 4–15 times from the ore, is then treated pyrometallurgically (roasting, smelting and converting) or by a flash smelting process to form a matte. The matte is treated by further flotation separation, matte refining processes or hydrometallurgical processes. Refining to produce nickel metal is done by electrolytic methods or by the nickel carbonyl process.

Laterites cannot be upgraded to the same degree as sulfides. After the removal of low-grade rocks, the ore is dried and may be smelted directly to produce ferronickel. Other processes involve smelting with gypsum (Société Anonyme le Nickel in New Caledonia) or with high-sulfur oil plus elemental S (Inco Indonesia) to produce matte which is then shipped to a refinery. Shipping a matte is much more economic than shipping an ore or a concentrate. However, with improved hydrometallurgical techniques, using either ammonia or acid leaching, laterites are becoming more and more economically attractive to mine and process than sulfides, particularly since mining costs are significantly lower.

The major miners/producers of nickel are Inco, Falconbridge and Sherritt (the last company receives most of its feed from Cuba at time of writing) in Canada, Outokumpu in Finland, Société Anonyme le Nickel in France (with feed from New Caledonia), Norilsk, Montschegorsk and Razno in Russia and Western Mining and Queensland Nickel Industries in Australia. Inco also has a large operation in Indonesia and Falconbridge has a significant ferronickel operation in the Dominican Republic. Lesser miners/producers include Tocantins and Codemin (Brazil), Cerro Matoso (Columbia), Matthey Rustenburg, Empress Nickel, Bindura and Impala (Southern Africa), Aneka Tambang (Indonesia), Larco (Greece), Kosovo (Yugoslavia), and Jinchuan (China). The Japanese have no mines, but there is some refining of matte in that country by Sumitomo, Pacific Metals, Nippon Yakin and Tokyo Nickel. Taiwan Nickel Refining produces utility nickel, a lower-grade product used in the making of low-alloy steels. Glenbrooke in the USA produced ferronickel, but has ceased production at time of writing. Cubaniquel in Cuba produces nickel oxide sinters.

This article will examine contemporary routine and umpire quality methods of analysis for nickel ores, nickel metal and nickel alloys.

2 HISTORY

Prior to the introduction of atomic absorption spectroscopy (AAS) into the analytical laboratory, nickel ores and metals were analyzed by tedious separation, gravimetric, titrimetric and spectrophotometric methods in the wet analytical laboratory and by X-ray fluorescence (XRF) spectroscopy. Flame atomic absorption spectroscopy (FAAS) and its companion, electrothermal atomic absorption spectroscopy (ETAS) have enabled the analyst to determine most of the elements required for the analysis of ores and metals, from macro to micro concentrations. As a result, the use of the ''classical'' techniques of separation and spectrophotometry are rarely, if ever, encountered now, although a recent paper described the spectrophotometric determination of Co.[6] Classical wet chemistry still finds a place in the analysis of standard reference materials (SRMs), and in settlement or umpire analysis. In recent years, inductively coupled plasma optical emission spectroscopy (ICPOES) has come into wide spread use, supplanting AAS in many laboratories. Most recently, inductively coupled plasma mass spectrometry (ICPMS) has been introduced, primarily for the determination of trace impurities. Recent advances in FAAS such as flow injection hydride generation and multi-lamp configurations coupled with computerized data retrieval have allowed this technique to become more competitive with ICPOES.

Nonmetallic elements of interest (carbon, sulfur, oxygen, nitrogen and hydrogen) are still commonly determined by combustion and fusion techniques. Here the major advances have been in furnace design and detection systems, but the basic technique has remained the same.

Refined nickel metal is produced in a number of grades of purity. These grades used to be defined in terms of nickel plus cobalt content. Today, refining methods permit much better separation and recovery of cobalt, and the current ISO standard 6283 regarding impurity limits for various grades of nickel refers to nickel content and not to nickel plus cobalt. American Society for Testing and Materials (ASTM) B39 is another widely used standard (e.g. used by the London Metal Exchange) for grading refined nickel. The British Standard (BS) 375 and the German specification Deutsche Institut für Normung (DIN) 1701 are also widely respected.

3 ANALYSIS OF NICKEL ORES

3.1 General Sample Preparation and Preservation

Drill core should be split before crushing, the uncrushed half being retained as a record and/or back-up in case samples are lost in transit from the drill site to the laboratory. The retention of half the core as a record for future inspection and verification of laboratory results also serves to help protect investors from fraudulent

mining promotions. Ore from a mine should be sampled using proper sampling methods, if the laboratory results are to be used for process metal balances. The size of the ore pieces may range from powder to 0.5 m.

Sulfide ores are generally hard. Laterites are soft, particularly near the surface where they are weathered, and these can often be split with a knife. As the depth of the laterite deposit increases, the ore becomes less weathered and consequently harder, especially if olivine is present.

In general, the bulk sample is crushed through various steps to smaller and smaller pieces. During the crushing stages, the sample size is reduced by half to one quarter at each step. A portion of the reject from each crushing step should be retained in case contamination occurs at a later step. The crushing continues until the sample has been reduced to about 10 mesh in size. A representative portion, about 200 g, is then pulverized to −100 to −200 mesh, usually in a ring and puck type of pulverizer. This sample is then placed in a suitable container and sent to the laboratory for analysis. If more than one laboratory is to conduct analyses, then the amount pulverized should be increased accordingly. There are concerns specific to the preparation and preservation of sulfides and laterites which are given below.

For laterites, the ore should be dried at 100 °C for 1–2 h (depending on the ore), and then crushed with a jaw crusher while hot to avoid gumming up the crusher. Moisture pick-up by laterites is a major concern and is dealt with in detail below (section 3.3.1).

Cross-contamination can occur during crushing if the equipment is not cleaned between samples. In practice, for exploration samples, the core is usually of similar grade and this cross-contamination is not too critical, and minimal cleaning such as a burst of compressed air is sufficient. For more critical samples, a barren mineral such as sand or nephelite silica is often used in addition to compressed air. For extremely critical samples, a small amount of the actual sample may be passed through the sample preparation stream to reduce any slight contamination from remnants of the cleaning material (sand).

3.1.1 Nickel Sulfide Ores

Care must be taken at the pulverizing step. The friction of the moving components of the equipment (such as ring and puck) in contact with the ores can lead to a heat buildup. When the powdered ore is dumped on to a sheet of paper, the sulfides may react with the oxygen in the air and become hot enough to ignite the paper. This problem is common to all sulfide ores, and sample preparation personnel should be made aware of this possibility. Often there is a smell of SO_2, indicating that oxidation is occurring. The pot should be placed in a small can, a lid put on and allowed to cool before opening the pot.

The sample should be packaged quickly to minimize oxidation of the sulfides. After the analysis, the sample should be stored in a sealed container and/or in a freezer to prevent oxidation from occurring. Storage under nitrogen has been used for certified reference materials.[7] All samples and rejects must be properly labeled.

3.1.2 Nickel Laterite Ores

Laterite ores present problems for the sample preparation laboratory. They are very similar to clays, and are extremely hygroscopic. This often results in smearing and caking, particularly in the pulverizing step. The best way to prepare nickel laterites is to moisten the crushed ore with a small amount of alcohol (methanol or ethanol). The alcohol absorbs moisture in the ore and prevents caking during pulverizing.

The pulverizing equipment, usually a steel ring and puck in a pot, must be well cleaned after each sample is ground. Greater care than normal is required owing to the clay-like material sticking to the surfaces.

Once pulverized, the samples are prone to gain weight by absorbing moisture from the air. Drying the sample at 105 °C and storing it in a desiccator will give inaccurate results. The ore is so hygroscopic that it will absorb moisture even from fresh desiccant. For routine analysis, the ore should be stored in an oven at 105 °C until the sample is to be weighed.

The best way to obtain an accurate analysis is to weigh two portions of the pulverized sample, one for the elemental analysis and the other portion for moisture, and correct to a dry weight basis. Because this is critical, the determination of the moisture content is described in some detail below (section 3.3.1).

The pulverized samples are relatively stable with the exception of moisture pick-up.

3.2 Analysis of Nickel Sulfide Ores

3.2.1 Pay Metals (Nickel, Copper, Cobalt, Gold, Silver, Platinum Group)

3.2.1.1 Determination of Nickel, Copper and Cobalt by Wet Chemical Methods

By Dissolution with Mineral Acids. A common method is to attack the sample (usually 1 g) with 1 : 1 hydrochloric acid and heating on a hot-plate. This has the advantage of converting the sulfide sulfur to the volatile H_2S, reducing the amount of dissolved solids. When dissolution ceases, the insoluble matter is filtered,

ignited and fused with sodium peroxide in a zirconium crucible in order to achieve the complete dissolution of the sample. The filtrate contains most of the base metals as well as others of interest (see section 3.4). The sodium peroxide melt is dissolved in water and acidified with hydrochloric acid. The fusion solution will contain small amounts of Ni, Cu and Co which can be occluded in the crystal lattice of silicates or as impurities in refractory minerals such as magnetite which is often found in the Sudbury ores. It is best not to combine the two solutions, but to dilute each to an appropriate volume for analysis by classical wet chemistry or by FAAS or ICPOES.

Inco developed a method for the microwave acid digestion of sulfides, followed by FAAS.[8] This procedure has been supplanted by fusion with sodium peroxide in a zirconium crucible followed by acid dissolution of the melt.[9,10] The latter procedure is currently in routine use in Inco laboratories.

A simple method for acid dissolution where only Cu, Ni, Co and Fe are to be determined is the four-acid method. A 1-g sample is heated in a glass beaker with concentrated HCl to decompose most of the sulfides. A few milliliters of HNO_3 are added, then a few drops of HF (*caution*) to break up silicates, then a few milliliters of H_2SO_4 are carefully added and the sample is heated gently to dryness. The sample is taken up with a few milliliters of 1 : 1 HCl and some water and heated to complete dissolution. After dilution to an appropriate volume, the sample is analyzed by FAAS or ICPOES. Perchloric acid (*caution*) can be used instead of sulfuric acid. The method can be applied to both sulfides and laterites.

If the determination of Ni is to be done by the classical gravimetric determination by precipitation with dimethylglyoxime (DMG) (see the next section), it is best to heat the sample solution to strong fumes of perchloric acid to dehydrate silica. This is necessary to prevent falsely high results owing to SiO_2 being occluded in the NiDMG precipitate.

Caution: Perchloric acid must only be used in a fume hood specifically designed to withstand perchloric fumes. The use of regular fume hoods has resulted in major explosions in laboratories.

Determination of Nickel by Classical Wet Chemistry. The determination of Ni by precipitation with DMG has been described in many undergraduate textbooks. For sulfide ores, an aliquot of the solution from section 3.2.1.1 containing 25–50 mg of Ni should be taken. Tartaric acid must be added to complex Fe, which is always present in large amounts. Sodium thiosulfate must also be added to complex Cu. The weight of $Na_2S_2O_3 \cdot 5H_2O$ should be 15 times the weight of Cu plus Fe, with a minimum of 3 g. The salt is usually added as a 20% w/v solution, and 20 ml is commonly used. Alternatively, the Cu may be removed and determined by electrolytic plating prior to the precipitation of Ni with DMG. The NiDMG precipitate may be filtered through a tared sintered-glass crucible, washed carefully with water and dried at 150 °C. The drying temperature is critical. It must be high enough to dry the precipitate and to volatilize any residual DMG, but not high enough to decompose the NiDMG precipitate. The crucible is then weighed and the percent nickel is calculated.

An alternative is to filter the NiDMG precipitate through a fast filter, wet ash the paper and precipitate and titrate the nickel with ethylenediaminetetraacetic acid (EDTA) using murexide as the indicator. This alternative is faster and avoids the risk of incorrect results owing to improper drying of the precipitate.

The DMG precipitation procedures are generally restricted to umpire analysis or SRMs certification.

Xiang et al. reported the potentiometric determination of Ni and Cu.[11]

Determination by Flame Atomic Absorption Spectroscopy or Inductively Coupled Plasma Optical Emission Spectroscopy. For FAAS and ICPOES it is imperative that the analysis be carried out using matrix-matched standards.

The filtrate solution from dissolution with mineral acids (see above) is usually fairly stable, but if older than a few days it should be inspected for any precipitate which may have appeared. Any such precipitate is unlikely to contain enough Ni, Cu or Co to affect the accuracy of their determination. The insoluble sodium peroxide fusion solution is much less stable, particularly if the sample contains large amounts of silica and/or alumina. It is best to analyze this solution as quickly as possible.

Another approach, developed by Inco,[9,10] is to fuse the sample with sodium peroxide in a zirconium crucible. For large production runs, an automatic sample fluxer is employed. The melt is then dissolved in water, acidified with hydrochloric acid, diluted to an appropriate volume and analyzed by ICPOES against matrix-matched standards. This procedure retains the sulfur, which is oxidized to sulfate by the sodium peroxide, and the total sulfur is also determined by ICPOES. This procedure is in routine use at Inco laboratories.

Caution: Sodium peroxide is a very strong oxidant. Sulfide is a strong reductant. Heating the two together can produce a violent reaction unless great care is taken.

Claisse[12] indicates that Ni sulfides can be fused in platinum crucibles if the sample weight is kept to a maximum of 0.25 g, and an oxidant such as $LiNO_3$ (to convert sulfides to sulfates) is used with the Li or Na tetraborate flux. Exceeding this sample weight may cause an alloy to form with the platinum crucible. LiBr is used as the releasing agent in place of the usual iodide owing

to the volatility of copper iodide. The melt may be poured into a dish to form a glassy disk for XRF analysis, or be dissolved in a suitable acid such as 10% v/v nitric acid for ICPOES analysis (see Claisse,[12] p. 44).

3.2.1.2 Determination of Nickel, Copper and Cobalt by X-ray Fluorescence Spectroscopy This technique has been used by analytical laboratories for almost 50 years to provide rapid and accurate results. Improvements in the hardware and, more important, in the software, have helped this technique withstand the challenges of FAAS and, more recently, ICPOES.

These metals can be determined fairly accurately by fusing a 0.2-g sample with 5 g of potassium pyrosulfate in a quartz crucible. The fusion homogenizes the sample, breaking up the various mineral species, and the potassium pyrosulfate dilutes the sample, reducing matrix effects. The fusion also oxidizes the sulfides to sulfates. The melt is ground finely in a porcelain mortar and pestle, then is briquetted in a mounting press for analysis. The XRF spectrometer may be calibrated using a variety of reference materials including pure metals, as long as they fuse completely with the pyrosulfate. If these are not commercially available, a suite of samples covering the expected concentration ranges must be analyzed carefully by wet chemical methods, in order to serve as in-house secondary standards. The briquette is stable almost indefinitely.

This general procedure has been used with confidence for many years at Lakefield Research Limited[13] and can be used for metal balances. XRF has been used elsewhere.[14]

The pressed powder technique, where the pulverized ore is mixed with a suitable binder and is pressed into a smooth, flat disk under very high pressure, is a powerful technique for the rapid analysis of exploration and mine development samples. Several hundred samples per day can be analyzed. The instrument is calibrated using secondary standards of the ore, carefully analyzed by wet chemical methods. These calibration curves are, however, dependent on the mineralogy of the ore. If samples from a different ore zone or ore body are submitted for analysis, additional secondary standards and calibration curves must be developed.

The use of XRF is not restricted to the analytical laboratory. On-stream analyzers[15] located in mills and flotation circuits are valuable tools to plant operators. These analyzers are best calibrated by personnel from the XRF analytical laboratory, who have first selected appropriate samples for calibration, and have derived the correction factors to be used. This method is applicable to both sulfide and laterite samples. Caution must be taken with the storage of laterite disks, since they can pick up moisture, easily swell and distort the geometry of the disk.

On-site XRF spectrometers can also speed up the assessment of exploration samples and ore samples.[16]

3.2.1.3 Miscellaneous Techniques γ-Ray spectrometry has been used for bulk ore analysis.[17,18]

3.2.1.4 Determination of Silver, Gold, and the Platinum Group Metals by Fire Assay Au, Pt and Pd are usually determined by the classical fire assay with lead collection, adding a small amount of silver to give the final bead enough size to be handled easily (a silver inquart). The fire assayer will experiment with the ore to determine the best fusion flux for that particular ore. After fusion and cupellation of the lead button, the silver bead containing the Au, Pt and Pd is dissolved in aqua regia and the metals are determined by ICPOES[19] or FAAS. This general procedure is in common use in most commercial assay laboratories. For a general text on fire assaying, see Bugbee[20] and Dietrich and Shepard.[21] For more detailed information of the chemistry of precious metals, see Beamish[22] or Beamish and Van Loon.[23]

The lead bead can also be analyzed for precious metals by ETAS,[24] solvent extraction–FAAS[25] or by polarography,[26] although these techniques are rarely used.

Rh, Ru, Ir and Os are not determined as frequently as Au, Pt and Pd, and their concentrations are usually much lower in ores. If their determination is required, however, the sample is fused with a mixture of Ni powder or Ni oxide powder and sulfur. This produces a nickel sulfide matte, similar to that produced in a smelter.[27] The precious metals are collected in the matte. Pt and Pd are also collected and can be added to the determination suite. Au is not collected completely,[27–30] recoveries being 75–90%. Results from this collection procedure should be used as a rough guide, but if the Au content is low, the incomplete recovery may not be economically significant. The NiS button is pulverized, dissolved in hydrochloric acid and the insoluble precious-metal sulfides are filtered off. They may be dissolved in a mixture of H_2O_2 and HCl, contacted with an ion-exchange resin to remove traces of Ni. Rh, Ru and Ir can then be determined by FAAS or ICPOES. The removal of traces of Ni is critical if ICPOES is the technique used, owing to spectral interferences. If FAAS is used, traces of Ni do not interfere. Os is lost during the dissolution. A separate fusion with a different dissolution is required for the determination of Os. Alternatively, the NiS button is divided after the pulverization step, with part being used for the determination of Rh, Ru and Ir and the other part for Os. Os is very insensitive by FAAS and ICPOES. It can be determined wet chemically by fusing the precious metal residue with sodium peroxide, dissolving the melt in nitric acid, extracting Os into toluene, then back-extracting into a solution of thiourea to form a colored

complex. This is very tedious. A better route, if available, is to determine Os in the precious metal residues by instrumental neutron activation analysis (INAA). While this requires more elapsed time, it requires much less bench time and is therefore significantly cheaper.

Shibuya et al.[31] analyzed the NiS button directly by laser ablation ICPMS.

Recently, Barefoot[32] reviewed the determination of precious metals by traditional fire assay (Pb collection), NiS fusion, acid leaching, etc. with the determinative step being ICPMS. For mineral samples, fire assay is necessary. Other attacks were incomplete, based on certified reference materials.

INAA has also been used for the determination of the other precious metals in precious metal residues with good accuracy.[33–35] The disadvantage is that irradiation, cooling and counting times are much longer than the times required in the procedures for dissolution and FAAS or ICPOES for Rh, Ru and Ir. Access to a nuclear reactor is also required.

Ag and Au can be determined by the classical fire assay lead collection method if no inquart is utilized. The bead is weighed on a microbalance, then dissolved in aqua regia and Au is determined by FAAS or ICPOES. Ag is determined by difference. If the bead is spongy or black, there may be other precious metals present, and the dissolved solution should be checked for Pt and Pd.

Faye and Moloughney[36] proposed tin as a collector for the determination of the platinum-group metals, gold and silver. This method requires considerable skill to use, and has not found wide acceptance by laboratories determining precious metals.

Zheng and Cui[37] determined Pd by absorbing Pd on polyurethane foam, wet ashing the foam and determining Pd by ETAS.

Kolosova et al. reported the determination of Os by ETAS.[38]

Kinetic methods for the determination of Pd,[39] Rh[40] and Os[41–43] have been reported.

The determination of platinum group elements and Au by accelerator mass spectrometry has been reported.[44]

Warning: Some laboratories have attempted to determine precious metals in ores by dissolution with aqua regia followed by analysis of the solution by ICPOES. False positives are inevitably reported owing to spectral interferences from iron. Occasionally, there are reports that the precious metals are present in a form (unknown) that prevents their determination by conventional assay techniques. In these cases, it is best to determine the precious metals by the nickel sulfide collection procedure, since this mimics the actual production step of converting the ore concentrate into a matte. If the precious metals cannot be recovered by the production process, then the ore alleged to contain precious metals is worthless to the smelter.

3.2.1.5 Determination of Silver by Wet Chemical Methods
This procedure is best suited for low concentrations of Ag ($0.5–800\,g\,t^{-1}$, although an upper limit of $200\,g\,t^{-1}$ is commonly observed). The sample (usually 2 g) is first oxidized with a small amount of bromine, then attacked with aqua regia and 1–2 mL of hydrofluoric acid (to help decompose silicates). The digested sample is transferred to a volumetric flask, acidified to 25% v/v hydrochloric acid and the Ag is determined using matrix-matched standards. The bromine oxidation step is needed to prevent low recoveries of Ag.[45]

3.2.2 Elements of Environmental Concern (Sulfur, Mercury, Selenium, Tellurium, etc.)

Sulfur emissions from smelters have been a concern for many years. Increasingly strict regulations have required smelter operators to invest heavily in metallurgical research and plant capital. The accurate determination of S is therefore critical in the evaluation of an ore. The most common method is by combustion of the sample at high temperature in a ceramic crucible or boat in a stream of oxygen. The S is converted to SO_2, which is determined by passing the combustion gases through a dedicated infrared (IR) cell. The instrument is calibrated using reference materials with known total sulfur content. The detection cell must be capable of reading high sulfur levels (in the range 5–25% by weight). If the instrument cell cannot determine such high levels without saturating the detector, then smaller than normal sample weights must be used, with a consequent decrease in accuracy. The instrument manual should be consulted if there is any doubt. Older instruments would pass the combustion gases, the sulfur now present as sulfur dioxide, into acidified starch–iodide solution and titrate with potassium iodate solution. This old method may still be encountered, and is well suited for massive sulfide ores. Another old technique which may be encountered is to absorb the sulfur dioxide in a 5% solution of sodium peroxide and titrate the acid formed with sodium hydroxide.

Inco[9,10] has developed a routine procedure whereby the sample is fused with sodium peroxide. The sulfides are converted to sulfates, which are not volatile upon acidification with hydrochloric acid. The sample solution is then analyzed by ICPOES.

If total sulfur is to be determined by the classical gravimetric procedure (precipitation as $BaSO_4$), then the sulfides must be oxidized to sulfates. This can be accomplished by allowing the sample to stand in contact with a few milliliters of bromine plus a little water for at least 15 min at room temperature. Nitric acid is then added

and the sample is heated *gently* on a hot-plate to complete the oxidation and to remove Br_2. When dissolution ceases, the insoluble matter is removed by filtration and the sulfur (now present as sulfate) is determined in the classical manner. This procedure is not used for routine analyses, but still finds a place in the analysis of samples to be used as reference materials. Details of this method can be found in classical analytical texts such as Kolthoff and Sandell.[46]

Sulfur has also been determined by XRF.[47]

Mercury is found in trace amounts in many ores. Even at very low levels ($<0.1\ g\ t^{-1}$), there are concerns owing to the high tonnages of ore processed by smelters. Owing to its volatility, Hg is assumed to be lost to the environment in stack emissions. Most laboratories use variations of the cold vapor method developed originally by Hatch and Ott.[48] Recovery studies are critical, since some elements often found in nickel sulfide ores (such as Se and Te) may inhibit the reduction of Hg to the elemental form. During the dissolution, the sample should be kept oxidized ($KMnO_4$), and the temperature kept below 90 °C, or reflux conditions should be used. Tartaric acid is used to complex base metals during the reduction of Hg. In the presence of vegetation, if charring occurs, Hg will be reduced to its elemental form and lost. Blanks are critical since even the reagents may contain Hg at concentrations similar to the samples.

Se and Te are often encountered in sulfide ores. Some may also be lost during smelting via stack gases, although much is retained in the matte. The presence of Se and Te in the matte is a concern to nickel refineries, which must separate them from the Ni. The hydride evolution AAS method does not work well for the determination of Se and Te owing to the presence of Ni and Cu. If the Ni concentration is fairly low (0.5–1%), then complexing agents such as hydroxylamine or cysteine may be added before the reduction step. It is better to oxidize the sulfide ore sample with bromine, then dissolve Se and Te with nitric acid and determine them by ETAS.

Te has also been determined by polarography.[49]

Depending on the ore body, other trace and minor elements may also be required. As, Cd and Sn are frequently required. In ores they are often present only in traces, and can be determined by ETAS after dissolution by the bromine–nitric acid attack described above.

3.2.3 Other Elements (Iron, Silicon, Aluminum, Magnesium, Calcium)

These elements are critical to the smelter, since they will all eventually report to slags. They may be determined wet chemically by FAAS or ICPOES by dissolutions described in section 3.2.1.1. They may also be determined by XRF on briquettes if the ore body is reasonably homogeneous. If more than one ore body is being processed, then each body must have its own XRF calibration curve.

Classical methods (redox titration for Fe, complexometric titrations with EDTA for Al, Ca and Mg,

Table 1 Recommended methods for the analysis of nickel sulfide ores

Element	Recommended method	Alternative method
Nickel	Fusion/ICPOES	Fusion/XRF
Copper	Fusion/ICPOES	Fusion/XRF
Cobalt	Fusion/ICPOES	Fusion/XRF
Gold	Fire assay/Pb collection	
Platinum	Fire assay/Pb collection	
Palladium	Fire assay/Pb collection	
Rhodium	Fire assay/NiS collection	
Iridium	Fire assay/NiS collection	
Ruthenium	Fire assay/NiS collection	
Osmium	Fire assay/NiS collection	
Silver	Acid dissolution/FAAS	Fire assay/Pb collection
Sulfur	Combustion	Fusion/ICPOES
Aluminum	Fusion/ICPOES	Fusion/FAAS
Iron	Fusion/ICPOES	Fusion/FAAS
Silica	Fusion/ICPOES	Fusion/FAAS
Calcium	Fusion/ICPOES	Fusion/FAAS
Magnesium	Fusion/ICPOES	Fusion/FAAS
Mercury	Cold vapor AAS	
Arsenic	Fusion/ICPOES	Fusion/FAAS
Selenium	Br_2–HNO_3 leach/ICPOES	Br_2–HNO_3 leach/ETAS or FAAS
Tellurium	Br_2–HNO_3 leach/ETAS or FAAS	
Miscellaneous traces (Sn, Cd, etc.)	Br_2–HNO_3 leach/ETAS or FAAS	

gravimetry for Si) can also be used, although these methods are relatively slow and labor-intensive. Reports occasionally appear in the literature (Al,[50] Ca and Mg[51]).

3.2.4 Recommended Methods

Table 1 sets out recommended methods of analysis for sulfide ores.

3.3 Analysis of Nickel Laterite Ores

Bibent and Bufferne[52] reviewed methods of analysis for New Caledonian laterites.

Owing to the clay-like nature of laterites, the determination of moisture is critical to accurate analyses. Lopez et al.[53] discuss the determination of moisture in Moa laterites.

For practical, routine purposes, the pulverized sample should be dried and stored in an oven at 105 °C. If, however, a very accurate analysis is required, e.g. for settlement or for validation of a reference material, then the best way to obtain an accurate analysis is to weigh two portions of the pulverized sample, one for the elemental analysis and, the other portion for moisture, and correct to a dry weight basis. Because this is critical, the determination of the moisture content is described in some detail below (section 3.3.1).

Since laterites are extremely hygroscopic, storage in a desiccator is not recommended. The dried ore will absorb moisture from the desiccant, even if the desiccant is fresh, and will absorb moisture from the air each time the desiccator is opened.

3.3.1 Pay Metals (Nickel, Copper, Cobalt)

Ni is the element of greatest concern. Usually only small amounts of copper and cobalt are present. If significant amounts of Co are found, it is wise to check to make sure that the crushing and grinding equipment is not made of cobalt alloy (e.g. Stellite™). This has caused problems in some laboratories, where falsely high Co results have been reported. Caution should also be exercised if tungsten carbide pots are used. These pots may contain up to 6% Co (used as a ''cement'' to tie the tungsten carbide particles together).

For accurate results, a correction for moisture must be made. It cannot be stressed enough that the conventional procedure of drying at 105 °C and storing the sample in a desiccator will yield low, inaccurate results owing to moisture pick-up by the sample from the desiccant. In fact, a dried or partially dried sample can be seen to gain weight on a balance if the laboratory atmosphere is at all humid.

The best way to obtain the moisture correction factor is as follows:

1. Weigh the analytical sample into a beaker (for wet chemical analysis) or crucible (for ICPOES or XRF analysis)
2. Immediately thereafter, weigh a 5-g sample for moisture content determination, into a tared, dry glass weighing bottle plus tared lid.
3. Place the weighing bottle in an oven at 105 °C with the lid ajar to allow moisture to escape. Heat for at least 2 h.
4. Remove the weighing bottle from the oven and place in a desiccator containing fresh desiccant. Tap the bottle to cause the lid to seat into the top, closing the bottle. This must be done as rapidly as possible on removing from the oven.
5. Close the desiccator and allow the bottle to cool to room temperature. The cooling will create a slight vacuum inside the closed bottle, pulling the lid down snugly and preventing moisture from the inside of the desiccator from reaching the sample.
6. When cool, weigh the bottle and contents without removing the lid. The moisture content of the sample can then be calculated, and used to correct the analytical results to a dry basis.

3.3.1.1 Determination of Nickel by Wet Chemical Methods

Dissolution of the Sample. The dissolution described in the section on dissolution with mineral acids above may be used. There will be more dissolved solids than with sulfide ores since there is little S present to be volatilized as H_2S.

Nickel by Classical Precipitation with Dimethylglyoxime. The procedures described in the section on determination of nickel by classical wet chemistry may be applied. For accurate results the procedure for moisture correction detailed in section 3.3.1 *must* be followed.

Nickel by Flame Atomic Absorption Spectroscopy or Inductively Coupled Plasma Optical Emission Spectroscopy. The procedures described in the section on determination by FAAS or ICPOES may be applied. For accurate results, the moisture correction must be made, and the final (uncorrected) result should be determined by using close-bracketing calibration standard solutions. For ICPOES these standard solutions must be matched in acid strength to the samples, and the acid strength of the samples must be tightly controlled.

Gedeon et al.[54] compared digestion techniques for FAAS analysis and concluded that a mixture of HF and perchloric acid was best to achieve complete dissolution of the sample. They also compared the results to those obtained by XRF.

Copper and Cobalt by Flame Atomic Absorption Spectroscopy or Inductively Coupled Plasma Optical Emission Spectroscopy. The solution used for the determination of Ni (see the previous section) may also be used for Cu and Co.

3.3.1.2 Analysis of Nickel Laterite Ores by X-ray Fluorescence Spectroscopy XRF has been a popular technique for the analysis of laterites.[56–61] In a commercial laboratory,[59] the samples are fused with lithium tetraborate in a platinum crucible at a ratio of 0.5 g of sample to 7 g of lithium tetraborate. A small amount (0.5 g) of LiBr is added to the mix as a releasing agent during the pouring of the melt. The melt is poured into a platinum dish and cools to a glassy disk. The use of platinum ware is possible since laterites contain only small amounts of sulfur. High (>3%) Ni ores can also cause problems owing to the formation of an alloy with the Pt crucible. This will result in carryover to subsequent samples. For high-Ni samples, a smaller sample weight should be used. The disk is then analyzed for Ni, Co, Fe, Cr, Ca, Mg, Na, K, Mn, Ti, Al and Si. The disk is very stable and can be stored indefinitely. Loss on ignition (LOI) is also performed on a dried sample at 1000 °C, and gives a measure of the water of crystallization, carbonates, sulfides and some graphite. If trace elements are desired, e.g. Cu, P, Zn, a briquette is made by pressing into a pellet with a suitable binder. The briquette is also fairly stable, but is prone to pick up moisture. If further analysis of the briquette is contemplated, it must be stored in a desiccator and inspected before it is analyzed again (the briquette will swell markedly on picking up moisture).

The pressed pellet technique described earlier is very suitable for the routine analysis of laterites. It is best, however, to let the ore equilibrate with the moisture in the laboratory atmosphere before making the pellet. As is the case with sulfide ores, the calibration curve is specific to the ore zone being analyzed.

3.3.2 Neutron Activation

This technique is popular for exploration and bulk sample analysis.[61–64]

3.3.3 Elements of Environmental Concern (Sulfur, Mercury, Selenium)

Sulfur is usually present in low concentrations (<1%) and can be determined by combustion instruments. Owing to the large amount of water of crystallization found in laterites, the moisture trap on the instrument must be checked frequently. A good indication of problems is if the SRM used to monitor performance (every tenth sample) gives a poor result.

Hg can be determined by the cold vapor technique, with relatively few interferences likely to be encountered. Dissolution can be effected with aqua regia and $KMnO_4$ as an oxidant. If vegetation is present in the sample, care must be taken to avoid charring of the organic matter, since this will instantly reduce Hg to the elemental state and it will be lost.

Se is also usually present in trace amounts ($<5\,g\,t^{-1}$). Se can be determined at these levels by hydride evolution AAS if precautions are taken to overcome interference by Ni (see above) or by ETAS. If the latter technique is used, beware that the large amounts of Fe present in the ore will interfere with the measurement if the deuterium background correction is used. There should be no such problem if the instrument uses the Zeeman or Smith–Hieftje corrections. Alternatively, Se may be collected using As and determined by conventional ETAS.

Other trace elements such as As, Cd and Te are not commonly determined in laterite ores.

3.3.4 Other Elements (Carbon, Iron, Silicon, Aluminum, Magnesium, Calcium)

3.3.4.1 Carbon by the Combustion Technique C is determined using combustion techniques, with the same caveat regarding the moisture trap as described in section 3.3.3.

3.3.4.2 Iron, Silicon, Aluminum, Magnesium and Calcium by Wet Chemistry For routine analyses, the sample may be dried and stored at 105 °C. If very accurate results are required (e.g. for the certification of reference materials), then the moisture correction procedure in section 3.3.1 should be followed.

Routine Analyses. A 0.25-g sample may be fused in a Zr crucible with Na_2O_2. The cooled melt is dissolved in water and then acidified to 20% v/v HCl in a 250-mL volumetric flask. The solution should be clear. Any insoluble matter (usually silica) should be filtered off and re-fused to obtain a completely dissolved sample. The sample solution is analyzed by either FAAS or ICPOES against matrix-matched standards.

Alternatively, dissolve the sample with HCl, filter off the insoluble matter, fuse the latter with Na_2O_2 in a Zr crucible, dissolving the cooled melt in water and acidifying with HCl as described above. The solutions are made up to volume separately and are analyzed

by FAAS or ICPOES using matrix-matched calibration solutions.

Miscellaneous. Ahmad and Morris[65] determined Cl by neutron activation.

Reference Material Quality Analyses. For Fe, Al, Ca and Mg, the sample may be dissolved in HCl, the insolubles filtered off and fused as described in the alternative procedure in the routine analyses section above. Fe may be determined on an aliquot of the sample solution by redox titration (e.g. with potassium dichromate). Ca and Mg may be determined by EDTA titrations after separating the Fe and Al by precipitation with ammonia solution (the classical R_2O_3 separation). Al may also be determined by EDTA titration after separating Fe by ion exchange. Residual Fe, Ca, Mg and Ca present in the acid insoluble matter may be determined by FAAS or ICPOES, and added to the values determined by titration.

Silica is best determined on a separate sample. The sample is attacked first with HCl, then taken to strong fumes with perchloric acid ($HClO_4$) to dehydrate the silica. The residue after fuming is taken up in dilute HCl, filtered, washed well to remove $HClO_4$, ignited in a platinum crucible, and silica determined by difference upon volatilization with hydrofluoric acid (HF).

3.3.5 Recommended Methods for the Analysis of Laterite Ores

Table 2 sets out recommended methods.

3.4 Quality Control for the Analysis of Nickel Ores

While exact quality-control procedures vary from laboratory to laboratory, at the least, the following controls should be run:

1. Certified reference materials or in-house secondary standards which have been validated against certified standards. For nickel sulfide ores, the Canadian Certified Reference Materials Project (CCRMP), organized by CANMET, a federal government agency, is a source of well-proven materials. For nickel laterites, the International Geological Survey (IGS) is a good source, particularly materials IGS 22 and IGS 23. See also Potts et al.[66] for a more complete listing of ore standards.
2. Randomly selected replicate samples at frequent intervals.
3. Procedure blanks.

Caution should be used with reference materials. The certified Ni content of each material is a consensus value based on the results from many laboratories using different analytical techniques. The analysts who generated the data possess varying degrees of experience in the analysis of nickel ores. Occasionally, further analysis by experienced laboratories will give a more reliable result. If there is any doubt as to the accuracy of a certified reference material, the certifying body should be contacted. In the case of the IGS laterites, not all of the participating laboratories may have been aware of the hygroscopic nature of the laterites. Consequently,

Table 2 Recommended methods for the analysis of nickel laterite ores

Element	Recommended method	Alternative method
Nickel	Fusion/XRF	Fusion/ICPOES or FAAS
Copper	Briquette/XRF	Fusion/ICPOES or FAAS
Cobalt	Fusion/XRF	Fusion/ICPOES or FAAS
Iron	Fusion/XRF	Fusion/ICPOES or FAAS
Silica	Fusion/XRF	Fusion/ICPOES or FAAS
Aluminum	Fusion/XRF	Fusion/ICPOES or FAAS
Calcium	Fusion/XRF	Fusion/ICPOES or FAAS
Magnesium	Fusion/XRF	Fusion/ICPOES or FAAS
Chromium	Fusion/XRF	Fusion/ICPOES or FAAS
Manganese	Fusion/XRF	Fusion/ICPOES or FAAS
Titanium	Fusion/XRF	Fusion/ICPOES or FAAS
Sodium	Fusion/XRF	Acid dissolution/FAAS
Potassium	Fusion/XRF	Acid dissolution/FAAS
LOI at 1000 °C	Gravimetry	
Moisture at 105 °C	Gravimetry	
Carbon	Combustion	
Sulfur	Combustion	
Phosphorus	Briquette/XRF	Acid dissolution/colorimetry
Zinc	Briquette/XRF	Fusion/ICPOES or FAAS
Mercury	Cold vapor AAS	
Selenium	Acid dissolution/ETAS	

the certified values of the IGS reference laterites may be slightly low.

4 ANALYSIS OF NICKEL METAL

Refined nickel metal is produced in various shapes and sizes: powder, briquettes and rondelles, granular nickel and nickel shot, carbonyl pellets and cathode nickel. High-purity refined nickel is used by stainless-steel and high-Ni alloy producers and in the plating industry. Less pure nickel may be used by steel producers to make low-alloy steel. The minor impurities will be diluted in these steels to the point where their effect on the steel will be negligible for the particular application.

4.1 Common Forms of Refined Nickel Metal

The most common forms of refined nickel are nickel pellets (produced by the carbonyl process) and nickel cathodes.

Carbonyl pellets are spherical and range in size from 0.5 to 2 cm in diameter, with most being about 1 cm. The growth is by layers, and a cross-section under a microscope resembles an onion.

Nickel cathodes are usually about 0.7 m wide and 1 m long. They are harvested from the plating tank when they are about 0.5 cm thick. They may be shipped in this form, but most often are sheared into squares or rectangles, the most common being 1- and 2-in squares. These squares are shipped in drums of 50, 250 or 1000 kg capacity. Another form of cathode nickel is produced primarily for the plating industry. It is most often a circular piece of metal, roughly one inch in diameter (smaller sizes are also produced), grown by electrolytic deposition onto a stainless-steel stud and then removed when the desired size is achieved. This type of product has no sharp edges (unlike the sheared cathode squares or rectangles), and can be loaded into anode baskets in plating shops without tearing the baskets. This product may also be made in a plating bath containing sulfur compounds which will react during the electrolytic deposition to form a sulfur-depolarized nickel product which has dissolution properties desired by some nickel platers.

Nickel powders of different degrees of purity are also produced either by the carbonyl process or by extraction/reduction processes. These may be delivered as the powder or may be compacted into pillow-shaped briquettes or rondelles weighing 50–100 g for easier handling in foundries.

Granular nickel (often relatively impure) is also produced. It is about 1–2 mm in diameter, roughly spherical, but larger, irregular blobs may be present.

Each form presents its own problems in sampling, but once a representative sample is obtained, the analysis remains much the same regardless of the original form. The sampling of the various common forms and the preparation of the analytical sample is discussed below. For a detailed description of sampling, see ISO 7156.

4.1.1 Sampling and Sample Preparation of Carbonyl Pellets

These are spherical, ranging in size from 5 to 20 mm in diameter. They are usually shipped in drums of 50, 250 or 1000 kg capacity. Owing to size variations, there can be segregation within the drum. This is usually not too important owing to the high purity of this material. A number of drums are randomly selected. Pellets are scooped out from each drum, blended and split down until a 200-g representative sample is obtained. Small pellets may be analyzed directly, but larger ones should be drilled or machined to obtain chips. The chips are blended to form the analytical sample.

If there is concern that the variation in purity is great within a drum, then the entire contents of the drums selected must be dumped out and properly sampled by quartering or alternate shoveling techniques. The final sample should then be prepared by drilling or machining chips and blending. A detailed sampling procedure is given in ISO 7156.

Carbonyl nickel is usually very pure, with the exception of carbon content, and, from some older refineries, iron can be found in the 50–150 g t^{-1} range. Care must be taken during sample preparation not to contaminate the sample. Small amounts of cutting tools can abrade off into the sample during preparation. No lubricants should be used. If, by mistake, a cutting lubricant has been used, it *must* be removed with an appropriate solvent, such as acetone, prior to analysis.

4.1.2 Sampling and Sample Preparation of Nickel Cathodes

Whole cathodes weigh about 50 kg. ISO 7156 provides a detailed sampling scheme. Briefly, a random number of whole cathodes are selected, and each is sampled in five places along the diagonal from the top corner to the center of the cathode. The samples may be obtained by drilling completely through the cathode, milling a hole through the cathode or punching a hole through the cathode and milling the punch piece. The recommended hole diameter is between 15 and 25 mm. The drillings or millings are then combined and blended. If more than one laboratory is to analyze the sample, the chips should be divided by riffling, alternate shoveling or coning and quartering.

Sheared cathode pieces and round pieces produced for the plating industry are shipped in drums of 50, 250 or 1000 kg capacity. A number of drums are randomly selected and one or more pieces are taken from each drum, machined or drilled, and the chips combined and blended to form the analytical sample.

Cathode nickel may be very pure, comparable to carbonyl nickel. Less-pure cathode nickel is also produced. Regardless of the source of the nickel, care must be taken during sample preparation not to contaminate the sample. Small amounts of cutting tools can abrade off into the sample during preparation. No lubricants should be used. If, by mistake, a cutting lubricant has been used, it *must* be removed with an appropriate solvent, such as acetone, prior to analysis.

4.1.3 Sampling and Sample Preparation of Briquettes and Rondelles

An appropriate number of pieces are taken from randomly selected drums (see ISO 7156). The pieces, weighing typically 50–100 g depending on the producer, are then machined or otherwise broken into chips and the chips blended and riffled to obtain a final sample of about 500 g. Care must be taken during sample preparation not to contaminate the sample. Small amounts of cutting tools can abrade off into the sample during preparation. No lubricants should be used. If, by mistake, a cutting lubricant has been used, it *must* be removed with an appropriate solvent, such as acetone, prior to analysis.

4.1.4 Sampling and Sample Preparation of Granular Nickel (Including Fine Shot, Nickel Metal Powders)

The material is usually shipped in drums, of 50, 250 or 1000 kg capacity. Depending on the number of drums in question, a random number are selected (see for example ISO 7156), the entire contents of the drums are emptied, blended as well as possible and the mass is reduced to a reasonable size by alternate shoveling or riffling. About 200 g should be sufficient for powders and 500 g for granular nickel and fine shot. This lengthy procedure is necessary since the contents of the drums will segregate in transit. A grab sample, obtained by scooping from a drum or drums, may suffice for spot checking. No particle size reduction is required.

4.2 Determination of Trace Impurities

Once in the laboratory, the analyst should reblend the entire sample to minimize segregation effects. A knowledge of the history of the sample is advantageous. The more homogeneous and purer the material, the smaller the actual analytical sample can be. The minimum sample size should be 1 g. For impure and/or heterogeneous material, a larger weight should be taken (10–15 g).

Trace elements in nickel metal are often <0.001% in concentration. For this reason, great care must be taken during the sample preparation and dissolution so as not to introduce contaminants. As an example, touching the sample with bare skin will contaminate it with sodium. If surface contamination is suspected, either through touching or from tools used to generate chips, a quick rinse with warm, dilute HCl (also known as pickling the sample) followed by a thorough rinsing with water will remove virtually all such contamination.

4.2.1 Dissolution of the Sample

The sample is dissolved with 1:1 nitric acid for the determination of trace-metal impurities. Care must be taken at this stage to add the acid in small increments, allowing the reaction to subside between additions. *Failure to follow this procedure can lead to an extremely vigorous reaction, foaming over the top of the beaker, and possibly spraying acid on to the analyst. This is particularly true for powders.* Very high-purity nickel chips may be much slower dissolving and gentle heating may be required to initiate and/or maintain the reaction. Occasionally a sample will dissolve slowly at first, then accelerate very rapidly with foaming. It is wise to have a squirt bottle of water at hand to quench vigorous reactions.

For impure materials, there will be some insoluble matter left after the metal has dissolved. This should be filtered off and retained for further analysis. For carbonyl powders, there may be considerable carbon which should also be removed by filtration. The dissolved portion should be transferred to a clean volumetric flask, so that the final nickel concentration is 2–10% w/v (this varies from laboratory to laboratory, depending on how the analysis is then conducted).

4.2.2 Determination of Trace Impurities

Once into solution, the analysis is commonly carried out using several different techniques, depending on the element and its concentration range. Before the introduction of ETAS into common use in the mid-1970s, trace elements were determined by FAAS or by solvent extraction/colorimetric methods [e.g. Co in nickel by solvent extraction/colorimetry,[67,68] Sn by ion exchange/colorimetry[69]]. These methods were proving unsatisfactory owing to their relative insensitivity (FAAS) and length of time required (colorimetry). Refiners were improving their processes to produce better quality nickel, and a number of attempts were made to improve analytical detection limits. Preconcentration methods were

proposed including extraction of traces with trioctylphosphine oxide (TOPO) into methyl isobutyl ketone (MIBK) and analyzing by FAAS,[70–73] coprecipitation of Pb and Bi with manganese dioxide[74] and extracting several trace elements with sodium diethyldithiocarbamate followed by FAAS.[75] As these procedures were gaining acceptance, reliable ETAS instruments and lamps (electrodeless discharge lamps for Bi, Se, Te, As and Sb) became available, and the preconcentration work was abandoned. Direct injection of the dissolved sample into an electrothermal atomic absorption spectrometer is very simple, eliminates sources of contamination, is relatively fast,[76,77] and can be automated. Calibration is usually done by the method of standard additions.

Today a combination of FAAS, ETAS, ICPOES, ultraviolet/visible (UV/VIS) spectrophotometry, combustion and fusion instruments is required for the complete analysis of nickel metal, and the concentration level of the impurity can dictate the technique chosen. For example, Shvarts et al.[78] used FAAS to determine Co, Cu, Fe, Pb and Zn. Kiyokawa et al. determined Si by distillation/colorimetry,[79] but this procedure is not suited for routine use.

At the time of writing, at least one laboratory was still using an optical emission spectrograph (the method detailed in Lewis et al.[1]) for several elements, as well as the other techniques.

Analysis by neutron activation[80,81] is not easy owing to interferences from ^{58}Co, necessitating the removal of the Co by solvent extraction and counting the aqueous phase.

Taylor et al.[82] determined low levels of Co by ion-exchange/ICPOES and ETAS. ICPMS has also been tried.[83] The difficulty with this technique is the low tolerance for dissolved solids.

Interestingly, the multielement analysis capabilities of ICPOES have led to renewed interest in preconcentration techniques with lanthanum hydroxide[84] or yttrium hydroxide to isolate many of the trace elements. The metal sample is dissolved in nitric acid and a small amount of La or Y is added. Excess ammonia solution is added to precipitate the La or Y as a floc and form a soluble nickel ammine complex. The floc adsorbs many trace impurities. The floc is filtered off, dissolved in a small amount of nitric acid and diluted to a relatively small volume. Frequently the equivalent of 10 g of initial sample may be concentrated in 10 mL. The solution is then analyzed by ICPOES. Inco prefers the use of yttrium since it has a simpler spectrum in ICPOES and first applied it to the analysis of high-purity copper.[85] Inco subsequently applied the Y collection/ICPOES procedure to the analysis of nickel. The yttrium hydroxide collection/ICPOES method has been in routine use at Inco's Copper Cliff refinery for a number of years and is under review by ISO TC155/SC3/WG5 as a proposed routine method of analysis at the time of writing this article. Inco is currently considering ICPMS[86] with Y collection as a substitute for ICPOES since it permits the analyst to use smaller samples to achieve comparable detection limits. Smaller samples require less handling time, particularly at the sample dissolution and precipitation steps. The collection procedure should be used with caution since not all of the traces are collected quantitatively. Recovery studies must be made and correction factors introduced in order to obtain accurate results. This is especially true for Pb, where the collection efficiency tends be erratic. Calibration curves are produced by the method of standard additions applied to solutions of high-purity nickel, and then carried through the collection steps.

ISO methods are available for the FAAS determination of Ag, Bi, Cd, Co, Cu, Fe, Mn, Pb and Zn (ISO 6351) and for the ETAS determination of Ag, As, Bi, Cd, Pb, Sb, Se, Sn, Te and Tl (ISO 7523).

Table 3 sets out the elements determined and recommended methods of analysis for high-purity nickel. For less pure samples, it may be necessary to dilute to sample solution, or to choose a less sensitive technique, e.g. FAAS instead of ETAS.

Carbon and sulfur (and oxygen, nitrogen and hydrogen if desired) are determined on separate analytical samples by combustion (or fusion) instruments. The detectors for most carbon and sulfur determinators are IR cells (see e.g. ISO 7524 for the determination of carbon and ISO 7526 and 7527 for the determination of S). For the gases, gas chromatography (GC) with dedicated gas chromatographs is used to measure the evolved gases after heating (fusing in a graphite crucible) at high temperatures. For pellets and cathode nickel, the gases are better determined on small pieces, about 1 g, cut from larger pieces or pellets. Several determinations should be made using samples from different pieces and the results averaged. Powders are not normally analyzed for gases owing to the large surface areas.

A sensitive method for the determination of sulfur is by dissolution of a sample in a mixture of nitric and hydrochloric acid, removal of nitrates, reduction of the sulfate to H_2S and distillation into a solution containing *N*,*N*-dimethyl-*p*-phenylenediamine to form methylene blue. Contamination from sulfide samples elsewhere in the laboratory is a major concern. Consequently, this method is not in widespread use. For details of the method, see ISO 7525.

Another possibility is the determination of S by ICPMS, although the literature survey did not uncover any reference for this.

Occasionally other elements may be requested, as listed in Table 4.

Table 3 Recommended methods for the analysis of nickel metal

Element	Recommended method	Alternative method
Aluminum	ICPOES/Y collection	ICPOES
Antimony	ICPOES/Y collection	ETAS
Arsenic	ICPOES/Y collection	ETAS
Barium	ICPOES	
Beryllium	ICPOES/Y collection	ICPOES
Bismuth	ICPOES/Y collection	ETAS
Boron	ICPOES/Y collection	
Cadmium	ETAS	
Calcium	ICPOES	
Carbon[a]	Combustion/IR	
Chromium	ICPOES/Y collection	ICPOES
Cobalt	ICPOES	
Copper	ICPOES	
Gallium	ICPOES/Y collection	ETAS
Germanium	ICPOES	
Hydrogen[a]	Fusion/GC	
Indium	ICPOES/Y collection	
Iron	ICPOES/Y collection	
Lead	ETAS	ICPOES/Y collection
Magnesium	ICPOES	
Manganese	ICPOES/Y collection	ICPOES
Molybdenum	ICPOES/Y collection	ICPOES
Nitrogen[a]	Fusion/GC	
Oxygen[a]	Fusion/GC	
Phosphorus	ICPOES/Y collection	Mo blue after extraction
Potassium	FAAS	
Selenium	ICPOES/Y collection	ETAS
Silicon	ETAS	ICPOES
Silver	ETAS	ICPOES
Sodium	FAAS	
Strontium	ICPOES	
Sulfur[a]	Combustion/IR	
Tellurium	ICPOES/Y collection	ETAS
Thallium	ICPOES/Y collection	ETAS
Tin	ICPOES/Y collection	ETAS
Titanium	ICPOES/Y collection	ICPOES
Vanadium	ICPOES/Y collection	ICPOES
Zinc	FAAS	ICPOES

[a] Determined on solid samples.

Refractory elements are rarely requested. ICPMS is the simplest technique to use. Otherwise, the sample must undergo a preconcentration step whereby a large sample weight is subjected to treatment with CO to remove nickel as the carbonyl, leaving the refractories behind in the residue. *This is extremely hazardous, both from a toxicity point of view and from the risk of explosions*.

4.3 Quality Control

Procedure blanks and reference materials should be analyzed with the batch of samples. Contamination is the greatest concern at the very low levels of trace impurities encountered in high-purity nickel. Reference materials with impurities at current production levels may be obtained commercially from Brammer Standards and Analytical Reference Materials Inc. Well-analyzed secondary standards (real samples of very high-purity nickel) may also be used. The National Institute of Standards and Technology (NIST) has three nickel oxide reference materials, but owing to their high impurity levels they are not recommended.

Table 4 Recommended methods for miscellaneous elements in nickel metal

Element	Recommended method	Alternative method
Chlorine	Turbidimetry using $AgNO_3$	
Hafnium	ICPMS	Preconcentration/ICPOES
Niobium	ICPMS	Preconcentration/ICPOES
Tantalum	ICPMS	Preconcentration/ICPOES
Thorium	ICPMS	Preconcentration/ICPOES
Tungsten	ICPMS	Preconcentration/ICPOES
Uranium	ICPMS	Preconcentration/ICPOES

5 ANALYSIS OF NICKEL OXIDES

Some refineries produce crude nickel oxide sinters as their final product. Others produce high-purity nickel oxide which is made for sale to the chemical market. The two must not be confused.

Crude nickel oxide sinters have varying amounts of nickel (70–90%) and considerable gangue, mainly silica, alumina, calcia and magnesia. The concentrations of other impurities, particularly cobalt and copper, are much higher than in refined nickel metal. Such material is sold in small irregular lumps, and is used to make low-alloy steels. The gangue will slag off, and the relatively high impurities will be well diluted.

High-purity nickel oxide is sold as a powder mainly into the chemical market and consequently has impurities at very low levels, comparable to those found in high-purity nickel metal.

5.1 Sampling and Sample Preparation

5.1.1 Nickel Oxide Sinters

The lumps may be sampled in a manner similar to that described in section 4.1.3. The lumps are generally friable and the analytical sample can be prepared by crushing and blending.

5.1.2 High-purity Nickel Oxide

This material can be sampled in the same way as nickel metal powder described in section 4.1.4.

5.2 Sample Dissolution and Analysis

5.2.1 Crude Sinters

Crude sinters are produced by high-temperature roasting of nickel sulfide and are very refractory. They may be attacked with warm, dilute (1 : 1) nitric acid. Considerable insoluble material may be left, which should be filtered off, fused with sodium peroxide and analyzed separately by FAAS or ICPOES for the gangue elements. The other trace elements can be analyzed similarly to nickel metal in the nitric acid-soluble portion, although there may be dilutions required, or alternative, less-sensitive techniques substituted (e.g. FAAS instead of ETAS). Another approach is to dissolve the sinter by fuming with perchloric acid, with filtration and fusion of the insoluble residue (mainly silica). The combined solution can be analyzed for major impurities and gangue elements by FAAS or ICPOES.

5.2.2 High-purity Nickel Oxide Powder

Dissolution is accomplished with warm, dilute (1 : 1) nitric acid. There should be very little, if any, insoluble residue left. The solution can be analyzed in the same manner as nickel metal above.

5.2.3 Determination of Nickel in Nickel Oxides

The material is sold on the basis of nickel content. Ni can be determined by the classical precipitation with DMG, with a gravimetric or EDTA titration finish. Ni can also be determined gravimetrically by the electrolytic deposition method. This method is in use in production laboratories, and has been accepted by ISO as an umpire method (ISO 12169).

5.2.4 Determination of Oxygen in Nickel Oxides

Nickel oxides, regardless of purity, have oxygen as a major element. Commercial oxygen determinators work best at low (parts per million) levels, not at 20%. A good way to determine oxygen is by gravimetry. A sample is reduced under hydrogen at 400 °C in a tube furnace, and the weight loss is assumed to be oxygen. Residual oxygen associated with gangue matter can then be determined by a commercial oxygen determinator, and the results combined give the total oxygen present in the sample.

5.3 Quality Control

For high-purity nickel oxide, the same quality control criteria cited under the analysis of nickel metal apply. For nickel oxide sinters, the NIST nickel oxide reference materials may have some use.

6 FERRONICKEL

Ferronickel is produced by smelting laterite concentrate and ranges from 15 to 50% nickel, the balance being iron, with minor and trace impurities. It is used in the stainless-steel industry, where the producers obtain high-quality iron units, often at no cost, along with the nickel. Ferronickel is produced as ingots, lumps or granules.

6.1 Sampling and Sample Preparation

6.1.1 Granules

ISO 8049 sets out a rigorous procedure for the sampling of ferronickel granules. Once the containers (trucks, drums, etc.) have been sampled, the granules are blended and divided by riffling or alternate shoveling until the mass is 1–5 kg, depending on the shipment size. The granules are melted in 1-kg batches, preferably in an induction furnace under argon, and sample bars are cast. These bars can be sliced to obtain disks for XRF analysis, or be drilled or milled to obtain chips for wet chemical analysis and for the determination of carbon and sulfur.

Caution: Some ferronickels are *very* hard, and appropriate tools must be selected. If cutting lubricants must be used, then the sample must be washed with a suitable solvent such as acetone to remove the lubricant, and then dried to remove the solvent.

6.1.2 Ferronickel Ingots or Pieces

ISO 8050 gives a good description of the sampling protocol. Briefly, if the sampling is done during casting, a spoon sample is poured into a suitable mold to form a bar. The bar may be sliced and/or milled or drilled as described in section 6.1.1.

When sampling ingots or pieces, there is a different protocol depending on whether or not the shipment is from the same heat or from more than one heat. If the shipment is from one heat, then only five ingots or pieces are randomly selected. If there is more than one heat present or suspected to be present in the shipment, then the number of ingots or pieces selected must be increased dramatically. The number depends on the lot tonnage. ISO 8050 provides a table for guidance.

Once the ingot or pieces have been selected, the surface must be cleaned before drilling or milling. Since ferronickels can be very hard, annealing the ingots or pieces may make drilling easier. If lubricants are used during the drilling or milling, the chips must be washed with a suitable solvent such as acetone, then dried and blended.

For lots comprised of only one heat, there is no need to do rigorous blending. In the case of more than one heat, the sample must be homogenized. If the nickel content is <35%, the sample may be brittle enough

to crush. This can be done in a laboratory vibratory crusher, preferably made of tungsten carbide. The fine particles are thoroughly blended. Ferronickels with nickel contents >35% are too ductile to be crushed and the drillings or millings must be blended as well as possible. In extreme cases, it may be necessary to melt the chips under argon to cast a homogeneous bar as described in section 6.1.1.

6.2 Analysis of Ferronickel

6.2.1 Determination of Nickel Content

Ferronickel is sold on the basis of nickel content. Consequently, the determination of nickel is the most critical analysis.

6.2.1.1 Routine Determination of Nickel For routine work, Ni is best determined by XRF using slices cut from sample bars. The calibration curve may be generated by using well-analyzed secondary standards. NIST SRM 1158, containing about 40% Ni (the actual Ni percentage varies depending on the version of the SRM), is an excellent standard.

The sample slices are often rough and must be polished to remove burrs and surface blemishes. This can be done using metallographic polishing wheels or belts and various grits of polishing paper. If other elements are to be determined, such as Si or Al, then diamond polishing paper should be used in order not to contaminate the surface of the disk.

Caution: The surface of the calibration standards and the surface of the samples must be almost identical. If the standards are polished to a matte finish, then that is what must be done to the samples. If a mirror finish is used, then all must be polished in the same way.

Khlopina et al.[55] described the use of XRF in a Yugoslavian ferronickel plant. Van Peteghem[87] used a fundamental parameters calibration to determine the nickel content of ferronickel samples. Aleksandruk et al.[88] determined Ni and several other elements by XRF. Lakefield Research Limited[89] has determined nickel in ferronickel disks routinely for several years, in excellent agreement with results obtained by the umpire method described in section 6.2.1.2.

If XRF is not an option, then chips should be dissolved in warm, dilute (1 : 1) nitric acid and the solution analyzed by FAAS or ICPOES after appropriate dilutions. This will not produce as accurate a result as the XRF procedure described above.

6.2.1.2 Umpire Determination of Nickel ISO 6352 describes the determination of nickel in ferronickel by the DMG gravimetric procedure. This is the classical method for the determination of nickel, and variations of the method can be found in textbooks. After dissolution with nitric acid, the sample solution is taken to strong fumes of perchloric acid to dehydrate silica, which is removed by filtration. Nickel is then determined on the filtrate. Tartaric acid is used to complex iron, and sodium thiosulfate is added to complex copper if it is thought to be present. The NiDMG precipitate is filtered through tared sintered-glass crucibles and dried to constant weight at 150 °C. This temperature is fairly critical. The drying temperature must be high enough to remove any absorbed excess DMG but not too high as to decompose the NiDMG precipitate.

An alternative to the gravimetric finish is to wet ash the NiDMG precipitate with nitric and perchloric acids and titrate the nickel with EDTA using murexide as the indicator. In this case, silica does not have to be removed.

Appropriate precautions should be taken when using perchloric acid.

In either case, the filtrates from the DMG precipitations must be checked for Ni by FAAS or ICPOES and a correction made to the final result.

6.2.2 Determination of Minor and Trace Elements

6.2.2.1 Determination of Carbon and Sulfur C and S are determined on chips using combustion techniques. Commercial instruments use IR detectors for the most part. Older sulfur determinators may use an iodimetric titration finish.

6.2.2.2 Determination of Minor Elements in Ferronickel Most ferronickels contain small amounts of Si, Co, Cu, Cr, Al, P and Zn. The most convenient way to determine these elements is by XRF on the same disk as is used for the routine determination of Ni. If Si and Al are determined on the disk, then the polish should be done with diamond paper to avoid surface contamination. If this option is not available, these elements can be determined by FAAS or ICPOES after dissolution with nitric acid. See ISO 7520 for the determination of Co in ferronickel by FAAS.

In some materials, the Si content is high (>1%), and not all may dissolve. Such samples are best dissolved in nitric acid, then taken to strong fumes with perchloric acid to dehydrate silica, which is removed by filtration. The silicon content of the material can then be determined gravimetrically (see ISO 8343) or by FAAS or ICPOES after fusion with sodium peroxide.

6.2.2.3 Determination of Trace Elements in Ferronickel The determination of trace elements differs from pure nickel metal in that hydroxide collection procedures are ineffective owing to the large amount of iron present in the material. The list of trace elements requested is usually shorter. In general, combustion/IR, ICPOES, FAAS and

ETAS techniques are used. Se can be a problem owing to the presence of large amounts of iron. ETAS instruments using deuterium background correction cannot determine Se below about $5\,g\,t^{-1}$ (0.0005%). Those equipped with Zeeman or Smith–Hieftje background correction do not suffer from the same interference from iron. For laboratories whose instruments only have deuterium background correction, it is necessary to separate Se by collection with As.

ISO 11438 gives details of the determination of Pb, Sb, Sn, Te, Tl, Ag and In by ETAS. Arsenic is often requested since at least one ferronickel producer used laterite ore with a high As content. That ferronickel contained approximately 0.1% As. Most other ferronickels contain <0.01% As.

Table 5 Major and minor elements in ferronickel routinely determined by XRF

Nickel	Copper
Aluminum	Phosphorus
Chromium	Silicon
Cobalt	Zinc

Table 6 Recommended methods for the determination of minor and trace elements in ferronickel

Element	Recommended method	Alternative method
Aluminum	ICPOES	FAAS
Antimony	ETAS	
Arsenic	ETAS	
Bismuth	ETAS	
Cadmium	ETAS	
Calcium	ICPOES	
Carbon[a]	Combustion/IR	
Chromium	ICPOES	FAAS
Cobalt	ICPOES	FAAS
Copper	ICPOES	FAAS
Gallium	ETAS	
Indium	ETAS	
Lead	ETAS	
Magnesium	ICPOES	
Manganese	ICPOES	FAAS
Nitrogen[a]	Combustion/GC	
Oxygen[a]	Combustion/GC	
Phosphorus	Mo blue after extraction	
Selenium	ETAS	
Silicon	ETAS	ICPOES
Silver	ETAS	
Sodium	FAAS	
Sulfur[a]	Combustion/IR	
Tellurium	ETAS	
Thallium	ETAS	
Tin	ETAS	
Zinc	FAAS	ICPOES

[a] Determined on solid samples.

6.3 Summary

Table 5 sets out elements routinely determined by XRF.

Aleksandruk et al.[88] determined Cu, Co, Cr, Si, P and S by XRF. If the concentration level is too low to give an accurate result by XRF, then wet chemical methods should be used. C and S are best determined by combustion methods.

Table 6 sets out minor and trace elements commonly determined wet chemically and the techniques used.

7 ANALYSIS OF NICKEL ALLOYS

7.1 Introduction

The term ''nickel alloys'' covers a wide range of materials, from stainless steels containing 10–20% Ni, to heat-resistant alloys containing 30–40% Ni, to alloys where Ni is the major (>50%) constituent. There are a large number of major, minor and trace elements associated with each alloy. Since this is a very broad topic, it will be discussed in general terms, focusing on concentration ranges. Owing to the wide range of alloys, there is no single acid or acid mixture which can be used to dissolve every alloy. Where acids are recommended, it must be understood that for a particular sample, a modification of the acid or acid mixture may achieve better results. Nickel alloys are often designed to be corrosion resistant, so the dissolution may be very slow.

Owing to the difficulties of dissolution and complexities of wet chemical analysis of nickel alloys, XRF is the most widely used technique for the routine analysis of nickel alloys.

7.2 Sample Preparation

7.2.1 Chips and Drillings

Chips or drillings are commonly taken for wet chemical analysis and for the determination of C and S. Nickel alloys are often hard, and tool steels or tungsten carbide bits are used. Care must be used so as not to introduce contamination from the bit or tool into the sample.

7.2.2 Solid Pieces

For XRF analysis, the sample is often a slice cut from a test bar cast during the pouring of the alloy heat. Otherwise, a piece is cut from the part or ingot and then prepared for analysis.

For spark analysis in the field, the surface of the piece or part is simply cleaned before applying the instrument.

7.3 Determination of Major Constituents

7.3.1 By Classical Wet Chemistry

The major elements usually determined are Ni, Fe, Co, Cr and Cu (for cupronickels). Although there is no universal solvent, the best general approach is to add a small amount of water to the sample, then add diluted (1 : 1) aqua regia. The addition of water prior to the addition of the aqua regia prevents passivation of the surface of the alloy. For some alloys, particularly those containing refractory elements, the addition of small amounts of hydrofluoric acid may be required.

7.3.1.1 Nickel Interestingly, at the time of writing, there is no consensus amongst the ISO committee on the analysis of nickel and nickel alloys (TC 155) on how best to determine Ni in nickel alloys. The best approach is to dissolve the sample in aqua regia (adding HF if refractory elements are present), and precipitating Ni with DMG as described in section 6.2.1.2. The NiDMG precipitate is often off-color, perhaps brown, and does not have the same fluffy texture of a clean precipitate. This is due to the sorption of impurities, particularly Co by the precipitate. After filtration, the precipitate is dissolved in HCl and the DMG precipitation is repeated. This usually gives a clean NiDMG precipitate which can be weighed or titrated as described above. The filtrates from the two precipitation steps are combined and are analyzed for the small amounts of Ni which do not completely precipitate. This method has been used for several years for a very wide variety of alloys by Lakefield Research Limited,[90] covering the concentration range 10–70% Ni.

A conductometric titration has also been used for the determination of nickel.[91]

7.3.1.2 Iron The sample is dissolved in aqua regia, Cr is oxidized to Cr(VI), and Fe is separated from much of the matrix by precipitation with ammonia solution. The Fe is determined by redox titration with potassium dichromate (ISO 7528, covering the range 1–50% Fe).

7.3.1.3 Cobalt The sample is dissolved in aqua regia, then fumed with perchloric acid, adding small amounts of HCl to remove Cr as the volatile chromyl chloride. Co is determined by potentiometric titration using excess potassium iron(II) cyanide and back-titrating the excess with standard Co (ISO 9389). The original ISO work covered the range 2–20%, but higher concentrations (at least 65% Co) can be determined by this method. Low levels of Mn (<1%) can be complexed with pyrophosphate. For higher levels of Mn, the procedure detailed in ASTM E 353 should be followed. Here, Co is separated from the matrix by absorption on an ion-exchange resin, whilst Mn, Ni, Fe, Cr, etc. elute off. The Co is subsequently eluted and titrated potentiometrically.

7.3.1.4 Chromium The sample is dissolved in aqua regia, then fumed with sulfuric acid to remove chlorides and nitrates. Cr is oxidized to Cr(VI) and is determined by potentiometric titration with ammonium iron(II) sulfate solution. Interference from Mn is overcome by reducing the Mn(VII) with HCl. Vanadium will also titrate, giving a positive error. A correction can be made by subtracting the V content multiplied by 0.340. V can be determined by FAAS, described below. ISO 7529 covers the range 1–25% Cr. The method can be extended to higher Cr concentrations by adjusting the sample weight.

The titration can also be carried out as a conventional redox titration, with diphenylamine sulfonate as the indicator. The end-point may be off-color, depending on the alloy being analyzed. The analyst may require practice with synthetic sample solutions to discern the end-point.

Some alloys, particularly those with about 35% Ni, 30% Cr and the balance Fe, can be dissolved by gentle heating in 1 : 4 H_2SO_4. This may take several hours, depending on the coarseness of the drillings or millings.

7.3.1.5 Copper Copper can be determined gravimetrically, e.g. in German silver, by electrodeposition on a platinum cathode.[92] Alternatively, Cu can be titrated iodimetrically, with any Fe present being complexed with fluoride.[93]

7.3.2 By X-ray Fluorescence Spectroscopy

XRF is by far the most widely used routine method for the analysis of nickel alloys. XRF gives fast and accurate results for the major (and minor) elements in nickel alloys. It is the preferred technique in foundries, since heats can be analyzed while still in the furnace, and adjustments in the chemistry can be made before pouring the heat. This is particularly true if a simultaneous spectrometer is used.

Since the need for elaborate wet-chemical facilities is obviated, many scrap dealers prefer energy-dispersive XRF spectrometers for the analysis of metals, and can sort them quickly into different grades.

The sample should be polished to remove burrs and surface blemishes, then taken to a mirror finish, if possible (for best results). If the alloy type is known, the instrument can be calibrated with commercial or in-house secondary standards of the same matrix. If the alloy is unknown, or is a scrap sample, perhaps a mixture of several alloys, then it is best to use a fundamental parameters type of calibration.[94] There are a variety of software packages available, and it is best to consult with the instrument manufacturer and/or a laboratory currently using such a calibration to see what software would be best for the application. In part this may depend upon whether the XRF spectrometer is wavelength- or energy-dispersive. The spectrometer manufacturer should be consulted about the application(s) and calibration standards.

Bettinelli et al. compared XRF to ICPOES.[95] Brachfeld et al. described the determination of Cr in Ni alloys.[96] Other workers have described the use of XRF for a wide variety of major and minor elements.[97,98]

For heterogeneous samples, a sufficient mass should be melted in an induction furnace, or in an arc furnace, in order to produce a homogeneous sample for analysis. If the surface area is too large relative to the mass of the sample (e.g. long, stringy drillings), it is common to start the melt with a ''heel'' or pieces of high-purity metal with sufficient mass to couple and melt. The drillings may then be added to the molten bath. The heel metal is an element not to be determined, and the amount added must be taken into consideration for dilution calculations. Common heel metals are Cu or Fe.[99,100] The melt is poured into a cylindrical mold, cooled and a disk is cut for analysis.

This is common in many scrap yards where a material of considerable variety, both in composition and in shape and size, may constitute a lot. For these ''alloys'', fundamental parameter XRF software programs are necessary, since no SRM will match the actual sample.

7.3.3 By Spark

This technique is most often used in the field to identify unknown or mislabeled material. The surface of the casting is cleaned by an appropriate means, and a small amount of material is sparked from the surface. The sparked particles are carried as an aerosol to an ICPOES.[101] Alternatively, the light emitted from the sparking is carried by fiber optics to a dedicated spectrograph. The material is analyzed for a suite of elements, and the alloy is determined by computer matching with known alloys. With homogeneous samples and good calibration standards, quantitative results can be obtained.

Okochi et al.[102] used the spark method and a spectrograph to determine 13 elements in Ni alloys, and later used spark ablation and ICPOES.[103]

7.3.4 By Inductively Coupled Plasma Optical Emission Spectroscopy and Flame Atomic Absorption Spectroscopy

As with any wet chemistry technique, the most time-consuming step in the analysis is often the dissolution of the sample. This is particularly true for nickel alloys which are designed to be corrosion resistant. For this reason, most foundries prefer XRF as the prime technique for the analysis of their materials. For laboratories without XRF, ICPOES and/or FAAS is a good second choice.

The sample is dissolved by adding a small amount of water, then dilute aqua regia. This prevents passivation of the metal surface and speeds dissolution. In some cases HF may also be required. The sample solution may then be analyzed. If HF has been used, a corrosion-resistant nebulizer should be used for ICPOES, or the HF must be removed by fuming with $HClO_4$ or H_2SO_4. If the solution is taken to fumes, it should not undergo prolonged heating, otherwise refractory element oxyanions may form and precipitate. These can be difficult to redissolve.

Owing to the ability of ICPOES to analyze samples over wide concentration ranges, the technique is reasonably popular for the analysis of alloys for major, minor and occasionally trace elements.[94,104–107] The limiting factor is the dissolution of the sample and the possible presence of HF in the solution, which will necessitate changing nebulizers.

FAAS does not present the same problems with respect to HF that affects ICPOES. With proper dilutions, major and minor elements can be determined.[108] Kometani discussed the analysis of Cu–Ni–Sn spinodal alloys by FAAS.[109]

7.4 Minor Elements in Nickel Alloys

For the purposes of this article, minor elements are those whose concentrations fall in the range 0.01–10%.

7.4.1 By Flame Atomic Absorption Spectroscopy and Inductively Coupled Plasma Optical Emission Spectroscopy

ISO 7530 describes the FAAS determination of Al, Co, Cr, Cu, Fe and Mn up to 4%, Si up to 2% and V up to 1% using an aqua regia dissolution with $SrCl_2$ added as an ionization suppressant. The sample solution could also be used for ICPOES analysis. Some alloys may require HF to achieve complete dissolution, in which case Si cannot be determined unless the dissolution is done in a closed vessel.[108]

Peterson reviewed the FAAS analysis of alloys, including Ni alloys.[110]

Papers have described the determination of B, Cr, Fe and Si[108] and Co, Cr, Al, Ti, Mo, W, V, Fe, Nb and Mn.[111] Ni has been determined,[112] and also W.[113]

Refractory elements are generally not well determined by FAAS, particularly W. This is owing to the formation of oxyanions in solution and the difficulty in breaking the metal–oxygen bond in the flame of the spectrometer.

ICPOES, however, can give reasonably good results for Hf, Mo, Nb, Ta, Ti, W and Zr. Some laboratories use Sc as an internal standard to improve accuracy. Sample dissolution in aqua regia or aqua regia–HF is used, depending on the alloy. If HF is used, a corrosion-resistant nebulizer must be used.

For levels below 1%, ICPOES is a good technique for the determination of Al, Ca, Co, Cu, Fe, Mn, Mg, Ti and V. These elements can also be determined by FAAS, but ICPOES has the advantage of rapid multielement analysis. There is no universal solvent for nickel alloys. For these elements, a mixture of aqua regia and HF, followed by fuming with $HClO_4$, is a good starting point.

ICPOES is also very good for the determination of Ce and rare-earth elements in Ni alloys.[114,115]

7.4.2 By X-ray Fluorescence Spectroscopy

The preparation and analysis are identical with those described in section 7.3.2.

Brachfeld[116] determined Al and W in Ni alloy powder. The powder was heterogeneous and was melted into a button with the addition of pure Ni metal.

Care must be taken to choose calibration standards with a similar matrix to that of the samples. British Chemical Standards (BCS) or the European Standards in Europe and NIST in North America are good sources of well-analyzed SRMs.

7.4.3 By Classical Methods

Classical methods may still find favor for umpire analyses. They give accurate results, but are far slower than the techniques discussed above.

ISO TR11434 describes the gravimetric determination of Nb and Ta. An aqua regia–HF sample solution is passed through an ion-exchange column to isolate Nb and Ta. Nb is then eluted, followed by Ta. The elements are precipitated using cupferron and ignited to the oxides. Alternatively, Ta may be determined colorimetrically using pyrogallol.

ISO 11433 describes the determination of Ti colorimetrically using diantipyrylmethane in the range 0.4–5% Ti. The samples were dissolved in aqua regia and chloride and nitrate were removed by fuming with H_2SO_4.

Colorimetry has been used to determine Zr,[117] Zr and Hf,[118] Mo[119] and Nb and Ta.[120] With the advent of ICPOES, these methods are rarely used today.

Mu and Zhang[121] titrated Mg with EDTA in nickel hydrogen storage battery material. Meryan et al.[122] determined Mg by solvent extraction followed by polarography.

Pilipenko et al.[123] determined Cu coulometrically after precipitation with cinnamoylphenylhydroxylamine.

7.5 Trace Elements in Nickel Alloys

For this article, trace elements are those with concentrations <0.01%.

Since trace elements can have a catastrophic effect on nickel alloys, there has been much reported in the literature, particularly with the development of ETAS.

ETAS is the method of choice for many trace elements (Se, Te, Bi, Pb, Tl, As, Sb, Ga, Al, etc.).[124–144] A typical dissolution is done with $HF–HNO_3–H_2O$ (1 : 1 : 1). It is best to prepare a fresh acid mixture for each shift.

To overcome problems with the dissolution of alloys, ETAS has been applied to solid chips.[145–148] If this procedure is used, the chips are weighed on a microbalance and then placed in the atomizer. It is critical to place the chips in the same place each time, since the tube has a heat gradient. Because of the difficulties in manipulating the solid chips, this procedure has not achieved the same popularity as the dissolution–injection procedures.

Electrothermal atomization/laser-excited atomic fluorescence spectroscopy using solid chips was reported by Irwin et al.[149] The authors indicated a long linear dynamic range and 1–4 orders of magnitude better sensitivity than conventional ETAS.

Recently, Hinds et al.[150] determined Bi, Pb and Te by volatilizing the solid sample in an electrothermal atomizer and analyzing the vapors by ICPMS. The results were reasonable for Bi and Pb, but some problems were encountered with Te.

In using ETAS, it must be noted that untreated graphite tubes may form refractory carbides if elements such as W, Ta, Nb, Hf, Zr are present in the alloy. This will change the heating characteristics of the tube over time. Therefore, it is recommended that a coated tube, preferably a carbide-coated type [see Zatka[76]] be used.

ISO 11437 describes the determination of Pb by ETAS.

Hydride generation furnace AAS has also been used.[151,152] Flow injection analysis speeds up the analysis compared with conventional ETAS.[124,153–155] Various chemical modifiers must be used to complex the major components in the alloy (Ni especially) in order to form the hydrides.

Hydride generation AAS has also been used,[156] and hydride generation ICPOES.[103]

FAAS can be used if a preconcentration step such as solvent extraction is introduced,[135,157,158] or directly if the element has sufficient sensitivity, such as Ca[159] or Zn.[160]

The use of the high-temperature hollow-cathode technique appears occasionally.[161–164]

ICPOES has been used to determine trace B in Ni alloys, although a separation step is required.[165]

Phosphorus is determined spectrophotometrically by extraction of the molybdate complex and subsequent reduction to Mo blue.[166,167] ISO methods have been

Table 7 Recommended methods for the analysis of nickel alloys

Component	Recommended method	Alternative method
Majors	XRF	Acid dissolution/ICPOES or FAAS
Minors	XRF	Acid dissolution/ICPOES or FAAS
Traces	Acid dissolution/ETAS	
C, S, gases	Combustion (C, S) or fusion	

developed for the Mo blue (ISO 9388) and the phosphovanadomolybdate complex (ISO 11400).

Boron may be determined spectrophotometrically using curcumin without distillation of boron as the methyl borate complex (ISO 11436). The sample must be digested in quartz vessels at a high enough temperature to decompose boron nitrides. The ISO method calls for fuming with a mixture of H_3PO_4 and H_2SO_4 in order to effect the decomposition. ISO 11436 describes the determination of B in Ni alloys by the curcumin spectrophotometric procedure.

Other applications of spectrophotometry are still encountered,[168–173] but for most cases the use of AAS or ICPOES is more practical.

De Pablos et al.[174] determined Ga by fluorimetry.

Electroanalytical techniques such as polarography for Cd,[175] voltammetry for Sn[176] and differential-pulse anodic stripping voltammetry for Zn, Cd, Pb and Bi[177] have been reported.

Spark-source mass spectrometry has been used for a number of trace elements, often those normally determined by ETAS.[178–180] Isotope dilution mass spectrometry has been used to determine B.[181]

XRF can be used for the upper range of trace-element concentrations (>0.002%), if the alloy standards match the samples. Below this nominal concentration, detection and accuracy suffer, owing to the heavy matrix.

Platinum is present in some superalloys. The sample, along with 15–20 mg of Te metal powder, is dissolved in a mixture of nitric, sulfuric and hydrofluoric acids. After boiling off the HF, Te is precipitated by the addition of $SnCl_2$. The Te precipitate collects the Pt. The precipitate is dissolved and Pt is determined by FAAS (La is added as an ionization buffer) or by ICPOES. Alternatively, the sample may be dissolved in a mixture of nitric, sulfuric and hydrofluoric acids, a small amount of copper is added and the Pt is cemented out with the addition of Zn powder. The precipitate is cleaned of excess Zn and is fused according to the nickel sulfide fire assay procedure or by the classical lead fire assay procedure and Pt is then determined as described in section 3.2.1.3.[182]

Danilova et al.[183] determined Ir in copper–nickel ores kinetically.

INAA has been used to determine Cl, P and S in alloys.[184]

7.6 Carbon and Sulfur in Nickel Alloys

The conventional combustion IR instrumentation is the most common method for the determination of C and S in alloys. The analysis is usually done using chips or drillings.

7.7 Gases in Nickel Alloys

The conventional fusion GC instrumentation is used to determine O_2 and N_2.[185,186] H_2 may also be determined in this way,[187] but is not requested as frequently as O_2 and N_2.

INAA has been used to determine O_2,[187] as has a discharge tube with a hot hollow cathode.[188]

Since high-nickel alloys are often melted in an inert atmosphere, that gas may have to be determined. Kondakova et al.[189] used vacuum fusion mass spectrometry to determine Ar in alloys.

7.8 Recommended Methods for the Analysis of Nickel Alloys

Table 7 lists recommended and alternative ways to analyze nickel alloys. Owing to the large and complex family of nickel alloys, it is impossible to be more specific within the constraints of this article.

ACKNOWLEDGMENTS

The author thanks Lakefield Research Limited for its generous assistance regarding the literature survey. Thanks are also due to his colleagues at Lakefield and to Dr J Bozic and Dr St J.H. Blakeley of Inco Ltd., who provided valuable information.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
ASTM	American Society for Testing and Materials
BCS	British Chemical Standards
BS	British Standard
CCRMP	Canadian Certified Reference Materials Project

DIN	Deutsche Institut für Normung
DMG	Dimethylglyoxime
EDTA	Ethylenediaminetetraacetic Acid
ETAS	Electrothermal Atomic Absorption Spectroscopy
FAAS	Flame Atomic Absorption Spectroscopy
GC	Gas Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICPOES	Inductively Coupled Plasma Optical Emission Spectroscopy
IGS	International Geological Survey
INAA	Instrumental Neutron Activation Analysis
IR	Infrared
ISO	International Organization for Standardization
LOI	Loss on Ignition
MIBK	Methyl Isobutyl Ketone
NIST	National Institute of Standards and Technology
SRM	Standard Reference Material
TOPO	Trioctylphosphine Oxide
UV/VIS	Ultraviolet/Visible
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Process Instrumental Methods **(Volume 9)**
Titration Techniques for Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Steel and Related Materials **(Volume 10)**
Noble Metals, Analytical Chemistry of

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals

Nuclear Methods **(Volume 14)**
Instrumental Neutron Activation Analysis

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Absorption Techniques in X-ray Spectrometry • Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Wavelength-dispersive X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Quantitative Spectroscopic Calibration • Traceability in Analytical Chemistry

LIST OF STANDARDS

International Organization for Standardization Documents

6283	Refined nickel
6351	Nickel – Determination of silver, bismuth, cadmium, cobalt, copper, iron, manganese, lead and zinc contents – Flame atomic absorption spectrometric method
6352	Ferronickel – Determination of nickel content – DMG gravimetric method
7156	Refined Nickel – Sampling
7520	Ferronickel – Determination of cobalt content – Flame atomic absorption spectrometric method
7523	Nickel – Determination of silver, arsenic, bismuth, cadmium, lead, antimony, selenium, tin, tellurium and thallium contents – Electrothermal atomic absorption spectrometric method
7524	Nickel, ferronickel and nickel alloys – Determination of carbon content – IR absorption method after induction furnace combustion
7525	Nickel – Determination of sulfur content – Methylene blue molecular absorption spectrometric method after generation of hydrogen sulfide
7526	Nickel, ferronickel and nickel alloys – Determination of sulfur content – IR absorption method after induction furnace combustion
7527	Nickel, ferronickel and nickel alloys – Determination of sulfur content – Iodimetric titration after induction furnace combustion
7528	Nickel alloys – Determination of iron content – Titrimetric method with potassium dichromate

7529 Nickel alloys – Determination of chromium content – Potentiometric titration method with ammonium iron(II) sulfate
7530 Nickel alloys – Flame atomic absorption spectrometric analysis (Part 1: general requirements and sample dissolution; Part 2: determination of cobalt content; Part 3: determination of chromium content; Part 4: determination of copper content; Part 5: determination of iron content; Part 6: determination of manganese content; Part 7: determination of aluminum content; Part 8: determination of silicon content; Part 9: determination of vanadium content)
8049 Ferronickel shot – Sampling for analysis
8050 Ferronickel ingots or pieces – Sampling for analysis
8343 Ferronickel – Determination of silicon content – Gravimetric method
9389 Nickel alloys – Determination of cobalt content – Potentiometric titration method with hexacyanoferrate(III)
9389 Nickel alloys – Determination of phosphorus content – Molybdenum blue molecular absorption spectrometric method
11 400 Nickel, ferronickel and nickel alloys – Determination of phosphorus content – Phosphovanadomolybdate molecular absorption spectrometric method
11 433 Nickel alloys – Determination of titanium content – Diantipyrylmethane molecular absorption spectrometric method
11 434 Nickel alloys – determination of niobium and tantalum – Ion exchange method
11 436 Nickel and nickel alloys – Determination of total boron content – curcumin molecular absorption method
11 437 Nickel alloys – Determination of trace-element content by electrothermal atomic absorption spectrometric method (Part 1: general requirements and sample dissolution; Part 2: determination of lead content)
11 438 Ferronickel – Determination of trace-element content by electrothermal atomic absorption spectrometric method (Part 1: general requirements and dissolution; Part 2: determination of lead content; Part 3: determination of antimony content; Part 4: determination of tin content; Part 5: determination of tellurium content; Part 6: determination of thallium content; Part 7: determination of silver content; Part 8: determination of indium content)
12 169 Nickel oxide – Determination of nickel content – Electrolytic deposition method

National Standards

ASTM B39 Standard Specification for Nickel
BS 375 Specification for Refined Nickel
DIN 1701 Deutsche Normen – Hüttennickel

REFERENCES

1. C.L. Lewis, W.L. Ott, N.M. Sine, *The Analysis of Nickel*, Pergamon Press, Oxford, 1966.
2. C.L. Lewis, W.L. Ott, *Analytical Chemistry of Nickel*, Pergamon Press, Oxford, 1970.
3. V.M. Peshkova, V.M. Savostina, *Analytical Chemistry of Nickel*, Ann Arbor-Humphrey, Ann Arbor, MI, 1969.
4. I.V. Pyatnitskii, *Analytical Chemistry of Cobalt*, Ann Arbor-Humphrey, Ann Arbor, MI, 1969.
5. B.W. Haynes, M.J. Magyar, 'Analysis and Metallurgy of Manganese Nodules and Crusts', *NATO ASI Ser., Ser. C Math. Phys. Sci.*, **194**, 235–246 (1985).
6. D.P. Xia, D.Z. Pan, *Metall. Anal. (China)*, **12**(2), 10–12 (1992).
7. CANMET Report 79–14, Canada Centre for Mineral and Energy Technology, Ottowa, Canada, 1979.
8. F. Smith, B. Cousins, J. Bozic, W. Flora, *Anal. Chim. Acta*, **177**, 243–245 (1985).
9. J. Bozic, D. Maskery, S. Maggs, H. Susil, F.E. Smith, *Analyst*, **114**, 1401–1403 (1989).
10. J. Bozic, R.L. Ethier, S.R. Maggs, D. Maskery, H.M. Susil, J. Le Coche, F.E. Smith, *Am. Lab.*, March, 32F–32I (1994).
11. Y. Xiang, L. Chen, J. Hua, W. Du, *Metall. Anal. (China)*, **9**(5), 30–32 (1989).
12. *Claisse Fluxy Instruction Manual, Part 2*, F. Claisse, Quebec City, Canada, 23 and 44, 1995.
13. XRF Method No. 9.6.6, Lakefield Research Limited, Lakefield, ON, 1997.
14. B.D. Wheeler, D.M. Bartell, J.A. Cooper, *Adv. X-ray Anal.*, **20**, 423–430 (1977).
15. V.L. Gravitis, J.S. Watts, *AMDEL Bull.*, **21**(October), 1–7 (1976).
16. J. Wu, W. Lai, Z. Li, X. Dong, H. Guo, B. Wang, *J. Chengdu Inst. Technol.*, **22**(2), 101–108 (1997).
17. B.D. Sowerby, W.K. Ellis, R. Greenwood-Smith, in *Nuclear Techniques and Mineral Resources 1997*, International Atomic Energy Agency, Vienna, 499–520, 1997.
18. M. Vobecky, Z. Janout, S. Pospisil, J. Dana, *J. Radioanal. Nucl. Chem.*, **83**(1), 123–128 (1984).
19. P.S. Wee, G.R. Agnes, S. McGeorge, Presented at the Pittsburgh Conference, 1989.
20. E.E. Bugbee, *A Textbook of Fire Assaying*, 3rd edition, John Wiley & Sons, 1940 (reprinted in 1991 by Legend Metallurgical Laboratory Inc.).

21. O.C. Dietrich, W.F. Shepard, *Fire Assaying*, Met-Chem Research Inc., 1989, from the 1st edition, 4th impression of 1940.
22. F.E. Beamish, *The Analytical Chemistry of the Noble Metals*, Pergamon Press, Oxford, 1966.
23. F.E. Beamish, J.C. Van Loon, *Recent Advances in the Analytical Chemistry of the Noble Metals*, Pergamon Press, Oxford, 1972.
24. N.K. Bel'skii, L.A. Nebol'sina, L.K. Shubochkin, *Zavod. Lab.*, **55**(9), 49–52 (1989).
25. Yu.S. Zolotov, A.N. Shkil', V.A. Vilenkin, E.P. Zdorova, E.G. Rukhadze, *J. Anal. Chem. USSR*, **43**(7), 1016–1022 (1989).
26. S.V. Mashkina, G.K. Budnikov, T.V. Beilina, N.A. Ulakhovich, *Zavod. Lab.*, **55**(10), 4–6 (1989).
27. R.V.D. Robért, E. Van Wyk, R. Palmer, *NIM Report*, **1371**, 1971, National Inst. Metallurgy, Johannesburg, South Africa.
28. K. Dixon, E.A. Jones, S. Rasmussen, R.V.D. Robért, *NIM Report*, **1714**, 1975, National Inst. Metallurgy, Johannesburg, South Africa.
29. S.E. Jackson, B.J. Fryer, W. Gosse, D.C. Healy, H.P. Longrerich, D.F. Strong, *Chem. Geol.*, **83**, 119–132 (1990).
30. R. Juvonen, E. Kallio, T. Lakomaa, *Analyst*, **119**, 617–621 (1994).
31. E.K. Shibuya, J.E.S. Sarkis, J. Enzweiler, A.P.S. Jorge, A.M.G. Figueiredo, *J. Anal. At. Spectrom.*, **13**(9), 941–944 (1998).
32. R.R. Barefoot, *J. Anal. At. Spectrom.*, **13**, 1077–1084 (1998).
33. G.M. Kolesov, D.Y. Sapozhnikov, *Zh. Anal. Khim.*, **51**(2), 234–241 (1996).
34. E.L. Hofman, A.J. Naldrett, J.C. Van Loon, A. Manson, *Anal. Chim. Acta*, **102**, 157–166 (1978).
35. I. Shazali, L. Van't Dack, R. Gijbels, *Anal. Chim. Acta*, **196**, 49–58 (1987).
36. G.H. Faye, P.E. Moloughney, *Talanta*, **19**, 269–284 (1972).
37. H.L. Zheng, X.H. Cui, *Metall. Anal. (China)*, **14**(2), 43–44 (1994).
38. L.L. Kolosova, L.A. Ushinskaya, T.N. Kopylova, *Zh. Anal. Khim.*, **48**(4), 644–647 (1993).
39. G. Chen, X. Yan, X. Nie, *Yejin Fengxi*, **8**(5), 4–7 (1988).
40. L.P. Tikhanova, I.P. Svarkovskaya, I.A. Fedotova, G.N. Solov'eva, *Zh. Anal. Khim.*, **46**(6), 1176–1180 (1991).
41. E.G. Khomutova, A.P. Rysev, *Zavod. Lab.*, **61**(11), 1–3 (1995).
42. N.A. Khvorostukhina, E.G. Khomutova, I.N. Samuelnkova, *Zavod. Lab.*, **52**(3), 4–6 (1986).
43. I.I. Alekseeva, N.A. Khvorostukhina, A.P. Rysev, E.G. Khomutova, *Zh. Anal. Khim.*, **35**(3), 505–510 (1980).
44. G. Chai, A.J. Naldrett, J.C. Rucklidge, L.R. Kilius, *Can. Mineral.*, **31**(1), 19–30 (1993).
45. R. Irwin, Lakefield Research Limited, personal communication, October 1998.
46. I.M. Kolthoff, E.B. Sandell, *Textbook of Quantitative Inorganic Analysis*, 3rd edition, Macmillan, New York, 333–335, 1952.
47. T.D. Aleksandruk, T.N. Khlopina, D.M. Svarts, *Zavod. Lab.*, **42**(2), 161–163 (1976).
48. W.R. Hatch, W.L. Ott, *Anal. Chem.*, **40**, 2085–2087 (1968).
49. Y.P. Mao, *Metall. Anal. (China)*, **13**(2), 19–22 (1993).
50. X.-C. Li, B.-L. Zeng, *J. Central-South Inst. Min. Metall. (China)*, **24**(1), 129–135 (1993).
51. A. Hitchen, G. Zechanowitsch, *Talanta*, **27**(3), 269–275 (1980).
52. B. Bibent, L. Bufferne, *Chim. Anal.*, **52**(6), 632–646 (1970).
53. J.M. Lopez, G.A. Alvarez, J.R.M. Gonzalez, L.M.F. Desdin, J.M. Frometa, *Rev. Mineria. Geol.*, **13**(2), 39–44 (1996).
54. A.Z. Gedeon, C.R.M. Butt, K.A. Gardener, M.K. Hart, *J. Geochem. Explor.*, **8**, 283–303 (1977).
55. T.N. Khlopina, T.D. Alexandruk, L.G. Kim, *Zavod. Lab.*, **50**(11), 29–31 (1984).
56. H. Schorin, *Adv. X-ray Anal.*, **25**, 127–131 (1982).
57. N. Damjanovic, Y. Bujuri, B. Pavlovic, in *National Conference on Atomic Spectroscopy with International Participation, Varna (Bulgaria)*, 340–342, 1986 (article available from NTIS).
58. A.C. Diaz, *Radiochem. Radioanal. Lett.*, **50**(3), 139–149 (1981).
59. Method No. 9.6.1, Lakefield Research Limited, Lakefield, ON, 1998.
60. J.R. Vizoso, Report IAEA-R-2070-F, International Atomic Energy Agency, Vienna, 1980.
61. D. Rubio, J. Ehsteves, H. Brukhertseifer, 'Highly Sensitive Nuclear-physical Methods for Nickel and Cobalt Content Analysis of Ferrous Laterites, with Previous Chemical Treatment of Samples', *Joint Inst. for Nuclear Research, Dubna (USSR), Dept. of Radiation Safety Rep.*, No. 18-82-245, 1982.
62. G. Blondiaux, M. Crousilles, J.L. Debrun, C. Dixsaut, *Bull. Bur. Rech. Geol. Minieres, Sect. 2*, **4**, 323–324 (1978).
63. J. Griffit, G.M. Kolesov, 'Complex Neutron Activation Analysis of Cuban Laterite Ores and the Products of Their Processing', in *All Union Conference on Activation Analysis and Other Radioanalytical Methods*, Tashkent (USSR), May 26, 1987, Vol. 2, Summaries of Reports (1987).
64. S. Ahmad, D.F.C. Morris, *Analyst*, **102**, 395–399 (1977).
65. S. Ahmad, D.F.C. Morris, *Radiochem. Radioanal. Lett.*, **28**(2), 165–170 (1977).
66. P.J. Potts, A.G. Tindle, P.C. Webb, *Geochemical Reference Materials Compositions*, CRC Press, Boca Raton, FL, 1992.
67. H. Flaschka, R. Speights, *Microchem. J.*, **14**(3), 490–499 (1969).

68. G. Bazco Banyasz, *Kohasz. Lapok*, **111**(8), 379–381 (1978).
69. A. Hofer, *Fresenius' Z. Anal. Chem.*, **240**(4), 229–232 (1968).
70. K.E. Burke, *Analyst*, **97**, 19–28 (1972).
71. K.T. Thornton, K.E. Burke, *Analyst*, **99**, 469–470 (1974).
72. K.E. Burke, *Appl. Spectrosc.*, **28**, 234–237 (1974).
73. K.E. Burke, *Talanta*, **21**, 417–423 (1974).
74. S.H. Blakeley, A. Manson, V.J. Zatka, *Anal. Chem.*, **45**, 1941–1943 (1973).
75. E. Jackwerth, H. Linke, *Erzmetall*, **31**, 275–277 (1978).
76. V. Zatka, *Anal. Chem.*, **50**, 538–541 (1978).
77. J.E. Forrester, V. Lehecka, J.R. Johnston, W.L. Ott, *At. Absorpt. Newsl.*, **18**(4), 73–76 (1979).
78. D.M. Shvarts, I.V. Konovalova, L.P. Abroskina, I.K. Fadeeva, E.L. Grinzaid, *Zavod. Lab.*, **57**(8), 30–33 (1991).
79. M. Kiyokawa, H. Yamaguchi, R. Hasegawa, *Bunseki Kagaku*, **43**(4), 289–293 (1994).
80. M. Franek, V. Krivan, *Fresenius' J. Anal. Chem.*, **338**(1), 28–33 (1990).
81. Y.Y. Yoon, K.Y. Lee, *J. Radioanal. Nucl. Chem.*, **210**(1), 56–63 (1996).
82. P. Taylor, B. Desmet, R. Dams, *Anal. Lett.*, **18**, 2477–2487 (1985).
83. P. Richner, *J. Anal. At. Spectrom.*, **8**(5), 927–931 (1993).
84. O. Kujirai, K. Yamada, *Fresenius' J. Anal. Chem.*, **348**(11), 719–723 (1994).
85. Z. Waszczylo, D. Maskery, J. Bozic, Presented at the 1985 Pittsburgh Conference, paper 1129.
86. J. Bozic, Inco Ltd., Copper Cliff Laboratory, personal communication, 1998.
87. J.K. Van Peteghem, *Can. J. Spectrosc.*, **30**(9), 109–114 (1985).
88. T.D. Aleksandruk, T.N. Khlopina, D.M. Shvarts, *Zavod. Lab.*, **44**(2), 167–169 (1978).
89. Method No. 9.6.8, Lakefield Research Limited, Lakefield, ON, 1997.
90. Method No. 9.11.2, Lakefield Research Limited, Lakefield, ON, 1997.
91. H. Gu, Z. Wang, *J. Mater. Eng. (China)*, **11**, 40–42 (1997).
92. H.H. Willard, N.H. Furman, C.E. Bricker, *Elements of Quantitative Analysis*, 4th edition, Van Nostrand, New York, 438–441, 1956.
93. H.A. Laitinen, *Chemical Analysis*, McGraw-Hill, New York, 414–417, 1960.
94. S. Itoh, K. Sato, J. Takahashi, H. Okochi, *J. Jpn. Inst. Met.*, **54**(5), 575–580 (1990).
95. M. Bettinelli, P. Taina, F. Belei, U. Baroni, G. Bizzarri, in *Progress of Analytical Chemistry in the Iron and Steel Industry (Luxembourg)*, The Commission of the European Community, 362–376, 1992.
96. B. Brachfeld, F. Cambria, M. Palumbo, *Appl. Spectrosc.*, **33**(1), 51–54 (1979).
97. J.M. Griffiths, H.R. Whitehead, *Rep. Natl. Phys. Lab. Chem.*, **49** (1976).
98. M. Roca Adell, J.P. Diaz-Guerra, 'Analysis of Structural Materials for Fast Reactors by X-ray Fluorescence Spectrometry', *Junta Energ. Nucl. [Rep.]*, JEN (Spain), **407**, 1978.
99. H. Okochi, K. Takahashi, S. Suzuki, K. Sato, E. Sudo, *J. Jpn. Inst. Met.*, **41**(5), 492–498 (1977).
100. H. Okochi, K. Takahashi, S. Suzuki, K. Sato, E. Sudo, *Trans. Natl. Res. Inst. Met. (Tokyo)*, **19**(6), 293–299 (1977).
101. J.Y. Marks, D.E. Fornwalt, R.E. Yungk, *Spectrochim. Acta, Part B*, **38**(1/2), 107–113 (1983).
102. H. Okochi, K. Takahashi, S. Suzuki, E. Sudo, *J. Jpn. Inst. Met.*, **44**(3), 324–330 (1980).
103. H. Okochi, O. Kujirai, F. Hirose, M. Kori, S. Suzuki, K. Takahashi, M. Kiyokawa, T. Yoshioka, K. Sato, T. Kobayashi, *Bull. Natl. Res. Inst. Met.*, **14**, 323–335 (1997).
104. C.W. Whitten, *At. Spectrosc.*, **8**(3), 81–83 (1987).
105. A. Gómez Coedo, M.T. Dorado López, A. Vindel Maeso, *J. Anal. At. Spectrom.*, **2**, 629–635 (1987).
106. M. Kohuri, O. Kujirai, K. Yamada, H. Okochi, *Anal. Sci.*, **4**(3), 293–297 (1988).
107. T.L. Thiem, J. Sneddon, *Spectrosc. Lett.*, **27**(10), 1437–1445 (1994).
108. B.D. Zehr, *At. Spectrosc.*, **15**(5), 213–215 (1994).
109. T.Y. Kometani, 'Impediments to Analysis', ASTM STP 708, ed. J.E. Foster, American Society for Testing and Materials, Philadelphia, PA, 32–44, 1980.
110. G.E. Peterson, *At. Absorpt. Newsl.*, **16**(6), 133–142 (1977).
111. J. Gierek, A. Wycislik, S. Gregorczyk, A. Kasperek, *Rudy Met. Niezelaz.*, **24**(1), 10–12 (1979).
112. J. Sire, J. Collin, J.A. Voinovitch, 'Study of Physical and Chemical Conditions for Determination of Vanadium in Steel by Atomic Absorption Spectrometry and Extension to Copper, Chromium, Nickel and Manganese', *Spectrochim. Acta B*, **33**, 31–45 (1978).
113. A.L. Kruglyak, O.L. Skorskaya, M.N. Gorlova, *Ind. Lab. (USSR)*, **57**(4), 395–397 (1991).
114. L.A. Romanshko, I.P. Kharmalov, N.N. Basargin, Yu.G. Rozovskii, Yu.F. Ziborova, N.P. Krivenkova, V.M. Zharova, *Ind. Lab. (USSR)*, **53**(12), 1128–1133 (1988).
115. L.A. Romanshko, I.P. Kharmalov, N.N. Basargin, *Zavod. Lab.*, **53**(12), 21–24 (1987).
116. B. Brachfeld, *Appl. Spectrosc.*, **28**(5), 449–454 (1974).
117. I.M. Kutyrev, G.M. Chernyshova, N.N. Basargin, N.I. Mikheev, *Zavod. Lab.*, **62**(11), 7–9 (1996).
118. T.R. Dulski, *Talanta*, **29**(6), 467–471 (1982).
119. M.R.P. Reddy, P.V.S. Kumar, J.P. Shyamasundar, Y. Anjaneyulu, *Proc. Indian Natl. Sci. Acad. A*, **56**(3), 255–280 (1990).
120. G. Preis, *DEW Techn. Ber.*, **13**(3), 217–222 (1973).
121. B.L. Mu, D.W. Zhang, *Metall. Anal. (China)*, **13**(6), 44–45 (1993).

122. V.T. Meryan, I.I. Vataman, V.N. Baskin, *Zavod. Lab.*, **53**(5), 11–12 (1987).
123. A.T. Pilipenko, E.A. Shpak, O.P. Dunaev, *Zh. Anal. Khim.*, **30**(9), 1701–1703 (1975).
124. M. Bettinelli, S. Spezia, U. Baroni, G. Bizzarri, *At. Spectrosc.*, **15**(3), 115–121 (1994).
125. Z. Liang, R.F. Lonardo, R.G. Michel, *Spectrochim. Acta, Part B*, **48**(1), 7–23 (1993).
126. B. Mile, C.C. Rowlands, A.V. Jones, *J. Anal. At. Spectrom.*, **7**(7), 1069–1073 (1992).
127. R.T. Simpson, *At. Spectrosc.*, **10**(3), 82–84 (1989).
128. C.P. Bosnak, G.R. Carrick, W. Slavin, *At. Spectrosc.*, **7**(5), 148–150 (1986).
129. K. Gralewska, T. Glenc, *Pr. Inst. Met. Zelaza*, **33**(3–4), 151–156 (1981).
130. A. Osojnik, M. Renko, T. Drglin, Z. Torkar, *J. Phys. IV, Colloq.*, **5**(C7), 189–197 (1995).
131. G.G. Welcher, O.H. Kriege, J.Y. Marks, *Anal. Chem.*, **46**(9), 1227–1230 (1974).
132. J.B. Headridge, R.A. Nicholson, *Analyst*, **107**, 1200–1211 (1982).
133. O. Kujirai, T. Kobayashi, E. Sudo, *Trans. Jpn. Inst. Met.*, **18**, 775–779 (1977).
134. M.S. Reichardt, *At. Spectrosc.*, **13**(5), 178–184 (1992).
135. W. Frech, *Talanta*, **21**, 565–571 (1974).
136. M. Kudo, T. Kobayashi, S. Hasegawa, H. Okochi, *Bunseki Kagaku*, **43**(3), 215–221 (1994).
137. J.Y. Yao, W.B. Xie, Q.L. Hu, C.Z. Lian, *Fengxi Huaxue*, **23**(3), 264–267 (1995).
138. J.Y. Yoa, Z.M. Yun, H.I. Duan, R.Y. Xiang, *Fengxi Huaxue*, **21**(7), 846–849 (1993).
139. T. Kobayashi, O. Kujirai, F. Hirose, H. Okochi, *J. Jpn. Inst. Met.*, **47**(8), 676–682 (1983).
140. T. Kobayashi, O. Kujirai, F. Hirose, H. Okochi, *J. Jpn. Inst. Met.*, **48**(5), 542–548 (1984).
141. T. Kobayashi, F. Hirose, S. Hasegawa, H. Okochi, *Nippon Kinzuko Gakkaishi*, **49**(8), 656–662 (1985).
142. S. Hasegawa, T. Kobayashi, R. Hasegawa, K.H. Cho, *Trans. Jpn. Inst. Met.*, **36**(9), 1157–1162 (1995).
143. T. Kobayashi, H. Okochi, *J. Jpn. Inst. Met.*, **53**(11), 1123–1128 (1989).
144. S.-I. Chang, H.-M. Liu, S.-J.J. Tsai, *J. Anal. At. Spectrom.*, **13**, 1123–1131 (1998).
145. J.Y. Marks, G.G. Welcher, R.J. Spellman, *Appl. Spectrosc.*, **31**(1), 9–11 (1977).
146. S. Bächman, R.W. Karlsson, *Analyst*, **104**, 1017–1029 (1979).
147. A.A. Baker, J.B. Headridge, R.A. Nicholson, *Anal. Chim. Acta*, **113**, 47–53 (1980).
148. J.B. Headridge, R.A. Nicholson, *Analyst*, **107**, 1200–1211 (1982).
149. R.L. Irwin, D.J. Butcher, J. Takahashi, G.-T. Wei, R.G. Michel, *J. Anal. At. Spectrom.*, **5**(7), 603–610 (1990).
150. M.W. Hinds, D.C. Gregoire, E.A. Ozaki, *J. Anal. At. Spectrom.*, **12**(2), 131–135 (1997).
151. H.D. Fleming, R.D. Ide, *Anal. Chim. Acta*, **83**, 67–82 (1976).
152. J. Yao, L.Q. Peng, *Fengxi Huaxue*, **22**(11), 1135–1137 (1994).
153. K. Tong, X. Guo, *Metall. Anal. (China)*, **16**(6), 4–6 (1996).
154. M. Bettellini, S. Spezia, G. Bizzarri, *At. Spectrosc.*, **13**(2), 75–80 (1992).
155. C.P. Hanna, J.F. Tyson, S.G. Offley, *Spectrochim. Acta, Part B*, **47**, 1065–1073 (1992).
156. F. Hirose, T. Kobayashi, S. Hasegawa, H. Okochi, *Trans. Natl. Res. Inst. Met. (Tokyo)*, **28**(3), 258–264 (1986).
157. S. Matysik, A. Wycislik, *Rudy Met. Niezelaz.*, **27**(4), 180–182 (1982).
158. N.V. Trofimov, B.I. Petrov, N.N. Nekhaev, A.I. Busev, *Zavod. Lab.*, **46**(1), 15–17 (1980).
159. F. Yu, *Metall. Anal. (China)*, **10**(4), 23–25 (1990).
160. F. Yu, *Metall. Anal. (China)*, **15**(5), 25–28 (1995).
161. K. Thornton, *Analyst*, **94**, 958–967 (1969).
162. B. Thelin, *Appl. Spectrosc.*, **35**(3), 302–307 (1981).
163. D.S. Lowe, *Analyst*, **110**, 583–585 (1985).
164. B. Larsson, B. Berglund, in *Progress of Analytical Chemistry in the Iron and Steel Industry (Luxembourg)*, the Commission of the European Community, 210–217, 1992.
165. R.K. Malhotra, K. Satyanarayana, G.V. Ramanaiah, G. Srinivasan, *At. Spectrosc.*, **16**(6), 235–242 (1995).
166. N.V. Stashkova, O.E. Safenova, A.S. Sirazetdinova, *Freib. Forschungsh. B*, **B266**, 82–86 (1988).
167. P. Pakalns, *Anal. Chim. Acta*, **51**(3), 497–501 (1970).
168. P.P. Kish, Ya. I. Studenyak, Ya. R. Bazel', *Ind. Lab. (USSR)*, **58**(2), 1120–1123 (1992).
169. Ya. R. Bazel', Ya. I. Studenyak, L.N. Kushnir, O.G. Voronich, *Zavod. Lab.*, **60**(10), 12–14 (1994).
170. V.V. Serbinovich, V.P. Antonovich, *Zavod. Lab.*, **58**(7), 5–6 (1992).
171. D. Rosales, I. Milan, J.L.G. Ariza, *Talanta*, **33**(7), 607–610 (1986).
172. A.V. Ishchenko, E.M. Shvarts, N.V. Stashkova, A.Ya. Dzene, A.A. Bernane, *Zavod. Lab.*, **51**(4), 8–10 (1985).
173. V.I. Simonenko, V.V. Soloshonok, T.L. Vonogradova, V.V. Sukhan, *Ukr. Khim. Zh.*, **58**(3), 175–177 (1992).
174. F. de Pablos, G. Galan, J.G. Ariza, *Talanta*, **34**(10), 835–838 (1987).
175. T. Zhu, J.L.H. Sai, L. Fu, *Metall. Anal. (China)*, **10**(6), 7–9 (1990).
176. K.Z. Brainina, I.N. Nikulina, V.I. Kurbatova, *Ind. Lab. (USSR)*, **39**(7), 1028–1033 (1973).
177. Z. Yuexia, *Kang T'ieh*, **18**(2), 51–54 (1983).
178. M. Saito, S. Itoh, E. Sudo, *Trans. Jpn. Inst. Met.*, **27**(4), 279–287 (1986).
179. M. Saito, K. Yamada, H. Okochi, F. Hirose, *Trans. Natl Res. Inst. Met. (Japan)*, **27**(3), 169–176 (1985).
180. G. Vidal, P. Galmard, P. Lanusse, *Mem. Sci. Rev. Met.*, **67**(2), 137–147 (1970).

181. S. Tamura, *Mass Spectrosc. (Japan)*, **21**(4), 283–292 (1973).
182. Method 9.3.6, Lakefield Research Limited, Lakefield, ON, 1999.
183. F.I. Danilova, I.A. Fedotova, N.V. Ustinova, *Ind. Lab. (USSR)*, **51**(8), 682–683 (1986).
184. I.R. MacArthur, D.F.C. Morris, *Metallurgia*, **80**, 37–42 (1969).
185. K. Thornton, *Chemical Characterization*, Institute of Metals, London, 61–83, 1988.
186. N.A. Zakorina, I.S. Lindstrem, A.A. Petrov, *Ind. Lab. (USSR)*, **50**(2), 144–147 (1984).
187. A.P. Brovko, Yu.P. Shapovalov, V.A. Danilkin, O.K. Nikolaenko, *Ind. Lab. (USSR)*, **44**(10), 1379–1381 (1978).
188. N.A. Zakorina, I.S. Lindstrem, *Zavod. Lab.*, **45**(9), 816–818 (1979).
189. K.V. Kondakova, G.G. Glavin, Yu.A. Karpov, G.G. Sikharulidze, *Ind. Lab. (USSR)*, **49**(12), 1260–1262 (1983).

Noble Metals, Analytical Chemistry of

Maria Balcerzak
Warsaw University of Technology, Warsaw, Poland

1 Introduction **2**
1.1 Discovery and Natural Occurrence 2
1.2 Uses 2
1.3 Physical Properties 4
1.4 Chemical Properties 4
2 Sample Preparation **5**
2.1 Sampling 5
2.2 Decomposition/Dissolution Techniques 5
3 Separation and Preconcentration Methods **6**
3.1 Volatilization 6
3.2 Precipitation and Coprecipitation 7
3.3 Solvent Extraction 7
3.4 Sorption 8
3.5 Chromatographic Techniques 8
4 Determination Methods **9**
4.1 Classical Methods 9
4.2 Spectrometric Methods 11
4.3 Catalytic Kinetic Methods 17
4.4 Electrochemical Methods 17
4.5 Neutron Activation Analysis 18
5 Quality Control and Assurance **18**
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **20**

Analytical methods for the determination of noble (precious) metals: ruthenium (Ru), rhodium (Rh), palladium (Pd), osmium (Os), iridium (Ir), platinum (Pt) and gold (Au) are presented in this article. Discovery, natural occurrence and main applications of the metals are described. Physical and chemical properties of noble metals are summarized.

The nobility and catalytic activity of precious metals are the main properties that allow their use in a wide variety of applications, e.g. as catalysts in various chemical processes, as autocatalysts, in the electrical and electronic industry and in jewellery. Recent applications of some platinum compounds (cisplatin and its derivatives) as anticancer drugs are important.

The large variety of complex matrices, wide analytical concentration range (from sub-ppb to >99.99%), low reactivity towards single chemical reagents, great chemical similarities (especially between the pairs Ru and Os, Rh and Ir, Pt and Pd), complexity of platinum group metals (PGMs) species in solutions and rates of reaction make the accurate determination of noble metals a difficult analytical problem. The use of direct instrumental methods is restricted owing to interferences caused by matrix elements and low analyte concentrations. Sampling, sample decomposition, separation and preconcentration are critical steps in the majority of analytical procedures used.

The choice of the digestion procedure used depends on the nature of the sample matrix and the analyte concentration. Fire assay (lead, iron, copper, nickel, tin or nickel sulfide as collectors), oxidizing fusion, acids treatment and chlorination are used to digest various materials. Precipitation, solvent extraction and chromatographic methods (ion-exchange and chelating resins, capillary electrophoresis) are applied to separate noble metals from associated base metals and to separate the individual precious metals. Preliminary isolation of ruthenium and osmium from the other noble and base metals, as well as from each other, by distillation or extraction in the form of RuO_4 and OsO_4, is often applied.

Spectrophotometric methods using the complexes with inorganic and organic reagents can be applied to the determination of precious metals at ppm levels. Atomic absorption spectroscopy (AAS) (flame and graphite furnace) is well suited to the determination of Au, Pd, Rh and Pt (ppm and ppb levels, respectively). Ultratraces (ppb and sub-ppb levels) of noble metals can be determined in a large number of complex matrices by inductively coupled plasma mass spectrometry (ICPMS) with or without separation and preconcentration steps. A wide range of PGM concentrations, from percentage to ppm levels, can be determined by X-ray fluorescence (XRF) directly in solid samples or after pretreatment procedures (fire assay, coprecipitation, chromatographic preconcentration). Nuclear techniques (mainly neutron activation) are favored for the determination of low (ppb and sub-ppb) levels of precious metals (high sensitivities for Au, Ir, Pd and Os) in geological samples. Voltammetric measurements in combination with sample preparation, including the transformation of noble metals into electrochemically active compounds, provide high sensitivity (ppm and ppb levels) for the determination. The use of standard reference materials is essential to check the accuracy of the results obtained by various analytical techniques.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The noble metals are the elements from the eighth group of the periodic table: ruthenium (Ru), rhodium (Rh), palladium (Pd), osmium (Os), iridium (Ir) and platinum (Pt) (so-called PGM) and gold (Au). These metals are also called ''precious metals''. The name reflects the economic value of the metals as well as their rare occurrence.

1.1 Discovery and Natural Occurrence

Gold is the only noble metal that has been known since prehistoric times. Ancient civilizations already recognized gold as a precious metal. Numerous legends, mythological stories, folk tales and even the bible show a special interest in gold. There were numerous attempts, in particular in the Middle Ages, to convert other elements into gold.

According to records platinum was known in the sixteenth century. It was used by the Indians of Ecuador to make small articles of jewellery. The first description of ''platina'' as a new infusible metal found in the mines of the Chocó district of New Granada, Colombia, was made by de Ulloa in 1748. The first scientific studies on platinum were carried out in Europe after Charles Wood brought some samples of native platinum to England in 1741. They were described in 1750 by William Watson. The name ''platina del Pinto'' (little silver of the Pinto River), derived from the first source of the metal, was replaced by ''platinum'' after establishing its elemental nature in the eighteenth century. Detailed histories of platinum were published in 1960[1] and 1982.[2]

Palladium and rhodium were discovered by W.H. Wollaston in London in 1803 and 1804, respectively. They were isolated from the solutions remaining after dissolving crude platinum in aqua regia followed by precipitation of platinum as chloroplatinate. The names ''palladium'' and ''rhodium'' were given after the planetoid Pallas (discovered in 1802) and from the Greek *rhodes* (a rose) due to the red color of the rhodium compounds.

Two further PGMs were isolated by Smithson Tennant in 1804 from the insoluble residue remaining after the aqua regia treatment of crude platinum. They were called iridium after the Greek *iris* (a rainbow) as a consequence of the variety of colors of iridium salts, and osmium after the Greek *osme* (a smell) owing to the characteristic odor of the volatile osmium tetroxide.

The discovery of ruthenium by K.K. Klaus, professor of chemistry at the University of Kazan, Russia, in 1844 completed the list of PGM. Ruthenium was isolated from the residue left after the treatment of platinum ore from the Central Ural Mountains with aqua regia. Klaus ignited the residue with potassium nitrate and separated osmium by distillation in the form of tetroxide. The name of the element has its roots in Ruthenia, the Latin name for Russia.

The abundances of noble metals in the earth's crust are at the level of 1 ppb for each of ruthenium, rhodium, osmium and iridium; 10 ppb for palladium; 5 ppb for platinum[3] and 3.5 ppb for gold.[4] Precious metals have also been found in meteorites.

The major sources of noble metals are in South Africa, Canada and Russia. They are all primary deposits, usually associated with ultrabasic rock formations and often with copper and nickel sulfide deposits.[5,6] They occur as native grains or in minerals such as sperrylite ($PtAs_2$), braggite (Pt, PdNi)S, cooperite (PtPd)S, arsenopalladinite (Pd_3As), stibiopalladinite (Pd_3Sb), laurite $(RuOs)S_2$, osmiridium (a syserskite with less than 60% of iridium and ca. 35% of osmium) and iridiosmium (a nevyanskite with over 60% of iridium and ca. 20% of osmium). Ruthenium as a major constituent (7–15% in iridiosmium and 9–14% in osmiridium) and small amounts of the other noble metals are associated with osmium and iridium deposits. Native gold is present in sedimentary and igneous rocks in many parts of the world. It is also a minor component of base-metal ores, especially copper. Electrum is a specific gold and silver mineral (AuAg) containing 35.4% of silver. Gold and silver occur in nature as calvarite ($AuTe_2$), sylvanite $(AgAu)Te_2$, petzite $(AgAu)_2Te$, krennerite $(AgAu)_xTe_y$ and nagyagite (containing gold, lead, antimony, sulfur and tellurium).

Since 1953 South Africa has dominated amongst the world's producers of noble metals. The Merensky Reef of the Bushveld Igneous Complex in Central Transvaal in South Africa is the richest source of precious metals (Figure 1). The composition of the sources in the Merensky Reef is roughly 60% Pt, 25% Pd, 4% Rh and 11% Ru, Ir and Os. Native platinum, occurring in the basic igneous rocks in the Merensky Reef contains 70–90% of platinum and small quantities of the other noble metals and iron. Other large deposits of noble metals were found in Sudbury, Great Lakes Region, Ontario in Canada and in Noril'sk Complex of Siberia, Russia. The concentration of noble metals in Sudbury deposits is less than 1 ppm. South African noble metal ores consist of smaller amounts of copper and nickel than any other deposits. All other sources are primarily nickel and copper ores. Noble metals remain there only as by-products of nickel and copper mining. Minor resources of the PGM occur in other areas of the world. Colombia, USA (including Alaska), China and Western Australia are minor producers of noble metals.

1.2 Uses

Nobility and catalytic ability are the main properties that allow use of precious metals in a wide range of applications.[6,7] Platinum, palladium and rhodium are extensively used as components of automotive catalysts

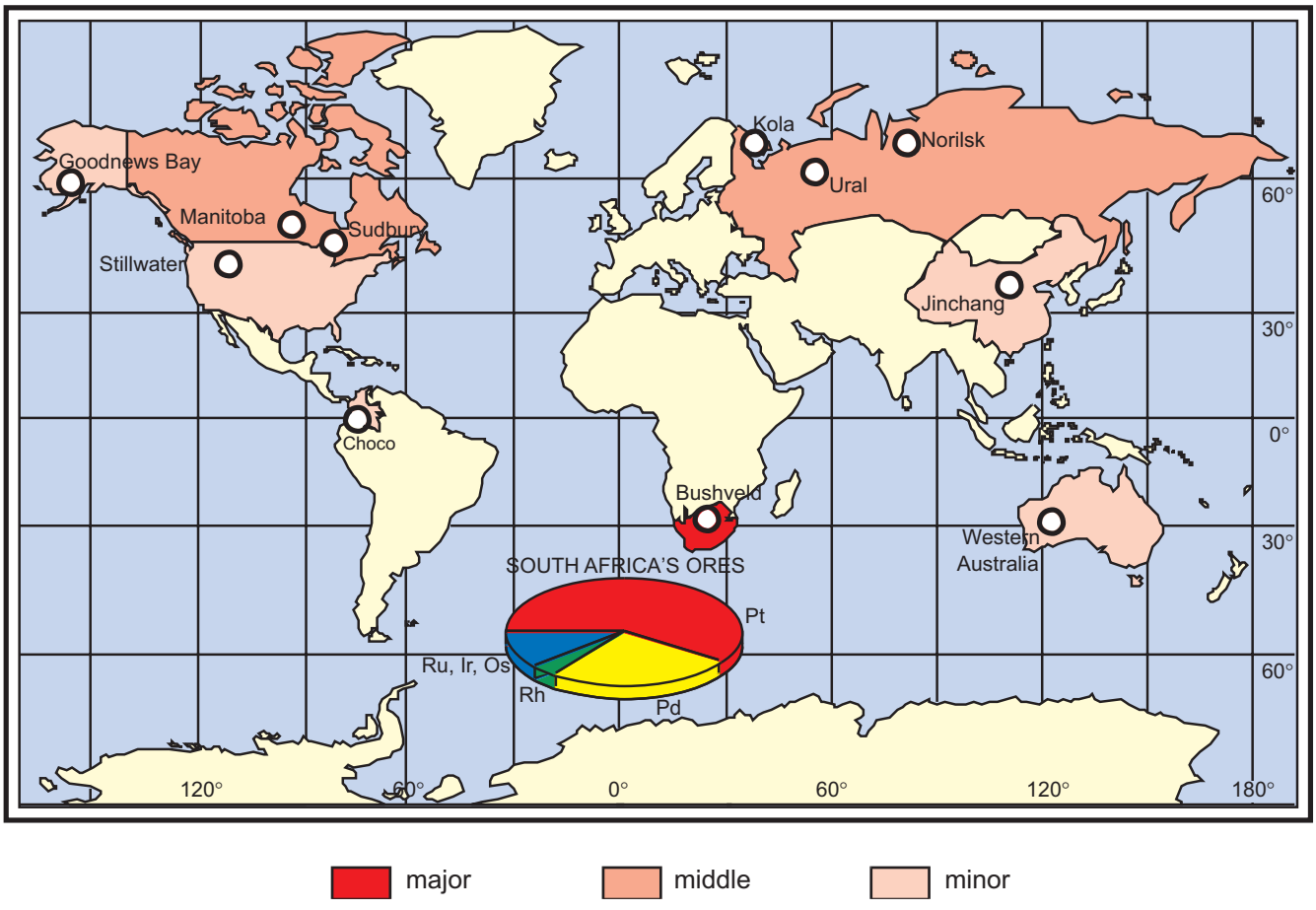

Figure 1 World sources of noble metals.

to reduce the emission of carbon monoxide, unburnt hydrocarbons and nitrogen oxides in exhaust gases. Platinum metals are also well established as catalysts in structural rearrangements and dehydrogenation of aliphatic hydrocarbons obtained by distillation of crude petroleum. Catalysts, in oxidation of ammonia to nitric oxides, in selective hydrogenation and dehydrogenation reactions for chemical intermediates in plastics, artificial fibers, rubber, pesticides, dyestuffs and pharmaceutical products provide other examples of the wide application of platinum metals in the chemical industry. There is a large demand for noble metals, in particular gold, palladium and platinum, from the jewellery industry.

Platinum, because of its low and temperature-dependent resistance, also finds numerous applications in the electrical and electronic industries, e.g. in thermocouples for temperature measurements, corrosion-resistant electrical contacts, electrodes, sensors for gas detection and optical data storage disks. It also has substantial use in the glass-making industry mainly for production of special glass and glass fiber. The importance of some platinum-containing compounds (cisplatin (*cis*-diamminedichloroplatinum(II)); carboplatin (*cis*-diammine(1,1-cyclobutanedicarboxylato)platinum(II)); iproplatin (*cis*-dichloro-*trans*-dihydroxo-*cis*-bis(isopropylamine)platinum(IV)) and their derivatives) as anticancer agents is rapidly increasing. Cisplatin was introduced into chemotherapy in 1978.

The main demand for palladium comes from the electrical and electronic industries. Multilayer ceramic capacitors for integrated circuits used, for example, in microcomputers, televisions, video recorders and compact disk players are made with the use of palladium. Palladium is also applied in the production of electrical contacts and conducting materials. Palladium alloys are extensively used in the chemical industry, mainly as chemical process catalysts. Jewellery and dentistry where gold is gradually replaced by less expensive palladium employ significant amounts of the world palladium supply as well.

A considerable increase in the use of rhodium in the world's industry has recently been observed. The main demand for rhodium (ca. 73% of the total supply)[6] comes from the production of autocatalysts. Rhodium as a component of platinum alloys is also used in production of

Table 1 Basic physical properties of noble metals

Property	Ru	Rh	Pd	Os	Ir	Pt	Au
Atomic number	44	45	46	76	77	78	79
Relative atomic mass	101.07	102.905	106.4	190.2	192.22	195.09	196.97
Electronic configuration	$4d^7 5s^1$	$4d^8 5s^1$	$4d^{10}$	$5d^6 6s^2$	$5d^7 6s^2$	$5d^9 6s^1$	$5d^{10} 6s^1$
Melting point (°C)	2310	1960	1554	3050	2443	1769	1063
Density (25 °C) ($g\,cm^{-3}$)	12.45	12.41	12.02	22.61	22.65	21.45	19.3
Electronegativity	1.42	1.45	1.35	1.52	1.55	1.44	1.42
Hardness (Mho's scale)	6.5	–	4.8	7.0	6.5	4.3	–
Crystal lattice	hexagonal	cubic	cubic	hexagonal	cubic	cubic	cubic

thermocouples, glass and glass fiber. The use of rhodium in the chemical industry in the production of nitric acid, hydroformylation of alkenes and carbonylation of methanol to yield acetic acid is growing.

The principal uses of iridium are related to its quite remarkable chemical and physical properties. Iridium is the most resistant of all metals to corrosion. It has a very high melting point (2443 °C) and it is the only metal that maintains good chemical properties in air at temperatures above 1600 °C. Thermocouples of pure Ir against 40% Rh–Ir are used to measure temperatures up to 2100 °C. Iridium-containing coatings of anodes in the electrochemical industry (e.g. in the production of chlorates) also have a wide range of applications. The production of crucibles suitable for growing oxide single crystals, such as gadolinium gallium garnet (GGG) and yttrium aluminum garnet (YAG), used in computer memory devices and in solid-state lasers represents the largest modern area of application of iridium. Iridium is suitable for making the crucibles in which single crystals of oxides can grow at temperatures above 2000 °C. Smaller, but essential demands for iridium come from other areas, e.g. medicine (for radioactive ^{192}Ir in cancer therapy), dentistry and jewellery (for iridium-containing alloys).

The major uses of ruthenium are in the electronics and electrochemical industries. Chip resistors, resistor networks containing ruthenium thick film pastes and electrodes coated with ruthenium dioxide provide the main examples of industrial applications of the metal. Ruthenium is used to harden platinum and palladium alloys. Ruthenium compounds are widely applied as chemical catalysts. Recent investigations of ruthenium complexes as an alternative to platinum cancer inhibitors are of special interest. *Cis*-dichlorotetrakis(dimethyl sulfoxide)ruthenium(II) (*cis*-$Ru(DMSO)_4Cl_2$), *cis*-tetraamminedichlororuthenium(III) chloride (*cis*-$[Ru(NH_3)_4Cl_2]Cl$) and *fac*-trisamminetrichlororuthenium(III) (*fac*-$Ru(NH_3)_3Cl_3$) have been recognized as highly active against some tumors with side effects lower than those from cisplatin.

Osmium alloys (with ca. 60% Os) are used as materials of extreme hardness, e.g. for instrument pivots. Osmium tetroxide is used as a strong oxidizing agent in organic oxidation processes and as a staining reagent for the microscopic examination of tissues.

Gold was primarily used in the jewellery industry and the investment world. For centuries dentistry has made use of gold. Gold compounds are used in medicine owing to their antirheumatoid activity.[8] There has been a growing demand for gold from the electronics and metal industries both searching for chemically highly resistant alloys.

1.3 Physical Properties

Noble metals are silvery white (Pt, Pd, Rh, Ir), gray-white (Ru), bluish white (Os) and yellow (Au) metals. Some general physical properties of precious metals are listed in Table 1. High melting point and hardness (in particular of Os, Ru and Ir) distinguish PGM. All metals are ductile and malleable. They can be obtained in compact, sponge, powdered and blacks forms. The ability to absorb high volumes of gases, in particular hydrogen by palladium (900 volumes of hydrogen in 1 volume of palladium), and oxygen by platinum (100 volumes of oxygen in 1 volume of platinum) is a characteristic property of these metals.

1.4 Chemical Properties

The high ionization potential of precious metals results in their noble character. The metals exhibit high resistance to single chemical reagents, particularly to single mineral acids, hydroxides and active metalloids. Molten alkali and alkaline hypochlorite solutions attack all noble metals.

Oxygen attacks noble metals at high temperature. Osmium shows the strongest affinity for oxygen. Finely divided osmium is oxidizable at room temperature yielding the volatile tetroxide, OsO_4, with its characteristic odor. Iridium exhibits the highest oxidation resistance. Ruthenium dioxide is formed at about 600 °C and converted into volatile RuO_3 and RuO_4 with further temperature increase. PtO_2 and PtO_3 formed during heating of platinum to a bright red heat in air can lead to its loss. Palladium dioxide formed at about 700 °C dissociates back to the free metal and oxygen at higher temperature.

Chlorine and fluorine attack all noble metals (osmium and ruthenium at about 100 °C and 300 °C, respectively; iridium above red heat; platinum and palladium slowly at room temperature) giving mixtures of simple chlorides and fluorides.

Noble metals exhibit numerous oxidation states in their compounds.[5] Ruthenium and osmium can occur at the highest oxidation states (+VIII), corresponding to the number of periodic table group. The (+VI) is the highest oxidation state for Rh, Ir and Pt, (+IV) for Pd and (+III) for Au. Ruthenium, osmium and iridium can occur at the lowest ((−II) and (−I)) oxidation states. The most stable oxidation states are (+IV) and (+II) for Pt and Pd, (+III) for Rh, (+IV) and (+III) for Ir, Ru and Os.

Apart from simple binary compounds, such as oxides, halides and sulfides, noble metals form a variety of complexes with inorganic and organic ligands. π-Bonding ligands, such as CO, NO, N_2, CN^-, SCN^-, PR_3, AsR_3 stabilize low metal oxidation states (−II, −I, 0, +I). Numerous complexes with halides, nitrogen, sulfur, oxygen and selenium as well as hydrides are known. Chloride complexes are of great importance in extraction and refining processes and also in the analytical chemistry of the noble metals. Strong complexes with oxygen are characteristic for ruthenium and osmium. The majority of the complexes of precious metals are kinetically inert.

All noble metal species occurring in solutions are extremely coordinated and complex. The metals can occur in numerous oxidation states and in many forms with different activity in a given oxidation state. Great chemical similarities (especially between the pairs: Ru and Os, Rh and Ir, Pt and Pd), the formation of compounds of similar composition and properties and the tendency for hydrolysis as well as formation of polynuclear complexes should be taken into consideration, in particular while aqueous solutions are examined. Strong reaction conditions, such as heating, high reagent concentration and long reaction time may be required for any ligand substitution, owing to the great inertness of many complexes. The complexity of the chemical properties of noble metals as well as the kinetics of the reactions in solutions cause serious problems in many analytical procedures. The availability of a standard solution containing a particular metal in a definite oxidation state and chemical form suitable for the analytical method being used is an important problem.

2 SAMPLE PREPARATION

2.1 Sampling

Sampling requires special attention in the analysis of noble metals. Standard methods of sampling can be used in the analysis of homogeneous materials, e.g. alloys, liquids, powders and so on. Noble metals, especially gold, often occur inhomogeneously in many matrices, e.g. in geological, environmental and biological materials. Powdering and the use of large sample size may sometimes be sufficient to ensure the representativeness of the sample in the case of minerals, ores and concentrates. Some preliminary operations, such as drying, crushing and grinding may be necessary. The formation of alloys, e.g. some electronic scrap with copper, lead and nickel, to produce homogeneous materials which can be sampled by standard methods is sometimes used. The history of the submitted sample must be taken into consideration prior to any reduction in weight or volume.

Special precautions are needed in sample collection and storage in the case of a readily reduced analyte (e.g. gold to elemental form) or an analyte easily oxidized to volatile compounds, (e.g. osmium to OsO_4). Careful treatment and storage of the sample may be necessary to avoid any changes in species of particular metals to be determined, e.g. in clinical samples.[8] The determination of ultratraces of noble metals requires special attention, to avoid contamination during the sample preparation step. Sampling and decomposition steps may become the main sources of contamination.

2.2 Decomposition/Dissolution Techniques

The choice of procedure used to digest the sample of noble metals depends on the nature of the sample matrix and the analytes' concentration.

Fire assay is the most popular procedure for isolation and concentration of noble metals from rocks, soils, sediments, minerals, ores and concentrates. Classical lead assay is the basis of the majority of digestion procedures used.[9] Fire assay involves fusion of a sample at high temperature with a flux (Na_2CO_3, $Na_2B_4O_7$), reducing agent (flour, starch) and collector (PbO). On fusion the analytes are extracted from a complex matrix into the lead button that is formed. The matrix elements react with the flux component to form a slag which is subsequently discarded. The lead button is placed on a cupel prepared from bone ash or magnetite. On heating to a temperature of about 800–850 °C in an oxidizing atmosphere, lead oxide and nonnoble metal oxides are produced and absorbed into the cupel where a bead of precious metals remains. The precious metals alloy is usually treated with a mixture of acids to extract the analytes.

The extraction of noble metals from a large sample size of a complex matrix into a relatively small bead of a simple metals alloy is the main advantage of the fire assay procedure. However, success in the quantitative recovery of precious metals requires an experienced and skilled

assayer to optimize both flux composition and fusion conditions. Another drawback of the classical fire assay using a lead collector is that it does not provide efficient recovery for all noble metals. Quantitative collection can be achieved for gold, silver, platinum and palladium. The recovery of ruthenium, iridium and osmium is seriously affected by flux composition and experimental conditions. Losses of osmium due to the high volatility of OsO_4 can occur. Rhodium and iridium can be determined using the classical fire assay followed by cupellation using a gold bead. Silver is added to the charge when analysis of a sample containing gold, platinum and palladium approaching the detection limit (DL) concentration is carried out. Losses of iridium and ruthenium can occur when cupellation involving a silver bead is used.

The other fire assay collectors, e.g. nickel sulfide, iron–nickel–copper alloys and tin, were found to be capable of quantitative recovery of all precious metals. Nickel sulfide fire assay has recently gained popularity.[10–12] Contamination due to large amounts of salts employed is a drawback of these sample pretreatment methods.

Wet acids treatment, alkaline oxidizing fusion and chlorination are the alternative digestion procedures used. The effectiveness of acids extraction depends on the chemical solubility of individual noble metals and sample composition.[13] Palladium and rhodium are the only metals attacked by hot nitric acid and boiling sulfuric acid, respectively. Aqua regia is used for the dissolution of gold and platinum. Ruthenium, osmium and iridium resist the attack of a mixture of acids, including aqua regia. Mixtures of acids (HCl, HNO_3 and $HClO_4$) in wet-, dry- and high-pressure-ashing, were recommended for digestion of clinical, biological and environmental materials. Hydrofluoric acid is sometimes added to the extractant to attack silicate phases and facilitate the liberation of analytes. However, uncertainties in the quantitative recovery of noble metals limit the practical application of direct wet digestion procedures, in particular to geological and environmental materials.

Alkaline, oxidizing fusion (NaOH, Na_2O_2) of the residue remaining after preliminary acid attack can be used to convert more resistant components into dissolved forms. Fusion with oxidizing flux can be effectively used to digest geological samples. The application of the method is restricted to small samples.

High recoveries (>90%) of all noble metals from the samples containing native metals, natural alloys and minerals can be achieved after quantitative conversion of analytes into chlorides that are easily dissolved in hydrochloric acid.[14] Three types of chlorination procedure are used, ''direct'' chlorination in the presence of large amounts of an alkali chloride, ''wet'' chlorination at elevated temperatures and pressures in a sealed tube containing HCl and oxidizing agent, and ''dry'' chlorination by hot chlorine passing over the sample usually mixed with a small amount of NaCl in an open tube at 500–600 °C. Dry chlorination has some advantages such as low concentration of salts and low blank levels. Moreover, large samples can be submitted for analysis and the level of contamination can be reduced. The use of bubbler apparatus avoids losses of volatile chlorides formed.

Platinum- and palladium-based alloys, containing up to 5% of the other noble metals undergo aqua regia treatment. Owing to the high volatility of osmium and ruthenium the decompositions must be performed in closed systems. The dissolution of alloys with a higher ($\geq$10%) content of more resistant metals (Ru, Os, Ir) may be achieved at higher temperature and in high-pressure systems. Silver, if present in the examined sample, forms an insoluble residue.

3 SEPARATION AND PRECONCENTRATION METHODS

The great chemical similarities of the PGM, the low concentrations to be determined and the large interfering effects from matrix elements generally require separation and preconcentration steps prior to the determination. Chemical methods of separation using volatilization, precipitation and coprecipitation, solvent extraction, sorption and chromatographic techniques are applied to separate particular noble metals from their mixtures as well as from the base metals.

3.1 Volatilization

In analysis of multicomponent noble metal samples the preliminary selective separation of ruthenium and osmium in the form of the volatile tetroxides, RuO_4 and OsO_4, is generally performed. Ruthenium and osmium are oxidized to Ru(VIII) and Os(VIII) and quantitatively transformed into tetroxides by treatment in acid solution with strong oxidizing agents, e.g. $HClO_4 + H_2SO_4$, $KMnO_4$, K_2CrO_4, NaClO, $NaBrO_3$. Both tetroxides can be isolated from the mixture of noble metals by distillation. Selective oxidation of osmium to OsO_4, which occurs in the presence of hydrogen peroxide or 5 M HNO_3, allows osmium to be isolated from ruthenium and from the other noble metals.

The form in which ruthenium and osmium are present in the solution examined affects the effectiveness of their conversion into volatile compounds. Chloride complexes require strong oxidizing conditions, e.g. the action of $HClO_4 + H_2SO_4$, to achieve quantitative oxidation to RuO_4 and OsO_4. Evaporation of the examined solution with sulfuric acid to remove hydrochloric acid prior

to distillation is often applied. Nitroso complexes of ruthenium exhibit high resistance to oxidants. They must be destroyed prior to any oxidation step.

Hydrochloric acid is most frequently used as the absorbing solution for volatile tetroxides. Ruthenium and osmium are reduced to various oxidation states (VI, IV, III and II) depending on the conditions used (hydrochloric acid concentration, temperature and reaction time).[15,16] In 6 M HCl the $RuCl_6^{2-}$ complex is the main reduction product of Ru(VIII). Reduction of Os(VIII) in hydrochloric acid proceeds slowly. Quantitative conversion of osmium into the stable $OsCl_6^{2-}$ complex requires the solutions to be heated for a minimum of 20 min at 90–100 °C. Water, organic solvents (e.g. CCl_4, $CHCl_3$) or sodium/potassium hydroxide can also be used as absorbing media for both ruthenium and osmium tetroxides.

3.2 Precipitation and Coprecipitation

Chemical methods of separation based on the precipitation of sparingly soluble forms are used to isolate a single element or a group of elements from a mixture of metals.[15,17,18] Silver is usually isolated from a mixture of noble metals by precipitation of silver chloride. Platinum, as a major component, can be separated by precipitation as the ammonium chloride $(NH_4)_2PtCl_6$. Platinum can also be separated from the other PGM by a hydrolytic method. At pH 6–8 all platinum metals form sparingly soluble hydrolysis products, whereas platinum remains in solution. The precipitation of palladium with dimethylglyoxime is widely used to separate palladium from the other metals.

Reduction to the elementary form under the action of numerous reductants, e.g. zinc, iron, copper, tin(II) chloride, hydrazine, formic acid and sodium tetrahydroborate, is the basis of separation of precious metals from a large number of base metals. The easier reduction of gold is used for its separation from platinum group elements. The use of oxalic acid, hydroquinone and nitrite as reductants leads to selective precipitation of gold from solutions containing the other noble metals. Platinum and palladium, in contrast to iridium and rhodium, are reduced to the elementary form and precipitated under the action of Hg_2Cl_2 or CuCl and during boiling with tellurium powder. Rhodium may be separated from iridium by reduction to the metal with Sb powder in hot H_2SO_4. Any precipitation of rhodium and iridium requires their preliminary separation from the other noble metals. Quantitative separation of particular noble metals may require a reprecipitation step owing to a tendency to occlude the other components present in the solution.

Reductive coprecipitation with a suitable collector is applied to separate noble metals from matrix elements and concentrate them to the level detected by instrumental techniques. Tellurium, selenium, arsenic, mercury and copper are most frequently used as collectors. Coprecipitation is effectively applied to quantitative separation of ultratrace amounts of noble metals from major and minor components of geological, biological and environmental samples.[19–21] Coprecipitation with a tellurium collector following a sodium peroxide fusion was used to improve the concentration of Pt, Pd, Ru and Ir from geological materials.[22]

3.3 Solvent Extraction

Solvent extraction is successfully used for separation of particular noble metals from their multicomponent samples, as well as from base metals.[23] Halide (chloride, bromide and iodide) complexes form the basis of most of separation methods. Chelate complexes with organic reagents, e.g. α-dioximes, dithizone, carbamates, 8-hydroxyquinoline, formazans, thiourea derivatives, and ion pairs with high-molecular-weight amines are also used in many separation procedures. Neutral organic compounds containing sulfur (sulfides and sulfoxides) and phosphorus (e.g. triphenylphosphine (TPP), trioctylphosphine oxide (TOPO) and tributyl phosphate (TBP)) effectively extract noble metals from various media. Solvating systems with oxygen-containing solvents are widely used.[24] Extraction of Au(III) from 4–8 M HCl (HBr) as $AuCl_4^-$ ($AuBr_4^-$) into, for example, diisopropyl ether (DIPE), methyl isobutyl ketone (MIBK), ethyl or amyl acetate is often applied to isolate gold from various matrices.

Differences in the labile character of the platinum metals towards complexing and chelating agents, differences in the charge of particular metal species and inertness to hydrolysis may lead to different affinities to extractant and make up the basis of separation methods. Palladium chloride complexes are kinetically labile and readily react with a number of complexing agents at room temperature. The great inertness of platinum towards complexation often requires the solutions to be heated or labilizing agent to be added. Tin(II) chloride and iodides are often used to reduce Pt(IV) to Pt(II) which more readily reacts with many complexing agents. Extraction of palladium prior to formation of an extractable platinum complex (e.g. with dithizone) can be used for quantitative separation of both metals. Chloride complexes of Ru(IV), Ir(IV) and Os(IV) provide an example of highly extractable species through either the anion-exchange or hydration-solvation mechanism, whereas the extraction of Ru(III), Ir(III) and Os(III) is poor. The labile character of rhodium, in comparison with iridium, allows the separation of both metals in some extraction systems. Rhodium forms extractable complexes with, for example, 2-thiobenzothiazole, 2-thiobenzimidazole and diphenylthiourea while iridium

remains in the solution. The high tendency of rhodium for hydrolysis may cause difficulties in obtaining a readily extractable complex.

Polyurethane foams were found to be effective extractants of noble metals from their thiocyanate solutions allowing selective separation of some binary mixtures, e.g. Rh and Ir,[25] Ru and Rh,[26] Os and Ru.[27]

Separation of ruthenium and osmium by extraction of RuO_4 and OsO_4 (CCl_4, $CHCl_3$) is an alternative to the distillation procedure of separation for both metals. Osmium tetroxide can be directly extracted from the samples treated with aqua regia.

3.4 Sorption

Sorption methods are particularly convenient for concentration of small amounts of noble metals from solutions containing high concentrations of nonnoble metals. In some systems quantitative separation of precious metals from 10^4-fold quantities of accompaning metals (especially Cu, Fe, Ni, Co, Al, Mg and Ca) and concentration factors of ca. 10^3 can be achieved.[28,29] Inorganic and organic sorbents, as well as chelating resins have found application in concentration of noble metals. The sorption procedures can be performed in a batch or a column mode. Desorption of analytes by subsequent elution with suitable agents usually precede their determination by conventional analytical techniques. The metals can also be determined directly in the sorbent phase by physical methods applicable to solid samples (e.g. XRF). Some sorbents are destroyed by ashing at high temperature or heating with a mixture of perchloric, sulfuric and nitric acid before the determination of analytes by a selective analytical technique.

Sulfides of copper, mercury, lead, cobalt and nickel, oxides of aluminum, zirconium, tin and titanium, phosphates of zirconium and cerium, and insoluble ferrocyanides were examined as inorganic sorbents for noble metals. The low solubility of platinum metal sulfides in strongly acid media favored the sorption on sulfide collectors. Activated charcoal is also used to concentrate noble metals from parts per billion levels, in particular in analysis of water samples.

Resins impregnated with chelating ligands or chemically modified by the introduction of a functional group are widely used as chelating sorbents. High selectivity and efficiency are characteristic properties of this group of sorbents. Chelating resins containing a functional group with sulfur (e.g. dithiocarbamate, thiosemicarbazide, 2-thiobenzothiazole, benzylthiourea derivatives) and nitrogen (e.g. aliphatic amines and polyamines, heterocyclic nitrogenous ligands) have found special reference in separation and concentration of precious metals.

The POLYORGS™ sorbents containing nitrogen–heterocycle ligands (heterocyclic amine and amidoxime groups) are highly selective for noble metals in the presence of base metals.[29] They are especially efficient in direct group separation and concentration of noble metals from solutions of complex composition obtained after digestion of rocks, ores, minerals, alloys and industrial process solutions. Complete separation of noble metals in the presence of $50\,g\,L^{-1}$ Cu(II), $100\,g\,L^{-1}$ Ni(II) and Co(II) and $10–20\,g\,L^{-1}$ Fe(III) was reported. Other base metals (e.g. Al, Ca, Mg, etc.) are not sorbed.

Preconcentration of ultratrace amounts of noble metals by sorption of some chelate complexes with, for example, 8-hydroxyquinoline[30] and 1-(2-pyridylazo)-2-naphthol (PAN)[31] on microcrystalline naphthalene has been reported.

3.5 Chromatographic Techniques

Ion-exchange chromatography is widely used for group separation of noble metals from base metals and for separation of precious metals from one another.[23,32] Both cation and anion exchangers can be used for separation and preconcentration of noble metals. Strong chloride complexes formed in dilute hydrochloric acid constitute the basis of the majority of separation procedures. Anionic chloride complexes pass through cation-exchange columns while base metals which exist as cations under the conditions used are retained on the resins. Precious metals are retained by anion exchangers which do not absorb the base metals. A variety of different organic and inorganic agents (e.g. thiourea, mineral acids of different concentrations and at various temperatures) can be used for selective elution of particular metals. It should be noted, however, that quantitative conversion of a particular metal into a definite complex is required prior to the chromatographic separation step. Different forms of analyte present in the solution may behave quite differently during the chromatographic separation step providing difficulties in complete separation of the metals. The labile character of complexes towards aquation and the ability to change the oxidation state of the analyte are the drawbacks of ion-exchanging procedures. The retention of anionic complexes is additionally affected by their charge. Doubly charged complexes (e.g. $PtCl_6^{2-}$, $RuCl_6^{2-}$, $IrCl_6^{2-}$) are strongly retained by the resins, whereas triply charged complexes (e.g. $PtCl_6^{3-}$, $RuCl_6^{3-}$, $IrCl_6^{3-}$) are only weakly bound. Strong affinity of precious metals to the resins may cause difficulties in quantitative elution of analytes. In some systems, the use of differences in the strength of electrostatic interactions of particular complexes with the resins or labile character in the formation of cationic and anionic species allow separation of multicomponent mixtures of the metals.[23] Dowex

1-X8 provides an example of a strong anion exchanger exhibiting a high selectivity for noble metals.[33]

The application of high-performance liquid chromatography (HPLC) using chelate complexes to separate and preconcentrate noble metals has gained popularity. 8-Hydroxyquinoline, PAN, 1-thiazolylazo-2-naphthol (TAN), 4-(2-thiazolylazo)resorcinol (TAR), diethyldithiocarbamate (DDTC), 2-(6-methyl-2-benzothiazolylazo)-5-diethylaminophenol and thiourea derivatives are most often used as precolumn derivatization agents. The retention behavior and separation of analytes may be considerably affected by the composition and properties of the mobile phase (pH, the kind and the content of the organic solvents used). The combination of a preliminary group extraction with a subsequent chromatographic separation forms the basis of effective and rapid separation procedures.

N,*N*-diethyl-*N*′-benzylthiourea complexes of platinum, palladium and rhodium have been used effectively for separation of binary mixtures, Pd–Pt and Rh–Pt, from platinum-based catalysts by thin-layer chromatography.[34] Size-exclusion chromatography has been applied to investigate platinum metabolites in plants[35] and gold drug metabolites in human blood.[36] Platinum and palladium were both effectively separated from the samples of catalytic converters by capillary electrophoresis after conversion into chloride complexes under treatment with a mixture of $HCl + HNO_3$.[37]

4 DETERMINATION METHODS

Numerous chemical and instrumental techniques have been applied to the determination of noble metals in a wide range of sample types. The choice of a suitable detection technique is to a large degree limited by the concentration level to be determined. Great chemical similarities between precious metals provide potential sources of mutual interference in their determination by the majority of the methods used.

Classical (gravimetric and volumetric) methods can be applied to analysis of samples containing a sufficient quantity (from milligram levels) of analytes. Such methods are favored when establishing the accurate concentration of the standard solution is required.

Spectrophotometric methods using complexes with inorganic or organic agents can be applied to the determination of precious metals at the ppm level. Atomic absorption spectroscopy (flame and graphite-furnace) is particularly suitable for determination of gold, palladium, rhodium and platinum (parts per million and parts per billion levels respectively). Plasma sources (ICP and microwave-induced plasma (MIP)) with atomic emission spectroscopy (AES) or MS detection are recommended for analysis of multicomponent samples and where large concentration ranges are determined. ICPMS (together with isotope dilution (ID)) is particularly suited to the determination of ultratrace amounts (ppb and sub-ppb levels) of noble metals in a large number of complex matrices. It can be applied with or without separation and preconcentration steps. A wide range of precious metal concentrations, from percentage to ppm levels, can be determined by XRF directly in solid samples or after pretreatment procedures (fire assay, coprecipitation, chromatographic preconcentration). Nuclear techniques (mainly neutron activation) are favored in the determination of low (ppb and sub-ppb) levels of noble metals (high sensitivities for Au, Ir, Pd and Os) in geological samples. Voltammetric measurements in combination with sample preparation, including the transformation of noble metals into electrochemically active compounds, provide high sensitivity (ppm and ppb) of the determination.

4.1 Classical Methods

4.1.1 Gravimetric Methods

Chemical methods of analysis are most widely accepted for the determination of palladium. Precipitation of palladium with dimethylglyoxime has found the widest application among gravimetric methods used. It allows separation of palladium (in the range from a few milligram to 1000 mg) from the other noble metals, from medium quantities of base metals and from small amounts of gold. The selectivity of the reaction and the definite composition of the isolated precipitate are very important factors. A number of other oximes (e.g. α-furildioxime, α-benzoinoxime, oxalenediamidoxime) have been examined as precipitation agents for palladium.

Gravimetric methods are relatively rarely used to determine other metals. Reduction in hydrogen to the elemental form is recommended for quantitative isolation of ruthenium, rhodium and iridium. Precipitation of sparingly soluble compounds with chosen, mostly organic, agents is also applied. However, the lack of specific reagents as well as difficulties in establishing the final composition of the isolated precipitate and coprecipitation are the main obstacles to developing methods suitable for quantification of particular metals.

Hydrolysis at pH 6–8 results in the conversion of platinum metals (except platinum) into sparingly soluble hydrated oxides. The hydrated oxides can be used as weighed forms for gravimetric determination after filtration and heating at ca. 450–600 °C. They may also be subsequently transformed into a metallic form by reduction in hydrogen. Hydrolytic precipitation allows platinum to be separated from the other precious metals. Platinum, in the form of Pt(IV), remains in solution

under the conditions used. The precipitation of platinum in the form of ammonium chloroplatinate followed by ignition to platinum metal is a very old method used for separation of the metal. In the case of a larger proportion of associated precious metals, in particular iridium, rhodium and palladium, reprecipitation may be required. Chlorination to the anhydrous trichlorides and subsequent ignition (in air and hydrogen) into a metallic form can be applied to produce weighed forms for rhodium and iridium. Oxalic acid (pH 1–2), hydroquinone (0.1–1.2 M HCl) and nitrite are used for selective reduction of gold to a weighed metallic form in solutions containing platinum metals.

Noble metals belong to the acid sulfide group. Various inorganic (hydrogen or sodium) and organic (e.g. phenylthiosemicarbazide, thiobenzothiazole, thiophenol) sulfides have been proposed for precipitation of metals. Weighed forms obtained by coprecipitation of univalent thallium with the sulfides of the platinum metals, $TlRu_2S_6$, $TlRh_2S_4$, $TlPtS_3$ and $TlPd_2S_3$, can be used for gravimetric determination. The method is limited to micro amounts of each metal.

A number of organic reagents (e.g. thionalide, thiourea, antipyrine derivatives, 2-thiobenzothiazol, benzidine) have been used to precipitate precious metals from various matrices. The complexes of ruthenium, osmium, palladium, rhodium and iridium with thionalide can be weighed. Partial reduction of Pt(IV) to Pt(II) and simultaneous oxidation of the reagent to dithionalide absorbed on the separated compound produce a mixed precipitate which limits the use of the reagent for gravimetric determination of platinum. Antipyrine derivatives, e.g. diantipyrilmethane, diantipyrilpropylmethane, form precipitates with osmium, platinum and palladium. Platinum, osmium and iridium can be precipitated with tetraphenylphosphonium and arsonium salts. Each platinum metal (except osmium) can be precipitated by benzidine.

Quantitative separation of noble metals in the form of sparingly soluble precipitates can be seriously affected by the form of the analytes existing in the solutions examined. Nitroso complexes present in the samples examined must be destroyed prior to any precipitation. Precipitation of ruthenium and osmium from the solutions of their tetroxides easily isolated from mixtures of precious metals by distillation is preferred.

A detailed discussion of gravimetric methods applied to the determination of noble metals has been presented.[38,39]

4.1.2 Titrimetric Methods

Redox reactions, the formation of stable complexes and precipitation of sparingly soluble compounds, form the basis of titrimetric methods used for the determination of noble metals.[15,17] The number of volumetric methods is relatively few compared with the gravimetric methods used. In general, the transformation of an analyte into a definite complex is required prior to the titration step. High inertness of noble metal complexes limits their use in titrimetric procedures.

The redox reactions Ru(IV $\rightleftarrows$ III, IV $\rightleftarrows$ VIII), Os(VIII $\rightleftarrows$ IV, VI $\rightleftarrows$ IV) and Ir(IV $\rightleftarrows$ III) are the only ones used in titrimetric determination of ruthenium, osmium and iridium. Titanium(III), copper(I) chloride, hydroquinone, ascorbic acid and iodides, are used as titrimetric agents in reduction methods. Reduction of OsO_4 by iodides followed by a titration of iodine with thiosulfate ions provides an effective and rapid method for the determination of osmium at milligram levels. Interferences from the other platinum metals should be taken into consideration in the presence of strong reduction agents. Redox reactions are seldom used to determine palladium and rhodium, mainly owing to the high values of redox potential of the systems Pd(IV)/Pd(II), Rh(IV)/Rh(II) and low values of potentials of the Pd(II)/Pd(0) and Rh(III)/Rh(0) systems. Potassium permanganate and Ce(IV) sulfate are often applied to oxidize analytes to higher oxidation steps. The reaction Pt(II) $\rightleftarrows$ Pt(IV) taking place under the action of a strong oxidizing agent is used for quantification of platinum.

Iodides, dithizone, thiourea, thionalid, DDTC, α-nitroso-β-naphthol and thiobenzothiazol provide examples of inorganic and organic agents used in titrimetric methods resulting in the precipitation of sparingly soluble compounds. A titrimetric method using iodides as precipitation agent is often applied to the determination of palladium (PdI_2) in alloys, catalysts and the other industrial products containing Pd, Pt, Au and Ag. The use of organic reagents generally allows lower amounts of analytes to be determined. Extractive titration with dithizone is an example of a method widely used for the determination of gold (at the 10^{-5}% level) in ores, rocks, soils, waters and organic materials. Such a method was also applied satisfactorily to the determination of platinum in platinum–silver alloys.

Titration by using an excess of ethylenediaminetetraacetic acid (EDTA) followed by back-titration with mercury(II), zinc or lead can be applied to the determination of milligram amounts of palladium, platinum and rhodium in mixtures with some base metals, e.g. Fe(III), Th, In and Ta. The catalytic effects of noble metals on some redox reactions are the basis of titrimetric procedures. Spectrophotometry and potentiometry are most frequently used as detection techniques for titrimetric methods.

4.2 Spectrometric Methods

4.2.1 Ultraviolet/Visible Spectrophotometry

Spectrophotometric methods were historically the first spectroscopic methods applied to the determination of small amounts of noble metals. The considerable progress in use of spectrophotometry as an analytical technique, the availability of the instrumentation, the wide range of concentrations that can be determined and satisfactory precision and accuracy make the technique very popular for the determination of many elements, including precious metals. Although the number of other spectroscopic methods developed for determining precious metals has significantly increased, spectrophotometry is still a very useful technique, in particular for the determination of ruthenium, osmium and iridium. The number of papers devoted to the determination of noble metals by spectrophotometry is still high. A marked trend towards the application of the methods developed to the analysis of real samples can be noted. Recent progress in simultaneous analysis of binary mixtures of metals by applying derivative spectrophotometric measurements should be emphasized.

The determination of precious metals, particularly in multicomponent mixtures, is very difficult because of the mutual interference caused by their great chemical similarity and their ability to form complexes of a similar composition and properties. The low selectivity of the color reagents generally requires the separation of the precious metals from each other as well as from the interfering matrix. Solvent extraction and chromatographic techniques are most often applied to separate and preconcentrate the analytes prior to spectrophotometric detection. The integration of the detection technique with the chemical pretreatment step applied is essential. Numerous inorganic and organic reagents have been proposed for the determination of noble metals in various materials.[40–42] In general, the methods involving organic chromogenic reagents provide better selectivity as compared with those employing inorganic ligands.

Colored complexes of precious metals with halide ions (chloride, bromide and iodide), thiocyanates and $SnCl_3^-$ ligands play an important role in quantitative methods of analysis. Complexes with $SnCl_3^-$ groups ($SnCl_2$ in HCl) are widely used for spectrophotometric determination of all platinum metals.[43] The high lability of various forms which can occur in aqueous solutions and the tendency for the composition to change with even small changes in the reaction conditions present difficulties in obtaining reproducible results. The ligand and acid concentrations, temperature and reaction time affect the final composition of the complex obtained. The most labile complexes containing various amounts of Cl^- and $SnCl_3^-$ ligands, gradually formed under addition of tin(II) chloride to the solution being examined, are characteristic of palladium. An osmium–tin(II) chloride complex (0.3 M $SnCl_2$, 2 M HCl) is the most stable among the complexes of platinum metals. The values of molar absorptivities (ε) are in the range from 2.4×10^3 L mol^{-1} cm^{-1} for osmium ($\lambda_{max} = 385$ nm) to 1.3×10^4 L mol^{-1} cm^{-1} for platinum ($\lambda_{max} = 403$ nm). An iridium–tin(II) bromide complex provides a higher sensitivity (ε at $\lambda_{max} = 402$ nm amounts to 4.96×10^4 L mol^{-1} cm^{-1}) compared with the complex with $SnCl_3^-$ ligands.

Ruthenate (RuO_4^{2-}) and osmate $OsO_4(OH)_2^{2-}$ anions, obtained after alkaline oxidation fusion of powdered metals, can be used for the spectrophotometric determination of larger amounts of ruthenium and osmium. Molar absorptivities are equal to 1.74×10^3 L mol^{-1} cm^{-1} (465 nm) and 2.75×10^3 L mol^{-1} cm^{-1} (340 nm) for ruthenium and osmium, respectively.[44] The selective determination of osmium in the presence of ruthenium can be carried out in iodide solutions.[45]

Attempts to enhance the selectivity of the spectrophotometric methods by applying derivative spectrophotometry result in methods being developed that allow the determination of particular metals in mixtures. The resolution of overlapping absorption spectra of particular components by applying analogue or numerical differentiation is a characteristic property of the derivative spectrophotometric technique.[46] The individual signal of particular analytes can be isolated from the examined mixtures by obtaining derivative spectra of the suitable order. The selective determination of metals using derivative spectrophotometry has been carried out for mixtures of ruthenium and osmium in the form of chloride[47] and tin(II) chloride[48] complexes; palladium, platinum and gold in chloride solutions;[49] gold, palladium and platinum as bromide complexes,[50] and palladium and platinum in iodide solutions.[51] Direct determination of ruthenium and osmium in aqueous solutions of their tetroxides has been presented.[52] Derivative spectrophotometry was successfully used to determine platinum and iridium in reforming catalysts[53] and platinum and ruthenium in Pt–Ru carbon supported catalysts for fuel cells[54] after conversion of analytes into chloride and tin(II) chloride complexes, respectively.

Numerous organic reagents have been applied to the spectrophotometric determination of noble metals. Many of them contain sulfur as donor atom, e.g. dithizone, thiourea and its derivatives, *p*-dimethylaminobezylidenerhodanine (rhodanine) and its derivatives, thio-Michler's ketone, derivatives of dithio-oxamide (rubeanic acid), thiosalicylamide, dithiobenzoic acid, 1,4-diphenylthiosemicarbazide and 2-thiobenzothiazole. Complexes

of osmium(VIII) and palladium(II) with allylthiourea,[55] platinum and palladium with dithizone[56] and ruthenium(III) and rhodium(III) with octadecyldithiocarbamate[57] were used for simultaneous determination of particular metals in mixtures by derivative spectrophotometry.

A group of methods for determining noble metals are based on azo-compounds, e.g. PAN, 4-(2-pyridylazo)resorcinol (PAR), TAR, chloro- and bromo-derivatives of pyridylazocompounds, arsenazo III, sulfonitrophenol M, 2-(2-thiazolylazo)-5-diethyl-*m*-aminophenol (TAAP) and sulfochlorophenolazorhodanine.

Nitroso compounds, such as 1-nitroso-2-naphthol and 2-nitroso-1-naphthol, *p*-nitrosodimethylaniline, *p*-nitrosodiethylaniline and nitroso-R salt have also been proposed as chromogenic agents for the determination of noble metals.

Triphenylmethane chelating reagents, e.g. xylenol orange, chrome azurol S, eriochrome cyanine R, pyrocatechol violet, bromopyrogallol red and pyrogallol red form the basis of sensitive spectrophotometric methods. A significant increase in sensitivity (ε ca. 1×10^5) can be reached in ternary systems containing cationic surfactants, mainly cetylpyridinium (CP) and cetyltrimethylammonium (CTA) ions.

Ion associates with basic dyes undergoing extraction or flotation preconcentration prior to spectrophotometric detection make the basis of very sensitive methods. Xanthene (rhodamine B and rhodamine 6G), triarylmethane (malachite green, brilliant green, crystal violet and victoria blue B) and azine (methylene blue and capri blue) dyes have been investigated as counter ions for the formation of sparingly soluble ion associates with anionic (chloride, bromide, iodide, thiocyanate and $SnCl_3^-$) complexes of noble metals. The ion associates, existing in aqueous solutions as colloids or suspensions, undergo extraction or flotation when shaken with a nonpolar organic solvent. The colored extract or solution obtained after dissolving a floated precipitate in an organic polar solvent (e.g. methanol, ethanol, acetone, dimethylformamide) form the basis of spectrophotometric measurement.[58] In particular, the methods applying the flotation preconcentration steps provide great sensitivity for the determination (ε in the range 2×10^5–1.14×10^6) owing to the high molar absorptivities of the dyes and the fact that more than one cation of the dye is involved in the formation of the sparingly soluble precipitate.[59] Flotation–spectrophotometric methods, however, are nonselective and preliminary separation of analytes by another separation technique is generally required. Solvent flotation methods applying basic dyes for preconcentration of ruthenium and osmium for spectrophotometry have been reviewed.[60]

4.2.2 Atomic Absorption Spectroscopy

AAS is widely used to determine noble metals in a large number of different materials: ores, rocks, minerals, concentrates, alloys, industrial products, biological and clinical samples, environmental samples, waste solutions and solids (including nuclear wastes). The technique can also be applied to the determination of impurities in high-purity noble metal alloys. It is particularly suitable for analysis of gold, palladium, rhodium and platinum. Relatively poor DLs are obtained for iridium, ruthenium and osmium.[61] Both flame (air–acetylene and nitrous oxide) and furnace atomizers were employed.

Sensitivities offered by flame atomic absorption spectroscopy (FAAS) at the most sensitive lines: 0.5–1 $\mu g\,mL^{-1}$ Ru (349.9 nm, air–C_2H_2); 0.1–0.2 $\mu g\,mL^{-1}$ Rh (343.5 nm, air–C_2H_2); 0.1 $\mu g\,mL^{-1}$ Pd (244.8 and 247.6 nm, air–C_2H_2); 1 $\mu g\,mL^{-1}$ Os (290.9 or 305.9 nm, N_2O–C_2H_2); 6–10 $\mu g\,mL^{-1}$ Ir (264.0 and 266.5 nm, C_2H_2); 0.5 $\mu g\,mL^{-1}$ Pt (265.9 nm, N_2O–C_2H_2) and 2 $\mu g\,mL^{-1}$ Au (242.8 nm, N_2O–C_2H_2) have been reported.[62]

The sensitivity of AAS measurement is seriously hampered by mutual noble metal interferences and interfering effects from a number of common elements. The components of the aqueous solution examined (aqua regia, HCl) may also strongly affect the AAS signal. The use of nitrous oxide–acetylene flame allows some chemical interferences to be overcome. Buffers and modifiers (e.g. V, U, Cu, La) have been proposed to improve the atomization efficiency. The examination of a solution of the sample in a suitable organic solvent often results in enhancement of the sensitivity. MIBK has found the widest application among the organic solvents tested for FAAS. Strong oxidants present in the samples examined lead to enhancement of the signal for ruthenium owing to improved transport and atomization efficiency of RuO_4 formed.

The use of the preconcentration steps provides the enhancement of the DLs offered by FAAS. A DL of 0.3 $ng\,mL^{-1}$ Pd was reported for the system comprising a microcolumn packed with activated carbon fiber to preconcentrate the analyte.[63] FAAS determination of platinum, palladium and rhodium in catalysts characterized by DLs of 5, 3 and 40 $\mu g\,L^{-1}$, respectively has been described.[64] A DL of 1 $\mu g\,mL^{-1}$ Pt in deactivated catalysts was reached.[65] Concentrations of 0.1–50 $\mu g\,g^{-1}$ Au were successfully determined in copper concentrates by combining reductive coprecipitation of gold using part of the matrix copper as collector.[66] Improvement in the DL in the determination of gold (35 $ng\,mL^{-1}$) was achieved by combination of tubular flow Donnan dialysis with FAAS.[67]

Graphite furnace atomizers offer sensitivities suitable for the determination of noble metals at submicrogram levels. Absolute DLs of 30 pg Ru, 10 pg Rh, 10 pg Pd,

ca. 1 ng Os, 0.2 ng Ir, 30 pg Pt and 10 pg Au were reported.[62] The mechanism of atomization of noble metals under an argon atmosphere using a graphite furnace atomizer was investigated.[68] The ''appearance temperature'' (the lowest temperature at which atoms of particular elements are released from the surface of the atomizer into a vapor phase) as a function of the nature of the atomizer surface, the chemical form of the analyte present on the atomizer surface and the reaction leading to conversion of analyte into atoms were studied. The appearance temperature was in close agreement while using HCl or HNO_3 media. The use of aqueous solutions of palladium, silver and gold or films of pure metals led to similar results. The appearance temperatures for aqueous solutions used were found to be lower than the melting point of some metals.

Because of serious interference from a number of elements, preliminary separation and preconcentration steps are generally applied prior to graphite furnace atomic absorption spectroscopy (GFAAS) detection. Solvent extraction,[69–72] ion-exchange chromatography,[73,74] sorbent exchangers[75–79] and coprecipitation[20,21,80] are widely used to separate noble metals from the interference elements and to preconcentrate them prior to GFAAS detection. Matrix extraction was applied to the determination of iridium, rhodium and ruthenium in high purity platinum.[81] The AAS signals of Ir, Rh and Ru are strongly suppressed in the presence of large amounts of platinum. Preliminary, double extraction of platinum into isoamyl alcohol–MIBK provides results with satisfactory precision and accuracy for a minimum of 2 $\mu g\,g^{-1}$ Ir and Ru and 1 $\mu g\,g^{-1}$ Rh.

The interferences caused by the presence of the other platinum metals, gold, silver, nickel and iron in the AAS measurement of palladium, platinum and rhodium signals using a nickel sulfide fusion preconcentration step were examined.[82] Strong interference of nickel on the determination of the three analytes was established. For a concentration of 250 mg L^{-1} Ni, as representative of the average content after the fusion step, in a solution of 10% HCl containing 10 $\mu g\,L^{-1}$ Pd, 50 $\mu g\,L^{-1}$ Pt and 10 $\mu g\,L^{-1}$ Rh, the signals of palladium, platinum and rhodium were increased by 102%, 50% and 43%, respectively. The palladium signal was stable when the content of Ni was >250 mg L^{-1}. The platinum signal seemed to be stable at 25–100 mg L^{-1} Ni and increased at 100–250 mg L^{-1} Ni. The rhodium signal was least affected for all the nickel concentrations used. The presence of 10 mg L^{-1} Fe caused 10% decrease in the signal for palladium and about 10% increase in the signals obtained for platinum and rhodium.

The mutual interference of PGM and gold when present in the proportions typically encountered in copper–nickel ores were studied after collective extraction of metals with a mixture of alkylaniline hypochlorite and petroleum sulfides and AAS examination of the extracts.[71] It was established that palladium had no effect on the signals of platinum, ruthenium and rhodium up to a 500-fold excess. The 300-fold excess of palladium decreases the signal of iridium by not more than 10%. Platinum does not interfere with the determination of palladium and rhodium up to a 300-fold excess. A 200-fold excess of platinum only insignificantly decreases the signals of ruthenium and iridium. The presence of a 10–15-fold excess of rhodium does not interfere with the determination of the other PGM. The most serious interferences were observed from ruthenium on the absorption of iridium and from ruthenium and iridium on the signal of osmium. These interferences were apparent at a 15- to 20-fold excess of the metals. DLs of ($\mu g\,g^{-1}$): 0.001 Au, 0.005 Pd and Rh, 0.01 Pt, 0.02 Ru and Ir and 0.06 Os (5 g sample and an extractant volume of 5 mL) were reported.

The effects of copper and iron on the determination of platinum, palladium and rhodium by AAS were studied using the Polyorgs™-XXVII complexing sorbent to preconcentrate analytes from HCl solutions.[77] Tolerable copper and iron concentrations amounted to 10 g L^{-1} and 2 g L^{-1}, respectively.

Matrix interferences (NH_4Cl, Al, Fe(III) and Cu(II)) on GFAAS determination of palladium, platinum and rhodium in some polluted biological materials were examined applying on-line sorbent extraction preconcentration with the use of bis(carboxymethyl)dithiocarbamate (CMDTC) chelates.[75] DLs (ng g^{-1}) of 0.03 for palladium, 0.1 for platinum and 0.01 for rhodium were reached.

Chelates with ammonium pyrrolidinedithiocarbamate together with liquid–liquid extraction (LLE) using 4-methyl-2-pentanone were applied for separation and enrichment of palladium, platinum and gold from the human urine of unexposed people prior to GFAAS with Zeeman-effect background correction.[72] DLs of 20 ng g^{-1} for palladium and gold and 70 ng g^{-1} for platinum were reported.

GFAAS is sufficiently sensitive for the determination of ultratrace amounts of noble metals in a large number of geological samples (rocks, minerals and ores), in particular when a suitable separation–preconcentration step is employed. Ion-exchange chromatography[73,74] and coprecipitation with Se, Te and Hg as collectors[20,21,73,74] are widely used to separate noble metals from matrix elements. Mercury is an ideal matrix for GFAAS because it is completely volatilized at low temperatures and causes no interferences.

An increase in the analytical signals of noble metals in the presence of organic substances (e.g. suspensions of chelating agents) was reported.[83] The reduced sample preparation time and the decreased possibility of analyte losses through volatilization using solid sampling GFAAS

are promising. A range of solid sampling techniques such as spark ablation, laser ablation and glow discharge has been examined to interface with AAS for analysis of high-purity precious metals.[84]

4.2.3 Atomic Emission Spectrometry

Simultaneous multielement capability and greater linear dynamic range are the main advantages of AES over the atomic absorption technique. However, the DLs offered by AES are poorer by up to two orders of magnitude compared with GFAAS. The DLs offered by inductively coupled plasma atomic emission spectrometry (ICPAES) at the most sensitive lines of particular noble metals are as follows: 30 ng mL^{-1} Ru (240.272 nm), 6 μg mL^{-1} Rh (343.5 nm), 0.5 ng mL^{-1} Pd (225.585 nm and 228.585 nm), from 0.5 pg to 1 μg mL^{-1} Os (225.585 nm), 0.03 μg mL^{-1} Pt (214.423 nm) and 0.02 and 0.03 μg mL^{-1} Au (242.795 nm and 267.595 nm, respectively).[62] The technique is more suitable for the determination of iridium, ruthenium and osmium compared with AAS.

The capabilities of emission flame spectrometric detection of noble metals using different atomizers were investigated.[85] DLs of 0.001–0.01 g t^{-1} were reached using $N_2O-C_2H_2$ flame. The effect of organic solvent (xylene, acetic aldehyde and nitrobenzene at a ratio of 4.5 : 4.5 : 1) introduced to the plasma on the ICPAES signals of iridium, rhodium, palladium and platinum has been studied.[86] No mutual effects of the analytes have been found. An increase in the signal of iridium was observed in the presence of copper. DLs (μg mL^{-1}): 0.029 for Ir, 0.11 for Rh, 0.024 for Pd and 0.073 for Pt were reported. The effect of HCl/HNO_3 acids on ICPAES signals of silver, palladium and gold in the analysis of Ag–Pd and Ag–Pd–Au–Cu–Zn alloys decreased with increase in acidity.[87]

Interfering effects of matrix elements may cause problems in the direct determination of low concentrations of analytes by the ICPAES technique. Base metals, (e.g. Fe, Cu, Ni, Cr, Ti, Mg and V) interfere with ICPAES signals of precious metals. In practice, the separation and preconcentration steps are generally applied prior to the detection of noble metals in rich matrices. Various separation and preconcentration procedures have been combined with ICPAES. Coprecipitation with CuS and 2-thiobenzothiazol followed by sorption on a suitable sorbent to separate copper was applied prior to the determination of Ir, Pt, Rh and Pd in copper-based minerals.[88] DLs of 5×10^{-6}% for Ir and Pt, 5×10^{-7}% for Rh and 1×10^{-7}% for Pd were reported. Concentrations of gold (5×10^{-8}%), platinum and palladium (2×10^{-7}%) in natural waters were detected using a Thiopan-13 fibrous complexing sorbent to preconcentrate the analytes.[89] The interference of iron and manganese on the ICPAES signal of gold has been eliminated by using the chelating anion exchanger SRAFION™ NMRR (containing thiourea active groups bound to styrene-divinyl-benzene matrix) to separate and preconcentrate gold from rich sediments.[90] A DL of 0.9 ng mL^{-1} Au (242.795 nm) was achieved. The combination of flow injection systems incorporating microcolumns of Amberlyst A-26 and sulfhydryl cotton with ICPAES results in DLs of about 1 μg L^{-1} Au.[91] The method was suitable for the determination of gold in waters and refinery effluent.

A 400- to 10 000-fold excess of Ca(II), Mg(II), Fe(III), Al(III), Zn(II), Ni(II), Cu(II), Mn(II), Zr(IV) to Ru(III) and Au(III) is tolerable while a poly(epoxy-melamine) chelating resin is used for separation and preconcentration of analytes at the ng mL^{-1} level.[92] Preliminary separation of osmium by extraction of OsO_4 is recommended to avoid common interferences (in particular from nickel and copper) in the AES determination of metal.[93] The direct injection of OsO_4 into a plasma torch provides an increase in sensitivity by a factor of 10 over conventional nebulization. Osmium when present in the (VIII) oxidation state gives the maximum atomic emission signal over a wide pH range.[94] The reduction of osmium to lower oxidation states results in lower volatility of the element and in a decrease in the intensity of atomic emission signal. Osmium, in high noble metal content alloys, can be determined together with other noble metals (without distillation) after decomposition under the action of potassium tetrafluorobromate and conversion of the complex fluorides obtained into completely soluble forms.[95] Direct ICPAES detection can be applied to monitor the levels of platinum complexes in blood after a chemotherapy treatment.[96]

Some approaches to the application of microwave-induced plasma atomic emission spectroscopy (MIPAES) to the analysis of noble metal samples have been presented.[97,98]

4.2.4 Inductively Coupled Plasma Mass Spectrometry

High speed, excellent DLs, wide dynamic range, the possibility of accurate multielement analysis and the unique capability of measuring isotopic ratios make the ICPMS technique the most promising for the determination of ultratrace levels of elements in complex matrices. ICPMS has found wide application to the determination of noble metals at ppb and sub-ppb levels in a large variety of geological, biological, environmental and industrial samples. A review of the application of ICPMS to analysis of noble metal samples has been published.[99]

Direct application of ICPMS to quantification of noble metals in complex, in particular in geological and environmental, matrices is limited owing to the low

concentrations determined, very often below the DLs offered, and matrix and spectral interferences. High salt concentrations can lead to suppression or enhancement of the ICPMS signal of analytes. Spectral interferences from ions and polyatomic groups, as well as from particular analytes, can occur owing to the limited resolution of the instrument used. Sample dilution, chemical separation, mathematical correction, calibration procedures such as standard addition and ID are used to eliminate interferences in ICPMS signals of noble metals.

The detection of precious metals at ppb and sub-ppb levels in complex matrices is generally preceded by a separation and preconcentration technique. The nickel sulfide fire assay is most frequently applied to separation and preconcentration of noble metals prior to the ICPMS measurement. Dry chlorination, acid dissolution and alkaline oxidizing fusion are also used. Direct sample introduction in the form of slurry, laser ablation and glow discharge technique provide the advantages of reduced sample preparation steps and lower amounts of chemicals used. However, the wide heterogeneity of the samples examined and the low concentrations to be determined seriously limit the use of direct sample introduction into the plasma.

The combination of nickel sulfide fire assay and ICPMS leads to the highest recovery of ultratraces of noble metals in geological materials.[100,101] Contamination due to the high amounts of salts employed is one of the drawbacks of the analytical procedure used. The effects of nickel and copper on the ICPMS signals of Pt, Pd, Ru and Ir were studied.[100] The addition of 500 ppm of nickel (as nitrate) resulted in up to ten-fold enhancement in the analyte ion (at 100 ppb level) count rate. Approximately 5000 ppm of nickel suppressed the platinum signal by 50%. A quite different interference of nickel was observed while electrothermal vaporization was used for sample introduction into the plasma. Enhancement factors of 4.6 for palladium, 8.1 for rhodium and 9.4 for platinum were observed when 500 ppm of nickel was added to the solutions containing 100 ppb each of analyte. At higher nickel concentrations the enhancement factor decreased to, for example, 4.4 for palladium at 10 000 ppm Ni. A smaller effect was observed from copper. Copper nitrate exhibited ca. 60% enhancement in the Pt signal compared with the chloride. The chloride is probably volatilized at temperatures below the vaporization temperature of the PGM whereas the nitrate is decomposed to the oxide and reduced to the metal. The use of electrothermal vaporization provides ca. 50% increase in the sensitivity of the Pt, Ru and Ir signals and 2.6% for Pd, compared with classical nebulization.

A comparison of aqua regia leaching and fire assay as separation and preconcentration methods for precious metals from rocks has been presented.[101] Partial leaching of iridium, osmium, rhodium and ruthenium by aqua regia was reported. Good recoveries were obtained for gold, palladium and platinum. DLs for Ru, Pd, Ir and Pt in rocks in the range of $0.3-2\,ng\,g^{-1}$ were reported for ID ICPMS combined with alkaline fusion and tellurium coprecipitation.[22] The anion-exchange separation step can be applied to eliminate potential isobaric interferences from some base metals on ICPMS determinations of Ru, Pd, Ir and Pt in geological materials.[102] DLs of $0.2-0.5\,ng\,g^{-1}$ for the metals examined (1 g sample) were achieved. Large samples (250 g) of rocks containing nanogram per gram levels of noble metals were analyzed by combining dry chlorination and ICPMS.[103] A review of the methods of separation of platinum metals from complex matrices (rocks, waters, plants), elimination of interferences, new introduction systems for ICPMS and measurement by ID has been presented.[104]

The ICPMS technique is particularly suitable for osmium isotopic analysis. A unique opportunity to evaluate the isotopic ratios of osmium and rhenium is of great value in geochronology where the rhenium–osmium decay scheme ($^{187}Re \longrightarrow {}^{187}Os$) is used to study the age of ore deposits and extraterrestrial samples.[105] A DL below 100 fg Os was reached for microelectrothermal vaporization ICPMS determination of osmium in meteorites.[106] Typical contents of 1000–2300 ppb of osmium and 220–260 ppb of rhenium in meteorites were estimated.[107]

ICPMS is the most suitable technique for the determination of noble metals at ppb and sub-ppb levels in a large variety of environmental samples. In 1975 environmental concern about the growing contamination from motorization led to the introduction of platinum metals as components of automotive catalysts in order to decrease the emission of carbon monoxide, unburnt hydrocarbons and nitrogen oxides in exhaust gases. Soon, it was established that those metals were released from the catalysts into the environment in the form of nanocrystals attached to alumina particles. Enriched, as compared with crustal abundances, platinum metal contents in the environmental materials (e.g. water, plants, airborne particles, dust, soils, urban gullypots, biotic materials and sediments) have been found.[108–115] Difficulties in the determination of ultratrace amounts of noble metals, particularly in complex environmental matrices, seriously limit the number of the published results. In practice, published data concern mainly platinum, which is the major noble metal component in catalysts and, from the analytical point of view, is much easier to examine than, for example, rhodium.

Oxides of hafnium, ytterbium and tungsten cause spectral interferences in ICPMS determination of

platinum. In particular hafnium, which occurs in the earth's crust at relatively higher concentrations than those of platinum, may enhance the platinum signal. The hafnium oxide signal overlaps the signals of the three most abundant Pt isotopes (^{194}Pt (32.9%), ^{195}Pt (33.8%) and ^{196}Pt (25.2%)). Hafnium should be removed from the samples before the detection of platinum. Extraction, ion-exchange chromatography, chelating resin and sorption on activated carbon were employed to separate platinum from interfering elements and to preconcentrate prior to ICPMS detection. The application of the ID technique to the evaluation of platinum concentration is a useful approach to eliminate the effects of interferents.

The small volume of effluents converted into aerosol by conventional nebulizers and the low efficiency of transport of the analyte into plasma are generally considered to be the weakest components in the successful coupling of plasma detectors to some pretreatment methods (e.g. liquid chromatography (LC)). The coupling of the thermospray system for the purpose of sample introduction into the plasma resulted in an increase in the sensitivity of the determination of platinum by a factor of 9.[115] Direct electrospray ionization (ESI) of the eluent into the plasma allowing the introduction of the total volume of the eluent into the plasma has been used in the analysis of some anticancer drugs.[116]

Analytical methods of determining platinum in environmental and biological samples have been reviewed.[117]

Nanogram levels of gold in natural waters (DL of 0.19 ng L^{-1})[118] and plant materials (DL of 0.04 ng mL^{-1})[119] were detected by the ICPMS technique.

The application of ICPMS to detection of noble metals in clinical samples involves relatively simple sample pretreatment steps. In practice, sample dilution in order to lower the concentration of dissolved solids is sufficient. The level of analytes in the reagents used, especially in acids, and the purity of the vessels used, require special attention. Significant and unpredictable memory effects can occur, particularly when the vessels are used to determine the noble metals within different concentration ranges.

ICPMS has been found to be an ideal technique for monitoring the level of platinum in the body fluids and tissues of patients treated with cisplatin and other anticancer drugs,[8,120] as well as physiological levels of noble metals in human body.[121,122] The use of high-resolution inductively coupled plasma mass spectrometry (HRICPMS) for the determination of physiological levels of Rh, Pd, Pt and Au in blood and detailed discussion of mass spectral interferences caused by Pb, Sr, Cu and Zn which can occur in blood have been presented.[123] The interfering effects of the elements examined (at $\mu g\,L^{-1}$ and $mg\,L^{-1}$ levels) on rhodium and palladium HRICPMS signals were recorded. The signals of gold and platinum were found not to be affected by spectral interference.

ICPMS was successfully used to determine platinum, palladium and rhodium in automotive catalysts.[124–127] The technique could be directly applied to the determination of platinum.[124] Instrumental correction was employed to determine palladium and rhodium owing to interferences from Pb^{2+} ions. The interferences from Zr and Y, present at appreciable amounts in modern autocatalysts, have been examined using glow discharge mass spectrometry (GDMS).[127] The effect of sample size on the results of the determination of Pt, Pd, Rh and Ti using a microwave digestion procedure employing a mixture of HF, HCl and HNO_3 or H_2SO_4, H_3PO_4 and aqua regia prior to the ICPMS measurement has been studied.[125] A comparison of the results obtained for rhodium and platinum using liquid nebulization (DL of 0.02 and 0.04 $\mu g\,g^{-1}$, respectively) and spark ablation (DL of 1.0 and 0.5 $\mu g\,g^{-1}$, respectively) ICPMS has been presented.[126]

4.2.5 X-Ray Fluorescence

XRF is widely used for qualitative and quantitative analysis of multicomponent samples. The high-speed, nondestructive nature of the analysis, the wide variety of the samples that can be examined (solutions, powders, alloys, etc.), the relatively simple sample preparation steps, the wide concentration ranges that can be determined (from parts per billion up to even 100%) and excellent precision are attractive features of the technique. XRF is widely applicable in analysis of noble metal samples.[128,129] It can be directly applied to the determination of analytes at concentration levels from $n \cong 10^{-4}$% to 70–80%, or to detection of lower concentrations after the preliminary separation and preconcentration steps. Fire assay and sorption methods are favored to precede the XRF measurement. Dilution methods, internal standards or mathematical correction are used to eliminate matrix effects on the XRF signals of metals. The requirement for standards with the same chemical composition and physical properties limits the wide application of the technique to the analysis of real samples.

Problems relating to the representativeness and homogeneity of the geological samples for XRF analysis of noble metals, as well as sample preparation steps applied have been discussed.[130] The effects of zinc and copper, the major matrix components in ores, rocks, minerals and concentrates on the determination of precious metals using a fire assay preconcentration step were investigated.[129] It was found, that the presence of more than 200 μg of Ni or Cu caused a reduction in the signals of Ru, Rh and Pd by 10%. A 20% reduction was observed for Os, Ir, Pt and Au while 400 μg of base metals were

present in the samples examined. The mutual effects of noble metals in their XRF signals can be neglected after selective extraction by Polyorgs™ X1-H as complexing sorbent.[131] XRF was successfully applied to the determination of the elemental composition of platinum metal complexes.[132]

Femtogram levels as an absolute DL of platinum were reported for XRF analysis of biological (blood, urine) and environmental (soil) samples.[133] The influence of acids, bases and salts on the XRF signal of platinum has been studied. XRF was found to be suitable for monitoring platinum in biopsy tissues of patients treated with cisplatin.[134] The technique was successfully applied to the determination of gold in various types of samples, e.g. ores,[135–137] rocks[20,138] and jewellery.[139,140]

4.3 Catalytic Kinetic Methods

The catalytic effects of noble metals on a large number of various reactions form the basis of kinetic methods of analysis. The low DLs (ppb and sub-ppb levels) and the high precision of the results are the main attractions of kinetic methods. However, poor selectivity and the large effect of the chemical form of analyte present in the sample examined on the results obtained strongly restrict the application of the methods to analysis of real samples. Transformation of the metals into a suitable form (e.g. Os(VIII), Ru(III), Rh(III), Ir(IV)) is generally required prior to the detection.[141]

The catalytic effects of precious metals on the oxidation of organic substances, mainly aromatic amines and phenol derivatives, by various inorganic reagents (e.g. IO_4^-, Ce(IV), H_2O_2, hexacyanoferrate(II)), as well as the redox reactions in systems containing inorganic oxidants (IO_4^-, ClO_3^-, BrO_3^-, MnO_4^-, Ce(IV), Fe(III)) and reductants (I^-, As(III), Sn(II)) form the basis of the majority of methods developed. The differing catalytic activity of particular metals in different media can provide the basis for their determination in each other's presence.[142,143]

Osmium and ruthenium are most frequently determined on the basis of the catalytic effects of Os(VIII) and Ru(III) on a large variety of redox reactions.[144] Methods are generally developed for pure solutions of both analytes easily obtained by separation of ruthenium and osmium by distillation. Extraction and ion-exchange chromatography are also applied to separate the noble metals from a large excess of interfering base metals.[145]

Spectrophotometry and chemiluminescence are the detection techniques most often used for monitoring the rate of the reactions examined. Polarographic methods based on catalytic currents recorded are also used to determine the metals.

4.4 Electrochemical Methods

Electrochemical methods are rarely applied to the analysis of complex noble metal samples. Difficulties in quantitative transformation of the analytes into electrochemically active compounds and severe interference from complex matrices limit the application of electrochemical methods to the analysis of real samples. Separation and preconcentration steps usually precede the final detection.[146]

Polarographic techniques are most widely applicable to the analysis of noble metal samples. Alternating current polarography, differential pulse polarography and anodic stripping voltammetry proved to be useful for determining sub-ppb levels of noble metals.

Adsorptive voltammetry based on catalytic reduction of protons by platinum–formazone (a condensation product of formaldehyde and hydrazine) complex provides a sensitive method for the determination of platinum in clinical[147–149] and environmental samples.[150] The technique was found to be suitable for the determination of natural levels of platinum ($0.1–2.8\,\mu g\,L^{-1}$[149] and $\leq 0.8–15\,ng\,L^{-1}$[150]) in human blood and $0.5–15\,ng\,L^{-1}$ in urine.[149] Large interfering effects from organic matrices can occur. A comparison of voltammetric results with those of GFAAS for platinum in clinical[147] and environmental samples[150] has been presented. Sub-nanogram per gram levels of platinum (DL of $21\,pg\,L^{-1}$) and rhodium (DL of $5.6\,pg\,L^{-1}$) in solutions containing formaldehyde and hydrazine in a H_2SO_4/HCl medium can be simultaneously determined.[151]

Adsorptive accumulation of the complexes of platinum and ruthenium with, for example, PAN on the surface of a hanging mercury drop electrode followed by reduction of the absorbed complex during the cathodic scan forms the basis of an extremely sensitive method (DL of 3.2×10^{-10} M Pt and 4.1×10^{-10} M Ru).[152] Different complexing ligands, various supporting electrolytes and the DLs achieved for the determination of platinum have been reviewed.[153] Detailed discussion of voltammetric methods for the determination of platinum metals has been presented.[154]

An alternating current polarographic method using the extraction of 8-thioquinolinates of Ru, Os and Ir allows their determination (DL of 2.2×10^{-3} M Ru, 6.1×10^{-7} M Os and 8×10^{-8} M Ir) in the presence of Pt(II), Ru(III), Au(III), Ag(I), Ni(II) and Cu(II).[155]

Coulometric methods applied to the determination of noble metals have been reviewed.[156] Coulometry was applied to the determination of platinum (in the range of 0.5–5 mg) in dead catalysts[157] and gold (>10%) in Au–Ag–Cu alloys.[158] Coulometric titration of platinum has been successfully used in the determination of large amounts of platinum (98–99.9%) in platinum alloys.[159]

4.5 Neutron Activation Analysis

The extremely high sensitivities offered by neutron activation analysis (NAA) make the technique particularly suitable for the determination of ppb and sub-ppb levels of noble metals in a large variety of materials. The analysis is accomplished by the production and detection of specific radionuclides of particular elements.

Thermal neutron activation of metals results in radionuclides of: ^{103}Ru ($t_{1/2} = 39.8$ day); ^{104}Rh ($t_{1/2} = 42$ s); ^{191m}Os ($t_{1/2} = 14$ h) with its daughters ^{191}Os ($t_{1/2} = 16$ day) and ^{193}Os ($t_{1/2} = 31$ h). Radiogold daughter isotope, one of the products of thermal neutron activation of platinum (^{198}Pt (n, γ) ^{199}Pt $\longrightarrow$ ^{199}Au ($t_{1/2} = 3.15$ day), is used as a measure of platinum content. NAA of palladium is based on the reaction ^{108}Pd$(n, \gamma)^{109}$Pd ($t_{1/2} = 13.5$ h). Gold can be determined with high sensitivity on the basis of the reaction: ^{197}Au$(n, \gamma)^{198}$Au producing ^{198}Au nuclide ($t_{1/2} = 2.7$ day).

Direct application of NAA to the analysis of real samples is restricted owing to their complexity and the numerous interfering effects from matrix elements. In particular, the interfering effects can be critical in the determination of very low concentrations of analytes. Separation and concentration are often applied, either before or after the irradiation. Fire assay preconcentration, ion-exchange chromatography and coprecipitation with a suitable collector are most frequently used to separate noble metals from an interfering matrix.

Numerous neutron activation procedures for the determination of noble metals in a variety of geological, biological and environmental materials have been developed.[160–168] The combination of fire assay with NAA, particularly suitable for analysis of geological samples, has been studied by many authors. DLs of 10^{-3}–10^{-1} ppm of noble metals;[168] 0.2 (Os), 0.005 (Ir), 1.0 (Ru), 0.35 (Rh), 2.5 (Pt), 0.8 (Pd) and 0.001 (Au) (μg kg^{-1}) (100 g sample);[166] 0.002 (Rh), 0.025 (Pd), 0.018 (Pt), 0.0002 (Ir), 0.002 (Au), 0.020 (Os) and 0.2 (Ru) (mg kg^{-1})[164] of noble metals in geological samples using fire assay and NAA have been reported.

5 QUALITY CONTROL AND ASSURANCE

Reliability of the results is the main concern in the analysis of noble metal samples. The accuracy assessment is of great importance in analysis of various samples containing analytes in different concentration ranges.

The choice of an appropriate analytical procedure is the fundamental factor affecting the quality of the results obtained. It should be noted that procedures suitable for the determination of noble metals in a specific matrix may not be applicable to their determination in a different material.

All steps in the analytical procedure used can provide severe sources of errors. High heterogeneity of the examined samples and low concentrations to be determined require special attention during sampling and sample handling. The possibility of losses or contamination during the sample preparation steps requires special care. High purity reagents (e.g. acids of Suprapur grade or purified by sub-boiling distillation) are necessary when ultratraces of precious metals (ppb and sub-ppb levels) are determined. Standards of an appropriate composition containing analytes in the definite form are required for calibration, especially when some analytical techniques are employed (e.g. LC and spectrophotometry). Spectral overlap and multiplicate interference (matrix suppression or enhancement) often occur. Estimation of the value obtained with the aid of statistics is very important in data assessment and presentation.

Analysis of certified reference materials (CRMs), interlaboratory comparisons, the use of definite analytical methods, comparison of the results obtained by different analytical techniques and recovery studies are used to check the quality of the results. Unfortunately, the use of CRMs in analysis of noble metal samples is considerably restricted owing to the limited number of materials available, their high costs and limited supplies. Geological materials, ores and concentrates provide the largest group of CRMs available. A majority of them have been supplied by Canada Center for Mineral and Energy Technology (CANMET).[3] The ore sample of certified values for all noble metals (SARM-7), is a product of National Institute for Metallurgy, Johannesburg, South Africa. The lack of CRMs for biological and environmental materials of different origin is a fundamental problem.

Calibration procedures, such as standard addition or ID, are recommended when the nature of the matrix is unknown or cannot be easily matched with the standards. Comparison of the results obtained by different analytical techniques is recommended. Interlaboratory studies may improve the quality of the determination.

ACKNOWLEDGMENTS

The author expresses special gratitude to Dr. Jan Kryński from the Institute of Geodesy and Cartography, Warsaw, Poland, for designing the map of world sources of noble metals.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AES	Atomic Emission Spectroscopy

CANMET	Canada Center for Mineral and Energy Technology
CMDTC	bis(Carboxymethyl)dithiocarbamate
CP	Cetylpyridinium
CRM	Certified Reference Material
CTA	Cetyltrimethylammonium
DDTC	Diethyldithiocarbamate
DIPE	Diisopropyl ether
DL	Detection Limit
DMSO	Dimethyl Sulfoxide
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray Ionization
FAAS	Flame Atomic Absorption Spectroscopy
GDMS	Glow Discharge Mass Spectrometry
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
GGG	Gadolinium Gallium Garnet
HPLC	High-performance Liquid Chromatography
HRICPMS	High-resolution Inductively Coupled Plasma Mass Spectrometry
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ID	Isotope Dilution
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
MIBK	Methyl Isobutyl Ketone
MIP	Microwave-induced Plasma
MIPAES	Microwave-induced Plasma Atomic Emission Spectroscopy
NAA	Neutron Activation Analysis
PAN	1-(2-Pyridylazo)-2-naphthol
PAR	4-(2-Pyridylazo)resorcinol
PGM	Platinum Group Metal
TAAP	2-(2-Thiazolylazo)-5-diethyl-*m*-aminophenol
TAN	1-Thiazolylazo-2-naphthol
TAR	4-(2-Thiazolylazo)resorcinol
TBP	Tributyl phosphate
TOPO	Trioctylphosphine oxide
TPP	Triphenylphosphine
XRF	X-Ray Fluorescence
YAG	Yttrium Aluminum Garnet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Atomic Fluorescence in Environmental Analysis • Biological Samples in Environmental Analysis: Preparation and Cleanup • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Detection and Quantification of Environmental Pollutants • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by Electrochemical Methods • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Luminescence in Environmental Analysis • Neutron Activation in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Soil Instrumental Methods • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Industrial Hygiene **(Volume 5)**
Dust, Measurement of Trace Elements in

Industrial Hygiene cont'd **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Mass Spectrometry in Pharmaceutical Analysis

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in • Nickel Ore and Metals Analysis • X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Glow Discharge Optical Spectroscopy and Mass Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical

Emission Spectrometry • Microwave-induced Plasma Systems in Atomic Spectroscopy

Chemometrics **(Volume 11)**
Signal Processing in Analytical Chemistry

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Ion Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Thin-layer Chromatography

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Liquid Chromatography/Mass Spectrometry

Nuclear Methods **(Volume 14)**
Elemental Analysis by Isotope Dilution • Instrumental Neutron Activation Analysis • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Mass Spectrometry of Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Sample Preparation for X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Microwave Techniques • Quality Assurance in Analytical Chemistry

REFERENCES

1. D. McDonald, *A History of Platinum*, Johnson Matthey, London, 1960.
2. D. McDonald, L.B. Hunt, *A History of Platinum and its Allied Metals*, Johnson Matthey, London, 1982.
3. J.C. Van Loon, R.R. Barefoot, *Determination of the Precious Metals: Selected Instrumental Methods*, John Wiley & Sons, Chichester, 1991.
4. A.I. Busev, V.M. Ivanov, *Analytical Chemistry of Gold* (in Russian), Nauka, Moscow, 1973.
5. S.E. Livingstone, *The Chemistry of Ruthenium, Rhodium, Palladium, Osmium, Iridium and Platinum*, Pergamon Press, Oxford, 1973.
6. F.R. Hartley (ed.), *Chemistry of the Platinum Group Metals*, Elsevier, Amsterdam, 1991.
7. N.G. Randolph, 'Precious Metals', *Pure Appl. Chem.*, **65**(12), 2411–2416 (1993).
8. R.R. Barefoot, J.C. Van Loon, 'Determination of Platinum and Gold in Anticancer and Antiarthritic Drugs and Metabolites', *Anal. Chim. Acta*, **334**(1–2), 5–14 (1996).
9. J.C. Van Loon, 'Accurate Determination of the Noble Metals. I. Sample Decomposition and Methods of Separation', *Trends Anal. Chem.*, **3**(10), 272–275 (1984).
10. G.S. Reddi, C.R.M. Rao, T.A.S. Rao, S. Vijaya Lakshmi, R.K. Prabhu, T.R. Mahalingam, 'Nickel Sulfide Fire Assay-ICPMS Method for the Determination of Platinum Group Elements: A Detailed Study on the Recovery and Losses at Different Stages', *Fresenius' J. Anal. Chem.*, **348**(5–6), 350–352 (1994).
11. H. Urban, F. Zereini, B. Skerstupp, M. Tarkian, 'The Determination of Platinum Group Elements (PGE) by Nickel Sulfide Fire-Assay: Coexisting PGE-Phases in the Nickel Sulfide Button', *Fresenius' J. Anal. Chem.*, **352**(6), 537–543 (1995).
12. A. Frimpong, B.J. Fryer, H.P. Longerich, Z. Chen, S.E. Jackson, 'Recovery of Precious Metals Using Nickel Sulfide Fire Assay Collection: Problems at Nanogram per Gram Concentrations', *Analyst*, **120**(6), 1675–1680 (1995).
13. Ch.J.B. Gowing, P.J. Potts, 'Evaluation of a Rapid Technique for the Determination of Precious Metals in Geological Samples Based on a Selective Aqua Regia Leach', *Analyst*, **116**(8), 773–779 (1991).
14. B.J. Perry, J.C. Van Loon, D.V. Speller, 'Dry-chlorination Inductively Coupled Plasma Mass Spectrometric Method for the Determination of Platinum Group Elements in Rocks', *J. Anal. At. Spectrom.*, **7**(6), 883–888 (1992).
15. S.I. Ginsburg, N.A. Ezerskaya, I.V. Prokof'eva, N.F. Fedorenko, V.I. Shlenskaya, N.K. Belskii, *Analytical Chemistry of Platinum Metals*, Nauka, Moscow, 1972.
16. M. Balcerzak, E. Święcicka, 'Determination of Ruthenium and Osmium in Each Other's Presence in Chloride Solutions by Direct and Third-order Derivative Spectrophotometry', *Talanta*, **43**(3), 471–478 (1996).
17. F.E. Beamish, J.C. Van Loon, *Recent Advances in the Analytical Chemistry of the Noble Metals*, Pergamon Press, Oxford, 1972.
18. S. Kallmann, 'A Survey of the Determination of the Platinum Group Elements', *Talanta*, **34**(8), 677–698 (1987).
19. J.S. Gupta, 'Determination of Trace and Ultra-trace Amounts of Noble Metals in Geological and Related Materials by Graphite Furnace Atomic Absorption Spectrometry after Separation by Ion-exchange and

Co-precipitation with Tellurium', *Talanta*, **36**(6), 651–656 (1989).
20. R. Eller, F. Alt, G. Tölg, H.J. Tobschall, 'An Efficient Combined Procedure for the Extreme Trace Analysis of Gold, Platinum, Palladium and Rhodium with the Aid of Graphite Furnace Atomic Absorption Spectrometry and Total-reflection X-Ray Fluorescence Analysis', *Fresenius Z. Anal. Chem.*, **334**(8), 723–739 (1989).
21. H. Niskavaara, E. Kontas, 'Reductive Coprecipitation as a Separation Method for the Determination of Gold, Palladium, Platinum, Rhodium, Silver, Selenium and Tellurium in Geological Samples by Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **231**(2), 273–282 (1990).
22. J. Enzweiler, P.J. Potts, K.E. Jarvis, 'Determination of Platinum, Palladium, Ruthenium and Iridium in Geological Samples by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry using a Sodium Peroxide Fusion and Tellurium Coprecipitation', *Analyst*, **120**(5), 1391–1396 (1995).
23. S.J. Al-Bazi, A. Chow, 'Platinum Metals–Solution Chemistry and Separation Methods (Ion-Exchange and Solvent Extraction)', *Talanta*, **31**(10A), 815–836 (1984).
24. M. Mojski, 'Solvent Extraction of Platinum Metals. Neutral Oxygen Containing Extractants', *Chem. Anal. (Warsaw)*, **25**(6), 919–944 (1980).
25. L. Jones, I. Nel, K.R. Koch, 'Polyurethane Foams as Selective Sorbents for Noble Metals. Quantitative Extraction and Separation of Rhodium from Iridium in HCl Containing $SnCl_2$', *Anal. Chim. Acta*, **182**(1), 61–70 (1986).
26. S.J. Al-Bazi, A. Chow, 'Extraction of Ruthenium Thiocyanate and its Separation from Rhodium by Polyurethane Foams', *Talanta*, **31**(3), 189–193 (1984).
27. A. Chow, S.J. Al-Bazi, 'Extraction of Osmium Thiocyanate and its Separation from Ruthenium by Polyurethane Foam', *Anal. Chim. Acta*, **157**(1), 83–89 (1984).
28. G.I. Malofeeva, G.V. Myasoedova, M.P. Volynets, 'Sorption Methods of Concentrating Noble Metals', *Mikrochim. Acta*, **I**, 391–398 (1978).
29. G.V. Myasoedova, I.I. Antokol'skaya, S.B. Savvin, 'New Chelating Sorbents for Noble Metals', *Talanta*, **32**(12), 1105–1112 (1985).
30. R.K. Dubey, A. Bhalotra, M.K. Gupta, B.K. Puri, 'Differential Pulse Polarographic Determination of Rh(III) and Ru(III) in Synthetic Samples after Preconcentration of Their Quinolin-8-olates Complexes onto Crystalline Naphthalene', *Microchem. J.*, **58**(1), 117–126 (1998).
31. M.A. Taher, S. Puri, R.K. Bansal, B.K. Puri, 'Derivative Spectrophotometric Determination of Iridium after Preconcentration of its 1-(2-Pyridylazo)-2-Naphthol Complex on Microcrystalline Naphthalene', *Talanta*, **45**(2), 411–416 (1998).
32. I.P. Alimarin, E.M. Basova, T.A. Bol'shova, V.M. Ivanov, 'Application of Column Liquid Chromatography to Preconcentration, Separation and Determination of Platinum Metals', *Zh. Anal. Khim.*, **41**(1), 5–34 (1986).
33. I. Jarvis, M.M. Totland, K.E. Jarvis, 'Assessment of Dowex 1-X8-based Anion Exchange Procedures for the Separation and Determination of Ruthenium, Rhodium, Palladium, Iridium, Platinum and Gold in Geological Samples by ICPMS', *Analyst*, **122**(1), 19–26 (1997).
34. R.S. Aygün, M. Merdivan, N. Külcü, 'Determination of Platinum, Palladium and Rhodium in Pt-catalysts by HPTLC using Densitometry', *Mikrochim. Acta*, **127**(3–4), 233–238 (1997).
35. D. Klueppel, N. Jakubowski, J. Messerschmidt, D. Stuewer, D. Klockow, 'Speciation of Platinum Metabolites in Plants by Size-exclusion Chromatography and ICPMS', *J. Anal. At. Spectrom.*, **13**(4), 255–262 (1998).
36. S.G. Matz, R.C. Elder, K. Tepperman, 'Liquid Chromatography with an ICPMS Detector for Simultaneous Determination of Gold Drug Metabolites and Related Metals in Human Blood', *J. Anal. At. Spectrom.*, **4**(8), 767–771 (1989).
37. B. Baraj, A. Sastre, M. Martinez, K. Spahiu, 'Simultaneous Determination of Chloride Complexes of Platinum(IV) and Palladium(II) by Capillary Zone Electrophoresis with Direct UV Absorbance Detection', *Anal. Chim. Acta*, **319**(1–2), 191–197 (1996).
38. F.E. Beamish, 'A Critical Evaluation of the Gravimetric Methods for the Determination of the Platinum Metals', *Talanta*, **1**(1–2), 3–29 (1958).
39. F.E. Beamish, 'A Critical Review of Gravimetric Methods for Determination of the Noble Metals – II', *Talanta*, **13**(6), 773–801 (1966).
40. F.E. Beamish, 'A Critical Evaluation of Colorimetric Methods for Determination of the Noble Metals – III. Palladium and Platinum', *Talanta*, **12**(8), 743–772 (1965).
41. F.E. Beamish, 'A Critical Evaluation of Colorimetric Methods for Determination of the Noble Metals – III. Rhodium, Iridium, Ruthenium, Osmium and Gold', *Talanta*, **12**(9), 789–816 (1965).
42. Z. Marczenko, *Separation and Spectrophotometric Determination of Elements*, Ellis Horwood, Chichester, 1986.
43. M. Balcerzak, 'Platinum Metals–Tin(II) Chloride Complexes. Analytical Applications. A Review', *Analusis*, **22**(7), 353–359 (1994).
44. Z. Marczenko, M. Balcerzak, S. Kuś, 'Spectrophotometric Determination of Ruthenium and Osmium', *Talanta*, **27**(11B), 1087–1089 (1980).
45. M. Balcerzak, K. Komar, E. Święcicka, K. Kasiura, 'Selective Spectrophotometric Determination of Osmium in Aqueous Iodide Solutions or after Extraction with Trioctylamine–Carbon Tetrachloride', *Anal. Sci.*, **13**(1), 33–36 (1997).
46. G. Talsky, *Derivative Spectrophotometry*, VCH, Weinheim, 1994.

47. M. Balcerzak, E. Święcicka, 'Determination of Ruthenium and Osmium in Each Other's Presence in Chloride Solutions by Direct and Third-order Spectrophotometry', *Talanta*, **43**(3), 471–478 (1996).
48. M. Balcerzak, S. Kuś, 'Determination of Ruthenium and Osmium with Tin(II) Chloride by Second-order Derivative Spectrophotometry', *Anal. Sci.*, **10**(1), 65–69 (1994).
49. R. Kuroda, Y. Hayashibe, K. Yoshitsuka, 'Determination of Palladium, Platinum and Gold by First-derivative Spectrophotometry', *Fresenius' J. Anal. Chem.*, **336**(6), 494–497 (1990).
50. N. Obarski, Z. Marczenko, 'Simultaneous Determination of Gold, Platinum and Palladium as Bromide Complexes using Derivative Spectrophotometry', *Chem. Anal. (Warsaw)*, **40**(3), 385–393 (1995).
51. Z. Marczenko, N. Obarski, 'Simultaneous Determination of Palladium and Platinum as Iodide Complexes using Third-order Derivative Spectrophotometry', *Chem. Anal. (Warsaw)*, **39**(3), 347–354 (1994).
52. M. Balcerzak, E. Święcicka, 'Rapid Simultaneous Determination of Ruthenium and Osmium in Aqueous Solutions of Their Tetroxides by Second-order Derivative Spectrophotometry', *Anal. Chim. Acta*, **349**(1–3), 53–57 (1997).
53. J. Tummavuori, S. Hankala, L. Siltaloppi, 'Determination of Platinum and Iridium in Reforming Catalysts by Second-derivative and Absorptiometric Spectrophotometry', *Fresenius' J. Anal. Chem.*, **339**(3), 145–150 (1991).
54. M. Balcerzak, E. Święcicka, E. Balukiewicz, 'Determination of Platinum and Ruthenium in Pt and Pt-Ru Catalysts with Carbon Support by Direct and Derivative Spectrophotometry', *Talanta*, **48**(1), 39–47 (1999).
55. B. Morelli, 'Determination of Osmium(VIII) and Palladium(II) in Mixtures by Derivative Spectrophotometry', *Anal. Lett.*, **18**(A19), 2453–2466 (1985).
56. S. Kuś, Z. Marczenko, 'Determination of Microgram Amounts of Platinum as Dithizonate in the Presence of Palladium by Second-derivative Spectrophotometry', *Analyst*, **114**(2), 207–210 (1989).
57. K. Malathi, M. Subbaiyan, 'Simultaneous Determination of Ruthenium(III) and Rhodium(III) Using Octadecyldithiocarbamate by First-derivative Spectrophotometry', *Talanta*, **42**(10), 1487–1491 (1995).
58. Z. Marczenko, 'New Type of Flotation of Ion-association Compounds of Complexes of Multicharged Anions with Basic Dyes', *Pure Appl. Chem.*, **57**(6), 849–854 (1985).
59. Z. Marczenko, K. Kalinowski, 'Flotation-spectrophotometric Methods for the Determination of Platinum Metals and Gold using Basic Dyes', *Chem. Anal. (Warsaw)*, **32**(4), 451–463 (1987).
60. M. Balcerzak, 'Solvent Flotation with Basic Dyes as a Method of Separation of Small Amounts of Ruthenium and Osmium for Spectrophotometry', *Analusis*, **20**(6), 295–301 (1992).
61. J.C. Van Loon, 'Accurate Determination of the Noble Metals. II. Determination Methods', *Trends Anal. Chem.*, **4**(1), 24–29 (1985).
62. R. Łobiński, Z. Marczenko, *Spectrochemical Trace Analysis for Metals and Metalloids*, Elsevier, Amsterdam, 1996.
63. S. Lin, Ch. Zheng, G. Yun, 'Determination of Palladium by Flame Atomic Absorption Spectrometry Combined On-line with Flow Injection Preconcentration Using a Micro-column Packed with Activated Carbon Fiber', *Talanta*, **42**(7), 921–926 (1995).
64. M. Merdivan, R.S. Aygun, N. Kulcu, 'Flame Atomic Absorption Spectrometric Determination of Platinum, Palladium and Rhodium in Catalysts', *At. Spectrosc.*, **18**(4), 122–126 (1997).
65. L.V. Zelentsova, V.A. Shchipachev, L.M. Levchenko, 'Atomic Absorption Determination of Platinum and Rhenium in Deactivated Catalysts Based on γ-Alumina', *Zh. Anal. Khim.*, **47**(10–11), 1889–1892 (1992).
66. S. Arpadjan, L. Jordanova, I. Karadjova, 'Reductive Coprecipitation and Extraction as Separation Methods for the Determination of Gold in Copper Concentrates by AAS and ICP–AES', *Fresenius' J. Anal. Chem.*, **347**(12), 480–482 (1993).
67. K. Pyrzyńska, 'Atomic Absorption Spectrophotometric Determination of Gold with Preconcentration by Donnan Dialysis', *Talanta*, **41**(3), 381–386 (1994).
68. W.B. Rowston, J.M. Ottaway, 'Determination of Noble Metals by Carbon Furnace Atomic Absorption Spectrometry. Part I. Atom Formation Processes', *Analyst*, **104**(1240), 645–659 (1979).
69. N.A. Borshch, Yu.A. Zolotov, L.N. Kolonina, O.M. Petrukhin, V.N. Shevchenko, O.A. Shiryaeva, 'Extraction–Atomic Absorption Determination of Platinum Metals', *Zh. Anal. Khim.*, **35**(12), 2369–2373 (1980).
70. S.A. Wood, D. Vlassopoulos, A. Mucci, 'Effect of Concentrated Matrices on the Determination of Trace Levels of Platinum and Gold in Aqueous Samples Using Solvent Extraction-Zeeman Effect Graphite Furnace Atomic Absorption Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **229**(2), 227–238 (1990).
71. V.G. Torgov, M.G. Demidova, T.M. Korda, N.K. Kalish, R.S. Shulman, 'Extraction–Atomic Absorption Spectrometric Method for the Determination of Platinum Group Elements and Gold in Copper–Nickel Ores Using an Autoclave Sample Decomposition Technique', *Analyst*, **121**(4), 489–494 (1996).
72. J. Begerow, M. Turfeld, L. Dunemann, 'Determination of Physiological Noble Metals in Human Urine Using Liquid–Liquid Extraction and Zeeman Electrothermal Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **340**(1–3), 277–283 (1997).
73. S.J.G. Gupta, 'Determination of Trace and Ultra-trace Amounts of Noble Metals in Geological and Related Materials by Graphite Furnace Atomic Absorption

Spectrometry after Separation by Ion-exchange or Co-precipitation with Tellurium', *Talanta*, **36**(6), 651–656 (1989).
74. S.J.G. Gupta, 'Determination of Noble Metals in Silicate Rocks, Ores and Metallurgical Samples by Simultaneous Multi-element Graphite Furnace Atomic Absorption Spectrometry with Zeeman Background Correction', *Talanta*, **40**(6), 791–797 (1993).
75. M.L. Lee, G. Tölg, E. Beinrohr, P. Tschöpel, 'Preconcentration of Palladium, Platinum and Rhodium by On-Line Sorbent-Extraction for Graphite Furnace Atomic Absorption Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **272**(2), 193–203 (1993).
76. P. Di, D.E. Davey, 'On-line Preconcentration and Separation of Palladium, Platinum and Iridium Using α-Aminopyridine Resin with Flame Atomic Absorption Spectrometry', *Talanta*, **42**(5), 685–692 (1995).
77. E.B. Stavnivenko, I.V. Kubrakova, N.I. Shcherbinina, G.V. Myasoedova, N.M. Kuz'min, 'Preconcentration of Platinum, Palladium and Rhodium on Polyorgs™ XXVII Sorbent and Their Subsequent Determination by Electrothermal Atomic Absorption Spectrometry', *Zh. Anal. Khim.*, **50**(12), 1243–1246 (1995).
78. I.V. Kubrakova, T.F. Kudinova, N.M. Kuz'min, I.A. Kovalev, G.I. Tsysin, Yu.A. Zolotov, 'Determination of Low Levels of Platinum Group Metals: New Solutions', *Anal. Chim. Acta*, **334**(1–2), 167–175 (1996).
79. Á. Lévai, A. Lásztity, K. Zih-Perényi, Zs. Horváth, 'Graphite Furnace Atomic Absorption Spectrometry Determination and On-line Preconcentration of Palladium', *Microchem. J.*, **58**(3), 272–280 (1998).
80. M.L. Rakhlina, M.Yu. Lomteva, G.I. Ermolina, 'Atomic Absorption Determination of Platinum Group Metals and Gold after Precipitation Concentration on Copper Sulphide', *Zavod. Lab.*, **51**(11), 91–93 (1985).
81. Z. Aneva, S. Arpadjan, I. Kalaidjeva, 'Graphite Furnace Atomic Absorption Spectrometric Determination of Iridium, Rhodium and Ruthenium in High-purity Platinum after Matrix Extraction', *Anal. Chim. Acta*, **236**(2), 385–389 (1990).
82. R. Boisvert, M. Bergeron, J. Turcotte, 'Re-examination of the Determination of Palladium, Platinum and Rhodium in Rocks by Nickel Sulphide Fire-assay Followed by Graphite Furnace Atomic Absorption Measurements', *Anal. Chim. Acta*, **246**(2), 365–373 (1991).
83. I.V. Kubrakova, G.M. Varshal, T.F. Kudinova, 'Peculiarities of Atomic Absorption Determination of Noble Metals by Direct Introduction of Organic Sorbents into a Graphite Furnace', *Zh. Anal. Khim.*, **42**(1), 126–131 (1987).
84. M.W. Hinds, V.V. Kogan, 'Assay of Precious Metals by Solid Sampling Atomic Spectrometry', *Spectroscopy*, **10**(5), 14–18 (1995).
85. E.D. Prudnikov, Yu.S. Shapkina, 'Emission Flame Spectrometry of Trace Platinum Group Metals, Gold and Silver Traces', *Zh. Anal. Khim.*, **43**(7), 1261–1265 (1988).
86. I.F. Seregina, G.L. Bukhbinder, L.N. Shabanova, E.N. Gil'bert, O.M. Petrukhin, Yu.A. Zolotov, 'Solvent Extraction Atomic Emission Determination of Platinum Metals Using ICP Spectrometry', *Zh. Anal. Khim.*, **41**(5), 861–869 (1986).
87. B.F. Abramović, K.S. Horváth, F.F. Gaál, O.P. Gaál, 'Determination of Silver, Palladium and Gold in Some Alloys by Titrimetric Methods and Inductively Coupled Plasma Atomic Emission Spectrometry', *Microchem. J.*, **48**(2), 137–150 (1993).
88. L.I. Pavlenko, I.A. Popova, E.V. Marcheva, G.V. Myasoedova, G.I. Malofeeva, 'Chemical-spectrographic Determination of Platinum Metals', *Zh. Anal. Khim.*, **35**(4), 717–724 (1980).
89. E.M. Tat'yankina, 'Atomic-emission Spectrophotometric Determination of Gold, Platinum and Palladium in Solutions after the Sorption Preconcentration with a Thiopan-13 Fibrous Complexing Sorbent', *Zh. Anal. Khim.*, **51**(5), 498–501 (1996).
90. M. Ochsenkühn-Petropulu, P. Schramel, 'Separation and Determination of Gold from Iron Rich Matrices by an Off/On-line Preconcentration System Coupled to ICP AES', *Mikrochim. Acta*, **119**(3–4), 265–276 (1995).
91. M.M. Gomez, C.W. McLeod, 'Trace Enrichment and Determination of Gold by Flow-injection Inductively Coupled Plasma Spectrometry. Part 1. Inductively Coupled Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **8**(3), 461–465 (1993).
92. Ch. Xi-Jun, G. Bo-Lin, S. Zhi-Xing, Y. Dong, L. Xing-Yin, 'ICP–OES Determination of Traces of Ruthenium, Gold, Vanadium and Titanium Preconcentrated and Separated by a New Poly(Epoxy-melamine) Chelating Resin from Solutions', *Fresenius' J. Anal. Chem.*, **360**(6), 736–739 (1998).
93. M. Akhoond, A.A. Ensafi, A. Massoumi, A. Safavi, 'Detection of Osmium by Flame Atomic Emission Spectrometry after Extraction as Osmium Tetroxide into MIBK', *Microchem. J.*, **45**(3), 365–369 (1992).
94. A.J. Lopez-Molinero, J.R. Castillo, J.M. Mermet, 'Observations on the Determination of Osmium by Inductively Coupled Plasma Atomic-emission Spectroscopy', *Talanta*, **37**(9), 895–899 (1990).
95. V.N. Mit'kin, S.V. Zemskov, B.I. Zaksas, E.A. Petrova, A.N. Kurskii, 'Analysis of Placer Platinum by ICP AES', *Zh. Anal. Khim.*, **46**(12), 2416–2422 (1991).
96. V. Di Noto, D. Ni, L. Dalla Via, F. Scomazzon, M. Vidali, 'Determination of Platinum in Human Blood Using Inductively Coupled Plasma Atomic-emission Spectrometry with an Ultrasonic Nebulizer', *Analyst*, **120**(6), 1669–1673 (1995).
97. K. Jankowski, 'Application of the Microwave-induced Plasma Atomic-emission Spectrometry to Platinum-

group Metals Analysis', *Chem. Anal. (Warsaw)*, **39**(3), 381–387 (1994).
98. F. Liang, D. Zhang, Y. Lei, H. Zhang, Q. Jin, 'Determination of Selected Noble Metals by MPT-AES Using a Pneumatic Nebulizer', *Microchem. J.*, **52**(2), 181–187 (1995).
99. B.J. Perry, R.R. Barefoot, J.C. Van Loon, 'Inductively Coupled Plasma Mass Spectrometry for the Determination of Platinum Group Elements and Gold', *Trends Anal. Chem.*, **14**(8), 388–397 (1995).
100. D.C. Gregoire, 'Determination of Platinum, Palladium, Ruthenium and Iridium in Geological Materials by Inductively Coupled Plasma Mass Spectrometry with Sample Introduction by Electrothermal Vaporisation', *J. Anal. At. Spectrom.*, **3**(2), 309–314 (1988).
101. R. Juvonen, E. Kallio, T. Lakomaa, 'Determination of Precious Metals in Rocks by Inductively Coupled Plasma Mass Spectrometry Using Nickel Sulphide Concentration. Comparison with Other Pre-treatment Methods', *Analyst*, **119**(4), 617–621 (1994).
102. Y.V. Yi, A. Masuda, 'Simultaneous Determination of Ruthenium, Palladium, Iridium and Platinum at Ultra-trace Levels by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry in Geological Samples', *Anal. Chem.*, **68**(8), 1444–1450 (1996).
103. B.J. Perry, D.V. Speller, R.R. Barefoot, J.C. Van Loon, 'Large Sample, Dry Chlorination, ICP MS Analytical Method for the Determination of Platinum Group Elements and Gold in Rocks', *Can. J. Appl. Spectrosc.*, **38**(5), 131–136 (1993).
104. T. Paukert, 'Determination of Noble Metals in Rocks, Waters and Plants by ICP MS', *Chem. Listy*, **87**(1), 8–15 (1993).
105. J.M. Richardson, A.P. Dickin, R.H. McNutt, J.I. McAndrew, S.B. Beneteau, 'Analysis of a Rhenium–Osmium Solid-solution Spike by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **4**(5), 465–471 (1989).
106. T. Hirata, T. Akagi, H. Shimizu, A. Masuda, 'Determination of Osmium and Osmium Isotope Ratios by Micro-electrothermal Vaporization Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **61**(20), 2263–2266 (1989).
107. J.W. Morgan, R.J. Walker, 'Isotopic Determination of Rhenium and Osmium in Meteorites by using Fusion, Distillation and Ion-exchange Separations', *Anal. Chim. Acta*, **222**(2), 291–300 (1989).
108. H. Mukai, Y. Ambe, M. Morita, 'Flow-injection Inductively Coupled Plasma Mass Spectrometry for the Determination of Platinum in Airborne Particulate Matter', *J. Anal. At. Spectrom.*, **5**(1), 75–80 (1990).
109. D.C. Colodner, E.A. Boyle, J.M. Edmond, 'Determination of Rhenium and Platinum in Natural Waters and Sediments, and Iridium in Sediments by Flow-injection Isotope Dilution Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **65**(10), 1419–1425 (1993).
110. G.E.M. Hall, J.C. Pelchat, 'Determination of Palladium and Platinum in Fresh Waters by Inductively Coupled Plasma Mass Spectrometry and Activated Charcoal Preconcentration', *J. Anal. At. Spectrom.*, **8**(8), 1059–1065 (1993).
111. P. Schramel, I. Wendler, S. Lustig, 'Capability of ICP MS (Pneumatic Nebulization and ETV) for Platinum Analysis in Different Matrices at Ecologically Relevant Concentrations', *Fresenius' J. Anal. Chem.*, **353**(2), 115–118 (1995).
112. S. Lusig, S. Zang, B. Michalke, P. Schramel, 'Platinum Determination in Nutrient Plants by Inductively Coupled Plasma Mass Spectrometry with Special Respect to the Hafnium Oxide Interference', *Fresenius' J. Anal. Chem.*, **357**(8), 1157–1163 (1997).
113. M.E. Farago, P. Kavanagh, R. Blanks, J. Kelly, G. Kazantis, L. Thornton, P.R. Simpson, J.M. Cook, H.T. Delves, G.E.M. Hall, 'Platinum Concentration in Urban Road Dust and Soil, and in Blood and Urine in the United Kingdom', *Analyst*, **123**(3), 451–454 (1998).
114. S. Lusig, B. Michalke, W. Beck, P. Schramel, 'Platinum Speciation with Hyphenated Techniques: High Performance Liquid Chromatography and Capillary Electrophoresis On-line Coupled to an Inductively Coupled Plasma Mass Spectrometer – Application to Aqueous Extracts from a Platinum Treated Soil', *Fresenius' J. Anal. Chem.*, **360**(1), 18–25 (1998).
115. M. Parent, H. Vanhoe, L. Moens, R. Dams, 'Determination of Low Amounts of Platinum in Environmental and Biological Materials Using Thermospray Nebulization Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **354**(5–6), 664–667 (1996).
116. H.C. Ehrsson, I.B. Wallin, A.S. Andersson, 'Cisplatin, Transplatin and Their Hydrated Complexes: Separation and Identification Using Porous Graphite Carbon and Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **67**(19), 3608–3611 (1995).
117. M. Balcerzak, 'Analytical Methods for the Determination of Platinum in Biological and Environmental Materials', *Analyst*, **122**(5), 67R–74R (1997).
118. M.M. Gómez Gómez, C.W. McLeod, 'Trace Enrichment and Determination of Gold by Flow-injection Inductively Coupled Plasma Mass Spectrometry. Part II. Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **10**(2), 89–91 (1995).
119. C.A. Williams, F.R. Abou-Shakra, N.I. Ward, 'Investigations into the Use of Inductively Coupled Plasma Mass Spectrometry for the Determination of Gold in Plant Materials', *Analyst*, **120**(2), 341–346 (1995).
120. P. Tothill, L.M. Matheson, J.F. Smyth, K. McKay, 'Inductively Coupled Plasma Mass Spectrometry for the Determination of Platinum in Animal Tissues and a Comparison with Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **5**(7), 619–622 (1990).
121. J. Begerow, M. Turfeld, L. Dunemann, 'Determination of Physiological Palladium and Platinum Levels in Urine

Using Double Focusing Magnetic Sector Field ICP MS', *Fresenius' J. Anal. Chem.*, **359**(4–5), 427–429 (1997).

122. J. Begerow, M. Turfeld, L. Dunemann, 'Determination of Physiological Palladium, Platinum, Iridium and Gold Levels in Human Blood Using Double Focusing Magnetic Sector Field ICP MS', *J. Anal. At. Spectrom.*, **12**(9), 1095–1098 (1997).
123. J. Begerow, L. Dunemann, 'Mass Spectral Interferences in the Determination of Trace Levels of Precious Metals in Human Blood Using Quadrupole Magnetic Sector Field and Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**(4), 303–306 (1996).
124. E.S. Beary, P.J. Paulsen, 'Development of High Accuracy ICP Mass Spectrometric Procedures for the Quantification of Platinum, Palladium, Rhodium and Lead in Used Autocatalysts', *Anal. Chem.*, **67**(18), 3193–3201 (1995).
125. O.V. Borisov, D.M. Coleman, R.O. Carter, 'Determination of Platinum, Palladium, Rhodium and Titanium in Automotive Catalytic Converters Using Inductively Coupled Plasma Mass Spectrometry with Liquid Nebulization', *J. Anal. At. Spectrom.*, **12**(2), 239–246 (1997).
126. O.V. Borisov, D.M. Coleman, R.O. Carter, 'Determination of Vanadium, Rhodium and Platinum in Automotive Catalytic Convertors Using Inductively Coupled Plasma Mass Spectrometry with Spark Ablation', *J. Anal. At. Spectrom.*, **12**(2), 231–237 (1997).
127. D.M. Wayne, 'Direct Determination of Trace Noble Metals (Palladium, Platinum and Rhodium) in Automobile Catalysts by Glow Discharge Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**(10), 1195–1202 (1997).
128. V.A. Shestakov, O.M. Petrukhin, 'X-Ray Fluorescence Determination of Noble Metals', *Zh. Anal. Khim.*, **42**(6), 965–983 (1987).
129. P.R. Oumo, E. Nieboer, 'Determination of Microgram Amounts of Precious Metals using X-Ray Fluorescence Spectrometry', *Analyst*, **104**(1244), 1037–1049 (1979).
130. Yu.P. Kolmogorov, A.S. Lapukhov, R.D. Melnikova, V.P. Khvostova, 'X-Ray Fluorescence Technique in Geochemical Studies of Noble Metals', *Nucl. Instrum. Methods Phys. Res., Sect. A.*, **282**(2–3), 692–694 (1989).
131. A.V. Dement'ev, I.I. Antokol'skaya, G.V. Myasoedova, N.M. Kuz'min, 'X-Ray Fluorescence Determination of Trace Noble Metals After Concentration on POLYORGS™ X1-H as Complexing Sorbent', *Zh. Anal. Khim.*, **44**(11), 2002–2006 (1989).
132. E.G. Khomutova, N.M. Sinitsyn, 'Elemental X-Ray Fluorescence Analysis of Complex Compounds of Platinum Metals in Solutions', *Zh. Anal. Khim.*, **46**(6), 1193–1198 (1991).
133. R.Q. Aucélio, V.N. Rubin, B.W. Smith, J.D. Winefordner, 'Ultratrace Determination of Platinum in Environmental and Biological Samples by Electrothermal Atomization Laser-excited Atomic Fluorescence Using a Copper Vapor Laser Pumped Dye', *J. Anal. At. Spectrom.*, **13**(1), 49–54 (1998).
134. K.L. Szluha, I. Uzonyi, J. Bascó, L. Lampé, I. Czifra, M. Péter, C. Villena, W. Szmidt, 'Determination of Platinum Concentration in Cervical Tumor Tissues Using Energy-dispersive X-Ray Fluorescence (EDXRF) Method', *Microchem. J.*, **51**(1–2), 238–244 (1995).
135. S. Perähiemi, J. Parkkinen, K. Smolander, H. Mustalahti, M. Ahlgrén, 'Determination of Gold in Ore Samples by EDXRF [Energy-Dispersive X-Ray Fluorescence Spectrometry] After Thiourea Leaching and Adsorption on Activated Charcoal', *Fresenius' J. Anal. Chem.*, **343**(3), 292–296 (1992).
136. V. Cojocaru, S. Spiridon, 'X-Ray Fluorescence Method for Gold Determination in Ore Samples', *J. Radioanal. Nucl. Chem.*, **170**(1), 259–264 (1993).
137. B. de Celis, 'X-Ray Fluorescence Analysis of Gold Ore', *Appl. Spectrosc.*, **50**(5), 572–575 (1996).
138. F.I. Lobanov, S.A. Logunova, A.M. Popov, E.N. Krasnopevtseva, 'Rapid Extraction and X-Ray Fluorescence Determination of Gold in Rock', *Zh. Anal. Khim.*, **49**(3), 295–297 (1994).
139. R.R. Lu, A.Q. Le, Y.M. Gu, G.D. Wu, J.Q. Zhu, 'Precious Metals Assay by Means of Microbeam XRF Technology', *Nucl. Instrum. Methods Phys. Res., Sect. B.*, **104**(1–4), 595–601 (1995).
140. V. Honkimaki, K. Hamalainen, S. Manninen, 'Quantitative X-Ray Fluorescence Analysis Using Fundamental Parameters: Application to Gold Jewellery', *X-Ray Spectrom.*, **25**(5), 215–220 (1996).
141. I.A. Fedotova, L.P. Tikhonova, N.V. Ustinova, 'Application of Kinetic Methods for Determination of Platinum Group Metals to Express Control of Technological Processes', *Zh. Anal. Khim.*, **48**(1), 19–27 (1993).
142. E.G. Khomutova, L.E. Romanovskaya, L.P. Zhitenko, A.P. Rysev, 'Kinetic Determination of Osmium, Ruthenium, Iridium and Rhodium in a Single Sample of Copper Alloy', *Zavod. Lab.*, **58**(10), 3–4 (1992).
143. Z.C. Gu, Z.L. Zhu, R.M. Chen, B.L. Lu, C.Q. Han, 'Simultaneous Determination of Catalysts Through a Single Catalytic Kinetic Run', *Analyst*, **118**(8), 1055–1059 (1993).
144. V.I. Shlenskaya, V.P. Khvostova, G.I. Kadyrova, 'Kinetic Methods for Determination of Osmium and Ruthenium', *Zh. Anal. Khim.*, **28**(4), 779–784 (1973).
145. E.G. Khomutova, A.P. Rysev, L.I. Fedorina, V.A. Solomonov, I.V. Ershova, 'Elimination of Iron Interferences in Kinetic Determination of Rhodium, Ruthenium and Iridium', *Zh. Anal. Khim.*, **48**(10), 1653–1657 (1993).
146. N.A. Ulakhovich, G.K. Budnikov, E.P. Medyantseva, 'Preconcentration in Voltammetric Analysis of Platinum Metal Materials', *Zh. Anal. Khim.*, **47**(9), 1546–1566 (1992).
147. P. Shearan, M.R. Smyth, 'Comparison of Voltammetric and Graphite Furnace Atomic Absorption Methods

for the Direct Determination of Inorganic Platinum in Urine', *Analyst*, **113**(4), 609–612 (1988).
148. O. Nygren, G.T. Vaughan, T.M. Florence, G.M.P. Morrison, I.W. Warner, L.S. Dale, 'Determination of Platinum in Blood by Adsorptive Voltammetry', *Anal. Chem.*, **62**(15), 1637–1640 (1990).
149. J. Messerschmidt, F. Alt, G. Tölg, J. Angerer, K.H. Schaller, 'Adsorptive Voltammetric Procedure for the Determination of Platinum Baseline Levels in Human Body Fluids', *Fresenius' J. Anal. Chem.*, **343**(4), 391–394 (1992).
150. K. Hoppstock, F. Alt, K. Cammann, G. Weber, 'Determination of Platinum in Biotic and Environmental Materials. Part II. A Sensitive Voltammetric Method', *Fresenius' Z. Anal. Chem.*, **335**(7), 813–816 (1989).
151. C. León, H. Emons, P. Ostapczuk, K. Hoppstock, 'Simultaneous Ultratrace Determination of Platinum and Rhodium by Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **356**(1), 99–104 (1997).
152. M.S. El-Shahavi, A.Z. Abu Zuhri, M.M. Kamal, 'Adsorptive Stripping Voltammetric Measurement of Trace Amounts of Platinum(II) and Ruthenium(III) in the Presence of 1-(2-Pyridylazo)-2-Naphthol', *Fresenius' J. Anal. Chem.*, **348**(11), 730–735 (1994).
153. Z. Zhao, H. Freiser, 'Differential Pulse Polarographic Determination of Trace Levels of Platinum', *Anal. Chem.*, **58**(7), 1498–1501 (1986).
154. N.A. Ezerskaya, 'Voltammetric Determination of Platinum Metals', *Zh. Anal. Khim.*, **46**(4), 675–682 (1991).
155. E.P. Medyantseva, N.A. Ulakhovich, O.N. Romanova, G.K. Budnikov, 'Alternating Current Polarography of Ruthenium, Osmium and Iridium Quinolin-8-thiolates after Solvent Extraction Separation', *Zh. Anal. Khim.*, **44**(4), 695–698 (1989).
156. O.L. Kabanova, V.I. Shirokova, 'Coulometry of the Noble Metals', *Zh. Anal. Khim.*, **50**(8), 818–826 (1995).
157. A.M. Demkin, L.V. Borisova, 'Coulometric Determination of Platinum and Rhenium in Dead Catalysts', *Zh. Anal. Khim.*, **50**(8), 875–878 (1996).
158. N.A. Ezerskaya, V.F. Sukhoverkhov, L.P. Zhitenko, I.N. Kiseleva, M.V. Afanas'eva, L.G. Podzolko, 'Coulometric Determination of Gold in Jewellery Alloys with High Silver Concentration after their Dissolution using F-containing Agents', *Zavod. Lab.*, **63**(13), 1–5 (1997).
159. D. Shouan, 'Precise Coulometric Titration of Precious Metals. II. Determination of Platinum(IV) with Electrogenerated Chlorocuprous Ion in Potassium Chloride-copper-EDTA Buffer Medium', *Talanta*, **42**(1), 49–53 (1995).
160. R.A. Nadkarni, G.H. Morrison, 'Neutron Activation Determination of Noble Metals Using a Selective Group Separation Scheme', *J. Radioanal. Chem.*, **38**(1–2), 435–449 (1977).
161. T.R. Sykes, L.G. Stephens-Newsham, M.J. Apps, A.A. Noujaim, 'Assay of Platinum and Gold in Biological Material by NAA', *J. Radioanal. Chem.*, **69**(1–2), 441–452 (1982).
162. I.M. Valente, M.J. Minski, P.J. Peterson, 'Neutron Activation Analysis of Noble Metals in Vegetation', *J. Radioanal. Chem.*, **71**(1–2), 115–127 (1982).
163. A. Cocherie, M. Volfinger, G. Meyer, 'Determination of the Noble Metals in Chromites and Other Geological Materials by Radiochemical NAA', *J. Radioanal. Nucl. Chem.*, **113**(1), 133–143 (1987).
164. M. Asif, S.J. Parry, H. Malik, 'Instrumental Neutron Activation Analysis of a Nickel Sulphide Fire Assay Button to Determine the Platinum Group Elements and Gold', *Analyst*, **117**(8), 1351–1353 (1992).
165. B. Rietz, K. Heydorn, 'Determination of Gold and Platinum in Biological Materials by Radiochemical Neutron Activation Analysis Using Electrolytic Separation of Gold', *J. Radioanal. Nucl. Chem.*, **174**(1), 49–56 (1993).
166. I. McDonald, R.J. Hart, M. Tredoux, 'Determination of the Platinum Group Elements in South African Kimberlites by Nickel Sulfide Fire-assay and Neutron Activation Analysis', *Anal. Chim. Acta*, **289**(2), 237–247 (1994).
167. C.A. Nogueira, A.M.G. Figueiredo, 'Determination of Platinum, Palladium, Iridium and Gold in Selected Geological Reference Materials by Radiochemical Neutron Activation Analysis: Comparison of Procedures Based on Aqua Regia Leaching and Sodium Peroxide Sintering', *Analyst*, **120**(5), 1441–1443 (1995).
168. G.M. Kolesov, D.Yu. Sapozhnikov, 'Neutron Activation Analysis of Noble Metals of Terestrial and Cosmic Origin Using Micro-fire Assay Concentration', *Analyst*, **120**(5), 1461–1464 (1995).

Nuclear Magnetic Resonance in Metals Analysis

Günther Neue
University of Dortmund, Dortmund, Germany

1 Introduction 1
2 History 2
3 Nuclear Magnetic Resonance Parameters in Metals 2
4 Applications 3
4.1 Pure Metals 3
4.2 Alloys 4
4.3 Liquid Metals 4
4.4 Ferromagnets and Superconductors 4
4.5 Metallic Interfaces 5
5 Method Development 5
Acknowledgments 6
Abbreviations and Acronyms 6
Related Articles 6
References 6

Nuclear magnetic resonance (NMR) is a spectroscopic technique which uses the magnetic component ($\boldsymbol{B}_1$) of radiofrequency (RF) electromagnetic fields to excite nuclear spins which are subject to a static magnetic field ($\boldsymbol{B}_0$). The resonance frequency is proportional to this field. Most stable elements possess at least one isotope with spin $I > 0$ and, hence, are observable by NMR. Magnetic field gradients make it possible to label spatial coordinates within a sample. Magnetic resonance imaging (MRI) uses this principle to reconstruct images from NMR signals. Skin effect enhanced imaging nuclear magnetic resonance (SEEING-NMR) derives its submicrometer resolution from the strong shielding properties of metals with respect to RF fields. Resonance frequencies of most isotopes are well separated from each other, making NMR an element-specific technique. Interactions of spins with their local environment lead to characteristic spectral changes that reflect the chemical state, the physical state or local geometry. The possibility of coherent manipulation and detection of spin states allows for an extreme flexibility in designing new and highly selective NMR experiments. Integrated signal intensities are strictly proportional to concentrations. The disadvantage of NMR is the small signal-to-noise ratio, especially for nuclei that have low resonance frequencies. Except for sensitive isotopes like ^{1}H or ^{19}F, concentrations should therefore be higher than 1% and samples weighing 10 to 100 mg are recommended. The sensitivity to local interactions often makes it impossible to predict the visibility of NMR signals in unknown substances.

1 INTRODUCTION

Metals and other electronic conductors are troublesome for NMR experiments because of their strong interaction with RF fields (skin effect). It is just this effect that is used in RF electronics like NMR spectrometers to shield cables and circuits from unwanted electromagnetic radiation. Nevertheless it is also possible to obtain NMR signals from metals.

Even the earliest investigations showed dramatic consequences for impedance matching, line shapes, etc.[1] As time domain NMR advanced to become a general tool in chemistry and solid-state physics, it appeared natural to study the behavior of nuclear spins in metals. Responses to common pulse sequences were discussed to form a basis for the evaluation of such experiments.[2,3]

It turned out that distributions of RF absorption and phase shifts within the samples spoiled almost any known experiment. Even for simple spectra the superposition of signals with different phases is a very inconvenient feature as no phase correction can restore true line shapes. More complicated pulse sequences that rely on π- or $\pi/2$-pulses suffer also from the distribution of RF amplitudes. Without model calculations no useful information could be extracted.

Facing all these problems it became common practice to avoid the skin effect as far as possible by using only fine metallic particles that are isolated from each other. This approach solves RF screening problems but by grinding every sample into fine powders one loses one of the most valuable properties of NMR – to obtain information without destroying the sample. Even worse, many metals show differences in their NMR signals by different mechanical treatments due to the ease of introducing defects into the respective crystal lattices.

In view of the strong influence of the skin effect on NMR measurements, one might ask whether it is possible to extract useful information by designing special experiments. A very interesting proposal used a series of a few systematically incremented single pulses to water that was both outside and inside a metallized glass tube.[4] It was shown that the thickness of the metallic films can be obtained by fitting the set of measurements with expected model functions.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 HISTORY

Towards the end of 1945 Purcell, Torrey, and Pound discovered the first NMR signal at Massachussets Institute of Technology.[5] Only a few weeks later another group at Stanford – Bloch, Hansen and Packard – independently also found a nuclear resonance signal.[6,7] The approaches of the two groups were so different that it took some time to realize that both had observed the same physical phenomenon.

NMR signals of metals were observed only 2 years later. Pound saw a Cu resonance which he identified as resulting from the copper wire of the RF coil that was used in his experiments.[5] A little bit later, Knight discovered the typical large shift of the resonance of metals when he compared the line positions of copper metal and CuCl.[8]

Soon after the discovery of metal resonances more systematic investigations were begun to learn more about the physics and chemistry of metals. Good examples are the study of diffusion in alkali metals[9,10] and of hydrogen absorption in Pd.[1] During these studies the pronounced effect of RF absorption in metals, the skin effect, complicated the experiments and the discussion of results.

The initial NMR studies were limited by the fact that only a few research groups in the world had the knowledge to build the necessary electronic equipment that was sensitive enough to detect the extremely small voltages induced by nuclear spins in metals.

3 NUCLEAR MAGNETIC RESONANCE PARAMETERS IN METALS

The most obvious macroscopic physical effect that occurs in NMR of metals is the above-mentioned screening of high-frequency electromagnetic fields. The electric component of the field generates strong currents at the surface of the metal. There are two important consequences of these currents.

The first is heat generated by the flow of current through the resistance of the metal. Rapid pulsing with complicated pulse sequences should be avoided or a big increase in sample temperature may occur. For the same reason it is not possible to do magic angle spinning experiments. In this standard procedure for solids to achieve line narrowing the sample is rotated very fast. The rotating metal ''sees'' the static magnetic field as an alternating electromagnetic field which also contains an electric field component. It is possible to melt a metal in this way. Another aspect of the heating is that the power is drawn from the transmitter. For the transmitter the conductive sample appears as an extra resistive load of the resonance circuit. This leads to a damping which not only changes the tuning and impedance matching but it also broadens and flattens the resonance curve as a function of frequency. As the same argument also holds for the receiver, the result is lower sensitivity with a poorer signal-to-noise ratio. It should be observed that this means that observed intensities depend not only on spin concentrations but also on the electric properties of the particular sample.

The second consequence of the surface currents during a pulse is the generation of electromagnetic fields that compensate the external RF field inside the sample. The result is that the RF field that is used for exciting the spins decays very rapidly towards the inner of the sample. It is present only in a thin ''skin''. Hence the name skin effect. The characteristic decay length, shown in Equation (1), is the skin depth δ

$$\delta = (\pi\mu\sigma\nu)^{-\frac{1}{2}} \tag{1}$$

where $\mu(=\mu_r\mu_0)$ is the magnetic permeability, σ is the electric conductivity and ν is the RF frequency.

Typical values are $\delta = 5\text{–}30\,\mu\text{m}$ for most metals and magnetic field strengths of modern NMR spectrometers.

Often it will be possible to approximate the sample as consisting of planar pieces where the magnetic component $\mathbf{B}_{1,0}$ of the applied RF field is parallel to the surface. This is the case, for example, for metallic foils or rods. Then the field $\mathbf{B}_1$ exhibits an exponential decay perpendicular to the surface. It is accompanied by a phase shift proportional to the distance from the surface x as shown in Equation (2)

$$\mathbf{B}_1 = \mathbf{B}_{1,0} \exp\left(\frac{-x}{\delta}\right) \exp\left(\frac{ix}{\delta}\right) \tag{2}$$

where δ is given by Equation (1). This depth-dependent phase shift makes it difficult but not impossible to interpret the line shapes observed in metals.[2,3]

A third effect of induced currents in large metallic samples is that they produce a magnetic field which interacts with the static field. The consequence is forced mechanical vibration of the sample near resonance frequency. This in turn leads to induced currents in the observation coil that mimic a signal and might obscure the real signal. Low resonance frequencies (low magnetic field) favor mechanical oscillations. Hence, if this effect turns out to be important, higher magnetic fields should be used.

Also on a microscopic scale NMR signals are influenced by conduction electrons. The Knight shift K[8] of the resonance frequency is very obvious. It is given by[11] Equation (3):

$$K = \frac{\delta\nu}{\nu_0} = \frac{8\pi}{3}\chi_p M \; |\psi_k(0)|^2_{\;F} \tag{3}$$

where $\langle|\psi_k(0)|^2\rangle_F$ corresponds to the conduction electron density at the nucleus, χ_p is the susceptibility per unit mass, and M is the atomic mass. Values for this shift can be found in tables.[12,13]

Fluctuations of conduction electrons disturb the nuclear spins and are an efficient mechanism for relaxing them after initial excitation in the NMR experiment. Typical decay times T_1 in metals at room temperature are often of the order of a few milliseconds or even shorter. A simple theory leads to the so-called Korringa relation which relates the relaxation time T_1 at temperature T to the Knight shift[14] as shown in Equation (4).

$$T_1 K^2 = \frac{\hbar}{4\pi kT}\frac{\gamma_e^2}{\gamma_n^2} \tag{4}$$

where γ_e and γ_n are gyromagnetic ratios of the electron and the nucleus, respectively. A more detailed discussion can be found in Slichter.[15]

As the ground state of conduction band electrons is sensitive to applied magnetic fields it should be noted that at very low temperatures the effective field for the nuclei becomes a complicated function of external fields. This is known as the de Haas–van Alphen effect. Typically, the susceptibility shows oscillations as a function of the applied magnetic field. As seen above this will in turn change the Knight shift accordingly. This effect may easily change the Knight shift by some 10 ppm if the field is changed by 1000 ppm.[16]

The majority of nuclei in metals have spin $I > 1$. For them the interaction of the nuclear electric quadrupole moment (eQ) with the local electric field gradient tensor ($\tilde{V}$) is the main one. It leads to multiplets and modifies the positions of the $2I$ lines. In first-order perturbation theory transition frequencies between energy levels with magnetic quantum numbers $(m-1) \leftrightarrow m$ are given by Equation (5):

$$\nu_m^{(1)} = \nu_0 - \nu_Q \left(m - \frac{1}{2} \right) \left(\frac{3\cos^2\theta - 1}{2} + \frac{1}{2}\eta \sin^2\theta \cos 2\varphi \right. \tag{5}$$

where ν_0 is the position of the unperturbed line, $\nu_Q = (3V_{ZZ}eQ)/[h2I(2I-1)], \eta = (V_{XX} - V_{YY})/(V_{ZZ})$ is the asymmetry parameter, V_{XX}, V_{YY}, V_{ZZ} are principal axes components of V, and θ and φ measure the polar angles of the field B_0 with respect to the principal axes frame (X, Y, Z).

Note that for half-integer spins ($I = \frac{3}{2}, \frac{5}{2}, \ldots$) Equation (5) predicts that the central line ($m = \frac{1}{2}$) is unperturbed. The orientational dependence often leads to extremely broad lines for the other transitions in the case of powders or polycrystalline materials. In those cases only the central line is observed which contains $1/(2I)$ of the total intensity. For strong quadrupolar interactions second-order perturbation also shifts the central line. Equation (6) shows that for $\eta = 0$ it is displaced by

$$\nu_{1/2}^{(2)} = -\frac{\nu_Q^2}{16\nu_0} \left[I(I+1) - \frac{3}{4} \right] (1 - \cos^2\theta)(9\cos^2\theta - 1) \tag{6}$$

The situation is different if there is no electric field gradient present at the position of the nucleus for symmetry reasons. The transitions degenerate and the whole intensity is visible in spectra. To distinguish between these possibilities the technique of quadrupolar nutation[17] can be used.

With small or vanishing electric field gradients the magnetic dipoles of neighboring atoms become a considerable source of line broadening and may dominate the line shape. This dipole–dipole interaction is also anisotropic and depends on the orientation of the line joining two nuclei with respect to the external magnetic field and their respective distances. Hence, line shapes due to dipole–dipole interactions can be calculated precisely if the geometry of the crystal lattice is known. Typical line widths due to this effect are 1–10 kHz.

4 APPLICATIONS

4.1 Pure Metals

The first observation of a metal resonance by Purcell et al.[5] where they saw the signal of the NMR coil itself, was soon followed by systematic studies of metals. One of these early experiments was devoted to looking at the line width in solid sodium at different temperatures.[9] Surprisingly it was found that at 77 K the line width was 2.5 gauss while at 200 K it is narrower by a factor of 10. The latter is due to fast translational lattice diffusion of sodium atoms with jump rates in the kilohertz range even 170 K below the melting point. The fast motion averages partially anisotropic interactions. A similar behavior was found by studying the isotope ^{7}Li in solid lithium.[18] Line narrowing sets in above 230 K. A more detailed study on the same system showed that this motion is present up to the melting point and is described by a simple Arrhenius behavior with an activation energy of $E_a = 55.2\,\text{kJ mol}^{-1}$.[10]

Another physical quantity studied in alkali metals is the volume dependence of the Knight shift.[19] The variation of the volume was achieved by applying pressures between 0 and 10 kbar. The effect is easily observable and amounts to $(\Delta K/K)/(\Delta V/V)$ values of +0.15, +0.1, −0.3, and −1.7 for Li, Na, Rb, and Cs respectively.

The sensitivity of the line shape and line position to changes in the local environment of the metal nucleus as well as to the electronic band structure makes it a valuable tool for analytical work. A textbook example for

this type of application is an early study of copper.[20] It was found that filings gave a signal that increased by a factor of 2.5 after annealing. Partial annealing or etching led to intensities in-between. Cold work introduces mechanically dislocations and crystal defects. This, in turn, disturbs cubic symmetry and electric quadrupole interaction dominates the spectra. Except for the central transition all other are broadened beyond detectability. For ^{63}Cu ($I = \frac{3}{2}$) the intensity contained in the central transition only is 0.4, in excellent agreement with the assumption that filing distorts the copper lattice to such an extent that every nucleus experiences a strong electric field gradient due to loss of symmetry in its immediate environment. Careful calibration of intensities allows for a quantitative determination of lattice defects.

4.2 Alloys

Intensity changes of the observable magnetization are also the basis for studies of alloys and the observation of alloying processes. Again, lowering the local symmetry is the reason for these changes. The effect is typically much bigger as compared to lattice defects due to direct electronic effects between the different metal atoms. For nearby neighbors the electric field gradient may become so large that the second-order shift of the central line (Equation 6) is sufficient to shift it beyond visibility. This creates an ''all-or-nothing'' effect and the intensity approaches zero at higher concentrations. NMR signals from copper–zinc and copper–silver alloys were the first that were interpreted in this way.[20] The study of the concentration dependence of ^{63}Cu signals for both alloys led to the conclusion that foreign atoms located in the first and second coordination shell (18 atoms) are able to wipe out the signal of a copper nucleus.

In aluminum–zinc alloys the effect on ^{27}Al signals is much smaller and resembles the behavior of mechanical defects in copper. Only a reduction of the central line is observed and with higher concentrations of foreign atoms it approaches a fraction of 9/35, the intensity of the central transition alone. Obviously, the smaller number of electrons around the aluminum nucleus is not able to create an electric field gradient that is sufficient for a big second-order effect. NMR studies on dilute aluminum alloys can be found in Stiles and Williams.[21]

Later Cu–Zn, Cu–Ag, Cu–Ga, Cu–Au, Cu–Pd, Cu–Ni, Cu–Mn, Cu–Ge, Cu–Pt, and Cu–Fe alloys were studied over the concentration range of 10–4500 ppm noncopper atoms.[22] A standard-addition technique is proposed to quantify the amount of impurities in industrial copper.

The method was also used to follow the progress of mechanical alloying of Cu and Zn in a ball milling process.[23] Owing to their slightly different Knight shifts a quantification of unalloyed copper, α-Cu(Zn) (fcc), and β-Cu(Zn) (bcc) is possible by deconvolution of the ^{63}Cu signal.

Studies of Knight shifts and longitudinal relaxation rates have been used to determine local order and long-range disorder in several amorphous metals (metallic glasses). It was found that nickel–boron alloys show no different local electronic structure in amorphous phases than in the crystalline state.[24] In more complex Pd–Cu–P systems similar studies exhibited structural differences between crystalline and amorphous forms.[25] In the latter phosphorus atoms are preferentially coordinated by Pd. Long-range disorder was studied by relaxation measurements in MoRuP glasses[26] and in NiPB.[27]

4.3 Liquid Metals

The above-mentioned study of ^{7}Li at several temperatures and magnetic field strengths includes also liquid lithium.[10] It is found that in liquid lithium a Korringa relationship (Equation 4) holds. This proves that conduction band electrons dominate relaxation of metal nuclei also in the liquid state. In contrast to the solid, molecular motion of metal atoms does not result in fluctuating fields that give rise to significant spectral densities around 10 MHz and, thus, does not contribute to spin-lattice relaxation.

A large variety of NMR and other studies of a very large number of different liquid alloys are reviewed and discussed in a more recent article.[28] While this theoretical article is not directly interesting in the context of this encyclopedia, it points to a very large database of NMR parameters that may be very helpful for future analytical work.

4.4 Ferromagnets and Superconductors

Strongly magnetic materials require different techniques to produce NMR signals. Internal fields exist in ferromagnets. In some cases these hyperfine fields even exceed the strongest fields that are used in NMR today. In pure cobalt the nucleus experiences a field of 21.34 T,[29] whereas the respective field strength in iron is 33 T.[30,31] Rather than putting these metals into NMR magnets it is much better to utilize their internal fields to produce the resonance signal. This avoids not only further complicated physics but also eliminates serious mechanical problems due to magnetic forces in the case of bigger samples. For the excitation of NMR signals in strong ferromagnets it has to be taken into account that the RF field is also strongly amplified, by a factor of 1000 in the case of cobalt.[29] It was found that the only spins observable are located at domain walls and the signal

of processing nuclear spins is greatly amplified by the hyperfine field.[32] This signal enhancement breaks down if strong external fields are applied that cause magnetic saturation.

The above examples represent pure metals with local cubic symmetry. Hence, the hyperfine field is essentially the same at all nuclei and reasonably sharp lines are observed. If the structure becomes more complicated, NMR spectra reflect the distributions of hyperfine fields. These may extend over a range of several tesla and NMR frequencies that span several tens of megahertz. In these cases detection has to be done by spin echoes and intensities determined individually point by point at several resonance frequencies, as typical NMR probeheads possess bandwidths of not more than a few hundred kilohertz, to obtain a reasonable signal-to-noise ratio. The interpretation of NMR spectra as a distribution of hyperfine fields was used to deduce the distribution of coordination numbers in amorphous CoBSi alloys.[25]

Isolated magnetic atoms in a nonmagnetic matrix can be studied by a technique called satellite NMR.[33] This method observes the resonance of matrix atoms. If distances between matrix and impurity atoms are well-defined, a series of weak lines at a distance from the strong pure matrix signal appear. The shift is a direct consequence of the magnetic field of the impurity. Though experimentally demanding, satellite lines were identified and explained in terms of coordination shells in a number of doped coppers (Sc, Ti, V, Cr, Mn, Fe, Co, Ni).[33]

Early NMR experiments on metallic superconductors were carried out in parallel with the appearance of the Bardeen–Cooper–Schrieffer theory of superconductivity.[34] The experimental results were inconclusive with respect to the validity of this theory, which assumed spin pairing. While longitudinal relaxation data of Al supported it, Knight shift measurements on Hg and Sn pointed clearly to residual electron paramagnetism even at $T = 0\,\mathrm{K}$.[35–37] Careful studies of ultrapure Al layers[38] resolved the puzzle by showing that the Knight shift is due to small paramagnetic impurities like oxygen and can be used to quantify these. According to Equation (3) the impurity's finite χ_p at $T = 0\,\mathrm{K}$ can produce a Knight shift under these conditions.

^{17}O studies of many high-temperature superconductors[39] helped to identify the role of different types of oxygen for superconductivity.

4.5 Metallic Interfaces

As RF fields penetrate a few micrometers into a metal, NMR offers the possibility to study destructionless buried metallic layers and interfaces. SEEING-NMR, which is described in more detail in the next section, uses the skin effect for spatial resolution below 1 μm and delivers separate NMR spectra as a function of depth.

In ferromagnetic layers the huge gain in intensity allows the detection of extremely thin layers and even monolayers if low temperatures (1 K) are used to enhance the population difference of nuclear energy levels. The different environment of interface atoms from those in the bulk leads typically to drastically changed hyperfine fields.

Because of its outstanding sensitivity ^{59}Co in ferromagnetic phases was the prime target to investigate local structure, symmetry, surface roughness, and strain of metallic interfaces by NMR. Application of magnetic fields in different directions may help to assign signals from anisotropic structures and often sharpen the lines through removal of large-scale magnetic inhomogeneities by magnetic saturation of the sample. In Co/Cu and Co/Ni multilayers differently coordinated interface atoms were clearly identified.[32] By modeling the spectra, quantitative information about mixing across the interface was obtained for both systems. Similar studies on Fe/V, Co/Sb, and Fe/Mn multilayers revealed several details.[32] In Fe/V mixing takes place over three to five atomic layers depending on the method of preparation.[32] At the interface of the Co/Sb system nonmagnetic Co and the compounds $CoSb_2$ or $CoSb_3$ were identified.[32] In Fe/Mn random mixing of the two phases takes place over 1.5 nm.[32] Systems like Co/Ru and Co/Cr[40] where the two metals can form solid solutions show, as expected, thick regions with interdiffusion. Even more complicated is the behavior of Co/Fe systems. These elements can form intermetallic compounds, and layer-thickness-dependent segregation into intermetallic and Co-rich phases was observed.[40]

In Co/Cu, Co/Ni, and Co/Pd multilayers strain due to crystallographic mismatch was determined by using the fact that the line shift is proportional to a volume change and, hence, strain.[32]

5 METHOD DEVELOPMENT

Responses of the nuclear magnetization in metals to common pulse sequences were discussed to form a basis for the evaluation of such experiments.[2,3]

It turned out that distributions of RF absorption and phase shifts within the samples (Equation 2) spoiled almost any known experiment. Even for simple spectra the superposition of signals with different phases is a very inconvenient feature as no phase correction can restore true line shapes. More complicated pulse sequences that rely on π- or $\pi/2$-pulses suffer also from the distribution of RF amplitudes. Without model calculations no useful information could be extracted.

Facing all these problems it became common practice to avoid the skin effect as far as possible by using small metallic particles that are electrically isolated from each other. But by grinding a sample into fine powders one loses one of the most valuable properties of NMR – the ability to obtain information without destroying the sample.

SEEING-NMR solves line shape and phase problems by looking at the response of a series of pulses with increasing pulse lengths. This is called a nutation experiment. Different nutation frequencies that belong to different RF field strengths and, hence, different depths are extracted by a Fourier transform with respect to the pulse length. As every pulse also excites the whole spectrum, the experiment is actually a two-dimensional (2-D) NMR experiment. The final 2-D spectrum represents a series of normal NMR spectra containing all of the above-discussed chemical and physical information about the metal separated with respect to depth. Spatial resolution may be as high as 100 nm. The observable depth is about twice the skin depth (Equation 1).

SEEING-NMR spectra of deuterium absorbed in Pd revealed a higher concentration of ^{2}H just below the surface.[41] Application of this technique to electrochemically produced Cu/Ag multilayers showed that the Knight shift of deposited copper is different from that in the annealed Cu that served as a base material for deposition. The amount of disorder in the copper was estimated.[42] Analysis of the observed intensity profile of the ^{63}Cu resonance perpendicular to the surface quantified the variation of layer thicknesses over the sample and elucidated the correlation of thicknesses among different layers.[43]

A logical extension of this technique is the addition of magnetic field gradients during the experiments to achieve spatial resolution in the other two directions parallel to the surface.[44] Work in this field is currently in progress.

ACKNOWLEDGMENTS

The author would like to thank S. Willamowski and K. Hesse for helping with the bibliography. Fonds der Chemischen Industrie is acknowledged for funding the development of SEEING-NMR.

ABBREVIATIONS AND ACRONYMS

MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
RF	Radiofrequency
SEEING-NMR	Skin Effect Enhanced Imaging Nuclear Magnetic Resonance
2-D	Two-dimensional

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction • Microscopy of Coatings • Nuclear Magnetic Resonance of Coating and Adhesive Systems

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance

REFERENCES

1. R.E. Norberg, 'Nuclear-magnetic Resonance of Hydrogen Absorbed into Palladium Wires', *Phys. Rev.*, **86**, 745–752 (1952).
2. M. Mehring, D. Kotzur, O. Kanert, 'Influence of the Skin Effect on the Bloch Decay in Metals', *Phys. Status Solidi B*, **53**, K25–K28 (1972).
3. D. Kotzur, O. Kanert, M. Mehring, 'Amplitude and Phase of Spin-echo Signals in Bulk Metals', *Z. Naturforsch.*, **28a**, 1607–1612 (1973).
4. S. Reich, A. Vega, 'Skin Depth Measurements in Thin Metal Films by Pulse Nuclear-magnetic-resonance Nutation Technique', *J. Appl. Phys.*, **58**, 2759–2760 (1985).
5. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
6. F. Bloch, W.W. Hansen, M. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
7. F. Bloch, 'Nuclear Induction', *Phys. Rev.*, **70**, 460–474 (1946).
8. W.D. Knight, 'Nuclear Magnetic Resonance Shift in Metals', *Phys. Rev.*, **76**, 1259–1260 (1949).
9. R.E. Norberg, C.P. Slichter, 'Nuclear Relaxation Times in Metallic ^{23}Na', *Phys. Rev.*, **83**, 1074–1075 (1951).
10. D.F. Holcomb, R.E. Norberg, 'Nuclear Spin Relaxation in Alkali Metals', *Phys. Rev.*, **98**, 1074–1091 (1955).

11. C.H. Townes, C. Herring, W.D. Knight, 'The Effect of Electronic Paramagnetism on Nuclear Magnetic Resonance Frequencies in Metals', *Phys. Rev.*, **77**, 852–853 (1950).
12. W.D. Knight, 'Electron Paramagnetism and Nuclear Magnetic Resonance in Metals', *Solid State Physics*, **2**, 93–136 (1956).
13. C.G. Carter, L.H. Bennett, D.J. Kahan (eds.), *Metallic Shifts in NMR*, Pergamon, Oxford, Vols. 1–4, 1977.
14. J. Korringa, 'Nuclear Magnetic Relaxation and Resonance Line Shift in Metals', *Physica*, **XVI**, 601–610 (1950).
15. C.P. Slichter, *Principles of Magnetic Resonance*, 3rd edition, Springer-Verlag, Berlin, 1990.
16. H. Broemer, 'Nuclear Resonance in High Magnetic Fields', *Int. J. Magn.*, 15–25 (1973).
17. A. Samoson, E. Lippmaa, 'Excitation Phenomena and Line Intensities in High-resolution NMR Powder Spectra of Half-integer Quadrupolar Nuclei', *Phys. Rev. B*, **28**, 6567–6570 (1983).
18. H.S. Gutowsky, B.R. McGarvey, 'Nuclear Magnetic Resonance in Metals. I. Broadening of Absorption Lines by Spin-lattice Interactions', *J. Chem. Phys.*, **20**, 1472–1477 (1952).
19. G.B. Benedek, T. Kushida, 'The Pressure Dependence of the Knight Shift in the Alkali Metals and Copper', *J. Phys. Chem. Solids*, **5**, 241–255 (1958).
20. N. Bloembergen, T.J. Rowland, 'On the Nuclear Magnetic Resonance in Metals and Alloys', *Acta Metall.*, **1**, 731–746 (1953).
21. J.A.R. Stiles, D.L. Williams, 'Nuclear Magnetic Resonance Studies of Electric Field Gradients in Single Crystal Dilute Alloys of Aluminum', *J. Phys. F: Metal Phys.*, **4**, 2297–2309 (1974).
22. K. Tompa, K.Z. Balla, 'NMR Method for Quantitative Analysis of Dilute Alloys', *Mikrochim. Acta*, **I**, 375–384 (1977).
23. B.Q. Li, Y.N. Wang, '^{63}Cu Nuclear Magnetic resonance Analysis of Mechanical Alloying Processes in Copper-Zinc Powders', *Philos. Mag. B*, **70**, 77–87 (1994).
24. I. Bakonyi, P. Panissod, R. Hasegawa, 'Magnetic and NMR Study of Amorphous and Crystalline Ni–B Alloys', *J. Non-Cryst. Solids*, 61–62, 1189–1193 (1984).
25. P. Panissod, 'Studies of the Electronic and Atomic Structure of Amorphous Metals by NMR', *Helv. Phys. Acta*, **58**, 60–75 (1985).
26. D.A. Guerra, J. Durand, W.L. Johnson, P. Panissod, 'NMR Study on ^{31}P in an Amorphous Superconducting $(Mo_{0.5}Ru_{0.5})_{80}P_{20}$ Alloy', *Solid State Commun.*, **31**, 487–491 (1979).
27. D.A. Guerra, P. Panissod, J. Durand, '^{31}P and ^{11}B in Amorphous NiPB Alloys Near the Paramagnetic–Ferromagnetic Transition', *Solid State Commun.*, **28**, 745–749 (1978).
28. D. Quitmann, 'Nuclear Magnetic Resonance Studies of Liquid Metals', *J. Phys. Condens. Matter*, **2**, SA245–SA249 (1990).
29. A.C. Gossard, A.M. Portis, 'Observation of Nuclear Resonance in a Ferromagnet', *Phys. Rev. Lett.*, **3**, 164–166 (1959).
30. A.C. Gossard, A.M. Portis, W.J. Sandle, 'Nuclear Magnetic Resonance in Ferromagnetic ^{57}Fe', *Phys. Chem. Solids*, **17**, 341–342 (1961).
31. C. Robert, J.M. Winter, 'Observation of the Nuclear Resonance of Iron-57 in Natural (not Enriched) Metallic Iron in the Absence of an External Field', *C. R. Acad. Sci.*, **250**, 3831–3833 (1960).
32. H.A.M. de Gronckel, W.J.M. de Jonge, 'NMR Studies in Layered Magnetic Structures', in *Magnetism and Structure in Systems of Reduced Dimension*, eds. R.F.C. Farrow et al., Plenum Press, New York, 309–321, 1993.
33. C.P. Slichter, 'Satellite NMR in Cu Doped with Transition Impurities', *J. Appl. Phys.*, **50**, 7678–7681 (1979).
34. J. Bardeen, L.N. Cooper, J.N. Schrieffer, 'Theory of Superconductivity', *Phys. Rev.*, **108**, 1175–1204 (1957).
35. L.C. Hebel, C.P. Slichter, 'Nuclear Spin Relaxation in Normal and Superconducting Aluminium', *Phys. Rev.*, **113**, 1504–1519 (1959).
36. F. Reif, 'Observation of Nuclear Magnetic Resonance in Superconducting Mercury', *Phys. Rev.*, **102**, 1417–1418 (1956).
37. G.M. Androes, W.D. Knight, 'Nuclear Magnetic Resonance in Superconducting Tin', *Phys. Rev.*, **121**, 779–787 (1961).
38. R.H. Hammond, G.M. Kelly, 'Nuclear Magnetic Resonance and Relaxation in Superconducting Aluminum', *Rev. Mod. Phys.*, **36**, 185–187 (1964).
39. E. Oldfield, C. Coretsopoulos, S. Yang, L. Reven, H.C. Lee, J. Shore, O.C. Han, E. Ramli, D. Hinks, '^{17}O Nuclear-magnetic-resonance Spectroscopic Study of High-T_c Superconductors', *Phys. Rev. B*, **40**, 6832–6849 (1989).
40. P. Panissod, J.P. Jay, C. Meny, M. Wojcik, E. Jedryka, 'NMR Analysis of Buried Metallic Interfaces', *Hyperfine Interact.*, **97/98**, 75–98 (1996).
41. U. Skibbe, G. Neue, 'A 2D-NMR Method to Study Near-surface Regions of Conductors', *Colloids Surf.*, **45**, 235–242 (1990).
42. G. Neue, U. Skibbe, C. Dybowski, 'Solid-state NMR as an Analytical Technique Offering Spatially Resolved Chemical Information', *Fresenius' J. Anal. Chem.*, **349**, 112–116 (1994).
43. U. Skibbe, 'SEEING-NMR', Thesis, Universitaet Dortmund, 1992.
44. P.C. Lauterbur, 'Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance', *Nature*, **242**, 190–191 (1973).

Thermal Evolution Methods for Carbon, Sulfur, Oxygen, Nitrogen and Hydrogen in Iron and Steel Analysis

Dennis A. Lawrenz and Joel Mitchell
LECO® Corporation, St. Joseph, USA

1 Introduction 1
2 History 1
3 Carbon and Sulfur Determination 2
3.1 Description of Methodology and Instrumentation 2
3.2 Detection Limits and Ranges 4
3.3 Analytical Procedures 5
3.4 Calibration 6
3.5 Sampling and Sample Preparation 7
3.6 Performance Data 7
4 Oxygen and Nitrogen Determination 8
4.1 Description of Methodology and Instrumentation 8
4.2 Detection Limits and Ranges 10
4.3 Analytical Procedures 10
4.4 Calibration 11
4.5 Sampling and Sample Preparation 12
4.6 Performance Data 12
5 Hydrogen Determination 12
5.1 Description of Methodology and Instrumentation 12
5.2 Detection Limits and Ranges 15
5.3 Analytical Procedures 16
5.4 Calibration 17
5.5 Sampling and Sample Preparation 17
5.6 Performance Data 17
Acknowledgments 18
Abbreviations and Acronyms 18
Related Articles 18
References 18

A thermal evolution method is a method of destructive analysis in which the sample is subjected to heating, combustion, or fusion[1] to quantify the chemical composition of iron and steel. Thermal evolution methods have been used for many years to determine carbon, sulfur, oxygen, nitrogen, and hydrogen contents of iron and steel. Carbon and sulfur are determined using combustion techniques. Oxygen and nitrogen are determined using inert gas fusion techniques. Hydrogen is determined using either inert gas fusion or hot extraction techniques. Analytical instrumentation for all of these techniques is commercially available.

1 INTRODUCTION

Thermal evolution methods have been used for many years to determine the carbon, sulfur, oxygen, nitrogen, and hydrogen contents of iron and steel. Thermal evolution refers to a method for measuring the chemical composition of iron and steel whereby a sample is subjected to a process of heating, combustion, or fusion.[1] These methods are considered to be destructive analyses. Combustion techniques are employed for the determination of carbon and sulfur, while inert gas fusion techniques are used for oxygen, nitrogen, and hydrogen. Hydrogen may also be determined via hot extraction. Analytical instrumentation based on thermal methods is commercially available.

2 HISTORY

For most of the twentieth century, thermal evolution methods have been used to determine carbon, sulfur, oxygen, nitrogen, and hydrogen in iron and steel products. Beginning early in the century, carbon was determined through the use of gravimetric methods. Samples were placed in an oxygen-purged resistance furnace. The combustion of carbon released from the sample formed carbon dioxide, which could be absorbed onto soda-asbestos. The carbon content of the sample was then determined by the mass gain of the soda-asbestos.[2] This procedure is commonly referred to as the gravimetric carbon method. Likewise, sulfur was determined by combusting a sample in oxygen using a resistance-heated furnace followed by absorption of the oxides of sulfur. The sulfur oxides then reacted with potassium iodate and potassium iodide to produce a colored solution. This solution was subsequently measured by either a color comparison or a titration technique.[3] During the 1960s, high-frequency (HF) induction furnaces began to replace resistance-heated furnaces for both carbon and sulfur determinations. HF induction furnaces were the furnaces of choice in the 1970s for the combustion of steel and iron samples. During that period, automated instrumentation based on HF furnaces and infrared (IR) detection systems that permitted simultaneous carbon and sulfur determination (i.e. from the same sample) developed rapidly. By the 1990s, the vast majority of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

carbon and sulfur determinators utilized HF induction furnaces coupled to IR detection systems.

Thermal evolution methods for oxygen, nitrogen, and hydrogen became prevalent in the 1930s. The vacuum fusion method was typically employed. This method uses a graphite crucible heated in a HF induction furnace to melt the sample under a vacuum. Furnace temperatures of about 1600 °C were reported. The evolved gases were measured either volumetrically or gravimetrically. Oxygen could be determined using hydrogen reduction methods as well. Hydrogen could also be determined by hot extraction heating a sample to around 600 °C in a vacuum.[4] Around 1970, inert gas fusion instruments were designed to provide both improved sensitivity and reduced analysis time. The inert gas fusion technique of today features a direct current electrode (impulse) furnace that heats a graphite crucible resistively to temperatures approaching 3000 °C in an inert carrier gas such as helium or argon. The sample is fused in the graphite crucible and the carrier gas sweeps the evolved carbon monoxide (CO), nitrogen (N_2), and hydrogen (H_2) to the detectors. If the carrier stream is split, oxygen and nitrogen can be determined simultaneously by thermal conductivity (TC) detectors. Oxygen can also be determined by IR detection as either CO or CO_2. Hydrogen is detected in a separate determinator using a TC detector. Since TC detectors are not selective, any gas other than the analyte or the carrier must be separated or removed to allow for accurate determination. Since 1980, simultaneous oxygen and nitrogen determinators have been using IR detection for oxygen and TC detection for nitrogen. Oxygen from the sample combines with carbon from the crucible to form CO. For additional sensitivity, the CO is usually converted to CO_2 via catalytic oxidation. CO or CO_2 is subsequently detected by IR. Modern hydrogen determinators employ resistance, HF induction, or electrode furnaces coupled to TC detectors for the determination of hydrogen by inert gas fusion or hot extraction.[5]

3 CARBON AND SULFUR DETERMINATION

3.1 Description of Methodology and Instrumentation

The carbon and sulfur contents of iron and steel products are routinely determined simultaneously using computer-controlled instrumentation based on the high-temperature combustion–IR detection method. Resistance-heated furnace units with temperature ramping features have been successfully used in determining surface carbon on steel products. However, resistance furnaces are of limited use for bulk (total) carbon and sulfur determinations in the myriad of steel and iron samples which require rapid and accurate determinations. Because of this shortcoming, the resistance-heated furnace has largely been replaced by the HF induction furnace. HF furnaces easily combust a variety of sample shapes and forms such as pins, slugs, chunks, chips, drillings, or powders. Figure 1 illustrates the combustion zone of a typical HF induction furnace. A HF induction furnace heats the sample by subjecting it to oscillating electric and magnetic fields. These fields heat electrically conductive samples by inducing electron motion.The combustion gas is carried out in a flow of pure oxygen to solid state IR detectors for accurate determination of the carbon as carbon dioxide (CO_2) and sulfur content as sulfur dioxide (SO_2).

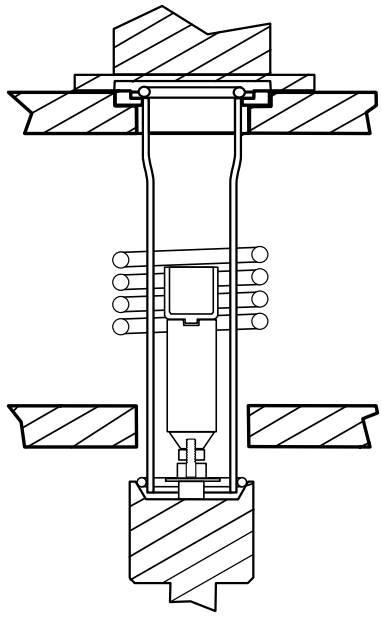

Figure 1 Combustion zone of a typical HF induction furnace.

A TC detector can also be used to measure carbon and sulfur. However, TC detectors are universal detectors (i.e. they respond to any gas that has a different TC from the carrier or reference gas), that rely on the separation or removal of gases other than the carrier gas and the gas being measured. Thus, analysis time and the potential for measurement interference are increased.[6] For this reason, the majority of carbon and sulfur combustion instrumentation designed for the analysis of iron and steel does not utilize TC detector technology.

In contrast, IR detectors are designed to measure only the specific molecule of interest, making them the preferred detectors for this instrumentation.[7] A typical IR detection cell consists of a cell body with a measure chamber, an IR source, a chopper motor with blades, a narrow bandpass (wavelength) filter, a

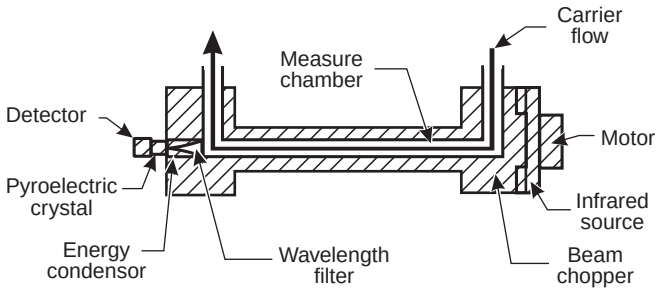

Figure 2 Typical IR detection system.

condensing cone, a solid state IR energy detector, and a preamplifier (Figure 2). The IR source, located at one end of the measure chamber, radiates visible energy as well as wavelengths in the IR spectrum. This energy is transmitted through the windows of a sealed measure chamber to the opposite end of the cell body. A chopper motor with blades rotates through the light beam, which presents the IR energy to the detector in pulses. CO_2 and SO_2 absorb IR energy at specific wavelengths within the IR spectrum, 4.27 μm and 7.37 μm, respectively. The pulsed IR energy passes through a wavelength filter selected for either CO_2 or SO_2; therefore, only the wavelength of interest passes on to the condenser cone that concentrates the energy at the detector. The detector is typically a pyroelectric device that responds to chopped or pulsed light[8] and is AC coupled to a preamplifier.

At the beginning of an analysis, a starting reference level or ''baseline'' is established with only the oxygen (IR inactive) carrier gas flowing through the cells. During the combustion of the sample, the concentration of CO_2 or SO_2 increases in the gas stream and upon reaching the respective IR cell, IR energy is absorbed. This absorption of IR energy results in a decrease in voltage (signal) transmitted by the detector to the preamplifier, which is proportional to the CO_2 or SO_2 concentration in the gas stream. The Lambert–Beer law describes this process. Once the analyte has passed through the cell, the signal returns to its starting level. The integration of the CO_2 or SO_2 signal is directly correlated to the carbon or sulfur concentration in the sample. Figure 3 depicts the signal obtained from the analysis of a low alloy steel. Note that the signal is inverted for simpler interpretation. The length of the measure chamber (distance between the IR source and detector) has a direct effect on the sensitivity and range of the IR cell. The amount of IR radiation absorbed by the analyte is directly proportional to IR cell pathlength. Longer IR cell pathlengths are more sensitive and provide excellent low-level detection but limit detection at higher concentrations (detector saturates). The opposite is true for shorter pathlength IR cells. The typical measure chamber of a CO_2 IR cell is approximately 2–15 cm in length. SO_2 IR cell measure chambers are approximately 33 cm long, since the molar absorptivity of SO_2 is only a fraction of the molar absorptivity of CO_2. The flow rate of the gas through the IR cell is closely monitored and controlled. Separate carbon-only and sulfur-only combustion instruments, based on the same concept as a simultaneous carbon and sulfur instrument, are also manufactured for laboratory use.

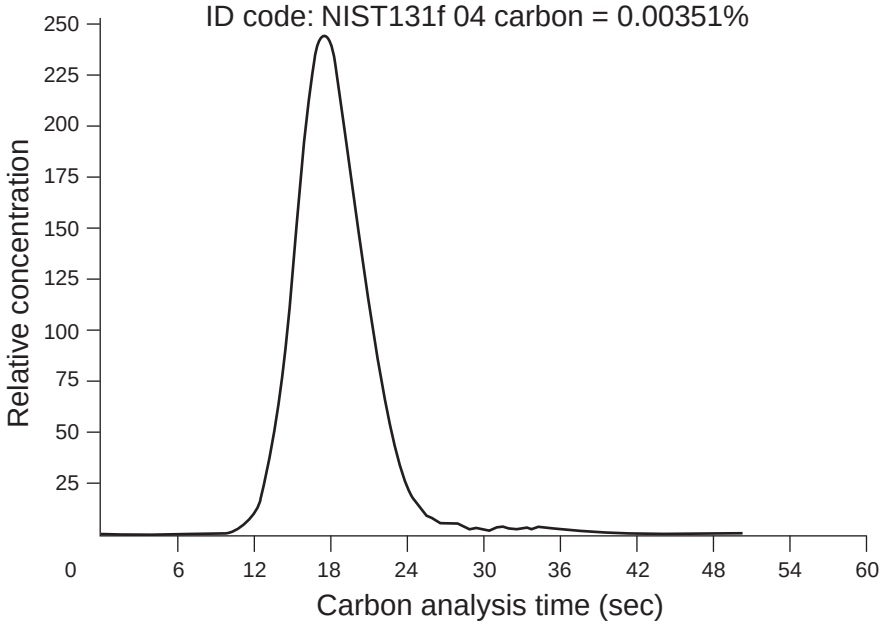

Figure 3 Carbon signal obtained from low-alloy steel.

3.1.1 Analysis Sequence

Figure 4 illustrates a typical flow diagram for a high-temperature combustion carbon and sulfur instrument. A preweighed sample of approximately 1 g is placed in a single-use high-purity ceramic crucible, and approximately 1 g of an appropriate metal accelerator is added. The crucible is placed on a ceramic pedestal. The pedestal is raised, the crucible is positioned in the center of the HF furnace work coil and sealed from the atmosphere. Purified oxygen is used to purge the combustion area. Power supplied to the induction coil establishes an alternating electric and magnetic field around the sample and accelerator. This combination rapidly heats the sample and accelerator causing them to combust. The carbon in the sample oxidizes primarily to CO_2, with some CO possibly produced depending on the sample type and carbon level. The sulfur in the sample is oxidized to SO_2. These gases are swept along with the oxygen carrier gas through a dust filter and drying agent into a solid state IR detector where sulfur is measured as SO_2. The gases are then passed through a platinized silica catalyst heated to 350 °C which converts any CO to CO_2 and most of the SO_2 to sulfur trioxide (SO_3). The gases are then routed through a cellulose filter where the SO_3 is removed, while the CO_2 passes through the filter and is measured by a separate IR detector. The results are displayed as weight percent carbon and sulfur. Analysis times of one minute or less are typical.

3.2 Detection Limits and Ranges

State-of-the-art high-temperature combustion carbon and sulfur instrumentation is capable of determining total carbon and sulfur in iron and steel products from parts per million (ppm) to weight percent levels. A typical instrument will be able to analyze 1 g samples with carbon contents from 0.0001–3.5% and sulfur contents from 0.0001–0.4%. The ranges can be expanded upward by reducing the sample weight (i.e. the absolute amount of carbon or sulfur is less, but the percentage remains the same). In addition, a variety of instruments are optimized for either low or high levels of carbon and sulfur depending on the user's needs. The detection limit (DL) is typically calculated from standard deviations of a number of consecutive measurements obtained on an individual instrument at or near the lowest measurable concentration of the element being determined. Instrument manufacturers may combine standard deviations from several individual instruments to determine DLs. One

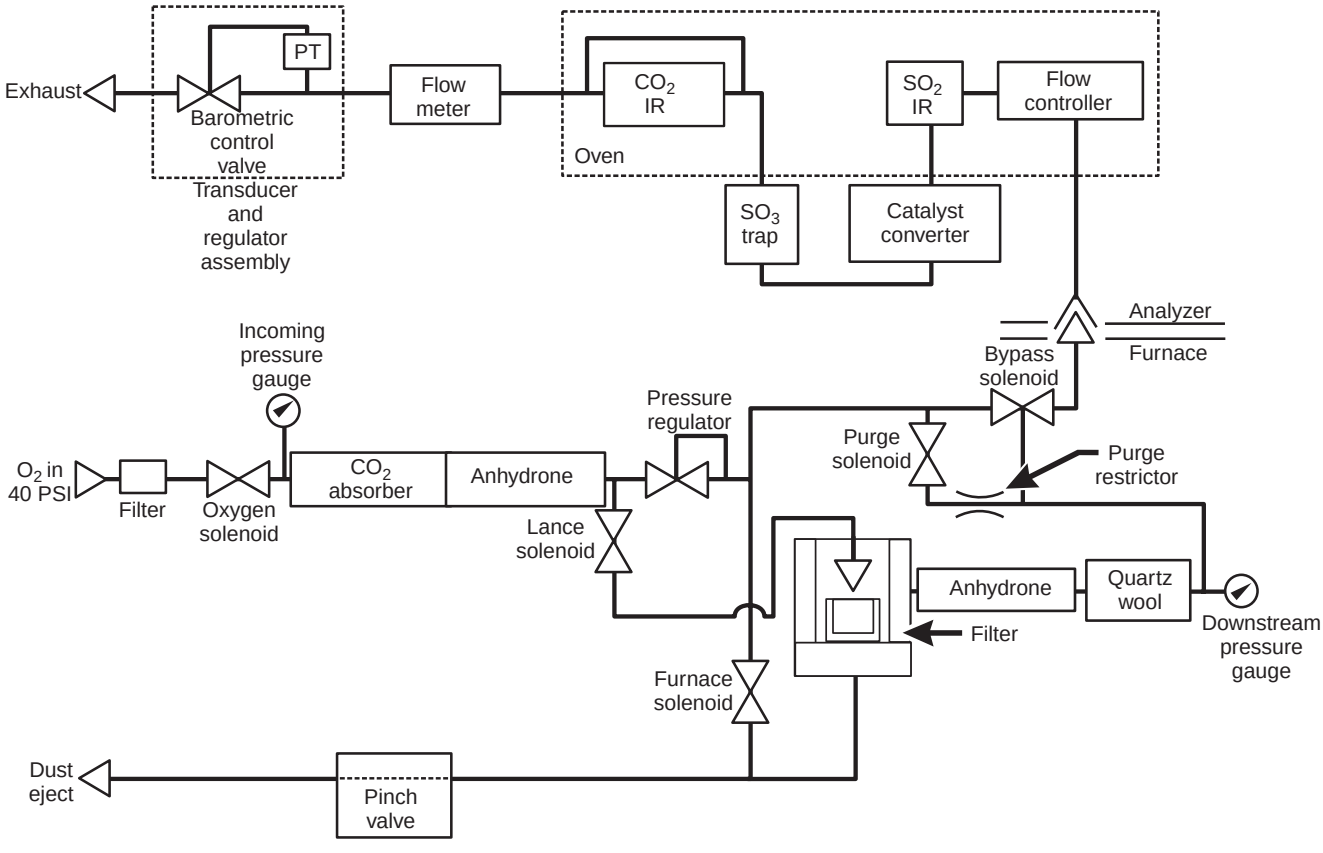

Figure 4 Flow diagram for combustion carbon and sulfur determinator.

way of accomplishing this is to dose known quantities of CO_2 and SO_2 gases directly into the measure stream of an instrument. Some manufacturers offer carbon and sulfur instruments with built-in gas dosing capabilities. This feature allows for evaluation of the instrument's ability to measure carbon and sulfur apart from the influence of variables such as sample homogeneity, combustion characteristics, and the contributions of the crucibles and accelerators. Mixtures of CO_2 and SO_2 in a carrier gas such as nitrogen (stable and IR transparent), dosed at various volumes, can be used to verify an instrument's performance, linearity, and measurement range.

3.3 Analytical Procedures

In this section, analytical parameters that have an important impact on the precision of carbon and sulfur determinations are presented.

3.3.1 Crucibles

The ceramic crucibles used to contain the sample and accelerator during the combustion cycle are a very important, yet often ignored, portion of the analytical technique. These crucibles, which are approximately 2.5 cm tall and 2.5 cm in diameter, are somewhat porous and can be contaminated during handling or storage. The crucibles typically weigh between 18 and 20 g each, while the nominal sample weight used is around 1 g. Therefore, just 1 ppm of carbon or sulfur impurity in or on the crucible will contribute 18–20 µg of carbon and or sulfur to the analysis. This contamination translates to an 18–20 ppm contribution to the final result for a 1 g sample. The most problematic characteristic of crucible contamination is consistency; some crucibles may have more, some less, and some none at all. This inconsistency has a direct impact on analytical performance and is especially noticeable in ultra-low-carbon and sulfur grades of steel. In order to eliminate this variability, the crucibles should be preheated to volatilize the contaminants. Decontamination is typically accomplished by placing the crucibles in a tube or muffle furnace at 1350 °C for at least 15 min, or at 1000 °C for not less than 40 min. The crucibles are removed from the furnace, allowed to cool for approximately 2 min, and then placed in a desiccator before use.[9] Once the crucibles have been removed from the furnace, they should only be handled with a clean pair of tongs. Table 1 contains a comparison of contaminated (unbaked) versus purified (baked) crucibles. It illustrates the impact that preheating the crucibles can have on the analytical result. The mean carbon result is decreased from 14.7 to 5.0 ppm by removing the crucible contaminants, while the sulfur results were reduced from 4.6 to 2.0 ppm. The baked crucible results demonstrate a three-fold improvement in low-level quantitation capabilities.

Table 1 Impact of preheating crucibles on carbon/sulfur results

	Baked		Not baked	
	C (ppm)	S (ppm)	C (ppm)	S (ppm)
High purity iron chip standard @ 5 ppm C 2 ppm S	4.6	2.5	14.9	4.0
	4.6	2.0	13.5	4.9
	4.8	1.8	12.7	4.1
	4.9	1.5	13.8	4.2
	5.8	2.7	14.1	5.0
	4.8	2.5	13.9	4.9
	4.5	2.2	12.8	5.3
	3.6	1.5	20.3	3.0
	6.0	2.0	13.2	5.1
	6.0	1.6	17.6	5.5
$\bar{x}$ (ppm)	5.0	2.0	14.7	4.6
s (ppm)	0.75	0.45	2.42	0.76

$\bar{x}$ = Mean, s = estimate of standard deviation.

3.3.2 Oxygen Supply

The purity of the oxygen that is used as both combustion and carrier gas can have an adverse effect on the analysis. Regular grades of oxygen may contain several parts per million of hydrocarbons such as methane (CH_4). A significant portion of these hydrocarbon gases will convert to CO_2 and H_2O during the combustion of a sample and bias the carbon results. Higher purity grades of oxygen can be used, although the cost may be significantly higher. Another option is to pass the incoming oxygen over copper oxide (CuO) heated to 600 °C to convert the hydrocarbon gases to CO_2 and H_2O. Then chemical reaction or adsorption with the appropriate reagents removes the CO_2 and H_2O. Some commercially available instruments have built-in heated catalyst furnaces designed to remove hydrocarbons from the incoming oxygen supply, but high-purity gases should be used to obtain peak performance from the instrument whenever possible.

3.3.3 Accelerators

Steel and iron will combust in oxygen if elevated to a sufficient temperature. In order to assure rapid and complete combustion using a HF induction furnace, an accelerator is typically employed. An accelerator is an igniter and a flux, which promotes both combustion and a fluid melt by effectively lowering the melting point of the sample. Several high-purity metals have been used as accelerators in HF induction furnaces over the years; however, only a few are routinely used. Like the crucible, any analyte contained in the accelerator will contribute to

the result. To compensate properly for this contribution, the analyte should be present in the accelerator at only the lowest of levels and be uniformly distributed throughout the material. Metals such as copper, iron, tungsten, and tin have all been used successfully either by themselves or in combination with other accelerators. These accelerators are in chip or granular form, normally in the −10 to +40 mesh size.

Copper accelerator is typically used when carbon-only determination is required, as it may cause sulfur recovery problems in iron-based alloys. Iron is normally used with tungsten and/or tin when nonferrous samples are analyzed. Tungsten only is used when plain carbon steels are analyzed, but caution must be used to assure that complete recovery is obtained when analyzing high-alloy steels. Tin is used in conjunction with tungsten and is the accelerator of choice for simultaneous carbon and sulfur determination in steel and iron. This combination will result in the best overall combustion and recovery for the widest variety of steel and iron alloys. Approximately 1–1.5 g of accelerator is used with a 1 g sample of steel.

In order to obtain the most accurate results, the carbon and sulfur content of the accelerator is compensated for by the blank subtraction feature of the instrument. Typically, the accelerator is placed in a preheated crucible and a 1.000 g weight is entered into the weight stack of the instrument. The accelerator is burned and the carbon and sulfur contents are recorded. This process is repeated a minimum of three times and the average result for carbon and sulfur is entered as a blank. Another method to determine the blank starts by calibrating the instrument with a steel standard at a concentration where the contribution of the blank is insignificant. Then 1.000 g of a low-carbon and sulfur steel standard along with the accelerator is analyzed a minimum of three times. The difference between the average total carbon and sulfur measured and the certified carbon and sulfur content of the standard is considered the blank. The advantage of this procedure is that it simulates the combustion of the unknown samples more closely. It achieves this by producing similar temperatures and penetration of the sample into the crucible rather than simply combusting the accelerator itself. Thus, a more representative blank is recorded. The disadvantage of this technique is that the variability of the standard is transferred to the precision that can be obtained on the blank.

3.3.4 Methodology

The methodology can also adversely affect the precision of carbon and sulfur determinations. The performance of a method can be determined by conducting an interlaboratory study (ILS). Method development organizations such as the American Society for Testing and Materials (ASTM) and the International Organization for Standardization (ISO) produce written procedures and use an ILS to determine the suitability and scope of a method. The ILS will establish longer-term sources of variability within and between laboratories and instruments.[10] For example, ASTM method E 1019 ''Determination of Carbon, Sulfur, Nitrogen and Oxygen in Steel and in Iron, Nickel, and Cobalt Alloys''[11] is a document that includes carbon and sulfur data obtained using an ILS to determine the scope of the combustion method.

3.4 Calibration

The accuracy of the combustion carbon and sulfur determination is dependent on proper calibration. Calibration is accomplished by analyzing reference materials with known carbon and sulfur contents. A major advantage of the combustion method for carbon and sulfur determination in steel and iron is that there are typically no matrix interference problems. Iron, cast iron, plain carbon steels, low-alloy steels, high-alloy steels, stainless steels, tool steels, specialty steels, and high-temperature alloys can all be analyzed using the same technique. Commercially available instrumentation is normally factory linearized (linear detector output from DL to saturation), which allows for single point calibrations. Typically, three to five replicates of a reference material are analyzed and the calibration slope is adjusted via the instrument's software program. The reference materials used to calibrate are normally ''consensus standards'', which means that the known or certified values are calculated from data obtained from various laboratories. Therefore, the accuracy of the standard is dependent on the accuracy of the method or methods used to determine these values. If several reference materials are analyzed on an individual instrument and the carbon and/or sulfur results do not agree with the accepted concentrations, either the instrument linearity or the accepted value of the reference is in question. Analyzing fractional weights of pure compounds such as calcium carbonate ($CaCO_3$) and barium sulfate ($BaSO_4$) can verify instrument linearity or, as mentioned in section 3.2, gas dosing may be used. A more practical approach incorporates two steel reference materials, one at the upper level of interest and one at the lower level of interest. For example, to verify the linearity of the carbon response of the instrument, steel reference samples with carbon contents of approximately 1.0% and 0.1% can be used. Each reference sample is analyzed several times using 1 g sample weights, with the proper accelerator, to determine an ''expected value''. Then approximately 0.9 g of the 1.0% sample is mixed with approximately 0.1 g of the 0.1% sample. A new expected value is determined by multiplying the actual weight of

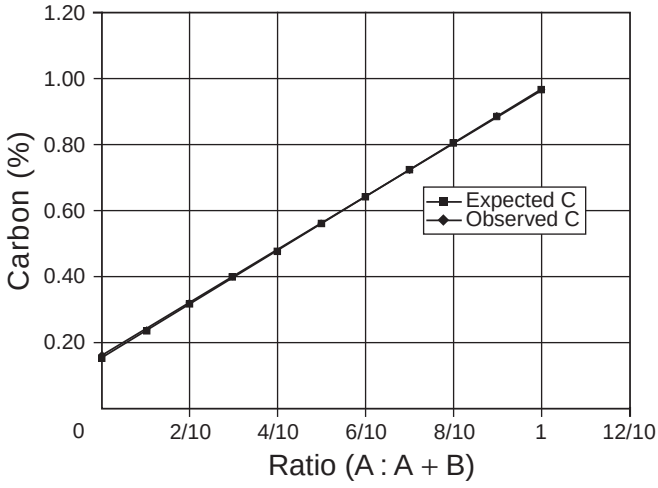

Figure 5 Linearity of carbon by induction furnace.

each sample by its expected value divided by the combined weight of both. The mixed sample is analyzed and the obtained value is compared with the expected value. The process is repeated with ca. 0.8 g of the high sample added to ca. 0.2 g of the lower sample and so on, until ca. 0.9 g of the low sample and ca. 0.1 g of the high sample are analyzed. The expected versus the obtained results are plotted to determine if the relationship is linear (Figure 5).

A linear relationship confirms the linearity of the instrument. Any reference sample employed in such a test must be homogeneous. The standard deviation of replicate analyses can be used to confirm homogeneity. Statistically, a minimum of five replicates of each level should be analyzed. This technique is also applicable to sulfur determination and it is a competent way to determine the linearity of a combustion carbon and sulfur instrument.

3.5 Sampling and Sample Preparation

Obtaining a suitable sample for analysis is very important. The sampling method must provide a sample that is representative of the material analyzed. Samples of liquid metal, cast products, wrought products, and other finished products are normally required. Samples may be in the form of solid slugs or pins, strips, chips, drillings, or powders. It is important that the sample to be analyzed is free of surface contamination. Solid samples, such as pins or slugs, can be abraded with a file, turned on a lathe, chemically etched, and/or washed in a suitable solvent such as acetone, and dried with warm air to remove surface contamination. Chips, drillings, and powders can be washed in acetone and dried with warm air. Care must be taken to assure that all traces of the solvent are removed. Surface contamination is an important factor for accurate analysis of low-carbon and sulfur grades of iron and steel. In some cases, surface carbon due to residual oils or lubricants can be removed by preheating the sample for 5–10 min at 420 °C in air. This procedure is typically applied to ultra-low carbon determination in high-purity iron and low-carbon steels.[12] ASTM E 1806, "Sampling Steel and Iron for Determination of Chemical Composition",[13] and ISO 14284, "Steel and Iron – Sampling and Preparation of Samples for the Determination of Chemical Composition",[14] are two comprehensive reference documents that provide detailed sampling procedures for steel and iron.

3.6 Performance Data

Table 2 contains typical carbon and sulfur results obtained on a modern combustion instrument with IR detection. Carbon and sulfur determinators are typically employed in a quality control capacity by the iron and steel industries. In the USA, the Ford Motor Company introduced several quality control procedures and indices, which are widely accepted by both producers and manufacturers.[15] These indices can be used to determine if the long-term performances of the iron and steel production processes are within their targeted requirements. One such index is the process capability (C_p) index, designed to compare

Table 2 Typical carbon and sulfur data obtained on a modern combustion instrument

Material	Carbon (%)				Sulfur (%)			
	Expected	$\bar{x}$	s	RSD	Expected	$\bar{x}$	s	RSD
White cast iron	3.33	3.33	0.0040	0.11	0.015	0.0150	0.00060	4.07
Gray cast iron	2.83	2.83	0.014	0.48	0.043	0.0430	0.0010	2.39
Plain carbon steel	0.97	0.970	0.0025	0.26	*	0.0234	0.00030	1.28
Low-alloy steel	0.160	0.158	0.00040	0.25	0.036	0.0360	0.00060	1.67
Silicon steel	0.028	0.0280	0.00022	0.78	0.008	0.0080	0.00018	2.30
Tool steel	0.719	0.722	0.0010	0.13	*	0.0050	0.00020	4.00
Stainless steel	0.310	0.308	0.0010	0.32	0.036	0.0354	0.00030	0.85
Low-C and S steel	0.0009	0.00089	0.00005	5.55	0.0006	0.00061	0.00004	6.67

* Indicates that sulfur concentration was not certified.
$\bar{x}$ = Mean, s = estimate of standard deviation, RSD = relative standard deviation.

the reproducibility of the measuring process to the process specification limits. In effect, the process divides the acceptable specification range for the product by the process range (typically ±3 standard deviations of the measurements). C_p is, assuming a normal distribution, the proportion of the process range within the specification range. For example, Table 2 shows the average results of a white iron standard as 3.33% carbon, with a standard deviation of ±0.004%. If the lower specification limit and upper specification limit for the sample are 3.20% and 3.45%, respectively, calculating the C_p index yields the following:

$$C_p = \frac{\text{upper spec. limit} - \text{lower spec. limit (specification range)}}{6\sigma\ \text{(process range)}}$$

$$C_p = \frac{3.45 - 3.20}{6(0.004)}$$

$$C_p = 10.42$$

where σ represents the standard deviation of a population.

Nominal C_p values in the US manufacturing industries during the 1980s averaged approximately 0.67. There has been an upward trend in the minimum acceptable C_p index.[14] Many companies are establishing targets of 2.0 or higher long term. When the C_p is applied to analytical instrumentation, sources of variation must be considered. These include quality of calibrants, homogeneity of material being analyzed, and the inherent reproducibility of the analytical method, including the detector. In order to measure the C_p of a carbon or sulfur detector alone, gas dosing may be employed. To measure the C_p of the method, well-characterized homogeneous reference samples, or pure compounds, may be employed. The C_p on production samples is defined by the deviations from the instrument, the method, and the sample. Descriptions of additional indices can be found in Bhote[16] and Chambers et al.[17] One procedure employed to characterize variation is gage repeatability and reproducibility analysis. This type of analysis can determine the proportion of measurement variability attributable to sample-to-sample variation, the instrument (reproducibility) and errors in the determination of the same sample by the same operator (repeatability). Further information concerning the use of gage reproducibility and repeatability (R&R) analysis is widely available.[18–22]

4 OXYGEN AND NITROGEN DETERMINATION

4.1 Description of Methodology and Instrumentation

The oxygen and nitrogen contents of iron and steel products are routinely determined from the same sample using the inert gas fusion method. A typical inert gas fusion oxygen and nitrogen determinator is computer controlled and has two main components: an electrode furnace and a measurement unit. The electrode furnace is equipped with water-cooled alloy-tipped copper electrodes and is capable of heating a variety of high-purity graphite crucibles to temperatures up to 3000 °C in a flowing helium stream. The measurement unit contains the detectors along with the supporting hardware and software to perform the actual measurement. Figure 6 is

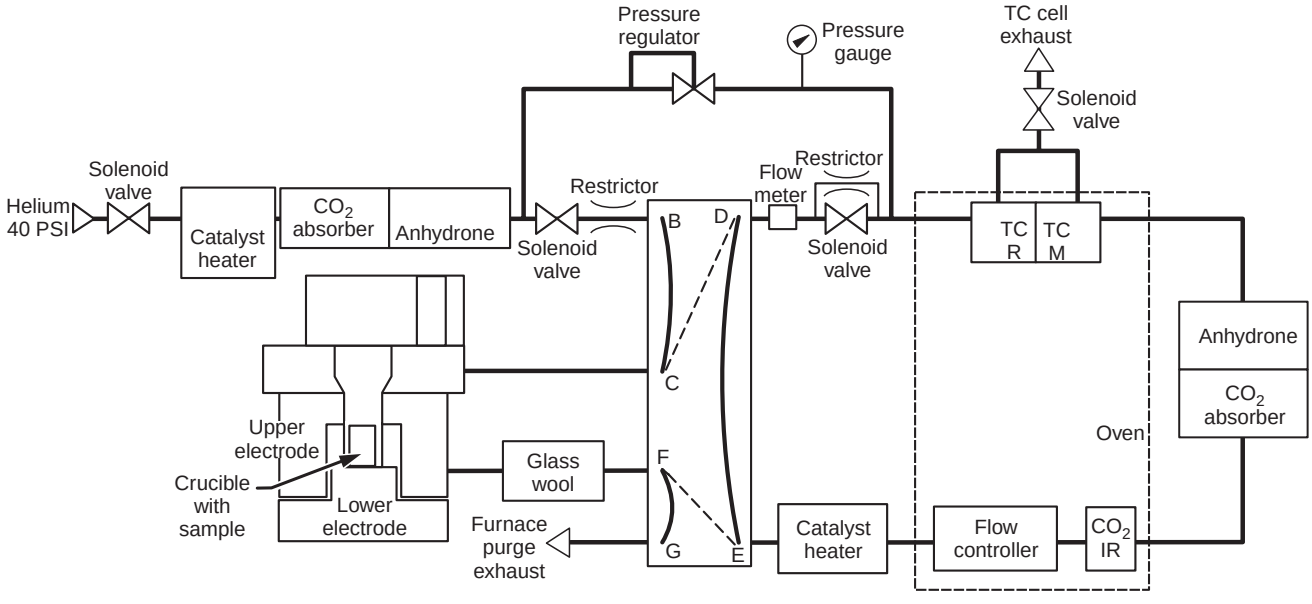

Figure 6 Flow diagram for inert gas fusion oxygen/nitrogen determinator.

an illustration of a simplified flow diagram for a modern inert gas fusion oxygen and nitrogen determinator.

Oxygen is determined as either CO or CO_2 using IR detection, while nitrogen is determined using TC detection. CO and CO_2 absorb IR energy at specific wavelengths within the IR spectrum, 4.68 μm and 4.27 μm, respectively. The same type of IR cells are used in the inert gas fusion instruments as the combustion carbon and sulfur determinators. A description of a typical IR cell is found in section 3.1 of this article.

Nitrogen is determined as N_2 using a TC detector. A typical TC detection cell consists of one or two pairs of tungsten–rhenium wire filaments carefully matched in length. These filaments are mounted in a metal block maintained at a controlled temperature, usually in an oven. The filaments are connected in a Wheatstone bridge circuit that heats the filaments above the ambient temperature of the oven. One filament, or pair of filaments, is used as the reference side of the TC detector (carrier gas only). The other filament (or pair) is used as the measure side of the TC detector. The reference filament(s) are maintained at a constant gas concentration, temperature, and flow. The measure filament(s) are maintained at a constant temperature and flow, but the gas composition is allowed to vary. Helium is used as the carrier gas for a typical oxygen and nitrogen inert gas fusion determinator since its TC is significantly different from that of nitrogen. At the start of an analysis, helium is flowing through both the reference and measure sides of the TC detector resulting in a balanced output between the filaments. During the analysis of a sample, nitrogen as N_2 enters the measure side of the TC detector along with the helium carrier gas. The temperature of the measure filament(s) will increase because N_2 has a lower TC than helium. This temperature increase changes the resistance of the measure filament(s) causing an imbalance in the bridge circuit. This in turn is used to determine the N_2 content of the sample. Figure 7 is an illustration of a TC detection cell.[23] Oxygen-only and nitrogen-only inert gas fusion instruments, based on the same concept as a simultaneous oxygen and nitrogen instrument, are also manufactured. If an oxygen-only inert gas fusion system uses an IR detector, then helium, argon, or nitrogen can be used as a carrier gas as these gases do not absorb IR radiation.

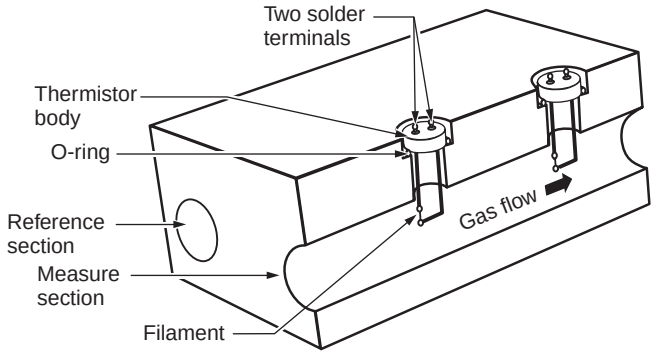

Figure 7 Typical TC detection cell.

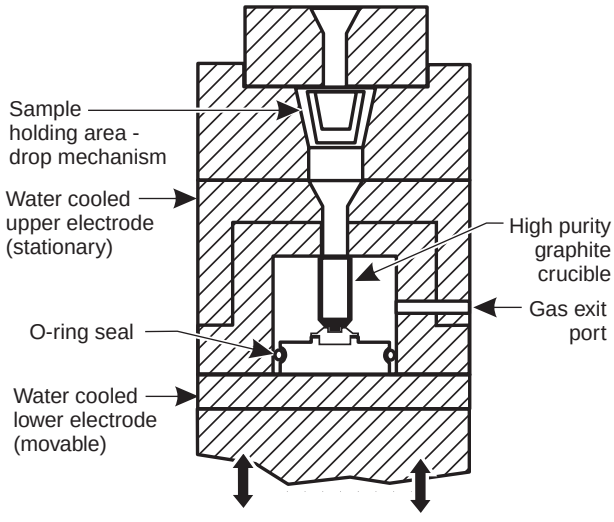

Figure 8 Cut away of electrode furnace crucible area.

4.1.1 Analysis Sequence

A preweighed sample of approximately 1 g is placed in the electrode furnace sample loading port. The sample is retained in a chamber directly above the graphite crucible area. A high purity graphite crucible is placed on the lower electrode of the furnace and the furnace is closed. Figure 8 is a cross-sectional illustration of a typical electrode furnace. After a short purge cycle (to eliminate atmospheric gases), an electric current is passed through the crucible, which heats it to temperatures up to 3000 °C. Helium is flowing through the furnace, ''sweeping out'' contaminants that are released from the crucible. This heat cycle is typically referred to as the outgas cycle. After the outgas cycle, the furnace current is lowered which reduces the crucible temperature. An electronic baseline is established for each of the detectors, after which the sample is automatically dropped into the hot crucible. The sample is fused in the graphite crucible. The oxygen in the sample reacts with carbon from the crucible forming CO, while the nitrogen is released as N_2 and any hydrogen is released as H_2. These gases are swept along with the helium through particle filters and into a heated (400–600 °C) CuO catalyst furnace to convert any CO to CO_2 and H_2 to H_2O while the N_2 passes through unchanged. The gases are then routed to a CO_2 IR cell where the CO_2 is measured. H_2O absorbs IR radiation at a different wavelength (6.6 μm). Therefore, it does not interfere with the

CO_2 determination. The CO_2 and H_2O are removed by sodium hydroxide on a clay base and magnesium perchlorate, since they will interfere with the N_2 measurement. The helium flow carries the N_2 to the TC cell where it is measured.[24] Oxygen and nitrogen are displayed as weight percent. A typical analysis cycle takes approximately 1 min. Some high-end instruments incorporate a feature that ''ramps'' the furnace temperature during the analysis cycle. The purpose of this technique is to separate forms of oxygen (oxides) and nitrogen (nitrides) in iron and steel. This feature is generally limited to research, since the separation of oxides and nitrides using this technique is somewhat complex and interpretation of the data requires detailed knowledge of the sample metallurgy.[25]

4.2 Detection Limits and Ranges

State-of-the-art inert gas fusion oxygen and nitrogen instrumentation is capable of determining total oxygen and nitrogen in iron and steel products from parts per million to weight percent levels. These instruments are able to analyze 1 g samples of iron and steel with oxygen contents ranging from 0.000 05% to 0.1% and nitrogen contents from 0.000 05% to 0.5%. The ranges can be extended upward by reducing the sample weight. A variety of instruments are available that have expanded range capabilities. The DL of the instrumentation is normally calculated from standard deviations of a number of measurements on an individual instrument at or near the lowest concentration of the element being determined. Some instrument manufacturers may determine their published DLs and ranges (specifications) for a particular model of an instrument by combining statistics from a number of individual instruments. Dosing known volumes of N_2, CO, or CO_2 directly into the measure stream is another competent way to determine the measurement precision of an inert gas fusion oxygen and nitrogen determinator. This permits evaluation of the instrument's detection system without the influence of other variables such as sample homogeneity and fusion characteristics. The instrument's performance throughout its range can be evaluated by changing the volume and concentration of the dose gas. Mixtures of N_2, CO, or CO_2 in helium can be used as well. Some instrument manufacturers produce instruments with built-in gas dosing capabilities.

4.3 Analytical Procedures

In this section, analytical parameters that have an impact on the precision of the oxygen and nitrogen determination when using inert gas fusion instrumentation will be reviewed.

4.3.1 Crucibles

The graphite crucibles used to contain the sample during the fusion cycle are an important part of the analysis. They not only hold the sample during the fusion cycle, they are essentially carbon resistors that supply the heat that fuses a sample and the carbon that reacts with oxygen from the sample to form CO. These crucibles are machined from high-purity graphite rods and are made in a variety of shapes and sizes. The crucibles contain impurities including moisture and gases that have been absorbed during handling and storage. Just prior to transferring the sample to the crucible, these impurities are removed during the outgas cycle. State-of-the-art inert gas fusion oxygen and nitrogen instruments offer multiple outgas cycle capabilities that are operator selectable. Multiple outgas cycles subject the graphite crucible to several hot/cold cycles to enhance the removal of impurities from the crucible subsequently improving precision. The shape and size of the graphite crucible will have some effect on the application. Figure 9 shows some basic crucible designs used by inert gas fusion oxygen and nitrogen instruments. The standard crucible is designed for the majority of applications and typically can obtain a maximum temperature of 2600 °C. The high-temperature version is designed to reach temperatures of 3000 °C and is typically used for high-temperature alloys or refractory metals. During the fusion of a sample, electric current is flowing through the graphite crucible. As the sample melts, it will wet the inner walls of the crucible and change the resistance and temperature of the crucible. The standard and high-temperature versions are single wall crucibles. In some cases the sample may actually penetrate the crucible sidewall causing a loss in temperature. A third type of crucible used in these instruments is a double-wall design with both an inner and outer crucible. The inner crucible is replaced after each analysis, while the outer crucible is reused several times. This crucible offers improved temperature control and has a maximum temperature of 2200 °C. The current flows through the outer crucible and is not affected as the sample melts and penetrates in the inner crucible. This crucible is primarily used for oxide characterization.

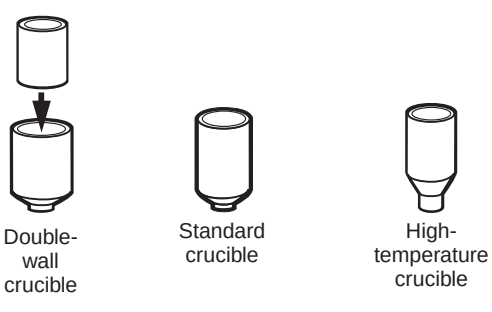

Figure 9 Graphite crucible designs.

With a 2200 °C maximum temperature, the double-wall crucible is not recommended for nitrogen determination owing to low nitrogen recovery.

4.3.2 Carrier Gas

The purity of the carrier gas is an important factor in obtaining suitable oxygen and nitrogen results on an inert gas fusion instrument. An ultra-pure carrier grade (UPC) of helium, with a minimum purity of 99.9995%, is typically required for this instrumentation. Trace amounts of H_2O, O_2, CO, or CO_2 in the helium carrier gas can cause some difficulties in this instrumentation. Most commercially available instruments will have reagent scrubbers such as sodium hydroxide on a clay base and magnesium perchlorate to remove CO_2 and H_2O, respectively, from the incoming gas supply. More advanced instruments will typically incorporate additional heated catalysts such as copper metal to remove O_2, and CuO to convert CO to CO_2, which is subsequently removed by the sodium hydroxide. During a normal analysis (as described in section 4.1), the sample is retained above the crucible until the outgas cycle is complete. The crucible temperature is lowered and electronic baselines for the detectors are established with the carrier gas flowing over the hot crucible. This allows the instrument to ''mask'' or subtract any contribution from impurities that are still in the carrier gas or the crucible. This mode of analysis is typically referred to as the automatic mode and it is the preferred mode of operation. In some instances, such as the analysis of powders or chips, the manual mode of operation is desirable. In the manual mode, the sample-loading chamber is bypassed. After the graphite crucible is outgassed and a short cool-down cycle has elapsed, the furnace is opened and the sample is loaded directly into the crucible. The furnace is closed and after a purge cycle the sample is fused. In this case, the electronic baselines for the detectors are taken from a cold crucible. Therefore, in the manual mode the instrument cannot ''mask'' the effects of impurities in the carrier gas. In addition, exposing the outgassed crucible to atmosphere causes higher and more erratic oxygen blanks.

4.3.3 Methodology

The methodology can also adversely affect the precision of oxygen and nitrogen. The performance of a method can be determined by conducting an ILS. Method development organizations such as the ASTM and the ISO produce written procedures and use an ILS to determine the suitability and scope of a method. The ILS will establish longer term sources of variability within and between laboratories and instruments.[26] For example, ASTM method E 1019 ''Determination of Carbon, Sulfur, Nitrogen and Oxygen in Steel and in Iron, Nickel, and Cobalt Alloys''[11] includes oxygen and nitrogen data obtained using an ILS to determine the scope of the method.

4.4 Calibration

The accuracy of the inert gas fusion oxygen and nitrogen determinator depends on proper calibration. Calibration is accomplished by analyzing reference materials with known oxygen and nitrogen contents. A definite advantage of the inert gas fusion method for oxygen and nitrogen determination in iron and steel is the lack of matrix interference problems associated with this method. Samples of iron, cast iron, plain carbon steel, low- and high-alloy steel, tool steel, specialty steel, stainless steel, and high-temperature alloys can be analyzed using the same technique.

Commercially available instrumentation is normally factory linearized, permitting single point calibrations. Typically, three to five replicates of a reference material are analyzed and the calibration slope is adjusted using the software program of the instrument. The accuracy of the calibration is dependent on the accuracy of the method or methods used to determine the values of the standard. If several reference materials are analyzed on an individual instrument and the oxygen or nitrogen results do not agree with the accepted concentrations, either the instrument linearity or the accepted value of the reference is in question. Instrument linearity can be verified by analyzing fractional weights of a reference material or by combining reference materials in the same fashion as described in section 3.4 regarding carbon and sulfur calibration and linearity.

Combining reference materials is more important when linearizing the induction furnace combustion method than it is with the electrode furnace inert gas fusion method. The purpose is to maintain approximately 1 g sample mass while changing the analyte content. This is important because the sample mass couples with the induction field producing and maintaining sufficient temperatures to assure complete oxidation of the carbon and sulfur in the sample. In the case of the inert gas fusion method, the graphite crucible temperature is controlled by the current passed through it, not by the mass of the sample. Therefore, fractional weights of a single sample can also be used to verify instrument linearity for the inert gas fusion method. The main problem with mixing reference materials while linearizing using the inert gas fusion method relates to oxygen determination. As with carbon and sulfur, there are a variety of iron and steel reference materials in chip or granular form with certified nitrogen values. This facilitates the mixing of materials to simulate differing concentrations of nitrogen. There are virtually

no chip or granular iron or steel reference materials with certified oxygen values. This is primarily due to surface area and oxidation potential. Surface oxygen must be removed or otherwise accounted for in order to obtain accurate and precise oxygen results on iron and steel samples. Chip and granular samples have varying surface areas which correspond to varying levels of oxygen. While removing surface oxidation from solid samples is feasible, removing surface oxidation from chips and/or granular samples is not practical.

4.5 Sampling and Sample Preparation

Samples will need to be taken from liquid metal, cast metal, wrought metal, and finished products. The sampling method must provide a representative sample. For nitrogen determination, the samples can be in the form of solid slugs, pins, wires, strips, chunks, drillings, shavings, or powders. For oxygen determination, the samples should be solids with a shape that permits the surface to be mechanically prepared to remove surface oxygen. Surface oxidation is a primary consideration when accurate oxygen results are required. Two documents that contain comprehensive sampling guidelines for iron and steel are ASTM E 1806[13] and ISO 14284.[14] Sample preparation of iron and steel for nitrogen determination can be as simple as washing in a solvent such as acetone and drying with warm air. Sample preparation for oxygen typically requires removing a portion of the surface to remove surface oxides, followed by washing in acetone and drying with warm air. If simultaneous nitrogen and oxygen determination is desired, the sample surface should be removed either by turning on a lathe or by abrading with a flat mill file, then washing in acetone and drying with warm air. Table 3 shows oxygen and nitrogen data obtained on a modern inert gas fusion instrument for a steel wire sample that was analyzed as received and after the surface was prepared by abrading with a flat mill file and washing in acetone.

Table 3 Comparison of O and N values as received versus abraded

Steel wire	Rinsed in acetone		Abraded and rinsed	
	Oxygen (%)	Nitrogen (%)	Oxygen (%)	Nitrogen (%)
	0.0435	0.0075	0.0109	0.0070
	0.0410	0.0075	0.0103	0.0074
	0.0416	0.0075	0.0102	0.0070
	0.0535	0.0075	0.0102	0.0070
	0.0244	0.0070	0.0103	0.0071
	0.0454	0.0074	0.0107	0.0072
	0.0360	0.0070	0.0103	0.0072
	0.0464	0.0074	0.0103	0.0072
	0.0415	0.0074	0.0105	0.0070
	0.0440	0.0073	0.0103	0.0070
$\bar{x}$ (%)	0.0417	0.0073	0.0104	0.0071
s (%)	0.0076	0.00020	0.00021	0.00012

$\bar{x}$ = Mean, s = estimate of standard deviation.

4.6 Performance Data

Table 4 shows oxygen and nitrogen data obtained on a variety of iron and steel samples using a modern inert gas fusion instrument. Process capability or gage R&R procedures are commonly used for measuring the R&R of a process. For a more detailed explanation, refer to section 3.6.

5 HYDROGEN DETERMINATION

5.1 Description of Methodology and Instrumentation

The hydrogen content of iron and steel is routinely determined by fusing or heating a sample in an inert carrier gas followed by measurement with a TC cell. There are two main methods typically used for hydrogen determination: high temperature inert gas fusion and hot extraction below the melting point (which is usually at a temperature of 1100 °C in an inert carrier gas stream). Like oxygen and nitrogen inert gas fusion instruments, a typical inert gas fusion hydrogen determinator is computer controlled and has a water-cooled electrode furnace coupled to a measurement unit. The electrode furnace has alloy-tipped

Table 4 Oxygen and nitrogen results obtained on an inert gas fusion instrument

Material	Oxygen (%)				Nitrogen (%)			
	Expected	$\bar{x}$	s	RSD	Expected	$\bar{x}$	s	RSD
Iron	0.0448	0.0448	0.00035	0.78	0.0007	0.00069	0.00006	8.70
Ingot iron	0.0436	0.0434	0.0005	1.15	0.0018	0.0018	0.0001	5.56
Plain carbon steel	0.0107	0.0108	0.0001	0.93	0.0071	0.0071	0.0001	1.41
Low-alloy steel	0.0008	0.00075	0.00003	4.00	0.0166	0.0166	0.00038	2.29
High-alloy steel	0.0212	0.0211	0.0001	0.47	0.0286	0.0290	0.0004	1.38
Stainless steel	0.0245	0.0244	0.0001	0.41	0.0504	0.0503	0.0004	0.80

$\bar{x}$ = Mean, s = estimate of standard deviation.

copper electrodes capable of heating a graphite crucible to temperatures of up to 3000 °C. The measurement unit contains a TC detection cell along with the necessary hardware and software to perform the measurement. Argon is used as a carrier gas instead of helium. Argon has a significantly lower TC than hydrogen, so that subsequently excellent detection capabilities are attained. Since these instruments use a TC detector, gases such as CO, CO_2, and N_2 released from the sample must be removed or separated from the hydrogen if accurate measurement is to take place. Removal of these gases is accomplished by a series of reagents designed to react with the CO and CO_2 and by a column to separate the N_2 (N_2 is retained on the column for a longer time) from the hydrogen.

Another instrument design incorporates an induction furnace (instead of an electrode furnace) coupled to a measurement unit. The induction furnace can be used to fuse iron and steel samples in a graphite crucible, or it can be used to heat samples below their melting points in a quartz crucible using the hot extraction technique. When using an induction furnace, nitrogen is used as the carrier gas because argon will couple with the induction field and form a discharge. However, there can be a problem associated with using a graphite crucible and a nitrogen carrier gas: lower hydrogen recoveries are usually obtained. This reduced recovery efficiency is likely to be due to a reaction between the carbon from the crucible, the hydrogen from the sample, and the nitrogen carrier gas at elevated temperatures. However, the use of an induction furnace permits larger samples to be analyzed. Sample weights of 5 g are typically analyzed on an induction furnace unit, compared with 1–2 g samples which can be analyzed on an electrode furnace unit. When using the hot extraction technique in a quartz crucible, the crucible can be used several times and it does not have to be outgassed, as is necessary with the graphite crucible. The majority of users of the induction furnace hydrogen determinators use the hot extraction technique.

A third instrument design uses a resistance-heated furnace with a quartz reaction tube, nitrogen or argon as the carrier gas, and TC detector. The furnace is maintained at 1100 °C. This instrument does not use a crucible, but rather the samples are placed in the open end of the reaction tube and the tube tilts back sending the sample to a built-in stop in the hot zone of the furnace. The reaction tube is purged with an excess amount of carrier gas which creates a ''gas blanket'' at the open end of the tube. Thus, atmospheric gases are prevented from entering the tube. At the end of the extraction cycle, the tube tilts downward and the sample is expelled. This design offers simplicity and ease of operation as well as the ability to analyze 5 g samples. The instrument is not designed to analyze chips or granular samples.

Some models of resistance furnace instruments incorporate a piercing unit designed to determine ''diffusible'' hydrogen when used with an appropriate sampling device. The sampling device is a dual-chamber sample probe designed to take a sample from molten iron or steel and to trap the hydrogen that diffuses out of the sample during solidification. The sampler is placed in the piercing unit of the determinator, pierced, and then the diffusible hydrogen trapped in the outer chamber of the sampler is delivered to the TC detector for measurement. After this measurement, the sampler is removed from the piercing unit and the outer chamber is removed to expose the solidified sample. This solid sample is placed in the resistance furnace for extraction of the residual hydrogen. The diffusible and residual hydrogen results are combined and displayed as total hydrogen content.

5.1.1 Analysis Sequence

Figure 10 is a simplified flow diagram for an electrode furnace hydrogen determinator. An electrode furnace hydrogen determinator analysis procedure is similar to that of an inert gas fusion oxygen and nitrogen unit (section 4.1). A preweighed sample of 1–2 g is placed in the furnace loading port. The sample is retained in a chamber directly above the graphite crucible area. A graphite crucible containing approximately 1 g of high-purity tin flux (usually two 0.5 g tin pellets) is placed on the lower furnace electrode and the furnace is closed. Addition of tin prevents the fused sample from penetrating the crucible, resulting in more consistent results. After a short purge cycle, electric current is passed through the crucible heating it to temperatures near 2300 °C. Argon is flowing through the furnace to an exhaust port which removes contaminates released from the crucible and flux. This is typically referred to as the outgas cycle. Following the outgas cycle, the crucible temperature is lowered by reducing the electrical current and the argon is routed through the TC cell. An electronic baseline is established for the TC cell and then the sample is automatically dropped into the crucible. The oxygen in the sample reacts with the carbon from the crucible forming CO, the nitrogen is released as N_2, and the hydrogen is released as H_2. These gases are carried along with the argon through particle filters to Schutze™ reagent (iodine pentoxide over silica gel) which converts CO to CO_2 without oxidizing the H_2 to H_2O. The gases are then passed through sodium hydroxide on clay followed by anhydrone to remove the CO_2. The remaining H_2 and N_2 are carried along with the argon to a molecular sieve column that separates the two. The H_2 elutes from the column first and is measured by the TC cell. The N_2 exits the column approximately 1 min after

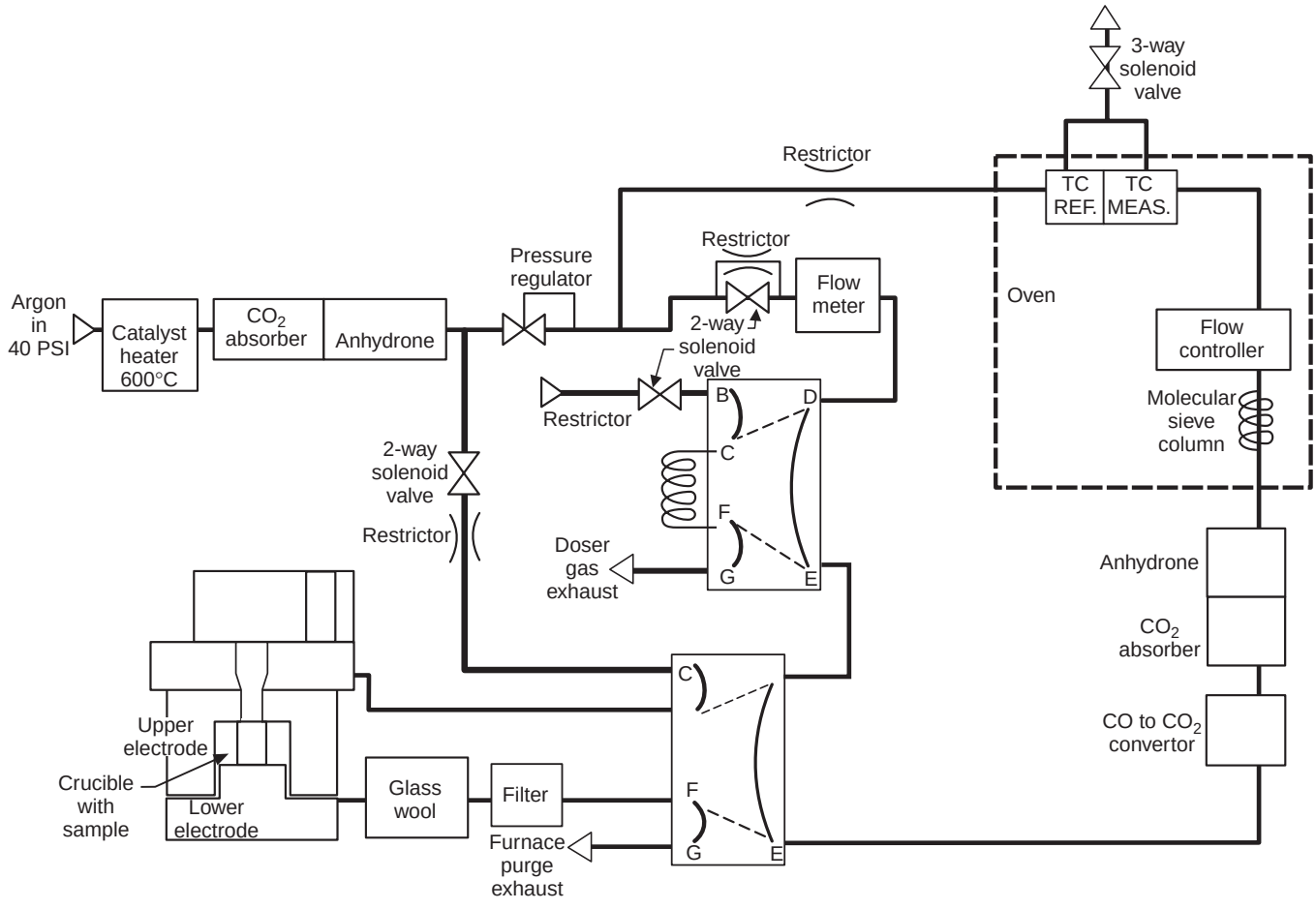

Figure 10 Flow diagram for electrode furnace hydrogen determinator.

the H_2 has been measured, passing through the TC cell without being recorded. Hydrogen results are displayed as weight parts per million. A typical analysis cycle takes approximately 3 min.

A hydrogen determinator equipped with an induction furnace (Figure 11) can be used either to fuse a sample in a graphite crucible or to heat a sample to just below its melting point using a hot-extraction technique. The analysis sequence used for the fusion of a sample in a graphite crucible is similar to that of an electrode furnace. A preweighed sample of approximately 5 g is placed in a sample loading head and a single-use graphite crucible containing approximately 2 g of tin flux (usually four 0.5 g tin pellets) is placed on the furnace pedestal. The furnace is closed and the pedestal positions the graphite crucible in the center of the HF coil. After a short purge cycle, power is applied to the HF coil and the graphite crucible is heated by induction to a temperature of up to 1800 °C. The nitrogen carrier gas sweeps out contaminants released from the crucible and the tin flux. Following the outgas cycle, the furnace power is shut off, the crucible is raised close to the sample loading port, and the sample is transferred to the crucible. The crucible is positioned in the center of the HF coil and power is applied. The crucible heats up rapidly, and the sample melts. Any hydrogen in the sample is released as H_2, the oxygen reacts with the carbon from the graphite crucible to form CO, while nitrogen is released as N_2. The CO is removed by passing through Schutze™ reagent converting it to CO_2, which is then removed with sodium hydroxide on a clay base. The nitrogen carrier gas carries the H_2 to the TC cell for measurement. Since nitrogen is used as the carrier gas, any nitrogen released from the sample has no effect on the measurement. Analysis time is approximately 4–5 min.

When used in the hot-extraction mode, an induction furnace hydrogen determinator analysis sequence is relatively simple. A preweighed sample of approximately 5 g is placed in a quartz crucible. The crucible, with the sample, is positioned on the furnace pedestal. The furnace is closed and the pedestal positions the crucible and sample in the center of the HF coil. After a purge

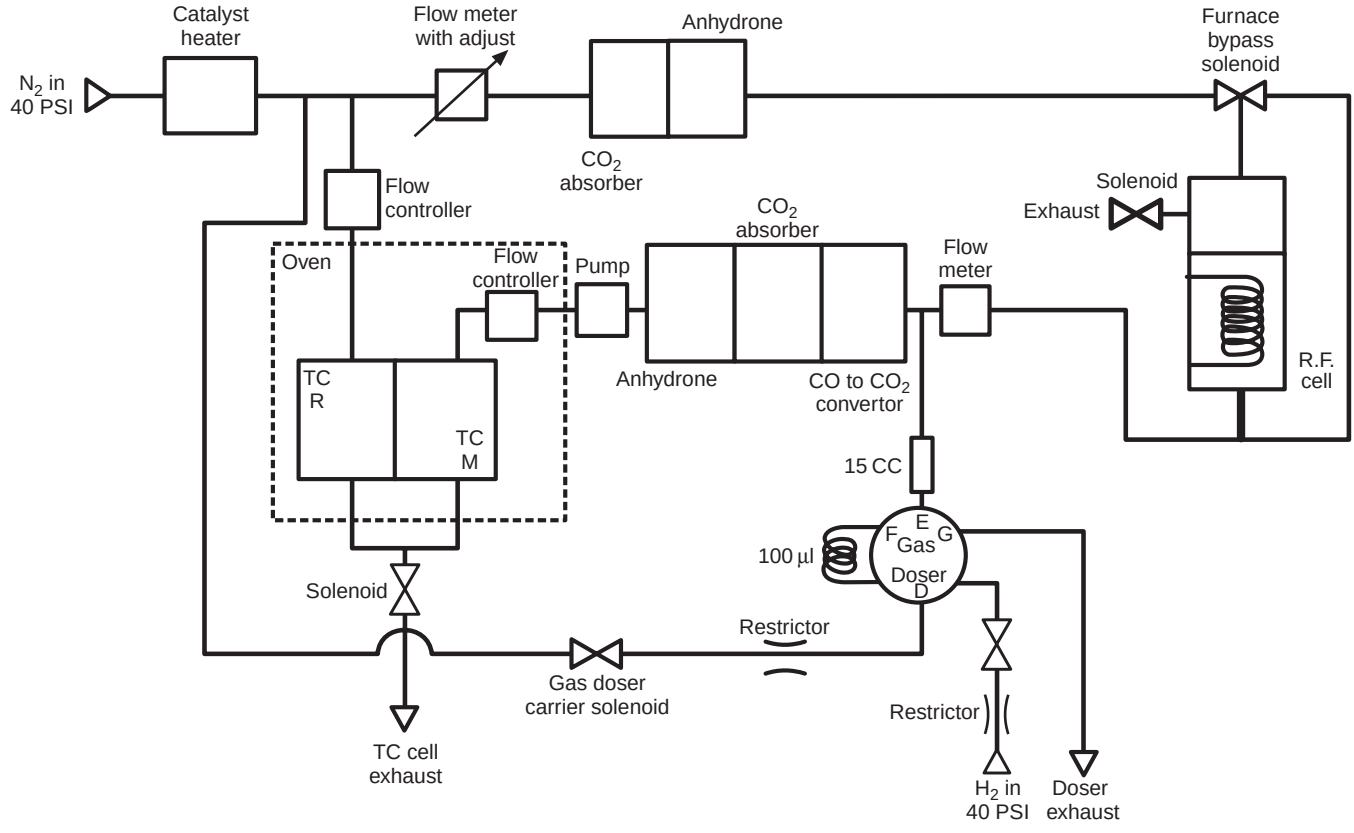

Figure 11 Flow diagram for induction furnace hydrogen determinator.

cycle, power is applied to the HF coil and the sample begins to heat up by coupling with the HF current. The computer controls the furnace power and the sample heats up to approximately 1100 °C. The hydrogen is released as H_2 and is carried to the TC detector by the nitrogen carrier gas. Oxygen and nitrogen are not released at this temperature. There is no outgas cycle and the sample loading port is not used. The analysis time is approximately 4–5 min. This is similar to a fusion analysis with an outgas cycle because it takes longer to extract the hydrogen when the sample is not melted. The extraction time is dependent on the cross-section and the weight of a sample.

A resistance-furnace-based hydrogen determinator (Figure 12) typically uses the hot-extraction method. A sample of approximately 5 g is placed in the open end of the reaction tube. The reaction tube tilts back and the sample slides to the center of the hot zone of the furnace, which is set at 1100 °C. Hydrogen is released and delivered to the TC detector by the carrier gas. The analysis time is 4–6 min.

Some commercially available resistance-furnace hydrogen determinators are equipped with a piercing device used in conjunction with special probe samplers. These probes are used to sample molten metal and are designed to capture diffusible hydrogen. The diffusible hydrogen is measured when the probe is inserted in the piercing unit. Then the probe is removed from the piercing unit and the outer chamber of the probe is removed exposing the solidified metal sample. The sample is subsequently placed in the furnace for the determination of residual hydrogen. The diffusible and residual hydrogen values are combined and then reported as total hydrogen. The combined analysis time for diffusible and residual hydrogen determination is 5–8 min.

5.2 Detection Limits and Ranges

Modern hydrogen determinators have analytical ranges from 0.01–400 ppm of hydrogen based on 1–5 g sample weights. This range is more than adequate, as the hydrogen content of iron and steel typically ranges from 0.1–20 ppm. The DLs of a hydrogen determinator are usually calculated from standard deviations of a number of measurements obtained on an individual instrument at or near the lowest concentration of hydrogen being determined. Some instrument manufacturers may determine their published DLs and ranges for a particular model by combining statistics from a number

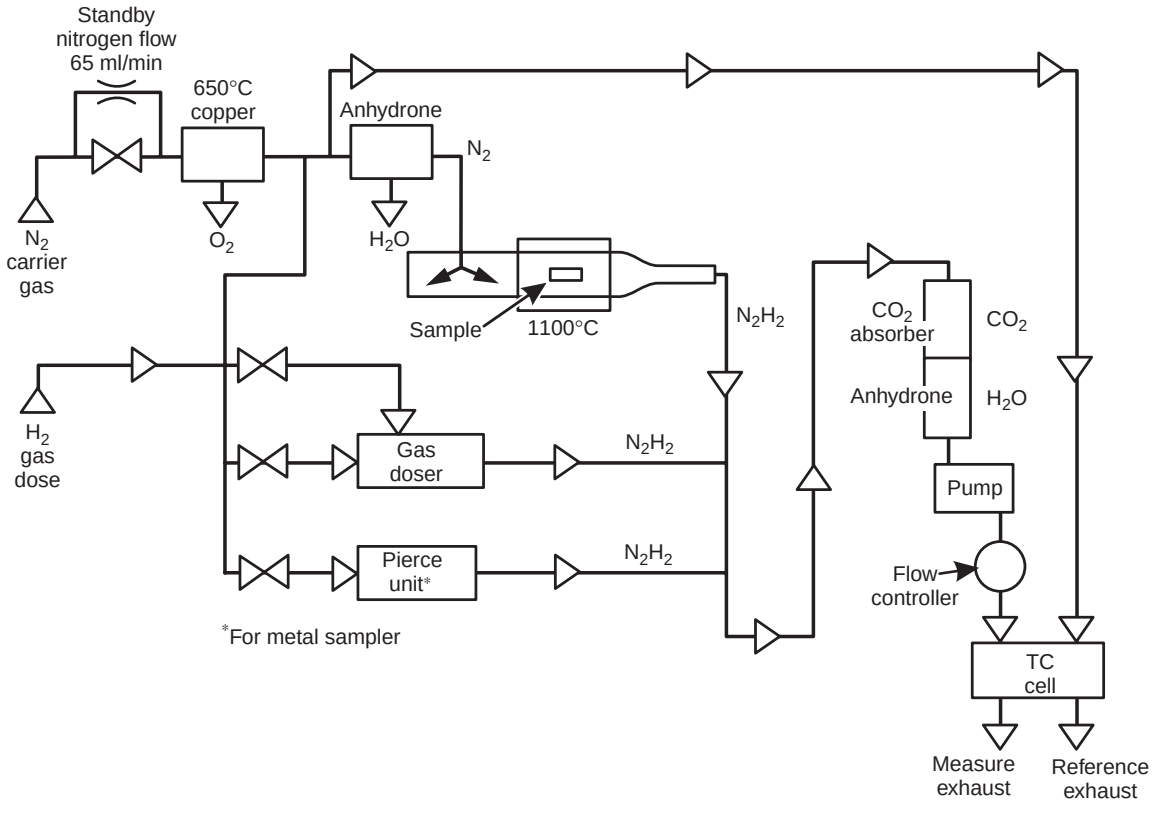

Figure 12 Flow diagram for resistance furnace hydrogen determinator.

of individual instruments. Dosing known volumes of H_2 directly into the measure stream is an excellent way to determine the measurement capabilities of a hydrogen determinator. This evaluates the measurement capabilities of the instrument's detection system without the influence of variables inherent in sampling, sample preparation, or actual analysis of a sample. Instrument performance throughout its range can be evaluated by changing the volume and concentration of the dose gas. Mixtures of H_2 in either nitrogen or argon can be used depending on the carrier gas of the instrument. Most manufactured hydrogen determinators have a built-in gas dose capability.

5.3 Analytical Procedures

Analytical parameters that have an impact on the precision and accuracy of hydrogen determination of iron and steel when using these instruments will be discussed in the following section.

5.3.1 Crucibles

Graphite crucibles are used when samples are fused in an electrode furnace and an induction furnace model in the fusion mode. They are an important component in obtaining optimum accuracy and precision. These crucibles are machined from high-purity graphite rods and may be further purified by vacuum degassing. Even so, they can contain hydrogen in the form of moisture from subsequent handling and storage. This contamination must be removed by outgassing the crucible just prior to transferring the sample to the crucible for analysis. Electrode furnace units typically have the capability to perform multiple outgas cycles. Multiple outgas cycles subject the graphite crucible to two or more hot/cold cycles which enhance the removal of impurities from the crucible. Induction furnace units usually do not offer this option because the crucible is evenly heated by the induction coil, which in turn removes the crucible contamination very effectively. The crucible temperature for hydrogen determination is typically lower than that used for oxygen and nitrogen determination. Temperatures of less than 2200 °C are normally used on an electrode furnace unit, depending on the crucible shape. A typical induction furnace unit will obtain temperatures less than 1800 °C during a fusion cycle. The quartz crucibles that are used during the hot-extraction technique on an induction furnace unit do not need to be degassed prior to use. However, they must be clean and should only be handled with tongs.

5.3.2 Flux

High-purity tin is used as a flux in the graphite crucibles. The tin is outgassed along with the crucible and trace contamination is not normally a problem. The tin dust that is generated due to volatilization during the outgas and analysis cycles must be removed periodically, otherwise problems with precision may occur.

5.3.3 Carrier Gas

The purity of the carrier gas used in these instruments is a critical component in obtaining accurate and precise hydrogen results. Typically, UPC grades with a minimum purity of 99.999% of argon and nitrogen are required. Moisture in the carrier gas and/or in the gas delivery lines is the main concern. All of the hydrogen determinators described in this section normally have incoming gas purification catalysts and reagents designed to remove small amounts of O_2, CO, and H_2O. However, sometimes trace amounts of H_2O and O_2 may make it through these reagents. Additional scrubbers, typically using lithium-based reagents, can be added further to remove trace amounts of H_2O and O_2.

5.4 Calibration

The accuracy of hydrogen determination using this method depends on proper calibration. There are few iron and steel reference materials with certified hydrogen values; therefore, gas dosing with hydrogen is the primary technique used to calibrate these instruments. The majority of manufactured instruments will have built-in gas dosing capabilities. Some instruments have multiple volumes, while others have a single volume with the capability to fill and dose several times during a single analysis to simulate various levels of hydrogen. Instrument linearity is normally not an issue with these determinators, because the hydrogen levels in iron and steel are generally less than 10 ppm.

5.5 Sampling and Sample Preparation

Sampling and sample preparation of iron and steel for hydrogen determination is extremely important. Hydrogen is mobile in iron and steel products and will diffuse out at room temperature. This diffusion can be very pronounced especially in samples with thin cross-sections. Hydrogen can be picked up by molten steel during production only to diffuse out after casting.[27] This can lead to problems such as bleeding ingots, blowholes, embrittlement, and low ductility. Hydrogen is typically removed from liquid steel by vacuum degassing. In solid steel, hydrogen can cause internal ruptures.[28] This tendency for hydrogen to diffuse out of steel makes sampling and sample preparation a definite challenge. Lowering the temperature of the sample reduces the diffusion of hydrogen from iron and steel. Samples taken from molten steel and iron must be quenched in ice water within 10 s after taking the sample. After the sample is cooled, it must be stored in a refrigerant such as liquefied nitrogen or a mixture of acetone and solid carbon dioxide. Likewise, samples taken from solid steel products should be kept cool at all stages of sampling, storage, and sample preparation. Filing, light grinding, or grit blasting can be used to prepare the surface of the test portion. The sample must be cooled at frequent intervals to minimize hydrogen diffusion. The test portion must be degreased by immersion in acetone, dried with a rough vacuum, and followed by immediate analysis. Evacuated sampling probes designed to capture diffusible hydrogen are available. These sampling devices are designed to take samples from molten metal and are used in conjunction with instruments equipped to handle them. When the probe is immersed in the melt, a fusible cap melts and liquid metal is pulled into an inner chamber. The probe is removed from the melt and seals as the metal solidifies.[14] The probe is quenched in cold water and delivered to the laboratory for analysis. ASTM E 1806[13] and ISO 14284[14] contain sampling and sample preparation recommendations for steel and iron and are excellent reference documents.

5.6 Performance Data

Table 5 contains hydrogen results typically obtained from the analysis of stainless steel on the electrode furnace, induction furnace, and resistance furnace hydrogen determinators. A comparison of the hot-extraction technique versus the fusion technique of hydrogen analysis on an induction furnace appears in Table 6. For a review of process capability or gage R&R procedures that are commonly used for measuring the R&R of a process, refer to section 3.6.

Table 5 Typical hydrogen data on the different types of furnace

	Material: stainless steel		
	Electrode furnace	Induction furnace (HF)	Resistance furnace
Expected (ppm)	4.80	4.74	4.74
$\bar{x}$ (ppm)	4.81	4.67	4.66
s (ppm)	0.184	0.082	0.068
RSD (%)	3.83	1.76	1.46
n	10	10	10

$\bar{x}$ = Mean, s = estimate of standard deviation, n = number of measurements.

Table 6 Typical hydrogen data: hot-extraction technique vs inert gas fusion technique

	Material: stainless steel	
	Hot-extraction technique	Inert gas fusion technique
Expected (ppm)	4.80	4.80
$\bar{x}$ (ppm)	4.84	4.81
s (ppm)	0.174	0.184
RSD (%)	3.59	3.82
n	10	10

$\bar{x}$ = Mean, s = estimate of standard deviation, n = number of measurements.

ACKNOWLEDGMENTS

Editorial review by Dave Valensi, Brian Pack, Rob Broadwater, and Angie Seger is gratefully acknowledged. The authors thank Bill Strzynski for this article's illustrations.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
DL	Detection Limit
HF	High-frequency
ILS	Interlaboratory Study
IR	Infrared
ISO	International Organization for Standardization
R&R	Reproducibility and Repeatability
RSD	Relative Standard Deviation
TC	Thermal Conductivity
UPC	Ultra-pure Carrier

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction • Metal Analysis, Sampling and Sample Preparation in

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. *Annual Book of ASTM Standards*, ASTM, West Conshohocken, PA, 607, Vol. 3.06, 1997.
2. G.E.F. Lundell, J.I. Hoffman, H.A. Bright, *Chemical Analysis of Iron and Steel*, Wiley, New York, 157, 1950.
3. G.E.F. Laundell, J.I. Hoffman, H.A. Bright, *Chemical Analysis of Iron and Steel*, Wiley, New York, 249–251, 1950.
4. G.E.F. Laundell, J.I. Hoffman, H.A. Bright, *Chemical Analysis of Iron and Steel*, Wiley, New York, 397–428, 1950.
5. D. Lawrenz, *The Determination of C, S, O, N and H in High Purity Base Metals*, Proc. UHPM-94, Kitakyushu-City, Japan, Japan Institute of Metals, Sendai, 183–190, 24–27 May, 1994.
6. H.A. Strobel, W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, 3rd edition, Wiley-Interscience, New York, 915, 1989.
7. *Annual Book of ASTM Standards*, ASTM, Pennsylvania, 282–284, Vol. 3.06, 1997.
8. H.A. Strobel, W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, 3rd edition, Wiley-Interscience, New York, 280–281, 1989.
9. D. Lawrenz, *Ultra Low Carbon and Sulfur Determination in High Purity Steel*, Proceedings CETAS-95, Luxembourg, 16–18 May, 1995, European Commission, Luxembourg, 1996.
10. *Annual Book of ASTM Standards*, ASTM, West Conshohocken, PA, 582–594, Vol. 3.06, 1997.
11. *Annual Book of ASTM Standards*, ASTM, West Conshohocken, PA, 282–300, Vol. 3.06, 1997.
12. ISO 15349-2, *Unalloyed Steel–Determination of Low Carbon Content–Part 2: Infrared Absorption Method After Combustion in an Induction Furnace (with preheating)*, ISO, Geneve, 1998.
13. *Annual Book of ASTM Standards*, ASTM, West Conshohocken, PA, 606–607, Vol. 3.06, 1997.
14. ISO 14284, *Steel and Iron–Sampling and Preparation of Samples for the Determination of Chemical Analysis*, ISO, Geneve, 15, 1996.
15. V.E. Kane, 'Process Capability Indices', *J. Quality Technol.*, **18**, 41–52 (1986).
16. K.R. Bhote, *World-class Quality*, AMA Membership Publications, New York, 1988.
17. L.K. Chan, S.W. Cheng, F. Spriring, 'A New Measure of Process Capability', *Cpm. J. Quality Technol.*, **20**, 162–175 (1988).
18. A.J. Duncan, *Quality Control and Industrial Statistics*, Richard D. Irwin, Homewood, IL, 1974.
19. D.C. Montgomery, *Design and Analysis of Experiments*, 3rd edition, John Wiley & Sons, New York, 1991.
20. J. Houston (ed.), *DataMyte Handbook: A Practical Guide to Computerized Data Collection for Statistical Process Control*, Datamyte, Minnetonka, MN, 1992.
21. American Society of Quality Control/Automative Industry Action Group, *Measurement Systems Analysis Reference Manual*, AIAG, Troy, MI, 1990.

22. American Society of Quality Control/Automative Industry Action Group, *Fundamental Statistical Process Control Reference Manual*, AIAG, Troy, MI, 1991.
23. H.A. Strobel, W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, 3rd edition, Wiley-Interscience, New York, 915, 1989.
24. *Annual Book of ASTM Standards*, ASTM, West Conshohocken, PA, 290–297, Vol. 3.06, 1997.
25. R. Prumbaum, K. Orths, 'Method and Apparatus for the Rapid Determination of Gases–Oxygen in Particular–and of the Combined Forms in Which They are Present in Metals', *Giess-Forch*, **2/3**, 71 (1979).
26. *Annual Book of ASTM Standards*, ASTM, West Conshohocken, PA, 582–594, Vol. 3.06, 1997.
27. H.E. McGannon, *The Making, Shaping, and Treating of Steel*, 10th edition, Herbick & Held, Pittsburgh, PA, 488, 1985.
28. H.E. McGannon, *The Making, Shaping, and Treating of Steel*, 10th edition, Herbick & Held, Pittsburgh, PA, 675, 1985.

X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Didier Bonvin
ARL Applied Research Laboratories, Ecublens, Switzerland

1 Introduction 1
2 Sampling 2
3 Specimen Preparation 4
3.1 Metals 4
3.2 Pressed Powders 4
3.3 Fused Beads 5
4 Calibration of the X-ray Fluorescence Instrument 6
4.1 Introduction 6
4.2 Setting-up Samples and Control Samples 6
5 Material Types and Analytical Requirements 7
5.1 Analysis of Iron Ores 7
5.2 Analysis of Sinters 8
5.3 Analysis of Slags 8
5.4 Analysis of Irons 9
5.5 Analysis of Steels 10
5.6 Analysis of Ferroalloys 12
5.7 Miscellaneous Analyses 13
6 Choice of Spectrometer 14
6.1 Sequential X-ray Fluorescence Spectrometer 15
6.2 Simultaneous X-ray Fluorescence Spectrometer 16
6.3 Simultaneous/Sequential X-ray Fluorescence Spectrometer 16
6.4 Simultaneous/Sequential X-ray Fluorescence Spectrometer with X-ray Diffraction Capability 17
7 Conclusion 17
Acknowledgments 18
Abbreviations and Acronyms 18
Related Articles 18
References 18

In reviewing the current status of wavelength-dispersive X-ray fluorescence (WDXRF) spectrometry in the iron and steel industry, attention is given to aspects such as sampling and specimen preparation, calibration requirements and analytical demands for the various material types and the choice of instrument configuration.

1 INTRODUCTION

WDXRF spectrometry is well established as an analytical technique in the iron and steel industry. WDXRF is routinely used for the analysis of raw materials, sinters, slags and coupled with optical emission spectrometry (OES) for determination of the elemental composition of cast iron and steels (Figure 1).

One of the great advantages of WDXRF is its ability to analyze both conducting and nonconducting solids and liquids, whereas OES is used mainly for the analysis of metals. Both techniques are really complementary in the iron and steel industry:

- Elements such as boron, carbon and nitrogen at levels down to parts per million (ppm) cannot be measured by X-ray fluorescence (XRF) spectrometry but OES allows the quick reporting of their composition in metals, which is essential in this industry.
- All other elements required (except hydrogen) can be measured both by XRF and OES. The latter technique is superior at low concentration levels and WDXRF allows better precision at high concentration levels, e.g. nickel and chromium in high-alloy steels, leading to substantial gains in the amount of expensive alloying elements that are added to make a given alloy grade.
- Analysis of nonconductive materials, such as slags and sinters, is done very effectively and quickly by WDXRF.

For the analysis of metals, this complementarity means that very often the same specimen is used for both analysis techniques.

The demands for lower limits of detection, improved reliability of trace element analysis and tight control of alloy compositions have been steadily increasing in recent years. In addition, there is a trend within the iron and steel industry to place the analytical capability closer to the process that uses the analysis values so as to facilitate and speed up any corrective actions during production. These demands have led manufacturers of WDXRF instrumentation to design instruments that are more sensitive, have greater flexibility and which can be easily incorporated into automated laboratories. In this industry, energy-dispersive X-ray fluorescence (EDXRF) spectrometry is generally not selected because of its poor

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

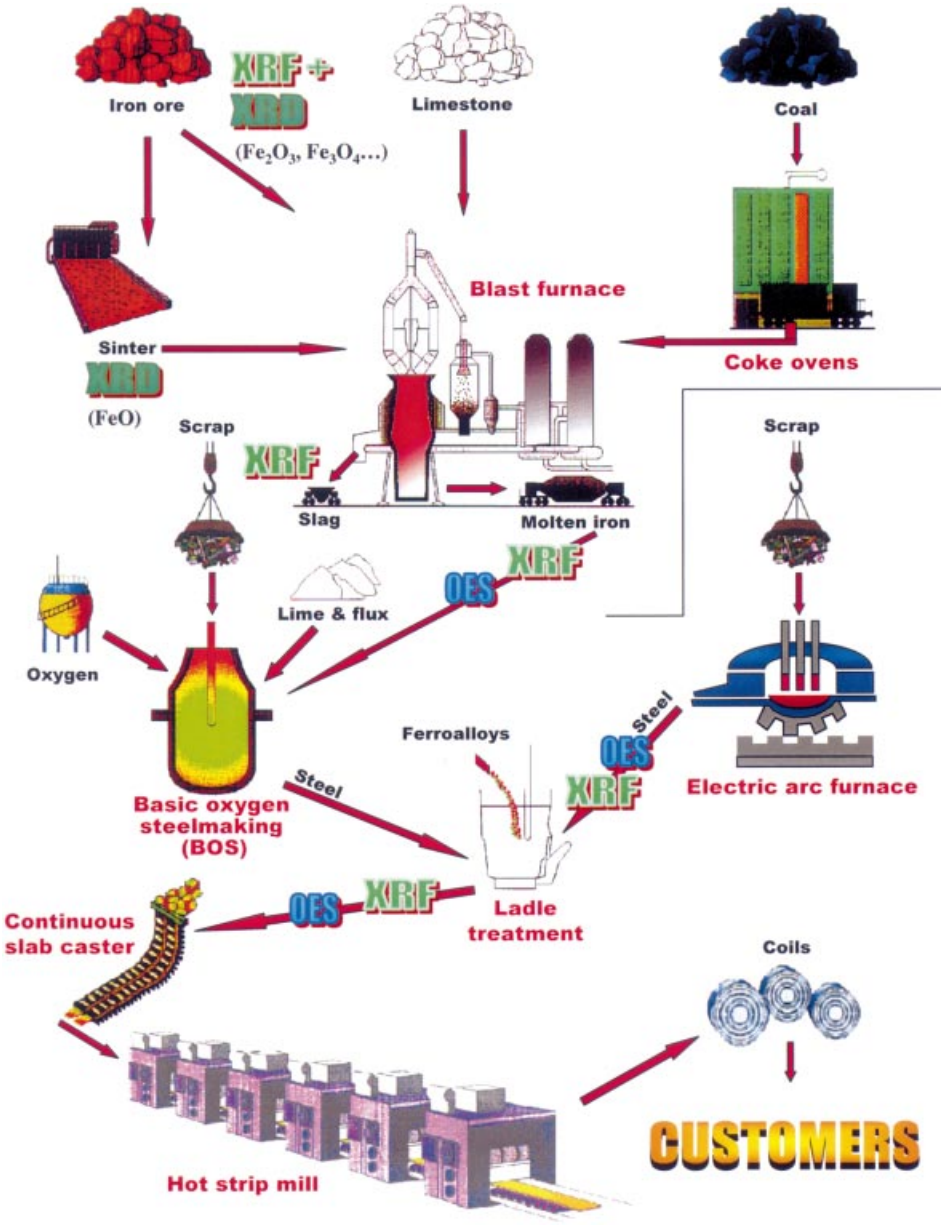

Figure 1 Processes in the iron and steel industry with various demands on analysis.

resolution and low performance with light elements (Na to Cl and particularly Mg, Al, Si and S) and its insufficient precision for major element analyses.

2 SAMPLING

It is well known that the analytical result is only as good as the sample submitted for analysis. All too often the taking of a representative sample is neglected while at the same time emphasis is placed on producing the most accurate analytical result using the most sensitive procedure or instrumentation. The effort put into achieving high-quality results is, of course, totally misplaced if an equivalent effort is not applied to sampling.

In the iron and steel industry, the samples submitted to the laboratory are intended to be representative of a shipload of iron ore, a stockpile of raw materials, intermediate and final products from the steel plant. This represents an enormous reduction in size to derive a laboratory-size sample of about 5 kg from many thousands of tons of original material. It is important that a sampling procedure is implemented which will

minimize the errors associated with sampling these large volumes of material. If possible, samples should be taken from moving streams using automatic cutters that remove samples from the conveyor belt in a controlled and regular manner.

The volume of material collected at the belt must be reduced in a systematic way so that the sample that ultimately reaches the laboratory is homogeneous and fully representative of the original sample.

Tests have shown that a totally automated system that takes and prepares the sample is capable of a precision that is typically twice as good as that of a manual system. The results of an automatic sampling system are compared with a manual system in Table 1.[1] Included in the comparison is the analytical precision obtainable by preparing and analyzing 10 replicate fusion beads. The results emphasize the point that sampling is the limiting factor in the achievement of precise analytical data when analyzing large sample volumes.

The quality of sampling being of the utmost importance with a view to obtaining representative results, International Standardization Organization (ISO) norms have been produced, e.g. ISO 3081, 3082 and 3083 for sampling and sample preparation of iron ores and ISO 14284 for sampling and preparation of steel and iron.

For processed final products such as cast iron and steel, a dip sample is usually taken from the molten

Table 1 Sampling precision trials on iron ores. (Reproduced by permission of Dr Ulrich Senff.)

Component	Average (%)	Automatic SD[a] (2σ) (%)	Manual SD[a] (2σ) (%)	Analytical precision SD[a] (2σ) (%)
Fe	66.12	0.54	1.1	0.2
SiO_2	2.68	0.52	1.22	0.02
Al_2O_3	1.04	0.2	0.44	0.01
Mn	0.12	0.06	0.16	0.01
MgO	0.09	0.04	0.08	0.01
CaO	0.08	0.08	0.12	0.01
TiO_2	0.05	0.02	0.02	0.01
P	0.036	0.004	0.01	0.002
K_2O	0.018	0.008	0.022	0.001
Zn	0.002	0.002	0.002	0.001
LOI[b]	1.16	0.16	0.32	0.06

[a] SD, standard deviation.
[b] LOI, loss on ignition.

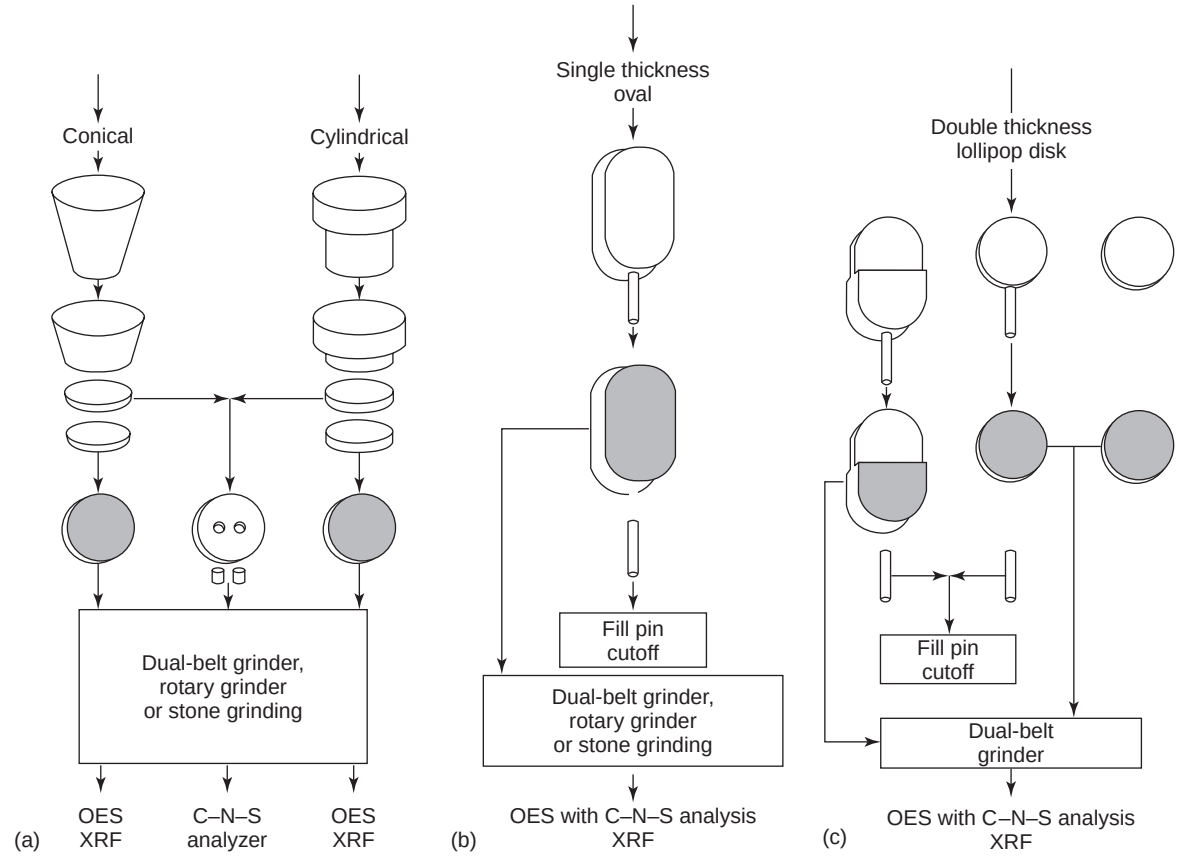

Figure 2 Typical sample shapes used for metal analysis in the iron and steel industry. (Reproduced by permission of Herzog Maschinenfabrik GmbH.)

mass. An appropriate sampling device is used and the molten material is cast into a mold that provides rapid solidification (see **Metal Analysis, Sampling and Sample Preparation in**). The aim of the exercise is to produce a homogeneous and representative sample through a methodology which will ensure as high a reproducibility as possible and promote the formation of a fine-grained metallographic structure of the specimen. For these reasons, the mold casting has nowadays been largely replaced by immersion probes, which allow better control and reproducibility of the conditions in which the samples are taken. The most popular type of immersion probes produce the so-called lollipop, in either oval or disk form, as shown in Figure 2(a–c). They produce samples which will fit into the sample holders of all modern WDXRF spectrometers (maximum diameter 52–60 mm) and which present a flat surface of at least 30 mm diameter as the aperture of the holder for presentation of the sample to the WDXRF spectrometer is typically 28–30 mm in diameter.

3 SPECIMEN PREPARATION

Specimen preparation involves the procedure for preparing the sample in a form that is acceptable for introduction into the XRF spectrometer. It is clear that here again a high reproducibility of the preparation method is essential in order finally to ensure high precision and accuracy of the analyses. As seen above, a minimum flat diameter of 30 mm is generally desired for the analysis.

3.1 Metals

Metal samples include pig irons, cast irons, carbon steels, stainless steels and tool steels. Samples taken at the production point are usually of the correct size and consequently only require surface preparation. Only in the presence of visible defects such as air bubbles, nonmetallic inclusions or cracks will the operator have to take a decision about the validity of the sample.

The surface preparation procedures generally involve the use of abrasive paper either on a rotary grinder or on a belt linisher. Polishing striations give rise to a shielding effect, which results in a decrease in fluorescence intensities, especially for lighter elements, depending on the orientation of the sample. As expected, the decrease in intensity is more important when the XRF analytical devices view the sample perpendicularly to the striations than when they look at the sample parallel to them. For this reason, modern spectrometers are equipped with spinning of the sample holders to smooth out the influence of sample orientation, resulting in reproducible intensities on samples and standards.

Surface preparation is often started with a coarse alumina paper, e.g. 30–80 grit. The final surface finish is obtained by using finer grits, e.g. 120–180 grit, but this additional fine grinding is not practised everywhere. Grinding with alumina paper will, of course, contaminate the surface of the specimen with aluminum. Therefore, accurate aluminum content is generally determined by the OES technique, for which the depth of analysis is larger. Alternatively, silicon carbide abrasive paper can be used, but this will give a similar problem with silicon. Other types of abrasive paper such as zirconium oxide have been shown always to contain some alumina and therefore cannot solve the problem entirely. An interesting alternative is being applied in France on stainless steels where lathe preparation is used, which produces a good quality of surface without contamination.

3.2 Pressed Powders

Pressed powders offer the quickest and simplest form of sample preparation for WDXRF analysis. There are, however, limitations to the procedure and as long as these are recognized the method can be used for the preparation of most of the oxides and ferroalloys associated with the iron and steel industry.

The most serious disadvantages of the pressed powder procedure are particle size effects and mineralogical effects. Particle size effects are most significant in the determination of light elements where long-wavelength XRF radiation is emitted from the upper layers of the specimen and is influenced by differences in particle size. To minimize these effects, the particle size for pressed powders should be $<50\,\mu m$. This is achievable with vibro-rotatory milling equipment. Care should be taken in the choice of mill vessel as contamination of the sample occurs during the milling process. In the iron and steel industry, vessels made of chrome steel are generally used because they are stronger and less expensive than the tungsten carbide (WC) versions. However, being less hard, they will wear quicker and also will contaminate the specimen with iron and chromium. This is why in cases where only minimum contamination can be tolerated WC should be chosen.

When the binding characteristics of the powder are not sufficient, a suitable binder/grinding aid must be added to the sample before milling the mixture. Generally a dilution of one part of binder to nine parts of sample is used. The powder can then be pressed into an aluminum cup or a steel ring.

One of the difficulties in carrying out an XRF calibration for pressed powders is that generally certified

reference materials (CRMs) cannot be used efficiently owing to mineralogical effects. The use of in-house standards of the same material types as those requiring analysis is recommended in order to produce calibration curves which will fit with the production samples.

3.3 Fused Beads

3.3.1 Introduction

Preparation of oxide materials as fused beads based on the original method of Fernand Claisse[2] is the best method of obtaining accurate results. Essentially, the fusion procedure consists of heating a mixture of sample and flux at high temperature (800–1200 °C) with a borate so that the flux melts and dissolves the sample. The overall composition and cooling conditions must be such that the end product after cooling is a one-phase glass. Heating of the sample–flux mixture is usually done in platinum alloy crucibles (e.g. Pt–5% Au), but graphite may also be used when conditions permit. Both particle size effects and mineralogical effects are removed as the original structure of the sample is completely destroyed by the fusion at high temperature and the elements are then embedded in a truly homogeneous matrix. Thus improvement in reproducibility and accuracy can be achieved with this method compared with the pressed pellets method, as shown in Table 2.[1]

3.3.2 Advantages of the Fusion Technique

The fusion technique also has additional advantages:

- Possibility of high or low specimen dilution: the dilution level is often dictated by the ability of the given oxide to be fused completely and without segregation. Lower dilution allows better limits of detection and decreases the preparation errors as a larger amount of sample is weighed.
- Possibility of adding compounds such as heavy absorbers or internal standards to decrease or compensate for matrix effects.
- Standards of desired composition can be prepared from pure oxide materials in order to generate calibration curves.
- CRMs can be used without restriction to generate calibration curves.
- Possibility of obtaining very large calibration ranges within the same calibration curve when the proper matrix corrections are used (Table 3).[3]

Figure 3 shows an example of such a calibration curve for Ca covering a concentration range from 200 ppm to 94%.

Table 2 Comparison of repeatability with blast furnace slag for pressed powder and fusion bead preparation: 10 specimens prepared from the same sample. (Reproduced by permission of Dr Ulrich Senff.)

Component	Concentration (%)	SD (2σ) (%)	
		Fused disk	Pressed powder
CaO	40.7	0.12	0.22
SiO_2	34.8	0.08	0.18
Al_2O_3	14.2	0.06	0.08
MgO	7.2	0.08	0.1
TiO_2	0.88	0.02	0.02
S	0.59	0.02	0.02
K_2O	0.51	0.01	0.02
MnO	0.42	0.01	0.02
FeO	0.39	0.02	0.12
P	0.005	0.002	0.002

Table 3 Calibration ranges achievable with a general calibration for oxides prepared as glass beads (data for ignited samples): the standard error of estimate (SEE) gives a measure of the accuracy achieved with such wide range calibration for oxides. (Reproduced by permission of ARL/Applied Research Laboratories SA.)

Element	Range (%)	Typical SEE (%)
CaO	0.02–94.4	0.21
SiO_2	0.35–99.7	0.17
Fe_2O_3	0.025–94.0	0.15
MgO	0.01–97.3	0.10
Al_2O_3	0.16–89.2	0.11
K_2O	0.006–15.4	0.05
MnO	0.005–8.0	0.04
Cr_2O_3	0.002–17.4	0.03
TiO_2	0.011–3.8	0.03
P_2O_5	0.014–1.0	0.04
SO_3	0.015–3.7	0.05
Na_2O	0.045–10.4	0.05

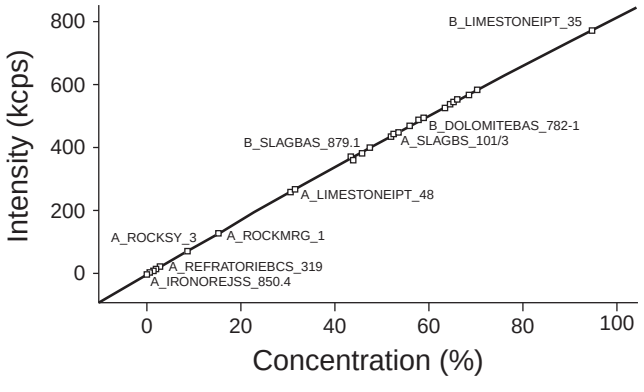

Figure 3 Calibration curve for Ca in various oxide materials using the fused-bead preparation technique. (Reproduced by permission of ARL/Applied Research Laboratories SA.)

3.3.3 Variables in the Method

There are still many decisions that need to be made in order to achieve accurate analyses with the fusion bead procedure.

3.3.3.1 Choice of Flux This is the most important aspect as the flux must be able to dissolve all the different material types that require analysis. Many preferences exist among the various laboratories engaged in major element analysis of oxide materials. The most frequently used fluxes are borates, namely lithium tetraborate, lithium metaborate and sodium tetraborate. Lithium tetraborate is widely used as it is suitable for many applications. Sometimes mixtures of these fluxes are more effective, e.g. fluxes with 12 parts of $Li_2B_4O_7$ to 22 parts of $LiBO_2$[4] or 65% $Li_2B_4O_7$–35% $LiBO_2$.[5]

Nonwetting additives are often used in order to reduce the retention of the melt in the crucible during pouring, reduce any sticking of the bead to the mold and therefore allow cooling without risks of cracking. Additives can be liquids such as solutions of HBr, NH_4I or LiF or solids such as LiBr.

Oxidizing agents, e.g. $LiNO_3$ or $NaNO_3$, are added when some metallic particles occur in the sample in order to save the platinum ware and obtain a homogeneous bead.

Because of all these additions, accurate weighing is essential in order not to introduce serious errors.

3.3.3.2 Fusion Apparatus Types of fusion apparatus range from simple gas burners or muffle furnaces to sophisticated automatic fusion machines with multiple burners or induction coils.

3.3.4 Disadvantages of the Method

The disadvantages of the method are relatively obvious and constitute the reason why pressed powders are still used in many instances:

- time of preparation;
- cost per specimen due to the consumables, e.g. flux, additives and platinum ware;
- loss of volatile elements during the fusion;
- possible errors when weighing flux and sample.

4 CALIBRATION OF THE X-RAY FLUORESCENCE INSTRUMENT

4.1 Introduction

In order to obtain the concentrations of the various elements present in the sample, the XRF spectrometer must be calibrated. This is done by measuring reference materials, either a CRM or a secondary reference material (SRM).

Overlap corrections are necessary when a spectral line of an element interferes with the analyte line. The contribution from the interference line to the intensity of the analyte line must be corrected. Such a correction depends on the sensitivity of the instrument for the given analysis lines, and therefore they have to be determined experimentally, for example by running binary standards.

In addition, it is essential that *matrix corrections* be used in order to achieve the desired accuracy. The correction coefficients can be determined experimentally with a large number of reference materials. However, the common practice nowadays is to generate theoretical α factors through specific software programs which take into account the conditions of excitation (anode type, kV, mA), the geometry of the instrument and the matrix considered. Finally, matrix correction must be applied to the intensities obtained on the XRF instrument through a multivariable regression program in order to obtain the corrected calibration curves.

4.2 Setting-up Samples and Control Samples

In order to maintain the calibration over time, a number of stable samples representing the high and low intensities (not the concentrations) found in the various calibration curves are chosen. They are called setting-up samples. The drift correction, also called recalibration or standardization, needs to be performed at given intervals. Modern high-power WDXRF instruments are very stable in the long term and standardization is often practised with intervals of 1 week or more. Procedures are now used in order to determine at which time the standardization must take place. On-line statistical process control (SPC) software performs the statistical analysis as a background task as soon as a stable control sample has been run. In case of nonstatistical behavior of a given analysis channel, an alarm appears on the instrument screen requesting a selective standardization. SPC allows guaranteeing of the precision and accuracy of the spectrometer, which is required by quality systems such as the ISO 9000 series of standards.

Small, low-power desktop XRF machines, in contrast, require a monitor sample (also called type standard) to be used generally before each analysis in order to correct for possible drift.

It is not necessary that the setting-up samples have the same matrix as the production samples, but it is essential that they are very stable and easy to clean. For example, for slags or sinters, flat, polished, glass–ceramic samples are very adequate whereas for calibration of iron or steel, diamond-polished metallic samples are selected.[6]

Table 4 Metallic materials analyzed by XRF

Type	Typical frequency	Typical response time (min)	Response time requested
Pig iron (blast furnace iron)	90 per day	3–15	3–5 min
Pig iron before desulfurization	30 per day	3–15	Asap[a]
Pig iron after desulfurization	30 per day	3–15	Asap
Steels	150–300 per day	3–6	Asap
Ferroalloys (each type)	1 per batch + 1 per week	15–45	Within the day

[a] As soon as possible.

5 MATERIAL TYPES AND ANALYTICAL REQUIREMENTS

In the iron and steel industry, analyses on materials and products are carried out for two very distinct purposes: control of the process and control of incoming material for invoicing purposes. Analysis requirements are generally expressed in term of limits of detection, of precision at given concentration levels and of accuracy. In view of the high performance of modern WDXRF instruments, accuracy depends more on the quality and number of standard samples used for the calibration and on the matrix and overlap corrections than on the intrinsic qualities of the instrument. This is why instrument manufacturers nowadays utilize numerous CRMs and are able to deliver instruments that are fully factory calibrated.

The materials that are analyzed by WDXRF in the iron and steel industry are metallic (Table 4) or nonmetallic (Table 5). A third category can be defined for the various coatings and other miscellaneous materials (Table 6).

Two parameters allow rating of the importance of a given material within a plant, its frequency of analysis and the desired reporting time. The response time varies as a function of the distance that the specimen has to travel to reach the analytical equipment, the degree of automation involved and the type of XRF spectrometer (sequential, simultaneous or simultaneous/sequential).

Table 5 Nonmetallic materials analyzed by XRF

Type	Typical frequency	Typical response time (min)	Response time requested
Iron ores	5–10 per day	10–30	2–4 h
Sinters	Every 2–4 h	10–60	2–4 h
Blast furnace slags	20–40 per day	6–30	Asap[a]
Converter slags	20–40 per day	6–30	Asap
BOS slags[b]	30–50 per day[c]	6–30	Asap
Ladle slags	20–40 per day	6–30	Asap

[a] As soon as possible.
[b] BOS, basic oxygen steelmaking.
[c] Depending on the process, sometimes none.

Table 6 Miscellaneous material analyzed by XRF

Type	Typical frequency	Response time requested
Bauxite, olivine, limestone, dolomite, serpentine, calcium hydroxide, ilmenite, casting fluxes, mold powder	2 per week	Within the day
Coke ash, coal ash	1 per incoming batch + 10 per day	2–4 h
Refractories	Occasionally	Within the day
Coatings (Zn–Al, Zn–Ni, Zn–Fe, Sn, Cr, plastic)	1 per hour	30 min

5.1 Analysis of Iron Ores

Producing steel from natural material is achieved through reduction of iron ores in a blast furnace or a direct reduced iron (DRI) process as opposed to the alternative method using steel scraps fused in an electric furnace (Figure 1).

Two procedures are available for the preparation of iron ores in a suitable form to allow XRF analysis. The preparation through fusion into glass beads (section 3.3) opens up the possibility of having a standardized method and is described in ISO 9516.[7]

Pressed powder briquettes of iron ore are employed in cases where ultimate accuracy in not required, when the number of samples to be tested is large,[8] when trace elements must be determined or when an additional X-ray diffraction (XRD) analysis is intended on the same specimen.

The XRD technique permits the determination of the compounds or phases present in a sample. However, only limited use has been made of quantitative XRD in process control as such instruments are usually housed in the Central Research Laboratory, well away from the process.

The latest instrumentation incorporating an XRD device into a conventional WDXRF instrument permits

the determination of both the elemental and phase composition of specimens.[9] The determination of the relative proportions of hematite (Fe_2O_3) and magnetite (Fe_3O_4) in iron ore, for example, results in better utilization and optimization of the available mineral resources. This kind of instrument can replace the time-consuming wet chemical methods employed up to now to determine these different phases.

More details on the analysis of iron ores can be found in **Iron Ore, Sample Preparation and Analysis of**.

5.2 Analysis of Sinters

Sinters represent the majority of the ferrous products charged into the blast furnace (around 75% or more). They are made of iron ores, coke and limestone heated to 1200 °C and sintered by aspiration of air through the mixture. This allows substitution of the raw iron ores by a man-made material with desired specifications and keeps the fluctuations of the firing conditions in the furnace as small as possible.[10]

The XRF technique only determines the presence and amount of elements. In the case of sinters and other oxidic materials, some elements are to be reported as oxides rather than elements, e.g. SiO_2, Al_2O_3. Hence the calibration of the XRF instrument for these elements is based on the oxide concentrations so as to obtain the desired oxidic reporting.

Determination of the oxides/elements (Table 7) is carried out by XRF after preparation either as pressed pellets[11] or as fused beads in the same manner as for iron ores.

The pressed pellet method is also desired when using low-power desktop XRF equipment installed at the sinter plant, so as to have a quick and simple preparation. In such a case an analysis is performed every hour and a moving average over three samples is taken, leading to acceptable results.[12]

Table 7 Typical elements or oxides determined in sinters, their concentration ranges and the typical reproducibility of analysis over 20 days using a high-power instrument with X-ray tube settings of 50 kV, 40 mA.[11] (Copyright ASTM. Reprinted with permission.)

Element/oxide	Typical range (%)	Concentration (%)	SD (%)
Fe	30–70	35.71	0.2
CaO	3–35	27.73	0.23
MgO	0.1–15	8.52	0.12
SiO_2	3–10	7.74	0.19
Mn	0.1–6	1.91	0.01
Al_2O_3	0.5–5	1.6	0.12
K_2O	0.01–1.0	0.1	0.003
P	0.01–1.0	0.15	0.006
Na_2O	0.01–1.0	0.03	0.002
S	0.01–1.0	0.07	0.006
Zn	0.01–0.1	Not tested	

One of the most important controls to be applied in sinter is the amount of FeO phase, which allows one to monitor the sintering process and to avoid problems in the blast furnace operation during oxide reduction. XRF does not allow a distinction between the various phases of iron oxide, i.e. Fe_2O_3, Fe_3O_4 and FeO; it only measures the total iron in the sample. Other methods are required for the FeO value, e.g. titration with potassium dichromate or XRD. A recent advance in instrumentation has been the integration of an XRD system into an XRF spectrometer, which allows both elemental and phase determination to be carried out on the same specimen.[9,13] The determination of FeO in sinter is possible with good precision with such an integrated XRD system with a counting time of <100 s.

5.3 Analysis of Slags

Slags originate from various stages in the iron and steel process, e.g. blast furnace, converter, basic oxygen furnace (BOF) (also called BOS), electric arc furnace or ladle.

In the blast furnace, slag is formed from the impurities in the iron ores (known as the gangue), the flux and the coke ashes. It is a complex silicate of aluminum, calcium and magnesium containing small quantities of oxides of manganese and iron and calcium sulfide.[10] Slag has a double role: it permits the removal of the gangue thanks to its fusibility and fluidity, and it allows reactions of exchange with the liquid metal and permits control of the process in order for the desirable elements to remain in the melt while the others are removed. As an example, in an electric arc furnace the slag formation process can be controlled by adding oxygen, carbon and slag formers to the melt. This will promote the formation of CO instead of MnO and FeO and help keep these elements in their metallic form in the melt. The basic slag formers such as lime (CaO) and magnesia (MgO) will help to neutralize the acidity of the slag in order to save the refractory bricks of the furnace.[14]

Slag samples are crushed and ground in a mill to obtain a small enough particle size. Magnetic separation of metallic residues is done on the milled fractions before further preparation. The pressed powder method is generally used for routine elemental determinations in slags, especially when fast reporting is important (converter and ladle slags). High-speed slag analysis is gaining importance as a shorter reporting time allows, for example, optimization of the desulfurization of steel.[15] However, when a certificate must be produced in order to sell the slags or if accuracy is more important than speed of response, the fusion bead method is used. The oxides of interest in the slags are shown in Table 8 with their

Table 8 Typical analytes determined in blast furnace and BOF slags, their concentration ranges and typical performance (pressed pellets, fixed channels, 30-s counting time, 30 kV, 80 mA)

Element/ oxide	Typical range (%)	Performance (%)	
		Precision (2σ)	Accuracy
Al_2O_3	0.1–30	0.03–0.34	0.05–0.3
CaO	15–65	0.1–0.4	0.4–0.7
FeO	0.1–40	0.02–0.37	0.03–0.4
MgO	1–15	0.05–0.2	0.1–0.5
MnO	0.1–25	0.03–0.3	0.03–0.4
P_2O_5	0.01–5	0.007–0.09	0.01–0.2
SiO_2	0.2–40	0.06–0.3	0.04–0.4
TiO_2	0.2–3	0.08–0.1	0.1–0.2

concentration ranges and typical precision and accuracy for pressed pellets.

Other elements oxides are determined in various types of slags are S up to 1.5%, V up to 5%, Ni up to 0.3%, K_2O up to 10%, Na_2O up to 5%, Cr_2O_3 up to 10%, Zn up to 0.1% and F up to 2%.

XRF calibration for slags is relatively straightforward with few interelement corrections to use, especially when a specific calibration is applied for each type of slag. With pressed powder preparation, SRMs originating from the process in question should be preferred. If slags are prepared as fused beads, CRMs and SRMs can be mixed and the same calibration can be used for iron ores, sinters and slags.

5.4 Analysis of Irons

Blast furnace iron, also called pig iron, is brought to the desulfurization station and then to the converter, where it is transformed into steel by oxidation of the excess carbon (Figure 1). At each of these stages samples are taken in order to control the quality of the product (Table 4).

Cast irons are iron–carbon base alloys containing various amounts of silicon, manganese, phosphorus, sulfur and trace elements. Wide variations in properties can be achieved by varying the balance between carbon and silicon, by alloying with various metallic or nonmetallic elements, and by varying casting and heat treatments.[16]

Two types of cast irons are to be considered in order to achieve good precision and accuracy of analysis: white irons and gray irons. Gray cast irons present an inhomogeneous structure with flakes of carbon, which will impair the results especially on phosphorus and sulfur. Therefore, it is recommended to obtain molds and cooling conditions which will produce a white cast iron structure. In addition, all CRMs for cast irons are prepared by chill casting, which promotes the formation of a white iron structure with finely dispersed iron carbides. The sampling is often done with an immersion probe, which delivers a compact coin-like specimen (Figure 2c).[17,18] Obtaining a slag-free specimen is very important as slag would lead to incorrect results for sulfur and silicon.

A proper surface can be achieved with a surface grinder (belt or disk) using an aluminum oxide abrasive paper. Although some referenced methods specify abrasive size up to 240 grit,[19] in practice coarser papers (e.g. 60- or 120-grit) are often selected in a production environment owing to time constraints. Another very effective specimen preparation is by using a 37-grit alumina stone wheel on a horizontal grinder. The specimen is fixed magnetically. Coolant should be used at first when grinding off the casting crust. The final grinding is done without coolant.[17] Fine polishing is only necessary when carbon is to be determined. Diamond polish surface preparation allows the best results in terms of accuracy, but it is time-consuming and seldom used in production control.[1]

The main elements that are to be determined in irons are listed in Table 9. The ranges given correspond also to factory calibrations that XRF instrument manufacturers can typically propose.[6]

Table 9 Typical elements, their analytical lines and their concentration ranges (%) as determined in various types of irons

Element	XRF analytical line	Blast furnace iron (pig iron)	Cast iron, nodular iron	High-alloyed cast iron
C	Kα	4.0–4.5	2.0–4.4	1.3–3.7
Si	Kα	0.3–5	0.4–3.8	0.3–1.8
Mn	Kα	0.1–2	0.009–2	0.2–2.0
P	Kα	0.01–0.1	0.009–1.0	0.03–0.4
S	Kα	0.001–0.17	0.005–0.2	0.007–0.08
Cr	Kα	0.001–1	0.02–2.5	12.0–32.0
Mo	Kα	0.001–1	0.007–1.5	0.5–4.0
Ni	Kα	0.1–1	0.025–3	0.2–16.0
Al	Kα		0.005–0.1	
As	Kβ	0.001–0.025	0.003–0.17	
Bi	Lα		0.002–0.01	
Ce	Lβ		0.003–0.06	
Co	Kα		0.004–0.15	
Cu	Kα	0.001–1	0.01–1.1	0.02–1.3
Mg[a]	Kα		0.003–0.1	
Nb	Kα		0.002–0.5	
Pb	Lα		0.003–0.04	
Sb	Kα		0.003–0.15	
Sn	Kα		0.003–0.42	
Te	Lα			
Ti	Kα	0.001–0.6	0.003–0.25	
V	Kα		0.002–0.6	
W	Lα		0.006–0.1	
Zn	Kα		0.002–0.06	

[a] Nodular iron only.

Calibration of the XRF equipment will require basically the same overlap and matrix corrections as are used for steel analysis (section 5.5).

Some qualities of steel require sulfur contents below 10 ppm and impose the use of iron with very low sulfur concentrations, typically around 20–30 ppm.[20] In that case, the WDXRF instrument must be especially accurate for low levels of sulfur. In addition, the analysis result is required within the shortest possible time, hence the use of simultaneous instruments fitted with fixed channels for the required elements.

The improved determination of carbon in cast iron by WDXRF has become possible as a result of several important developments. A superior light element analyzing crystal (synthetic multilayer structure),[21] better excitation of light elements using thinner X-ray tube windows, higher excitation conditions and closer tube to sample coupling have all contributed to the detection limit of carbon in steels being lowered to 50 ppm (in a 100-s counting time) in the most sensitive instruments. Significant improvements in the detection limits (Tables 18 and 19), as well as in terms of precision (Table 10) for all the other elements normally determined in iron and steel, have also been achieved as a result of these instrumental enhancements. Concerning the determination of carbon and other ultralight elements by XRF, it is important to understand that their characteristic XRF radiations are of low energy and can emerge only from a very thin layer at the surface of the specimen (0.03 μm for carbon in an iron matrix). Hence reproducibility of surface preparation is of the utmost importance in order to achieve a successful analysis of ultralight elements by XRF.

Table 10 Typical instrumental precision obtained on a cast iron sample: 10 replicate analyses of 25 s without preparation of the specimen between runs (30 kV, 80 mA)

Element	XRF analytical line	Concentration level in the specimen (%) (unless specified otherwise)	SD (1σ), 25-s analysis on monochromators (%) (unless specified otherwise)
C	Kα	3.8	0.035
Si	Kα	0.362	0.0008
Mn	Kα	0.12	0.0002
P	Kα	150 ppm	1.6 ppm
S	Kα	250 ppm	1.8 ppm
Cr	Kα	0.28	0.0006
Mo	Kα	0.45	0.0002
Ni	Kα	0.155	0.0004
Al	Kα	0.22	0.0012
Cu	Kα	0.14	0.0004
Nb	Kα	0.55	0.0002
Pb	Lα	100 ppm	3.6 ppm
Sb	Kα	470 ppm	3.6 ppm
Sn	Kα	200 ppm	3 ppm
Ti	Kα	0.12	0.0004
V	Kα	0.24	0.0004

5.5 Analysis of Steels

Examination of the output of large steel companies reveals that they can be producing up to 1500 different steel qualities. Preparation of a specific steel quality takes between 3 and 5 h from initial receipt of the hot metal to final production of the steel quality. During this time, samples are taken at regular intervals to monitor the progress of the process. All these samples require as quick a chemical analysis as possible owing to the short residence times of given charges at each treatment station for cost reasons.[22]

Specimens from all the production stages are nowadays generally transported by a pneumatic tube system into the laboratory, where they are prepared and analyzed. To save costs and increase speed of response, full automation of the whole preparation and analysis is being implemented in many iron and steel plants. This is discussed in **Automation of Analytical Control in the Steel and Metals Industry**.

The analytical steps from sample preparation to transmission of the results must be performed as quickly as possible. Specimens undergo grinding, cleaning and punching prior to analysis. Not more than 180-grit abrasive paper should be used for grinding the specimen surface otherwise variations from operator to operator become too large. During steel making, the carbon level must be controlled and adjusted down to parts per million levels. Low-level carbon, nitrogen and sulfur are measured by combustion analyzers on 1-g slugs, which are punched from the metal after fine grinding (Figure 2a). Alternatively, carbon determination can also be done by OES when the precision and limit of detection (LOD) that can be achieved by XRF are not sufficient. For some or all of the other elements which enter into the composition of the steels (Table 11), both the OES and XRF analysis results are often taken into account. Sometimes an average of the two determinations is used for accuracy at the expense of speed of analysis, as the specimen must be reprepared after having been burned by the spark of the OES spectrometer.

WDXRF is used extensively for the analysis of the high-alloy steels because of the very good precision achieved on major elements (Figure 4a and b). In addition to the elements listed in Table 11, cerium, magnesium, selenium and tellurium are sometimes determined. The typical LODs for two types of XRF instruments can be found in Tables 18 and 19, and precisions at various concentration levels are given in Table 12.

The typical precisions (Table 12) required by the steel industry for elemental determinations can be achieved with counting times ranging from 20 s on

Table 11 Elements, their analytical lines and their concentration (%) range for XRF analysis in various types of steels

Element	Line	Low-alloy steel (free cutting steel)	High-alloy steel	High-speed steel
C	OES[a,b]			0.03–2.0
Si	Kα	0.002–1.7	0.09–4.5	0.1–0.5
Mn	Kα	0.002–2.0	0.25–20	0.08–0.5
P	Kα	0.001–0.1	0.002–0.04	0.01–0.05
S	Kα[b]	0.001–0.08 (0.3)	0.001–0.04	0.002–0.05
Cr	Kα	0.001–5.15	6.6–32.0	2.0–5.0
Mo	Kα	0.001–1.5	0.06–7.0	0.15–9.0
N	OES[a,b]			
Ni	Kα	0.002–4.7	0.1–35.0	0.03–0.45
Al	Kα	0.002–1.2	0.002–1.5	
As	Kβ	0.003–0.17	0.003–0.01	
B	OES[a]			
Ca	Kα	0.001–0.01		
Ce	Lα	0.003–0.1		
Co	Kα	0.003–0.3	0.01–2.1	0.05–18.0
Cu	Kα	0.001–0.70	0.03–3.5	0.01–0.1
Fe	Kα	Rest		
N	OES[a,b]			
Nb	Kα	0.001–0.6	0.005–2.0	
O	OES[a,b]			
Pb	Lα	0.002–0.02 (0.3)		
Sb	Kα	0.003–0.07	0.014–0.18	
Sn	Kα	0.003–0.15	0.003–0.01	
Ta	Lβ	0.003–0.2	0.003–0.05	
Ti	Kα	0.001–0.4	0.003–2.7	
V	Kα	0.001–0.95	0.03–1.5	0.45–3.3
W	Lα	0.003–0.3	0.02–3.1	2.0–22.0
Zn	Kα	0.002–0.025		
Zr	Kα	0.001–0.2		

[a] Analysis done by OES for low levels of given element.
[b] Analysis done alternatively by combustion analyzer.

simultaneous instruments to 300 s on sequential machines (see Table 20).

Table 12 Typical precision (%) required for XRF analysis of steels at different concentration levels

Element	Concentration level (%)				
	0.01	0.1	1	5	20
Si	0.002	0.005		0.03	
Mn	0.002	0.005		0.03	
P	0.001	0.003		0.02	
S	0.001	0.003		0.02	
Cr	0.001	0.005		0.02	0.04
Mo	0.001	0.005		0.02	
Ni	0.001	0.005		0.02	0.04
Al	0.001	0.004	0.01		
As	0.002				
Ca	0.001				
Co	0.002	0.005		0.03	
Cu	0.002	0.005		0.03	
Nb	0.001	0.005	0.01		
Pb	0.002				
Sb	0.002				
Sn	0.002	0.005			
Ta	0.002				
Ti	0.001	0.003	0.01		
V	0.001	0.004		0.02	
W	0.002	0.003		0.03	
Zn	0.001				
Zr	0.001				

Concerning the calibration of the XRF instrument, overlap corrections are necessary notably for the following interfering element–analyte pairs: Mo–P, Cu–P, Mo–S, Ti–V, V–Cr, Cr–Mn, Fe–Co, Ni–Cu, Cu–Zn, W–Zn, Mo–Zn, W–As, W–Ta, Ni–Ta, Cu–Ta, Ni–W and Sn–Pb. These corrections depend on the intrinsic sensitivity of the instrument to the given analytical line, therefore they are determined experimentally using binary standards. Note that iron being the matrix is not reported but is measured in order

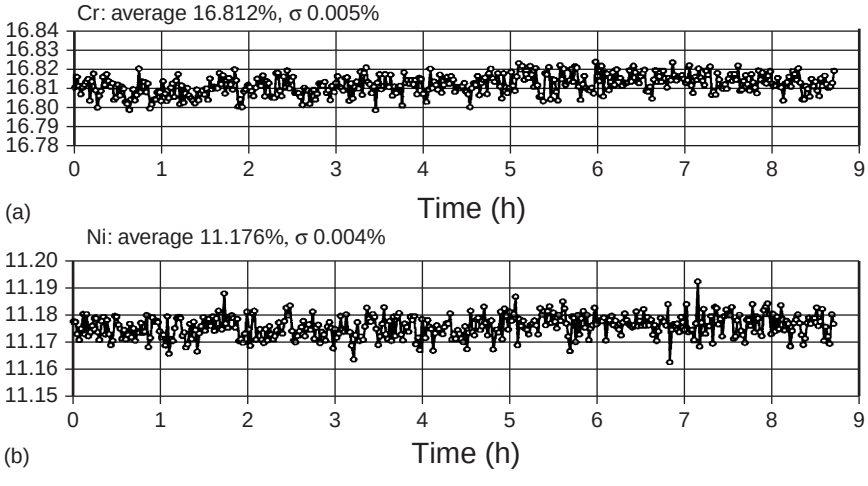

Figure 4 Example of overnight repeatability for (a) Cr and (b) Ni in stainless steel using a state-of-the-art WDXRF instrument; 30-s counting time on fixed channels (monochromators). (Reproduced by permission of ARL/Applied Research Laboratories SA.)

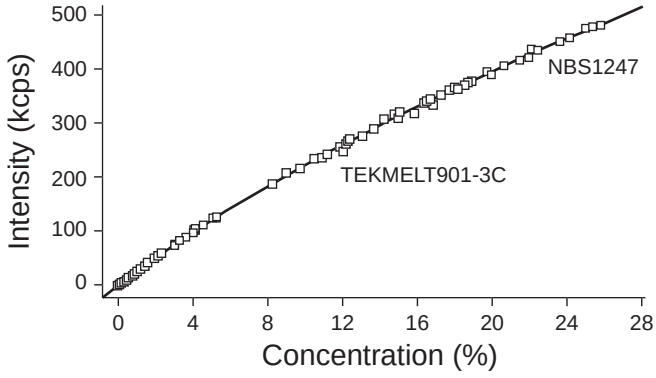

Figure 5 Calibration curve for Cr in steel. A second-order curve is used in order to obtain a wide concentration range on the same curve (from 60 ppm to 25%). (Reproduced by permission of ARL/Applied Research Laboratories SA.)

to be able to make overlap corrections, notably on cobalt.

Matrix corrections are an essential part of the calibration in order to achieve the desired accuracy. The correction coefficients can be determined experimentally with a large number of reference materials (either CRMs or SRM). However, the common practice nowadays is to generate these α factors theoretically through specific software programs which take into account the conditions of excitation (anode, kV, mA), the geometry of the instrument and the matrix considered. With adequate matrix corrections, excellent accuracy over large concentration ranges can be achieved, as shown in Figure 5 in the case of chromium.

A correction for carbon on chromium improves the accuracy for the latter element. The carbon value obtained from another technique such as using a combustion analyzer or OES can be entered either manually or through software connection into the analytical software of the instrument. The latest software developments allow for the final XRF analytical result to be recalculated when the carbon value is obtained from the OES equipment in cases where this analysis is done after the XRF analysis.

5.6 Analysis of Ferroalloys

Ferroalloys are important materials in the specialty steel industry as their addition to molten steel in predetermined quantities produces steels with specific physical and chemical properties. Ferroalloys are analyzed to confirm compliance with their chemical specifications and allow precise additions in the ladle. In addition, this analysis can be used as a control of the supplied materials for invoicing purposes. Analyses of ferroalloys are therefore carried out at least once a week and on each new supply. However, in view of the ISO 9000 certifications of many suppliers, there is a trend to diminish the number of analyses carried out on ferroalloys and to rely on the analysis certificates submitted by the suppliers.

Table 13 Types of ferroalloys employed in the iron and steel industry

Fe–B	Fe–P	Fe–Mn	Fe–Si–Mn
Fe–Cr	Fe–Nb	Fe–Mn–C	Fe–V
Fe–Cr–C	Fe–Ni	Fe–Mo	Fe–W
Fe–Cr–Si	Fe–Ti	Fe–Si	

There are a large variety of ferroalloys (Table 13) and their preparation for XRF analysis will vary depending on the hardness of the material and the demands in terms of analytical precision and accuracy.

5.6.1 Sample Preparation

Preparation of ferroalloys for analysis is not a simple task. Three main procedures can be employed.

5.6.1.1 Remelting with Pure Iron This method allows a metal button to be obtained by remelting the ferroalloy mixed with pure iron chips in a high-frequency furnace. Casting is done through centrifugation of the melted metal. Crucibles of 20 mL are made of refractory material and molds of copper–beryllium are often used. The most important parameters to be controlled are the dilution ratio and the melting and casting times and temperatures. The best results are obtained with dilution ratios of sample to pure iron varying from 3 : 5 to 3 : 17 for the most difficult ones. Melting temperatures are chosen between 1300 and 1900 °C with melting times of about 2 min. The disks obtained look like thick coins and are polished with abrasive paper in the same way as steel samples.

Using the remelting method, the grade element concentration of all the ferroalloys can be determined with tolerances of ± 0.2 and $\pm 0.4\%$ (2σ) and the relative standard deviation (RSD) values for the minor elements are below 5%.[23]

5.6.1.2 Pressed Pellets Ferroalloys are usually extremely hard and difficult to crush or mill, but this method is rapid and inexpensive. The grinding time and grain size (40–70 μm) must be strictly reproduced for the best results. Various binders are used, e.g. 10% methylcellulose, before pressing at 20–40 tons.

RSDs of 0.1–0.15% around the calibration curve can be achieved for the major elements, and between 3 and 5% for minor elements.

As the specimen preparation is the same for all ferroalloys a single analytical program can be built incorporating all of them (Table 14). This is a very useful starting point before refining the program into specific

Table 14 Concentration ranges for a calibration for ferroalloys using pressed powders

Element	Range (%)
Nb	43–68
Si	0.01–91
P (high)	0.2–26
P (low)	0.01–0.3
Ti	0.5–37
Mn	0.1–90
Cr	53–74
V	0.1–81
Mo	59–76
Al	0.6–7.2

calibration curves for each type of ferroalloy in the quest for the best results.

The use of specialized synthetic multilayer crystals and the enhanced sensitivity of modern WDXRF spectrometers allows much improved precisions and LODs to be achieved for elements, which used to be impossible in the past. An element such as boron can now be determined in ferroboron prepared as pressed powder with good precision (Table 15).[24]

5.6.1.3 Fusion Beads In order to fuse ferroalloys with a flux and cast the resultant melt into a mold, the ferroalloy must first be converted into an oxide. Oxidation of ferroalloys by simple heating is possible but the reaction is exothermic and if done directly in a platinum–gold crucible considerable damage to the crucible takes place. Various methods have been proposed[5,23,25–27] using muffle furnace or fusion machines. In some cases $Sr(NO_3)_2$ is used as oxidant, in others $NaNO_3$ with fluxes made of a mixture of $Li_2B_4O_7$ and Na_2CO_3. A recipe for Fe–Mo preparation proposes a flux made of $CaCO_3$, $NaCO_3$ and $Na_2B_4O_7$, while a mixture of only $CaCO_3$ and $Na_2B_4O_7$ is used for Fe–V and Fe–Ti. Dilution can be as high as a 1 : 35 sample to flux ratio, thus diminishing the matrix effects but also the intensities obtained on the XRF instrument, which can lead to difficulties with trace elements.

Table 15 Typical repeatability for boron analysis in ferroboron with WDXRF (conditions: AX20 crystal, 30 kV, 100 mA, 10-s counting time per run; instrument, ARL 9400 sequential spectrometer). (Reproduced by permission of ARL/Applied Research Laboratories SA.)

Run no.	Boron (%)
1	19.34
2	19.38
3	19.34
4	19.36
5	19.31
6	19.39
7	19.43
8	19.41
9	19.38
10	19.46
11	19.34
Average	19.38
SD	0.04

A successful method[28] protects the platinum crucible during oxidation of the ferroalloys by melting a known mass of lithium tetraborate ($Li_2B_4O_7$) and coating the walls of the crucible by rotating it during cooling. The oxidation is completed by chemical reaction with lithium carbonate (Li_2CO_3) and finally the oxidized ferroalloy is dissolved in $Li_2B_4O_7$ flux.

Using an appropriate fusion method, major elements present deviations around the calibration curve of about ±0.1% while RSDs for minor elements are better than 3%.[23]

Dissolution in acids followed by drying and then fusion of the residue with suitable fluxes into a glass disk has been used with success for main-element determinations on a limited number of ferroalloys. However, this procedure has the disadvantage of being time-consuming and labor intensive.[29]

5.7 Miscellaneous Analyses

5.7.1 Surface Analysis

For surface analysis of steel sheets,[30] specimens 50 mm in diameter are punched out of the metal sheet. They are analyzed directly for determination of the levels of nonmetallic elements present on the surface, such as K^+, Na^+, Ca^{2+}, P, S, Si and Cl^-. These can reveal or explain production anomalies or breakdowns. Stains and product appearance also depend on these surface elements. With constant and fast monitoring, remedial action can be taken in good time. Such determinations can be systematic in some cases.

Total surface iron is determined on the metal after pickling, rolling, annealing or degreasing. The specimen is taken by application of an adhesive tape to the sheet. The tape is removed from the sheet and adhered to a cellulose membrane. Adherent iron is measured on the membrane. These tests are performed occasionally in case of problems.

5.7.2 Coating Analysis

Deposition (hot-dipped or electrogalvanic) of different elements on steel sheet has become a common practice

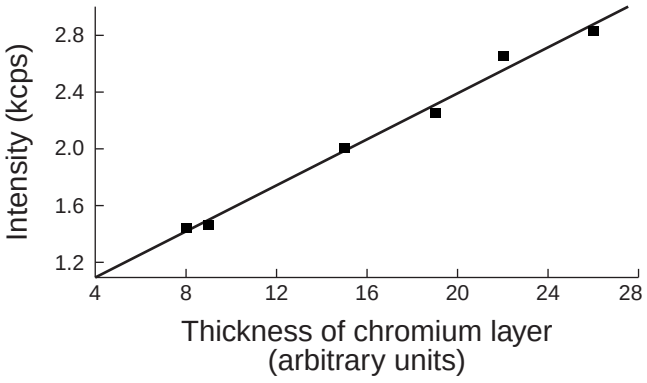

Figure 6 XRF regression curve for chromium layer thickness on a series of steel samples. (Reproduced by permission of ARL/Applied Research Laboratories SA.)

Table 16 XRF calibration results for chromium thickness on steel plates. (Reproduced by permission of ARL/Applied Research Laboratories SA.)

Sample no.	Intensity (kcps)	Chromium thickness (arbitrary units)	
		Nominal	Calculated
01	2.250	19	18.30
02	2.001	15	15.22
03	2.649	22	23.15
07	2.827	26	25.32
08	1.439	8	8.36
09	1.461	9	8.64
SEE			0.8

in order to achieve various properties, e.g. anticorrosion. The layers vary between, among others, Zn, Zn–Al, Zn–Ni, Zn–Fe, Sn, Cr and plastic, depending on the desired property.

XRF can be used to quantify the thickness or surface density of the layer deposited subject to certain conditions. As an example, Figure 6 shows the calibration curve obtained for a chromium layer deposited on a steel plate. Table 16 gives the corresponding numerical results. Other elements in the layer have also been measured. Surface densities down to 20 mg m^{-2} can be measured.[31]

Examples of analysis of other types of coating are shown in Table 17, where the determination of thickness and composition is reported; the term ''Given'' indicates the result obtained by gravimetric analysis (mass thickness) or by atomic absorption (concentration) and ''Found'' indicates the XRF result.[32]

Plastic coatings of different colors are used on steel sheets. Their thickness ranges between 25 and 200 μm and they contain inorganic elements that vary depending on the coating manufacturer and the date of delivery. Analysis by WDXRF provides evidence on claims about poor quality such as color changes or adhesion problems due to insufficient stabilizer or other necessary additives.[33]

5.7.3 Other Analyses

Systematic analyses per incoming batch of materials and weekly analysis are carried out for other types of materials which are used in the iron and steel industry, e.g. bauxite, olivine, dolomite, limestone, calcium hydroxide, casting fluxes and mold powder. Specimens are prepared either as pressed powder or fused beads depending on the plant.

Coal ash and coke ash are analyzed for the determination of elements and oxides such as SiO_2, Al_2O_3, CaO, MgO, Fe, S and P.

Determination of the presence of toxic heavy metals is required in the dust from electro-filters.

6 CHOICE OF SPECTROMETER

Various types of WDXRF spectrometer are employed in the analysis of samples from the iron and steel industry. The correct type of instrument depends very much on the analytical requirements of the individual application, e.g. speed of analysis, number of elements and throughput required. It is worth reviewing the capabilities of these instrument types as they each have specific applications.

Table 17 Analysis results for Sn and Zn coatings on steel plates by XRF.[32] (Reproduced by permission of Koninklijke Philips Electronics NV.)

Sample/side	Given ($g\,m^{-2}$)	Found ($g\,m^{-2}$)	Sample/side	Given	Found
1B/top	2.0	1.7	A6/top ($g\,m^{-2}$)	46	45.7
1B/bottom	2.0	1.7	Fe (%)	13.1	14.4
Sn25/top	11.2	10.6	Al (%)	0.23	0.96
Sn25/bottom	2.8	3.2	A9/bottom ($g\,m^{-2}$)	59.0	56.2
Sn60/top	5.6	5.2	Fe (%)	11.1	11.4
Sn60/bottom	5.6	5.4	Al (%)	0.17	0.8

6.1 Sequential X-ray Fluorescence Spectrometer

Modern sequential XRF spectrometers, with improved tube to sample coupling, high-power generator (up to 4.2 kW) and X-ray tube fitted with a thin beryllium window (down to 50 μm), are capable of producing very high count rates from solid samples. Faster electronics allow linearity of counting on detectors up to 2×10^6 counts per second, making it possible to exploit the better sensitivity of these instruments.

This high sensitivity permits the use of short counting times. In some instances the required precision can be reached with counting times as short as 2 s. Hence it is possible for sequential spectrometers to be used in the determination of up to 12 elements in metal samples, slags or sinters in less than 2 min. The limitation of the sequential spectrometer appears as soon as the number of elements requiring determination increases. Typically low-alloy steels require up to 22 elements, for which complete analysis can only be achieved in about 5 min (see Table 20). This is inadequate when the ladle holding time is 3 min.

LODs close to 1 ppm in a 100-s counting time can be reached for many of the important elements. As can be seen in Table 18, the limits of detection for a 4-s counting time are five times higher, but for many of the elements are still acceptable for the requirements of the iron and steel industry. Evidently the counting time can be selected depending on the element and the performance required.

Table 18 Typical LODs (3σ) for a sequential XRF spectrometer for iron base with different counting times (3-kW X-ray tube excitation). (Reproduced by permission of ARL/Applied Research Laboratories SA.)

Element	LOD (ppm)	
	In 100 s	In 4 s
Aluminum	2.7	13.5
Carbon	60	300
Chromium	2.1	10.5
Copper	2.2	11
Manganese	2.3	11.5
Molybdenum	1.2	6
Niobium	1.4	7
Nickel	1.8	7.5
Phosphorus	1.1	5.5
Lead	5.4	22
Sulfur	1.0	5
Silicon	3.6	18
Tantalum	7.0	35
Titanium	1.4	7
Vanadium	1.0	5
Tungsten	5.0	25
Zinc	2.2	11

In the manufacture of specialty steels, the group of elements requiring determination can change as the demand for the product changes. Thanks to its universal goniometer, the sequential XRF spectrometer is capable of determining up to 84 elements of the periodic table (beryllium to uranium) provided that the appropriate crystals and collimators are fitted. This versatility makes it an ideal choice when the number of elements amounts to about a dozen. In addition, large-capacity sample changers can be fitted. Batches of up to 170 samples can then be analyzed automatically. Mixtures of pressed powder briquettes and glass fusion beads are possible with these large-capacity changers.

6.1.1 Standardless Analysis

Although X-ray instruments in the iron and steel industry are predominantly used for routine analysis of process control samples, there is often a need for flexibility to handle nonroutine samples submitted to the laboratory. For example, new sources of raw materials, coal, fluxes or environmental dust and specialty products may need to be characterized. Difficulties arise when these nonroutine samples do not fit into any of the calibrated programs. The development of the so-called ''standardless'' or semiquantitative analysis packages has met this requirement. Two types of semiquantitative programs have been developed. One employs a global scanning procedure followed by spectral processing while the other measures individual peaks and backgrounds at predetermined spectral positions followed by intensity processing. These packages require a universal goniometer, as fitted in a sequential instrument, and the necessary crystals, collimators and detectors to cover the elements of the periodic table from fluorine to uranium. They are generally calibrated at the XRF manufacturer's plant using standards made as much as possible of a pure element or oxide and sophisticated spectral and matrix corrections.

With such a standardless analysis package, the spectrometer is ready for analysis of a wide range of material types and matrices upon delivery to site. Analysis of unknown samples can commence immediately. Loading of the prepared sample into the XRF spectrometer followed by activation of the program yields a comprehensive determination of all major, minor and trace elements present. Even shift operators with no XRF knowledge can obtain an analysis in 10–15 min. Although the accuracy and precision achieved with these programs are not as good as those obtained with quantitative programs calibrated with matching standards on limited ranges, their capabilities are improving as more efficient matrix correction algorithms, spectral line overlap and background correction models are developed.

6.2 Simultaneous X-ray Fluorescence Spectrometer

In a stainless-steel plant, the turnaround time is critical and analyses for up to 28 elements or more may be required. A high-power XRF spectrometer (up to 4 kW) fitted with fixed-element monochromators will do the job as the elements are measured simultaneously. Each monochromator is generally based on a curved crystal geometry where the XRF from one given element of the specimen is focused into the detector slit. The detector will be selected from a flow-proportional counter (FPC) for the lighter elements to a sealed detector with appropriate gas filling for the light and medium elements. A scintillation counter will be used for the analysis of the shorter wavelengths typical of the Kα or Kβ lines of heavier elements. For light elements, either a curved or a flat crystal geometry can be chosen. Typical counting times for elements in stainless steels are 30–40 s (Table 19), with loading and unloading of the sample taking 20 s, hence a turnaround time of 1 min is easily achievable for metal samples.

The disadvantage of the simultaneous system lies in its inflexibility: no element other than those chosen as monochromators at the time of purchase of the equipment are possible and no backup exists in case of failure of a monochromator.

Table 19 Typical LODs (3σ) obtained on a modern simultaneous WDXRF instrument in a 30-s counting time on the monochromators (3.6 kW)

Element	Line	Crystal	Detector	LOD (ppm)
C	Kα	ML[a]	FPC	170
Si	Kα	PET[b]	Sealed	7.5
Mn	Kα	LiF200	Sealed	5.9
P	Kα	Ge111	Sealed	2.8
S	Kα	Ge111	Sealed	1.5
Cr	Kα	LiF200	Sealed	3.1
Mo	Kα	LiF200	Scintillation	1.9
Ni	Kα	LiF200	Scintillation	6.1
Al	Kα	PET	Sealed	4.9
As	Kβ	LiF200	Scintillation	15.0
Ca	Kα	LiF200	Sealed	5.0
Co	Kα	LiF200	Sealed	10.3
Cu	Kα	LiF200	Scintillation	5.0
Nb	Kα	LiF200	Scintillation	1.4
Pb	Lβ	LiF200	Scintillation	7.1
Sb	Kα	LiF200	Scintillation	9.9
Sn	Kα	LiF200	Scintillation	9.5
Ti	Kα	LiF200	Sealed	5.6
V	Kα	LiF200	Sealed	4.6
W	Kα	LiF200	Scintillation	11.9
Zn	Kα	LiF200	Scintillation	5.0
Zr	Kα	LiF200	Scintillation	1.9

[a] ML, multi-layered synthetic microstructure pseudo-crystal.
[b] PET, pentaerythritol.

Sometimes low-power (200 W) simultaneous instruments are used exclusively for specific control of the oxides, e.g. iron ores, sinters and slags, when the response time is not critical. The disadvantage of these low-power systems is their inability to serve as backup of the OES equipment for the metals analysis.

6.3 Simultaneous/Sequential X-ray Fluorescence Spectrometer

The option to have fixed-element monochromators, which perform an analysis in parallel to a sequential device, is often of considerable benefit in the iron and steel industry. In general, these instruments have a high capacity of fixed channels, e.g. 28 or more, and are fitted with sequential devices, which can be scanners or goniometers. Scanners are limited in angular range and are fitted with only one or two crystals and a single detector. More than one scanner is always necessary to handle all the elements of the periodic table covered by the XRF technique and generally the elements below aluminum ($Z = 13$) are not included. In contrast, a true goniometer as used in sequential XRF instruments is universal. It can be fitted with several crystals and collimators and a minimum of two detectors in order to determine the elements of the periodic table from beryllium to uranium (with some restrictions for the ultralight elements). There are many advantages of having a simultaneous/sequential instrument fitted with a universal goniometer:

- the goniometer can be employed for the quantitative analysis of any elements that are not fitted as fixed monochromators;
- it can back up any of the fixed channels if needed;
- the goniometer can be employed for qualitative determination of any element through wavelength scanning;
- it can be used for ''standardless analysis'' of nonroutine samples (see section 6.1.1).

Table 20 presents the measurement time necessary to achieve the precision values listed in Table 12. Various configurations of instrument are considered:

- simultaneous, fitted only with fixed channels, a counting time of 40 s is used for all channels, fast instrument but without flexibility for any additional element;
- pure sequential, fitted with one universal goniometer, counting time is chosen depending on the element, flexible but slow;
- simultaneous/sequential system with one goniometer and 14 fixed channels, almost as fast as a simultaneous spectrometer and has the flexibility of adding extra elements whenever required.

Table 20 Measurement times (s) required to obtain the precision levels (2σ) generally needed in the iron and steel industry with various configurations of instruments

Element	Instrument type		
	Simultaneous: 22 fixed channels	Simultaneous/ sequential: 14 fixed channels + goniometer	Sequential: 1 goniometer
Si	40	Fixed	8
Mn	40	5	3
P	40	Fixed	14
S	40	3	2
Cr	40	Fixed	12
Mo	40	Fixed	8
Ni	40	Fixed	16
Al	40	Fixed	14
As Kβ	40	Fixed	48
Ca	40	6	4
Co	40	Fixed	12
Cu	40	6	4
Fe	40	2	2
Nb	40	7	6
Pb Lβ	40	Fixed	20
Sb	40	Fixed	30
Sn	40	Fixed	34
Ti	40	Fixed	10
V	40	7	6
W Lα	40	Fixed	18
Zn	40	Fixed	18
Zr	40	8	6
Total time	40	44	295

The complementarity between fixed monochromators and a universal goniometer in a simultaneous/sequential instrument is demonstrated: fixed channels are chosen for the elements requiring long counting times while the goniometer is used for the easier elements.

6.4 Simultaneous/Sequential X-ray Fluorescence Spectrometer with X-ray Diffraction Capability

Given the increasing demand for process integration and total characterization of materials, a wavelength-dispersive spectrometer has been developed which effectively combines WDXRF and XRD within the same instrument. Employing the same pressed powder briquette as prepared for XRF analysis, the integrated XRD system is capable of the quantitative analysis of specific phases or minerals present in the sample. The performance of the XRF monochromators and goniometers being in no way degraded by the addition of the XRD capability, such an instrument can bring some advantages for process control in the iron and steel industry, notably for the determination of iron phases in iron ores and of FeO in sinters and monitoring of the reduction in the DRI process.

6.4.1 The X-ray Fluorescence/X-ray Diffraction System for Direct Reduced Iron Production

New and improved processes for the conversion of iron ore to metallic iron involve direct reduction where iron ore is reduced to iron without going through the pig iron stage. The iron oxide can be in the form of ore, concentrate or pellet and the reduction process uses either gas or coal as the reducing agent. The advantages of this process are low levels of impurities and chemical and physical consistency.[31]

Close monitoring of the progress in the reduction of Fe_2O_3 (hematite) to Fe is important. The determination of the metallic component in the reduced material is possible by wet chemical procedures, but is time-consuming, needing at least 30 min for the analysis of one phase, and therefore cannot be classified as a process control analysis. Because hematite has a distinctive XRD peak, it is possible to monitor the progress of this conversion by observing the decrease in the iron oxide (hematite) concentration in the material. Therefore, the quantitative determination of the hematite content of the ore passing through the system can be achieved using the integrated XRF/XRD instrument.[34]

7 CONCLUSION

As we enter the 21st century, there are various trends in the iron and steel industry as concerns the XRF technique:

- Routine analyses are increasingly carried out by nonexperts. In addition, scientists wish to focus on the details of the chemistry rather than on the particularity of the measurement technique. Therefore, instruments must become more and more autonomous and include the required analytical expertise.
- The WDXRF technique is being used for more varied applications, which call for innovative methods and standardless analysis programs.
- There is a need for more complete information about the sample, e.g. elemental and phase information, which will promote instruments combining different analytical techniques.
- There is a trend to decentralize the measurements and to bring the analytical equipment closer to the process. This means very often that the instruments

are to be fully automated with connection to automatic sample preparation and located in protected cabins.

ACKNOWLEDGMENTS

The author thanks Fernand Hoffert of Sollac Florange, Ray Jowitt of British Steel and Dr Reeb of Dillinger Hüttenwerke for the useful discussions. Thanks are also due to Fritz Baumgartner for his help with the preparation of the manuscript.

ABBREVIATIONS AND ACRONYMS

BOF	Basic Oxygen Furnace
BOS	Basic Oxygen Steelmaking
CRM	Certified Reference Material
DRI	Direct Reduced Iron
EDXRF	Energy-dispersive X-ray Fluorescence
FPC	Flow-proportional Counter
ISO	International Standardization Organization
LOD	Limit of Detection
LOI	Loss on Ignition
ML	Multi-layered Synthetic Microstructure
OES	Optical Emission Spectrometry
PET	Pentaerythritol
RSD	Relative Standard Deviation
SD	Standard Deviation
SEE	Standard Error of Estimate
SPC	Statistical Process Control
SRM	Secondary Reference Material
WDXRF	Wavelength-dispersive X-ray Fluorescence
XRD	X-ray Diffraction
XRF	X-ray Fluorescence

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction • Automation of Analytical Control in the Steel and Metals Industry • Iron Ore, Sample Preparation and Analysis of • Metal Analysis, Sampling and Sample Preparation in

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Sample Preparation for X-ray Fluorescence Analysis • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. U.E. Senff, 'X-ray Fluorescence Spectrometry in the Iron and Steel Industry', presented at the Australian X-ray Analytical Association Conference, 1993.
2. F. Claisse, 'Accurate X-ray Fluorescence Analysis Without Internal Standards', *Norelco Rep.*, **4**, 3–7 (1957).
3. A. Kohler, P. Laederach, K. Juchli, D. Bonvin, R. Yellepeddi, 'Total Oxide X-ray Analyzer', *Int. Ceram.*, **1**, 84–88 (1999).
4. K. Norrish, J. Hutton, 'X-ray Spectrographic Method for the Analysis of a Wide Range of Geological Samples', *Ecochim. Cosmochim. Acta*, **33**, 431–453 (1969).
5. M. Davidts, 'Les Flux du Futur', Proceedings of the Rayons X et Matière Conference, Université de Strasbourg, Strasbourg, C4-683 to C4-691, 1996.
6. *XRF Calibration Specifications*, ARL, Ecublens, Switzerland, 1999
7. ISO 9516, 'Iron Ores – Determination of Silicon, Calcium, Manganese, Aluminum, Titanium, Magnesium, Phosphorus, Sulfur and Potassium – WDXRF Spectrometric Method', ISO, Geneva, 1996.
8. M. Veloso de Castilho, 'Development of a System for the Quality Assurance of Ore Shipments Based on Mapping the Chemical Composition of the Ore in the Mining Area', Proceedings of the CETAS Conference on *Progress of Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 408–416, 1992.
9. R.S. Yellepeddi, 'Ironmaking Using an XRF/XRD Spectrometer', *Adv. Steel*, July, 39–43 (1997).
10. J. Barralis, G. Maeder, *Précis Métallurgie: Elaboration, Structures–Propriétés, Normalisation*, Nathan, Paris, 1–7, 1997.
11. J. Martin, 'Suggested Method for X-ray Emission Spectrometric Analysis of Sinters', Method E-1 SM 10-39, ASTM, Philadelphia, PA, 1992.
12. J.C. Hoet, 'Sinter Plant Autocontrol by Using Low Power X-ray Spectrometer', presented at the 1998 CETAS Conference, Luxembourg.
13. T. Kitaide, T. Sugihara, 'Analysis of FeO in Sinters by Integrated X-ray Spectrometer Combining the Virtues of XRF and XRD', presented at the 1998 CETAS Conference, Luxembourg.
14. E. Uitbeijerse, 'Slag Formation and Analysis by ED-XRF Spectroscopy', *Int. Lab. News*, **26**, 23–24 (1996).
15. N. Bannenberg, 'Requirements of the Production Engineer in View of Analytical Chemistry', presented at the 1998 CETAS Conference, Luxembourg.
16. D. Stefanescu, 'Classification of Ferrous Casting Alloys', in *Metals Handbook*, ASM, Materials Park, Ohio, Vol. 15, 627–629, 1988.
17. V. Buhrke, R. Jenkins, D. Smith, *A Practical Guide for the Preparation of Specimens for XRF and XRD Analysis*, Wiley-VCH, Weinheim, 93–96, 1998.

18. ISO 14284, 'Steel and Iron – Sampling and Preparation of Samples for the Determination of Chemical Composition', ISO, Geneva, 1996.
19. ASTM Norm E322, ASTM, Philadelphia, PA, 1982.
20. P. Delaurie, D. Mutez, G. Butin, 'Automatic Analysis of Iron at Sollac Dunkerque Steelmaking Plant', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 311–318, 1992.
21. D. Bonvin, B. Adamson, K. Juchli, 'The Use of X-ray Fluorescence Spectrometry as an Analytical Technique in the Iron and Steel Industry', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 519–525, 1992.
22. W. Marschinke, H.-J. von Wachtendonk, 'The New Automated Laboratory at Thyssen Stahl, Ruhrort Works', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 264–274, 1996.
23. A. Gomez Coedo, J. Jimenez Seco, E.M. Lopez Rodriguez, 'Development of Methodologies for Instrumental Analysis of Ferroalloys', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 157–164, 1992.
24. D. Bonvin, K. Juchli, 'New Developments in XRF', presented at Anwendertreffen Röntgenfluoreszenz- und Funkenemissionsspektometrie Dortmund, 1999.
25. J. Pétin, A. Wagner, F. Bentz, 'Combination of Oxidation and Melt Treatment for Metal Sample Preparation for XRF Analysis', *Steel Res.*, **56**(4), 215–218 (1985).
26. H.L. Giles, G.M. Holmes, 'The X-ray Fluorescence Analysis of Ferroniobium by a Fusion Method', *X-ray Spectrom.*, **7**(1), 2–4 (1978).
27. G. Staats, 'Study of Ferroalloys with X-ray Fluorescence Analysis', *Arch. Eisenhüttenwes.*, **45**(10), 693–697.
28. R.T. Rutherford, 'Ferroalloy Analysis by X-ray Fluorescence Spectrometry', *X-ray Spectrom.*, **24**, (1993).
29. M.A. Guerrero, E. Andres, 'New Procedure of Preparation and Analysis of Ferroalloys with Additional Advantages Over Current Methods', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxemburg, 594–597, 1996.
30. C. Schwendimann, F. Hoffert, 'Rapid Analysis of the Surface of Coated and Uncoated Flat Products by GDS and XRF', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 125–131, 1992.
31. R. Yellepeddi, A. Bapst, D. Bonvin, B.W. Adamson, 'Process and Quality Control in the Iron and Steel Industry Using the ARL Total Iron X-ray Analyzer', ARL Lecture Transcript LT/XRF/0395, 1995.
32. H. de Koning, 'Speed and Flexibility of XRFS in the Steel Industry', Presented at the Ustron Conference on Analytical Spectroscopy (PL), Ustron, 1998.
33. J. Viiret, I. Ala-Vainio, 'Identification of Plastic-coated Steel Sheet by X-ray Fluorescence Spectrometry', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 210–215, 1996.
34. R. Yellepeddi, D. Bonvin, 'Process Integration of Modern X-ray Spectrometers: Future Perspectives in the Steel and Metal Industries', Proceedings of the CETAS Conference on *Progress in Analytical Chemistry in the Iron and Steel Industry*, Luxembourg, 700–707, 1999.

Surfaces: Introduction

G.E. McGuire
MCNC, Research Triangle Park, USA

When selecting a technique to characterize a material, one of the key factors to be considered is the region and volume of material to be analyzed. When the surface is selected for analysis, it presents some of the most challenging aspects of analytical chemistry. Compared to other types of analysis, the volume of material analyzed is extremely small. When the depth of analysis is only a few ångstroms deep and the spatial resolution is on the order of 500 Å, the number of atoms available for analysis may be below 1000. Impurity atoms that may be present in the parts per million or parts per billion level in this region typically fall below the detection limits of most surface analytical techniques. This is precisely the situation that one faces in trying to analyze the dopant concentration in the junction region of an advanced semiconductor device, for example. There is a tendency to want to increase the current density of the probe beam to enhance the signal-to-noise; however, this may lead to specimen damage in the region of analysis. This may be a result of localized heating, beam-stimulated desorption, bond breaking and chemical decomposition. In spite of these potential problems, surface analysis is accomplished routinely in many applications. There is an ongoing challenge to improve the sensitivity and spatial resolution of the techniques.

When is surface analysis important? It is widely recognized that surfaces are important in many technical areas. Surface reactions are involved in catalytic reactions; friction and wear are influenced by molecular interactions at surfaces; interfaces influence the electrical properties of semiconductor devices; and interfaces control the interactions between biological systems and man-made implants. Many other examples could be given, since the list of situations where surfaces play an important role is quite long. The range of materials that require surface analysis is quite broad, including metals, glasses, organics, biomaterials, semiconductors, etc. A wide diversity of surface analysis techniques have been developed, which that are capable of analyzing the many different types of materials and provide the range of information that is sought regarding the composition, microstructure and physical properties of specimens. These techniques are often complementary, such that two or more of them may be used in order to gain a thorough understanding of a surface.

The analysis techniques can be grouped in many different ways; however, one convenient way of examining them for comparison is the type of information they provide. For example, several of the techniques reviewed in the Surface Analysis section provide images of the surface. A scanning probe is one of the principal methods of generating spatially resolved images. Techniques that utilize scanned probes are discussed in the article **Scanning Probe Microscopy, Industrial Applications of**. Surface microstructure is important in understanding many of the interactions that occur at surfaces. One of the most broadly used surface imaging techniques is the scanning electron microscope (SEM), which generates surface images from secondary electrons produced by high-energy primary electron beam bombardment. This technique has been available for more than forty years and is discussed in the article **Scanning Electron Microscopy in Analysis of Surfaces**. However, steady improvement in the spatial resolution of SEMs has resulted in instruments that image areas of less than 15 Å. The versatility of the sample chamber and stage allows the analysis of specimens up to 300 mm in diameter. In comparison, scanning probe techniques (discussed in the article **Proximal Probe Techniques**), such as scanning tunneling microscopy (STM) reviewed in the article **Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces**, atomic force microscopy (AFM), and others, achieve atomic resolution but generally over a small field of view, on the order of hundreds of ångstroms. As a result, detailed information is revealed about surfaces that has heretofore not been available. This has increased our understanding of surfaces and surface interactions at a pace that was not anticipated just a few years ago. The additional information available from the scanning probes has opened up many new areas of research. The early efforts in STM and AFM were directed toward understanding the techniques and improving the instrumentation. The emphasis has now shifted toward application, as is illustrated in the articles **Scanning Tunneling Microscopy, In Situ, Electrochemical** in the section *Electroanalytical Methods* and **Atomic Force Microscopy in Analysis of Polymers** in the section *Polymers and Rubbers*. However, more traditional techniques such as SEM are just as popular as ever and are used in a wide diversity of fields, such as forensics, which is discussed in the article **Scanning Electron Microscopy in Forensic Science** in the section about *Forensic Science*. In one sense, the scanning probe techniques have become the instruments of choice for research. They provide detailed information on the surface, but a well-controlled surface is required. Many of the studies utilizing scanning probe techniques are conducted in vacuum or controlled ambients. Atomic level control is not necessary for many technologies where the surface plays an important role. For these surfaces the more versatile SEM may be a more appropriate characterization tool.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Information beyond the initial surface is essential for many investigations. The near surface and interfaces play an important role in many technologies. One method of probing beyond the outer surface is to utilize high-energy electron beams that penetrate through a specimen. The image is captured on photographic film or an electronic imaging device such as a charge-coupled device (CCD) camera. The high-energy primary beam may be scanned (optional) as in the SEM. Images of specimens up to several thousand ångstroms thick may be obtained with a spatial resolution below 2 Å. This is discussed in the article **Electron Microscopy and Scanning Microanalysis**. The high-energy probe is the basis for other complementary techniques such as electron energy loss spectroscopy (EELS). As the electrons pass through the specimen they lose discrete amounts of energy through the excitation of electrons of the atoms along their path. The energy loss can be measured using an electron energy analyzer, and can be used as a means to identify what elements are present. It is also possible to do EELS from the front surface of a specimen, using reflected low-energy electrons. This approach is discussed in the article **Electron Energy Loss Spectroscopy in Analysis of Surfaces**.

Chemical composition is another important aspect of surfaces that must be characterized. Energetic beams, such as electrons, ions and X-rays, are used to probe the surface. By exciting secondary radiation or particles, or by examining the energy lost during the excitation of the specimen, one can determine the elemental composition. Techniques such as Auger electron spectroscopy (AES) or X-ray photoelectron spectroscopy (XPS) use energetic electrons and X-rays, respectively, to excite secondary electrons. The Auger and photoelectron have characteristic energies which are utilized for elemental identification, while the signal intensity is an indication of the concentration. What distinguishes AES and XPS from other analytical techniques that rely upon high-energy electrons or X-rays for analysis is that the mean free path of the secondary electrons is very short, providing a surface analysis of the top few monolayers of material. The article **X-ray Photoelectron and Auger Electron Spectroscopy** in the section *X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy* reviews the fundamentals of these techniques, while the articles **X-ray Photoelectron Spectroscopy in Analysis of Surfaces** and **Auger Electron Spectroscopy in Analysis of Surfaces** review some of their applications. Applications that utilize the enhanced energy resolution when soft X-rays from a synchrotron are used for excitation are discussed in the article **Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces**. Likewise, surface analysis utilizing ion beams, such as in secondary ion mass spectrometry (SIMS) or ion scattering spectroscopy (ISS), relies on the limited path length of scattered ions or the secondary ions produced by energetic ion bombardment. The secondary or scattered ions provide elemental analysis of the surface. The sensitivity of the various techniques varies by several orders of magnitude and is usually best for the ion sputtering techniques such as SIMS, which has a sensitivity down to the parts per million level or less in most specimens. The application of SIMS is discussed in the article **Secondary Ion Mass Spectrometry as Related to Surface Analysis** in the section about *Mass Spectrometry*, while the application of ISS is discussed in **Ion Scattering Spectroscopy in Analysis of Surfaces**.

There are many other physical properties of surfaces that must be characterized. Optical techniques are a class of analysis techniques that provide some of this information. For example, optical techniques provide information on the thickness, composition and band gap of materials, information that is generally not available by other analysis techniques. For example, ellipsometry is utilized by the semiconductor industry to determine the thickness and separation of dielectric layers such as oxides and nitrides even when present in multiple layers. This can be accomplished in a nondestructive fashion, which makes it attractive for research as well as production applications. Film thicknesses as small as a few monolayers can be determined. This is discussed in the article **Ellipsometry in Analysis of Surfaces and Thin Films**. Differences in the optical properties of surfaces, such as in differential reflectance spectroscopy, reviewed in the article **Differential Reflectance Spectroscopy in Analysis of Surfaces**, can highlight subtle differences in surfaces, even though information on the exact compositional difference is not provided. In addition, optical techniques such as photoluminescence (PL) can be utilized to measure impurity concentrations down to the parts per trillion level, and defect concentrations in monocrystalline materials at depths of up to one micrometer. PL is reviewed in the article **Photoluminescence in Analysis of Surfaces and Interfaces**. Optical techniques also probe the chemical bonds formed at surfaces. This is accomplished using infrared and Raman spectroscopy which are discussed in the articles **Infrared and Raman Spectroscopy in Analysis of Surfaces** and **Infrared and Raman Spectroscopy and Imaging in Coatings Analysis**.

There are over one hundred surface analysis techniques. Some of the most widely utilized surface analysis techniques are described in this section. The most popular ones are more general in nature and can be utilized on a wide range of materials and specimens. All of the surface analysis techniques have a special niche for which they are best utilized.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
CCD	Charge-coupled Device
EELS	Electron Energy Loss Spectroscopy
ISS	Ion Scattering Spectroscopy
PL	Photoluminescence
SEM	Scanning Electron Microscope
SIMS	Secondary Ion Mass Spectrometry
STM	Scanning Tunneling Microscopy
XPS	X-ray Photoelectron Spectroscopy

Auger Electron Spectroscopy in Analysis of Surfaces

John M. Lannon, Jr
MCNC, Research Triangle Park, USA

Charter D. Stinespring
West Virginia University, Morgantown, USA

1 Introduction **1**
2 Fundamentals **2**
2.1 Auger Emission 2
2.2 Ionization and Auger Transition Probabilities 3
2.3 The Analysis Volume 4
3 Instrumental Methods **5**
3.1 Electron Excitation 5
3.2 Electron Energy Analysis 5
3.3 Vacuum Requirements 7
3.4 Signal-to-noise Factors 7
3.5 Point Analysis and Auger Mapping 8
3.6 Sample Preparation 9
3.7 Depth Profiling 9
3.8 Instrumentation Vendors and Systems 9
4 Data Analysis **10**
4.1 Elemental Identification 10
4.2 Quantification of Data 10
4.3 Calibration Using External Standards 11
4.4 Extraction of Chemical Bonding Information 11
5 Applications **11**
5.1 Basic Surface Science 12
5.2 Materials Science 12
Acknowledgments **13**
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

Auger electron spectroscopy (AES) is an electron energy spectroscopy which probes the electronic energy levels of atoms, using electron or X-ray beam stimulation. Irradiation of a sample with one of these stimuli causes electron emissions which can be energy analyzed to obtain an energy distribution. Since the energies of the Auger electrons are characteristic of the atomic core and/or valence level energies, they contain chemical information about their source atoms. As a result, the recorded energy spectra can be analyzed to determine the atomic composition (atomic percent) of the sample, and, in some cases, extract chemical bonding information. Further, since the escape depth of Auger electrons is limited to the first few monolayers of the sample, the technique has a high surface sensitivity. When used in conjunction with an ion source, AES can provide compositional information as a function of depth (depth profiling). Although AES was initially used primarily for fundamental surface science applications, the technique has developed into an important analytical tool for a broad range of surface and materials science applications, including surface reaction studies, surface segregation studies, and thin film growth studies.

1 INTRODUCTION

AES is an extremely useful analysis technique used to identify and quantify the elements present on solid surfaces. The AES sampling depth is on the order of 0.5 to 2 nm, while the detection limits are approximately 0.25 atomic percent. All elements except hydrogen and helium can be detected, and in many cases, information on the chemical bonding can be extracted from AES data.

AES probes the electron energy level of ions in a solid undergoing auto-ionization and is a subset of a broader field called secondary electron spectroscopy. The initial ions may be produced by X-ray irradiation or most commonly by a primary electron beam of energy E_p in the range of 1–10 keV. The electrons emitted or backscattered from the surface are analyzed as a function of their kinetic energy E, and the resulting electron energy distribution $N(E)$, or its first derivative, $dN(E)/dE$, is referred to as the spectrum.

Auger electrons, which are a component of the secondary electron spectrum, were first discovered by Pierre Auger in 1925.[1] The initial suggestion that these electrons could be used for analytical purposes was made in 1953 by Lander.[2] It was not until 1968 that Harris demonstrated the utility of the technique.[3] The period between 1967 and 1969 was marked by a rapid evolution of electron energy analysis instrumentation, first by Weber and Peria[4] and later by Palmberg.[5] Palmberg also introduced the idea of combining AES and inert ion etching to obtain compositional depth profiles.[5] Finally, in 1970, MacDonald used a focused electron beam rastered across the surface to obtain two-dimensional compositional analysis of the surface.[6] This gave rise to a subset of AES known as scanning Auger microscopy (SAM). Although improvements in hardware and data analysis tools have been introduced over the intervening years, the basic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

AES and SAM analysis tools used today were developed in the span of a few years between 1967 and 1970.

The remainder of this article is designed to provide a basic understanding of AES, the equipment required to implement the technique, and the type of information it can provide. Section 2 discusses the fundamentals of the Auger process. Section 3 describes the instrumental methods used to obtain AES spectra. Section 4 reviews the interpretation of AES data. Section 5 provides an overview of several of the major applications of AES.

2 FUNDAMENTALS

AES involves ionization of atomic core levels, the subsequent relaxation of the excited atom by the emission of a so-called Auger electron (auto-ionization), and the escape of the Auger electron into the vacuum (where the electron's kinetic energy is measured by an electron spectrometer). Either an incident photon or an electron beam may stimulate the initial ionization process. Thus, Auger electrons are commonly encountered in X-ray photoelectron spectroscopy (XPS). For this discussion, however, we will focus primarily on electron beam induced Auger electron emission.

When a primary electron beam of energy E_p, typically in the range of 1 to 10 keV, irradiates a solid surface, it produces secondary and backscattered electrons with an energy distribution $N(E)$. A typical $N(E)$ versus E plot or spectrum for these secondary electrons is shown in Figure 1. This spectrum is comprised of four distinct energy regimes. The peak at $E = E_p$ is due to the elastically backscattered primary electrons. These electrons may be used for low-energy electron diffraction (LEED) for sufficiently low primary energies and for reflection high-energy electron diffraction (RHEED) for sufficiently high primary energies.

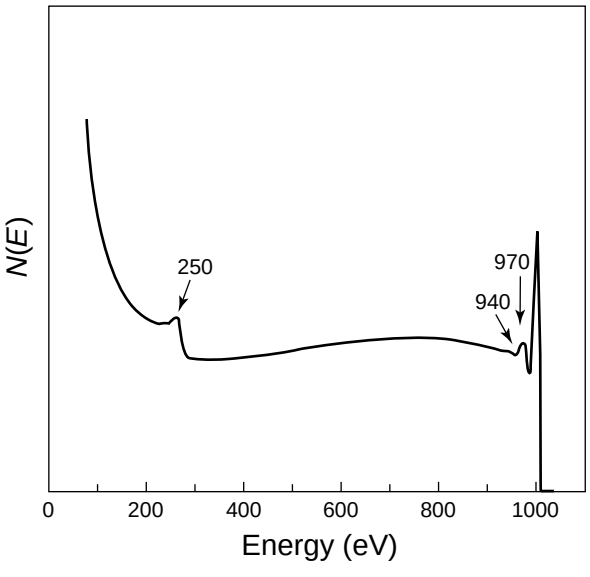

Figure 1 The secondary electron spectrum, $N(E)$ vs E, showing the distribution of electron energies for silicon produced by a primary electron beam having $E_p = 1000$ eV.

Features in the energy range $E_p - 50\,\text{eV} \leq E \leq E_p$ (or plasmon loss region) are associated with primary electrons which are inelastically scattered by plasmons. These electrons are studied in plasmon loss spectroscopy (PLS), electron energy-loss spectroscopy (EELS), and high-resolution electron energy-loss spectroscopy (HREELS).

At the opposite end of the spectrum in the energy range $0\,\text{eV} \leq E \leq 50\,\text{eV}$, the rapidly rising background is due to "true" secondary electrons. This feature is associated with electrons that have lost nearly all of their energy through inelastic collisions, and its intensity and cut-off energy are very sensitive to small changes in the work function.[7]

Peaks in the energy range $30\,\text{eV} \leq E \leq E_p - 50\,\text{eV}$ include those due to Auger electrons, plasmon loss features associated with the Auger electrons, ionization loss electrons (due to the initial ionization event), and diffraction features.[7] It is the Auger electrons in this region to which attention is now directed.

2.1 Auger Emission

Figure 2 illustrates the basic physics of the Auger emission process. The atomic levels shown here are identified using both standard spectroscopic and X-ray notation. As discussed by Briggs and Riviere,[8] it is the latter that is used most commonly in identifying Auger spectra. When a sample is irradiated by a primary electron beam having energy E_p, electronic levels with ionization potentials less than E_p may be ionized. In the case illustrated here, this is the K level (or $n = 1$ level). The excited atom quickly relaxes as an electron from a higher level (L_1 or $n = 2$ level in this case) fills the core hole. This may be accompanied by either the emission of a characteristic X-ray having an energy $E_K - E_{L_1}$ or by the radiationless emission of a second electron (L_2 in this case) having a kinetic energy $E_K - E_{L_1} - E_{L_{2^*}}$. The notation L_{2^*} is used to indicate that this is not the energy of the L_2 level in the ground-state atom, but rather, it is the energy of the L_2 level in the excited atom having a hole in the L_1 level.

The relaxation process involving X-ray emission is called X-ray fluorescence (XRF), while using X-ray notation is called Auger emission. These are competitive processes whose relative probabilities are illustrated in Figure 3 for an initial core hole in the K level. Similar considerations are true for all other initial core hole possibilities (L, M, N, etc.). The Auger electron emitted in the situation depicted in Figure 2 is identified as the KL_1L_2 Auger transition. Other possible Auger transitions

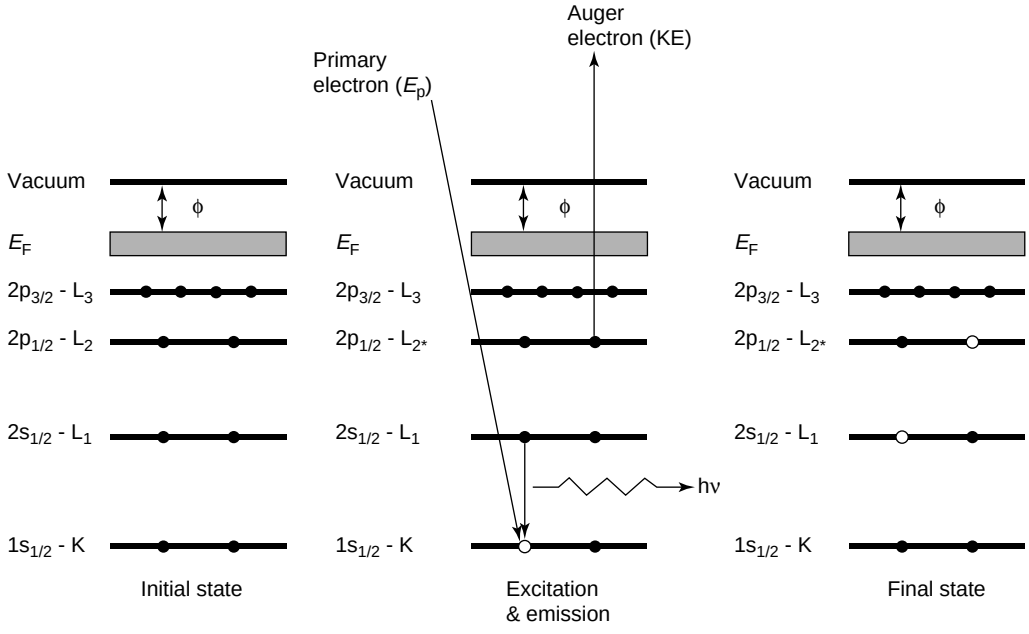

Figure 2 Schematic diagram showing the Auger emission process. In this figure the L_2 and L_3 levels are split. For situations where there is no splitting the combined level is referred to as $L_{2,3}$.

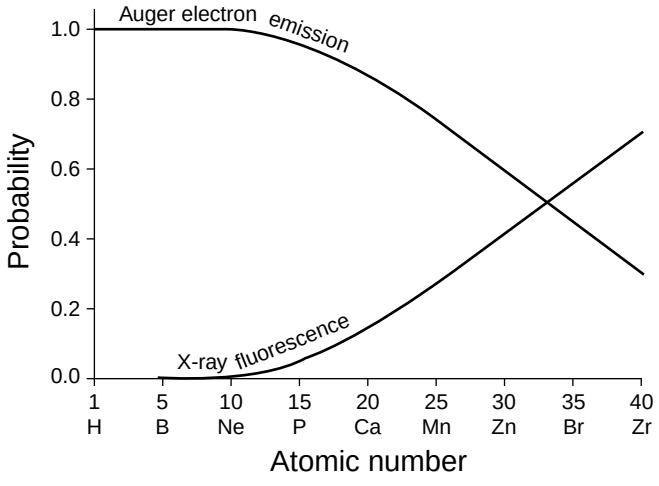

Figure 3 The probability of core hole relaxation by either XRF or Auger emission, plotted as a function of atomic number. (Reproduced with permission from D. Briggs, J.C. Riviere *Practical Surface Analysis: Volume 1: Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, © 1990 John Wiley & Sons Limited.)

involving the illustrated levels include KL_1L_1, KL_1L_3, KL_2L_2, KL_2L_3, KL_3L_3.

It has been noted that it is not possible for H or He to emit Auger electrons. Clearly, this is because the Auger process involves at least two different states and two electrons in states higher than the initial core hole. Likewise an isolated Li atom cannot emit an Auger electron, but Li in the solid state may do this by taking advantage of additional electrons in the valence level.

Following the empirical method developed by Chung and Jenkins,[9] the kinetic energy of the Auger transition ABC from an atom with atomic number Z is given by Equation (1)

$$E_{ABC}(Z) = E_A(Z) - \tfrac{1}{2}[E_B(Z) + E_B(Z+1)] - \tfrac{1}{2}[E_C(Z) + E_C(Z+1)] \quad (1)$$

where $E_i(Z)$ is the binding energy of the ith level in the element of atomic number Z and $E_i(Z+1)$ is the energy of the corresponding level in the element of atomic number $Z+1$. In this formulation, taking the averages of energies E_B and E_C for the elements Z and $Z+1$ is an attempt to account for the fact that these energies actually relate to the core level ions of element Z. A more accurate approach is described by Briggs and Riviere,[8] in which relaxation of the electron is taken into account, as is the interaction between the two final state holes.

Although these computational methods could be used to identify observed Auger transitions, compilations of Auger line positions are generally used for this purpose. Using data such as this, peak positions determined by the maximum negative excursion in the differentiated $(dN(E)/dE)$ spectrum provide sufficient information for the identification of all elements except H and He.

2.2 Ionization and Auger Transition Probabilities

As discussed later in section 3.4, two of the key factors in determining the intensity of a given Auger transition are the ionization cross-section ϕ_i and the Auger transition

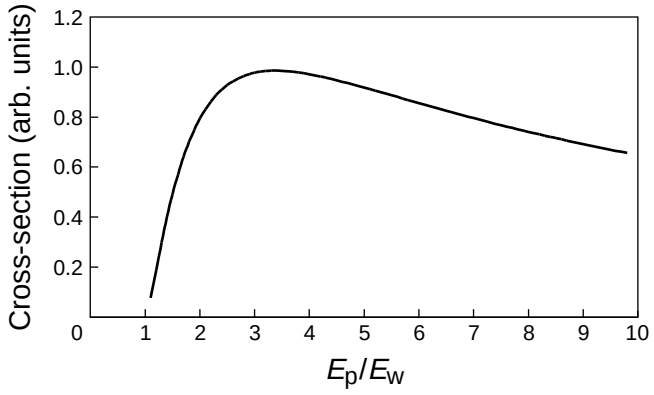

Figure 4 The variation of the ionization cross-section with the ratio E_p/E_w based upon the theoretical model developed by Gryzinsky.

probability ψ_i. The product of these two terms gives a measure of the probability that the *i*th level of an atom will be ionized and that, once ionized, an Auger electron will be emitted. The Auger transition probability has already been considered in the context of its relationship to the XRF probability and is illustrated in Figure 3. The ionization probability will now be discussed.

Figure 4 shows a plot of the ionization cross-section as a function of the ratio E_p/E_w where E_p and E_w are, respectively, the primary electron beam energy and the ionization energy of the initial level.[10] Although $E_p/E_w > 1$ is required, it is clear that $E_p/E_w \sim 3$ is most efficient at producing ions.

In addition to the primary electron beam, many secondary electrons have sufficient energy to ionize atoms within the solid. This has the overall effect of enhancing ionization, particularly for surface atoms and levels with low ionization energies.

2.3 The Analysis Volume

A critical requirement in the interpretation of AES data is an understanding of the volume sampled by the technique. As illustrated in Figure 5, the primary electron beam may penetrate several microns into the solid. A fraction of these primary electrons are backscattered and emitted with some losses (e.g. the plasmon loss features in the secondary electron spectrum). Auger electrons may be excited throughout the penetration volume. As discussed below, however, only those Auger electrons generated within the first few layers of the solid can be detected. This is because of the restricted escape depth for Auger electrons. This situation contrasts sharply with the analysis volume for XRF, which essentially comprises the entire penetration volume. Auger electrons can also be emitted from the near surface layers outside the area that was directly irradiated by the primary electron beam.

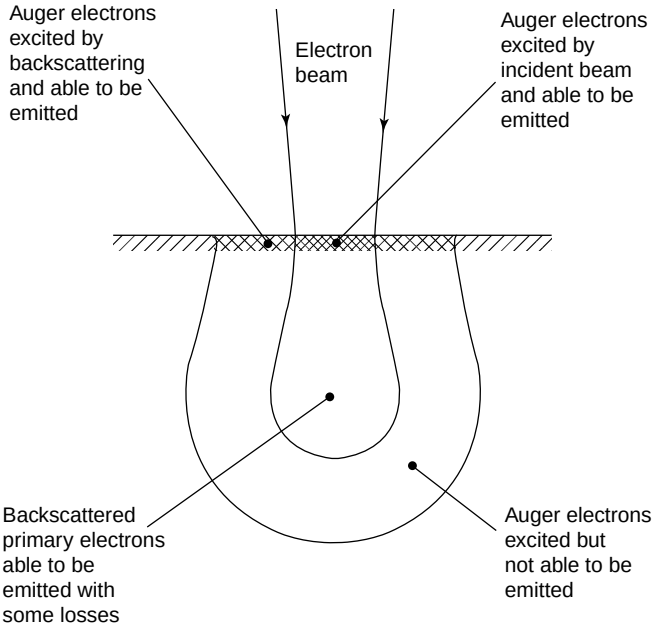

Figure 5 Schematic illustration of primary electron beam interactions with the solid and the resulting analysis volume for AES. (Reproduced with permission from D. Briggs, J.C. Riviere *Practical Surface Analysis: Volume 1: Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, © 1990 John Wiley & Sons Limited.)

This is due to dispersion of the backscattered electrons, which can also excite Auger electrons.

It is the limited escape depth or mean free path of the Auger electrons which is responsible for the surface sensitivity of the technique. Physically, this is due to the

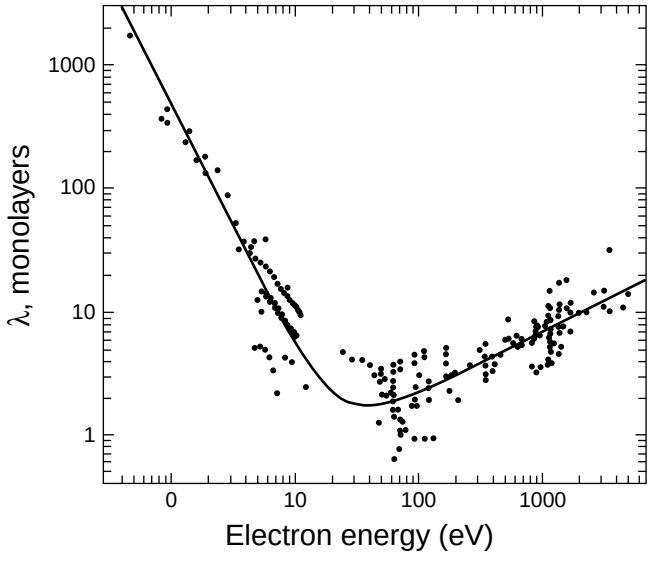

Figure 6 The variation of electron mean free path as a function of electron energy. (Reprinted with permission from Briggs and Riviere.[8])

strong interactions experienced by electrons as they pass through the solid. Figure 6 is a compilation of escape depth data for electron kinetic energies in the range of 2 to 1500 eV.[8] This universal curve is frequently used in quantitative analysis; however, this neglects the variation in escape depth for different materials.

From this discussion, it is evident that the Auger sampling volume is not strongly dependent on the energy of the primary electron beam. The angle of incidence is, however, an important factor. In general, decreasing the angle of incidence from near normal to a low angle (from the surface) increases the analysis area/volume.[11] Thus, varying the angle of incidence can increase the Auger signal for spot mode AES, but degrade the resolution in SAM.

3 INSTRUMENTAL METHODS

3.1 Electron Excitation

As mentioned earlier, the Auger process is initiated by irradiating a sample with electrons, photons, or ions. Since electrons are the most widely used Auger excitation source, this discussion will focus on electron sources (electron guns).

A simple electron gun design is shown in Figure 7. In this example, a filament or thermionic emitter is the source of electrons for the electron gun. Thermionic emitters produce electrons by heating a material (LaB_6, W, Ta, or thorium oxide) to temperatures that allow a portion of the material's electrons to overcome the work function barrier and escape into vacuum. Typical thermionic sources produce electron beam diameters in the millimeter range.[12,13]

A second, common, electron source is based upon field emission. These sources use an applied electric field to reduce the barrier to electron emission. Here, the electrons escape into vacuum by tunneling through a work function barrier of decreased width.[8,13] Consequently, field emission sources can operate at or near room temperature. The resulting electron beam diameter is typically <1 micron in size.[12] Although the smaller beam diameter would make field emission sources the source of choice, current stability issues can be a deterrent. Further, the work function and emission current are greatly affected by adsorption of gases on the field emitter tip.[8] As a result, field emission sources require a highly controlled environment for operation.

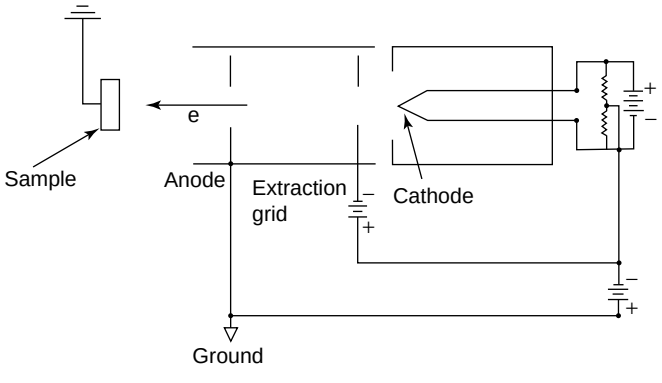

Figure 7 Schematic of a typical electron gun. Cathode emits electrons; extraction grid and anode accelerate electrons away from the cathode and toward the sample.

Once the electrons are emitted into vacuum, they are extracted and accelerated from the emitting area via an anode. Focusing of the beam is accomplished through a combination of electrostatic condenser and objective lenses. Such lenses focus thermionic source beams down to a few microns in size, while field emission sources can be focused down to a few tenths of microns in size. Once focused, deflection plates are used to steer the beam. This allows the beam to be scanned/rastered over a surface in a manner similar to that used in scanning electron microscopy (SEM).[12]

It should also be noted that electron guns utilizing electromagnetic focusing are available. These guns produce smaller beam diameters (20 nm for field emission sources and 50 nm for thermionic sources).[8]

3.2 Electron Energy Analysis

Auger electrons are energy analyzed with one of three types of energy analyzers: cylindrical mirror analyzers (CMAs), concentric hemispherical analyzers (CHAs), or retarding field analyzers (RFAs). The CMA is possibly the most widely used analyzer for AES.

The basic design of a CMA is shown in Figure 8. It consists of a coaxial electron gun, entrance aperture, two

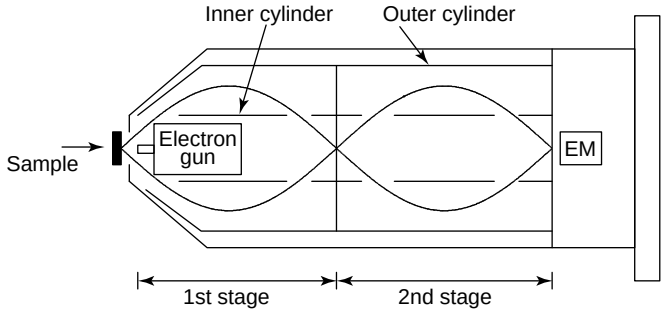

Figure 8 Schematic of a double pass CMA. The electron gun is coaxial with the cylindrical electrodes, with the electron multiplier (EM) located at the focus point of the second stage. The sine-wave lines indicate the path of electrons entering the analyzer. (Reproduced with permission from D. Briggs, J.C. Riviere *Practical Surface Analysis: Volume 1: Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, © 1990 John Wiley & Sons Limited.)

coaxial-cylindrical electrodes, and a detector. Electrons passing through the entrance aperture enter an applied electric field between the coaxial cylinders. The inner cylinder is held at ground, while the outer cylinder has a negative voltage. The resulting field directs electrons in a parabolic path toward the detector. For a given negative voltage, only electrons within a small kinetic energy window are focused onto the detector. In this manner, the concentric cylinder design acts as an energy filter for the analyzer. By varying the voltage on the outer cylinder, the CMA sweeps through a specified range of the electron energy spectra. For a single pass CMA, the detector is located at the exit of the first stage. The detector is either a channeltron or electron multiplier.[8,12,14]

Integral spectra ($N(E)$) are recorded by directly counting the number of electrons detected at each kinetic energy. If a small alternating current (AC) signal is superimposed on the ramp voltage of the outer cylinder, a lock-in amplifier can be used in conjunction with the detector to record derivative spectra (dN/dE).[12] Most modern Auger systems record data in the $N(E)$ mode and use computer algorithms to differentiate the spectra for peak identification and other data analysis.[15]

The acceptance angle, instrument transmission efficiency, and energy resolution are some of the key parameters associated with any energy analyzer. The single pass CMA has a large acceptance angle (42°). This contributes to its high instrument transmission. Energy resolution, $\Delta E/E$, is usually on the order of 0.3–2%, depending on the electron kinetic energy. Adding a second stage of coaxial cylinders in series with the first stage can enhance the energy resolution. In this case, the focal point of the first stage becomes the entrance point for the second stage and the detector is located at the focal point of the second stage. Since this modified CMA has two stages of focusing electro-optics, it is called a double pass CMA.[8,12,16]

The sample-to-analyzer or working distance is a very important parameter for CMAs because they are extremely sensitive to small changes in this parameter. This is because the sample must be positioned at the focal point to optimize the electron optics of the CMA. Studies have shown that changes as small as 0.5 mm in distance lead to noticeable shifts in the electron energy spectra and a significant drop in signal intensity.[8,16] Positioning a sample at the focal point of the analyzer is obtained by aligning the elastic peak of the primary electron beam. The kinetic energy spectrum in the region around the primary electron beam voltage is monitored while the sample is moved along the z-direction (axial direction) of the analyzer. Adjusting the sample's z-position causes shifts in the elastic peak's energy. The system is considered in alignment when the elastic peak's energy matches the primary electron beam voltage. For example, if a sample is positioned at the optimum working distance for an analyzer, the elastic peak generated from a 3 kV primary electron beam will be observed at 3000 eV. The working distance is also important because it may (or may not) limit the ability to simultaneously focus other diagnostic tools on the sample.

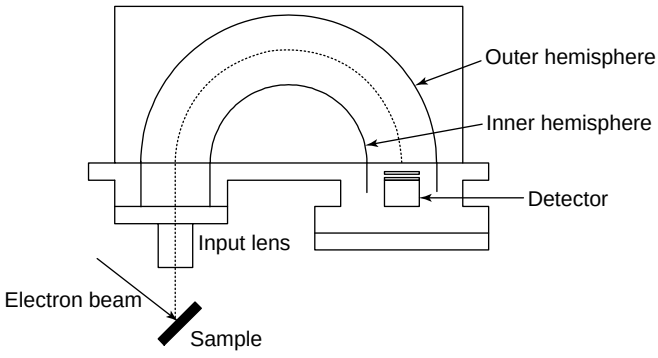

Figure 9 Basic design of a CHA. (Reproduced with permission from D. Briggs, J.C. Riviere *Practical Surface Analysis: Volume 1: Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, © 1990 John Wiley & Sons Limited.)

It should also be noted that not all CMAs are equipped with a coaxial electron gun. For vacuum systems where multiple techniques might be used, an external electron gun can provide more flexibility. An external electron gun also allows more flexibility for analyzing at different angles of incidence.

The CHA shown in Figure 9 consists of an input lens, concentric hemispherical electrodes, and a detector.[8,12,14] The electron gun is external for this energy analyzer design. The input lens retards (reduces the kinetic energy of) electrons and focuses them on the entrance aperture to the hemisphere. The electric field produced between the inner and outer hemispheres acts as an energy filter and focuses electrons on the detector at the exit aperture of the hemispheres.

CHAs can be operated in two different modes. In the constant analyzer energy (CAE) mode, the voltage difference between the hemispheres is fixed. Consequently, the kinetic energy of the electrons that can pass through the analyzer (i.e. the pass energy) and the range of passable energies (ΔE) is also fixed. In this case, a sweep through the energy spectra is obtained by ramping the retarding voltage on the input lens. The advantage of this mode is that analyzer contributions to peak widths will remain constant across the energy spectra.[8,14]

In the constant retard ratio (CRR) mode, the potential difference across the inner and outer hemispheres is varied. This means that the pass energy is varied and the range of passable energies (ΔE) also varies. However, the

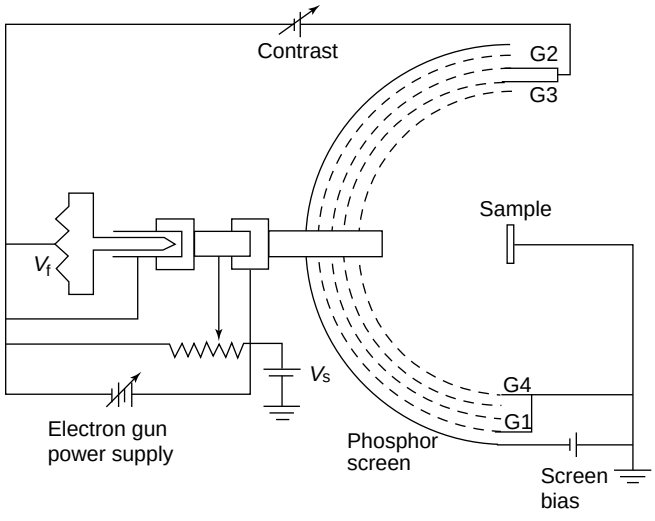

Figure 10 Schematic of a LEED/Auger RFA.

ratio of $\Delta E/E$ remains constant; the energy resolution is constant across the energy spectrum. Typically, CHAs used for AES are run in the CRR mode, while those used for XPS are run in the CAE mode. The acceptance angle and instrument transmission efficiency can be comparable to that of the CMA. The sensitivity to sample distance is lower than that of the CMA.[8,12,14]

An RFA is shown in Figure 10.[17] This design is based upon electron optics used for LEED from single crystal samples.[12] The LEED/RFA consists of an electron gun, four grids, and a phosphor screen. G1 and G4 are held at ground potential, while G2 and G3 have variable negative potentials carrying an AC signal. A lock-in amplifier provides data in the $E\,\mathrm{d}N/\mathrm{d}E$ mode by processing the signal detected at the phosphor screen.[17] This is the only mode of data acquisition for this type of energy analyzer. RFAs have good instrument transmission, CMA quality energy resolution, and large acceptance angles. The signal-to-noise ratio is rather poor, while the sensitivity to sample distance is comparable to a CHA.[7,12]

3.3 Vacuum Requirements

One of the basic requirements for operating an Auger system is an ultra-high vacuum (UHV) environment. A number of factors contribute to this. First, the electrons from the source or sample must be able to reach their desired destinations (sample or detector) before colliding with gas molecules.[12,16] Since the mean free path of a free electron is approximately 1 m at a pressure of 10^{-3} torr,[16] Auger analyses could be carried out under relatively poor vacuum conditions. However, such pressures would adversely affect the lifetime of the filaments used in the electron sources. As a result, pressures below 10^{-5} torr are recommended for extended electron source lifetimes.[16]

The reactivity of the sample being analyzed and the high surface sensitivity of AES are factors that must be considered. Reactive samples may be chemically altered by interactions with the ambient gas molecules (e.g. oxidation of metals). At 10^{-6} torr, potentially 1 monolayer (1 monolayer $= 10^{15}\,\mathrm{cm}^{-2}$) of ambient gas adsorption could occur in a period of one second.[18] Since the utility of the AES technique is based on its ability to examine only the first few atomic layers of a sample, this represents a serious problem. To minimize this effect, most Auger systems operate at pressures in the 10^{-8}–10^{-10} torr range.[8,12,16]

The instrumentation required to achieve such pressures is readily available from a variety of vacuum equipment vendors (see section 3.8). A typical vacuum system consists of a stainless steel chamber with a number of ports and attached pumping equipment. The dimensions of the ports are standardized, so equipment from virtually any vendor can be attached to the chamber. For a comprehensive review of standard vacuum equipment and instrumentation, please refer to the work of Harris.[18]

3.4 Signal-to-noise Factors

There are a number of instrumental and sample factors that contribute to the measured signal intensity of an Auger transition. Equation (2) relates these parameters to the Auger intensity, I, for a specific transition i.

$$I_i = AI_pX_iNB\phi_i\psi_iRT_i\lambda_i \qquad (2)$$

Here, A is the area irradiated by the electron beam, I_p is the primary electron beam current density (A/cm^2), X_i is the fraction of atoms producing transition i, N is the atom density (atoms/cm^3) of the material, B is the backscattering factor (>1) of the sample, ϕ_i is the ionization cross-section (cm^2/atom) for transition i, ψ_i is the transition probability, R is the surface roughness factor, T_i is the transmission factor of the energy analyzer, and λ_i is the energy dependent escape depth of the emitted electron.[19]

Increasing the primary beam current will increase the measured intensity. However, using an excessively high current may cause electron beam damage of the sample surface. The backscatter factor is a representation of how well a material causes backscattering of the primary electrons. Increasing the number of backscattered primary electrons leads to a small increase in the measured Auger intensity and a substantial increase in the background (noise) signal. Similarly, the roughness factor, R, is a representation of how rough the sample surface is. This is an important factor to consider because rough surfaces cause more electron scattering than smooth surfaces. A larger portion of primary electrons is scattered away from the sample and more Auger electrons are scattered back

into the sample. This has a two-fold effect on the electron energy spectra. The increased scattering of Auger electrons leads to a decrease in the Auger signal detected. The increased scattering of primary electrons leads to an increase in the background (noise) signal.

Sample charging is another factor that must be considered. During Auger analyses of insulating or semi-insulating samples, a negative charge can accumulate in the near surface regions. This adversely affects the electron energy spectra by causing Auger peak shifts and lineshape broadening. In addition, charging may significantly increase the background noise.[16,20]

Fortunately, a number of options for counteracting the effects of charging exist. The simplest option involves lowering the primary beam current. This may lead to significantly increasing acquisition times so that the desired signal-to-noise ratio can be obtained. If the insulating material is a thin film on a semiconducting or conducting substrate, increasing the primary beam voltage may help. By increasing the beam voltage, a larger portion of the primary electrons may reach the film/substrate interface and be transported away by the sample's connection to ground. The net effect is a reduction in the negative charge accumulation in the film. However, if the insulating film is thicker than the range of the primary electrons, increasing the beam voltage will only serve to enhance the charging effect. Another option is to perform the analysis at glancing incidence (i.e. primary beam incident at glancing angle with respect to sample surface). This option decreases the number of primary electrons that end up trapped in the insulating material.[7,12,15,20] Recent developments in electron spectroscopy have shown that simultaneous sample irradiation by low-energy (10 eV) inert ions can significantly reduce, and in most cases offset, the effects of sample charging.[21]

3.5 Point Analysis and Auger Mapping

The most common mode of Auger characterization is point analysis (spot mode analysis). Spot analysis usually provides the best signal and minimizes the effects of surface roughness. However, directing large currents into a small area can lead to electron beam induced chemistry/damage on the sample surface. Decomposition of polymers, desorption of adsorbates, oxidation, migration of mobile species, adsorption, chemical decomposition, and sample charging are some of the effects that have been reported.[7,15,16,22] In some cases, decreasing the primary beam current can alleviate these effects. Another option is scanning the beam (line or area scan) across a region of the sample.[20] For example, Pepper has shown that electron beam exposure of a diamond surface during Auger analysis can cause graphitization of the surface.[22] By rotating the sample during analysis, Pepper effectively avoided prolonged exposures of a given analysis spot to the electron beam and remedied this problem. The disadvantage with this approach is that the Auger signal may decrease due to varying surface roughness and/or varying secondary electron yields across the sample. Figure 11 is an example of two survey scans acquired in spot mode.

Rastering capabilities also make it possible to monitor the intensity of a given Auger peak as a function of position (Auger mapping or Auger line scans). These techniques are very useful for looking at the spatial distribution of specific elements across a sample. Figure 12 is an example of how Auger mapping was used to look at the distribution of elements in a cross-section of a multilayered device. The transition from one metal layer to another is more clearly defined in the Auger maps than in the secondary electron image. It should be noted

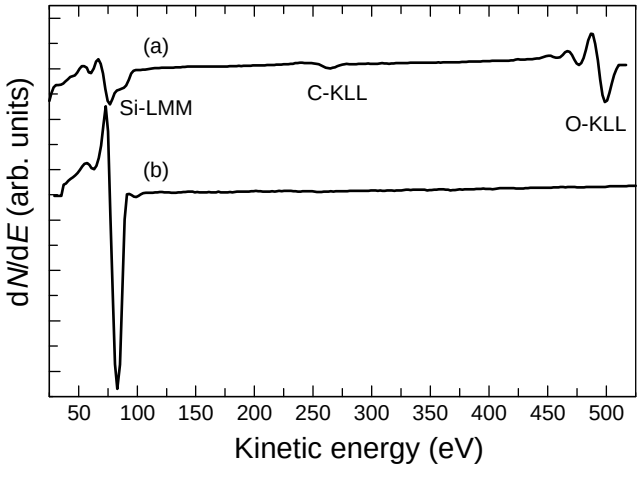

Figure 11 Survey spectrum of (a) as-received Si(100) and (b) cleaned Si(100). Data acquired in spot mode.

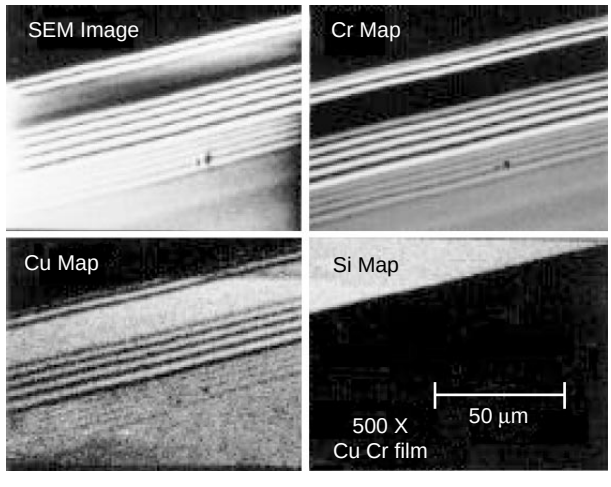

Figure 12 Auger map of Cu/Cr multi-layered device.

that Auger mapping can be extremely time-consuming. Each point on the map requires a number of sweeps over the Auger peak for each element of interest. The less time-consuming alternative is to perform an Auger line scan. Instead of mapping the intensity of specific elements for a whole field of view, the intensities of the elements are mapped for only a single line.

3.6 Sample Preparation

For most AES applications, the as-received surface of a sample has been exposed to air and is therefore covered with an ''adventitious layer'' of carbon and oxygen. As a result, a number of sample cleaning/preparation methods have been developed to produce a ''clean'' surface.

The most common method involves inert ion bombardment of the sample surface. Inert ions (Ar or Xe) are used because they will not chemically react with samples. The ions are created and accelerated toward the sample surface with an ion gun. The simple ion gun shown in Figure 13 creates ions by electron impact ionization of the inert gas. The gas is either introduced into the ionization volume of the gun directly or by introducing a backpressure of gas in the vacuum chamber. In the ionization volume of the ion gun, electrons are emitted from two hot filaments and accelerated toward an anode. Prior to reaching the anode, the energetic electrons collide with the gas molecules and ionize them. Once formed, the ions are accelerated away from the anode. The focusing lens(es) focus the ions into a beam and the deflection plates raster the beam over an area of the target. Most ion guns produce ions with energies in the 500–5000 eV range and beam diameters ranging from 5–100 microns.[12] Low-energy (10–100 eV) ion sources and radio frequency source ion guns are also commercially available from a variety of vendors (see section 3.8).

Some of the effects associated with ion sputtering are knock-on processes, enhanced diffusion of a sample's constituents, preferential sputtering, surface roughening, and chemical degradation of the sample.[8,12,16] Knock-on refers to ions causing surface atoms to be driven into the sample. Diffusion can be enhanced by the energy being deposited into the sample during bombardment. Preferential sputtering occurs when constituents of differing collision cross-sections and/or sputter yields are present in a sample. The element with the higher sputter yield will be sputtered away at a higher rate than the element with the lower sputter yield. Surface roughening is an issue because ion sputtering can increase the surface roughness over a period of time. This effect can be virtually eliminated by rotating the sample during ion bombardment.[12] In all cases, the resulting surface can give an erroneous representation of the original sample composition.

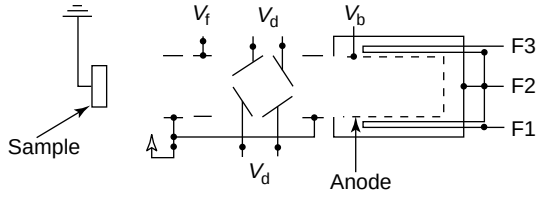

Figure 13 Schematic of an electron impact ion source. F1 and F3 are the filaments. V_b is the anode potential. V_d is the deflection plate voltage. V_f is the focusing lens voltage.

For soft metals, in situ cleaning can be accomplished by scraping the surface with a sharp edge or abrading the surface with a wire brush. For other samples, a fresh surface may be obtained by in situ fracturing/cleaving of the sample. Both methods require the use of additional vacuum accessories.[8,12,14]

3.7 Depth Profiling

Depth profiling is a technique that combines AES with ion sputtering to provide compositional information as a function of depth. Depending on the mode of operation, Auger data can be acquired either simultaneously with the ion sputtering (continuous mode) or in alternating stages (alternating mode), using preset sputter times. Depth profile data is then recorded as a function of sputter time. Converting sputter time to depth can be accomplished in several ways. Once the depth profile is completed, the resulting crater depth can be measured by a surface profilometer. Then, assuming a constant sputter rate, the relationship between depth and sputter time can easily be determined. Figure 14 is an example of a depth profile where the depth indicated on the x-axis was determined from a profilometer measurement. For samples made of a well-characterized material (e.g. SiO_2), sputter rates are readily available in the literature.[8] In other cases, standards may be used to calibrate the ion gun etch rate for a given material. Currently, oxides and metal film standards of known thickness are commercially available.

3.8 Instrumentation Vendors and Systems

The following (Table 1) is a list of vendors that sell AES systems, components, and related vacuum accessories. For further information about their products, please contact them through the indicated website addresses. Their listing here should in no way be construed as an endorsement of their products or services. This list is provided for informational purposes only.

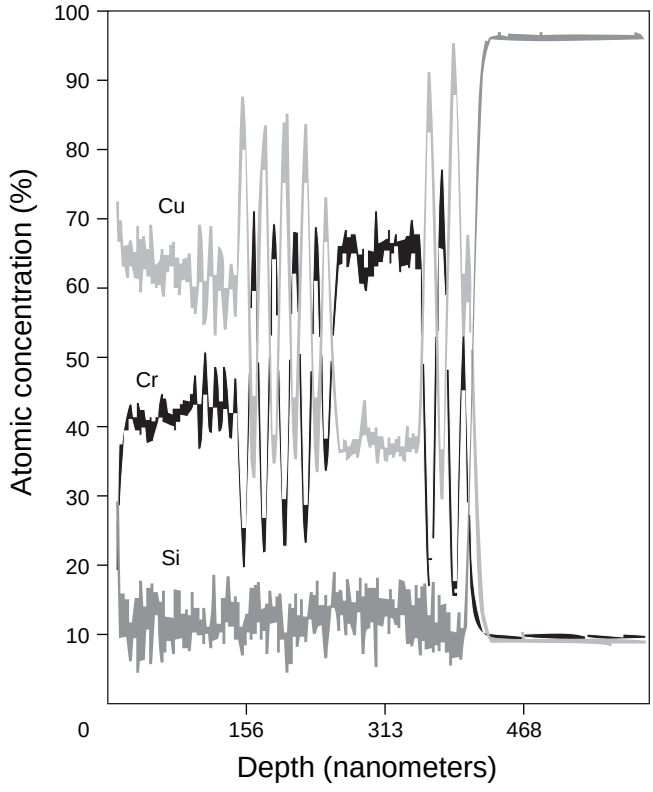

Figure 14 Depth profile of a Cu/Cr multi-layer film on Si.

4 DATA ANALYSIS

4.1 Elemental Identification

As was discussed in section 2.1, Auger spectra are fingerprints of the density of electron states (DOS) for the elements present in a sample. Since the DOS are unique for each element, the Auger transitions are also element specific. When combined with the large database of existing Auger spectra (e.g. PHI or JEOL Handbooks), elemental identification is possible without the use of calculations. The minor differences between reference spectra are not enough to create confusion in identifying elements. Since Auger spectra of many elements contain multiple peaks, looking at all of the peaks associated with an element removes most of the ambiguity in identifying the elements that are present. Ultimately, any uncertainty can be removed by acquiring standard spectra with one's own spectrometer.

4.2 Quantification of Data

Extracting quantitative information from Auger spectra can be a difficult process. The Auger intensity is dependent upon a number of parameters (see Equation 2), most of which are not sufficiently known.[12,15,16] So, in practice, an approach utilizing elemental sensitivity factors, derived from comparisons of standard spectra, is used. Most Auger systems are supplied with a database of sensitivity factors relevant for that particular instrument. These factors provide reasonably accurate quantification results when data is acquired under the same conditions as those used for deriving the sensitivity factors. Again, more accurate results can be obtained by acquiring standard spectra with one's own spectrometer (see section 4.3).

The most common method of determining peak intensities is the measurement of peak-to-peak height in the dN/dE spectra.[12,15,16] The relative intensity and atomic concentration of an element i is then calculated using Equations (3) and (4), respectively.

$$I_{\mathrm{ri}} = \frac{I_{\mathrm{mi}}}{S_{\mathrm{i}}} \tag{3}$$

$$AC_{\mathrm{i}} = 100 \times \frac{I_{\mathrm{ri}}}{\sum I_{\mathrm{rj}}} \tag{4}$$

Table 1 List of AES systems and related accessories

Vendor	Related products	Website
Physical Electronics (PHI)	Surface analysis systems (AES, XPS, SIMS) and components	www.phi.com
Omnicron	Energy analyzers, electron sources, ion guns	www.omnicron.de
VG Scientific	Custom and standard surface analysis systems	www.vgscientific.com
SPECS	Surface analysis systems, analyzers, ion sources, electron sources	www.specs.de
Staib Instruments	Energy analyzers, electron guns, custom vacuum systems	www.staib-instruments.com
Kimball Physics Inc.	Electron sources, ion sources	www.kimphys.com
FEI Beam Technology	Electron sources	www.feibeamtech.com
Kurt J. Lesker Company	Surface analysis equipment, vacuum systems, electron sources, ion sources, and vacuum accessories	www.lesker.com
LK Technologies	Electron sources, ion sources, Auger spectrometers	www.lktech.com
Commonwealth Scientific	Ion sources	www.ionbeam.com

For the ith transition, I_{ri} is the relative intensity, I_{mi} is the measured intensity, S_i is the relative sensitivity factor, and AC_i is the relative atomic concentration (atomic %). The sum is over all elements detected.[23]

While the use of peak-to-peak heights provides a reasonable estimate of the relative concentration of elements, a more accurate representation is achieved by integrating the area under an Auger peak in the $N(E)$ spectra.[12,16] Subtracting out the background and correcting for resolution and loss features are required prior to integration. It should be noted that background subtraction can be a difficult task. Since most sensitivity factors are generated from peak-to-peak height data, integrated intensities must be related to peak-to-peak heights via a proportionality factor[24,25] to avoid introducing more error in the quantification.[16]

4.3 Calibration Using External Standards

Calibration of sensitivity factors for an Auger spectrometer becomes necessary when the accuracy and precision of quantification data are an issue. Making up standards of known compositions that approximate the sample under study is useful. Standards should be chosen to cover a range of compositions, such that the unknown sample's composition falls within the calibrated range of the spectrometer. Analysis of the standards and the unknown sample should also be performed with all operating conditions held constant.[16] For example, in recent Auger studies of boron and phosphorus doped germanosilicate glass, results produced from use of the PHI Handbook sensitivity factors conflicted with results acquired from energy dispersive spectroscopy (EDS) analyses of the same samples. As a result, a SiO_2 and a boro-phospho-silicate glass (BPSG) standard were analyzed. The oxide standard was used to calibrate the O sensitivity factor. The BPSG standard was used to calibrate the sensitivity factors for B, P, and Si in a matrix similar to the sample under study. Since a good germanosilicate standard was not available, the handbook value for Ge was used. Re-evaluation of the unknown sample's data using the newly derived sensitivity factors and the handbook Ge factor provided results in good agreement with the EDS results.[26] It should be noted that, despite the rigor of this approach, errors associated with drifts in primary beam current and variations in sample roughness are still a possibility.[16]

4.4 Extraction of Chemical Bonding Information

The shape and position of an Auger peak are reflections of an element's core-level energies and valence band density of states. Altering the bonding environment of the element induces a shift in the core-level energies and a redistribution of electrons in the valence band that directly affects the Auger peak position and lineshape for that element. As a result, analyzing the peak position and lineshape of a transition provides some chemical bonding information.[7,12,16] Examples of shifts in peak position and lineshape changes for Si and C, respectively, are well illustrated in the literature.[7,8,23,27,28] The elemental Si KLL peak is typically observed at 1619 eV, while the silicon dioxide Si KLL peak is shifted down to 1609 eV.[23] This energy shift is large enough that identification of the bonding states for Si and SiO_2 are relatively easy. Auger spectra of various carbon compounds illustrate changes in lineshapes associated with changes in chemical bonding. Figure 15 shows the derivative spectra of graphite, diamond, and SiC standards. While the energy shifts are negligible, the lineshapes are distinctly different.

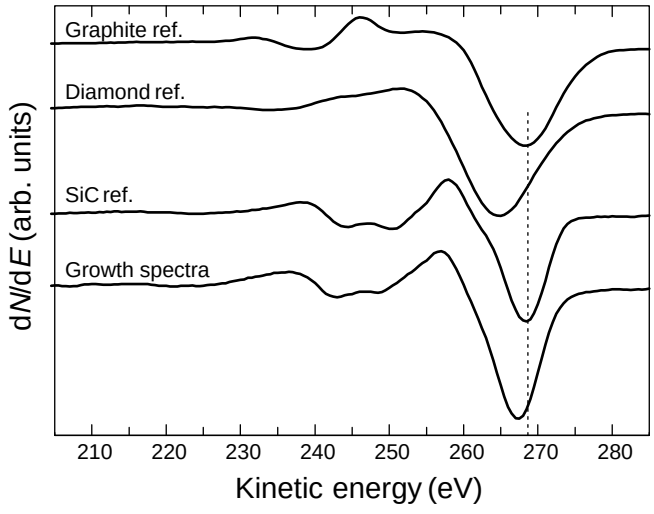

Figure 15 C-KLL Auger spectra for three carbon species (SiC, graphite, and diamond) and an as-grown carbon film. The differences in Auger lineshapes illustrate differences in the chemical bonding of the carbon atoms for each type of carbon.

Quantification and identification of the chemical bonding information contained in AES data is a relatively new field. Factor analysis is one technique that has been used to quantify chemical bonding information.[29,30] Establishing a database of literature in this area of analysis is an endeavor undertaken by a few technical journals (i.e. *Surface Science Spectra* or *Surface and Interface Analysis*). The reader is referred to these journals and the general literature (journals like *Surface Science, Surface and Interface Analysis, Journal of Electron Spectroscopy and Related Phenomena*, and *Journal of Vacuum Science and Technology*) as a source of future information in this regard.

5 APPLICATIONS

Initial use of AES was limited primarily to fundamental surface science applications of the technique. As

awareness and understanding of the utility of AES spread and as AES related equipment improved, the applications have expanded greatly. Through the past thirty years or so, AES has played an important role in the development of broad areas of surface and materials science. In the following discussion, examples of this work will be discussed.

5.1 Basic Surface Science

Surface studies using AES include analyses of adsorbed species bonding, sticking coefficients and surface reaction mechanisms, and surface segregation. In many cases, AES is combined with other analytical tools such as LEED, EELS, or thermal desorption spectroscopy (TDS). One example of this is the work of Kiskinova et al.[31] on the adsorption of oxygen on an alkali metal-covered Ni(100) surface. In this study, AES intensities were used to follow the kinetics of oxygen uptake for various alkali metals and alkali metal coverages of the Ni(100) surface. Also, AES lineshapes were used to deduce information concerning the nature of the O–Ni bonding. It was shown that the effect of the alkali metals was associated with an increase of the electron charge distribution in the near surface layers. The overall effect was to increase the surface oxygen concentration and induce the formation of an oxide phase at lower oxygen exposures than for clean nickel surfaces.

In addition to studies of metal surface chemistry, considerable attention has been given to semiconductors. In an excellent review article, Yates and coworkers[32] describe in detail the application of AES, TDS, and other surface science methods to the study of semiconductor surface chemistry. Further, there is an extensive literature relating specifically to silicon, gallium arsenide, and other semiconductors.

Drathen et al.[33] have, for example, studied the composition and structure of differently prepared GaAs(100) surfaces using LEED and AES. They were able to determine that different surface structures were uniquely and sensitively correlated to the arsenic surface coverage. In this work, the quantitative determination of the arsenic surface coverage was obtained from AES peak intensities using a layer model and a calibration measurement on a cleaved GaAs(110) surface.

Leone and coworkers[34] have characterized the structures and growth of gallium overlayers on Si(100) surfaces. By measuring AES peak intensities as a function of gallium exposure they were able to determine the growth mechanisms. They found that gallium grows epitaxially up to 5 monolayers at 300 K. For substrate temperatures in the range of 330–600 K, growth followed the Stranski–Krastanov mechanism. The types of surface structures observed were found to depend on the gallium surface coverage.

5.2 Materials Science

AES has played a vital role in the development of many aspects of materials science. This area spans a broad range including metallurgy, oxidation and corrosion science, polymer technology, ceramics and powder technology, composite materials, tribology, catalysis, and microelectronics. Even within these topics there are broad application areas which merit their own review article. Two representative AES applications of interest and significance to this general area are surface segregation and thin film growth.

An excellent overview of surface segregation and its influence on materials properties is provided by Johnson and Blakely.[35] As a result of differences in bonding and lattice strain at the surface and in the bulk, the surface composition of a multi-component system may differ substantially from its bulk composition. This is known generally as surface segregation and it may occur at grain boundaries as well as free surfaces. At a free surface, segregation may alter the sticking coefficient or reactivity. At an internal grain boundary, segregation may alter the mechanical properties of the solid.

Harris[3] was the first to use AES to observe the segregation of sulfur to the surface of nickel. In these studies, it was observed that sulfur, present at very low levels in the bulk, segregated to the surface when the sample was heated to temperatures in the range of 873 to 1173 K. Similar examples of impurity segregation are reported by Jenkins and Chung[36] (sulfur in copper) and Taylor[37] (sulfur and phosphorus in stainless steel).

Segregation of alloy components has also been of intense practical and theoretical importance. An example of this is the excellent work of Webber et al.[38] on the segregation of copper in nickel. For temperatures in the range of 850 to 920 K, they measured the equilibrium surface composition for alloys of 5% and 50% copper using a combination of AES and XPS. This combination of techniques, with their slightly differing sampling depths, allowed them also to non-destructively depth profile the near surface layers of the nickel. In addition to providing insight into the segregation process, these studies also demonstrated the difficulty of establishing equilibrium in experiments where evaporation of one component (Cu in this case) may occur. This observation led Lawson and Stinespring[39] to extend the segregation model of Hofman and Erlewein[40] to include the evaporation process.

The use of AES in the study of thin film growth has also been extensive. One example of this is the investigation of SiC formation by Yates and coworkers.[41] In these studies SiC was grown at temperatures in the range of 890–1140 K using C_2H_4 at an effective pressure of 4.4×10^{-5} torr. At

predetermined intervals, the growth was interrupted to perform AES, XPS, and ELS measurements. In later work, Stinespring and Wormhoudt[42] extended these studies to include the use of other gas species (CH_4 and C_3H_8) for growth at temperatures as high as 1495 K. Based on such studies, it is possible to determine fundamental parameters, such as reaction rates, and to gain some insights into the mechanisms of chemical vapor deposition (CVD) processes used to synthesize materials.

AES has also been used extensively to investigate diamond thin film growth. Stoner et al.[43] have used a combination of AES, XPS, and Raman spectroscopy to investigate bias enhanced nucleation of diamond during plasma assisted chemical vapor deposition (PACVD). By periodically stopping the growth and performing a variety of surface and bulk characterization measurements, they were able to gain considerable understanding of the overall sequence of events which occur as diamond is first nucleated and then grown on silicon substrates. In later studies, Lannon et al.[44,45] performed detailed surface studies using AES to characterize the initial stages of diamond nucleation on silicon. In these studies, C KLL line shapes were analyzed using factor analysis and standards for graphite, silicon carbide, and diamond. Use of factor analysis followed the methods developed by Fuchs et al.[46] to characterize the sp^2-C/sp^3-C ratios of a-C : H films and by Putz et al.[30] to determine the W−C/sp^2-C ratios of tungsten carbide films.

ACKNOWLEDGMENTS

The authors would like to thank Adam Holyoake at J. Wiley & Sons for his assistance during the various phases of preparing this manuscript. The authors would also like to recognize Gary McGuire at MCNC for his encouragement and recommendations during the preparation of this work. Finally, we would like to thank our coworkers and loved ones for their patience and support throughout the writing process.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
AES	Auger Electron Spectroscopy
BPSG	Boro-phospho-silicate Glass
CAE	Constant Analyzer Energy
CHA	Concentric Hemispherical Analyzer
CMA	Cylindrical Mirror Analyzer
CRR	Constant Retard Ratio
CVD	Chemical Vapor Deposition
DOS	Density of Electron States
EDS	Energy Dispersive Spectroscopy
EELS	Electron Energy-loss Spectroscopy
HREELS	High-resolution Electron Energy-loss Spectroscopy
LEED	Low-energy Electron Diffraction
PACVD	Plasma Assisted Chemical Vapor Deposition
PLS	Plasmon Loss Spectroscopy
RFA	Retarding Field Analyzer
RHEED	Reflection High-energy Electron Diffraction
SAM	Scanning Auger Microscopy
SEM	Scanning Electron Microscopy
TDS	Thermal Desorption Spectroscopy
UHV	Ultra-high Vacuum
XPS	X-ray Photoelectron Spectroscopy
XRF	X-ray Fluorescence

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
Steel and Related Materials: Introduction

Surfaces **(Volume 10)**
Surfaces: Introduction • Electron Energy Loss Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Mass Spectrometry **(Volume 13)**
Secondary Ion Mass Spectrometry as Related to Surface Analysis • Time-of-flight Mass Spectrometry

REFERENCES

1. P. Auger, 'Sur L'Effet Photoélectrique Composé', *J. Phys. Radium*, **6**, 205 (1925).
2. J.J. Lander, 'Auger Peaks in the Energy Spectra of Secondary Electrons from Various Materials', *Phys. Rev.*, **91**, 1382 (1953).
3. L.A. Harris, 'Analysis of Materials by Electron-Excited Auger Electrons', *J. Appl. Phys.*, **39**, 1419 (1968).
4. R.E. Weber, W.T. Peria, 'Use of LEED Aparatus for the Detection and Identification of Surface Contaminants', *J. Appl. Phys.*, **38**, 4355 (1967).
5. P.W. Palmberg, G.K. Bohn, J.C. Tracy, 'High Sensitivity Auger Electron Spectrometer', *Appl. Phys. Lett.*, **15**, 254 (1969).
6. N.C. MacDonald, 'Auger Electron Spectroscopy in Scanning Electron Microscopy: Potential Measurements', *Appl. Phys. Lett.*, **16**, 76 (1970).

7. C.C. Chang, 'Auger Electron Spectroscopy', *Surface Science*, **25**, 53 (1971).
8. D. Briggs, J.C. Riviere, *Practical Surface Analysis, Volume 1 – Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 1990.
9. M.F. Chung, L.H. Jenkins, 'Auger Electron Energies of the Outer Shell Electrons', *Surface Science*, **22**, 479 (1970).
10. M. Gryzinski, 'Classical Theory of Atomic Collisions. I. Theory of Inelastic Collisions', *Phys. Rev.*, **A138**, 336 (1965).
11. L.A. Harris, 'Angular Dependences in Electron-excited Auger Emission', *Surface Science*, **15**, 77 (1969).
12. N.H. Turner, 'Auger Electron Spectroscopy', in *The Handbook of Surface Imaging and Visualization*, ed. A.T. Hubbard, CRC Press, Boca Raton, FL, 1995.
13. J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, C. Fiori, E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*, Plenum Press, New York, London, 1981.
14. H. Luth, *Surfaces and Interfaces of Solid Materials Third Edition*, Springer-Verlag, Berlin, Heidelberg, 1995.
15. C.J. Powell, M.P. Seah, 'Precision, Accuracy, and Uncertainty in Quantitative Surface Analyses by Auger-electron Spectroscopy and X-ray Photoelectron Spectroscopy', *J. Vac. Sci. Technol.*, **A8**(2), 735 (1990).
16. P.H. Holloway, 'Fundamentals and Applications of Auger Electron Spectroscopy', in *Advances in Electronics and Electron Physics 54*, Academic Press, New York, 241, 1980.
17. *LEED – Auger Electron Optics Systems*, Perkin-Elmer Technical Manual for PHI Model 15-120, Physical Electronics Division, Eden Prarie, MN, 1972.
18. N.S. Harris, *Modern Vacuum Practice*, McGraw-Hill Book Company, London, 1989.
19. C.C. Chang, 'Analytical Auger Electron Spectroscopy', in *Characterization of Solid Surfaces*, eds. P. Kane, G. Larrabee, Plenum Press, New York, 1974.
20. W. Wei, 'Reduction of Charging Effects in Auger Electron Spectroscopy', *J. Vac. Sci. Technol.*, **A6**(4), 2576 (1988).
21. P.E. Larson, M.A. Kelly, 'Surface Charge Neutralization of Insulating Samples in X-ray Photoemission Spectroscopy', *J. Vac. Sci. Technol.*, **A16**(6), 3483 (1998).
22. S. Pepper, 'Electron Spectroscopy of the Diamond Surface', *Appl. Phys. Lett.*, **38**(5), 344 (1981).
23. L.E. Davis, N.C. MacDonald, P.W. Palmberg, G.E. Riach, R.E. Weber, *Handbook of Auger Electron Spectroscopy*, Physical Electronics Industries, Inc., 1976.
24. N.J. Taylor, 'Resolution and Sensitivity Considerations of an Auger Electron Spectrometer Based on Display LEED Optics', *Rev. Sci. Instrum.*, **40**, 792 (1969).
25. Y.E. Strausser, in *Applied Surface Analysis*, ASTM STP 699, Am. Soc. Test. Mater., Philadelphia, Pennsylvania, 1980.
26. R.T. Croswell, A. Riesman, D.L. Simpson, D. Temple, C.K. Williams, 'Planarization Processes and Applications Part III – As-deposited and Annealed Film Properties', in preparation for submission to *J. Electr. Chem. Soc.*
27. T. Haas, J. Grant, G. Dooley III, 'Chemical Effects in Auger Electron Spectroscopy', *J. Appl. Phys.*, **43**(4), 1853 (1972).
28. P.G. Lurie, J.M. Wilson, 'The Diamond Surface II: Secondary Electron Emission', *Surface Science*, **65**, 476 (1997).
29. S.W. Gaarenstroom, 'Principal Component Analysis of Auger Line Shapes at Solid–Solid Interfaces', *Appl. Surf. Sci.*, **7**, 7 (1981).
30. T. Putz, A. Fuchs, H. Erhardt, 'Factor Analysis Applied to Carbon Auger Spectra of Tungsten Carbide Coatings Deposited by Reactive Sputtering', *Surface Science*, **265**, 219 (1992).
31. M. Kiskinova, L. Surnev, G. Bliznakov, 'Oxygen Adsorption on an Alkali Metal Covered Ni(100) Surface', *Surface Science*, **104**, 240 (1981).
32. M.J. Bozack, L. Muehlhoff, J.N. Russell, Jr, W.J. Choyke, J.T. Yates, 'Methods in Semiconductor Surface Chemistry', *J. Vac. Sci. Technol.*, **A5**, 1 (1987).
33. P. Drathen, W. Ranke, K. Jacobi, 'Composition and Structure of Differently Prepared GaAs(100) Surfaces Studied by LEED and AES', *Surface Science*, **77**, L162 (1978).
34. B. Bourguignon, K.L. Carleton, S.R. Leone, 'Surface Structures and Growth Mechanisms of Ga on Si(100) Determined by LEED and Auger Electron Spectroscopy', *Surface Science*, **204**, 455 (1988).
35. W.C. Johnson, J.M. Blakely (eds.), *Interfacial Segregation*, American Society of Metals, Metals Park, OH, 1979.
36. L.H. Jenkins, M.F. Chung, 'LEED and Auger Investigations of Cu(111) Surfaces', *Surface Science*, **24**, 125 (1971).
37. N.J. Taylor, 'Auger Electron Spectrometer as a Tool for Surface Analysis (Contamination Monitor)', *J. Vac. Sci. Technol.*, **6**, 241 (1969).
38. P.R. Webber, C.E. Rojas, P.J. Dobson, D. Chadwick, 'A Combined XPS/AES Study of Cu Segregation to the High and Low Index Surfaces of a Cu–Ni Alloy', *Surface Science*, **105**, 20 (1981).
39. W.F. Lawson, C.D. Stinespring, 'A Theoretical Model for Combined Surface Segregation and Evaporation', *Surface Science*, **150**, 209 (1985).
40. S. Hofman, J. Erlewein, 'A Model of the Kinetics and Equilibria of Surface Segregation in the Monolayer Regime', *Surface Science*, **77**, 591 (1978).
41. F. Bozso, J.T. Yates, Jr, W.J. Choyke, L. Muehlhoff, 'Studies of SiC Formation on Si(100) by Chemical Vapor Deposition', *J. Appl. Phys.*, **57**, 2771 (1985).
42. C.D. Stinespring, J.C. Wormhoudt, 'Surface Studies Relevant to Silicon Carbide Chemical Vapor Deposition', *J. Appl. Phys.*, **65**(4), 1733 (1989).

43. B.R. Stoner, G.-H. Ma, S.D. Wolter, J.T. Glass, 'Characterization of Bias-enhanced Nucleation of Diamond on Silicon by In Vacuo Surface Analysis and Transmission Electron Microscopy', *Phys. Rev.*, **B45**, 11067 (1992).
44. J.M. Lannon, Jr, J.S. Gold, C.D. Stinespring, 'Hydrogen Ion Interactions with Silicon Carbide and the Nucleation of Diamond Thin Films', *J. Appl. Phys.*, **77**, 3823 (1995).
45. J.M. Lannon, Jr, J.S. Gold, C.D. Stinespring, 'Surface Studies Relevant to the Initial Stages of Diamond Nucleation', *Mat. Res. Soc. Symp. Proc.*, **339**, 63 (1994).
46. A. Fuchs, J. Scherer, K. Jung, H. Erhardt, 'Determination of sp^2/sp^3 Carbon Bonding Ratio in a-C:H Including Irradiation Damage by Factor Analysis of Auger Electron Spectra', *Thin Solid Films*, **232**, 51 (1993).

Differential Reflectance Spectroscopy in Analysis of Surfaces

R.E. Hummel
University of Florida, Gainesville, USA

1 Introduction 1
2 The Instrument 2
2.1 Optics and Electronics 2
2.2 Sample Holder 3
2.3 Other Designs 4
2.4 Sample Preparation 4
2.5 Related Techniques 5
2.6 Lineshape Analysis 6
3 Applications 9
3.1 Investigations of the Electronic Structure of Alloys 9
3.2 Atomic Ordering in Alloys 13
3.3 Study of Ion Implantation Damage and Epitaxial Regrowth in Semiconductors 14
3.4 Corrosion Optics 18
4 Conclusion 21
Acknowledgments 22
Abbreviations and Acronyms 22
Related Articles 22
References 22

Differential reflectance spectroscopy (DRS) is a surface analytical technique. It uses ultraviolet (UV), visible, or infrared (IR) light as a probing medium. The interaction of light with ''strongly absorbing materials'', such as metals, alloys, semiconductors, etc. occurs in the first 10–20 nm. Thus the differential reflectometer probes 50–100 atomic layers into nontransparent solid surfaces. Because of the specific probing depth of light, DRS fills the gap between other surface techniques such as ion-scattering, Auger spectroscopy, and ESCA (electron spectroscopy for chemical analysis), which probe 1, 5, or even 20 monolayers, and X-ray diffraction (XRD) which probes as deep as 1–50 μm into a bulk material. The information gained by DRS is somewhat different from that obtained by the other surface techniques mentioned. A ''differential reflectogram'' reveals details about the electron structure around the Fermi surface. Specifically, the instrument allows the exact measurement of the energies which electrons absorb from photons as they are raised into a higher, allowed energy states. Since each material has a specific electron-band structure the measurement of the characteristic energies for electron ''interband transitions'' serves as a means for identifying these materials. Furthermore, investigation of any shift of these characteristic energies, which may be caused by the addition of solute elements to a solvent, by transformations, lattice defects, ordering, or the like, provides a deeper insight into the nature of the solid state. The application of DRS is, of course, not restricted to strongly absorbing materials such as metals, alloys, or semiconductors. Its strength has likewise been demonstrated in the identification and characterization of transparent or semitransparent surface layers as observed in thin-film corrosion products on metal substrates. DRS is used to scan two samples whose properties differ slightly. Thus, the main advantage of DRS over conventional optical techniques lies in its ability to eliminate any undesirable influences of oxides, windows, electrolytes, instrumental variations, and the like, owing to the differential nature of the technique. No vacuum is needed unless measurements in the vacuum UV are desired. Thus, the formation of a surface layer due to environmental interactions can be studied in situ. The measurements are fast: a complete differential reflectogram, that is, an automatic scan from the UV through the visible into the IR region, is accomplished in about 3 min.

1 INTRODUCTION

DRS, also often called differential reflectometry (DR) or differential reflection spectrometry, is a surface analytical technique.[1,2] It uses photons as a probing medium (i.e. light, varying in wavelength from the near-UV through the visible to the IR region). The interaction of these photons with strongly absorbing materials such as metals, alloys, semiconductors, etc. occurs in the first 10–20 nm. Thus, DRS probes 50–100 atomic layers into nontransparent solid surfaces. Because of this characteristic probing depth, DRS fills the gap between other surface techniques such as ion-scattering, Auger spectroscopy, or ESCA, which penetrate 1, 5, or even 20 monolayers, and XRD, which probes as deep as 1–50 μm into a bulk material.

DRS scans two samples (rather than one) whose properties (e.g. the composition of an alloy) differ slightly. The light that is reflected from this sample pair eventually impinges on a light-sensitive device such as a photomultiplier tube (PMT). After electronic processing the signal thus obtained yields a ''differential reflectogram''

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

which has the essential features of an optical absorption spectrum.

The information obtained by DRS is somewhat different from that gained by the aforementioned surface analytical techniques. Specifically, the characteristic peaks in a differential reflectogram represent the energies which electrons absorb from photons as they are raised from a lower, filled energy state into an upper, unfilled state. The peaks thus obtained are reasonably sharp; still, they are less pronounced than X-ray emission or Auger peaks, for example. This stems from the fact that the latter spectra are obtained by electron transitions between distinct, inner electron *levels* whereas optical spectra result from electron transitions between broad, outer electron *bands*. In other words, optical techniques, such as DRS, disclose photon interactions with valence (or conduction) electrons rather than core electrons and thus reveal details about the electronic structure around the Fermi surface. Since each material has a specific electron-band structure, the measurement of the characteristic energies for electron ''interband transitions'' serves as a means for identifying these materials. Furthermore, investigations of any shifts of these characteristic energies, which may be caused by the addition of solute elements to a solvent, by transformations, lattice defects, ordering, ion-implantation surface states, or the like, provide a deeper insight into the nature of the solid state.

The application of DRS is, of course, not restricted to strongly absorbing materials such as metals, alloys, or semiconductors. Its strength has also been demonstrated in the identification and characterization of transparent or semitransparent surface layers as observed in thin-film corrosion products on metal substrates.

The main advantage of DRS over conventional optical techniques lies in its ability to eliminate any undesirable influences of oxides, windows, electrolytes, instrumental variations, etc. upon a differential reflectogram, owing to the difference-forming nature of the technique. Thus, the signal-to-noise ratio is quite high, which renders the method extremely sensitive. No vacuum is needed, unless measurements in the vacuum UV are desired. Therefore, the formation of a surface layer due to environmental interactions can be studied in situ. Moreover, the measurements are fast: a complete differential reflectogram, that is, a continuous and automatic scan from the UV through the visible into the IR region, is accomplished in about 3 min.

Conventional optical reflection spectra (i.e. graphs that depict the ratio between the reflected and the incoming light intensities as a function of the photon energy) are often relatively featureless. In contrast, differential reflection spectra are distinguished by a sequence of pronounced maxima and minima which represent in certain cases the first derivative of a conventional reflection spectrum. This feature is experimentally obtained as already mentioned, by utilizing two samples whose properties differ slightly, and by employing a lock-in amplifier which provides the difference in reflectivity between these two nearly identical specimens. This will become clearer in section 2.

The present contribution describes the differential reflection spectrometer and its application to problems in solid-state chemistry, solid-state physics, electrochemistry, and materials science. It also provides the theoretical background used to interpret the experimental results by relating them to pertinent electron-band structures.

2 THE INSTRUMENT

2.1 Optics and Electronics

The differential reflection spectrometer (also called the differential reflectometer) measures the normalized difference between the reflectivities of two specimens which are mounted side by side, having virtually no gap between them. Alternatively, two slightly different parts of the same specimen may be chosen. The sample configuration depends on the particular case under investigation, as shown in section 2.4.

Unpolarized, monochromatic light, having a continuously varying wavelength, emanates from a double monochromator (illuminated, for example, by a high-pressure xenon source) and is alternately deflected to one or the other sample by means of an oscillating mirror (see Figure 1). This mirror vibrates at a convenient frequency such as 50 or 60 Hz. The cross-section of the light beam is rectangular and parallel to the boundary that separates the two samples. This beam sweeps over the two specimens perpendicular to the boundary, as shown schematically in the insert of Figure 1. The total area scanned is generally $2 \times 4\,\text{mm}^2$, but can be varied in size by regulating the voltage to the vibrator coil of the mirror and by changing the gap size of the exit slit of the monochromator. The light impinges on the sample pair under near-normal incidence. The exit slit of the monochromator is focused on these samples. Another focusing mirror, which is placed after the specimens, produces a magnified image of the vibrating mirror on a ground plate of fused silica which is placed a short distance in front of the face of a PMT. This plate diffuses the light falling on the PMT (whose sensitivity may vary over the surface), thereby minimizing the effect of the slight motion of the light spot that results from the oscillations of the vibrating mirror.

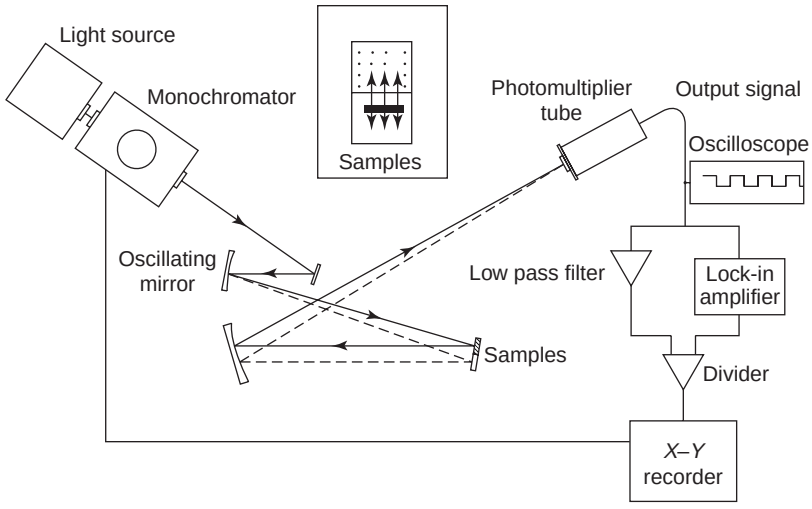

Figure 1 Schematic diagram of the differential reflectometer. For clarity, the angles of incident and reflected light beams on the sample are shown larger than they are in an actual instrument. The insert depicts the cross-section of the light beam as it scans between the two samples.

The signal from the PMT consists of a direct current (DC) component, modulated by a 60-Hz square wave. This output is split into two channels. One contains a low-pass filter which removes the 60-Hz square wave modulation from the original signal. Thus, the average of the reflectivity $\overline{R} = (R_1 + R_2)/2$ remains. This signal is held constant by a servo which regulates the high voltage applied to the PMT. The other channel is fed into a lock-in amplifier which yields an output proportional to $\Delta R = R_1 - R_2$. The lock-in is tuned to the frequency of the oscillations of the mirror. A divider circuit forms the ratio of the two signals, thus providing the normalized difference in reflectivity $\Delta R/\overline{R}$, which is fed to the Y-axis of an $X-Y$ recorder or to a computer. A potentiometer attached to the scanning gear of the monochromator delivers a DC voltage proportional to the wavelength of the light or to the photon energy. This signal is applied to the X input of the $X-Y$ recorder (or a computer). A differential reflectogram, that is, a continuous plot of $\Delta R/\overline{R}$ versus wavelength, is thus generated automatically. A scan between 200 and 800 nm (1.6–6 eV) requires about 1–3 min, depending on the scanning speed selected. The sensitivity in observing the normalized difference in reflectivity is better than 0.01%.

Measuring R_1 and R_2 at the same time and forming the ratio between ΔR and $\overline{R}$ eliminates possible influences from fluctuations of the line voltage. It also eradicates intensity variations in the spectral output of the light source, the spectral sensitivity of the detector and the spectral reflectivities of the mirrors. Further, owing to the difference-forming technique, all disturbances of the sample surfaces (such as roughness, oxidation, contamination, etc.) are eliminated as long as they are common to both sample parts.

An oscilloscope connected to the PMT output monitors the light which scans over the two samples. It allows a direct measurement of the normalized difference in reflectivities of the two specimens and reveals whether or not the light beam rests for an equal amount of time on each specimen. Any corrective measures can thus be performed.

In short, DRS is fast, sensitive and nondestructive and it provides a continuous absorption-type spectrum, does not require a vacuum or polarized light, and eliminates surface disturbances that are common to both specimen parts. The data can be recorded under near-normal incidence. The analog instrument described here can certainly be replaced by digital components in conjunction with a personal computer (PC).

2.2 Sample Holder

The samples are clamped or attached by double-sided tape on their back to a mounting stage which can be moved vertically and laterally. This allows a specific area on the sample surface to be selected and the two specimens to be centered about the scanning beam.

The circuit of the differential reflectometer is arranged to provide a positive peak in $\Delta R/\overline{R}$ if the sample on top of the mounting stage has a higher reflectivity at a given wavelength. (As a matter of consistency, an alloy with a higher solute concentration is always placed on top. In corrosion experiments, the reference specimen is placed in the upper position.)

2.3 Other Designs

A differential reflectometer using the configuration described here has been demonstrated to be the most versatile and sensitive design, though other approaches have been proposed. In one case, a stationary light beam and a rotating stage on which the two specimens are mounted is used.[3,4] The samples are then vapor-deposited on round substrates of equal size and placed in two matching holes in the spinning disk. The light beam scans a relatively large arc-shaped segment of the sample as well as a substantial portion of the blackened spinning disk. In another design, which utilizes a moving specimen stage, the samples are mounted on the vibrating ends of a tuning fork.[5] A stationary set of samples[1,2] has, however, several advantages: any sample size and sample configuration can be used; smaller areas can be investigated; the area to be studied can be specifically selected; measurements in vacuum or in an electrolyte (corrosion) are less of a problem; and differences in temperature, stress, or electric field can be effortlessly studied. Bulk specimens are essential when alloys of exact composition, high homogeneity, smooth surface, and small grain size need to be investigated, and if a good reproducibility of the results is required. Their solute content can be easily checked, using chemical or microprobe analysis. Bulk specimens may be polished by various methods, thus eliminating the rough surface that is often caused by annealing. Furthermore, small changes in compositional difference can be easily measured. (The latter point is particularly important to reduce possible variations caused by oxidation, and to obtain a more pronounced structure of the spectral dependence of $\Delta R/\overline{R}$.) Finally, bulk specimens can be heat-treated, rolled, and quenched to obtain ordered, disordered, or any other configuration.

2.4 Sample Preparation

The preparation of specimens to be studied by DRS varies somewhat with the type of experiments being carried out. For the investigations of alloys, for example, a combination of heat treatments and cold rolling is necessary to ensure absolute homogeneity of the alloy composition across the samples and to obtain a fine grain. Because of the high sensitivity of the differential reflectometer, even the smallest inhomogeneity may lead to structure in a differential reflectogram when the light beam is scanned across *one* sample only.[6] Thus, it is common to apply a solution heat treatment for about 10 days at slightly below the solidus temperature, followed by cold rolling and recrystallization at a lower temperature for a short time. Two specimens containing the same type of solute but of slightly different composition (generally, not more than 1 or 2 atom % difference) are usually embedded side by side in the same metallographic mount. The sample pair is then polished using standard metallographic procedures (ending, for example, with 1 μm diamond polishing compound on felt cloth). Since both alloys undergo identical preparation procedures at the same time, and the difference in solute is small, any possible changes of the surfaces such as deformation, oxidation, surface roughening, etc. are nearly identical and subtract out due to the differential technique. The measurements can therefore be performed in air. Any oxidation or surface roughening of the alloys decreases the peak height only slightly, but does not alter the position of the peaks on the energy scale. Embedding and polishing the two specimens in a common mount has another important advantage: the two alloys are situated in the same optical plane, causing the reflected beams to reach virtually the same point on the PMT.

For oxidation studies, a pure metal (or only *one kind* of alloy) is used. Half of this specimen is covered (after polishing) with a protective lacquer, while the other half is allowed to corrode. After that, the lacquer is peeled off, leaving the corroded and uncorroded specimens next to each other. Since the metal substrate is common to both sides, the contribution of the substrate to $\Delta R/\overline{R}$ cancels out due to the differential nature of the technique if the corrosion product film is thin enough. Thus, the features in the differential reflectogram are essentially only those of the corrosion product.

For electrochemical corrosion studies a more advanced method is used. Rather than utilizing *one* specimen, the metal slab in this case is divided into *two* parts which are electrically insulated from each other by a thin insulating foil. Different potentials can then be applied to each side. For example, one specimen half may be held at the protective potential (reference) and the other at the corrosion potential. The main advantage of this technique is that neither the corrosion cell window nor the electrolyte solution affects the results because of the nature of the differential reflectometer. In situ studies are therefore effortlessly achieved without the use of a protective layer on the reference half and without removing the specimen from the corrosion cell for measurements.

For high-temperature oxidation studies, the *entire* specimen is corroded. Subsequently, the corrosion product is removed from part of the specimen by a "half polishing" technique.

Samples suitable for the study of ion-implantation damage are prepared by covering half of a commercial silicon wafer with a solid mask during the implantation process to shield that part of the underlying wafer from implantation. As a result, an implanted specimen next to a reference is readily obtained without having to deal with two separate samples and possible alignment problems. The light from the differential reflectometer is then scanned between these two specimen parts.

2.5 Related Techniques

DRS is a so-called ''modulation spectroscopy'' technique; such techniques provide the derivative of the spectral reflectivity (or of the imaginary part of the dielectric constant, ε_2) with respect to an external parameter.[7] The features obtained by modulation techniques are restricted to so-called critical-point interband transitions in the electronic band structure, i.e. they emphasize specific electron transitions from an essentially featureless background. This background is caused by many allowed electron transitions at a wide range of points in the Brillouin zone. Most modulation techniques, such as DRS, wavelength modulation, thermoreflectance or piezoreflectance, are first-derivative techniques (Figure 2a). In semiconductor research, another modulation technique called electroreflectance is often used, which provides the third derivative of R or ε_2. It utilizes an alternating electric field which is applied to a semiconducting material during the reflection measurement through either a Schottky barrier or a semiconductor/electrolyte junction.[8] The third derivative provides a set of three peaks for each interband transition (Figure 2b) whereas a first-derivative technique yields essentially only one (Figure 2a). A relatively involved line-shape analysis of electroreflectance spectra eventually yields the interband transition energies. In a first-derivative modulation spectrum, the lattice periodicity is retained, the optical transitions remain vertical and the interband transition energy changes with the perturbation (see insert of Figure 2a). In electromodulation, the formerly sharp vertical transitions are spread out over a finite range of initial and final moments[9] (see insert of Figure 2b). This causes the previously mentioned three peaks in an electroreflectance spectrum (Figure 2b).

A technique similar to DRS has been named reflectance difference spectroscopy (RDS).[10] Here, polarized light at near-normal incidence ($\alpha \approx 10^\circ$) is reflected from *one* sample only which is spun at 50 or 60 Hz in front of the light beam. This way any optical anisotropies that may exist on the sample surface have been shown to produce essentially identical absorption spectra to those from the two-sample technique described previously. However, optical measurements that involve sample motion are particularly sensitive to macroscopic imperfections that scatter light[10] and their general applicability is limited, as outlined in section 2.3.

Another technique related to DRS is called surface differential reflectivity spectroscopy (SDRS);[11] in this technique the sample to be investigated is placed inside a high-vacuum chamber and the reference sample is located outside the chamber. By switching the light beam between these two samples, the difference in reflectivity between clean (cleaved) and gas (air)-exposed surfaces is obtained as a function of the impinging photo energy.

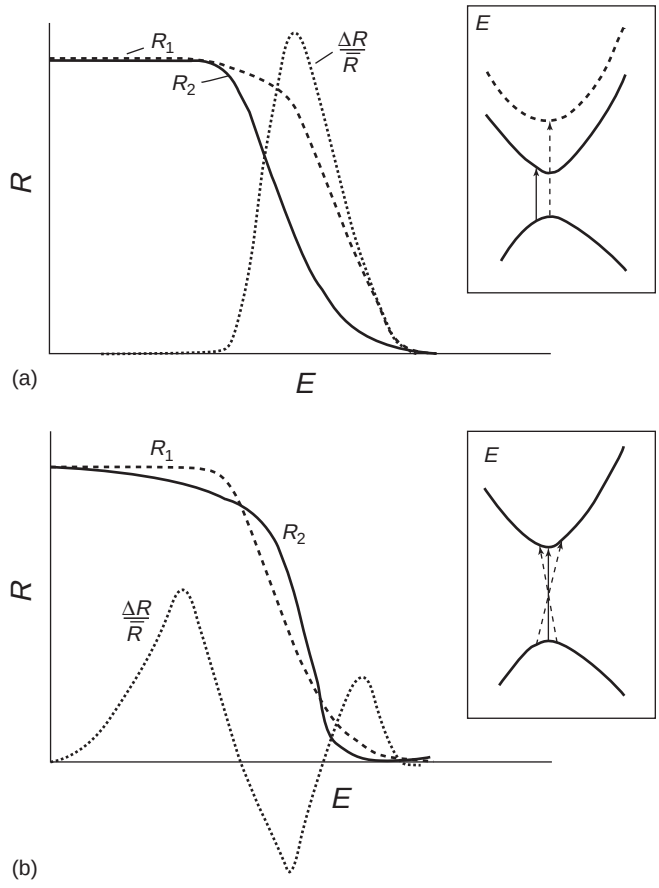

Figure 2 Schematic representation of (a) a first derivative and (b) a third derivative of a reflectivity spectrum. The equivalent interband transitions at a so-called M_0 symmetry point are shown in the inserts.

The window to the vacuum chamber needs to be of high quality and also needs to be optically isotropic, since only one beam interacts with this window and it may otherwise introduce a characteristic spectrum by itself.

In the past decade spectroscopic ellipsometry (SE), featuring for example a rotating analyzer, has come of age.[12] This technique analyzes the elliptically polarized light which is reflected from a material. It yields the phase difference between the two mutually perpendicular, linearly polarized waves and the angle between these components (the ''azimuth''). From these parameters the optical constants (index of refraction and damping constant or the real and imaginary dielectric constants) can be calculated for each wavelength. Besides being a nonderivative technique (section 2.1) this technique suffers from disturbances which affect the properties of the free surface, such as interactions with the environment, from mechanical deformation, polishing, native oxides, etc. As outlined in the literature[13] DRS offers some advantages over

SE, such as simplicity of the hardware, no need for polarization optics, and no need for precise optical alignment. Further, DRS requires only one optical window when in situ investigators are conducted, whereas SE needs two identical polarization-free optical-quality ports. DRS operates at near-normal incidence whereas in SE a 70° incidence is commonly used to increase the sensitivity.

2.6 Lineshape Analysis

As mentioned in section 2.1, the structure in differential reflectograms provides the energies which certain electrons absorb from photons when they are raised from lower, occupied energy states into higher, empty, and allowed energy states. Moreover, lifetime-broadening energies can be deduced from differential reflectograms. (The initial and the final states of interband transitions are generally not sharp, but are somewhat broadened because of the limited time an electron remains at an excited energy level.) The energies of the maxima in $\Delta R/\overline{R}$ give in many cases a relatively good and sufficiently accurate value for interband transition energies. However, for more refined interpretations, a thorough line-shape analysis needs to be performed.[14,15] This will be demonstrated now for the case of ''compositional modulation'' in which the solute content between two α-phase copper-based alloys is assumed to be different by a small amount (e.g. 1 atom %). The resulting differential reflectogram is then equivalent to the response from *one* sample having a composition $(x_1 + x_2)/2$ and which is modulated periodically

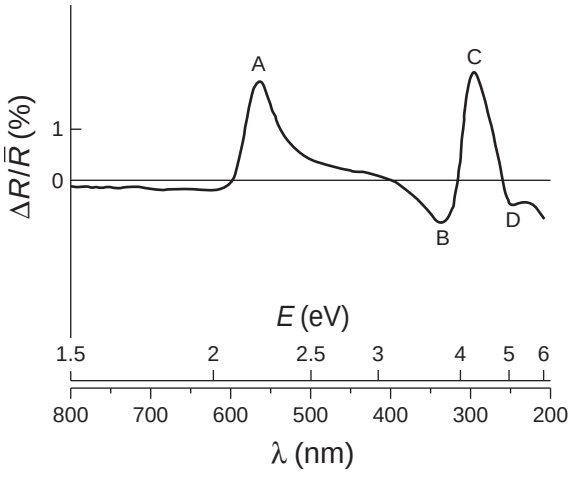

Figure 3 Experimental differential reflectogram ($\Delta R/\overline{R}$ versus photon energy E) for a Cu–1.5 atom % Ga alloy (average alloy composition: the two specimens consisted of a Cu–1 atom % Ga and a Cu–2 atom % Ga alloy respectively). (Reproduced from R.E. Hummel, R. Enderlein, *Phys. Rev. B*, **29**(4), 1529–1533 (1984). Copyright 1984 by the American Physical Society.)

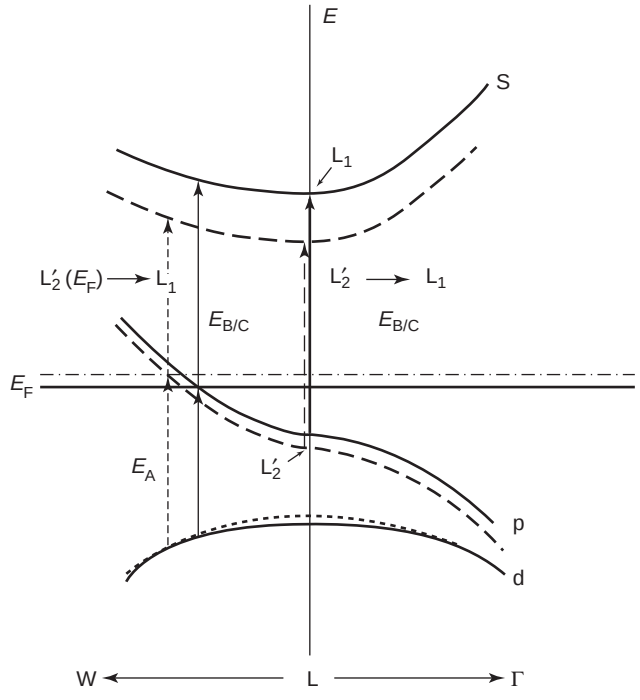

Figure 4 Schematic band structure around the L symmetry point for copper (solid lines) and an assumed dilute copper-based alloy (dashed lines). Pertinent electron interband transitions are shown by vertical arrows. Note: only the upper d-band is shown. (Reproduced from R.E. Hummel, *Electronic Properties of Materials*, 2nd edition, Figure 13.18, 1993. Copyright Springer-Verlag GmbH & Co.)

by a compositional increment $\pm\Delta x/2$. As an example, Figure 3 depicts a differential reflectogram typically found for compositional modulation of many α-phase copper-based alloys. Four distinct maxima or minima are observed, which are labeled as peaks A through D. They are caused by interband transitions from d-bands below the Fermi level to free-electron-like L_2' bands just above the Fermi level, and by transitions from the same L_2' band into the L_1 band well above the Fermi level, as shown in Figure 4.[16] The change Δx of an alloy having the composition x affects both types of transitions for the same reasons, which are (1) a shift of the energy bands, (2) a change of lifetime broadening, and (3) a shift of the Fermi level with respect to the bottom of the L_2' band. These shifts are indicated in Figure 4 by dashed lines. As a result of these perturbances, a change of the complex dielectric function occurs, as shown in Equation (1).

$$\Delta\varepsilon = \frac{\mathrm{d}\varepsilon}{\mathrm{d}x}\Delta x \tag{1}$$

This variation is related to the relative change $\Delta R/\overline{R}$ by means of Equation (2),

$$\frac{\Delta R}{\overline{R}} = \left(\alpha\frac{\mathrm{d}\varepsilon_1}{\mathrm{d}x} + \beta\frac{\mathrm{d}\varepsilon_2}{\mathrm{d}x}\right. \quad \Delta x \tag{2}$$

where α and β are weighting factors, called the Seraphin coefficients[8] and

$$\varepsilon_1 = n^2 - k^2 \tag{3}$$

$$\varepsilon_2 = 2nk \tag{4}$$

ε_1 and ε_2 are the real and imaginary parts of the complex dielectric function[16]

$$\varepsilon = \varepsilon_1 - \mathrm{i}\varepsilon_2 \tag{5}$$

Using the spheric parabolic band model one can write[16]

$$E_{\mathrm{d,s}}(\boldsymbol{k}, x) = E^0_{\mathrm{d,s}}(x) + \frac{\boldsymbol{k}^2\hbar^2}{2m_{\mathrm{d,s}}} \tag{6}$$

The Fermi energy is varied through solute additions according to

$$E_{\mathrm{F}}(x) = E^0_{\mathrm{F}}(x) + E^0_{\mathrm{s}}(x) \tag{7}$$

where $E^0_{\mathrm{F}}(x)$ is the Fermi energy with respect to the bottom of the s-band. The interband part of the dielectric function, ε (which is assumed to be the only part relevant in the spectral region under consideration), is expressed by Equation (8).

$$\varepsilon = -\frac{4\pi e^2 |p_{\mathrm{sd}}|^2}{\hbar w^2 m^2} \quad \frac{\mathrm{d}^3\boldsymbol{k}}{4\pi^3} \quad \frac{f(E_{\mathrm{d}}(\boldsymbol{k}, x)) - f(E_{\mathrm{s}}(\boldsymbol{k}, x))}{\hbar w - [E_{\mathrm{s}}(\boldsymbol{k}, x) - E_{\mathrm{d}}(\boldsymbol{k}, x)] + \mathrm{i}\hbar\Gamma} \tag{8}$$

Here, $f(E)$ is the Fermi distribution, p_{sd} is the interband momentum matrix element, and $\hbar\Gamma$ the lifetime broadening energy. The derivative $\mathrm{d}\varepsilon/\mathrm{d}x$ of Equation (8) in combination with Equation (2) yields Equation (9).

$$\frac{\Delta R}{\overline{R}} = A\frac{\mathrm{d}}{\mathrm{d}x}[E^0_{\mathrm{F}}(x) + E^0_{\mathrm{s,d}}(x)]F\left(\frac{w - w_{\mathrm{T}}}{\Gamma}, \theta\right. \tag{9}$$

where

$$F(s, \theta) = \sin\varphi\cos(\varphi - \theta) \tag{10}$$

$$\varphi = \arcsin\left(\frac{1}{\sqrt{1 + s^2}}\right. \tag{11}$$

$$A = \frac{kC}{\Gamma}\Delta x \tag{12}$$

$$\theta = \arctan\left(\frac{\beta}{\alpha}\right. \tag{13}$$

and

$$w_{\mathrm{T}} = \frac{E_{\mathrm{T}}}{\hbar} \tag{14}$$

C and θ are the absolute value and argument of the complex number $\alpha - \mathrm{i}\beta = C\mathrm{e}^{\mathrm{i}\theta}$, respectively, and s is proportional to the photon energy.

Equation (9) implies that $\Delta R/\overline{R}$ is proportional to the function $F(s, \theta)$. Figure 5 depicts $F(s)$ for different values of θ. Comparison of Figure 5 with Figure 3 shows that the experimentally obtained spectral dependence of $\Delta R/\overline{R}$ in the vicinity of peak A resembles the spectral dependence of $F(s)$ for θ values between 0° and 90°. For $\theta = 90^\circ$, $F(s)$ is symmetrical about $s = 0$ and the peak energy E_{m} is equal to the transition energy E_{T}. This type of curve is identical with the spectral dependence of ε_2 which is obtained by solving the classical Lorentz equation for harmonic oscillators.[12] For $\theta = 0^\circ$, $F(s)$ is asymmetrical about $s = 0$ and E_{T} lies in the center between minimum and maximum. This curve resembles the classical spectral dependence of ε_1 which is likewise obtained from the Lorentz equations. For features B and C in Figure 3, θ may be assumed to be close to 0°, which leads to a transition energy somewhere between minimum B and maximum C. Experimental observations suggest that features B and C are indeed part of *one* transition: when the energies of peaks B and C are plotted versus the solute concentration, both curves are essentially parallel to each other.[17] The inflection points in the curves of Figure 5 can then be used for the characterization of E_{T}.

Most experimental differential reflectograms have curve slopes which suggest θ values other than 0° and 90°. These θ values and the appropriate transition energies E_{T} can be found by curve fitting in the following way. In Figure 6, part of an experimental differential reflectogram (in the vicinity of peak A) is shown. The minimum of $(\Delta R/\overline{R})$ is used as the zero point of the vertical axis. Then three critical frequency values are extracted: w_{m} is the frequency at which $(\Delta R/\overline{R})$ has its maximum value; $w' < w_{\mathrm{m}}$ is the frequency where $(\Delta R/\overline{R})_{w'} = 1/2(\Delta R/\overline{R})_{\mathrm{m}}$. Finally, w'' is $w_{\mathrm{m}} + (w_{\mathrm{m}} - w')$. The ratio $(\Delta R/\overline{R})_{w''}/(\Delta R/\overline{R})_{\mathrm{m}}$ is a function of θ only and is plotted in Figure 7. Once θ is known, w_{T} (and thus E_{T}), can be found as a value between w' and w_{m} as shown in Equation (15).

$$\left(\frac{\Delta R}{\overline{R}}\right._{w\mathrm{T}} = \frac{1}{2}(1 + \sin\theta)\left(\frac{\Delta R}{\overline{R}}\right._{\mathrm{m}} \tag{15}$$

The lifetime broadening can be calculated using Equation (16).

$$\Gamma = (w_{\mathrm{m}} - w_{\mathrm{T}})\tan\left(\frac{1}{2}\theta + \frac{\pi}{4}\right. \tag{16}$$

For illustration, a lineshape analysis, as described above, has been applied to an experimental differential reflectogram for a copper–3.5 atom % aluminum alloy. Figure 8 shows the experimental spectral dependence of $\Delta R/\overline{R}$ in the vicinity of peak A (solid line) and some

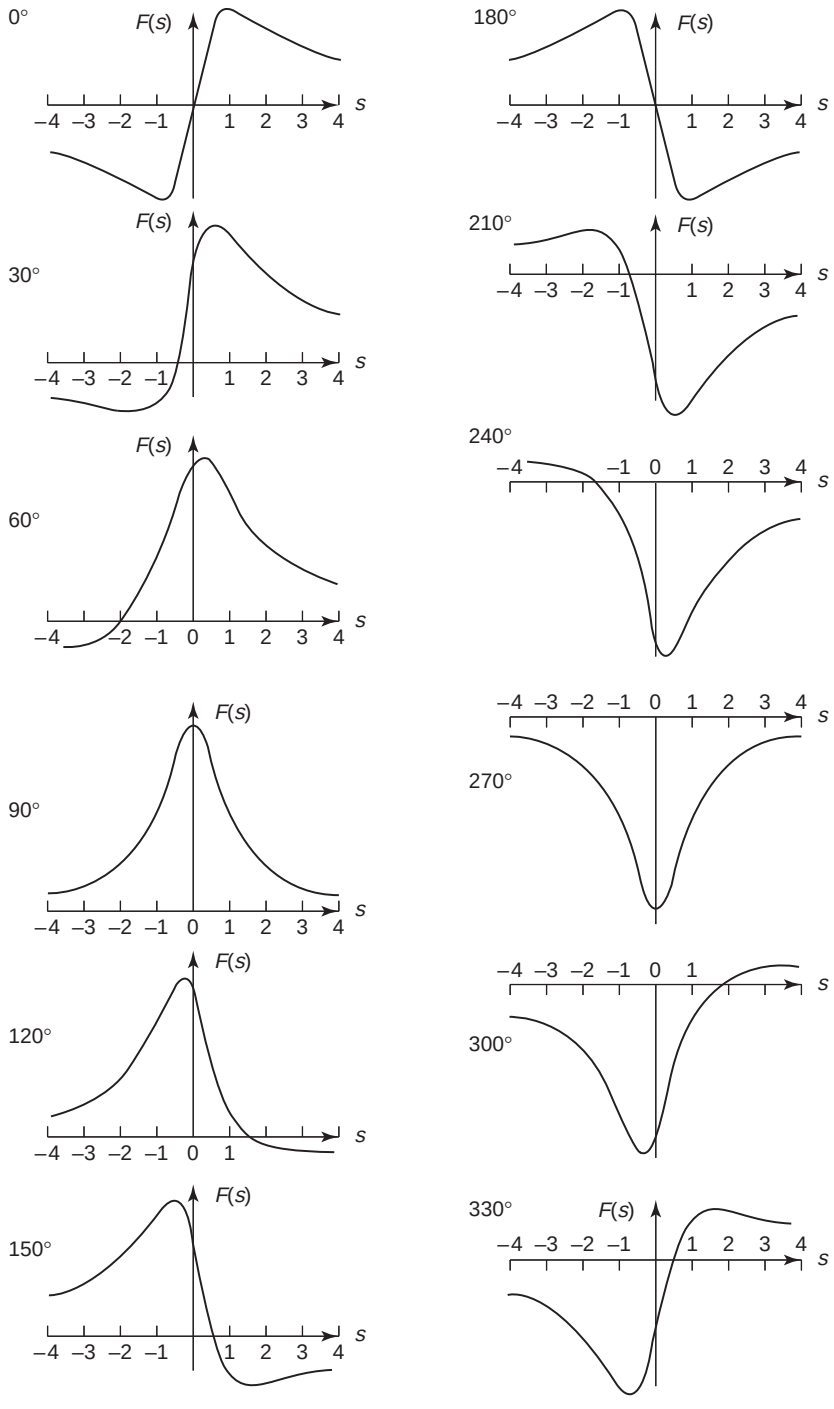

Figure 5 $F(s)$ versus s for selected θ values (s is proportional to the photon energy $E = \hbar w$). (Reproduced from Enderlein et al.[14] by permission of Wiley–VCH Verlag.)

calculated $\Delta R/\overline{R}$ values (dots) for $\theta = 55°$. The agreement is remarkably good. One can state therefore, with reasonable confidence, that peak A and the substructure around this peak are caused by electron transitions from d-band to s-states just above the Fermi energy level. Comparison of the experimental transitional energy E_T with that deduced from calculated band diagrams[18,19] suggests that peak A is caused by transitions from the *upper* d-bands to the Fermi energy level.

Experience has shown that the difference between E_m and E_T is usually only between 0.02 and 0.06 eV[20] – a difference which is too small to be recognized in current

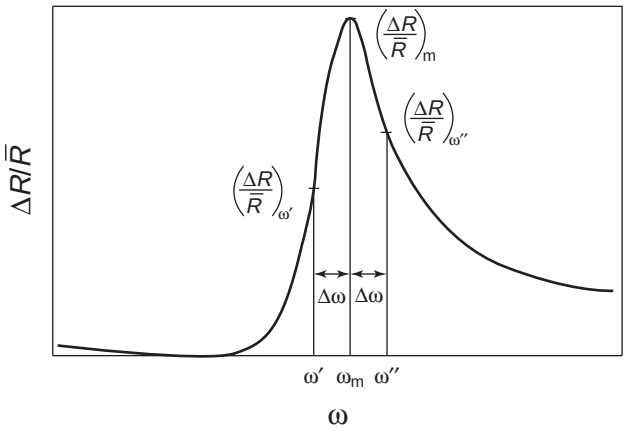

Figure 6 Schematic representation of a differential reflectogram in the vicinity of peak A. Specific w and $(\Delta R/\bar{R})$ values are shown, which are extracted from the differential reflectogram. This procedure allows θ to be determined. The photon energy is $E = \hbar w$. (Reproduced from Enderlein et al.[14] by permission of Wiley–VCH Verlag.)

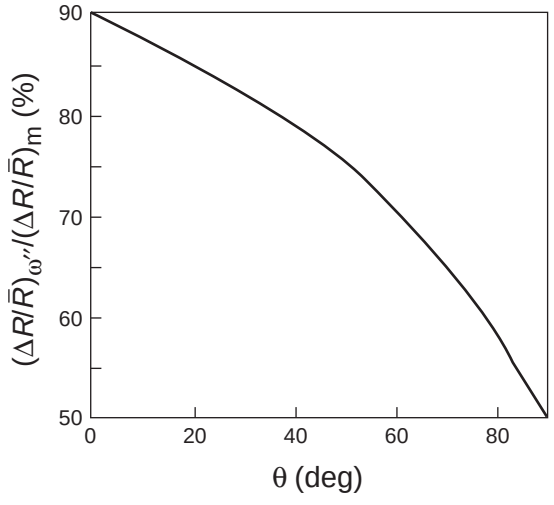

Figure 7 Calculated dependence between θ and $(\Delta R/\bar{R})_{w''}/(\Delta R/\bar{R})_m$. (Reproduced from Enderlein et al.[14] by permission of Wiley–VCH Verlag.)

band calculations. Thus, the peak energy E_m can be used in many cases as a good approximation for the true transition energy, E_T.

The lineshape in the vicinity of peaks B and C has also been analyzed by considering the $L_2'(E_F) \rightarrow L_1$ as well as the $L_2' \rightarrow L_1$ transitions (Figure 4). Using transition energies and other variables as adjustable parameters, experimental differential reflectograms can be essentially reproduced.[15] The ''hand fitting'' exemplified here can certainly be replaced by analytical least-square methods run on a PC.

In summary, this section has demonstrated, for dilute copper-based alloys, that the structure observed in differential reflectograms can be utilized to extract the energies for electron interband transitions. This technique is, however, much more widerly applicable, as will be demonstrated in the next section using a number of characteristic examples.

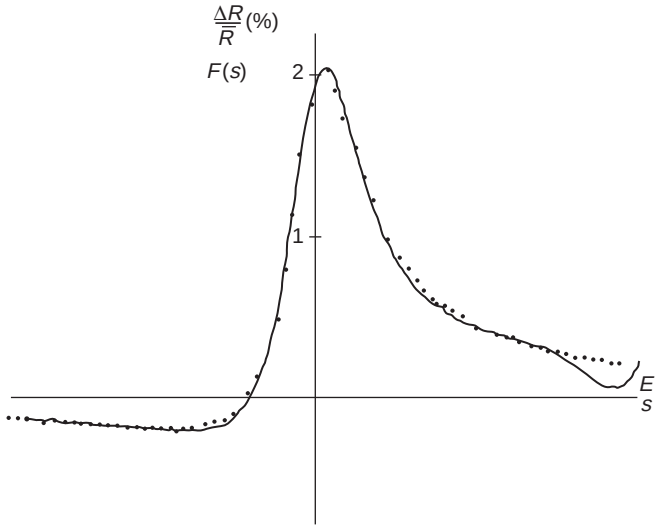

Figure 8 Experimental differential reflectogram, $\Delta R/\bar{R}$ versus E (solid line), and calculated $F(s)$ versus s (dots) for a Cu–3.5 atom % Al alloy. (Reproduced from Enderlein et al.[14] by permission of Wiley–VCH Verlag.)

3 APPLICATIONS

The differential reflectometer has been successfully used in a number of research fields such as the study of the electronic structure of alloys,[20] order–disorder transformations,[9,14] and investigations of surface films,[11] corrosion,[12] and the effects of ion-implantation in semiconducting materials.[17] Some selected applications are presented here as an illustration of the capabilities of DRS in scientific research and nondestructive industrial production control.

3.1 Investigations of the Electronic Structure of Alloys

The differential reflectograms for copper alloys containing zinc, gallium, aluminum, tin, silicon, or germanium are qualitatively similar;[20] thus reflectograms of only one of these binary systems are presented and discussed in detail here as an example. In Figure 9, a series of $\Delta R/\bar{R}$ versus E curves taken on copper–zinc alloys are shown, in which the zinc content increases from top to bottom. The characteristic peaks A through D, mentioned in section 2.6, can be readily recognized. One observes immediately that the energies of these peaks shift with variation of the solute

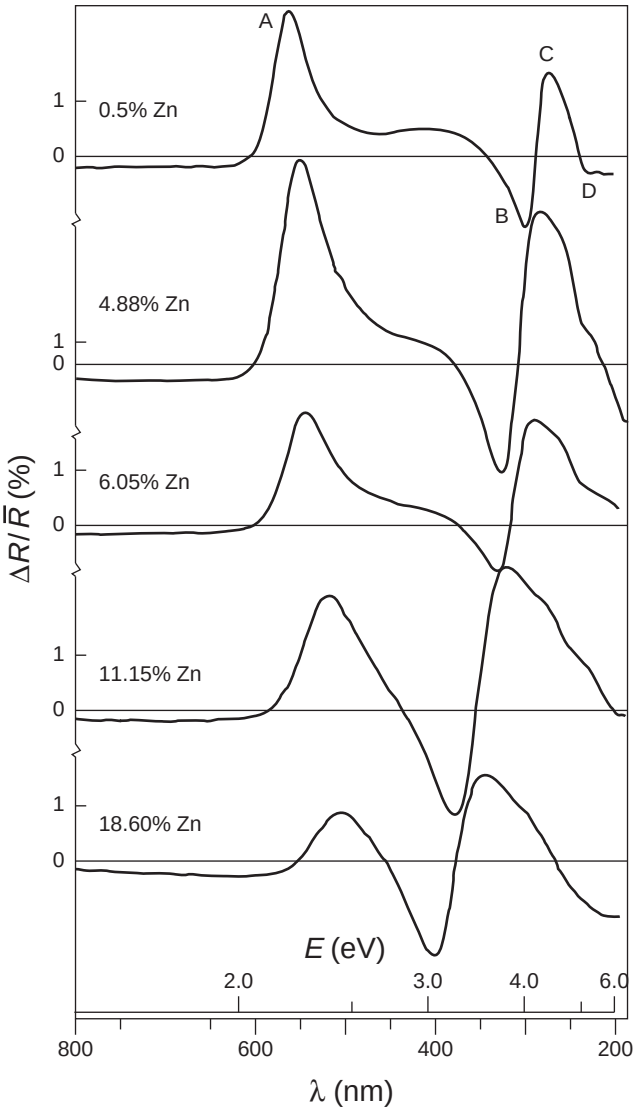

Figure 9 Experimental differential reflectograms for various bulk copper–zinc alloys. The parameter on the curves is the average zinc concentration of the two alloys in atom %. (The curve marked 0.5% for example was obtained by scanning the light beam between pure copper and a Cu–1 atom % Zn alloy. The compositional difference between the two alloys was characteristically 1 atom %.) The individual curves have been shifted for clarity. (Reproduced from R.J. Nastasi-Andrews, R.E. Hummel, *Phys. Rev. B*, **16**(10), 4314–4323 (1977). Copyright 1977 by the American Physical Society.)

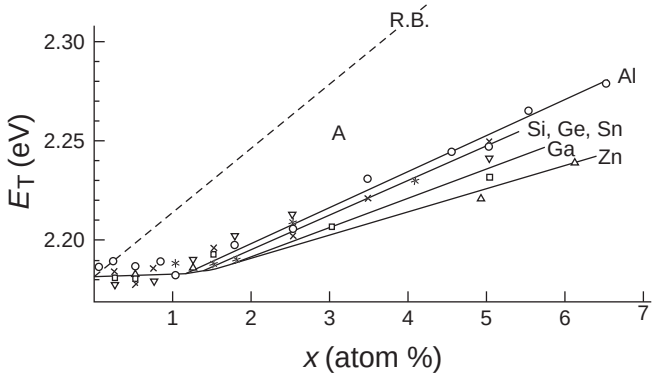

Figure 10 Threshold energies, E_T, for interband transitions for various copper-based alloys as a function of solute content, x (△ Zn, Ga, ○ Al, ▽ Sn, × Si, ∗ Ge). The E_T values are extracted from differential reflectograms similar to those shown in Figure 9. For this the structure around peak A is used in conjunction with a lineshape analysis, as shown in Figure 6. The difference between E_T and the energy corresponding to the maximum of peak A (E_m) is about 0.03 eV. The rigid band line (R.B.) for Cu–Zn is added for comparison. (Reproduced from R.J. Nastasi-Andrews, R.E. Hummel, *Phys. Rev. B*, **16**(10), 4314–4323 (1977). Copyright 1977 by the American Physical Society.)

concentration. In particular, peak A, at about 2.2 eV, is seen to shift to higher electron energies. This peak was identified in section 2.6 as resulting from electron transitions between the upper d-bands and the conduction bands just above the Fermi surface. Specifically, they occur near the L symmetry point and also, to a lesser extent, near X and other points.[21] In Figure 10, this shift in transition energy, E_T, is shown (along with E_T values measured on other copper-based alloys) as a function of solute content, x. Essentially a linear increase of E_T with increasing x is observed. The slopes are found to be steeper for solutes having successively larger electron-to-atom ratios (with the exception of copper–aluminum alloys).

The rise in energy difference between the upper d-band and the Fermi level caused by solute additions can be explained in a first approximation by suggesting a rise in the Fermi energy, which results when extra electrons are introduced into the copper matrix from the higher-valence solutes (rigid band model).[22] The slopes in the $E_T = f(x)$ curves in Figure 10 are, however, considerably smaller than those predicted by the rigid band model. These results suggest, therefore, that the d-bands are likewise raised with increasing x and/or that the Fermi level is shifted up much less than anticipated. Band calculations[23] substantiate this suggestion. They reveal that upon solute additions to copper, the d-bands become narrower (which results from a reduction of Cu–Cu interactions) and are lifted up as a whole. Furthermore, the calculations show that solute additions to copper cause a rise in E_F and a downward shift of the bottom of the s-band. Figure 4 reflects these results. Because of the lowering of the s-band, the Fermi level rises much less than would be expected if the s-band had remained constant.

An important characteristic of all $E_T = f(x)$ curves is that the threshold energy for interband transitions, E_T, does not vary appreciably for solute concentrations up to slightly above 1 atom % (Figure 10). Friedel[24] predicted

just this type of behavior and related it to ''screening'' effects. He argued that for the first few atomic percent solute additions to copper, the additional charge from the higher-valence solute is effectively screened and the copper matrix behaves as if the impurities were not present. The matrix remains essentially unperturbed as long as the impurities do not interact.

We turn now to electron transitions which cause peak D in the differential reflectograms and which take place when electrons absorb approximately 5 eV from the impinging photons. This peak involves transitions

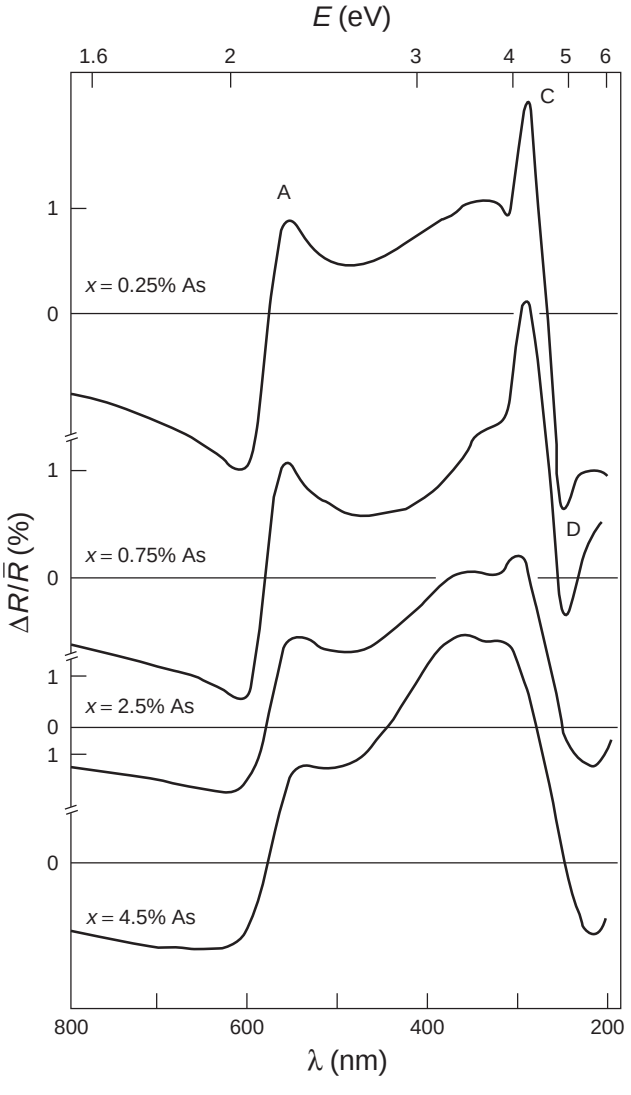

Figure 11 Experimental differential reflectograms for various bulk copper–arsenic alloys. The parameter is the average arsenic content of the two alloys in atom %. The difference between the two alloy compositions is equal to or smaller than 1 atom %. (Reproduced from R.J. Nastasi-Andrews, R.E. Hummel, *Phys. Rev. B*, **16**(10), 4314–4323 (1977). Copyright 1977 by the American Physical Society.)

from the lower d-bands to the Fermi energy level[19,25] (not shown in Figure 4). Thus, considerations can be employed that are similar to those used to explain the 2.2 eV peak. Two interesting variations should be pointed out, however. First, for photon energies around 5 eV, the β/α ratio is negative (see section 2.6) which results in a lineshape that is inverted with respect to the one observed in the vicinity of peak A. Second, it has been found[20,26] that in copper-based alloys in which the solvents possess a large valence electron concentration, the transition $E_{\mathrm{d,lower}} \rightarrow E_{\mathrm{F}}$ is much more pronounced than in alloys containing solutes with a small number of valence electrons. (This may be attributed to an increase in the density of states of the d-bands at this energy for alloys with large valence electron concentration.) For example, in Figure 11 (copper–arsenic), feature D is a sharp minimum, whereas in Figure 9 (copper–zinc) the same feature consists merely of a shoulder. Thus, copper alloys containing silicon, germanium, or arsenic are used preferentially to study the $E_{\mathrm{d,lower}} \rightarrow E_{\mathrm{F}}$ transition.

In Figure 12, the energy of minimum D is depicted as a function of solute concentration for various diluted Cu-based alloys. This plot is similar to the one in Figure 10 which suggests that the lower d-band $\rightarrow E_{\mathrm{F}}$ transitions behave in a similar (but not identical) manner to upper d-band $\rightarrow E_{\mathrm{F}}$ transitions. Specifically, one observes only a very small shift in transition energy up to about 1 atom % solute, followed by a steady rise in interband energy amounting 16×10^{-2} eV per atom % solute. The slope in Figure 12 is about ten times steeper than in Figure 10 ($E_{\mathrm{d,upper}} \rightarrow E_{\mathrm{F}}$ transition) in which E_{T} increases by only 1.2×10^{-2} eV per atom % solute. This variation can only be attributed to a difference in the behavior of the upper compared with the lower d-bands. Thus, the results seem to indicate that the lower d-bands (not shown in Figure 4), at the points of origin for the electron transitions, move only small amounts upon alloying.

One more structural feature is contained in the differential reflectograms shown in Figure 9. It is of the ε_1 type (see section 2.6) and involves peaks B and C. This structure around 4 eV has also been ascribed in section 2.6 to transitions near the L symmetry point (see Figure 4[19,27]). Chen and Segall[25] found that the $L_2'(E_{\mathrm{F}}) \rightarrow L_1$ transition energy for pure copper is 4.26 eV, which is exactly halfway between peaks B and C (for low solute concentrations). This confirms the suggestion made in section 2.6 that the transitional energy contained in features B and C has to be taken at the inflection point between these peaks.

In Figure 13 the energy of peak B is shown to decrease sharply with increasing solute concentration, suggesting a strong lowering of particularly the L_1 symmetry point due to alloying (Figure 4). Band calculations confirm this suggestion.[28] The transitions involving

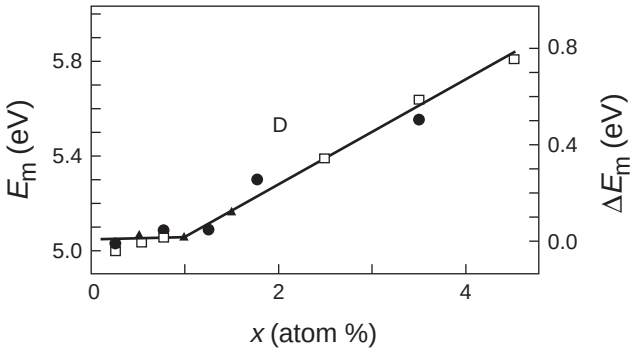

Figure 12 Energy (E_m) of peak D and shift in energy (ΔE_m) compared with pure copper as a function of composition for various copper-based alloys (● Si, As, Ge). (Reproduced from R.J. Nastasi-Andrews, R.E. Hummel, *Phys. Rev. B*, **16**(10), 4314–4323 (1977). Copyright 1977 by the American Physical Society.)

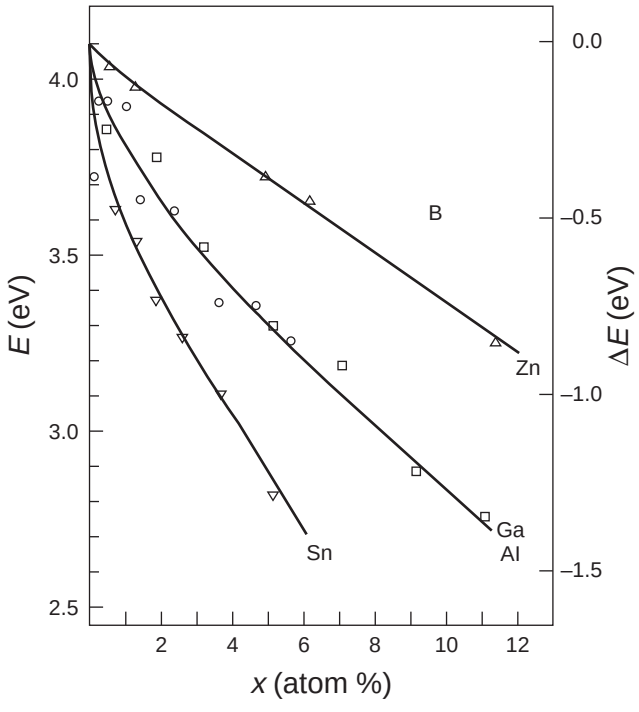

Figure 13 Energy of peak B and shift in energy compared with pure copper as a function of composition for various copper-based alloys (△ Zn, ○ Al, Ga, ▽ Sn). (Reproduced from R.J. Nastasi-Andrews, R.E. Hummel, *Phys. Rev. B*, **16**(10), 4314–4323 (1977). Copyright 1977 by the American Physical Society.) A similar shift is found for peak C.[17]

the conduction bands are affected by solute effects, for example lattice dilations. Different solute elements having different atomic radii must, therefore, influence the $L_2' \rightarrow L_1$ transitions in different ways. On the other hand, Figure 13 suggests that these transitions are also strongly influenced by the electron concentration of the solute: the decrease of the energy of peak B is more substantial for alloys which contain solutes with a larger valence electron concentration. Further, Cu–Ga and Cu–Al alloys, which possess identical electron concentrations per atom, behave alike.

It is important to note that in copper-based alloys, containing silicon, germanium, or arsenic,[20] having solute concentrations of more than 1 atom % minimum B (that is, the 4.2 eV absorption edge) becomes less distinguishable (Figure 11), which may suggest that for higher electron concentrations the $L_2(E_F) \rightarrow L_1$ transition loses some strength.

One further piece of information can be taken from the differential reflectograms. It pertains to the ''lifetime broadening energy''. The initial and the final states of any interband transition are generally not sharp, but are somewhat broadened because of the limited time an electron remains at an excited energy level (see section 2.6). This lifetime broadening can be calculated using Equation (16) and has been found to increase from $1.2 \times 10^{14}\,s^{-1}$ for pure copper to $1.9 \times 10^{14}\,s^{-1}$ for α-phase copper alloys containing about 9 atom % of the solute.[20] The lifetime broadening energy $\hbar\Gamma$ also increases with increasing solute concentration (after an initial flat portion) up to approximately 1 atom % solute (Figure 14).

In summarizing this section, it can be stated that differential reflectometry is easily capable of identifying the electron transitions for copper–zinc and similar alloys. The results presented here confirm convincingly the findings that have been obtained by theoretical means. Furthermore, the information contained in Figures 9 through 14 add quantitative, experimental data about finer details of the changes in the electron configuration which happen due to the alloying process. We have seen that in the energy range between 1.5 and 6.2 eV electron transitions occur between the upper d-bands and the

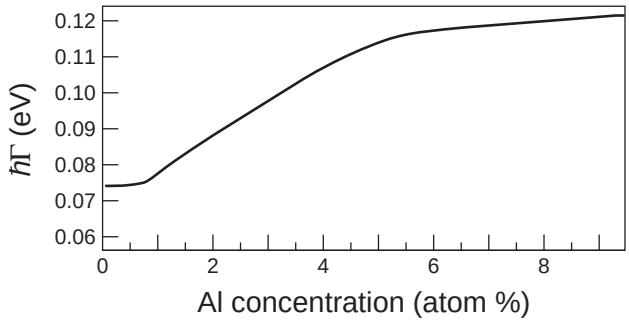

Figure 14 Lifetime broadening energy $\hbar\Gamma$ calculated from experimental compositional modulation spectra for copper–aluminum alloys, as a function of Al concentration (see section 2.6). (Reproduced from Enderlein et al.[14] by permission of Wiley–VCH Verlag.)

Fermi energy level, between the lower d-bands and the Fermi energy level, and between some conduction bands.

In closing, it should be mentioned that differential reflection investigations have been also performed on copper–gold alloys,[20] copper–nickel alloys,[17,29] gold–iron alloys,[30] nickel-based alloys,[31,32] silver–aluminum alloys,[33,34] and copper–cobalt alloys.[35]

3.2 Atomic Ordering in Alloys

DRS is quite useful for the study of order–disorder phenomena. For these investigations, two specimens of the same alloy are utilized. One is in the ordered state, the other in the disordered state. The light beam of the reflectometer scans the two samples. This creates differential reflectograms which show peaks whenever the ordered state allows additional interband transitions.

In long-range ordered alloys, such as in the extensively studied intermetallic phase Cu_3Au, the two kinds of atoms are periodically arranged on regular lattice sites and thus form a superlattice whose presence can be discovered by additional X-ray lines, or measured indirectly by changes in hardness, resistivity, Hall constant, or other physical properties.[36]

Ordering has a definite effect on the electronic structure of alloys.[37–39] It has been proposed[40] that the formation of a superlattice in Cu_3Au is accompanied by a folding of energy bands into a new, simple, cubic Brillouin zone which causes additional interband transitions.

Figure 15 depicts a differential reflectogram involving ordered versus disordered Cu_3Au. Several pieces of structure are observed. Firstly, a ''down–up'' curve between 3 and 4 eV, that is, an ε_1-type structure, is clearly visible involving a minimum at 3.29 eV and a maximum at 3.85 eV. As outlined in section 2.6, the energy for interband transitions for an ε_1-type structure is generally located near the center, i.e. between the minimum and the maximum, similar to those shown in Figure 5 for $\theta = 0^\circ$. For the case of Cu_3Au (Figure 15), this suggests a transition energy around 3.6 eV in complete agreement with conventional, that is, unmodulated, optical measurements.[37–39] This structure has been assigned to transitions from Γ_{25}^1 to X_4^1 (unfolded)[5] or from the top of the Cu 3d-band to conduction band states near Γ.[37]

Secondly, the maximum at 2.17 eV is of interest because its lineshape and peak energy closely resemble those observed for peak A for compositional modulation (Figure 3). It is suggested, therefore, that this peak is caused by transitions from the copper 3d-bands to s-states just above the Fermi level, similar to those depicted in Figure 4. Skriver and Lengkeek[37] arrive at a comparable conclusion based on their relativistic energy band calculations. However, splitting of the s-bands at

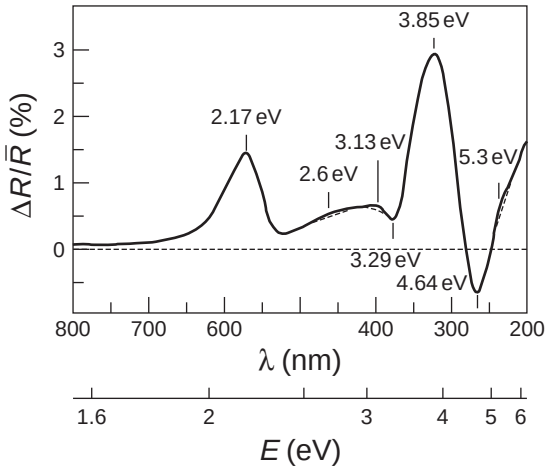

Figure 15 Experimental differential reflectogram of (long-range) ordered versus disordered Cu_3Au. Ordering was accomplished by vacuum annealing Cu_3Au at 340 °C and quenching into water at room temperature. The disordered state was achieved by quenching the sample from 500 °C. (Reproduced from Hummel et al.[6] by permission of Carl-Hanser Verlag.)

the point of the $d \rightarrow E_F$ transition is an alternative way for interpreting the peak at 2.17 eV.

Thirdly, weak structures occur near 3.13 and 5.3 eV in Figure 15. These peaks have comparable transition energies to similar peaks observed in compositional modulation spectra for copper–gold.[20] They have been assigned to transitions from the Au d-bands and the lower Cu d-bands to the Fermi level, respectively, that is, to newly created transitions from the copper and gold d-bands to the Fermi surface.[20]

Finally, it should be noted that $\Delta R/\bar{R}$ does not become zero at small energies, as it does for dilute copper alloys, which indicates electron transitions in the IR. This has indeed been observed,[5,37] and has been interpreted to be due to high-lying energy bands close to the Γ symmetry line.

A change of the degree of short-range order in solid solutions (for example in Cu–Al alloys) is more difficult to detect. Specifically, upon increasing short-range order only small variations in the previously mentioned physical properties occur. These variations take place concomitantly with a slight increase of the diffuse maxima in the continuous background of the diffuse X-ray scattering, typical for solid solutions.[41,42] Figure 16 shows a differential reflectogram obtained on two copper–17 atom % aluminum alloys, which have different degrees of short-range order.[43] Two pieces of structure are visible: a pronounced maximum at 2.64 eV and a broad ε_1-type shoulder in the vicinity of 5 eV. These features can be linked to interband transitions caused by short-range ordering since they are accompanied by the previously mentioned change

in the diffuse X-ray background and do not occur in compositional modulation of these alloys.[43] Comparison between Figures 15 and 16 indicates that the effect of short-range ordering on differential reflectograms (and thus upon the band structure) is about six times smaller than for long-range ordering.

Short-range order in Cu–Al alloys can also be characterized by a microdomain model in which the domains are 1–2 nm in size and are separated by diffuse boundaries that are enriched in one of the constituents.[44] The microdomains consist of a two-dimensional antiphase shift structure resembling that for long-range ordered alloys. Further, an increase in short-range order raises the average number of next-nearest-neighbor solute atoms in relation to a matrix atom. In the present case the results can be interpreted as an increase of the next-nearest-neighbor Al concentration by about 0.4 atom %.[42,45]

3.3 Study of Ion Implantation Damage and Epitaxial Regrowth in Semiconductors

Ion-assisted processes are widely used in semiconductor device fabrication, for example for doping. However, when highly energetic ions are implanted into a solid, some of the atoms near the surface are displaced from their regular lattice positions. This, in turn, causes a change in the electronic band structure. A differential reflectogram, obtained by scanning the light beam between an implanted and an unimplanted sample, therefore displays the characteristic interband transition peaks of the host material as we will see below. If no foreign atoms are involved, the peaks in the pertinent differential reflectogram are obviously caused by the lattice damage only.

Figure 17 shows a series of differential reflectograms taken on silicon samples, half of which were bombarded by silicon ions. The implantation energies varied between 60 and 180 keV while the implantation dose was kept constant.[46] Three characteristic peaks in the UV region (which maintain their position on the energy scale) are observed in all diagrams. They are known to be due to interband transitions designated as $L_3' \rightarrow L_1$ (3.4 eV), Σ region (4.2 eV) and $L_3' \rightarrow L_3$ (5.6 eV).[47] This indicates that for all chosen implantation energies, a fair amount of implantation damage was introduced. Some differences in the UV peaks among the five graphs are noted, nevertheless: the heights of the UV peaks for the first three reflectograms are essentially identical,

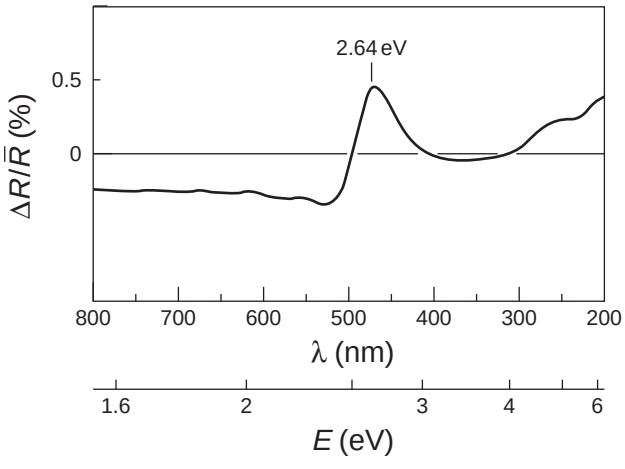

Figure 16 Experimental differential reflectogram of a more short-range ordered versus a less short-range ordered Cu–17 atom % Al alloy. The more short-range ordered state was achieved by quenching the alloy from 600 °C and storing it at room temperature for 42 h; the less short-range ordered state was accomplished by annealing at 290 °C after the 600 °C quench. (Reproduced from J.B. Andrews, R.J. Nastasi-Andrews, R.E. Hummel, *Phys. Rev. B*, **22**(4), 1837–1842 (1980). Copyright 1980 by the American Physical Society.)

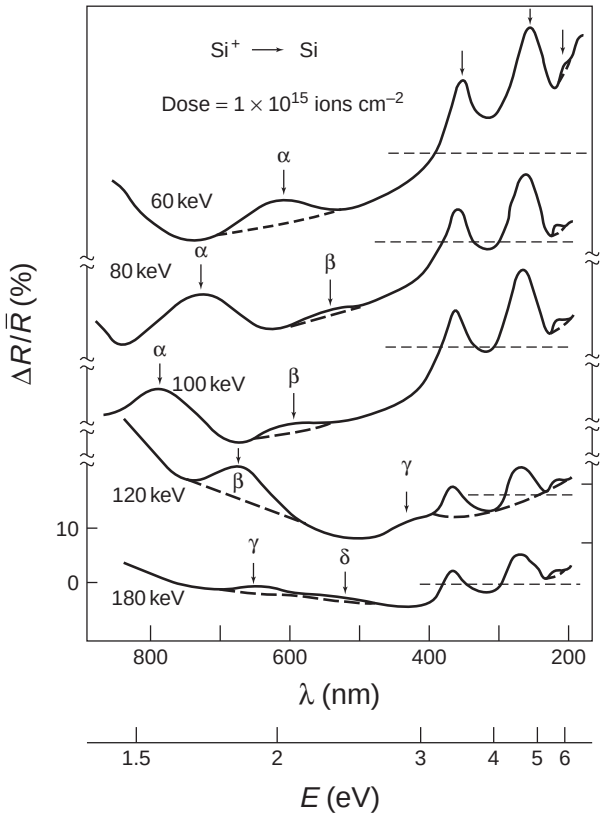

Figure 17 Experimental differential reflectograms of Si^+ ion-implanted silicon wafers. The dose was held constant at 10^{15} ions cm^{-2} whereas the implantation energy was varied as indicated. The individual curves have been shifted for clarity. The substrate consisted of a p-type Si single crystalline wafer whose resistivity was 5–10 Ω cm. (Reproduced from Hummel et al.[46] by permission of the American Institute of Physics.)

whereas the peak heights for larger implantation energies are observed to be smaller. This can be understood by knowing that with increasing implantation energies the radiation-damaged areas become amorphous, grow wider, and are eventually submerged below an essentially crystalline layer,[48] as shown schematically in Figure 18. In other words, the height of the interband transition peaks indicates whether the information stems from the surface of the material or whether the signal comes from a submerged area. In the latter case the light that reaches the photosensitive device after reflection from the amorphous layer first has to pass twice through an essentially unperturbed surface layer where it is attenuated. In the present case the transition between the occurrence of a surface amorphous layer and a submerged amorphous layer (Figure 18) seems to occur at implantation energies between 100 and 120 keV, as deduced from Figure 17. This is confirmed by cross-sectional transmission electron micrographs.[46]

The differential reflectograms depicted in Figure 17 contain additional vital information pertaining to the thickness of the amorphous layer. The broad peaks at low photon energies (labeled with Greek letters)

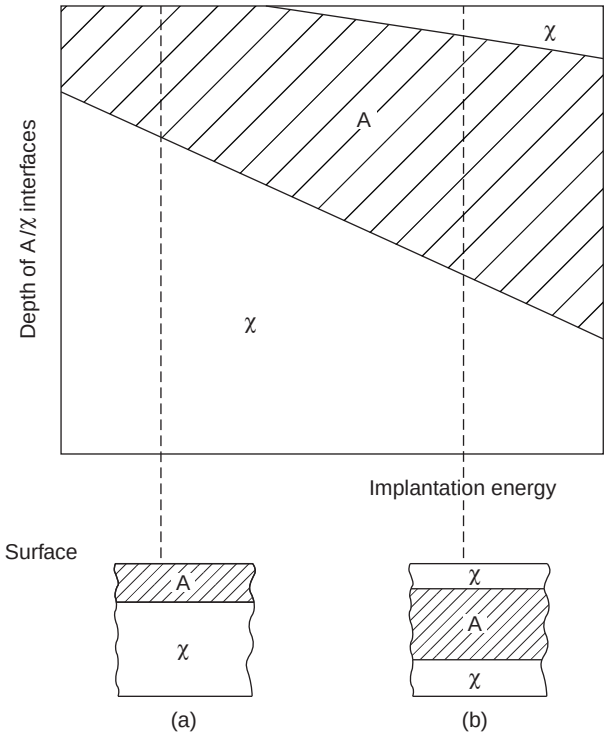

Figure 18 Schematic representation of the amorphous (A)/crystalline (χ) interfaces as a function of ion-implantation energy. Two examples are given in which the amorphous layer is (a) at the surface or (b) submerged below a crystalline layer. (Reproduced from Hummel et al.[46] by permission of the American Institute of Physics.)

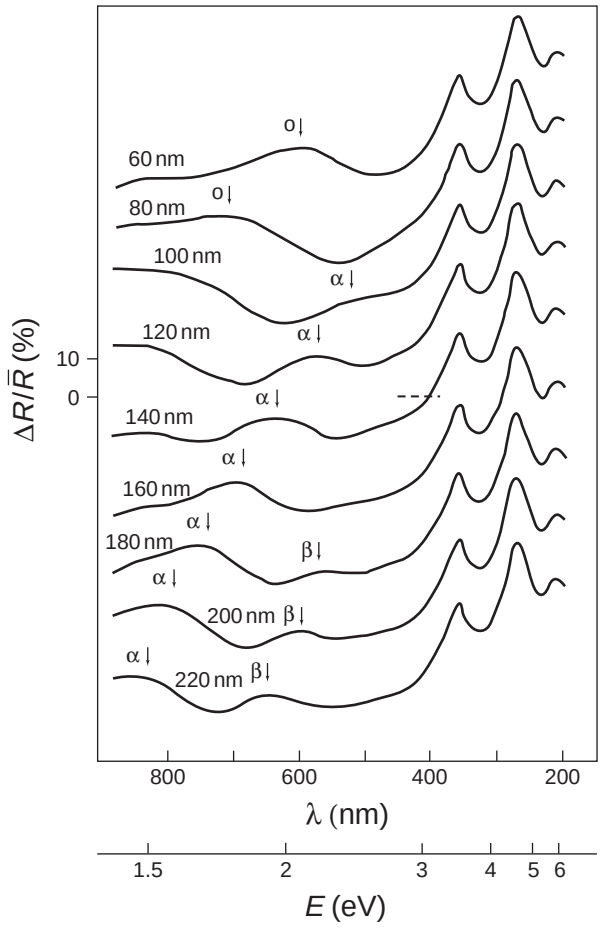

Figure 19 Calculated differential reflectograms obtained by applying the multiple-layer equations from classical optics and using literature values for the optical constants of crystalline and amorphous silicon. (Reproduced from Hummel et al.[51] by permission of The Electrochemical Society, Inc.)

originate from interference effects and are caused by the path difference between the beams which are reflected from the top and from the bottom interfaces of the amorphous layer, respectively. Figure 19 depicts calculated differential reflectograms which have been obtained by applying the multiple-layer formulae from classical optics[49] and by utilizing literature values for the optical constants of crystalline and amorphous silicon.[50] These calculated reflectograms[51] show that with increasing thickness of the amorphous layer, a given interference peak (e.g. α) shifts to lower energies. By comparing Figures 17 and 19, one can finally conclude that the amorphous layer grows in thickness with increasing implantation energy, quite in agreement with the current understanding and as depicted in Figure 18. A closer comparison reveals that the α-peak energy for the 60 keV implantation (Figure 17) and the α-peak energy for a 120 nm layer thickness (Figure 19) are nearly

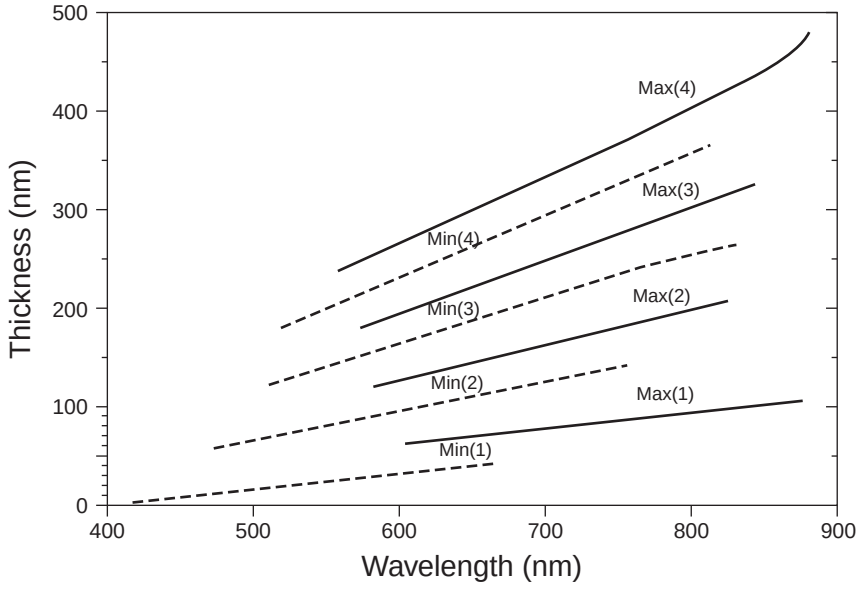

Figure 20 Calculated thickness values of surface amorphous layers for Si^+ implanted silicon, plotted as a function of the wavelength of interference maxima and minima in differential reflectograms (see Figure 19). The curves represent various orders of interference, as indicated by the numbers in parentheses. (Reproduced from Hummel et al.[51] by permission of The Electrochemical Society, Inc.)

identical, which suggests that in the present case a 60 keV bombardment causes a 120 nm thick amorphous layer, quite in agreement with cross-sectional electron micrographs.[46]

In order to obtain the thickness of an amorphous layer from an experimental differential reflectogram, a diagram as shown in Figure 20 is useful. Here interference minima and maxima belonging to different orders as obtained from calculated differential reflectograms are depicted as a function of the wavelengths of their occurrence. As can be seen in Figures 19 and 20, each order is observed in a limited wavelength range only. To obtain the desired thickness from an experimental differential reflectogram, all available maxima and minima (of various orders) need to be matched to provide a unique thickness value from Figure 20.

It should be noted that for amorphous layers less than 5 nm thick, the first-order minimum starts to approach the onset of the interband transition peaks.[52] In this case the thickness can be evaluated only with a lesser accuracy and it may be advantageous to replace the unimplanted reference sample by an amorphous one, thus obtaining an independent check of the thickness. This procedure eliminates the interband transition peaks (which mainly originate from the crystalline reference sample).

In summary, a single differential reflectogram provides several pieces of valuable information pertaining to the implantation-induced changes in a solid. Specifically, it is possible to conclude whether an implanted wafer is still crystalline, or damaged crystalline, or amorphous. Further, one can deduce the thickness of an amorphous layer and whether or not an amorphous layer is submerged. These parameters prove to be quite useful when a nondestructive production-line test needs to be conducted. A direct comparison of data obtained by DRS and by SE shows substantially comparable results. However, the presence of buried layers is not seen in SE[13] (see section 2.5).

Ion implantation is generally used to provide electronic devices with the necessary doping elements to facilitate n- or p-type semiconductors. Detailed studies have shown that DRS is capable of providing quick and nondestructive information about the finer details of these manufacturing processes.[51] As an example, Figure 21 depicts a series of differential reflectograms which were obtained after implanting As^+ ions of different energies into a p-type (100) silicon substrate while holding the dose constant at 10^{15} ions cm^{-2}. It is observed that the interband transition peaks (to the right) remain at constant height, indicating that for all applied ion energies the amorphous region stays close to the wafer surface. This is expected, since heavy ions, such as arsenic, already encounter a relatively high fraction of nuclear stopping at the surface, thus displacing a large number of silicon atoms.[53] We further observe in Figure 21 that the interference peaks, α, β, . . ., move to the left with increasing implantation energy, thus confirming that the thickness of an amorphous layer increases with higher accelerated ion bombardment.

Similar experiments have been conducted by implanting light elements, such as boron (p-type), or by varying

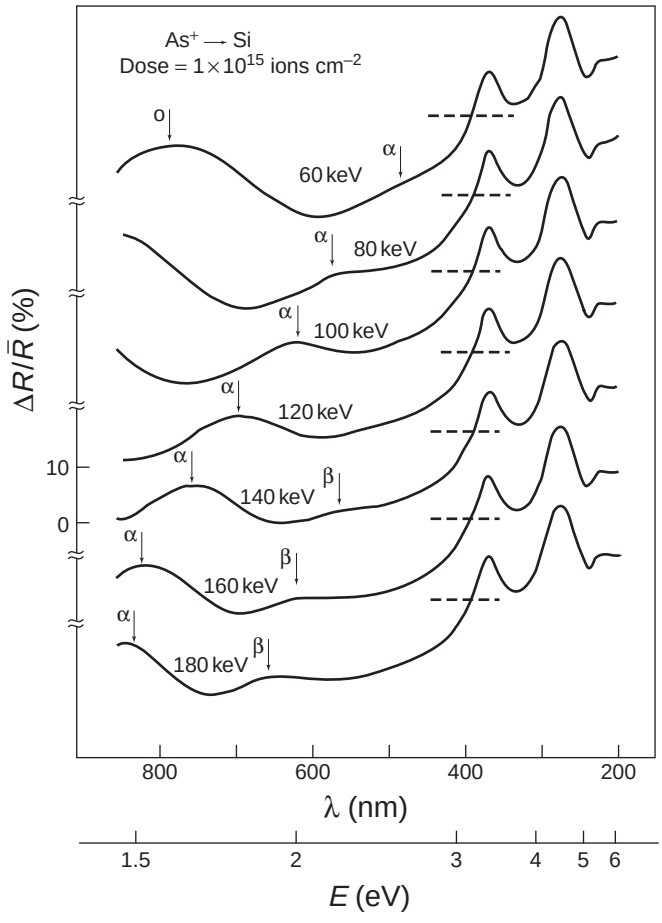

Figure 21 Experimental differential reflectograms of arsenic-implanted (100) silicon wafers. The implantation energies are given on each curve. The positions of the interference maxima are marked with arrows. (Reproduced from Hummel et al.[51] by permission of The Electrochemical Society, Inc.)

the implantation dose while keeping the implantation energy constant.[54] For brevity, these studies are not shown here.

In general, the doping process is not completed by mere ion implantation. Instead, a well-designed annealing procedure has to follow which removes the damage inflicted on the lattice and which also activates the dopants. Figure 22 depicts a series of differential reflectograms which were obtained after several pieces of a silicon wafer (which had been previously implanted with 120 keV Si^+ ions at 5×10^{15} cm^{-2}), were isochronally annealed for 1 h at various temperatures. It is well known that annealing results in recrystallization, i.e. in a solid-phase epitaxial regrowth (SPER). Specifically, heating causes the amorphous/crystalline interface to move progressively through the amorphous region. (This is in contrast to a possible recrystallization process that would occur spontaneously throughout the entire damaged zone.) A recrystallization at the interface is plausible considering that the number of nucleation sites is substantially higher at the crystalline interface. Figure 22 illustrates this in a vivid way: the shift of the interference peaks "α" to the right, with increasing annealing temperatures, indeed shows a steady reduction of the thickness of the amorphous layer. The SPER is essentially completed after annealing for 1 h at 550 °C. This is deduced by inspecting the appropriate differential reflectogram in Figure 22, which is observed to display no interference peaks in the low-energy region.

A supplemental piece of information can be obtained by studying the interband transition peaks in Figure 22. The heights of these peaks are observed to stay fairly constant up to 500 °C, indicating that the surface remains amorphous up to this temperature. For annealings above 500 °C, however, the interband transition peak heights become smaller, suggesting that the surface is now in the crystalline state. Some interband transitions are, however, still visible, implying that some residual defects remain. For temperatures higher than about 600 °C,

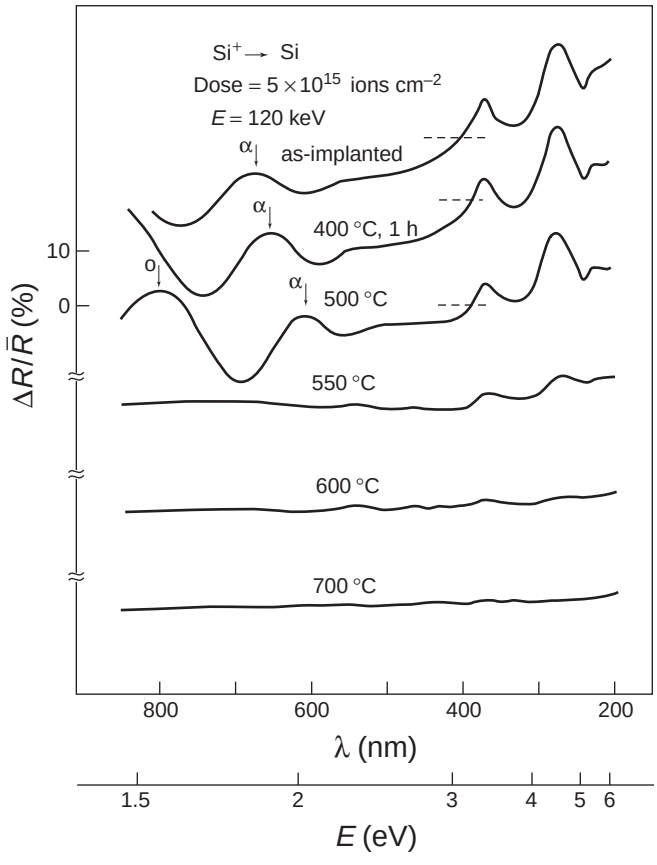

Figure 22 Experimental differential reflectograms for Si^+ ion-implanted (100) silicon wafers after isochronal annealing for 1 h at various temperatures as indicated. (Reproduced from Hummel et al.[51] by permission of The Electrochemical Society, Inc.)

the UV signals eventually disappear, indicating that any difference between the crystalline reference sample and the implanted and annealed specimen has finally ceased to exist. Similar experiments have been carried out for arsenic-implanted silicon in which several distinct annealing steps could be observed.[52]

A comparable study has been conducted by implanting GaAs with He^+ or Ar^+ ions.[55] Here, critical interband transitions at 1.42 eV (E_0), 1.7 eV ($E_0 + \Delta_0$), 2.9 eV (E_1) and at 3.2 eV ($E_1 + \Delta_1$) have been distinctly observed. To quantify the ion-implantation damage the area under the pertinent differential reflectograms has been evaluated around the ($E_1 + \Delta_1$) critical point. This procedure was repeated several times after the top 100 Å were chemically etched off in successive steps. (The interaction depth for light for the impinging plus the reflected beam is estimated at this wavelength to be about 60 Å.) The resulting damage-profile thus obtained compares favorably with a Monte-Carlo simulation of displaced atoms (trajectories of ions in materials, TRIM) as shown in Figure 23. In other words, DRS can be used to obtain implantation profiles in certain materials.

Morris et al.[56] systematically extended ion-implantation studies of As, B, and BF_2 into Si and compared their theoretical predictions of amorphous layer thicknesses (using the UT-MARLOWE software platform for Monte-Carlo simulations) with a number of techniques including DRS, transmission electron microscopy (TEM), scanning electron microscopy (SEM) and Rutherford backscattering spectroscopy (RBS). Good agreement has generally been found between the various techniques except for the B implants. It was assumed that the apparent measured thickness in the latter case may merely indicate a ''depth of damage'' rather than an amorphous layer.

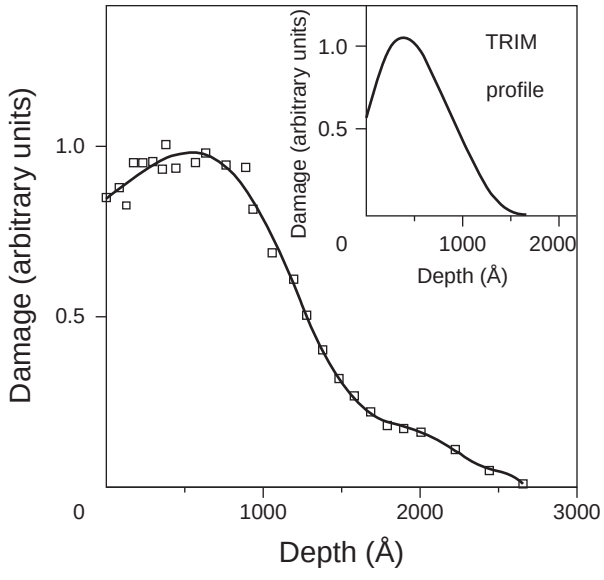

Figure 23 Damage profile in GaAs induced by implantation of 10^{13} ions cm^{-2} Ar^+ ions at 100 keV. The insert shows the corresponding theoretical damage profile calculated with TRIM. (Reproduced from Shwe and Gal[55] by permission of the American Institute of Physics.)

3.4 Corrosion Optics

For the study of corrosion by DRS, one sample is usually a clean (e.g. electropolished) metal (reference) while the other is the same material which is additionally covered by a corrosion product. Corrosion products, particularly in their early stages, are generally thin and relatively transparent. Thus, the light from the differential reflectometer interacts entirely with this transparent or semitransparent film. Owing to the difference-forming nature of the technique, the effect of the underlying metal substrate is essentially subtracted out if the corrosion layer is thin enough. Thus, the resulting optical spectrum represents mainly the properties of the film alone with only minor contributions from the substrate. Therefore, DRS particularly identifies and characterizes the early stages of corrosion phenomena, which are less accessible by other methods such as X-ray techniques. Moreover, corrosion studies can be performed in situ, that is, in an electrolyte by utilizing a corrosion cell into which the light enters and from which it exits through a quartz window, as shown in Figure 24.[57] For the experiments, a metal disk is divided into two halves which are electrically insulated from each other by a thin Teflon® foil, thus permitting the application of a different potential to each subspecimen. This allows to produce the required two samples, one in the corroded state and the other electropolished.

As an example, Figure 25(a) depicts the differential reflectogram for Cu_2O. It is characterized by a pronounced maximum near 380 nm (3.25 eV) and several other peaks and shoulders as indicated in the figure. In contrast, the differential reflectogram of CuO is relatively featureless, displaying only weak shoulders in the vicinity of 560 nm (2.2 eV) and 360 nm (3.4 eV) which are superimposed on a steady rise in $\Delta R/\overline{R}$ towards higher photon energies (Figure 25b). In short, the differential reflectograms of these two oxide species of copper are clearly different from each other. We shall make use of this information in a moment.

Figure 26 depicts a series of differential reflectograms demonstrating the evolution of Cu_2O on a copper substrate in a buffered electrolyte of pH 9.[60] After 10 min (0.17 h) the characteristic features of Cu_2O are already noticeable. The peak height near 380 nm, and thus the corrosion film thickness, initially grow rapidly. The growth rate, however, slows down as the film becomes thicker. The growth kinetics obey a logarithmic

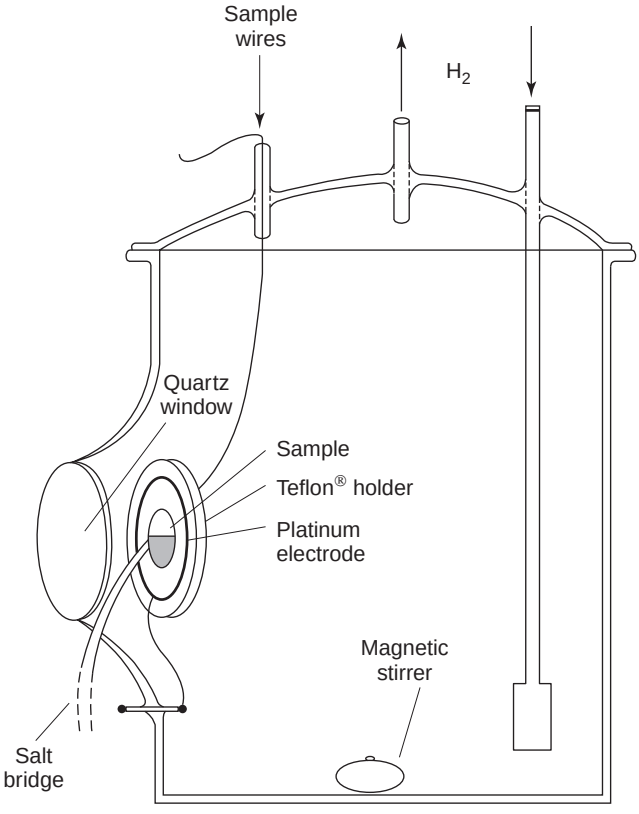

Figure 24 Schematic diagram of a cell used for in situ electrochemical corrosion experiments. (Reprinted from C.W. Shanley, R.E. Hummel, E.D. Verink, Jr, 'Differential Reflectometry – A New Optical Technique to Study Corrosion Phenomena', *Corros. Sci.*, **20**, 467–480 (1980). Copyright 1980, with permission from Elsevier Science.)

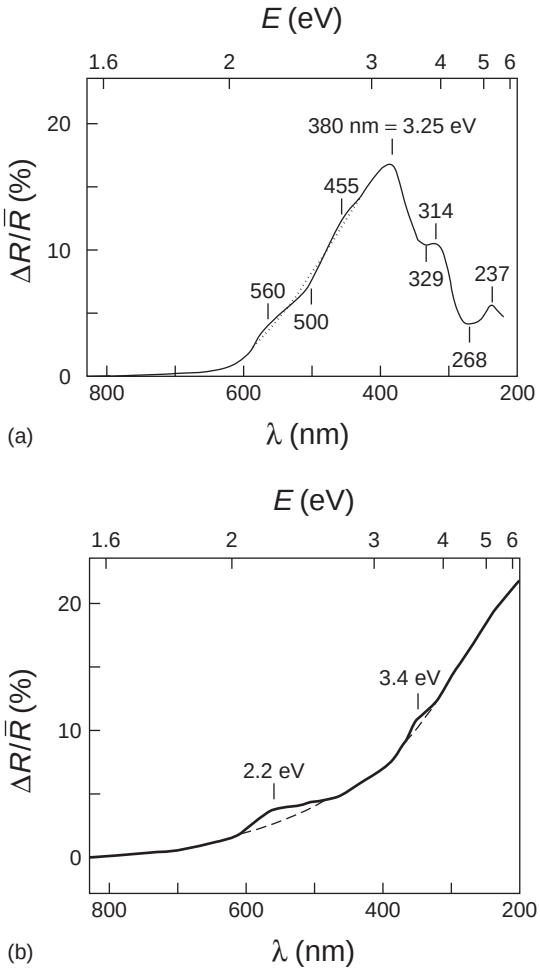

Figure 25 Differential reflectogram of (a) Cu_2O: a part of the copper was potentiostatically held for 53 h in the Cu_2O region of the Pourbaix diagram, i.e. at −200 mV (saturated calomel electrode) and pH 9.2. (b) CuO: a part of the copper was held potentiostatically in the CuO region of the Pourbaix diagram (+200 mV (saturated calomel electrode) and pH 9.2, for 18 h). (Reprinted from C.W. Shanley, R.E. Hummel, E.D. Verink, Jr, 'Differential Reflectometry – A New Optical Technique to Study Corrosion Phenomena', *Corros. Sci.*, **20**, 467–480 (1980). Copyright 1980, with permission from Elsevier Science.) Note that in both cases R of a copper oxide on a Cu substrate was measured versus R of unoxidized Cu.

relationship, as can be seen in Figure 27. This technique can be used to measure the activation energy for oxide formation under actual conditions.

The most common type of corrosion occurs when materials are exposed to the atmosphere or to water. For the investigation of this ''free corrosion'', half of a polished metal disk may be covered with a protective lacquer, while the other half is permitted to corrode. After a specific time, the lacquer is stripped off and the sample is transferred immediately into the differential reflectometer. For in situ studies, a protective coating which is transparent in the entire spectral range (such as collodion) is applied to the reference sample and remains there during the entire measurement.

Copper exposed to the laboratory atmosphere for one day yields a differential reflectogram which contains the multiple peak structure characteristic of Cu_2O (Figure 28).[58] The upward bend indicates, however, that some CuO is also present. With increasing time of exposure to air the reflectograms successively lose the Cu_2O peaks and tend to produce CuO patterns. This result is interesting, because the copper–oxygen phase diagram[59] indicates that CuO is the stable species if copper is exposed to dry air below 375 °C. On the other hand, it has been suggested[59] that CuO tends to be reduced to Cu_2O in the presence of metallic copper. The results obtained by differential reflectometry indicate the coexistence of different corrosion products, particularly in the early stages of corrosion, and the eventual formation of the equilibrium oxide. This is concluded from the superposition of the spectrum of one compound on the spectrum of the other.[57]

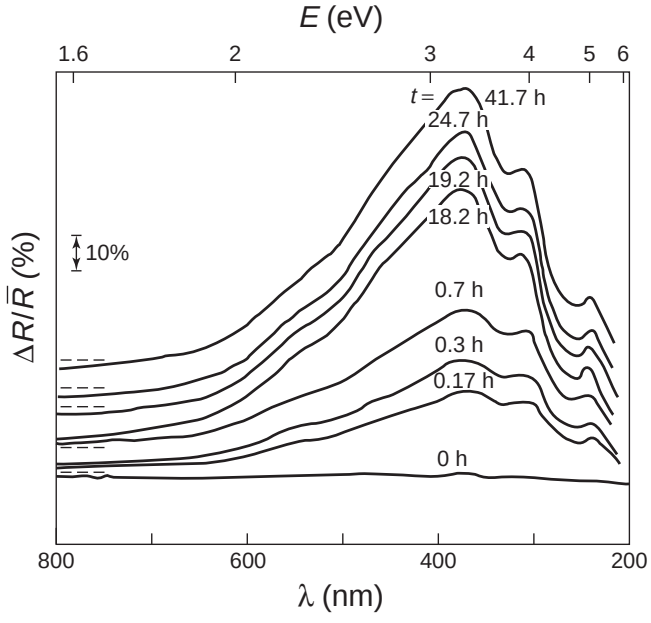

Figure 26 Differential reflectograms depicting the in situ evolution of Cu_2O on a copper substrate in a buffered electrolyte of pH 9. (Reproduced from Hummel[60] with kind permission from Kluwer Academic Publishers.) Both sample halves were initially held at the protective potential (−500 mV (saturated calomel electrode)) for 30 min (reflectogram marked 0 h). Subsequently, one sample half was held potentiostatically at −200 mV (saturated calomel electrode) and reflectograms were taken at the time intervals marked on the curves.

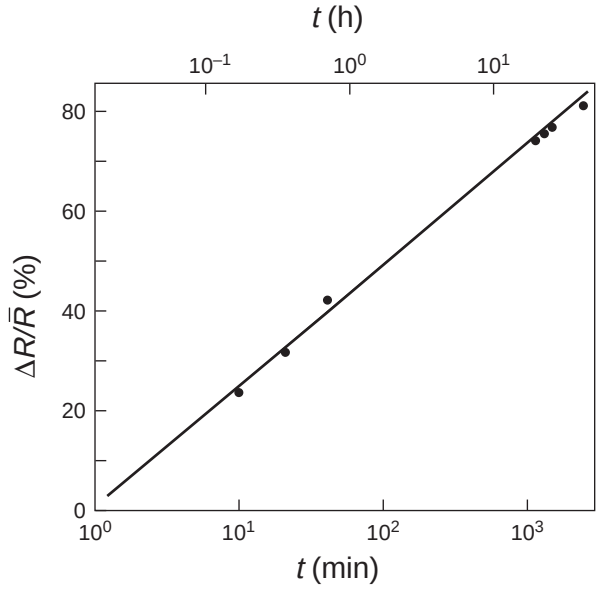

Figure 27 Growth kinetics of Cu_2O on copper substrate: evolution of the 380 nm peak of Figure 26 as a function of potentiostating time, *t*. (Reproduced from Hummel[60] with kind permission from Kluwer Academic Publishers.)

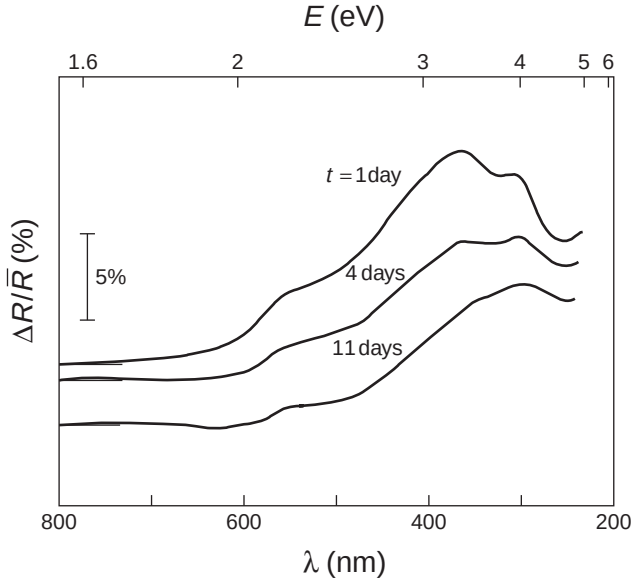

Figure 28 Differential reflectograms of corrosion products of copper. Copper was allowed to corrode freely for various times as indicated. The zero lines have been shifted for clarity. (Reprinted from C.W. Shanley, R.E. Hummel, E.D. Verink, Jr, 'Differential Reflectometry of Corrosion Products of Copper', *Corros. Sci.*, **20**, 481–487 (1980). Copyright 1980, with permission from Elsevier Science.)

Free corrosion of copper in distilled water again produces differential reflectograms which are indicative of mixed oxides of CuO and Cu_2O. The same observation is made when oxygen is continuously bubbled through water in which copper is immersed.[58]

The question arises as to whether or not the thickness of a corrosion product can be estimated from its differential reflectogram. This is indeed possible when the optical constants (index of refraction, *n*, and damping constant, *k*) of substrate and corrosion product are known. Then, by applying the (rigorous) two-layer formulae[49] the film-thickness-dependence of $\Delta R/\bar{R}$ can be calculated. This has been done for CuO on a copper substrate and for NiO on a nickel substrate for wavelengths of 600 nm and 550 nm respectively, and for a film thickness range of up to 200 Å (see Figure 29).[60] Utilizing this graph for the CuO film shown in Figure 25(b), one arrives at a film thickness of about 5 Å which is well within the accepted range. Some caution should be exercised, however, in applying Figure 29 too rigidly. The optical constants of thin films may not be exactly the same as those of the bulk. Nevertheless, an assumed error of 10% in the index of refraction also changes the film thickness by about 10%, which in the case described here is less than a monolayer and therefore negligible.

As a last example for the application of DRS in corrosion research, the phenomenon of dealloying is chosen.

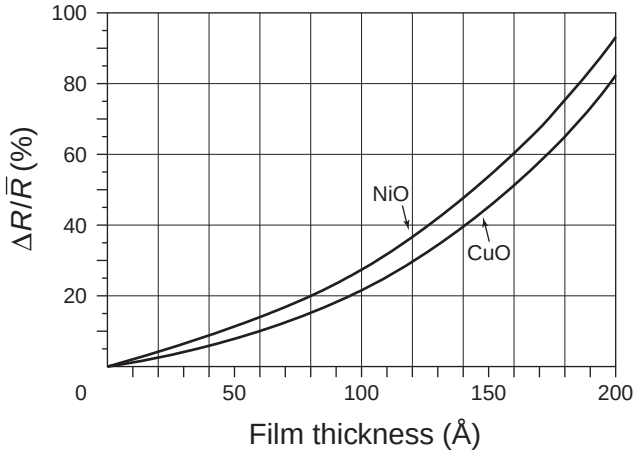

Figure 29 Calculated differential reflectivity as a function of film thickness for CuO on a copper substrate in aqueous solution ($\lambda = 600$ nm; $n_{Cu} = 0.17$; $k_{Cu} = 3.07$; $n_{CuO} = 2.6$; $k_{CuO} = 0.445$; $n_{H_2O} = 1.33$). (Reproduced from Hummel[60] with kind permission from Kluwer Academic Publishers.) For comparison $\Delta R/\bar{R}$ for NiO on Ni is also depicted ($\lambda = 550$ nm).

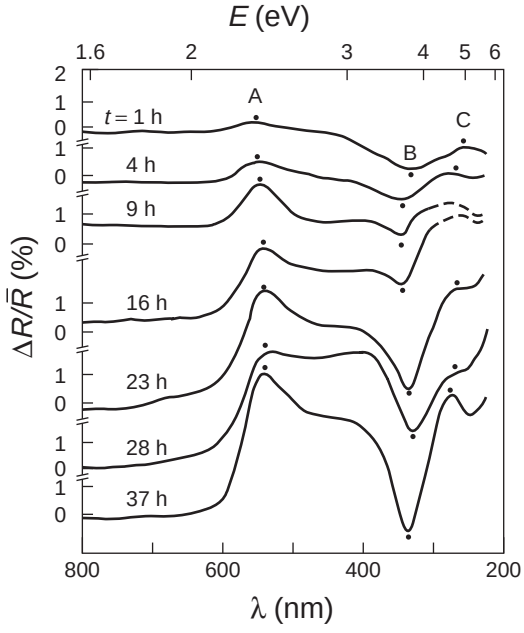

Figure 30 Differential reflectograms of a Cu–9.2 atom % Zn alloy (red brass), half of which was electrochemically treated for various times in a pH 4 solution at $E = -350$ mV (saturated calomel electrode) that is, at -105 mV (standard hydrogen electrode) (dezincification). The other half remained in the unexposed state. (Reproduced from Finnegan et al.[63] by permission of Nace International.)

Specifically, under certain conditions, such as exposure to aqueous solutions (e.g. a water faucet), copper–zinc alloys suffer a loss of zinc (by a mechanism called ''selective leaching'') which reduces their hardness. Based on visual inspection it has been generally assumed that copper alloys containing less than 15 atom % Zn are immune to dezincification.[61,62] Differential reflectometry can effortlessly aid in the decision as to whether or not this statement is correct. To investigate dealloying, a surface technique is superior to chemical analysis, which averages bulk and surface values.

For the experiments described here, the reflectivity of a freshly polished specimen of the original composition is compared with that of a dealloyed specimen. Any difference in composition between the two sample halves should provide differential reflectograms similar to those in Figure 9.

Figure 30 depicts the spectral dependence of the difference in reflectivity between two copper–zinc alloys (one dezincified as stated in the caption of Figure 30, and the other in the virgin state) whose zinc content initially was 9.2 atom % (red brass). The reflectograms possess all the major features of compositional modulation, which suggests that red brass indeed undergoes dezincification. From the shift of peak A, a zinc loss of 3.8 atom % is found. (On the other hand, a Cu–21.4 atom % Zn alloy dezincifies under comparable conditions by losing 4.4 atom % Zn.) This smaller loss of zinc in red brass is, however, not the only difference compared with yellow brass (that is, Cu–Zn having a zinc concentration in excess of 15 atom %). The kinetics of formation of a specific peak height, and thus the corrosion rate, are about 16 times larger in yellow brass.[63] Even a Cu–5.3 atom % Zn alloy dezincifies if potentiostated for long enough. The loss of zinc is in this case about 2.6 atom %.[63] These experimental findings demonstrate once more the usefulness and sensitivity of DRS.

4 CONCLUSION

DRS has been shown over the past 30 years to be a highly sensitive, nondestructive, fast surface analytical technique which probes into solids between some monolayers up to 1 μm (depending on the material under investigation). Moreover, DRS is capable of characterizing and measuring the thickness of thin layers of, for example, corrosion films on bulk substrates or ion-implanted layers, whether they are on the surface or submerged below the surface. Moreover, DRS is distinguished by a low background noise and by an immunity towards surface disturbances (e.g. mechanical polishing, interactions with the environment, etc.) due to its difference-forming features. No vacuum is needed unless measurements in the vacuum UV are desired. The probing light does not to be polarized. DRS provides a continuous absorption-type spectrum from the IR through the visible into the UV spectral region. Its peak structure is caused by specific

interband transitions of the valence electrons, involving the electron band structure of the material under consideration.

The examples for applications of DRS given in this review emphasize only some major broad areas. Needless to say, DRS has also been used in a number of other investigations which cannot be easily summarized under a given heading. For the benefit of the interested readers, some more papers which utilize DRS are cited below. Anodic oxide-film growth on GaAs has been studied in situ as a function of time utilizing only one wavelength (He–Ne laser).[64] Structure changes in GaAs caused by Ar^+ ion bombardment are reported by Ludwig and Hummel.[65] The DRS method has also been used to study InGaAs/GaAs and AlGaAs/GaAs quantum wells.[66] The surface-state band gap of InP (110) has been investigated by surface differential reflectometry[67–69] (see also section 2.5). Overall, it can be seen that in recent years DRS has been applied mainly to the study of semiconductors, particularly compound semiconductors.

ACKNOWLEDGMENTS

Dr V. Gerold (MPI Stuttgart) is thanked for his critical review of the manuscript. Likewise, Dr E.A. Irene (University of North Carolina at Chapel Hill) made several helpful suggestions. The development of DRS and its application in materials research would not have been possible without a steady stream of dedicated graduate students, such as J.A. Holbrook, J.B. Andrews, R.J. Nastasi-Andrews, W.M. Goho, W. Xi, C.W. Shanley, J.E. Finnegan, R.J. Smith, and F.K. Urban. The National Science Foundation has supported these investigations for many years.

ABBREVIATIONS AND ACRONYMS

DC	Direct Current
DRS	Differential Reflectance Spectroscopy
ESCA	Electron Spectroscopy for Chemical Analysis
IR	Infrared
PC	Personal Computer
PMT	Photomultiplier Tube
RBS	Rutherford Backscattering Spectroscopy
RDS	Reflectance Difference Spectroscopy
SDRS	Surface Differential Reflectivity Spectroscopy
SE	Spectroscopic Ellipsometry
SEM	Scanning Electron Microscopy
SPER	Solid-phase Epitaxial Regrowth
TEM	Transmission Electron Microscopy
TRIM	Trajectories of Ions in Materials
UV	Ultraviolet
XRD	X-ray Diffraction

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Surfaces: Introduction • Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Ellipsometry in Analysis of Surfaces and Thin Films • Infrared and Raman Spectroscopy in Analysis of Surfaces • Ion Scattering Spectroscopy in Analysis of Surfaces • Photoluminescence in Analysis of Surfaces and Interfaces

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Electroanalytical Methods **(Volume 11)**
Infrared Spectroelectrochemistry

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview

REFERENCES

1. R.E. Hummel, D.B. Dove, J. Alfaro Holbrook, 'Optical Reflectivity Measurements on Alloys by Compositional Modulation', *Phys. Rev. Lett.*, **25**(5), 290–292 (1970).
2. J. Alfaro Holbrook, R.E. Hummel, 'A Differential Reflectometer for Measurements of Small Differences in Reflectivity', *Rev. Sci. Instrum.*, **44**(4), 463–466 (1973).
3. D. Beaglehole, 'A Sensitive Single Beam Device for Continuous Reflectance or Transmittance Measurements', *Appl. Opt.*, **7**(11), 2218–2220 (1968).
4. A.A. Volfson, V.K. Subashiev, 'Differential Reflection Spectra of Heavily Doped Silicon and Germanium in the Ultraviolet Spectral Region', *Phys. Status Solidi B*, **33**(1), 149–159 (1969).
5. W.R. Scott, L. Muldawer, M.A. Graber, 'Sensitive Technique for Measuring Differences in Reflectivity', *Appl. Opt.*, **33**(8), 1956–1958 (1974).

6. R.E. Hummel, J.B. Andrews, J. Alfaro Holbrook, 'Application of the Differential Reflectometer for Materials Research such as Corrosion, Ordering, Alloying, etc.', *Z. Metallkd.*, **64**, 573–577 (1973).
7. M. Cardona, 'Modulation Spectroscopy', in *Solid State Physics*, Supplement 11, Academic Press, New York, 1969.
8. B.O. Seraphin, N. Bottka, 'Band-structure Analysis from Electro-reflectance Studies', *Phys. Rev.*, **145**(2), 628–636 (1966).
9. D.E. Aspnes, 'Third-derivative Modulation Spectroscopy with Low-field Electro-reflectance', *Surf. Sci.*, **37**, 418–442 (1973).
10. D.E. Aspnes, 'Above-bandgap Optical Anisotropies in Cubic Semiconductors: A Visible–Near Ultraviolet Probe of Surfaces', *J. Vac. Sci. Technol., B*, **3**(5), 1498–1506 (1985).
11. S. Selci, F. Ciccacci, G. Chiarotti, P. Chiaradia, A. Cricenti, 'Surface Differential Reflectivity Spectroscopy of Semiconductor Surfaces', *J. Vac. Sci. Technol., A*, **5**(3), 327–332 (1987).
12. D.E. Aspnes, A.A. Studna, 'High Precision Scanning Ellipsometer', *Appl. Opt.*, **14**(1), 220–228 (1975).
13. T.M. Burns, S. Chongsawangvirod, J.W. Andrews, E.A. Irene, G. McGuire, S. Chevacharoeukul, 'A Comparison of the Measurement of Ion Damage in Silicon Surfaces Using Differential Reflectance and Spectroscopic Ellipsometry', *J. Vac. Sci. Technol., B*, **9**(1), 41–49 (1991).
14. R. Enderlein, R.E. Hummel, J.B. Andrews, R.J. Nastasi-Andrews, C.W. Shanley, 'Interpretation of Compositional Modulation Spectra to Determine Some Optical Properties of Alloys', *Phys. Status Solidi B*, **88**, 173–178 (1978).
15. R.E. Hummel, R. Enderlein, 'New Look at the Lineshape of Differential Reflectograms for Dilute Alloys', *Phys. Rev. B*, **29**(4), 1529–1533 (1984).
16. R.E. Hummel, *Electronic Properties of Materials*, 2nd edition, Springer-Verlag, New York, 1993.
17. R.E. Hummel, J. Alfaro Holbrook, J.B. Andrews, 'Compositional Modulation of Cu–Zn, Cu–Al, and Cu–Ni Alloys', *Surf. Sci.*, **37**, 717–729 (1973).
18. B. Segall, 'Fermi Surface and Energy Bands of Copper', *Phys. Rev.*, **125**(1), 109–122 (1962).
19. F.M. Mueller, J.C. Phillips, 'Electronic Spectrum of Crystalline Copper', *Phys. Rev.*, **157**(3), 600–607 (1967).
20. R.J. Nastasi-Andrews, R.E. Hummel, 'Optical Properties and Electronic Structure of Dilute Cu–Au, Cu–Zn, Cu–Al, Cu–Ga and Cu–Si, Cu–Ge, Cu–Sn, and Cu–As Alloys', *Phys. Rev. B*, **16**(10), 4314–4323 (1977).
21. M. Guerrisi, R. Rosei, P. Winsemius, 'Splitting of the Interband Absorption Edge in Au', *Phys. Rev. B*, **12**(2), 557–563 (1975).
22. N.F. Mott, H. Jones, *The Theory of the Properties of Metals and Alloys*, Clarendon Press, Oxford, 1936.
23. A. Bansil, H. Ehrenreich, L. Schwartz, R.E. Watson, 'Complex Energy Bands in Alpha-brass', *Phys. Rev. B*, **9**(2), 445–465 (1974).
24. J. Friedel, 'The Absorption of Light by Noble Metals and Its Relation to the Van der Waals Contribution to the Cohesive Energy', *Proc. Phys. Soc. B*, **65**, 769–774 (1952).
25. A.B. Chen, B. Segall, 'Piezooptical Response Associated with the Interconnection Band Transitions in the Nobel Metals', *Solid State Commun.*, **18**(1), 149–152 (1976).
26. P.G. Pells, H. Montgomery, 'The Optical Properties of Alpha-phase Cu–Zn, Cu–Ga, Cu–Fe and Cu–As Alloys', *Metal Physics C Suppl.*, **3**, 330–340 (1970).
27. R. Rosei, S. Modesti, V. Prantera, I. Davoli, 'Shift of Electronic States at L in Alpha-phase Cu–Zn and Cu–Al Alloys', *J. Phys. F: Metal Physics*, **9**(11), 2275–2285 (1979).
28. H. Amar, K.H. Johnson, C.B. Sommers, 'Electronic Structure of Alpha Brass', *Phys. Rev.*, **153**(3), 655–658 (1967).
29. D. Beaglehole, 'Optical Absorption and s and p Phase Shifts of Virtual-bond-state Alloys', *Phys. Rev. B*, **14**(2), 341–346 (1976).
30. D. Beaglehole, T.J. Hendrickson, 'Reflectivity Studies of Dilute Gold–Iron Alloys', *Phys. Rev. Lett.*, **22**(4), 133–136 (1969).
31. M. Tokumoto, 'Optical Absorption of Ni-based Alloys', *Phys. Rev. B*, **22**(2), 638–644 (1980).
32. M. Tokumoto, H.D. Drew, A. Bagchi, 'Optical Absorption of Ni–Cu Alloys', *Phys. Rev. B*, **16**(8), 3497–3503 (1977).
33. R.E. Hummel, 'Reflectivity of Silver, and Aluminum-based Alloys for Solar Reflectors', *Solar Energy*, **27**(6), 449–455 (1981).
34. R.E. Hummel, 'Differential Reflectometry of Sputtered Silver–Aluminum Films Between 1.6 and 6.2 eV', *Opt. Commun.*, **39**(1,2), 55–58 (1981).
35. R.E. Hummel, W.M. Goho, 'Optical Properties and Electronic Structure of Copper–Cobalt Alloys', *Phys. Status Solidi B*, **147**(1), 417–420 (1988).
36. H. Warlimont (ed.), *Order–Disorder Transformations in Alloys*, Springer-Verlag, Berlin, 1974.
37. H.L. Skriver, H.P. Lengkeek, 'Band Structure and Optical Properties of Ordered $AuCu_3$', *Phys. Rev. B*, **19**(2), 900–910 (1979).
38. P.O. Nilsson, C. Norris, 'Order–Disorder Effects in $AuCu_3$ Observed in Reflectivity and Photoemission', *Phys. Lett. A*, **29**(1), 22–23 (1969).
39. M.M. Pant, S.K. Joshi, 'Green's-function Method for Energy Bands in Disordered Alloys. II. Band Structure of Disordered Cu_3Au', *Phys. Rev. B*, **1**(2), 506–508 (1970).
40. W. Scott, L. Muldawer, 'Optical Properties and Electronic Structure of Ordered and Disordered Cu_3Au', *Phys. Rev. B*, **9**(4), 1115–1125 (1974).
41. J.E. Epperson, P. Fürnrohr, C. Oritz, 'The Short-range-order Structure of Alpha-phase Cu–Al Alloys', *Acta. Crystallogr. A*, **34**(5), 667–681 (1978).

42. V. Gerold, R.E. Hummel, 'Der Einfluß der Nahordnung in Kupfer–Aluminum–Legierungen auf Elektrische und Optische Eigenschaften', *Z. Metallkd.*, **72**(11), 761–764 (1981).
43. J.B. Andrews, R.J. Nastasi-Andrews, R.E. Hummel, 'Optical Investigations of Short-range Ordering of α-Copper–Aluminum Alloys', *Phys. Rev. B*, **22**(4), 1837–1842 (1980).
44. W. Gaudig, H. Warlimont, 'The Structure of Short-range Ordered Alpha-Cu–Al Alloys and a New Superlattice Phase', *Acta Metall.*, **26**(5), 709–724 (1978).
45. W. Köster, H.P. Rave, 'Conductivity and Hall Constant XIII, Alpha-Copper–Aluminum Alloys', *Z. Metallkd.*, **52**, 158–161 (1961).
46. R.E. Hummel, W. Xi, P.H. Holloway, K.A. Jones, 'Optical Investigations of Ion Implant Damage in Silicon', *J. Appl. Phys.*, **63**(8), 2591–2594 (1988).
47. M.L. Cohen, T.K. Bergstresser, 'Band Structures and Pseudopotential Form Factors for Fourteen Semiconductors of the Zinc-blende Structures', *Phys. Rev.*, **141**(2), 789–796 (1966).
48. S. Prussin, D.E. Margolese, R.N. Tauber, 'Formation of Amorphous Layers by Ion Implantation', *J. Appl. Phys.*, **57**(2), 180–185 (1985).
49. O.S. Heavens, *Optical Properties of Thin Solid Films*, Dover, New York, 1965.
50. E.D. Palik (ed.), *Handbook of Optical Constants of Solids*, Academic Press, Orlando, 1985.
51. R.E. Hummel, W. Xi, D.R. Hagmann, 'Ion Implantation Damage and Annealing of Silicon as Characterized by Differential Reflectometry', *J. Electrochem. Soc.*, **137**(11), 3583–3588 (1990).
52. D.R. Hagmann, W. Xi, R.E. Hummel, 'A Study of the Solid-phase Epitaxial Regrowth of Amorphous Silicon Utilizing Differential Reflectometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **59/60**, 110–115 (1991).
53. T.E. Seidel, in *VLSI Technology*, ed. S.M. Sze, McGraw-Hill, New York, 219, 1983.
54. R.E. Hummel, W. Xi, 'Differential Reflectometry Studies of Implantation Damage and Annealing Behavior of Silicon Wafers Involving Light and Heavy Ions', in *Proc. Symp. Ion Implantation and Dielectrics for Elemental and Compound Semiconductors*, eds. S.J. Pearton, K.S. Jones, V.J. Kapoor, 176th Meeting of the Electrochemical Society, Hollywood, October 15–20, 1989, The Electrochem. Soc., Inc., Pennington, 196–206, Proc. Vol. 90–13, 1990.
55. C. Shwe, M. Gal, 'Measurement of Damage Profile in Semiconductors: A Sensitive Optical Technique', *Appl. Phys. Lett.*, **62**(5), 516–518 (1993).
56. M.F. Morris, S. Tian, S.-H. Yang, S.J. Morris, B. Obradovic, A.F. Tasch, R.E. Hummel, 'Comparison of UT-MARLOWE Predictions of Implant-induced Damage with Experimentally Measured Amorphous Layer Thicknesses', in *Proc. 11th International Conference on Ion Implantation Technology, Austin, June 16–21, 1996*, IEEE, New York, 551–558, 1997.
57. C.W. Shanley, R.E. Hummel, E.D. Verink, Jr, 'Differential Reflectometry – A New Optical Technique to Study Corrosion Phenomena', *Corros. Sci.*, **20**, 467–480 (1980).
58. C.W. Shanley, R.E. Hummel, E.D. Verink, Jr, 'Differential Reflectometry of Corrosion Products of Copper', *Corros. Sci.*, **20**, 481–487 (1980).
59. M. Hanson, K. Anderko, 'Constitution of Binary Alloys', McGraw-Hill, New York, 1958.
60. R.E. Hummel, 'Differential Reflectometry and Other Optical Techniques in Corrosion Research', in *Electrochemical and Other Techniques for the Study and Monitoring of Metallic Corrosion*, eds. M.G.S. Ferreira, C.A. Melendres, Kluwer Academic Publishers, Boston, 1991.
61. M. Fontana, N. Greene, *Corrosion Engineering*, McGraw-Hill, New York, 1978.
62. H.H. Uhlig, *Corrosion and Corrosion Control*, John Wiley & Sons, New York, 1971.
63. J.E. Finnegan, R.E. Hummel, E.D. Verink, Jr, 'Optical Studies of Dezincification in Alpha-brass', *Corrosion*, **37**(5), 256–261 (1981).
64. A. Moritani, H. Kubo, J. Nakai, 'Observation of Anodic Oxide-film Growth on GaAs by in situ Differential Reflectance', *J. Appl. Phys.*, **48**(6), 2638–2640 (1977).
65. M.H. Ludwig, R.E. Hummel, 'Differential Reflectance Studies of Structural Changes in GaAs Caused by Ar^+ Ion Bombardment', *Mater. Res. Soc. Symp. Proc.*, **300**, 345–349 (1993).
66. C. Shwe, M. Gal, 'Differential Reflectance Spectroscopy of InGaAs/GaAs and AlGaAs/GaGaAs Quantum Wells', *Appl. Phys. Lett.*, **57**(18), 1910–1912 (1990).
67. S. Selci, A. Cricenti, A.C. Felici, G. Chiarotti, 'Surface-state Band Gap of InP (110) by Polarized Surface Differential Reflectivity', *Phys. Rev. B*, **44**(15), 8327–8329 (1991).
68. A. Cricenti, S. Selci, A.C. Felici, L. Ferrari, A. Gavrilovich, C. Goletti, G. Chiarotti, 'Clean and Oxygen Covered InP (110) Surface Differential Reflectivity', *J. Vac. Sci. Technol., A*, **9**(3), 1026–1028 (1991).
69. S. Selci, A. Cricenti, A.C. Felici, L. Ferrari, C. Goletti, G. Chiarotti, 'Oxygen Chemisorption on Cleaved InP (110) Surfaces Studied with Surface Differential Reflectivity', *Phys. Rev. B*, **43**(8), 6757–6759 (1991).

Electron Energy Loss Spectroscopy in Analysis of Surfaces

J.L. Erskine
The University of Texas at Austin, Austin, USA

1 Introduction 1
2 History 3
3 Sample Preparation and Requirements 3
4 Electron Scattering Mechanisms and Basis of Technique 3
4.1 Scattering Kinematics 3
4.2 Scattering Mechanisms 5
4.3 Scattering Cross-sections and Selection Rules 6
5 Instrumentation 7
6 Applications to Intrinsic Properties of Surfaces 8
6.1 Fuchs–Kliewer Phonons 8
6.2 Bulk Phonon Resonances at Metal Surfaces 8
6.3 Surface Phonon Dispersion 8
6.4 Semiconductor Surfaces 8
6.5 Ultrathin Metal Films 9
6.6 High Critical Temperature Cuprate Superconductors 9
6.7 Electronic Excitations at Surfaces/ Surface Plasmons 9
7 Applications to Chemisorbed Atoms and Molecules 9
7.1 Atomic Adsorbates 9
7.2 Underlayer Formation at Metal Surfaces 10
7.3 Molecular Adsorbates 11
7.4 Technical Surfaces 11
8 Comparison with Other Spectroscopic Methods 12
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 13
References 13

Electron energy loss spectroscopy (EELS) is a versatile and sensitive probe of electronic and vibrational excitations at surfaces. The technique is extensively used as a research and analytical tool in a broad range of surface science applications including research on intrinsic physical and chemical properties of surfaces, chemisorption phenomena, and chemical processes at surfaces. The spectroscopy is based on measuring the energy loss and momentum transfer of electrons in a highly collimated, monochromatic (narrow energy spread) beam after the electrons scatter from a surface. Vibrational excitations of atoms and molecules adsorbed on surfaces, collective excitations of ordered chemisorbed structures, the intrinsic collective excitations of surface atoms (surface phonons), and collective electronic excitations (surface plasmons) can be studied using the technique. Under favorable conditions, vibrational losses from surface adsorbate concentrations below 10^{-3} monolayer (10^{12} atoms cm^{-2}) can be detected. EELS is a surface science probe, typically requiring an ultrahigh vacuum (UHV) environment, and is generally employed in conjunction with complementary surface analytical probes such as Auger electron spectroscopy (AES), low-energy electron diffraction (LEED) and thermal desorption mass spectroscopy (TDS). EELS is an ideal tool for determining surface chemical composition, and provides vibrational loss spectra similar to infrared reflection absorption spectroscopy (IRAS), Raman spectroscopy, and helium atom scattering spectroscopy (HAS).

1 INTRODUCTION

High-resolution EELS is an important technique for probing physical and chemical properties of surfaces.[1] In an EELS experiment, a monochromatic beam of electrons is scattered from a surface, and the energy and momentum of the scattered electrons are determined. The energy and momentum exchanged between the incident electrons and the surface is related to electronic and vibrational excitations at the surface through various conservation laws, scattering selection rules and scattering mechanisms. EELS therefore provides a means of probing vibrational modes of adsorbed atoms, collective vibrational excitations of crystal surfaces (surface phonons), local electronic effects (negative ion resonances) and collective electronic excitations (surface plasmons). Vibrational modes of atoms and molecules adsorbed on surfaces reflect not only the species present, but also the structural and chemical effects associated with the surface–atom or surface–molecule bonds. EELS is therefore an important technique for gaining insight into atomic level processes that underlie chemisorption and chemical reactions at surfaces.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

EELS experiments are generally carried out using incident electron kinetic energies in the range of a few to several hundred electron volts (eV). This is an important energy range for electron–solid interactions. The range includes energies sufficiently high to: (1) excite interband electronic transitions and collective electron excitations (bulk and surface plasmons), (2) observe and study energy dependencies of inelastic electron scattering mechanisms, and (3) observe effects resulting from electron diffraction; that is accessible energies include the energy range where the de Broglie wavelength of an electron, λ_e, becomes comparable to a typical crystal lattice constant a. When $\lambda_e \sim a$, and long-range surface atomic order exists, such as in a crystal, strong electron diffraction effects occur. Electron diffraction effects can affect EELS vibrational cross-sections, but more importantly, provide a basis for electron-based surface crystallography known as LEED. An important feature of electron–solid interactions is characterized by the inelastic electron mean free path, ξ_e, in the solid. In the energy range covered by elastic and inelastic electron scattering studies of solid surfaces, ξ_e is of the order of 5–10 Å and is characterized by a universal curve having a minimum in the neighborhood of 50 eV (see Figure 1). The relatively short elastic mean free path of low-energy electrons in solids accounts for the fact that vibrational and electronic losses detected by EELS occur very near the surface.

EELS is now a well known and widely used surface analysis technique. The development and extensive applications of EELS are documented in a book,[2] review articles[3–6] and the proceedings[7–14] of a series of international conferences focused on vibrational properties at surfaces. The utility and broad application of the EELS technique in surface science is based on the various levels at which experimental data can be interpreted, and on the important feature that when applied to adsorbed surface species, EELS spectra directly identify molecular or chemical composition. While AES offers a highly sensitive means of detecting surface atomic composition (all elements except hydrogen and helium exhibit characteristic Auger spectra), it is unable to determine the chemical identity of adsorbed species unambiguously. AES, like electron spectroscopy for chemical analysis (ESCA), is an electron spectroscopy of electron energy levels in atoms. EELS is a molecular sensitive spectroscopy because vibrational excitations are functions of atomic composition, masses and chemical bond strengths and bonding geometry.

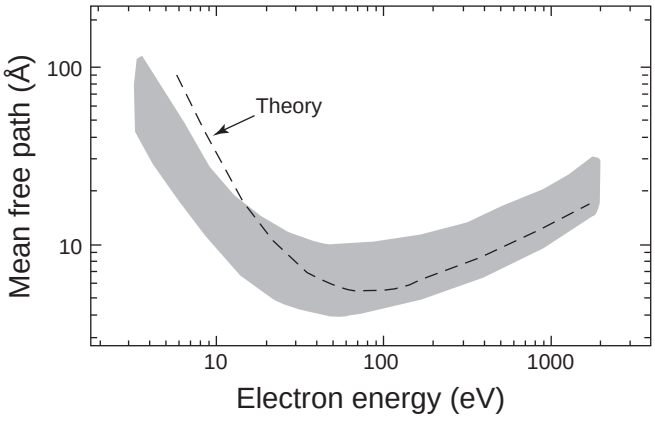

Figure 1 Mean free path of electrons in solids as a function of electron energy. Dashed line, theoretical prediction based on the Golden Rule (Penn, 1976, see Zangwill[1]); shaded region, range of experimental values for various elements. (Adapted from Zangwill.[1])

One level of interpretation of EELS spectra is based on the fact that many features of the vibrational excitations of molecules are essentially preserved when the molecules are adsorbed on a surface. This feature of characteristic vibrational signatures provides the basis for a powerful qualitative approach to EELS spectra analysis. Specific functional groups of atoms that make up more complex molecules exhibit characteristic vibrational signatures that can be used to identify these groups. The large body of vibrational data[15–21] obtained using EELS and other vibrational spectroscopic methods, including Raman scattering and IRAS, can be used to identify the chemical nature of adsorbed species based on characteristic sets of vibrational excitations.

A second level of interpretation of EELS experimental data exploits symmetry selection rules that govern the scattering intensities. In many cases of practical interest, the molecular orientation, the local symmetry associated with an adsorbate vibration or the symmetry of a collective vibrational excitation (phonon) at the surface can be determined by noting the number of modes detected, or by examining the scattering intensity under various scattering geometries indexed to the crystal symmetries of the surface. Application of these techniques requires certain flexibility in the instrumentation as well as the capability to determine and orient crystal symmetry directions in relation to the scattering directions.

A third level of interpretation of EELS experiments is based on analysis of the full range of quantitative information available from an EELS experiment. Analysis of EELS data on this level deals with vibrational loss line widths and shapes, energy-and angle-dependencies of the scattering cross-sections and the information obtained from full use of the kinematic scattering relationships. Interpretation of EELS experiments on this level can yield realistic values for interaction potentials, local coordination and symmetry of adsorbate sites, and information about coupling between vibrational and electronic

excitations. The approach generally requires large-scale computational exercises in which the lattice dynamics are simulated, or in which the electron scattering is evaluated based on multiple scattering analysis similar to dynamical analysis of LEED beam intensities.[22,23]

2 HISTORY

The detection of surface vibrational losses using inelastic electron scattering was first demonstrated by Propst and Piper[24] in 1967. While the energy resolution of their EELS spectrometer was only about 50 meV, the experiments, which involved chemisorption of H_2, N_2, CO and H_2O on W(100), clearly demonstrated the feasibility of probing vibrational excitations of chemisorbed species using electron scattering and established the high sensitivity of the method. Several years later, Ibach,[25,26] using an improved instrument, reported observation of surface phonons on ZnO($1\bar{1}00$), and on Si(111). This was followed by work by several experimental groups including Andersson,[27] Backx et al.,[28] Bertolini et al.[29] and others, as well as a number of theoretical papers[30–39] that presented formal electron scattering models providing a quantitative basis for exploiting the new technique fully. The advent of high-performance commercially available instruments based on those developed by Ibach,[40] and later by Kesmodel,[41] opened up the field to a broad range of surface scientists. The resulting enormous growth in published scientific work based on application of EELS techniques attests to the importance and popularity of this form of vibrational spectroscopy.

3 SAMPLE PREPARATION AND REQUIREMENTS

Application of EELS generally requires high quality single crystal surfaces. Collective surface vibrational modes (surface phonons) as well as the local vibrations of chemisorbed atoms and molecules are sensitive functions of the local structure and atomic coordination as well as the chemical environment. Lattice discontinuities, including surface steps, or the presence of impurity atoms, cause changes in vibrational excitation energies. EELS instruments are necessarily incorporated into UHV systems that generally also include surface preparation and characterization capabilities such as ion sputtering and sample heating for cleaning the crystal, and LEED and AES for probing surface crystal structure and chemical composition. A single crystal sample that yields a high quality LEED pattern characterized by sharp diffraction spots and low diffuse background will usually also yield a good EELS spectrum with narrow well-defined vibrational features. Many factors can broaden and shift vibrational loss peaks including substrate effects (steps, roughness) and chemical effects from impurity atoms, or coverage-dependent lateral interactions associated with chemisorbed atoms or molecules.

Clean and well-ordered surfaces can be prepared by a variety of methods. Many insulating or semiconducting crystals can be cleaved in vacuum yielding a clean well-ordered surface. Special flood gun techniques (described later) are required to deal with charging in cases where insulating samples are used. Some semiconducting surfaces can be passivated by chemical means and cleaned by simply heating the sample in UHV. Most metal crystals are prepared by cutting and aligning a suitable sample (typically 1 cm diameter $\times$ 1 cm thick) using spark wire or diamond saw technology followed by mechanical, chemical, electrochemical polishing and finally in situ cleaning using ion sputtering and/or high temperature annealing. Useful reviews of surface preparation,[42] mounting and manipulation techniques[43] are available.

4 ELECTRON SCATTERING MECHANISMS AND BASIS OF TECHNIQUE

The description of inelastic electron scattering can be conveniently divided into discussions that consider independently the scattering kinematics and conservation laws, and the scattering mechanisms.

4.1 Scattering Kinematics

Electron scattering from surfaces is governed by conservation laws and by several distinct scattering mechanisms. In certain cases involving high-symmetry geometry, selection rules can be used to draw conclusions about vibrational modes directly from general features of the scattering experiments. Single crystal surfaces are generally used in EELS experiments because translational invariance and geometrical uniformity of local adsorption sites are important requirements for interpreting experimental results within the framework of electron scattering models.

Figure 2 illustrates the parameters required to describe electron scattering from a crystal surface and introduces the notation to be used in following discussions. When an electron scatters elastically from the crystal (no energy exchange between the electron and the crystal), the conservation laws become the Laue diffraction conditions that govern diffraction processes, including LEED and reflection high-energy electron diffraction (RHEED),

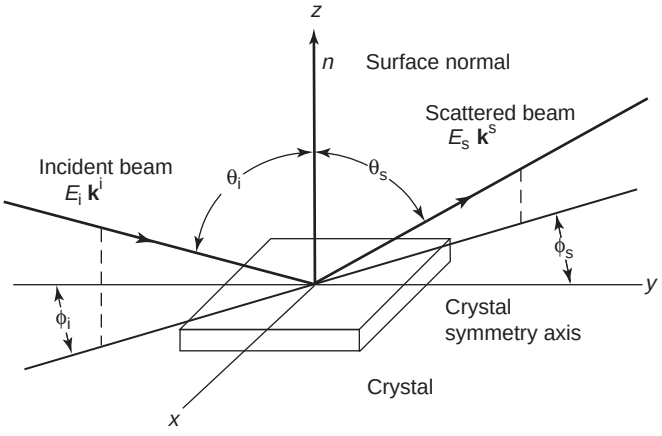

Figure 2 Scattering geometry and parameters required to specify an inelastic electron scattering event from a crystal surface.

Equations (1–3):

$$\mathbf{k}_{\parallel}^{s} = \mathbf{k}_{\parallel}^{i} + \mathbf{G}_{\parallel} \tag{1}$$

$$k_{\perp}^{s} = -k_{\perp}^{i} \tag{2}$$

$$k^{s} = k^{i} \tag{3}$$

where $\mathbf{k}_{\parallel}^{s}$ and $\mathbf{k}_{\parallel}^{i}$ represent the parallel component of scattered and incident electron wave vector, $k_{\perp}^{s}$ and $k_{\perp}^{i}$ represent the magnitude of the perpendicular component of scattered and incident electron wave vector and $\mathbf{G}_{\parallel}$ is a surface reciprocal lattice vector.[1]

When the scattering process involves the excitation of a single surface vibrational mode of energy $\hbar w$ and wave vector $\mathbf{q}_{\parallel}$, the conservation laws are, Equations (4) and (5):

$$E_s = E_i \pm \hbar w(\mathbf{q}_{\parallel}) \tag{4}$$

$$\mathbf{k}_{\parallel}^{s} = \mathbf{k}_{\parallel}^{i} \pm \mathbf{q}_{\parallel} + \mathbf{G}_{\parallel} \tag{5}$$

where E_s and E_i are, respectively, the scattered and incident electron energy. The scattering process can create or annihilate a phonon corresponding to the $-$ and $+$ signs. The relative probability of phonon creation and annihilation is governed by the Bose–Einstein statistics obeyed by phonon excitations. The gain-peak intensity at energy $\hbar w$ in an EELS experiment is proportional to the number of occupied vibrational states having energy $\hbar w$, which is proportional to the Bose factor $[\exp^{(-\hbar w/kT)} - 1]^{-1}$. The intensity of the corresponding loss peak is proportional to $1 + n$. Therefore the ratio of gain-peak to loss-peak intensity is given by $n/(1+n) = \exp^{(-\hbar w/kT)}$. Figure 3 presents a typical EELS spectra showing loss and gain peaks associated with excitation of a surface phonon on the Ni(100) surface at a specific value of $\mathbf{q}_{\parallel}$.

An electron can excite or annihilate more than one phonon through a sequence of single-phonon scattering

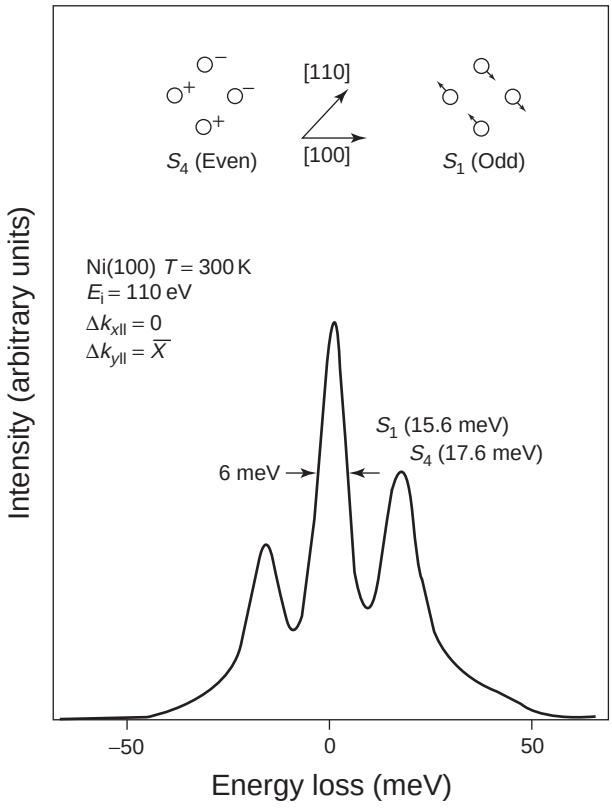

Figure 3 EELS spectrum from Ni(100) at room temperature under scattering conditions that probe the $\overline{X}$ point of the two-dimensional Brillouin zone. Inset at top illustrates direction of surface atom vibrational displacements. Both even and odd phonon modes (S_4 and S_1) which are nearly degenerate can contribute to the loss and gain peaks at 16 meV because an out-of-plane scattering configuration is used. Gain-to-loss peak ratio is approximately $\Delta E/kT \sim 0.64$ (refer to text); multiphonon background is approximately 18% of the elastic peak intensity.

events. A two-phonon scattering event is described by Equations (6) and (7):

$$E_s = E_i \pm \hbar w(\mathbf{q}_{\parallel}) \pm \hbar w(\mathbf{q}'_{\parallel}) \tag{6}$$

$$\mathbf{k}_{\parallel}^{s} = \mathbf{k}_{\parallel}^{i} \pm \mathbf{q}_{\parallel} \pm \mathbf{q}'_{\parallel} + \mathbf{G}_{\parallel} \tag{7}$$

where $\mathbf{q}'$ is the wave vector associated with a second scattering event. For multiphonon process involving excitation or annihilation of two or more phonons the number of variables exceeds the number of constraints represented by the conservation laws, and a continuum of loss/gain energies exists for each value of $\mathbf{k}_{\parallel}^{s}$. While single-phonon loss processes yield unique loss/gain peak energies for a prescribed scattering geometry ($\mathbf{k}_{\parallel}^{i}$ and $\mathbf{k}_{\parallel}^{s}$) corresponding to branches of the phonon dispersion curves, multiphonon processes yield a broad continuum. Fortunately, in most cases, single-phonon peaks are easily resolved above the multiphonon background (see

Figure 3), and measurements of the peak energies as a function of momentum transfer $\Delta\mathbf{k} = \mathbf{k}_{\parallel}^{s} - \mathbf{k}_{\parallel}^{i}$ can yield the energy–momentum (dispersion) relations of phonon excitations. The dispersion relation ($\hbar w$ versus $\mathbf{q}_{\parallel}$) contains useful information about interaction potentials associated with surface atoms.

4.2 Scattering Mechanisms

The conservation laws Equations (1–7) specify the constraints that govern a scattering event, but do not provide any information about the probability that a specific scattering event will occur. EELS peak positions are determined by phonon dispersion curves, and conservation laws and the relative strength of loss and gain peaks are governed by the temperature-dependent probability that a specific vibrational energy level is excited. In order to predict the peak intensities for phonon or other vibrational excitations at the surface, an understanding of the scattering mechanisms is required. Four unique scattering mechanisms have been identified that account for various observations of inelastic electron scattering from surfaces.[3–6]

4.2.1 Dipole Scattering

Most EELS experiments are carried out using specular scattering geometry ($\theta_i = \theta_s$ and $\phi_i = \phi_s = 0$) and using low incident energies, $E_i \sim 1$–10 eV. Under these scattering conditions the dipole scattering mechanism usually accounts for the major contributions to EELS peaks. Dipole scattering theory has been discussed extensively in the literature.[30–37] The basis of dipole scattering is the interaction of the incident electron and its image charge (which produce a dipole field) with long-range electric fields produced by electronic or vibrational excitations. An important characteristic of the dipole scattering mechanism is that the angular spread of scattered electrons is small and approximately equal to $\hbar w/E_i$, where $\hbar w$ is the excitation energy (loss energy) and E_i is the incident electron energy. The nature of the dipole scattering mechanism leads to inelastic scattering processes occurring away from the surface (typical distances ~ 100 Å) either before or after (specular) elastic scattering at the surface. The excitation probe depth of dipole scattering is of the order of the characteristic screening length for electrons in the material which is of the order of 1 Å for metals.

Dipole scattering formalisms can be applied to inelastic scattering resulting from adsorbate vibrations, surface phonons, and electronic excitations including electronic transitions and surface plasmons. The narrow angular spread of scattered electrons (or equivalently the small momentum transfer) associated with dipole scattering renders this mechanism of limited use in studying surface phonons away from the center of the two-dimensional Brillouin zone. When electronic excitations near the surface dominate the inelastic scattering, formulations of dipole scattering in terms of the dielectric response function[44–48] are useful. It has been shown that dielectric response determined by EELS compares well with that obtained from optical techniques. In semiconductor materials, where electron densities are low, loss energies are small ($\sim$1 meV) and measurements of the width of the quasi-elastic peak interpreted using dielectric response models provide a useful means of probing low-energy excitations at semiconductor surfaces.[44–48]

4.2.2 Impact Scattering

A second scattering mechanism that dominates surface inelastic electron scattering at higher incident energies and is characterized by large-angle scattering (large momentum transfer) is known as impact scattering.[38,39] The existence of this scattering regime is well known from inelastic scattering from gas phase molecules[49] where image charge effects and long-range dipolar interactions do not exist. The diffuse background observed in LEED experiments[22,23] is another manifestation of a large-angle scattering mechanism. The temperature-dependent diffuse background exhibits energy-dependent angular patterns that reflect the crystal symmetry, suggesting that useful information is contained in the diffuse scattering spectra and angular distributions. Figure 4 displays the calculated incident energy dependence of the inelastic electron scattering cross-section of S_1 and S_4 phonons (see Figure 3) for Ag(100) under specific scattering conditions prescribed by $\Delta\mathbf{q}_{\parallel}$. It is clear that the cross-section varies substantially (S_1 varies over a factor of 100) over the energy range of interest. Such calculations are very useful in interpreting experimental data when closely spaced phonon energies are probed as in Figure 3. The excitation probe depth for impact scattering is governed by the inelastic mean free path of electrons in the material at the incident energy, $\lambda_e(E_i)$.

4.2.3 Negative Ion Resonances

Temporary trapping of an incident electron is an important mechanism in vibrational excitation of a gas phase molecule. Strong coupling to various decay channels associated with a molecule adsorbed at a surface tends to oppose the formation of negative ion resonances in surface electron scattering;[50] however, various experiments[6,51–53] have shown that negative ion resonance scattering exists for both physisorbed and chemisorbed molecules. Negative ion resonance scattering is characterized by an incident threshold energy and broad angular distributions characteristic of short-range interactions. Negative ion resonance scattering can

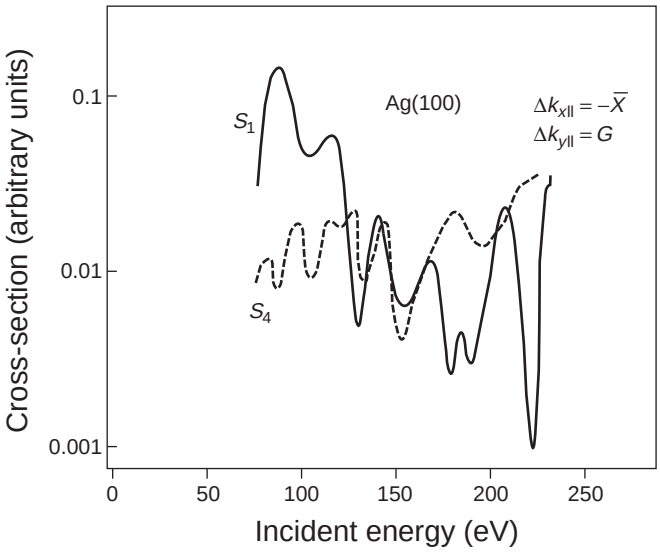

Figure 4 Calculated EELS impact scattering energy-dependent cross-section for even (S_4) and odd (S_1) symmetry phonons at the $\overline{X}$ point of Ag(100). (With permission from J.L. Erskine, E.-J. Jeong, J. Yater, Y. Chem and S.Y. Tong, 'Detection of Odd-symmetry Shear Modes at Metal Surfaces by Inelastic Electron Scattering: Experiment and Theory', *J. Vac. Sci. Technol.*, **A8**, 2649, 1990.) The rapid energy variation versus kinetic energy results from multiple scattering.

generally be identified by the characteristic threshold energy and the angular scattering patterns that can be calculated for oriented molecules. This type of scattering has been reviewed by Palmer and Rous.[6]

4.2.4 Surface Resonance Scattering

Electron diffraction from a crystal surface can lead to elastic (and inelastic) scattering resonances in which the electron is temporarily trapped in a propagating surface state above the vacuum level.[54–61] The Laue conditions, Equations (1–3), can be solved for the threshold incident beam energies at which LEED beams incident at θ_i emerge from the crystal, Equation (8):

$$E_{th} = \left(\frac{G}{0.512(1+\sin\theta_i)}\right)^2 (\text{eV}) \qquad (8)$$

where $(2m/\hbar^2)^{1/2} = 0.512\,\text{eV}\,\text{Å}^{-1}$ and $G = 2\pi/a\,\text{Å}^{-1}$ is a reciprocal lattice vector. Electrons in these surface resonance states can suffer inelastic losses. Experimental evidence for this loss mechanism is well established,[62–65] and theoretical treatments[66,67] of the surface resonance loss mechanism are available.

4.3 Scattering Cross-sections and Selection Rules

Analysis of the energy- and angular-dependent scattering cross-sections generally requires detailed numerical calculations. Multiple scattering is strong in the energy range (20–300 eV) generally used in inelastic electron scattering experiments, therefore, the phonon loss scattering cross-sections exhibit prominent and numerous peaks similar to those observed in LEED intensities. The symmetry of the crystal and the scattering conditions can often lead to predictions of scattering nulls without extensive calculations. The scattering nulls and general characteristics of scattering cross-sections imposed by symmetry are called selection rules.

Dipole scattering cross-sections[30–37] are proportional to the strength of surface dynamic dipole moments. Dipole scattering generally dominates inelastic loss mechanisms at low incident energies (1–10 eV) when losses are measured in specular scattering geometry. A dipole oriented perpendicular to a metal surface induces an image dipole that enhances the dipole field, whereas a dipole oriented parallel to the surface induces an image dipole that tends to cancel the field of the original dipole moment. This reasoning leads to the dipole selection rule that states: dipole scattering probes primarily surface vibrational losses associated with dynamic dipole moments oriented perpendicular to the surface.

The impact scattering mechanism,[38,39] in principle, permits probing vibrational modes of all symmetries and orientations using EELS. However, impact scattering cross-sections tend to be one to two orders of magnitude weaker than dipole scattering cross-sections. Selection rules[68] govern null scattering conditions for impact scattering. For example *x*-polarization modes (Figure 2) having dynamic displacements perpendicular to the scattering plane cannot be observed in a scattering experiment with $\phi_i = \phi_s = 0$ if the crystal has reflection symmetry about the scattering plane. Similarly, odd symmetry (S_1) phonon modes, for example (Figure 3), cannot be observed in a scattering geometry in which $\phi_i = \phi_s = 0$. Table 1 summarizes scattering selection rules for parallel (in-plane) mode vibrations.

Table 1 Impact scattering selection rules

X Modes (parallel out-of-plane vibrations)		
Symmetry	Rule	
Reflection (R_{yz})	$\sigma(k) = 0$ in yz plane	
Rotation $(R_\pi)z_{axis}$	$\sigma(k) = 0$	$\lvert k_i\rvert = \lvert k_f\rvert$
Y Modes (parallel in-plane vibrations)		
Symmetry	Rule	
Reflection (R_{yz})	$\sigma(k) = 0$	$\lvert k_i\rvert = \lvert k_f\rvert$
Rotation $(R_\pi)z_{axis}$	$\sigma(k) = 0$	$\lvert k_i\rvert = \lvert k_f\rvert$

R, rotation about an axis or reflection about a plane; i, initial state (incident wave vector); f, final state (scattered wave vector).

5 INSTRUMENTATION

Figure 5 presents a schematic diagram of a high-performance commercial EELS instrument (adapted from technical information furnished by L.L. Kesmodel, L.K. Technologies Inc., Bloomington, Indiana, USA). The monochromator is designed to deliver a beam of electrons having a narrow angular width and energy spread to the sample. Scattered electrons are detected in an analyzer with high angular and energy resolution. Important parameters that characterize EELS instrument performance include energy and angular resolution, range of energy and scattering angles, monochromator current, analyzer sensitivity, background counting rate, and stability. Numerous technical articles,[69–73] and a book[40] contain material covering all aspects of the theory, design, construction and optimization of EELS monochromators, analyzers and integrated instruments. This section only covers the most basic features of EELS instrumentation in relation to intended applications.

State-of-the-art commercially available instruments have achieved energy resolution exceeding 1 meV under ideal conditions. High-performance instruments currently available achieve specifications listed in Table 2 (adapted from technical information furnished by L.L. Kesmodel, L.K. Technologies Inc., Bloomington, Indiana, USA). These performance specifications cover essentially all of the requirements for EELS studies of adsorbate vibrations, electronic excitations, and surface phonons.

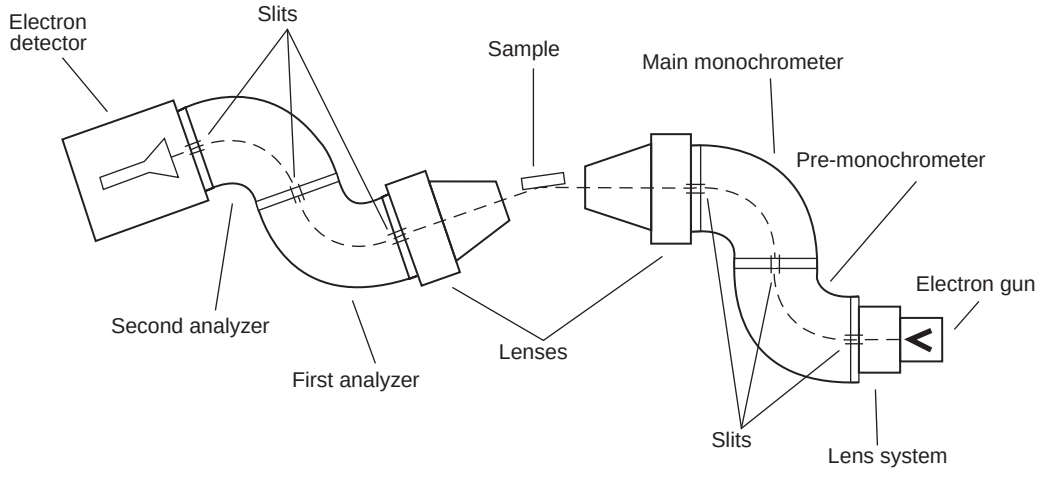

Figure 5 Schematic diagram of a commercial high-performance EELS instrument; refer to Table 2 for specification. (Adapted from L.K. Technologies Technical Data with permission.)

Table 2 Capabilities and specifications of a commercial high-performance EELS instrument (Ibach design manufactured by L.K. Technologies, Inc.) (see Figure 4)

Mechanical and electronic specifications		
Electron gun	LaB_6 cathode with repeller	
Monochromator/analyzer	Dual tandem 127° cylindrical sectors	
Lens system	Symmetric monochromator/analyzer, noncylindrical plate lenses. In-plane out-of-plane focus characteristics optimized to 127° sectors/sample distance	
Detector:	Single channel detection; channeltron	
Energy range	2–200 eV	
Scattering angle range	0–90° beam deflection in-plane only; $\phi_i = \varphi_s = 0$	
Resolution	**Current**	**Angular spread of beam**
fwhm	Detector current	Lens acceptance
2 meV	7 pA	±1.5° In dispersive plane ±1.7° Perpendicular to plane
4 meV	100 pA	±0.4° In dispersive plane ±1.7° Perpendicular to plane

fwhm, full width at half maximum.

EELS studies of surface adsorbates are generally carried out at low incident electron beam energies, $E_i < 10\,eV$, and using specular scattering geometry, $\theta_i = \theta_s$, $\phi_i = \phi_s = 0$. Under these conditions, the dipole scattering cross-section is maximized and very high detection sensitivity can be achieved, i.e. 10^{-3} monolayer of CO can be detected. Angular resolution is generally not a concern in this regime because dipole scattering is sharply peaked along the specular direction. EELS studies of surface phonon (and plasmon) dispersion relations require higher incident energies and also require attention to the angular spread of the incident beam and analyzer angular acceptance. Measurement of energy losses at large momentum transfer (corresponding to a Brillouin zone boundary $\Delta k_\parallel \sim \pi/a \sim 2.5\,\text{Å}^{-1}$) requires higher energies: the momentum transfer associated with a surface phonon loss (where $\Delta E/E_0 \ll 1$) is governed by Equation (9):

$$\sqrt{\frac{2m}{\hbar^2}} E_i(\sin\theta_i - \sin\theta_s) = \Delta k_\parallel \tag{9}$$

where E_i is the incident energy, and θ_i and θ_s are the incident and scattering angles. Evaluation of Equation (9) for a range of incident energies reveals that the entire Brillouin zone of a typical crystal can be accessed at $E_i = 200\,eV$ with a relatively small scattered electron angular range (15°) whereas for $E_i \sim 10\,eV$, a much larger (and partially inaccessible due to mechanical constraints) angular range $\sim 70°$ is required.[74] Formulas for calculating uncertainties in $\Delta k_\parallel$ as a function of the finite angular acceptance of the analyzer and monochromator beam width are given by Ho.[5]

Several groups have developed EELS instruments for specialized applications. Ho has developed an EELS instrument optimized for time-resolved studies.[75–77] This instrument incorporates a high-speed multichannel detection system that permits an EELS spectrum to be measured on short timescales (<100 ms). Erskine has developed an EELS instrument having more geometrical flexibility (monochromator and analyzer on independent orthogonal rotation goniometers) that permits inelastic scattering studies using a broad range of incident and scattering configurations.[70,78] High sensitivity (50 channels) and resolution of 2.5 meV is achieved using multichannel detection.

6 APPLICATIONS TO INTRINSIC PROPERTIES OF SURFACES

Principal applications of EELS techniques to surface analysis are outlined in this and the following section which divide the surface properties evaluated into intrinsic properties of a clean surface, and more complex systems in which a chemical adsorbate is present.

6.1 Fuchs–Kliewer Phonons

Polar surfaces of ionic crystals (ZnO $(1\bar{1}00)$ and GaAs$(1\bar{1}0)$ for example) exhibit infrared-active longitudinal optical surface phonons known as Fuchs–Kliewer phonons.[79] Some of the earliest work demonstrating the sensitivity of EELS to surface phonons was carried out on these surfaces.[25,26] The signature of these phonons is a series of intense, equally spaced loss peaks having intensities described by a Poisson distribution. Loss and gain peak intensities scale according to Bose–Einstein statistics as expected.

6.2 Bulk Phonon Resonances at Metal Surfaces

Certain metal surfaces (i.e. the (110) surfaces of face-centered cubic metals) exhibit strong surface resonances.[4,5,80,81] The resonance is a result of a longitudinal bulk phonon that produces a dipole active mode with a strong scattering cross-section at the surface due to the surface atomic geometry.

6.3 Surface Phonon Dispersion

The first direct measurements of the dispersion of surface phonons throughout the two-dimensional Brillouin zone were obtained using inelastic He atom scattering.[82] However, improvements in EELS spectrometers soon permitted similar measurements based on inelastic electron scattering.[74] Full measurements of phonon dispersion relations ($E(\mathbf{q}_\parallel)$ versus $\mathbf{q}_\parallel$) along high-symmetry directions of the two-dimensional Brillouin zone can be used to study surface interaction potentials by comparing the experimental phonon dispersion curves with those calculated using lattice dynamics methods.[83,84] A typical EELS spectrum showing loss and gain peaks from phonons on Ni(100) is shown in Figure 3. It is now possible to calculate lattice dynamics based on first principles approaches. Thus, measurements of surface phonon energies throughout the two-dimensional Brillouin zone supported by analysis of the surface lattice dynamics offers important opportunities for improving understanding of surface phenomena associated with interaction potentials, i.e. surface relaxation and reconstruction as well as film nucleation and growth.

6.4 Semiconductor Surfaces

EELS can be used as a probe of the transport properties of semiconductor surfaces.[3–5,85,86] The Fermi energy in semiconducting materials lies in a gap in the electronic structure. At low temperatures and low beam energies,

the elastic peak width for semiconductors (and insulators) is very narrow due to the absence of very low-energy electronic excitations (no free carriers). An increase in temperature (or equivalently doping) leads to a finite and increasing free carrier concentration. This in turn leads to increased electron scattering and excitation of low-energy plasmons; surface energy loss mechanisms become possible and contribute to the elastic-peak width. The manifestation of the temperature-dependent loss mechanisms is a strong temperature dependence of the elastic-peak width. Theoretical models of this quasielastic scattering[4,5,46–48] permit calculation of bulk carrier density and estimates of other important parameters such as band gaps and effective masses of carriers. In the case of Si(111) 7×7, analysis of the quasielastic peak has led to a model of the reconstructed surface in which two-dimensional metallic behavior is ascribed to the clean surface. The metallic character arises from surface states in the band gap. Adsorption of 1% of a monolayer of H on the Si(111) 7×7 was found to quench the metallic surface properties effectively as evidenced by suppression of elastic-peak broadening. Corresponding studies of GaAs(100) surfaces have explored similar surface conductivity effects.[48]

6.5 Ultrathin Metal Films

A thin metal film deposited on an insulating or semiconducting surface will behave as a two-dimensional conducting system. The same dielectric response formalisms developed to analyze quasielastic scattering in terms of surface conductivity apply to the thin film systems. EELS studies of ultrathin films deposited on semiconductor surfaces[85,86] have been interpreted within the dielectric models to yield the frequency-dependent film conductivity and used to characterize the effects of film microstructure on electron transport in the films.

6.6 High Critical Temperature Cuprate Superconductors

EELS has been used to explore electronic and vibrational properties of $YBa_2Cu_3O_7$ and $Bi_2Sr_2CaCu_2O_8$ single crystals. Initial experiments[87,88] detected temperature-dependent loss features that were interpreted as resulting from high-conductivity regions of the surface, and attributed to the superconducting energy gap. Subsequent experiments[89,90] supported by a model[90] for dipole inelastic electron scattering from a superlattice consisting of anisotropic-dielectric layers (modeling the cuprate materials) have shown that all loss features are nicely accounted for based on surface optical phonons. This application is a good example of the importance of using meaningful theoretical models to support the interpretation of EELS measurements, and an additional demonstration of the equivalence of dielectric response as measured independently by optical and inelastic electron scattering techniques.

6.7 Electronic Excitations at Surfaces/Surface Plasmons

Modern EELS instruments have recently been used to study collective surface electronic excitations (surface plasmons). Existing theoretical models[91–93] suggest that knowledge of the surface–plasmon dispersion relations can lead to better understanding of intrinsic surface properties such as electron screening and the charge-density profile normal to the surface. The measurements of surface plasmon dispersion are similar to those used to determine surface–phonon dispersion. In the case of surface–plasmon dispersion measurements, it is no longer valid to assume that the electron energy loss $\Delta E = E_i - E_s$ is much smaller than the incident energy E_i, and an exact form of Equation (9) is required, Equation (10):

$$\sqrt{\frac{2m}{\hbar}E_i}\ \sin\theta_i - \left(1 - \frac{\Delta E}{E_0}\right)^{1/2} \sin\theta_s = \Delta k_{\parallel} \qquad (10)$$

EELS experiments[94] studying surface plasmons on K and Na confirmed theoretical predictions based on the *jellium* model of the electron gas: the energy of the surface plasmon at small values of $k_{\parallel}$ was found to be equal to $w \approx w_p\sqrt{2}$ (where $w_p = 4\pi ne^2/m$ the bulk plasmon frequency) and independent of the surface charge profile. Also, the dispersion at small $\mathbf{q}_{\parallel}$ was found to be negative. Corresponding experiments on Ag[95,96] are beginning to clarify how surface plasmons are affected by more complex electronic structure.

7 APPLICATIONS TO CHEMISORBED ATOMS AND MOLECULES

Applications of EELS to surface chemical phenomena cover a very broad range of topics and the literature is extensive. This section highlights a few of the more important applications of EELS to the study of surface chemical phenomena by discussing selected ''textbook'' examples in sufficient detail to illustrate principal concepts.

7.1 Atomic Adsorbates

The simplest surface chemical systems involve dissociative chemisorption of diatomic molecules (or in some cases more complex gas phase species) that leave a single atomic species adsorbed at the surface. Many examples

of this type of system exist. Two systems will be discussed here: H on W(100)[97–99] and O on Ni(100).[100–103]

The saturated coverage of H on W(100) (β_1 phase) consists of H atoms bridge bonded to W surface atoms yielding two H atoms per surface unit cell. Saturated hydrogen coverage on W(100) stabilizes the surface W atoms into a $p(1 \times 1)$ bulk termination; i.e. the novel order–disorder surface reconstruction[81] observed on clean W(100) below room temperature does not occur on the β_1 H/W(100) surface. A single vibrational mode v_1 at 130 meV is measured by EELS under specular scattering conditions at low impact energies. The 130 meV (dipole) vibrational loss corresponds to the symmetric stretch (perpendicular vibration) mode. Off-specular scattering studies[97–99] reveal new (impact scattering) modes at $v_2 = 80$ meV, $v_3 = 160$ meV and $2v_1 = 260$ meV corresponding to vibrations parallel to the surface: v_2 and v_3 are assigned to vibrational displacements perpendicular and parallel to the bridge axis respectively, and the $2v_1$ mode is an overtone of the strong perpendicularly polarized mode. The energy and angular dependencies of these modes and similar results obtained for D_2 (isotope substitution yields the same modes shifted to lower frequencies by $\sqrt{2}$) have provided important insight into scattering mechanisms,[39,64] mode assignments (based on orientation and scattering angle dependencies of mode intensities) and selection rules.

EELS experiments carried out using scattering conditions under which LEED beams emerge from the crystal have yielded evidence of the surface resonance scattering mechanism.[62,99] Rapid variations in inelastic scattering cross-sections as a function of incident energy, E_i, or incident parallel wave vector, $\mathbf{k}_\parallel^i$, are apparent in EELS spectra obtained for scattering parameters close to LEED beam emergence conditions. Isotope substitution confirms the H adsorbate origin of the loss peaks.

Detailed analysis of the EELS peaks of β_1H and β_1D on W(100) reveal that the vibrational loss energies are functions of momentum transfer $\mathbf{q}_\parallel$ along high-symmetry directions of the two-dimensional Brillouin zone.[99] Dispersion of the loss peaks as a function of $\mathbf{q}_\parallel$ is a manifestation of a collective excitation (in this case an adsorbate phonon). When lateral interactions between chemisorbed H atoms on W(100) are taken into account, the degeneracy of the three vibrational modes associated with noninteracting bridge-bonded H is broken yielding a possible total of six phonon modes (two atoms per unit cell × three degrees of freedom per atom). A lattice dynamical analysis[99] of two- and three-body interactions for β_1H on W(100) accounts for the experimentally observed phonon modes in terms of weak (1/10 of the H-W force) repulsive lateral forces between the chemisorbed H atoms.

Similar EELS experiments and lattice dynamical analysis of $c(2 \times 2)$ and $p(2 \times 2)$ O on Ni(100), and other ordered chemisorbed layers,[101–103] have been carried out. These studies demonstrate the sensitivity of substrate and adsorbate phonon dispersion to overlayer structure (height of oxygen atoms over the substrate), as well as the technique of using lattice dynamical analysis of surface phonons to obtain adsorbate–adsorbate and adsorbate–substrate interaction potentials. Experimentally determined phonon dispersion curves interpreted within lattice dynamical models have been useful in exploring the types of interaction potential most suitable for analyzing surface phonon data (very good fits were obtained with nearest neighbor only interactions, and no angle-bending forces). The studies also suggest that various features in the dispersion curves can be interpreted in terms of several types of lateral interaction including short-range direct coupling between adsorbates, indirect coupling through the substrate, and longer range dipole–dipole coupling.[104]

7.2 Underlayer Formation at Metal Surfaces

In selected metal adsorbate systems, dissociative chemisorption can be accompanied by adsorbate place exchange or diffusion perpendicular to the surface normal direction yielding adsorbate atoms in subsurface sites near the surface. Systems that exhibit this type of behavior include O/Al(111),[105,106] H/Nb(100)[107–109] and H/Ni(111).[110] Underlayer modes can be probed via both dipole[105] and impact[110] scattering mechanisms. For dipole scattering by a dipole located below the surface, the important parameter is the strength of the screened dipole potential. Using the Thomas–Fermi model to characterize the screening length, by the Thomas–Fermi wave vector, k_{TF}, the potential of a dipole located a distance d below the surface scales as $\phi \sim 2\exp^{(-k_{TF}d)}$. For impact scattering, the important factor is the inelastic mean free path of the incident electron. The scattering intensity scales as $I \sim \exp^{(-2d/\xi_e \cos\theta)}$ where $\xi_e(E_i)$ is the inelastic mean free path of incident electrons of energy E_i.

Oxidation at the Al(111) surface is a complicated process involving the formation of a subsurface oxygen structure at the earliest stage of chemisorption. EELS studies[105,106] of this system yielded the first definitive evidence suggesting that the initial stage of oxidation at Al(111) involved the simultaneous formation of chemisorbed surface and subsurface oxygen. Oxygen dose and substrate temperature-dependent studies of EELS peak intensities (see Figure 6) revealed a tendency of subsurface sites to becomes populated at the expense of surface sites at fixed oxygen dose, as well as the evolution of the loss spectra signature toward the spectra characteristic of Al_2O_3 as dose and temperature were

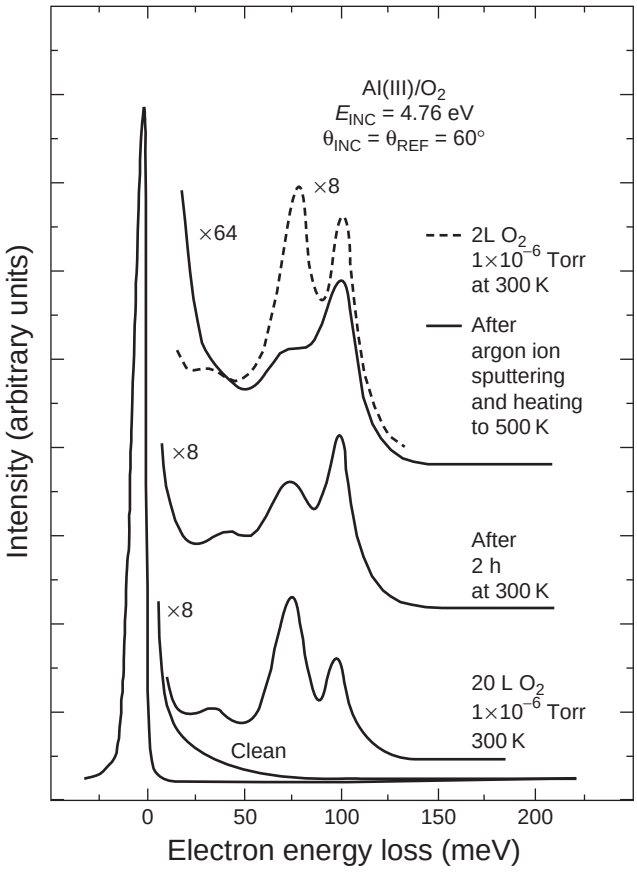

Figure 6 EELS spectra showing oxygen-induced loss features that accompany the oxidation of Al(111).

increased. Lattice-dynamical modeling[106] was able to account for the observed dipole-active modes in terms of a surface structural model that included both surface and subsurface oxygen atoms.

A second system in which subsurface atoms have been probed by EELS is H at Ni(111).[110] Dissociative adsorption of H_2 at Ni(111) leads to a $p(1 \times 1)$ H overlayer with H atoms occupying three-fold hollow surface sites. A suitably long exposure of Ni(111) to atomic H (dissociated by a 1800 K tungsten filament) results in the population of subsurface sites by H atoms. Manifestations of the subsurface adsorption sites include new peaks in the thermal desorption spectrum and a new peak in the off-specular scattering spectrum: the surface H yields peaks at 145 meV (symmetric stretch) and 118 meV (asymmetric stretch); the subsurface species produces a new peak at 99 meV. Off-specular scattering studies of the 99 meV loss peak revealed that it is produced by impact scattering. Impact energy dependence of the same peak confirms that its intensity is compatible with an inelastic mean free path of $\xi_e \sim 2.05$ Å, corresponding to rapid decay of the electron penetration depth at the surface. In this case, dipole scattering is apparently very weak for the subsurface site, and the impact scattering vanishes in specular scattering based on a selection rule assuming a site of octahedral symmetry.

7.3 Molecular Adsorbates

EELS has been used extensively as an analytical tool to probe the adsorption and dissociation of molecules at metal[104,111,112] and semiconducting[113–117] surfaces. Extensions of these studies include investigations of surface-mediated chemical reactions, studies of model-supported catalysts, alkali metal promotion, and a broad range of technologically important surface processes.[4,5] As the size and complexity of molecules increases, the number of vibrational modes increases and an EELS spectra can become complex. Analysis of complex EELS spectra are facilitated by the broad range of source material that tabulates molecular vibrational spectra,[15–21] and the use of isotopic substitution techniques and angle dependencies which permit exploiting selection rules. In many cases, distinct groupings of frequencies associated with stable molecular fragments can be used to simplify mode assignments. For details, refer to sections of reviews[4,5] dealing with complex molecules, scientific papers by analytical chemists, and the books containing tabulations of vibrational frequencies.[15–21]

7.4 Technical Surfaces

While most applications of EELS techniques have focused on studies of electronic and vibrational excitations of well-characterized single crystal surfaces, some recent efforts have expanded applications to a class of surfaces that can be defined as technical surfaces (see Figure 7). The term *technical* surface is a descriptively appropriate term for surfaces such as supported catalysts (metal clusters on an oxide substrate), thin SiO_2 insulators found in metal oxide semiconductor (MOS) technology, the surface of a Si wafer after a wet chemical step (HF dip cleaning) in a semiconductor Fab line, or a polymer surface. Several technical articles[118–121] and reviews[122,123] have addressed application of EELS to these types of surfaces. The same basic principles that govern scattering mechanisms and interpretation of loss peaks that apply to EELS studies of single crystal surfaces also apply to technical surfaces. However, at a practical level, there are significant differences: any surface involving an insulator is likely to exhibit effects associated with charging by the electron beam; diffuse elastic scattering will generally be much stronger (factor 10–100) due to the rougher nature of the surface; dipole scattering will generally be significantly reduced; and inhomogeneous broadening of loss

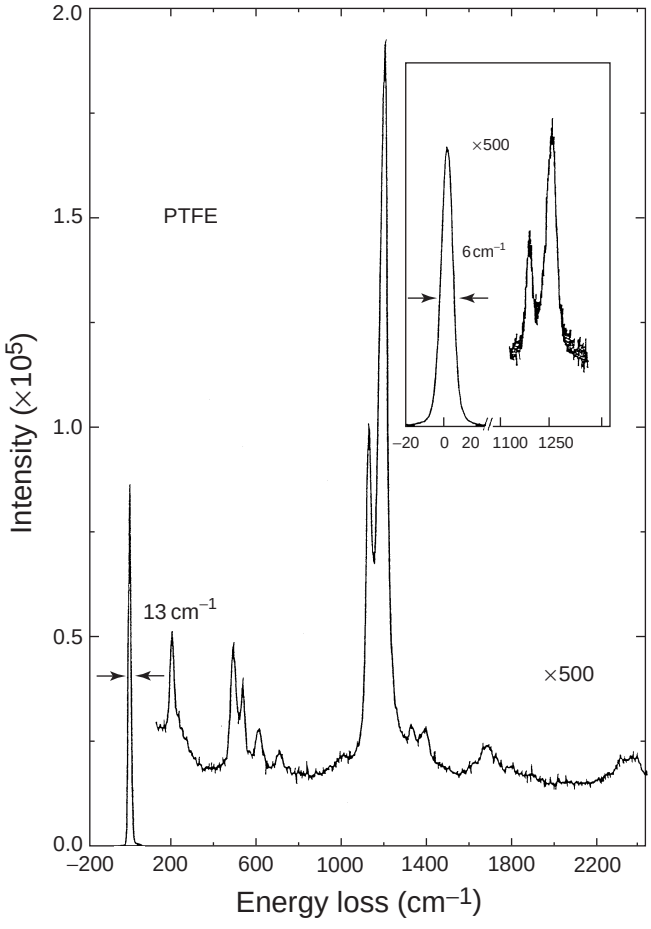

Figure 7 EELS spectra of polytetrafluoroethylene (PTFE). The figure illustrates the high-performance level achieved by modern commercial EELS instruments (inset shows energy resolution below 1 meV). From these results it is clear that useful vibrational loss features can be obtained from a technical surface. (Adapted from Akavoor et al.[121] with permission from the author.)

peaks will occur resulting from greater nonuniformity of adsorption sites.

Charging is generally not a problem in ultrathin insulating layers and semiconductor surfaces, but in bulk insulators and thick insulating films, electron beam charging must be neutralized by a secondary electron gun flooding the sample (1 kV energy, 100 nA cm^{-2} beam current).[118,119] The mechanism for surface charge elimination apparently involves the creation of a low-density surface plasma that conducts excess charge to an appropriate reference potential, the spectrometer ground.

EELS has been successfully applied to the analysis of surface chemistry involving a model supported catalyst.[120] A thin film of alumina was grown by reactive Al deposition on a metal substrate in an oxygen atmosphere. The film was thin enough for tunneling or other mechanisms to prevent charging during EELS experiments. Vapor-deposited platinum (of the order of one monolayer effective coverage) produced Pt clusters ranging from 1.0–1.6 nm diameter due to the nonwetting properties of Pt on alumina. Loss features for π and di-σ-bonded ethylene species were observed after ethylene adsorption at 165 K (the dominant scattering mechanism was impact scattering and intensities were low), and upon warming the sample to 325 K, principal loss features corresponding to the formation of ethylidyne were clearly observed.

EELS has also recently been applied to the analysis of Si and SiO_2 surfaces during wafer processing.[119] Analysis of chemicals on the SiO_2 surfaces required suppression of surface charging by using a secondary electron flood gun. The sensitivity of EELS to surface hydrogen was found to be very useful in examining the Si–H and Si–H_2 modes; SiC groups and other molecular groups on silicon surfaces that form during various cleaning and processing steps could also be monitored.

As a final example of the use of EELS in examining technical surfaces, recent work has developed useful applications to polymer surfaces.[121,122] The ability of EELS to serve as an analytical probe of selected characteristics (surface uniformity or other quality control parameter) is already evident. Additional improvements should broaden applications to polymer and other technical surfaces.

8 COMPARISON WITH OTHER SPECTROSCOPIC METHODS

The EELS technique probes surface species and intrinsic surface properties by measuring energy loss of scattered electrons resulting from vibrational or electronic excitations. A broad range of surface-sensitive probes are available that can detect surface species, and several alternate spectroscopic techniques are available that are sensitive to vibrational and electronic excitations. In most research efforts, a single technique is not able to provide all of the required structural, compositional, chemical, vibrational or electronic information, and a combination of probes is usually incorporated into the instrumentation.

Surface vibrational excitations can be probed by IRAS, Raman scattering spectroscopy, and by neutral atomic beam scattering, usually HAS. The optical probes, IRAS and Raman scattering, achieve higher energy resolution (typically 0.1–0.5 meV) and offer the ability to probe surfaces under high ambient pressures. The sensitivity of the optical probes tends to be lower than EELS due to weaker coupling of photons to vibrational (and electronic) excitations, although the stability and intensity of synchrotron radiation and new laser technology applied

respectively to IRAS and Raman spectroscopy now permit studies at low surface concentrations.

IRAS probes vibrational modes having dynamic dipole moments perpendicular to the surface and can only detect surface phonons at $\Delta\mathbf{q}_{\parallel} = 0$; EELS can, in principle, detect vibrational modes of all polarizations, and can measure the energy loss and momentum transfer associated with scattering from phonons throughout the two-dimensional Brillouin zone. HAS achieves high-energy resolution (0.1–0.5 meV) and also exhibits high-surface sensitivity resulting from the surface scattering mechanism. The energy range of the incident beam is limited to relatively low values ~0.1 eV, therefore, atom beam scattering is best suited for studying low-energy surface phonons (below ~60 meV). A large fraction of our experimental knowledge of surface lattice dynamics has been obtained using He scattering.

Surface structure and chemical composition can be probed by EELS, and by other surface sensitive techniques. Vibrational excitations are sensitive to symmetry and local geometry, and, in many cases, the number and symmetry of vibrational excitations can yield information on the local environment of a chemisorbed atom. Other probes such as LEED, photoelectron diffraction (PED), and scanning tunneling microscopy (STM) offer more direct structurally sensitive means of determining atomic level geometry at surfaces.

AES and ESCA are highly sensitive and accurate probes of atomic species at surfaces. Both AES and ESCA are atomic probes, and rely on relative strengths of electron emission from atomic levels to infer chemical composition. Auger spectroscopy is unable to detect hydrogen. Other elements can generally be probed by AES at a detection limit of ~0.01 monolayer equivalent coverage. EELS has the advantage of probing energy losses that are characteristic of a molecule (a vibrational excitation) and therefore generally provides a better chemical probe of surface species.

An additional and very important complementary surface sensitive technique that is often used in conjunction with EELS is TDS. The technique involves measurement with a mass spectrometer of the molecular (and in rare cases atomic) species that are desorbed from a surface as the temperature of the crystal is ramped. The TDS technique can be extremely sensitive (individual ions, molecules or atoms can be detected) and can provide detailed information about desorbed species, and their surface binding energies.

ACKNOWLEDGMENTS

The author wishes to thank Sylvia Ceyer, Wilson Ho, Larry Kesmodel and John Yates, Jr., for reprints of their work. This work was supported by the Robert A. Welch Foundation.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
EELS	Electron Energy Loss Spectroscopy
ESCA	Electron Spectroscopy for Chemical Analysis
fwhm	full width at half maximum
HAS	Helium Atom Scattering Spectroscopy
IRAS	Infrared Reflection Absorption Spectroscopy
LEED	Low-energy Electron Diffraction
MOS	Metal Oxide Semiconductor
PED	Photoelectron Diffraction
PTFE	Polytetrafluoroethylene
RHEED	Reflection High-energy Electron Diffraction
STM	Scanning Tunneling Microscopy
TDS	Thermal Desorption Mass Spectroscopy
UHV	Ultrahigh Vacuum

RELATED ARTICLES

Surfaces **(Volume 10)**
Surfaces: Introduction • Auger Electron Spectroscopy in Analysis of Surfaces • Differential Reflectance Spectroscopy in Analysis of Surfaces • Ellipsometry in Analysis of Surfaces and Thin Films • Infrared and Raman Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Infrared Spectroscopy **(Volume 12)**
Infrared Reflection–Absorption Spectroscopy

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

REFERENCES

1. A. Zangwill, *Physics at Surfaces*, Cambridge University Press, New York, 1990.
2. H. Ibach, D.L. Mills, *Electron Energy Loss Spectroscopy and Surface Vibrations*, Academic, New York, 1982.

3. H. Froitzheim, *Topics in Current Physics*, ed. H. Ibach, Springer-Verlag, New York, Vol. 4, 1976.
4. J.L. Erskine, 'High-resolution Electron Energy Loss Spectroscopy', *CRC Crit. Rev. Solid State Mater. Sci.*, **13**, 311–380 (1987).
5. W. Ho, 'High-resolution Electron Energy Loss Spectroscopy', in *Physical Methods of Chemistry*, 2nd edition, eds. B.W. Rossiter, R.C. Baetzold, Wiley-Interscience, New York, 209–320, Vol. IXA, 1993.
6. R.E. Palmer, P.J. Rous, 'Resonances in Electron Scattering by Molecules on Surfaces', *Rev. Mod. Phys.*, **64**, 383–440 (1992).
7. H. Ibach, S. Lehwald (eds.), 'Vibrations in Adsorbed Layers: Proceedings of an International Conference', Julich, Germany, June 12–14, 1978.
8. R. Caudano, J.-M. Gilles, A.A. Lucas (eds.), 'Vibrations at Surfaces: Proceedings of the Second International Conference, Namur, Belgium, September 10–12, 1980', Plenum, New York, 1982.
9. C.R. Brundle, H. Morawitz (eds.), 'Vibrations at Surfaces: Proceedings of the Third International Conference, Asilomar, CA, September 1–4, 1982', *J. Electron Spectrosc. Relat. Phenom.*, **29**, **30**, 1983.
10. D.A. King, N.V. Richardson, S. Holloway (eds.), 'Vibrations at Surfaces: Proceedings of the Fourth International Conference, Bowness-on-Windermere, United Kingdom, September 15–19, 1985', *J. Electron Spectrosc. Relat. Phenom.*, **38**, **39**, 1986.
11. A.M. Bradshaw (ed.) 'Vibrations at Surfaces: Proceedings of the Fifth International Conference, Grainau-Eibsee, Federal Republic of Germany, September 6–10, 1987', *J. Electron Spectrosc. Relat. Phenom.*, **44**, **45**, 1987.
12. Y.T. Chabal, F.M. Hoffmann, G.P. Williams (eds.), 'Vibrations at Surfaces: Proceedings of the Sixth International Conference, Shelter Island, N.Y., USA, September 10–14, 1990', *J. Electron Spectrosc. Relat. Phenom.*, **54**, **55**, 1990.
13. M. Rocca, G. Bracco, U. Valbusa (eds.) 'Vibrations at Surfaces: Proceedings of the Seventh International Conference, Genoa, Italy, June 14–17, 1993', *J. Electron Spectrosc. Relat. Phenom.*, **64**, **65**, 1993.
14. R.E. Palmer, J. Schmidt, D.M. Bird, K.W. Kolasinski, C.F. McCanville, R. Raval (eds.) 'Vibrations at Surfaces: Proceedings of the Eighth International Conference, Birmingham, UK, June 23–27, 1996', *Surf. Sci.*, **368**, 1996.
15. G. Herzberg, *Molecular Spectra and Molecular Structure I. Spectra of Diatomic Molecules*, 2nd edition, Van Nostrand Reinhold, New York, 1950.
16. G. Herzberg, *Molecular Spectra and Molecular Structure II. Infrared, Raman Spectra of Polyatomic Molecules*, Van Nostrand Reinhold, New York, 1945.
17. T. Shimanouchi, *Tables of Molecular Vibrational Frequencies, Consolidated*, Volume 1, Nat. Stand. Ref. Data Ser., Nat. Bur. Stand. (US), 1–160, **39**, June 1972.
18. L.J. Bellamy, *The Infrared Spectra of Complex Molecules*, 3rd edition, Chapman & Hall, London, 1975.
19. L.J. Bellamy, *Advances in Infrared Group Frequencies*, Chapman & Hall, London, 1975.
20. L.M. Sverdlov, M.A. Kovner, E.P. Krainov, *Vibrational Spectra of Polyatomic Molecules*, Halsted, Jerusalem, 1974.
21. K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, 3rd edition, John Wiley & Sons, New York, 1978.
22. J.B. Pendry, *Low-energy Electron Diffraction*, Academic Press, New York, 1974.
23. M.A. Van Hove, S.Y. Tong, *Surface Crystallography by LEED*, Springer-Verlag, New York, 1979.
24. F.M. Propst, T.C. Piper, 'Detection of the Vibrational States Adsorbed on Tungsten by Low-energy Electron Scattering', *J. Vac. Sci. Technol.*, **4**, 53–56 (1967).
25. H. Ibach, 'Optical Surface Phonons in Zinc Oxide Detected by Slow-electron Spectroscopy', *Phys. Rev. Lett.*, **24**, 1416–1418 (1970).
26. H. Ibach, 'Surface Vibrations of Silicon Detected by Low-energy Electron Spectroscopy', *Phys. Rev. Lett.*, **27**, 253–256 (1971).
27. S. Andersson, 'Surface Vibrational Excitations of O in the $p(2 \times 2)$O and $c(2 \times 2)$O Structures on Ni(100)', *Solid State Commun.*, **20**, 229–232 (1976).
28. C. Backx, B. Feuerbacher, B. Fitton, R.F. Willis, 'High-resolution Electron Energy Loss Spectroscopy of Vibrational States of Absorbates on (111) Tungsten', *Surf. Sci.*, **63**, 193–202 (1977).
29. J.C. Bertolini, G. Dalmai-Imelik, J. Rousseau, 'Benzene Adsorption on Nickel(100) and (111) Faces Studied by LEED and High-resolution Electron Energy Loss Spectroscopy', *Surf. Sci*, **67**, 478–480 (1977).
30. A.A. Lucas, M. Sunjic, 'Fast-electron Spectroscopy of Surface Excitations', *Phys. Rev. Lett.*, **26**, 229–232 (1971).
31. A.A. Lucas, M. Sunjic, 'Fast-electron Spectroscopy of Collective Excitations in Solids', *Prog. Surf. Sci.*, **2**, 76–137 (1972).
32. E. Evans, D.L. Mills, 'Theory of Inelastic Scattering of Slow Electrons by Long-wavelength Surface Optical Phonons', *Phys. Rev. B*, **5**, 4126–4139 (1972).
33. B.J.N. Persson, 'Theory of Inelastic Scattering of Slow Electrons by Molecules Absorbed on Metal Surfaces', *Solid State Commun.*, **24**, 573–575 (1977).
34. D. Sokcevic, Z. Lenac, R. Brako, M. Sunjic, 'Excitation of Absorbed Molecule Vibrations in Low-energy Electron Scattering', *J. Phys. B*, **28**, 273–281 (1977).
35. F. Delanaye, A. Lucas, G.D. Mahan, 'Inelastic Scattering of Electrons by Vibrational Motion of Molecule Absorbed at Metal Surfaces', *Surf. Sci.*, **70**, 629–642 (1978).
36. W.L. Schaich, 'Long-range Scattering of Electrons by Phonons at Metal Surfaces', *Phys. Rev. B*, **24**, 686–691 (1981).

37. W.L. Schaich, 'Calculation of EELS at a Doped Semiconductor Surface', *Surf. Sci.*, **122**, 175–179 (1982).
38. V. Roundy, D.L. Mills, 'Theory of the Inelastic Scattering of Low-energy Electrons from Crystal Surfaces by One-phonon Processes', *Phys. Rev. B*, **5**, 1347–1363 (1972).
39. G. Aers, J.B. Pendry, T.B. Grimley, K.L. Sebastian, 'Electron Energy Loss Spectroscopy: Calculation of the Impact Scattering from W(100)-$p(1 \times 1)$H', *J. Phys. C*, **14**, 3995–4007 (1981).
40. H. Ibach, *Electron Energy Loss Spectrometers*, Springer, Berlin, 1991.
41. L.L. Kesmodel, 'New High-resolution Electron Spectrometer for Surface Vibrational Analysis', *J. Vac. Sci. Technol. A*, **1**(3), 1456–1460 (1983).
42. R.G. Musket, W. McLean, C.A. Colmenares, D.M. Makowiecki, W.J. Siekhaus, 'Preparation of Atomically Clean Surfaces of Selected Elements: A Review', *Appl. Surf. Sci.*, **10**, 143–206 (1982).
43. J.T. Yates, Jr, *Experimental Innovations in Surface Science*, Springer-Verlag, New York, 1998.
44. M. Persson, J.A. Stroscio, W. Ho, 'Geometric Structure, Pseudoband Gaps, and Surface Vibrational Resonances on Metal Surfaces', *Phys. Scr.*, **36**, 548–558 (1987).
45. B.N.J. Persson, 'Relation Between Dynamical Processes at Surfaces and Electron Energy Loss Measurements', *Phys. Rev. Lett.*, **50**, 1089–1091 (1983).
46. J.A. Stroscio, W. Ho, 'Long-range Quasielastic Scattering of Low-energy Electrons by Conduction-band Surface Plasmons on Si(111)-(7×7)', *Phys. Rev. Lett.*, **54**, 1573–1576 (1985).
47. J.A. Stroscio, W. Ho, 'Quasielastic Electron Scattering as a Probe of the Silicon Surface Space-charge Region', *Phys. Rev. B*, **36**, 9736–9745 (1987).
48. L.H. Dubois, B.R. Zegarski, B.N.J. Persson, 'Electron Energy Loss Study of the Space-charge Region at Semiconductor Surfaces', *Phys. Rev. B*, **35**, 9128–9134 (1987).
49. G.J. Schulz, 'Resonance in Electron Impact on Atoms', *Rev. Mod. Phys.*, **45**, 378–422 (1973).
50. J.W. Davenport, W. Ho, J.R. Schrieffer, 'Theory of Vibrationally Inelastic Electron Scattering from Oriented Molecules', *Phys. Rev. B*, **17**, 3115–3127 (1978).
51. J.E. Demuth, D. Schmeisser, Ph. Avouris, 'Resonance Scattering of Electrons from N_2, CO, O_2 and H_2 Absorbed on a Silver Surface', *Phys. Rev. Lett.*, **47**, 1166–1169 (1981).
52. L.L. Kesmodel, 'Resonance Scattering of Electrons from Benzene Chemisorbed on Pd(100)', *Phys. Rev. Lett.*, **53**, 1001–1004 (1984).
53. R.E. Palmer, P.J. Rous, J.L. Wilkes, R.F. Willis, 'Determination of Absorbate Molecular Orientation from Resonance Electron-scattering Angular Distributions', *Phys. Rev. Lett.*, **60**, 329–332 (1988).
54. E.G. McRae, 'Electronic Surface Resonance of Crystals', *Rev. Mod. Phys.*, **51**, 541–568 (1979).
55. J.M. Baribeau, J.D. Carette, P.J. Jennings, R.O. Smith, 'Low-energy-electron-diffraction Fine Structure in W(001) for Energies from 0 to 35 eV', *Phys. Rev. B*, **32**, 6131–6137 (1985).
56. J.M. Baribeau, J.D. Carette, 'New Low-energy-electron-diffraction Fine Structure on W(110) and Effect of Hydrogen Chemisorption', *Phys. Rev. B*, **25**, 2962–2964 (1982).
57. J.M. Baribeau, J.D. Carette, 'Observation and Angular Behavior of Rydberg Surface Resonance on W(110)', *Phys. Rev. B*, **23**, 6201–6203 (1981).
58. R.O. Jones, P.J. Jennings, O. Jepsen, 'Surface Barrier in Metals: A New Model with Application to W(001)', *Phys. Rev. B*, **29**, 6474–6480 (1984).
59. R.O. Jones, P.J. Jennings, 'Fine-structure Analysis of Spin-polarized Low-energy-electron-diffraction from W(001)', *Phys. Rev. B*, **27**, 4702–4711 (1983).
60. R.C. Jaklevic, L.C. Davis, 'Band Signatures in the Low-energy-electron Reflectance Spectra of FCC Metals', *Phys. Rev. B*, **26**, 5391–5397 (1982).
61. R.C. Jaklevic, 'Epitaxial Growth and Electronic Structure of Au Films of FCC(111) Metals Studied by Electron Reflection Spectroscopy', *Phys. Rev. B*, **30**, 5494–5499 (1984).
62. R.F. Willis, 'Electronic Surface Resonance Enhancement of Vibrational Loss Intensities', Vibrations at Surfaces: Proceedings of Second International Conference, eds. R. Caudano, J.-M. Gilles, A.A. Lucas, Namur, Belgium, September 10–12, 1980, Plenum, New York, 151, 1982.
63. H. Conrad, M.E. Kordesch, R. Scala, W.J. Stenzel, 'Surface Resonances on Pd(111)/H Observed with HREELS', *J. Electron Spectrosc. Relat. Phenom.*, **38**, 289–298 (1986).
64. J.P. Woods, J.L. Erskine, 'Experimental Investigation of β_1-H on W(100) Using High-resolution Electron-energy-loss Spectroscopy: Bond Distances, Scattering Mechanisms, and Impact-scattering Selection Rules', *Phys. Rev. Lett.*, **55**, 2595–2598 (1985).
65. J.P. Woods, J.L. Erskine, 'High-resolution Low-energy-electron Reflection from W(100) Using the Electron-energy-loss Spectrometer: A Step Towards Quantitative Analysis of Surface Vibrational Spectra', *J. Vac. Sci. Technol.*, **4**, 435–438 (1987).
66. B.M. Hall, S.Y. Tong, D.L. Mills, 'Large-angle Electron-energy-loss Spectroscopy with the Inclusion of a Surface Image Potential', *Phys. Rev. Lett.*, **50**, 1277–1280 (1983).
67. M. Sunjic, 'Electron Spectroscopy of Adsorbate Vibrations: The Role of Electronic Surface Resonances', *Phys. Scr.*, **32**, 377–384 (1985).
68. B.M. Hall, S.Y. Tong, D.L. Mills, 'Large-angle Electron-energy-loss Spectroscopy with the Inclusion of a Surface Image Potential', *Phys. Rev. Lett.*, **50**, 1277–1280 (1983).
69. H. Froitzheim, 'Spectrometer Functions: Their Optimization and Output Current Limitations for

Charged-particle Spectrometers', *J. Electron Spectrosc. Relat. Phenom.*, **34**, 11–38 (1984).
70. F. Hadjarab, J.L. Erskine, 'Image Properties of the Hemispherical Analyzer Applied to Multichannel Energy Detection', *J. Electron Spectrosc. Relat. Phenom.*, **36**, 227–243 (1985).
71. C. Oshima, R. Franchy, H. Ibach, 'Numerical Calculations of Electron Trajectories in the 127° Analyzer Using a Position-sensitive Detector under Conditions of Fringing Fields', *Rev. Sci. Instrum.*, **54**, 1042–1046 (1983).
72. L.J. Richter, W. Ho, 'Position-sensitive Detector Performance and Relevance to Time-resolved Electron-energy-loss Spectroscopy', *Rev. Sci. Instrum.*, **57**, 1469–1482 (1986).
73. J.L. Erskine, *Electron Energy Analyzers*, Experimental Methods in the Physical Sciences, Academic Press, San Diego, 209–230, Vol. 29A, 1995.
74. S. Lehwald, J.M. Szeftel, H. Ibach, T.S. Rahman, D.L. Mills, 'Surface Phonon Dispersion of Ni(100) Measured by Inelastic Electron Scattering', *Phys. Rev. Lett.*, **50**, 518–521 (1983).
75. B.A. Gurney, W. Ho, L.J. Richter, J.S. Villarrubia, 'Multidetector Electron-energy-loss Spectrometer for Time-resolved Surface Studies', *Rev. Sci. Instrum.*, **59**, 22–44 (1988).
76. L.J. Richter, W.D. Mieher, L.J. Whitman, W.A. Noonan, W. Ho, 'Improved Multidetector for Time-resolved Electron-energy-loss Spectroscopy', *Rev. Sci. Instrum.*, **60**, 12–16 (1989).
77. P.W. Lorraine, B.D. Thomas, W. Ho, 'A Differentially Pumped Electron-energy-loss Spectrometer with Multichannel Detector for Time-resolved Studies at Intermediate Ambient Pressures', *Rev. Sci. Instrum.*, **63**, 1652–1670 (1992).
78. E. Jeong, J.L. Erskine, 'Multichannel Detection High-resolution Electron-energy-loss Spectrometer', *Rev. Sci. Instrum.*, **60**, 3139–3150 (1989).
79. R. Fuchs, K.L. Kliewer, 'Optical Modes of Vibration in an Ionic Crystal Slab', *Phys. Rev.*, **140**, 2076–2088 (1965).
80. J.A. Stroscio, M. Persson, S.R. Bare, W. Ho, 'Observation of Structure-induced Surface Vibrational Resonances on Metal Surfaces', *Phys. Rev. Lett.*, **54**, 1428–1431 (1985).
81. J.P. Woods, J.L. Erskine, 'Surface Vibrational Resonances and the Order–Disorder Transformation of the W(100) Surface', *J. Vac. Sci. Technol. A*, **4**, 1414–1415 (1986).
82. G. Brusdeylins, R.B. Doak, J.P. Toennies, 'Measurement of the Dispersion Relation for Rayleigh Surface Phonons of LiF(001) by Inelastic Scattering of He Atoms', *Phys. Rev. Lett.*, **46**, 437–439 (1981).
83. F.W. de Wette, *Dynamical Properties of Solids*, eds. G.K. Horton, A.A. Maradudin, Elsevier, New York, 1990.
84. M. Rocca, H. Ibach, S. Lehwald, T.S. Rahman, *Structure and Dynamics of Surface I Topics in Current Physics*, eds. I.W. Schommers, P. von Blauckenhagen, Springer-Verlag, 245–276, Vol. 41, 1986.
85. L.H. Dubois, G.P. Schwartz, R.E. Camley, D.L. Mills, 'Inelastic Scattering of Electrons from Ionic Crystals with a Highly Conducting Overlayer', *Phys. Rev. B*, **29**, 3208–3216 (1984).
86. J.E. Demuth, B.J.N. Persson, F. Holtzberg, C.V. Chandrasekhar, 'Surface and Superconducting Properties of Cleaved High-temperature Superconductors', *Phys. Rev. Lett.*, **64**, 603–606 (1990).
87. Y. Li, J.L. Huang, C.M. Lieber, 'Temperature Dependence of the Energy Gap in $Bi_2Sr_2CaCu_2O_8$ Superconductors by High-resolution Electron-energy-loss Spectroscopy', *Phys. Rev. Lett.*, **68**, 3240–3243 (1992).
88. Y. Li, J. Liu, C.M. Lieber, 'Dependence of the Energy Gap on T_c: Absence of Scaling in the Copper Oxide Superconductors', *Phys. Rev. Lett.*, **70**, 3494–3497 (1993).
89. R.B. Phelps, P. Akavoor, L.L. Kesmodel, A.L. Barr, J.T. Markert, J. Ma, R.J. Kelley, M. Onellion, 'Absence of an Energy Gap in Measurements of Cu–O Superconductors with High-resolution Electron-energy-loss Spectroscopy', *Phys. Rev. B*, **50**, 6526–6529 (1994).
90. D.L. Mills, R.B. Phelps, L.L. Kesmodel, 'Electron-energy-loss Spectrum of the Cuprate Superconductors', *Phys. Rev. B*, **50**, 6394–6401 (1994).
91. P.J. Feibelman, 'Surface Electromagnetic Fields', *Prog. Surf. Sci.*, **12**, 287–407 (1982).
92. J. Harris, A. Griffin, 'Surface Plasmon Dispersion', *Phys. Lett. A*, **34**, 51–52 (1971).
93. A.J. Bennett, 'Influence of the Electron Charge Distribution on Surface Plasmon Dispersion', *Phys. Rev. B*, **1**, 203–207 (1970).
94. Ku-Ding, E.W. Plummer, P.J. Feibelman, 'Surface–Plasmon Dispersion in Simple Metals', *Phys. Rev. Lett.*, **63**, 2256–2259 (1989).
95. S. Suto, K.D. Tsuei, E.W. Plummer, E. Burstein, 'Surface–Plasmon Energy and Dispersion on Ag Single Crystals', *Phys. Rev. Lett.*, **63**, 2590–2593 (1989).
96. M. Rocca, M. Lazzarino, U. Valbusa, 'Comment on Surface–Plasmon Energy and Dispersion on Ag Single Crystals' (and reply), *Phys. Rev. Lett.*, **67**, 3197–3198 (1991).
97. W. Ho, R.F. Willis, E.W. Plummer, 'Observation of Nondipole Electron Impact Vibrational Excitations: H on W(100)', *Phys. Rev. Lett.*, **40**, 1463–1466 (1978).
98. W. Ho, R.F. Willis, E.W. Plummer, 'Mechanisms for Low-energy-electron Vibrational Excitation of Adsorbates: H on W(100)', *Phys. Rev. B*, **21**, 4202–4222 (1980).
99. J.P. Woods, A.D. Kulkarni, J.L. Erskine, F.W. de Wette, 'Vibrational Properties of β_1H and β_1D on W(001): Electron-energy-loss Measurements and Lattice-dynamic Calculations', *Phys. Rev. B*, **36**, 5848–5860 (1987).

100. J.L. Erskine, J.P. Woods, A.D. Kulkarni, F.W. de Wette, 'Surface Vibrations on Clean and Hydrogen Saturated W(100)', *J. Electron Spectrosc. Relat. Phenom.*, **44**, 27–36 (1987).
101. R.L. Strong, J.L. Erskine, *Bull. Am. Phys. Soc.*, **27**, 372, 1982.
102. R.L. Strong, J.L. Erskine, 'Adsorbate Structure Determination Using Surface Vibrational Spectroscopy', *Phys. Rev. Lett.*, **54**, 346–348 (1985).
103. T.S. Rahman, D.L. Mills, J.E. Black, J.M. Szeftel, S. Lehwald, H. Ibach, 'Surface Phonons and the $c(2 \times 2)$ Oxygen Overlayer on Ni(100): Theory and Experiment', *Phys. Rev. B*, **30**, 589–603 (1984).
104. M.L. Colaianni, J.G. Chen, W.H. Weinberg, J.T. Yates Jr, 'Oxygen on Mo(110): Low-temperature Adsorption and High-temperature Oxidation', *Surf. Sci.*, **279**, 211–222 (1992).
105. J.L. Erskine, R.L. Strong, 'High-resolution Electron-energy-loss Spectroscopy Study of the Oxidation of Al(111)', *Phys. Rev. B*, **25**, 5547–5550 (1982).
106. R.L. Strong, B. Firey, F.W. de Wette, J.L. Erskine, 'Surface-site Determination Using Electron-energy-loss Spectroscopy and Lattice-dynamical Models', *Phys. Rev. B*, **26**, 3483–3486 (1982).
107. Y. Li, J.L. Erskine, A.C. Diebold, 'High-resolution Electron-energy-loss Spectroscopy of Hydrogen Chemisorption at Nb(100) Surfaces: Evidence for Subsurface Absorption Sites', *Phys. Rev. B*, **34**, 5951–5954 (1986).
108. B.S. Fang, C.A. Ballentine, J.L. Erskine, 'Bulk-plasmon-enhanced Photoemission from Nb(100) Surface Resonances', *Phys. Rev. B*, **38**, 4299–4302 (1988).
109. B.S. Fang, C.A. Ballentine, J.L. Erskine, 'Hydrogen Adsorption at Nb(100): Photoemission Evidence of Two-state Exchange Involving Subsurface States', *Surf. Sci. Lett.*, **204**, L713–L720 (1988).
110. A.D. Johnson, K.J. Maynard, S.P. Daley, Q.Y. Yang, S.T. Ceyer, 'Hydrogen Embedded in Ni: Production by Incident Atomic Hydrogen and Detection by High-resolution Electron-energy-loss', *Phys. Rev. Lett.*, **67**, 927–930 (1991).
111. P.J. Chen, M.L. Colaianni, J.T. Yates Jr, 'Effect of Complete Oxidation on the Vibrational Properties of Aluminum Oxide Thin Films: An Electron-energy-loss Spectroscopy Study', *Phys. Rev. B*, **41**, 8025–8032 (1990).
112. J.E. Crowell, J.G. Chen, D. Hercules, J.T. Yates Jr, 'The Adsorption and Thermal Decomposition of Water on Clean and Oxygen-predosed Al(111)', *J. Chem. Phys.*, **86**, 5804–5815 (1987).
113. P.J. Chen, M.L. Colaianni, J.T. Yates Jr, 'The Thermal Dissociation of Decaborane on Si(111)-(7×7) and Doping Effects in the Near Surface Region', *J. Appl. Phys.*, **72**, 3155–3160 (1992).
114. P.J. Chen, M.L. Colaianni, J.T. Yates Jr, 'Silicon Backbond Strain Effects on (NH_3) Surface Chemistry: Si(111)-(7×7) Compared to Si(100)-(2×1)', *Surf. Sci. Lett.*, **274**, L605–L608 (1992).
115. P.J. Chen, M.L. Colaianni, R.M. Wallace, J.T. Yates Jr, 'Dissociative Adsorption of PH_3/ on Si(111)-(7×7): A High-resolution Electron-energy-loss Spectroscopy Study', *Surf. Sci.*, **244**, 177–184 (1991).
116. M.L. Colaianni, P.J. Chen, H. Gutleiben, J.T. Yates Jr, 'Vibrational Studies of CH_3I on Si(100)-(2×1): Adsorption and Decomposition of the Methyl Species', *Chem. Phys. Lett.*, **197**, 561–568 (1992).
117. P.W. Lorraine, B.D. Thomas, R.A. Machonkin, W. Ho, 'Translationally and Vibrationally Activated Reaction of CO_2 on Si(111)-(7×7)', *J. Chem. Phys.*, **96**, 3285–3297 (1992).
118. M. Liehr, P.A. Thiry, 'High-resolution Electron-energy-loss Spectroscopic Characterization of Insulators for Si Technology', *J. Electron Spectrosc. Relat. Phenom.*, **54/55**, 1013–1032 (1990).
119. M. Liehr, P.A. Thiry, J.J. Pireaux, R. Caudano, 'Characterization of Insulators by High-resolution Electron-energy-loss Spectroscopy: Application of a Surface Potential Stabilization Technique', *Phys. Rev. B*, **33**, 5682–5697 (1986).
120. D.A. Hensley, L.L. Kesmodel, 'Application of HREELS to Model Catalysts: CO and C_2H_4 Adsorption on Pt/Al_2O_3', *J. Phys. Chem.*, **95**, 1368–1373 (1991).
121. P. Akavoor, W. Menezes, L.L. Kesmodel, G. Apai, W.P. McKenna, 'High-resolution Electron-energy-loss Spectroscopy and Infrared Spectroscopy of Polymer Surfaces: High-resolution and Orientation Effects of Polytetrafluor Polytetrafluoroethylene Films', *J. Vac. Sci. Technol.*, **14**, 95–103 (1996).
122. J.J. Pireaux, 'Electron Induced Vibrational Spectroscopy', in *Surface Characterization of Advanced Polymers*, eds. L. Sabbatini, P.G. Zambonin, to be published.
123. L.L. Kesmodel, 'High-resolution Electron-energy-loss Spectroscopy', *Handbook of Surface Imaging and Visualization*, CRC Press, Boca Raton, FL, 1995.

Electron Microscopy and Scanning Microanalysis

Vladimir Oleshko
Arizona State University, Tempe, USA

Renaat Gijbels and Severin Amelinckx
University of Antwerp (UIA), Antwerp, Belgium

1 Introduction **1**
1.1 Analytical Electron Microscopy: Subject, Advantages and Limitations 1
1.2 Short History and Stages of Development 2
2 Theory and Instrumentation **3**
2.1 Electron–Solid Interactions 3
2.2 Diffraction by Crystals 4
2.3 Inelastic Scattering Processes 12
2.4 Auger Electrons 13
2.5 X-rays 13
2.6 Cathodoluminescence 16
2.7 Overview of Instrumentation 16
3 Analytical Electron Microscopy Methods **20**
3.1 X-ray Emission Spectrometry in the Transmission Electron Microscopy, Scanning Transmission Electron Microscopy and Scanning Electron Microscopy Modes 20
3.2 Cathodoluminescence Spectroscopy in the Scanning Electron Microscopy, Scanning Transmission Electron Microscopy and Transmission Electron Microscopy Modes 21
3.3 Energy-filtering Transmission Electron Microscopy and Electron Energy-loss Spectroscopy 22
4 Digital Image Analysis **22**
5 Sample Preparation Techniques **22**
6 Quantification **23**
6.1 X-ray Microanalysis 23
6.2 Electron Energy-loss Spectroscopy 26
7 Applications **27**
7.1 Automated Population Analysis: Size–Thickness–Elemental Composition 27
7.2 X-ray and Electron Energy-loss Spectroscopic Elemental Mapping 28
7.3 Sample Thickness Determination 29
8 Current Trends **29**
Acknowledgments **30**
Abbreviations and Acronyms **31**
Related Articles **31**
References **31**

In analytical electron microscopy (AEM), several imaging, diffraction and analytical modes are integrated in a design to provide analytical synergism having obvious advantages over any single instrument. The subject of AEM is to determine the morphology, crystallinity, defect structure, phase and elemental compositions and electronic properties of a material using the focused electron beam and signals generated in the course of its interaction with the specimen. The article considers history and stages of development of AEM, theoretical aspects of electron beam–solid interactions, instrumentation and methodology of particular techniques and image analysis, sample preparation and some typical applications. Various electron–specimen interactions generate a great deal of structural and analytical information in the form of emitted electrons and/or photons and internally produced signals, such as elastically and inelastically scattered electrons, Auger electrons (AE), X-rays and cathodoluminescence (CL), which can be analyzed in different operating modes. Imaging of transmitted solid materials is essentially due to elastic scattering (diffraction) of electrons by the periodic arrangement of atoms in crystals (diffraction contrast) and/or interference of several diffracted and transmitted beams (phase contrast). High-resolution imaging in the scanning transmission mode is also possible by using incoherently scattered electrons (Z-contrast). Inelastic interactions are forming the basis for all chemical analytical techniques (energy-dispersive (X-ray) spectroscopy (EDS) and wavelength-dispersive (X-ray) spectroscopy (WDS), electron energy-loss spectroscopy (EELS)/energy-filtering transmission electron microscopy (EFTEM), AE and CL spectroscopy). Basic data characterizing state-of-the-art modern AEM and current trends in its development are presented.

1 INTRODUCTION

1.1 Analytical Electron Microscopy: Subject, Advantages and Limitations

There are two main essentially different classes of electron microscopy (EM) techniques. In the first class, the electron probe is a stationary beam incident along a fixed direction. This incident beam can be either parallel [transmission electron microscopy (TEM), high-resolution transmission electron microscopy (HRTEM), high-voltage transmission electron microscopy (HVTEM),

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

selected-area electron diffraction (SAED)] or convergent [convergent-beam electron diffraction (CBED), convergent-beam electron microscopy (CBEM)]. The resolution is determined by the quality of the imaging optics behind the specimen. Instruments implementing these techniques are conceptually related to classical light microscopes.

In the second class of methods, a fine electron probe is scanned across the specimen and transmitted electrons (TE) [scanning transmission electron microscopy (STEM)] or the desired excited signal such as secondary electrons (SE) or backscattered electrons (BSE) [scanning electron microscopy (SEM)], and/or AE [Auger electron spectroscopy (AES)/scanning Auger microscopy (SAM)] are selected, detected, amplified and used to modulate the intensity of another electron beam which is scanned synchronously with the first one over the screen of a TV monitor. The stationary beam methods are based on image formation processes whereas the scanning methods are essentially ''mapping'' techniques. Their resolution is mainly determined by the probe size, i.e. by the probe formation electron optics. The magnification is geometric, it is determined by the ratio of the areas scanned by the electron beam on the screen and synchronously by the electron probe on the specimen.

AEM is a logical extension of a concept referred to as an integrated modular EM system, which combines several electron imaging and diffraction modes and a number of analytical facilities. In this concept imaging modes [conventional transmission electron microscopy (CTEM), HRTEM, STEM, SEM, EFTEM], diffraction modes (SAED, CBED) and analytical modes [EDS and WDS, EELS and/or CL spectroscopy] are integrated in a design to provide analytical synergism having obvious advantages over any single instrument.

Moreover, the number of signals and contrast effects available in AEM may require the development of an optimum strategy for its application to the structural and analytical characterization of certain materials. This should take into account coordination of the methods being used, image processing, storage of data, separation and preparation techniques involved and mechanisms of the electron beam–specimen interactions and solid-state models of the material under study.[1] The potential of AEM only became exploited with the development of nanochemistry and high nanotechnologies related particularly to cluster-matter such as cluster-produced thin films, nanocrystalline materials and cluster–matrix systems (metal–nonmetal systems, granular superconductors, cluster tunnel junctions, cermets).

The purpose of AEM is to determine both qualitatively and quantitatively the morphology, crystallinity, defect structure, phase and elemental compositions and electronic properties of a material at lateral spatial resolutions down to 0.5 nm within a single instrument using the focused electron beam and signals generated in the course of its interaction with the specimen. In principle, owing to the common origin of the interactions of fast electrons with a solid, all electron beam-based analytical techniques such as electron probe X-ray microanalysis (EPXMA), analytical SEM and AES/SAM may be associated with AEM.[2,3]

Since inelastic scattering of the incident electrons causes radiation damage, AEM methods, and EM in general, are destructive, although the extent of the damage is influenced by the electron dose needed to collect a useful signal, which in its turn depends on the instrumental performance. Moreover, TEM requires very thin specimens, which often implies cutting and thinning.

1.2 Short History and Stages of Development

Microanalysis by electrons and beam-generated X-rays and its basic concepts were first developed in the 1940–60s.[2,4] Instruments for analytical SEM, EPXMA, TEM and STEM equipped with imaging filters, EDS and/or WDS and EELS were successfully utilized for the elemental and compositional characterization of advanced materials in a variety of fields of materials science. In the mid-1960s, when the advantage of using thin foils as specimens for X-ray analyses was recognized, Duncumb[5] developed the first electron microscopy microanalyzer (EMMA), which comprised a 100-keV transmission electron microscope and wavelength-dispersive X-ray spectrometers. However, the efficiency of detection for WDS was low, and a higher beam current was required, which unfortunately increased the probe size to more than 100 nm. Fortunately, an EDS instrument was invented in the early 1970s and was fitted to an Auger electron microscope. It was placed within 1–2 cm of the probe and thus enabled one to measure a sufficiently large fraction of the X-ray flux produced by electron probes even smaller than 10 nm in diameter.

EELS records the energy distribution of beam electrons: TE after having interacted with a thin specimen without reflection or absorption. Hence interaction takes place inside the specimen, and information about its structure could be obtained by the use of an electrostatic or a magnetic electron spectrometer. The first low-loss electron energy-loss (EEL) spectrum of 5.3-keV electrons transmitted through a thin self-supported Al foil was recorded by Rutherman.[6] Hillier and Baker[7] reported inner-shell EEL spectra K-ionization edges of several elements and Fe L- and M-edges using the first electron microanalyzer, which could focus 25–75-keV electrons into a 20-nm probe used as microprobe or shadow microscope. A 180° magnetic spectrometer was used to record

energy losses of TE. A prism–mirror–prism energy filter incorporated into the column of a TEM instrument as proposed by Castaing and Henry[8] was applied not only to acquire EEL spectra but also to display images and diffraction patterns of a specimen at a selected energy loss. An alternative method of energy filtering is the use of a STEM instrument equipped with an electrostatic energy analyzer.[9] During the last three decades various electrostatic and magnetic electron analyzers and energy filters (Mollenstedt analyzer, Wien filter, omega filter, single-prism spectrometers, prism–mirror–prism filter, postcolumn quadrupole and sextupole lens spectrometers) have been developed.[2,10,11]

Modern AEM systems are based on computer-controlled parallel electron energy-loss (PEEL) spectrometers and energy filters equipped with charge-coupled device (CCD) cameras in a combination with solid-state Si/Li and intrinsic Ge (EDS) and crystal (WDS) detectors as well as scanning attachments fitted with multidetector systems for imaging in bright field (BF) and dark field (DF) TEM/STEM and diffraction (SAED, CBED) modes, in SE/BSE/SEM, and in annular dark-field (ADF) STEM modes.

2 THEORY AND INSTRUMENTATION

2.1 Electron–Solid Interactions

AEM is based on the effects of elastic and inelastic scattering of an accelerated electron beam upon interaction with atoms and electrons of the material to be examined.[12] The final signals used for image formation and for analyses are normally not the result of single scattering but of some electron diffusion caused by the gradual loss of the electron energy and by some lateral spreading mainly due to multiple elastic large-angle scattering. Figure 1 describes schematically the most important interaction processes and signals detected in different operating modes of AEM and their information volumes (inset). The complete energy spectrum comprises primary electrons (PE) with energy E_0, and emitted electrons, ions, heat, quanta and internally generated signals such as the following: TE; wide-angle and narrow-angle elastically scattered and/or diffracted electrons (DE); BSE; SE; inelastically scattered electrons with loss of energy; sample current (SC) or absorbed electrons; AE; hard and soft characteristic and continuous X-ray photons; CL; and electron plasmons (EP) and lattice phonons (LP).

Usually only a comparatively small fraction of the characteristic X-rays isotropically emitted from the specimen is detected because of the small solid angle of collection (10^{-3}–10^{-1} sr). Moreover, inner-shell ionizations result in the emission not only of X-rays but also

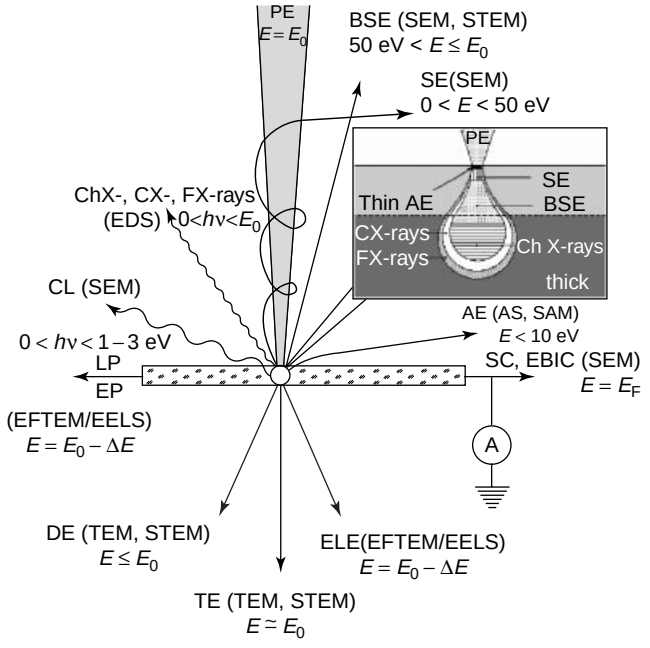

Figure 1 Main processes of electron beam interactions and signals detected in different operating and analytical modes of AEM and schematic arrangement of their information volumes: PE, primary electrons; TE, transmitted electrons; DE, diffracted electrons; ELE, electrons with losses of energy; BSE, backscattered electrons; SE, secondary electrons; the information depth of SE is 5–50 nm; AE, Auger electrons; the information depth of AE is about 1 nm; SC, sample current; ChX-rays, characteristic X-rays; CX-rays, continuous X-rays; FX-rays, secondary X-ray fluorescence; CL, cathodoluminescence; EP, electron plasmons; LP, lattice phonons.

of AE and the X-ray fluorescence yield decreases with decreasing sample atomic number Z, even below 1% for light elements. A large fraction of about 10–70% of the inner-shell ionization processes are associated with inelastically scattered electrons concentrated in small scattering angles which also pass through the objective diaphragm. Atomic electrons can be excited from an inner shell to unoccupied energy states above the Fermi level, resulting in a characteristic edge in the EEL spectrum. Complementary to X-ray spectroscopy, in specimens with thickness of about or smaller than the mean free electron path, the well defined ionization edges for elements with atomic number $2 \leq Z \leq 92$ can be analyzed.[2,13]

The plasmons and inelastic intra- and interband excitations of the outer shell electrons near the Fermi level that normally can be observed with energy losses ΔE smaller than 50 eV are influenced by chemical bonds and the electronic band structure by analogy with optical excitations. In semiconductors the electron impact results in the generation of electron–hole pairs and causes an electron beam-induced current (EBIC). Electron–hole pairs can recombine with emission of

luminescent photons, either in the ultraviolet, visible or infrared (IR) regions or by nonradiative LP. Some fraction of the E_0 that is lost in the course of a cascade of inelastic scattering processes is converted into phonons and/or heat and causes radiolysis, thermal damage, bond rupture and loss of mass and crystallinity by the sputtering of specimen matter.

Ionization of an inner electron shell by the inelastic impact results in a hole, which can be filled by an electron from a higher state. The energy difference can then be emitted either as a characteristic X-ray quantum or as an AE. Hence the X-ray spectrum represents the transition between two energy levels, and the generated intensities are the product of the state density and the transition rate governed by the selection rules.

Since the actual mechanisms of electron–solid interactions are complicated, estimations of the information volume of generated signals based on Monte Carlo simulations (Figure 2a–c) are of importance, in particular for understanding relationships between scattered electrons and produced X-rays.[14] The consequence of the gradual decrease of the electron energy is that the electrons have a finite depth range of the order of several nanometers up to tens of micrometers depending on the E_0 value and the thickness and density of the specimen. The depth and the lateral extent of the information volume governing the resolution of the corresponding operating modes contribute to each of the possible signals, decreasing considerably with a decrease in specimen thickness and an increase in its density. Hence the various electron–specimen interactions can generate a great deal of structural and analytical information in the form of emitted electrons and/or photons and internally produced signals. Inelastic interactions form the basis for all chemical analytical techniques.

2.2 Diffraction by Crystals

2.2.1 Diffraction Effects; General Considerations

The complete characterization of solid compounds, especially of crystalline matter, requires, apart from a knowledge of the chemical composition, also a description of the crystallographic features, i.e. lattice parameters, space group and crystal structures. Moreover, a number of technologically important properties are ''structure'' sensitive, i.e. are influenced by the state of perfection and purity of the crystals and may depend on the microtexture.

Images of solid materials produced in TEM are essentially due to elastic scattering, i.e. to diffraction of electrons by the periodic arrangement of atoms in crystals (diffraction contrast). Not only is diffraction responsible for the formation of the images, but also the diffraction pattern allows the determination of the crystallographic features.[15–18] In the same instrument it is possible easily

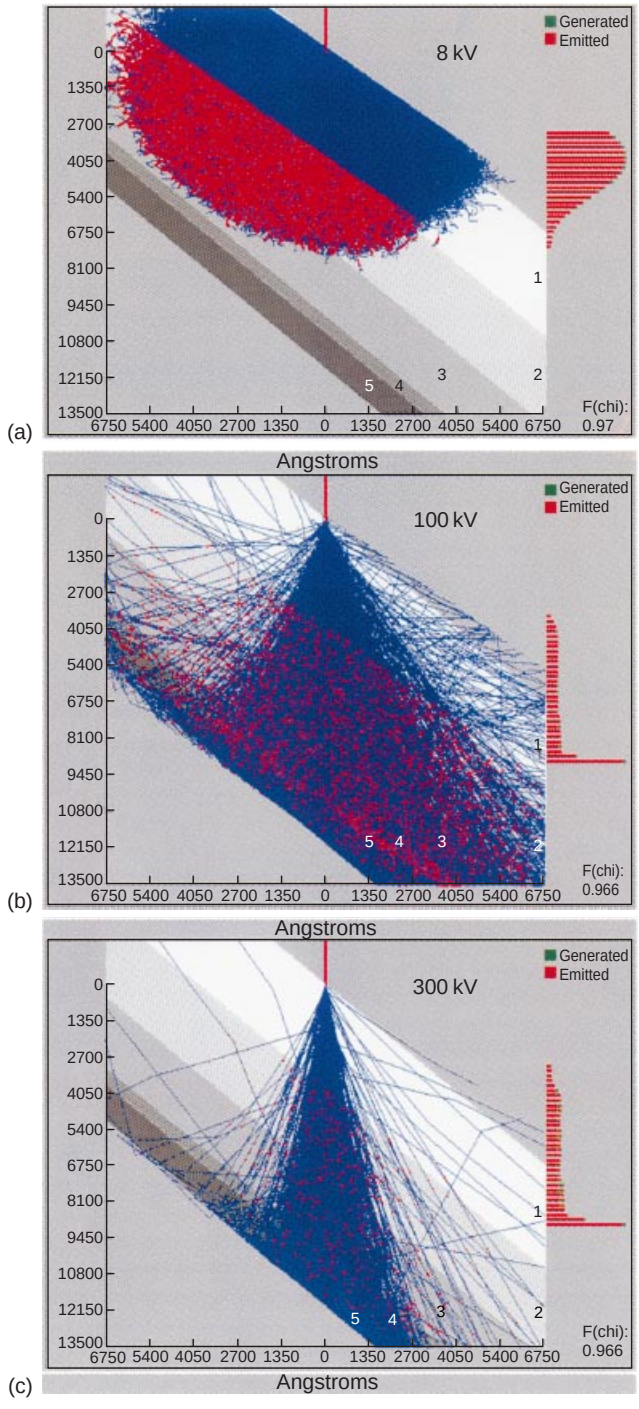

Figure 2 Monte Carlo simulations of a lateral electron distribution (blue) and Si Kα X-ray emission (red) in a multilayer composite film at (a) 8, (b) 100 and (c) 300 keV, 32 000 trajectories, 40° tilt. Histogram of normalized yield of Si Kα emission, $\phi(\rho z)$, is plotted on the right side of the graphs starting at the crystal surface; the length of bars shows the relative value of $\phi(\rho z)$ at that depth. 1, Al, 300 nm; 2, SiO_2, 250 nm; 3, Si_3N_4, 200 nm; 4, WSi_2, 50 nm; 5, SiO_2, 100 nm. Calculations indicate a possibility of analyzing the top layer by reducing the accelerating voltage to $\leq$8 kV and an improvement in lateral resolution and sensitivity at 300 kV compared with 100 kV.

to obtain the electron diffraction (ED) pattern of a small crystal fragment and the image of the same small area without changing the relative orientation of the specimen and the beam. The knowledge of the relative orientation of image and diffraction pattern is essential for a detailed interpretation of the image.

2.2.2 Scattering by Atoms

The elastic scattering of electrons by atoms is governed by Coulomb interaction; it contains two contributions of opposite sign. The first is due to the positively charged nucleus whereas the second is due to the negatively charged electron cloud. It is described by the atomic scattering factor according to Equation (1):[15]

$$f_e(\theta) = \frac{me^2\lambda}{2h^2} \quad \frac{Z - f_x(\theta)}{\sin^2\theta} \tag{1}$$

where Z is the atomic number, $f_x(\theta)$ is the scattering factor for X-rays, m is the electron mass, e is the electron charge, λ is the de Broglie wavelength of the electron and h is Planck's constant. The interaction with atoms is stronger for electrons than for X-rays. As a result, scattering processes in crystals are best described by dynamic diffraction theory, which takes multiple scattering events into account. However, for many applications kinematic diffraction theory is a sufficiently good approximation.

2.2.3 Geometry of Diffraction by Periodic Structures

The direct lattice of a crystal has node points at positions $\mathbf{r_L}$ (Equation 2):

$$\mathbf{r_L} = l_1\mathbf{a}_1 + l_2\mathbf{a}_2 + l_3\mathbf{a}_3 \tag{2}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are the basic lattice vectors and l_1, l_2, l_3 are integers. Reciprocal space, also called diffraction space, consists of reciprocal lattice nodes at positions (Equation 3):

$$\mathbf{b_H} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3 \tag{3}$$

where the $\mathbf{b}_i$ are the basic vectors of the reciprocal lattice defined by the relations $\mathbf{a}_i\mathbf{b}_j = \delta_{ij}$ and $\mathbf{H} = (h_1, h_2, h_3)$ are integers called Miller indices. The interplanar spacing of lattice planes with Miller indices $\mathbf{H}$ is $d_\mathbf{H} = 1/|\mathbf{b_H}|$.

2.2.3.1 Bragg's Law The condition for diffraction to occur can be expressed either in direct space by Bragg's law or in reciprocal space by Ewald's condition. Bragg's law (Figure 3a[19]) (Equation 4):

$$2d_\mathbf{H}\sin\theta_\mathbf{H} = n\lambda \tag{4}$$

(n = integer) states that a strongly diffracted beam will form whenever the incident beam with wavelength λ encloses a Bragg angle θ_H with the set of lattice planes with Miller indices **H**. It expresses the condition that the path difference Δ between beams "reflected" by successive lattice planes with spacing $d_\mathbf{H}$ should be an integral number n of electron wavelengths λ. In ED due to the small wavelength of the electrons, Bragg angles θ_n are only a few degrees at most.

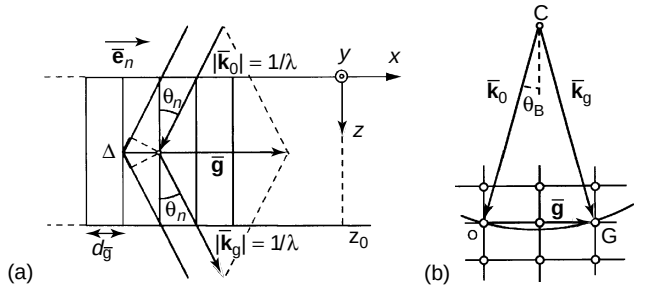

Figure 3 Illustration of the diffraction conditions. (a) Bragg's law. The diffracting planes with interplanar spacing d_g are perpendicular to the foil plane. Bragg's law follows from the requirement that the path difference $\Delta = n\lambda$. The diffraction vector is $\bar{\mathbf{g}}$; the Bragg angle θ_n is of order n. (b) Ewald construction ($\mathbf{k}$ = wavevector of incident beam; $\mathbf{k}_g$ = wavevector of diffracted beam; $\bar{\mathbf{g}}$ = diffraction vector); $\bar{\mathbf{k}}_g = \bar{\mathbf{k}}_0 + \bar{\mathbf{g}}$.

2.2.3.2 Ewald Construction The Ewald condition (Figure 3b) for strong diffraction by the set of planes with indices **H** can be formulated as Equation (5):

$$\mathbf{k}_g = \mathbf{k}_0 + \mathbf{g} \tag{5}$$

where $\mathbf{g} = \mathbf{b_H}$, $\mathbf{k}_g$ is the wavevector of the scattered wave and $\mathbf{k}_0$ that of the incident wave and $\mathbf{g}$ is the diffraction vector. The physical content of the Ewald condition is the same as that of Bragg's law since $|\mathbf{k}_g| = |\mathbf{k}_0| = 1/\lambda$ and $d_\mathbf{H} = 1/|\mathbf{b_H}|$. The Ewald condition expresses the conservation, on scattering, of energy and linear momentum of the incident electrons. It leads to a simple geometric construction of the scattered beam direction. Consider a sphere with radius $1/\lambda$ and with its center C in $-\mathbf{k}_0$ with respect to the origin O of the reciprocal lattice (Figure 3b). Diffraction will occur when the sphere touches a reciprocal lattice node. The wavevector of the scattered wave $\mathbf{k}_g$ is then obtained by joining the center of Ewald's sphere C with the excited reciprocal lattice node.

2.2.3.3 Thin Foil Effects The specimens used in TEM have to be thin (of the order of 100 nm). The actual useful thickness depends on the accelerating voltage of the electrons as well as on the material. For thin specimens, the diffraction conditions are relaxed (Figure 4b). Even if the Ewald condition is not exactly satisfied, diffraction occurs. In terms of reciprocal space, the nodes become "rods" (relrods) having a weight profile

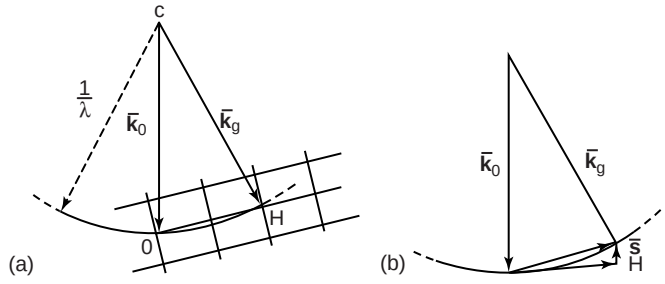

Figure 4 Diffraction condition in thin foil. (a) The Ewald condition is exactly fulfilled for the reciprocal lattice node $\overline{\mathbf{H}}$. (b) The Ewald condition is only approximately satisfied; the node $\overline{\mathbf{H}}$ is missed by the excitation error $\bar{s}$ along the ''relrod''.

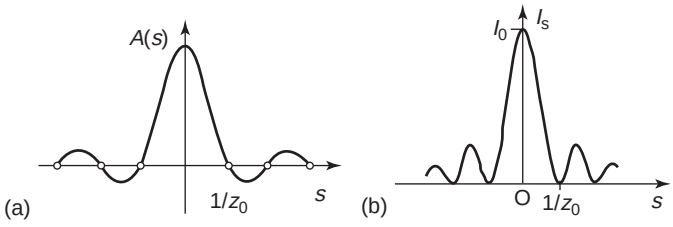

Figure 5 (a) Weight profile of the amplitude along the ''relrod'' as a function of the excitation error $\bar{s}$ (z_0 = foil thickness). (b) Rocking curve according to the kinematic diffraction theory: scattered intensity I_s as a function of the excitation error $\bar{s}$ (z_0 = foil thickness).

(Figure 5a) and they are perpendicular to the specimen plane. If Ewald's sphere intersects these ''relrods'', diffraction occurs, the amplitude being proportional to the local value of the weight function. The intersection point of Ewald's sphere with the rod is defined by its s value (s measures the deviation from the exact Bragg condition). The scattered amplitude (or intensity) as a function of s is called the ''rocking curve'' (Figure 5b).

The smallness of the Bragg angles in TEM and the small foil thickness together with the occurrence of multiple scattering cause the electrons to travel through the foil along narrow columns parallel to the incident beam (Figure 6). The intensity observed in a point along the exit plane of the specimen depends only on the scattering within the column centered on that point. In an undeformed foil of constant thickness and orientation, the intensity does not depend on the column position; it is uniform. In faulted crystals, the intensity does depend on the position of the column. The intensity distribution along the exit surface then acquires some structure: the image, which is in fact a map of the intensity distribution either in a given diffraction spot (DF image) or in the direct beam (BF image) magnified by the lens system (Figure 7).

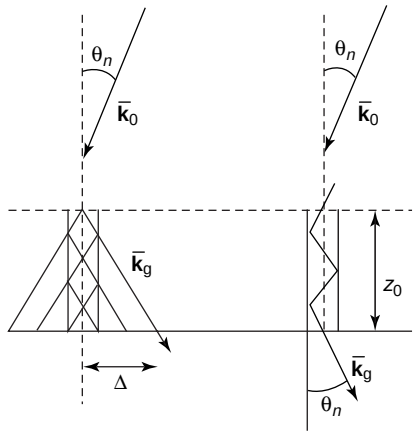

Figure 6 Propagation of electrons along a column. (Left) According to the kinematic theory: only single scattering events occur. (Right) According to the dynamic theory: multiple scattering events occur.

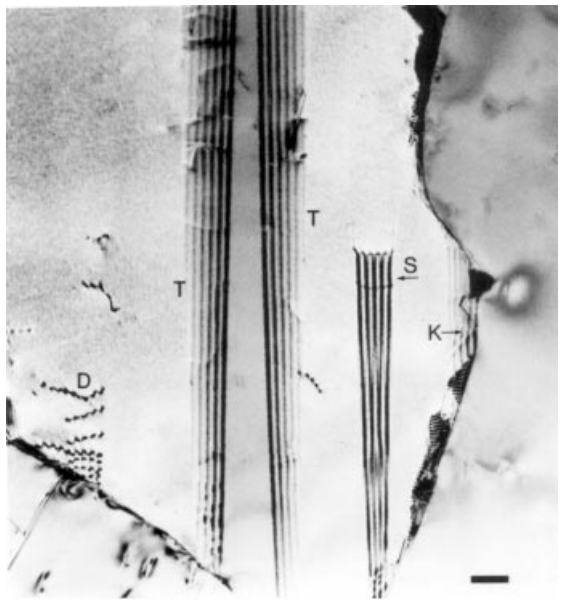

Figure 7 BF diffraction contrast image of a stainless-steel foil containing various types of defects: dislocations (D), twin boundaries (T), stacking faults (S) and grain boundaries (K).

2.2.4 Kinematic Diffraction Theory

According to the kinematic theory,[20] the scattered intensity $I(s, z_0)$ for a foil with a thickness z_0 is given by Equation (6):

$$I(s, z_0) = F_{\mathrm{H}}^2 \frac{\sin^2 \pi s_{\mathrm{H}} z_0}{(\pi s_{\mathrm{H}})^2} \tag{6}$$

where F_{H} is the structure factor of reflection $\mathbf{H}$ for which the corresponding excitation error is $s_{\mathbf{H}}$. This relation describes the ''rocking curve'' shown in Figure 5(b).

The loci corresponding to $s_{\mathbf{H}} = \text{constant}$ or $z_0 = \text{constant}$ values corresponding to maxima of $I(s, z_0)$ are imaged as bright fringes.

The electron wavevector [$k_0 \sim (1/0.004)\,\text{nm}^{-1}$ at 100 kV] is of the order of 50 times larger than the mesh size of the reciprocal lattice ($\sim 1/0.1\,\text{nm}^{-1}$); the Ewald sphere can thus be approximated by a plane perpendicular to $\mathbf{k}_0$ and the geometry of the ED is thus given by a planar section of reciprocal space. Usually the diffraction pattern will thus exhibit a planar lattice of spots (Figure 8a and b).

By tilting the specimen, an orientation can be achieved whereby one diffracted beam is much more intense than all others and comparable in intensity to that of the transmitted beam. Such a ''two-beam situation'' is desirable for a detailed interpretation of defect images since then the image producing the diffraction vector is well defined.

2.2.5 Dynamic Theory

In a two-beam situation, the diffracted and the incident beam play equivalent roles in the interior of the crystal, the diffracted beam being incident under the Bragg angle for the same set of planes as the incident beam. The dynamic interplay of incident and diffracted beams leads to a rocking curve for the scattered beam given by Equation (7):[20]

$$I_s = \frac{1}{(\sigma_{\mathbf{H}} t_{\mathbf{H}})^2} \sin^2 \pi \sigma_{\mathbf{H}} z \tag{7}$$

where $1/\sigma_{\mathbf{H}} = t_{\mathbf{H}}/(1 + s_{\mathbf{H}}^2 t_{\mathbf{H}}^2)^{1/2}$, $t_{\mathbf{H}}$ being the ''extinction distance'', which is inversely related to the structure factor $F_{\mathbf{H}}$; it is a measure for the strength of the reflection $\mathbf{H}$; a long $t_{\mathbf{H}}$ corresponds to a weak reflection; z is the level behind the entrance face of the foil. The rocking curve of the transmitted beam is then $I_T = 1 - I_S$, if absorption is neglected. For $s_{\mathbf{H}} = 0$, i.e. for the exact Bragg condition, I_S and I_T vary periodically in depth (z) with a period $t_{\mathbf{H}}$, I_S and I_T being in antiphase. This periodic character is attenuated by absorption (Figure 9).

2.2.6 Diffraction Contrast Images

Diffraction contrast images made under two-beam conditions are maps of the intensity distribution in a magnified diffraction spot: either the direct spot (BF image) or the diffracted beam (DF image). They reveal deviations from the perfectly periodic structure of the crystalline foil but they cannot provide information on the periodic structure.

Diffraction contrast arises because the intensity emerging from a column depends in general on its position. In a perfect undeformed foil of constant thickness, s is constant along a column and the same for all columns. The thickness and the s value determine the emerging uniform intensity, as given by the rocking curve. If the foil has constant thickness, but is bent, the image will exhibit bright contours corresponding to loci of constant s, so-called ''bent contours''. If, on the other hand, the foil is undeformed but has a wedge shape, the image will consist of loci of constant thickness called ''thickness or wedge

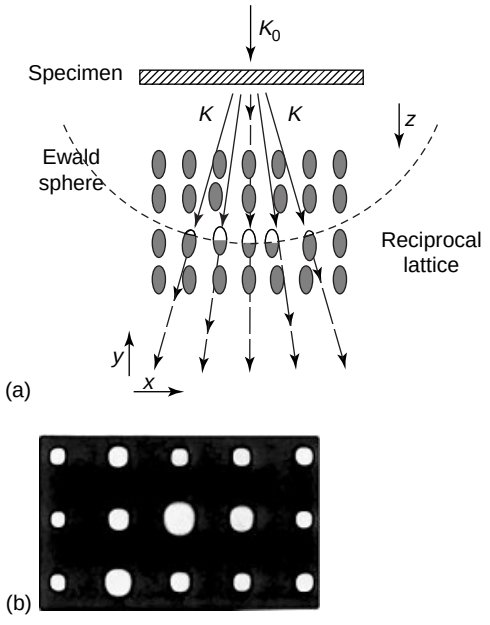

Figure 8 Formation of the ED pattern. Each diffracted beam corresponds to an intersection point of the ''relrods'' in reciprocal space with Ewald's sphere. In fact Ewald's sphere is well approximated by a plane normal to k_0.

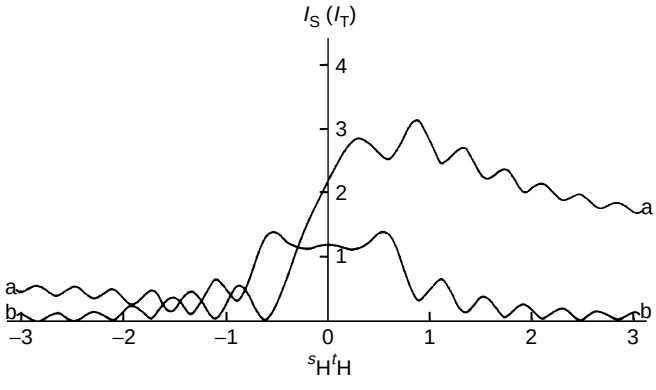

Figure 9 Rocking curves according to the dynamic diffraction theory taking absorption into account: I_T (curve a) and I_S (curve b) as a function of $s_{\mathbf{H}} t_{\mathbf{H}}$ ($s_{\mathbf{H}}$ = excitation error of reflection $\overline{\mathbf{H}}$; $t_{\mathbf{H}}$ = extinction distance of reflection $\overline{\mathbf{H}}$).

fringes''. Often, a foil will exhibit both types of contours simultaneously. They should be distinguished from defect images.

2.2.7 Defects Images and Their Interpretation

A dislocation line produces a line of contrast because in the vicinity of the dislocation the lattice planes change in orientation as a consequence of the associated strain field (Figure 10). On one side of the dislocation E a given set of lattice planes is inclined in one sense, whereas it is inclined in the opposite sense on the other side. On one side (E_1) of the dislocation the Bragg condition may thus be better satisfied than in the perfect part, whereas on the other side (E_2) the Bragg condition is then less well satisfied than in the perfect part (Figure 10). Along the columns on one side of the dislocation more intensity will then be diffracted than in the perfect part; the reverse is true on the other side of the dislocation. Hence, on one side of the dislocation a bright line is produced in a DF image, whereas on the other side a dark line is produced. The image is thus one-sided and the line image does not coincide exactly with the position of the dislocation line. Diffraction by lattice planes, which are not deformed by the presence of the dislocation, does not produce an image. A network of dislocations in graphite in which one set of dislocation segments remains invisible is shown in Figure 11(a) and (b).

A stacking fault parallel to the foil plane changes the intensity emerging from a column intersecting the fault plane. This intensity therefore depends on the level at which the fault occurs in the column. This intensity change results from the phase change of the diffracted electron wave caused by the parallel relative shift of the crystal structures above and below the fault plane. Inclined stacking faults are imaged as fringe patterns in the projected area of the fault plane (Figure 12a and b). The fringes are parallel to the intersection line of the fault plane and the foil surface; they result from the fact that the relative shift of the two crystal parts, separated by the fault plane, occurs at a varying level along the fault plane. The depth period of the fringes is equal to the ''extinction distance''.

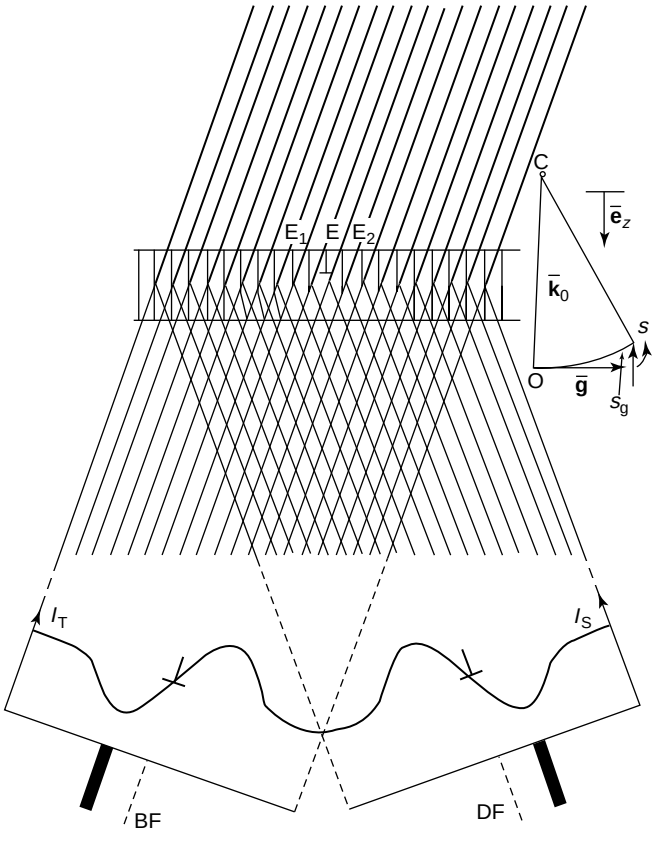

Figure 10 Formation of diffraction contrast image at an edge dislocation E. Left (at E_1) and right (at E_2) of the dislocation the considered lattice planes are inclined in opposite directions: the intensity distributions in the transmitted beam (I_T) and in the scattered beam (I_S). The resulting image profiles are denoted BF and DF, respectively. The two beam diffraction conditions are also represented in reciprocal space.

2.2.8 Lattice Fringes

Instead of providing an image by selecting a single beam, one can select both beams in a two-beam orientation (or more) by means of a large enough aperture. The interference of the beams produces a set of sinusoidal fringes with a spacing $d_{\mathbf{H}} = 1/\mathbf{g}$ equal to the interplanar spacing corresponding to the selected diffraction vector $\mathbf{g}$, the direction of the fringes being perpendicular to $\mathbf{g}$. An intuitive, purely geometric model allows one to understand the formation of the lattice fringes. In the foil F represented in Figure 13(a) and (b), the set of indicated lattice planes with diffraction vector $\mathbf{g}$ has an interplanar spacing of $d_{\mathbf{H}} = 1/|\mathbf{g}|$. The incident beam I satisfies the Bragg condition for diffraction by this set of

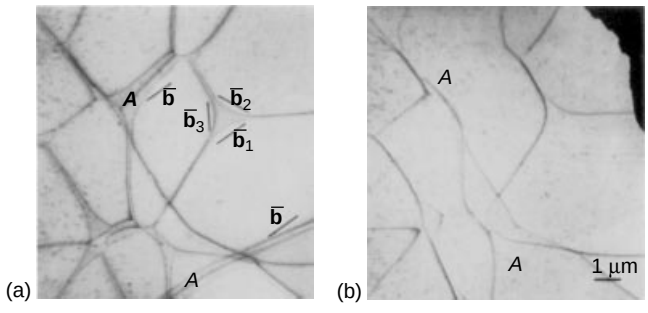

Figure 11 Network of dislocations in graphite as imaged in diffraction contrast. (a) All dislocations exhibit line contrast; (b) one family of dislocations is out of contrast.

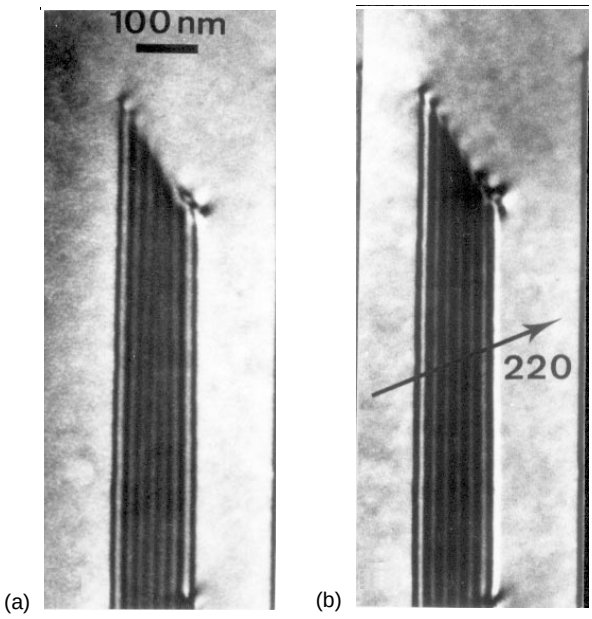

Figure 12 Diffraction contrast images of an intrinsic stacking fault in stainless steel situated in a plane which is inclined with respect to the foil plane. (a) BF image: the intensity profile of the fringe patterns is symmetric. (b) DF image: the intensity profile of the fringe patterns is asymmetric.

planes, i.e. $s = 0$. The transmitted beam is T and the Bragg scattered beam is S; they are assumed to represent plane waves with wavelength λ, which are perpendicular to the corresponding beams; they have maximum elongation along the heavy lines shown in Figure 13(a) and (b). The plane waves interfere behind the foil and intensity maxima will occur along the straight lines where the planes of maximum elongation of T and S intersect. The maxima are thus located along the straight lines seen end-on as black dots in Figure 13(a) and (b). As both waves propagate with the same speed, these lines of maxima move along planes P_1, P_2, P_3 with spacing Δ, parallel to the considered lattice planes. Simple geometric considerations show that $d_{\mathbf{H}} = \Delta$. The planes of maximum intensity form a stationary pattern and can thus be observed behind the exit plane of the foil. This pattern is an "image" of the lattice planes. The localization of the fringes (the image) relative to the lattice planes (the object) depends in reality on the diffraction conditions (in particular on s), but this has no practical significance in this simple case.

If we include next to the beam with diffraction vector **g** also the diffracted beam with diffraction vector 2**g**, the resulting interference pattern will also include parallel fringes with a spacing $1/(2|\mathbf{g}|) = 0.5d$, i.e. features within the basic period d are now imaged. As higher order reflections $n\mathbf{g}$ are included in the interference pattern, the resulting fringe pattern will contain contributions

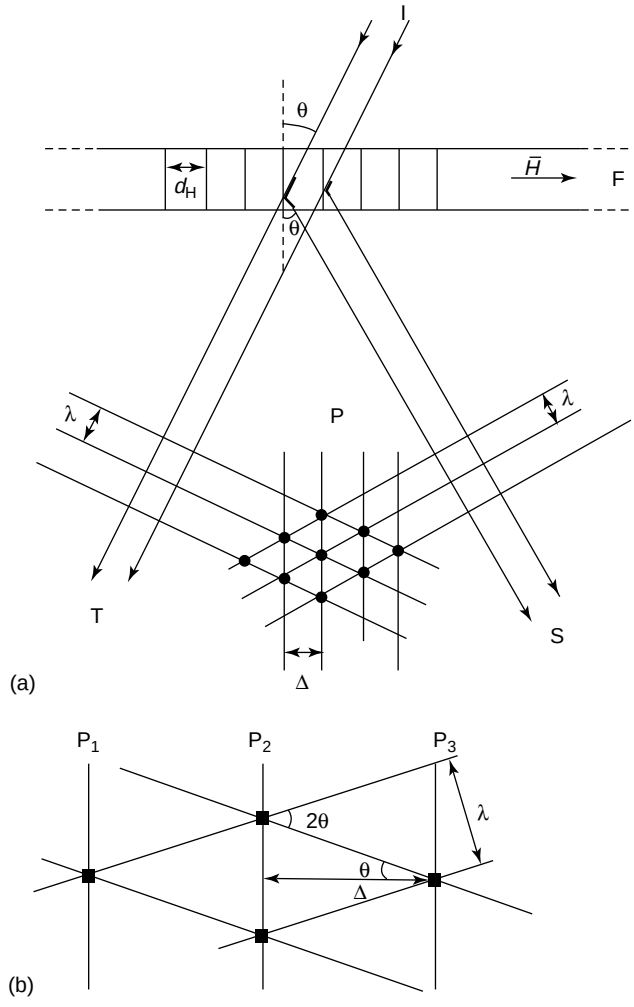

Figure 13 Geometric model illustrating the formation of two-beam lattice fringes associated with diffracting planes with indices $\bar{\mathbf{H}}$. The incident beam satisfies the Bragg condition (θ = Bragg angle) [F = foil; $d_{\mathbf{H}}$ = interplanar spacing; $\Delta(d_{\mathbf{H}})$ = interfringe distance]. Part (b) is a magnified view of part (a).

(i.e. Fourier components) with period d/n consisting of fringes perpendicular to $n\mathbf{g}$. The higher the order of the included reflections, the finer is the detail that can be imaged, provided that the resolution of the microscope is able to resolve these fine details. In practice, the lattice resolution of present-day HRTEM is routinely limited to 0.16 nm and exceptionally to 0.1 nm.

2.2.9 High-resolution Images

2.2.9.1 Fourier Approach If one selects a two-dimensional array of spots centered on the direct beam, each pair of spots determines a diffraction vector $\bar{\mathbf{g}}_j$, which will lead on interference to a set of parallel lattice fringes perpendicular to this vector (Figure 14a

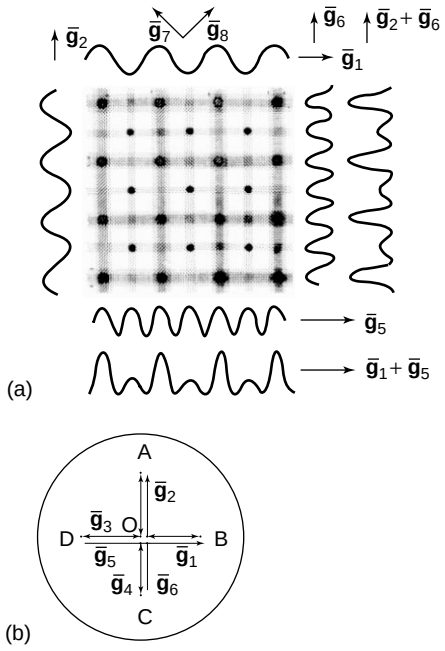

Figure 14 Formation of two-dimensional high-resolution image by the interference between the direct beam O and four diffracted beams (A, B, C, D). Each pair of beams produces a Fourier component with a wavevector given by the line segment connecting the two diffraction spots. Part (a) illustrates the superposition of the different Fourier components with wavevector $\bar{\mathbf{g}}_j$ giving rise to maxima in brightness, where the maxima of the different Fourier waves overlap.

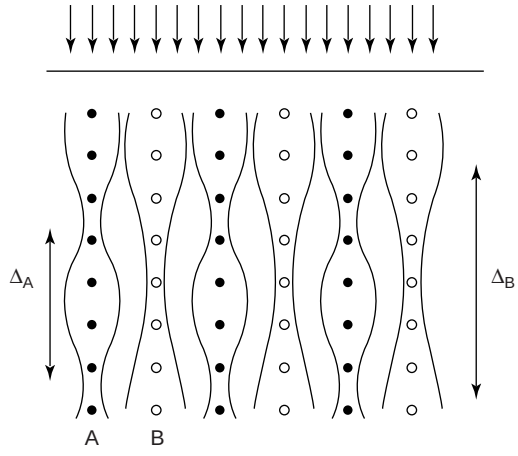

Figure 15 Formation of two-dimensional dot patterns as a result of channeling of the electrons along atom columns parallel to the incident beam. The periods Δ_A (or Δ_B) of the depth variation of the intensity depend on the atom species A (or B) along the column.

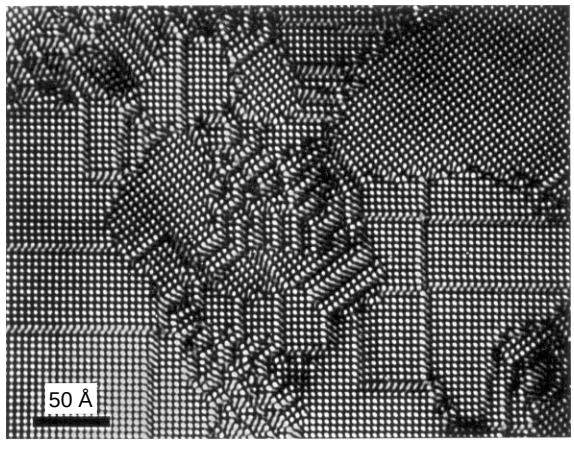

Figure 16 High-resolution image of the ordered alloy Au_4Mn. The bright dots correspond to the minority columns (consisting of manganese). Anti-phase boundaries are observed in two orientation structural variants. (Courtesy of Professor G. Van Tendeloo.)

and b). According to this model, the image will be formed by the superposition of all Fourier components with wavevectors corresponding to the diffraction spots admitted through the selector aperture. Unfortunately, the limited resolution due to the lens aberrations restricts the useful diffraction vectors $\bar{\mathbf{g}}_j$ and hence limits the detail achievable. This is due to the fact that beams enclosing increasing angles with the optical axis become increasingly out of phase with the paraxial beams and moreover become increasingly damped.

2.2.9.2 Channeling Approach An alternative intuitive model emphasizes the fact that one images in reality atom columns parallel to the incident beam rather than single atoms. It considers the propagation (channeling) of electrons in the cylindrically symmetric electrostatic potential well due to the linear strings of atoms forming the columns (Figure 15). On propagating along such a column, the electron wavefunction is centered on the column and oscillates periodically in lateral extent and in intensity with a characteristic depth period Δ_A (or Δ_B) (comparable to the extinction distance in dynamic diffraction theory). This period is the shorter the larger the average Z value of the atoms in the string; it varies in length in the range from 4 to 20 nm. Depending on the foil thickness, columns may thus exhibit either a maximum or a minimum in brightness at the level of the exit surface. At a given thickness, columns having a different composition will in general produce dot images of a different brightness (Figure 16). However, whether a bright or a dark dot will be produced in the final image depends not only on the foil thickness, but also on the imaging conditions, i.e. whether the microscope is either over- or underfocused. Images are mostly taken under somewhat underfocused conditions (which produces the best contrast), optimizing in this way the window in the image transfer function[20,21] (Figure 17a and b).

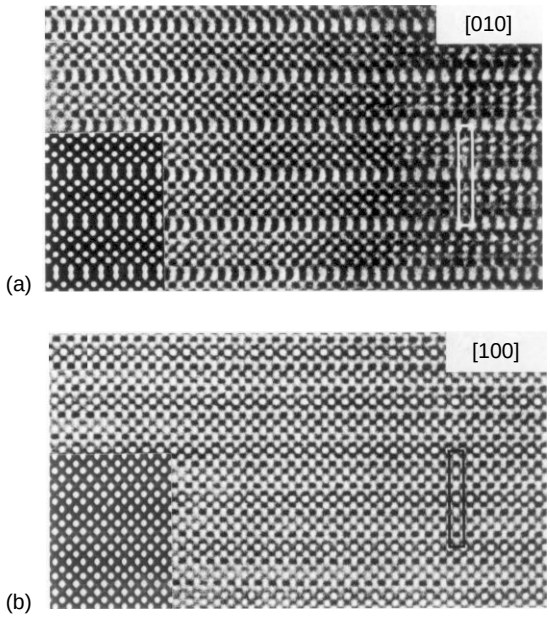

Figure 17 High-resolution image of the high-temperature superconductor $Y_2Ba_4Cu_7O_{15-\delta}$. The bright dots reveal the heavy atom columns as projected along two different zones. Simulated images are reproduced as insets. The projected unit cell is outlined (Courtesy of Professor G. Van Tendeloo). The images were made at optimum defocus. The two images refer to different specimens, with comparable thickness.

2.2.10 Image Interpretation

2.2.10.1 Trial and Error Digital simulation programs for HRTEM are usually based on the Fourier approach.[19,22] They allow one to compute the image of a given structure projected along a given zone axis as a function of the foil thickness and of the imaging conditions (defocus, ...) for a microscope with known characteristics (spherical aberration coefficient, accelerating voltage, beam convergence) (Figure 17a and b).

Identification of a structure or of a defect in a structure, proceeds by ''trial and error''. A model is proposed and a matrix of images, varying the two main independent variables (foil thickness and defocus), is computed (Figure 17a and b). These theoretical images are then compared with the observed images at the different thicknesses along a wedge-shaped part of the foil. A solution is considered acceptable when the correspondence between observed and computed images is judged visually to be ''good enough'' for all available thicknesses at a constant defocus. This comparison allows one at the same time to estimate the foil thickness and the defocus; these quantities are usually not known a priori. Numerical criteria to quantify the ''goodness of fit'' have been proposed[21] but have seldom been applied as yet. In recent developments, high-resolution images have been used to determine the chemical composition along individual columns in rather special circumstances such as along the interfaces in synthetic layer structures of semiconductors grown by molecular beam epitaxy. The composition profile across an interface can be obtained at an atomic scale. The application of these methods requires a certain amount of a priori knowledge concerning the structure.

2.2.10.2 Direct Retrieval Recently, ''direct retrieval'' methods have been developed which use as the input a series of images taken at closely spaced defocus values (focus variation method).[19,20] Knowing the microscope parameters describing the transfer function and its inverse, the projected wavefunction at the exit plane of the foil can be reconstructed. From this, the projected structure (the object) can be retrieved using an analytical formulation of the channeling model. Knowing the projected structure along more than one zone axis allows one to reconstruct the three-dimensional structure. Less a priori knowledge is required than for the methods based on ''trial and error''.

2.2.11 High-resolution Defect Images

From high-resolution images with well-resolved atom columns, the presence of translation interfaces, such as stacking faults and out-of phase boundaries, can be deduced directly from the relative shift of the column configurations in the two crystal parts. Only the component of the displacement vector perpendicular to the incident beam direction is obtained in this way. Since the lattice of the structure is represented directly by the dot configuration, also the orientation domain microstructure and its interfaces, for instance twins, can be imaged directly. The core configuration of dislocation lines parallel to the viewing direction can be imaged directly (Figure 18).

2.2.12 Selected-area Electron Diffraction

ED methods (SAED, CBED) are a useful complement to X-ray diffraction, especially when the material is available only as small microcrystals. Then only powder diffractometry can be applied when using X-rays. However, even fine powders usually contain single crystals of a sufficient size to produce a single-crystal ED pattern, and hence to allow the approximate (but unambiguous) determination of the lattice parameters. These approximate lattice parameters then in turn allow unambiguous indexing of the powder diffraction pattern and subsequently make possible a precise lattice parameter measurement. In addition, ED patterns often exhibit weak reflections due to superstructures, which are not visible in X-ray powder diffraction.

The spatial geometry of diffuse scattering can be reconstructed more easily from ED patterns than from

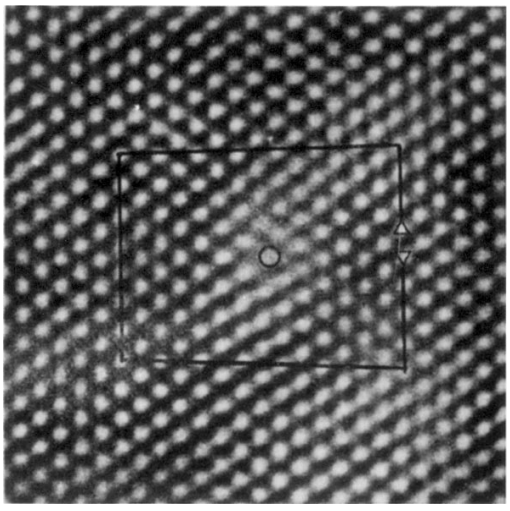

Figure 18 Core configuration of an edge dislocation in germanium as revealed by HRTEM. A Burgers circuit surrounds the dislocation.

X-ray diffraction patterns since ED patterns are planar sections of reciprocal space, which is not the case for X-rays.

2.2.13 Convergent-beam Electron Diffraction

In most commercial microscopes, it is possible to produce a ''stationary'' convergent incident beam rather than a parallel beam, by means of the condenser lens system. The electron beams are then incident along directions either within a cone of revolution or confined to a conical surface (hollow cone method), with a semiapex angle within the range 2–10 mrad. The apex of the cone is mostly situated close to the sample plane. The CBED pattern now consists of circular disks of which the size is proportional to the semiapex angle; they are centered on the positions of the Bragg spots that would be produced if the incident beams were parallel to the cone axis. If the semiapex angle is small enough, these disks do not overlap, but if the semiapex angle exceeds the Bragg angle of the lowest order reflections, these disks overlap and interference occurs between the electron waves in the overlap parts of adjacent disks. For a perfectly flat foil of constant thickness, the intensity distribution within well-separated disks images in a sense the rocking curve, associated with the corresponding Bragg spot. The interference between different, partly overlapping disks gives rise to complicated two-dimensional fringe patterns, which nevertheless still reflect the symmetry along the zone axis and are representative of the point group and of the space group of the crystalline sample, provided that the selected area is free from defects. A major application of CBED is the determination of crystal symmetry and of the strains in crystals, but also defects can be characterized using DF images and local strain patterns can be revealed. From intensity oscillations in the rocking curve obtained from DF images in a known disk, the foil thickness can be deduced with good accuracy.[19]

2.2.14 Z-contrast Images

The commonly used imaging modes (CTEM, HRTEM), being based on interference and diffraction, rely strongly on coherently scattered electrons. However, simultaneously with the coherent scattering, incoherently scattered electrons also are produced by thermal diffuse scattering and in particular by Rutherford scattering. In the STEM mode these incoherently scattered electrons can be used to image atom columns, provided that the proper electron optics are available.[23] A fine electron probe is obtained by focusing the incident convergent electron beam in the sample and scanning over the foil. Thereby, a significant fraction of the incoherently scattered electrons emerging from the sample at relatively large scattering angles is captured in an annular aperture and detected.

In the case of incoherent imaging conditions, the contrast transfer function is a monotonously decreasing function of the spatial frequency, whereas under coherent imaging conditions it is a rapidly oscillating function. This has important consequences. In the coherent case the brightness of a dot imaging a given atom column may change from bright to dark and vice versa as a function of defocus. In contrast, in the incoherent case, the relative brightness of a dot remains consistently the same, independent of defocus, i.e. there is no contrast reversal. Moreover, the dot brightness increases monotonously with increasing average Z value of the atoms in the column as a consequence of the contribution of the Rutherford scattering to the incoherently scattered electrons. Incoherent images can therefore be interpreted on an intuitive basis even for relatively large thickness. It can be shown that structure retrieval is in principle possible and simpler than in the case of coherent imaging.[19] The method is therefore well suited to the study of geometric defects in crystals.

2.3 Inelastic Scattering Processes

Part of the electron kinetic energy is converted in the course of inelastic scattering due to the following atom–electron excitations:[12]

1. Phonon excitations in solids ($0.02\,\text{eV} \leq E \leq 1\,\text{eV}$) normally superimposed with the beam energy spreading.
2. Volume and surface plasmon losses ($0\,\text{eV} < E \leq 50\,\text{eV}$) due to longitudinal oscillations of the plasma of valence and/or conduction electrons. The plasmon

energy loss, E_p, is expressed according to Equation (8):

$$E_p = \hbar w_p = \hbar \left(\frac{ne^2}{(\varepsilon_0 m)} \right)^{\frac{1}{2}} \quad (8)$$

where $w_p = [ne^2/(\varepsilon_0 m)]^{1/2}$ is the plasma resonance frequency, n is the electron density, e is the electron charge, ε_0 is the permittivity of vacuum and m is the electron mass. The low-loss region may also contain losses from the excitations of intra- and interband transitions and of Cerenkov radiation, sometimes overlapped by low-energy ionization edges of outer shells.

3. Quasi-free single-electron excitations (Compton scattering) due to an electron–electron collision, when the transferred energy is larger than the binding energy of excited electrons.
4. Inner-shell ionizations of electrons to a free state or to the continuum beyond the Fermi level resulting in the appearance of an elemental edge at $E = E_I$ followed by an intensity decrease in a long tail for $E > E_I$. The edge profiles (particularly for L- and M-edges) may often exhibit delayed maxima and narrow ''white'' lines. An energy-loss near-edge structure (ELNES), which is dependent on the local bonding and coordination of atoms, can be detected up to 50–60 eV beyond the ionization edge. An extended energy-loss fine structure (EXELFS) covering several hundred electron-volts beyond the edge may be observed owing to interference between the outgoing spherical ejected electron wave and reflected electron waves backscattered from neighboring atoms, thus giving information about interatomic distances.

2.4 Auger Electrons

AE emerge from a solid sample as a consequence of nonradiative rearrangement of the electrons in atoms in which a core hole has been created by the exciting radiation. For an Auger transition involving, for instance, the j, k and l levels, an electron leaves the solid with a certain kinetic energy, the atom being left behind in a doubly ionized state (Equation 9):[24]

$$E_{kin}(jkl;\ X) = E_b(j) - E_b(k) - E_b(l) - F(kl;\ X) + R_c \quad (9)$$

where X is the multiplet state resulting from the coupling of the two holes k and l in the final state, $E_b(j)$, $E_b(k)$ and $E_b(l)$ are corresponding ground-state one-electron binding energies, which take into account initial chemical shifts and one-hole relaxation phenomena, F is the two-electron term and R is a supplementary cross-relaxation energy containing both an atomic and extra-atomic contribution.

Since the primary energy of the exciting electrons in AES is normally in the range 3–10 keV, only K-shell core holes of the light elements can be created, giving rise to the K-series Auger lines. For medium and heavy elements one excites the L- and M-series Auger lines. The observed AE kinetic energies are usually situated in the range 0–2 keV, limiting the information depth. Electrons must escape from the sample surface without suffering inelastic collisions in order to be identified, and therefore the escape depth is limited to a few atomic layers only. Depending on the chemical environment of the ionized atom, shifts in the positions of AE peaks are observed. The shape of an AE line also strongly depends on the chemical environment. Both effects are important for chemical composition imaging.

2.5 X-rays

The X-ray emission spectra, shown in Figure 19(a–d) for a high-temperature superconductor $YBa_2Cu_3O_{7-x}$ ceramic, consist of a background (continuum X-rays, bremsstrahlung) which extends up to the energy of the incident beam, together with superimposed discrete characteristic lines of Cu (Cu K- and L-series), Y (Y K- and L-series), Ba (Ba L-series) and O (O K-series). Wavelength-dispersive X-ray spectra (b–d) demonstrate clearly much better resolution than energy-dispersive X-ray spectrum (a). The relationship between the energy of characteristic X-ray emission lines and the atomic number of the element of interest is described by Moseley's law (Equation 10):

$$E = A(Z - 1)^2 \quad (10)$$

where A is constant within the K-, L- and M-series and Z is the target atomic number. When an inner shell electron is ejected from the atom, the latter becomes ionized and goes to a higher energy state. An electron from one of the outer levels must fill the hole vacancy formed in this way. Electron transitions are regulated by the selection rule in which $\Delta n \neq 1$, $|\Delta l| = 1$, $|\Delta j| = 0$, where n, l and j denote principal quantum number, azimuthal quantum number and inner quantum number, respectively. An X-ray quantum can be emitted with a discrete energy corresponding to the difference in energy between the levels involved. Major X-ray emission lines together with their designation are shown in Figure 20. Actually, transitions which do not satisfy the selection rules, so-called ''forbidden'' transitions, can also occur and do produce some emission lines, but their intensities are usually low. Bonding also affects inner shell electrons owing to the change in the surrounding charge distribution. However, the range of characteristic lines,

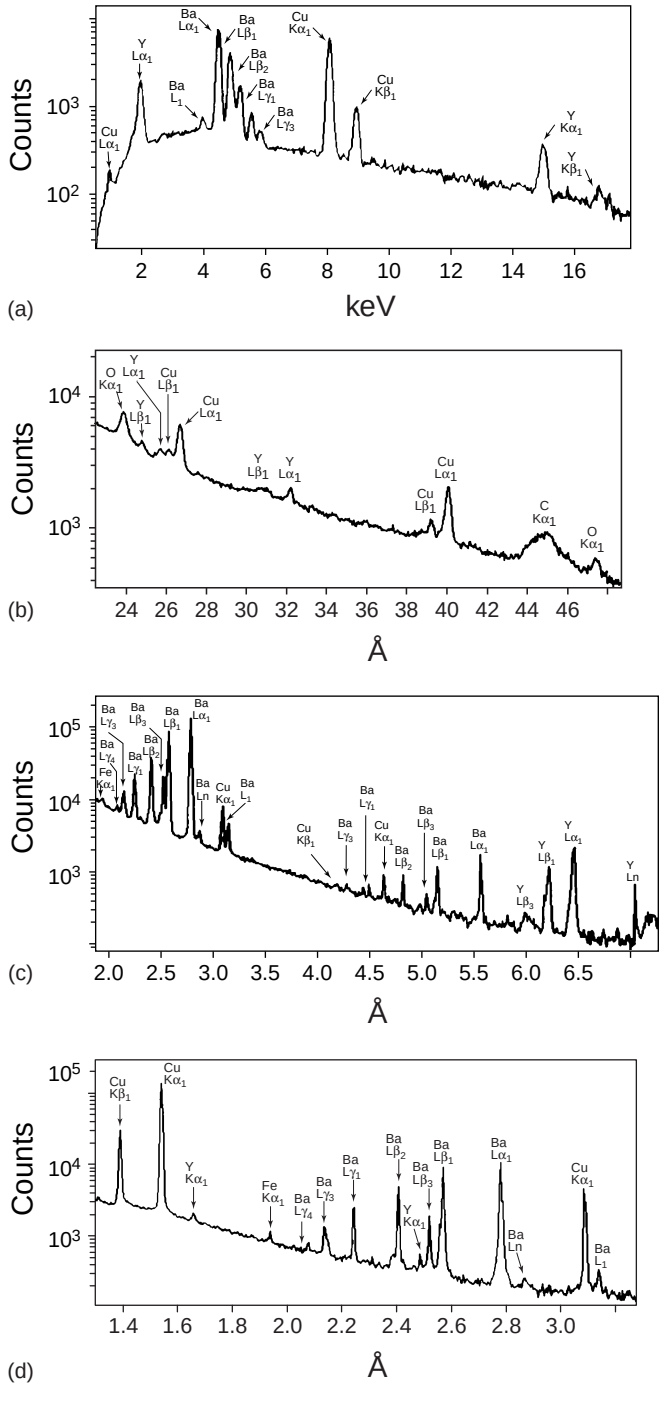

Figure 19 X-ray emission spectra of a high-temperature superconductor $YBa_2Cu_3O_{7-x}$ ceramic recorded at accelerating voltage of 25 kV by EDS [(a) 1 nA probe current] and WDS [(b–d), 200 nA probe current]. (b) Lead stearate crystal (STE); (c) PET crystal; (d) LiF crystal. The peak of C Kα at 4.47 nm (0.277 keV) in spectrum (b) belongs to a carbon conductive coating deposited on the specimen surface. X-ray spectra (c) and (d) exhibit also the presence of traces of Fe [Fe Kα at 0.194 nm (6.398 keV)] and Pt [Pt Lα at 0.131 nm (9.441 keV)], probably from crucible material.

which allow chemical effects to be observed, is fairly limited. This is because the spectrum does not always reflect changes in electron states: there are no changes in X-ray spectra if chemical effects on two different energy levels are the same.

The intensity of the total X-ray emission originating from depth z (Figure 21) below the surface of a specimen with density ρ including any fluorescence contribution may be expressed[4] by Equation (11):

$$I = \phi(\Delta\rho z) \int_0^\infty \phi(\rho z) \exp(-\chi\rho z)\, \mathrm{d}\rho z f(\chi)(1+\gamma+\delta) \tag{11}$$

where $\phi(\Delta\rho z)$ is the emission from an isolated thin film of mass thickness $\Delta\rho z$, $\chi = (\mu/\rho)\,\mathrm{cosec}\,\psi$, ρz is the specimen mass thickness and μ is the linear absorption coefficient; the absorption factor is defined by Equation (12):

$$f(\chi) = \frac{\int_0^\infty \varphi(\rho z) \exp(-\chi\rho z)\, \mathrm{d}\rho z}{\int_0^\infty \varphi(\rho z)\, \mathrm{d}\rho z} \tag{12}$$

The fluorescence correction factor $(1+\gamma+\delta)$ includes the ratio of the fluorescence intensity to the primary characteristic X-ray intensity, γ, and the corresponding ratio for the continuum fluorescence contribution, δ.

Although most of the characteristic X-ray emission can be explained on the basis of transitions allowed by the selection rules, weak lines may appear which occur as satellites of about 10–20% of the intensity of the parent line and may lie within ~20 eV of the principal line.[4] Their production has been explained by assuming that an atom may be doubly ionized by the incident radiation. The two ionizations have to occur virtually simultaneously since the lifetime of an excited state is very short, 10^{-14} s. Usually satellites are relatively more intense for lighter elements because the lifetime of an excited state is longer and the probability of double ionization is higher. Spectral deconvolution accounting for the presence of low- and high-energy satellites and instrumental distortions is required in order to process experimental asymmetric peaks of soft X-ray emission bands. Peak shape changes as a function of the excitation conditions and the matrix composition are related to self-absorption phenomena.

The soft X-ray range can be defined as extending from about 100 eV up to 1.5 keV. For the light elements ($4 \leq Z \leq 9$), in particular, there is no other alternative to using soft X-ray emission. However, soft X-rays may be produced not only owing to electron transitions involving inner orbitals of the atoms but also as a result of transitions associated with outer orbitals containing valence electrons. The overlap of the valence energy states leads to a decrease in X-ray intensity from ionized

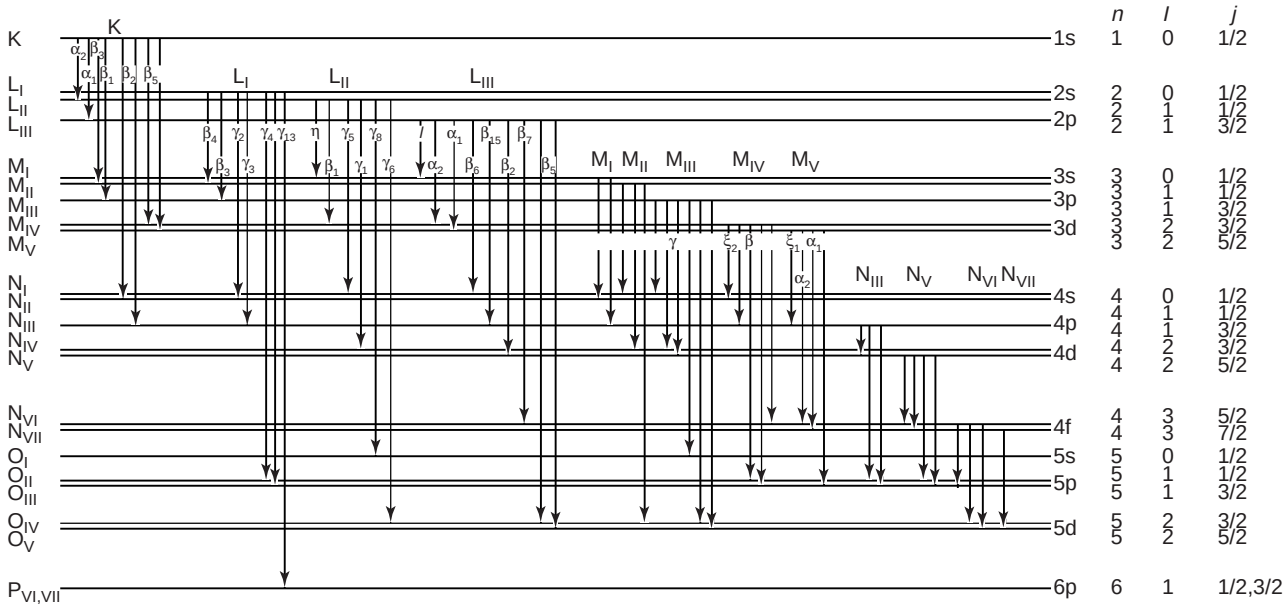

Figure 20 Schematic diagram showing common X-ray emission lines with their designations.

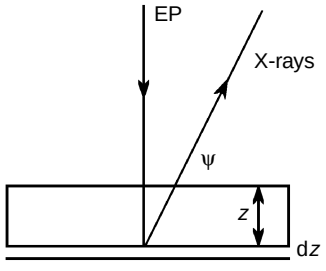

Figure 21 Intensity of the total X-ray emission originating from depth z. The angle ψ is called the take-off angle.

atoms. Moreover, since the inner levels are relatively discrete compared with the valence band transitions associated with outer orbitals, the low-energy X-ray lines sensitively reflect the energy states of valence electrons. Chemical bonding effects in the soft X-ray emission spectra are usually more pronounced in insulators than in conductors because the binding energies of valence electrons increase from metallic, through covalent, to ionic bonding. At the same time, the energy of the soft X-ray emission correspondingly decreases. Line shifts and change of shape due to the chemical bonding may be observed in the Kα series from the light elements ($Z < 10$), where the L-shells involved in K–L transitions are incomplete, and also in the L-series from transition metals and their compounds ($21 < Z < 28$). Analysis of the fine structure of soft X-ray spectra recorded with the appropriate resolution can give not only the elemental composition of an object under study but also important information on electronic structure and bonding.[3] Application of low-energy X-rays to layered specimens enables one to obtain a reduced depth of ionization. It also permits the elimination of most of the secondary X-ray fluorescence emission that occurs when lines of higher energy are employed, particularly for the EPXMA of multiphase specimens. However, quantitative analysis in this case may be complicated by a number of problems such as contamination, coating, background subtraction and line interference.

Bremsstrahlung is produced as a consequence of the slowing of electrons in the Coulomb field of atomic nuclei. The continuous X-rays form a background over an energy range $0 \leq E \leq \mathrm{eU}$ extending up to an energy corresponding to the conversion of the entire energy of an incident electron into a radiation photon in a single interaction. The energy distribution of bremsstrahlung may be expressed by Kramers' equation (Equation 13):[4]

$$N(E)\,\mathrm{d}E = bZ\frac{E_0 - E}{E}\,\mathrm{d}E \tag{13}$$

where $N(E)\,\mathrm{d}E$ is the number of photons within the energy interval E and $E + \mathrm{d}E$, $b = 2 \times 10^{-9}$ photons s^{-1} eV^{-1} $\mathrm{electron}^{-1}$ is Kramers' constant and $E_0 = eU$ is the incident electron energy in electron-volts, where e is the electron charge and U is the electrostatic potential. The angular distribution of the X-ray continuum is anisotropic and forward peaked. The bremsstrahlung intensity also contributes to the background below the characteristic X-ray peaks, thereby decreasing the peak-to-background ratio; the latter can be improved with increasing energy

of the incident electron beam. The bremsstrahlung can be used to calibrate the film thickness.

2.6 Cathodoluminescence

Electron bombardment of semiconductors results in the generation of electron–hole pairs leading to the CL emission of photons in the ultraviolet, visible and IR regions. The photons are emitted as a result of electronic transitions between the conduction and valence bands and levels lying in the band gap, which are ascribed to the crystal structure, chemical composition, doping and impurities.[25] Various defects, dopants and surface and external perturbations (temperature, electric field and stress) of the analyzed specimen drastically affect the CL signal. The CL wavelength strongly correlates with the electronic structures, while the CL intensity is a measure of the crystal perfection: it is attenuated when the band structure is perturbed.

2.7 Overview of Instrumentation

TEM/STEM, analytical SEM and EPXMA instruments and specially designed AE microanalyzers equipped with corresponding analytical facilities in order to detect X-rays, inelastically scattered electrons and/or AE are the most important electron-optical instruments for the analytical and structural investigation of the variety of bulk and thin samples. Table 1 contains basic data characterizing state-of-the-art modern AEM techniques.

2.7.1 Analytical Electron Microscopes

A remarkable capability of STEM is the formation of very small electron probes less than 1 nm in diameter by means of a field-emission gun (FEG) and a three-stage condenser–lens system. This enables the instrument to operate in the STEM mode with a resolution determined by the electron probe diameter and sample thickness. The main advantage of equipping a TEM system with a STEM attachment and especially with an FEG source is the possibility of producing a very small electron probe, with which X-ray analysis can be performed on extremely small areas, while switching to operation in the scanning observation modes allows the probe conveniently to be situated on points of interest for analysis.

X-ray generation in thin foils is confined to the small volume excited by the electron probe only slightly broadened by multiple scattering. Better spatial resolution is therefore obtainable for precipitates, or for segregation effects at crystal interfaces, than in an X-ray microanalyzer with bulk specimens, where the spatial resolution is limited to 100–1000 nm by the diameter of the electron interaction volume.

Figure 22(a) and (b) shows an analytical TEM/STEM system. Normally such an instrument involves an FEG, a probe-forming condenser–lens system, an objective lens and an electron detection system (EDS), often together with a PEEL spectrometer. Electron probes of 0.2–0.5 nm in diameter can be formed (Figure 23a and b), the spherical aberration of the lens being the limiting factor in this case. An advantage of STEM instruments is that the contrast can be enhanced by collecting several signals and displaying differences and/or ratios of these by analog or digital processing. Single atoms of heavy elements on an appropriate thin substrate can be imaged with a wide-angle ADF semiconductor detector (Z-contrast) with a higher contrast than in BF- or DF-CTEM modes.[23,26]

X-ray microanalysis in the STEM mode has a significant advantage over the classical EPXMA in that the lateral resolution can be improved by reducing the illumination area below 1 nm. The fraction of continuous X-ray emission is lower than for bulk samples because of preferential emission in the forward direction. Improvements in X-ray detection have been achieved by increasing the collection angle for a solid state detector up to 0.3 sr and the X-ray count rate up to 29 000 cps by the use of electrostatic blanking, simultaneous X-ray collection from two detectors with equivalent view points and increase in peak-to-background value for the intrinsic germanium detector up to 6300.[27] Determinations of the minimum mass fraction (MMF) for Cr measured on a standard thin film showed that an MMF below 0.1–0.05 wt% is possible. With dedicated STEM instruments, ultrasensitive analysis of a few atoms and at subnanometer lateral resolution has been realized.[3,27] Owing to the complementary nature of the information obtained, the simultaneous Z-contrast high-resolution imaging and EDS/EELS in a dedicated AEM instrument seem to provide a powerful tool to obtain a deeper insight into a fundamental understanding of correlations between the atomic and electronic structure of materials at the atomic scale.

2.7.2 Electron Probe X-ray Microanalyzers and Analytical Scanning Electron Microscopes

The main task of the EPXMA instrument (Figure 24) is to analyze the elemental compositions of flat, polished surfaces at normal electron incidence with a high analytical sensitivity. The ray diagram of such an instrument is similar to that in SEM, but it contains an additional optical microscope in order to perform fine alignments of the full-focusing wavelength-dispersive spectrometers and to select the specimen points, profiles and/or areas to be analyzed. It contains up to five wavelength-dispersive X-ray crystal spectrometers, which can record different characteristic X-ray wavelengths, and

Table 1 Basic characteristics of AEM methods[2–4,11,12,23–50]

Methods	Signals[a]	Types of contrast	Resolution (nm) (magnification)	Information obtained[a]	Information depth (nm)	Element range	Detection limit (at. %) (accuracy, %)
HRTEM	TE/DE	Phase	0.1–0.5 (point) 0.06–0.4 (lattice) ($\times 20$–2×10^7)	Structural (lattice imaging), elemental (binary systems)	1–100	–	0.1 (10)
CTEM	TE/DE	Amplitude (diffraction)	0.3–0.5 (point)	Structural (BF, DF imaging)	–	–	–
SAED, CBED	DE	–		Structural (point and ring nano- and microdiffraction patterns)	–		
STEM	*Z*-contrast			Structural (DF images)			
	TE, DE TE + DE	Diffraction	0.1–2 (point-to-point)	Structural (BF, DF images)			
	DE, TE + DE	Phase	($\times 300$–10^7)	Structural (lattice images)			
	BSE	Atomic number	3–10	Elemental	t(film)	$\Delta Z < 0.1$	2–3 (10)
SAED/STEM	DE	–	1000	Structural (point, ring diffraction patterns)	–	–	
CBED/STEM	DE	–	5	Structural (microdiffraction patterns)	–	–	
SEM	SE	Topography, atomic number, orientation	0.6–7 (edge-to-edge) ($\times 10$–10^6)	Structural (surface topography images)	5–50	–	–
	BSE	Atomic number "compo", topography "topo", orientation	1.5–15 ($\times 10$–10^4)	Elemental (composition images) Structural (surface topography images)	100	$\Delta Z < 0.1$	–
	SC	Atomic number, orientation	3–100	Elemental (composition images)	1000	–	–
				Structural			
WDS	X-ray photons	–	1000 (bulk)	Elemental (spot spectra, profiles, mapping)	1000 (bulk)	4–92	0.01 (2–6)
EDS	X-ray photons	–	1000 (bulk) 0.6–1 (film)	Elemental (spot spectra, profiles, mapping)	1000 (bulk) t(film)	4–92	0.1 (2–6)
EFTEM/EELS	ELE with $0 < E < 2$ keV	–	1–10	Elemental (spot spectra, 2D-spectral profiles, mapping) Chemical (ELNES, EXELFS) Structural (ESI images, diffraction patterns)	t(film)	(1) 3–92	1–5 (10–20)
AES/SAM	SE, AE	–	8–300	Surface Elemental (spot spectra, mapping) Chemical (spectra)	1	3–92	0.1 (10)
CL, SEM	UV/VIS, IR photons, $0 < h\nu < 1$–3 eV		30–500	Chemical (spectra) Structural (spectroscopic images)			

[a] ELE, electrons with loss of energy; ESI, electron spectroscopic imaging; UV/VIS, ultraviolet/visible.

often also an EDS instrument, which can detect X-rays in a wide energy range. An electron probe (from about 6 nm to 1 μm in diameter is governed by the acceptable probe current range from 10^{-12} to 10^{-5} A) is produced by a one-, two- or three-stage demagnification of the smallest cross-section of the electron beam after acceleration. Images are displayed on a cathode-ray tube (CRT) rastered in synchronism. The CRT beam intensity may be modulated by any of the different signals, i.e. SE, BSE, SC, CL or X-rays that result from the electron–specimen interactions.

The combined WDS/EDS microanalyzer is a new-generation EPXMA system that controls its EDS and WDS components with the aid of a powerful computer multitasking workstation and presents X-ray data acquired by both spectrometers and images as a unified analysis result. Combination of WDS and EDS systems can increase the number of simultaneously detectable elements to 13 (five with WDS and eight with EDS). TV display allows the optical microscope image and one of the SEM modes (SE or BSE images) to be observed simultaneously.

Traditionally, electron probe microanalysis (EPMA) or analytical SEM are operated at a high voltage in the range 15–40 kV when used with EDS and WDS.

Figure 22 (a) General view and (b) column schematic diagram of electron optics of an analytical TEM/STEM system. On the left in (a) are the high-resolution TV monitor and the electron optical column with side-entry airlock and windowless EDS detector. On the right are the scanning observation attachment, control panel and X-ray EDS system. (Courtesy of Philips Electron Optics.)

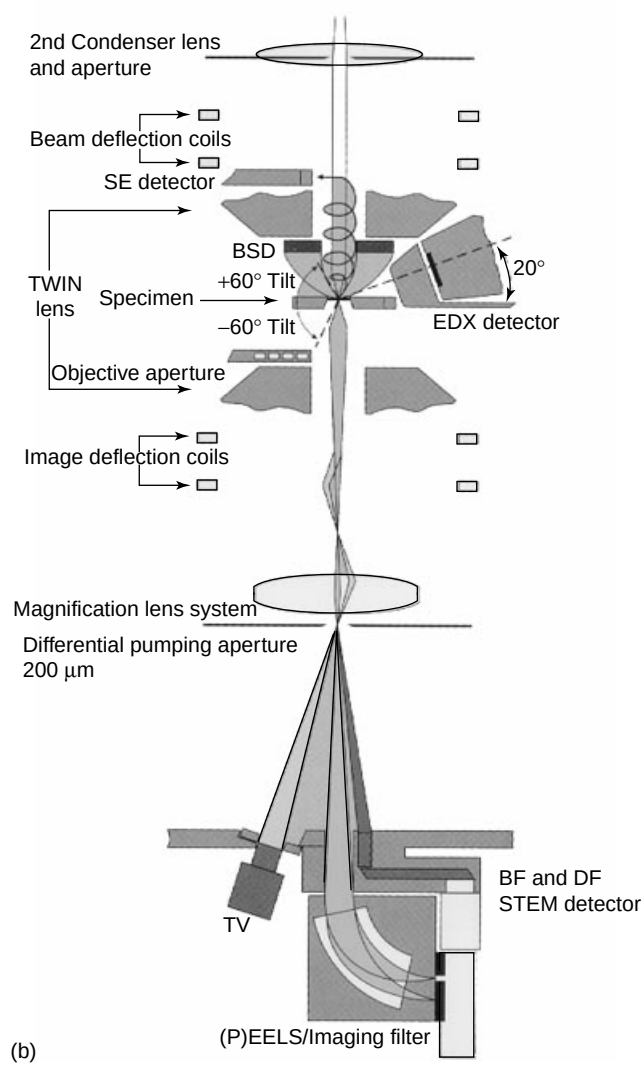

Figure 22 (*Continued*)

These conditions are sufficient to allow excitation of X-ray emission of all relevant elements. However, with the increasing interest in the detection of light elements and the availability of ultrathin window or windowless detectors, the importance of low voltages (<1 kV) is also emphasized. With FEGs that produce sufficient current density even in the low voltage range from 5 to 2 kV, the lateral resolution for X-ray analysis may be significantly improved.[28] As a consequence, thin layers and small particles can be examined by SEM without interference of the bulk support.

Utilizing the scanning beam facilities of the EPMA or SEM instrument, panchromatic CL imaging and CL emission spectroscopy may be easily combined with EDS/WDS and X-ray mapping and electron imaging to perform comprehensive spatially resolved microanalysis of point defects in minerals and ceramics and dopant impurities in semiconductors. Environmental scanning

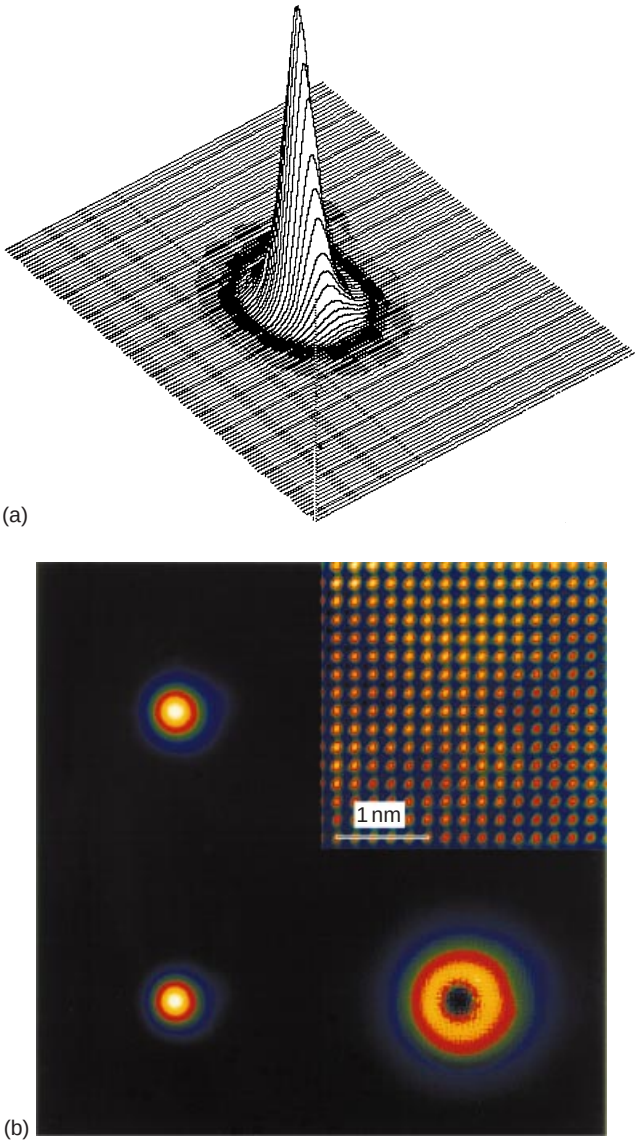

Figure 23 (a) Three-dimensional view of a 0.4 nm (FWHM) electron probe and (b) spot-size measurements at a current of 30 pA using a slow-scan CCD camera. (b) Top left: original image. Top right: lattice image of gold in [100] orientation recorded and displayed at exactly the same magnification as the image of the spot, showing the 0.2-nm spacing for calibration. Bottom left: original image corrected for the response of the CCD camera (beam spreading in scintillator), reducing the apparent spot size by 0.05–0.1 nm. Bottom right: difference between top left and bottom left, showing that the original image is wider (yellow–white ring) and has an extended tail (in blue to black) due to the CCD. (Courtesy of Dr M. Otten, Philips Electron Optics, The Netherlands.)

electron microscopy (ESEM) is the natural extension of conventional SEM that can operate from high vacuum up to a pressure level that can maintain fully wet specimens. This is promising for microanalytical studies of uncoated insulating specimens by EDS and CL

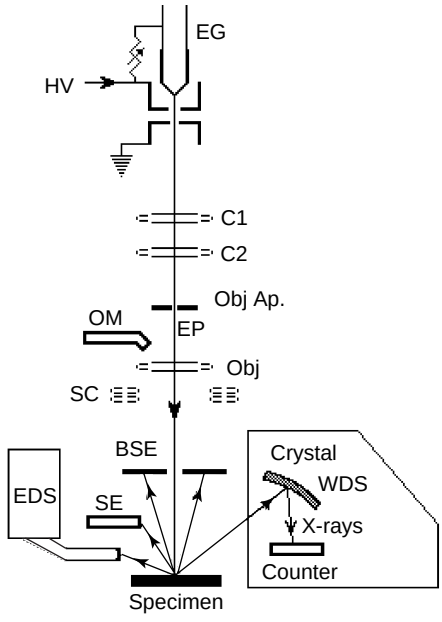

Figure 24 Ray diagram of an electron microprobe/analytical scanning electron microscope. EG = electron gun; HV = high voltage; C1 = first condenser lens; C2 = second condenser lens; Obj Ap. = objective lens aperture; OM = optical microscope; EP = electron probe; Obj = objective lens; SC = scanning coils; BSE = backscattered electron detector; SE = secondary electron detector.

in conjunction with morphological characterization by SE/BSE imaging.[29]

2.7.3 Universal Multipurpose Analytical Electron Microscopy Systems

AEM and EPXMA utilize electron excitation for carrying out chemical analysis. A microanalyzer may also be equipped with a cylindrical mirror AE spectrometer;[30] this, however, needs an ultrahigh vacuum. AE microanalyzers, in which the 1–10-keV electron gun is incorporated in the inner cylinder of the spectrometer, can work in the scanning mode so that an image of the surface is formed with SE or an element-distribution map, especially of light elements, is generated using AE (SAM). Digital image processing of multiple detector signals (SE/BSE/SC/AE/X-ray/CL and/or light image) and computer control with the aid of a powerful multitasking workstation, conversion and storage of data, on-line processing for immediate interpretation of images and spectra and feedback to the instrument increase significantly the capability of modern analyzers.

An advanced 300-kV field emission (FE)/AEM system (Argonne National Laboratory) has been designed to attain the best possible analytical sensitivity, resolution and versatility for EDS, EELS, AES, SAED, CBED, scanning transmission ED and reflected high-energy ED

consistent with state-of-the-art materials research and still provide moderate imaging capabilities in CTEM, HRTEM, STEM and SEM modes. Basically, the system comprises a conventional FEG with gun lens, a triple condenser, objective and quadruple projector. The gun brightness is $\sim 4 \times 10^9$ A cm^{-2} sr^{-1} at 300 kV and the nominal image resolutions in both BF/TEM/STEM modes should be better than 0.3 nm (point-to-point), and better than 0.2 nm in the high-resolution ADF/STEM mode.

Windowless EDS has been optimized to achieve a continuously variable solid angle up to a maximum of 0.3 sr. A hemispherical AES instrument with extraction optics is interfaced to the center of the objective lens and both serial and PEEL detection capabilities will be present. In addition, SE and AE spectrometers are installed within the objective prefield. This universal AEM system may be controlled either directly or using the personal computer and a mouse-directed interactive graphical user interface, thus providing telepresence microscopy remote control and operation over the Internet.[3]

3 ANALYTICAL ELECTRON MICROSCOPY METHODS

3.1 X-ray Emission Spectrometry in the Transmission Electron Microscopy, Scanning Transmission Electron Microscopy and Scanning Electron Microscopy Modes

There are two main spectroscopic systems which can be coupled to SEM, STEM or TEM instruments to record X-ray quanta, i.e. WDS and EDS. Their principal features are compared in Table 2.

WDS utilizes the Bragg reflection of X-ray emission dispersed from the lattice planes of an analyzing curved crystal ($2d \sin\theta = n\lambda$). Crystals with different lattice spacings (0.4–10 nm) (see Table 3) are used to analyze the whole wavelength range from below 0.1 nm (U L$\alpha \approx$

Table 2 Comparison of WDS and EDS[4,27,31,33,37]

Items compared	WDS	EDS	
Basic method	Wavelength dispersion by diffracting crystals	Energy dispersion by solid-state detector	Superconducting transition-edge X-ray microcalorimeter
Available elemental range	^{4}Be–^{92}U	^{4}Be–^{92}U	^{4}Be–^{92}U
Resolution (Mn Kα, FWHM) (eV)	2–20($\sim 0.7 \times 10^{-3}$)	100–150($\sim 6 \times 10^{-3}$)	3–10
Probe current range (A)	10^{-9}–10^{-5}	10^{-12}–10^{-9}	4×10^{-11}–2.5×10^{-7}
Detection limit (ppm)	50–100	1500–2000	–
Number of simultaneously analyzed elements	1	Up to 25	Up to 25
Maximum X-ray acquisition rate	10^2–10^5	10^3–3.10^4	1.5×10^2–2.5×10^3
Solid angle (msr)	8–25	150	0.01–4
Collection efficiency (%)	0.8–2.5	19–115	0.005–2
Typical active area of detector (mm^2)	Not applicable	10–30	
Time to acquire full spectrum (min)	30	1	
Artifacts	High-order lines	Escape peaks, sum peaks, Si K [Ge K (L)] internal fluorescence	–

Table 3 Analysis range of analyzing crystals

Crystal (abbrev.)	2d (nm)	Wavelength range (nm)	^{11}Na ^{20}Ca ^{30}Zn ^{40}Zr ^{50}Sn ^{60}Nd ^{70}Yb ^{80}Hg ^{90}Th
STE	10.04	2.22–9.3	^{5}B–^{8}O, ^{16}S–^{23}V
TAP	2.576	0.569–2.38	^{8}O–^{15}P, ^{23}Cr–^{41}Nb ^{46}Pd ——— ^{79}Au
PET	0.8742	0.193–0.81	^{13}Al–^{25}Mn, ^{36}Kr——— ^{65}Tb, ^{70}Yb——— ^{92}U
LIF	0.4027	0.0889–0.373	^{19}K ——— ^{37}Rb ^{48}Cd ——————— ^{92}U
MYR	8.0	1.77–7.41	^{5}B–^{9}F, ^{17}Cl–^{25}Mn
LDEI[a]			^{6}C–^{9}F
LDE2[a], LDEN			^{5}B–^{8}O
LDEB[a]			^{4}Be–^{5}B

[a] Superlattice crystals.[3,4]

0.091 nm) to above 11 nm (Be K$\alpha \approx 11.3$ nm). The spot irradiated by the electron beam on the specimen acts as an entrance slit, while the analyzing crystal and the exit slit are mounted on a Rowland circle. The lattice planes of the crystal are bent so that their radius is $2R$ and the crystal surface is ground to a radius R. Focusing allows a better separation of narrow characteristic lines and a solid angle of collection of nearly 10^{-3}–10^{-2} sr.

Behind the slit, X-ray photons are recorded by a proportional counter and their energy is converted to a voltage pulse. The detection efficiency of the Bragg reflection and of the proportional counter is about 10–30%.[4,12] The number of electron–ion pairs generated in the counter is proportional to the quantum energy $E = h\nu$. The pulses are further amplified, discriminated in a single-channel analyzer and counted by a scaler. The pulse intensity per second is indicated by a rateameter. Advanced counters with ultrathin Mylar windows are available to detect the weak Kα lines of light elements ($4 \leq Z \leq 11$) more efficiently than with EDS. WDS systems offer better energy (wavelength) resolution and higher count rates (above 50 000 cps) than EDS systems but require a higher beam current (see Figure 19a–d and Table 2). For scanning across a chosen spectral region, the counter and the analyzing crystal should be moved by a pivot mechanism. Usually X-ray microanalyzers are equipped with several WDS instruments, which enable different wavelengths to be recorded simultaneously.

With EDS, a solid-state detector is positioned before the dispersing system to collect the distribution of X-ray emission in a wide energy range [0.1–40 keV for a conventional Si(Li) detector and even up to 80 keV for an intrinsic Ge detector]. Dispersion of the signal detected then takes place by the electronic processing using pulse-height amplification, pileup rejection of possible coincidence of pulses and sorting by a multichannel analyzer, which relates the measured pulse height to the energy of the incoming photon. The entire assembly of a conventional detector, including a field-effect transistor preamplifier, is cryo-cooled under vacuum at near liquid nitrogen temperature to minimize any thermally induced signals. The detection efficiency of an Si(Li) detector is nearly 100% in the range 3–15 keV. The decrease at low energies is caused by the absorption of the X-rays in the thin Be window separating the high-vacuum microscope column from the detector. Both windowless and ultrathin window detectors can record Kα quanta from light elements down to Be. At energies higher than 15 keV the decrease in the efficiency is caused by the increasing probability of penetrating the sensitive layer of the detector without photoionization.

Recent developments in EDS detector fabrication have resulted in significant characteristic improvements in low-energy performance and sensitivities. Modern EDS detectors can also be exploited without permanent cooling; they can be kept at room temperature without any degradation in performance. Newly designed models are electrically cooled with an incorporated Peltier device or need no cooling at all, although still at the cost of some decrease in the resolution. The development of a new microcalorimeter EDS detector with an energy resolution of 3–10 eV allowed detection not only of closely spaced X-ray peaks but also of chemical shifts, although the counting rate is still rather low.[31]

Unlike WDS, where the irradiated point has to be carefully adjusted on the Rowland circle, an EDS instrument does not need any mechanical adjustment, and therefore can be used much more effectively for profile or area analyses by scanning over large and/or rough specimens. Since an EDS system occupies a smaller space than a WDS system, it is commonly used in analytical SEM and TEM/STEM set-ups. A further advantage of EDS is that most of the characteristic X-ray lines can be recorded simultaneously in a short time. For this reason, X-ray microanalyzers that work with WDS are often also equipped with an EDS instrument under the control of the same computer, thus forming an integrated WDS/EDS system in order to provide more rapid and accurate assessments of the specimen elemental composition. This includes: simultaneous data acquisition of up to 13 elements with the combination of five WDS and one EDS instruments, which is more than twice the number measurable with WDS alone; microanalysis of light elements and microvolume elements with a high beam current of 10^{-9}–10^{-5} A (WDS) and of heavy elements of a few percent under the same conditions as for SEM observations at a current of 10^{-12}–10^{-9} A (EDS) for increasing the total analysis efficiency; and preliminary evaluation of beam-sensitive samples by EDS.

3.2 Cathodoluminescence Spectroscopy in the Scanning Electron Microscopy, Scanning Transmission Electron Microscopy and Transmission Electron Microscopy Modes

CL is a powerful technique that provides microcharacterization of the optical and electronic properties of luminescent materials with a detection limit as low as 10^{14} atoms cm^{-3} and a spatial resolution down to 10 nm and even less for adsorbed or thin specimens. Owing to the scattering of the beam electrons and the large escape depth of light, the spatial resolution may not reach that of an SE image. CL is a well-established technique in SEM,[12,25] although the CL signal can, of course, also be collected in a TEM/STEM system equipped with a detector for the light quanta. An advantage of CL is the possibility of simultaneously imaging crystal defects including depth-resolved information by

varying the beam energy and examining their CL nature. Additional information concerning variations in the concentration of dopants acting as luminescence centers and recombination centers is obtainable. The CL intensity is usually low and requires a highly efficient light-collection system with a large solid angle and lateral selection of the irradiated area to avoid the contribution of diffusely scattered electrons.

3.3 Energy-filtering Transmission Electron Microscopy and Electron Energy-loss Spectroscopy

EELS measures the energy distribution of electrons which have been transmitted through a specimen. EFTEM/EELS using an imaging filter or magnetic sector spectrometer offer possibilities to select electrons with a well-defined energy loss (zero-loss, plasmon and elemental imaging on inner-shell ionization edges) and to record spatially and angularly resolved EEL spectra providing local analyses of plasmon losses and inner-shell excitations with ELNES and EXELFS.[2] The transferred energy can be recorded as an energy loss of the incident monochromatic electrons and reduces their kinetic energy to $E_0 - \Delta E$. This results from excitations from occupied states below the Fermi level to unoccupied states above it, allowed by selection rules. Indeed, EELS is spectroscopy of excited electron states beyond the Fermi level. An EEL spectrum (Figure 25a and b) is more complicated than an X-ray emission spectrum and contains information from several excitation processes.[10]

Always being delocalized to some extent, inelastically scattered electrons considerably reduce the image contrast. Furthermore, the chromatic aberration of the objective lens also reduces the resolution for thicker specimens. The great advantage of energy filtering is that it allows the separation of the contribution of inelastically scattered electrons from that of elastically scattered electrons and selection of inelastically scattered electrons with certain energy losses. With column-incorporated and postcolumn imaging filters, a new era of AEM began. The different operating modes of the EFTEM/EELS instrument usually are combined in order to switch easily between ESI and electron spectroscopic diffraction (ESD) to different EELS modes. Because electron images and diffraction patterns as projections of the reciprocal lattice represent two-dimensional intensity distributions, selection of energy losses with an appropriate energy window allows a new dimension in EM (spectrum imaging and image-spectrum techniques[10,11]). Furthermore, an upgrading of the initially developed serial detection system, which comprises a slit, a scintillator and a photomultiplier connected to a multichannel analyzer, by the parallel array of typically 1000 diodes, which measures all the spectral channels in parallel, became a major breakthrough in terms of detection efficiency.

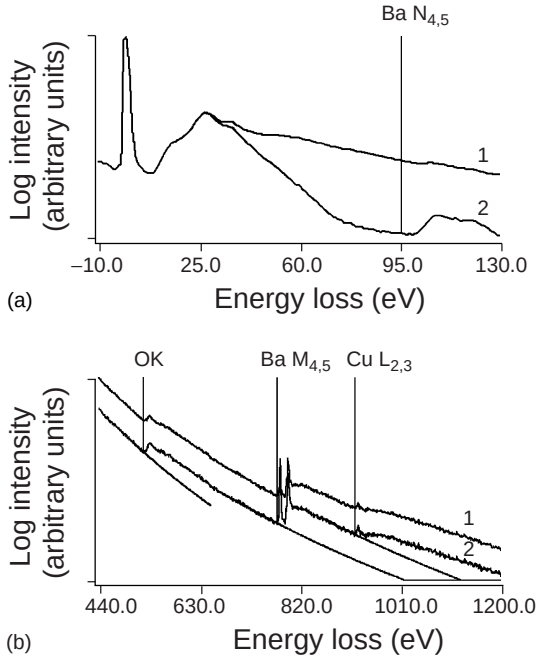

Figure 25 EEL spectra of a specimen of high-temperature superconductor $YBa_2Cu_3O_{7-x}$ ceramic: (a) low losses; (b) inner-shell region. Background assigned by a power-law function is shown. 1 = Experimental spectrum, 2 = single scattering distribution. Removal of multiple scattering improves the resolution of low-loss features and elemental ionization edges.

4 DIGITAL IMAGE ANALYSIS

Image analysis converts the information in the electron image into numerical form. In addition to on-line image acquisition, this is usually also off-line or on-line gray-scale processing, discrimination and binarization followed by quantitative evaluation of features of interest. Processing of data may vary in a wide range from size/shape population distributions, stereological measurements, image reconstruction and specimen restoration to fractal analysis.[1,12,32,33] The next important application of image analysis (see section 2.2.10) is simulation, and matching of conventional and HRTEM lattice images and diffraction patterns, particularly by multislice calculations, taking into account the contrast transfer function, astigmatism and beam tilt correction.[12,33] Some specific applications of image analysis in AEM will be considered in section 7.

5 SAMPLE PREPARATION TECHNIQUES

Sample and standard preparation for AEM includes a variety of techniques commonly used for specimen preparation in conventional EM, although with some

Table 4 Main material preparation techniques for AEM studies[24,33,51]

AEM technique	Samples	Standards	Preparation methods	Support
WDS/EDS/ SEM	Bulk specimens ($\geq$1 µm thick) mounted in beam-stable low-viscosity resin	Known compositions (inner and secondary standards) stable under the beam and homogeneous on the submicrometer scale, may be mounted in ordered arrays in the resin	Mechanical polishing + conducting coating with a 10–20-nm thick carbon layer (if necessary)	Not necessary
WDS/EDS/ STEM/TEM	Thin specimens deposited on thin (lacey, holey) carbon (SiO) films	Known compositions (inner and secondary standards) stable under the beam and homogeneous down to the nm scale with known thickness	(Cryo-)ultramicrotomy, ion thinning, mechanical polishing followed by electrochemical and/or ion thinning + conducting coating with thin carbon layer (if necessary); deposition on thin (holey, lacey) film supports	Self-supported specimens or ones deposited on thin, stable films (should not contain elements of interest)
EFTEM/EELS	1–1000-nm thick specimens deposited on thin (lacey, holey) carbon (SiO) films	Accurately known compositions, stable under the beam and homogeneous down to the nm scale	(Cryo-)ultramicrotomy, ion thinning, mechanical polishing followed by (electro)chemical and/or ion thinning + conducting coating with thin carbon layer (if necessary); deposition on thin (holey, lacey) films	Self-supported specimens or ones deposited on thin, stable films (should not contain elements of interest)
AES/SAM	Bulk specimens, thin films deposited on bulk substrates	Accurately known bulk and surface compositions, stable under the beam and homogeneous on the nm scale	In situ ultrahigh vacuum preparation (scraping, fracture, cleavage); surface cleaning by Ar^+ bombardment, degreasing using detergents followed by rinsing in clean water and ultrasonic bath treatment + careful drying	Not necessary

additional requirements aimed at providing correct qualitative assignment and further quantification of the data (see Table 4).

6 QUANTIFICATION

6.1 X-ray Microanalysis

In order to quantify X-ray spectra, the measured intensity of a particular characteristic X-ray line from the bulk specimen should be compared with the same one produced by a reference bulk standard of known composition. The procedure involving the use of bulk pure-element standards for EXPMA of bulk specimens can, in principle, also be adapted for the microanalysis of thin films.[4,12] Hence, by keeping instrumental settings (probe current, high voltage, detector efficiency, etc.) constant, while the X-ray intensity readings are being taken with dead time corrections related to the spectrometer, one can consider the ratio of the X-ray intensities of the i line of an element a, k_a^i, measured respectively in the specimen, I_{sp}^i, and a standard, I_{st}^i (Equation 14):

$$\frac{I_{sp}^i}{I_{st}^i} = k_a^i \tag{14}$$

where the intensity ratio can be related to the mass concentration, c_a, of the analyte element as $k_a^i = f(c_a)$ and $c_a = 1$.

An accurate quantitative evaluation can be performed if the k_a^i ratio is corrected for various effects: the atomic number correction for the differences between the electron scattering and penetration in the sample and the standard; the absorption correction for the difference in the absorption of the emitted X-rays as they pass through the sample or standard; and the fluorescence correction for the secondary X-ray fluorescence generated by the X-ray emission in the specimen and in the standard.

6.1.1 ZAF Method

The ratio of X-ray intensities emitted from element a in the specimen and in the pure standard is given by

Equation (15):

$$\frac{I^a_{sp}}{I^a_{st}} = \frac{\varphi(\rho z)^a_{sp}}{\varphi(\rho z)^a_{st}} \frac{\int_0^\infty \varphi(\rho z)\,\mathrm{d}\rho z}{\int_0^\infty \varphi(\rho z)\,\mathrm{d}\rho z} \frac{f(\chi)^a_{sp}}{f(\chi)^a_{st}} \frac{(1+\gamma+\beta)^a_{sp}}{(1+\gamma+\beta)^a_{st}} \tag{15}$$

The first ratio in Equation (15), which corresponds to the ratio of the emission of element a from an isolated thin layer of mass thickness $\mathrm{d}\rho z$ in the specimen and in the standard, respectively, is proportional to the number of ionizations produced by the electron beam, i.e. (Equation 16):

$$\frac{\varphi(\rho z)^a_{sp}}{\varphi(\rho z)^a_{st}} = \frac{Qw(Nc_a/A)\,\mathrm{d}\rho z}{Qw(N/A)\,\mathrm{d}\rho z} = c_a \tag{16}$$

where Q is the ionization cross-section of atoms a, w is the X-ray emission yield, N is Avogadro's number, A is the atomic weight of a and c_a is the mass concentration of element a in the specimen. Then one can easily deduce from Equation (15) that the next three ratios may be used to represent the atomic number, absorption and fluorescence correction factors, k_Z, k_A and k_F, respectively. The corresponding corrections are considered in the ZAF method as independent multiplicative factors of the k ratios (Equation 17):

$$k^i_a = k_Z k_A k_F c_a \tag{17}$$

The atomic number correction should be applied to the k ratio to compensate for the difference between the electron retardation and electron backscattering in the specimen and the standard. Because each correction factor is a function of c_a rather than k^i_a, iteration procedures (normalization, hyperbolic or Wegstein methods) should be employed in order to improve further the precision of the data.[4] The k_Z for element a can be written as Equation (18):

$$k^a_Z = \frac{R^a_{sp}}{R^a_{st}} \frac{\int_{E_c}^{E_0} (Q/S)^a_{sp}\,\mathrm{d}E}{\int_{E_c}^{E_0} (Q/S)^a_{st}\,\mathrm{d}E} \tag{18}$$

where the terms R and S refer to the electron backscattering and the electron stopping power, respectively, and Q is the ionization cross-section. The integral limits are from the incident electron beam energy E_0 to the critical excitation band E_c for the chosen X-ray line of a. The electron stopping power $S = -1/\rho\,\mathrm{d}E/\mathrm{d}x$, defined as the energy lost per unit electron path length in material of density ρ, may be approximated according to Equation (19):[4]

$$S = 78500 \frac{Z^a}{A^a} \frac{1}{E} \ln\left(\frac{1.166E}{J^a}\right) \tag{19}$$

where $J = [10.04 + 8.25\exp(-Z/11.22)]\,Z$. The expressions for the ionization cross-section Q have the general form of Equation (20):

$$Q = C\frac{\ln U}{UE^2_c} \tag{20}$$

where C is a constant and $U = E_0/E_c$ is the overvoltage.

The electron backscattering factor $R = (I_t - I_b)/I_t$, defined as the fraction of total generated X-ray intensity excluding loss due to BSE, may be calculated using empirical expressions. In a multielement system, the factor R for element i may be derived from Equation (21):

$$R_i = \sum_j C_j R_{ij} \tag{21}$$

where i represents the element being measured and j denotes the elements in the specimen including i; R_{ij} is the backscatter correction factor for element i in the presence of element j. Several tabulations and fitted polynomial expressions are available to estimate values of R for various elements as a function of Z and U.[36]

The absorption correction factor $k_A = f(\chi)^a_{sp}/f(\chi)^a_{st}$ reflects the attenuation of the X-ray intensity measured by the detector. The basic formulation for the absorption term $f(\chi)$ is given by the Philibert–Duncumb–Heinrich equation (Equation 22):[4]

$$\frac{1}{f(\chi)} = 1 + \frac{\chi}{\sigma}\left(1 + \frac{h}{1+h}\frac{\chi}{\sigma}\right. \tag{22}$$

where $\sigma = 4.5 \times 10^5/(E_0^{1.65} - E_c^{1.65})$ is Lenard's constant and $h = 1.2A/Z^2$.

The variable h is dependent on the composition and must be averaged over the various elements present in multielement specimens according to Equation (23):

$$h_i = \sum_j c_j h_j \tag{23}$$

In addition, the mass absorption coefficient μ/ρ for the characteristic line of element i must be the weighted sum over all elements, i.e. (Equation 24):

$$\left(\frac{\mu}{\rho}\right)^{sp}_i = \sum_j c_j \left(\frac{\mu}{\rho}\right)^j_i \tag{24}$$

The calculation of $f(\chi)$ from Equation (22) is most accurate for values greater than 0.7.

The characteristic fluorescence correction is necessary when the energy E_j of the characteristic X-ray line from an element j is greater than the critical excitation energy $E_{c,i}$ for an element i and when the difference $E_j - E_{c,i} < 0.5\,\mathrm{keV}$. This may result in excitation of the characteristic emission of element i by the corresponding

characteristic line of element j. The basic formulation of the characteristic fluorescence correction is given by Equation (25):

$$F_i = \frac{1 + \sum_j I^{\mathrm{f}}_{i,j}/I_i}{1 + \sum_j I^{\mathrm{f}}_{i,j}/I_i}\Bigg|_{\mathrm{st}}^{\mathrm{sp}} \tag{25}$$

where $I^{\mathrm{f}}_{i,j}/I_i$ is the ratio of the characteristic intensity of element i excited by fluorescence to the electron-excited intensity. The total correction should be summed over all the elements in the specimen. In practice, when the standard is either a pure element or there is no significant fluorescence of element i by other elements in the standard, Equation (25) can be simplified to Equation (26):

$$F_i = \left[1 + \sum_j I^{\mathrm{f}}_{i,j}/I_i\right]_{\mathrm{sp}} \tag{26}$$

In addition, the calculation of the continuum fluorescence is relatively complicated, involving integration over the range from E_0 to E_c for each element in the sample. When $f(\chi) > 0.95$, $c_i < 0.5$ and $\overline{Z}_{\mathrm{sp}} \neq \overline{Z}_{\mathrm{st}}$, the correction can be as large as 2–4%, and it should be included for highest accuracy.

6.1.2 $\phi(\rho z)$ Methods

Quantitative microanalysis can be carried out more suitably by treating the atomic number and absorption factors together rather than as separate entities. The correction by the $\phi(\rho z)$ procedure is based on the integration of X-ray distributions, $\phi(\rho z)$ curves; it is attractive for the analysis of low-energy X-ray lines, where $f(\chi)$ is much less than 0.7 and the accuracy of the ZAF method is low. However, results of the analysis based on the $\phi(\rho z)$ methods are dependent on the accuracy of description of the experimental $\phi(\rho z)$ curves (especially in the case of systems with unknown X-ray depth distributions) and/or fitting to the corresponding results of Monte Carlo simulation.[4]

For instance, the Packwood–Brown model[34] utilizes a modified surface-centered Gaussian function to describe the shape of $\phi(\rho z)$ curves. The Gaussian nature of the curve implies some random distribution of electron scattering events. The surface modification should be applied to take into account that the incoming accelerated electrons require some finite distance range before becoming randomized. The general expression for $\phi(\rho z)$ is given by Equation (27):

$$\phi(\rho z) = \gamma \exp[-\alpha^2(\rho z)^2] \left[1 - \left(\frac{\gamma - \phi(0)}{\gamma}\right) \exp(-\beta \rho z)\right] \tag{27}$$

The Gaussian is modified by a transient function to model an increase of the X-ray production away from the near-surface region. The coefficient α relates to the width of the Gaussian and γ relates to the amplitude. The term β in the transient is related to the slope of the curve in the subsurface region: this means the rate at which the focused electron beam is randomized through scattering in the sample. The intercept $\phi(0)$ is related to the surface ionization potential. The α, β and γ parameters are expressed in terms of several different experimental terms, i.e. elemental composition of the sample, incident electron energy, atomic number, etc.

6.1.3 Standardless Analysis

In standardless analysis, the measured standard intensities are substituted with calculated values based on atomic data and empirical adjustments to experimental data. Better quantitative procedures were developed in the last decade to validate better standardless calculations. In particular, with the K-lines a relative accuracy of 1–5% is reached.[35] Some uncertainties exist with the L-lines and especially with the M-lines where atomic data sets are still incomplete or not accurate enough. Further study of the effect of the incompletely filled inner shells is necessary to obtain better atomic data.

The X-ray intensity generated in a bulk sample (standard) for the simpler case of the K-line[36] may be expressed by Equation (28):

$$I = g\varepsilon p w f(\chi) \frac{N}{A} R \int_{E_0}^{E_c} \frac{Q}{\mathrm{d}E/\mathrm{d}\rho s}\,\mathrm{d}E \tag{28}$$

where g is a normalization factor that depends on the experimental parameters (beam current, acquisition time, solid angle of the detector, etc.), ε is the detector efficiency, p is the transition probability, w is the fluorescence yield, $f(\chi)$ is the absorption correction, R is the backscatter factor and Q is the ionization cross-section given in general form by Equation (20). The additional intensity induced by the secondary X-ray fluorescence is not considered here. When calculating the number of ionizations for the L- and M-shells, both direct ionizations induced by the bombarding electrons and indirect ionizations caused by the nonradiative Coster–Kronig transitions between subshells prior to the emission of the X-ray line must also be taken into account. The Si(Li) detector efficiency may be given by Equation (29):

$$\varepsilon = \left\{1 - \exp\ -\left(\frac{\mu}{\rho}\right)_{\text{Si}} \rho d_{\text{Si,act}}\right\} \exp\ -\left(\frac{\mu}{\rho}\right)_{\text{Be}} \rho d_{\text{Be}} \times \exp\ -\left(\frac{\mu}{\rho}\right)_{\text{Au}} \rho d_{\text{Au}}\ \exp\ -\left(\frac{\mu}{\rho}\right)_{\text{Si}} \rho d_{\text{Si,dead}} \tag{29}$$

where Be, Au, Si,act and Si,dead refer to the beryllium window, the gold contact and the active and the dead silicon layer of the detector, respectively.

The k ratios obtained by comparing experimental spectra with calculated standard intensities may be used as concentrations of first approximation following the iteration correction until convergence is achieved.[4,35]

6.1.4 Analysis of Thin Films and Particles

Thin films and particles are conveniently analyzed in AEM by high-energy electron beams with accelerating voltages not less than 80 keV. Under these conditions, the effect of electron backscattering is minimal and electron energy losses in the specimen are rather small. For specimens in which the thin-film criterion is fulfilled, i.e. the X-ray absorption is <3% and/or fluorescence <5%, the atomic number correction can be neglected and the characteristic X-ray intensity of element i produced may be expressed according to Equation (30):[37]

$$I_i = K\varepsilon_i c_i w_i Q_i a_i \frac{t}{A_i} \tag{30}$$

where K is a constant, ε_i is the efficiency of the EDS detector to detect X-ray emission from element i, c_i is the mass concentration of element i, w_i is the fluorescence yield for the analyzed X-ray peak of element i, a_i is the measured fraction of total series intensity, t is the film thickness and A_i is the atomic weight of element i. To avoid the influence of the substrate, the particles should be deposited on thin-film supports of low atomic number such as carbon. Further, the particles should be scanned in a raster because in general the uniformity of the beam density cannot be assured when the probe diameter is comparable to or larger than the particle size.

In practice, it is suitable to measure the relative concentration of element i to another element j which may be given by Equation (31):[38]

$$\frac{c_i}{c_j} = k_{ij}\frac{I_i}{I_j} \tag{31}$$

where c_i and c_j are the mass concentrations of elements i and j, respectively, I_i and I_j are their corresponding X-ray line intensities and k_{ij} is the Cliff–Lorimer sensitivity factor. The sensitivity factors can be related to an inner standard element by Equation (32):

$$k_{ij} = \frac{k_{i,\text{is}}}{k_{j,\text{is}}} = \frac{k_i}{k_j} \tag{32}$$

where $k_{i(j),\text{is}} = k_{i(j)}$ are the relative sensitivity factors. The index is usually referred to silicon and iron commonly used as internal standard (is) elements. The tabulated experimental values of the sensitivity factors for K-, L- and M-line emission from various elements are available.[38–41] It is also possible to calculate the sensitivity factors using a set of parameterized equations.[37,42]

Equation (31) may be expanded to take into account the effects of absorption and fluorescence as follows (Equation 33):

$$\frac{c_i}{c_j} = k_{ij}\frac{I_i}{I_j}k_{\text{a}}\frac{1}{1 + I^i_{\text{f}}/I^i_0} \tag{33}$$

where I^i_{f} and I^i_0 are the fluorescent and primary X-ray intensities of i, respectively, and k_{a} is the absorption factor given by Equation (34):[36]

$$k_{\text{a}} = \frac{(\mu/\rho)^i_{\text{sp}}}{(\mu/\rho)^j_{\text{sp}}}\frac{1 - \exp[-\chi^i(\rho t)]}{1 - \exp[-\chi^j(\rho t)]} \tag{34}$$

The parameterless extrapolation method proposed by Van Cappellen[43] is based on measurements of the relative intensities I_i/I_j at various thicknesses following extrapolation to zero thickness ($I_i = 0$) by plotting I_i/I_j versus I_i. Then, if the secondary emission of the elements is much smaller than the primary emission, the mass concentration ratio can be determined, such as (Equation 35):

$$\frac{c_i}{c_j} = k_{ij}\lim_{t\to 0}\frac{I_i}{I_j} \tag{35}$$

Other methods based particularly on convergence between two or more different measurements for simultaneous determination of film composition and thickness and Monte Carlo techniques are discussed in the literature.[4,14,36,37]

6.2 Electron Energy-loss Spectroscopy

Multiple scattering can alter the EEL spectrum (Figure 25a and b, curve 1), taking the form of one or more low-loss excitations if the specimen thickness is of the order of or more than the estimated inelastic mean free path. Multiple scattering effects can be removed from the spectra using a number of procedures, although the most commonly employed procedures are based on fast Fourier transforms and are known as Fourier-log and Fourier-ratio techniques.[44]

For Fourier-log deconvolution, the recorded spectral intensity may be written in the form of Equation (36):[45]

$$J(E) = \sum_n C_n[R(E) \otimes S_n(E)] \tag{36}$$

where $R(E)$ is a unit-area resolution function, $\otimes$ represents a convolution with respect of energy loss, $S_n(E)$ is the n-fold self-convolution of the single scattering distribution (SSD), $S(E)$, and the Poisson coefficient $C_n = I_0^{n-1}/n!$. The Fourier transform of Equation (36) gives a simple product, and taking into account that the zero-loss profile $Z(E) = I_0 R(E)$, one can obtain Equation (37):

$$j(\nu) = \sum C_n r(\nu)[s(\nu)]^n = I_0 r(\nu) \exp \frac{s(\nu)}{I_0}$$

$$= z(\nu) \exp \frac{s(\nu)}{I_0} \tag{37}$$

The SSD (Figure 25a and b, curve 2) may be computed as the inverse Fourier transform of Equation (38):

$$s(\nu) = I_0 \ln \frac{j(\nu)}{z(\nu)} \tag{38}$$

where $j(\nu)$ and $z(\nu)$ are the Fourier transforms of the experimental EEL spectrum and of the zero-loss peak, respectively.

For quantitative EELS, the number n_a of atoms of element i per unit area may be obtained as Equation (39):[2]

$$n_a = \frac{1}{\sigma_i(\alpha, \Delta)} \frac{N_i(\alpha, \Delta)}{N_0(\alpha, \Delta)} \tag{39}$$

where $\sigma_i(\alpha, \Delta)$ is the partial cross-section for ionizations which result in scattering of the PE within $0 \leq \theta \leq \alpha$ and the energy-loss region with an energy window of width $\Delta E_I \leq \Delta E \leq E_I + \Delta$; N_0 and N_i are the number of electrons in the range $0 \leq \Delta E \leq \Delta$ and $E_I \leq \Delta E \leq E_I + \Delta$, respectively.

As in X-ray analysis, a ratio method is utilized to determine the relative number of atoms n_i/n_j of two elements (Equation 40):

$$\frac{n_i}{n_j} = \frac{N_i}{N_j} \frac{\sigma_j(\alpha, \Delta)}{\sigma_i(\alpha, \Delta)} \tag{40}$$

7 APPLICATIONS

Examples selected in this section demonstrate some methodological innovations and practically important applications of AEM and scanning microanalysis such as advanced EPXMA particle analysis of populations of mixed silver halide microcrystals of contemporary industrial photographic emulsions under cryogenic conditions (section 7.1), combined X-ray compositional mapping (section 7.2.1), EELS spectrum-imaging and image-spectrum (section 7.2.2) and local specimen thickness determination (section 7.3).

7.1 Automated Population Analysis: Size–Thickness–Elemental Composition

Figure 26(a–c) shows (a) a typical X-ray spectrum, (b) some results on elemental compositions and

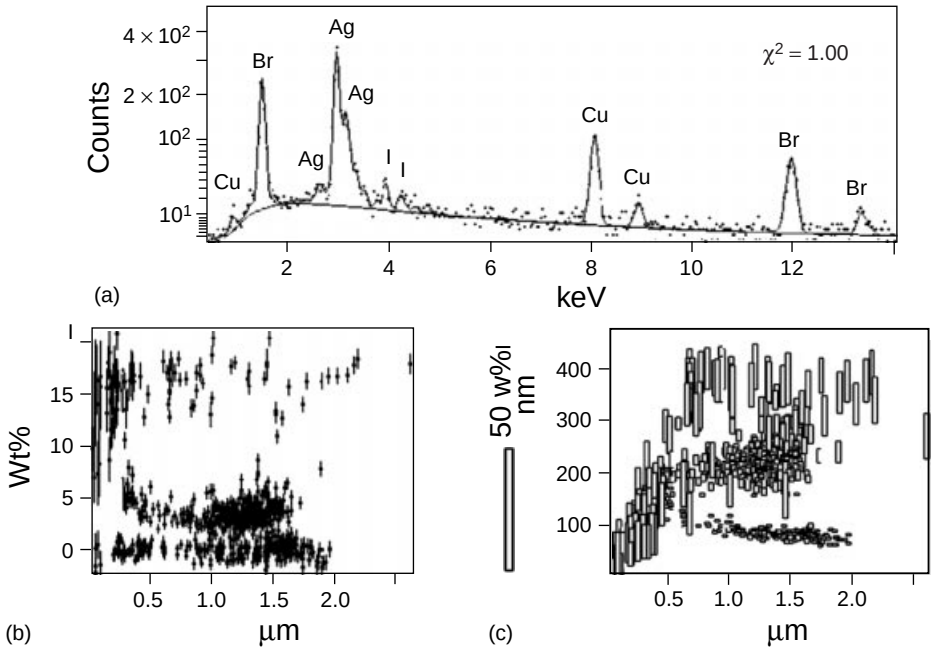

Figure 26 (a) X-ray spectrum of an Ag (Br, I) microcrystal, (b) iodine content (wt%) of 700 crystals as a function of crystal diameter (μm) showing standard deviations on iodine concentrations and (c) iodine content of the population as a function of crystal diameter (μm) and thickness (nm). Background fitted by least squares is shown in (a).

(c) thickness analyses obtained by cryo-EPXMA for 700 Ag (Br, I) microcrystals of a contemporary industrial photographic emulsion.[46] A JEOL JXA-733 electron microprobe operating at a 25 kV accelerating voltage and 1 nA beam current was equipped with a Noran Si(Li) EDS detector placed at a take-off angle of 40°, a Wolf–Everhart BSE detector, an Everhart–Thornley SE detector, a scintillator TE detector and a transmission specimen holder. The crystal size was determined by SE/BSE imaging, the thickness was estimated from the BSE intensity and the X-ray spectra provided the required composition information. However, in order to perform automated analyses, several serious problems had to be overcome. First, the AgX (X = Br, I) microcrystals are not only sensitive to light irradiation, but also degrade rapidly under the influence of the electron beam. Under normal analysis conditions, the crystals are often destroyed within seconds owing to the partial evaporation of halogens. Lowering the sample temperature to about −190 °C by the use of a liquid-nitrogen-cooled Oxford C10001 cryo-stage increased the resistance against radiation damage, thus allowing an X-ray analysis during 20 s or more.

Sample cooling inevitably caused sample drift; large temperature differences inside the microprobe instrument led to metal shrinkage, which in turn resulted in unpredictable motion of the sample under the electron beam. This drift continued for several hours after sample insertion and was far too large (about 50 nm min^{-1}) for the usual particle analysis software; indeed, the detected particles had moved away from their detected position before their X-ray spectrum acquisition even began. The problem was resolved by employing an original algorithm, which detects and analyzes particles sequentially during (and not following) the BSE image acquisition.

Particles were distinguished from the background by selecting a threshold value. To avoid missing small or very thin crystals, the threshold value was set as close to the background as possible. Owing to the limited sharpness of the BSE image, however, this might sometimes cause an overestimation of the projected area or diameter of the crystal. An accurate area determination was achieved by setting a second threshold value, different for every particle, halfway between the background image brightness and the maximum particle brightness. If the number of particle pixels rejected by this second threshold was too large, this was considered as an indication that the detected structure was in fact composed of overlapping particles. Therefore, it was eliminated from the analysis. The occurrence of large shape factor (perimeter$^2/4\pi \times$ area) values indicates touching or overlapping of crystals; they were also eliminated.

Since the crystals were very thin and high beam currents were undesirable because of possible crystal destruction, the X-ray emission was relatively low. Therefore, a large 30 mm^2 area Si(Li) detector placed as close as possible to the sample was used. The crystals were mounted on a very thin supporting foil (e.g. carbon foil on copper grid) using a transmission holder in order to reduce the bremsstrahlung background. Net peak intensities used for the accurate determination of the crystal composition were obtained by means of least-squares fitting.

7.2 X-ray and Electron Energy-loss Spectroscopic Elemental Mapping

7.2.1 X-ray Mapping

X-ray maps provide valuable information on two-dimensional elemental distributions over bulk or thin samples.[3,4,33,36] X-ray mapping can be undertaken with both EDS and WDS, where incoming counts for elements of interest are fed back into the SEM or STEM instrument. The X-ray analyzer takes control over the electron beam via a special interface, driving the beam around in a rectangular frame and collecting the X-ray emission for each pixel separately. An external beam deflection interface allows X-ray mapping by beam rocking also on TEM instruments using the standard hardware and software. In the case shown in Figure 27, combined WDS/EDS X-ray composition maps of an Na–Al–Si–Ca–Fe–Zr–Sn–O ceramic were acquired on a JEOL JXA-8900 WDS/EDS combined automated microanalyzer operating at 15 kV accelerating voltage. The simultaneous composition measurements of eight constituent elements by EDS (Al, Si, Fe, Sn) and by WDS (O, Na, Ca, Zr) and BSE imaging in a "compo" mode were performed on a $140 \times 140\,\mu m^2$ area.

Color compositional maps represent spatially resolved X-ray multielemental analysis within the chosen region of the sample. However, it is often difficult to visualize the compositional ranges and resulting interelement correlations from the map, particularly for minor or trace constituents. As an alternative method, a composition–composition histogram displaying the numerical relationship between concentrations of the components at various points of the sample can be proposed.[36]

The SE/BSE images are normally used to select regions for subsequent X-ray mapping. If features of interest are not differentiated in the monochrome black–white electron image then they can easily be overlooked for further analysis. Using the X-ray spectrum detected by EDS, one can construct a color response directly related to the underlying elemental composition where the spectrum from each compound has a characteristic

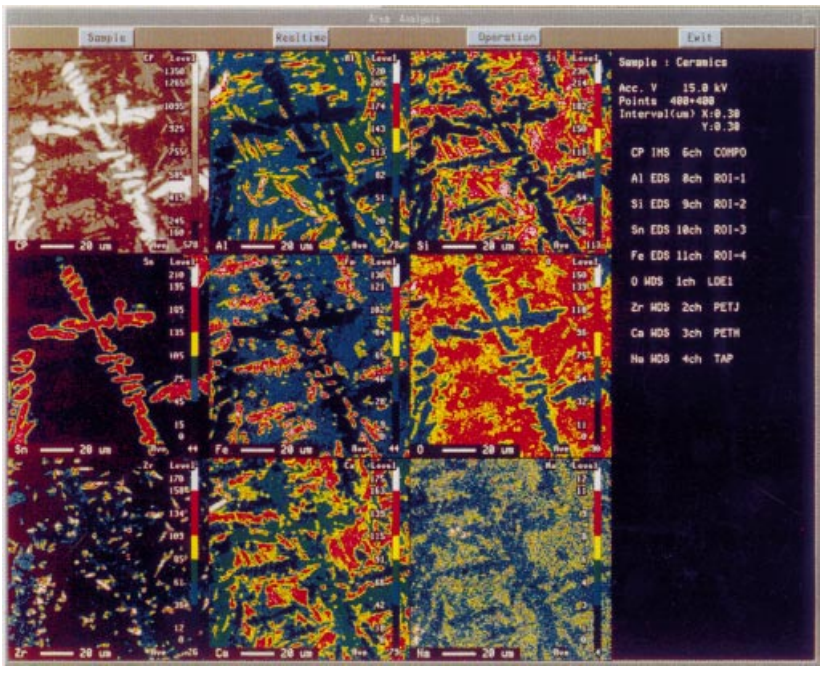

Figure 27 WD/ED combined 140 × 140 μm^2 area analysis of an Na–Al–Si–Ca–Fe–Zr–Sn–O ceramic, 400 × 400 points. X-ray maps of eight constituent elements and backscattered electron image (''compo'' mode, CP) recorded at an accelerating voltage of 15 keV. Mapping was performed simultaneously by EDS (Al, Si, Fe, Sn) and WDS (O, Na, Ca, Zr). (Courtesy of JEOL.)

color. Thus, topography and elemental composition of a specimen are compressed into a single view. Furthermore, using this technique it is not necessary to monitor the X-ray spectrum, set windows or collect X-ray maps. Compositional data even in samples with a rough surface are automatically acquired into the SEM images because the colors displayed always relate to the sample composition.[3]

7.2.2 Electron Energy-loss Spectroscopic Spectrum-imaging and Image-spectrum

Using a PEEL spectrometer (and/or EDS) fitted to a dedicated STEM instrument (FEG-VG H501) and/or TEM/STEM (a Hitachi H700H and a Philips EM430 were first used), a large four-dimensional data array may be recorded, which corresponds to electron intensity within a three-dimensional (x, y, E) space, one independent scattering angle dimension and a dependent counts dimension.[2,11,47,48] Here x and y are image coordinates and E is the selected energy loss. A similar series of energy-filtered images at different energy losses can be also acquired using EFTEM (Zeiss computerized EM902 and Zeiss (Leo) EM912 Omega) equipped with an imaging filter (image-spectrum[11,49]).

These techniques enable one to collect a very large amount of information for subsequent off-line processing in order to subtract the pre-edge background, to calculate local thickness and, to perform mean least-squares fitting, deconvolution and Kramers–Kronig analyses. The data may be displayed as line scans, two-dimensional arrays of spectra or two-dimensional images of elemental concentration, thickness, complex dielectric permittivity, etc.

7.3 Sample Thickness Determination

A need for the determination of the local specimen thickness is associated with absorption and/or fluorescence corrections and estimation of the spatial analysis resolution (EDS) and with multiple scattering corrections (EELS). Several means for measuring specimen thickness have been proposed (Table 5), but no one technique has been recognized as universal, although methods based on absorption of X-rays, EELS and CBED may be recommended because they provide possibilities for automated on-line procedures.

8 CURRENT TRENDS

The last decade has shown considerable progress in the theoretical description of fundamental electron–solid interactions and methodology of AEM, especially on standardless and fully quantitative procedures, automatic instrumentation, low-voltage microanalysis

Table 5 EM methods for the measurement of specimen thickness[2,33,45]

EM method	Signals used	Samples	Thickness–signal relation
CTEM	TE	Cross-sections	Direct measurement
CTEM	TE	Thin films	Beam current measurement vs thickness using a Faraday cup followed by densitometric calibration
CTEM	TE	Shadowing of latex particles, shadowed replicas, small fold	Direct measurement
CTEM	TE	Tilting of small latex spheres deposited on both sides	Measurement of parallax shift between balls on the top surfaces and bottom sides
CTEM	BSE	Thin films	Measurements of backscattering coefficient vs thickness using the equation $\eta = \text{constant} \times N_A Zt$
SEM	SE	Cross-sections, fractures	Direct measurement
SEM	Characteristic X-rays	Thin films on bulk support	Measurement of X-ray intensity vs thickness, accelerating voltage followed by calibration with standards
CTEM/STEM	TE	Contamination spots	Measurement of separation between spots on the top and bottom surfaces
CTEM/STEM	TE	Features intersecting both surfaces (faults)	Projection measurements of tilted and untilted features against the beam
CTEM/STEM/EELS	Zero-loss/inelastically scattered electrons	Thin films	Relative and absolute measurements using a parameterized equation, $I_t/I_0 = t/\lambda$. Bethe sum rule. Kramers–Kronig sum rule
CBED	TE	Thin crystalline films	Determination of deviations of diffraction fringes from the exact Bragg position

of light elements and super-high-sensitivity analysis at nanometer and subnanometer lateral resolution as well, which have been briefly discussed in this article.

Nowadays, there is a tendency to develop integrated multipurpose AEM systems based on TEM/STEM and/or SEM systems equipped with multiple detector systems and to detect the effects of inelastic interactions of electron beams within the object (WDS, EDS, PEEL spectroscopy, ESI, AES/SAM). On the other hand, standard modes, in particular electron and X-ray diffraction methods (SAED, CBED, electron channeling, Kikuchi patterns), DF and stereo observations, are, of course, still widely used. As a result, conventional TEM, STEM and SEM instruments have evolved from a purely imaging/diffraction instrument into a true AEM tool using most of the signals produced in the course of electron beam–sample interactions, although for ultimate performance it is mostly better to optimize an instrument for that particular purpose (for instance, as in the case of a dedicated STEM instrument). In HRTEM and STEM *Z*-contrast modes spatial resolution ($\sim$0.1 nm) has steadily reached the atomic level. In high-resolution FE/SEM one can achieve resolution of medium-resolution TEM but for bulk specimens. Further developments indicate rapid progress towards atomic-level analytical resolution. These significant improvements became possible by the introduction of very bright electron sources (FEG, $\sim 10^8 - 10^9$ A cm^{-2} sr^{-1}), capable of producing fine electron probes less than 0.2 nm in diameter. This capability has dramatically boosted both imaging and analytical AEM modes.

Development of PEEL detection systems based on CCD offers better detection limits; they allow one to investigate with much higher efficiency not only the low losses but also the more local and less intense core losses up to 3000 eV. An integral energy filter is fast becoming a standard option on the latest generation of AEM instruments owing to the rising importance of EFTEM/EELS for novel imaging techniques such as elastic contrast, energy-loss tuning and high-resolution electron spectroscopic elemental mapping. The development of a new microcalorimeter EDS system with an energy resolution of 3–10 eV allowed the detection not only of closely spaced X-ray peaks but also of chemical shifts. Hence the versatility of AEM, combining high-resolution imaging in different modes and powerful analytical facilities, the variety of signals and contrast effects available in the course of electron beam–specimen interactions, establish it as an outstanding tool for universal applications in chemistry and materials science.

ACKNOWLEDGMENTS

The authors are grateful to Dr R. Herstens, JEOL (Europe) BV, Belgium, and Dr M. Otten, Philips (Eindhoven), The Netherlands, for providing valuable scientific and technical information.

ABBREVIATIONS AND ACRONYMS

ADF	Annular Dark-field
AE	Auger Electrons
AEM	Analytical Electron Microscopy
AES	Auger Electron Spectroscopy
BF	Bright Field
BSE	Backscattered Electrons
CBED	Convergent-beam Electron Diffraction
CBEM	Convergent-beam Electron Microscopy
CCD	Charge-coupled Device
CL	Cathodoluminescence
CRT	Cathode-ray Tube
CTEM	Conventional Transmission Electron Microscopy
DE	Diffracted Electrons
DF	Dark Field
EBIC	Electron Beam-induced Current
ED	Electron Diffraction
EDS	Energy-dispersive (X-ray) Spectroscopy
EEL	Electron Energy-loss
EELS	Electron Energy-loss Spectroscopy
EFTEM	Energy-filtering Transmission Electron Microscopy
ELNES	Energy-loss Near-edge Structure
EM	Electron Microscopy
EMMA	Electron Microscopy Microanalyzer
EP	Electron Plasmons
EPMA	Electron Probe Microanalysis
EPXMA	Electron Probe X-ray Microanalysis
ESD	Electron Spectroscopic Diffraction
ESEM	Environmental Scanning Electron Microscopy
ESI	Electron Spectroscopic Imaging
EXELFS	Extended Energy-loss Fine Structure
FE	Field Emission
FEG	Field-emission Gun
HRTEM	High-resolution Transmission Electron Microscopy
HVTEM	High-voltage Transmission Electron Microscopy
IR	Infrared
LP	Lattice Phonons
MMF	Minimum Mass Fraction
PE	Primary Electrons
PEEL	Parallel Electron Energy-loss
SAED	Selected-area Electron Diffraction
SAM	Scanning Auger Microscopy
SC	Sample Current
SE	Secondary Electrons
SEM	Scanning Electron Microscopy
SSD	Single Scattering Distribution
STEM	Scanning Transmission Electron Microscopy
TE	Transmitted Electrons
TEM	Transmission Electron Microscopy
UV/VIS	Ultraviolet/Visible
WDS	Wavelength-dispersive (X-ray) Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Microscopy of Coatings

Forensic Science **(Volume 5)**
Scanning Electron Microscopy in Forensic Science • X-ray Fluorescence in Forensic Science

Industrial Hygiene **(Volume 5)**
Aerosols and Particulates Analysis: Indoor Air

Particle Size Analysis **(Volume 6)**
Particle Size Analysis: Introduction • Velocimetry in Particle Size Analysis

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Differential Reflectance Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

Mass Spectrometry **(Volume 13)**
Secondary Ion Mass Spectrometry as Related to Surface Analysis

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

General Articles **(Volume 15)**
Archaeological Chemical Analysis • Spot Test Analysis

REFERENCES

1. V. Oleshko, R. Gijbels, W. Jacob, M. Alfimov, 'Characterization of Complex Silver Halide Photographic

Systems by Means of Analytical Electron Microscopy', *Microbeam Anal.*, **4**, 1–29 (1995).
2. R.F. Egerton, *Electron Energy-loss Spectroscopy in the Electron Microscope*, 2nd edition, Plenum Press, New York, 1–29, 31–129, 245–299, 301–406, 1996.
3. V. Oleshko, R. Gijbels, 'Scanning Microanalysis', in *Handbook of Microscopy*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, VCH, Weinheim, 661–690, Vol. 2, 1997; in *Electron Microscopy, Principles and Fundamentals*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, Wiley-VCH, Weinheim, 427–456, 1997.
4. V.D. Scott, G. Love, S.J.B. Reed, *Quantitative Electron-probe Microanalysis*, 2nd edition, Ellis Horwood, New York, 1–75, 111–211, 1995.
5. P. Duncumb, 'Precipitation Studies with EMMA – a Combined Electron Microscope and X-ray Microanalyzer', in *The Electron Microprobe*, eds. T.D. McKinley, K.F.J. Heinrich, D.B. Wittry, Wiley, New York, 490, 1966.
6. G. Rutherman, 'Diskrete Energieveluste Schneller Elektronen in Festkörpern', *Naturwissenschaften*, **29**, 648 (1941).
7. J. Hillier, R.F. Baker, 'Microanalysis by Means of Electrons', *J. Appl. Phys.*, **15**, 663–675 (1944).
8. R. Castaing, L. Henry, 'Filtrage Magnetique des Vitesses en Microscopies Electronique', *C.R. Acad. Sci.*, **B255**, 76–78 (1962).
9. M.T. Browne, 'Electron Energy Analysis in a Vacuum Generators HB5 STEM', in *Scanning Electron Microscopy*, SEM Inc., AMF O'Hare, IL, Part 2, 827–834, 1979.
10. L. Reimer, 'Introduction' and 'Electron Spectroscopic Imaging', in *Energy-filtering Transmission Electron Microscopy*, ed. L. Reimer, Springer Series in Optical Sciences, Springer, Berlin, 1–42, 347–400, Vol. 71, 1995.
11. Ch. Colliex, 'Electron Energy-loss Spectroscopy Imaging', in *Handbook of Microscopy*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, VCH, Weinheim, 425–445, Vol. 1, 1997.
12. L. Reimer, *Transmission Electron Microscopy. Physics of Image Formation and Microanalysis*, 4th edition, Springer, Berlin, 1–15, 157–196, 251–264, 404–420, 423–461, 1997.
13. F. Hofer, 'Inner-shell Ionizations', in *Energy-filtering Transmission Electron Microscopy*, ed. L. Reimer, Springer Series in Optical Sciences, Springer, Berlin, 225–268, Vol. 71, 1995.
14. D.E. Newbury, 'Monte Carlo Electron Trajectory Simulation for Electron Probe Analysis: Present Utility and Future Possibilities', *Microbeam Anal.*, **1**, 9–17 (1992).
15. P.B. Hirsch, R.B. Nicholson, A. Howie, D.W. Pashley, M.J. Whelan, *Electron Microscopy of Thin Crystals*, 2nd edition, Butterworths, London, 1977.
16. G. Thomas, *Transmission Electron Microscopy of Metals*, John Wiley and Sons, New York, 1962.
17. H. Bethge, J. Heydenreich (eds.), *Elektronenmikroskopie in der Festkörperphysik*, Springer, Berlin, 1982.
18. S. Amelinckx, R. Gevers, J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science*, North-Holland, Amsterdam, 1978.
19. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo (eds.), *Electron Microscopy, Principles and Fundamentals*, Wiley-VCH, Weinheim, 1997.
20. S. Amelinckx, D. Van Dyck, 'Diffraction Contrast and High Resolution Electron Microscopy of Structures and Structural Defects', in *Electron Diffraction Techniques*, IUC Monographs on Crystallography 4, ed. J. Cowley, Oxford University Press, Oxford, 1–222, Vol. 2, 1993.
21. J.C.H. Spence, *Experimental High Resolution Electron Microscopy*, Monographs on the Physics and Chemistry of Materials, Oxford Science Publications, Clarendon Press, Oxford, 1981.
22. A.K. Head, P. Humble, L.M. Clarebrough, A.J. Morton, G.T. Forwood, 'Computed Electron Micrographs and Defect Identification', in *Defects in Crystalline Solids*, eds. S. Amelinckx, R. Gevers, J. Nihoul, North-Holland, Amsterdam, 1–400, Vol. 7, 1973.
23. S. Pennycook, 'Scanning Transmission Electron Microscopy: *Z* Contrast', in *Handbook of Microscopy*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, VCH, Weinheim, 595–620, Vol. 2, 1997.
24. L. Fiermans, R. De Gryse, 'Scanning Auger Microscopy (SAM) and Imaging X-ray Photoelectron Spectroscopy (XPS)', in *Handbook of Microscopy*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, VCH, Weinheim, 621–659, Vol. 2, 1997.
25. B.G. Yacobi, D.B. Holt, *Cathodoluminescence Microscopy of Inorganic Solids*, Plenum Press, New York, 89–118, 121–268, 1990.
26. N.D. Browning, S.J. Pennycook, 'Atomic Resolution Spectroscopy for the Microanalysis of Materials', *Microbeam Anal.*, **2**, 81–89 (1993).
27. C.E. Lyman, J.I. Goldstein, D.B. Williams, D.W. Ackland, S. Von Harrach, A.W. Nicholls, P.J. Statham, 'High-performance X-ray Detection in a New Analytical Electron Microscope', *J. Microsc.*, **176**, 85–98 (1994).
28. C.E. Nockolds, 'Low-voltage Electron Probe X-ray Microanalysis in an In-lens Field Emission Gun SEM: an Evaluation', *Microbeam Anal.*, **3**, 185–189 (1994).
29. G.D. Danilatos, 'Environmental Scanning Electron Microscopy and Microanalysis', *Mikrochim. Acta*, **114/115**, 143–155 (1994).
30. M.A. Baker, J.E. Castle, 'Scanning Auger Microscopy', in *Materials Science and Technology: a Comprehensive Treatment*, eds. R.W. Cahn, P. Haasen, E.I. Kramer, VCH, Weinheim, Chapter 13, 220–239, Vol. 2B, 1994.
31. D.A. Wollman, K.D. Irvin, G.C. Hilton, L.L. Dulcie, D.E. Newbury, J.M. Martinis, 'High-resolution, Energy-dispersive Microcalorimeter for X-ray Microanalysis', *J. Microsc.*, **188**, 196–223 (1997).
32. V. Oleshko, V. Kindratenko, R. Gijbels, P. Van Espen, W. Jacob, 'Study of Quasi-fractal Many-particle Systems by Zero-loss Spectroscopic Imaging, Electron Energy-loss

Spectroscopy and Digital Image Analysis', *Mikrochim. Acta*, Suppl. 13, 443–451 (1996).
33. D.B. Williams, C.B. Carter, *Transmission Electron Microscopy. A Textbook for Materials Science*, Plenum Press, New York, 553–703, 1996.
34. R.H. Packwood, J.D. Brown, 'A Gaussian Expression to Describe $\phi(\rho Z)$ Curves for Quantitative Electron Probe Microanalysis', *X-Ray Spectrom.*, **10**, 138–146 (1981).
35. J.L. Pouchou, 'Standardless X-ray Analysis of Bulk Specimens', *Mikrochim. Acta*, **114/115**, 33–52 (1994).
36. J.A. Small, 'Electron-induced X-ray Emission', in *Handbook of X-ray Spectrometry: Methods and Techniques*, eds. R.E. Van Grieken, A.A. Markowicz, Marcel Dekker, New York, Chapter 12, 583–655, 1993.
37. J.I. Goldstein, D.B. Williams, 'X-Ray Microanalysis and Electron Energy Loss Spectrometry in the Analytical Electron Microscope: Review and Future Directions', *Microbeam Anal.*, **1**, 29–53 (1992).
38. G. Cliff, G.W. Lorimer, 'The Quantitative Analysis of Thin Specimens', *J. Microsc.*, **103**, 203–207 (1975).
39. J.I. Goldstein, 'Principles of Thin Film X-ray Microanalysis', in *Introduction to Analytical Electron Microscopy*, eds. J.J. Hren, J.I. Goldstein, D.C. Joy, Plenum Press, New York, Chapter 3, 83–120, 1979.
40. J.E. Wood, D.B. Williams, J.I. Goldstein, 'Experimental and Theoretical Determination of k_{Afe} Factors for Quantitative X-ray Microanalysis in the Analytical Electron Microscope', *J. Microsc.*, **133**, Pt. 1, 255–274 (1984).
41. R.J. Graham, J.W. Steeds, 'Determination of Cliff–Lorimer k Factors by Analysis of Crystallized Microdroplets', *J. Microsc.*, **133**, Pt. 1, 275–280 (1984).
42. N. Zaluzec, 'Quantitative X-ray Microanalysis: Instrumental Considerations and Applications to Materials Science', in *Introduction to Analytical Electron Microscopy*, eds. J.J. Hren, J.I. Goldstein, D.C. Joy, Plenum Press, New York, Chapter 4, 121–167, 1979.
43. E. Van Cappellen, 'The Parameterless Correction Method in X-ray Microanalysis', *Microsc. Microanal. Microstruct.*, **1**, 1–22 (1990).
44. R.F. Egerton, 'Quantitative Analysis of Electron Energy-loss Spectra', *Ultramicroscopy*, **28**, 215–225 (1989).
45. R.F. Egerton, R.D. Leapman, 'Quantitative Electron Energy-loss Spectroscopy', in *Energy-filtering Transmission Electron Microscopy*, ed. L. Reimer, Springer, Berlin, 269–290, 1995.
46. C. Gregory, H.A. Nullens, R. Gijbels, W. Jacob, P.J. Van Espen, I. Geuens, R. De Keyzer, 'Automated Particle Analysis of Populations of Silver Halide Microcrystals by Electron Probe Microanalysis under Cryogenic Conditions', *Anal. Chem.*, **70**, 2551–2559 (1998).
47. C. Jeanguillaume, Ch. Colliex, 'Spectrum-image: the Next Step in EELS Digital Acquisition', *Ultramicroscopy*, **28**, 252–257 (1989).
48. J.A. Hunt, D.D. Williams, 'Electron Energy-loss Spectrum Imaging', *Ultramicroscopy*, **38**, 47–73 (1991).
49. J.-L. Lavergne, C. Foa, P. Bongrand, D. Seux, J.M. Martin, 'Application of Recording and Processing of Energy-filtered Image Sequences for the Elemental Mapping of Biological Specimens: Image-spectrum', *J. Microsc.*, **174**, Pt. 3, 195–206 (1994).
50. D.C. Joy, 'Scanning Electron Microscopy', in *Handbook of Microscopy*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, VCH, Weinheim, 539–561, Vol 2, 1997.
51. A.W. Robards, A.J. Wilson (eds.), *Procedures in Electron Microscopy*, John Wiley and Sons, Chichester, 15:1.4–15:1.5, 15:2.10–15:2.24, 15:5.1–15:5.6 (1993, 1996, 1997).

Ellipsometry in Analysis of Surfaces and Thin Films

Robert W. Collins
The Pennsylvania State University, USA

1 Introduction **2**
2 History **4**
3 Measurement Principles **5**
3.1 Polarization and the Optical Properties of Materials 6
3.2 Ellipsometry Measurement of a Single Interface 6
3.3 Ellipsometry Measurement of a One-layer Structure 8
3.4 Ellipsometry Measurement of a Multilayer Structure 9
4 Static Measurement Applications **10**
4.1 Instrument Designs, Operation, and Performance 10
4.2 Spectral Range 15
4.3 Spatial Resolution 17
4.4 Sample Preparation and Limitations 18
4.5 Data Analysis 21
4.6 Applications in the Characterization of Materials, Thin Films, and Structures 22
5 Real Time Measurement Applications **27**
5.1 Instrument Designs and Selection 27
5.2 Data Analysis 28
5.3 Applications in Gaseous and Plasma Environments 30
5.4 Applications in Liquid Environments 33
6 Comparison with Other Techniques **35**
7 Method Development **38**
Acknowledgments **39**
Abbreviations and Acronyms **39**
Related Articles **39**
References **39**

An ellipsometry measurement consists of five steps.

1. *A light beam is generated in a known polarization state.*
2. *The beam is reflected from or transmitted through a sample having specularly reflecting plane-parallel surfaces, leading to a linear (frequency-conserving) interaction that changes the polarization state.*
3. *The polarization state of the reflected or transmitted beam is measured.*
4. *Parameters are determined that characterize the interaction in terms of the change in polarization state.*
5. *From these parameters, information about the sample is deduced, including its optical properties, or the thickness and optical properties of a thin film on the surface of the sample.*

Steps 1–3 involve set-up and operation of instrumentation whereas step 4, data reduction, is generally performed on-line by computer during measurement. Step 5 is usually performed after measurement, unless ellipsometry is being applied for real time process control.

Several variations of the ellipsometry experiment have been developed. The most common is reflection ellipsometry, in which the polarization state change is measured upon oblique reflection of the light beam from a specularly reflecting sample. For an isotropic sample, the parameters describing this interaction are (ψ, Δ)*, defined by* $\tan\psi \exp(i\Delta) = r_p/r_s$*, where* r_p *and* r_s *are the complex amplitude reflection coefficients for linear p and s polarization states. In these states, the electric field vibrates parallel and perpendicular to the plane of incidence, respectively. In spectroscopic ellipsometry (SE),* (ψ, Δ) *are measured continuously versus wavelength over a spectral range designed for sensitivity to the physical and chemical features of interest. In real time ellipsometry (RTE),* (ψ, Δ) *are measured versus time at fixed wavelength using a time interval designed to follow the kinetic processes of interest. The latter two modes can be combined to yield real time spectroscopic ellipsometry (RTSE), utilizing parallel detection at many wavelengths simultaneously. In imaging ellipsometry (IE),* (ψ, Δ) *are measured over a two-dimensional area of a nonuniform sample surface using a resolution designed to detect feature sizes of interest.*

Because the polarization change is determined in an ellipsometry measurement, rather than the irradiance change as in a reflectance measurement, ellipsometry is a very high precision technique for characterizing the optical properties of materials and thin films, as well as the thicknesses of thin films. The $p-s$ *phase shift difference* Δ *provides exceptionally high sensitivity to changes in the thickness of a thin film on a specular surface. In fact, changes on the order of hundredths of a monolayer in surface coverage can be detected. A primary advantage of ellipsometry is its ability to obtain data noninvasively from samples immersed in any medium that is transparent to the light beam. Thus, numerous in situ and real time applications exist in disciplines such as physics and chemistry, materials and surface science, electrical and chemical engineering, electrochemistry, and*

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

biochemistry. In contrast, analytical techniques based on electron spectroscopies, for example, are only effective in high vacuum environments; thus, in vacuo instrumentation is required.

Limitations of ellipsometry arise from restrictions on the nature of the sample, as well as from the difficulties of data analysis, particularly for complex samples e.g. those exhibiting nonuniformities over the area of the probe beam. To avoid measurement difficulties as a result of light scattering and to simplify data analysis, reflecting surfaces should be specular and free of optical nonuniformities on an in-plane scale greater than about one-tenth of the wavelength of the incident beam. Nonuniformities on a scale greater than the wavelength can be detected by IE; in other measurement modes the sample should be uniform over the beam area for ease of data analysis. In addition, characterization of films that are thinner than the thickness of the roughness modulations on the substrate surface is difficult since data analysis relies on optical models consisting of multilayer stacks with plane-parallel interfaces. Recent progress has been made in ellipsometry research that addresses the analysis of complex samples.

The most widely used instruments for SE and RTSE span the spectral range from the ultraviolet (200–300 nm) to the near-infrared (800 nm). Over this spectral range, the optical properties deduced from SE provide information on the processes of absorption and dispersion originating from the valence electron transitions. Specialized SE equipment designed for the infrared range can be more effective for chemical identification of thin films, based on the absorption processes originating from the bond vibrational modes.

1 INTRODUCTION

The analytical technique of ellipsometry derives its name from the measurement of the ellipse of polarization that is generated when a polarized light beam interacts with a sample under analysis.[1] The instrument that performs this measurement is called an ellipsometer. The ellipsometry experiment entails five steps:

1. generation of a light beam in a known state of polarization using optical elements such as polarizers and compensators (or retarders),
2. frequency and wavefront-conserving interaction of the beam with a sample leading to an emergent beam in a second polarization state,
3. full or partial characterization of the second polarization state,
4. determination of parameters that characterize the interaction from the information on the two polarization states and
5. deduction of information on the sample, such as optical properties and film thicknesses, from the interaction parameters.

As a result of the interaction, the light beam may emerge by transmission through the sample (i.e. transmission ellipsometry) or by oblique reflection from its specular surface (i.e. reflection ellipsometry). The original and widest usage of the term ellipsometry, however, is to describe the reflection experiment performed to obtain the optical properties of specular surfaces, the optical properties and thicknesses of thin films, and the nature of specular interfaces. Transmission ellipsometry is used for analysis of molecular structure through natural anisotropy, however, this experiment is most often identified, not as ellipsometry, but in terms of the specific information extracted such as optical rotation or circular dichroism.

The most general polarization state of a light beam is specified by the Stokes parameters.[1] These four parameters describe (i) the irradiance (or power flux) I in the beam, (ii) the elliptical polarization state of the beam, characterized by the tilt angle Q ($-90° < Q \leq 90°$) and the ellipticity e ($-1 \leq e \leq 1$), which is the ratio of the length of the minor axis to that of the major axis of the ellipse, and (iii) the degree of polarization p ($0 \leq p \leq 1$). Measurement of the incident and reflected irradiances I_i and I_r, respectively, yields the reflectance $R = I_r/I_i$ of the sample in a measurement called reflectometry. This is distinct from the ellipsometry measurement, which involves measurement of (Q_i, e_i) and (Q_r, e_r) for the incident and reflected beams. In some instruments, reflectometry and ellipsometry are performed simultaneously; then the measurement is sometimes called three-parameter ellipsometry.[2] Other instruments provide Q, e, and p of the reflected beam, a capability that is important when the sample depolarizes the incident beam. This measurement is often called polarimetry. In standard reflection ellipsometry on a uniform isotropic sample, rather than expressing the outcome in terms of the beam parameters Q and e, it is conventional to use the sample parameters ψ and Δ, where $\tan\psi \exp(i\Delta) = r_p/r_s$. In this expression, r_p and r_s are the complex amplitude reflection coefficients of the reflecting sample for p and s linear polarization states in which the electric field vibrates parallel (p) and perpendicular (s) to the plane of incidence.

Because two data values ψ and Δ are determined in a standard ellipsometry experiment, a measurement at one wavelength can provide two sample parameters.[1] For example, for a sample consisting of an atomically smooth film-free material, the complex index of refraction N of the material can be obtained at the wavelength λ of the incident beam. N provides a complete description

of the linear optical properties of an isotropic material and includes as its real part the real index of refraction n and as its imaginary part the extinction coefficient k, i.e. $N = n - ik$. For a sample consisting of a transparent film (i.e. $k = 0$) on a substrate of known complex index of refraction, n of the film and its thickness d can be determined within limitations. In contrast to ellipsometry, only one data value R is determined in a reflectometry experiment. Hence, reflectometry at selected λ on a film-free material cannot provide N. As a result, ellipsometry is the preferred method for measuring N for λ within the range of conventional ellipsometers, from the ultraviolet (200–300 nm) to the near-infrared (800 nm), for strongly absorbing materials, e.g. metals and semiconductors, and for weakly absorbing materials that exist only in thin film form. One limitation of ellipsometry is its inability to measure low k values or low absorption coefficients α, where $\alpha = 4\pi k/\lambda$. Typical ellipsometers are limited by accuracy to $\alpha > 10^2\,\mathrm{cm}^{-1}$. This is in contrast to transmission spectroscopy in which the ratio of the transmitted to incident irradiance is determined, and the sample thickness may be increased to measure low α values.

The term ellipsometry was coined by Rothen in 1944, and the precursor of present day instruments was developed soon after when the photomultiplier tube (PMT) became available.[3] The primary limitation of early manually operated instruments was the time taken to perform one (ψ, Δ) measurement, typically ~10 min. Thus, the early instruments could be used only to study static or quasistatic samples. Upon introduction of laboratory computers for data collection and reduction, a number of automatic ellipsometers were invented, starting in the late 1960s.[4] Design differences arose from the diverse methods used to generate and detect polarized light. A common feature of these automatic instruments is the 10^3–10^8 increase in measurement speed compared with the manual ellipsometer. As a result, two important capabilities emerged, RTE and SE. In RTE, measurements at fixed λ are performed versus time on dynamic samples, revealing the kinetics of film growth or materials and surface modification. In SE, measurements closely spaced versus λ are performed on static or quasistatic surfaces. The latter measurements yield the λ-dependence of the optical properties of the sample, allowing one access to a wealth of information including sample composition, crystallinity, microstructure, and temperature, providing that suitable databases exist. Recent advances have led to an expansion of the capabilities of automatic ellipsometers.[2] For example, the measurement time for a 100-point spectrum from the ultraviolet to near-infrared has been reduced from ~15 min for a conventional automatic spectroscopic ellipsometer to 16 ms for a multichannel ellipsometer with a photodiode array detector. The latter advance permits real time spectroscopy on dynamic surfaces.

The ellipsometers described in the previous paragraph were designed for selected wavelength or spectroscopic operation from the ultraviolet (200–300 nm) to the near-infrared (800 nm). The optical properties of materials over this range provide information on absorption and dispersion associated with electronic transitions between the valence and conduction bands.[5] Thus, the density of electronic states within $\sim\pm5$ eV of the Fermi level can be probed. Parallel efforts have led to ellipsometers with broad-band infrared sources covering the range from 700 to 7000 cm^{-1}.[6] The optical properties over this range provide information on the chemical bonding of a variety of materials as well as the free electron characteristics of conductors and doped semiconductors. Finally, ellipsometers have also been developed to study materials outside the ultraviolet to mid-infrared range.[2] These specialized instruments exploit synchrotron sources to cover (i) the far-infrared range from 20 to 800 cm^{-1} in studies of phonons and plasmons for superconductors, (ii) the vacuum ultraviolet from 50 to 200 nm in studies of the energy band structures of wide gap materials, and (iii) the extreme ultraviolet, ~10 nm in studies of core electronic levels.

In conventional reflection ellipsometry, the surface is probed with a collimated 1–5 mm diameter beam at an angle of incidence of 50–75°.[1] Thus, the physical parameters of the sample must be uniform over the probed area if the (ψ, Δ) data are to be readily interpretable using simple optical models. However, samples are often not perfect over the full beam area. In comparison to reflectometry, however, ellipsometry provides greater immunity to localized surface defects that scatter light out of the specular beam. However, if the sample is very rough or inhomogeneous with a correlation length of the structure greater than $\sim0.1\lambda$, then conventional ellipsometry may fail due to light scattering that can depolarize the beam ($p < 1$) and lead to unmanageable complexities in data analysis. If the correlation length of the inhomogeneity is greater than $\sim10\lambda$, then IE or microellipsometry (ME) can be applied to characterize the sample.[2] In IE, a two-dimensional charge-coupled device (CCD) is used as a detector to collect images in (ψ, Δ) over a spot size of ~3 mm with a resolution of 10 μm. In ME, a spot size of ~10 μm is scanned over the sample surface to collect images in (ψ, Δ). Because of its sensitivity to relative phase changes, ellipsometry is more strongly affected than reflectometry by surface layers such as oxides that are sometimes neglected in data interpretation. Thus, the problem of obtaining accurate optical properties of materials by ellipsometry becomes one of preparing atomically clean smooth surfaces.

The popularity of ellipsometry arises from numerous applications that rely on its high surface sensitivity and its ability to probe samples in situ and in real time in varied environments.[1,7] Because of the high sensitivity to changes in the relative phases of the orthogonal p and s polarization components of the beam, through the parameter Δ, ellipsometry can detect changes in the average thickness of surface layers equivalent to 0.01 monolayer. Importantly, this level of sensitivity is possible for samples immersed in any medium that is transparent to the beam. First, ellipsometry can be used to characterize thin films, surfaces, interfaces and bulk solids prepared and processed in ultrahigh vacuum (UHV). It can also be used to probe processes occurring in various levels of vacuum or controlled gaseous environments, including physical and chemical adsorption, oxidation, annealing and crystallization, physical and chemical vapor deposition, and plasma etching. Finally, ellipsometry can also be applied to study processes occurring in laboratory ambients as well as in liquids. Such applications are in the field of electrochemistry, including ionic adsorption, anodic oxidation, passivation, corrosion, electropolishing, and electrodeposition and in the fields of biochemistry and medicine, including coagulation, immunological reactions, and cell surface characterization. Thus, ellipsometry has been used widely by chemists and physicists, by materials, electrical, and chemical engineers, and by biochemical and medical researchers.

2 HISTORY

The first studies in ellipsometry are credited to Paul Drude, although the measurement technique was not called ellipsometry until 1944.[3] As part of his dissertation research performed at Gottingen University and published from 1887 to 1890, Drude developed the first theory describing reflection and refraction of light at the surfaces of absorbing solids.[8–12] In attempts to test the theory experimentally, Drude found that the polarization state of light reflected from the surface of an optically absorbing crystal changed gradually with time after it was cleaved in air.[9] This change was attributed to oxidation of the surface, and Drude was able to relate quantitatively the observed polarization change to the complex index of refraction N of the crystal and the optical thickness (nd) of the oxide layer. In doing so, Drude utilized the parameters (ψ, Δ) to characterize reflection from the surface, and, applying Maxwell's equations, he derived expressions for (ψ, Δ) for the single-layer model (see section 3.3) consisting of an oxide on the crystal surface.[10,11] Drude also developed first order expansions of these expressions in (d/λ), valid for very thin films ($d < 1$ nm). Having identified the origin of the changes in polarization with time, Drude measured N for several metals.[12] His results are remarkably close in many cases to modern values.[13]

In Drude's study of the oxidation of crystals, he was unable to verify independently that the deduced thicknesses were correct. Much later in 1935, Leif Tronstad at the Norwegian Institute of Technology in Trondheim measured the thicknesses of monolayers of fatty acids on mercury surfaces using Drude's method and the first order expansions of the single layer expressions.[14] The thicknesses were found to be in accord with molecular models for the monolayer. Tronstad also performed pioneering studies of film formation on electrically biased metal surfaces immersed in solution.[15] In the early instruments, the polarization state of the reflected beam was determined through tedious visual observations and much effort was devoted to simplifying this process.

In 1944, Rothen first coined the term ellipsometer to describe an instrument that he developed to measure the thicknesses of protein films on slides.[3,16] In the instrument of Rothen (see Figure 1), incident monochromatic light was polarized by a polarizer with its transmission axis set at a known tilt angle P, measured with respect to the plane of incidence. The beam was reflected from a Cr-coated glass slide which was in turn coated with barium stearate monolayers. One-half of the illuminated area was covered by a single monolayer while the other half was covered by three monolayers. This generated an 4.8 nm step at the center of the viewing area. Thus, the reflected beam was elliptically polarized with the major axes of the ellipses from the two halves forming a 3° angle for optimal viewing. This reflected beam was analyzed first with a quarterwave retarder and then with a second polarizer (called the analyzer). The retarder imposes a 90° phase shift between the electric field components along orthogonal axes of the device, designated the fast (F) and slow (S) axes. When the retarder was oriented so that the fast axis bisected the angle formed between the major axes of the two ellipses, then the output of the waveplate consisted of two nearly linear polarization states with small right and left ellipticities for the two halves of the beam. In this situation, the analyzer was set such that if one looked through it at the sample, one saw a weak light output (nearly ''null'') of equal irradiances in the two halves of the beam. Finally, the film to be studied was added to the stearate-coated slide. This led to a brightening of one half of the beam and a dimming of the other half. By measuring the angle by which the analyzer had to be rotated to equalize the irradiance, the film thickness could be calculated. Since this rotation angle could be determined to within $\pm 0.01^\circ$, a change in thickness of 0.01 nm over the surface of the slide could

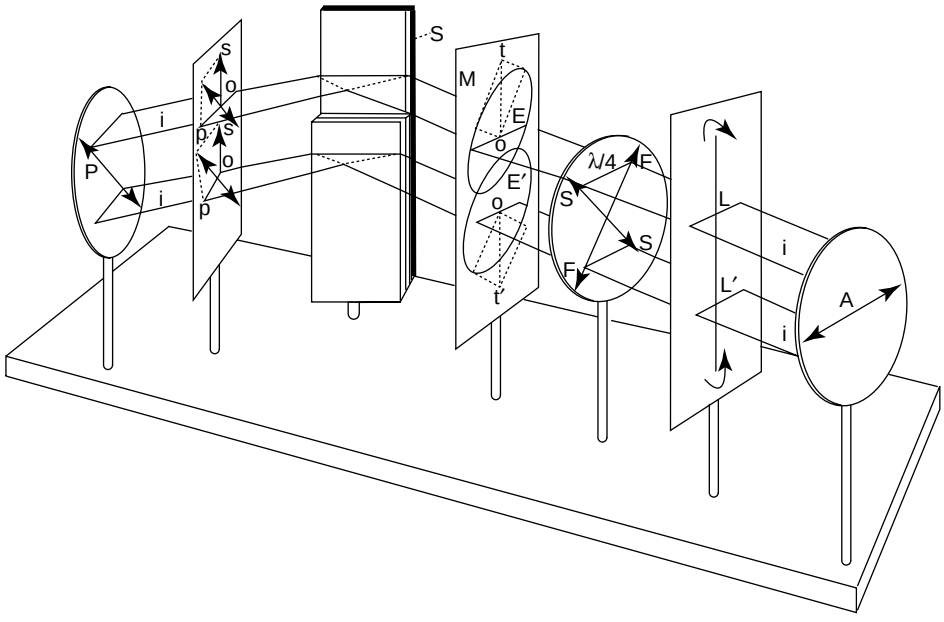

Figure 1 The first instrument to be called an ''ellipsometer''. The following are depicted: (i) polarizer at angle P, (ii) incident beam polarization states in the p–s coordinate frame, (iii) stepped sample S, (iv) reflected beam polarization states M, (v) quarterwave plate, $\lambda/4$, with the fast (F) and slow (S) axes shown, (vi) polarization states L and L′ for the beam after transmission through the quarterwave plate, (vii) analyzer at angle A. (Reproduced by permission of the American Institute of Physics from A. Rothen, *Rev. Sci. Instrum.*, **16**, 26–30 (1945).[16])

be detected. Rothen used such an instrument to follow immunological reactions.

The sequence of optical elements in Rothen's ellipsometer is identical to that used in today's state-of-the-art photometric ellipsometers.[2] The cumbersome half-shade detection, however, that relied on Langmuir–Blodgett film deposition methods, became obsolete soon after the development of the PMT in the late 1940s.[3] With the PMT used as a detector in place of the eye, the instrument configuration of Rothen shown in Figure 1 was adapted as one variant of the popular null ellipsometer.[1] In this case special substrates were not needed, and the polarizer and analyzer were rotated alternately to reach a null PMT output. The thickness and optical properties of the film were determined from the angular settings of the optical components at null (see section 4.1.1). In the 1950s and 1960s efforts were initiated to develop automatic ''recording'' ellipsometers that could collect (ψ, Δ) at sufficiently high speed to follow the kinetics film growth in real time.[17] These RTE efforts were motivated by the introduction of the first laboratory computers that could be used for instrument control, repetitive data collection, and data reduction. By 1969, several automatic ellipsometer designs based on null (or compensating) principles and on photometric principles had been developed,[4] as described in section 4.1. Starting in 1984, more advanced instruments were developed along the same basic principles but incorporating either rapid scanning sources or detection systems so that SE could be performed on dynamic surfaces in real time.[18] Such instruments for RTSE are continuing to be improved to this date.[2]

The advancement of the instruments and applications for ellipsometry can be appreciated by consulting the series of seven proceedings of conferences devoted to ellipsometry.[2,4,17,19–22] The first conference was held in 1963 in Washington DC, USA,[17] and the most recent one as of this writing was held in 1997 in Charleston SC, USA.[2]

3 MEASUREMENT PRINCIPLES

The theoretical basis for the ellipsometry measurement is given here. First, the electric field associated with polarized light will be described, and the optical properties of isotropic media will be defined. Then, the reflection of light from (i) a film-free isotropic material, (ii) a single-layer structure consisting of an isotropic ambient medium, a thin film, and a semi-infinite substrate, and finally, (iii) an *m*-layer structure consisting of *m* isotropic thin films between the ambient medium and substrate will be treated in detail. Treatments of the more difficult problem of anisotropic component materials can be found in Azzam and Bashara[1] and Collins et al.[2]

3.1 Polarization and the Optical Properties of Materials

The electric field, **E**, associated with a polarized monochromatic or quasimonochromatic light wave propagating along **z** in an isotropic medium is given by Equations (1) and (2):[1,23]

$$\mathbf{E}(\mathbf{r}, t) = (E_{0x}\mathbf{x} + E_{0y}\mathbf{y}) \exp\left\{ \mathrm{i}w \left(t - \frac{N\mathrm{z}}{c} \right) \right\} \quad (1)$$

$$= (E_{0x}\mathbf{x} + E_{0y}\mathbf{y}) \exp \left[\mathrm{i}w \left(t - \frac{n\mathrm{z}}{c} \right) \right] \times \exp\left\{ -\frac{wk\mathrm{z}}{c} \right\} \quad (2)$$

where E_{0x} and E_{0y} are defined by Equations (3) and (4):

$$E_{0x} = |E_{0x}| \exp(\mathrm{i}\delta_x) \quad (3)$$

$$E_{0y} = |E_{0y}| \exp(\mathrm{i}\delta_y) \quad (4)$$

These expressions follow the sign convention used most widely by the optics and ellipsometry communities. In Equations (1–4), w is the angular frequency of the light wave, c is the speed of light in vacuum, and E_{0x} and E_{0y} are the complex electric field amplitudes along the orthogonal x and y directions whose ratio specifies the polarization. In addition, $N = n - \mathrm{i}k$ is the complex index of refraction, where n is the (real) index of refraction and k is the extinction coefficient.

The polarization state of the plane wave of Equations (1–4) can be described in terms of the complex field amplitudes.[24] By defining Equations (5) and (6):

$$\tan\gamma = \frac{|E_{0y}|}{|E_{0x}|} \quad (5)$$

$$\delta = \delta_y - \delta_x \quad (6)$$

then the ellipticity angle $\chi = \tan^{-1} e$ $(-45^\circ \leq \chi \leq 45^\circ)$ and the tilt angle Q $(-90^\circ < Q \leq 90^\circ)$ associated with the ellipse of polarization are given by Equations (7) and (8):

$$\sin 2\chi = -\sin 2\gamma \sin\delta \quad (7)$$

$$\tan 2Q = -\tan 2\gamma \cos\delta \quad (8)$$

Within the convention used here, positive values of χ and e denote right-handed polarization states (clockwise rotation of E at a fixed point in space looking toward the source); negative values denote left-handed polarization states.

The index of refraction n describes the factor by which the speed of the wave in the medium is reduced relative to the speed in vacuum.[23] Thus, $v_\mathrm{p} = c/n$, where v_p is the speed of the planes of constant phase for the wave of Equations (1–4). The extinction coefficient k describes the exponential rate at which the electric field decays within the medium due to dissipative absorption. In fact, the absorption coefficient α of the medium is given by Equation (9):

$$\alpha = \frac{4\pi k}{\lambda} \quad (9)$$

α describes the exponential attenuation of the irradiance I with distance z_0 that the wave travels, according to $I = I_0 \exp(-\alpha z_0)$. An alternative approach for describing the optical properties is the dielectric function, $\varepsilon = \varepsilon_1 - \mathrm{i}\varepsilon_2 = N^2$, where ε_1 and ε_2 are the real and imaginary parts. The two descriptions of the optical properties are related through Equations (10–13):

$$\varepsilon_1 = n^2 - k^2 \quad (10)$$

$$\varepsilon_2 = 2nk \quad (11)$$

$$n = \left\{ \frac{\varepsilon_1 + (\varepsilon_1^2 + \varepsilon_2^2)^{1/2}}{2} \right\}^{1/2} \quad (12)$$

$$k = \left\{ \frac{-\varepsilon_1 + (\varepsilon_1^2 + \varepsilon_2^2)^{1/2}}{2} \right\}^{1/2} \quad (13)$$

Finally, it should be mentioned that for an anisotropic material, **E** is no longer orthogonal to the propagation direction, **z**. In addition, ε is a second-rank tensor, and there exist two values of N appropriate when the displacement, $\mathbf{D} = \varepsilon\mathbf{E}$, is along two mutually orthogonal directions.[23]

3.2 Ellipsometry Measurement of a Single Interface

In describing the reflection of a polarized, quasimonochromatic light wave from an ideal specular interface between two isotropic media, the electric field vectors of the incident (i) and reflected (r) waves are resolved along the orthogonal p and s directions, which are parallel (p) and perpendicular (s) to the plane of incidence. This yields the pairs of components $(E_{0\mathrm{p}})_\mathrm{i}$ and $(E_{0\mathrm{s}})_\mathrm{i}$ for the incident wave, and $(E_{0\mathrm{p}})_\mathrm{r}$, and $(E_{0\mathrm{s}})_\mathrm{r}$ for the reflected wave (assuming no depolarization). The plane of incidence contains both incident and reflected wave propagation vectors (see Figure 2), the latter being normal to the planes of constant phase for the two waves. In the reflection process, the key parameters that characterize the interface are the complex amplitude reflection coefficients, r_p and r_s. These coefficients are defined separately for the p- and s-directions as the ratio of reflected and incident electric field components by Equations (14–17):[5]

$$r_\mathrm{p} = |r_\mathrm{p}| \exp(\mathrm{i}\delta_\mathrm{p}) \equiv \frac{(E_{0\mathrm{p}})_\mathrm{r}}{(E_{0\mathrm{p}})_\mathrm{i}} \quad (14)$$

$$= \frac{\varepsilon_\mathrm{s} N_{\mathrm{a}\perp} - \varepsilon_\mathrm{a} N_{\mathrm{s}\perp}}{\varepsilon_\mathrm{s} N_{\mathrm{a}\perp} + \varepsilon_\mathrm{a} N_{\mathrm{s}\perp}} \quad (15)$$

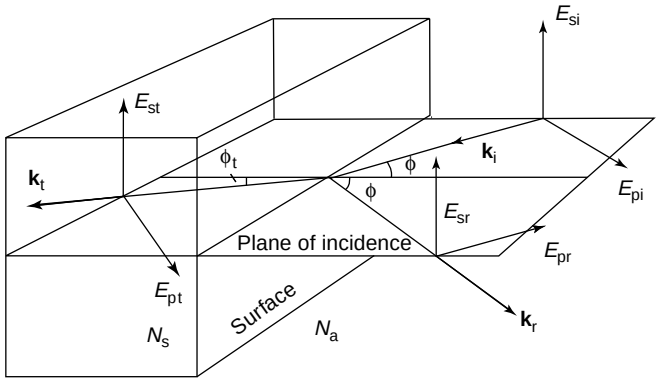

Figure 2 Electric field vectors resolved into p and s components, shown in the positive sense, for light beams incident (i), reflected (r), and transmitted (t) at an interface between an ambient and a substrate The propagation vectors for the beams are labeled $\mathbf{k}_i$, $\mathbf{k}_r$, and $\mathbf{k}_t$.

$$r_s = |r_s| \exp(i\delta_s) \equiv \frac{(E_{0s})_r}{(E_{0s})_i} \tag{16}$$

$$= \frac{N_{a\perp} - N_{s\perp}}{N_{a\perp} + N_{s\perp}} \tag{17}$$

where

$$N_{a\perp} \equiv N_a \cos\phi_i \tag{18}$$

$$N_{s\perp} \equiv N_s \cos\phi_t = (\varepsilon_s - \varepsilon_a \sin^2\phi_i)^{1/2} \tag{19}$$

$N_a = n_a - ik_a$ and $N_s = n_s - ik_s$ are the complex indices of refraction of the media containing the incident wave (a, ambient) and the transmitted wave (s, substrate), respectively, and ε_a and ε_s are the dielectric functions. ϕ_i and ϕ_t are the angles of incidence and transmission (see Figure 2), related by Snell's law. In Equations (14) and (16), δ_p and δ_s are the phase changes that occur upon reflection, and $|r_p|$ and $|r_s|$ are the ratios of the amplitudes for the complex fields associated with the reflected and incident waves. Figure 2 depicts positive sign conventions for the p and s electric field components. The different phase and amplitude changes for the p and s components lead to a change in the polarization state of the wave upon reflection, and it is this change that is measured by ellipsometry. Equations (14) and (16) can be applied to surfaces with multiple interfaces whereas Equations (15) and (17–19) are valid only for a single interface.

Assuming no depolarization by the sample, the polarization states of the incident and reflected waves are specified by the complex ratios $[(E_{0p})_i/(E_{0s})_i]$ and $[(E_{0p})_r/(E_{0s})_r]$, respectively.[1] In ellipsometry, the change in polarization state upon reflection is measured as a ratio of reflected and incident polarization states. Thus, for a single interface between isotropic media, the object of conventional reflection ellipsometry is ρ, the complex amplitude reflection ratio given by Equations (20–22):[5]

$$\rho \equiv \frac{r_p}{r_s} = \frac{(E_{0p})_r (E_{0s})_i}{(E_{0p})_i (E_{0s})_r} = \frac{|r_p|}{|r_s|} \exp[i(\delta_p - \delta_s)] \tag{20}$$

$$\equiv \tan\psi \exp(i\Delta) \tag{21}$$

$$= \frac{\sin^2\phi_i - [(\varepsilon_s/\varepsilon_a) - \sin^2\phi_i]^{1/2} \cos\phi_i}{\sin^2\phi_i + [(\varepsilon_s/\varepsilon_a) - \sin^2\phi_i]^{1/2} \cos\phi_i} \tag{22}$$

Here ψ and Δ are the ellipsometry angles that are commonly used to represent the interaction of the incident wave with the reflecting interface. Although Equations (20) and (21) can be applied to isotropic surfaces having any number of interfaces, Equation (22) is valid for a single interface only.

When one or both media are anisotropic, incident electric field components in either the p or s directions lead to components of the reflected wave in both directions. Thus, in general, four complex amplitude reflection coefficients, r_{pp}, r_{ss}, r_{ps}, and r_{sp}, can characterize reflection at such an interface.[1] Ellipsometric measurements for different incident polarization states, angles of incidence, or surface orientations can yield the three complex amplitude reflection ratios.[25]

The simplest system measurable by reflection ellipsometry is the ideal interface between two media. An example is an opaque substrate that presents to the ambient an atomically smooth abrupt termination of the bulk structure. Real systems inevitably deviate from this mathematical idealization owing to the presence of surface roughness and/or native oxide or other contaminant layers. Even if atomically smooth clean surfaces could be generated and measured in situ under UHV, surface reconstruction reorders the top monolayer or more, resulting in a measurable difference in the optical response of the monolayer(s) compared with the bulk. Ignoring such problems for the moment, the dielectric function of the substrate ε_s can be deduced from ψ and Δ if the dielectric function of the ambient ε_a and angle of incidence ϕ_i are known (Equation 23):[5]

$$\varepsilon_s = \varepsilon_a \sin^2\phi_i \left\{1 + \tan^2\phi_i \frac{(1-\rho)^2}{(1+\rho)^2}\right\} \tag{23}$$

For a two-medium structure with $|\varepsilon_s/\varepsilon_a| \geq 1$, ψ and Δ are constrained to $0^\circ \leq \psi \leq 45^\circ$ and $0^\circ \leq \Delta \leq 180^\circ$, within the sign convention adopted here. The right side of Equation (23) is often applied to experimental data in ρ when no effort has been taken to ensure that the sample is a two-medium structure. The left side is then called the pseudodielectric function, written as $\langle\varepsilon\rangle = \langle\varepsilon_1\rangle - i\langle\varepsilon_2\rangle$. $\langle\varepsilon\rangle$ is even used for samples known to consist of multiple layers. In this case, $\langle\varepsilon\rangle$ provides an alternative to (ψ, Δ) for describing interaction of the incident beam with

the surface, but does not convey any direct physical information.

Two angles of incidence are of interest in the ellipsometry experiment for a single interface.[5] The first is the pseudo-Brewster angle ϕ'_B at which $\tan\psi$ and the p-component reflectance $|r_p|^2$ are minimized. This angle is identical to the Brewster angle $\phi_B = \tan^{-1}(n_s/n_a)$ when $k_a = k_s = 0$. The second angle of incidence of interest is the principal angle ϕ_P at which $\Delta = 90^\circ$. The principal angle is usually within $\sim 1^\circ$ of the pseudo-Brewster angle for absorbing materials and is equal to it for transparent materials. The relative uncertainty in ε_s, $\delta\varepsilon_s/\varepsilon_s$, owing to uncertainties ψ and Δ is minimized at ϕ_P. For this reason, if one is interested in measuring the dielectric function ε_s of a material, then the ellipsometer should be operated with $\phi_i = \phi_P$. Under these optimum conditions, experimental fluctuations of $\delta\psi \cong 0.001^\circ$ and $\delta\Delta \cong 0.002^\circ$, achievable with an automatic ellipsometer, lead to relative uncertainties in ε_s of 1×10^{-4}. In section 3.3, it will be shown that sample surface preparation, not ellipsometer precision, limits the ability to deduce ε_s accurately. Precision is paramount, however, if SE data are to be differentiated with respect to the photon energy in order to characterize the composition or temperature of a crystalline solid based on the positions of the critical points in the electronic density of states (see section 5.3).

3.3 Ellipsometry Measurement of a One-layer Structure

The p and s complex amplitude reflection coefficients, defined as the ratio of the reflected and incident electric field components in the p and s directions, can also be used to characterize a system consisting of two or more interfaces. For a one-layer sample consisting of two ideal atomically smooth and abrupt interfaces and three different media (ambient/film/substrate), the reflection coefficients denoted r_{p2} and r_{s2} are given by:[1,5]

$$r_{p2} = \frac{r_{p,af} + r_{p,fs}Z_f}{1 + r_{p,af}r_{p,fs}Z_f} \tag{24}$$

$$r_{s2} = \frac{r_{s,af} + r_{s,fs}Z_f}{1 + r_{s,af}r_{s,fs}Z_f} \tag{25}$$

where

$$Z_f = \exp\left\{\frac{4\pi i d_f}{\lambda}(\varepsilon_f - \varepsilon_a \sin^2\phi_i)^{1/2}\right\} \tag{26}$$

The ellipsometry angles for the one-layer sample are expressed as:

$$\rho_2 \equiv \frac{r_{p2}}{r_{s2}} \equiv \tan\psi \exp(i\Delta) \tag{27}$$

In Equations (24–26), $\{r_{p,af}, r_{s,af}\}$ and $\{r_{p,fs}, r_{s,fs}\}$ are the p and s reflection coefficients for the ambient/film (af) and film/substrate (fs) interfaces, obtained through expressions analogous to Equations (15) and (17–19). In addition, $\varepsilon_f = \varepsilon_{1f} - i\varepsilon_{2f}$, d_f, and λ, are the dielectric function of the film, the thickness of the film, and the wavelength of the incident light, respectively.

The basic data analysis problem in the characterization of thin films by ellipsometry involves determining ε_{1f}, ε_{2f}, and d_f from the measured (ψ, Δ) values using the presumed known values of ε_a, ε_s, ϕ_i, and λ.[1,5] Because three thin film parameters cannot be uniquely deduced from only two measured interaction parameters, multiple measurements of (ψ, Δ) are needed. These can be obtained at different λ or ϕ_i, or under certain limitations for different ambients, substrates, or film thicknesses. One of these limitations requires that ε_f does not depend on the ambient, substrate, or thickness. Even for the simpler case of a nonabsorbing film with $\varepsilon_{2f} = 0$, Equations (24–27) cannot be inverted to provide a unique solution for d_f from a single measurement of (ψ, Δ). The periodicity of Z_f in Equation (26) shows that if $d_{f,0}$ is the minimum solution, then Equation (28)

$$d_{f,j} = d_{f,0} + \frac{m\lambda}{2(\varepsilon_f - \varepsilon_a \sin^2\phi_i)^{1/2}} \tag{28}$$

also provides solutions for all integers $m \geq 1$. In addition, as there are no closed-form expressions for either ε_{1f} or $d_{f,0}$, determining these parameters from (ψ, Δ) often involves Newton's method of numerical inversion, although more efficient algorithms are also available.[2] For a transparent film, (ψ, Δ) values at two wavelengths, λ_0 and λ_1, can be inverted to obtain $\varepsilon_{1f}(\lambda_0)$ and $\varepsilon_{1f}(\lambda_1)$ and two families of solutions for d_f. The correct d_f is the value common to the two families. This approach has entrenched ellipsometry as the method of choice for determination of thickness and index of refraction of dielectrics on semiconductors for use in microelectronics technology.

For semiconductor films that are nonabsorbing at λ_0 and λ_1, corresponding to photon energies $E_0 = hc/\lambda_0$ and $E_1 = hc/\lambda_1$ below the band gap, and absorbing at λ_2, with $E_2 = hc/\lambda_2$ above the gap, the approach of the previous paragraph can be used to deduce $\varepsilon_{1f}(\lambda_0)$, $\varepsilon_{1f}(\lambda_1)$, and d_f from (ψ, Δ) at λ_0 and λ_1. Then, numerical inversion can be applied to deduce $\varepsilon_{1f}(\lambda_2)$ and $\varepsilon_{2f}(\lambda_2)$ from (ψ, Δ) at λ_2, using the known value of d_f. Thus, a light source with the ability to vary λ is a benefit for the analysis of discrete data. Other approaches for solving the one-layer problem with discrete data on absorbing films involve measurements at multiple incidence angles[1,26] and combined ellipsometry and reflectance.[2,27] In fact, the most successful method for characterizing absorbing

films is with SE.[5] SE provides an additional advantage in that ε_f is obtained as a continuous function of λ. More advanced analysis techniques for nonabsorbing and absorbing films that exploit SE and RTSE data will be described in sections 4.5 and 5.2.

The complex amplitude reflection ratio of Equation (27) can be expressed as a power series in d_f, and for d_f less than about 1 nm all terms beyond first order can be neglected as described by Equation (29):[5]

$$\rho_2 \cong \rho \left\{ 1 + \left(\frac{4\pi i d_f n_a \cos\phi_i}{\lambda} \times \left(\frac{\varepsilon_s(\varepsilon_f - \varepsilon_a)(\varepsilon_s - \varepsilon_f)}{\varepsilon_f(\varepsilon_s - \varepsilon_a)(\varepsilon_s \cot^2\phi_i - \varepsilon_a)} \right) \right\} \tag{29}$$

Here ρ is the bare substrate amplitude reflection ratio from Equation (22). The result of Equation (29) shows that fluctuations of $\delta\psi \cong 0.001^\circ$ and $\delta\Delta \cong 0.002^\circ$ achievable with automatic instruments correspond to $\delta d_f \cong 0.001$ nm. Thus, minute changes in monolayer coverage of a surface can be detected with an automatic ellipsometer. This demonstrates that the sample surface condition rather than the ellipsometer precision is a controlling factor in the accurate determination of ρ and thus ε_s. Equation (29) also demonstrates, for ultrathin dielectric film formation on a dielectric substrate, that ψ remains constant and only Δ changes. Thus, a single interaction parameter is obtained and it is impossible to deduce both ε_{1f} and d_f.

3.4 Ellipsometry Measurement of a Multilayer Structure

The analogs of Equations (14–19) can be derived for an m-layer system using matrix methods to perform the required manipulations.[1,28] The sample structure is shown in Figure 3. The method consists of determining the 2×2 scattering matrix S_v (v designating either p or s) as a product of $(m+1)$ interface matrices $I_{v,j,(j+1)}$ $(j = 0, \ldots, m)$ and m layer matrices L_j $(j = 1, \ldots, m)$, according to Equation (30)

$$S_{v,(m+1)} = I_{v,01} L_1 I_{v,12} L_2 \ldots L_j I_{v,j,(j+1)} L_{j+1} \ldots L_m I_{v,m,(m+1)} \tag{30}$$

Here the subscript "0" designates the ambient medium from which the light wave is incident and "$m+1$" designates the substrate medium. The two matrices are given by Equations (31) and (32):

$$I_{v,j,(j+1)} = \frac{1}{t_{v,j,(j+1)}} \begin{matrix} 1 & r_{v,j,(j+1)} \\ r_{v,j,(j+1)} & 1 \end{matrix} \tag{31}$$

$$L_j = \begin{matrix} Z_j & 0 \\ 0 & Z_j^{-1} \end{matrix} \tag{32}$$

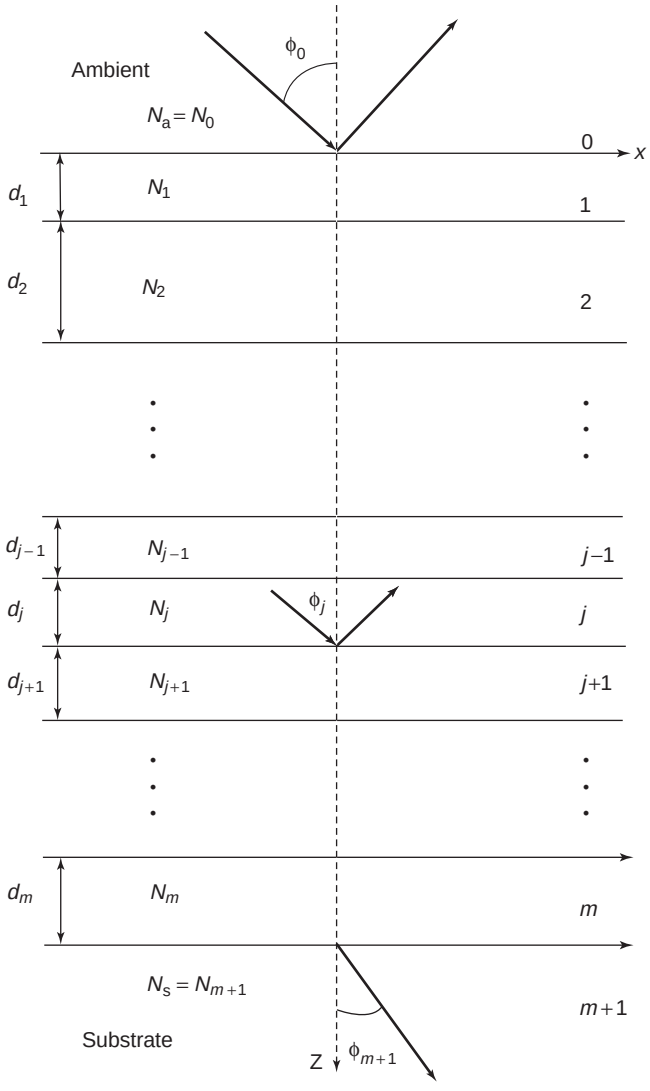

Figure 3 Schematic diagram of an m-layer structure. N_a, N_j, and N_s are the complex indices of refraction for the ambient, the jth layer of thickness d_j, and the substrate, respectively. $\phi_i = \phi_0$ is the angle of incidence on the first layer and ϕ_j is the angle of incidence at the interface between the jth and $(j+1)$st layers.

Here $r_{v,j,(j+1)}$ and $t_{v,j,(j+1)}$ are the amplitude reflection and transmission coefficients for the single interface between media j and $j+1$, respectively. The transmission coefficients are defined in the same way as the reflection coefficients, but as the ratio of transmitted and incident electric field components across one or more interface (see Equations 14 and 16). For the specific case of a single interface between media j and $j+1$, these coefficients are given by Equations (33) and (34)

$$t_{p,j,(j+1)} = \frac{2N_j N_{j+1} N_{j,\perp}}{\varepsilon_{j+1} N_{j,\perp} + \varepsilon_j N_{j+1,\perp}} \tag{33}$$

$$t_{s,j,(j+1)} = \frac{2N_{j,\perp}}{N_{j,\perp} + N_{j+1,\perp}} \tag{34}$$

Equations (35) and (36) where $N_{j,\perp}$ and $N_{j+1,\perp}$ are given by

$$N_{j,\perp} \equiv N_j \cos\phi_j \tag{35}$$

$$N_{j+1,\perp} \equiv N_{j+1}\cos\phi_{j+1} = (\varepsilon_{j+1} - \varepsilon_j \sin^2\phi_j)^{1/2} \tag{36}$$

The angle of incidence for the interface between media j and $j+1$, denoted ϕ_j is obtained from $\phi_i \equiv \phi_0$ through successive application of Snell's law, Equation (37):

$$N_0 \sin\phi_0 = N_1 \sin\phi_1 = \cdots = N_j \sin\phi_j = N_{j+1}\sin\phi_{j+1} = \cdots = N_m \sin\phi_m = N_{m+1}\sin\phi_{m+1} \tag{37}$$

Finally, the parameter Z_j for the jth layer is given by Equation (38)

$$Z_j = \exp\left\{\frac{2\pi i d_j}{\lambda} N_{j,\perp}\right\} \tag{38}$$

where d_j is the thickness of layer j. Once the scattering matrix is determined, then the complex amplitude reflection coefficients, $r_{p,(m+1)}$ and $r_{s,(m+1)}$, and (ψ, Δ) can be determined from Equations (39–41):

$$r_{p,(m+1)} = \frac{[S_{p,(m+1)}]_{21}}{[S_{p,(m+1)}]_{11}} \tag{39}$$

$$r_{s,(m+1)} = \frac{[S_{s,(m+1)}]_{21}}{[S_{s,(m+1)}]_{11}} \tag{40}$$

$$\rho_{m+1} \equiv \frac{r_{p,(m+1)}}{r_{s,(m+1)}} \equiv \tan\psi \exp(i\Delta) \tag{41}$$

Such computations are performed by computer from input values of λ, ϕ_0, $\{N_0, \ldots, N_{m+1}\}$, and $\{d_1, \ldots, d_m\}$. The matrix equations can sometimes be inverted numerically to deduce one unknown complex index and one or more thicknesses if all other parameters are known. To deal with problems of this nature continuous SE data are required at least. If the unknown complex index can be expressed in terms of λ-independent parameters, then least squares regression analysis can be applied as will be described in section 4.5.

4 STATIC MEASUREMENT APPLICATIONS

Ellipsometry applications fall into two broad categories, namely, those in which measurements on a static sample yield its properties in a single stable state and those in which measurements on a dynamic sample surface yield the time evolution of its properties. For the former applications, the static sample can be either inside a controlled environment used for sample preparation or removed from the environment (i.e. in situ or ex situ ellipsometry, respectively). For measurements of dynamic surfaces, the term ''RTE'' is often used. It is important to note that ''in situ'' does not necessarily imply ''real time''; however, ''real time'' does imply ''in situ''. In general, ellipsometry measurements requiring several seconds or less to determine the interaction parameters are compatible with real time applications, whereas those that require a minute or more are not. Because most ellipsometry experiments are now performed in a multiple-measurement mode, usually continuously versus wavelength (i.e. SE), the instrument designs optimized for measurements on static and dynamic surfaces are notably different.[2] In the most common case of SE from the ultraviolet to near-infrared, measurements on static surfaces are performed wavelength-by-wavelength using a white light source and monochromator with a detection system consisting of a single PMT. In contrast, SE measurements on dynamic surfaces are best performed in parallel for a range of wavelengths simultaneously using a white-light source with a detection system consisting of a spectrograph and rapid scanning photodiode array.[18] Because of these design differences, the discussion of instruments and applications is divided into two sections. In this section ellipsometry on static surfaces is described and in section 5 RTE and RTSE are described.

4.1 Instrument Designs, Operation, and Performance

It is instructive to consider the generic ellipsometer configuration in two parts as shown in Figure 4.[1] The polarization generator part consists of a light source, focusing optics, a monochromator, collimating optics, a polarizer, and a compensator. The polarizer passes the electric field component of the incident light wave that is aligned along a specific axis of the device called the transmission axis, whereas it removes the electric field component of the wave that is aligned along the orthogonal extinction axis. The transmission axis, extinction axis, and propagation direction of the wave form a right-handed coordinate system (see Figure 4). The polarizer is inserted into a mounting that allows it to be rotated about an axis parallel to the propagation

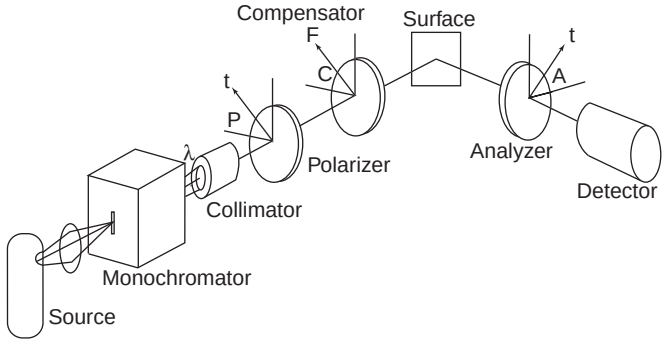

Figure 4 Configuration of a generic reflection ellipsometer.

direction. Thus, the transmission axis can be set at any angle P, measured with respect to the plane of incidence in a counterclockwise-positive sense facing the source. The compensator is a device that imposes a phase shift δ_C on the electric field component along the slow axis of the device, relative to the component along the orthogonal fast axis. In compensators that employ birefringent crystals, δ_C is approximately inversely proportional to λ. The compensator is also inserted into a mounting that allows it to be oriented such that the fast axis is tilted at any angle C, measured with respect to the plane of incidence. When used in transmission with an incident collimated quasimonochromatic light beam, the polarizer and compensator can generate light in any desired polarization state. If the compensator is set for quarterwave operation so that $\delta_C = \pm 90^\circ$, for example, and a polarization ellipse with ellipticity angle χ and tilt angle Q is desired, then the polarizer and compensator angles are set at $P = Q + \chi$ and $C = Q$.

The sample is mounted between the polarization generator and polarization detector parts of the instrument. The polarization detector consists of a second polarizer called an analyzer and a light detector. This combination allows the change in polarization imparted by the sample to be deduced. The analyzer is also rotatable such that the angle of the transmission axis with respect to the plane of incidence is A. In this configuration, the field amplitude transmitted by the analyzer is:[1,5]

$$\begin{aligned} E_{0A} = E_{0P} r_s t_F [&\rho \cos C \cos(P - C) \cos A - \rho \rho_C \sin C \\ &\times \sin(P - C) \cos A + \sin C \cos(P - C) \sin A \\ &+ \rho_C \cos C \sin(P - C) \sin A] \end{aligned} \quad (42)$$

The detected irradiance is calculated as $I = (1/2) c \varepsilon_0 |E_{0A}|^2$, where ε_0 is the free space permittivity. In Equation (42), E_{0P} is the field amplitude at the polarizer output. The sample parameters r_s and ρ, are given by Equations (16), (20), and (21), and the compensator parameters are given by Equation (43):

$$\rho_C = \frac{t_S}{t_F} = \frac{E_{0CS} E_{0PF}}{E_{0PS} E_{0CF}} \equiv T_C \exp(i\delta_C) \cong \exp(i\delta_C) \quad (43)$$

t_S and t_F are the transmission coefficients for field components along the compensator slow and fast axes, which are defined as ratios of the transmitted component, E_{0Cj} (j = F, S) to the incident component, E_{0Pj} (j = F, S). Compensators can exhibit dichroism; thus T_C can deviate from unity.

In null or compensating ellipsometry, ρ is determined by adjusting the polarizer and analyzer angles P and A to obtain the condition $I = 0$.[1] In photometric ellipsometry, ρ is determined by measuring the time dependence of the detector irradiance $I(t)$ that results when one of the angles, P, C, A, or δ_C is continuously varied either linearly or sinusoidally with time. Equation (42) describes the ideal operation of the ellipsometer of Figure 4 for an isotropic reflecting surface. Calibration procedures specific to each type of instrument must be employed to determine corrections to the nominal scale values read from the device mountings (or the phases of the angular variations) for the polarizer, compensator, and analyzer. In addition, the compensator must also be calibrated to determine ρ_C for all measurement wavelengths. For accurate operation, correction terms may need to be included in Equation (42) that account for imperfect optical elements. For example, polarizers and compensators may exhibit strain birefringence or optical activity, or the sample may be enclosed in an optical cell whose windows are slightly birefringent. Ellipsometer alignment is also important for achieving high accuracy.[29] For proper alignment, the optic and rotational axes of the components on each of the source and detector sides of the instrument must all be coincident. The resulting axes on the two sides must lie in the plane of incidence and must intersect one another on the surface of the sample. The light beam should be well collimated and follow the path defined by the intersecting optical axes. The angle of incidence can be determined by autocollimation of reflections from the surfaces of a metallized prism at the sample position.[24]

4.1.1 Manual Null Ellipsometers

The manual null ellipsometer was the most popular instrument in the period from 1950 to 1970, before the widespread use of laboratory computers for instrument automation. The applications of this instrument today include (i) single wavelength measurement for determination of the thickness and index of refraction of a nonabsorbing layer on a substrate of known N_s using a one-layer model, and (ii) assessment of the accuracy of photometric instruments by measurement of the same sample by photometric and four-zone null ellipsometry.

For the manual null ellipsometer used in the configuration of Figure 4, the polarizer angle, P, is adjusted so that the elliptical polarization state generated by the compensator is converted back into a linear state by the sample. Then the condition, $I = 0$, can be achieved by adjustment of the analyzer angle, A. To achieve high precision, the polarizer and analyzer are alternately rotated for minimum detector signal until the null is reached. The sample characteristic ρ is obtained in terms of the angles at null, P_n and A_n, Equation (44):[1]

$$\rho = -\tan A_n \frac{\tan C + \rho_C \tan(P_n - C)}{1 - \rho_C \tan C \tan(P_n - C)} \quad (44)$$

Null ellipsometers are operated with a quarterwave compensator setting ($\delta_C = 90^\circ$). For an ideal compensator ($\rho_C = \mathrm{i}$) with $C = 45^\circ$, either of two inequivalent polarizer and analyzer null settings, (P_2, A_2) and (P_4, A_4), can establish (ψ, Δ). The two nulls are designated "zone 2" and "zone 4" according to Equations (45–48):

$$\psi = A_2 \geq 0 \tag{45}$$

$$\Delta = 360^\circ m - 90^\circ - 2P_2; \quad m = 0 \text{ for } -90^\circ < P_2 < 45^\circ; \quad m = 1 \text{ for } 45^\circ \leq P_2 \leq 90^\circ \tag{46}$$

$$\psi = -A_4 \geq 0 \tag{47}$$

$$\Delta = -360^\circ m + 90^\circ - 2P_4; \quad m = 0 \text{ for } -45^\circ \leq P_4 \leq 90^\circ; \quad m = 1 \text{ for } -90^\circ < P_4 < -45^\circ \tag{48}$$

Zone 1 and zone 3 nulls, (P_1, A_1) and (P_3, A_3), are obtained with $\delta_C = 90^\circ$ and $C = -45^\circ$. (ψ, Δ) are obtained from the zone 1 null through Equations (45) and (46), but with an overall sign reversal on the right side of Equation (46) for Δ. (ψ, Δ) are obtained from the zone 3 null through Equations (47) and (48) with the analogous sign reversal. Equations (45–48) use the ranges $0^\circ \leq \psi \leq 90^\circ$ and $-180^\circ < \Delta \leq 180^\circ$, and equivalent nulls outside $-90^\circ < P \leq 90^\circ$ and $-90^\circ < A \leq 90^\circ$ are not considered.

Among different designs, the null ellipsometer can provide the highest accuracy, independent of the values of (ψ, Δ). When (ψ, Δ) are obtained as a four-zone average, they are free of errors owing to a variety of component and sample imperfections.[1] In addition, the null measurements are influenced neither by detector nonlinearity, as are photometric measurements, nor by source polarization and detector polarization sensitivity. The technique of azimuth modulation, whereby the null angles, (P_n, A_n), are determined by averaging angles of equal signal level on opposite sides of the null, has been applied to enhance the precision of the method. Under usual operating conditions with an Xe arc lamp source, the precision of the manual null ellipsometer caused by shot noise and detector fluctuations is estimated to be $2\delta\psi \cong \delta\Delta \cong 0.01^\circ$, which is near the limit of the resolution of typical scales on the component mountings.[15] The primary disadvantage of the manual null instrument is the inordinate amount of patience and time required for the measurement (~10 min for single zone (ψ, Δ) by azimuth modulation). This is not a major concern if a single measurement on a static surface is desired. More often, however, measurements at many wavelengths and/or angles of incidence or data on dynamic surfaces are desired, and automation is essential to reduce the effort and increase the speed.

4.1.2 Automatic Null Ellipsometers

The null ellipsometer can be automated by attaching stepping motors to the polarizer and analyzer angular drives, allowing alternate adjustment of the angles P and A under computer control.[30] Azimuth modulation can provide an error signal that directs the controller to the null settings. In this way, the measurement time can be reduced to ~1 s. Azimuth modulation can also be performed with Faraday cells, which are placed just after the polarizer and just before the analyzer as shown in Figure 5.[31] In a Faraday cell, a magnetic field H is applied to a transparent core material parallel to the direction of propagation of a

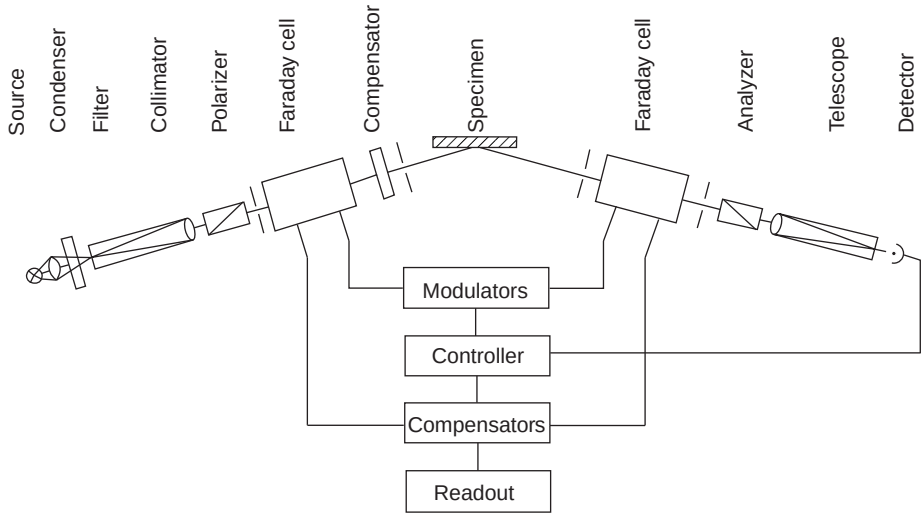

Figure 5 Simplified schematic of the optical and electronic components for an automatic nulling ellipsometer that uses Faraday cells for modulation and compensation. (Reproduced by permission of the Society of Photo-Optical Instrumentation Engineers from R. Muller, *Proc. Soc. Photo-Opt. Instrum. Eng.*, **112**, 68–73 (1977).[32])

linearly polarized beam that passes through the material. The electric field vector for this beam is rotated by an angle proportional to the magnetic field (i.e. $\Delta Q \propto H$). If the polarizer and analyzer Faraday cells are driven by ac currents at w_P and w_A and $|r_p| \cong |r_s| \cong 1$, then the detector irradiance near null can be written as:[5]

$$I(t) \cong I_0\{(A - A_n)A_m \sin(w_A t + \delta_A) + (P - P_n) \times P_m \sin(w_P t + \delta_P)\} \quad (49)$$

In Equation (49), $P_m \sin(w_P t + \delta_P)$ and $A_m \sin(w_A t + \delta_A)$ represent the polarizer and analyzer azimuth modulations. The terms in $I(t)$ serve as error signals that can be detected and used to direct the system to null with stepping motors on the polarizer and analyzer. Higher speed and precision can be obtained using Faraday cells both for modulation and polarization rotation. The instrument in Figure 5 uses concentric windings driven by high frequency (10 kHz) modulation and low frequency (1 kHz) nulling currents, the latter providing angular readout. Nulling is obtained in about 1 ms, more than 10^5 times faster than the manual instrument.[31] The ultimate precision of automatic null ellipsometers that use Faraday cells is an order of magnitude better than the manual instruments, with $2\delta\psi \cong \delta\Delta \cong 0.001°$, and higher than any other automatic instrument.[5] Because the automatic instruments are generally not designed to measure nulls in multiple zones, the measurements are less accurate than four zone-averaged null measurements.

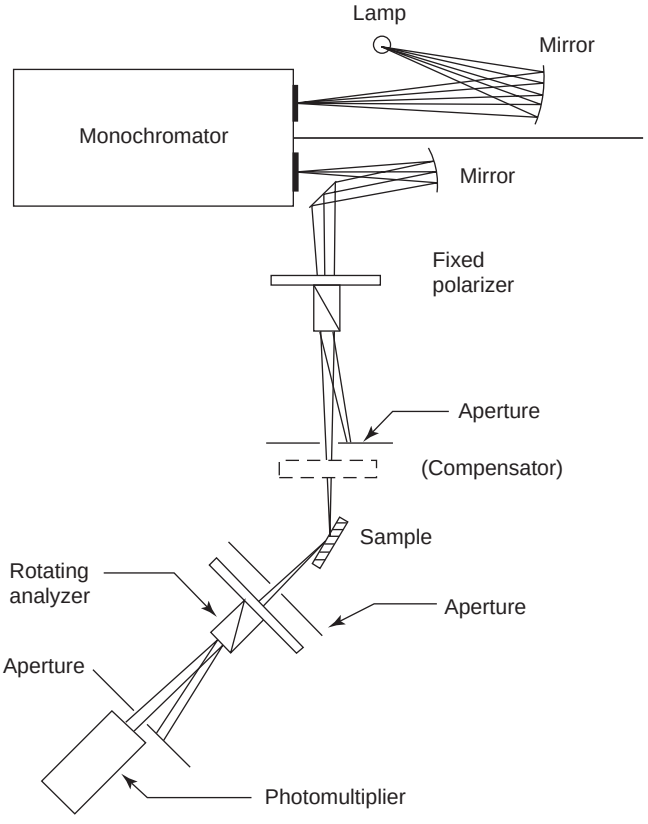

Figure 6 Schematic diagram of optical elements for a rotating analyzer spectroscopic ellipsometer. (Reproduced by permission of the Optical Society of America from D.E. Aspnes, A.A. Studna, *Appl. Opt.*, **14**, 220–228 (1975).[33])

4.1.3 Rotating Element Ellipsometers

Rotating element ellipsometers are constructed so that either the polarizer, compensator, or analyzer is mechanically rotated at frequency w about an axis coincident with the propagation direction of the beam.[33,34] For the rotating analyzer and polarizer instruments, the compensator is optional. Owing to its polarization sensitivity, the monochromator is usually placed before the fixed polarizer in the rotating analyzer configuration (see Figure 6) and after the fixed analyzer in the rotating polarizer configuration. For both configurations, the detector irradiance obeys Equation (50):

$$I(t) = I_0\{1 + \alpha \cos 2(wt + \varphi) + \beta \sin 2(wt + \varphi)\} \quad (50)$$

where $wt + \varphi$ represents the time dependence of the polarizer or analyzer angle, φ is a phase angle determined in calibration, and α and β are normalized $2w$ Fourier coefficients. The coefficients are deduced by Fourier analysis of the detector signal which is sampled, using the output of an encoder mounted on the rotating element as a trigger. For the rotating analyzer instrument with $C = 0°$, α and β are related to ψ and Δ through Equations (51) and (52):

$$\tan\psi = \left\{\frac{1+\alpha}{1-\alpha}\right\}^{1/2} \tan P \quad (51)$$

$$\cos(\Delta - \delta_C) = \frac{\beta}{(1-\alpha^2)^{1/2}} \quad (52)$$

Without the compensator, $\delta_C = 0$ in Equation (52). For the rotating polarizer instrument, P in Equation (51) is replaced by A. For typical rotation rates $w/2\pi$ in the range of 10–100 Hz, a single measurement of (ψ, Δ) can be obtained in 5–50 ms, i.e. from a single half rotation.

The advantage of the rotating polarizer and analyzer instruments is that in the absence of the compensator, the sample is the only λ-dependent element.[34] Thus, it requires no calibrations versus λ and can be designed for a wide spectral range. Because the instrument measures $\cos(\Delta - \delta_C)$, the sign of the cosine argument is indeterminate, and for samples with one or more layers where $-180° < \Delta \leq 180°$ the ambiguity must be resolved in data analysis. The precision of the instrument depends on the reflecting surface, with an optimum for the detection of circularly polarized light ($\alpha = \beta = 0$). A compensator

is needed when $\Delta \cong 0^\circ$ or $\pm 180^\circ$ (e.g. bulk dielectrics) because of the poor sensitivity in measuring linear polarization. The theoretical precision of the rotating analyzer instrument is somewhat poorer than the modulated null instrument. Detector nonlinearity is one source of inaccuracy for rotating element ellipsometers. For the rotating polarizer instrument the source must be unpolarized, whereas for the rotating analyzer the detector must be polarization insensitive. Because the latter can be achieved more readily, the rotating analyzer ellipsometer is more accurate. Comparisons of measurements by rotating analyzer and manual four-zone null techniques show that the former can achieve an accuracy better than 0.1° relative to the latter.

In the rotating compensator ellipsometer, the fixed polarizer is mounted before the sample, and the rotating compensator and fixed analyzer are mounted after the sample.[35,36] Several advantages of this design over the simpler rotating analyzer design justify the added complexity of the required λ-dependent calibration for ρ_C. These include the ability to (i) determine the sign of Δ, (ii) obtain accurate Δ values for nearly linear polarization states, and (iii) deduce the degree of polarization p of the light beam reflected from the sample. This latter parameter assists in characterizing nonideal samples having macroscopic nonuniformities such as roughness or thickness gradients that generate a distribution of polarization states in the beam reaching the detector. For the rotating compensator system, the detector irradiance obeys Equation (53):

$$I(t) = I_0\{1 + \alpha_2 \cos 2(wt + \varphi) + \beta_2 \sin 2(wt + \varphi) + \alpha_4 \cos 4(wt + \varphi) + \beta_4 \sin 4(wt + \varphi)\} \quad (53)$$

where $wt + \varphi$ represents the time dependence of the compensator angle, φ is the phase angle, and $\{(\alpha_m, \beta_m); m = 2, 4\}$ are normalized $2w$ and $4w$ Fourier coefficients. These coefficients are deduced by Fourier analysis of the detector signal. From the Fourier coefficients, the three parameters $\{Q, \chi, p\}$ that characterize the normalized Stokes vector of the light beam just after reflection from the sample can be obtained by solving Equations (54–57):[37]

$$\alpha_2 = \frac{p \sin \delta_C \sin 2\chi \sin 2A}{1 + \{p \cos^2(\delta_C/2) \cos 2\chi \cos 2(A - Q)\}} \quad (54)$$

$$\beta_2 = \frac{-p \sin \delta_C \sin 2\chi \cos 2A}{1 + \{p \cos^2(\delta_C/2) \cos 2\chi \cos 2(A - Q)\}} \quad (55)$$

$$\alpha_4 = \frac{p \sin^2(\delta_C/2) \cos 2\chi \cos 2(A + Q)}{1 + \{p \cos^2(\delta_C/2) \cos 2\chi \cos 2(A - Q)\}} \quad (56)$$

$$\beta_4 = \frac{p \sin^2(\delta_C/2) \cos 2\chi \sin 2(A + Q)}{1 + \{p \cos^2(\delta_C/2) \cos 2\chi \cos 2(A - Q)\}} \quad (57)$$

For samples in which $p = 1$, i.e. no polarization occurs, (ψ, Δ) can be deduced from (Q, χ) using Equations (58–60):

$$\cos 2\psi = \frac{\cos 2P - \cos 2Q \cos 2\chi}{1 - \cos 2Q \cos 2\chi \cos 2P} \quad (58)$$

$$\sin \Delta = \frac{\sin 2\chi (\cos 2\psi \cos 2P - 1)}{\sin 2\psi \sin 2P} \quad (59)$$

$$\cos \Delta = \frac{\cos 2\chi \sin 2Q \, (1 - \cos 2\psi \cos 2P)}{\sin 2\psi \sin 2P} \quad (60)$$

The importance of the rotating compensator system is its ability to maintain high precision, independent of (ψ, Δ), a feature also common to the null ellipsometer. Its accuracy is also much improved over the rotating polarizer and analyzer systems when the reflected beam is nearly linearly polarized. In addition, the rotating-compensator system is not affected by errors caused by source polarization or detection system polarization sensitivity since both the polarizer and analyzer are fixed. The main difficulty of the rotating-compensator ellipsometer is in the development of a first-order compensator for error-free spectroscopic operation. Compensators constructed from quartz or MgF_2 biplates or mica sheets can exhibit oscillatory spectral artifacts due to plate misalignment and multiple reflections between parallel surfaces of the device.[38]

4.1.4 Phase Modulation Ellipsometers

For phase modulation ellipsometers, the fixed compensator in the generic ellipsometer design of Figure 4 is replaced by a birefringent phase modulator, e.g. a Pockels cell or silica piezobirefringent element (see Figure 7).[39–41] This provides a sinusoidal relative retardation δ between orthogonal field components given by $\delta(t) = \delta_m \sin w_m t$, where $w_m/2\pi$ is typically 50 kHz and δ_m is proportional to V_m/λ, where V_m is the modulator drive voltage. In practice, A and the polarizer angle relative to the phase modulator angle, $P - M$, are both set to 45°. Assuming no depolarization, this leads to the detector irradiance given by Equation (61):

$$I(t) = I_0\{1 + \sin 2\psi \sin \Delta \sin[\delta(t)] + (\sin 2M \cos 2\psi + \cos 2M \sin 2\psi \cos \Delta) \cos[\delta(t)]\} \quad (61)$$

Two different modulator angles, $M = 0^\circ$ and 45°, referred to as configurations II and III, respectively, have been used most often. To determine (ψ, Δ) from $I(t)$, the ratio of the w and $2w$ components to the dc component, R_w and R_{2w}, respectively, are obtained using a low pass filter (dc) and phase sensitive detection (w and $2w$). Alternatively, the detector output is sampled throughout the modulation period by an ADC, triggered by the

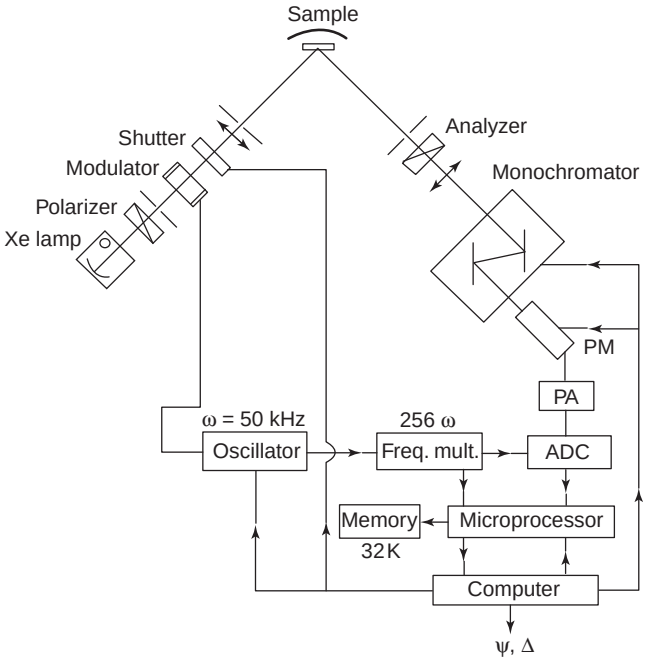

Figure 7 Schematic of the optical and electronic components for a phase modulation ellipsometer that employs Fourier analysis to extract the harmonic constituents of the detector output. PM, photomultiplier tube; PA, preamplifier; ADC, analog-to-digital converter. (Reproduced by permission of the American Institute of Physics from B. Drevillon, J. Perrin, R. Marbot, A. Violet, J.L. Dalby, *Rev. Sci. Instrum.*, **53**, 969–977 (1982).[40])

frequency-multiplied modulator reference as shown in Figure 7. The resulting raw data are Fourier analyzed in a scheme that permits the collection of a single (ψ, Δ) in a time of 20 μs. In configuration II, (ψ, Δ) are determined from Equations (62) and (63):

$$C_{\mathrm{II}} = \sin 2\psi \cos \Delta = \frac{R_{2w}}{2J_2(\delta_m) - R_{2w}J_0(\delta_m)} \tag{62}$$

$$S_{\mathrm{II}} = \sin 2\psi \sin \Delta = \frac{R_w J_2(\delta_m)}{J_1(\delta_m)\{2J_2(\delta_m) - R_{2w}J_0(\delta_m)\}} \tag{63}$$

Alternatively in configuration III, one can apply Equations (64) and (65):

$$N_{\mathrm{III}} = \cos 2\psi = \frac{R_{2w}}{2J_2(\delta_m) - R_{2w}J_0(\delta_m)} \tag{64}$$

$$S_{\mathrm{III}} = \sin 2\psi \sin \Delta = \frac{R_w J_2(\delta_m)}{J_1(\delta_m)\{2J_2(\delta_m) - R_{2w}J_0(\delta_m)\}} \tag{65}$$

where $J_n(\delta_m)$ is the nth order Bessel function. Best results are obtained when the modulator drive voltage is set to give $J_0(\delta_m) = 0$, whereupon $\delta_m = 138°$. To maintain this condition for SE, V_m must be increased linearly versus λ.

Because only $\sin 2\psi$ is measured in configuration II, both ψ $(0° \leq \psi \leq 90°)$ and $90° - \psi$ are solutions to Equations (62) and (63). Because only $\sin \Delta$ is measured in configuration III, both Δ $(-180° < \Delta \leq 180°)$ and $\pm 180° - \Delta$ are solutions to Equation (65). In configuration III with Δ near 90°, such as may occur for bulk metal surfaces, the precision and accuracy in Δ is poor. Measurements in both configurations solve these problems and, in addition, when depolarization occurs yield p:

$$p = (N_{III}^2 + C_{II}^2 + S_{II}^2)^{1/2} \tag{66}$$

The quantities in Equation (66) can also be obtained using a polarization state detector design that incorporates a Wollaston prism in place of the analyzer. In this design, the reflected beam is decomposed into orthogonal linearly polarized beams that are measured by two PMTs.

The precision of the phase modulation and rotating analyzer ellipsometers are comparable when measured under optimum conditions for both instruments.[5] The advantage of the former is its high speed when operated using the Fourier analysis scheme. A high speed is useful in a number of real time studies of reactions at interfaces, film growth at high rates, and ion, laser, and thermal processing of materials. Unfortunately, the precision of the instrument degrades rapidly under these measurement conditions and the monolayer sensitivity is lost. An advantage of the phase modulation ellipsometer over the rotating polarizer or analyzer systems is its immunity to source polarization and detector polarization sensitivity. The disadvantage of the phase modulation ellipsometer is its greater complexity. Nonideal behavior of the modulator, e.g. effects of multiple harmonics and static birefringence, imply that greater care is required to achieve high accuracy operation in comparison to the rotating polarizer and analyzer systems.

4.2 Spectral Range

The spectral range of an ellipsometer determines the physical and chemical information accessible from the measured dielectric function.[23] The most common ellipsometers operate from the ultraviolet to the near-infrared and provide information on the electronic transitions between states in the valence and conduction bands located within $\Delta E \sim \pm 5$ eV of the Fermi level.[5] The optical absorption and dispersion associated with these transitions allow the material density, phase (e.g. crystalline versus amorphous), composition, and temperature to be established, assuming that suitable databases exist. Ellipsometers operating over a wide spectral range of the infrared ($700–7000\,\mathrm{cm}^{-1}$) have increased in popularity recently.[6,42] The dielectric functions of materials in the infrared provide information on vibrational excitations that allow the chemical bonds to be identified at the surface or within the bulk of the materials. In addition, the dielectric functions in the infrared allow

free electron transitions to be characterized in conducting materials such as metals and doped semiconductors. Finally, specialized instruments utilizing synchrotron radiation sources have also been developed to span spectral ranges outside the ultraviolet to mid-infrared. With such instruments, materials have been measured in the far-infrared (20–800 cm^{-1}),[43,44] the vacuum ultraviolet (50–200 nm),[45] and the extreme ultraviolet (~10 nm).[46] In the far infrared, the dielectric functions of solids reveal excitations of vibrational modes, excitations across the energy gaps of superconductors, as well as collective excitations of charge and spin density waves. In the vacuum and extreme ultraviolet, the dielectric functions provide information on core electron transitions. For the different spectral ranges, the ellipsometer components, including the source, the polarizers, the compensator, and the spectrometer, must be chosen for optimum performance.

In the most common ultraviolet to near-infrared ellipsometer, a broad-band Xe arc lamp is used as the source.[33] This can provide useful output from 200–300 to 1000 nm. High quality polarizers for this range are constructed from calcite prisms with a usable lower wavelength limit of ~220 nm. Quartz or magnesium fluoride polarizers can extend the response to less than 200 nm. The problem with quartz polarizers is their optical activity. As a result, the polarization state after transmission through the polarizer is not linear, but exhibits a slight ellipticity which must be corrected in data reduction to achieve an accuracy better than 0.1°.[33,34] Birefringent compensators used in transmission over the ultraviolet to near-infrared range consist of pairs of fixed or adjustable (Babinet–Soleil) quartz or MgF_2 plates; a mica sheet can also be used, but with a restricted spectral range (300–800 nm). For such birefringent compensators, the phase shift δ_C is approximately inversely proportional to λ. Three-reflection compensators are also available made from vitreous silica with the advantage of having a nearly λ-independent δ_C.[47] Phase modulators for the ultraviolet to near-infrared range are also usually constructed from a vitreous silica element that is subjected to a periodic stress.[39–41] This device can cover the spectral range from 200 to 1000 nm providing that the drive voltage is adjusted to cancel the λ-dependence of the phase shift. The monochromators or spectrometers used in the ultraviolet to near-infrared employ either rotatable prisms, gratings, or combinations of the two in order to disperse the incident light. The most popular detector for this spectral range is the PMT, which is typically sensitive from 200 to 800 nm. Silicon photodiodes can extend the spectral range to 1000 nm in the near-infrared. For photometric ellipsometers, one must ensure that the detector remains within its linear range of operation throughout the wavelength scan.[33]

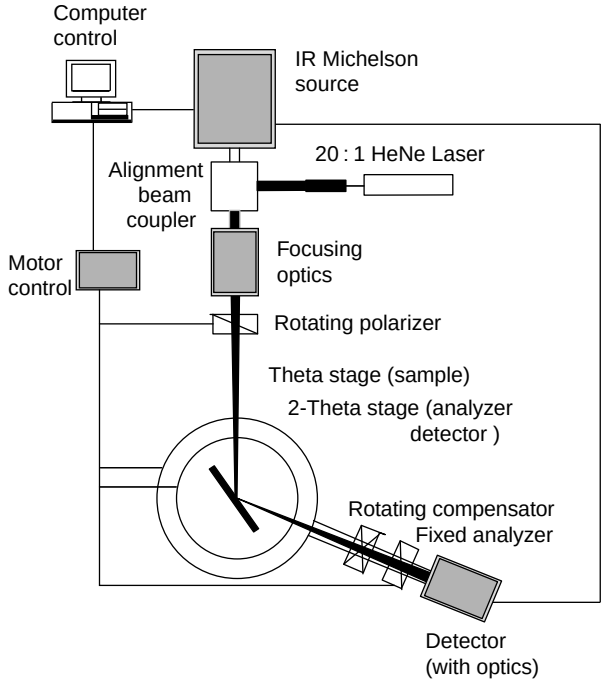

Figure 8 Schematic of a rotating element infrared spectroscopic ellipsometer that uses Fourier transform interferometry for signal processing. (Reproduced from T.E. Tiwald, D.W. Thompson, J.A. Woollam, S.V. Pepper, *Thin Solid Films*, **313–314**, 718–721. Copyright (1998) with permission of Elsevier Science.[48])

Ellipsometers operating in the mid-infrared range have been designed based on the rotating-element[6] and phase-modulation principles.[42] For both instruments, the radiation source is a SiC globar and Michelson interferometer for wavelength scanning based on Fourier transform principles, and the detector employs a cooled mercury cadmium telluride ($Hg_{1-x}Cd_xTe$) element for highest sensitivity. Figure 8 shows a typical schematic of a rotating-element infrared spectroscopic ellipsometer.[48] Because the frequency of wavelength scanning for the instrument in Figure 8 (~100 Hz) is comparable with the rotation frequency of the rotating elements (e.g. the polarizer and compensator), then the waveform $I(t)$ at the detector must be reconstructed from wavelength scans performed at selected settings P and C. This approach is in contrast to the rotating element instrument operating from the ultraviolet to near-infrared in which the optical elements are rotated continuously at a number of settings λ. In phase-modulation infrared SE (IRSE), however, the frequency of wavelength scanning is much lower than the phase modulation frequency.[42] As a result, both the interferometer and the modulator can be operated simultaneously, and the waveform of Equation (61) can be extracted using double modulation principles. Phase-modulation IRSE is claimed to be faster and thus more

suitable for real time characterization of film surfaces. Sensitivities at the partial monolayer level can be obtained in both cases with long averaging times. For example, fluctuations in Δ of $\delta\Delta \sim \pm 0.01^\circ$ have been measured in phase-modulation IRSE with a 15 min integration time.

The optical elements used in IRSE must be fabricated from materials with high infrared transmission. The most common polarizer consists of a ZnSe plate coated with a wire grid.[6] The problem with such a polarizer is its poor attenuation ratio, i.e. the ratio of the irradiances transmitted by identical polarizers in parallel and crossed orientations, which ranges from ~ 0.002 at $700\,\mathrm{cm}^{-1}$ to ~ 0.05 at $6000\,\mathrm{cm}^{-1}$.[49] For comparison, the attenuation ratios of MgF_2 and calcite polarizers operating in the ultraviolet to near-infrared are 10^{-4} to 10^{-6}. Polarizers based on Brewster angle reflections at the surfaces of a pair of Ge crystals provide a much improved attenuation ratio of less than 2×10^{-4}; however, such polarizers are difficult to align. A reasonable trade-off in rotating-element IRSE is to employ a fixed Brewster angle polarizer at the output of the interferometer and a rotatable wire grid analyzer in front of the detector.[49] This configuration simplifies correction for the attenuation ratio of the analyzer, which is necessary in order to obtain an accuracy better than $\pm 1^\circ$. A compensator is desirable in the rotating analyzer IRSE since many samples of interest exhibit Δ near 0° or $\pm 180^\circ$ over a wide spectral range. Compensators in the transmission geometry are generally not used in infrared ellipsometers because of their limited spectral range.[6] Instead, compensators are employed based on the principles of external reflection from a thin film or total internal reflection within a prism. For example, a ZnSe prism in which the beam reflects from an internal surface can serve as a compensator. Finally, the piezobirefringent modulator used in phase-modulation IRSE is constructed from ZnSe as well and is operated at a frequency of 37 kHz.[42]

4.3 Spatial Resolution

Conventional ellipsometers operating from the ultraviolet to near-infrared probe the sample with a typical 1–5 mm diameter beam at $\phi_i = 50 - 75^\circ$.[1] All physical parameters of the sample including the layer thicknesses and dielectric functions must be uniform over the probed area, unless provisions for nonuniformity are included in data analysis. In many situations, information is sought on macroscopically nonuniform surfaces, specifically, surfaces whose properties differ on scales ranging from greater than the probe wavelength ($>5\,\mu\mathrm{m}$) to less than the beam size ($<5\,\mathrm{mm}$). For example, in studies of the oxidation of polycrystalline materials, the oxide growth kinetics, and hence the layer thicknesses, are different on differently oriented grains.[50] In addition, in studies of corrosion, passive layer breakdown on a metal surface may occur locally, signaling the onset of pitting.[51] Finally, in studies of the processing of semiconductor structures used in microelectronics, the surfaces are patterned with different multilayer stacks on different areas of the surface.[52]

If the nonuniformity over the surface of a particular sample is sufficiently well known, then conventional ellipsometry can be performed and the nonuniformity can be extracted in the data analysis. For an incident beam with lateral spatial coherence greater than the scale of the nonuniformity, the amplitude reflection coefficient for the interfaces can be calculated as the surface area-weighted average of the individual reflection coefficients for the regions of the sample with different properties.[24] Such an approach has been applied to characterize nucleating films consisting of large well-separated crystallites. For an incident beam with lateral coherence less than the scale of the nonuniformity, the irradiance at the detector can be calculated as a surface area-weighted average of individual irradiances.[24] This method has been used to characterize film thickness distributions extending over the beam area. Macroscopically nonuniform surfaces often lead to pseudodepolarization of the reflected beam, meaning that the reflected beam is characterized by an incoherent superposition of polarization states each associated with an area element of the surface having a unique structure and optical properties. Thus, in measurements in which a single beam probes a nonuniform surface, it is also important to measure p and include this parameter in the data analysis.

In spite of the recent advances in data analysis for characterization of macroscopically nonuniform samples, such samples are best studied with an ellipsometer specifically designed for this purpose. Two different approaches for performing spatially resolved ellipsometry have been developed and a resolution of about 10 μm has been achieved in both cases.

The first instrument uses focusing systems mounted before the polarizer and after the analyzer in the rotating analyzer configuration to obtain a spot size of 10 μm at the sample for ME.[52] An ellipsometric image of a surface can be generated serially by sample translation. The advantage of this approach is that, for selected points on the surface, spectra from the ultraviolet to near-infrared can be collected, which are usually necessary for the analysis of complex sample structures. Such a capability makes this instrument design best suited for semiconductor wafer mapping. A disadvantage is the inordinate time required for serial mapping combined with serial collection of spectra using a single PMT. Recently, a similar instrument has been developed in the rotating polarizer configuration, incorporating a photodiode array that

allows spectra to be collected in parallel, thus reducing the mapping time by two orders of magnitude.[53]

The second instrument for spatial resolution uses the generic ellipsometer configuration of Figure 4 but with a source consisting of a laser, coherence scrambler, and collimator.[54] A lens after the sample images the illuminated surface through the analyzer onto a two-dimensional CCD for parallel detection. Images collected at three polarizer or analyzer angles permit construction of a 480-line × 512-pixel (ψ, Δ) image in 40 s. The advantage of the CCD-based instrument for IE lies in the dynamic imaging capability that can be achieved through improvements in speed made possible by advances in CCD technology and image processing hardware and software. The disadvantage is the single wavelength operation which makes the instrument less suitable for the interpretation of complex sample structures.

The speed of IE for the analysis of simple sample structures (e.g. a single nonabsorbing film of known n_f on a substrate of known N_s) can be improved by using the configuration of Figure 4 in the off-null mode.[55] In this mode, the angles P and A are set to obtain a null image for the bare substrate. During nonuniform film growth on the substrate, the irradiance that develops at a given pixel is proportional to the square of the film thickness at the conjugate point on the sample. As a result, the image in the irradiance provides a thickness map over the surface. The off-null imaging ellipsometer has been applied as a ''proof-of-concept'' affinity biosensor.[55] An example of IE results from such an instrument is given in Figure 9. The sample in this case is a 15 × 25-mm hydrophobic silicon wafer onto which three different proteins have been adsorbed, including fibrinogen, HSA, and human immunoglobin. The different plateau heights are due to the differences in thicknesses that arise from the different sizes of the protein molecules. For example, the HSA layer is the thinnest at 2 nm. In the affinity biosensor, the Si wafer can be sensitized in small areas by incubation with solutions containing different molecules with specific affinities. Then the wafer is exposed to solutions containing the molecules to be detected. The differences in the ellipsometric signals in the different areas provide the biosensor output.

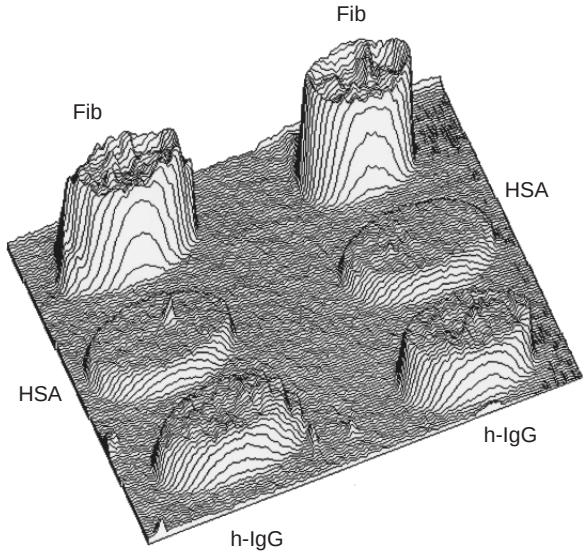

Figure 9 Ellipsometric image of a 15 × 25-mm hydrophobic silicon wafer onto which three different proteins have been adsorbed in pairs of 4-mm spots. The three proteins are fibrinogen (Fib), human serum albumin (HSA) and human immunoglobin G (h-IgG). The thickness of the HSA film is 2 nm. (Reproduced from H. Arwin, *Thin Solid Films*, **313–314**, 764–774. Copyright (1998) with permission of Elsevier Science.[55])

4.4 Sample Preparation and Limitations

Various complications may be encountered in applying the conventional ellipsometric measurement and analysis procedures of section 3 to real samples. Two important sample-related issues need to be addressed. The first issue is related to sample structure, inhomogeneity, and nonuniformity. Samples exhibit features on scales ranging from the atomic to macroscopic. Five scales of sample features can be discussed in relation to measurements performed on a conventional ellipsometer: (i) the atomic scale (~0.1–1 nm); (ii) much greater than the atomic scale, but less than the wavelength of the probe light (~10–100 nm); (iii) the probe wavelength scale (~200 nm–1 μm), (iv) much larger than the probe wavelength, but less than the beam size (~5 μm–1 mm); and (v) larger than the beam size (>1 mm). Whether these features can be studied by ellipsometry and, if so, how the information can be extracted is one issue addressed in this section. The second issue concerns the deviations from ideality in the nature of real surfaces and interfaces. Obviously the concept of a mathematically abrupt interface as employed in section 3 is an idealization; however, surfaces can be prepared to conform as closely as possible to this idealization. Alternatively, nonidealities can be taken into account in data analysis.

First the effect of atomic structure, microstructure, inhomogeneity, and nonuniformity of the sample on the ellipsometry measurements will be discussed. From section 4.3, it should be clear that if the sample is nonuniform on a scale larger than the beam size, then the conventional analysis is valid, and the sample can be mapped by translating it in the plane of the surface. If the sample is nonuniform on a scale smaller than the beam size, but larger than λ and the lateral spatial coherence of

the beam, then the nonuniformity can be detected through its effect on p. In this case, averaging methods using incoherent summation of irradiances may be effective in characterizing the nonuniformity, providing that its origin is sufficiently well known.[24] A better approach to characterize such nonuniformity is with IE or ME (see section 4.3).

Inhomogeneity on the scale of λ is perhaps the most difficult to characterize. If the inhomogeneity is extensive, then there may be no significant irradiance in the specularly reflected beam and the measurement may fail. At the other extreme, if the inhomogeneity is caused by a low density of defects in the sample or on its surface, then the defect-free regions will generate a well-defined specular beam while the defects will scatter light. An ellipsometry measurement is relatively insensitive to such defects, whereas a reflectance measurement is much more strongly affected since the scattering generates a loss. As defects increase in density on the surface, it is useful to determine the beam parameter p along with the sample parameters (ψ, Δ), e.g. with a rotating compensator or dual-channel phase modulation ellipsometer, to assess the possible effect of overlapping scattered and specular light at the detector.[56] In such cases, analysis of the (ψ, Δ) data can provide the specular sample properties, which can be applied to predict the reflectance R that would be observed if the sample was free of inhomogeneities. An analysis of the deviation between the measured and predicted reflectance can provide information on the nature of the inhomogeneity.[57] Thus, an ellipsometer that can simultaneously provide the sample parameters $\{(\psi, \Delta), R\}$ and the reflected beam parameter p is beneficial.

Sample microstructure much larger than the atomic scale, but smaller than λ can be characterized with ellipsometry by applying effective medium theories.[58] Such theories allow the optical properties of a composite to be calculated from the optical properties of the component materials and their volume fractions. The simplest effective medium theories have the form:

$$\frac{\varepsilon - \varepsilon_h}{\varepsilon + \kappa\varepsilon_h} = f_a \frac{\varepsilon_a - \varepsilon_h}{\varepsilon_a + \kappa\varepsilon_h} + f_b \frac{\varepsilon_b - \varepsilon_h}{\varepsilon_b + \kappa\varepsilon_h} + \cdots \qquad (67)$$

where ε is the dielectric function of the composite, which is a mixture of two or more components having volume fractions $f_a, f_b, \ldots$ and dielectric functions $\varepsilon_a, \varepsilon_b, \ldots$. In Equation (67), ε_h is the host medium dielectric function, and κ is a parameter that describes the screening at the phase boundaries. Effective medium theories used for solid state applications include the Bruggeman theory in which case $\varepsilon_h = \varepsilon$ and $\kappa = 2$, and the Maxwell–Garnett theory in which case ε_h is the dielectric function of the dominant component and $\kappa = 2$. The choice $\kappa = 2$ in both cases corresponds to spherical inclusions, i.e. isotropic microstructure in three dimensions.

The Bruggeman theory has been applied most widely to model the optical properties of thin films containing regions devoid of atoms (i.e. voids) or films consisting of multiple phases (e.g. an amorphous solid with crystalline inclusions).[58] In these cases, an analysis provides the void or crystalline volume fraction. The Bruggeman theory has also been applied to model surface or interface roughness as one or more layers consisting of a mixture of the underlying and overlying materials such that the volume fraction(s) and layer thickness(es) are to be determined. Finally, effective medium theories have been used to model nucleating thin films as a mixture of bulk material and voids, representing the clusters and the free space between clusters, respectively. The Bruggeman theory has been applied for nucleating semiconductor films, whereas a form of the Maxwell–Garnett theory has been applied for nucleating metal films. Using such a model, the thickness and void volume fraction of the nucleating film is determined. Further discussion of the data analysis procedures will appear in section 4.5.

Effective medium theories begin to lose their validity when the scale of the microstructure approaches within an order of magnitude of λ.[59] In this regime, light scattering comes into play as described earlier, and the assumptions on which the theories are based break down. In the opposite limit, if the microstructure becomes small enough so that the individual inclusions no longer exhibit the bulk dielectric functions, then size-modified versions must be used instead in Equation (67). This lower limit depends on the material. For example, crystalline silicon exhibits size-dependent optical effects for grain sizes as large as 100 nm, whereas in amorphous silicon, size-dependent effects have not been identified even for the smallest clusters.[60] Thus, in the latter case, the optical properties of amorphous silicon containing a distribution of atomic scale voids can be predicted from the optical properties of dense amorphous silicon. In contrast, for a crystalline material, atomic scale heterogeneities alter the basic electronic structure and require that the optical properties of the materials be characterized individually as a function of their composition. The same strategy is needed in the case of crystalline and amorphous alloys, an example being $Si_{1-x}Ge_x$. In general, size effects render the analysis extremely difficult in all but the simplest situations.

This section will conclude with a discussion of deviations exhibited by real samples from mathematically exact behavior in the case of single interfaces. This problem has a long history starting with Paul Drude and Lord Rayleigh. Drude noticed the gradual change over time in the polarization state of light reflected from a freshly cleaved Sb_2S_3 crystal and attributed it to contamination

arising from the reaction of the top layers of the crystal with the ambient medium.[9] Rayleigh noticed that the light reflected from a water surface at the Brewster angle showed residual ellipticity and attributed the effect to a thin ''greasy contamination'' layer.[61] He concluded that ''it is even possible that there would be no sensible ellipticity for the surface of a chemically pure body''. Thus, the basic problem of sample preparation and cleanliness in polarization state measurements was recognized more than a century ago. Today researchers struggle with the same issues and recognize not only contamination, but also surface roughness, and even surface reconstruction as effects that distort the experimental data from results predicted on the basis of models assuming mathematically abrupt interfaces.

One might expect that the best way to prepare a bulk material for SE measurement of its dielectric function is to exploit the methods of surface science.[62] This would entail mounting an UHV system at the ellipsometer axis and measuring the sample in situ through UHV windows before and after desorption of native oxides through annealing or sputter etching. Such treatments can generate microscopic roughness or damaged surface regions, however, whose effects on the SE data are often larger than that of the native oxide. Thus, a superior approach is to prepare the material in situ and measure it under UHV prior to exposure to air. In this case, one need only worry about the roughness on the as-prepared surface and any surface reconstruction that occurs. In fact, measuring the material during preparation as described in section 5 often allows surface layers that have different optical properties from the underlying bulk to be deduced, and then taken into account in the analysis. Unfortunately, the technical complexity has limited the application of this approach. In the absence of UHV capabilities, in situ chemical cleaning and chemomechanical polishing have been used most widely to prepare sample surfaces.[63] In such experiments, an open-windowed cell is fixed at the ellipsometer axis which allows the sample to be mounted and its surface to be exposed to various treatments while being measured in situ. In fact, a ''biggest is best'' criterion in $\langle\varepsilon_2\rangle$ has been developed to monitor the success of the treatments. With this criterion, the treatments leading to the largest value of $\langle\varepsilon_2\rangle$ at a particular wavelength provide the surface that is the closest to a mathematically abrupt termination of the bulk structure and, thus, the closest representation of ε_2 for the bulk material. When these conditions are reached, then the surface can be measured by SE under a N_2 flow to maintain the clean surface. The resulting spectrum in $\langle\varepsilon\rangle$ deduced from (ψ, Δ) is as close as possible to the true dielectric function ε.

Figure 10 provides an example of this approach for crystalline Si.[64,65] In this case different crystal faces are

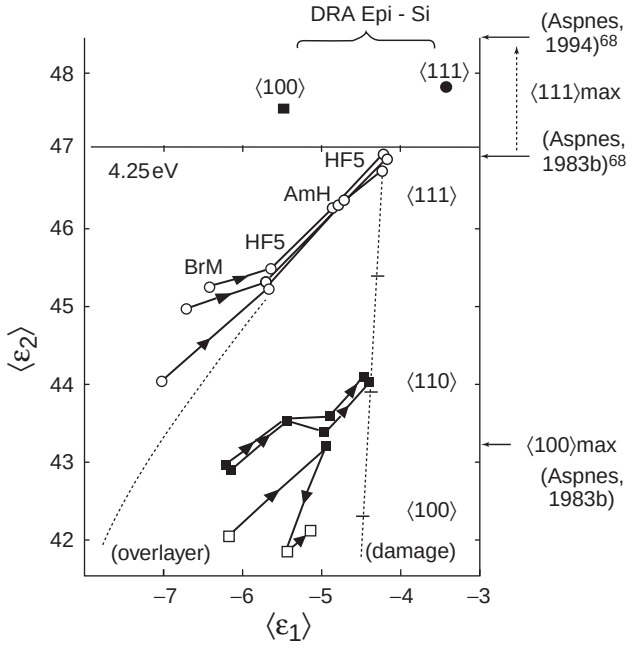

Figure 10 Summary of pseudodielectric function values from the ⟨100⟩, ⟨110⟩, and ⟨111⟩ surfaces of single crystal Si measured at a photon energy of 4.25 eV. The lower panel shows the results of surface treatments performed in an attempt to maximize $\langle\varepsilon_2\rangle$. The broken lines show treatment trajectories expected for oxide removal (overlayer) and surface smoothening (damage). The upper panel shows measurements of epitaxial Si. DRA Epi-Si, epitaxial Si prepared at the Defence Research Agency; BRM, treatment with 0.02 vol.% bromine in methanol; HF5, treatment with 5 vol% HF in methanol; AmH, treatment with 1 : 1 $NH_4OH : H_2O$ by volume. (Reproduced from C. Pickering, *Thin Solid Films*, **313–314**, 406–415. Copyright (1998) with permission of Elsevier Science.[65])

exposed to chemical and chemomechanical treatments in an attempt to obtain the highest value of $\langle\varepsilon_2\rangle$ at $\lambda = 291.8$ nm, where the ''biggest is best'' criterion holds. Because Si is cubic, mathematically abrupt terminations of the bulk should provide the same data for ε_2 for all crystal faces. Figure 10 shows that it is easiest to obtain an abrupt ⟨111⟩ surface, whereas it is most difficult to obtain an abrupt ⟨100⟩ surface. The trajectory labeled ''damage'' indicates the expected trend in the $(\langle\varepsilon_1\rangle, \langle\varepsilon_2\rangle)$ data if roughness developed on the treated ⟨111⟩ surface. A comparison of this trend with the data for different crystal faces shows that roughness can account for much of the differences between the best surfaces. The trajectory labeled ''overlayer'' is the expected trend if an SiO_2 layer developed on the treated ⟨111⟩ surface. A comparison of this trend with the data for individual crystal faces shows that much of the improvement in the surfaces upon cleaning can be attributed to oxide removal. The arrow at the top right indicates the most recent results for the ⟨111⟩ Si surface. The points at the top show that much better data for the ⟨100⟩ surface and comparable data for the

⟨111⟩ surface can be obtained from epitaxially grown Si measured immediately upon removal from the deposition chamber.

Techniques such as those shown in Figure 10 have led to the development of an extensive database for the optical properties of various materials and this database continues to be improved.[13,66,67] With an accurate database, information can be obtained on the nonidealities in more complex single layer and multilayer structures. For example, using the database one can predict the multiple (ψ, Δ) data sets to be expected for a thin film of arbitrary thickness d_f which presents ideal interfaces to the ambient medium and to the underlying substrate. Inevitably, an exact fit to the data for a real film cannot be obtained for any one choice of film thickness. Additional layers can then be introduced into the model in an attempt to understand the origin of the deviations in the data. For example, native oxides or roughness layers can be added to model a nonideal film surface, and interface roughness or chemical interaction regions can be added to model a nonideal film/substrate interface. In situ surface cleaning techniques similar to those mentioned above can also be applied in an attempt to reduce the effects of a nonideal ambient/film interface. In the next section, the discussion will focus on the analysis methods for multiple (ψ, Δ) data sets collected on complex structures such as these.

4.5 Data Analysis

The analysis of multiple (ψ, Δ) values has been described briefly in sections 3.3 and 3.4 for single and multilayer samples. Next, general data analysis procedures will be described in greater detail. Owing to the widespread use of SE, it is assumed that the minimum data set consists of continuous spectra of 50–300 (ψ, Δ) values.[5] Multiple spectra can also be collected for different angles of incidence, in a method called variable angle SE,[26] or for different incident polarization states, in a method called generalized ellipsometry.[25] The availability of multiple spectra further enhances the analytical capabilities of SE. We assume that SE is performed ex situ, so that only the final sample state is accessible. Analysis techniques specific to RTSE data will be described in section 5.2. Here, two categories of the data analysis method will be described. In the first category, the analysis of sample optical properties and structure is reduced to the determination of several λ-independent quantities. Then least squares regression analysis can be applied to find the parameter values that minimize the deviations between the experimental data and the simulated results. In the second category, the dielectric function of one component of the sample is determined by exact numerical inversion of the experimental data. In addition, by constraining the form of this dielectric function, as many as three λ-independent structural parameters, usually thicknesses, can be deduced.

In the first general data analysis procedure, a model for the sample is proposed at the outset.[69] The model includes the number of layers m and the expected nature of these layers. The unknown λ-independent free parameters in the analysis will include the layer thicknesses, $d_j; j = 1, \ldots, m$. The optical properties of the layers can be incorporated into the model by applying one or more of at least four different strategies.[28]

1. Dielectric functions from previously determined databases can be used. In this case, no additional free parameters are introduced.
2. An effective medium theory can be used for any layers that are composite materials. In this case, the additional λ-independent parameters are the component volume fractions. The dielectric functions of the component materials can be obtained using any of the other three strategies.
3. A parameterization formula for one or more of the dielectric functions can be used. In this case, the λ-independent parameters are defined by the formula.
4. The λ-independent parameters from the strategy 3 can be expressed as functions of the composition of the material, thus allowing the composition to be used directly as a free parameter.

Once the model and the variable parameters are identified, often with a simulation program, initial guesses for the parameters are made by performing a grid search within the simulation program. Then a least squares regression analysis routine provides the successive parameter adjustments required to minimize the error function, which can be computed as the biased estimator of the mean square deviation between the data and calculated fit.[28] Once the best fit is obtained for a given model, complexities are added or other models are tested in an attempt to identify the correct model. Such identification is made on the basis of (i) the reasonableness of the model and best fit parameters, (ii) the biased estimator, i.e. the magnitude of the error function relative to the accumulated experimental errors in the data, and (iii) the statistical information deduced from the best fit, including the confidence limits on the parameters and the elements of the parameter correlation matrix. The power of this technique has been extolled while potential pitfalls resulting from its abuse have been outlined.[28]

As an example of this data analysis approach, it is helpful to consider the case of a known substrate coated with an absorbing film of unknown optical properties,

such that the dielectric function of the film must be parameterized. One of the simplest ways of doing this is to assume a sum of m_o damped harmonic oscillators (often referred to as Lorentz oscillators) (Equation 68):[23]

$$\varepsilon(E) = \varepsilon_\infty + \sum_{j=1}^{m_o} \frac{A_j^2}{E_j^2 - E^2 + i\Gamma_j E} \tag{68}$$

Here $E = hc/\lambda$ is the photon energy, ε_∞ is a constant contribution to the real part of the dielectric function, and $\{A_j, E_j, \Gamma_j\}$ denote the amplitude, resonance energy, and broadening parameter for the jth oscillator. Thus, if analysis assuming a single oscillator is attempted, then least squares regression involves fitting to determine five free parameters, d_f, ε_∞, and $\{A_1, E_1, \Gamma_1\}$. For each additional oscillator incorporated into the dielectric function, an additional three parameters must be determined. If a roughness layer is known to exist on the film, then one additional parameter needs to be determined, the roughness layer thickness (assuming a void volume fraction of 0.5). For example, the dielectric functions of bulk $Al_{1-x}Ga_xAs$ crystals have been modeled using Equation (68) with 7 oscillators and 22 free parameters, ε_∞ and $\{A_j, E_j, \Gamma_j\}; j = 1, \ldots, 7$, each of which could be expressable as simple quadratic functions of the alloy composition x.[70] As a result, a database of 66 parameters is sufficient to parameterize the dielectric function of the alloys fully and thus calculate $\varepsilon(E)$ for any x. With this database, analysis of SE measurements of a heteroepitaxial film of $Al_{1-x}Ga_xAs$ on GaAs can provide d_f and x. This demonstrates the importance of dielectric function parameterization and the development of databases. More sophisticated methods than that of Equation (68) have been developed for dielectric function parameterization based on an analysis of the energy band structures of the materials.[71]

In the second general data analysis procedure, a single unknown dielectric function is extracted from the (ψ, Δ) spectra, without making any assumptions about its functional form. In addition, one or more λ-independent structural parameters can be determined.[72] At first sight, this procedure may seem impossible since more parameter values are deduced than data available. The additional structural parameters can be determined, however, based on the constraint that the dielectric function $\varepsilon = \varepsilon_1 - i\varepsilon_2$ extracted from (ψ, Δ) by exact numerical inversion must show physically reasonable characteristics. One may stipulate that ε_2 vanish over some region of the spectrum if the material is a semiconductor or dielectric, or that ε show no artifacts due to features from the dielectric function of the substrate or from the interference pattern in the (ψ, Δ) data. Thus, this procedure is performed by (i) guessing the unknown λ-independent parameters, (ii) applying a nonlinear numerical analysis method (such as that of Newton) to determine the remaining two unknowns $(\varepsilon_1, \varepsilon_2)$ from (ψ, Δ), (iii) inspecting $(\varepsilon_1, \varepsilon_2)$ for the desired characteristics, and (iv) adjusting the initial guesses successively in order to obtain these characteristics. A specific application of this analysis procedure is provided in the next section.

4.6 Applications in the Characterization of Materials, Thin Films, and Structures

In this section, selected applications will be described in which SE has been employed to determine the dielectric function of bulk materials and the dielectric function and microstructure of thin films. In the first part of this section, a novel method is described for determining the infrared optical properties of liquids using an infrared spectroscopic ellipsometer combined with an attenuated total reflection cell. In the second part, the application of different approaches for the complete analysis of SE data on thin and thick films will be provided. The first two examples in the second part will focus on the analysis of relatively thin (<50 nm) polymeric and protein films, while the second two examples will focus on thick (>500 nm) inorganic films.

4.6.1 Bulk Surfaces: Optical Property Measurement

The analysis of SE data to obtain the dielectric function ε_s of a bulk material is straightforward; one simply substitutes the angle of incidence ϕ_i, the ambient medium dielectric function ε_a, and the measured (ψ, Δ) spectra into Equation (23). As discussed in section 4.4, however, the major difficulty rests in the uncertainty about the nature of the interface.

Here, an application of IRSE is described, namely the measurement of the optical properties of liquids using an attenuated total reflection cell.[48] The instrument is shown in Figure 8, and the measurement configuration is shown in Figure 11(a). The ambient medium is a ZnSe 45° prism that encloses the liquid within a reservoir. The infrared beam enters the first prism face at normal incidence, strikes the ZnSe/liquid interface at $\phi_i = 45°$, and exits through the second prism face. Total internal reflection occurs for a liquid index of refraction of $n_s < 1.7$; however, IRSE does not rely on this condition. The major errors in this measurement arise from sample nonidealities. For example, the ZnSe/liquid interface is most likely to be far from ideal even if the liquid itself is contaminant free. Surface roughness, a native oxide, and contamination are likely to be found on the ZnSe side of the interface. The advantage of the inverted configuration in Figure 11(a) is that a measurement performed in the absence of a liquid can be used to obtain information on the perfection of the interface. In such analyses, the optical properties of the ZnSe ambient medium from a

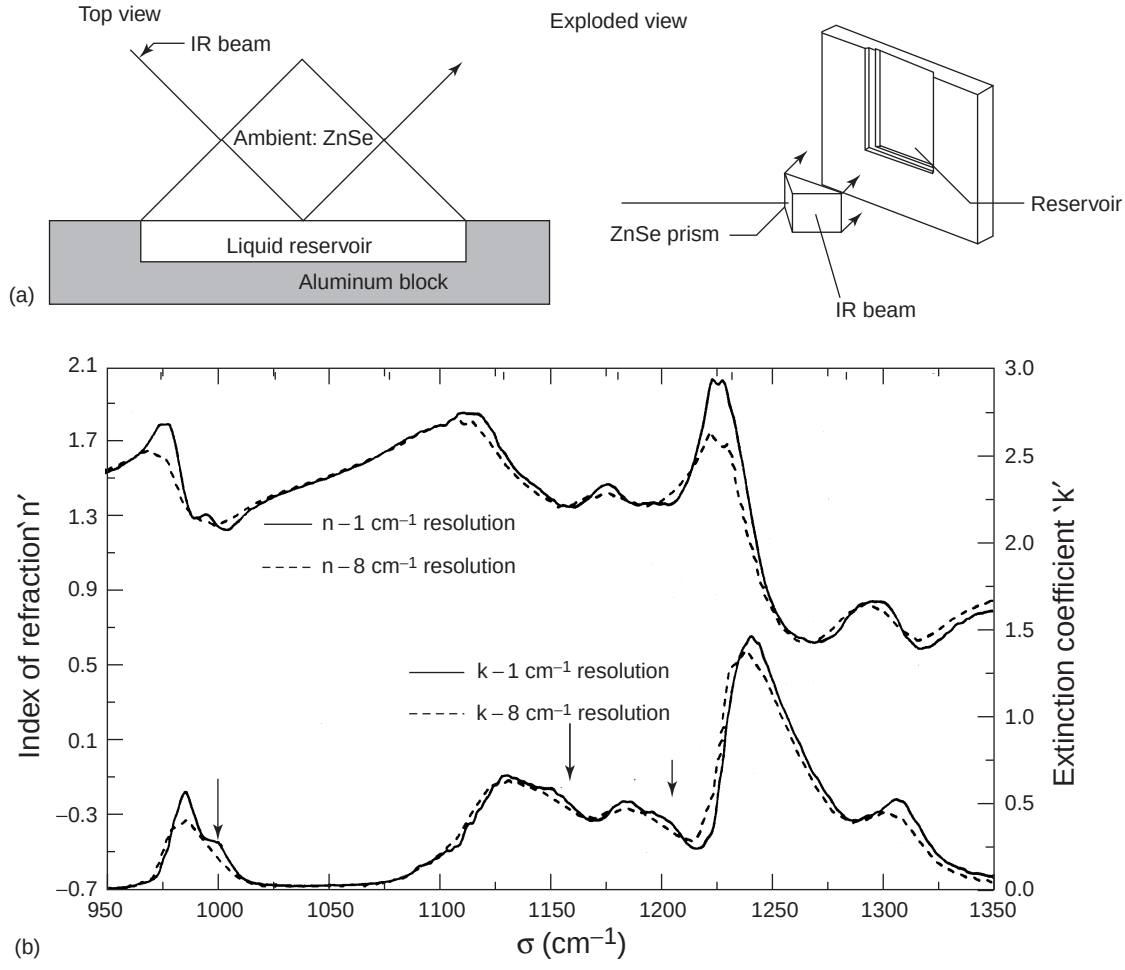

Figure 11 (a) Top and exploded views of a cell used to determine the infrared optical properties of a liquid in the reservoir. (b) Infrared optical properties of Krytox 16256N at 1 and 8 cm^{-1} resolution obtained from the instrument in Figure 8 using the cell of (a). (Reproduced from T.E. Tiwald, D.W. Thompson, J.A. Woollam, S.V. Pepper, *Thin Solid Films*, **313–314**, 718–721. Copyright (1998) with permission of Elsevier Science.[48])

pre-existing database are employed. Figure 11(b) shows typical results obtained for Krytox 16256N, a liquid lubricant made by DuPont, described in terms of (n_s, k_s). Once the optical properties of the bulk lubricant are available, the results can be added to a database that will allow the thickness and surface coverage of a thin lubricant layer to be determined. For example, partial surface coverage by a lubricant can be modeled as a layer consisting of a mixture of bulk lubricant and void using the Bruggeman theory (see section 4.4).

4.6.2 Thin Films: Thickness and Optical Property Measurement

The analysis of SE data for thin film structures is more difficult than for bulk materials. General data analysis approaches for single and multilayer structures have been described in section 4.5. Here the approaches are demonstrated for the characterization of a thin polypyrrole film and a much thinner film of lactoperoxidase (LPO), both on gold substrates.

Electrically conducting polypyrrole films for SE studies were prepared by electrodeposition on opaque Au substrates in an electrochemical cell using an aqueous solution of the monomer with a KNO_3-based electrolyte.[18] The (ψ, Δ) spectra were measured in situ to avoid reaction of the film with ambient air. SE was performed using a rotating polarizer multichannel ellipsometer with an acquisition time of 3.2 s (see section 5.1). The (ψ, Δ) spectra were analyzed using a single layer model to determine the dielectric function ε_f and thickness d_f of the polypyrrole film. The well-defined features in the dielectric function ε_s of the Au substrate near 2.5 eV facilitated such an analysis. Thus, d_f was chosen by trial and error to eliminate the Au-related

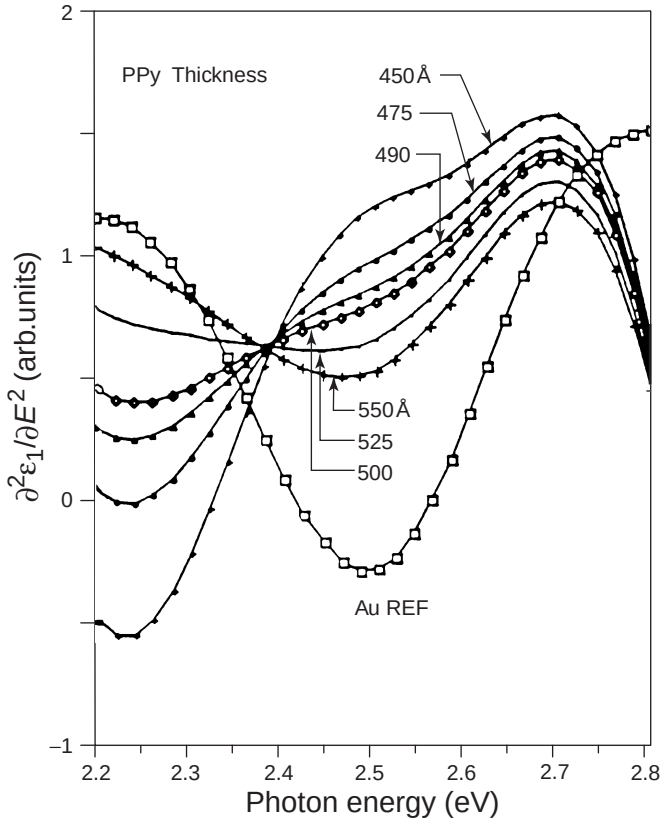

Figure 12 Second derivatives of the real parts of trial dielectric functions for a polypyrrole (PPy) thin film calculated assuming a single-layer optical model consisting of H_2O/PPy/Au using different PPy thicknesses. Corresponding results for the Au substrate are also shown. The correct thickness of the PPy film, 49 nm, minimizes the artifact in the derivative spectrum arising from the dielectric function of the substrate. (Reproduced from Y.-T. Kim, D.L. Allara, R.W. Collins, K. Vedam, *Thin Solid Films*, **193–194**, 350–360. Copyright (1990) with permission of Elsevier Science.[73])

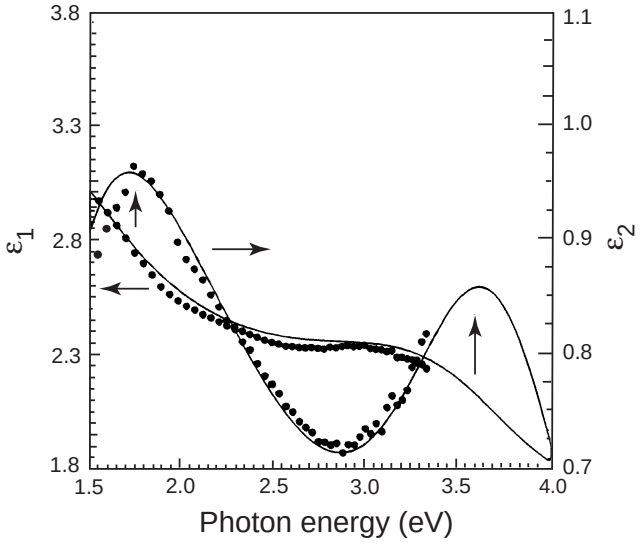

Figure 13 Dielectric function of a 49 nm polypyrrole film deduced assuming a single-layer optical model (H_2O/PPy/Au) and the method of Figure 12 (•). The solid line is obtained by fitting the (ψ, Δ) spectra using a model for the dielectric function that includes three Lorentz oscillators. •, Inversion; ——, oscillator model; $\varepsilon_1 = n^2 - k^2$; $\varepsilon_2 = 2nk$. (Reproduced from Y.-T. Kim, D.L. Allara, R.W. Collins, K. Vedam, *Thin Solid Films*, **193–194**, 350–360. Copyright (1990) with permission of Elsevier Science.[73])

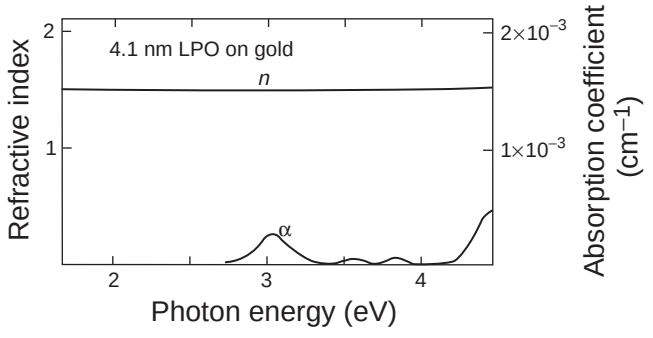

Figure 14 Spectra of the refractive index n and absorption coefficient α for a 4.1 nm thick monolayer film of the protein LPO measured at the solid–liquid interface. (Reproduced from H. Arwin, *Thin Solid Films*, **313–314**, 764–774. Copyright (1998) with permission of Elsevier Science.[55])

artifacts at 2.5 eV in ε_f, which was obtained by exact numerical inversion of the (ψ, Δ) spectra, as described in section 4.5. Figure 12 shows derivative spectra plotted versus energy for several trial ε_{1f} spectra, obtained by inversion assuming different d_f values. Inspection of these curves identifies $d_f = 49 \pm 0.5$ nm as the correct choice. The resulting ε_f for the polypyrrole is shown in Figure 13 (•). The solid line results in Figure 13 were obtained in a least squares regression analysis of the same (ψ, Δ) data assuming that ε_f can be described as a sum of three Lorentz oscillators. Thus, this analysis requires extraction of 11 λ-independent parameters, ten associated with Equation (68), and one thickness. The resulting best fit d_f by least squares regression was found to be 47.5 ± 2 nm, in good agreement with the value from artifact minimization. The agreement between the dielectric functions is also very good. Oscillators with $E_1 = 1.65$ eV and $E_2 = 2.3$ eV appear to represent sub-gap transitions from the valence band to defect states associated with anion (NO_3^-) doping whereas the oscillator at $E_3 = 3.8$ eV represents the band gap transition.

As another example, Figure 14 shows the optical properties, expressed in terms of the index of refraction n_f and the absorption coefficient α_f, plotted versus photon energy for a much thinner film ($d_f = 4.1$ nm) of the protein LPO adsorbed onto a Au substrate.[55] These results were obtained serially by in situ SE using

a single-channel instrument, and were extracted from (ψ, Δ) data using numerical inversion. For such films, the correct thickness d_f can be chosen to ensure first that no Au substrate-related features appear in the LPO film optical properties and second that α_f vanishes in the regions where absorption bands are absent. As in the case of polypyrrole, well-defined features appear in the LPO optical properties that are specific to the protein and can assist in its identification or thickness determination when the optical properties are added to a database.

Next, applications of SE for characterization of thick semitransparent films will be described. In this case, the spectra in (ψ, Δ) show interference patterns since Z_f in Equation (26) passes through several periods as λ increases from the ultraviolet to near-infrared. Figure 15 provides results for a ZnSe layer prepared by molecular beam epitaxy on GaAs.[74] These results were obtained with a rotating analyzer instrument (see Figure 6), and the sample was mounted in an open-windowed cell for measurement. Studies of native oxide removal were performed by exposing the surface to a $NH_4OH : CH_3OH$ solution, and then to dry N_2, while monitoring in real time at 3.5 eV. If no further increases in Δ were detected with additional cycling, it was assumed that the oxide was completely removed. Then the (ψ, Δ) spectra were obtained on the oxide-free, stable surfaces under flowing N_2. For photon energies above the bandgap where the ZnSe film is opaque, (ψ, Δ) provide ε_f for the ZnSe directly according to Equation (23). Below the gap where the film is nonabsorbing, a rapidly varying interference pattern dominates (ψ, Δ). In this regime, ε_f can be deduced by numerical inversion using a single layer model and the known ε_s for the GaAs substrate. In this case, the layer thickness was chosen to eliminate interference artifacts in ε_f, leading to $d_f = 1363.3 \pm 7.6$ nm. Figure 16 shows the dielectric functions of the ZnSe along with two $Zn_{1-x}Cd_xSe$ layers on GaAs obtained using the same methods. The solid lines are fits to these results using 27 parameters that are each quadratic functions of x. This parameterization allows SE data to be modeled for any $Zn_{1-x}Cd_xSe$ layer to determine d_f and x. Finally, the inset in Figure 16 shows the variability in the ε_{2f} spectra for ZnSe near the bandgap, as obtained by different researchers. This variability is attributed to the differences in the surface conditions, with the lowest ε_2 being most representative of the correct results in this case.

The analysis of the (ψ, Δ) data for the ZnSe/GaAs heterostructures of Figure 15 is relatively straightforward owing to the perfection of the structure once the oxide is removed. Here, an application at the other extreme is described, a thick transparent conducting $SnO_2 : F$ film prepared by chemical vapor deposition on glass. Figure 17 shows $\{(\psi, \Delta), R\}$ of the sample and p of the

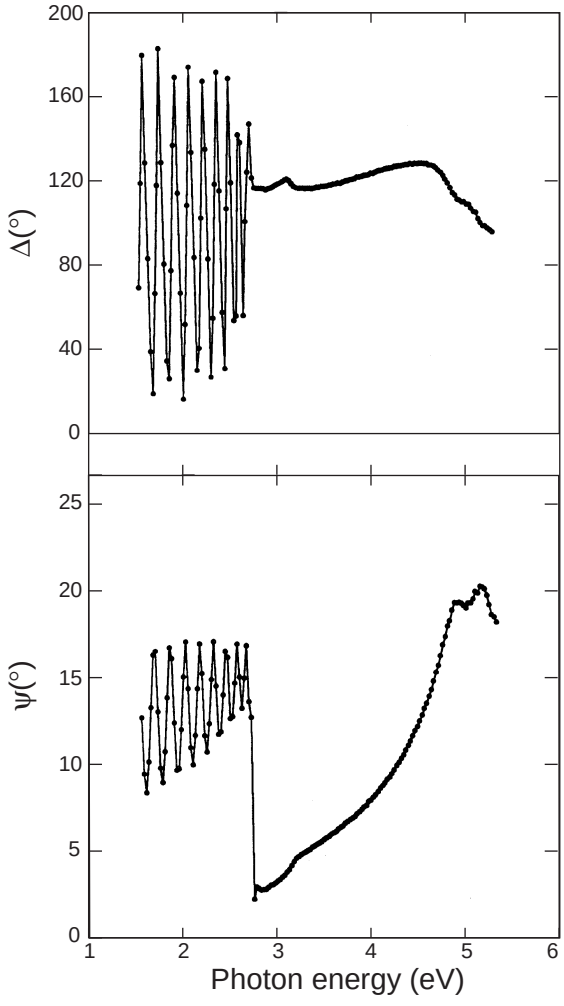

Figure 15 Ellipsometric (ψ, Δ) spectra for a 1.363 μm thick ZnSe layer on a (100) GaAs substrate measured with an angle of incidence of 70°. Surface cleaning was performed to remove the native oxide. The solid line connects the data points. (Reproduced by permission of the Optical Society of America from J. Lee, F. Flack, N. Samarth, R.W. Collins, *Appl. Opt.*, **36**, 5372–5382 (1997).[75])

reflected beam for a $SnO_2 : F$ film plotted versus energy, measured in parallel in 2.5 s using a rotating compensator multichannel ellipsometer.[57] Least squares regression analysis of the (ψ, Δ) data is performed from 1.5–3.75 eV where $p > 0.9$. In the analysis, the dielectric function of the $SnO_2 : F$ is simulated using two oscillators, the first due to free electrons with $E_1 = 0$ eV and the second due to interband transitions with E_2 greater than the bandgap of $SnO_2 : F$. This second oscillator is described as a convolution between a Lorentz oscillator and a Tauc law that generates a well-defined absorption onset at the $SnO_2 : F$ bandgap. As a result, the dielectric function of the $SnO_2 : F$ is defined by seven free parameters. Figure 17 includes the best fit to the (ψ, Δ) spectra, and Figure 18 shows the resulting dielectric function along

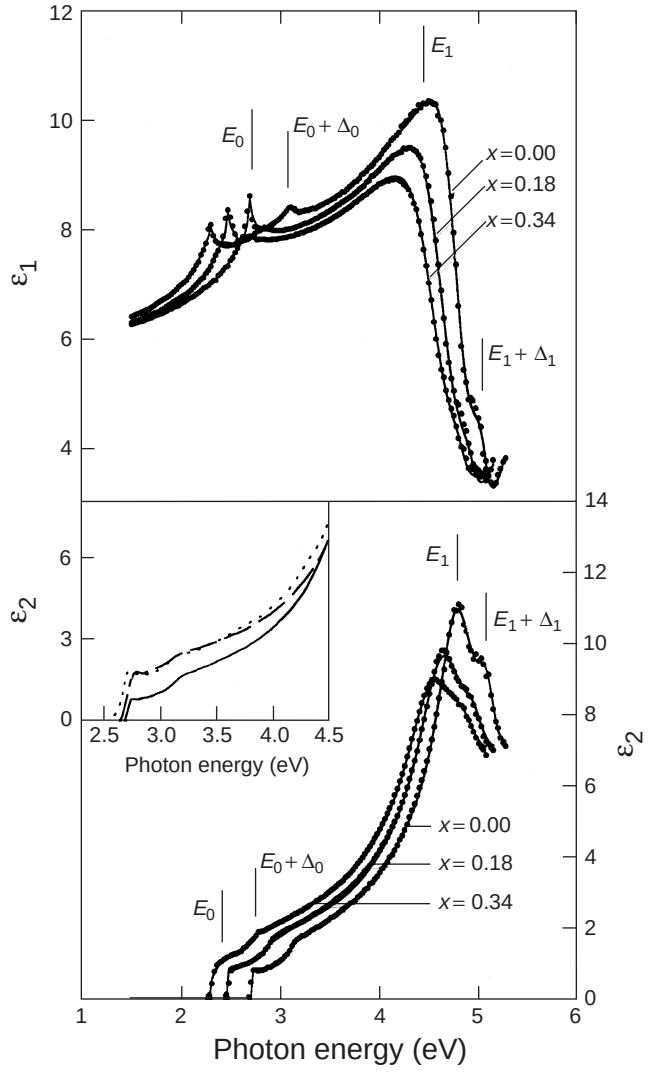

Figure 16 Dielectric functions measured for ~0.5–1.5 μm thick $Zn_{1-x}Cd_xSe$ layers prepared with $x = 0$, 0.18, and 0.34 onto (100) GaAs. The solid lines are fits to the experimental spectra (•) using a parameterization scheme. The inset shows a comparison among ε_2 spectra for ZnSe near the bandgap energy as measured by different researchers. (Reproduced by permission of the American Institute of Physics from J. Lee, R.W. Collins, A.R. Heyd, F. Flack, N. Samarth, *Appl. Phys. Lett.*, **69**, 2273–2275 (1996).[74])

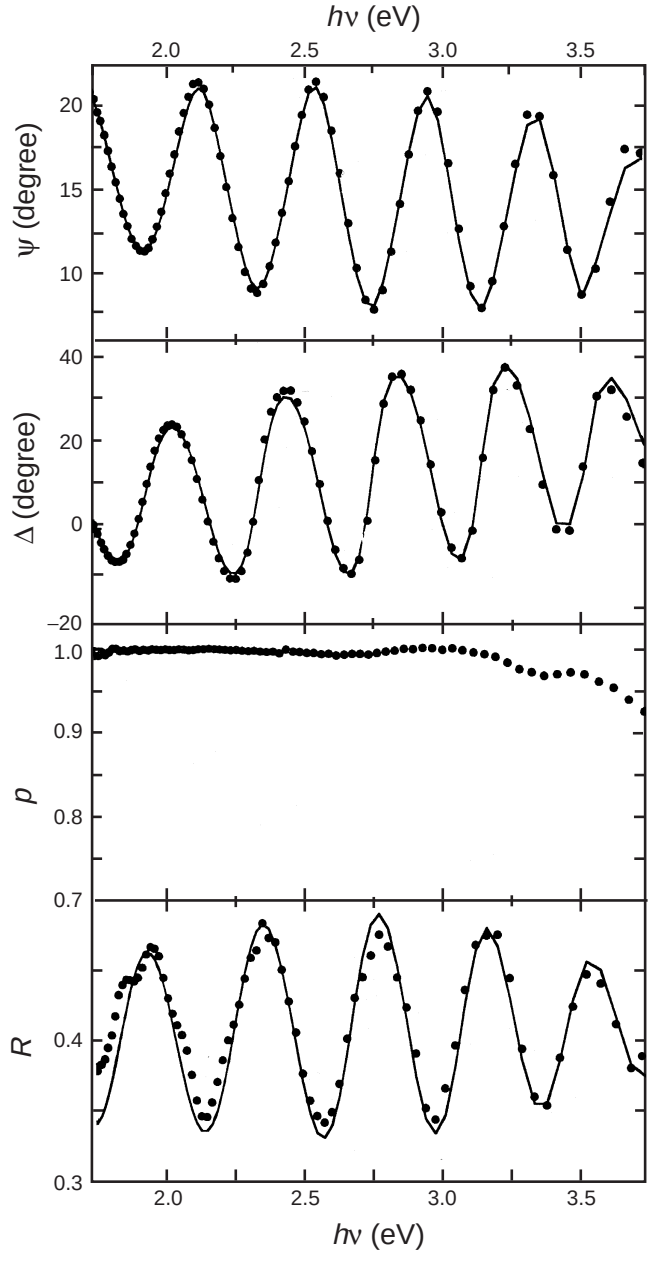

Figure 17 Spectra in (ψ, Δ), the degree of polarization p of the reflected beam, and the reflectance R for a 0.75 μm thick fluorine-doped SnO_2 film on glass. All four spectra were collected in 2.5 s using the rotating compensator multichannel ellipsometer of Figure 19. The solid lines are fits to the data using the two-layer model of Figure 18, along with a macroscopic roughness layer required to fit R (Unpublished data from P.I. Rovira, R.W. Collins, (1998).)

with the deduced sample structure. The latter includes a 752.4 ± 3.0 nm bulk layer and a 14.0 ± 1.6 nm surface roughness layer consisting of 0.48 ± 0.14 volume fraction of voids. The reflectance spectrum predicted on the basis of the best fit optical model for (ψ, Δ) shows systematic deviations from the data in Figure 17. The deviations can be accounted for by assuming additional roughness that is undetectable by SE because its feature scale is too large (see section 4.4). As a result, this macroscopic roughness component generates light scattering that appears as a reflectance loss. The best fit to R in Figure 17 is obtained by including, in addition to the microscopic roughness layer, a macroscopic roughness layer having a 15 nm wide Lorentzian distribution for the surface profile normal to the film plane with a distribution full-width at half-maximum of 3 nm.

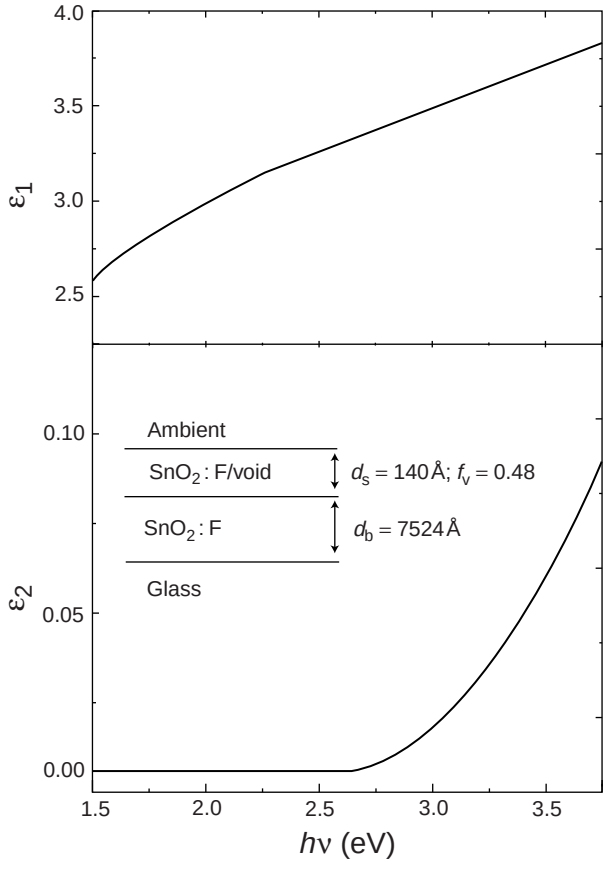

Figure 18 Dielectric function of the 0.75 μm thick bulk layer of a fluorine-doped SnO_2 film (see Figure 17) obtained in a best fit to the (ψ, Δ) spectra assuming a parametric model. The best fit structural model for the sample is shown in the inset. (Unpublished data from P.I. Rovira, R.W. Collins, (1998).)

5 REAL TIME MEASUREMENT APPLICATIONS

Each of the automatic instruments in section 4.1 can be used for real time studies at a single wavelength. In this mode, minimum (ψ, Δ) measurement times range from 20 μs to 5 ms. Because of the limitations in the analysis of data collected in real time at one λ, the focus in the first part of this section will be on instruments designed for high speed SE from the ultraviolet to near-infrared. Such instruments have been designed on the same principles as the slower scanning instruments of section 4.1, namely nulling, rotating element, and phase modulation. Two different approaches are used for rapid spectral scanning. In the first approach, a rotating grating or interference filter can replace the slower stepping-motor controlled monochromator. In the second approach, a spectrograph and photodiode array for parallel SE data acquisition can replace the conventional monochromator and PMT. In sections 5.1 and 5.2, the instruments for high speed SE and the data analysis procedures unique to RTSE will be described, respectively. In sections 5.3 and 5.4, applications of these instruments will be presented.

5.1 Instrument Designs and Selection

The first rapid-scanning system for RTSE has been designed based on the automatic nulling principle using Faraday cell modulators as in Figure 5.[76] The instrument also uses a Xe source and rotating interference filter as a rapid-scanning monochromator, and a single PMT as the detector. Successive ''nulling angles'' are measured repetitively with an acquisition time of 1 ms, and related to the wavelength scan through an encoder on the interference filter. A spectrum consisting of ~400 λ values can be measured with a 3 s repetition time. The orders of magnitude difference between the single (ψ, Δ) acquisition time and the full spectra repetition times is characteristic of a serial-scanning system. The advantages of this self-nulling instrument include uniform sensitivity for all reflecting surfaces and the capability of measuring (ψ, Δ) over their full ranges without ambiguity.

A different instrument for RTSE has been developed based the rotating polarizer principle.[18,34] Its design exploits the fact that the mechanical period accessible to a continuously rotating polarizer is ~10 times longer than the scanning time of a multichannel detector (5–10 ms). The instrument utilizes a Xe source to generate a white light beam that is linearly polarized by a rotating polarizer and reflected from the sample. The detection system, mounted after a fixed analyzer, consists of a spectrograph and a 1024-pixel linear Si photodiode array. The 1024 pixels can be grouped by eight for 128 spectral positions. The array is operated in an integrating mode, and readout is triggered at least three times per half-rotation using an encoder output from the polarizer. This yields at least three integrals over the irradiance waveform registered at each pixel group. From a minimum of three readouts, spectra in (α, β) and (ψ, Δ) can be determined (see Equations 51 and 52). With this design, the minimum acquisition and repetition times for SE data are both equal to the polarizer optical period π/w. The integrating mode implies that all pixel groups are measuring the irradiance at all times so that instrument operation is fully parallel. Thus, the rotating polarizer system has demonstrated the highest speed reported for RTSE, acquisition and repetition times of 16 and 32 ms, respectively, for 64 spectral points. Under these conditions, standard deviations of $\delta\psi \cong 0.06^\circ$ and $\delta\Delta \cong 0.12^\circ$ have been obtained at $\lambda = 496$ nm under film growth conditions. Acquisition and repetition times of 3 s have yielded $\delta\psi \cong 0.003^\circ$ and $\delta\Delta \cong 0.007^\circ$ at $\lambda = 496$ nm.[77]

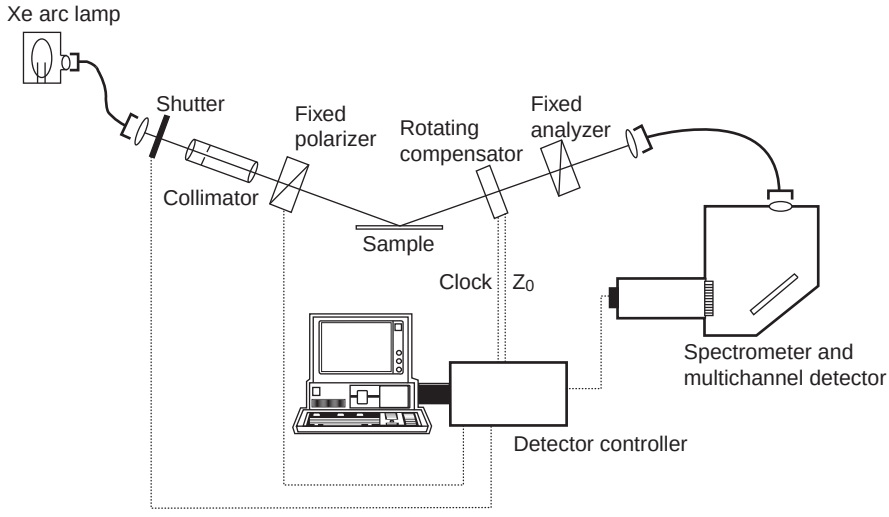

Figure 19 Schematic of a rotating-compensator multichannel ellipsometer.

The third and most recent rapid-scanning ellipsometer has been developed based on the phase modulation principle.[78] With the exception of the detection system, this instrument design is similar to that in Figure 7. A Xe source and polarizer provides a white light beam whose polarization is modulated by a piezobirefringent element. As in the rotating polarizer multichannel ellipsometer, the polarization detector consists of an analyzer, spectrograph, and photodiode array (40 pixels in this case). Because the phase-modulator period (20 μs) is ~100 times faster than the array scanning time, however, the detector cannot be operated in an integrating mode. Instead, the pixel outputs are multiplexed into a common signal processor. The output of each pixel is digitized at 1 MHz, and a fast Fourier transform provides the dc component, and first and second harmonics of the irradiance waveform from which (ψ, Δ) can be determined (see Equations 61–65). The disadvantage of the multiplexing scheme is that, although the photodiode array is a parallel detector, the pixels are sampled serially. Thus, the repetition time is 40 times longer than the acquisition time.

Finally, advanced versions of the rotating element multichannel ellipsometer have been developed as well that allow high speed collection of spectra in the reflectance R and/or degree of polarization p of the beam in addition to (ψ, Δ).[79,80] For the rotating polarizer multichannel ellipsometer, at least three readouts of the photodiode array every polarizer half-rotation are sufficient to obtain (ψ, Δ) and R. In this case, it is necessary to measure the ellipsometer system response to obtain absolute reflectance spectra in real time studies of film growth. This is done using a starting substrate of known optical properties and structure. Finally, a rotating compensator multichannel ellipsometer has also been developed.[57] With this instrument, $\{(\psi, \Delta)$, and $R\}$ for the sample and p for the reflected beam can be derived from raw data collected in a time as short as 32 ms using a minimum of five readouts of the array every compensator half-rotation (see Figure 19). An application of this instrument for ex situ analysis of a transparent conducting oxide film was described in section 4.6.2.

5.2 Data Analysis

RTSE data analysis procedures follow the same principles as the ex situ analysis procedures of section 4.5. The first method, least squares regression analysis, allows extraction of the λ-independent parameters that describe the sample structure and optical properties at a given time during film growth or surface modification. With this approach, a database for the optical properties of one or more components of the sample may be required at the actual processing temperature of the sample. This requirement can lead to difficulties in the analysis particularly for processing in which the sample temperature changes. With the high speed data collection capability of instruments for RTSE, however, databases for materials of different composition and temperature, for example, can be established expeditiously. The second method of data analysis, exact mathematical inversion, yields the optical properties of one material as well as one or more λ-independent sample structural parameters characteristic of a given time in the process.

In many cases, the optical properties of one of the layers of the sample can be assumed to remain unchanged during some stage of film growth. In such cases, least squares regression and exact inversion methods can be combined in a global error function minimization procedure.[81] First, (ψ, Δ) spectra are chosen at a specific time t_1

during the deposition of the layer. Then trial values are chosen for the structural parameters associated with this layer, e.g. bulk, surface roughness, or interface roughness layer thicknesses. Next the (ψ, Δ) spectra at $t = t_1$ are subjected to exact inversion to deduce a trial dielectric function ε_j of the growing layer. ε_j is applied in least squares regression analysis of the (ψ, Δ) spectra collected for times $t_0 \leq t \leq t_1$ during which ε_j is assumed to be constant. The output of the regression analysis is the time dependence of the λ-independent structural parameters for $t_0 \leq t \leq t_1$. The error functions between the measured and best fit SE data, when summed for all spectra collected versus time $t_0 \leq t \leq t_1$, provide a measure of the quality of the global fit. Finally, the initial guesses for the structural parameters at $t = t_1$ are adjusted and the process is iterated until the sum of the error functions is minimized. Under the global best fit conditions, ε_j obtained by inversion at $t = t_1$ is correct, and the time dependences of the structural parameters obtained by least squares regression for $t_0 \leq t \leq t_1$ are also correct.

One might expect ideal uniform growth with thickness-independent optical properties for the preparation of heteroepitaxial crystalline films. Thickness dependences in the optical properties can occur, however, due to quantum size effects in thin layers or to strain relaxation in thicker layers. In the case of the quantum effects, the optical properties of the layer at a given time are uniform throughout the thickness, but these properties change, i.e. become more bulk-like, with increasing thickness. In the case of strain relaxation, the optical properties exhibit a gradient with thickness, such that the near-substrate and near-surface properties are characteristic of strained and relaxed material, respectively. Analogous problems occur in the deposition of nonepitaxial and noncrystalline films. The thickness dependences in the optical properties can occur in thin layers due to clustering during nucleation and to size effects. In thicker layers, the crystallite size or void fraction within the film may evolve with thickness. The optical properties during nucleation can be modeled in a similar manner to the size effects. Specifically, the layer optical properties at a given time are assumed to be uniform throughout the thickness, but these properties change continuously as the clusters increase in size and coalesce. In contrast, when structural evolution occurs in coalesced layers, the properties exhibit a gradient with thickness.

Here, analysis of RTSE data acquired during graded layer growth will be discussed. Graded layers can result from the natural structural evolution of a film (as just described) or from intentional compositional variations imposed continuously during growth. Intentional variations can be obtained in chemical vapor deposition, for example, by varying the source gas flows. The simplest graded layer analysis employs a one-layer model based on the assumption that the interface to the ambient is perfect and that a ''virtual interface'' exists at a depth d_o below the film surface.[82] The virtual interface separates the most recently deposited material of thickness d_o called the outerlayer, from the underlying material which includes the past history of the deposition. Thus, the underlying material is treated as a ''pseudo-substrate'' characterized by a pseudodielectric function. The one-layer or pseudosubstrate approximation is exact in the limit of very thin outerlayers on opaque uniform substrate films. The goal of graded-layer analysis is to extract the deposition rate r_d and the outerlayer composition x, assuming that a parameterization scheme exists that allows the dielectric function to be determined given x.

The graded-layer analysis begins by considering the first N (odd) spectra $\{(\psi_m, \Delta_m); m = 1, \ldots, N\}$ collected in a time δt starting from t_0 at which graded layer growth begins.[82,83] These spectra are converted into pseudodielectric functions $\{\langle\varepsilon_m\rangle; m = 1, \ldots, N\}$, using Equation (23). First, trial values are chosen for r_d and x appropriate for the interval δt. Then the derivative of $\langle\varepsilon\rangle$ with respect to outerlayer thickness d_o, denoted $\langle\varepsilon\rangle'$ and evaluated at $t = t_0 + \delta t/2$, is determined from $\langle\varepsilon_m\rangle$ by N-point linear regression. Next, the pseudosubstrate approximation

$$\langle\varepsilon\rangle' = \frac{4\pi i \langle\varepsilon\rangle_c (\langle\varepsilon\rangle_c - \varepsilon_0)(\varepsilon_0 - 1)(\langle\varepsilon\rangle_c - \sin^2\phi_i)^{1/2}}{\lambda\varepsilon_0(\langle\varepsilon\rangle_c - 1)} \quad (69)$$

is applied where ε_0 is the outerlayer dielectric function defined by x. Equation (69) provides $\langle\varepsilon\rangle_c = \langle\varepsilon\rangle_\mu$ [$\mu = (N+1)/2$], the calculated pseudodielectric function at center (c) of the data collection range ($t = t_0 + \delta t/2$), which is in turn converted into (ψ_μ, Δ_μ). The trial values of x and r_d are adjusted to minimize an error function between the measured and calculated (ψ_μ, Δ_μ) spectra. The overall procedure is repeated for successive N-member pairs of real time (ψ, Δ) spectra to determine the best-fit x and r_d versus time from which the depth profile in x can be deduced.

This analysis has two limitations that have been addressed in refinements and alternative approaches. First, it is assumed that the top surface of the sample is atomically smooth. In a refinement of the above procedure, a two-layer analysis has been developed that includes a surface roughness layer on top of the outerlayer.[83] The roughness layer is modeled as a 0.5/0.5 volume fraction mixture of the outerlayer and void using the Bruggeman theory, and its thickness is determined in the analysis. The results of such a depth-profiling study for graded amorphous silicon carbon alloys is given in section 6. Second, for Equation (69) to be valid, the underlying film structure must be opaque. Thus, the procedure is not valid for graded weakly absorbing materials.

In order to handle such situations, in general, one must resort to an analysis in which film growth is divided into small segments within which the optical properties are assumed to be constant.[84] Then, a procedure based on inversion and least squares regression can be applied.

5.3 Applications in Gaseous and Plasma Environments

In this section, the focus will be on applications of RTSE in film growth environments that preclude other real time probes used in high vacuum (see section 6). Such environments include chemical vapor deposition that utilizes relatively high pressure reactive gases. As an example, recent studies of diamond film growth by microwave plasma-enhanced chemical vapor deposition will be described.[85] Such studies were performed using rotating-element multichannel ellipsometers, and analysis by least squares regression has provided a number of parameters of interest including (i) the near surface temperature of the substrate, (ii) the bulk layer deposition rate, (iii) the evolution of the surface roughness layer thickness, (iv) the composition of the final film, and (v) the thickness uniformity of the film over the beam area. In fact, the unique ability to deduce near surface temperature and film growth rate permits an improved understanding of the kinetics and chemical mechanisms for film growth using different gas mixtures.

Temperature measurements of the silicon substrates were performed by RTSE during their exposure to the plasmas used for diamond film growth. The induction time for diamond nucleation on the substrate was longer than the time required for the substrate to reach thermal equilibrium with the plasma. Thus, in the analysis to determine the substrate temperature T a one-layer structure could be assumed, consisting of a native oxide on Si. The Si dielectric function was deduced from the (ψ, Δ) spectra using exact numerical inversion in which the oxide thickness was also determined. To obtain T, the second derivative of the Si dielectric function in the neighborhood of the sharp feature near 3.3 eV, reflecting the band structure critical point at Λ in k-space, was fitted to an analytical form for an excitonic transition, namely,

$$\frac{d^2\varepsilon}{dE^2} = \frac{2A_1 \exp(i\theta_1)}{(E - E_1 + i\Gamma_1)^3} \tag{70}$$

in order to extract the amplitude A_1, phase θ_1, resonance energy E_1, and broadening parameter Γ_1. A prior study has reported that E_1 decreases linearly with T according to $E_1(\text{eV}) = 3.375 - [4.07 \times 10^{-4} T(^\circ\text{C})]$, from which T can be obtained by substituting E_1.[86] At $E = E_1$, the absorption depth (α^{-1}) is $\sim$20 nm, and $\sim 3\alpha^{-1}$ is the surface region from which T is measured. As a result, the temperature obtained by SE is a true characteristic of the Si surface. Figure 20 shows results, in which the error ΔT, i.e. the difference between the true temperature T determined by RTSE and the nominal value from a thermocouple, is plotted versus T for microwave plasmas at 0.5 kW power using different CO/H_2 gas flow ratios at a pressure of 10 Torr.

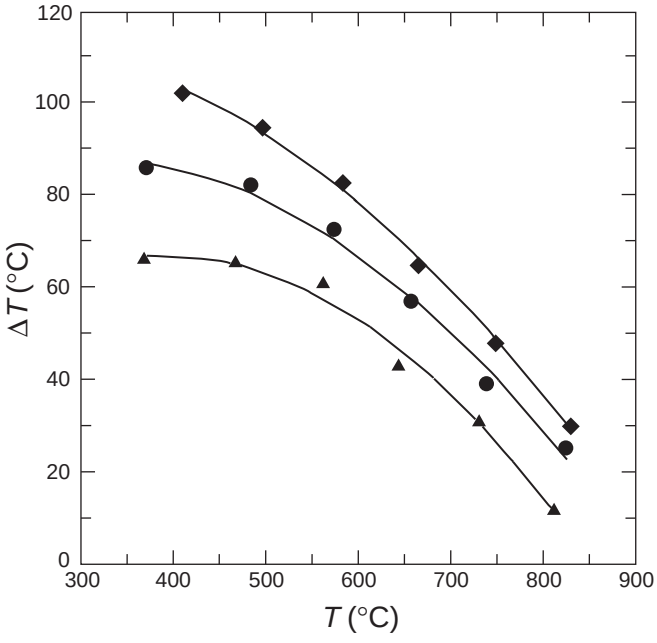

Figure 20 The RTSE determined correction ΔT that must be added to the thermocouple readout in order to obtain the true near-surface temperature T of Si wafers used for diamond growth. The correction is plotted versus T for different $[CO]/[H_2]$ gas flow ratios: , 0.2; ●, 2; , 8. (Reproduced from J. Lee, B. Hong, R. Messier, R.W. Collins, *Thin Solid Films*, **313–314**, 506–510. Copyright (1998) with permission of Elsevier Science.[85])

Figure 21 shows RTSE analysis results for diamond film growth using a CO/H_2 flow ratio of 4.7, a temperature of 375 °C, and a pressure of 10 Torr.[85] The data were collected with a rotating-polarizer multichannel ellipsometer. The substrate in Figure 21 was a diamond film with a $\sim$16 nm thick roughness layer prepared on a Si substrate. Because of the roughness layer on the underlying film, analysis of the RTSE data by least squares regression employed first a two-layer model [ambient/(surface roughness)/(interface roughness)/substrate] that simulates interface formation between the underlying structure and the new film, and then a three-layer model [ambient/(surface roughness)/bulk/(interface roughness)/substrate] that simulates bulk film growth. In the interface formation regime, the two free parameters include (i) the volume fraction of new diamond (f_{di}) that forms within the modulations of the substrate roughness and (ii) the thickness of the diamond layer (d_s) that forms above the substrate roughness. In the bulk growth regime, the four free parameters include (i) the bulk layer

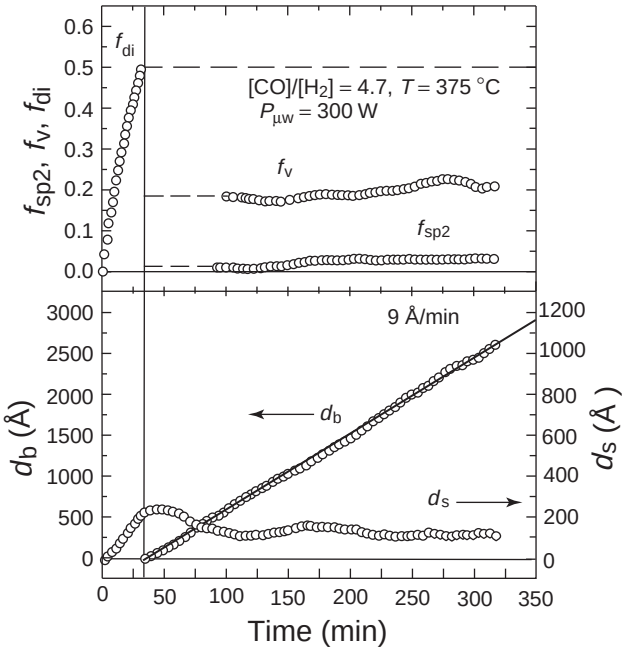

Figure 21 Structural evolution of diamond film growth on diamond, deduced from RTSE data collected with a $[CO]/[H_2]$ gas flow ratio of 4.7 and $T = 375\,^{\circ}C$. Structural parameters include (top panel) the volume fraction of diamond within the modulations of the substrate surface roughness (f_{di}), the volume fractions of sp^2 C (f_{sp^2}) and void (f_v) in the bulk layer, and (bottom panel) the thicknesses of the bulk (d_b) and surface roughness (d_s) layers. The horizontal broken lines (top panel) indicate fixed compositions. (Reproduced from J. Lee, B. Hong, R. Messier, R.W. Collins, *Thin Solid Films*, **313–314**, 506–510. Copyright (1998) with permission of Elsevier Science.[85])

thickness (d_b), (ii) the surface roughness layer thickness (d_s), (iii) the bulk layer void volume fraction (f_v), and (iv) the bulk layer graphitic or sp^2 C carbon volume fraction (f_{sp^2}). The latter undesirable component in diamond films is simulated with glassy carbon (pure sp^2 C) whose dielectric function is known from a previous database.

In the two-layer film structure, the interface roughness layer is modeled using the Bruggeman theory as a two-component diamond/void composite in which the composition varies with time. This approach simulates the filling of the modulations in the roughness layer of the substrate film upon growth of the new film. In the three-layer structure, the surface roughness layer is modeled as a 50/50 vol.% mixture of [diamond + void], whereas the bulk layer is modeled as a [diamond + (sp^2 C) + void] composite. For bulk layer thicknesses <50 nm, the bulk layer composition was fixed (broken lines in Figure 21) to ensure stability of the analysis. The results in Figure 21 are consistent with a process in which the substrate interface fills in as the diamond film roughness layer develops to ~20 nm. In the bulk regime, the roughness layer relaxes to 11 nm as the bulk layer thickness increases linearly with time at $0.9\,nm\,min^{-1}$. In this regime, the bulk composition is stable (0.78 ± 0.04 diamond, 0.20 ± 0.03 void, and 0.02 ± 0.01 sp^2 C).

The analysis of Figure 21 can be performed for each of a series of diamond films deposited on the same substrate under different conditions.[85] The results of such analyses provide the time evolution of the diamond mass thickness, i.e. the diamond volume per area, as a function of the conditions. The mass thickness is given by $d_m = f_{di}d_i + f_{db}d_b + f_{ds}d_s$, where f_{dv} is the fraction of diamond within the layer of thickness d_v, where (v = i,b, and s) designate the interface roughness, bulk, and surface roughness layers, respectively. Figure 22 shows the mass thickness rate versus the true near surface temperature for microwave plasma-enhanced chemical vapor deposition of diamond using different gas mixtures at a pressure of 10 Torr and a plasma power of 0.5 kW. When CO or CH_4 is highly diluted in H_2, the rate increases with T in accordance with an $8\,kcal\,mol^{-1}$ activation energy (solid lines) that appears to suggest a common hydrocarbon film precursor such as CH_3. As the CO/H_2 ratio is increased to 18, the T-dependence of the rate is no longer activated and a sharp maximum develops near 500 °C. This behavior has suggested a new diamond growth mechanism from nonhydrocarbon precursors.

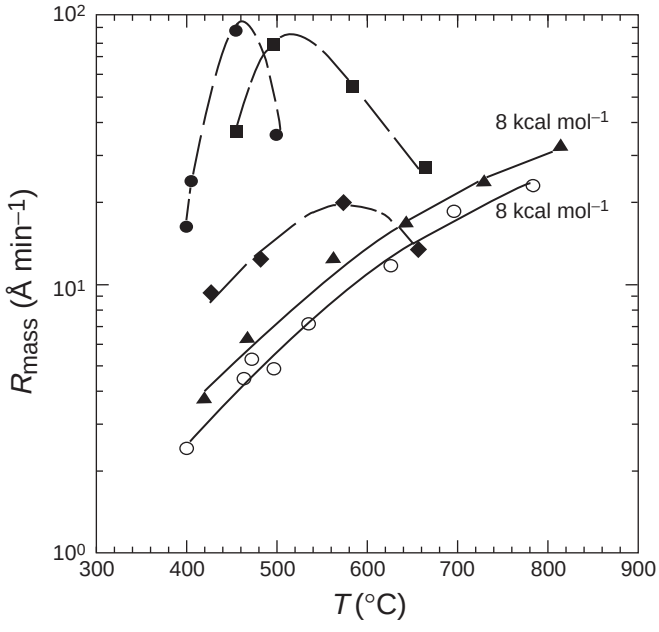

Figure 22 Growth rate of diamond as a function of near-surface temperature using microwave plasma-enhanced chemical vapor deposition with different $[CO]/[H_2]$ (, 0.2; , 2; ■, 8; ●, 18) and $[CH_4]/[H_2]$ (○, 0.01) gas flow ratios. The growth rate is obtained from RTSE analysis results (see Figure 21). (Reproduced from J. Lee, B. Hong, R. Messier, R.W. Collins, *Thin Solid Films*, **313–314**, 506–510. Copyright (1998) with permission of Elsevier Science.[85])

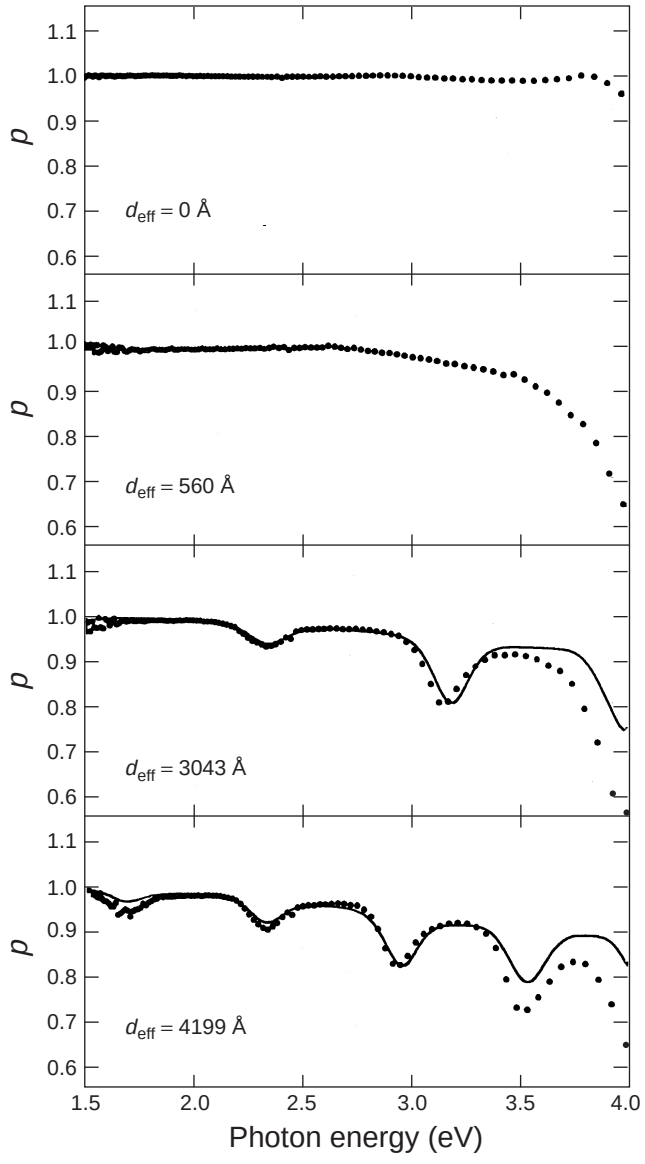

Figure 23 Reflected beam degree of polarization measured in real time for a Si substrate and for three effective thicknesses of diamond. The solid lines are simulations employing a single variable parameter, the thickness gradient of the diamond film over the beam area. (Reproduced with permission of American Institute of Physics from J. Lee, P.I. Rovira, I. An, R.W. Collins, *Rev. Sci. Instrum.*, **69**, 1800–1810 (1998).[80])

Figure 23 shows spectra in the reflected beam degree of polarization p for the Si substrate and for selected mass thicknesses of diamond obtained using a rotating-compensator multichannel ellipsometer in real time during film growth.[56] The diamond film was deposited at $T = 812\,^{\circ}\mathrm{C}$ with a CO/H_2 flow ratio of 0.022. Upon initial nucleation of diamond, there is a significant decrease in p from unity over the higher energy range. This can be attributed to multiple scattering of the beam within the largest diamond crystallites which are on the order of the wavelength. The features in the two spectra in the thick film regime in Figure 23 can be attributed to thickness variations over the surfaces of the films. A simplified one-layer model for the film was used to simulate the spectra in p, assuming a linear thickness gradient over the area probed by the beam as shown in the inset of Figure 24.[56] Using this model, the probed area was partitioned into $2N_d + 1$ slices having thicknesses $d = d_{av} + M\Delta d$, where d_{av} is the average thickness, $\Delta d = 0.3$ nm is the thickness increment between slices, and M is an integer whose value ranges from $-N_d$ to N_d. Irradiance waveforms were generated for all $2N_d + 1$ thicknesses using the optical properties of the media as deduced from the (ψ, Δ) spectra along with trial values for the thickness variation, $2N_d\Delta d$, and the average thickness d_{av}. The irradiance waveform for each thickness was weighted by the area fraction, and then all waveforms were added, assuming incoherent superposition. The calculated spectra for p were then compared with the data in an attempt to match the features. Best fit simulations are shown in Figure 23. Figure 24 includes the resulting film thickness gradient, derived from the beam diameter of 3.5 mm, versus the mass thickness. When one or more well-defined minima in p are observed, the uncertainty in the gradient is $\sim\pm 1.5\,\mathrm{nm\,cm^{-1}}$.

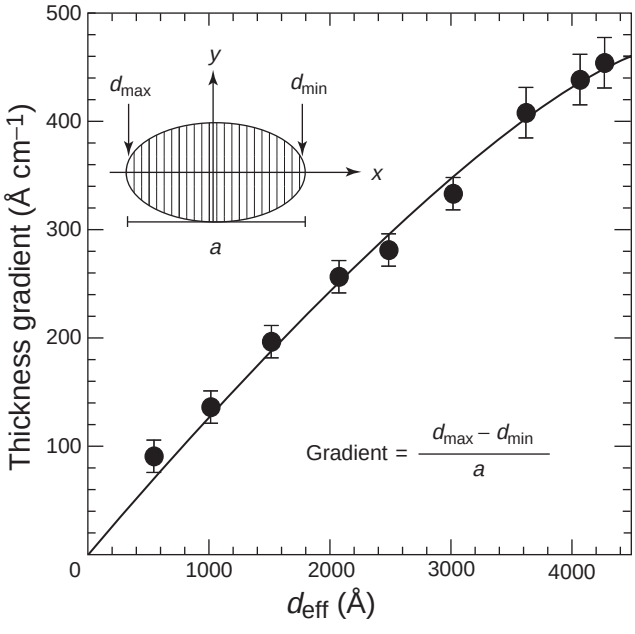

Figure 24 Diamond film thickness gradient over the area probed by the incident beam, plotted versus effective film thickness. The inset shows a schematic of the cross-section of the beam on the film surface (x–y plane) and the orientation of the assumed linear thickness gradient. (Reproduced with permission of American Institute of Physics from J. Lee, P.I. Rovira, I. An, R.W. Collins, *Rev. Sci. Instrum.*, **69**, 1800–1810 (1998).[80])

5.4 Applications in Liquid Environments

Applications of RTSE will be described next for analysis of film growth and modification in liquid environments. Such environments preclude in situ probes based on conventional electron and ion spectroscopies. Although liquid ambient media often reduce its spectral range, RTSE provides sufficient sensitivity for characterizing film formation at the monolayer level as well as small changes in the composition of bulk materials (±0.2 vol.%). Most applications of RTSE for studies of surfaces in liquids are in the field of electrochemistry and involve research into processes of adsorption, oxidation, passivation, corrosion, polishing, and deposition. Here, three applications are described, including electrodeposition of lead on a copper substrate, electrochemical oxidation of a silver crystal surface, and electrochromic reduction of WO_3.

The first application of RTSE using the rapid scanning, self-nulling instrument was for studies of plating additives in electrodeposition.[87] Figure 25 shows RTSE data plotted versus λ for potentiostatic deposition of Pb on a polycrystalline Cu electrode. An aqueous electrolyte containing $Pb(NO_3)_2$ and $NaClO_4$ without additives was used for electrodeposition and each pair of spectra in (ψ, Δ) was collected in 3 s. The values (0, 31, 60, 110) alongside the curves in Figure 25 represent the expected thicknesses (in nm) of the Pb film as measured by coulometry, i.e. by charge balance assuming a film of single crystal density. An analysis of the RTSE data, however, found that this was not the true physical thickness of the layer due to a complicated, porous

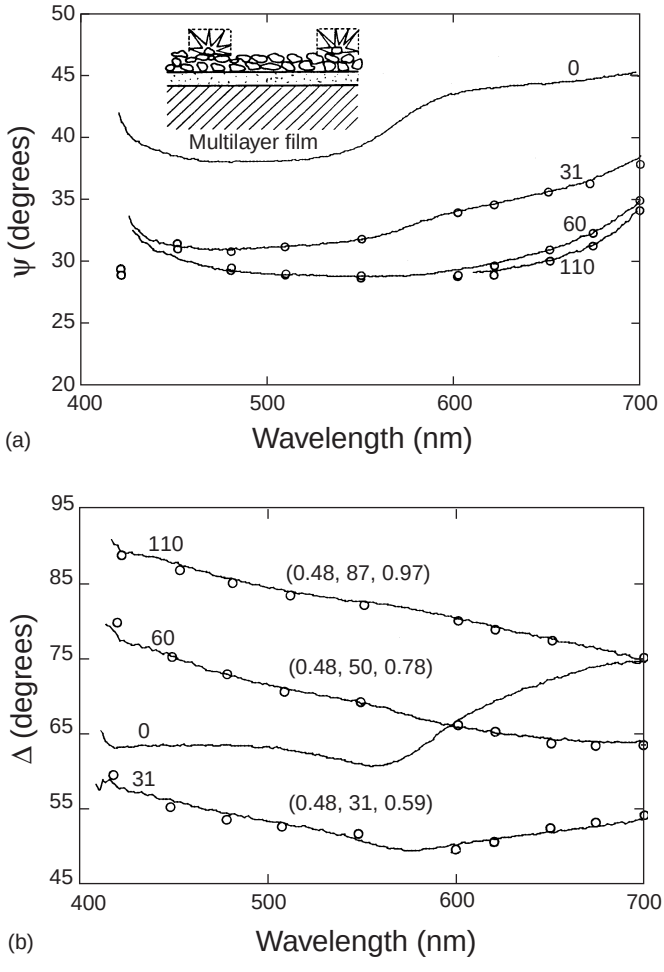

Figure 25 RTSE data in (a) ψ and (b) Δ (lines) and best fit modeling results (○) for the electrodeposition of Pb on Cu. The single values alongside the curves are the equivalent deposit thicknesses (in nm) deduced from coulometry, and the three values in parentheses are the thicknesses (in nm) of the near-substrate compact and overlying porous Pb layers, and the porous layer Pb volume fractions deduced from RTSE. The inset shows the model consistent with RTSE and electron microscopy. (Reproduced with permission of The Electrochemical Society from J.C. Farmer, R.H. Muller, *J. Electrochem Soc.*, **132**, 313–319 (1985).[87])

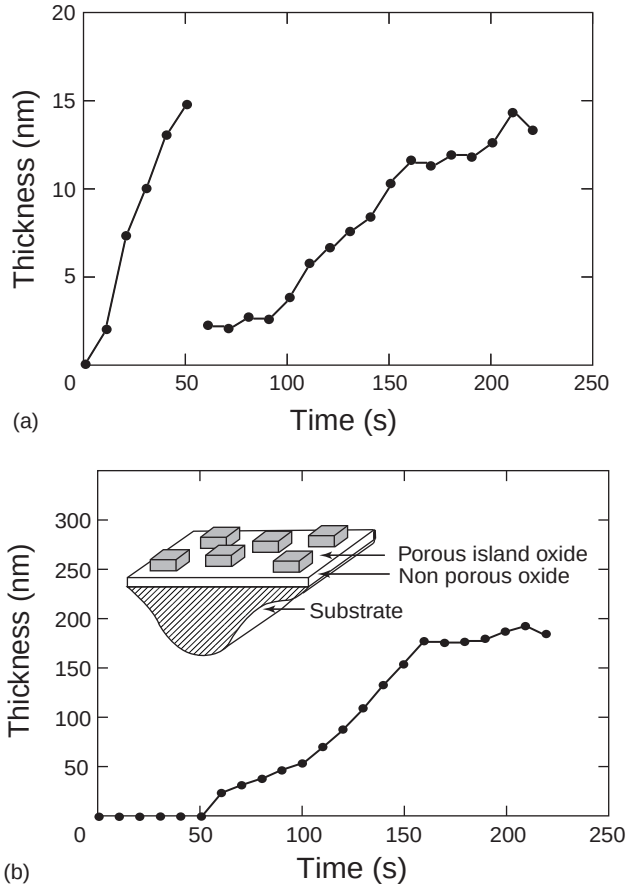

Figure 26 Thicknesses of (a) dense primary and (b) porous secondary Ag_2O layers versus time deduced from RTSE data using a two-layer model for anodic oxidation of a Ag(111) surface. The optical model used to interpret the RTSE data is shown in the inset. (Reproduced with permission of The Electrochemical Society from S.T. Mayer, R.H. Muller, *J. Electrochem Soc.*, **135**, 2133–2142 (1988).[88])

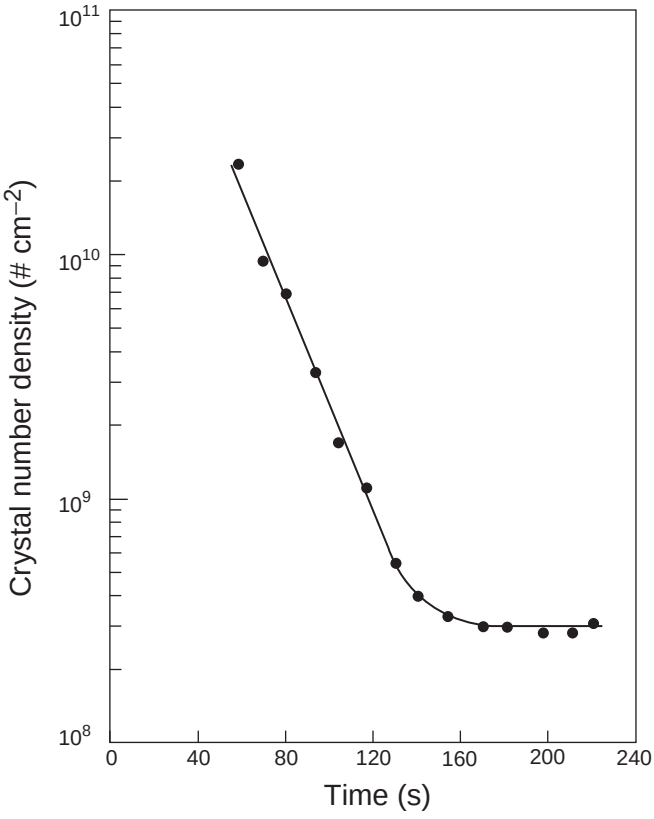

Figure 27 Number density of crystallites making up the porous Ag_2O layer plotted versus time during the anodic oxidation of a Ag(111) surface at $i = 0.10\,\text{mA cm}^{-2}$. These results were deduced from RTSE data using a two-layer model for oxidation. (Reproduced with permission of The Electrochemical Society from S.T. Mayer, R.H. Muller, *J. Electrochem Soc.*, **135**, 2133–2142 (1988).[88])

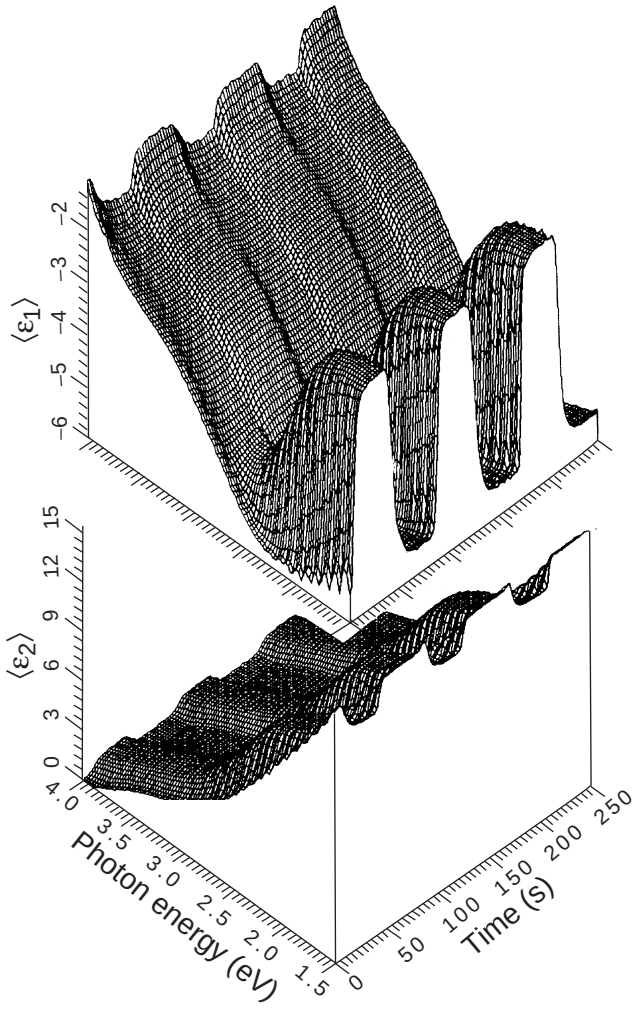

Figure 28 RTSE data expressed in terms of the real and imaginary parts of the pseudodielectric function collected during three color–bleach cycles for a 15 nm WO_3 film on a Cr substrate film. (Reproduced with permission of the American Institute of Physics from H. Witham, P. Chindaudom, I. An, R.W. Collins, R. Messier, K. Vedam, *J. Vac. Sci. Technol.*, A, **11**, 1881–1887 (1993).[89])

microstructure. The inset in Figure 25(a) shows the optical model that provides the best fit to the data. This model includes (i) a compact layer of Pb at the substrate interface, (ii) a porous Pb layer with optical properties computed as a mixture of bulk Pb and electrolyte using the Bruggeman theory, and (iii) islands of Pb at the top surface whose optical effect is calculated as a coherent superposition of two polarization states from surface elements with and without island coverage. Overall, the best fit model results were found to be in good agreement with the data and consistent with coulometry and scanning electron microscopy. Since the coverage of the porous layer by islands was found to be low (~5–10%) and the size of the islands was <10 μm, the optical effect of this layer on the polarization was much weaker than the other layers (see section 4.4). The values in parentheses alongside the data and fits in Figure 25(b) are the best fit thicknesses (in nm) of the compact and porous layers, and the volume fraction of Pb in the porous layer, respectively. From the information on the distribution of Pb between compact, porous, and island structures, the effect of plating additives could be characterized in detail.

The same instrument for RTSE was also used to study initial oxidation of Ag(111) surfaces to the Ag_2O phase upon application of a potential step.[88] Oxidation was performed in an aqueous electrolyte with KOH, and each pair of (ψ, Δ) spectra consisting of 40 points was collected in 3.5 s. The optical model found to provide the best overall fit to the RTSE data is shown in the inset of Figure 26(b) and consists of a dense oxide film (primary layer) at the interface to the Ag, modeled as a homogeneous layer, and a surface layer of crystalline nuclei (secondary layer) modeled assuming coherent superposition. In Figure 26(a) and (b), the evolution

of the two layer thicknesses are shown versus time after application of the potential step. The thickness of the dense oxide layer initially increases, but decreases sharply at the onset of crystallite nucleation. This suggests that oxide crystallites nucleate at sites within the dense layer and then oxide mass is redistributed by dissolution of the dense layer, followed by precipitation at the crystallites. Figure 27 shows the time dependence of the crystallite number density on the surface. These results can be derived from the crystallite layer thickness and surface coverage, assuming hemispherical crystallites. The behavior in Figure 27 also appears to be another manifestation of dissolution and precipitation via an Ostwald ripening phenomenon. Through these studies RTSE has shown that the oxidation process is more complex than had been previously proposed.

In the electrochromic effect for WO_3, the optical properties of the material immersed in an electrolyte are altered by dual injection (''coloring'') or emission (''bleaching'') of electrons and ions. Figure 28 shows RTSE data collected using a rotating-polarizer multichannel ellipsometer (with acquisition and repetition times of 0.2 s and 1 s) during three 82 s color–bleach cycles for a 15 nm WO_3 film prepared on a Cr contact layer by sputtering.[89] The electrolyte was an aqueous solution containing H_2SO_4, and the coloring voltage was applied for 1 s and held for 40 s. From these data, the dielectric function for the bleached state (WO_3) can be deduced by exact inversion using a single-layer model in which the thickness is chosen to ensure that $\varepsilon_2 = 0$ in the visible as shown in Figure 29. The dielectric function for the colored state (H_xWO_{3-y}) in Figure 29 is obtained assuming that the thickness does not change upon coloring. Using these results, the kinetics of coloring can be determined as shown in Figure 30. In this figure, the correct kinetic model is identified from the evolution of the error function, computed in this case as the unbiased estimator of the mean square deviation between the experimental and best fit ($\tan\psi$, $\cos\Delta$) spectra. In the diffusion model, a colored layer in contact with the electrolyte is assumed to increase in thickness at the expense of the underlying bleached layer. The interface between the layers represents the diffusion front of H^+ ions that react with electrons in WO_3 to form tungsten bronze, H_xWO_{3-y}. In the reaction-limited model, a composite of the colored (H_xWO_{3-y}) and bleached (WO_3) phases is assumed to exist in which the optical properties of the composite are calculated from the Bruggeman theory. The better agreement with a reaction limited model is consistent

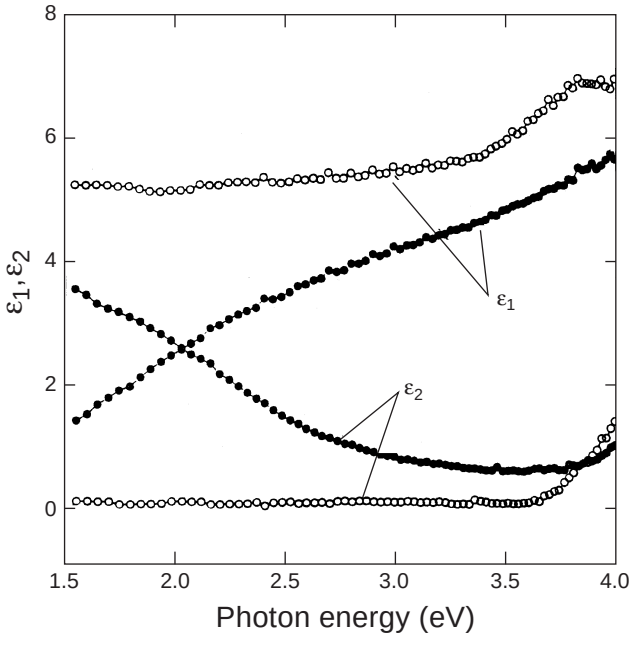

Figure 29 Dielectric functions of a 15 nm electrochromic WO_3 film in the bleached (○) (WO_3) and colored (•) (H_xWO_{3-y}) states, obtained from a one-layer analysis of the RTSE data (see Figure 28). (Reproduced with permission of the American Institute of Physics from H. Witham, P. Chindaudom, I. An, R.W. Collins, R. Messier, K. Vedam, *J. Vac. Sci. Technol.*, A, **11**, 1881–1887 (1993).[89])

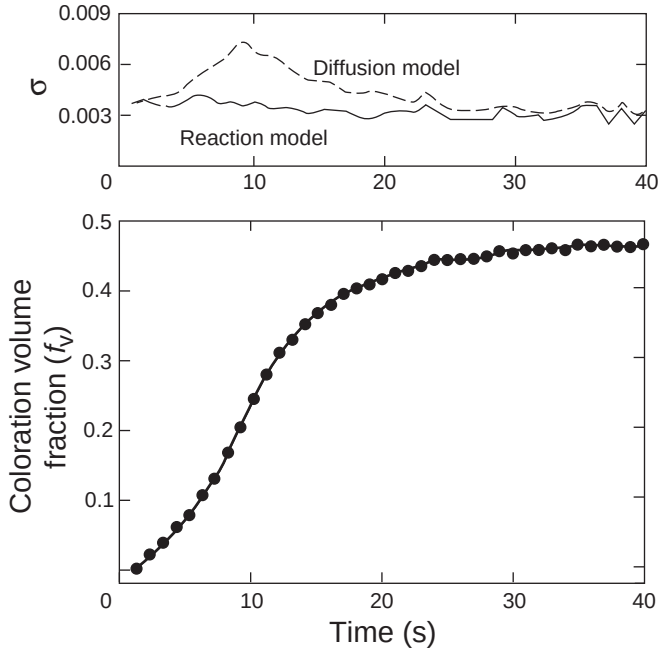

Figure 30 Time evolution of the volume fraction of the H_xWO_{3-y} phase during electrochemical coloration, obtained from an analysis of RTSE data applying a reaction-limited model $[(1-f_v)WO_3 + (f_v)H_xWO_3]$ for the formation of the colored state. At the top the error function is shown versus time for RTSE analyses assuming reaction and diffusion limited models for coloration. (Reproduced with permission of the American Institute of Physics from H. Witham, P. Chindaudom, I. An, R.W. Collins, R. Messier, K. Vedam, *J. Vac. Sci. Technol.*, A, **11**, 1881–1887 (1993).[89])

with a mechanism whereby coloring is limited by diffusion of ions parallel to the film plane from void channels that extend throughout the film.

6 COMPARISON WITH OTHER TECHNIQUES

Ellipsometry in its various forms is used most widely to determine film thicknesses from partial monolayer coverage of a substrate surface (~0.01 nm) to thick coatings (~10 μm). The uniqueness of ellipsometry is its ability to perform such measurements in situ and in real time with the sample immersed in any medium that is transparent to the incident beam. It should be recalled, however, that the entire optical problem must be solved (i.e. the dielectric functions of all the media must be known or determined) in order to obtain accurate thickness values. Several alternative techniques are also available for both ex situ and real time thickness measurement.[90]

Among ex situ measurements, stylus profilometry is the most widely used method for film thickness determination owing to its simplicity. Because it can provide thickness without the need to extract other information, nonideal and nonspecular thin films can be measured. Furthermore, it can be used when optical measurements of opaque films fail. A disadvantage of profilometry is that it cannot provide accurate thicknesses less than about 5 nm, and even for uniform films >10 nm, ellipsometry under optimum conditions is more than an order of magnitude more accurate. In addition, an abrupt step must be generated on the sample surface; thus, profilometry can be considered an invasive or destructive method. Cross-sectional transmission electron microscopy (XTEM) or edge-on scanning electron microscopy can also be used as a destructive method of thickness determination. The advantage of microscopy is that it provides a direct image; the disadvantage is the time-consuming sample preparation. As an example, Figure 31 shows the agreement between XTEM and SE analyses of the layered structure of an ion-implanted Si wafer.[91,92] Nondestructive optical techniques that do not rely on polarization state measurements can also be applied to determine film thicknesses. Transmittance and reflectance spectroscopy can be used; however, for optically thin films ($\alpha d \ll 1; nd \ll 1$), transmittance is insensitive to the presence of the film. In this case, reflectance alone cannot provide sufficient information to solve the entire thin film problem unless the optical properties of the film are known.

Alternative techniques are also available for in situ and real time thickness measurements. The most widely used is the quartz crystal oscillator that measures mass accumulation during film growth. The advantage of this method is that it yields the mass accumulation at the monolayer level without the need to extract other information. The disadvantage is that the measurement is performed for deposition on the quartz oscillator surface, not on the substrate of interest. For cases in which film growth rates are sensitive to the substrate, the usefulness of the oscillator method is limited. In

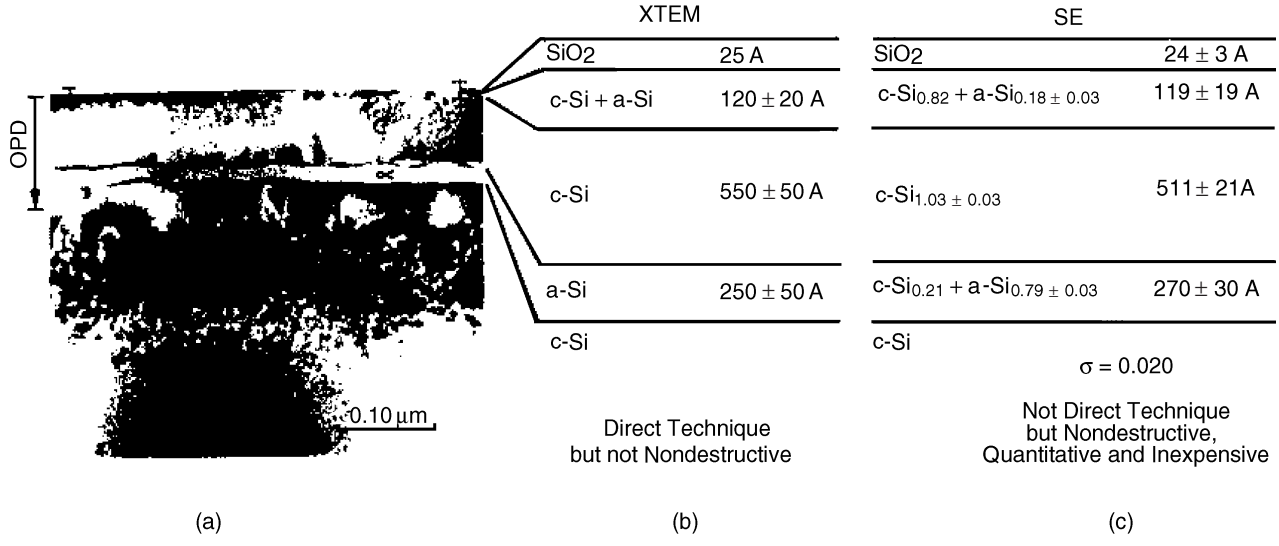

Figure 31 Comparison of (a, b) XTEM and (c) ex situ SE analysis of a Si wafer implanted with 80 keV Si^+ ions at a rate of 160 mA cm^{-2} with a total dose of 1×10^{16} cm^{-2}. OPD designates the maximum optical penetration depth accessible by SE. c-Si and a-Si designate the crystalline and amorphous phases of silicon and the subscripts indicate the volume fractions of the phases (Reproduced with permission of the American Institute of Physics from K. Vedam, P.J. McMarr, J. Narayan, *Appl. Phys. Lett.*, **47**, 339–341 (1985).[92])

addition, one may be interested in the physical thickness, which is different from the mass thickness deduced by the oscillator, owing to film porosity. In contrast, SE can provide the physical thickness as well as microstructural information. Another disadvantage of the oscillator method is that it cannot be used in reactive, high pressure environments such as exist in plasma-enhanced chemical vapor deposition. A second popular real time technique for determining film thickness is reflection high energy electron diffraction. This technique is used widely to characterize monolayer-by-monolayer growth of single crystal layers in UHV; however, it fails for noncrystalline materials and at high pressures. Finally, reflectance can also be used for determining film thickness evolution in real time. This method is advantageous because of its simplicity; however, it is less sensitive than ellipsometry for monolayer level films and the information on the optical properties of the deposited material is limited. The sensitivity of reflectance can be enhanced by using a *p*-polarized incident beam at an angle of incidence equal to the pseudo-Brewster angle.

As has been made clear in sections 4.5 and 5.2, the usefulness of ellipsometry is not restricted to bulk film thickness determination. This technique can also provide information on roughness, structure, composition, and temperature. Detailed comparison of alternative techniques for these characteristics is beyond the scope of this discussion; however, general comments are in order. For determining surface roughness and bulk film microstructure, ellipsometry is an indirect method and is often described as model dependent because the analysis results depend on which effective medium theory is applied. Thus, it is prudent to corroborate the results using an alternative technique. As an example, Figure 32 shows the correlation between the surface roughness layer thickness deduced by RTSE at the end of film deposition and by AFM ex situ after deposition (the latter expressed as the root mean square roughness).[93] The same set of amorphous silicon–carbon alloy films were used for both measurements. A good linear correlation is observed, particularly for microscopy performed in the tapping mode, indicating that the SE-determined roughness thickness is 1.5 times larger than the rms roughness. Such results promote confidence in the RTSE method for situations in which AFM cannot be applied, such as for characterization of the roughness evolution during deposition. Other parameters deduced by SE using indirect methods, such as the bulk film void volume fraction, may be difficult to corroborate by alternative techniques. In this case, special efforts must be undertaken to ensure that the effective medium theory is valid, e.g. by collecting SE data over the widest possible spectral range and selecting the theory that provides the best fit.

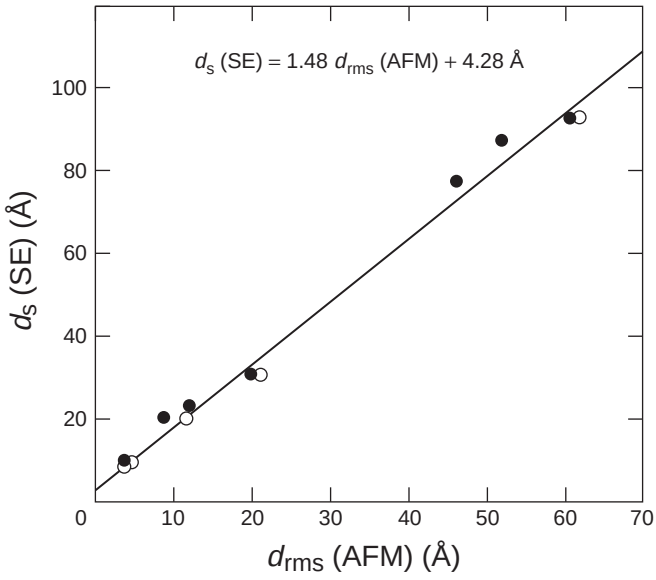

Figure 32 Surface roughness measurements for a-$Si_{1-x}C_x$: H films on c-Si ($d_s < 35$ Å) and SnO_2 : F ($d_s > 70$ Å) substrates. Roughness layer thicknesses were deduced independently from analyses of RTSE data d_s (SE) and ex situ atomic force microscopy images d_{rms} (AFM) for the same samples. AFM images were obtained with microscopes operated in contact (●) and tapping (○) modes. (Reproduced with permission of the American Institute of Physics from J. Koh, Y. Lu, C.R. Wronski, Y. Kuang, R.W. Collins, T.T. Tsong, Y.E. Strausser, *Appl. Phys. Lett.*, **68**, 1297–1299 (1996).[93])

In contrast to composites in which effective medium theory can be applied, mixtures on the atomic scale (i.e. alloys) cannot be analyzed by ellipsometric methods to determine composition directly. For alloys, a series of samples must be measured first using X-ray diffraction or electron spectroscopy to obtain the composition and second using SE to obtain the optical properties. Then the dielectric function (or its derivatives) can be expressed analytically in terms of λ-independent parameters that in turn can be described as quadratic functions of composition. Only with such a database, can SE be applied as a routine analytical tool for alloy composition. As an example, Figure 33(a) provides the depth-profile in the composition x, deduced from RTSE data for a graded amorphous silicon–carbon alloy (a-$Si_{1-x}C_x$: H) prepared by plasma-enhanced chemical vapor deposition using SiH_4 and CH_4 gases.[94] The graded layer is obtained using a triangular-shaped profile for the gas flow ratio $z = [CH_4]/\{[SiH_4] + [CH_4]\}$ versus time. Also shown in Figure 33(a) is the result of a simulation based on reactor modeling. Good agreement between the simulation and data supports the validity of the analysis, based on Equation (69). For further corroboration, Figure 33(a) also includes a depth-profile for the same sample deduced in an ex situ destructive analysis by SIMS. SIMS yields a broader profile as a result of the

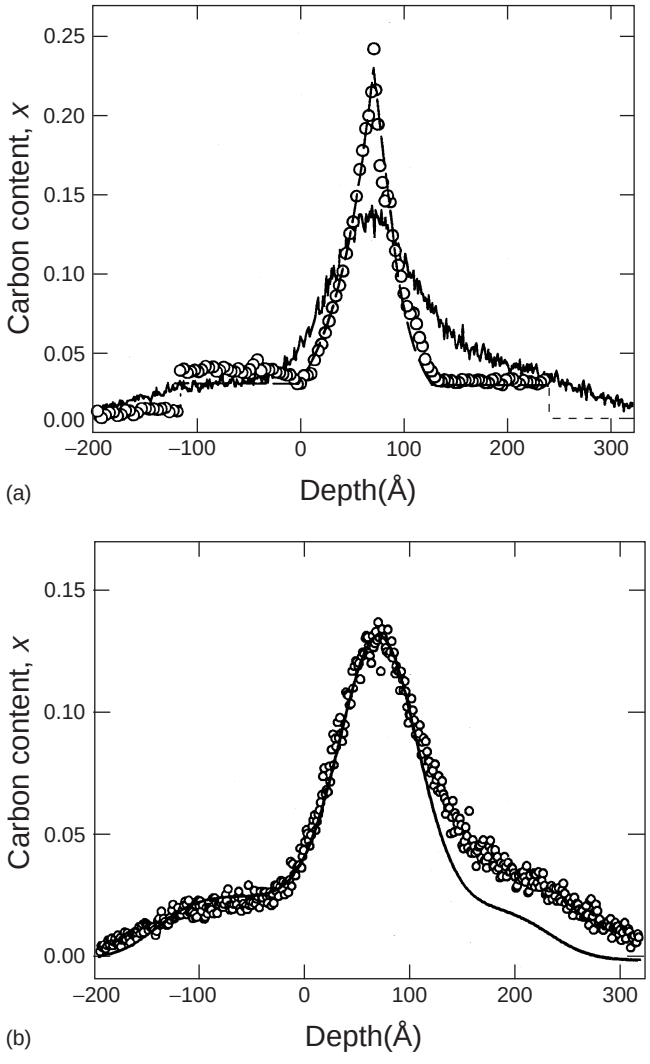

Figure 33 (a) Depth-profiles in the C-content x for an a-$Si_{1-x}C_x$: H graded layer structure obtained from RTSE (○) and from ex situ secondary ion mass spectrometry (SIMS) (solid line). The broken line is a fit to the RTSE profile using a deposition model. (b) Comparison of a Gaussian-broadened profile from RTSE (line) with the SIMS results (○). (Reproduced with permission of the American Institute of Physics from H. Fujiwara, J. Koh, C.R. Wronski, R.W. Collins, J.S. Burnham, *Appl. Phys. Lett.*, **72**, 2993–2995 (1998).[94])

limited resolution generated by the sputtering process used for depth-profiling. In Figure 33(b), the SIMS result is compared with a RTSE depth-profile Gaussian broadened by 44 Å. The excellent agreement provides further support for the validity of such SE analyses.

7 METHOD DEVELOPMENT

Although ellipsometry is a century-old analytical technique, most applications described here have become possible as a result of instrumentation advances since the late 1980s. Several of the most notable advances will be enumerated here. (i) The incorporation of linear photodiode array detectors into ellipsometers enable high-speed parallel spectroscopy. As a result RTSE from the ultraviolet to near-infrared can be employed to characterize thin film growth dynamics and spectroscopic ME can be employed for high resolution mapping of static surfaces. (ii) The incorporation of CCD cameras into ellipsometer designs enable imaging of surfaces at high resolution for real time studies of selective adsorption, deposition, oxidation, and corrosion. (iii) The development of higher speed instrumentation for use in the infrared range allows researchers to characterize the dynamics of surface reactions with a chemical specificity not possible from measurements in the ultraviolet to near-infrared range. (iv) The determination of other parameters in addition to the angles (ψ, Δ), such as the reflectance and the degree of polarization, can assist in studies of complex nonideal surfaces. (v) Finally, the last decade has seen numerous advances in data analysis capabilities to accompany the new measurement capabilities. Characterization of complex samples using ellipsometry, although not routine, is now possible, including those samples that exhibit anisotropy, structural inhomogeneity and compositional gradients throughout the thickness, nonuniformity over the beam profile, and light scattering due to structure on the scale of the wavelength.

In the absence of unforseen new directions in the field of ellipsometry, the limitations encountered in these recent advances are expected to set the focus for future developments. (i) Ellipsometers employing photodiode arrays will benefit from improvements in multichannel detector technology that will provide higher sensitivity, thus, enabling measurements of higher quality over a wider spectral range. Furthermore, the future may see a joining of multichannel ellipsometric and near-field microscopic techniques for much higher resolution in spectroscopic ME. (ii) Continuous advances in the speed of CCD cameras and associated software for image processing are expected to lead to improvements in IE. Thus, incorporation of a rapid scanning monochromator into the imaging ellipsometer may be possible for real time spectroscopic IE. Such anticipated advances are confined to the ultraviolet to near-infrared spectral range. (iii) Advances in measurement capability in the infrared range, desirable due to the chemical bonding information available, pose even greater challenges owing to the insensitivity of detection systems. In this case, brighter sources such as provided by free electron lasers are expected to make the greatest impact. (iv) Advances in multiparameter approaches may culminate with the development of a multichannel ellipsometer that provides spectra in the 16 elements of the sample's Muller matrix.

The Muller matrix provides a complete description of the effect of the sample on an incident polarization state of arbitrary Stokes vector. (v) Finally, future advances in data analysis are expected to exploit continued improvements in computational speed for the application of conventional data analyses in real time monitoring and control. Improved databases and more efficient algorithms are also expected for carrying out such analyses.

ACKNOWLEDGMENTS

The preparation of this article was supported in part by the U.S. National Science Foundation under Grant No. DMR-9622774. The author appreciates the technical assistance of Dr Joohyun Koh and Mr Pablo I. Rovira.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
AFM	Atomic Force Microscopy
CCD	Charge-coupled Device
HSA	Human Serum Albumin
IE	Imaging Ellipsometry
IRSE	Infrared Spectroscopic Ellipsometry
LPO	Lactoperoxidase
ME	Microellipsometry
PA	Preamplifier
PMT	Photomultiplier Tube
RTE	Real Time Ellipsometry
RTSE	Real Time Spectroscopic Ellipsometry
SE	Spectroscopic Ellipsometry
SIMS	Secondary Ion Mass Spectrometry
UHV	Ultrahigh Vacuum
XTEM	Cross-sectional Transmission Electron Microscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction • Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Peptides and Proteins **(Volume 7)**
Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction • Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis • Ultraviolet/Visible Spectroscopy in Process Analyses

Surfaces **(Volume 10)**
Surfaces: Introduction • Differential Reflectance Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Infrared and Raman Spectroscopy in Analysis of Surfaces • Photoluminescence in Analysis of Surfaces and Interfaces

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Infrared Spectroelectrochemistry • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques • Ultraviolet/Visible Spectroelectrochemistry

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Circular Dichroism and Linear Dichroism • Surface Measurements using Absorption/Luminescence

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

REFERENCES

1. R.M.A. Azzam, N.M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam, 1977.
2. R.W. Collins, D.E. Aspnes, E.A. Irene (eds.), *Spectroscopic Ellipsometry: Proceedings of the Second International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam, 1998.
3. A. Rothen, 'Measurements of the Thickness of Thin Films by Optical Means, from Rayleigh and Drude to Langmuir, and the Development of the Present Ellipsometer', in *Ellipsometry in the Measurement of Surfaces and Thin Films*, eds. E. Passaglia, R.R. Stromberg, J. Kruger, National Bureau of Standards Misc. Publ. 256, NBS, Washington, 7–21, 1963.
4. N.M. Bashara, A.B. Buckman, A.C. Hall (eds.), *Proceedings of the Symposium on Recent Developments in Ellipsometry*, North-Holland, Amsterdam, 1969.
5. D.E. Aspnes, 'Spectroscopic Ellipsometry of Solids', in *Optical Properties of Solids: New Developments*, ed. B.O. Seraphin, North-Holland, Amsterdam, 799–846, Chapter 15, 1976.
6. A. Roseler, *Infrared Spectroscopic Ellipsometry*, Akademie, Berlin, 1990.

7. K. Riedling, *Ellipsometry for Industrial Applications*, Springer, New York, 1988.
8. P. Drude, 'Ueber die Gesetze der Reflexion und Brechung des Lichtes an der Grenze Absorbirender Krystalle', *Ann. Phys. Chem. N.F. (Leipzig)*, **32**, 584–625 (1887).
9. P. Drude, 'Beobachtungen ueber die Reflexion des Lichtes am Antimonglanz', *Ann. Phys. Chem. N.F. (Leipzig)*, **34**, 489–531 (1888).
10. P. Drude, 'Ueber Oberflaechenschichten. I. Theil', *Ann. Phys. Chem. N.F. (Leipzig)*, **36**, 532–560 (1889).
11. P. Drude, 'Ueber Oberflaechenschichten. II. Theil', *Ann. Phys. Chem. N.F. (Leipzig)*, **36**, 865–897 (1889).
12. P. Drude, 'Bestimmung der Optischen Constanten der Metalle', *Ann. Phys. Chem. N.F. (Leipzig)*, **39**, 481–554 (1890).
13. E.D. Palik (ed.), *Handbook of Optical Constants of Solids*, Academic, Orlando, 1985.
14. L. Tronstad, 'The Validity of Drude's Optical Method of Investigating Transparent Films on Metals', *Trans. Faraday Soc.*, **31**, 1151–1158 (1935).
15. L. Tronstad, 'Optische Untersuchungen zur Frage der Passivitat des Eisens und Stahls', in *Det Kongelige Norske Videnskabers Selskabs Skrifter 1931*, F. Bruns Bokhandel, Trondheim, 1–248, 1932.
16. A. Rothen, 'The Ellipsometer, an Apparatus to Measure Thicknesses of Thin Surface Films', *Rev. Sci. Instrum.*, **16**, 26–30 (1945).
17. E. Passaglia, R.R. Stromberg, J. Kruger (eds.), *Ellipsometry in the Measurement of Surfaces and Thin Films*, National Bureau of Standards Misc. Publ. 256, NBS, Washington, 1963.
18. R.W. Collins, Y.-T. Kim, 'Ellipsometry for Thin-film and Surface Analysis', *Anal. Chem.*, **62**, 887A–900A (1990).
19. N.M. Bashara, R.M.A. Azzam (eds.), *Ellipsometry; Proceedings of the Third International Conference on Ellipsometry*, North-Holland, Amsterdam, 1976.
20. R.H. Muller, R.M.A. Azzam, D.E. Aspnes (eds.), *Ellipsometry; Proceedings of the Fourth International Conference on Ellipsometry*, North-Holland, Amsterdam, 1980.
21. F. Abeles (ed.), *Ellipsometry and Other Optical Methods for Surface and Thin Film Analysis*, Les Editions de Physique, Les Ulis, France, 1983.
22. A.C. Boccara, C. Pickering, J. Rivory (eds.), *Spectroscopic Ellipsometry: Proceedings of the First International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam, 1993.
23. R.W. Collins, K. Vedam, 'Optical Properties of Solids', in *Encyclopedia of Applied Physics*, ed. G.L. Trigg, VCH, New York, 285–336, Vol. 12, 1995.
24. R.H. Muller, 'Principles of Ellipsometry', in *Advances in Electrochemistry and Electrochemical Engineering*, ed. R.H. Muller, Wiley-Interscience, New York, 167–226, Vol. 9, 1973.
25. M. Schubert, 'Polarization Dependent Parameters of Arbitrarily Anisotropic Homogeneous Epitaxial Systems', *Phys. Rev. B*, **53**, 4265–4274 (1996).
26. J.A. Woollam, P.G. Snyder, 'Fundamentals and Applications of Variable Angle Spectroscopic Ellipsometry', *Mater. Sci. Eng.*, **B5**, 279–283 (1990).
27. W.-K. Paik, J.O'M. Bockris, 'Exact Ellipsometric Measurement of Thickness and Optical Properties of a Thin Light-absorbing Film Without Auxiliary Measurements', *Surf. Sci.*, **28**, 61–68 (1971).
28. G.E. Jellison, Jr, 'Data Analysis for Spectroscopic Ellipsometry', *Thin Solid Films*, **233–234**, 416–422 (1993).
29. D.E. Aspnes, 'Geometrically Exact Ellipsometer Alignment', *Appl. Opt.*, **10**, 1024–1030 (1971).
30. J.L. Ord, 'An Ellipsometer for Following Film Growth', *Surf. Sci.*, **16**, 155–165 (1969).
31. H.J. Mathieu, D.E. McClure, R.H. Muller, 'Fast Self-Compensating Ellipsometer', *Rev. Sci. Instrum.*, **45**, 798–802 (1974).
32. R. Muller, 'Ellipsometry of Electrochemical Surface Layers', *Proc. Soc. Photo-Opt. Instrum. Eng.*, **112**, 68–73 (1977).
33. D.E. Aspnes, A.A. Studna, 'High Precision Scanning Ellipsometer', *Appl. Opt.*, **14**, 220–228 (1975).
34. R.W. Collins, 'Automatic Rotating-element Ellipsometers: Calibration, Operation, and Real Time Applications', *Rev. Sci. Instrum.*, **61**, 2029–2062 (1990).
35. P.S. Hauge, F.H. Dill, 'A Rotating-compensator Fourier Ellipsometer', *Opt. Commun.*, **14**, 431–437 (1975).
36. D.E. Aspnes, 'Photometric Ellipsometer for Measuring Partially Polarized Light', *J. Opt. Soc. Am.*, **65**, 1274–1278 (1975).
37. R.W. Collins, I. An, H. Fujiwara, J. Lee, Y. Lu, J. Koh, P.I. Rovira, 'Advances in Multichannel Spectroscopic Ellipsometry', *Thin Solid Films*, **313–314**, 18–32 (1998).
38. J.M. Bennett, H.E. Bennett, 'Polarization', in *Handbook of Optics*, ed. W.G. Driscoll, McGraw-Hill, New York, 1–164, Chapter 10, 1978.
39. S.N. Jasperson, S.E. Schnatterly, 'An Improved Method for High Reflectivity Ellipsometry Based on a New Polarization Modulation Technique', *Rev. Sci. Instrum.*, **40**, 761–767 (1969).
40. B. Drevillon, J. Perrin, R. Marbot, A. Violet, J.L. Dalby, 'Fast Polarization Modulated Ellipsometer Using a Microprocessor System for Digital Fourier Analysis', *Rev. Sci. Instrum.*, **53**, 969–977 (1982).
41. G.E. Jellison, Jr, F.A. Modine, 'Two-Channel Polarization Modulation Ellipsometer', *Appl. Opt.*, **29**, 959–974 (1990).
42. B. Drevillon, 'Spectroscopic Ellipsometry in the Infrared Range', *Thin Solid Films*, **313–314**, 625–630 (1998).
43. K.-L. Barth, D. Bohme, K. Kamaras, F. Keilmann, M. Cardona, 'Far-IR Spectroscopic Ellipsometer', *Thin Solid Films*, **233–234**, 314–317 (1993).
44. E. Wold, J. Bremer, O. Hunderi, B.-O. Fimland, 'Far Infrared Ellipsometric Measurements of $(GaAs)_n$/

$(Al_xGa_{1-x}As)_n$ Superlattice Films by Means of Synchrotron Radiation', *Thin Solid Films*, **313–314**, 649–655 (1998).
45. R.L. Johnson, J. Barth, M. Cardona, D. Fuchs, A.M. Bradshaw, 'Spectroscopic Ellipsometry with Synchrotron Radiation', *Rev. Sci. Instrum.*, **60**, 2209–2212 (1989).
46. M. Yamamoto, M. Furudate, 'Ellipsometry in the Extreme Ultraviolet Region with Multilayer Polarizers', *Thin Solid Films*, **313–314**, 751–755 (1998).
47. P. Chindaudom, K. Vedam, 'Optical Characterization of Inhomogeneous Transparent Films on Transparent Substrates by Spectroscopic Ellipsometry', in *Optical Characterization of Real Films and Surfaces*, ed. K. Vedam, Academic, New York, 191–247, 1994.
48. T.E. Tiwald, D.W. Thompson, J.A. Woollam, S.V. Pepper, 'Determination of the Mid-IR Optical Constants of Water and Lubricants Using IR Ellipsometry Combined with an ATR Cell', *Thin Solid Films*, **313–314**, 718–721 (1998).
49. M. Luttmann, J.-L. Stehle, C. Defranoux, J.-P. Piel, 'High Accuracy IR Ellipsometer Working with a Ge Brewster Angle Reflection Polarizer and Grid Analyzer', *Thin Solid Films*, **313–314**, 631–641 (1998).
50. A. Michaelis, M. Schweinsberg, 'An Anisotropy Microellipsometry Study of Anodic Film Formation on Ti and Zr Single Grains', *Thin Solid Films*, **313–314**, 756–763 (1998).
51. K. Sugimoto, S. Matsuda, 'Analysis of Passive Films on Austeno-ferritic Stainless Steel by Microscopic Ellipsometry', *J. Electrochem. Soc.*, **130**, 2323–2328 (1983).
52. M. Erman, J.B. Theeten, 'Spatially Resolved Ellipsometry', *J. Appl. Phys.*, **60**, 859–873 (1986).
53. C. Pickering, J. Russell, V. Nayar, J. Imschweiler, H. Wille, S. Harrington, C. Wiggins, J.-L. Stehle, J.-P. Piel, J. Bruchez, 'Evaluation of Automated Spectroscopic Ellipsometry for In-line Process Control', *Thin Solid Films*, **313–314**, 446–453 (1998).
54. R.F. Cohn, J.W. Wagner, J. Kruger, 'Dynamic Imaging Microellipsometry: Theory, System Design, and Feasibility Demonstration', *Appl. Opt.*, **27**, 4664–4671 (1988).
55. H. Arwin, 'Spectroscopic Ellipsometry and Biology; Recent Developments and Challenges', *Thin Solid Films*, **313–314**, 764–774 (1998).
56. J. Lee, P.I. Rovira, I. An, R.W. Collins, 'Rotating-Compensator Multichannel Ellipsometry for Characterization of the Evolution of Non-Uniformities in Diamond Film Growth', *Appl. Phys. Lett.*, **72**, 900–902 (1998).
57. P.I. Rovira, R.W. Collins, 'Analysis of Specular and Textured SnO_2:F Films by High Speed Four-parameter Stokes Vector Spectroscopy', *J. Appl. Phys.*, **85**, 2015–2025 (1999).
58. D.E. Aspnes, 'Optical Properties of Thin Films', *Thin Solid Films*, **89**, 249–262 (1982).
59. G.A. Niklasson, C.G. Granqvist, O. Hunderi, 'Effective Medium Models for the Optical Properties of Inhomogeneous Materials', *Appl. Opt.*, **20**, 26–30 (1981).
60. H.V. Nguyen, Y. Lu, S. Kim, M. Wakagi, R.W. Collins, 'Optical Properties of Ultrathin Crystalline and Amorphous Silicon Films', *Phys. Rev. Lett.*, **74**, 3880–3883 (1995).
61. Lord Rayleigh, 'On Reflexion from Liquid Surfaces in the Neighbourhood of the Polarizing Angle', *Philos. Mag.*, **33**, 1–19 (1892).
62. D.E. Aspnes, A.A. Studna, 'An Investigation of Ion-bombarded and Annealed ⟨111⟩ Surfaces of Ge by Spectroscopic Ellipsometry', *Surf. Sci.*, **96**, 294–306 (1980).
63. D.E. Aspnes, A.A. Studna, 'Chemical Etching and Cleaning Procedures for Si, Ge, and Some III-V Compound Semiconductors', *Appl. Phys. Lett.*, **39**, 316–318 (1981).
64. D.E. Aspnes, 'Surface Transition Regions and Visible–Near UV Optical Properties of Some Semiconductors', *Physica (Amsterdam) B*, **117–118**, 359–361 (1983).
65. C. Pickering, 'Complementary In-Situ and Post-deposition Diagnostics of Thin Film Semiconductor Structures', *Thin Solid Films*, **313–314**, 406–415 (1998).
66. E.D. Palik (ed.), *Handbook of Optical Constants of Solids II*, Academic, New York, 1991.
67. D.E. Aspnes, A.A. Studna, 'Dielectric Functions and Optical Parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV', *Phys. Rev. B*, **27**, 985–1009 (1983).
68. T. Yasuda, D.E. Aspnes, 'Optical Standard Surfaces of Single-crystal Silicon for Calibrating Ellipsometers and Reflectometers', *Appl. Opt.*, **33**, 7435–7438 (1994).
69. D.E. Aspnes, 'Microstructural Information from Optical Properties in Semiconductor Technology', *Proc. Soc. Photo-Opt. Instrum. Eng.*, **276**, 188–195 (1981).
70. M. Erman, J.B. Theeten, P. Chambon, S.M. Kelso, D.E. Aspnes, 'Optical Properties and Damage Analysis of GaAs Single Crystals Partly Amorphized by Ion Implantation', *J. Appl. Phys.*, **56**, 2664–2671 (1984).
71. C.C. Kim, J.W. Garland, P.M. Raccah, 'Modeling the Optical Dielectric Function of Semiconductors: Extension of the Critical-point Parabolic-band Approximation', *Phys. Rev. B*, **45**, 11 749–11 767 (1992).
72. D.E. Aspnes, 'Analysis of Semiconductor Materials and Structures by Spectroellipsometry', *Proc. Soc. Photo-Opt. Instrum. Eng.*, **946**, 84–97 (1988).
73. Y.-T. Kim, D.L. Allara, R.W. Collins, K. Vedam, 'Real Time Spectroscopic Ellipsometry Study of Electrodeposition of Polypyrrole', *Thin Solid Films*, **193–194**, 350–360 (1990).
74. J. Lee, R.W. Collins, A.R. Heyd, F. Flack, N. Samarth, 'Spectroellipsometry for Characterization of $Zn_{1-x}Cd_xSe$ Multilayered Structures on GaAs', *Appl. Phys. Lett.*, **69**, 2273–2275 (1996).
75. J. Lee, F. Flack, N. Samarth, R.W. Collins, 'Composition and Temperature Dependence of the Optical Properties of $Zn_{1-x}Cd_xSe$ ($0 \leq x \leq 0.34$) below the Fundamental Bandgap', *Appl. Opt.*, **36**, 5372–5382 (1997).

76. R.H. Muller, J.C. Farmer, 'Fast Self-compensating Spectral-scanning Ellipsometer', *Rev. Sci. Instrum.*, **55**, 371–374 (1984).
77. I. An, Y.M. Li, H.V. Nguyen, R.W. Collins, 'Spectroscopic Ellipsometry on the Millisecond Time Scale for Real Time Investigations of Thin Film and Surface Phenomena', *Rev. Sci. Instrum.*, **63**, 3842–3848 (1992).
78. S.A. Henck, W.M. Duncan, L.M. Lowenstein, S.W. Butler, 'In Situ Spectral Ellipsometry for Real Time Thickness Measurement: Etching Multilayer Stacks', *J. Vac. Sci. Technol. A*, **11**, 1179–1185 (1993).
79. I. An, H.V. Nguyen, A.R. Heyd, R.W. Collins, 'Simultaneous Real Time Spectroscopic Ellipsometry and Reflectance for Monitoring Thin-film Preparation', *Rev. Sci. Instrum.*, **65**, 3489–3500 (1994).
80. J. Lee, P.I. Rovira, I. An, R.W. Collins, 'Rotating-compensator Multichannel Ellipsometry: Applications for Real Time Stokes Vector Spectroscopy of Thin Film Growth', *Rev. Sci. Instrum.*, **69**, 1800–1810 (1998).
81. R.W. Collins, I. An, H.V. Nguyen, Y. Li, Y. Lu, 'Real Time Spectroscopic Ellipsometry Studies of the Nucleation, Growth, and Optical Functions of Thin Films, Part I: Tetrahedrally Bonded Materials', in *Optical Characterization of Real Films and Surfaces*, ed. K. Vedam, Academic, New York, 49–125, 1994.
82. D.E. Aspnes, 'Minimal Data Approaches for Determining Outer-layer Dielectric Responses of Films from Kinetic Reflectometric and Ellipsometric Measurements', *J. Opt. Soc. Am. A*, **10**, 974–983 (1993).
83. S. Kim, J.S. Burnham, J. Koh, L. Jiao, C.R. Wronski, R.W. Collins, 'Real Time Spectroellipsometry Characterization of Optical Gap Profiles in Compositionally-graded Semiconductor Structures: Applications to Bandgap Engineering in Amorphous Silicon–Carbon Alloy Solar Cells', *J. Appl. Phys.*, **80**, 2420–2429 (1996).
84. J. Rivory, 'Characterization of Inhomogeneous Dielectric Films by Spectroscopic Ellipsometry', *Thin Solid Films*, **313–314**, 333–340 (1998).
85. J. Lee, B. Hong, R. Messier, R.W. Collins, 'Application of Real Time Spectroscopic Ellipsometry for the Development of Low Temperature Diamond Growth Processes', *Thin Solid Films*, **313–314**, 506–510 (1998).
86. P. Lautenschlager, M. Garriga, L. Vina, M. Cardona, 'Temperature Dependence of the Dielectric Function and Interband Critical Points in Silicon', *Phys. Rev. B*, **36**, 4821–4830 (1987).
87. J.C. Farmer, R.H. Muller, 'Effect of Rhodamine B on the Electrodeposition of Lead on Copper', *J. Electrochem. Soc.*, **132**, 313–319 (1985).
88. S.T. Mayer, R.H. Muller, 'Nucleation of Silver(I) Oxide Investigated by Spectroscopic Ellipsometry', *J. Electrochem. Soc.*, **135**, 2133–2142 (1988).
89. H.S. Witham, P. Chindaudom, I. An, R.W. Collins, R. Messier, K. Vedam, 'Effect of Preparation Conditions on the Morphology and Electrochromic Properties of Amorphous Tungsten Oxide Films', *J. Vac. Sci. Technol. A*, **11**, 1881–1887 (1993).
90. R.W. Collins, 'Real Time Diagnostics of Thin Film Growth: Introduction and General Discussion', in *Handbook of Thin Film Process Technology*, ed. D.A. Glocker, S.I. Shah, IOP, Philadelphia, 1–10 Part D0, (1995).
91. K. Vedam, 'Spectroscopic Ellipsometry: A Historical Overview', *Thin Solid Films*, **313–314**, 1–9 (1998).
92. K. Vedam, P.J. McMarr, J. Narayan, 'Nondestructive Depth Profiling by Spectroscopic Ellipsometry', *Appl. Phys. Lett.*, **47**, 339–341 (1985)
93. J. Koh, Y. Lu, C.R. Wronski, Y. Kuang, R.W. Collins, T.T. Tsong, Y.E. Strausser, 'Correlation of Real Time Spectroscopic Ellipsometry and Atomic Force Microscopy Measurements of Surface Roughness on Amorphous Semiconductor Thin Films', *Appl. Phys. Lett.*, **68**, 1297–1299 (1996).
94. H. Fujiwara, J. Koh, C.R. Wronski, R.W. Collins, J.S. Burnham, 'Optical Depth Profiling of Bandgap Engineered Interfaces in Amorphous Silicon Solar Cells at Monolayer Resolution', *Appl. Phys. Lett.*, **72**, 2993–2995 (1998).

Infrared and Raman Spectroscopy in Analysis of Surfaces

Zhong-Qun Tian and Bin Ren
Xiamen University, Xiamen, China

1 Introduction **2**
1.1 Vibrational Properties of Surface Species 2
1.2 Surface Selection Rules for Infrared and Raman Spectroscopies 3
1.3 Surface-enhanced Raman Scattering and Surface-enhanced Infrared Absorption 5
1.4 Surface Infrared and Raman Spectroscopy 9
2 History **9**
2.1 History of Surface Infrared Spectroscopy 9
2.2 History of Surface Raman Spectroscopy 10
3 Instrumental Methods **11**
3.1 Preparation of Samples 11
3.2 Experimental Set-up and Operation Modes 15
3.3 Newly Developing Methods 22
3.4 Troubleshooting 28
4 Applications **30**
4.1 Advanced Materials and Surface Processing 30
4.2 Catalysis 32
4.3 Electrochemistry and Corrosion 32
4.4 Biology and Sensor 34
5 Prospective and Future Developments **35**
5.1 Surfaces 35
5.2 Instrumentation 35
5.3 Methods 36
Acknowledgments **37**
Abbreviations and Acronyms **37**
Related Articles **37**
References **38**

Infrared (IR) spectroscopy and Raman spectroscopy are important analytical instrumental techniques for surface science. An analytically useful fingerprint spectrum can be recorded from almost any surface. As a consequence, surface IR and Raman spectroscopies find extensive application in qualitative analysis, providing either functional group or structural information about a surface, and determining surface bonding, conformation and orientation. More importantly, these two vibrational spectroscopies can be applied under in situ conditions of solid–gas, solid–liquid and solid–solid interfaces of both fundamental and practical importance. They can also very successfully be used for studies of surface dynamics and ill-defined high-surface-area porous materials, to which many surface techniques are not applicable. Therefore, IR and Raman spectroscopies are among the most promising and most widely used methods for surface science. However, there is great limitation when these two techniques are applied to the detection of surface species of ten times less than a monolayer. Their surface sensitivities, especially for Raman spectroscopy, are remarkably lower than that of ultra high vacuum (UHV)-based electron energy loss spectroscopy (EELS) technique. Fortunately, surface-enhanced Raman spectroscopy (or scattering) (SERS) and surface-enhanced infrared absorption (SEIRA) can improve the sensitivity significantly by several orders of magnitude for roughened surfaces of noble and transition metals.

This article first concisely outlines the basic principles and opportunities of surface IR and Raman spectroscopy followed by a description of SERS and SEIRA phenomena as well as techniques in more detail. A brief survey of the histories of surface IR spectroscopy and surface Raman spectroscopy is then given. Comparisons are made between surface IR and surface Raman spectroscopies on the sensitivities, surface enhancement, instrumental methods and applications. Emphasis is placed on their ability to deliver specific chemical identifications, coupled with the wide range of instrumental and sampling methodologies available and related phenomena in terms of the fundamental and instrumental parameters which affect the surface signal intensity. Important factors are recognized and prioritized to extend the general utilities of surface IR and surface Raman spectroscopy to in situ studies on solid–gas, solid–liquid and solid–solid interfaces under various conditions with changing temperature, pressure, media and electrode potential. Some newly developing techniques for surface science, such as single molecule Raman spectroscopy, IR and Raman microscopies and imaging techniques, time-resolved IR and Raman spectroscopies, and ultraviolet (UV) resonance Raman spectroscopy, are introduced with representative examples. The troubleshooting

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

section deals with some problems commonly encountered in the study, e.g. low detection sensitivity, surface heating and damage by laser, fluorescence elimination, coupling with transient measurement, and spectral data analyzing. Applications are exemplified on extensively studied areas such as in advanced materials, catalysis, electrochemistry, corrosion, biology and sensor. Finally, prospects and further developments of this field are given with emphasis on the emerging as well as promising methodologies in view of surface preparation, instrumentation and hyphenated techniques.

1 INTRODUCTION

1.1 Vibrational Properties of Surface Species

The vibrational properties of surface species are significantly different from that of the same species in bulk phases. The individual molecules may be distorted and their symmetry altered as compared to the free molecules, thus leading to modifications of their vibrational behaviors. There are various processes contributing to the bandwidths, intensities and positions, including change of molecular symmetry, dipole coupling, inhomogeneous broadening, intramolecular vibrational relaxation, dephasing, electron–hole pair creation, and phonon coupling. It is necessary to first briefly describe some key features of a molecule interacting (or adsorbing) with a surface and/or with the same or other surface species, in terms of molecular symmetry,[1] dipole coupling interactions,[2] and the electrochemical Stark effect.[3]

A systematic and good illustration requires a model molecule as adsorbate interacting with a surface. There are several important reasons for using carbon monoxide (CO) as an adsorbate. CO possesses a minimum internal vibrational mode and can easily be isotopically labeled at either end. Besides, the ν_{CO} band position and shape are sensitive to the chemical and electrostatic environment as well as to the mode of surface bonding. The relatively large dynamic dipole moment of the C−O vibration, ν_{CO}, for chemisorbed CO, facilitates its detection by surface IR spectroscopy. The strong chemisorption found on transition metals enables adsorbed CO to be examined over a wide range of coverage, and under ambient and higher temperatures. Moreover, a rich base of knowledge about CO adsorption at well-defined metal surfaces in UHV can be used to probe the additional complexities of catalysts' surfaces and electrochemical interfaces.

1.1.1 Molecular Symmetry of Surface Species[1]

It is well known that the number of dipole active vibrations of a molecule is determined by degrees of vibrational freedom and symmetry. The freedom of the molecule depends on the molecule's physical state. A free molecule with N atoms has $3N$ degrees of freedom, with $3N-6$ ($3N-5$ for a linear molecule) degrees of vibrational freedom. When a molecule becomes attached to the surface, the translational motion and rotational motion are restricted and the degree of vibrational freedom equals $3N$. That means, an adsorbed molecule has an excess of 6 (or 5) degrees of vibrational freedom. To illustrate what can happen when gaseous species adsorbed on the surface, let us consider CO adsorbed on a single crystal (1 0 0) surface. CO can adsorb either linearly or by bridge mode on the adsorption center of the surface, or both, see Figure 1. Linearly adsorbed species belong to the C_{4V} point group with two totally symmetric vibrational modes belonging to the A_1 representation; one corresponds to the stretching mode of free CO molecule, another comes from the restricted translational motion of the whole CO molecule relative to the surface, i.e. the stretching mode of the surface atom with the C atom. All other four modes are twisted vibrations, are parallel to the surface and belong to the degenerate E representation. The freedom for these vibrations comes from the restricted molecular translation and rotation. The bridge-bonded species has C_{2V} symmetry, with two A_1 stretching modes perpendicular to the surface. The degeneration of the parallel mode as in the case of the on-top mode leaves two B_1 modes parallel to the bridged plane and two B_2 modes normal to the bridged plane (the molecular models for the B_2 vibrations are given as side views in Figure 1).

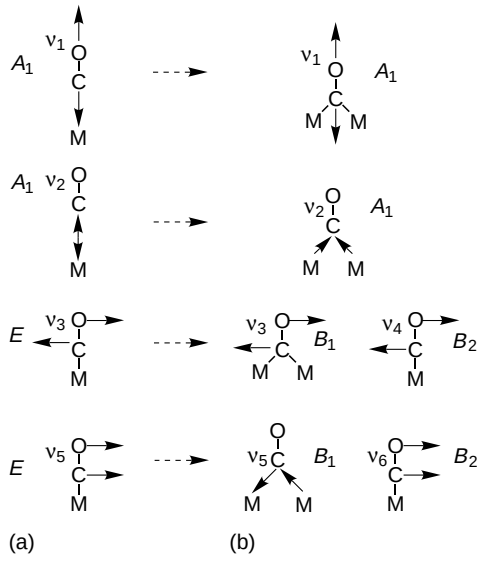

Figure 1 Possible structural features (vibrational modes) of CO on a metal surface: (a) on-top site; (b) bridge site.

1.1.2 Dipole Coupling[2]

In an adsorbed layer at surfaces, electrostatic interactions between the dipoles induced in neighboring molecules serve to couple together the molecules' vibrational motion. The effects of such dipole coupling interactions have been investigated in considerable detail on single crystal surfaces. Briefly, coupling has two important consequences: coupling between identical molecules gives rise to a single IR-active mode, having a vibrational frequency somewhat higher than that of the isolated single molecule; coupling between two species with different singleton frequencies yields two active modes. Each of these has a vibrational frequency close to that of the original species, so there are no significant shifts in the positions of the absorption bands. There is, however, a characteristic transfer of intensity from the low frequency mode to its higher frequency counterpart.

1.1.3 Electrochemical Stark Effect[3]

Spectra of adsorbed species on an electrode are often characterized by changes in their vibrational frequencies with potential. Two mechanisms have been proposed in an effort to explain this shift in band position with potential:

1. The first involves molecular orbital consideration, in which electrons can be donated from filled adsorbate orbitals of proper symmetry to empty metal orbitals through σ overlap. The metal can then back-donate electrons from filled d orbitals to empty π^* adsorbate anti-bonding orbitals. The frequency of an adsorbed moiety (with respect to solution species) will decrease if the π-bonding interaction dominates (such as CO), while the frequency will increase if the σ interaction is more important (as in the case of CN^-). When the electrode potential is made more negative, the adsorbate–metal bond is weakened, owing to charge donation from the metal into adsorbate π^* orbitals, thus the frequency shifts to lower wavenumber and vice versa.
2. The second is the electric field mechanism that involves coupling of highly polarizable electrons of the adsorbate with the strong electric field across the electrochemical double layer. If the electrolyte concentration is high enough (ca. 1 mol L^{-1}), most of the potential drop will be approximately linear with distance and will occur within a few angstroms of the electrochemical interface. An adsorbed species can act as a dielectric across which the largest portion of the potential drop will occur. Very high electric field (on the order of 10^7 V cm^{-1}) is predicted to exist in the double layer. Interaction of this electric field with the dipole moment of an adsorbed molecule can lead to changes in the molecular vibrational frequency.

1.2 Surface Selection Rules for Infrared and Raman Spectroscopies

As vibrational properties of surface species are so different from that of the same species in the bulk phase, surface selection rules are extremely important for analyzing surface IR and surface Raman spectra correctly. These rules can be the basis of a valuable technique for the determination of the molecular orientation and conformation on surfaces.

1.2.1 Surface Selection Rule of Infrared Spectroscopy[3]

The special properties of the reflection of electromagnetic radiation from metal surfaces in the IR region make it possible to obtain information about the orientation of molecular species at the interface. When the IR light is reflected from a metal surface, s-polarized radiation (whose electric field vector component is polarized perpendicular to the plane of incidence) undergoes a phase shift of nearly 180° at all angles of incidence. The incident and reflected waves cancel at the surface, giving rise to a standing wave that has a node near the surface. p-Polarized IR radiation (whose electric field vector component is parallel to the incidence plane) undergoes a phase shift that is a function of the angle of incidence. A phase shift of about 90° is attained at a glancing angle of incidence, and a standing wave that has measurable amplitude near the surface results from the interaction of incident and reflected light. Amplitude and angle of incidence relationships for interfaces with known optical parameters may be obtained by solving Frensnel's equations.

Since p-polarized light is the only component of IR radiation with appreciable field strength at the metal surface, it alone carries information on vibrations of surface species. Greenler was the first to propose the p-polarized IR radiation carrying virtually all of the vibrational absorption information of species adsorbed at a metallic reflecting surface. Since the electric field intensity increases sharply with angle of incidence up to near grazing (for p-polarized radiation), the absorbance by an oscillator could be expected to increase as well. The absorption of radiation by an oscillator varies as the square of the cosine of the angle between the electric field vector of the radiation and the oscillating dipole. Hence, optimal absorption is obtained when the component of the dipole derivative with respect to the normal coordinate of an absorbing species is in the same direction as incident p-polarized radiation. If the IR beam is incident at very glancing/grazing angles, a sizeable component of the electric field will lie in a direction perpendicular to

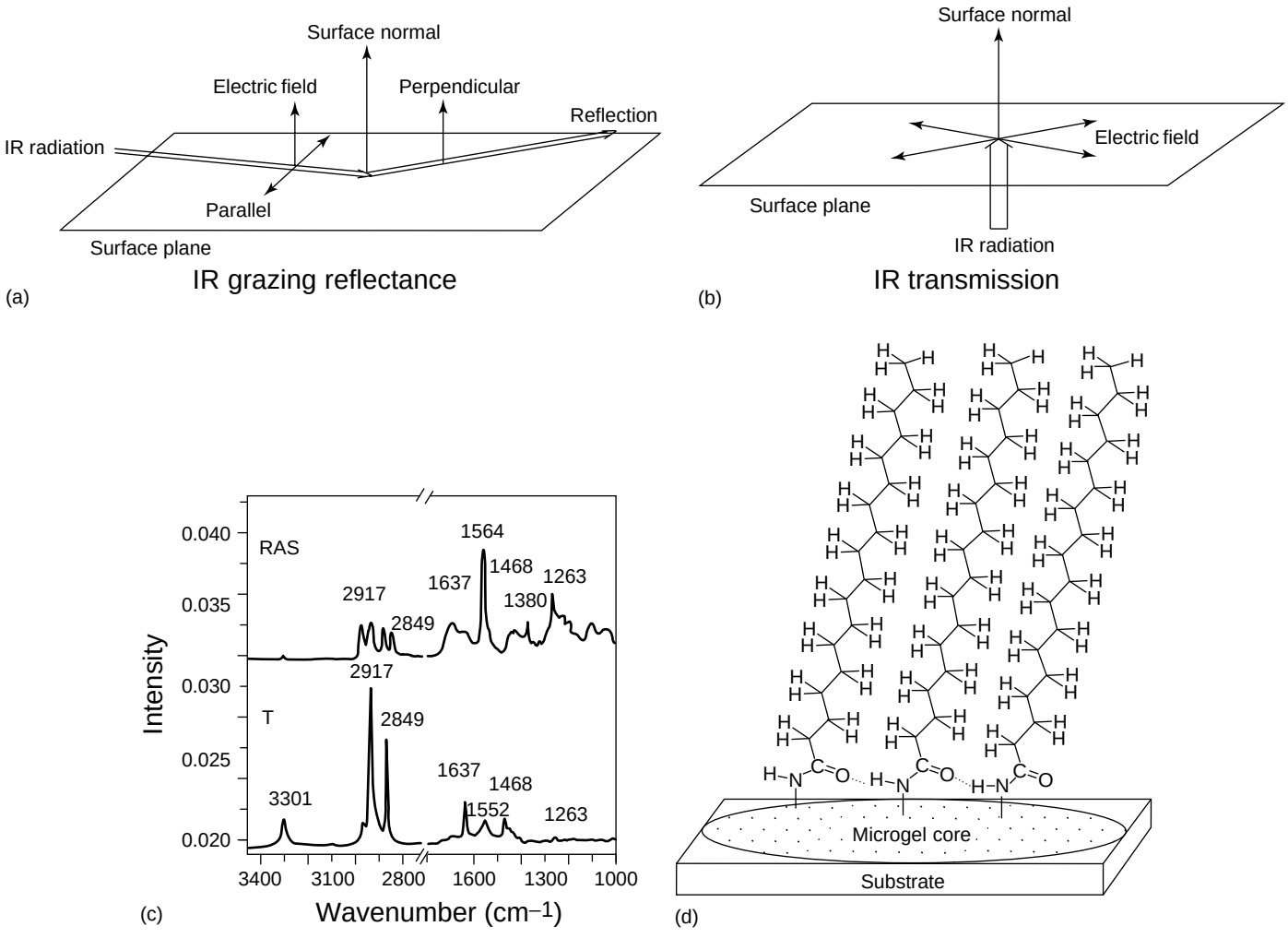

Figure 2 The electric field vectors for IR light incident with (a) grazing angle or (b) perpendicular to the surface. (c) A comparison of IR transmission spectrum (T) and RA spectrum (RAS) of an LB film. (d) Molecular orientation at the surface.

a metal surface, see Figure 2(a). Oscillators with their dipole derivative component more parallel to the surface will have smaller probability of IR absorption. In this way the relative intensities of IR absorption bands can give information regarding molecular orientations on the surface. This distinction between absorbances due to oscillators oriented parallel and perpendicular with respect to the surface is known as the surface selection rule.

To summarize, for the absorption of light wavenumber ν by a molecule immobilized at a metal surface using the reflection mode, three conditions must be met: (i) the molecule must have a nonzero transition dipole moment at ν; (ii) the magnitude of the mean-square electric field at the interface must be nonzero (with conditions for high detectability at large angles of incidence and p-polarized light); (iii) the transition dipole moment must have a component oriented along the surface normal.

In the case of transmission mode, see Figure 2(b), the molecular monolayer investigated is deposited uniformly onto a flat and IR transparent substrate such as CaF_2. Since both the p-polarized and s-polarized light is parallel to the surface, oscillators with their dipole derivative component more parallel to the surface will have larger probability of IR absorption. Based on the surface selection rule, the combined study of IR transmission spectroscopy and reflection–absorption (RA) spectroscopy can provide detailed structural information about molecular orientation of films and adsorbed layers at the surface after considering that in RA spectroscopy, vibrational modes with their transition moments perpendicular to the surface are enhanced in the reflectance spectrum. An example to determine the molecular orientation of the alkyl chains of a Langmuir–Blodgett (LB) film of an amphiphilic polymer, was shown in Figure 2(c) and (d). The intensities of CH_2 symmetric and antisymmetric stretching bands

(at 2849 cm^{-1} and 2917 cm^{-1} respectively) are much stronger in the transmission spectra than in the RA spectra. Therefore, it indicates that the hydrocarbon chains of the LB film are nearly perpendicular to the substrate surface. The same tendency of the N–H stretching band (3301 cm^{-1}) and the C=O stretching band (1637 cm^{-1}) in the transmission spectrum indicates that both groups are nearly parallel to the substrate surface. The N–H bending band (~1560 cm^{-1}) and the band of in-plane C–N stretching coupled with N–H bending mode (1263 cm^{-1}) are much stronger in the RA spectra than in the transmission spectra, indicating that the amide planes are nearly perpendicular to the substrate surface.[4]

1.2.2 Surface Selection Rule of Raman Spectroscopy[5,6]

For extending the surface selection rule to Raman spectroscopy of metal surfaces, let us consider a molecule placed at a flat surface. The molecule illuminated by the incident light at a certain angle can be thought to be illuminated by two beams, a direct beam and one reflected beam at the surface, which superimpose coherently at the molecule. It is shown that the molecule could be polarized by the local electric field on a flat surface and the induced dipole of the molecule radiates the scattered light. The relationships between the local electric field caused by the incident and reflected light and the scattered light can be described by a polarizability tensor, which contains nine components, such as $\alpha_{xx}, \alpha_{yy}, \alpha_{zz}, \alpha_{xy}, \alpha_{yx}, \alpha_{xz}, \alpha_{zx}, \alpha_{yz}$, and α_{zy}. The polarizability tensor is written as follows:

$$\boldsymbol{\alpha} = \begin{matrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{matrix}$$

in which $\alpha_{xy} = \alpha_{yx}, \alpha_{xz} = \alpha_{zx}, \alpha_{yz} = \alpha_{zy}$. Therefore there are only six independent tensor elements. The intensity of the Raman scattering light on the surface is directly related to the derivative of the polarizability tensor with respect to the given normal coordinates. In order to determine these tensor elements, it is necessary to consider the system including the molecule and surface atoms, and to define the Cartesian coordinate axis z along the surface normal and others along the surface parallel.

Based on the above statement of the molecule–surface system and coordinate frame, one is able to determine the symmetry point group of the molecule adsorbed on the surface. However, it should be emphasized that for a chemisorbed molecule, the symmetry of the point group is that of the surface molecule, possibly having a lower symmetry than the molecule in the bulk phase such as of the solution. In particular, for the strongly chemisorbed molecule, its geometry may be changed significantly by the interaction between the molecule and the surface. After obtaining the symmetry point group of the adsorbed system, it is easy to determine which symmetry representations of the vibrations of the molecule will be Raman active by looking in the group character tables for the given point group. The symmetry representations appearing in the character tables form the irreducible representation of the point group. Different vibrational modes of the molecule belong to different irreducible representations of the point group to which it belongs. Raman-active modes are those that possess nonzero components of the Raman polarizability tensor. In general, modes belonging to different representations will have different components of the nonvanishing polarizability tensor.

The surface Raman spectrum can be significantly different from the ordinary Raman spectrum. It is reasonable that the tangential component of the surface electric field is nearly zero due to the superimposition of the incident field with an out-of-phase reflected field, while the normal component of the incident field is reinforced by the reflected field. So the field component normal to the surface has an important contribution to totally symmetric vibration modes. Accordingly, it is possible to design Raman experiments such that the directions of the incident and scattering light and their polarization directions are chosen so as to enable the determination of the particular polarizability tensor elements from the observed Raman intensities, which are related directly to the surface orientation of the molecule. In addition to metals, the surface selection rules should apply to insulators at frequencies where they have metallic reflectance.

1.3 Surface-enhanced Raman Scattering and Surface-enhanced Infrared Absorption

The most important component of any reliable and general surface analytical method is the sensitivity, which is unfortunately the disadvantage of both IR spectroscopy and Raman spectroscopy in comparison with EELS and many other surface techniques. Raman scattering is a second-order process, with very low cross-sections. In the absence of any resonance Raman process, the differential Raman cross-sections $(d\sigma/d\Omega)_{NRS}$ are smaller than 10^{-29} cm^2 sr^{-1}. The corresponding surface Raman intensities expected for a monolayer of adsorbates are less than 1 count s^{-1} when using standard Raman spectrometer systems. Not surprisingly, there was no attempt to apply Raman spectroscopy to adsorbates at surfaces until the mid-1970s, when SERS was discovered.

In 1977, it was discovered largely through the work of Fleischmann, Van Duyne, Creighton, and their coworkers that molecules adsorbed on electrochemically roughened silver surfaces produce a Raman spectrum

that is at times a millionfold more intense than what is expected.[7,8] This tremendous effect was dubbed SERS. SERS differs in a number of ways from ordinary Raman spectroscopy of molecules and even from unenhanced surface Raman spectroscopy. A large number of experimental characteristics of SERS have been discovered, and the key features are summarized as follows:

1. SERS occurs for a very large number of molecules adsorbed on or near to the surface of metals in a variety of morphologies. The largest enhancements (with surface enhancement factor (SEF) $> 10^6$) are observed from a few coinage metals of silver, gold and copper. An enhancement as high as 10^{14} is obtained from single silver nanoparticles of suitable size (see below). Little work has been reported on the alkali metals. Recently, it has been confirmed that transition metals such as Pt, Pd, Ru, Rh, Ni, Co and Fe have surface enhancements, although relatively weak. Their SEF values range from 10^2 to 10^4, depending on the nature of the metal and surface roughness.
2. A surface roughness with features of submicroscopic dimensions (about 20–500 nm) and/or atomic scale roughness (adatoms, steps, kinks, or vacancies) is required. The largest enhancements occur for surfaces which are rough at the nanoscale (20–150 nm). These mainly include electrode surfaces roughened by one or more oxidation–reduction cycles (ORCs), surfaces roughened by chemical etching in acids, island films deposited on glass surfaces at elevated temperatures, films deposited by evaporation or sputtering in vacuum onto cold (100 K) substrates or by electrochemical deposition in solutions, colloids (especially aggregated colloids) generated by chemical method or laser ablation, single nanoparticles filtered from the colloids, and nanoparticle and nanorod arrays prepared by lithographic or template techniques. It has been thought that surface roughness, either atomic scale or nanoscale, was required for a strong SERS effect. However, recent results show that roughness is not a requirement for some specific molecules.[8]
3. SERS is interface-sensitive. Molecules adsorbed in the first layer at the surface show the largest enhancements. However, the enhancement also has a long-range aspect with molecules separated from the surface by tens of nanometers showing some enhancement, depending upon the substrate morphology and physical environments.
4. The Raman bands from rough surfaces of SERS activity are completely depolarized unless single nanoparticles are employed, in contrast to those taken from molecules adsorbed on atomically smooth surfaces. Overtone and combination bands are not prevalent. Selection rules are relaxed resulting in the appearance of normally forbidden Raman modes in the surface spectra. SERS active surfaces also display a continuum inelastic background scattering and are very effective in quenching fluorescence.
5. The excitation profile (scattering intensity versus exciting frequency) deviates from the fourth-power dependence of normal Raman scattering. Both vibrational band frequencies and SERS intensities are a function of the applied electrode potential in electrochemical experiments, and the functionality may be different for different sets of vibrational modes.

Based on the experimental evidence, a number of SERS mechanisms have been proposed to explain the experimental characteristics; however, no one mechanism can explain all of the observed characteristics. There are at least two major types of mechanism that contribute to the SERS effect:[8–10] an electromagnetic effect associated with large local fields caused by electromagnetic resonance; and a chemical effect involving a scattering process associated with chemical interactions between the molecule and the metal surface. In addition to the chemical effects, there are other synonyms to describe the nonelectromagnetic mechanism such as charge transfer, short-range effect and SERS-active sites.

1.3.1 The Electromagnetic Effects in Surface-enhanced Raman Scattering[9]

Electromagnetic effects have been extensively studied and can be further divided into several enhancement processes. A major contribution to the electromagnetic enhancement is made by surface plasmons for the coinage metals. The electromagnetic field of the light at the surface can be greatly enhanced under conditions of surface plasmon excitation (collective electron resonance). At the plasmon frequency, the metal results in large field-induced polarizations and thus large local fields at the surface by the incident radiation. These local fields increase remarkably the Raman scattering intensity of the surface species, which is proportional to the square of the applied field. The rough surface not only enhances the incident laser field but also the Raman scattered field, therefore, the overall enhancement of the Raman intensity scales roughly the four orders of magnitude of the field.

Electromagnetic enhancement mechanisms are characterized by the following features: the effects are long range in nature, since the dipole fields induced in polarizable

metal particles vary as the inverse cube of the distance to the center of the particle; electromagnetic effects are generally independent of the adsorbed molecule; the enhancements depend on the electronic structure of the substrate and the roughness of the surface, since the frequency of the surface plasmon resonance depends on these factors.

Finally, it should be noted that surface plasmons are not the only sources of enhanced local electromagnetic fields. Other types of electromagnetic field lie near high-curvature points on the rough surface, e.g. Fresnel reflection effects and the ''lightning rod'' effect. They may become the major contribution to the enhancement for transition metals.

1.3.2 The Chemical Effects in Surface-enhanced Raman Scattering[10]

The chemical effect is associated with the electronic structure (state) of metal and adsorbate, which leads to resonance Raman scattering through photon-driven charge-transfer processes. In the charge-transfer model, an electron–hole pair of the metal is created by the incident photon, then the electron or hole is excited and tunnels into a charge-transfer excited state of the adsorbed molecule. In the first case it results in the negative ion (adsorbate molecule-electron) having equilibrium geometry differing from that of the original adsorbate molecule. Therefore, the charge-transfer process induces a nuclear relaxation in the adsorbate molecule which, after the return of the electron to the metal, leads to a vibrationally excited neutral molecule and to emission of a Raman-shifted photon. Additional Raman enhancement for adsorbates at special active sites of atomic-scale roughness may facilitate charge-transfer enhancement mechanisms, i.e. the electron–phonon couplings are increased by microscopic surface roughness features (adatoms).

In general, the chemical effects contribute to SERS on a short-range molecular scale. This mechanism depends on the adsorption site, the geometry of bonding, and the energy levels of the adsorbate molecule. For specific adsorbate–surface systems, the enhancement may be large. The contribution of charge-transfer processes to SERS has been estimated to be approximately $10–10^3$. Although the chemical effect is not a general mechanism and is restricted by its chemical specificity, it can provide useful information uniquely on chemisorptive interactions between metal and adsorbate. In comparison with electromagnetic effects, the contributions of chemical effects are less known and are currently topics of extensive research. For recent progress in SERS, readers may refer to the special issue of *Journal of Raman Spectroscopy*.[11]

1.3.3 Calculation of the Surface Enhancement Factor[7,12]

The calculation of the SEF is very important to the study of the SERS mechanism and SERS applications. It is calculated generally by comparing the integrated Raman intensity obtained from the surface with that from the solution phase. Thus, the SEF can be defined by Equation (1):

$$G = \frac{I_{\text{surf}}/N_{\text{surf}}}{I_{\text{bulk}}/N_{\text{bulk}}} \tag{1}$$

where I_{surf} and I_{bulk} denote the integrated intensities for the strongest band of surface and solution species, respectively, which can be obtained directly from Raman measurement, and N_{surf} and N_{bulk} represent the number of the corresponding surface and solution molecules effectively excited by a laser beam.

N_{surf} can be calculated from Equation (2) after considering the contribution of surface roughness:

$$N_{\text{surf}} = \frac{RA}{\sigma} \tag{2}$$

where R is the roughness factor of an electrode, A is the area of the focal spot of the laser, and σ is the surface area occupied by an adsorbed molecule (this may be obtained for a model molecule by other techniques, such as temperature-programmed desorption, and the electrochemical and radiotracer methods).

The calculation on N_{bulk} is complex and should be considered very carefully in order to obtain the correct value. Because confocal Raman microscopy is getting more popular, the calculation of SEF based on the confocal system is given here. Figure 3(a) shows the waist

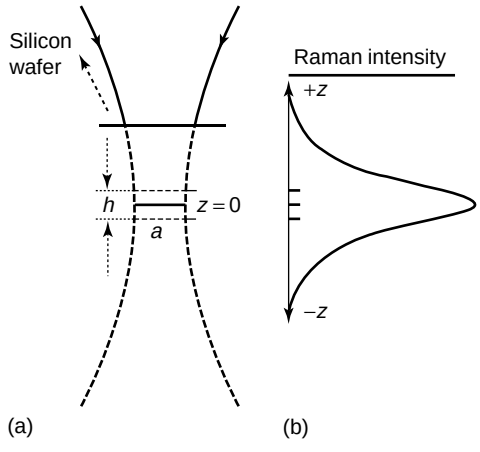

Figure 3 Schematic diagram showing (a) the waist profile of the laser beam of Raman microscopy and (b) the integrated intensity–depth profile of solution species along the laser waist. The plane at $z = 0$ is defined as the laser focusing plane. The contribution of solution species from each illuminated plane to I_{bulk} is a function of the confocal depth (see text).

profile of a focused laser beam in aqueous solution. In principle, all molecules within the illuminated volume of solution contribute to N_{bulk}. However, the collection efficiency of scattered photons from molecules in each plane of solution varies with confocal depth (z), i.e. the distance deviated from the ideal focal plane, where z equals zero and molecules therein contribute the most to overall intensity, the integrated intensity of which is defined to be I_{max}. Experimentally, a single crystal silicon wafer immersed in solution can be used to provide the confocal depth profile. The integrated intensity of the strongest band for Si at 520.6 cm^{-1} was measured as the Si wafer plane was moved up and down vertically, while the confocal position was kept stationary. Figure 3(b) presents the intensity–distance profile that simulated the confocal depth profile. It shows an abrupt decrease in both sides of the ideally focused plane. For the illuminated volume of the solution, a plane of solution can be imaginarily considered to move up and down vertically, just like the realistically moveable Si wafer. It can be seen that the contribution from molecules outside the region of $|z| > 75\,\mu m$ is negligible to N_{bulk}. Now consider an ultra-thin layer of solution with a thickness h and a volume hA in the vicinity of the ideally focused plane, in which all the molecules have the same contribution to the overall signal as that in the ideal focal plane. The overall signal can be obtained by integrating the signal on the intensity–depth profile. Thus, the thickness h can be calculated by Equation (3):

$$h = \frac{\int_{-\infty}^{\infty} I(z)\,dz}{I_{max}} \tag{3}$$

The h value (in micrometers) depends on the pinhole size and the objective lens of the Raman microscope. This value can be considered as the key parameter of the confocal characteristic, depending on the pinhole size and the magnification of the objective lens of the Raman instrument. After systematic study on the water overlayer effect on the spatial resolution for the solid–liquid interface we found that the spatial resolution for the confocal system at the same pinhole size and slit width using the same microscopy objective is quite different with and without the water overlayer. The effectively illuminated number of molecules in aqueous solution, N_{bulk}, can be written as Equation (4):

$$N_{bulk} = AhcN_A \tag{4}$$

where c is the concentration of adsorbate, and N_A is Avogadro's number. Using Equations (2) and (4), Equation (1) can be rearranged to Equation (5):

$$G = \frac{cN_A\sigma h I_{surf}}{RI_{bulk}} \tag{5}$$

The value of G can be obtained after substituting all the known data and constants in the equation.

1.3.4 Surface-enhanced Infrared Absorption[13]

Molecules adsorbed on evaporated thin and rough metal films can also exhibit enormously strong IR absorption, which is named SEIRA. Since normal infrared absorption spectroscopy (IRAS) has a sensitivity high enough to detect monolayer adsorbates, and the SEF value is about 50 at the present stage, one might think that the difference in sensitivity is not very important. In fact, only a few groups have been engaged in this field. However, it must be noted that the signal-to-noise ratio of a spectrum is proportional to the square root of the co-added number of interferograms in Fourier transform infrared (FTIR) spectrometry. In IRAS measurements of monolayer adsorbates, hundreds to thousands of interferograms are required to be averaged to enhance the signal-to-noise ratio of the spectrum. In surface-enhanced infrared absorption spectroscopy (SEIRAS) measurements, on the other hand, the same signal-to-noise ratio as IRAS is expected to be obtained without interferogram averaging or with averaging of a smaller number of interferograms, on the basis of simple calculation. Therefore, SEIRAS is suitable not only for obtaining high-quality spectra from submonolayer adsorbate, but also for time-resolved monitoring of reaction processes at surfaces and interfaces.

The IR absorption enhancement has been observed for a variety of molecules and ions adsorbed onto vacuum evaporated thin metal films such as Ag, Au, Cu, In, Pt, Pd, Ni and Al. In general, Ag and Au exhibit the most intense enhancement.[13] The enhancement factor changes greatly depending upon the island structure (shapes and size) of the metal film. The enhancement is significant for the chemisorbed first layer and decays as the distance from the surface increases. The distance-dependence was investigated in more detail by using LB monolayers as spacers, and it was shown that the enhancement diminishes about 5 nm from the surface. The enhancement is the largest around the Brewster angle for p-polarization. The enhanced IR bands are superposed on a very broad and strong background absorption by metal island films ranging from the visible to IR regions. The enhanced absorption of the adsorbed molecule is proportional to the background absorption in the mid-IR region. The nature of SEIRA mentioned above is very similar to that of SERS observed on island films, suggesting that the enhancing mechanisms of SEIRA are similar to those of SERS.

The intimate relationship between the enhancement and electric field intensity strongly argues that a certain electromagnetic mechanism is contributing to the

Table 1 Characteristics of surface IR and Raman spectroscopy

	IR	Raman
Cross-section (cm^2)	$\sim10^{-20}$	$\sim10^{-29}$
Sensitivity	High	Low[a]
SEF for coinage metals	~80	$\sim10^{6}$[b]
SEF for transition metals	~30	~100–1000
Surface selection rules	Strict, simple	Relax, complex
Determination of molecular orientation and structure	Simple and clear	Difficult and complex (especially for SERS)
Spatial resolution (μm)	~30	~1
Spectral windows (cm^{-1})	Normally 900–4000[c]	Normally 30–4000
Band shape	Broad	Sharp
Background	High	Low
Interference by fluorescence	No	Yes
Intensity–concentration	Exponential	Linear
Sampling methods	Transmission Reflection, Diffuse reflection ATR	180° backscattering 90° scattering collection ATR
Sample preparation	Some requirements	No special requirement[d]
Sample destruction	Nondestructive	Sample heating or annealing
Interference from ambient environment	Strong interference from water and CO_2	Water and CO_2 are very weak Raman scatterers

[a] Traditionally, Raman analysis has been viewed as a relatively insensitive technique. Indeed, the probability of a molecule producing Raman scattering is much less than the probability of absorption in the IR region. However, advances in instrumentation and lasers have narrowed the gap considerably, to the point where, in some optimal cases, the sensitivity of normal (i.e. unenhanced) Raman scatter compares favorably with IR absorption. In addition, the Raman scattering efficiency can be greatly increased by use of surface-enhanced and resonance Raman effects.
[b] For a single silver nanoparticle or colloid with suitable size, the SEF value can be up 10^{14}.
[c] A current limitation is that the IR technique becomes difficult to apply at wavenumbers below about 800 cm^{-1}, because of the low intensity of the thermal IR source in this region. This difficulty can be overcome by the use of a synchrotron radiation source.
[d] To prepare highly active SERS substrates, proper surface roughening procedures are necessary.

IR enhancement.[13] In the visible region, a short-ranged strong electric field is produced around metal islands via the excitation of localized plasmon or metal islands, which dominantly contributes to the SERS on evaporated metal films. The enhanced local field around the islands is expected to play a role also in SEIRA. Localized plasmon is excited only along the major axis of the ellipsoids in the visible and near-IR regions. Since the major axis is parallel to the substrate surface, the enhancement is observed in the transmission geometry. In the attenuated total reflection (ATR) geometry, the electric field parallel to the substrate surface can give the absorption enhancement. In the external reflection geometry, the substrate is highly reflective; however, no enhancement is observed in the external reflection geometry because the electric field that excites metal islands is polarized normal to the substrate surface. IR absorption can be enhanced through a vibronic coupling of vibrational modes with the charge-transfer between the adsorbed molecule and metal surface. Despite several studies, however, the chemical effect has not been fully elucidated.

Another kind of enhanced IR phenomenon has been reported by Lu et al.,[14] who observed the enhancement of the surface IR signal on film surfaces of transition metals (e.g. Pt, Pd and Rh). The observation was made in an external reflection mode. The enhancement factors calculated fall in the range of that for SEIRA, typically 10 to 30. The maximum intensity is related to the film thickness and the metal (~12 nm). The key difference for the so-called abnormal IR effect with SEIRAS is the inversion of the absorption peak of the adsorbate. Although the detailed mechanism of this interesting phenomenon is still unknown, the high sensitivity facilitates the measurement of various surface species (e.g. CO, SCN^-, CN^-) on transition metals.

1.4 Surface Infrared and Raman Spectroscopy

Table 1 compares the features of IR and Raman spectroscopy.

2 HISTORY

2.1 History of Surface Infrared Spectroscopy[15,16]

The use of IR spectroscopy to study adsorbed species dates back to the 1940s. Terenin employed it to investigate

adsorbate on porous glass. Its application to adsorbates on metal surfaces was pioneered in the mid-1950s. The first attempts to obtain spectra from adsorbed layers on bulk metal samples were made in 1959, and in the same year LB films on silver were studied. During the next decade a mass of experimental data appeared and the late 1960s witnessed the expansion of this technique. The theory of reflection at a metal surface, that the optimum angle of incidence is close to grazing, was developed by Greenler in 1966.[15] Since then, IR spectroscopy has been applied to an ever-increasing range of adsorbates and surface intermediates on many metallic and nonmetallic substrates with the rise of modern surface science, including the study under UHV and of adsorbates on single crystal substrates. In the 1970s the technique came to prominence with much better signal-to-noise ratios and its combination with other modern surface analytical probes. In recent years, the application of Fourier transform techniques has enhanced the performance very significantly, and IRAS is now one of the most sensitive methods of the surface spectroscopic techniques. The limitation of difficulty in the low wavenumber region is now partially solved by the development of synchrotron radiation sources.

IRAS can readily discriminate between adsorbed and gas-phase species and spectra of adsorbed species have been recorded under pressures much higher than atmospheric using the transmission or diffuse reflectance technique with properly designed apparatus. Thus, it is able to bridge the ''pressure gap'' between surface science and catalytic studies. In the 1990s, sophisticated in situ cells enabled IR study to be carried out in situ over a wide range of temperatures (from liquid nitrogen temperature to 1073 K) and pressures (1 torr to 100 atm).

The application of IR spectroscopy to electrochemistry was first carried out by Bewick et al.,[16] and has undergone an explosive growth. The applications vary from the characterization of adsorbate orientation at submonolayer concentration at electrode surfaces to the determination of short-lived radical ions in the diffusion layer, and from aqueous to nonaqueous solvents. Efforts to improve the sensitivity of these measurements have been made in the late 1980s, and opened the way to probe the reaction dynamics of a variety of heterogeneous and homogeneous chemical processes. Since the 1990s, the emergence of synchrotron sources has promised to yield information regarding interactions between adsorbate and metal substrate, as well as changes in the vibrational modes of species confined within the electrical double layer. Thus various techniques have been proposed to enhance the sensitivity of monolayer adsorbates.

Enhanced IR absorption of organic thin films on Si substrate was reported by Hartstein et al. in 1980 by evaporating Ag or Au thin films onto organic films (metal-overlayer geometry). The enhancement was also observed by forming organic films on very thin metal films evaporated on Si (metal-underlayer geometry). Since the enhancement is rather weak for SEIRA in comparison with SERS, few papers were published in the 1980s. Recently, SEIRA has received renewed interest due to the development of its applications. This interest was triggered by experimental and theoretical studies of the enhancing mechanisms by Osawa,[13] and by the work of Nishikawa et al. on the application to the trace analysis of chemical compounds.

2.2 History of Surface Raman Spectroscopy[7,8,17,18]

Raman spectroscopy of adsorbed molecules on clean metal surfaces is a relatively new experimental technique in contrast to the related subject of IR spectroscopy of surfaces, which has been actively studied since the pioneering investigations by Terenin (1949) and Eischens (1954). Although cross-sections for Raman scattering are much smaller than those for IR absorption (e.g. $\sim 10^{-29}\,cm^2$ compared to $\sim 10^{-20}\,cm^2$), Raman spectroscopy enjoys the advantage of quantum detectors and continuous wave (CW) laser excitation of small samples. Burstein et al. studied Raman scattering from a variety of surfaces, beginning with studies of doped InSb crystals in 1968. As early as 1973, Raman spectroscopy excited by red laser has been applied in the study of carriers in p-doped silicon. A review about Raman spectroscopy application in semiconductors and their surfaces presented by Richter appeared in 1976.[17] In the 1970s, Raman spectra of organic species absorbed on high-area catalysts began to appear but the interest dwindled as fluorescence was found to be a persistent nuisance. Thus in this period, most Raman studies were concerned with the structures of the bulk catalyst itself. The study of adsorbate on the catalyst surface became prominent from the early 1980s. In the mid-1980s near-IR Raman spectroscopy and in the mid-1990s UV Raman spectroscopy were developed, and the renaissance of Raman applications in catalysis began. The first in situ Raman spectroscopic study on an electrochemical system was reported for thin metal oxide and metal halide film electrodes in 1973. Soon after that, the resonance Raman effect was employed to study surface reactions and time-resolved studies on a millisecond scale became possible to detect solution species produced by surface reactions.

The field of SERS was inaugurated in the mid-1970s. The first observation of Raman scattering at a silver electrode surface was made in 1974 by Fleischmann et al. and the full implications of their observations were not clearly recognized. Not until 1977 was it verified that Raman signals arising from pyridine on Ag electrodes were enhanced between four and six orders

of magnitude by Van Duyne et al. and Creighton et al., respectively. Since these first observations, there have been thousands of papers dealing with this phenomenon either theoretically or experimentally.[8–11,18] Extending SERS to the study of transition metals as well as single crystal surfaces has proved successful in the last 5 years, and single molecule detection by Raman spectroscopy has been accomplished in the past 2 years.[8,11]

3 INSTRUMENTAL METHODS

3.1 Preparation of Samples

In spite of the fact that IR spectroscopy and Raman spectroscopy can be used to study any solid surface, it is essential to choose the appropriate experimental apparatus, optical geometry, light source and sample preparation for the specific surface studied. In this section, we will discuss various ways of surface sample preparation. Because rough surfaces are more practically important and widely used (the preparation of active surfaces for SERS and SEIRA is vitally important to obtain high quality surface spectra), they will be described and discussed in more detail.

Basically, the solid surface samples can be classified into two categories: the flat surfaces include atomically smooth single crystal surfaces, polished mirror-like surfaces, ultrathin films such as polymer film, LB film and self-assembled monolayers (SAMs) on smooth metallic or nonmetallic substrates; the rough surfaces include micro- or nanoscale rough or porous surfaces, powders and colloids.

Two points may be stressed as to why powders and colloids are discussed at length while the main topic of this article is concerned with the surface: for the films, powders and colloids mentioned in this article, specific attention is paid only to their surface structures, the surface orientations and bonding of adsorbates instead of their bulk structures; the rough, porous and even powder materials discussed here are mainly for significantly increasing the surface area available for adsorption and reaction, and/or generating the SERS and SEIRA activity.

3.1.1 Preparation of the Flat Surfaces

Preparation of the flat surfaces is relatively easy in both IR and Raman measurements. For the mechanical polishing procedure, the solid sample, such as a metal disk sheet or metal rods, is polished first using emery paper followed by polishing with a polishing cloth with wet alumina powder of decreasing size from 5 μm down to 0.05 μm. The flat surfaces undergoing this procedure are either used directly or further treated depending on the experiment. If a well-defined single crystalline surface is being used, either a chemical etching procedure or flame annealing procedure should be used to produce atomically smooth surfaces.

For the ultrathin film samples such as LB films and SAMs, preparation is dependent on the technique. For example, in IR transmission spectroscopy the film is deposited on a transparent substrate such as CaF_2, while in reflectance spectroscopy the film is deposited on a gold film coated substrate. In some cases the flat surface samples are held by adhesion to a spindle capable of rotation about its axis in order to avoid the sample overheating.

In Raman measurements, the laser can be focused on a small area with the aid of a microscope. This means that a very small surface sample, e.g. with diameter of 0.5 mm or less, can be used. However, it may be experimentally difficult if the sample is a very small disk or a thin wire embedded in the holder and its surface needs to be well polished mechanically.

3.1.2 Preparation of the Rough Surfaces

Preparation of the rough surface for surface IR measurement depends on the optical geometry used. In the transmission mode most powder samples are run as potassium bromide disks, as mulls. Although this sampling method provides an easy and adequate way of obtaining spectra, difficulties may arise as a result of water absorption or undesired reactions with the salts. So, some powders and polymer samples are pressed to make a self-supporting film. The thickness is typically between 20 and 1000 μm. In the diffuse reflectance mode samples can be measured as loose powders, with the advantages that not only the tedious preparation of wafers is unnecessary but also that diffusion limitations associated with tightly pressed samples are avoided. It becomes an important advantage in some cases, especially for in situ studying of some catalysis reactions involving diffusion processes. Diffuse reflectance is also the indicated technique for strongly scattering or absorbing particles. ATR or attenuated multiple internal reflectance is also used for the study of dark solids. The sample, prepared as a fine powder, is dispersed on the two faces of a KRS5 (TlBr + TlI) single crystal, which has a high refractive index. Samples for the internal reflectance mode should be dealt with care to ensure that the powders and the surface are in very good contact.

Preparation of the rough surface for ordinary surface Raman (non-SERS) measurements is quite simple. Powders, colloids and rough solid surfaces can all be measured directly in the container in which they are supplied provided that the container is transparent to

the laser and Raman radiation. Powdered catalysts are prepared generally by grinding and calcining. Some powders can be simply held by pressing them into a sample holder, e.g. put on the filter in a capillary tube, or by using a piston to compress them into a soft pellet.

3.1.3 Preparation of Surface-enhanced Raman Scattering-active Substrates[7–11]

It is well known that for obtaining good-quality SERS or SEIRA spectra, the surface must be pretreated properly. Figure 4 illustrates the distinct effect of the surface roughening procedure on the Raman intensity of pyridine adsorbed on Ni electrode surfaces that only have weak SERS activity. The surface signal is extremely weak from the mirror-like surface after mechanical polishing with fine grades of alumina powders down to 0.05 μm. Even the strongest band intensity for the adsorbed pyridine is extremely weak, only about 0.5 counts s^{-1}, Figure 4(a). The spectrum is certainly too weak to be investigated in detail. The signal-to-noise ratio was improved after the electrode was chemically etched in 1.0 mol L^{-1} solution of HNO_3. The signal intensity increased doubly if the electrode was further treated in a 0.1 mol L^{-1} solution of KCl by a double-step ORC. An even more intense signal was obtained when the additional ORC was performed in situ in the spectroelectrochemical cell prior to the measurement. These results show that there are several ways to roughen the electrode surface and the correct surface roughening procedure is the key to obtaining reasonably good spectra. As some surface preparations for SERS-active substrates are unique and complicated, they are described in detail in the following.

3.1.3.1 Oxidation–Reduction Cycle In general the electrode is made of a metal wire, rod or disk sealed into an inert holder (glass, Kel-F or Teflon). The electrode surface is then polished mechanically using decreasing sizes of alumina powders (normally down to 0.3 μm or 0.05 μm). Before the SERS activation procedure

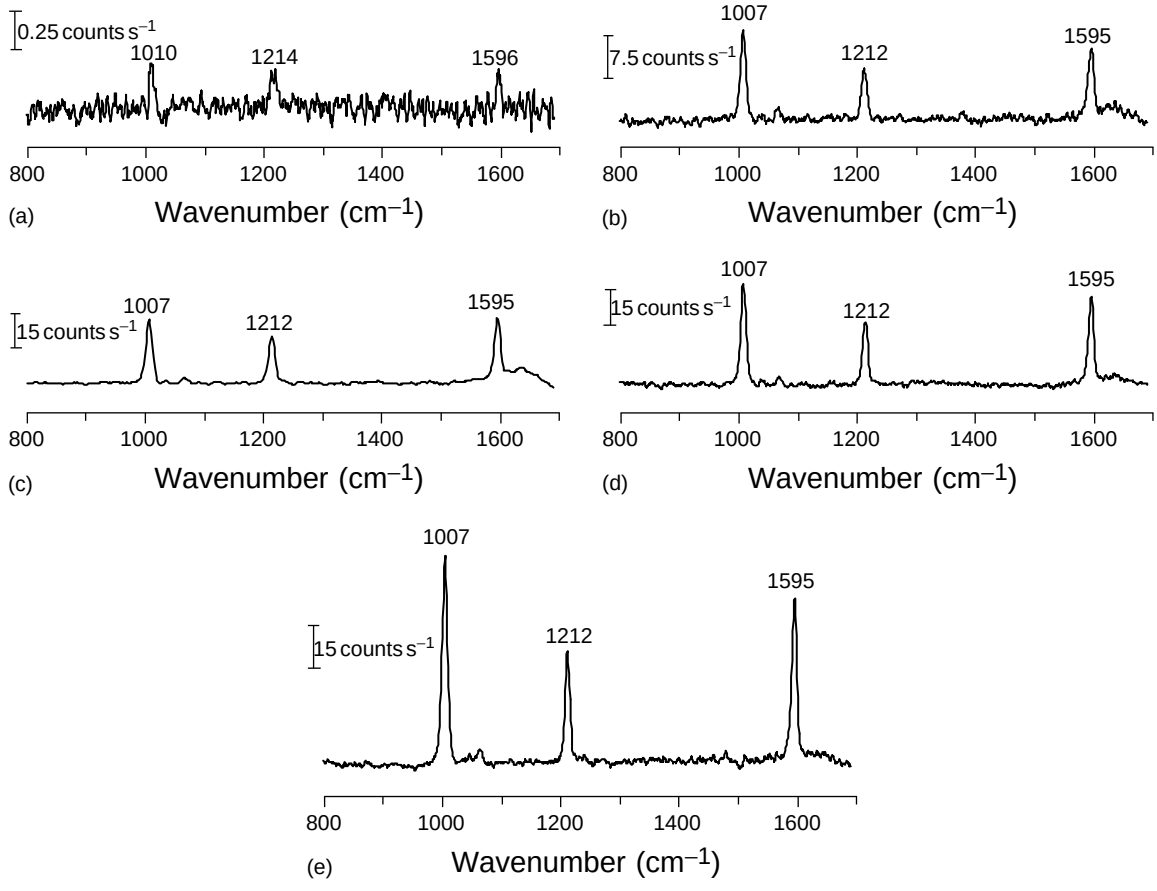

Figure 4 Raman spectra of pyridine adsorbed on nickel surfaces with different roughening procedures in a solution of 0.01 mol L^{-1} pyridine and 0.1 mol L^{-1} KCl: (a) mechanical polishing; (b) chemical etching in 1 mol L^{-1} HNO_3; (c) electrodeposition of nickel onto Ni electrode in 0.1 mol L^{-1} $NiSO_4$; (d) ex-situ ORC in 0.1 mol L^{-1} KCl by a double-step ORC from −0.4 V to 0.4 V where the potential is held for 3 s and then back to −0.4 V; (e) in situ ORC in 0.01 mol L^{-1} pyridine + 0.1 mol L^{-1} KCl in the spectroelectrochemical cell prior to the measurement. The laser excitation line is 632.8 nm.

Table 2 Electrochemical roughening procedures for various metal electrodes

	Pretreatment	Electrolyte	Potential waveform	Parameters	Cycles	Appearance
Ag	Chemical etching (optional)	0.1 M KCl	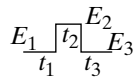	$E_1 = E_3 = -0.25$ V $E_2 = 0.18$ V $t_1 = 15$ s, $t_2 = 8$ s, $t_3 = 60$ s	1	Milky yellow
Au	Electrochemical cleaning (optional)	0.1 M KCl	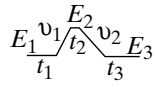	$E_1 = E_3 = -0.3$ V $E_2 = 1.2$ V $t_1 = t_3 = 30$ s, $t_2 = 1.2$ s $\upsilon_1 = 1$V s^{-1}, $\upsilon_2 = 0.5$ V s^{-1}	25	Dark brown
Cu	Electrochemical cleaning (optional)	0.1 M KCl	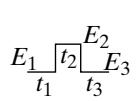	$E_1 = E_3 = -0.4$ V $E_2 = 0.4$ V $t_1 = 15$ s, $t_2 = 3$–5 s $t_3 = 60$ s	1–5	Dark brown
Pt	Electrochemical cleaning	0.5 M H_2SO_4	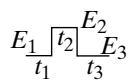	$E_1 = E_3 = -0.2$ V $E_2 = 2.4$ V $f = 1.5$ kHz $(t_1 + t_2 = 1/f)$	30–900	Gray to dark yellow
Fe	Chemical etching (optional)	0.5 M H_2SO_4	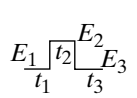	$E_1 = E_3 = -0.7$ V $E_2 = 0.35$ V $t_1 = t_3 = 60$ s, $t_2 = 15$ s	1	Gray

performed by the electrochemical ORC, the electrode is sometimes treated by a chemical etching procedure for further smoothening the surface and/or by an electrochemical cleaning procedure by controlling the electrode potential to severe hydrogen evolution. Experimental variables in the ORC include the oxidation and reduction potentials, the type of potential–time function utilized (triangular-wave potential sweep or double potential steps), and the amount of charge passed during the oxidation step. The applied potential, especially the oxidation potential, is dictated by the electrolyte and the electrode used. For example, in the case of 0.1 mol L^{-1} K_2SO_4, oxidation of Ag is accomplished near +0.50 V vs saturated calomel electrode (SCE), the optimum SERS signals are obtained following the charge passage of 20–50 mC cm^{-2}. This is equivalent to ca. 100–200 monolayers of Ag atoms. For some electrodes such as Pt and Fe, the same SERS signal intensity resulted when the adsorbate was present during the ORC (i.e. in situ ORC) as when the adsorbate was added after the ORC (i.e. ex situ ORC). It has been observed that illumination of some electrodes such as Ag, Cu and Ni during the ORC in the presence of halide can yield a stronger SERS signal. Table 2 is a summary of typical experimental variables for various metal electrodes. Figure 5 shows a roughened Pt electrode and a Au nanorod array that exhibit high SERS activity.

3.1.3.2 Chemical Etching Although chemical etching looks quite simple, it still requires experimental skill to optimize the SERS activity for different metal samples. Copper foils were immersed into 2 mol L^{-1} HNO_3 solution at 303 K, and after about 10 s bubbles were formed near the copper surface. Vigorous agitation was then applied. After 2 min, a sponge-type surface with substantial roughness was created. In the case of Ag foils, a 5–6 mol L^{-1} HNO_3 solution was used and the etching time was prolonged to 5–10 min until the surface became milky. For Ni surface, a dilute solution of HNO_3 with concentration about 1 mol L^{-1} was used, and the etching time was about 5–7 min. For the Fe surface, the best result was obtained by etching in an ultrasonic bath of 2 mol L^{-1} H_2SO_4 for 5 min. The surface should be gray. After the etching the metal foils or surfaces should be thoroughly rinsed with water and kept wet in order to protect the roughened electrode and to prevent contamination from the atmosphere.

3.1.3.3 Film Deposition Silver and gold form hemispheroidal particles when thin layers (ca. 5 nm) are vacuum-deposited on various substrates, including silica slides. A new procedure for preparing SERS-active Ag substrate is sputter-depositing Ag onto various substrates (metal, graphite, alumina, and even filter paper). This method produces surfaces that are stable in air and can be SERS-activated by laser irradiation in the presence of an adsorbate that complexes with Ag. Recently, a chemical procedure was developed for preparing silver island films using Tollens' reagent; this is rapid, simple

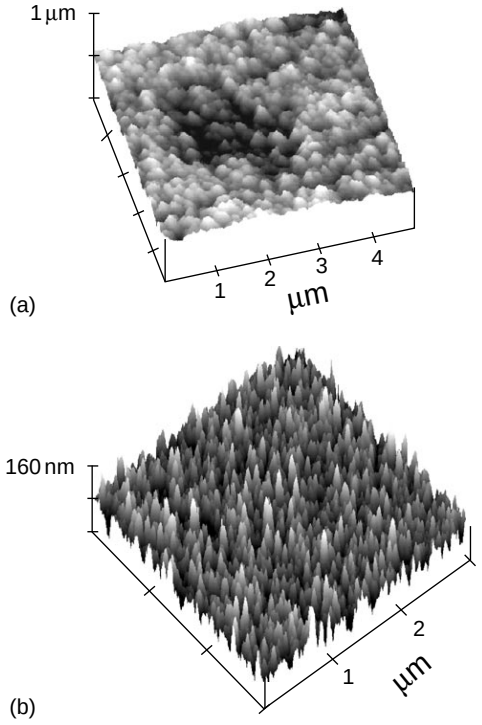

Figure 5 (a) STM (scanning tunneling microscopy) image of a Pt electrode surface after treatment of electrochemically periodic roughening in a sulfuric acid solution. (b) AFM (atomic force microscopy) image of a Au surface fabricated by electrochemical deposition into alumina porous film.

and highly reproducible. Very recently, some rough films of transition metals such as Ru, Rh, Pd, Ni, Co and Fe, exhibiting weak SERS effect, have been prepared by a simple method of electrochemical deposition onto carbon electrode surfaces in the solution containing the metal ions.

3.1.3.4 Chemical Reduction and Laser Ablation Metal sols are prepared by chemically reducing a dissolved metal salt by means of an appropriate reducing agent in either an aqueous or a nonaqueous medium. Typically, silver sol can be obtained by adding 1 mL 2×10^{-3} mol L^{-1} $AgNO_3$ into 50 mL 5.5×10^{-3} mol L^{-1} $NaBH_4$ solution (prepared at 277 K) drop by drop slowly over an ice bath by continuous stirring for 1 h. There is a UV absorption at 380 nm for this kind of sol, and it is yellowish and the solution should be clear. Then the sol is kept in the dark for measurement. It should be noted that the size, shape and aggregation of metal sols could be controlled critically by the solution concentration and composition, mixing rate of the two solutions and the temperature. Metal colloids could also be prepared by exposing metal foil to the focused radiation of high power Nd : YAG (neodymium : yttrium aluminum garnet) laser (55 mJ $pulse^{-1}$) pulse for 10 min for silver and gold and 15 min for platinum and copper in 15 mL solvent solution. The solvent could be organic solvent or pure water. Metal colloids prepared by this method are free of interference from the additive.

3.1.3.5 Self-assembled Monolayers Technique To obtain SERS substrates with ease of preparation, good reproducibility, stability and compatibility with biomolecules, a novel procedure has been developed by Natan et al. based upon the SAMs technique. They arrange monodisperse colloid gold and silver particles regularly onto functionalized metal or glass substrates. Another important aspect of this method is the ability to control the particle size and density, and the space between the substrate and the particle, and thus to tune the electromagnetic characteristics of the surface for systematic study of the SERS mechanism.

3.1.3.6 Lithography SERS substrate having a very narrow distribution of roughness feature is very important for both fundamental and applied studies. Such SERS substrates have been prepared by nanosphere lithography, in which SERS-active metal is evaporated into preformed arrays of nanospores which are then removed, leaving behind the metal particles formed in the interstices. An alternative way to obtain regularly structured SERS substrates is to use metal film or particle deposition over periodic arrays of polymer nanospheres or silica posts, which are generated by electron-beam lithography. This allows the optimization of SERS enhancement and investigation of the relationship between the parameters and the enhancement in comparison with conventional island films and also with regularly arranged nanoparticles as mentioned above. The possibility of generating two-dimensional ordered arrays with SERS-active structures with different structural parameters on one substrate in a single production cycle further increases the usefulness of the electron-beam method in the fast development of optimum SERS substrates. By evaporating the metal at a different angle along a direction in which the posts shadow each other, isolated metal particles are obtained on top of each post, with different shape.

3.1.3.7 Overlayer Technique To extend SERS study to many metals or semiconductors, which only have weak- or non-SERS activity, two important approaches related to the overlayer technique have been made since 1987. The first one is to coat SERS-active Ag or Au electrodes with ultrathin films of metals or semiconductors by electrochemical deposition or evaporation, as developed by Fleischmann et al. and Weaver et al., respectively. With the aid of the long-range effect of the electromagnetic enhancement created by the SERS-active substrate underneath, SERS spectra of adsorbates on these films can be obtained. This method provides

a unique way to study various metal surfaces so as to offer enormous potential for the surface Raman spectroscopic study of catalytically important reactions at surfaces. It can also be extended to study solid–solid interfaces, through which metal oxides, semiconductors, polymer or LB films are coated by different ways onto the SERS-active substrates. Strong SERS signals provide some meaningful information of solid–solid interfacial structures that are difficult to be obtained by other means.

The second is to use the reverse geometry of the overlayer technique, in that the SERS-active metal (e.g. Ag and Au) islands are deposited electrochemically or chemically onto non-SERS-active substrates including semiconductors, metal oxides or polymers, as first carried out by Van Duyne et al. By taking the advantage of the long-range effect of SERS, the structures and properties of the substrates themselves of non-SERS activity can be investigated.

3.1.4 Preparation of Surface-enhanced Infrared Absorption Substrates

Preparation of SEIRA substrates is relatively simple. Very thin metal films are evaporated on IR transparent substrates to form discontinuous layers consisting of small islands. The average dimensions of the islands are about 30 nm in diameter and 10 nm in height for a 10 nm thick Ag film evaporated on a Si substrate. The enhancement factor greatly depends upon the island structure (shapes and size) of the metal film. Slower deposition rate and lower temperature give the larger enhancement.

3.2 Experimental Set-up and Operation Modes

The structure and dynamics of various interfaces, including solid–gas, solid–liquid and solid–solid interfaces, play an increasingly important role in modern science and technology. The ramifications of the subject extend into areas as diverse as catalysis, material corrosion, semiconductor processes, batteries, electric synthesis of organic and inorganic compounds, and biological electron transfer processes. The application of spectroscopy to specific in situ surface (interface) analysis depends on its ability to discriminate between molecules at the surface (interface) and in gas phase, liquid or solid phase. In order to use IR and Raman spectroscopy to study the above systems in situ, it is essential to design the sample cells and apparatus for different chemical and physical environments. In fact, the interfaces can be classified into solid–gas, solid–liquid, solid–solid and solid–UHV interfaces, as discussed below.

3.2.1 Solid–Gas Interfaces[19–21]

3.2.1.1 Cells There are many types of IR and Raman cells specially designed for the study of solid–gas interfaces as shown in Figure 6. Here we describe two examples for IR and Raman measurement, respectively, for heterogeneous catalysis with the capability of controlling temperature–pressure–vacuum–atmosphere. These complicated cells can be simplified for the purpose of investigating solid–air and solid–UHV interfaces accordingly. The transmittance IR cells shown in Figure 6(a) and 6(d) are specially designed to treat the sample in situ. The cell is fitted with a furnace, cooling water jackets, gas inlet and outlet, and with BaF_2 windows. The cell material should be inert to the catalytic reaction, and can be quartz or stainless steel. The latter cannot be used at high temperature or in some environments where stainless steel can react with the gas. The gas inlet and outlet of the IR cell are connected to two-way and three-way valves, respectively. By switching these valves, the sample in the IR cell can be treated under vacuum exposed to different gas atmospheres or react with a flow of gas mixture. The sample holder of the cell can be heated up to 1073 K. Before the IR experiment, the catalyst was pressed into a self-supported disk then fixed in a slide of ceramic tube that is held in the cell. The sample is first treated in O_2 at high temperature, e.g. about 1023 K, for 30 min followed by evacuation under vacuum (10^{-3} torr or lower) at the same temperature until surface contaminants such as carbonate species are completely removed. The treated sample is then cooled down under vacuum to the desired temperature for adsorption or reaction. The reported absorbances were obtained from spectra referenced to the background spectra which were taken

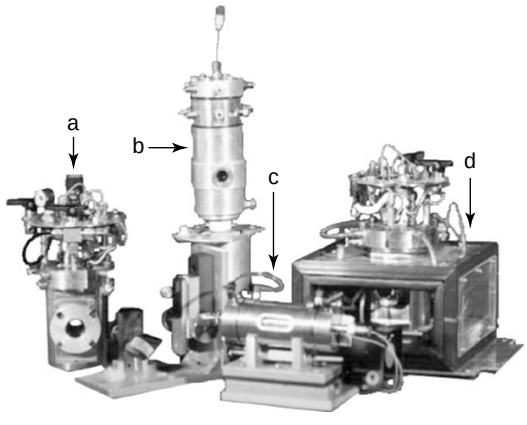

Figure 6 A set of cells for in situ study on solid–gas interfaces: (a) IR cell for transmission mode and with reactor; (b) and (c) Raman cells for various optical collecting configurations; (d) IR cell for transmission mode. (Reproduced by permission from Y.Y. Liao.)

under vacuum prior to the introduction of the reactants to the catalyst.

The Raman cell for catalysis usually uses backscattering mode because many catalysts are dark in color. Quartz and Pyrex glass are used as the cell materials as they are sufficiently nonreactive to retain many reactive samples like oxidizing agents. For surface Raman measurements, the exciting and scattered frequencies are close to each other and are in the visible range (except for UV Raman measurements). Thus there is generally no problem of cell transparency and quartz is used as the window material. Figure 7 shows a detailed schematic diagram of the Raman cell for Raman microscopy with high performance, including high scattering light throughput, for in situ studies. The Raman cell consists of three parts: sampling, heating (cooling) and sealing. The solid (powder) sample is put on the filter in the capillary tube. The cell is heated by tubular heater, which is inserted in the hole of the cylindrical block (furnace) or it is cooled by streaming nitrogen pretreated through a copper coil in a liquid nitrogen Dewar trap. The temperature ranges from 170 to 1000 K with accuracy of ±1 K. The pressure can reach 1.5 MPa or more and can be vacuumed down to 10^{-5} Pa. The quartz sample capillary tube and outside quartz sleeve are sealed by assembly of the O-rings, gaskets and screw caps. When working at high temperature, high efficient cooling jackets are very important to keep the O-rings at lower temperature and protect the objective when Raman microscopy is used. Moreover, some cells are designed to be moved up and down using mechanical or pneumatic apparatus for minimizing the thermal effects and damage of the sample by laser beam, see Figure 6(b) and (c).

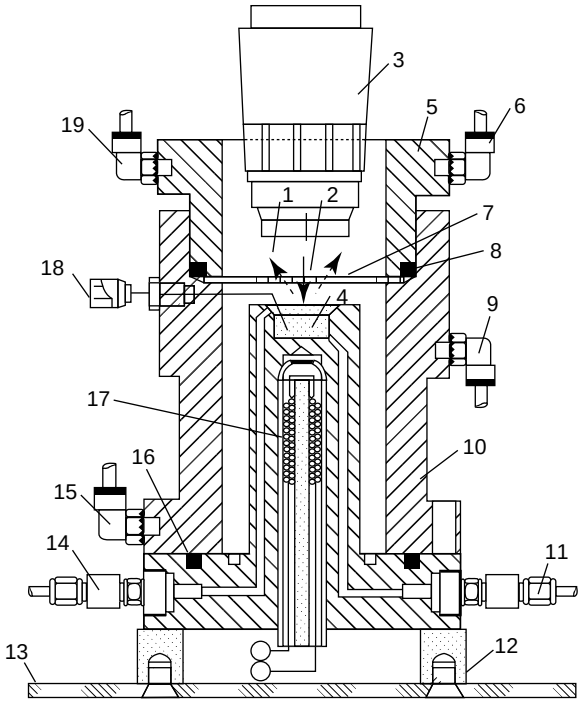

Figure 7 Schematic diagram of a detailed in situ Raman cell for a Raman microscope: 1, laser beam; 2, scattering light; 3, microscope lens; 4, sample chamber; 5, top cap; 6, 9, 15, 19, Legris push-in fitting for cooling system; 7, quartz windows; 8,16, O-ring; 10, stainless steel outside sleeve; 11, 14, swagelok tube fitting; 12, electrical insulator block; 13, bottom plate; 17, heater; 18, sheathed thermocouple (K-type). (Reproduced by permission from Y.Y. Liao.)

It is necessary to mention, although very briefly, the cell for ATR or attenuated multiple internal reflectance modes for the study of dark solids. The IR beam enters through the input side of the crystal and reflects from the two faces. The crystal is put into a suitable sample holder, which is placed in a metallic cell, allowing spectra to be obtained under various temperatures, pressures, and gas compositions. This method has disadvantages because the crystal may react chemically at temperatures higher than 523 K.

3.2.1.2 Modes[1,19,21] Figure 8 illustrates optical geometries for IR spectroscopy, which are the most commonly utilized in surface analysis. They include the transmission, single RA, diffusion reflectance, single internal reflection, (multiple) attenuated total reflectance and emission modes. The techniques are only briefly referred to in this section, since they are covered in detail elsewhere in the Encyclopedia. The emphasis will be laid

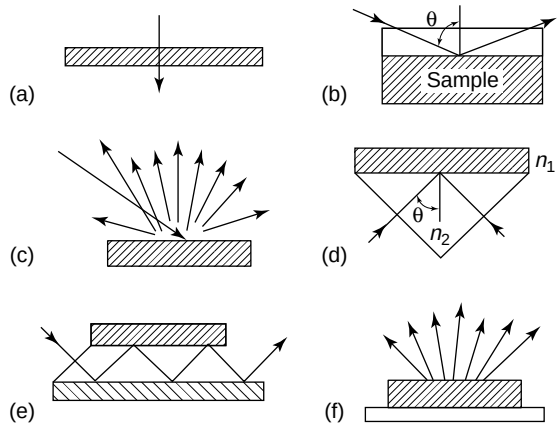

Figure 8 (a) IR transmission mode: the light penetrates through a thin film sample and is analyzed by the detector. (b) Single RA mode: light penetrates an ultrathin surface layer and is reflected by the mirror-like metal. (c) Diffuse reflectance mode: diffusely scattering light is scattered and collected by mirrors and redirected to the detector. (d) Single internal reflection mode: light passes through the IR spectroscopy element and is totally reflected. (e) Multiple-reflection (attenuated total reflectance) mode. (f) Emission mode: the sample is heated and the emitted light is analyzed by the IR detector.

on the experimental set-up for the in situ study of various interfaces.

Transmission Mode. Transmission is the most commonly used IR technique to obtain the spectra of powders. In many circumstances transmission measurements also provide the more direct, robust and reproducible quantitative methods for single-component analyses, operating over a wide dynamic range and at high sensitivity. The powder sample is dispersed in the IR-transparent powder (KBr, CsI, NaCl, etc.) as routine, then is pressed together to form a pellet. However, for in situ investigations on a catalyst this technique is not applicable, and pressed self-supporting wafers have to be used. The applicability of the transmission technique is determined by the properties of the solid powder under study. If the sample is a thin, free-standing film, the transmission technique provides an excellent means of obtaining spectra. One of the disadvantages of transmission IR for supported catalysts is that the spectral range accessible is limited by the transparency of the support. Samples which exhibit only weak bulk absorption, with average particle size d smaller than the wavelength of the IR radiation in the region of interest, will be optimally suited for the transmission mode. The particle size condition ($d < \lambda$) which determines the wavelength range of suitable low scattering losses, is usually met in the mid- and far-IR regions, whereas scattering losses become strongly involved in the near-IR region. However, most samples show strong bulk absorption in the low wavenumber region (roughly $<800\,\mathrm{cm}^{-1}$). As a consequence, the low frequency range, in which the vibrational modes of interest for structural characterization of the solid powder material are frequently located, is often not accessible to in situ studies using the pressed-wafer technique in transmission mode. Another drawback of transmission work is that the use of compressed disks limits the diffusion rate of reactants and products, so that kinetic studies are difficult.

Reflectance Modes. The external reflection technique of RA is primarily used for observations on thin films and adsorbates on reflective substrates. Although specular reflectance can be of value as a rapid, nondestructive generic ''fingerprinting'' tool, and quite extensively used in electrochemistry, its applicability is very limited in solid–gas interfaces. The normal procedure for RA spectroscopy involves the adjustment of the incidence and reflectance angle of the IR beam by about 60–80°. In order to get optimum sensitivity, a grazing angle procedure is used, usually an incidence and reflectance angle of about 80° is employed. The IR beam therefore traverses a considerably longer path through the sample than it does with small angles and produces a stronger spectrum. The diffuse reflectance mode is a convenient means of examining many finely powdered or highly scattering solid samples. It has particular importance for a wide range of near-IR applications. Internal reflection spectroscopy is a technique that measures the optical spectrum of a sample that is in contact with an optically dense and transparent medium. It is most appropriately used to study the surface layer characteristics of a continuous solid sample. This technique was later modified experimentally by increasing the number of reflections, giving rise to attenuated total reflectance spectroscopy, which was originated to study surfaces and thin films. Finally, in specific circumstances, emission spectroscopy provides an alternative and more favorable approach to surface or thin film studies.

Polarization Modulation Techniques. This technique is used to separate surface from bulk-phase absorption in metal–gas interfaces, and the polarization of the incident light is modulated between p- and s-polarization in the reflectance mode. In practice, s-polarized radiation, which has virtually no field strength at the metal surface, is usually filtered since it contains no useful information for adsorbed species. Since only p-polarized light is surface-sensitive, the modulation between the two polarizations yields an AC signal due to the species at the surface. This results from the fact that the intensity for p- and s-polarized light will, on average, be attenuated to the same degree by any randomly oriented gas phase molecules.

In the measurement the detector signal is divided into two channels for signal normalization. One signal proceeds normally, and the interferogram, which is equivalent to the integrated beam intensity in time $(I_p + I_s)$ is stored. The other signal is demodulated at the photoelastic modulator frequency, giving an interferogram proportional to $(I_p - I_s)$, which is in turn stored in the instrument computer. These stored interferograms are then Fourier transformed and ratioed to give a normalized intensity $(I_p - I_s)/(I_p + I_s)$, which is proportional to absorbance. Sum and difference interferograms are alternately sampled during the same interferometer mirror scan.

3.2.1.3 Surface-enhanced Raman Scattering Study on Solid–Gas Interfaces[20] Only a technique with high sensitivity can perform kinetic measurements using real-time measurements, enabling the role of the adsorbed species in the catalytic reaction mechanism to be clarified in situ. Clearly, SERS offers a means to overcome the obstacles of sensitivity, allowing the receipt of surface vibrational data during catalyst reaction. However, SERS has not been applied extensively to study gas–solid interfaces at high gas pressures although it has served as a valuable surface probe in electrochemical environments. Two key stumbling blocks in applying SERS at gas–solid interfaces are often perceived to be the limitation

of metals to be only silver, gold and copper and the lack of thermally stable SERS-active materials of catalytic importance. Based on overlayer strategy (see section 3.1.3.7), SERS study has been readily extended to a range of other materials by depositing them as ultra-thin films on roughened gold substrate. Williams et al.[20] have demonstrated that these electrochemically prepared films also exhibit stable, intense SERS activity in the gas phase over a range of elevated temperatures. The chemical inertness of the gold substrate, especially at elevated temperatures in the gas phase, is a further fortunate circumstance in that SERS signals usually then are associated with adsorption on the catalytic overlayers, rather than on the gold support. SERS can provide in situ information about vibrational modes of adsorbed species over a wide range of gas pressures, without associated problems of bulk-phase interference. Moreover, it allows real-time sequences of Raman spectra for interfacial species to be obtained with a rapidity ($\leq$10 s) that is commensurate with temperature programmed or other experimental perturbations commonly employed to yield catalytic information.[20]

3.2.2 Solid–Liquid Interfaces[16,18,22–24]

3.2.2.1 Cells Most IR and Raman cells for the in situ study of solid–liquid interfaces at room temperature and ambient pressure are made of glass or Teflon. They are sealed with an optical window with an epoxy resin or with an O-ring seal. The latter method is preferable to eliminate the possible contamination of the electrode and solution from the slow dissolution of the epoxy. Incident radiation passing through the window and solution layer onto the electrode surface is reflected or scattered and is subsequently focused onto the detector. The choice of window material depends on the spectral region of interest and on the solvent. Examples of IR window material for the mid-IR region include calcium fluoride, silicon, and zinc selenide, each of which may be used with a variety of aqueous and nonaqueous electrolytes. For the far-IR region polyethylene is usually employed.

A widely used IR cell and the schematic diagram of the experimental set-up are shown in Figure 9. The working electrode is mounted on a piston constructed either from glass, Teflon or Kel-F. Electrical connection to the working electrode is made through a channel in the plunger. The flat electrode surface is usually polished to a specular finish with an alumina slurry. As a final step, the piston is inserted into the cell and positioned concurrently to ensure that all surfaces are parallel with each other. A Luggin reference probe is positioned about 1 mm from the edge of the working electrode. The counter electrode is typically a platinum wire placed behind the working electrode in a configuration that minimizes

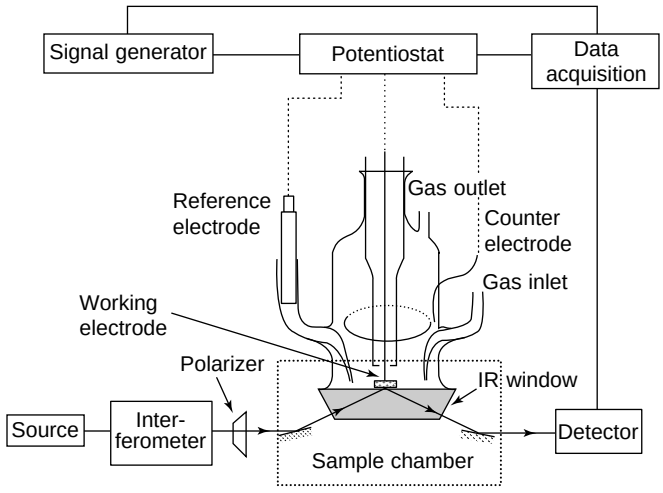

Figure 9 Schematic diagram of the experimental set-up for in situ electrochemical IR measurements.

solution resistance and provides as symmetric a current distribution pattern to the working electrode as possible.

It is vitally important to avoid strong interference from the same species in the solution phase or solvent by an extremely small quantity of surface (interface) species. This is especially serious for the IR study if the solvent is water, which absorbs strongly throughout most of the mid- and far-IR regions. This problem is minimized by using a thin-layer cell. To obtain the best spectra in IR or non-SERS measurements, one should position the electrode surface about 10 μm behind the optical window. Further discussion in this aspect will be given in the following sections. It is inferred that SERS has a great advantage in the investigation of two condensed phase interfaces as the thickness of the solution phase can range from 0.5 mm to 5 mm. Therefore, the cell configuration and electrochemical behavior is quite similar to that of the conventional electrochemical cell. For situations where the solution contains strong Raman interference, such as from an organic solvent, a resonance Raman-active species, or a fluorescing species, the working electrode which is parallel to the cell windows in the backscattering configuration can be placed against the windows to greatly reduce the solution pathlength.

Some cells require special design, such as in situ examining of an electrochemical corrosion process at high temperatures and pressures. Melendres[22] designed more complicated cells to study the corrosion and passivation of metals in solution from 298 to 563 K at 100 bar pressure. The design contains all the essential features for this type of work including solution exchange, and could be adapted to suit most instrument arrangements. In Raman spectroscopy, the signal is in proportion to the solid angle of the collecting optics as well as the throughput of the system in use. Pettinger[23] developed

a special electrochemical-optical device which consists of an elliptic mirror with a high eccentricity to which an electrochemical cell is attached from the back. The window for the cell is a half spherical window. For an optimum optical configuration, the electrode surface and center of the half sphere should be located at the first focal point of the elliptic mirror. This mirror serves as incident and collection optics, and the scattered light is focused to the second focal point of the mirror. The efficiency of this device in collecting and transmitting the scattered light through the entrance slit is about 50%. This is surprisingly higher than that collected by a microscopic system where with a $\times 50$ objective (numerical aperture, NA = 0.5) the value is only about 13.4%.

3.2.2.2 Methods For studying solid–liquid interfaces, electrochemical instrumentation is generally necessary for the set-up, including a potentiostat and waveform generator (see Figure 9). The latter should be capable of generating a potential pulse and a square triangular waveform superimposed on an initial potential. It should be noted that the thin-layer cell arrangement still exhibits large solvent and electrolyte absorption. There are several methods having signal-to-noise enhancement for surface (interface) species.

Electrochemically Modulated Infrared Spectroscopy. Electrochemically modulated infrared spectroscopy (EMIRS)[3,16] is the first IR technique applied in electrochemistry that employs electrode potential difference tactics. A change in IR reflectance can result from a change in potential for a number of reasons. These include changes in the coverage of species adsorbed on the electrode, changes in the nature of bonding between adsorbed molecule and the electrode surface, migration of ions into or out of the optical path, and electron-transfer reaction that consume and/or generate species in solution or on the electrode surface. A change in band intensity or the appearance of new bands may result. A band position may change due to changes in the frequency of the oscillator as the strength of bonding to the surface is changed with electrode potential. Since the spectra of the species in bulk solution, electrolyte, solvent and atmosphere generally do not change with potential, they are effectively nulled when two spectra obtained at two potentials respectively are ratioed.

EMIRS measurement is accomplished through the use of phase-sensitive detection and/or signal-averaging techniques. When a dispersive spectrometer is used, the potential is modulated at a frequency dictated by the type of detector used (10 Hz to about 1 kHz). Phase-sensitive detection is used to synchronously demodulate the measured reflectance at a fixed wavelength. Amplification of only that component of the optical signal having the same frequency as the modulation and appropriately related in-phase with the potential modulation increases the sensitivity to the level adequate for detection of IR bands of submonolayer quantities of adsorbates. The wavelength is incrementally scanned over the desired spectral region and the signal integrated at each wavelength. Signal averaging of several spectra further enhances the signal-to-noise ratio. The difference in the IR reflectivities normalized to the absolute reflectivity, $\Delta R/R$, as a function of wavelength effectively removes the wavenumber dependence of instrumental energy throughput, detector sensitivity, and electronic response throughout the beam path. It yields the difference in the IR spectra between the two electrode potentials. Figure 10 illustrates some possible difference spectra, by which different vibrational characteristics on potential changing can be deducted. EMIRS offers great sensitivity under favorable conditions, with $\Delta R/R$ values as small as 10^{-6} being detectable.

Electrochemical Infrared Absorbance Spectroscopy. Electrochemical infrared absorbance spectroscopy (EC-IRAS)[3] is now the most commonly used IR technique in electrochemistry, that also involves the acquisition of $\Delta R/R$ values, usually between two electrode potentials, but now by acquiring a suitable number of interferograms at fixed potential between each modulation, then

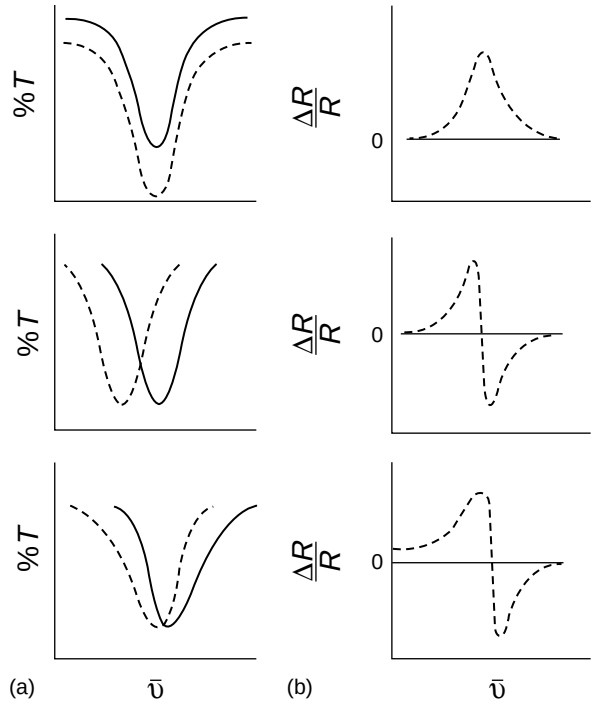

Figure 10 Examples of (a) possible IR reflection spectra at the two different potentials and (b) the corresponding EMIRS difference spectra.

co-adding those obtained at E_1 and E_2. The multiplex advantage of the Fourier transform technique allows for very rapid spectral acquisition, but there is lower noise discrimination per scan than in EMIRS. The time to obtain spectra by each technique at the same sensitivity is therefore approximately the same. Consequently, the rate of potential modulation is considerably less than in EMIRS, typically 10^{-3}–0.1 Hz. A schematic diagram for this technique is shown in Figure 9, from which one can deduce many potential dependent characteristics of surface species. This method was originally dubbed 'subtractively normalized interfacial Fourier transform infrared spectroscopy' (SNIFTIRS). As this method is now so commonly used in electrochemistry, many people refer to it as EC-IRAS for simplicity. To stress the employment of potential-difference tactics in electrochemistry, some groups also refer to it as potential difference infrared spectroscopy (PDIRS).

EC-IRAS offers considerably greater flexibility in the electrode potential waveform that can be employed, broadening the applicability of the method to encompass reactive as well as stable electrochemical systems. Data accumulation may be accomplished by various means. For example, sequences of potential-difference spectra can be obtained during a single electrode potential-step or -sweep waveform, enabling irreversible voltammetric conversions of adsorbed species to be monitored in a fashion that is not feasible with EMIRS. The simple potential-step method has several advantages. One is that electronic and mechanical drifts over long periods of time are cancelled effectively. Also, interference from atmospheric absorbances due to water and carbon dioxide may be easily eliminated by cancellation in the ratioing of the spectra. Next, since substantial product accumulation may occur in some electrochemical systems over long time periods, the measured difference signal could be decreased if the potential is held at a fixed value for a long period of time.

Polarization Modulation Techniques.[3] As has been mentioned for the solid–gas interface systems, the polarization of the incident light can be modulated between p- and s-polarization so the measurement can be performed without an alteration of the applied potential in electrochemical systems. Polarization modulation at a fixed electrode potential is an important technique in that it differentiates between adsorbed and dissolved species, but the sensitivity is relatively low. Moreover, This separation of surface and bulk phase is usually incomplete since the p/s interfacial demarcation extends the order of a wavelength from the metal surface. A distance comparable to the thickness of the electrochemical thin layers (ca. 10^{-3} mm) usually is employed. It may be noted that some authors use the acronym IRAS to denote specifically the p/s modulation technique.

Methods for Surface Raman Measurement.[22–24] Three methods have been used for surface Raman spectroscopic studies, the 90° geometry, the backscattering geometry and the ATR configuration. In the first geometry, a focusing lens is used to focus the laser beam at a spot on the surface, the incident light making an angle of 90° with the optic axis of the monochromator. In this configuration, the electrochemical cell must have an optically flat cell bottom through which the excitation light enters and an optically flat side through which the scattered light exits. In the backscattering geometry, only one side of the electrochemical cell must have an optically flat window. In the configuration, the focusing lens may be a cylindrical lens which produces a line image for fitting the slits of the Raman spectrometer. An angle of incidence of around 45° can be used between the incident laser light and the optic axis of the monochromator. This geometry has some advantages over the 90° geometry when the reflected light does not come near the entrance slit of the monochromator.

Attenuated Total Reflection Configuration for Raman Measurement. SERS of adsorbates at a metal surface is mostly restricted to roughened surfaces showing a larger surface enhancement. The application of the ATR method in Raman spectroscopy provides a novel and essential way to study mirror-like polished metals,[24] in particular single crystalline surfaces without additional mild-roughening procedure. The ATR method is used to excite surface plasmon polaritons (SPPs) at the smooth metal surface to improve the sensitivity in Raman spectroscopy by electromagnetic field enhancement. The enhancement of the ATR configuration with respect to the normal external reflection geometry ranges from one to three orders, depending on the electrodes and their crystallographic orientations.

Figure 11 shows the ATR configuration under electrochemical working condition. The Weierstrass prism made of ZnSe (radius $r = 5$ mm, refractive index $n = 2.6$ at 647 nm excitation) which is longer by r/n than a hemispherical prism is mounted in the bottom of the electrochemical cell. The experimental procedure is much more difficult than that of normal SERS measurement. The single crystal electrode is carefully pressed against the prism with a micrometer and a spring, so that the optimum gap size d for the SPP excitation (about half a wavelength) can be adjusted. By moving the p-polarized laser beam with a small prism perpendicular to the optical axis, the angle of incidence α with respect to the surface normal and therefore the momentum of the incident photon can be controlled. The laser beam is reflected at

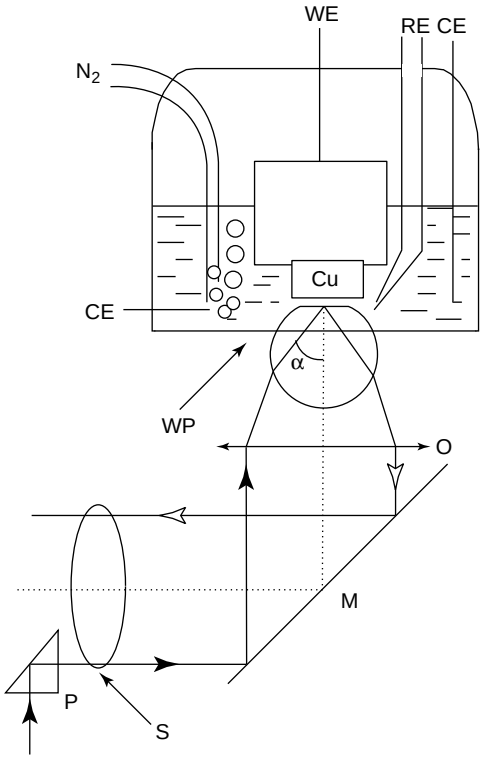

Figure 11 Schematic diagram of the electrochemical ATR cell for studies with single crystal surfaces: WE, working electrode; RE, reference electrode; CE, counter electrode; WP, Weierstrass prism; O, collection objective; M, mirror; P, prism; S, scattered light. (A. Bruckbauer, A. Otto, 'Raman Spectroscopy of Pyridine Adsorbed on Single Crystal Copper Electrodes', *J. Raman. Spectrosc.*, **29**, 665–672, Copyright (1998) John Wiley & Sons Ltd. Reproduced with permission.)

the prism base and the intensity of the reflected beam can be measured by a photo diode. When the angle of incidence and the gap size are close to the optimum conditions for SPP excitation, a bright cone of scattered light is emitted and appears as a sharp ring behind the 1 : 1 collection objective. The intensity of the reflected light goes to zero and the Raman intensity has a maximum under optimum conditions, so that a fine adjustment of the experimental parameters is possible. For example, the optimum condition for SPP excitation at $\lambda = 647\,\text{nm}$ was found to be at $\alpha = 34°$. The scattered light is collected by an additional objective which provides a good matching to the monochromator. With the extra enhancement given by the ATR method, it was possible to study the Raman signal of adsorbates at electropolished single crystal electrodes, including Ag, Cu, Pt and Ni as a function of the electrochemical potential.

With the polarization analyzer in place, polarization studies can be made by measuring the scattered light with the analyzer oriented so that only light with its electric vector parallel to the scattering plane is passed. Then the measurement is made with the analyzer rotated exactly 90° so that only plane-polarized light with the electric vector perpendicular to the scattering plane is passed. The scattering plane can be defined as the plane made by the direction of propagation of the incident laser light. In any optical configuration, the angle of incidence can be varied by rotating the electrode orientation with the micropositioner. To facilitate alignment, the cell can be mounted on an x-y-x-ϕ micropositioner which has provisions for translations as well as rotation in the vertical plane that the optic axis of the monochromator makes with the entrance slit.

To improve the detection sensitivity, the main consideration for the collection lens is that it should have as low an f number as feasible in order to collect as much scattered light as possible. Also the ratio of the collection lens radius to the distance between the lens and the entrance slit should be equal to the ratio of the radius of the first optical element in the monochromator to its distance to the entrance slit. This condition will ensure that the scattered light will fill the spectrometer optics, giving optimum sensitivity and resolving power.

3.2.3 Solid–Solid Interfaces[25–27]

Monolayers and multilayer (or polymer) films are ubiquitous in nature, forming at the interface between two phases and affecting the surface properties in a variety of ways. They may be present as a reaction front between two materials, as in corrosion or rusting of metals, or they can be specifically formed to prevent such corrosion from occurring. An understanding of the role of the physical chemistry of the film–metal and polymer–metal interfaces is critical to the control of the final properties of the composites. Molecular level characterization of the solid–solid interfaces by IR and Raman spectroscopy is therefore highly desirable. The IR and Raman study on solid–solid interfaces is much less extensive in comparison with that on the solid–gas and solid–liquid interfaces. As a consequence, the cell and method employed are relatively simple, and will not be described here in detail. However, it should be noted that for probing solid–solid interfaces, at least one solid phase must be thin enough to be transparent to the light beam.

IR spectroscopy is an ideal probe of the structure and orientation in deposited films because of its sensitivity to chemical and physical properties.[25,26] Recent examples in the literature are extremely diverse. When LB and SAM films are deposited upon metal surfaces, the effect of the metal surface selection rule upon various bands in IR RA spectra can be used to determine the average orientation with respect to the surface normal. In fact, the calculation of chain alignment with respect to a surface

is also of interest to protein and polymer science and a number of computational techniques used in these areas may be of relevance to surface thin-film structure. Most of these studies were carried out in an ATR configuration in order to provide sufficiently high sensitivity.

Raman study of film–metal and polymer–metal interfaces usually involves a certain kind of enhancement to prevent strong interference by the solid phase. Signal enhancement can usually be achieved by depositing films onto the surfaces of noble metals such as Ag and Au, to make use of the SERS effect.[27] For materials with electronic transitions in the visible, the resonance Raman effect can be utilized to enhance certain bands. Furthermore, surface-enhanced resonance Raman spectroscopy of various dye layers has been reported, and the polarized Raman scattering has been proved to be useful in determining the orientation of the molecular orientation of the molecules in films. However, direct measurements of the interfacial structure have long been a difficulty. Polymers adsorbed from a solution usually form multilayers on a solid surface. A ''very thin layer'' of polymer formed on a metal is usually at least tens or hundreds of angstroms thick. Thus, the film of polymer on metal must be a few hundreds of angstroms thick in order to cover the surface entirely while the interfacial region relates only to a few or a few tens of angstroms of polymer adjacent to the metal surface. It is therefore difficult to measure the interfacial region without the interference of the polymer bulk. SERS is ideal for the study of polymer–metal interfaces since it enhances almost exclusively the first monolayer of molecules adjacent to the surface. That makes it possible to examine the interfacial region between metal and polymer as long as the polymer films are not so thick that normal Raman scattering is stronger than that from the interface. Another advantage of SERS for polymer studies is its fluorescence-quenching property. Normal Raman spectra cannot be obtained from many samples because of the curtain of fluorescence that hides them. In the visible excitation region, most polymers are not amenable to Raman analysis because of fluorescence. Adsorption of molecules on the SERS-active metal surface causes fluorescence quenching in highly fluorescent compounds. Fourier transform Raman spectroscopy is considered as a proper technique for studying such specific interfaces.[27]

3.2.4 Solid–Ultra High Vacuum Interfaces[28]

In spite of the fact that increasing emphasis in recent years has been placed on in situ vibrational spectroscopy in studying various interfaces under working conditions, as has been stated above, ex situ methods of surface characterization under UHV conditions are still of substantial fundamental and practical value. As can be seen in this Encyclopedia, the highly sensitive UHV-based EELS technique can yield detailed surface vibrational data as an ex situ method. Similar remarks also apply to surface IR and surface Raman spectroscopy, although these techniques have been used even more rarely in this way.

Ex situ IR and Raman spectroscopies are considered as complementary techniques to EELS, particularly for studying some complex interfaces. Taking the investigation on an electrochemical interface as an example, a general limitation of the ex situ method is that the ambient-temperature transfer into UHV usually removes solvent and other volatile interfacial components, so that the double-layer structure will be significantly or even drastically altered. An alternative UHV-based approach, pioneered by Sass in 1988, entails forming electrochemically relevant interfaces by sequential gas-phase dosing onto a clean metal surface of some or all of the various components (solvent, solute, electrolyte ions, etc.) which collectively form a double layer. This strategy is often known as ''UHV double-layer modeling''. It can yield unique insight into the nature of double-layer interactions, and also provides an integrated stepwise link between the multi-component electrochemical interfaces and constituent metal–UHV systems. An additional but important advantage of this approach is to overcome the problem of strong interference due to the bulk solution phase for surface IR spectroscopy. It should be noted that low temperatures (typically $<150\,K$) are required at the solid–UHV interfaces to prevent evaporation of volatile components, such an interfacial structure involving water as the solvent is very different from that of solid–electrolyte interfaces. However, the fundamental nature of ion–solvent and other interactions should be insensitive to temperature. This unique approach has the obvious virtue of enabling UHV-based techniques such as EELS to be exploited and combined with IR and Raman techniques for interfacial characterization. Vibrational data obtained along these lines are considered as a link between the structural properties of in situ electrochemistry or catalysis and metal–UHV interfaces. Weaver and Zou[28] have carried out systematic work on this important approach, and have published a review recently.

3.3 Newly Developing Methods

In the 1990s, especially in the last 5 years, several important techniques of surface IR and Raman spectroscopies have been developing very rapidly, including IR and Raman microscopies, time-resolved IR and Raman spectroscopies, UV-resonance Raman spectroscopy, and scanning near-field optical microscopy (SNOM). These techniques have distinct advantages individually either in high detection sensitivity, or high spatial-resolution, or

high time-resolution. Combination of these techniques to maximize the advantages will certainly make them increasingly important and become key tools in surface analysis in the 21st century. Therefore, we include these newly developing techniques in this section.

3.3.1 Single Molecule Spectroscopy[29,30]

Single molecule imaging and spectroscopy is of basic scientific and practical interest and has been the subject of considerable effort over the past few years. SERS is a phenomenon resulting in strongly increased Raman signals of molecules that have been attached to nanometer-sized metallic structure. The enormous surface enhancement must be utilized to make single molecule Raman spectroscopy a reality. Recently, two groups have achieved this goal using different approaches. Kneipp et al.[29] detected successfully a single molecule of crystal violet adsorbed on aggregated silver colloids. They used near-IR excitation which is not resonant with any intramolecular optical transitions of the dye but efficiently excites the plasmons of the aggregated colloids. On the other hand, Nie and Emory[30] combined SERS effect and normal resonance effect to produce the required sensitivity to detect a dye molecule adsorbed on the surface of a single silver particle.

Nie et al. used standard citrate reduction techniques to produce silver sols which comprised a heterogeneous collection of mostly unaggregated particles of various sizes and shapes. Surprisingly, a small number of these particles showed extraordinarily high enhancements, that are labeled ''hot'' particles. These special particles can be imaged by bandpass filtering the Raman scattering of adsorbed Rhodamine 6G. Combined optical and atomic force microscopy (AFM) measurements showed that most of these particles were isolated with dimensions of ca. 80–100 nm and have shapes which ranged from spherical to rod-like. This result is in general agreement with the electromagnetic theory of SERS that the surface plasmon resonance frequency is a function of particle size. However, the large enhancement factors observed cannot be accounted for solely by electromagnetic theory.[7,9] Figure 12 shows some particularly interesting spectra. It indicates that much more intense scattering was observed when the laser polarization was aligned along the long axis of an ellipsoidal particle which is the most polarizable direction. This result is in contrast to what is commonly recognized – that SERS is depolarized. Further, in a series of Raman spectra taken from a single molecule at single particles at different times, fluctuations both in frequency and intensity were observed. These results clearly illustrate the high potential of SERS among single molecule detection methods, not only for its tremendously high sensitivity, but also its capability to characterize the molecular structure, orientation and dynamics.

3.3.2 Infrared and Raman Microscopy and Spatially Resolved Measurements[31,32]

IR microscopy and Raman microscopy are hybrids of optical microscopy and IR spectroscopy or Raman spectroscopy and, as a consequence, have all the concomitant advantages of both techniques. The surface signal is analyzed by an optical microscope coupled to the IR spectrometer or Raman spectrometer. The vibrational frequency shift is analyzed at different points of the surface. IR microscopy and Raman microscopy can resolve parts with different chemical composition in a surface and are sometimes referred to as chemical imaging. They have been applied to the study of inhomogeneous surfaces, thin films, coatings, microelectronic integrated circuits, biological tissues, and others. Their advantages and some technical details related to surface analysis are discussed.

3.3.2.1 Infrared and Raman Microscopy For increasing the spatial resolution and detection sensitivity in surface analysis, two important characteristics of a microscope lens are its magnification and its NA. Magnification is simply the ratio of the image size to the object size. NA is given by $NA = n \sin\theta$. The magnification is important only to the visual operation of the microscope, since it does not affect energy throughput. However, the NA is directly proportional to the energy throughput and so should be as high as practical. Typical NAs of IR microscopic lenses are 0.5–0.7. In IR microscopy the source of radiation is the normal source of the IR spectrometer, namely a heated body, often just a nichrome resistance wire. All IR microscopes use Cassegrains as objective lenses, but there is variation between microscopes as to whether off-axis or on-axis elements are used for other purposes. The reflection mode is feasible with the microscope by the utilization of specialized accessories. A grazing angle objective and an ATR objective that utilizes various high-density hemispheres (ZnS, KRS-5, germanium, diamond, etc.) are also commercially available and simply replace the normal Cassegrain lens in the microscope for such studies, which could be valuable for analyzing some surfaces.

For a Raman microscope, the main purposes of the microscope are to excite and collect Raman radiation very efficiently from the sample and to provide a means for sample positioning and viewing at high magnification. Raman scattered radiation is collected 180° to the incident excitation by the same objective and transmitted back through the beam splitter. In most Raman microscopes holographic notch filters are employed as the beam splitter of the microscope, and the means by which Rayleigh rejection is achieved. This achievement allows the use of a single short focal length monochromator (0.25 m). However, for extremely low wavenumber

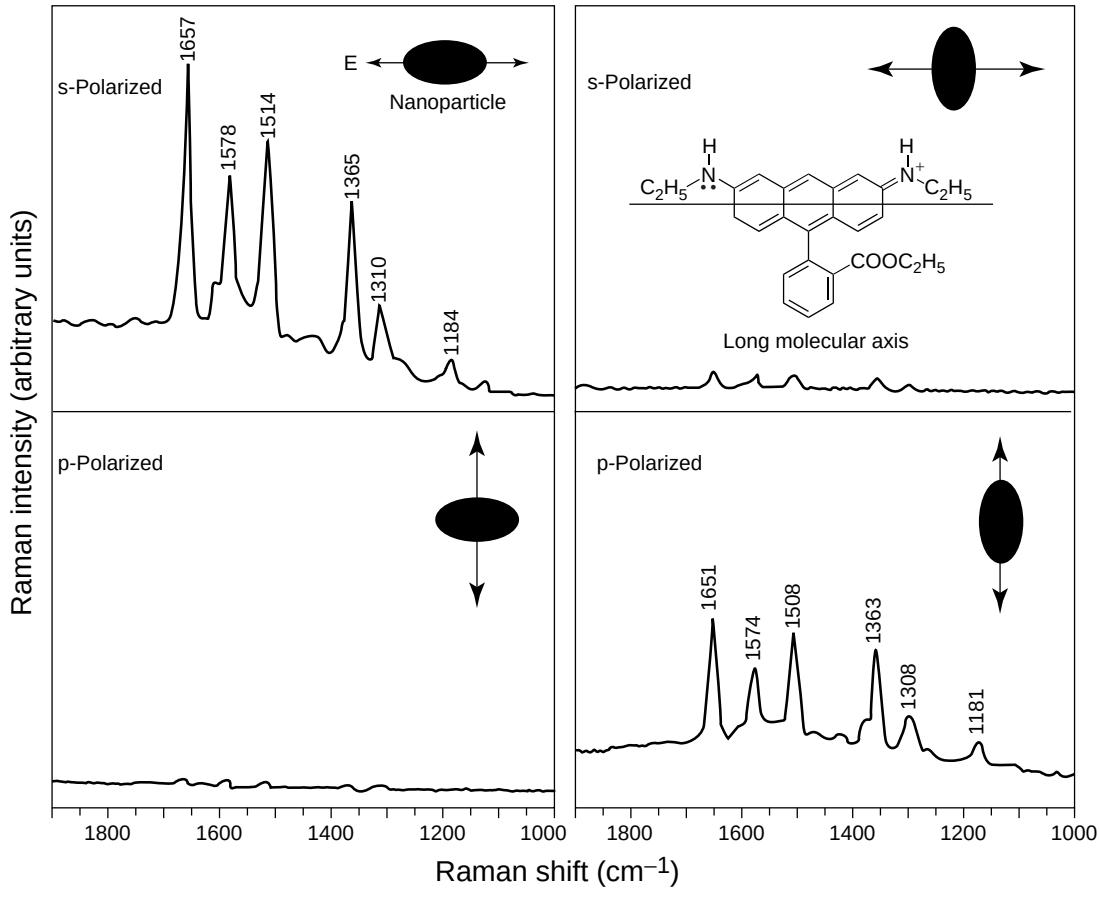

Figure 12 Surface-enhanced Raman (SER) spectra of R6G obtained with a linearly polarized confocal laser beam from two Ag nanoparticles. The R6G concentration was 2×10^{-11} M, corresponding to an average of 0.1 analyte molecule per particle. The direction of laser polarization and the expected particle orientation are shown schematically for each spectrum. Laser wavelength, 514.5 nm; laser power, 250 nW; laser focal radius, ~250 nm; integration time, 30 s. (Reproduced by permission from S. Nie, S.R. Emory, 'Probing Single Molecules and Single Nanoparticles by Surface-enhanced Raman Scattering', *Science*, **275**, 1102–1106, Copyright (1997) American Association for the Advancement of Science.)

shift information (i.e. 2 to 20 cm^{-1} shift), a double monochromator may be required. Additional gains have been achieved by employing the notch filter as a beam splitter. The filter reflects the excitation wavelength with an efficiency of better than 90% and transmits the Raman scattered radiation with a similar efficiency. No compromise is made between excitation and collection of the Raman scattered light. The overall benefit is that the size and cost of Raman microprobes have been reduced considerably and the surface sensitivity increased remarkably. The latter is extremely important to surface (interface) studies. The intensity of Raman scattering is proportional to the irradiance (power/area of cross-section) of the exciting beam. The Raman intensity depends critically on the solid angle Ω over which the scattered light is collected.

3.3.2.2 Raman and Infrared Imaging Techniques[33]
There are two types of sampling technique in Raman and IR imaging, global imaging and point-by-point mapping. One unique feature of the Raman or IR microscope is the ability to obtain a microscopic image of a sample through the use of discrete wavelength segments of the spectrum of Raman or IR signal originating from a light-irradiated sample. The system utilizes the multiplex advantage of a two-dimensional low light level television-type electronic camera detector. The principle involved is that the sample is placed under the microscope and the light beam illuminates the whole field of view of the microscope. From Raman imaging, this was achieved by rotating the laser beam to form an annular beam of light which is directed to the condenser of a dark-field microscope objective. Thus the sample is uniformly irradiated with monochromatic radiation. The microscope then transfers the image of the sample to a monochromator which selects the desired wavelength to be imaged onto a television camera placed at the exit focal plane. If the wavelength selected corresponds to a Raman band frequency of a component of the

surface, the micrographic image indicates the distribution of the component throughout the illuminated area. The spatial resolution of the system is about 1 μm. This system works well for surfaces which are good Raman scatterers but tends not to work with poor Raman scatterers and could not easily discriminate Raman features from fluorescent backgrounds. The big advantage of this method of Raman imaging is that the detector observes the entire field simultaneously.

Another approach is to use a computer-controlled stepping motor driven sample stage on the microscope relative to the stationary focused laser beam. In this procedure the spectra are detected at several hundred positions on a sample surface. A computer is programmed to record a set of spectra point by point. This procedure is repeated until all of the necessary data are recorded and stored from the area studied. In this way either one-dimensional mapping line profiles or two-dimensional maps of the species present on the surface are built up.

A different approach is to scan the focused laser beam over the stationary sample and record the spectrum at different locations on the surface. This method is achieved by a new kind of transfer optics placed between the microscope and the spectrometer, which enables an optimized coupling. The coupling optics consists of a pair of lenses. One lens, optically coupled to the back aperture of the objective can be moved in two orthogonal directions perpendicular to the laser beam. Thus this lens can focus the light beam on any point in the microscope field and therefore on any point of the sample. The second lens is mechanically coupled to the first and placed in the scattered beam to balance any shift in the image probed area and focus the scattered beam on the entrance slit of the spectrometer. The optical arrangement has proved to be convenient for a variety of cumbersome surface samples, such as variable temperature or pressure cells.

3.3.3 Time-resolved Spectroscopic Measurements

IR and Raman spectroscopy are in principle amenable to the dynamic study. The conceptually most straightforward extensions into this field are time-resolved investigations. Here, changes in molecular structure and interaction at the surface initiated by some external action at a given time are monitored. This is surely an important advantage over other vibrational spectroscopies including EELS and neutron scattering techniques. During the last 15 years several technical developments in the light source and detector made time-resolved IR spectroscopy and Raman spectroscopy an important branch of time-resolved surface spectroscopy. It should be emphasized that there are two kinds of time-resolved studies: (i) the measurement is triggered by a certain surface process (reaction). By applying a potential pulse to the electrode or injecting reactive gas to the catalyst, one can monitor the whole dynamic process from both spectral and temporal aspects; (ii) the measurement is performed at steady state of the surface process that can be disturbed by a pump light beam and probed by another light beam. The first method is relatively simple, but has attracted more attention as it can obtain a lot of information, such as the reaction intermediate and pathway as well as reconstruction (or phase transition) of the surface adlayer. The second method requires expensive pulse lasers, and is suitable to probe surface charge transfer processes and characterize vibrational energy transfer rates for surface species.

3.3.3.1 Time-resolved Infrared Study As in all kinds of time-resolved spectroscopy, the principal quantity to be measured is the spectral and temporal dependence of the radiation intensity (here IR radiation). Femtosecond- to millisecond-scale time-resolved surface IR studies have been achieved. They involve the investigation of energy transfer at the surface (femtosecond),[34] the photochemical process (picosecond), the structure and dynamics of the excited states (nanosecond), structural change during temperature, pressure or electric stimulation (microsecond), and the chemical or electrochemical reaction (microsecond to millisecond even down to seconds). The instrumentation is totally different to provide a time resolution higher or lower than 1 ns.

When the time-resolution is higher than 1 ns, because of the very fast signals involved, it is no longer possible to record electronically the complete time course of the intensity change. Instead, the so-called pump-probe technique is employed. The essence of this technique is that the system is perturbed from equilibrium by a very short laser pulse. The system's response to the perturbation as well as the decay back to equilibrium are monitored by a second, time-delayed laser pulse. By repeating the experiments with different delays, the time course of the intensity change of the monitoring pulse caused by the exciting pulse is obtained. Evidently, this scheme allows for the direct observation of the dynamic behavior of the system. To obtain a precise time correlation between the exciting and monitoring pulses, both are derived from the same master laser pulse. IR pulses are generated by nonlinear methods from visible pulses with a pulse-width well below 1 ps using present laser technology. In vibrational spectroscopy, which usually requires a considerably higher spectral resolution than UV/visible spectroscopy, one has to take into account the time-energy uncertainty principle. A transform-limited pulse with a duration of 0.1 ps exhibits a spectral bandwidth of about 27 cm^{-1}.

It should be noted that most nanoscale or picoscale analyses have been carried out at static condition

(e.g. potential, temperature or pressure) after ultrafast perturbation. To bridge the kinetic and structural analyses and to obtain a deeper insight into the dynamic processes occurring in the interface, time-resolved surface analyses are required to be carried out at the same timescale (microseconds to milliseconds), i.e. coupled with the surface dynamic process.[35,36] There are two kinds of technique to perform time-resolved IR study with time resolution lower than nanoscale, the continuous-scan technique and the step-scan technique. The IR intensity changes caused by the external distortion can now be measured with a fast detector and the corresponding signal-acquisition electronics. The continuous scan has a varying time resolution from nanoseconds to tens of milliseconds in different sources (tunable laser or broad-band IR source), and acquisition mode (rapid scan and stroboscopic) with different triggering methods (zero-crossings of the helium–neon laser interferogram, asynchronous or fast digitalization triggering).

The step-scan technique is a newly developed technique very useful for time-resolved studies. The step-by-step optical retardation change is achieved by a combination of a constant motion of the moving mirror and a sawtooth-like motion of the fixed mirror for the FTIR spectrometer. The minimum time resolution is limited by the analog-to-digital converter which is about 5 μs. Step scan time-resolved spectroscopy can be used to collect spectral and kinetic information on repeatable processes that are typically initiated by electronic pulse, laser pulse, temperature jump or rapid mixing.

Although time-resolved IR spectroscopy provides dynamic and kinetic information on surface sorption and reactions, the interpretation of the spectra is not easy if IR bands of reactants, intermediates and final products are highly overlapping. The two-dimensional IR technique could be used to differentiate overlapping bands and to highlight dynamic and kinetic information involved in the time-resolved spectra. In two-dimensional IR, a spectrum defined by two independent wavenumbers is generated by a cross-correlation analysis of dynamic fluctuations of IR signals induced by a sample perturbation (e.g. potential modulation). The cross-correlation analysis includes synchronous and asynchronous correlations which characterize the coherence of dynamic and the independent fluctuations of IR signals, respectively. The sign of the synchronous correlations indicates the relative direction of change in dynamic spectral response, while the asynchronous correlation provides the temporal relationship between the intensity fluctuations at different wavenumbers.[13]

3.3.3.2 Time-resolved Raman Studies Raman spectroscopy can also provide quite high time resolution, which affords additional information about various surface (interface) processes. However, due to the intrinsic low sensitivity it provides, conventional surface Raman spectroscopy is not considered suitable for time-resolved investigations for timescales shorter than several tens of seconds. With the application of SERS effect or resonance Raman effect, high time resolution reaching the picosecond scale with reasonable signal-to-noise ratio has been obtained.

The detailed operation of time-resolved Raman measurement depends on what kind of study is being performed. It can be divided into three categories: (i) using a CW laser, time resolution depends on the detector; (ii) using a pulse laser, the time resolution depends on the pulse width of the laser; (iii) combining with electrochemical transient technique, the time resolution depends neither on the detector nor the laser but on the electrochemical set-up (see section 3.3.3.3).

For those using CW laser, several kinds of detectors have been employed. For the tunable monochromatic technique a single channel detection system, such as a photon multiplier tube, is used. The time-resolved Raman investigation can only be done with the grating set at the desired peak position (frequency). The spectral band of the surface species is in general quite broad and the entrance and exit slits are tuned at a wider width for detecting the signal of whole band. The signal is collected after the external signal pulse in a standstill mode (also called the intensity–time curve). Provided there is sufficient signal intensity, the time-resolved value depends on the response rate of the detector.

For the multichannel detection system normally two kinds of detectors are used; the intensified charge coupled device (ICCD) detector and the intensified photodiode array (PDA), which have a very fast response rate. Without a gating system, the best time resolution provided is only about tens of milliseconds. In this case, the external perturbation (external temperature, pressure, or electrode potential change) is applied to the sample, triggering the Raman spectra to be recorded successively, giving a series of time-dependent spectra.[37] With the use of the gating system, as low as 1 ps is possible. However, in order to synchronize the process studied and the detection system, the pump and probe technique has to be employed to acquire the signal of the process occurring at a fixed potential after perturbation by a light or magnetic field. The time resolution value is determined by the signal and the response rate of the detector. The higher the time resolution value anticipated, the stronger the Raman signal should be. Thus, in the nanosecond resolved Raman study, a molecule with very strong resonance and/or surface enhanced Raman effect is adopted in order to provide a strong enough signal. For example, Shi et al.[38]

using the gating technique, performed a nanosecond time-resolved Raman study on the photochemical process of flavin. For the study using a pulse laser, the detector needs to be set at the gatable mode. After the desired time delay of external perturbation to the system, the laser pulses were delivered on the sample and triggered simultaneously the detector so that the Raman signal was collected. Normally repeatedly acquiring spectra is anticipated in order to provide sufficient intensity. However, it is necessary to point out that for time-resolved IR and Raman spectroscopic studies on surface kinetics or dynamics, the repeatedly acquiring method can only be applied to those surface changes possessing good reversibility.

3.3.3.3 Potential-averaged Raman Spectroscopy[39]

Because the time resolving value is generally limited by the detector, as mentioned above, it has to be equipped with the costly gatable ICCD for the time-resolved Raman study. For those who do not have pulse lasers and fast response rate detectors, a unique method named potential-averaged Raman spectroscopy (PARS) is applicable in spectroelectrochemical studies. The PARS method is based on the application of a CW potential modulation to the electrode during the spectral acquisition period. A PARS spectrum is recorded while the two electrode potentials of interest are being rapidly modulated, and the spectrum contains the sum of the information of the surface species at the two modulated potentials. Later, each spectrum at individual potential can be extracted by deconvolution. By properly choosing a series of modulating frequencies in sequence (e.g. ranging from 100 kHz to 1 Hz), one can obtain a set of time-resolved (ranging from 5 μs to 0.5 s) spectra of surface species at the two electrode potentials modulated. The time-resolving value is no longer dependent on the detector but only on the electrochemical experimental parameters. To reduce the charging time of the electrochemical double layer, ultramicroelectrodes are required in order to have sub-microsecond resolution with a simple nongatable charge coupled device detector. Therefore, taking advantage of easily switching the electrode potential, it is feasible to use PARS to obtain electrochemical dynamic information on a microsecond scale with prolonged acquisition time (e.g. 60 s) in order to obtain time-resolved spectra of better signal-to-noise ratio. This technique cannot be used to study the surface process with irreversible change between the two electrode potentials modulated.

3.3.4 Ultraviolet Resonance Raman Spectroscopy[40]

UV Raman spectroscopy has recently proved to be a very powerful technique for catalyst characterization. Generally, the application of Raman spectroscopy to catalyst characterization has been limited (i) by the small Raman scattering cross-sections typical for excitation at the visible wavelengths available from commercial, CW lasers and (ii) by interference from sample fluorescence and luminescence, which often produce a huge background compared with the weak Raman bands. Typical catalyst surfaces are often covered by carbonaceous residues (coke) during reaction or deposition. These samples often exhibit strong fluorescence, which makes detection of the weak Raman signals impossible. Therefore, both an increase in the intensity of Raman signals and a decrease in the intensity of background fluorescence would make a substantial advancement in the application of Raman spectroscopy to practical material characterization. This has recently been achieved by the application of UV resonance Raman spectroscopy. Using UV excitation, the variety of samples where useful Raman spectra can be obtained has been dramatically increased. For example, good-quality Raman spectra that are nearly free of fluorescence interference have been obtained from deeply colored, heavily coked catalysts and chemical vapor deposited diamond film. Moreover, since the Raman scattering cross-section is generally larger using UV excitation compared to visible or near-IR excitation, the sensitivity of UV Raman spectroscopy is correspondingly greater. In addition, the intensity of Raman bands can be greatly enhanced owing to the resonance Raman effect when the excitation line is accessible to the electronic absorption of the surface or surface species. As a consequence, it is possible to obtain spectra of supported oxide catalysts at loadings below the detection limit of conventional Raman spectroscopy.

3.3.5 Surface-enhanced Resonance Raman Spectroscopy[7,41]

At present, this technique seems not to be so promising as those mentioned above. However, along with the rapid development in laser technology, wavelength tunable lasers ranging from near-IR to UV will be adopted in most laboratories in future so that surface resonance Raman spectroscopy together with surface-enhanced resonance Raman spectroscopy will be probably developed into a powerful tool. A molecule adsorbed at a surface or within the diffusion layer, when excited by light within its absorption band, has enough scattering intensity even at very low concentration to give a resonance Raman spectrum. This type of spectrum will be potential dependent if the Raman resonance-active species is generated by the surface process (reaction).[7] With the resonance Raman effect, a resonance Raman spectrum of good quality can be obtained on smooth non-SERS-active substrates, even atomically smooth single crystal surfaces.

If the resonance Raman scattering can also be enhanced by a SERS-active metal concurrently, i.e. by a surface-enhanced resonance Raman scattering (SERRS) process, SERRS spectroscopy has extremely high sensitivity with an enhancement factor larger than 10^{10}. It should be stressed that due to the participation of the metal, for the molecules near the surface, the resonance Raman process is damped by energy transfer to the metal, with a more severe damping of the fluorescence at the surface. It is possible to obtain SERRS in the absence of an interfering fluorescence background. The combination of advantages of SERS and SERRS effect will make this technique one of the most sensitive methods in surface science, which could be valuable for trace analysis and even for single molecular detection.[30]

3.4 Troubleshooting

3.4.1 Detection Sensitivity

As the number of surface species under study is usually extremely small, ten times less than a monolayer (a few tenths of a nanomole in quantity), the most serious problem, especially for studying the interface of two condensed phases, is the detection sensitivity. Thus it is vitally important to optimize the experimental set-up from every aspect. It has been found that most trouble is caused by optical alignment related to the cell and sample.

3.4.1.1 Raman Measurements of Solid–Liquid Interfaces For a microprobe Raman system, the solution layer of the electrochemical system could affect dramatically the collection efficiency of the microscope and the optical alignment. An investigation on the effect of the thickness of the solution layer on the Raman signal detected revealed that decreasing the solution layer thickness from about 2 mm, 1 mm, 0.5 mm to 0.2 mm, the Raman intensity increases 35%, 52%, 71% and 89% of the maximum intensity obtained on a silicon wafer without a water overlayer. At a thickness of 0.2 mm, the Raman intensity suffers only about 10% loss and the solution layer with this thickness will not affect the electrochemical process much. This thickness was used as the optimized condition after considering both the Raman intensity and the electrochemical processes. Furthermore, in the electrochemical system, most electrolytes are corrosive, so to protect the objective a cover glass or quartz window has to be employed between the electrolyte and objective; this results in 50% loss of the signal. An alternative way to protect the objective is to wrap the objective with a very thin poly(vinyl chloride) film (benefit from confocal set-up: no signal from poly(vinyl chloride) could be detected in this configuration). Using this approach, the Raman signal only suffers a 20% loss. The disadvantage of this approach is that without the quartz window, the solution has to be exposed to the air, and impurities in air will contaminate the solution studied. For a system very sensitive to impurities, the thin quartz window has to be employed at the cost of intensity.

The water overlayer in an electrochemical system greatly affects the vertical spatial resolution of the confocal system. For example, with smaller pinhole size (300 μm), the spatial resolution is about 17 μm and 19 μm without and with solution layer, while when the size is about 600 μm, the spatial resolution amounts to 21 and 53 μm for the surface without and with solution layer, respectively. This indicates that in an experiment high spatial resolution is desired, decreasing the pinhole size is a very effective way to eliminate the distortion of the solution layer to the optical path.

3.4.1.2 Electrochemical Infrared Absorption Spectroscopy Measurements In EC-IRAS experiments, crucial to the success of this technique is the formation of a very thin solution layer between the IR window and the surface by pressing the electrode against the window carefully and firmly. This approach reduces the background absorption due to the bulk solution which is very large in the case of aqueous solutions. To position the electrode surface exactly parallel to the window, it is better to use a spring as a ''soft'' connection of the electrode to the holder so that the optimum position can be adjusted. It should be also pointed out that the higher the reflectivity of the surface for reflectance IR, the higher the signal-to-noise ratio could be. Hence the electrode surface must be polished very carefully.

Additional problems may be caused by the thin-layer geometry for EC-IRAS measurement, in particular time-resolved measurement. Mass transport in and out of the thin layer is strongly restricted, resulting in the change in the interface including the local pH and concentrations. In addition, the time constant of the spectroelectrochemical cell is very large due to the high resistance. Reactions cannot quickly respond to the externally applied potential changes. Furthermore, reactions occur first at the edge of the electrode and then gradually reach the center of the electrode. Therefore, time-resolved IR monitoring often becomes meaningless except for very slow reactions. To unravel this problem, the Kretschmann ATR configuration could be used, so that the IR radiation is passing through the ATR prism to irradiate on the backside of the ultrathin metal film surface evaporated on the prism. Since the evanescent wave that penetrates into the solution phase is damped by the metal films, the solution background is reduced significantly, so the solution phase does not need to be thin.[13]

Alternatively, if the system studied is not restricted to single crystalline surfaces, one may consider changing the optical geometry from specular reflectance mode to diffuse reflectance so as to use rough surfaces. The increase of the surface area will inevitably increase the concentration of the surface adsorbed species. For certain kinds of surface, such as some noble and transition metal surfaces, this increasing of the surface area sometimes means the emergence of surface enhancement (SERS or SEIRAS). It is a quite common strategy that in the study of catalysis and solid–liquid interfaces, people find ways to increase the surface area in order to improve the surface sensitivity.

3.4.2 Surface Heating and Damage

Laser-induced sample damage is probably the second largest cause of unsuccessful sample analysis in surface Raman spectroscopy, especially for Raman microscopy. For a conventional Raman spectroscopic measurement, a 100 mW laser beam with a cross-sectional area of 0.785 mm^2 (diameter 1 mm) has power per unit area of 127 $mW\,mm^{-2}$. For a Raman microscope, the objective characteristically yields a 1-μm diameter beam waist at the sample focus. The power density at the focus is then in the order of $127 \times 10^6\,mW\,mm^{-2}$. This power confinement represents a million-fold increase in the power per unit area. The extent of these problems can range from subtle color or crystalline changes in the materials, to the complete ablation of the sample beneath the microscope objective. The heating problem is the main reason that high time-resolved Raman spectroscopy with a timescale of femtoseconds has not been applied to surface analysis because the extremely high power density could easily burn the sample surface. Even for the conventional Raman measurement, the heating may cause corresponding spectral changes, including subtle shifts in Raman transitions, the appearance of new transitions, or complete loss of signals. In order to prevent these possibilities from occurring, the following plan of action is suggested.

Prior to collection of a spectrum, a photograph or mental note of the sample's appearance should be made. This photograph or note can then be compared to the appearance of the sample after a spectrum has been collected. If any changes have occurred, the sample will appear different. More subtle heating effects are best detected by collecting several spectra at different scan rates and/or power levels. If these spectra are not reproducible, the sample may be changing as a result of irradiation. If care is not taken during the analysis, the results can be easily misinterpreted. On some Raman microprobe systems it is possible to view the sample while it is being illuminated with the laser. For these systems one can view the laser focus and determine whether it changes as a function of time. Should material be ablated from the sample, the size of the focused laser will appear large, since the remaining sample surface is below the microscope's focus. Due to the fact that the objective is directly above the sample and in close proximity to the sample, any sample that is vaporized will condense on the cell window or the microscope objective. This material must be removed, since it will affect the focal characteristics of the laser and could yield a parasitic background in the Raman spectrum.

In the event that the sample undergoes degradation even at the lowest power levels, one of the more obvious solutions is to reduce the laser power still further. This solution can be implemented by defocusing the laser beam. Defocusing the laser is usually accomplished by adjusting the beam expander in the prefiltering optics of the system so that the objective's aperture is under filled, so a large beam waist is obtained at the objective's focus. If lowering the power of the laser decreases the sensitivity too much, an alternative way is to make a cell that can be moved horizontally or vertically using mechanical or pneumatic apparatus to minimize the thermal effects and damage very efficiently.

3.4.3 Fluorescence Elimination[40]

One of the major problems associated with surface Raman spectroscopy is the elimination of parasitic fluorescence. Raman is a scatter-based phenomenon, inherently a two-photon effect. Fluorescence, on the other hand, consists of two sequential photon-induced transitions, and absorption followed by spontaneous emission. The timescale of scattering is on the order of 10^{-14} s, while fluorescence lifetimes are typically on the order of 10^{-8} s. Thus, the problem can be tackled using temporal fluorescence rejection techniques with a fast response rate detector. Fluorescence quenching is also a technique in which the sample is irradiated for a long period of time. After that, the fluorescence signal has been eliminated or reduced to a tolerable level. The use of an aperture at a remote image plane (confocal microspectroscopy) has also been employed in the elimination of fluorescence from near-neighbor locations. Although Fourier transform Raman spectroscopy has relatively low sensitivity, it allows Raman spectra to be recorded of many samples that fluoresce. This is because the samples are excited using near-IR excitation which does not induce sample fluorescence. Moreover, UV-Raman spectroscopy has been used in some situations very efficiently. An excellent example will be given in more detail in section 4.

3.4.4 Problems of Analyzing Spectral Data

In any process of interaction of electromagnetic radiation with matter, it is important to realize not only the advantages of each situation but also the limitations. There are potential traps where band distortions or intensity changes might have been interpreted as chemical or physical changes within the studied systems. In reality, they might have been simply artifacts attributed to various reflection or scattering processes. For instance, the IR radiation beam makes a double-pass through the film, having been reflected by the substrate. It must be remembered that superimposed on the essentially double-pathlength transmission spectrum will be a weaker specular reflectance spectrum which will mostly distort the stronger absorption bands, and in some circumstances cause band inversion. For thinner films the more specialized technique of grazing-incidence RA is instigated.

One often forgotten problem with ATR surface depth profiling is the fact that the relationships for depth profiling are valid only for homogeneous surfaces, that is, when the concentration of the surface species does not change. To overcome this problem, reflection theory for stepwise-stratified surfaces was successfully used to determine concentration changes at various depths. However, a fundamental difference between ATR and transmission spectra is the position of the bands and their intensities. The frequency shifts along with the sensitivity difference between high- and low-wavenumber regions, and spectral distortions near critical angle conditions, are the primary concerns in ATR quantitative analysis.

4 APPLICATIONS

IR spectroscopy and Raman spectroscopy are among the most widely used methods for surface science. Their ability to deliver specific chemical identification, coupled with the wide range of instrumental and sampling methodologies available, has led to their continuing use in both new and old venues in surface science. They are of prime importance in both fundamental research and various industrial fields related to surface oxidation, adhesion, corrosion and catalytic processes, and even open up the possibility of single molecular analysis. Their applications are reviewed well but briefly in the special issue of *Analytical Chemistry* published every two years. It is beyond the scope of this article to cover all applied fields. Only some typical applications in advanced materials, surface processing, catalysis, electrochemistry, corrosion, biology and sensor, which are related to future applicable areas, are given.

4.1 Advanced Materials and Surface Processing

4.1.1 Semiconductor Surface Processing[42–44]

With the rapid development of materials science, IR and Raman spectroscopy has been used extensively to characterize various material surfaces as well as the physical and chemical processes occurring there. For semiconductors, surface structure change, electronic state mixing, nanoscale behavior (quantum well) and photoluminescence related materials have become routine measurements. More importantly, both techniques can be applied to in situ study of the chemical etching process of various semiconductor surfaces, in particular silicon surfaces, in varying etching solutions to determine the surface bonding and structure during or after the etching treatment or post-treatment of etched surfaces. The multiple internal IR reflectance with its high surface sensitivity and spectral resolution is sensitive to the atomic bonding configuration of the hydrogen in the vicinity of the surface, and permits an easy identification of bonding configurations of hydrogen on silicon surfaces.[42] While the backscattering Raman measurement can be performed in a relatively convenient way, it enables the scrutinization of the etching mechanism of silicon surfaces.[43] To increase the sensitivity, SERS studies of semiconductor ultrathin films (several nanometers) coated on SERS-active gold substrate have been reported, allowing investigation into the relationship between structure and physical properties.[44]

4.1.2 Polymer Interfaces[45–47]

Polymer–metal composites, which include engineering materials, adhesives and coatings, are of great interest in both science and technology. The study of functional polymers and other advanced materials provides numerous opportunities for the application of IR and Raman spectroscopy. Many reviews on this area have been given elsewhere.[27,45,46] The rapid-scanning FTIR technique enables in situ time-resolved analyses, hyphenated studies of IR and mass spectroscopy, and on-line measurements to be carried out. It allows for the determination of crystallinity, composition, conformation, degradation and molecular interactions of polymers at surfaces and monitoring electropolymerization processes. For gaining high Raman sensitivity, SERS studies have been applied to determine the molecular structure of polymer–metal, polymer–solution and polymer–air interfaces. The polymers include polyaniline, polypyrrole, polythiophene, poly(vinyl pyridine) and poly(L-histidine), etc. The approach is achieved by covering the SERS-active surface with the thin polymer film or depositing SERS-active particles (islands) over the polymers chemically or electrochemically.

A good complementary SERS and IR study has been performed on a polythiophene film deposited electrochemically onto stainless steel.[47] It displays a very interesting anisotropy property in electric conductivity, and the film parallel to the film surface is more than 10^4 times that across the film thickness. On the basis of the surface selection rules for SERS, the appearance of two new SERS bands is assigned to the molecular arrangement for the thiophene ring parallel to the metal surface for polymerization. According to the IR surface selection rule, the suppression of these modes parallel to the ring in the IR RA spectra strongly supports the same conclusion. The results suggest that the metal surface was coated with a layer of polymer in which the thiophene rings of polymer chains are nearly parallel to the surface.[47]

4.1.3 Self-assembled Monolayers and Langmuir–Blodgett Films[25,26,48]

There are numerous technological applications of thin films, such as electroluminescent devices, molecular devices and molecular sensing elements. Surface IR and Raman spectroscopy is one of the most widely used tools to characterize various organized monolayers and multilayers such as SAMs and LB films bound to or deposited on smooth surfaces. Because vibrational information can be obtained from all locations within the molecule of SAMs and LB films, for example, of self-assembled alkanethiol monolayers, the ordering and defect structure of the alkane chains can be directly investigated by observing G conformations of the carbon backbone.[48] It has been shown that the alkane chain of adsorbed butanethiol is mostly in the all-T conformation, and the alkane chain of the adsorbed octadecanethiol appears totally free of G conformers. In contrast, adsorbed dodecanethiol shows a large number of G defects in the alkane chain. Based on the surface selection rule, these alkanethiols are proposed to be oriented with their alkane chain largely perpendicular to the surface with the S head group bonded to the surface. A chain tilt of ca. 15° with a rotation of ca. 45° about the chain axis is consistent with the surface Raman spectral intensities.[48] The spectra quality obtained by confocal Raman microscopy is excellent. This will extend the study of ultrathin films, even monolayers, without the aid of SERS. Nevertheless, because of the relatively high sensitivity, IR spectroscopy has been used more extensively in this area.[25,26] It is very useful for the study of the air–water interface of monolayer films, including fatty acids, fatty alcohols, fatty esters, octadecanol and fluorinated undecanol. Raman spectroscopy can investigate the monolayers at the air–water interface with the aid of the SERS active colloidal Ag electrochemically grown underneath lipid monolayers. As the imaging measurement is very time-consuming, SERS imaging with a much faster acquisition time can be easily performed on a UV-photopatterned *p*-nitrophenol SAM and a SAM patterned by microcontact printing.

4.1.4 Zeolites and Molecular Sieves[49,50]

As an advanced material, molecular sieves with atoms substituted by transition metal atoms in their framework have been considered as a new class of catalysts showing remarkable activity and selectivity for a number of oxidation reactions. The catalytic property is mainly attributed to the isolated transition metal atoms in the framework of the molecular sieves. Therefore, characterization of the transition metal atoms incorporated in the framework is the most important aspect of the study. However, it is extremely difficult to know how and whether the transition metal atoms are incorporated into the framework of a molecular sieve as special surface active sites *inside* zeolites. A new approach, UV resonance Raman spectroscopy, is used to identify the transition metal atoms in the framework of molecular sieves. This successful work is based on the resonance Raman effect since there are charge-transfer transitions between the framework oxygen and transition metal atoms in the UV region, see Figure 13. The characteristic Raman bands solely associated with the framework transition metal atoms are selectively enhanced so that the transition metal atoms in the framework of a molecular sieve can be definitely determined. For example, Ti atoms in titanium-substituted silicalite-1 (TS-1) were successfully identified using UV resonance Raman spectroscopy excited with

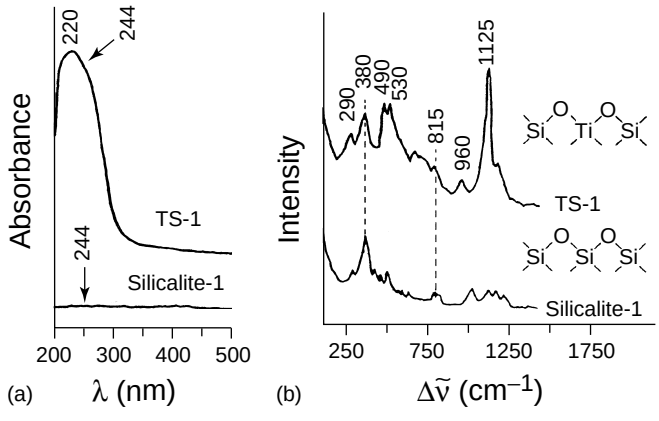

Figure 13 (a) UV/visible diffuse reflectance spectra of TS-1 and silicalite-1. The band in the 200–300 nm range is centered at 220 nm. (b) UV resonance Raman spectra of TS-1 and silicalite-1 with excitation at 244 nm. (Reproduced by permission from C. Li, G. Xiong, Q. Xin, J.K. Liu, P.L. Ying, Z.C. Feng, J. Li, W.B. Yang, Y.Z. Wang, G.R. Wang, X.Y. Liu, M. Lin, X.Q. Wang, E.Z. Min, *Angew. Chem., Int. Ed. Engl.*, **38**(15), 2220–2222, Copyright (1999) Wiley-VCH Verlag.)

a 244 nm laser. The vibrational information of Ti–O–Si bonding is not detectable with a normal Raman system excited by 355 and 488 nm laser excitation (Figure 13). This demonstrated that titanium actually exists in the framework and gives powerful evidence for the predicted structure. Therefore UV resonance Raman spectroscopy has opened up the possibility of identifying the framework transition metal atoms in molecular sieves.[49]

Another important approach, picosecond IR spectroscopy, is used to study the dynamics of zeolitic protons and adsorbates. In the experiment, one specific vibration in the zeolite system is selectively excited with an ultrashort (tunable mid-IR) pulse. The effect of this excitation and the subsequent energy relaxation is monitored in real time, providing information on the structure of the bare zeolite and adsorption complexes. With this technique the picosecond energy flow at the catalytic site and the dynamics of the catalyst–adsorbate interaction can be investigated. Short-lived transient species are observed and the picosecond relaxation rates and pathways at the catalytic site render insights into the dynamics of the interaction between the zeolite and its adsorbate at a molecular level. This illustrates the potential of time-resolved IR spectroscopy in the investigation of zeolite systems.[50]

4.2 Catalysis[51–54]

4.2.1 Adsorption and Reaction on Catalysts

In situ application of IR and Raman spectroscopies to catalytic study benefits from the state-of-the-art design of the in situ cell (or reactor) and the continuous development of instruments and light sources operating from near-IR to the UV region. It offers a number of potential advantages over other vibrational spectroscopy techniques in the characterization of real catalysts since it could obtain information about both surface adsorbed species and the structure of catalysts under working conditions. In particular, Raman spectroscopy has been applied extensively in determining the low frequency vibrations of oxide and oxide-supported catalysts. Furthermore, it is useful in identifying phases that are present in precursor stages of the catalyst during its preparation. Recently these two techniques have been used to characterize the surface species present and its reactivity with other gas or adspecies on catalysts at high temperatures.[51] For instance, syngas reaction reforming of methane on catalysts has been the subject of extensive investigations. Raman scattering has been particularly invaluable in three aspects: determination of catalyst oxidation state, evaluation of the role of the support, and identification of adsorbed intermediates. The time-resolved technique is widely employed to determine catalytic reaction kinetics, and identification of the reaction intermediates.[52] The adsorption of oxygen on the catalyst surface and the interactions of oxygen species (O_2^{2-}, O_2^- and $O_2^{\delta-}$ $(0 < \delta < 1)$) with methane, ethane and ethene from room temperature to high temperatures (up to 1073 K) were studied with the aid of a high sensitivity confocal microprobe Raman system. The high signal-to-noise ratio provided enables the time-dependent study of the oxidation process, thus the adsorbed species and the change in the structure of catalyst can be investigated.[53]

4.2.2 Probe Molecules for Catalytic Study

In addition to identifying adsorbed species, IR spectroscopy and Raman spectroscopy are also used to monitor and analyze surface adsorption sites and reaction sites indirectly by observing the spectral change of an adsorbate known as the probe molecule.[54] In practice, CO is most frequently employed for this purpose, although NO and N_2 have also proved suitable on occasions. Wavenumber and band shape of the ν_{CO} are very sensitive to the chemical and electrostatic environment as well as to the mode of surface bonding that is dependent on the nature of adsorption site, coadsorbed species and surface coverage. This characterization of surface structure constitutes a major in situ application of IR, since it can be used to monitor changes in the catalyst during reaction. Moreover, examining the overall effects of electrode potential, together with the more specific influences of the solvent medium, electrolyte ions, and other adsorbates upon the vibrational spectrum of adsorbed CO provides a means of elucidating structure and bonding in metal–solution systems. To examine some more complex surface systems, two different probe molecules such as CO and NO conjugally adsorbed on the surface with different concentration together with the isotropic experiments are used for gaining detailed structural information of surface active sites.

4.3 Electrochemistry and Corrosion

4.3.1 Electrocatalysis[55–57]

Electrochemistry is one of the most ''hot'' areas because the vibrational data provide insight not only into the structure and orientation of adsorbates, but also regarding the nature of local interactions with the surrounding solvent, ions, and the surface, as well as with the variable electrostatic fields. Electrocatalysis is the central topic in the study due to its importance both fundamentally and industrially, which is related to fuel cells, batteries and electrochemical synthesis. IR spectroscopy has been shown to be very useful for the characterization of catalysts, determination of

catalytic reaction kinetics, and identification of the reaction intermediates. The richest information has been gained from the electrochemical behavior of small organic molecules (e.g. alcohol, formaldehyde, formic acid) on adsorption, dissociation and oxidation on transition metal surfaces (Pt, Pd, Ru, etc.).[55,56] Although EC-IRAS can be applied to any material surfaces, its drawback is that only smooth or mild rough surfaces with reasonably good reflectivity can be used when using the RA mode, which is the most commonly used method.

Recently, SERS study has been successfully extended to many bare transition metals of practical importance, in particular for electrocatalysis.[57] The remarkable advantages should be pointed out. The sample studied can be very dark in color and with highly rough surfaces. This means that practical catalysts can be directly investigated in situ. Moreover, the metal–adsorbate vibrational modes located in the low-wavenumber region can be easily observed, providing significant information about surface bonding of the adsorbate surface. The complementary study of IR and Raman will ensure a better level of understanding of adsorbate structure and surface bonding, and eventually of electrocatalytic reactivity.[28]

4.3.2 *Electrochemical Double Layer*[13,28,57]

Understanding the composition and structure of the electric double layer formed at the electrochemical interface constitutes one of the major objectives in electrochemistry. In particular, detailed knowledge of water molecules at the interface is an essential prerequisite to understanding electrocatalytic reactions. Unfortunately, the signal from interfacial water is overwhelmed by that from bulk water. SEIRAS with ATR configuration can probe the interface without the interference from the bulk solution and, hence, has been used to investigate the potential-dependent reorientation of water molecules at the Au(111) surface. Figure 14(a) shows a series of IR spectra from the electrode surface in perchloric acid, collected during a potential sweep from 0.1 to 1.3 V. The intensities and shapes of the νOH and δHOH modes of water change significantly as the potential changes, indicating different orientation of water molecules at the interface, see Figure 14(b). In the potential region around 0.6 V, the flat-lying water molecules do not give IR absorption because of the surface selection rule. Furthermore, the dynamics of the water reorientation caused by a potential step was studied at a time resolution of 10 µs. The spectra showed that the reorientation of water occurs within the initial 1 ms and is followed by the adsorption of electrolyte anion. Very recently, successful extension of SERS to the study of transition metals invoked substantial interest in electrochemical adsorption of hydrogen, oxygen, (pseudo-) halide ions and their coadsorption on transition metals.[58]

4.3.3 *Corrosion Processes*[22,59]

Early Raman study of the corrosion process involved the borrowing of the SERS effect of silver overlayers or sublayers to investigate the passivation film and the adsorbed corrosion inhibitors. The results become complicated with the introduction of Ag. The sensitivity provided by Raman and IR microscopy makes in situ investigation of the corrosion process under working conditions

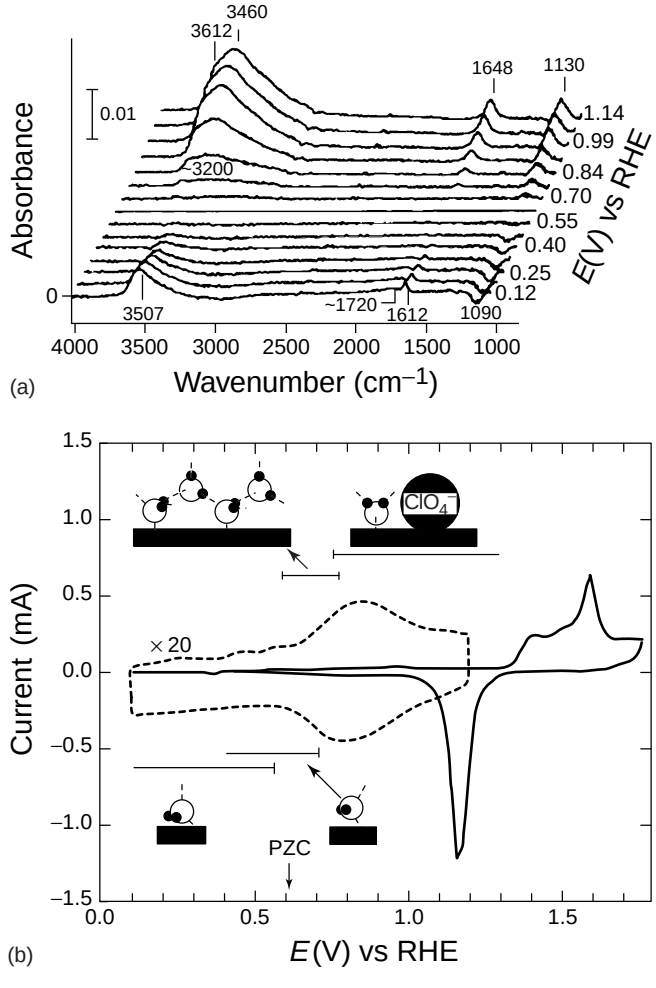

Figure 14 (a) A series of IR spectra of a highly ordered Au(111) surface in 0.5 M $HClO_4$ as a function of applied potential. The reference potential is 0.62 V vs RHE. (b) Cyclic voltammogram of a vacuum-evaporated Au thin film electrode in 0.5 M $HClO_4$ at a sweep rate of 50 mV s^{-1}. The insets show the orientation of water at the interface deduced from IR spectra. RHE, relative hydrogen electrode; PZC, potential of zero change. (Reproduced by permission from M. Osawa, *Bull. Chem. Soc. Jpn.*, **70**(12), 2861–2880 (1997).)

feasible.[59] With higher spatial resolution, the Raman imaging technique enables the two-dimensional distribution of several chemical composites from the corroded sample to be obtained, and it provides a means of rapidly obtaining surface composition information over a much larger analysis area. It is able to identify the anticorrosion effect of the coating on different substrates, and to study and compare the effect of corrosion inhibitors, such as thiourea and benzotriazole on the corrosion of iron surfaces. Two advantages of Raman microscopy are its higher spatial resolution both horizontally and vertically and its ability to record spectra in the low-frequency region. Hence, it is suitable to investigate the pitting corrosion process of metals and alloys involving various oxides and complexes with the chloride ion.[22]

4.4 Biology and Sensor

4.4.1 Biology and Biochemistry[60,61]

The main purpose of surface Raman spectroscopy to be applied in biological, biochemical and medical systems is to utilize the unique SERS effect with dramatically high sensitivity and molecular (or even chemical bond) selectivity as a powerful analytical tool, which has been well reviewed.[60,61] There is a good example of SERS study on humic substances adsorbed to noble metal substrates. A variety of SERS substrates have been proposed for studies of humic acid, including Ag sols, Au electrodes, and silver nitrate modified Fe–C–Cr–Ni surfaces. A SERS-active sample has also been prepared in the form of a pellet prepared from an aqueous mixture of humic acid and silver powder. Examples are neurotransmitters, immunoglobulin G, enzymes and amino acids, nucleic acids, eye lens pigments, and intermediate states of rhodopsin. Cytochrome c is a popular protein for SERRS investigations due to its intense heme absorbance and good adsorption characteristics. The orientation of this protein at colloidal Ag surfaces has been investigated under a variety of conditions. Photochemical charge transfer between flavin mononucleotide and a Ag electrode has been investigated by time-resolved SERS. Incorporation of anthracyclines into lipid mono- and bilayers has been proposed as a method for investigation of drug transportation through model or biological membranes by measurement of arrival times at the surface. SERS is becoming an increasingly popular method for investigation of drug–nucleic acid interactions, such as binding of acridine drugs to DNA and topoisomerase II, and antitumor drug action within living cancer cells. SERS has been used to identify covalent adducts between another antitumor drug [*N*(2)-methyl-9-hydroxyellipticinium] and amino acids and nucleic acids. In addition, the structure of adsorbed RNA triple helices has been investigated. It is necessary to emphasize that when SERS or any other surface technique is applied, care should be taken to ensure that the biological molecule keeps its activity and is not denaturalized when adsorbing or bonding at the surface.

4.4.2 Chemical Sensors[62]

Since SERS delivers vibrational spectra of the adsorbed analyte, which are highly specific for the detected substance, SERS has the potential of dual selectivity arising from both adsorption and detection. This can be used for the development of extremely selective sensors with low cross sensitivities or of multi-substance sensors with the identification of individual compounds by their spectra. Since SERS usually uses a metallic substrate for adsorption which will exhibit chemical reactivity and low selectivity, organic modification of the metal surface can solve this problem. Sensors utilizing SERS detection have been developed for a wide range of analytes, including organic molecules in aqueous solutions, nerve agents, and molecules of environmental importance. A SERS-based enzyme-linked sandwich immunoassay has also been developed, whereby the reaction of a peroxidase with *o*-phenyldiamine and hydrogen peroxide yielded azoaniline which was then detected at a SERS substrate. Subpicogram quantities of the explosive trinitrotoluene have been detected by SERS at colloidal Ag and Au.

4.4.3 Hyphenated Analytical Techniques[63]

Although there are many hyphenated analytical techniques to couple a separate method with IR and Raman spectroscopy, the application of surface IR and surface Raman methods is quite limited and restricted almost to SERS. SERS has been used as a detector for liquid chromatography by flowing a silver hydrosol as the SERS substrate through the chromatography column eluant in a windowless flow cell. The system was optimized with respect to integration time, eluant flow rate, and hydrosol flow rate, and the detection limits for five eluants were determined. Several papers focused on combining SERS with other analytical techniques such as gas chromatography, thin-layer chromatography, and flow injection analysis have appeared. A new gas chromatography detector composed of SERS substrate organic analytes at the surface, and a method for drug analysis by separating a mixture via thin-layer chromatography, followed by SERS detection from the chromatography plate, have been developed.

5 PROSPECTIVE AND FUTURE DEVELOPMENTS

Because of significant developments in materials science, laser and synchrotron technology, and nanotechnology, many opportunities are provided for the fundamental progress for IR spectroscopy and Raman spectroscopy to further enhance our understanding of surface science. The prospects will be given briefly and divided into three parts involving the surfaces, instrumentation and methodology.

5.1 Surfaces

5.1.1 Nanoparticle (Nanorod) Arrays

Enormous enhancements with SEF of up to 10^{14} have been found, by which high-quality SERS or SERRS spectra from a single molecule adsorbed on the surface of a single silver particle or aggregated silver colloids can be obtained.[29,30] It is very interesting to note that only substrate with particle size of 80–100 nm is optimal for SERS. Comparing with the normal SEF of 10^6, this phenomenon indicates SERS (maybe also SEIRA) substrates with much higher activity can be prepared, in particular with the aid of rapid development of nanofabrication technology. It is very important that the SERS substrate has a very narrow distribution of roughness features, i.e. optimized particle size and shape. Highly ordered periodic arrays of nanodots or nanorods can be prepared by lithography. This technique apparently is a good way to find suitable particle size and shape to generate SERS for various metals. Another approach is to make ordered nanorod arrays as very good SERS-active substrates especially for transition metals (e.g. Ni, Co and Fe); this has been realized in the authors' laboratory. The arrays are fabricated by electrodepositing metals into the nanoholes of the alumina film as the template.[64] The SERS intensities depend critically on the length of the rods emerging from the surface. This strategy will be very helpful for obtaining a deeper insight into complicated surface enhancement mechanisms of Raman and IR spectroscopy. The reasonable reproducibility of the regularly structured SERS and SEIRS substrates on a nanometer scale improves the applicability of these sensitive techniques in quantitative chemical analysis and sensors, and also offers the possibility for comparisons with theoretical approaches.

5.1.2 Single Crystal Surfaces and Non-surface-enhanced Raman Scattering Active Surfaces[24,65,66]

Although there have been quite extensive IR studies on atomic flat single crystal surfaces, only a few surface Raman studies have been reported due to its low sensitivity. The early Raman studies were restricted to vacuum–solid and air–solid interfaces. Only recently has a SERS study on true smooth single crystal surfaces been accomplished.[8,24,65] It should be noted that, at present, there are only a very limited number of molecules with large Raman cross-section that can be used successfully for this investigation. To extend the study to a variety of molecules is one of the major challenges to the further application of surface Raman spectroscopy. The instrumentation needs to be improved and the method needs to be developed including the use of the ATR mode.[24] Therefore, a detailed and comparative study on both the ordered monocrystalline and rough surfaces will achieve a better level of understanding of SERS mechanisms. Moreover, this approach is very helpful for linking the structural properties of in situ electrochemical or catalysis and metal–UHV interfaces, aiding and enriching the structural interpretation of both Raman and IR vibrational data of structural well-defined surfaces.

Starting with Campion's work in 1982, the development of high quantum efficiency, low noise multichannel detectors and high-throughput spectrometers led to a large increase in the fraction of the Raman signal that can be collected from a given sample. This has enabled several groups to obtain the Raman spectra of nonresonant adsorbates in the monolayer regime on low surface area substrates that do not support SERS. In a recent work by Kagan and McCreery,[66] the Raman study was extended to some physisorbed molecules on mechanically polished glassy carbon surface at monolayer level with and without the help of resonance Raman effect with reasonable signal-to-noise ratio. This allows insight into both the nature of the adsorbate–substrate interaction and its effect on both the vibrational and electronic properties of the adsorbate.

5.2 Instrumentation

5.2.1 Light Sources[26]

One advantage of surface IR spectroscopy in the mid-IR region over other vibrational techniques such as EELS or neutron scattering arises from its good frequency resolution and over Raman spectroscopy arises from its relatively high sensitivity. However, very few surface IR studies have successfully probed the far-IR region below 400 cm^{-1}, mainly because of the difficulty in obtaining high brightness radiation sources operating in this range. Traditional thermal sources tend to run out of energy below 300 cm^{-1}, excluding the important region where many adsorbate–substrate vibrations absorb. This effectively precludes direct observation by IR signals of the low wavenumber vibrational modes which have been shown to be important in vibrational energy transfer.

This brightness problem has recently been addressed using synchrotron radiation. The principal advantage of synchrotrons over blackbody sources arises because of their high brightness in a highly collimated beam. It has been shown that the intensity advantage from the use of synchrotron source can exceed that of conventional sources by a factor of 5 at 1000 cm^{-1}, rising to a factor of over 300 at 100 cm^{-1}. This permits spectra to be recorded in the far-IR with signal-to-noise ratios comparable to those obtained in the mid-IR region. Moreover, further development of tunable lasers in various wavelength regions from near-IR to UV will inevitably stimulate wider use of Raman spectroscopy.

5.3 Methods

5.3.1 Surface-enhanced Hyper-Raman Spectroscopy[67]

Hyper-Raman scattering is a nonlinear three-photon energy conversion process. It offers complementary information to IR Raman spectroscopy and has some advantages over Raman and IR spectroscopies due to its relaxed selection rules. In principle, the hyper-Raman active vibrational modes of a centrosymmetric molecule are complementary to its Raman-active vibrational modes. As a result, compared with its SERS spectrum, the surface-enhanced hyper-Raman (SEHR) spectrum for a centrosymmetric molecule shows some new vibrational bands and significant relative intensity changes. By contrast, the SEHR spectrum for a noncentrosymmetric molecule resembles its SER spectrum with the absence of new vibrational bands. However, the low scattering efficiency has single-handedly impeded the development of hyper-Raman to a prominent analytical technique. For surface species, one has to use surface enhancement to circumvent the low intensity problem. Surface-enhanced hyper-Raman spectroscopy (or scattering) (SEHRS) can greatly increase the surface detection sensitivity with an enhancement factor up to 10^{10}. It is possible to obtain rich vibrational information and observe both IR and Raman inactive vibrational modes. Most SEHRS can only obtain signals from adsorbates with large Raman cross-section (such as pyrazine) and dye sample, and most studies were done on colloid surfaces. The aim of research in this area is to find a stable SEHRS substrate with very high activity.

5.3.2 Hyphenated Technique of Raman and Infrared Spectroscopy with Scanning Probe Microscopy[68]

It is increasingly desirable to couple vibrational data with the corresponding information on surface morphology at microscopic level to probe surface adsorption and reaction sites. There are more than 20 techniques in surface science, including spectroscopies, diffraction techniques and scanning probe microscopy (SPM). Each

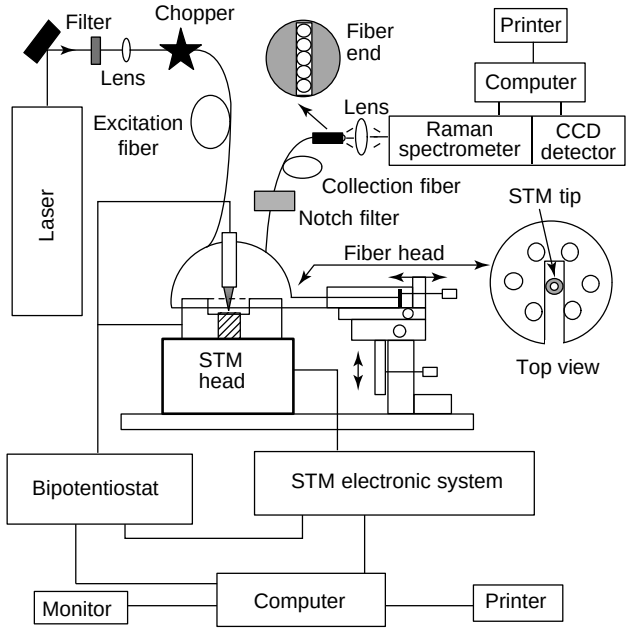

Figure 15 Schematic diagram of hyphenated Raman and STM technique.

technique has its own advantages and disadvantages in probing surfaces and interfaces. For instance, spatial resolution for Raman spectroscopy which provides chemical identification falls onto micrometer scale only as mentioned above, while scanning tunneling microscopy has extremely high spatial resolution up to atomic scale but without chemical identification. These two techniques are obviously complementary for probing interfaces for optical and morphological characterization. Several studies have been directed to investigate Ag films and colloids at higher resolution by combining SERS with SPM (STM or AFM) separately. There is no doubt that much useful and reliable information can be obtained if the advantages of these two techniques are used at the same time to study the interface. Thus, a hyphenated Raman/STM system was set-up in the authors' laboratory after designing the STM head, modifying the Raman optical system, and setting up a new high efficient collection system with the help of fiber optics and proper use of the detection device (Figure 15). Simultaneous STM and Raman measurement has been made on Ag surface and silver island covered highly oriented pyrolytic graphite electrodes to correlate the SERS activity and surface morphology. Further development of a very stable SPM system combined with a highly sensitive Raman system (perhaps an IR system later) makes this technique promising.

5.3.3 Scanning Near-field Optical Microscopy

Another important approach of the hyphenating technique is to develop SNOM that allows optical imaging

with a spatial resolution (~20 nm) well beyond the diffraction limit. This technique is described in detail elsewhere in this Encyclopedia. Combination of SNOM with Raman spectroscopy will provide a powerful tool to obtain chemical information of surfaces (interfaces) with nanometer spatial resolution. Since cross-sections of Raman scattering are usually extraordinary small, however, the combination of SNOM and Raman spectroscopy will be facilitated by the use of resonance Raman scattering and SERS. High spatial resolution SNOM equipment has been developed by using high efficiency optical fiber probes, and Raman spectra have been obtained of copper phthalocyanine film with a thickness of about 50 nm on a prism. However, it should be noted that the sensitivity of this technique is still quite low compared to the conventional Raman system due to its extremely low output of laser power. Further sophisticated development of the instrumentation and use of tunable IR lasers, with wavelengths covering mid-IR and even some far-IR regions, for IR study are expected to broaden the application of this method.

ACKNOWLEDGMENTS

This work has been made possible by financial support of the Natural Science Foundation of China (to Z.Q.T. and B.R.) and the Ministry of Education of China (Z.Q.T.). We are grateful to K.Q. Huang, D.Y. Wu, C.X. She, J.M. Wu and F.M. Liu who have provided much help and assistance in editing the article. We are also grateful to B.W. Mao, Y.Y. Liao, C. Li, S.G. Sun, M.J. Weaver and W.Z. Weng for their valuable comments and suggestions. Whenever the work from the authors' group is mentioned in the article, it is the contribution of self-motivated and hard working students and all other group members.

ABBREVIATIONS AND ACRONYMS

AFM	Atomic Force Microscopy
ATR	Attenuated Total Reflection
CW	Continuous Wave
EC-IRAS	Electrochemical Infrared Absorbance Spectroscopy
EELS	Electron Energy Loss Spectroscopy
EMIRS	Electrochemically Modulated Infrared Spectroscopy
FTIR	Fourier Transform Infrared
ICCD	Intensified Charge Coupled Device
IR	Infrared
IRAS	Infrared Absorption Spectroscopy
LB	Langmuir–Blodgett
NA	Numerical Aperture
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
ORC	Oxidation–Reduction Cycle
PARS	Potential-averaged Raman Spectroscopy
PDA	Photodiode Array
PDIRS	Potential Difference Infrared Spectroscopy
RA	Reflection–Absorption
SAM	Self-assembled Monolayer
SCE	Saturated Calomel Electrode
SEF	Surface Enhancement Factor
SEHR	Surface-enhanced Hyper-Raman
SEHRS	Surface-enhanced Hyper-Raman Spectroscopy (or Scattering)
SEIRA	Surface-enhanced Infrared Absorption
SEIRAS	Surface-enhanced Infrared Absorption Spectroscopy
SER	Surface-enhanced Raman
SERRS	Surface-enhanced Resonance Raman Scattering
SERS	Surface-enhanced Raman Spectroscopy (or Scattering)
SNIFTIRS	Subtractively Normalized Interfacial Fourier Transform Infrared Spectroscopy
SNOM	Scanning Near-field Optical Microscopy
SPM	Scanning Probe Microscopy
SPP	Surface Plasmon Polariton
STM	Scanning Tunneling Microscopy
UHV	Ultra High Vacuum
UV	Ultraviolet

RELATED ARTICLES

Particle Size Analysis **(Volume 6)**
Surface Area and Pore Size Distributions

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis

Electroanalytical Methods **(Volume 11)**
Infrared Spectroelectrochemistry • Scanning Tunneling Microscopy, In Situ, Electrochemical • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Microspectroscopy • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation • Raman Microscopy and Imaging • Raman Scattering, Fundamentals

REFERENCES

1. M.K. Urban, *Vibrational Spectroscopy of Molecules and Macromolecules on Surfaces*, Wiley & Sons, New York, 1993.
2. P. Hollins, J. Pritchard, 'Infrared Studies of Chemisorbed Layers on Single Crystals', *Prog. Surf. Sci.*, **19**(4), 275–350 (1985).
3. A. Bewick, S. Pons, 'Infrared Spectroscopy of the Electrode–Electrolyte Solution Interface', in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Wiley Hayden, Chichester, 1–63, Vol. 12, 1985.
4. B. Zhao, H. Li, X. Zhang, J. Shen, Y. Ozaki, 'Infrared Study on Molecular Orientation and Phase Transition in Langmuir–Blodgett Films of an Amphiphilic Microgel Copolymer with the Branching of Octadecyl Groups', *J. Phys. Chem. B*, **120**(34), 6515–6520 (1998).
5. M. Moskovits, 'Surface Selection Rules', *J. Chem. Phys.*, **77**, 4408–4412 (1982).
6. J.A. Creigton, 'The Selection Rules for Surface-enhanced Raman Spectroscopy', in *Spectroscopy of Surface*, eds. R.J.H. Clark, R.E. Hester, Wiley & Sons, New York, 37–89, 1988.
7. R.P. Van Duyne, 'Laser Excitation of Raman Scattering from Adsorbed Molecules Electrode Surfaces', in *Chemical and Biochemical Applications of Lasers*, ed. C.B. Moore, Academic Press, New York, 101–105, Vol. 4, 1979.
8. A. Campion, P. Kambhampati, 'Surface-enhanced Raman Scattering', *Chem. Soc. Rev.*, **27**, 241–249 (1998).
9. M. Moskovits, 'Surface-enhanced Raman Spectroscopy', *Rev. Mod. Phys.*, **57**(3), 783–826 (1985).
10. A. Otto, I. Mrozek, H. Grabhorn, W. Akemann, 'Surface-enhanced Raman Scattering', *J. Phys.: Condens. Matter*, **4**, 1143–1212 (1992).
11. J. Corset, J. Aubard, 'Surface Enhanced Raman Scattering: New Trends and Applications', *J. Raman Spectrosc., Special Issue*, **29**(8) (1999).
12. W.B. Cai, B. Ren, X.Q. Li, C.X. She, F.M. Liu, X.W. Cai, Z.Q. Tian, 'Investigation of Surface-enhanced Raman Scattering from Platinum Electrodes Using a Confocal Raman Microscope: Dependence of Surface Roughening Pretreatment', *Surf. Sci.*, **406**, 9–22 (1998).
13. M. Osawa, 'Dynamics Processes in Electrochemical Reactions Studied by Surface-enhanced Infrared Absorption Spectroscopy (SEIRAS)', *Bull. Chem. Soc. Jpn.*, **70**(12), 2861–2880 (1997).
14. G.Q. Lu, S.G. Sun, L.R. Cai, S.P. Chen, Z.W. Tian, K.-K. Shiu, 'In Situ FTIR Spectroscopic Studies of Adsorption of CO, SCN^-, and Poly(*o*-Phenylenediamine) on Electrodes of Nanometer Thin Films of Pt, Pd, and Rh: Abnormal Infrared Effects (AIRES)', *Langmuir*, **16**, 778–786 (1999).
15. R.G. Greenler, 'Infrared Study of Adsorbed Molecules on Metal Surfaces by Reflection Techniques', *J. Chem. Phys.*, **44**, 310–315 (1966).
16. A. Bewick, K. Kunimatsu, B.S. Pons, J.W. Russell, 'Electrochemically Modulated Infrared Spectroscopy (EMIRS)', *J. Electroanal. Chem.*, **160**, 47–61 (1984).
17. W. Richter, 'Resonant Raman Scatters', in *Solid State Physics*, ed. G. Hohler, Springer, Berlin, 121–272, Vol. 78, 1976.
18. M. Fleischmann, I.R. Hill, 'Raman Spectroscopy', in *Comprehensive Treatise of Electrochemistry*, eds. R.E. White, J. O'M Bockris, B.E. Conway, E. Yeager, Plenum Press, New York, 373–443, Vol. 8, 1985.
19. J.W. Niemantsverdriet, 'The Vibrational Spectroscopies', in *Spectroscopy in Catalysis*, VCH, Weinheim, 193–218, 1993.
20. C.T. Williams, A.A. Tolia, H.Y.H. Chan, C.G. Takoudis, M.J. Weaver, 'Surface-enhanced Raman Spectroscopy as an In Situ Real-time Probe of Catalytic Mechanisms at High Gas Pressures: the CO–NO Reaction on Platinum and Palladium', *J. Catal.*, **163**, 63–76 (1996).
21. G. Mestl, H. Knozinger, 'Vibrational Spectroscopies', in *Handbook of Heterogeneous Catalysis*, eds. G. Ertl, H. Knozinger, J. Weitkamp, VCH, Weinheim, 539–574, Vol. 2, 1997.
22. C.A. Melendres, 'Laser Raman Spectroscopy in Studies of Corrosion and Electrocatalysis', in *Spectroscopic and Diffraction Techniques in Interfacial Electrochemistry*, eds. C. Gutierrez, C. Melendres, Kluwer Academic Publishers, Boston, 181–222, 1988.
23. B. Pettinger, 'In Situ Raman Spectroscopy at Metal Surfaces', in *Adsorption of Molecules at Metal Electrodes*, eds. J. Lipkowski, P.N. Ross, VCH, New York, 285–345, 1992.
24. A. Bruckbauer, A. Otto, 'Raman Spectroscopy of Pyridine Adsorbed on Single Crystal Copper Electrodes', *J. Raman. Spectrosc.*, **29**, 665–672 (1998).
25. M.C. Petty, 'Characterization and Properties', in *Langmuir–Blodgett Films*, ed. G. Roberts, Plenum, New York, Chapter 4, 1990.
26. A. Horn, 'Infrared Spectroscopic Technique for the Study of Thin Interfacial Films', in *Spectroscopy for Surface Science*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, New York, 273–339, Vol. 26, 1998.

27. G. Xue, 'Fourier Transform Raman Spectroscopy and Its Application for the Analysis of Polymeric Materials', *Prog. Polym. Sci.*, **22**, 313–406 (1997).
28. M.J. Weaver, S.Z. Zou, 'Vibrational Spectroscopy of Electrochemical Interfaces: Some Walls and Bridges to Surface Science Understanding', in *Spectroscopy for Surface Science*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, New York, 219–272, Vol. 26, 1998.
29. K. Kneipp, Y. Wang, H. Kneipp, L.T. Perelman, I. Itzkan, R.R. Dasari, M.S. Feld, 'Single Molecule Detection Using Surface-enhanced Raman Scattering (SERS)', *Phys. Rev. Lett.*, **78**(9), 1667–1670 (1997).
30. S. Nie, S.R. Emory, 'Probing Single Molecules and Single Nanoparticles by Surface-enhanced Raman Scattering', *Science*, **275**, 1102–1106 (1997).
31. J.E. Katon, 'Infrared Microspectroscopy', in *Modern Techniques in Applied Molecular Spectroscopy*, ed. F.M. Mirabella, John Wiley & Sons, New York, 267–289, 1998.
32. A.J. Sommer, 'Raman Microspectroscopy', in *Modern Techniques in Applied Molecular Spectroscopy*, ed. F.M. Mirabella, John Wiley & Sons, New York, 291–322, 1998.
33. X.M. Yang, D.A. Tryk, K. Hashimoto, A. Fujishima, 'Surface-enhanced Raman Imaging (SERI) as a Technique for Imaging Molecular Monolayers with Chemical Selectivity Under Ambient Conditions', *J. Raman Spectrosc.*, **29**, 725–732 (1998).
34. R.R. Cavanagh, T.A. Germer, J.C. Stephenson, 'Ultrafast Time-resolved Infrared Probing of Energy Transfer at Surfaces', *Vibrat. Spectrosc.*, **9**, 77–83 (1995).
35. M. Osawa, K. Yoshii, K. Ataka, T. Yotsuyanagi, 'Real-time Monitoring of Electrochemical Dynamics by Submillisecond Time-resolved Surface-enhanced Infrared Attenuated-total-reflection Spectroscopy', *Langmuir*, **10**, 640–642 (1994).
36. S.G. Sun, Y. Li, 'Kinetics of Isopropanol Oxidation on Pt(111), Pt(110), Pt(100), Pt(610) and Pt(211) Single Crystal Electrode-studies of In Situ Time-resolved FTIR Spectroscopy', *Electrochim. Acta*, **44**, 1153–1162 (1998).
37. P. Gao, D. Gosztola, M.J. Weaver, 'Surface-enhanced Raman Spectroscopy as a Probe of Electroorganic Reaction Pathways. I Processes Involving Adsorbed Nitrobenzene, Azobenzene, and Related Species', *J. Phys. Chem.*, **92**, 7122–7130 (1988).
38. C. Shi, W. Zhang, J.R. Lombardi, R.L. Birke, 'Nanosecond Timescale Kinetics in the Flavin Mononucleotide on an Electrode Surface Using Time-resolved Surface-enhanced Raman Spectroscopy', *J. Phys. Chem.*, **96**, 10093–10096 (1992).
39. Z.Q. Tian, W.H. Li, S.Z. Zou, B.W. Mao, J.S. Gao, 'Potential Averaged Surface Enhanced Raman Spectroscopy', *Appl. Spectrosc.*, **50**, 1569–1577 (1996).
40. P.C. Stair, C. Li, 'Ultraviolet Raman Spectroscopy of Catalysts and Other Solids', *J. Vac. Sci. Technol. A*, **15**, 1679–1684 (1997).
41. R.P. Van Duyne, J.P. Haushalter, 'Surface-enhanced Resonance Raman Spectroscopy of Adsorbates on Semiconductor Electrode Surface: Tris(bipyridine)ruthenium (I) Adsorbed on Silver-modified n-GaAs(100)', *J. Phys. Chem.*, **87**, 2999–3003 (1983).
42. M. Niwano, T. Miura, Y. Kimura, R. Tajima, N. Miyamoto, 'Real-time, In-situ Study of Etching of Si(100) and (111) Surfaces in Dilute Hydrofluoric Acid Solution', *J. Appl. Phys.*, **79**, 3708–3713 (1996).
43. B. Ren, F.M. Liu, J. Xie, B.W. Mao, Y.B. Zu, Z.Q. Tian, 'In-situ Monitoring of Raman Scattering and Photoluminescence from Silicon Surfaces in HF Aqueous Solutions', *Appl. Phys. Lett.*, **72**, 933–936 (1998).
44. S.Z. Zou, M.J. Weaver, 'Surface-enhanced Raman Scattering of Ultrathin Cadmium Chalcogenide Films on Gold Formed by Electrochemical Atomic-layer Epitaxy: Thickness-dependent Phonon Characteristics', *J. Phys. Chem. B*, **103**, 2323–2326 (1999).
45. M.B. Mitchell, 'Structure–Property Relations in Polymers', *Adv. Chem.*, **236**, 351–375 (1993).
46. L.J. Fina, 'Depth Profiling of Polymer Surfaces with FTIR', *Appl. Spectrosc. Rev.*, **29**, 309–365 (1994).
47. G. Xue, Y. Lu, G. Shi, 'Surface-enhanced Raman Scattering Used as an In Situ Analytical Method for Studies of Polymer–Metal Interface', *Polymer*, **35**, 2488–2494 (1994).
48. M.A. Byant, J.E. Pemberton, 'Surface Raman Scattering of Self-assembled Monolayers Formed from 1-Alkanethiols at Ag', *J. Am. Chem. Soc.*, **113**, 3629–3637 (1991).
49. C. Li, G. Xiong, Q. Xin, J.K. Liu, P.L. Ying, Z.C. Feng, J. Li, W.B. Yang, Y.Z. Wang, G.R. Wang, X.Y. Liu, M. Lin, X.Q. Wang, E.Z. Min, 'UV Resonance Raman Spectroscopic Identification of Titanium Atoms in the Frame Work of TS-1 Zeolite', *Angew. Chem., Int. Ed. Engl.*, **38**(15), 2220–2222 (1999).
50. M. Bonn, H.J. Bakker, K. Domen, C. Hirose, A.W. Kleyn, R.A. Van Santen, 'Dynamical Studies of Zeolitic Protons and Adsorbates by Picosecond Infrared Spectroscopy', *Catal. Rev. Sci. Eng.*, **40**, 127–173 (1998).
51. H. Knozinger, 'Characterization of Heterogeneous Catalysts by Vibrational Spectroscopies', in *Fundamental Aspects of Heterogeneous Catalysis Studied by Particle Beams*, eds. H.H. Brongersma, R.A. Van Santen, Plenum Press, New York, 167–189, 1991.
52. M. Colombo, M.W. George, J.N. Moore, D.I. Pattison, R.N. Perutz, I.G. Virrels, T.Q. Ye, 'Methanol Synthesis Pathway over Cu/ZrO_2 Catalysts: a Time-resolved DRIFT C-13-Labelling Experiment', *Catal. Lett.*, **54**, 41–48 (1998).
53. R.Q. Long, Y.P. Huang, H.L. Wan, 'Surface Oxygen Species over Cerium Oxide and Their Reactivities with Methane and Ethane by Means of In Situ Confocal Microprobe Raman Spectroscopy', *J. Raman Spectrosc.*, **28**, 29–32 (1996).

54. V.M. Browne, S.G. Fox, P. Hollins, 'Infrared Spectroscopy as an In Situ Probe of Morphology', *Catal. Today*, **9**, 1–14 (1991).
55. T. Iwasita, F.C. Nart, 'In-situ Fourier Transform Infrared Spectroscopy: A Tool to Characterize the Metal–Electrolyte Interface at Molecular Level', in *Surface Infrared and Raman Spectroscopy: Methods and Applications*, ed. W. Suetaka, Plenum, New York, 126–157, 1995.
56. S.G. Sun, 'Studying Electrocatalytic Oxidation of Small Organic Molecules with In-situ Infrared Spectroscopy', in *Electrocatalysis*, eds. J. Lipkowski, P.N. Ross, Wiley-VCH, New York, 243–291, 1998.
57. Y. Shingaya, M. Ito, 'Modeling of the Electrochemical Double-layer in Ultrahigh Vacuum: Adsorption of Sulfuric Acid Species on Pt(111)', in *Interfacial Electrochemistry*, ed. A. Wieckowski, Marcel Dekker, New York, 287–300, 1999.
58. Z.Q. Tian, J.S. Gao, X.Q. Li, B. Ren, Q.J. Huang, W.B. Cai, F.M. Liu, B.W. Mao, 'Can Surface Raman Spectroscopy be a General Technique for Surface Science and Electrochemistry?', *J. Raman Spectrosc.*, **29**, 703–711 (1998).
59. N. Boucherit, A. Hugot-Le Goff, 'Localized Corrosion Processes in Iron and Steels Studied by In Situ Raman Spectroscopy', *Faraday Discuss.*, **94**, 137–147 (1992).
60. T.M. Cotton, 'The Application of Surface-enhanced Raman Scattering to Biochemical Systems', in *Spectroscopy of Surfaces*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, New York, 91–153, Vol. 16, 1988.
61. R. Nabiev, K.V. Sokolov, M. Manfait, 'Surface-enhanced Raman Spectroscopy and its Biomedical Applications', in *Biomolecular Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, London, 267–338, 1993.
62. W. Hill, B. Wehling, V. Fallourd, D. Klockow, 'Application of Surface-enhanced Raman Scattering for Chemical Sensors', *Spectrosc. Eur.*, **7**(5), 20–24 (1995).
63. R.L. Garrel, 'Surface-enhanced Raman Spectroscopy', *Anal. Chem.*, **61**, 401A–411A (1989).
64. J.L. Yao, K.H. Xue, G.P. Pan, B. Ren, Y. Xie, B.W. Mao, Z.Q. Tian, 'Surface Enhanced Raman Scattering from Metal Nano-rod Arrays', *Pure Appl. Chem.*, **72**(1), 221–228 (2000).
65. M. Futamata, 'Highly-sensitive Raman Spectroscopy Using Surface-plasmon-polariton', in *Proceedings of the Sixteenth International Conference on Raman Spectroscopy*, ed. M. Heyns, John Wiley & Sons, Chichester, 356–359, 1998.
66. M.R. Kagan, R.L. McCreery, 'Quantitative Surface Raman Spectroscopy of Physisorbed Monolayers on Glassy Carbon', *Langmuir*, **11**, 4041–4047 (1995).
67. W.H. Li, X.Y. Li, N.T. Yu, 'Surface-enhanced Hyper-Raman Spectroscopy (SEHRS) and Surface-enhanced Raman Spectroscopy (SERS) Studies of Pyrazine and Pyridine Adsorbed on Silver Electrodes', *Chem. Phys. Lett.*, **302**, 303–310 (1999).
68. Z.Q. Tian, W.H. Li, B. Ren, B.W. Mao, J.G. Chen, J.Q. Mu, X.D. Zhuo, D. Wang, 'Simultaneous STM and Raman Measurements on Electrochemical Interfaces', *J. Electroanal. Chem.*, **401**, 247–251 (1996).

Ion Scattering Spectroscopy in Analysis of Surfaces

Robert A. Langley
Kingston, TN, USA

1 Introduction 1
2 Equipment 2
2.1 Beam Delivery System 2
2.2 Reaction/Scattering Chamber 2
2.3 Scattered Particle and Reactant Radiation Measurement 2
3 Ion Beam Analysis Techniques 3
3.1 Elastic Scattering 3
3.2 Inelastic Scattering 5
3.3 Channeling and Blocking 7
3.4 Miscellaneous 7
4 Data Analysis Techniques, Data Sources, Summary and Future Directions 7
Abbreviations and Acronyms 7
Related Articles 8
References 8

Ion scattering spectroscopy (ISS) is used to analyze materials for their atomic constituents for depths from a monolayer to many microns. The probing depth and depth resolution are mainly determined by the energy and mass of the projectile. The projectile energy ranges from several thousand to many million electron-volts. For ISS there are many options for analysis, e.g. scattered incident projectiles, knock-on target atoms, reactants from nuclear reactions between the projectile and target atom, and X-rays from excitation of target atoms by the projectiles. There are also numerous experimental techniques for studying various particular material characteristics, e.g. crystal structure, crystal damage, lattice deformation, and impurity content and location. Each of these areas is addressed, as well as the equipment needed for their implementation. Review articles and books on specific relevant subjects are noted, as well as data analysis techniques and computer analysis and simulation programs.

1 INTRODUCTION

The use of ion beam analysis (IBA) techniques allows the determination of the atomic constituents of a sample (including impurities) as a function of depth and also offers a method to obtain information on the structure of the sample (crystallinity, crystal damage and lattice deformation). The availability of IBA techniques has lead to surface modification techniques and multilayer thin-film structures. IBA techniques also stimulated the development of thin-film deposition techniques and the study of oxidation, reduction and analytical reactions under well-controlled conditions. Although IBA can easily distinguish between depth profile atomic concentrations, chemical bonding effects are generally not observed.

When ions bombard material, the incident particles are scattered or react with the material to give off secondary particles (ions, electrons, etc.) and/or electromagnetic radiation. A fraction of the reacting particles and/or radiation emerge from the sample and a small percentage enters the detector and is detected. Identification of the constituent elements is based on the energy of the emerging particles or radiation, whereas the atomic concentration is derived from the intensity of the emerging particles or radiation.

For IBA the near-surface region is regarded as a bulk region because physical properties generally correspond to those of the bulk material, whereas the outermost atomic layers – the surface – may exhibit significantly different properties from the bulk region.

Section 2 discusses how particles are accelerated and presents examples of the types of accelerators commonly used. Section 3 discusses the various IBA techniques, including their applications. IBA techniques are generally divided into several classes depending on the energy of the beam particles and the type of reaction that the incident energetic ion undergoes. Low-energy ion scattering (LEIS) and secondary ion mass spectrometry (SIMS) use projectiles up to several thousand electron-volts, medium-energy ion scattering (MEIS) includes projectile energies of 10–500 keV and high-energy ion scattering (HEIS) ranges in energy from one to many million electron-volts per atomic mass unit. The latter two energy-range categories include Rutherford backscattering spectroscopy (RBS), elastic recoil detection (ERD), particle-induced X-ray emission (PIXE), and nuclear reaction analysis (NRA). This section (section 3) includes a discussion of the techniques of channeling and blocking that are used in highly structured crystals to determine structural parameters in surfaces and interfaces and the atomic position of impurity atoms. Each of these techniques will be discussed in detail and their major applications given.

Section 4 presents analysis techniques for application of the experimental data and provides references for many of the necessary cross-sections used in the analysis, closing with a brief summary and an indication of future directions.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 EQUIPMENT

This section is divided into three parts, technically noted by: ionization of an atom, its acceleration and energy selection; scattering of the projectile ion or reaction of the projectile ion with the target nuclei; and energy analysis of the emerging particles and radiation. These three parts will be discussed along with a few ancillary matters. For a more complete discussion of this subject, see Tesner and Nastasi[1] and Chu et al.[2]

2.1 Beam Delivery System

This section is devoted to an explanation of the main types of accelerators used to produce and accelerate charged particles for the study of materials. Charged particles are created in an ion source and can be accelerated to large velocities by passing them through a large potential difference. In order to accomplish this, a high voltage must be produced; Cockcroft and Walton used Greinacher's idea of multiplying a given voltage by a suitable arrangement of condensers and rectifiers, and in 1932 they built a cascade generator for use in nuclear reaction studies.[3] Cockcroft–Walton-type accelerators are now widely used, especially in ion implanters in the energy range up to 500 kV. Another method to produce high voltages was discovered by van de Graaff in 1931:[4] a fast-moving insulating belt carries charge that is sprayed onto a belt at one electrode and removed at another electrode. In this manner, potentials up to tens of millions of volts can be obtained. The van de Graaff accelerator has become the workhorse for IBA studies in the last few decades and is still used extensively. A modification of this technique has been applied by replacing the insulating belt with an insulating chain. An additional modification has been the creation of the 'tandem' accelerator, which initially injects negative ions into a van de Graaff type of accelerator where the high-voltage terminal, positively charged, is in the center of the accelerating column and as the negative ions pass through this high-voltage center one or more electrons are stripped, i.e. removed, in a gas or thin foil to form a positively charged ion that is accelerated further from the positive high-voltage terminal in the center of the accelerator to ground potential. There are two distinct advantages of the tandem accelerator: the ion source is at or near ground potential and allows various kinds of ion sources to be used; and substantially higher energy ions can be obtained because multiply-charged positive ions can be created in the high-voltage region of the tandem accelerator, thereby creating ions with energies much larger than the voltage on the center terminal of the accelerator.

The accelerator types discussed thus far are linear electrostatic accelerators, but there are other types of accelerators that have been used previously. One of these types is a circular accelerator known as a cyclotron,[5] which accelerates ions through a small potential difference many times as they travel in a nearly circular path. Put simply, there are two semicircular electrodes, electrically isolated from each other, with an oscillating voltage applied between the two electrodes. As the ions travel through one semicircular electrode, the voltage between the electrodes is switched and the ions accelerate by twice the applied voltage for each revolution in the cyclotron. The ions follow circular orbits in a homogeneous magnetic field and pass through the alternating voltage many times in order to reach the energy range required for the types of studies considered here.

In most cases the beam leaving the accelerator is not monoenergetic nor necessarily of the same mass. In order to separate the useful beam from the other beam components, either magnetic or electrostatic deflection is used. This deflection usually provides a near-monoenergetic beam of a single mass. Then, the deflected and analyzed beam usually is focused and steered into the scattering chamber and onto the target that is to be analyzed. For a complete description of these aspects, see Tesner and Nastasi[1] and Chu et al.[2]

2.2 Reaction/Scattering Chamber

The reaction chamber contains the material to be analyzed, i.e. the sample, a sample manipulator onto which the sample is mounted and used to position the portion of the sample to be analyzed in front of the incident beam, and detectors to analyze both the mass and the energy of the scattered particles or reactant radiation. In addition, the chamber may have attached to it any necessary equipment to perform ancillary experiments in situ, such as oxidation or diffusion. For a more complete discussion of this subject and papers pertaining to the individual techniques and experiments that are referenced in section 3, see Tesner and Nastasi.[1]

2.3 Scattered Particle and Reactant Radiation Measurement

Electrostatic energy analyzers coupled with a Faraday cup or an electron multiplier are generally used for LEIS systems, whereas solid-state semiconductor detectors are used in higher energy scattering experiments. The electrostatic energy analyzer is of the same construction as the analyzers used in electron spectrometers but with opposite polarities applied to all lens and analyzer elements. In most circumstances the energy resolution ($\Delta E/E$) is 0.05–0.5%. The ions are usually measured with a position-sensitive microchannel-plate detector. For more energetic ions, i.e. $E > 100$ keV, gold barrier silicon

detectors generally are used. These detectors have energy resolutions that vary from about 5 keV to a few tens of keV, depending on many factors, but usually the largest contributing factor is the active area of the detector, with the resolution increasing as the detector area increases. For a complete description of these detectors and others noted later in this contribution, see Knoll.[6]

3 ION BEAM ANALYSIS TECHNIQUES

3.1 Elastic Scattering

3.1.1 Rutherford Backscattering Spectroscopy

The most widely used ion scattering technique for the analysis of surfaces is elastic ion scattering, also known as RBS after Sir Ernest Rutherford's proposal of the new model for an atom in 1911.[7] Geiger and Marsden[8] performed an experiment to test Rutherford's theory and unambiguously confirmed the validity of the proposed model. In doing the experiment they came upon a method of analyzing samples. The basic concept of this technique is equivalent to bouncing a billiard ball of one mass off another ball of different mass. If the mass of the projectile ball is less than the mass of the target ball, the projectile ball will bounce backward from the collision with an energy that is dependent on its initial energy, the angle of scattering, and the ratio of the projectile mass to the target atom mass. If the projectile mass is equal to or greater than the target mass, only forward scattering is allowed by the laws of conservation of energy and momentum. A schematic representation of such an elastic collision between a projectile of mass m_1, velocity V_0, initial energy E_0 and final energy E_1, and a target of mass m_2, initially at rest, is shown in Figure 1.

In Figure 1 the scattering angle of the projectile is denoted by θ and the recoil angle of the target with respect to the incoming direction of the projectile is denoted by ϕ. All quantities are in the laboratory frame of reference. There is a direct correspondence between the energy of the projectile before the collision and after the collision, dependent on the mass of the projectile m_1, the mass of the target, m_2, and the scattering angle θ. The ratio E_1/E_0 is called the kinematic factor K, and is dependent on m_1, m_2, and θ, as shown by Equation (1):

$$K = \frac{(m_2^2 - m_1^2 \sin^2\theta)^{1/2} + m_1 \cos\theta}{m_1 + m_2}^2 \qquad (1)$$

Thus, if one knows the mass and energy of the projectile, then the mass of the target can be determined if the scattered projectile energy is measured for a particular scattering angle. The probability of a scattering event

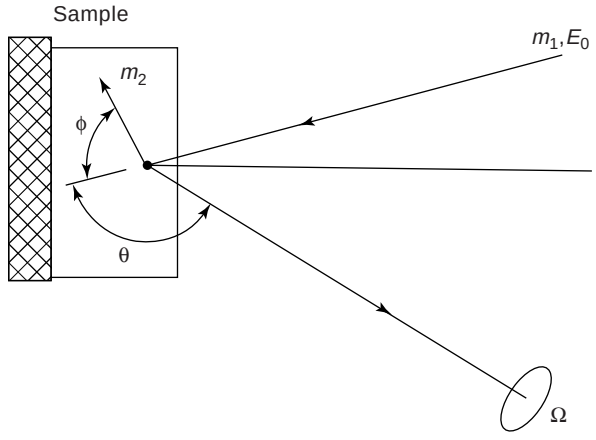

Figure 1 Schematic representation of an elastic collision between a projectile of mass m_1, with velocity V_0 and energy E_0 and a target of mass m_2 at rest; θ is the scattering angle of the projectile and ϕ is the recoil angle of the target atom.

occurring is directly related to the cross-section for elastic scattering σ. More precisely, the differential cross-section $d\sigma$ is directly proportional to the probability of a scattering event of angle θ into a solid angle Ω. There are two implicit assumptions: that the projectile energy is much greater than the binding energy of the atoms in the target, i.e. of the order of 10 eV; and that the projectile energy is less than the energy for any nuclear reactions or resonances of the incident particles with target nuclei.

The differential elastic scattering cross-section is given by Equation (2):

$$\sigma(E,\theta) = \left(\frac{1}{Nt} \quad \frac{dQ(E)}{Q} \frac{1}{\Omega(\theta)}\right. \qquad (2)$$

where θ is the scattering angle, Q is the total number of projectile ions, dQ is the number of projectile ions scattered into the detector that subtends a solid angle Ω, N is the volume density of atoms in the target and t is the target thickness. Therefore, if the differential cross-section, the solid angle subtended by the detector, and the number of scattered projectiles are known, then the areal density, Nt, is determined. For a complete derivation of these equations, see Chu et al.[2]

As the projectiles proceed through the target they encounter many bound electrons that interact with the projectile, causing very small energy loss of the projectile, so that as the projectile penetrates deeper into the target material its energy decreases. The same effect also occurs on the outbound path of the projectile or the reactant nucleus. This leads to an energy loss on both the inbound and outbound paths. From this energy loss and the scattering geometry, the thickness of the target can be determined. A typical RBS spectrum for a 2.0-MeV He^4 ion backscattered from a gold film overlayed on a

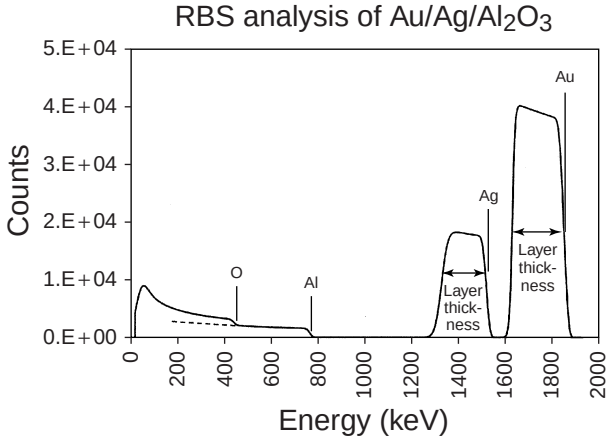

Figure 2 Rutherford backscattering spectrum for a 2.0-MeV He^4 ion backscattered from a gold film overlayed on a silver film overlayed on a sapphire substrate. The thickness of each film is 10^{14} atoms m^{-2} and the scattering angle is 165°.

silver film overlayed on a sapphire substrate is shown in Figure 2, where the thickness of each film is determined to be 10^{14} atoms m^{-2}. The high-energy scattering edge is indicated for both Au and Ag. From this spectrum it can be seen that the Au and Ag films have not interdiffused and so it is easy to determine the thickness of each film.

The reactant target atom receives energy from the projectile and may be dislodged from its lattice site if the imparted energy is greater than the binding energy of the target atom. As the dislodged target atom moves through the lattice, it creates damage sites along its path.

In the geometric arrangement shown in Figure 1, multilayer films may be analyzed as well as interdiffusion between layers. Low-Z impurities are difficult to analyze in this arrangement but can be measured using ERD and NRA.

Because the elastic scattering cross-section is well known from Rutherford's work,[7] absolute measurement of individual constituents in a sample can be made if the geometry of the scattering system is known, if measurement of the beam parameters is made and if measurement of the scattered projectile is determined. This applies only if the nuclear scattering event is such that the nuclear potentials of the scattering nuclei do not overlap. The onset of this overlap and its effect on the scattering cross-section are well covered by Tesner and Nastasi.[1]

Using this technique, depth resolutions of a few tens of nanometers are possible and concentrations of target constituents as low as a few tens of parts-per-million for medium-Z elements and 10 ppm for high-Z elements are possible, although under special conditions smaller depth resolutions and lower concentrations may be obtained. Low-Z ($Z < 10$) elements are difficult to measure using this technique but other avenues are available and will be discussed in later sections.

Applications of this technique are commonly used in the study of multilayer film structures, interdiffusion, impurity analysis of medium- and low-Z materials with high-Z impurities, oxidation of materials and accurate determination of stoichiometry.

RBS has, through the years, been divided into subsets defined by the maximum energy of the incident projectile: LEIS (maximum projectile energy = 10 keV), MEIS (50–500 keV), HEIS (many millions of electron-volts per atomic mass unit).

LEIS deals with analyzing surface phenomena, where surface means the first atomic layer or two; therefore, it is imperative that the analysis be performed under clean vacuum conditions, i.e. 10^{-8} Pa or better. During analysis the surface being analyzed is changed because the surface is damaged and implantation of the beam occurs; therefore LEIS is normally considered a destructive analysis procedure. Both composition and structure can be determined. LEIS has been used to make direct measurements of interatomic spacings and adsorption sites. Two types of energy analyzers generally are used for these measurements: electrostatic energy analyzers and time-of-flight (TOF) analyzers. Use of the electrostatic analyzer requires the entering particle to be charged, whereas TOF analyzers do not. For a more complete description of the technique and its applications, consult O'Connor and Macdonald,[9] O'Connor,[10] Niehus et al.[11] and Rabalais.[12]

MEIS and HEIS will be addressed together because the techniques and particle detection are similar. The method is nondestructive and provides elemental composition as a function of depth spanning from the first few tens of atomic layers up to a depth of tens of microns. The main advantages are the short analysis times and the direct and simple manner for analyzing the data. The technique is amenable to simple calibration procedures and yields quantitative results. Mainly inorganic compounds, such as multilayer thin films and bulk samples, are analyzed for composition variation or impurity distribution as a function of depth. The depth resolution typically can be a few tens of nanometers. Biological samples also can be analyzed but they are limited by radiation and thermal damage to the specimen. Elemental identification can be poor for high-Z target material due to small $\Delta m/m$ differences that lead to small differences in scattered incident beam energy. Also, the sensitivity for light element detection can be limited because of the low scattering cross-section for backscattered beam particles. Interpretation of data can be ambiguous if the sample is not homogenous.

Alpha particles, He^4, are commonly used as the incident particles, with energies up to a few million electron-volts. Protons are also used for thicker targets because

the stopping power for protons is significantly smaller than for He^4 and the penetration depths are therefore correspondingly greater, but the depth resolution and mass resolution for incident protons is degraded from that of He^4. The spectrum shown in Figure 2 is typical for HEIS. Data analysis programs are available that cannot only analyze experimental data but also predict spectra for specific target configurations.[1,2,13–16]

3.1.2 Elastic Recoil Detection

ERD is a technique devised to measure low-Z atoms in medium- to high-Z samples. In this technique the incident beam particles collide with the target and impart a significant fraction of their energy to low-mass atoms in the target; these exit the target and are detected and energy-analyzed. The angle of incidence θ, i.e. the angle of the incoming beam to the target surface normal, is nominally 75° and the scattering angle for detection of the low-Z components of the target, ϕ, is nominally 30°; see Figure 3 for a representation of the scattering geometry.

These reactions are elastic but are significantly enhanced over the cross-section expected for hard-sphere scattering, i.e. Rutherford scattering. The enhancement can be more than 1000, yielding cross-sections as big as a barn. One major problem with this technique is that elastic scattering of the incident beam particles from the high-Z substrate atoms is also quite large and can exceed the capabilities of the counting system. In order to facilitate use of the technique, the beam particle must have a mass greater than the low-Z atom to be detected, so that it imparts significant energy to the low-Z nucleus. By using a thin foil in front of the detector, the scattered beam particles can be stopped in the foil with the recoiling low-Z target nuclei passing through the foil, but with some energy loss. A detection system called a telescope detector can also enhance the usefulness of this technique; see Knoll[17] for a complete description. A spectrum of 11.48 MeV He^4 incident on a target of ErHD is shown in Figure 4. For this spectrum the angle of incidence is 75° and the scattering angle is 30°.

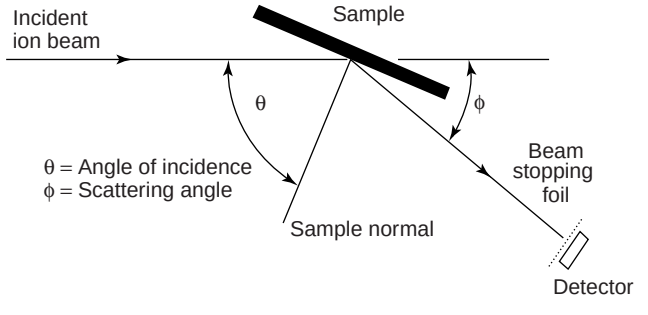

Figure 3 Schematic representation of the scattering geometry for the ERD technique.

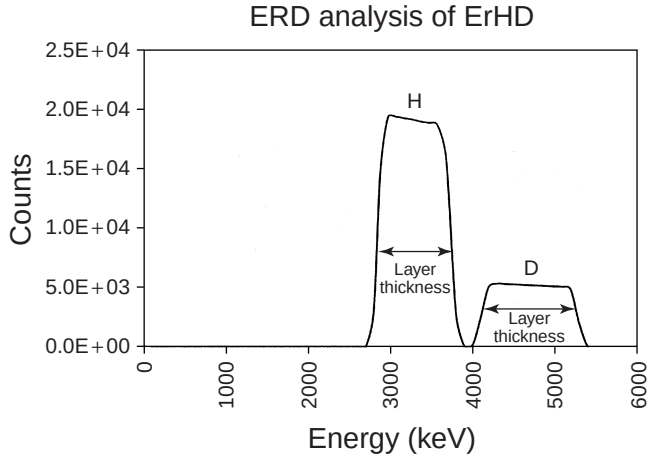

Figure 4 ERD spectrum for an 11.48-MeV He^4 ion beam incident on a target of ErHD. The angle of incidence is 75° and the recoil angle is 30°.

The peaks shown in Figure 4 result from recoil H and D from the target. For ERD analysis, depth resolution depends on many factors but it is generally larger than for RBS analysis. Excellent discussions of this technique are presented by Tesner and Nastasi[1] and Browning et al.[18]

3.2 Inelastic Scattering

3.2.1 Nuclear Reaction Analysis

NRA is based on nuclear reactions that are specific to atomic number and mass. This technique is limited to those elements that exhibit strong nuclear (resonant) reactions and have reactant particles or radiation that can be measured easily. In general, low-Z nuclei, e.g. H, D, T, He^3, He^4, incident on specific isotopes of low- to medium-Z nuclei (up to $Z = 30$) exhibit resonant reactions that are pertinent to this technique. Lists of the relevant reactions have been compiled and can be found in the work by Tesner and Nastasi.[1] It is important to point out that reverse reactions can be used, i.e. not only can the incident beam particles be the low-Z nuclei but they can also be the reactant low- to medium-Z nuclei. A typical reaction is $D(He^3,p)He^4$ with $Q = 18.352$ MeV. This reaction can be used to detect either D or He^3. A typical geometrical set-up for this analysis is similar to that shown in Figure 3, with θ and ϕ dependent on the particular reaction used. Figure 5 shows the spectrum for a 1.0-MeV He^3 beam incident on a sample of ErD_2 for an angle of incidence of 15° and a reaction product angle of 15°. The measured count from the He^3 elastically scattered from Er is quite large compared with the count from the nuclear reaction, but the nuclear reaction signal is measured easily because it is at a much higher energy. In many cases a thin foil

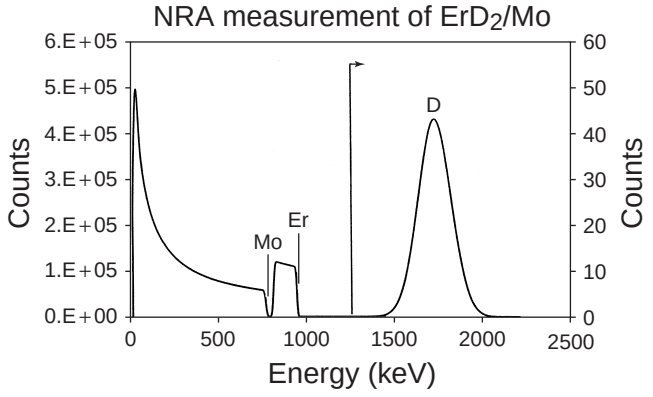

Figure 5 NRA spectrum for a 1.0-MeV He^3 beam incident on an ErD_2 sample. The angle of incidence is 15°, the reaction product angle is 15° and Q is 18.352 MeV.

is placed in front of the detector, which completely stops the elastically scattered particles.

There are many advantages to this method:

- isotope analysis is easy for those isotopes where nuclear reactions exist;
- there is no natural background;
- Usually the scattered beam particles are well separated from the reaction products.

There are also many drawbacks to the use of this method; the most outstanding are:

- the composition of the target usually cannot be measured during the same analysis;
- the spectra are sometimes difficult to interpret because peaks of different elemental isotopes can overlap;
- the cross-sections are less than for RBS, so much higher beam fluences must be used to obtain comparable counting statistics;
- usually there is no analytical form for the reaction cross-section.

An excellent description of the various uses of NRA is given by Tesner and Nastasi,[1] as well as the most useful nuclear reactions.

3.2.2 Particle-induced X-ray Emission

PIXE uses accelerated particles to ionize the atoms of a sample. The incident particles create vacancies in the inner electron shells of the target atoms as they traverse the target, and the electrons from the outer shells fill the inner shells, giving off characteristic X-rays and/or Auger electrons. The energy for each X-ray is characteristic of the energy difference between two electron states in the target atom, which allows identification of the atom originally ionized, and the specific energy released is called the X-ray line. There are many X-ray lines for each element and the number of X-ray lines per element increases with increasing atomic number. Sample analysis using PIXE is limited almost exclusively to ionization of the K- and L-shells. For excellent discussions on PIXE, see Mayer and Rimini[13] and Munnik.[19]

There are nonaccelerator-based techniques that also analyze materials using induced target X-rays: electron probe microanalysis (EPMA) and X-ray fluorescence (XRF). These two methods are also based on detection of the induced X-rays and only the ionization methods are different from PIXE. These two complementary techniques are covered in other areas of this publication.

For PIXE there are two methods to measure the energy of the X-rays: wavelength-dispersive analysis and energy-dispersive analysis. For the wavelenth-dispersive system, X-rays emitted from the sample are reflected from a Bragg crystal and detected using a proportional counter. The angle of scattering determines the energy of the X-rays and the energy resolution is partially determined by the solid angle of the detection system. This technique has superior energy resolution (2–40 eV) for X-ray energies below 20 keV. This analysis technique creates a limitation in that only one X-ray line can be measured for each setting of the Bragg crystal and a small solid angle is necessary for good resolution. These limitations decrease the efficiency of the system. For the energy-dispersive system the X-rays are detected using a solid-state detector, usually a lithium-drifted silicon detector. This system has the advantage that many X-ray lines can be measured using the same set-up with a relatively large detection solid angle so that the detection efficiency is high except for low-energy X-rays. The energy resolution for these types of detectors is about 160 eV. For most analyses the advantages of the energy-dispersive system outweigh the advantages of the wavelength-dispersive system.

From the X-ray intensities the amounts of the constituent atoms in the sample can be determined. Analysis of thin samples is relatively easy but for thick targets the analysis is considerably more complicated due to two effects: as the incident particles traverse the target they lose energy through electronic collisions and the ionization cross-section becomes a function of the sample depth; and the X-rays emitted in the sample have an increasing probability of being absorbed in the sample as the depth of X-ray emission increases. These effects are a function not only of depth in the sample but also of the constituent atoms within the sample matrix. Analysis of thick-target PIXE can be complicated; see Munnik[19].

Sensitivity for PIXE can reach 1 $\mu g\, g^{-1}$ with an absolute detection limit of 10^{-14} g and it has good sensitivity for elements such as S, P, Cl, K, and Ca. It is this sensitivity

that gives PIXE its great importance in biological and medical fields.

3.3 Channeling and Blocking

The basic idea of both channeling and blocking is best understood if a well-collimated beam of ions is envisaged as incident on a high-symmetry direction of a highly ordered crystal, usually a single crystal. Most of the ions propagate in the large ''channels'' between the target nuclei and lose only a small fraction of the energy in traversing the crystal compared with the energy that they would lose in traversing an amorphous target or nonaligned crystal. For foreign atoms located in these channels the interaction with the incident ions is very large compared with the native crystal atoms, so there is excellent discrimination even for low concentrations of the foreign atoms. And because the incident beam can be aimed down many different crystal axes, a very accurate measurement of the location of the foreign atom can be made (see, for example, Tesner and Nastasi,[1] Chu et al.[2] and Morgan[20]). In addition to the above application, amount and depth distribution of lattice disorder, especially surface disorder, and composition and thickness of amorphous surface layers can be determined.

3.4 Miscellaneous

There are three subjects that should be mentioned briefly: analysis with microbeams; external sample analysis; and analysis of polymer films.

For samples that are nonuniform laterally, e.g. grain boundaries, it is necessary to use beams of dimensions as small as 0.25 μm. As beam-focusing components get better, smaller beam dimensions are expected; see Doyle[21] for a complete discussion on the subject and Doyle[22] for applications.

Some types of samples cannot be placed in a vacuum and therefore must be analyzed in air. A technique has been devised to accomplish this analysis, known as external beam analysis (EBA). A review of this technique and some of its applications are given by Doyle et al.[23]

IBA of polymers can be extremely difficult. This subject has been addressed in a review paper by Green and Doyle,[24] which discusses the techniques, problem areas, and areas of application.

4 DATA ANALYSIS TECHNIQUES, DATA SOURCES, SUMMARY AND FUTURE DIRECTIONS

There are numerous analysis techniques that apply to very specific applications but there are only two generalized analysis programs that are commonly used for the IBA of experimental data. These two programs also can be used to predict a particular spectrum for known input parameters. The programs are SIMNRA (Version 4.0) and RUMP (Version 4.0); see Mayer[25] and Doolittle[26] for information on acquiring these programs. There are also data sets available on stopping power, range, and energy-loss partition that are necessary for the analysis programs above; the two most prominent data sets are SRIM-2000 (V.09) and TRIM.SP (see Ziegler et al.[27] and Eckstein,[28] respectively, for information on acquiring these data sets).

This article has dealt with the major aspects in the use of energetic ion beams for the analysis of solid samples, although there are new programs to analyze both liquids and biological samples. Many different approaches have been pointed out for the analysis of sample thickness, composition and state of crystallinity. Specific techniques have been given for the analysis of hydrogen isotopes and isotopes of many other light- to medium-mass elements. Major applications have been given for each technique discussed, and advantages and limitations of the technique have been presented. Data analysis programs have been presented, as well as the data sets needed for the analysis.

The future of IBA appears to grow both in the number of laboratories available for performing IBA and in the number of industries that demonstrate a need for IBA. As existing equipment becomes more sensitive and new equipment comes to market, more techniques are expected to evolve, thereby expanding the breadth and usefulness of IBA.

ABBREVIATIONS AND ACRONYMS

EBA	External Beam Analysis
EPMA	Electron Probe Microanalysis
ERD	Elastic Recoil Detection
HEIS	High-energy Ion Scattering
IBA	Ion Beam Analysis
ISS	Ion Scattering Spectroscopy
LEIS	Low-energy Ion Scattering
MEIS	Medium-energy Ion Scattering
NRA	Nuclear Reaction Analysis
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy
SIMS	Secondary Ion Mass Spectrometry
TOF	Time-of-Flight
XRF	X-ray Fluorescence

RELATED ARTICLES

Surfaces **(Volume 10)**
Surfaces: Introduction • Electron Microscopy and Scanning Microanalysis • Scanning Probe Microscopy, Industrial Applications of

Mass Spectrometry **(Volume 13)**
Secondary Ion Mass Spectrometry as Related to Surface Analysis

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Elastic Recoil Detection Analysis • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission) • Rutherford Backscattering Spectroscopy

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis

REFERENCES

1. J.R. Tesner, M. Nastasi (eds.), *Handbook of Modern Ion Beam Materials Analysis*, MRS, 1995.
2. W.-K. Chu, J.W. Mayer, M.-A. Nicolet, *Backscattering Spectrometry*, Academic Press, New York, 1978.
3. D. Cockcroft, E.T.S. Walton, 'Experiments with High Velocity Positive Ions: I. Further Developments in the Method of Obtaining High Velocity Positive Ions', *Proc. R. Soc. A*, **136**, 619–630 (1932).
4. R.J. van de Graaff, 'A 1,500,000 Volt Electrostatic Generator', *Phys. Rev.*, **38**, 1919–1920 (1931).
5. M.S. Livingston, 'Early History of Particle Accelerators', *Adv. Electronics Electron Phys.*, **50**, 1–88 (1980).
6. G.F. Knoll, *Radiation Detection and Measurement*, John Wiley and Sons, New York, 1987.
7. E. Rutherford, 'The Scattering of a and b Particles by Matter and the Structure of the Atom', *Philos. Mag.*, **21**, 669–688 (1911).
8. H. Geiger, E. Marsden, 'The Laws of Deflexion of a Particles Through Large Angles', *Philos. Mag.*, **25**, 606–623 (1913).
9. D.J. O'Connor, R.J. Macdonald, 'Low Energy Ion Scattering from Surfaces', in *Ion Beams for Materials Analysis*, eds. J.R. Bird, J.S. Williams, Academic Press, Sydney, 373–414, 1989.
10. D.J. O'Connor, 'Low Energy Ion Scattering', in *Surface Analysis Methods in Material Science*, eds. D.J. O'Connor, B.A. Sexton, R.St.C. Smart, Springer-Verlag, Berlin, 245–262, 1992.
11. H. Niehus, W. Heiland, E. Taglauer, 'Low-energy Ion Scattering at Surfaces', *Surf. Sci. Rep.*, **17**, 213–303 (1993).
12. J.W. Rabalais, 'Low Energy Ion Scattering and Recoiling', *Surf. Sci.*, **299/300**, 219–232 (1994).
13. J.W. Mayer, E. Rimini, *Ion Beam Handbook for Material Analysis*, Academic Press, New York, 1977.
14. L.C. Feldman, 'High Energy Scattering', *Surf. Sci.*, **299/300**, 233–245 (1994).
15. J.P.F. Sellschop, S.H. Connell, 'Elemental Isotopic and Structural Analysis Using Ion Beams', *Nucl. Instrum. Methods Phys. Res.*, **B85**, 1–19 (1994).
16. S.H. Sie, 'Rutherford Backscattering Spectrometry and Nuclear Reaction Analysis', in *Surface Analysis Methods in Material Science*, eds. D.J. O'Connor, B.A. Sexton, R.St.C. Smart, Springer-Verlag, Berlin, 203–219, 1992.
17. *Radiation Detection and Measurement*, ed. G.F. Knoll, John Wiley & Sons, New York, 380–382, 1987.
18. J.F. Browning, R.A. Langley, B. Doyle, J.C. Banks, W.R. Wampler, 'High Accuracy High Energy He ERD Analysis of H, D, and T', *Nucl. Instrum. Methods Phys. Res. B*, (2000).
19. F. Munnik, 'Accurate Evaluation of Absolute Calibration in Thick-target PIXE', *Thesis*, Department of Physics, N-lagg, Eindhoven University of Technology, 1994.
20. D.V. Morgan (ed.), *Channeling*, John Wiley & Sons, New York, 1973.
21. B.L. Doyle, 'Multidimensional Ion Microbeam Analysis', *Nucl. Instrum. Methods B*, **15**, 654–660 (1986).
22. B.L. Doyle, 'Nuclear-microprobe Analysis: Applications', in *Microbeam Analysis*, eds. A.D. Romig, Jr, W.F. Chambers, San Francisco Press, San Francisco, 15–21, 1986.
23. B.L. Doyle, D.S. Walsh, S.R. Lee, 'External Micro-ion-beam Analysis', *Nucl. Instrum. Methods B*, **54**, 244–257 (1991).
24. P.F. Green, B.L. Doyle, 'Ion Beam Analysis of Thin Polymer Films', in *New Characterization Techniques for Thin Polymer Films*, eds. H.-M. Tong, L.T. Nguyen, John Wiley and Sons, New York, 139–179, 1990.
25. SIMNRA (V.4.0) can be obtained from M. Mayer at mam@ipp.mpg.de.
26. RUMP (V.4.0), L.R. Doolittle, 'A Semiautomatic Algorithm for Rutherford Backscattering Analysis', *Nucl. Instrum. Methods Phys.*, **B15**, 227–231 (1986).
27. SRIM-2000 (V.09) can be obtained from J.F. Ziegler at ziegler@watson.ibm.com; J.F. Ziegler, J.P. Biersack, U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, New York, 1999.
28. W. Eckstein, *Computer Simulation of Ion–Solid Interactions*, Springer-Verlag, Berlin, 1991.

Photoluminescence in Analysis of Surfaces and Interfaces

Timothy H. Gfroerer
Davidson College, Davidson, USA

1 Introduction **1**
2 Photoluminescence Excitation **3**
2.1 Excitation Energy: Penetration Depth and Stoke's Shift 3
2.2 Excitation Intensity: Interface State Density and Distribution 4
3 Photoluminescence Spectrum **7**
3.1 Photoluminescence Peak Positions: Energy Levels 8
3.2 Photoluminescence Line Width and Splitting: Alloy Disorder and Interface Roughness 10
4 Photoluminescence Intensity **12**
4.1 Dependence on Applied Bias: Surface Potential 13
4.2 Spatial Dependence: Interface Uniformity and Carrier Diffusion 14
4.3 Time Dependence: Recombination Rates 16
4.4 Temperature Dependence: Thermal Population 18
5 Photoluminescence Polarization **19**
6 Conclusion **20**
Abbreviations and Acronyms **20**
Related Articles **21**
References **21**

Photoluminescence (PL) is the spontaneous emission of light from a material under optical excitation. The excitation energy and intensity are chosen to probe different regions and excitation concentrations in the sample. PL investigations can be used to characterize a variety of material parameters. PL spectroscopy provides electrical (as opposed to mechanical) characterization, and it is a selective and extremely sensitive probe of discrete electronic states. Features of the emission spectrum can be used to identify surface, interface, and impurity levels and to gauge alloy disorder and interface roughness. The intensity of the PL signal provides information on the quality of surfaces and interfaces. Under pulsed excitation, the transient PL intensity yields the lifetime of nonequilibrium interface and bulk states. Variation of the PL intensity under an applied bias can be used to map the electric field at the surface of a sample. In addition, thermally activated processes cause changes in PL intensity with temperature.

PL analysis is nondestructive. Indeed, the technique requires very little sample manipulation or environmental control. Because the sample is excited optically, electrical contacts and junctions are unnecessary and high-resistivity materials pose no practical difficulty. In addition, time-resolved PL can be very fast, making it useful for characterizing the most rapid processes in a material. The fundamental limitation of PL analysis is its reliance on radiative events. Materials with poor radiative efficiency, such as low-quality indirect bandgap semiconductors, are difficult to study via ordinary PL. Similarly, identification of impurity and defect states depends on their optical activity. Although PL is a very sensitive probe of radiative levels, one must rely on secondary evidence to study states that couple weakly with light.

1 INTRODUCTION

Multilayer material systems are increasingly important in the development of smaller, faster, and more efficient electronic and optoelectronic devices. The primary motivation for using multilayer structures is to change the potential energy of electrons and holes at the material interfaces. Because phenomena at surfaces and interfaces tend to dominate the behavior of excitations in these heterostructures, the performance of many microelectronic devices is limited by the nature of heterojunctions. Smooth and atomically abrupt interfaces are necessary for good optical and electrical reflection, uniform quantum confinement, and high carrier mobility. Even more importantly, defects and impurities at interfaces provide new states for electrons and holes, altering their motion, lifetime, and transition energies.

When light of sufficient energy is incident on a material, photons are absorbed and electronic excitations are created. Eventually, these excitations relax and the electrons return to the ground state. If radiative relaxation occurs, the emitted light is called PL. This light can be collected and analyzed to yield a wealth of information about the photoexcited material. The PL spectrum provides the transition energies, which can be used to determine electronic energy levels. The PL intensity gives a measure of the relative rates of radiative and nonradiative recombination. Variation of the PL intensity with external parameters like temperature and applied voltage can be used to characterize further the underlying electronic states and bands.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

PL depends on the nature of the optical excitation. The excitation energy selects the initial photoexcited state and governs the penetration depth of the incident light. The PL signal often depends on the density of photoexcited electrons, and the intensity of the incident beam can be adjusted to control this parameter. When the type or quality of material under investigation varies spatially, the PL signal will change with excitation position. In addition, pulsed optical excitation provides a powerful means for studying transient phenomena. Short laser pulses produce virtually instantaneous excited populations, after which the PL signal can be monitored to determine recombination rates.

Because PL often originates near the surface of a material, PL analysis is an important tool in the characterization of surfaces. The utility of PL for this purpose is derived from its unique sensitivity to discrete electronic states, many of which lie near surfaces and interfaces. Using the techniques noted above, the nature of these states can be probed in detail. The energy distribution and density of interface states can be ascertained by studying the excitation intensity dependence of the PL spectrum. The presence of surface adsorbates alters the intensity of the PL signal. When the states serve as long-lived traps, the depth of the trap can be determined by observing thermal activation in temperature-dependent PL. In fact, even if interface states are nonradiative, which is usually the case, the states alter the time-resolved PL of radiative transitions in the material. Nonradiative traps dominate the transient PL signal at low carrier density.

PL is simple, versatile, and nondestructive. The instrumentation that is required for ordinary PL work is modest: an optical source and an optical power meter or spectrophotometer. A typical PL set-up is shown in Figure 1. Because the measurement does not rely on electrical excitation or detection, sample preparation is minimal. This feature makes PL particularly attractive for material systems having poor conductivity or undeveloped contact/junction technology. Measuring the continuous wave PL intensity and spectrum is quick and straightforward. On the other hand, investigating transient PL is more challenging, especially if recombination processes are fast. Instrumentation for time-resolved detection, such as single photon counting, can be expensive and complex. Even so, PL is one of the only techniques available for studying fast transient behavior in materials.

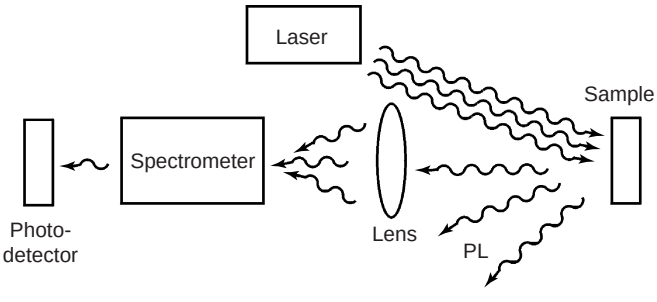

Figure 1 Typical experimental set-up for PL measurements.

Because PL can be used to study virtually any surface in any environment, it can be used to monitor changes induced by surface modification in real time. For example, unlike most surface characterization techniques, PL is generally not sensitive to the pressure in the sample chamber. Hence, it can be used to study surface properties in relatively high-pressure semiconductor growth reactors. Although PL does depend quite strongly on temperature, liquid helium temperatures being required for the highest spectral resolution, room-temperature measurements are sufficient for many purposes. In addition, PL has little effect on the surface under investigation. Photoinduced changes and sample heating are possible, but low excitation can minimize these effects. In situ PL measurements do require optical access to the sample chamber. Yet, compared with other optical methods of characterization like reflection and absorption, PL is less stringent about beam alignment, surface flatness, and sample thickness.

The advantages of PL analysis listed above derive from the simplicity of optical measurements and the power to probe fundamental electronic properties. The chief drawback of PL analysis also follows from the reliance on optical techniques: the sample under investigation must emit light. Indirect-bandgap semiconductors, where the conduction band minimum is separated from the valence band maximum in momentum space, have inherently low PL efficiency. Nonradiative recombination tends to dominate the relaxation of excited populations in these materials. This problem can be augmented by poor surface quality, where rapid nonradiative events may occur. Nevertheless, once a PL signal is detected, it can be used to characterize both radiative and nonradiative mechanisms. Although it may not be possible to identify directly the nonradiative traps via PL, their signature is evident in several types of PL measurements.

Another shortcoming of PL is the difficulty in estimating the density of interface and impurity states. When these states have radiative levels, they are readily identified in the PL spectrum, and the size of the associated PL peaks provides a relative measure of their presence in the sample. However, measuring the absolute density of these states is a far more formidable task and typically requires an exhaustive analysis of the excitation intensity dependence of the PL signal.

2 PHOTOLUMINESCENCE EXCITATION

The choice of excitation is critical in any PL measurement. The excitation energy and intensity will have profound effects on the PL signal. Although the excitation conditions must be considered carefully, the strength of the PL technique relies heavily on the flexibility that these adjustable parameters provide. Because the absorption of most materials depends on energy, the penetration depth of the incident light will depend on the excitation wavelength. Hence, different excitation energies probe different regions of the sample. The excitation energy also selects the initial excited state in the experiment. Because lasers are monochromatic, intense, and readily focused, they are the instruments of choice for photoluminescence excitation (PLE). For many applications, the excitation energy is not critical. Here, a relatively inexpensive HeNe or diode laser will often satisfy the basic requirement of light exceeding the bandgap energy. In more demanding experiments, the laser is chosen carefully to probe a particular depth or to excite a particular species in the sample.

Unlike the excitation energy, which may or may not be important, the excitation intensity will influence the result of any PL experiment. The excitation intensity controls the density of photoexcited electrons and holes, which governs the behavior of these carriers. Each electron–hole recombination mechanism has a distinct functional dependence on carrier density. For example, the number of interface and impurity states is finite, and recombination at these sites will saturate at high excitation. In addition, the photoexcited carriers themselves can alter the distribution of interface states. Thus, the excitation intensity must be calibrated accurately and controlled precisely.

2.1 Excitation Energy: Penetration Depth and Stoke's Shift

The absorption of a material depends strongly on the energy of the incident light. In the analysis of surfaces, it is useful to describe this wavelength-dependent absorption by a penetration depth (the inverse of the absorption coefficient), which is a measure of the thickness of the layer that is probed. In direct-bandgap semiconductors, above-bandgap excitation has a penetration depth of the order of 1 μm. The diffusion of photoexcited carriers can vary widely but is typically in the range 1–10 μm. Hence, PL with above-gap excitation is very sensitive to surface effects. In indirect-bandgap semiconductors, or direct-gap semiconductors with sub-bandgap excitation, absorption is weaker and the light penetrates deeper into the sample. Here, the PL is dominated by bulk recombination. If multiple excitation wavelengths are available, this property of PL can be used to distinguish surface and bulk contributions.

Garcia-Garcia et al. have used such an approach to study the surface layer of chemically etched CdTe.[1] When the surface structure is poor, the PL spectrum changes with increasing excitation energy, the strength of defect-related lines increasing relative to bulk features. Correlation of these changes with penetration depth indicates the location of near-surface defects. Furthermore, the authors found that the PL spectrum of a high-quality surface is independent of excitation energy. In an extension of this technique, Wang et al. used the widely divergent penetration of one and two photon absorption processes to distinguish surface and bulk effects.[2] In contrast to relatively slow recombination with two-photon excitation, a fast recombination mechanism dominated the one-photon excitation experiment. This result indicates the presence of rapid recombination sites near the surface of the semiconductor, where the single-photon excitation is absorbed.

Dangling bonds at a semiconductor surface or interface give rise to electronic states within the bandgap. These mid-gap states fill up to the Fermi level with electrons that originate in the bulk of the material. The accumulation of charge at the surface creates an electric field – a depletion region – that leads to bending of the valence and conduction band edges. According to a simple dead-layer model,[3] electron–hole pairs that are generated in this region are swept apart by the electric field, prohibiting radiative recombination. The phenomenon is described quantitatively by a depletion thickness that characterizes this ''dead layer''.

Although the dead-layer picture of carrier kinetics at the surface of a semiconductor overlooks the complicated role of surface states in the recombination process, it is useful when PL quenching has an electrostatic origin. For example, Ryswyk and Ellis have studied the effect of I_2 on n-GaAs in the context of this model.[4] They used PL depth profiling with multiple excitation wavelengths to determine the depletion thickness in the I_2-treated sample. Uosaki et al.[5] have used a similar experimental approach to characterize p-GaAs/electrolyte interfaces. Results were analyzed within the context of the more sophisticated model proposed by Mettler,[6] which accounts for changes in charge separation *and* surface state recombination. It should be noted that these phenomena are intricately connected, and resolving their contributions to PL quenching is difficult. Although changes in the electric field at the surface should not affect the surface state density, altering the nature or number of surface states can affect depletion and band bending. This point will be discussed in more detail in section 4.

In addition to controlling the penetration depth, the excitation energy selects the initial photoexcited state in the sample. This state is short-lived because thermalization of photoexcited carriers via phonon emission is rapid. Relaxation to within kT of the lowest available states is usually orders of magnitude faster than recombination. Thus, ordinary PL emission only reveals the lowest energy states. Nevertheless, it is possible to obtain spectral information on excited states by collecting the low-energy emission while tuning the excitation energy across higher energy states. The PL signal is enhanced when the excitation energy is resonant with an excited state. The technique, called photoluminescence excitation spectroscopy, is similar to ordinary absorption subject to the condition that there exists a relaxation channel from the excited state to the emission state being monitored. PLE is frequently used to study epilayers on opaque substrates because absorption is not feasible in this case.

Compared to band continua, discrete interface states are relatively sparse. PL is easily dominated by interface states because such states tend to have levels below the intrinsic bands, and carriers are remarkably efficient at finding the lowest available states. However, PLE relies on optical absorption, which is relatively weak for isolated interface sites. Hence, it is usually difficult to uncover interface states in PLE spectra. Nevertheless, PLE has been applied successfully to the characterization of certain surface species. For example, Zhu et al. utilized PLE to identify groups chemisorbed on the surface of TiO_2 ultrafine particles.[7] Similarly, Anpo et al. used the dependence of PL on excitation energy to resolve differences between anchored and impregnated Mo/SiO_2 catalysts.[8]

An important feature of optical transitions in semiconductor quantum wells (QWs) is the energy difference between the absorption (or PLE) and emission peaks – Stoke's shift. This shift is a measure of interface fluctuations because it indicates that the thickness of the QW varies with position. Carriers are photoexcited uniformly in the QW but they diffuse to regions of larger well width, where the confinement energy is smaller, before recombining. Deveaud et al. have studied the PL and PLE spectra of a series of multiple QW samples.[9] In samples prepared under optimal growth conditions, they observed two sharp PL peaks, which they attribute to recombination in regions where the QW thickness differs by exactly one atomic monolayer. PLE is used to test whether excitation at the higher energy peak yields emission from the lower energy peak and to confirm that thermalization between the two regions occurs. It has been argued that splittings in PL from QWs like those discussed above can be misleading because impurities can produce similar effects.[10] The ambiguity can be resolved by using PLE where impurity transitions are weak. The structure in PL spectra from QWs will be discussed in more detail in section 3.2.

2.2 Excitation Intensity: Interface State Density and Distribution

The intensity of the incident light controls a critical property of the PL experiment: the density of photoexcited electrons and holes. When the carrier density is low, the measurement is dominated by discrete defect and impurity sites at the interfaces and within the bulk of the material. Recombination at these energetically favored sites is frequently referred to as Shockley–Read–Hall (SRH) recombination, in recognition of early work on the statistics of this phenomenon. The SRH rate is proportional to the dominant carrier density n. The dominant carrier density is the greater of two quantities: the equilibrium carrier concentration (i.e. the doping level) and the photoexcited carrier concentration. Under intermediate excitation, the discrete states are filled and bulk radiative recombination plays a greater role. Neglecting excitonic effects, these transitions are two-body events between free electrons and holes, so the radiative rate varies as n^2. It should be noted that low temperatures or quantum confinement can make Coulomb binding of electron–hole pairs energetically favorable, leading to the formation of excitons that recombine radiatively at a rate proportional to n. At the highest carrier densities, three-body Auger recombination becomes important because the rate of this mechanism varies as n^3. In the analysis of interfaces, the transition between SRH and radiative recombination is the most important because it depends on the density of defect and impurity sites.

In systems with high interface-to-volume ratios and active layers that are thin relative to the carrier diffusion length, interface effects dominate SRH recombination. In planar structures, the interface-to-volume dependence of SRH recombination can be separated from actual interface phenomena via the interface recombination velocity S. The interface recombination rate is defined as $R_s = (S/d)n$, where d is the thickness of the optically active layer. When two interfaces are present, as in a double heterostructure, the two interface contributions are combined: S is replaced by $S_1 + S_2$ (or $2S$ for equivalent interfaces). Because S is independent of carrier density and layer thickness, it can be used to compare interface recombination in a variety of materials. S depends strongly on the nature of the interface. For example, S is of the order of $10^7\,\mathrm{cm\,s^{-1}}$ at the interface of GaAs and air, whereas high-quality GaAs/GaInP interfaces can have $S < 10\,\mathrm{cm\,s^{-1}}$.[11]

This reduction of the interface recombination velocity at the edge of a material is called surface passivation.

Because the performance of many optoelectronic devices is limited by interface recombination, interface passivation has been studied extensively. Passivation techniques include epitaxial growth of lattice-matched semiconductor alloys and a variety of chemical surface treatments. Qualitative evaluation of these techniques is straightforward: an increase in the PL signal indicates a decrease in nonradiative interface recombination. Quantitative information on S requires more sophisticated techniques, many of which involve investigations of the excitation intensity dependence of PL.

Komiya et al. measured the excitation-intensity-dependent PL signal from a series of InGaAsP/InP double heterostructures.[12] Although the PL signal was usually linear with excitation power, some samples showed superlinear dependence under intermediate excitation (see Figure 2), indicating a transition between nonradiative and radiative recombination regimes. The lack of such a transition in the other samples suggests that radiative recombination dominated at all excitation levels, which gave an upper limit for S. In a very different approach, Mullenborn and Haegel studied single AlGaAs/GaAs and GaInP/GaAs heterostructures where the buried GaAs layer is excited indirectly via either carrier diffusion or recycling (i.e. reabsorption and emission) of photons.[13] The excitation intensity dependence of PL from both layers was used to distinguish between the two mechanisms for carrier generation in the buried layer. Because the two processes are competing, and react differently to interface quality, their relative strengths can be modeled to characterize interface recombination.

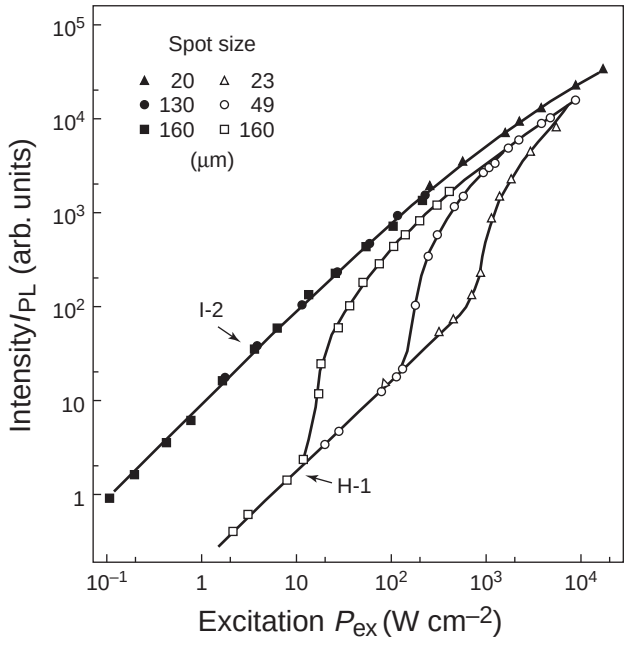

Figure 2 Excitation power dependence of the PL intensity for two different samples. The I-2 PL signal is approximately linear with excitation, whereas the H-1 signal increases superlinearly in the intermediate excitation regime.[12]

The experiments described above combine relative measurements of the PL intensity with considerable assumptions or modeling to obtain quantitative information on semiconductor interfaces. Absolute measurements of the PL intensity provide for a more direct analysis of interfaces. Evaluating the absolute PL intensity requires an estimate of what fraction of the emitted light is collected by the detector. However, the PL emission profile is a complicated function of sample geometry, refraction, internal reflection, and reabsorption. Hence, large uncertainties in this estimate are inevitable. Integrating spheres can be used to ensure that most of the emission is collected, but it is difficult to control environmental parameters such as temperature in these devices.

A very accurate alternative technique for measuring the absolute PL intensity from light-emitting materials with unspecified emission profiles has been developed.[14] The sample can be studied in a controlled environment, and the measurement does not require knowledge of the fraction of emission collected. Determination of the absolute PL intensity relies on the fact that simultaneous relative measurements of radiative and nonradiative recombination can be combined to determine the absolute radiative efficiency (the ratio of photons emitted to photons absorbed). Because the excitation power must be converted into either light or heat, an increase in one must be accompanied by a proportional decrease in the other. The intensity of the emitted light and the temperature change are measured as a function of excitation power to evaluate the relative strength of the radiative and nonradiative mechanisms. Then, extrapolating the temperature change to zero heating, the light signal equivalent to 100% radiative efficiency is obtained.

This technique has been used to measure the absolute radiative efficiency of an InGaAs/InP double heterostructure with an integral substrate reflector.[14] A plot of the normalized light signal vs. temperature change is shown in Figure 3. At low excitation, the relative intensity of the emitted light increases, and the relative temperature change decreases with increasing excitation. However, at high excitation the radiative efficiency rolls off (see Figure 4), and the data begin to show the opposite trend. As predicted by the model described above, the points fall on a straight line and do not depend on whether the efficiency is increasing or decreasing with excitation. Extrapolation to zero heat is not shown in Figure 3 because a false origin is required to give sufficient detail. Nevertheless, the y-intercept of the linear fit in Figure 3 can be combined with the excitation and average emission energies to calibrate the relative light signal data. The result of this calibration is presented in Figure 4.

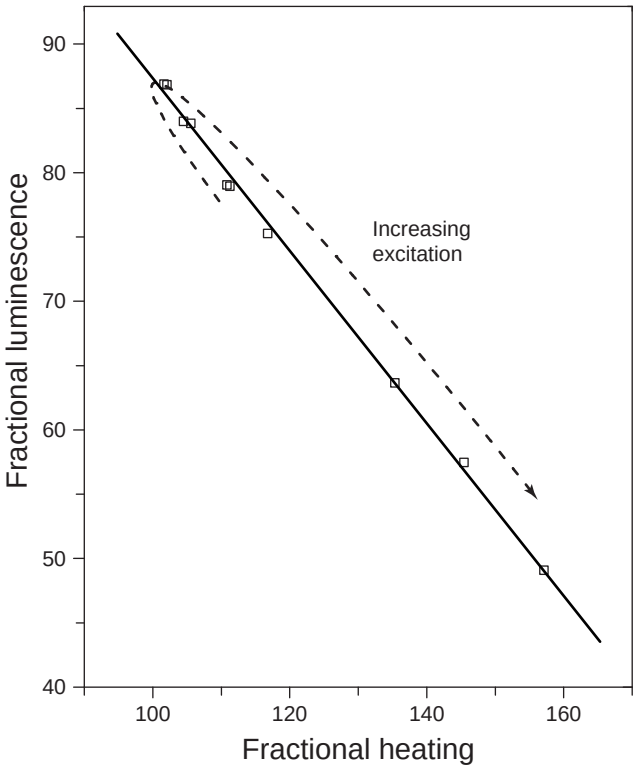

Figure 3 Relative measurements of luminescence signal and temperature change, normalized by the excitation intensity and plotted against each other. The dotted line indicates how the result evolves as the excitation power is increased.[14]

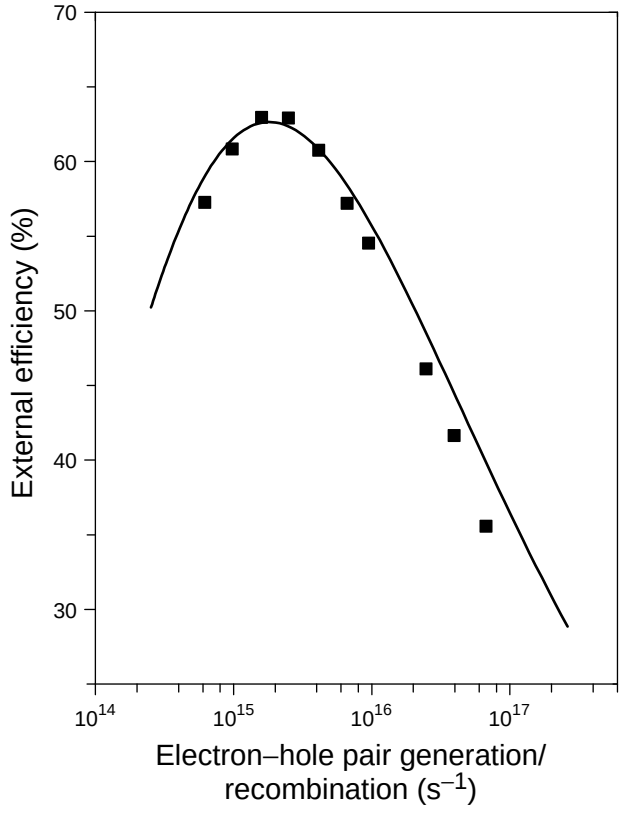

Figure 4 External radiative quantum efficiency vs. the rate of electron–hole pair generation and recombination in steady state. The solid line is a fit to the data that accounts for interface, radiative, and Auger recombination events.[14]

The fit in Figure 4 relies on the fact that, in steady state, the rate of carrier generation equals the rate of recombination, as shown in Equation (1):

$$\frac{P_{\text{abs}}}{E_{\text{ex}}} = V\left(\frac{2S}{d}n + \frac{B}{N}n^2 + Cn^3 \right. \tag{1}$$

Here, P_{abs} is the absorbed laser power, E_{ex} is the laser energy, and V is the volume. The photon recycling factor N is the average number of radiative recombination events required for a photon to escape the semiconductor, and B and C are parameters that describe the rates of radiative and Auger recombination respectively. Because the fit is very sensitive to the coefficients that characterize the different mechanisms for recombination, it can be used to evaluate parameters such as the interface recombination velocity S.

Saitoh et al. have argued that the simple picture of an excitation-independent surface recombination velocity is not sufficient to describe the excitation dependence of the PL signal.[15] In particular, when the excitation is intermediate between the SRH and radiative recombination-dominated regimes, the quasi-Fermi levels for electrons and holes split away from the low-excitation pinned position and move toward their respective band edges. The quasi-Fermi levels determine the energy range of surface states that contribute to recombination. Thus, the PL efficiency in this transition region is very sensitive to the density and energy distribution of surface states. The authors have developed an elaborate model of carrier behavior near the surface of the semiconductor that accounts for carrier transport, surface state occupation, band bending, and recombination. Analysis of the excitation-intensity-dependent PL signal in the context of this model is referred to as photoluminescence surface-state spectroscopy (PLS^3).

The shape of the radiative efficiency curve in the SRH to radiative recombination transition region depends on the distribution of surface states that participate in recombination. If the distribution consists of a discrete set of states with similar energetic positions within the gap, complete saturation of the SRH recombination process will lead to a PL efficiency slope of unity. Efficiency curves with slopes less than unity indicate that the surface-state distribution is continuous and that an increasing number of surface states are participating in recombination as the quasi-Fermi levels move through the bandgap. Hence, variation in the slope of the efficiency curve with

increasing excitation can be used to map the surface-state distribution within the gap. Two distinct surface-state distributions and their corresponding efficiency curves are shown in Figure 5.[16] Since its development in 1991, PLS[3] has been applied to a wide number of semiconductor systems. Although some of these studies have uncovered surface states with discrete energies, most have revealed that the surface-state distribution is U-shaped, with increasing density near the band edges.

Excitation-intensity-dependent PL can also be used to study the properties of QWs. Ding et al. measured the excitation intensity I_{ex} dependence of the integrated PL signal from asymmetric coupled QWs.[17] They note that the PL intensity should vary as I_{ex}^2 when recombination is dominated by nearly saturated interface traps, but should be proportional to I_{ex} when radiative recombination is more important. This behavior was clearly observed in their results. They used the transition intensity to estimate the nonradiative decay time, which suggested that traps were located at the interface of the coupling barrier and the larger well. Excitation-dependent saturation of PL can also be used to determine the relative density of different radiative states. For example, when splittings are observed in the PL from QWs, as described in section 2.1, the lower energy peak often saturates with increasing excitation because of state filling. In this case, the saturation intensity can be used to estimate the size of interface islands in the QW where the lower energy PL is believed to originate.[18]

A third application of excitation-intensity-dependent PL is in the evaluation of interface band alignment. Vignaud et al. observed a dramatic excitation-dependent blueshift in the PL from InAlAs/InP heterostructures. They attributed this blueshift to spatially indirect type II

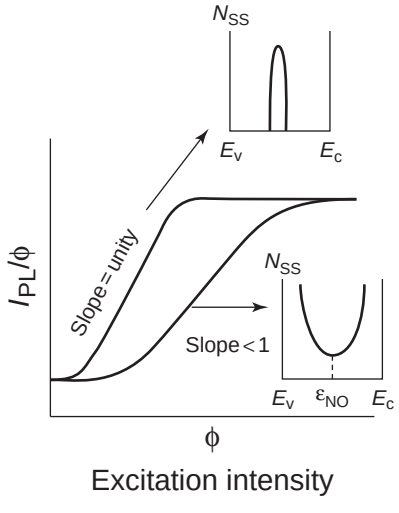

Figure 5 Relative PL efficiency I_{PL}/Φ vs. excitation intensity Φ for discrete and U-shaped surface-state distributions.[16]

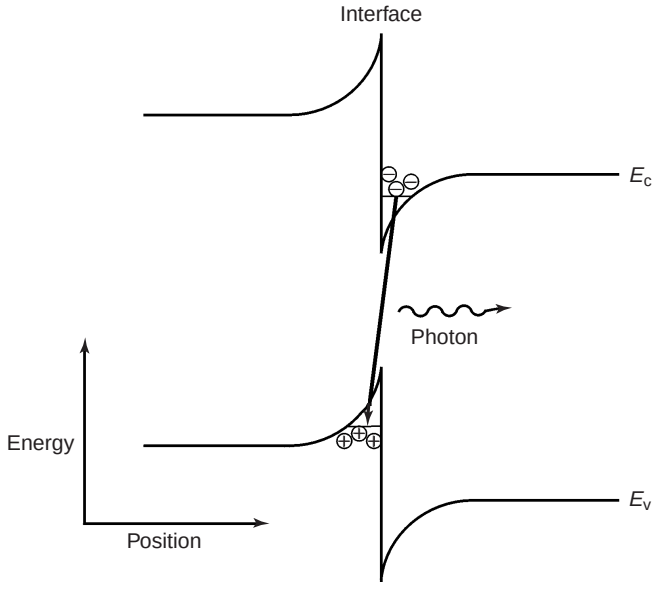

Figure 6 Type II band alignment where electrons and holes collect on opposite sides of a semiconductor interface. The small QWs that form near the interface lead to a substantial excitation-dependent blueshift in the PL.

transitions across the interface.[19] The blueshift is explained as follows: when the conduction band minimum and valence band maximum occur on opposite sides of a semiconductor interface, electrons and holes are trapped in different regions of the heterostructure. Because the electrons and holes are attracted to each other, narrow QWs are formed adjacent to the interface, as shown in Figure 6. As the excitation is increased, carriers accumulate in these wells, increasing the confinement energy of the wells and the dipole energy across the interface. Both of these mechanisms lead to substantial blueshifts in the PL. It is interesting to note that type II transitions, which occur across the interface itself, are especially sensitive to interface quality.

3 PHOTOLUMINESCENCE SPECTRUM

Optical transitions provide direct access to the energy level structure of a system. Photons of a particular energy that are absorbed or emitted by a sample provide evidence of electronic states differing by that energy within the material. Absorption is a good probe of the overall band structure of a system because bands have a relatively high density of states. PL emission, on the other hand, tends to favor sparse low-lying states because photoexcited carriers rapidly thermalize through bands and closely spaced states to within kT of the lowest available levels. This feature of PL makes it particularly effective in the analysis of interfaces where discrete defect and impurity states

abound. If the state is radiative, it will generate unique peaks in the PL spectrum. Thus, the PL measurement is a very sensitive and selective probe of such states.

In addition to identifying discrete states, PL peak positions can be used to evaluate the composition of semiconductor alloys. Here, the energy of the band-edge emission is correlated with the composition-dependent bandgap of the alloy. This application is also useful in interface analysis where atomic interdiffusion leads to interface alloying. Interface alloys can form narrow wells or barriers that have an impact on the behavior of carriers at heterojunctions. Evidence of interface wells and barriers is often found in the PL spectrum.

Other features of the PL spectrum, including line widths and splittings, provide important information on QW interfaces. These systems are very sensitive to interface roughness because fluctuations as small as one atomic monolayer can alter the carrier confinement energy considerably. When the QW thickness varies substantially within the sampled region, a general broadening of the PL spectrum is observed. In samples with higher quality interfaces, variation in the QW thickness is limited, and recombination in different regions yields sharp, well-resolved peaks in the PL.

3.1 Photoluminescence Peak Positions: Energy Levels

In the bulk of a crystalline material, translational symmetry leads to the formation of electronic energy bands. Defects and impurities break the periodicity of the lattice and perturb the band structure locally. The perturbation usually can be characterized by a discrete energy level that lies within the bandgap. Depending on the defect or impurity, the state acts as a donor or acceptor of excess electrons in the crystal. Electrons or holes are attracted to the excess or deficiency of local charge due to the impurity nucleus or defect, and Coulomb binding occurs. The situation can be modeled as a hydrogenic system where the binding energy is reduced by the dielectric constant of the material. Because electrons and holes have different effective masses, donors and acceptors have different binding energies.

When the temperature is sufficiently low, carriers will be trapped at these states. If these carriers recombine radiatively, the energy of the emitted light can be analyzed to determine the energy of the defect or impurity level. Shallow levels, which lie near the conduction or valence band edge, are more likely to participate in radiative recombination, but the sample temperature must be small enough to discourage thermal activation of carriers out of the traps. Deep levels tend to facilitate nonradiative recombination by providing a stop-over for electrons making their way between the conduction and valence

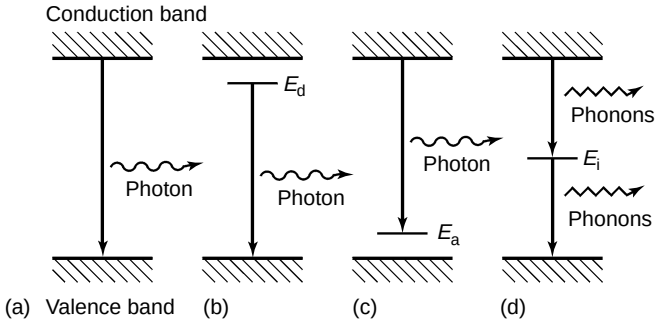

Figure 7 (a–c) Radiative recombination paths: (a) band-to-band; (b) donor to valence band; (c) conduction band to acceptor. (d) Nonradiative recombination via an intermediate state.

bands by emitting phonons. Several intrinsic and impurity transitions are illustrated in Figure 7.

Surfaces and interfaces usually contain a high concentration of impurity and defect states. As the physical termination of the sample, the surface represents a drastic interruption of the material itself. Dangling bonds often provide numerous midgap states that facilitate rapid nonradiative recombination. They also act as getters for ambient impurities. Dangling bonds can be accommodated by a variety of surface treatments and lattice-matching semiconductor epilayers. Even so, impurities accumulate on surfaces before treatment and during growth interruptions. In addition, interface defects due to slight lattice mismatch at semiconductor heterojunctions are inevitable.

In systems where the interfaces and bulk material contribute to the PL, the two components can be distinguished via depth profiling. Thewalt et al. combined PL with spreading resistance analysis (SRA) depth profiling to identify the type and location of several impurities in ultrahigh-purity epitaxial silicon.[20] PL spectra of the substrate and three epitaxial samples are shown in Figure 8. The strong Al line in sample (c) and the weaker line in sample (b) correlate well with an interfacial contaminant detected in SRA. Destructive depth profiling can be accomplished by step-etching and measuring the PL vs. etch depth. Akimoto et al. used this approach to connect specific PL peaks with vacancy complexes at a GaAs/AlGaAs interface.[21]

Some good examples of PL spectral analysis in the characterization of bare surfaces are in the field of semiconductor etching. Reactive ion etching (RIE), where etching is accomplished by bombarding the surface with energetic reactive ions, is an important procedure in the fabrication of optoelectronic devices with submicron features. However, RIE can produce various types of defects in the near-surface region. For optimal device performance, the damage and contamination caused by

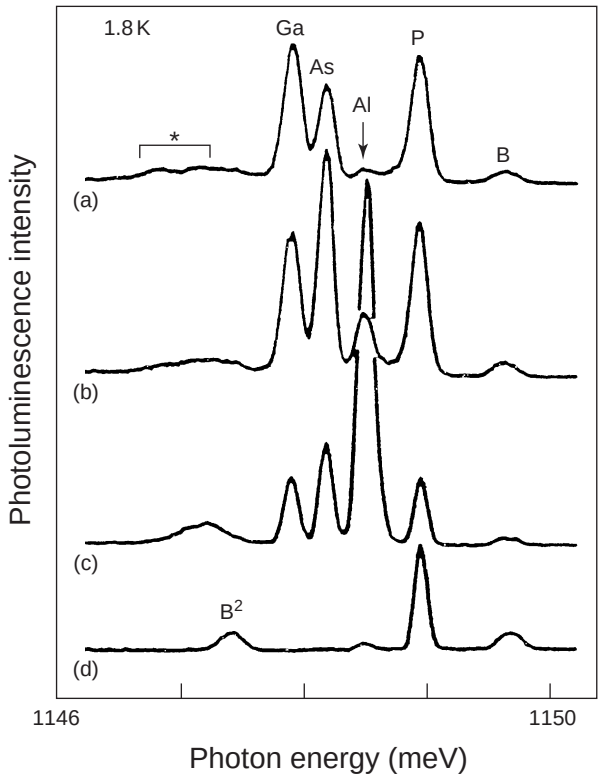

Figure 8 PL spectra of three epitaxial Si samples (a–c) and the substrate material (d) showing spectral lines associated with Ga, As, Al, P, and B impurities.[20]

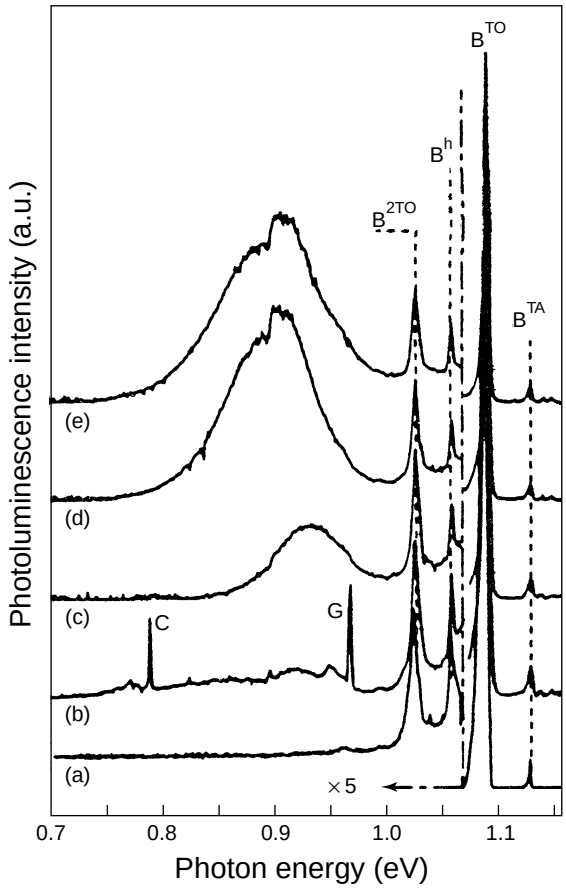

Figure 9 PL spectra of B-doped Si after the following plasma exposures: (a) 80% He–20% HBr; (b) CF_4; (c) Ar; (d) Ar-D_2; (e) D_2.[22]

RIE should be minimized. PL analysis is very useful in identifying and controlling RIE-induced defects.

Henry et al. used PL to study the effects of various plasmas on exposed surfaces of phosphorus- and boron-doped silicon.[22] As shown in Figure 9, the PL spectrum depended strongly on the plasma composition. Before treatment, the spectrum consisted of a number of lines just below the bandgap that are associated with the boron dopant in the material (B^{TA}, B^{TO}, B^{h}, and B^{2TO}). No deep luminescence was observed in the control sample. After treatment, several deep luminescence features emerged in the PL spectrum. The features can be divided into two categories: sharp lines and broad bands. The sharp lines are characteristic of radiative transitions at deep neutral defects in silicon, which are known from other studies involving high-energy particle irradiation. For example, the lines labeled C and G in spectrum 9(b) are recognized as transitions associated with carbon-related defects in silicon. The origin of the broad bands observed in spectra 9(c–e) is not completely understood. It is attributed to recombination at extended defects where the strained silicon lattice creates local potential wells that trap electrons and holes.

Broad, deep PL emission is a common characteristic of damaged semiconductors. Foad et al. have investigated changes in the PL spectrum of ZnTe/GaAs epitaxial layers after RIE and subsequent annealing.[23] RIE leads to a factor of 100 reduction in the near-band-edge emission, accompanied by the appearance of very broad new features well below the gap. One of the low-energy peaks coincides with a specific defect found in bulk ZnTe, but the peak is much broader than observed previously. Upon annealing, near-band-edge PL recovers by one order of magnitude and the deep emission shows more structure, revealing that the very broad, deep emission seen prior to annealing was actually composed of several phonon replicas (transitions assisted by one or more phonons). The annealing results suggest that the strain around RIE-induced defects is reduced.

Deep emission has also been attributed to dislocation networks at the interface of lattice-mismatched heterostructures. The effects of strain and dislocations can be separated in InGaAs/GaAs heterostructures where accommodation of lattice mismatch depends on the InGaAs layer thickness. Strain dominates when the InGaAs is thin, but the strain is relaxed by the formation

of misfit dislocations when the InGaAs layer exceeds a critical thickness. Joyce et al. observed broad, deep PL in the InGaAs/GaAs system and measured the spectrum as a function of InGaAs thickness.[24] They found that the intensity of the deep emission increases rapidly above the critical thickness, suggesting that interface dislocations are responsible for the broad, sub-bandgap PL.

As with surface damage, relative changes in surface state and band-edge emission can be used to evaluate surface passivation. For example, Xu et al. have used PL to study the properties of CuSe coatings on CdSe nanocrystals.[25] The uncoated CdSe nanocrystals produced broad, sub-bandgap PL, which they attributed to recombination at deep surface traps. With increasing CuSe fraction on the CdSe cores, they observed a monotonic decrease in this emission accompanied by steadily increasing band-edge PL. Ordinarily, materials with larger bandgaps are required for surface passivation so that carriers see potential barriers at the interface, shielding them from the surface. Because CuSe has a smaller bandgap than CdSe, surface passivation with this material is unusual. The authors suggest that the CuSe coating is very thin so that the CuSe gap is enhanced by quantum confinement.

Although spectral analysis of sub-bandgap emission is useful for characterizing interfaces, the peak position of the band-edge PL itself provides important information on atomic interdiffusion and interface alloying. The bandgap of a semiconductor alloy depends directly on alloy composition. When heterojunctions are not abrupt, unintentional alloy layers are formed at the interface. Depending on composition, such layers can have bandgaps larger, smaller, or intermediate between the adjoining materials. For example, Guimaraes et al. observed an anomalous emission band below the GaAs and GaInP gaps in GaInP/GaAs heterostructures.[26] They postulated that the band was due to recombination in GaInPAs intermediate layers, and confirmed their hypothesis by observing an enhancement of the band with an As-rich growth interruption at the interface.

Exchange or accumulation of atomic species at interfaces has a drastic impact on recombination in semiconductor QWs where individual monolayers of material are significant. GaInP/GaAs QW energies reported in the literature vary by as much as 100 meV for well widths below 10 monolayers. Mesrine et al. have shown that this scatter can be explained by two growth-related mechanisms: As/P exchange at interfaces and In surface segregation on GaInP layers.[27] Limiting cases incorporating a single monolayer of InAs or GaP at an interface more than account for the dispersion in PL energies obtained by different groups.

3.2 Photoluminescence Line Width and Splitting: Alloy Disorder and Interface Roughness

Whereas graded interfaces are likely to shift transition energies, interface roughness tends to produce line broadening and splitting in QWs. QW PL peaks are almost always broader than bulk PL. The line broadening is attributed to unintentional variation in the confinement energy in different regions of the well. Two distinct mechanisms contribute to lateral variation in the QW properties: alloy disorder in the well or barrier layers, and interface roughness. When the line width increases rapidly with decreasing well width L, the broadening is usually attributed to interface roughness. The confinement energy depends heavily on L in narrow wells and, assuming that the scale of interface fluctuations does not depend on layer thickness, decreasing the well width means that the fractional variation in L is larger. Alloy disorder can also lead to well-width-dependent PL broadening, especially when the disorder is concentrated in the barrier layers. In this case, reducing the well width causes the wavefunction of confined carriers to penetrate deeper into the barriers, thus sampling increasing amounts of disorder.

When optimal growth conditions are achieved, interface fluctuations can be restricted to a single monolayer of atoms and the average lateral extent of atomically smooth interfaces is larger than the wavefunction of the confined state. When PL measurements are conducted at liquid helium temperatures, the confined states are usually Coulomb-bound electron–hole pairs, or excitons. Several possible scenarios for excitons in QWs with single-monolayer fluctuations are shown in Figure 10. Typically, the lateral extent of atomically smooth islands is smaller

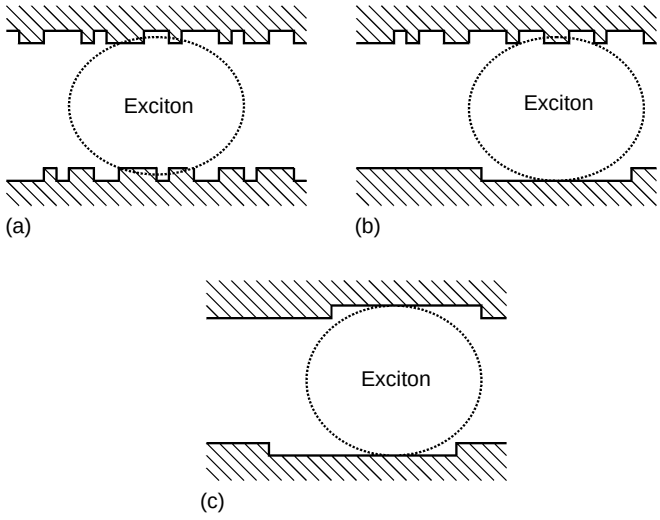

Figure 10 Models of interface structure relative to the size of the exciton wavefunction: (a) two rough interfaces; (b) one smooth and one rough interface; (c) two smooth interfaces.

than the exciton wavefunction, so the exciton senses a position-dependent distribution of well widths. This phenomenon tends to broaden the PL as excitons recombine in regions with different average well widths. In contrast, when interface island dimensions exceed those of the exciton, most of the recombination occurs within a limited set of QW regions having smooth walls and fixed well width. In this case, discrete lines may be resolved in the QW spectrum.

Because fine structure in low-temperature PL can have many origins, including a variety of impurity transitions, sharp lines in QW spectra must be interpreted with care. Indeed, the growth schedule for the highest quality QWs usually includes growth interruptions at interfaces when excess residual impurities could accumulate. Elman et al. observed splitting in the low-temperature PL of GaAs/AlGaAs QWs, suggesting recombination in well regions differing by one monolayer.[10] However, the lower energy peak (labeled I in Figure 11) vanished with increasing temperature, showing that it was impurity related. Meanwhile, a third peak at slightly higher energy (labeled E_{HH2}) gained strength with temperature, identifying it as an intrinsic QW transition. Thermal activation of carriers out of the relatively sparse impurity-related traps quenched the lower energy peak, whereas thermal population of the higher energy intrinsic state raised its PL to a level comparable with that of the lower intrinsic state.

The measured splitting between the intrinsic peaks was found to equal the energy difference expected for wells differing by one monolayer. The PLE spectrum correlates well with the intrinsic states (the impurity state does not appear) and time-resolved spectra demonstrate that they are coupled, confirming the interpretation that the doublet originates in adjacent QW regions with monolayer-thick deviations. Another expected signature of PL multiplets due to QW regions differing by single monolayers is the increased splitting with confinement energy. Deveaud et al. have compiled experimental results demonstrating this phenomenon.[9] This group also uses temperature-dependent PL and PLE to confirm that PL peaks are intrinsic and that excitons created in the narrow regions thermalize to the wider zones.

As described above, observation of PL splitting in QWs is unusual because very smooth interfaces are required. In ordinary QWs, PL broadening is more likely. Because broader lines imply a wider distribution of well widths, narrow PL lines generally indicate high-quality QWs. However, this interpretation can be misleading. Ferguson et al. have conducted an experimental and theoretical study of the effects of a misoriented substrate and growth interruption on the variation in QW line width against well thickness L.[28] In samples with uninterrupted growth, the increase in line width with decreasing L corresponded to

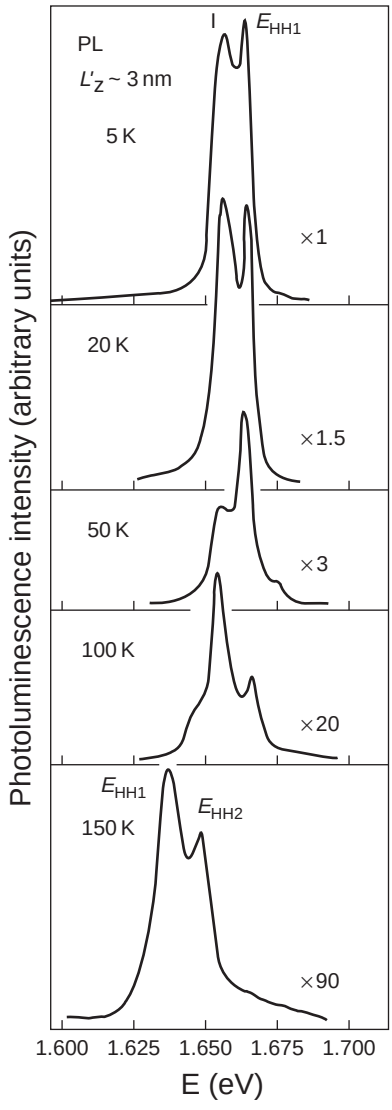

Figure 11 PL spectra of a GaAs/AlGaAs single QW measured at the temperatures indicated. The peak labeled I disappears with increasing temperature, whereas E_{HH2} rises steadily.[10]

an effective interface roughness of 0.2 monolayers. When growth interruptions are used, the PL line width increased by a factor of 3–4, indicating well-width fluctuations of approximately one monolayer. The increase in line width accompanying growth interruption is attributed to an increase in the lateral extent of interface islands, which is generally regarded as an improvement of the interface. The idea here is that when islands are small the exciton sees a constant average well thickness, but when islands are large the well width depends on exciton position.

In their analysis, Ferguson et al. assumed that the alloy composition of the layers was constant and that the line width reflected interface roughness alone. This interpretation was supported by the strong dependence

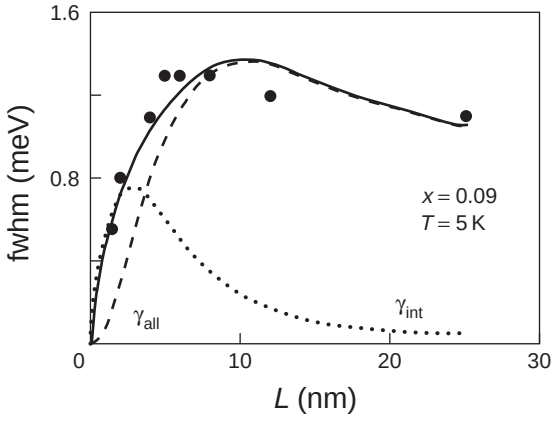

Figure 12 Experimental (dots) and theoretical (solid line) full width at half-maximum (fwhm) of the exciton recombination line in InGaAs/GaAs QWs. Interface roughness (γ_{int}, dotted line) and alloy disorder (γ_{all}, dashed line) contributions to the theoretical result are shown separately.[30]

of line width on L. Woods et al. studied the PL of InGaAsP/InGaAsP QWs and came to the opposite conclusion.[29] Because the line width did not increase in the narrow wells, the broadening was attributed to alloy composition variation in the well and barrier layers. In general, of course, both mechanisms will contribute to the PL line width. For example, Patane et al. found that alloy disorder *and* interface roughness are required to fit the L dependence of $In_xGa_{1-x}As/GaAs$ PL lines.[30] Although alloy disorder is sufficient to explain the line width in wide wells, interface roughness is necessary to fit the data at small L (see Figure 12). When $x = 1$, alloy broadening is absent and interface roughness alone must account for the L-dependent line width.

4 PHOTOLUMINESCENCE INTENSITY

Of all the properties that characterize PL, the intensity of the PL signal has received the most attention in the analysis of interfaces. This interest is due to the fact that, although several important mechanisms affect the PL response, it is generally found that large PL signals correlate with good interface properties. A useful review of the dominant mechanisms and the relationship between them has been provided by Chang et al.[31] In particular, they discuss the roles of the surface recombination velocity S and band bending at the surface in the PL measurement. Because surface recombination is usually nonradiative, and band bending can lead to the formation of a depletion region or ''dead layer'' where PL is effectively quenched, both of these phenomena tend to reduce the PL intensity. Distinguishing between the two effects is difficult and usually requires a supplementary measurement of the surface potential.

All else being constant, the surface recombination velocity is proportional to the density of surface states. However, changes in the surface-state density can affect the accumulation of charge at the surface, thereby altering the depletion thickness. If increasing the surface-state density enhances the depletion layer, both mechanisms suppress the PL intensity and the surface recombination velocity increases rapidly. Conversely, if the space-charge region is reduced by an increase in the density of surface states, the two mechanisms will have opposite effects on the PL signal and tend to cancel each other out. It should also be noted that, even though changes in surface-state density usually affect band bending, the inverse is not necessarily true. For example, the electric field at the surface can be modified by adsorption of molecules that shift the distribution of electrons between bulk and surface states but leave the surface-state density unchanged. Hence, the coupling of the two phenomena is quite complex.

PL intensity measurements have been used to evaluate a wide variety of surface treatments, including etching, oxidation, hydrogenation, adsorption of gases, deposition of coatings, and heteroepitaxy. The results are usually interpreted within the context of one of the above models: changes in surface-state density or changes in depletion thickness. Occasionally, when supplementary measurements of the surface Fermi level or dependence on applied bias are performed, more sophisticated models that account for both mechanisms are used. Otherwise, restricting the discussion to changes in an effective surface recombination velocity, which contains both effects, can accommodate the ambiguity of the phenomena.

Because a strong PL response is widely regarded as an indicator of a high-quality surface, and PL measurements are nondestructive and environment-insensitive, PL intensity measurements are an important in situ evaluation tool. The PL signal is monitored in real time while the surface is physically or chemically modified. Hence, the advantage of in situ methods is that processing steps can be controlled and optimized precisely. The relative ease of the experiment, combined with the broad implications of results for the optoelectronics industry, has made in situ PL a very popular technique. Indeed, the method has been used to characterize the development of virtually every technologically important surface undergoing almost every technologically relevant process. In these experiments, it is important to bear in mind that PLE itself can induce or accelerate chemical interactions.[32]

As an example, in situ PL has been used to assess InP surfaces during various cleaning and etching steps, oxidation, ambient gas flow, plasma exposure, and

heating.[31,32] In some cases, laser-induced interactions were found to alter etch rates, surface morphology, and surface contamination.[32] The results are explained by the availability of photoexcited electrons and holes that catalyze chemical reactions. In situ PL has also been used to probe the passivation of GaAs with heteroepitaxial AlGaAs monolayer by monolayer.[33] Conducting a similar study outside the growth chamber would have required a long sequence of separate growth runs, where run-to-run scatter in growth conditions could be problematic. Timoshenko et al. have extended the in situ PL technique to evaluate electrochemical treatments of indirect semiconductor surfaces, where pulsed excitation is required to obtain a sufficient PL signal.[34]

An interesting application of surface-adsorbate-induced changes in PL intensity has been described in a series of publications by Ellis et al.[35] Molecular species adsorbed onto semiconductor surfaces can be divided into two categories: Lewis acids that have a large electron affinity, and Lewis bases that have a small electron affinity relative to the work function of the material. Lewis acids draw electrons from the bulk to surface electronic states, and Lewis bases push electrons from surface states into the bulk. These shifts in surface charge expand or contract the depletion thickness of the semiconductor, quenching or enhancing the PL intensity. Hence, the partial pressure of a gaseous species that adsorbs to a semiconductor surface can be inferred by monitoring the PL signal, thus forming the basis for a novel chemical sensor.

4.1 Dependence on Applied Bias: Surface Potential

As discussed in section 2.1, the accumulation of charge in low-energy states near the surface leads to curvature of the conduction and valence bands. Representative pictures of this phenomenon are shown in Figure 13. If the surface states trap electrons, negative charge collects at the surface and the bands bend upward. If the surface states tend to lose electrons (trap holes), then positive charge accumulates, curving the bands downward. The electric field associated with this space-charge region sweeps electrons and holes in opposite directions, forming a depletion region free of electrons and holes. Because electrons and holes are spatially separated and cannot recombine, the depleted region is referred to as a dead layer. The PL intensity depends on the thickness of the dead layer, which relies on the magnitude of band bending. Increased curvature extends the recombination-free region deeper into the material, quenching the PL signal.

The magnitude of band curvature is characterized by the surface potential V_s indicated in Figure 13. One can compensate for or accentuate the surface potential V_s by applying an external bias to the surface. The resulting change in the depletion thickness can be evaluated by

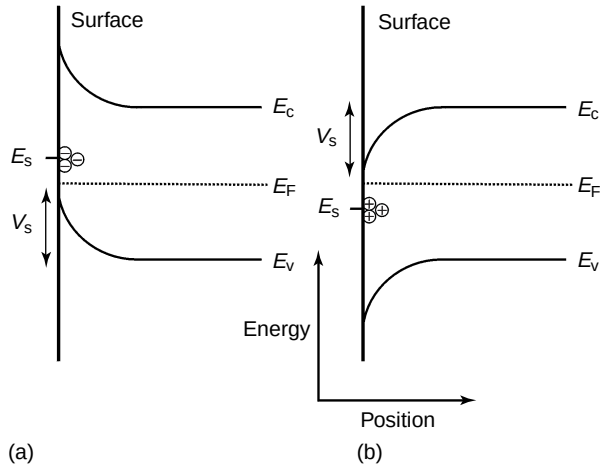

Figure 13 (a) When electrons accumulate in surface levels (denoted E_s), the conduction and valence bands bend upward near the surface. (b) Accumulation of holes bends the bands downward. The magnitude of the effect is characterized by the surface potential V_s.

monitoring the PL intensity. Using this approach, surface recombination and band bending contributions to the effective surface recombination velocity can be separated. Indeed, when the external bias exactly cancels the surface potential, the so-called flatband condition is obtained. At flatband, the depletion region is reduced to zero thickness and the intrinsic surface recombination velocity S_{in}, excluding dead-layer quenching, can be determined.

For a high-quality surface with a low density of surface states, the dead-layer model alone can explain the dependence of the PL intensity on applied bias. In contrast, when the surface-state density is large, PL quenching is dominated by surface recombination and the dead-layer model does not suffice. Chang et al. have observed both of these scenarios in their study of InP surface treatments.[31] However, they restrict their interpretation to qualitative statements about the relative role of the two mechanisms in PL extinction.

The difference between the bias-dependent PL signal and the dead-layer fit can be used to make a quantitative estimate of the intrinsic surface recombination velocity. Because the dead-layer model assumes that the intrinsic surface recombination velocity is effectively infinite, it underestimates the PL signal for finite S_{in}. Kauffman et al. have used this approach to analyze the excitation intensity dependence of the flatband surface trapping velocity at a GaAs/electrolyte interface.[36] The excitation intensity dependence of S_{in} is attributed to the filling of surface states. Hence, the analysis distinguishes between carrier trapping and carrier recombination, which occur at very different rates.

Excitation-dependent shifts in the flatband potential itself have been reported also for the GaAs/electrolyte

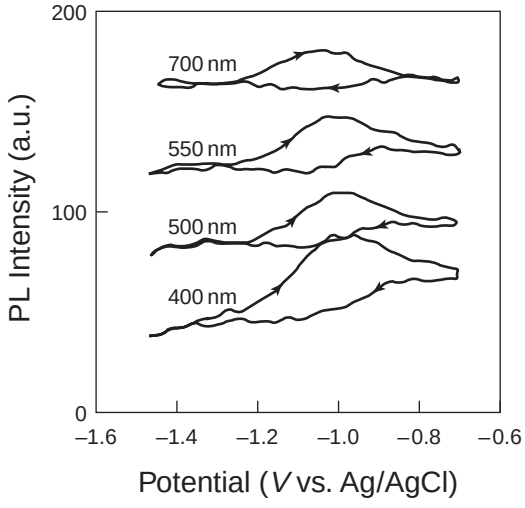

Figure 14 PL signal at 870 nm vs. applied potential for an etched p-GaAs electrode in 1 M NaOH. Excitation wavelengths are indicated in the plot. The hysteresis in the measurement is attributed to a fundamental change in the nature of the surface during the potential sweep.[5]

interface. Charge and potential distributions at interfaces can be altered dramatically by illumination. One manifestation of the flatband shift is hysteresis in the measurement of PL intensity versus applied bias.[5] It appears that the nature of the surface itself changes under specific biasing and illumination conditions, altering the flatband potential and the associated depletion layer thickness. This phenomenon leads to different bias-dependent PL signals for positive and negative voltage sweeps, as shown in Figure 14.

Under high excitation, surface states are saturated and the accumulation of photogenerated carriers near the surface screens the space-charge field. In this case, a decrease in the PL efficiency reflects the presence of additional surface states. High excitation measurements have been used in this context to study the formation of corrosion-induced surface states at a GaAs/electrolyte interface.[37] After exposure, the PL signal quenches when the applied bias exceeds a value of 1 V above the flatband potential. This result indicates that the applied bias bends the bands and that the new surface states become populated only when the minority carrier quasi-Fermi level reaches the energy of the states. Thus, from the dependence on applied bias, the energetic position of the new levels can be ascertained.

4.2 Spatial Dependence: Interface Uniformity and Carrier Diffusion

Because PL intensity is an indicator of interface quality, measurements of the PL signal vs. position provide information on the spatial uniformity of interface properties. The nondestructive and environment-insensitive features of PL make this application particularly useful in the evaluation of substrate surfaces, where detection of electrically active features may help to control problems in epitaxial devices. Such features might be overlooked in mechanical investigations of surface morphology. Although the spatial resolution in a PL scanning application is ultimately dictated by surface area and scan time, the diffraction limit of approximately 1 μm can be achieved in the best experimental arrangements. These high-resolution schemes must address carrier diffusion, which can smear images on a much larger scale.

Spatially resolved PL measurements are usually accomplished by scanning the optical excitation spot relative to the sample surface and detecting the PL signal in the far field. One of the first demonstrations of this approach was made by Krawczyk et al. in an investigation of InP surface treatments.[38] By coupling the excitation into one end of an optical fiber and scanning the other end relative to a focusing objective, they achieved resolution of the order of a few microns. They observed wide variation in the PL signal on a microscopic scale. For example, PL images of the effect of annealing on NH_4OH-treated InP are shown in Figure 15. The PL topography evolves from a random distribution of depressions and peaks to a flat response with randomly distributed PL islands. The annealing results are attributed to the presence of small oxide islands that protect the surface from thermal degradation.

Scanning the excitation laser with stepping motors or galvanometer mirrors usually requires a few seconds to accumulate a PL image. However, high-speed rastering with resonant mirrors or acousto-optic devices can generate frames at standard video rates.[39] In this case, PL images can be observed in real time. An additional advantage of video-rate laser scanning is the short dwell time (less than 1 μs) at each excitation position, which minimizes the possibility of photodegradation.

Although spatially resolved PL usually focuses on the band edge or integrated PL signal, spectral selectivity can be incorporated to map the distribution of particular surface states. Tajima has used this approach to plot the deep-level distribution in the near-surface region of GaAs and Si wafers.[40] Because these states tend to saturate at high excitation, he emphasizes the importance of using a low laser power and stabilizing the system mechanically to accommodate long dwell times. He also points out that the surface finish must be controlled carefully to avoid surface-treatment-related phenomena like those discussed above.

The spatial dependence of the PL spectrum itself can be used to evaluate uniformity of alloy composition, epilayer thickness, and a variety of other material properties that affect PL spectra. For example, the PL spectrum has

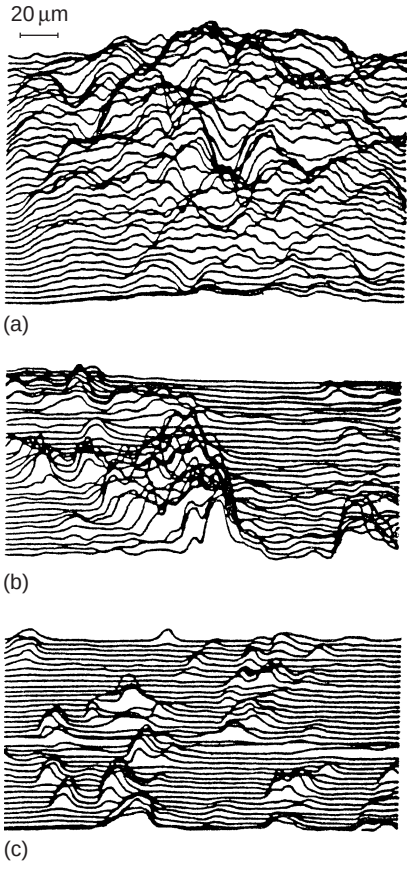

Figure 15 PL images of InP treated with NH_4OH: (a) as treated; (b) annealed at 350 °C; (c) annealed at 450 °C. The PL intensity scale in (b) and (c) is 10 times smaller than in (a).[38]

been recorded as a function of position to look for signs of dislocations or residual stress in laterally overgrown InP on InP-coated Si substrates.[41] Dislocations are expected to reduce the PL signal, and stress tends to shift and broaden PL peaks. Hence, the spatially resolved InP/Si PL was compared with that of lattice-matched InP layers grown homoepitaxially on InP substrates. The measured PL peak was as strong and narrow as that of the homoepitaxial InP with little shift in peak wavelength, suggesting that the overgrowth layer was dislocation and stress free. In addition, the PL spectrum was independent of position, indicating good uniformity in layer quality.

When a laser is focused on a surface, the minimum spot size is determined by the diffraction limit, which is approximately equivalent to the wavelength of the laser light. In the analysis of semiconductors, the optical excitation energy must exceed the bandgap, which usually corresponds to wavelengths in the visible or near-infrared. Hence, the diffraction-limited spot size is of the order of 1 μm. In the absence of carrier diffusion, the excitation spot size governs the spatial resolution of PL measurements. However, photogenerated carriers often diffuse average distances much larger than 1 μm before recombining, so the PL is generated in a broader region than the original excitation spot. Restricting the PL collection can circumvent diffusion-limited spatial resolution.

The most straightforward approach to this problem is a PL extension of confocal microscopy. Using matched illumination and collection optics with back focal plane apertures, enhanced spatial resolution, improved depth profiling, and insensitivity to scattered light are obtained. Only light originating in the focal spot is imaged by the collection optics. In the context of PL experiments, the configuration rejects luminescence due to recombination of carriers outside the illuminated area. A typical experimental set-up is shown in Figure 16. Fong et al. have used confocal PL to study a GaAs/AlGaAs QW grown on a grooved substrate.[42] The improved spatial resolution permitted them to profile abrupt fluctuations (on a scale of 1 μm) in alloy composition and QW thickness. In the confocal measurement, if the detection

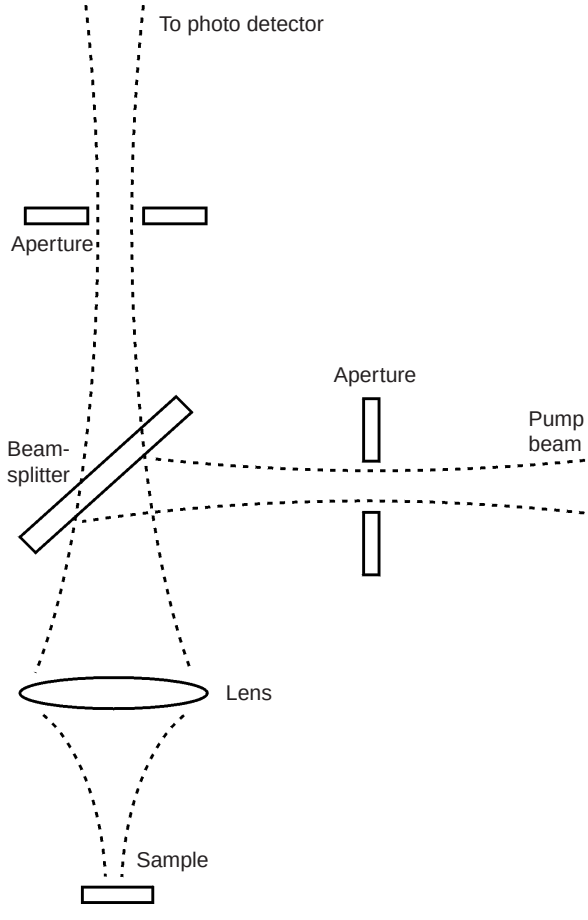

Figure 16 Typical experimental set-up for confocal PL measurements.

aperture is translated laterally in the image plane, the corresponding detection spot can be maneuvered relative to the excitation spot. This technique can be used to monitor the diffusion process itself. For example, Hubner et al. have measured the diffusion length along semiconductor quantum wire structures.[43] A variation in carrier transport with wire width is explained by changes in sidewall recombination due to different surface-to-volume ratios.

Finally, it should be pointed out that the diffraction limit itself can be surmounted when optical measurements are made in the near-field. If a subwavelength aperture is positioned in the near-field region of the optical emission, resolution comparable to the aperture size can be obtained. The most straightforward manifestation of this idea is to place a mask with a tiny hole very close to the sample surface such that excitation and PL must pass through the aperture. A more popular technique known as near-field scanning optical microscopy utilizes the tip of an extruded optical fiber to excite and/or collect the PL emission. Both of these approaches have produced spatial resolution of the order of 100 nm. Yet, with emission areas this small, generating sufficient PL signal under appropriate excitation conditions can be quite challenging.

4.3 Time Dependence: Recombination Rates

When CW excitation is used in a PL experiment, the system quickly converges on steady state. The rate of excitation equals the rate of recombination, and the photogenerated carrier density is constant in time. In contrast, when a material is excited by a series of short laser pulses, the concentration of carriers depends strongly on time. Because the laser pulse can be much shorter than the average recombination time, a specific carrier density can be generated almost instantaneously. The photoexcited carriers then recombine at a rate that is characteristic of the recombination path they follow. Time-resolved PL measurements can be used to determine carrier lifetimes, and to identify and characterize various recombination mechanisms in the material.

Photogenerated carrier lifetimes are obtained by monitoring the transient PL signal after pulsed excitation. Although the experimental apparatus required to make such a measurement depends on the desired resolution, the most common detection scheme is time-correlated single photon counting. When a photon is incident on a photodetector, an electrical pulse is generated. This pulse and an excitation reference pulse are fed into a constant fraction discriminator, which is designed to create output pulses that are timed correctly independent of input pulse size. Next, the signal and reference pulses are sent to a time-to-amplitude converter. This device uses the two input pulses as start and stop triggers for a linearly charged timing capacitor and produces an output pulse whose amplitude is governed by the capacitor charge. Thus, the output pulse height is proportional to the delay between input pulses. The output pulses are sorted according to amplitude and counted by a multichannel analyzer, yielding the transient PL decay.

As discussed in section 2.2, there are three general mechanisms for recombination in semiconductors: SRH transitions via intermediate states, radiative events, and Auger scattering. Equation (1) can be rewritten in terms of the nonequilibrium carrier lifetime τ as shown in Equation (2):

$$\frac{1}{\tau} = \frac{2S}{d} + \frac{B}{N}n + Cn^2 \tag{2}$$

It is important to recall that this expression relies on two assumptions: SRH recombination is dominated by interface phenomena and two equivalent interfaces are present in the system. In the analysis of interfaces, we are primarily interested in the SRH term, and in the carrier-density-dependent transition from SRH to bulk recombination. As the excitation level is increased, radiative transitions usually become important well before Auger scattering must be considered. Hence, the Auger term in Equation (2) can be ignored for our purposes. It should be noted, however, that it is possible for low-quality small-bandgap semiconductors with large C coefficients to have Auger rates comparable to the radiative rate when interface recombination begins to saturate.

Neglecting this possibility and assuming that the intrinsic carrier concentration is small, we can identify two important regimes in the carrier kinetics. At low excitation, the carriers are dilute and radiative recombination, which is proportional to n, is weak. Hence, the lifetime $\tau \approx d/2S$ is independent of excitation intensity. In this regime, measurements of τ vs. the active layer thickness d provide direct access to the interface recombination velocity. At the opposite extreme, when interface states are saturated and the radiative rate is fast, $1/\tau \approx (B/N)n$ and the lifetime decreases with increasing excitation. In this context, it is important to realize that n is not constant in time-resolved PL measurements. Rather, it decreases steadily after pulsed excitation via recombination processes. Therefore, at this excitation level one often observes both of the above phenomena in the transient PL signal: fast n-dependent decay at early times is followed by a slow, exponential SRH component. Some representative decay curves for GaAs/AlGaAs double heterostructures of varying thickness are shown in Figure 17.[44]

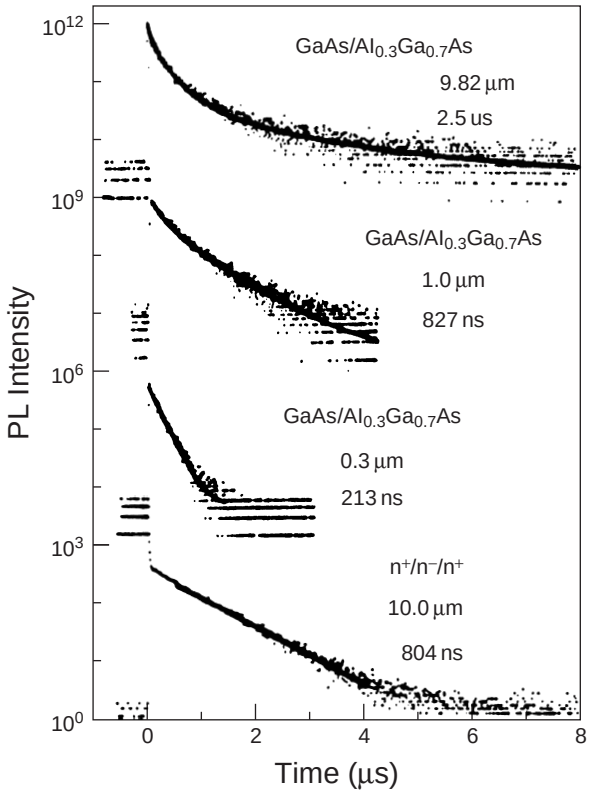

Figure 17 PL decay curves for a set of GaAs/AlGaAs double heterostructures and an $n^+/n^-/n^+$ GaAs homostructure. Active layer thickness (μm) and carrier lifetime (ns) are indicated in the plot (lifetimes are obtained by fitting the exponential tail of the curves).[44]

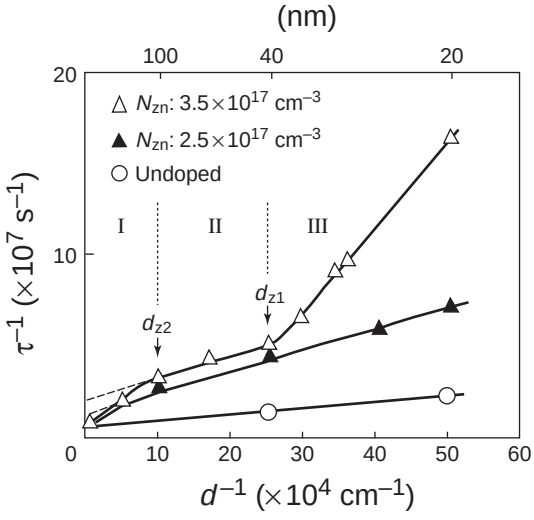

Figure 18 Recombination rate vs. the reciprocal of the active layer thickness in Zn-doped AlGaInP/GaInP double heterostructures. Symbols representing different doping levels are specified in the plot.[46]

Quantitative analysis of such nonexponential multicomponent decay curves is difficult. The situation is simplified in the case of doped materials with relatively high equilibrium carrier concentrations. Here, the photoexcitation can be kept well below the doping level so that the fixed majority carrier density controls the recombination kinetics. Because n is essentially constant throughout the transient measurement, the lifetime is independent of excitation even in the radiative and Auger regimes. Lifetimes in a set of samples covering a wide range of doping concentrations can yield accurate recombination coefficients.[45] In particular, a discrete transition from n-independent lifetimes to steadily decreasing n-dependent lifetimes should be observed when the carrier concentration reaches the threshold for significant bulk contributions.

Returning to the intrinsic (low equilibrium carrier concentration) case, if the pulsed excitation is kept sufficiently weak then even the initial photoexcited carrier density is within the SRH regime. Thus, an exponential PL decay with a constant lifetime is expected. As described above, the variation in τ with layer thickness yields the interface recombination velocity S. An excellent example of this type of experiment is the investigation by Domen et al. of Zn doping in AlGaInP/GaInP double heterostructures.[46] Zinc was intentionally incorporated into the upper cladding layer and the resulting carrier lifetime in the active layer was evaluated via time-resolved PL. Measured lifetimes for different doping concentrations are plotted against active layer thickness in Figure 18. The results show the expected behavior ($1/\tau \propto 1/d$) when separated into three regimes that are distinguished by the size of the active layer relative to the Zn-diffused region. In region I, only the top interface is degraded by the presence of Zn. In region II, an increasing fraction of the active layer suffers from diffused Zn. In region III, the high-concentration Zn penetrates all the way through the thin active layer and attacks the bottom cladding interface. The recombination velocities at the relevant interfaces are quickly obtained from the slope of the linear fits.

Time-resolved PL can also be used to distinguish between surface-state and depletion-layer contributions to PL quenching. A decrease in the integrated PL signal implies that the nonradiative recombination rate is enhanced relative to the radiative rate. Such an enhancement can be attributed to one of two mechanisms: an increase in the density of interface states, or an increase in the dead layer thickness. In transient PL, accelerated nonradiative recombination is manifested as a decrease in the photoexcited carrier lifetime. For example, wet-etch surface modification of GaInP has been characterized by CW and time-resolved PL.[47] Etchants yielding enhanced CW PL signals produced corresponding increases in carrier

lifetime, indicating that the nonradiative recombination rate is effectively reduced.

In this particular experiment, it is not clear whether the etching procedure removes nonradiative recombination sites or depletion-inducing surface charge. This shortcoming in the interpretation can be addressed by conducting the transient PL measurements at high excitation, where photoexcited carriers screen the electric field in the space-charge region. If the injection level is sufficiently high, the bands are essentially flat at the surface and changes in the carrier lifetime can be unambiguously attributed to changes in the surface-state density. Lunt et al. have demonstrated this approach in a comprehensive investigation of GaAs surfaces exposed to a variety of sulfur, nitrogen, and oxygen donors.[48] Some treatments produced an increase in the flatband carrier lifetime, indicating a decrease in the intrinsic surface recombination velocity. Meanwhile, other reagents produced substantial increases in the steady-state PL signal but did not affect the high-injection recombination rate. In this case, the results suggest that the enhanced PL is primarily due to changes in the equilibrium band bending. Thus, the two phenomena are effectively separated.

One final application of transient PL to note is the identification of transitions involving type II band alignment. Because these transitions occur between electrons and holes that are separated in real space (see Figure 6), the electron–hole wavefunction overlap is small. Thus, type II optical transitions are expected to proceed more slowly than transitions in type I systems.[19]

4.4 Temperature Dependence: Thermal Population

When light is incident on a semiconductor, photoexcitation of electrons from the valence to conduction band can occur for any photon energy exceeding the bandgap. In this process, the difference between the excitation energy and the bandgap goes into the kinetic energy of the photoexcited carriers. Thus, the initial distribution of electron and hole energies depends on energy conservation, not temperature, and the original carrier populations are nonthermal. For example, if the excitation energy is much larger than the bandgap, the electrons and holes will be created high in their respective bands with kinetic energies far exceeding the characteristic thermal energy of the lattice kT. The electrons and holes then thermalize with the lattice by emitting phonons and, in the absence of Fermi filling effects, settle into Boltzmann distributions above their respective band edges. Because this process almost always occurs on a timescale much faster than that of recombination, thermal distributions are usually observed in ordinary PL. Thus, in PL experiments, temperature is primarily used to tune the thermal occupation of available states.

At the lowest temperatures, PL is dominated by the lowest energy levels. For instance, excitons (electrons and holes bound by the Coulomb interaction) and shallow impurity traps often appear in low-temperature PL. These states are characterized by binding energies of the order of a few milli-electron-volts, which is small relative to kT at ordinary temperatures. As the sample temperature and corresponding thermal energy are increased, excitons dissociate and carriers vacate shallow traps, reducing the intensity of these features in the PL spectrum. It is important to note here that the PL signal from any level depends on two parameters: the fractional population, and the density of participating states. Thus, relatively sparse low-energy traps appear in low-temperature PL because the thermal population of high-density band levels is very small. However, when the thermal population of band levels is appreciable, band transitions dominate the PL because of the abundance of states taking part.

In addition to the population of discrete states, thermal distributions are manifested in the PL line width of band-edge transitions. When kT exceeds the inhomogeneous broadening in the sample, a high-energy Boltzmann tail appears on the PL peak. This feature can be useful for estimating the local temperature of the sample. However, because thermal quenching can hide sparse low-energy states and thermal broadening can obscure important details in the spectrum, certain PL experiments must be conducted at relatively low (i.e. liquid helium) temperatures.

Relative changes in state population with temperature provide evidence that PL peaks originate in the same part of the sample and that carriers are free to move between the available states. As discussed in section 3.2, multiple peaks in the PL of high-quality QWs are frequently attributed to states in regions of the well differing by single monolayers. In these systems, thermal redistribution among states rules out an important competing explanation: the peaks being due to recombination in separate regions of the sample where carriers do not mix. The temperature-dependent PL spectra in Figure 11 clearly show thermal redistribution among states. The peak labeled E_{HH2} gains strength with increasing temperature, as expected for thermal population of a higher energy state. In contrast, the peak labeled I vanishes at high temperature, demonstrating the importance of the other parameter that controls PL intensity – the density of states. Peak I is strong at low temperature because carriers are trapped at these sites and do not have enough thermal energy to escape, but it disappears at high temperature because the states are sparse relative to the intrinsic bands.

In the characterization of discrete low-energy states, quantitative analysis of the decrease in PL intensity with

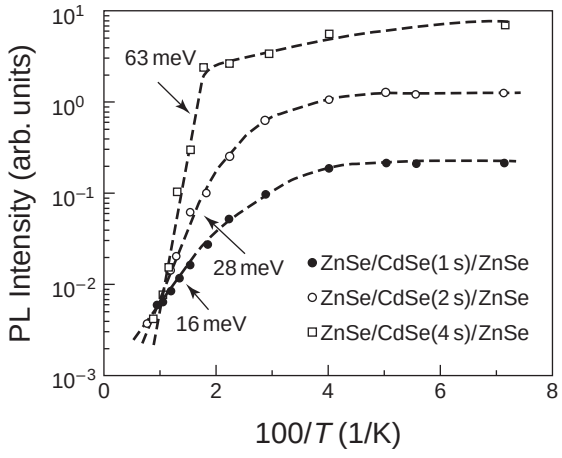

Figure 19 Arrhenius plots of the temperature-dependent PL intensity from ZnSe/CdSe submonolayer QWs. Deposition times of 1 s, 2 s, and 4 s, as indicated in the plot, correspond to well thicknesses of 1/4, 1/2, and 1 monolayer, respectively.[49]

temperature can be used to measure the depth of the trap. Plotting the log of the PL intensity vs. the reciprocal of the temperature, the slope yields the activation energy for exciting carriers out of the traps. These graphs, often referred to as Arrhenius plots, have been used to study interface alloy formation in ZnSe/CdSe QWs.[49] Results for three submonolayer wells (1, 1/2, and 1/4 monolayer thick) are displayed in Figure 19. The linear fit at high temperature gives the activation energies shown for delocalization of carriers from the submonolayer well region. When less than one monolayer of CdSe is deposited on ZnSe, two scenarios must be considered. Either distinct CdSe islands form within the ZnSe bulk or atomic interdiffusion leads to ZnCdSe alloying. Here, the increasing well depth with layer thickness indicates that ZnCdSe alloys, not CdSe islands, are formed.

Temperature is also an important parameter in carrier dynamics. The temperature dependence of the carrier lifetime relies on the recombination mechanism and can be used to distinguish between SRH and radiative events. Radiative recombination slows with increasing temperature for two reasons. At the lowest temperatures, excitons dominate radiative transitions and thermal energy in this regime leads to difficulties in momentum conservation. Because photons have small momentum, only low-momentum excitons can directly recombine. Rising temperatures confound recombination by increasing the fraction of excitons with excessive momentum. Beyond this regime, the excitons themselves begin to dissociate and the oscillator strength for free carriers is usually much smaller than that for excitons. In this case, recombining carriers must have equal and opposite momenta, a condition that decreases in likelihood as the average energy increases. In contrast, nonradiative recombination processes tend to accelerate with increasing temperature. In particular, nonradiative interface recombination usually involves thermally activated multiphonon events.[50]

A good example of this behavior is the single- and two-photon excitation studies of ZnSe conducted by Wang et al.[2] As discussed in section 2.1, single-photon above-gap excitation is absorbed very close to the surface, whereas two-photon excitation penetrates deep into the bulk. In the single-photon excitation experiments, the transient PL reveals short carrier lifetimes that diminish with increasing temperature, as expected for nonradiative surface recombination. In contrast, relatively long lifetimes are observed under two-photon excitation and the lifetime is extended at higher temperatures, indicating that radiative transitions dominate in the bulk.

5 PHOTOLUMINESCENCE POLARIZATION

Polarization is another important degree of freedom in optical measurements. In PL experiments, the emission polarization depends on the orientation of the dipole oscillators. Ordinarily, all orientations of electron–hole pair recombination events are equally probable, so the PL is not polarized. However, a variation in PL intensity with electric field orientation is sometimes observed. The maxima and minima in the polarization-angle-dependent PL signal usually occur for electric fields aligned with crystallographic axes. Polarization anisotropy in PL can be attributed to bond asymmetries, alloy composition modulation, and strain.

Investigations of PL polarization in the analysis of interfaces are relatively sparse. Nevertheless, they can be quite useful for identifying qualitative features of interfaces. Vignaud et al. observed large polarization anisotropy in the PL from InAlAs/InP heterostructures.[19] They found that the polarization angle yielding the maximum PL signal depended on the specific interface structure. When a thin well of InAs was present at the interface, the PL signal peaked with the polarizer parallel to the [011] axis. On the contrary, when a thin InAlP barrier was present, the PL signal was strongest along $[01\bar{1}]$. Although the exact origin of this phenomenon was not fully understood, the two scenarios agreed with other measurements that distinguished between the two types of interfaces. The polarization anisotropy was tentatively attributed to local interface bond asymmetry.

Another system that has displayed marked PL polarization anisotropy is ZnCdSe/ZnSe QWs grown on GaAs(110) surfaces.[51] The variation in PL intensity with polarization angle for one such QW is shown in Figure 20. The ratio of maximum to minimum polarized PL signals can be calculated using Luttinger parameters,

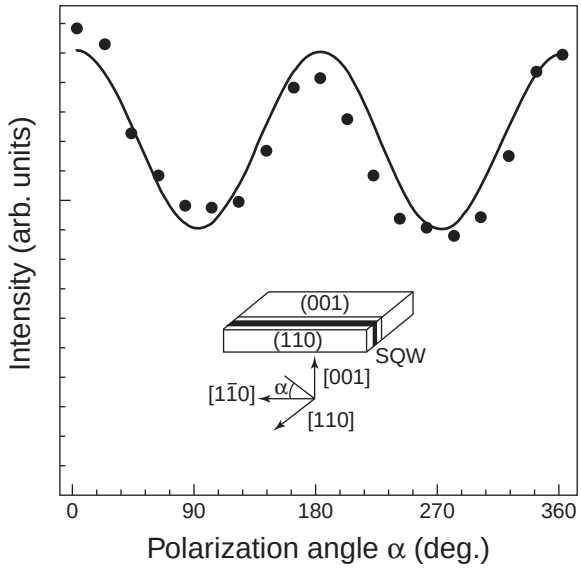

Figure 20 Polarization angle dependence of the PL intensity from a (110)-oriented ZnCdSe/ZnSe single QW. The solid line is a simple sinusoidal fit.[51]

which describe the valence band dispersion in the semiconductor. Assuming that the Luttinger parameters of ZnCdSe are similar to those of ZnSe, the calculated ratio is much smaller than that found experimentally. The enhanced polarization anisotropy in this system is attributed to compressive strain in the well layer. This interpretation is consistent with the observation of a very narrow PL line width in this sample. Because the lattice mismatch is accommodated by uniform strain rather than dislocations at the interface, the distribution of states in the QW should be relatively sharp.

6 CONCLUSION

PL analysis is a powerful tool in the characterization of surfaces and interfaces. Although a number of experimental techniques can provide detailed mechanical information about interfaces, the optoelectronics industry that drives most interface investigations is ultimately concerned with optical and electronic properties. Mechanical information is useful because it is closely correlated with these properties, but PL measurements explore electronic features directly. Other techniques can provide similar access, but they typically require more sophisticated excitation or detection schemes.

PLE is simple, but it is also quite versatile. The excitation energy and optical intensity can be chosen to study different regions and recombination mechanisms near interfaces. Because absorption of the incident light depends on the excitation energy, this parameter determines the depth of the PL probe. Using PLE, absorption and emission energies are observed simultaneously to evaluate the distribution of electronic states. The excitation intensity is even more important, controlling the density of photoexcited electrons. This density is critical in the interpretation of recombination dynamics.

The PL signal itself is characterized by three essential features: energy, intensity, and polarization. Because PL is the result of optical transitions between electronic states, the PL spectrum gives precise information on the energy levels available to electrons in the material. The intensity of the PL signal depends on the rate of radiative and nonradiative events, which depends in turn on the density of nonradiative interface states. Although investigations of PL polarization are still relatively sparse and are not well understood, they can identify unique anisotropic features in the underlying crystal. Future work in this area should lead to a more complete picture of asymmetry at surfaces and interfaces.

PL measurements are not sensitive to the pressure in the sample chamber and can be performed at virtually any temperature. These features make PL an excellent in situ probe of surface or interface modification. Even so, variation of the PL signal with external parameters such as temperature and applied bias can provide additional information on the nature of interfaces. Temperature-dependent thermal activation of electronic states can be used to estimate their depth below the intrinsic bands. An applied bias shifts the bands at the surface, permitting evaluation of the zero-bias band bending. Thus, the availability of PL under a wide variety of experimental conditions provides for advanced measurement opportunities.

Applications of PL analysis range from simple spatial scans of epitaxial wafers to exhaustive investigations of excitation-intensity-dependent PL in novel materials. Furthermore, new PL techniques continue to emerge, expanding the arsenal of PL analysis. Because PLE is usually absorbed near the surface, and interfaces tend to dominate electronic behavior in layered systems, PL is especially well suited to surface and interface investigations. Interfaces are increasingly important in new optoelectronic materials where layered structures are becoming thinner and more complex. Thus, although PL measurements have been useful for the characterization and refinement of such materials, they can be expected to play an even greater role in the future.

ABBREVIATIONS AND ACRONYMS

PL	Photoluminescence
PLE	Photoluminescence Excitation
PLS^3	Photoluminescence Surface-state Spectroscopy

QW Quantum Well
RIE Reactive Ion Etching
SRA Spreading Resistance Analysis
SRH Shockley–Read–Hall

RELATED ARTICLES

Electronic Absorption and Luminescence **(Volume 12)**
Surface Measurements using Absorption/Luminescence

REFERENCES

1. J. Garcia-Garcia, J. Gonzalez-Hernandez, J.G. Mendoza-Alvarez, E.L. Cruz, G. Contreras-Puente, 'Photoluminescence Characterization of the Surface Layer of Chemically Etched CdTe', *J. Appl. Phys.*, **67**, 3810–3814 (1990).
2. H. Wang, K.S. Wong, B.A. Foreman, Z.Y. Yang, G.K.L. Wong, 'One- and Two-photon-excited Time-resolved Photoluminescence Investigations of Bulk and Surface Recombination Dynamics in ZnSe', *J. Appl. Phys.*, **83**, 4773–4776 (1998).
3. R.E. Hollingsworth, J.R. Sites, 'Photoluminescence Dead Layer in p-Type InP', *J. Appl. Phys.*, **53**, 5357–5358 (1982).
4. H. Van Ryswyk, Arthur B. Ellis, 'Optical Coupling of Surface Chemistry. Photoluminescent Properties of a Derivatized Gallium Arsenide Surface Undergoing Redox Chemistry', *J. Am. Chem. Soc.*, **108**, 2454–2455 (1986).
5. K. Uosaki, Y. Shigematsu, S. Kaneko, H. Kita, 'Photoluminescence and Impedance Study of p-GaAs/Electrolyte Interfaces Under Cathodic Bias: Evidence for Flat-band Potential Shift During Illumination and Introduction of High-density Surface States by Pt Treatment', *J. Phys. Chem.*, **93**, 6521–6526 (1989).
6. K. Mettler, 'Photoluminescence as a Tool for the Study of the Electronic Surface Properties of GaAs', *Appl. Phys.*, **12**, 75–82 (1977).
7. Y. Zhu, C. Ding, G. Ma, Z. Du, 'Electronic State Characterization of TiO_2 Ultrafine Particles by Luminescence Spectroscopy', *J. Solid State Chem.*, **139**, 124–127 (1998).
8. M. Anpo, M. Kondo, S. Coluccia, C. Louis, M. Che, 'Application of Dynamic Photoluminescence to the Study of Active Surface Sites on Supported Mo/SiO_2 Catalysts. Features of Anchored and Impregnated Catalysts', *J. Am. Chem. Soc.*, **111**, 8791–8799 (1989).
9. B. Deveaud, A. Regreny, J.-Y. Emery, A. Chomette, 'Photoluminescence Study of Interface Defects in High-quality GaAs–GaAlAs Superlattices', *J. Appl. Phys.*, **59**, 1633–1640 (1986).
10. B.S. Elman, E.S. Koteles, C. Jagannath, Y.J. Chen, S. Charbonneau, M.L.W. Thewalt, 'Optical Characterization of Single Quantum Wells Fabricated Under Conditions of Interrupted Growth', *Proc. SPIE*, **794**, 44–49 (1987).
11. J.M. Olson, R.K. Ahrenkiel, D.J. Dunlavy, B. Keyes, A.E. Kibbler, 'Ultralow Recombination Velocity at $Ga_{0.5}In_{0.5}P$/GaAs Heterointerfaces', *Appl. Phys. Lett.*, **55**, 1208–1210 (1989).
12. S. Komiya, A. Yamaguchi, I. Umebu, 'Characterization of Radiative Efficiency in Double Heterostructures of InGaAsP/InP by Photoluminescence Intensity Analysis', *Solid State Electron*, **29**, 235–240 (1986).
13. M. Mullenborn, N.M. Haegel, 'Recombination Model for Heterostructure Interfaces', *J. Appl. Phys.*, **74**, 5748–5753 (1993).
14. T.H. Gfroerer, E.A. Cornell, M.W. Wanlass, 'Efficient Directional Spontaneous Emission from an InGaAs/InP Heterostructure with an Integral Parabolic Reflector', *J. Appl. Phys.*, **84**, 5360–5362 (1998).
15. T. Saitoh, H. Iwadate, H. Hasegawa, 'In Situ Surface State Spectroscopy by Photoluminescence and Surface Current Transport for Compound Semiconductors', *Jpn. J. Appl. Phys.*, **30**, 3750–3754 (1991).
16. T. Yoshida, T. Hashizume, H. Hasegawa, 'Characterization of Interface Electronic Properties of Low-temperature Ultrathin Oxides and Oxynitrides Formed on Si(111) Surfaces by Contactless Capacitance–Voltage and Photoluminescence Methods', *Jpn. J. Appl. Phys.*, **36**, 1453–1459 (1997).
17. Y.J. Ding, C.L. Guo, J.B. Khurgin, K.-K. Law, J.L. Merz, 'Characterization of Recombination Processes in Multiple Narrow Asymmetric Coupled Quantum Wells Based on the Dependence of the Photoluminescence on Laser Intensity', *Appl. Phys. Lett.*, **60**, 2051–2053 (1992).
18. Y.J. Ding, D.C. Reynolds, S.J. Lee, J.B. Khurgin, W.S. Rabinovich, D.S. Katzer, 'Evidence for Strong Spatially Localized Band-filling Effects at Interface Islands', *Appl. Phys. Lett.*, **71**, 2581–2583 (1997).
19. D. Vignaud, X. Wallart, F. Mollot, B. Sermage, 'Photoluminescence Study of the Interface in Type II InAlAs–InP Heterostructures', *J. Appl. Phys.*, **84**, 2138–2145 (1998).
20. M.L.W. Thewalt, A.G. Steele, J.E. Huffman, 'Photoluminescence Studies of Ultrahigh-purity Epitaxial Silicon', *Appl. Phys. Lett.*, **49**, 1444–1446 (1986).
21. K. Akimoto, K. Tamamura, J. Ogawa, Y. Mori, C. Kojima, 'Photoluminescence Study of the Interface of a GaAs/$Al_xGa_{1-x}As$ Heterostructure Grown by Metalorganic Chemical Vapor Deposition', *J. Appl. Phys.*, **63**, 460–464 (1988).
22. A. Henry, B. Monemar, J.L. Lindstrom, T.D. Bestwick, G.S. Oehrlein, 'Photoluminescence Characterization of Plasma Exposed Silicon Surfaces', *J. Appl. Phys.*, **70**, 5597–5603 (1991).

23. M.A. Foad, M. Watt, A.P. Smart, C.M. Sotomayor Torres, C.D.W. Wilkinson, W. Kuhn, H.P. Wagner, S. Bauer, H. Leiderer, W. Gebhardt, 'High-resolution Dry Etching of Zinc Telluride: Characterization of Etched Surfaces by X-ray Photoelectron Spectroscopy, Photoluminescence and Raman Scattering', *Semicond. Sci. Technol.*, **6**, A115–A122 (1991).
24. M.J. Joyce, M. Gal, J. Tann, 'Observation of Interface Defects in Strained InGaAs–GaAs by Photoluminescence Spectroscopy', *J. Appl. Phys.*, **65**, 1377–1379 (1989).
25. L. Xu, X. Huang, H. Huang, H. Chen, J. Xu, K. Chen, 'Surface Modification and Enhancement of Luminescence due to Quantum Effects in Coated CdSe/CuSe Semiconductor Nanocrystals', *Jpn. J. Appl. Phys.*, **37**, 3491–3494 (1998).
26. F.E.G. Guimaraes, B. Elsner, R. Westphalen, B. Spangenberg, H.J. Geelen, P. Balk, K. Heime, 'LP-MOVPE Growth and Optical Characterization of GaInP/GaAs Heterostructures: Interfaces, Quantum Wells, and Quantum Wires', *J. Cryst. Growth*, **124**, 199–206 (1992).
27. M. Mesrine, J. Massies, E. Vanelle, N. Grandjean, C. Deparis, 'Photoluminescence Energy and Interface Chemistry of GaInP/GaAs Quantum Wells', *Appl. Phys. Lett.*, **71**, 3552–3554 (1997).
28. I.T. Ferguson, C.M.S. Torres, T.M. Kerr, 'The Optical Properties of Thin $Al_{0.3}Ga_{0.7}As$–GaAs Quantum Wells on Misoriented Substrates with [110] Terraces. A Study of Interface Roughness Using Photoluminescence', *Semicond. Sci. Technol.*, **7**, 892–899 (1992).
29. L.M. Woods, P. Silvestre, P. Thiagarajan, G.A. Patrizi, G.Y. Robinson, K.M. Jones, M. Al-Jassim, 'Photoluminescence and Interface Abruptness in InGaAsP/InGaAsP Quantum Wells', *J. Electron. Mater*, **23**, 1229–1233 (1994).
30. A. Patane, A. Polimeni, M. Capizzi, F. Martelli, 'Linewidth Analysis of the Photoluminescence of $In_xGa_{1-x}As$/GaAs Quantum Wells ($x = 0.09, 0.18, 1.0$)', *Phys. Rev. B*, **52**, 2784–2788 (1995).
31. R.R. Chang, R. Iyer, D.L. Lile, 'Surface Characterization of InP Using Photoluminescence', *J. Appl. Phys.*, **61**, 1995–2004 (1987).
32. H.J. Frenck, W. Kulisch, R. Kassing, 'Characterization of InP Surfaces Using Integral Photoluminescence Measurements', *Proc. SPIE* , **1144**, 250–258 (1989).
33. C.J. Sandroff, F.S. Turco-Sandroff, L.T. Florez, J.P. Harbison, 'Recombination at GaAs Surfaces and GaAs/AlGaAs Interfaces Probed by In Situ Photoluminescence', *J. Appl. Phys.*, **70**, 3632–3635 (1991).
34. V.Yu. Timoshenko, J. Rappich, Th. Dittrich, 'Express Characterization of Indirect Semiconductor Surfaces by In Situ Photoluminescence During Chemical and Electrochemical Treatments', *Appl. Surf. Sci.*, **123/124**, 111–114 (1998).
35. E.J. Winder, D.E. Moore, D.R. Neu, A.B. Ellis, J.F. Geisz, T.F. Kuech, 'Detection of Ammonia, Phosphine, and Arsine Gases by Reversible Modulation of Cadmium Selenide Photoluminescence Intensity', *J. Cryst. Growth*, **148**, 63–69 (1995), and references therein.
36. J.F. Kauffman, C.S. Liu, M.W. Karl, 'Surface Recombination Kinetics at the GaAs/Electrolyte Interface via Photoluminescence Efficiency Measurements', *J. Phys. Chem. B*, **102**, 6766–6773 (1998).
37. B.A. Balko, G.L. Richmond, 'Corrosion-induced Surface States on n-GaAs as Studied by Photoluminescence Versus Voltage Scans and Luminescence Decays', *J. Phys. Chem.*, **97**, 9002–9008 (1993).
38. S.K. Krawczyk, M. Garrigues, H. Bouredoucen, 'Study of InP Surface Treatments by Scanning Photoluminescence Microscopy', *J. Appl. Phys.*, **60**, 392–395 (1986).
39. G.E. Carver, 'Scanned Photoluminescence with High Spatial Resolution in Semi-insulating GaAs and InP: Aspects of Surface Passivation and Photodegradation', *Semicond. Sci. Technol.*, **7**, A53–A58 (1992).
40. M. Tajima, 'Characterization of Semiconductors by Photoluminescence Mapping at Room Temperature', *J. Cryst. Growth*, **103**, 1–7 (1990).
41. S. Naritsuka, T. Nishinaga, 'Spatially Resolved Photoluminescence of Laterally Overgrown InP on InP-coated Si Substrates', *J. Cryst. Growth*, **174**, 622–629 (1997).
42. Y.-C. Fong, E.A. Armour, S.D. Hersee, S.R.J. Brueck, 'A Confocal Photoluminescence Study of Metalorganic Chemical Vapor Deposition Growth on Patterned GaAs Substrates', *J. Appl. Phys.*, **75**, 3049–3055 (1994).
43. B. Hubner, R. Zengerle, A. Forchel, 'Direct Optical Analysis of Carrier Diffusion in Semiconductor Wire Structures', *Mater. Sci. Eng.*, **B35**, 273–277 (1995).
44. D.J. Wolford, G.D. Gilliland, T.F. Kuech, L.M. Smith, J. Martinsen, J.A. Bradley, C.F. Tsang, R. Venkatasubramanian, S.K. Ghandi, H.P. Hjalmarson, 'Intrinsic Recombination and Interface Characterization in 'Surface-free' GaAs Structures', *J. Vac. Sci. Technol.*, **B9**, 2369–2376 (1991).
45. R.K. Ahrenkiel, R. Ellingson, S. Johnston, M.W. Wanlass, 'Recombination Lifetime of $In_{0.53}Ga_{0.47}As$ as a Function of Doping Density', *Appl. Phys. Lett.*, **72**, 3470–3472 (1998).
46. K. Domen, M. Kondo, T. Tanahashi, 'Analysis for Deterioration of Radiative Efficiency in Zn-doped AlGaInP/GaInP Heterostructures Using Time-resolved Photoluminescence', *Inst. Phys. Conf. Ser.*, **129**, 447–452 (1992).
47. S.S. Kocha, M.W. Peterson, A.J. Nelson, Y. Rosenwaks, D.J. Arent, J.A. Turner, 'Investigation of Chemical Wet-etch Surface Modification of Ga0.5In0.5P Using Photoluminescence, X-ray Photoelectron Spectroscopy, Capacitance Measurements, and Photocurrent–Voltage Curves', *J. Phys. Chem.*, **99**, 744–749 (1995).
48. S.R. Lunt, G.N. Ryba, P.G. Santangelo, N.S. Lewis, 'Chemical Studies of GaAs Surface Recombination Using Sulfides and Thiols', *J. Appl. Phys.*, **70**, 7449–7467 (1991).

49. Z. Zhu, H. Yoshihara, K. Takebayashi, T. Yao, 'Characterization of Alloy Formation at the ZnSe/CdSe Quantum-well Interface by Photoluminescence Spectroscopy', *J. Cryst. Growth*, **138**, 619–624 (1994).
50. L.W. Molenkamp, H.F.J. van't Blik, 'Very Low Interface Recombination Velocity in (Al,Ga)As Heterostructures Grown by Organometallic Vapor-phase Epitaxy', *J. Appl. Phys.*, **64**, 4253–4256 (1988).
51. H.-C. Ko, D.-C. Park, Y. Kawakami, S. Fujita, S. Fujita, 'Fabrication and Optical Properties of ZnCdSe/ZnSe Single Quantum Wells on GaAs(110) Surfaces Cleaved in UHV by Molecular Beam Epitaxy', *Semicond. Sci. Technol.*, **11**, 1873–1877 (1996).

Proximal Probe Techniques

Roland Wiesendanger
University of Hamburg, Germany

1 Introduction **1**
2 Tunneling Microscopy-based Techniques **2**
2.1 Theory and Methods 2
2.2 Scanning Tunneling Spectroscopy of Dangling Bonds, Surface States and Image States 6
2.3 Local Work Function Mapping 9
2.4 Thermovoltage Mapping 10
2.5 Spin-polarized Scanning Tunneling Spectroscopy 11
2.6 Inelastic Electron Tunneling Spectroscopy 12
2.7 Scanning Tunneling Spectroscopy with Laser Excitation 13
2.8 Subsurface Information by Scanning Tunneling Microscopy on Semiconductors 13
3 Scanning Near-field Optical Spectroscopy **14**
4 Force Microscopy-based Techniques **16**
4.1 Theory and Methods 16
4.2 Lateral Force Microscopy 17
4.3 Adhesion Force Mapping 19
4.4 Force Modulation Microscopy 20
4.5 Noncontact Force Microscopy 20
4.6 Dynamic Force Microscopy with True Atomic Resolution 20
4.7 Magnetic Resonance Force Microscopy 21
5 Related Scanning Probe Techniques **21**
6 Perspective and Future Developments **22**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Proximal probe techniques are based on sharp tips which are brought within a small distance of a sample to be investigated by means of a stepper motor. The various interactions of these probe tips with the sample, which might be of mechanical, electrical, magnetic, optical or thermal origin, can be used to probe the sample's properties locally and to control the separation between tip and sample surface.

If the distance dependence of a particular type of interaction is strong, the distance control based on that interaction is very sensitive to small changes in tip–sample separation. By scanning the probe tip over the sample surface while keeping the interaction strength constant using a feedback loop, the surface contours can be followed by the tip with high accuracy. By monitoring the vertical position of the tip as a function of the lateral position, a three-dimensional image of the sample surface is obtained. The motion of the tip both laterally and vertically with respect to the sample surface can be realized with subatomic accuracy by means of piezoelectric drives.

In contrast to electron microscopy and vacuum-based surface analytical methods, proximal probe techniques can be operated in air and in liquids in addition to in vacuum. Therefore, proximal probe techniques offer great potential for investigations of solid–liquid interfaces or for in vivo investigations of biological specimens. Furthermore, proximal probe techniques can be operated in a wide temperature range (from millikelvin up to more than a 1000 K) and under extreme conditions (e.g. in high magnetic fields or high-pressure conditions). The most attractive feature of proximal probe techniques is, however, that they combine ultrahigh spatial resolution (down to the atomic scale) with high-resolution spectroscopic measurements so that detailed information about the physical, chemical or biological state of the sample can be extracted. Limiting factors are the finite size and often ill-defined state of the probe tips and the limited information depth of proximal probe techniques. It has also found to be difficult to identify elemental species based solely on proximal probe measurements although chemical-selective imaging even down to the atomic scale is nowadays routinely achieved.

1 INTRODUCTION

Scanning tunneling microscopy (STM) was the first proximal probe technique to be developed.[1–3] It is based on vacuum tunneling of electrons between an electrically conducting probe tip and an electrically conducting sample. The schematic set-up for an STM experiment is shown in Figure 1. An atomically sharp tip is brought within a distance of only a few ångströms ($1\,\text{Å} = 0.1\,\text{nm} = 10^{-10}\,\text{m}$) from the sample surface by means of a piezoelectric drive. If a bias voltage U has been applied between tip and sample, a tunneling current I will flow due to the quantum mechanical tunneling effect even before mechanical point contact is reached. Since the tunneling current is strongly (exponentially) dependent on the tip–surface separation, it can be used very efficiently for distance control. By scanning the tip

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

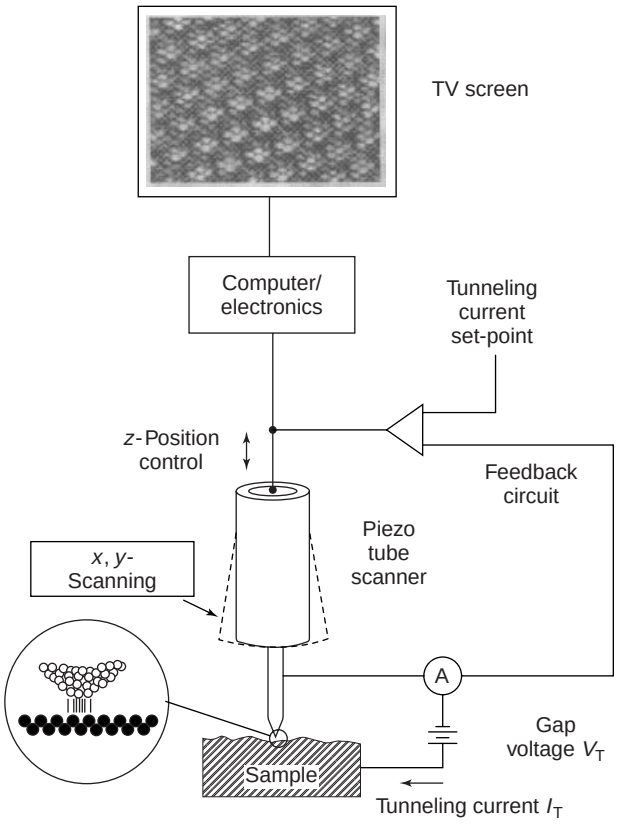

Figure 1 Schematic diagram of a scanning tunneling microscope.

over the sample surface while keeping the tunneling current constant, the surface contours can be traced by the tip. A ''constant-current topograph'' is finally gained by monitoring the vertical position z of the tip as a function of lateral position (x, y) which is again controlled by piezoelectric drives. Besides its use to control the tip–surface separation, the tunneling current contains valuable information about the local electronic properties of the sample surface.

Alternatively, different probes, e.g. a sharp optical probe, may be brought within close proximity to the sample at a distance s which is smaller than the optical wavelength λ used. In this so-called near-field regime the spatial resolution, which can be achieved, is no longer limited by diffraction, but rather by geometric parameters: the distance s between the probe tip and the sample surface and the effective radius of curvature R of the probe tip. In addition to the achievement of super-resolution beyond the diffraction limit, scanning near-field optical microscopy (SNOM) provides valuable information about local optical properties.[4,5] Various optical contrast mechanisms may be exploited, similarly to conventional optical microscopy.

Another type of proximal probe technique, scanning force microscopy (SFM), has found extremely wide fields of application.[6] In SFM, the forces or force gradients acting between a sharp tip, mounted on a cantilever, and a sample are detected by the bending of the cantilever. Depending on the particular type of SFM operation, with the tip and sample either in contact or not, short-range (e.g. ion–ion repulsion) or long-range [e.g. van der Waals (VDW), electrostatic, magnetic] forces, respectively, can be measured. SFM can be applied to any sample, including insulators, and can be combined with STM or SNOM operation by using appropriate probe tips. It also constitutes the basic set-up for various other proximal probe techniques, such as scanning near-field acoustic microscopy, scanning near-field thermal microscopy, scanning capacitance microscopy (SCM) and scanning ion conductance microscopy.[7]

2 TUNNELING MICROSCOPY-BASED TECHNIQUES

2.1 Theory and Methods

In this section, the basic principles of STM will be outlined and section 3 will focus on various modes of STM operation for chemical-specific imaging.

STM is based on the quantum-mechanical tunnel effect. A simple one-dimensional model is depicted in Figure 2. An electron of energy E is approaching a potential barrier of height V_0 and width s from the left. Even if we assume that $V_0 > E$, there exists a finite probability for the electron to traverse the potential barrier due to its quantum-mechanical wave nature. If the barrier is sufficiently high and wide ($\kappa s \gg 1$), then this probability is given by Equation (1):

$$T = \frac{16k^2\kappa^2}{(k^2+\kappa^2)^2} e^{-2\kappa s} \tag{1}$$

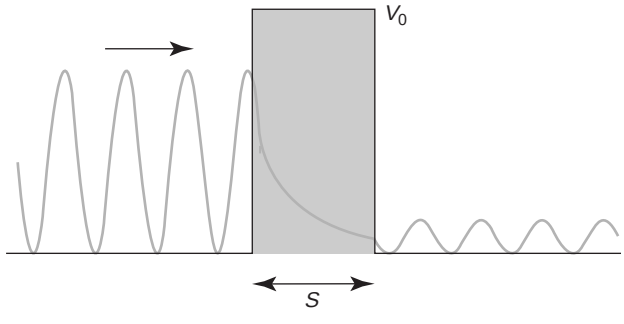

Figure 2 Schematic diagram of a one-dimensional tunnel barrier and the wave function for an electron traversing the barrier with a finite probability although its energy E is less than the barrier height V_0.

where (Equation 2)

$$k = \frac{\sqrt{2mE}}{\hbar} \tag{2}$$

and (Equation 3)

$$\kappa = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \tag{3}$$

with m being the electron mass and $\hbar$ Planck's constant divided by 2π.

The dominant contribution to the transmission coefficient T comes from the factor $\exp(-2\kappa s)$. This strong exponential dependence of T on the barrier width s and the square root of the effective barrier height $(V_0 - E)$ is typical for tunneling, independent of the exact shape of the barrier. Changing the barrier width by only 1 Å typically leads to a change of the barrier transmission by one order of magnitude.

In the three-dimensional case to be considered for describing the STM geometry correctly, a perturbative treatment of tunneling, as introduced by Bardeen,[8] provides an appropriate insight into the physics of the tunneling process as well as a connection to specific tip and sample surface properties. Within Bardeen's transfer Hamiltonian approach, the tunneling current can be evaluated in first-order time-dependent perturbation theory according to Equation (4):

$$I = \frac{2\pi e}{\hbar} \sum_{\mu,\nu} \{f(E_\mu)[1 - f(E_\nu + eU)] - f(E_\nu + eU) \times [1 - f(E_\mu)]\} \, |M_{\mu\nu}|^2 \delta(E_\nu - E_F) \tag{4}$$

where $f(E)$ is the Fermi function, U is the applied sample bias voltage, $M_{\mu\nu}$ is the tunneling matrix element between the unperturbed electronic states ψ_μ of the tip and ψ_ν of the sample and $E_\mu(E_\nu)$ is the energy of the state $\psi_\mu(\psi_\nu)$ in the absence of tunneling. The delta function describes the conservation of energy for the case of elastic tunneling. The tunneling matrix element is given by Equation (5):

$$M_{\mu\nu} = \frac{-\hbar^2}{2m} \int \mathrm{dS}(\psi_\mu^* \nabla \psi_\nu - \psi_\nu \nabla \psi_\mu^*) \tag{5}$$

where the integral has to be evaluated over an arbitrary surface lying entirely within the vacuum barrier region separating tip and sample. To derive the matrix element $M_{\mu\nu}$ from Equation (5), explicit expressions for the wave functions ψ_μ and ψ_ν of the tip and sample, respectively, are required, but ψ_μ is normally not known. Tersoff and Hamann,[9,10] who first applied the transfer Hamiltonian approach to STM, used the simplest possible model for the tip with spherical symmetry. In this model, the tunneling matrix element is evaluated for an s-type tip wave function ψ_μ, whereas contributions from tip wave functions with orbital quantum number $1 \neq 0$ have been neglected. Within this s-wave tip model the tunneling current in the limits of low temperature and small applied bias voltage becomes (Equation 6):

$$I \propto U n_t(E_F) \exp(2\kappa R) \sum_\nu |\psi_\nu(\mathbf{r}_0)|^2 \delta(E_\nu - E_F) \tag{6}$$

where $n_t(E_F)$ is the density of states at the Fermi level of the tip, R is the effective tip radius and $\mathbf{r}_0$ is the center of curvature of the tip. The quantity (Equation 7)

$$n_s(E_F, \mathbf{r}_0) = \sum_\nu |\psi_\nu(\mathbf{r}_0)|^2 \delta(E_\nu - E_F) \tag{7}$$

can be identified with the surface local density of states (LDOS) at the Fermi level (E_F), i.e. the charge density from electronic states at E_F of the sample, evaluated at the center of curvature $\mathbf{r}_0$ of the effective tip. The STM images obtained at low bias in the constant current mode therefore respresent contour maps of constant surface LDOS at E_F evaluated at the center of curvature of the effective tip, provided that the s-wave approximation for the tip can be justified. Since the wave functions decay exponentially in the z-direction normal to the surface (Equation 8):

$$\psi_\nu(\mathbf{r}) \propto \exp(-\kappa z) \tag{8}$$

it follows that (Equation 9)

$$|\psi_\nu(\mathbf{r}_0)|^2 \propto \exp[-2\kappa(s + R)] \tag{9}$$

where s denotes the distance between the sample surface and the front end of the tip. Therefore, the tunneling current, given by Equation (6), becomes exponentially dependent on the distance s (Equation 10):

$$I \propto \exp(-2\kappa s) \tag{10}$$

similarly to the case of a one-dimensional tunnel barrier (see Equation 1). Unfortunately, the simple interpretation of constant current STM images as given by Tersoff and Hamann is no longer valid for tip wave functions with angular dependence or for high applied bias voltages. Since STM tips are usually made from tungsten or platinum–iridium alloy wires for which the density of states at the Fermi level is dominated by d-states rather than by s-states,[11] the evaluation of the tunneling current according to the Equations (4) and (5) requires the calculation of the tunneling matrix element for tip wave functions with angular dependence ($1 \neq 0$). Chen[12] has shown that generally the tunneling matrix element can simply be obtained from a ''derivative rule'': the angle dependence of the tip wave function in terms of x, y and

z has to be replaced according to Equations (11–13):

$$x \longrightarrow \frac{\partial}{\partial x} \tag{11}$$

$$y \longrightarrow \frac{\partial}{\partial y} \tag{12}$$

$$z \longrightarrow \frac{\partial}{\partial z} \tag{13}$$

where the derivatives have to act on the sample surface wave function at the center of the apex atom. For instance, the tunneling matrix element for a p_z-tip state is proportional to the z-derivative of the sample surface wave function at the center of the apex atom at $\mathbf{r}_0$.

In terms of a microscopic view of the STM imaging mechanism[13] illustrated in Figure 3, a dangling-bond state at the tip apex atom is scanned over a two-dimensional array of atomic-like states at the sample surface. Overlap of the tip state with the atomic-like states on the sample surface generates a tunneling conductance which depends on the relative position of the tip state and the sample state. The atomic corrugation Δz depends on the spatial distribution in addition to the symmetry of tip and sample surface states. Generally, for non-s-wave tip states, the tip apex atom follows a contour, determined by the derivatives of the sample surface wave functions, which exhibit much stronger corrugation than the contour of constant surface LDOS at E_F.

Next, we consider effects due to the finite size of the applied bias voltage. First, the bias determines the energy window of those electronic states which can contribute to the tunneling current. Second, a finite bias and the corresponding electric field ($\sim 10^9\,\mathrm{V\,m^{-1}}$) can lead to a distortion of the tip and sample surface wave functions, ψ_ν and ψ_ν, and also to a modification

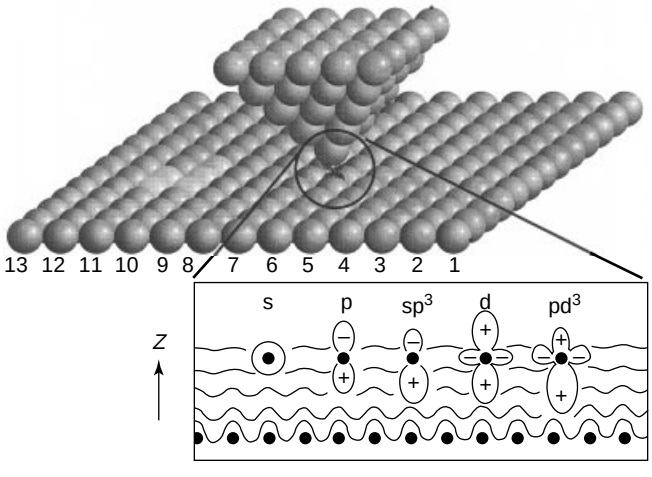

Figure 3 Microscopic view of the STM imaging mechanism which can be described by a process of bond forming and bond breaking between atomic-like tip and sample surface states.

of the energy eigenvalues E_μ and E_ν.[14] The derivation of these distorted tip and sample surface wave functions and energy eigenvalues under the presence of an applied bias is, however, a difficult problem. Therefore, as a first approximation, the undistorted zero-voltage wave functions and energy eigenvalues are usually taken. Consequently, the effect of a finite bias U only enters through an energy offset $\Delta E = eU$ of the undistorted surface wave functions or density of states relative to the tip states. Under this assumption, Equation (14) for the tunneling current as a generalization of the result of Tersoff and Hamann may be used:

$$I \propto \int_0^{eU} n_t(\pm eU \mp \varepsilon) n_s(\varepsilon) T(s, \varepsilon, eU)\, d\varepsilon \tag{14}$$

with an energy- and bias-dependent transmission coefficient $T(s, \varepsilon, eU)$ given by Equation (15):

$$T(s, \varepsilon, eU) = \exp\left\{ -2(s+R) \times \frac{2m}{\hbar^2}\left(\frac{\phi_t + \phi_s}{2} + \frac{eU}{2} - \varepsilon\right)^{1/2} \right\} \tag{15}$$

where $\phi_t(\phi_s)$ denotes the tip (sample surface) work function and $\varepsilon = E - E_F$ is the energy measured with respect to the Fermi level of the sample. By writing Equations (14) and (15), matrix element effects in tunneling are expressed in terms of a modified decay rate κ including a dependence on energy and applied bias voltage.

The experimental investigation of the bias dependence allows the extraction of various spectroscopic information at high spatial resolution, ultimately down to the atomic level. The spectroscopic capability of STM combined with its ultrahigh spatial resolution is its most powerful feature and makes it the most outstanding proximal probe technique for basic science studies. Figures 4(a–d) show a simplified one-dimensional potential energy diagram at zero temperature for the system consisting of the tip (left electrode) and the sample (right electrode) which are separated by a small vacuum gap. For zero applied bias (Figure 4b) the Fermi levels of tip and sample are equal at equilibrium. If a bias voltage U is applied to the sample, the main consequence is a rigid shift of the energy levels downward or upward in energy by an amount $|eU|$, depending on whether the polarity is positive (Figure 4c) or negative (Figure 4d). For positive sample bias, the net tunneling current arises from electrons that tunnel from the occupied states of the tip into unoccupied states of the sample (Figure 4c), whereas at negative sample bias, electrons tunnel from occupied states of the sample into unoccupied states of the tip. Consequently, the bias polarity determines

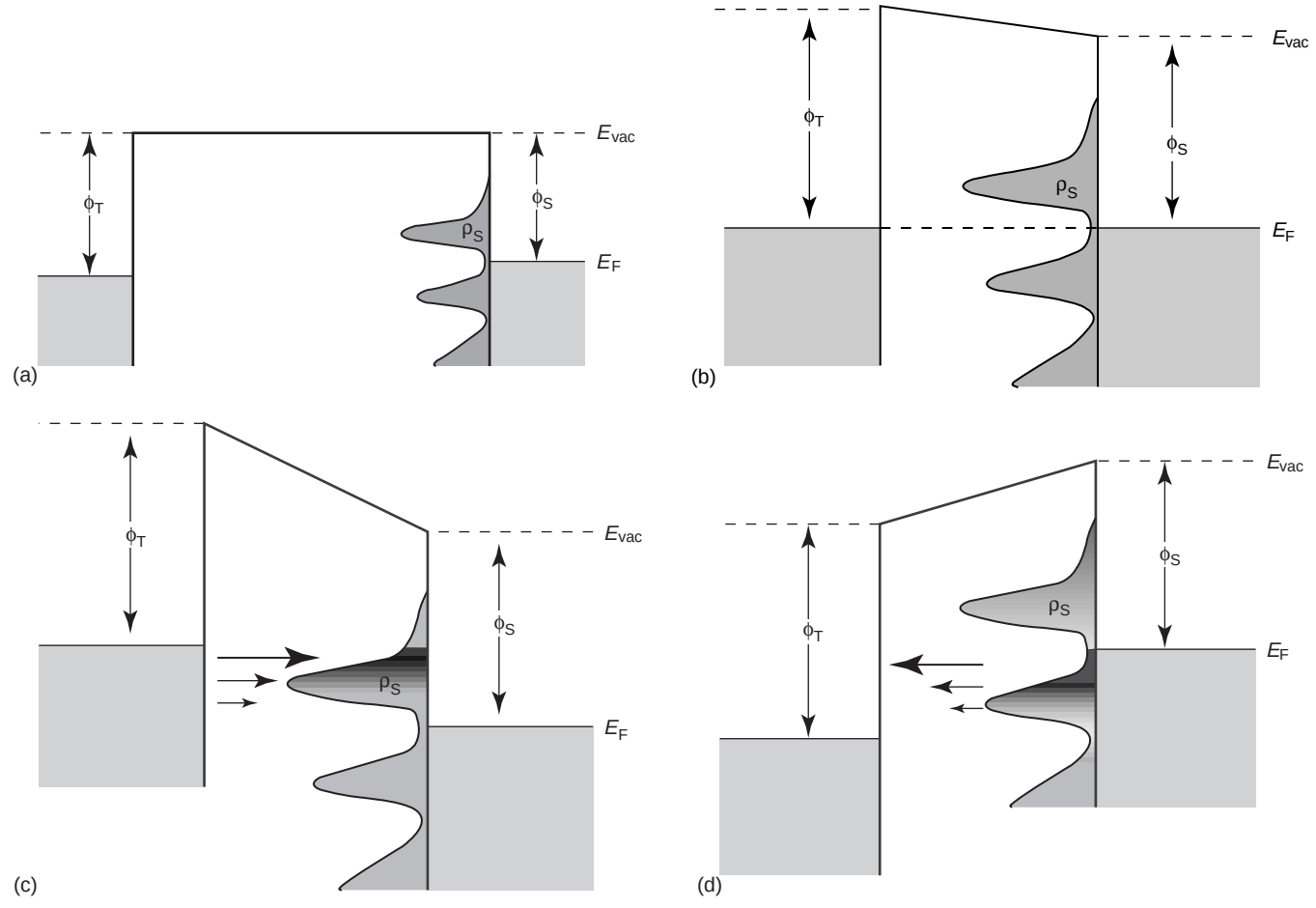

Figure 4 Energy level diagrams for tip (T) and sample (s). (a) Independent tip and sample; (b) tip and sample at equilibrium, separated by a small vacuum gap; (c) positive sample bias, electrons tunnel from the tip to the sample; (d) negative sample bias, electrons tunnel from the sample into the tip.

whether unoccupied or occupied sample electronic states are probed.

By varying the amount of the applied bias voltage, one can select the electronic states that contribute to the tunneling current and, in principle, measure the LDOS. For instance, the current increases strongly if the applied bias voltage allows the onset of tunneling into a maximum of the unoccupied sample electronic density of states. Therefore, the first derivative of the current–voltage characteristic, $\mathrm{d}I/\mathrm{d}U(U)$, should reflect the electronic density of states to a first approximation. By differentiation of Equation (14) one obtains Equation (16):

$$\frac{\mathrm{d}I}{\mathrm{d}U}(U) \propto n_t(0)n_s(eU)T(s, eU, eU) + \int_0^{eU} \frac{\mathrm{d}}{\mathrm{d}U}[n_t(\pm eU \mp \varepsilon)]n_s(\varepsilon)T(s, \varepsilon, eU)\,\mathrm{d}\varepsilon + \int_0^{eU} n_t(\pm eU \mp \varepsilon)n_s(\varepsilon)\frac{\mathrm{d}}{\mathrm{d}U}T(s, \varepsilon, eU)\,\mathrm{d}\varepsilon \qquad (16)$$

If we assume that the electronic density of states for the tip is featureless in the energy window determined by the applied bias, i.e. $\mathrm{d}n_t/\mathrm{d}U \approx 0$, then we can neglect the second term in Equation (16). Furthermore, at a fixed location the increase of the transmission coefficient with the applied bias voltage is smooth and monotonic. Therefore, structure in the $\mathrm{d}I/\mathrm{d}U(U)$ curve can usually be attributed to structure in the state density n_s via the first term in Equation (16). Under these assumptions we can write Equation (17):

$$\frac{\mathrm{d}I}{\mathrm{d}U}(U) \propto n_s(eU)T(s, eU, eU) \qquad (17)$$

with

$$T(s, eU, eU) = \exp\left\{-2(s+R)\times\frac{2m}{\hbar^2}\left(\frac{\phi_t+\phi_s}{2}-\frac{eU}{2}\right)^{1/2}\right\} \qquad (18)$$

The dominant contribution to the tunneling conductance comes from electrons near the Fermi level of the negatively biased electrode because the transmission coefficient becomes maximum in this case (Equation 19):

$$T_{\max}(s, eU, eU) = \exp \quad -2(s+R) \times \frac{2m}{\hbar^2}\left(\frac{\phi_t + \phi_s}{2} - \frac{|eU|}{2}\right)^{1/2} \Bigg\} \quad (19)$$

This fact has been indicated in Figures 4(b) and (c) by arrows of differing size.

As an important consequence of the dominant contribution of tunneling from states near the Fermi level of the negatively biased electrode, tunneling from the tip to the sample (Figure 4c) mainly probes the sample's empty states with negligible influence of the tip's occupied states. On the other hand, tunneling from the sample to the tip is much more sensitive to the electronic structure of the tip's empty states which sometimes prevents detailed spectroscopic STM studies of the sample's occupied states.[15]

STM can nowadays provide atomic or molecular resolution on many different kinds of materials routinely. On the other hand, the analytical capabilities of STM are not yet fully developed. The determination of chemical species by STM or tunneling spectroscopy is inherently prevented by the fact that valence electronic states accessible by the STM cannot unequivocally be related with a particular element in a multicomponent solid. However, considerable progress has been made in recent years to exploit various contrast mechanisms to differentiate between different atomic or molecular species. In the following, several examples of chemical contrast mechanisms in STM will be presented and discussed.

2.2 Scanning Tunneling Spectroscopy of Dangling Bonds, Surface States and Image States

Scanning tunneling spectroscopy (STS) can be a powerful experimental method for atom-selective imaging, as first demonstrated at the example of compound semiconductors such as GaAs.[16] Charge transfer from Ga to As results in an occupied electronic state centered at the As sites and an empty state centered at the Ga sites. Consequently, bias-polarity-dependent STM images selectively show the As atoms at negative sample bias and the Ga atoms at positive sample bias. A similar atom-selective imaging mechanism is found for other III–V semiconductor compounds.

Another example is given by the Si(111)–$(\sqrt{3} \times \sqrt{3})$Al system.[17,18] STM images of this surface obtained at positive sample bias reveal the $(\sqrt{3} \times \sqrt{3})$ lattice together with defect sites which appear less bright (Figure 5a). At negative sample bias, the contrast reverses (Figure 5b). Based on the observed dependence of the defect density on Al coverage, it was concluded that the bright protrusions seen at positive bias must correspond to Al adatoms, whereas the defect sites were attributed to Si adatoms substituting for Al in the $(\sqrt{3} \times \sqrt{3})$ structure below an Al coverage of one-third of a monolayer. The Si adatoms give rise to an extra dangling-bond defect state near −0.4 eV, which causes the Si adatoms to appear brighter than the Al adatoms for negative sample bias. Based on the polarity-dependent contrast, an atom-selective imaging can be achieved for the Si(111)–$(\sqrt{3} \times \sqrt{3})$Al system (similar to that for the GaAs(110) surface). For the general case of an arbitrary multicomponent solid, however, the details of charge transfer between different atomic species are usually not known and therefore an assignment of ''topographic'' features in STM images to particular atomic species cannot be made without additional information about the system.

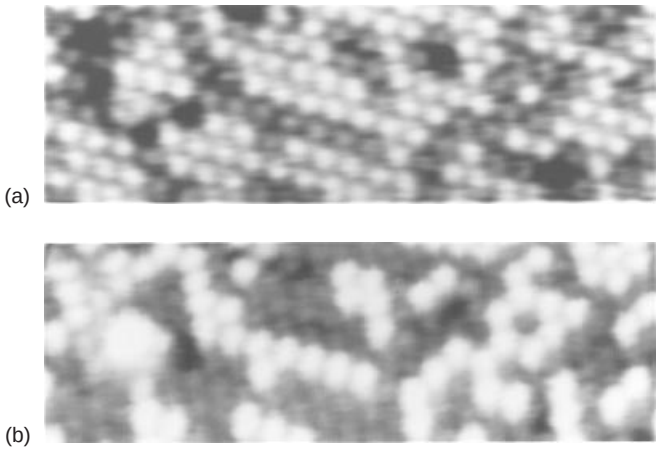

Figure 5 STM images of the same region of Si(111) with a sixth of a monolayer Al coverage. The local nature of the $\sqrt{3} \times \sqrt{3}$ bonding permits selective imaging of (a) Al adatoms at positive sample bias (+2 V) and (b) Si adatoms at negative sample bias (−2 V). (Image courtesy of R.J. Hamers.)

Alternatively, the spectroscopic capabilities of STM have been exploited to differentiate between different chemical species in multicomponent metal systems based on characteristic features in the local tunneling spectra arising, e.g., from surface states, surface resonances or adsorbate-induced states. An example is presented in Figures 6(a–c) showing a submonolayer Fe film grown at room temperature on a stepped W(110) substrate.[19,20] The topography (Figure 6a) is dominated by iron islands of monolayer height and a lateral extent of about 5 nm, although a small amount of the deposited material tends to decorate the monoatomic steps of the W

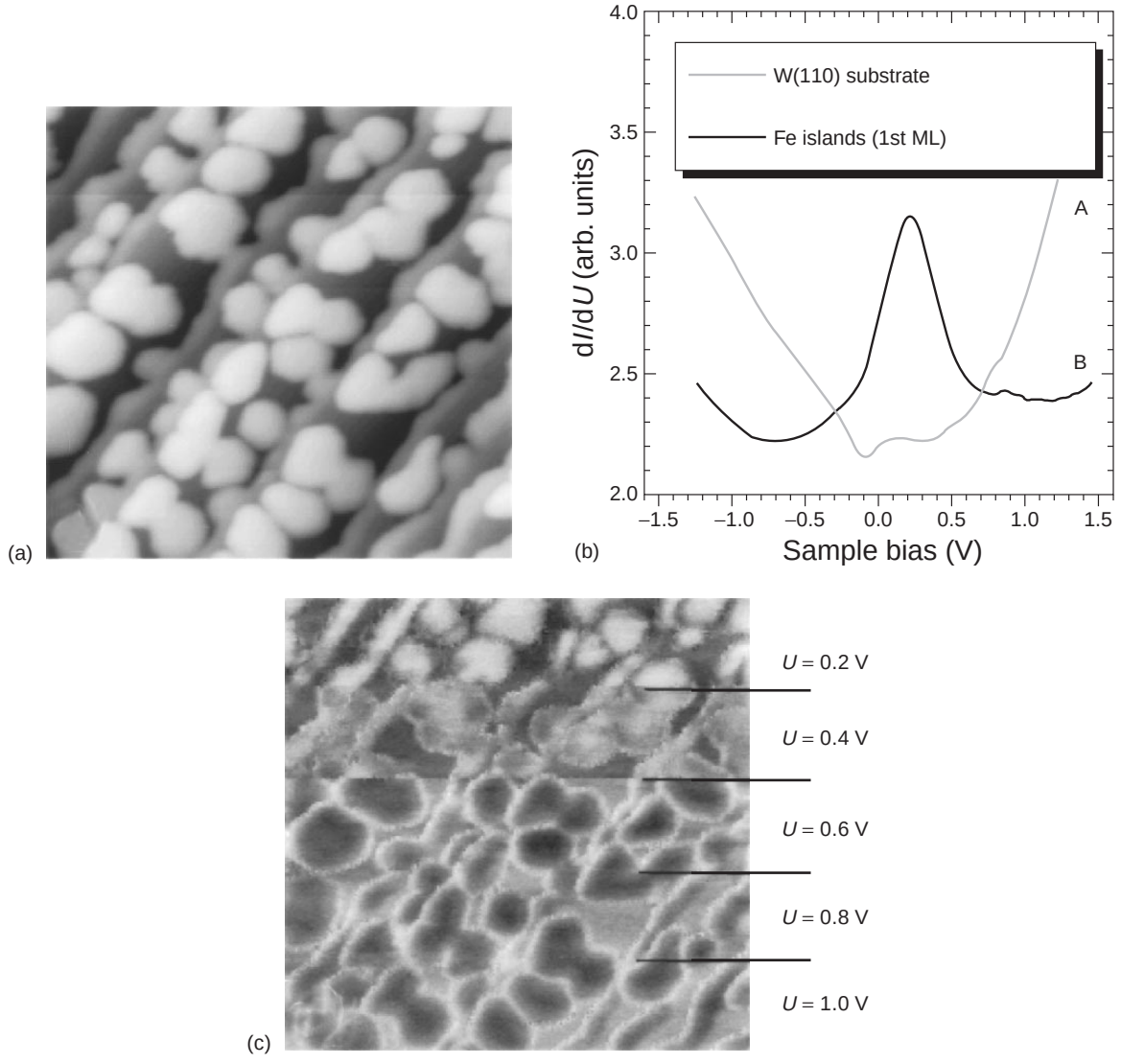

Figure 6 (a) STM image of half a monolayer of Fe on a W(110) substrate (scan size: 250 × 250 nm). The deposited material forms monolayer islands and a small amount decorates the step edges. (b) Averaged local spectra of the differential conductivity dI/dU taken above Fe and W sites from the former image. Obviously one monolayer of Fe on W(110) exhibits an empty-state peak which is centered at $U = 0.22$ V while the spectrum of the bare W(110) substrate is relatively featureless in this bias regime. (c) Map of the differential tunneling conductivity dI/dU (x, y). During the scan from bottom to top, the gap voltage was decreased every 80 lines by 0.2 V. The contrast between Fe and W flips between 0.6 and 0.4 V. While W has the higher differential conductivity (bright areas) in the range $0.6\,\text{V} < U < 1.0\,\text{V}$ (lower part of the image), the opposite is true in the upper part.

substrate. A distinction between Fe and W is difficult to make based on such constant-current topographs, particularly close to step edges where the difference in the apparent topographic height between Fe and W almost vanishes. Therefore, chemical-specific fingerprints are needed which can be extracted from the local tunneling spectra measured either above the bare W(110) substrate (Figure 6b, spectrum A) or above the Fe islands (Figure 6b, spectrum B). While the dI/dU–U characteristic measured above the Fe islands exhibits a pronounced peak centered at $U = +0.2$ V corresponding to an empty d-state, the spectrum measured above the bare W(110) substrate shows no significant feature in this bias voltage regime. Based on the pronounced difference between the tunneling spectra measured above Fe and W, a chemical-specific imaging of the ultrathin film system can be achieved by spatially resolved measurements of the differential tunneling conductivity dI/dU (x, y) for selected values of the applied dc bias voltage. This is demonstrated in Figure 6(c), which shows a map of dI/dU (x,y) where the applied bias voltage was changed every 80 scan lines in steps of $\Delta U = 0.2$ V. At

$U = +0.2$ V (Figure 6c, top part) the Fe islands appear bright compared to the W substrate corresponding to the high differential tunneling conductivity of Fe at $U = +0.2$ V as known from the local tunneling spectrum (Figure 6b, curve B). On the other hand, at $U = +1.0$ V (Figure 6c, bottom part) the Fe islands appear dark compared with the W substrate, which is explained by the much lower differential tunneling conductivity of Fe compared with W for high sample bias voltage (Figure 6b). Accordingly, a contrast flip in the dI/dU (x, y) image (Figure 6c) is observed between $U = +0.2$ and $+1.0$ V, corresponding to the crossing point of the two local tunneling spectra presented in Figure 6(b). To obtain high chemical contrast in the dI/dU (x,y) map, it is therefore important to select an appropriate dc bias voltage for which the difference in the differential tunneling conductivity between different chemical species appears to be the highest. This information can easily be extracted from local $dI/dU-U$ spectra as demonstrated for the Fe/W(110) system in the present case.

The method for chemical-selective imaging of multicomponent metal surfaces at the nanometer scale outlined above is applicable quite generally. It is known that most metal surfaces exhibit some pronounced density-of-states feature within the bias voltage regime accessible by STM. Other examples include Gd(0001),[21,22] or Fe(001)[23,24] and Cr(001)[23] for which pronounced d-like electronic states close to the Fermi level have been observed by tunneling spectroscopy. As an important application, the formation of a Cr–Fe alloy could be studied at the atomic level (Figures 7a and b), resulting from the growth of submonolayer Cr on Fe(001) at elevated temperatures (290 °C).[25]

One of the limitations in the application of tunneling spectroscopy to chemical-selective imaging based on pronounced density-of-states features of the sample surface is given by occasionally strong effects of tip states on the spectroscopic results, particularly with regard to the occupied-state density probed at negative sample bias voltages.[15] On the other hand, special tip states obtained, e.g. by an adsorbate at the front end of the tip, can sometimes be very useful for chemical-selective imaging at the atomic level. This was first demonstrated by a direct observation of surface chemical order on a $Pt_{25}Ni_{75}(111)$ surface[26] based on a difference in apparent topographic height between Pt and Ni atoms of 0.3 Å as seen under special tunneling conditions (Figure 8). It was proposed that an adsorbate on the tip can explain the chemically resolved images at extremely low bias voltage because an adsorbate, such as oxygen, may tend to form a chemical bond more likely with one element (in this case Ni) than with the other (Pt), leading to the conclusion that the bright species in Figure 8 are Ni atoms. A similar discrimination between

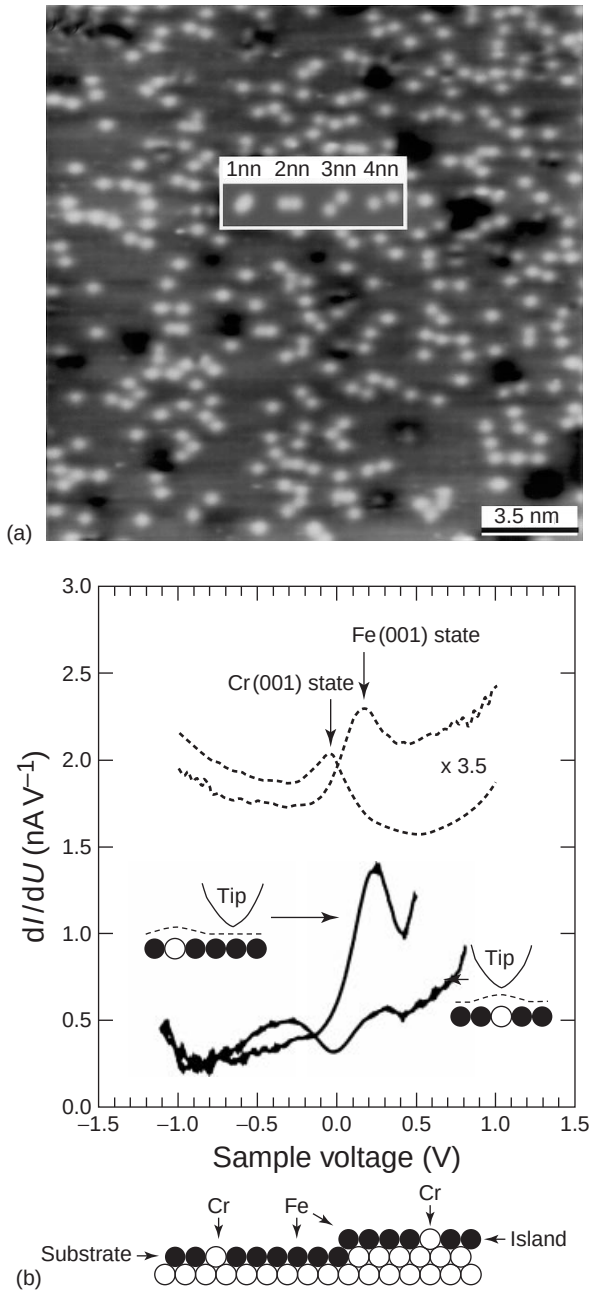

Figure 7 (a) STM image of a submonolayer Cr coverage on a Fe(001) single-crystal substrate showing the alloyed Cr impurity atoms. Simulated images of Cr pairs separated by first, second, third and fourth nearest-neighbor (nn) distances are shown in the inset for comparison. (b) Tunneling conductance spectra from clean Cr(001) and Fe(001) surfaces (a) and submonolayer Cr on Fe(001) alloyed surfaces (b). A model of the Cr–Fe alloy is shown in the lower part of the figure. (Image courtesy of J. Stroscio.)

different elements based on a special tip state was also demonstrated for several other metal alloy systems[27,28] and for oxygen-induced reconstructions of Cu and Ni surfaces.[29,30]

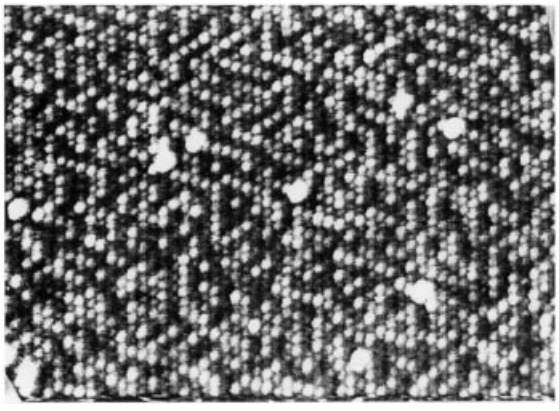

Figure 8 Chemically resolved STM image of a (111) surface of a $Pt_{25}Ni_{75}$ single crystal (scan size: 12.5 × 10 nm). The dark and bright atomic species correspond to Pt and Ni, respectively. The surface composition is different from the bulk composition. (Reproduced by permission from P. Varga.[26])

Chemical contrast on metal surfaces can also be achieved via spectroscopic STM studies of image states.[31–33] Image states can provide even higher elemental contrast than do surface states because they are narrower and have a higher amplitude. Figures 9(a) and (b) show an example of submonolayer Cu on W(110). The Cu stripes grown at the W(110) substrate steps cannot be distinguished in the topographic STM image. However, they clearly show up if the sample bias voltage is tuned to an image-state resonance of Cu. The difference picture subtracts out the topography and provides an elemental map of the surface. Since the image state positions are depending on the surface work function, the contrast obtained reflects the magnitude of the work function difference between the different elemental species.

2.3 Local Work Function Mapping

Instead of using image-state resonances which require high bias voltages ($U > 5\,\mathrm{V}$) leading to an increase in tip–surface separation and therefore to a reduced spatial resolution, it is also possible to map out local work function variations by measuring the spatially resolved local tunneling barrier height ϕ.[34] Motivated by Equation (10), $I \propto \exp(-2\kappa s)$, with (Equation 20)

$$\kappa = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \tag{20}$$

for a rectangular potential barrier of effective height $(V_0 - E)$, a local barrier height ϕ can be introduced in the general case by Equation (21):

$$\phi \propto \frac{\hbar^2}{8m}\left(\frac{\mathrm{d}\ln I}{\mathrm{d}s}\right)^2 \tag{21}$$

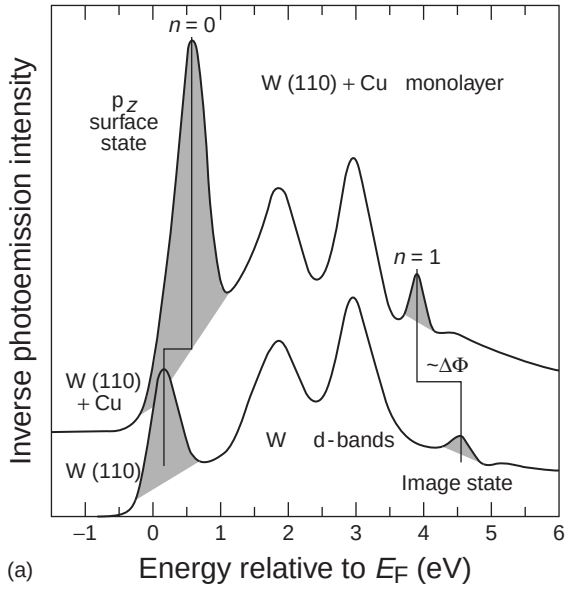

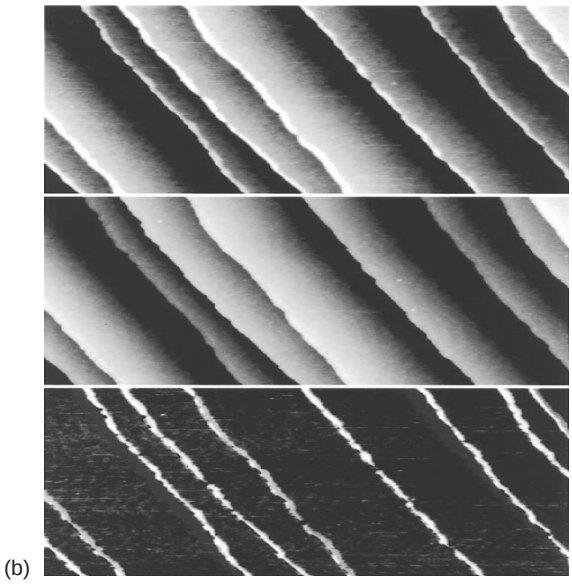

Figure 9 (a) Surface states on clean and Cu-covered W(110). A p_z-like surface state near the Fermi level and an image state are seen in both cases. The image state is used to obtain elemental contrast in STM images (b). By tuning into a Cu-specific image state, the Cu patches appear bright (top) compared with a simple topographic image showing the step-and-terrace structure (middle). The difference between the two pictures (bottom) provides an elemental map of the surface with the topography removed. Scan size: 200 × 200 nm. (Reproduced by permission from T. Jung.[33])

Consequently, by modulating the tip–surface separation s by ds while scanning at a constant average tunneling current I, the modulation of the ln I signal provides a direct measure of the local barrier height. For a large tip–surface separation, the local barrier height reflects the surface local work function ϕ_s which has a chemical component

determined by the chemical nature and structure of the solid only, in addition to an electrical component determined by the chemical nature of the solid and its surface crystallographic orientation. Therefore, spatially resolved measurements of ϕ can be used to map chemical inhomogeneities at surfaces [Figures 10(a) and (b)] and also different crystallographic facets of small crystallites.[35]

On the atomic scale, it is more appropriate to relate the measured local barrier height with the decay rates of the wave functions describing the sample surface and the tip. Lateral variations of ϕ then have to be interpreted as lateral variations in the decay rate of the surface wave function. Figure 11 shows an example of different atomic-scale superstructures of Gd on W(110) at submonolayer coverage. The atomic-scale variation of the measured local barrier height reflects the atomic-scale variations of the surface wave-functions whereas the spatial dependence of the average value of the measured barrier height reflects the difference in the local work function for the different types of superstructures.[36]

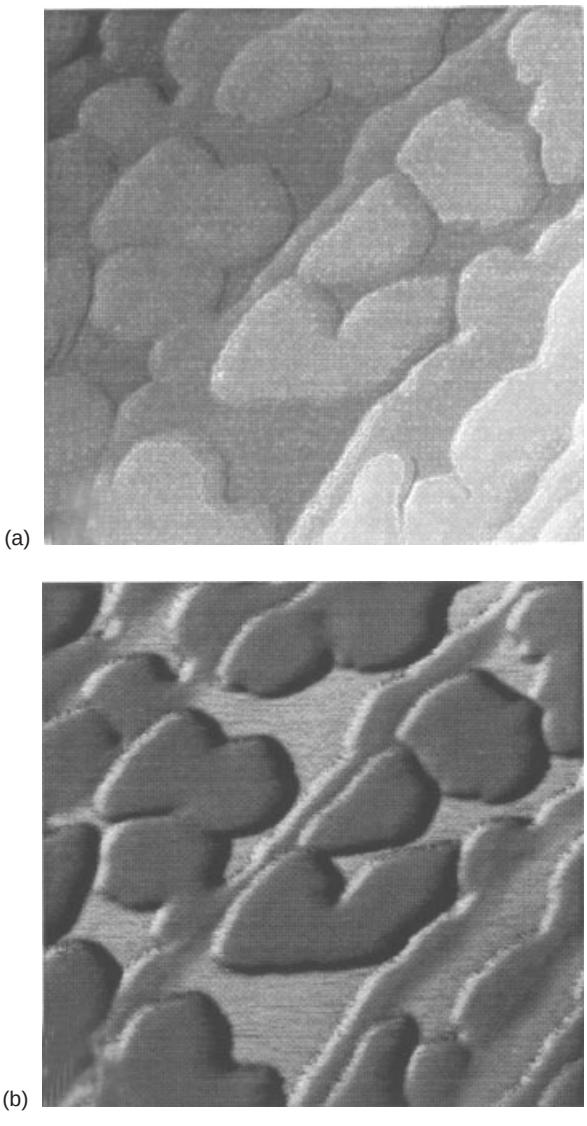

Figure 10 (a) Topographic STM image and (b) a corresponding map of the local tunnel barrier height for half a monolayer of Fe on W(110). Owing to the lower local work function of Fe(110) compared with W(110), the Fe patches close to the step edges of the W(110) substrate can easily be distinguished in the local barrier height map (scan size: 30 × 30 nm).

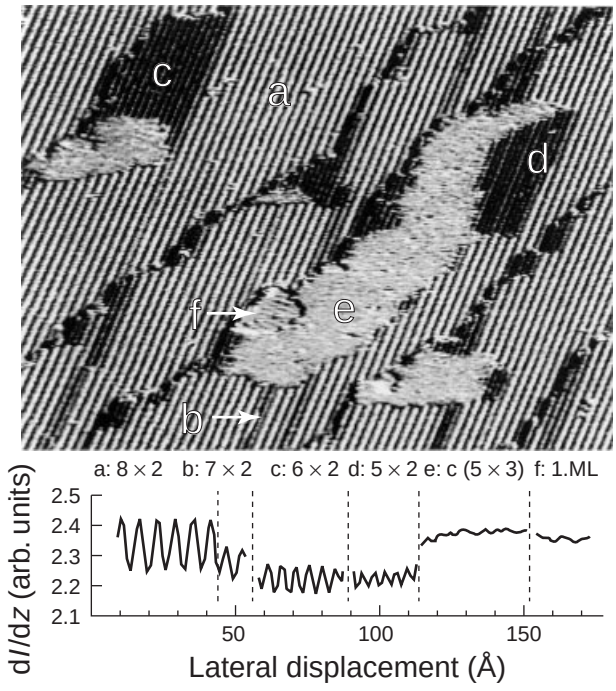

Figure 11 Spatially resolved map (80 × 60 nm) of the local tunnel barrier height for different superstructures of Gd on W(110) in the submonolayer coverage regime. The lowest barrier height is measured above (5 × 2) and (6 × 2) superstructures, in agreement with macroscopic work function measurements. The bottom part of the figure displays section lines through the barrier height map for the different superstructures.

2.4 Thermovoltage Mapping

Another contrast mechanism which has been exploited for chemical mapping is based on the thermovoltage across the tunneling barrier which is built up if a temperature difference between tip and sample exists.[37] In case of different Fermi distributions of tip and sample, a net thermally driven current will result even for zero bias according to Equation (4). However, if the external current is zero, an electric charge is induced associated with a bias voltage leading to a tunneling current which is opposed to the thermally driven current. At equilibrium the two cancel exactly. However, spatial variations of the electronic structure may modify the balance which is responsible for the finite measured thermovoltage across the vacuum barrier. To first approximation the

thermovoltage U_{th} is given by Equation (22):[37]

$$U_{th} = \frac{\pi^2 k_B^2}{6e}(T_t^2 - T_s^2) \quad [\ln n_t]' + [\ln n_s]' + \frac{s}{\hbar}\left(\frac{2m}{\phi}\right)^{1/2} \Bigg\} \tag{22}$$

where k_B is Boltzmann's constant, T_t and T_s are the temperatures of tip and sample and s is the tip–sample separation. The first and second terms in Equation (22) contributing to the thermovoltage are determined by the logarithmic derivatives of the electronic density of states of tip and sample whereas the third term depends on the tip–sample distance and the local work function ϕ.

As an example, the early stages of growth of Ag on Au(111) have been studied by thermovoltage imaging. Figures 12(a) and (b) show the topography (a) and the simultaneously obtained thermovoltage map of submonolayer Ag on the Au substrate (b). Based on the topographic image it cannot be determined to what extent the Au substrate is covered by Ag. On the other hand, the thermovoltage map provides a clear contrast between the two different elements with the Au substrate appearing bright (thermovoltage of $-25\,\mu V$). Obviously, some Ag has diffused towards the step edges of the Au(111) substrate which is hardly seen in the topographic image.

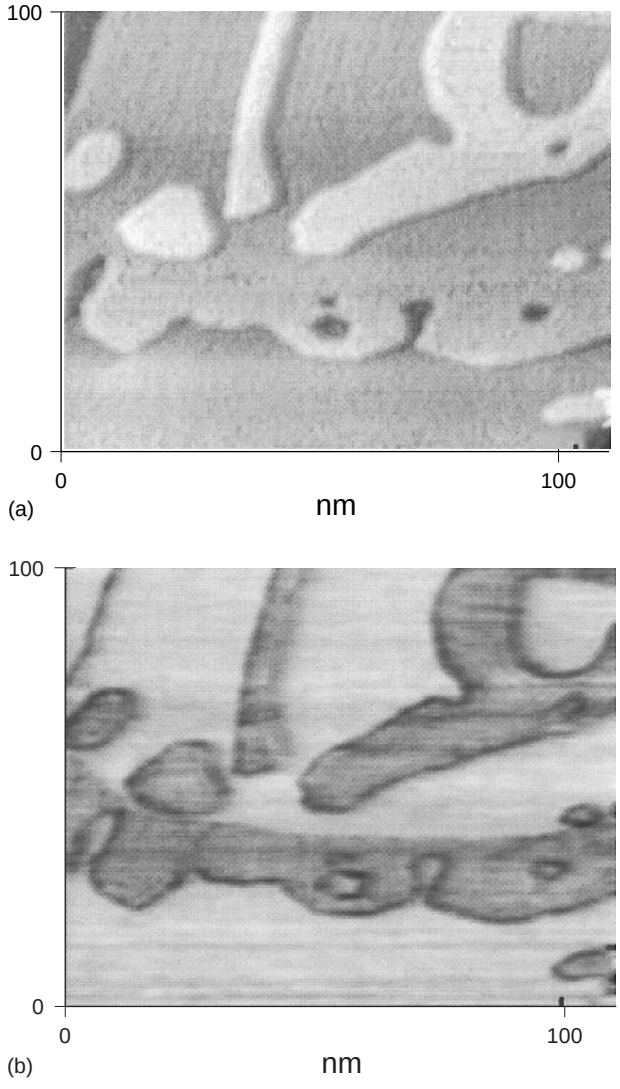

Figure 12 (a) Topographic STM image and (b) corresponding map of the thermovoltage for a fraction of a monolayer of Ag on Au(111) (scan size: 110×110 nm). The Ag patches close to the step edges are easily visible as dark areas in the thermovoltage map while they are hardly distinguished in the topographic image. (Reproduced by permission from R. Möller.[37])

2.5 Spin-polarized Scanning Tunneling Spectroscopy

So far, we have considered the dependence of the tunneling current on the tip–sample separation s, the sample bias voltage U and the local barrier height ϕ. However, we have not yet considered the spin of the tunneling electrons and the additional spin dependence of the tunneling current if magnetic electrodes are involved (Equation 23):

$$I = I(s, U, \phi, \uparrow) \tag{23}$$

By exploiting the spin dependence of the tunneling current, magnetic contrast down to the atomic level can be achieved.[38–40]

In the case of two ferromagnetic electrodes for which the directions of the internal magnetic fields differ by an angle θ, it can be shown within a free-electron model and in the limit of small applied bias voltage that the tunneling conductance σ is given by Equation (24):[38]

$$\sigma = \sigma_{fbf'}(1 + P_{fb}P_{f'b}\cos\theta) \tag{24}$$

where, P_{fb} and $P_{f'b}$ denote the effective spin polarizations of the ferromagnet–barrier interfaces and $\sigma_{fbf'}$ is a mean conductance which is proportional to $\exp(-2\kappa s)$. It is easily verified that the effective polarization for the whole tunnel junction introduced by Equation (25):

$$P_{fbf'} := P_{fb} \cdot P_{f'b} \tag{25}$$

becomes Equation (26):

$$P_{fbf'} = \frac{\sigma_{\uparrow\uparrow} - \sigma_{\uparrow\downarrow}}{\sigma_{\uparrow\uparrow} + \sigma_{\uparrow\downarrow}} \tag{26}$$

where $\sigma_{\uparrow\uparrow}$ and $\sigma_{\uparrow\downarrow}$ denote the tunneling conductances in the case of parallel ($\theta = 0°$) and antiparallel ($\theta = 180°$) alignment of the internal magnetic field directions, respectively. Spin-polarized scanning tunneling microscopy

(SPSTM) performed with ferromagnetic probe tips on magnetic samples allows the mapping of spatial variations of the effective spin polarization of the tunnel junction as well as spatially resolved studies of the spin-resolved electronic density of states of the sample surface provided that the spin-resolved electronic density of states of the probe tip is known (at least in the energy window relevant for tunneling). SPSTM even allows the discrimination between different magnetic ions at the atomic scale.[39,40] This was first demonstrated by the example of Fe_3O_4 (magnetite) for which the Fe^{2+} and Fe^{3+} sites in the (001) plane of octahedrally coordinated Fe sites of Fe_3O_4 could be distinguished by using a ferromagnetic Fe probe tip [Figures 13(a) and (b)]. Since SPSTM is sensitive to the surface magnetization rather than the magnetic stray field, as for magnetic force microscopy,[7] it can even be applied to antiferromagnetic samples.[41] Another important application of SPSTM is the observation of magnetic domains and domain walls on a nanometer scale, as was first demonstrated for ultrathin Gd films.[42] In this case, spin-polarized vacuum tunneling into an exchange split surface state allowed a clear separation of topographic, electronic and magnetic structure information.

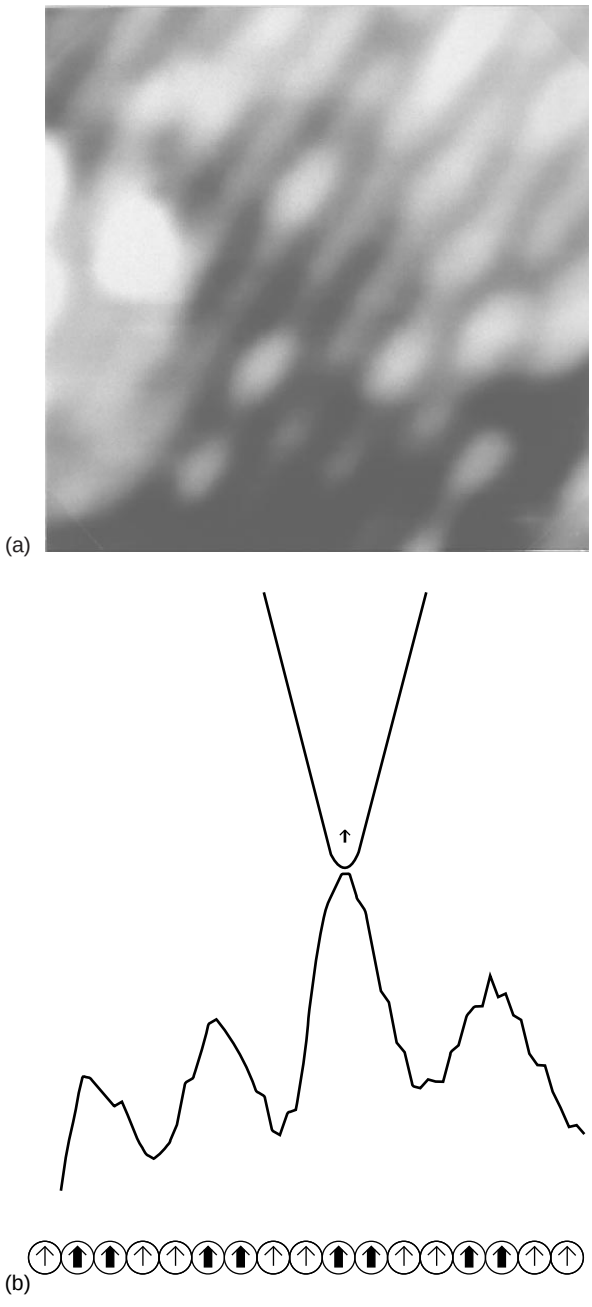

Figure 13 Spin-resolved STM image of the Fe–O(001) plane of magnetite obtained with an Fe probe tip. The rows of octahedrally coordinated Fe sites of spacing 6 Å and with a modulation along these rows can clearly be seen. A period of 12 Å is preferentially observed along the Fe rows as shown in the line section (b). The period of 12 Å corresponds to the repeat pattern of Fe^{2+} and Fe^{3+} in the Fe–O(001) plane of magnetite.

2.6 Inelastic Electron Tunneling Spectroscopy

In addition to elastic tunneling processes, in which the energy of the tunneling electrons is conserved, electrons may also tunnel inelastically due to interactions with elementary excitations such as molecular vibrations or collective lattice vibrations (phonons), as schematically depicted in Figures 14(a–d).

For molecular species IETS[43,44] combined with STM might become a powerful technique to identify functional groups. A first experiment toward this goal was performed with sorbic acid adsorbed on graphite.[45] A series of strong peaks was observed in the tunneling spectrum with the energetic positions of the peaks corresponding approximately to the vibrational modes of the molecule. In another experiment,[46–48] IETS traces of an isolated acetylene (C_2H_2) molecule adsorbed on a Cu(100) surface showed an increase in tunneling conductance at an energy corresponding to the excitation of the C–H stretch mode [Figures 15(a) and (b)]. Additionally, an isotopic shift was observed for deuterated acetylene (C_2D_2), which supported the interpretation of the observation of inelastic tunneling strongly. Spatially resolved maps of the inelastic tunneling channels allowed the distinction and characterization of the two isotopes. It is believed that vibrational microscopy will have great potential for the determination of the identity and the arrangement of functional groups within a single molecule, and also for the characterization of small molecules in more complex chemical or biological environments.

Signatures of inelastic electron tunneling processes can also be observed by the detection of photons emanating

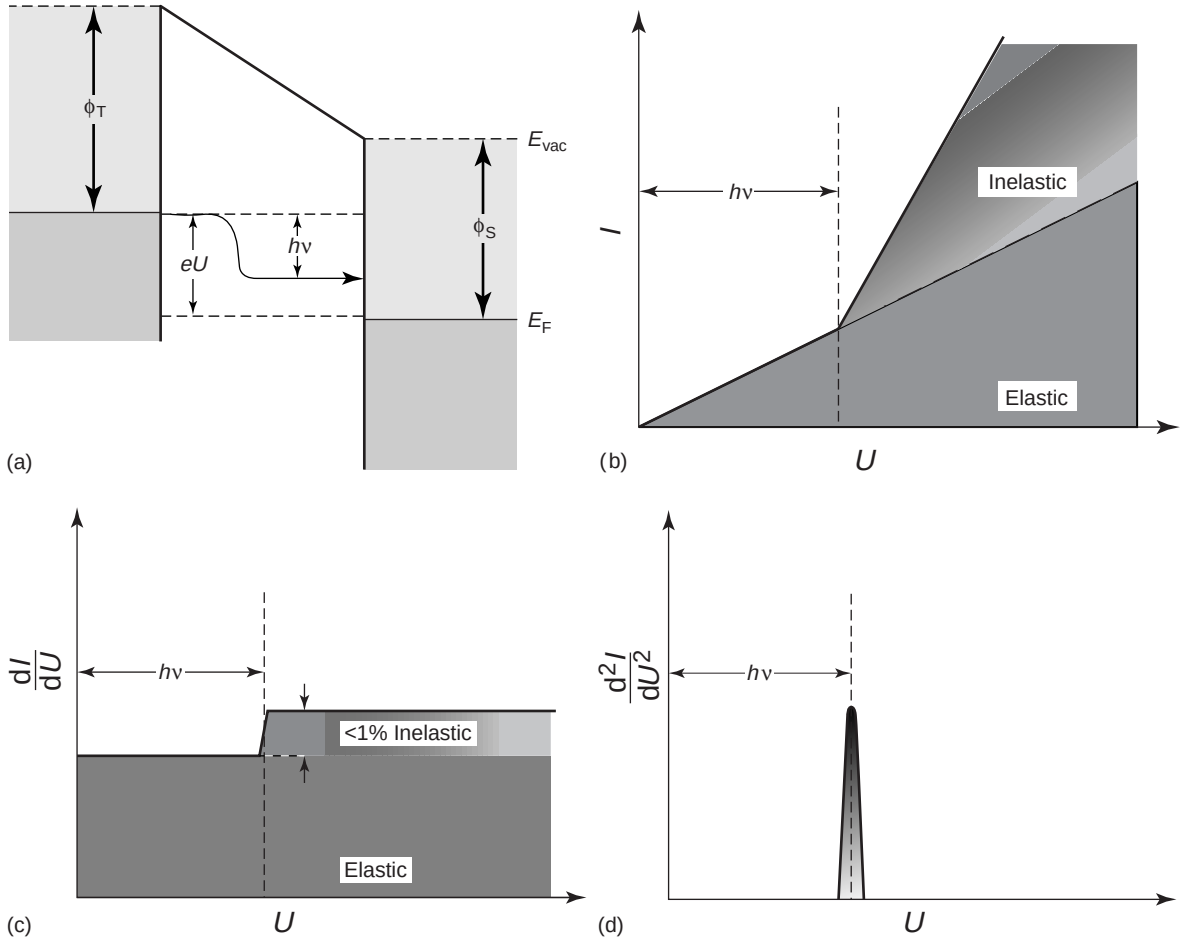

Figure 14 Tunneling electrons can excite a molecular vibration of energy $h\nu/e$. The increase in conductance at this threshold is typically small (<1%). Therefore, inelastic electron tunneling spectroscopy (IETS) is usually performed by measuring the second derivative of the current–voltage curve: $d^2I/dU^2(U)$.

from an STM-type tunnel junction.[49] The combination of STM with optical methods offers many attractive features with regard to spectroscopic contrast mechanisms which may add ''true color'' to the STM images.[50]

2.7 Scanning Tunneling Spectroscopy with Laser Excitation

STM with simultaneous laser excitation can reveal chemical contrast on multicomponent semiconductor surfaces down to the atomic level.[51,52] The use of coordinated electronic and photon biasing has been demonstrated to provide atom-specific imaging, e.g. for $CuInSe_2$, a covalently bonded semiconductor with applications for thin-film photovoltaic technologies. It is assumed that the photon bias at a particular wavelength directly excites the atom, with the wavelength corresponding to the atomic transition probability. This method has been applied to study the local chemical order in the vicinity of defects, such as grain boundaries.

2.8 Subsurface Information by Scanning Tunneling Microscopy on Semiconductors

In contrast to metals, electric fields in semiconductor samples are not screened within a near-surface region but penetrate several nanometers into the sample. Therefore, subsurface information, e.g. about the distribution and electronic properties of dopants, can be obtained. As a rule of thumb, the depth up to which dopants are visible in STM images is given by the screening length in the semiconductor which itself is determined by the dopant concentration.

Figure 16 shows an atomically resolved STM image of a cleaved n-InAs(110) surface. Atomic rows of the As sublattice are visible because a negative sample bias was applied (see section 2.2). In addition, a 0.5 Å high elevation with a full width at half-maximum (fwhm) of 10 nm is observed. Such elevations are caused by charged dopants,[53,54] which can easily be verified since the number of the elevated features scales with the dopant

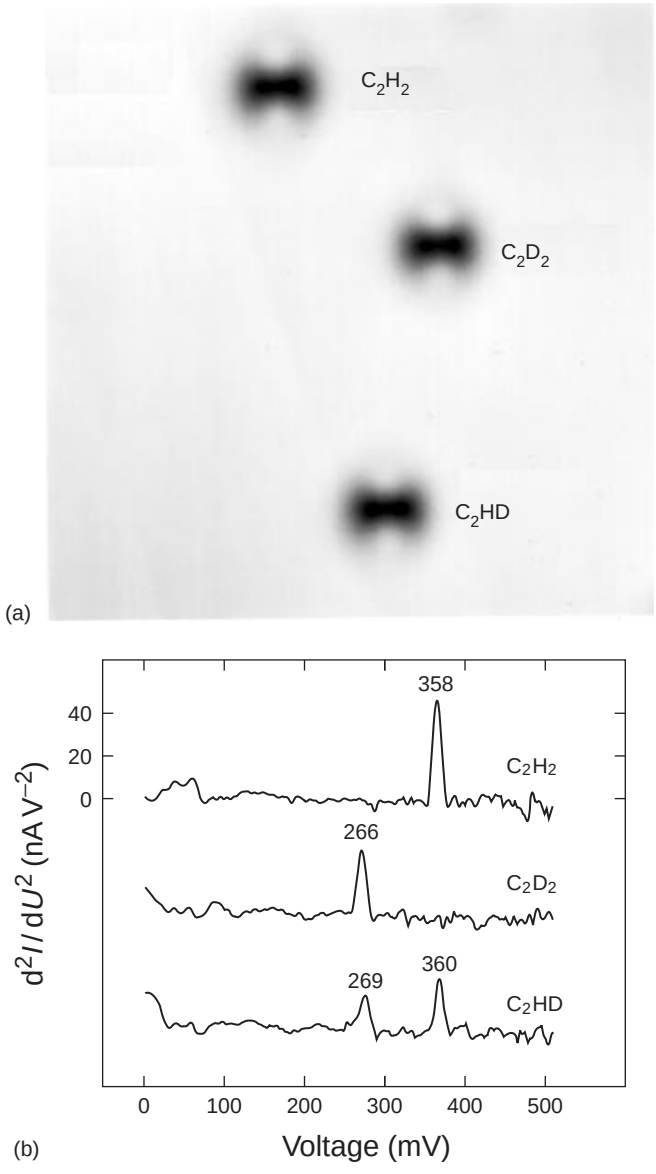

Figure 15 (a) STM image (56 × 56 Å) and (b) single-molecule vibrational spectra of three acetylene isotopes on Cu(100) at 8 K. The inelastic tunneling spectra are the average of 16 scans. Background spectra over the bare substrate surfaces have been subtracted. (Reproduced by permission from W. Ho.)

concentration. A simple model is able to reproduce the elevations qualitatively.[55] It assumes that the screened Coulomb potential of the ionized dopants leads to a local band bending which shifts the LDOS $n_s(E)$ of the sample according to Equation (27):

$$n_s(E, x, y) \propto \sqrt{E - E_{CBM}(x, y)} \qquad (27)$$

where E_{CBM} is the energy of the local conduction band minimum. The resulting apparent height change Δz in constant-current images can be calculated to a first

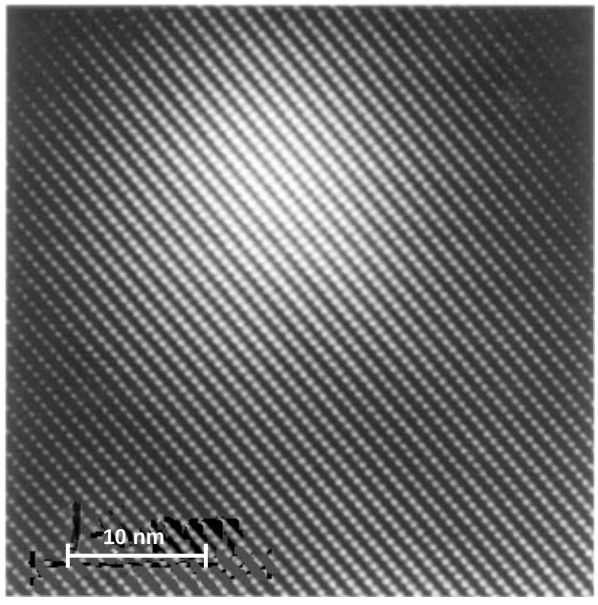

Figure 16 STM image of a cleaved n-InAs(110) surface showing the As sublattice of the top surface plane and a broad elevated feature originating from the perturbed LDOS due to the presence of a subsurface dopant (scan size: 25 × 25 nm).

approximation by solving Equation (28):

$$I \propto \int_0^{eU} n_s(E, x, y) \times \exp\left(-2 \sqrt{\frac{2m}{\hbar^2}\left(\phi + \frac{eU}{2}\right)}\, z(x, y)\right) \mathrm{d}E \qquad (28)$$

where ϕ is the local tunneling barrier height, which has to be determined independently (see section 2.3).

By making use of the spectroscopic capabilities of the STM (see section 2.1) at low temperatures it is even possible to observe the scattering of electron waves at the attractive potential of ionized dopants,[56] as shown in Figure 17.

3 SCANNING NEAR-FIELD OPTICAL SPECTROSCOPY

SNOM combined with optical spectroscopy probably offers the richest variety of experimental methods for chemical-selective imaging by scanning probe methods (SPM) in the nanometer-scale regime.[4,5] It is well known that the spatial resolution achieved in classical optical microscopy is limited by diffraction to about $\lambda/2$, where λ is the optical wavelength. This limit arises because the evanescent waves carrying the information about the high-spatial-frequency components of the diffracted

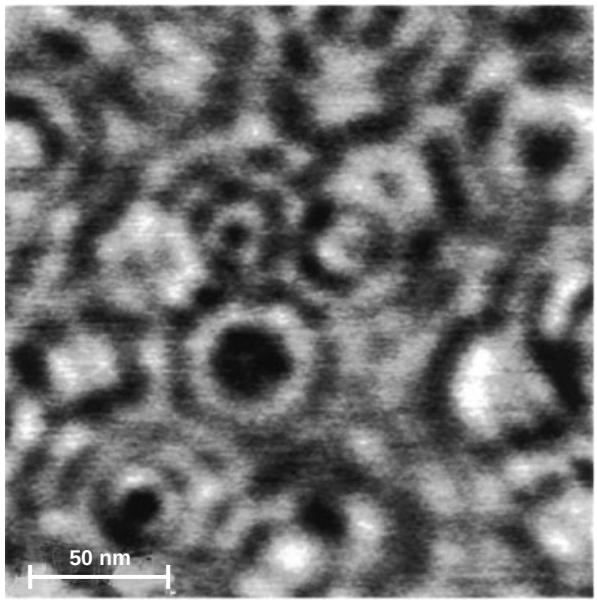

Figure 17 Spectroscopic dI/dU image of a cleaved n-InAs (110) surface obtained at low temperature (8 K) showing interfering concentric ring structures originating from the scattering of electron waves at ionized dopants. The diameter of the ring structures depends on the electron wavelength and therefore on the energy determined by the applied bias voltage (scan size: 200 × 200 nm).

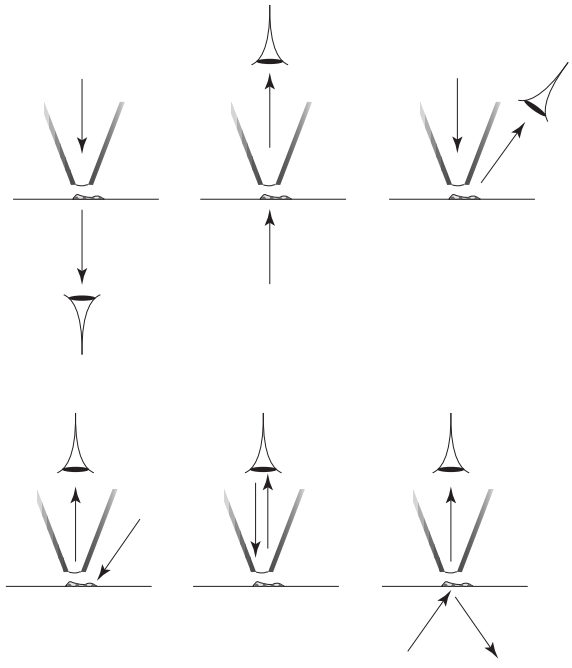

Figure 18 Different configurations for SNOM.

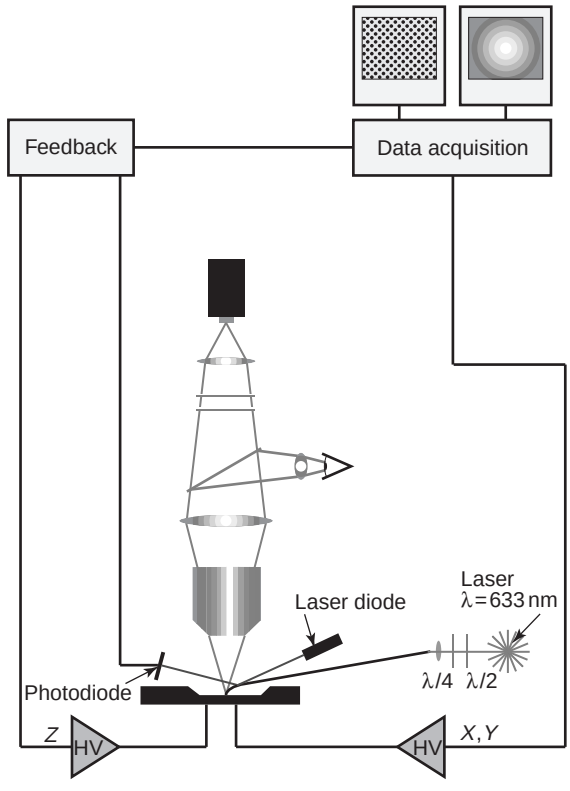

Figure 19 A schematic diagram of a scanning near-field optical microscope.

waves decay rapidly with distance s from the object. As a result information about subwavelength features of the object cannot be retrieved in the far-field regime of classical optics. On the other hand, by approaching a sharp optical probe within the near-field regime ($s < \lambda$), the diffraction limit of classical optics can easily be surpassed.

In the first SNOM experiments a tiny aperture, illuminated by a laser beam from the rear side, was scanned across a sample surface and the intensity of the light transmitted through the sample was recorded. The optical probes are usually formed by a sharpened glass or fiber tip coated with a thin metallic layer, leaving a submicrometer aperture at the apex. The capability of fabricating nanometer size apertures is essential to obtain high spatial resolution which, however, is gained at the expense of signal intensity. As an alternative to glass or fiber tips, sharpened micropipets coated with a thin metallic layer can serve as optical probes as well. The micropipet apertures can either be illuminated from the rear side by a laser beam, acting as a light source, or they can be used as collectors for radiation from a small area of the sample which itself is illuminated as a whole. To enhance the transmission of electromagnetic energy through the narrow micropipet tube, it has been proposed to fill the micropipet with a fluorescent dye.

SNOM can be performed in reflection, in addition to transmission, which is of great practical importance (Figure 18). Furthermore, SNOM can be combined with all techniques known in classical optical microscopy

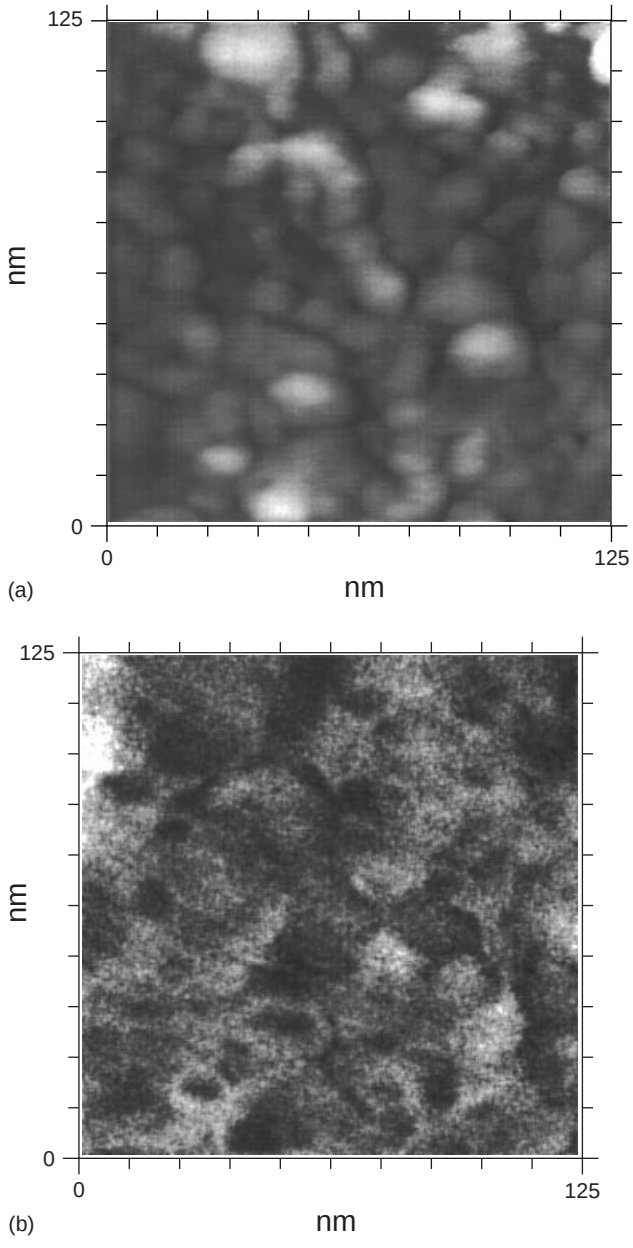

Figure 20 (a) Topographic STM image of a mixed film of silver and gold and (b) simultaneous SNOM image obtained with a tetrahedral tip (scan range: 125 × 125 nm). The contrast between silver and gold grains can be explained by the difference in the dielectric constants of silver and gold at optical frequencies. (Reproduced by permission from U. Fischer.[20])

including the exploitation of luminescence and polarization and also phase contrast (Figure 19).

Chemical contrast by SNOM has recently been demonstrated down to the nanometer scale.[5] Figures 20(a) and (b) show an STM image (a) and a simultaneously obtained SNOM image of a mixed film of silver and gold (b). Grains as visible in the topographic STM image appear either dark or bright in the corresponding SNOM image which has been explained by the difference in the complex dielectric constants of silver and gold at optical frequencies. The transition from dark (Ag) to bright (Au) regions is limited to a width of less than 1 nm in this case.

Other important applications of SNOM include the detection, microscopy and spectroscopy of single molecules[57] and the local spectroscopic studies of individual semiconductor quantum well, quantum wire and quantum dot structures.[58]

4 FORCE MICROSCOPY-BASED TECHNIQUES

4.1 Theory and Methods

The SFM[6] instrument is based on a miniaturized force sensor consisting of a cantilever-type spring with integrated sharp tip. The force interaction between the tip and the sample after approaching each other causes the cantilever to deflect by Δz according to Hooke's law (Equation 29):

$$F = c\Delta z \tag{29}$$

where c is the spring constant. These deflections can be measured with high sensitivity by optical methods (laser interferometry or laser beam deflection), achieving a resolution of 0.1 Å or better (Figure 21). For a given spring constant of the order of $1\,\mathrm{N\,m^{-1}}$, the

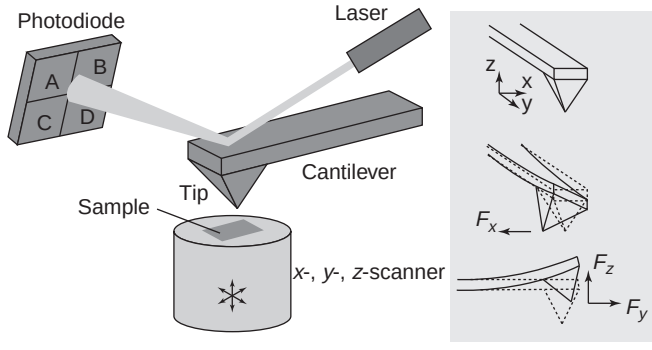

Figure 21 Principle of the scanning force microscope. Bending and torsion of the cantilever are detected simultaneously by measuring the lateral and vertical deflection of a laser beam while the sample is scanned in the x–y plane. The laser beam deflection is determined using a four-quadrant photodiode: (A + B) − (C + D) is a measure of the bending and (A + C) − (B + D) a measure of the torsion of the cantilever, if A, B, C and D are proportional to the intensity of the incident light of the corresponding quadrant. The torsion of the cantilever (middle) is solely due to lateral forces acting in x-direction, whereas both forces acting normal to the surface (F_z), in addition to acting in-plane in the y-direction (F_y), cause a bending of the cantilever (bottom).

measurable forces are as small as 10^{-11} N. A topographic map of a sample surface is obtained by keeping the interaction force constant while scanning the sample relative to the tip. SFM can be operated either in the contact regime, similar to the use of a stylus profilometer but with considerably lower forces, or in the noncontact regime. In the contact mode, short-range ion–ion repulsion forces are probed whereas in the noncontact mode SFM is exclusively sensitive to long-range forces, such as VDW or electrostatic forces. Two basically different modes of operation of SFM exist, i.e. quasistatic or dynamic operation mode.[7] In the quasistatic mode, the interaction forces are directly detected via measurements of the cantilever deflection Δz and the use of the relationship expressed by Equation (29). In the dynamic mode the cantilever is oscillated close to its resonance frequency and changes in the resonance frequency are measured resulting from tip–sample force interactions. Such changes can be explained to first order by a modification of the effective spring constant of the cantilever according to Equation (30):

$$c_{\text{eff}} = c - F' \tag{30}$$

where $F' = \partial F/\partial z$ is the gradient of the force acting between tip and sample. An attractive tip–sample force interaction with $F' > 0$ will soften the effective spring ($c_{\text{eff}} < c$), whereas a repulsive tip–surface force interaction with $F' < 0$ will strengthen the effective spring ($c_{\text{eff}} > c$). The change in the effective spring constant causes, in turn, a shift in the resonance frequency w of the cantilever according to Equation (31):

$$w = \left(\frac{c_{\text{eff}}}{m}\right)^{1/2} = \left(\frac{c - F'}{m}\right)^{1/2} = w_0\left(1 - \frac{F'}{c}\right)^{1/2} \tag{31}$$

where m is an effective mass and w_0 is the resonance frequency of the cantilever in the absence of a force gradient. An attractive force will therefore lead to a decrease of the resonance frequency ($w < w_0$) whereas a repulsive force will lead to an increase ($w > w_0$). However, Equation (31) assumes that the force gradient is constant over the whole distance range of the oscillating cantilever. This is usually not the case, leading to more complicated relations between frequency shift and tip–sample interaction.

In the case of frequency modulation (FM) detection of oscillation frequency changes, the minimum detectable force gradient F'_{min} is given by Equation (32):

$$F'_{\text{min}} = \frac{1}{A}\left(\frac{4ck_{\text{B}}TB}{w_0 Q}\right)^{1/2} \tag{32}$$

where A is the r.m.s. amplitude of the cantilever oscillation, Q is the quality factor of the cantilever and B is the detection bandwidth. To achieve the highest possible detection sensitivity for a given oscillation amplitude A and detection bandwidth B, the mass of the cantilever should be as small as possible and the quality factor as high as possible (e.g. by operation under high-vacuum conditions). In addition, operation at low temperatures is preferable. Under these conditions, forces well below 10^{-12} N can be measured.

Various modes of SFM operation have been exploited in order to achieve the goal of chemical-specific imaging. In the following, three examples for the contact regime and two examples for the noncontact regime will be discussed.

4.2 Lateral Force Microscopy

Besides measuring the forces normal to the sample surface by detecting the induced bending mode deflections of the cantilever in this direction, it is also possible to study the lateral (frictional) forces with SFM by detecting the induced torsion mode deflections of the cantilever while scanning.[7] According to Amontons' law, as derived for macroscopic contacts, the frictional force F_{f} should be proportional to the loading force F_{l} in the normal direction (Equation 33):

$$F_{\text{f}} = \mu F_{\text{l}} \tag{33}$$

where μ is the friction coefficient which is material dependent and therefore chemical specific.[59] However, for a microscopic tip–sample contact, as realized in SFM, Equation (33) is no longer valid and the dependence of the frictional force on the applied loading force becomes nonlinear (Figure 22j). As a result, the contrast in lateral force microscopy (LFM) depends on the loading force in addition to the chemical species involved.[60] This is illustrated in Figures 22(a–i), where the lateral or frictional force maps of a C_{60} thin film on a GeS(001) substrate are presented for a series of loads F_{l} applied to the area of contact between tip and sample. Dark colors in LFM images represent low friction force areas whereas bright colors represent high friction force areas.

The LFM maps in Figures 22(a–i), obtained with loads below 10 nN, exhibit lower frictional forces on the GeS substrate compared with the C_{60} islands. At approximately 10 nN (Figure 22f) this contrast vanishes and finally flips as the load rises [Figures 22g–i]. At high loads, the C_{60} islands therefore exhibit lower frictional forces compared with the GeS substrate. It is clear that the "chemical contrast" achieved in this model case critically depends on an external parameter, the applied loading force, and that an assignment of chemical species is not possible without understanding the details of the tip–sample contact mechanics. To study the origin of the

observed friction contrast reversal as a function of the applied load in more detail, the quantitative dependence of the frictional force on the applied loading force for the GeS substrate and the C_{60} islands have to be measured as shown in Figure 22(j). The behavior of the frictional force on the C_{60} islands (2/3-power law dependence) is markedly different compared with the behavior on the substrate (linear dependence). As a result, the two data sets exhibit a crossing point, which corresponds to the contrast reversal in the spatially resolved measurements.

Consequently, the interpretation of friction force maps in terms of chemical contrast requires a careful analysis of

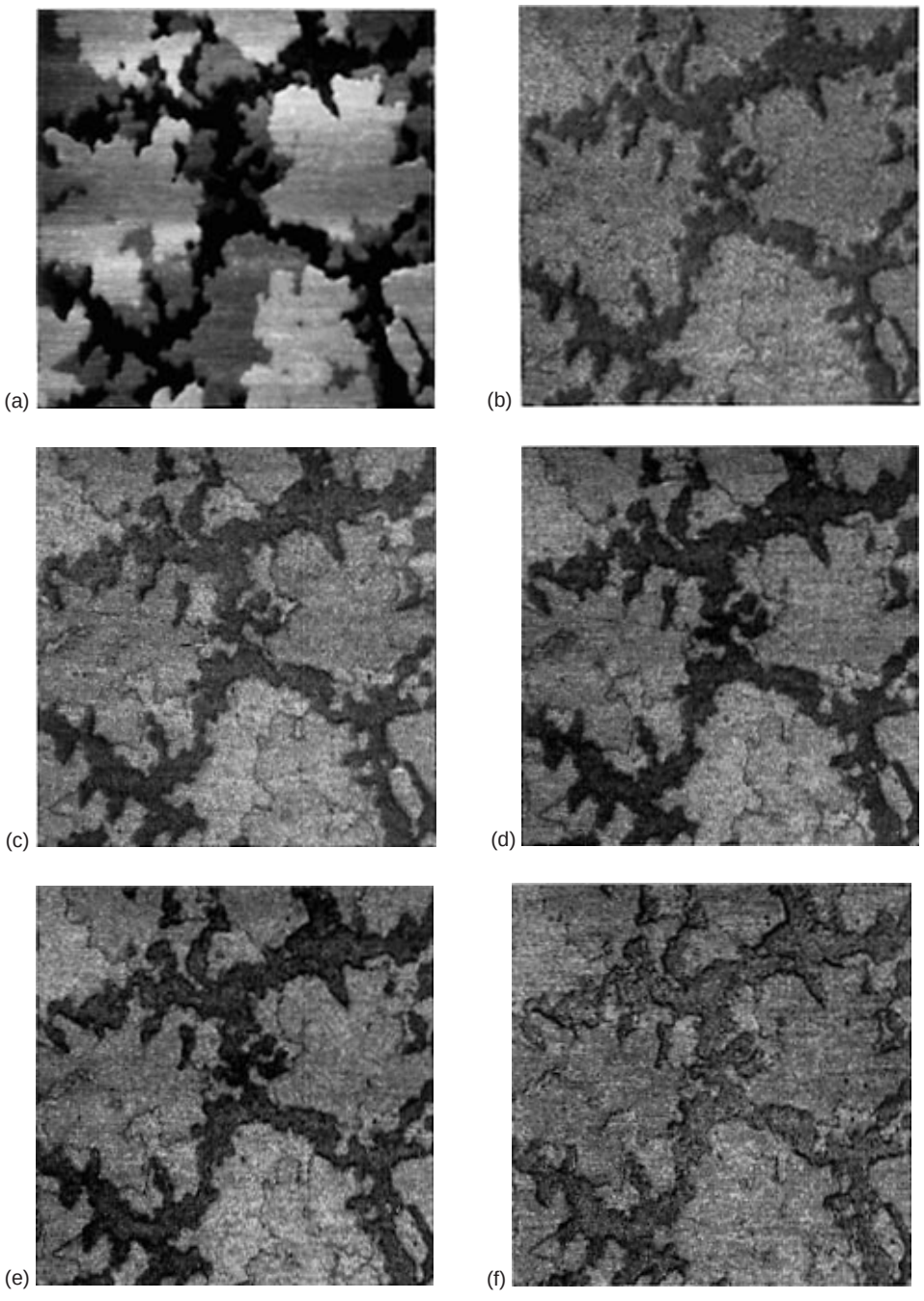

Figure 22 (a) Constant-force image of a 1.2 monolayer thin film of C_{60} on a GeS(001) substrate (scan size: $2 \times 2\,\mu m$) (b–i) Corresponding friction force maps for the same area at different loading forces of (b) 1.3, (c) 2.6, (d) 4.5, (e) 6.7, (f) 10, (g) 15, (h) 20 and (i) 30 nN. A friction contrast reversal is observed between images (f) and (g) which can be understood by a different friction-load dependence (j) for the C_{60} thin film and the GeS(001) substrate. (j) Friction-load dependence for a C_{60} thin film and a GeS(001) single-crystal substrate.

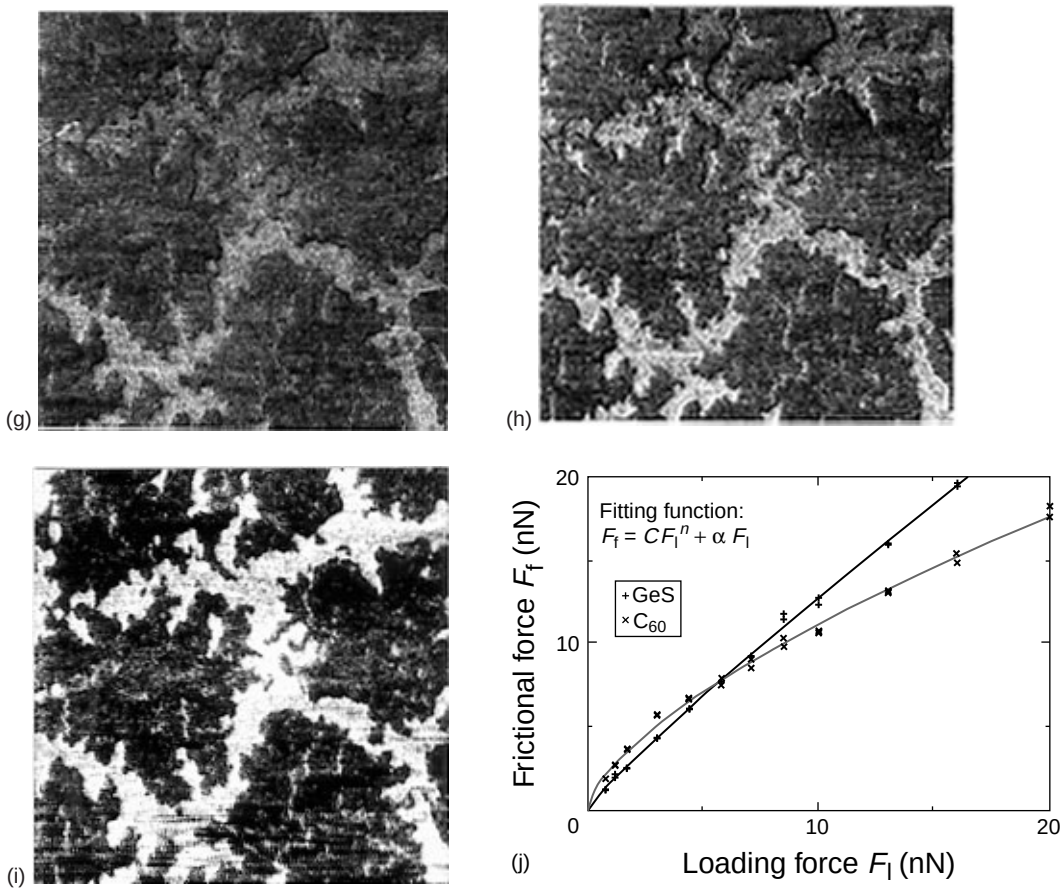

Figure 22 (*Continued*)

the local friction force spectroscopic data. The situation is analogous to STM or STS: the applied loading force as an important external parameter in LFM studies plays a similar role as the applied sample bias voltage in STM. Again, a combination of local spectroscopic measurements with spatially resolved imaging is required in order to obtain data with clear chemical contrast and clear assignment of the chemical species.

Lateral forces also play an important role for the interpretation of atomic-scale SFM images obtained in contact mode (Figure 23). Since typical loading forces exerted by the tip on the sample during measurement are of the order of 1–100 nN, up to several hundred tip atoms are in contact with the sample surface. Owing to a ''stick-slip''-type motion of the tip the surface periodicity of a crystalline sample is still resolved. However, the unit cells often show an artificial shape (Figure 23). In addition, the atomic structure within each unit cell is not resolved. On highly reactive surfaces, e.g. on semiconductor surfaces with unsaturated dangling bonds, it is particularly difficult to obtain high resolution in contact mode. In this case, dynamic force microscopy (see section 4.6) has to be applied.

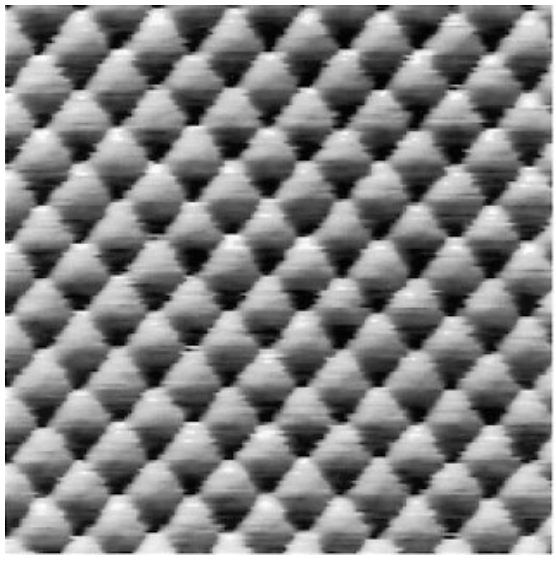

Figure 23 SFM image of a cleaved LiF(001) surface with unit-cell resolution obtained in the static contact mode. The contrast is governed by lateral forces acting on the tip in the contact region leading to a square-like appearance of the unit cell (scan size: 3 × 3 nm).

4.3 Adhesion Force Mapping

Adhesion forces can be determined by retracting the tip upon contact formation while monitoring the force necessary to break the contact. By repeating this procedure at different locations adhesion force maps can be obtained which can help to distinguish between different chemical species with different adhesion properties.[61] This method can also be applied to molecular species for which specific molecular interaction mechanisms based on the existence of geometrically complementary surfaces of recognition sites provide a means to obtain molecular fingerprints.[62–65] Examples of such SFM studies include molecular recognition between receptor and ligand, antibody and antigen and complementary strands of DNA.

4.4 Force Modulation Microscopy

By modulating the tip–sample force interaction in the contact regime, spatial variations in local surface elasticity can be probed. This method is based on modulating the sample height (z-position) by a fixed amount Δz_s. Motion of the sample causes the cantilever to deflect periodically by an amount Δz_c. For a given applied loading force, a soft surface region deforms more than a hard one. Consequently, the measured cantilever deflection Δz_c is smaller over a soft surface region. By plotting the normalized quantity $\Delta z_c/\Delta z_s$ as a function of surface location, a spatial map of surface elasticity can therefore be obtained. Applications include chemical-selective imaging of composite materials[66] and biological specimens. Alternatively, SFM can be operated in the so-called tapping mode, being sensitive also to mechanical material properties.[67]

4.5 Noncontact Force Microscopy

In the noncontact regime, the tip–sample interaction is governed by long-range VDW forces if tip and sample are neither electrostatically nor magnetically charged. The VDW forces acting between two macroscopic bodies can be calculated to a first approximation by assuming the VDW forces to be additive. For a sphere of radius R (tip) held at a distance s from a surface, the total force is given by Equation (34):

$$F_{\mathrm{VDW}}(s) = -\frac{HR}{6s^2} \tag{34}$$

where H is the Hamaker constant, which is material dependent and therefore chemical specific. Thus, by operating at a constant tip–sample distance, it is possible to distinguish between different chemical components due to the difference in the corresponding Hamaker constants.[68]

4.6 Dynamic Force Microscopy with True Atomic Resolution

In dynamic SFM with true atomic resolution, a stiff cantilever with integrated atomically sharp tip is oscillated with a fixed amplitude at its resonance frequency in ultrahigh vacuum slightly above the sample surface (i.e. the amplitude chosen is close to the equilibrium tip–sample separation). The resonance frequency w_0 is affected by the force field acting between tip and sample, which includes both short- and long-range (VDW) force contributions. Image acquisition is accomplished by profiling the sample surface with a constant frequency shift Δw. The strong distance dependence of the short-range forces allows true atomic resolution to be obtained by SFM comparable to the quality of STM data.

Figure 24 shows a dynamic SFM image of a cleaved n-InAs(110) surface with "true atomic resolution", i.e. both atomic species, In and As, are resolved within each unit cell. In contrast to contact-mode SFM, the dynamic operation mode additionally allows the observation of single point defects such as As vacancies on the InAs(110) surface. Other important applications of dynamic SFM include the investigation of noble gas crystals and biological samples.

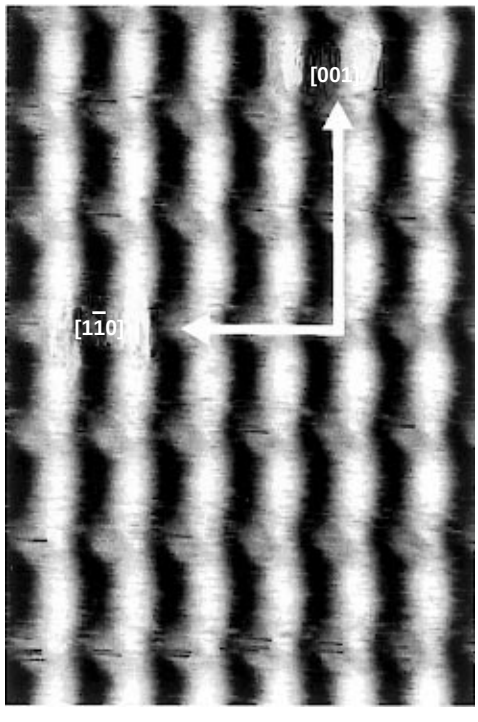

Figure 24 SFM image of a cleaved n-InAs(110) surface with "true atomic resolution" obtained in the dynamic noncontact mode at a temperature of 14 K. Within each unit cell two protrusions are observed corresponding to the As sites (bright protrusions) and the In sites (lower-lying and therefore less-bright protrusions). Scan size: 2.5 × 3.2 nm.

4.7 Magnetic Resonance Force Microscopy

Another SFM-derived technique which additionally offers great potential for three-dimensional (3D) chemical analysis is magnetic resonance force microscopy (MRFM).[69–71] In MRFM a force signal is generated by modulating the sample magnetization with standard magnetic resonance techniques. Samples of only a few nanograms generate force signals of the order of 10^{-14}–10^{-16} N, which can be detected by highly-sensitive microfabricated force sensors operated preferably under high-vacuum conditions and at low temperatures (Equation 32). Possible applications of MRFM include 3D imaging of biological specimens.

5 RELATED SCANNING PROBE TECHNIQUES

In addition to STM, SNOM and SFM, several other techniques exist which also have the potential for chemical-selective imaging.[7] For instance, the scanning thermal profiler makes use of different thermal properties of different materials whereas the scanning near-field acoustic microscope exploits their different mechanical properties. Even in solution, chemical contrast can be achieved by scanning electrochemical microscopy based on the dependence of the Faradaic currents on the particular type of material.

SCM and spectroscopy has probably become the most important technique for the nanoscale analysis of semiconductor devices. SCM is based on the measurement

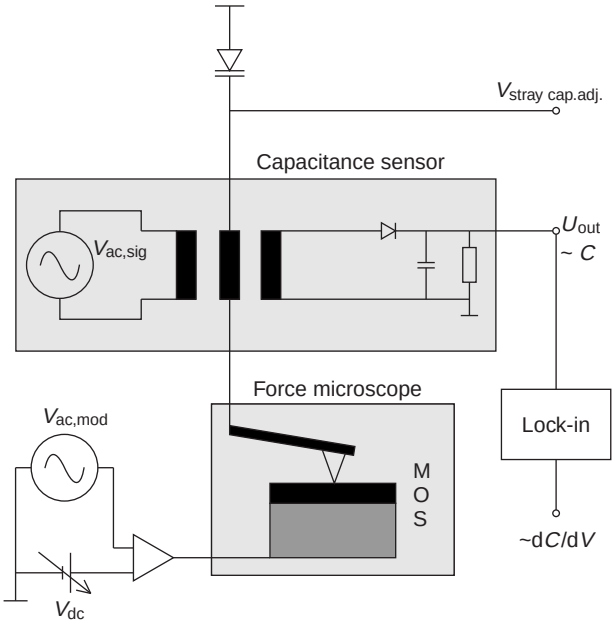

Figure 25 Schematic diagram of a scanning capacitance microscope.

of the capacitance between a doped Si tip (which might additionally be metal coated) and a semiconductor sample with a thin surface oxide layer (Figure 25). Spatial variations in the dopant concentration or type of doping cause changes in the measured capacitance. Since the electrostatic interaction involved has a long-range nature, information about subsurface properties can be obtained without the need for depth profiling.[72,73] Figures 26(a)

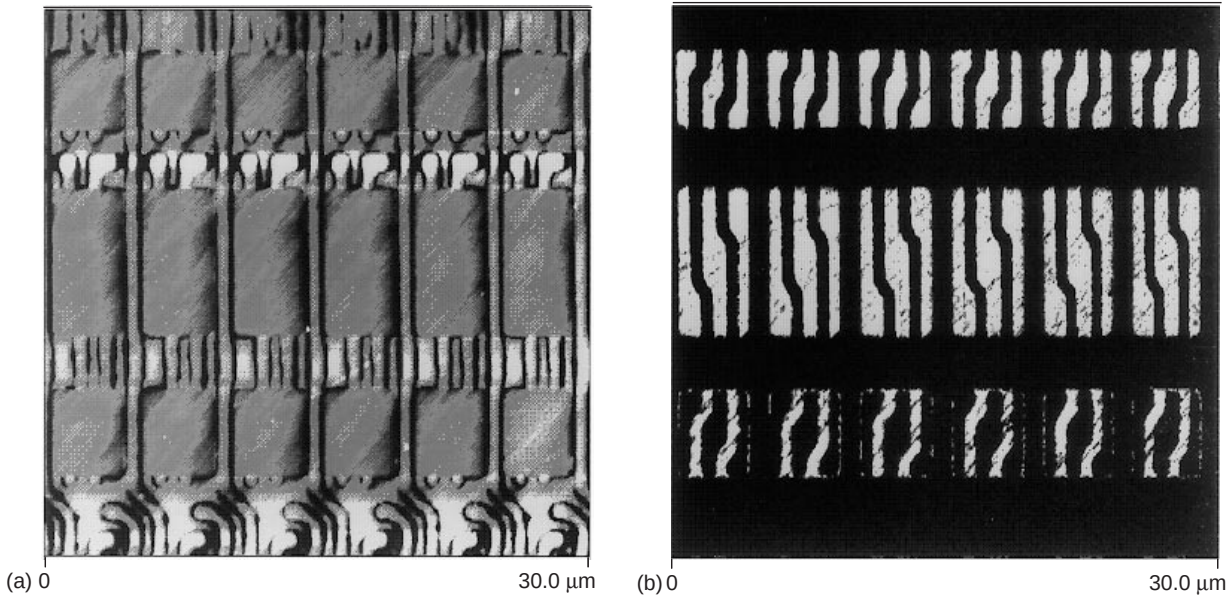

Figure 26 (a) Topographic SFM and (b) SCM images of refresh transistors of a dynamic random access memory (DRAM). The contrast in the SCM image is determined by differently doped regions (scan size: 30 × 30 µm).

and (b) show an example of a DRAM device. The SFM image reveals the surface topography whereas the simultaneously obtained SCM image provides information about the locations of gate strips of transistor structures which cannot be deduced from the SFM image.

Although all STM-related scanning probe techniques offer additional complementary information about chemical inhomogeneities, the identification of the elemental species remains an unsolved problem.

6 PERSPECTIVE AND FUTURE DEVELOPMENTS

SPM have undoubtedly contributed significantly in recent years to correlate structural with local physical and chemical properties down to the atomic scale. The lack of elemental specificity is, however, a major challenge for future developments in which SPM might be combined with established methods for chemical analysis such as mass spectrometry or core-level spectroscopic techniques.

ACKNOWLEDGMENTS

The author would like to thank his co-workers W. Allers, M. Bode, A. Born, R. Dombrowski, M. Getzlaff, M. Morgenstern, R. Pascal, A. Schwarz, U. D. Schwarz and Chr. Wittneven for their contributions and U. Fischer, R.J. Hamers, W. Ho, T. Jung, R. Möller, J. Stroscio and P. Varga for providing figures for this article. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

DRAM	Dynamic Random Access Memory
FM	Frequency Modulation
fwhm	Full Width at Half-maximum
IETS	Inelastic Electron Tunneling Spectroscopy
LDOS	Local Density of States
LFM	Lateral Force Microscopy
MRFM	Magnetic Resonance Force Microscopy
SCM	Scanning Capacitance Microscopy
SFM	Scanning Force Microscopy
SNOM	Scanning Near-field Optical Microscopy
SPM	Scanning Probe Methods
SPSTM	Spin-polarized Scanning Tunneling Microscopy
STM	Scanning Tunneling Microscopy
STS	Scanning Tunneling Spectroscopy
VDW	Van der Waals

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Atomic Force Microscopy in Analysis of Polymers

Surfaces **(Volume 10)**
Scanning Probe Microscopy, Industrial Applications of • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

REFERENCES

1. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, 'Surface Studies by Scanning Tunneling Microscopy', *Phys. Rev. Lett.*, **49**, 57–60 (1982).
2. G. Binnig, H. Rohrer, 'Scanning Tunneling Microscopy', *Helv. Phys. Acta*, **55**, 726–735 (1982).
3. G. Binnig, H. Rohrer, 'Scanning Tunneling Microscopy – from Birth to Adolescence', *Rev. Mod. Phys.*, **59**, 615–625 (1987).
4. D.W. Pohl, 'Nano-optics and Scanning Near-field Optical Microscopy', in *Scanning Tunneling Microscopy II*, 2nd edition, eds. R. Wiesendanger, H.-J. Güntherodt, Springer Series in Surface Sciences, Springer, Berlin, 233–271, 1995.
5. C. Fischer, 'Scanning Near-field Optical Microscopy', in *Scanning Probe Microscopy: Analytical Methods*, ed. R. Wiesendanger, Springer Series in Nano Science and Technology, Springer, Berlin, 161–210, 1998.
6. G. Binnig, C.F. Quate, Ch. Gerber, 'Atomic Force Microscope', *Phys. Rev. Lett.*, **56**, 930–933 (1986).
7. R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, Cambridge, 1994.
8. J. Bardeen, 'Tunneling from a Many-body Point of View', *Phys. Rev. Lett.*, **6**, 57–59 (1961).
9. J. Tersoff, D.R. Hamann, 'Theory and Application for the Scanning Tunneling Microscope', *Phys. Rev. Lett.*, **50**, 1998–2001 (1983).
10. J. Tersoff, D.R. Hamann, 'Theory of the Scanning Tunneling Microscope', *Phys. Rev. B*, **31**, 805–813 (1985).
11. S. Ohnishi, M. Tsukada, 'Effect of the Microscopic Electronic States of the Tip on the Scanning Tunneling Microscopy Images', *J. Vac. Sci. Technol. A*, **8**, 174–176 (1990).
12. C.J. Chen, 'Tunneling Matrix Elements in Three-dimensional Space: the Derivative Rule and the Sum Rule', *Phys. Rev. B*, **42**, 8841–8857 (1990).
13. C.J. Chen, 'Microscopic View of Scanning Tunneling Microscopy', *J. Vac. Sci. Technol. A*, **9**, 44–50 (1991).

14. C.J. Chen, 'Theory of Scanning Tunneling Spectroscopy', *J. Vac. Sci. Technol. A*, **6**, 319–322 (1988).
15. T. Klitsner, R.S. Becker, J.S. Vickers, 'Observation of the Effect of Tip Electronic States on Tunnel Spectra Acquired with the Scanning Tunneling Microscope', *Phys. Rev. B*, **41**, 3837–3840 (1990).
16. R.M. Feenstra, J.A. Stroscio, J. Tersoff, A.P. Fein, 'Atom-selective Imaging of the GaAs(110) Surface', *Phys. Rev. Lett.*, **58**, 1192–1195 (1987).
17. R.J. Hamers, J.E. Demuth, 'Electronic Structure of Localized Si Dangling-bond Defects by Tunneling Spectroscopy', *Phys. Rev. Lett.*, **60**, 2527–2530 (1988).
18. R.J. Hamers, 'Characterization of Localized Atomic Surface Defects by Tunneling Microscopy and Spectroscopy', *J. Vac. Sci. Technol. B*, **6**, 1462–1467 (1988).
19. M. Bode, R. Pascal, R. Wiesendanger, 'Chemical-specific Imaging of Multicomponent Metal Surfaces on the Nanometer Scale by Scanning Tunneling Spectroscopy', *Appl. Phys. A*, **62**, 571–573 (1996).
20. M. Bode, R. Pascal, M. Dreyer, R. Wiesendanger, 'Nanostructural and Local Electronic Properties of Fe/W(110) Correlated by Scanning Tunneling Spectroscopy', *Phys. Rev. B*, **54**, R8385–R8388 (1996).
21. R. Pascal, Ch. Zarnitz, M. Bode, M. Getzlaff, R. Wiesendanger, 'Surface Electronic Structure of Gd(0001) Films on W(110)', *Appl. Phys. A*, **65**, 603–606 (1997).
22. M. Getzlaff, M. Bode, S. Heinze, R. Pascal, R. Wiesendanger, 'Temperature-dependent Exchange Splitting of the Magnetic Gd(0001) Surface State', *J. Magn. Magn. Mater.*, **184**, 155–165 (1998).
23. J.A. Stroscio, D.T. Pierce, A. Davies, R.J. Celotta, 'Tunneling Spectroscopy of bcc(001) Surface States', *Phys. Rev. Lett.*, **75**, 2960–2963 (1995).
24. A. Biedermann, O. Genser, W. Hebenstreit, M. Schmid, J. Redinger, R. Podloucky, P. Varga, 'Scanning Tunneling Spectroscopy of One-dimensional Surface States on a Metal Surface', *Phys. Rev. Lett.*, **76**, 4179–4182 (1996).
25. A. Davies, J.A. Stroscio, D.T. Pierce, R.J. Celotta, 'Atomic-scale Observations of Alloying at the Cr–Fe(001) Interface', *Phys. Rev. Lett.*, **76**, 4175–4178 (1996).
26. M. Schmid, H. Stadler, P. Varga, 'Direct Observation of Surface Chemical Order by Scanning Tunneling Microscopy', *Phys. Rev. Lett.*, **70**, 1441–1444 (1993).
27. P.T. Wouda, B.E. Nieuwenhys, M. Schmid, P. Varga, 'Chemically Resolved STM on a PtRh(100) Surface', *Surf. Sci.*, **359**, 17–22 (1996).
28. D.D. Chambliss, S. Chiang, 'Surface Alloy Formation Studied by Scanning Tunneling Microscopy: Cu(100) + Au-c(2 × 2)', *Surf. Sci. Lett.*, **264**, L187–L192 (1992).
29. L. Ruan, F. Besenbacher, I. Stensgaard, E. Laegsgaard, 'Atom Resolved Discrimination of Chemically Different Elements on Metal Surfaces', *Phys. Rev. Lett.*, **70**, 4079–4082 (1993).
30. F. Besenbacher, E. Laegsgaard, L. Pleth Nielsen, L. Ruan, I. Stensgaard, 'Scanning Tunneling Microscopy Studies of Metal Surfaces: Surface Reactions, Discrimination of Chemically Different Elements, and Surface Alloying', *J. Vac. Sci. Technol. B*, **12**, 1758–1763 (1994).
31. Y.W. Mo, F.J. Himpsel, 'Spectroscopic Signature of Cu on W(110) from Scanning Tunneling Microscopy and Inverse Photoemission', *Phys. Rev. B*, **50**, 7868–7871 (1994).
32. T. Jung, Y.W. Mo, F.J. Himpsel, 'Identification of Metals in Scanning Tunneling Microscopy via Image States', *Phys. Rev. Lett.*, **74**, 1641–1644 (1995).
33. T. Jung, R. Schlittler, J.K. Gimzewski, F.J. Himpsel, 'One-dimensional Metal Structures at Decorated Steps', *Appl. Phys. A*, **61**, 461–474 (1995).
34. G. Binnig, H. Rohrer, 'Scanning Tunneling Microscopy', *Surf. Sci.*, **126**, 236–244 (1983).
35. R. Wiesendanger, M. Bode, R. Pascal, W. Allers, U.D. Schwarz, 'Issues of Atomic-resolution Structure and Chemical Analysis by Scanning Probe Microscopy and Spectroscopy', *J. Vac. Sci. Technol. A*, **14**, 1161–1167 (1996).
36. R. Pascal, Ch. Zarnitz, M. Bode, R. Wiesendanger, 'Atomic and Local Electronic Structure of Gd Thin Films Studied by STM and STS', *Phys. Rev. B*, **56**, 3636–3639 (1997).
37. R. Möller, 'Thermovoltages in Scanning Tunneling Microscopy', in *Scanning Probe Microscopy: Analytical Methods*, ed. R. Wiesendanger, Springer, Series in Nano Science and Technology, Springer, Berlin, 49–69, 1998.
38. R. Wiesendanger, 'Spin-polarized Scanning Tunneling Microscopy', in *Scanning Probe Microscopy: Analytical Methods*, ed. R. Wiesendanger, Springer Series in Nano Science and Technology, Springer, Berlin, 71–95, 1998.
39. R. Wiesendanger, I.V. Shvets, D. Bürgler, G. Tarrach, H.-J. Güntherodt, J.M.D. Coey, 'Evidence for Selective Imaging of Different Magnetic Ions on the Atomic Scale by Using a Scanning Tunnelling Microscope with a Ferromagnetic Probe Tip', *Europhys. Lett.*, **19**, 141–146 (1992).
40. R. Wiesendanger, I.V. Shvets, D. Bürgler, G. Tarrach, H.-J. Güntherodt, J.M.D. Coey, S. Gräser, 'Topographic and Magnetic-sensitive Scanning Tunneling Microscope Study of Magnetite', *Science*, **255**, 583–586 (1992).
41. R. Wiesendanger, H.-J. Güntherodt, G. Güntherodt, R.J. Gambino, R. Ruf, 'Observation of Vacuum Tunneling of Spin-polarized Electrons with the Scanning Tunneling Microscope', *Phys. Rev. Lett.*, **65**, 247–250 (1990).
42. M. Bode, M. Getzlaff, R. Wiesendanger, 'Spin-polarized Vacuum Tunneling into the Exchange-split Surface State of Gd(0001)', *Phys. Rev. Lett.*, **81**, 4256–4259 (1998).
43. T. Wolfram (ed.), *Inelastic Electron Tunneling Spectroscopy*, Springer, Berlin, 1979.
44. P.K. Hansma (ed.), *Tunneling Spectroscopy*, Plenum Press, New York, 1982.
45. D.P.E. Smith, M.D. Kirk, C.F. Quate, 'Molecular Images and Vibrational Spectroscopy of Sorbic Acid with the Scanning Tunneling Microscope', *J. Chem. Phys.*, **86**, 6034–6038 (1987).

46. B.C. Stipe, M.A. Rezaei, W. Ho, 'Single-molecule Vibrational Spectroscopy and Microscopy', *Science*, **280**, 1732–1735 (1998).
47. B.C. Stipe, M.A. Rezaei, W. Ho, 'Coupling of Vibrational Excitation to the Rotational Motion of a Single Adsorbed Molecule', *Phys. Rev. Lett.*, **81**, 1263–1266 (1998).
48. B.C. Stipe, M.A. Rezaei, W. Ho, 'Localization of Inelastic Tunneling and the Determination of Atomic-scale Structure with Chemical Specificity', *Phys. Rev. Lett.*, **82**, 1724–1727 (1999).
49. R. Berndt, R. Gaisch, J.K. Gimzewski, B. Reihl, R.R. Schlittler, W.D. Schneider, M. Tschudy, 'Photon Emission at Molecular Resolution Induced by a Scanning Tunneling Microscope', *Science*, **262**, 1425–1427 (1993).
50. R. Berndt, 'Photon Emission from the Scanning Tunneling Microscope', in *Scanning Probe Microscopy: Analytical Methods*, ed. R. Wiesendanger, Springer Series in Nano Science and Technology, Springer, Berlin, 97–134, 1998.
51. L.L. Kazmerski, 'Specific Atom Imaging, Nanoprocessing, and Electrical Nanoanalysis with Scanning Tunneling Microscopy', *J. Vac. Sci. Technol. B*, **9**, 1549–1556 (1991).
52. L.L. Kazmerski, 'Spectroscopic Scanning Tunneling Microscope Studies of Hydrogenated Amorphous Silicon Solar Cells', in *Proceedings of the 9th E.C. Photovoltaic Solar Energy Conference*, Freiburg, Germany, 25–29 September, 589–594, 1989.
53. M.B. Johnson, O. Albrektsen, R.M. Feenstra, H.W.M. Salemink, 'Direct Imaging of Dopants in GaAs with Cross-sectional Scanning Tunneling Microscopy', *Appl. Phys. Lett.*, **63**, 2923–2925 (1993).
54. Ph. Ebert, M. Heinrich, M. Simon, C. Domke, K. Urban, C.K. Shih, M.B. Webb, M.G. Lagally, 'Thermal Formation of Zn–Dopant–Vacancy Defect Complexes on InP(110) Surfaces', *Phys. Rev. B*, **53**, 4580–4590 (1996).
55. M. Morgenstern, R. Dombrowski, Chr. Wittneven, R. Wiesendanger, 'Landau Level Quantization Measured by Scanning Tunneling Spectroscopy on n-InAs(110)', *Phys. Status Solidi*, **210**, 845–851 (1998).
56. Chr. Wittneven, R. Dombrowski, M. Morgenstern, R. Wiesendanger, 'Scattering States of Ionized Dopants Probed by Low Temperature Scanning Tunneling Spectroscopy', *Phys. Rev. Lett.*, **81**, 5616–5619 (1998).
57. T. Basché, W.E. Moerner, M. Orrit, U.P. Wild (eds.), *Single-molecule Optical Detection, Imaging and Spectroscopy*, John Wiley & Sons, New York, 1997.
58. H.F. Hess, E. Betzig, T.D. Harris, L.N. Pfeiffer, K.W. West, 'Near-field Spectroscopy of the Quantum Constituents of a Luminescent System', *Science*, **264**, 1740–1745 (1994).
59. R.M. Overney, E. Meyer, J. Frommer, D. Brodbeck, R. Lüthi, L. Howald, H.-J. Güntherodt, M. Fujihira, H. Rakano, Y. Gotoh, 'Friction Measurements on Phase-separated Thin Films with a Modified Atomic Force Microscope', *Nature*, **359**, 133–135 (1992).
60. U.D. Schwarz, W. Allers, G. Gensterblum, R. Wiesendanger, 'Low-load Friction Behavior of Epitaxial C_{60} Monolayers Under Hertzian Contact', *Phys. Rev. B*, **52**, 14976–14984 (1995).
61. H.A. Mizes, K.-G. Loh, R.J.D. Miller, S.K. Ahuja, E.F. Grabowski, 'Submicron Probe of Polymer Adhesion with Atomic Force Microscopy: Dependence on Topography and Material Inhomogeneities', *Appl. Phys. Lett.*, **59**, 2901–2903 (1991).
62. K.O. van der Werf, C.A.J. Putman, B.G. de Grooth, J. Greve, 'Adhesion Force Imaging in Air and Liquid by Adhesion Mode Atomic Force Microscopy', *Appl. Phys. Lett.*, **65**, 1195–1197 (1994).
63. E.-L. Florin, V.T. Moy, H.E. Gaub, 'Adhesion Forces Between Individual Ligand–Receptor Pairs', *Science*, **264**, 415–417 (1994).
64. C.D. Frisbie, L.F. Rozsnayai, A. Noy, M.S. Wrighton, C.M. Lieber, 'Functional Group Imaging by Chemical Force Microscopy', *Science*, **265**, 2071–2074 (1994).
65. G.U. Lee, L.A. Chrisey, R.J. Colton, 'Direct Measurement of the Forces Between Complementary Strands of DNA', *Science*, **266**, 771–773 (1994).
66. P. Maivald, H.J. Butt, S.A.C. Gould, C.B. Prater, B. Drake, J.A. Gurley, V.B. Elings, P.K. Hansma, 'Using Force Modulation to Image Surface Elasticities with the Atomic Force Microscope', *Nanotechnology*, **2**, 103–106 (1991).
67. R.G. Winkler, J.P. Spatz, S. Sheiko, M. Möller, P. Reineker, O. Marti, 'Imaging Material Properties by Resonant Tapping-force Microscopy: a Model Investigation', *Phys. Rev. B*, **54**, 8908–8912 (1996).
68. I.Y. Sokolov, 'On the Recovery of the Spectroscopic Image in Atomic Force Microscopy', *J. Vac. Sci. Technol. A*, **14**, 2901–2904 (1996).
69. J.A. Sidles, J.L. Garbini, K.J. Bruland, D. Rugar, O. Züger, S. Hoen, C.S. Yannoni, 'Magnetic Resonance Force Microscopy', *Rev. Mod. Phys.*, **67**, 249–265 (1995).
70. D. Rugar, C.S. Yannoni, J.A. Sidles, 'Mechanical Detection of Magnetic Resonance', *Nature*, **360**, 563–566 (1992).
71. D. Rugar, O. Züger, S. Hoen, C.S. Yannoni, H.-M. Vieth, R.D. Kendrick, 'Force Detection of Nuclear Magnetic Resonance', *Science*, **264**, 1560–1563 (1994).
72. A. Born, R. Wiesendanger, 'Scanning Capacitance Microscope as a Tool for the Characterization of Integrated Circuits', *Appl. Phys. A*, **66**, S421–S426 (1998).
73. A. Born, A. Olbrich, M. Maywald, R. Wiesendanger, 'Analysis of Electrical Breakdown Failures by Means of SFM-based Methods', *Appl. Phys. A*, **66**, S1063–S1065 (1998).

Scanning Electron Microscopy in Analysis of Surfaces

Michael K. Lamvik
MCNC, Research Triangle Park, USA

1 Introduction **2**
1.1 Overview of Scanning Electron Microscopy 2
1.2 Capabilities 3
1.3 Limitations 3
2 History **3**
3 Sample Preparation **4**
3.1 Cleaning and Coating 4
3.2 Embedding and Sectioning 4
3.3 Lapping and Polishing 4
3.4 Freezing and Cleaving 4
3.5 Modification of Ambient Conditions 4
4 Microscopy of Surfaces **4**
4.1 Resolution Limits with Tungsten Sources 4
4.2 Resolution Limits with Hexaboride Sources 5
4.3 Resolution Limits with Field Emission Sources 5
4.4 Resolution Limits with Traditional Scanning Electron Microscope Objective Lenses 6
4.5 Resolution Limits with Immersion and Semi-immersion Objective Lenses 6
4.6 Contrast and Resolution Limits with Secondary Electrons 6
4.7 Contrast and Resolution Limits with Backscattered Electrons 7
5 Applications to Minerals **7**
6 Applications to Metals **8**
7 Applications to Fibers **9**
8 Applications to Polymers **9**
9 Applications in Biology **10**
10 Instrument Performance **10**
10.1 Calibration 10
10.2 Resolution Tests 11
11 Comparison with Other Types of Microscopy **11**
11.1 Optical and Confocal Microscopes 11
11.2 Transmission Electron Microscope 11
11.3 Atomic Force and Scanning Tunneling Microscopy 11
Acknowledgments **12**
Abbreviations and Acronyms **12**
Disclaimer **12**
Related Articles **12**
References **12**

Microscopy is the art and science of visually observing small objects. Electrons are electrically charged subatomic particles typically responsible for the flow of electrical current. Scanning is a method of observation where the focus of attention is moved in a chosen pattern and desired information is gained by assembling observations in sequence. In a scanning electron microscope (SEM), a beam of electrons is scanned in a regular pattern over a small object, and the image of the object appears on a screen by displaying the image points in synchrony with the movements of the electron beam on the object.

A familiar example of image scanning is a television receiver, where a picture is created on a screen by rapidly scanning an electron beam across the phosphor face of the television tube. The beam striking the phosphor produces light, which is seen as a picture. The signal that modulates the electron beam in a television set is created in a television camera, which is scanned in the same way that the television screen is scanned. In a SEM, the signals come from special detectors in the microscope, not from an optical camera. Otherwise, the image display in a SEM is very similar to television. Although it is typical for SEM images to be scanned at a slower rate than are television signals, some SEMs display images at the same speed as television.

The most common signal for detection in a SEM is the secondary electron signal. Secondary electrons are low-speed electrons (having energies less than about 50 eV) that result from the interaction of the primary (i.e. beam) electrons with the sample. The secondary electrons are detected, usually by a scintillator/photomultiplier detector, and the resulting signal modulates the display screen.

Secondary electrons are popular because they provide a kind of shading that gives a three-dimensional appearance to an image, making it easy to visualize the structural relationships of parts of the sample (Figure 1).

Other signals provide information of a more analytical nature. The signal produced by backscattered electrons is directly (although not exactly proportionally) related to the atomic number of the material being struck by the primary beam. Diffracted electrons provide information about crystal spacings and orientations in the sample.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

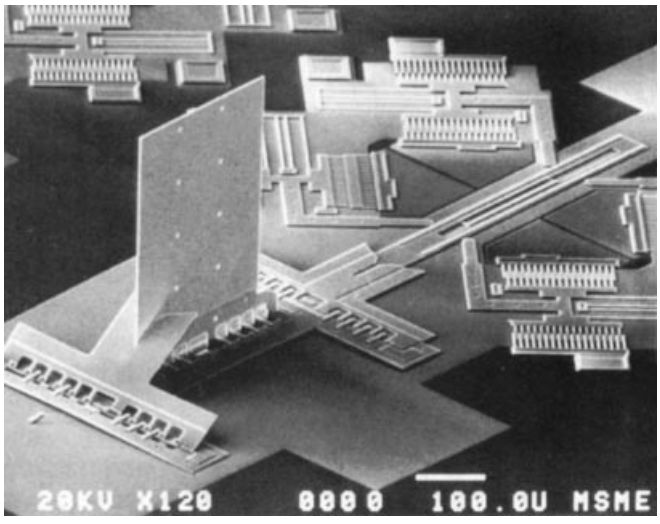

Figure 1 This photograph illustrates the essential features of scanning electron microscopy: large depth of field, high resolution, and an easily interpreted shading that gives the impression of a diffuse light source in the distance. The high resolution and large depth of field depend on the small wavelength of electrons and the small aperture angles that are used to define the electron beam. The appearance of diffuse illumination does not depend on the illuminating beam at all, but rather on the way secondary electrons are detected. (Reproduced by permission from A. Friedberger, R.S. Muller, 'Improved Surface-micromachined Hinges for Fold-out Structures', *J. Microelectromech. Systems*, **7**(3) 315–319, copyright © 1998 IEEE.)

Induced-current signals are particularly interesting for studying the response of active electronic components.

The resolution of a SEM image depends on the quality of the optical system that forms the electron beam, including the nature of the electron source. Most SEMs can typically resolve objects as small as 10 nm. The resolution of specialized SEMs is better than 1 nm. In contrast, the resolution of light microscopes is limited by the wavelength of light to about 500 nm.

The magnification of a SEM is not directly related to the optical system, very unlike a light microscope. The magnifying power of a SEM is related to the ratio between the size of the scan on the object compared to the size of the scan on the image display. In this way a SEM functions like a mechanical pantograph that was used, typically in older days, for copying drawings. If the arms of the pantograph were of equal length, the drawing would be copied the same size. If the arms were adjusted to different lengths, the drawing could be magnified or reduced.

The primary advantage of a SEM is that it provides familiar-looking images of surfaces with higher resolution and greater depth-of-field than an optical microscope. Although the images are typically familiar looking, the contrast mechanism in a SEM is not the same as the contrast provided by light, and an observer expecting always to interpret a SEM image the same way as a light image could easily be fooled.

As the electron beam consists of electrically charged particles, the sample for SEM observation must either be inherently electrically conducting or must be made electrically conducting to yield good images. Otherwise, electrical charge builds up on the sample, the yield of secondary electrons is changed, and the scanning beam can become erratic. Low-voltage or low-vacuum SEMs can be used to reduce the charging problem.

1 INTRODUCTION

1.1 Overview of Scanning Electron Microscopy

A SEM is a versatile and efficient instrument for obtaining magnified images of small objects with a resolution better than that available with a light microscope.[1] SEMs are available with a wide range of costs and capabilities.

Scanning microscopes generally have an electron source, which creates a beam of electrons, an optical system, which focuses the electrons, a scanning system, which directs the beam over the sample, one or more detector systems, which create signals dependent on interactions of the beam electrons with the sample, and display systems, which allow the microscope operator to interpret visually the information produced by the microscope.

Light microscopes reached a fundamental limit to their ability to see small objects once the resolution of their optical systems approached the size of the wavelength of light. No matter how good the microscope lenses were, light itself could not reproduce an image of something smaller than its own wavelength. (Near-field scanning optical microscopes use light photons to produce resolution better than the light wavelength, but here the resolution is created by the scanning mode rather than by the light optics. This is a good analogy with the SEM, where magnification is provided by the scanning mode, not by the optical system.)

Among alternatives to light, electrons have a very small wavelength, much smaller than that of light, and they are relatively easy to produce and to focus. X-rays also have a small wavelength, but they are very difficult to focus. Electrons have an electric charge, so they can be focused and deflected with electric and magnetic fields. Protons and ions are charged, and therefore can be focused, but they are typically more difficult to produce than electrons. Electrons can be ejected from a solid by heating, such as commonly occurs with a tungsten filament or lanthanum hexaboride crystal, or they can be drawn out of a solid by field emission, a quantum-mechanical tunneling phenomenon. Field emission electron sources

are typically the smallest and the brightest, and they usually offer the highest resolution. Tungsten filaments produce the highest total electron current, although they allow a rather limited image resolution. Lanthanum hexaboride sources offer brightness and resolution that are less than field emission sources but more than tungsten filaments.

Once the electron beam is produced, it is generally focused by a series of cylindrically symmetric magnetic lenses to form a small spot on the sample. Because the image is created from the interaction of the beam with the sample, the highest image resolution comes from the smallest interaction spot. The objective lens is crucial in forming a small spot. Lens aberrations are directly related to the focal length of the lens, so the smallest focal length lens generally has the highest resolution. This fact typically leads to a trade-off, because a large working distance is very useful in providing flexibility in observing a variety of large samples, but a large working distance usually means a long focal length and hence a reduced resolution.

After the electron beam hits the sample, a large range of interactions can occur. Some electrons interact with atomic nuclei and are deflected back at large angles (called backscattered electrons). Some primary (beam) electrons create lower-energy secondary electrons that escape from the sample. Atomic and molecular electrons in the sample can de-excite, with the release of X-rays, light, and Auger electrons. Because the number of electrons leaving the sample is generally not equal to the number of electrons incident on the sample, there is a net electric current through the sample induced by the beam, and this current can be used as a signal to form an image.

In a scanning microscope, the beam typically generates a signal at one point on the sample and then moves to an adjacent point to obtain another reading of the same signal. Then the beam continues to move by small increments and continues to record the signal. A spot on a display screen is moved in synchrony with the spot on the sample. If the signal is low at one spot on the sample and high at another spot, typically the display screen will be dim at the first spot and bright at the second. If the spot on the display screen moves 1 mm every time the spot on the sample moves 1 μm, then the screen will show a magnified image of the sample (with a magnification of 1000× in this case). Unlike a light microscope, the lenses of the scanning microscope do not need to be changed to allow the magnification to be changed. It is only necessary to change the ratio between the distance moved on the sample and the distance moved on the screen. Because of the simplicity of the scanning mechanism, devices for driving the scans and recording images can be constructed very easily and inexpensively.[2]

1.2 Capabilities

A SEM will typically produce an image of higher resolution and depth-of-field than a light microscope. Because of the higher resolution, a SEM can produce an image with a higher useful magnification than an optical microscope. A SEM generally produces a surface image, because of the nature of the detectors that are typically used. A SEM generally can handle a larger sample than a transmission electron microscope (TEM), because a TEM typically has a smaller space in the objective lens available to hold the sample (the smaller space correlates with a small focal length which implies higher resolution). A SEM can make use of a wide range of electron-induced signals from the sample to offer a variety of different kinds of information about the sample.

1.3 Limitations

The sample in a SEM must be observed in vacuum or, for specialized instruments, in a significantly reduced-pressure chamber. The sample must be conductive or must be made conductive for good images (except in a low-voltage SEM, where there can be a closer balance between the numbers of incident and emitted electrons, or an environmental SEM where charge neutralization can occur). Although SEM images often look realistic, the imaging modes are not the same as in a light microscope and misinterpretations can occur when too strong an analogy is assumed between imaging with light and electrons. Fundamental resolution limits are caused by properties of the electron source and electron optics in the microscope. Except for very specialized instruments, the resolution of a TEM will be better than the resolution of a SEM, primarily because of the optical properties of the respective objective lenses.

2 HISTORY

Electron microscopes, including SEMs, were demonstrated in Germany in the early part of the 1930s. According to Wells,[3] the idea of a device using a scanning beam was patented in Germany as early as 1927 by Stintzing. A SEM of a kind that would be familiar now was developed from cathode ray tubes by Knoll in 1935.[4] A similar kind of mechanism for making enlarged images was demonstrated in the USA by Burnett in 1938 at the RCA labs using television tubes.[3] Goldstein et al.[1] credit von Ardenne[5] with the earliest description of the construction of a SEM. By 1942, Zworykin et al. at RCA had developed a SEM with many of the features of current designs,[6] where electron optical elements focused a beam on the sample

and a secondary electron signal was picked up with a photomultiplier detector. This group reported other far-sighted developments, including work with field emission sources. But there was always difficulty in obtaining an adequate signal-to-noise ratio in the images. By 1953, McMullan in England had made significant progress in improving image quality by using an improved electron detector.[7] Improvements continued to be made,[8] and a commercial SEM was announced in 1965 by Cambridge Instruments.[1] Numerous other commercial designs followed.

3 SAMPLE PREPARATION

3.1 Cleaning and Coating

A SEM is usually used to provide a realistic-looking image of the surface of a solid, so in many cases it is not desirable to modify the sample. Many vacuum-compatible materials can simply be placed in the SEM for observation without further preparation. In an environmental scanning electron microscope (ESEM), samples can continue to include some volatile components and cleaning is not needed. Some samples need to be cut to fit in the SEM chamber. Samples that are contaminated with dirt or oil are usually cleaned prior to observation. Depending on the solubility properties of the sample, it is commonly possible to wash the sample with detergent and water or with organic solvents.

Nonconducting samples typically need to be made conductive to produce good images in a high-vacuum SEM. Samples are usually sputter coated with a heavy metal, such as gold or platinum, to make them conducting. The sputter-coating method usually provides a more uniform coating over irregular surfaces than does vacuum evaporation. When it is desirable not to coat the sample with a metal, it is possible to coat the sample with carbon by vacuum evaporation. It would not be good to use a heavy metal coating, for example, if the goal is to use the backscattered electron signal to detect metal clusters in a polymer matrix.

3.2 Embedding and Sectioning

Sometimes a SEM image of the interior of a solid sample is required. In this case, the sample can be cut to reveal the area required. Small samples, or soft samples, may need to be embedded in a hard epoxy resin before cutting in order to provide enough support to hold the sample. Some epoxy resins have conductive fillers that improve the conductivity of the sample, and these can be helpful for samples that will be examined by SEM. However, it can be helpful to use a transparent (nonconductive) epoxy resin when the location of the cut has to be determined visually. Other mounting materials are also available. Pressed mounts are used for samples that can tolerate higher pressure. After embedding, hard samples will usually be cut and polished to yield a cross-section. Softer materials can be cut with a microtome.

3.3 Lapping and Polishing

The microstructure of a sample can easily be obscured by a random surface texture, such as at a cut made by a saw, so it can be helpful to make a smooth surface. In these cases, it is useful to lap and polish the sample surface that is intended for observation. Plastic lapping films containing diamond abrasives are very good for high-quality lapping. Sample surfaces can be lapped with films down to 0.5-μm abrasive particle size, and further polished with materials such as colloidal silica for an extremely fine finish. Methods such as this are particularly useful for visualizing the interior of semiconductor integrated circuits.

3.4 Freezing and Cleaving

Some soft materials are not easily cut, or they suffer plastic deformation that yields a cut surface that is not useful for observation. In some cases, a more useful surface is obtained if the sample is frozen (for example, in liquid nitrogen), and is cleaved by a blow from a hard knife or by shear from clamps on opposite sides of the sample.[9,10] Obviously care must be exercised to ensure safety. Shields and thermal insulators need to be used in observing and handling cold materials. It is important to avoid splattering cryogenic liquids.

3.5 Modification of Ambient Conditions

ESEMs have been developed that require less stringent vacuum requirements in the sample region than typical SEMs. By controlling the atmosphere surrounding the sample in these instruments and the temperature of the sample stage, it is often possible to allow volatile components to remain in the sample and even to view hydrated samples directly. By controlling the environment around the sample, it is possible to reduce the necessity for coating samples with metal.

4 MICROSCOPY OF SURFACES

4.1 Resolution Limits with Tungsten Sources

The whole purpose of microscopes is to make a magnified image, so it might seem strange that the optical system of a SEM has nothing to do with magnification. As noted

above, the scanning system of a SEM is what creates the magnified image. The optical system of a SEM is actually a demagnifying system. The lenses of a SEM are used to focus the electron beam to a small point on the sample and, the smaller the point, the better the resolution. As the size of a tungsten filament is a substantial fraction of a millimeter, and the desired resolution of the microscope is a small fraction of a micrometer, the demagnification of a SEM optical system has to be at least 1000×. For 10 nm resolution with a tungsten source, the demagnification is typically around 1/30 000.

The ultimate limitation on the resolution of a SEM is the electron source, because no amount of demagnification of the source can make the image point brighter than the source. (In classical mechanics that fact would be expressed as a consequence of Liouville's theorem on the conservation of phase space, but in electron optics reference is made to Langmuir's equation.[3]) The brightness of the electron source[11] is the current density per solid angle, which can be given by Equation (1)

$$\beta = \frac{i}{s\Omega} \tag{1}$$

where β is the brightness, i is the beam current, $s = \pi r^2$ is the area of the beam, r is the radius of the beam, and $\Omega = \pi\alpha^2$ is the solid angle subtended by the beam, where α is the divergence angle (semi-apex angle) of the beam (measured, for example, at the gun cross-over, beyond which the electron energy is constant). Demagnification cannot increase the brightness. Also, because the aberrations of electron lenses require the acceptance angle to be limited in order to maintain resolution (see below), beam intensity is actually lost on the aperture stops used to limit the lens acceptance angles. With additional demagnification the beam becomes dimmer and dimmer until the signal is inadequate to make a picture in a reasonable amount of time. The statistical standards for deciding what are adequate signal levels were worked out first for television pictures,[12] but the same statistics apply as well to scanning microscopes.

A tungsten filament is the largest of the electron sources typically used in electron microscopy. Although such a filament can produce large amounts of current, the current is spread over a large source size, so the maximum brightness is limited. A tungsten hot filament may have a brightness β of $10^4\,\mathrm{A\,cm^{-2}\,sr^{-1}}$ at 25 kV.[3] A certain amount of brightness is needed to provide an image that is good enough to allow the microscope to be focused and an image to be recorded. In practical terms, resolutions better than 10 nm are difficult to achieve with a tungsten hot filament without facing serious intensity limits.

4.2 Resolution Limits with Hexaboride Sources

Lanthanum hexaboride filaments are an improvement over tungsten, first because of higher brightness. The source size is smaller and the resulting electron beam is brighter. That produces better image resolution.

Another advantage of a hexaboride source is that it operates at a lower temperature, causing less evaporation of the filament and hence increasing lifetime. For instruments that are heavily used, this is an operational advantage. This advantage was noted by Broers, who was primarily responsible for the development of practical lanthanum hexaboride guns.[13] A disadvantage is that the hexaboride crystal can be easily poisoned by contaminants, requiring a better vacuum system than is typical with tungsten filaments.

The brightness β of a lanthanum hexaboride gun can be $10^6\,\mathrm{A\,cm^{-2}\,sr^{-1}}$ at 25 kV.[3] Instrumental resolution of 1 nm is possible with lanthanum hexaboride sources.

4.3 Resolution Limits with Field Emission Sources

Field emission tips are the smallest of the electron sources typically used in electron microscopy, and the brightest. The first practical SEM with a field emission source was developed by Crewe et al.[14] Because of the curvature of the electric field lines around the tip, the effective source size is a few tens of nanometers, so a demagnification of 1000× can produce a resolution approaching 0.1 nm.[15] With such a small demagnification, there is adequate brightness for recording high-resolution pictures.[16] The brightness β of a field emission gun can be $10^8\,\mathrm{A\,cm^{-2}\,sr^{-1}}$ at 25 kV.[3]

Another advantage of field-emission tips is that they can be operated with little heating, or even at room temperature. Field emission is a quantum-mechanical effect. The electrical potential at the surface of a metal usually keeps electrons inside, and energy usually has be added as heat to allow the electrons to escape (which happens with tungsten filaments, for example). With field emission, an extremely high electric field is applied to the tip to make the electrostatic barrier very thin. Although the barrier is still there, the quantum-mechanical uncertainty principle makes it possible for electrons to appear on the other side of the barrier, where they can form the electron beam that is used in the microscope.

A disadvantage of the field emission source is that vacuum conditions must be excellent. An ordinary vacuum would allow the formation of heavy ions that could bombard the tip and destroy it. Adsorbed atoms can adversely effect the local work function of the metal, leading to increased instability of the beam.

4.4 Resolution Limits with Traditional Scanning Electron Microscope Objective Lenses

Although the ultimate resolution of a SEM is limited by the electron source, the useful resolution of a SEM over practical magnification ranges is set by the probe-forming lens (sometimes called the objective lens, by analogy with other kinds of microscopes). Lens aberrations and diffraction are lens defects that limit resolution. Electron-optical lenses used for focusing electrons are much worse than glass lenses for light. Whereas chromatic and spherical aberration can be corrected in the construction of glass lenses, it has been proved that spherical aberration can never be removed from the cylindrically symmetric magnetic fields typically used as electron lenses (although it is possible to build noncylindrical correctors – see below). The spherical aberration coefficient C_s, which indicates the unsharpness of a focal point, is related to the focal length of an electron lens. The longer the focal length, the larger the aberration. Spherical aberration is also related to the aperture size of the lens. The minimum focal spot size d is given by Equation (2)

$$d = 0.5\, C_s\, \alpha^3 \tag{2}$$

where α is the beam angle defined by the aperture. The wider the aperture, the larger the unsharpness. To limit the aberration, the aperture must be kept small (thus limiting the beam intensity, which is why it is not practical to use extreme demagnification to make a high-resolution microscope).

The focal length of the lens has to be adjusted to accommodate the working distance between the lens pole piece and the sample. To look at large irregular samples, it is often necessary to keep the sample away from the lens, so the beam has to focus far from the pole piece. This long distance creates larger aberrations.

To allow the most flexibility in dealing with large and irregular samples, the most common lens for SEM instruments was designed to take an electron beam coming from one side of the lens and focus it to a point on the other side of the lens, outside of the major part of the magnetic field. Keeping the sample outside of the magnetic field also makes possible the most efficient collection of secondary electrons, which are commonly used for imaging samples in a SEM. Such a design has the longest focal length, and the lowest resolution.

4.5 Resolution Limits with Immersion and Semi-immersion Objective Lenses

For high resolution, the best place for the sample is directly inside the lens, where the magnetic field of the lens is highest. Here the curvature of the electron beam is high, and the focal length is very short. This produces the smallest aberration coefficients, but only small samples can fit inside such lenses.

An interesting disadvantage of this high-resolution immersion design in a SEM is that secondary electrons, which are typically detected in a SEM, cannot escape from the magnetic field of the lens. In such a microscope, a secondary electron detector at the side of the chamber, which is the usual location, would detect few electrons. This was solved by placing the secondary detector above the lens, where the electrons could spiral up the lens field to reach it. There are also other detector designs to solve this problem.

A more recent semi-immersion lens design, the snorkel lens, is a compromise between the fully exterior probe and the immersion lens. The exit pole piece of the lens is constructed to squeeze the magnetic field out toward the sample, so the magnetic field will be near the sample even though the sample is outside of the lens. This allows a short focal length with a sample that is exterior to the lens, allowing larger samples than the full-immersion lens and much higher resolution than the probe lens.[17] Each of these two modes (immersion and probe) have their advantages, so one manufacturer has introduced an electron optical column with extra lenses that allows switching between semi-immersion mode and external-probe mode.[18]

To go beyond the fundamental resolution of the cylindrically symmetric electron objective lens, it is necessary to correct the spherical aberration inherent in such lenses. As noted above, for a SEM the optics needed to focus on the sample are totally independent of the mechanism of producing magnification. The improvement of the sharpness of the lens can be done before the beam is scanned to produce magnification. For that reason, an aberration corrector for a SEM only needs to correct the on-axis aberrations, whereas for a TEM it is necessary to correct off-axis aberrations. For a SEM, a corrector has been constructed that has improved the resolution of an instrument operating at 1 kV from 5 nm to below 2.5 nm.[19]

Combining a field-emission source with a full-immersion lens can produce impressive resolution at higher voltages. One commercial producer offers a field emission scanning electron microscope (FESEM) with 0.24 nm resolution at 200 kV,[20] designed primarily to detect electrons passing through the samples and intended for the semiconductor industry.[21]

4.6 Contrast and Resolution Limits with Secondary Electrons

Because the scanning beam only strikes one point of the sample at a time, it is not necessary to focus the electrons as they exit the sample to make an image on a screen. Any

group of exiting electrons can be detected to produce an image. The scanning mode produces the image, point-by-point. The detector only needs to tell the microscope how bright each point should be. Not only does this give great freedom in selecting what signal is used to produce the image, it allows the possibility of detecting many signals simultaneously. Secondary electrons and backscattered electrons are examples of such groups of electrons. Electron-induced signals of other types can also be detected, such as cathodoluminescence photons or elementally specific X-rays (these signals are discussed in section 4.6 and **Electron Microscopy and Scanning Microanalysis**).

Secondary electrons are electrons of less than 50 eV energy that are induced by the primary electrons (beam electrons). Because of their low energy, secondary electrons have a small escape depth (of around 3 nm[22] in carbon), so they are capable of yielding images with resolution of this same scale. (However, secondary electrons created at the exit surface of backscattered electrons can confuse this resolution, as distinguished from the secondary electrons resulting from the incident primary beam.)

The low energy of the secondary electrons also makes it easy for the electrical potential on the secondary electron detector to pull the secondary electrons around small obstructions to reach the detector. This effect is one of the main reasons why the secondary electron signal produces such a pleasing three-dimensional appearance in the micrographs. It is very easy to look at a SEM secondary electron image and imagine viewing a landscape from a high vantage point.

The limited escape depth of the secondary electrons is also involved in another contrast effect that enhances the appearance of edges. When a primary electron is incident perpendicular to a surface, it generates secondary electrons from only a short distance inside the solid. When a primary electron enters a surface at a glancing angle, it can travel inside the solid at a position still near the surface, so secondary electrons can escape for a considerable length of the primary electron's path. As the glancing electron develops more secondary electrons than a perpendicular primary electron, the signal corresponding to the glancing electron is higher, and the image point is brighter. This kind of situation happens particularly at edges, where primary electrons can generate many more secondary electrons, so edges are frequently bright in secondary electron images.

In addition to geometrical effects, contrast in secondary electron images depends on the properties of the materials being struck by electrons. The secondary electron yield is different for different materials, causing significant contrast between substances. One of the reasons that gold is often preferred as a coating material for SEM samples, rather than carbon for example, is that gold has a particularly high secondary electron yield.

4.7 Contrast and Resolution Limits with Backscattered Electrons

Backscattered electrons have a relatively high energy, so they can scatter from the atoms of a solid and travel a considerable distance in the solid. For that reason, the resolution of images from backscattered electrons can be limited. This degree of scattering is dependent on the accelerating voltage of the primary beam, so resolution can be improved at lower voltages. If the backscattered electrons are energy filtered, the resolution can be improved by limiting the collection of electrons to those that have scattered only once or a few times (by elastic scattering).[23] As backscattering is heavily dependent on elastic electron scattering from the electrostatic field of the nucleus of atoms, the backscattered signal is predominantly atomic-number dependent. The backscattered signal can be used to make an image show contrast that is specific to the materials being illuminated by the electron beam.

As the backscattered electrons are of higher energy, they mostly travel in straight lines and they do not yield softly lighted images, such as those generated by secondary electrons. Typically a backscattered-electron image is more dependent on material contrast than on geometrical contrast. Some manufacturers provide a segmented backscattered electron detector that allows backscattered electrons from one direction to be contrasted against backscattered electrons traveling in another direction. This system introduces some geometrical contrast into an image produced by backscattered electrons.

Electrons can also be reflected back to a detector by diffraction from a crystal lattice (backscattered electron diffraction), or the escape depth can be modified by the direction of travel through a crystal lattice (electron channeling). These signals can be useful for studying crystalline materials or localizing grain boundaries in metals.

5 APPLICATIONS TO MINERALS

Being quick and efficient instruments with which to produce images with better resolution than light will allow, SEMs have been in routine use for decades. There have been so many applications of SEMs that it would be impossible to provide an adequate review in a limited space. In this and following sections, a variety of

recent applications are described that may provide some interesting and useful examples of possible uses of SEMs.

Minerals have a variety of compositions and a variety of crystal structures. One striking advantage of SEMs is that the beam can provide an image of sample morphology at the same time as providing a measure of elemental composition through energy-dispersive X-ray spectrometry. This important capability of elemental analysis is described in section 5.

A controversy over the age of a wall painting in the American Southwest provided a good example of the value of SEM in analyzing the microstructure of mineral products. The pictograph, called the All American Man because of its red, white and blue decoration, may have been painted on the wall of a cave at a time before the USA was created and before the well-known red, white and blue flag was adopted by the USA. However, it was also possible that the painting was made, or was redecorated, after the founding of the USA, because the shield motif of the painting was so reminiscent of the stars and stripes common in the USA, and the colors were so similar to the red, white and blue of the flag. To determine if the design predated the USA, it was necessary to identify and date the pigmented layers of the pictograph.

Analyzing flakes that fell off the wall (so the wall painting itself would not be further damaged) a SEM was used to characterize the layers that made up the paint decoration. Once the layers were confirmed morphologically, carbon dating techniques were used to demonstrate that the pigments of the pictograph dated from the 1400s, prior to the founding of the USA.[24]

In addition to observing natural crystallite morphology, the fine structure of mechanically produced products has also been a common application of SEMs. Some products, such as decorative beads, have been made for centuries. Such beads have been important consumer products and trade goods in ancient civilizations as well as in more recent times. Artisans have a strong motivation to develop distinctive products for wealthy customers and patrons. In order to make necklaces and other jewelry from beads consisting of minerals and semiprecious stones, it is generally necessary to make a hole in the bead for stringing the necklace. It was not an easy task to produce such holes in stones in ancient times, and different cultures developed different methods. Diamond-tipped drills were used to make the holes in such beads, even in Roman times, long before metal drills were available. Tool marks in the walls of bead holes give clues to the type of drills used to make the holes, whether diamond drills or drills based on other abrasives. Often it is not desirable to break a valuable sample, such as ancient jewelry, to observe the interior of a structural feature. In such a case, it is common to inject material into the feature to make a solid cast. The cast material can then be removed from the feature and be observed by SEM. Such casts have been used in a study of ancient drills, where a SEM was used to produce information about the drill types used in the Indus Valley in ancient times.[25]

The surface structure of minerals may yield clues to reactivity, and the enhanced resolution of a FESEM is useful in detailing the surface structure. A recent study considered the use of rice-husk ash as a silica source for the formation of zeolites.[26] X-ray diffraction was used to identify amorphous silica, but not all of the silica identified this way proved to be equally reactive. Nuclear magnetic resonance was used to investigate the local structure of the silica, finding the atomic bonding arrangement that was associated with the most reactive samples. The surface structure of the most reactive silica was examined by a FESEM, and these samples showed clearly defined parallel layers of sharp flakes interleaved with smoother surface features. The smoother features were also present in the less reactive samples but the sharp flakes were not prominent, suggesting that the sharp flakes were indicative of the structure of the more reactive silica.

Another specialized SEM, the ESEM, allows the sample to be held in a special environment of temperature and pressure instead of in the high vacuum that is typical for SEMs. The structural changes involved in the curing of cement can only be observed in real time when hydration is allowed, which is possible with an ESEM. By using an ESEM, the evolution of microstructure was observed in high-strength high-alumina cement.[27] In the presence of phosphate, small irregular shapes, such as polygons, plates and fibers, were observed to form between the crystallites of calcium aluminate hydrate, and appeared to be involved in enhancing the strength of the cement.

6 APPLICATIONS TO METALS

As the backscattered electron signal is related to the atomic number of the material illuminated by the electron beam, it is a useful signal for distinguishing areas of metals in other matrices, or in discriminating different metals. In a study of cosmetics from ancient Palestine, a protrusion was observed from a container that was being studied. A SEM was used to study cross-sections of the container, identifying the protrusion as a metal rod and discovering a second corroded rod nearby.[28] Again, as in other studies, energy-dispersive X-ray spectrometry was used to identify the metals. Other techniques, such as Fourier transform infrared (FTIR) and X-ray diffraction, figured prominently in the study. However, in contrast to many other techniques, the backscattered electron

image provides an exceptionally rapid method for finding small areas of metal in lighter matrices. Such areas might not even be suspected without the initial screening by a SEM with the backscattered electron detector. After an anomalous area is discovered by a SEM, the composition of the area can be further studied by other more detailed, but slower, techniques.

The SEM provides morphological information with higher resolution than optical microscopes. In studying forensic evidence for wildlife enforcement activities, it is often useful to identify markings on bullets, shells or shotgun pellets that are obtained as evidence in the investigation of poaching or killing of protected animals, or other illegal hunting. By matching such marks with ammunition fired by suspected weapons, a case can be supported against individuals charged with illegal activities. In a study of the use of SEMs for such investigations, SEMs provided clearer matches of firing pin imprints than optical microscopy.[29] In this same study, a SEM was also used to assist in the correlation of surface features on shotgun pellets.

Some surface features are so small that a FESEM is required to obtain the necessary resolution to see the objects. Nanocrystals of InAs were observed in a FESEM at 100 000× magnification in order to determine the size, orientation and distribution of the nanocrystals on a modified silicon surface.[30] Such an observation is relatively quick to accomplish with a high-resolution SEM, whereas the same observation with a TEM would require the lengthy preparation of a plan-view sample. However, cross-section TEM was used in this study to illustrate the uniformity of the crystal lattice of the nanocrystals.

7 APPLICATIONS TO FIBERS

Fibers used for a variety of purposes have varying size, structure and orientation with respect to neighboring fibers. The surface structure of natural fibers often give clues to their plant source. The SEM is very useful in examining such surface structure, and the source of fibers in a given sample of fabric can often be deduced. It is a particular challenge to make such an identification of the fibers in an object as old as an ancient Egyptian mummy. By checking fiber surface substructure in the wrappings of an Egyptian mummy with a SEM, flax and ramie have been identified as constituents of the fabric.[31] The study correlated the choice of fibers with the properties helpful for preserving the mummy.

The structural properties of products that contain fibers depend on the choice of fibers and also on the arrangement of the fibers. The strength of materials containing woven fibers is often nonisotropic, and the type of weave often determines some of the directional variables of the strength of the material. A SEM can be used to examine such questions. In the case of dental prostheses, the arrangement of the internal strengthening fibers have been studied by a SEM. By removing the solid matrix material surrounding the fibers, the strengthening fibers used in materials for the prostheses were observed by a SEM. The size of the fibers was measured and the weaves used in producing layers of fibers were identified.[32]

Studies involving hydration are often essential in studying fiber products. Studies of the effect of water on paper, for example, can include the application of ink to printing paper or a water drop on absorbent paper. It would be impossible to study liquid water drops in a typical high-vacuum SEM. Such studies are perfect examples of opportunities for the use of an ESEM, where samples can be kept in a hydrated environment. The fibers of the paper do not need to be coated or otherwise treated to avoid charging when an ESEM is used. A variety of paper components and sizing agents were investigated by an ESEM in a recent study,[33] yielding information about the spreading and absorption of water drops.

8 APPLICATIONS TO POLYMERS

The surfaces of special-purpose polymer layers can often be observed easily by a SEM. Testing a polymer surface for binding of platelets was needed when a NO-releasing membrane was proposed as a way to limit the fouling of sensors exposed to blood flow. Platelets are blood cells involved with blood clotting, and they attach easily to a wide range of polymers exposed to blood flow. Sensors designed to function while exposed to the blood circulation can often have a limited lifetime because of platelet binding. In a study comparing a NO-releasing membrane with other polymers, SEM images showed cells and biological materials bound to both membranes, but with far fewer cells attached to the active membranes.[34]

Interior surfaces of solids can be observed in a SEM by preparing fractured samples. The strength of polymer mixtures is often correlated with the nature of the mixing of the components, and the mixing can be revealed by SEM observations of the domains formed in solids by the different components. Polymer mixtures can show different sizes and distributions of components, depending on the compatibilizers that are included in the mixtures. Boundaries of different components have different binding strengths, and a fracture through a mixed solid can give clues to the

strength of binding. If a fracture follows the interface between components, the binding of those components is less than the bulk; if the fracture goes through the bulk of a component, the binding to the nearby components is stronger than that within the bulk. Scanning electron microscopy of fractured samples can determine whether fractures follow interfaces between components or pass through components.[35] A study of polystyrene–polypropylene mixtures showed that there was weak binding between components under some compatibilizer conditions, better adhesion in other conditions.[36] By using an ESEM to allow higher vapor pressures than are typical of SEMs, the kinetics of the formation of layers from polymer mixtures can be studied.[37]

Microelectromechanical systems (MEMS) devices are generally tiny machines that are constructed using methods that were derived from fabrication techniques developed originally for the microelectronics industry. Microelectromechanical structures often have surface features at the micrometer level (high resolution), whereas the devices can extend over millimeters (large depth-of-field). Light microscopes have too little depth of field at the necessary resolution to produce good images of many of these structures. The SEM is particularly valuable for observing MEMS devices,[38] because of extensive depth of field and good resolution. For pop-up structures, where components are fabricated on a plane and then are raised to a perpendicular or oblique angle, the need for a good depth of field is particularly important.[39]

In the construction of micromechanical structures, it is typical to print patterns onto polymer resists by exposing the resist to light prior to development and then etching. Printing the patterns on the resist by electron-beam writing promises to yield higher resolution and smaller features. The ESEM is a valuable device for prototyping electron-beam lithography methods without needing to metal-coat or otherwise prepare the resist.[40]

9 APPLICATIONS IN BIOLOGY

SEMs have been widely used in biology and the number of applications is vast. Observations have been made of organisms (from insects to bacteria), cells (including blood cells and tissue cells), and even large molecular arrays (like nuclear pore complexes and muscle filaments). For the smallest objects, a FESEM is required. When hydrated samples need to be examined, an ESEM is appropriate For many samples, a typical SEM is very productive. An excellent way to survey biological applications of SEMs is by using one of several atlases dedicated to this topic.[41]

For biological objects, hydration is often extremely important. Although hydration can be maintained in an ESEM, sometimes it is necessary to have more resolution than is available with an ESEM, or it might be necessary to prevent movement of the sample. In such a case, hydration can be maintained in high vacuum with frozen samples.[10] Insects feeding on plant fluids can be observed when the sample is preserved in the SEM by freezing.[42]

The search for specific molecules is often important in biology, and high-atomic-number elements can be bound to compounds that can be used for labeling specifically targeted molecules in a biological sample. The backscattered signal in a SEM is particularly useful for detecting and localizing heavy-atom label compounds in biological materials.[43]

Determining the molecular weight is crucial in understanding the structure and function of components of molecular complexes, and a special kind of FESEM with a full-immersion lens is capable of making such measurements. The molecular weight of individual components can often be measured without removing them from their complexes, an advantage over biochemical techniques where purification is usually necessary. The mass of a molecule is determined by counting the number of electrons scattered from the molecule using a transmitted-electron detector; for that reason, this kind of instrument is often called a scanning transmission electron microscope (STEM). The repeating unit of some *Neurospora* helical filaments was found to weigh 90 kDa by the transmitted-electron method, suggesting a different ratio of molecular components than was indicated by quantitative gel electrophoresis.[44]

Another study, which compared structural observation with molecular weight measurement, determined that the globular particle of a helical filament core in *Leishmania* consisted of a polypeptide dimer.[45] It is very useful when the molecular weight can be determined in the microscope simultaneously with the observation of structure, as was also done in the determination of phage secretin multimer composition.[46] Measurement of mass per unit area was found to be useful in modeling the molecular structure of the cell envelope of terminally differentiated keratinocytes (skin cells).[47] A special STEM uses low temperature to preserve samples better while measuring molecular weight.[48]

10 INSTRUMENT PERFORMANCE

10.1 Calibration

The standard calibration for SEM magnification in the USA is done using the NIST Standard Reference Material SRM 484G, which contains metal lines with certified

spacings. As an alternative, it is common to perform magnification calibration with secondary standards from a variety of commercial manufacturers and private laboratories. Secondary standards offer some advantages over the NIST reference, such as lower cost, higher contrast, or a greater variety of test patterns. To calibrate the magnification, typically the reference standard is photographed with the SEM in the conventional way, and the magnification is determined by dividing the measured size of the feature on the photograph by the certified size of the feature on the reference standard. It is valuable to calibrate the horizontal and vertical dimensions of the SEM image separately, because different chains of amplifiers are used to generate the scans in these perpendicular dimensions. In many cases SEMs have automatic readouts and scale bars that print the magnification on the photographs. In these cases it is useful to adjust the scan amplifiers until the printed magnification is equal to the calibrated magnification.

10.2 Resolution Tests

Because of the high contrast of gold in the secondary-electron image, it is common to observe the resolution of a SEM by photographing a surface covered with islands of evaporated gold of various sizes. Resolution is estimated by measuring the separation of the smallest features that can be determined to be distinct. A variety of both natural and fabricated test structures are also available for resolution testing.

11 COMPARISON WITH OTHER TYPES OF MICROSCOPY

11.1 Optical and Confocal Microscopes

Optical microscopy can quickly yield magnified images of samples, and the varying colors of different materials helps in making quick identifications. Also, light microscopes can be used to observe samples in liquid solutions, in air, or under a variety of other conditions that still allow an optical path to reach the samples. However, the resolutions of light microscopes are fundamentally limited by the wavelength of light itself.

An exception to this general rule is the near-field scanning optical microscope, where the resolution is set by a tiny aperture rather than by the wavelength of light. Here, like in a SEM, the image is created by scanning the aperture across the sample, which is why the resolution of the near-field scanning optical microscope is disconnected from the magnification. In a conventional light microscope, the optical system of lenses that create the magnification is also responsible for creating, and limiting, the resolution.

Because electrons have a substantially smaller wavelength than light, SEMs are fundamentally capable of higher resolution than light microscopes. Also, because of the small wavelengths of electrons, SEMs operate with a much smaller objective aperture angle than light microscopes, giving a large depth of focus that provides a satisfying depth of view into three-dimensionally extensive samples. This effect will be familiar to photographers who use manual cameras with control of the lens aperture. When photographing someone standing in front of the Grand Canyon, at f/16 (a small aperture) both the person and the canyon will be in focus, whereas at f/2 (a large aperture) the person will be in focus and the canyon will be blurry.

An exception to the limited depth of field of optical microscopes is another kind of scanning light microscope, the confocal microscope. In this instrument a point of light is scanned through the volume of a transparent sample while the illuminated point is imaged by the objective. The scanning mode creates depth information that can be displayed in three dimensions. Although this mode of image recording increases the distance into the sample that can be imaged with good sharpness, it does not significantly improve the fundamental resolution. Because the objective still images the illuminated point, the resolution of the confocal microscope is still limited by the wavelength of light.

11.2 Transmission Electron Microscope

TEMs typically have higher resolution than SEMs primarily because transmission microscopes use high-field immersion objective lenses with very small aberration coefficients. However, because electrons need to be transmitted through samples in a TEM, generally samples need to be very thin and substantial three-dimensional information is not available. Thinning the samples is a time-consuming process, so preparation of samples for a SEM is typically much quicker than for a TEM.

11.3 Atomic Force and Scanning Tunneling Microscopy

Atomic force microscopy and scanning tunneling microscopy are examples of the general class of scanning probe microscopy. These methods are like scanning electron microscopy, except that instead of an electron beam traveling over the sample, a physical mechanical probe is driven over the sample. The interaction of the probe with the sample makes the signal that makes the image. The signal is reproduced on a screen in synchrony with the probe's position on the sample, in a way that is very similar to the process in a SEM. Because of the physical size of the probe, the depth of field of probe microscopes is typically much less than that of SEMs. However, because

small points on the probe are generally responsible for a majority of the signal, the scanning probe microscope can often have atomic-level resolution on flat samples.

ACKNOWLEDGMENTS

I thank Dr Gary E. McGuire of MCNC for encouraging this work, Ray Gundersdorff and Nancy Chrestensen of JEOL for providing information about JEOL instruments, Evan Slow of FEI for providing information about Philips instruments, and Elisa Carr of Hitachi Scientific for providing information about Hitachi instruments.

ABBREVIATIONS AND ACRONYMS

ESEM	Environmental Scanning Electron Microscope
FESEM	Field Emission Scanning Electron Microscope
FTIR	Fourier Transform Infrared
MEMS	Microelectromechanical Systems
SEM	Scanning Electron Microscope
STEM	Scanning Transmission Electron Microscope
TEM	Transmission Electron Microscope

DISCLAIMER

Mention of specific commercial products or vendors within this article does not imply endorsement.

RELATED ARTICLES

Forensic Science **(Volume 5)**
Scanning Electron Microscopy in Forensic Science

Surfaces **(Volume 10)**
Electron Microscopy and Scanning Microanalysis

REFERENCES

1. J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, *Scanning Electron Microscopy and X-ray Microanalysis: A Text for Biologists, Materials Scientists, and Geologists*, 2nd edition, Plenum Publishing Corp., New York, 1992.
2. R.F. Egerton, L. Wan, 'Low-cost Image-capture System for a Scanning Electron Microscope', *J. Microsc.*, **191**(2), 113–115 (1998).
3. O.C. Wells, *Scanning Electron Microscopy*, McGraw-Hill Book Company, New York, 1974.
4. M. Knoll, 'Static Potential and Secondary Emission of Bodies Under Electron Irradiation', *Z. Tech. Phys.*, **11**, 467–475 (1935).
5. M. von Ardenne, 'Das Elektronen-Rastermikroscop', *Z. Tech. Phys.*, **19**, 407–416 (1938).
6. V.K. Zworykin, J. Hillier, R.L. Snyder, 'A Scanning Electron Microscope', *ASTM Bull.*, **117**, 15–23 (1942).
7. D. McMullan, 'An Improved Scanning Electron Microscope for Opaque Specimens', *Proc. IEE*, **100**(II), 245–259 (1953).
8. T.E. Everhart, 'Persistence Pays Off: Sir Charles Oatley and the Scanning Electron Microscope', *J. Vac. Sci. Technol. B*, **14**(6), 3620–3624 (1996).
9. A. Boyde, 'Freezing, Freeze-fracturing, and Freeze-drying in Biological Specimen Preparation for the SEM', in *Scanning Electron Microscopy*, eds. O. Johari, I. Corvin, IIT Research Institute, Chicago, 1044–1046, 1974.
10. P. Echlin, *Low-temperature Microscopy and Analysis*, Plenum Press, New York, 1992.
11. L. Reimer, in *Scanning Electron Microscopy*, 2nd edition, ed. P.W. Hawkes, Springer-Verlag, Berlin, Springer Series in Optical Sciences, Vol. 45, 1998.
12. A. Rose, 'Television Pickup Tubes and the Problem of Vision', *Adv. Electronics Electr. Phys.*, **1**, 131–166 (1948).
13. A.N. Broers, 'Some Experimental and Estimated Characteristics of the Lanthanum Hexaboride Rod Cathode Electron Gun', *J. Phys. E: J. Scient. Instrum.*, **2**, 273–276 (1969).
14. A.V. Crewe, M. Isaacson, D. Johnson, 'A Simple Scanning Electron Microscope', *Rev. Sci. Instrum.*, **40**, 241–246 (1969).
15. A.V. Crewe, 'Scanning Microscopes: The Approach to 1 Å.', in *Quantitative Scanning Electron Microscopy*, eds. D.B. Holt, M.D. Muir, P.R. Grant, I.M. Boswarva, Academic Press, New York, 65–73, 1974.
16. P.S.D. Lin, M.K. Lamvik, 'High Resolution Scanning Electron Microscopy at the Subcellular Level', *J. Microsc.*, **103**, 249–257 (1975).
17. Y. Yamamoto, A. Yamada, T. Miyokawa, N. Tamura, 'Development of a High Resolution Semi-in-lens Digital Field Emission Scanning Electron Microscope: JSM-6340F', *JEOL News*, **32E**(1), 39–41 (1996).
18. Anonymous, 'New Hexalens Column Views Uncoated and Untreated Specimens', in *Scope*, Philips Electron Optics, Eindhoven, 6–7, Vol. 8, 1998.
19. J. Zach, M. Haider, 'Correction of Spherical and Chromatic Aberration in a Low Voltage SEM', *Optik*, **98**(3), 112–118 (1995).
20. 'HD-2000 Ultra-thin Film Evaluation System', Product Brochure, Hitachi Scientific Instruments, Gaithersburg, MD.
21. Anonymous, 'Hitachi to Market a New Semiconductor Ultra-thin Film Evaluation System Featuring High

Sample Throughput and High Image Contrast', News Release, Hitachi Asia Ltd, Singapore, 1997.
22. D. Voreades, 'Secondary Electron Emission from Thin Carbon Films', *Surf. Sci.*, **60**, 325–348 (1976).
23. O.C. Wells, A.N. Broers, C.G. Bremer, 'Method for Examining Solid Specimens with Improved Resolution in the Scanning Electron Microscope (SEM)', *Appl. Phys. Lett.*, **353**, 353–355 (1973).
24. S.D. Chaffee, M. Hyman, M.W. Rowe, N.J. Coulam, A. Schroedl, K. Hogue, 'Radiocarbon Dates on the All-American Man Pictograph', *Am. Antiquity*, **59**(4), 769–781 (1994).
25. J.M. Kenoyer, M. Vidale, 'A New Look at Stone Drills of the Indus Valley Tradition', in *Materials Issues in Art and Archeology III*, eds. P.B. Vandiver, J.R. Druzik, G.S. Wheeler, I.C. Freestone, Materials Research Society, San Francisco, CA, 495–518, 1992.
26. H. Hamdan, M.N.M. Muhid, S. Endud, E. Listiorini, Z. Ramli, '^{29}Si MAS NMR, XRD and FESEM Studies of Rice Husk Silica for the Synthesis of Zeolites', *J. Non-Cryst. Solids*, **211**(2), 126–131 (1997).
27. W. Ma, P.W. Brown, 'Hydration of Sodium Phosphate-modified High Alumina Cement', *J. Mater. Res.*, **9**(5), 1291–1297 (1994).
28. W.D. Blanchard, E.M. Stern, L.P. Stodulski, 'Analysis of Materials Contained in Mid-4th to Early 7th Century A.D. Palestinian Kohl Tubes', in *Materials Issues in Art and Archaeology III*, eds. P.B. Vandiver, J.R. Druzik, G.S. Wheeler, I.C. Freestone, Materials Research Society, San Francisco, CA, 239–254, 1992.
29. M.-J. Mann, 'Scanning Electron Microscopy in Wildlife Forensic Science', *Scanning*, **15**(Suppl. III), III-11–III-13 (1993).
30. T. Mano, H. Fukioka, K. Ono, Y. Watanabe, M. Oshima, 'InAs Nanocrystal Growth on Si (100)', *Appl. Surf. Sci.*, **132**, 760–764 (1998).
31. M.L. Proefke, K.L. Rinehard, M. Raheel, S.H. Ambrose, S.U. Wisseman, 'Probing the Mysteries of Ancient Egypt: Chemical Analysis of a Roman Period Egyptian Mummy', *Anal. Chem.*, **64**(2), 105A–111A (1992).
32. P.K. Vallittu, 'Compositional and Weave Pattern Analyzes of Glass Fibers in Dental Polymer Fiber Composites', *J. Prosthodont.*, **7**(3), 170–176 (1998).
33. A. Liukkonen, 'Contact Angle of Water on Paper Components: Sessile Drops Versus Environmental Scanning Electron Microscope Measurements', *Scanning*, **19**(6), 411–415 (1997).
34. C. Espadas-Torre, V. Oklejas, K. Mowery, M.E. Meyerhoff, 'Thromboresistant Chemical Sensors Using Combined Nitric Oxide Release/Ion Sensing Polymeric Films', *J. Am. Chem. Soc.*, **119**(9), 2321–2322 (1997).
35. S. Horiuchi, N. Matchariyakul, K. Yase, T. Kitano, H.K. Choi, Y.M. Lee, 'Compatibilizing effect of Maleic Anhydride Functionalized SEBS Triblock Elastomer Through a Reaction Induced Phase Formation in the Blends of Polyamide 6 and Polycarbonate: 2. Mechanical Properties', *Polymer*, **38**(1), 59–78 (1996).
36. Z. Horák, V. Fořt, D. Hlavatá, F. Lednický, F. Vecerka, 'Compatibilization of High-impact Polystyrene/Polypropylene Blends', *Polymer*, **37**(1), 65–73 (1996).
37. J.L. Keddie, P. Meredith, R.A.L. Jones, A.M. Donald, 'Film Formation of Acrylic Lattices with Varying Concentrations on Non-film-forming Latex Particles', *Langmuir*, **12**(16), 3729–3801 (1996).
38. H. Becker, K. Lowack, A. Manz, 'Planar Quartz Chips with Submicron Channels for Two-dimensional Capillary Electrophoresis Applications', *J. Micromech. Microeng.*, **8**, 24–28 (1998).
39. A. Friedberger, R.S. Muller, 'Improved Surface-micromachined Hinges for Fold-out Structures', *J. Microelectromech. Syst.*, **7**(3), 315–319 (1998).
40. B.K. Paul, 'Effects of Gas Pressure on Low-pressure Electron-beam Lithography', *Scanning*, **19**(7), 466–468 (1997).
41. G. Shih, R. Kessel, *Living Images: Biological Microstructures Revealed by Scanning Electron Microscopy*, Science Books International, Boston, 1982.
42. L.J. Crews, M.E. McCully, M.J. Canny, C.X. Huang, L.E.C. Ling, 'Xylem Feeding by Spittlebug Nymphs: Some Observations by Optical and Cryo-scanning Electron Microscopy', *Am. J. Botan.*, **85**(4), 449–460 (1998).
43. R.F. Stump, J.R. Pfeiffer, J. Seagrave, J.M. Oliver, 'Mapping Gold-labeled IgE Receptors on Mast Cells by Scanning Electron Microscopy: Receptor Distributions Revealed by Silver Enhancement, Backscattered Electron Imaging, and Digital Image Analysis', *J. Histochem. Cytochem.*, **36**, 493–502 (1988).
44. G. Hutchinson, W. Tichelaar, H. Weiss, K. Leonard, 'Electron Microscope Characterization of Helical Filaments Formed by Subunits I and II (core proteins) of Ubiquinol: Cytochrome C Reductase from *Neurospora* Mitochondria', *J. Struct. Biol.*, **103**, 75–88 (1990).
45. Y.-D. Stierhof, M. Wiese, T. Ilg, P. Overath, M. Häner, U. Aebi, 'Structure of a Filamentous Phosphoglycoprotein Polymer: The Secreted Acid Phosphatase of *Leishmania mexicana*', *J. Mol. Biol.*, **282**(1), 137–148 (1998).
46. N.A. Linderoth, M.N. Simon, M. Russel, 'The Filamentous Phage pIV Multimer Visualized by Scanning Transmission Electron Microscopy', *Science*, **278**(5343), 1635–1638 (1997).
47. M. Jarnik, M.N. Simon, A.C. Steven, 'Cornified Cell Envelope Assembly: A Model Based on Electron Microscopic Determinations of Thickness and Projected Density', *J. Cell Sci.*, **111**(8), 1051–1060 (1998).
48. C. Tribet, D. Mills, M. Haider, J.L. Popot, 'Scanning Transmission Electron Microscopy Study of the Molecular Mass of Amphipol/Cytochrome b_6f complexes', *Biochemie*, **80**, 475–482 (1998).

Scanning Probe Microscopy, Industrial Applications of

Axel Born
University of Hamburg, Hamburg, Germany

1 Introduction **1**
2 Examples of Industrial Applications of Scanning Probe Microscopy **4**
2.1 Scanning Force Microscopy of Hair 4
2.2 Magnetic Force Microscopy of Tapes 5
2.3 Scanning Thermal Microscopy of Hot Spots 6
2.4 Kelvin Force Microscopy of Metals 8
2.5 Scanning Capacitance Microscopy and Scanning Spreading Resistance Profiling of Dopants 9
2.6 Conducting Atomic Force Microscopy of Oxides 11
3 Perspective and Future Developments **12**
Acknowledgments **13**
Abbreviations and Acronyms **13**
Related Articles **14**
References **14**

Scanning probe microscopy (SPM) has developed to become an integral part of the hierarchy of analytical tools in industrial surroundings, owing to its excellent spatial resolution and its ability to work under ambient conditions. The purpose of this article is to give an insight into the potential of SPM techniques in industrial applications today. Owing to the wide variety of possible applications, this article can present only a small selection of uses.

The given examples highlight typical aspects of the current analytical work of different industrial laboratories. First, an introduction to the specific industry and its analytical obstacles is given. Second, a solution using a SPM method is demonstrated and the particular SPM method is explained in detail. Finally, an actual measurement is given for each example.

To summarize, with the necessary precautions concerning stable conditions during the measurement, SPMs are a useful tool for industrial applications which fulfill the demands of industry today. SPMs are commercially available, rapid, quantitative and reproducible. Furthermore, they are available with a user-friendly interface. The increasing number of companies that are using SPMs in the laboratory and manufacturing every day provides evidence of the rising popularity of SPM.

1 INTRODUCTION

The twenty-first century may be called the century of nanoscience. Chemistry, biology and physics, and especially semiconductor devices, will come together on the nanometer length scale. Physical, chemical and biological processes must be studied in real space and under environmental conditions. These requirements will push forward a new type of analytical tool, namely SPM.[1–3] This technique is based on scanning tunneling microscopy (STM),[4–7] which was developed in 1982 at the IBM Zurich Research Laboratory (Switzerland). Within 10 years it could be found in almost every laboratory around the world, and now, nearly 20 years later, SPM is being used not only in laboratories, but also in production. The purpose of this article to give an insight into the potential of SPM techniques in industrial applications today. Owing to the wide variety of possible applications, this article can present only a small selection of uses.

Since the early days of STM, many different types of SPM (Table 1) have been developed, all of which are based on the same principle: a small probe is brought in close proximity to the surface of a sample in order to measure near-field interaction (Figure 1). In most cases, feedback keeps the interaction strength constant during the passing of the probe across the sample. Scanning is achieved by means of piezoelectric actuators which define the set-up of the probe and sample with an accuracy of the order of a few picometers (10^{-12} m). We are then able to scan a probe across a sample at a distance smaller than the characteristic wavelength of the specific interaction (for STM and the wavelength electrons, this is a fraction of a nanometer, depending on the energy). The so-called near-field interaction regime permits a spatial resolution which is beyond the Abbe limitation.

STM gives us a panorama of conductive surfaces with atomic resolution. SPM has been used in biology, chemistry and physics.[8] In biology, it is valued for its ability to measure in an ambient environment or an ionic solution. It is therefore possible to observe viruses, bacteria or cells in action (e.g. the pulsing of a heart cell) or the interaction between DNA and specific proteins. This cannot be done by using scanning electron microscopy (SEM) or transmission electron microscopy (TEM), because of the need for ultrahigh vacuum (UHV) conditions and/or metal coating. However, the achievable resolution of standard optical microscopy is limited to

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Promising SPM methods

Abbreviation[a]	Name	Measured property
c-AFM	Conducting atomic force microscopy	FN tunneling current
EFM	Electrostatic force microscopy	Electrostatic field
KFM	Kelvin force microscopy	Surface potential
MFM	Magnetic force microscopy	Magnetic stray field
SCM	Scanning capacitance microscopy	Capacitance (dopant density)
SFM	Scanning force microscopy	Force
SNOM	Scanning near-field optical microscopy	Light
SSRP	Scanning spreading resistance profiling	Spreading resistance (dopant density)
SThM	Scanning thermal microscopy	Temperature, thermal conductivity
STM	Scanning tunneling microscopy	Tunneling current

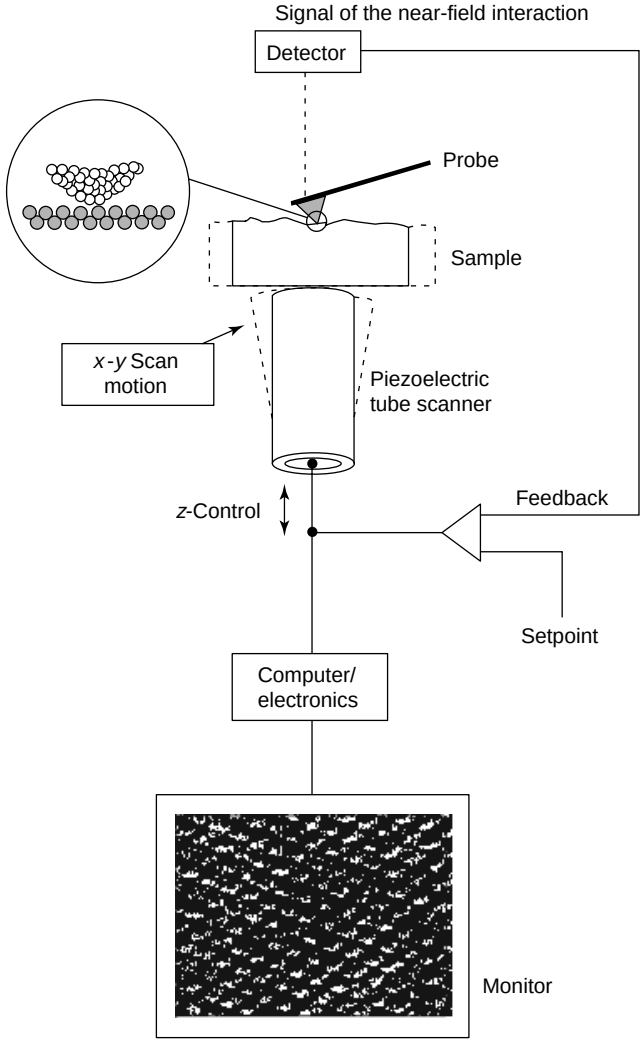

Figure 1 Schematic diagram of an experimental set-up for a typical scanning probe microscope.

0.5 μm and far removed from the atomic resolution available by SPM.

A resolution of better than 1 nm is required for the imaging of almost all chemical processes. SPM has been used to observe rotational and vibrational states of single molecules.[9] It allows the identification of the molecule and a detailed understanding of the chemistry on the nanometer scale. In particular, in the interesting field of the self-assembled monolayer (SAM), SPM has shown great potential.[8] Whereas the self-arrangement of a monolayer of an Au_{55} cluster can be observed by SEM or TEM, SPM can give additional information about the electronic structure of the cluster (e.g. coulomb blockade and staircase or quantum size effects). SPM allows the observation of the electronic spectra of a single cluster depending on the environment (charge fluctuation).[10] Such investigations are of utmost importance for the use of this cluster for single electron devices and therefore for the future successor of the metal-oxide semiconductor field-effect transistor (MOSFET).[11]

The gate length of this dominant structure of the semiconductor industry of the order of 180 nm today and should break through the 100 nm "frontier" in a couple of years. The semiconductor industry association (SIA) roadmap (a selection is given in Table 3), which represents an agreement with regard to future developments in the semiconductor industry, predicts a gate length of 25 nm by 2015. MOSFETs with a gate length down to 14 nm can be fabricated in the laboratory,[12] but until now not at the mass-production level. Owing to the problems arising from dopant fluctuations or thin-gate dielectrics, we shall need new MOSFET designs with gate length down to 50 nm.[13] These new designs require a precise knowledge of physical properties (e.g. voltage, current or dopant density) at the nanometer length scale. This is exactly the field for SPM application, whereas most

conventional analytical techniques [e.g. secondary ion mass spectrometry (SIMS)] fail at a resolution below 100 nm.[14]

Besides high spatial lateral resolution and operation under environmental conditions, there are normally some additional demands required for an industrial application: commercial availability; user-friendly interface; rapidity; quantitativeness; and reproducibility. These points will be discussed in detail in the following sections.

Most common SPMs for industrial application are commercially available from an increasing number of companies (Table 2) in a price range from US$ <50 000 for laboratories to several million dollars for operation in clean-room conditions and production.

These commercial SPMs are provided with a user-friendly interface that permits a rapid acquaintance with the new tool. A ''plug and measure'' approach is possible for each user today. However, the main problem is interpreting and understanding the images.

SPMs are a time-consuming analytical tool owing to the small scan range (100 μm) and the relatively slow scanning speed of around 1 Hz per line. An image with a resolution of 512 × 512 pixels takes roughly 8 min. Some developments will produce arrays of cantilevers (e.g. 32 × 32 = 1024 cantilevers) or high-speed cantilevers, which should permit scan speeds of several meters per second.[15,16] However, SPM has to be used within a hierarchy of analytical tools. No single instrument can give a complete set of data for a sample. There are without doubt faster analytical tools (albeit with a poor spatial resolution) which can be used to examine whole samples before going into detail into specific areas of the sample by means of SPM. For example, in failure analysis in the semiconductor industry, a first step is to localize failure by means of conventional electrical measurements. This localizes failure to an area of tens of micrometers. In a second step, one can use analytical tools for the evaluation of the physical failure, and by involving computer-aided test stands, the area can be automatically scanned by SPM. At this level, SPM can be used in product monitoring. An increasing number of SPMs are being used in mass production conditions. Most are used to monitor topographic features from particular areas of interest of a wafer of semiconductor devices. Consequently, slow scan rates are no longer a handicap.

Table 2 SPM manufacturers (selection)

Manufacturer	Country	Web page url: http://www.
Burleigh Instruments	USA	burleigh.com
Danish Micro Engineering	Denmark	dme-spm.com
Digital Instruments	USA	di.com
JEOL	Japan	jeol.com
Molecular Imaging	USA	molec.com
Nanonics	Israel	nanonics.co.il
NT-MDT	Russia	ntmdt.ru
Omicron	Germany	omicron.de
Quesant Instruments	USA	quesant.com
RHK Technology	USA	rhk-tech.com
Surface Imaging Systems	Germany	sis-gmbh.com
ThermoMicroscopes (formerly Park Scientific Instruments and TopoMetrix Corporation)	USA	thermomicro.com
Triple-O Microscopy	Germany	triple-o.de
Win Instruments	Germany	nanoscience.de
WiTec	Germany	witec.de

Quantitativeness and reproducibility are based on identical circumstances. Both require stable conditions for the probe and sample. The probe (in the case of a topographical measurement of a cantilever) suffers from wear between the apex of the cantilever tip and the samples, which in normal application in the semiconductor industry are extremely hard materials (e.g. silicon dioxide or silicon nitride). A change of the tip shape may lead to a change of the SPM response because the measured topography is a convolution between tip and sample. Knowledge of the tip shape can be gained from calibration structures or SEM measurements and allows the calculation of the ''real'' topography of the sample. However, a change of the tip shape during the measurement causes problems. This leads to the demand for new materials for cantilevers, e.g. diamond.

In the above-mentioned automated test stand, the tip shape is also controlled automatically between a pair of measurements. A deviation from a certain value leads to an exchange of the cantilever and, as a consequence, tip wear is no longer a problem. More difficult to achieve are stable conditions for the sample. The setting out of an instruction manual for a definable sample preparation is the first step to overcome this problem. Further, the ambient conditions must be taken into consideration, because a change of the relative humidity could lead to different results of SPM measurements on the same sample. The measurement of electrical, thermal or tribological properties depends on the thickness and the properties (e.g. conductivity) of the water layer of the sample. Defined conditions can be realized by placing the SPM in a glove box or flow box with an atmosphere of nitrogen or argon, which permits reproducible measurements, the basis for quantitative results. Furthermore, in almost all cases, a test sample is used to calibrate the probe. For some SPM techniques the measured data (e.g. voltage) must be converted into another physical property (e.g. dopant density).

This conversion requires a detailed understanding of the phenomena on the nanometer scale, and especially the interaction of the probe and sample. The near-field regime gives us resolution that is no longer limited by diffraction, but by geometrical properties (e.g. effective probing area). These properties have to be taken into account to obtain a model for the response of the SPM. The regime, sometimes called the regime of mesoscopic physics,[17] brings together the physics of the bulk material (macro) with the physics of atoms and molecules (micro) and is, although of great interest, not completely understood today. The modeling of the tip–sample interaction may therefore be complicated. Nevertheless, by using some simplifications, one can obtain a useful model for a first conversion with estimable failure. For some types of SPMs, e.g. STM[18] or SFM,[19] there are theoretical models that fit the experimental results very well. Other types suffer from a short lifetime and/or complexity, but improvements are in progress. For a great number of industrial applications, the operator has to distinguish between good and bad samples. This can be done by using SPMs in a more qualitative way.

Finally, with the necessary caution concerning the aforementioned conditions, SPMs are a useful tool for industrial applications which fulfill the demands of the industry today. SPMs are commercially available, rapid, quantitative and reproducible. Furthermore, they are available with a user-friendly interface. The increasing number of companies that are using SPMs in the laboratory and manufacturing every day provides evidence of the rising popularity of SPM.

2 EXAMPLES OF INDUSTRIAL APPLICATIONS OF SCANNING PROBE MICROSCOPY

The following examples highlight typical aspects of the analytical work of different industrial laboratories today. First, an introduction to the specific industry and its analytical obstacles is given. Second, a solution using a SPM method is demonstrated and the particular SPM method is explained in detail. Finally, an actual measurement is given for each example.

2.1 Scanning Force Microscopy of Hair

There is a large market centered around human and animal (e.g. lamb's wool) hair. Additionally, there are a large number of companies in the cosmetic and textile industries who deal with products for the treatment of hair. Consequently, a detailed understanding of hair and the influences of the products concerning the treatment of hair is the foundation of a market leader. In the past, investigation of hair was done by SEM, TEM and optical microscopy. SEM and TEM require high-vacuum conditions and/or a metallic coating of the specimen. That means that it is not possible to observe the same area of interest of a hair before and after the treatment. Optical microscopy, however, lacks a suitable resolution. Only SPM combines the possibility of working in environmental conditions with appropriate resolution. Furthermore, it enables in situ study of the processes during the treatment of the hair.

To measure the topography of an insulating surface, a SFM is used.[20–22] The probe, in this type of SPM, is a small cantilever with a sharp needle at the end. The cantilever is brought in close proximity to the sample, so that the tip of the needle touches the surface and the cantilever is bent out of shape for repulsive mode imaging. Imaging can also be done in the attractive mode, where deflection of the tip toward the surface by dispersion or other forces is measured. The deflection of the cantilever perpendicular to the sample is measured, (in all commercial SFMs) by an optical method. Feedback keeps the deflection constant during the movement of the tip across the sample, by means of a piezo actuator. This mode is called the constant force mode, because a constant deflection of the cantilever results in a constant force between the probe and sample. With the implementation of microfabricated silicon or silicon nitride cantilevers (with spring constants of the order of $0.01–1000\,\mathrm{N\,m^{-1}}$) and an optimized optical detection system, a cantilever deflection ranging from 0.01 nm to a few micrometers can be detected. This corresponds to forces between 10^{-13} and 10^{-4} N. Cantilevers with spring constants of $0.01–1\,\mathrm{N\,m^{-1}}$ are used in the contact mode, where the tip remains in contact with the sample during the entire measurement. An investigation under ambient conditions produces a water layer at the sample, which induces a large attractive force (capillary force) between tip and sample. As a result, the minimum possible force that can be used for SFM in air is of the order of 10^{-9} N. Even for such small forces, some structures of soft samples, e.g. biological samples, could be destroyed. In such cases, an intermittent mode, or so-called Tapping Mode™, can be used, where the cantilever touches the surface with a frequency near to the resonant frequency of the cantilever. Typically, the resonant frequency is in the range 50–500 kHz, corresponding to values of the spring constant of the cantilever of $1–50\,\mathrm{N\,m^{-1}}$, with an amplitude of 10 nm. By using the Tapping Mode™, minimal loading forces of 10^{-12} N can be realized. Hence the Tapping Mode™ is the most popular mode for the investigation of biological samples in air or in liquids. In the following, some examples are given for the analysis of human hair.

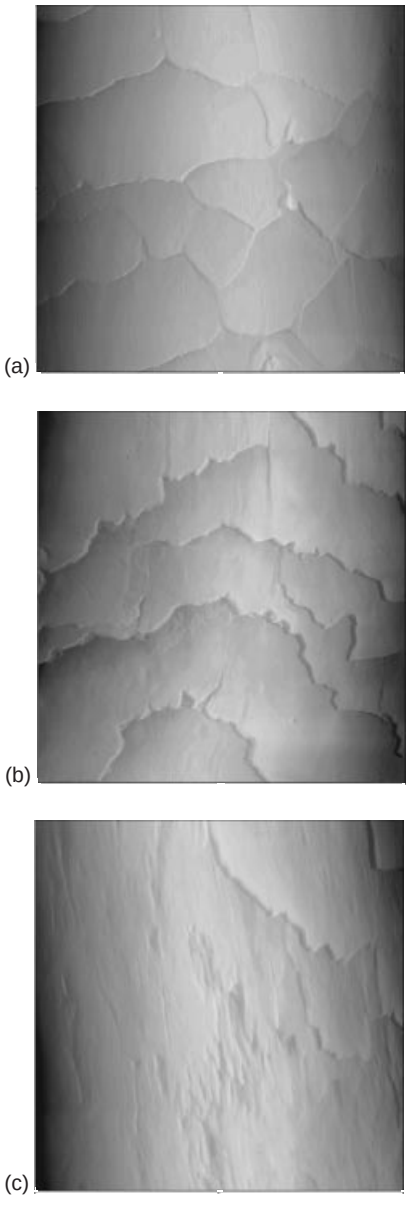

Figure 2 $25 \times 25\,\mu m$ images (topography) of an untreated female hair which were taken at different distances from the root: (a) 1; (b) 25; and (c) 45 cm.

A special problem concerning products for the treatment of hair is the existence of many different phases of the hair, depending for example on the length, which correlates with the age of the fragment of a specific hair. Close to the root, i.e. young hair (Figure 2a), the edges of the cuticle cells are smoother than for older ones (Figure 2b). Near the tip (Figure 2c) of the specific hair under investigation with a length of 45 cm and an age of about 3 years, almost all cuticle cells are broken away and the cortex of the hair can be seen. It is therefore apparent that any investigation of the effects of a treatment must

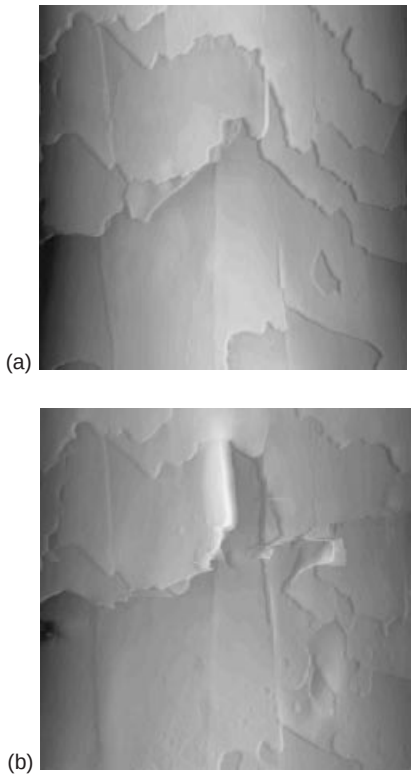

Figure 3 $30 \times 30\,\mu m$ images (topography) of the same area of a female hair which was taken (a) before and (b) after a bleaching process.

be realized by the observation of the same area of the same hair. An example is given in Figure 3(a) and (b), which shows the same area before and after a bleaching process. Parts of the cuticle cells are broken away owing to the oxidation process.

The results shown cannot be achieved merely by using SEM or TEM; rather, they can only be achieved by using a larger set of samples and SEM or TEM combined with statistical evaluation methods. Therefore, alternative analytical tools are more time-consuming, and it can be seen that for in situ analysis, there is no alternative to SPM today.

These results demonstrate the value of SFM as an analytical instrument within the cosmetic and textile industry.

2.2 Magnetic Force Microscopy of Tapes

Hard disk capacity is increasing more rapidly while the physical size decreases, owing to the shrinking of the bit size. The magnetic structure of current commercial devices (which represents the physical unit of the information) is of the order of some 100 nm. This structure has to be written and read with a frequency of more than 100 MHz, which places great demands on the recording

head. The immense market for storage devices, together with intense competition, force technology to produce higher data densities and transfer rates. A technological advantage can always be seen as an advantage in terms of competition.

The only analytical method for the investigation of magnetic structures of less than 1 μm in industrial surroundings is MFM.[23–25] All other methods such as electron holography, superconducting quantum interference device (SQUID), Hall bar and TEM suffer because of extraordinary sample preparation methods and/or lack of the appropriate resolution. There have been some developments to fabricate ultrasmall SQUID and Hall bar magnetometers and electromagnetic force coils on the top of a cantilever, but they remain commercially unavailable so far.

MFM is based on SFM (which is outlined in detail in section 2.1). It permits the measurement of the stray field of a magnetic structure with a spatial resolution of <50 nm. All MFM realizations measure either the force itself or the force gradient between a magnetic tip and the sample. Most of the commercial tips for MFM are silicon cantilevers with a thin magnetic coating such as cobalt, iron or nickel. A common MFM mode is the so-called ''lift mode'', which permits the virtually simultaneous measurements of the topography and the magnetic stray field of a sample. Most true noncontact methods show a mixture of topography and magnetic signal. To overcome this disadvantage, we need an independent measurement of both quantities. In the lift mode (Figure 4), initially, the cantilever measures the topography for one scan line and stores the information. The tip is then raised 30–60 nm and the stored data are used to retrace the topography, but vertically displaced. A deflection of the cantilever in the second scan line is an effect of a long-range interaction such as a magnetostatic field and is interpreted as the magnetic signal in MFM. By using a nonmagnetic tip, the deflection signal can be explained by electrostatic interaction. In MFM, the influence of an electrostatic field is avoided by the grounding of the tip and cantilever. In comparison with other SPM methods, the resulting data reflect the interaction of a small probe and a sample. A strong magnetic field from the tip can destroy a soft magnetic sample such as garnet films or permalloy, whereas a strong magnetic field from a sample can change the direction of the magnetization of the tip, which leads to a serious disturbance in the measurement. Therefore, it is important to make the right choice for the type of magnetic tip for a particular sample. For a successful interpretation of a MFM image, it is vital to know the precise magnetization of the tip. This can be done by using a calibration structure, such as a hard disk with a well-known magnetic arrangement.

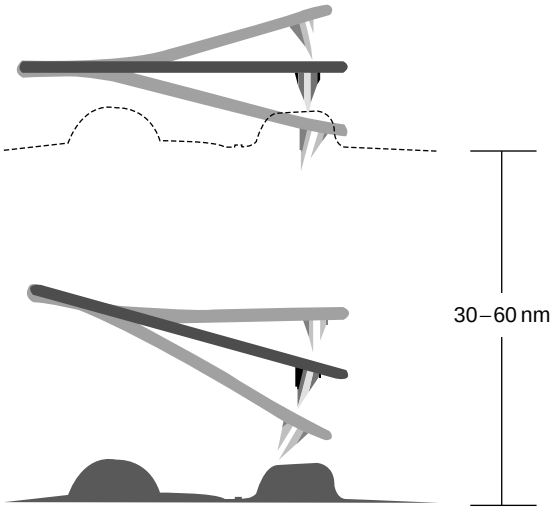

Figure 4 Schematic diagram of the lift-mode which is used to measure the long-range interactions.

MFM allows the direct observation of recording heads in operation with frequencies up to and even beyond 1 GHz.[26] However, the resulting structure, which was written by using recording heads, is the subject of many investigations concerning hard disks and any other magnetic storage devices. Figure 5 shows the topography and magnetic signal of a magnetic tape written with 6 MHz. A comparison of different tapes (not shown here) which were written with different parameters allows a check of the performance of the specific configuration.

Another industrial application is the use of MFM for measuring the current of a specific point on an integrated circuit (IC). Owing to the increasing integration of devices in IC fabrication, there is an equally increasing urgency to measure currents of a few hundred microamps with a lateral resolution of <100 nm. Hence some effort is being put into the use of MFM as a standard tool for failure analysis within the semiconductor industry. Here, quantitative results concerning the current of a specific wire are needed. This can be realized by using a calibration standard for the evaluation of the magnetic structure of the tip and an algorithm to convert the measured data into current values considering the specific geometrical properties of the sample.[27,28]

To summarize, MFM is a widespread technique within the magnetic storage industry, and is also being used within the semiconductor industry.

2.3 Scanning Thermal Microscopy of Hot Spots

The global semiconductor market has doubled with every second generation of technology. Business with ICs is more strongly linked with technological progress than in any other market. Worldwide market forces drive manufacturers to produce ever more highly integrated

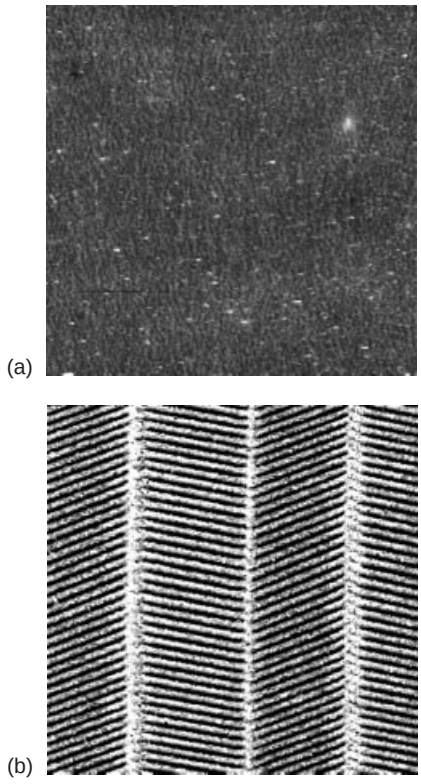

Figure 5 $40 \times 40\,\mu m$ images [(a) topography and (b) corresponding magnetic signal] of a magnetic tape (by courtesy of Exabyte) which had been written with 6 MHz before the MFM measurement was taken.

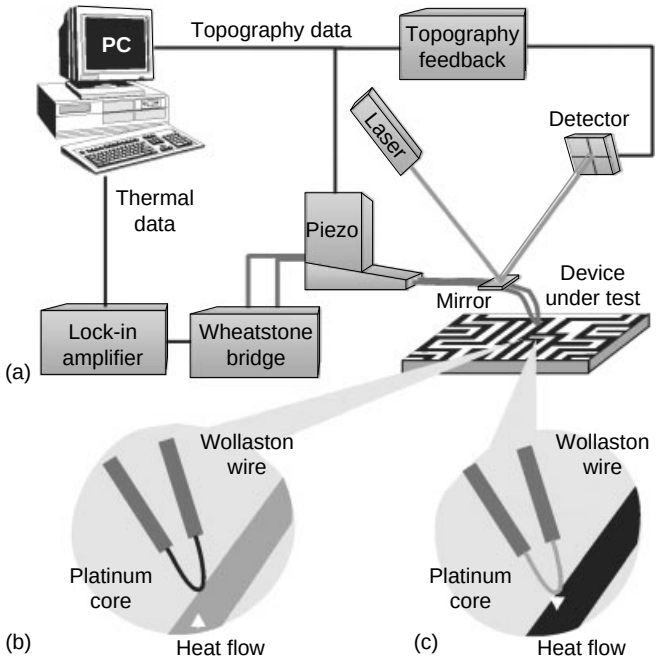

Figure 6 Schematic diagram of an experimental set-up for SThM. [Reprinted from G.B.M. Fiege, V. Feige, J.C.H. Phang, M. Maywald, S. Görlich, L.J. Balk, 'Failure Analysis of Integrated Devices by Scanning Thermal Microscopy (SThM)', in *Proceedings of the ESREF 98, Microelectron. Reliab.*, **38**, 957–961, Elsevier, Amsterdam, 1998, Copyright 1998, with permission from Elsevier Science.]

products, from the very large-scale integrated (VLSI) towards the ultralarge-scale integrated (ULSI) IC generation. This lays the foundation for the amazing law of Moore, which predicts that the numbers of transistors on an IC quadruples every 3 years. The degree of integration leads (besides the shrinking size of the unit devices) to thermal problems concerning local heat sources on an IC: so-called hot spots. A sophisticated design may avoid these hot spots, but it is an important requirement to be able to detect them with a lateral resolution of less than a few hundred nanometers. A standard tool in failure analysis for the localization of hot spots is liquid-crystal thermography (LCT). By means of thermal activation of the liquid crystal, it is possible to detect hot spots by optical microscopy. The radiation of hot electrons can be determined by emission microscopy (EM) or rather spectroscopic photon emission microscopy (SPEM). Thus the thermal effects of leakage currents in gate oxides can be measured. Both techniques (EM and SPEM) have a maximum spatial resolution of 0.5 µm. For higher magnification, SThM must be used.[29–31]

Useful SThM for the semiconductor industry combines SFM (see section 2.1) with a highly sensitive thermal detector at the end of the cantilever.[32,33] Detectors based on thermocouples, Schottky diodes and resistors have been developed. Figure 6(a–c) shows an schematic set-up of an SThM which is based on the measurement of a resistor consisting of a Wollaston wire with a platinum core. The wire is used as a cantilever to measure simultaneously the topography and the thermal properties of the sample. By applying a small current, the temperature distribution can be measured. Figure 7(a) and (b) shows a SThM measurement of an *n*-channel MOSFET operating in breakdown. A relative increase of the temperature (of the order of 1 K) can be determined close to the gate below the drain, where we can receive no signal by using SPEM, because of the shielding of the metallic drain contact.

A larger current can be used to probe the local thermal conductivity. A calibration process permits a quantitative determination of this value (in $W\,cm^{-1}\,K^{-1}$) with a spatial resolution of $<1\,\mu m$.[34] This gives the ability to track down defects which are covered by a passivation layer.

Finally, SThM gives us a chance to detect defects from the rear side of a thinned sample. This can be a great advantage for the use of SThM in failure analysis owing to the tendency for an increasing number of layers above the actual devices.

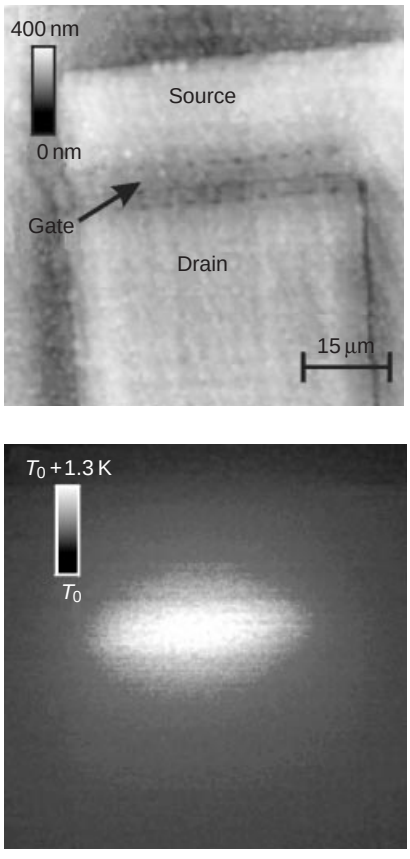

Figure 7 $80 \times 80\,\mu m$ images [(a) topography and (b) corresponding temperature micrograph] of an *n*-channel MOSFET biased at saturation. [Reprinted from G.B.M. Fiege, V. Feige, J.C.H. Phang, M. Maywald, S. Görlich, L.J. Balk, 'Failure Analysis of Integrated Devices by Scanning Thermal Microscopy (SThM)', in *Proceedings of the ESREF 98, Microelectron. Reliab.*, **38**, 957–961, Elsevier, Amsterdam, 1998, Copyright 1998, with permission from Elsevier Science.]

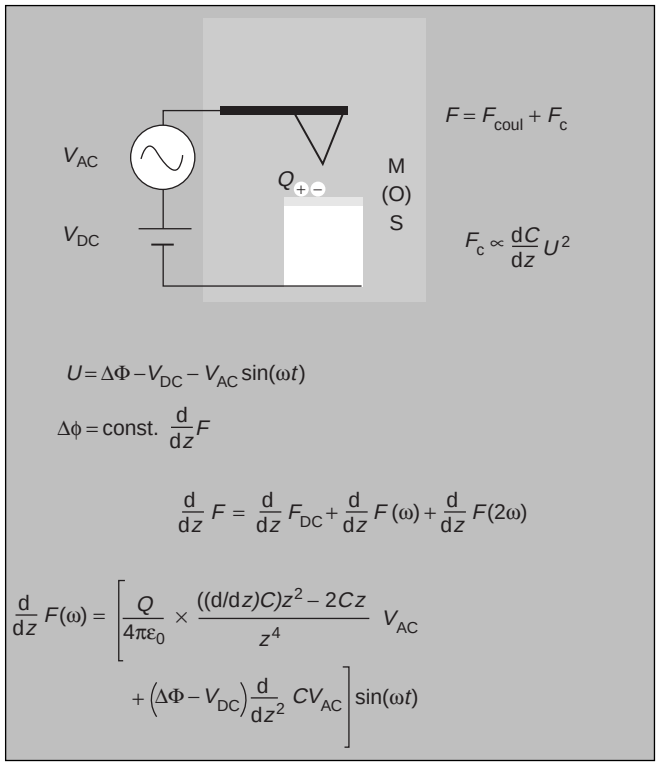

Figure 8 Schematic diagram of an experimental set-up for KFM and useful equations concerning the measured signal.

2.4 Kelvin Force Microscopy of Metals

The steel industry is sometimes thought to be outdated. Nevertheless, it is one of the biggest markets today. Metallurgy is a very important science for heavy industry. One of the most important areas of metallurgy is the investigation of corrosion. Because corrosion is well described by the surface potential, it can be measured by KFM[35–37] with a spatial resolution of <100 nm. Conventional techniques cannot be used in air or they suffer from poor lateral resolution.

This derivative of SFM is based on the lift mode, which was explained in section 2.2. A first scan gives us the topographical data and the second scan, at a defined scan height above the sample, gives us information about the long-range interaction. In the case of a non-magnetic metal, this is the electrostatic field between the sample and the metallic tip. A quantitative value is reached by applying a voltage to the tip which is composed of a direct current (DC) and an alternating current (AC) part. The frequency of the AC part is near the resonant frequency of the cantilever, which allows an extremely sensitive measurement. By measuring the first harmonic oscillation of the deflection signal of the cantilever by a lock-in amplifier, it is possible to balance this signal to zero because of an adjustment of the DC voltage (see Figure 8). This means that the applied DC voltage equals the reversed surface potential. Hence KFM produces a two-dimensional (2-D) map of quantitative values for the surface potential and the topography with a spatial resolution of <100 nm. By using a UHV/KFM system, it is possible to observe single dopant atoms.[38] On the other hand, the measurement is based on the long-range interaction between tip and sample, therefore each point probing is an average. Only 50% of the measured forces can be located within a region with a radius of 100 nm underneath the tip, depending on the tip shape. Hence KFM gives a high spatial resolution, but the value of the measured structure depends for small structures (<100 nm) on the size and the value of the surface potential of the surrounding.

Nevertheless, KFM is an advantageous tool for the analysis of corrosion.[39] KFM allows a time-dependent study of the process of corrosion in air or in a specific solution, such as salt water.

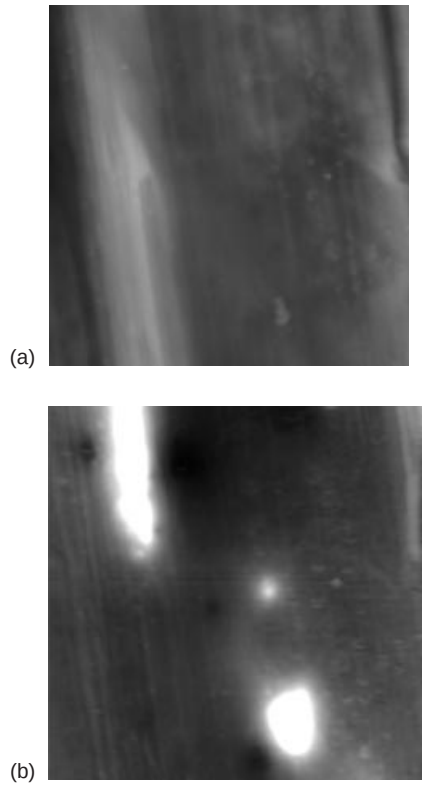

Figure 9 $20 \times 20\,\mu m$ images [(a) topography and (b) corresponding surface potential map] of a metallic foil.

In addition to corrosion, the local detection of the accumulation of elements, which results from a separation process of an alloy, is also an important area in metallurgy that can be investigated with KFM. An extreme treatment of metal, such as rolling it to a foil, can lead to significant modifications of the surface and, furthermore, to a modification of the properties of the metal in an electrochemical environment. Figure 9(a) and (b) shows the results of a KFM measurement of a metallic foil. The dots of the surface potential image seem to be accumulations of lead and can be located at the maxima in the topography. This can be a meaningful hint for the optimization of the treatment process.

To summarize, KFM can be used as an analytical tool in metallurgy and combines the possibility of in situ measurements with a high lateral resolution.

2.5 Scanning Capacitance Microscopy and Scanning Spreading Resistance Profiling of Dopants

Following the comments above, the semiconductor market is growing exponentially and requires technological progress. To acquire a feeling for the direction of this progress and to coordinate the research, the SIA created the National Technology Roadmap for Semiconductors (NTRS), which brings into focus the required progress over the next 15 years (Table 3).[40] A consequence of the NTRS is the need for an analytical tool for the assessment of 2-D dopant profiles with a spatial resolution of <10 nm. It is an urgent requirement to calibrate technology computer-aided design (TCAD) process simulators for future chip design. The failure analysis of semiconductor devices requires analytical tools with a spatial resolution of <50 nm today.

Lateral SIMS suffers from the necessity to design and prepare a special sample geometry. 'Selective etching' combined with SEM or TEM is lacking for quantification. Both methods need UHV conditions and are destructive procedures. A further candidate, electron holography, suffers from a large-scale sample preparation method and a low sensitivity concerning the doping concentration. Finally, the most promising tools are SCM (Figure 10)[41–44] and SSRP (Figure 11).[45–49] The above-mentioned KFM can also be used for 2-D dopant profiling,[50] but KFM is very sensitive to surface contamination and lacks the required lateral resolution. KFM imaging is a weighted averaging over the sample, due to the measured long-range electrostatic interaction between the tip and sample. Figure 12(a) and (b) shows a KFM measurement in comparison with an SCM measurement on the same sample.

SCM and SSRP are based on SFM operating in constant-force and contact mode (for details see section 2.1). Both instruments need conductive

Table 3 SIA NTRS (selection)

Parameter	Year of first product shipment					
	1999	2001	2003	2006	2009	2012
Dense lines (DRAM half-pitch) (nm)	180	150	130	100	70	50
DRAM (capacity) (Gb)	1	1	4	16	64	256
MPU (transistors/chips) (1×10^6)	21	40	76	200	520	1400
DRAM (chip size) (mm^2)	400	455	560	790	1120	1580
Frequency (MHz)	1200	1400	1600	2000	2500	3000
Max. wiring levels	6–7	7	7	7–8	8–9	9
Min. logic (supply voltage) (V)	1.5–1.8	1.2–1.5	1.2–1.5	0.9–1.2	0.6–0.9	0.5–0.6

DRAM, dynamic random-access memory; MPU, microprocessor unit.

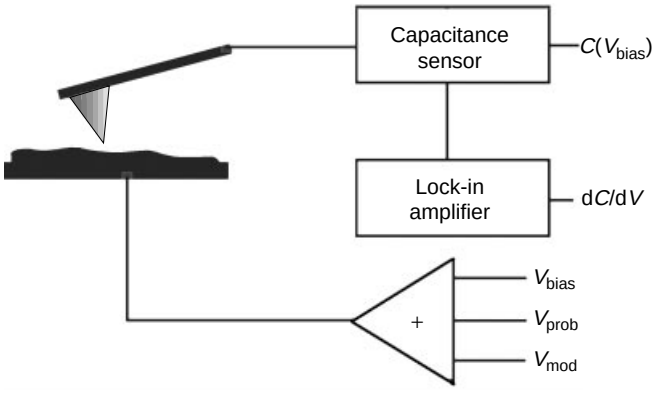

Figure 10 Schematic diagram of an experimental set-up for SCM.

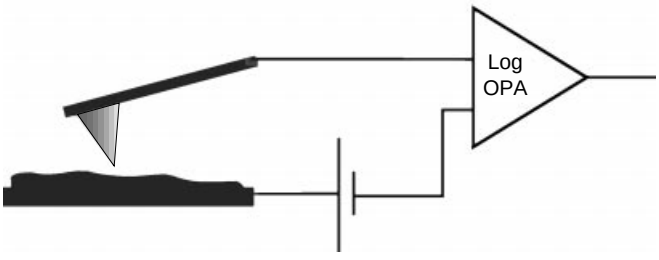

Figure 11 Schematic diagram of an experimental set-up for SSRP. OPA, operational amplifier.

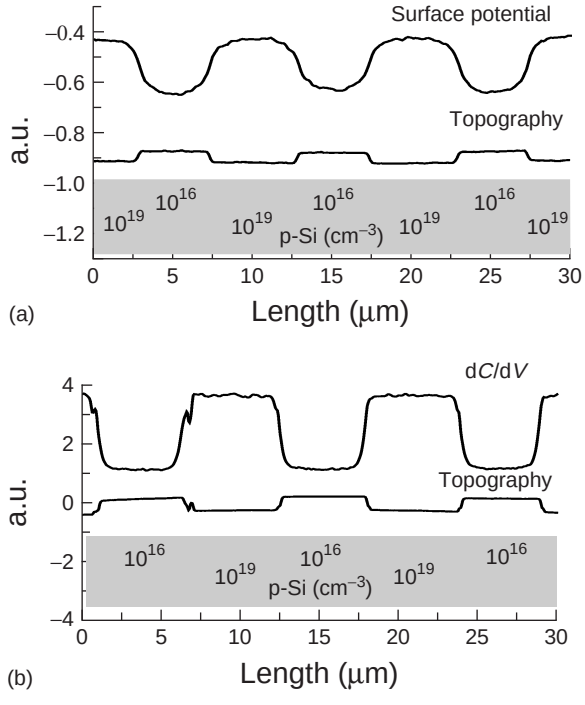

Figure 12 A line scan of the same location of a test structure by (a) KFM and (b) SCM (above, measured signal; below, measured topography; shaded gray, assumed dopant concentration). SCM demonstrates a better lateral and dopant resolution.

cantilevers to connect the apex of the tip with a special detector. The SCM sensor allows the detection of a capacitance of <1 aF (10^{-6} pF). The probing voltage has a frequency of 915 MHz and a typical amplitude of 2 V. To reduce the influence of stray capacitance, we adopt a lock-in technique to produce a signal that is proportional to dC/dV.[51] Like conventional capacitance–voltage measurements, the capacitance measurements by means of an SCM depend on the doping concentration. The probing voltage results in a change of the depletion region underneath the tip. Therefore, the spatial resolution depends on the doping concentration and on the tip shape, or rather the radius of the area of the tip which is in direct contact with the sample. Simulations, which are carried out by a device simulator (DESSIS™), indicate a spatial resolution of <20 nm. By using the right sample preparation method in combination with the right tip, it should be possible to achieve this resolution experimentally. For an SCM measurement, silicon cantilevers with a metallic coating are used. One may utilize any other conductive cantilever, e.g. (doped) silicon or (doped) diamond cantilevers, but the influence of the doped probes, concerning the depletion region into the apex of the probe, must be taken into account.

Unlike SCM, SSRP needs extremely hard probes, such as doped diamond cantilevers. The resistance between the tip and the back contact of the sample is measured by means of a logarithmic amplifier. The potential spatial resolution of <20 nm results from the measurement of the spreading resistance. This is the resistance directly underneath the tip–sample contact and requires an ohmic contact between the tip and sample. By the pressure of the cantilever against the silicon probe, the native oxide is penetrated. At a pressure of 10 GPa, it is possible to obtain an ohmic contact. This necessary high pressure is crucial to SSRP. With the utmost caution concerning the contact force, a doped-diamond cantilever can be used for one or two measurements. Figure 13(a) and (b) gives an SCM and an SSRP image of the same sample.

Both methods need extremely smooth samples, with a roughness of <0.3 nm (RMS). Whereas the cantilever is crucial to SSRP, in SCM it is the requirement for a thin oxide layer. This layer must be of high quality, with a thickness of the order of 3 nm, and it must be produced by a low-temperature process to avoid diffusion of the dopants. Owing to the application of such thin oxides for future MOSFETs, there are many investigations concerning the production process, but they are not yet adequately understood. However, by applying small values of the DC voltage of no more than 1 V, the native oxide can be used for SCM measurements.

TCAD systems and each operator need the values of the local doping concentration, whereas SSRP and SCM measure the local carrier concentration and give

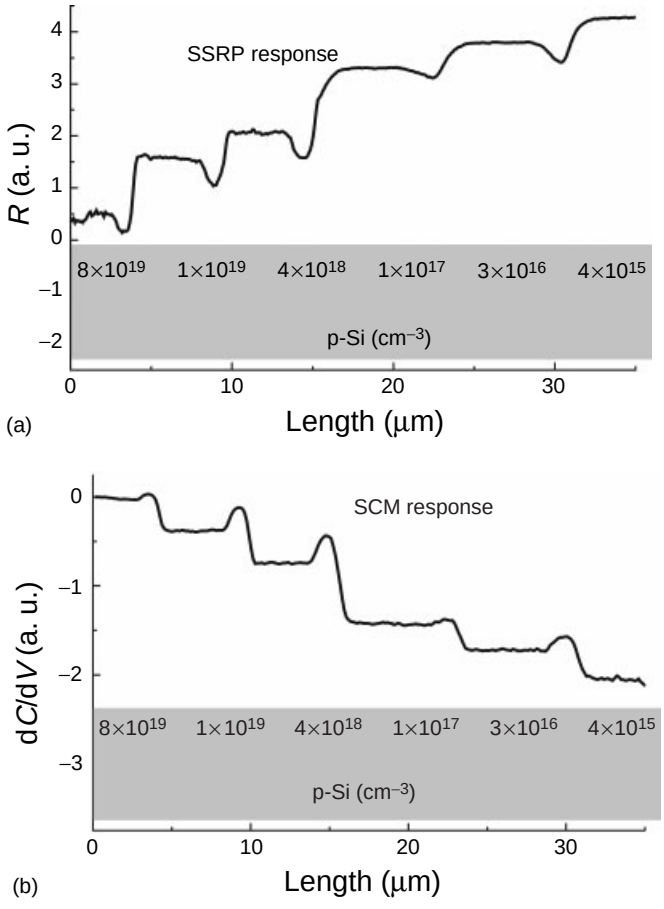

Figure 13 A line scan of the same location of a test structure by (a) SSRP and (b) SCM [above, measured signal; below (shaded gray), assumed dopant concentration]. Both methods show great potential for 2-D dopant profiling.

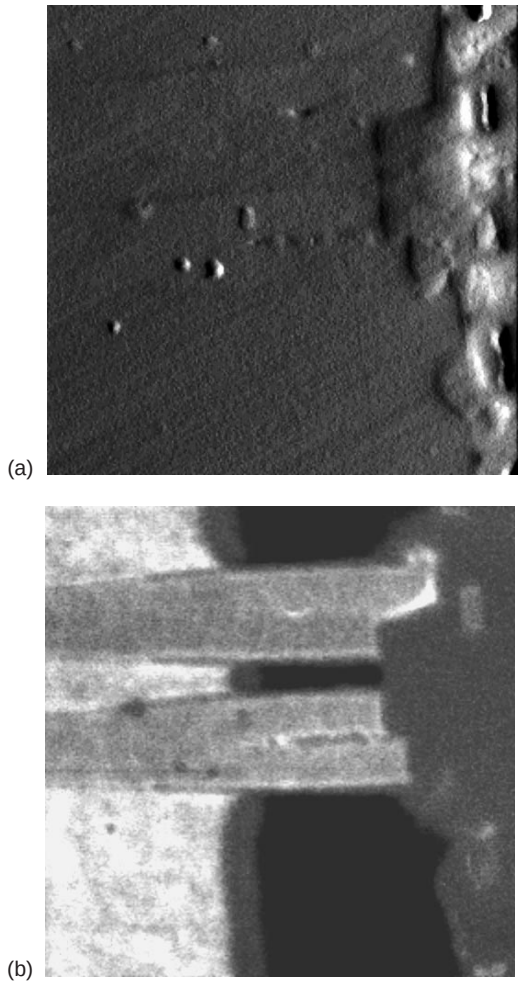

Figure 14 Cross-sectional area [2×2 m; (a) topography and (b) corresponding capacitance signal] of a DRAM.

a signal that is related to the spreading resistance and the capacitance, respectively. Therefore, the measured value must be converted into the doping concentration. Normally (besides a calibration procedure for each tip) the conversion is done by means of a look-up table. At the moment, it works only for one-dimensional (1-D) doping profiles, but some efforts are going into the conversion of 2-D profiles.

Figure 14 shows a SCM image of a cross-section of a DRAM, prepared by a lift-off procedure. Whereas the topography shows the remainder above the transistor structures, the capacitance image locates the different dopant concentrations (n-Si light, p-Si dark).

To summarize, SCM[52] and SSRP can be used to advantage in semiconductor failure analysis today.

2.6 Conducting Atomic Force Microscopy of Oxides

The above-mentioned NTRS predicts a gate oxide thickness of <5 nm for the next generation of ICs. A simple extrapolation of the 250 nm MOSFET design to the 25 nm generation would suggest a gate oxide thickness of 0.5 nm, which means that it consists of two monolayers and the device will not be viable. Therefore, we need new designs and/or new gate oxide materials. These confirm the significance of a method for the investigation of thin oxides with a spatial resolution of <10 nm. Today, oxide thinning due to the local oxidation of silicon (LOCOS) process can lead to a leakage current. The observation of a cross-section, by means of TEM, can show us only part of the whole LOCOS boundary. On the other hand, SPEM suffers from a deficient resolution.

Today, c-AFM[53–56] is the best choice for the characterization of thin oxides. It is based on the SFM (see section 2.1) and the well-known Fowler–Nordheim (FN) tunneling process (see Figure 15). The current through the oxide can be measured with a noise level of <20 fA. A fit of the measured current–voltage curve, using the FN equation, gives the required thickness of the oxide with

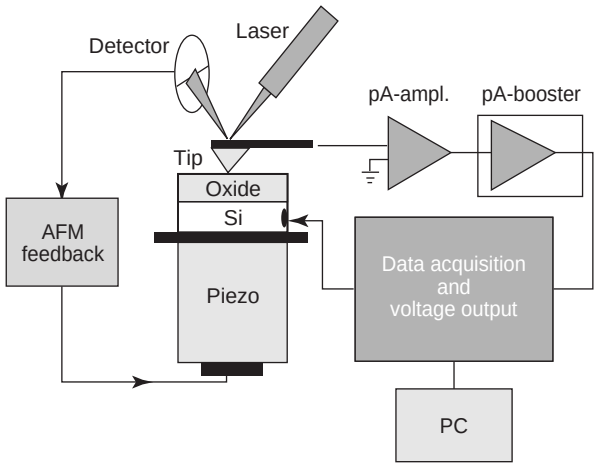

Figure 15 Schematic diagram of an experimental set-up for c-AFM. [Reproduced by permission of the American Institute of Physics from A. Olbrich, B. Ebersberger, C. Boit, 'Conductive Atomic Force Microscopy (c-AFM) for Nanoscale Electrical Characterization of Thin SiO_2', *Appl. Phys. Lett.*, **73**, 3114–3116 (1998).]

a resolution of <0.5 nm. This simple approach requires a homogeneous, high-quality oxide. Interfacial traps and impurities in the oxide can change the current significantly. The sensitivity to the oxide quality determines c-AFM as a useful tool in the failure analysis of semiconductor devices. For oxides, with a thickness of <20 nm, a spatial resolution of <10 nm can be achieved, owing to the exponential nature of the FN current. Crucial to this technique is the tip. It has to be able to conduct, and must be as sharp and as hard as possible, because of the great forces of the high electrical fields between the tip and sample. The best choice is a doped-diamond cantilever, sharpened by a focused-ion beam (FIB) system.

Figure 16 shows an analysis of the tunneling-gate oxide (TOX) of an electrically erasable programmable read-only memory (EEPROM). The measurement was carried out by a bias voltage of 10 V and gives us as a result the thickness of the oxide (of the order of 8.5 nm for TOX and 50 nm for field oxide (FOX)), and the quality of the oxide, which refers to an excellent oxidation process.

Finally, c-AFM gives us significant hints concerning the analysis of thin oxides, because of its high spatial resolution with regard to the thickness and quality of an oxide.

3 PERSPECTIVE AND FUTURE DEVELOPMENTS

All above-mentioned SPM methods reach a spatial resolution of <100 nm and are able to operate in air. The majority of them can also be used in almost any liquid. The

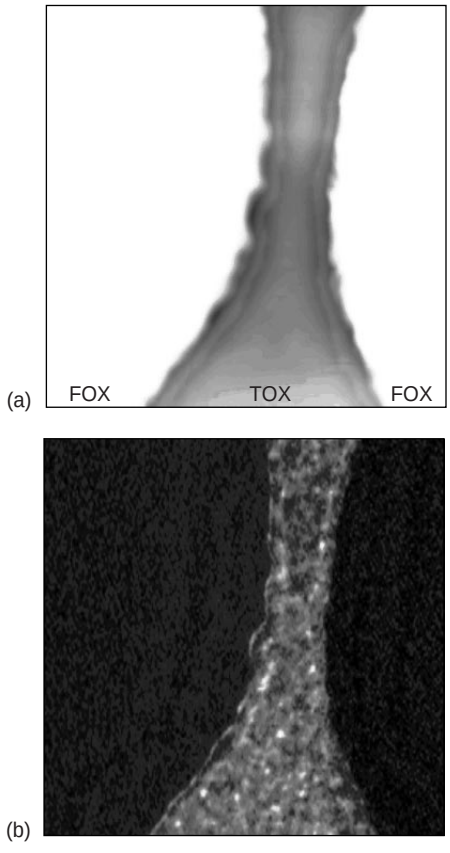

Figure 16 $2 \times 2\,\mu m$ images [(a) topography and (b) corresponding FN current map] of a TOX of an EEPROM.

difficulties concerning the demands for an analytical tool in an industrial environment are commercial availability, user-friendly interface, rapidity, quantitativeness and reproducibility. They are highlighted in section 1.

Further progress must be made in combining different SPM methods, an automatic calibration procedure for an actual probe and the development of specific probes. These points will be explained in detail in the following.

Figure 17 gives as an example an overview of possible SPM methods for operating in failure analysis in the semiconductor industry. Today, all methods are commercially available, but difficult to combine. For example, the combination of SCM and c-AFM seems to be very useful for the analysis of semiconductor devices. Whereas SCM is less sensitive to the oxide thickness and extremely sensitive to the dopant concentration, c-AFM shows the opposite behavior. Therefore, a combination would appear to be a great advantage. An automatic procedure for the calibration of the actual probe would be another worthy goal. These are challenges for all SPM researchers.

Another key issue for SPMs today is the development of special probes. Because of the large number of new innovative features added to the third generation of SPM

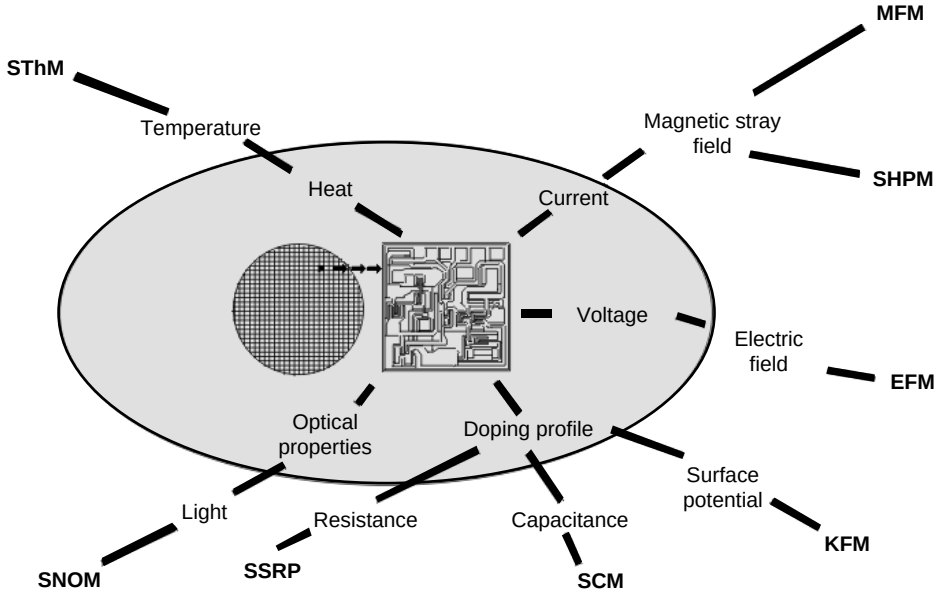

Figure 17 An overview of useful SPMs for the analysis of ICs. SHPM, scanning Hall probe microscope.

methods, following on from the work on STM and SFM, not all special probes are commercially available. This is a challenge for all researchers working on SPM probe development.

Finally, the main goal is a multifunctional SPM system combined with an expert system to give appropriate guidelines concerning the interpretation or quantification of the resulting measurements for each operator in an industrial setting. This requires the cooperation of researchers of SPMs, inventors and users in the universities and industrial companies, and some efforts are now being made in this direction. SPM has developed to become an integral part of the hierarchy of analytical tools in industrial settings, which may be due, in part, to its excellent spatial resolution and its ability to work under ambient conditions.

ACKNOWLEDGMENTS

It is a pleasure to thank G. Fiege (SThM), A. Kelch (SFM) and A. Olbrich (c-AFM) for supplying SPM images. Moreover, I acknowledge the SXM group at the University of Hamburg, who provided assistance and support.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
c-AFM	Conducting Atomic Force Microscopy/Microscope
DC	Direct Current
DRAM	Dynamic Random-access Memory
EEPROM	Electrically Erasable Programmable Read-only Memory
EFM	Electrostatic Force Microscopy/Microscope
EM	Emission Microscopy/Microscope
FIB	Focused-ion Beam
FN	Fowler–Nordheim
FOX	Field Oxide
IC	Integrated Circuit
KFM	Kelvin Force Microscopy/Microscope
LCT	Liquid-crystal Thermography
LOCOS	Local Oxidation of Silicon
MFM	Magnetic Force Microscopy/Microscope
MOSFET	Metal-oxide Semiconductor Field-effect Transistor
MPU	Microprocessor Unit
NTRS	National Technology Roadmap for Semiconductors
OPA	Operational Amplifier
SAM	Self-assembled Monolayer
SCM	Scanning Capacitance Microscopy/Microscope
SEM	Scanning Electron Microscopy/Microscope
SFM	Scanning Force Microscopy/Microscope
SHPM	Scanning Hall Probe Microscope
SIA	Semiconductor Industry Association
SIMS	Secondary Ion Mass Spectrometry
SNOM	Scanning Near-field Optical Microscopy/Microscope

SPEM	Spectroscopic Photon Emission Microscopy/Microscope
SPM	Scanning Probe Microscopy/Microscope
SQUID	Superconducting Quantum Interference Device
SSRP	Scanning Spreading Resistance Profiling
SThM	Scanning Thermal Microscopy/Microscope
STM	Scanning Tunneling Microscopy/Microscope
TCAD	Technology Computer-aided Design
TEM	Transmission Electron Microscopy/Microscope
TOX	Tunneling-gate Oxide
UHV	Ultrahigh Vacuum
ULSI	Ultralarge-scale Integrated
VLSI	Very Large-scale Integrated
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Atomic Force Microscopy in Analysis of Polymers

Surfaces **(Volume 10)**
Surfaces: Introduction • Electron Microscopy and Scanning Microanalysis • Proximal Probe Techniques • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

REFERENCES

1. R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, 1994.
2. R. Wiesendanger, H.-J. Güntherodt (eds.), *Scanning Tunneling Microscopy II*, Springer Series in Surface Sciences, 1st edition, Springer, Berlin, Vol. 28, 1992.
3. R. Wiesendanger, H.-J. Güntherodt (eds.), *Scanning Tunneling Microscopy III*, Springer Series in Surface Sciences, 1st edition, Springer, Berlin, Vol. 29, 1993.
4. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, 'Surface Studies by Scanning Tunneling Microscopy', *Phys. Rev. Lett.*, **49**, 57–61 (1982).
5. H.-J. Güntherodt, R. Wiesendanger (eds.), *Scanning Tunneling Microscopy I*, Springer Series in Surface Sciences, 2nd edition, Springer, Berlin, Vol. 20, 1994.
6. C.J. Chen, *Introduction to Scanning Tunneling Microscopy*, Oxford University Press, Oxford, 1993.
7. C. Bai, *Scanning Tunneling Microscopy and its Application*, Springer Series in Surface Sciences, 1st edition, Springer, Berlin, Vol. 32, 1995.
8. R.J. Colton, A. Engel, J.E. Frommer, H.E. Gaub, A.A. Gewirth, R. Guckenberger, J. Rabe, W.M. Heckl, B. Parkinson (eds.), *Procedures in Scanning Probe Microscopies*, Wiley, Chichester, 1998.
9. B.C. Stipe, M.A. Rezaei, W. Ho, 'Inducing and Viewing the Rotational Motion of a Single Molecule', *Science*, **279**, 1907–1909 (1998).
10. J.G.A. Dubois, J.W. Gerritsen, S.E. Shafranjuk, E.J.G. Boon, G. Schmid, H. van Kempen, 'Coulomb Staircases and Quantum Size Effects in Tunneling Spectroscopy on Ligand-stabilized Metal Clusters', *Europhys. Lett.*, **33**(4), 279–284 (1996).
11. H. Ahmed, 'Single Electron Electronics: Challenge for Nanofabrication', *J. Vac. Sci. Technol. B*, **15**(6), 2101–2108 (1997).
12. J. Sone, J. Fujita, Y. Ochiai, S. Manako, S. Matsui, E. Nomura, T. Baba, H. Kawaura, T. Sakamoto, C.D. Chen, Y. Nakamura, J.S. Tsai, 'Nanofabrication Toward Sub-10 nm and its Application to Novel Nanodevices', *Nanotechnology*, **10**, 135–141 (1999).
13. C. Hu, 'Silicon Nanoelectronics for the 21st Century', *Nanotechnology*, **10**, 113–116 (1999).
14. R. Subrahmanyan, 'Methods for the Measurement of Two-dimensional Doping Profiles', *J. Vac. Sci. Technol. B*, **10**(1), 358–368 (1992).
15. T. Sulchek, S.C. Minne, J.D. Adamis, D.A. Fletcher, A. Atalar, C.F. Quate, D.M. Adderton, 'Dual Integrated Actuators for Extended Range High Speed Atomic Force Microscopy', *Appl. Phys. Lett.*, **75**, 1637–1639 (1999).
16. G. Binnig, M. Despont, U. Drechsler, W. Haeberle, M. Lutwyche, P. Vettiger, H.J. Mamin, B.W. Chui, T.W. Kenny, 'Ultrahigh-density Atomic Force Microscopy Data Storage with Erase Capability', *Appl. Phys. Lett.*, **74**, 1329–1331 (1999).
17. T. Ando, Y. Arakawa, K. Furuya, S. Komiyama, N. Nakashima (eds.), *Mesoscopic Physics and Electronics*, Nano-Science and Technology, 1st edition, Springer, Berlin, 1998.
18. J. Tersoff, D.R. Hamann, 'Theory and Application of the Scanning Tunneling Microscope', *Phys. Rev. Lett.*, **50**, 1998–2001 (1983).
19. S. Ciraci, A. Baratoff, I.P. Batra, 'Tip–Sample Interaction Effects in Scanning-tunneling and Atomic-force Microscopy', *Phys. Rev. B*, **41**, 2763–2775 (1990).
20. D.P.E. Smith, G. Binnig, C.F. Quate, 'Atomic Point-contact Imaging', *Appl. Phys. Lett.*, **49**, 1166–1168 (1986).
21. D. Sarid, 'Scanning Force Microscopy', 1st edition, Oxford University Press, New York, 1991.
22. U. Schwarz, 'Scanning Force Microscopy', in *Handbook of Microscopy. Methods II*, eds. S. Amelinckx, D. van Dyck, J. van Landhuyt, G. van Tendeloo, VCH, Weinheim, 827–844, 1997.

23. Y. Martin, H.K. Wickramasinghe, 'Magnetic Imaging by Force Microscopy with 1000 Å Resolution', *Appl. Phys. Lett.*, **50**, 1455–1457 (1987).
24. A. Wadas, 'Magnetic Force Microscopy', in *Handbook of Microscopy. Methods II*, eds. S. Amelinckx, D. van Dyck, J. van Landhuyt, G. van Tendeloo, VCH, Weinheim, 845–853, 1997.
25. U. Hartmann, 'Magnetic Force Microscopy', *Annu. Rev. Mater. Sci.*, **29**, 53–87 (1999).
26. R. Proksch, P. Neilson, S. Austvold, J.J. Schmidt, 'Measuring the Gigahertz Response of Recording Heads with the Magnetic Force Microscope', *Appl. Phys. Lett.*, **74**, 1308–1310 (1999).
27. T. Göddenhenrich, H. Lemke, M. Mück, U. Hartmann, C. Heiden, 'Probe Calibration in Magnetic Force Microscopy', *Appl. Phys. Lett.*, **57**, 2612–2614 (1990).
28. K.L. Babcock, V.B. Elings, J. Shi, D.D. Awschalom, M. Dugas, 'Field-dependence of Microscopic Probes in Magnetic Force Microscopy', *Appl. Phys. Lett.*, **69**, 705–707 (1996).
29. C.C. Williams, H.K. Wickramasinghe, 'Scanning Thermal Profiler', *Appl. Phys. Lett.*, **49**, 1587–1589 (1986).
30. M. Nonnenmacher, H.K. Wickramasinghe, 'Scanning Probe Microscopy of Thermal Conductivity and Subsurface Properties', *Appl. Phys. Lett.*, **61**, 168–170 (1992).
31. A. Majumdar, 'Scanning Thermal Microscopy', *Annu. Rev. Mater. Sci.*, **29**, 505–585 (1999).
32. L.J. Balk, M. Maywald, R.J. Pylkki, 'Nanoscopic Detection of the Thermal Conductivity of Compound Semiconductor Materials by Enhanced Scanning Thermal Microscopy', *Inst. Phys. Conf. Ser.*, **146**, 655–658 (1995).
33. G.B.M. Fiege, V. Feige, J.C.H. Phang, M. Maywald, S. Görlich, L.J. Balk, 'Failure Analysis of Integrated Devices by Scanning Thermal Microscopy (SThM)', in *Proceedings of the ESREF 98, Microelectron. Reliab.*, Elsevier, Amsterdam, 957–961, Vol. 38, 1998.
34. V.M. Asnin, F.H. Pollak, J. Ramer, M. Schurman, I. Ferguson, 'High Spatial Resolution Thermal Conductivity of Lateral Epitaxial Overgrown GaN/Sapphire (0001) Using a Scanning Thermal Microscope', *Appl. Phys. Lett.*, **75**, 1240–1242 (1999).
35. M. Nonnenmacher, M. O'Boyle, H.K. Wickramasinghe, 'Kelvin Probe Force Microscopy', *Appl. Phys. Lett.*, **58**, 2921–2923 (1991).
36. M. Fujihira, 'Kelvin Probe Force Microscopy of Molecular Scales', *Annu. Rev. Mater. Sci.*, **29**, 353–380 (1999).
37. H.O. Jacobs, P. Leuchtmann, O.J. Homan, A. Stemmer, 'Resolution and Contrast in Kelvin Probe Force Microscopy', *J. Appl. Phys.*, **84**, 1168–1173 (1998).
38. Ch. Sommerhalter, Th.W. Matthes, Th. Glatzel, A. Jäger-Waldau, M.Ch. Lux-Steiner, 'High-sensitivity Quantitative Kelvin Probe Microscopy by Noncontact Ultra-high-vacuum Atomic Force Microscopy', *Appl. Phys. Lett.*, **75**, 286–288 (1999).
39. H. Nanjo, T. Inoue, H. Yokoyama, 'The Relation Between Corrosion and Surface Potential Measured with the Scanning Maxwell Stress Microscope', *Nanotechnology*, **9**, 316–320 (1998).
40. The NTRS, http://www.semichips.org/, SIA, San Jose, CA, 1998.
41. J.R. Matey, J. Blanc, 'Scanning Capacitance Microscopy', *J. Appl. Phys.*, **57**, 1437–1444 (1985).
42. C.C. Williams, W.P. Hough, S.A. Rishton, 'Scanning Capacitance Microscopy on a 25 nm Scale', *Appl. Phys. Lett.*, **55**, 203–205 (1989).
43. R.C. Barrett, C.F. Quate, 'Charge Storage in a Nitride–Oxide–Silicon Medium by Scanning Capacitance Microscopy', *J. Appl. Phys.*, **70**, 2725–2733 (1991).
44. C.C. Williams, 'Two-dimensional Dopant Profiling by Scanning Capacitance Microscopy', *Annu. Rev. Mater. Sci.*, **29**, 471–504 (1999).
45. C. Shafai, D.J. Thomson, M. Simard-Normandin, G. Mattiussi, P.J. Scanlon, 'Delineation of Semiconductor Doping by Scanning Resistance Microscopy', *Appl. Phys. Lett.*, **64**, 342–344 (1994).
46. J.N. Nxumalo, D.T. Shimizu, D.J. Thomson, 'Cross-section Imaging of Semiconductor Device Structures by Scanning Resistance Microscopy', *J. Vac. Sci. Technol. B*, **14**, 386–389 (1996).
47. T. Clarysse, P. de Wolf, H. Bender, W. Vandervorst, 'Recent Insight into the Physical Modeling of the Spreading Resistance Point Contact', *J. Vac. Sci. Technol. B*, **14**, 358–368 (1996).
48. P. de Wolf, T. Clarysse, W. Vandervorst, L. Hellemans, Ph. Niedermann, W. Hänni, 'Cross-sectional Nanospreading Resistance Profiling', *J. Vac. Sci. Technol. B*, **16**, 355–361 (1998).
49. P. de Wolf, T. Clarysse, W. Vandervorst, L. Hellemans, 'Low Weight Spreading Resistance Profiling of Ultrashallow Dopant Profiles', *J. Vac. Sci. Technol. B*, **16**, 401–405 (1998).
50. T. Hochwitz, A.K. Henning, C. Levey, D. Daghlian, J. Slinkman, J. Never, P. Kaszuba, R. Gluck, R. Wells, J. Pekarik, R. Finch, 'Imaging Integrated Circuit Dopant Profiles with the Force-based Scanning Kelvin Probe Microscope', *J. Vac. Sci. Technol. B*, **14**, 440–446 (1996).
51. A. Born, R. Wiesendanger, 'Scanning Capacitance Microscope as a Tool for the Characterization of Integrated Circuits', *Appl. Phys. A*, **66**, S421–S426 (1998).
52. A. Born, A. Olbrich, M. Maywald, R. Wiesendanger, 'Analysis of Electrical Breakdown Failures by Means of SFM-based Methods', *Appl. Phys. A*, **66**, S1063–S1065 (1998).
53. S.J. O'Shea, R.M. Atta, M.P. Murrell, M.E. Welland, 'Conductive Atomic Force Microscopy Study of Silicon Dioxide Breakdown', *J. Vac. Sci. Technol. B*, **13**, 1945–1952 (1995).
54. T.G. Ruskell, R.K. Workman, D. Chen, D. Sarid, S. Dahl, S. Gilbert, 'High Resolution Fowler-Nordheim Field

Emission Maps of Thin Silicon Oxide Layers', *Appl. Phys. Lett.*, **68**, 93–95 (1996).

55. A. Olbrich, B. Ebersberger, C. Boit, 'Conductive Atomic Force Microscopy (c-AFM) for Nanoscale Electrical Characterization of Thin SiO_2', *Appl. Phys. Lett.*, **73**, 3114–3116 (1998).

56. A. Olbrich, B. Ebersberger, C. Boit, Ph. Niedermann, W. Haemm, J. Vancea, H. Hoffmann, 'High Aspect Ratio All Diamond Tips Formed by Focused Ion Beam for Conducting Atomic Force Microscopy', *J. Vac. Sci. Technol. B*, **17**, 1570–1574 (1999).

Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

C.V. Dharmadhikari
Department of Physics, University of Pune, Pune, India

1 Introduction 1
2 History 2
3 Scanning Tunneling Microscopy/ Spectroscopy: Basic Principles 2
3.1 Phenomenon of Tunneling 2
3.2 Theory and Operation of Scanning Tunneling Microscope 2
3.3 Resolution and Magnification 3
3.4 Imaging Modes 3
3.5 Fundamentals of Tunneling Spectroscopy 4
4 Instrumentation and Methodology 5
4.1 Scanning Tunneling Microscopy 5
4.2 Control Electronics 5
4.3 Data Acquisition, Image Processing and Analysis 5
4.4 Preparation and Characterization of Tip and Sample 5
4.5 Vibration and Acoustic Isolation 6
4.6 Scanning Tunneling Spectroscopy Experimentation 7
5 Illustrative Applications 7
5.1 Structure of Crystal Surfaces 7
6 Comparison with Other Methods 14
6.1 Scanning Tunneling Microscopy Compared with High-resolution Electron Microscopies 14
6.2 Scanning Tunneling Spectroscopy Compared with Other Electron Spectroscopies 14
7 Scanning Probe Microscopy: Recent Trends and Future Prospects 14
Acknowledgments 15
Abbreviations and Acronyms 15
Related Articles 15
References 15

Scanning tunneling microscopy (STM), scanning tunneling spectroscopy (STS) and related techniques with their unique capability to resolve topological and electronic structures at atomic level have revolutionized the power of experimental techniques in high resolution imaging of surfaces. In STM, a sharp conducting tip is mechanically scanned over a conducting or semi-conducting sample at the distance of few angstroms from the surface. A voltage bias applied between the tip and the sample leads to flow of tunnel current, which decreases exponentially as the gap increases. By keeping the tunnel current constant with an electronic feedback controller the tip is maintained at a fixed distance during the scanning by means of a 3-D scanner made of a piezoelectric transducer. The trajectory of the tip then traces out a profile of the surface, including the bumps due to individual atoms. In the context of atomic resolution imaging capability, there are only a few techniques such as scanning transmission electron microscopy (STEM), transmission electron microscopy (TEM) and field ion microscopy (FIM) which perform as well as STM, but they do so under extremely special circumstances. The traditional spectroscopic techniques that can be compared with STS are those of conventional (elastic) electron tunneling spectroscopy (ETS) and inelastic electron tunneling spectroscopy (IETS), X-ray photoelectron spectroscopy (XPS), ultraviolet photoemission spectroscopy (UPS) and inverse photoemission spectroscopy (IPS). But all these techniques detect and average the data from a relatively large area, a few microns to a few millimeters across. The STM/STS, in contrast, can take spectra localized to as small an area as an individual atom. The most important characteristic that distinguishes STM/STS from other techniques is ability to operate in a variety of different environments such as air, reactive gases, liquids, electrolytes and biological fluids, leading to its applications not only in basic research in surface science but also in such diverse fields as lithography, electrochemistry, biology and medicine. An inherent limitation of STM is that it always operates at high resolution and, therefore, works better for atomically flat surfaces.

1 INTRODUCTION

Surface studies are of intrinsic interest because most solids interact with the external environment through their surfaces; they are also important in a large number of phenomena of great technological importance.[1] Traditionally, surfaces have been studied by conventional techniques such as TEM,[2] scanning electron microscopy (SEM),[3] low energy electron diffraction (LEED),[4] reflection electron microscopy (REM),[5] X-ray diffrac-

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

tion and scattering experiments,[6] and field electron microscopy (FEM)[7] and field ion microscopy (FIM).[8] STM, with its unique capability to resolve topographic and electronic structures at atomic level in vacuum, atmospheric and fluid environments, has removed the limitations of some earlier techniques and revolutionized the experimental techniques not only in basic research in surface science but also in applied research in such diverse fields as lithography, electrochemistry, metrology, tribology, biology and medicine.

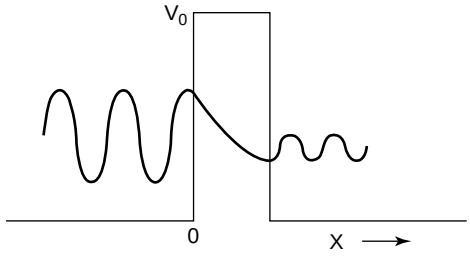

Figure 1 Schematic diagram of quantum mechanical tunneling through a rectangular barrier.

2 HISTORY

The first glimpse of the STM concept appeared in the 1970s in the form of an instrument called ''topographiner'' based on the principle of the scanning field emission microscope.[9] Although it was recognized that atomic resolution could be achieved by improving the tip, precision and stability of the instrument, it was not experimentally realized until 1982, when a group at IBM, Zürich, observed that the tunnel conductivity plots on surfaces involved sharp repeated steps and faint bumps of atomic dimensions.[10–12] Although the origin of STM can be traced back to the early 1980s, it came into the limelight only after the announcement of the Nobel Prize to its inventors, G. Binnig and H. Rohrer in 1986, which led to an outburst of worldwide interest. In recent years, STM activity has grown beyond the original expectations, leading to a totally new branch called scanning probe microscopy (SPM), which has established itself as an indispensable tool in the realm of materials science and microelectronics.[13–21] Several historical overviews can be cited for an absorbing account of the invention and development of STM and related techniques.[22–27]

3 SCANNING TUNNELING MICROSCOPY/SPECTROSCOPY: BASIC PRINCIPLES

3.1 Phenomenon of Tunneling

According to quantum mechanics, a subatomic particle can pass through a spatial region in which the particle's kinetic energy is less than its potential energy. The phenomenon is known as tunneling.[28] Figure 1 shows an energy diagram for a typical tunneling process. The observation of the tunneling effect in a variety of systems has offered direct evidence of quantum mechanics in action. Some of the landmarks in the science of tunneling are:[29] observation of field emission from metal, ionization of hydrogen atoms by electron tunneling, explanation of field emission, alpha particle decay process in heavy atoms, the development of FEM[7] and FIM[8] tunneling in the degenerate p–n junction, experimental verification of Josephson effects (tunneling Cooper pairs),[30] IETS,[31,32] the development of the topographiner[9] and STM.[22–27]

3.2 Theory and Operation of Scanning Tunneling Microscope

As the name suggests, STM makes use of the phenomenon of tunneling discussed in the preceding section; according to this, if two metals are brought close enough, i.e. they are separated by only a few angstroms, the wave functions of electronic quantum states of the two surfaces overlap. The overlap allows electrons to jump across the vacuum gap between the tip and the sample, leading to a tunnel current when the voltage is applied between them. The tunneling current expression for low bias based on the free electron model[33] is as stated in Equation (1) below:

$$J = \frac{3}{4\pi}\frac{e^2}{h}V\frac{k_0}{s}\exp(-2k_0 s) \qquad (1)$$

where the V is applied voltage, s is the barrier width and k_0 is the decay constant of the wave function in the barrier ($\hbar^2 k_0{}^2/2m = \phi$, the work function). This expression indicates that $I \propto V/s\exp(-A\phi^{1/2}s)$, if V is in volts, s in nm, the average work function ϕ in eV, then $A = 10.25\,(\text{nm})^{-1}\,\text{eV}^{-1/2}$. The exponential dependence was measured by Young et al.[34] and Teague,[35] and later by Binnig et al.[10–12] The main feature of this equation is its extreme sensitivity of tunnel current to changes in s. A change of electrode spacing of 0.1 nm produces approximately an order of magnitude change in current for a barrier height of 5 eV. In STM (Figure 2), a metal wire (called a tip), with a finely ground or electrochemically etched point, is positioned near the sample surface to be investigated so that a measurable tunnel current (nA to pA) flows for a reasonable bias voltage (a few millivolts to several volts). The tip is then mechanically scanned across the sample surface by

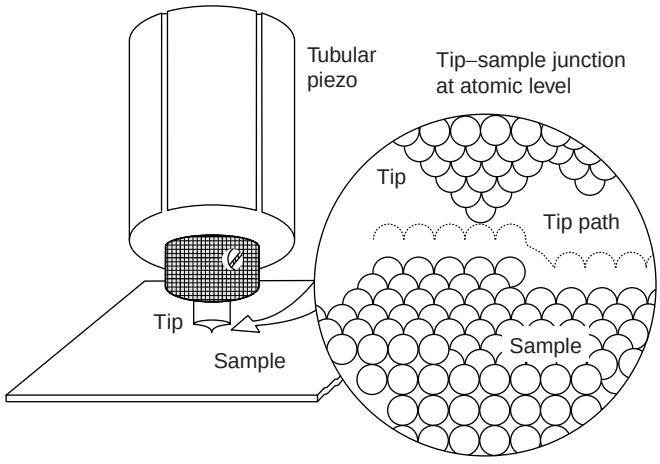

Figure 2 Principle of STM.

means of a 3-D scanner made of piezoelectric transducers. If the surface is rough even at atomic level, the tunnel current will change with the tip location on the surface. The voltage at z-piezo required to maintain the current constant as a function of x, y coordinates derived from voltages at x- and y-piezos yields a topographic profile of the surface.

3.3 Resolution and Magnification

There are no special resolution criteria in STM. The vertical resolution is governed by the stability of the tunnel junction, since height details of the sample surface smaller than the tip to sample vibration amplitudes are usually masked. The lateral resolution is determined by the width of the tunnel current. The approximate expression for tunnel current $I \propto V \exp(-2k_0 s)$ can be used to get an order of magnitude estimate of the vertical and horizontal resolution. In Figure 3, at a given lateral position $\delta/2$ relative to the center of the tip, the height of the corresponding point on the tip is $x = s + \delta^2/8R$, assuming a curved tip of radius of curvature R and distance of closest approach s. Thus the current has a Gaussian profile, with root-mean-square width $2(R/k_0)^{1/2}$. Since $2k_0 \sim 1\text{Å}^{-1}$ ($10\,\text{nm}^{-1}$), the lateral resolution of $L = 50$ Å (5 nm) can be estimated for a typical smoothed field emission tip of ca. 1000 Å (100 nm). The estimation of resolution can be improved by preparing ultra-sharp tips by a factor of 2, but is still smaller than atomic resolution. It turns out that most of the chemically etched or mechanically ground tips have asperities of near atomic dimensions and because of the strong dependence of tunnel current on distance only asperities closest to the sample will contribute substantially to the tunnel current. It is clear that atomic resolution would require atomically sharp asperities. More elaborate estimation of lateral

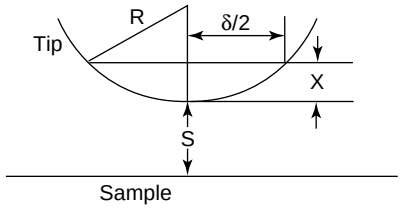

Figure 3 Illustration of role of radius of curvature of the tip in the lateral resolution of STM.

resolution based on quantum mechanical calculation yields $L_{\text{eff}} = 1.66[(s + R)/k_0]^{1/2}$. The actual resolution can be smaller than this if the apex of the tip has an atom with a directed orbital pointing outwards. It appears that[36,37] p-orbital type of element of elemental semiconductors or d-orbital from d-band metals such as tungsten, etc. are favorable for atomic resolution on metal surfaces.

3.4 Imaging Modes

There are two main modes of STM-imaging: constant height mode and constant current mode.

3.4.1 Constant Height Mode

A straightforward application of Equation (1) for imaging would involve scanning the tip at constant height for a fixed bias across the tunnel junction and monitoring the tunnel current either on the oscilloscope or via a video-processor. The method is therefore called current imaging. This has an obvious advantage that relatively high speeds (1 kHz) can be used, enabling real-time video display of the surface.[38] The disadvantage is that the method can only be used for atomically flat surfaces, where the surface roughness is less than the tunneling gap.

3.4.2 Constant Current Mode

This is the most widely used method for acquiring STM images. As in the previous method, the tip is scanned across the surface at constant voltage. However, the time-constant of the servo-system is selected to maintain the tunnel current, hence the height of the tip above the surface is kept constant by changing the voltage to the z-piezo. The feedback voltage for z-piezo, which is proportional to the z-displacement of the tip, is recorded as a function of x, y coordinates of the tip and therefore generates a topographical image of the surface. The constant current topography maps the local density of states (LDOS) at the position of the tip.[39,40] The scan rate in this method is typically low (<10 Hz) and is limited by the response time of the feedback loop and the inertia of the motion.

According to Equation (1) the current is a measure of topography (distance-dependence) only if work function ϕ and pre-factor are constant over the scanned area. If both ϕ and s vary, the STM-image would involve convolution of topographic and electronic information. In such cases it is possible to record barrier image by plotting $\partial \ln I/\partial s$ instead of I. The comparison of two images facilitates extraction of topographic information.[22–27]

3.5 Fundamentals of Tunneling Spectroscopy

The tunneling expression used for understanding STM operation in the previous sections was derived by assuming low bias voltage, implying that the electron current was coming essentially from a small band of electron states near Fermi-energy.

If the applied voltages are large compared with the Fermi-energy, the resulting image may depend dramatically on the bias voltage, and the approximate tunnel current expression for nonzero voltage V between tip and sample can be written as[41]

$$I \propto \int_{o}^{eV} \rho_1(E)\rho_2(E - eV)T(E, V)\,\mathrm{d}E \qquad (2)$$

where $\rho_1(E)$ and $\rho_2(E - eV)$ are the surface state densities of sample and tip respectively at energies relative to their Fermi-levels and $T(E, V)$ is the transmission term for tunneling electrons.

In WKB approximation, the maximum transmission has the form shown in Equation (3):

$$T_{\max}(E, V) = \exp\left\{-2s\left[\frac{2m}{\hbar^2}\bar{\phi} - \frac{|e\mathrm{V}|}{2}\right]^{1/2}\right\} \qquad (3)$$

For a small bias this reverts to Equation (1). It can be seen that the tunnel current varies by about an order of magnitude for a unit change in bias voltage. As the dominant contribution of tunneling current comes from states near the Fermi-level of a negatively biased electrode, tunneling from tip to sample mainly probes empty states of the sample, with negligible influence from the occupied states of the tip, whereas the tunneling from sample to tip is much more sensitive to the electronic structure of empty states of tip, which often prevents detailed spectroscopic STM studies of the occupied states of the sample.

There are three modes in which to perform tunneling spectroscopy: constant current mode, constant separation mode and current imaging tunneling spectroscopy (CITS).

3.5.1 Constant Current Mode

The simplest way of obtaining spectroscopic information is to compare constant current topological images acquired in the conventional constant current mode at different bias voltages. If the electrons tunnel from sample to tip (negative sample bias), the occupied (bonding) states of the sample are imaged. If they tunnel from tip to sample (positive sample bias), the empty (antibonding) states appear. To obtain qualitative information, a high-frequency sinusoidal modulation voltage can be superimposed on the constant bias voltage between tip and sample. The modulation frequency is chosen to be higher than the cut-off frequency of the feedback loop, which maintains a constant tunneling current. By recording the in-phase component of modulated tunnel current using a lock-in amplifier, a spatially resolved spectroscopy signal $\mathrm{d}I/\mathrm{d}V$ can be obtained simultaneously with the topographic image.

3.5.2 Constant Separation Mode

The constant current mode based on the modulation technique experiences problems in cases where no current is flowing i.e. at zero bias voltage or in band gap and in cases where the conductivity diverges (metals) as the bias voltage approaches zero. These limitations are overcome in constant separation modes where the tip is temporarily frozen above the fixed height by breaking the feedback circuit for certain time interval using a sample-and-hold circuit while the I-V curves are recorded. The I-V curves are repeated several times at each point and the signal averaged to reduce the noise. The method is particularly useful for probing bandgap states in semiconductors because the small currents which lead to crashing of the tip into the sample are avoided by breaking the feedback action. The plot of normalized differential conductance $(\mathrm{d}I/\mathrm{d}V)/(I/V)$ versus the bias voltage mimics the density of states spectrum. The agreement with theoretical density of states can be expected only for the positions of the peak as the intensities may differ significantly.

3.5.3 Current Imaging Tunneling Spectroscopy

The combination of scanning ability of STM with the spectroscopic capability, i.e. measurement of local I-V curves leads to the third mode of spectroscopy, CITS. In this method, the feedback control circuit used for constant current STM is gated using a sample-and-hold circuit so that it is active for only 20 to 30% of the time. When the feedback is active, a constant voltage that is applied to the sample and feedback control adjusts the tip height to maintain the tunnel current at a constant level. When inactive, the position of the tip is held stationary and the tunnel current is measured at various different bias voltages. By acquiring the I-V curves rapidly as compared to the scan speed of the tip, both the sample "topography" and spatially resolved tunneling characteristics can be

obtained either at each location or at only a few selected locations.

The spatial resolution of tunneling spectroscopy is similar to STM and is determined by the size of the probe. The energy resolution is determined by the range of electron energies that can contribute to the tunnel current. Inelastic tunneling spectroscopy, which involves interaction of tunneling electrons with vibrational modes of molecules adsorbed on a surface can be used to obtain a vibrational spectrum of an individual molecule or molecular functional group. At low voltage, the effective energy resolution is $\approx 5.4\,kT$.[42]

4 INSTRUMENTATION AND METHODOLOGY

The important components of STM are: STM head, control electronics, tip and sample, vibration and acoustic isolation and data acquisition system.

4.1 Scanning Tunneling Microscopy

The STM head consists of the sample, tip and a 3-D micromanipulator which controls the precision positioning of the tip. The tip must be brought within few angstroms of the sample surface. This is generally done in two steps. The coarse positioning involves bringing the sample towards the probe from a distance of several millimeters to a few hundred angstroms. The fine positioning involves moving the probe from a few hundreds of angstroms to the tunneling distances. Many types of tip approaches have been used, mostly based on piezo fabrications in the form of a ''louse'', inertial sliders which move the sample towards the tip by a slip-stick mechanism or by using the bending motion of the piezoelectric tubes.[43–46] A much simpler approach involves use of micrometers or fine pitched screws, whose motion is demagnified by use of differential springs or levers. The fine positioning approach mainly involves tripod structures made up of three piezoelectric bars, whose motion can be independently controlled by applying voltages to individual electrodes or a single tube scanner with segmented electrodes.[46] A schematic diagram of a typical STM head is shown in Figure 4.

4.2 Control Electronics

The current from the tunnel junction is converted into voltage by a preamplifier firmly mounted on the STM head to minimize electrical interference. Since the tunnel current is exponentially dependent on distance, the signal is linearized by a logarithmic amplifier to improve the dynamic range. This voltage is given to the input of the feedback controller, which consists of a difference amplifier, an analog integrator with variable time constant, a proportional amplifier, a low pass filter and a voltage driver.[47] Figure 5 shows a block diagram of STM instrumentation.

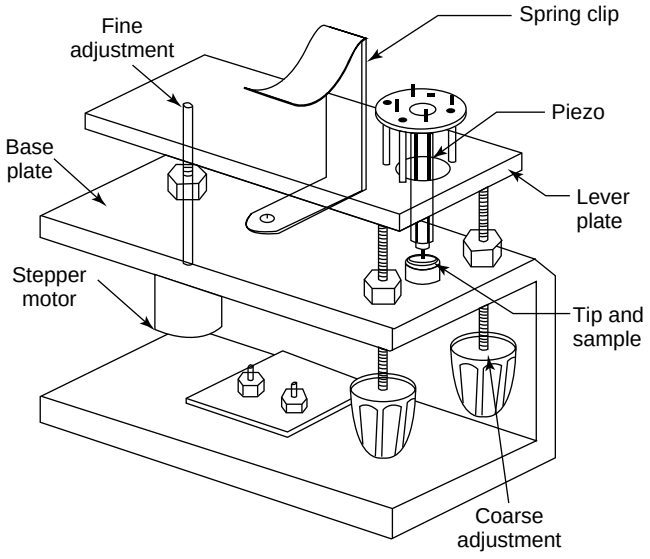

Figure 4 Schematic diagram of STM head.

4.3 Data Acquisition, Image Processing and Analysis

The electrical signals from the inputs of x, y and z (error signal from feedback) piezos, the tip voltage and tunneling current are stored in an image file on the PC or workstation.[48,49] Graphical routines are used to display the data and for image processing. The display usually involves 3-D line plots or a top view gray or color image (Figure 6). Several image-processing techniques can be used to enhance the quality of the data. For instance, the tilt of the sample can be compensated by a background plane subtraction. The image can be corrected for thermal drift and hysteresis. Histogram equalization of gray or color images is possible. Noise reduction is possible by using several filtering techniques such as smoothing, median filtering, sharpening, edge detection and correlation-averaging in the spatial domain or Weiner filtering in the frequency domain. Typical analysis facilities would involve line profiling (cross-sectional view), height-histograms, fast Fourier transform (FFT), roughness data and grain analysis.

4.4 Preparation and Characterization of Tip and Sample

The tip is the most important, yet least understood, part of the STM. The size, shape and chemical identity of the tip all seem to influence the resolution. The tips prepared by chemical etching, mechanical grinding or simply

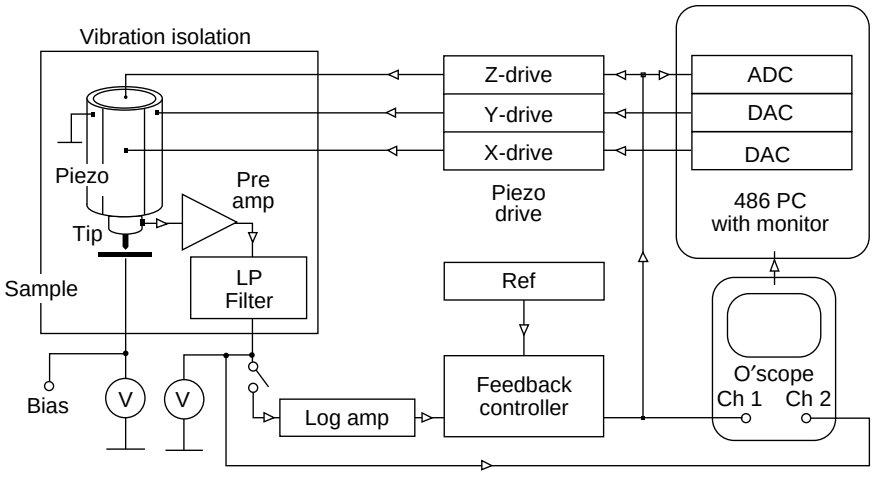

Figure 5 A block diagram of STM instrumentation.

cutting at an angle have been found to produce atomic resolution without further treatment. However, such tips may give erratic performance. The most popular method for making STM tips is based on an electrochemical technique which uses tungsten wire of 0.1 mm to 0.5 mm in diameter as a working electrode in an approximately 1 or 2M NaOH or KOH aqueous solution with carbon or stainless steel as a counter electrode. The immersed portion (2–3 mm) of the wire is etched as alternating current (AC) or direct current (DC) is passed through it at typically 5 to 15 V. Because of the preferential etching at the meniscus in the solution, neck formation takes place at the liquid interface. When the neck becomes very thin it can no longer support the lower portion, which eventually falls off due to brittle fracture. The etching is stopped at this point by an electronic control circuit, which senses the change in the etching current when the etched lower portion of the wire drops down to avoid further electrochemical reactions. The tips are then washed in distilled water, and are ready for use.[50] The AC etched tips tend to have larger cone angles than the tips prepared by DC methods, which produce sharp tips. A two-step method involving a combination of AC and DC methods can be used to get cusp-shaped tips, which are sharp and stubby. They are sometimes dipped in hydrogen fluoride to remove tungsten oxide prior to washing in distilled water, particularly for application in air. Atomic resolution can be reproducibly obtained from electrochemically etched tips by repeated tunneling or scanning or crashing the tip against the sample, or by further processing either by heating with or without electric fields, bombarding them with electrons or ions, or by operating the tip in the field emission mode. Tungsten tips are generally used for ultrahigh vacuum (UHV) applications because of the ease with which they can be etched to a radius of less than 1000 Å and cleaned in UHV by annealing or field evaporation. Platinum and platinum-based alloys (Pt–Rh, Pt–Ir) are often used for STM imaging in air because of their oxidation resistance. Due to the inherent chemical inhomogenity, tips of alloys have a greater tendency to form microasperities than single element tips after etching. In spite of the arbitrariness in the formation of microasperities, several methods have been developed empirically to process the tips which produce atomic resolution routinely.[13–21] The sample preparation for UHV involves application of standard techniques in surface analysis, namely heating, sputtering, cleaning and so forth. In air, the samples can be studied as received, but the surface is often contaminated with such substances as carbon and water vapor and one has to be careful in interpreting the results.

4.5 Vibration and Acoustic Isolation

The exponential dependence of tunnel current on distance makes vibration isolation extremely critical in STM design and installation.[43–46] Since vertical resolution depends on the mechanical stability of the tunnel junction, the noise must be well below 0.01 nm for attaining atomic resolution. There are two main approaches adopted for vibration isolation of STM. The first is spring suspension with eddy current damping and the second is a stack of metal plates separated by rubber spacers. The first one is very effective for resonance frequencies close to 1 Hz. The second one has the advantage of simplicity and ease of operation. While most of the commercial STMs come with adequate vibration isolation, care must be exercised to make sure that STM is not installed in noisy areas.

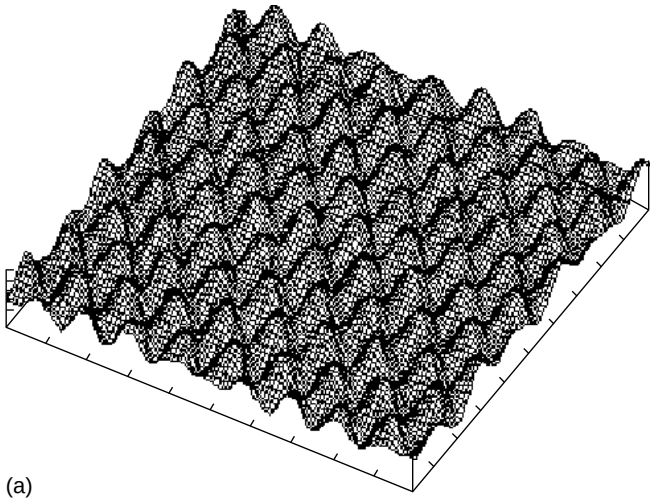

(a)

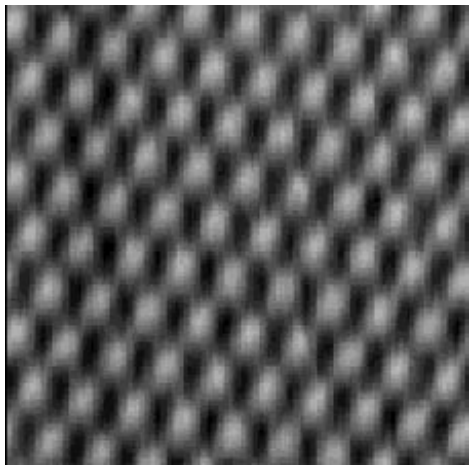

(b)

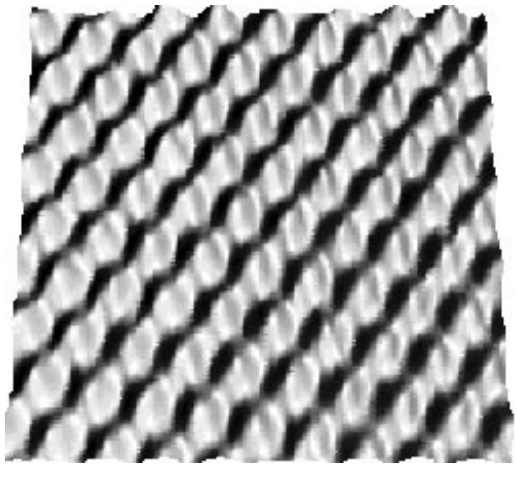

(c)

Figure 6 Three different ways of displaying of STM images: (a) line scans; (b) gray scale; (c) 3D perspective. [Reproduced with permission from Dr. K. Kelly (M.Sc. Thesis, Rice University, 1996, *Functionalization of Scanning Tunneling Microscope Probes with Buckminsterfullerenes*).]

4.6 Scanning Tunneling Spectroscopy Experimentation

The instrumentation for tunneling spectroscopy is simple and requires few additional digital-to-analog (D/A) and analog-to-digital (A/D) channels and software for the data acquisition system in addition to sample and hold circuit to the feedback loop. A/D channels are required to measure dI/dV and D/A channels for programming the tunneling voltage and tip position. The tip position can be programmed by adding a summing input to z-piezo high voltage amplifier. The timing sequence for obtaining the basic tunneling spectrum involves moving the tip to the desired location while the feedback is on, and recording the I-V curve by ramping the voltage. For a better signal-to-noise ratio, the bias voltage can be modulated for the lock-in detection. The dynamic range of the measurement can be increased by freezing the feedback and moving the probe toward the sample during the voltage sweep. Basic concepts underlying the other methods have already been discussed in section 3.5. Unlike STM imaging, where sharp tips are desirable, STS requires blunt tips as STS from sharp tips often exhibit non-reproducible tunneling spectra. Blunt tips are usually produced by positioning the tip at 100 Å from a metal substrate and drawing a few μA of field emission current at a few hundred volts. This procedure not only cleans the tip by local heating, but also melts it, leading to blunting. Such tips tend to have a reproducible density of states.

5 ILLUSTRATIVE APPLICATIONS

5.1 Structure of Crystal Surfaces [1]

Because of the inherent ability of STM to obtain real-space, atomic resolution images of smooth surfaces, it is widely used in the field of surface science for the investigation of single crystal surfaces. If a surface characterized by Miller indices (hkl) does not relax or re-orient after formation, then the resultant surface is called ideal. Surfaces of low index planes such as (110) body-centered cubic or (111) face-centered cubic are densely packed. These planes are generally atomically smooth and referred to as ''singular'' planes because they exhibit local minima in free energy plots. A surface whose orientation differs only slightly from the singular planes is termed vicinal. Such a surface is characterized by flat terraces separated by widely spaced atomic steps or ledges. The steps are straight or jagged (kinked) depending on their vicinality. The surfaces have a tendency to reconstruct and the effect is more pronounced in semiconductors due to the presence of dangling bonds. The various reconstructions are denoted by $(m \times n)$, where $m \times n$ refers to the two-dimensional Miller indices needed to

describe the surface unit cell in terms of bulk lattice vectors. At elevated temperatures or in the presence of gases, a phenomenon known as faceting occurs, where the original surface of uniform step density decomposes into hill-and-valley structures consisting of two types of plane. One of the planes has low Miller indices and other is more complex, involving crowded steps of several atomic layers with an orientation necessary to keep the average surface orientation parallel to that of the original surface. STM and STS have been extensively used for the direct observation of a variety of surfaces at the atomic scale. A wealth of information is available in a great many papers and several books.[13–21] It will not be possible to review all these results to any degree. In the following section a brief discussion of the general approach adopted for application will be described to highlight the experimental achievements and future potential of the techniques.

5.1.1 Metals, Semiconductors, Superconductors and Oxides

5.1.1.1 Metals The simplest materials that can be investigated by STM are metals.[51–53] Metals require relatively low bias (a few millivolts) because there are states at Fermi levels into or from which the electrons can tunnel. Since metals are good conductors, applied bias voltage across the barrier can be used directly as an energy parameter in spectroscopy. Because wavefunctions at Fermi energy are generally S-like, the charge density clearly follows the atomic corrugation and therefore the images can be directly related to the surface topography. Atomic resolution is generally difficult on closely packed surfaces, due to lack of surface states. The observed STM images generally show atomic steps and reconstructions. As an example, Au (110) has been found to consist of channels in [110] direction with lengths of several hundred angstroms. The channels could be interpreted as (1×2), (1×3) and sometimes (1×4) reconstructions. Figure 7 shows the STM image of a (1×2) reconstructed domain on a Au (110) surface in 0.1 M $HClO_4$ where potential induced reversible transition between nonreconstructed (1×1) and a reconstructed surface exhibiting (1×2) and (1×3) symmetries occur.[54] Similar observations have been made in UHV conditions.[11,55,56] It was found that kink sites on the Au (110) (2×1) surface are highly mobile, which appears to be important for nucleation and growth of this phase.

5.1.1.2 Semiconductors STM images of reconstructed semiconductors have been described.[12,22–27,57–59] These images are often beautiful, with clearly resolved atomic features. However, there are several difficulties in their imaging; for instance, because of low conductivity there is a considerable voltage drop over the sample.

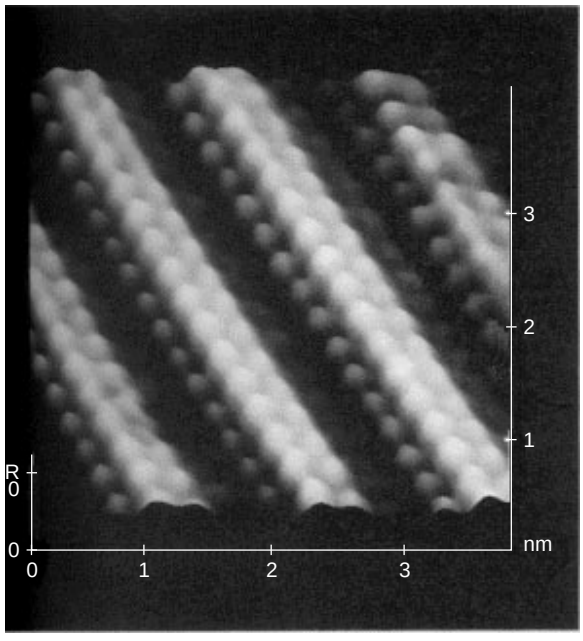

(a)

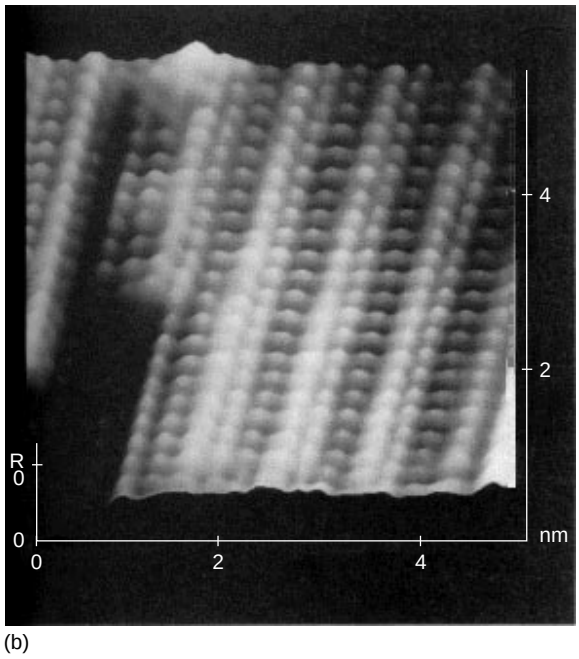

(b)

Figure 7 (a) STM image of a (1×2) reconstructed domain on a Au (110) surface in 0.1 M $HClO_4$ at an electrochemical potential of -0.3 V. (b) STM image of a (1×3) reconstructed domain. [Reproduced with permission from X. Gao, A. Hamelin, M.J. Weaver.[54]]

Despite the above restrictions, STM has been used successfully for the investigation of semiconductor surfaces, particularly Si and GaAs, revealing the details of many reconstructions. Unlike metals, the STM images of semiconductor surfaces reveal pronounced bias dependence.

The topographic maxima are usually associated with surface dangling bonds. According to their energy, certain images may be seen at some particular bias voltage. It is therefore necessary to perform voltage dependent STM/STS studies. A clean Si (111) has been known to reconstruct into a (7×7) structure upon heating typically above 1000 °C in UHV conditions ($<10^{-10}$ mbar) and subsequent cooling to ambient temperatures. STM images of Si (111) (7×7) surface reveal 12 adatoms per unit cell, as shown in Figure 8. The dangling bonds on the adatoms are partially filled and therefore contribute to both empty and filled states. Si (111) (7×7) is a surface for which STM images directly provide geometric information about the positions of the surface atoms.

5.1.1.3 Superconductors STM/STS has been used for the investigation of the local atomic and electronic structures of high-temperature superconductors (HTSC).[60]

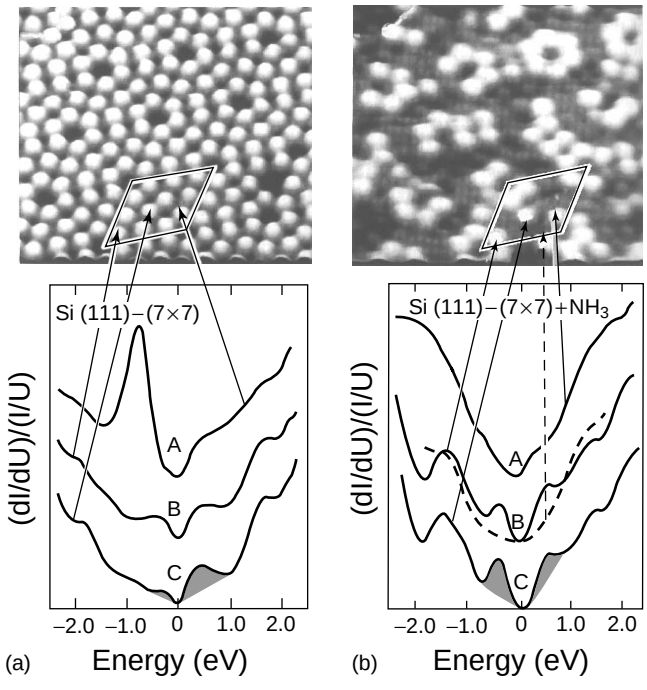

Figure 8 (a) STM topograph of the unoccupied states of the clean (7×7) surface (top), and the atom-resolved tunneling spectra (below). Curve A gives the spectrum over a rest atom site, curve B gives the spectrum over a corner adatom site and curve C gives the spectrum over a center adatom site. Negative energies indicate occupied states, while positive energies indicate empty states. (b) STM topograph of the unoccupied states (top) and adatom-resolved tunneling spectra (below) of an NH_3 exposed surface. Curve A gives the spectrum over a reacted rest atom site, curve B (dashed line) gives the spectrum over a reacted corner adatom, while curve B (solid line) and C give the spectra over an unreacted corner and a center adatom, respectively. [Reproduced by permission of the American Physical Society and the authors, Ph. Avouris and R. Wolkow.[105]]

Among the different HTSC which are oxides, the Bi- and Tl-based oxides have been widely investigated because clean surfaces are easily obtained by cleaving the samples. STM images of $Bi_2Sr_2CaCu_2O_{8+x}$ reveal the atomic lattice and a superimposed incommensurate superstructure consisting of a sinusoidal modulation with a periodicity of 9–10 unit cells along the [010] direction and a periodicity of 4.5–5 unit cells. Local tunneling spectroscopy measurements on Bi-based compounds have revealed a lack of density of states at Fermi level, indicating that the surface of the BiO layer is not metallic. The other types of HTSC have also been investigated, namely $YBa_2Cu_3O_{7-x}$ and Pb_2Sr_2 (Y, Ca) $Cu_3O_{8+\delta}$, although with some difficulty because they do not cleave easily. Figure 9 shows a typical top view of two edge dislocations on a single crystal of $Pb_2Sr_2(Y,Ca)Cu_3O_{8+\delta}$.[61,62] STM has also been widely used for surface morphology of sputter and laser deposited $YBa_2Cu_3O_{7-x}$ thin films[63] (Figure 10).

5.1.1.4 Oxides In STM, there is an obvious requirement that the sample should have sufficient conductivity so that trouble-free imaging is possible. It turns out that metallic oxides have the required conductivity and they have been successfully investigated using STM.[64] The oxides that have been imaged with atomic resolution include the HTSC discussed in the preceding section,[65–68] ReO_3,[69] WO_2,[70] sodium tungsten bronzes Na_xWO_3,[71] together with oxide bronzes such as blue bronzes $M_{0.3}MoO_3$[72] and Fe_3O_4.[73] For these materials, STM imaging is possible at either positive or negative sample bias. Stoichiometric transition metal oxides such as TiO_2,[74,75] WO_3,[76] $SrTiO_3$[77,78] and $BaTiO_3$[79] are intrinsically poor conductors; however, annealing in UHV or a partial pressure of hydrogen results in bulk oxygen loss, rendering them extrinsic

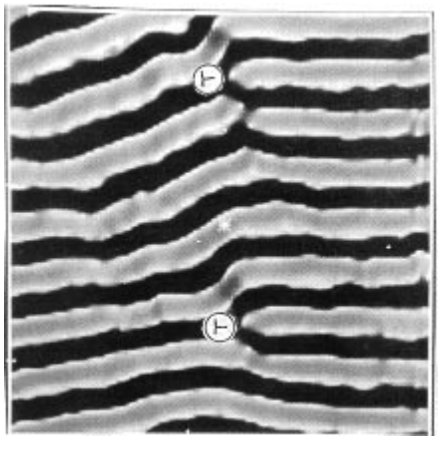

Figure 9 A top view of two edge dislocations of the same sign. (Scan size: 2.25 nm × 2.25 nm, bias voltage: 1800 mV, current: 0.23 nA, low pass filtering) [Reproduced by permission from P.K. Dutta et al.[62]]

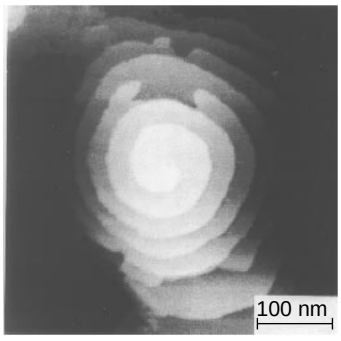

Figure 10 STM image of an $YBa_2Cu_3O_7$ thin film showing a growth hill with a screw dislocation. [Reproduced from H.P. Lang, H. Haefke, G. Leemann, H.J. Güntherodt, *Physica C*, **194**, 81, Copyright (1992) with permission from Elsevier Science.]

(n-type) semiconductors. It is possible to image these surfaces by tunneling into empty states at negative sample bias. Other alternatives to circumvent the conductivity problem involve the deposition of a thin conducting layer of metal such as Au, carrying out of experiments at elevated temperatures, or growing an epitaxial layer on the metal substrates. A typical STM image of an oxide, like metals and semiconductors, will exhibit an ordered array of maxima and minima in addition to steps, dislocations and other such defects. In metals, the maxima invariably correspond to atomic position; in semiconductors, they may correspond to atomic position or may involve electronic structure. In oxides, the interpretation of STM images is not always unambiguous. A major issue, therefore, is to establish a correspondence between maxima in gray scale images and atomic positions within the unit cell.[64]

5.1.2 Adsorption, Surface Diffusion and Surface Chemistry

5.1.2.1 Adsorption Adsorption[80] is a necessary step in numerous reactions at solid surfaces. Research work on the phenomenon of adsorption has been directed at determining the nature of the binding sites and interactions between adsorbed atoms and molecules, which often lead to the formation of ordered layers. STM images of adsorbate-covered metal surfaces depend strongly on the applied bias voltage, electronic structure and surface mobility of the adsorbate at the temperature of the experiment. Adsorption of oxygen has been studied for a variety of metal substrates. For instance, at room temperature, the highly mobile oxygen adatoms, originating from dissociative chemisorption, can diffuse on terraces of the unconstructed Cu (110) surface. Cu adatoms evaporating from surface step edges are highly mobile and also diffuse across the terraces of the substrate surface. This results in nucleation of the (2×1) phase, consisting of single strings of copper atoms glued together by oxygen atoms.[80,81] Similar anisotropic island formation has been reported for other planes and different adsorbate substrate combination. The periodicities of superstructures are found to depend on the oxygen coverage and the substrate temperature. Adsorption of other species such as carbon, sulphur, carbon monoxide and halogens on metals has also been investigated. Investigation of the initial stages of oxidation of semiconductor surfaces is important and has been investigated using STM. These studies have shown that defect sites on the clean Si (111) (7×7) surfaces, such as vacancies or (7×7) domain boundaries, act as nucleation centers for the oxidation process.[82,83] The straight atomic steps were found to be relatively insensitive to oxygen exposure. Atomic hydrogen adsorption on a Si (111) (7×7) surface has been investigated with a view to understanding the formation and disruption of the silicon–hydrogen bond, due to its importance in many technologically valuable processes. The direct bonding of hydrogen atoms to existing surface dangling bonds is observed at low coverage.[84] Increasing hydrogen exposure results in the removal of the uppermost silicon adatom layer, accompanied by binding of hydrogen atoms to the dangling bonds created. The STM images reveal (1×1) structures in the background of (7×7) unit cells.

5.1.2.2 Surface Diffusion Surface diffusion of adatoms is central to the understanding of a variety of phenomena, such as adsorption, chemical reactions and surface phase transitions. Direct atomic scale observation of surface diffusion and the determination of the diffusion coefficient, activation energies and atomic migration frequency is possible using STM. Further analysis of the images yields quantitative information about diffusion anisotropy and its correlation to atomic sites. Diffusion coefficients as low as $10^{-17}\,cm^2\,s^{-1}$ ($10^{-21}\,m^2\,s^{-1}$) can, in principle, be measured using STM. Before the advent of STM, such low diffusion constants were measured using FIM, and valuable information is available on diffusion of metal atoms on metal surfaces. However, the techniques for diffusion of atoms on semiconductor surfaces did not exist prior to STM invention. Although some experimental difficulties still prevent accurate measurement of the absolute value of the diffusion coefficient, investigations have been carried on the atomic-scale dynamics of adsorbed atoms, giving a glimpse of the future potential of this technique. Recent real-time observation of the dynamics of single Pb atoms on Si (111) suggest that single Pb atoms are mobile within each half (7×7) unit cell.[85] Individual jumps of single atoms between different half-cells have been resolved, with a measured activation energy of $0.64 \pm 0.07\,eV$. In

contrast, the motion of Pb atoms on Ge (111) C (2×8) structures takes place between well-defined equilibrium positions.[86,87] STM observations for Si and Ge adatoms on top of Si (100) (2×1) indicate a strong anisotropy in the surface diffusion process due to the dimer-type (2×1) reconstruction. Diffusion is typically a thousand times faster along the surface dimer rows than perpendicular to them.[17]

5.1.2.3 Surface Chemistry Understanding of a catalyst's structures as well as the structure of adsorbed molecules under catalytic reaction conditions is important for elucidating how a reaction progresses on the atomic scale.[88,89,90] STM investigation of a Pt (111) crystal surface in the presence of a hydrogen–propylene mixture (10 : 1) at 1 atmosphere and room temperature reveals featureless terraces separated by straight mono-atomic height steps, even though the surface is catalytically active and propylene is hydrogenated to propane during STM imaging. The most significant result of the in situ STM images was that the underlying platinum atoms were not significantly reconstructed by the adsorbed species. Controlled experiments performed by annealing the sample to temperatures > 770 K in propylene to form carbon clusters result in substantial reconstruction, and STM was sensitive enough to observe it. The steps were no longer straight but exhibited a propensity for kinks or protrusions in the steps that are pinned on the top of some clusters. These results indicate that STM is especially suitable for imaging defects at high pressures and elevated temperatures and is capable of detecting stationary species and adsorbate-induced surface restructuring as it occurs during chemisorption or a catalytic reaction. STM, however, cannot at present image surface species that are mostly small chemisorbed organic molecules at 300 K or above. STM can be combined with surface-specific vibrational spectroscopy such as sum-frequency generation (SFG), which can operate under dynamic high-pressure reaction conditions to get additional information about the molecular structure of the adsorbates and reaction intermediates.[91]

5.1.3 Thin Film Morphology, Nucleation, Growth, etc.

A study of nucleation and growth of thin films has been an active area of research for the last several decades because its outcome has a direct bearing on a variety of processes of technological importance. Typical nucleation and growth experiments involve characterization of films obtained under a variety of controlled deposition conditions. The experiments using STM during the last 10 years have provided quantitative information on island distribution within each layer, their dependence on external parameters, temperature and deposition flux and their strong dependence on interlayer transport. The importance of STM lies in its ability to actually observe islands, thin shapes, size, coverage and spatial distribution on atomic length scales.[92] Traditionally, three different growth mechanisms have been identified based on thermodynamic consideration: Frank–Van der Merwe for monolayer growth, Volmer–Weber for initial film nucleation by 3-D crystallite formation and Stranski–Krastanov for formation of an initial uniform layer followed by 3-D crystallite growth.[93] Recently, it has been realized that these thermodynamic models need to be modified to account for kinetic limitations. The solid films grown under far from equilibrium conditions are consistently predicted to have self-affine surfaces and are therefore candidates for a recently developed dynamical scaling approach. The experimental determination of growth exponents has been an area of intense activity during this decade and STM, because of its ability to measure roughness with unprecedented resolution, has contributed enormously to this field.[94] The field of nucleation and growth investigation using STM has advanced considerably, and several exhaustive review articles and monographs are available on this topic.[80,93,94] The atomic scale mechanisms involved in crystal growth strongly depend on crystal growth conditions and crystallographic orientation of the substrate. Even if the direct observation of atomic scale growth processes may not be possible in every such case, the morphology of layers formed during the growth can provide information about growth processes like surface diffusion, island formation and growth at step edges[95] (Figure 11).

STM has also been used for the investigation of films grown under most real-life environments. The growth of diamond by chemical vapor deposition (CVD) is one such example. In recent years, the subject of diamond films prepared by CVD has emerged as a technologically useful and scientifically interesting field of research.[96]

The diamond films are usually prepared by the hot-filament method, the microwave plasma CVD method, or by a combination of both. Several investigations have been reported on the application of STM for

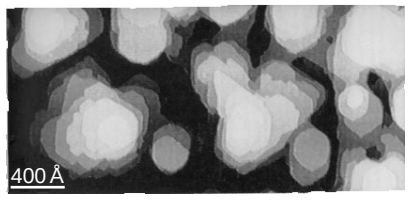

Figure 11 STM image (300 nm × 150 nm) of a high cover age (>5M_L) Cu film after room temperature deposition. Multilayered islands of characteristic pyramidal shape are formed, whose edges are oriented along the closed-path directions of the substrate. [Reproduced with permission from G. Pötshke, J. Schröder, C.Günther, R.Q. Hwang, R.J. Behm, *Surf. Sci.*, **251/252**, 592 (1991).[95]]

nucleation-growth investigation of diamond films.[97–99] In one of the recent works,[100] the diamond films grown by the conventional hot filament CVD method (HF–CVD) using a hydrogen–methane (100 : 1) gas mixture (gas flow of 100 sccm), with a chamber pressure of 30 torr at substrate temperature T_c ca. 850 °C to 900 °C for a duration of 3 to 4 h were investigated using STM. Because of the high band gap (5.4 eV) of diamond, the films had to be doped by irradiating them with boron ions of 500–800 keV with doses up to 10^{14} ions cm^{-2}, and annealed at high temperature for successful STM imaging. The SEM images of these films showed diamond crystallites with predominantly (100) facets. The STM images showed a large number of pyramidal nanofacets exhibiting layered structure (100) symmetry commensurate with the underlying facet. The smallest terrace top was typically ca. 50 Å and showed no evidence of dislocations. The quantitative analysis of results indicates that the transport processes such as surface diffusion are more dominant than evaporation, condensation processes in the growth of (100) diamond, leading to layer-by-layer growth (Figure 12).

5.1.4 Surface Electronic Structure

The origin of electronic structure effects in STM imaging has been discussed in section 3.5. The electronic structure effects have been investigated for a variety of metals and semiconductor surfaces.[13–27] Only two of the many celebrated applications will be considered here to illustrate the type of electronic information available from STM/STS results. It is well known that cleaving of Si (111) surface in UHV results in a Si (111) (2 × 1) structure, in contrast to (7 × 7) structures after annealing cycles at >800 °C. This is a metastable surface and its STM image is dominated by electronic effects.[101–103] The voltage-dependent images in Figure 13 show that the maximum in the image taken at +1 V appears as a minimum at −1 V because, as seen in the center of the cross-hairs in the two images indicating shift in topographic maxima by half a unit cell in the $[0\bar{1}\bar{1}]$ direction. The results are consistent with π-bonded chain model.[104] The two adatoms in the unit cell of this model are nonequivalent due to the underlying symmetry of the lattice and also possibly due to buckling of the chain. Tunneling at positive sample bias then accesses the empty states while at negative bias it probes the filled states.

One of the striking examples of electronic structure studies using STM is the adsorption of NH_3 on the

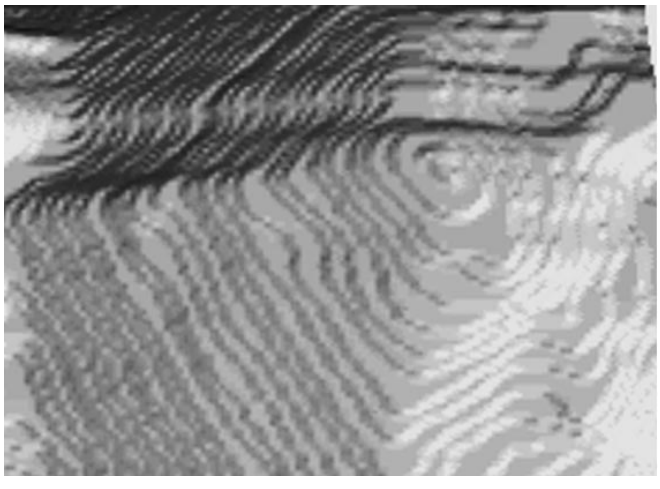

Figure 12 STM images (1000 × 600 Å) of (100) diamond facets exhibiting layered growth with pyramidal shape. [Reproduced by permission of the American Institute of Physics from the article by V.P. Godbole et al.[100]]

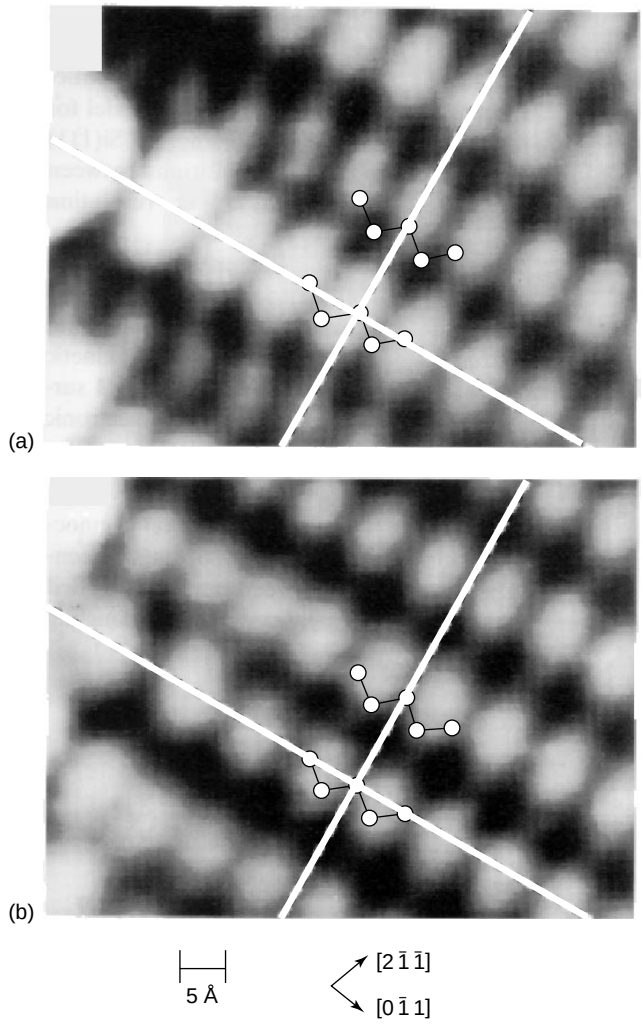

Figure 13 Two STM images of the Si (111) (2 × 1) surface, acquired at sample bias voltages of (a) +1 V and (b) −1 V. [Reproduced by permission of the American Institute of Physics from the article by J.A. Stroscio, R.M. Feenstra, D.M. Newns and A.P. Fein.[102]]

Si (111) (7 × 7) surface. Figure 8 shows STM images of clean (a) and NH_3 exposed surfaces. (b) NH_3 is known to dissociate on the Si (111) (7 × 7) surface, producing Si–H and Si–NH_2 groups. The reacted atoms appear dark because of the saturation of their dangling bonds by the chemical reaction process, leading to decrease in the LDOS near E_F. It turns out that the products of surface reaction can also be imaged at negative sample bias, probing the occupied electronic states. These images are very rich in electronic information.[105] Firstly, STM images of clean Si (111) (7 × 7) clearly show two nonequivalent types of surface Si adatoms, called ''corner adatoms'' and ''center adatoms''. It can be seen that the majority of unreacted sites are ''corner adatoms'', while the ''center adatoms'' exhibit a significantly higher degree of local reactivity. The influence of the reaction process on the local electronic structure can be directly studied by comparing local $(dI/dV)/(I/V) - V$ characteristics, where I is tunnel current and V is the bias voltage. It must be noted that such detailed information would not be available from conventional surface analytical techniques averaged over large surface areas. Even in less ideal environments, namely those for the investigation of oxides, diamond films and so forth, the special resolution of STM is far better than the conventional techniques.[106–108]

5.1.5 *Organic Molecules, Biological Molecules and Electrochemistry*

5.1.5.1 Organic Molecules STM has been extensively used for imaging molecules in organic conducting salts or organic molecules adsorbed on various conducting substrates.[19] Although STM images of organic adsorbed molecules are spectacular and clearly resolve complex multilayer structural details at the molecular level, the structural characterization of the organic adsorbate has not yet met with the same success as metals or semiconductors. In most cases, the molecular order at the liquid–solid interface is found in the vicinity of the substrate surface. The topmost layers are usually etched away by the tip, which makes it difficult to characterize their surface properties away from the interface. In almost all cases, one is imaging the interface and therefore the observed ordered layers are sensitive to substrate topography. On the atomically flat surface of graphite or specially prepared gold or silver films, organic molecules tend to form a lamellar structure. The mismatch between the molecular shape and surface topography also influences the ordered layer formation. One solution would be to develop more sensitive instruments which would operate at low current (picoamperes) and probably high bias (>10 V). Under this condition STM works in field emission mode.

Organic conducting salts are prepared as high-quality single crystals that are stable in air, and the crystals are characterized by well-defined flat surfaces. The molecular arrangements of these surfaces have been resolved using STM.[19] Although there are difficulties in the interpretation of images as they appear to depend on imaging conditions, there is a strong indication that tip–sample force interaction exists, leading to the removal of the surface layers, the migration of molecular defects and changes in the positions of the topmost molecular group. Some of the organic molecular systems that have been extensively investigated are: liquid crystals, alkanes and alkane-derived molecules, Langmuir–Blodgett (LB) films, polymers and fullerene films.

5.1.5.2 Biological Molecules The greatest incentive to use STM for the investigation of biological molecules is its ability to operate in an atmospheric and fluid environment and observe the in vivo dynamics of processes as they occur in real time. In addition, STM techniques are less destructive than most of the conventional electron microscopy techniques, which involve bombardment with high-energy electrons. However, there are major challenges involved in the successful application of STM to biological molecules, such as specimen preparation, location of features of interest, and the understanding of image contrast for thick micromolecules. In spite of these problems, imaging of biological specimens is one of the most rapidly growing applications of STM and has been extensively used for the investigation of nucleic acids, proteins and membranes.[17]

5.1.5.3 Electrochemistry The techniques for the investigation of solid–vacuum or solid–gas interfaces at atomic level have advanced considerably, although corresponding techniques for liquid–solid interfaces are still scarce. With the advent of STM, it has been possible for the first time to investigate the microstructure of a solid–liquid interface and its relation to other properties. The two STM-based techniques that operate in electrolyte solutions are: scanning electrochemical microscopy (SECM)[109] and scanning ion conductance microscopy (SICM).[110] The SECM uses the faradic current for stabilization of the tip–sample distance. The resulting image combines topographic information and the electrical and chemical properties of the sample. In SICM, a micropipet filled with electrolytes scans the sample. The conductance between micropipet and sample depends on the cross-section of the ion path and thus on the distance between micropipet and sample. The resolution depends on the micropipet opening but can be as small as 1000 Å.[110] As an example of an application of STM in electrochemistry, it has been directly shown for a Au (110) surface in 0.1 M $HClO_4$ solution that a

reversible potential-induced transition between a nonreconstructed (1 × 1) and a reconstructed surface exhibiting (1 × 2) symmetry occurs.[54] The results obtained for a Au (110) surface structure are largely consistent with the missing-row model.

5.1.6 Atom Manipulation and Nanolithography

The basic geometry of STM, a sharp tip held in close proximity to the sample surface with nanometer precision, lends itself as a powerful tool for nanofabrication, i.e. fabrication at the nanometer level by controlling the interaction necessary for imaging, manipulation and modification by adjusting the tip–sample distance or by changing the strength of interaction, that is by changing the bias voltage across the junction. A tantalizing glimpse of atom-by-atom manipulation by STM was demonstrated by Eigler and Schweizer[111] in which xenon atoms were moved into the desired location and the pattern thus formed was subsequently imaged at liquid helium temperatures in UHV conditions. In another application, Xe atoms were reversibly transferred between tip and the sample using high-voltage pulses of opposite polarity at low temperatures.[112] The scratching and indentation by the STM tip has also been used for mechanical modification of surfaces.[113,114] The localized electron transport provides an extremely effective focusing of electrons both by tunneling and field emission modes. This, along with the inherent electric field in the tunnel junction, renders STM a versatile tool for electron lithography.[115] A variety of electron beam resists, such as hydrocarbons, metal halides, poly(methyl methacrylate) (PMMA), and LB films have been exposed by STM operated in field emission mode (bias voltage higher than work function).[115,116] Line widths below 100 nm have been routinely achieved. The other methods based on chemical surface modifications using low-energy electrons offered by STM include: local oxidation passivated semiconductor surfaces,[117] electron enhanced etching,[118] electron stimulated desorption,[119] decomposition of organometallic gases[120] and electrochemical etching and decomposition.[121]

6 COMPARISON WITH OTHER METHODS

6.1 Scanning Tunneling Microscopy Compared with High-resolution Electron Microscopies

In the context of atomic resolution capability, there are very few techniques, namely STEM, TEM, and FIM which perform as well as STM, but they do so under special circumstances. For instance, STEM and TEM use high-energy electron beams, require very thin film samples, underlying periodic structures and large mass contrast such as heavy atoms on carbon film. FIM is usable only for refractory metals which can withstand the high electric fields required for imaging. In addition, only edge atoms on flat parts of a few hundred angstroms size are visible. In STM, the energy of the tunneling electrons is very low, typically a few meV to a few volts and electric fields three to four orders of magnitude less than FIM. Unlike REM, which has near atomic resolution only in the vertical direction, STM has very high vertical and lateral resolution, yielding a true 3-D topography of planar surfaces on the atomic scale. An inherent limitation of STM is that it always operates at high resolution and, therefore, works better for atomically flat surfaces. Such surfaces, however, can be easily prepared by standard techniques such as annealing, sputtering and cleaving in UHV.

The most important characteristic of STM that distinguishes it from other techniques is its somewhat surprising ability to operate in a variety of environments, such as air, reactive gases, liquids, electrolytes, greases and biological fluids. In addition, imaging with a lateral resolution of nanometers is important in the context of potential applications of STM in technologically important areas such as thin film device fabrication, preparation and characterization of industrially important surfaces, and surface roughness measurement.

6.2 Scanning Tunneling Spectroscopy Compared with Other Electron Spectroscopies

The surface spectroscopic techniques that can be compared with STS are: ETS, IETS, UPS and IPS. In contrast to conventional ETS and IETS, STS does not require high-quality oriented films. Also, there is no electric breakdown problem because a vacuum or air forms the barrier instead of an oxide, leading to a larger accessible energy range (typically an energy window of $E_F \pm 3$ eV). In addition, it can be performed locally at a preselected position under well-defined conditions and can be used in combination with other surface sensitive techniques, and provides spectroscopic images of electron states. In comparison with UPS and IPS, which probe a depth comparable to the escape depth of electrons (5 nm), STS is more surface sensitive because of its strong dependence on wave function overlap, which takes place only at a separation of a few angstroms. Moreover, it is able to detect both occupied and unoccupied states by simply reversing the tunnel bias.

7 SCANNING PROBE MICROSCOPY: RECENT TRENDS AND FUTURE PROSPECTS

The advent of STM has revolutionized our ability to see objects never seen before, and with minimum or no

sample preparation. The STM originally gained attention as a simple instrument for imaging atoms. Just a few years after its invention, it has established itself in surface science as an indispensable tool and is now being used in combination with other surface science techniques such as Auger electron spectroscopy (AES) and LEED. In the early years, most STMs were built in-house. Now they are available commercially; however, the trend is still to build them for very specific applications. From its modest beginnings, STM has instigated the growth of a family of instruments leading to a totally new means of investigation, SPM. In its most basic form, SPM senses some interaction between the proximal probe and the sample as a function of the distance between the object and the probe. If the interaction probe is either repulsive or attractive van der Waals force then the investigation is by atomic force microscopy (AFM). If frictional force is measured, then it is called lateral force microscopy. If electric force or magnetic force is monitored, then it is called electric force microscopy or magnetic force microscopy. If the spatial variation of the evanescent field is probed by a sharp scanning fiber optical probe, the technique is called scanning near field optical microscopy. The derivatives of each of the above methods have led to an arsenal of different techniques.[13–21] In just a decade or so of its existence, SPM has been transformed from a laboratory curiosity into one of the most powerful and versatile techniques in the optics, semiconductor, and computer industries.

ACKNOWLEDGMENTS

The financial assistance given to the STM project by BRNS, DAE, and the government of India, is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
A/D	Analog-to-digital
AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
CITS	Current Imaging Tunneling Spectroscopy
CVD	Chemical Vapor Deposition
D/A	Digital-to-analog
DC	Direct Current
ETS	Electron Tunneling Spectroscopy
FEM	Field Electron Microscopy
FFT	Fast Fourier Transform
FIM	Field Ion Microscopy
HTSC	High-temperature Superconductors
IETS	Inelastic Electron Tunneling Spectroscopy
IPS	Inverse Photoemission Spectroscopy
LB	Langmuir–Blodgett
LDOS	Local Density of States
LEED	Low Energy Electron Diffraction
PMMA	Poly(methyl methacrylate)
REM	Reflection Electron Microscopy
SECM	Scanning Electrochemical Microscopy
SEM	Scanning Electron Microscopy
SFG	Sum-frequency Generation
SICM	Scanning Ion Conductance Microscopy
SPM	Scanning Probe Microscopy
STEM	Scanning Transmission Electron Microscopy
STM	Scanning Tunneling Microscopy
STS	Scanning Tunneling Spectroscopy
TEM	Transmission Electron Microscopy
UHV	Ultrahigh Vacuum
UPS	Ultraviolet Photoemission Spectroscopy
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Polymers and Rubbers **(Volume 8)**
Atomic Force Microscopy in Analysis of Polymers

Surfaces **(Volume 10)**
Surfaces: Introduction • Electron Microscopy and Scanning Microanalysis • Proximal Probe Techniques • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Probe Microscopy, Industrial Applications of • Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

REFERENCES

1. J.M. Blakely, *Introduction to the Properties of Crystal Surfaces*, Pergamon, Oxford, 1973.
2. L. Reimer, *Transmission Electron Microscopy*, (Springer Series in Optical Sciences), Berlin, 1984.
3. C.W. Oatley, *Scanning Electron Microscopy*, Cambridge University Press, Cambridge, 1972.
4. G. Ertl, J. Kuppers, *Low Energy Electron and Surface Chemistry*, Verlag Chemie, Weinheim, 1974.
5. E. Bauer, *Rep. Prog. Phys.*, **57**, 895 (1994).
6. I.K. Robinson, *Rep. Prog. Phys.*, **55**, 599 (1992).
7. R. Gomer, *Field Emission and Ionization*, Harvard University Press, Cambridge, Mass., 1961.
8. E.W. Muller, T.T. Tsong, *Field Ion Microscopy*, Elsevier, New York, 1969.

9. R.D. Young, J. Ward, F. Scire, *Rev. Sci. Instrum.*, **43**, 999 (1972).
10. G. Binnig, H. Rohrer, C.H. Gerber, E. Weibel, *Appl. Phys. Lett.*, **40**, 178 (1982).
11. G. Binnig, H. Rohrer, C.H. Gerber, E. Weibel, *Phys. Rev. Lett.*, **49**, 57 (1982).
12. G. Binnig, H. Rohrer, C.H. Gerber, E. Weibel, *Phys. Rev. Lett.*, **50**, 120 (1983).
13. R. Behm, N. Garcia, H. Rohrer (eds.), *Scanning Tunneling Microscopy and Related Methods*, Kluwer Academic Publishers, Dordrecht, 1990.
14. D.A. Bonell, *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques and Applications*, VCH Publishers, New York, 1993.
15. C.J. Chen, *Introduction to Scanning Tunneling Microscopy*, Oxford University Press, Oxford, 1993.
16. R. Wiesendanger, H.J. Guntherodt (eds.), *Scanning Tunneling Microscopy, Vols I, II and III*, (Springer Series in Surface Science Vols 21, 28, 29), Springer-Verlag, Berlin, 1994, 1995, 1996.
17. R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, 1994.
18. N.J. DiNardo, *Nanoscale Characterization of Surfaces and Interfaces*, VCH Publishers, New York, 1994.
19. S.N. Magonov, *Surface Analysis with STM and AFM: Experimental and Theoretical Aspects of Image Analysis*, VCH Publishers, Weinheim, 1996.
20. J.A. Stroscio, J. Kaiser (eds.), *Scanning Tunneling Microscopy* (Methods in Experimental Physics, Vol. 27), Academic Press, San Diego, 1993.
21. C. Bai, *Scanning Tunneling Microscopy and its Applications* (Springer Series in Surface Science, Vol. 32), Springer-Verlag Berlin, 1995.
22. G. Binnig, H. Rohrer, *Helv. Phys. Acta*, **55**, 72 (1982).
23. G. Binnig, H. Rohrer, *Rev. Mod. Phys.*, **59**, 615 (1987).
24. C.F. Quate, *Phys. Today*, **39**, 26 (1986).
25. H. Rohrer, *Ultramicroscopy*, **42–44**, 1 (1992).
26. H. Rohrer, *Surf. Sci.*, **299/300**, 956 (1994).
27. G. Binnig, H. Rohrer, *Rev. Mod. Phys.*, **71**, 5324 (1999).
28. A.P. French, E.F. Taylor, *An Introduction to Quantum Physics*, Van Nostrand Reinhold, New York, Chapter 9, 1978.
29. D.K. Roy, *Quantum Mechanical Tunneling and its Applications*, World Scientific, Singapore, 1986.
30. C.B. Duke, 'Tunneling in Solids', in *Solid State Physics*, Supplement 10, eds. F. Seitz, D. Turnbull, Academic Press, New York, 1969.
31. E.L. Wolf, *Principles of Electron Tunneling Spectroscopy*, Oxford University Press, New York, 1989.
32. C.B. Duke, *Tunneling in Solids*, Plenum Press, New York, 1969.
33. J.G. Simmons, *J. Appl. Phys.*, **34**, 1793 (1963).
34. R.D. Young, J. Ward, F. Scire, *Phys. Rev. Lett.*, **27**, 922 (1971).
35. E.C. Teague, *Bull. Am. Phys. Soc.*, **23**, 290 (1978).
36. C.J. Chen, *J. Vac. Sci. Technol., A*, **9**, 44 (1991).
37. C.J. Chen, *Phys. Rev. Lett.*, **65**, 448 (1990).
38. A. Bryant, D.P.E. Smith, C.F. Quate, *Appl. Phys. Lett.*, **48**, 832 (1986).
39. J. Tersoff, D.R. Hamann, *Phys. Rev. B, Condens. Matter*, **31**, 805 (1985).
40. J. Tersoff, D.R. Hamann, *Phys. Rev. Lett.*, **50**, 1998 (1983).
41. R.M. Feenstra, in *Scanning Tunneling Microscopy and Related Methods*, eds. R. Behm, N. Garcia, H. Rohrers, Kluwer Academic Publishers, Dordrecht, 211, 1990.
42. P.K. Hansma, *Tunneling Spectroscopy: Capabilities, Applications and New Techniques*, Plenum, New York, 1982.
43. D.W. Pohl, *IBM J. Res. Develop.*, **30**, 417 (1986).
44. Y. Kuk, P.J. Silverman, *Rev. Sci. Instrum.*, **60**, 165 (1989).
45. Y. Kuk, P.J. Silverman, *Scanning Tunneling Microscopy and its Applications* (Springer Series in Surface Science, Vol. 32), Springer-Verlag, Berlin and Heidelberg, Chapter 4, 1995.
46. D.P.E. Smith, G. Binnig, *Rev. Sci. Instrum.*, **57**, 2630 (1986).
47. D.P. Dilella, J.H. Wandass, R.J. Colton, C.R. Marrian, *Rev. Sci. Instrum.*, **60**, 997 (1989).
48. M. Aguilar, P.J. Pascual, A. Santisteban, *IBM J. Res. Develop.*, **30**, 525 (1986).
49. P.H. Schroer, J. Becker, *IBM J. Res. Develop.*, **30**, 543 (1986).
50. A.J. Melmed, *J. Vac. Sci. Technol., B*, **9**, 601 (1991).
51. Y. Kuk, in *Scanning Tunneling Microscopy* (Methods in Experimental Physics, Vol. 27), eds. J.A. Stroscio, J. Kaiser, Academic Press, San Diego, Chapter 6, 1993.
52. Y. Kuk, *Scanning Probe Microscopy and Spectroscopy*, ed. R. Wiesendanger, Cambridge University Press, Cambridge, Chapter 4, 1994.
53. L.L. Soethout, H. Van Kempen, G.F.A. van de Walle, *Adv. Electron Phys.*, **79**, 155 (1990).
54. X. Gao, A. Hamelin, M.J. Weaver, *Phys. Rev. B, Condens. Matter*, **44**, 10 983 (1991).
55. Y. Kuk, P.J. Silverman, F.M. Chua, *J. Microscop.*, **152**, 449 (1988).
56. H.Q. Nguyen, Y. Kuk, P.J. Silverman, *J. de. Physique.*, **49**, 7988 (1989).
57. R. Becker, R. Wolkow, in *Scanning Tunneling Microscopy*, (Methods in Experimental Physics, Vol. 27), eds. J.A. Stroscio, J. Kaiser, Academic Press, San Diego, Chapter 5, 1993.
58. R. Becker, R. Wolkow, in *Scanning Probe Microscopy and Spectroscopy*, ed. R. Wiesdanger, Cambridge University Press, Cambridge, Chapter 4, 1994.
59. R. Wolkow, P. Avouris, *Phys. Rev. Lett.*, **60**, 1049 (1988).
60. Z. Zhang, C.M. Lieber, *J. Phys. Chem.*, **96**, 2030 (1992).
61. A.V. Narlikar, P.K. Dutta, S.B. Samanta, O.N. Srivastava, *J. Cryst. Growth*, **121**, 527 (1992).
62. P.K. Dutta, S.B. Samanta, A.V. Narlikar, C. Chen, J.W. Hodby, B.W. Wanklyn, *Philos. Mag. A*, **66**, 507 (1992).

63. H.P. Lang, R. Wiesendanger, V. Thommeuson-Geiser, H.J. Güntherodt, *Phys. Rev. B, Condens. Matter*, **45**, 1829 (1992).
64. R.G. Egdell, F.H. Jones, *J. Mater. Chem.*, **8**, 469 (1998).
65. C.K. Shih, R.M. Feanstra, et al., *Phys. Rev. B, Condens. Matter*, **40**, 2682 (1989).
66. M.D. Kirk, D.P.E. Smith, et al., *Science*, **242**, 1673 (1988).
67. X.L. Wu, Z. Zhang, Y.L. Wang, C.M. Lieber, *Science*, **248**, 1211 (1990).
68. X.L. Wu, C.M. Lieber, D.S. Ginley, R.J. Bangham, *Appl. Phys. Lett.*, **55**, 2129 (1989).
69. S. Ikebe, K. Suzuki, H. Nishikawa, *Jpn. J. Appl. Phys.*, **31**, 2221 (1992).
70. F.H. Jones, R.G. Egdell, A. Brown, F.R. Wondre, *Surf. Sci.*, **374**, 80 (1997).
71. F.H. Jones, et al., *Surf. Sci.*, **336**, 181 (1995).
72. W. Lu, N. Nevins, M.L. Norton, G.S. Rohrer, *Surf. Sci.*, **291**, 395 (1993).
73. N.G. Condon, F.M. Leibsle, T. Parkar, A.R. Lennie, D.J. Vaughan, G. Thorton, *Phys. Rev. B, Condens. Matter*, **55**, 15 885 (1997).
74. G.S. Rohrer, V.E. Henrich, D.A. Bonnell, *Science*, **250**, 1239 (1990).
75. P.W. Murray, J. Shen, N.G. Condon, S.J. Pang, G. Thorton, *Surf. Sci.*, **380**, L455 (1997).
76. F.H. Jones, R.A. Dixon, A. Brown, *Surf. Sci.*, **369**, 343 (1996).
77. Y. Liang, D.A. Bonnell, *Surf. Sci.*, **285**, L510 (1993).
78. Q. Jiang, J. Zegenhagen, *Surf. Sci.*, **367**, L42 (1996).
79. H. Bando, T. Shimitzu, et al., *J. Vac. Sci. Technol., B*, **14**, 1060 (1996).
80. F. Basenbacher, *Rep. Prog. Phys.*, **59**(10), 1737 (1996).
81. D. Coulman, J. Winrelin, R.J. Behm, G. Ertl, *Phys. Rev. Lett.*, **64**, 1761 (1990).
82. F.M. Leibsle, A. Samsavar, T.C. Chiang, *Phys. Rev. Lett.*, **38**, 5780 (1988).
83. J.P. Pelz, R.H. Koch, *J. Vac. Sci. Technol., B*, **9**, 775 (1991).
84. J.J. Boland, *J. Vac. Sci. Technol., B*, **9**, 764 (1991).
85. J.M. Gomez-Rodriguez, J.J. Saenz, A.M. Baro, J.Y. Yeuillen, R.C. Cinti, *Phys. Rev. Lett., B*, **76**, 799 (1996).
86. E. Ganz, S.K. Thesis, I.S. Hwang, J. Golovchenko, *Phys. Rev. Lett.*, **68**, 1567 (1992).
87. I.S. Hwang, J. Golovchenko, *Science*, **265**, 490 (1994).
88. G.A. Somarjai, *Chem. Rev.*, **96**, 1223 (1996).
89. P.S. Cremer, B.J. McIntyre, M. Salmeron, et al., *Catal. Lett.*, **34**, 11 (1995).
90. B.J. McIntyre, M.B. Salmeron, G.A. Somorjai, *Rev. Sci. Instrum.*, **64**(3), 687 (1993).
91. Y.R. Shen, *Nature*, **337**, 519 (1989).
92. R.Q. Hwang, M.C. Bartelt, *Chem. Rev.*, **97**, 1063 (1997).
93. J.A. Venables, G.D.T. Spiller, M. Handbucken, *Rep. Prog. Phys.*, **47**, 399 (1984).
94. A.L. Barabasi, H.E. Stanely, *Fractal Concepts in Surface Growth*, Cambridge University Press, Cambridge, 1995.
95. G. Pötshke, J. Schröder, C. Günther, R.J. Behm, *Surf. Sci.*, **251/252**, 592 (1991).
96. R.F. Davis (ed.), *Diamond Films and Coatings*, Noyes Publications, Park Ridge, NJ, 1993.
97. T. Tsuno, T. Tomikawa, S. Shikata, N. Fujimori, *J. Appl. Phys.*, **75**, 1526 (1994).
98. M.P. Everson, M.A. Tamor, *J. Mater. Res.*, **7**, 1438 (1992).
99. T.J. Kreutz, R.E. Clausing, et al., *Phys. Rev. B, Condens. Matter*, **51**, 14 554 (1995).
100. V.P. Godbole, A.V. Sumant, R.B. Kshirsagar, C.V. Dharmadhikari, *Appl. Phys. Lett.*, **71**, 2626 (1997).
101. J.A. Stroscio, R.M. Feanstra, A.P. Fein, *Phys. Rev. Lett.*, **57**, 2579 (1986).
102. J.A. Stroscio, R.M. Feanstra, D.M. Newns, A.P. Fein, *J. Vac. Sci. Technol. A*, **6**, 499 (1988).
103. J.A. Stroscio, R.M. Feanstra, A.P. Fein, *J. Vac. Sci. Technol. A*, **5**, 838 (1987).
104. K.C. Pandey, *Phys. Rev. Lett.*, **47**, 1913 (1981).
105. Ph. Avouris, R. Wolkow, *Phys. Rev. B, Condens. Matter*, **39**, 5091 (1989).
106. G.S. Rohrer, D.A. Bonnell, *J. Am. Ceram. Soc.*, **73**, 3026 (1990).
107. F.-R.F. Fau, A.J. Bard, *J. Phys. Chem.*, **94**, 3761 (1990).
108. J.M. Perez, W. Rivera, C. Lin, et al., 'Scanning Tunneling Microscopy of the Structural and Electronic Properties of Chemical Vapour Deposited Diamond Films', in *Atomic Force Microscopy/Scanning Tunneling Microscopy*, ed. S.H. Cohen, Plenum Publishing, 203–210, New York, 1994.
109. A.J. Bard, F.R. Fan, J. Kwak, O. Leu, *Anal. Chem.*, **61**, 132 (1989).
110. P.K. Hansma, B. Drake, O. Marti, S.A.C. Gould, C.B. Prater, *Science*, **243**, 641 (1989).
111. D.M. Eigler, E.K. Schweizer, *Nature*, **344**, 534 (1990).
112. D.M. Eigler, C.P. Lutz, W.E. Rudge, *Nature*, **352**, 600 (1991).
113. M.A. McCord, R.F.W. Pease, *Appl. Phys. Lett.*, **50**, 569 (1987).
114. E.J. van Loenen, D. Dijkkamp, A.J. Hoeven, J.M. Lerissinck, J. Dieleman, *Appl. Phys. Lett.*, **55**, 1312 (1989).
115. C.R.K. Marrian, *Naval Research Reviews*, **VLIX**, 5 (1997).
116. C.R.K. Marrian, R.J. Colton, *Appl. Phys. Lett.*, **56**, 755 (1990).
117. J.A. Dagata, W. Tseng, J. Bennett, J. Schneir, H.H. Harary, *Appl. Phys. Lett.*, **59**, 3288 (1991).
118. T.R. Albrecht, M.M. Dovek, M.D. Kirk, C.A. Lang, C.F. Quate, D.P.E. Smith, *Appl. Phys. Lett.*, **55**, 1727.
119. R.S. Becker, G.S. Higashi, Y.J. Chabal, A.J. Becker, *Phys. Rev. Lett.*, **65**, 1917 (1990).
120. M.A. McCord, D.D. Awschalom, *Appl. Phys. Lett.*, **57**, 2153 (1990).
121. D. Mandler, A.J. Bard, *J. Electrochem. Soc.*, **137**, 2468 (1990).

Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Jory A. Yarmoff
University of California, Riverside, USA

1 Introduction **1**
2 Photoelectron Spectroscopy **2**
3 Surface Sensitivity **4**
4 Energy Distribution Curves **5**
5 Curve Fitting **6**
6 Chemical Shifts **7**
6.1 Clean Surfaces 8
6.2 Adsorption 9
6.3 Quantitative Analysis 9
7 Other Modes of Soft X-ray Photoelectron Spectroscopy **10**
7.1 Constant Final State Photoemission 10
7.2 Constant Initial State Photoemission 11
7.3 Resonant Photoemission 11
8 Other Information Obtained from Soft X-ray Photoelectron Spectroscopy **11**
8.1 Band Bending/Fermi Level Pinning 11
8.2 Work Function Measurements 12
8.3 Polarization Effects 12
9 Novel Applications of Soft X-ray Photoelectron Spectroscopy **13**
9.1 High-resolution Photoelectron Spectroscopy 13
9.2 Soft X-ray Photoelectron Spectroscopy Microscopy 13
9.3 Photoelectron Diffraction/ Holography 14
9.4 Spin-polarized Photoelectron Spectroscopy 15
10 Instrumentation for Soft X-ray Photoelectron Spectroscopy **16**
10.1 Storage Ring 16
10.2 Monochromator 17
10.3 Electron Energy Analyzer 17
Abbreviations and Acronyms **17**
Related Articles **18**
References **18**

X-ray photoelectron spectroscopy (XPS) is widely used to measure the composition and chemical states of species in the near-surface region of a solid material in ultrahigh vacuum (UHV). In XPS, the core-level, or inner-shell, electrons that are emitted via the photoelectric effect are measured. The spectra provide information related to the electronic structure and chemical composition of the near-surface region and to the bonding of adsorbates onto the surface. Because the mean free path for electrons traveling in solids is small, the spectra reflect the composition of the outermost few atomic layers. When synchrotron light is used as the excitation source, the technique is known as soft X-ray photoelectron spectroscopy (SXPS). There are many advantages of synchrotron radiation over conventional XPS. For example, the photon energy is tunable over a wide range, which allows for the enhancement of particular spectral features and for unique types of spectroscopy in which the photon energy is continuously varied. The spectral resolution is also greatly enhanced, which allows for the observation of features not otherwise discernible. Also, the brightness of the beam is much greater than that of conventional X-ray sources, thereby allowing for the detection of very small photoelectron signals. SXPS has been applied to a wide variety of materials, including metals, semiconductors, adsorbates on surfaces and nanoparticles. The information obtained from SXPS includes determining the core-level binding energies of individual surface atoms, the position of the Fermi level within the gap of a semiconductor, and the surface work function. Shifts of the core-level binding energy are observed for the outermost atoms on clean surfaces because they are in a different chemical environment than the atoms in the bulk. When an adsorbate attaches to a surface atom, there is a shift in binding energy that is related to the degree of charge transfer during bonding, i.e. the oxidation state. Analysis of the intensity of the shifted components provides the coverage and chemical state distribution of each adsorbate. Novel uses of SXPS include ultrahigh-resolution spectroscopy, SXPS microscopy, photoelectron diffraction and spin-polarized photoemission. The equipment utilized for SXPS includes an electron storage ring, an X-ray monochromator, and an electron spectrometer.

1 INTRODUCTION

Photoelectron spectroscopy (PES) is widely used for the chemical analysis of solid surfaces. When ultraviolet (UV) light or X-rays are incident upon a solid surface, electrons are emitted via the photoelectric effect. PES involves measurements of the kinetic energy distribution of the emitted electrons. If a monochromatic light source is employed, then the kinetic energy distribution reflects the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

filled density-of-states (DOS) of the material. The DOS contains information related to the electronic structure and chemical composition of the near-surface region and to the bonding of adsorbates. Because the mean free path for electrons traveling in solids is fairly short, of the order of atomic dimensions,[1] a PES spectrum reflects the DOS of the outermost few atomic layers of the solid.

When PES is used to measure the core-level, or inner-shell, electrons using a conventional X-ray tube, it is known as XPS or electron spectroscopy for chemical analysis (ESCA). The latter term has been coined to reflect the fact that chemical oxidation states are routinely obtained with this technique. The X-rays commonly employed for XPS are the Mg Kα (1253.6 eV) or the Al Kα (1486.6 eV) lines. For general information on XPS, see the article **X-ray Photoelectron Spectroscopy in Analysis of Surfaces** in this encyclopedia.

More recently, synchrotron radiation light sources have been developed to provide high-resolution, high-intensity and tunable sources of electromagnetic radiation.[2,3] PES performed with synchrotron radiation has significant advantages over conventional XPS. One of the advantages is the ability to use smaller photon energies (longer wavelengths). These energies correspond to the soft X-ray and vacuum ultraviolet (VUV) regions of the electromagnetic spectrum. Hence, SXPS is often used to denote XPS performed with synchrotron radiation. In addition, the tunability of synchrotron radiation allows for types of measurements not possible with a fixed energy X-ray source.

Some of the major advantages in employing synchrotron radiation for XPS are:

- The photon energy is tunable over a wide range (typically 10–1000 eV for SXPS). This allows the excitation energy to be adjusted in order to enhance a particular spectral feature, as well as for additional types of spectroscopy in which the photon energy is continuously varied.
- The spectral resolution is greatly enhanced over that of a conventional X-ray source. This allows for the identification and quantification of spectral features not otherwise discernible.
- The brightness of the beam can be much greater than that of conventional X-ray sources, thereby allowing for the detection of very small photoelectron signals. For example, the photoelectrons emitted from a microscopic region of a surface can be detected, thereby enabling chemically sensitive microscopy. The brightness available with synchrotron radiation has been further increased with the recent advent of third-generation light sources, which provide much more intense radiation.
- Synchrotron radiation is naturally polarized, which allows for the dipole selection rules to be exploited for spectroscopy. For example, the angular dependence of the photoelectron signal on the incident beam direction can be used to determine the orientation of molecules adsorbed onto a surface.

Since much of the theory and analysis of SXPS data is the same as for XPS, except that the resolving power of SXPS is much greater, this chapter emphasizes those aspects of SXPS that are peculiar to synchrotron radiation. SXPS has been applied to a wide variety of materials, including metals, semiconductors, adsorbates on surfaces and nanoparticles. Note that many of the examples given in this chapter are from the author's own laboratory. These particular examples were chosen because of their usefulness in illustrating the technique, and are not intended to be representative of the variety of materials that can be studied with SXPS.

2 PHOTOELECTRON SPECTROSCOPY

In PES, a photon is absorbed by direct excitation of a bound electron to a final state above the vacuum level, E_{vac}. This process is illustrated in Figure 1, which shows a photon of fixed energy, $h\nu$, being absorbed by the system, and the subsequent emission of an electron from a bound atomic core level. If the electron is excited above E_{vac}, it is emitted from the surface where it can be measured by an electron spectrometer.

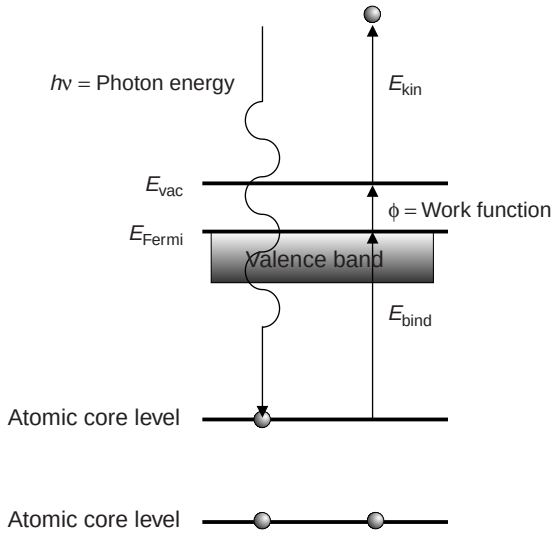

Figure 1 Schematic diagram illustrating the photoemission process. An electron bound with energy E_{bind} absorbs a photon of energy $h\nu$. The electron is emitted from the surface with kinetic energy E_{kin}.

The binding energy (E_{bind}) is the minimum energy required to remove an electron from a bound state. In other words, E_{bind} is the energy needed to excite the electron to the first empty state of the solid, which is just above the Fermi level. In order to escape into vacuum, the electron also needs sufficient energy to overcome the surface work function, ϕ. Thus, as illustrated in Figure 1 and shown in Equation (1), conservation of energy implies that

$$h\nu = E_{\text{bind}} + E_{\text{kin}} + \phi \tag{1}$$

The outcome of the photoelectric effect is to translate the filled DOS, represented by the valence band and core levels, to energies above the vacuum level. Figure 2 illustrates how the photoemission process creates a photoelectron spectrum by translating the entire distribution of filled states to positions above the vacuum level. If the probability for photoemission, i.e. the cross-section, were constant, then the measured photoelectron spectrum would be an exact representation of the filled DOS. Note that a real photoelectron spectrum would reside on top of a background of secondary electrons, which is not shown in Figure 2.

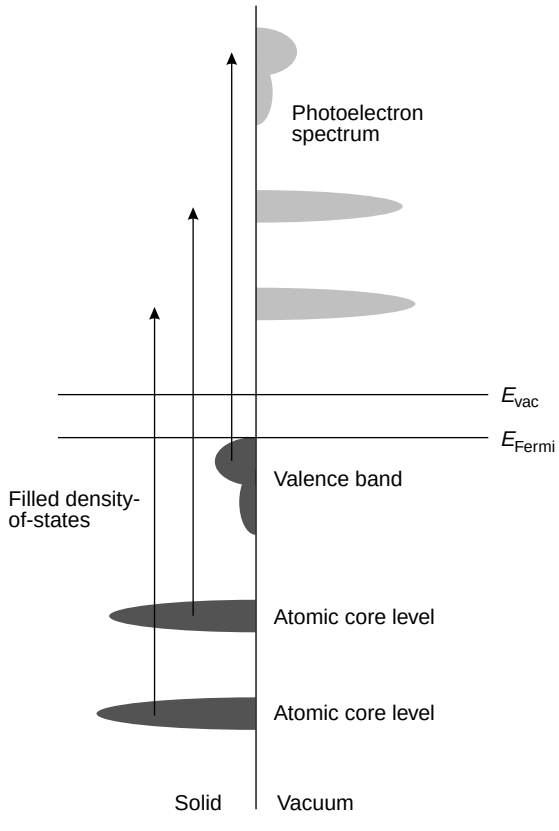

Figure 2 Schematic diagram illustrating how photoemission translates the filled DOS into a photoelectron spectrum. The length of the vertical arrows represents the photon energy.

Peaks are seen in PES spectra that correspond to the valence band and to each bound core-level electron in the material. The intensity of each core-level peak is proportional to the abundance of the corresponding element in the near-surface region, while the precise binding energy of each peak is a function of the chemical oxidation state and the local environment of the emitting atoms. The Perkin–Elmer *Handbook of X-ray Photoelectron Spectroscopy* contains sample spectra, collected with traditional XPS, of each element and a table of binding energies for certain compounds, and is useful as a reference.[4]

Note that core levels are often split into two resolvable components by the spin–orbit interactions. This spin–orbit splitting is an atomic effect, and is changed little by the presence of the surface or by the formation of chemical bonds. The result is that a doublet in a PES spectrum often represents a single atom in a single chemical environment. A spin–orbit doublet is described by two parameters: the spin–orbit splitting and the branching ratio. The spin–orbit splitting is the difference in binding energy between the two components. The branching ratio is the relative areas of the two components. To first approximation, the branching ratio will be that determined by statistics. For example, a p-level contains four electrons in the $p_{3/2}$ orbital and two electrons in the $p_{1/2}$ orbital. Thus, the $p_{3/2}$ and $p_{1/2}$ components have a statistical branching ratio of 2 : 1. Often, however, final state effects (discussed below) will slightly alter the measured branching ratio away from the statistical value.

The photoemission process involves the excitation of an electron from its bound initial state to an unbound final state. The matrix element for the transition determines the photoemission cross-section. This cross-section depends on the photon energy employed. There are tables, which are useful for quantitative analysis, that list the atomic cross-section as a function of the photon energy.[5] The presence of the surface has only a minor effect on the overall cross-section. Note, however, that in performing quantitative analysis one must also take into account how the sensitivity of the electron spectrometer varies as a function of the measured kinetic energy,[6] as well as the electron escape depth, which is discussed below.

The transition probability, w_{fi}, can be calculated from Fermi's Golden Rule[7,8] as shown in Equation (2)

$$w_{\text{fi}} = \frac{2\pi}{\hbar}|\langle\psi_{\text{f}}| \quad |\psi_{\text{i}}\rangle|^2\delta(E_{\text{f}} - E_{\text{i}} - \hbar w) \tag{2}$$

where ψ_{i} and ψ_{f} are the initial and final state wave functions. The simplest approximation in applying this rule is to assume one-electron wave functions for the initial and final states, and that the final state has a free electron with energy E_{kin}. Much more sophisticated

theoretical approaches have also been applied to this problem.

The fundamental shape of a single core-level photoelectron component is that of a Lorentzian that has been broadened by a Gaussian. The Lorentzian component is the natural shape of an atomic state, whose width derives from the Heisenberg uncertainty principle. The longer the lifetime of the excited state produced by photoemission, the more certain is the energy of that state. Thus, the Lorentzian width is often called the lifetime broadening. The Gaussian contribution to the width of a core-level feature comes from the instrumental resolution and from inhomogeneous broadening in the material. Either or both of these sources of broadening can be significant and often dominate an experimental spectrum. Instrumental broadening will affect each component in a spectrum in a similar manner, whereas inhomogeneous broadening could be different for different components. Inhomogeneous broadening includes phonon broadening and broadening due to atoms occupying a range of local environments.

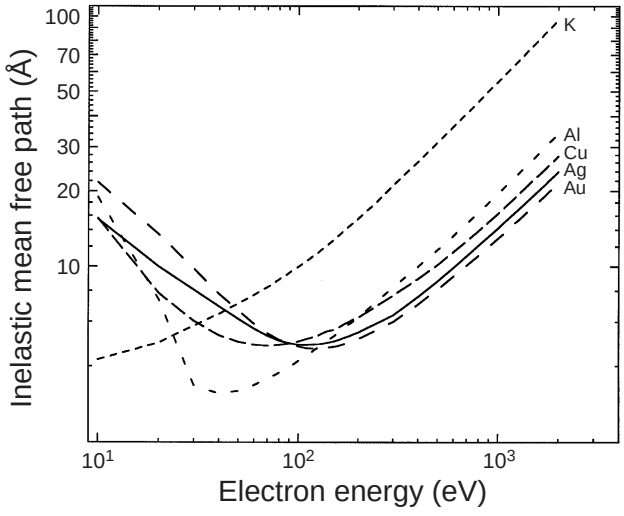

Figure 3 Calculations of the IMFP for electrons traveling in Al, Cu, Ag, Au and K. (Reproduced from Powell et al., *J. Electron Spectrosc. Relat. Phenom.*, **98/99**, 1–14 (1999) by permission of Elsevier Science.)

3 SURFACE SENSITIVITY

Electrons are extremely useful as surface probes because the distances that they travel within a solid are very short. This insures that any electrons created deep within a sample do not escape into vacuum. Techniques that rely on measurements of low-energy electrons emitted from solid surfaces therefore provide information from just the outermost few atomic layers. Because of this inherent surface sensitivity, the various electron spectroscopies are probably the most useful and popular techniques in surface science.

The reason that electrons interact strongly with solid materials is that they readily scatter from the various components of a solid. When electrons scatter inelastically, they lose energy to the solid so that the kinetic energy is reduced and the electrons no longer contribute to a spectroscopic feature. The most important inelastic scattering mechanism is the excitation of plasmons.[9] A plasmon is a collective excitation of the electrons in a material. Electrons can also scatter elastically, which may cause them to change direction so that they are no longer emitted. Elastic scattering can also, however, increase the photoelectron yield due to diffraction, which is discussed below.

The inelastic mean free path (IMFP) characterizes the average distance that an electron travels through a solid before it is annihilated by inelastic scattering. Early efforts were made to derive a "Universal Curve" for the IMFP as a function of the electron kinetic energy.[1] It was later realized, however, that the IMFP varies from material to material.[10] Figure 3 shows the IMFP plotted as a function of the electron kinetic energy for a few selected materials.[10] In general, the IMFP has a minimum that corresponds roughly to the plasmon energy. When the kinetic energy is too small, electrons can no longer excite plasmons. As the kinetic energy increases above the plasmon energy, the probability for excitation decreases. Thus, the electrons with the smallest mean free path for most materials are those with approximately 30 to 100 eV of kinetic energy. The specific location of the minimum and the detailed shape of the IMFP function differ from one material to the next. SXPS will generally be, however, much more surface sensitive than traditional XPS owing to the lower kinetic energy of the photoelectrons that are typically measured.

There have been various attempts at quantifying the penetration depth of electrons in order to use XPS for analytical purposes. The IMFP accounts for inelastic scattering, but electrons can also scatter elastically. Thus, the electron escape depth has been defined, which is the quantity that specifies how far an electron has traveled within a sample and includes losses by all possible mechanisms.[11] For quantitative analysis, however, what is actually needed is the mean distance below the surface from which the photoelectrons have originated. This depends on the angle of emission, roughly as the cosine of the angle, since electrons emitted at more grazing angles spend more time below the surface. The mean escape depth (MED) takes into account the emission angle in order to define the photoelectron yield in terms of the depth of origin, rather than the distance traveled through the solid.[10] Although the values will differ, the

general shape of the MED curve is similar to that of the IMFP shown in Figure 3. In general, values of the escape depth for 10–200 eV electrons, which are the important electrons for SXPS, range from 4 to 10 Å, which is of the order of the interlayer spacings of a solid.

Electron attenuation is most often modeled by assuming that the yield of electrons originating from a particular depth within the sample decreases exponentially with increasing depth, as shown in Equation (3)

$$\text{Number of electrons} = \exp\left(\frac{-d}{\Delta}\right) \tag{3}$$

where Δ is the MED for electrons in the material, and d is the distance below the surface from which the electron originated. This exponential attenuation model is often used for quantitative analysis, although it has some limitations. A measurement can be difficult to analyze, for example, if the MED changes as electrons travel from one region of an inhomogeneous sample to another.

A consideration of electron attenuation allows a determination of the depth distribution of elements within a sample by changing either the emission angle or the electron kinetic energy in order to vary Δ. Varying the emission angle, which is also routinely employed in conventional XPS measurements, changes the path length through which the photoelectrons travel within the solid. Thus, at more grazing emission angles, the surface sensitivity is increased. Then, a comparison of spectra obtained at a grazing angle to spectra collected at normal emission provides a measure of the depth distribution. At a synchrotron light source, the tunability of the radiation can be exploited to change the escape depth. For example, in order to maximize the surface sensitivity of the measurement, the photon energy can be selected to be approximately 20–50 eV above the binding energy of the desired core level. Or, in order to compare spectra that sampled different depths below the surface, spectra could be collected employing a surface-sensitive photon energy and one that produced electron kinetic energies that were not as surface sensitive. For example, a common method for obtaining a bulk-sensitive spectrum is to use kinetic energies well below the plasmon energy. In addition, if levels from different elements are being compared, the photon energies employed for the collection of each spectrum could be selected such that all of the electron kinetic energies are nearly equal. In this way, each spectrum would probe the same depth into the sample.

4 ENERGY DISTRIBUTION CURVES

The standard SXPS spectrum is collected as an energy distribution curve (EDC). For an EDC, the photon energy is fixed and the kinetic energy distribution of the emitted photoelectrons is measured. This is the same manner in which traditional XPS spectra are collected. EDC spectra can be used to probe either the valence band or core-level states.

Note that EDC spectra are often displayed with respect to the binding energy, rather than the electron kinetic energy, in order to reflect the properties of the material, rather than those of the spectrometer. Because spectroscopists are accustomed to viewing spectra plotted in terms of increasing kinetic energy, however, the convention has arisen to plot PES spectra in reverse, i.e. from high binding energy on the left to low binding energy on the right.

For valence band spectroscopy, photon energies in the range of 20 to 50 eV are usually employed, as this represents the region of maximum cross-section for valence band photoemission. Traditional ultraviolet photoelectron spectroscopy (UPS) employing an He discharge lamp is limited to the He I and He II lines at 21.2 and 40.8 eV. An advantage to the use of synchrotron radiation for valence band PES is that a wide range of photon energies is available. As the photon energy changes, the relative contributions to the spectra of various atomic and molecular orbitals change, which is a great aid in delineating the spectroscopy of a system.

Synchrotron radiation is also particularly useful for angle-resolved photoelectron spectroscopy (ARPES), also called angle-resolved ultraviolet photoelectron spectroscopy (ARUPS).[12–14] In this technique, measurements are made of the valence band photoelectrons emitted into a small angle as a function of the electron emission angle or photon energy. This allows for the simultaneous determination of the kinetic energy and momentum of the photoelectrons with respect to the two-dimensional surface Brillouin zone. From this information, the bulk and surface electronic band structure of a single crystal material can be experimentally determined via the procedure called ''band mapping''.

SXPS usually refers to core-level spectroscopy. Each core-level EDC contains peaks that correspond to emission from a particular inner-shell atomic level of an element in the near-surface region of the material. The intensity of each core-level peak is related to the abundance of that element in the near-surface region. The exact position of the peak, i.e. the binding energy, is related to the local chemical environment of the emitting atoms.

A single core-level peak often contains multiple components, in which the different components arise from atoms in inequivalent chemical configurations. Sometimes the individual components will be clearly resolved from each other. Often, however, a particular peak will be composed of a sum of unresolved components. These components

are then delineated from each other by a numerical curve-fitting procedure, as described below. In either case, the positions of the various core-level components are used to identify the chemical species present on the surface, and the relative intensities are then used to determine the distributions of the various chemical states.

5 CURVE FITTING

In order to interpret high-resolution SXPS data, it is often necessary to deconvolve the spectra into their individual core-level components. This allows for a determination of how many separate states exist for each element in the surface region, the binding energy associated with each of these states, and their relative abundance. Care must be taken in fitting, however, to make sure that the result is physically reasonable. It is easy to get a good fit by introducing extra components into an analysis. Usually, the least number of components that are required to provide a good fit should be used.

Before a spectrum can be fitted, the secondary electron background signal must be removed from the raw data. The shape of the secondary electron distribution depends on the details of the electronic structure of the material, as it results from photoelectrons that have lost energy by inelastic scattering from various excitations in the solid. This produces a cascade of secondary electrons whose intensity increases towards lower kinetic energies. The background underneath any particular photoelectron feature consists of secondary electrons that were produced both extrinsically and intrinsically. The extrinsic secondary electrons originated as higher energy photoelectrons, i.e. from the valence band and from lower-lying core levels. The intrinsic secondary electrons originate from the photoelectron peak itself. The process of background subtraction is illustrated in Figure 4, which shows SXPS data collected from Si(100) reacted with chlorine gas.[15]

The background due to the extrinsic secondary electrons can be determined experimentally. An experimental background spectrum is collected by measuring the electron yield over the same kinetic energy region as was used for the SXPS data, but employing a lower photon energy. In this manner, the direct photoemission peaks are moved to kinetic energies below the region over which the spectrum is collected. This gives a good experimental measurement of the shape of the background due to higher energy photoelectrons. As an example, Figure 4(a) shows a raw photoelectron spectrum, collected with $h\nu = 130\,\text{eV}$, along with a measured background collected with $h\nu = 120\,\text{eV}$. Figure 4(b) shows the spectrum after the experimental background was subtracted. Note that the experimental background was approximated by a smooth polynomial prior to subtraction so as not to introduce additional noise into the data.

After removal of the experimental background, there is still a background arising from the intrinsic secondary electrons. This is seen in the data of Figure 4(b), where a clear background signal is visible at kinetic energies below the peaks, while there is no background above the peaks. This intrinsic background signal must be removed numerically. The most common method used to numerically model the secondary electron background in SXPS is that devised by Shirley.[16] Note that often a Shirley-type background is calculated directly from the raw data, i.e. removal of an experimental background is not essential. In the Shirley method, the region on both sides of the peak is considered separately from the region underneath the peak. Away from the peak, the background signal can be fitted to a polynomial. Underneath the peak, the yield of intrinsic secondary electrons is considered to be proportional to the integral of the peak. The implicit assumption here is that the loss of photoelectrons is constant throughout the spectrum. Figure 4(b) shows data along with a Shirley-type background calculated from the data, and Figure 4(c) shows the data after the background has been removed.

After background removal, the data are composed of a sum of components that correspond to the different chemical states present in the surface region. A single component, however, is often composed of a spin–orbit doublet. In Figure 4(c), for example, it might initially appear that there are three distinct states of Si. The spin–orbit splitting can be numerically removed, however, which simplifies the analysis. In order to remove the spin–orbit splitting, the background must first be

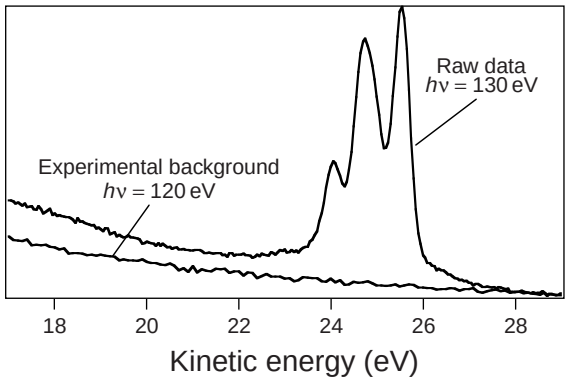

Figure 4 SXPS spectra collected from Cl_2 gas adsorbed on Si(100), illustrating background subtraction and spin–orbit deconvolution. (a) Raw data collected with $h\nu = 130\,\text{eV}$ are shown along with a background collected with $h\nu = 120\,\text{eV}$. (b) Data are shown following subtraction of the experimental background along with a calculated background. (c) Data are shown following subtraction of all backgrounds. (d) Data are shown following removal of the $2p_{1/2}$ spin–orbit component.

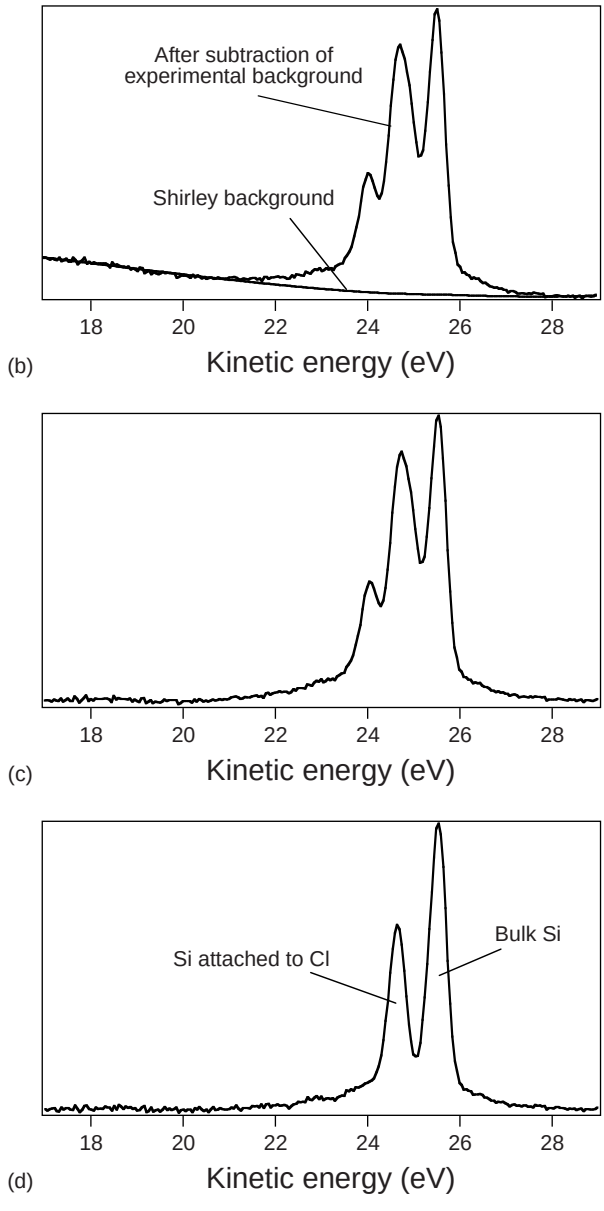

Figure 4 (*Continued*)

subtracted, and the spin–orbit parameters must be known and assumed to be the same for each component.[17] Figure 4(d) shows the result of spin–orbit deconvolution of the data in Figure 4(c). After removal of the spin–orbit splitting, it is seen that there are actually only two major components: one that represents Si in the substrate and the other that represents surface Si atoms that are attached to Cl.[15]

After background subtraction, the data are numerically fitted to a sum of Gaussian-broadened Lorentzian lineshapes using a least-squares procedure. The fit can be performed either with or without first removing the spin–orbit splitting.

6 CHEMICAL SHIFTS

The binding energy of a core-level component depends on the bonding configuration and local environment of the emitting atom. The nominal energy of the core-level state is determined by the atomic orbital, but the binding energy is modified in forming a molecule or solid. In addition, the exact binding energy associated with a particular arrangement within a solid depends on many other factors. The amount that the binding energy changes when an atom goes from one chemical state or environment to another is referred to as the chemical shift.

When the atoms at the outermost surface differ in binding energy from the bulk, i.e. substrate, atoms, the difference in binding energy is called a surface core-level shift (SCLS). A surface shift could arise simply because an atom is located at the surface, where its chemical environment differs from that of the bulk, or due to the chemical bonding of an adsorbate to the surface. Note that often only the chemical shifts between the substrate and surface atoms are reported, rather than the absolute binding energies.

Photoemission occurs by excitation of a ground state electron from its initial state to an excited final state. The final state produced by the photoemission process lacks one electron as compared to the initial state. Thus, the binding energy that is measured is actually the difference between the neutral ground state system and the ionized final state. Note that there is no direct way in which to experimentally measure the binding energy of an electron as it exists in the neutral ground state.

Chemical shifts result from a combination of initial state and final state effects, as changes to the chemical environment of an atom alter both the ground state and the excited state energies.[18] Initial state effects result from phenomena such as charge transfer, Madelung energy and rehybridization. The energy of the final state depends on the interaction of the photoelectron with the ionized system. Thus, final state effects contribute to chemical shifts via alterations of the screening that photoelectrons experience at the surface.

To first approximation, changes in the initial state configuration can be considered to be the major contributor to chemical shifts. The largest factor that contributes to determining the binding energy of the initial state is the chemical oxidation state of the atom. As the oxidation state increases, the density of valence electrons that surround an atom decreases due to charge transfer away from the atom. The remaining core electrons then become more tightly bound to the nucleus since fewer valence electrons are there to screen the interaction. Thus, the more positive the oxidation state, the higher the binding energy will be. Note that in many XPS and

SXPS studies, it is only the oxidation states of the surface atoms that are of interest.

There has been theoretical work employing a Born–Haber cycle in order to calculate core-level shifts and to elucidate the relative importance of initial and final state effects.[19–21] Since the charge seen by the photoelectron is a combination of the nuclear charge plus the core hole, the final state produced by photoemission from an atom of atomic number Z can be approximated by a neutral atom of atomic number $Z + 1$. Then, for example, a chemical shift due to an adsorbate can be related to the difference in bond energy of that adsorbate to elements Z and $Z + 1$.

The theory is, however, not sufficiently developed that it can be used to reliably calculate binding energies. Thus, for most applications, binding energies are determined empirically regardless of the nature of the chemical shifts. In general, the amount that initial or final state effects contribute to a chemical shift is unimportant in performing spectroscopy for chemical analysis.

Note that in most SXPS studies, the photoelectric cross-section of a particular component is assumed to be unchanged when the binding energy of that component is altered by chemical bonding or other effects. There have been some reports, however, in which the cross-section for photoemission has been shown to be affected by changes in the oxidation state due to chemical bonding.[22] These effects are likely due to changes in the final state DOS that occurs with increasing oxidation state. An increase in the final state DOS will increase the probability for photoemission, as can be seen from Equation (2). Shape resonances, which are enhancements of the final state DOS, are well known from X-ray absorption studies, and are often larger for atoms in higher oxidation states. Thus, it is not surprising that photoemission cross-sections are sometimes affected by the chemical state of the emitting atom. Also, such a nonconstant final state DOS can act to slightly alter the spin–orbit branching ratio measured in an EDC spectrum. Although these effects are usually not too severe, they can be avoided by using constant final state (CFS) photoemission, as described below.

6.1 Clean Surfaces

The outermost atoms on a clean surface are chemically and structurally different from the bulk atoms of a material. Thus, there are often clearly resolvable SCLS components in spectra collected from clean surfaces.[23] In a bulk-terminated or relaxed surface, the SCLS component would have an intensity corresponding to one monolayer of atoms. In a reconstructed surface, however, there can be a number of SCLS components, each corresponding to atoms in different structural positions.

Figure 5 shows an SXPS spectrum collected from clean W(110), which has a bulk-terminated surface structure.[24]

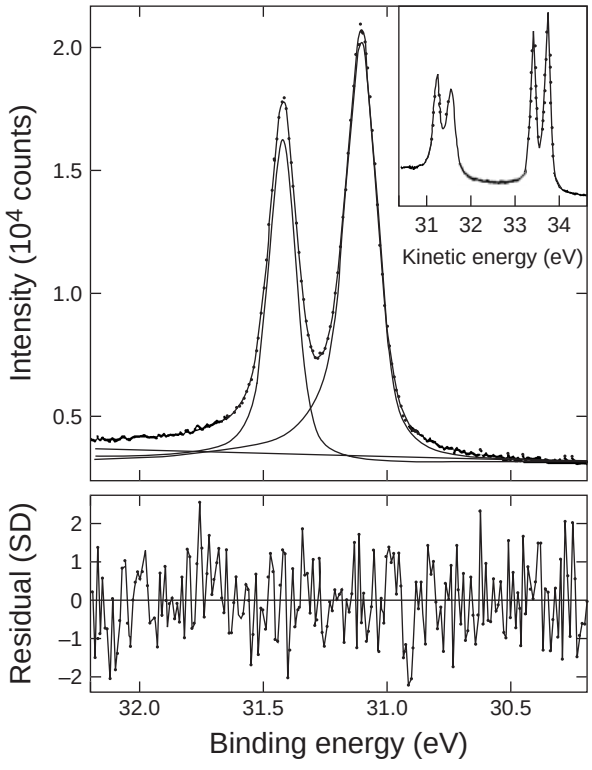

Figure 5 W $4f_{7/2}$ spectrum collected from a clean W(110) surface.[24] The dots show the data, while the solid lines show a numerical fit to the data. The lower panel shows residuals, which is the difference between the data and the fit. The inset shows the entire W 4f spectrum, which encompasses both the $4f_{7/2}$ and $4f_{5/2}$ levels. (Reproduced from Citrin et al., *Phys. Rev. Lett.*, **63**, 1976 (1989) by permission of the American Physical Society.)

The inset shows the entire W 4f spectrum, which includes the $4f_{7/2}$ and $4f_{5/2}$ spin–orbit-split levels, which are completely resolved from each other. It is seen that each level is composed of two features, the one at lower binding energy (rightmost) being the bulk component and the higher binding energy feature (leftmost) being the SCLS component. The main figure shows a blow-up of the $4f_{7/2}$ level along with a numerical fit. The SCLS is nearly as intense as the bulk component in this spectrum, which attests to the surface sensitivity of this measurement. Because the surface atoms are in a different environment from the bulk atoms, their binding energy is lower by 0.32 eV, even though there is no change in the formal oxidation state.

An example for a reconstructed semiconductor surface is given in Figure 6, which shows an SXPS spectrum collected from clean Si(100)-2×1 along with a numerical fit.[17] In this spectrum, there is a bulk component and two SCLS components. One SCLS is shifted to higher binding energy and the other to lower binding energy. The nature of these shifts can be understood in simple

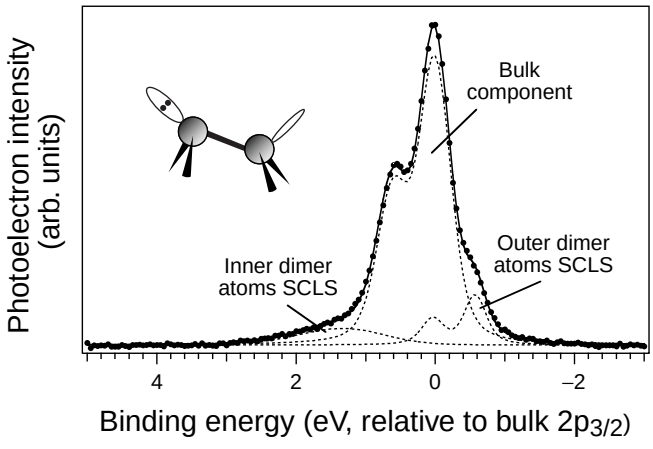

Figure 6 Si 2p SXPS spectrum collected from clean Si(100)-2×1 with a photon energy of 130 eV. The data, following background subtraction, are shown as dots. The solid line shows the result of a numerical fit to the data, while the dashed lines show the individual components of the fit. Each individual chemical component is split into the Si $2p_{3/2}$ and $2p_{1/2}$ contributions by the spin–orbit interaction. The inset shows an illustration of an asymmetric surface dimer that is believed to terminate this surface.

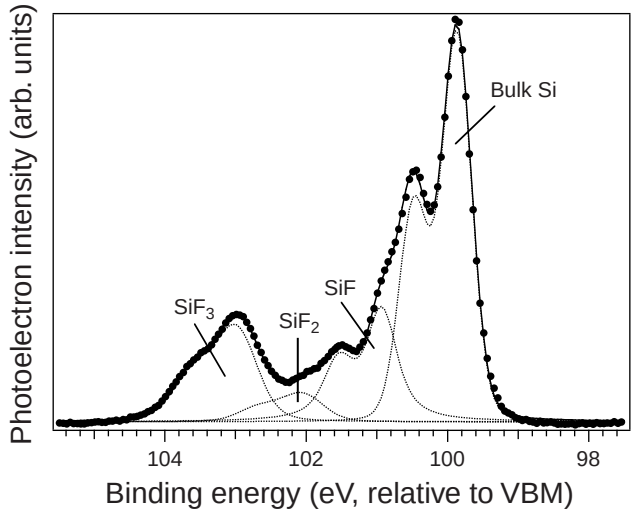

Figure 7 Si 2p SXPS spectrum collected with a photon energy of 130 eV from Si(111) following reaction with 50 L of XeF_2. The raw data are shown as dots. The data were fit numerically. Each chemically distinct component is shown as a dotted line, while the sum of all of the fit components is shown as a solid line. Each individual chemical component is split into the Si $2p_{3/2}$ and $2p_{1/2}$ contributions by the spin–orbit interaction.

terms by consideration of the geometry at the Si(100)-2×1 surface. The Si surface atoms form asymmetric dimers, as shown in the inset, in which the outer dimer atom has a filled orbital and is nearly sp^3 hybridized, while the other atom has an empty orbital and is nearly sp^2 hybridized. Because the outer atom has more electron density associated with it, the binding energy of the core electrons is decreased. Hence, it is associated with the low binding energy component. The inner dimer atom has a lower electron density surrounding it, and hence it is associated with the high binding energy SCLS component.

6.2 Adsorption

When adsorbates attach to surface atoms, chemical shifts are often observed in the SXPS spectra.[25] The usual cause for this is the transfer of charge to (from) electronegative (electropositive) adsorbates. In some cases, the effects of adsorption are to simply remove the SCLS components associated with the clean surface. The identification, concentration and bonding configuration of surface adsorbates is obtained from analysis of SXPS spectra.

As an example, Figure 7 shows an SXPS spectrum that was collected from a silicon surface following reaction with XeF_2 (a source of atomic fluorine).[26] The photon energy was chosen to be 130 eV so that the electron kinetic energy would be approximately 25 eV, i.e. near the minimum escape depth. The raw data, following background subtraction, are shown along with a numerical fit to the data. There are four resolvable spin–orbit doublets in the spectrum, which correspond to the substrate Si and to surface Si atoms that are attached to one, two and three fluorine atoms, i.e. Si, SiF, SiF_2 and SiF_3. It is seen that the bonding of each fluorine atom causes an approximately 1 eV shift to the binding energy. In this case, the main contribution to the chemical shift is the change in oxidation state. As additional fluorine atoms attach to each Si atom, the oxidation state is increased and charge is removed from the Si atom. The more charge that is removed, the more the binding energy increases. The resolution and surface sensitivity afforded by synchrotron radiation were necessary in order to resolve the SiF_x features. Results such as these are extremely useful in determining the pathways of surface chemical reactions.

The example shown in Figure 8 shows how SXPS can be used to determine the binding site on a binary material. Figure 8(b) and (d) show In and As spectra collected from the clean InAs(001)-c(8×2)/(4×2) surface, in which the clean surface SCLS components S_1 and S_2 can be seen. In Figure 8(a) and (c) the spectra collected following iodine adsorption are shown. It can be clearly seen that iodine attaches solely to the surface In atoms.[27]

6.3 Quantitative Analysis

In order to calculate the coverage of an adsorbate from the ratio of two SXPS core-level peaks, the photoemission cross-section, the electron escape depths and the incident photon flux must all be known. The cross-section can be found from a table.[5] The escape depth can be obtained

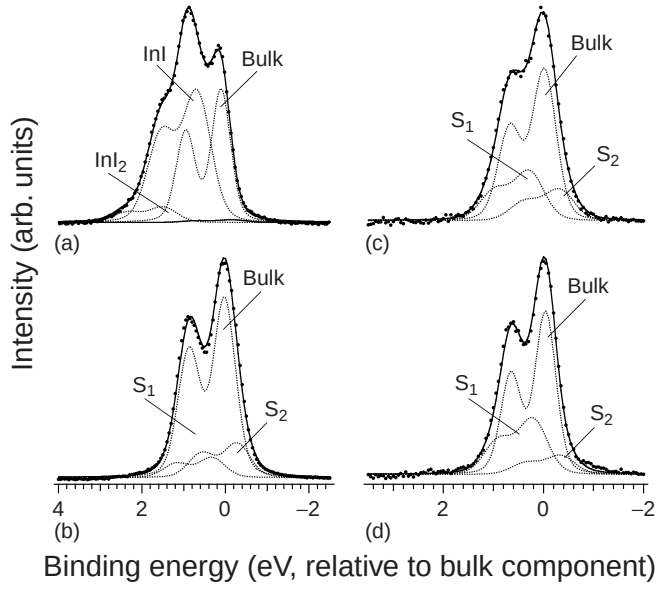

Figure 8 SXPS spectra collected from clean InAs(001) and from InAs(001) following reaction with I_2. (b) and (d) show In 4d and As 3d spectra, respectively, from the clean surface, while (a) and (c) show In 4d and As 3d spectra, respectively, collected from the surface saturated by iodine. The data, following background subtraction, are shown as dots. The solid line shows the result of a numerical fit to the data, while the dotted lines show the individual components of the fit.

from calculations[10] or calibrated from a known structure. The photon flux is usually measured during a run. This is important, as the flux of radiation from a synchrotron light source decays over time, and is also strongly dependent on the photon energy.

Quantitative analysis involves using this information along with the assumption of exponential attenuation, i.e. Equation (3), and a model structure. For example, if it is assumed that an adsorbate lies completely on top of the surface, then the SXPS signal from the adsorbate would not be attenuated by inelastic scattering. On the other hand, if the adsorbate formed multilayers or was intermixed with the substrate, then the SXPS signal would not fully represent the total number of adsorbate atoms, as some of those photoelectrons would have to travel through solid material in order to escape the surface. Thus, the calculation of coverages from SXPS areas depends on a model of the distribution of atoms in the surface region.

One of the best ways to perform quantitative analysis is to use two components from within a single spectrum.[22] In this manner, the cross-section, photon flux and differences in escape length will all cancel, as they will be the same for both components. For example, suppose that one is interested in determining the surface coverage of an adsorbate that causes a measurable chemical shift to a substrate core level, such as the spectrum of Figure 7. The intensity ratio between each shifted core-level component and the bulk component can then be used to calculate the coverage. The only information required in this case is the escape depth. In fact, the escape depth can be calibrated from a sample with a known coverage. A convenient method is to use a clean surface SCLS from a known structure. For example, in the Si(100) spectrum shown in Figure 6, each SCLS represents 0.5 monolayers of surface Si atoms.

7 OTHER MODES OF SOFT X-RAY PHOTOELECTRON SPECTROSCOPY

Because of the tunability of synchrotron radiation, there are other ways in which SXPS and related data can be obtained. The following methods can only be performed with synchrotron radiation, as it is necessary to continuously vary the photon energy.

7.1 Constant Final State Photoemission

Photoelectron spectra can be collected by setting the electron energy analyzer to measure electrons at a fixed kinetic energy and then varying the photon energy.[2] When the photon energy is such that Equation (1) holds for a particular filled state, a peak will appear in the spectrum. In this manner, the final state that is involved in the transition is fixed over the entire spectrum, while the initial state is varied. Thus, this method of collecting spectra is called CFS spectroscopy. The analysis of CFS data is somewhat simpler than EDC, as both the initial and final states vary across an EDC spectrum. Another advantage to the CFS technique is that cross-section changes due to final-state shape resonances do not influence the lineshapes within a spectrum. In addition, the electron escape depth remains constant across the spectrum. CFS photoemission has not been commonly employed despite these advantages, however, as most researchers are more familiar with EDC spectra. Also, it takes longer to collect a CFS spectrum, as the monochromator needs to be adjusted continuously, and many software packages do not have this as an option.

The same experimental conditions used for CFS data collection are also used for X-ray absorption spectroscopy. In X-ray absorption, the probability for the absorption of photons is measured as the photon energy is varied. For surface work, the yield of secondary electrons is monitored, which is assumed to be proportional to the absorption probability. If all of the emitted secondary electrons are collected, such a spectrum is labeled as total electron yield (TEY). Often, however, only a portion of the secondary electron distribution is monitored, i.e. the

partial electron yield (PEY) is collected. A PEY spectrum might be collected, for example, by monitoring the Auger electrons emitted from a particular atom or the yield at a surface sensitive kinetic energy, such as 25 eV. A PEY spectrum can be collected using the same instrumentation employed for SXPS.

X-ray absorption spectroscopy is commonly used to measure the electronic and geometric structures of surfaces by varying the photon energy and collecting a TEY or PEY spectrum. When the energy corresponding to a transition from a filled core level to the Fermi energy is reached, the probability for absorption greatly increases and a core-level edge is observed in the spectrum. When probing the region close to an edge, the technique is referred to as near edge X-ray absorption fine structure (NEXAFS) or X-ray absorption near edge spectroscopy (XANES).[28] NEXAFS provides a good measure of the unoccupied DOS, and is particularly useful for studies of molecular adsorbates. When probing the absorption of a surface over a wide range of photon energies above a core-level edge, the technique is called surface extended X-ray absorption fine structure (SEXAFS). SEXAFS, which is the surface analog to the popular EXAFS technique, provides a measure of the nearest neighbor bond lengths and coordination for atoms at the surface.

7.2 Constant Initial State Photoemission

In constant initial state (CIS) spectroscopy, the photon energy and electron kinetic energy are ramped together in such a way that $E_{\text{kin}} - h\nu - \phi = E_{\text{bind}}$ is kept constant.[2] In this way, the intensity of a single photoelectron feature is measured as the final state is varied. For valence band spectroscopy, this is useful for mapping out the empty DOS of the conduction band in order to compare with electronic structure calculations. By combining EDC and CIS spectra, a complete picture of the occupied and unoccupied DOS can be obtained. For core-level spectra, CIS spectroscopy is one method that can be used for mapping out changes to the photoelectron yield due to diffraction, as discussed below.

7.3 Resonant Photoemission

Valence band EDC spectra contain contributions from the various atomic and molecular orbitals that comprise the electronic structure of a material. As the photon energy is changed, the relative contributions of each of these levels to the spectra change. Monitoring these changes is one manner in which the details of the electronic structure of the surface region can be elucidated.

A more dramatic enhancement of certain valence band features occurs, however, when the photon energy employed also corresponds to the excitation energy of an inner-shell electron, i.e. at a core-level edge. Such an enhancement of a particular valence band feature is known as resonant photoemission.[29] When the photon energy is just above a core-level edge, the incident beam induces direct photoemission from the valence band while at the same time it creates core holes. The core holes will be filled by an Auger process, in which one valence electron falls down to fill the hole, while a second valence electron is emitted from the surface in order to satisfy conservation of energy. Resonant photoemission occurs because the Auger electron that is released will have the same kinetic energy as the valence band photoelectrons affiliated with that element. This is because the valence electrons that have the highest probability for making these transitions are the ones associated with the atom whose inner-shell orbital is being filled. Thus, element-specific valence band features are enhanced at energies corresponding to the core-level edges of that element.

Resonant photoemission has been applied to the determination of the electronic structure of a number of materials, such as metals,[30] oxides,[31] high temperature superconductors[32,33] and other materials that display correlations in the behavior of their electrons.[29]

8 OTHER INFORMATION OBTAINED FROM SOFT X-RAY PHOTOELECTRON SPECTROSCOPY

In addition to concentrations and chemical shifts, there is other information that can be obtained by analysis of SXPS data.

8.1 Band Bending/Fermi Level Pinning

For a semiconductor or insulating sample, the Fermi level can be positioned anywhere within the band gap, i.e. between the valence band maximum (VBM) and the conduction band minimum (CBM). The location of the Fermi level within the band gap is known as the pinning position, and the precise value is of critical importance in electronic devices as it determines the height of the Schottky barrier at an interface.[34] Because of the importance of Fermi level pinning, there have been many studies in which changes in the pinning position have been monitored as a function of metal overlayer growth.

The Fermi level is pinned by the existence of partially filled states in the band gap, such as surface electronic states, defect states, or states induced by an adsorbate or metal overlayer. Note that some materials, such as clean III–V semiconductor (110) surfaces, are unpinned, i.e. they have no states in the gap. For an unpinned

surface, the Fermi level is located near the CBM for n-type material and at the VBM for p-type. When an adsorbate attaches to the surface of one of these unpinned materials, states form that act to pin the Fermi level. Thus, these surfaces are ideal for studies of Fermi level pinning mechanisms.

When the Fermi level is pinned at the surface, all of the electronic energy levels bend as they move deeper into the sample.[17] This is illustrated in Figure 9. The slope of this band bending is usually shallow enough so that SXPS measures only the photoelectrons that originate from the outermost surface, i.e. the region in which the Fermi level is pinned. Note that if the band bending were steep with respect to the electron escape depth, the SXPS peaks would be broadened. The measured kinetic energies of all of the photoelectrons are affected by the pinning position. When the pinning position is altered, the bands bend and the kinetic energies of the photoelectron peaks move up or down accordingly. The movement of a sharp SXPS peak is therefore used to monitor the position of the Fermi level as a surface is altered. There is a caveat to using SXPS to measure the pinning position, however, as the photon beam itself can induce movement of the Fermi level due to the surface photovoltage (SPV) effect.[35,36]

8.2 Work Function Measurements

Electron spectrometers measure kinetic energy with respect to their own vacuum level. This is illustrated in Figure 10, which shows a sample with a slightly larger work function than that of the analyzer. The thick vertical arrow indicates the lowest kinetic energy electron that can be emitted from the sample, and how it would appear to have a finite kinetic energy when measured by the analyzer. This is essentially another way of stating

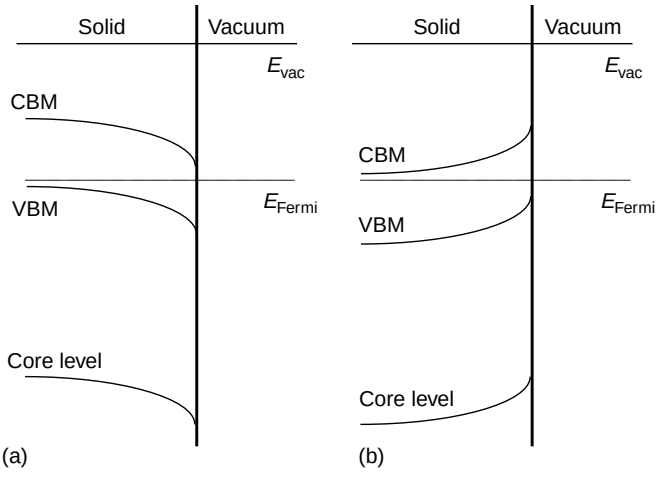

Figure 9 Energy diagram used to illustrate band bending for (a) p-type and (b) n-type semiconductors in which the Fermi level is pinned at the surface near mid-gap.

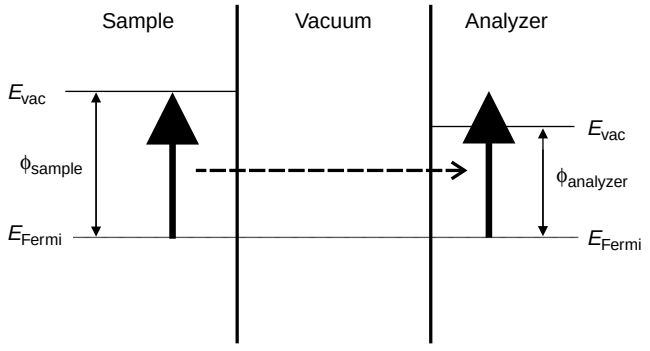

Figure 10 Energy diagram used to illustrate a work function measurement. The Fermi levels are assumed to be the same for the sample and analyzer, while the analyzer work function is assumed to be smaller than that of the sample. The thick vertical line represents a zero kinetic energy electron that is leaving the surface and arriving at the detector.

that there is an electric field between the sample and analyzer due to the work function difference, and this field accelerates the electrons into the detector. Thus, a measurement of the position of the low energy cutoff of the secondary electron distribution provides a measure of the surface work function. Note that it is common to apply a small (~ -5 to 10 V) bias to the sample so that the cutoff of secondary electrons is always well above the analyzer vacuum level. The position of this secondary electron cutoff shifts if the sample work function is changed, for example by the adsorption of alkali atoms.[37] Monitoring shifts of the cutoff energy does not provide an absolute measure of the surface work function, but it can be calibrated by comparison to a cutoff measured from a known surface. Changes to the work function that occur as a surface is altered are important in understanding many kinds of surface phenomena.

8.3 Polarization Effects

The light from a synchrotron radiation source is linearly polarized in a direction perpendicular to the beam.[38] A particular photoelectron transition can be either allowed or forbidden, depending on the projection of the matrix element for the transition, i.e. Equation (2), onto the polarization vector.[2] Thus, for example, the relative angle between the polarization direction and the direction of a chemical bond on a surface can affect the intensity of a given photoelectron feature. PES can then be used to determine the orientation of adsorbates on surfaces by monitoring a valence band signal as a function of angle. Note that this cannot be used for core-level studies since diffraction effects (see below) will dominate the angular distributions. However, the polarization of synchrotron radiation is often exploited in NEXAFS for determining the bond angle of a molecular adsorbate.[28]

The polarization directions are usually defined as s-polarized and p-polarized. In the s-polarized configuration, the polarization vector of the light is parallel to the sample, while for p-polarization it is perpendicular to the sample. Thus, for s-polarization the beam is incident normal to the surface, while it is incident at a near-grazing angle for p-polarization. Note that it is possible to get a completely s-polarized configuration, but it is not possible to be 100% p-polarized since the incident beam cannot be aligned perfectly along the plane of the surface and still induce photoelectron emission.

9 NOVEL APPLICATIONS OF SOFT X-RAY PHOTOELECTRON SPECTROSCOPY

In addition to collecting standard SXPS spectra for chemical analysis, there are many other unique uses of this technique. In addition, the advent of insertion devices and third-generation synchrotron light sources, which are described below, has furthered the development of some of these novel applications by providing more intense radiation sources. Some of the latest developments are described here.

9.1 High-resolution Photoelectron Spectroscopy

The high brightness available from third-generation light sources has improved the performance of extremely high-resolution SXPS.[39] The high brightness is necessary so that when the spectral width of a beam is made extremely narrow, there is still a sufficient photon flux remaining to be able to produce a measurable signal. In addition, the sample is usually cooled to reduce the phonon broadening in order to obtain ultrahigh-resolution SXPS spectra from solids.

For example, ultrahigh-resolution SXPS spectra collected from the Si(111)-7×7 surface enabled the delineation of at least six SCLS components, instead of the usual two components measured under more conventional conditions.[40] An example is given in Figure 11, which shows a comparison of spectra collected at liquid nitrogen and room temperatures. In these spectra, the Gaussian broadening was reduced to the point that the spectra actually appear by eye to have a Lorentzian shape. Ultrahigh-resolution SXPS has also been applied to measurements of the electronic structure of quasicrystals.[41] In this work, the Fermi edge was measured as a function of surface temperature using a 5-meV resolution. This was necessary in order to be able to compare the results with theoretical predictions, as the edge is extremely sharp.

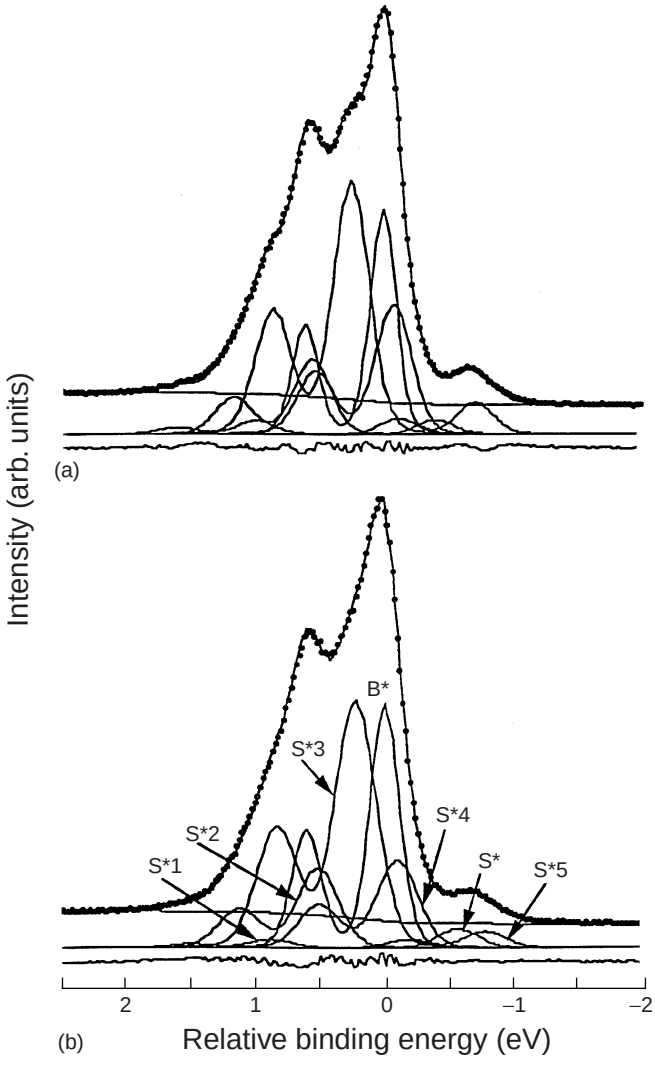

Figure 11 Ultrahigh-resolution SXPS spectra collected at (a) 100 K and (b) 300 K from Si(111)-7×7 with a photon energy of 140 eV, shown along with a numerical fit to the data. (Reproduced from Le Lay et al., *Phys. Rev. B*, **50**, 14 277 (1994) by permission of the American Physical Society.)

9.2 Soft X-ray Photoelectron Spectroscopy Microscopy

One of the more recent advances in SXPS is the development of photoelectron microscopy.[42] SXPS microscopy allows for the collection of chemically resolved images. For example, the spatial variations of the photoelectron yield of each element in a material can be independently collected or an element- or oxidation-state-specific image can be obtained for systems with a resolvable core-level shift. This technique has many technological and environmental applications. For example, the chemical makeup of micromechanical and microelectronic devices can be viewed directly on the scale of the device dimensions.

There are two modes in which spatially resolved SXPS is performed. In ''spectromicroscopy'', electrostatic lenses are employed to image the spatial distribution of the emitted photoelectrons. In ''microspectroscopy'', the incident X-ray beam is focused onto a small spot and an image is acquired by moving the sample. High brightness is an integral part of the development of SXPS microscopy, as it is needed in order to provide sufficient photoelectron signal from a microscopic region. Thus, most of this work requires an insertion device beamline. The limit to the spatial resolution is currently of the order of 1000 Å, but is expected to reach 100 Å in the near future.

Figure 12 shows an example of SXPS microscopy images collected from a naturally occurring ilmenite sample.[43] Figure 12(a) shows spectra that were collected from two different regions of the sample surface. One region shows an enhancement of Fe, while the other shows an enhancement of Ti. Figure 12(b) and (c) shows micrographs collected by monitoring either Fe or Ti photoelectrons. These images clearly show the Fe- and Ti-rich domains on the surface.

9.3 Photoelectron Diffraction/Holography

X-ray photoelectron diffraction (XPD) is routinely used to obtain the atomic structure of the surface region of a crystalline sample.[44,45] When electrons of a well-defined energy are created at a particular atomic site, such as in SXPS, the emitted electrons will interact with other atoms in the crystal structure prior to leaving the surface. The largest effect is ''forward scattering'', in which the intensity of an electron wave emitted from one atom is enhanced when it passes through another atom. Thus, the angular distribution of the emitted electron intensity provides a ''map'' of the surface crystal structure. Forward scattering dominates at high photon energies. At lower photon energies, however, there is a more complex diffraction behavior in which the intensity modulations do not necessarily lie along the atomic bond directions. This kind of multiple scattering in electron diffraction produces variations of the emitted electron intensity with respect to angle and energy. In order to determine a surface structure from such diffraction data, the measured angular and/or energy distributions of the photoelectrons are compared to a theoretical prediction for a given structure. The structure employed for the calculation is varied until the best fit to the data is found.

One of the advantages of using synchrotron radiation for XPD studies is that the photon energy, and not just the emission angle, can be adjusted. Another advantage is that the large photon flux allows for spectroscopy to be performed on a much quicker timescale. This is particularly important for photoelectron diffraction,

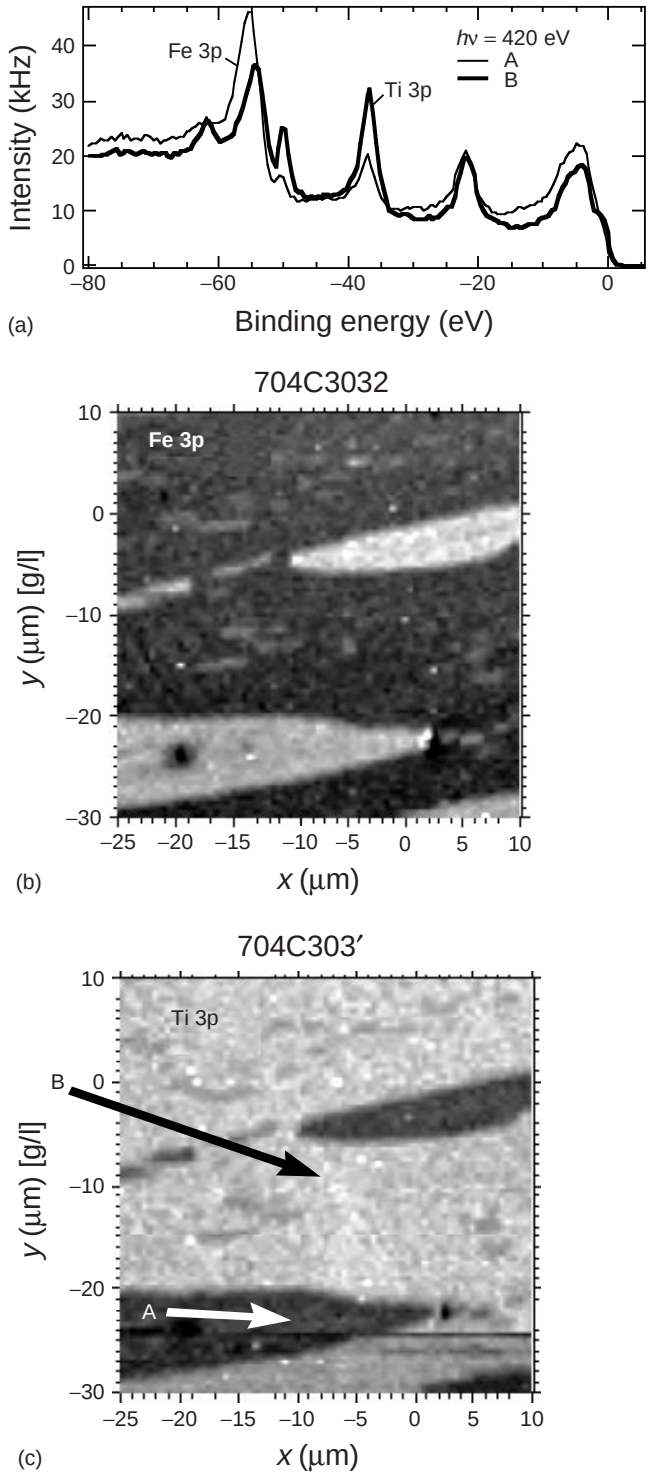

Figure 12 Scanning SXPS microscopy of a natural ilmenite sample.[43] (a) SXPS spectra collected from two different regions, A and B, of the surface using a photon energy of 420 eV. (b) SXPS micrograph image acquired by monitoring the Fe 3p photoelectrons. (c) SXPS micrograph acquired by monitoring the Ti 3p photoelectrons. The regions used in (a) are indicated in (c). (Reproduced from Tonner et al., *Surf. Interface Anal.*, **27**, 247–258 (1999) by permission of John Wiley & Sons.)

as there are typically a large number of spectra that must be collected in order to properly determine a surface structure. As an example, Figure 13 shows angular distributions of the S 2p core level, each collected with a different photon energy from a clean PbS(001), i.e. galena, surface (B.P. Tonner, unpublished results). Each pattern is marked with the electron kinetic energy used for measurement.

In photoelectron holography, the diffraction pattern represented by the angular distribution of a particular photoelectron feature is measured and then directly converted into the real-space image of a surface structure.[46] Thus, no model of the structure needs to be assumed, and there is no trial-and-error type of analysis in which artificial spectra are fabricated based on the model structures. This is similar to the manner in which X-ray diffraction is used to determine bulk crystal structures. Photoelectron holography is hampered by the difficulty in accurately measuring photoelectron angular distributions, however, as well as by the more complex scattering behavior of electrons as compared with X-rays. However, this technique is currently under development and much progress has been made in recent years.

9.4 Spin-polarized Photoelectron Spectroscopy

It is possible to measure not only the kinetic energy of a photoelectron, but also its spin.[47] Spin-polarized studies have applications in the study of magnetic materials, and have led to an improved understanding of the origins of magnetic anisotropy. Spin-resolved valence and core-level spectroscopy have been performed on clean surfaces, adsorbate-covered and oxidized surfaces, and thin magnetic films. The inherent strength of spin-resolved measurements is the direct probing of the imbalance in the spin-up and spin-down population, which is the basis of the magnetic behavior. It has also been found that spin is important in describing photoemission from non-magnetic materials, in particular when using chiral experimental geometries, for example by employing polarized photon beams.

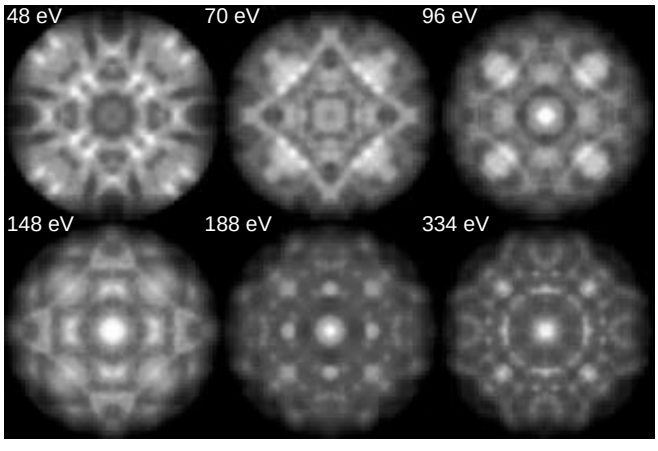

Figure 13 XPD angular distribution images of S 2p photoelectrons emitted from a clean PbS(001) surface. (Reproduced by permission of B.P. Tonner.)

The spin of a photoelectron is measured with the use of a specialized spin-resolving detector, which is most often attached to the exit of the electron energy analyzer. The majority of today's detectors are based on the spin-dependence in electron–solid scattering. The most commonly used is the Mott detector, which operates in the 10–100 keV range and uses Au as a target. The spin dependence in a Mott detector is produced by the spin–orbit part of the scattering potential, which is due to coupling of the electron spin and the angular momentum acquired in the scattering process. Since the spin–orbit part of the potential is proportional to the vector product of the spin and the angular momentum, the electrons scattered to the left acquire the opposite angular momentum from the ones scattered to the right, thereby producing different scattering probability for a given electron spin direction. The distribution of the spins is then measured with the use of two separate electron detectors and it is proportional to the asymmetry of the two intensities. Spin detectors are inherently inefficient (the figure of merit for Mott detectors is only $\sim 10^{-4}$) because of losses in the scattering process ($\sim 10^{-3}$) and because there is only a limited spin asymmetry in the scattering (~ 0.3). Thus, it is necessary to have a large incident photon flux in order to create a sufficient number of photoelectrons so that the signal can be detected. For this reason, spin-resolved PES is usually carried out at specialized insertion device beamlines, which are described below.

As an example, Figure 14 shows Fe 3p spectra collected from clean Fe(001) and after a small exposure to O_2.[48] Figure 14(a) shows the spin-up photoelectrons and Figure 14(b) shows the spin-down photoelectrons both before and after oxidation. It is seen that oxygen adsorption produces a new chemical shift towards higher binding energy, and that the signal arising from the oxidized overlayer is not symmetric with respect to spin, indicating some net oxide spin polarization. Furthermore, this polarization is opposite to the polarization of the Fe metal indicating antiparallel coupling between the metal and oxide magnetic moments. This is effectively a spin-resolved version of SXPS, which illustrates the ability to separately identify the magnetic states of thin Fe oxide layers on top and Fe metal below.

Another method of probing magnetic properties is to use circularly polarized light to observe magnetic circular dichroism (MCD).[49] In the MCD technique, photoelectron spectra are compared for light that is right-hand vs left-hand polarized. The difference in the spectra

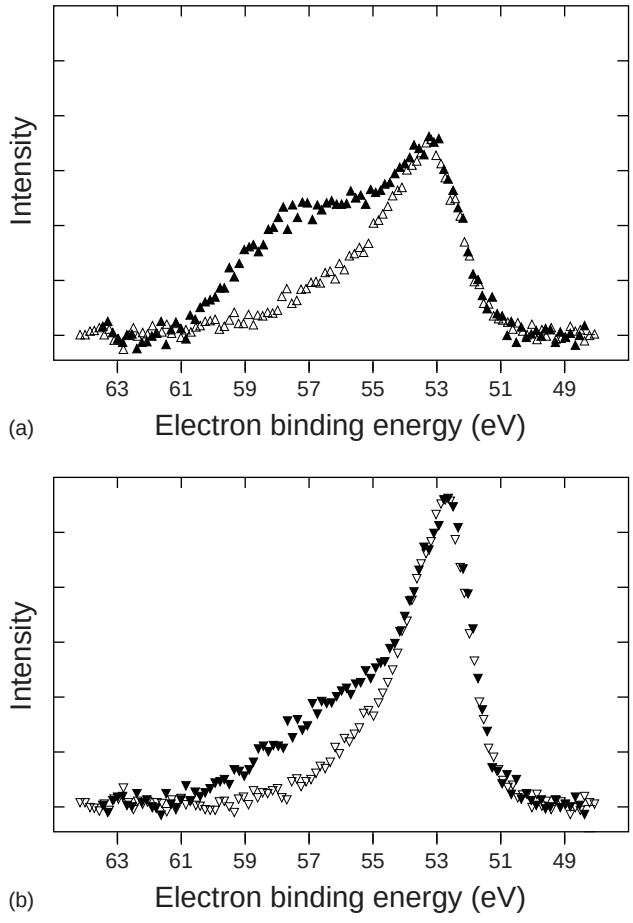

Figure 14 Fe 3p SXPS spectra collected with a photon energy of 90 eV from clean Fe(001) (open triangles) and after exposure to 16 L of O_2 (filled triangles). The majority spin photoelectrons are shown in (a), and the minority spin photoelectrons in (b). (Reproduced from Sinkovic et al., *Phys. Rev. Lett.*, **65**, 1647 (1990) by permission of the American Physical Society.)

reflects the spin character of the initial states. Circularly polarized light is required in order to perform MCD. Although synchrotron radiation is linearly polarized in the plane at the center of the beam, the light is elliptically polarized at the upper and lower portions of the beam. Thus, these portions of the beam are utilized for some MCD experiments. For more sophisticated measurements, however, specialized insertion devices are used to directly produce a high flux of circularly polarized light. Since the dichroism is based on the spin–orbit coupling in the photoemission process, the interpretation of MCD spectra is much less straightforward than the spin-resolved measurements. Nevertheless, MCD techniques allow for a much more efficient investigation of magnetic effects since they do not require a spin-detector, and have thus complemented the spin-resolving measurements.

10 INSTRUMENTATION FOR SOFT X-RAY PHOTOELECTRON SPECTROSCOPY

Because of the need for intense, tunable radiation, SXPS is carried out at synchrotron light source facilities that employ an electron storage ring for the production of synchrotron radiation. In the USA, these include the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, the Stanford Synchrotron Radiation Laboratory (SSRL), the Advanced Photon Source (APS) at Argonne National Laboratory, and the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory, among others. There are also other synchrotron light sources in operation, and new ones under construction, throughout the world. Most of these function as user facilities, so that any researcher with a good idea can get beamtime via the submission of a short proposal.

Aside from the storage ring, the instrumentation needed for SXPS spectroscopy consists of a monochromator and a spectrometer. The spectrometer is normally housed inside a UHV chamber. Like any surface science technique, SXPS must be carried out under UHV in order to maintain sample cleanliness. Synchrotron radiation storage rings must also be kept under UHV conditions in order to be able to maintain a stored electron beam for long periods of time. Thus, the entire storage ring, and each beamline attached to it, is essentially one huge interconnected vacuum system.

In the following sections, the three major components of an SXPS experiment are described. Note, however, that each beamline must also contain the instrumentation normally used to prepare surfaces for UHV studies, such as a sputter gun, sample manipulator, and low-energy electron diffraction (LEED) optics.

10.1 Storage Ring

The major piece of instrumentation for SXPS is the electron storage ring used to produce synchrotron radiation.[2] An electron storage ring is a large donut-shaped vacuum chamber inside which electrons travel at relativistic velocities around a nearly circular path. The electrons are steered around the ring by a number of ''bending magnets'' and can be stored in a stable orbit for hours. When any charged particle is accelerated, there is a perturbation to the electric field that induces the emission of radiation. As each bending magnet alters the direction of the beam in a storage ring, synchrotron radiation is emitted primarily in the direction tangential to the electron path. At a typical storage ring that is being used as a light source, there are a large number of bending magnets, with each one attached to a ''beamline''. The radiation is focused down the beamline by a combination

of grazing incidence X-ray mirrors so that it can be used for various experiments. Note that SXPS is only one of the many uses of synchrotron radiation.

The earliest, or first-generation, synchrotron radiation experiments were performed using accelerators that were built for high-energy physics experiments. These ''parasitic'' experiments provided the earliest SXPS results, but had as a disadvantage the fact that the available beamtime was dependent upon the experimental schedule of the particle physicists. The second-generation machines, such as the NSLS, were built specifically for the production of synchrotron radiation. These machines have become the cornerstones of synchrotron radiation research in providing stable, reliable beams for a large variety of experiments. The latest technology in synchrotron radiation light sources is represented by the third-generation machines, such as the ALS.[50] The designs of these machines have been optimized for the use of insertion devices to increase the photon flux and brightness.

Insertion devices, such as undulators and wigglers, are placed directly in the path of the electrons traveling through the storage ring. Insertion devices use magnetic fields to rapidly accelerate the electron beam perpendicular to the propagation direction, which produces a strong enhancement of the photon yield at a particular energy by focusing all of the synchrotron energy into a narrow spectral region. Insertion devices can increase the brightness up to four orders of magnitudes above that available from a conventional bending magnet beamline.

10.2 Monochromator

A monochromator is used to select a single photon energy from the white light synchrotron radiation beam.[2] The monochromators that are used for the soft X-ray region are based on the use of diffraction gratings of the grazing incidence reflection type. Grazing incidence is required so that the material of the grating does not absorb the X-ray beam. These gratings are of three general types, classified by the shape of the reflecting surface: (1) plane grating monochromator (PGM); (2) spherical grating monochromator (SGM); and (3) toroidal grating monochromator (TGM). Typical beamlines contain a selection of gratings, each optimized for a different spectral region. Before the beam enters the grating chamber, and also after it leaves, it is collimated by slits that allow only a narrow spectral region to pass. In more sophisticated beamlines, the slits move in a direction parallel to the beam as the photon energy is adjusted to insure that they are always located at the focal point of the grating. In addition, each beamline contains X-ray reflection mirrors used to steer and focus the beam onto the grating and then onto the sample.

10.3 Electron Energy Analyzer

Electron energy analyzers come in various types, but most use electrostatic deflection to measure electron kinetic energies.[6] The most common is the hemispherical energy analyzer. These analyzers are constructed from concentric hemispherical sectors, and electrons travel between the two hemispheres. At the entrance to the analyzer, there is an electrostatic lens that is used to focus the electrons. At the exit of the analyzer, there is a channeltron or microchannelplate electron detector. The outer hemisphere is held at a more negative potential than the inner hemisphere, so that the electrons of the proper energy will follow the path between the hemispheres. Hemispherical analyzers can efficiently collect electrons emitted into a small angular region, and measure them with high spectral resolution. Cylindrical mirror analyzers (CMAs) are also occasionally used for SXPS. In a CMA, electrostatic fields are used to focus electrons along cylindrical paths. CMAs collect a wider angular region than hemispherical analyzers, but are not as efficient and cannot achieve as high an energy resolution. There are also other designs for electrostatic analyzers,[51] as well as time-of-flight analyzers.

ABBREVIATIONS AND ACRONYMS

ALS	Advanced Light Source
APS	Advanced Photon Source
ARPES	Angle-resolved Photoelectron Spectroscopy
ARUPS	Angle-resolved Ultraviolet Photoelectron Spectroscopy
CBM	Conduction Band Minimum
CFS	Constant Final State
CIS	Constant Initial State
CMA	Cylindrical Mirror Analyzer
DOS	Density-of-states
EDC	Energy Distribution Curve
ESCA	Electron Spectroscopy for Chemical Analysis
IMFP	Inelastic Mean Free Path
LEED	Low-energy Electron Diffraction
MCD	Magnetic Circular Dichroism
MED	Mean Escape Depth
NEXAFS	Near Edge X-ray Absorption Fine Structure
NSLS	National Synchrotron Light Source
PES	Photoelectron Spectroscopy
PEY	Partial Electron Yield
PGM	Plane Grating Monochromator
SCLS	Surface Core-level Shift

SEXAFS	Surface Extended X-ray Absorption Fine Structure
SGM	Spherical Grating Monochromator
SPV	Surface Photovoltage
SSRL	Stanford Synchrotron Radiation Laboratory
SXPS	Soft X-ray Photoelectron Spectroscopy
TEY	Total Electron Yield
TGM	Toroidal Grating Monochromator
UHV	Ultrahigh Vacuum
UPS	Ultraviolet Photoelectron Spectroscopy
UV	Ultraviolet
VBM	Valence Band Maximum
VUV	Vacuum Ultraviolet
XANES	X-ray Absorption Near Edge Spectroscopy
XPD	X-ray Photoelectron Diffraction
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Surfaces **(Volume 10)**
Surfaces: Introduction • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

REFERENCES

1. M.P. Seah, W.A. Dench, 'Quantitative Electron Spectroscopy of Surfaces: A Standard Data Base for Electron Inelastic Mean Free Paths in Solids', *Surf. Interface Anal.*, **1**, 2–11 (1979).
2. G. Margaritondo, *Introduction to Synchrotron Radiation*, Oxford University Press, New York, 1988.
3. E.-E. Koch, *Handbook on Synchrotron Radiation*, North Holland, Amsterdam, Vol. 1A, 1983.
4. C.D. Wagner, W.M. Riggs, L.E. Davis, J.F. Moulder, G.E. Mullenberg, *Handbook of X-ray Photoelectron Spectroscopy*, Perkin-Elmer Corporation, Eden Prairie, MN, 1979.
5. J.J. Yeh, I. Lindau, 'Subshell Photoionization Cross-sections', *At. Data Nucl. Data Tables*, **32**, 1–155 (1985).
6. D.P. Woodruff, T.A. Delchar, *Modern Techniques of Surface Science*, 2nd edition, Cambridge University Press, Cambridge, 1994.
7. H. Lüth, *Surfaces and Interfaces of Solid Materials*, 3rd edition, Springer-Verlag, Berlin, 1995.
8. S. Hüfner, *Photoelectron Spectroscopy: Principles and Applications*, 2nd edition, Springer, Berlin, 1996.
9. C.B. Duke, 'Interaction of Electrons and Positrons with Solids: From Bulk to Surface in Thirty Years', *Surf. Sci.*, **299/300**, 24–33 (1994).
10. C.J. Powell, A. Jablonski, I.S. Tilinin, S. Tanuma, D.R. Penn, 'Surface Sensitivity of Auger-electron Spectroscopy and X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **98–99**, 1–14 (1999).
11. C.J. Powell, 'Inelastic Interactions of Electrons with Surfaces: Applications to Auger-electron Spectroscopy and X-ray Photoelectron Spectroscopy', *Surf. Sci.*, **299/300**, 34–48 (1994).
12. E.W. Plummer, W. Eberhardt, 'Angle-resolved Photoemission as a Tool for the Study of Surfaces', *Adv. Chem. Phys.*, **49**, 533–656 (1982).
13. N.V. Smith, F.J. Himpsel, 'Photoelectron Spectroscopy,' in *Handbook on Synchrotron Radiation*, ed. E.E. Koch, North-Holland Publishing Company, Amsterdam, 905–954, 1983.
14. S.D. Kevan, 'Angle-resolved Photoemission: Theory and Current Applications', *Studies in Surface Science and Catalysis*, Elsevier, Amsterdam, Vol. 74, 1992.
15. T.D. Durbin, W.C. Simpson, V. Chakarian, D.K. Shuh, P.R. Varekamp, C.W. Lo, J.A. Yarmoff, 'Stimulated Desorption of Cl^+ and the Chemisorption of Cl_2 on Si(111)-7×7 and Si(100)-2×1', *Surf. Sci.*, **316**, 257–266 (1994).
16. D.A. Shirley, 'High-Resolution X-ray Photoemission Spectrum of the Valence Bands of Gold', *Phys. Rev. B*, **5**, 4709–4714 (1972).
17. F.J. Himpsel, B.S. Meyerson, F.R. McFeely, J.F. Morar, A. Taleb-Ibrahimi, J.A. Yarmoff, 'Core Level Spectroscopy at Silicon Surface and Interfaces', in *Enrico Fermi School on 'Photoemission and Absorption Spectroscopy of Solids and Interfaces with Synchrotron Radiation'*, eds. M. Campagna, R. Rosei, North Holland, Amsterdam, 203–236, 1990.
18. A. Flodström, R. Nyholm, B. Johansson, 'Surface Core Level Spectroscopy', in *Synchrotron Radiation Research*, ed. R.Z. Bachrach, Plenum Press, New York, 199–251, Vol. 1, 1992.
19. B. Johansson, N. Mårtensson, 'Core-level Binding-energy Shifts for the Metallic Elements', *Phys. Rev. B*, **21**, 4427–4457 (1980).
20. A. Rosengren, B. Johansson, 'Calculated Transition-metal Surface Core-level Binding-energy Shifts', *Phys. Rev. B*, **22**, 3706–3709 (1980).
21. G.K. Wertheim, 'Core-Electron Binding Energies in Atoms and Monolayers Adsorbed on Metallic Surfaces', *Appl. Phys. A*, **41**, 75–81 (1986).
22. F.J. Himpsel, F.R. McFeely, A. Taleb-Ibrahimi, J.A. Yarmoff, 'Microscopic Structure of the SiO_2/Si Interface', *Phys. Rev. B*, **38**, 6084–6096 (1988).
23. G.V. Hansson, R.I.G. Uhrberg, 'Photoelectron Spectroscopy of Surface States on Semiconductor Surfaces', *Surf. Sci. Rep.*, **9**, 197–292 (1988).

24. D.M. Riffe, G.K. Wertheim, P.H. Citrin, 'Different Core-hole Lifetime and Screening in the Surface of W(110)', *Phys. Rev. Lett.*, **63**, 1976–1979 (1989).
25. N. Mårtensson, A. Nilsson, 'High-resolution Core-level Photoelectron Spectroscopy of Surfaces and Adsorbates', in *Applications of Synchrotron Radiation. High-resolution Studies of Molecules and Molecular Adsorbates on Surfaces*, ed. W. Eberhardt, Springer-Verlag, Berlin, 65–126, 1995.
26. C.W. Lo, D.K. Shuh, V. Chakarian, T.D. Durbin, P.R. Varekamp, J.A. Yarmoff, 'XeF_2 Etching of Si(111): The Geometric Structure of the Reaction Layer', *Phys. Rev. B*, **47**, 15 648–15 659 (1993).
27. W.K. Wang, W.C. Simpson, J.A. Yarmoff, 'Passivation Versus Etching: Adsorption of I_2 on InAs(001)', *Phys. Rev. Lett.*, **81**, 1465–1468 (1998).
28. J. Stöhr, *NEXAFS Spectroscopy*, Springer-Verlag, Berlin, 1992.
29. J.W. Allen, 'Resonant Photoemission of Solids with Strongly Correlated Electrons', in *Synchrotron Radiation Research*, ed. R.Z. Bachrach, Plenum Press, New York, 253–323, Vol. 1, 1992.
30. M. Iwan, F.J. Himpsel, D.E. Eastman, 'Two-Electron Resonance at the 3p Threshold of Cu and Ni', *Phys. Rev. Lett.*, **43**, 1829–1832 (1979).
31. K.E. Smith, V.E. Henrich, 'Resonant Photoemission in Ti_2O_3 and V_2O_3: Hybridization and Localization of Cation 3d Orbitals', *Phys. Rev. B*, **38**, 9571–9580 (1988).
32. R.L. Kurtz, R.L. Stockbauer, D. Mueller, A. Shih, L.E. Toth, M. Osofsky, S.A. Wolf, 'Resonant Photoemission Study of Superconducting Y–Ba–Cu–O', *Phys. Rev. B*, **35**, 8818–8820 (1987).
33. M. Onellion, Y. Chang, D.W. Niles, R. Joynt, G. Margaritondo, N.G. Stoffel, J.M. Tarascon, 'Photoemission Resonances of the High-temperature Superconductor $Ba_2YCu_3O_{7+x}$', *Phys. Rev. B*, **36**, 819–821 (1987).
34. L.J. Brillson, 'The Structure and Properties of Metal–Semiconductor Interfaces', *Surf. Sci. Rep.*, **2**, 123–326 (1982).
35. K. Horn, 'Semiconductor Interface Studies Using Core and Valence Level Photoemission', *Appl. Phys. A*, **51**, 289–304 (1990).
36. M. Alonso, R. Cimino, K. Horn, 'Surface Photovoltage Effects in Photoemission from Metal–GaP(110) Interfaces: Importance for Band-bending Evaluation', *Phys. Rev. Lett.*, **64**, 1947–1950 (1990).
37. T. Aruga, Y. Murata, 'Alkali Metal Adsorption on Metals', *Prog. Surf. Sci.*, **31**, 61–130 (1989).
38. H. Winick, S. Doniach, *Synchrotron Radiation Research*, Plenum Press, New York, 1980.
39. G. Le Lay, V.Y. Aristov, M. Fontaine, 'High-resolution Synchrotron Radiation Core-level Spectroscopy of Semiconductor Surfaces', *Prog. Surf. Sci.*, **48**, 145–156 (1995).
40. G. Le Lay, M. Göthelid, T.M. Grehk, M. Björkquist, V.Y. Aristov, U.O. Karlsson, 'Surface Core-level Shifts of Si(111)-7×7: A Fundamental Reassessment', *Phys. Rev. B*, **50**, 14 277–14 282 (1994).
41. Z.M. Stadnik, D. Purdie, M. Garnier, Y. Baer, A.-P. Tsai, A. Inoue, K. Edagawa, S. Takeuchi, K.H.J. Buschow, 'Electronic Structure of Quasicrystals Studied by Ultra-high-energy-resolution Photoemission Spectroscopy', *Phys. Rev. B*, **55**, 10 938–10 951 (1997).
42. B.P. Tonner, D. Dunham, T. Droubay, J. Kikuma, J. Denlinger, E. Rotenberg, A. Warwick, 'The Development of Electron Spectromicroscopy', *J. Electron Spectrosc.*, **75–322**, 309–322 (1995).
43. B.P. Tonner, T. Droubay, J. Denlinger, W. Meyer-Ilse, T. Warwick, J. Rothe, E. Kneedler, K. Pecher, K. Nealson, T. Grundl, 'Soft X-ray Spectroscopy and Imaging of Interfacial Chemistry in Environmental Specimens', *Surf. Interface Anal.*, **27**, 247–258 (1999).
44. W.F. Egelhoff, Jr, 'X-ray Photoelectron and Auger Electron Forward Scattering: A New Tool for Surface Crystallography', *CRC Crit. Rev. Solid State Mater. Sci.*, **16**, 213–235 (1990).
45. C.S. Fadley, 'Diffraction and Holography with Photoelectrons and Auger Electrons: Some New Directions', *Surf. Sci. Rep.*, **19**, 231–264 (1993).
46. P.M. Len, J.D. Denlinger, E. Rotenberg, S.D. Kevan, B.P. Tonner, Y. Chen, M.A. Van Hove, C.S. Fadley, 'Holographic Atomic Images from Surface and Bulk W(110) Photoelectron Diffraction Data', *Phys. Rev. B*, **59**, 5857–5870 (1999).
47. P.D. Johnson, 'Spin-polarized Photoemission', *Rep. Prog. Phys.*, **60**, 1217–1304 (1997).
48. B. Sinkovic, P.D. Johnson, N.B. Brooks, 'Magnetic Structure of Oxidized Fe(001)', *Phys. Rev. Lett.*, **65**, 1647–1650 (1990).
49. C.M. Schneider, J. Kirschner, 'Spin- and Angle-resolved Photoelectron Spectroscopy from Solid Surfaces with Circularly Polarized Light', *Crit. Rev. Solid State Mater. Sci.*, **20**, 179–283 (1995).
50. G. Comelli, M. Kiskinova, G. Paolucci, K.C. Prince, A. Savoia, G. Margaritondo, 'Latest Photoemission Developments: Time Resolution, Microanalysis, Order Parameters', *Surf. Rev. Lett.*, **4**, 695–701 (1997).
51. D.E. Eastman, J.J. Donelon, N.C. Hein, F.J. Himpsel, 'An Ellipsoidal Mirror Display Analyzer System for Electron Energy and Angular Measurements', *Nucl. Instrum. Methods*, **172**, 327–336 (1980).

X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Steffen Oswald
Institut für Festkörper- und Werkstofforschung Dresden, Dresden, Germany

1 Introduction 1
2 Sample Requirements, Sample Preparation 2
3 Measuring Strategy 3
3.1 Elemental Information 3
3.2 Chemical Information 6
3.3 Local Information 8
3.4 Valence Band Information 10
4 Applications for Typical Material Classes 11
4.1 Metals and Alloys 11
4.2 Semiconductors 12
4.3 Insulators 14
4.4 Polymers and Organic Materials 18
5 Special Methodologies 22
5.1 Valence Band Spectroscopy 22
5.2 Loss Spectroscopy 22
5.3 Photoelectron Diffraction 23
5.4 In Situ Preparation, Special Equipment 24
6 Comparison with Other Techniques 24
6.1 Auger Electron Spectroscopy 24
6.2 Secondary Ion Mass Spectrometry 24
6.3 Atomic Force Microscopy and Raster Tunneling Microscopy 25
6.4 Low-energy Electron Diffraction and Reflection High-energy Electron Diffraction 25
6.5 Ion Surface Scattering 25
6.6 Infrared Spectroscopy 25
6.7 Electron Energy Loss Spectroscopy 25
6.8 Rutherford Backscattering Spectroscopy 25
6.9 Glow Discharge 25
7 Outlook 26
Abbreviations and Acronyms 26
Related Articles 26
References 26

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

X-ray photoelectron spectroscopy (XPS) is an analytical technique that uses photoelectrons excited by X-ray radiation (usually Mg Kα or Al Kα) for the characterization of surfaces to a depth of 2–5 nm. Elemental identification and information on chemical bonding are derived from the measured electron energy and energy shifts, respectively. The use of ultrahigh vacuum (UHV) during analysis requires special sample handling. Depth profiling is possible using ion sputtering.

In contrast to the most popular surface analytical technique, Auger electron spectroscopy (AES), nonconducting material can be investigated, little material damage occurs, and the chemical shifts are easier to interpret. However, the lateral resolution of AES (typically 50 nm) is much greater than for XPS (50 μm for standard equipment, down to lower than 3 μm for dedicated instruments). The detection limit (of about 0.1 atom %) is not as low as for mass spectroscopic techniques, but the quantification from XPS peak area measurements is much better, even disclaiming standard sample measurements.

This article briefly discusses the principles of the technique, sample requirements, and the typical measuring strategy. Possible information sources (elements, chemical bonding, depth and lateral distributions) are described and their quantification principles are summarized. The main part deals with typical applications relating to several material classes (metals, semiconductors, insulators, polymers) to give a feeling for the effectiveness of the method over a wide range of research and technological problems. A comparison with similar techniques is given as a summary.

Technological developments in electronics, nanotechnology, polymers, biotechnology and medicine are all concerned with surface-related phenomena, suggesting sustained interest in XPS in the foreseeable future.

1 INTRODUCTION

XPS is an analytical technique which uses photoelectrons excited by X-ray radiation and released from the material in to vacuum for the characterization of surfaces. The electrons in the (solid) sample are characterized by their binding energies (BEs) which depend on the element. Using only the elastically scattered electrons (electrons without energy losses) emitted from the sample, information on the composition of the sample may be derived from their energy spectrum. The energy spectrum is measured in an electron spectrometer and the measured kinetic energies (KEs) of the electrons are determined by their BE and the energy of the X-ray photons used for excitation.

The surface sensitivity of the technique is determined by the relatively low escape depth (0.5–2 nm) of the

elastically scattered electrons. All elements of the periodic table can be detected by different characteristic BE peak positions (with the restriction that the light elements H and He may only be found with very low intensity in the valence band). The detection limits depend on the element and are typically in the range of 0.1–0.5 atom %. As the excitation does not use electrically charged particles, the method is dedicated to the measurement of electrically nonconducting materials. Radiation damage of the surface is small in comparison with electron beam excitation. Chemical information may be obtained by analyzing typical BE shifts of the photoelectron peaks, which depend on the chemical bonding of the elements under investigation.

However, because of the high surface sensitivity and to avoid energy losses of electrons in the analyzer region, the measurements require UHV conditions typically in the range of 10^{-9}–10^{-8} Pa. Sample information is often disturbed by surface contamination (mostly hydrocarbons or hydroxides) which requires special sample preparation or preconditioning. In multicomponent samples spectral overlap from several elements may occur, further complicated by X-ray excited Auger electron (XAES) lines.

The origin of the technique is connected with the discovery of the photoelectric effect by Hertz in 1887. In the early years of the twentieth century additional theoretical and experimental work was done. In 1914, Rutherford determined Equation (1),

$$E_k = h\nu - E_b \quad (1)$$

which describes the measured electron energy E_k as the difference of X-ray energy $h\nu$ and the electron BE E_b. Step by step the instrumentation was improved by various research groups; however, the modern XPS era is closely connected with the work of Siegbahn et al., dating from about 1954. They also introduced the acronym electron spectroscopy for chemical analysis (ESCA) which is often used.[1] Industrial equipment became available in the 1970s and today XPS machines are among the most used surface analytical tools.

Valence band spectroscopy, which is often concerned only in context with ultraviolet (UV) excitation, should be regarded as an integral part of the XPS method. Although the interpretation of valence band spectra needs a lot of expertise, this technique is also briefly discussed.

The XPS method and its applications are described in considerable detail in monographs and review articles that summarize both the earlier[1–7] and the more recent[8–12] years of XPS development. Additionally the article by Turner (**X-ray Photoelectron and Auger Electron Spectroscopy**) in this *Encyclopedia* covers the whole field of XPS: physical principles, instrumentation and quantification. Consequently, this article covers the physics and theory only to the extent necessary to acknowledge the general application strategy and to interpret the examples discussed.

2 SAMPLE REQUIREMENTS, SAMPLE PREPARATION

The sample requirements are determined from the design of the particular electron spectrometer used and the physical principles of the technique itself. The minimum sample size is determined from the acceptance area of the analyzer, and the maximum sample size by the construction of the entry lock and the sample manipulating system (today up to 200 mm). Most modern spectrometers have variable acceptance areas of between about 50 μm and 1 mm, therefore the typical sample size (also for handling purposes) is in the range of 5 mm × 5 mm. The samples should be as smooth as possible in order to have a defined surface region for electron emission.

Usually the XPS measurements have to be done under UHV conditions. Remember that for a gas molecule at room temperature and a pressure of 10^{-4} Pa, and assuming a sticking probability of 1, the time for monolayer adsorption is only of the order of some seconds! Thus for minimum electron collision with the gas molecules, to avoid energy losses, and for minimizing the recontamination of the prepared surfaces, pressures in the range of 10^{-8} Pa should be reached at least. Therefore samples should generally have a low vapor pressure. Sample preparation with organic glue or epoxy resin should be avoided. Dedicated adhesive materials are available and can be used in minimum quantities for special preparation purposes. Powders may be impressed into ductile metal foils (In or Au). An alternative is the use of special UHV-compatible double-sided adhesive tape for sample fixing. Generally the samples should be outgassed in the fast entry chamber of the spectrometer for enough time, especially in the case of porous material. Biological samples may be freeze dried or, if sensitive to dehydration, the samples may be cooled in situ (by liquid nitrogen) during analysis.

Surface contamination is one of the main problems of sample preparation for surface analysis, unless the topmost natural surface layer itself is the topic of investigation (e.g. adhesion, catalysis, corrosion). In particular so-called ''aventitious carbon'' contamination from hydrocarbons plays an important role. Therefore clean conditions (gloves, tweezers, exsiccators, glove boxes, laminar boxes, etc.) and special cleaning procedures (rinsing in solvents, ultrasonic cleaning) should be routinely used. Despite such sample precleaning, contamination

never can be avoided completely. Because even parts of monolayers of adsorbed species can negatively influence the surface-specific electron spectroscopic measurements, in situ cleaning is necessary in many cases. Medium-energy (1–5 keV) noble gas (Ar, Kr) ion sputtering is generally used; however, the ion bombardment induces morphological, structural and chemical changes in the surface. For the preparation of clean single-crystal surfaces sputtering–heating cycles are often applied. As an alternative, mechanical methods can be used to prepare fresh surfaces under UHV conditions if the sample properties and geometry are suitable. In principle cleaving (of single crystals), scraping (with diamond-coated tools) or fracturing (special sample geometry, cooling with liquid nitrogen) may be used. Guidelines for specimen handling and preparation may be found in some reviews.[13,14]

3 MEASURING STRATEGY

3.1 Elemental Information

Information about the elements incorporated in the surface region is obtained from the measured BE values of the separate lines in the electron spectrum. Complete series of spectra are given in several publications (a standard reference book is that by Moulder et al.[15]). Table 1 summarizes the BE values of the two most intense lines for each element. Note that XPS nomenclature uses the quantum numbers of the electronic core levels. In addition, note that the p, d and f levels are split by spin–orbit coupling into doublets with characteristic energy differences (in Table 1 energy splitting is given) and intensity ratios (1:2, 2:3, 3:4). Commonly Mg Kα (1253.6 eV) or Al Kα (1586.6 eV) are used for excitation because their natural line width is relatively small, photoelectron lines of all elements may be measured, and the relatively low KE of the measured photoelectrons allows high-energy resolution and good surface sensitivity. Only for special reasons, especially to obtain more bulk-specific information or to generate effectively higher energy Auger lines, X-ray sources with higher energy (e.g. Ag, Si, Ti, synchrotron radiation) have been used. The energy range in Table 1 is therefore restricted to the energy region up to 1400 eV.

As can be seen from this table there is a wide spread of energies which may be used for elemental identification. However, there may be spectral overlaps for specific combinations of elements, especially because Auger electron lines (not included in Table 1) are also excited by X-rays. The effect of overlapping Auger lines may be mostly avoided by an energy shift when changing the X-ray source. However, these XAES lines are often useful for both elemental identification and as a source of chemical information[16] (section 3.2).

As an example, Figure 1 shows an Al Kα excited spectrum of a Sb_2O_4 powder coated with a thin Au film for energy calibration. Besides a complete overlap of O1s and $Sb2p_{3/2}$, one can see the Auger lines, C contamination, Au lines and satellite lines (from satellite energies of nonmonochromatized Al Kα radiation). Despite the dependence on the element (see Equation 3), the measured intensities vary systematically depending on the transmission function of the electron analyzer, here with $1/E_k$ for the hemispheric analyzer used. An increasing background to lower KE (higher BE) comes from both nonelastic scattered electrons and the electrons excited by bremsstrahlung radiation from the X-ray source. Using a monochromatized source instead, the bremsstrahlung radiation and satellite peaks can be avoided, and additionally sharper peaks are obtained.

For quantitative analysis the BE scale has to be calibrated.[3,5,8] This is done experimentally using the Fermi energy level of the spectrometer $E_F = 0$ as a reference and by measuring (a) the well-defined peak positions of core levels of clean metals (e.g. Au4f = 84.0 eV, $Cu2p_{3/2} = 932.67$ eV) or (b) the valence band cut-off of metals with a high density of states near $E_F = 0$ (e.g. Ni). Thus the basic Equation (1) is modified to Equation (2),

$$E_k = h\nu - E_b - \Phi_S \tag{2}$$

with Φ_S as the work function of the spectrometer.

Concentration quantification is done using the intensity of a suitable (sufficient intensity, no spectral overlap with other elements) characteristic line as a measure for the number of atoms of a certain element in the analysis region. The basic relationship to calculate the total photoelectron intensity I_{ix} of the core level peak i of element x for a homogeneous sample is given by Equation (3):[9]

$$I_{ix} = B\sigma_{ix}\lambda_{ix}T_{ix}n_x \tag{3}$$

where B is the instrumental factor (X-ray flux, angular acceptance and total transmission of the spectrometer), σ is the photoabsorption cross-section for the level including asymmetry parameter, λ is the total electron escape depth for the core level energy and sample material, T is the transmission coefficient of the electrons through the surface, and n is the atomic density of element x.

For exact calculations all these terms have to be assumed to be dependent on space coordinates and the appropriate integrals have to be solved.[8,10]

For routine concentration calculations some assumptions have to be made:

Table 1 BE values (in eV, values rounded) for the two most intense XPS lines in the order of the elements of the periodic system (values are taken from Moulder et al.[15])

Atomic number	Element	Line	BE	Splitting	Line	BE	Splitting
3	Li	1s	56				
4	Be	1s	112				
5	B	1s	189				
6	C	1s	285				
7	N	1s	398				
8	O	1s	531		2s	23	
9	F	1s	685		2s	30	
10	Ne	1s	863		2s	41	
11	Na	1s	1072		2s	64	
12	Mg	1s	1303		2s	89	
13	Al	2s	118		2p	73	
14	Si	$2p_{3/2}$	99	1	2s	151	
15	P	$2p_{3/2}$	130	1	2s	188	
16	S	$2p_{3/2}$	164	1	2s	228	
17	Cl	$2p_{3/2}$	199	2	2s	271	
18	Ar	$2p_{3/2}$	242	2	2s	320	
19	K	$2p_{3/2}$	294	3	2s	380	
20	Ca	$2p_{3/2}$	347	4	2s	440	
21	Sc	$2p_{3/2}$	399	5	2s	499	
22	Ti	$2p_{3/2}$	454	6	2s	561	
23	V	$2p_{3/2}$	512	8	2s	626	
24	Cr	$2p_{3/2}$	574	9	3p	43	
25	Mn	$2p_{3/2}$	639	11	3p	48	
26	Fe	$2p_{3/2}$	707	13	3p	53	
27	Co	$2p_{3/2}$	778	15	3p	60	
28	Ni	$2p_{3/2}$	853	17	3p	67	
29	Cu	$2p_{3/2}$	933	20	$3p_{3/2}$	75	2
30	Zn	$2p_{3/2}$	1022	23	$3p_{3/2}$	89	2
31	Ga	$2p_{3/2}$	1117	27	3d	19	
32	Ge	$2p_{3/2}$	1217	31	$3d_{5/2}$	29	1
33	As	$2p_{3/2}$	1324	35	$3d_{5/2}$	42	1
34	Se	$3d_{5/2}$	56	1	$3p_{3/2}$	163	6
35	Br	$3d_{5/2}$	69	1	$3p_{3/2}$	182	7
36	Kr	$3d_{5/2}$	87	1	$3p_{3/2}$	208	8
37	Rb	$3d_{5/2}$	111	2	$3p_{3/2}$	240	9
38	Sr	$3d_{5/2}$	134	2	$3p_{3/2}$	270	11
39	Y	$3d_{5/2}$	156	2	$3p_{3/2}$	299	12
40	Zr	$3d_{5/2}$	179	2	$3p_{3/2}$	330	13
41	Nb	$3d_{5/2}$	202	3	$3p_{3/2}$	361	15
42	Mo	$3d_{5/2}$	228	3	$3p_{3/2}$	394	18
43	Tc	$3d_{5/2}$	253	4	$3p_{3/3}$	425	20
44	Ru	$3d_{5/2}$	280	4	$3p_{3/2}$	462	22
45	Rh	$3d_{5/2}$	307	5	$3p_{3/2}$	497	24
46	Pd	$3d_{5/2}$	335	5	$3p_{3/2}$	533	27
47	Ag	$3d_{5/2}$	368	6	$3p_{3/2}$	573	31
48	Cd	$3d_{5/2}$	405	7	$3p_{3/2}$	618	34
49	In	$3d_{5/2}$	444	8	4d	17	
50	Sn	$3d_{5/2}$	485	8	4d	25	
51	Sb	$3d_{5/2}$	528	9	4d	33	
52	Te	$3d_{5/2}$	573	10	$4d_{5/2}$	41	1
53	I	$3d_{5/2}$	619	11	$4d_{5/2}$	49	2
54	Xe	$3d_{5/2}$	670	13	$4d_{5/2}$	61	2
55	Cs	$3d_{5/2}$	726	14	$4d_{5/2}$	77	3
56	Ba	$3d_{5/2}$	781	15	$4d_{5/2}$	90	3
57	La	$3d_{5/2}$	836	17	$4d_{5/2}$	103	3
58	Ce	$3d_{5/2}$	884	18	$4d_{5/2}$	109	3
59	Pr	$3d_{5/2}$	932	20	4d	115	

Table 1 (*continued*)

Atomic number	Element	Line	BE	Splitting	Line	BE	Splitting
60	Nd	$3d_{5/2}$	981	22	4d	121	
61	Pm	$3d_{5/2}$	1034	26	4d	129	
62	Sm	$3d_{5/2}$	1081	27	4d	129	
63	Eu	$3d_{5/2}$	1126	29	4d	128	
64	Gd	$3d_{5/2}$	1186	32	4d	140	
65	Tb	$3d_{5/2}$	1241	35	4d	146	
66	Dy	$3d_{5/2}$	1296	37	4d	152	
67	Ho	4d	160		$4p_{3/2}$	309	44
68	Er	4d	167		$4p_{3/2}$	321	47
69	Tm	4d	175		$4p_{3/2}$	333	51
70	Yb	4d	182		$4p_{3/2}$	341	48
71	Lu	$4f_{7/2}$	7	2	$4d_{5/2}$	196	10
72	Hf	$4f_{7/2}$	14	2	$4d_{5/2}$	211	11
73	Ta	$4f_{7/2}$	22	2	$4d_{5/2}$	226	12
74	W	$4f_{7/2}$	31	2	$4d_{5/2}$	243	13
75	Re	$4f_{7/2}$	40	2	$4d_{5/2}$	260	14
76	Os	$4f_{7/2}$	51	3	$4d_{5/2}$	279	14
77	Ir	$4f_{7/2}$	61	3	$4d_{5/2}$	297	15
78	Pt	$4f_{7/2}$	71	3	$4d_{5/2}$	315	17
79	Au	$4f_{7/2}$	84	4	$4d_{5/2}$	335	18
80	Hg	$4f_{7/2}$	101	4	$4d_{5/2}$	361	20
81	Tl	$4f_{7/2}$	118	4	$4d_{5/2}$	385	21
82	Pb	$4f_{7/2}$	137	5	$4d_{5/2}$	412	22
83	Bi	$4f_{7/2}$	157	5	$4d_{5/2}$	440	24
90	Th	$4f_{7/2}$	333	9	$4d_{5/2}$	676	37
92	U	$4f_{7/2}$	377	11	$4d_{5/2}$	736	43

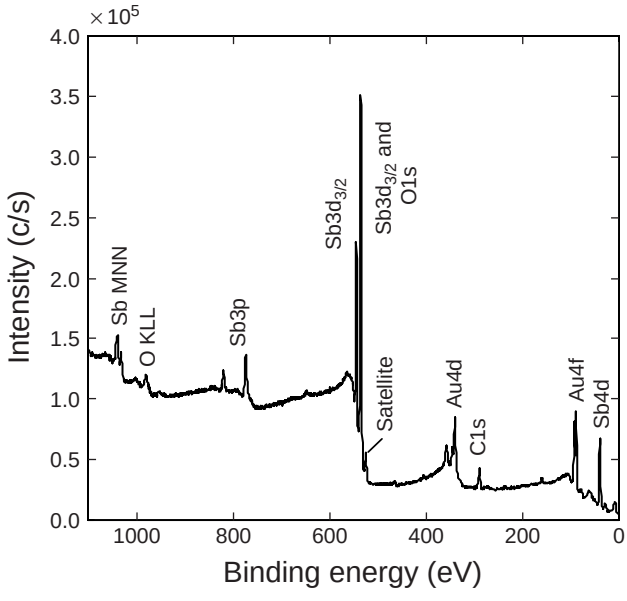

Figure 1 Spectrum of an Sb_2O_4 sample. The sample was coated with a thin Au film for energy calibration.

- the analysis region should be a homogeneous mixture of the elements;
- the excitation probability of the core level used does not depend on the environment of the atom (low matrix effect);
- the electron background near the selected peak has to be removed before the intensity calculation is carried out.

The background subtraction can be done in several ways, from simple linear dependence to physically determined functions.[17] Most used up to now is the method after Shirley[18] which calculates an integrated background under the peaks. An example is given in Figure 2 for the complex Mo3d peak structure for Mo oxide.

Even if both the peak height and the peak area can be used as intensity values for calculations,[15] only the use of peak area can be seriously recommended (see also **X-ray Photoelectron and Auger Electron Spectroscopy**). Peak area calculations lead to more realistic values, because the peak shape is changed by chemical effects (see Figure 2 and as discussed below) and will produce completely wrong results for components with different bonding states for the peak-height method. The peak area may be estimated by simple integration over the background or by using more sophisticated peak-fitting procedures (see section 3.2 and **X-ray Photoelectron and Auger Electron Spectroscopy**). Using these assumptions the simplest calculation for the concentration c is given by Equation (4):

$$c_x = \frac{I_x/S_x}{\sum_i I_i/S_i} \tag{4}$$

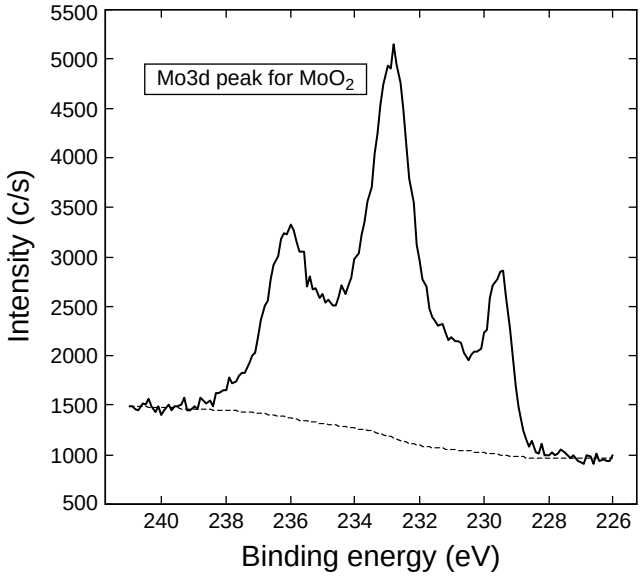

Figure 2 Demonstration of a Shirley background subtraction for an MoO_2 sample which was measured by XPS without further in situ sample treatment.

where I_x is the intensity (peak area or peak height) and S_x is a relative sensitivity factor of the element x.

In a first approximation the sensitivity factor is taken from an element or standard sample measured under equivalent conditions; alternatively, values from reference measurements for the appropriate analyzer type (using the same transmission function)[15] or from theoretical calculations[19] can be used. Analytical experience derived from working with a wide range of materials shows that, based on this simple strategy, a precision of about 10% can be reached; however, dramatic errors may occur if the sample is inhomogeneous within the range of electron escape depth.

Further improvements to the results may be obtained applying the following steps:[8]

1. measuring element standards under exactly the same measuring conditions as for the unknown sample;
2. measuring similar well-defined compound standards;
3. performing escape depth correction and correction of changed atomic density (matrix correction);
4. simulation of concentration gradients if they are well known (overlayer correction).

A critical review of the limits of quantitative analysis is given by Powell and Seah[20] which covers all the experimental, physical and mathematical aspects.

3.2 Chemical Information

One of the main advantages of XPS is the possibility of obtaining chemical information relatively easily by analyzing changes of the BE of the photoelectrons. This is possible because only one electron level is involved in the emission process, in comparison to AES where three energy levels have to be considered. The general shape of the photoelectron peaks, as shown in Figure 2 for a multicomponent peak, is determined by experimental and physical effects. First, the detected BE state is determined by the final ionized state of the emitting atom. Experimental broadening occurs typically as the result of the limited energy resolution of the spectrometer and the energetic width of the X-ray source, and may be described by a Gaussian function. Broadening due to the lifetime of the resulting hole is given by a Lorentzian function, and has to be added to the experimental broadening. Especially for metals, there is an asymmetric broadening to higher BE, generated by an interaction with unfilled electron levels of the valence band above the Fermi level (discussed in section 5.2). The resulting width is usually defined as full width at half maximum (FWHM) and, for optimum measuring conditions and sharp metallic lines (e.g. $Ag2p_{3/2}$), is around 0.5 eV.

The exact energetic peak position of an emitted core-level electron on the BE scale now depends on the environment of the emitting atom. In single-element (e.g. single-crystalline or polycrystalline) materials used for standardization, this environment is defined by the same element as its characteristic bonding state and near-neighbor distances. In multi-element materials (alloys, oxides, carbides, etc.) this environment is altered and so is the electronic structure of the emitting atom. Several more or less empirical models were proposed during the early stages of XPS development, such as the ''charge potential method'',[1,21] the ''valence potential method'',[22] and the ''equivalent cores approach''.[23,24] These models, summarized by Brundle and Baker[3] and by Briggs,[5] consider the chemical characteristics of the bonding partners as electronegativity, charge transfer or reaction energies. These original approaches have been further developed and extended, also using quantum mechanical methods.[9,10] The principles of chemical shift estimation are discussed here for the charge potential method, which describes the change ΔE in the BE of a core level by the changes in the electron density and the resulting potential changes:

$$\Delta E = k\Delta q + \Delta V \tag{5}$$

In Equation (5) Δq is the difference in valence charge of the atom (assumed to be located on a hollow sphere) multiplied by a factor k, and ΔV is the change in the effective potential assuming the surrounding atoms to be point charges. It is clear from the first term that the decrease in valence charge (e.g. oxidation) leads to higher BE values and the increase (e.g. reduction) to lower BE values.

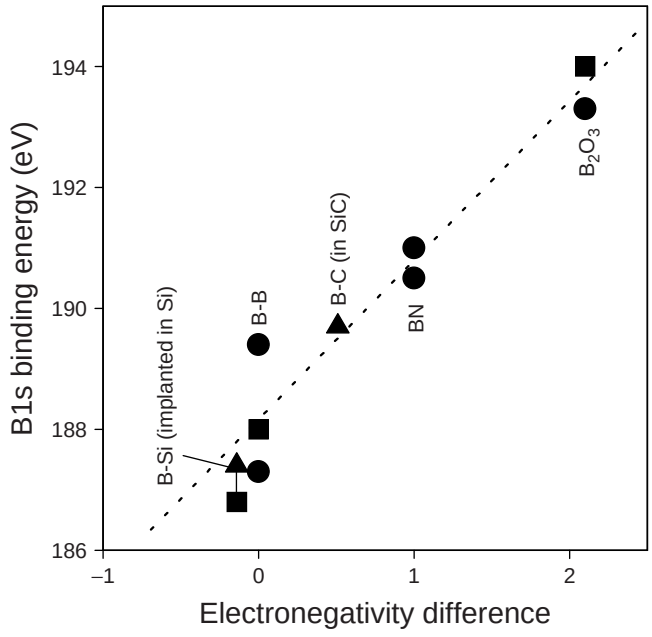

Figure 3 Dependence of B1s BE on electronegativity difference of the binding partners in several compounds. (Taken from S. Oswald et al.[26])

However, the second term must also to be considered; it has an opposite sign. This charge transfer situation has to be modeled using complex calculations; fast estimations are often possible using simple considerations of electronegativity differences,[21,25] as shown in Figure 3 for some boron compounds.[26]

From this simple picture some general rules can be derived:

1. Most core levels of metals are shifted to higher BE in oxides (or fluorides, chlorides, etc.) because of the high electron affinity of O (F, Cl, etc.). These shifts will usually increase with increasing formal oxidation numbers.
2. The carbon line is shifted to higher BE in oxygen-containing compounds and to lower BE in carbides.
3. The core-level shifts in similar compounds will decrease with higher atomic radius, i.e. higher Z in the same group of the periodic table.

However, for real-world samples the specific electronic and crystallographic structure of the compound must always be considered. Thus, in practice it is not generally possible (although it is proposed from time to time) to apply BE reference energy values from a standard sample with well-defined valence states to an unknown compound for quantitative chemical state characterization.

Another approximation for the characterization of chemical effects is the so-called (modified) Auger parameter method. By measuring an additional XAES level (the three electron energy levels participating in the Auger process should be indexed *lmn*) of the same element and doing some approximations,[27,28] the difference of this KE $E_k(lmn)$ and the measured core level BE $E_b(i)$ may be assigned to the parameter α (Equation 6):

$$\alpha = E_k(lmn) - E_b(i) \tag{6}$$

and the change in this parameter may be determined by Equation (7),

$$\Delta\alpha = 2\Delta R_{ea} \tag{7}$$

where ΔR_{ea} is the so-called extra atomic relaxation energy, where static and dynamic relaxation have to be considered.[9] This relaxation is associated with the polarization of a dielectric medium (here by the core-hole states), and thus the XPS experiment is closely connected with the chemical surrounding of the emitting atoms. However, the parameter $\Delta\alpha$ is not influenced by surface charging (section 4.3) or other factors affecting BE reference levels. This can be a useful method for fingerprinting sample states. In a so-called Auger parameter plot (Figure 4)[15] bonding types are characterized by clusters and may differ in both Auger parameter values and BE values.

For quantification of the described chemical shifts in sample series, mathematical methods may be used. The quantities of overlapping parts of core-level peaks may be estimated by peak-fitting procedures[29] which are based mostly on nonlinear least-squares (NLLS) algorithms, as summarized by Gans.[30] The starting parameters (number of peaks, energy separation, peak shape model, etc.) influence the results and require operator knowledge. Procedures are implemented in the manufacturers' measurement software and in standard graphic software, which is critically reviewed by Siegbahn (**X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction**). Figure 5 demonstrates such a fitting for the MoO_2 sample given in Figure 2. The peak area ratio ($Mo3d_{3/2} : Mo3d_{5/2} = 1 : 1.5$) and the peak separation (here 3.1 eV) are fixed for the calculation. Three 3d doublets, which arise from different oxidation states, are necessary for good reproduction of the spectrum.

Changes occurring in a series of measurements may be typically classified by mathematical methods as linear least-squares (LLS) fitting, principal component analysis (PCA) or factor analysis (FA).[29–33] The mathematical principles are summarized elsewhere.[34,35] The LLS algorithms use standard measurements or typical sample states included in the measured data set as target spectra, whereas FA (commonly based on PCA algorithms) can detect chemical changes without (or with little) a priori information about the sample. Figure 6 demonstrates this

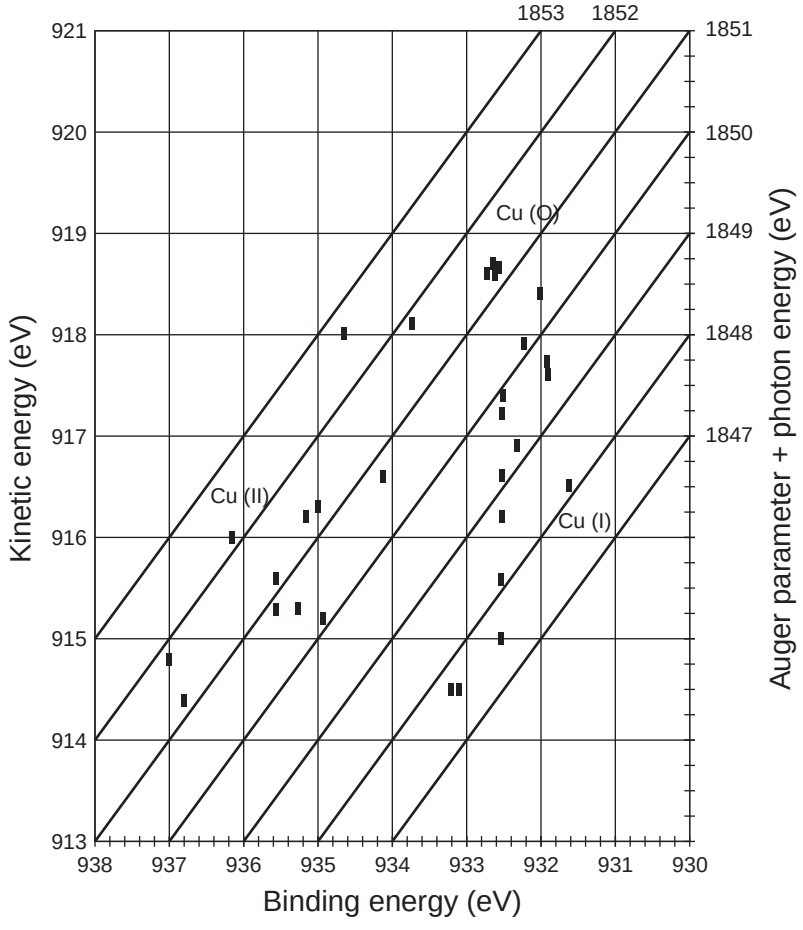

Figure 4 An Auger parameter plot for various copper compounds. Typical Cu valence states are shown. (Reproduced from Moulder et al.[15] by permission of Physical Electronics, Inc.)

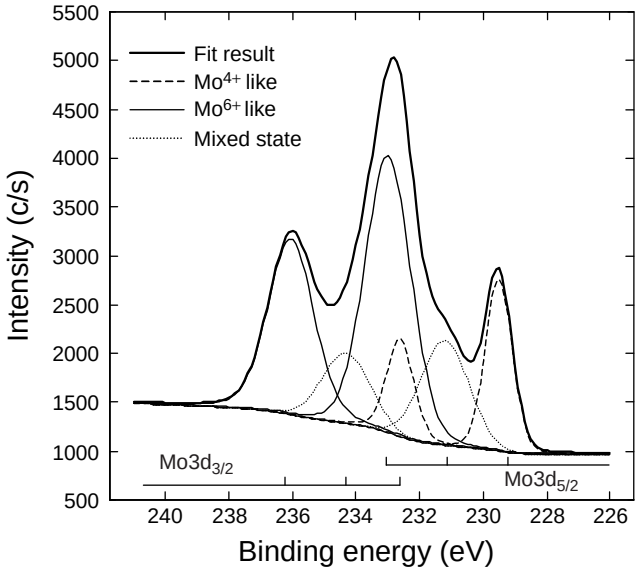

Figure 5 Demonstration of a multicomponent fit for the MoO_2 measurement given in Figure 2. Three 3d-doublet spectra are necessary for a good reproduction of the measured spectrum.

for a depth profile of an Re–Si–Re layer system.[36] Interface phases connected with silicide formation, mainly induced by the ion sputtering effects, can be derived by such a mathematical procedure. The calculated spectra are characterized by both changes in intensity ratios and small energy shifts.

3.3 Local Information

Local information from a surface analytical technique such as XPS can be both depth information (because of the small information depth) and lateral information (using spectrometers with appropriate lateral resolution). The depth profiling methods can be classified into destructive (using depth erosion methods and lateral resolving spectroscopy) and nondestructive methods (only in the near-surface region).

For destructive depth profiling, medium energy (1–5 keV) noble gas (Ar, Kr) ion sputtering is mostly used. Depending on the ion source and the necessary raster size (typically 2 mm × 2 mm, determined by the

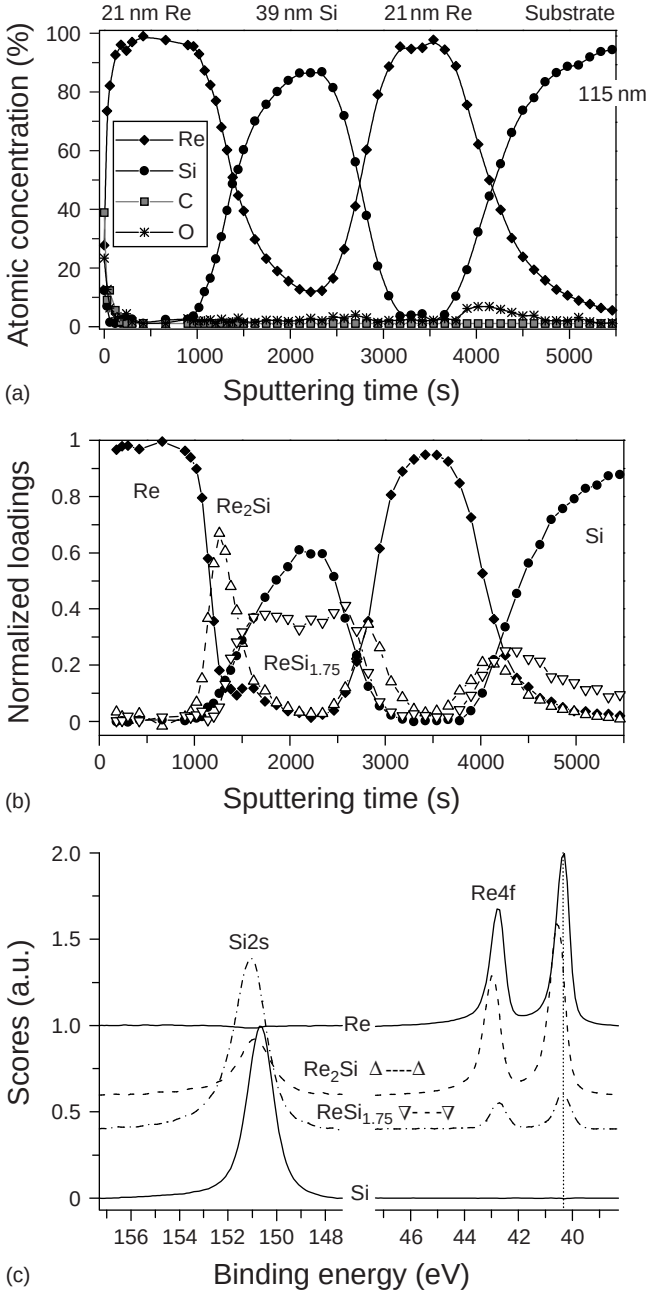

Figure 6 Depth profile of an Re–Si–Re–Si-substrate layer system (a) decomposed by FA into several phases (i.e. principal components). (b) The portions (loadings) of these several spectral components show strong changes in the interface regions. (c) The spectral shape of the components (scores) varies in peak energy/shape and peak area ratio and can be assigned to several silicide phases, in addition to Re and Si.

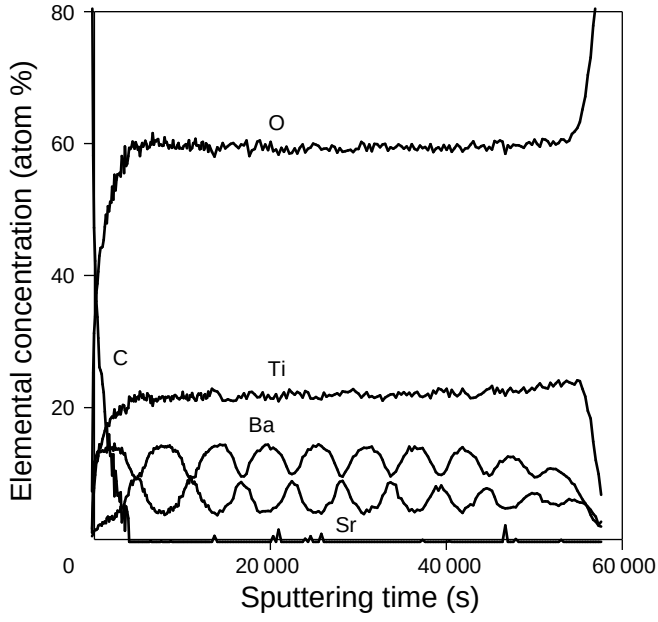

Figure 7 Kr^+ sputtered (1.5 keV ion energy) depth profile of a Ba-titanate–Sr-titanate superlattice. The double layer thickness of the nonconducting structure is only 21 nm.

lateral resolution of the spectrometer) the erosion rates are in the range of 1–10 nm min^{-1}. Thus elemental depth profiles can be determined by sequential measurements of elemental intensities after several sputtering depth intervals. The depth resolution is determined by the escape depth of the electrons and sputtering effects, and lies in the range of 3–5 nm. In practice, the measuring depth is limited to approximately 1 μm, because the total measuring time is not only determined by the time of the sputtering cycles, but mainly by the time for sequential detection of the electron spectra. Effectively it depends therefore on the sputter-cycles-per-depth density, the number of measured elements, the desired spectral resolution, and the detection limit. As an example, Figure 7 shows a depth profile of a nonconducting Ba–Sr-titanate superlattice.[37]

Depth calibration of the intensity versus sputtering time is difficult because the sputtering rates depend on both material (composition, microstructure) and the sputtering conditions (ions, angles, current density). The conversion of the sputtering timescale into a depth scale, if necessary, has to be done using depth (material)-dependent sputter rates for the particular experimental conditions.[38] Single-layer sputter rates are usually determined for layers of known thickness or from crater depth measurements by surface profilometry.[39,40] Alternative methodologies have also been discussed.[41,42]

One of the main problems of depth profiling is the sputtering artifacts, which include preferential sputtering, atomic mixing, recoil implantation, bond breaking, amorphization, and crystallization.[43–46]

Preferential sputtering arises from the different sputtering yields of the individual elements in multicomponent samples. As a result, the surface composition (which is measured by XPS) is altered and does not reflect the

bulk (or layer) composition. Atomic mixing and recoil implantation additionally change the concentration gradients in the region influenced by ion implantation (up to 10 nm depth depending on ion sputtering conditions). These three effects can be modeled with reasonable results by Monte-Carlo computer simulations for a wide range of materials.[47] The possible bond-breaking and structural changes limit the chemical information obtainable from depth profiling; however, in many cases ion depth profiling is the only method to get such information from thin layers or buried interfaces.

In summary, with ion sputtering the virgin (bulk) sample state is not necessarily being analyzed, and the nature and extent of damage strongly depends on the sample. Therefore, the influence of the sputtering effects has to be established for every new kind of material.

Nondestructive depth profiling techniques make use of the limited escape depth of the electrons (summarized in detail by Seah and Dench[48]) and are therefore limited to the first few nanometers of the surface region. However, this region is often of particular interest in surface science (contamination, segregation, passivation, adsorption, etc.).

Angle-resolved X-ray photoelectron spectroscopy (ARXPS) is the most used technique.[49–51] By sampling a series of measurements for different emission angles, the effective information depth d (for 95% of signal intensity) varies with the take-off angle β of the electron analyzer to the surface normal (Equation 8):

$$d = 3\lambda \cos\beta \tag{8}$$

The relative changes of peak area and peak shape give qualitative information on changes in both elemental composition and chemical bonding. Using appropriate surface layer models and mathematical algorithms one can estimate the surface layer structure.[52,53]

A second method uses changes in the electron background coming from the nonelastic scattered electrons.[54] The low-KE tail of a characteristic core level of one selected element is measured over a range of tens of electron volts, because these losses are characteristic of the depth distribution of this element in the surface region. Approximating the shape of this energy region with a computer program, by assuming a characteristic depth dependence (e.g. QUASES[55]), allows the estimation of some typical parameters (thickness, concentration, nature of surface layer growth) of the surface structures.

Lateral information is rarely used in XPS because of the relatively poor lateral resolution (50–1 μm, depending on the instrumentation) at relatively long measuring times. Several experimental principles have to be considered. Traditionally, for large-area X-ray radiation, either a small-area lens (with deflection unit) prior to the electron analyzer or an imaging analyzer with channel plates is used; currently, focused X-ray sources with beam scanning or mechanical movement of the sample are also applied. Thus, line-scan measurements or mapping of surface areas can be obtained. Although most of the interesting phenomena regarding modern nanostructured materials are associated with nanometer-scale dimensions, some applications arise from the possibility of chemical imaging, which is the detection of complete spectra at each measuring point and spectral interpretation with respect to bonding changes. This is demonstrated in Figure 8 for a map of laser-treated Si_3N_4 ceramics.[56] Spectral analysis by FA (section 3.2) shows that in addition to Si–N/Si–O bonding (not distinguished at the energy resolution used, which was optimized with respect to reasonable measuring time for the small spot measurements) in the center of the laser spots Si–Si bondings resulting from molten and decomposed Si_3N_4 were found.

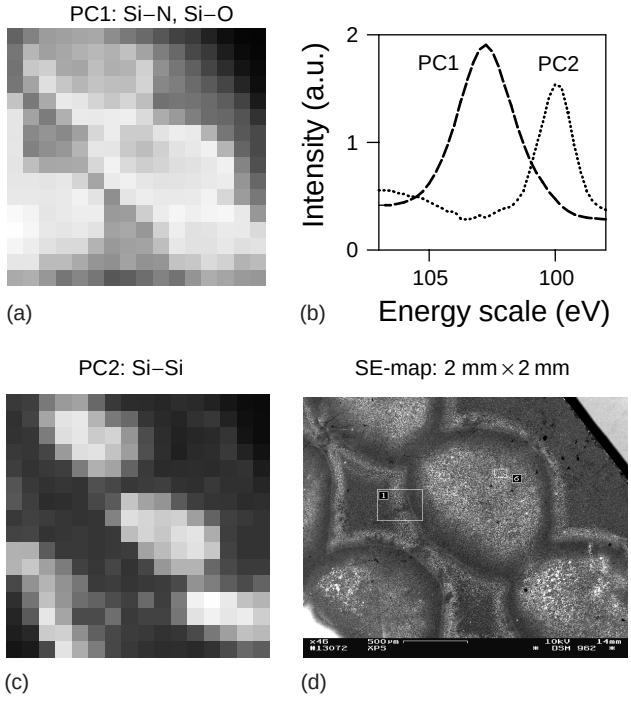

Figure 8 FA results from chemical mapping of Si2p spectra. The secondary electron (SE) micrograph (d) shows the investigated area of an Si_3N_4 sample with separated laser spots. The spectral components PC1 (oxidized Si, silicon-nitride) and PC2 (elemental Si) are characterized by FA (b). As shown in the lateral distributions of the two components (a and c), calculated from a 16 × 16 point XPS surface mapping, nonbonded elemental (Si–Si) silicon is found in the centers of the spots.

3.4 Valence Band Information

Although early XPS[1,11] work mainly dealt with the elemental and chemical information from core level

photoelectron lines, valence band photoemission studies soon became increasingly important.[3,5,57] This went hand in hand with the application of UV radiation as an excitation source, establishing the method of ultraviolet photoelectron spectroscopy (UPS). The valence band region (generally in the BE range from 0 to about 20 eV) covers the information about the outermost electron shell(s). The spectra in this region are usually rather complex, and therefore are less easy to interpret than the core-level peaks. In molecules the spectra reflect changes in the chemical bonding between the different atoms, of particular interest in organic materials.[58] In the solid state the valence region has a band structure and the spectra characterize the density of the occupied electronic states (DOS), of considerable importance in the investigation of electronic transport properties in materials. Such band structure investigations may be complemented by the method of inverse photoemission spectroscopy (IPES) that measures X-rays emitted from electrons excited to the unoccupied density of states near the Fermi level.[16] Valence band measurements with UV radiation are characterized by high excitation cross-sections, good energy resolution, and k (i.e. momentum) conservation (angular measurements at single crystals), by very high surface sensitivity (leading to problems with contamination), and by dependence on the unoccupied density of states at the final state energy.[5] For excitations with conventional high-energy X-ray sources the intensities and energy resolution of the valence band measurements are weak; however, a direct measurement of the DOS is possible that does not depend on final-state effects.[9] Because of the complexity of this topic only a small number of application examples are discussed in section 5.1.

4 APPLICATIONS FOR TYPICAL MATERIAL CLASSES

Many years of XPS usage have identified a broad spectrum of applications in nearly all scientific fields. The aim of this section is to give an idea of typical applications in several fields of natural and materials science, rather than a comprehensive account, together with some complementary techniques. Most examples are taken from recent work. Much of the pioneering work, taken from worldwide sources, has been well reviewed elsewhere.[3–5,8,16]

4.1 Metals and Alloys

Measuring the elemental composition of metal surfaces (layers or bulk material) is not a typical XPS task because elemental information for electrically conducting material can be obtained more quickly by other methods, such as AES. However, if chemical changes are of interest, XPS is of advantage, because the peak shifts are often much easier to interpret in terms of chemical bonding. Metals have been studied for a long time with respect of the chemical potential during alloying,[59–61] work function measurements,[5,62] and segregation phenomena,[63–65] chemisorption studies,[66,67] electrochemistry and corrosion science,[68–70] or fundamental studies in catalysis.[71,72]

4.1.1 Segregation

Segregation of surfaces or interfaces is a phenomenon that strongly influences a material's properties, such as brittleness or reactivity. Some XPS studies provide information about the segregant elements as well as information about their binding. Zhang et al.[73] have shown for a W–Ni system how a combination of ARXPS and depth profiling can help to understand the chemical nature of the segregated surfaces. Danoix-Souchet et al.[74] have observed Cu–Ni elemental and chemical changes during annealing – in addition to enrichment of the minor elements, a phase transition of the residual surface oxides from hydroxide to oxide could be found. Different $MoSi_2$ composite materials were studied by Yi et al.[75] by in situ fracturing. As shown in Figure 9, at the fractured surfaces an increasing amount of oxygen and Si–O bonding was observed. With information derived from scanning electron microscopy (SEM) of the fractured surfaces it was concluded that the grain boundaries were covered with SiO_2. Boron segregation, forming borides, was also found.

4.1.2 Interfaces

Thin interlayer regions or interfaces determine the properties of most layer systems and can be studied by depth profiling analysis. Metal–metal interfaces are often used in electronic contacts. Marinova et al.[76] showed an example for an Al–Ni–Al on SiC contact system where both the elemental changes and peak shape changes connected with phase formation were explained on the basis of depth profiling measurements. Working with Pd–Co superlattices, Lesiak et al.[77] determined the interface phases using a pattern recognition method.

For an Re–Si–Re–Si-substrate system, interface silicide phases were found and quantitatively extracted with help of FA,[36] as shown in Figure 6. These phases were produced mainly by the ion sputtering process because they were not found in transmission electron microscopy (TEM) cross-sections of the virgin sample.

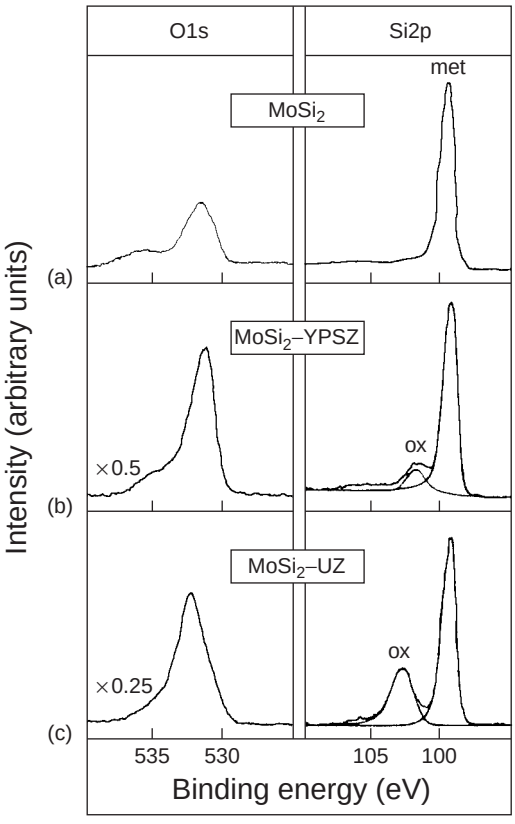

Figure 9 Spectra recorded at in situ fractured surfaces of several $MoSi_2$ composite materials: (a) pure $MoSi_2$, (b) and (c) with Zr oxide. The grain boundaries at (b) and (c) are covered with an SiO_2 film. (Taken from Yi et al.[75])

4.1.3 Corrosion

The investigation of corrosion is a typical XPS application because chemical information plays an important role. Research activity in this field ranges from basic research on well-defined surfaces to practical applications relating to construction materials. A review of the application of surface analytical techniques (including XPS) for corrosion investigation is given by Quaddakkers and Viefhaus.[78] The XPS technique is characterized as a well-established method for the detection of corrosion products. Also useful is AES if applied with the high-energy resolution and with the higher lateral resolution for local analysis. The application of XPS in corrosion science covering many metals has been reviewed by McIntyre et al.[79]

An example of a systematic study of the formation of passivation layers is given by Maurice et al.[80] for the Fe–^{18}Cr–^{13}Ni system. Starting from the analysis of the natural oxide layer, the passivation layer formed in an electrochemical cell, for various exposure times, was studied ex situ in the XPS apparatus. Both depth dependence of the layer by ARXPS and chemical states (with respect to oxide and hydroxide state) were estimated by curve fitting (Figure 10). The investigations were accompanied by AFM (atomic force microscopy) measurements of the surface structure.

4.1.4 Catalytic Reactions

Catalysis can be studied with XPS in UHV only remotely. Gas adsorption investigations are of principal interest in this context. The motivation behind the use of surface analytical techniques is that in heterogeneous catalysis the mechanisms of interest are surface reactions. Most investigations in this field are dedicated to the study of the basic reaction mechanisms of catalytic processes. However, there is a strong interest in getting closer to more realistic conditions, which will be discussed later, in section 5.4.

Adsorption studies are usually performed under well-defined conditions using single-crystal surfaces.[81–83] A typical example is given by Sandell et al.[81] with a study of adsorption and reaction mechanisms of several gases (CO, NO, O_2, CO_2) on an ordered Pd–Mn single crystal surface prepared by Mn deposition on a Pd(100) single crystal and subsequent heating. Figure 11 shows changes in the $Pd3d_{5/2}$ spectra during adsorption. The multiple components (see the curves at 20 L CO) were separated by peak fitting and assigned to several surface and bulk contributions.

Besides such adsorption studies, investigations using catalysts with more technological backgrounds were also done on precious metals (e.g. Pd[84,85]). From the technological point of view, oxide catalysts are of more interest and will be discussed later in section 4.3.

4.2 Semiconductors

Because of the very small lateral structures of semiconductor devices, XPS is not a favored technique for failure analysis or process control. Thus, AES, as a surface analytical method with high lateral resolution, was used early in this field.[86] XPS with its poorer lateral resolution is, nevertheless, used for investigations of new materials and their electronic properties, and for the development of new thin film deposition techniques. Some early examples are XPS studies of surface treatment[87–89] and of new device technologies.[90–92]

An important problem with semiconductors is the BE calibration. On one hand semiconductors have sufficient electrical conductivity for stable measurement using charged-particle methods, on the other hand the position of the Fermi energy E_F, which is the energy reference in XPS (section 3.1), is not always well defined. In the undoped semiconductor E_F lies in the middle of the band gap, whereas when doped E_F is shifted to the acceptor or donor states at the edge of the valence band or conducting

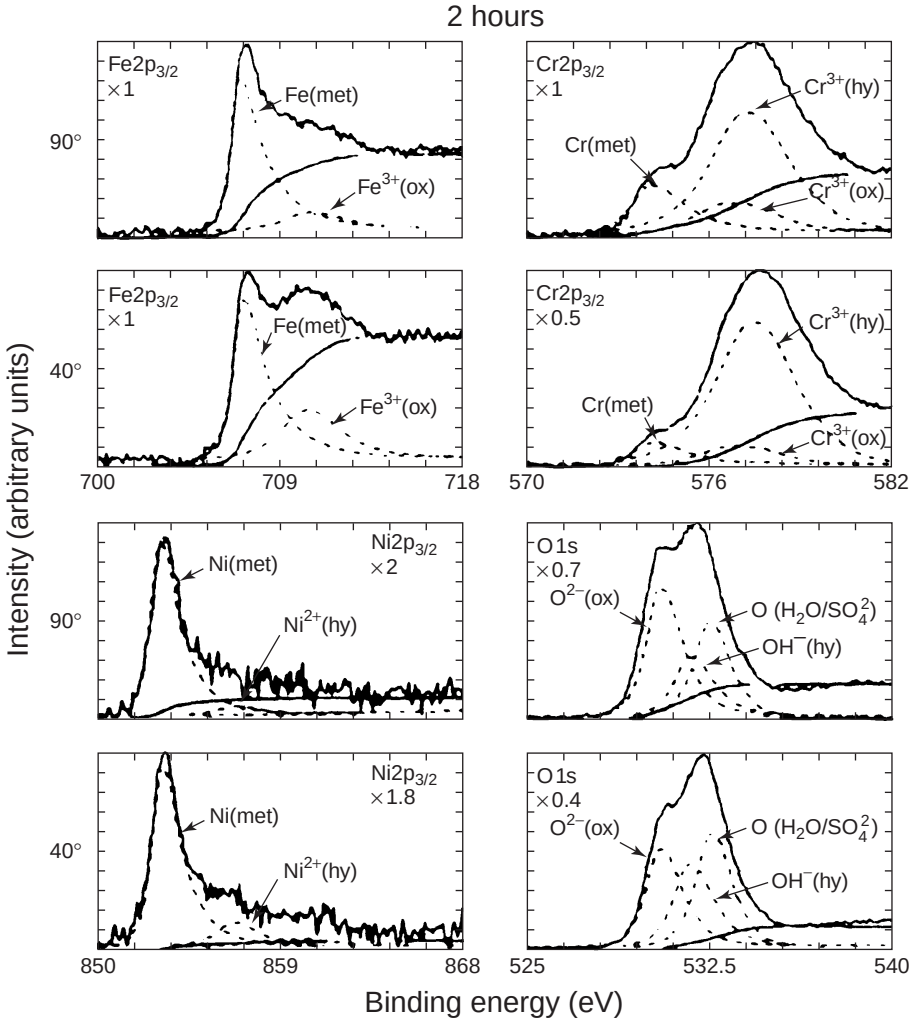

Figure 10 Spectra after 2 h of electrochemical passivation of an Fe−^{18}Cr−^{13}Ni sample. The ARXPS measurements, supplemented by the peak fit, show differences in the oxide/hydroxide ratio in the uppermost passive layer. (Reproduced from: V. Maurice, W.P. Yang, P. Marcus, 'X-ray Photoelectron Spectroscopy and Scanning Tunneling Microscopy Study of Passive Films Formed on (100) Fe−^{18}Cr−^{13}Ni Single-crystal Surfaces', *J. Electrochem. Soc.*, **145**(3), 909–920 (1998), by permission of The Electrochemical Society, Inc.)

band, respectively.[93,94] The degree of shift depends on doping concentration and temperature. As discussed in detail by Egelhoff,[95] when in contact with the metallic part of the spectrometer a (more or less well defined) Schottky contact may be formed; however, at the vacuum side of the sample (which is under investigation), band bending may occur which can considerably influence the measured BE. This problem has been discussed for B- and Al-doped SiC single crystals.[26] UHV fractured and sputtered surfaces there have been compared. Energy shifts that could be assigned to the influence of surface states occurring at the amorphized sputtered surfaces were corrected. Thus realistic BE values for Al and B could be obtained with respect to their chemical bonding in the SiC environment. It was shown that the Ar2p energy reference, from the noble gas atoms implanted by ion sputtering, was misleading in this case.

4.2.1 New Electronic Materials

The material SiC (discussed above) has been assessed as a candidate for high-temperature applications.[96,97] Also of interest are the potential and problems for studying its electronic properties by XPS. Other material, such as InP, AlGaAs, or GeSi, are studied typically with respect to their surface and interface reactions.[98–103] In InP, for example, the surface Fermi level is separated at cleaved surfaces for p- and n-type doping due to the Fermi level pinning.[98] Figure 12 shows how this changes as a result of exposure to several gases and subsequent heating.

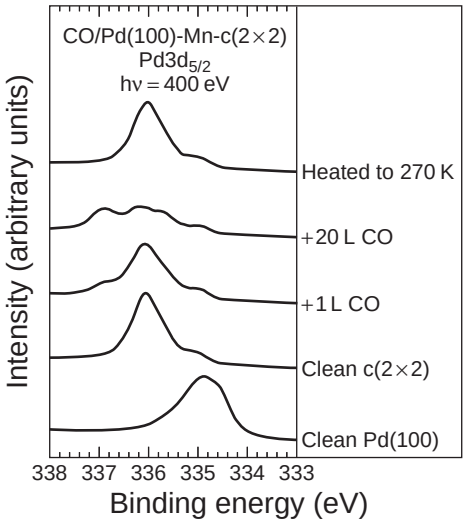

Figure 11 CO adsorption on Pd. The adsorption started from a clean Pd(100)-Mn-c(2×2) surface. This ordered surface structure is equivalent to the (100) surface of an ordered Pd_3Mn bulk alloy. Coverage using 1 and 20 L (1 L = 1 Langmuir, a dose of 10^{-6} Torr s) of CO completely changes the spectra. By heating, the CO coverage can be removed. (Reprinted from: A. Sandell, A.J. Jaworowski, A. Beutler, M. Wiklund, 'Adsorption and Reactions on a Surface Alloy: CO, NO, O_2 and CO_2 on Pd(100)-Mn-c(2×2)', *Surf. Sci.*, **421**, 116–134 (1999), with permission of Elsevier Science.)

Chemical core level shifts after treatment with S_2Cl_2 were examined by Peisert et al.,[99] including the use of the Auger parameter method. Figure 13 shows peak shape changes during heating that have been estimated by peak fitting.

Investigations of new deposition methods[104,105] allow the study of film formation or new interface properties.

4.2.2 Contact Structures

These play an important role in determining electrical properties and microelectronic technology. Rinta-Möykky et al.[106] showed the selective oxidation of Te at the interface for a PdTe alloy contact on ZnSe during interface formation, i.e. during its preparation. A Pd/SiC Schottky structure was measured with depth profiling, including peak shape analysis.[107] Silicide formation could be observed during heating.

4.2.3 Surface Contamination

The characterization and minimization of surface contamination is also an application of XPS. Moon et al.[108] studied a cleaning procedure for organic contamination. The effectiveness of the cleaning process, especially when using UV radiation, can be observed through analysis of the C1s peak (Figure 14).

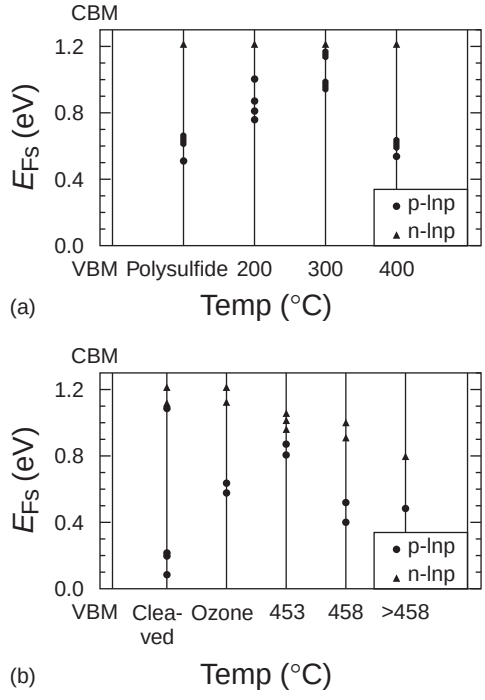

Figure 12 Changes in surface Fermi level for p- and n-doped InP. Differences for (a) hydrogen polysulfide/hydrogen sulfide mixture and (b) ozone exposure are shown. (Reprinted from: W.M. Lau, R.W.M. Kwok, S. Ingrey, 'Controlling Surface Band-bending of InP with Polysulfide Treatments', *Surf. Sci.*, **271**, 579–586 (1992), with permission of Elsevier Science.)

The XPS technique has been used to optimize etching processes so as to obtain low defect densities,[109] and for studying Fermi level pinning for doped Si wafers.[110]

Quantitative analysis of such thin overlayers and of the material underneath is assisted by mathematical methods. These have been further developed for ARXPS, including the effects arising from surface topography.[111–114]

4.3 Insulators

Surface analytical techniques that use charged particles, such as AES or SIMS (secondary ion mass spectrometry), generally do not work for insulators because their electrical nonconductivity results in surface charging. However, XPS uses electromagnetic waves for excitation and is therefore favored for the investigation of surface properties of insulators. These insulators include inorganic material such as various oxides,[115,116] oxidic catalysts,[71] glasses, and ceramics.[117]

The emitted electrons will also lead to surface charging of the sample, a feature that has to be considered in any investigation. In most cases, however, the use of nonmonochromatized X-ray excitation leads to relatively small and stable energy shifts at the BE scale. This is

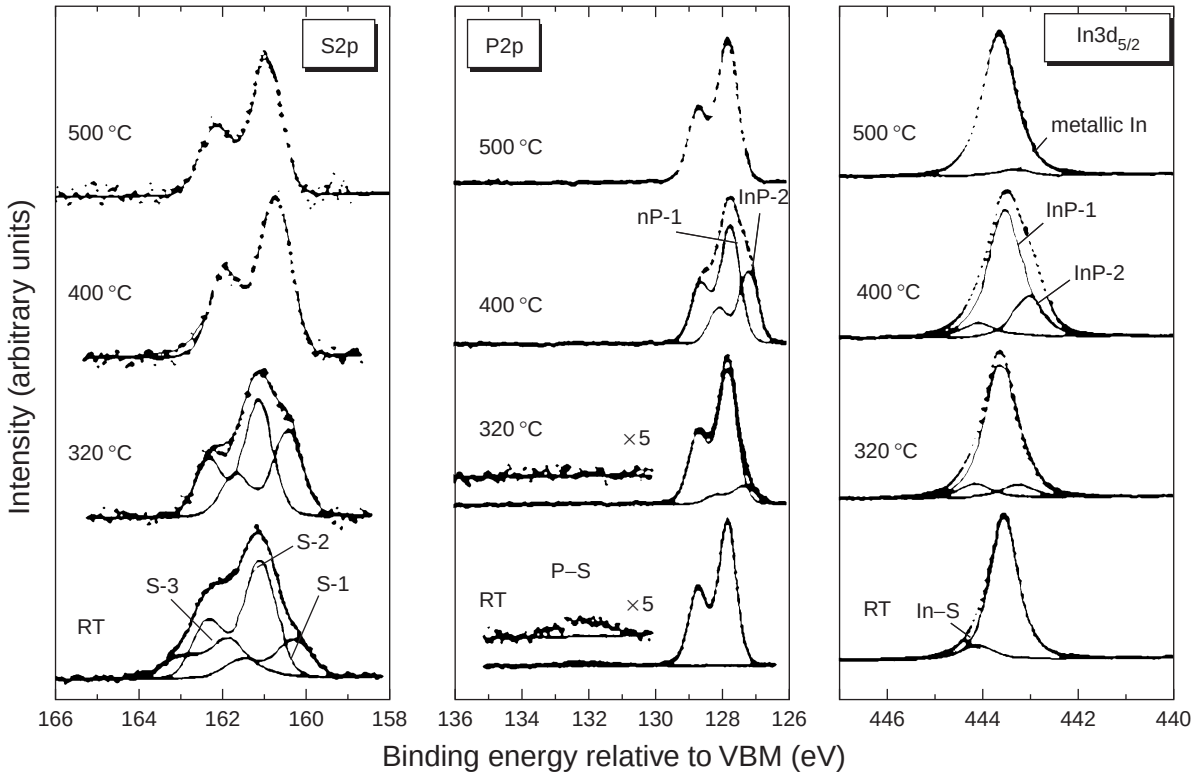

Figure 13 Results of surface treatment of InP(001) with S_2Cl_2 at room temperature (RT) and with subsequent heating. The BE scale is referenced to the valence band maximum (VBM) and not, as usually, to E_F to minimize the influence of band gap changes on the results. (Taken from H. Peisert et al.[99])

because secondary electrons emitted from the X-ray tube effectively stabilize the surface charging situation. External flood guns, which produce low-energy electrons, can be used for charge stabilization. This is necessary especially when using monochromatized X-ray radiation.[118] This approach also minimizes the shift observed when using nonmonochromatic radiation. Another method is to use a metallic grid (e.g. from TEM) or metallic aperture on top of the insulating sample for charge exchange. The problems with surface charging are complex, and depend on sample geometry, differential charging, and peak shape changes.[119–121]

For quantification of chemical information, the residual shift of the BE scale has to be corrected with respect to a reference BE. There are several traditional methods:[8] (a) adventitious carbon contamination, C1s at 285 ± 0.2 eV, although there can be relatively broad peaks from various C–X species;[122] (b) internal reference lines of well-defined energy; (c) implanted noble gas, mostly Ar, although errors may arise from electronic interaction with the surrounding chemical environment; and (d) artificial mixtures/overlayers, where the problems are homogeneity, contamination, and particle size. A promising method is the deposition of Au nanoparticles from a suspension onto the surface (Au4f = 84.0 eV).[123] Such referencing techniques have also been successfully used for automatic shift correction of measurement series[124] (such as depth profiles with varying surface conductivity) prior to further data analysis by FA or LLS.

Insulator analysis is a wide-ranging application area, because of the many material classes involved. It includes materials with covalent bonds (e.g. oxides, nitrides, and carbides) and with ionic bonding (e.g. chlorides and fluorides).

4.3.1 Catalysts

Catalysts, discussed in section 4.1, are often applied as their oxides. Mathematical methods (curve fitting, FA, deconvolution) are used for quantification of small peak shifts or peak shape changes in the XPS spectra.[125–128]

In many investigations standard spectra are compared with spectra of chemical states formed during treatment of the catalytic surfaces. Davidson et al.[125] investigated changes of Ni oxide species on Ce-based material in comparison with NiO, Ni_2O_3, and $Ni(OH)_2$ standard materials. A transformation from $Ni(OH)_2$ to NiO during heating in air was observed. The chosen BE reference was C1s at 285 eV.

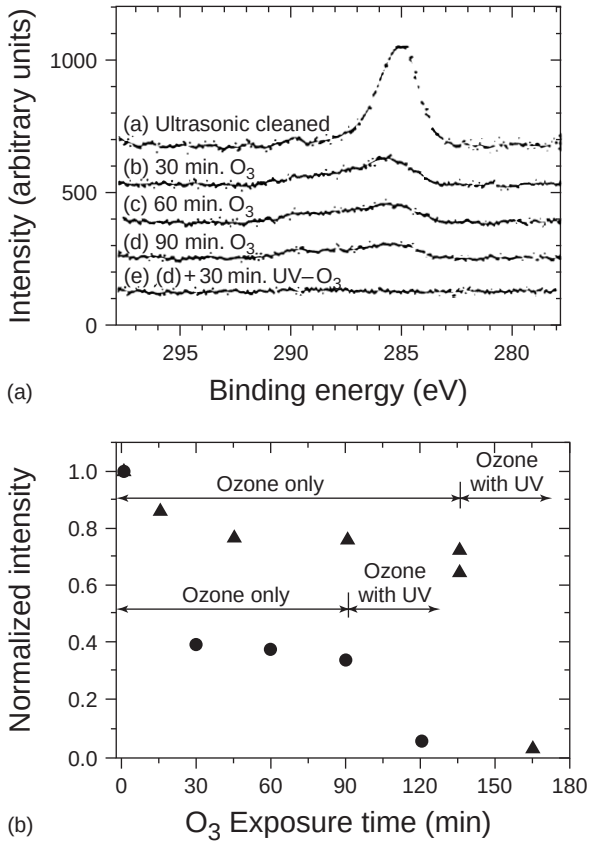

Figure 14 Efficiency of ozone surface cleaning for hydrocarbon contamination of Si wafers: (a) spectra of residual C1s contamination, (b) C1s intensities for several exposure conditions. (• = taken from experiment (a), = experiment after a period of one year with another wafer). The positive effect of UV radiation is clearly shown. (Reprinted with permission from: D.W. Moon, A. Kurokawa, S. Ichimura, H.W. Lee, I.C. Jeon, 'Ultraviolet-ozone Jet Cleaning Process of Organic Surface Contamination Layers', *J. Vac. Sci. Technol. A*, **17**(1), 150–154 (1999), Copyright 1999, American Institute of Physics.)

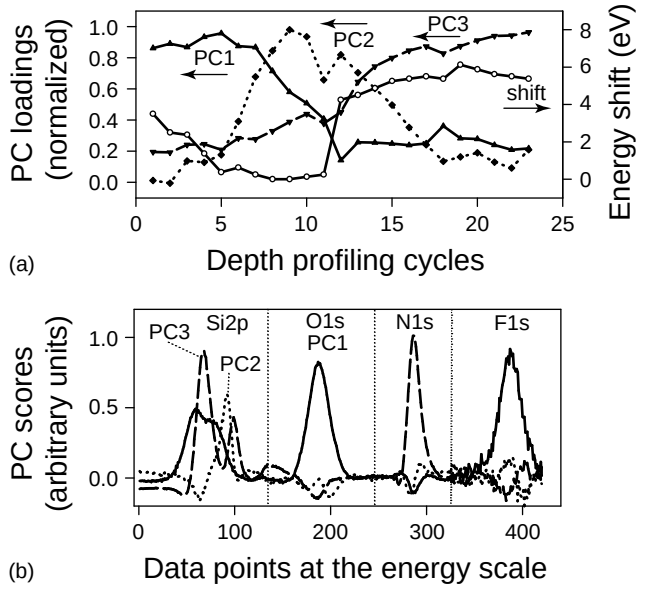

Figure 15 Results of FA at depth profiling of a laser-treated Si_3N_4 ceramic sample. (a) The shift necessary for charge correction in connection with the chemical depth profile of the three identified compounds (PC1–3). (b) The spectral shapes of these components point to the following identification: PC1 = surface oxifluoride, PC2 = intermediate Si–Si phase, PC3 = bulk Si nitride.

4.3.2 Ceramic Materials

These are investigated to identify changes in their surface properties. Figure 15 shows a laser-ablated surface of an Si_3N_4 ceramic sample. In this example large-area Nd : YAG radiation in a halogenic atmosphere was being tested for future micromachining applications. A depth profile of the changed surface region of about 2 µm depth was measured. Because of changing conductivity, different peak shifts were observed during depth profiling. An automatic correction with respect to Ar2p was done, using the method proposed by Oswald et al.[56,124] which confirms peak shifts of up to 6 eV. The result, which was derived by FA, shows three compounds with different depth profiles: PC1 is surface oxifluoride, PC2 is an Si-rich interlayer formed from molten and decomposed silicon nitride, and PC3 is bulk silicon nitride.[56] However, the resulting spectra (scores) do not have ideal peak shapes because of the relatively high charging shifts, in which there is a possibility of differential charging. The decomposition of silicon nitride as observed in PC2 was also found during local laser radiation in the center of the laser spots. A chemical map of a sample area about 2 mm × 2 mm measured with 100 µm lateral resolution is shown above in Figure 8 and discussed in section 3.3.

Ion-implanted silicide ceramic has been investigated to detect modification of its tribological properties.[129] Line scan and depth profile measurements established that the wear reduction mechanism is not solid lubrication, as previously assumed, but comes from structural changes in the surface region.

4.3.3 Oxide Materials

Mostly oxides are insulating materials. However, the electrical properties of oxides vary, depending on structure, from wide-gap insulating, to semiconducting, to superconducting properties at low temperatures.

As an important dielectric material silicon oxide has been studied with respect to it's surface stoichiometry.[130] Several methods (peak area calculation, lineshape fitting, Auger parameter) were compared, showing the Auger parameter method as being easy, fast and very reliable. This fact has been confirmed for many oxide systems.[131,132] Cluster formation and oxidation of Ge,

implanted in silicon dioxide films after thermal treatment, were studied by XPS and the redistribution processes followed by the characteristic peak shape changes that occurred.[133] These were probed further by FA.[134]

Stoichiometric SiO_2 is a very stable oxide and is used as a passivation layer. It is also very stable during ion sputtering procedures, whereas other oxides show marked changes after ion sputtering. Okude et al.[135] demonstrated this for tin oxide thin films. Using ion beam sputtering with an oxygen beam, however, XPS has proved that stoichiometric SnO_2 films can be prepared.

As a semiconducting oxide doped with other components, SnO_2 is used for electrical applications such as gas sensors and indium–tin oxide (ITO) transparent contacts. Examples of gas sensors are given by various authors.[136–138] Figure 16 shows XPS results[137] for a Pt-doped material. Peak fitting procedures show that several compounds are formed at the surface, changing the electrical properties.

It is necessary to know the formal valence states of oxides in order to study their properties and structure. To derive this information from XPS measurements, many problems have to be considered. Although BE reference values from external standards[139,140] can be used, there may be variations arising from the BE not being in correlation with a formal chemical valence.[115] Another problem is that the valence states at the sample surface may differ from bulk, as a result of contamination or preferential oxidation.[141,142] By way of illustration, Figure 17 shows spectra for Mo compounds prepared under different conditions. The species Mo^{6+} and Mo^{4+} in the MoO_3 and MoO_2 (scraped) samples are well defined. In the $Mn_2Mo_3O_8$ (scraped) sample Mo should also be Mo^{4+} ($2MnO-3MoO_2$) but the peak is significantly separated from MoO_2 due to the MnO environment. It may be that this is an intermediate oxide state that also occurs at the MoO_2 surface (shoulders to high BE side). Rapid oxidation to Mo^{6+} is found after short exposure to air of the Mo^{4+}-containing samples.

Barr[10] has shown that analysis of the oxygen O1s peak shift can give good information about the metal–oxygen bonding. This is also true for O2s information about TiAlO films.[132]

Complex oxides are of interest for new electronic applications and as oxide superconductors. The depth profile of the Sr–Ba titanate superlattice shown in Figure 7 can only be measured while using a charge compensation (here a low-energy electron flood gun). Despite the sputtering effects, relatively sharp depth profile tails point to sharp interfaces. Calculations[37] show that the interface roughness is about 2 nm. High T_c superconductors (especially with a critical temperature, T_c, higher than 77 K, the boiling point of liquid nitrogen) are materials attracting much basic research

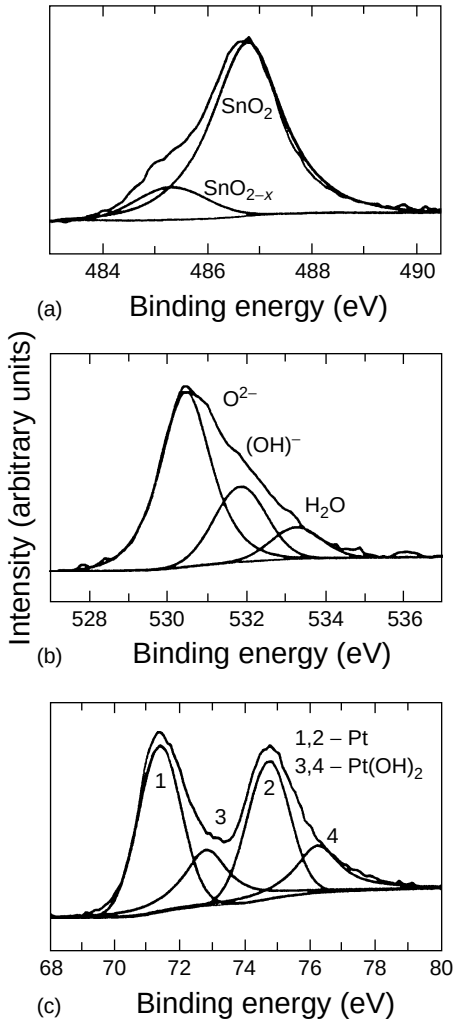

Figure 16 Spectra of a Pt/SnO_x film prepared by dc magnetron sputtering. Synthetic peak fit spectra are given in comparison with the measured spectra. The peak fit results at O1s peak (b) and the Pt4f peak (c, peak doublet 3,4) show that hydroxide formation takes place at Pt. At the O1s peak some adsorbed water and an O^{2-} component attributed to SnO_2 were also found. (Reprinted with permission from: G. Gaggiotti, A. Galdikas, S. Kaciulis, G. Mattogno, A. Setkus, 'Surface Chemistry of Tin Oxide Based Gas Sensors', *J. Appl. Phys.*, **76**(8), 4467–4471 (1994), Copyright 1994, American Institute of Physics.)

and are of technological significance. Padalia et al.[143] have reviewed XPS studies of YBCO oxide superconductors. The main topic of these investigations is the detailed study of the Cu2p spectra including the satellite structures typical of the active Cu–O superconductor layers in comparison to Cu oxide standard samples.

In this context it should be mentioned that radiation damage may also occur during XPS measurement of particular samples (see section 4.4). This may arise

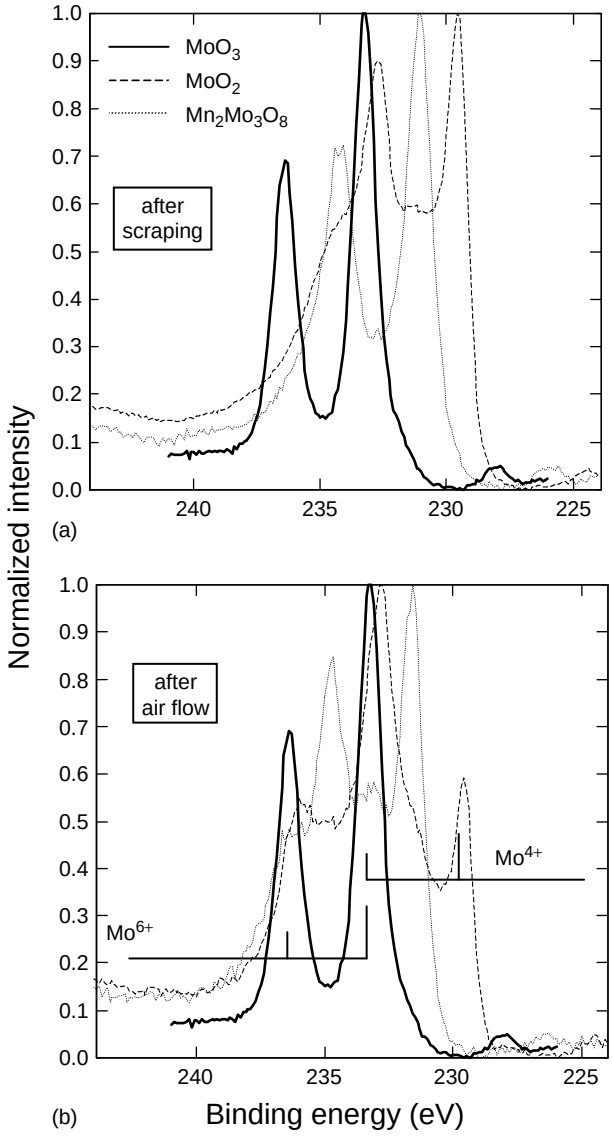

Figure 17 Spectra of different Mo-containing oxide single crystals (a) scraped under UHV conditions and (b) after short treatment in air. Relatively well defined valence states (the sharp peaks) are found on the scraped samples. Oxidation to Mo^{6+} occurs immediately in air.

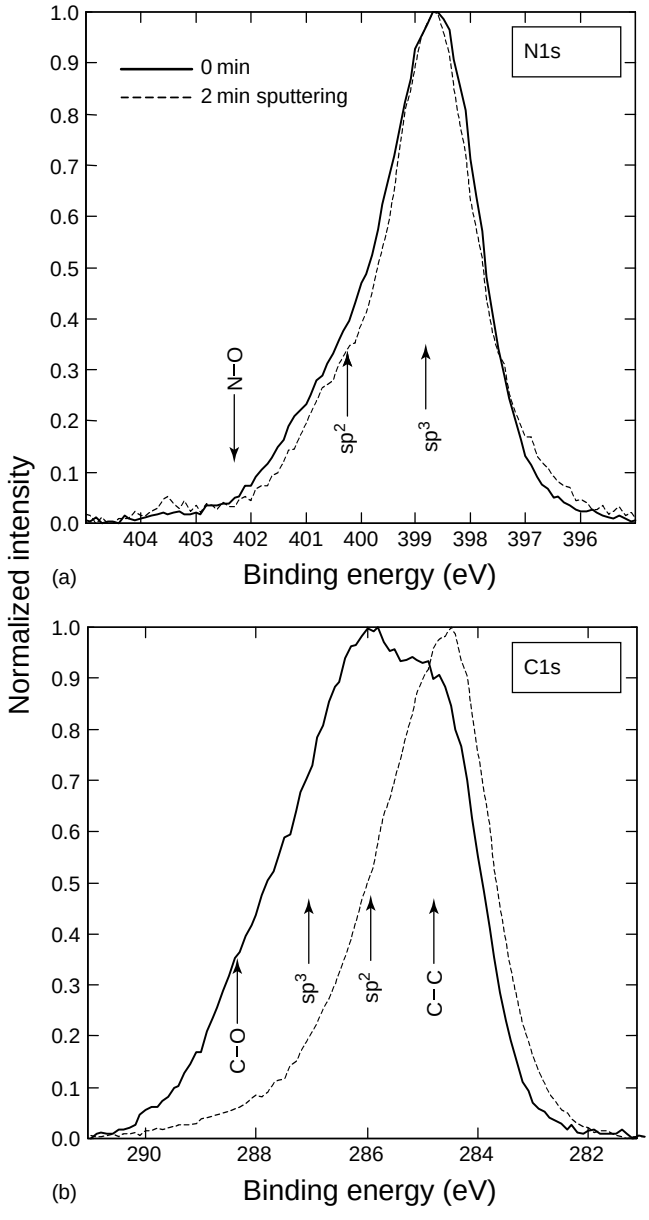

Figure 18 Spectra of sputter deposited CN layers (surface composition C : N ≈ 1 : 1) before and after low-dose Ar^+ ion sputtering for surface cleaning. The N1s spectra (a) and C1s spectra (b) are normalized and thus do not represent the true intensity relations. The sputtering leads to a strong decrease of N (to 20 atom %) but the N1s peak shape is not influenced. Strong peak shape changes are found for C1s. Whereas in the as-deposited state a mixture of C–C, sp^2, sp^3 and C–O (from surface contamination) bonds with a preference for sp^2 was found, drastic changes occur after sputtering and the surface is dominated by the C–C bonds.

from both secondary electrons from the X-ray window (nonmonochromatized source) and from photoelectrons and Auger electrons emitted from the sample (also with monochromator). Iijima et al.[144] discussed this for the partial reduction of CuO surfaces in different experimental conditions.

4.4 Polymers and Organic Materials

The materials discussed here can be insulating, semiconducting, or electrically conducting. The central topic is material systems based on carbon and its compounds. This covers a wide range of materials, including new polymers widely used as construction and functional materials, organic materials including problems of biocompatibility, and new inorganic carbon-based materials such as diamond or a : C–H (hydrogenated amorphous carbon)

layers, carbon nitride or the conjugated carbon systems (fullerenes, nanotubes, etc.). Because of the relatively low damage by X-rays to polymers or organic materials, XPS was applied to this field early in its development. Clark et al. did a lot of experimental and theoretical work in the 1970s (e.g. on core levels[145]) which covered a wide range of materials (summarized by Clark[146] and by Dilks[147]). The usefulness of valence band information to distinguish different functional groups was also recognized early on.[148–152] In biology and medicine XPS also found application.[153–155]

Investigations of changes in the C1s peak are the main interest. During the early stages of XPS development[21] a wide spread of BE was reported,[15] from 281 eV (carbide) to 293 eV (fluorine). However, these material systems consist mainly of the light elements (C, H, N, O, etc.) and thus are sensitive to radiation damage.[58] Toth et al.[156] discussed this for polyvinyltrimethylsilane with respect to surface ion cleaning and electron flooding for charge compensation. The example in Figure 18 shows results from CN layers demonstrating that only a very small ion dose (Ar^+, 3.5 keV, 2 min, 0.3 μA, 2 mm × 2 mm) leads to drastic peak changes and a strong degradation of nitrogen.

If handled carefully, XPS and UPS provide one of the most powerful methods for analysis of such sensitive materials, in contrast to electron beam techniques such as AES and EPMA (electron probe microanalysis). Because of its surface sensitivity XPS is mostly used for investigation of surface treatments or interface properties in compound materials. An overview of XPS applications to polymers up to the late 1980s is given by Briggs.[58]

4.4.1 Polymers

Polymers and their surface treatments is one main area of XPS application. The cleaning of polyester products is discussed by Briggs et al.[157] Figure 19 demonstrates how the complex C1s spectrum of a commercial product can be fitted to the base material spectrum of the polyester and to the spectrum of residuals of a Permalose polymer. This was also confirmed by time-of-flight/secondary ion mass spectrometry (TOF/SIMS) measurements. The thickness of the Permalose contamination is estimated to 2.4 nm from attenuation length consideration.

Similar surface modification experiments have been published with respect to plasma polymerization,[158]

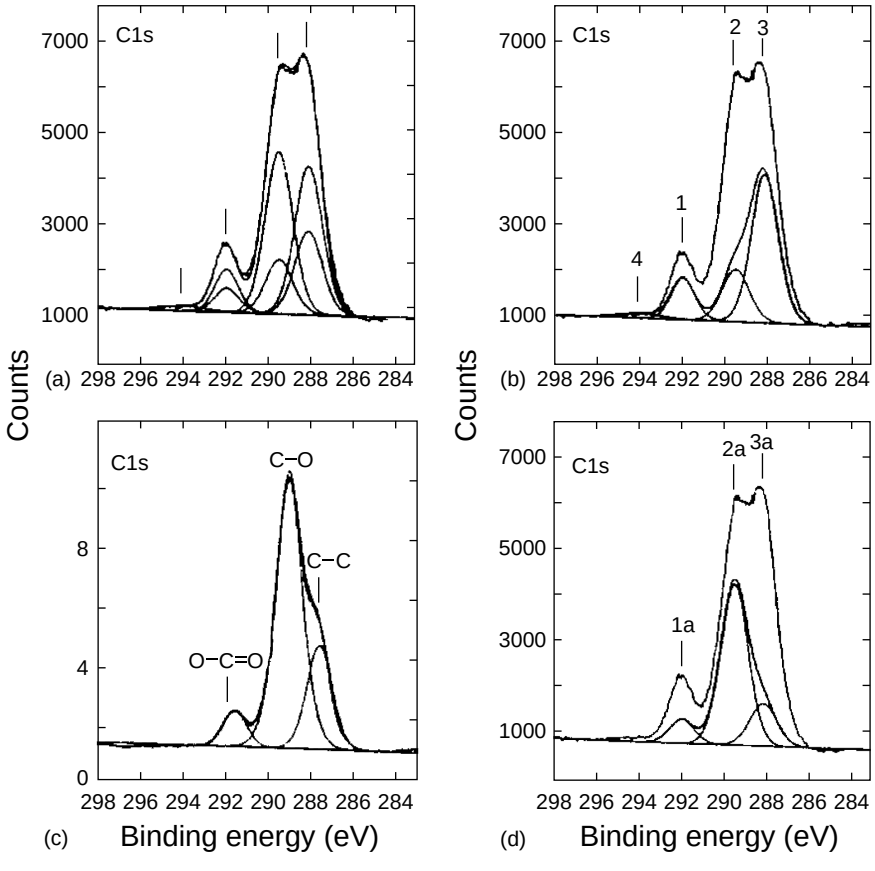

Figure 19 (a) Fingerprint spectra of a polyester-based fabric. The peak overlay procedure shows it contains both (b) the PET base spectrum and (c and d) Permalose polymer spectral components. (Taken from Briggs et al.[157])

surface segregation[159] and laser treatment[160] for different polymers. Mähl et al.[161] have dealt with the fitting of multicomponent C1s peak structures to polymers. Polymer–metal interfaces are of growing interest. During evaporation of aluminum on polyaniline films Lim et al.[162] studied the reaction mechanism by evaluating peak shape changes of the N1s peak. Similar work has been done for the metals Cu, Ti,[163] Cu, Fe, Al,[164] Al,[165] and Cr.[166] Riga et al.[167] used valence band spectroscopy to identify the specific structure of well-oriented polyethylene mats after adsorption on Au. Structural information can also be derived for Langmuir–Blodgett films.[168] Adhesion phenomena are closely connected with the interaction of organic polymers and metal oxides.[169]

Conducting and doped polymers have been studied by valence band XPS or UPS to establish changes of electronic structure and thus conducting or doping mechanisms.[170–172] Salanek[173] has summarized investigations on conjugated polymers. As an example the UPS valence band spectra (see section 5.1) of Figure 20 show how the electronic properties of a polymer may be controlled by sodium doping. The charge carrier density near Fermi energy is increasing with dopant concentration, leading to higher conductivity.

The possibilities of chemically imaging for polymers were demonstrated by Forsyth and Coxon.[174] Using the different C1s energies for C–C, C–O, and C–F as the source for mapping, a surface contamination of polytetrafluoroethene could be detected with a lateral resolution below 5 μm.

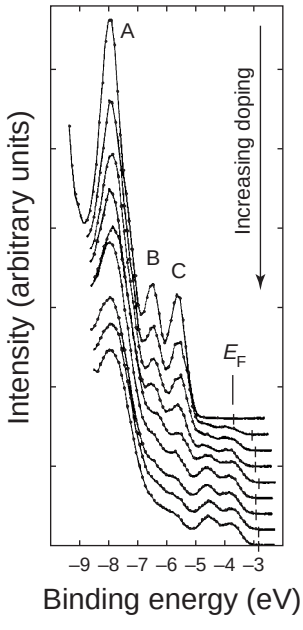

Figure 20 Valence band spectra of an Na-doped polymer. Increasing doping leads to a higher DOS at the Fermi edge and a higher conductivity. The energy scale is calibrated with respect to the vacuum level by comparison with spectra of the free polymer molecules considering the typical spectral features (A, B, C). Therefore, changes in E_F can be observed. (Taken from Salaneck[173].)

4.4.2 Biomaterials

Investigations in this area mostly include the problem of dehydration in a vacuum. Little reported work deals with the measurement of hydrated samples at low temperatures.[58,175] More successful are investigations into naturally dehydrated materials, such as wool,[176] bone, teeth, or cellulose-containing material from plants including wood.[177] Also demonstrated in this work was the combination of XPS and TOF/SIMS, which is often used for polymer analysis.

The chemical changes of flax fibers after NaOH treatment have been studied.[178] The flax fibers were fixed on the sample holder for analysis with UHV-compatible adhesive tape; however, a longer outgassing period was necessary. Strong oxidation with a peak shift to C–O was found. Other examples in this field deal with proteins on surfaces[179,180] and aerosol particles.[181]

Many applications concern biocompatible surfaces[182,183] because of their increasing importance for modern medical implants. The sol–gel technique seems to be promising for oxide interlayer deposition on Ti. Cirilli et al.[184] have compared sol–gel prepared and conventional plasma-sprayed Ti oxide layers. By studying the O1s peak shape it was found that in the sol–gel prepared oxide layer fewer OH and H_2O components are incorporated, which indicates that these films consist of better defined stoichiometric oxides.

4.4.3 Inorganic Carbon-based Materials

The most studied feature here is the ratio of sp^3 to sp^2 hybridization, because this controls the material structure. As demonstrated by Meral et al.,[185] for diamond-like carbon films using XPS with high-energy resolution, the sp^2 (284.4 eV, graphite) and sp^3 (285.2 eV, diamond) spectra may be separated directly and quantified by curve fitting. Figure 21 shows that the proportions of two components depend on the power density used for the laser deposition technique. This separation was used to detect the sp^3 content, which had to be maximized for optimum layer properties. For the carbon nitride system with respect to the theoretically predicted C_3N_4 phase, the sp^2 to sp^3 ratio of the C–N bondings plays an important role (sp^3-like carbon bonds are expected in C_3N_4). As concluded by Boyd et al.[186] for the C–N system, the sp^2 (C1s = 285.9 eV, N1s = 400.6 eV) and sp^3 (C1s = 287.0 eV, N1s = 398.8 eV) bondings are also well separated. From Figure 18, it can be seen that this reactively

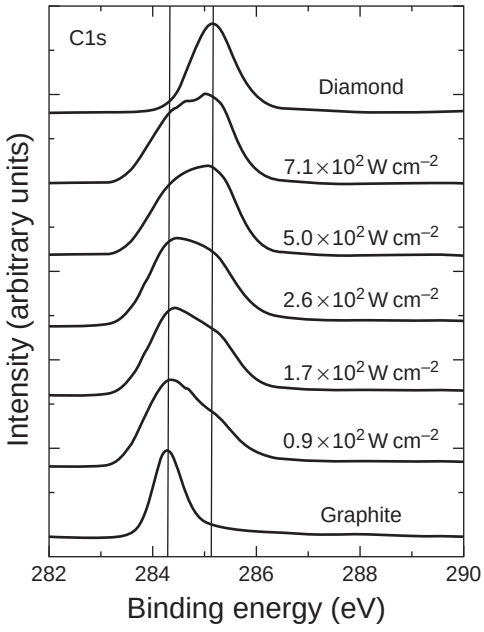

Figure 21 C1s spectra of laser-deposited diamond-like carbon films compared with graphite and diamond. With increasing laser power density a higher quantity of sp^3 bonds are found in the films. (Reprinted from: P. Merel, M. Tabbal, M. Chaker, S. Moisa, J. Margot, 'Direct Evaluation of the sp^3 Content in Diamond-like-carbon Films by XPS', *Appl. Surf. Sci.*, **136**, 105–110 (1998), with permission of Elsevier Science.)

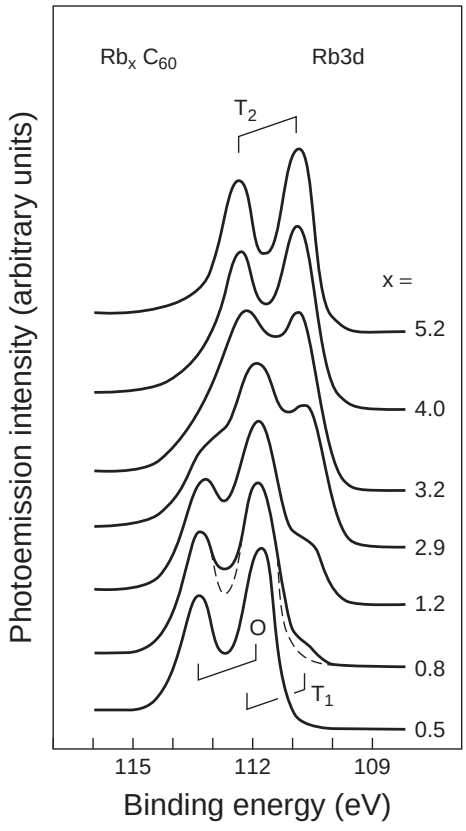

Figure 22 Spectral changes of Rb3d-doped fullerenes depending on Rb concentration, as demonstrated by the coordination geometry: O = fcc octahedral sites, T_1 = tetrahedral sites, T_2 = body centered phase. (Reprinted with permission from: D.M. Poirier, T.R. Ohno, G.H. Kroll, P.J. Benning, F. Stepniak, J.H. Weaver, 'X-ray Photoemission Investigations of Binary and Ternary C_{60} Fullerides of Na, K, Rb and Cs', *Phys. Rev. B*, **47**(15), 9870–9877 (1993), Copyright 1993, the American Physical Society.)

sputtered layer consists mainly of sp^2 carbon–nitrogen bonds and that the degradation due to sputtering leads to a substantial increase of C–C bonding by preferential sputtering of N.

One area of basic research is connected with conjugated carbon systems, namely the fullerenes, nanotubes, and onions (so-called because of their appearance as multiwalled carbon spheres). Photoelectron spectroscopy is used to study changes in the electronic structure during doping or phase formation. Using core-level measurements, Poirier et al.[187] observed structural changes depending on dopant concentration. Figure 22 demonstrates this feature for Rb doping.

Valence band measurements are used to study the occupied states near the Fermi energy. For RbC_{60} changes were found in the lowest unoccupied molecular orbital (LUMO) states, depending on phase structure.[188] Additional studies[189–191] combining photoemission measurements with electron energy loss spectroscopy (EELS) are providing new information about both occupied and unoccupied electronic states in this new class of materials.

4.4.4 Spectral Databanks

Spectral databanks should become useful for polymers and organic materials because of the great variability in their structure and chemical bonding (i.e. for fingerprint investigations). As already in place for static SIMS for several years,[192] collections of XPS spectra are now being offered. However, the expectations should not be too high because the spectral shape, especially for complicated systems, depends on many experimental factors, such as X-ray source parameters, take-off angle, spectrometer type and pass energy, and BE reference/charge neutralization. The reliability of such systems depends on the analytical problem and has to be assessed by the individual user.

For some time XPS spectra, including sample and measuring parameters, have been published periodically in a specialist journal.[193] A database specializing in organic polymers has also been published.[194] Two commercial databases in electronic form come from the National Institute of Standards and Technology, USA[195] (with free access after year 2000), and from

XPS International, Japan.[196] Most activities also are offered on the Worldwide Web: (ojps.aip.org/ssso/), (www.nist.gov), (www.xpsdata.com), (www.uwo.ca/ssw), (www.specs.de); naturally, this trend will increase in coming years.

A free database from the Surface Analysis Society of Japan is available, including a data processing system from the Internet (sekimori.nrim.go.jp).[197] This database can be complemented with own spectra. In this context it should be also mentioned that there exists a project for implementing a common data format[198,199] for the transfer of spectral data between different machines in surface science.

5 SPECIAL METHODOLOGIES

5.1 Valence Band Spectroscopy

The method of valence band spectroscopy (section 3.4) is an integral part of the XPS method. However, it has a particular role in terms of chemical analysis, because chemical bonding is directly involved in the valence band structure and the valence spectra are quite complex. Therefore, no direct insight into the elemental composition of the investigated sample is possible. Consequently, some of its applications are discussed here in this separate section. Because of the higher excitation probability UV radiation from discharge lamps (He I = 21.2 eV, He II = 40.8 eV) is often used. UPS has been used for the investigation of light elements, gases, small molecules or hydrides,[200] oxidation/chemisorption processes,[66,67,201] and conducting polymers,[148,202] often in conjunction with XPS results. Valence band investigations (also done by XPS) are useful for the identification of different structures in polymers where the C1s core-level peak shows no change.[151,203] These UPS results have been compared with DOS calculations for simple metallic systems.[204,205] Using approximation calculations for the molecular electronic structure commonly based on multiple scattered wave methods (usually described as Xα calculations)[206,207] valence band spectra estimations are possible for larger molecules. The combination of such measurements and DOS calculations are performed routinely to confirm theoretical considerations.[208,209] Figure 23 gives an example for different dimers of doped fullerenes.

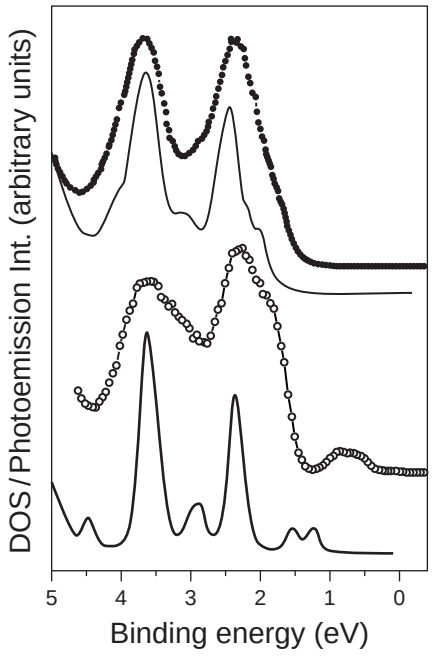

Figure 23 Valence band spectra for $(C_{59}N)_2$ (full circles) and $(Rb_1C_{60})_2$ (open circles) in comparison with density-of-states (DOS) calculations (solid lines). The normalized curves of the calculated DOS and the measured photoelectron intensities can then be used for qualitative comparison of the doping mechanism. (Reprinted with permission from: T. Pichler, M. Knupfer, M.S. Golden, S. Haffner, R. Friedlein, J. Fink, W. Andreoni, A. Curioni, M. Keshavarz-K, C. Bellavia-Lund, A. Sastre, J.-C. Hummelen, F. Wudl, 'On-ball Doping of Fullerenes: The Electronic Structure of $C_{59}N$ Dimers from Experiment and Theory', *Phys. Rev. Lett.*, **78**(22), 4249–4252 (1997), Copyright 1997, the American Physical Society.)

5.2 Loss Spectroscopy

Because the X-ray-induced photoelectrons or Auger electrons have to cross the surface region before they are emitted into the vacuum, nonelastic scattering may influence their energy. If due to the structural, electronic or chemical nature of the material one specific energy (region) of these losses is dominating, additional peaks may be identified and used for analysis. Processes for such additional (secondary) structures are very complex and are described in more detail elsewhere.[8,9,15] In this section, although not completely correct, all additional structures at the low-KE (high-BE) side of the photoelectron peak are discussed briefly as losses:

- *Plasmon losses* – mainly occurring on metals, these are collective interactions with the conduction band electrons, and are divided into bulk and surface plasmons. These losses may also be discussed in connection with conventional EELS investigations.
- *Shake-up satellite lines* – may be observed if a significant part of the emitting atoms is in the final state, not in the ionic ground state but in an excited state (valence electrons excited in higher unfilled levels).
- *Asymmetric core levels* – are found in metals due to the excitation of conducting band electrons into unfilled bands above the Fermi energy (similar to

the shake-up process). In non-metals vibrational broadening may be observed.

- *Multiplet splitting* – may occur if the atom has unpaired electrons in the valence level. Because of the residual spin the final ionic state may have a different energy and an additional splitting may be observed.

In connection with the main XPS features an analysis of such structures can lead to additional information. Cremer et al.[132] have discussed the possibilities of loss analysis for ceramic material. There is also a direct comparison of XPS and EELS loss analysis.

A classic example of loss spectroscopy is boron nitride,[210] because its cubic and hexagonal modifications may not be distinguished by XPS core-level analysis, but these are clearly identified by changes of the $\pi-\pi^*$ shake-up satellite. By way of illustration, Figure 24 shows spectra of hexagonal BN and cubic BN material.

5.3 Photoelectron Diffraction

As well as the nonelastic processes discussed in section 5.2, naturally elastic scattering of the emitted electrons is a principal process in the surface region. Because of the wave nature of the electrons in the quantum-mechanical description, diffraction and interference phenomena are to be expected. By measuring the peak intensity for the BE of a core-level photoelectron peak in the hemisphere above the sample, diffraction patterns can be observed if the surface region is well ordered (e.g. for single crystals or epitaxial layers). One of the first detailed studies was the work on copper by Fadley et al.[50] In practice the sample is usually rotated with respect to the analyzer (and the X-ray source) for such measurements.[50,211] Photoelectron diffraction (PED) is a complementary technique to extended X-ray absorption fine structure (EXAFS) investigations.[9,212]

Band dispersion properties can be investigated using UV excitation and photoelectron energies near the Fermi energy with angular dependent measurements. Using a small energy window, Fermi surface mapping can be done – as a spherical cut through the three-dimensional Fermi surface.[213] This work demonstrates the powerful combination of X-ray and UV excited techniques. As an example, Figure 25 shows diffraction patterns of Au to study the growth mode on Bi-2212 ($Bi_2Sr_2CaCu_2O_{8+x}$) superconductors where a 8.4 nm thick Au film was deposited onto Bi-2212. The measured Au4f diffraction pattern (Figure 25a) can be explained by a structural

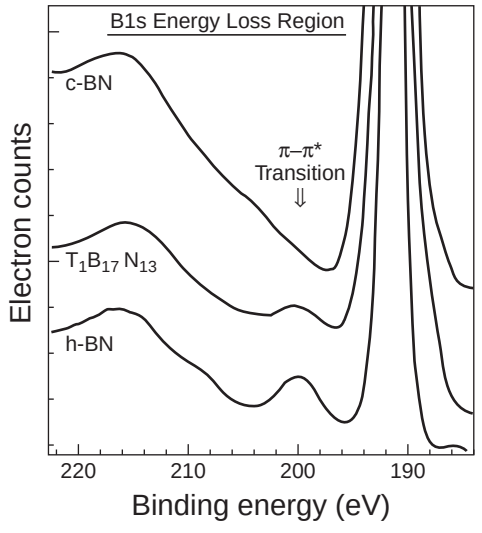

Figure 24 Difference between hexagonal (h-) BN and cubic (c-) BN by investigation of the $\pi-\pi^*$ transition 9 eV from the B1s peak. (Reprinted with permission from: M.A. Baker, T.P. Mollart, P.N. Gibson, W. Gissler, 'Combined X-ray Photoelectron/Auger Electron Spectroscopy/Glancing Angle X-ray Diffraction/Extended X-ray Absorption Fine Structure Investigations of TiB_xN_y Coatings', *J. Vac. Sci. Technol. A*, **15**(2), 284–291 (1997), Copyright 1997, American Institute of Physics.)

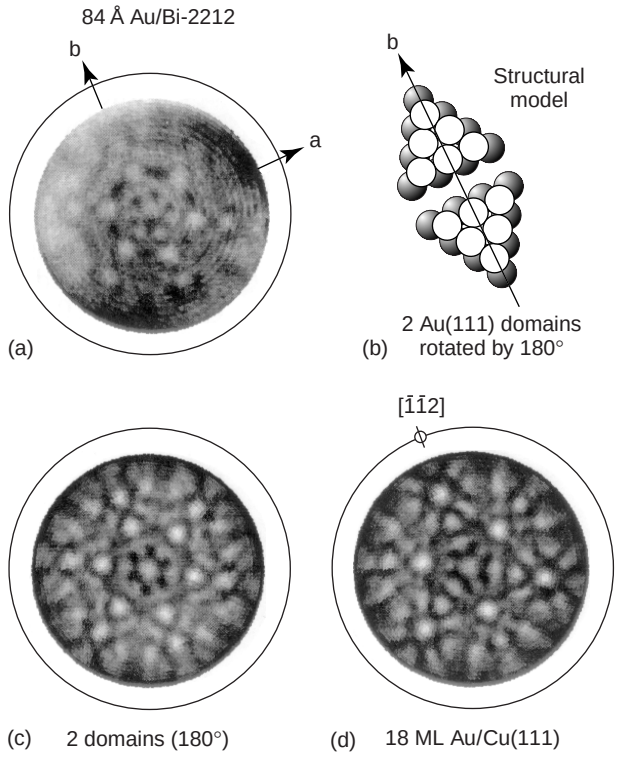

Figure 25 Examples for Au4f PED patterns in connection with the study of contact structures on Bi-2212 superconductors. (a) Measured Au4f diffraction pattern of a thin Au film on Bi-2212, (b) proposed structural model: 2 Au(111) domains rotated by 180°, (c) pattern very similar to (a) reconstructed from two Au(111) diffraction patterns rotated by 180° as measured at 18 monolayers Au on Cu(111) shown in part (d). (Taken from Aebi et al.[213])

model (Figure 25b) of two Au(111) domains rotated by 180°. This was concluded from Figure 25(c), which shows a very similar hexagonal pattern to that in (a). This pattern was reconstructed from an overlay of two Au(111) diffraction patterns as measured on Cu(111) (Figure 25d), but rotated by 180°.

Other work[214,215] deals with structural effects in GaAs-based epitaxy. As shown theoretically by Nefedov et al.[216] it should also be possible to learn about overlayer thicknesses by investigating the shape of the diffraction patterns.

5.4 In Situ Preparation, Special Equipment

Because of the problems that can arise concerning surface contamination, sample preparation is an important issue. Classical preparation techniques such as sequential sputtering and heating (for single-crystal preparation[13]), scraping, fracturing or cleaving (for bulk or grain-boundary investigations) can often be used directly in the analysis chamber under UHV conditions.[14] The trend towards more realistic experimental conditions for surface and interface modification has led to the use of special preparation chambers, to protect the analytical system from damage or contamination. Such chambers are separated from the analysis chamber by gate valves. This makes it possible to conduct procedures such as layer deposition with various deposition techniques, heating to higher temperatures, treatments with reactive gases and even electrochemistry without affecting the XPS measuring system. For contamination-free transfer of the sample after treatment, the preparation chamber is evacuated to UHV.[14]

Electrochemical experiments and other solid–liquid interaction studies have been tried, using specialized equipment[217–223] to deal with the non-UHV conditions. Another way to avoid contamination is the use of special transfer chambers.[224–226] Such equipment can be used for changing samples between different vacuum processes or to extract them from experiments under glove-box conditions.

Efforts to obtain X-ray sources with higher brightness have led to the use of synchrotron radiation. Using special monochromators, it is possible to use monoenergetic X-rays over a wide energy range,[16,227,228] in addition to the high-intensity bremsstrahlung radiation preferably used in XAES. Thus, the equipment for surface analysis has been installed at individual beam lines, or the experimental teams go with their own equipment to the beam line. Synchrotron experiments[229–231] are often focused on microscopic studies,[230] real-time measurements,[81,232] PED,[233,234] or resonant photoemission studies.[235]

6 COMPARISON WITH OTHER TECHNIQUES

Comparison with other techniques is a difficult task, because of the wide range of methods and their specifications. For the confident solution of most problems a skillful combination of several techniques is usually necessary.

As a basis for comparison with other analytical techniques, the main characteristics of standard XPS are as follows:

- *Features* – high surface sensitivity (2–5 nm), elemental information from core levels, detection limit about 0.1 atom %, chemical information from peak shifts, sputter depth profiling, direct measurement of band structure (valence band).
- *Advantages* – high surface sensitivity, chemical information easily to understand, nonconducting material possible, low sample damage.
- *Disadvantages* – low lateral resolution and thus low depth profiling sputter rates, problems with surface contamination, difficult preparation of bulk states, difficult BE scale calibration at insulators and semiconductors.

The techniques outlined below and other surface-related techniques are summarized regularly in several monographs.[2,11,16,236–238]

6.1 Auger Electron Spectroscopy

AES is the most complementary technique to XPS because of the similarity of the physical process employed. The Auger electrons are usually excited by a focused electron beam (3–10 keV), but X-ray excitation is also possible. In principle, AES has the same surface sensitivity and detection limits, although it is used more routinely for surface elemental analysis. High lateral resolution (50 nm) and faster depth profiling is possible. Chemical shifts are also present here, but they are more complicated to interpret. The AES technique is not suitable for insulating samples. Often multitechnique equipment AES/XPS is available on one machine using the same electron analyzer.

6.2 Secondary Ion Mass Spectrometry

This technique, which uses ion sputtering and detection of the emitted secondary ions from the surface, is a method for local trace analysis (ppb range). Its characteristics include lateral resolution down to 50 nm (with liquid metal ion sources), monolayer depth resolution (additionally broadened by mixing effects), and detection of all elements (including H), and isotope separation is possible.

Quantification is difficult because of strong matrix effects. With additional postionization, using secondary neutral mass spectrometry (SNMS), better quantification may be obtained. In static mode (low ion dose) and with TOF (time-of-flight) mass spectrometry (offering high mass resolution and mass range), molecular information may also be obtained; SIMS is often used in connection with XPS for polymer analysis.

6.3 Atomic Force Microscopy and Raster Tunneling Microscopy

These are methods for studying the local surface structure. Atomically sharp tips mounted on a very thin cantilever are used in close contact with the investigated surfaces. Mechanical (i.e. atomic forces) or electrical (i.e. tunnel current) interaction is used as the detected signal. Fine positioning of the cantilever is done by piezo ceramics and the resulting deflections are measured with laser optics. Structural information of the surfaces down to atomic resolution is routinely available. Using special measuring modes (modulation, special tips), additional local (mechanical, magnetic, chemical) information may be obtained. These methods are often used in connection with XPS to derive additional molecular and structural information. Atomic force measurements are also possible in the atmosphere or even in liquids.

6.4 Low-energy Electron Diffraction and Reflection High-energy Electron Diffraction

These are both electron diffraction techniques that can be used for surface structure investigations during sample cleaning (single crystals), thin-film preparation (epitaxy), and adsorption studies. They mainly differ in the impact energy and in the impact angle of the electrons. Because of the very low penetration depth of one or two atomic layers, low-energy electron diffraction (LEED) applications mainly concern adsorption or single-crystal studies. In reflection high-energy electron diffraction (RHEED) the penetration depth is larger and may be controlled additionally by the reflection geometry. This method is used also in combination with glancing angle X-ray diffraction.

6.5 Ion Surface Scattering

This is a rarely used method with extreme surface sensitivity (topmost layer). The elastic scattering of light ions (typically He^+ or Ne^+ in the energy range from 500 eV to a few keV) depends on the masses of surface elements. An energy spectrum of the scattered primary ions, depending on the angular situation, is recorded. The surface damage is low, but not always negligible. The main applications are surface segregation and adsorption studies. It is usually combined with XPS because of the use of the same analyzer (with changed polarity). Ion surface scattering (ISS) is sometimes called low-energy ion scattering (LEIS).

6.6 Infrared Spectroscopy

In surface analysis, infrared spectroscopy is useful for investigation of the bonding of molecules on the surface. Two methods should be mentioned here: Fourier-transform infrared spectroscopy (FTIR, based on the absorption of the radiation) and Raman spectroscopy (based on characteristic inelastic scattering of light). Lateral information may be obtained using focused laser radiation. Surface analysis is done preferably using reflection geometry. Vibrational infrared modes give information on the bonding types in UHV, under any gas pressure or in the liquid phase, which allows in situ experiments even in electrochemical cells.

6.7 Electron Energy Loss Spectroscopy

This has been discussed earlier in this article and is not a typical surface analytical technique. However, it is often used in connection with valence band XPS and UPS for basic studies of electronic band structure. Because high-energy electrons are used, EELS is commonly done in transmission on thin samples (typically 100 nm) in both standard transmission electron microscopes (mainly for local element or chemical analysis) and special UHV equipment (basic band structure studies).

6.8 Rutherford Backscattering Spectroscopy

This is a widely used nondestructive depth profiling method with very good quantification capabilities using the scattering of high-energy (as opposed to ISS), low-mass ions (typically He^+ in the 1 MeV range). The depth information has to be recalculated from the broadening of the elemental scattering peaks on the KE scale. The method has no lateral resolution (typically in the millimeter range, or using special beam collimators down to some micrometers) and is not suitable for light elements with high-Z samples.

6.9 Glow Discharge

Glow discharge optical emission spectroscopy (GDOES) and glow discharge mass spectrometry (GDMS) are developing from bulk-only to depth-profiling methods because of new source design and fast data collection. Relatively low ion energies from the glow discharge enable good depth resolution; however, the surface sensitivity is limited by high erosion rates and high sample temperatures. No lateral information is available.

7 OUTLOOK

Photoelectron spectroscopy is an analytical technique that, during its four decades of use, is assuming an increasingly important role among the surface-related techniques. Both experimental and theoretical aspects are now well understood. For routine analysis further development in terms of quantification, promoted by the revolution in computer techniques, is envisaged – more sophisticated escape depth models, calculation of chemical shifts and valence band features, use of large spectral data banks. Demands on future applications will be connected with new trends in biotechnology, environmental research, life sciences, biology, and nanotechnology. New instrumentation is likely to address wishes for lower spatial resolution, higher brightness, good energy resolution and variable X-ray energies, and will be closely connected with synchrotron radiation and X-ray optics.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
ARXPS	Angle-resolved X-ray Photoelectron Spectroscopy
BE	Binding Energy
DOS	Density of the Occupied Electronic States
EELS	Electron Energy Loss Spectroscopy
EPMA	Electron Probe Microanalysis
ESCA	Electron Spectroscopy for Chemical Analysis
EXAFS	Extended X-ray Absorption Fine Structure
FA	Factor Analysis
FTIR	Fourier-transform Infrared Spectroscopy
FWHM	Full Width at Half Maximum
GDMS	Glow Discharge Mass Spectrometry
GDOES	Glow Discharge Optical Emission Spectroscopy
IPES	Inverse Photoemission Spectroscopy
ISS	Ion Surface Scattering
ITO	Indium–Tin Oxide
KE	Kinetic Energy
LEED	Low-energy Electron Diffraction
LEIS	Low-energy Ion Scattering
LLS	Linear Least-squares
LUMO	Lowest Unoccupied Molecular Orbital
NLLS	Nonlinear Least-squares
PCA	Principal Component Analysis
PED	Photoelectron Diffraction
RHEED	Reflection High-energy Electron Diffraction
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectrometry
SNMS	Secondary Neutral Mass Spectrometry
TEM	Transmission Electron Microscopy
TOF	Time-of-flight
TOF/SIMS	Time-of-flight/Secondary Ion Mass Spectrometry
UHV	Ultrahigh Vacuum
UPS	Ultraviolet Photoelectron Spectroscopy
UV	Ultraviolet
XAES	X-ray Excited Auger Electron
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Pulp and Paper **(Volume 10)**
X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Infrared and Raman Spectroscopy in Analysis of Surfaces • Ion Scattering Spectroscopy in Analysis of Surfaces • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces • Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Atomic Spectroscopy **(Volume 11)**
Glow Discharge Optical Spectroscopy and Mass Spectrometry

Mass Spectrometry **(Volume 13)**
Secondary Ion Mass Spectrometry as Related to Surface Analysis • Time-of-flight Mass Spectrometry

Nuclear Methods **(Volume 14)**
Rutherford Backscattering Spectroscopy

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

REFERENCES

1. K. Siegbahn, C. Nordling, A. Fahlmann, R. Nordberg, K. Hamrin, J. Hedman, G. Johansson, T. Bergmark, S.-E. Karlsson, I. Lindgren, B. Lindberg, 'ESCA – Atomic, Molecular and Solid State Structure Studied by Means

of Electron Spectroscopy', *Nova Acta Regiae Soc. Sci. Upsaliensis, Ser. IV*, 20 (1967).
2. A.W. Czanderna (ed.), *Methods of Surface Analysis*, Elsevier, Amsterdam, 1975.
3. C.R. Brundle, A.D. Baker (eds.), *Electron Spectroscopy: Theory, Techniques and Applications*, Academic Press, London, Vol. 1, 1977.
4. C.R. Brundle, A.D. Baker (eds.), *Electron Spectroscopy: Theory, Techniques and Applications*, Academic Press, London, Vols. 2–5, 1978–1984.
5. D. Briggs (ed.), *Handbook of X-ray and Ultraviolet Photoelectron Spectroscopy*, Heyden, London, 1977.
6. M. Cardona, L. Ley (eds.), *Photoemission in Solids I: General Principles*, Springer-Verlag, Berlin, 1979.
7. L. Ley, M. Cardona (eds.), *Photoemission in Solids II: Case Studies*, Springer-Verlag, Berlin, 1979.
8. D. Briggs, M.P. Seah (eds.), *Practical Surface Analysis, Volume 1, Auger and X-ray Photoelectron Spectroscopy*, Wiley, Chichester, 1990.
9. S. Hüfner, *Photoelectron Spectroscopy, Principles and Applications*, Springer-Verlag, Heidelberg, 1995.
10. T.L. Barr, *Modern ESCA–The Principles and Practice of X-ray Photoelectron Spectroscopy*, CRC Press, Boca Raton, 1993.
11. G.C. Smith, *Quantitative Surface Analysis for Materials Science*, The Institute of Metals, London, 1991.
12. J.F. Watts, 'X-ray Photoelectron Spectroscopy', *Vacuum*, **45**(6/7), 653–671 (1994).
13. R.G. Musket, W. McLean, C.A. Colmenares, D.M. Makowiecki, W.J. Siekhaus, 'Preparation of Atomically Clean Surfaces of Selected Elements: A Review', *Appl. Surf. Sci.*, **10**, 143–207 (1982).
14. A.W. Czanderna, C.J. Powell, T.E. Madley (eds.), *Specimen Handling, Preparation, and Treatments in Surface Characterization*, Kluwer Academic, New York, 1998.
15. J.F. Moulder, W.F. Stickle, P.E. Sobol, K.D. Bomben, *Handbook of X-ray Photoelectron Spectroscopy*, Physical Electronics Inc., Eden Prairie, MN, 1995.
16. J.C. Riviere, *Surface Analytical Techniques*, Clarendon Press, Oxford, 1990.
17. S. Tougaard, 'Composition Depth Information from the Inelastic Background Signals in XPS', *Surf. Sci.*, **162**, 875–885 (1985).
18. D.A. Shirley, 'High-resolution Photoemission of the Valence Band of Gold', *Phys. Rev. B*, **5**(12), 4709–4714 (1972).
19. J.H. Scofield, 'Hartree-Slater Subshell Photoionization Cross-sections at 1254 and 1487 eV', *J. Electron Spectrosc. Rel. Phenom.*, **8**, 129–137 (1976).
20. C.J. Powell, M.P. Seah, 'Precision, Accuracy, and Uncertainty in Quantitative Surface Analyses by Auger-electron Spectroscopy and X-ray Photoelectron Spectroscopy', *J. Vac. Sci. Technol. A*, **8**(2), 735–763 (1990).
21. K. Siegbahn, C. Nordling, G. Johansson, J. Hedman, P.F. Heden, K. Hamrin, U. Gelius, T. Bergmark, L.O. Werme, R. Manne, Y. Baer, *ESCA Applied to Free Molecules*, North-Holland, Amsterdam, 1969.
22. M.E. Schwartz, 'Correlation of 1s Binding Energy with the Average Quantum Mechanical Potential at a Nucleus', *Chem. Phys. Lett.*, **6**(6), 631–636 (1970).
23. W.L. Jolly, D.N. Hendrickson, 'Thermodynamic Interpretation of Chemical Shifts in Core-electron Binding Energies', *J. Am. Chem. Soc.*, **92**(7), 1863–1870 (1970).
24. W.L. Jolly, 'A New Method for the Estimation of Dissociation Energies and its Application to the Correlation of Core-electron Binding Energies Obtained from X-ray Photoelectron Spectra', *J. Am. Chem. Soc.*, **92**(11), 3260–3265 (1970).
25. T. Kazahaya, M. Hirose, 'Coordinate Number of Doped Boron Atoms in Photochemically-deposited Amorphous Silicon Studied by X-ray Photoelectron Spectroscopy', *Jap. J. Appl. Phys.*, **25**(1), L75–L77 (1986).
26. S. Oswald, H. Wirth, 'Core-level Shifts at B- and Al-doped 6H–SiC Studied by XPS', *Surf. Interface Anal.*, **27**, 136–141 (1999).
27. C.D. Wagner, 'Chemical Shifts of Auger Lines, and the Auger Parameter', *Farad. Discuss. Chem. Soc.*, **60**, 291–300 (1975).
28. C.D. Wagner, 'The Auger Parameter, its Utility and Advantages: A Review', *J. Electron Spectrosc. Rel. Phenom.*, **47**, 283–313 (1988).
29. R.O. Ansell, T. Dickinson, A.F. Povey, P.M.A. Sherwood, 'X-ray Photoelectron Spectroscopic Studies of Electrode Surfaces Using a New Controlled Transfer Technique, Part II. Results for a Molybdenum Electrode and the Curve Fitting Procedure', *J. Electroanal. Chem.*, **98**, 79–89 (1979).
30. P. Gans, *Data Fitting in the Chemical Sciences*, Wiley, Chichester, 1990.
31. S.W. Gaarenstroom, 'Chemical Characterization from Carbon Auger Spectra by Application of Pattern Recognition and Factor Analysis', *J. Vac. Sci. Technol.*, **16**(2), 600–604 (1979).
32. M.F. Koenig, J.T. Grant, 'Comparison of Factor Analysis and Curve-fitting for Data Analysis in XPS', *J. Electron Spectrosc. Rel. Phenom.*, **41**, 145–156 (1986).
33. H.J. Steffen, S. Hofmann, 'Quantitative Studies of the Initial Oxidation Stages of FeCrNi Alloys Using Factor Analysis and Least-squares Fitting Methods Applied to Low-energy AES Spectra', *Surf. Interface Anal.*, **19**, 157–160 (1992).
34. M.S. Srivastava, E.M. Carter, *An Introduction to Applied Multivariate Statistics*, North-Holland, New York, 1983.
35. M. Otto, *Chemometrics–Statistics and Computer Application in Analytical Chemistry*, Wiley-VCH, Weinheim, 1999.
36. R. Reiche, S. Oswald, D. Hofmann, J. Thomas, K. Wetzig, 'Bombardment-induced Silicide Formation at Rhenium-silicon Interfaces Studied by XPS and TEM', *Fresenius' J. Anal. Chem.*, **365**(1–2), 76–82 (1999).

37. S. Oswald, W. Häßler, R. Reiche, J. Lindner, F. Weiss, 'XPS Depth Profile Analysis of a Thin Non-conducting Titanate Superlattice', *Mikrochim. Acta.*, **133**(1–4), 303–306 (2000).
38. S. Hofmann, 'Quantitative Depth Profiling in Surface Analysis: a Review', *Surf. Interface Anal.*, **2**, 148–160 (1980).
39. P. Laty, D. Seethanen, F. Degreve, 'Microroughness Induced on Solids by Ion Bombardment, I. Experimental Results on Sputtering of Aluminium by A^+ (15 keV); Quantification of Microroughness', *Surf. Sci.*, **85**, 353–364 (1979).
40. J.M. Morabito, R.K. Lewis, 'Secondary Ion Emission for Surface and In-depth Analysis of Tantalum Films', *Anal. Chem.*, **45**(6), 869–880 (1973).
41. H. Oechnser (ed.), *Thin Film and Depth Profile Analysis*, Springer-Verlag, Berlin, 1984.
42. K. Wetzig, S. Baunack, V. Hoffmann, S. Oswald, F. Präßler, 'Quantitative Depth Profiling of Thin Layers', *Fresenius' J. Anal. Chem.*, **358**, 25–31 (1997).
43. G. Betz, G.K. Wehner, 'Sputtering of Multicomponent Materials', in *Sputtering by Particle Bombardment II*, ed. R. Behrisch, Springer-Verlag, Berlin, 11–90, 1983.
44. N.Q. Lam, 'Ion Bombardment Effects on the Near-surface Composition During Sputtering', *Surf. Interface Anal.*, **12**, 65–77 (1988).
45. P. Sigmund, N.Q. Lam, 'Alloy and Isotope Sputtering', *Det Kongelige Danske Videnskabernes Selskab Matematik-fysike Meddelelser*, **43**, 255–349 (1993).
46. S. Hofmann, 'Sputter Depth Profile Analysis of Interfaces', *Rep. Prog. Phys.*, **61**, 827–888 (1998).
47. W. Eckstein, *Computer Simulation of Ion–Solid Interactions*, Springer-Verlag, Berlin, 1991.
48. M.P. Seah, W.A. Dench, 'Quantitative Electron Spectroscopy of Surfaces: A Standard Data Base for Electron Inelastic Mean Free Paths in Solids', *Surf. Interface Anal.*, **1**, 2–11 (1979).
49. K. Siegbahn, U. Gelius, H. Siegbahn, E. Olson, 'Angular Distribution of Electrons in ESCA Spectra from a Single Crystal', *Phys. Lett. A*, **32**(4), 221–222 (1970).
50. C.J. Fadley, R.J. Baird, W. Siekhaus, T. Novakov, S.A.L. Bergstrom, 'Surface Analysis and Angular Distribution in X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Rel. Phenom.*, **4**, 93–137 (1974).
51. H. Iwasaki, R. Nishitani, S. Nakamura, 'Determination of Depth Profiles by Angular Dependent X-ray Photoelectron Spectra', *Jap. J. Appl. Phys.*, **17**(9), 1519–1523 (1978).
52. I. Olefjord, H.J. Mathieu, P. Marcus, 'Intercomparison of Surface Analysis of Thin Aluminium Films', *Surf. Interface Anal.*, **15**, 681–692 (1990).
53. P.J. Cumpson, 'Angle-resolved XPS and AES: Development in Modelling for Reliable, Practical Quantification', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 627–634, 1996.
54. S. Tougaard, 'Quantitative Analysis of the Inelastic Background in Surface Electron Spectroscopy', *Surf. Interface Anal.*, **11**, 453–472 (1988).
55. S. Tougaard, *QUASES–Software Package for Quantitative XPS/AES, Users Guide*, QUASES Tougaard ApS., Odense, 1994.
56. S. Oswald, S. Baunack, 'XPS-investigations in Combination with Data Analysis for the Characterization of Non-conducting Material', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 844–847, 1997.
57. J.A.R. Samson, *Techniques of Vacuum Ultraviolet Spectroscopy*, Wiley, New York, 1967.
58. D. Briggs, 'Applications of XPS in Polymer Technology', in *Practical Surface Analysis, Volume 1, Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, 437–483, 1990.
59. R.E. Watson, J. Hudis, M.L. Perlman, 'Charge Flow and Compensation in Gold Alloys', *Phys. Rev. B*, **4**(12), 4139–4144 (1971).
60. M.F. Friedman, J. Hudis, M.L. Perlman, R.E. Watson, 'Electronic Behavior in Alloys: Au–Sn', *Phys. Rev. B*, **8**(6), 2433–2440 (1973).
61. J.C. Fuggle, E. Källne, L.M. Watson, D.J. Fabian, 'Electronic Structure of Aluminium-noble-metal Alloys Studied by Soft-X-ray and X-ray Photoelectron Spectroscopy', *Phys. Rev. B*, **16**(2), 750–761 (1977).
62. S. Evans, 'Work Function Measurements by X-PE Spectroscopy and their Relevance to the Calibration of X-PE Spectra', *Chem. Phys. Lett.*, **23**(1), 134–138 (1973).
63. B. Egert, G. Panzner, 'Electron Spectroscopic Study of Phosphorus Segregated to α-Iron Surfaces', *Surf. Sci.*, **118**, 345–368 (1982).
64. G. Panzner, W. Dieckmann, 'The Bonding State of Carbon Segregated to α-Iron Surfaces and on Iron Carbide Surfaces Studied by Electron Spectroscopy', *Surf. Sci.*, **160**, 253–270 (1985).
65. S. Hofmann, 'Surface Analysis Techniques in the Study of Segregation Phenomena', *Vacuum*, **40**(1/2), 9–17 (1990).
66. S. Evans, E.L. Evans, D.E. Parry, M.L. Tricker, M.J. Walters, J.M. Thomas, 'Ultra-violet and X-ray Photoelectron Spectroscopy Studies of Oxygen Chemisorption on Copper, Silver and Gold', *Faraday Discuss. Chem. Soc.*, **58**, 97–105 (1974).
67. S. Evans, J.M. Thomas, 'Electronic Structure of the Oxides of Lead', *J. Chem. Soc. Faraday Trans. 2*, **71**, 313–328 (1975).
68. J.E. Castle, D.C. Epler, 'The Correlation of XPS Analysis with Electrochemical History in Aqueous Corrosion', *Surf. Sci.*, **53**, 286–296 (1975).
69. J.E. Castle, 'The Use of X-ray Photoelectron Spectroscopy in Corrosion Science', *Surf. Sci.*, **86**, 583–602 (1977).

70. I. Olefjord, B. Brox, U. Jelvestam, 'Surface Composition of Stainless Steels during Anodic Dissolution and Passivation Studied by ESCA', *J. Electrochem. Soc.*, **132**(12), 2854–2861 (1985).
71. T.E. Madey, C.D. Wagner, A. Joshi, 'Surface Characterization of Catalysts using Electron Spectroscopies: Result of a Round-robin Sponsored by ASTM Committee D-32 on Catalysts', *J. Electron Spectrosc. Rel. Phenom.*, **10**, 359–388 (1977).
72. C.J. Powell, N.E. Erickson, T.E. Madey, 'Result of a Joint Auger/ESCA Round Robin Sponsored by ASTM Committee E42 on Surface Analysis, Part I. ESCA Results', *J. Electron Spectrosc. Rel. Phenom.*, **17**, 361–403 (1979).
73. L. Zhang, D.D. Macdonald, 'Segregation of Alloying Elements in Passive Systems – I. XPS Studies on the Ni–W System', *Electrochimica Acta*, **43**(18), 2261–2671 (1998).
74. R. Danoix-Souchet, A. D'Huysser, 'XPS Study of the Segregation of Minor Elements at the Surface of a Commercial Copper–10wt.% Nickel Alloy', *Surf. Interface Anal.*, **25**, 747–754 (1997).
75. D.-Q. Yi, U. Jevestam, I. Olefjord, 'ESCA and AES Analyses of the Fracture Surface of $MoSi_2$ Base Material', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 699–702, 1997.
76. T. Marinova, A. Kakanakova-Georgieva, C. Hallin, R. Yakimova, R. Kakanakov, L. Kassamakova, E. Janzen, 'Interface Chemistry of an Al/Ni/Al/SiC Ohmic Contact', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 659–662, 1997.
77. B. Lesiak, J. Zemek, P. de Haan, A. Jowzwik, 'XPS Study of Pd/Co Multilayers: Identification of Pd and Co Interface Bonding States by Pattern Recognition Methods', *Surf. Sci.*, **346**, 79–90 (1996).
78. W.J. Quadakkers, H. Viefhaus, 'The Application of Surface Analysis Techniques in High Temperature Corrosion Research', *High Temperature Corrosion Research and Testing*, **13**, 189–217 (1997).
79. N.S. McIntyre, T.C. Chan, 'Uses of Auger Electron and Photoelectron Spectroscopies in Corrosion Science', in *Practical Surface Analysis, Volume 1, Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, 485–529, 1990.
80. V. Maurice, W.P. Yang, P. Marcus, 'X-ray Photoelectron Spectroscopy and Scanning Tunneling Microscopy Study of Passive Films Formed on (100) Fe–^{18}Cr–^{13}Ni Single-crystal Surfaces', *J. Electrochem. Soc.*, **145**(3), 909–920 (1998).
81. A. Sandell, A.J. Jaworowski, A. Beutler, M. Wiklund, 'Adsorption and Reactions on a Surface Alloy: CO, NO, O_2 and CO_2 on Pd(100)-Mn-c(2×2)', *Surf. Sci.*, **421**, 116–134 (1999).
82. X. Bao, U. Wild, M. Muhler, B. Pettinger, R. Schlögl, G. Ertl, 'Coadsorption of Nitric Oxide and Oxygen on the Ag(110) Surface', *Surf. Sci.*, **425**, 224–232 (1999).
83. C.M. Whelan, C.J. Barnes, C.G.H. Walker, N.M.D. Brown, 'Benzenethiol Adsorption on Au(111) Studied by Synchrotron ARUPS, HREELS and XPS', *Surf. Sci.*, **425**, 195–211 (1999).
84. E.J. Sturrock, A.G. Fitzgerald, J. Thomson, A.C. Hourd, J.A. Cairns, P.R. Drapacz, 'An Investigation into the Palladium Metal Function of γ-Alumina Supported Palladium/Zinc Oxide Catalysts', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 73–76, 1996.
85. M. Valden, J. Aaltonen, N. Xiang, J. Pere, M. Pessa, 'Influence of Chemical State of Pd on the Interaction of Co with Supported Palladium Catalysts', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 81–84, 1996.
86. D.W. Harris, R.S. Nowicky, 'Application of AES in Microelectronics', in *Practical Surface Analysis, Volume 1, Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, 257–310, 1990.
87. C.C. Chang, P.H. Citrin, B. Schwartz, 'Chemical Preparation of GaAs Surfaces and their Characterization by Auger Electron and X-ray Photoemission Spectroscopy', *J. Vac. Sci. Technol.*, **14**(4), 943–952 (1977).
88. J.H. Thomas, III, S.P. Sharma, 'ESCA Study of the Passive Layer on Sn–Ni Alloy', *J. Vac. Sci. Technol.*, **14**(5), 1168–1172 (1977).
89. P. Alnot, A.M. Huber, J. Olivier, 'Application of SIMS, SAES and XPS to Problems in the Semiconductor Industry', *Surf. Interface Anal.*, **9**, 283–291 (1986).
90. P. DiBona, D.A. Wiegand, J. Sharma, 'XPS Studies of Surfaces and Thin Films of Small-band-gap Inorganic Azides', *J. Vac. Sci. Technol.*, **13**(1), 45–47 (1976).
91. R.W. Grant, J.G. Pasko, J.T. Longo, A.M. Andrews, 'ESCA Surface Studies of $Pb_{1-x}Sn_xTe$ Devices', *J. Vac. Sci. Technol.*, **13**(4), 940–947 (1976).
92. T.N. Wittberg, J.R. Hoenigman, W.E. Moddeman, C.R. Cothern, M.R. Gullet, 'AES and XPS of Silicon Nitride Films of Varying Refractive Indices', *J. Vac. Sci. Technol.*, **15**(2), 348–352 (1978).
93. R.M. Warner, B.L. Grung, *Semiconductor-device Electronics*, Holt, Rinehart and Winston, Philadelphia, 1991.
94. B.K. Tanner, *Introduction to the Physics of Electrons in Solids*, Cambridge University Press, Cambridge, 1995.
95. W.F. Egelhoff, Jr, 'Core-level Binding-energy Shifts at Surfaces and in Solids', *Surf. Sci. Rep.*, **6**, 253–415 (1987).
96. W.J. Choyke, H. Matsunami, G. Pensl (eds.), 'Fundamental Question and Applications of SiC (Part II)', *Phys. Status Solidi A*, **162**(1) (1997).
97. W.J. Choyke, H. Matsunami, G. Pensl (eds.), 'Fundamental Question and Applications of SiC (Part I)', *Phys. Status Solidi B*, **162**(1) (1997).

98. W.M. Lau, R.W.M. Kwok, S. Ingrey, 'Controlling Surface Band-bending of InP with Polysulfide Treatments', *Surf. Sci.*, **271**, 579–586 (1992).
99. H. Peisert, P. Streubel, T. Chasse, R. Szargan, 'A New Method of Sulfurization of InP(001) Surface with S_2Cl_2 studied by XPS and XAES', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 319–322, 1996.
100. S. Kaciulis, G. Mattogno, V. Olevano, P. Castrucci, M. De Crescenzi, I. Davoli, N. Pinto, 'Characterization of Si/Si_xGe_y Heterostructures and Si/Ge Superlattices', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 252–255, 1996.
101. A.C. Simonsen, M. Schleberger, S. Tougaard, J.L. Hansen, A.N. Larsen, 'Nanostructure of Ge Deposited on Si(001): A Study by XPS Peak Shape Analysis and AFM', *Thin Solid Films*, **338**, 165–171 (1999).
102. J. Olivier, S. Delande, R. Majorel, P. Collot, J. Nagle, 'Surface Sulfurization by Dipping $Al_xGa_{1-x}As$ in Ammonium Sulfide after Removal of Oxide and Element Arsenic by Wet Na_2S-based Cleaning: a Passivation Study by Monochromatized X-ray Photoelectron Spectroscopy', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 479–482, 1997.
103. J. Olivier, S. Delande, R. Majorel, P. Collot, J. Nagle, 'Study by Monochromatized X-ray Photoelectron Spectroscopy of the Improvement of Conventional Cleanings of $Al_xGa_{1-x}As$ (100) Surfaces: Deoxidation and Removal of Elementary Arsenic by Wet Na_2S-based Treatment', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 515–518, 1997.
104. R. Merica, J. Prokop, S. Veprek, 'In Situ XPS-studies of the Deposition of Amorphous Germanium by Organometallic CVD', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 315–318, 1996.
105. G.M. Ingo, G. Padeletti, S. Viticoli, C. Lanzieri, F. Fumi, 'Small-area XPS Depth Profiling and AFM Investigations on Plasma-enhanced CVD SiN_x Films Grown on GaAs(100) Surfaces', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 323–326, 1996.
106. A. Rinta-Möykky, P. Uusimaa, S. Suhonen, M. Valden, A. Salokatve, M. Pessa, 'Study of Ohmic Multilayer Metal Contacts to p-type ZnSe', *J. Vac. Sci. Technol. A*, **17**(2), 347–353 (1999).
107. L.-Y. Chen, G.W. Hunter, P.G. Neudeck, D. Knight, 'X-ray Photoelectron Spectroscopy Study of the Heating Effects on Pd/6H-SiC Schottky Structure', *J. Vac. Sci. Technol. A*, **16**(5), 2890–2895 (1998).
108. D.W. Moon, A. Kurokawa, S. Ichimura, H.W. Lee, I.C. Jeon, 'Ultraviolet-ozone Jet Cleaning Process of Organic Surface Contamination Layers', *J. Vac. Sci. Technol. A*, **17**(1), 150–154 (1999).
109. T. Ohno, A. Ohki, T. Matsuoka, 'Surface Cleaning with Hydrogen Plasma for Low Defect Density ZnSe Homoepitaxial Growth', *J. Vac. Sci. Technol. A*, **16**(4), 2539–2545 (1998).
110. R. Schlaf, R. Hinogami, M. Fujitani, S. Yae, Y. Nakato, 'Fermi Level Pinning of HF Etched Silicon Surfaces Investigated by Photoelectron Spectroscopy', *J. Vac. Sci. Technol. A*, **17**(1), 164–169 (1999).
111. W.H. Gries, 'The Angle-resolved Self-ratio Technique for Surface Depth Profile Investigations by XFS, EMA, XPS and AES', *Mikrochim. Acta*, **107**, 117–135 (1992).
112. I.S. Tilinin, J. Zemek, S. Hucek, 'Escape Probability of Signal Auger and Photoelectrons from Ultrathin Overlayers', *Surf. Interface Anal.*, **25**, 683–687 (1997).
113. P.C. Zalm, 'Overlayer Correction in XPS', *Surf. Interface Anal.*, **26**, 352–358 (1998).
114. R.C. Chatelier, H.A.W. St John, T.R. Gegenbach, P. Kingshott, H.J. Griesser, 'Incorporation of Surface Topography in the XPS Analysis of Curved or Rough Samples Covered by Thin Multilayers', *Surf. Interface Anal.*, **25**, 741–746 (1997).
115. T. Birchall, J.A. Connor, I, H. Hiller, 'High-energy Photoelectron Spectroscopy of Some Antimony Compounds', *J. Chem. Soc. Dalton Trans*, 2003–2006 (1975).
116. T.L. Barr, 'An XPS Study of Si as it Occurs in Adsorbents, Catalysts and Thin Films', *Appl. Surf. Sci.*, **15**, 1–35 (1983).
117. C.G. Pantano, 'Surface and In-depth-analysis of Glass and Ceramics', *Am. Ceram. Soc. Bull.*, **60**(11), 1154–1163, 1167 (1981).
118. P.E. Larson, M.A. Kelly, 'Surface Charge Neutralization of Insulating Samples in X-ray Photoemission Spectroscopy', *J. Vac. Sci. Technol. A*, **16**(6), 3483–3489 (1998).
119. G. Vereecke, P.G. Rouxhet, 'Surface Charging of Insulating Samples in X-ray Photoelectron Spectroscopy', *Surf. Interface Anal.*, **26**, 490–497 (1998).
120. B.J. Tielsch, J.E. Fulghum, 'Differential Charging in XPS. Part I: Demonstration of Lateral Charging in a Bulk Insulator Using Imaging XPS', *Surf. Interface Anal.*, **24**, 28–33 (1996).
121. B.J. Tielsch, J.E. Fulghum, 'Differential Charging in XPS. Part III: Comparison of Charging in Thin Polymer Overlayers on Conducting and Non-conducting Substrates', *Surf. Interface Anal.*, **25**, 904–912 (1997).
122. P. Swift, 'Adventitious Carbon–The Panacea for Energy Referencing?', *Surf. Interface Anal.*, **4**, 47–52 (1982).
123. T. Gross, M. Ramm, H. Sonntag, W. Unger, H.M. Weijers, E.H. Adem, 'An XPS Analysis of Different SiO_2 Modifications Employing a C1s as well as an $Au4f_{7/2}$

Static Charge Reference', *Surf. Interface Anal.*, **18**, 59–64 (1992).
124. S. Oswald, S. Baunack, 'Application of XPS and Factor Analysis for Non-conducting Materials', *Surf. Interface Anal.*, **25**, 942–947 (1997).
125. A. Davidson, J.F. Tempere, M. Che, H. Roulet, G. Dufour, 'Spectroscopic Studies of Nickel(II) and Nickel(III) Species Generated upon Thermal Treatments of Nickel/Ceria-supported Materials', *J. Phys. Chem.*, **100**, 4919–4929 (1996).
126. A.R. Gonzalez-Elipe, J.P. Holgado, R. Alvarez, G. Munuera, 'Use of Factor Analysis and XPS to Study Defective Nickel Oxide', *J. Phys. Chem.*, **96**, 3080–3086 (1992).
127. D.M. Hercules, A. Proctor, M. Houalla, 'Quantitative Analysis of Mixed Oxidation States in Supported Catalysts', *Acc. Chem. Res.*, **27**(12), 387–393 (1994).
128. D. Robba, C. Casale, M. Notaro, D.M. Ori, 'Reduction Behaviour of Vanadia in VO_x/TiO_2 Catalysts Poisoned with K_2O', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 169–172, 1997.
129. F. Brenscheidt, S. Oswald, A. Mücklich, E. Wieser, W. Möller, 'Wear Mechanism in Titanium Implanted Silicon Nitride Ceramics', *Nucl. Instrum. and Methods Phys. Res. Sect. B*, **129**, 483–486 (1997).
130. R. Alfonsetti, G. De Simone, L. Lozzi, M. Passacantando, P. Picozzi, S. Santucci, 'SiO_x Surface Stoichiometry by XPS: a Comparison of Various Methods', *Surf. Interface Anal.*, **22**, 89–92 (1994).
131. J.A. Mejias, V.M. Jimenez, G. Lassaletta, A. Fernandez, J.P. Espinos, A.R. Gonzalez-Elipe, 'Interpretation of the Binding Energy and Auger Parameter Shifts Found for TiO_2 Supported on Different Surfaces', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 272–275, 1996.
132. R. Cremer, M. Witthaut, D. Neuschütz, 'Electron Spectroscopy Applied to Metastable Ceramic Solution Phases', *Fresenius' J. Anal. Chem.*, **365**(1–2), 28–37 (1999).
133. K.H. Heinig, B. Schmidt, A. Markwitz, R. Grötzschel, M. Strobel, S. Oswald, 'Precipitation, Ripening and Chemical Effects During Annealing of Ge^+ Implanted SiO_2 Layers', *Nucl. Instrum. and Methods Phys. Res. Sect. B*, **148**, 969–974 (1999).
134. S. Oswald, B. Schmidt, K.-H. Heinig, 'XPS Investigation with Factor Analysis for the Study of Ge Clustering in SiO_2', *Surf. Interface Anal.*, **29**(4), 249–254 (2000).
135. N. Okude, M. Nagoshi, H. Kato, 'Chemical States of Surface Oxides on Electrodeposited Tin Studied by X-ray Photoelectron Spectroscopy', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 867–870, 1996.
136. W. Fliegel, G. Behr, J. Werner, G. Krabbes, 'Preparation, Development of Microstructure, Electrical and Gassensitive Properties of Pure and Doped SnO_2 Powders', *Sens. Actuators, B*, **18–19**, 474–477 (1994).
137. G. Gaggiotti, A. Galdikas, S. Kaciulis, G. Mattogno, A. Setkus, 'Surface Chemistry of Tin Oxide Based Gas Sensors', *J. Appl. Phys.*, **76**(8), 4467–4471 (1994).
138. R. Sanjines, D. Rosenfeld, F. Gozzo, Ph. Almeras, L. Perez, G. Magaritondo, W.H. Schreiner, 'ESCA Investigations of SnO_x Films Used as Gas Sensors', *Surf. Interface Anal.*, **22**, 372–375 (1994).
139. T. Hashemi, A.W. Brinkman, 'X-ray Photoelectron Spectroscopy of Nickel Manganese Oxide Thermistors', *J. Mater. Res.*, **7**(5), 1278–1282 (1992).
140. N. Morita, T. Endo, T. Sato, M. Shimada, 'The Oxidation State of Antimony and Electrical Properties in Antimony-doped Rutil', *J. Solid State Chem.*, **68**, 106–111 (1987).
141. J. Subrt, J. Halova, M. Skokanek, Z. Bastl, 'Characterization of Lead Oxides by X-ray Photoelectron Spectroscopy and Thermal Analysis Methods', *J. Mater. Sci.*, **30**, 437–442 (1995).
142. A.F. Orchard, G. Thornton, 'Evidence of 'Mixed-valency' Character in the X-ray Photoelectron Spectra of α-Diantimony Tetraoxide and Barium Bismuth Trioxide', *J. Chem. Soc., Dalton Trans*, 1238–1240 (1977).
143. B.D. Padalia, P.K. Mehta, *X-ray Photoelectron Spectroscopic Studies of High-T_c Oxide Superconductors, Studies of High Temperature Superconductors*, Nova Science, New York, 283–319, Vol. 1, 1989.
144. Y. Iijima, N. Niimura, K. Hiraoka, 'Prevention of the Reduction of CuO during X-ray Photoelectron Spectroscopy Analysis', *Surf. Interface Anal.*, **24**, 193–197 (1996).
145. D.T. Clark, A. Harrison, 'ESCA Applied to Polymers. XXXI. A Theoretical Investigation of Molecular Core Binding and Relaxation Energies in a Series of Prototype Systems for Nitrogen and Oxygen Functionalities in Polymers', *J. Polym. Sci., Polym. Chem. Ed.*, **19**, 1945–1955 (1981).
146. D.T. Clark, 'ESCA Applied to Organic and Polymeric Systems', in *Handbook of X-ray and Ultraviolet Photoelectron Spectroscopy*, ed. D.Briggs, 211–247, 1977.
147. A. Dilks, 'X-ray Photoelectron Spectroscopy for the Investigation of Polymeric Materials', in *Electron Spectroscopy–Theory, Techniques and Applications*, eds. C.R. Brundle, A.D. Baker, Academic Press, London, 278–359, Vol. 4, 1981.
148. W.D. Grobman, R.A. Pollak, D.E. Eastman, E.T. Maas, Jr, B.A. Scott, 'Valence Electronic Structure and Charge Transfer in Tetrathiofulvalinium Tetracyanoquinodimethane (TTF-TCNQ) from Photoemission Spectroscopy', *Phys. Rev. Lett.*, **32**(10), 534–537 (1974).
149. D.T. Clark, H.R. Thomas, 'Application of ESCA to Polymer Chemistry. X. Core and Valence Energy Level

of a Series of Polyacrylates', *J. Polym. Sci., Polym. Chem. Ed.*, **14**, 1671–1700 (1976).
150. D.T. Clark, H.R. Thomas, 'Application of ESCA to Polymer Chemistry. XI. Core and Valence Energy Level of a Series of Polymethacrylates', *J. Polym. Sci., Polym. Chem. Ed.*, **14**, 1701–1713 (1976).
151. J.J. Pireaux, J. Riga, R. Caudano, J. Verbist, 'Electronic Structure of Polymers, X-ray Photoelectron Valence Band Spectra', *ACS Symp. Ser.*, **162**, 169–201 (1981).
152. W.R. Salanek, 'Photoelectron Spectroscopy of the Valence Electronic Structure of Polymers', *CRC Crit. Rev. Solid State Mater. Sci.*, **12**(4), 267–296 (1985).
153. N. Larsson, P. Steinius, J.C. Eriksson, R. Maripuu, B. Lindberg, 'ESCA Studies of Sulfated Polyethylene Surfaces Modified by Adsorption of Polyethyleneimine and Colloidal Silica Particles', *J. Colloid. Interface Sci.*, **90**(1), 127–136 (1982).
154. B.D. Ratner, 'Surface Characterization of Biomaterials by Electron Spectroscopy for Chemical Analysis', *Ann. Biomed. Eng.*, **11**, 313–336 (1983).
155. E.D. Goddard, W.C. Harris, 'An ESCA Study of the Substantivity of Conditioning Polymers and Hair Substrates', *J. Soc. Cosmet. Chem.*, **38**, 233–246 (1987).
156. A. Toth, I. Bertoti, V.S. Khotimsky, 'Ion Bombardment Induced Modification of Polyvinyltrimethylsilane Studied by XPS', *Surf. Interface Anal.*, **22**, 551–555 (1994).
157. D. Briggs, S.F. Davies, P. Marsh, A. Broadhurst, 'Detection and Quantification of the Finishing Agent Permalose TM on the Surface of Polyester Fabric', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 509–512, 1996.
158. A.J. Ward, R.D. Short, 'A Spectroscopic Analysis of Plasma Polymers Prepared from a Series of Vinyl Sulphones', *Surf. Interface Anal.*, **22**, 477–482 (1994).
159. C. Jama, O. Dessaux, P. Goudmand, L. Gengembre, J. Grimblot, 'Characterization of Polyoxymethylene by XPS: Evidence of Surface Segregation and Composition Evolution During Analysis', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 771–774, 1997.
160. S.S. Perry, G.A. Somorjai, 'Characterization of Organic Surfaces', *Anal. Chem.*, **66**, 403A–415A (1994).
161. S. Mähl, M. Neumann, V. Schlett, A. Baalmann, 'Some Aspects of the Fitting of XPS Core Spectra of Polymers', *Surf. Interface Anal.*, **26**, 204–212 (1994).
162. S.L. Lim, K.L. Tan, E.T. Kang, 'Interaction of Evaporated Aluminium Atoms with Polyaniline Films: An X-ray Photoelectron Spectroscopic Study', *J. Vac. Sci. Technol. A*, **16**(1), 13–20 (1998).
163. M. Du, R.L. Opila, C. Case, 'Interface Formation Between Metals (Cu, Ti) and Low Dielectric Constant Organic Polymer (FLARE TM 1.0)', *J. Vac. Sci. Technol. A*, **16**(1), 155–162 (1998).
164. N. Tegen, S.A. Morton, J.F. Watts, G.K. Wolf, 'XPS Studies of Ion Implanted and As-deposited Metal/Polymer Interfaces', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 3–6, 1996.
165. M. Keil, C.S. Rastomijee, A. Rajagopal, C.L.A. Lamont, H. Sotobayashi, A.M. Bradshaw, D. Gador, E. Umbach, 'Interface Processes in the Adhesion of Aluminium with Untreated and Plasma-activated Polycarbonate Surfaces', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 739–742, 1997.
166. G. Polzonetti, G. Iucci, M.V. Russo, G. Infante, G. Paolucci, B. Brena, D. Cocco, 'X-ray Photoemission Study of the Interface Between Chromium and a Platinum-organometallic π-conjugated Polymer', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 125–128, 1997.
167. J. Riga, J. Delhalle, M. Deleuze, J.J. Pireaux, J.J. Verbist, 'An XPS Valence Band Study of Alkane Chains Secondary Structure', *Surf. Interface Anal.*, **22**, 507–510 (1994).
168. J.-M. Ouyang, W.-J. Zheng, N.-X. Huang, Z.-H. Tai, 'X-ray Photoelectron Spectroscopy Study of Langmuir-Blodgett Films of N-octadecyl-8-hydroxy-2-quinoline Carboxamide Deposited from Subphases Containing Metal Ions', *Thin Solid Films*, **340**, 257–261 (1999).
169. J.F. Watts, 'The Application of Surface Analysis to Studies of the Environmental Degradation of Polymer-to-metal Adhesion', *Surf. Interface Anal.*, **12**, 497–503 (1988).
170. M.V. Russo, G. Polzonetti, A. Furlami, A. Bearzotti, I. Fratoddi, P. Altamura, 'X-ray Photoelectron Spectroscopy and Scanning Electron-microscopy Characterization of Novel Poly(monosubstituted)acetylenes Containing Doping Species', *J. Vac. Sci. Technol.*, **16**(1), 35–44 (1998).
171. F. Garbassi, M. Morra, E. Occiello, L. Pozzi, 'The Use of XPS Valence Band Spectra in the Study of Plasma-deposited Polymeric Films of Tetramethyldisiloxane of Hydroxyethylmethacrylate', *Surf. Interface Anal.*, **22**, 485–490 (1994).
172. D. Schattka, T. Reier, J.W. Schultze, S. Winkels, 'ESCA-investigations for the Through Hole Plating of Printed Circuit Boards by Means of Conducting Polymers', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 787–790, 1997.
173. W.R. Salaneck, 'Electronic and Chemical Structure of Conjugated Polymer Surfaces and Interfaces', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 501–508, 1996.

174. N.M. Forsyth, P. Coxon, 'Use of Parallel Imaging XPS to Perform Rapid Analysis of Polymer Surfaces with Spatial Resolution <5 μm', *Surf. Interface Anal.*, **21**, 430–434 (1994).
175. B.D. Ratner, 'Surface Characterization of Polymeric Biomaterials: A Brief Tutorial', *Polym. Prepr.*, **34**(2), 50–51 (1993).
176. R.H. Bradley, I.J. Clackson, D.E. Sykes, 'XPS of Oxidized Wool Fiber Surfaces', *Surf. Interface Anal.*, **22**, 497–501 (1994).
177. P.J. de Lange, J.W.G. Mahy, 'ToF-SIMS and XPS Investigations of Fibers, Coatings and Biomedical Material', *Fresenius' J. Anal. Chem.*, **353**, 487–493 (1995).
178. S. Oswald, 'XPS Investigations on Flax Fibers', Internal Report IFS/29/1998, Institut für Festkörper- und Werkstofforschung (IFW), Dresden, 1998.
179. F.G. Zaugg, N.D. Spencer, P. Wagner, P. Kernen, A. Vinckier, P. Groscurth, G. Semenza, 'Microstructured Bioreactive Surfaces: Covalent Immobilization of Proteins on Au(111)/Silicon via Aminoreactive Alkanethiolate Self-assembled Monolayers', *J. Mater. Sci.: Mater. Med.*, **10**, 255–263 (1999).
180. P. Gröning, O. Gröning, M. Collaud-Coen, 'AFM and XPS Investigations on the Binding of Immunoglobulin G to Protein A Immobilized on TiO_2', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 131–138, 1997.
181. F. Faude, J. Goschnick, 'XPS, SIMS and SNMS Applied to a Combined Analysis of Aerosol Particles from a Region of Considerable Ar Pollution in the Upper Rhine Valley', *Fresenius' J. Anal. Chem.*, **358**, 67–72 (1997).
182. L.-D. Piveteau, M.I. Girona, L. Schlapbach, P. Barboux, J.-P. Boilot, B. Gasser, 'Thin Films of Calcium Phosphate and Titanium Dioxide by a Sol-gel Route: A New Method for Coating Medical Implants', *J. Mater. Sci.: Mater. Med.*, **10**, 161–167 (1999).
183. K. Schenk-Meuser, H. Duschner, 'ESCA-Analysis of Tin Compounds on the Surface of Hydroxyapatite', *Fresenius' J. Anal. Chem.*, **358**, 265–267 (1997).
184. F. Cirilli, S. Kaciulis, G. Mattogno, G. Righini, F. Ferrari, A. Montenero, G. Gnappi, 'Surface Analysis of Biocompatible Hydroxoapatite Coating on Titanium', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 151–154, 1997.
185. P. Merel, M. Tabbal, M. Chaker, S. Moisa, J. Margot, 'Direct Evaluation of the sp^3 Content in Diamond-like-carbon Films by XPS', *Appl. Surf. Sci.*, **136**, 105–110 (1998).
186. K.J. Boyd, D. Marton, S.S. Todorov, A.H. Al-Bayati, J. Kulik, R.A. Zuhr, 'Formation of C–N Thin Films by Ion Beam Deposition', *J. Vac. Sci Technol. A*, **13**(4), 2110–2122 (1995).
187. D.M. Poirier, T.R. Ohno, G.H. Kroll, P.J. Benning, F. Stepniak, J.H. Weaver, 'X-ray Photoemission Investigations of Binary and Ternary C_{60} Fullerides of Na, K, Rb and Cs', *Phys. Rev. B*, **47**(15), 9870–9877 (1993).
188. D.M. Poirier, C.G. Olson, 'Electronic States and Stability of the Insulating RbC_{60} Dimer Phase', *Phys. Rev. B*, **52**(16), R11 662–R11 664 (1995).
189. M.S. Golden, M. Knupfer, J. Fink, J.F. Armbruster, T.R. Cummins, H.A. Romberg, M. Roth, M. Sing, M. Schmidt, E. Sohmen, 'The Electronic Structure of Fullerenes and Fullerene Compounds from High-energy Spectroscopy', *J. Phys.: Condens. Matter*, **7**, 8219–8247 (1995).
190. T. Pichler, M. Knupfer, M.S. Golden, J. Fink, J. Winter, M. Haluska, H. Kuzmany, M. Keshavarz-K., C. Bellavia-Lund, A. Sastre, J.C. Hummelen, F. Wudl, 'The Electronic Structure of Polymerized Fullerenes and Dimerized Hetereofullerenes', *Appl. Phys. A*, **64**, 302–305 (1997).
191. S. Haffner, T. Pichler, M. Knupfer, B. Umlauf, R. Friedlein, M.S. Golden, J. Fink, M. Keshavarz-K, C. Bellavia-Lund, A. Sastre, J.C. Hummelen, F. Wudl, 'The Electronic Structure of $(C_{59}N)_2$ from High Energy Spectroscopy', *Eur. Phys. J. B*, **1**, 11–17 (1998).
192. D. Briggs, A. Brown, J. Vickerman, *Handbook of Static Secondary Ion Mass Spectrometry*, Wiley, Chichester, 1989.
193. *Surface Sci. Spectra*
194. G. Beamson, D. Briggs, *High Resolution XPS of Organic Polymers – The Scienta ESCA 300 Database*, Wiley, Chichester, 1992.
195. *NIST X-ray Photoelectron Spectroscopy Database, Version 2.0*, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.
196. *XPS Spectral Libraries #1 – Common Materials and #2 – Practical Studies*, XPS International, Kawasaki, Japan.
197. K. Yoshihara, M. Yoshitake, 'Sharing AES and XPS Data Through the Internet', *J. Vac. Sci. Technol. A*, **16**(3), 1388–1393 (1998).
198. W.A. Dench, L.B. Hazell, M.P. Seah, 'VAMAS Surface Chemical Analysis Standard Data Transfer Format with Skeleton Decoding Programs', *Surf. Interface Anal.*, **13**, 63–122 (1988).
199. *ISO 14976: 1998, Surface Chemical Analysis – Data Transfer Format*, International Organization for Standardization (ISO), Geneva, 1998.
200. W.C. Price, 'Ultraviolet Photoelectron Spectroscopy: Basic Concepts and the Spectra of Small Molecules', in *Electron Spectroscopy: Theory, Techniques and Applications*, eds. C.R. Brundle, A.D. Baker, Academic Press, London, 151–203, 1977.
201. S. Evans, 'Oxidation of the Group Ib Metals Studied by X-ray and Ultraviolet Photoelectron Spectroscopy', *J. Chem. Soc., Faraday Trans. 2*, **71**, 1044–1057 (1975).

202. P. Nielsen, A.J. Epstein, D.J. Sandman, 'Photoemission Studies of TCNQ Compounds', *Solid State Comm.*, **15**, 53–58 (1974).
203. E. Heilbronner, J.P. Maier, 'Some Aspects of Organic Photoelectron Spectroscopy', in *Electron Spectroscopy: Theory, Techniques and Applications*, eds. C.R. Brundle, A.D. Baker, Academic Press, London, 205–292, Vol. 3, 1977.
204. Y. Baer, G. Busch, 'X-ray Photoemission from Aluminium', *Phys. Rev. Lett.*, **30**(7), 280–282 (1973).
205. G.A. Rooke, 'Interpretation of Aluminium X-ray Band Spectra, I. Intensity Distribution', *J. Phys. C*, **1**, 767–775 (1968).
206. S. Thomas, P.M.A. Sherwood, 'Valence Band Spectra of Aluminium Oxides, Hydroxides, and Oxyhydroxides Interpreted by Xα Calculations', *Anal. Chem.*, **64**, 2488–2495 (1992).
207. P.M.A. Sherwood, 'Surface Analysis for the Investigation of Electrochemical and Corrosion Systems', *Anal. Chim. Acta*, **283**, 52–61 (1993).
208. M. Kudo, T. Hikita, T. Hanada, R. Sekine, M. Kawai, 'Surface Reactions at the Controlled Structure of $SrTiO_3$(001)', *Surf. Interface Anal.*, **22**, 412–416 (1994).
209. T. Pichler, M. Knupfer, M.S. Golden, S. Haffner, R. Friedlein, J. Fink, W. Andreoni, A. Curioni, M. Keshavarz-K, C. Bellavia-Lund, A. Sastre, J.-C. Hummelen, F. Wudl, 'On-ball Doping of Fullerenes: The Electronic Structure of $C_{59}N$ Dimers from Experiment and Theory', *Phys. Rev. Lett.*, **78**(22), 4249–4252 (1997).
210. M.A. Baker, T.P. Mollart, P.N. Gibson, W. Gissler, 'Combined X-ray Photoelectron/Auger Electron Spectroscopy/Glancing Angle X-ray Diffraction/Extended X-ray Absorption Fine Structure Investigations of TiB_xN_y Coatings', *J. Vac. Sci. Technol. A*, **15**(2), 284–291 (1997).
211. D. Naumovic, A. Stuck, T. Greber, J. Osterwalder, L. Schlapbach, 'Full-hemispherical Photoelectron-diffraction Data from Cu(001): Energy Dependence and Comparison with Single-scattering-cluster Simulation', *Phys. Rev. B*, **47**(12), 7462–7479 (1993).
212. D.C. Konigsberger, R. Prins (eds.), *Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*, Wiley, New York, 1988.
213. P. Aebi, P. Schwaller, J. Osterwalder, R. Fasel, H. Berger, C. Beeli, L. Schlapbach, 'Angle-scanned Photoemission on Bi-type High T_c Superconductors: Fermi Surfaces and Thin Metal Films', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 733–740, 1996.
214. M.G. Proiretti, S. Turchini, J. Garcia, M.C. Arsenio, C. Casado, F. Martelli, T. Prosperi, 'Soft X-ray Photoelectron Study of Epitaxial InGaAs/GaAs(001)', *J. Vac. Sci. Technol. A*, **16**(4), 2318–2325 (1998).
215. M. Seelmann-Eggebert, H. Zimmermann, H. Obloh, R. Niebuhr, B. Wachtendorf, 'Plasma Cleaning and Nitridation of Sapphire Studied by X-ray Photoelectron Diffraction', *J. Vac. Sci. Technol. A*, **16**(4), 2008–2015 (1998).
216. V.I. Nefedov, I.S. Fedorova, 'The Influence of the Photoelectron Elastic Scattering on the Photoelectron Diffraction Patterns, In-depth Profiling and Thin Film Thickness Determination', *J. Electron Spectrosc. Rel. Phenomena*, **88–91**, 509–516 (1998).
217. B. Baron, P. Chartier, P. Delahay, R. Lugo, 'Photoelectron Emission by Solutions', *J. Chem. Phys.*, **51**(6), 2562–2572 (1969).
218. H. Aulich, P. Delahay, L. Nemec, 'Energy Distribution and Cross Section for Photoionization in Polar Liquids', *J. Chem. Phys.*, **59**(5), 2354–2364 (1973).
219. H. Siegbahn, K. Siegbahn, 'ESCA Applied to Liquids', *J. Electron Spectrosc. Rel. Phenom.*, **2**, 319–325 (1973).
220. P. Delahay, 'Photoelectron Emission Spectroscopy of Liquids and Solutions', in *Electron Spectroscopy: Theory, Techniques and Application*, eds. C.R. Brundle, A.D. Baker, Academic Press, New York, 124–196, Vol. 5, 1984.
221. D. Schaepers, H.-H. Strehblow, U. Künzelmann, 'Electrochemical, Ellipsometrical and Surface Analytical Investigations of Passive Layers on Binary Fe/Al Alloys', *Surf. Interface Anal.*, **21**, 342–348 (1994).
222. D. Schaepers, H.-H. Strehblow, 'Surface Analytical Investigations of Electrochemical Formed Passive Layers on Binary Fe/Al Alloys', *J. Electrochem. Soc.*, **142**(7), 2210–2218 (1995).
223. S. Södergren, H. Rensmo, H. Siegbahn, 'Electron Spectroscopy Applied to Electrochemical Interfaces: Intercalation Processes in Metal Oxide Surfaces', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 3–12, 1997.
224. J.P. Hobson, 'Abstract: First Intercontinental Test of Vacuum Transfer Device', *J. Vac. Sci. Technol.*, **15**(2), 659 (1978).
225. T. Fleisch, A.J. Shepard, T.Y. Ridley, W.E. Vaughn, N. Winograd, W.E. Baitinger, G.L. Ott, W.N. Delgass, 'A System for Transferring Samples between Chambers in UHV', *J. Vac. Sci. Technol.*, **15**(5), 1756–1760 (1978).
226. J.-M. Abels, D. Hecht, H.-H. Strehblow, 'A Closed Specimen Transfer for Surface Analytical Application', *Surf. Interface Anal.*, **24**, 332–338 (1996).
227. I. Lindau, P. Pianetta, S. Doniach, W.E. Spicer, 'X-ray Photoemission Spectroscopy', *Nature*, **250**, 214–215 (1974).
228. P. Pianetta, I. Lindau, C. Garner, W.E. Spicer, 'Determination of the Oxygen Binding Site on GaAs(110) Using Soft-X-ray-Photoemission Spectroscopy', *Phys. Rev. Lett.*, **35**(20), 1356–1359 (1975).
229. C.S. Fadley, 'The Study of Surface Structures by Photoelectron Diffraction and Auger Electron Diffraction', in *Synchrotron Radiation Research: Advances in Surface and Interface Science Vol. 1: Techniques*, ed. R.Z. Bachrach, Plenum Press, New York, 421–518, 1992.

230. G. Margaritondo, 'The New Perspectives in Surface and Interface Analysis', *Surf. Interface Anal.*, **22**, 1–8 (1994).
231. R.P. Walker, 'Seeing Things in a New Light: Synchrotron Radiation and Free Electron Laser Sources', in *ECASIA 95, 6th European Conference on Applications of Surface and Interface Analysis*, eds. H.J. Mathieu, B. Reihl, D. Briggs, Wiley, Chichester, 705–712, 1996.
232. N. Zacchetti, O. Tassa, S. Bellini, G. Paolucci, G. Comelli, A. Baraldi, S. Lizzit, G. Polzonetti, 'Oxidation Study of FeAl, NiAl and Ni_3Al Intermetallic Compounds by Real-time Synchrotron Radiation Photoemission', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 883–886, 1997.
233. L.J. Terminello, X.S. Zhang, Z.Q. Huang, S. Kim, A.E. Schach von Wittenau, K.T. Leung, D.A. Shirley, 'c(2 × 2)S/Cr(001) Surface and Near-surface Structure Determined Using Angle-resolved Photoemission Extended Fine Structure', *Phys. Rev. B*, **38**(6), 3879–3891 (1988).
234. C.S. Fadley, S. Thevuthasan, A.P. Kaduwela, C. Westphal, Y.J. Kim, R. Ynzunza, P. Len, E. Tober, F. Zhang, Z. Wang, S. Ruebush, A. Budge, M.A. Van Hove, 'Photoelectron Diffraction and Holography: Present Status and Future Prospects', *J. Electron Spectrosc. Rel. Phenom.*, **68**, 19–47 (1994).
235. M.D. Crapper, D. Brown, M. Petty, K.H. Bedwell, S.J. Guilfoyle, A.E.R. Malins, M.T. Butterfield, 'Investigation of the Electronic Structure of Intermetallic Compounds Utilising Synchrotron Radiation', in *ECASIA 97, 7th European Conference on Applications of Surface and Interface Analysis*, eds. I. Olefjord, L. Nyborg, D. Briggs, Wiley, Chichester, 891–894, 1997.
236. M. Grasserbauer, H.J. Dudek, M.F. Ebel, *Angewandte Oberflächenanalyse mit SIMS, AES und XPS (Applied Surface Analysis with SIMS, AES and XPS)*, Springer-Verlag, Berlin, 1986.
237. D. Brune, R. Hellborg, H.J. Whitlow, O. Hunderi, *Surface Characterization – A User's Sourcebook*, Wiley, Weinheim, 1997.
238. J.C. Rivere, S. Myhra (eds.), *Handbook of Surface and Interface Analysis – Methods for Problem Solving*, Marcel Dekker, New York, 1998.

Atomic Spectroscopy: Introduction

Gary M. Hieftje
Indiana University, Bloomington, USA

1 Systems for Atom Formation in Atomic Spectroscopy 1
2 Detection Methods in Atomic Spectroscopy 3
Abbreviations and Acronyms 5
References 5

Atomic spectroscopy (also termed atomic spectrometry) is one of the oldest and most well established of the analytical methods. Its roots can be traced to very early observations in which the presence of specific salts in a chemical sample imparted characteristic colors to a luminous flame. Atomic spectrometry is perhaps the most prominent and widely used of the family of methods employed for elemental analysis.

All methods for elemental analysis, including atomic spectrometry, exploit quantized transitions characteristic of each individual element. For example, X-ray fluorescence spectrometry utilizes inner-shell electron transitions. Because inner-shell electrons are not involved in bonding, the transitions they undergo are perturbed very little by the presence of nearby atoms, by chemical bonds, or by the external environment of an atom. As a consequence, X-ray fluorescence spectra consist of narrow spectral lines that are indicative of each element. Moreover, the sample can be in virtually any state – solid, liquid, or gas. However, X-ray fluorescence spectrometry has its shortcomings. Detection limits are seldom below the level of one part per million in either a solid or liquid, it is difficult to measure elements of low atomic number, and X-radiation emitted by one element can be absorbed by certain others present in the same sample. A consequence of this last factor is rather severe matrix (interelement) interferences. That is, the signal that is observed from one element depends not only on its concentration, but also on the concentration of other elements in the sample.

Neutron activation analysis (NAA) is another approach for elemental analysis that relies upon discrete transitions typical of each element. In this case, however, the transitions are in the atomic nucleus. Thus, they too are unaffected by chemical bonding or by the presence of other nearby atoms. Unfortunately, most samples do not lend themselves immediately to analysis by means of nuclear transitions, since they do not contain appreciable amounts of naturally radioactive nuclei. As a consequence, the samples must be ''activated'' by placing them in an intense flux of slow neutrons. The resulting activated nuclei can then be employed for chemical analysis. However, the requirement of a slow neutron source limits the widespread applicability of the method.

Unlike these alternative methods, atomic spectrometry utilizes quantized transitions in the conveniently accessible ultraviolet, visible, and near-infrared regions of the electromagnetic spectrum. As a result, the spectroscopic instrumentation is relatively inexpensive, readily available, and easy to use. Of course, it is valence electronic transitions that occur in these spectral regions. For narrow-band, characteristic spectrum to be generated, each atom must be isolated from all others, so the transitions are not perturbed by neighboring atoms or by bonding effects. If this requirement is not met, the resulting spectra are representative more of molecules or molecular fragments than of atoms themselves.

Accordingly, the underlying requirement for all atomic methods of analysis is that a sample be decomposed to the greatest extent possible into its constituent atoms. Ideally, this atomization step should be quantitative; there should be no residual bonding in the gas-phase atomic cloud. Anything less than complete atomization will understandably yield lower sensitivities in any atomic method. Even more importantly, changes in the fraction of atomization from sample to sample or from sample to standard will cause errors in calibration. Thus, it is less important that complete atomization be achieved than that the fractional atomization be extremely consistent. To the extent that this condition is not met, interelement (matrix) interferences in atomic spectrometry can be extremely troublesome.

It is not surprising, then, that a consistent theme in the history of atomic spectrometric analysis is a search for improved methods of atomization for samples in solid, solution, or gaseous form. It is therefore appropriate that we consider such systems in some detail here.

1 SYSTEMS FOR ATOM FORMATION IN ATOMIC SPECTROSCOPY

Schemes for the atomization of various samples have followed several traditional paths. The most common first step is to dissolve the sample if it is not already in solution form. Although this step is an inconvenient and often time-consuming one, it also offers a number of important benefits. First, after a sample is dissolved, the principal constituent in the sample solution is the solvent. Consequently, most sample solutions look more or

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

less the same and similarly resemble standard solutions that are prepared. Secondly, samples in solution form are relatively easy to handle and lend themselves readily to automation. Third, sample solutions permit relatively simple and straightforward background correction, simply by use of a solvent or reagent blank. Lastly, other constituents can readily be added to sample and standard solutions to simplify such procedures as standard additions (spiking) and internal standardization.

Among the methods widely employed for automation of the solution-introduction step in atomic spectrometry is flow injection analysis (FIA). Unlike many other automation schemes, FIA can be employed not only to introduce solution samples directly, but to dilute them in an automatic and predictable fashion, to mix the sample solution with other reagents, to exploit standard-additions and internal-standardization approaches, and to speed up the sample introduction process. Through use of FIA, it becomes possible routinely to employ sample solution volumes as low as 10 μl, to reduce build-up on sprayers and other equipment in an atomic spectrometric system, and to improve precision. Some of the benefits and capabilities of FIA are described in **Flow Injection Analysis Techniques in Atomic Spectroscopy**.

Once a sample solution is prepared, it can be converted to free atoms in several alternative ways. One is to pipette a small aliquot of the sample solution into an electrothermal atomizer (ETA) such as a graphite furnace. This method is discussed in greater detail in **Graphite Furnace Atomic Absorption Spectrometry**. In the graphite furnace, the solvent is first evaporated from the sample at a moderate temperature. The furnace temperature is then raised so organic material is ashed, and the temperature next increased rapidly to the point where the sample is vaporized and ultimately atomized. Not surprisingly, the furnace temperature and the sample composition are extremely important in ensuring the efficiency and consistency of this atomization.

A second path to the atomization of solution samples is through a nebulization (spraying) process. In simple terms, nebulization serves to increase the surface area of the solution sample, so solvent evaporation (desolvation) can proceed more rapidly and so the resulted dried solute particles can be volatilized better. This scheme, which is probably the most common in analytical atomic spectrometry, is employed in flame atomic absorption, flame emission, and plasma emission spectrometry and is detailed in **Flame and Vapor Generation Atomic Absorption Spectrometry** and **Inductively Coupled Plasma/Optical Emission Spectrometry**. Once formed, droplets in the nebulized spray are sent into a high-temperature environment such as a chemical flame or flowing rare-gas plasma. There, desolvation and solute-particle vaporization take place, and the resulting vapor converted more or less efficiently into free atoms. Indeed, the environment in these discharges is often hot enough that many of the atoms that are formed wind up as positively charged ions. Also, the environment in these atomization sources is often energetic enough to yield strong emission from either the freed atoms or their ionic counterparts. Correspondingly, this general scheme lends itself well to a number of different detection approaches that will be discussed shortly: atomic emission spectrometry (AES), atomic absorption spectrometry (AAS), atomic fluorescence spectrometry (AFS), and atomic mass spectrometry (AMS).

For some elements, it is possible to employ a more straightforward means of generating free atoms. In the case of mercury, for example, free atoms can be formed simply by the chemical reduction of inorganic mercury in solution to the free atomic form. The neutral mercury atoms can then be driven from solution merely by passing an appropriate carrier gas (for example, air or argon) through it. The liberated atoms, present in a relatively cool environment, can then be measured alternatively by such techniques as atomic absorption or AFS.

For other elements, a chemical reaction will yield not free atoms but rather other volatile species that can be dissociated into free atoms at moderate temperatures. The most common example here is the use of a chemical reduction to form a stable hydride of such elements as tin, antimony, and arsenic. As in the case of mercury atoms, these volatile hydrides can be driven from solution by bubbles of an appropriate gas, which will carry them into a moderate-temperature flame or furnace for atomization. Once formed, those atoms can then be measured by atomic absorption, atomic fluorescence, or certain other spectrometric methods. These schemes are discussed in greater detail in **Flame and Vapor Generation Atomic Absorption Spectrometry**.

Naturally, it would be desirable in many cases to be able to analyze solid samples directly. Even more beneficial in special cases would be the possibility of measuring sample concentrations in a solid on a three-dimensional spatial basis. Some methods employed for sample volatilization and atomization have been aimed at exactly that goal.

Of such methods, the best established employs a glow discharge, detailed in **Glow Discharge Optical Spectroscopy and Mass Spectrometry**. In a glow discharge, ordinarily operated at pressures in the range of 1 Torr, the sample surface is bombarded by energetic rare-gas ions, usually of argon. Because of this steady bombardment, the surface is eroded on a layer-by-layer basis. Thus, atoms freed from the surface as a function of time are taken from successively deeper layers within the sample. Recording the spectrometric signal as a function of time then permits a ''depth profile'' of the sample to be obtained. In early work, glow-discharge devices were

applicable only to conductive solid samples because of the need to impart a negative charge to the sample surface, in order to attract argon ions to it. More recent developments, however, have shown it possible to utilize radiofrequency-sustained discharges for the same purpose; such discharges permit the depth-resolved analysis of nonconductive samples as well.

In recent years, intense laser beams also have been exploited for depth resolution. Unlike the glow discharge, however, a laser beam can be focused to discrete, extremely small spots on the sample surface, so that not only depth resolved but also laterally resolved information can be obtained. In its most straightforward version, the laser is employed not as an atomization source but rather as a sampling device, as is described in greater detail in **Laser Ablation in Atomic Spectroscopy**. In recent years, the importance of the laser's wavelength and power density on the ablation process has been characterized more fully and has led to a burgeoning array of applications. Significantly, sample material ablated by a laser can be fed into any of several sources for further atomization. The most attractive combinations have been between laser ablation and either inductively coupled plasma (ICP) AES or ICP mass spectrometry.

A more direct approach to solid-sample analysis, described in **Laser-induced Breakdown Spectroscopy**, is to exploit the laser beam not only for sample vaporization but also to generate a plasma in which the sample vapor can be atomized, excited, and ionized. This technique, termed laser-induced breakdown spectroscopy (LIBS), is applicable to samples of many types and can also be applied to surfaces that are relatively remote from both the laser and the spectrometric measurement equipment.

Some sources employed in atomic spectrometry are not as good at generating atoms as they are at exciting or ionizing them. The microwave-induced plasma (MIP), discussed in **Microwave-induced Plasma Systems in Atomic Spectroscopy**, typifies such sources. Interestingly, the MIP (like some other spectroscopic sources) is well removed from local thermodynamic equilibrium. That is, the source appears to have several temperatures that are linked not by equilibrium processes but rather by kinetically controlled events. Thus, the temperature that would define the energy distribution of electrons moving in a MIP is generally much higher than the temperature that would pertain to the motion of heavier particles such as rare-gas atoms or ions. An important consequence of this dis-equilibrium is that features driven by the electrons (most importantly from our standpoint, ionization and excitation of atoms) are extremely efficient whereas other events that are controlled by gas-kinetic (thermal) operations are less efficient. Because solvent evaporation and solute-particle vaporization occur mostly through these latter thermal events, they are not efficient in a MIP.

The result of this situation is that the MIP is exceptionally useful for exciting and ionizing atoms, but that it cannot as readily effect vaporization. In fact, many MIP systems tolerate only a small loading of solvent vapor before they are visibly perturbed. The MIP therefore finds an important role in atomic spectrometry in the dissociation of gas-phase samples and in the production of atomic emission and mass spectra. It is therefore particularly attractive as a source for detection of gas chromatographic effluents, of gas-phase species generated by a chemical-based hydride-formation apparatus, or for atoms volatilized from an auxiliary source such as a carbon furnace.

2 DETECTION METHODS IN ATOMIC SPECTROSCOPY

Once atoms are in the gas phase, they can be probed by any of several spectrometric techniques, including AES, AAS, AFS, AMS, coherent forward scattering spectrometry, photothermal deflection spectrometry, atomic magneto-optic rotation spectrometry, and several others. Of these alternatives, the most common have become atomic absorption, atomic emission, and AMS.

AAS dates from the earliest observations by Fraunhofer of dark lines in the sun's spectrum and, as an analytical method, from the pioneering work by Walsh and by Alkemade and Milatz in 1955.[1,2] It has become a workhorse in atomic spectrometric analysis, in large part because of the simplicity of the instrumentation that it requires. To perform atomic absorption requires only a primary light source, usually a hollow cathode lamp, an atomization cell such as any of those described in the preceding paragraphs, a spectral isolation device such as a monochromator, an appropriate photodetector, and associated electronics. Working curves are usually constructed in the form of a Beer's-law plot familiar from molecular spectrophotometry. However, for Beer's law to be followed closely (that is, for linear calibration curves to be obtained), the band of detected radiation must be narrower than the atomic absorption line that is being measured. Because atomic absorption lines are extraordinarily narrow, the primary light source (hollow cathode lamp) must emit a band of light that is at least as narrow. It is for this reason that the hollow cathode lamp was chosen. Significantly, this choice also simplifies the measurement, since spectral lines emitted by the hollow cathode lamp are typical of the element being measured and are naturally locked onto the narrow atomic absorption lines of interest. Thus, spectral mismatch is almost impossible and spectral interferences are relatively uncommon because of the narrowness of the spectral lines that are being probed.

Unfortunately, it is not particularly convenient to measure more than one element at a time by means of AAS. There is a necessary straight-line geometry among the primary light source (hollow cathode lamp), the atom cell (flame or furnace) and the detection equipment. This alignment makes it difficult to incorporate more than a single source into the system; because each hollow cathode lamp emits efficiently the spectrum of only one, two, or three elements at a time, measuring additional elements requires substituting a new hollow cathode lamp. Although recent advances have been made in continuum-source atomic absorption, even those arrangements are somewhat limited: continuum sources that extend well into the important ultraviolet region of the spectrum are not widely available. Moreover, when furnace atomization is employed, it is often necessary to employ a different temperature program for each element.

One of the problems encountered in AAS is the generation of a broad-band spectral background. This background arises from absorption caused by residual molecules or molecular fragments in the atom source and by scattering from smoke or other airborne particulate matter. Because a hollow-cathode lamp is utilized as a light source, this molecular absorption or scattering cannot be distinguished from atomic absorption, since it simply attenuates the beam. If left uncorrected, this ''nonspecific absorption'' then makes element concentrations appear to be higher than they really are.

Not surprisingly, nonspecific absorption is more of a problem when furnace atomizers are employed than when a chemical flame is utilized. In a furnace, a dense smoke cloud often arises during atomization of the sample, whereas in a chemical flame the particulate matter is more thoroughly volatilized. Nevertheless, it is considered to be prudent to employ some form of background correction with either a furnace or flame atomizer.

Several alternative schemes for background correction have been developed for AAS and are covered in some detail in **Background Correction Methods in Atomic Absorption Spectroscopy**. The most common of these schemes, termed ''continuum-source background correction'' is particularly ingenious and is the first of such methods to be developed. It exploits the spectral difference between narrow-band atomic absorption and broad-band molecular absorption or scattering. However, two sources are required, a hollow cathode lamp and an auxiliary continuum source, usually a deuterium arc lamp. Still, it remains the most widely used method.

Another technique for background correction in atomic absorption utilizes the Zeeman effect. The Zeeman effect is a splitting of atomic lines that occurs when the atoms are present in a magnetic field. Accompanying this spectral splitting is a change in polarization of the spectral lines that are generated. By use of appropriate instrumentation, it is possible to distinguish atomic absorption from the background features. Although no auxiliary source is required, a magnetic field and polarizers usually are.

Another approach employed for background correction in atomic absorption utilizes a pulsed hollow cathode and is similar in concept to the continuum-source procedure. In essence, the same hollow cathode lamp is used as both a narrow-band and broad-band source. The broad-band spectrum arises from spectral line broadening that occurs at unusually high hollow cathode operating currents.

In contrast to AAS, AES is inherently a multielement procedure. In a high-temperature flame or plasma (such as the ICP), atoms are not only formed extremely efficiently, but also they generate intense atomic emission. The emission is isotropic and also occurs from all elements at the same time. It is therefore possible to perform multielement determinations simply by means of a multichannel detection system. Multichannel devices are routinely being introduced that employ two-dimensional spectral dispersion coupled with two-dimensional arrays of detector elements of unprecedented sensitivity and low noise.

Like AES, AMS is inherently a multielement approach. Moreover, AMS can provide information about isotopes present in a sample. Although a small spectral shift occurs in atomic emission spectra because of the ''isotope effect'', for most elements in the periodic table that shift is too small to be detected with a conventional spectrometer. In contrast, AMS can readily distinguish one isotope from another; one of the important applications of AMS has therefore become the measurement of isotope ratios in various kinds of samples.

AMS offers also the benefit of extraordinarily low background levels. In the simplest sense, this low background results from the fact that ions in a mass spectrometer can possess energies that are considerably greater than energies corresponding to photons in the ultraviolet or visible region of the spectrum. For a photon of such low energy to be detected with high sensitivity requires naturally that the detector be very sensitive and accordingly sensitive to thermal effects. Thermal noise generated in such detectors therefore makes it more difficult to observe the atomic transitions of interest. In contrast, detectors employed in mass spectrometry can be considerably less sensitive to thermal noise because they can employ detection surfaces with a high ''work function''. The result is background count rates in AMS that can be below one count per second, whereas background count rates in AES are of the order of 10^4 cps. The result is that typical detection limits in AMS are lower than those in AES by a factor of 100 or so. Recent manufacturers' literature suggests that it is now routinely

possible to measure solution concentrations in ICP mass spectrometry at the level of one part per quadrillion or so.

For the ultimate in sensitivity, however, none of the above-mentioned techniques can better AFS. Indeed, a number of AFS studies over the last decade or so have demonstrated the detection of single atoms of a desired element. The reason for this extraordinary sensitivity is quite straightforward: in AFS, each atom can be detected many times, since it can be excited over and over again. Each time the atom emits a photon, it can be re-excited by an incident beam of photons, probably from a laser. In contrast, detection in AMS is usually destructive, so there is only a single chance to observe each atomic ion. Although AES also offers the opportunity to collect several photons from each atom, the background levels are usually high enough that it is difficult to distinguish the few photons an atom emits from others that a detector registers.

Several properties of a laser make it useful not only as a source for AFS, but for other measurements in atomic spectrometry as well. These features are documented in greater detail in **Laser Spectrometric Techniques in Analytical Atomic Spectrometry**. In short, the laser can deliver light to a sample very effectively, so it can be used not only as a source for laser sampling (laser ablation) and LIBS, as indicated earlier, but also as a source for atomic magneto-optic rotation spectrometry, for AAS, and for other more exotic measurements. To date, the widespread acceptance of these alternative methods has been limited by the cost and complexity of current laser systems. However, recent advances and engineering developments suggest that such impediments might be overcome in the future.

From this brief narrative, it should be clear that atomic spectrometric analysis is not only an extremely important area of chemical analysis, but also one that continues to evolve rapidly. Over the past decade, the nature of instrumentation for atomic spectrometry has changed dramatically. Benchtop units are now available that measure virtually the entire atomic emission spectrum of a complex chemical sample at once. Similarly, alternative kinds of mass spectrometric instrumentation (quadrupole mass filters, sector-field instruments, time-of-flight mass spectrometers, ion traps, and Fourier transform mass spectrometers) have been applied to AMS. These alternative approaches enable samples to be analyzed with unprecedented speed, accuracy, and precision. They permit isotope ratios to be determined on samples having elemental concentrations below 1 ppt and at a speed that would have been considered impossible only a few years ago. Because of this rapid development, the reader should consider the treatments in this section to be an excellent introduction to the field, but that continuing acquaintance will be necessary to keep abreast of ongoing developments.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
AFS	Atomic Fluorescence Spectrometry
AMS	Atomic Mass Spectrometry
ETA	Electrothermal Atomizer
FIA	Flow Injection Analysis
ICP	Inductively Coupled Plasma
LIBS	Laser-induced Breakdown Spectroscopy
MIP	Microwave-induced Plasma
NAA	Neutron Activation Analysis

REFERENCES

1. A. Walsh, 'The Application of Atomic Absorption Spectra to Chemical Analysis', *Spectrochim. Acta*, **7**, 108–117 (1955).
2. C.T.J. Alkemade, J.M.W. Milatz, 'Double-beam Method of Spectral Selection with Flames', *J. Opt. Soc. Am.*, **45**, 583–584 (1955).

Background Correction Methods in Atomic Absorption Spectroscopy

Margaretha T.C. de Loos-Vollebregt
Delft University of Technology, Delft, The Netherlands

1 Introduction 1
2 Two-line Background Correction 2
3 Continuum-source Background Correction 3
3.1 History 3
3.2 Principle 3
3.3 Instrumentation 3
3.4 Applications 5
4 Zeeman Background Correction 5
4.1 Zeeman Effect 5
4.2 Principle of Zeeman Background Correction 7
4.3 Instrumentation 9
4.4 Applications 10
5 Pulsed Lamp Background Correction 12
5.1 Principle 12
5.2 Instrumentation 12
5.3 Applications 12
6 Wavelength-modulation Correction Methods 13
6.1 Principle 13
6.2 Instrumentation 13
6.3 Applications in Continuum-source Atomic Absorption Spectroscopy 13
7 Comparison of Methods 14
7.1 Sensitivity and Dynamic Range 14
7.2 Frequency of the Background Correction System 15
7.3 Background Correction Performance 16
8 Background Correction Problems and How to Avoid Them 16
Abbreviations and Acronyms 17
Related Articles 17
References 17

Atomic absorption spectroscopy (AAS) is based on the absorption of element-specific primary source radiation by analyte atoms. If part of the radiation is absorbed by molecules or lost due to scattering, a higher gross absorbance is measured. The difference between the net absorption of the analyte atoms and the measured gross absorbance is called background absorbance. Background absorption and scattering effects have much more serious effects on the results produced in electrothermal atomization atomic absorption spectroscopy (ETAAS) than flame atomic absorption spectroscopy (FAAS).

The amount of incident light deflected or absorbed by nonatomic species must be measured to obtain the correct, net absorbance of the analyte atoms only. Perfect background correction can be obtained only when the background absorbance measurement corresponds exactly in space, time and wavelength with the atomic absorbance measurement. Since exact coincidence of all three parameters is obviously impossible, it is customary to give priority to the equality in space and to make the difference in time and/or wavelength as small as possible. Although molecular absorption and radiation scattering are both broad-band phenomena, there is no spectral range where constant background attenuation can be guaranteed.

In all background correction systems two measurements are made. The total or gross absorbance is measured at the wavelength of the resonance line emitted by the hollow cathode lamp (HCL). The background attenuation is then subtracted from the measured analyte absorbance to obtain the analyte absorbance. There are several ways in which the nonatomic absorption at the resonance wavelength can be estimated, including two-line background correction, continuum-source or deuterium lamp background correction, Zeeman background correction, pulsed lamp or Smith–Hieftje background correction and wavelength-modulation correction methods. The principle of each method, the instrumentation and applications are discussed. Continuum-source background correction is widely used in FAAS and ETAAS, whereas Zeeman background correction and pulsed lamp background correction are often preferred in ETAAS.

1 INTRODUCTION

AAS is based on the absorption of element-specific primary source radiation by analyte atoms. If part of the radiation is absorbed by molecules or lost due to scattering, a higher gross absorbance is measured. The difference between the net absorption of the analyte atoms and the measured gross absorbance is called background absorbance. In fact, we do not know which part of the background absorbance arises from scattering and which part is due to molecular absorption.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Scattering of primary source radiation from non-volatilized particles much smaller than the wavelength (λ) follows Rayleigh's law.The intensity of the scattering depends on the wavelength as λ^{-4}, so that short wavelengths are scattered more than than long wavelengths. Scattering is directly proportional to the number of scattering particles per unit volume and to the square of the particle volume. Consequently, radiation scattering occurs more strongly with increasing particle size. In FAAS, well-designed nebulizers and spray chambers are used and therefore virtually no unvolatilized particles pass into the optical beam. In graphite furnace AAS, considerable numbers of particles can be observed. They may arise from organic sample materials during the pyrolysis step or from volatilized inorganic materials that condense in the cooler parts of the atomizer. Small solid particles can also appear owing to boiling, sputtering and incomplete vaporization. Broad-band molecular absorption of radiation is caused by molecular species formed or vaporized in the atomizer during the atomization step, particularly alkali and alkaline earth metal halides. Sharp fine structure of the vibrational bands in electronic spectra of the molecules can be superimposed on broad-band background absorption. The effects of molecular absorption in FAAS have long been recognized. The $CaOH^+$ band has a maximum near the barium resonance line at 553.6 nm. Molecular absorption spectra of a number of halides, nitrates and sulfates have been reported. As an example, Figure 1 shows the continuous dissociation spectra of SO_3 and of SO resulting from volatilization of sulfates in a graphite furnace, reported by Massmann and Güçer[1] and Newstead et al.[2] Massmann and Güçer[1] also measured emission and absorption spectra resulting from the CN system with band heads at 386.2, 387.1 and 388.3 nm using nitrogen as the inert gas in ETAAS. Background absorption and scattering effects have much more serious effects on the results produced in ETAAS than in FAAS.

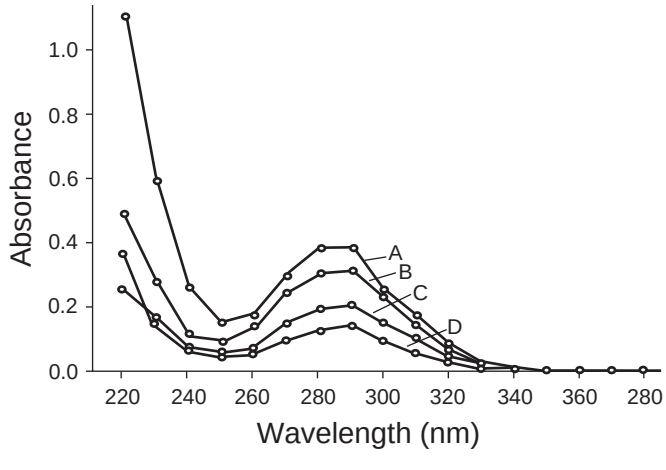

Figure 1 Continuous dissociation spectra of SO_3 (260–330 nm) and SO ($\lambda < 245$ nm) resulting from volatilization of sulfates in a graphite tube furnace. A, 0.01 g mL^{-1} Mg; B, 0.01 g mL^{-1} Fe; C, 0.01 g mL^{-1} Ni; D, 0.01 g mL^{-1}; Zn. Sample volume, 20 mL. (Reprinted from R.A. Newstead, W.J. Price, P.J. Whiteside, *Prog. Anal. At. Spectrosc.*, **1**, 267–298 (1978), Copyright 1978, with permission from Elsevier Science.)

The amount of incident light deflected or absorbed by nonatomic species must be measured to obtain the correct, net absorbance of the analyte atoms only. In electrothermal atomization the spatial and temporal distribution of the atoms and the background-generating species may be entirely different.[3] Perfect background correction can be obtained only when the background absorbance measurement corresponds exactly in space, time and wavelength with the atomic absorbance measurement. Since exact coincidence of all three parameters is obviously impossible, it is customary to give priority to the equality in space and to make the difference in time and/or wavelength as small as possible. Although molecular absorption and radiation scattering are both broad-band phenomena, there is no spectral range where constant background attenuation can be guaranteed.

In all background correction systems two measurements are made. The total or gross absorbance (A_{measured}) of the atoms and the nonatomic species is measured at the wavelength of the resonance line emitted by the HCL. The background attenuation (A_{bg}) is then subtracted from the measured absorbance to obtain the analyte absorbance (A_{analyte}) (Equation 1):

$$A_{\text{analyte}} = A_{\text{measured}} - A_{\text{bg}} \tag{1}$$

There are several ways in which the nonatomic absorption at the resonance wavelength can be estimated.

2 TWO-LINE BACKGROUND CORRECTION

In the two-line background absorption approach, the total absorbance is measured at the resonance line emitted by the HCL and the nonspecific absorbance is measured at a nearby line of the lamp, where the analyte atoms do not absorb.

Two-line background correction was first proposed by Willis,[4] using a nonresonant analyte line or a line from an element that is emitted by the source but not present in the sample. Argon or neon lines can also be used. Alternatively, another source, of an element which is not found in the sample, can be inserted for the background measurement. The background spectral character must be known to be the same at the second line as it is at the primary line and therefore the closer the second line is to the primary line, the better will be the correction. These

conditions cannot be met for all elements and all matrices. In some wavelength regions, the steep slope of molecular absorption bands makes it difficult to find a nonabsorbing line close enough to the analyte line to ensure an accurate background correction. The atomization conditions for sequential measurements should be as similar as possible, even if an appropriate nonabsorbing line is available from the primary source itself. Dual-channel AAS instruments offer the possibility of simultaneous measurement at two lines. The monochromator of the first channel is set to the chosen absorption line for the analyte element and the second channel monochromator is set to the chosen line for the background correction. The signal-handling system of the instrument subtracts the absorbance measured in the second channel from the absorbance signal measured in the first channel. Two-line background correction went into practice earlier than the other methods. However, it has not been widely used. Lines have been recommended for background correction for the most commonly determined elements.[2] It is important that the chosen line should be close to the line of the analyte element. Because the absorption signal from background scatter varies with wavelength and increases with decreasing wavelength, the background correction error may become substantial at wavelengths below 220 nm if the line chosen is more than 1 nm from the analyte line. The use of an automatic system, chopping rapidly between both lines, is essential if the system is used with an electrothermal atomizer because measurements made on two successive furnace cycles are unlikely to show the same background signal.

3 CONTINUUM-SOURCE BACKGROUND CORRECTION

3.1 History

To improve the sensitivity of AAS, Koirtyohann and Picket[5] used a 40 cm long, 10 mm i.d. tube that was heated in a hydrogen–oxygen flame, providing a much longer effective absorption path than was previously used in the more conventional burners. They observed that, at the higher sensitivities, absorption by matrix salts at the wavelength of an elemental resonance line can cause significant errors and proposed to use a hydrogen lamp to correct for matrix absorption. In fact, this background correction system is still widely used to correct for background absorption in FAAS and ETAAS.

3.2 Principle

The gross absorbance A_{measured} is the sum of the atomic absorbance of the analyte atoms A_{analyte} and the background attenuation A_{bg}. The gross absorbance is measured with the narrow HCL emission line of the element at the selected specific wavelength of the resonance line. An estimate of the background absorbance is obtained as an average over the spectral band-pass of the monochromator using a continuum-source of radiation, e.g. a hydrogen lamp or deuterium lamp in the ultraviolet (UV) wavelength region or a tungsten lamp in the visible part of the spectrum. The width of the atomic absorption line is about 100 times smaller than the band-pass of the monochromator, so that the analyte absorption of the continuum-source radiation can be neglected.

3.3 Instrumentation

For the correction of background attenuation, Koirtyohann and Picket[5] used a dual-channel system in which the radiation from an HCL was passed alternately with the radiation of a continuum hydrogen lamp through the flame. Nowadays, a deuterium lamp is most frequently used as the continuum-source. Figure 2 shows the basic schematic configuration of a continuum-source background correction system for FAAS. By using lamp pulsation or by means of a rotating chopper with a sector mirror, the radiation from the HCL with intensity I^0_{HCL} and the radiation from the deuterium lamp with intensity $I^0_{\text{D}_2}$ are passed alternately through the atomizer. After passing the monochromator, both radiation beams reach the same detector and the net atomic absorbance is derived.

The operation of the continuum-source background correction system is explained in Figure 3(a–f). The monochromator separates the resonance line from the emission spectrum of the element-specific HCL. The width of this line is a few picometers. From the deuterium lamp emission spectrum the monochromator isolates a band equivalent to the spectral band-pass, usually 0.2–1 nm. The primary intensities of the two beams of radiation are equalized to obtain $I^0_{\text{HCL}} = I^0_{\text{D}_2}$. When analyte atoms absorb HCL radiation, the intensity I^0_{HCL} decreases with an amount corresponding to the atom concentration. Naturally, $I^0_{\text{D}_2}$ is also attenuated at the resonance wavelength, but since the half-width

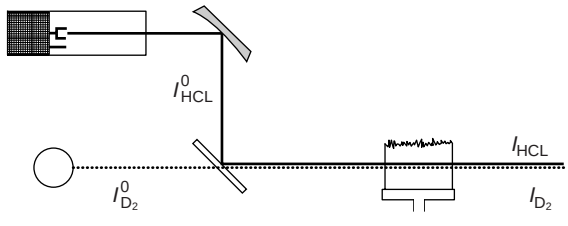

Figure 2 Schematic configuration of a continuum-source background correction system. Radiation from the HCL (I_{HCL}) is passed alternately with radiation from the deuterium lamp (I_{D_2}).

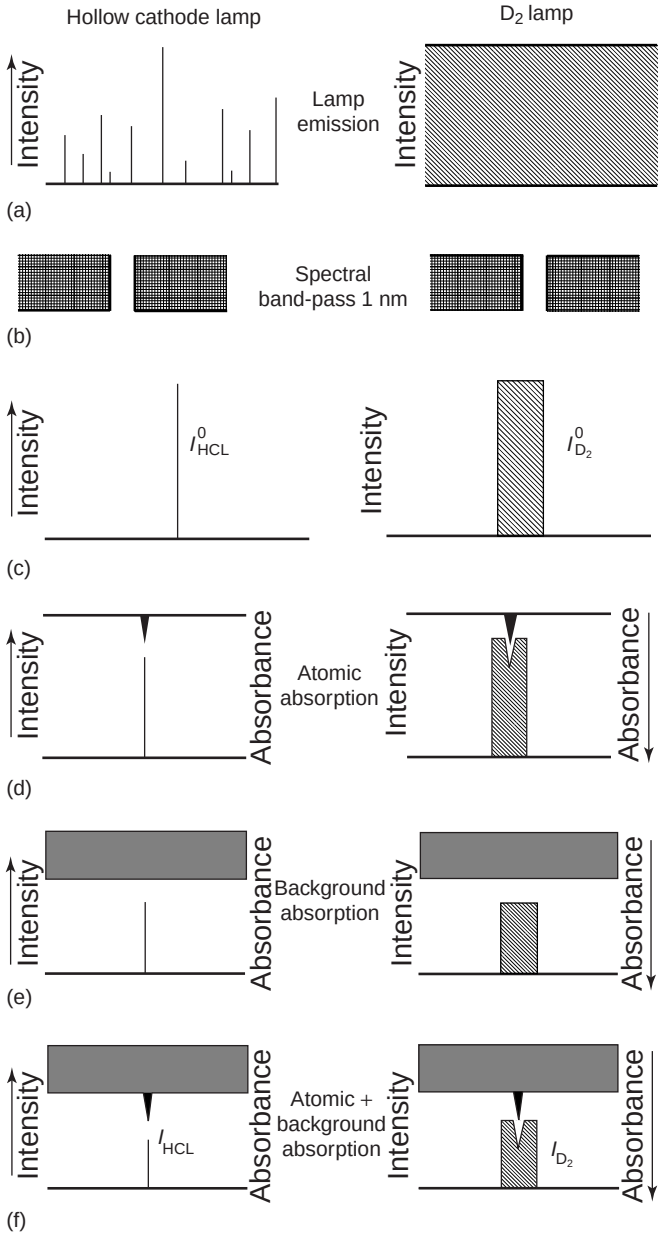

Figure 3 Mode of operation of a continuum-source background correction system. (a) The line emission spectrum of the HCL and the continuum emitted by the deuterium lamp; (b) the monochromator isolates the resonance line from the spectrum of the HCL and passes a band of radiation from the deuterium lamp corresponding with the spectral band-pass; (c) the intensities I^0_{HCL} and $I^0_{D_2}$ are equalized; (d) for atomic absorption by the analyte element I^0_{HCL} is attenuated by an amount corresponding to its concentration whereas $I^0_{D_2}$ is not significantly attenuated; (e) broad-band background attenuates the intensity of both sources to the same degree; (f) atomic absorption by the analyte in addition to the background attenuates the HCL intensity further as in (d) whereas the deuterium lamp intensity is not further attenuated. (Reproduced by permission of Wiley-VCH from B. Welz, *Atomic Absorption Spectrometry*, 2nd edition, VCH, Weinheim, 1985.)

of the atomic absorption line is about 5 pm whereas the continuum has a width equal to the spectral band-pass of 0.2–1 nm, the absorption of continuum radiation by the analyte atoms can be neglected. If nonspecific attenuation of the radiation occurs, whether through scattering or molecular absorption, both radiation beams will be attenuated to the same extent since background attenuation is usually a broad-band phenomenon. In continuum-source background correction it is assumed that the background is constant over the observed spectral range, i.e. the attenuation of $I^0_{D_2}$ is measured as the average over the spectral band-pass and used for background correction at the wavelength position of the analyte absorption line. If the background attenuation is not continuous within the observed spectral range, positive or negative background correction errors will occur.

The two lamps are observed by the detector alternately in time. Instrument electronics separate the signals and compare the absorbances from both sources. A net absorbance will be displayed only when the absorbance of the two lamps differs. Since the background attenuates both sources equally, it is ignored. Analyte atoms absorb the HCL radiation and negligibly absorb the broad-band continuum-source emission and it is therefore still measured and displayed as usual without background correction. Most modern AAS instruments permit almost simultaneous measurement of the uncorrected signal or background signal and the corrected signal. The speed of the background correction system should be high enough to follow the fast transient signals in ETAAS (see section 7.2).

Background correction with a continuum-source has some disadvantages. The alignment and superposition of the HCL or electrodeless discharge lamp and the continuum-source are critical. Both lamps are located at different positions in the instrument, the shape of the beams is different and the intensity distribution in the beam is not the same. Even if it is possible to align the beams exactly, different absorption volumes may be irradiated. The application of two light beams requires a more complicated optical system and in many instruments double-beam operation is available to compensate for intensity changes of the lamps. The deuterium lamp emits sufficient radiation in the short wavelength range from 190 nm to 330 nm, whereas a halogen lamp is used for background corrections above this range. In general, the signal-to-noise ratio (S/N) is worsened through the use of two radiation sources and the more complicated optical system that is used to measure their intensities. If a continuum-source is used outside its optimum range, the radiant intensity of the line source must be reduced, leading to further detoriation of the S/N. Finally, it should be noted that the principle

of operation of continuum-source background correction limits successful use to continuous background absorption only.

3.4 Applications

Many commercial AAS systems are based on deuterium lamp background correction. In FAAS and hydride generation AAS it is the most widely applied background correction system. It is fully adequate for all applications of FAAS, except in some very unusual circumstances. During methods development, background correction is considered. In many applications, there is no need to use background correction at all. In general, in the lower UV wavelength region (<220 nm) background correction is required.

Background correction systems employing continuum-sources are incapable of correcting the background attenuation of electronic excitation spectra because these comprise many narrow lines. The actual background correction that is required depends on the degree of overlap between the elemental spectral line and the individual molecular rotational lines. An example is the absorption line of gold at 267.6 nm that lies exactly in the middle between two rotational lines of indium chloride so that the actual background attenuation is much lower than the mean background absorbance found over the spectral range. Similar interferences have been reported for the determination of Bi at 306.8 nm and Mg at 285.2 nm through absorption bands of OH in an air–acetylene flame and also for Fe at 247.3 nm, Pd at 244.8 and 247.6 nm and Yb at 246.4 nm due to PO molecular absorption.[6] Another possible source of error with continuum-source background correction may arise from absorption by concomitant atoms from the matrix. This type of spectral interference has been reported by Manning[7] for the determination of Se in the presence of Fe. Problems with continuum-source background correction in FAAS seldom occur. The background correction system is an extremely useful device for FAAS that frequently leads to the correct results.

However, the continuum-source background correction system is not free from pitfalls. The limited energy of the deuterium lamp lowers the S/N and the correction becomes inaccurate when the background is structured. Most users distrust their results when the background level exceeds 0.5 absorbance unit. The limitations are probably related to the different geometries and possible misalignment of the optical beams of the HCL and the deuterium lamp. The increasing interest in ETAAS as a selective and highly sensitive analytical method has stimulated the development of improved background correction systems.

4 ZEEMAN BACKGROUND CORRECTION

4.1 Zeeman Effect

In 1897, the Dutch physicist Pieter Zeeman[8] (Figure 4) reported the observation that if a source of light is placed between the poles of a magnet the lines appear to broaden and, if the magnetic field is strong enough and the resolving power of the spectroscope is sufficiently high, to spit into a number of components. The first splitting patterns studied by Zeeman follow from lines of a singlet series. A single line was split into three polarized components.

Both splitting and polarization were explained by Lorentz with the classical electromagnetic theory. A few years later it was demonstrated that the lines of other series do not split into the simple triplet but to a group of four or more components. The classical theory was unable to account for this. Therefore, the two types of splitting were labeled the normal and the anomalous Zeeman effects. An example of the

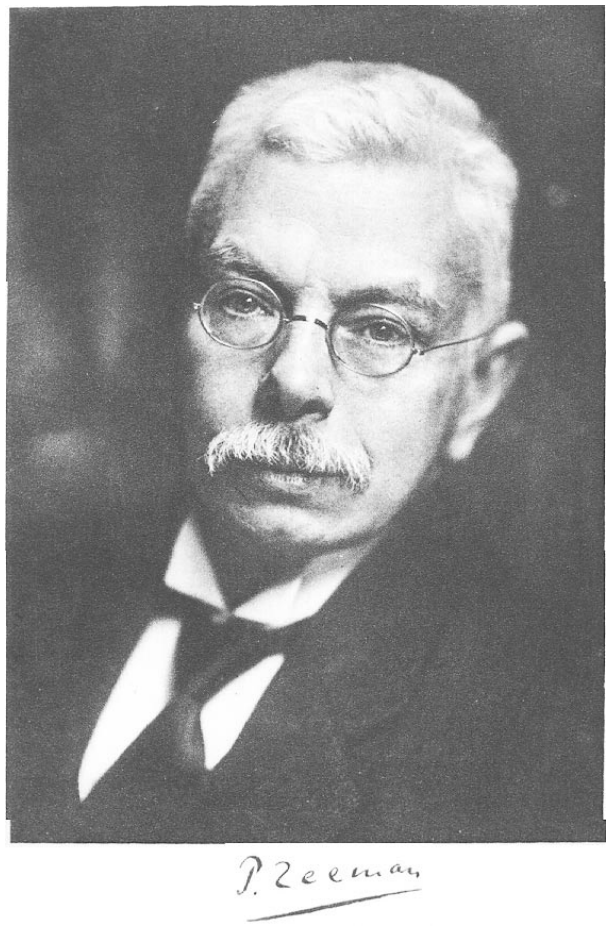

Figure 4 Dr Pieter Zeeman (1865–1943). Dutch physicist who discovered the influence of a magnetic field on radiation in 1896.

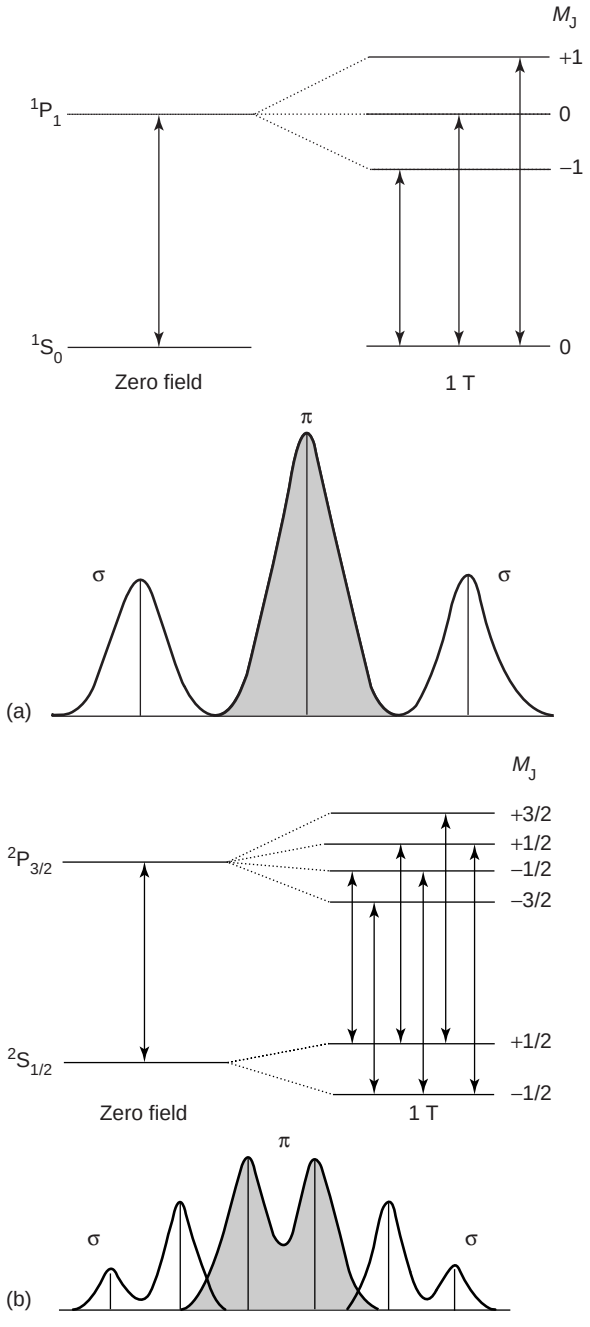

Figure 5 Magnetic splitting of atomic energy levels and the corresponding Zeeman splitting patterns. (a) $^1S_0 - {}^1P_1$ transition, normal Zeeman effect; (b) $^2S_{1/2} - {}^2P_{3/2}$ transition, anomalous Zeeman effect. (Reprinted from M.T.C. de Loos-Vollebregt, L. de Galan, *Prog. Anal. At. Spectrosc.*, **8**, 47–81 (1985), Copyright 1985, with permission from Elsevier Science.)

normal Zeeman effect is presented in Figure 5(a) and Figure 5(b) shows an example of the anomalous Zeeman effect. The splitting pattern is always symmetrical about the position of the nondisplaced line. The components polarized parallel to the magnetic field are referred to as π-components and the components polarized perpendicular to the magnetic field are referred to as σ-components (σ deriving from the German senkrecht). In the case of a normal Zeeman triplet, the π-component is located at the position of the original line at frequency ν and two σ-components are found at frequencies $\nu \pm \Delta\nu$. Such a Zeeman splitting is observed from the direction perpendicular to the magnetic field (transverse Zeeman effect). In the direction of the magnetic field, the central component is not observed whereas the two displaced components are circularly polarized and rotate in opposite directions (longitudinal Zeeman effect). Lines that do not belong to a singlet series split into more than three components. Their pattern is always symmetrical with respect to the position of the original line, with a central group of π-components and groups of σ-components on either side as illustrated in Figure 5(b) for a $^2S_{1/2} - {}^2P_{3/2}$ transition. All Zeeman splitting patterns are observed in both absorption and emission.

The sum of the intensities of the π-components is equal to the sum of the intensities of the σ-components and if the various components are combined, the resulting beam is nonpolarized and the intensity is equal to the intensity observed without the magnetic field. The energy states generated around an original energy level E^0 are described by Equation (2):

$$E = E^0 + \mu_B B M_J g \tag{2}$$

where μ_B is the Bohr magneton ($9.274 \times 10^{-24}\,J\,T^{-1}$), B is the magnetic flux density (tesla), M_J is the magnetic quantum number and g is the Landé factor. Both emission and absorption transitions between energy levels are governed by the selection rule $\Delta M_J = 0, \pm 1$, where $\Delta M_J = 0$ represents the π-components and $\Delta M_J = \pm 1$ the σ-components. The displacement of each Zeeman component is proportional to the strength of the magnetic field. In a magnetic field of 1 T the displacement of the σ-components in the normal Zeeman effect is $0.467\,cm^{-1}$, which is equal to 4 pm at a wavelength of 300 nm. The normal Zeeman effect is shown by Be 234.9 nm, Mg 285.2 nm, Ca 422.7 nm, Sr 460.7 nm, Ba 553.6 nm, Cd 228.8 nm and Zn 213.9 nm. Several lines that belong to a triplet series also split into one π-component and two σ-components. However, the shift of their σ-components varies from half to twice the shift in the normal Zeeman effect, e.g. Ge 265.2 nm (1.5), Hg 253.7 nm (1.5), Hg 184.9 nm (1.0), Pb 217.0 nm (0.5), Pb 283.3 nm (1.5), Pd 244.8 nm (1.0), Si 251.6 nm (1.5) and Sn 286.3 nm (1.5). All other transitions used in AAS show anomalous Zeeman splitting patterns that consist of at least two π-components and at least two σ-components. The Zeeman splitting pattern in Figure 5(b) is observed for Ag 328.1 nm, Cu 324.8 nm, Au 242.8 nm, Na 589.0 nm and K 766.5 nm.

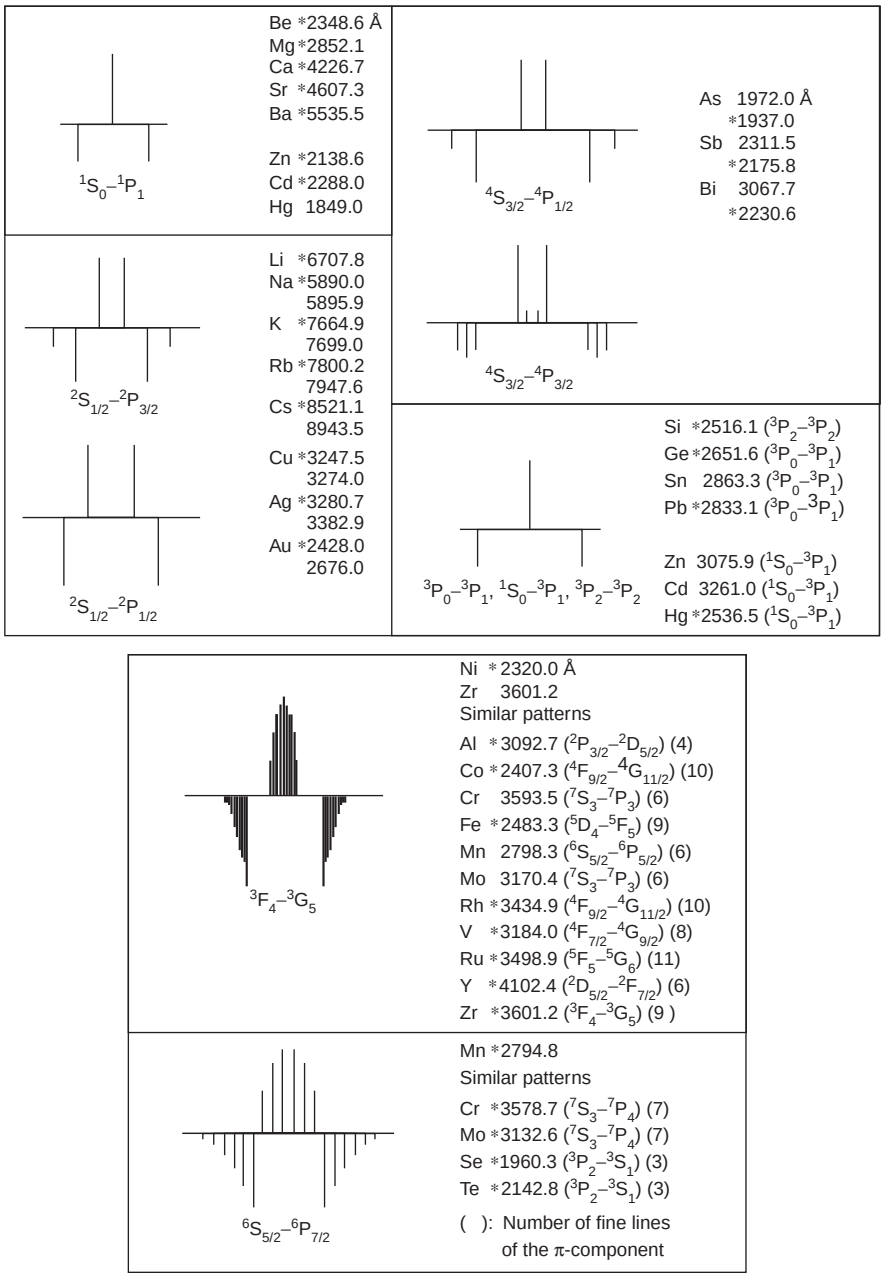

Figure 6 Classification of Zeeman patterns on various elements. * indicates lines most frequently used in AAS. 1 Å = 10^{-10} m. (Reprinted from H. Koizumi, K. Yasuda, *Spectrochim. Acta, Part B*, **31**(10–12), 523–535 (1976), Copyright 1976, with permission from Elsevier Science.)

Zeeman splitting patterns of the absorption lines used in AAS of various elements were classified in several groups by Koizumi and Yasuda.[9] Figure 6 shows the Zeeman splitting patterns for various transitions that are used in AAS.

4.2 Principle of Zeeman Background Correction

In AAS, the non- or slightly shifted π-components are used to measure the total absorbance of the analyte atoms and the background whereas the shift of the (groups of) σ-components from the original line position is used to measure the background absorbance. Figure 7(a) shows the resonance line emitted by the HCL and the absorption line of the atoms in the graphite furnace at zero magnetic field strength. Note that the width of the absorption line is about twice the width of the lamp emission line owing to the higher temperature and higher pressure in the atomizer compared to the HCL.

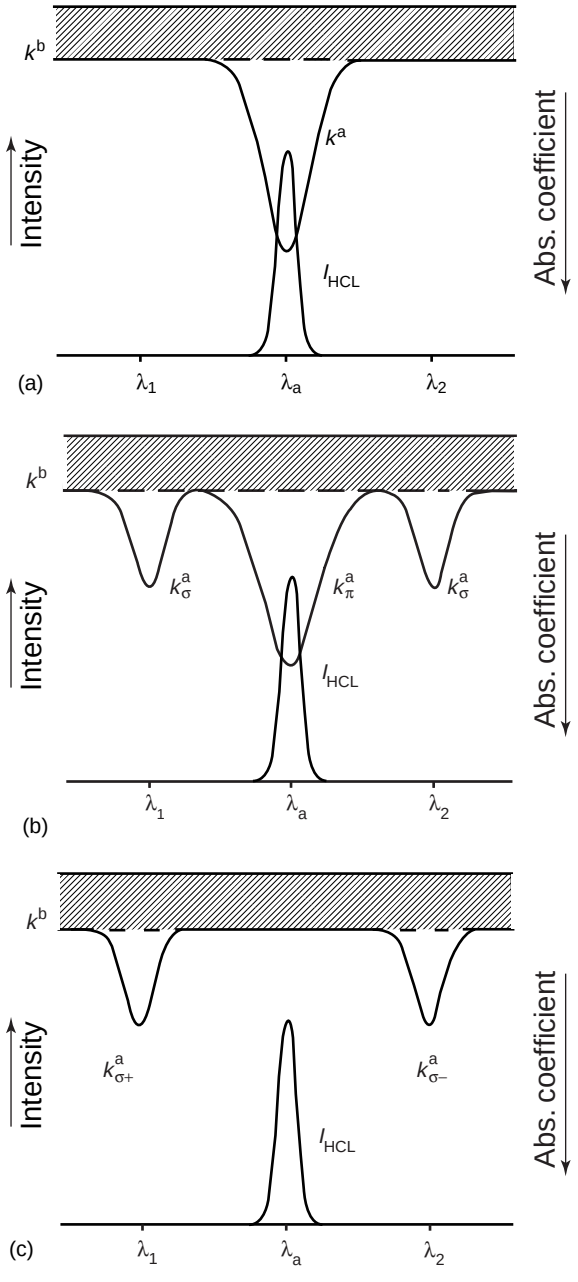

Figure 7 Profiles of the HCL analyte emission line (I_{HCL}) and the absorption coefficient of the analyte (k^a) at wavelength λ_a in the presence of background absorbance (k^b) for the situation of normal Zeeman effect. (a) Conventional AAS, no magnetic field; (b) in the presence of a transverse magnetic field; (c) in the presence of a longitudinal magnetic field.

Figure 7(b) shows the situation where the graphite furnace is in a transverse magnetic field of 1 T. Absorption takes place within the π-components of the absorption line profile k^a_π whereas the σ-components of the absorption line profile k^a_σ are shifted away to wavelengths λ_1 and λ_2. When a π-component rejecting polarizer is positioned in the optical beam and when the magnetic field is switched on and off at high enough frequency, e.g. a 50 Hz alternating current (AC) magnetic field, the gross absorbance and the background absorbance can be measured alternately. The net analyte absorbance follows then from Equation (1). The background absorbance k^b is measured exactly at the wavelength of the analyte resonance line λ_a and within the width of the HCL emission line profile. Consequently, the background absorbance is measured correctly for any type of background, even if it is highly structured. The alternative approach is to use a constant magnetic field from a permanent magnet or a direct current (DC) magnet and to rotate the polarizer to permit alternate measurements with π-polarized radiation and σ-polarized radiation. The π-polarized radiation is then absorbed by the π-component of the analyte absorption line profile k^a_π and the background k^b providing the gross absorbance. The σ-polarized radiation is absorbed by the background only because the σ-components of the analyte absorption line profile are shifted to wavelengths λ_1 and λ_2 and they are too far away to absorb the lamp emission line. Combination of the two measurements again provides the net analyte absorption.

Another possibility is to use a longitudinal magnetic field rather than a transverse magnetic field. Figure 7(c) shows that in the direction of the magnetic field only the circularly polarized σ-components of the analyte absorption line profile $k^a_{\sigma+}$ and $k^a_{\sigma-}$ are observed. In a longitudinal AC modulated magnetic field the gross absorbance is measured at zero-field (see Figure 7a) and the absorbance of the background is measured at exactly the wavelength position of the lamp emission line during the magnet on period. A longitudinal AC Zeeman background correction system does not require a polarizer and therefore the full intensity of the HCL resonance emission line can be used in both measurements.

In all AAS instruments using Zeeman background correction, two alternate intensity measurements are performed. One refers to the unshifted zero-field analyte absorption line (in an AC field) or the slightly shifted and broadened π-components (in a constant field). The intensity is given by Equation (3):

$$I_1 = I_1^0 \exp(-k_1^a) \exp(-k_1^b) \tag{3}$$

where I_1^0 is the incident intensity, k_1^b is the absorption coefficient of the background and k_1^a is the analyte absorption coefficient. The other measurement refers to the situation when the σ-components are shifted away from the original line position (AC field) or when σ-polarized radiation is used (constant field). The intensity is given by Equation (4):

$$I_2 = I_2^0 \exp(-k_2^a) \exp(-k_2^b) \tag{4}$$

where I_2^0 is the incident intensity, k_2^b is the absorption coefficient of the background and k_2^a is now the remaining analyte absorption coefficient of the shifted σ-components, which should be close to zero. A simple subtraction of the absorbances measured in both situations corresponds to taking the log ratio of the intensities (Equation 5):

$$\ln\left(\frac{I_2}{I_1}\right. = \ln\left(\frac{I_2^0}{I_1^0}\right. + (k_1^a - k_2^a) + (k_1^b - k_2^b) \qquad (5)$$

This expression forms the basis for all Zeeman background correction systems in AAS. When the incident intensities I_1^0 and I_2^0 are equal and when the same background attenuation is measured, the net analyte absorption follows from Equation (6):

$$A = \log\left(\frac{I_2}{I_1}\right. = 0.43(k_1^a - k_2^a) \propto N_A \qquad (6)$$

It is clear that a linear analytical response proportional to the number of analyte atoms N_A is obtained. The assumption that I_1^0 is equal to I_2^0 is reasonable because it is the same beam of radiation in the case of the AC magnetic field, or the π-polarized and the σ-polarized parts of the same beam in the case of the constant magnetic field. The background absorbance is measured within the analyte emission line profile in all Zeeman background correction systems with the magnetic field applied to the atomizer. The only requirement is that the background is not influenced by the magnetic field when an AC magnet is used. From Equation (6) it can be concluded that the σ-components in the Zeeman splitting pattern should be shifted as far away as possible from the original line position to reduce k_2^a to a minimum, i.e. a strong magnetic field should be applied, especially when the analyte absorption line is broad by nature or when its Zeeman splitting pattern is relatively narrow. In contrast, maximum sensitivity also requires that k_1^a is large, i.e. close to the conventional AAS absorption coefficient. Optimum sensitivity is obtained when a strong AC magnetic field is used. In a constant magnetic field, the optimum magnetic field strength compromises a large shift for the σ-components simultaneous with minimum shift in the π-components. Table 1 shows a comparison of relative sensitivities for several elements obtained in a constant magnetic field of 1 T, the best sensitivity that can be obtained for the element at optimized constant magnetic field strength and, for comparison, also the corresponding sensitivity obtained in an AC magnetic field of 1 T.[10] Even at optimum field strength, the sensitivity is reduced by about a factor of two for many of the analyte absorption lines that show the anomalous Zeeman effect. This explains why the AC modulated magnetic field is preferred in many of the commercially available AAS instruments with Zeeman background correction.

In principle, Zeeman background correction systems can have the magnetic field either around the atomizer or around the primary source of radiation. When the primary source of radiation is in the magnetic field, σ-components of the lamp emission line are located at the wavelength positions λ_1 and λ_2. These lines are used to measure the background absorbance at both sides of the analyte atomic absorption line. Equations (3–6) apply also to this situation.

4.3 Instrumentation

The application of the Zeeman effect for background correction in AAS was proposed by Prügger and Torge[11] in 1969. The authors claim that a shift of the primary

Table 1 Comparison of relative sensitivities in different Zeeman background correction systems[9] (conventional AAS sensitivity is equal to 1)

Element	Wavelength (nm)	Zeeman pattern[a]	Constant field 1 T	Constant field optimized[b]	AC field 0.8 T
Ag	328.1	A	0.28	0.61	0.91
Al	309.3	A	0.97	0.97	0.90
As	193.7	A	0.43	0.66	0.89
Cd	228.8	N	0.94	0.94	0.98
Cr	357.9	A	0.42	0.68	0.88
Cu	324.8	A	0.49	0.49	0.53
Fe	248.3	A	0.73	0.76	0.92
Mn	279.5	A	0.61	0.65	0.91
Pb	283.3	N	0.86	0.86	0.83
Sb	217.6	A	0.68	0.81	0.95
Se	196.0	A	0.45	0.51	0.88
Sn	286.3	N	0.97	0.97	0.94
Zn	213.8	N	0.90	0.90	0.88

[a] Zeeman splitting pattern: N = normal, A = anomalous.
[b] Variable constant magnetic field up to 1 T.

source resonance line of about 10 pm is required to allow successive measurements of the gross AAS signal and the background absorbance and propose a system based on magnetic splitting of the source line in a DC or AC magnetic field. The first commercial instrument became available in 1975, incorporating a 1.1 T permanent magnet around the atomizer.[12] The plane of polarization is rotated at 100 Hz to provide for background correction whereas the intensity of the HCL radiation is modulated at 1.5 kHz to eliminate the signal caused by the emission of radiation from the atomizer. Since that time, many other companies have marketed Zeeman AAS instruments. Only a few instruments have the magnetic field around the primary source of radiation. The main reason is that normal HCLs and electrodeless discharge lamps that are widely used in AAS do not fit into the gap of a magnet that has to provide a magnetic field strength close to 1 T at reasonable size, power requirements and price of the magnet. In addition, the direction of the discharge in the lamp should be parallel to the direction of the magnetic field to provide a stable beam of radiation. Special lamps have been designed for stable operation in a magnetic field. For nonvolatile elements, instruments were operated with glow discharge lamps in a permanent magnet, whereas low-frequency, high-voltage gas discharge lamps were used for volatile elements.

The more useful alternative is to apply the magnetic field to the atomizer, as discussed in the previous section. Only one manufacturer decided to apply Zeeman background correction to FAAS and therefore constructed a magnet with a gap as long as the conventional 10-cm burner that is routinely used in FAAS. In general, there is no need for Zeeman background correction in FAAS. The continuum-source background correction system works satisfactorily. In graphite furnace AAS, the transverse Zeeman background correction system in Figure 8(a) is marketed by various manufacturers of AAS instruments and most of them use an AC modulated magnetic field of about 0.8 T[10,13] in combination with a fixed polarizer to reject π-polarized radiation and transmit only σ-polarized radiation. The longitudinal Zeeman background correction system presented in Figure 8(b) was introduced by De Loos-Vollebregt et al.[14] and is also commercially available.[15] In this system, a transverse heated atomizer is placed within a longitudinal magnetic field and there is no polarizer in the optical system. This gives a considerable improvement in light throughput and simplifies the optical system so that detection limits are lower.

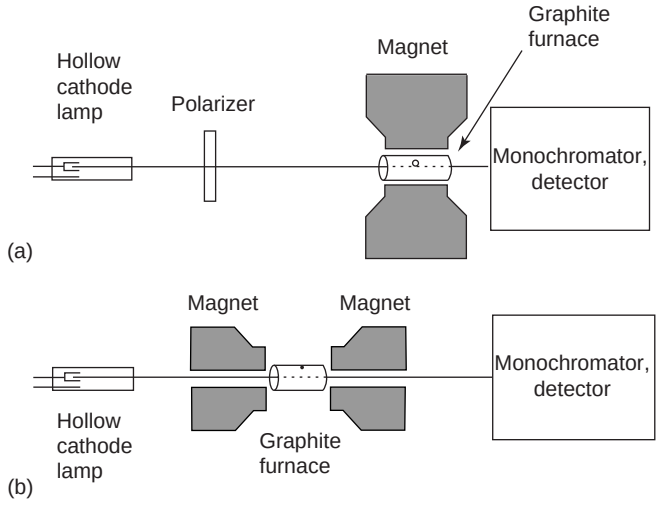

Figure 8 Schematic diagrams of different Zeeman background correction systems for AAS. (a) Transverse magnetic field, i.e. the magnetic field is perpendicular to the optical axis of the spectrometer. The polarizer provides alternately π- and σ-polarized radiation if the magnetic field is constant or transmits only σ-polarized radiation if an AC modulated magnetic field is applied; (b) longitudinal magnetic field oriented parallel to the optical axis of the spectrometer. No polarizer is required in this configuration.

4.4 Applications

Zeeman background correction is nowadays widely used in ETAAS. The background correction is performed exactly at the wavelength of the atomic absorption line. High nonspecific absorbances (up to about 2) can be corrected and in the AC Zeeman system sensitivities are similar to those for conventional AAS. Nevertheless, there are a few elements that show somewhat reduced sensitivity because their lines are relatively broad and/or the Zeeman splitting pattern is relatively narrow. An overview of the early applications[16] of Zeeman background correction shows that a wide variety of elements have been measured in different samples and certified reference materials and in the recent years many applications followed. The relatively volatile elements, e.g. Pb and Cd, have been measured by many analytical chemists using ETAAS with Zeeman background correction. This is not surprising because in the determination of volatile elements the pyrolysis temperature is not high enough to remove most of the matrix compounds that are responsible for the background absorption during atomization. Pb and Cd are often determined in a rather difficult matrix such as blood and urine and the low concentrations of the analyte elements do not permit dilution of the samples. The results obtained with Zeeman background correction have been compared with those obtained with continuum-source background correction, looking at the baseline of the net AAS signals. Zeeman background correction seems to be superior, although a fair comparison would require that the same atomizer and atomization conditions are used. There are an increasing number of solid

sampling applications where the Zeeman background correction system is the key to success.

In general, the dynamic range is slightly limited when Zeeman background correction is used. In a strong magnetic field the σ-components are shifted almost completely away from the absorption line profile so that, according to Equation (6), the shape of the AC Zeeman AAS analytical curve is completely determined by the zero-field absorbance. The analytical curve will therefore be very similar to the corresponding conventional AAS analytical curve measured without background correction. The relatively small contribution of the σ-components to the measurement of the background has only a minor influence on the slope and the curvature of the AC Zeeman AAS analytical curve. This is equally valid for the normal Zeeman effect and the anomalous Zeeman effect. The same is essentially true when a constant magnetic field is used as long as the spectral line displays a normal Zeeman effect. For the majority of the transitions that show an anomalous Zeeman effect, the splitting in the π-components broadens the absorption line profile and therefore the linearity of the Zeeman AAS analytical curve is somewhat improved, although the slope is lower.

At very high analyte concentrations, all the ETAAS signals show a dip (see Figure 9) when Zeeman background correction is applied, which can be explained from the phenomenon that is known as roll-over. From the very early days when Zeeman background correction was used, rumors were around that the analytical curve first increases, then reaches the maximum absorption level at relatively high analyte concentration and beyond the maximum decreases again towards the concentration axis. Such a double-valued behavior has indeed been reported for FAAS systems with Zeeman background correction. From a theoretical analysis of the phenomenon[16] it became clear that the nonabsorbed radiation that is emitted by the primary source and reaches the detector is responsible for this. With increasing analyte concentration, the zero-field absorbance or π-component absorbance, k_1^a in Equation (6), increases much more rapidly than the σ-component absorbance, k_2^a. With increasing analyte concentration, k_1^a levels off. The σ-component absorbance is subject to the same curvature, but at much higher concentration. At a certain concentration a situation is reached where k_1^a and k_2^a show the same rate of increase, or equal derivative with respect to concentration. At that concentration, the difference signal according to Equation (6) shows zero slope, i.e. the analytical curve reaches a maximum. Thereafter, a FAAS analytical curve would bend back to the concentration axis. The absorbance and concentration of roll-over are indicated in Figure 9, where the dotted line illustrates the decreasing absorbance values. The height and the position of the maximum depend on the strength of the magnetic field and the amount of nonabsorbed source radiation (so-called stray light). A list of typical roll-over absorbances for ETAAS published by Slavin and Carnrick[17] is given in Table 2. The characteristic masses are also presented.

Figure 9 shows the net ETAAS analyte absorbance signals. At the lower analyte concentrations, the signals are similar to those in conventional AAS. However, beyond the concentration where the maximum peak absorbance is reached, the signals show a dip. The dip becomes deeper with increasing concentrations. In peak absorbance, the analytical curve levels off when the absorbance of roll-over is reached whereas in integrated absorbance the analytical curve continues to increase but the slope becomes much smaller. It is clear that there is no risk of misunderstanding the measured peak absorbances or integrated absorbances in ETAAS. At high analyte

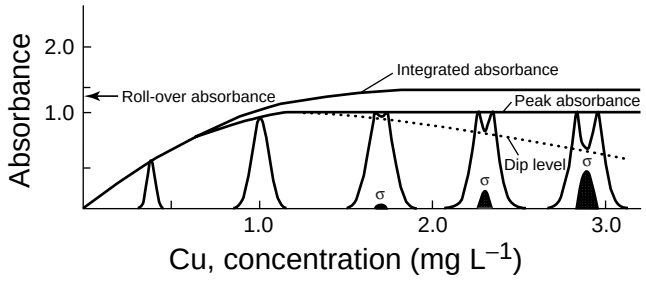

Figure 9 Analytical curve for Cu 324.8 nm measured in ETAAS with Zeeman background correction at magnetic field strength 0.8 T. The peak absorbance reaches a maximum at high analyte concentration whereas the integrated absorbance increases slowly beyond the roll-over absorbance.

Table 2 Roll-over absorbance and characteristic mass data[16] measured in an AC magnetic field of 0.9 T

Element	Wavelength (nm)	Absorbance of roll-over	Characteristic mass[a] (pg)
Ag	328.1	>2	1.2
Al	309.3	0.8	10
As	193.7	1.3	17
Cd	228.8	0.8	0.4
Cr	357.9	1.6	1.8
Cu	324.8	0.7	3.0
Fe	248.3	0.8	2.8
Mn	279.5	1.1	1.7
Pb	283.3	1.5	7.0
Sb	217.6	1.4	38
Se	196.0	1.4	19
Sn	286.3	>2	20
Zn	213.8	1.0	0.4

[a] Mass of analyte in picograms that produces an integrated absorbance signal of 0.0044 s.

concentrations, the signals are analytically useless but there is no double-valued behavior. In FAAS, a warning system for roll-over is used that is based on the detection of the short periods of high absorbance at the start of nebulization and at the end where the atom density is increasing beyond the value that corresponds with the roll-over concentration. Detection of these peaks alerts the analyst that sample dilution is necessary to obtain the correct concentration values.

5 PULSED LAMP BACKGROUND CORRECTION

5.1 Principle

Ling[18] first advanced the idea of background correction based on the subtraction of absorbances measured alternately with a normal lamp emission line and a broadened lamp emission profile showing a dip in the center. He described a simple method to correct for nonatomic absorption in a portable mercury photometer that included two mercury vapor lamps emitting the mercury resonance line broadened to different extents. Smith and Hieftje[19] proposed a method for background correction in AAS based on the broadening which occurs in an HCL spectral line when the lamp is operated at very high currents. Under such conditions, the absorbance measured for a narrow atomic line is low, whereas the apparent absorbance caused by a broad-band background contributor remains the same as when the lamp is operated at conventional current levels. Background correction can therefore be effected by taking the difference in absorbances measured with the lamp operated at high and low currents. The so-called Smith–Hieftje background correction system applies its correction very near the atomic line of interest.

5.2 Instrumentation

The HCL power supply and drive circuitry[19] generate a 9-ms low-current pulse of 5–10 mA, followed by a 0.3-ms high-current pulse of 200–300 mA, as presented in Figure 10(a). The atom cloud in front of the hollow cathode dissipates rapidly for some elements whereas other elements (e.g. Ag, Cd, Cu, Pb) exhibit a persistence of the vapor cloud for more than 6 ms. The total pulse cycle repetition time of 50 ms allows the atomic cloud generated in the HCL during the high-current mode to clear before measurement is made again in the low-current mode for a broad range of elements. Figure 10(b) presents the conventional lamp emission line profile observed at low lamp current and the broadened and self-reversed lamp emission line during the high-current operation of the lamp. It is clear that the

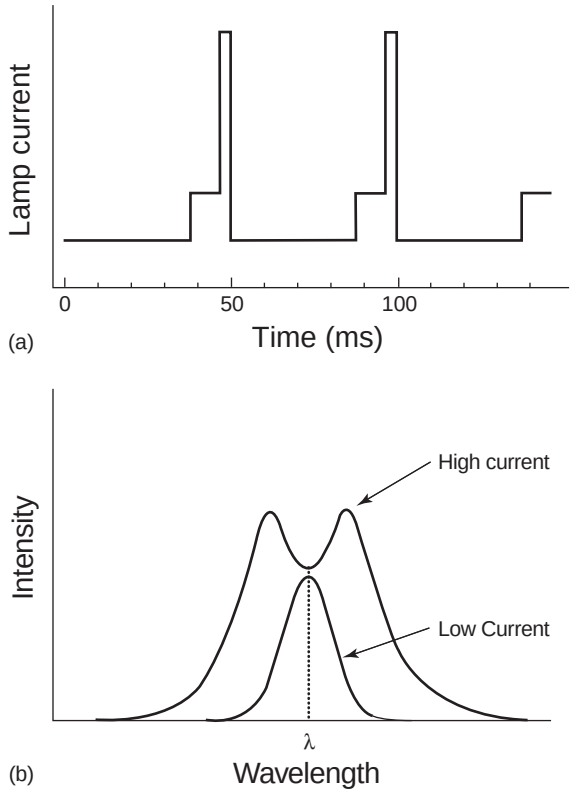

Figure 10 Pulsed HCL background correction. (a) Modulation of the lamp current that drives the HCL; (b) corresponding lamp emission line profiles generated under low- and high-current operation.

narrow line profile emitted under low-current operation is affected equally by atomic absorption or a broad-band spectral feature. In contrast, the broadened and self-reversed profile obtained at high current is affected by a broad-band absorber or scatterer but there is not so much overlap with the atomic absorption line profile. Absorbances calculated under low- and high-current operation are subtracted to yield a difference value which is free of any broad-band contribution.

5.3 Applications

The pulsed lamp background correction method is equally applicable to FAAS, hydride generation AAS and graphite furnace AAS. For effective operation it is necessary to broaden appreciably the emission line of each element under investigation. However, the degree of broadening is different for each element and also depends upon the peak current at which the HCL is driven. To achieve the desired degree of line broadening and line reversal while maintaining HCL reliability and acceptable lifetime, a special design of the lamp modulation circuitry and optimum high lamp current adjustment for each element are required. The slope

of the analytical curves is decreased to some extent when pulsed HCL background correction is applied. The measurement at high lamp current is intended to measure the background. Unfortunately, the broadening of the lamp emission line profile is limited and self-reversal is not complete. The remaining radiation in the central part of the lamp emission line profile is therefore absorbed by analyte atoms and subtracted from the gross absorbance together with the background. For 30 elements most commonly determined by AAS the loss of sensitivity is 47%, ranging from 13% for Cd to 84% for Hg. Similarly to the Zeeman background correction system, the ETAAS signals show a dip at very high analyte concentrations when pulsed HCL background correction is applied. The ability of the pulsed HCL background correction system to overcome background interferences in real samples has been demonstrated. Like the Zeeman approach, the pulsed lamp background correction method requires only a single primary source of radiation and single beam optics so that alignment is simplified.

6 WAVELENGTH-MODULATION CORRECTION METHODS

6.1 Principle

The broad-band spectral distribution of a continuum-source conveniently permits off-line background measurements to be performed in AAS similar to the procedures employed traditionally in atomic emission spectroscopy. Background with a flat spectral distribution requires simply a two-point background correction procedure in which the background absorption at the analytical absorption line is accounted for by a second measurement with the monochromator tuned to an arbitrary off-line wavelength position. The difference between the two measurements provides the net analytical signal. For an absorption line superimposed on a linearly sloping background, two background measurements, one on either side of the absorption line, would be required. The background absorption at the analytical absorption line in this case would be equivalent to the interpolated value of the two off-line measurements. Clearly, with increasing complexity of the background in the immediate vicinity of the analyte absorption line, manual off-line correction procedures would become exceedingly tedious and time-consuming. The use of wavelength-modulation in AAS in conjunction with a continuum primary source of radiation was first described by Snelleman.[20]

6.2 Instrumentation

A vibrating mirror, mounted on the axis of a milliammeter, is positioned in the optical path between the focusing mirror and the exit slit and driven by an AC current which vibrates the mirror at a fixed frequency. The analytical absorption line, when scanned across the exit slit, generates an AC component in the photosignal with a frequency twice that of the mirror vibration frequency. Phase-sensitive signal detection discriminates against DC and first-harmonic AC components of the photosignal due to flat and linearly sloping background distributions, respectively, over the narrow wavelength interval.

Zander et al.[21] combined a high-intensity continuum-source and a wavelength-modulated high-resolution échelle monochromator. A quartz refractor plate modulator was mounted vertically in front of the exit slit. Sinusoidal modulation of the quartz refractor plate scanned the image of the entrance slit rapidly back and forth across the exit slit. In later studies,[22] the refractor plate was positioned immediately behind the entrance slit in order to wavelength modulate the incident nondispersed radiation, permitting wavelength modulation of multiple predetermined wavelengths in the use of a multichannel échelle polychromator. Figure 11 shows a block diagram of a single-channel continuum-source AAS instrument based on a wavelength-modulated high-resolution échelle spectrometer. In this system the exit slit is replaced by a multichannel exit slit unit consisting of 20 prealigned exit slits for different atomic lines and a photomultiplier is positioned behind each slit. Later, more sophisticated detection systems as a linear photodiode array and a segmented array charge-coupled device detector[23] have been used in combination with the échelle optics.

6.3 Applications in Continuum-source Atomic Absorption Spectroscopy

Simultaneous multielement AAS based on wavelength-modulation background correction provides the basic raw data to cover the wide range of concentrations likely to be encountered in different samples because the entire absorption profile can be observed. The intensities at several points across the modulation interval are measured. Using the points at the extreme ends of the modulation interval as I^0 and those near the center and along the sides of the line as I, several absorbances with different sensitivities are calculated. This set of absorbances is measured and stored for every sample, blank and standard solution. From the set of absorbances measured for a series of standard solutions, a set of calibration graphs with different slopes is obtained. The set of absorbances measured for each sample and blank are converted to concentrations by reference to the corresponding calibration graphs. This procedure yields a set of concentration estimates for each element in each sample. The final concentration estimate is a weighted average based

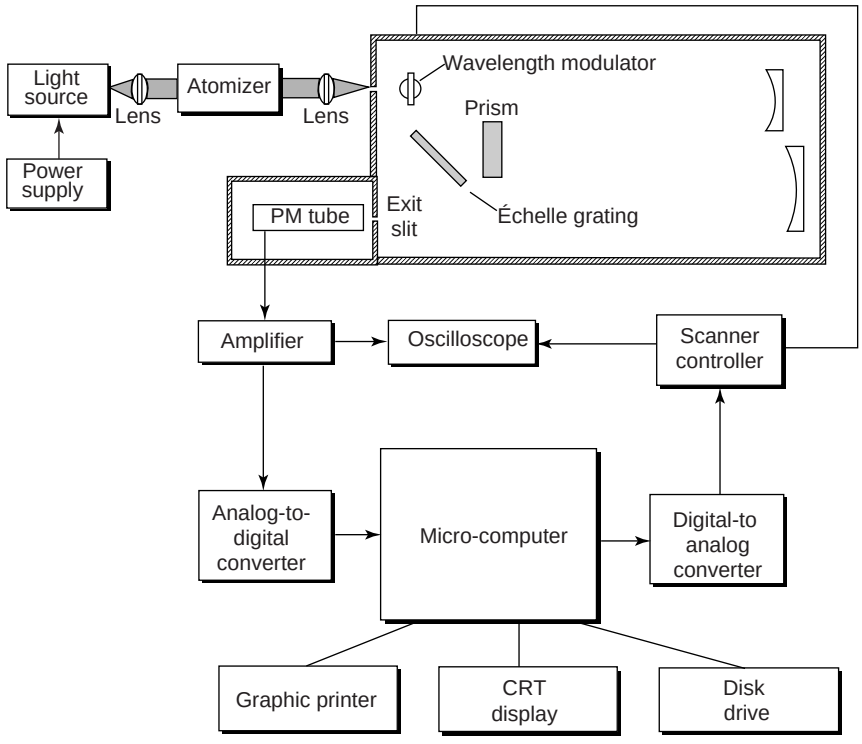

Figure 11 Block diagram of a single-channel continuum-source AAS spectrometer based on a wavelength-modulated, high-resolution échelle spectrometer. (Reprinted from T.C. O'Haver, J.D. Messman, *Prog. Anal. Spectrosc.*, **9**, 483–503 (1986), Copyright 1986, with permission from Elsevier Science.)

on the concentration S/N of each individual estimate. The line-center absorbance is utilized at low concentrations and, as the concentration increases, the lower sensitivity absorbances are progressively utilized. The individual calibration graphs are not linear over the entire range. Therefore, a certain number of standard solutions are required to characterize each calibration curve adequately. The wavelength-modulation background correction approach shares with the Zeeman and pulsed lamp methods the advantage of using only a single light source, so that lamp alignment is not so much of a problem as in the continuum-source method, and of making the background measurement at wavelengths very close to the analytical wavelength. Consequently, errors due to background structure and matrix atomic absorption are less likely to occur. The background correction performance of the continuum-source wavelength-modulation method has been evaluated experimentally for a variety of samples.

7 COMPARISON OF METHODS

7.1 Sensitivity and Dynamic Range

From the principle of continuum-source background correction (section 3.1), it is clear that the conventional AAS sensitivity is not influenced by the background correction system. The additional measurement with the deuterium lamp provides the background absorbance and the analyte atoms do not contribute significantly to the background signal because of their narrow absorption line profile.

In all Zeeman background correction systems, the measurement that is intended to provide the background absorbance is slightly influenced by the absorption of the analyte atoms owing to the incomplete shift of the Zeeman σ-components away from the lamp emission line profile. The contribution of the remaining σ-components decreases the net analyte absorbance by about 10% for most elements and by up to 45% for a few elements (Cu, Be, Bi). A further decrease in sensitivity is observed in Zeeman background correction systems based on a constant magnetic field from a permanent magnet or a DC magnet for all elements that show anomalous Zeeman splitting patterns. The sensitivity reduction is more severe (about 50%) when a central π-component is missing. The best Zeeman AAS sensitivity for all elements is obtained in a strong AC magnetic field. The best result in a constant magnetic field of optimum strength is about half the conventional AAS sensitivity, except for those elements that show normal Zeeman splitting patterns.

Pulsed lamp background correction provides sensitivity close to conventional AAS for a few elements only. It is hard to achieve sufficient broadening and self-reversal in the lamp emission line profile during the high-current pulse. Similarly to Zeeman AAS, the contribution of the remaining part of the lamp line profile that is insufficiently shifted is subtracted from the gross absorbance together with the estimated background. The loss of sensitivity varies from 13% for Cd to 84% for Hg.

In wavelength-modulation background correction, analytical curves of different sensitivity are obtained simultaneously. The sensitivity can be selected between the normal value for conventional AAS and much lower values when the analytical information comes from the wings of the line profile. The various calibration graphs become available simultaneously. In comparison with atomic emission spectroscopy, the dynamic range of AAS is limited. Starting from the detection limit, absorbances can be measured over about three orders of magnitude, i.e. from 0.001 up to 1 absorbance unit. This range is not influenced by the continuum-source background correction system. In Zeeman and pulsed lamp background correction systems the dynamic range is further limited owing to the roll-over phenomenon. The analytical curves in peak absorbance level off at somewhat lower absorbances than in conventional AAS whereas the analytical curves in integrated absorbance continue to increase, slowly, up to very high analyte concentrations. In contrast, the wavelength-modulation background correction system offers various analytical curves of different slopes and using them all together the dynamic range is significantly increased in comparison with conventional AAS.

7.2 Frequency of the Background Correction System

All background correction systems involve sequential measurements of the gross absorbance and the background according to Equation (1). In the continuum-source background correction system this requires switching, at a certain frequency, between the HCL and the continuum source. The frequency should be high enough to obtain a good estimate of the background at the time of the measurement of the gross absorbance. In ETAAS, the atomic and background absorption signals may rise to a maximum in times ranging from 0.1 s or less to several seconds, depending on the furnace design, the heating rate and the composition of the sample. The beam switching frequency should be high enough to enable the fastest rates of change of the atomic and background absorption signals to be followed. The upper limit to the chopping frequency follows from practical considerations whereas HCLs can be pulsed at very high frequencies. A high enough background correction frequency of 200–400 Hz can easily be achieved.

In Zeeman background correction systems the frequency is related to the modulation frequency of the AC magnetic field or the frequency of rotating the polarizer in the case of a constant magnetic field. In most of the instruments the background correction frequency is 50–100 Hz. It is complicated to rotate a polarizer at high speed and it is expensive to modulate an AC magnetic field at a frequency higher than the mains frequency. The pulsed lamp background correction system also is not very fast (20 Hz) because the lamp needs time to recover after the high lamp current pulse. In all background correction approaches, an increase in the frequency decreases the time available to measure the signals and consequently decreases the S/N.

Figure 12 shows that for a short interval during the atomization step in ETAAS, each background correction system makes repeated, rapid, sequential measurements of the gross sample absorbance and background absorbance.[24] These measurements are usually spaced evenly in Zeeman background correction and in wavelength-modulation background correction whereas an uneven spacing of the measurements of gross sample absorbance and the background is found in pulsed lamp background correction systems. For accurate

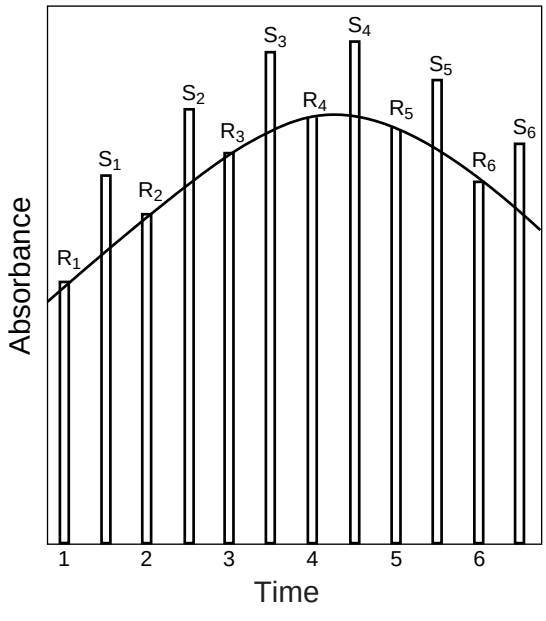

Figure 12 Illustration of repetitive, sequential data aquisition of the gross absorbances of the sample (S_1–S_6) and background absorbances (R_1–R_6) as a function of time for a short interval during the atomization of a sample in ETAAS with analyte and background present. (Reproduced by permission of The Royal Society of Chemistry from J.M. Harnly, J.A. Holcombe, *J. Anal. At. Spectrom.*, **2**, 105–113 (1987).)

background corrections, the background absorbance must be known at the same time as the gross sample absorbance. Ideally, a mathematical function with the same shape as the background absorbance function would be fit to the background measurements and the background absorbance could be predicted at any time. Various interpolation procedures are used in ETAAS instruments to avoid or reduce background correction errors when the background is changing rapidly.

7.3 Background Correction Performance

The accuracy of the correction for background absorbance must be perfect up to high levels of background absorption. In the absence of analyte, the baseline should be a perfectly straight line even at high background absorbance levels. Three conditions should be fulfilled to accomplish this. In the previous section we have seen that the transient ETAAS signal requires a good approximation of the background at the time of measurement of the gross absorbance signal. In addition, the background must be measured at the same position in the atomizer. This is easily achieved if only one source of radiation is used as in the Zeeman background correction system, the pulsed lamp background correction system and the wavelength-modulation system. In continuum-source background correction, lamp alignment is always difficult because of the different positions of the lamps in the instrument, the different optical paths, the different geometries of the beams and the different intensity distributions. Since the atomic vapor in both the flame and the graphite furnace is not truly homogeneous, nor is the radiant cross-section of either source, perfect matching is difficult. Background absorbance in FAAS is usually low and it can be corrected easily with the continuum-source background correction system.

In ETAAS, the continuum-source background correction system is often also successful. Background correction problems have been observed when the background is higher than about 0.5–1 absorbance unit, often owing to misalignment of the radiation beams. Many instruments are fitted with both a deuterium lamp for the UV and a tungsten lamp for the visible wavelength region. The intensity of the deuterium lamp is quite low when a low spectral band-pass is selected, which may be a limiting factor for the HCL intensity also because most instruments require more or less equal intensities of the two radiation beams.

Continuum-source background correction systems are not able to correct for structured background. The background is measured as an average background absorbance level over the spectral band-pass of the monochromator, i.e. within 0.2–2 nm. Numerous over- and undercorrections have been reported. Several of them are well known and can be found in text books on AAS, others are only observed under very special experimental conditions. Zeeman-, pulsed lamp- and wavelength-modulation-based background correction systems measure the background at exactly the wavelength position of the analyte absorption line or very close to line, i.e. within an interval of 10 pm. Although perfect background correction is not guaranteed, there is a better chance that the right background absorbance is measured at a wavelength position close to the absorption line in comparison with the average value over the full spectral band-pass. The Zeeman background correction system that applies the magnetic field to the atomizer is the only system that provides correction for the background exactly at the wavelength of the analyte absorption line. Even here, perfect background correction is not guaranteed because, in a very few cases, the background may change with the strength of the magnetic field in AC Zeeman systems. Most of the background correction errors reported for continuum-source background correction are not found when Zeeman, pulsed lamp or wavelength-modulation background correction is applied.

8 BACKGROUND CORRECTION PROBLEMS AND HOW TO AVOID THEM

The most frequent spectral interference in continuum-source background correction results from atomic absorption of the background radiation by massive amounts of metals in the matrix, usually at secondary absorption lines of these elements. However, some of the structured absorption errors result from molecular absorption in the vapor phase. Probably the first identified example of background correction error was the influence of large amounts of Fe in the determination of Se at the 196-nm line. A list of potential interferences of this type[17] shows errors that may occur, based on the wavelengths of the various lines. Some of the problems are very unlikely to be serious in FAAS. In the graphite furnace the amount of matrix is typically much larger relative to the analyte and therefore interferences can be more severe. However, during method development much time and effort are usually spent in obtaining an acceptable program for heating of the graphite furnace and the application of chemical modifiers is also considered. Owing to the improved graphite furnace design, the temperature program used and the application of chemical modifiers, potential interferences and background correction errors are not always found in real analysis.

Background correction errors are hard to identify when the background absorbance takes place simultaneously

with the atomic absorption. If there is some time delay between the two processes, a positive or negative disruption of the baseline may be observed just before the atomic absorption starts or just after the atomic absorption is finished. It should be noted that the method of standard additions does not control background correction errors.

Several problems of background overcorrection have been noted that have not yet been fully explained whereas others are well understood. For example, the effect of phosphate on Se and As is related to molecular P_2 that is generated in the vapor phase of the furnace and the molecular band absorbs the continuum radiation.

With Zeeman background correction, the pulsed lamp background correction system or wavelength-modulation, most of the interferences reported for continuum-source background correction are not observed. Nevertheless, a few interferences have been reported.[16] In most of the commercial Zeeman instruments, background correction is achieved by using an AC magnetic field transverse to the optical axis. If matrix absorption lines or molecular bands that are very close to the analyte wavelength exhibit Zeeman splitting, correction errors may occur. A matrix absorption line that does not overlap with the lamp emission line at zero-field may do so at maximum field strength, or the other way around. An example of such problems is the PO molecular absorption band that causes an overcorrection for lead in the presence of phosphate. The overcorrection error is influenced by the strength of the magnetic field. It is diminished by decreasing the amount of phosphate or by replacing the standard transverse-heated graphite atomizer with an end-capped tube. Accurate results for the determination of Pb in bone show that such background correction errors can be avoided by optimization of the experimental conditions during method development.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AC	Alternating Current
DC	Direct Current
ETAAS	Electrothermal Atomization Atomic Absorption Spectroscopy
FAAS	Flame Atomic Absorption Spectroscopy
HCL	Hollow Cathode Lamp
S/N	Signal-to-noise Ratio
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry

REFERENCES

1. H. Massmann, S. Güçer, 'Physikalische und chemische Prozesse bei der Atomabsorptionsanalyse mit Grafitrohröfen – I. Messbedingungen, Versuchsanordnung und Temperaturmessungen', *Spectrochim. Acta, Part B*, **29**(11/12), 283–300 (1974).
2. R.A. Newstead, W.J. Price, P.J. Whiteside, 'Background Correction in Atomic Absorption Analysis', *Prog. Anal. At. Spectrosc.*, **1**, 267–298 (1978).
3. A.K. Gilmutdinov, T.M. Abdullina, S.F. Gorbachev, V.L. Makarov, 'Concentration Curves in Atomic Absorption Spectroscopy', *Spectrochim. Acta, Part B*, **47**(9), 1075–1095 (1992).
4. J.B. Willis, in *Methods of Biochemical Analysis*, ed. D. Flick, Wiley-Interscience, New York, 1–67, Vol. 11, 1963.
5. S.R. Koirtyohann, E.E. Picket, 'Background Corrections in Long Path Atomic Absorption Spectrometry', *Anal. Chem.*, **37**, 601–603 (1965).
6. H. Massmann, Z. El Gohary, 'Analysestörungen durch strukturierten Untergrund in der Atomabsorptionsspektrometrie', *Spectrochim. Acta, Part B*, **31**(7), 399–409 (1976).
7. D.C. Manning, 'Spectral Interferences in Graphite Furnace Atomic Absorption Spectroscopy. I. The Determination of Selenium in an Iron Matrix', *At. Absorpt. Newsl.*, **17**(5), 107–108 (1978).
8. P. Zeeman, 'On the Influence of Magnetism on the Nature of the Light Emitted by a Substance', *Philos. Mag.*, **43**(5), 226–237 (1897).
9. H. Koizumi, K. Yasuda, 'A Novel Method for Atomic Absorption Spectroscopy Based on the Analyte–Zeeman Effect', *Spectrochim. Acta, Part B*, **31**(10–12), 523–535 (1976).

10. F.J. Fernandez, W. Bohler, M.M. Beaty, W.B. Barnett, 'Correction for High Background Levels Using the Zeeman Effect', *At. Spectrosc.*, **2**(3), 73–80 (1981).
11. M. Prügger, R. Torge, 'Vorrichtung zur Spektrochemischen Feststellung der Konzentration eines Elementes in einer Probe', *Ger. Pat.*, 1964469, filed 23 December, 1969.
12. H. Koizumi, K. Yasuda, 'New Zeeman Method for Atomic Absorption Spectrophotometry', *Anal. Chem.*, **47**, 1679–1682 (1975).
13. P.R. Liddell, K.G. Brodie, 'Application of a Modulated Magnetic Field to a Graphite Furnace in Zeeman Effect Atomic Absorption Spectrometry', *Anal. Chem.*, **52**, 1256–1260 (1980).
14. M.T.C. de Loos-Vollebregt, L. de Galan, J.W.M. van Uffelen, 'Longitudinal a.c. Zeeman AAS with a Transverse Heated Graphite Furnace', *Spectrochim. Acta, Part B*, **43**(9–11), 1147–1156 (1988).
15. I. Shuttler, 'The Application of a Transversely Heated Electrothermal Atomizer with Longitudinal Zeeman-effect Background Correction to the Determination of Vanadium in Urine', *At. Spectrosc.*, **13**(5), 174–177 (1992).
16. M.T.C. de Loos-Vollebregt, L. de Galan, 'Zeeman Atomic Absorption Spectrometry', *Prog. Anal. At. Spectrosc.*, **8**, 47–81 (1985).
17. W. Slavin, G.R. Carnrick, 'Background Correction in Atomic Absorption Spectrometry', *CRC Crit. Rev. Anal. Chem.*, **19**(2), 95–134 (1988).
18. C. Ling, 'Sensitive Simple Mercury Photometer Using Mercury Resonance Lamp as a Monochromatic Source', *Anal.Chem.*, **39**, 798 (1967).
19. S.B. Smith, Jr, G.M. Hieftje, 'New Background-correction Method for Atomic Absorption Spectrometry', *Appl. Spectrosc.*, **37**, 419–424 (1983).
20. W. Snelleman, 'An a.c. Scanning Method with Increased Sensitivity in Atomic Absorption Analysis Using a Continuum Primary Source', *Spectrochim. Acta, Part B*, **23**(6), 403–411 (1968).
21. A.T. Zander, T.C. O'Haver, P.N. Keliher, 'Continuum-source Atomic Absorption Spectrometry with High-Resolution and Wavelength-Modulation', *Anal. Chem.*, **48**, 1166–1175 (1976).
22. T.C. O'Haver, J.D. Messman, 'Continuum-source Atomic Absorption Spectrometry', *Prog. Anal. Spectrosc.*, **9**, 483–503 (1986).
23. J.M. Harnly, C.M.M. Smith, D.N. Wichems, J.C. Ivaldi, P.L. Lundberg, B. Radziuk, 'Use of a Segmented Array Charge Coupled Device Detector for Continuum-source Atomic Absorption Spectrometry with Graphite Furnace Atomization', *J. Anal. At. Spectrom.*, **12**, 617–627 (1997).
24. J.M. Harnly, J.A. Holcombe, 'Mathematical Correction of Systematic Temporal Background-correction Errors for Graphite Furnace Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **2**, 105–113 (1987).

Flame and Vapor Generation Atomic Absorption Spectrometry

M.D. Amos
Emerald, Vic., Australia

1 Introduction **2**
1.1 The Origins of Atomic Spectra 2
1.2 First Analytical Applications of Atomic Spectra 2
1.3 Suggestions for the Analytical Use of Atomic Absorption Spectra 2
2 Theoretical Basis of Atomic Absorption Spectrometry **3**
2.1 General Comments 3
2.2 Atomization and Excitation 3
2.3 Absorption and Emission Lines 4
2.4 The Shape of Spectral Lines 4
2.5 Measuring Atomic Absorption 5
2.6 Spectral Interferences in Atomic Absorption Spectrometry 6
3 Instrumentation for Atomic Absorption Spectrometry **8**
3.1 Light Sources 9
3.2 Atomizers for Atomic Absorption Spectrometry 11
3.3 Wavelength Selection 11
3.4 Detectors 11
3.5 Amplifier and Read-out 11
4 Atomization Systems and Their Use **12**
4.1 Historical 12
4.2 Flame Atomization 12
4.3 Vapor Generation Techniques 15
4.4 Graphite Furnace Atomization 16
5 Optical Design of Instruments for Atomic Absorption Spectrometry **16**
5.1 Atomic Absorption Spectrometer Design 16
5.2 Optical Design for Background Correction 17
6 Practices of Atomic Absorption Spectrometry **19**
6.1 General 19
6.2 Optimization of Instrumental Parameters 19
7 Interferences in Atomic Absorption Spectrometry **20**
7.1 Spectral Interferences 20
7.2 "Chemical" Interferences 20
7.3 Ionization Interferences 20
7.4 "Physical" Interferences 21
8 Validation of Results Obtained by Atomic Absorption Spectrometry **21**
9 Automation in Analysis Using Atomic Absorption Spectrometry **21**
10 Comparison with Other Analytical Techniques **22**
10.1 Flame Emission 22
10.2 Inductively Coupled Plasma Atomic Emission Spectrometry 22
10.3 Inductively Coupled Plasma Mass Spectrometry 22
10.4 Choice of Methods 22
11 Future Directions **22**
Abbreviations and Acronyms **22**
Related Articles **23**
Note **23**
References **23**

Atomic absorption spectrometry (AAS) is a means of specific analysis for most of the metallic and metalloid elements. It achieves its specificity from the highly selective absorption, by free atoms of the analyte, of atomic emission of the same element at characteristic wavelengths. In its simplest form of application the free atoms of the analyte are created by aspiration of a solution of the sample to be analysed into a suitable flame, flame atomic absorption spectrometry (FAAS). The other principal means employed for the production of atoms (atomization) are graphite furnace atomization (GFAAS) and vapor generation atomization (VGAAS). Sample volumes of a few milliliters are normally needed for FAAS and VGAAS, but smaller volumes are needed for GFAAS. For most of the approximately 65 elements amenable to measurement by AAS, detection limits are from parts per billion to parts per million by flame methods and from parts per trillion to parts per billion levels by GFAAS and VGAAS. FAAS is inherently simple and rapid to use, whereas GFAAS and VGAAS are less rapid. The instrumentation used for AAS is relatively inexpensive Since all of these methods generally require the sample to be in solution form, the need for dissolution may be seen as a limitation in some cases. In the most common instrumentation employed, elements are only measured one at a time. It follows that for the determination of many elements in a large group of samples the method must be seen

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

to be slow by comparison with multielement techniques such as inductively coupled plasma mass spectrometry (ICPMS) and inductively coupled plasma atomic emission spectrometry (ICPAES). However, the equipment costs involved for these techniques are generally higher and they are more complex to employ.

1 INTRODUCTION

1.1 The Origins of Atomic Spectra

Although astronomical observations led to the discovery of the 'Fraunhofer (atomic absorption) lines' in the solar spectrum, it was not until about 1850 that the origin of these dark lines in an otherwise continuous bright spectrum started to be properly deduced. It appears that several workers, among them Foucault, Balfour Stewart and Stokes started to realize that these lines were the result of absorption of the continuous emission from the sun and other astronomical bodies by the atomic vapor of some elements occurring in their atmospheres. It was not until 1860 that Bunsen and Kirchhoff first comprehensively demonstrated experimentally, and then elucidated, the origin of such atomic emission and absorption spectra and drew attention to their interrelationship.

In what are now seen as classical experiments they demonstrated that atomic vapors of sodium and lithium, produced in a flame, would each not only emit radiation at specific wavelengths but would also absorb radiation at precisely the same wavelengths.[1,2] Kirchhoff's original paper[1] is given only for completeness but the paper, and of course the title, are in German. However an English translation of this paper is also available.[2] From these results Kirchhoff elucidated that atomic transitions between different energy levels might be observed either using absorption or emission.

He found that if, for example, a solution of a sodium salt was heated in a flame it would then spontaneously emit or absorb radiation at wavelengths characteristic *for sodium only*, the famous 'D lines'. Likewise a solution of a lithium salt would also, under the same circumstances, emit or absorb radiation at a wavelength different from those of sodium but equally characteristic *for lithium only*. By doing this he essentially drew attention to the potential usefulness of such observations to identify the presence or absence of sodium and lithium (qualitative analysis). This work was later extended, by Bunsen and Kirchhoff, to the qualitative determination of the presence of other elements (potassium, calcium, strontium and barium).[3]

As an interesting example of the powerful usefulness of atomic spectra, these workers also first discovered the existence of the elements rubidium and cesium by observing the presence of other characteristic (and unexpected) lines when employing solutions of natural salt for their experiments.[4] The salt they used was not produced from the evaporation of seawater but rather was natural salt from underground deposits.

(Note that the Kirchhoff articles[3,4] are also translations of articles originally published in *Annalen der Physik und Chemie*.)

Bunsen and Kirchhoff only employed these observations in a qualitative mode and seem to have drawn no attention to the possible quantitative use of the phenomena of atomic emission and absorption.

This does not seem too surprising given that in 1860 means of quantitative measurement of emission or absorption signals were not readily available.

1.2 First Analytical Applications of Atomic Spectra

Later, starting with the further development of the photographic process, the means of recording intensities for at least semiquantitative purposes was to lead to the development of emission spectrography using (initially) electric arcs as the means of excitation.

Flames were also seen to be a potential means of excitation for observing emission and by 1928 Lundegardh developed suitable burners (using acetylene as the fuel gas and air as the support gas) and methods for sample introduction.[5] From this time on, and for the period to the 1950s, developments continued in the above methods but the use of methods utilizing atomic absorption spectra were almost totally ignored. As will be described in section 2, this was in major part because the means of measuring atomic absorption with good sensitivity was not yet properly developed.

1.3 Suggestions for the Analytical Use of Atomic Absorption Spectra

After a significant period of research into atomic emission spectrometry (AES), Walsh became convinced that this neglect of the use of atomic absorption had gone on for too long. He recognized that there would be potential advantages in using the measurement of atomic absorption instead of atomic emission as a means of measuring the concentration of elements in samples.

The major advantages he postulated for the use of atomic absorption methods were as follows:

- Since ground-state atoms would be used for the measurement, the method should be independent of the excitation potential of the element to be measured.
- The method should be much less influenced by the temperature of the atomizing device used.

- There should be much less likelihood of spectral interferences impairing the accuracy of the method.

In 1955, Walsh published a landmark paper explaining, in major detail, these theories Additionally, however, he also defined a practical approach to the necessary key elements required for instrumentation to measure atomic absorption.[6]

Completely independently, Alkemade and Milatz also published an article suggesting the use of atomic absorption spectra.[7] They, however, did not develop the practical means of performing this measurement to the same extent as Walsh and this limited the importance of their paper in a practical sense.

In the following section, the theoretical basis of atomic absorption and its measurement will be dealt with in more detail.

2 THEORETICAL BASIS OF ATOMIC ABSORPTION SPECTROMETRY

2.1 General Comments

At this juncture, any comparisons made will be solely between flame emission and flame atomic absorption and no further reference will be made to emission measurements made following excitation by electric arc or spark or by inductively coupled plasma. These are all established methods in widespread use, although arc and spark spectrometry is currently in much less use than in previous years.

In order to measure atomic absorption, it is first necessary to produce free atoms of the element to be measured. However, it should be realized that it is also necessary to produce free atoms of any element to be measured by atomic emission. Atoms cannot be excited unless they are first converted to free atoms from any compound associations.

2.2 Atomization and Excitation

In most atomic absorption, or flame emission, methods atoms are produced from a sample in solution form by the application of heat. If a solution sample is introduced, for instance, into a flame then atomization will often occur. Depending on the excitation potential of the element in question, some of these atoms produced may also then be excited. It is this atomization, and subsequent excitation, which forms the basis of flame emission spectrometry. However, as Walsh pointed out, there are some important factors which should be considered. These are best expressed in some figures taken from a table in his first paper, shown here in Table 1.

Table 1 Values of N_j/N_0 for various resonance lines[a]

Element	Wavelength of resonance line (nm)	N_j/N_0	
		2000 K	3000 K
Cs	852.1	4×10^{-4}	7×10^{-3}
Na	589.0	1×10^{-5}	6×10^{-4}
Ca	422.7	1×10^{-7}	4×10^{-5}
Zn	213.9	7×10^{-15}	6×10^{-10}

[a] N_j is the number of atoms in the excited state at any given time and N_0 is the number of atoms in the ground (unexcited) state at any given time, and consequently the ratio N_j/N_0 is the fraction of atoms in the excited state at any given time. The figures quoted are taken from a table which appears in Walsh.[6] They represent only some of the entries in that table and have been rounded off for convenience.

The figures in Table 1 show that

- the fraction of excited atoms at any time is always a very small fraction of the total atoms present;
- the fraction of excited atoms changes by a huge amount with any major temperature change;
- since the fraction of ground-state atoms is equal to one minus the fraction of excited atoms, this remains sensibly constant at very close to 1, regardless of temperature change.

Since atomic emission signals originate from excited atoms and atomic absorption signals from ground-state atoms, the consequences of the above are the following:

- Atomic emission signals will be dramatically affected by temperature changes but absorption signals will be affected to a much lesser extent.
- For many harder to excite elements (such as Zn in Table 1), atomic absorption will be much more sensitive than atomic emission at such temperatures.
- All excited states will, at any given time, have a very small population of atoms. As a result, sensibly only those transitions which originate in the ground state will be able to be measured by atomic absorption. It follows, then, that since there are only a very limited number of such transitions for each element, the atomic absorption spectrum of all elements will be very much simpler than the atomic emission spectrum of the same element. The major consequence of this is that the likelihood of spectral interference will be much lower for absorption measurements than for emission measurements.

Properly, the resonance line for any element is the line associated with the transition between the ground state and the first (lowest energy) excited state. There is therefore commonly only one true resonance line per element. In the years since atomic absorption analysis was

introduced, the convention has arisen that all lines which are associated with transitions involving the ground state are called resonance lines. This atomic absorption convention will generally be followed throughout this article.

2.3 Absorption and Emission Lines

A purely diagrammatic representation of the energy levels involved in the production of an emission and absorption line is shown in Figure 1.

The relationship between the difference of the two energy levels involved in the transition and the frequency of the radiation emitted or absorbed is given by Equation (1):

$$E_1 - E_0 = h\nu \tag{1}$$

where E_1 and E_0 are the two energy states, h is Planck's constant and ν is the frequency of the radiation emitted or absorbed. In spectroscopy in the ultraviolet and visible regions of the spectrum wavelengths are more commonly used rather than frequency, so Equation (1) may be rewritten as Equation (2):

$$E_1 - E_0 = \frac{hc}{\lambda} \tag{2}$$

where c is the velocity of light and λ is the wavelength of emission or absorption. Rearranging, we obtain Equation (3):

$$\lambda = \frac{hc}{E_1 - E_0} \tag{3}$$

Since h and c are both constants, it follows that the wavelength corresponding to the change of energy of the atom, from one energy level to another, is inversely proportional to the difference between the two energy levels. Since the same two energy levels are involved, it follows that the emission and the absorption wavelengths will be identical since the calculation is the same in both cases.

2.4 The Shape of Spectral Lines

The emission and absorption wavelengths are generally quoted in texts as being, for example, Cu 324.754 nm, and it might often be assumed that this means a line of virtually no shape, i.e. an infinitely narrow line. Such is not the case, however, and lines do have a finite width and shape, both of which will vary under certain conditions. The wavelength designated might be seen as the central wavelength, but there will be present under all conditions a distribution of wavelengths around that central wavelength (called the 'natural' line width). Various factors may increase the line width further ('broaden' the line).

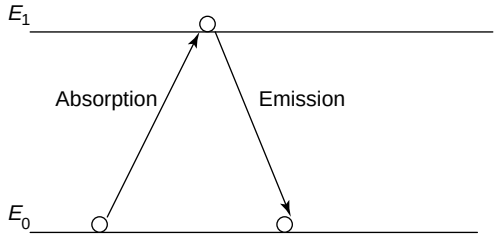

Figure 1 Relationship of emission and absorption lines.

In a short survey such as this article, only the two principle causes of line broadening will be considered. The first of these is Doppler broadening, caused by the fact that in a vapor cloud the atoms will be moving in all directions. Observations of emission from an atom coming towards the point of observation at the time of emission will record a wavelength shorter than that recorded from an atom going away from the point of observation at the time it emits. Obviously, as the temperature is increased the velocity of motion of the atoms is increased and this effect will therefore increase with increasing temperature, producing an increased line width.

The second cause is pressure broadening due to the concentration of atoms and molecules in the vapor in which the emission is occurring. The higher these concentrations are, the greater is the broadening which will occur.

The combined effects of these two (and some other minor) causes of broadening is to produce an emission line with an approximately Gaussian profile, as shown in Figure 2(a). The usual manner of describing the width of this line is the 'half-width', that is, the width at half the maximum peak intensity. The use of this manner of describing line width is necessary because the maximum width which exists, as zero intensity is approached, becomes progressively harder to measure and is dependent on the sensitivity of measurement available and the concentration of the atoms responsible for the absorption or emission.

Under the conditions prevailing in a flame, burning at atmospheric pressure at a temperature in the 2000–3000 K range, the half-width of the absorption lines will be of the order of 0.002–0.01 nm. An absorption line may be drawn, as shown in Figure 2(b), following the convention of drawing absorption lines downwards and emission lines upwards. For the understanding of this diagram, and for descriptions which will follow later, it is necessary to describe just what it illustrates, as far as atomic absorption measurement is concerned.

For any absorption spectrometric method, the absorption signal is always designated in terms of "absorbance" (A) and this is defined as $\log_{10}(I_0/I_t)$ where I_0 is the intensity of the original source radiation and

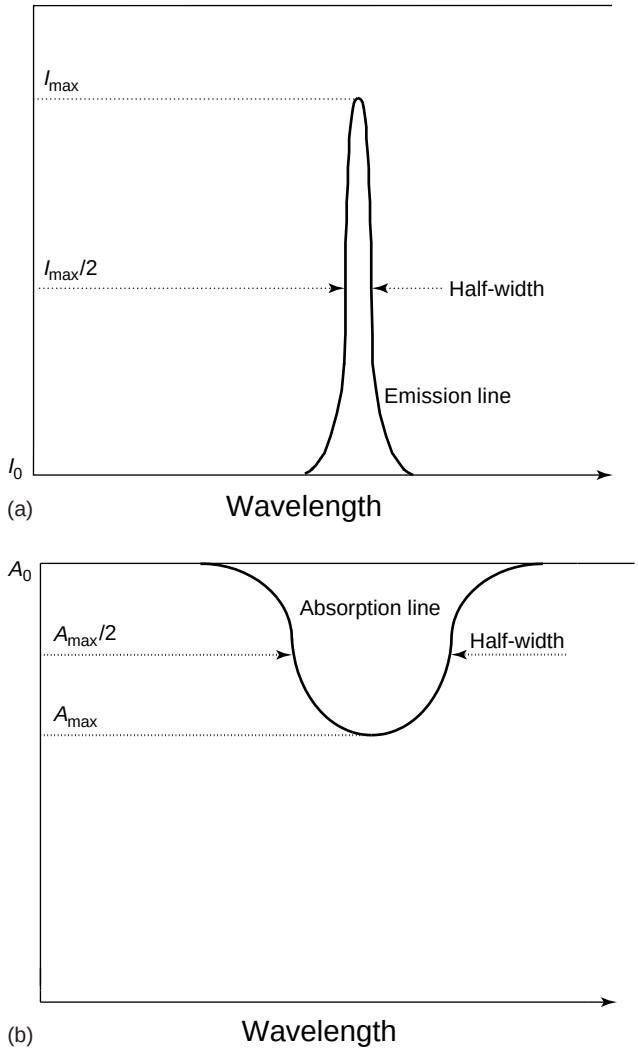

Figure 2 Representation of (a) an emission line and (b) an absorption line.

I_t the intensity of the same radiation after it has passed through the absorbing medium.

The absorbance (A) which will occur follows Beer's law, which may be expressed as Equation (4):

$$A = kcl \tag{4}$$

where k is the absorption coefficient, c the concentration of the absorbing species and l the absorption pathlength.

Returning to Figure 2(b), any point on the curve shown may be seen to represent the absorption coefficient at the particular wavelength. As a result, it may easily be understood that at the two wavelengths represented at the arrow points, where the half-width is defined, the absorption coefficient will be exactly half that at the peak of the absorption line, designated A_{max}. From the same reasoning, the absorption coefficient at the extreme left and right of the absorption line will be approaching zero.

2.5 Measuring Atomic Absorption

2.5.1 Using Continuum Source Emission

Early concepts of measuring atomic absorption were based on the idea of measuring the absorption of light from a continuum source, but this approach has limited value for the following reason. The spectral band-pass of a monochromator of modest resolution, say 0.2 nm, can be represented as shown in Figure 3 with a representation of an absorption line with a half-width of approximately 0.01 nm superimposed on it. It can be seen that the sensitivity of this measurement will be limited, since even if the peak absorption at the center of the absorption line is, for example, 40%, perhaps less than 1% of the total energy reaching the detector will be absorbed. In addition, it is obvious that changes of absorption with concentration of the atoms of the analyte will give rise to a nonlinear function.

Naturally, the sensitivity may be improved by using a higher resolution monochromator as represented, for example, as 0.05 nm in Figure 3. However, the absorption will still probably be less than 2–3% of the total radiation. Also, and equally important, with such a narrow slit opening, the total intensity of radiation from a typical continuum source would be very low and the precision of the measurement function would therefore be limited by this low intensity.

2.5.2 Using a Sharp Line Emission Source

One of Walsh's major contributions to the subject was to suggest that peak absorption could be measured if use was made of a line source, providing emission at precisely

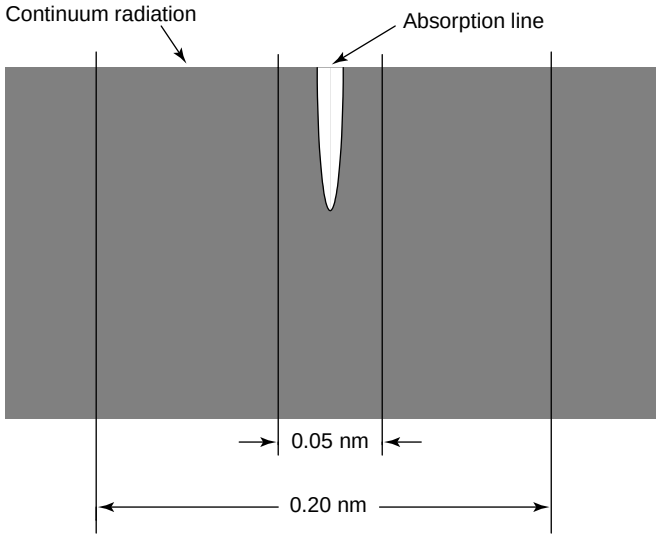

Figure 3 Measurement of atomic absorption using a continuum source.

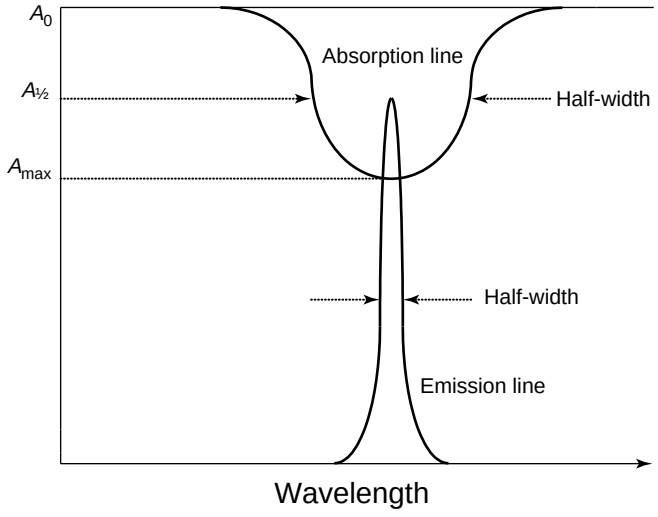

Figure 4 Measurement of atomic absorption using a sharp-line source.

the same wavelength as the absorption, and having a line width much less than that of the absorption line. This situation is represented in Figure 4.

The means of providing such a sharp line emission, at precisely the same wavelength as the absorption, was to use an individual hollow-cathode lamp (HCL) to provide the emission spectrum for each element to be determined. The construction and operation of such lamps will be discussed more in the next section, but at this stage it is useful to indicate why this sharp line emission occurs.

An HCL operates at reduced pressure, typically 1–2 kPa, so the concentration of atoms and molecules within the tube will be much less than in a flame at atmospheric pressure and this will reduce pressure broadening. Additionally, the temperature of the discharge will be only perhaps 500–600 K, compared with flame temperatures of 2000 K or more, and this will reduce Doppler broadening. Under these conditions, a HCL produces emission lines with half-widths of the order of 0.0004–0.002 nm.

As shown in Figure 4, it should be understood that the major part of the emission signal will be capable of absorption at or near the center of the absorption line, that is, at the wavelength where the absorption coefficient is at its maximum. This will result in maximum sensitivity of measurement. Quantitatively, the amount of energy at the maximum and minimum wavelengths of the emission line (the left- and right-hand flanks of the line as shown) will be very small. As a result, the fact that at these wavelengths the absorption coefficient will be slightly smaller will have only a small effect on the overall sensitivity achieved.

2.6 Spectral Interferences in Atomic Absorption Spectrometry

True spectral interferences may be defined as spectral interferences which, if undetected and uncontrolled, may lead to an error in the results obtained. There is another phenomenon, here termed ''pseudo-spectral interference'', which can affect measurements and this will be described first. It is not really a spectral interference since it only contributes to the quality of the measurement function and does not cause incorrect results. This minor problem will be dealt with first.

2.6.1 Pseudo-spectral Interferences

As illustrated in Figure 4, it may be seen that the emitted line from the light source used to make the measurement may be absorbed by the atomic absorption process, the amount of absorption being dependent on the concentration of the analyte present. If for whatever reason there is an emission produced by the light source which results in a line within the spectral band-pass of the monochromator, but which is not overlapped by the absorption line, as shown in Figure 5, no interference will be produced.

However, since the extraneous line will not be absorbed, regardless of how great the concentration of the analyte is, some loss of sensitivity of the measurement will occur and the calibration function will be nonlinear. Two means are available which may successfully overcome this problem:

- employ a narrower spectral band-pass from the monochromator by altering the slit opening;
- remove the element which gives rise to the extraneous line.

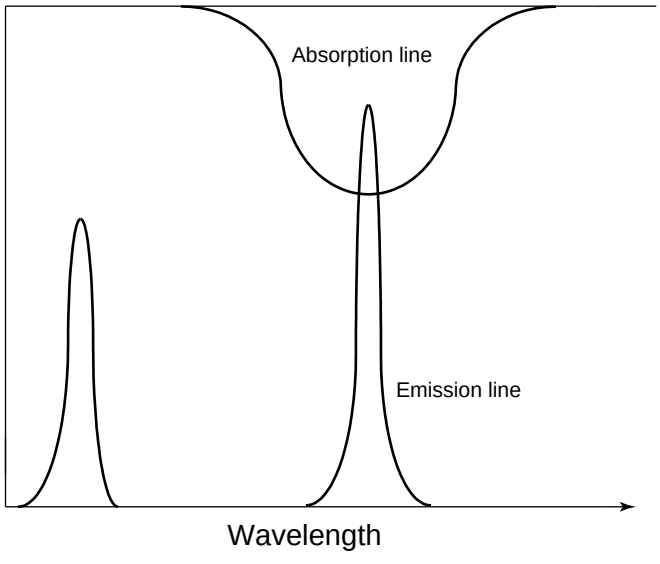

Figure 5 Atomic absorption measurement with a nearby emission line.

With regard to the first of these solutions, it should be recognized that in Figure 5 the spectral band-pass is roughly 0.04 nm (if we assume that the half-width of the absorption line is about 0.01 nm). To eliminate the effect of the "interfering" line the slit opening would need to be closed to a spectral band-pass of about 0.02 nm and this would require a relatively high resolution monochromator.

The second solution will be possible, for example, if the foreign line is due to the fill gas of the HCL by replacing it with a different fill gas. If the foreign line is due to an element other than the analyte, and which is introduced in the construction of the cathode, then the construction of the cathode might be modified to remove this element.

Should the foreign line be a nonresonance line of the analyte itself, then the second solution is not possible and only the first solution can be used either to remove or at least to minimize this curvature of calibration. It must be emphasized again, however, that curvature of calibration is only a minor operational problem and does not in itself cause errors in the analysis.

2.6.2 True Spectral Interferences

By comparison, again with Figure 4, if another element has an absorption line within the spectral band-pass of the monochromator, as illustrated in Figure 6, then no interference will occur. This is because there is no overlap of the other absorption line with the emission from the sharp line source at the resonance wavelength.

Should there be, however, a much closer absorption line, such as shown in Figure 7, then interference is possible since the left-hand flank of the absorption line

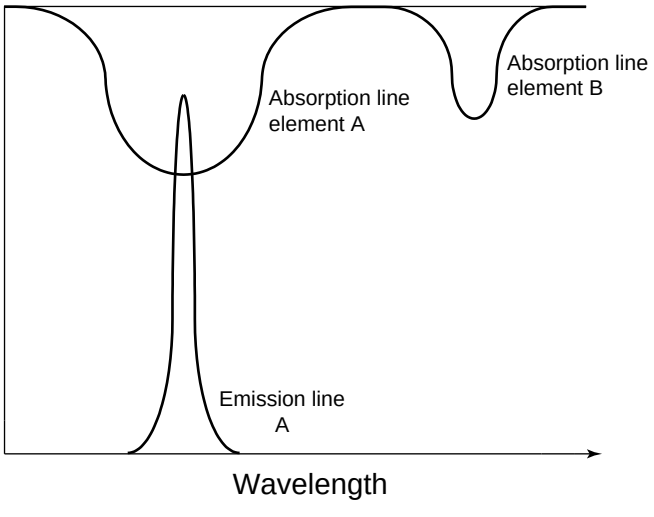

Figure 6 Atomic absorption measurement with a nearby absorption line.

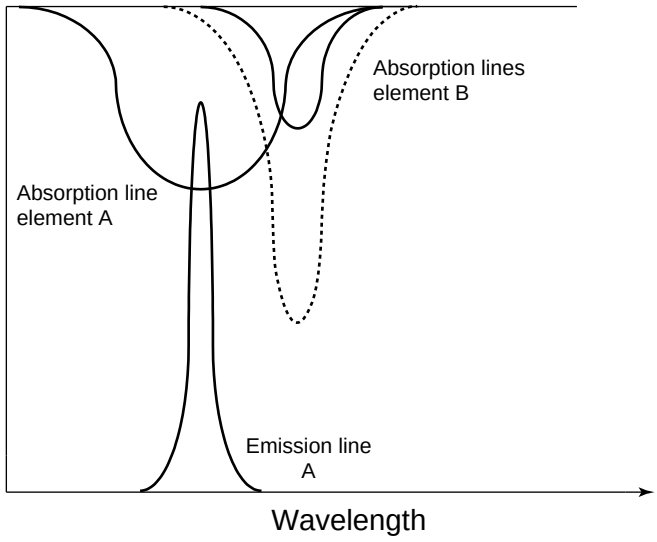

Figure 7 Interfering atomic absorption line.

of B does, as represented here, slightly overlap the right-hand flank of the emission line of the analyte (A).

It should be appreciated, however, that, in addition to the separation of absorption lines, the relative concentrations of the two elements A and B will affect this situation. In the absorption line for element B, at low concentration (shown as the full line) there is a small overlap but as represented it would cause only a slight absorption of a very small part of the total emission signal. However, the absorption due to a higher concentration (represented by the dotted line) is at a greater absorption coefficient, and the overlap is greater. This means that the potential interference will be more significant. Note that the absorption line, as represented, is actually effectively wider. This is because the concentration of atoms absorbing at these flank wavelengths is greater and so the effect will be more observable. Remembering that $A = kcl$, this may be seen as the effect that even though k may be very low as zero absorption coefficient is approached, the effect of multiplying a low k by a high c will result in a larger signal.

Nothing can be done about this type of interference except to use a different resonance line of A for the measurement, since clearly limiting the spectral band-pass of the monochromator will not remove the problem.

When the method was first proposed, it was suggested that since the total atomic absorption spectra of all elements were relatively simple, such interferences might be expected to be rare. This has proven to be the case and no more than 10–12 such interferences have been reported, often rather exotic and unlikely to be encountered in practice. For example, such an interference does occur between copper and europium, as illustrated in Figure 8. The most sensitive line of copper at

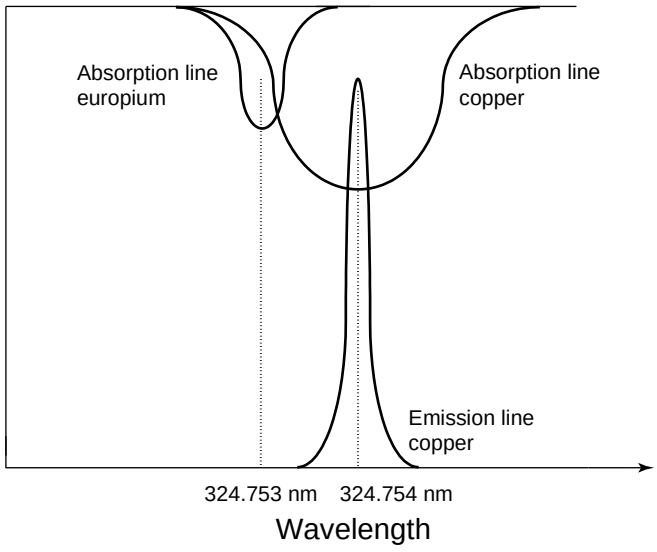

Figure 8 Atomic absorption measurement of copper in the presence of europium.

324.754 nm is very close to an absorption line of europium at 324.753 nm. Referring to Figure 8, if we assume that the half-width of the emission line is 0.0002 nm, then the separation of the central wavelengths of the two absorption lines shown is about 0.001 nm. As may be seen, the copper emission line from the HCL is slightly overlapped by the tail of the europium absorption line. This line is not the most sensitive line for europium so its absorption coefficient would be very low, and additionally the overlap is very small.

In fact, significant interference will only be observed if the Eu–Cu ratio is greater than about 500 : 1. However, if copper needed to be determined, in a sample which contained a very high concentration of europium an error would occur. In what is, after all, a rare analytical situation, the solution to the problem would be to measure copper at the alternative resonance wavelength of 327.4 nm where no such interference occurs and where the sensitivity for copper measurement is reduced by a factor of only about two.

Another case of true spectral interference is the possible absorption of the resonance line of the analyte by molecular absorption. Some molecular species, which might be produced in the atomizer from certain sample types, might have absorption bands which overly the resonance wavelength. Such nonspecific absorption can occur in the presence of high salt concentrations, particularly at wavelengths in the low-ultraviolet region of the spectrum (below 300 nm).

This type of interference is seldom very major in flame atomization but may be needed for the most precise analyses, particularly for measurements at very short wavelengths. It is, however, very important in furnace atomization methods and for this reason background correctors have been developed and will be discussed briefly in the next section.

3 INSTRUMENTATION FOR ATOMIC ABSORPTION SPECTROMETRY

Following the publication of his first paper,[6] Walsh, with some colleagues, described a suitable instrument for the practical measurement of atomic absorption spectra.[8]

The essential component sections of an atomic absorption spectrometer are shown in Figure 9. As illustrated these are as follows:

- a light source (usually an HCL) which provides a sharp line emission at the resonance wavelength for the element to be measured (analyte);

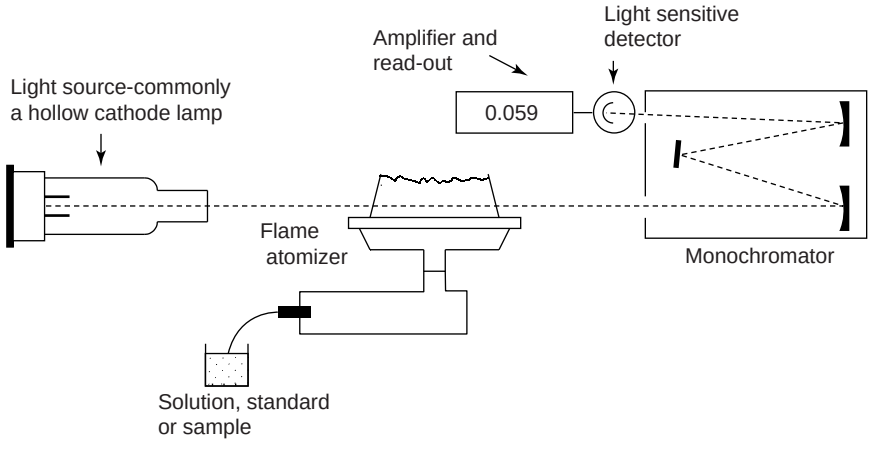

Figure 9 Diagrammatic representation of an atomic absorption spectrometer. (Reproduced by permission of Varian Australia Pty Ltd. from *Introducing Atomic Absorption*, 1983.)

- an atomizer (shown here as a flame) which produces free atoms of the analyte;
- a wavelength selection device (usually a monochromator) which allows the isolation of this resonance line from other lines produced in the spectrum of the lamp;
- a detector (usually a photomultiplier tube) which measures the intensity of the signal which is passed by the wavelength selection device;
- an amplifier–read-out device which processes the signal from the photomultiplier and converts it into a suitable form from which the result is obtained.

Each of these component sections will now be examined in some detail.

3.1 Light Sources

3.1.1 Hollow-cathode Lamp

The principal light sources which were first used in the development of a practical atomic absorption instrument, in the early 1950s, were HCLs. Such lamps were, at the time, relatively uncommon and an initial task of Walsh and his colleagues was to perfect the means of making such lamps which were both stable and exhibited adequately long lives.[9] Possibly surprisingly, nearly 50 years later, such lamps still remain the basis of most atomic absorption instruments. A sketch of such a lamp and an illustration of its operation appear in Figure 10(a) and (b), respectively.

The lamp consists of a glass envelope into which are sealed the contacts leading to the anode and cathode. The cathode is either made from the element whose spectrum is sought, an alloy containing this element, or an insertion of the element or a compacted powder mixture containing the element.

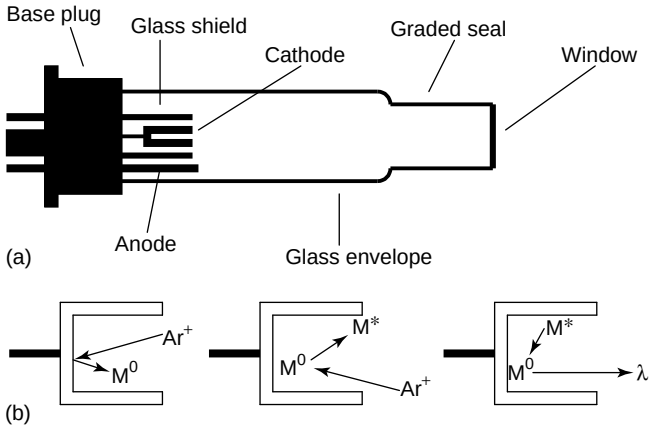

Figure 10 (a) An HCL and (b) its operation. (Reproduced by permission of Varian Australia Pty Ltd. from *Introducing Atomic Absorption*, 1983.)

A fused-silica window is sealed to the end of the lamp with a 'graded seal' joining the window to the envelope. This is necessary as the coefficient of expansion of the glass in the envelope is much greater than that of the silica and cracking would result following the heating and cooling which take place during lamp operation.

The lamp is filled with a monatomic inert gas such as argon, helium or neon, at a low pressure (1–2 kPa). Early lamps generally used argon as the fill gas but for some time neon has been used as the 'standard' fill gas. Argon is used for a few lamps, principally for lamps of elements where the presence of neon causes the emission of lines which are closely adjacent to the resonance lines commonly used for the measurement of that element. In operation, a potential difference of 400–500 V is applied between the anode and the cathode. Initially this ionizes some of the neon (or argon) and once current flow commences the potential difference across the electrodes will fall to about 250 V. The positively charged ions of the inert gas are attracted to the (negatively charged) cathode by electrostatic forces, causing the ions to be rapidly accelerated towards the cathode.

The rapidly moving ions of the inert gas, on striking the interior of the cathode, dislodge atoms from the cathode material (''cathodic sputtering''), creating a cloud of atoms at the mouth of the cathode. Many of these atoms are excited by collision with more of the rapidly moving ions of the inert gas and then, on spontaneously returning to the ground state, emit the emission spectrum of the element.

The total spectrum produced by this process will consist of atomic and ionic emission lines of the inert gas and atomic and ionic emission lines of the element(s) contained in the cathode. Of course, for any element the strongest atomic emission lines will be the resonance lines, since the resonance transitions will be the most likely to occur.

Such HCLs can be made for all of the elements which may be determined by atomic absorption, and are the standard light sources used for most analytical atomic absorption measurements. They do, however, have finite lifetimes due to a process which is called ''clean-up''.

As described earlier, HCLs operate by some of the material of the cathode being removed, as atoms, by a sputtering process. After these atoms have been excited and given off their emission signals, they eventually condense on a cooler part of the lamp as a thin film of the metal. In general, most of this thin film of metal appears on the portion of the inner wall of the lamp which is immediately adjacent to the cathode itself as a mirror like layer. Since this film is literally deposited atom by atom, it has a very large surface area and adsorbs

some small amount of the fill gas used in the lamp. Eventually this adsorption process reduces the fill gas pressure sufficiently that the lamp will no longer sustain a hollow-cathode discharge and the signal of the resonance line is no longer strongly produced. At this stage the lamp must be replaced.

Operating lifetimes of lamps were initially a problem, particularly for some volatile elements, but, with progressive development, today are typically from 1000 to as much as 10 000 h or longer.

Light sources other than HCLs have been used through the period of development of the technique but only two of these are commercially available and in serious use today. These are the boosted discharge HCL (initially called the ''high-intensity lamp'') and the so-called electrodeless discharge lamp.

3.1.2 Boosted-discharge Hollow-cathode Lamp

This has the essential construction of an HCL, but with additional auxiliary electrode(s) sealed into the lamp beside the open end of the hollow cathode, as shown in Figure 11.[10] Similar designs have been used by various manufacturers, variously called ''high-intensity lamps'' and ''super lamps'', but the principle remains essentially the same.

For probability reasons, such as described earlier for flame excitation, most of the atoms in a hollow-cathode discharge are in the ground state at any given time. Many of these atoms may be excited by the passage of a low-voltage discharge from the auxiliary electrode(s) through the atom cloud which exists in front of the cathode. The result is a more intense emission signal and, because the auxiliary discharge is of low energy (voltage), a greater proportion of the extra emission signal occurs at the resonance lines.

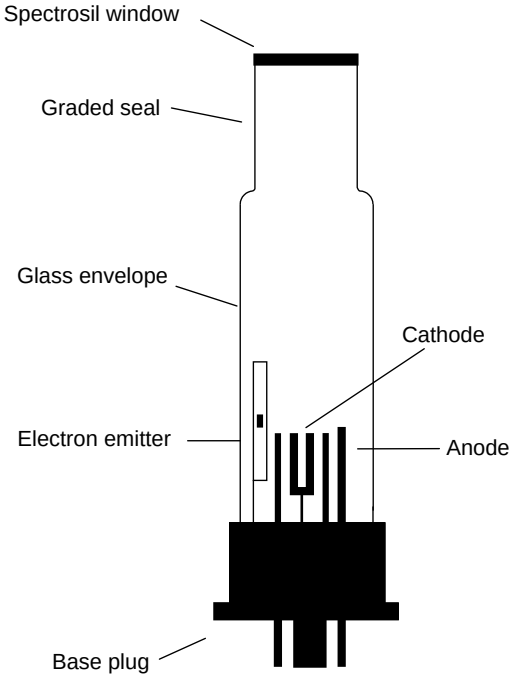

Figure 11 A boosted discharge HCL. (Reproduced by permission of Photron Pty Ltd. from *Photron Lamps*.)

3.1.3 Electrodeless Discharge Lamps

A typical commercially available electrodeless discharge lamp is illustrated in Figure 12.

An electrodeless discharge lamp consists of a small, sealed-off silica bulb containing a very small mass of the element whose emission spectrum is required and a low pressure of inert fill gas. Excitation is by a radiofrequency signal coupled into the contents of the bulb via an induction coil surrounding the bulb. Such lamps are only readily produced for relatively volatile elements such as As, Se, Na and K, but for these elements more intense signals are produced compared with HCLs. Additionally, since HCLs for these volatile elements have shorter lifetimes, as described above, the electrodeless discharge lamps often also have the advantage of longer operating lifetimes.

3.1.4 Operation of Light Sources

Since the most sensitive wavelengths for atomic absorption are generally also the most sensitive wavelengths for atomic emission, some interferences may occur for elements which are significantly excited in the flame atomizer. This is because the flame atomic emission signal would add to the total signal incident on the detector. To overcome this, the signal from the light source is modulated, and an alternating current (AC) amplifier is then able to discriminate this signal from the (essentially) direct current (DC) signal produced by flame atomic emission. Such a modulated signal may be produced by applying a modulated current to the light source or by

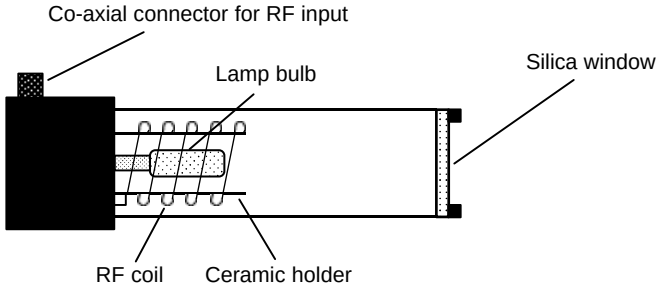

Figure 12 An electrodeless discharge lamp. (Reproduced by permission of Perkin-Elmer Corporation from *Concepts, Instrumentation and Techniques in Atomic Absorption Spectrophotometry*, 1978.)

mechanical chopping. Both techniques are in use in commercial instruments, as will be shown later in this section. A stable power supply is necessary, regardless of which of these modulation procedures are used, in order to produce a stable output from the light source.

3.2 Atomizers for Atomic Absorption Spectrometry

Without the production of free atoms, atomic absorption cannot be measured. It is in this section of the instrumentation that the operator of an atomic absorption spectrometer has the most control over the actual analysis and its veracity. Since the choice of the means of atomization employed and its detailed operation are of such fundamental importance to the technique, more detailed discussion will be postponed until later.

3.3 Wavelength Selection

As has already been discussed, the real resolution of atomic absorption measurement is provided by the absorption process itself. The function of the monochromator in the technique is therefore only to 'tidy up' the emission spectra from the light source. This is illustrated diagrammatically in Figure 13.

Here it may be seen that the light source is producing a spectrum, in the vicinity of the desired resonance wavelength R, containing a number of other lines. These lines may consist of other lines of the analyte, lines produced by other elements in the cathode material and lines of the fill gas of the lamp. Separating the resonance line (R) from all other lines in the vicinity is desirable in order to maximize sensitivity and limit calibration curvature.

Generally a modest monochromator will achieve this fairly readily, with a spectral band-pass of 0.7–1 nm usually being adequate. In a few cases where the analyte itself has another line near the resonance line (and this occurs with Fe, Co and Ni), a spectral band-pass of 0.2 nm will be advantageous in the interests of achieving a calibration graph which is as linear as possible.

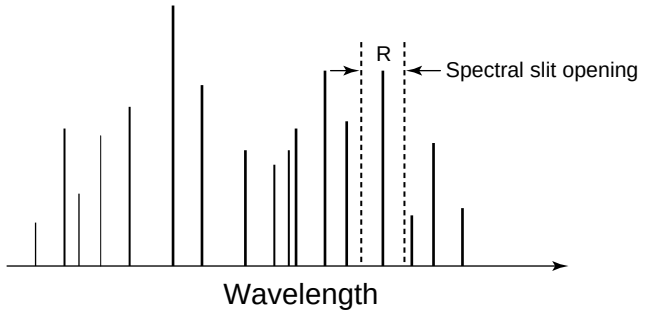

Figure 13 Isolation of a resonance line.

Initially monochromators which employed prisms for dispersion were commonly employed. However, for several years now most atomic absorption spectrometers have employed a monochromator with a diffraction grating as the means of wavelength dispersion. Such monochromators are more desirable than those employing prisms for dispersion as the wavelength settings are less prone to drift with temperature change. Obviously this is an important property when hot flames are being used as the means of atomization.

3.4 Detectors

The universal detectors of atomic absorption spectrometers for many years have been photomultiplier tubes. In the formative years of the technique it was often considered necessary, for the best results, to employ two different types. The range of AAS, as normally used, is from 193.7 nm for arsenic to 852 nm for cesium. In the 1950s there were no photomultipliers that covered this entire range with good sensitivity. Users who wished to measure arsenic and selenium as well as potassium needed two different types (rubidium and cesium, if needed, made this situation even more acute, but very few analysts measured these elements).

Two factors have altered this situation since that time. First, more intense light sources, including electrodeless discharge lamps, made the problem less acute. Second, the technology of photomultiplier tubes has improved and much better tubes with a wider wavelength range have been developed. Both of these improvements, of course, had a compound effect and today a single photomultiplier tube is all that is needed.

3.5 Amplifier and Read-out

As indicated before, the amplifier should respond only to a modulated AC signal and ignore any essentially DC signal resulting from flame emission. For the best performance this is accomplished by employing an amplifier tuned as closely as possible to the frequency at which the light source signal is modulated. Preferably it should be a synchronously demodulated amplifier which is ''locked in'' to the frequency of the lamp modulation.

The relationship between the concentration of the analyte and its absorption is logarithmic so that the amplifier will incorporate a logarithmic function to achieve a linear, or close to linear, read-out. Modern amplifier systems also incorporate mathematical processing of curved calibration functions to achieve direct concentration read-out of the result, regardless of whether the calibration function is linear or otherwise.

4 ATOMIZATION SYSTEMS AND THEIR USE

4.1 Historical

Historically the first actual use of atomic absorption measurements for quantitative analysis occurred in the late 1930s, long before Walsh's first suggestions. The British company Adam Hilger (later Hilger and Watts) developed an instrument to measure the concentration of mercury in air samples.[11] This was a specialized approach to serve a particular need and was, of necessity confined to mercury. Of all the elements which may be measured by AAS, mercury is the only one which can exist at room temperature and pressure in the stable form of free atoms. Later discussions will concentrate on mercury individually but, for the moment, more general approaches need to be considered.

4.2 Flame Atomization

4.2.1 Early History of Flame Atomic Emission

Flame atomization methods had been in use for many years when atomic absorption methods started to be developed in the 1950s. They had been used in flame emission analysis since the beginning of the century although probably seriously only after the important work of Lundegardh in 1928.[5]

The fact that these were indeed means of flame atomization was obscured by the fact that they were also means of excitation and were viewed more in this vein. However, as stated earlier, in flame emission methods the process is first of all the production of atoms and only then the excitation of those atoms produced.

From what has been said before, it should be clear that since the wavelength of radiation resulting from atomic emission is inversely proportional to the energy level of the excited state, then low-energy excitation sources can only produce emission for elements which emit radiation at long wavelengths.

Although Lundegardh's development of the use of the air–acetylene flame did extend the range of elements which could be determined by atomic flame emission, the temperature of this flame (ca. 2500 K) was not high enough to excite very many elements significantly. Since inadequate temperature was obviously the major limitation, investigators in the 1940s and later were drawn to the use of gas mixtures which would give rise to hotter flames, notably the readily available oxygen–acetylene mixture, which produced a temperature of about 3300 K. However it was found that the use of this gas mixture in a premix system could easily lead to serious explosion hazards and so alternative systems were developed. Particularly this led to the development of the so-called "total consumption burner", in which the gases were only mixed with each other near the burner mouth. As will be mentioned later, such a burner system, although providing higher temperatures, introduced many problems to the technique.

4.2.2 Initial Developments of Flame Atomizers of Atomic Absorption

When investigations into the use of atomic absorption spectra for chemical analysis first commenced, the emphasis was initially on those elements which, although easy to atomize, were difficult to excite.[8] These include Zn (213.8 nm), Cd (228.8 nm), Ni (232.0 nm) and Pb (283.3 nm). Because of the short wavelengths associated with these resonance lines, the energy required to excite the atoms is relatively high.

Initial use was made of burner systems from existing (emission) flame photometers which employed premix systems and employed relatively cold flames such as air–coal gas (ca. 2100 K). It was immediately found that it was possible to measure elements such as Zn and Cd with excellent sensitivity, comparable, for example, to the sensitivity obtainable for Na and K. This immediately proved one of the major benefits, initially claimed for atomic absorption, that it would be independent of the excitation energy required to populate the first excited state. It was also realized that since the method did measure absorption, a longer absorption pathlength increased the sensitivity, and suitable elongated slot burners were soon developed. Such burners had to employ slot dimensions which were matched, as closely as possible, to the geometry of the beam of light from the HCL.

4.2.3 The Use of the Air–Acetylene Flame

It was soon realized that although high energy was not needed for excitation purposes, it was, for some elements, needed to achieve atomization. The development of elongated slot burners suitable for use with the air–acetylene flame (in the early 1960s) extended the technique to a number of additional elements so that by this period about 35 elements could be determined by AAS using this flame. (The scientist most instrumental in the first development of air–acetylene burners, suitable for atomic absorption, was Allan[12]).

Relatively soon this flame became virtually the standard flame in use,[13] and this remains the situation today. Several of the additional elements, however, either had modest sensitivities or suffered from a number of matrix interferences, or both.

4.2.4 Development of Hotter Flames

It had become clear that a number of additional elements, although potentially amenable to measurement by atomic

absorption, were not readily available because they were very difficult to atomize. These included important elements such as aluminum, titanium and silicon. These, and many other, elements readily formed very stable monoxides in the flame and at the temperature of the air–acetylene flame produced virtually no free atoms. Attempts were made to use total consumption burners, with various modifications, in order to use the oxygen–acetylene flame, but these proved to be of limited success.

The design problem lay in the very high burning velocity of the oxygen–acetylene flame compared with the air–acetylene flame.

In a premixed system, where the fuel and the oxidant gases are mixed in a chamber before passing through the burner slot, it is of vital importance that the velocity of the mixed gases through the burner slot is never exceeded by the burning velocity of the mixture of gases. Should it do so then a ''flashback'' of the flame into the chamber will result in a severe explosion.

In designing a suitable burner for safe operation with an air–acetylene flame, the slot opening had to be reduced compared with the opening in use with air–coal gas. However, examination of the burning velocities for the acetylene flames, supported by air and oxygen, respectively, made it clear that a huge (and impractical) reduction in slot dimensions would be required to burn oxygen–acetylene safely in a premixed system.

Some developments were first made using physical mixtures of air and nitrogen, as the support gas, and successful determinations of aluminum, titanium, silicon and many other elements were achieved with this approach. Mixtures containing as much as 85% oxygen were investigated, albeit with some hazards.[14]

The use of nitrous oxide as a support gas, however, was a major breakthrough as it permitted the safe burning of nitrous oxide–acetylene using a burner with a sensibly reduced slot opening.[15] As may be seen in Table 2, this flame provided a high temperature with a moderate flame velocity.

Compared with the potentially variable composition of a physical mixture of nitrogen and oxygen, nitrous oxide, on thermal decomposition in the flame, provides a mixture containing 67% (v/v) nitrogen and 33% (v/v) oxygen. However, it produces a flame more equivalent in temperature to that achieved with 50% N_2–50% O_2 owing to the exothermic decomposition of nitrous oxide.

With the additional use of this flame, a total of about 65 elements could, from 1965, be determined by AAS using flame atomization. More detailed examination of the flame atomization process will be presented later in this article.

Table 2 Characteristics of various types of flames

Gas mixture	Burning velocity ($cm\,s^{-1}$)	Maximum temperature (K)	Burner slot (mm)
Air–coal gas	55	2100	100 × 1.5
Air–propane	85	2200	100 × 1.5
Air–acetylene	160	2500	100 × 0.5
Nitrous oxide–acetylene	280	3100	50 × 0.5
Oxygen–acetylene	1130	3300	–

4.2.5 Nebulizer–Burner Systems

Many different types of flame nebulizer–burner systems have been employed for atomic absorption. Some of these have been very useful for certain applications but for the purposes of this article the description will be confined to those commonly employed in available commercial instrumentation. A typical system is illustrated in Figure 14.

This system operates as follows. A flow of support gas (air or nitrous oxide), under pressure, passes into the nebulizer. The gas is expanded through a venturi throat and develops a pressure difference which draws a flow of sample solution through the inlet capillary. This solution, on emerging from the mouth of the venturi throat at a very high velocity, is shattered into droplets ranging in size from as small as 1–2 μm up to 100 μm or more, and this spray of droplets passes into the spray chamber. In this chamber, the oxidant gas and the droplets of sample solution are mixed with the fuel gas (acetylene). The flow spoiler, in this case a set of baffles, encourages the precipitation of the larger and heavier droplets of solution and these are drained from the chamber to waste. Most droplets larger than 10 μm in diameter will precipitate in the chamber. The mixture of oxidant, fuel and fine

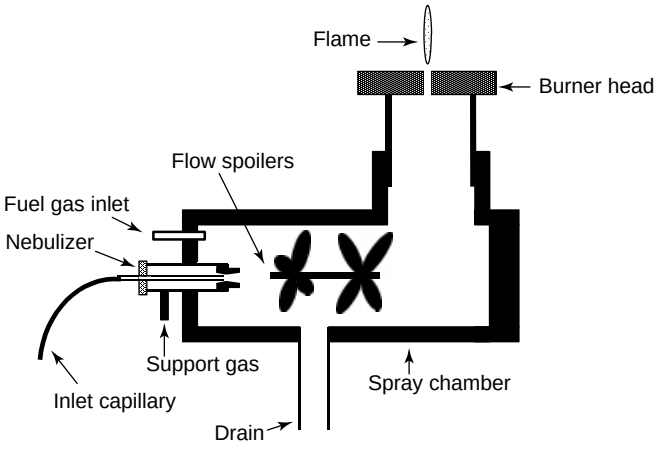

Figure 14 A nebulizer–burner system with flow spoilers. (Reproduced by permission of Perkin-Elmer Corporation from *Concepts, Instrumentation and Techniques in Atomic Absorption Spectrophotometry*, 1978.)

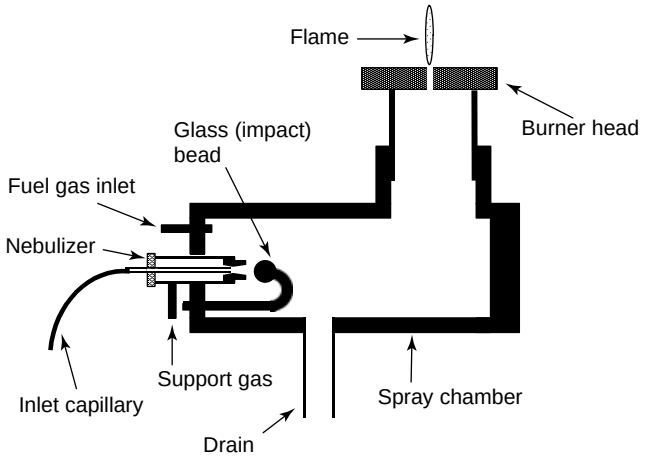

Figure 15 A nebulizer–burner system with a glass impact bead. (Reproduced by permission of Varian Australia Pty Ltd. from *Introducing Atomic Absorption*, 1983.)

spray (aerosol) then passes into the burner head and through the burner slot to the flame, burning above the slot.

A minor variation of the system illustrated in Figure 14 is shown in Figure 15. In this system, an impact bead, adjustable in position, is set in front of the nebulizer throat. When this is adjusted to its optimum position (which is critical), larger signals are produced, indicating that a greater amount of fine aerosol is being produced, presumably owing to the bead causing a more efficient shattering of the larger droplets into smaller ones. Both of these systems are in use in commercial instruments currently available.

Overall, the total nebulizer–burner system has changed little in principle since the 1950s when Walsh and his colleagues first used such a system to demonstrate the method.

Nebulizers were developed which were more corrosion resistant when it was realized that the solutions required to keep some elements in stable solution were very corrosive.

Early burner heads were of relatively light construction, but the development of the nitrous oxide–acetylene flame made additional demands. In particular, since the flame was much hotter, more robust construction was needed to avoid distortion after protracted use. Also, a more massive construction was needed to provide a better "heat sink" to avoid overheating, which could detonate the combustible mixture of gases in the spray chamber. In general practice today the above system is used with two different types of burner head. First, for use with an air–acetylene flame, a burner with a slot of dimensions 100×0.5 mm is used. For burning a nitrous oxide–acetylene flame, a burner with slot dimensions of 50×0.5 mm is suitable.

4.2.6 The Atomization Process in the Flame

As the aerosol produced from the sample solution passes into the flame, the first process will be "desolvation", that is, the boiling off of the solvent (which is generally water). This will result in the production of extremely minute solid particles of the constituents of the dissolved material in the solution.

The effect of the higher temperature, as these particles move into the hotter part of the flame, will then be, in many cases, to dissociate the salts present and produce atoms of the constituent elements. However this is not always as simple as this might sound. Simple compounds such as NaCl and $ZnCl_2$ will very readily dissociate in the simplest manner possible, e.g. as shown in Equation (5):

$$NaCl + \text{heat} \longrightarrow Na^0 + Cl^0 \qquad (5)$$

where Na^0 and Cl^0 represent free atoms of sodium and chlorine. For many elements this simple process will apply in the air–acetylene flame. However, for some elements the process will be more complex and this may be represented for aluminum as Equation (6):

$$Al(NO_3)_3 + \text{heat} \longrightarrow AlO + (\text{various gases}) \qquad (6)$$

This is due to the great affinity for aluminum with oxygen and will occur regardless of which compound of aluminum is present. At the temperature of the air–acetylene flame, AlO is so stable it is virtually completely undissociated, so that no Al atoms are formed, and no atomic absorption signal can therefore be measured. However, at the temperature of the nitrous oxide–acetylene flame the next stage (Equation 7):

$$AlO + \text{heat} \longrightarrow Al^0 + (O) \qquad (7)$$

will occur. Most of those elements which demand the nitrous oxide–acetylene flame for atomization will react in this manner as most of them form very stable monoxide compounds.

Some elements occupy an intermediate position between these two extremes, and this may be typified with calcium. In this case it is appropriate to represent the reactions as Equation (8):

$$Ca(NO_3)_2 + \text{heat} \longrightarrow (\text{various gases}) + CaO \rightleftharpoons Ca^0 + (O) \qquad (8)$$

When air–acetylene is employed as the flame, the equilibrium represented is very much an incomplete reaction and only a portion of the calcium is converted to atoms. At the higher temperature of the nitrous oxide–acetylene flame, however, the equilibrium will be very much pushed

in the direction of almost complete production of atoms and greater sensitivity of measurement will result.

It is worth adding some comments here about the above description as it relates to the so-called ''total consumption'' burner. Such burners were developed to provide safe oxygen–acetylene flames for flame atomic emission measurement. In such burners, constructed as shown in Figure 16, all of the solution droplets pass into the flame, since there is no intermediate spray chamber.

Referring to the description above it follows that, since very many large droplets will pass to the flame, relatively large solid particles are formed following desolvation. These very large solid particles are poorly dissociated in the obviously very limited residence time available. Additionally, such burners provide a much lower temperature than the theoretical oxygen–acetylene flame temperature owing both to the lack of adequate mixing of the two gases and to the cooling effect of the larger solvent volume. All of these factors make such burners very limited as atomizers, whether used for atomic emission or atomic absorption, compared with premix systems such as those described previously.

It is worth pointing out that in the discussions above the accent has almost solely on the temperature of the flame atomizer. This is almost certainly the major factor, but it is not the only factor. The chemical environment of the flame is extraordinarily complex, and of course the chemistry of every single element which may be measured by atomic absorption is different from that of almost every other element. Flames used for atomization

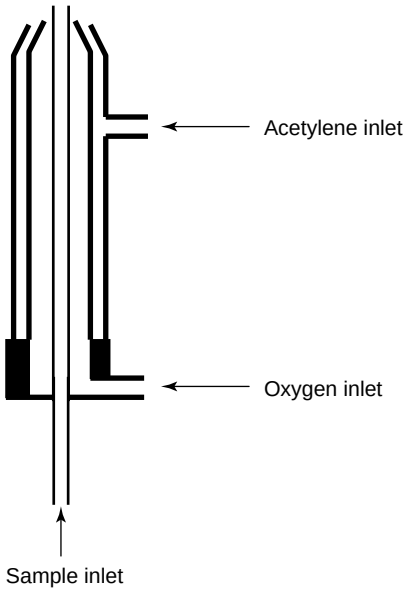

Figure 16 A total consumption burner. (Reproduced by permission of Wiley-VCH from B. Welz, *Atomic Absorption Spectrometry*, VCH, Weinheim, 1985.)

may, for simplicity, be classified as stoichiometric, fuel lean or fuel rich. A stoichiometric flame may be defined as a flame where the amount of oxygen present is exactly the amount required to totally burn the hydrocarbon of the fuel in conformity to the equation of the combustion process. For example (Equation 9):

$$C_2H_2 + 5N_2O \longrightarrow 2CO_2 + H_2O + 5N_2 \qquad (9)$$

For any given fuel and oxidant combination the highest temperature will be provided by such a stoichiometric flame.

A flame which has an excess of the fuel gas present, compared with the amount of oxidant present, is classified as fuel rich, is sometimes also called a reducing flame and will be somewhat cooler than the stoichiometric flame.

However, for many elements, better atomization, as indicated by larger atomic absorption signals, will occur for mildly or even strongly reducing flames. Chromium in the air–acetylene flame is a good example of this, as is silicon in the nitrous oxide–acetylene flame. The full understanding of all of the processes involved probably still evades the scientific investigators of flame chemistry, but as will be discussed later, a practical means of achieving optimum adjustments is still attainable.

4.3 Vapor Generation Techniques

4.3.1 Mercury Cold Vapor

It has been pointed out earlier that mercury was measured by atomic absorption in the 1930s. Using flame atomic absorption, the sensitivity obtained for this element is very poor, particularly when viewed against the background that since mercury is a very toxic element it often needs to be determined at very low concentrations. However, it was found that mercury can be released from any compound associations in solution, and converted to elemental mercury, by adding a reducing agent (such as stannous chloride) to the solution.[16] By bubbling a stream of a gas, such as nitrogen, through the solution the mercury atomic vapor is carried from the solution in this stream. The flow of gas is then passed through a flow cell placed in the light path of the atomic absorption apparatus, which then records the absorption signal. Because the reaction is fairly rapid, a very high transient concentration of atomic vapor of mercury is achieved, giving rise to a large absorption signal peak. Whereas flame methods only allow the detection of about 1 mg L^{-1} of mercury, the ''cold vapor'' method permits the detection of small fractions of 1 μg L^{-1}.

In general, the equipment commonly used for this type of measurement is the same as that used for hydride methods employed for some other elements, and which will be described below.

In the case of stannous chloride the reduction reaction, in solution, is as shown in Equation (10):

$$Hg^{2+} + Sn^{2+} \longrightarrow Hg^{0} + Sn^{4+} \qquad (10)$$

It is also possible to use sodium borohydride ($NaBH_4$), as a means of generating hydrogen, as a reductant as described below for hydride methods.

4.3.2 Hydride Methods

Another group of metals, which exhibit poor sensitivities by flame atomization methods, may also be determined by a different vapor generation method. First suggestions of such an approach were made in 1969.[17] As now employed, the addition of a sodium borohydride solution to an acid solution of the analyte causes the production of hydrogen, and this in turn induces volatilization of hydrides of these elements. Such hydrides may be typified by arsine (AsH_3), which has a very low boiling point and so is readily evolved from solution once formed. These hydrides are readily decomposed, in a quartz flow cell, at modest temperatures (800–1000 °C) to produce free atoms of the analyte. A diagrammatic representation of the principle is shown in Figure 17.

The quartz cell may be heated either by mounting it above a flame produced using a normal burner system or by electrically heating the cell. (Initially the method used was to add metallic zinc powder to the acidic solution, but sodium borohydride gives a faster and more reproducible reduction of the metal to the hydride. In both cases the reduction depends upon the production of nascent hydrogen.) This method is only applicable to those elements which readily produce hydrides of very low boiling point, viz. arsenic, antimony, bismuth, selenium, tellurium and tin.

Using this method, most of these elements have detection limits of $<0.1\ \mu g\ L^{-1}$, whereas in the flame their detection limits are at best around 1000 times poorer at ca. $0.1\ mg\ L^{-1}$.

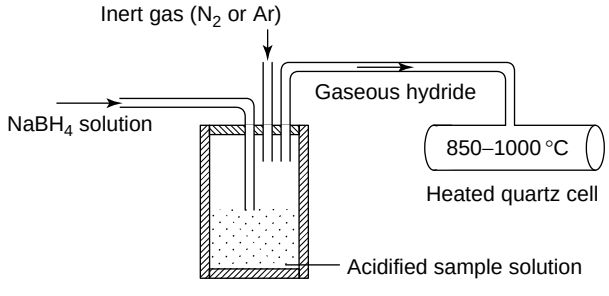

Figure 17 Diagrammatic representation of a vapor generation system. (Reproduced by permission of Wiley-VCH from B. Welz, *Atomic Absorption Spectrometry*, VCH, Weinheim, 1985.)

Lead has been determined by this technique but with some operational difficulty, which limits the appeal of the method for this element.

4.4 Graphite Furnace Atomization

This method was first described in 1961[18] and has been in use with commercial equipment since about 1970. It produces detection limits for many elements which are several hundred times better than flame detection limits of the same elements. It also has the advantage that only very small sizes of sample are needed (a few microliters compared with a few milliliters typically required for flame methods). It has the disadvantages of requiring a moderately expensive accessory apparatus and of being a very slow method. However, it is readily automated, which overcomes at least some of the disadvantages of the slowness. This method is outside the scope of this article and will be dealt with in detail elsewhere.

5 OPTICAL DESIGN OF INSTRUMENTS FOR ATOMIC ABSORPTION SPECTROMETRY

5.1 Atomic Absorption Spectrometer Design

Two basic types of optical system have been employed since the first atomic absorption instruments were produced commercially. These are the single-beam system and the double-beam system and simplified diagrams of such types are illustrated here.

In Figure 18 is shown a very simple single-beam system. This contains an absolute minimum of optical components and is therefore optically very efficient, and also relatively inexpensive. The major disadvantage of such a system is that any instability in light source output will result in a variable base line reading.

In Figure 19 is shown a simple double-beam system where the beam is split and two paths of light, one through the atomizer (sample beam) and the other around it (reference beam), are produced. These are subsequently recombined and both signals measured and compared with one another.

The obvious advantage is that since comparison of the two signals is continually and frequently being made, variations in source intensity are cancelled out. Two means of achieving this are in use in commercial instruments. That illustrated in Figure 19 employs a beam splitter and recombiner. The same result can be achieved by employing a system of chopping with a rotating mirror, as shown in Figure 20. In both cases the systems are not as simple as the single-beam system and consequently the cost is greater.

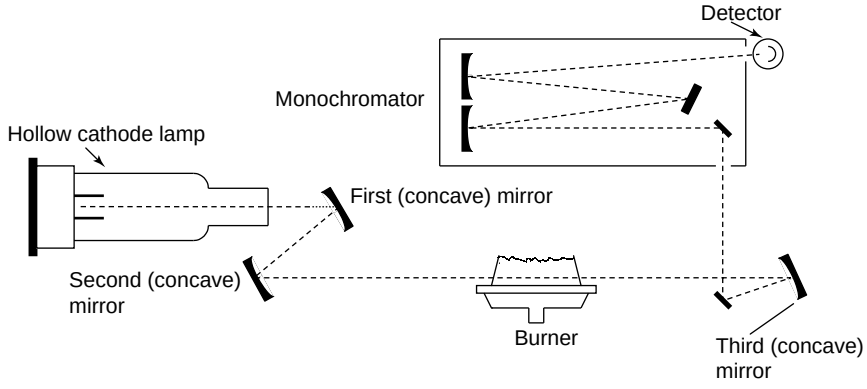

Figure 18 An optical system for a single-beam atomic absorption spectrometer. (Reproduced by permission of Varian Australia Pty Ltd. from *Introducing Atomic Absorption*, 1983.)

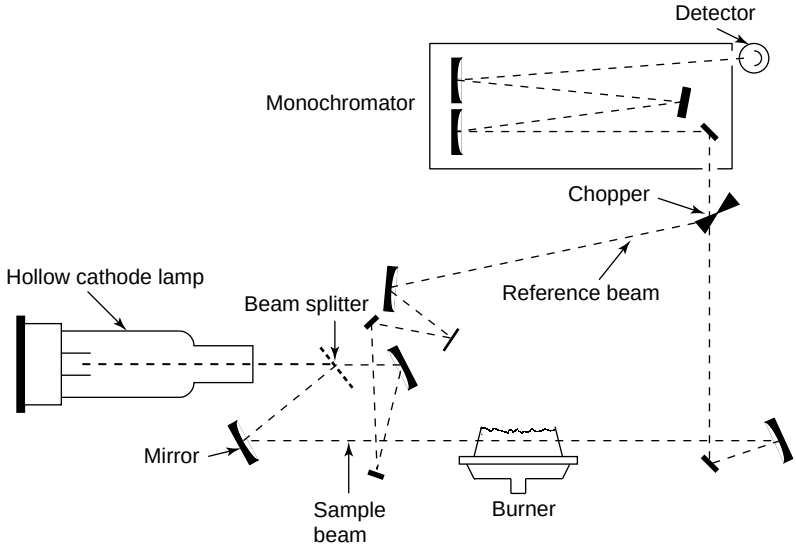

Figure 19 An optical system for a double-beam atomic absorption spectrometer with beam splitters. (Reproduced by permission of Varian Australia Pty Ltd. from *Introducing Atomic Absorption*, 1983.)

5.2 Optical Design for Background Correction

As mentioned previously, attenuation of the signal at the resonance wavelength by molecular absorption can cause errors in some cases, particularly when GFAAS is used. However, even in flame atomic absorption if the concentration of the analyte is low (giving rise to a very small atomic absorption signal), and the concentration of species producing molecular absorption at the analyte wavelength is high, errors will be introduced if correction is not made.

Figure 21 shows how a simple means of correction using a continuum light source may overcome this problem, a method first described in 1965.[19] This is illustrated by comparison with the double-beam instrument shown in Figure 19, but such an approach can be similarly used with the systems illustrated in Figures 18 and 20.

Alternative measurements with the sharp line source and the continuum source will measure as indicated in Equations (11) and (12):

$$\text{sharp line source} = AA + MA \tag{11}$$

$$\text{continuum source} = MA \tag{12}$$

where AA = atomic absorption and MA = molecular absorption. Subtracting the second reading from the first then gives (Equation 13):

$$(AA + MA) - MA = AA \tag{13}$$

Some of the radiation from the continuum source will be absorbed by atomic absorption. However, the instrument will be looking at the radiation from the continuum across a spectral band-pass of, say, 0.7 nm.

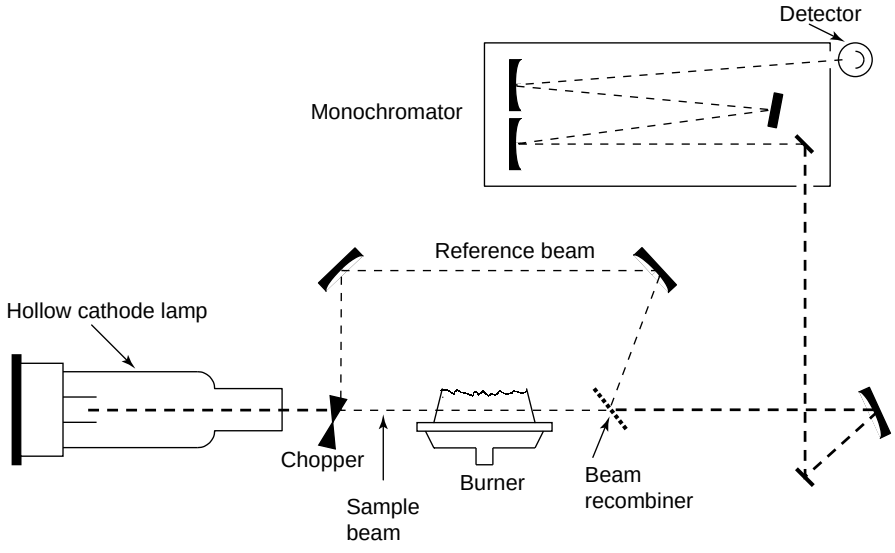

Figure 20 An optical system for a double-beam atomic absorption spectrometer with a reflecting chopper. (Reproduced by permission of Perkin-Elmer Corporation from *Concepts, Instrumentation and Techniques in Atomic Absorption Spectrophotometry*, 1978.)

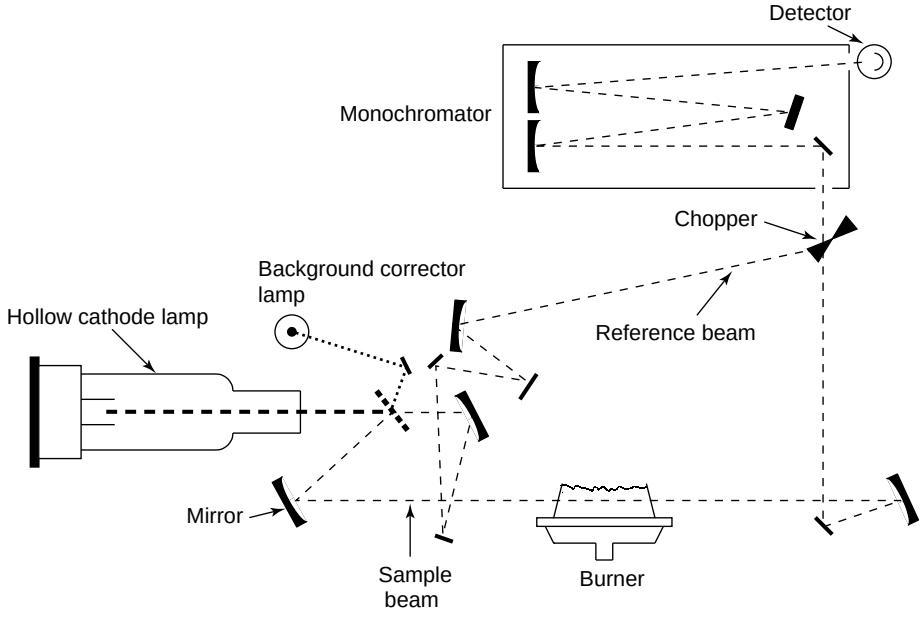

Figure 21 An optical system for a double-beam atomic absorption spectrometer with background correction. (Reproduced by permission of Varian Australia Pty Ltd. from *Introducing Atomic Absorption*, 1983.)

Since the atomic absorption will only be over an absorption profile maybe 0.01 nm wide, less than 1% of this radiation will be absorbed and as a first approximation this may be ignored with very little resultant error.

The situation is completely different for GFAAS methods. Here it may be commonplace for the background absorption to be of a magnitude which is greater than that of the atomic absorption signal. Very careful design of the background corrector and its very careful adjustment are then most important. For this reason, a much better means of background correction is the use of the Zeeman effect background correction system. This is outside the scope of this article since it is applied only to the graphite furnace method.

6 PRACTICES OF ATOMIC ABSORPTION SPECTROMETRY

6.1 General

Like most spectroscopic methods, AAS is a comparative method, that is, it requires the calibration of the instrument with standards of known concentration before the analytical measurement of the samples may be made. However, as will be discussed later, it is generally not necessary to match the composition of the calibration standards to the samples with regard to elements other than the analyte. The calibration standards may generally be simple aqueous solutions containing nothing but accurately controlled concentrations of the analyte. Like almost all instrumental methods, the performance is limited by the signal-to-noise ratio, with regard to both the detection limit and the precision achieved. In the following sections these factors will be discussed more and attention given to means of optimization in order to obtain the best possible detection limits and precision.

6.2 Optimization of Instrumental Parameters

Because the optimization process involves two factors, signal (size) and noise, it often involves some sensible compromise. This will be briefly discussed in relation to the various parameters which may be set by the operator of an atomic absorption spectrometer.

6.2.1 Hollow-cathode Lamp Currents

Referring to earlier discussion on line shapes, the selection of lamp currents is very much a compromise. As the HCL current is increased, two things occur. A stronger signal is produced but inevitably the line width of the emission signal of the resonance line is also increased. The first of these changes will allow the use of a lower gain in the photomultiplier and amplifier system and this will decrease the ''noise'' produced from the photomultiplier and amplifier system. However, the broader emission line will decrease the sensitivity and this will reduce the absorbance for a given concentration and may also increase the curvature of calibration. In method development, the best compromise should be experimentally determined with reference to obtaining the best detection limit and/or the best precision.

6.2.2 Wavelength Selection

In general, for a majority of elements, the most sensitive line will be used for the measurement. However, this is not always the case and some brief examples will be mentioned.

In the measurement of lead, the most sensitive line is at 217.0 nm and this will produce a signal, for a given concentration, which is about twice as large as that obtained if the 283.3-nm line is used. However, the signal at 217.0 nm has a much greater noise component because of two factors:

1. The photomultiplier response at 217.0 nm is much poorer than at 283.3 nm so the photomultiplier will have to be operated at higher dynode volts, giving rise to more noise.
2. At 217.0 nm the flame gases absorb more radiation than they do at 283.3 nm and so minor fluctuations in the flame will generate some noise.

As a result, the best signal-to-noise ratio and hence the best detection limits and precision will be achieved using the 283.3-nm line even though it has the lesser sensitivity.

For the measurement of low concentrations of nickel, the 232.0-nm line (which is the most sensitive line for nickel) is preferable. However, owing to a very closely adjacent (and virtually non-absorbing) line of nickel, curved calibration graphs are produced for this element, this effect becoming very pronounced at high concentrations. Obviously, at the extreme of curvature when the calibration graph becomes almost flat, measurement is no longer possible. If samples containing high concentrations of nickel are to be measured, the less sensitive line at 341.5 nm will provide almost linear calibration graphs over a wide range of concentration.

Many similar cases may be cited – the rule should be that for optimum results each element and each application need to be investigated.

6.2.3 Monochromator Settings

The slit opening selected is generally a matter of recommendation by the equipment manufacturer as a function of the element to be measured and the wavelength to be employed. For example, considering the case of nickel discussed above, if this is to be measured at the 232.0-nm line a slit opening of, say, 0.2 nm should be chosen, rather than a wider slit to maximize the exclusion of the adjacent nonabsorbing line of nickel. However, if the 341.5-nm line is used, a larger slit opening should be chosen to increase the light throughput as this will improve the signal strength. Since there are no lines closely adjacent to this line, this may be done with impunity.

6.2.4 Flame Selection

For those elements which are easy to atomize, the air–acetylene flame should be used. It is less expensive to operate and will give rise to better sensitivity than if the

nitrous oxide–acetylene flame is used (unnecessarily). In general, nitrous oxide will be used for those elements which are either atomized poorly, or not at all, with the air–acetylene flame.

There are some elements which occupy a central ground in this case. Elements such as calcium and strontium are only modestly atomized in air–acetylene and are subject to some interferences. By employing the nitrous oxide–acetylene flame, not only is the sensitivity improved but also the interferences are overcome. There are a number of such cases and again for each element the recommendations of the instrument manufacturers need to be accepted or the actual element and sample type need to be studied for each analysis by the analytical chemist.

6.2.5 Flame Optimization

It should be understood that each element has its own optimum atomization conditions, although for some elements the differences between them may be relatively small. This applies specifically to the composition of the flame used and the position of the burner in the light path.

Some elements give a more efficient atomization, resulting in a larger signal, if the flame is very ''lean'', that is, oxidizing or deficient in fuel. On the other hand, certain elements, such as chromium, require very 'fuel-rich' flames, that is, reducing and having an excess of fuel. Additionally for each element, and for each flame composition, the portion of the flame in which the maximum population of atoms occurs will be different. Since the maximum absorption signal will result when the beam from the HCL passes through this region of maximum atom population, efforts must be made to adjust the burner position to achieve this situation.

The best means of achieving both of the above two optimizations is, while aspirating a solution of the analyte, first to adjust the gas mixture to achieve maximum absorption and then to adjust the burner position. Sometimes, with elements for which the atomization conditions are critical, more than one iteration of these two adjustments might be necessary.

7 INTERFERENCES IN ATOMIC ABSORPTION SPECTROMETRY

7.1 Spectral Interferences

This subject has been already covered in some detail. True spectral interferences will only occur when an absorbing line of another element overlaps the emission line of the analyte from the light source and the interfering element is also present in the sample. This is, as previously stated, a rare occurrence and probably all such cases have, by now, been documented in the literature.

7.2 "Chemical" Interferences

Such interferences are sometimes also called ''matrix interferences'', and arise when an element, or compound, present in the sample (but not in the calibration standards) interferes with the production of atoms. A classical case has already been mentioned in passing.

If a simple solution of calcium of a given concentration is sprayed into the flame a certain reading will result. If, instead of the simple calcium solution, one containing the same calcium concentration, but also a significant concentration of phosphorus, is sprayed into the flame a much lower reading will be obtained. This is caused by the formation in the flame of some calcium phosphate type of compound which is harder to convert to atoms than the CaO usually formed. Obviously one could match the standards with the same P concentration but this would require the prior determination of P in each sample. Similar interferences will occur for all of the alkaline earth group – Ca, Mg, Sr, Ba and Mg – in the presence of P, Al and Si.

Such interference may be overcome in two ways:

1. By adding a large concentration of strontium or lanthanum a competing mechanism is set-up. Because the Sr or La is present in much greater concentration than Ca, and because Sr and La are chemically similar to Ca, the P will preferentially associate with the Sr or La rather than the Ca. In practice, the Sr or La is added both to the samples and to the calibration standards.
2. If the nitrous oxide–acetylene flame is used, the interference disappears because the flame is hot enough to dissociate readily the compounds formed.

These are probably the most severe cases but minor interferences of a similar nature occur with some other element combinations and the literature should always be consulted before embarking on any totally new application.

7.3 Ionization Interferences

At the elevated temperatures occurring in flame atomizers, the atoms of some elements will be ionized, that is, the equilibrium shown in Equation (14) will occur:

$$M^0 \rightleftharpoons M^+ + e \qquad (14)$$

where M^0 is the ground-state atom of element M, M^+ is the ion of the same element and e is an electron. Since

this particular atom is now an ion and not an atom it will not absorb radiation at the same wavelength.

This effect will occur with elements that have ionization potentials which are low relative to the energy of the flame used for atomization. With air–acetylene flames significant ionization will occur only for the alkali metals, Cs, Rb, K and Na, all of which have very low ionization potentials.

In the much hotter nitrous oxide–acetylene flame, however, many elements will be significantly ionized. The most extreme are the alkaline earths, Ca, Sr and Ba, but ionization will also occur significantly for many elements such as Al and the rare earth elements.

The ionization reaction shown above, being an equilibrium, is subject to the law of mass action (Equation 15):

$$K = \frac{[M^+][e]}{[M^0]} \qquad (15)$$

where K is an equilibrium constant, $[M^+]$ is the concentration of the ion of the atom in the flame, $[M^0]$ is the concentration of the atom of the element in the flame and [e] is the concentration of electrons in the flame. If there is present in a sample, apart from the analyte, any element which readily ionizes (such as Na), then the additional electrons contributed to the flame by the ionization process will cause the equilibrium to shift back to the left in the equation and favor the production of more ground-state atoms. This will increase the atomic absorption reading. If the calibration standards have no sodium present then the readings obtained for the standards will be relatively lower. The net result will be a high result for the sample. In the case of the measurement of a number of samples with varying Na content, this will produce errors which will also vary.

To overcome this problem all solutions, calibration standards and samples should have added to them a large excess of a salt of an easily ionized element as an ionization suppressor. The most common such suppressor used is cesium, generally added as the chloride, which has the lowest ionization potential of all elements. This suppressor will ionize very substantially and in doing so will generate a high concentration of electrons. As described before, this high concentration of electrons will push the equilibrium back very strongly in the direction of simple (nonionized) atoms.

7.4 "Physical" Interferences

Some errors may also occur due to the physical composition of the sample solutions, by comparison with the calibration standard solutions. For example, if solutions of a solid sample are prepared by using high concentrations of sulfuric and/or phosphoric acid, the viscosity of the solutions will be high compared with simple aqueous solutions. This will diminish the uptake rate of the solution by the nebulizer and also will diminish the yield of fine spray to the burner. The error caused by this effect can only be prevented either by avoiding the use of such viscous acids or, if this is not possible, preparing the standards with the same concentrations of these acids. Similar viscosity effects may occur with very high concentrations of dissolved solids and again such problems can only be avoided in the same way.

The presence of a miscible organic compound may have the effect of reducing the viscosity of a solution and thus enhance the efficiency of nebulization. Ethanol, for example, will cause this effect. If an element in a sample of wine is measured by comparison with aqueous standards of the same element, high results will be obtained. Again, the solution to the problem will be to match, or at least approximate, the alcohol concentration of the calibration standards to that of the wine.

8 VALIDATION OF RESULTS OBTAINED BY ATOMIC ABSORPTION SPECTROMETRY

As may be seen from much of the preceding discussion, many factors may influence the veracity of results obtained using atomic absorption. In order to avoid errors due to an unexpected influence, it is wise to accompany any batches of samples analyzed with one or more standard reference materials, as close as possible in type to the samples. Control charts should be kept of the results obtained and remedial action taken if any deterioration in the quality of results is observed.

9 AUTOMATION IN ANALYSIS USING ATOMIC ABSORPTION SPECTROMETRY

Many automatic or semi-automatic atomic absorption spectrometers are now available. It is possible with such instruments, generally computer controlled, to store various parameters such as lamp currents, wavelengths, monochromator slit openings and flame gas flows. Although clearly such instruments are more costly, they do permit analyses to be performed by semiskilled operators once the parameters have been determined and stored by an expert analyst.

Automatic sample presentation to the instrument permits unattended operation, once a suitable type of automatic sampler has been loaded with samples and the analysis programmed. With instruments such as those described above, it is possible to measure sequentially a large batch of samples for more than one element as part of an automatic process.

Caution is suggested, however, with the totally unattended use of flame atomic absorption instruments for obvious safety reasons.

One of the most important developments for automated analysis by AAS is the use of flow injection analysis (FIA). This method represents a means of faster automated analysis and also permits minimization of the sample volume. Suitable FIA methods and equipment have been developed for use with flame and vapor generation methods of atomization. Details of such methodology properly belongs in an article on FIA and so are not considered further here.

10 COMPARISON WITH OTHER ANALYTICAL TECHNIQUES

10.1 Flame Emission

At the time when atomic absorption methods were first introduced, the obvious comparison was with flame atomic emission. Today it can safely be said that the atomic absorption method has all but replaced flame atomic emission in practical analytical chemistry.

FAAS provides a method which is rapid and easy to use, combined with simplicity of operation, and is not very prone to unexpected errors. It has a good sensitivity for most elements, adequate for many purposes. However, it measures only one element at a time and so is not particularly rapid for multielement analysis.

10.2 Inductively Coupled Plasma Atomic Emission Spectrometry

By comparison with FAAS, ICPAES is more sensitive for the more difficult to atomize elements such as B, W and Si, but is not as sensitive as some of the easy to atomize elements such as Zn and Cd. ICPAES methods, however, are ideally suited for simultaneous, or rapid sequential, multielement analysis and are thus faster than FAAS for this purpose.

In general FAAS instruments will cost less than ICPAES equipment and are easier to use. For the determination of low concentrations of elements such as As, Se and Hg, vapor generation AAS is more sensitive than ICPAES methods.

10.3 Inductively Coupled Plasma Mass Spectrometry

In more recent times, ICPMS has offered a much faster means of determining extremely small traces of most elements, and provides real competition for GFAAS. However ICPMS requires very expensive equipment and very skilled operation.

10.4 Choice of Methods

As discussed immediately above, there is major competition for AAS as a means of analysis. Certainly, for the determination of very low concentrations of several elements in large numbers of samples, ICPMS is probably the method of choice by comparison with graphite furnace AAS. The very high cost of the equipment and the high operating costs, plus the skills requirement, however, will keep the former technique out of the reach of most small laboratories.

In a similar manner, but to a lesser extent, ICPAES instruments offer some advantages for some users but remain less attractive to the smaller laboratory. It seems unlikely that major developments will occur in AAS to alter this scene. Rather, the inevitable lowering of costs for ICPAES and ICPMS will make it increasingly hard for expensive types of AAS to survive. A versatile and major analytical service laboratory may certainly justify the installation of ICPMS, ICPAES and AAS with flame, vapor generation and furnace modes.

In such a situation, the desirability of having individual AAS instruments for flame and furnace operation should be considered for the sake of convenience and the avoidance of contamination of the clean environment needed for graphite furnace methods. With such a versatile combination, the advisability of having a Zeeman background-corrected furnace instrument might also be considered.

11 FUTURE DIRECTIONS

Because of the lower costs of equipment and the simplicity of operation, AAS appears to be likely to continue to be attractive to most analytical laboratories, particularly the smaller ones. It also seems, to the author, that low-cost atomic absorption modules, serving as specialized detectors, in association with sample preparation and handling techniques, such as those associated with flow injection methods, are one of the important directions in which development might occur. Such an approach would continue to use the highly specific capabilities of the atomic absorption method, which have made it so popular since its inception.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AC	Alternating Current

AES	Atomic Emission Spectrometry
DC	Direct Current
FAAS	Flame Atomic Absorption Spectrometry
FIA	Flow Injection Analysis
GFAAS	Graphite Furnace Atomization
HCL	Hollow-cathode Lamp
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
VGAAS	Vapor Generation Atomization

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Optical Emission Inductively Coupled Plasma in Environmental Analysis

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flow Injection Analysis Techniques in Atomic Spectroscopy • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry

NOTE

The number of references available in AAS is, by now, huge and the following represent only a judicious selection. There are many detailed text books available and those readers requiring a more detailed treatment of particular aspects are advised to consult any of these. One particular text book is, however, recommended as being very detailed in its blend of theoretical and practical details.[20]

REFERENCES

1. G. Kirchhoff, 'Über das Verhältniss zwischen dem Emissionsvermögen und dem Absorptionsvermögen der Körper für Wärme und Licht', *Annalen der Physik und Chemie Band CIX*, 275–301 (1860).
2. G. Kirchhoff, 'On the Relation between the Radiating and Absorbing Powers of Different Bodies for Light and Heat', *Phil. Mag.*, S4, No. 130, **20**, 1–21 (1860).
3. G. Kirchhoff, R. Bunsen, IX, 'Chemical Analysis by Spectrum-observations', *Phil. Mag.*, S4, No. 131, **20**, 89–109 (1860).
4. G. Kirchhoff, R. Bunsen, XLII, 'Chemical Analysis by Spectrum-observations – Second Memoir', *Phil. Mag.*, S4, No. 148, **22**, 329–349 (1861).
5. H. Lundegardh, *Die quantitative Spektralanalyse der Elemente*, Fischer, Jena; Part I, 1929 and Part II, 1934.
6. A. Walsh, 'The Application of Atomic Absorption Spectra to Chemical Analysis', *Spectrochim. Acta*, **7**, 108–117 (1955).
7. C.T.J. Alkemade, J.M.W. Milatz, 'A Double Beam Method of Spectral Selection with Flames', *Appl. Sci. Res.*, **B4**, 289–299 (1955).
8. B.J. Russell, J.P. Shelton, A. Walsh, 'An Atomic Absorption Spectrophotometer and its Application to the Analysis of Solutions', *Spectrochim. Acta*, **8**, 317–328 (1957).
9. W.G. Jones, A. Walsh, 'Hollow Cathode Discharges – the Construction and Characteristics of Sealed-off Tubes for Use as Spectroscopic Light Sources', *Spectrochim. Acta*, **16**, 249–254 (1960).
10. J.V. Sullivan, A. Walsh, 'High-intensity Hollow Cathode Lamps', *Spectrochim. Acta*, **21**, 721–726 (1965).
11. T.T. Woodson, 'A New Mercury Vapour Detector', *Rev. Sci. Instr.*, **10**, 308–311 (1939).
12. J.E. Allan, 'Atomic Absorption Spectrophotometry with Special Reference to the Determination of Magnesium', *Analyst*, **83**, 466–471 (1958).
13. B.M. Gatehouse, J.B. Willis, 'Performance of a Simple Atomic Absorption Spectrophotometer', *Spectrochim. Acta*, **17**, 710–718 (1961).
14. M.D. Amos, P.E. Thomas, 'The Determination of Aluminum in Aqueous Solution by Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **32**, 139–147 (1965).
15. M.D. Amos, J.B. Willis, 'Use of High-temperature Premixed Flames in Atomic Absorption Spectroscopy', *Spectrochim. Acta*, **22**, 1325–1343 (1966).
16. N.S. Poluektov, R.A. Vitkun, Yu.V. Zelyukova, 'Determination of Milligamma Amounts of Mercury by Atomic

Absorption in the Gaseous Phase', *Zh. Anal. Khim.*, **19**, 937–942 (1964).
17. W. Holak, 'Gas Sampling Technique for Arsenic Determination by Atomic Absorption Spectrophotometry', *Anal. Chem.*, **41**, 1712–1713 (1969).
18. B.V. L'vov, 'The Analytical Use of Atomic Absorption Spectra', *Spectrochim. Acta*, **17**, 761–770 (1961).
19. S.R. Koirtyohann, E.E. Pickett, 'Background Corrections in Long Path Atomic Absorption Spectrometry', *Anal. Chem.*, **37**, 601–603 (1965).
20. B. Welz, M. Sperling, *Atomic Absorption Spectrometry*, 3rd edition, Wiley-VCH, Weinheim, 1999.

Flow Injection Analysis Techniques in Atomic Spectroscopy

Rosario Pereiro and Alfredo Sanz-Medel
University of Oviedo, 33006 Oviedo, Spain

1 Introduction **2**

2 Basic Concepts and Operation Principles **3**
2.1 Basic Principles of Dispersion 3
2.2 Chemical Reactions in a Flow 4
2.3 Flow Injection and Atomic Spectrometry Combination 4

3 Instrumentation **5**
3.1 Elemental Components of the Flow System 5
3.2 Assembled Flow Systems 7
3.3 Interfaces between Flow Systems and Atomic Detectors 8

4 Flow Injection Analysis Strategies Involving One-phase Liquid Flow Systems **9**
4.1 Automated Variable Dilution 9
4.2 Automated Addition of Reagents to Improve Analytical Performance 10
4.3 Standardization and Implementation of the Standard Additions Method 10

5 Flow Injection Systems for On-line Solid–Liquid Separation and Preconcentration **10**
5.1 Sorption 11
5.2 Precipitation and Co-precipitation 12
5.3 Anodic Stripping Voltammetry 12

6 Flow Systems for On-line Gas–Liquid Separation and Preconcentration **13**
6.1 Flow Systems for the Formation of Volatile Hydrides of Metals and Metalloids 13
6.2 Alternative Strategies: Cold Vapors, Ethylderivatives and Halines 14
6.3 Approaches for Preconcentration in the Gas Phase 14

7 Flow Systems for On-line Liquid–Liquid Separation and Preconcentration and Tandem On-line Combinations **15**
7.1 Flow Systems for Liquid–Liquid Extraction 15
7.2 Integration of Two Continuous Separation Units: The Tandem On-line Concept 16

8 Flow Systems for Field Sampling and On-line Digestion of Samples **17**
8.1 Minicolumn Field Sampling and Flow Injection Analysis 17
8.2 On-line Photo-oxidation Flow Systems 17
8.3 On-line Microwave-assisted Digestion 18

9 Potential Applications Preview and Future Developments **19**

Abbreviations and Acronyms **20**

Related Articles **21**

References **21**

Flow injection analysis (FIA), developed originally for the automation of serial assays, has become a powerful tool most adequate for performing on-line any sample preparations before final measurement (e.g. sample dissolution, dilutions, matrix removal, preconcentration, etc.). It is not surprising that the combination of FIA with atomic spectrometric techniques has enlarged the analytical potential of atomic methods and expanded their wide field of applications. The collection of sample manipulation processes which can be covered today by flow operation procedures is amazing and therefore the general instrumentation required is reviewed in this contribution.

The description of the different flow strategies is carried out according to a hierarchy going from simple dilutions (including isotopic dilution of particular interest when using mass spectrometric detection), reagent mixing or standard additions, to more sophisticated flow manifolds such as those based on the use of two phases (e.g. gas–liquid, liquid–liquid or solid–liquid) for separation/preconcentration purposes. Modern approaches allowing for on-line decomposition/dissolution of solid samples (e.g. on-line chemical oxidation, photo-oxidation and microwave heating) are also described. Finally, manifolds for the in situ uptake of the sample and its complete on-line pretreatment are discussed.

The coupling of the above flow methodologies to a variety of atomization/excitation/ionization sources (flames, quartz tubes, graphite furnaces, inductively coupled plasmas (ICPs), microwave coupled plasmas, glow discharges (GDs), etc.) is detailed, in order to understand the usefulness of this combination both in atomic techniques based on photon measurements (absorption, emission and fluorescence spectrometry techniques) and on ions measurement mass spectrometry (MS).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The progress and implementation of new techniques for the automation of sample preparation has improved the performance of sample pretreatment methods in the everyday work of trace element analytical laboratories. The processing of a larger number of samples per unit time, the improvement of the precision by minimizing sample contamination and human errors, the small sample volumes required and the decreasing costs (attributed to the reduction of both human participation and consumption of reagents), are among the goals which can be accomplished by automating the sample pretreatment step.[1]

Automation can be achieved by one of the following three approaches, or by combination of two of them:[2] using robots which imitate the way that samples are manipulated by human operators, by discrete systems (batch analyzers) in which each sample preserves its integrity in a vessel mechanically transported (by a belt) to various zones of the analyzer (where the different sample pretreatment steps are carried out), and by flow manifolds in which samples are transported in flowing streams.

Flow systems based on the use of a continuous stream of liquid or, much less often, a gas have reached a special interest. There are two types of flow analyzers: (a) gas segmented flow analyzers (originally developed in the 1950s and first commercialized by Technicon under the name of ''AutoAnalyzer''), in which the flow is segmented by air bubbles aiming at preserving the integrity of samples, and (b) unsegmented flow analyzers, which can be classified according to whether samples are discretely injected FIA or continuously introduced into the system. Nowadays, the unsegmented flow analyzers are much more popular than the gas segmented flow ones.

The name of FIA was coined by Ruzicka in 1975[3] and he is generally credited as father of this technique, although Sarbeck et al.[4] published, back to 1972, a sample introduction approach for flame atomic absorption spectrometry (FAAS) which would be called today flow injection (FI) in its strict sense. The first reports on techniques explicitly defined as FI for atomic absorption spectrometry (AAS) were published in 1979.[5,6] Since then, the connection of FI to atomic spectrometric detectors has been extremely fruitful in terms of instrumentation versatility and variety of applications, giving rise to a great number of publications.[7–10]

FIA can be defined[11] as ''information-gathering from a concentration gradient formed from an injected, well-defined zone of fluid, dispersed into a continuous unsegmented stream of a carrier''. Since this definition does not facilitate visualization of the instrumentation required and the analytical interest of this versatile sample manipulation strategy, both aspects will be emphasized in the following paragraphs.

The basic equipment needed in FIA is relatively low cost, mechanically simple, robust and easily incorporated in automated systems. Figure 1 shows a general diagram of the most basic FI manifold, including the most commonly used instrumental units. The carrier solution and reagents are pumped continuously through a long, narrow tube, while a well-defined sample volume is introduced into the carrier stream (e.g. with the help of an injection valve). After the reactions occurring in the FIA system, the product of the reaction, which is monitored at the detector, gives rise to a transient response. Precise pumping and injection allows for identical treatment of every sample and standard in the FIA manifold, and this has the important consequence that there is no need to be restricted to reactions in which the reactants are stable. Also, precise timing is not necessary to obtain stable products nor to form them instantly. The use of FIA systems brings about several important advantages as compared to their manual batch analogs, both in terms of enhanced sample manipulation and of analytical performance (see Figure 2).

The sample manipulation processes collected in Figure 3, covered today by flow operation procedures extends over a wide scope and the instrumentation required for most of them will be reviewed in this contribution. Such flow strategies can be classified according to a hierarchy going from simple sample dilutions before final measurement to the in situ uptake of the sample and its complete on-line pretreatment.

Although processes like dilution (including isotopic dilution of particular interest when using mass spectrometric detection) and reagents mixing or standard

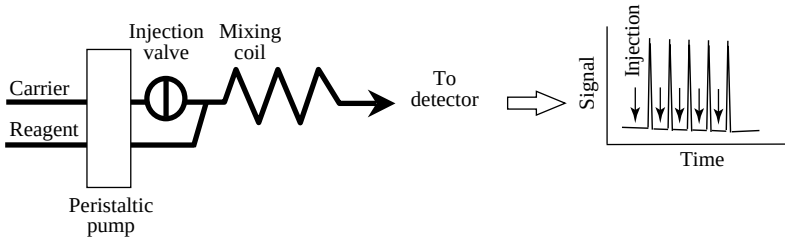

Figure 1 Schematic of a basic two-channel FI manifold including the most commonly used components.

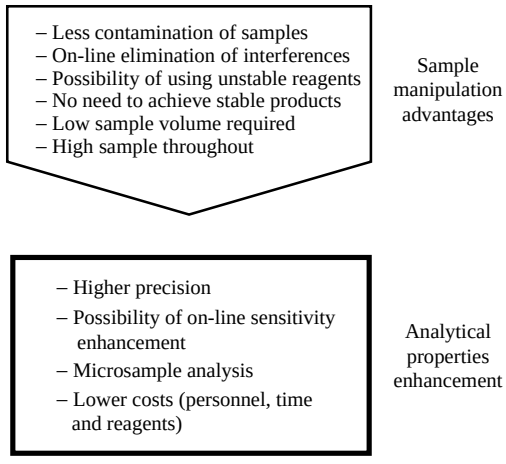

Figure 2 Analytical advantages of FIA systems in comparison with batch sample pretreatment procedures.

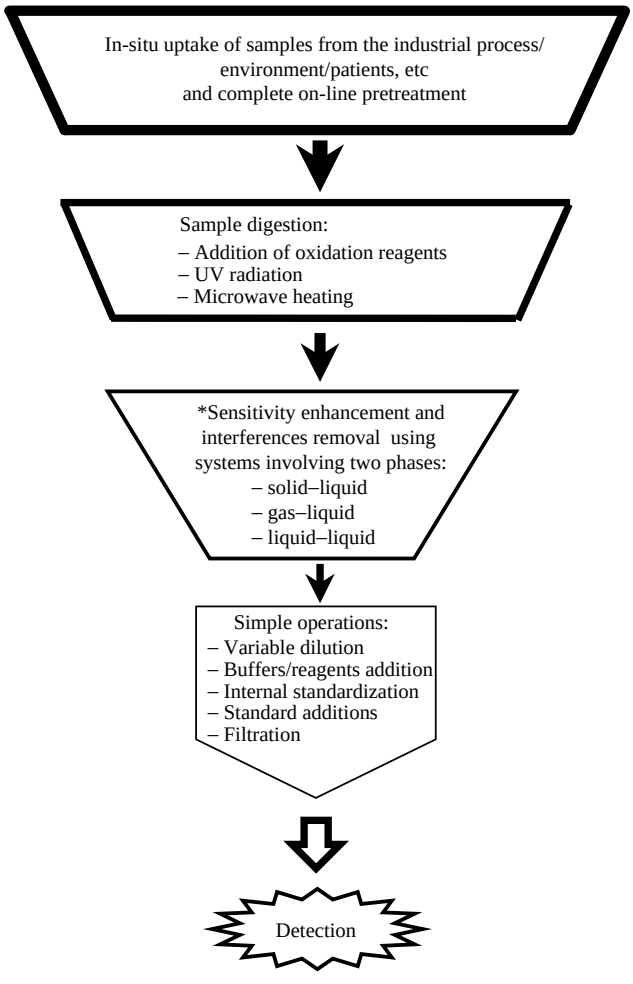

Figure 3 Scope of the sample pretreatment processes covered by flow procedures.

additions, can be considered as the simpler pretreatment operations which can be carried out in flowing systems, however, they are none the less important and nowadays robust on-line flow manifolds are routinely used to face these sample pretreatment stages.[12,13]

A most interesting advantage of flow systems derives from their ability to integrate non chromatographic separation/preconcentration techniques based on the use of two phases (e.g. gas–liquid, liquid–liquid or solid–liquid) with a continuous detector in a straightforward manner. The general aim of these on-line flow separation/preconcentration techniques is to provide two sample fractions: one would contain the analyte(s) enriched and free from potential matrix interferences, while the other fraction would contain the matrix. In some particular cases this fractionation can offer a simple means of speciation;[14,15] however, general speciation problems demanding information on several compounds or species of a given element require more powerful separation techniques (e.g. chromatographic or electrophoretic in nature). Finally, it should be highlighted that the use of simple two-phase flow separation systems allowed a dramatic increase of analytical reproducibility achievable by atomic spectrometry indirect determinations of organic pharmaceutical products, as compared to batch procedures.[16]

In a further extension of sample pretreatment processes enhanced by flow manifolds and keeping in mind that most atomic detectors are designed for the introduction of liquid samples, it is not surprising that a great deal of interest is attracted at present by the development of on-line flow systems allowing for the decomposition/dissolution of solid samples. Different approaches including on-line chemical oxidation,[17] photo-oxidation[18] and microwave heating[19] have been successfully exploited for these purposes.

Moreover, the in situ uptake of samples straightforward from their source and their on-line pretreament before atomic detection can be considered as the ideal to be pursued in terms of automation (see Figure 3). This goal has already been addressed and interesting manifolds (some already available commercially) have been proposed to carry out in situ environmental analysis, to wholly automate analytical measurements in the clinical laboratories (e.g. in vivo sample uptake of blood samples[20]) and for industrial process control.

2 BASIC CONCEPTS AND OPERATION PRINCIPLES

2.1 Basic Principles of Dispersion

The basis of FIA analyses relies on three aspects which should be kept controlled and reproducible: (i) volume of sample injected into the flow manifold, (ii) dispersion of the injected sample throughout the conduits of the system

(precise pumping is required), and (iii) timing for each sequence carried out in the flow system (i.e. mixing with other streams, time to reach the detector, etc.).

Precise pumping and injection leads to a precise flow movement in a given manifold, allowing for reproducible chemical and physical manipulations. This is the fundamental feature that makes FIA so useful as an analytical technique.[11] In FIA, every sample and standard is subjected to identical treatment and measured after the same time interval, therefore, there is no need to wait to complete reactions.

In FIA analysis the profile ''concentration versus time'' observed at the detector, for an injected sample with an analyte concentration C_o, is a peak. This peak is the result of the dispersion processes occurring to the sample in the flow. The ratio of the injected analyte concentration, C_o, to the instantaneous concentration corresponding to any point of the sample volume reaching the detector, C_g, is known as the dispersion coefficient, D_g. In many cases, the value of the dispersion coefficient at the peak maximum, D, is used as the parameter to characterize the extent of mixing in a given FIA system.

The main experimental variables increasing the D value include: increases of the length and diameter of the tube, increases of the average flow rate and detector volume, and increases of the corresponding molecular diffusion coefficient. On the other hand, the most common way to obtain decreased D values is to increase the sample volume injected (other methods include the use of packed tubes or the modification of the viscosity). In most FIA systems D values are higher than unity. However, whenever a preconcentration step has taken place, the resulting D values are lower than unity.

Laminar flows are obtained under the experimental conditions used in FIA systems, i.e. on average, the molecules on the flow follow stream lines parallel to the walls of the narrow bore tubing. The walls of the tubing produces a drag on the flow. Thus, a gradient between the center of the stream and the wall is formed (the plot of the velocity profile from wall to the tube center is a parabola) giving rise to concentration gradients.

Besides, the concentration gradients will produce the diffusion of analyte molecules during the movement inside the tube. This diffusion will be axial (along the length of the tube) and radial (across the tube). It is important to note that, in the case of radial diffusion, the sample molecules on the leading boundary which diffuse towards the walls enter slower moving stream lines while molecules of sample close to the walls diffuse towards the middle of the tube and thus moving faster (see Figure 4). As a consequence of these processes, an interdiffusion between sample and carrier takes place which dilutes the sample or, in the case that appropriate reagents

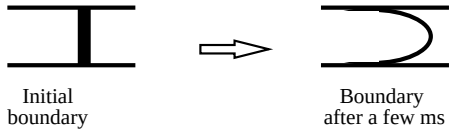

Figure 4 Effect of laminar flow and diffusion on the flow boundaries between sample and carrier after a few ms. Similar effects occur for the carrier and the sample.

are introduced, produces chemical reactions between components of sample and carrier.

In those cases requiring efficient mixing, other mixing mechanisms are added to this inherent interdiffusion process, such as the addition of other channels which merge with the sample at confluence points, the inclusion of mixing chambers, etc. Also, in some instances additional flow patterns to aid mixing are included, like bending or knotting the tube, packing a tube with glass beads, etc. Mixing of carrier streams in FIA procedures can be aimed at achieving on-line sample dilution, internal standardization, standard additions, chemical reactions, etc.

2.2 Chemical Reactions in a Flow

In many FIA procedures, a reaction between a component of the carrier stream and the analyte is needed to form a detectable product or when interference eliminations by chemical complexation are required. In both cases, a compromise between residence time and percentage of product formation should be chosen for maximum sensitivity: lower residence times would allow D decreases, however, product formation can be too low at short reaction times (in any case, it should be stressed again that, in many cases, chemical reactions do not need to be completed in FIA procedures).

In atomic spectrometry, on-line chemical reactions are increasingly exploited: (i) to form a volatile derivative of the analyte (e.g. hydride generation (HG)), (ii) to produce a hydrophobic analyte derivative able to be extracted on an organic solvent or on a hydrophobic solid surface, (iii) to fix and later release an analyte from an active solid support with chemical bonding groups, and (iv) to digest samples on-line (e.g. by chemical oxidation, with the aid of ultraviolet (UV) radiation, microwaves, etc.). There are also situations where the addition of a given reagent is required to promote a desirable reaction in the atomizer (e.g. the addition of lanthanum as releasing agent in FAAS).

2.3 Flow Injection and Atomic Spectrometry Combination

There are two basic modes for detection in atomic spectrometry: photons measurement: AAS, atomic emission

spectrometry (AES) or atomic fluorescence spectrometry (AFS) techniques, and ions measurement: MS techniques. A great variety of atomization/excitation/ionization sources is available to the analytical chemist in order to transform the analyte adequately for further measurement as ions or photons, including: flames, graphite furnaces and plasmas [e.g. ICP, microwave induced plasma (MIP), direct current plasma (DCP), stabilized capacitive plasma (SCP), GD, etc.].

Basic concepts and instrumentation of atomic spectrometry are beyond the scope of this contribution, but such concepts and general applications can be found in general text books as well as in monographs.[21–23] The most widely used instruments in atomic spectrometry are the FAAS, the graphite furnace atomic absorption spectrometry (GFAAS), also known as electrothermal atomic absorption spectrometry (ETAAS), and the inductively coupled plasma atomic emission spectrometry (ICPAES) while flame atomic fluorescence spectrometry (FAFS) is used to a lesser extent. Also, the continuously growing technique of inductively coupled plasma mass spectrometry (ICPMS) should be highlighted. Developments of all aspects of atomic spectrometry are still taking place exploiting FIA combinations with atomic spectroscopy and may be easily followed via the Atomic Spectrometry Updates review articles appearing in the *Journal of Analytical Atomic Spectrometry* of the Royal Society of Chemistry.

The combination of FIA strategies to atomic spectrometry has improved the analytical performance of such atomic instruments in terms of: (i) lower detection limits (through on-line preconcentration processes and matrix separations), (ii) greater accuracy (sample pretreatment is carried out in a highly reproducible fashion and contamination is reduced), (iii) high variety of samples can be directly processed (e.g. by on-line digestion) and, of course, (iv) higher sample throughput.

3 INSTRUMENTATION

3.1 Elemental Components of the Flow System

The three basic components of a flow manifold are the sample introduction system, the propulsion unit and the connecting tubes. Moreover, as was commented in the Introduction, special devices to allow for on-line separation/preconcentration/sample digestion, are becoming more and more popular inserted or integrated in the basic manifold.

3.1.1 Sample Introduction Systems

Two basic categories of sample introduction modes in FIA can be distinguished, namely volume-based and time-based. In the first case, samples are introduced in the FI manifold through an injection port, which consists of a septum or, most frequently, a valve. When time-based injection systems are used, the injected sample volume depends on the time period during which the sample is allowed to enter into the flow manifold.

In the early stages of development of FIA, the liquid samples were introduced with a syringe through a rubber septum.[24] However, important disadvantages of this sample introduction mode were distortion of the stream movement, lack of reproducibility in the manual injection and leakage after repeated injections on the same spot. The number of designs described to improve the performance of the septum has been considerable and different volume-based systems were proposed such as double proportional slider valves,[25] six[26] or eight[27] ports rotary injection valves, and hydrodynamic injectors (valveless systems),[28] as well as time-based injection systems including approaches based on two variable-speed computer-controlled peristaltic pumps and a commutator at the merging point[29] or timed solenoid injectors.[30]

In Figure 5(a) the schematics of a commercially available six-way rotary valve are shown. The sliding rotary valve has an external loop determining the sample volume to be injected. To introduce a different sample volume the external loop has to be changed. This type of valve is the most frequently used.

Figure 5(b) shows the principle of hydrodynamic injection:[28] the system is based on the introduction of the sample by a second propulsion system into a well defined volume of conduit of the main flow manifold and Figure 5(c) shows the schematics for a time-based/sample dilution system[29] allowing for the injection of a sample plug having an exactly predetermined concentration profile for the measured analyte.

Common insertion ports for the introduction of gaseous samples are gas-tight rotary valves and also the so-called 'exponential dilutors'[31,32] consisting of a magnetically stirred chamber of known volume, where the sample is diluted in the carrier gas. Injections are made through a septum by means of a gas-tight syringe. A set of two valves (or a three-way stopcock) permits the chamber to be isolated from the carrier gas flow during injections (see Figure 6). After the sample is stirred for some seconds in the isolated chamber, the valves are open giving rise to a signal peak with an exponential decay as a function of time.

3.1.2 Propulsion Systems

The propulsion system has to provide a continuous and reproducible flow rate of the solutions passing through

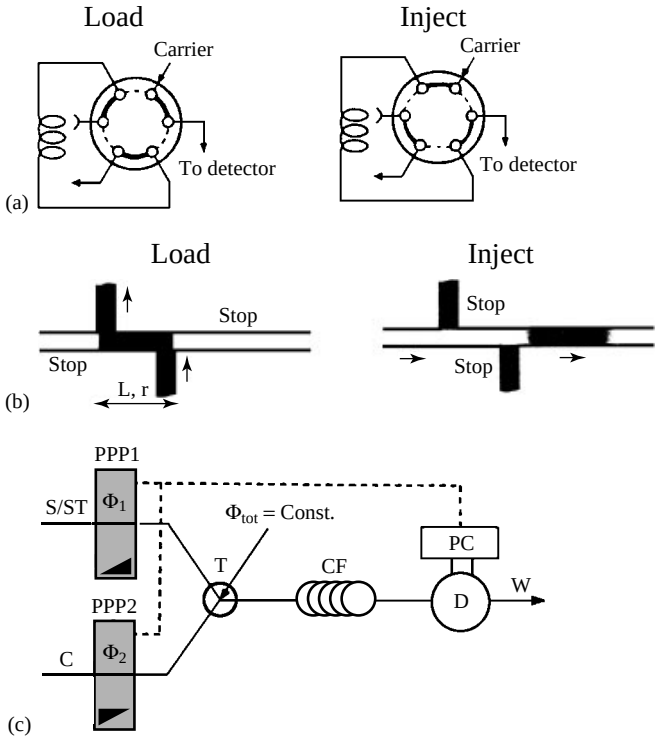

Figure 5 Examples of liquid sample introduction devices in FIA. (a) Six-way rotary valve. In the ''load'' position the loop is filled with the sample while in the 'injection' position the loop content is dragged through the manifold by the carrier solution. (b) Principle of hydrodynamic injection. A fixed volume of sample solution is metered into the conduit of length L and internal radius r (Load), and this volume is subsequently propelled downstream by the carrier (Injection). During the sampling cycle, the carrier stream circuit is stopped, and vice versa. (Reproduced with permission of Elsevier Science Ltd. from Ruzicka and Hansen[28] © 1983.) (c) Time-based/sample dilution system. S/ST: sample/standard; C: carrier; PPP1, PPP2: computer controllable peristaltic pumps; T: mixing point; CF: chemifold; D: detector; PC: personal computer; W: waste. (Reproduced with permission of Elsevier Science Ltd. from Novic et al.,[29] © 1999.)

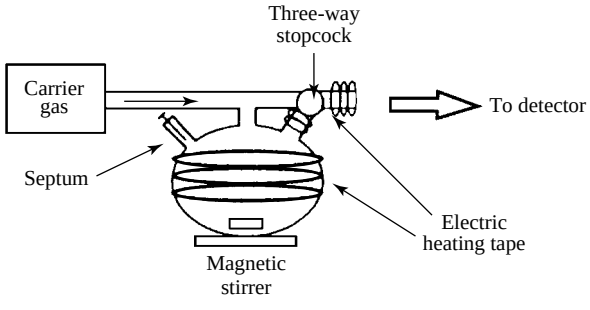

Figure 6 Schematic of an exponential dilutor for gas sample introduction.

the manifold. The ideal propulsion system should give a pulsed-free flow with an easily selected rate. Besides, in some cases, more than one solution has to be propelled and, therefore, multichannel capabilities are advisable.

Three main types of propulsion systems have been described:

1. Propulsion-less manifolds, which rely in the negative pressure generated by the nebulizer of an atomic detector to draw the liquid solution to the instrument.[33]
2. Gas-pressurized carrier reservoirs.[34] In this case, to keep a constant flow rate the pressure has to be maintained by means of a regulator. Although this system gives rise to a pulsed-free flow, drawbacks such as the consumption of gases or the difficulties to achieve constant flow rates have limited its use.
3. Pumps. Positive displacement pumps have been described as propulsion systems in flow manifolds; however, they introduce pulses in the analytical system and when two or more reagents are delivered in parallel these pressure pulses cause a non-uniform mixing of reagents reducing the accuracy and precision of the analysis. Although several depulsing systems have been described,[35] this type of pump does not enjoy great popularity. The peristaltic pump is by far the most used type of propulsion system. These pumps allow to use more than one channel and to achieve a constant and accurate flow rate, which can be easily modified according to the analyst needs. The pumping rate provided depends on the rotation rate and the inner diameter of the pumping tubes. Peristaltic pumps tend to give slight pulses which can be almost eliminated by an adequate adjustment of the clamps pressing the peristaltic tubes or by the use of coils located in the manifold right next to the propulsion unit.

3.1.3 Connecting Tubes

Inert and flexible tubing with inner diameters in the range from 0.3 to 1.5 mm allows the connection of the different units of the flow manifold. In some parts of the system, the tubing could be bent to allow for an efficient mixture of reagents and sample. Also, multiway connectors are used to mix flows from different channels.

3.1.4 Two-phase Separation Units

In this section, only a brief introduction to the basic instrumentation needed for gas–liquid, liquid–liquid and solid–liquid separations in a flowing system will be presented. More details will be given in sections 5, 6 and 7 of this chapter.

3.1.4.1 Gas–Liquid Separation The co-occurrence of a gas and a liquid phase in a hydrodynamic system

to achieve a two-phase separation gives rise to problems which call for ingenious technical solutions.[36–38] Gas liquid separators (GLSs) to be used in continuous separation systems should work smoothly and regularly to avoid lack of reproducibility and signal fluctuations and they should induce minimal dispersion or dilution of the gas phase (vaporized analyte). Many different designs have been evaluated as GLSs. The most common ones consist of a chamber, typically made of glass, in which a smooth separation between gases and liquids is aided by employing an inert purge gas. The use of gas permeable membranes, either flat[37] or concentric hollow cylinders[38] has been also described to separate both phases in flow systems. Examples of different GLSs used in flow systems are depicted in Figure 7.

3.1.4.2 Liquid–Liquid Separation A flow liquid–liquid extractor consists of a device able to merge two non miscible liquid phases to produce a single segmented flow, within which mass transfer takes place; the segmented flow passes through a phase separator, emerging from it as two separate streams, containing one phase each. The liquid–liquid phase separator must work efficiently and rapidly. There is a large variety of phase separators, based on two main principles, namely: density differences between the phases and selective wetting of appropriate inner surfaces. The most popular designs are based on the use of hydrophobic membranes which are only permeable to the organic solvent.

3.1.4.3 Solid–Liquid Separation Three main approaches have been described: sorption on a solid phase, generally consisting of a minicolumn, inserted in the flowing system,[26,39] precipitation and co-precipitation, based on the combination in a flow of three processes: precipitation of the analyte, its filtration and final dissolution,[40] and anodic stripping, consisting of the use of flow electrolytic cells to eliminate nondepositing sample matrix components and to preconcentrate analytes which are deposited at the working electrode.[15]

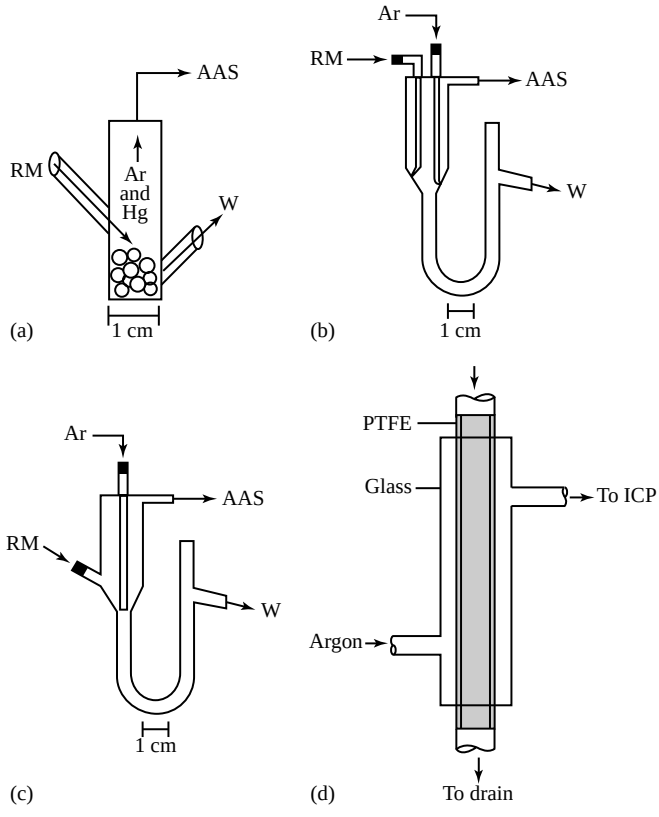

Figure 7 Examples of gas–liquid separators used in flow systems (not drawn at the same scale). RM: reaction mixture, W: waste. (a) Perkin-Elmer GLS; (b) PS Analytical GLS for sodium tetrahydroborate reductions; (c) PS Analytical GLS for tin chloride reductions; (d) Tubular membrane GLS. (Parts (a), (b) and (c) reproduced from Hanna et al.,[36] © 1993; part (d) reproduced from Barnes and Wang,[38] © 1988, with permission from the Royal Society of Chemistry.)

3.2 Assembled Flow Systems

Several companies are marketing sample introduction flow systems allowing for automated preparation of the most varied sample matrices. Representative examples of manifolds allowing for several sample pretreatment operations are reviewed below and in the next sections. However, it is important to emphasize that the modular character of these manifolds allows to adapt them to almost every particular sample pretreatment needed.

Figure 8 presents several diagrams of different configurations of basic flow manifolds. The addition of a second channel to the system depicted in Figure 1 allows some manipulations or pretreatment of the sample (Figure 8a) such as dilution and addition of reagents or masking agents, while the set-up depicted in Figure 8(b), called ''merging zones'', permits important savings in reagent consumption[41] and is also of interest for on-line dilution or standard additions.

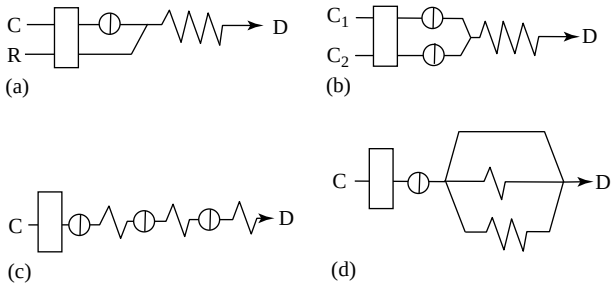

Figure 8 Examples of configurations of basic systems. C: carrier; R: reagent; D: detector. (a) Two line; (b) merging zones; (c) sequential injection flow system; (d) three-branch on-line sample dilution system.

Figure 8(c) shows a manifold to inject simultaneously several plugs of a given solution into the same carrier stream to achieve overlapped zones as a consequence of dispersion and thus permitting the analysis of a wide range of analyte concentrations[42] and Figure 8(d) collects the schematics of a system to obtain a variety of dilution factors for a single injection without the need for controlled timing of any operation.[43] The manifold consists of a three-branch network giving three partially overlapping peaks and five measurement points, three maxima and two minima.

Detailed instrumentation for separation/preconcentration manifolds including a solid–liquid, liquid–liquid or gas–liquid phases separation units as well as tandem on-line combinations of more than one separation strategy is thoroughly reviewed in sections 5 to 7 and the reader is addressed to those sections for specific information.

The possibility of carrying out decomposition/dissolution of solid samples on-line with the detector is attracting also a great deal of interest. Besides, in order to achieve volatile species generation from different analyte compounds, some organospecies need first to be decomposed, being this approach particularly interesting in metal speciation for species decomposition at the interface between the exit of a liquid chromatographic column and the atomic detector. Different approaches for on-line digestion of samples will be illustrated in section 8 in more detail.

3.3 Interfaces between Flow Systems and Atomic Detectors

The advantages of using flow manifolds as sample introduction systems for atomic detectors have been demonstrated for a variety of techniques including different atomic sources and/or detectors characterized by continuous operation (e.g. FAAS; ICPAES; MIP, SCP, and GD combined with AES; ICPMS, etc.) and even interesting approaches have been proposed for GFAAS, notwithstanding its discontinuous operation.

3.3.1 On-line Coupling to Continuous Atomic Spectrometric Detectors

The interfacing of flow manifolds to continuous atomic spectrometric detectors for the introduction of liquid samples requires usually a nebulizer and a spray chamber to produce a well-defined reproducible aerosol whose small droplets are sent to the high temperature atomic cell.

In FAAS, the concentric pneumatic nebulizer is the most frequently used.[44,45] Appleton and Tyson have given a theoretical modeling approach for nebulizer behavior in FIA coupled with FAAS.[44] The suitability of the response kinetics of the FAAS as well as the contribution of individual components to the overall dispersion of an injected sample has been investigated by Fang et al.,[45] being observed that the dispersion effects in the nebulizer – burner system generally showed a very limited contribution to the dispersion of the injected samples.

The following guidelines for optimum performance in terms of sensitivity and precision when using a barrel nebulizer and FAAS detection were given by Brown and Ruzicka:[46] (a) the flow rate of the carrier pumped into the nebulizer, should always be greater than the natural aspiration rate of the nebulizer; (b) the sensitivity is a function of the flow-rate of the carrier stream entering the nebulizer and the aspiration rate of the latter; (c) to achieve an optimum flow through the nebulizer, an additional stream may be added to augment the flow of the carrier without a significant decrease in sensitivity due to dilution.

Concerning plasma-based atomic sources, the more commonly employed is, by far, the ICP which was first used for AES and, more recently, for MS detection. A variety of nebulizers have been described for ICP work[47] including cross-flow pneumatic nebulizers, pneumatic concentric nebulizers, Babington pneumatic nebulizers, ultrasonic nebulizers, microwave thermospray nebulizers, jet impact nebulizers, electrospray, microconcentric nebulizers, oscillating capillary nebulizers, direct injection nebulizers, etc.

In the 1980s much reasearch work was devoted to the search of the ''ideal'' nebulizer, most of these designs also being tested as FIA interfaces to the plasma. It should be perhaps highlighted the direct injection nebulizer, ideally 100% efficient at transporting the introduced analyte into the ICP[48] (the relative detection limits found for 30 μL sample injection volumes were generally comparable to those obtained for continuous sample introduction into a conventional cross-flow nebulizer), the microconcentric nebulizer, inserted directly into the tip of a conventional sample introduction tube on an ICP torch to allow for a low dead volume interface,[49] or the thermospray interface allowing for 10 fold better FIA detection limits as compared to FIA cross flow.[50]

Probably the most frequently used spray chamber in ICP is the Scott type. However, a new type of spray chamber combining gravitational, centrifugal, turbulent and impact loss mechanisms in one apparatus to remove large droplets, increase transport efficiency and reduce memory effects was designed and evaluated by Wu and Hieftje:[51] this chamber has at least 30% higher sample utilization efficiency, 2–3 times shorter sample clean-out time, half the cost, and simpler construction than the Scott type; moreover, it offers better detection limits and precision and when it was investigated for FI combined with plasma AES, the same sensitivity of

detection was achieved by continuous nebulization than injecting volumes of 200 μL.

Considering alternative plasma sources to the ICP, the low-power MIP should be highlighted. This excitation source is inexpensive and simple to operate, but its major limitation is the low tolerance to liquid samples and susceptibility to interferences caused by easily ionized elements. However, a MIP source called microwave plasma torch (MPT), unlike more conventional MIP supporting torch/cavities structures, tolerates aqueous aerosols and molecular gases introduced into the discharge. A study exploring the potential of the FI mode for sample presentation to the MPT/AES with an ultrasonic nebulizer[52] allowed the reduction of memory effects of this nebulizer without loss of sensitivity or precision. Furthermore, by appropriate choice of sample dispersion, a significant reduction of the Na and K interferences were observed.

The introduction of analytes as gaseous derivatives offers special advantages in terms of sensitivity in atomic detection since 100% of the volatilized analyte goes into the atomic detector as compared to 1–5% of analyte introduced by liquid nebulization. The earliest flow volatile analyte generation systems connected on-line to ICPAES and AAS detector systems for the analysis of some metals and metalloids were reported in 1978[53] and 1982,[54] respectively, and they have been also coupled to other atomic excitation sources such as the MIP[55] and GD[56] for AES detection. Besides, AAS and AES, AFS[57] and MS[58] detection has been successfully employed as detectors for these manifolds. Using quadrupole-MS detectors, the elimination of matrix interferences in some particular applications is required and flow systems based on the formation of a gaseous analyte derivative can overcome these problems: for example, selenium is one of the most difficult elements to be determined sensitively by ICP combined with quadrupole MS because of isobaric interferences on its most abundant isotopes. However, using on-line flow HG coupled to ICPMS the sensitive analysis of Se is easily carried out.[58]

Plasmas such as MIPs, SCPs and GDs using He as plasma gas offer good sensitivity for non-metals (e.g. halogens) provided that the samples are introduced as a gas phase. The determination of halides with these plasmas and AES detection, by on-line previous chemical oxidation to the corresponding volatile haline, has been successfully investigated.[59,60]

3.3.2 Semi On-line Coupling to Graphite Furnace Atomic Absorption Spectrometry

The intrinsic discontinuous nature of the GFAAS technique has limited the interest of interfacing basic continuous flow manifolds to this detector. However, several flow approaches which will be further detailed in the next sections, offer some capabilities favoring their exploitation as important aids in GFAAS work. These approaches include:

- Separation and preconcentration by on-line column sorption, coprecipitation and solvent extraction, as a consequence of the compatibility of organic solvents with GFAAS.[39,40]
- Formation of volatile derivatives of the analyte and their preconcentration on a graphite tube (''in situ trapping'').[61,62]
- Slurry sampling: although solid samples may be placed directly in the graphite furnace, the coupling of slurry sample introduction systems to GFAAS is proving to be very promising and versatile.[63]

4 FLOW INJECTION ANALYSIS STRATEGIES INVOLVING ONE-PHASE LIQUID FLOW SYSTEMS

4.1 Automated Variable Dilution

Most FI procedures give rise to an inherent dilution of the sample, i.e. dispersion coefficient higher than one. This dilution can be particularly notorious if additional channels (and the corresponding mixing points) or strategies such as mixing chambers are introduced into the manifold.

There are a large number of possibilities for the use of FI techniques to control dilution of stock solutions or samples. According to Fang,[64] FIA dilution systems are based on two basic mechanisms: sample dispersion (the measurement of the dispersion coefficient at different points of the peak profile is equivalent to the measurement of different dilutions of the injected sample) and flow manipulations (e.g. merging-flow, split-flow, etc.). Examples of strategies proposed in atomic spectrometry for on-line dilution, very useful also for standardization using a single standard, are summarized below:[65]

- Variation of sample volume: changing the sample loop, a linear dilution of the standard solution injected can be obtained.
- Variation of the mixing coil dimensions: higher dispersion of the sample bolus, and therefore higher dilution of the sample, can be achieved by increasing the length and the inner diameter of the mixing coils.
- Merging flows: merging the sample carrier with another carrier provides a linear dilution of the sample which depends on the sample injected volume, the manifold volume and the ratio between the two mixing flow-rates.

- Time-based electronic dilution: consists of timing the readout delay following sample dispersion. This procedure is the basis of gradient techniques for standardization with a single standard.
- ''Partial overlapping zones'' strategy: based on multiple injections by using multiple loop valves or network manifolds, to obtain partial overlaps of sequentially injected sample zones which penetrate in each other. It offers a series of minima between neighboring zones where the injected samples are diluted to different degrees by the carrier.
- Mixing chambers: the exponential concentration gradient obtained by the dispersion of a discrete volume of a concentrated standard through a well-stirred mixing chamber provides a high dilution capacity.
- ''Zone sampling'' procedure: a portion of the dispersed sample is re-sampled from a section of the gradient and dispersed again.

4.2 Automated Addition of Reagents to Improve Analytical Performance

Addition of special reagents to the sample to improve the analytical performance characteristics of atomic spectrometry procedures could be classified in two groups:

1. Strategies for one phase flow manifolds, such as the use of releasing agents or ionization suppressors to overcome interferences in flames or plasmas, the method of isotopic dilution in ICPMS to improve precision and accuracy, the addition of stabilizers or the formation of slurries and the on-line digestion of samples.
2. Strategies involving two phase flow manifolds, needed for on-line chemical reactions to form a gaseous or a non water soluble analyte derivative or to retain/elute the analyte from a solid phase.

The measurement of isotope ratios in ICPMS can compensate for matrix effects and instrumental drift. In the isotopic dilution method, the isotopic ratio in a sample aliquot is modified by adding a spike enriched in one of the isotopes of the element to be measured. From the measurement of the relative abundance, R, of each one of the two considered isotopes, A and B, in the unspiked and the spiked sample, the concentration of the analyte can be calculated according to the formula:

$$C = \frac{(A_{sp} - R\,B_{sp})W_{sp}}{(R\,B_s - A_s)W_s}$$

being C the analyte concentration, A_s and B_s the mass percentage of A and B in the sample, A_{sp} and B_{sp} the mass percentage of A and B in the spike, W_s the weight of sample and W_{sp} the weight of the spike.

The on-line isotope dilution[66,67] can be carried out by merging two carrier streams or by the simultaneous injection of two solutions into the carrier stream. Advantages of the use of the isotopic dilution procedures using an on-line flow mode include: (i) only the sample volume consumed for measurement is mixed with the spike (the rest remaining intact), (ii) ease of selection of the proper sample to spike ratio in unknown sample concentrations, and (iii) possibility to achieve simultaneously on-line sample dilution and on-line isotope dilution.

The on-line flow addition of components such as anticoagulants in whole blood analysis, acidic solutions to form slurries, chemical modifiers for ETAAS analyses such as the mixture 0.1 mol L^{-1} $Mg(NO_3)_2$ + 0.01 mol L^{-1} $Pd(NO_3)_2$[20] has also been described. Samples, in acidic media, can be on-line digested using microwaves. The degree of dissolution in the microwave cavity is mainly governed by three variables: microwave power, length and diameter of the tubing inside the microwave oven, and the acid slurry strength. This methodology has been already applied to a great variety of samples of clinical, biological and environmental interest, using FAAS, GFAAS, FAFS, ICPAES or ICPMS detection modes. Details about on-line microwave assisted digestion are given in section 8.3.

4.3 Standardization and Implementation of the Standard Additions Method

Besides the traditional use of a series of standards with increasing concentrations of the analyte, in FIA systems external calibration can be also carried out employing a single standard by following any of the on-line dilution strategies already described under the ''automated variable dilution'' section.

The standard additions method to correct for matrix interferences and the internal standard method to alleviate instrumental drifts can be also incorporated via FIA systems, giving rise to automated methods where the amount of solutions employed and the time and cost of the analyses are reduced. Procedures based on merging zones, partial overlapping zones, zone sampling, etc. can be readily implemented for these purposes.[65]

5 FLOW INJECTION SYSTEMS FOR ON-LINE SOLID–LIQUID SEPARATION AND PRECONCENTRATION

Nowadays on-line solid–liquid extraction procedures, and in particular, methods based on sorption of the

analyte in columns and its subsequent release, are experiencing a great popularity for sample pretreatment because they can be coupled in a most straightforward manner to most atomic detectors. Simplified handling, reduced solvent consumption and increased preconcentration factors are among the favorable features of solid phase extraction procedures. Of course, the elimination of matrix interferences (e.g. alkaline and earth-alkaline elements in high salt content samples), is also another advantage of these couplings, which is important for ICP analysis (for example seawater, dialysis fluids, etc.) or in the frequent cases of eliminating polyatomic interferences using the popular quadrupole ICPMS.[58]

FI on-line solid–liquid separation and preconcentration methods for atomic spectrometry may be classified into three main groups, according to the separation principle used for retention of the analyte, namely: sorption, precipitation/coprecipitation and anodic stripping voltammetry (ASV).

5.1 Sorption

Many sorbents have been developed so far and some of them are marketed as prepacked minicolumns. The characteristic features of the on-line preconcentration flow systems demand some properties of the packing material which may be only of minor importance in batch or traditional column procedures, since for flowing systems the column material has to be reusable, the kinetic processes or reactions have to be rapid and the flow uniform. Therefore, the special requirements may at least include: high mechanical resistance (in order to withstand high flow rates through the column and to maintain long column lifetimes), good kinetic properties (analytes have to be rapidly retained and readily eluted) and low degree of swelling and shrinking when being transformed from one form to another or with change of solvent conditions (the swelling/shrinking of some resins will give rise to backpressure and non uniform flow patterns).

Sorption preconcentration methods can be divided into two general groups: (a) analyte ions are collected directly by immobilized counter ions (ion-exchange) or immobilized chelating functions (sorption of ions), and (b) metal ions are adsorbed on suitable phases (e.g. activated carbon, octadecyl functional groups bonded silica gel, etc.) as hydrophobic metal chelates, previously formed by reactions in solution (sorption of metal chelates). The eluents more commonly used in the first case are diluted acids, bases or complexing agents, while for the second, the chelates are most frequently eluted with hydroorganic solvents.

The sorption of ions involves the use of ''solid active phases'' containing a suitable ion exchanger, a chemical reagent or even bioorganisms, immobilized on the solid support packed in a minicolumn. The solid active phases can be purchased (e.g. Chelex 100, Muromac A1, EDTrA-cellulose™, 8-hydroxyquinoline (8Q) bound to controlled pore glass (CPG)-8Q, 8Q bound to cellulose, etc.) or prepared at the working laboratory. The most common chemical mechanisms for reagent immobilization on the solid support to prepare the solid active phase are: adsorption (frequently used for hydrophobic reagents such as dodecanylquinolin-8-ol which is effectively fixed to nonionic supports), ion-exchange (the reagent is bound to the support by ionic groups which are not involved in the reaction with the analyte, such as the binding of 5-sulfonic oxine to resins with ammonium quaternary groups) and covalent binding (the reagent is covalently bound to solid supports such as controlled pore glass, silica gel, cellulose, etc.).

Chelex 100 was the first packing material used for on-line column preconcentration[68] and it has been widely used since then. This resin contains paired iminodiacetate ions coupled to a styrene divinylbenzene support. Despite its success, Chelex 100 does not fully meet the requirements mentioned at the beginning of the section for an ideal packing since the resin is subjected to some troublesome swelling properties: the particle size depends very much upon the counterion. These problems have been successfully overcome by the use of countercurrent flow elution. The Japanese product Muromac A1 which has also iminodiacetic acid functional groups was reported to be free from the troublesome swelling properties of the latter and has been used in a number of FAAS and ICPAES applications;[69,70] however, the material does not seem to be widely available and it was used mostly by Japanese workers.

The controlled pore glass-oxine is other frequently column packing material used in on-line preconcentration.[71] Its popularity can be probably attributed to its good mechanical and good kinetic properties derived by the immobilization of the oxine functional groups on easily accessible porous glass surface. Fibrous cellulose with EDTA groups (EDTrA-cellulose™)[72] and oxine immobilized on CM-cellulose (cellulose-8Q),[73] are also commercially available. Both materials show high capacity, high exchange kinetics and good mechanical and chemical stability.

Also, it is interesting that a minicolumn made of sulfhydryl cotton has been proposed for the speciation of inorganic mercury and methylmercury:[74] at pH 3–4 only the organomercury species are claimed to be retained and not the inorganic mercury. On the other hand, several applications of activated alumina in continuous flow-systems have been described both in the acidic form for the preconcentration of oxyanions and analysis by ICPAES[75] or in the basic form. In the latter case, the

alumina has been applied, for example, for the separation of Cr(III) from Cr(VI).[76]

The ability of bioorganisms such as algae to accumulate trace metals by biosorption has been known for some years. This effect was used in water treatment systems, however, only recently it has been exploited for analytical measurement methods. In 1991 the first application was described of columns with immobilized algae on-line coupled to atomic detectors for trace metal preconcentration being analytes eluted with diluted acids.[77] Later, other applications such as the speciation of Cr(III) and Cr(VI) using an on-line column with immobilized algae has been also described.[78]

The sorption of metal chelates opens also interesting possibilities.[39,79–81] In this case the chelate is on-line formed in the flow system and then retained on a hydrophobic adsorbent, being later eluted with a hydroorganic solvent. The most commonly used reagents are dithio-carbamate derivatives, possessing very active sulfur binding sites. The most popular detectors for these applications seem to be FAAS and GFAAS (more compatible with hydroorganic solvents than plasmas).

Whenever the difference in the sorbent selectivity between the free chelating reagent and the formed metal chelate is small, the sorbent will absorb also free ligand (in addition to the metal chelate) during preconcentration. Once the capacity of the sorbent has been overcome, any remaining metal chelate still passing through the column will not be adsorbed. To avoid this problem, it has been proposed as strategy sorbing first a small amount of the chelating agent and then the analyte is preconcentrated by direct complexation with the solid sorbent agent.[82] In this case, since both the chelate and the reagent are eluted, the column has to be fed with the reagent for each analysis. A comparative view of general advantages and disadvantages of metal chelates sorption, as compared to the sorption of ions, is presented in Figure 9.

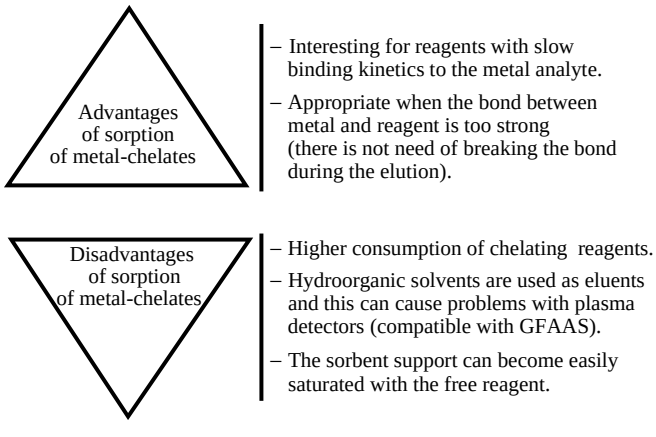

Figure 9 Comparison of "sorption of ions" and "sorption of metal chelates" procedures.

5.2 Precipitation and Co-precipitation

Continuous precipitation – filtration – dissolution processes can be described as continuous liquid–solid systems in which the second phase is first generated and then disappears in situ. The earliest flow on-line precipitation system with FAAS detection was reported by Valcarcel's group[83] employing on-line stainless filters to retain the precipitate. Usually, the precipitate is formed in the reaction coil, retained in the filter and subsequently dissolved by a suitable dissolution agent.

In 1991 Fang et al. presented an interesting approach based in analyte coprecipitation in the flow system. The analyte was coprecipitated quantitatively with the iron(II) – hexamethyleneammonium hexamethylenedithiocarbamate complex and collected in a "knotted reactor" made of ethyl vinyl acetate tubing. The precipitate was dissolved in isobutyl methyl ketone (IBMK) and introduced directly in the nebulizer of an FAAS.[84] Similar approaches have been exploited successfully using this or other collectors in combination with FAAS and GFAAS.[40,85] In comparison with on-line precipitation methods, coprecipitation procedures are less demanding on the solubility of the precipitate formed.

It is important to highlight how chemistry in precipitation and coprecipitation systems has some resemblance to metal–chelate sorption (the filter replacing the sorbent minicolumn), except that for precipitation/coprecipitation no column capacity limitations exist. Conversely, practical problems associated with continuous precipitation/coprecipitation systems can be encountered during the running of such manifolds.

5.3 Anodic Stripping Voltammetry

The use of flow-cells based on the principle of ASV coupled on-line to atomic detectors has also been investigated to obtain improved analysis of selected species (those which can be deposited at a working electrode and later released for atomic specific detection). These cells eliminate sample matrix components that are not electroactive and do not deposit during passage of the sample through the cell. Moreover, signal enhancement of analytes, as they may be preconcentrated at the electrode[86] are obtained. Besides, in some cases ASV can allow for oxidation state speciation of elements: for example As(III) and Se(IV) are electrochemically responsive in most electrolytes while As(V) and Se(VI) are not.[15] Thus, on-line ASV has allowed the speciation of both analytes in urine using ICPMS detection. Similar approaches have been also proposed for the speciation of Cr(VI) and V(V).[87]

6 FLOW SYSTEMS FOR ON-LINE GAS–LIQUID SEPARATION AND PRECONCENTRATION

For some elements it is possible to increase the sensitivity of their atomic spectrometric determinations by generating a volatile derivative of the analyte outside the instrument and transporting it most efficiently to the atomizer.[88] Strategies for such chemical vapor generation (CVG) include mainly: HG, cold vapor generation (CV), formation of alkyl derivatives and formation of halines.

Despite their advantages in atomic spectrometry, sample introduction methods based on classical batch analyte vapor formation procedures have a number of pitfalls, resulting in their replacement by flow-based methods. Their relatively modest throughputs, their bad precision (e.g. poor control over reaction conditions such as time), their lack of applicability to non stable analyte vapor derivatives, the use of quite large volumes of sample, and the generation of by-products, particularly hydrogen (when released as a sudden burst, it can affect negatively the flame or plasma) are detrimental to their analytical performance. Most of these problems can be eliminated, or at least reduced, by the use of flow procedures allowing for on-line vapor generation and gas–liquid separation. It is worth noting that such flow systems require reactions exhibiting high analyte volatilization kinetics (e.g. fast chemical reactions) and efficient mass transfer between both phases.

6.1 Flow Systems for the Formation of Volatile Hydrides of Metals and Metalloids

Chemical formation on a FIA system of covalent volatile hydrides of the analyte, has been widely applied to the atomic spectroscopic determinations of elements such as As, Bi, Ge, Pb, Sb, Se, Sn and Te.[89] The earliest continuous HG systems connected on-line to ICPAES and AAS were reported by Thompson et al.[53] and by Astrom,[54] respectively. In flow hydride systems, the sample (introduced either continuously or in the FIA mode) is on-line mixed with a reducing agent. In some cases additional channels are included for the introduction of other reagents (e.g. for on-line prereductions). The volatile species formed are separated from the liquid in a GLS with the aid of a continuous flow of a stripping gas (see Figure 10). The stripping gas and the volatile species (analyte derivatives plus by-products continuously produced such as H_2, CO_2 or water vapor) are continuously swept to the detector, while the liquid is driven to waste. Argon, at constant flow rate, is most frequently used as the stripping gas (examples of designs of GLSs are shown in Figure 7).

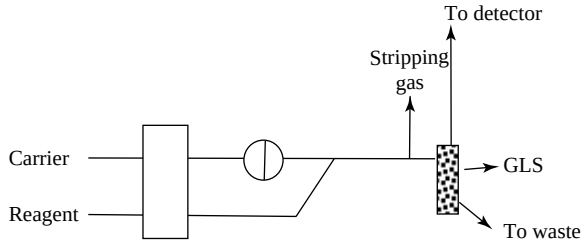

Figure 10 Basic FIA manifold to achieve gas–liquid separation.

On-line flow hydride procedures have been coupled to AAS,[54] AES,[53,55,56] AFS[57] and MS[58] specific detectors. For AES, the coupling of these manifolds has even been described to MIP[55] or GD;[56] however, the most commonly used excitation source is, by far, the ICP.[53] The use of this source in combination with mass spectrometric detection gives rise to extremely sensitive procedures due to the MS detector intrinsic sensitivity as compared to AES. However, on a routine basis, the more common detection systems used for these applications are AAS and AFS, where the analyte is swept to a heated quartz cell located on the optical path. Of course, it should be stressed that AAS and AFS have much less running costs than the ICP and they offer frequently excellent detection limits.

The most frequently used chemical reduction agent is tetrahydroborate. For As and Sb determinations both trivalent and pentavalent forms of analytes can yield hydrides but prereduction is highly desirable since the inorganic species As(III) and Sb(III) provide better sensitivity. Reagents such as L-cysteine or iodide have been used for on-line prereduction. For selenium, the best results are obtained by on-line compounds decomposition transforming the analyte species in its selenite state, and in this case it is frequently resorted to the use of 'strong' prereduction steps, which better correspond to section 8, such as the use of microwaves in $KBrO_3$/HBr medium. Unlike the situation for the hydride-forming elements of groups 14 and 15 of the periodic table, lead hydride is formed from a precursor having the metal in its highest oxidation state, Pb(IV). Among the different reaction conditions investigated for the best generation of plumbane, the greatest sensitivity, using FI with HG and AAS with quartz tube or graphite furnace atomization, was obtained in a hexacyanoferrate (III)/hydrochloric acid medium.[27,28]

Generation of volatile species for atomic spectrometry can also be facilitated by resorting to surfactant based ''organized media'', such as micelles and vesicles.[90] Analytical sensitivity and selectivity achieved with HG/atomic detection systems could be improved by the use of surfactants for the generation of the volatile species. In

fact, surfactant aggregates can concentrate reactants at a molecular level and, therefore, change thermodynamic and kinetic reaction constants. Moreover, surfactants can solubilize, in a selective manner, analytes and reactants in the self-assembled ''aggregates''. As a result, the special microenvironment existing in or on these aggregates may change the reactions (and so the observed interferences) in aqueous media.

6.2 Alternative Strategies: Cold Vapors, Ethylderivatives and Halines

The capability of mercury to form a monoatomic vapor at room temperature has been widely exploited in atomic spectrometry for its sensitive determination in the most varied samples. The reduction of mercury ions to Hg^o CV was first utilized in 1968[91] to develop a CV with AAS ''batch'' analytical method for mercury determination. In recent years, most of the Hg analyses by CV technique are carried out by flow systems. Usually AAS detection,[36,37] but also AES[92] or MS[93] are being employed for detection. Tetrahydroborate or tin(II) chloride are common reduction agents. The use of tin(II) chloride avoids the massive production of molecular hydrogen, which often deteriorates the performance of the spectrochemical plasma source. Conversely, it is important to point out that $SnCl_2$ is not effective enough to reduce organomercury species. Therefore such organomercury species should be first oxidized to the inorganic form, preferably by resorting to a continuous on-line flow system.[17,19]

Mercury seemed to be the only metal proven to be able to form monoatomic vapor at room temperature. This property has been widely used to develop CV-atomic spectrometry techniques for very low levels of mercury as shown before. In 1995 evidence was shown[94] that cadmium is also able to produce 'cold atomic vapor' in appropriate conditions of reduction (e.g. surfactant-based vesicles). It appears that volatile cadmium hydride can be produced in the presence of several media and catalysts.[90] In this form, the metal can be transported to an absorption measurement cell where it dissociates spontaneously to atomic vapor of cadmium. More than 20 times improvements in Cd detectability have been demonstrated in this way by CV using flameless quartz tube with AAS,[94] in situ preconcentration and GFAAS,[62] CV with ICPAES[95] and recently by CV with ICPMS.[96]

Ethylation as a means of CVG is a promising (yet not fully explored) alternative to common HG and CV techniques.[88,97] The formation of volatile derivatives with sodium tetraethylborate is particularly interesting to enhance the detectability of some nonhydride generating elements (e.g. Zn, Cu, etc.) able to form ethylderivatives. Moreover, this type of derivatization is very useful to volatilize organospecies of Hg, Pg, Se, Sn, etc. (e.g. for speciation purposes), and is also of interest in the determinations of some hydride[98] forming elements which can be considered as 'problematic' such as Sn (high reagent blanks), Pb (low yield of HG generation) or Cd (unstable volatile derivatives). Recent work has also proposed the use of ethylation reactions in the presence of ''organized media'' (micelles and vesicles) in flow systems.[99]

Low power helium plasmas such as the He-MIP, the He-SCP or the He-GD possess high electronic temperatures allowing for good sensitivity by atomic emission determination of elements with high lying excitation levels in the visible region (e.g. most of the non-metals). Unfortunately, their rotational (kinetic) temperatures are so low that only gases or vapors are usually allowed in such plasmas. Thus, the chemical generation of chlorine, bromine, iodine, as a means of analyte introduction to these plasmas in flow mode has been successfully evaluated.[59,60] Different oxidizing mixtures were assayed such as H_2O_2/H_2SO_4, $K_2Cr_2O_7/H_2SO_4$, $K_2S_2O_8/H_2SO_4$ and $KMnO_4/H_2SO_4$, the latter being the most promising for the simultaneous generation of the three above mentioned halines. Halide detection limits achieved using such flow generation mode lie in the low ng mL^{-1} level using MIP,[59] SCP[59] or GD/AES[60] detection.

6.3 Approaches for Preconcentration in the Gas Phase

Further increases of analytical sensitivity can be obtained by in situ trapping of the volatile form of the analyte transported to an adequate trap of the vapor. The most frequently used approaches for the in situ trapping/concentration of analytes after chemical vaporization are: (i) amalgamation of Hg in gold and release by heating, (ii) retention of mercury or volatile hydrides on a slightly heated surface (normally graphite coated with noble metals) and release by further heating, (iii) cryogenic trapping of hydrides and subsequent release by removal of the cryotrap.

The amalgamation of mercury CV with noble metals, and its posterior release by heating, is a well established method (commercial equipment is available) to further concentrate mercury vapor. The trap usually consists of gold-coated silica powder, gold coated sand, or amalgams such as gold-platinum gauze.[36]

Other approaches to trap the vaporized analytes involve the use of cryogenic traps[100] or the retention of volatile hydrides on a slightly heated surface (normally graphite coated with noble metals) followed in both cases by metal release by further heating. This latter approach has found an important range of applications

in connection with final atomization by GFAAS.[61,89] In this case, the graphite furnace is used as both the hydride trapping medium and the atomization cell. The hydride purged from the generator is trapped in the pre-heated furnace, usually in the range 300–600 °C, until the evolution of hydride from the sample is completed. The trapped analyte is subsequently atomized at temperatures generally over 2000 °C. The scope of this technique has been steadily expanding in the last decade comprising, for example, most of the volatile hydride forming elements, CV of mercury and cadmium, ethylated derivatives of Pb, Se, Cd, alkyl derivatives of As, Pb, Sn, etc.

The technique FI using CVG with GFAAS, which is now fully automated and is commercialized, enhances the sensitivity significantly and eliminates effectively the possible influence of the HG kinetics on the signal shape. The nature of the graphite tube is expected to affect greatly the efficiency of hydride adsorption. It has been shown that the coating of the graphite tube with Pd, Ir Zr, Ag or mixtures, improves the sensitivity and precision significantly.

The sensitivities achieved with the above mentioned gas phase preconcentration methods are very high and in most cases they are limited by the blank values, being a major contribution to the contamination by the analyte in the reagents. In this sense, approaches such as the use of immobilized tetrahydroborate could open a way for further improvements in detection limits.[27]

7 FLOW SYSTEMS FOR ON-LINE LIQUID–LIQUID SEPARATION AND PRECONCENTRATION AND TANDEM ON-LINE COMBINATIONS

7.1 Flow Systems for Liquid–Liquid Extraction

Traditional set-ups (separation funnels) for liquid–liquid extraction usually require large vessels, large volume of reagents and laborious handling, being very time-consuming methodologies. To overcome most of these problems, the use of flow liquid–liquid extraction procedures has been investigated. A continuous liquid–liquid extractor includes a solvent segmentor, an extraction coil and a phase separator. The segmentor receives the streams of two non miscible phases and gives rise to a segmented flow; analyte mass transfer takes place through the multiple interfaces established throughout an extraction coil; finally, two separate streams (one with each phase) emerge from the liquid–liquid phase separator.[101–104]

The liquid–liquid phase separator is probably the most important unit in the system. This unit must separate efficiently and rapidly the aqueous phase from the organic solvent in a continuous manner and its operation should be smooth and constant throughout.

Making use of differences in density of the two liquid phases and the selective wetting of inner surfaces, a variety of phase separators has been described. Figure 11 collects four representative examples of such devices. The separator depicted in Figure 11(a) is exclusively based on density differences between the phases: it

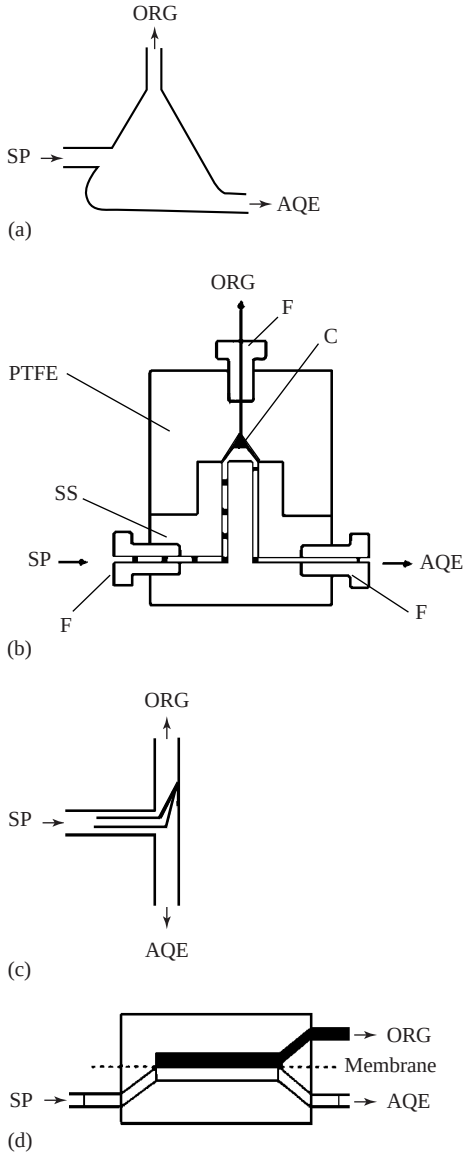

Figure 11 Diagrams of examples of liquid–liquid separation units (not drawn at the same scale). sp: segmented phase; aqe: aqueous phase; org: organic phase. (a) Gravity-based separator; (b) PTFE-stainless steel gravitational phase separator, C: conical cavity, F: threading fittings, SS: stainless steel; (c) T-type separator; (d) membrane type separator. (Parts (a), (c) and (d) reproduced from Lin et al., *Spectrochim. Acta*, Part B **51**, 1769–1775 © 1996; part (b) reproduced from Tao and Fang,[101] © 1995, with permission from Elsevier Science Ltd.)

consists of a conical device into which the segmented flow penetrates to form a single interface in the center. The flow of the lighter phase leaves the separator from its top while the heavier phase goes through the bottom of the device. To avoid large dead volumes in FIA systems, the volume of the inner cavity is of the order of tens of microliters. Figure 11(b) shows an improved gravitational phase separator.[101] The separation unit was composed of two half blocks. The upper block was made from PTFE (poly(tetrafluoroethylene)), with a conical cavity and a single outlet for the separated phase, while the lower was of stainless steel, and furnished with an inlet and outlet for the segmented phase and waste flows, respectively. The dead volume of the conical cavity was about 45 μL. Other improved models, all-glass made based on density differences have been also proposed.[102]

The T-type phase separator depicted in Figure 11(c) relies on both gravitational and selective wetting effects (as can be seen, it contains a whisker of fluoroplastics to assist the separation of the organic flow). The basic design of the membrane-based separators are shown in Figure 11(d). They are constructed with membranes which are permeable to most organic solvents but which repel aqueous solutions. The membrane is placed between two blocks. Once the segmented aquo/organic phase enters the lower part of the minichamber, the organic phase (lighter) passes across the membrane while the aqueous phase continues; thus, the two phases are separated. The membrane separator provides perhaps the best performance in terms of separation efficiency, reproducibility and applicability to a great variety of organic solvents; however, the short lifetime of the membrane is one of its serious practical limitations.

Important pitfalls of the liquid–liquid approach include: (i) the limited compatibility of organic effluents with plasma atomic sources, (ii) comparatively low enrichment factors for preconcentration purposes as compared to solid–liquid system (the actual enrichment factors depend on the flow-rate ratio for the two liquids involved and operational impediments preclude the use of highly different flow rates of the two phases), (iii) frequent problems associated to poor long term stability and robustness of the system requiring special operator skills.

Notwithstanding those limitations, flow liquid–liquid extraction systems have found important fields of application in connection with atomic detectors which withstand organic solvent introduction, such as FAAS and GFAAS. Finally, the use of these continuous separation systems for indirect analysis of organic pharmaceutical products should be highlighted here. Using the flow liquid–liquid separation strategy, it is possible to increase the analysis reproducibility as compared with the batch analogs, and some analyses would be only feasible when using these flow systems. For example, the determination of alkaloids has been described by formation of ion pairs between the sought drug and the inorganic complexes, BiI_4^- or $Co(SCN)_4^{2-}$, followed by liquid–liquid separation and measurement of bismuth or cobalt in the organic phase[103] or the determination of amphetamines based on their reaction with carbon disulfide to yield dithiocarbamic acids; these acids reacted with Cu(II), Ni(II) and Zn(II) and were extracted with IBMK being the salts or complexes on-line detected with FAAS.[104]

7.2 Integration of Two Continuous Separation Units: The Tandem On-line Concept

Occasionally, the complexity of the sample matrix or the sensitivity requirements may demand the use of more than one separation/preconcentration strategy to obtain appropriate analytical performance. Coupling two flow separation techniques in a tandem mode for sample preparation may be an effective solution to deal with such samples.

For example, we have seen that the sensitivity of analytical determinations by atomic spectroscopic methods can be enhanced by generation of volatile derivatives of the analyte. However, the chemical generation of these derivatives can be subjected to interferences (e.g. some metal transition ions in HG). Such interferences could be eliminated, for example, by using liquid–liquid or liquid–solid strategies. In this direction, a high degree of automation can be achieved by resorting to the 'tandem on-line' concept, i.e. the combination in a single on-line configuration of two continuous separation units: first, either a flow liquid–liquid extraction or a solid–liquid extraction step of the analyte; second, a volatile species generation from the second phase continuously fed to a gas–liquid separation unit; the first part allows improvements in both, selectivity and sensitivity, while the second permits the introduction of the analyte in gas phase to the detector, giving rise to further increases of the analytical sensitivity achievable.

This idea of ''tandem on-line'' separations for sample introduction in atomic spectrometry, was first applied to arsenic determinations in different samples by coupling in an on-line mode, a liquid–liquid and a gas–liquid separation device to an ICPAES.[105] A limitation of the technique could be the degree of compatibility of the plasma source with the organic solvent used for the liquid–liquid extraction (a certain amount of volatile organic solvents can reach the plasma in the gas separation step). However, this problem is eliminated by using xylene to extract the analyte in the liquid–liquid separation process and by employing $NaBH_4$ in dimethylformamide for the CVG process and final gas–liquid separation step. A similar ''tandem on-line'' concept has been extended to

indirect determinations of organic compounds[16] and to metal speciation, such as the analysis of inorganic Hg(II) and methylmercury based on the selective liquid–liquid extraction of methylmercury, into xylene, as bromide and CV generation of atomic mercury from the organic phase.[106]

Other on-line two-step approaches involve the elimination of interferences or the preconcentration of the analyte (prior to the HG step), using an ion exchange column (solid–liquid system), such as the manifold described for the determination of As, Sb and Se in a cobalt matrix.[107]

8 FLOW SYSTEMS FOR FIELD SAMPLING AND ON-LINE DIGESTION OF SAMPLES

8.1 Minicolumn Field Sampling and Flow Injection Analysis

The problem of securing during sample storage, an unchanged concentration of low levels of analyte or preserving the integrity of the species has received scarce attention from the analytical community both, for total metal and for metal speciation analyses. In recent years progress has been made in the development of true "field sampling" methods for the analysis of waters (seawater, river, lakes, etc.), which are mainly based on the isolation of the desired species in a stable form on appropriate minicolumns from which the analyte is later eluted for direct determination in the laboratory. A lateral advantage of this strategy is the easiness to transport, from the sampling source to the laboratory, small minicolumns containing the analyte (instead of transporting large containers with the liquid samples). This approach has been reported for the field sampling of Cr(III)/Cr(VI),[76] inorganic and organomercury species,[74] the determination of the 'fast reactive' aluminum fraction in waters,[108] etc.

Minicolumn field sampling is readily implemented using basic FI equipment. The sample delivery system can be a syringe, in the simplest case, or a peristaltic pump (battery-operated pumps are available). Although different approaches are possible, the minicolumns are generally transported back to the laboratory after sampling and on-line eluted to the detector. On-line elution reduces the risk of sample contamination or loss; on the other hand, since small elution volumes are used, high preconcentration factors can be achieved.

Different solid active phases can be used as column packing materials for field sampling (reviewed in section 5.1). However, it is important to point out that for minicolumn field sampling, additional parameters such as column to column variability and stability of immobilized species adsorbed in the column are of particular relevance and should be studied carefully.

8.2 On-line Photo-oxidation Flow Systems

The use of UV radiation to decompose organic matter is widely described in the literature; in particular, the oxidation reaction that combines a strong oxidizing agent with UV radiation has been described as an effective mode to destroy organic matter.

Concerning flow systems for analytical purposes, several photoreactor designs have been proposed to obtain the highest efficiency in such on-line decompositions. The favorable properties of PTFE for constructing tubular photoreactors is one of the most interesting features of this development, since in addition to its low UV absorption this material shows significant advantages over quartz coils, such as lower cost, easier handling and lower fragility. In atomic spectroscopy, most on-line photoreactors have been post-column coupled for metal speciation in hyphenated techniques using liquid chromatography for separation.[109]

FI systems involving on-line photo-oxidation schemes have been also proposed. Figure 12 shows the schematics of one such manifold used for the determination of organoarsenic compounds by on-line arsine

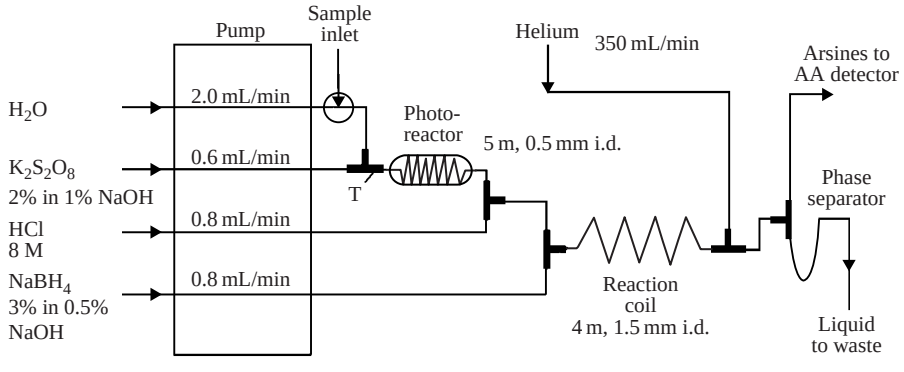

Figure 12 On-line flow sample digestion with UV radiation from a high intensity mercury lamp. (Reproduced with permission of Elsevier Science Ltd. from Atallah and Calman,[18] © 1991.)

generation.[108] A mercury lamp (OD 1.5 cm, length 20 cm) is used as UV source. The lamp is wrapped with a coil of 5 m of 0.56 mm ID. PTFE tubing through which the sample flows. To increase the light intensity reaching the coil and to prevent eye exposure to UV radiation, the unit was enclosed in aluminum foil.

8.3 On-line Microwave-assisted Digestion

A priori, continuous microwave digestion in a flow was expected to be associated with serious problems derived from the vigorous chemical conditions (e.g. excess of acids), resulting elevated temperature, high pressure, and long digestion times, frequently required to obtain complete decomposition. On the other hand, an additional problem to be faced is the evacuation of the gases produced during the digestion step. However, interesting instrumental designs have been developed trying to solve creatively these problems.

The microwave oven can be incorporated into flow manifolds which may be off-line or on-line connected to the atomic detector. For example, in order to analyze samples requiring long digestion times, the flow can be interrupted for a period of time, while the sample is in the oven, resulting in stopped-flow digestion systems. A prototype of this type was developed by Karanassios et al. In their system, a sample plug was pumped into the center of the coil located inside the oven, but leaving about 50 cm of air at both ends of the tubing.[110] During digestion, the sample slowly rotates inside the coiled tube; this rotation served as a stirring mechanism and also helped to reduce the effects of non-uniform microwave heating, due to the formation of 'hot' spots. The oven was modified by placing an electric fan to vent hot air during operation and to help cooling the tube at the end of the digestion.

Figure 13 illustrates the instrumental set-up for an on-line FI-microwave digestion system described by Haswell and Barclay for the analysis of biological tissues using FAAS detection.[111] Here, samples were prepared as slurries in 5% v/v nitric acid and sample volumes of 1 mL were introduced into the flow manifold. In the microwave oven, a PTFE coil of 20 m length and 0.8 mm i.d. was used as digestion loop, being the flow rate around 5 mL min^{-1}. Various types of animal and botanical tissues were analyzed and element recoveries for Ca, Fe, Mg and Zn were typically found to be in the interval 94–107% with sample throughputs of 1–2 min per sample. As can be seen in Figure 13, a backpressure regulator was used to remove the fumes produced during acid decomposition of organic materials. Other alternatives described for this purpose were the use of an ice-bath or a diffusion cell connected to a vacuum pump.

As already mentioned in section 6.1, prior to HG the destruction of the organic matter and the transformation of the analytes into appropriate oxidation steps are frequently needed. On-line continuous flow microwave digestions are ideally suited for FI using CVG sample introduction systems and have been extensively used to shorten the pre-reduction steps and to decrease the problems associated with the volatility and adsorption losses of analytes such as As, Sb, Sn, Sn, Pb or Hg during traditional extensive pretreatment procedures: microwave energy allows the rapid conversion of the analyte from its form bound to the organic matrix to a free inorganic appropriate oxidation step. The general procedure consists of first mixing the samples with an appropriate reagent, being the mixture transported in the flow to the microwave oven; the resulting digest is on-line mixed with sodium borohydride to form the corresponding volatile species which are swept into the detector. A bromination mixture (bromate/bromide) and persulfate in acidic media proved effective for the on-line speciation of selenium and determination of arsenic, antimony, tin and lead.[112] Table 1 collects several examples of applications of on-line FI with microwave digestion coupled to a variety of atomic detection systems, which have been proposed by different authors.[19,20,110–114]

Perhaps the more advanced degree of automation using flow systems has been demonstrated with an on-line microwave assisted mineralization manifold designed for in vivo sample uptake of whole blood samples (Figure 14).

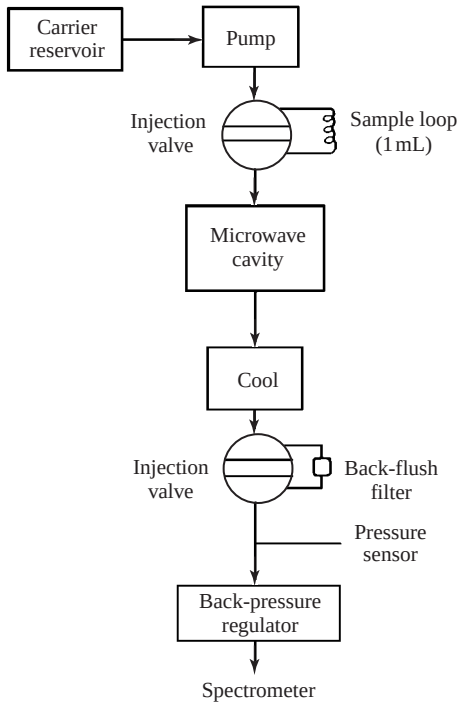

Figure 13 Experimental set-up for an on-line microwave digestion FI system for FAAS detection. (Reproduced with permission of the Royal Society of Chemistry from Haswell and Barclay,[111] © 1992.)

Table 1 Selected examples of FI digestion procedures with microwave heating and atomic detection

Sample	Analyte	Comments	Detector	Ref.
Powdered botanical and biological reference materials	Al, Ba, Ca, Cu Fe, Mg, Mn, Zn	Off-line stopped-flow digestion system	ICPAES	110
Animal and botanical tissues	Ca, Fe, Mg, Zn	Samples are introduced as slurries in 5% HNO_3	FAAS	111
Urine and waters	Hg, As, Bi, Pb, Sn	Bromate–bromide and peroxidisulfate as oxidants	HG/AAS	112
Sediments and biological tissues	Hg	Organomercury is oxidized to inorganic Hg(II)	CV/AFS	19
Waters	Se(IV) and total inorganic Se	Microwaves were used only for total inorganic Se analyses	HG/AFS	113
Waters	Inorganic arsenic and organoarsenic	First report on FI/HG interface to ICPMS allowing for instantaneous reduction of both inorganic and organometallic As species	HG/ICPMS	114
Whole blood	Co	In-vivo uptake of specimens and on-line treatment of the samples	GFAAS	20

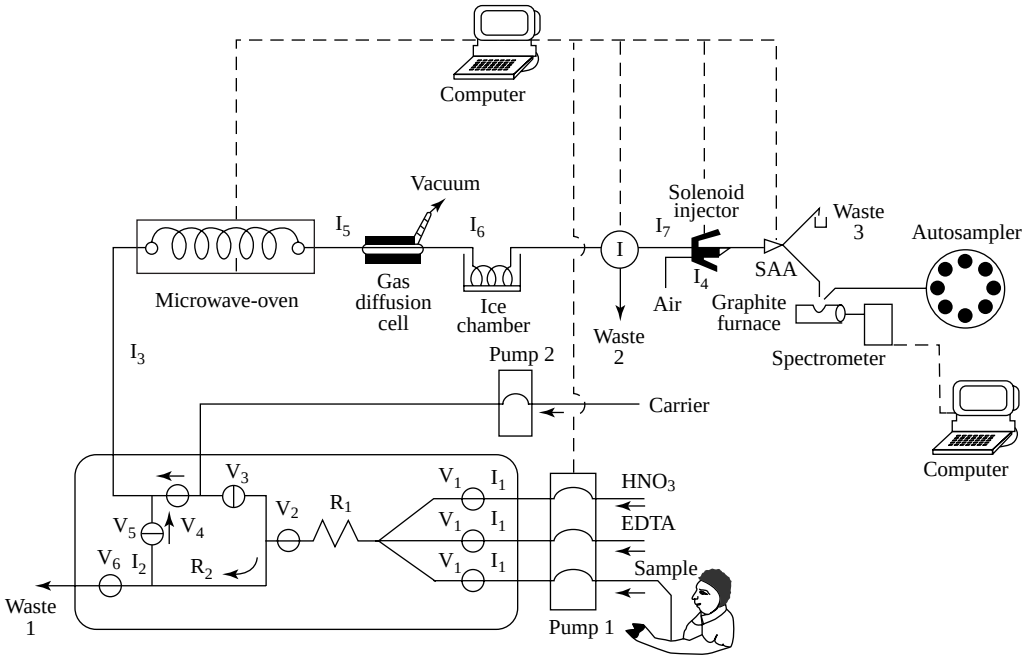

Figure 14 Diagram of a FI microwave oven ETAAS manifold for in vivo sample uptake of blood samples. I_1–I_7: tubing length; V_1–V_6: valves of the time-based solenoid injector; R_1: mixing tube; R_2: sample/reagent entrapment tubing; I: valve injector; SAA: sampling arm assembly. (Reproduced with permission of the Royal Society of Chemistry from Burguera et al.,[20] © 1995.)

The samples were drawn and pumped directly from the patient's forearm to a timed injector. This injector was automatically controlled to inject into the carrier stream a mixture on-line formed between the sample, an acidic solution and an anticoagulant.[20] Using this type of ''closed-circuit'' manifold, sample contamination problems from the environment should be reduced; besides, from the clinical point of view, it has also to be stressed that this type of manifold should eliminate risks of infection to nurses and analysts during the sample manipulation steps.

Before concluding this section it is important to point out that, although important progress has taken place in this field of on-line flow microwave digestion, the available instrumentation so far, does not yet seem to provide reliable performance. On a routine basis this technique needs further technical development to solve frequent malfunctions which will be probably overcome in the near future.

9 POTENTIAL APPLICATIONS PREVIEW AND FUTURE DEVELOPMENTS

As a consequence of the ever-growing number of chemical analyses routinely carried out, a major demand

in the area of research and development in chemical analysis is the need to make methods more cost-effective. On the other hand, the concept of quality is being fortunately installed in almost any aspect of our lives, analytical chemistry being a metrological discipline where the quality of results is of paramount importance. Hopefully, throughout this chapter it has become clear how both aspects could be improved by FIA methodologies.

There is a vast literature on the use of FIA techniques coupled to atomic detectors for a great variety of applications in different fields such as clinical, biological, environmental, etc. and several FIA methodologies, like volatile HG, are well implemented in routine analyses. However, the implementation of FIA strategies developed in research laboratories, in routine analysis is taking place at a slower pace, particularly for flow systems with higher degrees of complexity, even if they solve important problems. Since routine analyses demand robust instrumentation and procedures, this gap between routine practice and research laboratory work is being gradually reduced as reliable FI analyzers become commercially available.

From the upcoming research point of view, the great potential of the minicolumn 'field sampling' strategy should probably be highlighted; its merits have yet to be fully appreciated and exploited. Prospects for new developments and routine analysis in this area would be stimulated if custom sampling kits and minicolumn chemistries were commercially available. Also, the capability of flow systems to develop 'uncontaminating' methodologies, allowing sample contamination to decrease for modern ICPMS determinations and speciation cannot be overemphasized. The extreme sensitivity provided by this multielemental technique offers special attraction for its use in the analysis of very low level concentrations of any element and is unavoidable today for trace element speciation studies, particularly in biological systems. Besides, the ability of the ICPMS to accurately measure different isotopes opens new avenues in different fields such as in nutrition (e.g. to investigate the uptake of trace elements or their target organs and for identification of the source of environmental exposure to toxic elements). The simplification of the sample preparation steps and the reduction of contamination risks offered by FI methodologies, suggests that FI strategies will become the ''natural'' sample introduction mode in the ICPMS as in other detectors.

The field of analytical atomic spectrometry has undergone profound changes during the last decade. A sort of ''metamorphosis'' is taking place at the sunset of this millennium: instead of looking for information on the atomic/elemental composition of matter (for many decades the goal of atomic spectrometry), we are witnessing a return of many atomic spectroscopists to look for molecular information (as aimed at for present element speciating problems). Of course, as atomic methods are intrinsically non-speciating (i.e. they destroy molecular information) the most popular approach to overcome the problem consists of coupling a separating unit (for compounds separation) on-line with an atomic detector (for final element-specific detection). Particularly the use of MS analyzers has revolutionized this field by offering unprecedented sensitivity and isotope measurement capability.

It is clear that atomic spectrometry will expand its ''natural'' scope in the next millenium into molecular information gathering as well. To do so, flow analysis techniques (FIA conventional strategies, chromatography and also capillary electrophoresis) will be synergically combined/integrated with the rich array of atomic/specific detectors at our disposal nowadays.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AES	Atomic Emission Spectrometry
AFS	Atomic Fluorescence Spectrometry
ASV	Anodic Stripping Voltammetry
CV	Cold Vapor
CVG	Chemical Vapor Generation
DCP	Direct Current Plasma
ETAAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Flame Atomic Absorption Spectrometry
FAFS	Flame Atomic Fluorescence Spectrometry
FI	Flow Injection
FIA	Flow Injection Analysis
GD	Glow Discharge
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
GLS	Gas Liquid Separator
HG	Hydride Generation
IBMK	Isobutyl Methyl Ketone
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
MIP	Microwave Induced Plasma
MPT	Microwave Plasma Torch
MS	Mass Spectrometry
PTFE	Poly(tetrafluoroethylene)

SCP	Stabilized Capacitive Plasma
UV	Ultraviolet

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Automation in the Clinical Laboratory

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Flow-injection Techniques in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis

Atomic Spectroscopy **(Volume 11)**
Flame and Vapor Generation Atomic Absorption Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry • Microwave-induced Plasma Systems in Atomic Spectroscopy

REFERENCES

1. M. Valcárcel, M.D. Luque de Castro, M.T. Tena, 'Preliminary Operations: A Pending Goal of Today's Analytical Chemistry', *Anal. Proc.*, **30**, 276–279 (1993).
2. M. Valcarcel, M.D. Luque de Castro, *Automatic Methods of Analysis*, Elsevier, Amsterdam, 1988.
3. J. Ruzicka, E.H. Hansen, 'Flow Injection Analyses, Part I. A New Concept of Fast Continuous Analyses', *Anal. Chim. Acta*, **78**, 145–157 (1975).
4. J.R. Sarbeck, P.A. St. John, J.D. Winefordner, 'Measurement of Microsamples in Atomic Emission and Atomic Fluorescence Flame Spectrometry', *Mikrochim. Acta*, **149**, 55–60 (1972).
5. W.R. Wolf, K.K. Stewart, 'Automated Multiple Flow Injection Analysis for Flame Atomic Absorption Spectrometry', *Anal. Chem.*, **51**, 1201–1205 (1979).
6. N. Yoza, Y. Aoyagi, S. Ohashi, A. Tateda, 'Flow Injection System for Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **111**, 163–167 (1979).
7. Z. Fang, S. Xu, G. Tao, 'Developments and Trends in Flow Injection Atomic Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1–24 (1996).
8. Z. Fang, *Flow Injection Atomic Absorption Spectrometry*, Wiley, Chichester, 1995.
9. *Flow Injection Atomic Spectroscopy*, ed. J.L. Burguera, Marcel Dekker, New York, 1989.
10. *Flow Analysis with Atomic Spectrometric Detectors*, ed. A. Sanz-Medel, Elsevier, Amsterdam, 1999.
11. J. Ruzicka, E.H. Hansen, *Flow Injection Analysis*, Wiley, New York, 2nd edition, 1988.
12. I. Lavilla, B. Perez-Cid, C. Bendicho, 'Use of Flow-injection Sample-to Standard Addition Methods for Quantification of Metals Leached by Selective Chemical Extraction from Sewage Sludge', *Anal. Chim. Acta*, **381**, 297–305 (1999).
13. H. El Azouzi, M.Y. Pérez-Jordán, A. Salvador, M. de la Guardia, 'Extension of the Dynamic Range of Flame Atomic Absorption Spectrometry using Flow Injection Analysis with Variable-volume Dilution Chambers', *Spectrochim. Acta*, **51B**, 1747–1752 (1996).
14. L. Campanella, K. Pyrzynska, M. Trojanowicz, 'Chemical Speciation by Flow-injection Analysis. A Review', *Talanta*, **43**, 825–838 (1996).
15. J.R. Pretty, E.A. Blubaugh, J.A. Caruso, 'Determination of As(III) and Se(IV) Using an On-line Anodic Stripping Voltammetry Flow Cell with Detection by Inductively Coupled Plasma Atomic Emission Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **65**, 3396–3403 (1993).
16. A. Menéndez García, E. Sánchez-Uría, A. Sanz-Medel, 'Rapid Indirect Determination of Very Low Levels of Cocaine by Tandem On-line Continuous Separation and Inductively Coupled Plasma Atomic Emission Spectrometry Detection', *J. Anal. At. Spectrom.*, **11**, 561–565 (1996).
17. C.P. Hanna, J.F. Tyson, S. McInthosh, 'Determination of Total Mercury in Waters and Urine by Flow Injection Atomic Absorption Spectrometry Procedures Involving On and Off-line Oxidation of Organomercury Species', *Anal. Chem.*, **65**, 653–656 (1993).
18. R.J. Atallah, D.A. Kalman, 'On-line Photo-oxidation for the Determination of Organoarsenic Compounds by Atomic-absorption Spectrometry with Continuous Arsine Generation', *Talanta*, **38**, 167–173 (1991).

19. K.J. Lamble, S.J. Hill, 'Determination of Mercury in Slurried Samples by Both Batch and On-line Microwave Digestion-Cold Vapor Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **11**, 1099–1103 (1996).
20. M. Burguera, J.L. Burguera, C. Rondón, C. Rivas, P. Carrero, M. Gallignani, M.R. Brunetto, 'In Vivo Sample Uptake and On-line Measurements of Cobalt in Whole Blood by Microwave-assisted Mineralization and Flow Injection Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **10**, 343–347 (1995).
21. *Atomic Absorption Spectrometry. Theory, Design and Applications*, ed. S.J. Haswell, Elsevier, Amsterdam, 1991.
22. A. Montaser, D.W. Golightly, *Inductively Coupled Plasmas in Atomic Analytical Atomic Spectrometry*, VCH, New York, 2nd edition, 1992.
23. *Plasma Source Mass Spectrometry. New Developments and Applications*, eds. G. Holland, S.D. Tanner, The Royal Society of Chemistry, Cambridge, 1999.
24. P.W. Alexander, R.J. Finlayson, L.E. Smythe, A. Thalib, 'Rapid Flow Analyses with Inductively Coupled Plasma Atomic Emission Spectroscopy Using a Micro-Injection Technique', *Analyst*, **107**, 1335–1342 (1982).
25. F.J. Krug, H. Bergamin, E.A. Zagatto, 'Commutation in Flow Injection Analysis', *Anal. Chim. Acta*, **170**, 103–118 (1986).
26. M.R. Pereiro García, A. López García, M.E. Diaz García, A. Sanz-Medel, 'On-line Aluminium Pre-concentration and its Application to the Determination of the Metal in Dialysis Concentrates by Atomic Spectrometric Methods', *J. Anal. At. Spectrom.*, **5**, 15–20 (1990).
27. J.F. Tyson, 'High-performance, Flow-based, Sample Pretreatment and Introduction Procedures for Analytical Atomic Spectrometry', *J. Anal. At. Spectrom.*, **14**, 169–178 (1999).
28. J. Ruzicka, E.H. Hansen, 'Recent Developments in Flow Injection Analysis: Gradient Techniques and Hydrodynamic Injection', *Anal. Chim. Acta*, **145**, 1–15 (1983).
29. M. Novic, I. Berregi, A. Rios, M. Valcárcel, 'A New Sample-injection/Sample-dilution System for the Flow-injection Analytical Technique', *Anal. Chim. Acta*, **381**, 287–297 (1999).
30. J.L. Burguera, M. Burguera, C. Rivas, M. de la Guardia, A. Salvador, 'Simple Variable-volume Injector for Flow-injection Analysis', *Anal. Chim. Acta*, **234**, 253–257 (1990).
31. T.K. Starn, R. Pereiro, G.M. Hieftje, 'Gas-sampling Glow Discharge for Optical Emission Spectrometry. Part I: Design and Operating Characteristics', *Appl. Spectrosc.*, **47**, 1555–1561 (1993).
32. J.F. Camuña-Aguilar, R. Pereiro-Garcia, J.E. Sánchez-Uría, A. Sanz-Medel, 'A Comparative Study of Three Microwave-induced Plasma Sources for Atomic Emission Spectrometry-II. Evaluation of their Atomization/Excitation Capabilities for Chlorinated Hydrocarbons', *Spectrochim. Acta*, **49B**, 545–554 (1994).
33. A.S. Attiyat, G.D. Christian, 'Nonaqueous Solvents as Carrier on Sample Solvent in Flow Injection Analysis Atomic Absorption Spectrometry', *Anal. Chem.*, **56**, 439–442 (1984).
34. S. Olsen, L.C.R. Pessenda, J. Ruzicka, E.H. Hansen, 'Combination of Flow Injection Analysis with Flame Atomic-absorption Spectrophotometry: Determination of Trace Amounts of Heavy Metals in Polluted Seawater', *The Analyst*, **108**, 905–917 (1983).
35. K.K. Stewart, 'Depulsing System for Positive Displacement Pumps', *Anal. Chem.*, **49**, 2125–2126 (1977).
36. C.P. Hanna, P.E. Haigh, J.F. Tyson, S. McIntosh, 'Examination of Separation Efficiencies of Mercury Vapor for Different Gas–Liquid Separators in Flow Injection Cold Vapor Atomic Absorption Spectrometry with Amalgam Preconcentration', *J. Anal. At. Spectrom.*, **8**, 585–590 (1993).
37. J.C. de Andrade, C. Pasquini, N. Baccan, J.C. Van Loon, 'Cold Vapor Atomic Absorption Determination of Mercury by Flow Injection Analysis Using a Teflon Membrane Phase Separator Coupled to the Absorption Cell', *Spectrochim. Acta*, **38B**, 1329–1338 (1983).
38. R.M. Barnes, X. Wang, 'Microporous Polytetrafluoroethylene Tubing Gas Liquid Separator for Hydride Generation and Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **3**, 1083–1089 (1988).
39. R. Ma, W.V. Mol, F. Adams, 'Selective Flow Injection Sorbent Extraction for Determination of Cadmium, Copper and Lead in Biological and Environmental Samples by Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **293**, 251–260 (1994).
40. Z. Fang, L. Dong, 'Flow Injection On-line Coprecipitation Preconcentration for Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **7**, 439–445 (1992).
41. E.A. Zagatto, F.J. Krug, H. Bergamin, S.S. Jorgensen, B.F. Reis, 'Merging Zones in Flow Injection Analysis. Part 2. Determination of Calcium, Magnesium and Potassium in Plant Material by Continuous Flow Injection Atomic Absorption and Flame Emission Spectrometry', *Anal. Chim. Acta*, **104**, 279–284 (1979).
42. E.A. Zagatto, M.F. Giné, E.A.N. Fernandes, B.F. Reis, F.J. Krug, 'Sequential Injections in Flow Systems as an Alternative to Gradient Exploitation', *Anal. Chim. Acta*, **173**, 289–297 (1985).
43. J.F. Tyson, S.R. Bysouth, 'Network Flow Injection Manifolds for Sample Dilution and Calibration in Flame Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **3**, 211–215 (1988).
44. J.M.H. Appleton, J.F. Tyson, 'Flow Injection Atomic Absorption Spectrometry: The Kinetic of Instrument Response', *J. Anal. At. Spectrom.*, **1**, 63–74 (1986).

45. Z. Fang, B. Welz, M. Sperling, 'Contribution of System Components to Dispersion in the Analysis of Microvolume Samples by Flow Injection Flame Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**, 179–189 (1991).
46. M.W. Brown, J. Ruzicka, 'Parameters Affecting Sensitivity and Precision in the Combination of Flow Injection Analysis with Flame Atomic-absorption Spectrophotometry', *The Analyst*, **109**, 1091–1094 (1984).
47. 'Sistemas de introducción de muestras líquidas en espectrometría atómica', J.L. Todoli, J. Mori, V. Hernandis, A. Canals, Secretariado de Publicaciones, Alicante University, Spain, 1996.
48. K.E. LaFreniere, G.W. Rice, V.A. Fassel, 'Flow Injection Analysis with Inductively Coupled Plasma-atomic Emission Spectroscopy: Critical Comparison of Conventional Pneumatic, Ultrasonic and Direct Injection Nebulization', *Spectrochim. Acta*, **40B**, 1495–1503 (1985).
49. K.E. Lawrence, G.W. Rice, V.A. Fassel, 'Direct Liquid Sample Introduction for Flow Injection Analysis and Liquid Chromatography with Inductively Coupled Argon Plasma Spectrometric Detection', *Anal. Chem.*, **56**, 289–292 (1984).
50. J.A. Koropchak, D.H. Winn, 'Thermospray Interfacing for Flow Injection Analysis with Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Chem.*, **58**, 2558–2561 (1986).
51. M. Wu, Y. Madrid, J.A. Auxier, G.M. Hieftje, 'New Spray Chamber for Use in Flow-injection Plasma Emission Spectrometry', *Anal. Chim. Acta*, **286**, 155–167 (1994).
52. Y. Madrid, M. Wu, Q. Jin, G.M. Hieftje, 'Evaluation of Flow Injection Techniques for Microwave Plasma Torch Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **277**, 1–8 (1993).
53. M. Thompson, B. Pahlavanpour, S.J. Walton, G.F. Kirkbright, 'Simultaneous Determination of Trace Concentrations of As, Sb, Bi, Se and Te in Aqueous Solution by Introduction of the Gaseous Hydrides into an Inductively Coupled Plasma Source for Emission Spectrometry', *Analyst*, **103**, 568–573 (1978).
54. O. Astrom, 'Flow Injection Analysis for the Determination of Bismuth by Atomic Absorption Spectrometry with Hydride Generation', *Anal. Chem.*, **54**, 190–193 (1982).
55. F. Lunzer, R. Pereiro-García, N. Bordel-García, A. Sanz-Medel, 'Continuous Hydride Generation Low-pressure Microwave Induced Plasma Atomic Emission Spectrometry for the Determination of Arsenic, Antimony and Selenium', *J. Anal. At. Spectrom.*, **10**, 311–315 (1995).
56. J.A.C. Broekaert, R. Pereiro, T.K. Starn, G.M. Hieftje, 'A Gas-sampling Glow Discharge Coupled to Hydride Generation for the Atomic Spectrometric Determination of Arsenic', *Spectrochim. Acta*, **48B**, 1207–1220 (1993).
57. W.T. Corns, P.B. Stockwell, L. Ebdon, S.J. Hill, 'Development of an Atomic Fluorescence Spectrometer for the Hydride-forming Elements', *J. Anal. At. Spectrom.*, **8**, 71–77 (1993).
58. H. Tao, J.W.H. Lam, J.W. McLaren, 'Determination of Se in Marine Certified Reference Materials by Hydride Generation Inductively Coupled Plasma – Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**, 1067–1073 (1993).
59. J.F. Camuña, M. Montes, R. Pereiro, A. Sanz-Medel, C. Katschthaler, R. Gross, G. Knapp, 'Determination of Halides by Microwave Induced Plasma and Stabilised Capacitive Plasma Atomic Emission Spectrometry after On-line Continuous Halogen Generation', *Talanta*, **44**, 535–544 (1997).
60. J. Rodriguez, R. Pereiro, A. Sanz-Medel, 'Glow Discharge Atomic Emission Spectrometry for the Determination of Chlorides and Total Organochlorine in Water Samples via On-line Continuous Generation of Chlorine', *J. Anal. At. Spectrom.*, **13**, 911–915 (1998).
61. H. Matusiewicz, R.E. Sturgeon, 'Review. Atomic Spectrometric Detection of Hydride Forming Elements Following in situ Trapping Within a Graphite Furnace', *Spectrochim. Acta*, **51B**, 377–397 (1996).
62. H. Goenaga Infante, M.L. Fernández Sánchez, A. Sanz-Medel, 'Ultratrace Determination of Cadmium in Atomic Absorption Spectrometry Using Hydride Generation With In Situ Preconcentration in a Palladium-coated Graphite Atomizer', *J. Anal. At. Spectrom.*, **11**, 571–575 (1996).
63. M.A.Z. Arruda, M. Gallego, M. Valcárcel, 'Automatic Preparation of Milk Dessert Slurries for the Determination of Trace Amounts of Aluminium by Electrothermal Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **10**, 55–59 (1995).
64. Z. Fang, in *Flow Injection Atomic Spectroscopy*, ed. J.L. Burguera, Marcel Dekker, New York, 1989.
65. M. De la Guardia, in *Flow Analysis with Atomic Spectrometric Detectors*, ed. A. Sanz-Medel, Elsevier Science, Amsterdam, The Netherlands, Chapter 4, 1999.
66. M. Viczián, A. Lásztity, S. Wang, R.M. Barnes, 'On-line Isotope Dilution by Flow Injection and Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **5**, 125–133 (1990).
67. J. Goossens, L. Moens, R. Dams, 'Determination of Lead by Flow-injection Inductively Coupled Plasma Mass Spectrometry Comparing Several Calibration Techniques', *Anal. Chim. Acta*, **293**, 171–181 (1994).
68. S. Olsen, L.C.R. Pessenda, J. Ruzicka, E.H. Hansen, 'Combination of Flow Injection Analysis with Flame Atomic-absorption Spectrometry: Determination of Trace Amounts of Heavy Metals in Polluted Seawater', *The Analyst*, **108**, 905–917 (1983).

69. S. Hirata, K. Honda, T. Kumamaru, 'Trace Metal Enrichment by Automated On-line Column Preconcentration for Flow-injection Atomic Absorption Spectrometry', *Anal. Chim. Acta*, **221**, 65–76 (1989).
70. S. Hirata, Y. Umezaki, M. Ikeda, 'Determination of Cr(III), Ti, Va, Fe(III), and Al by Inductively Coupled Plasma Atomic Emission Spectrometry with On-line Preconcentrating Ion-exchange Column', *Anal. Chem.*, **58**, 2602–2606 (1986).
71. B. Mohammad, A.M. Ure, D. Littlejohn, 'On-line Preconcentration of Aluminium with Immobilized 8-Hydroxyquinoline for Determination by Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **7**, 695–699 (1992).
72. P. Schramel, L.Q. Xu, G. Knapp, M. Michaelis, 'Application of an On-line Preconcentration System in Simultaneous Inductively Coupled Plasma – Atomic Emission Spectrometry', *Mikrochim. Acta*, **106**, 191–201 (1992).
73. G. Knapp, K. Müller, M. Strunz, W. Wegscheider, 'Automation in Element Pre-concentration with Chelating Ion Exchangers. Plenary Lecture', *J. Anal. At. Spectrom.*, **2**, 611–614 (1987).
74. W. Jian, C.W. McLeod, 'Field Sampling Technique for Mercury Speciation', *Anal. Proc.*, **28**, 293–294 (1991).
75. C.W. McLeod, I.G. Cook, P.J. Worsfold, J.E. Davies, J. Queay, 'Analyte Enrichment and Matrix Removal in Flow Injection Analysis – Inductively Coupled Plasma – Atomic Emission Spectrometry: Determination of Phosphorus in Steels', *Spectrochim. Acta*, **40B**, 57–62 (1985).
76. A.G. Cox, C.W. McLeod, 'Preconcentration and Determination of Trace Chromium (III) by Flow Injection – Inductively Coupled Plasma-atomic Emission Spectrometry', *Anal. Chim. Acta*, **179**, 487–490 (1986).
77. H.A.M. Elmahadi, G.M. Greenway, 'Immobilized Algae as a Reagent for Preconcentration in Trace Element Atomic Absorption Spectrometry', *J. Anal. At. Spectrom.*, **6**, 643–646 (1991).
78. H.A.M. Elmahadi, G.M. Greenway, 'Speciation and Preconcentration of Trace Elements with Immobilized Algae for Atomic Absorption Spectrophotometric Detection', *J. Anal. At. Spectrom.*, **9**, 547–551 (1994).
79. J. Ruzicka, A. Arndal, 'Sorbent Extraction in Flow Injection Analysis and its Application to Enhancement of Atomic Spectrometry', *Anal. Chim. Acta*, **216**, 243–255 (1989).
80. M. Fernández García, R. Pereiro García, N. Bordel García, A. Sanz-Medel, 'On-line Preconcentration of Inorganic Mercury and Methylmercury in Seawater by Sorbent-extraction and Total Mercury Determination by Cold Vapor Atomic Absorption Spectrometry', *Talanta*, **41**, 1833–1839 (1994).
81. R.E. Santelli, M. Gallego, M. Valcarcel, 'Preconcentration and Atomic Absorption Determination of Copper Traces in Waters by On-line Adsorption-elution on an Activated Carbon Minicolumn', *Talanta*, **41**, 817–823 (1994).
82. H.L. Lancaster, G.D. Marshall, E.R. Gonzalo, J. Ruzicka, G.D. Christian, 'Trace Metal Atomic Absorption Spectrometric Analysis Utilizing Sorbent Extraction on Polymeric-based Supports and Renewable Reagents', *Analyst*, **119**, 1459–1465 (1994).
83. R. Martinez-Jiménez, M. Gallego, M. Valcarcel, 'Preconcentration and Determination of Trace Amounts of Lead in Water by Continuous Precipitation in an Unsegmented Flow Atomic Absorption Spectrometric System', *Analyst*, **112**, 1233–1236 (1987).
84. Z. Fang, M. Sperling, B. Welz, 'Flame Atomic Absorption Spectrometric Determination of Lead in Biological Samples Using a Flow Injection System with On-line Preconcentration by Coprecipitation Without Filtration', *J. Anal. At. Spectrom.*, **6**, 301–306 (1991).
85. H. Chen, S. Xu, Z. Fang, 'Determination of Copper in Water and Rice Samples by Flame Atomic Absorption Spectrometry With Flow Injection On-line Adsorption Preconcentration Using a Knotted Reactor', *Anal. Chim. Acta*, **298**, 167–173 (1994).
86. J.R. Pretty, E.A. Blubaugh, E.H. Evans, J.A. Caruso, T.M. Davidson, 'Determination of Copper and Cadmium using an On-line Anodic Stripping Voltammetry Flow Cell With Detection by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **7**, 1131–1137 (1992).
87. J.R. Pretty, E.A. Blubaugh, J.A. Caruso, T.M. Davidson, 'Determination of Cr(VI) and V(V) Using an On-Line Anodic Stripping Voltammetry Flow Cell with Detection by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chem.*, **66**, 1540–1547 (1994).
88. J. Dedina, D.L. Tsalev, 'Hydride Generation Atomic Absorption Spectrometry', Wiley, Chichester, 1995.
89. D.L. Tsalev, 'Hyphenated Vapor Generation Atomic Absorption Spectrometric Techniques. Invited Lecture', *J. Anal. At. Spectrom.*, **14**, 147–162 (1999).
90. A. Sanz-Medel, M.R. Fernández de la Campa, E. Blanco González, M.L. Fernández-Sánchez, 'Review. Organised Surfactant Assemblies in Analytical Atomic Spectometry', *Spectrochim. Acta*, **54B**, 251–287 (1999).
91. W. Ronald, W.L. Ott, 'Determination of Sub-Microgram Quantities of Mercury by Atomic Absorption Spectrophotometry', *Anal. Chem.*, **40**, 2085–2087 (1968).
92. J.M. Costa-Fernández, R. Pereiro-García, A. Sanz-Medel, N. Bordel-García, 'Effect of Plasma Pressure on the Determination of Mercury by Microwave-induced Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **10**, 649–653 (1995).
93. A. Stroh, V. Völlkopf, 'Optimization and Use of Flow Injection Vapor Generation Inductively Coupled Plasma Mass Spectrometry for the Determination of Arsenic, Antimony and Mercury in Water and Seawater at Ultratrace Levels', *J. Anal. At. Spectrom.*, **8**, 35–40 (1993).

94. A. Sanz-Medel, M.C. Valdés-Hevia y Temprano, N. Bordel Garcia, M.R. Fernández de la Campa, 'Generation of Cadmium Atoms at Room Temperature Using Vesicles and Its Application to Cadmium Determination by Cold Vapor Atomic Spectrometry', *Anal. Chem.*, **67**, 2216–2223 (1995).
95. M.C. Valdés-Hevia y Temprano, M.R. Fernández de la Campa, A. Sanz-Medel, 'Generation of Volatile Cadmium Species With Sodium Tetrahydroborate From Organized Media: Application to the Cadmium Determination by Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **8**, 847–852 (1993).
96. H. Goenaga Infante, M.L. Fernández Sánchez, A. Sanz-Medel, 'Vesicle-assisted Determination of Ultratrace Amounts of Cadmium in Urine by Electrothermal Atomic Absorption Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 899–903 (1998).
97. S. Rapsomanikis, 'Derivatization by Ethylation With Sodium Tetraethylborate for the Speciation of Metals and Organometallics in Environmental Samples. A Review', *Analyst*, **119**, 1429–1439 (1994).
98. J. Ashby, S. Clark, P.J. Craig, 'Methods for the Production of Volatile Organometallic Derivatives for Application to the Analysis of Environmental Samples', *J. Anal. At. Spectrom.*, **3**, 735–736 (1988).
99. M.C. Valdés-Hevia y Temprano, M.R. Fernández de la Campa, A. Sanz-Medel, 'Sensitive Inductively Coupled Plasma Atomic Emission Spectrometry Determination of Cadmium by Continuous Alkylation with Sodium Tetraethylborate', *J. Anal. At. Spectrom.*, **9**, 231–236 (1994).
100. M. Burguera, J.L. Burguera, M.R. Brunetto, M. De la Guardia, A. Salvador, 'Flow-injection Atomic Spectrometric Determination of Inorganic Arsenic(III) and Arsenic(V) Species by Use of an Aluminium-column Arsine Generator and Cold-trapping Arsine Collection', *Anal. Chim. Acta*, **261**, 105–113 (1991).
101. G. Tao, Z. Fang, 'On-line Flow Injection Solvent Extraction for Electrothermal Atomic Absorption Spectrometry Determination of Nickel in Biological Samples', *Spectrochim. Acta*, **50B**, 1747–1755 (1995).
102. S. Lin, Q. Shuai, H. Qiu, Z. Tang, 'Design and Application of a New Phase Separator for Flow Injection Liquid–Liquid Extraction', *Spectrochim. Acta*, **51B**, 1769–1775 (1996).
103. M. Eisman, M. Gallego, M. Valcárcel, 'Coupling of a Continuous Liquid–Liquid Extractor to a Flame Atomic Absorption Spectrometer for the Determination of Alkaloids', *J. Anal. At. Spectrom.*, **7**, 1295–1298 (1992).
104. R. Montero, M. Gallego, M. Valcárcel, 'Determination of Amphetamines by Use of Liquid–Liquid Extractor Coupled On-line to an Atomic Absorption Spectrometer', *Anal. Chim. Acta*, **252**, 83–88 (1991).
105. A. Menéndez García, E. Sánchez Uría, A. Sanz-Medel, 'Tandem on-line Continuous Separation and Determination of Arsenic by Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Chim. Acta*, **234**, 133–139 (1990).
106. A. Menéndez García, M.L. Fernández Sánchez, J.E. Sánchez Uría, A. Sanz-Medel, 'Speciation of Mercury by Continuous Flow Liquid–Liquid Extraction and Inductively Coupled Plasma Atomic Emission Spectrometry Detection', *Mikrochim. Acta*, **122**, 157–166 (1996).
107. H. Wang, Y. Chen, J. Wang, 'Determination of Arsenic, Antimony and Selenium in Cobalt by On-line Injection, Cation Exchange, Hydride-generation, Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Proc.*, **31**, 357–359 (1994).
108. B. Fairman, A. Sanz-Medel, P. Jones, 'Field Sampling Technique for the 'Fast Reactive' Aluminium Fraction in Waters Using a Flow Injection Mini-column System with Inductively Coupled Plasma Atomic Emission Spectrometric and Inductively Coupled Plasma Mass Spectrometric Detection', *J. Anal. At. Spectrom.*, **10**, 281–285 (1995).
109. R. Rubio, J. Albertí, A. Padró, G. Rauret, 'On-line Photolytic Decomposition for the Determination of Organoarsenic Compounds', *Trends Anal. Chem.*, **14**, 274–279 (1995).
110. V. Karanassios, F.H. Li, B. Liu, E.D. Salin, 'Rapid Stopped-flow Microwave Digestion System', *J. Anal. At. Spectrom.*, **6**, 457–463 (1991).
111. S.J. Haswell, D. Barclay, 'On-line Microwave Digestion of Slurry Samples With Direct Flame Atomic Absorption Spectrometric Elemental Detection', *Analyst*, **117**, 117–120 (1992).
112. D.L. Tsalev, M. Sperling, B. Welz, 'On-line Microwave Sample Pretreatment for Hydride Generation and Cold Vapor Atomic Absorption Spectrometry', *Analyst*, **117**, 1735–1741 (1992).
113. Y. He, H. El Azouzi, M.L. Cervera, M. De la Guardia, 'Completely Integrated On-line Determination of Dissolved Selenium (IV) and Total Inorganic Selenium in Seawater by Flow Injection Hydride Generation Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **13**, 1291–1296 (1998).
114. D. Beauchemin, 'Hydride Generation Interface for the Determination of Inorganic Arsenic and Organoarsenic by Inductively Coupled Plasma Mass Spectrometry Using Open-focused Microwave Digestion to Enhance the Pre-reduction Process', *J. Anal. At. Spectrom.*, **13**, 1–5 (1998).

Glow Discharge Optical Spectroscopy and Mass Spectrometry

Robert E. Steiner
Los Alamos National Laboratory, Los Alamos, USA

Christopher M. Barshick
Oak Ridge National Laboratory, Oak Ridge, USA

1 Introduction **1**
2 The Glow Discharge **2**
2.1 Fundamental Glow Discharge Processes 2
2.2 Radiofrequency-powered Glow Discharge Operation 5
2.3 Pulsed Operation of the Glow Discharge 6
2.4 Applications 7
3 Spectrochemical Methods of Analysis **8**
3.1 Basic Requirements Necessary to Obtain Optical Information 8
3.2 Atomic Emission Spectroscopy 10
3.3 Atomic Absorption Spectroscopy 13
3.4 Atomic Fluorescence Spectroscopy 14
3.5 Optogalvanic Spectroscopy 15
4 Mass Spectrometric Methods of Analysis **15**
4.1 Basic Requirements Necessary to Obtain Mass Abundance Information 16
4.2 Magnetic Sector Mass Analyzers 18
4.3 Quadrupole Mass Filters 19
4.4 Ion Trap and Fourier Transform Ion Cyclotron Resonance Devices 20
4.5 Time-of-flight Mass Spectrometers 21
5 Conclusions **22**
Disclaimer **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Glow discharge (GD) sources provide analytically useful gas-phase species from solid samples. These sources can be interfaced with a variety of spectroscopic and spectrometric instruments for both quantitative and qualitative analysis. Optical (atomic absorption spectroscopy, AAS; atomic emission spectroscopy, AES; atomic fluorescence spectroscopy, AFS; and optogalvanic spectroscopy) and mass spectrometric (magnetic sector, quadrupole mass analyzer, QMA; quadrupole ion trap, QIT; Fourier transform ion cyclotron resonance, FTICR; and time-of-flight, TOF) instrumentation are well suited for coupling to the GD.

The GD is a relatively simple device. A potential gradient (500–1500 V) is applied between an anode and cathode. In most cases the sample is also the cathode. A noble gas (e.g. Ar, Ne, Xe) is introduced into the discharge region prior to power initiation. When a potential is applied, electrons are accelerated toward the anode. As these electrons accelerate they collide with gas atoms. A fraction of these collisions are of sufficient energy to remove an electron from a support gas atom, forming an ion. These ions are, in turn, accelerated toward the cathode. These ions impinge on the surface of the cathode, sputtering sample atoms from the surface. Sputtered atoms that do not redeposit on the surface diffuse into the excitation/ionization regions of the plasma where they can undergo excitation via a number of collisional processes.

GD sources offer a number of distinct advantages that make them well suited for specific types of analyses. These sources afford direct analysis of solid samples, thus minimizing the sample preparation required for analysis. The nature of the plasma also provides mutually exclusive atomization and excitation processes that help to minimize the matrix effects that plague so many other elemental techniques. Unfortunately, the GD source functions optimally in a dry environment, making analysis of solutions difficult. These sources also suffer from difficulties associated with analyzing nonconducting samples.

In this article, the principles of operation of the GD plasma will first be reviewed, with an emphasis on how those principles relate to optical spectroscopy and mass spectrometry. Basic applications of the GD techniques will next be considered. These include bulk analysis, surface analysis, and the analysis of solution samples. The requirements necessary to obtain optical information will be addressed following the analytical applications. This section will focus on the instrumentation needed to make optical measurements using the GD as an atomization/excitation source. Finally, mass spectrometric instrumentation and interfaces will be addressed as they pertain to the use of a GD plasma as an ion source.

1 INTRODUCTION

We have always questioned the physical make-up of our universe, but only recently has science advanced to the point that some of the most fundamental questions can

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

be answered. While all fields of science play a role in information gathering, this is the specific charge of the analytical chemist. The field of analytical chemistry is broad, and although no one area can claim to be more valuable than another, the area of elemental analysis has always been critical to our understanding of the chemical and physical properties of materials.

Many years ago chemists characterized the world in terms of the basic elements of fire, earth, air, and water. Today, we classify our world in a similar manner, but in terms of carbon, nitrogen, oxygen, hydrogen, etc.; all that has really changed are the techniques we use – techniques that allow us to measure elemental and isotopic composition, elemental speciation, and spatial distribution of atoms and molecules.

This article focuses on two techniques used in elemental analysis: GD optical and mass spectrometries. We begin by describing the fundamental operation of a GD, including the sputtering and ionization processes. Next, we examine optical techniques that take advantage of the GD as an atomization source. Finally, we conclude by describing GD mass spectrometry, a powerful tool for multielement ultratrace analysis.

2 THE GLOW DISCHARGE

Although not always apparent, aspects of our everyday lives are permeated with technologies originally developed to solve scientific problems. An excellent example is the GD. Developed in the early 20th century[1–3] as a spectrochemical ionization source used for fundamental studies of atomic structure, a variation of the same GD now illuminates storefronts in the guise of the common neon sign.

The GD is a simple device. Before delving into the fundamental processes that characterize the glow, however, it is useful to define a few terms. A GD is one example of a general class of excitation/ionization sources known as ''plasmas''. This term refers to a partially ionized gas with equal numbers of positive and negative ions, and a larger number of neutral species.[4] Often the terms GD and gas discharge are interchangeable. A gas discharge is formed by passing an electric current through a gaseous medium.[5] For current to flow through this medium, a fraction of the gas must be ionized. In practice one applies a potential gradient between a cathode and an anode; in doing so, electrons are accelerated toward the anode. As the electrons move through the gas, they collide with gas phase atoms. A fraction of these collisions will be of sufficient energy to remove an electron from an atom, thus producing an ion and a secondary electron. The positively charged gas ions will, in turn, be accelerated toward and impinge upon the negatively biased sample cathode. Upon impact a variety of species will be liberated from the surface. Gas phase sample atoms are then free to diffuse into the plasma where they can undergo excitation and ionization. These dynamic processes are depicted in Figure 1.

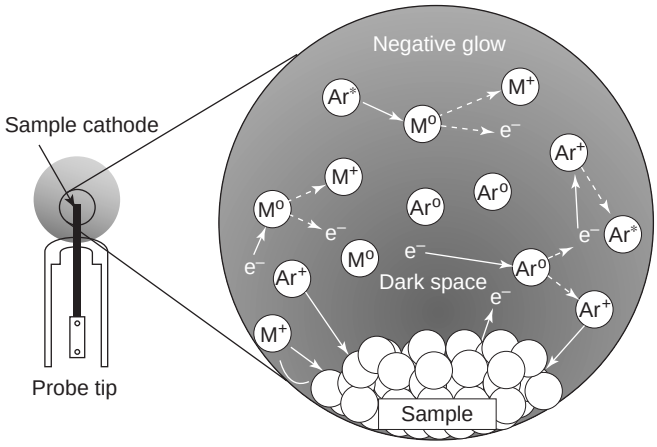

Figure 1 Illustration of various steady-state plasma processes. (Reproduced by permission from F.L. King, J. Teng, R.E. Steiner, *Glow Discharge Mass Spectrometry: Trace Elemental Determinations in Solid Samples*. Copyright (1995) John Wiley & Sons Ltd.)

A GD can be formed in virtually any vacuum cell that can be equipped with an inlet for a support gas. A variety of geometries have been investigated in the past. Table 1 lists some of the most common. The most popular source is based on a cross-shaped vacuum housing that provides ports for spectroscopic viewing, sample introduction, gas introduction, pressure monitoring, and ion extraction.

2.1 Fundamental Glow Discharge Processes

The fundamental processes occurring in the discharge define a number of discrete regions. For the current discussion, only three regions – the cathode dark space, the negative glow, and the Faraday dark space – will be defined, because the other regions, while important, are often not visible due to limited separation of the cathode and anode. For a complete treatise on this subject the reader is directed to several excellent articles.[6–8] The dark space regions are characterized by low light intensities relative to other regions of the plasma. This lack of luminosity arises from an absence of collisions and consequently the absence of excitation and the radiative relaxation events that produce photons. The cathode dark space is located between the cathode surface and the negative glow region. The presence of a large positive space charge in the negative glow causes the development of a potential gradient. The bulk discharge potential decreases rapidly through this region leading to

Table 1 GD ion sources and their characteristics

Source type	Voltage (V)	Current (mA)	Pressure (Torr)	Advantages	Disadvantages
Coaxial cathode	800–1500	1–5	0.1–10	Can conform to various sample shapes and sizes Penning ionization dominated	Powders must be pressed
Grimm	500–1000	25–100	1–5	Depth profiling Compacted powder samples	Flat samples only High gas flow rates
Hollow cathode	200–500	10–100	0.1–10	High sputter rate Intense ion beams for mass spectometry Large localized atom populations for optical spectroscopy	Complicated geometry Charge exchange mechanism is important
Jet-enhanced	800–1000	25–30	1–5	High sputter rate Compacted powder samples	Flat samples only High gas flow rates

its common name, the cathode fall. This large potential gradient affects the acceleration of electrons that can ionize the discharge gas species and liberate secondary electrons in the negative glow[4] which help to sustain the plasma. Radiative relaxation of species excited in the negative glow region yields the characteristic emission for which it is named. Because the major charge carriers in this region are electrons, the net space charge is negative. As electrons collisionally cool, they slow, decreasing their cross-section for excitational collisions with atoms. This smaller excitation cross-section results in the Faraday dark space close to the anode. Within this region the net space charge is zero and the potential gradient approaches a constant.

2.1.1 The Sputtering Process

The GD is of particular utility when analyzing solid conducting samples. Most competing techniques used for elemental applications (i.e. inductively coupled plasma mass spectrometry (ICPMS), flame spectroscopies, and graphite furnace techniques) require a sample in solution form to facilitate aspiration or introduction into the ionization source. The GD source possesses the inherent characteristic of producing gas phase analyte atoms directly from the solid conducting sample material. This phenomenon, known simply as cathodic sputtering, can be most easily described using a basic billiard ball analogy. Positively charged discharge gas ions are accelerated toward the negatively biased sample cathode. Prior to impact, these high-energy ions recombine with Auger electrons released from the cathode surface. The resulting high energy neutral species impact the surface of the cathode, transforming their kinetic energy (KE) into the lattice of the sample, thus causing a cascade of collisional events, much like the breaking of a racked set of billiard balls. If the resulting energy transfer is sufficient to overcome a surface atom's binding energy, the atom will be released into the gas phase. The sputtering process liberates not only individual cathodic atoms, but also electrons, ions, and clusters of atoms and molecules. This process is illustrated in Figure 2.

Sputtered electrons are accelerated across the cathode dark space into the negative glow where they can contribute to excitation and ionization of gas phase atoms. The ions formed by the sputtering event do not travel far from the cathode but are returned to the surface by the effects of the electric field. Once in the gas phase, the analyte atoms and neutral clusters are free to undergo collisions that may dissociate clusters and redeposit atoms at the surface. A fraction of these neutrals

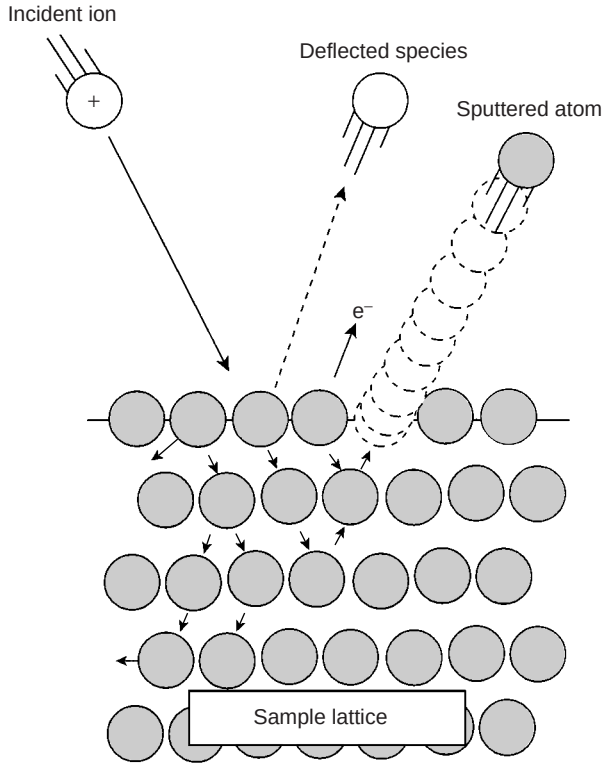

Figure 2 Illustration of cathodic sputtering process.

diffuse into the negative glow where they undergo excitation/ionization.

The effect of an ion's impact on the sample lattice can be measured by the sputter yield, S, as shown in Equation (1),[9]

$$S = (9.6 \times 10^4)\left(\frac{W}{M}\right) i^+ t \quad (1)$$

where W is the measured weight loss of the sample in grams, M is the atomic weight of the sample, i^+ is the ion current in amperes, and t is the sputtering time in seconds. The ion current is related to the total current, i, by Equation (2),

$$i^+ = \frac{i}{1+\gamma} \quad (2)$$

where γ is the number of secondary electrons released, on average, by a single ion.

Much of the previous research involving sputtering has used secondary ion mass spectrometry (SIMS). Typically, ion beams generated from these types of sources are more tightly focused and of much higher energy than in GDs. This is important to note since there are a number of characteristics that affect the sputter yield for a particular system including the nature of the target, the nature of the incident species, the energy of the incident ions, and the angle of the incident ion beam.

2.1.2 Excitation/Ionization Processes

Although the atomization or sputtering process creates species essential for atomic absorption and fluorescence spectroscopies, it does not supply the excited species (such as ions) needed for atomic emission and mass spectrometric analyses. These excitation/ionization processes occur in the collision-rich environment of the negative glow. Collisions that occur within this region not only provide analytically useful species, but are also integral in maintaining the stability of the plasma. Figure 3 illustrates

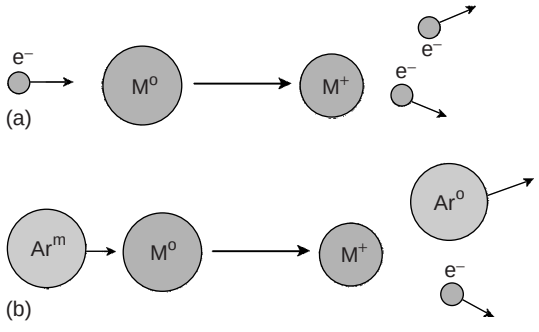

Figure 3 Illustration of the two major ionization pathways available in the GD plasma: (a) electron ionization; (b) Penning ionization.

the two principal types of collisions that occur within the negative glow involving electrons, excited atoms, and ions. Excitation is dominated by electron ionization (Figure 3a), while ionization is governed by both electron ionization and Penning ionization (Figure 3b). These two processes account for roughly 90% of ionization occurring within the plasma.[10] Electron ionization involves collision of an energetic electron with a gas phase atom. If the electron is of sufficient energy, it can interact with an electron in the valence shell of the atom, transfer enough energy to eject the electron from the atom's electron cloud, and form an ion and a secondary electron.[11,12] There is only a small probability that such a collision will result in ionization; this is zero below a threshold level, increasing at a rate of $C^{1.127}$, where C is the excitational cross-section, as the electron energy increases.[13] This can be explained using classical collision theory. At the threshold, only the complete transfer of energy will result in ionization. As the electrons increase in energy, collisions with only partial transfer of energy will result in ionization and the cross-section will increase. At high electron energies the time and wavefunction overlap is too short for ionization to occur, and the cross-section begins to decrease.[5,13] The average energy of electrons found within the GD is not great enough to ionize most elements or the discharge gas.[4,14,15] However, a Boltzmann distribution of energies predicts the presence of a small percentage of electrons with sufficient energy to ionize all elements. Chapman[4] has performed calculations of the Maxwell–Boltzmann distribution of electrons to determine the percentage of electrons with energies above 15.76 eV, the first ionization potential of argon, the most common discharge gas. At an average energy of 2 eV, 0.13% of the electrons present are of sufficient energy to ionize argon. At an average electron energy of 4 eV, the percentage increases to 5.1%. A more thorough presentation of these calculations is given by Chapman.[4]

Penning ionization, named after F.M. Penning who discovered the effect in 1925,[16,17] involves the transfer of potential energy from a metastable discharge gas atom to another atom or molecule. If the first ionization potential of an atom or molecule is lower than the energy of the metastable atom, ionization will occur when they collide. Ionization cross-sections for most elements are similar for the Penning process, resulting in somewhat uniform ionization efficiencies. Metastable states are reached through either the activation of a discharge gas atom to an excited state from which radiative decay is forbidden or by the radiative recombination of discharge gas ions with thermal electrons. For argon, the most common discharge gas, the metastable levels are the $^3P_2$ and $^3P_0$ states with energy levels of 11.55 eV and 11.72 eV, respectively. Table 2 lists the metastable levels and energies of the discharge gases most commonly used

Table 2 Metastable spectroscopic notations, energies and first ionization potentials for common discharge support gases

GD support gas	Spectroscopic notation	Metastable energy (eV)	First ionization potential (eV)
Helium (He)	2^1S	20.6	24.5
	2^3S	19.8	
Neon (Ne)	$^1P_0$	16.7	21.6
	$^3P_2$	16.6	
Argon (Ar)	$^3P_0$	11.7	15.8
	$^3P_2$	11.5	
Krypton (Kr)	$^3P_0$	10.5	14.0
	$^3P_2$	9.9	
Xenon (Xe)	$^3P_0$	9.4	12.1
	$^3P_2$	8.3	

to support GD plasmas. Metastable species are relatively long lived, existing for milliseconds under normal plasma conditions.[18] Their longevity within the plasma, along with a relatively large ionization cross-section and energy sufficient to ionize most elements, make this process a major contributor to ion production. Investigations of ionization processes in steady-state DC-powered discharges have indicated that 40–80% of ionization occurring within these plasmas can be attributed to the Penning ionization process.[19,20] Penning ionization also affords the unique advantage of discriminating against the ionization of discharge contaminants whose ionization potentials are greater than the metastable energy of the discharge gas atoms.

There are a number of other processes that play minor roles in ionization within the GD plasma. These processes involve charge exchange or associative ionization. Resonant charge exchange involves transfer of charge from one ion to an atom of the same species, while nonresonant charge exchange involves the transfer of charge from an ion to an atom of a different species. Associative ionization involves the combination of a metastable species with a gas phase atom to form a molecular ion with the liberation of a secondary electron. These processes contribute only marginally to ionization within the plasma but are the principal mechanisms by which interfering metal argide ions are formed.[4]

2.2 Radiofrequency-powered Glow Discharge Operation

Traditionally, most GD devices have used a steady-state DC-powered source; recently however, the utility of radiofrequency (RF)-powered discharges as sources for atomic spectroscopies and mass spectrometry has been investigated.[21,22] A major advantage offered by the RF-powered plasma is the ability to analyze directly nonconducting samples such as ceramics and glasses.[23] In the past, these types of samples were powdered, mixed with a conductive matrix, and pressed into a pin or disk to allow analysis by the DC GD.[24–26] Utilization of an RF plasma allows the sputtering and subsequent analysis of these sample types without this added sample preparation step.

Current cannot propagate through an insulating material; when subjected to an applied DC potential, therefore, an insulator behaves in a fashion similar to a capacitor and begins to accumulate charge. With the application of a negative voltage, the surface potential of the insulating material decays to a more positive potential with time. This decay can be attributed to charge neutralization at the cathode surface. The DC discharge will sustain itself until its threshold voltage is reached and the plasma is extinguished. Application of a high frequency potential to a conductive material adjacent to the insulator will allow positive charges that accumulate at the nonconductor surface to be neutralized by electrons during the positively biased portion of the cycle. Therefore, application of a high frequency potential will allow a negative potential to be maintained on the nonconductor's surface.

Figure 4 illustrates the phenomenon known as cathode self-biasing that allows the discharge to be maintained for lengthy periods of time.[4] Self-biasing is based upon the mobility difference between ions and electrons.

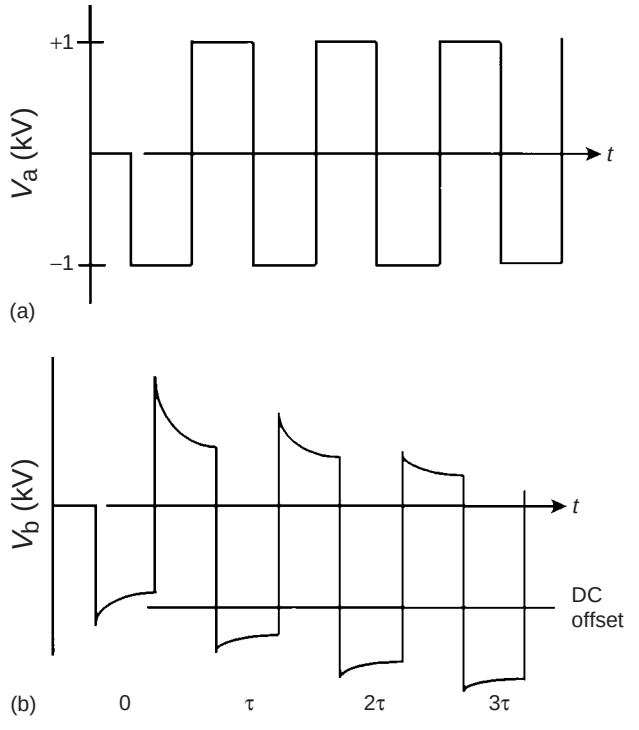

Figure 4 Electrode response to an applied square wave potential: (a) V_a, applied voltage; (b) V_b, response voltage. (Reproduced by permission from B. Chapman, *Glow Discharge Processes, Sputtering and Plasma Etching*. Copyright (1980) John Wiley & Sons Ltd.)

As a negative potential is initially applied, the surface charges quickly, reaches a maximum, and begins to decay (Figure 4b). When the potential is switched, electrons are accelerated and bombard the surface much like the positive ions during the negatively biased portion of the cycle. The electrons, however, have a greater mobility than the more massive positive ions, therefore the surface potential decays more quickly. After a number of cycles, the waveform will reach a steady DC offset. This DC offset potential is approximately one half of the applied peak-to-peak voltage and sustains the sputtering ion current. The nonconducting cathode is alternately bombarded by high-energy positive ions and low-energy electrons that support the sputtering process. Operation in this mode has proven useful for the analysis of materials such as nonconducting alloys, oxide powders, and glass samples.[27]

2.3 Pulsed Operation of the Glow Discharge

The GD plasma can be operated in a modulated power mode as well as the steady-state mode described previously. Pulsed power operation offers some distinct analytical advantages that make it attractive to the analyst.[28–32] Steady-state discharges are power limited because increased power application results in resistive heating of the cathode, eventually jeopardizing sample integrity.[33] Modulation of the applied discharge power permits operation at higher instantaneous power while keeping the average power at an acceptable level. This operation mode serves to increase the sputter yield by increasing the average energy and/or the number of incident ions while allowing the sample to cool during the off portion of the discharge cycle. Modulated operation also provides temporal segregation of discharge processes.

Modulated GD operation relies on a microsecond to millisecond square wave power pulse, with a duty cycle of 10–50%. These parameters allow sufficient time for cooling and for the removal of species from one pulsed event before the next one is initiated. Figure 5 depicts a typical pulse sequence showing both discharge gas (Ar) and analyte ion signal profiles. It is apparent that signal behaviors for discharge gas and analyte species differ dramatically. Upon power initiation, the discharge gas ion signal exhibits a sharp rise in intensity to a maximum[30,34] (Figure 5a). This ''prepeak'' results from the electrical breakdown of the discharge gas species upon power application. The short delay occurs because the acceleration of electrons and subsequent electron ionization of the discharge gas are not instantaneous processes. In contrast, the analyte ion profile behaves differently (Figure 5c). A much longer delay occurs before a more gradual signal increase is observed. This delay arises because sample atoms must first be sputtered from the cathode surface and diffuse into the negative glow region before they can undergo ionization. The temporal correlation between the observation of discharge gas ions and the appearance of analyte signal can be seen in the figure. Both signals reach equilibrium conditions about half way through the applied power pulse during the ''plateau'' region. During this time regime the plasma most closely approximates the behavior of a steady-state plasma.

Upon applied power termination, the two ion profiles again show markedly different behavior. The discharge gas ion profile decreases quickly ($\approx$500 μs) upon power termination. Previous studies have explained this signal decay as the rapid recombination of discharge gas ions with thermal electrons to form metastable discharge gas atoms.[35,36] The analyte ion profile quickly increases during the ''afterpeak'' region, reaching a maximum shortly after pulse power termination. This signal then decreases gradually to the baseline. Figure 5(b) represents the first derivative of the discharge gas ion signal profile. The maximum and minima observed in this portion of this figure represent the temporal location of the respective maximum signal intensity increases and decreases

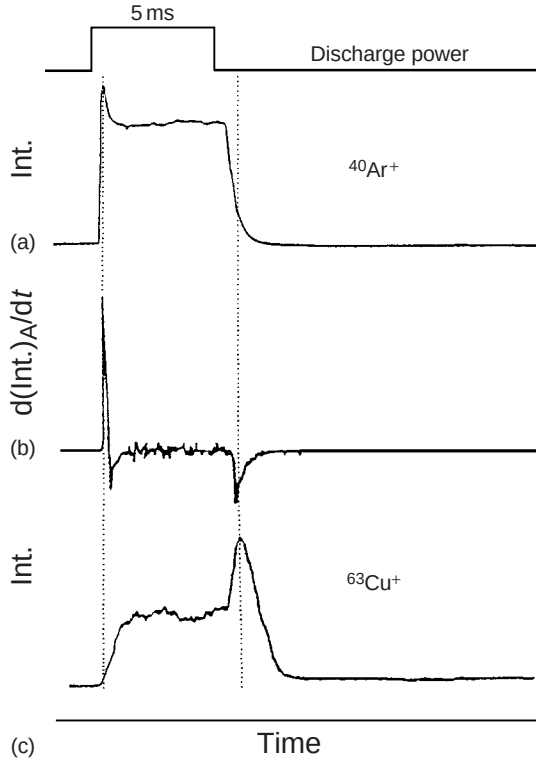

Figure 5 Temporal ion signal profiles: (a) discharge gas ions, $^{40}Ar^+$; (b) the first derivative of the $^{40}Ar^+$ profile; (c) sputtered analyte ions, $^{63}Cu^+$. (Reproduced by permission from King and Pan.[28] Copyright (1993) American Chemical Society.)

for the discharge gas signal. Interestingly, the maximum decrease in the discharge gas ion signal corresponds to the afterpeak maximum for the analyte species.[33] This correlation supports the theory that at power termination electrons are collisionally cooled and recombined with argon ions, forming metastable species. This increase in metastable argon atoms increases the probability of ionization via the Penning process, thus providing enhancements in the analyte ion signal.

Temporally gated separation and detection of species found in these distinct plasma regions increases the utility of the GD devices.[33,37] The most analytically useful region is the afterpeak. Data acquired within this time regime offer two advantageous characteristics. As illustrated in Figure 5(c), the signal intensity for the analyte species increases substantially, thus potentially enhancing the sensitivity for analyte species in this region. The other advantage arises from the suppression of electron-ionized interfering species.[29,30] The first ionization energies of the discharge gas and molecular contaminants (e.g. H_2O, N_2, O_2) are too high to allow ionization through the Penning process. Upon power termination species that are ionized via electron impact are no longer excited. Thus, if the acquisition gate is moved far enough into the afterpeak, contributions from these species will be minimized.

2.4 Applications

In the early 1970s, as GD instrumentation and techniques were being transitioned from research applications to routine sample analysis, spark source techniques reigned as the analytical tool of choice for trace analysis of solids. Spark source spectrometry was limited, however, by its expense, complexity, and unreliability. These disadvantages associated with spark source spectrometry facilitated the acceptance of the GD for the analysis of solid samples.

2.4.1 Bulk Analysis

Historically, GD mass spectrometry and optical spectroscopies have proven most valuable for the analysis of bulk conductive solid samples. As explained earlier, GD methods provide a representative gas phase analyte population directly from samples in the solid form, thus dramatically simplifying sample preparation. GD techniques can also be used for the analysis of nonconducting samples such as glasses, polymers, and ceramics.[38–40] The first analyses of nonconducting materials involved mixing a powdered sample with an easily sputtered, conducting matrix, such as high purity copper or silver powder.[41–44] The mixture was pressed into a sample cathode using a die and hydraulic press system. Unfortunately, this complicated sample preparation, especially for samples that were not in powdered form. Homogeneity of the mixed sample also became of concern for the analyst. Two other methods for analysis of nonconducting samples have also be used. These methods do not require sample mixing but rather use an RF-powered or a secondary cathode GD source to sample directly solid nonconducting materials. Section 2.2 describes plasmas powered by RF sources. Secondary, or sacrificial, cathode GD systems utilize a monoisotopic conductive mask (e.g. Ta, Pt) that is placed on top of the sample to be analyzed. A potential is applied to the mask, which thus assumes the role of the cathode. As material is sputtered from the mask, a large portion of it is redeposited on the surface of both the secondary cathode and the nonconducting sample. At this point, the layer of cathode material deposited onto the sample becomes conductive and thus assumes the applied potential and attracts impinging ions. As ions impinge upon the cathodic layer, they ablate both the cathodic material and the underlying nonconductor, introducing both species into the gas phase for subsequent analysis. It is clear that the mask material must not contain species that are of interest because this will contaminate the sample and preclude accurate measurements. A number of references are available describing in detail the use of secondary cathodes.[45–47]

2.4.2 Surface Analysis

Recently, surface analysis by GD spectrometry has aroused great interest in the analytical community. In a sense, GD is always a surface analysis technique, acting as an atomic mill to erode the sample surface via the sputtering process. Atoms sputtered from the surface are subsequently measured using either optical or mass spectrometric techniques. GD sputtering consumes relatively large quantities of sample in a relatively short time period (up to milligrams per minute), making analysis of thin films (<500 Å) difficult to virtually impossible. Plasma conditions, along with discharge gas choice, can be adjusted to slow the ablation process, thus allowing the analysis of films with micrometer thicknesses. Figure 6 is a graphical illustration of a typical spectral intensity–time profile of a multilayer coating prepared by chemical vapor deposition of carbon steel. The outer layer consists of vanadium carbide and the inner layer chromium carbide. It is obvious from Figure 6 that signal responses do not follow an ideal square wave pattern for appearance and disappearance. Clearly, the metal sample layer is not all eroded at a well-defined time, but is scattered across a diffuse region that is gradually removed. This signal tailing is a manifestation of the redeposition that occurs from the plasma. It has been estimated[48] that up to 67% of the sputtered atom population is returned to the surface by collisions with argon atoms, to

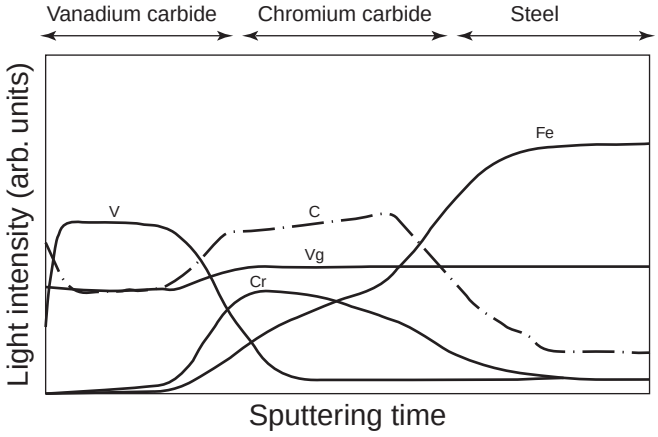

Figure 6 Qualitative analysis of multilayer coating. Vg, voltage. (Reproduced by permission from Hocquaux.[59])

be resputtered before eventually escaping permanently from the surface. Redeposition, along with a relatively high sputter rate, prevents GD from being used to profile very thin films. By calibrating erosion rates using standard layered samples, the thickness of layers can be determined by analyzing the signal–time profiles, allowing the full characterization of sample layers.

2.4.3 Analysis of Solution Samples

Analysis of solutions has been performed using GD techniques even though their major advantage lies in the ability to directly analyze solid-state samples. It is important to note that aspiration of a solution directly into the GD will cause quenching of the plasma, making removal of any solvent from the sample preferable. The simplest and most direct method of doing this is to evaporate a solution onto a conducting cathode, leaving a dried residue as the sample. A more elegant method for depositing solution samples onto a cathode involves electrodeposition. This approach also permits analytes in large sample volumes to be preconcentrated before analysis. Solutions containing samples have also been mixed with a conductive powder and dried for subsequent analysis. This approach is similar to that used for the analysis of nonconductive powders.

Although limited by its difficulty, direct analysis of solution samples by GD techniques has been performed for specialized applications. Strange and Marcus[49] have used a particle injection system to introduce a solution sample into the GD. Steiner et al.[50] have also utilized a pulsed plasma TOF system to obtain concurrent elemental and molecular information for high vapor pressure liquid samples. For the foreseeable future however, GD techniques will likely remain focused on the analysis of solid-state samples.

3 SPECTROCHEMICAL METHODS OF ANALYSIS

The goal of the analytical chemist is often to identify a particular chemical species or to quantify the amount of that species in a sample. In a spectrochemical analysis, the chemist uses the intensity of radiation emitted, absorbed, or scattered by a particular species versus a quantity related to photon energy, such as wavelength or frequency, to make such measurements.[51]

In this section, the basic requirements necessary to obtain a spectrochemical analysis are reviewed, along with three spectrochemical methods that are used to effect the measurement: AAS, AES, and luminescence spectroscopy. Although there are many variations of each of these methods, we will review only the classical approaches and the variations that have been used with GD devices. Table 3 summarizes the quantity measured and gives examples associated with each measurement technique.[51]

3.1 Basic Requirements Necessary to Obtain Optical Information

For thousands of years scientists have been performing qualitative analyses based on color, smell, taste, size, and shape. Although first year college chemistry students

Table 3 Classification of spectrochemical methods[51]

Class	Quantity measured	Examples
Emission	Radiant power of emission, Φ_E	Flame emission, DC arc emission, spark emission, inductively coupled plasma and direct current plasma emission, GD emission
Absorption	Absorbance or ratio of radiant power transmitted to that incident, $A = \log(\Phi/\Phi_0)$	UV/visible molecular absorption, IR absorption, atomic absorption
Luminescence	Radiant power of luminescence, Φ_L	Molecular fluorescence and phosphorescence, atomic fluorescence, chemiluminescence

UV, ultraviolet; IR, infrared.

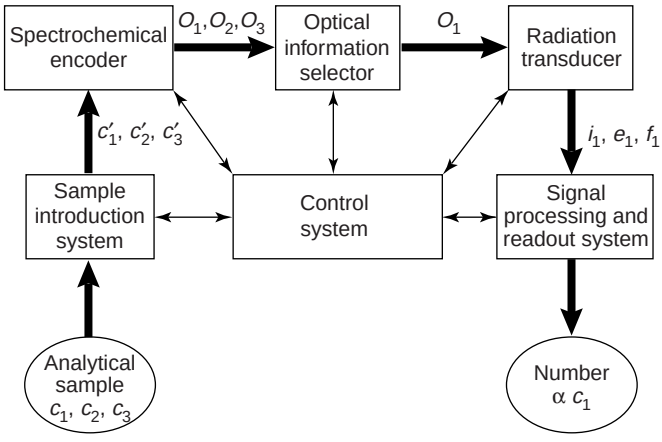

Figure 7 Spectrochemical measurement process. (Reproduced by permission from Ingle and Crouch.[51])

are still taught to use their senses to help identify substances in qualitative laboratory, for the rest of us these less precise approaches have been replaced with chemical and instrumental methods that can measure not only pure materials but trace components in complex mixtures. Most materials are not willing to give up this information spontaneously, however. Instead, to obtain chemical information about a sample, it is necessary to perturb the sample through the application of energy in the form of heat, radiation, electrical energy, particles, or a chemical reaction.[51] The application of this energy often causes electrons in analytes to be excited from their lowest energy or ground state to a higher energy or excited state. Several spectroscopic phenomena depend on these transitions between electronic energy states.[51] Information can be obtained by measuring the electromagnetic radiation emitted as the electron returns to its ground state from an excited state (emission), by measuring the amount of radiation absorbed in the excitation process (absorption), or by measuring the changes in the optical properties of the electromagnetic radiation that occur when it interacts with the analyte (e.g. ionization or photochemical reactions).[51] Qualitative information is extracted by observing a particular element-specific transition, while quantitative information is obtained by quantifying the amount of radiation emitted or absorbed in that transition. Figure 7 is an illustration used by Ingle and Crouch[51] to depict the many processes involved in converting concentration information into a number – the analytical chemist's goal in making a measurement. A sample introduction system presents the sample to the encoding system, which converts the concentrations c_1, c_2, c_3 into optical signals O_1, O_2, O_3. The GD is somewhat unusual (compared with other atomic sources like the flame or the inductively coupled plasma) because it serves as both the sample introduction system (through sputter atomization) and the spectrochemical encoder. The information selection system (often a monochromator) selects the desired optical signal O_1 for presentation to the radiation transducer or photodetector. This device converts the optical signal into an electrical signal that is processed and read out as a number. All spectrochemical techniques that operate in the UV/visible and IR regions of the spectrum employ similar instrumentation;[51] the only differences lie in the arrangement and type of sample introduction system, encoding systems, and information selection system.

3.1.1 Spectrometers

The optical information selector in Figure 7 sorts the desired optical signal from the many signals produced in the encoding process. Although it is possible to discriminate against background signals on the basis of time and position, most often discrimination is based on optical frequency (wavelength). The most widely used wavelength selection system is the monochromator, although there are a variety of other systems, including polychromators and spectrographs, and nondispersive systems like the Fabry-Perot, Michelson, Mach-Zender, and Sagnac interferometers. Our discussion will be limited to monochromators; for a thorough discussion of other types of wavelength selection systems, the reader is directed elsewhere.[51]

Monochromators isolate one wavelength from the countless number of wavelengths found in polychromatic sources. A monochromator consists of two principal components: a dispersive element and an image transfer system. Light is transferred from an entrance slit to an exit slit by a series of mirrors and lenses; along the way it is dispersed into its various wavelengths by a grating (or sometimes a prism). To change wavelengths, one rotates the dispersive element; the result is that different wavelength bands are brought through the exit slit in succession. One can easily calculate the angular dispersion (D_a) (i.e. the angular separation ($d\beta$) corresponding to the wavelength separation ($d\lambda$)) of a grating by knowing the angle of incidence, the angle of diffraction, the order of diffraction, and the groove spacing of the grating. For practical purposes, however, it is more important to calculate the linear dispersion, $D_l = dx/d\lambda$ (a value that defines how far apart in distance two wavelengths are separated in the focal plane), or the reciprocal linear dispersion, R_d (the number of wavelength intervals contained in each interval distance along the focal plane).

A monochromator's resolution is closely related to its dispersion in that dispersion determines how far apart two wavelengths are separated linearly while an instrument's resolution determines whether the two

wavelengths can be distinguished. In many cases the resolution is determined by the monochromator's spectral bandpass, s, defined as the half-width of the wavelength distribution passed by the exit slit. If the slit width is large enough to ignore aberrations and diffraction, a scan of two closely spaced monochromatic lines of peak wavelengths λ_1 and λ_2 will be just separated (baseline resolution) if $\lambda_2 - \lambda_1 = 2s$. Therefore, the slit-width-limited resolution $\Delta\lambda_s$ is given by Equation (3):

$$\Delta\lambda_s = 2R_d W \tag{3}$$

where W is the slit width.[51] By adjusting the monochromator so that $\Delta\lambda_0 = \Delta\lambda_1$ and $W = \Delta\lambda/R_d$, the image of λ_1 will be completely passed, while that of λ_2 will be at one side of the exit slit.[51]

3.1.2 Detectors

Two detectors are used most often in atomic spectroscopy: photomultipliers (the most common) and multichannel detectors. A photomultiplier is a more sophisticated version of a vacuum phototube. A cascade of electron collisions with dynodes of increasing potential and the subsequent ejection of electrons from each dynode's surface leads to the formation of an electrical current proportional to the number of photons striking the detector. The process begins with a photon striking a cathode made of a photoemissive material (e.g. alkali metal oxides, AgOCs, CS_3Sb). If the energy of the photon is above some threshold value, an electron is ejected from the cathode. Only a certain fraction of the photons with energy greater than threshold produce photoelectrons with sufficient KE to escape the photocathode. This fraction is called the quantum efficiency and is the ratio of the number of photoelectrons ejected to the number of incident photons. After leaving the photocathode, a photoelectron strikes the first dynode of the multiplier; this causes the subsequent ejection of two to five secondary electrons which are in turn accelerated by an electric potential to a second dynode where they cause the release of two to five more electrons. This multiplication process continues until the electrons reach the last dynode and impinge on the anode. A modern photomultiplier tube might have 5–15 dynodes (made of a secondary emission material like MgO or GaP) in a cascade. The result of this photomultiplication is a large charge packet a few nanoseconds in width produced at the anode for each photoelectron collected by the first dynode. Photomultipliers can be operated in either analog mode, where the average current that results from the arrival of many anodic pulse is measured, or in photon counting mode, where the number of anodic pulses, and not photons, is counted per unit time.

A wide variety of photomultipliers is available with both end-on and side-on viewing for adaptation to a wide variety of monochromators. Care must be given to the wavelength range over which one is working to ensure a uniform response. Other concerns for the spectroscopist, all of which are beyond the scope of this article, include the quantum efficiency of the photomultiplier, the multiplication factor of each dynode, the operating (accelerating) voltages applied to the dynodes, and the dark current generated when a potential is applied between the anode and cathode, with no photons hitting the photocathode.

The second type of detector that is widely used in atomic spectroscopy is the multichannel detector. These devices include early photographic detectors like photographic film or plates, as well as modern detectors such as photodiode arrays and charge-coupled and charge-injection devices. The idea behind the multichannel detector is simultaneous detection of dispersed radiation. Modern multichannel detectors usually take the form of some sort of solid state *pn*-junction diode device packaged in integrated-circuit form with a large number (e.g. 256, 512, or 1024) of elements arranged in a linear fashion. These devices often have linear dynamic ranges of two to four orders of magnitude. Limitations at the low end result from the noise associated with readout of a given dynode. Limitations at the upper end are the result of saturation; this is determined by the number of electron–hole pairs that can be created. A typical saturation charge is 1–10 pC. The reader is referred to several excellent references for a more thorough description of multichannel detectors.[52–54]

3.2 Atomic Emission Spectroscopy

AES is the simplest spectroscopic method for determining the elemental composition of a sample, and is the logical place to start talking about atomic spectrometry. Optical emission results from electron transitions occurring within the outer electron shells of atoms. These transitions give rise to line spectra where the wavelength of the lines relates to the energy difference of the levels according to Equation (4):

$$\Delta E = \frac{hc}{\lambda} \tag{4}$$

Spectroscopists often categorize spectral transitions according to term symbols. For a complete discussion of term symbols, the reader is referred elsewhere;[55] a brief discussion follows.

Each electronic state of an atom has five quantum numbers that define its electronic configuration. These include the principal quantum number, n, the orbital angular momentum quantum number, l, the orbital

magnetic quantum number, m_l, the electron spin quantum number, s, and the spin magnetic quantum number, m_s. According to Ingle and Crouch,[51] for many-electron atoms, the hydrogen quantum numbers can be thought of as describing the individual electrons, but they are not ''good'' quantum numbers for the entire atom. Good quantum numbers are associated with operators that commute with the total atom Hamiltonian. These include the resultant orbital angular momentum quantum number, L, produced by coupling the orbital angular momenta of each electron, and the resultant spin quantum number, S. For atoms with weak spin–orbit interactions, L and S couple to produce a total angular momentum quantum number, J. A multiplet of closely spaced states with the same L and S values but different J values is called a spectroscopic term, and is designated as $n^{2S+1}\{L\}_J$, where n is the principal quantum number for the valence electrons, $2S+1$ defines the multiplicity, and J is the total angular momentum quantum number. Of all the possible transitions between states, only a fraction of them are observed. From quantum mechanical principles, it is possible to derive selection rules that tell which transitions are allowed (i.e. those that occur with high probability and give reasonably intense lines) and which are forbidden (i.e. those that occur with low probability and give weak lines); this is beyond the scope of this article. For this discussion it is sufficient to note that term symbols for almost all practical configurations have been tabulated[55] and tables of spectral line intensities have been assembled for nearly all the elements.[56]

In AES, the information relevant to an analysis can be found in the radiation emitted by excited analyte atoms decaying from a nonradiational activation event. The radiant power of this emission is a function of several factors, including the population density of the excited atoms, the number of photons emitted per second by each atom, the energy of each photon, and the volume of the emitting system. The reader is directed elsewhere for a more complete discussion of each of these factors.[51]

AES has the power to provide rapid, qualitative, and quantitative multi-element analyses. Although a qualitative survey of the elements (i.e. a plot of the analytical signal versus wavelength) in a sample may be useful, more often the desired information is the concentration of an analyte. Unfortunately, this is almost never obtained directly as the result of an absolute measurement of an optical signal because amplitude and elemental concentration are seldom related in a simple way. Obtaining the desired concentration from an optical measurement usually involves calibration, subtraction of blanks, comparison with standards, and other similar procedures.[51] Quantitative analysis of a single element is most easily accomplished in AES by monitoring the emission intensity as a function of the analyte's concentration under a given set of conditions (e.g. constant discharge gas pressure, voltage, and current at a given wavelength). Standards, often provided by the National Institute of Standards and Technology, provide a range of concentrations over which a calibration curve can be developed. The signal intensity of an unknown concentration is then compared with the intensity of the standards, thus providing the concentration of the analyte in question.

Although simple in principle, quantification is complicated by a number of factors, including spectral background, incomplete wavelength separation, self-absorption, peak broadening, etc., most of which are beyond the scope of this article. When one considers these complicating issues, it is clear why it is important to control conditions precisely and use standards for the most accurate quantification.

Instrumentation used for emission spectroscopy includes an excitation source, a sample container, a wavelength selector, and a radiant power monitor. Depending upon the spectrochemical method, the excitation source and sample container may be separate components or they may be combined, as is the case with the GD. Figure 8 illustrates one instrumental configuration used in our laboratory at the Oak Ridge National Laboratory for AES. A 0.5-m monochromator serves as the wavelength selector and the combination of a photomultiplier tube, preamplifier, and readout photometer comprise the radiant power monitor. Figure 8 also illustrates three common sources used with GD emission spectroscopy – a planar cathode discharge and two versions of the hollow cathode discharge.

The planar cathode discharge is thus termed because the portion of the sample exposed to the discharge is flat. It is often contrasted with the coaxial or pin-type cathode used more commonly with mass spectrometry and described elsewhere in this chapter. With the development of the Grimm lamp in 1968[57] and its eventual commercialization, the planar cathode discharge gained widespread use for emission spectroscopy. Although other planar cathode discharges have been developed in the past 30 years, the Grimm source still finds the greatest application today. One interesting feature of the Grimm source is that it is an obstructed discharge (i.e. the discharge is confined to the sample by the extension of the anode into the cathode dark space). Moreover, the vacuum in the anode–cathode inner space is lower than in the discharge region itself, necessitating a dual outlet pump with a larger throughput for the inner electrode space.[58] Another interesting feature of the Grimm source is that the cathode is located outside of the source itself; this provides for easy sample interchange, and means that the Grimm source is particularly amenable to the analysis of any flat conducting surface that can be brought up to the source opening, such as metal sheets or disks.

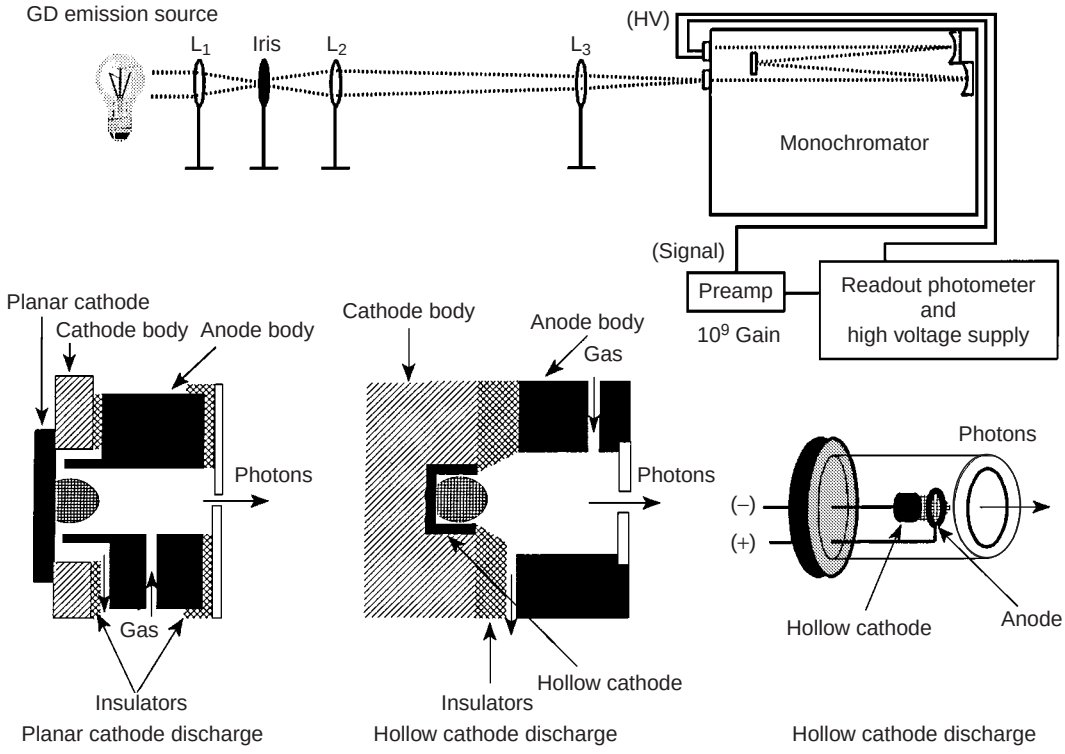

Figure 8 Schematic of an instrumental arrangement used at the Oak Ridge National Laboratory for AES. (Reproduced by permission from Harrison et al.[7] Copyright (1990) American Chemical Society.)

Typical operating conditions for the Grimm source are 500–1000 V, 25–100 mA, and 1–5 Torr. The relatively high power produced by the source (12.5–100 W) means that the cathode is often water-cooled; this usually is not a problem in emission spectroscopy, but makes interfacing the Grimm source with a mass spectrometer (with its high vacuum requirements) more difficult.

Planar cathode discharges have been interfaced to a variety of commercial emission instruments.[7] Grimm-type sources find their greatest use in trace elemental analysis of solids and in depth profiling of layered metal samples.[58] Detection limits by emission spectroscopy are of the order of 0.1 ppm.[7] Precision of the order of 0.5–5% has been obtained for concentrations in the 0.01–10 mg g^{-1} range.[58] Ablation rates range from 0.1 to 3 mg min^{-1} depending upon the element, discharge area, current, and voltage.[58] At these rates, the Grimm source cannot be used for thin film analysis, but it is ideal for thin layer analysis or in-depth profiling where it may be necessary to profile from a few nanometers to several tens of micrometers in a relatively short time.[59] Recently, Hocquaux[59] wrote an excellent chapter on thin film analysis by GD emission spectroscopy.

The other two sources shown in Figure 8 are hollow cathode discharges. Although the hollow cathode discharge appears similar to the planar cathode physically, the shape of the cathode cavity provides some properties that make it appealing for atomic spectroscopy. This discharge derives its name from a cathode that has been drilled out to form a cylindrical cavity closed at one end.[60] The so-called hollow cathode effect can be visualized as a GD with two parallel cathode plates being brought sufficiently close to each other until the two cathode glow regions coalesce.[61] The result of this coalescence is an increase in current density that can be several orders of magnitude larger than a single planar cathode at the same cathode fall potential.[62] Coupling this increase in current density with the longer residence time that the analyte experiences in the negative glow region (due to the cathode's shape) results in a dramatic increase in the intensity of radiation emitted compared to a planar cathode. In addition, background intensities are low because electron number densities are low, resulting in a very high signal-to-background ratio.[63] To perform an analysis using a hollow cathode discharge, it is necessary to machine the sample into the shape of a cylinder, or to place powder or metal chips into a hollow cathode made of some inert material such as graphite. One can also analyze solutions by drying a residue on the hollow cathode surface. Operating conditions vary widely, but typically range from 200 to 500 V, 10 to 100 mA, and 0.1 to 10.0 Torr.[7] Detection limits have been reported in

the picogram range,[58] but more typical results are in the nanogram range. Although hollow cathode discharges are used widely in atomic spectroscopy, the majority of these devices are light sources for AAS (see below). A typical hollow cathode lamp (HCL) is depicted at the extreme bottom right of Figure 8.

3.3 Atomic Absorption Spectroscopy

In absorption spectroscopy, spectrochemical information can be found in the magnitude of the radiant power from an external light source that is absorbed by an analyte.[51] To obtain information relevant to measuring an element's concentration, however, it is necessary for the frequency of the incident radiation to correspond to the energy difference between two electronic states of the analyte atoms being measured. Often, but not always, the atoms start in their electronic ground state and are excited to a higher lying electronic state by the incident radiation. The adsorption of this radiation usually follows Equation (5):

$$A = -\log T = \frac{-\log \Phi}{\Phi_0} = abc \qquad (5)$$

where A is the absorbance, T is the transmittance, a is the absorptivity, b is the path length of absorption, and c is the concentration of the absorbing species. This equation is commonly referred to as Beer's law. To calculate the concentration, one measures the incident radiation (M_0) and the transmitted radiation (M), calculates the absorbance, and relates the absorbance to concentration using a series of standards and calibration curves, similar to emission spectroscopy.

A typical instrumental configuration for an atomic absorption spectrometer is shown in Figure 9. A hollow cathode, fabricated from the elements of interest, is often used as the source of incident radiation, although an electrodeless discharge lamp may be used for some elements such as As, Se, or Te, where the emission from an HCL may be low. The HCL is focused to a point inside the discharge and then refocused into the entrance aperture of the monochromator. To obtain the background signal, one can use a shutter, or alternatively modulate the HCL and measure the background during the off period. In the arrangement shown in Figure 9, a mechanical chopper is used to facilitate background subtraction by providing a reference signal to a lock-in amplifier. Transmission is

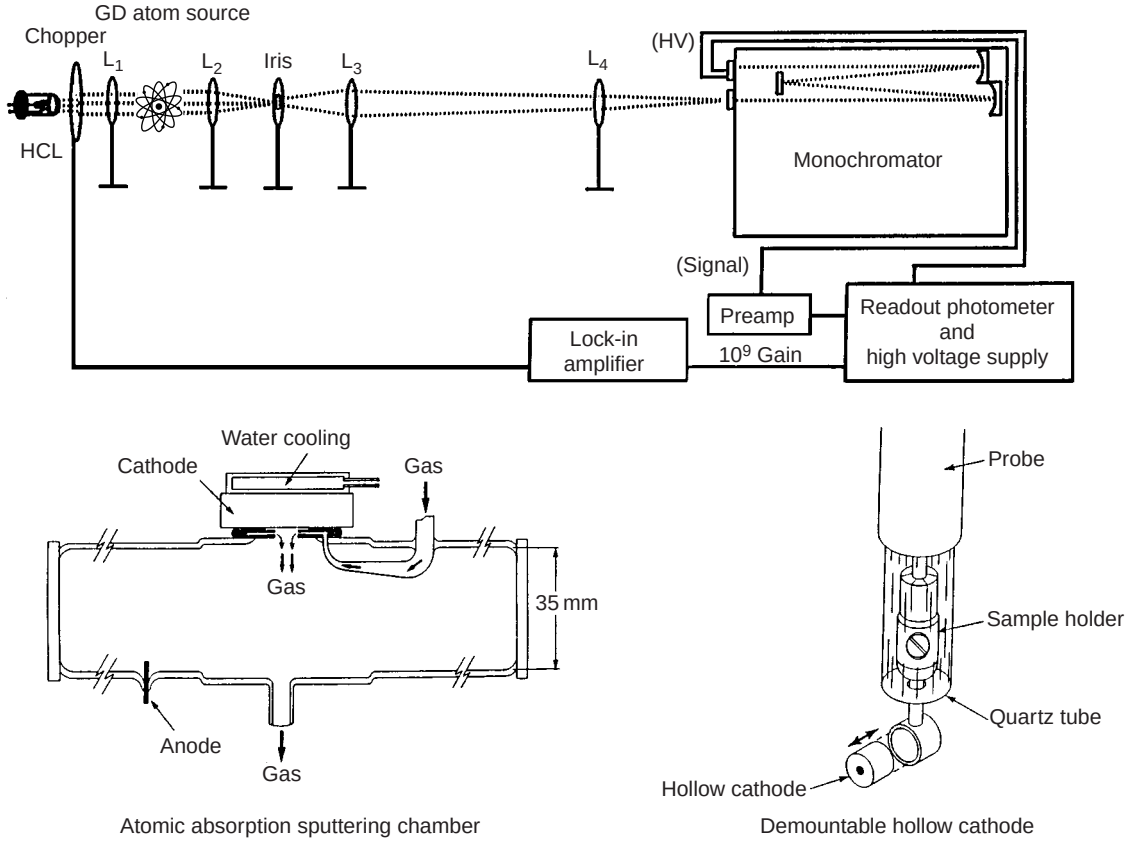

Figure 9 Schematic of an instrumental arrangement used at the Oak Ridge National Laboratory for AAS. (Reproduced by permission from Gough.[64] Copyright (1976) American Chemical Society.)

measured first with the discharge off and then with the discharge on, often for a range of currents and voltages. Using Beer's law, absorbance is calculated for a series of standards to produce a calibration curve; the absorbance of an unknown is then correlated with its concentration.

Two different discharge configurations are shown in Figure 9 for atomic absorption. The one on the left is an atomic absorption sputtering chamber developed by Gough.[64] A planar cathode is mounted near the top by pressing the sample against an O-ring that provides the vacuum seal. The gas flow of the cell was designed to provide transport of sputtered atoms into the observation zone, 1–2 cm from the sample. In 1987 Bernhard[65] took the idea of gas-assisted transport of atoms one step further, reporting on a design that used gas jets aimed at the sputtering surface to increase significantly the sampling rate as well as the absorption signal in a sputtering chamber. A commercial atomic absorption cell was designed based on this principle (Atomsource,[65] Analyte Corporation, Medford, OR), renewing the interest in AAS that began with Walsh[66] more than four decades ago. The source on the right is a much simpler atom generator. It is based on the direct insertion probe (DIP) design of King.[67] The coaxial cathode in King's original design has been replaced by a stainless steel ring that accommodates a demountable hollow cathode (4.82 mm in diameter × 2.54 mm in length with a 3.18-mm hole in the center). The DIP facilitates alignment of the HCL emission, which is focused through the orifice (i.e. the region of highest atom density) and collected after it passes through a window in a six-way vacuum cross. Typical operating conditions for this source are 1.0–3.0 Torr, 500–2000 V, and 2–15 mA. Detection limits for GD atomic absorption are in the low parts per million range. Although GD atomic absorption is not as widely used as flame or graphite furnace atomic absorption, it has found its niche in applications where analysis by other atomic absorption methods (primarily solution-based) are difficult (e.g. the analysis of materials that are difficult to dissolve).

3.4 Atomic Fluorescence Spectroscopy

Atomic fluorescence is similar to AAS in that both rely on an external light source to produce an analytical signal from an atomic vapor. In fluorescence spectroscopy, the signal is contained in the emission of photons from the atom population after absorption of the incident energy. There are five basic types of fluorescence: resonance, direct-line, step-wise, sensitized, and multiphoton fluorescence.[51] For this discussion, it is not important to define these five types, only to say that the differences lie in the excitation and relaxation pathways that each follow to produce fluorescence. Resonance fluorescence (where the same upper and lower levels are involved in the excitation–de-excitation process so absorption and emission wavelengths are the same) finds the most widespread use in analytical spectroscopy because the transition probabilities and the source radiances are the greatest for resonance fluorescence when conventional line sources are used.

Figure 10 shows a conventional instrumental arrangement for a single-beam atomic fluorescence spectrometer. Radiation from the source is focused into the GD. Fluorescence photons are imaged onto the entrance aperture of a monochromator that isolates the analyte fluorescence from background emission and fluorescence from other species. Fluorescence is usually viewed at an angle of 90° with respect to the excitation source to minimize the collection of scattered source radiation.

One critical component of an atomic fluorescence spectrometer is the excitation source. HCLs, electrodeless discharge lamps, and metal vapor discharge lamps have all been used successfully, although today most fluorescence experiments use a laser. Lasers are superior sources for atomic fluorescence because they provide a fluence several orders of magnitude greater than other sources, are tunable over a wide wavelength range, have spectral bandwidths much narrower than absorption line widths, and can be focused to very small spot sizes. Both continuous wave (with chopping) and pulsed lasers have been used, with dye lasers finding the most use because they can be tuned over a large number of wavelengths. An inductively coupled plasma has also been used as an excitation source; here, the excitation wavelength is governed by the analyte that is aspirated into the torch (usually at high concentrations).

Like AES and AAS, the ideal atomizer for fluorescence would produce a stable population of atoms of sufficient number density to make quantification of small concentrations practical. A fluorescence atomizer should also produce minimum thermal excitation to limit analyte and interference emission, a potential source of background. When this stipulation is met, Rayleigh scattering from atoms or molecules determines the fundamental limitation for background noise.

Most atomic fluorescence measurements have been made with flame atomizers,[51] but recently inductively coupled plasma has been used. Plasmas generally provide better atomization efficiency and a larger population of free atoms than flames. When analyzing a solid sample, however, an atomizer, like the one shown in Figure 10, is more practical.[68] This simple design consists of a Pyrex glass housing into which a 6.35-mm diameter sample rod is inserted through a ceramic sleeve and sealed to the cell with O-rings. Quartz windows are glued into the cell to allow a laser beam to pass. A fill port and an evacuation port are also provided, and the entire

cell is pumped by a single rotary vacuum pump. A DC power supply (Electronic Measurements, Eatontown, NJ) provides voltages up to 600 V and currents up to 200 mA. A typical operating voltage is 570 V at 25 mA. The cell is pressurized with argon to between 900 and 1000 Pa (7–8 Torr).

More recent applications of atomic fluorescence by this group[69] have produced GD cells constructed from high vacuum components like ConFlat® crosses and flanges with sapphire windows, but the principles of operation remain the same – fluorescence is detected 90° to the laser beam path by a photomultiplier tube. The fluorescence signal is amplified by a wideband amplifier and processed by a gated integrator and boxcar averager; the result is a fluorescence spectrum as a function of wavelength which is indicative of those elements in a GD cathode for which a fluorescence transition is allowed.

3.5 Optogalvanic Spectroscopy

The final optical technique to be discussed is optogalvanic spectroscopy. The optogalvanic effect was first observed using a weak, incoherent light source in 1928;[70] light from one neon discharge affected the electrical characteristics of a nearby discharge. The process is quite simple; in a discharge there is an equilibrium established between the neutral species and the corresponding ions. If some means of energy is added to the discharge, the equilibrium position can be displaced and the fraction of ions altered; this is the case when a photon is absorbed by the gaseous atom or molecule. This permits one to measure optical absorption by an all-electronic means, that is, without the use of photodetectors. The electrical circuit employed to monitor the optogalvanic effect commonly includes a ballast resistor in series with the discharge resistance; the discharge impedance change is usually monitored as a change in the discharge voltage. The light source is modulated, and a lock-in amplifier is employed to measure the alternating current component of the discharge voltage induced by absorption of light as the source wavelength is scanned. In theory, the atomization source could be any of the discharges discussed thus far; in practice, however, we have found that the demountable hollow cathode operating in the same fashion as it does for atomic absorption provides the greatest flexibility for optogalvanic spectroscopy.[71]

4 MASS SPECTROMETRIC METHODS OF ANALYSIS

Much like the spectrochemical techniques described above, mass spectrometry offers the analyst a method for determining the identity and quantity of a particular species in a sample. This technique however, provides

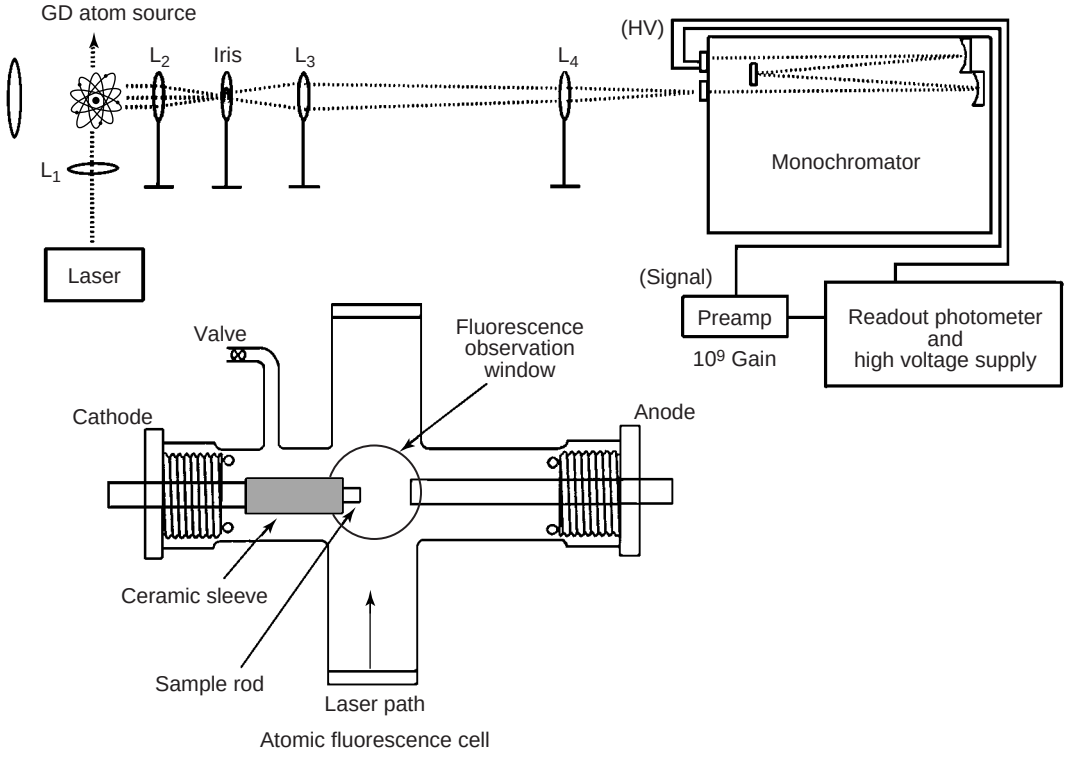

Figure 10 Schematic of a conventional instrumental arrangement for a single-beam atomic fluorescence spectrometer.

Table 4 Mass analyzers and considerations for elemental analysis

Mass analyzer type	Advantages	Disadvantages
Magnetic sector mass spectrometer	Commercially available Reasonable resolution	Scan speed Complex Cost
Quadrupole mass filter	Commercially available Robustness Scan speed Peak hopping mode	Limited resolution
QIT mass spectrometer	Cost CID to remove interferences	Complicated interface to GD source
FTICR mass spectrometer	High resolution CID capability	Cost Complex Complicated interface to GD source Space charge limited
TOF mass spectrometer	Simplicity Cost Speed Simultaneous data acquisition Good resolution when operated in reflectron mode	Some ion extraction biases Poor isotope ratio measurements

CID, collision-induced dissociation.

analytical information through the separation and subsequent detection of charged species associated with the sample. Ions are generated in the source region and selected by mass-to-charge ratio (m/z) usually using electrostatic or magnetic fields.

In this section, the basic requirements necessary to obtain mass abundance information are described, along with five types of mass spectrometers: quadrupole mass filters, magnetic sector mass analyzers, QITs, FTICR devices, and TOF mass spectrometers. These reviews will focus only on variants that have been coupled to the GD source. Table 4 provides basic information and characteristics that are unique to each of the mass spectrometric systems.

4.1 Basic Requirements Necessary to Obtain Mass Abundance Information

Mass spectrometric systems in general require a number of fundamental components to create, transport, separate, and detect ions, manipulate the resulting signal to account for system inconsistencies, and provide useful information to the analyst (see Figure 11). A number of these components will be considered in the following section.

4.1.1 Ion Source

There are a number of GD source geometries that can be implemented for a variety of analyses (see Table 1). Some types were designed, and are well suited, for optical applications as outlined in the previous section. Many of the advantages afforded by the optically applicable sources stem from an the enhancement of the atom population, often a result of increased sputtering rate or a confined viewing region that facilitates optical viewing. Mass spectrometry requires a different set of criteria, however. Sources designed for mass spectrometric applications must provide a reasonable population of analyte ions (not atoms). These ions then must be extracted from the source region into the mass analyzer. For these reasons, one appropriate source for mass spectrometry is the coaxial cathode. Ions generated by the coaxial cathode are extracted through an ion exit orifice in the anode. Ionization is dominated by the Penning process that leads to an ion population with a narrow KE spread relative to other GD source geometries.[72] This geometry also facilitates sample introduction via a DIP, making trivial the appropriate adjustment of the plasma relative to the ion exit orifice.

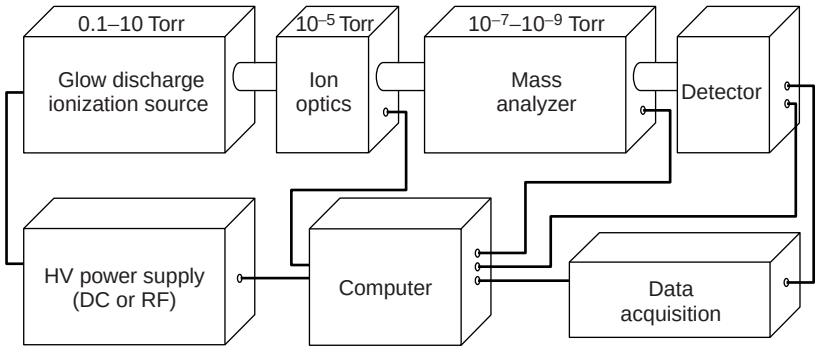

Figure 11 Block diagram of typical GD mass spectrometer system components.

4.1.2 Vacuum Systems

Mass spectrometric techniques impose stringent requirements on a vacuum system. One advantage of interfacing any of the currently available mass analyzers to the GD as compared to an atmospheric pressure source arises from the operating pressure of the ion source itself. Most mass spectrometers operate at pressures from 10^{-5} to 10^{-9} Torr. The GD source also operates at a reduced pressure (0.1–10 Torr with support gas), although not nearly as low as those required for mass analysis. The pressure differences associated with GD mass spectrometric implementation are overcome using differential pumping schemes. These types of systems employ a series of pumping regions to reduce the effect of the required pressure drop. A typical scheme would involve pumping the ion source with a rotary-vane roughing pump. This would evacuate the discharge cell to a pressure of approximately 10^{-3} Torr in the absence of discharge gas. The next differentially pumped region facilitates extraction of ions from the ion source. This region is often evacuated using a turbo molecular pump or oil diffusion pump and maintains pressures of 10^{-4}–10^{-5} Torr during discharge operation. The analyzer region, which includes the mass separation components as well as the detector, is evacuated using a turbo molecular, oil diffusion, cryo, or ion pump and maintains pressures of 10^{-7}–10^{-9} Torr.

4.1.3 Ion Optics

Ions generated in the GD plasma must be transported efficiently and indiscriminantly to the mass separation region to facilitate accurate and precise analytical measurements. The transfer of ions from the source to the mass analyzer is usually accomplished using an optics system. One type of system used for ion transport is the Einzel lens.[12,73,74] This type of lens is based on three conducting tubular lenses of similar dimension mounted in series. Typically, similar potentials are placed on the first and third lenses while that of the middle lens is adjustable. Many times the middle lens is held at ground potential. As an ion beam passes through this lens system it is focused in a manner analogous to an optical beam, reaching a focal point on the opposite side. It is important to note that cylindrical lens systems like the Einzel lenses produce a cylindrical ion beam rather than a planar one. This characteristic may limit their use for some applications such as ion guides for magnetic sector instruments which require a ribbon-shaped ion beam.

A second type of ion optic is the Bessel box.[75,76] This technology is used in conjunction with QMAs to remove photons and neutral species from the ion beam while simultaneously limiting the KE spread of ions entering the quadrupole lens region. If the energy spread is too great quadrupole performance deteriorates, resulting in degradation of mass resolution, peak splitting, and asymmetrical peak shapes.[77,78] A Bessel box is constructed from a square entrance and exit electrode surrounded by sets of electrodes on each of the other four sides. A center plate or cone is located within the box parallel to the entrance and exit electrodes. Potentials are applied to each electrode. These potentials can be varied to permit the transmission of ions with a discrete KE. Photons and neutral species will not be affected by these potentials and will proceed into the box linearly and collide with the center stop, removing them from the beam. Ions entering the box with too little KE will be repelled toward the entrance plate. If the energy of the ion is too great it will not be steered around the center stop and will collide with the side electrode or exit plate. Only ions with the selected KE will travel around the center stop, through the exit aperture and into the quadrupole region In most cases, one or two lenses are located behind the exit aperture to focus and transport the selected ions.

Quadrupole lenses can also be used as ion guides when operated in an RF-only mode.[79,80] It will be shown later in this chapter that the application of both DC and RF potentials provides a notch filter that can be adjusted and scanned to provide mass unit resolution. When only an RF potential is applied, the quadrupole acts as an ion guide, focusing all ions through quadrupole lenses.

4.1.4 Detection Systems

Three types of detection systems are routinely used for GD mass spectrometric measurements: Faraday cups, electron multipliers, and microchannel plates (MCPs). Detector selection is often independent of the mass analyzer in use.

The detector used most often for applications with high ion abundance is the Faraday cup. Modern Faraday detectors are extremely quiet. When operated using high grade resistors and amplification components, these detection systems offer state-of-the-art measurements with respect to signal-to-noise ratio. Although not currently available on commercial GD mass spectrometric systems, recent developments in multi-Faraday array detection systems offer increasingly precise measurements for scanning instruments by negating the effects of source fluctuations on measurements. Each Faraday cup in the array is dedicated and positioned to measure a single isotope at a given dispersion setting. This allows simultaneous collection of the selected ions during the acquisition sequence without scanning the mass dispersion device, virtually removing any dependence on fluctuations in ion beam intensity arising in the source. Minimization

of source fluctuation effects afforded by multicollector arrays is most important for applications involving the measurement of isotope ratios.

For applications requiring optimum sensitivity, discrete dynode electron multipliers operated in a pulse counting mode are required. Operation in this mode registers a signal pulse for every ion impinging on the first dynode. Each impinging ion generates a number of secondary electrons that are successively amplified by each dynode. Overall gains of 10^6–10^8 are common when using multipliers with 14–20 dynodes. After the pulse of electrons leaves the multiplier, it is amplified and proceeds to a discriminator that is set to remove pulses arising from dark noise. The signal is then sent to a universal counter that records each pulse, stores it for a given time, and passes it to a computer-based data acquisition system that presents the data in a usable form. It is important to keep the count rate low enough ($<10^6$ counts s^{-1}) to maintain the integrity of the pulse-counting system and to ensure that any pulses, not measured because of time lag in the electronics, will be statistically insignificant.

Daly detection systems behave in a manner similar to electron multipliers.[81] In the Daly system, the ion beam is accelerated to 10–20 kV and directed onto a highly polished aluminized steel electrode positioned directly behind the collector slit. Approximately eight electrons are liberated from the aluminized surface for every impinging ion. These electrons are repelled by the high negative potential applied to the electrode and directed onto a scintillator that produces a photon for each electron. These photons are counted by a photomultiplier located outside of the vacuum system. The resulting pulses are treated in a fashion analogous to that of the electron multiplier. Both Daly and electron multiplier-based detection systems can be operated in a mode that integrates the current of the impinging ion beam. This mode of operation is used for applications that do not impose such stringent sensitivity requirements.

A third detection system is used most often with TOF instruments and is built around an MCP detector. The MCP is characterized by a large, flat active area, high gain, and excellent time response. These operating parameters make it an ideal detector for TOF mass spectrometry. These detectors are fabricated from very thin glass wafers or plates perforated with microscopic channels oriented parallel to the impinging ion beam. The nature of the detector material is such that, in the presence of a potential bias (up to 1200 V), an ion impinging on the entrance to one of the channels will liberate one or more electrons which will, in turn, cascade through the channel, liberating further electrons with each wall collision. Amplification of the order of 10^4 is routinely realized for a single MCP. After emerging from the MCP, these electrons are collected by a positively charged electrode positioned parallel to the MCP. The resulting signal is further amplified and manipulated using a fast digitizing oscilloscope or any number of computer-based flash analog-to-digital converter (ADC) computer boards. Two or more MCPs can be arranged in a stack orientation to amplify further the signal thus increasing system performance.

4.2 Magnetic Sector Mass Analyzers

Initially used by Aston[82] for his studies of gaseous discharges, the magnetic sector mass analyzer, or mass spectrograph as Aston called it, is based on the spatial dispersion of ions with different m/z that is effected when they traverse an electromagnetic field. The magnetic field acts as a prism dispersing monoenergetic ions of differing m/z values across a focal plane; see Figure 12(a).[83] The radius of the curved flight path of an ion through a magnetic field is given by Equation (6):

$$r_m = \left(\frac{144}{B}\right)\left(\frac{mV}{z}\right)^{1/2} \qquad (6)$$

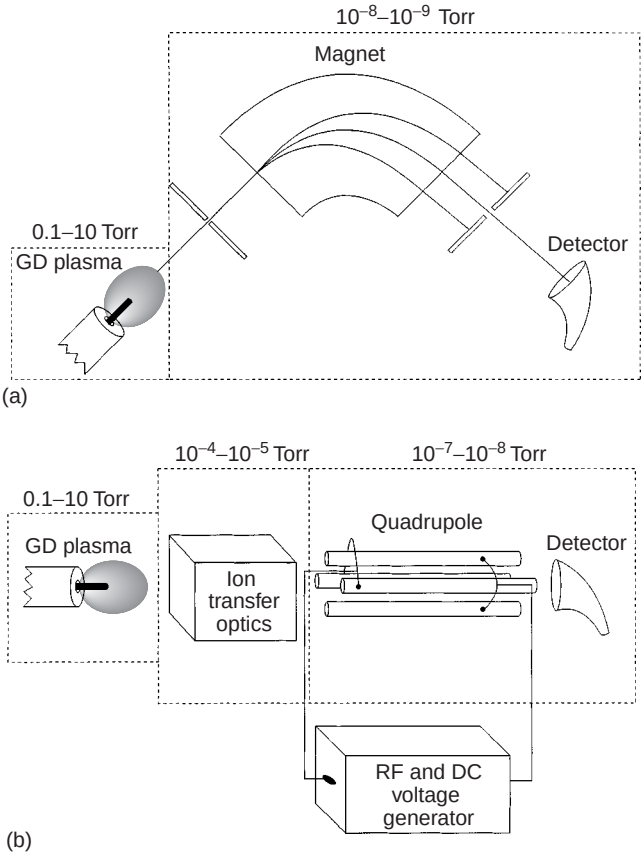

Figure 12 Schematic of (a) a magnetic sector mass spectrometer and (b) a QMA.

where B is the magnetic field strength in gauss, m is the atomic mass of the ions in amu, V is the acceleration voltage of the ion prior to entrance into the magnetic field in volts, z is the charge of the ion, and 144 is a constant prescribed by the units.[83] Because a sector instrument can be made to focus ions onto a plane, it can be designed as either a single or multi-collector instrument. A single electronic detector is used when operating in the sequential acquisition mode. Different m/z ions are brought into focus on the detector by varying either B or V. Simultaneous acquisition instruments utilize either a photographic plate or a detector array oriented in the focal plane. These systems allow the detection of a suite of different m/z ions at a given B and V setting. Although more expensive, the array-based systems offer shorter analysis times and the potential for more precise measurements because source fluctuation effects are minimized.

Mass resolution is another parameter that influences the ability of a system to solve an analytical problem. Mass resolution is a measure of the instrument's ability to separate ions having small mass differences. Mass resolution in magnetic sector instruments is defined by ion beam focusing in the focal plane. Single-focusing instruments rely on direction focusing to increase resolution. This is accomplished by narrowing the entrance and exit slit widths, thus reducing the width of the ion beam.[83] Mass resolution becomes more important when ions of the same nominal mass must be separated.

Multi-sector instruments are quite complex and expensive to build and maintain. They provide adequate to excellent resolving power, especially when an electrostatic sector is coupled to a magnetic sector to provide double (momentum and energy) focusing. When operated in the sequential detection mode, sector instruments are hindered by relatively slow scanning speeds which adversely affect analysis time and sample throughput. In the past, a number of commercial instruments were available from a variety of vendors based on the sector design. Currently, however, their availability has become somewhat limited.

4.3 Quadrupole Mass Filters

Since its development in the 1950s and early 1960s, the quadrupole mass spectrometer has become a powerful tool for the analysis of a variety of materials. Much of its popularity stems from the time of its development when it was viewed as a more rugged, more compact, and more cost-effective alternative to magnetic sector mass spectrometry systems, albeit with compromised performance. The quadrupole mass filter is a variable bandpass filtering ion optic, analogous to an optical bandpass filter. The quadrupole system is capable of transmission of all ions when operated in the RF-only mode (as described earlier), or of measuring only one m/z at a time as a sequential mass analyzer. The quadrupole offers the ability to scan the entire mass range very rapidly or to "peak-hop" among a series of selected isotopes.[84]

A quadrupole mass filter consists of four high precision, cylindrical, conducting rods or poles arranged in a square configuration, as shown in Figure 12(b). Mass filtering is accomplished by applying steady state DC and pulsed RF potenials to these poles. The application of these voltages results in the formation of hyperbolic electric fields, with the ideal quadrupole defined by Equation (7):

$$\frac{r}{r_0} = 1.148 \tag{7}$$

where r is the radius of each rod and r_0 is the radius enclosed by the electrodes.[85] The effects that the applied potentials and resulting electric fields have on a charged particle are best described by the ion trajectory in the $x-z$ and $y-z$ planes. The set of poles in the $x-z$ plane have a positive, time-independent DC voltage and a time-dependent RF voltage applied to them. The poles in the $y-z$ plane have a negative, time-independent DC voltage and a time-dependent RF voltage applied to them. The RF potential applied in the $y-z$ plane is 180° out of phase with the RF voltage in the $x-z$ plane. Ions enter and travel between the poles along the z-axis.

In the $x-z$ plane, larger mass ions are focused along the z-axis by the positively biased DC field, while the smaller mass ions are destabilized by the RF field. In the $y-z$ plane, larger mass ions are deflected away from the z-axis by the negatively biased potential while ions of smaller masses are stabilized by the RF field. The net result is a high-pass filter in the $x-z$ plane and a low-pass filter in the $y-z$ plane, allowing the stabilization and transmission of ions above a selected m/z, and the stabilization and transmission of ions below a selected m/z, respectively. When these two types of filters coexist, a narrow bandpass mass filter results. The magnitude of the DC potential and the frequency of the RF potential can be varied to allow transmission of different m/z ions through the quadrupole lenses, thus providing a means of m/z selection.

Quadrupole-based GD systems have been used extensively.[28,30] The relatively low cost and robustness of these instruments have made them an excellent choice for both the researcher and routine sample analyst. A major limitation of the quadrupole system is its relatively low resolving power, significantly increasing the deleterious effects of overlapping polyatomic interferences. Appropriate selection of operating conditions and discharge gas can minimize some of these concerns, but performance of quadrupoles does not match that of sector-based instruments.

4.4 Ion Trap and Fourier Transform Ion Cyclotron Resonance Devices

While sector and QMAs were being developed for elemental and isotopic applications, the QIT technology was being driven by needs in the organic community. FTICR mass spectrometry, another form of ion trapping technology, has filled a niche in the biological mass spectrometry community because of its unrivaled mass resolution.[86] Recently, these devices have been characterized and used by the inorganic mass spectrometric community. A number of external ionization sources have been implemented, including the GD.[86,87] This section will describe the operating principles of both the QIT and FTICR systems and cover some of their unique properties.

The operation of the QIT is very similar to that of the quadrupole mass filter. Two of the opposing rods in the quadrupole mass filter are connected to form a ring, and each of the remaining pair of rods is replaced by a hyperbolic endcap, as shown in Figure 13(a). The result is a three-dimensional quadrupole field that is symmetric with respect to rotation about the center. The end caps are oriented along the former z-axis, and the x- and y-axes become a plane, r, symmetric about z. In the most common inorganic applications, externally generated ions are collected and stored within the trap. Typically the internal pressure of the QIT is 10^{-3}–10^{-4} Torr He. Helium is introduced to cool collisionally the precessing ions, thus allowing them to relax toward the center of the trap and increase trapping efficiency.

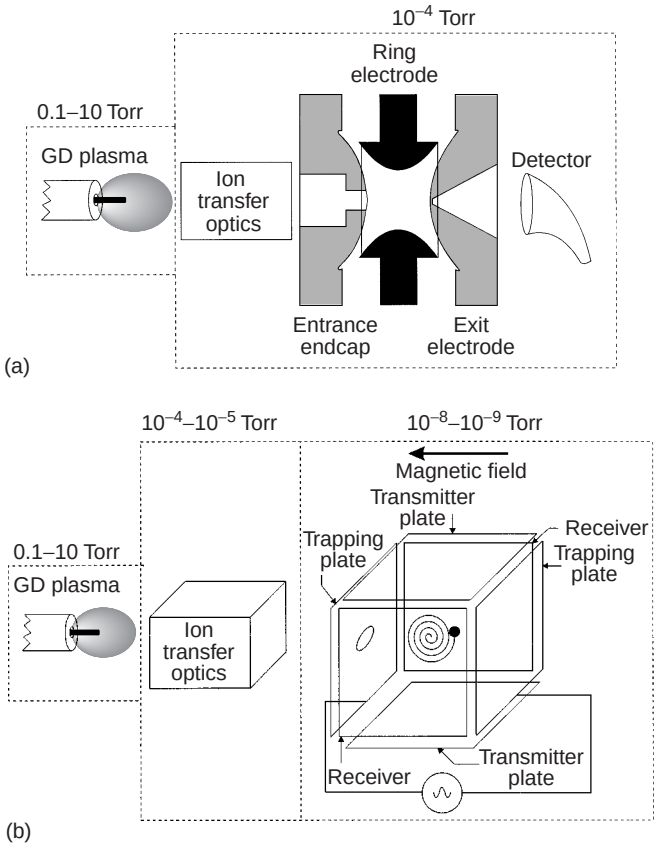

Figure 13 Schematic of (a) a QIT mass spectrometer and (b) an FTICR mass spectrometer.

Stored ions precess with a frequency that is m/z dependent. Ions that are stored in the trapping fields have a fundamental secular frequency along the axial or z-axis. When a supplemental AC signal is applied to the end-cap electrodes, ions whose secular frequencies are in resonance with the applied frequency are excited to higher translational energies. The magnitude of this resonance excitation is directly proportional to the amplitude of the applied signal. At appropriately high amplitudes, the ions can either be lost in collisions with an electrode or ejected through apertures in the exit end-cap electrode. Mass analysis can thus be performed by scanning the frequency of the AC excitation signal. During a scan, ions become destabilized through the excitation process and are selectively ejected through the exit end-cap electrode and are then detected using an electron multiplier.

The relatively high operating pressure of the QIT makes it well suited for coupling to ion sources that operate at higher pressures, such as the GD. Ions from the GD can be directly introduced into the QIT without the need for an elaborate differential pumping scheme. The high pressure and trapping nature of the QIT also facilitate the use of ion–molecule interactions to provide a number of advantageous results, most notably the suppression of unwanted contributions from isobaric interferences.

The FTICR mass spectrometer was initially considered by the elemental mass spectrometry community for its high resolving power. This characteristic allows the physical separation of the analyte signal from interferences without actually removing the interfering species from the cell. High resolution does not come without a cost, however: FTICR systems are one of the most expensive types of mass spectrometers available. The basic operation of the FTICR system is similar to that of the QIT. Precessing ions are constrained spatially within a cubic cell using both electric and magnetic fields. A homogeneous magnetic field confines ions radially, while electrostatic potentials are applied to the end-caps of the cell to trap the ions axially; see Figure 13(b). Ions trapped by these fields are characterized by three motions: one that confines the ions between the two end-cap electrodes, magnetron motion, and cyclotron motion.[88] The ions orbit perpendicular to the applied magnetic field at the ion characteristic cyclotron frequencies, T_c, that are inversely proportional to their m/z values and proportional to the magnetic field strength, B,[89] as shown

in Equation (8)

$$w_c = \frac{zB}{m} \tag{8}$$

Ions are excited to larger cyclotron orbits by the application of a resonant RF potential to transmitter plates. An image current is generated as the coherent ion packets of a given m/z come into close proximity to the receiver electrodes. This current is converted to a voltage, amplified, digitized, and stored as a transient signal. The Fourier transform converts these transient signals into their frequency components, T_c, which are, in turn, related to the m/z as shown in Equation (8). Extremely high mass resolution can be achieved because the basis for mass measurement lies in measuring frequency, which can be done with great precision.[90] It should also be noted that mass resolving power is inversely proportional to the ion m/z,[91] a fortunate circumstance for elemental analysis because all masses of interest are below 250 Da.

4.5 Time-of-flight Mass Spectrometers

Perhaps the simplest type of mass spectrometer, the TOF mass spectrometer, is depicted in Figure 14. Its operation is based on the KE equation, where KE is a function of mass, m, and velocity, v; alternatively it can be expressed in terms of charge, z, and accelerating potential, V, as shown in Equation (9):

$$\mathrm{KE} = \tfrac{1}{2}mv^2 = zV \tag{9}$$

From this equation one can deduce that ions of different m/z accelerated to a common KE will have different velocities. The TOF instrument operates on the principle that if all ions leave the extraction grid at the same point in space and time with equal KEs, they will travel with different velocities, v_y, that are inversely proportional to their respective masses, m_y, as shown in Equation (10):

$$v_y = \left(\frac{2\mathrm{KE}}{m_y}\right)^{1/2} \tag{10}$$

The time needed to traverse the flight path distance D and to arrive at the detector is related to the m_y/z ion by Equation (11):

$$t_y = D\left(\frac{m_y}{2zV}\right)^{1/2} \tag{11}$$

Monitoring the current at the detector (often an MCP) yields a time-dependent signal that can be correlated to the m/z using Equation (11). Mass resolution is determined directly from the temporal resolution $\Delta t/t$, which is determined by the initial KE spread of the ions and the speed of the detection electronics.

The effects of this spread in initial energy can be minimized by using a reflectron TOF mass spectrometric instrument. This instrument geometry has a series of electrostatic lenses located at the end of the flight tube. Positive potentials applied to these lenses increase toward the detection end of the flight tube. Ions traveling down the flight tube enter this potential gradient and penetrate, slowing until they reach a point in the gradient equal to their initial KE. They are then accelerated in the opposite direction, traversing the flight tube a second time. They are then detected by an MCP located at the base of the flight tube. Ions of higher KE penetrate further into the reflectron field, increasing their flight distance (time) and effectively minimizing the impact on mass resolution.

TOF instruments are ideally suited for pulsed ion sources such as lasers or pulsed ion beam sputtering. SIMS using TOF technology has gained wide acceptance as a technique for the characterization of a wide variety of materials. TOF mass spectrometry is also well suited for use with the pulsed GD source.[37] The TOF mass spectrometer provides simultaneous detection of all ions with each injection pulse. This quality affords an excellent diagnostic tool for plasmas and is essential for the temporal characterization of millisecond and microsecond GD pulses.[37,92] Although no commercial GD TOF mass spectrometric systems are

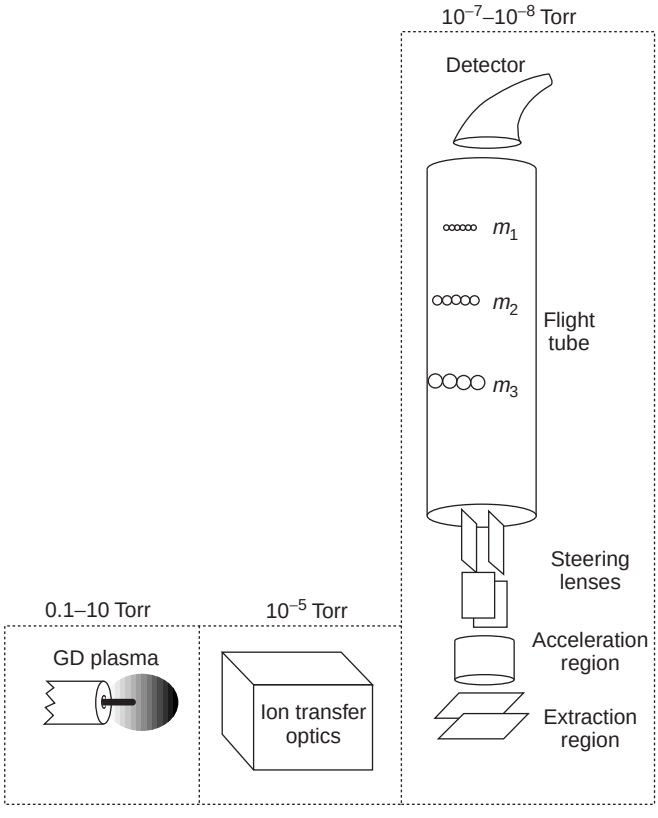

Figure 14 Schematic of a TOF mass spectrometer.

currently available, a number are being used for research activities.[37,38]

5 CONCLUSIONS

Optical and mass spectrometric techniques using GD ion sources are now established as routine methods for the direct analysis of solid samples of widely different origin and composition. This article has focused on the fundamental operation of the GD source in a number of operating modes as well as the optical and mass spectrometric instrumentation used for elemental determination. These source–instrument combinations will undoubtedly continue to offer advantages essential for specific applications in the future.

DISCLAIMER

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the US Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the US Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for US Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the US Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
ADC	Analog-to-digital Converter
AES	Atomic Emission Spectroscopy
AFS	Atomic Fluorescence Spectroscopy
CID	Collision-induced Dissociation
DIP	Direct Insertion Probe
FTICR	Fourier Transform Ion Cyclotron Resonance
GD	Glow Discharge
HCL	Hollow Cathode Lamp
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
KE	Kinetic Energy
MCP	Microchannel Plate
QIT	Quadrupole Ion Trap
QMA	Quadrupole Mass Analyzer
RF	Radiofrequency
SIMS	Secondary Ion Mass Spectrometry
TOF	Time-of-flight
UV	Ultraviolet

RELATED ARTICLES

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Laser Mass Spectrometry in Trace Analysis

Environment: Water and Waste **(Volume 3)**
Atomic Fluorescence in Environmental Analysis • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Luminescence in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications • Mass Spectrometry for Forensic Applications

Atomic Spectroscopy **(Volume 11)**
Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Ablation in Atomic Spectroscopy

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. F. Paschen, 'Bohrs Heliumlinien', *Ann. Phys.*, **50**, 901–940 (1916).

2. H. Schuler, 'Uber die Anregung von Spektren zur Untersuchung von Hyperfeinstrukturen', *Z. Phys.*, **59**, 149–153 (1929).
3. H. Schuler, J.E. Keystone, 'Hyperfeinstrukturen und Kernmomente des Quecksilbers', *Z. Phys.*, **72**, 423–441 (1931).
4. B. Chapman, *Glow Discharge Processes, Sputtering and Plasma Etching*, John Wiley & Sons, New York, 1980.
5. A.M. Howaston, *An Introduction to Gas Discharges*, 2nd edition, Pergamon Press, New York, 1976.
6. R.K. Marcus (ed.), *Glow Discharge Spectroscopies*, Plenum Press, New York, 1993.
7. W.W. Harrison, C.M. Barshick, J.A. Klingler, P.H. Ratliff, Y. Mei, 'Glow Discharge Techniques in Analytical Chemistry', *Anal. Chem.*, **62**, 943A–949A (1990).
8. R.K. Marcus, T.R. Harville, Y. Mei, C.R. Shick, 'RF-powered Glow-discharges: Elemental Analysis Across the Solids Spectrum', *Anal. Chem.*, **66**, 902A–911A (1994).
9. P.W.J.M. Boumans, 'Studies of Sputtering in a Glow Discharge for Spectrochemical Analysis', *Anal. Chem.*, **44**, 1219–1228 (1972).
10. F.L. King, J. Teng, R.E. Steiner, 'Glow Discharge Mass Spectrometry: Trace Elemental Determinations in Solid Samples', *J. Mass Spectrom.*, **30**, 1061–1075 (1995).
11. T.D. Mark, G.H. Nunn, *Electron Impact Ionization*, Springer-Verlag, New York, 1985.
12. L. Valyi, *Atom and Ion Sources*, John Wiley & Sons, New York, 1977.
13. F.H. Field, J.L. Franklin, *Electron Impact Phenomena*, Academic Press, New York, 1970.
14. J. Delcroix, C. Ferreira, A. Richard, 'Metastable Atoms and Molecules in Ionized Gases', in *Principles of Laser Plasmas*, ed. G. Bekefi, John Wiley & Sons, New York, 159–233, 1976.
15. V.S. Bordin, Y.M. Kagan, 'Excitation of Helium in a Hollow-cathode Discharge', *Opt. Spectrosc.*, **23**, 108–110 (1967).
16. F.M. Penning, 'Uber die Erhehung der Zundspannung von Neon–Argon Gemischen durch Bestrahlung', *Z. Phys.*, **57**, 723–738 (1929).
17. F.M. Penning, 'Uber den Einfluss Sehr Gerginger Beimischungen auf die Zundspannung der Edelgase', *Z. Phys.*, **46**, 335–348 (1925).
18. D.H. Stedman, D.W. Stetser, 'Chemical Applications of Metastable Rare Gas Atoms', *Prog. React. Kinet.*, **6**, 193–238 (1971).
19. R.L. Smith, D. Serxner, K.R. Hess, 'Assessment of the Relative Role of Penning Ionization in Low-pressure Glow Discharges', *Anal. Chem.*, **61**, 1103–1108 (1989).
20. M. Saito, 'Relative Sensitivity Factors in Direct Current Glow Discharge Mass Spectrometry Using Kr and Xe Gas – Estimation of the Role of Penning Ionization', *Fresenius' J. Anal. Chem.*, **351**, 148–153 (1995).
21. S. Degent, W.W. Harrison, 'Parameter Evaluation for the Analysis of Oxide-based Samples with Radio-frequency Glow Discharge Mass Spectrometry', *Anal. Chem.*, **67**, 1026–1033 (1995).
22. C. Lazik, R.K. Marcus, 'Effect of Excitation-frequency on Source Characteristics in Radio-frequency Glow-discharge Atomic-emission Spectrometry, 1.2–30 MHz', *Spectrochim. Acta*, **49B**, 649–663 (1994).
23. D.C. Duckworth, D.L. Donohue, D.H. Smith, T.A. Lewis, R.K. Marcus, 'Design and Characterization for a Radio-frequency-powered Glow-discharge Source for Double-focusing Mass Spectrometers', *Anal. Chem.*, **65**, 2478–2484 (1993).
24. J. Teng, C.M. Barshick, D.C. Duckworth, S.J. Norton, D.H. Smith, F.L. King, 'Factors Influencing the Quantitative Determination of Trace Elements in Soils by Glow-discharge Mass Spectrometry', *Appl. Spectrosc.*, **49**, 1361–1366 (1995).
25. D.C. Duckworth, C.M. Barshick, D.A. Bostick, D.H. Smith, 'Direct Measurement of Uranium Isotopic-ratios in Soils by Glow-discharge Mass Spectrometry', *Appl. Spectrosc.*, **47**, 243–245 (1993).
26. D.C. Duckworth, C.M. Barshick, D.H. Smith, 'Analysis of Soils by Glow-discharge Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **8**, 875–879 (1993).
27. D.C. Duckworth, R.K. Marcus, 'Radio-frequency Powered Glow Discharge Atomization Ionization Source for Solids Mass Spectrometry', *Anal. Chem.*, **61**, 1879–1886 (1989).
28. F.L. King, C. Pan, 'Time-resolved Studies of Ionized Sputtered Atoms in Pulsed Radio-frequency Powered Glow-discharge Mass Spectrometry', *Anal. Chem.*, 3187–3193 (1993).
29. J.A. Klingler, P.J. Savickas, W.W. Harrison, 'The Pulsed Glow-discharge as an Elemental Ion-source', *J. Am. Soc. Mass Spectrom.*, **1**, 138–143 (1990).
30. J.A. Klingler, C.M. Barshick, W.W. Harrison, 'Factors Influencing Ion Signal Profiles in Pulsed Glow-discharge Mass-spectrometry', *Anal. Chem.*, **63**, 2571–2576 (1991).
31. M.R. Winchester, R.K. Marcus, 'Emission Characteristics of a Pulsed, Radio-frequency Glow Discharge Atomic Emission Device', *Anal. Chem.*, **64**, 2067–2074 (1992).
32. M. Glick, B.W. Smith, J.D. Winefordner, 'Laser-excited Atomic Fluorescence in a Pulsed Hollow-cathode Glow Discharge', *Anal. Chem.*, **62**, 157–161 (1990).
33. C. Pan, F.L. King, 'Ion Formation Processes in the Afterpeak Time Regime of Pulsed Glow Discharge Plasmas', *J. Am. Soc. Mass Spectrom.*, **4**, 727–732 (1993).
34. W.W. Harrison, K.R. Hess, R.K. Marcus, F.L. King, 'Glow-discharge Mass-spectrometry', *Anal. Chem.*, **58**, 341A–356A (1986).
35. M.A. Biondi, 'Studies of the Mechanism of Electron–Ion Recombination', *Phys. Rev.*, **129**, 1181–1188 (1963).
36. M.A. Biondi, 'Diffusion, De-excitation and Ionization Cross-sections for Metastable Atoms', *Phys. Rev.*, **88**, 660–665 (1952).

37. R.E. Steiner, C.L. Lewis, F.L. King, 'Time-of-flight Mass Spectrometry with a Pulsed Glow Discharge Ionization Source', *Anal. Chem.*, **69**, 1715–1721 (1997).
38. C.R. Shick, P.A. DePalma, R.K. Marcus, 'Radio-frequency Glow-discharge Mass-spectrometry for the Characterization of Bulk Polymers', *Anal. Chem.*, **68**, 2113–2121 (1996).
39. X.H. Pan, R.K. Marcus, 'Direct Analysis of Glass Powder Samples by Radio-frequency Glow-discharge Atomic Emission Spectrometry (RF-GD-AES)', *Mikrochim. Acta*, **129**, 239–250 (1998).
40. J.A.C. Broekaert, T. Graule, H. Jenett, G. Tölg, P. Tschopel, 'Analysis of Advanced Ceramics and their Basic Products', *Fresenius Z. Anal. Chem.*, **332**, 825–838 (1989).
41. M. Dogar, K. Laqua, H. Massmann, 'Spektrochemische Analysen mit einer Glimmentladungslampe als Lichtquelle-II Analytische Anwengdungen', *Spectrochim. Acta*, **27B**, 65–88 (1972).
42. G.S. Lomdahl, R. McPherson, J.V. Sullivan, 'The Atomic Emission Spectrometric Determination of Non-conducting Materials with a Boosted Output Glow-discharge Source', *Anal. Chim. Acta*, **148**, 171–180 (1983).
43. T.J. Loving, W.W. Harrison, 'Dual-pin Cathode Geometry for Glow-discharge Mass-spectrometry', *Anal. Chem.*, **55**, 1526–1530 (1983).
44. S. Caroli, A. Almonti, K. Zimmer, 'Applicability of a Hollow-cathode Emission Source for Determining Trace-elements in Electrically Non-conducting Powders', *Spectrochim. Acta*, **38B**, 625–631 (1983).
45. W. Schelles, R.E. Van Grieken, 'Direct Current Glow Discharge Mass Spectrometric Analysis of Macor Ceramic Using a Secondary Cathode', *Anal. Chem.*, **68**, 3570–3574 (1996).
46. D.M. Wayne, R.K. Schulze, C. Maggiore, D.W. Cooke, G. Havrilla, 'Characterization of Tantalum Films on Analytical Surfaces: Insights into Sputtering of Nonconductors in a Direct-current Glow Discharge Using Secondary Cathodes', *Appl. Spectrosc.*, **53**, 266–277 (1999).
47. W. Schelles, R.E. Van Grieken, 'Quantitative Analysis of Zirconium-oxide by Direct Current Glow-discharge Mass-spectrometry Using a Secondary Cathode', *J. Anal. At. Spectrom.*, **12**, 49–52 (1997).
48. C.M. Barshick, W.W. Harrison, 'The Laser as an Analytical Probe in Glow-discharge Mass-spectrometry', *Mikrochim. Acta*, **III**, 169–177 (1989).
49. C.M. Strange, R.K. Marcus, 'Aqueous Sample Introduction into a Glow-discharge Device via a Particle Beam Interface', *Spectrochim. Acta*, **46B**, 517–526 (1991).
50. R.E. Steiner, C.L. Lewis, V. Majidi, 'Consideration of a Millisecond Pulsed Glow Discharge Time-of-flight Mass Spectrometer for Concurrent Elemental and Molecular Analysis', *J. Anal. At. Spectrom.*, **14**, 1537–1541 (1999).
51. J.D. Ingle, Jr, S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, Englewood Cliffs, 1988.
52. Y. Talmi, *Multichannel Image Detectors*, ACS Symposium Series 236, American Chemical Society, Washington, DC, Vol. 2, 1983.
53. Q.S. Hanley, C.W. Earle, F.M. Pennebaker, S.P. Madden, M.B. Denton, 'Charge Transfer Devices in Analytical Instrumentation', *Anal. Chem.*, **68**, 661A–667A (1996).
54. J.H. Giles, T.D. Ridder, R.H. Williams, D.A. Jones, M.B. Denton, 'Selecting a CCD Camera', *Anal. Chem.*, **70**, 663A–668A (1998).
55. E.U. Condon, G.H. Shortley, *The Theory of Atomic Spectra*, Cambridge University Press, Cambridge, 1963.
56. W.F. Meggers, C.H. Corliss, B.F. Scribner, *Tables of Spectral-line Intensities: Part II – Arranged by Wavelengths*, US Government Printing Office, Washington, 1975.
57. W. Grimm, 'Eine neue Glimmentladungslampe fur die Optische Emissionsspektralanalyse', *Spectrochim. Acta*, **23B**, 443–454 (1968).
58. J.A.C. Broekaert, 'Atomic Emission Spectroscopy', in *Glow Discharge Spectroscopies*, ed. R. K. Marcus, Plenum Press, New York, 113–174, 1993.
59. H. Hocquaux, 'Thin Film Analysis', in *Glow Discharge Spectroscopies*, ed. R.K. Marcus, Plenum Press, New York, 329–372, 1993.
60. E.H. Daughtrey, D.L. Donohue, P.J. Slevin, W.W. Harrison, 'Surface Sputter Effects in a Hollow-cathode Discharge', *Anal. Chem.*, **47**, 683–688 (1975).
61. W.W. Harrison, B.L. Bentz, 'Glow-discharge Mass-spectrometry', *Prog. Anal. Spectrosc.*, **11**, 53–110 (1988).
62. P.F. Little, A. von Engel, 'The Hollow-cathode Effect and the Theory of Glow Discharges', *Proc. R. Soc. London*, **224A**, 209–227 (1954).
63. S.L. Mandelstam, V.V. Nedler, 'On the Sensitivity of Emission Spectrochemical Analysis', *Spectrochim. Acta*, **17**, 885–894 (1961).
64. D.S. Gough, 'Direct Analysis of Metals and Alloys by Atomic-absorption Spectrometry', *Anal. Chem.*, **48**, 1926–1931 (1976).
65. A.E. Bernhard, 'Atomic Absorption Spectrometry Using Sputtering Atomization of Solid Samples', *Spectroscopy*, **2**, 24–27 (1987).
66. A. Walsh, 'The Application of Atomic Absorption Spectra to Chemical Analysis', *Spectrochim. Acta*, **7**, 108–117 (1955).
67. F.L. King, PhD Dissertation, University of Virginia, 1989.
68. B.M. Patel, B. Smith, J.D. Winefordner, 'Laser-excited Fluorescence of Diatomic Lead in a Glow-discharge Source', *Spectrochim. Acta*, **40B**, 1195–1204 (1985).
69. J.B. Womack, E.M. Gessler, J.D. Winefordner, 'Atomic Fluorescence in a Pulsed Hollow-cathode Discharge with a Copper Vapor Pumped Dye Laser', *Spectrochim. Acta*, **46B**, 301–308 (1991).
70. F.M. Penning, 'Demonstration of a New Photoelectric Effect', *Physica*, **8**, 137–140 (1928).

71. C.M. Barshick, R.W. Shaw, J.P. Young, J.M. Ramsey, 'Isotopic Analysis of Uranium Using Glow-discharge Optogalvanic Spectroscopy and Diode Lasers', *Anal. Chem.*, **66**, 4154–4158 (1994).
72. W.A. Mattson, B.L. Bentz, W.W. Harrison, 'Coaxial Cathode Ion-source for Solids Mass Spectrometry', *Anal. Chem.*, **48**, 489–491 (1976).
73. O. Klemperer, *Electron Optics*, Cambridge University Press, Cambridge, 1971.
74. A. Septier, *Focusing of Charge Particles*, Academic Press, New York, 1967.
75. B.L. Bentz, PhD Dissertation, University of Virginia, 1980.
76. *Electrostatic Energy Analyzer Model 616-1 Instruction Manual*, Etranuclear Laboratories, Inc., Pittsburgh, 1977.
77. S.S. Medley, 'Energetic Ion Mass Analysis Using a Radio-frequency Quadrupole Filter', *Rev. Sci. Instrum.*, **49**, 698–706 (1978).
78. W. Paul, H.P. Reinhard, U. VanZahn, 'Das Elektrische Massenfilter als Massenspektrometer und Isotopentrenner', *Z. Phys.*, **152**, 143–182 (1958).
79. M. Szilagyi, *Electron and Ion Optics*, Plenum Press, New York, 1988.
80. H. Wollnik, *Optics of Charged Particles*, Academic Press, Inc., Orlando, 1987.
81. N.R. Daly, 'Scintillation Type Mass Spectrometer Ion Detector', *Rev. Sci. Instrum.*, **31**, 264–267 (1960).
82. F.W. Aston, *Isotopes*, 2nd edition, Longman, New York, 1924.
83. J. Roboz, *Introduction to Mass Spectrometry: Instrumentation and Techniques*, Interscience, New York, 1968.
84. R.S. Houk, 'Mass-spectrometry of Inductively Coupled Plasmas', *Anal. Chem.*, **58**, 97A–105A (1986).
85. P.H. Dawson, *Quadrupole Mass Spectrometry and its Applications*, Elsevier Scientific, New York, 1976.
86. M.V. Buchanan, R.L. Hettich, 'Fourier-transform Mass-spectrometry of High-mass Biomolecules', *Anal. Chem.*, **65**, 245A–259A (1993).
87. S.A. McLuckey, G.J. Van Berkel, D.E. Goeringer, G.L. Glish, 'Ion-trap Mass-spectrometry of Externally Generated Ions', *Anal. Chem.*, **66**, 689A–696A (1994).
88. A.G. Marshall, P.B. Grosshans, 'Fourier-transform Ion-cyclotron Resonance Mass-spectrometry: the Teenage Years', *Anal. Chem.*, **63**, 215A–229A (1991).
89. N.M.M. Nibbering, 'Gas-phase Ion Molecule Reactions as Studied by Fourier-transform Ion-cyclotron Resonance', *Acc. Chem. Res.*, **23**, 279–285 (1990).
90. D.C. Duckworth, C.M. Barshick, 'Ion Traps: What do they Hold for Elemental Analysis', *Anal. Chem.*, **70**, 709A–717A (1998).
91. M.B. Comisarow, A.G. Marshall, 'Theory of Fourier-transform Ion-cyclotron Resonance Mass-spectrometry 1. Fundamental Equations and Low Pressure Line-shape', *J. Chem. Phys.*, **64**, 110–119 (1976).
92. W.W. Harrison, W. Hang, 'Pulsed Glow Discharge Time-of-flight Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 835–840 (1996).

Graphite Furnace Atomic Absorption Spectrometry

James A. Holcombe
University of Texas at Austin, Austin, USA

1 Introduction and Overview 1
2 Brief History 2
3 Theory 5
3.1 Absorption Fundamentals 5
3.2 Production of the Atomic Absorption Signal 6
3.3 Basic Vaporization (Desorption) Process 6
3.4 Preatomization Reactions 8
4 Instrumentation 10
4.1 Light Sources 10
4.2 Background Correction 11
4.3 Atomizer 11
4.4 Platforms and Probes 12
4.5 Spectrometer and Optics 12
4.6 Signal Processing 12
4.7 Autosamplers and Other Accessories 13
5 Matrix Modifiers 13
6 Practical Application of Graphite Furnace Atomic Absorption Spectrometry to Analysis 14
6.1 Setting Up Temperature Programs 14
6.2 Sample Types 15
6.3 Calibration Curves 15
6.4 Absolute Analysis 16
7 Conclusion 16
Acknowledgments 16
Abbreviations and Acronyms 16
Related Articles 17
References 17

Graphite furnace atomic absorption spectrometry (GFAAS) is an atomic spectroscopic technique in which a small sample is placed inside a graphite tube that is then resistively heated to accomplish sample desolvation (for liquid samples), ashing or charring (to decompose the sample and volatilize some of the matrix) and finally atomization. The light from a line source characteristic of the element being determined is passed longitudinally through the tube and the absorbance resulting from the presence of free analyte atoms in the gas phase is measured. The signal is transient in character, lasting approximately 1–5 s, and the area under this transient peak is generally used in the construction of a calibration curve. Modern instrumentation provides high levels of automation with capabilities of background correction as well as routine methods of sample analysis, for example quality assurance/quality control (QA/QC), standard additions, matrix modification, etc. Since the technique was first introduced in 1969, a lot of progress has been made in understanding the processes occurring within the graphite tube atomizer that ultimately produces the absorbance signal. An understanding of the free atoms formation process has facilitated the application of the technique to the analysis of a variety of complex samples.

Graphite furnace (also known as an electrothermal atomizer or ETA) atomic absorption (AA) is generally considered an ultratrace and microtrace analytical technique with limits of detection (LODs) in the low picogram range, precision of a few percent (relative standard deviation) and a dynamic range of about three orders of magnitude. In addition to its excellent sensitivity, it is unique in its ability to handle microsamples including aqueous solutions, viscous liquids, slurries and even solids. In general, there is considerable literature detailing methods and procedures for the determination of a variety of analytes in complex matrices that can be used by the analyst to apply the approach to new, complex analytical needs. When used correctly, this analytical tool can provide precise, accurate analysis for a wide range of sample types.

1 INTRODUCTION AND OVERVIEW

What is commonly referred to as GFAAS has been referred to as flameless AA in the past and is officially designated by IUPAC as electrothermal atomic absorption spectrometry (ETAAS). Typically, a commercial ETA consists of a graphite tube approximately 2.5–3 cm in length and 4–6 mm inside diameter. There is a small (1–2 mm diameter) hole in the center of the tube for introducing the sample, which is generally a solution aliquot of up to ca. 50 μL and is more typically 20–25 μL. Figure 1 shows a simplified schematic of the furnace and some of its typical operating features. Commercial tube furnaces are ca. 2.5 cm long and 0.6 cm in diameter and are composed of graphite with a relatively impervious pyrographite coating. The tube is resistively heated by an external supply and is protected from oxidation by a flow of sheath gas (typically Ar) around the furnace (not shown in Figure 1). The light source passes through the furnace and Beer–Lambert's law relates the absorbance to the concentration. Shown in the graphical inserts are the temperature (heating) program of the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

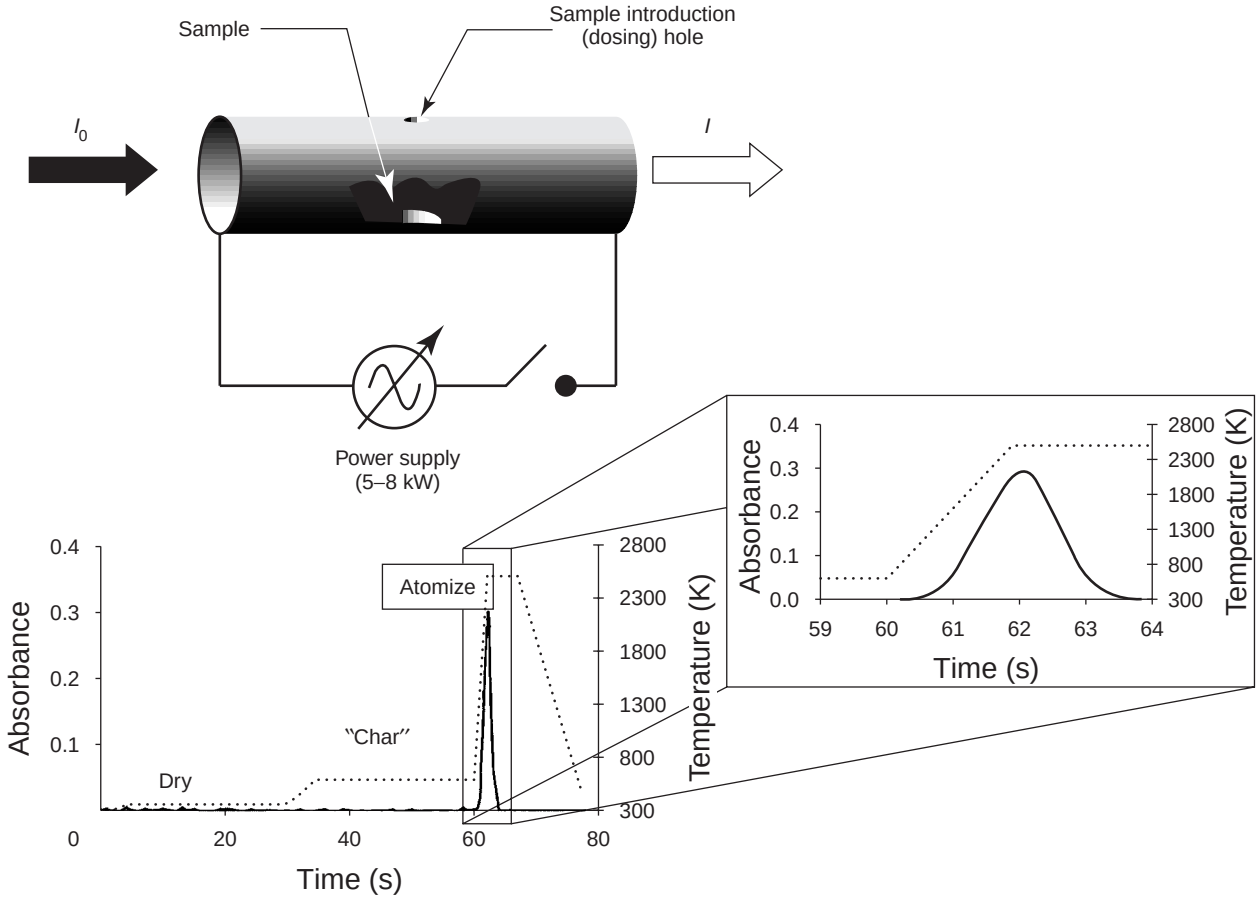

Figure 1 Conceptual diagram of a tubular ETA or graphite furnace atomizer. The atomizer is connected to a variable voltage power source and is resistively heated by current passing through the electrothermal vaporizer. Shown are the three basic heating regimes: dry, char and atomization cycles. The light source is passed through the tube and during the high temperature ''atomization cycle'' the analyte is vaporized and atomized. As a consequence an absorbance signal is registered.

furnace and the absorbance signal that results once the ''atomization cycle'' is reached. As can be seen, the signal is transient in character and can last up to several seconds.

2 BRIEF HISTORY

The ability of free atoms in the gas phase to absorb characteristic wavelengths of light is not a new concept. In fact, the absence of discrete wavelengths in the yellow region of the solar spectrum was attributed to the absorption of these wavelengths by sodium as early as 1860 by Kirchoff. Walsh[1] is credited with the first practical employment of AA spectrometry as a quantitative analytical tool using a flame as the ''atomic cell'' where free gaseous atoms are produced. Light from a forerunner of today's hollow cathode lamp (HCL) was passed through the flame, the correct wavelength of radiation was isolated and detected. The attenuation of the light was then used for determining the concentration of analyte in the sample solution.

Prior to Wash's publication, a large, heated graphite tube (''King furnace'') had been used for basic studies of the absorption characteristics of free atoms and molecules which were produced continuously from a large amount of material located within the furnace.[2] However, there was no attempt to use this for quantitative or trace analysis, and there was no suggestion that this might even be possible.

L'vov[3] was the first to combine the concepts of Walsh with a significantly altered version of the King furnace to demonstrate that quantitative analysis by AA could be conducted on discrete sample amounts by pulsed vaporization of the sample. Figure 2 shows a picture of this first graphite furnace atomizer for spectrochemical analysis by AA. This approach not only provided quantitative results but gave significantly improved sensitivities with LODs in the picogram range for many elements because of the significantly longer

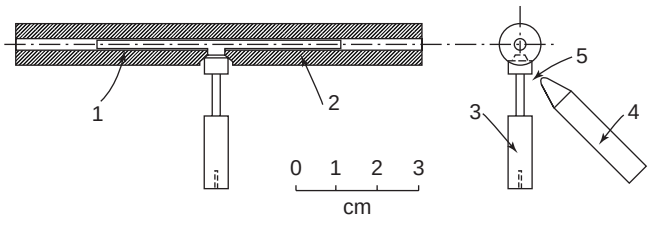

Figure 2 L'vov's original ''graphite furnace atomizer. The sample was deposited on the ''plug'' (3) which was inserted into the preheated graphite tube (1). An arc (5) was then struck with an auxiliary electrode (4) to rapidly heat the plug and release the sample into the tube and optical path. A tantalum liner (2) was used in this original furnace to minimize analyte loss from vapor diffusion through the graphite (Modern furnaces use pyrolytically coated graphite through which minimal vapor diffusion occurs). (Reprinted from B.V. L'vov, 'The Analytical Use of Atomic Absorption Spectra', *Spectrochim. Acta*, **17**, 761–770, Copyright (1961), with permission from Elsevier Science.)

residence time within the furnace in comparison to a flame. This paper and the promising results spurred a great deal of activity in ''flameless'' atomizers.

Much of the initial work employed graphite-based devices that were resistively heated, although many were in the form of rods rather than the original tubular design of L'vov. This was probably a result of the ease of in-house manufacturing of making low resistance contacts to carry the currents needed to heat the rods. Much activity persisted in the 1970s in trying different atomizer designs that ranged from rods, to braids to tubular designs. The tubular design ultimately dominated because of its ability to provide a higher temperature, semi-enclosed environment. The elevated temperature minimized compound formation and condensation while the semi-enclosed nature of the tubular design enhanced the residence time of the analyte within the optical path to improve sensitivity.

The atomizer material was not limited to graphite, and a number of high melting metals (e.g. W, Ta, Pt, etc.) were evaluated. The metals often provided the advantage of higher resistance and lower mass, thereby permitting lower powers and the use of higher voltages (with lower currents) to bring about rapid heating. For analytes that tended to form refractory carbides on graphite, these metal atomizers permitted complete vaporization at lower temperatures. However, there were also a large number of metals that dissolved in the metal atomizer or formed solid solutions that produced memory effects as they were slowly and continuously distilled from the atomizer with each subsequent firing. Graphite probably remained the material of choice for a number of reasons, although the most readily apparent were its advantage of being available at high purity, having a very high (ca. 3400 K) sublimation temperature, and oxidizing to form a surface ''oxide'' ($C-O_{(ads)}$) that was readily removed as $CO_{(g)}$ and $CO_{2(g)}$ from the surface with simple heating. The size of the tubular atomizers ranged from very small devices that heated very rapidly and could only accommodate a few microliters of sample, to the ''Woodriff furnace''[4,5] which approached 25 cm in length and was preheated before the sample was introduced.

Significant among the developments was Massmann's[6] early design of a directly heated tube into which the sample was introduced. This design precluded the striking of an auxiliary arc as was required with the L'vov design. Most modern furnaces are smaller versions of this design with tube lengths typically ca. 2.5–3 cm with a 4–7 mm inner diameter (Figure 3). As shown in Figure 3, the furnace is heated from the ends (i.e. ''end-heated'') by direct passage of current through the furnace.

With the tube furnace, it became readily apparent that analyte elements were often ''stuck'' near the ends of the tube since the tube did not heat uniformly. Greater heat loss was experienced where the ends of the furnace contacted the water-cooled endcones and could result in a temperature gradient as large as 1000 K between tube center and end.[7] The heating of the furnace was more uniform during a rapid deposition of energy into it (e.g. during a rapid thermal ramp) and the ends tended to cool as heating rates decreased or once a steady state temperature in the tube center was achieved. This often resulted in material vaporizing from the initial deposition site and recondensing on the now cooler ends of the furnace. As a result, the next heating of the atomizer produced a secondary release of these condensed metals, giving rise to memory effects. This was not uncommon, especially for the more refractory analyte elements. To minimize this problem, it was common to use higher

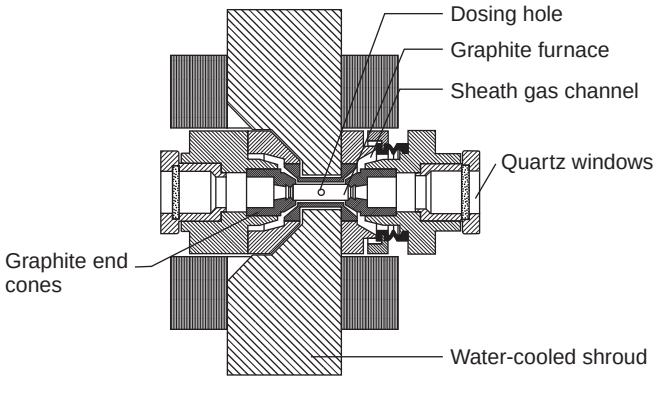

Figure 3 This is a typical, modern end-heated graphite furnace and workhead. The graphite furnace atomizer is in the center of the workhead that includes the electrical, gas and cooling connections. The furnace itself is ca. 2.5 cm in length with a 0.5 cm inner diameter. (Reproduced by permission of Varian, Inc.)

atomization temperatures and ''clean cycles'' (i.e. high-temperature heating of the furnace after the analytical measurement was complete). Another alternative was a redesign of the tube so that the electrical contacts did not also serve as a heat sink that produced this longitudinal temperature gradient. Isothermal cuvettes (or ''transversely heated graphite atomizers'') were first introduced by Frech et al.[8] to minimize thermal gradients along the tube. Because of the more isothermal environment and encouraging preliminary results, this concept has made its way into the commercial market and is another option to the end-heated furnace concept.

Another major breakthrough in use of furnaces was initially suggested by L'vov et al.[9] and involves the use of a platform inserted into the furnace (Figure 4). Since the formation of free atoms is favored at elevated temperatures, it becomes problematic if a volatile analyte (or analyte-containing compound) vaporizes at a low temperature, which permits molecular species to form (or remain). A small graphite flat in the furnace onto which the sample is placed has minimal thermal contact with the heated tube walls. As a result of poor conductive heating of the platform by the wall, the temperature of the platform lags that of the furnace walls. Depending on the heating rate of the tube wall and the heat transfer rate between wall and platform, this lag can be as large as several hundred degrees. As a consequence, vaporization of the sample from the platform into the furnace takes place when the gas phase is at a higher temperature. This is illustrated in Figure 4. Generally, the higher temperature will sponsor molecular dissociation (atomization) and both improve atomization efficiency as well as minimize many interference effects. Because most platforms never reach the wall temperature, the simple design depicted in the figure is often not useable for the more refractory elements such as V, Mo or even Ni and Cr. Alternative designs of the platform where better thermal contact with the wall exists are starting to appear on the market and even determination of refractory metals with this new design is feasible.

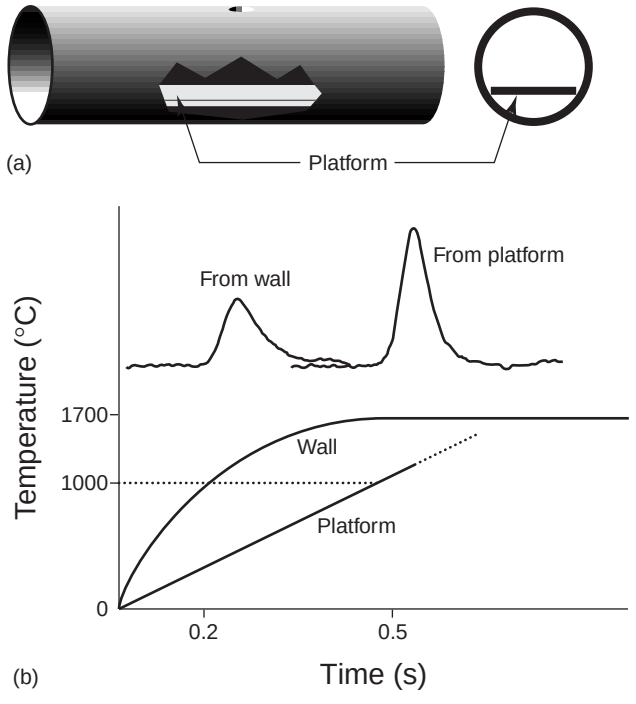

Figure 4 (a) Schematic diagram of a platform located inside a graphite furnace atomizer. The sample is initially deposited on the platform. (b) Representation of the temperature of the furnace wall and platform temperature during the pulse heating of the atomizer, that is ''atomization cycle''. (Reproduced by permission from K.W. Jackson (ed.), *Electrothermal Atomization for Analytical Atomic Spectrometry*, Wiley, New York, 13, 1999.) The horizontal dashed line at 1000 °C represents the appearance temperature. It should be noted that vaporization from the platform presents the analyte to the gas phase when the wall temperature is several hundred degrees hotter, thus promoting free atom formation.

In the early days of flameless AA, carbon rod atomizers, etc. the technique was plagued by extreme interference effects. Some of the modifications noted above (i.e. tube-type construction, isothermal heating and platform use) ameliorated many of these problems. In spite of these unique challenges, it was readily recognized that vapor formation using an ETA was not limited by factors that often plagued other spectroscopic techniques where nebulizers were used for sample introduction. For example, samples that were viscous, contained high concentrations of total dissolved solids, or contained suspended solids were often unusable with conventional nebulization but could be dealt with using ETA. As a result, expectations of high accuracy and precision were appearing for samples containing extremely complex matrices. These complex matrices placed heavy burdens on the ability of the ETA to ''synthesize'' atoms within the furnace consistently and with nearly 100% efficiency.

The use of chemicals added to the sample as ''matrix modifiers'' became common place as a means of minimizing matrix effects by permitting, for example, the removal of the offensive matrix during a thermal pretreatment step prior to the high temperature atomization heating cycle. Ediger[10] is often recognized as one of the first to introduce such a concept with the suggestion that Ni could be added to a sample that was being analyzed for Se. His suggestion was that the Ni bound the Se and ''stabilized'' it on the graphite surface to a higher temperature while offending matrix concomitants could be vaporized during the ash or thermal pretreatment cycle. The delayed vaporization of the selenium until the ETA was at a higher temperature also improved the atomization efficiency. The search for and employment of matrix modifiers continues to dominate the literature

as an important means of achieving reproducible and accurate analytical results in ETAAS. Many of the features alluded to above will be discussed in more detail in subsequent sections.

It is important to stress at this point that ETAAS is a powerful analytical technique that is widely used for ultratrace and microtrace analysis. It has the advantages of a high degree of elemental selectivity, and once a method is developed, analysis is relatively simple. Modern instrumentation has empowered the user with a high degree of automation, which further simplifies the use of ETAAS for a large number of samples. Inherent to AA, however, is its traditional single elemental analytical capabilities, that is, one element at a time can be determined. However, recent work in the area of using multiple sources and continuum sources may arm the approach with simultaneous multielement capabilities in future instruments.

The first sections discuss the theory behind the technique in an attempt to explain what is taking place within the atomizer at the molecular level. In this chapter there is also a brief discussion of ''interferences'' that can occur and why they are occurring, again to provide a complete picture of what is taking place fundamentally. However, the practical analytical utility of the approach should not be covered up by these discussions since, it will be shown, most of these potential problems can be circumvented with proper choice of instrument operating conditions and, in many cases, correct selection of matrix modifiers.

3 THEORY

3.1 Absorption Fundamentals

The basics of AA are also covered in **Atomic Spectroscopy: Introduction**. In brief, the absorption of radiation by the analyte within the furnace volume for the purpose of conducting analysis assumes the existence of free, gas-phase atoms. Analyte atoms present as molecular species (e.g. PbO, MnCl, etc.) or as ions (e.g. Na^+) will not absorb at wavelengths used to detect the neutral atomic species. The preferred analytical lines for conducting AA in the ETA are identical to those used with flame atomic absorption (FAA) and are generally absorption transitions which originate in the unexcited atomic ground state and terminate in a higher lying electronic state, that is, resonance lines. For many atoms, there exists more than a single excited state that can be populated by an absorptive transition from the ground state. In these instances, the strength of the absorption is determined by the oscillator strength (or Einstein B coefficient for stimulated absorption). A larger value implies a higher probability of photon absorption and, hence, greater sensitivity. In some instances, the wavelength of the source wavelength may be noisier at this more strongly absorbing line and, as a consequence, the weaker line is used since it may provide an improved signal-to-noise ratio. For example, the 217-nm line for Pb is ca. 2.4 times more sensitive than the 283-nm resonance line, but the shorter wavelength is often noisier and more prone to interferences by scatter and molecular absorption. Hence, it is not uncommon to see the 283-nm line used in preference.

Lines which originate above the ground state generally show weaker signals since the strength of the absorption is directly dependent on the population of the low lying state. The magnitude of the absorption is governed by both the number of atoms within the optical path and the population of the absorbing state for those atoms. Generally 99+% of the atoms will exist in the lowest excited electronic state (i.e. ground state) even at temperatures of 3000 K. Any of the higher lying states will be less populated and vary with temperature as dictated by the Boltzmann distribution. Hence, if absorbance measurements are being taken as the temperature within the furnace is changing, the atom density is not strictly proportional to the absorbance since the population of the originating state is changing.

It is possible that the absorbance measurement can be too large, that is it lies outside the dynamic range. When a given analyte concentration produces too large an absorbance signal, the sample can be diluted or an alternate absorbing line can be used. An alternate line that originates in the ground state but terminates in an upper state with a smaller oscillator strength is preferred since it will show minimal temperature dependence in comparison to an absorption that originates in a state above the ground state.

In general, the absorbance signal is proportional to the gaseous analyte mass within the furnace volume at any given instant in time. If the rate of analyte release into the gas phase is constant from standard-to-sample and from sample-to-sample, then the height of the absorbance signal (''peak height absorbance'') could be used to construct a calibration curve. However, the peak shapes may change because of changes in vaporization rates caused by altered ETA heating rates, changes in vaporization mechanisms, etc. Therefore, the peak area is generally used in constructing the calibration curve and results in improved analytical accuracy. The peak area is calculated by integrating the absorbance signal over the lifetime of the peak and is expressed as peak area and has units of ''per second'' (since absorbance is dimensionless).

As with all atomic absorbance techniques, the linearity of the calibration curve is limited and covers only two to three orders of magnitude in concentration. The cause of the nonlinearity and apparent violation of Beer's

law, can be a consequence of stray light,[11,12] the finite width of the source line relative to the absorbing line[13] or other more subtle spectral line shape alterations[14] and spatial distributions of source radiation and analyte vapor.[15] Lasers (e.g. diode lasers[16,17]) can be more intense with less noise than conventional HCLs or electrodeless discharge lamps (EDLs). As a result, the lower concentration end of calibration curves may be extended, which results in improved detection limits and an enhancement in the dynamic range of ETAAS.

Like any AA technique, the likelihood of atomic spectral line interferences is low. This is a result of the narrow line widths of both the source and the absorber. Additionally, unlike emission spectroscopy, the number of atomic spectral lines that are capable of producing a significant absorbance signal is quite small. This is primarily a result of the need for a significant population of the originating state.

For any given radiation source (e.g. HCL), ETA is considered 100 to 1000 times more sensitive than FAA when comparing characteristic masses, that is the mass deposited in the furnace that is needed to produce a peak area of 0.0044 s. While some gain in sensitivity is garnered by the narrower absorbing line profile for atoms present in the furnace, most of the improvement results from the longer residence times of the atoms within the radiation beam with ETAAS. In a typical flame with a linear flame velocity of ca. 100 cm s^{-1}, the analyte atoms reside in the optical path for less than 10 ms. Atom residence times within the furnace are several tenths of a second to several seconds!

3.2 Production of the Atomic Absorption Signal

The typical sequence of events in the practical analysis of a sample involves the introduction of a sample (typically 10–25 μL of a solution) into the furnace through the dosing hole. A gentle dry cycle is initiated by heating the furnace to ca. 80–150 °C for 30–40 s to desolvate the sample. The proper choice of time and temperature is generally selected by viewing the sample using an inspection mirror (e.g. a small ''dentist's mirror'') through the furnace while drying to ensure that the droplet disappears completely and slowly, that is, no ''splattering'' from heating too rapidly. One or more thermal pretreatment cycles (sometimes referred to as an ''ash'', ''char'', or ''pyrolysis'') are used to decompose some of the matrix and possibly to even remove the matrix if it is more volatile than the analyte. The temperature and times for this step vary considerably depending on the analyte, sample and matrix. For example, a simple sample containing a volatile element such as Cd, Zn or Pb may be heated to no more than 300–400 °C while a more refractory element such as Ni might be heated to 1200 °C or higher without fear of analyte loss. Finally, the atomization cycle is initiated where the furnace is rapidly heated (e.g. 1000–2000 °C s^{-1}) to a final temperature that is sufficient to remove all of the analyte from the furnace (e.g. 2000–2700 °C). It is during this cycle that the absorbance is recorded. Following the ramp, the furnace is often held for a few seconds at this elevated temperature. A clean cycle (2800–2900 °C) is sometimes initiated before cooling down to ensure complete vaporization and removal of any residuals from the sample.

In the following sections are discussions of the processes occurring during these heating cycles that ultimately lead to the generation of the transient absorbance signal from which quantitative analysis will be conducted. The discussion will begin with a look at the last cycle first (namely, the atomization cycle) and discuss the key steps leading to the production of gaseous atoms within the furnace ETA.

3.3 Basic Vaporization (Desorption) Process

The basic analytical signal is shown in Figure 5. As one might expect it is transient in character because the vapor generated by vaporization is not ''trapped'' in the furnace but is lost through the dosing hole and the ends of the furnace. There are some characteristics of the peak that are often alluded to and these are labeled in the diagram. The appearance time and temperature of the signal are t_{app} and T_{app} respectively. This is typically the time or temperature when the signal level has reached a magnitude that is three times the baseline noise level. The values are not thermodynamic in their origin, but do give an indication of when a signal will first appear and provide guidance regarding the atomization temperature that should be used. The time designations of τ_1 and τ_2

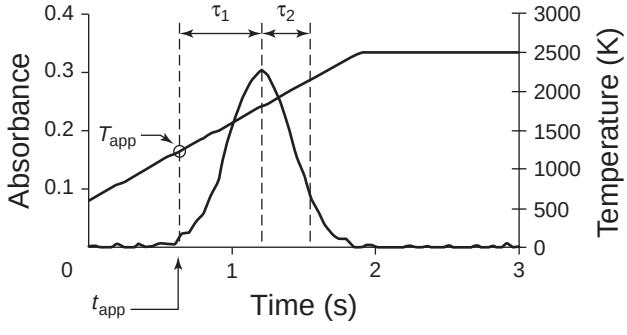

Figure 5 Typical analytical absorbance signal. The width of the peak can vary depending on the analyte, heating rate and furnace geometry. The appearance temperature is determined as the temperature or time where the analytical signal reaches a value three times the standard deviation of the baseline noise level. The time periods designated as τ_1 and τ_2 are times required to reach the peak absorbance signal and for the signal to decay to 1/e of its peak maximum, respectively.

provide some indication of the shape of the signal and have their origins in the original papers of L'vov who made the first attempts to explain the atom production process. τ_1 is the ''atomization time'' and is measured from t_{app} to the peak time. τ_2 is the time required for the peak absorbance to decay to 1/e of its value. In the initial treatment, L'vov suggested that minimizing τ_1 and maximizing τ_2 would place the maximum number of analyte atoms in the furnace volume and in the analytical volume. He proposed that rapid heating combined with a longer furnace would favor the total containment of the sample vapor in the furnace during the measurement period. Since this earliest work, we find now that the processes are somewhat more complex in many instances.

The mathematics describing the production of the analytical signal is a convolution of a time-dependent supply function, $S(t)$, and a removal (or dissipation) function, $R(t)$, as shown in Equation (1):

$$N(t) = \int_0^\infty S(t) - R(t)\,\partial t \tag{1}$$

where $N(t)$ is the number of atoms in the analytical volume, that is, the furnace. $R(t)$ must incorporate loss by: diffusion out of the ends of the furnace and the dosing hole; reactions in the gas phase that form analyte-containing molecules; readsorption, condensation, or sticking of the analyte to the furnace wall; nucleation in the gas phase to form condensed phase particles; and any other process which depletes the free atom density within the volume where the absorption measurement is made. $S(t)$ includes the time-dependent production of gaseous analyte atoms by: vaporization (or desorption) from the surface; dissociation of analyte-containing molecules in the gas phase or on the surface (followed by metal vaporization); and any other source responsible for adding more atomic analyte to the gas phase.

3.3.1 Generation

Since the furnace is generally being heated while the transient absorption signal is produced, generation and loss functions must also account for the time-dependent temperature change, that is, nonisothermality of the furnace in time. Additionally, the furnace and the gas phase within the furnace may not be spatially isothermal, which must also be considered for an exact description of the shape of the signal. Thus, it is easy to see that even if the exact processes that contributed to $N(t)$ were known, the solution to Equation (1) could be quite complex. Using overly simplified assumptions, however, does produce an $N(t)$ versus t plot that is a reasonable approximation to what is observed. For example, one can assume that the number of gaseous atoms is proportional to the partial pressure of the analyte and assume that this pressure is dictated by equilibrium vapor pressures of the heated metal solid or liquid. Additionally, one can consider loss as a simple diffusive transport of material from the center of the furnace out the ends and ignore the dosing hole and any other loss pathway.

Since the number of analyte atoms on the surface is quite small, in many instances the analyte may not take on ''bulk'' thermodynamic characteristics. For example, if 20 pg of copper were evenly dispersed over a 2-mm-diameter area where the sample was initially dried, only about 5% of the surface would be covered. Thus, the energy required to produce copper vapor may not depend on the heat of vaporization (ΔH_v) of $Cu_{(l)}$ but instead the energy needed to desorb adsorbed copper from graphite. In some instances where the interatomic forces are strong, it is possible that the metals on the surface diffuse and form small droplets of the ''pure metal'' in which case ΔH_v is applicable to describe vapor production.

The mechanism of vapor production in an ETA has been an intense area of research interest for a number of years and is still not totally understood. One of the more interesting diagnostics that is often used attempts to extract the energy associated with the vaporization process and, by inference, to deduce the pathway by comparing the experimental value to literature values. The commonly cited approach was proposed by Sturgeon et al.[18] and the resultant graph is often referred to as an ''Arrhenius plot'', because its derivation is similar to that used in deducing reaction kinetics. In brief, it may be given by Equation (2):

$$\ln(A) = \frac{-E_a}{kT} + C \tag{2}$$

where A is the absorbance signal, E_a is the activation energy of release (often taken as ΔH), k is the Boltzmann's constant and T is the furnace temperature in Kelvin at the time the absorbance measurement is made. C is often assumed to be a constant and incorporates a number of terms including spectroscopic constants that relate an absorbance measurement to atom density (or pressure) as well as thermodynamic (or kinetic) values that are assumed to be temperature independent. In particular entropy change (or the pre-exponential factor when viewing the process as a kinetic process) is involved in this term.

Since the derivation of Equation (2) assumes that the analyte surface coverage (or condensed state activity) is not changing, the equation is most applicable (and most linear) at the very beginning of the absorbance signal. Smets[19] added a modification to this basic equation that attempted to account for changing surface coverage and extends the linear portion of this plot later in time. The temperature is usually measured using optical pyrometry

by viewing the inside of the furnace wall through the dosing hole. This approach permits coverage starting at ca. 1100 K. Once a value for E_a (or ΔH) is obtained, interpretation to extract the vaporization step can often be tenuous.

It is also possible that gaseous atoms originate directly from molecular species. This includes thermal decomposition of metal-containing compounds in the gas phase or on the surface. In some instances, the surface reaction can be thought of as a reduction or ''thermoreduction'' process in which reactive carbon of the furnace participates in the free atom formation as the temperature increases, for example, reaction of a metal oxide with C to form free gaseous metal atoms and CO or CO_2.

3.3.2 *Dissipation*

Analyte loss from the analytical volume of the ETA can take a number of forms, but in all cases the loss must be irreversible. For example, metal gas colliding with the wall of the ETA and sticking but revaporizing at a later time during the measurement cycle would not be considered ''irreversible loss''. The simplest loss or dissipation is diffusion of analyte from the ends of the furnace or out of the dosing hole.

A simple equation, Equation (3), that considers the continuous diffusive loss from the furnace ends was proposed by L'vov

$$\tau_2 = \frac{L^2}{8D} \tag{3}$$

where τ_2 is the time required for the absorbance signal to decrease to 1/e of its peak value, L is the furnace length and D is the temperature-dependent diffusion coefficient. This provides a reasonable approximation to the loss rate by diffusion, but does not include loss from the dosing hole. It also ignores the finite time at the onset of vaporization for material to diffuse from the center to the ends of the furnace to set-up a pseudo-steady-state concentration. It also does not take into account the more subtle loss mechanisms of recombination of the analyte metal to form nonabsorbing molecular species.

3.3.3 *Putting it Together*

Use of the more complex geometry of the furnace and the presence of a small, finite amount of sample makes the solution to Equation (1) more complicated. One approach to modeling the more realistic system has employed the Monte Carlo technique.[20] This is a stochastic approach that permits complex geometries and boundary conditions to be used in the construction of a model. Figure 6 shows the results using Monte Carlo simulation techniques to illustrate the time-dependent

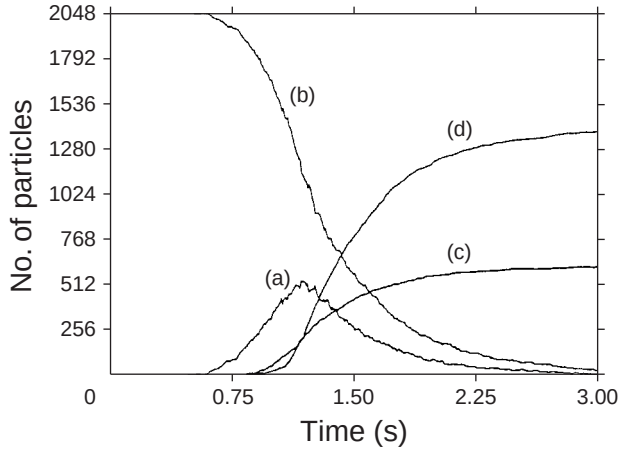

Figure 6 Monte Carlo simulation showing the time-dependent location of analyte atoms during a thermal heating cycle: (a) in the furnace gas phase; (b) on the furnace wall; (c) lost out of the dosing hole; and (d) lost out of the ends of the furnace. Data are for a small (3 mm diameter × 1 cm long) furnace. (Reproduced from Güell and Holcombe[20] by permission of the American Chemical Society.)

location and loss of analyte during a furnace heating cycle. This simulation is for the production of copper vapor. As can be seen, the leading edge of the gas phase signal rises exponentially as the temperature ramp increases linearly. As the amount of copper on the furnace wall decreases, the generation rate falls off. The peak of the absorbance signal represents the point where generation and loss rates are equal, that is, $S(t) = R(t)$. It should be noted that a significant (ca. 20% in this case) amount of the analyte is lost from the dosing hole and that analyte loss occurs from the dosing hole before it is lost from the furnace ends. The decrease in the atoms found on the tube wall also does not conform to a simple exponential evolution as might be surmised from a linear temperature ramp and the assumption of ''bulk amounts'' of metal on the surface (i.e. activity of metal is unity). In the case of copper, this is a result of the release rate being dependent on the amount of copper (i.e. surface coverage) on the graphite, which changes with time (i.e. its activity decreases as material is vaporized). As the surface coverage drops, the rate of evolution decreases below that expected when surface coverage is not taken into account.

3.4 Preatomization Reactions

Many reactions occur on the surface prior to atom production. These can be as simple as prevolatilization of metal compounds and migration of material below the surface. Some processes only impact the appearance temperature and the shape of the analytical signal while others can affect the number of free atoms produced. The latter, of course, directly impacts analytical accuracy

while the former may have little impact on accuracy when using the area under the absorbance signal. Most of these problems can be circumvented through the proper selection of temperature programming, a platform and chemical modifier.

Analyte prevolatilization involves the low temperature release of analyte-containing molecules at a temperature that is lower than that needed to thermally decompose these compounds. As a result, the absorbance signal is depressed. This is not uncommon for metal halides which are often relatively volatile but form stable metal halides in the gas phase. Some elements (e.g. selenium) also have volatile oxides that can also be lost from the furnace.

3.4.1 Decomposition

The form of the analyte salt residue that exists on the graphite surface after desolvation (''dry cycle'') can often be predicted by the dominant matrix anions and relative solubilities of the possible salts that could form. For example, the metal nitrate salt is likely in a nitric acid matrix, metal halides in a HCl matrix, etc. Depending on the solution pH, metal hydroxides are not unexpected. During the dry or char cycle, many of these salts decompose to the metal or metal oxides as shown in Equations (4) and (5):

$$M(NO_3)_{2(s)} \longrightarrow MO_{(s)} + 2NO_2 + \tfrac{1}{2}O_2 \qquad (4)$$

$$M(OH)_{2(s)} \longrightarrow MO_{(s)} + H_2O \qquad (5)$$

Easily reducible metals (e.g. Au, Ag, Cu, etc.) are likely present on the surface in the elemental form before atomization. There is little doubt that the graphite also serves as a reducing agent to assist this process for a number of metals as the furnace is heated.

Graphite also serves as a potentially problematic reactant as well as a reductant for elemental formation. Many metals bind very strongly to carbon to form stable surface species or even stoichiometric metal carbide compounds. These species are often very refractory and can be difficult to vaporize completely unless very high atomization temperatures are employed. Examples of these metals include W, Mo, Ta and Si. These stable carbides often precluded the use of platform technology (see section 4.4) since the platform may be incapable of reaching the wall temperature even with sustained heating times. Several workers have successfully used metal liners (e.g. Ta, W, etc.) placed inside the graphite furnace to circumvent this problem. However, metal liners cannot be permanently used because they limit the final atomization temperature to values substantially lower than that for graphite and can react with some analyte metals to form relatively stable solid solutions (e.g. alloys).

While the pyrolytic coating used on most commercial furnaces greatly reduces permeation of the sample liquid and gaseous vapor through the furnace walls, it is not single crystalline in character and has a considerable number of surface imperfections, dislocations and ''cracks'' in the surface. These features permit samples to migrate below the surface, perhaps as early as during the drying cycle and often as far as several micrometers.[21,22] Majidi et al. have also shown that all metals do not behave the same. In many instances it is thought that the analyte forms intercalation compounds that are located in the interlamellar space between the graphitic sheets.

When metals that form intercalation compounds are present in large amounts such as quantities that might be present in a matrix or used in a chemical modifier, extensive damage to the furnace can occur. This is a result of these compounds causing an expansion in the interplanar spacing of the graphite which ultimately causes delamination and exfoliation of the material. The extreme result can include bending/warping of platforms and destruction of the furnace.

It should be noted in closing that it is not uncommon for materials to migrate on surfaces. Depending on the forces of attraction present, it is possible that adsorbed atoms, analyte-containing molecules and even small particles can move on the surface by surface diffusion. This diffusion can bring about the formation of small microdroplets of metals even if the metals were initially quite dispersed on the surface. Similarly, metals present as small droplets or islands initially may disperse on the surface as the temperature increases and entropy provides a driving force.

In some instances the appearance of migration occurs as material vaporizes (or desorbs) and is readsorbed at some distance away from the original site. This process occurs quite readily and can precede the appearance of the absorbance signal by several hundred degrees. The net result is a picture of the release of atoms into the gas phase that is not as simple as ''vaporize, enter the gas phase and diffuse from the ETA''. With a mean free path of fractions of millimeters, there is ample opportunity for multiple collisions with the surface and readsorption at a measurable distance from the original sample deposition site.

3.4.2 Graphite–Oxygen Reaction

The importance of oxygen in forming metal oxides on the surface and in the gas phase has made studies of this ''interferent'' of special importance. Sources of oxygen include decomposition of oxyanion salts, the sample solvent (water), impurities in the sheath gas and diffusion of atmospheric oxygen into the furnace through the dosing hole. Reduction of oxygen by the graphite furnace is thermodynamically favorable at all temperatures and should produce CO and CO_2, with

the $CO:CO_2$ ratio increasing at elevated temperatures if equilibrium is achieved. However, Sturgeon et al.[23] were the first to suggest that this equilibrium was not likely achieved in the ETA until temperatures exceeded ca. 1500 K. This was later confirmed and presented in more detailed models.[24] Sturgeon and Falk[25] were the first to present experimental evidence of the partial pressures of oxygen in the furnace which was not based on inferences from the metal AA signal. Interestingly, the values agreed well with values chosen earlier by Cedergren et al.[26] in their thermodynamic modeling of reactions within the furnace. In their studies they selected values that produced reasonable appearance temperatures for the metals under study.

As seen in the work of Gilmutdinov et al.[24] the reactions of oxygen with the graphite present a complex time and spatial picture within the furnace. However, when combined with the authors' spatial viewing of free atom densities within the ETV,[27] it explains many of the unusual distributions that were observed. For example, the disappearance of many AA signals near the top of the furnace could be attributed to the infusion of air into the furnace through the dosing hole. Similarly, with the platform in place, some metals showed a particularly strong absorbance signal under the platform, which is consistent with a rarified oxygen atmosphere that would favor metal oxide dissociation.

In some analytical situations, oxygen is intentionally added to the sheath gas during the dry and thermal pretreatment stage. This is referred to as ''oxygen ashing'' and is most commonly done when there is a significant amount of a combustible material in the sample, e.g. a biological or low volatility organic matrix. Pyrolysis (i.e. heating to high temperatures in the absence of oxygen) of a biological matrix, for example, can leave copious amounts of carbon in the furnace even after the atomization cycle. Using oxygen admixed with the Ar, this matrix undergoes combustion and consumes the carbon as CO and CO_2. It is obviously critical that the oxygen be removed from the furnace if it is heated above ca. 800 °C since oxidation (and destruction!) of the graphite atomizer will occur.

3.4.3 Gas-phase Reactions

The presence of gas-phase reactions with the analyte will effectively reduce the free atom density and produce a signal depression. The depressed signal is of analytical importance if the source of the reacting species originates in the sample and is not present in the standard solutions. In many instances, standard addition is still capable of producing accurate results under these circumstances. The gas-phase reactions that are significant are, of course, metal-dependent, but as a general guide one must be cautious of halides and oxides, both of which can form stable gas-phase diatomic species with many metals. In most instances, dissociation of these diatomic species is favored at elevated temperatures. As a consequence platform technology is often recommended. Another alternative is to ''thermally stabilize'' the analyte with the use of a matrix modifier. This accomplishes two objectives. By delaying the analyte vaporization until a higher temperature, dissociation of analyte molecules on the surface or in the gas phase is enhanced. Additionally, if the analyte is not released until a higher temperature, then thermal pretreatment can be conducted at a higher temperature without fear of analyte loss. In many cases, this may permit the removal of the problematic matrix during this thermal stage prior to the atomization heating cycle.

The other option to reduce gas-phase interferences is to minimize the concentration of interferents in the sample through the use of matrix modifiers. As an example, the simple addition of HNO_3 to a chloride-containing sample results in a significant loss of the chloride as HCl during the dry cycle. Ammonium salts (e.g. $NH_4H_2PO_4$) are also used since the resulting $NH_4Cl_{(s)}$ is relatively volatile and provides a pathway to remove the chloride at relatively low thermal pretreatment temperatures.

4 INSTRUMENTATION

Modern ETAAS instrumentation includes a light source, background correction, the atomizer, the spectrometer and signal processing. Accessories would include autosampling systems and other attachments that maximize the performance and flexibility of ETAAS in the analytical laboratory. While other articles also discuss some aspects of the instrumentation (**Atomic Spectroscopy: Introduction; Background Correction Methods in Atomic Absorption Spectroscopy; Flow Injection Analysis Techniques in Atomic Spectroscopy**), this section will highlight aspects specific to ETAAS and provide some indication of current instrumentation as well as approaches that are on the horizon.

4.1 Light Sources

The traditional light source is the HCL which is a glow discharge source whose cathode is fabricated from the analyte metal(s) of interest and is contained within a sealed envelope surrounded by a few torr of an inert gas. The choice of fill gas is element-dependent. The glow (or Grimm) discharge is a low current (i.e. few milliamps) source that relies on sputtering and electron impact excitation for sampling and excitation, respectively. The source is not under local thermodynamic equilibrium and exhibits a very low kinetic temperature but a large electron temperature. As a consequence, a large number

of elemental lines are emitted but they have a very narrow half width, an ideal situation as a source for AA. Generally, a lamp provides the spectra for a single element. While multielement lamps are available, they usually provide reduced intensity for the elements within the lamp due to compromised excitation conditions and reduced sputtering efficiency because of fractional composition of the cathode material.

When low intensity from the HCL at the preferred wavelength presents a relatively high noise level, EDLs often become the source of choice. These devices are radiofrequency or microwave excited discharges which generally emit a more intense spectra. This is particularly useful for elements whose resonance lines are at shorter wavelengths such as As and Se.

More recently, tunable diode lasers have emerged as replacements for HCLs or EDLs. Because of their improved stability, narrow line width and reduced noise, diode laser sources have been used to record absorbance signals as low as 1×10^{-6} absorbance units. As a consequence, larger dynamic ranges and significantly reduced detection limits appear possible. A commercial source of these lamps is currently available. Widespread use may be contingent on the development of diode lasers with fundamental wavelengths deeper into the blue region of the spectra so that simple frequency doubling is only needed to reach the short wavelengths (ca. 180–200 nm) where the resonance lines of many elements of interest lie.

A continuum source with a high resolution spectrometer has also been explored. Uniquely, it offers the possibility of simultaneous multielement analysis for ETAAS. However, the weak intensity of the source at low wavelengths has traditionally made this source less sensitive than traditional HCLs or EDLs, especially at shorter wavelengths. Consequently, its use has been primarily confined to the research laboratory. Recent work with diode array detectors suggests that comparable sensitivities are available even at the shorter wavelengths. This approach may become another viable alternative to traditional line sources.

4.2 Background Correction

This topic has been discussed in greater detail in the article **Background Correction Methods in Atomic Absorption Spectroscopy** in this publication. In general, background correction is much more critical in ETAAS than in FAA because of the higher local density of sample vaporized and the presence of cooler areas near the dosing hole and furnace ends where molecular species may form and/or condensation may take place. Additionally, ETAAS is often employed with samples that have a very high concentration of matrix components, with the extreme case of using the method for direct solids analysis.

4.3 Atomizer

The atomizer has been discussed in general detail earlier in this chapter. In most cases the furnaces are fabricated from a high density ''amorphous graphite'', often referred to as ''electrographite''. This material is made from graphitized carbon at elevated temperatures but the crystals in the material are small and their orientation is random. While the density of the material is high (i.e. low porosity), it is not as great as pure graphite. This material is then coated with a 50 to 100-μm thick coating of ''pyrolytic graphite'' which is laid down by chemical vapor deposition techniques at elevated temperatures using a hydrocarbon gas as the source. (The temperatures and gas composition are usually proprietary information.) The coating makes the surface less permeable to diffusion by sample liquids and gaseous vapors. The coating is also more crystalline in nature with general orientation of the graphitic planes parallel to the surface. The reactivity of the coating is generally significantly less than the electrographite substrate.

Most furnaces are of a Massmann-type design with power supplied to the furnace via contacts at the tube ends. As noted earlier, Frech et al.[8] proposed an ''isothermal cuvette'' to minimize a temperature gradient that is known to persist from center-to-end in the more traditional furnace. A modification of the concept is commercially available as a ''transversely heated graphite atomizer''. Figure 7 shows an example of these two furnace types. As noted in earlier sections, platforms or probes are a routine part of modern analysis by ETAAS. Depending on the design and intimacy of the contact between the platform and the tube wall, refractory metals may or may not be vaporized completely from the platform. In those designs where the platform is incapable of reaching the needed temperature, wall vaporization is used.

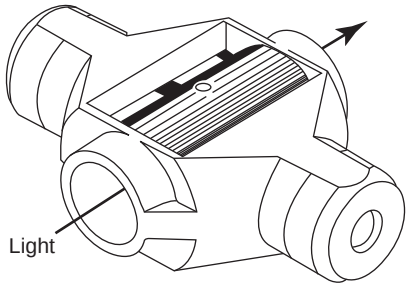

Figure 7 Transversely heated graphite atomizer where the current to heat the furnace is not passed axially through the furnace from the ends. The design is intended to eliminate thermal gradients along the length of the tube to minimize cool spots in the tube where condensation refractory metals and/or molecular formation might occur. This particular design also has a platform machined into the furnace. (Reproduced by permission of PerkinElmer Instruments.)

More recently, the addition to the ends of the furnaces of circular caps with holes in the centers has been suggested.[28] These caps would reduce the rate of diffusive loss of analyte from the furnace and provide a more isothermal environment that may enhance sensitivity and reduce interferences. An even more dramatic change in furnace design has been proposed by Katskov et al.[29] who use a ''filter furnace''. In brief, the sample is deposited in a small chamber between the conventional filter wall and an insert within the tube that is fabricated from porous graphite. Upon heating, the sample diffuses through the hot porous graphite into the analytical volume. Results suggest fewer interferences while maintaining good sensitivity for the limited number of elements that they have considered.

4.4 Platforms and Probes

As was discussed at the start of this chapter and shown in Figure 4, the use of inserts (e.g. platforms or probes) is a common part of modern use of ETAAS. Platforms that have a more intimate thermal contact with the furnace wall will exhibit less of a thermal lag and may not prove as effective in dissociation of analyte-containing gas-phase molecules. However, this type of platform can be used for most analytes, including the more refractory elements.

4.5 Spectrometer and Optics

In general, the spectrometer is a medium or low resolution monochromator since it is only needed to isolate the resonance line of interest from other lines of the same element emitted from the line source, typically an HCL. Spectral bandpasses in the range of 0.1–5 nm are generally all that are required. The exception to this occurs when a continuum source is being considered in place of the HCL or other line sources. In this instance, a high throughput, high resolution spectrometer is preferred. To this end, an echelle spectrometer used in very high orders is the spectrometer of choice.

There are instruments which use several line sources which are alternately passed through the ETA and detection is made at several wavelengths. In this fashion the ETAAS system becomes capable of multielement analysis for a limited (four to eight) number of elements with a single firing of the atomizer.

Most systems are operated in a double beam mode where a chopper is used to alternately pass the source radiation through and around the ETA. In the case of the pulsed background correction, chopping is not needed since the same source provides on-line and off-line measurements (see **Background Correction Methods in Atomic Absorption Spectroscopy**). The path around the ETA is used primarily to correct for any drift in the source intensity. Additionally, the source may be modulated between the line source and the source used to register a background correction signal. For example, if an HCL is used with a D_2 continuum source, the signals measured are typically: (i) the line source through the ETA measuring line and background absorbance; (ii) the continuum source through the ETA measuring background equivalent absorbance (and scatter); (iii) the line source around the ETA; and (iv) the continuum around the ETA to use as a measure of source stability and to register a zero absorbance level. Additionally, there is a fraction of the measurement cycle when no source radiation passes. During this period any emission from the furnace (blackbody, scatter, line or molecular) is registered and corrected for.

A photomultiplier tube is employed as the detector in most cases because of its low noise and high sensitivity. Solid-state detectors and array-type detectors may find greater use in the future but are minimally used in current AA systems.

4.6 Signal Processing

Because of the transient nature of the ETAA signal, faster detection and processing is required than might be needed when the flame is used for atomization. Millisecond response times are often needed from the system to accurately reflect the transient signal. Collection rates of 20–120 Hz are typical for most modern instruments, yielding more than 100 measurement points to define a typical ETA signal of ca. 5 s duration. While faster modulation and collection could be used, the added noise from the more rapid collection generally does not compensate for the improved time resolution.

Most instruments provide the user with output of the time varying background-corrected atomic signal as well as the background signal. This can be useful in making an evaluation of whether the background level is of sufficient level that one should question the accuracy of the correction made by the instrument. Again, the article **Background Correction Methods in Atomic Absorption Spectroscopy** discusses the use and interpretation of the background signals.

Output is generally sent to a computer or ''data station'' where elaborate work up of the data takes place. Most instruments with a computer interface and high level software package permit the user to specify a number of analytical options, for example number and concentration of standards, number of replicates, options for standard additions and frequency of QA/QC checks. Additionally, most have capabilities of adding matrix modifiers from a separate reservoir to each standard and sample. Most software packages also provide complete reports on the analysis including information on the analytical precision

as well as the accuracy and precision of the calibration curve. Additionally, these same software packages furnish full statistical analysis of the results and permit use of such analytical tools as standard additions.

4.7 Autosamplers and Other Accessories

ETAAS can be highly automated. The key ingredient to such automation is the use of an autosampler for sample introduction. The autosampler (under computer control) not only frees up the operator, but it also provides one of the most precise means of micropipetting the sample into the furnace (i.e. $<\pm 3\%$). Many modern systems permit user input to the autosampler regarding the number of elements to be determined and number of replicates in each sample. The instrument then analyzes each sample and controls the spectrometer wavelength drive and source selector to cover all the elements of interest. In such an instrument, of course, the spectrometer must be set-up with a multilamp turret so that the lamps can be changed as one moves from sample to sample. Because of the improved precision and flexibility provided by the autosampler, most laboratories consider it as an essential part of a modern ETAAS system.

Other analytical schemes including flow injection analysis (see **Flow Injection Analysis Techniques in Atomic Spectroscopy**), hydride generation techniques (see **Flame and Vapor Generation Atomic Absorption Spectrometry**), slurries (see **Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses**) and solids directly are all available for use with ETAAS. In all the cases cited above, including direct solids analysis, high degrees of automation and sample handling are commercially available.

5 MATRIX MODIFIERS

A ''matrix modifier'' originally referred to a species that was added to the sample to improve the analytical accuracy and/or precision. It could be as simple as the addition of HNO_3 to assist in elimination of chloride as HCl during the dry cycle. The use of modifiers represents a cornerstone in practical utilization of ETAAS for chemical analysis. While there are literally hundreds of papers expounding on the virtues of various modifiers for particular analyte/matrix scenarios, most have been arrived at empirically and in most cases there is a minimal amount known about the actual physico-chemical interactions responsible for the observed improvements. Regardless of this unfortunate fact, many modifiers have sufficient documentation that one can feel comfortable that they will assist – if not ensure – the achievement of reliable analytical results. In all cases there are one or more fundamental functions that the modifier is intended to serve.

Matrix Removal. Some modifiers assist in the prevolatilization of the matrix prior to the appearance temperature of the analyte in the gas phase. The removal may minimize coincidence of analyte and a matrix component in the vapor that would otherwise form a stable analyte-containing molecule and lead to a depressed signal. Examples include HNO_3 or NH_4^+ for halide removal or use of oxygen ashing for combustion of biological compounds.

Analyte ''Stabilization''. By forming a more stable condensed phase species on the atomizer surface, higher thermal pretreatment temperatures may be used to remove unwanted matrix components. Examples include the addition of phosphate salts or the use of Pd (usually as the nitrate) to the sample and standards.

One of the more commonly used modifiers is Pd (usually as the nitrate salt). This appears to serve as a stabilizer for a number of metals. Obviously, it should not ''stabilize'' all species or it really serves no useful purpose except to delay vaporization of every sample component. The inclusion of $Mg(NO_3)_2$ in this solution has also been suggested to assist in the performance of the Pd modifier by acting as a ''bulking agent''.

Surface Modification. A number of studies have modified the graphite surface in an attempt to enhance the atomization conditions of various analyte metals. Surface modification often involves the addition of large amounts of a metal salt (as a solution) to the furnace followed by thermal pretreatment to form, for example, a metal carbide coating on the surface. Metals have also been sputtered onto the graphite surface to accomplish the same purpose. This modified surface has been reported to improve atomization efficiencies and provide immunity to matrix problems. Metals that have been used include Ta and Ir. In contrast to the use of more conventional modifiers, these surface modification techniques need not be done prior to each firing of the atomizer.[30]

There are certainly a large number of modifiers that have been used to satisfy certain analyte/matrix combinations. Some modifiers appear to be useful only for selected analyte/matrix combinations (e.g. Ni for semimetals) while others (e.g. Pd) have a broader application base. Tsalev and Slaveykova[31] used multivariate techniques in an attempt to place some order in those that are reported to have worked. There is little doubt that many can be understood based on their chemical properties, but the mechanism for others remains elusive.

6 PRACTICAL APPLICATION OF GRAPHITE FURNACE ATOMIC ABSORPTION SPECTROMETRY TO ANALYSIS

Most instruments come equipped with a methods manual or ''cookbook'', which provides information on analyte-specific determinations. In many instances, procedures are often included for specific matrices; and in the newer instruments, this manual is on-line, that is, part of the data station or computer system that operates the system. Additionally, there are a large number of method development papers in the literature discussing specific analyte determinations in specific matrices that can be employed. These procedures can be used as an excellent starting point to initiate the analysis. This encyclopedia also presents some key analytical areas where ETAAS (also known as GFAAS) are employed (see **Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses; Atomic Spectroscopy: Introduction; Background Correction Methods in Atomic Absorption Spectroscopy; Flame and Vapor Generation Atomic Absorption Spectrometry;** and **Flow Injection Analysis Techniques in Atomic Spectroscopy**.)

As applied and fundamental research progressed, several instrumentation and methodological characteristics were generally acknowledged as adding significantly to analytical accuracy and precision. Slavin et al.[32] attached the label of 'stabilized temperature platform furnace' conditions to these features. The embodiment of the general approach advocated the use of pyrocoated furnaces with a platform, rapid heating, fast responding digital electronics to accurately capture the transient signal, peak area (in place of peak height) for quantitation, Zeeman-effect background correction, and chemical modifiers. Some debate may persist regarding the universal applicability for all of these guidelines; but they are, in general, quite sound.

To follow is a generic approach to the set-up of a new method.

6.1 Setting Up Temperature Programs

6.1.1 Sample Deposition and Dry Cycle

While depositing the sample, one can look through the end of the furnace toward the light source and see a shadow of the autosampler tubing that can be manually positioned in its deposition position. The tube should not make contact with the dosing hole sides, graphite wall or platform. When the droplet is deposited it should be ''pulled'' away from the tubing so that the tubing is not resting in the liquid after the sample is dispensed. An angled cut on the tubing often assists in improved sample deposition. The drying temperature should be set to a ramp of 5–10 s duration to ca. 80 °C (10–20 °C higher with platform) to begin and held at this temperature for 20–30 s. The criterion for adequate drying is to see the droplet gradually disappear over a period of 5–10 s with no noticeable boiling or bumping. Variations in temperature and time can be adjusted to achieve this. Drying is one of the more important steps since improper drying will lead to accuracy and precision problems as well as tube degradation. Bearing in mind the temperature setting and display on most instruments is not measured during the cycle, the error could be several tens of degrees at the low temperature settings.

6.1.2 Thermal Pretreatment (''Char'')

It is almost always advisable to have a thermal pretreatment step if for no other reason than to eliminate any residual water and acids as well as some of the volatile matrix components in the sample. Additionally, this step is useful in decomposing many of the inorganic salts, for example, nitrates.

It is most critical that the analyte is not lost during this step. To evaluate this, ''char curves'' can be constructed. Simply set the dry to the predetermined setting and the atomization to a sufficiently high value to feel confident in achieving atomization, for example from standard procedure. Next, make a series of firings using your sample where the thermal pretreatment temperature is changed by ca. 100 °C but the time program is kept constant. A typical program might be a 5 s ramp to the temperature and hold for 20 s. The absorbance peak area is then recorded for each char temperature used. A curve similar to that shown in Figure 8(a) will result. The highest temperature is then selected where there is no attenuation of the peak area for the thermal pretreatment, ~900–1000 °C in this case.

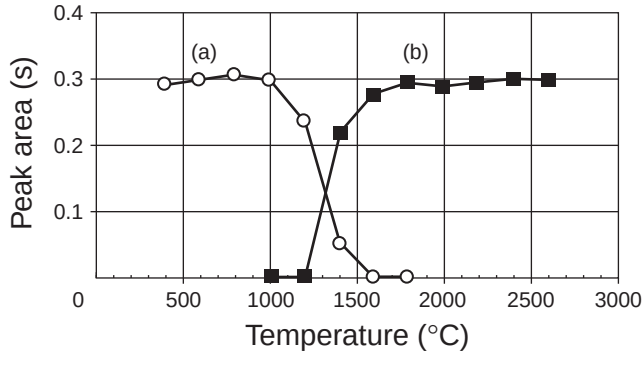

Figure 8 (a) Char and (b) atomization curves collected for the purpose of determining the correct thermal pretreatment (i.e. ''char'') and atomization temperature settings.

6.1.3 Atomization Step

While a small gas flow through the atomizer is generally employed during the dry and thermal pretreatment steps, the atomization step is generally used in a ''gas stop'' mode (i.e. no flow through the furnace) to maximize residence time and to ensure maximum gas-phase temperatures.

The temperature setting for this step is critical. Too low an atomization temperature can produce a weak signal with memory effects while too high a temperature can cause excessive degradation of the furnace. Signs of too low a temperature include a badly tailing absorption peak. In most instances ''step heating'' (0 s ramp) is used for the atomization cycle. If a platform is used it is best to cool the atomizer to ca. 100 °C before initiating the atomization cycle. This provides the maximum time delay in the heating of the platform and should provide optimal results. After the ramp, the atomization temperature is usually held for ca. 5 s, sometimes longer if a platform is used and there is difficulty in removing the sample. If possible, it is always good to check the time-dependent temperature and absorbance signals to verify that a sufficient but not excessive hold time is being used.

Similar to the construction of a char curve, an atomization curve can be constructed. Using the previously determined dry and thermal pretreatment values and varying the atomization temperature, and noting the peak area, Figure 8(b) will result. An atomization temperature has to be selected where a minimum temperature yields the maximum area, ~1900–2000 °C in this case. If the area continues to rise even with the maximum temperature used, it is possible that the analyte is sticking to the graphite. If a platform is being used, you may have to use wall atomization (i.e. remove the platform). If wall atomization is used and the problem seems to persist at the highest temperature available, you can verify complete vaporization by checking for memory effects by running a blank after a sample (or standard solution). The presence of a signal in the blank suggests that the analyte is not being completely removed, assuming that the blank is analyte-free. Some modifier or change in sample preparation may be needed.

6.1.4 Clean Cycle

For some samples a short (5–20 s) clean cycle near the maximum temperature of the atomizer is advisable. This is essential if an analyte signal or background signal persists when a blank is fired immediately after a sample solution. Many analysts use this cycle as a routine part of their thermal program.

It is not uncommon to conduct a quick check of the instrument and its settings at the start of the analysis by computing the experimentally determined characteristic mass, m_o, with that provided by the manufacturer. The characteristic mass is the mass of analyte needed to register a peak area of 0.0044 s. Some manufacturers quote other figures of merit (e.g. sensitivity values), which can be used in a similar fashion to that of characteristic mass. Most instrument manuals or on-line documentation provide these values for a large number of elements, and the value is element and instrument-dependent. As an example, m_o values range from a few tenths of a picogram to several tens of picograms, depending on the element and the instrument. The easiest way to determine m_o is to analyze a standard that provides a peak area of ca. 0.05–0.2. By multiplying the analyte mass in this sample by 0.0044 and dividing by the peak area of the signal, the resulting value will be m_o. While a value checking sensitivity often lies within ±20% of the manufacturer's tabulated value, a value that is no more than about two times the quoted value is reasonable assurance that there is no significant error in the methodology, instrument settings or lamp performance. Values that are much too high suggest that something is not quite right, for example lamp current is too high, temperature program is not correct, etc.

6.2 Sample Types

In most instances, ETAAS is used with 0–50 μL aliquots (20 μL being typical) of aqueous solutions, although analysis of slurries (see **Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses**) and solids are routinely conducted using this technique. Both of these approaches minimize sample handling and minimize or eliminate contamination because they minimize reagents needed to prepare a homogeneous solution sample. As alluded to earlier, gases (e.g. hydrides, as alluded to and discussed in **Flame and Vapor Generation Atomic Absorption Spectrometry**) and even aerosols have been analyzed by direct deposition into the furnace.

6.3 Calibration Curves

When a large number of samples are anticipated, the use of calibration curves is generally preferred over the method of standard additions. The method of standard additions can correct for enhancements or depressions that are multiplicative but cannot correct for additive problems such as background correction errors. It is not uncommon to encounter nonlinear calibration curves and most exhibit a dynamic working range of two to three orders of magnitude. With Zeeman background correction, roll-over may limit this range even more. Like FAA, the nonlinearity in the calibration curves is a result of a combination of factors including the finite width of the absorbing and emitting line, secondary lines within the

bandpass of the spectrometer and stray light reaching the detector. It has been suggested that contributions from nonuniform analyte distribution in the furnace combined with nonuniform spatial intensities from the line source may also contribute.[15] Because of the nonlinearity, one must exercise caution in using standard additions since this analytical approach assumes linear extrapolation of the standard additions data to the intercept on the concentration axis. Most commercial software packages permit curve fitting to nonlinear curves, which optimizes the useable dynamic range.

Critical to all analyses by ETAAS is the use of high purity reagents (e.g. acids, water, etc.) because of the sensitivity inherent to the technique. As a standard analytical procedure, it is also advisable to employ standard reference materials (SRMs) as a quality control check during methods development and as a routine part of the analyses.

6.4 Absolute Analysis

The general approach of absorption spectrometry has possibilities of conducting analysis without the use of standards, that is, absolute analysis. As an example, ultraviolet/visible solution spectrophotometry with known solvent and solution environment can be used for analysis working with Beer's law and knowledge of the proportionality between absorbance and concentration, that is, the molar absorptivity. It is acknowledged that there are certain caveats to do this successfully, but it can be done both in theory and practice. The success rests, in part, because absorption spectroscopy is a ratioing technique of the incident intensity with and without an absorber in the light path, that is, transmittance, and concentration is proportional to $\log(1/T)$.

Variables in nebulizer efficiency, flame gas flows (as associated dispersion of analyte in the flame), optical path of source light through the flame, etc., eliminated the application of absolute analysis to FAA. However, L'vov[33] recognized in his earliest work that this may be possible using ETAAS since a fixed volume was deposited in the furnace, atomization efficiency could be very nearly 100% and the full analyte could be retained in the analysis volume. (The latter fact was later modified to account for the continuous loss, and replaced by an assumption that the time spent by the analyte in the optical path could be constant and dependent only on furnace geometry if the analysis was conducted properly.) In essence, the approach suggested that one could calculate the peak area of the absorbance signal from fundamental spectroscopic constants combined with furnace geometry, thereby alleviating the need for preparing standard solutions and calibration curves. The potential of this concept was supported by experimental data from several laboratories.[34,35] Agreement between calculated absorbance values and experimental values were generally within 20% of each other. This should be very encouraging especially when such accuracy may really be all that is needed in many samples where parts per billion levels are being determined. However, the conventional use of calibration curves with ETAAS often yields accuracy and precision values of $<5\%$, which many regulatory agencies and ''customers'' of the analyst demand. To date, this potentially powerful feature of absolute analysis by ETAAS has not been routinely employed in any major segment of the analytical community.

7 CONCLUSION

Graphite furnace AA exists in the modern trace metal arsenal as a reliable and rugged analytical tool. With modern use of the technique, high precision (3–5% relative standard deviation) and accurate results can be obtained on a routine basis. With absolute LODs in the low picogram to subpicogram range (low parts per billion for relative LODs), it also becomes one of the most sensitive elemental analytical techniques. The ability to handle micro amounts of sample as well as complex samples (e.g. slurries, high dissolved salt matrices and even solids) has made ETAAS a very attractive option when limited sample amounts or complex matrices are encountered. Like all atomic spectroscopic techniques, the production of free atoms is essential to the analytical success of ETAAS. With modern furnace designs, matrix modifiers and a wide range of published procedures, the analyst has the key ingredients for obtaining reliable results.

Since its inception, a large number of fundamental studies have been conducted on the ETA and considerable knowledge gained on the processes preceding the formation of the free atoms. Many of the improvements in this technique have resulted from judicious use of this information. Similarly, this basic information is often of use in efficiently extracting solutions to analytical problems that are encountered with a new analyte or matrix.

ACKNOWLEDGMENTS

I would like to thank N.J. Miller-Ihli, Scott Baker and Debra Bradshaw for the comments and suggestions provided during the preparation of this manuscript.

ABBREVIATIONS AND ACRONYMS

AA	Atomic Absorption
EDL	Electrodeless Discharge Lamp

ETA	Electrothermal Atomizer
ETAAS	Electrothermal Atomic Absorption Spectrometry
FAA	Flame Atomic Absorption
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HCL	Hollow Cathode Lamp
LOD	Limit of Detection
QA/QC	Quality Assurance/Quality Control
SRM	Standard Reference Material

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges)

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy

REFERENCES

1. A. Walsh, 'The Application of Atomic Absorption Spectra to Chemical Analysis', *Spectrochim. Acta*, **7**, 108–117 (1955).
2. A.S. King, 'The Production of Spectra by an Electrical Resistance Furnace in Hydrogen Atmosphere', *Astrophys. J.*, **27**, 353 (1908).
3. B.V. L'vov, 'The Analytical Use of Atomic Absorption Spectra', *Spectrochim. Acta*, **17**, 761–770 (1961).
4. R. Woodriff, G. Ramelow, 'Atomic Absorption Spectroscopy with a High-temperature Flame', *Spectrochim. Acta, Part B*, **23**, 665–671 (1968).
5. R.W. Woodriff, 'Atomization Chambers for Atomic Absorption Spectrochemical Analysis: A Review', *Appl. Spectrosc.*, **28**, 413–416 (1974).
6. H. Massmann, 'Vergleich von Atomabsorption und Atomfluoreszenz in der Graphitkuvette', *Spectrochim. Acta, Part B*, **23**, 215–226 (1968).
7. B. Welz, M. Sperling, G. Schlemmer, N. Wenzel, G. Marowsky, 'Spatially and Temporally Resolved Gas Phase Temperature Measurements in a Massmann-type Graphite Tube Furnace Using Coherent Anti-stokes Raman Scattering', *Spectrochim. Acta, Part B*, **43**, 1187–1207 (1988).
8. W. Frech, D.C. Baxter, B. Hutsch, 'Spatially Isothermal Graphite Furnace for Atomic Absorption Spectrometry Using Side-heated Cuvettes with Integrated Contacts', *Anal. Chem.*, **58**, 1973–1977 (1986).
9. B.V. L'vov, L.A. Pelieva, A.I. Sharnopolsky, *Zh. Prikl. Spektrosk.*, **27**, 395 (1977).
10. R.D. Ediger, 'Atomic Absorption Analysis with the Graphite Furnace Using Matrix Modification', *At. Absorpt. Newsl.*, **14**, 127–130 (1975).
11. R.F. Lonardo, A.I. Yuzefovsky, J.X. Zhou, J.T. McCaffrey, R.G. Michel, 'Extension of Working Range in Zeeman Graphite Furnace Atomic Absorption Spectrometry by Nonlinear Calibration with Prior Correction for Stray Light', *Spectrochim. Acta, Part B*, **51**, 1309–1323 (1996).
12. M.T.C. d. Loos-Vollebregt, L. de Galan, 'Stray Light in Zeeman and Pulsed Hollow Cathode Lamp Atomic Absorption Spectrometry', *Spectrochim. Acta, Part B*, **41**, 597–610 (1986).
13. P.L. Larkins, 'Atomic Line Profiles – Their Measurement and Importance in Analytical Atomic Spectroscopy', *J. Anal. At. Spectrom.*, **7**, 265 (1992).
14. R.J. Lovett, 'The Influence of Temperature on Absorbance in Graphite Furnace Atomic Absorption Spectrometry. I. General Considerations', *Appl. Spectrosc.*, **39**, 778–786 (1996).
15. A.K. Gilmutdinov, K.Y. Nagulin, Y.A. Zakharov, 'Analytical Measurement in Electrothermal Atomic Absorption Spectrometry. How Correct is It?', *J. Anal. At. Spectrom.*, **9**, 643–650 (1994).
16. C. Schnurer-Patschan, A. Zybin, H. Groll, K. Niemax, 'Improvement in Detection Limits in Graphite Furnace Diode Laser Atomic Absorption Spectrometry by Wavelength Modulation Technique', *J. Anal. At. Spectrom.*, **8**, 1103–1107 (1993).
17. H. Groll, K. Niemax, 'Multielement Diode Laser Atomic Absorption Spectrometry in Graphite Tube Furnaces

and Analytical Flames', *Spectrochim. Acta, Part B*, **48**, 633–641 (1993).
18. R.E. Sturgeon, C.L. Chakrabarti, C.H. Langford, 'Studies on the Mechanism of Atom Formation in Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chem.*, **48**, 1792–1807 (1976).
19. B. Smets, 'Atom Formation and Dissipation in Electrothermal Atomization', *Spectrochim. Acta, Part B*, **35**, 33–41 (1980).
20. O.A. Güell, J.A. Holcombe, 'Analytical Applications of Monte Carlo Techniques', *Anal. Chem.*, **62**, 529A–542A (1990).
21. V. Majidi, J.D. Robertson, 'Investigation of High Temperature Reactions on Solid Substrates with Rutherford Backscattering Spectrometry: Interaction of Palladium with Selenium on Heated Graphite Surfaces', *Spectrochim. Acta, Part B*, **46**, 1723–1733 (1991).
22. C.C. Eloi, J.D. Robertson, V. Majidi, 'Rutherford Backscattering Spectrometry Investigation of the Effects of Oxygen and Hydrogen Pretreatment of Pyrolytically Coated Graphite on Pb Atomization', *Anal. Chem.*, **67**, 335–340 (1995).
23. R.E. Sturgeon, K.W.M. Siu, G.J. Gardner, S.S. Berman, 'Carbon–Oxygen Reactions in Graphite Furnace Atomic Absorption Spectrometry', *Anal. Chem.*, **58**, 42–50 (1986).
24. A.K. Gilmutdinov, C.L. Chakrabarti, J.C. Hutton, 'Three-dimensional Distributions of Oxygen in Graphite and Metal Tube Atomizers for Analytical Atomic Spectrometry', *J. Anal. At. Spectrom.*, **7**, 1047–1062 (1992).
25. R.E. Sturgeon, H. Falk, 'Spectroscopic Measurement of Carbon Monoxide in a Graphite Furnace', *Spectrochim. Acta, Part B*, **43**, 421–428 (1988).
26. A. Cedergren, W. Frech, E. Lundberg, 'Estimation of Oxygen Pressure in Graphite Furnaces for Atomic Absorption Spectrometry', *Anal. Chem.*, **56**, 1382–1387 (1984).
27. A.K. Gilmutdinov, Y.A. Zakharov, V.P. Ivanov, A.V. Voloshin, K. Dittrich, 'Shadow Spectral Filming: A Method of Investigating Electrothermal Atomization. Part 2. Dynamics of Formation and Structure of the Absorption Layer of Aluminum, Indium and Gallium Molecules', *J. Anal. At. Spectrom.*, **7**, 675–683 (1992).
28. N. Hadgu, W. Frech, 'Performance of Side-heated Graphite Atomizers in Atomic Absorption Spectrometry Using Tubes with End Caps', *Spectrochim. Acta, Part B*, **49**, 445–457 (1994).
29. D.A. Katskov, R.I. McCrindle, R. Schwarzer, P.J.J.G. Marais, 'The Graphite Filter Furnace: A New Atomization Concept for Atomic Spectroscopy', *Spectrochim. Acta, Part B*, **50**, 1543–1555 (1995).
30. C.J. Rademeyer, B. Radziuk, N. Romanova, N.P. Skaugset, A. Skogstad, Y. Thomassen, *J. Anal. At. Spectrom.*, **10**, 739–745 (1995).
31. D.L. Tsalev, V.I. Slaveykova, 'Chemical Modification in Electrothermal Atomic Absorption Spectrometry. Organization and Classification of Data by Multivariate Methods', *J. Anal. At. Spectrom.*, **7**, 147–153 (1992).
32. W. Slavin, D.C. Manning, G.R. Carnrick, 'The Stabilized Temperature Platform Furnace', *At. Spectrosc.*, **2**, 137–143 (1981).
33. B.V. L'vov, 'Potentialities of the Graphite Crucible Method in Atomic Absorption Spectroscopy', *Spectrochim. Acta, Part B*, **24**, 53–70 (1969).
34. W. Frech, D.C. Baxter, 'Temperature Dependence of Atomization Efficiencies in Graphite Furnaces', *Spectrochim. Acta, Part B*, **45**, 867–886 (1990).
35. W. Slavin, G.R. Carnrick, 'The Possibility of Standardless Furnace Atomic Absorption Spectroscopy', *Spectrochim. Acta, Part B*, **39B**, 271–282 (1984).

Inductively Coupled Plasma-Optical Emission Spectrometry

Xiandeng Hou
Sichuan University, Chengdu, China

Bradley T. Jones
Wake Forest University, Winston-Salem, NC, USA

1 Introduction **1**
2 Theory **2**
2.1 Inductively Coupled Plasma Operation 2
2.2 Inductively Coupled Plasma Characteristics 3
3 Sample Introduction **4**
3.1 Nebulizers 4
3.2 Hydride Generation 6
3.3 Electrothermal Vaporization 6
3.4 Chromatographic Couplers 7
4 Torch Configuration **7**
4.1 Radial View 7
4.2 Axial View 8
4.3 Dual View 8
5 Detection of Emission **8**
5.1 Gratings 8
5.2 The Photomultiplier Tube 10
5.3 Array Detectors 12
5.4 Simultaneous versus Sequential Detection 13
6 Analytical Performance **13**
6.1 Analytical Wavelength 13
6.2 Quantitative Methods 14
6.3 Analytical Figures of Merit 15
6.4 Interferences 17
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **18**
Further Reading **19**
References **19**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

Inductively coupled plasma-optical emission spectrometry (ICP-OES) is a powerful tool for the determination of metals in a variety of different sample matrices. With this technique, liquid samples are injected into a radiofrequency (RF)-induced argon plasma using one of a variety of nebulizers or sample introduction techniques. The sample mist reaching the plasma is quickly dried, vaporized, and energized through collisional excitation at high temperature. The atomic emission emanating from the plasma is viewed in either a radial or axial configuration, collected with a lens or mirror, and imaged onto the entrance slit of a wavelength selection device. Single element measurements can be performed cost-effectively with a simple monochromator-photomultiplier tube (PMT) combination, and simultaneous multielement determinations are performed for up to 70 elements with the combination of a polychromator and an array detector. The analytical performance of such systems is competitive with most other inorganic analysis techniques, especially with regard to sample throughput and sensitivity.

1 INTRODUCTION

ICP-OES is one of the most powerful and popular analytical tools for the determination of trace elements in a myriad of sample types (Table 1). The technique is based upon the spontaneous emission of photons from atoms and ions that have been excited in an RF discharge. Liquid and gas samples may be injected directly into the instrument, while solid samples usually require extraction or acid digestion so that the analytes are present in a solution. The sample solution is converted to an aerosol and directed into the central channel of the plasma. At its core, the inductively coupled plasma (ICP) sustains a temperature of approximately 10 000 K, so the aerosol is quickly vaporized. Analyte elements are liberated as free atoms in the gaseous state. Further collisional excitation within the plasma imparts additional energy to the atoms, promoting them to excited states. Sufficient energy is often available to convert the atoms to ions and subsequently promote the ions to excited states. Both the atomic and ionic excited state species may then relax to the ground state via the emission of a photon. These photons have characteristic energies that are determined by the quantized energy level structure for the atoms or ions. Thus, the wavelength of the photons can be used to identify the elements from which they originated. The total number of photons is directly proportional to the concentration of the originating element in the sample.

The instrumentation associated with an ICP-OES system is relatively simple. A portion of the photons emitted by the ICP is collected with a lens or a concave

Table 1 Survey of elemental application areas of ICP-OES[1–8]

Categories	Examples of samples
Agricultural and food	Animal tissues, beverages, feeds, fertilizers, garlic, nutrients, pesticides, plant materials, rice flour, soils, vegetables, wheat flour
Biological and clinical	Brain tissue, blood, bone, bovine liver, feces, fishes, milk powder, orchard leaves, pharmaceuticals, pollen, serum, urine
Geological	Coal, minerals, fossils, fossil fuel, ore, rocks, sediments, soils, water
Environmental and water	Brines, coal fly ash, drinking water, dust, mineral water, municipal wastewater, plating bath, sewage sludge, slags, seawater, soil
Metals	Alloys, aluminum, high-purity metals, iron, precious metals, solders, steel, tin
Organic	Adhesives, amino acids, antifreeze, combustion materials, cosmetics, cotton cellulose, dried wood, dyes, elastomers, epoxy, lubricant, organometallic, organophosphates, oils, organic solvent, polymers, sugars
Other materials	Acids, carbon, catalytic materials, electronics, fiber, film, packaging materials, paints and coatings, phosphates, semiconductors, superconducting materials

mirror. This focusing optic forms an image of the ICP on the entrance aperture of a wavelength selection device such as a monochromator. The particular wavelength of light exiting the monochromator is converted to an electrical signal by a photodetector. The signal is amplified and processed by the detector electronics, then displayed, and stored by a personal computer.

The characteristics of the ICP as an analytical atomic emission source are so impressive that virtually all other emission sources (such as the flame, microwave-induced plasma (MIP), direct current plasma (DCP), laser-induced plasma (LIP), and electrical discharge) have been relegated to specific, narrowly defined application niches. Indeed, even much of the application field originally assigned to flame atomic absorption spectrometry (FAAS) and graphite furnace atomic absorption spectrometry (GFAAS) has been relinquished to the ICP. Compared to these other techniques, ICP-OES enjoys a higher atomization temperature, a more inert environment, and the natural ability to provide simultaneous determinations for up to 70 elements. This makes the ICP less susceptible to matrix interferences, and better able to correct for them when they occur. In cases where sample volume is not limited, ICP-OES provides detection limits as low as, or lower than, its best competitor, GFAAS, for all but a few elements. Even for these elements, the simplicity with which the ICP-OES instrument is operated often outweighs the loss in sensitivity.

2 THEORY

The ICP was developed for OES by Fassel et al. at Iowa State University in the United States and by Greenfield et al. at Albright & Wilson, Ltd. in the United Kingdom in the mid-1960s.[1,9,10] The first commercially available ICP-OES instrument was introduced in 1974. The ICP is now not only the most popular source for OES but it is also an excellent ion source for mass spectrometry: inductively coupled plasma mass spectrometry (ICPMS).[11] ICP-OES is a proven commercial success, and the future is still bright for ICP-based spectroscopic techniques. Detectability has been continuously and dramatically improved over the past 45 years. Detection limits, for example, have improved by a factor of 4–6 orders of magnitude for many elements. Nevertheless, research and commercial opportunities for further development of ICP-OES remain intriguing.[11]

2.1 Inductively Coupled Plasma Operation

As shown in Figure 1, the so-called ICP torch is usually an assembly of three concentric fused-silica tubes. These are frequently referred to as the *outer, intermediate*, and *inner gas tubes*. The diameter of the outer tube ranges from 9 to 27 mm. A water-cooled, two- or three-turn copper coil, called the *load coil*, surrounds the end of the torch and is connected to an RF generator. The outer argon flow (10–15 L min^{-1}) sustains the high-temperature plasma, and positions the plasma relative to the outer walls and the induction coil, preventing the walls from melting and facilitating the observation of emission signals. The plasma under these conditions has an annular shape. The sample aerosol carried by the inner argon flow (0.5–1.5 L min^{-1}) enters the central channel of the plasma and helps to sustain the shape. The intermediate argon flow (0–1.5 L min^{-1}) is optional and may serve the functions of pushing the plasma slightly away from the end of the central sample tube, or diluting the inner gas flow in the presence of organic solvents.

The ICP is generated as follows. RF power, typically 700–1500 W, is applied to the load coil and an alternating current oscillates inside the coil at a rate corresponding to the frequency of the RF generator. For most ICP-OES instruments, the RF generator has a frequency of either 27 or 40 MHz. The oscillation of current at this high

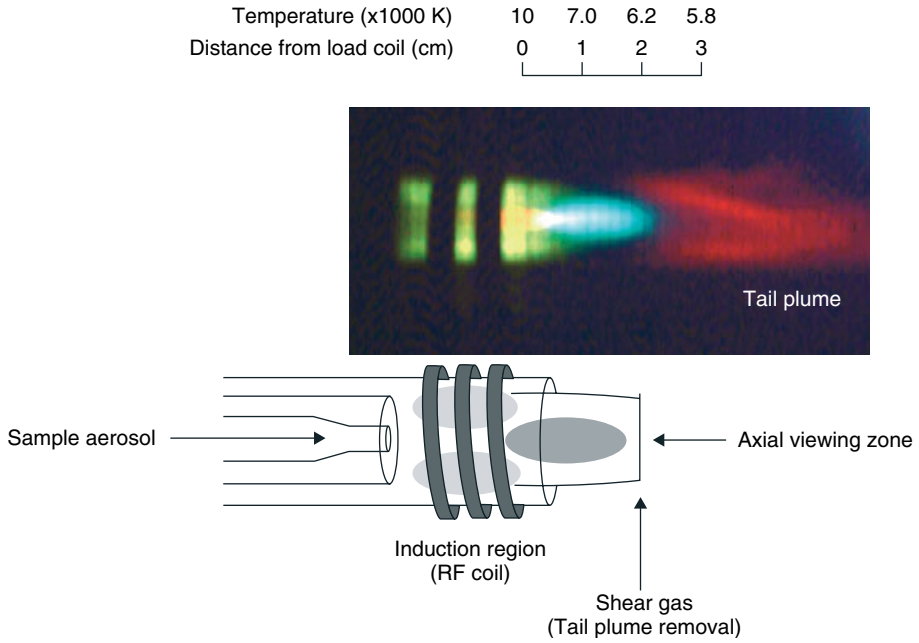

Figure 1 Schematic diagram and actual photograph of an ICP assembly showing the three concentric tubes composing the torch, the RF coil, the different plasma regions, and the temperature as a function of distance from the load coil.

frequency causes the same high-frequency oscillation of electric and magnetic fields to be set up inside the top of the torch. With argon gas flowing through the torch, a spark from a Tesla coil is used to produce ''seed'' electrons and ions in the argon gas inside the load coil region. These ions and electrons are then accelerated by the magnetic field, and collide with other argon atoms, causing further ionization in a chain-reaction manner. This process continues until a very intense, brilliant white, teardrop-shaped, high-temperature plasma is formed. Adding energy to the plasma via RF-induced collision is known as *inductive coupling*, and thus the plasma is called an *ICP*. The ICP is sustained within the torch as long as sufficient RF energy is applied.[1] In a cruder sense, the coupling of RF power to the plasma can be visualized as positively charged Ar ions in the plasma gas attempting to follow the negatively charged electrons flowing in the load coil as the flow changes direction 27 million times/s. The multitude of collisions produced in this environment gives rise to the energy transferred to the sample. No combustion reactions occur in the plasma.

Figure 1 shows the temperature gradient within the ICP with respect to distance from the load coil. It also gives the nomenclature for some of the different zones of the plasma as suggested by Koirtyohann et al.[12] The induction region (IR) encircled by the load coil is ''doughnut-shaped'' as described above, and it is the region where the inductive energy transfer occurs. This is also the region of highest temperature and it is characterized by a bright continuum emission. From the IR outward toward the tail plume, the temperature decreases.

An aerosol, or very fine mist of liquid droplets, is generated from a liquid sample by the use of a nebulizer. The aerosol is carried into the center of the plasma by the argon gas flow through the IR. Upon entering the plasma, the droplets undergo three processes. The first step is desolvation, or the removal of the solvent from the droplets, resulting in microscopic solid particulates, or a dry aerosol. The second step is vaporization, or the decomposition of the particles into gaseous-state molecules. The third step is atomization, or the breaking of the gaseous molecules into atoms. Finally, excitation and ionization of the atoms occur, followed by the emission of radiation from these excited species. The analytical signal is normally observed through an end-on (axial) viewing position.

2.2 Inductively Coupled Plasma Characteristics

The main analytical advantages of the ICP over other excitation sources originate from its capability for efficient and reproducible vaporization, atomization, excitation, and ionization for a wide range of elements in various sample matrices. This is mainly due to the high temperature, 6000–8000 K, in the observation zones of the ICP. This temperature is much higher than the maximum temperature of flames or furnaces (3300 K). The high temperature of the ICP also makes it capable of exciting refractory elements and renders it less prone to matrix interferences. Other electrical-discharge-based

sources, such as alternating current and direct current arcs and sparks, and the MIP, also have high temperatures for excitation and ionization, but the ICP is typically less noisy and better able to handle liquid samples. In addition, the ICP is an electrodeless source, so there is no contamination from the impurities present in an electrode material. Furthermore, it is relatively easy to build an ICP assembly and it is inexpensive, compared to some other sources, such as a LIP. The following is a list of some of the most beneficial characteristics of the ICP source:

- high temperature (6000–8000 K);
- high electron density (10^{14}–10^{16} cm^{-3});
- appreciable degree of ionization for many elements;
- simultaneous multielement capability (over 70 elements including P and S);
- low background emission, and relatively low chemical interference;
- high stability leading to excellent accuracy and precision;
- excellent detection limits for most elements (0.03–10 $ng\,mL^{-1}$);
- wide linear dynamic range (LDR) (4–6 orders of magnitude);
- applicable to the refractory elements;
- cost-effective analyses.

3 SAMPLE INTRODUCTION

An introduction system is used to transport a sample into the central channel of the ICP as a gas, vapor, aerosol of fine droplets, or solid particles. The general requirements for an ideal sample introduction system include amenity to samples in all phases (solid, liquid, or gas), tolerance to complex matrices, the ability to analyze very small amounts (<1 mL or <50 mg), excellent stability and reproducibility, high transport efficiency, simplicity, and low cost.[2] A wide variety of sample introduction methods have been developed, such as nebulization, hydride generation (HG), electrothermal vaporization (ETV), laser ablation, and direct sample insertion (DSI).[1,13–17] The analysis of liquid microsamples, at flow rates less than 0.1 $mL\,min^{-1}$, is also possible.[18,19] The most popular sample introduction methods are described below.

3.1 Nebulizers

Nebulizers are the most commonly used devices for solution sample introduction in ICP-OES. With a nebulizer, the sample liquid is converted into an aerosol and transported to the plasma. Both pneumatic and ultrasonic nebulizers (USNs) have been successfully used in ICP-OES. Pneumatic nebulizers make use of high-speed gas flows to create an aerosol, while the USN breaks liquid samples into a fine aerosol by the ultrasonic oscillations of a piezoelectric crystal. The formation of aerosol by the USN is therefore independent of the gas flow rate.

Only very fine droplets (about 8 μm in diameter) in the aerosol are suitable for injection into the plasma. A spray chamber is placed between the nebulizer and the ICP torch to remove large droplets from the aerosol and to dampen pulses that may occur during nebulization. Thermally stabilized spray chambers are sometimes adopted to decrease the amount of liquid introduced into the plasma, thus providing stability especially when organic solvents are involved. Pneumatic nebulization is very inefficient, however, because only a very small fraction (less than 5%) of the aspirated sample solution actually reaches the plasma. Most of the liquid is lost down the drain in the spray chamber. However, the pneumatic nebulizer retains its popularity owing to its convenience, reasonable stability, and ease of use. Efficiency may only be a concern when sample volumes are limited or measurements must be performed at or near the detection limit.

Three types of pneumatic nebulizers are commonly employed in ICP-OES: the concentric nebulizer, the cross-flow nebulizer, and the Babington nebulizer (Figure 2). The concentric nebulizer is fashioned from fused silica. The sample solution is pumped into the back end of the nebulizer by a peristaltic pump. Liquid uptake rates may be as high as 4 $mL\,min^{-1}$, but lower flows are more common. The sample solution flows through the inner capillary of the nebulizer. This capillary is tapered so that flexible tubing from the pump is attached at the

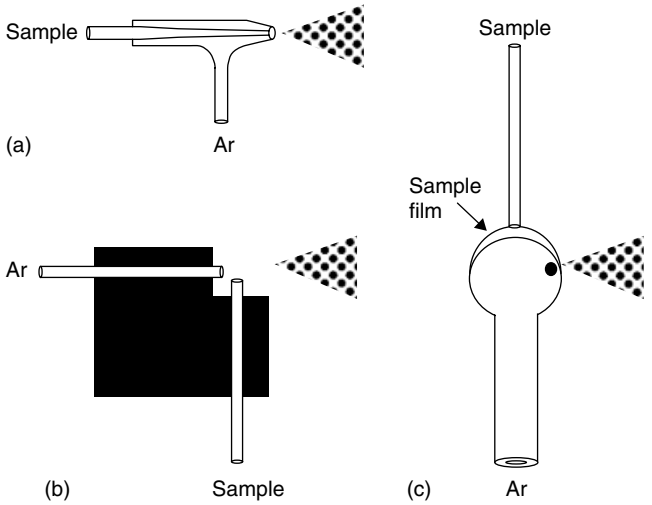

Figure 2 Schematic diagrams of three types of pneumatic nebulizer: (a) the concentric nebulizer; (b) the cross-flow nebulizer; and (c) the Babington nebulizer.

entrance (4-mm outer diameter) and the exit has a narrow orifice approaching 100 μm or less in inner diameter. Ar gas (0.5–1.5 L min^{-1}) is supplied at a right angle into the outer tube. This tube is also tapered so that the exit internal diameter approaches the outer diameter for the sample capillary. As the Ar gas passes through this narrow orifice, its velocity is greatly increased, resulting in the shearing of the sample stream into tiny droplets. Concentric nebulizers have the advantages of excellent sensitivity and stability, but the small fragile fused-silica orifices are prone to clogging, especially when aspirating samples of high salt content. Concentric nebulizers also require a fairly large volume of sample, given the high uptake rate. The microconcentric nebulizer (MCN) is designed to solve this problem. The sample uptake rate for the MCN is less than 0.1 mL min^{-1}. The compact MCN employs a smaller diameter capillary (polyimide or Teflon) and poly(vinylidine difluoride) body to minimize the formation of large droplets and to facilitate the use of hydrofluoric acid.

A second type of pneumatic nebulizer, the cross-flow nebulizer, is designed to reduce the clogging problem. In contrast to concentric nebulizers, cross-flow nebulizers use a high-speed stream of argon perpendicular to the tip of the sample capillary. Again the sample solution is broken into an aerosol, as shown in Figure 2. The drawbacks of the cross-flow nebulizer include lower sensitivity and potential capillary misalignment.

The third type of pneumatic nebulizer used for ICP-OES is the Babington nebulizer that allows a film of the sample solution to flow over a smooth surface having a small orifice (Figure 2).[20] High-speed argon gas emanating from the hole shears the sheet of liquid into small droplets. In some cases, the liquid may be guided over the argon orifice by positioning both the orifice and the sample introduction port at the base of a V-shaped groove (V-groove nebulizer). The essential feature of the Babington nebulizer is that the sample solution flows freely over a small aperture, rather than passing through a fine capillary, resulting in a high tolerance to dissolved solids. In fact, even slurries can be nebulized with a Babington nebulizer.[21] This type of nebulizer is the least susceptible to clogging and it can nebulize very viscous liquids.

The Hildebrand grid nebulizer (HGN) (Figure 3) may be considered a specialized version of the Babington nebulizer with many orifices. Often the nebulizer has a screw-cap design. The outer member (cap) of the nebulizer holds two parallel platinum screens or grids. The grids are separated by approximately 2 mm. The inner body of the nebulizer (screw) has a single sample channel. Liquid is pumped through this channel at rates up to 1 mL min^{-1}. The inner body also has a circular V-groove that allows the liquid to contact the entire

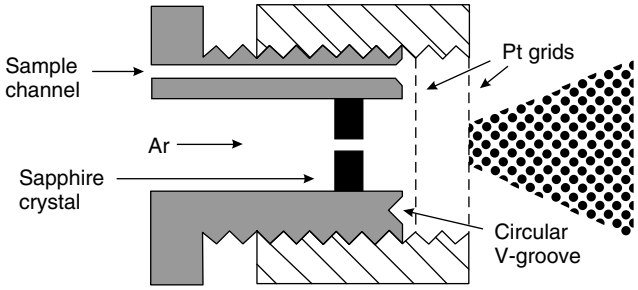

Figure 3 The HGN.

perimeter of the inner platinum grid. In this fashion, the liquid completely wets both grids. A high-velocity stream of argon (1 L min^{-1}) blows through the center of the nebulizer. Often the velocity of the argon is increased by placing a sapphire crystal containing a small orifice (0.2 mm) in the center of the gas stream. The high-velocity argon forces the liquid through the tiny openings in the screens producing a fine aerosol. The HGN is characterized by clog-free operation, high efficiency, and excellent stability.

With the USN, sample solution is first introduced onto the surface of a piezoelectric transducer that is operated at a frequency between 0.2 and 10 MHz. The longitudinal wave, which is propagated in the direction perpendicular to the surface of the transducer toward the liquid–air interface, produces pressure that breaks the liquid into an aerosol.[22] The efficiency of an USN is typically between 10 and 20%. This nebulizing efficiency is greater than that of a pneumatic nebulizer, and it is independent of argon flow rate. Therefore, a slower gas flow rate can be used to transport the aerosol to the plasma, thus prolonging the residence time of analyte in the plasma. This can result in improved sensitivity, and the limit of detection (LOD) may be lowered by a factor of 8–200, depending upon the element. However, the USN is more complicated, more expensive, and more susceptible to matrix effects, memory effects, and high solid loading. The USN is not compatible with hydrofluoric acid.

While the efficiency of each type of nebulizer depends upon the specific ICP-OES instrument employed (and its power, gas, and flow parameters), some comparisons are possible. Table 2 compares the relative magnitude of LODs observed for 11 elements using four different nebulizers. Since the USN provides the lowest LOD in each case, its value is set to 1, and the other LODs are higher by the factor listed in the table. Using these 11 elements as representative cases, on an average, LODs follow the trend USN < HG < concentric < V-groove.[3,23] Various other means of nebulization have been tested with limited degrees of success.[22]

Table 2 Relative ICP-OES LOD observed using different nebulizers[3,23]

Element	Wave length (nm)	USN[a]	HGN	Concentric	V-groove
Ag	328.1	1	9	9	17
Al	396.2	1	3	10	14
As	193.7	1	6	4	8
Cd	226.5	1	2	8	19
Co	228.6	1	3	6	11
Cr	267.7	1	3	7	11
Cu	324.8	1	4	12	25
Mn	267.6	1	7	8	12
Ni	231.6	1	2	3	6
Pb	220.4	1	7	5	9
Zn	213.9	1	4	6	8
Average		**1**	**5**	**7**	**13**

[a] The relative LOD is the ratio of the LOD using a given nebulizer to the LOD for the USN.

3.2 Hydride Generation

HG is a very effective sample introduction technique for some elements. These elements include arsenic, bismuth, germanium, lead, antimony, selenium, tin, and tellurium. In this method, the sample in diluted acid solution is mixed with a reducing agent, usually a solution of sodium borohydride in dilute sodium hydroxide. The reaction of sodium borohydride with the acid produces hydrogen. The hydrogen then reduces the analyte metal ions to hydrides, which are gaseous at ambient temperatures. The chemical reactions are shown in Equations (1) and (2).

$$NaBH_4 + 3H_2O + HCl \longrightarrow H_3BO_3 + NaCl + 8H \quad (1)$$

$$8H + E^{m+} \longrightarrow EH_n + H_2 \text{ (excess)} \quad (2)$$

where E is the hydride forming element of interest and m may or may not equal n.

The advantages of the HG technique include:

- physical separation of the analyte from possible matrix interferents;
- higher efficiency than conventional pneumatic nebulization;
- preconcentration of the analyte for better LOD;
- capability for inorganic and/or organic speciation;
- ease of automation when used with flow injection techniques.

On the other hand, several disadvantages of this technique may include:

- interference from those contaminants that reduce HG efficiency;
- slow reactions, necessitating hydride trapping prior to introduction;
- critical control of experimental conditions such as pH and reagent concentrations;
- extra influential factors, such as the oxidation state of the analyte of interest.

LODs for the hydride generating elements are listed in Table 3, together with those observed for conventional nebulization techniques. The LODs of HG-ICP-OES in Table 3 are the best LODs cited in Nakahara's review paper.[13] Compared with conventional pneumatic nebulization techniques, LODs achieved by the HG technique are enhanced by a factor between 10 and 1000, depending upon the element.

3.3 Electrothermal Vaporization

Electrothermal vaporization ETV has also been used to solve problems associated with pneumatic nebulization. Graphite furnaces or other electrothermal devices, such as carbon rods, carbon cups, graphite boats, graphite tubes, tungsten wire, and other metal filaments, have been used in research laboratories to electrothermally vaporize a liquid or solid sample for introduction into the ICP.[14,15,24] Other vaporization methods, such as arc/spark vaporization and laser ablation/vaporization, have also been used as a means for sample introduction in ICP-OES. Even the ICP itself has been used to vaporize samples into a second ICP for analytical measurements.[25] In a typical experiment, a low current is applied to the ETV to remove the sample solvent. A small portion of the sample is then vaporized by the device through the application of a high current. An optional ''ash'' step may be used to remove some of the matrix prior to the analyte vaporization step. The resulting dense cloud of the analyte vapor is then efficiently swept into the center of the plasma by a flow of argon gas.

A commercial graphite furnace designed for atomic absorption spectrometry (AAS) is most frequently used

Table 3 LOD observed by hydride generation, HG-ICP-OES[1,13]

Elements	LOD by HG ($ng\,mL^{-1}$)	LOD by conventional nebulization ($ng\,mL^{-1}$)	Ratios (conventional/HG)
As	0.03	20	667
Bi	0.06	20	333
Ge	0.3	20	67
Pb	1.0	10	10
Sb	0.07	10	143
Se	0.04	50	1250
Sn	0.2	30	150
Te	0.04	10	250

in ETV-ICP-OES. The major advantage of ETV as a means of sample introduction is that the transportation efficiency is dramatically improved over a pneumatic nebulizer, from less than 5% to over 60%. Consequently, the LODs are improved by at least an order of magnitude. Some difficult-to-analyze samples, those with high total dissolved solids (TDS), for example, can be introduced by the ETV. However, since these devices are generally not of a continuous-flow nature, the ICP instrument has to be capable of recording transient signals. Also, the simultaneous multielement capability of the system could be limited owing to this transient nature of signals. Furthermore, when graphite material is used for the ETV, carbide formation could be a problem for some elements, resulting in lowered sensitivity and memory effects for refractory elements.

In an attempt to eliminate the problems associated with graphite, metal filaments have been employed for ETV-ICP-OES. For example, a tungsten coil from a commercial slide projector bulb can be used to vaporize liquid samples prior to their introduction into the ICP.[24] A small volume, typically 20 μL, of sample solution is delivered to the tungsten coil and dried at low current. Then, a higher current is applied to atomize the sample from the coil. The vapor is then rapidly introduced into the plasma as a dense plug by a flow of argon/hydrogen gas. The LOD is typically improved by 100–1500 times compared with pneumatic nebulization. These LODs are comparable to those obtained by GFAAS, but with the capability of simultaneous multielement measurement, and at a low cost. A tungsten loop has also been used as an in-torch vaporization (ITV) means for sample introduction to the ICP, and the operation can possibly be automated.[14] These approaches still share some of the other disadvantages associated with graphite furnace ETV, and commercial systems have not yet appeared.

3.4 Chromatographic Couplers

The combination of the separation power of chromatography and the detection power of atomic emission spectroscopy results in many advantages. One of the primary advantages of chromatography over conventional sample introduction is the ability to obtain speciation information.[26,27] When used as a detector for chromatographic methods, the ICP offers good sensitivity, wide LDR, and multielement detection capability. The multielement capability of the ICP, in turn, enhances the performance of chromatographic methods. Both gas chromatography (GC) and high-performance liquid chromatography (HPLC) can be coupled with ICP-OES.[28–30] Compared with HPLC, however, the GC-ICP coupling is less common because the analytical performance of ICP-OES is often not adequate for the typical non-metal-containing analytes separated by GC. Undoubtedly, GC-ICP-OES is still useful in the analyses of volatile organometallics, as demonstrated in the determination of methylmercury species.[31] The successful combination of these two techniques is realized through the use of chromatographic couplers. Fortunately, most of the interface systems currently in use are relatively inexpensive and easy to construct, and they require few, if any, modifications to commercial ICP-OES instruments.

Direct connections between the end of the HPLC column and the nebulizer suffer from poor transport efficiency and low tolerance to many of the organic solvents commonly employed in mobile phases for HPLC, particularly when a pneumatic nebulizer is used. To improve the transport efficiency and to minimize the influence of organic solvents on the stability of the ICP, USNs, water-cooled thermospray chambers, and glass-frit nebulizers have been utilized for sample introduction in HPLC-ICP-OES. The solvent load on the plasma can also be decreased by aerosol thermostating, increasing the incident RF power, application of a condenser, or by use of a micro-HPLC column.

Other major sample introduction methods, such as thermospray, DSI, and laser ablation, have also been used for ICP-OES.[16,17,32–36] Each of these has its advantages and disadvantages. For example, laser ablation can be used to vaporize any solid samples into the ICP, but generally it has poor reproducibility and high cost.

4 TORCH CONFIGURATION

The atomic emission from the plasma, as shown in Figure 1, is sampled for spectrometric measurements. Two configurations may be employed for observing emission from the ICP. One is referred to as a *radial* or *side-on viewing of the plasma*, and the other is known as an *axial* or *end-on viewing of the plasma* (Figure 1). A third viewing mode is the combination of these two basic modes, and is known as *dual view*. These are all commercially available, and each of them has advantages and disadvantages.

4.1 Radial View

The radial view is the classical operation mode for ICP-OES. With radial viewing, the plasma is operated in a vertical orientation, and the analytical zone is observed from the side of the plasma. This mimics the classical viewing orientation for atomic emission from a flame. Radial viewing constrains the observation volume to the distance through the diameter of the plasma. Since this pathlength is limited, sensitivity is also limited, but

so are the effects of potential spectral and background interferences.

4.2 Axial View

With the axial view, the plasma is rotated to a horizontal position and the emission signal is observed from the end of the plasma. The axial view provides better LODs than radial view.[37] This may be attributed to the longer viewing path available down the axis of the plasma. Thus, a better sensitivity and a 5- to 10-fold improvement in the LOD can be achieved. The disadvantages of the axial view include the increased potential for spectral interference and matrix-induced interferences. Moreover, self-absorption effects can be quite severe because the observations are made through the much cooler tail plume of the plasma. These effects can be significantly reduced by use of a shear gas,[38] or an appropriate optical interface, which displace the tail plume from the optical path, thus reducing the self-absorption (Figure 1). Spectral interferences may be either corrected or minimized by improving spectral resolution, using an alternate analytical line with less or no interference, or by applying an interelement correction (IEC) factor.[1]

4.3 Dual View

In cases of very complicated sample matrices having a wide range of elemental concentrations, the axial view may be inappropriate. Several commercial instruments combine the axial view and radial view configurations into a single unit, known as *dual view*. This dual view system allows the user to optimize the appropriate configuration for the type of sample without the expense of two separate ICP-OES systems.[1]

5 DETECTION OF EMISSION

5.1 Gratings

ICP-OES is characterized by remarkably rich spectra. For example, the 70 elements most commonly determined by the technique give rise to at least 70 000 total emission lines in the 200–600-nm wavelength range. A consequence of this high density of spectral information is the need for high resolving power. The low-resolution dispersive systems typically employed with atomic absorption spectrometers will not suffice. Spectral interferences occur in this case if only a small number of elements are present at moderate concentrations in the sample. Much higher resolution is desirable in ICP-OES, with spectral bandpass ($\Delta\lambda_s$) 0.01 nm or lower, if possible. Traditionally, this degree of resolution has been accomplished using plane grating monochromators with large focal lengths ($f = 0.5$ m or more).

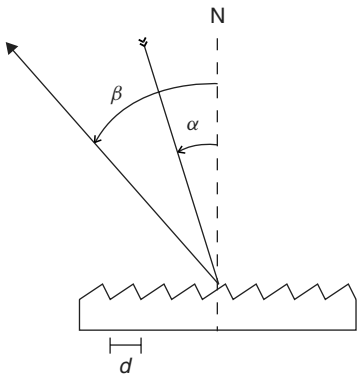

Figure 4 Diagram of the plane-ruled grating. N is the normal to the grating surface, α is the angle of incidence measured with respect to N, and β is the angle of diffraction measured with respect to N, and d is the width of a single groove.

Figure 4 depicts a plane-ruled grating. The normal to the grating surface (N) is shown as a dashed line. A light ray incident to the grating approaches at angle α measured with respect to N. The diffracted ray leaves the grating surface at angle β. Parallel rays striking the grating on different adjacent groove facets travel a different distance before reaching a common position beyond the grating. If the difference in distance traveled is a multiple of the wavelength of the light incident upon the grating surface, then the rays undergo constructive interference. Otherwise, destructive interference occurs. Relatively simple geometrical considerations result in the grating formula shown in Equation (3):

$$d(\sin\alpha + \sin\beta) = m\lambda \tag{3}$$

This equation shows the relationship between α, β, the groove spacing (d), the wavelength of light (λ), and the order of diffraction (m). The order of diffraction may take any integer value including zero. At zero order, all wavelengths undergo constructive interference at the same diffraction angle. In the first order, one particular wavelength undergoes constructive interference at the angle β that corresponds to the specular reflection angle for rays incident at angle α. This wavelength is called the *blaze wavelength* for the grating, and it is determined by the angle at which the grooves are cut with respect to the surface of the grating. A grating is most efficient at its blaze wavelength. Typically, a grating may have an efficiency as high as 70% at its blaze wavelength, so the intensity measured at the blaze wavelength is 70% of the intensity that would be measured at the specular reflectance angle for a polished mirror of the same coating and material as the grating.

Normally, when a plane grating is employed, the angle of incidence is nearly 0, so $\sin\alpha$ approaches 0. In this case, the grating formula may be further reduced, as shown in Equation (4):

$$\sin\beta = \frac{m\lambda}{d} \tag{4}$$

The angular dispersion of the grating ($d\beta/d\lambda$) may be found, as shown in Equations (5) and (6), by taking the derivative of both sides of the above equation with respect to λ:

$$(\cos\beta)\frac{d\beta}{d\lambda} = \frac{m}{d} \tag{5}$$

$$\frac{d\beta}{d\lambda} = \frac{m}{d\cos\beta} \tag{6}$$

The angular dispersion therefore increases with larger order, smaller groove spacing, or larger β. Angular dispersion may be converted to linear dispersion along the exit focal plane of the monochromator by simply multiplying by the focal length (f) of the monochromator. The spectral bandpass ($\Delta\lambda_s$) in wavelength units is then determined by dividing the slit width of the monochromator by the linear dispersion. So $\Delta\lambda_s$ gets smaller (higher resolution) for larger f, larger order, larger β, smaller d, and smaller slit width. A typical plane grating is operated in the first order. The groove density of the grating might be as high as 3600 grooves/mm, so the groove spacing (d) might be as small as 0.0003 mm. Therefore, assuming β is 45° and the slit width is 25 μm, a monochromator with a focal length of 500 mm provides a spectral bandpass of 0.01 nm. Such a system effectively isolates most ICP emission lines. Simultaneous multielement determinations are not performed effectively with such a system, however. The focal plane for this type of monochromator is 5-cm long at best, so the entire spectral window is approximately 20 nm in width. Unless the analytes of interest exhibit emission lines within 20 nm of one another, the wavelength must be scanned to detect multiple elements. So, two other optical approaches have become more popular for simultaneous determinations: the concave grating placed on a Rowland circle, and the echelle grating coupled with a prism order-sorting device.

Shortly after designing his grating ruling engine in 1881, Rowland first conceived the idea of ruling gratings on a spherical mirror of speculum metal.[39] The most important property of such a concave grating was also observed by Rowland. If the source of light and the grating are placed on the circumference of a circle, and the circle has a diameter equal to the radius of curvature of the grating, then the spectrum is always brought to a focus on the circle. Hence, the focal "plane" is curved,

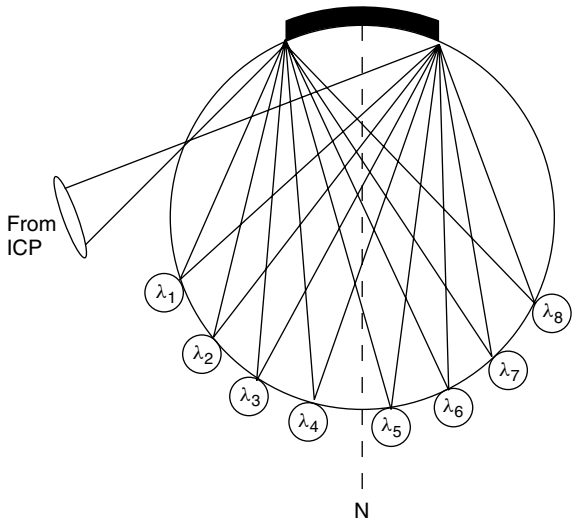

Figure 5 The Paschen–Runge mounting of a concave grating on a Rowland circle. The grating, entrance slit, and focused images of the diffracted wavelengths are all positioned on the perimeter of the circle. N is the grating normal.

and of considerable length (Figure 5). In this case, one entrance slit is placed on the circle for introduction of the source radiation, and multiple exit slits may be placed around the circle at the analytical wavelengths of interest. Hence, Rowland's circle is ideally suited for multielement ICP emission spectrometry. A further advantage of the Rowland geometry is the elimination of the need for any collimating or focusing lenses or mirrors. A Rowland circle spectrometer with the same groove density, slit width, and focal length as the plane grating system described above, provides similar spectral bandpass but with a much larger spectral window.

The echelle grating is a coarsely ruled grating, typically having a groove density of 70 grooves/mm, so $d = 0.014$ mm. The increase in spectral bandpass due to the increase in d is overcome by operating the grating in higher orders ($m = 25$–125) and by using steeper angles of diffraction ($\beta > 45°$). Figure 6 demonstrates how the steeper sides of the groove facets are used with the echelle grating. If the steep sides of the grooves are blazed such that specular reflectance occurs when $\alpha = 60°$ and $\beta = 50°$, then each wavelength exhibits a peak in grating efficiency at a particular order as determined by the grating formula. For example, for the 70 grooves/mm grating described above, the order of maximum efficiency (m_{max}) occurs as shown in Table 4. The efficiency of the grating for a given wavelength at its optimum order can be as high as 65%. This level of efficiency is typically attained across the free spectral range for a given order. The free spectral range ($\Delta\lambda_f$) is defined as the *range of wavelengths* over which no overlap from adjacent orders

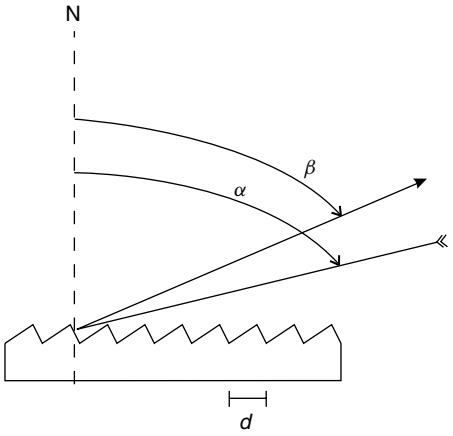

Figure 6 Diagram of the echelle grating. N is the normal to the grating surface, α is the angle of incidence measured with respect to N, and β is the angle of diffraction measured with respect to N. The light is incident to the steeper sides of the groove facets and approaches the grating at nearly grazing angles (62°); d is the width of a single groove.

occurs, and is given by Equation (7):

$$\Delta\lambda_f = \frac{\lambda}{m+1} \tag{7}$$

As indicated in Table 4, $\Delta\lambda_f$ is very small for large values of m. Obviously, then, severe spectral overlap occurs with an echelle grating. The overlap does not simply involve adjacent orders, but all orders are dispersed in multiple layers along the same focal plane. This overlap is corrected most often with an order-sorting prism. This prism is placed between the echelle grating and the focal plane (Figure 7). The prism is positioned so that it disperses the light in a direction perpendicular to the direction of dispersion of the grating. As a result, the focal plane has wavelength dispersed in the horizontal direction and order sorted in the vertical direction. The free spectral range (and the region of maximum efficiency) has a roughly triangular shape centered horizontally on the plane (Figure 8). An echelle monochromator often provides a spectral

Table 4 The order of maximum efficiency and the free spectral range at that order for an echelle grating having 70 grooves/mm, angle of incidence (α) of 60°, and blaze angle (β) of 50°

λ (nm)	m_{max}	$\Delta\lambda_f$ (nm)
200	117	1.7
250	93	2.7
300	78	3.8
350	67	5.2
400	58	6.7

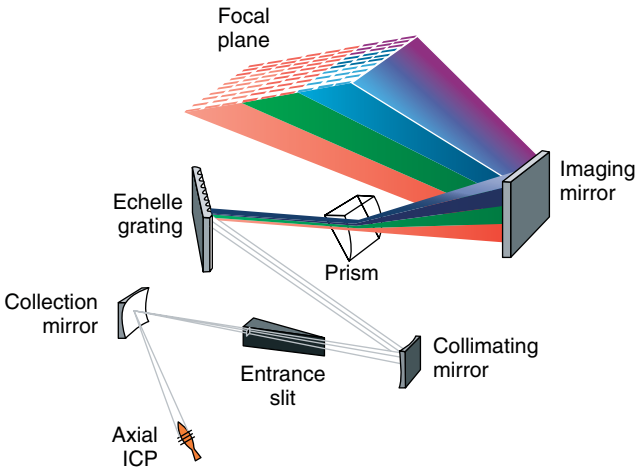

Figure 7 Schematic diagram of an echelle polychromator. (Adapted from the schematic diagram of the Prodigy ICP provided by Teledyne Leeman Labs, Inc., Hudson, NH).

bandpass nearly 10 times smaller than that of a typical grating monochromator with a similar focal length. In addition, the echelle system provides high efficiency at many wavelengths rather than a single blaze wavelength. Finally, both the high efficiency and superior resolution are available over a very broad spectral window. As a result, the echelle configuration is the predominant optical arrangement in modern ICP-OES instruments.[40]

5.2 The Photomultiplier Tube

Figure 9 is a schematic representation of the PMT. Like its predecessor, the vacuum phototube, the heart of the PMT consists of two electrodes sealed in a fused-silica envelope. The cathode has a relatively large surface area, usually in the shape of a vertical, hollow "half cylinder". The cathode is made from a photoemissive material such as an alkali metal oxide. The anode is simply an electron collection wire or grid. Unlike the phototube, however, the PMT has up to 14 secondary emission dynodes placed between the cathode and the anode. Typically, the anode is fixed to the ground potential and the dynodes are at potentials that are successively more negative, by about 100 V/dynode. The potential of the cathode is typically −1000 V.

A photon generated in the ICP and passing through the wavelength selection device may pass through the fused-silica envelope of the PMT, through a baffle-type grill, and strike the photocathode. If the energy of the photon is higher than the work function of the photocathode material, then an electron may be ejected from the cathode. The fraction of photons with energy greater than the work function that actually produce a photoelectron is called the *quantum efficiency of the*

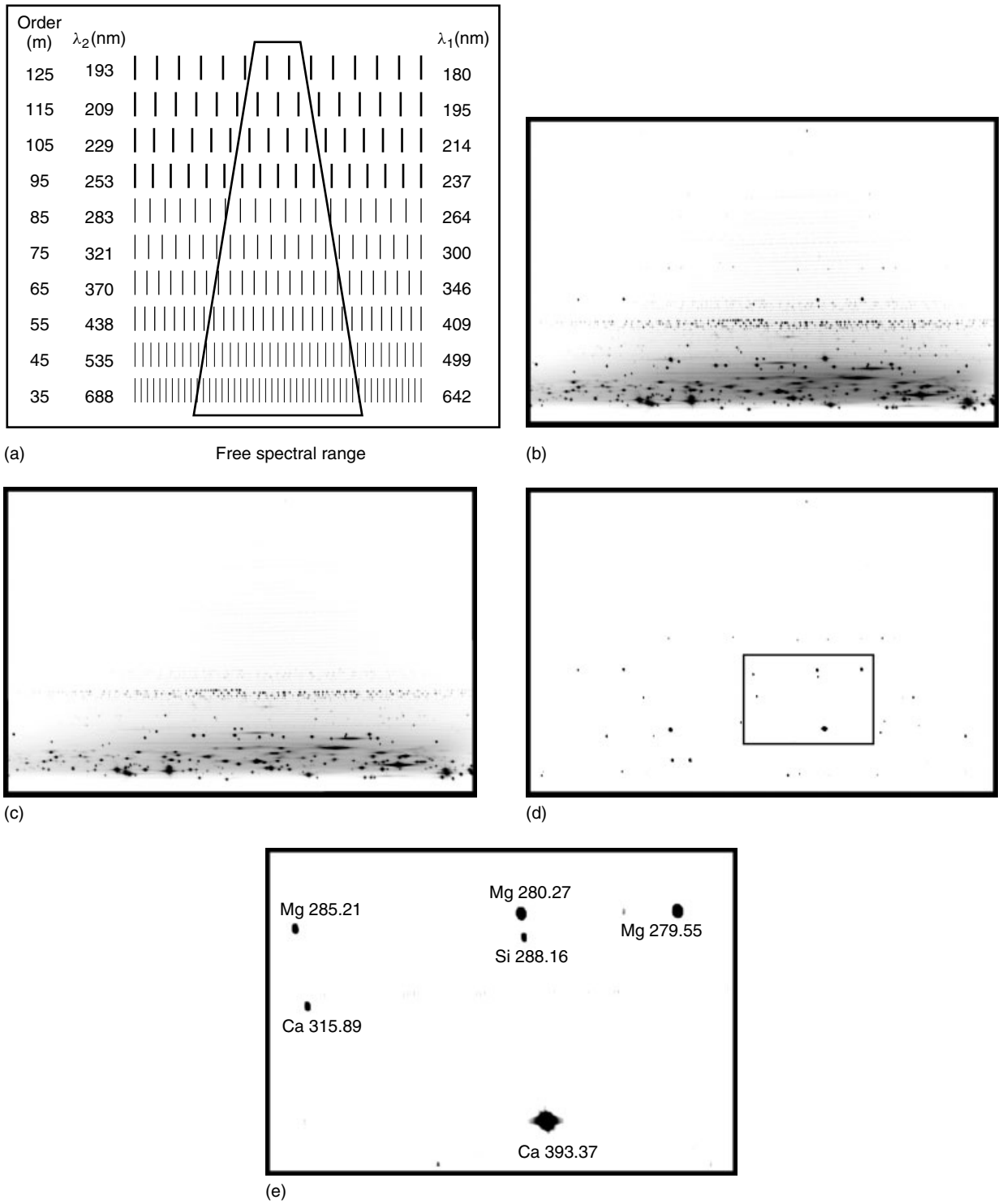

Figure 8 Diagram representing the two-dimensional focal plane provided by the echelle polychromator (a). The triangle-shaped free spectral range is the region of highest grating efficiency. The distance between adjacent vertical lines is 1 nm, and the beginning (λ_1) and ending (λ_2) wavelengths depicted by the lines are listed for each order. Only 1 out of every 10 orders is shown for clarity. An actual negative image collected during the aspiration of tap water (b) shows atomic emission lines as dark spots against a white background. Close inspection reveals the different orders as faint horizontal lines across the image, with the roughly triangular shape of the free spectral range. An image for distilled, deionized water (c) may be subtracted from the tap water image resulting in a background-subtracted image (d). The region of interest may be enlarged so that emission lines from Ca, Mg, and Si in the tap water may be identified (e).

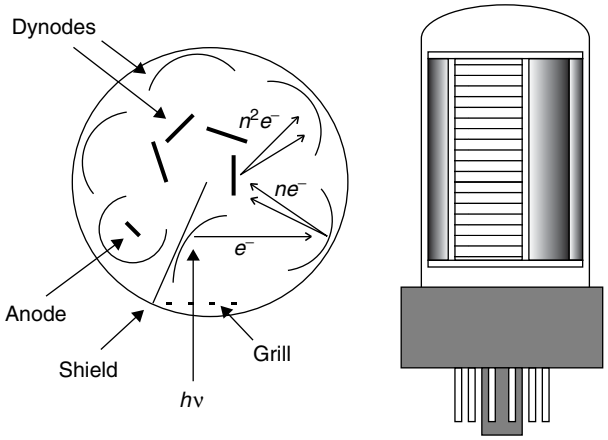

Figure 9 Cross-sectional and front views of a side-on PMT. In the cross section, two types of dynodes are depicted: curved (outer) and flat (inner). The outside surfaces of the curved dynodes are seen in the front view.

photocathode. The quantum efficiency may be as high as 0.5, and it depends upon the photocathode material and wavelength. A plot of quantum efficiency versus wavelength is called the *spectral response curve* for the PMT. The spectral response curve is usually supplied by the manufacturer.

Once an electron has been ejected by the photocathode, it is accelerated toward the first dynode. Upon impact, the first dynode releases x secondary electrons (where x is typically between 2 and 5). This process continues at each dynode, with each electron impact imparting x new secondary electrons. So, by the time that the pulse generated by a single photon reaches the anode it has been greatly multiplied, and hence the name *multiplier phototube* or *PMT*. The gain, G, for a PMT can be defined as shown in Equation (8):

$$G = x^n \tag{8}$$

where n is the number of dynodes. The gain depends upon the voltage across the PMT and it may be as high as 10^8. One of the best features of the PMT is that the gain is acquired with almost no increase in noise. Thus, the PMT is ideally suited for the detection of small analytical signals against a relatively dark background, as is the case near the detection limit in high-resolution atomic emission spectrometry.

Usually the *entrance aperture*, as defined by the grill on the PMT, is very large compared to a single resolution element at the focal plane of a monochromator. This disparity is corrected by placing a mechanical slit on the focal plane in front of the PMT. The width of this slit then defines the range of wavelengths ($\Delta\lambda_s$) that are allowed to strike the PMT. Sequential detection of multiple elements can be accomplished by changing the grating angle in a conventional monochromator, thus scanning the wavelengths that are detected. A second approach is to move the PMT rapidly along the focal curve of a Rowland circle spectrometer. Such a spectrometer may have several prealigned exit slits along the focal curve, each corresponding to a particular element (Figure 5). A still more efficient method for sequential multielement determinations involves the echelle polychromator (Figure 7). A mask with many slits, each corresponding to a different element, may be placed upon the two-dimensional focal plane. The PMT is held in a mechanical arm that quickly positions the detector at the appropriate $x-y$ coordinates for a given element. This design allows very fast sequential determinations by ''wavelength hopping'' or direct reading rather than scanning linearly through all wavelengths to reach a select few. One commercial instrument using this design is called a *direct reading echelle* (*DRE*) *ICP instrument*.

Simultaneous multielement determinations may be performed with multiple PMTs, but such designs quickly become limited by the size of the PMTs and the geometry of the polychromator. Multiple PMTs may be positioned either along the focal curve of a Rowland circle spectrometer or along the exit plane of an echelle polychromator. Solid-state detectors, with their relatively small size and their intrinsic multielement nature, are usually more effective for simultaneous determinations.

5.3 Array Detectors

Charge transfer devices (CTDs) include a broad range of solid-state silicon-based array detectors.[41] They include the charge injection device (CID) and the charge-coupled device (CCD). The CCD has found extensive use in nonspectroscopic devices such as video cameras, bar code scanners, and photocopiers. With the CTDs, photons falling on a silicon substrate produce electron–hole pairs. The positive electron–holes migrate freely through the p-type silicon semiconductor material, while the electrons are collected and stored temporarily by an array of metal oxide semiconductor (MOS) capacitors (Figure 10). Each

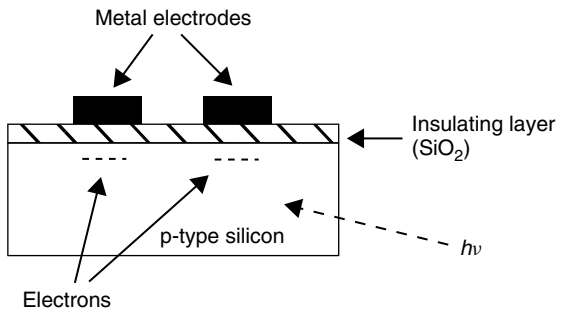

Figure 10 Cross-sectional diagram of two adjacent pixels in a CCD detector.

MOS capacitor is composed of a small metal electrode and a thin layer of insulating SiO_2 material on top of the p-type silicon substrate. A positive potential is applied to the metal electrode, so the electrons generated in a given region are trapped just below the insulating layer. Each MOS capacitor (or pixel) has a width in the 5- to 50-μm range and a height that may be as large as 200 μm. A two-dimensional array of pixels is easily prepared by proper placement of the metal electrodes. Such arrays may vary in size, even beyond 4000 pixels on an edge. The CCD differs from the CID mainly in the readout scheme. The CCD is read out in a sequential charge shifting manner toward the output amplifier. The CID, on the other hand, may be read out in a nondestructive manner by shifting charge between adjacent electrodes, and then shifting it back again. The CID thus benefits from quick random access, even during long integration periods.

Spectroscopic applications of CTDs have been hampered by the physical mismatch between the relatively small surface area of the detector and the large, sometimes two-dimensional, focal plane associated with polychromators. This mismatch may be overcome, however, and one commercial ICP spectrometer employs a CID detector having more than 250 000 pixels positioned upon an echelle focal plane.

Alternative approaches have been successful with the CCD detector. In one case, a group of several CCD arrays are arranged around a circular optical system (CIROS) based upon a Rowland circle design. Rather than monitoring discrete wavelengths, as is the case with the multiple PMT Rowland circle systems, the CIROS system provides total wavelength coverage from 120 to 800 nm, with resolution on the order of 0.009 nm.

A second multiple CCD array detector has been available commercially. This detector, called a *segmented array charge-coupled device detector* (*SCD*), employs over 200 small subarrays of 20–80 pixels each.[41] The subarrays are positioned along the two-dimensional focal plane of an echelle polychromator. The position of each subarray corresponds to one of the 236 most prominent ICP emission lines of the most commonly determined 70 ICP elements. This design allows for discrete wavelength determinations as seen with the multiple PMT designs, but it also provides additional spectral information around the vicinity of each emission line without exhaustively recording data at all wavelengths.

Another approach to correcting the mismatch between conventional imaging CTD arrays and the focal plane of an echelle spectrometer is to specifically design a CTD array to exactly match the spectrometer image.[3,42] This process, called *image-mapping the detector* or *large format programmable detection*, is available in a couple of commercial products. One device called the *VistaChip* consists of a series of 70 diagonalized linear arrays (DLAs) of pixels that are designed to exactly match the individual diffraction orders present in the focal plane of the echelle spectrometer. The 70 DLAs correspond to orders 19–88, and the length of an individual DLA is set to match the free spectral range for the corresponding order. In this fashion, continuous wavelength coverage is provided across the range 167–363 nm, and selected coverage is provided in the range 363–784 nm where the diffraction orders are wider than the selected width of the detector (although no atomic emission lines of relevance miss the detector). The overall dimensions of the VistaChip are 15×19 mm, and a total of 70 908 pixels are packed inside the 70 DLAs. Other commercial CTD detectors are now as large as 27×27 mm, and any wavelength point on the chip may be accessed, from 160 to 1100 nm.[3] As a result of this broad spectral coverage, the combination of echelle polychromator and CTD has become the system of choice for ICP-OES.[40]

5.4 Simultaneous versus Sequential Detection

In the final analysis, the detection system most appropriate for an ICP emission system depends upon the application. In cases where only one or two elements are determined routinely, the traditional scanning sequential detection system may be sufficient. The high sensitivity provided by a PMT coupled with the flexibility of interrogating any wavelength region may outweigh the need for rapid determinations. On the other hand, if the application may vary between the determination of a few elements to the determination of many, the reasonable cost and high sensitivity of the PMT-based direct reading systems may be attractive. Finally, if a large suite of elements must be determined on a routine basis, one of the array-based detection systems might be most suitable.

6 ANALYTICAL PERFORMANCE

6.1 Analytical Wavelength

The ultraviolet and visible regions (160–800 nm) of the electromagnetic spectrum are most commonly used for analytical atomic spectrometry. In ICP-OES, the number of elements that can be determined is related to the wavelength window that can be covered by the optical system. With current echelle-CTD combinations, this entire spectral window may be surveyed simultaneously (Figure 8). For the tap water sample image in Figure 8(b), note the relative crowding of the emission lines in the visible region (in the bottom third of the image). In this negative image, individual emission lines due to elements in the sample appear as dark spots against a white background. Many of these spots may also arise

from Ar emission lines, and hence are clustered in the visible region. In addition, the continuum emission from the plasma is relatively intense in the visible region, as evidenced by the gray shading across this region. These emission signals arising from the background may be effectively corrected by subtracting a blank: the image of a distilled-deionized water sample (Figure 8c). The resulting corrected spectrum (Figure 8d) shows only those lines due to the elements present in the tap water but not in the blank. Upon expansion of the region of interest marked by a rectangle in Figure 8(d), one may identify the metals of interest in the sample (Figure 8e). In this case, multiple lines are observed for both Mg and Ca. As a rule of thumb, in the absence of significant interferences, wavelengths above 500 nm are employed when alkali metals need to be determined, whereas wavelengths below 190 nm or even below 160 nm are employed when elements such as chlorine, bromine, nitrogen, or arsenic must be determined. There are several criteria for selecting the specific analytical lines for analysis. First, the wavelengths must be accessible by both the dispersive system and the detector. Second, the wavelengths must exhibit signal levels appropriate for the concentrations of the respective elements in the sample. Third, the wavelengths selected must be free from spectral interferences. When this is not possible, emission lines whose intensities can be corrected to account for spectral interferences should be chosen. Fourth, if an internal standard scheme is used, it may be preferable to match the analyte ion lines with an internal standard ion line, and analyte neutral atom lines with an internal standard neutral atom line. Most modern instruments offer several wavelength suggestions for each element, and they are capable of simultaneously monitoring multiple lines for a single element to reduce the potential for interference.

6.2 Quantitative Methods

The calibration curve method using aqueous standard solutions is easily employed in ICP-OES. Aqueous stock solutions for single metals or aqueous mixtures containing multiple metals are available from several commercial vendors. In addition, stock solutions may be prepared in the laboratory with little difficulty.[43]

In those cases where the analytical signal is noisy owing to variations such as small changes in solution flow rates, gas flow rates, or plasma position, the internal standard method may be employed. The internal standard element may just as easily be added to each solution by the analyst. The choice of the internal standard element may be critical. The element should not be present in the sample type of interest, should not interfere spectrally with any of the analytes of interest, and in the best case should have physical and chemical properties similar to those of the analyte.[44] The perfect internal standard element experiences the same matrix effects and spectral background as the analyte. In practice, many internal standard elements are suitable, and published analytical procedures may offer a suite of choices for a given analyte. In some cases, a single internal standard could be effective for a large number of different analyte elements. This approach is most successful under robust ICP conditions: high ICP power (~1.5 kW) and low carrier gas flow rate (~0.6 L min^{-1} Ar).[45] The procedure is straightforward. One simply adds the internal standard element with a constant concentration to each standard and sample solution to be analyzed. For example, a Pb calibration curve in the range of 0–100 ng mL^{-1} may be constructed by preparing 5 Pb standard solutions spanning this concentration region, each containing 100 ng mL^{-1} Cu. The Pb emission line at 217.0 nm may be monitored along with a nearby Cu emission line at 216.5 nm (Figure 11). If spectra such as this are collected for each solution, a plot of the raw Pb signals generate the calibration points depicted by the black diamonds in Figure 12. If on the other hand, those same Pb signals are divided by the emission signal for Cu in each spectrum, a better linear fit is observed (circles in Figure 12). When the curves are scaled to the same magnitude, note that the slopes are nearly identical, since, in theory, the internal standard plot is equivalent to the raw plot divided by a constant. In this example, not only is the linear fit improved but the signal-to-noise ratio for the individual measurements is increased by nearly a factor of 10.

In extreme cases, where the sample matrix actually affects the slope of the calibration curve, the standard addition method should be employed. Urine samples have been analyzed by ETV-ICP-OES after simple 1:1 dilution with water by this technique.[46] The urine sample was divided into five separate 1-mL aliquots. Successively, greater amounts of a Cd standard solution were added to

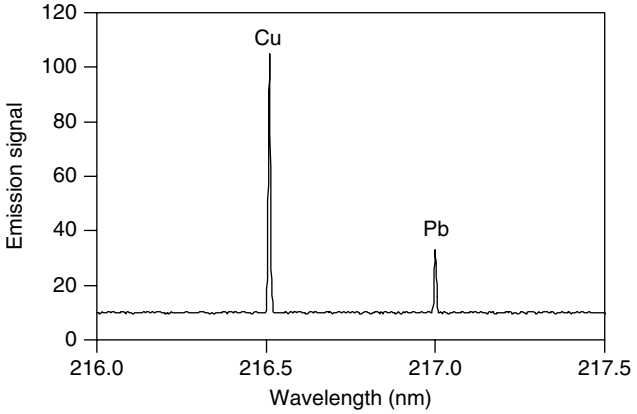

Figure 11 ICP-OES spectrum of a solution containing 60 ng mL^{-1} Pb and 100 ng mL^{-1} Cu.

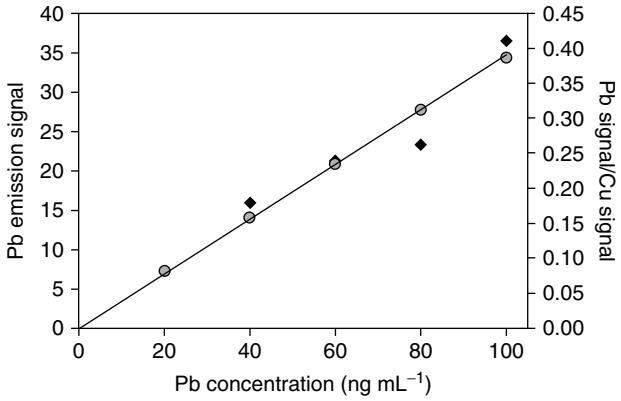

Figure 12 Internal standard correction for a Pb calibration curve. The fit using the raw Pb emission signal (diamonds) is greatly improved when ratioed to the signal measured for a given amount of Cu (100 ng mL^{-1}) added to each sample (circles). The same Pb ICP-OES emission signals were used in each case.

the five aliquots, resulting in a series of solutions having 0, 2.5, 5.0, 7.5, and 10.0 ng mL^{-1} added Cd. Each solution was also spiked with 1 μg mL^{-1} Bi as an internal standard, and the final volume was made up to 2 mL with water. No further sample preparation was employed, and the ETV-ICP-OES signals from Cd (diamonds) and Cd/Bi ratio (circles) are plotted in Figure 13. Note that the internal standard correction is necessary to improve the fit such that an accurate x intercept may be determined. The value for the x intercept is -1.1 ng mL^{-1}, and thus the original urine sample was found to contain 1.1 ± 0.2 ng mL^{-1} Cd. Analysis of the same sample by GFAAS following microwave acid digestion gave a value of 1.3 ± 0.5 ng mL^{-1} Cd.[46]

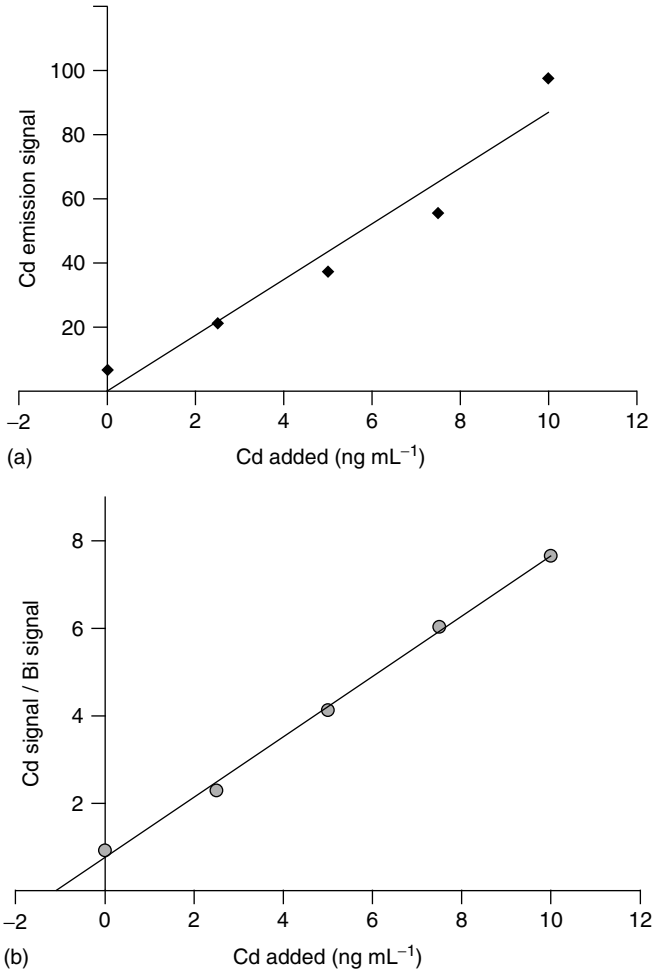

Figure 13 Standard addition curves for the determination of Cd in urine by ETV-ICP-OES. The x axis represents the concentration of Cd added to separate 1 mL aliquots of a urine sample. Each sample was also spiked with 1 μg mL^{-1} Bi. The diamonds (a) represent the raw Cd signal, and the circles (b) include internal standard correction using the ratio of the Cd and Bi emission signals. The urine sample was found to contain 1.1 ng mL^{-1} Cd using the internal standard. Without the internal standard, the amount found (x intercept) was zero.

6.3 Analytical Figures of Merit

For ICP-OES, the analytical figures of merit include the number of elements that can be determined, selectivity, reproducibility, long-term stability, susceptibility to matrix interferences, LOD, and accuracy.[47] The number of elements that can be measured by ICP-OES is often more than 70 out of a total of 92 naturally occurring elements, as listed in Table 5. Routine determination of 70 elements can be accomplished by ICP-OES at concentration levels below 1 mg L^{-1}. As can be seen from Table 5, almost all naturally occurring elements, with the

Table 5 A list of elements that can be determined by ICP-OES

Alkaline and alkaline earth	Rare earth	Transition metal	Others
Li, Na, K, Rb, Cs, Be, Mg, Ca, Sr, Ba	Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Th, U	Sc, V, Ti, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Nb, Zr, Mo, Ru, Th, Pd, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg	B, C, N, Al, Si, P, S, Cl, Ga, Ge, As, Se, Br, In, Sn, Sb, Te, I, Tl, Pb, Bi

exception of hydrogen, oxygen, fluorine, and inert gases, can be determined by ICP-OES. The elements that are not usually determined by ICP-OES fall into three basic categories. The first category includes those elements that occur either as trace contaminants in the argon gas used in the ICP-OES (C from CO_2), constituents of the sample solvent (C, O, H), or as contaminants from the environment or atmosphere (N, for example). The second category encompasses those elements that require high excitation energy, such as the halogens. These elements could be determined with poor LOD, however. The third category is the family of short-lived radioactive elements that are commonly determined by γ-ray spectrometry.[1]

Selectivity is important to minimize the spectral overlap interferences resulting from elements with rich line-emission spectra (tungsten, cobalt, niobium, molybdenum, tantalum, and rare earth elements) and to improve the signal-to-background ratio (SBR).[47] Selectivity is largely decided by the practical resolution of the wavelength dispersive system of the ICP-OES instrument. High selectivity is usually achieved with a sacrifice in sensitivity and the wavelength coverage range. For the best commercial ICP-OES instruments, a resolution of less than 5 pm is possible.[47]

The LODs of ICP-OES are generally in the nanogram per milliliter range. The LOD is usually defined as the *analyte concentration* that produces an analytical signal equivalent to three times the standard deviation observed for 16 measurements of a blank solution.[48] Another definition for the LOD of ICP-OES is related to the SBR of the analyte line at a given concentration, c, and the relative standard deviation (RSD) of the background, RSD_B as shown in Equation (9)[49]:

$$LOD = \frac{3 \times c \times RSD_B}{SBR} \tag{9}$$

The LOD is determined, therefore, by the sensitivity of the measurement and the noise level, or stability, of the ICP-OES instrument. The high degree of stability of an ICP was identified when Greenfield et al.[10] first used an ICP for analytical atomic spectrometry. For the best ICP-OES instruments, a long-term stability of less than 1% RSD has been achieved.[47] The atomic emission signals from the ICP are larger than those from other sources, such as a flame. This occurs because the high-temperature and inert-argon environment of the ICP leads to more efficient atomization, ionization, and excitation. In fact, the temperature of the ICP is so high that the largest signals are usually from the ionic lines. There are many other factors that may influence LOD, such as nebulizer type, view mode, and sample matrix.[50,51] Table 6 shows representative ICP-OES LODs obtained for 27 different elements.[3] LODs

Table 6 LOD (μg L^{-1}) observed for ICP-OES using a concentric nebulizer with different viewing modes[3]

Element	Wavelength (nm)	Radial view	Axial view
Ag	328.07	1	0.4
Al	396.15	1	0.2
As	193.70	12	3
B	249.77	2	0.1
Ba	455.40	0.1	0.01
Be	313.04	0.06	0.04
Ca	317.93	2	0.6
Cd	226.50	0.7	0.2
Co	228.62	2	0.3
Cr	267.72	1	0.1
Cu	324.75	0.7	0.2
Fe	259.94	0.7	0.1
K	766.49	10	0.8
Mg	279.55	7	0.01
Mn	267.61	0.3	0.02
Mo	202.03	1	1
Na	589.00	1	0.08
Ni	231.60	2	0.4
P	177.43	30	10
Pb	220.35	5	1
Sb	206.83	9	3
Se	196.03	20	4
Si	288.16	10	0.8
Sr	407.77	0.05	0.03
Tl	190.86	20	2
V	292.40	0.9	0.07
Zn	213.86	0.7	0.09

using an axially viewed plasma are typically better than those observed by radial viewing by a factor of 5–10,[1] as shown in Table 6. Table 7 shows LODs for ETV-ICP-OES in comparison with those of GFAAS. LODs for many elements by ETV-ICP-OES are better than or equivalent to those achieved by GFAAS.[52,53] Notice that some elements that cannot be determined by GFAAS can be measured by ICP-OES.

Often the background equivalent concentration (BEC) is also used to check instrumental performance in ICP-OES. The BEC is defined as the *concentration of a solution* that results in an analyte emission signal equivalent in intensity to that of the background emission signal at the measurement wavelength. The BEC can be used as an indicator of relative sensitivity for an emission line. An unusually high BEC often indicates problems with the efficiency of the sample introduction system.

The LDR of calibration curves for ICP-OES is usually 4–6 orders of magnitude wide, starting from the LOD on the low-concentration side. These LDRs are significantly larger than the 2–3 orders of magnitude observed for competing techniques such as AAS and arc/spark OES. The wide LDRs in ICP-OES translate into simple preparation of calibration curves. Very often,

Table 7 LOD (pg) for ETV-ICP-OES compared with those of GFAAS[2,52,53]

Elements	ETV-ICP-OES	GFAAS
Ag	1	0.5
Al	0.5	4
As	60	20
Au	10	10
Ba	0.3	10
Be	1	1
Bi	200	10
B	4	NA
Ca	150	5
Cd	1	0.3
Co	12	1
Cu	2	1
Er	34	300
Eu	12	10
Fe	10	2
Ga	10	10
Ge	10	20
Hg	4	100
In	20	5
K	1200	2
Li	2	5
Lu	54	4000
Mg	0.1	0.4
Mn	0.3	1
Mo	7	4
Na	400	5
Ni	27	10
P	100	3000
Pb	4	5
Pt	25	50
Rb	2800	5
Re	100	NA
Ru	95	40
Sb	100	20
Sc	17	40
Se	450	20
Si	100	40
Sn	20	20
Sr	5	2
Te	50	10
Ti	6	100
Tl	300	10
U	3	2.4 μg
V	6	20
W	160	NA
Y	25	NA
Yb	18	4
Zn	0.6	1

NA, not available.

a single standard together with a blank solution is enough to produce an accurate calibration. Multiple sample dilutions are seldom needed prior to the analysis.

In general, the accuracy of the ICP-OES technique gets poorer as the analyte concentration approaches the LOD. For semiquantitative analysis (accuracy ±10%), the analyte concentration should be at least five times higher than the LOD. For accurate quantitation (±2%), the concentration should be 100 times greater than the LOD. At this concentration level, the precision is typically better than 1% RSD. This precision is considered sufficient for most trace element determinations. Better precision can be achieved, if necessary, by sacrificing analysis speed or with a more complex instrumental design. The concept of limit of quantitation (LOQ) has been defined as *a concentration* for which the precision, expressed as RSD, would be below a given threshold, for instance, 10 or 5%.[54] A 5%-based LOQ normally occurs at concentrations approximately 10 times greater than the LOD (based upon three standard deviations).

6.4 Interferences

Among all commonly used analytical atomic spectrometry techniques, ICP-OES is probably the one with the fewest number of interferences. The argon plasma is inert when compared to the chemical reactivity of a flame. Also, the high temperature of the plasma helps to reduce chemical interferences. The temperature is high enough to break down most species into atoms or ions for excitation and subsequent emission. In contrast, in a low-temperature flame, chemical interferences can be a severe problem. For example, a small amount of aluminum will interfere with the determination of calcium in flame AAS, but even at aluminum concentration 100 times higher than this, interference is not observed in ICP-OES. However, chemical interferences do exist in the ICP. Sometimes higher RF power and/or lower inner argon flow rates are used to reduce these interferences. One particular type of chemical interference is the so-called easily ionized element (EIE) effect. The EIEs are those elements that have low ionization potentials, such as alkaline elements. High concentrations of EIEs can suppress or enhance emission signals, depending upon the analyte species. One way to reduce the EIE effect is to dilute the sample solution to the point that the EIE effect is not measurable. Sometimes, higher RF power or mathematical correction may be used to compensate for EIE interference. Instrumental conditions such as slit width, viewing height, and viewing volume can be chosen to minimize interference and to optimize emission signal collection in either axial or radial configuration.[55]

The most common interference problem in ICP-OES is spectral interference (also referred to as *background interference*). Ironically, this type of interference arises because of the multielement nature of the plasma. Since the ICP is capable of exciting almost any element that is introduced into the plasma, spectra are likely to be rich, especially for highly complex and concentrated samples. The solution to the spectral interference

problem, as discussed previously, is the use of high-resolution spectrometers. Some spectral overlap may even exist with the best commercial system. In these cases, advanced background correction techniques are employed or a different analytical wavelength for the element(s) of interest is chosen. Spectral interferences can be categorized into four categories: simple background shift, sloping background shift, direct spectral overlap, and complex background shift. The simple background shift is defined as *a shift in background intensity* that is essentially constant over a given wavelength range on either side of the analytical line. The background may shift up or down. There are two approaches to deal with this background problem. The first is to select a different analytical line at a wavelength with no background interference. The second is to correct for the background by measuring it somewhere near, but not falling on, the profile of the analytical line of the analyte element. Two background correction points, one on each side of the profile of the analytical line, are used to correct for the sloping background shift. This means that the average signal measured at the two points is subtracted from the total signal measured at the analytical line. In the worst case, direct spectral overlap occurs. This can be corrected if the magnitude of the interference is known as a *function of the concentration* of the interfering element. A correction factor can be calculated and used to correct the signal measured at the analyte wavelength. This can be best achieved by making simultaneous measurements of both the signal at the analyte wavelength and at a different wavelength for the interfering element. This method is referred to as the *concentration ratio method* or *inter-element correction method*. The basic requirement is that the concentration of the interfering element can be accurately measured at another wavelength. With advanced detector systems, all spectral lines are present, so IEC is possible. A complex background shift is a shift in a background intensity that varies significantly on both sides of the analytical line. This is usually caused by the occurrence of a number of intense, closely spaced emission lines nearby, and perhaps directly overlapping the analyte wavelength. In this case, a different analytical wavelength should be chosen if possible.[1]

ACKNOWLEDGMENTS

This work was funded in part by Teledyne Leeman Labs, Hudson, N.H. Teledyne Leeman Labs also provided the photograph in Figure 1, and the schematics in Figures 3 and 7. The images in Figure 8 were acquired with a Prodigy ICP-OES system provided by Teledyne Leeman Labs as well.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
BEC	Background Equivalent Concentration
CCD	Charge-coupled Device
CID	Charge Injection Device
CIROS	Circular Optical System
CTD	Charge Transfer Device
DCP	Direct Current Plasma
DLA	Diagonalized Linear Array
DRE	Direct Reading Echelle
DSI	Direct Sample Insertion
EIE	Easily Ionized Element
ETV	Electrothermal Vaporization
FAAS	Flame Atomic Absorption Spectrometry
GC	Gas Chromatography
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
HG	Hydride Generation
HGN	Hildebrand Grid Nebulizer
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICP-OES	Inductively Coupled Plasma-optical Emission Spectrometry
IEC	Interelement Correction
IR	Induction Region
ITV	In-torch Vaporization
LDR	Linear Dynamic Range
LIP	Laser-induced Plasma
LOD	Limit of Detection
LOQ	Limit of Quantitation
MCN	Microconcentric Nebulizer
MIP	Microwave-induced Plasma
MOS	Metal Oxide Semiconductor
OES	Optical Emission Spectrometry
PMT	Photomultiplier Tube
RF	Radiofrequency
RSD	Relative Standard Deviation
SBR	Signal-to-background Ratio
SCD	Segmented Array Charge-coupled Device Detector
TDS	Total Dissolved Solids
USN	Ultrasonic Nebulizer

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Laser-induced Breakdown Spectroscopy, Elemental Analysis

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Atomic Fluorescence in Environmental Analysis • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Heavy Metals Analysis in Seawater and Brines • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure • Surface and Dermal Monitoring

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis • Metal Analysis, Sampling and Sample Preparation in

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction

FURTHER READING

V. Karanassios, K.P. Bateman, 'Electrically Heated Wire-loop, In-torch Vaporization Sample Introduction System for Inductively Coupled Plasma Atomic Emission Spectrometry with Photodiode Array Detection', *Spectrochim. Acta*, **49B**(9), 847–865 (1994).

REFERENCES

1. C.B. Boss, K.J. Fredeen, *Concept, Instrumentation and Techniques in Inductively Coupled Plasma Optical Emission Spectrometry*, 2nd edition, Perkin-Elmer, Norwalk, 1997.
2. J.M. Carey, J.A. Caruso, 'Electrothermal Vaporization for Sample Introduction in Plasma Source Spectrometry', *Crit. Rev. Anal. Chem.*, **23**(5), 397–439 (1992).
3. Teledyne Leeman Labs, Inc., Hudson, 2007 www.leemanlabs.com.
4. M.D.A. Korn, D.S.S. dos Santos, B. Welz, M.G.R. Vale, A.P. Teixeira, D.D. Lima, S.L.C. Ferreira, 'Atomic Spectrometric Methods for the Determination of Metals and Metalloids in Automotive Fuels – A Review', *Talanta*, **73**(1), 1–11 (2007).
5. A.C. Davis, P. Wu, X.F. Zhang, X.D. Hou, B.T. Jones, 'Determination of Cadmium in Biological Samples', *Appl. Spectrosc. Rev.*, **41**(1), 35–75 (2006).
6. K.L. Linge, 'Recent Developments in Trace Element Analysis by ICP-AES and ICP-MS with Particular Reference to Geological and Environmental Sample', *Geostand. Geoanal. Res.*, **29**(1), 7–22 (2005).
7. J. Meija, S. Mounicou, J.A. Caruso, 'Plasma Spectrometry for Elemental Speciation and Characterization in Beverages', *J. Assoc. Off. Anal. Chem. Int.*, **87**(1), 205–224 (2004).
8. M. Aceto, O. Abollino, M.C. Bruzzoniti, E. Mentasti, C. Sarzanin, M. Malandrino, 'Determination of Metals in Wine with Atomic Spectroscopy (flame-AAS, GF-AAS and ICP-AES); A Review', *Food Addit. Contam.*, **19**(2), 126–133 (2002).
9. R.H. Wendt, V.A. Fassel, 'Induction-coupled Plasma Spectrometric Excitation Source', *Anal. Chem.*, **37**(7), 920–922 (1965).
10. S. Greenfield, I.L. Jones, C.T. Berry, 'High-pressure Plasma as Spectroscopic Emission Sources', *Analyst*, **89**(11), 713–720 (1964).
11. V.A. Fassel, 'Analytical Inductively Coupled Plasma Spectroscopies – Past, Present, and Future', *Fresenius Z. Anal. Chem.*, **324**(6), 511–518 (1986).
12. S.R. Koirtyohann, J.S. Jones, C.P. Jester, D.A. Yates, 'Use of Spatial Emission Profiles and a Nomenclature System as Acids in Interpreting Matrix Effects in the Low-power

Argon Inductively Coupled Plasma', *Spectrochim. Acta*, **36B**(1), 49–59 (1981).

13. T. Nakahara, 'Hydride Generation Techniques and Their Applications in Inductively Coupled Plasma–Atomic Emission Spectrometry', *Spectrochim. Acta Rev.*, **14**(1/2), 95–109 (1991).
14. H. Bin, S. Li, G.Q. Xiang, M. He, Z.C. Jiang, 'Recent Progress in Electrothermal Vaporization – Inductively Couplet Plasma Atomic Emission Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc. Rev.*, **42**(2), 203–234 (2007).
15. T. Kantor, 'Electrothermal Vaporization and Laser Ablation Sample Introduction for Flame and Plasma Spectrometric Analysis of Solid and Solution Samples', *Spectrochim. Acta*, **56B**(9), 1523–1563 (2001).
16. R.E. Russo, 'Laser Ablation', *Appl. Spectrosc.*, **49**(9), 14A–28A (1995).
17. R. Sing, 'Direct Sample Insertion for Inductively Coupled Plasma Spectrometry', *Spectrochim. Acta*, **54B**(3–4), 411–441 (1999).
18. J.L. Todoli, J.M. Mermet, 'Sample Introduction Systems for the Analysis of Liquid Microsamples by ICP-AES and ICP-MS', *Spectrochim. Acta*, **61B**(3), 239–283 (2006).
19. J.L. Todoli, J.M. Mermet, 'Elemental Analysis of Liquid Microsamples Through Inductively Coupled Plasma Spectyrochemistry', *TRAC – Trends Anal. Chem.*, **24**(2), 107–116 (2005).
20. R.S. Babington, 'It's Superspray', *Pop. Sci.*, **202**, 102–104 (1973).
21. M.C. Santos, J.A. Nobrega, 'Slurry Nebulization in Plasmas for Analysis of Inorganic Materials', *Appl. Spectrosc. Rev.*, **41**(4), 427–448 (2006).
22. J.G. Williams, 'Instrument Options', in *Inductively Coupled Plasma Mass Spectrometry*, eds. K.E. Jarvis, A.L. Gray, R.S. Houk, Blackie, New York, 58–80, 1992.
23. J. Borkowska-Burnecka, A. Lesniewicz, W. Zymicki, 'Comparison of Pneumatic and Ultrasonic Nebulizations in Inductively Coupled Plasma Atomic Emission Spectrometry – Matrix Effects and Plasma Parameters', *Spectrochim. Acta*, **61**, 579–587 (2006).
24. K. Levine, K.A. Wagner, B.T. Jones, 'Low-cost, Modular Electrothermal Vaporization System for Inductively Coupled Plasma Atomic Emission Spectrometry', *Appl. Spectrosc.*, **52**(9), 1165–1171 (1998).
25. G.M. Allen, D.M. Coleman, 'Segregated Sampling and Excitation with a Dual Inductively Coupled Plasma', *Anal. Chem.*, **56**(14), 2981–2983 (1984).
26. R. Lobinski, F.C. Adams, 'Speciation Analysis by Gas Chromatography with Plasma Source Spectrometric Detection', *Spectrochim. Acta*, **52B**(13), 1865–1903 (1997).
27. D.J. Butcher, 'Environmental Applications of Arsenic Speciation Using Atomic Spectrometry Detection', *Appl. Spectrosc. Rev.*, **42**(1), 1–22 (2007).
28. S.J. Hill, M.J. Bloxham, P.J. Worsfold, 'Chromatography Coupled with Inductively Coupled Plasma Atomic Emission Spectrometry and Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**(6), 499–515 (1993).
29. J.C.A. Wuilloud, R.G. Wuilloud, A.P. Vonderheide, J.A. Caruso, 'Gas Chromatography/Plasma Spectrometry – an Important Analytical Tool for Elemental Speciation Studies', *Spectrochim. Acta*, **59B**(6), 755–792 (2004).
30. H.L. Peters, B.T. Jones, 'Determination of Non-Metals by High Performance Liquid Chromatography with Inductively Coupled Plasma Detection', *Appl. Spectrosc. Rev.*, **38**(1), 71–99 (2003).
31. T. Kato, T. Uehiro, A. Yasuhara, M. Morita, 'Determination of Methylmercury Species by Capillary Column Gas Chromatography with Axially Viewed Inductively Coupled Plasma Atomic Emission Spectrometric Detection', *J. Anal. At. Spectrom.*, **7**(2), 15–18 (1992).
32. T.S. Conver, J. Yang, J.A. Koropchak, 'New Developments in Thermospray Sample Introduction for Atomic Spectrometry', *Spectrochim. Acta*, **52B**(8), 1087–1104 (1997).
33. V. Karanassios, G. Holick, 'Direct Sample Insertion Devices for Inductively Coupled Plasma Spectrometry', *Spectrochim. Acta Rev.*, **13**(2), 89–166 (1990).
34. C.D. Skinner, M. Cazagou, J. Blaise, E.D. Salin, 'Direct Insertion of Trace Metals Preconcentrated on Activated Charcoal into an Inductively Coupled Plasma', *Appl. Spectrosc.*, **53**(2), 191–196 (1999).
35. C.D. Skinner, E.D. Salin, 'Preconcentration and Adsorption of Metal Chelates with Analysis by Direct Sample Insertion Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **18**(5), 495–500 (2003).
36. L. Moenke-Blankenburg, 'Laser-ICP-spectrometry', *Spectrochim. Acta Rev.*, **15**(1), 1–38 (1993).
37. F.E. Lichte, S.R. Koirtyohann, *Induction Coupled Plasma Emission from a Different Angle*, Paper 26, Federation of Analytical Chemistry and Spectroscopy Society, Philadelphia, 1976.
38. D.R. Demers, 'Evaluation of the Axially Viewed (End-on) Inductively Coupled Argon Plasma Source for Atomic Emission Spectrometry', *Appl. Spectrosc.*, **33**(6), 584–591 (1979).
39. E.C.C. Bailey, *Spectroscopy*, Longman Green, London, Chapter 1, 1905.
40. J.W. Olesik, 'Echelle Grating Spectrometers for Inductively Coupled Plasma – Optical Emission Spectrometry – A Review of Basic Equations and Operating Principles', *Spectroscopy*, **14**(10), 36–41 (1999).

41. C.W. Earle, M.E. Baker, M. Bonner Denton, R.S. Pomeroy, 'Imaging Applications for Chemical Analysis Utilizing Charge Coupled Device Array Detectors', *Trends Anal. Chem.*, **12**(10), 395–403 (1993).

42. A.T. Zander, R.L. Chien, C.B. Cooper, P.V. Wilson, 'An Image Mapped Detector for Simultaneous ICP-AES', *Anal. Chem.*, **71**(16), 3332–3340 (1999).

43. B.W. Smith, M.L. Parsons, 'Preparation of Standard Solutions: Critically Selected Compounds', *J. Chem. Educ.*, **50**(10), 679–681 (1973).

44. W.B. Barnett, V.A. Fassel, R.N. Kniseley, 'Theoretical Principles of Internal Standardization in Analytical Emission Spectroscopy', *Spectrochim. Acta*, **23B**(10), 643–664 (1968).

45. G.J. Schmidt, W. Slavin, 'Inductively Coupled Plasma Emission Spectrometry with Internal Standardization and Subtraction of Plasma Background Fluctuations', *Anal. Chem.*, **54**, 2491–2495 (1982).

46. A.C. Davis, B.W. Alligood, C.P. Calloway, B.T. Jones Jr, 'Direct Determination of Cadmium in Urine by ETV-ICP Analysis Using a Tungsten Coil Atomizer', *Appl. Spectrosc.*, **59**(10), 1300–1303 (2005).

47. J.M. Mermet, E. Poussel, 'ICP Emission Spectrometers: 1995 Analytical Figures of Merit', *Appl. Spectrosc.*, **49**(10), 12A–18A (1995).

48. G.L. Long, J.D. Winefordner, 'Limit of Detection: A Closer Look at the IUPAC Definition', *Anal. Chem.*, **55**(7), 712A–724A (1983).

49. P.W.J.M. Boumans, 'Detection Limits and Spectral Interferences in Atomic Emission Spectrometry', *Anal. Chem.*, **66**(8), 459A–467A (1994).

50. M. Thompson, R.M. Barnes, 'Analytical Performance of Inductively Coupled Plasma–Atomic Emission Spectrometry', in *Inductively Coupled Plasma in Analytical Atomic Spectrometry*, 2nd edition, eds. A. Montaser, D.W. Golightly, VCH Publishers, New York, 249–297, 1992.

51. T.C. Johnson, R.S. Perry, L.J. Fick, H.B. Fannin, 'An Examination of Relative Trends of Atomic Detection Limits in the Inductively Coupled Plasma', *Spectrochim. Acta*, **52B**(1), 125–129 (1997).

52. *Guide to Atomic Spectroscopy Techniques and Applications: AA, GFAA, ICP, ICP-MS*, Perkin-Elmer Instruments, Norwalk, 2000.

53. W. Slavin, ed, *Graphite Furnace AAS: A Source Book*, Perkin-Elmer, Ridgefield, 1984.

54. M. Carre, S. Excoffier, J.M. Mermet, 'A Study of the Relation between the Limit of Detection and the Limit of Quantitation in Inductively Coupled Plasma Spectrochemistry', *Spectrochim. Acta*, **52B**(14), 2043–2049 (1997).

55. P.J. Galley, G.M. Hieftje, 'Easily Ionizable Element (EIE) Interference in Inductively Coupled Plasma Atomic Emission Spectrometry-II. Minimization of EIE Effects by Choice of Observation Volume', *Spectrochim. Acta*, **49B**(7), 703–724 (1994).

Laser Ablation in Atomic Spectroscopy

Richard E. Russo, Xianglei Mao, Oleg V. Borisov, and Haichen Liu
Lawrence Berkeley National Laboratory, Berkeley, USA

1 Introduction **1**
1.1 Characteristics of Laser Ablation 1
1.2 Solid Sampling Techniques: Comparison 2
2 Fundamental Characteristics **3**
2.1 Ablation Processes 3
2.2 Processes Affecting Accuracy 4
2.3 Processes Affecting Sensitivity 5
3 Analytical Characteristics **7**
3.1 Ablation Chamber and Transport Tubing 7
3.2 Laser Systems and Optics 9
3.3 Inductively Coupled Plasma Atomic Emission Spectroscopy and Inductively Coupled Plasma Mass Spectrometry Instrumentation 9
3.4 Detection Limits 10
3.5 Calibration and Optimization 10
4 Applications **12**
4.1 Environmental and Oceanography 12
4.2 Geochemistry and Cosmochemistry 12
4.3 Forensics and Authentication 14
4.4 Waste-sample Analysis 15
5 Noninductively Coupled Plasma Applications **15**
5.1 Laser Ablation Ion-storage Time-of-flight Mass Spectrometry 15
6 Perspectives and Future Trends **15**
Acknowledgments **16**
Abbreviations and Acronyms **17**
Related Articles **17**
References **17**

Laser ablation (LA) is a unique technique to transform a solid sample into vapor-phase constituents, which then can be chemically analyzed by atomic spectroscopy. Ablation brings many exciting capabilities to the field of chemical analysis, primarily because of the laser-beam properties. The ability to analyze directly any solid sample without sample preparation and minimal sample-quantity requirements are just some of the unique capabilities. This article discusses current issues related to using LA in atomic spectroscopy. A general introduction to LA sampling is presented along with a comparison with other solid sampling techniques. The critical issues for all analytical techniques are calibration, accuracy, and sensitivity. Techniques that have been demonstrated to address these analytical characteristics and define these parameters for LA are investigated in detail. Finally, many unique applications are described, ranging from dating geological materials to providing crime-scene evidence. Most of these applications could not be performed without the use of a laser beam. The inductively coupled plasma (ICP) will be emphasized in this article because it is currently the most prevalent excitation and ionization source for chemical analysis using LA.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Want to know the composition of an unknown solid sample? Don't want to deal with acids, wastes, and laborious procedures to conduct the analysis? What is the best way to transform the unknown solid sample into vapor-phase constituents for chemical analysis? The answer: explode the sample with a high-power laser beam! This is precisely the role of LA in atomic spectroscopy. Complex solid samples, including environmental wastes, geochemical materials, coatings, extraterrestrial samples, and others dictate the development of a sampling approach for chemical analysis that does not rely on complicated, laborious dissolution procedures; a direct approach for analyzing unknown samples has been a quest of analytical chemistry for a long time. In some cases, the desire is to analyze a sample chemically without destruction, especially if that sample is the Shroud of Turin! These are some of the characteristics that make ablation sampling attractive for atomic spectroscopy.

This article will discuss current issues related to LA in atomic spectroscopy, including calibration and optimization, accuracy, sensitivity, and particle transport. In addition, many unique applications will be described, in some cases applications that cannot be performed without the use of a laser. The ICP is emphasized in this article because it is the most prevalent excitation and ionization source for chemical analysis using LA at this time.

1.1 Characteristics of Laser Ablation

One of the most unique characteristics of LA is that any solid sample can be directly ablated into vapor-phase constituents (vapor and particles), be it a rock, hair sample, extraterrestrial sample, or priceless piece

of artwork. Only a very small portion of the sample is required for the analysis – sometimes of the order of picograms or less. Damage at these levels is generally not visible without the use of a microscope. In addition, there is no requirement for sample preparation; the laser beam can be used to ablate the surface contamination, as well as dig a crater and provide bulk analysis.

Because the laser beam directly converts the solid sample into the vapor phase, liquid reagents are not required to digest the sample. Therefore, there is no solution waste, no worries about loss of volatile species, and minimal sample handling. In addition, when liquid sample introduction is used, a larger portion of the sample is required because of dilution. Another significant and unique advantage of LA is the ability to perform spatial (micrometer) characterization. Since the laser beam can be guided to an exact location on a surface, the analyst can choose where the explosion and, therefore, the sampling occur. If the sample of interest is the microscopic inclusions in a bulk material, or particles on a filter paper, it is not necessary to analyze the entire sample; most of the mass in this case is not related and leads to large background signals in the analytical source. Finally, if one wants the analysis performed immediately, LA is the way to go – there are no time-consuming sample preparation or dissolution procedures; just place the sample in a simple chamber and hit it with the laser beam. With capabilities and advantages such as these, LA atomic spectroscopy is now utilized routinely in many industries, using commercially available and laboratory-built systems.

1.2 Solid Sampling Techniques: Comparison

The development of any ''new'' technique for elemental analysis requires parameters such as sampling, excitation, ionization, and detection to be critically investigated and optimized. An ''ideal'' sample introduction technique has to satisfy certain criteria, which include:

- remove a sufficient quantity of sample;
- reproducible sampling;
- minimal interference on the performance of the detection system;
- sampled material must be representative of the bulk composition (stoichiometry);
- absence of matrix effects (no variation of analyte signal in different matrices);
- sample transport to the excitation or ionization source without losses;
- absence of memory effects and sample carryover;
- adjustment of sampling parameters to satisfy detection requirements.

Among the most popular direct solid sampling methods are LA, glow discharge (GD), and spark/arc ablation. Direct analysis of solids with GD sources has been used with both atomic emission and mass spectrometric detection. GD converts a solid sample directly into the atomic phase by sputtering processes. Because sputtering is a primary sampling mechanism, heating is minimal and preferential vaporization is minimized. GD is best suited for conductive samples, although nonconductive samples can be analyzed with a radiofrequency discharge. A GD source coupled to a high-resolution mass spectrometer is a powerful technique for achieving excellent depth profiling and detection characteristics with a limit of detection (LOD) in the range $10–100\,\mathrm{ng\,g^{-1}}$.

For arc/spark ablation, mass is eroded from the sample in the form of atoms, molecules, vapor, droplets, solid flakes, and large particles. Normally, only conductive samples can be used with a spark discharge; however, nonconductive materials can be analyzed by mixing the sample with a conductive matrix.

Because of unique benefits and capabilities for direct solid sampling, LA has outpaced development of these other techniques. Laser sampling has been employed with flame and graphite furnace atomic absorption spectroscopy, with typical LODs in the micrograms per gram range. A description of these methods can be found in a review by Darke and Tyson.[1] The direct current plasma (DCP) and microwave-induced plasma (MIP) have been used for the analysis of solid samples with LA sampling.[2,3] The DCP and MIP offer good detection limits in the micrograms per gram range, with linear calibration curves over several orders of magnitude of concentration.

The ICP has become the prevalent source for chemical analysis with LA sampling.[4] The ICP plasma is robust, with high temperature and electron number densities, and is in most cases unperturbed by small amounts of sample. A separate article (**Inductively Coupled Plasma/Optical Emission Spectrometry**) is dedicated to the ICP. Slurry nebulization of powders, electrothermal vaporization (ETV), direct sample insertion (DSI), spark/arc and LA techniques are among the most popular methods investigated for solid sampling into the ICP. Each of these techniques has unique features: sensitive analysis of powders (slurry nebulization, ETV); analysis of small volumes of solids and liquids with selective solvent or matrix removal (ETV); direct sampling and atomization in the ICP plasma (DSI); analysis of conductive samples (spark ablation); and analysis of any sample (LA). A comparison of slurry nebulization, ETV, and LA methods was reviewed by Darke and Tyson.[5] DSI for ICP was recently reviewed by Sing.[6]

2 FUNDAMENTAL CHARACTERISTICS

Because of the numerous capabilities and advantages for atomic spectroscopy, there is a tremendous demand to understand fundamental LA processes. LA is complex, consisting collectively of many nonlinear mechanisms, each operative on different timescales, from femtoseconds to seconds. When the laser beam irradiates the surface, electrons will be released on the femtosecond to picosecond timescales, followed by atomic and molecular mass on the nanosecond timescale, and then the eruption of large particles, microseconds after the laser pulse has ceased (Figure 1a–c). With different processes occurring over so many orders of magnitude in time, it is no wonder the collective phenomena have never been combined into a unifying theory. However, it is critical to point out before discussing fundamental issues that LA sampling is easier than liquid nebulization. In fact, there have been many years of science devoted to proper dissolution procedures in order to achieve accurate analysis – and still one must be very careful when dissolving a sample that is truly 'unknown'. The underlying fundamentals of LA should not preclude its development as the method of choice for direct solid sampling in atomic spectroscopy.

Studying fundamental LA mechanisms will provide further improvements and benefits for atomic spectroscopy. By understanding ablation, it will be possible to control process variables, such as the ability to couple the laser beam efficiently into the sample, ablate a reproducible quantity of mass, control the amount of mass ablated, minimize preferential ablation, produce stoichiometric ablation, and control the particle size distribution. By controlling these variables, LA in atomic spectroscopy will become a routine chemical analysis technology for environmental, health, forensics, nonproliferation, and other applications where the primary sample is in the solid phase.

2.1 Ablation Processes

Many studies have been dedicated to understanding fundamental ablation mechanisms. Research studies have shown that LA of a solid sample consists of several stages, in which different kinds of 'vapor products' are ejected. The initial stage is electronic excitation inside the solid, accompanied by ejection of electrons at the sample surface, due to both photoelectric and thermionic emission. During this time, energetic electrons in the bulk of the solid also transfer energy to the lattice through a variety of scattering mechanisms; the sample target then undergoes melting and vaporization, followed by ionization and formation of a plasma plume consisting of the sample constituents. The expanding plume interacts with the surrounding gas to form a shock wave, causing the ambient gas to become further ionized. The expanding high-pressure plasma exerts a force back to the target, which flushes out the melted volume. This recoil pressure and flushing mechanism can produce large-sized particles (several micrometers). There is evidence of explosive boiling, occurring microseconds after the laser is finished, which also can produce micrometer-sized particles. Atomic- and micrometer-sized particles will both be transported to and digested in the ICP. Therefore, it is important to understand mechanisms of atomic- and

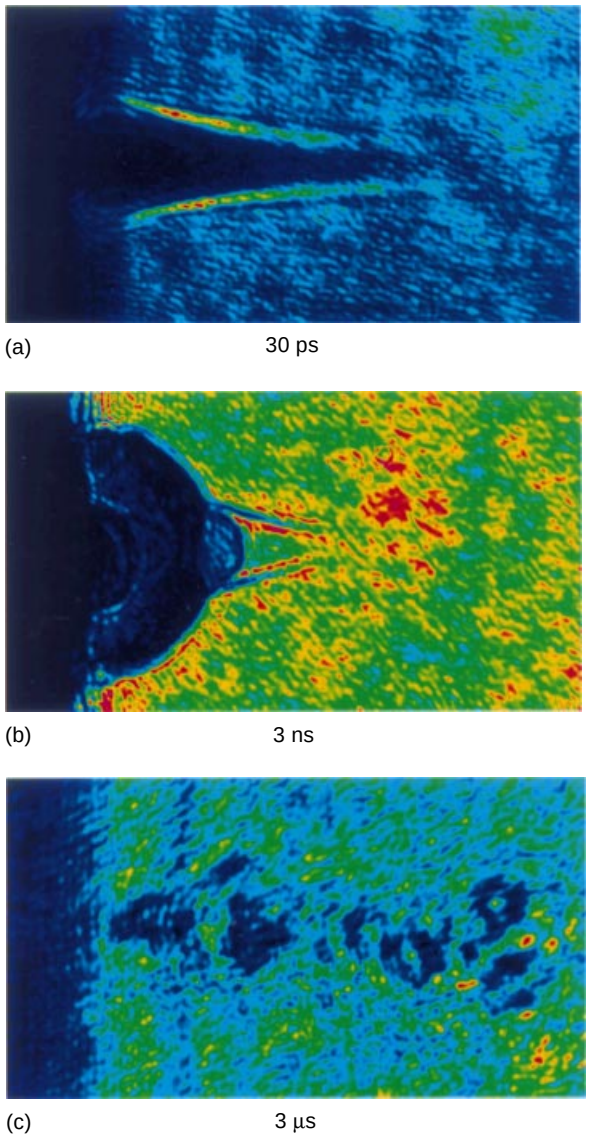

Figure 1 LA involves complex and collective phenomenon, exhibiting significant nonlinearity in space and time. The images show the evolution of the laser explosion from a solid surface. (a) At 30 ps after the laser hits the sample, electrons leave the surface and collide with air to form the plasma; (b) at 3 ns, atomic vapor escapes from the surface; and (c) at 3 μs, large (>20 μm) particles are spalled from the bulk.

micrometer-sized particle generation; particles in each of these size ranges can have different chemistries, affecting analytical accuracy. The particle size distribution will also affect transport efficiency and therefore analytical sensitivity.

To achieve accurate LA sampling, the composition of the ablated mass should be the same as the sample composition. To achieve high sensitivity, all of the ablated mass should be transported to and digested in the ICP. There are four primary processes that influence the accuracy and sensitivity in laser ablation/inductively coupled plasma (LA/ICP). The first is the LA process itself. For different laser conditions, the composition and quantity of the ablated mass can change significantly. The particle size distribution and particle composition can also change with laser conditions. The second process is transport, which includes the sample chamber design and tubing used to carry the ablated mass to the ICP. Transport efficiency will be different for different sized particles. The third process is sample digestion and excitation in the ICP. The conditions in the ICP (electron temperature and number density) which are controlled by ICP power, gas flow rate, and height above the load coil will influence vaporization, atomization, and ionization of the analyte. However, the amount and composition of the analyte and the matrix can change the ICP conditions (Figure 2). The fourth process includes detection, using the mass spectrometer for inductively coupled plasma mass spectrometry (ICPMS) or the spectrometer and photon detector for inductively coupled plasma atomic emission spectroscopy (ICPAES). These processes are discussed separately below as to their influence on accuracy and sensitivity.

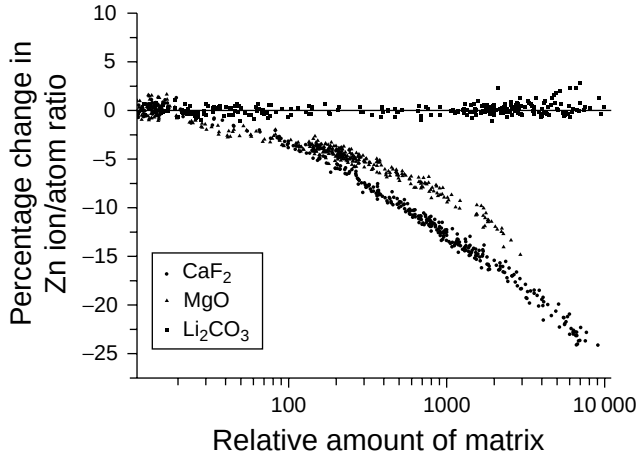

Figure 2 Percentage ratio change of Zn ionic to atomic emission lines in the ICP versus amount of matrix. The ratio of ionic to atomic line intensity correlates to the ICP conditions. LiF, CaF_2 and MgO were the matrices.

2.2 Processes Affecting Accuracy

An expectation of any analytical procedure is accurate analysis; the detected mass composition must be the same as the sample composition. For LA sampling, fractionation (preferential mass removal during LA), which can cause inaccurate analysis, can occur under some conditions. Numerous studies have shown that fractionation can be minimized or eliminated, depending on the sample and laser properties.[7–13] The distribution of elements in the matrix, their status (inclusions), and zone migration will influence fractionation.[11,12] Fractionation at the sample is a function of wavelength; ultraviolet (UV) LA has been found to provide better stoichiometry than infrared (IR) LA.[14–16] The number of laser pulses at a fixed location on the sample, the laser irradiance, and laser beam spot size (related to focus position) also influence fractionation (Figure 3).[13,16,17] For fundamental purposes to describe fractionation at the sample, attempts have been made to correlate ablation behavior with melting, boiling, vapor pressure, atomic or ionic radius, charge, and speciation.[1,15,18]

The mechanism for atomic mass ejection can be thermal vaporization and/or plasma sputtering. The sputtering process involves the high-temperature, high-pressure plasma colliding with the sample surface. The ablation rate in these two processes will depend on the bond energy; the stronger the bond, the less mass is removed. Fractionation on the nanosecond timescale can be related to thermal properties of the sample constituents. Fractionation in alloys depends on the latent heat of vaporization; the higher the latent heat, the more difficult it is to vaporize a constituent. For example, in brass, the latent heat of vaporization of Cu is greater than that of Zn;[10] Zn can be preferentially ablated compared with Cu in these samples. By using a calibrated ICPAES instrument, fractionation based on thermal vaporization was demonstrated by measuring Zn/Cu mole ratios during ablation of brass. With a 30-ns, 248-nm excimer laser, the Zn/Cu mole ratio initially decreased with increasing laser irradiance, then stabilized at irradiance greater than approximately 0.3 GW cm^{-2}. The initial decreasing Zn/Cu ratio was due to thermal vaporization.[14] For higher laser irradiance, there are likely several competing mechanisms involved in the ablation processes. Melt flushing from the crater, plasma shielding, and/or radiative heating by the laser-induced plasma can contribute to mass ejection.[7,10,19,20] Picosecond LA is even more esoteric; nonthermal mechanisms must be operative. Using a 35-ps pulse (Nd : YAG, $Nd^{3+} : Y_3Al_5O_{12}$) laser, some type of nonthermal mechanism appears to govern the LA process because Cu is enriched in the vapor at lower irradiance. The Zn/Cu ratio approaches the stoichiometric value at higher irradiance. These experiments demonstrate

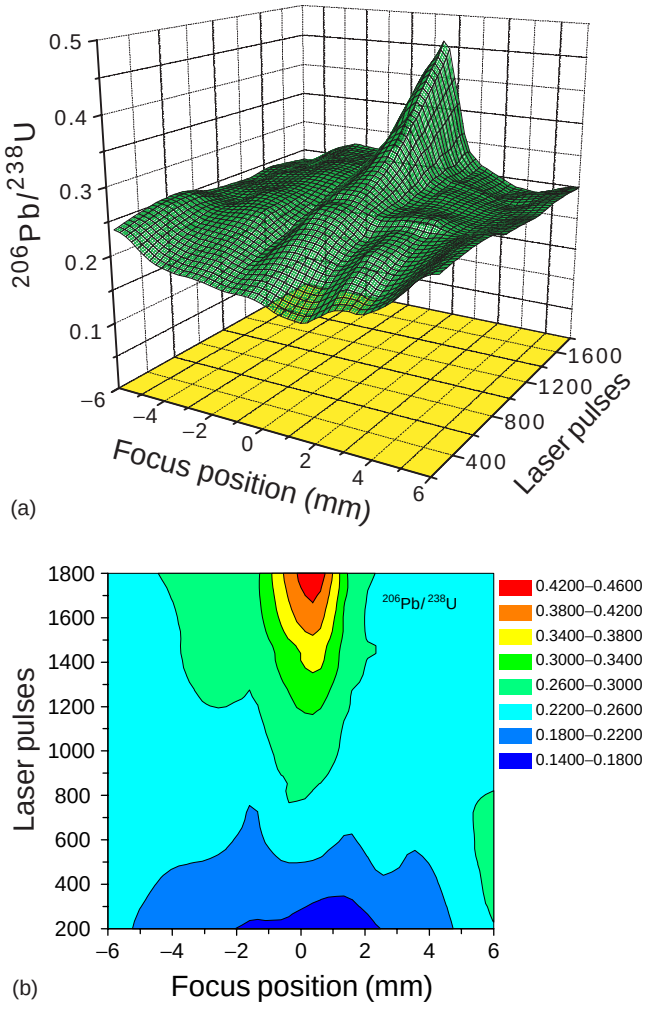

Figure 3 Pb–U fractionation versus number of laser pulses and lens-to-sample focal position (a) three-dimensional plot; (b) contour plot. Neodymium:yttrium aluminum garnet (Nd:YAG) laser, $\lambda = 213$ nm; pulse duration = 6 ns; energy = 1.15 mJ; sample: NIST SRM 610 glass in which the $^{206}Pb/^{238}U$ stoichiometric ratio is 0.2249. The positive and negative values represent the laser focus above and below the sample surface, respectively.

that different mechanisms contribute to fractionation, although identification of the mechanisms is still a large research effort.

Fractionation during transport and/or in the ICP can occur.[21,22] Fractionation during transport involves selective vapor condensation on the tubing walls or the selective nucleation of species on different-sized particles.[12] These effects are governed by the vapor-phase morphology (droplet/particle shape and size) and chemical composition. Because transport efficiency and chemical composition are particle size dependent, fractionation can occur during transport. Figg et al.[22] demonstrated this effect by inserting a coiled Tygon tube into the transport path; both particle size distribution and fractionation changed significantly. It is believed that many of the large particles come from flushing out the molten liquid layer. The composition of the molten liquid could be significantly different from the bulk because of preferential vaporization and re-deposition from the plasma plume. For "large" particles entering the ICP and undergoing sublimation (vaporization), fractionation may exist in the ICP itself.[21,22] It is important to state here that fractionation is not a problem that precludes the use of LA for accurate chemical analysis, especially when matrix-matched standards exist. This section was designed to make the reader aware that such issues can exist. There are many applications of LA, as will be discussed in a later section, in which fractionation does not influence the analysis.

2.3 Processes Affecting Sensitivity

The signal intensity in atomic emission spectroscopy (AES) or mass spectrometry (MS) corresponds directly to the quantity of mass ablated and transported to the ICP. Depending on the concentration of the elements in the sample, it may be necessary to increase the quantity of ablated sample to achieve better signal-to-noise ratios. Laser-beam properties (such as wavelength, pulse duration, energy, fluence, irradiance, and temporal and spatial profiles) and the ambient gas influence the quantity of mass ablated per laser pulse.[7,15,16,21,23–33]

2.3.1 Laser Energy

The ablated mass can be increased by increasing the laser-beam energy or fluence (energy per unit area). However, the relationship of ablated mass to laser-beam energy or irradiance is not linear;[7,18,19,34] mass was found to follow a power law with irradiance (I^m).[18,19,34] Using nanosecond laser pulses, with irradiance less than approximately 0.3 GW cm^{-2}, m had values ranging from 2 to 5 for many samples. When the irradiance was >0.3 GW cm^{-2}, m became <1. However, as the irradiance was increased further, another change in the mass ablation rate occurred, in which m increased to >2. In the lower irradiance region, thermal vaporization was found to be a dominant process,[14] as evidenced also by fractionation (discussed earlier). In the middle laser irradiance range, plasma shielding may be a major factor governing ablation. The laser-induced plasma formed above the target surface can absorb and/or reflect incident laser energy, thereby decreasing the efficiency of laser energy available for mass ablation.[35] A plasma can transmit just enough energy to the surface to sustain itself. The density and temperature of the plasma can adjust in such a manner that the optical thickness remains

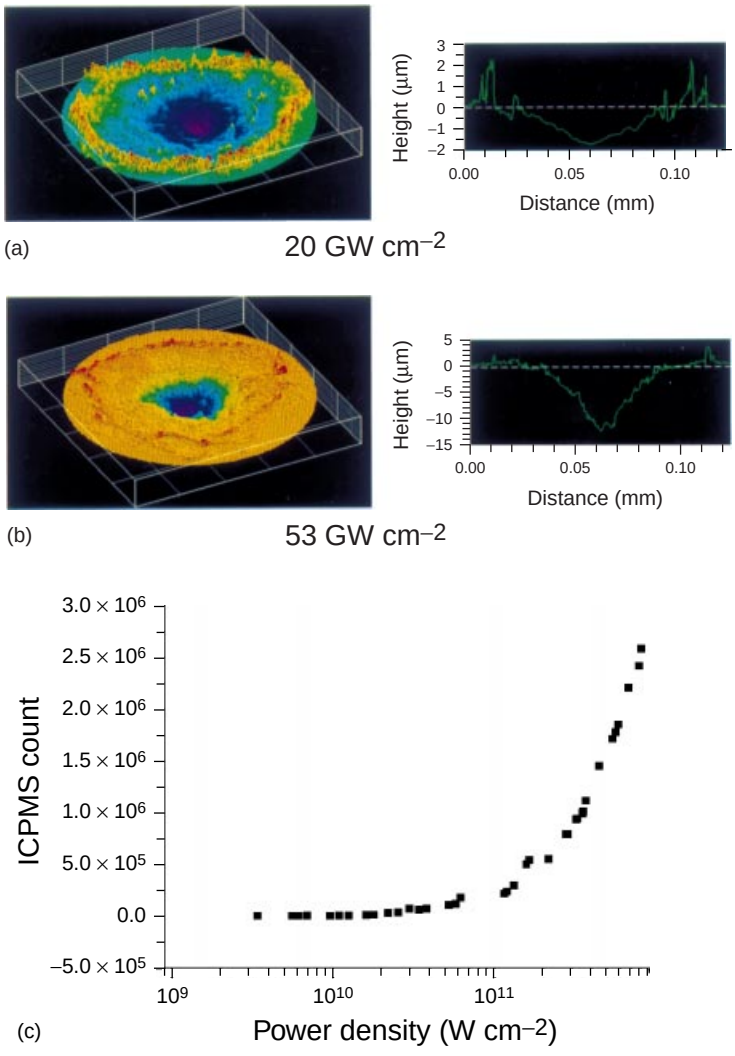

Figure 4 Crater profiles (a) before and (b) after an explosive boiling threshold. The crater depth changes dramatically from 1 μm to more than 10 μm. (c) ICPMS signal versus laser power density. The ICPMS intensity increases dramatically across the threshold.

constant; the final proportion of the laser energy absorbed and transmitted is constant. It is important to point out that this plasma is the basis of another chemical analysis technique, discussed in the article **Laser-induced Breakdown Spectroscopy**.

In the third irradiance region, mechanisms such as phase explosion and spallation may be dominant. Both crater depth and ICPMS intensity (Figure 4a–c) show dramatic increases at a threshold irradiance of 20–30 GW cm^{-2}. This dramatic change may be caused by explosive boiling.[36] The superheated molten liquid can experience an increased fluctuation in its density[37] when the temperature and pressure approach the critical point. Near the critical point, this fluctuation can generate vapor bubbles in the superheated liquid. For vapor bubbles larger than a critical radius, bubble growth will occur; bubbles smaller than the critical radius will collapse.[37] Once bubbles of a critical radius have been generated in the superheated liquid, the volume undergoes a rapid transition into a mixture of vapor and liquid droplets. During explosive boiling, rapid expansion of the high-pressure bubbles in the liquid leads to a violent ejection of the molten droplets from the sample. The shadowgraph images of liquid droplets ejected from the silicon surface (see Figure 1a–c) indicate that the onset of the explosive boiling is at an irradiance of approximately 2.2×10^{10} W cm^{-2}.

2.3.2 *Laser Pulse Width*

Laser wavelength and pulse duration also influence the quantity of ablated mass; the shorter the wavelength and pulse duration, the better the mass removal efficiency.[15,16,23–26,28–30,38,39] Using an Nd : YAG laser with a 3-ns pulse duration, the ICPAES intensity was

found to be almost an order of magnitude greater for UV than IR ablation sampling, with the same fluence. The absolute enhancement was found to be a function of laser irradiance.[39] When plasma shielding does exist, a lower wavelength is better because there is less absorption by the plasma; more energy is used to remove mass instead of heating the plasma. The shorter the laser pulse duration, the more efficient is the ablation process; picosecond laser pulses provided an order of magnitude greater signal intensity in ICPAES compared with a nanosecond pulsed laser with the same fluence. An explanation for increased mass ablation using picosecond pulses is that plasma shielding may be weaker; more laser energy is coupled to the sample than is absorbed by the plasma. Another possible mechanism is that more laser energy is converted to ablated mass instead of being lost in the sample through thermal dissipation, which is a function of pulse duration.[40] If the pulse duration is femtoseconds, the ablation process should be even more efficient.[32,33] For femtosecond LA, the generation of vapor and a plasma will occur after the laser pulse is finished. Therefore, there should be no plasma shielding on the femtosecond timescale. Thermal diffusion to the solid also should be negligible. Currently, there are no reports of using femtosecond LA for chemical analysis. However, as femtosecond laser technology becomes more reliable, it is expected that new benefits to chemical analysis (and other applications) will be forthcoming.

2.3.3 Gas Environment

The gas atmosphere in the sample chamber can have a dramatic effect on ablation behavior.[7,30,31,34,41,42] The use of different gases has been found to enhance sensitivity either by reducing plasma shielding or by influencing the ablated particle size distribution. With He in the ablation chamber, the mass ablation rate was 2–5 times and 10 times greater than that with Ar in the chamber for nanosecond and picosecond laser pulses, respectively.[13,43] He and Ne provided an increased mass ablation efficiency compared with a decrease with Xe and Kr, relative to Ar for the nanosecond-pulsed LA. The enhancement or depression was found to be dependent on the laser irradiance. For picosecond LA sampling, only He provided an enhancement; there was very little influence on the mass ablation behavior by the other noble gases.

3 ANALYTICAL CHARACTERISTICS

A typical experimental system for LA sampling with the ICP is shown in Figure 5. In general, mass ablated from a target surface inside the sample chamber is entrained into an Ar gas flow and transferred into the ICP source, where particles are first vaporized and then atomized and ionized. In most cases, the sample chamber is placed on a motorized or manually controlled micrometer translation stage. This is especially convenient for the analysis of heterogeneous samples, where different areas on a sample surface are to be analyzed. The individual components are described separately below.

3.1 Ablation Chamber and Transport Tubing

Various types of chambers have been described in the literature; a summary can be found in a review by Moenke-Blankenburg.[44] Typically, the sample is placed inside a chamber. However, there are cases in which the sample itself was the bottom of the chamber.[45] The simplest chamber can be a glass tube with two ports for gas flow, and a quartz window (or other transparent material at the laser wavelength) for laser beam delivery. Argon gas flow dynamics inside the chamber play a significant role in particle entrainment. With laminar flow in the chamber, entrainment efficiency and thus detection characteristics are improved; turbulent flow contributes to losses by trapping particles in stagnant flow regions.[46] The volume of the chamber is an important parameter; a larger internal volume can lead to sample dilution and memory effects. Because LA processes are transient in nature, larger chamber volumes may be advantageous for signal averaging during repetitive sampling experiments. Another important parameter is chamber length; if the laser beam window is located too close to the sample surface, vapor deposition can occur, reducing transmission and therefore the laser energy. Another issue concerning the chamber length is that different times are required for particles of different sizes to be entrained by the flow; larger particles will travel further before becoming entrained.[47] Depending on the flow velocity inside the chamber, large particles can be lost owing to collisions with the walls before entrainment.

Dispersion of different-sized particles in the ablation chamber depends on the Ar gas flow velocity in the chamber and can contribute to broadening of signal peaks (recorded during single laser pulse experiments). Further dispersion of ablated particles in the transport tube can occur and depends on the transfer tube length and internal diameter. To minimize dispersion, a short, narrow tube should be used. Although particle dispersion in the chamber and transport tube may be important for analysis during single laser pulse experiments, there are no significant advantages to using a short tube during LA sampling with repetitive pulses.

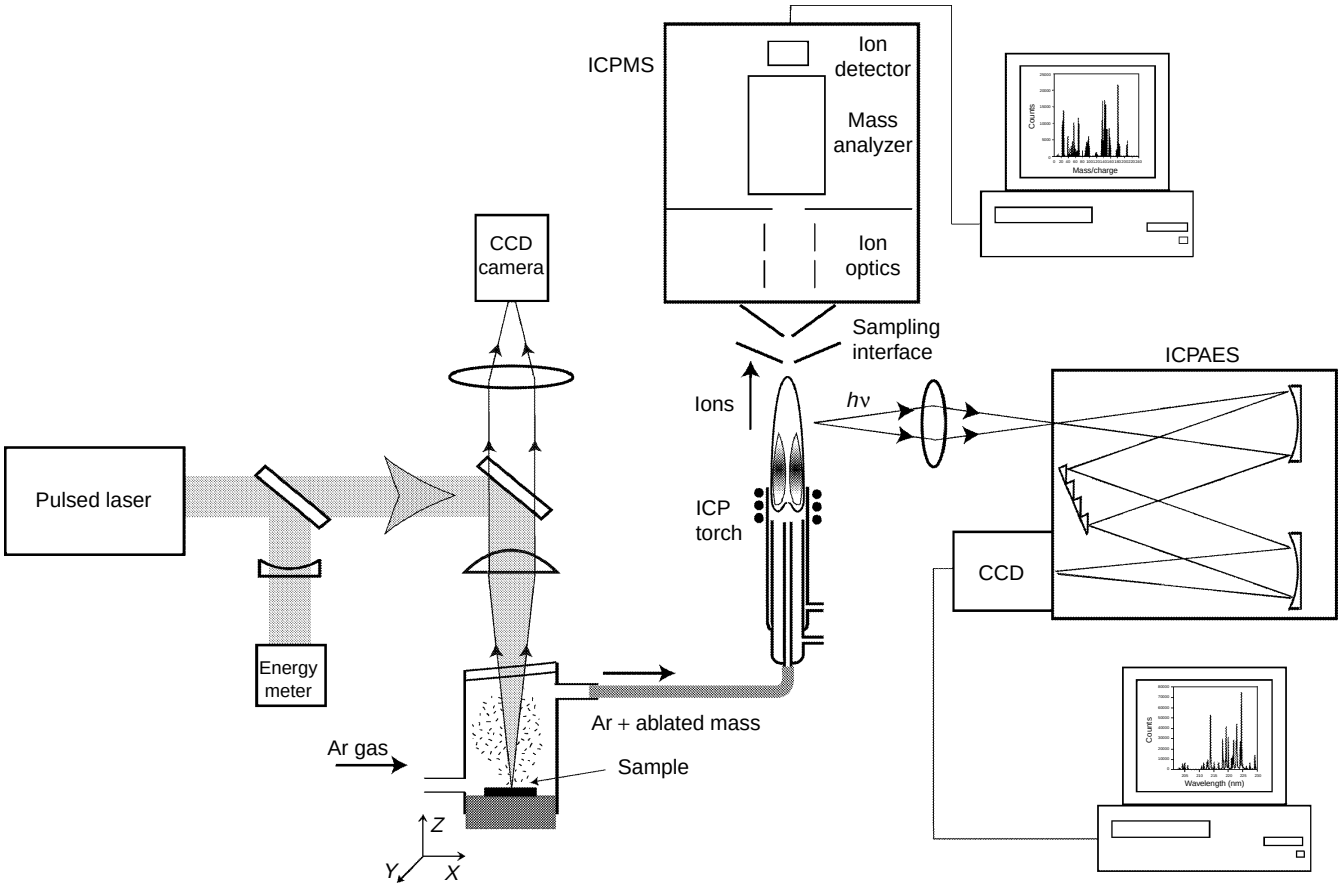

Figure 5 Schematic diagram of LA sampling into an ICP source. CCD, charge-coupled device.

3.1.1 Mass Transport Efficiency

For a given ablation chamber/tube configuration, particle transport efficiency depends on the ablated particle size distribution. The number of particles and their size distribution depends on the laser and sample properties.[22,47,48] The volume distribution of laser-generated particles changes with respect to the laser wavelength; there are more large particles generated with an IR laser than with a UV laser.[48] Since melting and melt-fusing may be responsible for the generation of particles, a thicker molten layer may produce a greater fraction of large particles. The optical absorption depth in most solid samples is greater for IR than UV wavelengths. Therefore, a larger molten liquid volume may be responsible for large particle production using IR LA.

Entrainment of ablated mass into the gas stream and transport to the ICP are particle size dependent. Large particles may not be entrained and those that are may not completely vaporize in the ICP. Particle sizes should be less than about 2 μm for efficient transport to and excitation in the ICP.[45,49,50] Particle entrainment efficiency can be defined as the ratio of mass entering the ICP to the total mass ablated from the sample. Particle entrainment efficiency has been found to decrease with increasing laser irradiance. Entrainment efficiency was about 25% at low irradiance and decreased to about 5% at high irradiance.[47] One possible reason for the small entrainment efficiency at high irradiance may be the formation of excessively large particles (>5 μm); large particle ejection is observed in Figure 1(a–c). The removal of large fractured pieces (>∼50 μm), possibly due to increased thermal stress and pressure on the sample surface, was also observed after ablation at high irradiances (Figure 6). These very large particles will not be entrained into the argon gas flow, but instead will settle in the ablation chamber due to gravity. Transport efficiency as a function of particle-size distribution needs to be critically studied for improving LA sampling in atomic spectroscopy.[45,51] A few studies have addressed particle transport, with preliminary data suggesting that particles in the 0.1–1.0 μm range reach and vaporize in the ICP.[22,45,48]

Literature values for LA transport efficiency are in the range of about 5–40%.[45,47] The large variation represents the effects of laser beam conditions, sample material properties, and ablation chamber/transfer

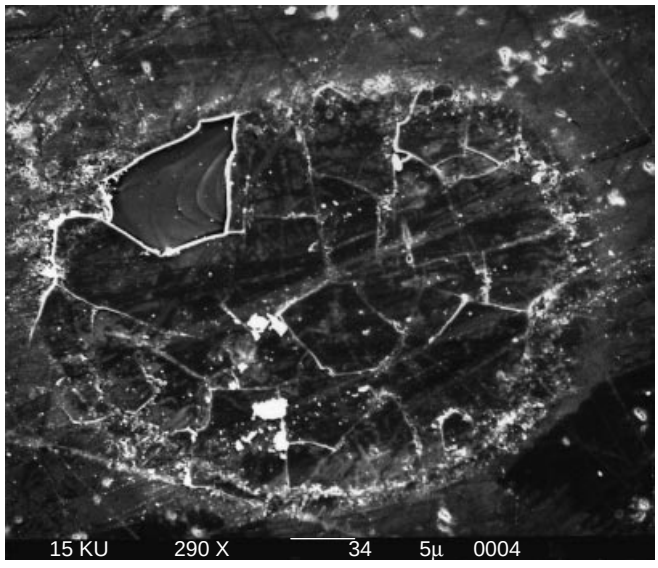

Figure 6 Scanning electron microscope image of an ablated glass surface. A large piece (~50 µm) of the cracked sample was removed from the ablation spot.

tube geometries. In general, particle losses in the sample chamber and transport tube are mainly due to gravitational settling or inertial impact to the walls for large particles, and diffusion to the walls for smaller particles. These processes depend on the tube length and internal diameter. Transport efficiency for larger particles, which represent most of the ablated mass and thus are responsible for most of the ICP signal, can be improved by utilization of a short, narrow tube. Generally, particles in the range 0.1–1.0 µm are most efficiently transported to the ICP. For a given tubing length and diameter, theoretical calculations predict losses to be flow dependent. In practice, however, no significant differences in the ICPMS signal count rates were observed when the Ar flow rate was varied from 0.1 to 0.9 L min^{-1} during ablation of Zr metal. Particle losses inside the sample chamber and transport tube can contribute to memory effects. To minimize these effects, it is recommended that the transport tube be cleaned regularly or replaced, and the system flushed with Ar gas at increased flow rates between measurements.

3.2 Laser Systems and Optics

Over the years, several pulsed lasers have been tested for ablation. Ruby (694.3 nm), CO_2 (10.6 µm), free-running Nd : YAG and excimer lasers have been shown to efficiently ablate solid samples.[1] However, Q-switched Nd : YAG lasers with nanosecond pulses have become the most prevalent systems used for LA today. These lasers are relatively inexpensive, easy to operate, and compact, have good pulse-to-pulse and long-term stability characteristics, and deliver sufficient energy for ablation. The fundamental harmonic of 1064 nm can be easily doubled (532 nm), tripled (355 nm), quadrupled (266 nm), and quintupled (213 nm). The Q-switched Nd : YAG laser operated at the fourth harmonic is currently the most popular for LA chemical analysis. Recently, it has been shown that samples with relatively low absorption at 266 nm, such as calcite and garnet, can be more efficiently ablated with the Nd : YAG at 213 nm[52] or the ArF excimer (193 nm).[53]

The laser beam is transferred to the sample chamber by means of mirrors, beam splitters, and/or prisms. In the simplest system, focusing is achieved with a plano-concave singlet lens, which has a low degree of spherical aberration. Laser beam spot size at a sample surface can be easily adjusted by translating the lens relative to the sample. The minimum spot size with radius, w, that can be achieved by focusing a Gaussian diffraction-limited laser beam can be approximated by Equation (1):

$$w \propto \frac{\lambda f}{\pi w_0} \tag{1}$$

where λ is the laser wavelength, f is the lens focal distance, and w_0 is the initial laser beam radius. By using a lens with a short focal distance and a wide initial beam diameter, the smallest LA spots (best spatial resolution) can be achieved. An optical microscope equipped with a CCD camera is an excellent addition to experimental systems, allowing easy focusing of the laser beam on to a selected sample location, with beam spot sizes in the order of several micrometers. Such systems are widely used, especially for the analysis of geological samples where a high degree of spatial resolution is required.

3.3 Inductively Coupled Plasma Atomic Emission Spectroscopy and Inductively Coupled Plasma Mass Spectrometry Instrumentation

In contrast to liquid nebulization sample introduction, LA sampling is transient in nature. Thus, simultaneous detection of elements is needed. Single-channel instruments equipped with a photomultiplier tube (PMT) require scanning of a monochromator grating. Such systems are not very useful for transient signal detection when detection of more than one element is required. Direct-read polychromators can employ more than 60 exit slits with PMTs for detection of numerous wavelengths, providing simultaneous multielement capability. The photodiode array (PDA) and more recently the charge transfer device (CTD) used with a conventional Czerny–Turner-type spectrometer can cover spectral windows of several tens of nanometers simultaneously. Spectrometers with

Paschen–Runge polychromators and with échelle gratings, equipped with solid-state detectors, have multielement capability and provide simultaneous UV and visible spectral coverage.

Most commercially available ICPMS instruments utilize radiofrequency quadrupole mass analyzers. Although quadrupole mass analyzers typically have a low resolution capability (ca. 0.5 u), their low cost, ease of coupling with an ICP source, and pseudosimultaneous mass detection make them attractive for elemental analysis with LA applications. The complete mass range from 1 to 250 u can be scanned in about 0.1 s.

Many LA applications require high-precision isotopic ratio measurements; high-resolution double-focusing ICPMS instruments equipped with multicollector array detectors are well suited for this purpose. ICPMS instruments with time-of-flight (TOF) analyzers were recently demonstrated with LA sampling.[54] The TOF approach allows the collection of several thousand complete mass spectra per second. Owing to the transient nature of LA, ICPMS with a TOF analyzer is potentially advantageous for multielemental determinations, although to date only limited research has been conducted.

The detection capabilities of LA with ICPAES and ICPMS techniques depend significantly on the experimental conditions and equipment. ICPMS is typically used for minor and trace elemental analysis, whereas ICPAES has lower sensitivity and is primarily used for the analysis of major and minor constituents. Most geological applications, which require a high degree of spatial resolution for accurate microanalysis of inclusions and grains in minerals, utilize ICPMS instrumentation.

3.4 Detection Limits

One of the advantages of dry ICP conditions from LA sample introduction is that plasma excitation/ionization temperatures and electron number densities are typically higher than with a wet plasma produced during liquid nebulization. Dry plasma conditions enhance ionization and excitation processes. This benefit, along with simplified mass spectra (due to reduction of polyatomic interferences), provides improved detection characteristics. Absolute detection limits (absolute detectable amount of analyte mass) for rare-earth elements in silicate samples were shown to be two orders of magnitude better for LA ICPMS than liquid nebulization sample introduction.[55] However when relative detection limits (detectable analyte concentration in a sample) are compared, liquid nebulization sample introduction is better. This situation only exists because the amount of mass nebulized into the ICP is much greater than that ablated by the laser. Assuming a 2% efficiency for conventional pneumatic nebulizers with a $1\,mL\,min^{-1}$ sampling rate, roughly $20\,mg\,min^{-1}$ of solution is introduced into the ICP. In contrast, the amount of ablated mass per laser pulse is typically only 1 ng–1 μg. Hence for a laser repetition rate of 10 Hz and a particle transfer efficiency of 40% (maximum value reported in the literature), the amount of mass introduced into the ICP is only from $0.2\,\mu g\,min^{-1}$ to $0.2\,mg\,min^{-1}$. Hence the mass per unit time introduced into the ICP as a result of LA is about 10^2–10^5 times less than that from liquid nebulization. Relative detection limits depend on the amount of mass and are therefore better for liquid nebulization sample introduction.

Typical detection limits, determined as three times the standard deviation of the blank, are in the low micrograms per gram and even nanograms per gram range for ICPMS detection with LA, in contrast to the picograms per gram level for liquid nebulization. For LA sampling with ICPAES, laser energies of 10–100 mJ are required, compared with only few milli- or even microjoules for sampling into the ICPMS. Typical LODs for LA sampling with ICPAES are in the micrograms per gram range and are comparable to those available from liquid nebulization sampling with ICPAES. Better detection limits can be achieved by increasing the amount of ablated mass per laser pulse, which can be realized by improvement in the laser energy coupling efficiency to the sample and/or improvement in entrainment/transport efficiency. When a sample was ablated inside the ICP torch, just below the discharge, particle losses associated with mass entrainment and transport were eliminated, and a significant improvement in detection characteristics was achieved.[56]

3.5 Calibration and Optimization

The signal intensity recorded by ICPAES and ICPMS during sample introduction (both liquid nebulization and solid sampling) is proportional to the concentration of an element of interest in the sample. Absolute measurements require calibration procedures to be established, which remains an essential issue for chemical analysis.[1,5,44,57] For liquid sample introduction, standards are relatively easy to obtain in the form of single- or multielemental solutions; linear calibration curves can be generated over three to six orders of magnitude for ICPAES and eight orders of magnitude for ICPMS.[57] Assuming similar viscosities for the diluted sample and the standard solutions, the amount of aerosol aspirated into the ICP is determined by the nebulizer/spray chamber parameters. In contrast, during LA the amount of mass ablated and transported into the ICP may be different for samples and standards, if the standards are not matrix matched. For LA, an internal or external standardization procedure is required to compensate for changes in the quantity of mass ablated, even when the analyte concentration

remains constant. Matrix-matched solid standards are generally required for instrument calibration.

3.5.1 Internal Standardization and Matrix Matching

The ablated mass is determined by the properties of the matrix, and close matching of calibration standards to samples is preferred. However, the chemical form in which the analyte is present in the sample was shown to influence the elemental response.[58] Compensation for this effect often can be performed by normalization of analyte intensity to that of another element (internal standard). This procedure requires that the concentration of the internal standard, usually the matrix element, be either known [determined from independent methods, such as X-ray fluorescence (XRF) spectroscopy] or be constant in the sample and standards. In most cases, an internal standard provides excellent compensation for differences in ablation behavior for samples and standards. For trace elements in homogeneous samples, this approach has been used to improve the measurement precision to better than 1%.[18,49,59] Internal standardization is especially useful for bulk analysis with samples and standards pressed into pellets,[60] fused into glass beads,[49] or preconcentrated in an NiS button[61] where the internal standard is added during sample preparation. A single calibration graph was used for analysis of silicate rocks and limestones fused with Sc and Y oxides as internal standards.[62] For cases in which relatively ''good'' standards were available, LA proved to be a reliable and accurate chemical analysis procedure. In some cases, however, internal standardization can be limited by differences in the ablation behavior of the sample and the standards, especially when standards are not matrix matched. For example, elemental fractionation of W relative to the internal standard Ca was shown to be different for a scheelite sample and a silicate glass calibration standard.[63]

Calibration without an internal standard has been shown to be possible, especially in cases where only trace level impurities differ among the standards and sample; the trace impurities do not effect ablation behavior. Linear calibration curves were established and accurate trace elemental analysis was demonstrated for U,[64] Au and Ag,[65] and glass[66] without internal standardization. In contrast, if the matrix properties change because of changes in the analyte concentration, the LA behavior is affected. For example, for the analysis of Zn in brasses, where Zn is a major element, linear calibration cannot be achieved, and other signal normalization procedures are needed for accurate calibration.[67]

When internal standardization procedures are used for analyte characterization in an unknown sample, Equation (2) is applied:

$$C_{sa}^{M} = C_{sa}^{IS} \times \frac{I_{sa}^{M}}{I_{sa}^{IS}} \times \frac{C_{st}^{M}}{C_{st}^{IS}} \times \frac{I_{st}^{IS}}{I_{st}^{M}} \qquad (2)$$

where C is the concentration and I is the ICP signal intensity. Subscripts sa and st denote sample and standard and superscripts M and IS correspond to the unknown analyte and the internal standard, respectively. This relationship assumes that the concentration of the internal standard is known and that fractionation between the analyte and internal standard is not significant, or is the same for analyte and standards.

3.5.2 Calibration with Liquid Standards

For many samples, matrix-matched standards will not be available. In these cases, external calibration can be performed using nebulization of liquid standards; a dual sample introduction method can be used to establish instrumental calibration with a series of standard solutions. To analyze the unknown sample by LA using liquid standards, wet ICP plasma conditions must be maintained.[68] However, the use of liquids for calibration defeats two important advantages of LA and dry ICP: elimination of isobaric interferences in ICPMS and the generation of solvent waste.[12] In addition, optimum ICP conditions will not be the same for LA and solution nebulization. For nebulization, the analyte dries from liquid droplets to form small particles in the plasma. For LA sampling, larger dry particles are introduced directly into the ICP.[45,48] Atomization and excitation processes in the ICP are expected to be different for these two cases. Even if the water content is the same in the ICP, it is still possible that the excitation characteristics will be different because of the different vaporization mechanisms for nebulized solution and laser-ablated particles. On the other hand, calibration with dried solution aerosol[69] or by direct LA of liquids[70,71] does not significantly perturb dry ICP conditions.

LA of liquids has been shown to be effective for the analysis of microscopic fluid inclusions in minerals. In this method, direct ablation of a standard solution was used for external ICPMS calibration with Na as an internal standard.[71] In another study, instrument calibration was done by using artificial fluid inclusions prepared by drawing a standard solution into a microcapillary tube[72] or with microwells containing aqueous solutions.[73]

3.5.3 External Standardization

For some samples, internal standardization may not be a viable option. In such cases, external standardization must be used to compensate for differences in ablation of the sample versus the standards. Several external

procedures have been demonstrated to compensate for changes in the quantity of ablated mass. They include light scattering,[48,74,75] acoustic emission in the sample[19,76,77] or in the ambient medium,[76,78] the use of a mass monitor to collect a portion of the laser-ablated aerosol,[79] and the measurement of spectral emission intensity in the ICP and laser-induced plasma simultaneously.[34,77] An absolute method to quantify the amount of mass ablated is to weigh the sample before and after ablation.[80] However, such direct mass measurements are tedious and may not be accurate because of the small quantity (<1 μg) of mass ablated for each laser pulse, and because transport efficiency is not included.[22,45] Determination of mass ablation and transport efficiencies is possible by collecting particles on a nonporous filter, which is an indication of the total mass transported to the ICP.

4 APPLICATIONS

Because of the unique properties of the laser, many novel applications for chemical analysis using the ICP have been demonstrated. This section presents a brief overview of applications that would be difficult or impossible to perform without the use of LA. Most of the applications utilize ICPMS because of its enhanced sensitivity.

4.1 Environmental and Oceanography

4.1.1 Tree Rings

Spatial patterns in the chemical content of tree rings can be used as a monitor for changes in atmospheric conditions, changes in soil chemistry, and the pollution history of an area. Laser ablation inductively coupled plasma mass spectrometry (LA/ICPMS) is an excellent approach to determine chemical content in tree samples because of the high spatial resolution provided by the focused laser beam, and the excellent sensitivity of ICPMS to measure very low detection levels for many elements.[81–84] Watmough et al.[83] obtained the quantitative multielement analysis of 11 elements in red maple tree rings; significant changes in these elements were measured for trees grown in contaminated soils adjacent to a metal smelter versus trees grown in unpolluted soils. GarbeSchongberg et al.[84] analyzed pine and birch tree rings from Norway and Russia and discussed the elemental relationship with the pollution history of these areas.

4.1.2 Sea Shells

Trace-element fluctuations in sea shells represent environmental changes and major pollution events. In the work of Raith et al.,[85] the inner to the outer walls of a shell were analyzed; the elemental changes between growth bands of the shell showed a history of heavy-metal pollution over the years. Vander Putten et al.[86] measured spatial variations of Mg, Mn, Sr, Ba, and Pb in the calcite layer of *Mytilus edulis* sea shells.

4.1.3 Airborne Particulates

The analysis of trace elements in airborne particulates provides unique signature information for monitoring air quality and air pollution. For example, arsenic, considered the major ''marker element'' of air pollution, is one of the most hazardous anthropogenic air pollutants affecting humans globally. Traditionally, membrane filters with small pore size have been used for collecting particulate samples. The entire filter is then digested and the total volume analyzed. LA is a perfect alternative for such analyses; the particulates can be ablated directly from the filter media. Tanaka et al.[87] and Wang et al.[88–90] analyzed airborne particulates for more than 20 major, minor, and trace elements using this approach.

4.1.4 High-resolution Analysis of Coral

The concentration of trace elements in coral skeletons can provide information about changes in seawater properties; calcification in reef-building corals is strongly affected by environmental factors such as temperature, light, water motion, and pollution.[91] As coral grows, it secretes a calcareous skeleton into which trace elements are partitioned from the ambient seawater. Spatial analysis of the coral skeleton allows a detailed investigation of seasonal composition changes.[92,93] Compared with the traditional method of sample milling, with processing and analysis by solution nebulization, LA/ICPMS provides in situ analysis of corals with spatial resolution less than 20 μm. Sinclair et al.[92] analyzed corals collected from Australia and showed that the elements B, Mg, Sr, and U exhibited seasonal variations, as shown in Figure 7. These fluctuations coincided with the changes of sea-surface temperature.

4.2 Geochemistry and Cosmochemistry

4.2.1 Uranium–Lead Geochronology

Zircon U–Pb geochronology is one of the principal dating tools used in the earth sciences; ages are calculated by measuring $^{206}Pb/^{238}U$, $^{207}Pb/^{235}U$, and $^{207}Pb/^{206}Pb$ ratios. The conventional method for U–Pb isotopic analysis has been by thermal ionization mass spectrometry (TIMS) with chemical separation. However, very low blank values are required and the data represent the ''average'' of the bulk grains. LA/ICPMS has the ability to perform spatially resolved in situ analysis of U–Pb

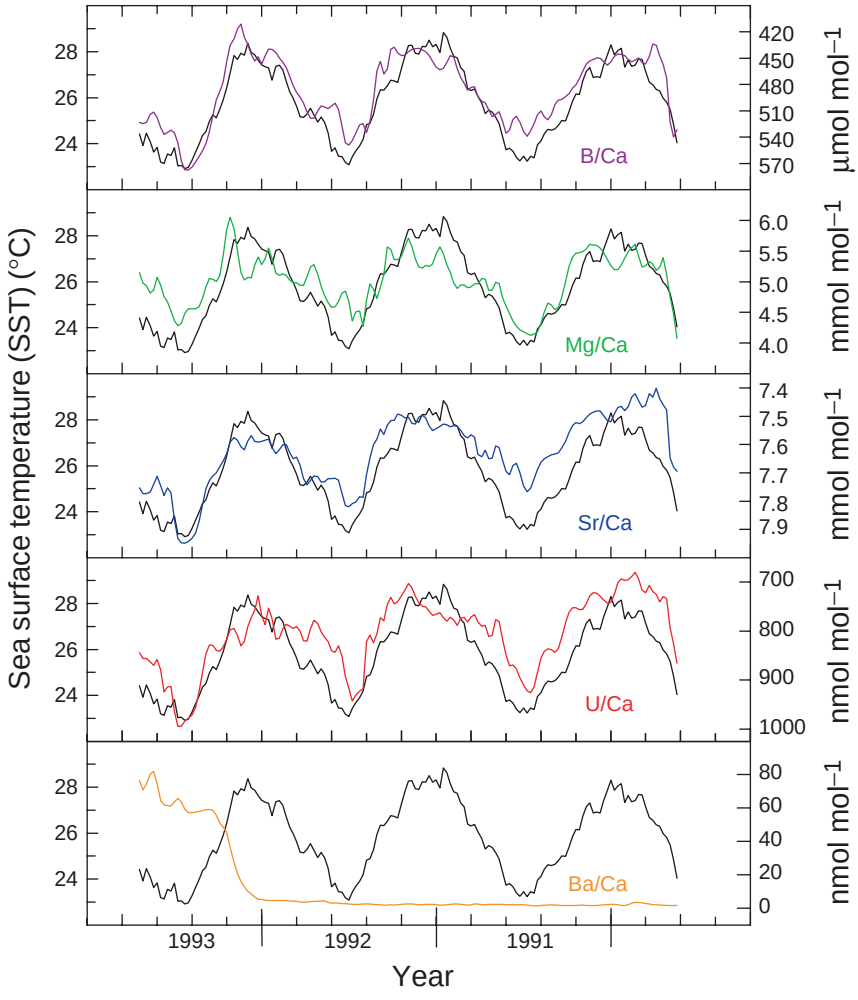

Figure 7 Trace element and sea surface temperature (SST) profiles for coral from Australia's Great Barrier Reef. Colored lines represent an average of three trace element profiles taken over the same coral track. Black lines are the instrumental SST data taken from the weather station. B, Mg, Sr, and U all display clear seasonal variation. Ba does not show a seasonality but displays a marked enrichment in the tissue zone. (Figure reproduced from original data with permission by Dr Daniel J. Sinclair[92].)

isotopic compositions in zircons and similar minerals. Several groups[94–97] have studied zircon and monazite samples using LA/ICPMS. Most of these studies only show $^{207}Pb/^{206}Pb$ data and a few include the $^{206}Pb/^{238}U$ ratio, because of a fractionation problem. Without the $^{206}Pb/^{238}U$ ratio, the analysis cannot be extended to "young" (<600 Ma) zircon dating. In order to minimize Pb–U fractionation, methods such as active focusing, line scanning, and soft ablation (increasing the laser power when ablation is progressing) were attempted. Solving Pb–U fractionation is necessary to the success of LA/ICPMS for this application.

4.2.2 *Inclusion Analysis*

Microscopic inclusions in minerals contain direct evidence of the composition of fluids associated with large-scale material transport in the Earth's interior. Quantitative knowledge of the elements and isotopic composition of these fluid inclusions is a prerequisite for understanding and modeling fluid–rock interactions. Detailed chemical information is difficult to obtain because of the very small size of these inclusions (typically 10^{-11}–10^{-9} g).[70] Crush–leach analysis or bulk analysis of quartz containing fluid inclusions can provide concentration ratios averaged over many inclusions, but most samples contain multiple generations of fluid inclusions of different compositions. LA/ICPMS for the analysis of individual microscopic inclusion has, therefore, received considerable attention. Good quantitative results have been published mainly for synthetic fluid inclusions,[73,98] where a heavy trace element of a known concentration (e.g. Sr, U) was added as an internal standard. Günther et al.[70] reported a method for measuring complex polyphase inclusions

using a stepwise opening procedure. A series of inclusions representing the fluid before, during, and after the deposition of cassiterite (SnO_2) in a tin deposit in Australia were analyzed by this method; physical and chemical mechanisms of ore precipitation were proposed based on these data.[99]

4.2.3 Precise In Situ Analysis of Hafnium, Tungsten, Strontium, Lead, and Osmium Isotopes

Many elements are of significant interest for isotope geochemistry and need to be measured at trace levels with excellent precision. Multiple collector inductively coupled plasma mass spectrometry (MC/ICPMS) with a magnetic sector is a new technology for the measurement of isotopic compositions with very high precision. It is particularly suited for elements with high first ionization potentials such as Hf, W, and Os, which cannot be measured with good precision using conventional TIMS. Combined with LA sampling, in situ isotopic measurements at the microscopic scale are possible. Although this technology is still in its infancy, diverse applications have already led to a number of important scientific developments.[100]

The initial Hf isotopic composition is more reliable than initial Nd as a geochemical tracer owing to the immobility of Hf. The Hf isotopic composition in zircon samples was analyzed by Thirlwall and Walder.[101] In low-temperature geochemistry, Hf isotopes may provide the most reliable and sensitive isotopic proxy for hydrothermal activity in the ocean. The Hf isotopic compositions in iron–manganese nodules and crusts were studied by Godfrey et al.[102] and the data reflected the concomitant growth from seawater and pore fluids. There is strong interest in measuring W and Hf isotopic compositions with high precision using LA with MC/ICPMS. The $^{182}Hf-^{182}W$ system is a method for constraining timescales of accretion and metal–silicate differentiation in planets; the age of the Earth's core, the Moon, and Mars have been measured using this new chronometer.[100]

Sr and Pb isotopic compositions are significant for geochemistry and oceanography research. Analysis of these elements traditionally requires complicated chemical separation procedures. LA with MC/ICPMS provided accurate and precise measurement of $^{87}Sr/^{86}Sr$ isotopic ratios in geological materials.[103] Pb isotopic compositions in a ferromanganese crust from the Pacific ocean were analyzed by Christensen et al.[104] and the authors concluded that the Pb isotopic data could be used to probe climate-driven changes in ocean circulation. Analysis of Os isotopic ratios for iridosmine samples was reported using LA with MC/ICPMS.[105] The mineral iridosmine has been used for the definition of the Os isotopic evolution of the mantle.

4.2.4 Bulk and Microbeam Analysis of Rocks and Minerals

Bulk and spatial analyses of rocks and minerals for elemental and isotopic compositions can provide fundamental information to help solve diverse geological and environmental problems. The application of LA/ICPMS for the analysis of whole-rock geological samples, such as pressed power pellets and lithium metaborate fusions, has been described by several authors.[106,107] In most of these studies, the measurement of rare-earth elements (Zr, Hf, U, Th, Sr Rb, Ba, Nb, Ta) was emphasized. LA/ICPMS was also used to determine the platinum group elements (Ru, Rh, Pd, Os, Ir and Pt) and gold, which have very low natural abundance but great economic and geological importance.[108]

The spatially resolved analysis of elements within minerals provides crystal-growth information and the variation in the physical and chemical nature of environments in which they grew. LA/ICPMS analysis, either on a single mineral grain or on a thin section, can provide such information. LA/ICPMS analyses of minerals such as calcite, zircon, olivine, plagioclase, feldspar, titanite, apatite, clinopyroxence, amphibole, and garnet have been reported.[109–111]

4.3 Forensics and Authentication

4.3.1 Authentication of Antique Objects

Chemical analysis is an excellent approach for verifying the authenticity of valuable artifacts. Obviously, the analytical method for authenticity verification of precious antiques should be either nondestructive or require extremely small sample quantities. Using LA/ICPMS, visible damage can be restricted to an acceptable minimum. Such studies were reported by Devos et al.,[112] in which a specially designed sample chamber was used, and almost invisible 100-μm craters were produced on antique silver objects. The elements Zn, Cd, Sn, Sb, Au, Pb, and Bi were measured, and their contents were used to distinguish forgery in silver antiques. Similar studies were reported by Wanner et al.[113] for trace element analyses of archaeological samples such as ancient coins, various antique silver items, and ancient iron. Owing to surface roughness of the samples, an autofocus system was used to achieve reproducible ablation conditions. In these experiments, a lateral resolution of 50 μm and absolute detection limits of 1–1.4 pg were achieved.

4.3.2 Fingerprinting Crime Scene Evidence

Many criminal activities result in the generation of debris or other materials, which become available to investigating authorities as physical evidence of the crime. However, the generation of traditional analytical and

forensic chemical data is often costly and time-consuming. LA/ICPMS offers the potential for producing fast, definitive, and cost-effective forensic chemical analysis for use in identifying physical evidence that relates a suspect to the scene of a crime. Watling et al.[114] examined several kinds of glass and steel samples as physical evidence of a crime. They developed software to facilitate an intercomparison of three elements simultaneously (ternary plots) for large groups of samples. This approach established both the reproducibility of the ''fingerprint'' and the uniqueness of interelement associations. A similar idea was used to source the provenance of cannabis. Certain elements from a specific area or geological environment characterize uniquely the source of the plant. Watling[115] showed the trace element association ''fingerprint'' patterns of cannabis crops and the potential tracing of these crops to specific geological environments. The association of elements formed the basis for determining the provenance of cannabis crops and samples recovered during police drug raids.

4.4 Waste-sample Analysis

LA has many advantages when used to analyze radiological contaminated samples. For example, organic solvents or concentrated acids that are required for classical separation procedures will not be needed for LA sampling. Also, with LA, much less total sample ($<1\,\mu g$) will be required, greatly reducing the risks associated with sample handling and contamination, and personnel exposure. Finally, elemental and isotopic analysis can be obtained entirely within a hot cell environment, further reducing the risk of contamination.[116] High-level waste analysis using LA/ICPMS was detailed by Smith et al.[117] A unique LA facility has been established in a hot cell environment at the Hanford Site for direct characterization of tank waste samples. Applications of LA and high-resolution ICPAES in the nuclear industry, especially high-resolution isotopic analysis of U and determination of lanthanides, were reported by Giglio et al.[116] and Zamzow et al.[118]

5 NONINDUCTIVELY COUPLED PLASMA APPLICATIONS

There are many LA applications that do not rely on particle digestion by an analytical source (ICP, flame, etc.), but instead directly measure the ablated atomic mass. Laser-induced breakdown spectroscopy (LIBS), described in more detail in the article **Laser-induced Breakdown Spectroscopy**, involves monitoring spectroscopic emission intensity directly in the laser plume. Other applications involve direct MS detection of the ablated mass, with and without secondary ionization, e.g. resonance ionization mass spectrometry (RIMS) and matrix-assisted laser desorption/ionization (MALDI). It would be impossible to discuss these numerous applications without doubling the length of this article. Instead, one particular technique employed by our research group is presented.

5.1 Laser Ablation Ion-storage Time-of-flight Mass Spectrometry

A new technique for single particle analysis was developed that uses ion-trap and TOF mass spectrometers. Commercial LA systems with imaging capabilities involve a chamber in which the ablated mass is entrained into a gas stream and transported to the analytical source by several meters of tubing, thereby diluting the mass vapor and influencing detection sensitivity. For single-particle analysis, excellent sensitivity must be available because of the limited absolute mass from a micrometer-sized particle. To overcome the dilution and sensitivity limitations, an instrument was developed using LA inside an ion-storage time-of-flight mass spectrometry (IS/TOFMS) system with an imaging camera to observe, select, vaporize, ionize, and analyze individual particles or spatial locations on a solid surface.

An ion trap was designed such that a pulsed Nd : YAG laser beam could ablate a sample affixed to a probe inserted radially through the ring electrode (Figure 8). Once the particle is selected using the imaging system, it is directly ablated/ionized by the laser beam. A primary advantage of this system is that the ions are generated directly inside the ion trap. The ion trap is not used as the mass spectrometer, because space-charge effects would critically influence the mass resolution. Instead, the ion trap is used only as a storage device and the TOF performs the MS. In this way, space charging is significantly reduced except in cases when using extremely high laser fluence, which is not necessary for excellent sensitivity. In preliminary experiments, trace contaminants of Ag, Sn, and Sb were measured in a Pb target with a single laser shot. Reproducible spectra could be measured with only approximately $10\,\mu J$ of laser energy. The laser-beam focusing system provided a spatial resolution of approximately $12\,\mu m$, with imaging capability to guide the laser beam to a specific location on the sample. The photographs in Figure 9(a) and (b) shows images of the laser-induced craters in hair and alloy samples, respectively.

6 PERSPECTIVES AND FUTURE TRENDS

LA and atomic spectroscopy have become a mature partnership providing significant benefits to both techniques. On the one hand, LA provides the best method

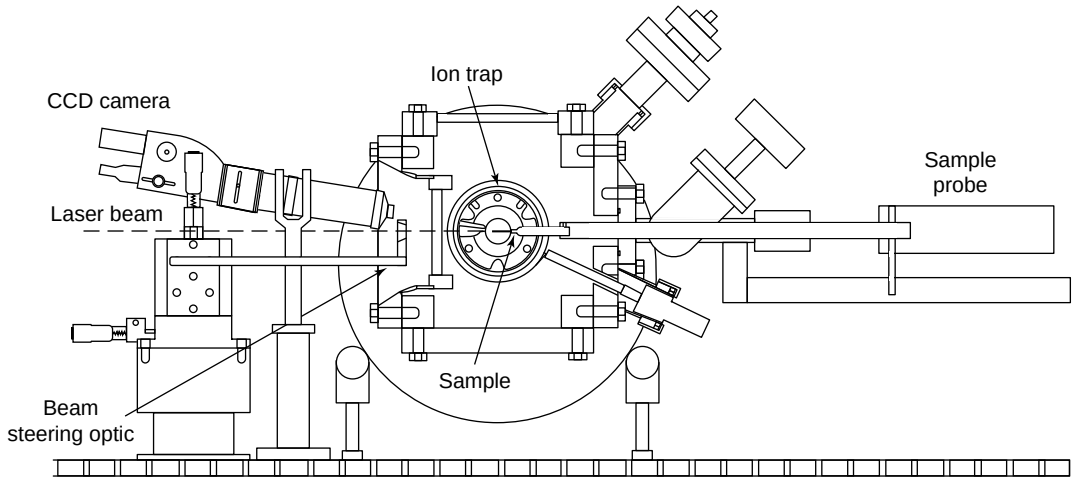

Figure 8 Schematic diagram of ion-trap TOF mass spectrometer with imaging capabilities. The system was designed such that a pulsed Nd : YAG laser beam could ablate a selected sample inside the ion trap mass spectrometer.

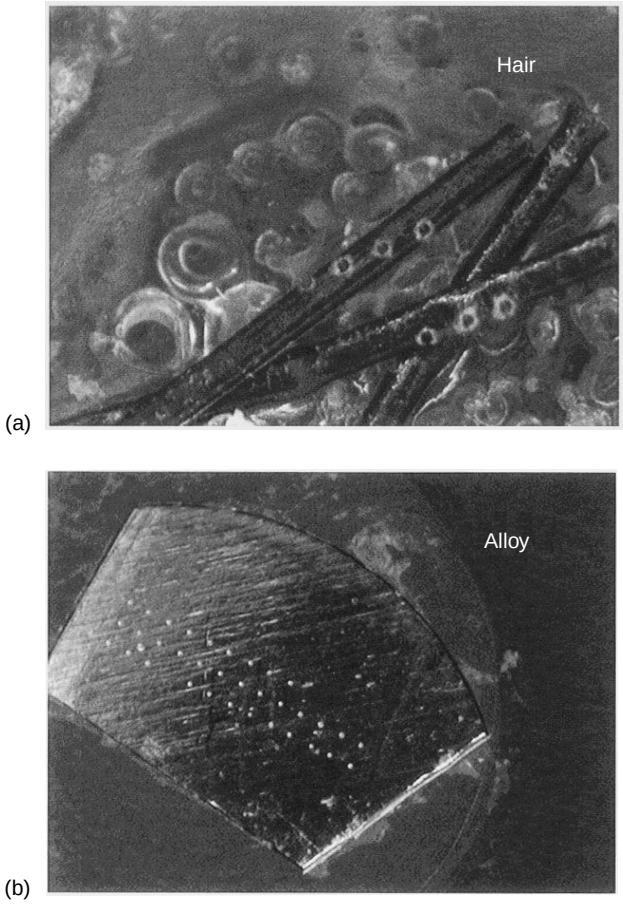

Figure 9 Images of the laser-induced craters in (a) hair and (b) alloy samples.

for directly converting a solid-phase sample into gas-phase constituents. On the other hand, atomic spectroscopy brings to the relationship a powerful method for understanding the fundamental mechanisms of ablation processes. For atomic spectroscopy, the relationship is continuing to flourish, as can be seen by the increasing number of applications, number of published papers and conference symposia, and new commercial systems. This article has primarily addressed LA as a sampling technique for atomic spectroscopy. There are numerous papers in the literature in which atomic spectroscopy is used to study the fundamental behavior of ablation processes; an example is the work by Bushaw and Alexander using high-resolution time-resolved atomic absorption spectroscopy for the investigation of LA plume dynamics.[119] A literature search using the two keywords ''laser'' and ''ablation'' provided approximately 2500 published papers in the *Current Contents* database, increasing continuously from 1990 to the present. Interest in LA continues to flourish because of applications in atomic spectroscopy, and also medical, semiconductor, materials, and other areas. When matrix-matched standards are available, LA is an ideal quantitative method with excellent accuracy and precision. In cases when matrix-matched standards are not available, LA with atomic spectroscopy is suitable for semiquantitative analysis. Further development of LA and its maturation into an accurate and precise sampling technique for atomic spectroscopy will occur through understanding of fundamental processes and the means to control them. The benefits of LA sampling and atomic spectroscopy warrant success.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the efforts of the Lawrence Berkeley National Laboratory (LBNL)

researchers whose papers are referenced in this work, with special thanks to Sungho Jeong, Sam Mao, George Chan, and Jong Yoo for contributing data. R.E.R. also acknowledges G.L. Klunder, P.G. Grant, and B.D. Andresen for the work on IS/TOFMS at Lawrence Livermore National Laboratory (LLNL). This work was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, and by the Environmental Waste Management Science Program (EMSP) under a joint grant from the Office of Energy Research and Office of Environmental Management, through the LBNL under Contract No. DE-AC03-76SF00098.

ABBREVIATIONS AND ACRONYMS

AES	Atomic Emission Spectroscopy
CCD	Charge-coupled Device
CTD	Charge Transfer Device
DCP	Direct Current Plasma
DSI	Direct Sample Insertion
EMSP	Environmental Waste Management Science Program
ETV	Electrothermal Vaporization
GD	Glow Discharge
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
IS/TOFMS	Ion-storage Time-of-flight Mass Spectrometry
LA	Laser Ablation
LA/ICP	Laser Ablation/Inductively Coupled Plasma
LA/ICPMS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
LBNL	Lawrence Berkeley National Laboratory
LIBS	Laser-induced Breakdown Spectroscopy
LLNL	Lawrence Livermore National Laboratory
LOD	Limit of Detection
MALDI	Matrix-assisted Laser Desorption/Ionization
MC/ICPMS	Multiple Collector Inductively Coupled Plasma Mass Spectrometry
MIP	Microwave-induced Plasma
MS	Mass Spectrometry
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
PDA	Photodiode Array
PMT	Photomultiplier Tube
RIMS	Resonance Ionization Mass Spectrometry
SST	Sea Surface Temperature
TIMS	Thermal Ionization Mass Spectrometry
TOF	Time-of-flight
UV	Ultraviolet
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Laser-induced Breakdown Spectroscopy, Elemental Analysis • Laser Mass Spectrometry in Trace Analysis • Laser- and Optical-based Techniques for the Detection of Explosives

Environment: Water and Waste **(Volume 3)**
Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis • Slurry Sampling Graphite Furnace Atomic Absorption Spectrometry in Environmental Analyses • Soil Sampling for the Characterization of Hazardous Waste Sites

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Spectrometric Techniques in Analytical Atomic Spectrometry • Laser-induced Breakdown Spectroscopy

REFERENCES

1. S.A. Darke, J.F. Tyson, 'Interaction of Laser Radiation with Solid Materials and Its Significance to Analytical Spectrometry', *J. Anal. At. Spectrom.*, **8**, 145–209 (1993).
2. A. Ciocan, J. Uebbing, K. Niemax, 'Analytical Applications of the Microwave Induced Plasma Used with Laser Ablation of Solid Samples', *Spectrochim. Acta, Part B*, **47**(5), 611–617 (1992).
3. P.G. Mitchell, J. Sneddon, L.J. Radziemski, 'Direct Determination of Copper in Solids by Direct Current Argon Plasma Emission Spectrometry with Sample Introduction Using Laser Ablation', *Appl. Spectrosc.*, **41**(1), 141–148 (1987).

4. A. Montaser, 'Assessment of the Potentials and Limitations of Plasma Sources Compared to ICP Discharges for Analytical Spectrometry', in *Inductively Coupled Plasmas in Atomic Spectrometry*, 2nd edition, eds. A. Montaser, D.W. Golightly, VCH, New York, 1–8, 1992.
5. S.A. Darke, J.F. Tyson, 'Review of Solid Sample Introduction for Plasma Spectrometry and a Comparison of Results for Laser Ablation, Electrothermal Vaporization, and Slurry Nebulization', *Microchem. J.*, **50**, 310–336 (1994).
6. R. Sing, 'Direct Sample Insertion for Inductively Coupled Plasma Spectrometry', *Spectrochim. Acta, Part B*, **54**, 411–441 (1999).
7. R.E. Russo, 'Laser Ablation', *Appl. Spectrosc.*, **49**, A14–A28 (1995).
8. L.A. Allen, H.M. Pang, A.R. Warren, R.S. Houk, 'Simultaneous Measurement of Isotope Ratios in Solids by Laser Ablation with a Twin Quadrupole Inductively Coupled Plasma Mass Spectrometer', *J. Anal. At. Spectrom.*, **10**(3), 267–271 (1995).
9. W.-T. Chan, X.L. Mao, R.E. Russo, 'Differential Vaporization During Laser Ablation/Deposition of Bi–Sr–Ca–Cu–O Superconducting Materials', *Appl. Spectrosc.*, **46**(6), 1025–1031 (1992).
10. X.L. Mao, W.T. Chan, M. Caetano, M.A. Shannon, R.E. Russo, 'Preferential Vaporization and Plasma Shielding during Nanosecond and Picosecond Laser Ablation', *Appl. Surf. Sci.*, **96–98**, 126–130 (1996).
11. E.F. Cromwell, P. Arrowsmith, 'Fractionation Effects in Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **49**(11), 1652–1660 (1995).
12. P.M. Outridge, W. Doherty, D.C. Gregoire, 'The Formation of Trace Element-enriched Particulates During Laser Ablation of Refractory Materials', *Spectrochim. Acta, Part B*, **51**(12), 1451–1462 (1996).
13. S.M. Eggins, L.P.J. Kinsley, J.M.G. Shelley, 'Deposition and Element Fractionation Processes During Atmospheric Pressure Laser Sampling for Analysis by ICPMS', *Appl. Surf. Sci.*, **127–129**, 278–286 (1998).
14. X.L. Mao, A.C. Ciocan, R.E. Russo, 'Preferential Vaporization During Laser Ablation Inductively Coupled Plasma/Atomic Emission Spectroscopy', *Appl. Spectrosc.*, **52**, 913–918 (1998).
15. T.E. Jeffries, N.J.G. Pearce, W.T. Perkins, A. Raith, 'Chemical Fractionation During Infrared and Ultraviolet Laser Ablation Inductively Coupled Plasma Mass Spectrometry – Implications for Mineral Microanalysis', *Anal. Commun.*, **33**(1), 35–39 (1996).
16. D. Figg, M.S. Kahr, 'Elemental Fractionation of Glass Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **51**(8), 1185–1192 (1997).
17. V. Kanicky, J. Musil, J.-M. Mermet, 'Determination of Zr and Ti in 3-μm-thick ZrTiN Ceramic Coating Using Laser Ablation Inductively Coupled Plasma Atomic Emission Spectroscopy', *Appl. Spectrosc.*, **51**(7), 1037–1041 (1997).
18. R.E. Russo, X.L. Mao, W.-T. Chan, M.F. Bryant, W.F. Kinard, 'Laser Ablation Sampling with Inductively Coupled Plasma Atomic Emission Spectrometry for the Analysis of Prototypical Glasses', *J. Anal. At. Spectrom.*, **10**(4), 295–301 (1995).
19. M.A. Shannon, X.L. Mao, A. Fernandez, W.-T. Chan, R.E. Russo, 'Laser Ablation Mass Removal Versus Incident Power Density During Solid Sampling for Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Chem.*, **67**(24), 4522–4529 (1995).
20. S. Witanachchi, K. Ahmed, P. Sakthivel, P. Mukherjee, 'Dual-laser Ablation for Particulate-free Film Growth', *Appl. Phys. Lett.*, **66**(12), 1469–1471 (1995).
21. P. Goodall, S.G. Johnson, E. Wood, 'Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry of a Uranium–Zirconium Alloy – Ablation Properties and Analytical Behavior', *Spectrochim. Acta, Part B*, **50**(14), 1823–1835 (1995).
22. D.J. Figg, J.B. Cross, C. Brink, 'More Investigations into Elemental Fractionation Resulting from Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry on Glass Samples', *Appl. Surf. Sci.*, **127–129**, 287–291 (1998).
23. C. Geertsen, A. Briand, F. Chartier, J.-L. Lacour, P. Mauchien, S. Sjostrom, J.-M. Mermet, 'Comparison Between Infrared and Ultraviolet Laser Ablation at Atmospheric Pressure – Implications for Solid Sampling Inductively Coupled Plasma Spectrometry', *J. Anal. At. Spectrom.*, **9**(1), 17–22 (1994).
24. W. Sdorra, J. Brust, K. Niemax, 'Basic Investigations for Laser Microanalysis. 4. The Dependence on the Laser Wavelength in Laser Ablation', *Mikrochim. Acta*, **108**(1–2), 1–10 (1992).
25. Y.-I. Lee, K. Song, H.-K. Cha, J.-M. Lee, M.-C. Park, G.-H. Lee, J. Sneddon, 'Influence of Atmosphere and Irradiation Wavelength on Copper Plasma Emission Induced by Excimer and Q-switched Nd:YAG Laser Ablation', *Appl. Spectrosc.*, **51**(7), 959–964 (1997).
26. A. Dupont, P. Caminat, P. Bournot, J.P. Gauchon, 'Enhancement of Material Ablation Using 248, 308, 532, 1064 nm Laser Pulse with a Water Film on the Treated Surface', *J. Appl. Phys.*, **78**(3), 2022–2028 (1995).
27. S. Nakamura, K. Midorikawa, H. Kumagai, M. Obara, K. Toyoda, 'Effect of Pulse Duration on Ablation Characteristics of Tetrafluoroethylene Hexafluoropropylene Copolymer Film Using Ti-sapphire Laser', *Jpn. J. Appl. Phys., Part 1*, **35**(1A), 101–106 (1996).
28. W. Sdorra, K. Niemax, 'Basic Investigations for Laser Microanalysis. 3. Application of Different Buffer Gases for Laser-produced Sample Plumes', *Mikrochimica Acta*, **107**(3–6), 319–327 (1992).
29. Y.-I. Lee, T.L. Thiem, G.-H. Kim, Y.-Y. Teng, J. Sneddon, 'Interaction of an Excimer-laser Beam with Metals. 3. The Effect of a Controlled Atmosphere

in Laser-Ablated Plasma Emission', *Appl. Spectrosc.*, **46**(11), 1597–1604 (1992).
30. Y. Iida, 'Effects of Atmosphere on Laser Vaporization and Excitation Processes of Solid Samples', *Spectrochim. Acta, Part B*, **45**(12), 1353–1367 (1990).
31. X.L. Mao, W.T. Chan, M.A. Shannon, R.E. Russo, 'Influence of Gas Medium on Plasma Shielding During Laser–Material Interactions in the Picosecond Time Regime', *J. Appl. Phys.*, **74**, 4915–4922 (1993).
32. A. Semerok, C. Chaleard, V. Detalle, J.L. Lacour, P. Mauchien, P. Meynadier, C. Nouvellon, B. Salle, P. Palianov, M. Perdrix, G. Petite, 'Experimental Investigations of Laser Ablation Efficiency of Pure Metals with Femto-, Pico- and Nanosecond Pulses', *Appl. Surf. Sci.*, **139**, 311–314 (1999).
33. C. Momma, B.N. Chichkov, S. Nolte, F. Vonalvensleben, A. Tunnermann, H. Welling, B. Wellegehausen, 'Short-pulse Laser Ablation of Solid Targets', *Opt. Commun.*, **129**(1–2), 134–142 (1996).
34. A. Fernandez, X.L. Mao, W.-T. Chan, M.A. Shannon, R.E. Russo, 'Correlation of Spectral Emission Intensity in the Inductively Coupled Plasma and Laser-induced Plasma During Laser Ablation of Solid Samples', *Anal. Chem.*, **67**(14), 2444–2450 (1995).
35. M. von Allmen, 'Evaporation and Plasma Formation', in *Laser-beam Interactions with Materials – Physical Principles and Applications*, 1st edition, ed. A. Mooradian, Springer, Berlin, 193–197, 1987.
36. J.H. Yoo, S.H. Jeong, R. Greif, R.E. Russo, 'Explosive Growth of Crater Volume during High Power Nanosecond Laser Ablation of Silicon', *J. Appl. Phys.*, to be published.
37. V.P. Carey, 'Phase Stability and Homogeneous Nucleation', in *Liquid–Vapor Phase Change Phenomena: An Introduction to the Thermophysics of Vaporization and Condensation Processes in Heat Transfer Equipment*, ed. V.P. Carey, Hemisphere, Washington, DC, 127, 1992.
38. M. Caetano, X.L. Mao, R.E. Russo, 'Laser Ablation Solid Sampling: Vertical Spatial Emission Intensity Profiles in the Inductively Coupled Plasma', *Spectrochim. Acta, Part B*, **51**, 1473–1485 (1996).
39. X.L. Mao, O.V. Borisov, R.E. Russo, 'Enhancements in Laser Ablation Inductively Coupled Plasma/Atomic Emission Spectrometry Based on Laser Properties and Ambient Environment', *Spectrochim. Acta, Part B*, **53**(5), 731–739 (1998).
40. A.M. Prokhorov, V.I. Konov, I. Ursu, I.N. Mihilescu, in *Laser Heating of Metals*, Adam Hilger, New York, 42, 1990.
41. N. Bloembergen, 'Fundamentals of Laser–Solid Interactions', in *Laser–Solid Interactions and Laser Processing – 1978*, eds. S.D. Ferris, H.J. Leamy, J.M. Poate, American Institute of Physics, New York, 1–9, 1979.
42. J.F. Ready, 'Gas Breakdown', in *Effects of High-power Laser Radiation*, Academic Press, New York, 213–272, 1971.
43. A.P.K. Leung, W.T. Chan, X.L. Mao, R.E. Russo, 'Influence of Gas Environment on Picosecond Laser Ablation Sampling Efficiency and ICP Conditions', *Anal. Chem.*, **70**(22), 4709–4716 (1998).
44. L. Moenke-Blankenburg, 'Laser–ICP Spectrometry', *Spectrochim. Acta Rev.*, **15**(1), 1–37 (1993).
45. P. Arrowsmith, S.K. Hughes, 'Entrainment and Transport of Laser Ablated Plumes for Subsequent Elemental Analysis', *Appl. Spectrosc.*, **42**(7), 1231–1239 (1988).
46. G. Su, S. Lin, 'Studies on the Complete Laser Vaporization of the Powdered Solid Samples into an Inductively Coupled Plasma for Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **3**, 841 (1988).
47. S.H. Jeong, O.V. Borisov, J.H. Yoo, X.L. Mao, R.E. Russo, 'Effects of Particle Size Distribution on Inductively Coupled Plasma Mass Spectrometry Signal Intensity During Laser Ablation of Glass Samples', *Anal. Chem.*, **71**, 5123–5130 (1999).
48. M.L. Alexander, M.R. Smith, J.S. Hartman, A. Mendoza, D.W. Koppenaal, 'Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Appl. Surf. Sci.*, **127–129**, 255–261 (1998).
49. A.A. van Heuzen, 'Analysis of Solids by Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LA/ICPMS) – I. Matching with Glass Matrix', *Spectrochim. Acta, Part B*, **46**(14), 1803–1817 (1991).
50. M. Thompson, S. Chenery, L. Brett, 'Nature of Particulate Matter Produced by Laser Ablation – Implications for Tandem Analytical Systems', *J. Anal. At. Spectrom.*, **5**(1), 49–55 (1990).
51. K.N. DeSilva, R. Guevremont, 'Direct Powder Introduction Inductively Coupled Plasma Atomic Emission Spectrometry with a Photodiode Array Spectrometer', *Spectrochim. Acta, Part B*, **46**(11), 1499–1515 (1991).
52. T.E. Jeffries, S.E. Jackson, H.P. Longerich, 'Application of a Frequency Quintupled Nd:YAG Source ($\lambda =$ 213 nm) for Laser Ablation Inductively Coupled Plasma Mass Spectrometric Analysis of Minerals', *J. Anal. At. Spectrom.*, **13**, 935–940 (1998).
53. D. Günther, R. Frischknecht, C.A. Heinrich, H.-J. Kahlert, 'Capabilities of an Argon Fluoride 193 nm Excimer Laser for Laser Ablation Inductively Coupled Plasma Mass Spectrometry Microanalysis of Geological Materials', *J. Anal. At. Spectrom.*, **12**, 939–944 (1997).
54. P.P. Mahoney, G. Li, G.M. Hieftje, 'Laser Ablation/Inductively Coupled Plasma Mass Spectrometry with a Time-of-flight Mass Analyzer', *J. Anal. At. Spectrom.*, **11**, 401–405 (1996).
55. T. Mochizuki, A. Sakashita, H. Iwata, 'Laser Ablation for Direct Elemental Analysis of Solid Samples by ICP–Atomic Emission Spectrometry and ICP–Mass Spectrometry', *Nippon Kagaku Kaishi*, **58**, 19–27 (1990).
56. X.R. Liu, G. Horlick, 'In Situ Laser Ablation Sampling for Inductively Coupled Plasma Atomic Emission

Spectroscopy', *Spectrochim. Acta, Part B*, **50**(4–7), 537–548 (1994).
57. M. Thompson, R.M. Barnes, 'Analytical Performance of Inductively Coupled Plasma/Atomic Emission Spectrometry', in *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*, 2nd edition, eds. A. Montaser, D.W. Golighty, VCH, New York, 276, 1992.
58. M. Motelica-Heino, O.F.X. Donard, J.-M. Mermet, 'Laser Ablation of Synthetic Geological Powders Using ICP/AES Detection: Effects of the Matrix, Chemical Form of the Analyte and Laser Wavelength', *J. Anal. At. Spectrom.*, **14**, 675–682 (1999).
59. M. Ducreux-Zappa, J.-M. Mermet, 'Analysis of Glass by UV Ablation Inductively Coupled Plasma Atomic Emission Spectrometry. Part 2. Analytical Figures of Merit', *Spectrochim. Acta, Part B*, **51**, 333–341 (1996).
60. A.A. van Heuzen, J.B. Morsink, 'Analysis of Solids by Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LA/ICPMS) – II. Matching with a Pressed Pellet', *Spectrochim. Acta, Part B*, **46**(14), 1819–1828 (1991).
61. K.E. Jarvis, J.G. Williams, S.J. Parry, E. Bertalan, 'Quantitative Determination of the Platinum-group Elements and Gold Using NiS Fire Assay with Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LA/ICPMS)', *Chem. Geol.*, **124**, 37–46 (1995).
62. V. Kinicky, J.-M. Mermet, 'Use of a Single Calibration Graph for the Determination of Major Elements in Geological Materials by Laser Ablation Inductively Coupled Plasma Atomic Emission Spectrometry with Added Internal Standard', *Fresenius' J. Anal. Chem.*, **363**, 294–299 (1999).
63. P.J. Sylvester, M. Ghaderi, 'Trace Element Analysis of Scheelite by Excimer Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (ELA/ICPMS) Using a Synthetic Silicate Glass Standard', *Chem. Geol.*, **141**, 49–65 (1997).
64. C. Leloup, P. Marty, D. Dall'ava, M. Perdereau, 'Quantitative Analysis for Impurities in Uranium by Laser Ablation Inductively Coupled Plasma Mass Spectrometry: Improvements in the Experimental Set-up', *J. Anal. At. Spectrom.*, **12**, 945–950 (1997).
65. V.V. Kogan, M.W. Hinds, G.I. Ramendik, 'The Direct Determination of Trace Metals in Gold and Silver Materials by Laser Ablation Inductively Coupled Plasma Mass Spectrometry Without Matrix Matched Standards', *Spectrochim. Acta, Part B*, **49**(4), 333–343 (1994).
66. A. Raith, J. Godfrey, R.C. Hutton, 'Quantitation Methods Using Laser Ablation ICPMS Part 2: Evaluation of New Glass Standards', *Fresenius' J. Anal. Chem.*, **354**, 163–168 (1996).
67. O.V. Borisov, X.L. Mao, A. Fernandez, M. Caetano, R.E. Russo, 'Inductively Coupled Plasma Mass Spectrometric Study of Non-linear Calibration Behavior During Laser Ablation of Binary Cu–Zn Alloys', *Spectrochim. Acta, Part B*, **54**(9), 1351–1365 (1999).
68. H.F. Falk, B. Hattendorf, K. Krengel-Rothensee, N. Wieberneit, S.L. Dannen, 'Calibration of Laser-ablation ICPMS. Can We Use Synthetic Standards with Pneumatic Nebulization?', *Fresenius' J. Anal. Chem.*, **362**, 468–472 (1998).
69. J.J. Leach, L.A. Allen, D.B. Aeschliman, R.S. Houk, 'Calibration of Laser Ablation Inductively Coupled Plasma Mass Spectrometry Using Standard Additions with Dried Solution Aerosols', *Anal. Chem.*, **71**, 440–445 (1999).
70. D. Günther, A. Audétat, R. Frischknecht, C.A. Heinrich, 'Quantitative Analysis of Major, Minor, and Trace Elements in Fluid Inclusions Using Laser Ablation/Inductively Coupled Plasma/Mass Spectrometry (LA/ICPMS)', *J. Anal. At. Spectrom.*, **13**, 263–270 (1998).
71. D. Günther, R. Frischknecht, H.-J. Müschenborn, C.A. Heinrich, 'Direct Liquid Ablation: A New Calibration Strategy for Laser Ablation/ICPMS Microanalysis of Solids and Liquids', *Fresenius' J. Anal. Chem.*, **359**, 390–393 (1997).
72. A.M. Ghazi, T.E. McCandless, D.A. Vanko, J. Ruiz, 'New Quantitative Approach in Trace Elemental Analysis of Single Fluid Inclusions: Applications of Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA/ICPMS)', *J. Anal. At. Spectrom.*, **11**, 667–674 (1996).
73. A. Moissette, T.J. Shepherd, S.R. Chenery, 'Calibration Strategies for the Elemental Analysis of Individual Aqueous Fluid Inclusions by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **11**, 177–185 (1996).
74. T. Tanaka, K. Yamamoto, T. Nomizu, H. Kawaguchi, 'Laser Ablation/Inductively Coupled Plasma Mass Spectrometry with Aerosol Density Normalization', *Anal. Sci.*, **11**(6), 967–971 (1995).
75. R.J. Watling, 'In-line Mass Transport Measurement Cell for Improving Quantification in Sulfide Mineral Analysis Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **13**, 927–934 (1998).
76. M.A. Shannon, B. Rubinsky, R.E. Russo, 'Mechanical Stress Power Measurements During High-power Laser Ablation', *J. Appl. Phys.*, **80**(8), 4665–4672 (1996).
77. C. Chaléard, P. Mauchien, N. Andre, J. Uebbing, J.L. Lacour, C. Geertsen, 'Correction of Matrix Effects in Quantitative Elemental Analysis with Laser Ablation Optical Emission Spectrometry', *J. Anal. At. Spectrom.*, **12**(2), 183–188 (1997).
78. V. Kanicky, V. Otruba, J.-M. Mermet, 'Characterization of Acoustic Signals Produced by Ultraviolet Laser Ablation Inductively Coupled Plasma Atomic Emission Spectroscopy', *Fresenius' J. Anal. Chem.*, **363**, 339–346 (1999).

79. D.P. Baldwin, D.S. Zamzow, A.P. D'Silva, 'Aerosol Mass Measurement and Solution Standard Additions for Quantitation in Laser Ablation Inductively Coupled Plasma Atomic Emission Spectroscopy', *Anal. Chem.*, **66**(11), 1911–1917 (1994).
80. J. Perez, B.R. Weiner, 'The Laser Ablation of Gold Films at the Electrode Surface of a Quartz Crystal Microbalance', *Appl. Surf. Sci.*, **62**(4), 281–285 (1992).
81. S.A. Watmough, T.C. Hutchinson, R.D. Evans, 'The Quantitative Analysis of Sugar Maple Tree Rings by Laser Ablation in Conjunction with ICPMS', *J. Environ. Qual.*, **27**(5), 1087–1094 (1998).
82. T. Prohaska, C. Stadlbauer, R. Wimmer, G. Stingeder, C. Latkoczy, E. Hoffmann, H. Stephanowitz, 'Investigation of Element Variability in Tree Rings of Young Norway Spruce by Laser-ablation–ICPMS', *Sci. Total Environ.*, **219**(1), 29–39 (1998).
83. S.A. Watmough, T.C. Hutchinson, R.D. Evans, 'Development of Solid Calibration Standards for Trace Elemental Analyses of Tree Rings by Laser Ablation Inductively Coupled Plasma/Mass Spectrometry', *Environ. Sci. Technol.*, **32**(14), 2185–2190 (1998).
84. C.D. GarbeSchonberg, C. Reimann, V.A. Pavlov, 'Laser Ablation ICPMS Analyses of Tree-ring Profiles in Pine and Birch From N Norway and NW Russia – a Reliable Record of the Pollution History of the Area?', *Environ. Geol.*, **32**(1), 9–16 (1997).
85. A. Raith, W.T. Perkins, N.J.G. Pearce, T.E. Jeffries, 'Environmental Monitoring on Shellfish Using UV Laser Ablation ICPMS', *Fresenius' J. Anal. Chem.*, **355**(7/8), 789–792 (1996).
86. E. Vander Putten, F. Dehairs, L. Andre, W. Baeyens, 'Quantitative In Situ Microanalysis of Minor and Trace Elements in Biogenic Calcite Using Infrared Laser Ablation/Inductively Coupled Plasma Mass Spectrometry: A Critical Evaluation', *Anal. Chim. Acta*, **378**(1–3), 261–272 (1999).
87. S. Tanaka, N. Yasushi, N. Sato, T. Fukasawa, S.J. Santosa, K. Yamanaka, T. Ootoshi, 'Rapid and Simultaneous Multi-element Analysis of Atmospheric Particulate Matter Using Inductively Coupled Plasma Mass Spectrometry with Laser Ablation Sample Introduction', *J. Anal. At. Spectrom.*, **13**(N2), 135–140 (1998).
88. C.J. Chin, C.F. Wang, S.L. Jeng, 'Multi-element Analysis of Airborne Particulate Matter Collected on PTFE-membrane Filters by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **14**(4), 663–668 (1999).
89. C.F. Wang, S.L. Jeng, F.J.R. Shieh, 'Determination of Arsenic in Airborne Particulate Matter by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**(1), 61–67 (1997).
90. C.F. Wang, C.J. Chin, S.K. Luo, L.C. Men, 'Determination of Chromium in Airborne Particulate Matter by High Resolution and Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **389**(N1–3), 257–266 (1999).
91. R. Vago, E. Gill, J.C. Collingwood, 'Laser Measurements of Coral Growth', *Nature (London)*, **386**(6620), 30–31 (1997).
92. D.J. Sinclair, L.P.J. Kinsley, M.T. McCulloch, 'High Resolution Analysis of Trace Elements in Corals by Laser Ablation ICPMS', *Geochim. Cosmochim. Acta*, **62**(11), 1889–1901 (1998).
93. S. Fallon, 'Corals at Their Limits: Laser Ablation Trace Element Systematics in Porites from Shirigai Bay, Japan', *Earth Planet. Sci. Lett.*, to be published.
94. R. Feng, N. Machado, J.N. Ludden, 'Lead Geochronology of Zircon by Laser Probe–Inductively Coupled Plasma Mass Spectrometry (LP/ICPMS)', *Geochim. Cosmochim. Acta*, **57**, 3479–3486 (1993).
95. B.J. Fryer, S.E. Jackson, P. Longerich, 'The Application of Laser Ablation Microprobe–Inductively Coupled Plasma Mass Spectrometry (LAM/ICPMS) to In Situ (U)–Pb Geochronology', *Chem. Geol.*, **109**, 1–8 (1993).
96. T. Hirata, R.W. Nesbitt, 'U–Pb Isotope Geochronology of Zircon: Evaluation of the Laser Probe–Inductively Coupled Plasma Mass Spectrometry Technique', *Geochim. Cosmochim. Acta*, **59**(12), 2491–2500 (1995).
97. H.C. Liu, B.Q. Zhu, Z.X. Zhang, 'Single Zircon Dating by LAM/ICPMS Technique', *Chin. Sci. Bull.*, **44**(2), 182–186 (1999).
98. T.J. Shepherd, S.R. Chenery, 'Laser Ablation ICPMS Elemental Analysis of Individual Fluid Inclusions – An Evaluation Study', *Geochim. Cosmochim. Acta*, **59**(19), 3997–4007 (1995).
99. A. Audetat, D. Günther, C.A. Heinrich, 'Formation of a Magmatic–Hydrothermal Ore Deposit: Insights with LA/ICPMS Analysis of Fluid inclusions', *Science*, **279**(27), 2091–2094 (1998).
100. A.N. Halliday, D.-C. Lee, J.N. Christensen, M. Rehkamper, W. Yi, X.Z. Luo, C.M. Hall, C.J. Ballentine, T. Pettke, C. Stirling, 'Application of Multiple Collector–ICPMS to Cosmochemistry, Geochemistry, and Paleoceanography', *Geochim. Cosmochim. Acta*, **62**(6), 919–940 (1998).
101. M.F. Thirlwall, A.J. Walder, 'In Situ Hafnium Isotope Ratio Analysis of Zircon by Inductively Coupled Plasma Multiple Collector Mass Spectrometry', *Chem Geol.*, **122**, 241–247 (1995).
102. L.V. Godfrey, D.-C. Lee, W.F. Sangrey, A.N. Halliday, V.J.M. Satlers, J.R. Hein, W.M. White, 'The Hf Isotopic Composition of Ferromanganese Nodules and Crusts and Hydrothermal Manganese Deposits: Implications for Seawater Hf', *Earth Planet. Sci. Lett.*, **151**, 91–105 (1997).
103. J N. Christensen, A.N. Halliday, D.-C. Lee, C.M. Hall, 'In Situ Sr Isotopic Analyses by Laser Ablation', *Earth Planet. Sci. Lett.*, **136**, 79–85 (1995).
104. J.N. Christensen, A.N. Halliday, L.V. Godfrey, J.R. Hein, D.K. Rea, 'Climate and Ocean Dynamics and

the Lead Isotopic Records in Pacific Ferro-Manganese Crust', *Science*, **277**, 913–918 (1997).
105. T. Hirata, M. Hattori, T. Tanaka, 'In-situ Osmium Isotope Ratio Analyses of Iridosmines by Laser Ablation Multiple Collector Inductively Coupled Plasma Mass Spectrometry', *Chem. Geol.*, **144**(3–4), 269–280 (1998).
106. S.A. Baker, M. Bi, R.Q. Aucelio, B.W. Smith, J.D. Winefordner, 'Analysis of Soil and Sediment Samples by Laser Ablation Inductively Coupled Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **14**(1), 19–26 (1999).
107. M. Odegard, S.H. Dundas, B. Flem, A. Grimstvedt, 'Application of a Double-focusing Magnetic Sector Inductively Coupled Plasma Mass Spectrometer with Laser Ablation for the Bulk Analysis of Rare Earth Elements in Rocks Fused with $Li_2B_4O_7$', *Fresenius' J. Anal. Chem.*, **362**(5), 477–482 (1998).
108. E.K. Shibuya, J.E.S. Sarkis, J. Enzweiler, A.P.S. Jorge, A.M.G. Figueiredo, 'Determination of Platinum Group Elements and Gold in Geological Materials Using an Ultraviolet Laser Ablation High-resolution Inductively Coupled Plasma Mass Spectrometric Technique', *J. Anal. At. Spectrom.*, **13**(9), 941–944 (1998).
109. F. Bea, P. Montero, G. Garuti, F. Zacharini, 'Pressure-dependence of Rare Earth Element Distribution in Amphibolite- and Granulite-grade Garnets. A LA/ICPMS Study', *Geostand. Newsl. J. Geostand. Geoanal.*, **21**(2), 253–270 (1997).
110. L.P. Bedard, 'Re-evaluation of the Homogeneity of REE, Th And U in Geochemical Reference Zircon 61.308b', *Geostand. Newsl.*, **20**(2), 289–293 (1996).
111. M.D. Norman, N.J. Pearson, A. Sharma, W.L. Griffin, 'Quantitative Analysis of Trace Elements in Geological Materials by Laser Ablation ICPMS – Instrumental Operating Conditions and Calibration Values of NIST Glasses', *Geostand. Newsl.*, **20**(2), 247–261 (1996).
112. W. Devos, C. Moor, P. Lienemann, 'Determination of Impurities in Antique Silver Objects for Authentication by Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA/ICPMS)', *J. Anal. At. Spectrom.*, **14**(4), 621–626 (1999).
113. B. Wanner, C. Moor, P. Richner, R. Bronnimann, B. Magyar, 'Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA/ICPMS) for Spatially Resolved Trace Element Determination of Solids Using an Autofocus System', *Spectrochim. Acta, Part B*, **54**(2), 289–298 (1999).
114. R.J. Watling, B.F. Lynch, D. Herring, 'Use of Laser Ablation Inductively Coupled Plasma Mass Spectrometry for Fingerprinting Scene of Crime Evidence', *J. Anal. At. Spectrom.*, **12**(2), 195–203 (1997).
115. R.J. Watling, 'Novel Application of Laser Ablation Inductively Coupled Plasma Mass Spectrometry in Forensic Science and Forensic Archaeology', *Spectroscopy*, **14**(6), 16–34 (1999).
116. J.J. Giglio, P.S. Goodall, S.G. Johnson, 'Application of High-resolution Inductively Coupled Plasma Atomic Emission Spectroscopy in the Nuclear Field', *Spectroscopy*, **12**(7), 26–37 (1997).
117. M.R. Smith, J.S. Hartman, M.L. Alexander, A. Mendoza, E.H. Hirt, M.R. Stewart, M.A. Hansen, W.R. Park, T.J. Peters, B.J. Burghard, J.W. Ball, C.T. Narquis, D.M. Thornton, R.L. Harris, 'Laser Ablation/Inductively Coupled Plasma Mass Spectrometry: Analysis of Hanford High-level Waste Materials', in *Science and Technology for Disposal of Radioactive Tank Wastes*, eds. W.W. Schullz, N.J. Lombardo, Plenum Press, New York, 135–158, 1998.
118. D.S. Zamzow, D.P. Baldwin, S.J. Weeks, S.J. Bajic, A.P. D'Silva, 'In Situ Determination of Uranium by Laser Ablation/Inductively Coupled Plasma Atomic Emission Mass Spectrometry', *Environ. Sci. Technol.*, **28**(2), 352–358 (1994).
119. B.A. Bushaw, M.L. Alexander, 'Investigation of Laser Ablation Plume Dynamics by High-resolution Time-resolved Atomic Absorption Spectroscopy', *Appl. Surf. Sci.*, **129**, 935–940 (1998).

Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Ove Axner
Umeå University, Umeå, Sweden

1 Introduction **3**
1.1 Conventional Techniques for Analytical Atomic Spectrometry – Emission and Absorption 3
1.2 Lasers as Powerful Light Sources for Analytical Atomic Spectrometry 3
1.3 Use of Lasers for Analytical Spectrometry 4
1.4 Laser Spectrometric Techniques for Analytical Atomic Spectrometry – Fluorescence, Ionization and Absorption 4
1.5 Organization of this Review 5
2 Laser Instrumentation **5**
2.1 Tunable Continuous-wave Laser Systems 5
2.2 Tunable Pulsed Laser Systems 6
2.3 Conclusions 7
3 Interaction of Atoms with Laser Light **7**
3.1 Fluorescence and Ionization 8
3.2 Two-level Atoms 9
3.3 Multilevel Atoms Exposed to One Laser Field 13
3.4 Three-level Atoms Exposed to Two Laser Fields 19
3.5 Conclusions – Justification of the Use of the Rate Equation Formalism 21
4 Laser-induced Fluorescence Spectrometry **22**
4.1 Introduction 22
4.2 Previous Reviews 23
4.3 Instrumentation 23
4.4 Analytical Performance 27
4.5 Diagnostic Capabilities 37
4.6 Conclusions and Outlook 37
5 Laser Ionization Spectrometry **38**
5.1 Laser-enhanced Ionization versus Resonance Ionization Spectrometry 38
5.2 Laser-enhanced Ionization Spectrometry – An Introduction 38
5.3 Previous Reviews 39
5.4 Instrumentation 39
5.5 Analytical Performance 44
5.6 Diagnostic Capabilities 54
5.7 Conclusions and Outlook for Laser-enhanced Ionization 56
5.8 Resonance Ionization Spectrometry 56
5.9 Conclusions and Outlook for Resonance Ionization Spectrometry/ Resonance Ionization Mass Spectrometry 59
6 Frequency- and Wavelength-modulation Diode Laser Absorption Spectrometry **59**
6.1 Introduction to Diode Lasers and Their Use for Analytical Applications 59
6.2 Frequency-modulation versus Wavelength-modulation 60
6.3 Wavelength-modulation Spectrometry 61
6.4 Conclusions and Outlook 67
7 Conclusions **67**
Acknowledgments **68**
Abbreviations and Acronyms **68**
Related Articles **69**
References **69**

Laser light has a number of spectacular properties that make it useful for analytical spectrometry. One is that it has a high directionality (i.e. it looks like a real ''beam''). This implies, among other things, that it can be focused down to micrometer-sized spots. Another is that pulsed lasers can emit large amounts of light in very short pulses (often with a duration of 10^{-9}–10^{-8} s). These two properties imply that laser light often can reach high irradiance ($W\,m^{-2}$), which is of importance for a number of applications, not least when laser light is used for vaporization and/or atomization purposes of solid material. The most important attribute for spectroscopic applications, however, is that it often has a narrow frequency width (in the MHz–GHz range). This implies that laser light can induce one specific transition in one particular species at a time. The narrow frequency width is thus the basis for the high species selectivity that laser spectroscopic techniques possess. In addition, the combination of high irradiance and a narrow frequency width often gives laser light such staggering spectral irradiance ($W\,m^{-2}\,Hz^{-1}$) that a significant fraction of the atoms under illumination can produce at least one detectable event during the interaction time (one or several

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

photons or an ion–electron pair). This explains why laser spectroscopic techniques can benefit from high species sensitivity.

The main purpose of this review article is to describe the theory and instrumentation for the field of laser spectroscopic techniques for analytical atomic spectrometry. This implies that techniques that use laser light for nonspectroscopic purposes, such as vaporization, and/or atomization purposes, e.g. laser ablation, laser-induced plasma spectrometry, laser mass spectrometry and laser-induced breakdown spectroscopy, will not be covered here. In addition, because atomic spectra in general consist of a few strong and narrow-band transitions (whose widths often are comparable to those of the light from tunable laser systems and therefore seldom overlap with those from other atomic species) whereas those of molecules comprise a few broader, weaker and to a certain extent structured transition bands (which thus overlap more often with those of other molecular species), laser spectrometric techniques often show the highest sensitivity and selectivity when atomic species are detected. This is the main reason why this review focuses upon the use of laser spectrometric techniques for analytical atomic spectrometry. This implies, in turn, that laser-based spectrometric techniques that predominantly detect molecules, e.g. Raman, thermal lensing and photoacoustic spectrometry, and spectrofluorometry, will not be discussed.

This theory and instrumentation review focuses upon the most useful and versatile laser spectroscopic techniques for analytical atomic spectrometry. These techniques are based upon the concepts of fluorescence, ionization or absorption. Those based upon fluorescence are often referred to as either laser-induced fluorescence (LIF) or laser-excited atomic fluorescence spectrometry (LEAFS), whereas those based upon ionization are termed either laser-enhanced ionization (LEI) or resonance ionization spectrometry (RIS). The working principles, instrumentation and present status of laser-induced fluorescence/laser-excited atomic fluorescence spectrometry (LIF/LEAFS) and LEI are covered in some detail in two separate sections, whereas RIS is covered more briefly. The reason for this is the broader versatility and applicability of LEI: it can be performed in a variety of atomizers, including those working under atmospheric pressure, whereas RIS has to be carried out in a vacuum and thus with significantly more complex instrumentation. However, because the laser instrumentation and the theoretical basis for all these techniques are quite similar, a short introduction to laser instrumentation and a theory section that outlines the common basic features of excitation of atoms by laser light precede the detailed descriptions. Those laser spectroscopic techniques that are based upon absorption and used for analytical atomic spectrometry are, nowadays, all performed in conjunction with some sort of modulation methodology (often wavelength-modulation (WM)), frequently by the use of diode lasers owing to their rapid tunability. They are therefore often referred to as WM diode laser (atomic) absorption techniques and are covered in a separate section.

Typical analytical qualities, e.g. limits of detection (LOD) and selectivity, as well as the typical strengths and limitations of each of these techniques, are given or discussed. It is concluded among other things that the most impressive performance of the LIF/LEAFS technique has been obtained together with the graphite furnace (GF) as atomizer, resulting in detection limits in the femtogram range and a very high selectivity for a large variety of elements. LEI, which has found its best use with the flame as atomizer, can provide LOD in the pg mL^{-1} range for many elements. It does not, however, show the same high selectivity as the LIF/LEAFS technique because it suffers from background effects when samples with high concentrations of easily ionized elements (EIEs) are analyzed. RIS combines a high sensitivity with an extraordinary selectivity, but instead has to pay the price of a more complex instrumentation. It was demonstrated earlier, for example, that the RIS technique is able to provide single atom detection (SAD). Because RIS in general is performed in a vacuum it can also provide good isotopic selectivity. Although not yet applied to a broad range of elements for analytical applications, it has shown impressive LOD (in the attogram range) and excellent isotopic selectivity ($>10^9$) under a few specific conditions. The wavelength-modulation diode laser absorption spectrometry (WM-DLAS) technique shows good promise for becoming a user-friendly and widespread detection technique because it requires less complex instrumentation than most other laser spectroscopic techniques. Used with the GF as atomizer, LOD in the femtogram range have been achieved. The WM-DLAS technique also has good ability to correct for various types of unstructured background absorption signals that might appear when samples with complex matrices are analyzed. One drawback, however, is that its applicability is still restricted owing to the limited availability of diode lasers that emit light in the visible and ultraviolet (UV) region. The WM-DLAS technique, therefore, has been applied to only a limited number of elements so far. A common denominator for all these laser spectroscopic techniques is that they often have a large linear dynamic range (LDR). The LIF/LEAFS technique, for example, has demonstrated an LDR of 5–7 orders of magnitude. Finally, in addition to being used as sensitive and selective tools for analytical assessments, the laser spectroscopic techniques often show an excellent applicability to diagnostic studies (e.g. for the determination of processes such as atomization, diffusion, collision or ionization).

1 INTRODUCTION

Analytical spectrometry refers to a family of techniques by which a sample is characterized with regard to its content of atoms (or molecules) by spectroscopic means. It is customary to distinguish between optical (e.g. laser) and mass spectrometric techniques. The underlying foundation of optical spectrometric techniques is the quantization of energy levels, which gives each type of atomic or molecular species its own unique set of transition wavelengths, whereas mass spectrometric techniques rely on the fact that each type of species has a unique mass. The present review is solely concerned with the use of optical spectrometric techniques. The reader is referred to other articles in this encyclopedia for a description of mass spectrometric techniques (**Mass Spectrometry: Overview and History; Time-of-flight Mass Spectrometry; Secondary Ion Mass Spectrometry as Related to Surface Analysis; Tandem Mass Spectrometry: Fundamentals and Instrumentation; Quadrupole Ion Trap Mass Spectrometer; Literature of Mass Spectrometry**).

1.1 Conventional Techniques for Analytical Atomic Spectrometry – Emission and Absorption

The two most commonly used conventional (optical) spectrometric techniques for analytical atomic spectrometry are emission and absorption spectroscopy.[1] In both of these techniques, the sample to be analyzed is introduced into a hot environment in which it is vaporized and subsequently atomized. Emission spectroscopy relies upon the fact that the heat of the atomizer thermally excites free atoms, which subsequently decay to lower-lying states by spontaneous emission. The wavelength distribution and the intensity of the spontaneously emitted light are measures of the analytical content of the sample. In absorption spectrometry, the wavelength distribution and the amount of the light absorbed when it is passing the atomized sample are measures of the analytical content of the sample.

Although both of these techniques have been developed to a considerable level of sophistication and applicability throughout the years (with a variety of atomizers), they suffer from a few inevitable inherent limitations. Emission spectrometry has a limited sensitivity (the thermal population of the excited states is usually rather low) and selectivity (given solely by the dispersive power of the detection instrumentation). In addition, the precision and accuracy are limited (the amount of light emitted for a given amount of analyte varies strongly with temperature, so any minor fluctuation of the atomizer conditions, e.g. from the matrix composition of the sample, can affect the sensitivity). The absorption technique is limited by influences from unspecific background absorption and stray light (although a variety of detection procedures, e.g. the D_2 lamp and Zeeman background correction, have been developed throughout the years in order to overcome these limitations). The present status of the emission and absorption spectrometry techniques can be found elsewhere in the literature[2–7] or in this encyclopedia (**Atomic Spectroscopy: Introduction; Inductively Coupled Plasma/Optical Emission Spectrometry; Flame and Vapor Generation Atomic Absorption Spectrometry; Graphite Furnace Atomic Absorption Spectrometry; Background Correction Methods in Atomic Absorption Spectroscopy; Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis**).

1.2 Lasers as Powerful Light Sources for Analytical Atomic Spectrometry

With the advent of the first laser (an acronym for light by amplified spontaneous emission (ASE) of Radiation) in 1960,[8] many of the limitations of the conventional analytical spectrometry techniques were expected to be overcome. Laser light has a number of spectacular properties that make it exceptionally useful for analytical spectrometry, the most important of which are: high directionality (i.e. it looks like a real ''beam'' of light), which implies that it has a high irradiance ($W\,m^{-2}$); and an extremely narrow frequency width (i.e. it can often be considered to consist of only ''one'' wavelength), implying that it has an even more impressive spectral irradiance ($W\,m^{-2}\,Hz^{-1}$). In addition, pulsed lasers can produce very short laser pulses, thus temporarily producing very high peak powers and thereby sometimes even staggering spectral irradiances (which is often of importance for analytical applications).

These properties give the laser-based spectrometric techniques both high sensitivity and excellent selectivity. The high sensitivity, which predominantly stems from the high irradiance of the laser light, implies that a significant fraction of the atoms under illumination can produce a detectable event (e.g. a high production of photons/ions/electrons per illuminated atom). Laser-based techniques, for example, can be used successfully for SAD.[9–11] The high selectivity, which mainly originates from the narrow bandwidth of the light, comes from the fact that it is often possible to induce one (and often only one) specific transition in one particular type of species at a time.

Several types of lasers with a variety of qualities have been developed since the first laser saw the light of day. The fact that some types of lasers can be scanned (or tuned) over the absorption profile (referred to as tunable lasers) makes them useful for analytical spectrometry (as well as indispensable for atomic physics[12–15]). With access to the tunable high-irradiance light sources

that lasers constitute, a whole new family of detection techniques for analytical spectrometry has emerged during the last decades.[15–21]

1.3 Use of Lasers for Analytical Spectrometry

Laser light can be used in a variety of ways for analytical spectrometric purposes. It is customary to distinguish between spectrometric and nonspectrometric applications.

1.3.1 Lasers for Spectrometric Applications – Detection of Atoms versus Molecules

Laser-based spectrometric techniques refer to those techniques in which the narrow bandwidth of the laser light is used to provide a species-selective interaction with the sample. Although laser spectrometric techniques can be used successfully for the detection of both atoms and molecules, they show the highest sensitivity and selectivity when atomic species are detected because atomic spectra generally consist of a few strong and narrow-band transitions (whose widths often are comparable to those of the light from a tunable laser system) as opposed to those of molecules, which comprise a few broader, weaker and to a certain extent structured transition bands. This is the main reason why this review is devoted to the use of laser spectrometric techniques for analytical atomic spectrometry. This implies, in turn, that laser-based spectrometric techniques that predominantly detect molecules, e.g. Raman,[22] thermal lensing and photoacoustic spectrometry,[23–28] spectrofluorometry,[29–31] optoacoustic detection[32] and degenerate four-wave mixing,[33] will not be covered in this review. The reader is referred to the literature or other articles in this encyclopedia for further information about these types of technique (**Raman Spectroscopy: Introduction; Raman Scattering, Fundamentals; Dispersive Raman Spectroscopy, Current Instrumental Designs; Fourier Transform Raman Instrumentation; Raman Microscopy and Imaging**).

1.3.2 Lasers for Nonspectrometric Analytical Applications

Lasers can also be used as a tool for vaporization and/or atomization of the sample in conjunction with various types of conventional detection techniques. There are a number of closely related techniques in this field: e.g. laser ablation, laser-induced plasma spectrometry, laser mass spectrometry and laser-induced breakdown spectroscopy. However, because this use of lasers does not contribute to the elemental selectivity through their narrow bandwidth, none of these types of technique will be covered in this review. The reader is again referred to other articles in this encyclopedia (e.g. **Laser-induced Breakdown Spectroscopy**) or to the literature for this type of application of lasers in analytical spectrometry.[34]

1.4 Laser Spectrometric Techniques for Analytical Atomic Spectrometry – Fluorescence, Ionization and Absorption

Although a number of different laser-based spectroscopic techniques have been scrutinized for their analytical applicability throughout the years, the techniques based upon *fluorescence*, *ionization* and *absorption* have been shown to be the most useful and versatile for analytical atomic spectrometry.

1.4.1 Laser-induced Fluorescence Spectrometry

Tuned to a particular transition in one specific element, the narrow-band laser light will selectively increase the number of excited analyte species far above the thermal levels. This implies that the emission will be increased considerably. Such laser-enhanced emission is generally termed LIF. By detecting this fluorescent light by suitable means (most often a spectrometer and a light-sensitive detector), an assessment of the amount (or concentration) of the analyte can be made.[35] LIF is a very useful and versatile technique that has found applications in a number of areas (e.g. atomic physics, quantum optics, environmental monitoring, combustion analysis and analytical spectrometry[36]). When applied to analytical spectrometry for the detection of atoms or molecules, it appears frequently under the name LEAFS[37,38] or laser-excited molecular fluorescence spectrometry (LEMOFS).[38]

1.4.2 Laser Ionization Spectrometry

Laser-excited atoms cannot only decay to lower-lying states by emission of photons (i.e. fluorescence) or by collisions with surrounding species (i.e. by inelastic or so-called quenching collisions), but they can also be ionized as a direct (or indirect) consequence of the laser illumination. Techniques that are based on a monitoring of the production of charged species (i.e. ions or electrons) that result because of laser illumination are therefore termed *laser ionization spectrometry techniques*. A laser ionization technique can thus be seen as an optical-to-electrical transducer whose output (number of charges produced) depends on the amount (or concentration) of the analyte to be assessed. In analytical applications, the ionization techniques are most often referred to as either LEI spectrometry[39,40] or RIS.[11,19,41] The main difference between these two types of technique is that RIS is performed under a vacuum but LEI needs to be

performed in an environment in which there is already an omnipresent thermal ionization. The reason for this is that the laser light is used for excitation as well as for ionization of the analytes in RIS, whereas the atoms only need to be promoted to an excited state by the laser light in the LEI technique (because the ionization is provided by an existing thermal ionization process in the atomic reservoir in this technique; see further discussion below). LEI has therefore the advantage over RIS that it can be performed in a variety of atomizers suitable for analytical purposes (flames, furnaces, etc.).

1.4.3 Laser Absorption Spectrometry

Tuned to a particular transition in an element, the laser light will not only increase the number of excited species but it will also experience an attenuation (one photon will be absorbed for each excitation event). This implies that the light will be attenuated as it passes a cloud of absorbing species. This is the basis for absorption techniques.

Laser-based absorption techniques have found best use for analytical spectrometry in combination with various modulation methodologies. Correctly used, the modulation reduces the noise in the system. Hence, although fluorescence and ionization techniques benefit from a high production of photons or ions/electrons, the laser-based absorption techniques benefit from a reduction of the noise in the system that the modulation brings. In addition, because diode lasers have the unique property that they can be tuned (i.e. modulated) at significantly higher rates than other tunable lasers, modulation absorption techniques often incorporate diode lasers and are therefore often referred to as wavelength-modulation (*WM*) or *frequency-modulation (FM) diode laser absorption techniques*.[42,43]

1.4.4 Coverage of this Review

Because these three laser spectroscopic techniques (fluorescence, ionization and absorption spectrometry) combine high sensitivity, high selectivity and broad versatility for the detection of a large variety of elements in trace concentrations/amounts in various types of samples, this review will focus upon their foundations, theory and instrumentation.

1.5 Organization of this Review

Section 2 gives some typical properties of modern laser systems suitable for analytical spectrometry. Section 3 covers the most important basic theoretical concepts related to the excitation of atoms by laser light and gives an introduction to the nomenclature and terminology of the field. Section 4 reviews the LIF/LEAFS technique in analytical applications. Section 5 deals with ionization techniques. After a short comparison of the RIS and the LEI techniques, most of the section is concerned with LEI owing to its broader versatility, whereas RIS is described more briefly at the end. Section 6 first makes some comments about FM and WM techniques in general, before it focuses the presentation on the WM diode laser absorption technique due to its suitability for analytical spectrometry. Finally, section 7 makes some concluding remarks about the field of laser spectrometric techniques for analytical atomic spectrometry and speculates briefly about its future.

2 LASER INSTRUMENTATION

Virtually all lasers used for analytical spectrometry are tunable (i.e. their wavelength can be tuned continuously over at least a certain part of the spectrum). With the exception of diode laser systems, a tunable laser system consists of a pump laser (producing nontunable high-power light) and a tunable laser (converting the pump laser light to preferably narrow-band tunable laser light).

Most tunable laser systems can be classified as either *continuous-wave* or *pulsed* systems, dictated by the qualities of the pump laser. Tunable continuous-wave laser systems most often consist of ion-laser-pumped dye laser or Ti : sapphire laser systems, or the diode lasers that have gained increasing interest during latter years. Tunable pulsed laser systems, which have dominated the area of laser analytical spectrometry for a number of years, consist most often now of excimer- or neodymium : yttrium aluminum garnet (Nd : YAG)-pumped dye laser or optical parametric oscillator (OPO) systems (although nitrogen- or flashlamp-pumped dye laser systems were common during the first years). Atomic-vapor-based systems, which have been used only occasionally for analytical spectrometry, take an intermediate position with pulse frequencies significantly exceeding and pulse energies significantly falling short of those of ordinary pulsed laser systems. A thorough and well-updated overview of the properties of laser systems in terms of the requirements of the field of analytical spectrometry has been given recently by Sneddon et al.[44]

2.1 Tunable Continuous-wave Laser Systems

2.1.1 Continuous-wave Dye Laser or Ti : Sapphire Laser Systems

Typical continuous-wave pump lasers produce light of a few (or tens of) watts. A tunable continuous-wave dye laser or Ti : sapphire laser system can therefore typically produce narrow-band laser light in the milliwatt to watt

range in the visible part of the spectrum. Light in the UV region, which often is needed for excitation of atomic species, can be produced at microwatt powers by use of frequency-doubling instrumentation (frequency-doubling is done in nonlinear crystals in which two photons of the incoming light beam are being combined to one photon of twice the energy, i.e. half the incoming wavelength).

Although laser beams most often are produced and used with typical diameters of a few millimeters, they can be focused down to spots of micrometer-size dimensions. This implies that laser irradiances from tunable continuous-wave laser systems can, in theory, range from fractions of $W\,m^{-2}$ to $10^{12}\,W\,m^{-2}$, with "typical" numbers in the 10^2–$10^3\,W\,m^{-2}$ range (for the case of 1–10 mW of light and a beam area of $10\,mm^2$).

In addition, monochromatic continuous-wave laser light can be produced with a variety of bandwidths, often in the megahertz range (from a few to some hundreds of megahertz, depending on the type of system, type of stabilization, etc.). This implies that tunable continuous-wave laser light can be produced with spectral irradiances up to $10^5\,W\,m^{-2}\,Hz^{-1}$ under strongly focused conditions, with "typical" values in the 10^{-6}–$10^{-4}\,W\,m^{-2}\,Hz^{-1}$ range for unfocused situations.

Comparing these numbers with those of an ordinary light source, e.g. the light from the Sun on the surface of the Earth (around $10^3\,W\,m^{-2}$), as is done in Table 1, one finds that the irradiances in fact are comparable. However, because sunlight is distributed over a large part of the spectrum (over approximately 10^{14} Hz), the typical spectral irradiance of tunable continuous-wave laser light (10^{-6}–$10^{-4}\,W\,m^{-2}\,Hz^{-1}$) is several orders of magnitude larger than that of the sunlight ($10^{-11}\,W\,m^{-2}\,Hz^{-1}$).

Table 1 Some "typical" (order-of-magnitude estimates) irradiances and spectral irradiances for three different light sources

	Irradiance[a] ($W\,m^{-2}$)	Spectral irradiance[a] ($W\,m^{-2}\,Hz^{-1}$)
The Sun[b]	10^3	10^{-11}
Continuous-wave laser system[c]	10^2–10^3 (10^{12})	10^{-6}–10^{-4} (10^5)
Pulsed laser system[d]	10^8–10^{10} (10^{18})	10^{-1}–10^1 (10^9)

[a] The values for the laser systems are given for a "typical" experimental situation with a beam diameter of a few millimeters. Values in parentheses refer to strongly focused situations (with a spot size of micrometer size dimensions).
[b] At the surface of the Earth.
[c] Values are given for a "typical" argon ion laser-pumped dye laser system.
[d] Values are given for a "typical" Nd:YAG- or excimer-pumped dye laser system with nanosecond pulse duration but are also in reasonable agreement with those for OPO systems.

2.1.2 Diode Lasers

Modern diode lasers suitable for analytical spectrometry (i.e. single-mode lasers) can produce light at powers of tens of milliwatts directly (mostly in the red or infrared region). By using external frequency-doubling instrumentation, powers in the microwatt region (and occasionally in the milliwatt region) can be produced.

In addition to the aforementioned properties, diode lasers possess yet another unique property: namely, an extraordinary rapid and versatile tunability. Diode lasers are controlled by their temperature and injection current.[45] A modulation of the wavelength can be accomplished by a rapid modulation of the injection current. Although modern diode laser systems can be modulated at gigahertz frequencies for communication purposes, most diode laser systems for analytical purposes are modulated in the megahertz range (for FM techniques) and the kilohertz range (for WM techniques).[42,46]

2.2 Tunable Pulsed Laser Systems

2.2.1 Pump Lasers

The comparison between laser light and conventional light becomes even more spectacular when pulsed lasers are considered. Nd:YAG lasers can in general produce light with higher pulse energies (about 1 J per pulse) than can excimer lasers (typically a few hundred millijoules per pulse). However, because the fundamental wavelength of Nd:YAG-produced light is 1064 nm, it has to be converted to more suitable wavelengths in order to make pumping of dye lasers or OPOs possible. Light at 532- and 355-nm wavelengths can be produced by frequency doubling of the 1064-nm light and frequency-mixing of the 532- and 1064-nm light, respectively, in nonlinear crystals. Crystals dedicated for frequency conversion of Nd:YAG laser light have rather high conversion efficiencies, implying that frequency-doubled and frequency-tripled light can be produced at pulse energies comparable to that of excimer laser systems (which produce light useful for dye laser pumping directly, most often at 308 nm when XeCl is used as the laser firing medium).

Typical pulse lengths for these types of laser are in the nanosecond range (typically 5 ns for Nd:YAG-pumped systems and often around 10–30 ns for excimer-pumped systems). The shorter pulse lengths of Nd:YAG lasers therefore give them an edge over excimer lasers when it comes to irradiances but also a handicap when it comes to the illumination time (as will be discussed in the following section). The excimer systems, however, have an irrefutable advantage when it comes to repetition rates. Although many Nd:YAG lasers can produce light only at a repetition rate of 10 or 30 Hz (and some at 50 Hz),

it is not uncommon for excimer lasers to run at several hundred hertz (some even up to 500 Hz).

2.2.2 Pulsed Dye Laser Systems

A pulsed dye laser system can produce tunable laser light in the low to medium millijoule range in the visible part of the spectrum. Because one particular dye can produce light only over a certain part of the spectrum (typically a few tens of nanometers), a number of different dyes are needed to cover the entire spectrum.[44] Typical conversion efficiencies of pulsed dye lasers depend on the choice of dye and the position within the laser firing range of the dye. Peak conversion efficiencies of various dyes range normally from a few per cent to a few tens per cent, with the highest efficiencies for coumarins (in the 450–600 nm range) for excimer-pumped dye lasers and for rhodamines (in the 560–650 nm range) for Nd : YAG-pumped systems. The fundamental wavelength ranges covered by these systems (i.e. using no nonlinear frequency-doubling or frequency-mixing instrumentation after the dye laser) are roughly 330–980 nm for XeCl-based excimer laser systems and 400–900 nm for Nd : YAG laser-based systems.

In order to create tunable UV light, the dye laser light has to be frequency-doubled in a nonlinear crystal. The conversion efficiencies of such crystals range from a fraction of a per cent to some tens per cent. Hence, UV light can most often be produced by pulsed dye laser systems with pulse energies in the micro- to low-millijoule range. The total wavelength ranges of pulsed dye laser systems depend on the particular nonlinear frequency-doubling (or frequency-mixing) schemes used. In general, however, virtually the entire wavelength region from below 200 nm up to around 1 μm can nowadays be covered with pulsed dye laser systems.

The irradiance from excimer- or Nd : YAG-pumped dye laser systems can be as high as 10^{18} W m^{-2} (assuming 25 mJ of light distributed over 5 ns and focused down to an area of 5 μm^2), with ''typical'' values in the 10^8–10^{10} W m^{-2} range (for 10 μJ–1 mJ of light distributed over 10 ns and a beam area of 10 mm^2). Unless any special precautions are taken, tunable light produced by pulsed dye laser systems has typical spectral bandwidths of a few gigahertz. This implies that the corresponding spectral irradiance can become as high as 10^9 W m^{-2} Hz^{-1}, but is under more normal conditions around 0.1–10 W m^{-2} Hz^{-1}, i.e. several orders of magnitude larger than from tunable continuous-wave laser systems (see Table 1).

2.2.3 Pulsed Optical Parametric Oscillator Systems

Tunable OPO laser systems have been developed during the latter years as a viable alternative to dye lasers. This all-solid-state laser has considerably improved the creation of pulsed tunable laser light. An OPO needs to be pumped by a pump laser in a similar way to dye lasers. An OPO does not, however, need any organic dye or wavelength-dispersive element because the laser light is being produced in a nonlinear crystal that splits the incoming photons into two (whose energies thus sum up to that of the incoming photon).[21] The two beams of light produced are referred to as signal and idler beams and have wavelengths that are above and below twice the wavelength of the incoming light, respectively. Pumped with the third harmonics of a Nd : YAG laser (355 nm), an OPO can therefore produce tunable laser light in two wavelength regions (approximately 410–690 nm and 730 nm–2 μm). Equipped with a frequency-doubling crystal, these lasers are tunable by computer control over the entire wavelength region of 220–2000 nm (with no need for any change of dye solution, etc.).

The output characteristics of an OPO depend greatly on the pump laser but in most situations are superior to those of dye lasers. Laser pulses with energies of several tens of millijoules can normally be produced in the entire visible part of the spectrum, whereas pulse energies of several millijoules can be produced in the UV region. Because there is no dispersive element in an OPO, the bandwidth is mainly determined by the bandwidth of the pump laser. In order to produce laser light with a sufficiently narrow bandwidth, the pump laser should therefore also have a narrow bandwidth. This requires the use of injection-seeded pump lasers (e.g. a Nd : YAG laser whose light is originally produced by a narrow-band diode laser system placed in the cavity).

2.3 Conclusions

In conclusion, modern tunable laser sources can produce narrow-band light with significantly higher spectral irradiance than any conventional light source. Because a narrow frequency width and a high spectral irradiance are the basis for the high selectivity and sensitivity, respectively, of laser-based spectroscopic techniques, laser light can, if correctly used, enhance the signal-to-noise (S/N) ratio and increase the detection power in many types of applications,[47–52] not least for analytical spectrometry.[53–55]

3 INTERACTION OF ATOMS WITH LASER LIGHT

Although all laser-based spectroscopic techniques have in common the fact that narrow-band laser light is used to interact selectively with the analyte atoms, the interaction has a different objective for the three techniques

scrutinized in this review: fluorescence, ionization and absorption. Although the main aim of the fluorescence technique is to produce as many fluorescence photons as possible, the ionization techniques strive for ionizing (selectively) as large a fraction of the analyte atoms as possible. The common denominator for these two techniques is therefore to excite as many analyte species as possible. The absorption techniques, on the other hand, strive for minimizing, by the use of various modulation techniques, the amount of noise that affects the measurements, rather than maximizing the number of excited analyte species. This necessarily leads to different theoretical descriptions of, on the one hand, the fluorescence and ionization techniques and, on the other hand, the absorption technique. In this section we will give a theoretical basis for the fluorescence and ionization techniques (in terms of the number of atoms excited); the theoretical description of the absorption techniques is referred to in section 6.

3.1 Fluorescence and Ionization

Before scrutinizing in detail how free atoms interact with laser light, let us first conclude which entities are of importance for the fluorescence and ionization techniques.

For the fluorescence technique, fluorescence is assumed to be measured from an excited level, here denoted level 2 for simplicity, down to a lower-lying level, denoted 1 (see Figure 1). The rate of photon emission from level 2 to level 1 in a volume V of excited atoms, $\dot{N}_{\mathrm{ph}}(t)$, is directly proportional to the number density of atoms in the excited state, $n_2(t)$ (in units of m^{-3}), as shown in Equation (1):

$$\dot{N}_{\mathrm{ph}}(t) = VA_{21}n_2(t) \tag{1}$$

where A_{21} represents the spontaneous emission rate from level 2 to level 1. The number of fluorescence photons emitted over a time t, $N_{\mathrm{ph}}(t)$, can then be written as in Equation (2):

$$N_{\mathrm{ph}}(t) = \int_0^t \dot{N}_{\mathrm{ph}}(t)\,\mathrm{d}t' = VA_{21}\int_0^t n_2(t')\,\mathrm{d}t' \tag{2}$$

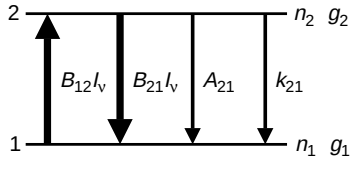

Figure 1 Schematic representation of excitation and de-excitation processes in a two-level atom: $B_{12}I_\nu$ and $B_{21}I_\nu$ represent the light-induced absorption and stimulated emission rates; A_{21} represents the spontaneous emission rate; k_{21} is the collisional de-excitation rate; and n_1 and n_2 are the populations (in units of m^{-3}) and g_1 and g_2 the degeneracies of the two levels respectively.

When ionization techniques are used, the entity of interest is the number of atoms that are ionized as a consequence of the laser illumination. As will be discussed in detail below, a prerequisite for the LEI technique is that there exists a thermal collisional ionization rate in the medium.[39] In such a system, species in all states, i (and not only those in one particular state, as is the case for the fluorescence technique), will contribute to the thermal ionization rate, $\dot{N}_{\mathrm{ion}}(t)$, by their own (state-specific) collisional ionization rate, $k_{i,\mathrm{ion}}$. This implies that the total ionization rate can be written as in Equation (3):

$$\dot{N}_{\mathrm{ion}}(t) = V\dot{n}_i(t) = V\sum_i k_{i,\mathrm{ion}}n_i(t) \tag{3}$$

where $\dot{n}_i(t)$ is the ionization rate density (in units of $\mathrm{m}^{-3}\,\mathrm{s}^{-1}$) and where the sum goes over all levels in the analyte.[56,57] The total number of ions produced during a time t, $N_{\mathrm{ion}}(t)$, can therefore be expressed by Equation (4):

$$N_{\mathrm{ion}}(t) = \int_0^t \dot{N}_{\mathrm{ion}}(t')\,\mathrm{d}t' = V\sum_i k_{i,\mathrm{ion}}\int_0^t n_i(t')\,\mathrm{d}t' \tag{4}$$

For the RIS technique, for which photoionization most often is the dominating ionization mechanism,[11,41] Equations (3) and (4) are still valid, with the only difference that the state-specific collisional ionization rate, $k_{i,\mathrm{ion}}$, needs to be interpreted as a state-specific photoionization rate, $\phi_{i,\mathrm{ion}}$, given by Equation (5):

$$\phi_{i,\mathrm{ion}} = I\frac{\sigma_{i,\mathrm{ion}}}{h\nu} \tag{5}$$

where I is the irradiance of the laser light (in units of $\mathrm{W\,m}^{-2}$), $\sigma_{i,\mathrm{ion}}$ is the cross-section for photoionization from state i, and $h\nu$ is the photon energy. The sum in Equation (4) should in this case run over all states that have an energy deficit to ionization limit that is smaller that the photon energy.

The presentation below has, however, mostly been restricted to techniques utilizing atmospheric pressure atomizers, such as flames, furnaces or plasmas, owing to their versatility and user-friendliness. This presentation will therefore focus more upon the situation for the LEI technique than the RIS technique (although the two techniques have many aspects in common) and thereby consider the ionization to be mainly caused by collisions. The reader is referred to the literature for a more detailed presentation of the theory for the RIS technique.[19]

Hence, in order to estimate the number of photons emitted, $N_{\mathrm{ph}}(t)$, or the number of ions created, $N_{\mathrm{ion}}(t)$,

from a given atomic system we need to determine the number densities (often referred to as the ''populations'') of the excited states ($n_2(t)$ for the LIF technique and all $n_i(t)$ for the LEI technique).

As will be shown by the treatment below, in order to get a high sensitivity for the LIF technique it is often sufficient to excite the analyte with light from one laser (commonly referred to as one-step excitation). One-step excitations in two- or three (multi)-level atoms will therefore be covered first. A one-step excitation is, however, normally not the optimum situation for the LEI technique. The highest performance (sensitivity as well as selectivity) is normally achieved when the atoms are excited in two separate steps using two individually tuned laser beams (using so-called two-step excitations). A two-step excitation process will therefore be modeled briefly thereafter. As will be discussed further below, two-step excitations can, under certain conditions, be beneficial also for the LIF technique. In addition, this presentation will be based upon the rate equation formalism and be restricted to the cases when the laser light is tuned exactly on resonance. This implies, among other things, that there will be virtually no discussion about the spectral shape of various excitation profiles.

3.2 Two-level Atoms

3.2.1 General Description of a Two-level System Exposed to a Resonant Laser Field

When laser light is tuned to a transition between the ground state and an excited state in an atom and the atomic structure is such that it is warranted to neglect the presence of other excited levels, the analyte atoms can often be described as two-level atoms. The two-level model can well be used, for example, for descriptions of excitations of the first resonant transition in elements that have no fine-structure splitting of any of the laser-connected levels and very few other excited states close to the laser-connected state (as, for example, is the case for Zn, Cd and Hg). The two-level model is sometimes appropriate also for other more complicated atomic systems.

3.2.1.1 Rate Equations for a Two-level System The excitation process in a two-level atom exposed to resonant laser light (as displayed in Figure 1) can be modeled in the so-called rate equation formalism as depicted in Equations (6) and (7):

$$\frac{dn_2}{dt} = B_{12}I_\nu n_1 - (B_{21}I_\nu + R_{21})n_2 \tag{6}$$

$$n_{\text{atom}} = n_1 + n_2 \tag{7}$$

where $B_{12}I_\nu$ and $B_{21}I_\nu$ represent the light-induced excitation and de-excitation rates between the two laser-connected states, respectively; R_{21} represents all the nonlight-induced de-excitation processes (in this case the sum of the spontaneous emission rate, denoted A_{ij}, and the collisional de-excitation rate, denoted k_{ij}); and n_{atom} is the total number density of (neutral) analyte atoms in the interaction region.

The light-induced excitation and de-excitation rates, $B_{12}I_\nu$ and $B_{21}I_\nu$, consist of a product of the Einstein coefficients for absorption and stimulated emission (B_{12} and B_{21}, respectively, in units of $m^2\,Hz\,J^{-1}$) and the spectral irradiance of the light (in units of $W\,m^{-2}\,Hz^{-1}$). The Einstein coefficients for absorption and emission are related to the rate of spontaneous emission, A_{21}, by the relations shown in Equation (8):

$$B_{12} = \frac{g_2}{g_1}B_{21} = \frac{g_2}{g_1}\frac{\lambda^3}{8\pi hc}A_{21} \tag{8}$$

where λ is the wavelength of the light, h is Planck's constant, c is the speed of light, and g_1 and g_2 are the degeneracies of the two levels.

3.2.1.2 Solution of the Rate Equations When solving rate equations, it is often convenient to work with normalized entities, i.e. to define the fraction of neutral atoms in a state i as $\bar{n}_i$ according to Equation (9):

$$\bar{n}_i = \frac{n_i}{n_{\text{atom}}} \tag{9}$$

Using this nomenclature, the fraction of neutral atoms in state 2 can be written according to Equation (10):

$$\bar{n}_2(t) = \bar{n}_2^{\text{sat}}\frac{B_{12}I_\nu}{B_{12}I_\nu + \bar{n}_2^{\text{sat}}R_{21}}\;1 - \exp\left(\frac{-t}{\tau}\right. \tag{10}$$

where $\bar{n}_2^{\text{sat}}$ and τ are the saturation population fraction of level 2 (Equation 11) and the time constant for establishment of a steady state (i.e. a time-independent situation; Equation 12) respectively.

$$\bar{n}_2^{\text{sat}} = \frac{g_2}{g_1 + g_2} \tag{11}$$

$$\tau = \frac{\bar{n}_2^{\text{sat}}}{B_{12}I_\nu + \bar{n}_2^{\text{sat}}R_{21}} \tag{12}$$

This implies that when the two-level system is exposed to a laser excitation, the fraction of excited atom will increase with time (in an exponential manner with a time constant τ) towards a steady-state value given by the product of the two first factors of Equation (10). The product of these two entities is therefore referred to as the *steady-state excitation population fraction*. In order to estimate a typical time constant for the establishment of a steady-state condition as well as the steady-state excitation population fraction of an atomic system exposed to laser light under conditions typically

prevailing in the most commonly used atomizers, the various rates involved need to be estimated.

3.2.1.3 Typical Excitation and De-excitation Rates Typical values for the spontaneous emission rate, A_{21}, range generally from $10^7\,s^{-1}$ to a few times $10^8\,s^{-1}$. This implies that the Einstein coefficients for absorption and stimulated emission, B_{12} and B_{21}, are of the order of 3×10^{10}–$3 \times 10^{11}\,m^2\,Hz\,J^{-1}$ for a UV transition (at 250 nm), assuming equal degrees of degeneracy of the two levels. Because typical spectral irradiances of UV light from pulsed laser systems are in the 0.1–$1\,W\,m^{-2}\,Hz^{-1}$ range (see Table 1), typical absorption and laser-induced stimulated emission rates for pulsed laser excitations are often in the gigahertz range. For example, excitation of atoms using a UV transition with a spontaneous emission rate of $10^8\,s^{-1}$ using laser pulses with a 10-μJ pulse energy, a 10-ns duration and a spectral width of 10 GHz, focused down to a diameter of 6 mm, gives rise to an excitation rate around 1 GHz. The corresponding excitation rate for a similar transition in the visible region can often become one to three orders of magnitude larger, owing to larger B-factors and higher laser pulse energies.

The collisional de-excitation rates depend on factors such as pressure, temperature and type of collision partner, and differ therefore among various atomizers. Typical collision rates between radiatively coupled levels in atoms in inert gas atmospheric pressure atomizers, e.g. flames with Ar as buffer gas,[58] often referred to as a high quantum efficiency flame, or electrothermal atomizers (ETAs), are often slightly below the gigahertz range. In air-supported atomizers, the rates are often in the low gigahertz range, with particularly high values for rates between closely spaced states.[59] This implies that collisional de-excitation rates often can dominate over spontaneous emission (especially in high collisional media such as flames). On the other hand, because light-induced excitation and de-excitation rates often can be well into the gigahertz range, they can, in turn, often exceed all nonlaser-connected de-excitation rates in the system.

3.2.1.4 Justification of the Steady-state Approximation Because excitation rates often are in the gigahertz range when pulsed laser systems are used, Equation (12) shows that the time constant for establishment of a steady-state condition in a two-level system becomes very short, often significantly less than 1 ns. A couple of examples of the time dependence of the population of the upper level of a two-level system are shown in Figure 2(a) for two different laser excitation rates (0.5 and 5 GHz, respectively) for a "typical" situation with a nonlaser-induced de-excitation rate of 1 GHz. Because most lasers used for LIF and LEI spectrometry are pulsed with pulse lengths in the 5–30 ns range, and the saturated

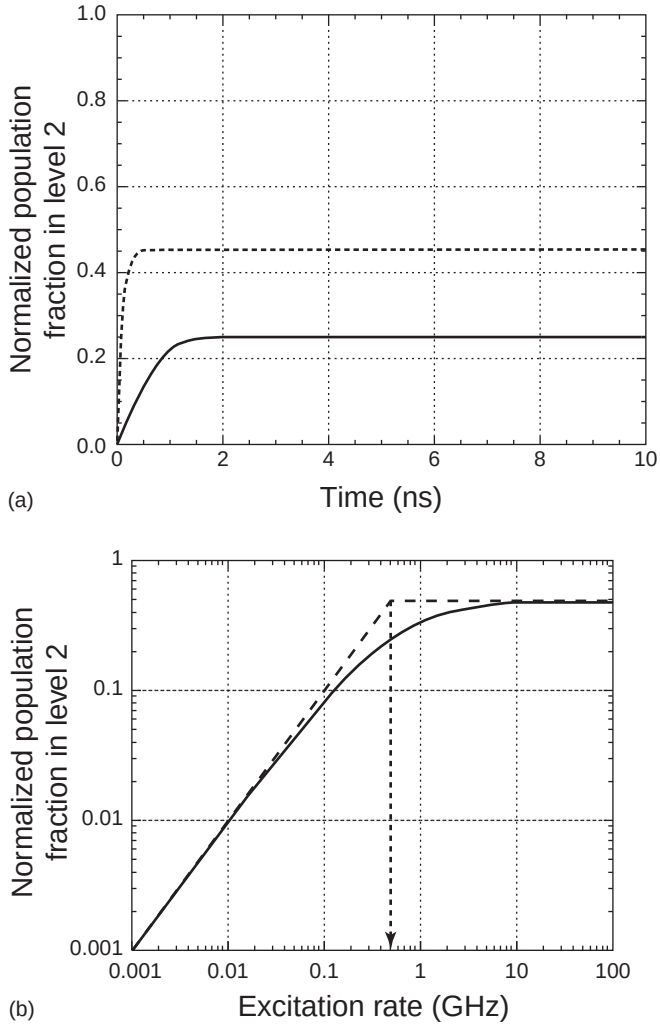

Figure 2 (a) The time dependence of the excited fraction of atoms of a two-level system exposed to laser-induced excitation rates of 0.5 (———) and 5 GHz (- - - -). (b) A saturation curve for the excited fraction of atoms (under steady-state conditions) as a function of laser-induced excitation rate. We have, for simplicity, assumed equal degeneracy of the two levels and a total nonlight-induced de-excitation rate of 1 GHz. The saturation rate for the system (0.5 GHz) is specifically marked.

excited fraction of atoms reaches its steady-state value within about 1 ns, one can conclude that any two laser-connected levels will reach their steady-state values within a fraction of the pulse length. This warrants the often-used approximation (valid for pulses whose duration is significantly longer than the time constant τ) that the number densities of atoms in the two laser-connected levels can be considered to be *locked* to each other (to their steady-state population fractions) during the laser illumination period.

3.2.1.5 Optical Saturation Moreover, Equation (10) above shows that the steady-state number density

of excited atoms in the interaction volume will be proportional to the laser spectral irradiance (i.e. to the intensity) for low light levels and that it will level off towards a plateau for higher light levels, i.e. the saturated excited fraction of atoms (given by $\bar{n}_2^{\text{sat}}$) shown in Figure 2(b). This particular type of population density versus irradiation curve is called a *saturation curve*. The reason for the leveling off of the population density for high light levels is that the rate of stimulated emission dominates the combined rate of spontaneous emission and collisional de-excitation. When this takes place, the system is said to be *optically saturated*. The second term in the denominator of Equation (12), i.e. $\bar{n}_2^{\text{sat}} R_{21}$, is called the *saturation rate* (or *saturation parameter*). This rate has an important role when laser-based techniques are used for diagnostic purposes.[60] The low- and high-intensity asymptotes are drawn with dashed lines in Figure 2(b). The saturation rate (which is the rate at which these two asymptotes meet, or where the saturated excited fraction of atoms reaches half of its maximum value) is also indicated for clarity.

The intensity (or spectral irradiance) for which the excitation rate is equal to the saturation parameter is termed *saturation intensity* (or *spectral saturation irradiance*).[61–68] The spectral saturation irradiance for a two-level system can be written according to Equation (13):

$$I_\nu^{\text{sat}} = \bar{n}_2^{\text{sat}} \frac{R_{21}}{B_{12}} = \frac{g_1}{g_1 + g_2} \frac{8\pi hc}{\lambda^3} \frac{1}{Y_{21}} \tag{13}$$

where Y_{21} represents the *fluorescence yield* (also referred to as *quantum yield* in the literature). The fluorescence yield can be defined as the ratio of the spontaneous emission rate to the total nonlaser-induced de-excitation rate (Equation 14), i.e.

$$Y_{21} = \frac{A_{21}}{R_{21}} = \frac{A_{21}}{A_{21} + k_{21}} \tag{14}$$

and is an entity of importance for the amount of fluorescence emitted from a system (see discussion below).

Because the collisional de-excitation rate differs among various atomizers (and depends on the running conditions to a certain extent), the fluorescence yield and thereby the saturation spectral irradiance will differ. Depending on details of the transitions and the collision partners, the fluorescence yield can range from a few percent to values close to unity in atmospheric pressure atomizers.[59] As an example, Hannaford reported on fluorescence yields in the 0.24–0.87 range for the strongest resonance lines of a number of elements (Li, Na, K, Rb, Cs, Ga, In, Tl, Au, Sn, Pb, etc.) in an argon-diluted stoichiometric oxygen–hydrogen flame.[69] Fluorescence yields in this range give "typical" spectral saturation irradiances for a two-level system in the 10^{-4}–10^{-3} W m^{-2} Hz^{-1} range for UV transitions and in the 10^{-3}–10^{-2} W m^{-2} Hz^{-1} range for transitions in the visible spectrum. By comparison with the "typical" spectral irradiance from pulsed as well as continuous-wave laser systems (see Table 1), one can conclude that optical saturation of a two-level system in atmospheric pressure atomizers can easily be obtained with light from pulsed laser systems. For a continuous-wave system, on the other hand, optical saturation is normally only achieved with some degree of focusing of the laser beam.

3.2.2 Fluorescence from a Two-level System

Let us define the number of photons emitted per atom under a time t, $\overline{N}_{\text{ph}}(t)$, according to Equation (15):

$$\overline{N}_{\text{ph}}(t) = \frac{N_{\text{ph}}(t)}{N_{\text{atom}}} \tag{15}$$

where N_{atom} is the number of illuminated atoms (given by Vn_{atom}). The rate of photon emission per atom, $\dot{\overline{N}}_{\text{ph}}(t)$, can then be written (using Equations 1, 9 and 15) according to Equation (16):

$$\dot{\overline{N}}_{\text{ph}}(t) = A_{21} \bar{n}_2(t) \tag{16}$$

3.2.2.1 Fluorescence from an Unsaturated Two-level System The rate of photon emission per atom, for a two-level system exposed to low light levels (i.e. for laser light intensities below the saturation intensity, $I_\nu \ll I_\nu^{\text{sat}}$), under steady-state conditions (i.e. for $t \gg \tau$), can then (using Equations 10, 14 and 16) be expressed as shown in Equation (17):

$$\dot{\overline{N}}_{\text{ph}} = Y_{21} B_{12} I_\nu \tag{17}$$

This shows that the rate of photon emission per atom from a two-level system (illuminated by a laser pulse whose duration is longer than τ) is independent of time and that it depends, in the unsaturated case, on the product of the excitation rate, $B_{12} I_\nu$, and the fluorescence yield, Y_{21}. The number of photons emitted per atom under a time t is therefore given by Equation (18):

$$\overline{N}_{\text{ph}}(t) = Y_{21} B_{12} I_\nu t \tag{18}$$

Although these expressions look handy, they show that the fluorescence technique indeed has a weakness. The fact that the signal depends on the fluorescence yield implies that the sensitivity of the LIF technique (the number of photons produced from a given amount of

atoms) varies, under unsaturated conditions, not only with the amount of laser light but also with the quenching collisional rate. This implies, in turn, that the sensitivity will vary in time and in space in a manner that is not fully controlled if the local environment in the atomizers differs between various parts of the interaction region or if it is dependent on time (e.g. if it is affected by the matrix in the sample).

3.2.2.2 Fluorescence from a Saturated Two-level System For a high laser light irradiance (i.e. for $I_\nu \gg I_\nu^{\text{sat}}$), the fraction of atoms excited (and thereby also the number of emitted photons) becomes independent of the laser light intensity as well as of the quenching rate. This implies that the rate of emission of fluorescence photons per atom under saturated conditions can be written as shown in Equation (19):

$$\dot{\overline{N}}_{\text{ph}} = \bar{n}_2^{\text{sat}} A_{21} \tag{19}$$

The number of photons emitted per atom from a saturated two-level system under a time t is therefore given by Equation (20):

$$\overline{N}_{\text{ph}}(t) = \bar{n}_2^{\text{sat}} A_{21} t \tag{20}$$

This implies that the number of photons emitted per (two-level) atom can be made independent of the fluorescence yield as well as the laser irradiation under saturated conditions.

3.2.2.3 Number of Photons Emitted per Two-level Atom Using ''typical'' numbers from the discussion above, one finds that the number of photons emitted per two-level atom, excited by light from an excimer system under saturated conditions, can be close to unity (e.g. for a situation with a spontaneous emission rate of $10^8\,\text{s}^{-1}$, equal degeneracies of the two levels, and a pulse duration of 20 ns).

The number of photons emitted per atom under saturated conditions will exceed unity for sufficiently long laser illumination times (i.e. for $t > (\bar{n}_2^{\text{sat}} A_{21})^{-1}$).

The interaction time for continuous-wave illumination is often determined by the time the atoms spend in the interaction volume, which, in turn, depends on properties of the particular instrumentation used: primarily the atom drift velocity and dimensions of the laser beam. The number of photons emitted per two-level atom can therefore significantly exceed unity for continuous-wave excitations (even if optical saturation is not always reached). However, because illumination with continuous-wave light implies that detection has to be performed over long periods of time, more noise will be detected than under pulsed conditions. This does not necessarily imply, therefore, that the detectability of atoms with continuous-wave light is better than with pulsed laser sources.

3.2.3 Practical Aspects of Optical Saturation

3.2.3.1 Analytical Aspects of Optical Saturation The advantages of optical saturation are that it maximizes the fluorescence signal, minimizes the signal fluctuations (which implies that it improves on the precision), and makes the signal independent of the fluorescence yield (improves on the accuracy). Although optical saturation seems to solve many problems related to quantitative spectrometry, it has some drawbacks. The main disadvantage is that optical saturation can deteriorate the S/N ratio by introducing high levels of scattered light or background signals.[68] The phenomenon of optical saturation also leads to a broadening of the profiles when scanning the laser wavelength. This is referred to as saturation broadening, as has been reviewed in detail in the literature.[65] It has therefore been suggested that a suitable working methodology for the fluorescence technique (when applied to analytical applications) is to work at laser intensities close to (but not really at) optical saturation.[66]

3.2.3.2 Diagnostic Aspects of Optical Saturation It was prophesied earlier that the LIF technique, owing to the existence of optical saturation as well as the tunability, narrow bandwidth and high directionality of the laser light, also has a unique diagnostic capability.[35,67] Numerous publications have therefore analyzed the properties of the optical saturation phenomenon in detail.[35,61–63,70] It should be possible, for example, to use the concept of saturated fluorescence to determine a variety of physical or chemical entities in the system under study. Such an experiment would consist of operating the laser at a sufficiently high spectral irradiance (so that optical saturation is guaranteed) and measuring the amount of fluorescence. By then attenuating the laser light in known fractions (e.g. using neutral density filters) and measuring the fluorescence, it should be possible to determine the saturation spectral irradiance. Entities such as the fluorescence yield, the de-excitation rate, or the quenching cross-section could then be evaluated from the saturation spectral irradiance through Equation (13).

History has shown, however, that the analysis of the experimental data is not always as unambiguous as suggested above. The reason is that the saturation curves very seldom conform to the predicted ideal saturation curve predicted by Equation (10).[61–63,71] Instead, experimentally measured saturation curves often continue to increase with increasing light levels instead of reaching the irradiance-independent plateau. There are

numerous reasons for this.[62] Two important ones are spatial and temporal nonhomogeneities of the laser beam. A laser beam with a higher intensity in its center part than in its periphery will saturate the atoms in its center parts to a larger extent that in its periphery.[62,71] A pulse with a temporal inhomogeneity will saturate the atoms more strongly during its high-intensity part than in its weak-intensity parts. Other possible reasons for nonideal saturation curves are that the atomic system is not sufficiently well represented by simple two-level atoms or that the steady-state condition is not valid. For further details, the reader is referred to an excellent review by Alkemade about anomalous saturation curves in LIF.[62]

3.2.4 Limitations of the Two-level Representation

The situation described by this two-level model, in which the fluorescence is measured at the same transition wavelength as is used for the excitation, is referred to as *resonance fluorescence*. However, this means of detecting fluorescence has little practical use for the most demanding analytical applications because the detection system, which necessarily has to be tuned to detect fluorescence at the laser wavelength, has no possibility to reject scattered laser light (which originates from laser light being reflected at various surfaces or by particles or molecules in the atomizer). It is therefore of importance to find viable alternatives to resonance fluorescence. One such alternative is to excite the atoms to levels from which they can fluoresce at more than one wavelength. This implies that it is of importance to investigate to which extent other levels (i.e. nonlaser-connected levels) interact with the laser-connected levels. The following section is therefore concerned with an analysis of a multilevel atomic system, represented by various types of three-level system for simplicity, exposed to one laser field. This issue is thus of special importance to all modes of detection that are not well represented by a two-level system. A three-level model will also enable us to describe the most basic phenomena for the LEI technique. As will be apparent, the existence of various nonlaser-connected levels will significantly affect the atomic system (and thereby also the signal strength and the detection capability).[72–74]

3.3 Multilevel Atoms Exposed to One Laser Field

3.3.1 General Description of a Three-level System Exposed to One Laser Field

3.3.1.1 Rate Equation System for a Three-level System Exposed to One Laser Field The rate equations for a three-level atomic system exposed to one laser field connecting levels 1 and 2, and with the nonlaser-connected level denoted 2′, can be written as Equations (21–23):

$$\frac{dn_2}{dt} = (B_{12}I_\nu + R_{12})n_1 - (B_{21}I_\nu + R_{21} + R_{22'})n_2 + R_{2'2}n_{2'} \quad (21)$$

$$\frac{dn_{2'}}{dt} = R_{12'}n_1 + R_{22'}n_2 - (R_{2'1} + R_{2'2})n_{2'} \quad (22)$$

$$n_{\text{atom}} = n_1 + n_2 + n_{2'} \quad (23)$$

where the various R_{ij}, as before, represent the sum of the collisional de-excitation rate, k_{ij}, and the spontaneous emission rate, A_{ij}, for the case of dipole-allowed exothermic transitions, but only the collisional excitation or de-excitation rates (i.e. k_{ij}) for the other transitions.

3.3.1.2 Various Types of Three-level Systems Although it is possible to solve the set of rate equations once and for all,[59,63,73,75] it is more illustrative and often more convenient to solve the system under a few typical conditions so that the specific influence of each type of nonlaser-connected level can be appraised clearly. There are three conceptually different three-level systems of specific interest for the LIF and LEI techniques, namely when the nonlaser-connected state is:

1. representative of the huge number of highly excited states (Rydberg levels[76]) that exist in all atoms (see Figure 3a);
2. an excited level in close proximity to the upper laser-connected level, which also is radiatively coupled to the ground state, e.g. the ^{2}P-states in alkali atoms (see Figure 3b);
3. a metastable state, i.e. one that is radiatively coupled to the upper laser-connected level but not to the ground state, e.g. the upper fine-structure component of the ground configuration in elements such as Al, In, Ga and Tl (see Figure 3c).

These three systems will be referred to below as cases 1, 2 and 3 respectively.

3.3.1.3 Three-level System with a High-lying Nonlaser-connected State The influence of a highly excited state (case 1) or a state in close proximity to the upper laser-connected level (case 2) can be investigated by solving the set of rate equations (Equations 21–23) under steady-state conditions. The fraction of atoms in the two excited states (under steady-state conditions) can be written according to Equations (24) and (25):

$$\bar{n}_2 = \bar{n}^{\text{sat}}_{2(3)} \frac{B_{12}I_\nu}{B_{12}I_\nu + \bar{n}^{\text{sat}}_{2(3)}(R_{21} + R_{2'1}\chi_{2'})} \quad (24)$$

$$\bar{n}_{2'} = \bar{n}^{\text{sat}}_{2'(3)} \frac{B_{12}I_\nu}{B_{12}I_\nu + \bar{n}^{\text{sat}}_{2(3)}(R_{21} + R_{2'1}\chi_{2'})} \quad (25)$$

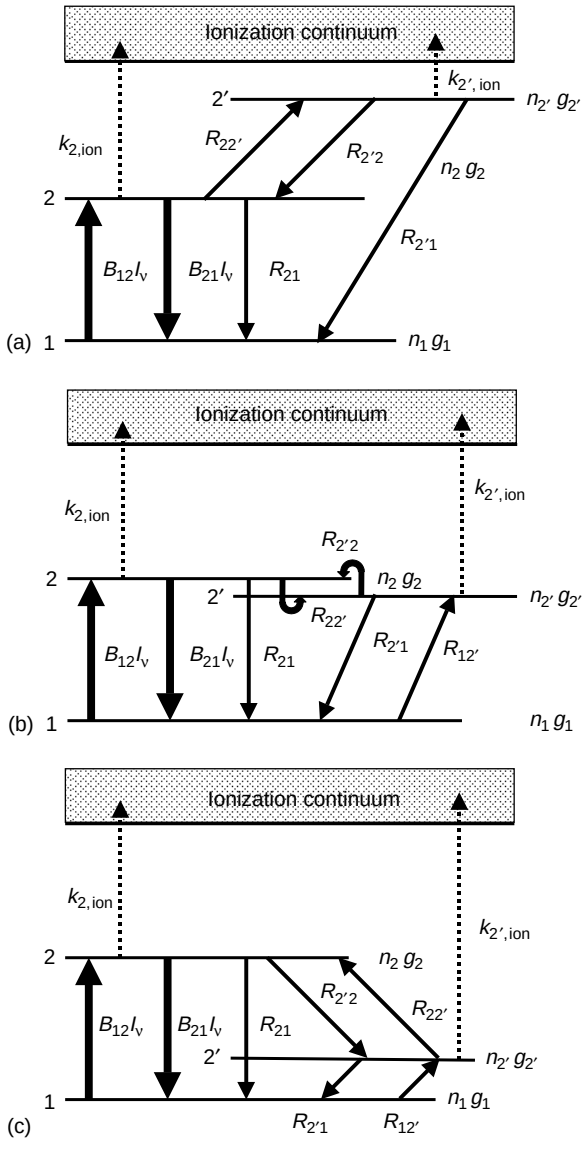

Figure 3 Three different three-level systems exposed to one laser field: (a) the nonlaser-connected level represents a highly excited state (a Rydberg state); (b) the nonlaser-connected state represents an excited state in close proximity to the upper laser-connected level; (c) the nonlaser-connected level represents a low-lying metastable state. The nonlaser-connected rates, R_{ij}, represent collisional excitation rates, collisional de-excitation rates or the sum of collisional de-excitation rates and the spontaneous emission rate, depending on the type of atom. The dashed lines represent collisional ionization rates that are assumed to interact weakly with the three-level systems.

where $\bar{n}^{\text{sat}}_{2(3)}$ and $\bar{n}^{\text{sat}}_{2'(3)}$ are the saturation population fractions of levels 2 and 2′ in a three-level system, respectively, given by Equations (26) and (27):

$$\bar{n}^{\text{sat}}_{2(3)} = \frac{g_2}{g_1 + g_2(1 + \chi_{2'})} \tag{26}$$

$$\bar{n}^{\text{sat}}_{2'(3)} = \chi_{2'}\bar{n}^{\text{sat}}_{2(3)} = \frac{g_2\chi_{2'}}{g_1 + g_2(1 + \chi_{2'})} \tag{27}$$

and where $\chi_{2'}$, in turn, represents the ratio of the filling and emptying rates of the nonlaser-connected level (Equation 28):

$$\chi_{2'} = \frac{R_{22'}}{R_{2'1} + R_{2'2}} \tag{28}$$

These expressions show clearly that not only the density of atoms in the upper laser-connected level is being affected by the laser excitation but also the densities of atoms in the nonlaser-connected levels (represented by level 2′) are being influenced.[77–80] This implies, among other things, that although the laser excitation only induces a transition to one specific state, LIF will in fact be emitted from atoms in virtually all excited states.[78,80] This opens up viable alternatives to resonance fluorescence.

The fraction of atoms in a specific nonlaser-connected excited state can be calculated by inserting the appropriate values for all the quantities involved. The collisional excitation rates can be related to the corresponding de-excitation rates by the concept of detailed balance, as shown in Equation (29):

$$k_{ij} = \frac{g_j}{g_i}\exp\left(\frac{-\Delta E_{ij}}{kT}\right) \quad k_{ji} \tag{29}$$

where ΔE_{ij} is the energy difference between the two states for which $E_j > E_i$ (in J), k is Boltzmann's constant and T is the temperature (in K). However, because not all such rates used to be known, some approximations or generalizations are often used.

One way to estimate the fraction of atoms in a specific nonlaser-connected excited state for cases 1 and 2 is to assume that the collisional rates from level 2′ dominate over any possible spontaneous emission rate from that particular level and that the decay of the nonlaser-connected level to the ground state is smaller than that to the excited state (i.e. $R_{2'1} < R_{2'2}$). One then finds that $\chi_{2'}$ takes a particularly simple form, as shown in Equation (30):

$$\chi_{2'} = \frac{g_{2'}}{g_2}\exp\left(\frac{-\Delta E_{22'}}{kT}\right. \tag{30}$$

Equation (27) then shows that the fraction of atoms in highly excited states (i.e. those that have a large energy deficit to the upper laser-connected level; case 1 atoms) will be significantly lower than that of the upper laser-connected level (because $\chi_{2'} \ll 1$).

For atoms with an excited level in close proximity to the upper laser-connected level (for which the energy deficit normally is small), on the other hand, i.e. for case 2 atoms, the fraction of atoms in this nonlaser-connected level can be comparable to that of the upper laser-connected level (because $\chi_{2'} \approx 1$).

3.3.1.4 Three-level System with a Metastable State The third case of special importance (case 3) is when the nonlaser-connected state represents a metastable state (i.e. a state that is radiatively coupled to the upper laser-connected level but not to the ground state). Elements that have a metastable state are, for example, those with fine-structure splitting of the ground configuration, e.g. Al, In, Ga and Tl, or other low-lying states of the same parity as the ground state, e.g. Cu, Ag and Au. It has been found that the actual lifetimes of metastable states are significantly longer than the typical pulse lengths from the most commonly used laser systems (μs versus ns), even in such highly collisional media as atmospheric pressure atomizers.[72,81–85] These long lifetimes preclude any steady-state condition to be established with the nonlaser-connected level within the laser pulse duration.[59,63] The system of rate equations (Equations 21–23) therefore needs to be solved with its full temporal behavior for atoms with metastable states.[59,63,72,73,85]

A simplified solution for the population of the uppermost laser-connected level in the presence of a lower-lying metastable level can be obtained by assuming that a steady-state condition first is established between the two laser-connected levels within a fraction of the pulse length (following the discussion in section 3.2.1.4). Under this assumption, atoms in the upper laser-connected level will decay to the metastable state at such a rate that the relation between the populations of the two laser-connected levels is not altered.[72,73] The fraction of atoms in the upper laser-connected level then has a time dependence that can be written as shown in Equation (31):

$$\bar{n}_2(t) = \bar{n}_2^{\mathrm{ss}} \quad 1 + \left(\frac{\bar{n}_2^{\mathrm{pss}}}{\bar{n}_2^{\mathrm{ss}}} - 1 \quad \exp\left(\frac{-t}{\tau_{\mathrm{m}}} \right. \right. \tag{31}$$

where $\bar{n}_2^{\mathrm{pss}}$ is the fraction of atoms in the second level directly after the onset of steady state between the two laser-connected levels but before any substantial population has been transferred to the metastable state (i.e. $\bar{n}_2(t) \approx \bar{n}_2^{\mathrm{pss}}$ when $t \approx 0^+$), $\bar{n}_2^{\mathrm{ss}}$ is the steady-state value of the fraction of atoms in level 2 after infinite time (i.e. $\bar{n}_2(t) \approx \bar{n}_2^{\mathrm{ss}}$ when $t \to \infty$) and τ_{m} is the decay time of the laser-connected pair of transitions into the metastable state. These three entities are given by Equations (32–34):

$$\bar{n}_2^{\mathrm{pss}} = \bar{n}_2^{\mathrm{sat}} \frac{B_{12} I_\nu}{B_{12} I_\nu + \bar{n}_2^{\mathrm{sat}} (R_{21} + R_{22'})} \tag{32}$$

$$\bar{n}_2^{\mathrm{ss}} = (\chi_{2'})^{-1} \frac{B_{12} I_\nu}{B_{12} I_\nu + (\chi_{2'})^{-1} (R_{21} + R_{22'})} \tag{33}$$

$$\tau_{\mathrm{m}} = \frac{1}{\bar{n}_2^{\mathrm{sat}} R_{22'}} \tag{34}$$

where $\bar{n}_2^{\mathrm{sat}}$ and $\chi_{2'}$ are the saturation population fraction in a two-level system (Equation 11) and the ratio of the filling and emptying rates of the nonlaser-connected level (Equation 28), respectively. It important to realize that $\chi_{2'} \gg 1$ (and thereby $\chi_{2'}^{-1} \ll 1$) for a metastable state, which implies the condition in Equation (35):

$$\bar{n}_2^{\mathrm{pss}} \gg \bar{n}_2^{\mathrm{ss}} \tag{35}$$

which, in turn, and in practice, implies that the fraction of atoms in the upper laser-connected level rapidly (in this approximation, instantaneously after the onset of the laser excitation) reaches a value similar to what would be the case if there were not any metastable state $\bar{n}_2^{\mathrm{pss}}$ (for saturated condition, the fraction of atoms in this state is momentarily $\bar{n}_2^{\mathrm{sat}}$). The decay to the metastable state will, however, rapidly drain this population of the upper laser-connected state. This is done with a rate roughly given by the decay rate from the upper laser-connected level to the metastable state ($R_{22'}$). Because this rate can be substantial, this implies that the fraction of atoms in the upper laser-connected level can decay to significantly lower values even within the duration of a laser pulse from a pulsed laser system. For excitation with a continuous laser light, a steady-state condition given by $\bar{n}_2^{\mathrm{ss}}$ will prevail. The fraction of atoms in the upper laser-connected level will, under these conditions, only be a fraction ($\chi_{2'}^{-1}$) of the value reached momentarily directly after onset of the excitation.

3.3.1.5 Conclusions This analysis of a three-level system exposed to a single laser field shows clearly that each nonlaser-connected level (irrespective of whether it is a Rydberg state, a closely situated state, or a lower-lying metastable state) influences the populations of the laser-connected levels as well as the requirement of optical saturation (and thereby also the fluorescence and ionization rates). It is not possible, however, to investigate the consequences of this in any detail here. The reader is referred to the literature for a more thorough treatment.[62,63,73,85,86] It can be concluded, however, that the multilevel structure of some atoms explains some (but far from all) of the unusual findings about optical saturation in the literature. Bolshov et al., for example, found that the $6p^2\,{}^3P_0 - 6p\,7s\,{}^3P_1$ transition (at 283.31 nm) in Pb atoms in a GF became saturated at a laser irradiance of 20 kW cm^{-2}, whereas the $3d^6\,4s^2\,{}^5D_4 - 3d^7\,4p\,{}^5F_5$ transition (at 296.69 nm) in Fe showed no sign of saturation up to 300 kW cm^{-2}.[85] As can be concluded from a comparison with the treatment of the two-level system above, both of these saturation irradiances are larger than expected from a simple two-level atom model, possibly indicating influences of nonlaser-connected levels.

The fact that atoms will be promoted also to nonlaser-connected states by the laser excitation has significant but

different implications for the fluorescence and ionization techniques. The effects on the fluorescence technique therefore will be discussed separately from those on the ionization techniques. More detailed analysis of these situations can again be found in the literature.[63,72–74]

3.3.2 *Implications of Nonlaser-connected Levels for Laser-induced Fluorescence*

3.3.2.1 Laser-induced Fluorescence from a Three-level System with a Nonlaser-connected Rydberg State (Case 1) Equations (24–27) and (30) show that the number density of atoms in highly excited nonlaser-connected states is significantly smaller than that of the upper laser-connected level. The density of atoms decreases by one order of magnitude for each 4000 cm^{-1} that separates the two excited states in an atomizer with a temperature of 2500 K. Hence, although LIF in principle will be emitted from virtually all excited states, very few photons will be emitted from excited levels significantly above the upper laser-connected level. Hence, highly excited nonlaser-connected levels generally play a small or insignificant role in the LIF technique.

3.3.2.2 Laser-induced Fluorescence from a Three-level System with a Split Upper Level (Case 2) The situation is different for those atoms that have a nonlaser-connected level in close proximity to the upper laser-connected level. Two levels in close proximity not only have a small energy difference, i.e. $\Delta E_{ij} \ll kT$, but often they also have a fast mixing rate, i.e. $R_{2'2}$ and $R_{22'} \gg R_{2'1}$. This justifies the assumption above, leading to Equation (30), which, in turn, implies that $\chi_{2'}$ often takes a value equal to the ratio of the degeneracy factors of the two levels, i.e. $g_{2'}/g_2$. This implies, furthermore, that the population density on the nonlaser-connected level, in fact, can be comparable to (or even exceed) that of the upper laser-connected level. If this nonlaser-connected level also is radiatively coupled to the ground state, this implies that LIF often can be detected from levels in close proximity to the upper laser-connected level without any significant loss of signal strength. This is of special importance for those elements that have a ground state with no fine-structure splitting, or no other suitable low-lying levels to which fluorescence can be measured from the upper laser-connected level (for which resonant fluorescence otherwise would have been the only possibility). In conclusion, detection of fluorescence from nonlaser-connected excited states in close proximity to the upper laser-connected level might often be a viable alternative to resonance detection.

3.3.2.3 Laser-induced Fluorescence from a Three-level System with a Metastable State (Case 3) The rate of emission of photons per atom for a system with a metastable state is given directly by the time evolution of the fraction of atoms in the upper laser-connected level (Equation 31) and can therefore, in general, be written according to Equation (36):

$$\dot{\overline{N}}_{\mathrm{ph}}(t) = \bar{n}_2^{\mathrm{ss}} A_{22'} \quad 1 + \left(\frac{\bar{n}_2^{\mathrm{pss}}}{\bar{n}_2^{\mathrm{ss}}} - 1 \quad \exp\left(\frac{-t}{\tau_{\mathrm{m}}} \right. \right. \tag{36}$$

This simplifies considerably under saturated conditions to the form given in Equation (37):

$$\dot{\overline{N}}_{\mathrm{ph}}(t) = \frac{A_{22'}}{\chi_{2'}} \quad 1 + \bar{n}_2^{\mathrm{sat}} \chi_{2'} \exp\left(\frac{-t}{\tau_{\mathrm{m}}} \right. \tag{37}$$

where $\chi_{2'} \gg 1$ (and hence $\bar{n}_2^{\mathrm{pss}} \gg \bar{n}_2^{\mathrm{ss}}$) has been assumed for a metastable state. This implies that the rate of emission of fluorescence from a metastable system rapidly decreases following the onset of laser illumination, as shown in Figure 4.

Although the initial rate of emission of fluorescence photons from a saturated metastable system (i.e. for $0 < t \ll \tau_{\mathrm{m}}$) is given by Equation (38):

$$\dot{\overline{N}}_{\mathrm{ph}}(t \approx 0^+) = \bar{n}_2^{\mathrm{sat}} A_{22'} \qquad (\approx A_{22'}) \tag{38}$$

which is similar to that of a saturated two-level system, the rate of emission becomes significantly lower when the system has reached steady-state conditions (i.e. for

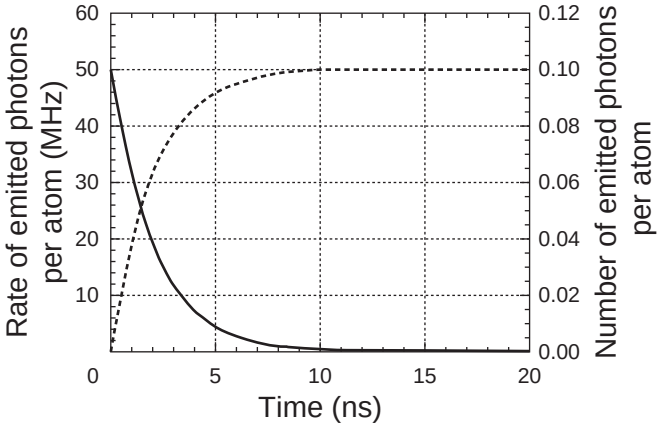

Figure 4 The rate of emitted photons per atom, $\dot{\overline{N}}_{\mathrm{ph}}(t)$ (———), and the number of emitted photons per atom, $\overline{N}_{\mathrm{ph}}(t)$ (- - - -), for the three-level system with a metastable state displayed in Figure 3(c) as a function of exposure time (according to Equations 37 and 41 respectively). It is assumed that the populations on the laser-connected levels become locked to each other instantaneously (to their relative steady-state population fractions). The following parameters have been used: $g_1 = g_2 = g_{2'}$, $A_{22'} = 10^8\ \mathrm{s}^{-1}$, $R_{22'} = 1\ \mathrm{GHz}$ (implying that $\tau_{\mathrm{m}} = 2\ \mathrm{ns}$ and $Y_{22'} = 0.1$), $R_{2'1} = 1\ \mathrm{MHz}$ and $R_{2'2} \ll R_{2'1}$ (implying that $\chi_{2'} = 10^3$).

$t \gg \tau_m$), according to Equation (39):

$$\dot{\overline{N}}_{ph}(t \gg \tau_m) = \frac{A_{22'}}{\chi_{2'}} \qquad (\ll A_{22'}) \tag{39}$$

This implies that the photon emission rate will be significantly smaller for continuous-wave excitations than for pulsed excitations. Because no steady-state condition will be established within the time period of a laser excitation from a pulsed laser, it is more appropriate to discuss the concept of the number of emitted fluorescence photons per atom within a time period t of laser excitation, $\overline{N}_{ph}(t)$, according to Equation (2). One then finds that $\overline{N}_{ph}(t)$ can be written as shown in Equation (40):

$$\overline{N}_{ph}(t) = \bar{n}_2^{pss} \left\{ \left(1 - \frac{\bar{n}_2^{ss}}{\bar{n}_2^{pss}} \quad \tau_m \quad 1 - \exp\left(\frac{-t}{\tau_m} \right. \right. \right. \\ \left. \left. \left. + \frac{\bar{n}_2^{ss}}{\bar{n}_2^{pss}} t \right\} A_{22'} \right. \tag{40}$$

for a system with a metastable state. This expression again simplifies under pulsed, saturated conditions, as shown in Equation (41):

$$\overline{N}_{ph}(t) = Y_{22'} \quad 1 - \exp\left(\frac{-t}{\tau_m} \right. \tag{41}$$

where τ_m, as above, is the decay time of the laser-connected pair of transitions given by Equation (34) and we again have used $\chi_{2'} \gg 1$ for a metastable state, assumed that the fluorescence is measured primarily over the most intensive part of the fluorescence pulse (i.e. over a time period t that does not substantially exceed the lifetime of the metastable state) and introduced $Y_{22'}$ as the fluorescence transition yield for the transition between the upper laser-connected level and the metastable state, defined according to Equation (42):

$$Y_{22'} = \frac{A_{22'}}{R_{22'}} \tag{42}$$

Equation (41) (see Figure 4) shows that the number of photons emitted per atom from this atomic system under pulsed conditions is limited to $Y_{22'}$. This value is reached (within 63%) after a time equal to the decay time of the laser-connected pair of transitions, i.e. τ_m, which, in turn, is roughly given by the inverse of the decay rate from the upper laser-connected state to the metastable state. Because many decay rates are slightly below or in the low gigahertz range, it implies that the actual length of typical pulses from pulsed laser systems (5–30 ns) corresponds fairly well to an efficient detection of fluorescence from atoms with metastable states in atmospheric pressure atomizers. This implies that no atom with a metastable state can emit more than one photon per laser pulse under excitation with pulsed laser light. It also implies that the fluorescence yield for the entire detection process of atoms with metastable states is given by the fluorescence transition yield for the transition between the upper laser-connected level and the metastable state, $Y_{22'}$. For continuous-wave excitations, on the other hand, atoms with metastable states give significantly poorer detectability because the number of photons per atom does not increase as much with increased illumination time as does the amount of noise detected.

3.3.3 *One-step Excitation Laser-enhanced Ionization*

The influence of the various nonlaser-connected levels will be somewhat different for the LEI technique.

3.3.3.1 Ionization Through Nonlaser-connected Rydberg States Although the high-lying levels play a minor role in the fluorescence technique, they are of major importance for the LEI technique. As stated above (near Equations 3 and 4), the total ionization signal depends on the ionization rate density, $\dot{n}_{ion}$, which in turn is given by a sum over all state-specific ionization rates. Adapting this reasoning to the current model with two excited states (one laser-connected and one not), we find that the ionization rate density can be written as shown in Equation (43):

$$\dot{n}_{ion}(t) = k_{2,ion} n_2(t) + k_{2',ion} n_{2'}(t) \tag{43}$$

It is commonplace to assume that the collisional ionization rate scales in a similar way to that of exciting collisions between two bound states, as was discussed in proximity to Equation (29), i.e. as $\exp(-\Delta E_{i,ion}/kT)$, where $\Delta E_{i,ion}$ is the energy deficit to the ionization limit.[57,87,88] This implies that even without any detailed knowledge about each of the two ionization rate constants in the expression above, they can be related to each other (assuming that $E_{2'} > E_2$) according to Equation (44):

$$k_{2,ion} \approx k_{2',ion} \exp\left(\frac{-\Delta E_{22'}}{kT} \right. \tag{44}$$

This implies that although the collisional ionization rate constant from level 2 is smaller than that of level 2′ by the factor $\exp(-\Delta E_{22'}/kT)$, Equations (24–27) and (30) show that the population of level 2′ is smaller than that of level 2 by almost the same factor (given by $\chi_{2'}$), which is expressed as Equation (45):

$$\bar{n}_{2'} \approx \bar{n}_2 \frac{g_{2'}}{g_{2'}} \exp\left(\frac{-\Delta E_{22'}}{kT} \right. \tag{45}$$

This implies, in turn, that all excited levels whose populations have an approximate relation to the upper laser-connected level by the activation energy factor (i.e. as given by Equation 30) also contribute approximately

equally to the total ionization rate. This can be interpreted as if the ionization simultaneously takes place through a large number of ''channels''. Following the treatment above, it would be easy to conclude that all levels whose energy is higher than that of the upper laser-connected level would qualify for being such a channel.

However, this argument gives rise to a contradiction. Because there are, in principle, an infinite number of excited states above the upper laser-connected level, this model predicts that the ionization rate would be infinite. In reality this is, of course, not the case. It has been argued in the literature that atoms that are not too highly excited can (or will) exist in atmospheric pressure atomizers (one reason is that their size makes them very vulnerable in the highly collisional environment). It has therefore been found suitable to introduce a certain cutoff principal number, representing the highest excited state that can contribute to the ionization.[74] This is equivalent to introducing an *effective* collisional ionization rate that would act from only one excited state, most conveniently the upper laser-connected level.[57] Hence, although the excited atoms will redistribute rapidly among a large number of excited states due to collisions,[78,80] all of which will have their own state-specific ionization rate, we will model LEI by an effective collisional ionization rate, $k_{i,\mathrm{ion}}^{\mathrm{eff}}$, acting only from the upper laser-connected level, as shown in Equation (46):

$$\dot{n}_{\mathrm{ion}}(t) = k_{2,\mathrm{ion}}^{\mathrm{eff}} n_2(t) \tag{46}$$

3.3.3.2 Ionization Yield for One-step Laser-enhanced Ionization The electrical field across the interaction region in LEI will rapidly separate ions from electrons. The recombination rate will therefore be insignificant during the laser pulse. If we allow for a substantial ionization during the laser pulse (so that the number density of neutral atoms, n_{atom}, will decrease during the laser pulse), the number density of ions, n_{ion}, needs to be expressed in terms of the number density of neutral atoms, n_{atom}, as shown in Equation (47):

$$n_{\mathrm{tot}} = n_{\mathrm{atom}} + n_{\mathrm{ion}} \tag{47}$$

where n_{tot} is the total number density of analyte species. This implies that the equation for the number density of ions can be written according to Equation (48):

$$\dot{n}_{\mathrm{ion}}(t) = k_{2,\mathrm{ion}}^{\mathrm{eff}} \bar{n}_2 [n_{\mathrm{tot}} - n_{\mathrm{ion}}(t)] \tag{48}$$

where $\bar{n}_2$ is given by Equation (10) for atoms lacking metastable states.

Following the convention above of working with normalized entities, it is convenient to define an ionization yield, $\overline{N}_{\mathrm{ion}}(t)$ (often also denoted Y^{ion} in the literature) as the fraction of analyte species that ionize as a consequence of the laser illumination, as shown in Equation (49):

$$\overline{N}_{\mathrm{ion}}(t) = \frac{n_{\mathrm{ion}}(t)}{n_{\mathrm{tot}}} \tag{49}$$

Solving Equation (48) for the number density of ions, one finds that the ionization yield in this case is given by Equation (50):

$$\overline{N}_{\mathrm{ion}}(t) = 1 - \exp(-k_{2,\mathrm{ion}}^{\mathrm{eff}} \bar{n}_2 t) \tag{50}$$

(under the assumption that $\bar{n}_2$ takes its steady-state value within a time that is short with respect to the ionization depletion time).

We can see here that the ionization yield depends strongly on the effective collision rate, which in turn depends on the energy deficit to the ionization limit of the upper laser-connected level, and that a substantial degree of ionization can be obtained when the interaction time is equal to, or larger than, the inverse of this effective collision rate.

Although the ionization limit of various elements differs, we can conclude that most elements have an ionization limit that is around $50\,000–70\,000\,\mathrm{cm}^{-1}$ above the ground state (with an important exception for the alkali elements). Typical UV light (with 250–330 nm wavelength) excites ground-state atoms to states with energies around $30\,000–40\,000\,\mathrm{cm}^{-1}$. This implies that the energy deficit of atoms excited by UV light often is in the $20\,000–30\,000\,\mathrm{cm}^{-1}$ region (i.e. a few electronvolts). It is reasonable to assume that the effective collisional ionization rate is in the $10^4–10^5\,\mathrm{s}^{-1}$ range for states a couple of (or a few) electronvolts below the ionization limit.[57,88] Because $\bar{n}_2$ is close to unity for a saturated transition (on an orders of magnitude basis) and a typical laser illumination time is 10 ns, we get an ionization yield of about $10^{-4}–10^{-3}$ for a typical one-step LEI excitation. This shows that it is not possible in general to reach any high degrees of ionization from one-step LEI (unless the atoms have unusually low ionization limits).

3.3.3.3 Means to Improve on the Low Ionization Yield for One-step Laser-enhanced Ionization An obvious remedy to this is to excite the atoms to levels closer to the ionization limit.[88–90] Although this can be done by a one-step excitation using far-UV light, it has been found to be less suitable. One reason is that it is getting increasingly difficult to create far-UV photons the further down in wavelength one goes. Another reason is that far-UV photons have a greater tendency to give rise to background signals from flame molecules and concomitant elements than visible photons have (which thus would counteract the expected increase in S/N ratio). It has been found more convenient (and considerably more beneficial in terms of selectivity) to excite the atoms

by light from two simultaneously pumped tunable lasers, alternatively referred to as stepwise excitation or two-step excitation (although also a rich variety of more complex denotations exists in the literature).[91]

3.4 Three-level Atoms Exposed to Two Laser Fields

There are several treatments of two-step (or even multistep) excitation of atoms in the literature, some of them with slightly different approaches.[73,74,92–94] Although the presence of metastable states might require a more sophisticated description and solution than those given here, we will describe two-step excitations of atoms in the simplest possible manner that can still correctly account for ionization depletion. This implies that we will only consider direct two-step excitations (i.e. those that share a common intermediate level), neglect the influence of metastable trapping levels, and solve the system of rate equations solely under steady-state conditions. The reader is referred to the literature for analyses of other configurations.[94]

3.4.1 General Description of a Three-level System Exposed to Two Laser Fields

Consider a situation in which the atoms are described as three-level atoms (with levels denoted 1, 2 and 3 as shown in Figure 5) between which two resonantly tuned laser fields are acting. In line with the discussion above, it is often a good approximation to assume that the laser excitations rapidly (i.e. within the first nanosecond of the laser pulse) create a population equilibrium among the three laser-connected states. The ionization is then assumed to act on this strongly coupled three-level system.

3.4.1.1 Rate Equation System for a Three-level System Exposed to Two Laser Fields The rate equations for a three-level system exposed to two resonant laser fields, as depicted in Figure 5, can in general be written as in Equations (51–53):

$$\frac{dn_3}{dt} = R_{13}n_1 + (R_{23} + B_{23}I_\nu^{23})n_2 - (R_{31} + R_{32} + B_{32}I_\nu^{23})n_3 \quad (51)$$

$$\frac{dn_2}{dt} = (R_{12} + B_{12}I_\nu^{12})n_1 - (B_{21}I_\nu^{12} + R_{21} + B_{23}I_\nu^{23} + R_{23})n_2 + (R_{32} + B_{32}I_\nu^{23})n_3 \quad (52)$$

$$n_{atom} = n_1 + n_2 + n_3 \quad (53)$$

3.4.1.2 Solution of the Rate Equation System for a Three-level System Exposed to Two Laser Fields Because ionization rates most often are considerably smaller than typical de-excitation and laser-induced rates, this three-level system can be solved under steady-state conditions. Under the assumption that smaller exciting collisions can be neglected in comparison with large de-exciting collision rates or laser-induced rates, the fraction of neutral atoms in the uppermost laser-connected state (level 3) can be written according to Equation (54):

$$\bar{n}_3 = \bar{n}_3^{sat}\frac{B_{12}I_\nu^{12}B_{23}I_\nu^{23}}{(B_{12}I_\nu^{12} + a)(B_{23}I_\nu^{23} + b) + c} \quad (54)$$

where the various coefficients are given in Table 2.

3.4.2 Two-step Excitation Laser-enhanced Ionization

As long as there is any appreciable laser illumination on the upper transition, the main contribution to the ionization rate will be from the uppermost laser-connected level. Hence, for a two-step excitation, we can write the ionization rate as shown in Equation (55):

$$\dot{n}_{ion}(t) = k_{3,ion}^{eff}n_3(t) \quad (55)$$

This implies that the ionization yield for a two-step excitation, $\overline{N}_{ion}^{2s}(t)$, can be written in an analogous way

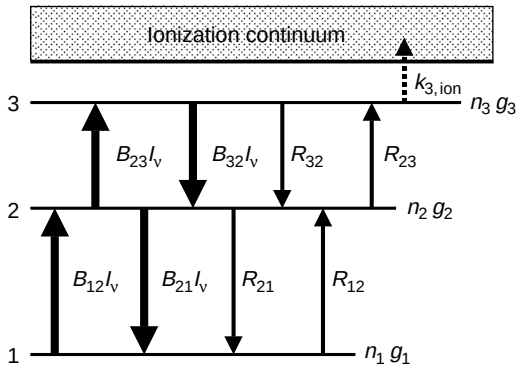

Figure 5 A three-level system exposed to two resonant laser fields.

Table 2 Coefficients for Equation (54)

Coefficient	Expression
$\bar{n}_3^{sat}$	$\frac{g_3}{g_1 + g_2 + g_3}$
a	$\bar{n}_3^{sat}R_{31} + \bar{n}_2^{sat}R_{21}$
b	$\frac{\bar{n}_3^{sat}}{\bar{n}_{2,off}^{sat}}(R_{31} + R_{32})$
c	$(\bar{n}_3^{sat})^2(R_{31} + R_{32})\left(R_{21} - \frac{R_{31}}{\bar{n}_{2,off}^{sat}}\right.$
$\bar{n}_{2,off}^{sat}$	$\frac{g_2}{g_1 + g_2}$

to that for a one-step excitation, namely according to Equation (56):

$$\overline{N}_{\text{ion}}^{2s}(t) = 1 - \exp(-k_{3,\text{ion}}^{\text{eff}}\bar{n}_3 t) \tag{56}$$

Because the second excitation step often can excite the atoms to states a couple of electronvolts closer to the ionization continuum than can be achieved with one-step excitation (a second-step photon with a wavelength of 500 nm corresponds to a decreased energy deficit of 2.5 eV), the (effective) ionization rate can be increased by two to three orders of magnitude by using two-step excitation. In addition, because the fraction of neutral atoms in the uppermost laser-connected state can still take values close to unity under saturated conditions, this implies that the ionization yield can approach unity within the time of a laser pulse excitation.[72,95] Two-step excitation is therefore the most efficient means to achieve strong LEI signals. As will be discussed below, two-step excitation not only increases the sensitivity but also improves the selectivity.

3.4.3 Two-step Excitation Laser-induced Fluorescence

Although one-step excitation often gives rise to a sufficiently high sensitivity (number of photons produced per atom) for the fluorescence technique, there are situations when a one-step excitation is not appropriate. One such example is when the atomic structure is such that resonance fluorescence is the only possibility when the atom is excited by light from one laser (as is the case for Cd, for example). Another example is when the fluorescence appears in the visible part of the region, a part of the spectrum in which substantial amounts of background emission often exist (e.g. from the blackbody radiation in ETAs). In this case the technique will not reach its optimum performance. A remedy for these situations is to use two-step excitation also for the LIF technique.[41,63] Because the fluorescence from a two-step excitation often appears in the UV or far-UV regions (thus, far from the exciting wavelengths, as well as blackbody radiation from the atomizer), two-step excitation LIF used to be characterized by freedom from scattered laser light and background radiation (see below).[72,94]

The expressions above (Equation 54 and Table 2) show that a strong fluorescence signal is expected from the upper laser-connected state (or from any collisionally coupled excited state) directly following the onset of laser excitation. However, as is shown by Equations (55) and (56), the atomic system can also suffer from an appreciable ionization depletion during the laser pulse if the uppermost laser-connected level has too small an energy deficit compared to the ionization limit. Taking this drain of the number of neutral atoms during the laser pulse into account, the rate of emission of fluorescence photons per atom (in this case, from the uppermost laser-connected level to any lower-lying level, here denoted i) can be expressed as shown in Equation (57):

$$\dot{\overline{N}}_{\text{ph}}(t) = A_{3i}\bar{n}_3(t)\exp(-k_{3,\text{ion}}^{\text{eff}}\bar{n}_3 t) \tag{57}$$

This shows that the emission rate of fluorescence photons is highest directly following the onset of laser illumination, after which it decays with a decay time given by the inverse of the product of the effective ionization rate and the fraction of atoms in the uppermost excited level. This behavior resembles that for the LIF signal from a three-level system with a metastable state exposed to one laser field (Equation 37). This is not unexpected because the drain to ionization acts in an equivalent manner to that to a metastable state on these timescales.

Using Equation (2), the total number of photons emitted per atom from such a system (in a time t) can then be expressed as shown in Equation (58):

$$\overline{N}_{\text{ph}}(t) = \frac{A_{3i}}{k_{3,\text{ion}}^{\text{eff}}}[1 - \exp(-k_{3,\text{ion}}^{\text{eff}}\bar{n}_3 t)] \tag{58}$$

This shows that for laser excitation to states whose (effective) collisional ionization rate is smaller than the inverse of the laser pulse illumination time (i.e. $k_{3,\text{ion}}^{\text{eff}} < t^{-1}$), the number of photons emitted per atom is approximately given by Equation (59):

$$\overline{N}_{\text{ph}}(t) = A_{3i}\bar{n}_3 t \tag{59}$$

which is equivalent to neglecting ionization depletion during the laser excitation.

If the excitation is performed to a state whose (effective) collisional ionization rate is significantly larger than the inverse of the laser pulse illumination (i.e. $k_{3,\text{ion}}^{\text{eff}} \gg t^{-1}$), the number of photons emitted per atom cannot exceed that shown in Equation (60):

$$\overline{N}_{\text{ph}}\left(t \gg (k_{3,\text{ion}}^{\text{eff}})^{-1}\right) = \frac{A_{3i}}{k_{3,\text{ion}}^{\text{eff}}} \tag{60}$$

due to ionization losses, which again is a similar behavior to that of the metastable three-level system exposed to one laser field (Equation 42). This shows that if the upper laser-connected level experiences too large an effective ionization rate, ionization losses can put a limitation on the number of photons emitted per atom from two-step excitation.

3.4.4 Fluorescence Dip

If the system of rate equations for a three-level system exposed to two laser fields (sharing a common intermediate level; Equations 51–53) is solved for the population

of the intermediate state, one can see clearly that this population (and hence the fluorescence from the intermediate state) will experience a decrease with the onset of the second-step laser.[94,96] This phenomenon, which is similar to the ionization depletion phenomenon discussed above, has been termed *fluorescence dip* and has been prophesied to have extraordinary diagnostic capabilities, e.g. for the determination of transition probabilities between excited states or photoionization cross-sections.[97] The first investigations of the fluorescence dip phenomenon were made by Omenetto et al.[97,98]

The relative decrease of the fluorescence from the intermediate state is termed relative fluorescence dip, ζ_2. It was concluded that the maximum relative fluorescence dip for an optically saturated two-level system (such as that depicted in Figure 5) under steady-state conditions (i.e. neglecting the influence of ionization depletion) can be written as shown in Equation (61):

$$\zeta_2 = \frac{\bar{n}_2[I_\nu^{12} > (I_\nu^{12})^{\text{sat}}, I_\nu^{23} = 0) - \bar{n}_2(I_\nu^{12} > (I_\nu^{12})^{\text{sat}}, I_\nu^{23} > (I_\nu^{23})^{\text{sat}}]}{\bar{n}_2[I_\nu^{12} > (I_\nu^{12})^{\text{sat}}, I_\nu^{23} = 0]} = \frac{g_3}{g_1 + g_2 + g_3} \tag{61}$$

For the case of Sr (with a $5s^2$–5s 5p–5s 6d excitation, an upper state about 6000 cm^{-1} below the ionization limit, and degeneracy factors of 1, 3 and 5 for the three states, respectively), the optically saturated relative fluorescence dip should be 55% under the assumption that a steady-state situation prevails. It was found experimentally by Axner et al., however, that significantly larger optically saturated relative fluorescence dips could be obtained in an acetylene–air flame (up to 85%).[99] Simultaneous measurement of the fluorescence dip and the increase in the ionization signal, coupled to an analysis based upon time-dependent rate equations for the fluorescence dip, revealed that the system was exposed to significant ionization depletion during the laser illumination time. This finding does not only show that substantial depletion of the intermediate laser-connected level follows the onset of the second excitation step, but it is also an example of the importance of taking ionization depletion into account when excitations of atomic systems to high-lying states in high-collision media are considered. This is in line with the theoretical predictions of ionization and fluorescence depletion made by Omenetto et al.[100]

3.5 Conclusions – Justification of the Use of the Rate Equation Formalism

Section 3 has given the basic mechanisms for the production of fluorescence photons and charges (ions and electrons) primarily for the LIF and the LEI techniques, respectively, for both one- and two-step excitations in high-collision media. The description was given in terms of the number of photons emitted per atom in time t, $\overline{N}_{\text{ph}}(t)$, and the fraction of analyte species that ionizes as a consequence of the laser illumination, $\overline{N}_{\text{ion}}(t)$, commonly also termed the ionization yield. It has been shown that both of these entities can reach values close to unity under a rather broad range of optimized pulsed conditions.

The description above was given in the rate equation formalism with no particular justification of its applicability. A correct representation of the excitation and de-excitation processes of atoms by narrow-band laser light can normally only be accomplished by using the so-called density matrix formalism.[101–103] In this formalism, the atomic system is described by its quantum mechanical wavefunction. Solved with its full temporal dependence, the density matrix formalism is capable of describing both the time dependence of the dipole moments as well as the atomic level populations. The density matrix formalism is therefore able to describe effects that appear as a result of coherence between the induced dipole moments and the populations, e.g. rapid population oscillations between various states, commonly termed Rabi oscillations. Solved under steady-state conditions for the dipole moments, the density matrix equations (often termed Bloch equations in this context) revert to pure population equations of similar type to those of the rate equation formalism. These population equations are able to describe how an atomic system responds to narrow-band laser light of an arbitrary wavelength (i.e. including wing excitations) and can therefore predict, among other things, profile shapes of atomic transitions. They can also describe coherent contributions in the excitation process, e.g. two-photon excitations and dynamic Stark effects, such as broadening, splitting, and shifts of transitions induced by the laser light.[73,102–109]

The density-matrix-based population equations, in turn, revert to the rate equations for sufficiently large laser bandwidths and collisional rates in the system (rather rapidly for one-step excitations but more reluctantly for two-step excitations).[73,103,107–109] The rate equation formalism assumes, among other things, that there is a constant irradiance of the light across the width of the transition. This implies, in turn, that the formalism assumes tacitly that the light is fully in resonance with the transition. It is therefore unable to predict excitation profiles correctly. The high collisional rates in atmospheric pressure atomizers, together with the fact that light with bandwidths larger than that of the analytes often is used, often justify (although not always) the use of the simpler rate equations for calculations of various-level populations.[73,103,107] It has been commonplace, however, to use the rate equation formalism for description of excitation processes in more or less all fields of analytical spectrometry, even under situations

when not all of the requirements for its use are fully fulfilled. The most common justification of the use of the rate equation formalism seems too often to be its simplicity rather than its appropriateness. As a result, the main part of the theoretical description in the literature of the interactions between laser light and atoms for analytical applications (as well as the description given above) has been cast in the framework of the rate equation formalism.[35,59–63,71–74,94,99,103,110–112] One exception to this is the work by Boudreau et al.[113] These authors have developed a user-friendly program, based upon a fully time-dependent density matrix model that describes stepwise excitations of atoms with degenerate states under collision-dominated conditions,[108] that can predict phenomena such as dynamic Stark effects (splitting shifting and broadening) and two-photon transitions.[113] The reader is referred to the literature for a proper description of how density matrix formalism is adapted to the conditions applicable to analytical spectrometry.[103–109,114,115]

4 LASER-INDUCED FLUORESCENCE SPECTROMETRY

4.1 Introduction

LIF spectrometry, commonly also termed LEAFS when applied to analytical spectrometry, is based upon resonant absorption of laser radiation by the analyte and detection of the subsequently emitted fluorescence radiation.

Because laser light has a high spectral irradiance, the population of excited analytes can be made significantly larger when laser light is being used for excitation than when other light sources are used. This implies a high sensitivity and low detection limits for laser-based spectroscopic techniques in general and for LIF/LEAFS in particular. LOD in the low picogram per milliliter range, and in the femtogram range, have been obtained for a number of elements when used with a GF as atomizer. The technique has also been applied to other atomizers (e.g. flames and plasmas) but not with the same impressive LOD. In addition, because laser light consists of narrow bands, laser-based spectroscopic techniques generally have good selectivity. Because the LIF/LEAFS technique often uses a dispersive element for detection of the fluorescence (most often a spectrometer), the elemental selectivity is excellent. The LIF/LEAFS technique also has a large dynamic range (often five to seven orders of magnitude).

The first time fluorescence was used for analytical purposes was in 1964 by Winefordner et al.[116,117] The first laser-based instrumentation for fluorescence was constructed by Fraser and Winefordner in 1971 and 1972.[118,119] The first time frequency-doubling was used to extend the applicability of laser-based spectroscopic techniques for analytical applications was in 1973 when Kuhl and Spitschan detected Mg, Ni, and Pb in a flame by the LIF/LEAFS technique.[120] Since then, a large number of papers have been published in the field, with significant contributions by the groups of Winefordner, Omenetto and Michel as well as others.

The range of elements that can be detected conveniently by the LIF/LEAFS technique has been restricted mostly to metals, although a few nonmetal elements also have been studied. In addition, the fluorescence technique has been used for the detection of molecules; it is then referred to as LEMOFS[38] or occasionally LEMFS[121]). Moreover, in hot plasmas, where the degree of ionization is fairly high, ionic fluorescence has been used as an alternative to atomic fluorescence in order to improve the sensitivity or extend the number of accessible elements. The isotope selectivity of the LIF/LEAFS technique, on the other hand, is in general not very impressive under normal conditions because most optical isotope shifts are smaller than typical laser bandwidths or Doppler and collisional broadening mechanisms in most atomizers (although a few exceptions exist, e.g. among the actinides and lanthanides[122]). Pulsed laser systems (primarily those with pulse duration in the nanosecond range, i.e. excimer-pumped dye laser systems and, to some extent, Nd:YAG-pumped dye laser or OPO systems) have shown the highest applicability for the LIF/LEAFS technique, although some impressive works also have been done with continuous-wave diode lasers. In addition, a variety of excitation–detection schemes have been investigated. Although the most common excitation–detection scheme is one-step excitation followed by Stokes direct-line or stepwise detection (mainly due to its simplicity and high fluorescence yield), some elements require two-step excitation in order to obtain the most favorable conditions.

This implies that in order to characterize fully the LIF/LEAFS technique with respect to its properties and applicability to analytical spectrometry, one can thus envisage a huge number of possible combinations of laser, atomization and detection instrumentation, excitation and detection schemes, as well as detection procedures. This naturally leads to a large spread in performance capabilities. Numerous detection limits of the LIF/LEAFS technique for a variety of elements have therefore been published in the literature throughout the years. It is not possible to cite all individual contributions to the continuously ongoing development of the LIF/LEAFS technique in a presentation like this so we have chosen to present the specific analytical attributes and characteristics of the LIF/LEAFS technique (sensitivity, types of noise, LOD, background reduction techniques, LDR, etc.) by referring to only a selection of papers in the field

(but including those that report on the lowest LOD for each element).

Prior to that, however, a survey of previous reviews of the LIF/LEAFS technique is given, followed by a description and discussion of ''typical'' LIF/LEAFS instrumentation. Because the LIF/LEAFS technique can be used also as a diagnostic probe for specific physical or chemical phenomena (e.g. state-specific ionization rates, fluorescence yields, or recombination rates), some comments on the diagnostic capabilities of the LIF/LEAFS technique are given in a special section.

4.2 Previous Reviews

The field of conventional-source-excited atomic fluorescence spectrometry (i.e. not using laser light for excitation) has been reviewed in detail by Butcher et al.[123] Smith et al. have made an extensive compilation of a variety of atomic fluorescence spectrometry results (up to 1989) arranged by elements.[124] The field of laser-induced/excited fluorescence, i.e. LIF/LEAFS, has been reviewed by several authors. An early review (from 1978) that serves as a good introduction to the practical aspects and early applicability of the LIF/LEAFS technique is that of Winefordner.[36] Contemporary reviews that are focused on the basic properties of the fluorescence technique are those of Omenetto and Winefordner.[35,125] Another early landmark review paper is that of Butcher et al.[37] Omenetto has written many interesting papers about various aspects of the LIF/LEAFS technique[35,48,66,125–129] but of special importance is the work by Omenetto and Human[66] in which many important issues for the analytical use of LEAFS in flames, furnaces and inductively coupled plasmas (ICPs) are discussed in detail (e.g. the influence of peak power, pulse duration, spectral bandwidth and repetition rate on the detectability, together with the choice of excitation line and the optical arrangement of the detection system). This paper also contains many ''handy recipes'' for the LIF/LEAFS technique. Sjöström[130] has written a comprehensible review of LIF/LEAFS in ETAs and, together with Mauchien, has given a critical view of the entire field of laser-based spectrometric techniques for analytical spectrometry.[131] Interesting reviews not only of LIF/LEAFS but also of other laser-based spectroscopic techniques have been given by Falk,[132,133] as well as by Omenetto.[134] The most recent review of the LIF/LEAFS technique is that by Hou et al.[38] The group of Winefordner has presented a series of papers critically discussing the ultimate possibilities of laser-based techniques in general and LIF/LEAFS in particular for analytical spectrometry, as well as their possibilities to achieve SAD.[53–55,129,135]

4.3 Instrumentation

4.3.1 Typical Set-up

A typical experimental set-up for LIF/LEAFS is shown in Figure 6. Laser light is directed into an atomizer in which the sample is atomized. The emitted fluorescence is collected by an optical system and directed into a spectrometer. The transmitted light is detected by a sensitive detector – most often a photomultiplier tube

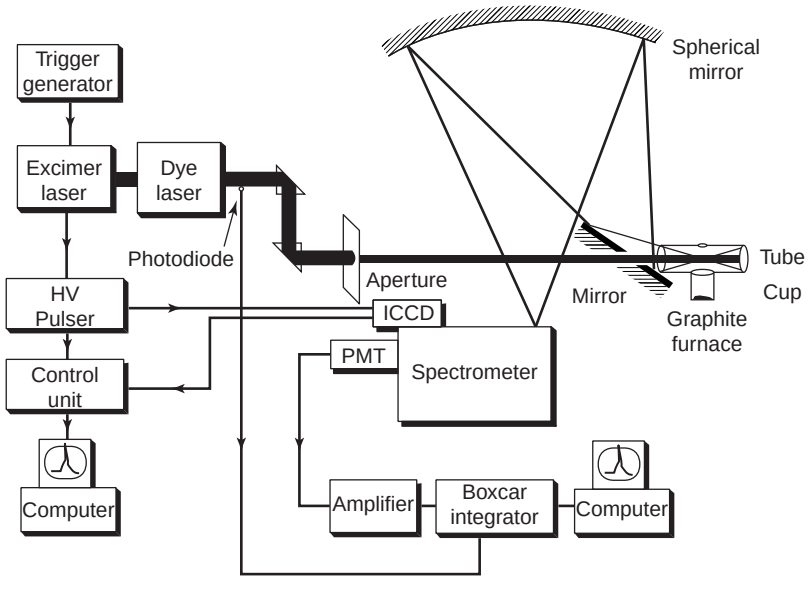

Figure 6 A typical experimental set-up for LIF/LEAFS in a GF. (Reproduced from A. Marunkov, N. Chekalin, J. Enger, O. Axner, *Spectrochimica Acta*, **49B**, 1385–1410, Copyright 1994, with permission from Elsevier Science.)

(PMT), whose signal is subsequently amplified, and infrequently an intensified charge-coupled device (ICCD). When pulsed lasers are employed, gated detection, most often synchronously triggered with light from the laser, needs to be used (accomplished by a boxcar integrator for the case of a PMT). The output signal from the boxcar integrator is finally fed into a computer for data storage and handling.

4.3.2 Atomizers

4.3.2.1 Flames The first LIF/LEAFS measurements for analytical purposes during the 1970s used flames as atomizers.[35,36,70,71,86,118–120,136–147] The prime reasons were the suitability of flames for analysis of liquid samples, their simplicity and reliability, and the fact that they have a continuous mode of operation (and hence provide a continuously refreshing distribution of atoms). However, although LIF/LEAFS in flames was prophesied early on to be a technique with extraordinary detectability and selectivity,[35] the actual detection limits turned out not to be as impressive as first anticipated.[37,38,60,64–67,78,81,115,148–158] The main reason is that the flame itself is far from the ideal atomizer for the LIF/LEAFS technique. Flames give rise to a rather large amount of background radiation and act as a source of scattering of laser light, which reduce the detection limits (see discussion below). However, LIF/LEAFS has been found to be a useful tool for the diagnostics of flames.

4.3.2.2 Electrothermal Atomizers The ETA, mostly represented by the GF, has several advantages in comparison with flames. The one that affects sensitivity the most is that the vaporized atomic cloud is held more concentrated, which implies that the atomic residence time (and thereby the interaction time) is longer. Other advantages are that only microliter volumes of samples are needed and that staged heating in an inert atmosphere makes analysis of samples with high concentrations of concomitant elements possible. A specific advantage with the ETA for the LIF/LEAFS technique is that it provides a rather quiet environment with a minimum amount of background radiation. It was therefore found early on that the LIF/LEAFS technique had its largest potential for analytical spectrometry in combination with the ETA.[37,66,77,79,80,85,121,130,132,156–227]

4.3.2.3 Other Atomizers The development of the ICP during the last decade has considerably facilitated the determination of trace elements in many types of matrices using conventional detection techniques. The most significant advantages with the ICP are that it provides an efficient atomization, excitation and ionization environment. It has therefore taken over the role of the flame as the workhorse for a large variety of analytical investigations in many laboratories. This has implied that the ICP has been of interest also for the community of laser spectroscopists. LIF/LEAFS has therefore been applied to various types of plasmas.[37,60,66,94,122,157,228–230] In addition, other atomizers, e.g. discharges[231–235] and cells,[178] have been of interest for analytical laser-based spectrometric techniques. The analytical performance of LIF/LEAFS systems containing these types of atomizers, however, has not been found to rival those of ETA systems. Therefore, only a limited number of studies in this area have been pursued.[37,60,66,94,122,157,178,228–235]

4.3.3 Laser Sources

A variety of lasers have been used for the LIF/LEAFS technique over the years. Pulsed laser systems (primarily those with pulse duration in the nanosecond range, i.e. excimer-pumped dye laser systems,[37,66,67,80,120,137,153,155–189,191–199] N_2-pumped dye laser systems,[118,119,138–141,148,152,214–224] and Nd:YAG-pumped dye laser[37,77,79,85,190,200–209] or OPO systems[154,227,236]) have shown the highest applicability for the LIF/LEAFS technique, although some impressive studies also have been done with Cu-vapor lasers,[209–213] ion-pumped dye laser systems,[142–145,151] flashlamp-pumped dye lasers[146,147,149,150] and continuous-wave diode lasers.[225]

The prime reasons for the success of pulsed lasers for LIF/LEAFS (as well as the ionization techniques, see below) are: versatility in producing UV light (because tunable UV laser light is normally produced by frequency-doubling visible light in a nonlinear crystal using processes that are strongly nonlinear, then, with a conversion efficiency that increases with irradiance, pulsed lasers, which have a significantly higher irradiance than continuous-wave systems, can more easily produce UV wavelengths); the possibility of producing a large signal over a short period of time (thereby maximizing the S/N ratio); and the fact that they can saturate the transitions more easily than continuous-wave systems.

A disadvantage with pulsed laser systems is that the duty cycle is rather poor, so that only a small fraction of all atoms entering the atomic reservoir will be exposed to laser light. Consider, for example, the case of a 10-Hz Nd:YAG laser illuminating an air–acetylene flame (whose rise time is around $10\,m\,s^{-1}$ and width is around 1 cm). The analyte atoms are being distributed over an area of $100\,cm^2$ in-between each laser firing. If the laser beam diameter is between 3 and 4 mm, it covers an area of about $10\,mm^2$. Hence, not more than 10^{-3} of all the atoms in the flame can interact with the laser light. The situation improves if excimer-pumped laser systems are used.

Modern excimer lasers can produce laser pulses at repetition rates of up to a few hundred hertz (although often at a reduced pulse energy). An excimer-pumped dye laser system running at 100 Hz can then potentially still interact with only a few percent of all the atoms in the flame using an enlarged beam. This is, however, not such a devastating drawback for LIF/LEAFS as might be expected because the pulsed detection also reduces the noise in the system (noise is measured only over $2\,\mu s\,s^{-1}$ using a 100-Hz system and a 20-ns fluorescence detection gate).[237]

Copper vapor lasers, on the other hand, which produce pulses at kilohertz repetition rates, can potentially interact with all atoms in a sample. However, a severe drawback of metal-vapor-pumped laser systems is that not the entire spectral region can be accessed.[238]

The most commonly used type of laser for LIF/LEAFS is the XeCl excimer-pumped dye laser, mainly due to a combination of a high repetition rate (up to 500 Hz), a high peak power (typically up to 1 MW in the visible spectrum and tens of kilowatts in the UV range) and pulse durations that are a few times longer than those of Nd:YAG lasers.[191] Another issue of importance in this respect is that there are commercial excimer-pumped dye laser systems in which the dye laser cuvette and dye pump form a combined unit that can be removed/exchanged rather easily. With access to a few sets of dye laser cuvette and pump units, a change of dye can be accomplished rather conveniently, i.e. within a few minutes (a procedure that often takes many hours with other dye laser systems).

The recent development of new pulsed tunable laser sources has had some impact on the field of analytical spectrometry. The development of OPO systems, with a rapidly scanned wide spectral tuning range (220–2200 nm) and pulse energies of tens of millijoules per pulse in the visible region and several millijoules in the UV region, takes the laser-based spectroscopic techniques one step further towards multielement determinations.[154,227,236] It is not clear, a priori, however, that pulsed dye laser systems with nanosecond pulse lengths provide the optimum excitation source for the LIF/LEAFS technique for *all* elements in *all* instances. As will be discussed further below, they can, under certain conditions, be rivaled by other types of laser systems (depending on the dominating type of noise).

The development of diode lasers has, in particular, opened up the possibility of constructing new, powerful and compact LIF/LEAFS instrumentation. Although most applications of diode lasers for analytical purposes so far have been pursued within the field of FM or WM absorption techniques (which is described in section 6), Zybin et al. established 10- and 20-fg LOD for Li and Rb in a graphite tube, respectively, detected by diode-laser-excited LIF/LEAFS.[225]

4.3.4 Detection Equipment

4.3.4.1 Optical Systems As was concluded by Yuzefovsky et al.[239] and clearly summarized by Hou et al.,[38] there are five main requirements for an efficient optical system, namely: a high light collection efficiency in both the UV and visible regions; a good ability to discriminate between background and analytical signals; minimal losses of the analytical signal at the detection wavelength; easy-to-align optics; and inexpensive and commercially available optical components.

The optical systems that have been used the most with the LIF/LEAFS technique over the years are the three that are depicted schematically in Figure 7 (exemplified using a graphite tube as the atomizer). Most early experimental set-ups directed the fluorescence into the spectrometer in a right-angle mode of detection, as displayed in Figure 7(a) (see references 35, 37, 65, 66, 77, 85, 115–121, 132, 137–153, 157–174, 203–208, 213–222). Later set-ups have instead had front-surface detection using a pierced mirror (see

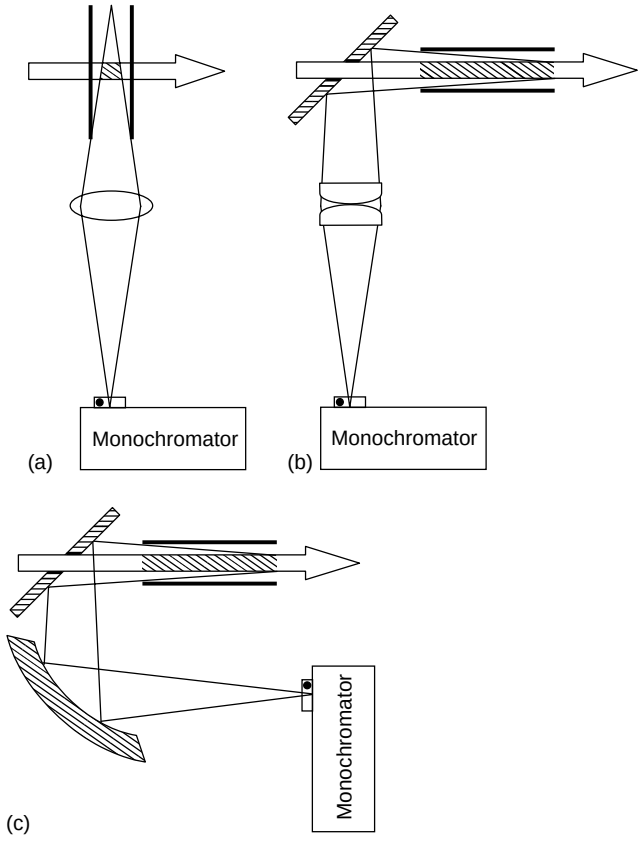

Figure 7 Three different types of optical systems that have been used with the LIF/LEAFS technique: (a) right-angle detection using refractive optics; (b) front-surface detection using refractive optics; (c) front-surface detection using reflective-only optics (i.e. a curved mirror).

references 37, 79, 80, 154–156, 158, 171, 173–202, 209–213, 223–227), either flat (together with a suitable set of lenses) (see references 37, 79, 156, 158, 173–185, 190–198, 200–202, 209–212, 224–226) as in Figure 7(b), or curved[80,155,171,186–193,199,209,223] (ellipsoidal[171,191–193,199,209] or spherical[80,155,186–189]) as in Figure 7(c). The front-surface mode of detection has been found to increase the volume of illuminated atoms seen by the spectrometer and hence the sensitivity of the system.[37,156,209] A front-surface mode of detection also has the advantage that it reduces so-called post-filtering effects (see below).

Most early experimental set-ups used a single biconvex lens for directing the fluorescence light into the spectrometer, whereas later ones have used either multilens systems (Figure 7b) or reflective optics (Figure 7c). It was calculated by Farnsworth et al. (using computer modeling of the collection efficiency of LIF) that the use of a single biconvex lens for collection of fluorescence (as shown in Figure 7a) would not provide the best collection of the fluorescence owing to aberrations (mainly spherical aberration).[240] They found that single-point collection efficiencies in general differ significantly from those that would be expected based on the solid angle subtended by the lens. An optical system consisting of a pair of planoconvex lenses, as indicated in Figure 7(b), was found to be far more effective in imaging the interaction volume onto the entrance slit of the spectrometer than a single biconvex lens. The best performance was obtained by a pair of achromatic doublets. The single-point collection efficiency was found to range from below 15% of the ideal value for a biconvex lens to nearly 100% for a pair of the achromats. Some of these predictions have been confirmed experimentally by Liang et al.[183]

In addition, because all optical material has dispersion (i.e. a wavelength-dependent index of refraction), instrumentation using refractive components (e.g. lenses) for the collection of fluorescence light will experience chromatic aberrations (implying that the light collection efficiency will have a wavelength dependence). The use of achromatic lenses therefore decreases significantly the chromatic aberrations in the system. Because there is no dispersion in fully reflective optical systems, the use of curved mirrors instead of lenses for focusing the light onto the spectrometer slit has the advantage that there are no chromatic aberrations.[187]

4.3.4.2 Detectors The light filtered by the spectrometer/monochromator is being detected by a sensitive detector. Because the vast majority of all the LIF/LEAFS measurements have been done using pulsed excimer- or Nd:YAG-pumped laser systems, and typical de-excitation rates in atmospheric pressure atomizers are in the range of hundreds of megahertz to a few gigahertz, the duration of the emitted fluorescence will be of the same order of magnitude as the duration of the laser pulses (i.e. most often a few to some tens of nanoseconds). Furthermore, because these types of laser have repetition rates of some tens or hundreds of hertz, the duty cycle is low (typically 10^{-6}–10^{-5}). Gated detection therefore has to be used in order to obtain the optimum S/N conditions.

By far the most common detector used is the PMT,[37,65,66,77,79,85,116–121,132,137–150,152–182,190–224,227] because it has a number of useful properties: it is sensitive, inexpensive and easy to use. The only disadvantage with a PMT is that it is not capable of providing any fluorescence spectrum – it can only give rise to one data point per laser firing (corresponding to the fluorescence integrated over the bandpass of the spectrometer at the center wavelength chosen). When PMTs are used the gating is done by a boxcar integrator, whereas diode arrays and charge-coupled device (CCD) detectors, which sometimes have been used with the LIF/LEAFS technique, need to be equipped with an intensifier in order to provide the necessary gating.

ICCD detectors[80,185–190] or diode arrays[151,226] have the capability of detecting an entire fluorescence spectrum for each laser shot, which can be an advantage when samples with complex matrices are to be detected. One example of the versatility of an ICCD detector is shown in Figure 8. The figure shows a pair of fluorescence spectra from Ti atoms excited by pulsed 264.108-nm light in a GF. Figure 8(a) displays the situation in a normal Ar atmosphere whereas Figure 8(b) shows the corresponding spectrum when Ar was substituted by N_2.[80] As can be concluded from the figure, the fluorescence spectrum is fairly rich. To aid in the identification of the fluorescence peaks, however, Ar was occasionally substituted for N_2, which is a better quencher of excited atoms. The peaks that decrease the most (labeled 1–11 in Figure 8) originate from ''indirect'' transitions (i.e. originating from upper levels that are different from that accessed by the laser), whereas the others refer to ''direct'' transitions (i.e. those that originate from the upper laser-connected level). This study led to the conclusion that most of the fluorescence spectra from Ti are dominated by ''indirect'' transitions. In addition, many direct transitions were found to be inferior to other indirect transitions, making it virtually impossible even to make a qualified guess of the optimum excitation–detection combination prior to the experiment. The results indicate also that the collisional redistribution processes amongst the excited levels are, in general, faster than typical fluorescence rates for Ti in GFs (which supports the discussion about rapid collisional redistribution processes among excited states in section 3.3.1.1 above).

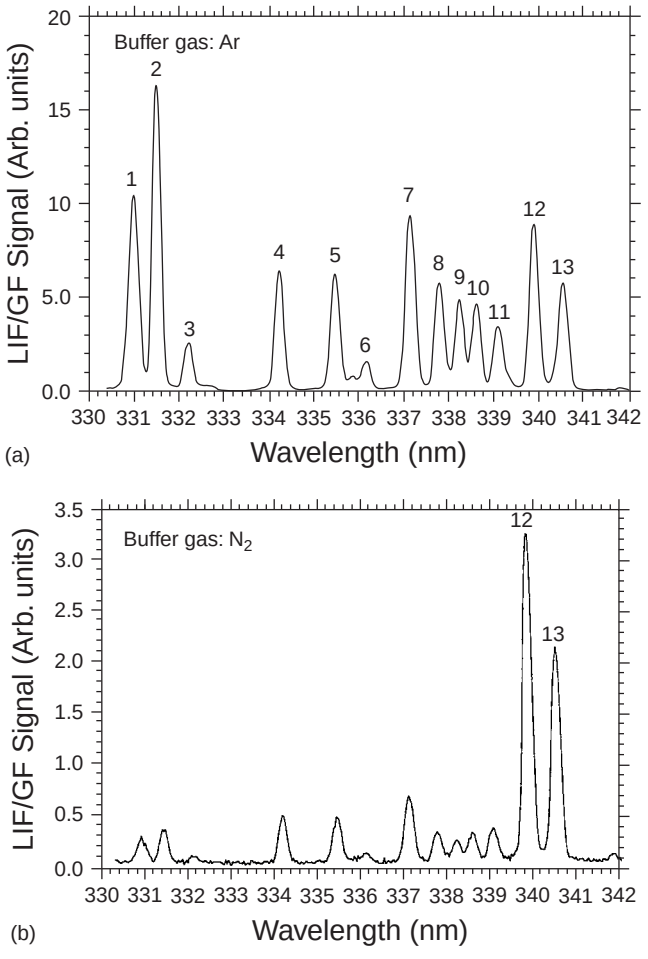

Figure 8 Two LIF/LEAFS/ICCD spectra from Ti atomized in a GF excited by 264.108-nm light to the $3d^2\,4s\,4p\,^3D_1$ state at 37 852.021 cm^{-1} when Ar (a) and N_2 (b) have been used as buffer gas. A full identification of the fluorescence lines is given in Ljung et al.[80] N_2 is a stronger quencher than Ar, so peaks 12 and 13 can be referred to as ''direct'' transitions. Note that several of the indirect peaks are equally as strong as the direct ones in the Ar atmosphere. (Reproduced from P. Ljung, E. Nyström, J. Enger, P. Ljungberg, O. Axner, *Spectrochimica Acta*, **52B**, 675–701, Copyright 1997, with permission from Elsevier Science.)

4.3.5 Types of Fluorescence Transitions

A variety of different excitation–detection modes have been used throughout the years, mainly dictated by the atomic structure of the analyte to be investigated. Although attempts have been made to define a clear nomenclature for these various excitation–detection modes,[119,241] a slight disorder has been prevalent. Lately, however, a new International Union of Pure and Applied Chemistry (IUPAC) notation for the description of various excitation–detection processes, applicable to fluorescence as well as ionization techniques, has been released.[242] A few of the most commonly used excitation–detection schemes for the LIF/LEAFS technique are given in Figure 9. The reader is referred to the literature, e.g. any of the works by Omenetto and Winefordner,[35,241] for a more extensive description of possible excitation–detection schemes for the LIF/LEAFS technique.

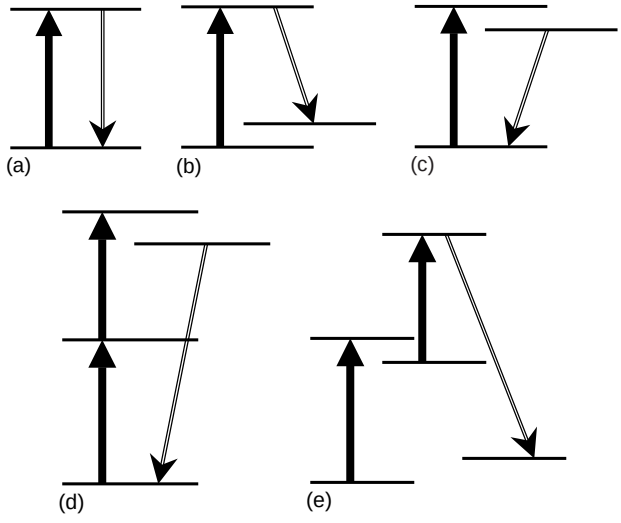

Figure 9 The most commonly used excitation–detection schemes for the LIF/LEAFS technique. Thick and thin arrows represent laser excitations and fluorescence, respectively. (a) Resonance fluorescence ($R_{12}F_{21}$); (b) Stokes direct line fluorescence ($R_{13}F_{32}$); (c) Stokes step-wise line fluorescence ($R_{13}F_{21}$); (d) connected double-resonance stepwise line fluorescence ($R_{12}R_{24}F_{31}$); (e) disconnected double-resonance direct line fluorescence ($R_{14}R_{35}F_{52}$). Both conventional notation[241] and the new IUPAC terminology (within parentheses)[242] are used.

4.4 Analytical Performance

The detectability of a certain species in a given atomizer depends on a number of factors. On the one hand, we have the sensitivity, i.e. the relation between the measured signal and the concentration (or amount) of the analyte in the sample. On the other hand, there is always noise in the system. The relation between the two determines the LOD for the system. Because both the sensitivity and limiting noise will differ between various types of instrumentation, the LOD for a given element will depend on the particular atomizer and laser system used, as well as their mode of operation.

4.4.1 Sensitivity

The sensitivity for a given instrumentation is normally defined as the rate of change of the signal with amount/concentration of analyte. Because PMTs are often capable of detecting single events and LIF/LEAFS measurements very seldom are limited by detector noise

(see below), it is often more convenient to define an overall efficiency of a system: as the number of events detected per concentration unit of the analyte in the sample for continuous-mode atomizers; and as the number of events detected per mass unit of the analyte or per atom in the atomizer for discontinuous atomizers (furnaces, discharges etc.). There are, however, several closely related entities defined and discussed in the literature (e.g. the efficiency of detection and efficiency of measurement).[37,65,129,174,214,243,244] The detection efficiency, for example, introduced by Alkemade,[243,244] refers to the probability that a given atom appearing in the probed volume produces an event during the probing time. However, we will not scrutinize these issues in detail here. It is sufficient to conclude that the overall sensitivity can be written as a product of several individual efficiency terms, e.g. the sample introduction efficiency, the atomization efficiency, the illumination efficiency (the fraction of atomized atoms that is illuminated by the laser light), the number of photons emitted per atom (as defined in section 3), the solid angle over which fluorescence is collected, the light collection efficiency of the optical system, the transmission of the spectrometer, and the quantum efficiency of the detector.[37,129,174,243,244]

One illustrative example of the concept of sensitivity is the determination by Wei et al.[174] of the overall efficiency for detection of Pb atoms in an ETA by the LIF/LEAFS technique for three different experimental set-ups: right-angle detection combined with dispersive detection; front-surface detection combined with dispersive detection; and front-surface detection combined with nondispersive detection.[174] These authors found overall efficiencies for their three set-ups of $4.3 \times 10^{-7}, 2.0 \times 10^{-6}$ and 1.2×10^{-5}, respectively. The major reason for the relatively low numbers for the overall efficiency is the small solid angle over which fluorescence is collected by the detection system. Their results show that although front-surface detection has a higher overall efficiency than right-angle detection, nondispersive detection systems can, in turn, have a higher overall efficiency than dispersive detection.

4.4.2 Noise and Background Signals

Noise in a measurement can be classified as either extrinsic or intrinsic. Extrinsic noise is the noise that arises from a nonspecific background signal that is present even in the absence of analyte. Extrinsic sources of noise comprise, for example, thermal noise, background emission, stray light and concomitant scatter. The intrinsic noise, on the other hand, is due to the inherent statistical fluctuations in the number of atoms present in the volume probed by the laser.[35,132,244] Under the most favorable conditions, only the intrinsic noise will limit the detectability of a certain species. Although discussed in the literature in the assessment of the ultimate detection power of laser-based techniques (in particular, in connection with the concept of SAD)[54,129,243,244] and because of the existence of many other types of noise (see below), the intrinsic noise will only occasionally limit an actual measurement under real conditions.

4.4.2.1 Thermal Noise Thermal noise consists of electronic noise from the instrumentation, e.g. detector noise and preamplifier noise, and is an omnipresent source of noise that only occasionally (when a small number of events is to be detected) is the limiting factor for the LIF/LEAFS technique. Because PMTs can detect a single event above its noise level, a low number of signal photons more often creates a situation limited by the intrinsic noise than by the thermal noise in experiments in which very few background photons are created. Hence, it is the consensus that PMT-based systems are seldom limited by the thermal noise.

A gated CCD detector (i.e. an ICCD), on the other hand, which consists of many thousands of individual detector elements each with its own noise, can, when used in a binned or integrating mode (as is most often the case when hooked up to a spectrometer), give rise to a higher amount of detector noise. Hence, although not yet fully confirmed, especially regarding the most modern ICCD systems, it seems most plausible that the detection of weak signals using ICCDs in binned modes can be limited by the thermal noise from the detector rather than the intrinsic noise.[187]

4.4.2.2 High-frequency (Radio-) Interferences High-power pulsed laser systems can produce significant amounts of high-frequency (radio-) interferences that can interfere with the detection equipment. Although this type of noise is more of a nuisance for the LEI technique (in which minuscule signal currents are measured in unshielded environments), it will only occasionally limit the detectability for the LIF/LEAFS technique (especially when laser or detector systems are used that are not fully shielded electrically).

4.4.2.3 Nonlaser-induced Background Signals Non-laser-induced background signals from the atomizer comprise emission from constituents in the atomizer or blackbody radiation. Although flame emission can be significant at the wavelength regions where the most frequent flame species emit (e.g. OH and C_2), the general trend is that the atomizer background signals decrease with wavelength.[38] For ETAs, the nonlaser-induced background signals consist mostly of the blackbody radiation, which can limit the detectability primarily at wavelengths above 300 nm.[130] A remedy for this is thus

to strive to obtain detection wavelengths as far as possible into the UV region.[176]

4.4.2.4 Stray Light Stray light is defined as laser light reflected into the detection system from various parts of the instrumentation (but excluding scattering from the sample). Most of the stray light originates from scattering from various optical surfaces. This source of noise is therefore most severe when closed atomizers (e.g. a GF) are used (in which the light has to pass through a window before it enters the interaction volume). Various attempts to design the instrumentation (i.e. the atomizer[132,137,164] or the optical system[239,240]) so that light originating from the entrance and exit windows does not reach the detector have therefore been made. One possibility is to use an optical system with a short depth-of-focus and the atomizer with its windows at the Brewster angle as far from the interaction region as possible. This implies that only a small part of the stray light will be imaged onto the spectrometer slit.

Stray light is always the dominating source of noise for resonant LIF/LEAFS but can also affect the measurements when other types of excitation–detection modes are used (especially when the detection wavelength is close to the excitation wavelength).[144,149,159,176,207,214,245] Remedies for this are to use excitation–detection wavelengths that are as widely separated as possible and to use a double monochromator (which has a better stray-light rejection ratio than ordinary monochromators/spectrometers).[200] However, these measures are insufficient if the stray light originates from ASE from the laser (which is a broad emission background from dye lasers that appears in addition to the stimulated light that is produced in the system) or if it originates from so-called ''environmental'' (or ''white'') fluorescence. Environmental fluorescence arises when UV light impinges upon optical components that generate fluorescence photons in the 290–420 nm region.[209] An efficient solution to this problem is to use double-resonance fluorescence (as was shown in Figures 9d and 9e), because the fluorescence light then can have a significantly shorter wavelength than the excitation light (described in more detail below).[94,176] The ordinary spectrometer instrumentation can then be complemented with either a suitable filter (which thus only transmits the fluorescence wavelength and not the laser excitation wavelengths) or with a solar-blind photomultiplier (which is sensitive only to UV light).[72,176]

4.4.2.5 Concomitant Scatter Scattering of laser radiation off particles introduced by the sample is referred to as concomitant scatter.[245] This phenomenon is closely related to the concept of molecular fluorescence, which may be generated either by the species commonly present in the atomizer (e.g. OH, NO, or C_2 molecules in flames) or by those being produced by the sample (metal chlorides, metal oxides or metal hydroxides).[38,139–141,145,246,247] These types of background can therefore be significant (and severely limit the detectability) when samples containing large amounts of nonvolatile and undissociated matrices are detected. Concomitant scatter is generally more severe in flames than in ETAs because the flame, with its chemical environment, has a larger tendency to produce nonvolatile species than has an inert gas atmosphere.[38] Atomizers with large temperature gradients (e.g. graphite filaments or rods) also give rise to potentially larger concomitant scatter and molecular fluorescence than those with good temperature homogeneity (e.g. transversely heated graphite atomizers), because the cooler region may allow recombination of matrix elements or the formation of scattering species from the analyte.[38,208]

One of the very few documented examples of the existence of concomitant scatter in a GF is shown in Figure 10 in which Sb is detected in riverine and estuarine water by LIF/LEAFS in a GF using an ICCD detector.[189] Figure 10(a) shows fluorescence spectra from (tenfold-diluted) estuarine water (containing 540 pg mL^{-1} Sb and 2×10^7 times more Na, i.e. 11 600 μg mL^{-1}) for three different laser wavelengths (in resonance with the Sb transition as curve A, and detuned −30 and +30 pm as curves B and C, respectively). The peak in curve A at around 289 nm, marked with an arrow in the figure, originates from Sb, whereas the other regularly spaced peaks in Figure 10(a) belong to some unidentified molecular species in the sample. Figure 10(b), in which riverine water (containing 275 pg mL^{-1} Sb and about 10^4 times less Na, i.e. 1.86 μg mL^{-1}) is investigated, shows no such background, indicating that it originates from the matrix in the estuarine water. Despite the clearly visible background signal from the matrix in the estuarine water at this particular excitation wavelength, the most prevalent situation for the LIF/LEAFS technique is the situation shown in Figure 10(b), i.e. in which there is no matrix background signal.

4.4.2.6 Nonanalyte Atomic Fluorescence Nonanalyte atomic background fluorescence is a rather uncommon phenomenon for the LIF/LEAF technique in ETAs (although it exists to a certain degree in flames owing to their richer chemical composition).[140] The main reason for this is the inherent double selectivity of the LIF/LEAFS technique. A potentially interfering species needs to have a wavelength overlap in both the excitation and detection steps in order to interfere with the analyte signal. Because atoms generally have very narrow excitation profiles, background signals from

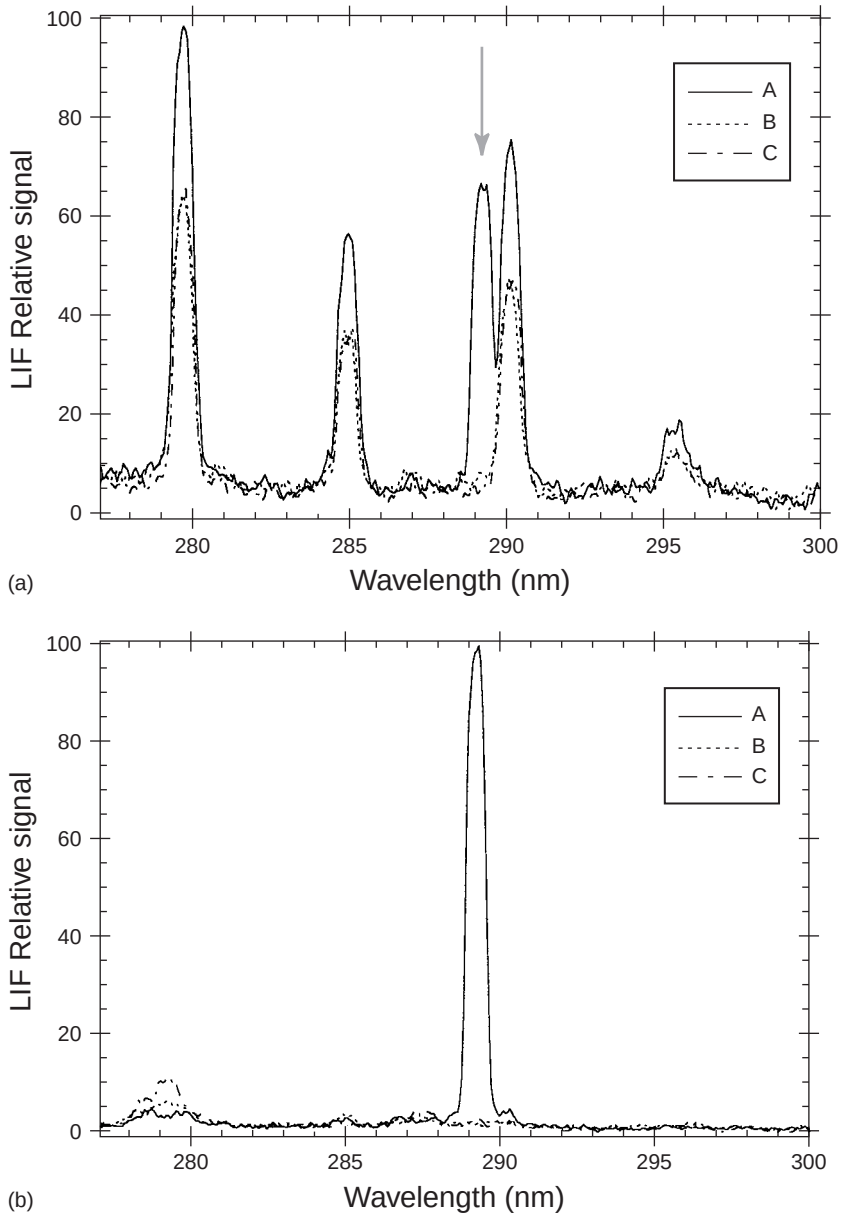

Figure 10 Two LIF/LEAFS/GF fluorescence spectra from tenfold-diluted estuarine water (a) and riverine water (b), respectively. The laser wavelength is 231.147 nm, which is fully resonant with the $5p^3\,{}^4S_{3/2}-5p^2\,6s\,{}^4P_{1/2}$ transition in Sb for the two curves A. Curves B and C correspond to detuned laser excitation (−30 and +30 pm, respectively). The Sb contents in the two samples were assessed to be 0.275 ± 0.01 and 0.54 ± 0.07 ng mL^{-1}, whereas the Na contents were 1.86 and 11 600 μg mL^{-1}, respectively. The peak in curve A at around 289 nm originates from Sb, whereas the other regularly spaced peaks in (a) belong to some unidentified molecular species in the estuarine sample. (Reproduced by permission of the Royal Society of Chemistry from J. Enger, A. Marunkov, N. Chekalin, O. Axner, *Journal of Analytical Atomic Spectrometry*, **10**, 539–549 (1995).)

nonanalyte atoms are extremely rare, although molecular background signals occasionally have been found (as was exemplified in Figure 10).

4.4.2.7 Conclusions It can be concluded that the laser stray light (e.g. ''environmental'' fluorescence), atomizer emission, or concomitant scatter limits the detectability of LIF/LEAFS in most practical situations.

4.4.3 Means to Reduce Background Signals

In order to reduce the influence of background signals, a variety of background-reducing or background-correction techniques have been developed and investigated.

4.4.3.1 Double-resonance Fluorescence As was concluded above, the limiting noise in many situations

is laser stray light, concomitant scattering or atomizer background emission noise (e.g. blackbody radiation in furnaces). A possible approach to decrease the influence of atomizer background emission noise is to bring the fluorescence far into the UV region. This often requires laser photons of very short wavelength (sometimes even below 220 nm), which can be difficult or cumbersome to produce. An alternative approach, which can reduce the influence of all these sources of noise significantly, is to use a double-resonance (two-step) excitation (i.e. to excite the atoms simultaneously with two laser pulses), as was shown in Figure 9(d) and (e). The greatest advantage of double-resonance excitation is that the fluorescence light can have a significantly shorter wavelength than the excitation light.[72,94,176]

Vera et al.[79] defined the ideal excitation–detection conditions, as shown in Equation (62):

$$\begin{cases} \lambda_1, \lambda_2 > 320\,\text{nm} \\ \lambda_F < 320\,\text{nm} \end{cases} \quad (62)$$

where λ_1 and λ_2 are the two excitation wavelengths and λ_F is the wavelength at which the fluorescence is detected. This allows for the use of a solar-blind PMT (i.e. one that does not respond to light above 320 nm) and implies that the detector will respond neither to any laser stray light nor to any concomitant scatter. Double-resonance excitation has thus a greater potential to bring the LIF/LEAFS technique closer to the predicted intrinsic LOD than ordinary one-step excitation–detection schemes.

The first double-resonance excitation LIF/LEAFS measurement was made on Pb by Miziolek and Willis in 1981 under the name of saturated optical nonresonant-emission spectrometry (SONRES).[77] Lead atoms were excited by 283.3- and 600.2-nm light. They concluded that very short wavelength photons (e.g. down to 202 nm) could be generated by collisional energy transfer among excited states. They studied a number of transitions and found the 261.4-nm transition to be the strongest. They quoted an order-of-magnitude LOD of 1 fg for Pb. Because this work was performed under an unconventional name, it took the community of scientists several years to realize the importance of this mode of detection for LIF/LEAFS. A number of two-step excitation LIF/LEAFS measurements were made in 1988 by Omenetto et al. (establishing LOD for Tl, Pb and Cd in a GF of 2, 5 and 18 fg, respectively),[176] by Leong et al. (who studied both direct and disconnected double-resonance excitations in Pb and made a comparison with one-step excitation),[200] and in 1989 by Vera et al. (who assessed LOD for In, Ga and Yb of 2, 1, and 220 fg, respectively).[79] Additional two-step excitation experiments have been performed by several authors thereafter, e.g. Sjöström et al. when detecting V,[194] Axner and Rubinsztein-Dunlop when investigating Cr,[173] and Petrucci et al. when detecting Au.[158]

4.4.3.2 Multichannel Background Monitoring Various approaches for the monitoring of background signals have been developed and scrutinized.

A multichannel background correction technique for pulsed LEAFS with a GF was developed by Sjöström (a poor man's multichannel analyzer but with the benefit of using PMTs).[130] The background was measured simultaneously with the analyte signal using three optical fibers positioned at the exit slit of the monochromator and coupled to two different PMTs (one for the center fiber and one for the two flanking fibers). Blackbody radiation, scattered laser light and molecular fluorescence could be corrected for by this simultaneous background-correction technique. Gallium was detected with an LOD of 50 fg.

An alternative approach was taken by Remy et al.[201] They split the fluorescence light, using a beamsplitter, into two different monochromators – one tuned to the analytical transition and one slightly detuned so as to measure the background from scattered laser light, furnace emission, concomitant scatter, etc. The authors conclude that this leads to improvements in LOD and precision. The drawback is that the beamsplitter reduces the analytical signal by a factor of 2.

A more powerful approach to this concept is to use an ICCD to detect the fluorescence.[80,186–189] An ICCD detector has several advantages over a PMT. It can detect several wavelengths simultaneously, so it can be used for fast and convenient investigations of the fluorescence spectra to find the most sensitive excitation–detection wavelength combination from atoms with complex atomic structure. It can be used to increase the absolute sensitivity of the LIF technique in a GF as compared to PMT detection by allowing simultaneous detection of the fluorescence at several wavelengths. It can monitor and correct for background signals from matrix interferences, blackbody radiation and scattered laser light both at and around the wavelength of detection, which is of special importance when samples with high concentrations of matrices are being analyzed. This leads to an improved spectral selectivity. In addition, owing to the delayed read-out of the ICCD detector (delayed in time with respect to the laser pulse), it has a high immunity from radiofrequency pick-ups emitted from the pump laser. Moreover, by storing a number of consecutive fluorescence spectra from within one furnace heating, the time development of spectra from the analyte as well as from any matrix constituent in the sample giving rise to background signals can be studied. Finally, a two-dimensional ICCD detector can be used for spatial studies of atomization and diffusion processes in the

GF if combined with an imaging spectrometer. There are a few disadvantages of ICCD detectors, the most significant of which are: they can have a lower sensitivity than a PMT, resulting from a lower quantum efficiency of the photocathode; and they have a limited read-out rate (around one or a fraction of a hertz for a full two-dimensional read-out, or up to a few hundred hertz for binned read-outs). They produce a huge amount of data, which necessitates the use of computers with large data-storage capabilities, and they are relatively expensive.

The first use of an ICCD together with LIF/LEAFS in an ETA was the investigation of Ni.[187] The most sensitive and versatile excitation and detection wavelengths were identified. The LOD of Ni by LIF/LEAFS in an ETA could thereby be improved by two orders of magnitude. The ICCD detector was also compared with an ordinary PMT. The LOD were found to be 15 and 10 fg for ICCD and PMT detection, respectively. The simultaneous monitoring of entire fluorescence spectra by the use of an ICCD made possible the detection of Ni in various aqueous standard reference samples with sodium concentrations ranging from micrograms to tens of milligrams per milliliter (riverine water and estuarine water) with good accuracy and precision.[187] ICCD detectors have also been used for studies of Ti (as was shown in Figure 8 above)[80,188] and Sb in environmental samples and human blood.[189]

4.4.3.3 Zeeman Background Correction Another technique for background correction is the Zeeman technique, extensively used for atomic absorption spectrometry (AAS) in an ETA. Use of the Zeeman background-correction technique together with LIF/LEAFS in an ETA (termed Zeeman electrothermal atomization/laser-excited atomic fluorescence spectrometry (ZETA/LEAFS)[171]) has been studied by the group of Michel. ZETA/LEAFS was first investigated with Co as a pilot element by Dougherty et al.[161] ZETA/LEAFS was found to correct for furnace blackbody radiation, scatter and stray light. However, the Zeeman effect has a drawback in that it degrades sensitivity. Dougherty et al. found an LOD of 0.3 pg with and 0.7 pg without Zeeman correction. It was later found that the Zeeman correction degraded the LOD for elements such as Ag, Co, In, Mn, Pb, and Tl by a factor between 1 and 10, whereas no successful Zeeman correction could be done for Cu.[171] The Zeeman correction technique did not affect the LOD but enabled correction for scatter in resonance fluorescence determination of Mn in a zinc chloride matrix and brain tissue. A diagnostic study of the Zeeman effect was made by Preli et al.[163] They investigated the influence of Zeeman splitting, applied field strength, laser excitation line width, and atomic spectral profile on the LOD and LDR for six elements (Ag, Co, Cu, In, Pb, Tl). All these investigations used a longitudinal Zeeman correction system. Transverse ZETA/LEAFS has been investigated by Irwin et al. on Pb and Co.[193]

4.4.3.4 Wavelength-modulating Laser-induced Fluorescence/Laser-excited Atomic Fluorescence Spectrometry Techniques Various WM methodologies have been investigated for the LIF/LEAFS techniques by Goff et al.[143] and by Su et al.[248] The light from a dye laser was modulated periodically either by rapid tuning of an etalon or by tilting a mirror in the cavity by electrical means. Both of these groups reported on the reduction of the laser scatter background in various types of measurements. A potential drawback of this type of wavelength-jumping method is that it cannot correctly compensate for a structured background.

4.4.4 Limits of Detection

LOD are normally defined as the concentration (or absolute amount) of analyte in a sample that produces a signal equivalent to three times the standard deviation of the noise associated with the blank. For steady-state atomizers (e.g. flames and plasmas) LOD are expressed in concentration units (e.g. ng mL^{-1}), whereas for nonsteady-state atomizers (e.g. ETAs and vapor cells) absolute amounts are used (e.g. pg). A convenient way to determine LOD for a given instrumentation is to divide the amount of noise (three times the standard deviation of the noise associated with the blank) by the sensitivity. This implies that LOD depend both on the sensitivity and the noise of the technique. Because both of these two entities in turn depend on most other parts of the instrumentation (type of laser, atomizer, detector, as well as sample constituents), LOD can vary widely between various elements and instrumentation. This also implies that one given instrumentation does not necessarily provide the optimum conditions for all elements. This can be exemplified by the works by Wei et al.[174] and Vera et al.[202,209]

Wei et al. determined the LOD for two elements (Tl and Pb) atomized in a GF for the three different optical systems examined for their overall efficiency above (i.e. right-angle detection combined with dispersive detection; front-surface detection combined with dispersive detection; and front-surface detection combined with nondispersive detection).[174] They found that although the sensitivity increased with front-surface detection (in comparison with right-angle detection) for Tl (for dispersive detection), the noise increased by almost the same amount. Hence, the front-surface and the right-angle approaches gave the same LOD for Tl (3 fg). For Pb, on the other hand, the sensitivity increased significantly

more than did the noise, and LOD were improved by almost one order of magnitude for front-surface detection (from 7 to 1 fg). They attributed this difference to the fact that the dominating source of noise for Pb was blackbody emission, whereas scattered laser light noise dominated for Tl. They therefore concluded that the front-surface mode of detection will offer a significant improvement of the LOD over right-angle mode detection for situations where blackbody emission dominates but not necessarily when noise from scattered laser light is dominating.

The same authors also found that nondispersive detection (e.g. narrow-bandpass filters) had an even better light-gathering power, and hence a better sensitivity, than dispersive detection (using a spectrometer).[174] They therefore claimed that nondispersive detection should reduce noise and lead to improved LOD. It remains to be proven, however, whether nondispersive systems can be as versatile as dispersive systems because they generally have a lower selectivity.

Vera et al.[209] evaluated three different laser systems (nitrogen-, copper-vapor- and Nd : YAG-pumped dye laser systems, respectively) for their Pb detection power by LIF/LEAFS in a graphite tube. They found that the best LOD of Pb could be achieved by the high-repetition-rate (6 kHz) Cu-vapor laser (0.5 fg, laser scatter limited), whereas a 20-Hz repetition-rate nitrogen laser and a Nd : YAG laser yielded virtually identical LOD (3 fg).

However, in an accompanying work by the same authors, concerned with the detection of Fe and Ga, the Cu-vapor laser did not yield the same advantage over the other systems.[202] The reason was attributed to the low pulse energy of the Cu-vapor laser (a few hundred nanojoules per pulse), which did not allow for optical saturation of the transitions in these elements. LOD for Ga, for example, were found to be almost two orders of magnitude lower when detected by the Nd : YAG laser than with the Cu-vapor laser (25 versus 2000 fg). Despite the appealingly high repetition rate, Cu-vapor lasers have not been used extensively for LIF/LEAFS in an ETA, with the powerful exception of the direct determination of Pb in Great Lakes waters by Cheam et al.[210]

The above examples show that it is *not* clear, a priori, that pulsed dye laser systems of nanosecond pulse length *always* provide the most optimum excitation source for the LIF/LEAFS. They can, under certain conditions, be rivaled by other types of laser systems. Theoretical considerations for (and discussions of) S/N ratios and LOD for the LIF/LEAFS technique have been given by a number of authors.[53–55,129,243,244] The actual LOD for the LIF/LEAFS technique that have been achieved in various experiments are presented below (with respect to the type of atomizer used).

Table 3 LOD for LIF/LEAFS in flames (ng mL^{-1})

Element	LOD	Refs.	Element	LOD	Refs.
Ag	4	145	Mo	12	145
Al	0.6	145	Na	0.1	144, 145
Au	4	152	Nb	1500	36, 136
Ba	0.7	151	Nd	2000	36, 144
Bi	3	145	Ni	0.5	150
Ca	0.08	145	Os	150 000	36, 136
Cd	8	145	Pb	0.02	157
Ce	500	36	Pd	1	152
Co	2	154	Pr	1000	36
Cr	1	145	Pt	0.7	152
Cu	1	145	Rh	100	36
Dy	300	36	Ru	2	152
Er	500	36	Sb	50	36
Eu	20	36	Sc	10	36, 136
Fe	0.2	149	Sm	100	36
Ga	0.9	145	Sn	3	150
Gd	800	36	Sr	0.3	145
Hf	100 000	136	Tb	500	36
Ho	100	36	Ti	2	145
In	0.2	145	Tl	0.8	157
Ir	9	152	Tm	100	36
Li	0.5	145	U	500 000	144
Lu	3000	36	V	30	145
Mg	0.2	145	Yb	10	36
Mn	0.4	145			

4.4.4.1 Laser-induced Fluorescence/Laser-excited Atomic Fluorescence Spectrometry in Flames Table 3 gives a compilation of the best LOD for the flame LIF/LEAFS technique.[36,136,144,145,149–152,154,157] As was mentioned above, although the LIF/LEAFS technique was first applied to flames as atomizers and a large number of studies have been done in the field,[35–38,60,64–67,70,71,78,81,86,115,118–120,136–158] the actual detection limits have turned out not to be as impressive as first anticipated.[35,36,125] The assessment of sub-nanogram per milliliter detection limits of Pb and Tl in an acetylene–air flame (0.02 and 0.8 ng mL^{-1}, respectively) by Human et al. in 1984 is still among the most impressive results by flame LIF/LEAFS.[157] Most other elements show detection limits in the ng mL^{-1} range (ranging from some parts of ng mL^{-1} to several μg mL^{-1}).

One important reason for the rather high LOD is that flames give rise to significant amounts of background radiation, which acts as a source of scattering for laser light. Many types of flames, supported on a variety of burners, have therefore been scrutinized for their applicability to LIF/LEAFS. Even if slot burners (which were specially designed for AAS by providing a long pathlength) were handy to use, they were found not to be suitable for the LIF/LEAFS technique.[35] A variety of alternative burners have therefore been developed for the LIF/LEAFS technique – some with small holes or capillary tubes, and some surrounded by a sheet of inert

gas (most often N_2 or Ar) in order to reduce the effects of quenching that can result from the entraining of air and reduce the fluorescence yield.[150]

Another way to improve the sensitivity and LOD of conventional flame LIF/LEAFS is to use a multipass configuration (using a flat mirror with a pierced hole in the perpendicular position together with a toroidal mirror).[149] Using a flashlamp-pumped dye laser for excitation, Epstein et al. demonstrated that the LOD of Fe could be improved from 0.6 to 0.2 ng mL^{-1}.

The field of flame LIF/LEAFS has been quite stagnant during latter years but an impressive exception is the multi-element analysis of river sediment for five elements using a modern tunable OPO system. Comparable or improved flame LEAFS detection limits over previous literature values were obtained for Co, Cu, Pb, Mn and Tl (2, 2, 0.4, 0.2 and 0.9 ng mL^{-1}, respectively).[154]

4.4.4.2 Laser-induced Fluorescence/Laser-excited Atomic Fluorescence Spectrometry in Graphite Furnaces If the actual detection limits turned out not to be as impressive as first anticipated when flames were used with the LIF/LEAFS technique, the results were soon found to be significantly more impressive when the ETA was used as atomizer. Detection limits well into the picogram per milliliter range and in the femtogram range were rapidly established. There are a number of developmental achievements related to furnaces, the optical system, detectors and detection strategies that have brought the LIF/LEAFS technique in an ETA to where it stands today.

Following the first LIF/LEAFS measurements in ETAs, which were made on graphite rods (see references 85, 157, 216, 218, 220–222), filaments[222] and boats,[217,219] open cups rapidly became popular (see references 132, 160, 165–170, 178, 198, 203–208, 213–215, 221) owing to their better performance. For example, Goforth and Winefordner compared a graphite rod, a plain graphite cup and a slotted graphite cup for a number of elements and found that the rod worked well for volatile elements but gave insufficient atomization for less-volatile elements, and that the plain graphite cup gave the best overall results of these atomizers. They also compared the LOD for pyrolytic coating, a tantalum foil liner and a tantalum carbide coating of a graphite cuvette, and found the pyrolytic coating to give the best results.[221]

It was meanwhile found by AAS, however, that open atomizers (e.g. rods, boats and cups) are impractical due to diffusion losses[249] and vapor-phase interferences[250] that occur in the cool zone above the atomizer. Similar effects were soon found for the LIF/LEAFS technique, for example by Bolshov et al.[203] It was therefore concluded, first by Human et al.[157] and Dittrich and Stärk,[218] and later by others (e.g. Falk and Tilch[214]), that graphite tubes provided the best atomization environment for analysis (see references 77, 79, 80, 121, 158–164, 171–177, 179–197, 199–202, 209–212, 214, 218, 220, 224–226). Dittrich and Stärk,[218] and shortly after, Falk and Tilch,[214] for example, demonstrated that the tube atomizer had an increased sensitivity over the rod atomizer. They attributed this to an increased residence time and a higher degree of atomization. Other advantages of the tube are a more homogeneous temperature distribution and a higher heating rate.[174,218]

The most important development of the optical system for LIF/LEAFS in an ETA is the front-surface detection (as was shown in Figure 7), which not only provides freedom from post-filter effects (see below) and the imaging of a larger interaction volume but it also has yet another distinct advantage over right-angle detection when it comes to ETAs – it can be applied directly to the atomizer without any modification of the furnace system. In addition, the use of reflective instead of refractive optics has also improved the performance of LIF/LEAFS in an ETA (because it provides chromatic-free light collection).

Regarding detectors, most LOD have been determined using PMTs. The ICCDs have not proven themselves to yield any better LOD than the PMTs under normal conditions. Their main advantage instead is that they can monitor the fluorescence background significantly more effectively than the PMTs and hence yield more reliable results when samples with complex matrices are detected.[189] The ICCDs have also proven to be indispensable when efficient excitation–detection transitions are to be found in atoms with complicated atomic structure.[80,187]

The state-of-the art performance of the LIF/LEAFS technique in an ETA is shown in Table 4, which displays the best LOD published so far for each element (see references 79, 158–160, 169, 170, 173, 174, 178, 179, 183, 187–190, 194–196, 199, 201, 203–206, 224, 225). Table 4 is based upon a compilation of around 200 LOD from the literature for the LIF/LEAFS technique in an ETA.

As can be seen from Table 4, quite a number of elements have LOD in the low femtogram range. For many elements, improvements of two to four orders of magnitude (with respect to the conventional AAS technique in an ETA) have been obtained. Elements that have poor LOD in the AAS technique because of furnace atomization problems (in general those that require high atomization temperatures, normally above 2500 °C) also have relatively poor LOD in the LIF/LEAFS technique. The LOD for these elements are limited either by memory effects in the graphite cuvette[80,173,188,194] or by blackbody radiation from the atomizer. Although the majority of the LOD have been measured using one-step excitation,

Table 4 LOD for the LIF/LEAFS technique in ETAs (fg)

Element	First-step excitation	Second-step excitation	Fluorescence wavelength	ETA[a]	Mode of detection[b]	Excitation source[c]	LOD	Refs.
Ag	328		338	T/P	90°	E	20[d]	159, 160
Al	308		394/6	T/W	180°	E	100	174
As	193.7		245.7	T/P	180°	Y	54	190
Au	267.6	406.5	200	T/W	180°	E	3	158
Ba	597.2		611.1			F	40 000 000	205
Bi	223.1		299.3	C	90°	E	2.5	170
Cd	228.8		228.8	C	90°	E	0.5	169
Co	308.3		345.4	T/W	180°	Y	4	201
Cr	427.5	529.8	360	T/W	180°	E	1400	173
Cu	324.7		510.5	C	90°	Y	150	203
Eu	287.9		536.1	C	90°	Y	300 000	203, 206
Fe	296.7		373.5	T/W	180°	E	70	174
Ga	403.3	641.4	250.0	T	180°	Y	1	79
Hg	253.7	435.8	546.1	C	180°	E	90	178
Ir	237.3		254.4	T/W	180°	E	18	179
Li	670.8		670.8	T/W	180°	D	10	225
Mn	279.5		279.5	T/P	90°	E	80	159
Mo	313.3		317.0	T/W	180°	N	100 000	224
Ni	224.5		231.4	T/W	180°	E	10	187
P	213.6		253.4/6	T/W	180°	E	8000	195, 199
Pb	283.3		405.8	T/W	180°	E	0.2	174
Pd	244.8		343.3	T/W	180°	E	4	179
Pt	264.7		270.2/6	T/W	180°	E	70	179
Rb	780.0		780.0	T/W	180°	D	20	225
Rh	236.2		381.5	T/W	180°	E	190	179
Ru	287.5		366.3	C/LP	90°	Y	3000[e]	204
Sb	287.8		372.2	T/W	180°	E	5	189
Se	196		204/6	T/P	180°	Y	15	190
Sn	286		318	T/W	180°	E	30	174
Te	214.3		238.3/6	T/W	180°	E	20	183
Ti	264.6		295	T/W	180°	E	1000	188
Tl	277		353	T/W	180°	E	0.1	196
V	458.0	578.6	256	T/W	180°	E	10 000[f]	194
Yb	398.8	666.8	246.4	T/W	180°	Y	220	79

[a] T = tube; P = platform; W = wall; C = cup.
[b] Direction of detection of the fluorescence.
[c] E = excimer-pumped dye laser system; N = nitrogen-pumped dye laser system; Y = Nd : YAG-pumped dye laser system; D = diode laser.
[d] Contamination-limited LOD reported: estimated LOD without contamination: 8 and 10 fg.
[e] A volume of 30 μL has been assumed.
[f] Contamination-limited LOD reported: estimated LOD without contamination: 2500 fg.

there are a few elements whose LOD have been obtained by using double-resonance fluorescence.

Molecular detection has found a certain (although a limited) use for the assessments of some elements whose excitation wavelengths are difficult to access with laser light, e.g. halogens.[121,177,192,195,223,251] In the case of fluorine, for example, a matrix of Mg was added to the sample (analyzed in a GF), leading to the formation of MgF, which could be detected by LEMOFS.[177,192]

4.4.4.3 Laser-induced Fluorescence/Laser-excited Atomic Fluorescence Spectrometry in Other Atomizers LIF/LEAFS has also been used in conjunction with other types of atomizers. Owing to the success of the ICP with conventional detection techniques, a number of investigations have been devoted to investigations of the use of this atomizer together with LIF/LEAFS.[37,60,66,94,122,157,228–230] An early, mostly diagnostic, investigation of the use of the LIF/LEAFS technique with an ICP was performed by Omenetto et al.[228] A more extensive demonstration of the analytical capability of the LIF/LEAFS technique applied to an ICP was made by Human et al. in 1984.[157] The authors assessed LOD for a range of elements (Al, B, Ba, Ga, Mo, Pb, Si, Sn, Ti, Tl, V, Y, Zr and U). The LOD were found to be between 0.4 and 20 ng mL^{-1}. Atomic fluorescence from Ag, Au, Hf, Ir, Mo, Nb, Pd, Pt, Ru, Ta and Zr in an ICP was measured by Huang et al.[229]

LOD found were in the 1.3–58 ng mL^{-1} range, with an LDR of over four orders of magnitude for most elements. Double-resonance *ionic* fluorescence has been used as an alternative to atomic fluorescence from Ca, Sr, Ba and Mg atoms in an ICP in order to improve on the sensitivity or to extend the number of accessible elements.[94] The LOD for these four elements were 0.007, 1, 1, and 0.05 ng mL^{-1}, respectively. Simeonsson et al.[230] used both single- and double-resonance excitation for detection of atomic or ionic fluorescence from Ag(I), Au(I), Co(I), Cu(I), Ni(I), Pb(I), Pd(I), Pt(I) and Sc(II). The LOD for single excitations for these elements were found to be 0.8, 4.2, 15, 3, 5, 0.7, 3.9, 3.3 and 0.2 ng mL^{-1}, respectively. The corresponding LOD for double excitation were 1.7, 6.0, 530, 5, 170, 8, 45, 29 and 13 ng mL^{-1}, respectively. Their results show that double-resonance excitation in fact yields higher LOD than single-resonance excitation. The authors stressed, however, that the spectral selectivity for the double-resonance excitation technique should widely supersede that of the single-resonance excitation technique.

These studies show that the LOD for the LIF/LEAFS technique with an ICP are generally in the low nanogram per milliliter range, with a few exceptions such as ionic detection of Ca and Mg, which are well into the picogram per milliliter range.[94]

Although the LIF/LEAFS technique is generally not very impressive for isotope detection, because most optical isotope shifts are smaller than the typical laser bandwidths or Doppler and collisional broadening mechanisms in most atomizers (with the exception for some actinides and lanthanides), it has been used for isotopic detection of uranium isotopes in complex matrices in an ICP.[122]

LIF/LEAFS has also been applied to a hollow-cathode glow discharge, which served the purpose of being the atomizer (aqueous solutions were dried in graphite electrodes used as disposable hollow cathodes).[231] LOD for Pb and Ir were 500 fg and 20 pg, respectively. A commercial hollow-cathode lamp (which has a very low background emission and excellent atomization characteristics) has also been investigated for its use for SAD (referring to the number of atoms in the interaction volume).[232] An LOD for Pb of 1.8 ag (i.e. 0.0018 fg) in the laser beam was established. Work with a hot, hollow-cathode atomizer has also been performed.[233] LOD for dried solutions of Co and Ni were 15 pg mL^{-1} and 5 ng mL^{-1}, respectively. LIF/LEAFS measurements for trace detection in conducting solids (Fe in brass) have been performed by Travis et al.[234] The LOD were below the microgram per gram level. LIF/LEAFS has also been applied to direct detection of U(VI) in solution, with a LOD of 40 pg mL^{-1}.[235]

4.4.5 Linear Dynamic Range

Although the LIF/LEAFS signal is linear with analyte concentration for small concentrations, pre-filter, self-absorption, and post-filter effects can degrade the linearity at higher concentrations.[35,37,70,137,164,241] Pre-filter effects refer to the reduction in fluorescence signal that occurs because of absorption of *laser light* by analyte atoms on its way to (i.e. prior to) the interaction volume (the interaction volume is defined as the overlapping volumes of laser irradiation, atomic density and detection), whereas self-absorption refers to reabsorption of *fluorescence light* by analyte atoms *within* the laser-irradiated volume. Finally, post-filter effects refer to reabsorption of *fluorescence light* by atoms *outside* the irradiated volume (i.e. absorption of fluorescence by atoms within the detection volume but outside the illuminated volume). The risk for self-absorption and post-filter effects to occur is naturally higher for transitions that terminate in highly populated states (e.g. the ground state or in low-lying excited states) than for those transitions that terminate in short-lived excited states.[38,164] Self-absorption can occur also for transitions terminating in relatively long-lived excited states mainly when continuous-wave excitation is used.

Because both pre-filter and self-absorption originate from absorption of light by atoms within the irradiated volume of atoms, these effects can be reduced by optically saturating the transition because optical saturation occurs when the rates of absorption and stimulated emission balance each other, a medium with optically saturated atoms is transparent at that particular wavelength.[145] Post-filter effects, on the other hand, are not affected by the laser irradiation. Hence, they can cause significant reduction in the LDR even when the atomic transition is saturated. On the other hand, post-filter effects can be eliminated by using front-surface detection (because all atoms viewed by the detector are illuminated by laser light).[35,37,38,156,164,174,176,195,221,224,248] This implies that the front-surface mode of detection yields smaller post-filter effects and hence a larger LDR than do systems with right-angle detection.[156,164,174,176,195,221,224,248] Dynamic ranges of five to seven orders of magnitude have been demonstrated for several elements using front-surface detection.[164,175,212,221]

4.4.6 Selectivity

The selectivity of the LIF technique is excellent due to a double selectivity for the conventional single-resonance fluorescence spectrometry technique and a triple selectivity for the double-resonance technique. As was discussed in connection with concomitant scatter and nonanalyte atomic fluorescence above, a concomitant species needs to have a spectral overlap both in the excitation as well as in the de-excitation processes in order

to yield a background signal. There are very few such examples documented in the literature. One exception was the study of Sb in estuarine water, as shown above in Figure 10.

4.4.7 Applications

Although the limited space in this theory and instrumentation review precludes any detailed analysis of the applicability of LIF/LEAFS, a few examples can be given.

4.4.7.1 Laser-induced Fluorescence/Laser-excited Atomic Fluorescence Spectrometry in Flames The flame LIF/LEAFS technique has been used for a limited number of applications. Compilations can be found in the literature.[36,68] One example is Fe detected in simulated fresh water, unalloyed copper and fly ash. Other examples are the detection of Pb in blood[155] and the determination of Ni and Sn in various reference materials (riverine water, unalloyed copper and fly ash).[150] A multi-element LIF/LEAFS analysis of river sediment in flames for Co, Cu, Pb, Mn and Tl using a modern tunable OPO system, encompassing 640 measurements in 6 h with triplicate measurements of all solutions and aqueous calibration curves, with a relative standard deviation (RSD) precision better than 5%, was done by Zhou et al.[154] Flame LIF/LEAFS has also been coupled to high-performance liquid chromatography (HPLC) for speciation. Walton et al., for example, detected various organomanganese and organotin species by HPLC and flame LIF/LEAFS. The detection limits for organomanganese species ranged from 8 to 22 pg of Mn.[156]

4.4.7.2 Laser-induced Fluorescence/Laser-excited Atomic Fluorescence Spectrometry in Eletrothermal Atomizers There are a large number of applications of the LIF/LEAFS technique in ETAs. A recent compilation is given in the review by Hou et al.[38] An early example is the detection of Cd and Pb in Antarctic and Greenland snow and ice by Bolshov et al.[169] (with an LOD of 0.5 fg). Thallium, manganese and lead were detected by slurry sampling in food and agriculture standard reference materials by Butcher et al.[175] The technique has also been used for the assessment of trace levels of gold in size-segregated, atmospheric particulate samples.[158] The simultaneous monitoring of entire fluorescence spectra by the use of an ICCD made possible the detection of Ni in various aqueous standard reference samples with sodium concentrations ranging from micrograms to tens of milligrams per milliliter (riverine water and estuarine water) with good accuracy and precision.[187] The Zeeman correction technique has been used for assessment of Mn in a zinc chloride matrix and brain tissue using resonance fluorescence.[171]

4.5 Diagnostic Capabilities

The LIF/LEAFS technique has been used by numerous authors for a variety of diagnostic studies.[60,252] Example of entities that have been studied are temperatures,[60,253–256] number densities,[35,87,257–259] quantum efficiencies,[260] lifetimes of atomic states in flames,[69,261] spatially resolved atomic distributions in flames,[67] laser-induced flame chemistry processes,[262,263] collisional effects in flames,[264] collisional mixing rates of excited atomic states in flames,[78,86] the role of chemical reactions on saturation curves,[265] excitation–detection wavelength combinations of atoms with complex atomic structure by the use of an ICCD,[80] memory effects in GFs,[188] and the Zeeman effect.[163,193] Of special importance is the fluorescence dip,[97–99,266] which has been used for studies of quantum efficiencies,[267] radiative transfer,[268] ionization depletion[99] and autoionizing cross-sections.[269] A thorough study of the capabilities of performing time-resolved investigations for diagnostic purposes was made by Omenetto and Matveev.[270] Further discussions about the diagnostic capability of the LIF/LEAFS technique are given by Omenetto.[127,266]

4.6 Conclusions and Outlook

LIF/LEAFS is a technique that has clearly demonstrated its usefulness for analytical applications. The most impressive performance has been achieved with the GF as an atomizer. The most versatile detection systems for LIF/LEAFS today are based upon a combination of front-surface detection and chromatic-free reflective optics (either by a pierced curved mirror or a pierced flat mirror followed by a curved mirror). Double-resonance fluorescence can bring high detection powers to some elements. The use of an ICCD when samples with complex matrices are analyzed is advantageous. LOD in the low femtogram range have been obtained for a number of elements. Owing to its double (or triple) selectivity, the LIF/LEAFS technique shows an extraordinary selectivity. Moreover, the LDR has been found to be large (five to seven orders of magnitude has been demonstrated). This implies that the LIF/LEAFS technique in an ETA can rival most other analytical techniques for trace element analysis.

Work has also been done with regard to the SAD capability of the LIF/LEAFS technique. It is clear that the LIF/LEAFS technique is capable of detecting the presence of one single atom in quiet surroundings (e.g. in a vacuum). However, the SAD capability of the technique seems more difficult to address when the technique is coupled to practically useful atomizers. It is clear that the graphite tube (and in particular the transversely heated tube) so far provides the best (although not the ideal) atomizer for practical analytical work. Work with other,

quieter atomizers, e.g. a hollow-cathode lamp, might bring down the LOD of the technique, although at the price of less versatile and less user-friendly instrumentation.

Looking back at the development of the LIF/LEAFS technique in an ETA – with the improved knowledge about the underlying principles for the technique that has been gained throughout the years, which in turn has resulted in more efficient experimental set-ups and detection methodologies – and speculating into the future, one can conclude that it would be surprising if the technique were not to be quoted with even better LOD, selectivity and LDR values a few years from now.

5 LASER IONIZATION SPECTROMETRY

5.1 Laser-enhanced Ionization versus Resonance Ionization Spectrometry

As was alluded to in the Introduction (section 1.4.2), there are basically two different laser-based spectroscopic ionization techniques with good potential for analytical applications: LEI and RIS. The main difference between the two is that LEI is performed in conventional atmospheric pressure atomizers (flames, furnaces, etc.) and RIS is performed under vacuum conditions.

Because atoms in atmospheric pressure atomizers are exposed to a high collision rate, they also experience a certain thermal ionization. As was discussed earlier, this thermal ionization ionizes excited atoms at a rate faster than ground-state atoms. Hence, this thermal ionization can be increased significantly by exciting atoms in high-collision media by laser light. By applying an electrical field across the interaction volume, the increased ionization rate is then detected as an increase in the current through the atomizer, which constitutes the signal in LEI spectrometry.[40] In the environment to which RIS is applied, there is no such thermal ionization. This implies that although the laser light has only to excite selectively the analytes in the LEI techniques (the ionization is provided by the existing thermal ionization process in the atomic reservoir), the laser light has to be used for both selective excitation and ionization of the analytes in RIS. Although the RIS technique has shown some extraordinary properties (i.e. SAD and extremely high elemental selectivity, especially when coupled to a mass spectrometer for detection of the ions produced), the fact that atomization of the sample in general is performed under vacuum conditions puts some restrictions on the practical applicability of the technique. It will therefore be covered in less detail in section 5.8. The reader is referred to the literature for a detailed description of the technique.[10,11,19,41,271–285]

5.2 Laser-enhanced Ionization Spectrometry – An Introduction

LEI is a technique that originates from a finding by a group of scientists at the National Institute of Standards and Technology (NIST, formerly National Bureau of Standards NBS), Gaithersburg, Maryland, USA, in 1976.[286] Green et al. found that atoms in a flame irradiated with laser light experience an enhanced ionization rate if the laser light is resonant with a transition in the atoms. This increased ionization could be detected most easily by applying an electrical field across the flame and measuring the corresponding increase in current. This effect was originally termed ''optogalvanic detection'' for a few years by the American community,[286–291] a name that rapidly became used exclusively for the phenomenon in discharges.[288] The name LEI was used for the ''optogalvanic'' phenomenon in flames for the first time in 1979.[91,292,293] The technique has also appeared under a variety of other names throughout the years, e.g. (stepwise) photoionization in flames (primarily by the Russian community),[294] dual laser ionization,[295–297] laser-assisted ionization,[298] laser-induced ionization,[299,300] laser step-photoionization,[301] or just stepwise photoionization.[294]

As was alluded to previously, it was found early on that laser light in fact does not need to ionize the atoms under study in LEI spectrometry; it is sufficient to *excite* them (e.g. promote ground-state atoms to an excited state) in order to create an increased rate of ionization. This led to the conclusion that the laser light enhances an already existing thermal ionization process, sustained by the heat of the flame, which ionizes excited atoms at a higher rate than it does for unexcited (ground-state) atoms. A more concise terminology of the technique would therefore have been laser-enhanced collisional ionization,[302] an expression that never came into use.

The common denominator for all atomic reservoirs (atomizers) to which LEI can be applied is, therefore, that they must possess a certain omnipresent thermal ionization that converts photons absorbed by the atoms under study into freely movable charges (electrons and ions). This implies, in turn, that the atomizer performs two functions simultaneously in LEI spectrometry. It renders an atomization of the sample and provides a collisional excitation of the laser-excited atoms.

The LEI technique has most often been used with flames as atomic reservoirs, although GFs, plasmas, discharges, cells and thermionic diodes also have been used but to a significantly lesser extent than flames. We will therefore direct most of this presentation of LEI to its use in flames and only briefly mention its use together with other atomizers (e.g. GFs and plasmas). Moreover,

in a perspective of analytical chemistry, LEI spectrometry with optogalvanic detection in discharges or thermionic diodes is of less interest because these types of atomizers do not provide for any simple means of introducing and atomizing an unknown sample into the atomic reservoir. Optogalvanic detection in discharges has become of more interest for spectroscopic investigations and as a technique for wavelength calibration, and therefore will not be discussed further here. The only exception is the development by Petrucci and Winefordner of a sensitive photon detector based upon the optogalvanic effect.[303] The reader is therefore directed to the literature for more information about optogalvanic spectrometry.[304,305]

5.3 Previous Reviews

The field of LEI spectrometry has been reviewed previously by a number of authors.[39,40,237,289,304,306–320] An early but very illustrative review is that by Travis et al.[39] The work by Green and Seltzer is a thorough and easy-reading review of the status of the field up to 1989.[314] A contemporary review, although somewhat more focused upon the principles of the technique, is that of Axner and Rubinsztein.[309] Butcher has recently provided an updated complement to the review of Green and Seltzer.[315] The most comprehensive review, however, is given by a recent book dedicated to the field.[40] This book covers the fundamental mechanisms of LEI in terms of the production of ions[57] and the signal detection;[316] the analytical performance of LEI in flames;[237] applications of LEI;[317] nonflame reservoirs for LEI spectrometry;[318] and the interplay of LIF and LEI in different atomic reservoirs.[319] Although mostly used for the detection of atoms, the LEI technique has also been used for the detection of molecules, which has been reviewed specifically by Webster and Rettner.[304]

5.4 Instrumentation

5.4.1 Typical Set-up

Because LEI mostly has been used with flames as atomizers, the flame LEI set-up shown in Figure 11 will serve the purpose of illustrating a ''typical'' LEI instrumentation. In short, light from a laser system, most often a pulsed dye laser system, is directed into a flame at a height where there is a sufficient atomic population of the analyte species to be detected. An electrical field is established in the interaction volume by applying a potential (most often ranging from −1 to −2 kV) between

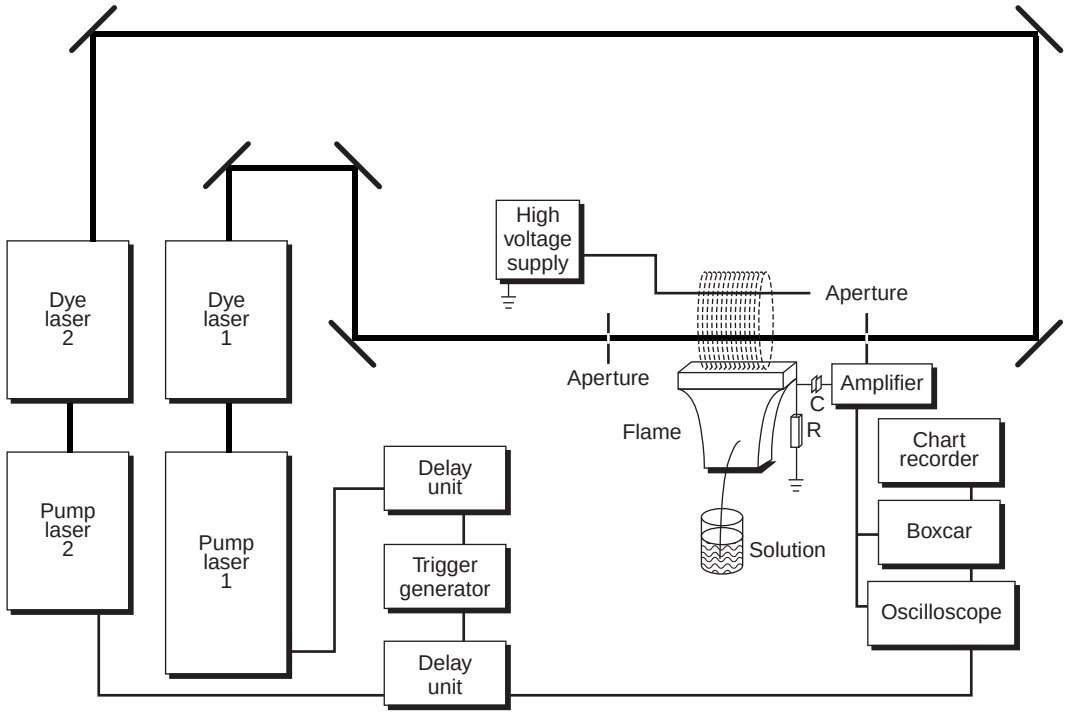

Figure 11 A set-up for LEI spectrometry in flames. The solution to be analyzed is aspirated into a flame. The atomic vapor is illuminated by light from either one or two tunable lasers. In conventional one-step LEI only one laser system is needed. For two-step excitation, the pump laser from the pulse laser is normally split and directed into two dye lasers simultaneously (not shown in the figure). By using two independent laser systems, as shown here, pump-and-probe experiments (e.g. for measurements of the lifetime of excited states) can be performed. (Reproduced by permission of Springer Verlag from O. Axner, P. Ljungberg, Y. Malmsten, *Applied Physics B*, **54**, 144–155 (1992).)

an electrode immersed in the flame (or alternatively between two external electrodes placed on each side of the flame) and the burner head (which is the head at ground potential). In order to be able to detect the charges created, the burner head has to be isolated electrically with respect to ground, with only a controlled electric path (accomplished by a resistor) to ground potential. When pulsed excitation sources are used, as is most often the case, a pulsed LEI current will result in this path. This current pulse is extracted from the direct current (DC) (originating from the thermal ionization) by a capacitor before it is converted to a voltage pulse by a current-to-voltage amplifier of suitable trans-impedance gain. The output signal is most often monitored by using a boxcar integrator whose detection window is set equal to or slightly longer than the resulting voltage pulse (often around 1 μs). The output from the boxcar is most conveniently fed into a laboratory computer.

5.4.2 Atomizers

5.4.2.1 Flames Most flames used for LEI have been supported on commercially available slot-burner units (including nebulizer units) developed for the AAS technique. The elongated form of the slot-burner-supported flame (normally 5 or 10 cm) is suitable for LEI because it gives a long interaction region (which in turn makes interaction with a large number of analytical species per laser pulse possible). In addition, the use of commercially available slot-burner units has also meant that the LEI technique has been able to benefit from the development of burner and nebulizer units. A typical pneumatic nebulizer, which converts the sample solution into a liquid aerosol, has a liquid sample uptake rate of around 5 mL min^{-1} and an efficiency of about 10% (the rest being drained to waste).

All standard types of flame used for AAS have been used for LEI (e.g. air–acetylene, nitrous oxide–acetylene, air–hydrogen), with the air–acetylene flame being the most popular. The reason for this is that this flame constitutes a good compromise between a high degree of atomization, relatively low occurrence of chemical interferences and a high degree of thermal ionization.

The higher-temperature flames (e.g. the nitrous oxide–acetylene flame) have a better degree of atomization of refractory elements and a higher degree of thermal ionization (and thereby collisional ionization), which would justify their use in LEI spectrometry. A disadvantage is that the higher degree of thermal ionization also produces a larger background signal (from which the LEI signal has to be extracted). This implies that the detection power of atoms that atomize readily in air–acetylene or air–hydrogen flames decreases when being detected in a nitrous oxide–acetylene flame.[237] For nonvolatile elements, however, the hotter nitrous oxide–acetylene flame offers an advantage.[321]

The cooler flames (e.g. hydrogen flames), on the other hand, benefit from a lower background ionization, which in principle should give opportunities to lower detection limits for those elements that atomize readily in such flames. The lower background ionization rate of the hydrogen flames, however, is not such an advantage in reality as might be expected. The reason is that when samples containing ''normal'' concentrations of EIEs are detected, the background ionization rate is dominated by thermal ionization from the EIEs rather than the flame background ionization. The cooler flames are instead limited by a lower degree of atomization and a lower collisional ionization rate, resulting in a poorer LEI detection power for most elements. In addition, the cooler hydrogen flames are more susceptible to chemical interferences, which limits their use for practical purposes. The cooler flames, however, are useful for LEI detection of alkali metals, because these ionize readily also in the cooler flames (and in fact exist with a larger fraction of neutral atoms).[91,237,322]

5.4.2.2 Plasmas Only a limited number of studies concerned with the use of plasmas as atomizers for LEI have been carried out.[300,323–327] The reason is that the plasma does not offer the same suitable environment for the LEI technique as does the flame. The main reasons for this are the following: firstly, the ion fractions in ICPs are much larger than in flames, which implies that the neutral atom population (i.e. those that are available for excitation) is smaller than in a flame for the atoms that already atomize readily in a flame; secondly, the electrical environment around a plasma is harsher than around a flame, implying that the radiofrequency pick-ups become more severe; thirdly, the electron density in a plasma is several orders of magnitude larger than in the flame, which makes it more difficult to detect the laser-induced creation of charges. LEI detection in an ICP has therefore been performed either in the tail end of the plasma[323] or in a power-modulated plasma.[324] Attempts have also been pursued to detect the laser-produced ions with a mass spectrometer rather than electrically.[325,327]

5.4.2.3 Electrothermal Atomizers The increased LOD that followed when the flame was substituted with the GF for the AAS and the LIF/LEAFS techniques also encouraged the field of LEI spectroscopists to move to the GF. Early attempts using the GF as an atomizer for LEI were therefore made by Gonchakov et al.,[301] Bykov et al.,[328] Magnusson et al.[329–331] and Sjöström.[332]

One drawback with LEI in a GF is that the GF does not allow for any apparently efficient geometry for the

charge collection as does the flame in flame LEI (see section 5.4.4.2). Therefore, although the discussion about electrode design for flame LEI has been conveniently placed in section 5.4.4.3, together with a discussion about the principles for the charge collection process, this is not possible in the case of LEI in a GF because the performance of this technique is strongly dependent on the interplay between the GF and the charge-collection electrode(s).

Gonchakov et al., who used a graphite cup as an atomizer,[301] used a circular wire probe placed above the cup. Magnusson et al., on the other hand, who used a large graphite tube (HGA-72), placed the charge-collecting probe inside the tube parallel with the laser light.[329,330]

Both Magnusson et al.[329,330] and Gonchakov et al.[301] investigated the influence of various voltages and potentials applied to the probe. Although the first work by Magnusson et al. concluded that the best reproducibilities and S/N ratios were obtained with a negative potential applied to the electrode,[329] they found contrary results in their second study.[330] With a negative voltage, the signal duration was determined by the time constant of the resistance–capacitance (RC) high-frequency cut-off filter (1μs or 100 ns, depending on the resistor value). The total signal (integrated over time), however, was found to be independent of the RC cut-off frequency. With a positive voltage, the signal lasted around 20 μs, and was found to be significantly larger than that obtained with a negative probe potential. It is as yet unclear as to whether this can be explained by the conventional model for charge collection and signal apportionment (based upon the difference in mobility between electrons and ions) developed by Travis and Turk (section 5.4.4.2).[40] Magnusson et al. suggested that signal-enhancing space–charge effects, similar to what exists in thermionic diodes, might play a role. Also Gonchakov et al. found a signal-enhancing effect. They state that: ''The signal amplification coefficient arising from the avalanche ionization of the probe was about 100. This was verified by replacing the tungsten probe with a stainless steel probe, for which the signal proved to be 100 times smaller. A polarity change at the probe reduced the signal by four to five orders.''[301]

In addition, it was concluded early on that the long-term reproducibility was not satisfactory for LEI in a GF. Magnusson concluded, for example, that the reproducibility could be improved by replacing the wire electrode with one of a solid graphite rod.[331] In this work, the laser beam was enlarged so that it filled the entire tube, and the electrode was placed in the center and held at zero potential. They reported on a large variation in the duration of the signal pulses when a positive potential was applied. These and similar results indicate that the signal collection process is not yet fully understood in LEI spectrometry in a GF.

It was found early on that the large currents from the heating system induced currents to the probe. In order to minimize these effects, Magnusson et al. synchronized the triggering of the laser with the 50-Hz power line. In this way, the fluctuations of the induced disturbances on the signal could be virtually eliminated.[329]

Both thermionic and photoelectric emission of electrons from surfaces (e.g. the electrode) had been noticed previously from flame LEI measurements when an uncooled metal surface was inserted into the flame or when intense UV light struck a surface, respectively. The same effects were found also for the LEI technique in a GF. Because a large electronic emission rate can lead to suppression of the analytical signal or electrical breakdown, it is not impossible that this might be one reason for the poor reproducibility of the LEI in a GF. Irrespective of whether this is the case, it will limit the applicability of the technique considerably.

Magnusson et al. showed that the effects of photoelectric emission could be reduced by carefully aligning the laser beam sufficiently far from the electrode.[330] It was noticed, however, that the heating of the furnace made the atmosphere act as a thermal lens, thus focusing the light during heating and defocusing it during cool-off, which can be a source of signal fluctuation.[330] Further investigations reported on the existence of dual peaks for LEI in a GF.[330] The same work also investigated, among other things, the influence of different protective gases on the thermionic emission of electrons and found significantly lower thermionic emission in N_2 than in Ar.[330]

An alternative approach to the careful aligning of the laser beam away from the electrode, first suggested by Magnusson[331] and then investigated practically by Sjöström,[332] was to separate spatially the atomization volume from the charge detection volume so as to overcome some of the problems occurring in the previous construction. A T-shaped furnace was constructed in which the atomization took place in the ''base'' of the T while the detection was carried out in the ''bar'' part of the furnace. This measurement section consisted of two stainless-steel plates, separated by Teflon insulation, with the outer plate held at −500 V. By separating the atomization and detection regions, less thermionic emission of electrons resulted.

Chekalin and Vlasov achieved a separation of the atomization and detection regions by placing the charge-collecting electrode above the dosing hole of a graphite tube while the laser beams were directed outside the tube (between the electrode and tube).[333]

The common denominator of all these LEI measurements in a GF is that none of them made use of the most modern and up-to-date furnace technology. Butcher et al. used a modern HGA-500 (Perkin-Elmer) furnace

and constructed a system with a moveable probe that was introduced automatically into the furnace shortly after the onset of the atomization so as to minimize the amount of thermionic emission.[334] They investigated the LEI and GF instrumentation with respect to a number of parameters in order to find the optimum conditions. Optimization of the atomization temperature for iron showed that no signal was obtained at temperatures exceeding 2000 °C. The loss of signal was attributed to the presence of a large thermionic emission of electrons. As a consequence, only temperatures up to 1900 °C could be used. This implied that relatively involatile elements such as Li, Mg, Fe and Mn required a long atomization period (20 s was used but this still did not ensure complete atomization).

5.4.2.4 Rod–Flame and Graphite Furnace–Flame Systems Attempts to combine the properties of the GF with the flame for LEI measurements have been pursued by Chekalin et al.[333,335–338] as well as by Winefordner, Smith et al.,[339,340] but with different approaches.

Chekalin et al. development a rod–flame system (see Figure 12).[335,336] The sample is placed onto a graphite rod placed in close proximity to a flame. By electrically heating the rod, the sample can be dried and charred before the rod is swung into the flame where the dried sample is evaporated. Because evaporation normally takes place at a lower temperature than atomization, atomization is delayed until the sample is well into the flame. Excitation and ionization then take place at a suitable height above the rod (determined by the height of the laser beams). By this procedure, the regions of evaporation, atomization and excitation–ionization–detection are separated from each other. This reduces, among other things, problems associated with thermionic emission

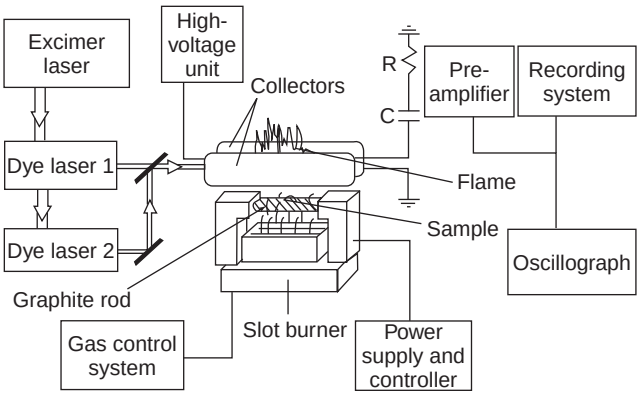

Figure 12 An early version of a set-up for flame–rod LEI spectrometry. In a more developed system, the rod can be swung in and out of the flame to gain better control of the vaporization process. (Reproduced by permission of the Royal Society of Chemistry from N.V. Chekalin, I.I. Vlasov, *Journal of Analytical Atomic Spectrometry*, **7**, 225–228 (1992).)

from the graphite rod that otherwise limit the detection power and matrix effects.

Smith et al. pursued an alternative route in order to separate the vaporization from the ionization detection.[339] The output of a GF was coupled into the argon flow of a miniature flame system.

5.4.2.5 Cells Detection of mercury differs from most other elements in that it already has a high vapor pressure at room temperature. No heated atomizer is therefore needed for mercury – a room-temperature atomic reservoir suffices. Mercury can therefore be measured directly in a cell. The groups of Winefordner and Omenetto have made extensive studies of the applicability of detection of mercury by the LEI technique (section 5.5.3.6).[178,270,341–346]

5.4.2.6 Thermionic Diodes Yet another alternative to the detection of laser-excited atoms by electrical means is the thermionic diode. Thermionic diodes are known to be extremely sensitive detectors for ions,[347] having a large dynamic range (over four orders of magnitude) and a very high gain (10^6 or larger).[318] The thermionic diode consists basically of a cylindrical anode in which an axially mounted cathode filament, heated by a DC, is mounted. No bias is needed to run the thermionic diode in a space-charge-limited mode. The ions produced by laser irradiation will be trapped within the negative space-charge cloud. Their presence reduces the space charge, which can be detected as a change in the diode current.[348–352]

Owing to its construction, the thermionic diode is most suitable for continuous-wave laser excitations. Although the first experiments were made with continuous-wave dye lasers, diode lasers have been used more frequently lately.[353–359] An improved type of thermionic diode that is suitable for analytical trace element analysis in samples with low vapor pressure has been constructed by Franzke et al.[354] By inserting a grid between the atomizer and the thermionic diode detector, and with appropriate biasing of grid and atomizer, the perturbation of the space charge due to the emission of thermal ions and electrons from the surface of the atomizer could be eliminated. The thermionic diode is also suitable for Doppler-free spectroscopy, which allows for precise isotope determinations.[355–357] It has also been found suitable for diagnostic purposes, i.e. the determination of collisional cross-sections and energy transfer rates.[353–363]

5.4.3 Laser Sources

As the selectivity of the LEI technique lies solely in the excitation step(s), narrow-band tunable laser systems must be used. Although the first demonstration of the LEI technique utilized a continuous-wave dye

laser system (based upon an argon ion laser),[277] most subsequent work has been done using pulsed dye laser systems pumped by excimer lasers (see references 72, 74, 82–84, 89, 90, 95, 99, 109, 173, 181, 327, 329, 336, 364–381), Nd:YAG lasers (see references 294, 322, 328, 365, 382–386), nitrogen lasers (see references 91, 322, 328, 387, 388), flashlamps (see references 74, 293, 306, 389, 390) or copper vapor lasers[238] (with the two first types being by far the most used). A smaller number of investigations have used continuous-wave lasers (tunable diode lasers,[391] argon- or krypton-ion-laser-pumped dye lasers,[392] or atomic line lasers[368]). As was alluded to earlier, the main reasons for the popularity of pulsed lasers is that they produce UV light (most atoms require UV light for efficient excitation) and saturate the transitions more easily than do continuous-wave systems, and they can produce a large signal over a short period of time (thereby maximizing the S/N ratio). A disadvantage with pulsed laser systems is the poor duty cycle.

5.4.4 Detection Equipment

5.4.4.1 Principles The most distinctive aspect of LEI is the charge-collection process. In LEI, the charge created as a consequence of laser illumination (i.e. the enhanced ionization rate) is detected electrically by applying an electrical potential across the volume of interaction. Hence, the atomic reservoir (e.g. the flame) is part of an electrical circuit. The applied potential gives rise to an electrical field that separates the charges created. The movement of the charges in the electrical field, in turn, gives rise to an induced current in the rest of the electrical circuit. This current is converted to a voltage by a current-to-voltage transducer (often termed amplifier), which thus constitutes the output signal of the LEI instrumentation. This implies that LEI differs from most other laser-based spectrometric techniques (e.g. LIF/LEAFS) because it does not require any light-detecting equipment.

5.4.4.2 Charge-collecting Electrical Field Although the electrical field strength between a cathode and an anode is constant under vacuum conditions, the thermal production of charges in a flame will give rise to a nonuniform electrical field strength. The cathode will attract the thermally created positively charged ions and the anode will attract the electrons. However, because the mobility of the electrons is about two to three orders of magnitude larger than that of the ions, they will be extracted more rapidly from the flame than the ions, leaving a certain density of positively charged ions behind (distributed with their highest density close to the cathode, decreasing in the direction of the anode). Hence, the flame will exhibit a certain positive charge. The presence of a net charge will give rise to a gradient in the electrical field. Because the highest density of ions is found close to the cathode, the largest field gradients will also be found there. This implies that the electrical field strength will decrease monotonically in a direction towards the anode. Depending on the potential applied, the distance between the burner head and the immersed cathode, and the amount of thermally produced charges, there will either be a finite but monotonically decreasing electrical field strength in all positions between the cathode and the burner head, or an electrical field strength that will be reduced to zero at a position before the anode. In the latter case, which can occur when there is a high thermal production of charges in the flame, the flame can be considered to be composed of two distinct regions: a positive ion space-charge region, often referred to as the cathode sheath or simply the sheath, which extends from the cathode to some intermediate position in the flame at which the electrical field strength becomes zero; and a part extending from the anode (the burner head) to the position in the flame in which there is no electrical field (hence constituting an unperturbed region in the flame). In order to be able to detect the increased ionization rate, the charges have to be created in a region where there is an electrical field that can separate the charges. This implies that only charges created in the sheath will contribute to the detectable signal.[316]

In addition, and as discussed in the literature in more detail, the LEI signal consists of two parts: one related to the movement of the electrons and one to that of the ions.[307,316] Because the mobility of the electrons is two or three orders of magnitude larger that of the ions, the electronic signal will last only a fraction of the ionic signal. Moreover, because the area of the signal versus time curve represents the relative potential passed by each type of particle, they are of roughly equal size, implying that the electronic signal is not only shorter but also about two or three orders of magnitude higher than the ionic signal. Detection of only the electronic part of the LEI signal (whose duration is about 1 μs) has been shown to yield better S/N than detection of the entire signal (with a microsecond duration) owing to the significantly smaller amount of noise detected with the smaller gate.

5.4.4.3 Electrode Design A variety of electrode designs for flame LEI have been used and evaluated throughout the years. The first set-ups used electrodes (plates or rods) placed in close proximity to, but outside, the flame. The electrodes had either different polarity (giving rise to a mainly horizontally directed electrical field across the flame) or the same polarity (acting as a common cathode with respect to the burner head so as to produce mainly a vertically directed electrical field in the flame). It was soon found, however, that the best

practical performance (for best immunity to interferences from so-called EIEs; see section 5.5.4.1 for details) could be obtained using a cathode immersed in the flame and directing the laser beam close to the electrode. It has been demonstrated repeatedly that the immersed cathode clearly outperforms the external cathodes as soon as samples with any substantial amounts of EIEs are analyzed.[316]

5.5 Analytical Performance

5.5.1 Sensitivity

5.5.1.1 Optimum Laser-enhanced Ionization Transitions A great deal of work in the field of LEI has been performed with the aim of finding the best/most suitable excitation wavelengths for various elements. It is obvious that whenever there are a number of transitions that all can be optically saturated by the laser light, the highest sensitivity will be obtained for the one that excites the atoms to the highest possible energy level (i.e. as close as possible to the ionization limit). However, because the transition probability (and hence the laser-induced excitation rate) rapidly decreases as a function of principal quantum number of the upper state, transitions to high-lying states will in general not be saturated by light from normal pulsed dye laser systems (which is illustrated in Figure 13). The question of which transition is the most sensitive in a given element then arises.

A systematic study of this concept was made by Axner.[89] Based upon a simple model (assuming that the transition probability from the ground state to excited states decreases with the principal quantum number cubed, as it does for transitions to Rydberg states, assuming no optical saturation, and a particularly simple form of the ionization efficiency, namely a pure Boltzmann factor), he predicted that the transitions to states positioned approximately $1.7\,kT$ below the ionization limit should show the highest LEI sensitivity for one-step excitations. In an acetylene–air flame with a temperature of 2500 K, this corresponds roughly to $3000\,\mathrm{cm}^{-1}$, which for the case of Na and Li (the pilot elements in many investigations) corresponds to the 7p and 6p states, respectively. Measured LEI sensitivities confirm these simple estimates fairly well. The highest LEI sensitivity (for unsaturated transitions) for Na has been found for the 3s–6p transition (i.e. to a state approximately $4150\,\mathrm{cm}^{-1}$ below the ionization limit), whereas the 2s–5p and 2s–6p transitions for Li have been shown to have approximately the same sensitivity. The small discrepancy between prediction and findings was explained by the use of an oversimplified expression for the effective ionization rate for excited atoms.[89] The same work also predicted the most sensitive unsaturated one-step transitions for a number of other elements (K, Rb, Cs, Mg, Ca, Sr, Ba, Al, Ga, In and Tl) using real transition probabilities rather than asymptotically extrapolated values. Because many elements have approximately the same level structure among the excited states, one can conclude that the most sensitive one-step transitions for LEI (for unsaturated transitions) are those that excite the atoms to states approximately $2–3\,kT$ below the ionization limit.

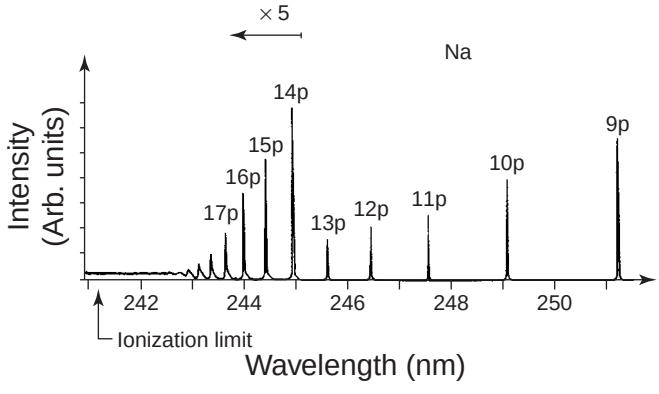

Figure 13 An LEI scan over a number of transitions (3s to higher-lying *n*p states, $n > 9$) in Na in a flame. (Reproduced by permission of the Society for Applied Spectroscopy from O. Axner, T. Berglind, *Applied Spectroscopy*, **43**(6), 940–951 (1989).)

5.5.1.2 Ionization Yield The ionization yield, defined as the fraction of illuminated atoms that are ionized as a consequence of laser irradiation, Y^{ion}, is a convenient quantitative measure of the sensitivity of the LEI excitation–ionization process. It has been found that the ionization yield can approach unity for pulsed as well as continuous-wave excitations for elements with low ionization potential.

For example, Smith et al. determined the ionization yield of Li atoms excited to the 2p and 4d states by pulsed lasers in an acetylene–air flame.[96] The authors concluded that approximately 60% of all the atoms were ionized following a stepwise excitation to the 4d state. Axner et al. measured ionization yields of up to 80% for Na atoms excited to the 7d state in a two-step process (3s–3p–7d) using a pulsed excimer laser system.[95] Unity (or close to unity) ionization depletion from states further away from the ionization limit (the 3p state in Na in this particular case) has been demonstrated using continuous-wave excitation by Schenck et al.[393] The increased illumination time for continuous-wave excitations compensates for the lower ionization rate of the 3p state.

5.5.1.3 Two-step versus One-step Laser-enhanced Ionization As was alluded to earlier, a one-step excitation

is not sufficient to obtain a high ionization yield for elements with higher ionization potential. An alternative then is to excite the atoms under study in a two-step excitation process, as was discussed in section 3.4. The two-step versus one-step signal enhancement in LEI, defined as S_{12}/S_1, where S_{12} and S_1 are the two-step and one-step LEI signals, respectively, has been found to range from unity (i.e. no enhancement) to a few thousand, depending on element, choice of transitions, irradiances, etc. The use of two-step LEI will thus not only provide a larger sensitivity to the elements with a moderate or a high ionization limit but will also improve on the selectivity because it enhances the signal from the analyte significantly more than that of the background (see section 5.5.5.2).

5.5.1.4 Ionization Efficiency Axner and Berglind determined a closely related entity, referred to as the ionization efficiency (defined as the probability that an excited atom will ionize before returning to the ground state), for excited *n*p states in Na and Li.[90] They found that the ionization efficiencies for these elements were considerably higher than those given by a simple Boltzmann factor (close to unity for states within $1\,kT$ from the ionization limit and around 50% for states $2.5\,kT$ below the ionization limit), thus supporting the discussion above (section 3) about an effective ionization rate that is larger than a single state-specific ionization rate.

5.5.1.5 Overall Efficiency There have been several attempts to model the overall efficiency of LEI based upon the basic assumptions given above. The total number of charges created (per laser pulse), for example, can, in general, be written as a product of the density of analyte species, the interaction volume, and the ionization yield. It has been found, however, that such descriptions do not always agree with the experimental findings with expected accuracy.[74] Although the dependence of most parameters (e.g. laser pulse energy and analyte concentration) is well understood, not all dependencies have yet found a satisfactory explanation. The area dependence is such an unexplained feature. In fact, the discrepancy between predicted and experimental behavior has been so extensive that not one single work has yet been able to address successfully this seemingly simple matter for the LEI technique.

It is rather straightforward to argue that there should be an optimum laser beam area for a given laser pulse energy.[57] As long as the transition is saturated, and the cross-section of the laser beam is smaller than the distribution of the analyte atoms in the part of the atomizer in which there is a sufficient charge collection efficiency, the total number of charges created can be increased by enlarging the laser beam area. As the beam area becomes larger, the degree of optical saturation will decrease, eventually implying that the linear excitation regime will be reached. In this case, any further defocusing will no longer increase the number of charges created (for a one-step LEI case) because the increasing number of atoms illuminated will be balanced by a decreasing fraction of atoms excited in the interaction volume. Eventually, when the beam area becomes larger than the distribution of the analyte atoms in the atomizer (or the part of the atomizer in which there is a sufficient electrical field for collection of the charges created), the total number of charged created or detected should decrease. However, this behavior has never been verified experimentally.

5.5.1.6 Influence of "Scattered Laser Light" A possible explanation as to why this discrepancy exists was given by the findings of Sjöström and Axner when they found that the signal strength of the LEI technique could be influenced significantly by what they referred to as "scattered laser light".[394]

Figure 14 shows the results from an experiment on Sr atoms in an acetylene–air flame. A laser with a small beam diameter (in this case 2 mm) was scanned across the $5s^2\,{}^1S_0 - 5s\,5p\,{}^1P_1$ transition around 460.733 nm for four different laser pulse energies in Figure 14(a). A distinct peak from Sr atoms is clearly visible in all the four scans. Worth noticing is that even though the irradiance of the light was increased by a factor of 45 from 12 to 550 $kW\,cm^{-2}$, the peak signal only increased by a factor of 2. This behavior is normally interpreted as a manifestation of optical saturation. However, a simple modeling of the behavior of a two-level system shows a clear discrepancy with respect to the experimental findings. The reason is that an optically saturated system shows saturation broadening when the laser wavelength is being scanned over the transition. Saturation broadening originates from the fact that a laser beam that strongly saturates the transition on resonance also saturates the transition when being slightly detuned. This implies that the profile broadens. The result of a modeling of a two-level system (with some photoionization) that replicates the peak values of the experimental curves in Figure 14(a) are shown in Figure 14(b). It can be seen clearly that there is a significant discrepancy between prediction and experimental findings when it comes to the width of the curves.

In order to understand this discrepancy, additional experiments were made. Figure 14(c) shows the results from a corresponding set of experiments for a larger beam area. In this case, the beam area was enlarged 16 times (the beam diameter was expanded from 2 to 8 mm), as was the beam pulse energy (so as to preserve the irradiance of the light and degree of optical saturation). It was then found unexpectedly that the signals were only marginally larger

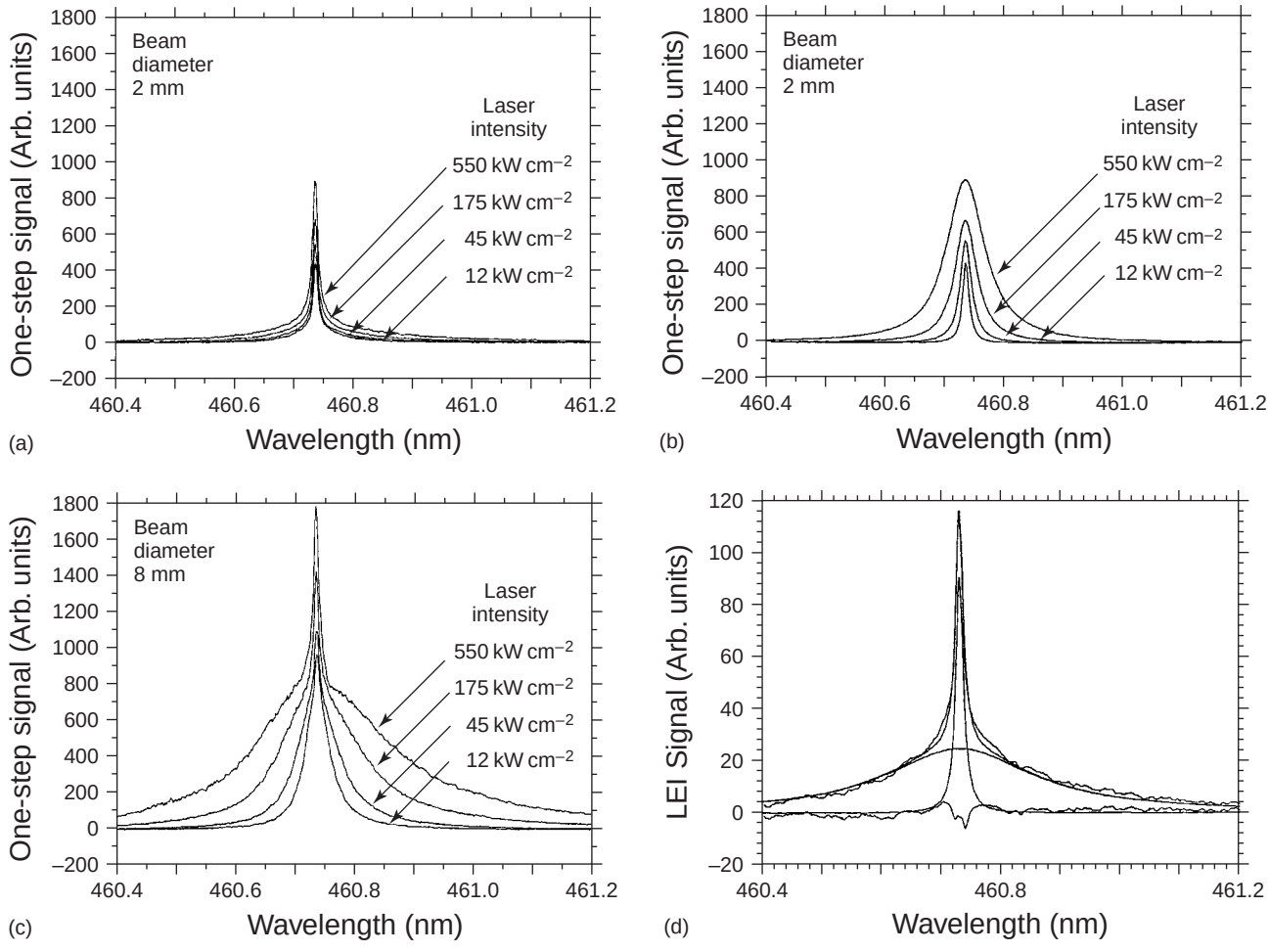

Figure 14 A set of one-step LEI scans over the $5s^2\,{}^1S_0 - 5s\,5p\,{}^1P_1$ transition in Sr in an acetylene–air flame. (a) The situation when light with four different laser irradiances (intensities) was scanned over the transition (12, 45, 175, and 550 kW cm^{-2} as the four curves, from bottom to top, respectively) with a beam diameter of 2 mm. (b) The results from the simplest possible model of the situation that can reproduce the peak values (a two-level model of a Lorentzian absorption profile with some photoionization). (c) A situation corresponding to (a), with a 16-fold larger area (a beam diameter of 8 mm) but with the same irradiances, from which one can see clearly that the LEI one-step signal from Sr is composed of at least two parts. (d) A one-step spectrum (top curve) decomposed into the sum of two Lorentzian-shaped peaks (next two curves), with the bottom curve being the residual error of the decomposition. It can be concluded from a comparison between the widths of the curves in (a–d) that it is the broad peak that behaves as expected from theory. Hence, the narrow peak, which dominates the curves in (a), originates from ''scattered'' laser light. (Parts (a) and (c) are reproduced by permission of the Society of Applied Spectroscopy from O. Axner, S. Sjöström, *Applied Spectroscopy*, **44**(5), 864–870 (1990).)

than with the 16 times smaller beam area (approximately 40–50% larger) and that a significant broadening of the lower parts of the peaks could be discerned. This gave some insight into the true physical origin of the LEI signal in this particular experiment. Analysis of the data showed that the LEI signal, in fact, can have several origins. It was found that each curve could be decomposed into a sum of two curves with significantly different widths, as is shown by Figure 14(d). The top experimental curve can be decomposed into a curve with broad shoulders and one narrow-band feature (next two curves). The curve with broad shoulders was found to follow rather well the predicted behavior of the LEI signal from an optically saturated system, but the narrow-band feature did not. This led to the conclusion that the signal with broad shoulders originates from atoms within the geometrically confined interaction region (referred to as the ''true'' interaction region), whereas the narrow-band peak has to be attributed to the signal from nonsaturated atoms in another part of the atomizer (i.e. outside the ''true'' interaction region) affected by what was referred to as ''scattered laser light''. It was thus concluded that a significant (sometimes even a dominating) part of the signal from Sr can originate from ''scattered laser

light'' under a variety of conditions, as is exemplified by Figure 14(a).

A number of possible mechanisms for the existence of ''scattered laser light'' outside the ''true'' interaction region have been proposed: the light might be due to diffraction of laser light from an aperture positioned outside the flame; it might consist of fluorescence light from excited analyte atoms in the interaction region; or it might originate from light scattered from particles of flame molecules in the ''true'' interaction region of the flame. Despite attempts to understand the physical origin of this ''scattered laser light'', no clear conclusion about its origin has yet been drawn. Furthermore, it has not been investigated to what extent this affects other elements detected by LEI. It is clear, however, that the influence of ''scattered laser light'' is smaller in two-step than in one-step experiments.[394] This also explains why there has been a better agreement between measured and predicted signal strengths for two-step excitations than for one-step excitations in LEI spectrometry.[74]

5.5.2 Noise

The most thorough description of the various sources of noise in LEI spectrometry has been given by Turk.[237] He divides the various types of noise into two categories: multiplicative and additive noise. Multiplicative noise is defined as the noise that is proportional to the concentration of the analyte, and additive noise comes from the other components. This implies that as the concentration of the analyte decreases, the influence of the multiplicative noise decreases accordingly, so that only the additive noise determines the limit of detection. Multiplicative noise, on the other hand, can affect the precision.

5.5.2.1 Multiplicative Noise Multiplicative noise can be due to fluctuations in the atomic population (affected by properties of the nebulizer/flame), the ionization yield (affected by properties of the laser), and the detection efficiency (affected by properties of the applied potential and the flame). Although pulse-to-pulse variation in the laser power can be substantial, it has been found that there is little correlation between the fluctuations of the LEI signal and those of the laser power.[395] The main reason is attributed to optical saturation. Instead, in a simultaneous measurement of LEI and LIF in a flame, it was found that the correlation between the noise in the LEI and LIF signals was substantial. This provides evidence that the major source of multiplicative noise in this particular case resulted from fluctuations of the atomic population within the laser-irradiated volume.[395]

5.5.2.2 Additive Noise Additive noise can consist of electronic noise, thermal background ionization noise and laser-induced background ionization noise. Electronic noise can in turn consist of either Johnson noise in resistors or radiofrequency interferences emitted from pulsed lasers synchronous with the laser pulses. As was concluded by Turk, the electrodes, burner head and preamplifier seem to form an excellent antenna for radiofrequency interferences.[237] Care in shielding and grounding the LEI instrumentation is therefore essential in LEI spectrometry. A comparison between the nonlaser-induced sources of noise for two flames (acetylene–air and hydrogen–air) and two different samples (synthetic drinking water with a few to some tens of ppm of EIEs and citrus leaves, diluted and digested to $10\,mg\,mL^{-1}$ with one to two orders of magnitude higher concentrations of EIE) were made by Turk.[237] It was found that although the root mean square of the electronic noise (no flame on) was 1.3 nA (unfiltered), the flames themselves contributed very little to this noise level. When drinking water was aspirated, however, the noise increased to approximately twice the value of the electronic noise (to 2.7 and 1.7 nA for the two flames, respectively). When the high-EIE samples were aspirated (citrus leaves), on the other hand, the noise increased considerably (to 23 and 6.2 nA, respectively). This shows that the dominating source of noise originates from the increased thermal ionization from EIE when samples with high concentrations of such elements are detected. In the absence of EIE most of the noise can be attributed to either electronic noise or flame current noise. Turk concluded furthermore that when the noise is dominated by flame noise, the noise magnitude is consistent with shot noise from the flame DC. When also considering the effect of signal averaging, the noise levels given above were found to be decreased significantly. Assuming a time constant of 1 s and exponential averaging over 10 Hz (as typical for Nd : YAG systems) the noise levels in the acetylene–air flame decreased to 0.45 and 3 nA (for the two samples, respectively), whereas the corresponding levels for a 300-Hz excimer system become 0.07 and 0.52 nA.

The relative importance of laser-induced background ionization noise is more difficult to assess because this type of noise depends on the amount of background signal generated by the laser (which normally is considered to be a type of interference). Irrespective of its origin (it can, for example, be due to wing excitation or multiphoton ionization of flame molecules), one can conclude, however, that this type of noise is particularly destructive for the LEI technique because in general it is affected by the fluctuations of the laser power (because wing excitation and multiphoton ionization are seldom optically saturated) as well as the atomic population in the flame, two entities that are known to carry a significant amount of noise.

5.5.3 Limits of Detection

5.5.3.1 Laser-enhanced Ionization in Flames As was alluded to above, the most commonly used atomizer for LEI has been the flame. A rather large number of detection limits for LEI in flames have therefore been published throughout the years. The best LOD for each element have been collected in Table 5 (see references 72, 293, 306, 321, 322, 328, 336, 364, 366, 367, 370, 371, 383–385, 387, 388, 392, 396–399). The LOD have been stated as given in the original references. This implies that although some elements have LOD that are limited by the actual noise of the background (i.e. including background signals from contamination, etc.), some have been estimated from the instrumentation noise (i.e. obtained when detuning the laser(s) from resonance so as to avoid the influence of contamination). In addition, although most LOD have been given with an S/N of 3, measured over 1 s, some are taken under slightly different conditions (or under unspecified conditions).

As can be seen from Table 5, the majority of the LOD are in the picogram per milliliter region with a few even below, i.e. in the femtogram per milliliter range (In, Li, Na, Pb and Rb), and a limited number above, i.e. in the nanogram per milliliter range (As, Eu, Mo, Sb, Si and W).

The best LOD correspond fairly well to the realistically estimated best LOD for the LEI technique. Turk calculated the best possible LOD for two different

Table 5 LOD for the LEI technique in flames (ng mL^{-1})

Element	First-step excitation (nm)	Second-step excitation (nm)	Laser[a]	Flame[b]	LOD	Refs.	Additional Refs.[c]
Ag	328.1	421.1	E	AA	0.05	72	
Al	308.2/309.2		F	AN	0.2	321	371
As	278.0		E	AA	3000	364	
Au	242.8	479.3	Y	AA	1	383	364
Ba	307.2		F	AA	0.2	293	364
Bi	227.7		E	AA	0.2	371	
Ca	227.6		E	AA	0.006	371	387, 396
Cd	228.8	466.2	Y	AA	0.1	383	271
Co	240.8		E	AA	0.06	371	383
Cr	240.9		E	AA	0.2	371	
Cs	455.5		N	AA	0.002	328	322, 388
Cu	324.8	453.1	Y	AA	0.07	383	336
Eu	459.4	564.0	N	AA	4000	397	
Fe	271.9		E	AA	0.1	364	364, 371
Ga	241.9		E	AA	0.03	371	306, 364, 366
In	303.9	532	Y	AA	0.0004	336, 398	328, 364
Ir	266.5	562.0 + 642.0	E	AA	0.3	367	
K	404.4		N	PBA	0.1	322	392
Li	670.8	460.3	E	AA	0.0003	72	293
Mg	285.2		E	AA	0.005	366	
Mn	279.5	521.5	Y	AA	0.02	384	364
Mo	319.4		F	AN	10	321	
Na	589.0	568.8	N	AA	0.0006	328	399
Ni	229.0/232.0		E	AA	0.02	371	
Pb	282.3	600.2 + 1064	Y	AA	0.0007	398	
Rb	420.2		N	AA	0.0006	387	
Sb	287.8		E	AA	50	364	
Si	288.2		F	AN	40	321	
Sn	284.0	597.0	Y	HA	0.3	383	321
Sr	230.7		E	AA	0.006	371	
Ti	318.6/319.2/320.0		F	AN	1	321	
Tl	276.8		E	AA	0.006	364	366, 370
V	318.4/318.5		F	AN	0.9	321	
W	283.1		E	AA	300	364	
Yb	555.6	581.2	Y	AA	0.1	328	
Zn	213.9	396.5	Y	AA	1	385	

[a] E = excimer-pumped dye laser system; N = nitrogen-pumped dye laser system; Y = Nd : YAG-pumped dye laser system; F = flashlamp-pumped dye laser system.
[b] AA = acetylene–air; AN = acetylene–nitrous oxide; PBA = propane–butane–air; HA = hydrogen–air.
[c] Additional references with LOD that are within a factor of three from the best. These works in general do not use the same type of laser, wavelength or flame as that specified.

situations.[337] Under the somewhat unrealistic condition that the detection power should be limited by the shot noise of the thermal ionization, he found that the LOD should be 3 fg mL^{-1}. This is far below any experimentally found LOD. When basing the calculation upon a more realistic situation (assuming that the detection power is limited by the actually measured noise values), the LOD was found to be of the order of 0.1 pg mL^{-1}. As can be seen from Table 5, there are a few elements that have their LOD within one order of magnitude from this estimated optimum LOD (primarily In, Li, Na, Pb and Rb). This indicates that there is still room for improvements for most of the detected elements for LEI in flames.

5.5.3.2 Laser-enhanced Ionization in Plasmas The detection power of laser-based ionization techniques has not been improved as much as for conventional techniques when going from flames to ICPs, for reasons given in the instrumentation section (section 5.4).

The first attempts to combine an ICP with a laser-based ionization technique were made by Turk and Watters. In order to circumvent these problems, the authors used an extended plasma torch and placed two charge-collecting electrodes 19 cm above the load coil (in the tail flame, so as to benefit from the higher neutral atom fraction that exists in the cooler part of the plasma and to minimize the influence of radiofrequency interferences).[323] The detection powers were clearly inferior to those previously attained by the LEI technique in flames (e.g. LOD for Cu of 7 μg mL^{-1}).

In order to increase the detection power of LEI in an ICP and to perform measurements in the analytically most preferable position in the plasma, Turk et al. power-modulated an ICP by temporarily turning the plasma off (reducing the influence of radiofrequency interferences and decreasing the plasma background ion/electron concentrations).[324] Ionization measurements were made 1.4 ms after complete interruption of plasma. LOD in the nanogram per milliliter range were achieved: for Fe and Ga, 80 and 20 ng mL^{-1}, respectively. It was concluded that the primary mode of ionization of the laser-excited atoms was photoionization rather than collisional ionization (i.e. more in-line with RIS).

Other atomizers have been investigated for possible application in LEI spectrometry, but with limited success. The microwave-induced plasma, for example, has been investigated by Seltzer and Green.[326] Another example is the use of an atmospheric pressure micro-arc plasma with LEI, which was investigated by Churchwell et al.[300]

5.5.3.3 Laser-enhanced Ionization in Graphite Furnaces As was alluded to above (in section 5.4), a variety of experimental configurations of furnaces and probes have been scrutinized for LEI in a GF. The common experiences are that the sensitivity is not as large as expected and that signal fluctuations are considerable.

For example, already the first measurements indicate that there is a somewhat lower collisional ionization rate (or yield) in the furnace than in the flame. One such indication was the finding by Gonchakov et al. that the LEI signal from two-step excitation of Na could be increased significantly (by a factor of 20–40) by actively photoionizing the atoms with infrared light from a Nd : YAG laser.[301] Magnusson et al. compared signal strengths (on a sensitivity basis) from flame LEI and LEI in a GF and concluded that the Na signal from the latter was smaller than expected and that from Mg was larger.[329] The smaller Na signal strength was attributed to a lower collisional ionization rate in the furnace than in the flame, whereas the larger Mg signal was explained partly as photoionization.

Moreover, many authors have noticed that there was always a background signal, fully resonant with the species detected, as soon as any liquid was injected into the furnace (corresponding to that from a few picograms of analyte).[301,329,330] Magnusson et al. showed explicitly that there were no background signals when the furnace was heated empty (with the lasers on) and concluded that the background signal originates from contamination.

Many LOD for LEI in a GF are in the picogram range. Magnusson et al. quoted contamination-limited LOD that are in the picogram range (Co, 5 pg; Cr, 5 pg; Mg, 1 pg; Mn, 1 pg; Na, 5 pg; Ni, 7 pg; Pb, 0.5 pg).[329,330] The contamination-free LOD for Mg and Na were estimated to be 0.03 and 0.5 pg, respectively (obtained from measurements when no sample was injected into the furnace).[329] Gonchakov et al. had previously estimated the LOD for Na in the absence of contamination to be in the low femtogram range.[301] Moreover, the LOD for Mn and Sr measured with the T-furnace by Magnusson et al. were assessed to be 1 and 2 pg, respectively.[331] Detection of Yb and In, using two-step excitation with the probe above the dosing hole, yielded LOD of 1 pg and 0.8 fg, respectively.[333]

Butcher et al., finally, found that the limiting noise for their LEI system in a modern HGA-500 furnace with a moveable probe was furnace noise and thermionic emission. LOD were established for a number of elements (Tl, 2 pg; In, 0.7 pg; Pb, 60 pg; Li, 1 pg; Mg, 10 pg; Mn, 30 pg; Fe, 50 pg; the LOD were determined on-line for all elements except Li and Mg). In agreement with previous authors, they found that the precision was poor (between 12 and 16%). They attributed this to poor precision of the manual pipetting, but gave no conclusive evidence as to whether this originated from any parts of the LEI or GF instrumentation (e.g. fluctuations in the charge-collection efficiency) as had been suggested earlier.

It can be concluded that the furnace does not bring the same advantages for the LEI technique as it does for the LIF/LEAFS technique, for the reasons described in section 5.4.2.3.

5.5.3.4 Laser-enhanced Ionization in Rod–Flame Systems As was alluded to above, an alternative means of atomization is the rod–flame system developed by Chekalin et al.[333,335–338] The most significant advantages with this system come from its ability to handle samples with high concentrations of concomitant elements and for analyses of solid materials. It was, for example, concluded that the background signal from the matrix decreased by over two orders of magnitude when samples in solid form were analyzed in comparison with samples in liquid form. This allowed for detection limits in solid samples down to the picogram per gram level. LOD for the rod–flame system are generally low (Au, 200 fg;[336] Co, 1 pg;[338] Cr, 2 pg;[338] Cu, 2 pg;[335,336] In, 4 fg;[335,336] Mn, 3 pg;[338] Na, 2 fg;[336] Ni, 800 fg[338]).

The rod–flame instrumentation has been used for detection of: In in CdHgTe solutions,[335,336] Au in $AgNO_3$,[333] Cu in Ge,[335,336] Na and Cu in orthophosphoric acid,[333] Li, K, Na, Sb, Cu and Ag in phosphoric acid, acetone and isopropanol,[337] Ni, Co, Cr and Mn in fiber-optic fluoride materials,[337] and Cr, Co, Mn and Ni in NH_4F and NaF.[338]

5.5.3.5 Laser-enhanced Ionization in Graphite Furnace–Flame Systems Two-step excitation from a pulsed Nd : YAG-pumped dye laser system was used to detect Mg, Tl and In in the previously described combined GF–flame system developed by Smith et al.[339] After optimization of various parameters (e.g. carrier gas flow), the LOD for these three elements could be assessed as 1, 120 and 260 fg, respectively.

Riter et al. determined experimentally the vaporization, transport, atomization, probing, and detection efficiencies of Mg atoms in an LEI system in which a GF was coupled to a miniature flame.[340] The authors estimated the overall efficiency of the system to be 0.0025%. The experimental LOD was 20 pg. The LOD for Mg was estimated to be 6 fg in the absence of the blank and after a reduction of radiofrequency noise.[340] The same group has subsequently used an LEI–GF–flame system for the determination of Pb in whole blood. An LOD of 90 pg mL^{-1} (0.9 pg absolute) was found for lead in whole blood.[400]

5.5.3.6 Laser-enhanced Ionization in Cells Mercury has been studied by ionization techniques directly in a cell.[178,270,341–346] Clevenger et al. studied the temporal behavior of the LEI signal of Hg in a quartz cell under low buffer gas pressure. Using fast electronics and a short laser pulse, it was possible to distinguish, in one single time-resolved ionization waveform, the nonselective photoionization component of the signal from that due to collisional ionization from selected levels. Experimental results were shown to agree with those obtained by computer simulations, and optimal conditions for deconvolution of the two components were assessed.[343] A spectroscopic study of Hg Rydberg states ($n = 10–42$) was also made. Broadening and splitting caused by the influence of Stark effects and of increasing buffer gas pressure were observed. The optimal operating conditions, in terms of pressure and applied high voltage, for obtaining the best sensitivity and LOD by analytical LEI spectroscopy were identified.[344]

A new method for the detection of Hg atoms in an inert gas atmosphere by LEI, based on the avalanche amplification of the signal resulting from the ionization from a selected Rydberg level reached by a three-step laser excitation of mercury vapor in a simple quartz cell, was developed by Clevenger et al.[345] An avalanche amplification effect was found when Ar and P-10 gases were used at atmospheric pressure. The authors estimate that an LOD of 15 Hg atoms per laser pulse in the interaction region is achievable under amplifier noise-limited conditions.[345] A comparison of various means to detect Hg (including laser-based techniques) is given by Clevenger et al.[346]

5.5.4 Interferences

Interferences are phenomena that adversely affect the performance of a technique. There are two types of interferences in LEI spectrometry: space-charge and laser-induced interferences. Although space-charge interferences are exclusively associated with EIEs and affect the charge-collection process, spectral interferences can originate from either matrix elements (to a large extent EIEs) or flame native molecules and therefore affect primarily the excitation–ionization process.

5.5.4.1 Space-charge Interferences Space-charge interferences originate from the fact that (for a given potential applied across the flame) the electrical field distribution depends strongly on the thermal ionization rate.[307,316,401]

As was described in the instrumentation section (section 5.4), applying a potential across a weakly ionized plasma (as LEI atomizers can be characterized as) gives rise to two distinctly different regions: the sheath, in which there is a finite electrical field; and a fieldless region (close to the anode). Only charges created in the sheath will contribute to the signal. Because the extent of the sheath depends on the amount of thermally produced charges, the active interaction volume (i.e. the volume in the flame in which the laser beam overlaps with a nonzero external

field) will decrease (i.e. shrink towards the cathode) with increasing amount of EIEs (because EIEs contribute significantly to the thermal ionization in the flame). Hence, when samples with high concentrations of EIEs are investigated, only the charges created close to the cathode will contribute to the signal. Moreover, because normally only the electronic part of the signal is being measured, the creation of a pair of charges close to the cathode (from which the electronic part of the total LEI signal is larger than the ionic part) will give rise to a larger LEI signal than if the charges were created further away from the cathode. This implies that the laser beam preferably should be directed close to (i.e. just below) the immersed electrode in order to have the highest immunity to EIEs and yield the highest signal. This also explains why the externally positioned cathodes are not capable of handling samples with high concentrations of EIEs (they have their cathode sheath across a cold region in which no free atoms exist). Although the influence of space-charge interference can be reduced by a proper choice of electrode and probe volume, this type of interference is a considerable limitation to the applicability of the LEI technique.

A flow injection system has been interfaced to an LEI apparatus to handle samples with high concentrations of EIEs by Wang et al.[402] With the use of the flow injection/laser-enhanced ionization (FI/LEI) apparatus, the electrical interference induced by the matrix could be effectively diminished. It was found that the FI/LEI technique is capable of detecting indium in a solution containing an Na matrix of more than 40 ppm ($\mu g\,mL^{-1}$), which was about 20 times higher than what the conventional LEI apparatus could tolerate. The authors state that it also has a much larger LDR than conventional LEI. Additional advantages are good reproducibility and a rapid sampling rate.[402,403]

5.5.4.2 Spectral Interferences Spectral interferences can be separated into the following categories: ionization of native flame molecules; matrix spectral interferences (i.e. wing excitations or direct overlap between the spectral line of the analyte species with those of the matrix elements); and ionization of matrix elements by laser ASE.

Although there are several molecular species present in a flame, there are only a few that give rise to ionization signals. The NO molecule is the flame species that gives rise to the most severe flame background signals. With its strongest bandhead around 227 nm, and with others at around 237 and 248 (etc.) nm, it is a source of background signals primarily in the UV region. It has been found, however, that the ionization signal from NO is due to a two-photon excitation, in which case the signals can be held at moderate levels by avoiding focusing too strongly in the flame. Most other flame molecules that contribute to background signals in fluorescence or emission techniques, e.g. OH, have a larger tendency to dissociate than to ionize when being excited under the conditions typically prevailing in LEI spectroscopy, so they do not contribute to the LEI background signal.

Matrix spectral interferences, on the other hand, are in general a more severe problem in LEI spectrometry. Direct overlaps between a transition in an analyte and that of a matrix are rather scarce. Wing excitations of transitions in matrix elements, on the other hand, are much more frequent. It has been found that wing excitations of ubiquitous alkali and alkaline earth elements give rise to substantial background signals in LEI spectrometry. One such example is seen in Figure 15, which shows the LEI signal from a $20\,\mu g\ mL^{-1}$ Na solution in the vicinity of the 3s–3p transition. As can be seen, wing excitations exist even several nanometers from the transition wavelength.

In addition, alkali elements often have quite a low ionization limit. When such an element is exposed to laser light with wavelengths below their threshold wavelength for direct photoionization from the ground state, a structureless LEI background signal appears.[372,377] The threshold wavelengths for direct photoionization for Li, Na, K, Rb and Cs are 230, 241, 286, 297 and 318 nm, respectively. Hence, whenever UV light is used for investigating samples with large amounts of alkali matrix elements, a significant and unstructured LEI background signal will appear. As with all background signals, this will add to the existing noise in the system, thus reducing the detection power of the LEI technique.

A thorough study of background signals (direct overlap, wing excitations and photoionization) from Na in the

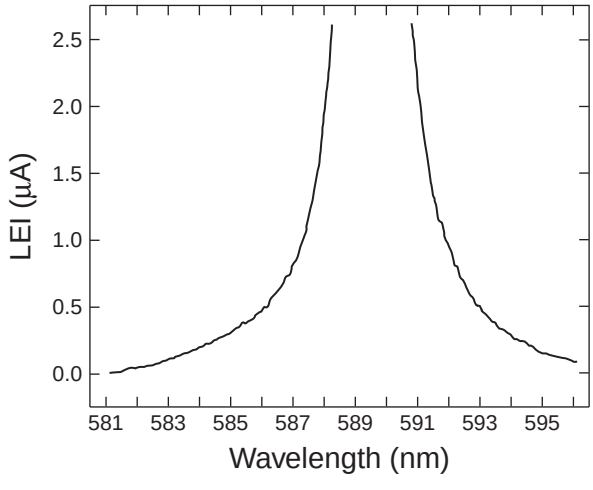

Figure 15 One-step LEI signal from $20\,\mu g\,mL^{-1}$ of Na in the vicinity of the 3s–3p transition. (Reproduced by permission of John Wiley & Sons, from G.C. Turk, in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, 161–211, 1996.)

entire 200–450 nm region was conducted by Axner et al.[377] They found that even when the laser was detuned as much as possible from the discrete transitions, there was an Na signal from wing excitation that will create a background signal whenever samples with large amounts of Na are detected, as is shown in Figure 16. It was found that the background signal originates mainly from direct photoionization from the ground state below 240 nm; excitation to a number of closely spaced Rydberg levels gives rise to a strongly structured background signal in the 240–260 nm region; direct overlap or wing excitations of the strong 3s–4p, 3s–5p, 3s–6p and 3s–7p transitions dominate between 260 and 360 nm, whereas photoionization of thermally excited 3p atoms predominantly takes place slightly below 420 nm.

Moreover, a part of the light emitted from a dye laser consists of ASE. This broad-band light, which occurs within the tuning curve of the dye used, is usually of much lower power than the laser light but nevertheless can be responsible for excitation of concomitant elements. Because the amount of ASE emitted from a laser is not strongly dependent on the laser light wavelength (except at the edges of the tuning curves where it increases), it gives rise to a rather wavelength-independent background signal.

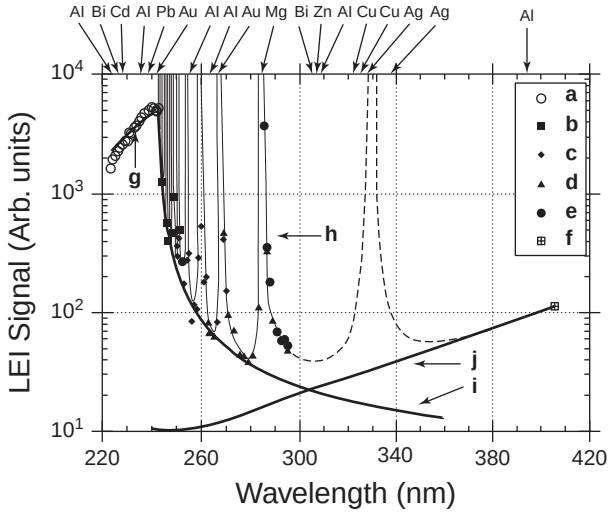

Figure 16 Experimentally measured (intensity- and concentration-normalized) LEI signal strengths for Na versus wavelength. The open and solid markers represent individual measurements done by various dyes in the dye laser: **(a)** Coumarin 47; **(b)** Coumarin 102; **(c)** Coumarin 307; **(d)** Coumarin 153; **(e)** Rhodamine 6G; **(f)** DPS. The predicted LEI signal from direct photoionization from the ground state (below 240 nm) is indicated by solid curve **(g)**. Various discrete resonance transitions are indicated with the solid and dashed line **(h)** (the peak at around 185 nm corresponds to the 3s–5p transition). The lowest contribution from wing excitations of the 3s–*n*p transitions in each wavelength region is indicated by solid line **(i)**. The predicted contribution from photoionization of the thermally populated 3p state is indicated by the solid curve **(j)**. (Reproduced by permission of the Society of Applied Spectroscopy from O. Axner, M. Norberg, H. Rubinsztein-Dunlop, *Applied Spectroscopy*, **44**(7), 1124–1133 (1990).)

5.5.5 Selectivity

It has been corroborated frequently that LEI is a technique that has a high sensitivity (implying that the number of charges created per photon absorbed is high, in some cases close to unity). This is especially the case for the alkali elements because they have a rather low ionization potential and are easily atomized in a flame. This is, however, also the Achilles heel of the LEI technique. As was shown for Na above, all alkali elements will give rise to various types of spectral interferences (wing excitations or photoionization from lower-lying states). Any analyte to be detected in a matrix of alkali element therefore has to be detected on top of a background signal. The high sensitivity of the alkali elements thus reduces the detection power of other analytes and reduces the selectivity of the LEI technique. The selectivity of LEI therefore has been found to be lower than that of the LIF/LEAFS technique. This is an unrelenting consequence of the fact that LEI is a technique in which the selectivity provided by the laser source is limited by the amount of resolution it can provide when coupled with an undiscriminating charge-collecting detector.

5.5.5.1 Means to Improve Selectivity – Proper Choice of Laser Irradiance One possible remedy to the aforementioned problem of interferences and limited selectivity is to excite the analyte atoms with as low an irradiance as possible in order to avoid excess optical saturation of the transitions of the analyte (and minimize the effect of background signals from matrix elements).

An example is shown in Figure 17. So far, the most severe spectral interference encountered for LEI spectrometry occurs when Mg is to be detected in the presence of Na. The reason is that Mg has its only practically accessible transition ($3s^2$–3s 3p) at 285.21 nm whereas Na has one of its strongest transitions (3s–5p) at 285.28 nm, which is only 0.07 nm away. It is therefore clear that significant wing excitations of Na atoms will occur when the laser is tuned to the Mg resonance. Figure 17(a) shows the situation for one-step excitation of 10 ng mL^{-1} Mg in a matrix of 10 times more Na, i.e. 100 ng mL^{-1}, using a laser irradiance of 25 kW cm^{-2}. As can be seen from Figure 17(a), there is a spectral interference at the wavelength of the Mg transition already for this rather moderate amount of Na matrix under these conditions.

One way to increase the selectivity is to adjust the laser irradiance so as to avoid excess optical saturation of the Mg transition. Figure 17(b) shows the situation when the

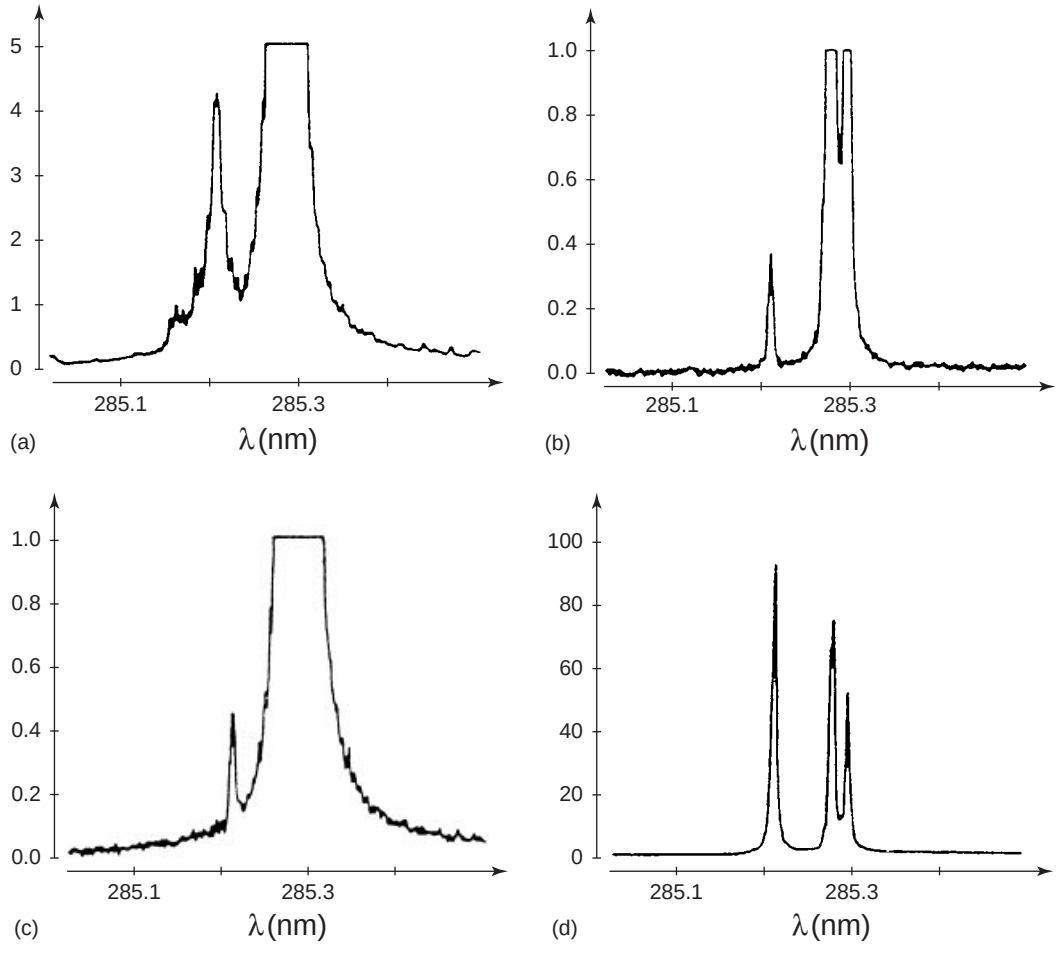

Figure 17 LEI signals from solutions of 10 ng mL^{-1} Mg in various matrices of Na for a variety of experimental conditions: (a) one-step excitation using 25 kW cm^{-2} irradiance laser light in a matrix of 100 ng mL^{-1} Na; (b) one-step excitation using 0.5 kW cm^{-2} irradiance laser light in the same Na matrix as in (a); (c) one-step excitation using the same laser light irradiance as in (b) in a matrix of 1 μg mL^{-1} Na; (d) two-step excitation using the same laser light irradiance in the first step as in (b) and (c) and an irradiance of approximately 4 MW cm^{-2} fixed on the second-step transition in Mg (at 470.30 nm) as the second step in the same Na matrix as in (c). (Reproduced by permission of The Royal Swedish Academy of Sciences from I. Magnusson, O. Axner, H. Rubinsztein-Dunlop, *Physica Scripta*, **3**, 429–433 (1986).)

laser irradiance has been attenuated to 0.5 kW cm^{-2}. The interference from Na at the Mg transition has decreased considerably.

The increase in selectivity is not sufficient, however, when samples with higher Na concentrations are to be analyzed. Figure 17(c) shows the situation when the Na concentration has been increased by one order of magnitude (i.e. to 1 μg mL^{-1}).

5.5.5.2 Means to Improve Selectivity – Use of Two-step Excitations By illuminating the atoms from the example above also with 470.30-nm light, in order to excite the Mg atoms closer to the ionization limit by a two-step process ($3s^2$–3s 3p–3s 5d), the Mg signal will be increased and thereby also the selectivity. In this particular case, the signal enhancement was about 300, as is shown in Figure 17(d), and thus virtually all of the Na interference at the Mg transition has been eliminated.

However, although an optimization of the working conditions (regarding laser irradiance, laser beam area, etc.) and the use of two-step excitation can increase the selectivity considerably, they cannot completely eliminate the spectral interferences. This is illustrated in Figure 18, where the influence of wing excitation on the detection of Ni in an Na matrix is displayed. Nickel atoms are excited by a two-step excitation (at 300.3 and 561.5 nm) for best sensitivity. The second-step laser is scanned across the second-step excitation in Ni for two water solutions containing 100 ng Ni mL^{-1} with and without the presence of 100 μg Na mL^{-1} of matrix. A background signal is clearly visible for the sample containing the Na matrix. In this case, the background signal originates

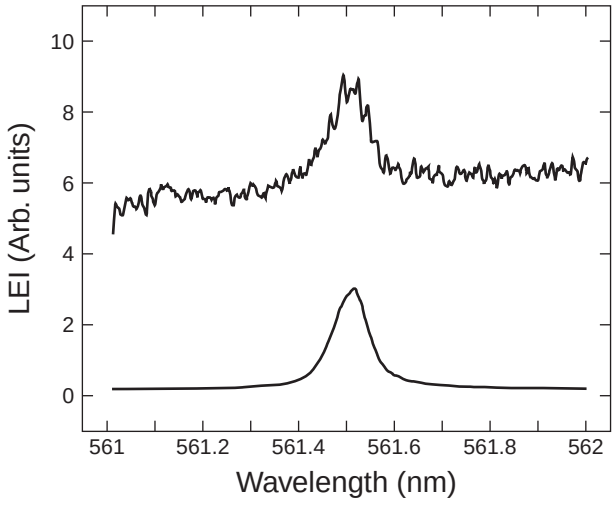

Figure 18 Two-step LEI spectra for 10 ng mL^{-1} Ni in pure deionized water and in a matrix of 100 μg mL^{-1} Na as the lower and upper curves, respectively. The first-step laser is fixed on the 300.249 nm transition in Ni and the second-step laser is scanned over the second-step excitation at around 561.5 nm. (Reproduced by permission of John Wiley & Sons, from G.C. Turk, in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, 161–211 (1996).)

from a wing excitation for the 3p–4d transition with a central wavelength of 568.8 nm. Although the presence of 100 μg Na mL^{-1} does not influence the charge collection efficiency (the Ni peaks are virtually the same), it considerably decreases the S/N ratio.

5.5.6 Applications

The LEI technique has been used for a variety of applications, both analytical as well as diagnostic. A thorough review of the analytical applications of LEI has been given by Green.[317] A common analytical application is the assessment of traces in standard reference samples. Another is to determine the metal content in rocks.[404] The LEI technique has also been used with liquid chromatography in order to create a very sensitive method for speciation: organolead species were studied by Epler et al.,[405] and Cr speciation was studied by Paquet et al.[381]

Another example of an application of LEI is the development of a resonance photon ionization detector,[303,406–410] which is a narrow-band detector for photons or collisional energy transfer. In short, the detector consists of a two-step LEI instrumentation in which the second-step excitation is provided with light from a laser system. This system then acts as a sensitive and narrow-band detector of light whose wavelength corresponds to that of the first excitation step. One area of application for this technique is to serve as a detector for Raman light;[303,406,407,411,412] another area is to measure temperatures in a flame with high spatial resolution.[408,409]

A field-deployable flame-based LEI spectrometry system (termed laser optogalvanic spectrometry) has been developed to measure the concentration of metal species present in the near-atmospheric pressure off-gases of mixed-waste thermal treatment systems.[413]

The technique has also been used for assessment of unidentified transitions from Fe(I) from solar spectra, resulting in the possibility for several high-lying states in Fe(I) to be identified and labeled correctly.[380]

5.6 Diagnostic Capabilities

The high sensitivity of the LEI technique makes it useful as a diagnostic tool. A number of various entities and phenomena have been investigated by the LEI technique throughout the years and a few examples in this field follow.

5.6.1 Laser-enhanced Ionization in Flames

Studies concerned with the excitation and ionization processes (including energy pooling effects, charge exchange rates, etc.) in a variety of atomic reservoirs have been performed by a number of authors.[270,357,360–362,414] For example, Turk developed a methodology that can determine where within a flame LEI actually takes place.[414a,414b] He found that spatial control of the flame volume sampled by the ionization detection electrodes could be obtained by varying the applied voltage across the flame. This procedure was used to observe anomalous contributions to LEI caused by scattering or fluorescence from the laser beam to other regions in the flame. Axner and Berglind demonstrated that the LEI technique can be used to measure the electrical field distribution in a flame by monitoring the Stark-splitting Rydberg states that are exposed by the electrical field.[373] The ionization efficiency of Na atoms in an acetylene–air flame (defined as the probability than an excited atom ionizes rather than returning to the ground state) was measured by the same authors in another work.[90] The ionization yield (defined as the fraction of atoms in the interaction region that ionize) has been measured for a few elements in acetylene–air flames.[415] The LEI technique has also been used for measuring the atomization efficiency of atoms in an acetylene–air flame,[416,417] ionization rates,[418] flame temperatures[418,419] and dynamic Stark effects (i.e. effects on the energy levels induced by the laser light).[109]

A demonstration of the applicability of a delayed two-step LEI technique for measurement of the lifetimes of metastable states (in Pb and Tl) in an acetylene–air flame was first demonstrated by Omenetto et al.[81] Axner et al. made a more through study of nine different

metastable states under a variety of conditions in an acetylene–air flame.[82] A few such examples are given in Figure 19. They concluded, among other things, that metastable states that are forbidden to decay to lower states because of a parity violation generally have longer lifetimes (85 ns–3.1 μs) than those that are forbidden to decay by violation of spin conservation (5–33 ns).[82] They also showed that the lifetime of a metastable state depends strongly on the local stoichiometry in the interaction volume (the lifetime of a specific metastable state in Au was found to range between 600 ns and 8.8 μs, depending on the local stoichiometry), which led to the finding that the LEI technique could be used to map the local stoichiometry in an acetylene–air

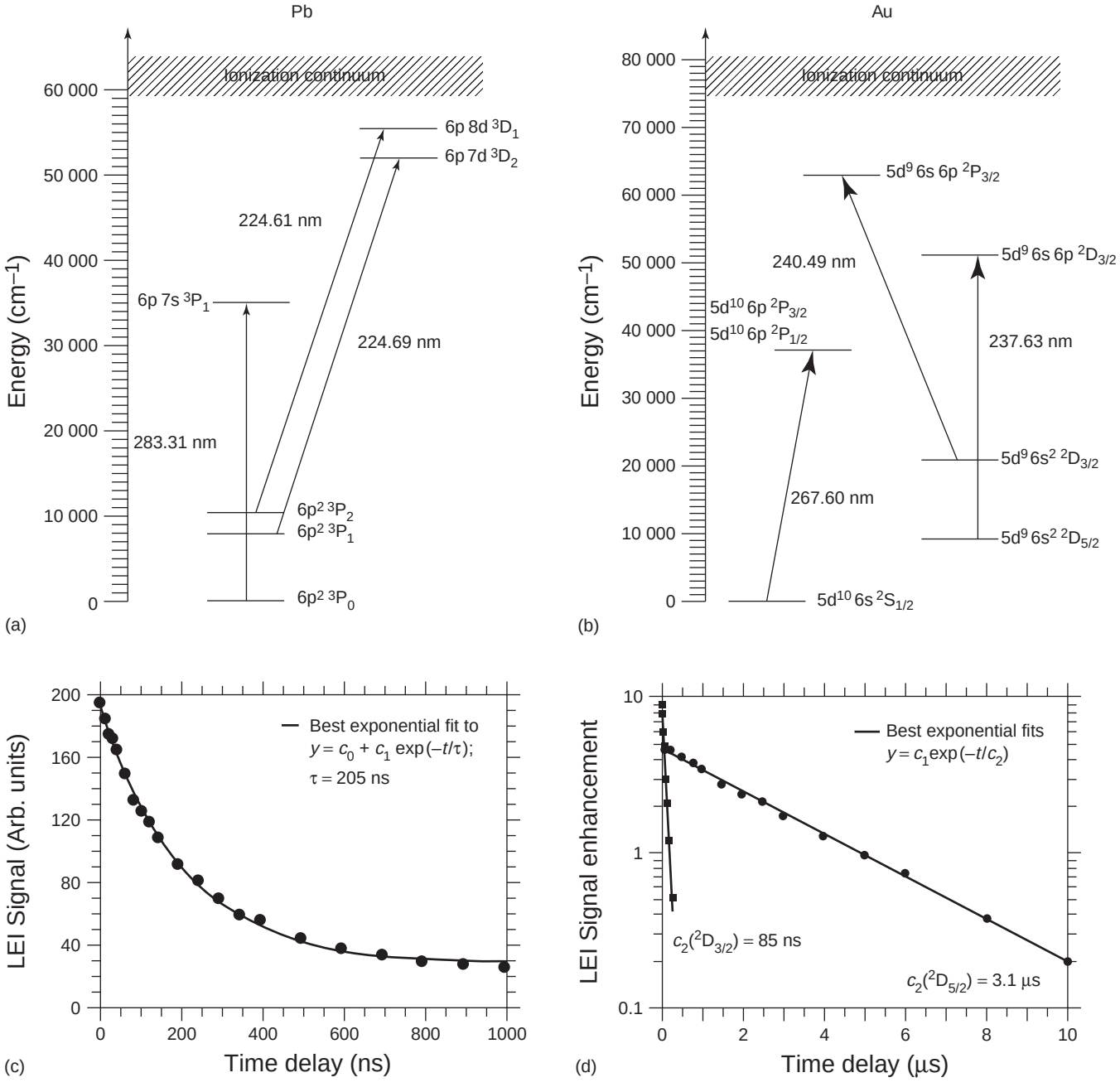

Figure 19 (a, b) Simplified energy level diagrams of Pb and Au in which some pump-and-probe excitations are marked. (c, d) Display decay curves for the $6p^2\ ^3P_1$ metastable state of Pb and the $5d^9\ 6s^2\ ^2D_{3/2,5/2}$ metastable states of Au, respectively, in an acetylene–air flame. (Reproduced by permission of Springer Verlag from O. Axner, P. Ljungberg, Y. Malmsten, *Applied Physics B*, **54**, 144–155 (1992).)

flame by monitoring the lifetimes of atomic metastable states.[82–84]

The two-step LEI technique has also been used to measure flame flow velocities or gas velocities in a shock tube.[420–423] One laser system is used to deplete the density of neutral atoms locally in the medium while another is used to probe the time development of this depleted region. The fluorescence dip technique, finally, has been used to study laser-enhanced emission processes.[268]

5.6.2 Laser-enhanced Ionization in Plasma

Because ICP provides a rather different ionization environment than do flames, it is not clear, a priori, how excited atoms ionize in this type of atomizer. Laser power studies were therefore made on several elements (Cu, Na, Fe and Mn) to investigate the ionization mechanisms of the excited atoms.[323] However, the results were not fully conclusive. Both Fe and Mn, which were excited to states whose energy is slightly above half of the ionization energy, showed a linear laser power dependence. Assuming that the optical transition was saturated at the intensities used, this indicates that the major ionization route is photoionization for these elements. Copper was found to be excited in two-photon transitions to states close to the ionization limit from which ionization is assumed to proceed, with no further involvement of any light-induced process. For Na (excited to the 3p state), collisional ionization was the most probable ionization mechanism.

Laser-induced ionization in an ICP has also been performed by Turk et al. in order to measure the ionization yield and the fate of the laser-produced Sr ions in the ICP. Neutral Sr atoms were excited and photoionized by one laser system, while the number of Sr ions produced was monitored by another laser system (using the LIF technique).[327] By inducing a delay between the two ionizing and probing laser systems, it was found, among other things, that the recombination time constant of Sr ions in the ICP was 15.5 μs (but this decreased with increasing numbers of EIEs).

In another study concerned with the fate of ions in an ICP, Yu et al. found that the laser-induced enhancement of Sr^+ ions in a modified inductively coupled plasma mass spectrometry (ICPMS) instrument lasted for approximately 0.25 ms and peaked 0.2 ms after the laser pulse at a value of 11% relative to the Sr^+ ion signal due to the plasma ionization alone. The authors found that the laser-induced signal was limited by the high degree of direct plasma ionization and ion–electron recombination of laser-produced ions.[325]

With the use of a simplified model of the magnesium atom, several possible approaches to the measurement of charge exchange rates between Mg and Ar in the ICP have been developed.[414] Two of the approaches, which used pulsed dye lasers to populate the Mg(II) levels that are close to resonance for charge exchange, were tested experimentally. The experiments yielded an effective rate constant for the transfer of charge from Mg^{2+} to Ar of $1.1 \times 10^8\ s^{-1}$. This value is a factor of 50 lower than previously published estimates.[414]

A commercial ICPMS instrument, modified for use with a flame rather than an ICP, was used to sample and detect the LEI ions. Using double-resonance LEI with pulsed dye lasers, an Na^+ signal that was 350 times larger than that induced by thermal ionization could be achieved. Using a 5-mm laser beam diameter, the LEI signal ion pulse was found to last for 0.53 ms. Spatial studies, in which the position of the laser beam relative to the mass spectrometer sampler cone was varied, demonstrated that the ions produced by LEI travel with the flame velocity into the mass spectrometer, with no significant losses due to recombination from as far as 13 mm from the interface.[424]

5.7 Conclusions and Outlook for Laser-enhanced Ionization

It is clear that LEI is a highly sensitive laser-based spectroscopic technique that benefits from using a simple flame as atomizer (which allows for an uncomplicated sample introduction). LOD in the picogram per milliliter range have been obtained for a large number of elements (and with a few LOD even below this range). The most significant drawback of the technique is its sensitivity to EIEs. When samples with considerable numbers of EIEs are analyzed, significant background effects that limit the detection power of the analyte appear. This has limited the use of LEI primarily to situations when samples with uncomplicated matrices are to be analyzed. The technique has a great potential for diagnostics (e.g. of atomizer-specific physical and chemical processes) as well as for basic studies.

5.8 Resonance Ionization Spectrometry

5.8.1 Historical Development of Resonance Ionization Spectrometry

RIS was historically developed as a means of probing excited (metastable)-state populations in He as part of a study of the intricate interplay between various types of energy transfer processes in noble gas systems in the middle of the 1970s.[19] At that time it was conceived that photoionization could provide a more unambiguous probing of the excited-state populations than emission because the emission processes from metastable states in noble gas atoms are indirect. It was soon found, however, that the real importance of the photoionization approach

was that it could provide a quantum-state selectivity if a tunable pulsed laser were used to first excite the metastable atom to another excited state, after which photoionization could proceed by absorption of a second photon from the same laser. The use of an intermediate state ensured that the ionization process could take place only if the laser wavelength was resonant with an existing transition in the atoms, which thus gave the process its name: RIS.[19]

5.8.2 *Early Acquaintances for Resonance Ionization Spectrometry with the Field of Analytical Applications – Single Atom Detection*

It was soon recognized that the RIS technique had an extraordinary potential for analytical applications by tuning the laser light to transitions originating from the highly populated ground state. It was even prophesied early on that SAD capabilities were within reach. The SAD capability of RIS was demonstrated by Hurst et al. in 1977 in a proportional counter detecting individual Cs atoms.[9,10] Since then the field developed rapidly, with the aim of meeting the various requirements of an analytical technique.[10,11,19,41,271–285] The RIS technique was first applied to species like alkali atoms, whose ionization limit was so low that the RIS process was feasible with the laser systems available at that time. As tunable laser systems became more advanced and more prevalent, however, RIS was also applied to other types of species.[9–11,19,41,271–285]

5.8.3 *Estimated Analytical Properties of Resonance Ionization Spectrometry*

The greatest advantage of RIS is undoubtedly the unmatched combination of high sensitivity (a high ionization yield for the laser-illuminated atoms) and high selectivity. It has been found, for example, that almost 100% of the illuminated atoms can be ionized. The selectivity (calculated as a product of the estimated selectivity of each excitation step) was estimated to be of the order of $10^9 – 10^{12}$ under quite general conditions (e.g. from RIS using two resonant excitation steps followed by photoionization from the uppermost laser-connected level), and even higher under most favorable conditions. In addition, if the laser-produced ions were detected by a mass spectrometer (most often a time-of-flight mass spectrometer for pulsed excitations and a quadrupole mass spectrometer for continuous-wave excitations), the selectivity could be even larger. The technique is then often referred to as resonance ionization mass spectrometry (RIMS). This high selectivity gives the resonance ionization spectrometry/resonance ionization mass spectrometry (RIS/RIMS) technique good applicability to isotopic selective assessments, an application area in which most other optical techniques are not as amenable owing to the increased transition widths in finite-pressure atomizers (collisionally broadened). This extraordinary selectivity has truly given the RIS/RIMS technique extraordinary promise for analytical applications.

5.8.4 *Practical Limitations of Resonance Ionization Spectrometry/Resonance Ionization Mass Spectrometry*

Reality showed that not all of the expected properties of RIS/RIMS could be realized readily as first anticipated. The main limitation of the technique is a poor temporal and/or spatial overlap between a thermally atomized sample and the interaction volume. This is particularly evident when pulsed laser sources are used for excitation (which was the prevalent situation for the first years, mainly dictated by the availability of UV laser light) together with thermal atomization. For example, the spatial and temporal overlap of a thermal atomic beam and the light from a pulsed laser system is of the order of 10^{-5} using a 25-Hz repetition rate laser (and typical values for other parameters). Means to improve on this consist mainly of using a high-duty-cycle laser system (i.e. atomic-vapor-pumped dye laser systems[425] or continuous-wave lasers, either ion-laser-pumped dye laser systems[426] or, lately, diode laser systems[427]) or pulsed atomization sources (e.g. laser ablation[428,429] and sputter-initiated atomization[430]).[431] The use of a 6.5-kHz atomic-vapor-pumped dye laser system instead of a conventional pulse laser with a 25-Hz repetition rate, for example, increases the spatial and temporal overlap by two or three orders of magnitude to a few tenths of a percent.[432]

5.8.5 *Practical Applicability of Resonance Ionization Spectrometry/Resonance Ionization Mass Spectrometry*

An assessment of the analytical capabilities of the RIS/RIMS technique can most easily be performed by studying the practical applicability of the technique. The RIS/RIMS technique has been used for a variety of applications in both the basic and applied sciences throughout the years.

5.8.5.1 Basic Studies with Resonance Ionization Spectrometry/Resonance Ionization Mass Spectrometry The basic studies have often been concerned with measurements of atomic physics entities, e.g. hyperfine structures,[433] ionization potentials,[434] photoionization and continuum structure[435,436] (including studies of negative ions[437,438]), and for high-precision measurements,[439] etc. but also for studies of basic quantum phenomena,[440] e.g. dynamic Stark shift[441,442] and population dynamics.[443]

5.8.5.2 Analytical Applications of Resonance Ionization Spectrometry/Resonance Ionization Mass Spectrometry
The analytical applications of RIS/RIMS were originally devoted to studies of atomic species but molecular species have been assessed also during latter years.[285] The most recent activity within the field of analytical atomic spectrometry has been on isotopic issues. For example, double-resonance RIMS using single-frequency continuous-wave dye lasers has demonstrated detection limits in the attogram range for ^{90}Sr[444] and an optical isotope selectivity above 10^9 in measurements of Pb.[445]

One particular problem of using RIS/RIMS for isotope determinations, however, is that various isotopes can have different sets of (i.e. different numbers of) hyperfine components of a given atomic level and thereby different degeneracy factors of the laser-connected states. Because the ionization rate depends on details of the level structure of the atoms under study, the ionization fraction will vary between various isotopes. This is often referred to as the odd–even effect because isotopes with an even mass number generally lack nuclear spin ($I = 0$) and thereby hyperfine structure, whereas isotopes with odd mass number have a finite nuclear spin (i.e. $I \neq 0$) and show hyperfine structure.[446] This effect is now understood (theoretical predictions have been verified by experiments[447]) but the effect has to be taken into account properly when analyzing experimental data. There are several ways to reduce the odd–even effect in practical RIS/RIMS measurements, e.g. reducing the resonance laser intensity well below saturation, increasing the ionization intensity to full saturation so that all isotopes are ionized, or working with depolarized laser light or short pulses.[448]

One of the most important uses of RIMS applies its ability to detect low concentrations of long-lived radioisotopes, prompted by the fact that radiometry faces various problems in the detection of some of the most important and hazardous radioisotopes. For example, for very long-lived α- and β-emitters (e.g. ^{239}Pu with $t_{1/2} = 2.41 \times 10^4$ years and ^{99g}Tc with $t_{1/2} = 2.1 \times 10^5$ years), the measuring time is very long using radiometric methods. In addition, detection limits are impaired by the background. Some isotopes cannot be distinguished by α-spectroscopy owing to the very similar energies of the α-lines (e.g. the pairs $^{239}Pu/^{240}Pu$ and $^{238}Pu/^{241}Am$). Furthermore, because the continuous energy spectra of the β-emitters ^{89}Sr and ^{90}Sr overlap, they cannot be measured separately by β-spectroscopy. ^{90}Sr ($t_{1/2} = 28.5$ years) therefore often has to be detected via the daughter isotope ^{90}Y, which includes a waiting time of about 10–14 days.[432] This situation has justified the RIS/RIMS technique to seek applications in the detection of long-lived radioisotopes. Erdmann et al., for example, demonstrated detection of Pu with LOD of only 10^6 atoms using RIMS.[448] Isotopic Pu compositions could be determined with sufficient accuracy to assess the isotopic composition of Pu in soil from the Chernobyl area.[448] Passler used RIMS to detect long-lived radioisotopes like Pu, Tc, and $^{89,90}Sr$.[432] The LOD of ^{89}Sr and ^{90}Sr were 5×10^7 and 3×10^6 atoms, respectively.[432]

RIS/RIMS has also been used for precise determination of Ca isotope ratios, primarily ^{41}Ca,[427] for purposes of geological and anthropological dating. Other uses of techniques capable of a precise assessment of widely different Ca isotope ratios include the evaluation of exposure histories of extraterrestrial materials, measurements of ratios of the minor stable isotopes in meteorite inclusions for tests of nucleosynthetic models, and investigations of isotope traces in medical studies.[427] The analytical requirements for these applications vary widely, but for dating measurements with ^{41}Ca they include, among other things, an isotopic selectivity of 10^{15}.[427] In order to accomplish this, and therefore with the objective of finding the most optimal excitation schemes for precise determination of Ca isotope ratios, a series of studies concerned with isotope shifts and hyperfine structure,[449] as well as line shapes and optical selectivity issues in Ca,[450] have recently been performed.

Although many laser sources produce light with an extremely narrow bandwidth, the combination of small optical isotope shifts and finite broadening mechanisms (Doppler shifts and natural broadening) restricts the isotopic selectivity of the RIS/RIMS technique in certain applications. As a means to improve on this, collinear RIS/RIMS on accelerated atoms has been performed. The use of accelerated atoms eliminates Doppler broadening and enlarges the isotope shifts, whereas the collinear approach increases the interaction volume. After a long series of methodology improvements, LOD of 3×10^6 atoms of ^{90}Sr (corresponding to an activity of 2 mBq) in the presence of up to 10^{17} atoms of stable ^{88}Sr could be achieved.[426]

The combination of pulsed atomization and RIS/RIMS not only provides a higher temporal and spatial overlap between atomized sample and the interaction volume, but also provides the possibility to perform microanalysis on solid samples. RIS/RIMS has been used successfully for selective post-ionization of sputtered neutral species, proving powerful ultratrace analysis capabilities below atomic fractions of 10^{-9} by removing only a few monolayers of the substrate.[451] One such application is the detection of Al in brain tissue homogenates with a spatial resolution of 100 µm.[452] When RIMS is combined with techniques for pulsed atomization, however, it becomes closely related to many other analytical techniques and thus under these situations has

appeared under a large variety of names, e.g. sputter-initiated RIS or laser ablation RIS.

In addition to the few examples of the applicability of the RIS/RIMS technique to analytical atomic spectrometry given above, there are also a number of other applications of the RIS/RIMS technique that prove its applicability to ultrasensitive trace element or isotope analysis. The latest conference proceedings of the series of ''International Symposia on Resonance Ionization Spectroscopy and its Applications'' provide an unusually good coverage of the status of the field, and the reader is referred to these for a more in-depth description of the RIS/RIMS technique.[281–285]

5.9 Conclusions and Outlook for Resonance Ionization Spectrometry/Resonance Ionization Mass Spectrometry

Because RIS/RIMS normally is performed under vacuum conditions (thus with a minimum of broadening mechanisms), it can make full use of the narrow frequency width of laser light. By the use of double (or sometime even triple) resonant excitations followed by detection of the produced ions using mass spectrometry, it can provide impressive selectivity. This huge selectivity also makes isotopic determinations possible, even in cases when one isotope has an abundance that exceeds another by nine or ten orders of magnitude. One drawback is that the RIS/RIMS technique has difficulties in providing a high temporal and spatial overlap between a thermally atomized sample and the interaction volume, which can put practical restrictions on the sensitivity of the technique. Combined with pulsed atomization (e.g. laser ablation or ion sputtering) it can, however, be used for a variety of applications (e.g. surface analysis) with high sensitivity. Another drawback is that the RIS/RIMS technique requires rather complex instrumentation. The current development of continuous-wave diode lasers might be a useful solution to this problem. On the other hand, the technique has an excellent potential for studies of basic phenomena, e.g. within the fields of quantum optics or atomic physics.

6 FREQUENCY- AND WAVELENGTH-MODULATION DIODE LASER ABSORPTION SPECTROMETRY

6.1 Introduction to Diode Lasers and Their Use for Analytical Applications

6.1.1 Historical Development of Diode Lasers

Although the diode laser was invented before the dye laser,[453–456] for a long period of time it was not able to compete successfully with the dye laser for spectroscopic applications. The reason was the poor performance of the early diode lasers, which made them unattractive in comparison with dye lasers. They needed to be operated at liquid nitrogen temperatures, the spatial profile of their emission was not stable, their spectral emission was spread over many cavity modes, the output intensity was not always linear with injection current, and their lifetimes were not very impressive.[457] The development of diode lasers during the 1970s and 1980s has implied, however, that most of the problems associated with early diode lasers have been overcome. Today, there are handy single-mode room-temperature diode lasers with a number of unique and attractive properties, of which rapid tunability and low noise are two of the most important.

6.1.2 Spectroscopic Applications of Diode Lasers

Diode lasers have found use mainly for fundamental studies – atomic and molecular spectroscopy;[352,458–466] environmental monitoring, particularly remote sensing;[467–475] and analytical applications[43,45,225,476–480] – but also for a variety of other purposes (e.g. studies of effects of collisions[462,463] and evaluation of absolute number densities of atoms in atomic reservoirs[481]). Although the general properties of diode lasers have been reviewed repeatedly,[457,482–489] the spectroscopic properties of commercial laser diodes firing in the visible and near-infrared regions have been specially addressed by a handful of authors,[45,391,457,459,474,486,490–493] of which the reviews by Franzke et al.[45] and Fox et al.[391] are of particular interest for spectrochemical analysis. Pioneering work towards bringing diode lasers into the field of spectrochemistry has been done primarily and extensively (but not exclusively) by the groups of Niemax et al.[43,225,352,359,460,477–479,494–503] and Winefordner et al.[504–507]

Despite a number of advantages of diode lasers, a crucial drawback is the lack of lasers that can operate directly at visible or UV wavelengths. The number of elements today that can be studied efficiently by diode laser spectroscopy is therefore somewhat limited. Only a few elements can be detected using diode laser light directly and around 50 can be detected utilizing frequency-doubled diode laser light.[43] The atomic species that have been detected so far by diode laser spectroscopy in various atomizers (primarily flames or furnaces) for analytical purposes are Ba,[477,479,494] Br,[498] Ca,[479] Cl,[497,498] Cr,[508] Cs,[494,508] Eu,[479] K,[494] Li,[479,494,509] Pb,[479] Rb,[477,494,510–513] Sr[479,494] and Ti,[508] amongst which Li and Rb are the most frequently studied, primarily due to strong transitions from the ground state at wavelengths that can be produced rather easily by diode lasers. Issues that have been addressed are the multi-element capability

of diode laser spectrometry,[225,494] the isotope selectivity of the technique,[359,461,479,492,509] and the use of diode laser techniques for assessment of the elemental content in samples with complex matrices.[513] Also, the concept of frequency-doubling of diode laser light has received a certain amount of attention lately.[514–516]

A number of molecular species have been detected by diode laser light for remote-sensing applications.[467,468,473,517,518] A somewhat more limited number of molecules have been detected for analytical applications. They have been detected either directly (e.g. water vapor[519,520]) or indirectly (e.g. $C_2F_4Cl_2$, CCl_4, CHF_3 and O_2 by detection of Cl, F and O in a DC discharge[496]). A good review of the field of analytical molecular spectroscopy with diode lasers has been given by Imasaka.[480]

The detection capabilities for various species differ considerably, depending on the type of atomizer, the transition (i.e. population of the lower laser-connected level and transition probability) and the detection technique used. Diode lasers have been used for spectrochemical purposes with a variety of techniques, e.g. fluorescence techniques,[225,504,510] ionization techniques,[278,352,460–463,476,479,509,521] noise detection techniques[511] and, last but not least, absorption techniques.[477,478,481,494,496–498,505,508,512,522] Although very low detection limits can be obtained by applying diode lasers to the RIS/RIMS technique (which has to be performed under low-pressure conditions in rather complex instrumental set-ups),[11,19,274,278] it has been found that FM or WM techniques significantly improve the detectability of various species using absorption spectrometry under far more versatile conditions.[42,46] The following presentation of the use of diode lasers for analytical spectrometric applications will therefore be devoted to modulated (mostly WM) diode laser absorption techniques.

6.2 Frequency-modulation versus Wavelength-modulation

The general approach of modulation techniques comprises modulating the frequency (or the wavelength) of the laser light (most conveniently done by modulating the injection current to the laser diode), transmitting the light through the species cloud to be analyzed, and analyzing the detector signal at the modulation frequency (or at some frequency related to this). This process significantly reduces noise from the laser (primarily $1/f$ noise) by shifting the detection to higher frequencies where the noise is less significant. It also removes baseline offsets, slopes or even some curvature of the background (depending on the mode of operation).

Although being basically the same, the techniques utilizing a modulation of the frequency (or wavelength) are referred to as WM or FM techniques, depending on the relative sizes of the various modulation parameters and the means of detection.[42,46]

WM techniques are characterized by a modulation frequency, f_m, that is much *smaller* than the half-width at half-maximum of the absorption peak, $\Delta\nu_p$ (i.e. $f_m \ll \Delta\nu_p$).[42,523] The signal is detected at a suitable harmonic of the modulation frequency (most often at $2f_m$) using a lock-in amplifier. The optimum conditions for WM techniques include, among other things, a modulation amplitude, ν_a, that is a few times larger than the half-width at half-maximum of the absorption profile ($\nu_a = 2.2\Delta\nu_p$ for $2f$ detection, $\nu_a = 3.9\Delta\nu_p$ for $4f$ detection, etc.).[42] This implies, in turn, that the modulation frequency is smaller than the modulation amplitude (i.e. $f_m \ll \nu_a$).

When atmospheric pressure-broadened absorption profiles are detected (whose widths often are in the gigahertz range), the modulation amplitude therefore needs to be in the gigahertz range but the modulation frequency (often limited by the performance of the lock-in amplifier) is in the 10–100 kHz range. The use of modulation frequencies in the mid-kilohertz range implies that the influence of laser excess noise has been reduced significantly (although not necessarily fully eliminated).

FM techniques, on the other hand, pioneered by Bjorklund in 1980,[524] and investigated by numerous authors since then,[46,524–532] utilize, under optimum conditions, a modulation frequency that is *larger* than the width of the absorption peak (i.e. $f_m > \Delta\nu_p$). Such a high modulation frequency will impose sidebands (in frequency space) whose spacing is equal to the modulation frequency. By detuning the center frequency of the laser light by an amount equal to the sideband spacing, a sideband will be placed atop the absorption profile. One of the two sidebands will then be absorbed by the analyte. Detection is then achieved by feeding the signal from the detector to a mixer in which this signal is mixed with a reference signal (at frequency f_m), thus producing the sums and differences of these two signals (often referred to as heterodyning). The resulting signal, which is proportional to the difference in electrical field amplitude of the two sidebands, and therefore also the density of analyte species, can then often be detected directly by ordinary (fast) electronics. In order to optimize the signal strength, preferably only one sideband on each side of the carrier should be produced. This can be achieved if the modulation frequency also is *larger* than the FM amplitude (i.e. $f_m \gg \nu_a$).

This technique has found its best use for the detection of various molecular species under low-pressure conditions (often in multipass cells). Typical modulation frequencies for the FM techniques are in the high megahertz range

(or occasionally around a few gigahertz), which is larger than the typical (often Doppler-broadened) absorption profiles of molecules.[533] Because excess laser noise is negligible at such detection frequencies, it should, in principle, be possible to obtain shot-noise-limited detection limits with FM techniques.[528,529] FM techniques in most cases should therefore outperform WM techniques when it comes to detection power.[46] However, this has not always been the case in practice. The high modulation frequencies involved put large demands on the instrumentation and increase the complexity of the system considerably. The instrumentation used therefore does not always fulfill all simultaneous requirements for optimum detectability. This is particularly the case when atomic species are to be detected under atmospheric pressure conditions (in which absorption profiles can have widths in the 5–10 GHz range). The rather complex instrumentation is also the prime drawback to the widespread use of FM techniques. Another reason why FM techniques do not outperform WM techniques (despite a lower inherent noise) is the existence of background signals (which, in many practical cases, is the limiting factor for analytical applications). Hence, despite the inherently higher detectability of FM techniques, WM techniques are still often used for the detection of species for spectrochemical purposes[495–498,508,510,513,522,523,534–539] whereas FM techniques have been used almost exclusively for gas species analysis.[517,520,526–533,540] We will therefore only discuss the WM technique below. The reader is referred to the literature for a discussion of FM techniques.[42,43,520,524–533,540] Thorough investigations and comparisons of (theoretical as well as experimental) various FM and WM techniques have been done by Silver[42] as well as Bomse et al.[46]

6.3 Wavelength-modulation Spectrometry

6.3.1 Instrumentation

A typical experimental set-up for WM spectrometry is shown in Figure 20. Wavelength-modulated light from a diode laser is directed through an atomizer onto a detector. The detector output is fed to a lock-in amplifier (through a preamplifier when applicable). The output of the lock-in amplifier constitutes the final signal to be registered and stored by a computer.

There are a few different modes of detection for the WM techniques. The most common is $2f$ detection (i.e. detecting the absorption at twice the modulation frequency, $2f_m$),[43,495–499,508,510,513,522,531,534–538,541–544] although occasionally higher harmonics ($4f_m$ and $6f_m$) have been used.[467] Reasons for this are that: the $2f$ output is intuitively the harmonic that carries the main part of the analytical signal in WM spectrometry (see discussion below); it produces the largest WM signals; and it eliminates (or reduces significantly) both constant and linearly sloping baselines. Another attractive feature is that the signal peaks at the absorption line center.

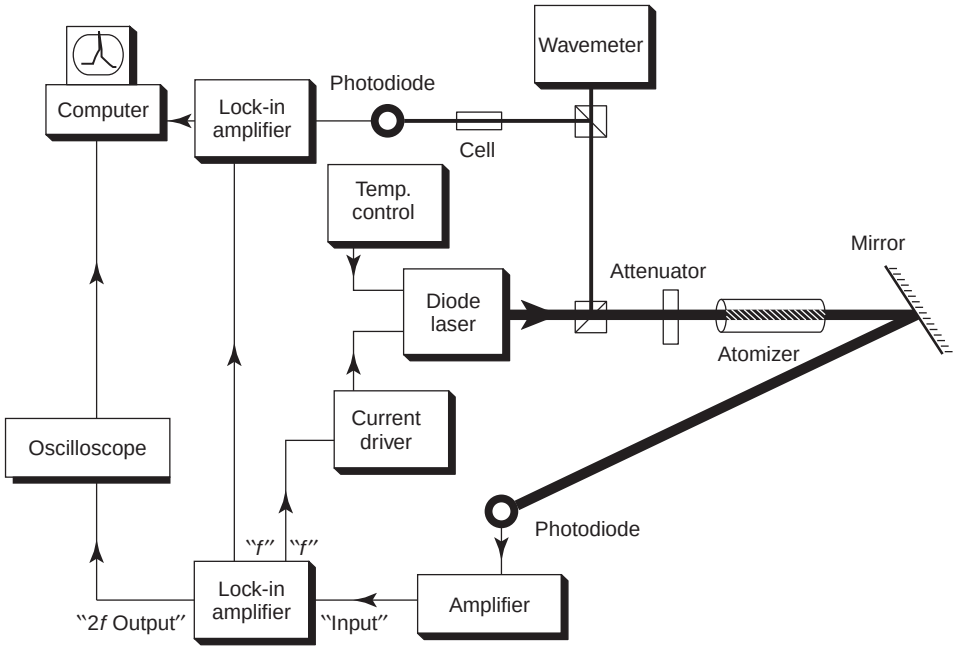

Figure 20 A typical set-up for WMDLAS spectrometry in a GF. (Reproduced from P. Ljung, O. Axner, *Spectrochimica Acta*, **52B**, 305–319, Copyright 1997, with permission from Elsevier Science.)

6.3.2 Theory

6.3.2.1 Absorption Process and Nomenclature For absorption measurements, the detector signal, $S_D(\nu, \nu_0)$, can be written as shown in Equation (63):

$$S_D(\nu, \nu_0) = \kappa I_0 e^{-\rho L \sigma(\nu - \nu_0)} \tag{63}$$

where ν is the frequency of the laser light, ν_0 is the center frequency of the absorption profile, κ is an instrumentation constant (including a photodiode response and an amplification function) that relates the light intensity impinging upon the photodiode to the measured signal, I_0 is the laser light intensity prior to the interaction region, ρ is the density of species to be detected (in m^{-3}), L is the interaction length, and σ is the absorption cross-section of the species to be detected at the particular wavelength (frequency) used. Because the absorption cross-section depends on the wavelength of the light (i.e. the frequency), it can be written conveniently as in Equation (64):

$$\sigma(\nu - \nu_0) = \sigma_0 \bar{\chi}(\nu - \nu_0) \tag{64}$$

where σ_0 represents the absorption cross-section at the peak of the profile, i.e. $\sigma(\nu = \nu_0)$, and $\bar{\chi}(\nu - \nu_0)$ is the (peak-normalized) line-shape function for the species to be detected.

6.3.2.2 Wavelength Modulation WM techniques build upon modulation of the laser light wavelength, λ, by a sinusoidal modulation frequency, f_m, of the injection current, as shown in Equation (65):

$$\lambda(t) = \lambda_c + \lambda_a \sin(2\pi f_m t) \tag{65}$$

where λ_c and λ_a are the center wavelength and amplitude of the WM, respectively. Because it is more convenient to work with broadening and distribution functions in frequency space, the modulation can be written in terms of frequency (i.e. in Hz), as in Equation (66):

$$\nu(t) = \nu_c + \nu_a \sin(2\pi f_m t) \tag{66}$$

where ν_c and ν_a are the corresponding entities in frequency space, i.e. the center frequency and the modulation amplitude, respectively.

6.3.2.3 The nf *Wavelength-modulation Signal* This implies that the detector signal has a rather complex time dependence. An intuitive understanding of the origin of the 2*f* WM signal in resonance can be obtained from Figure 21. Figure 21(a) displays: the detector signal as a function of frequency for an analyte species with its absorption profile centered around ν_0; the time evolution of the laser light frequency (modulated at a frequency f_m); and the time-dependent detector signal constructed from both of the former. The absorption profile has been turned through 90° to correlate the frequency axis with that of the laser light modulation. Because the center frequency of the laser light coincides with that of the analyte absorption profile (at λ_0), the detector signal will experience an absorption maximum twice per modulation period, as shown in Figure 21(a). This explains clearly why the detector output signal will contain a significant amount of $2f_m$ frequency components. Figure 21(b) shows the same entities for a linear background absorption. As can be seen, a linear (or constant) background absorption produces no $2f_m$ frequency components.

There is, however, no need to study the time dependence of the detector signal directly. A lock-in amplifier is normally used to extract the *n*th harmonic of this detector signal, $\overline{S}_{nf}(\nu_d, \nu_a)$, according to Equation (67):

$$\overline{S}_{nf}(\nu_d, \nu_a) = \frac{1}{\tau} \int_0^\tau S_D[\nu_c + \nu_a \sin(2\pi f_m t), \nu_0] \times \sin(2\pi n f_m t + \phi)\, dt \tag{67}$$

where ϕ is a suitable chosen phase and where we, for convenience, have introduced ν_d as the detuning of the modulation center frequency with respect to that of the absorption profile, $\nu_d = \nu_c - \nu_0$. This implies in practice that $\overline{S}_{nf}(\nu_d, \nu_a)$ is the *n*th Fourier component of the time-dependent detector signal, $S_D(\nu_c + \nu_a \sin(2\pi f_m t), \nu_0)$.[523]

Arndt[545] has shown that the signal strength of the *n*th harmonic can be expressed in terms of an integral as shown in Equation (68):

$$\overline{S}_{nf}(\nu_d, \nu_a) = \frac{(-1)^n (2 - \delta_{n0})}{\pi} \int_{-\nu_a}^{\nu_a} S(\nu_c - \nu', \nu_0) \times \frac{T_n(\nu'/\nu_a)}{\sqrt{(\nu_a)^2 - (\nu')^2}}\, d\nu' \tag{68}$$

where δ_{n0} is the Kronecker delta and T_n is the Chebyshev polynomial of degree *n*.

An *nf* WM spectrum of the species to be detected, $\overline{S}_{nf}(\nu_d)$, can be created by slowly scanning the center frequency across the absorption profile (i.e. the detuning) for a given modulation amplitude.

For the case when the medium is optically thin, the exponential factor in Equation (63) can be approximated with the first two terms in its series expansion; thus, the *nf* WM signal, $\overline{S}_{nf}(\nu_d, \nu_a)$, can be written in terms of the *n*th harmonic of the line-shape function, $\bar{\chi}_n(\nu_d, \nu_a)$, according to Equation (69):

$$\overline{S}_{nf}(\nu_d, \nu_a) = -\kappa I_0 (\rho L \sigma_0) \bar{\chi}_n(\nu_d, \nu_a) \tag{69}$$

The *nf* lock-in output signal depends not only on the detuning, the amplitude of the modulation, ν_a, and the width of the absorption profile, $\Delta\nu_p$, but also on the particular form of the absorption profile. The general

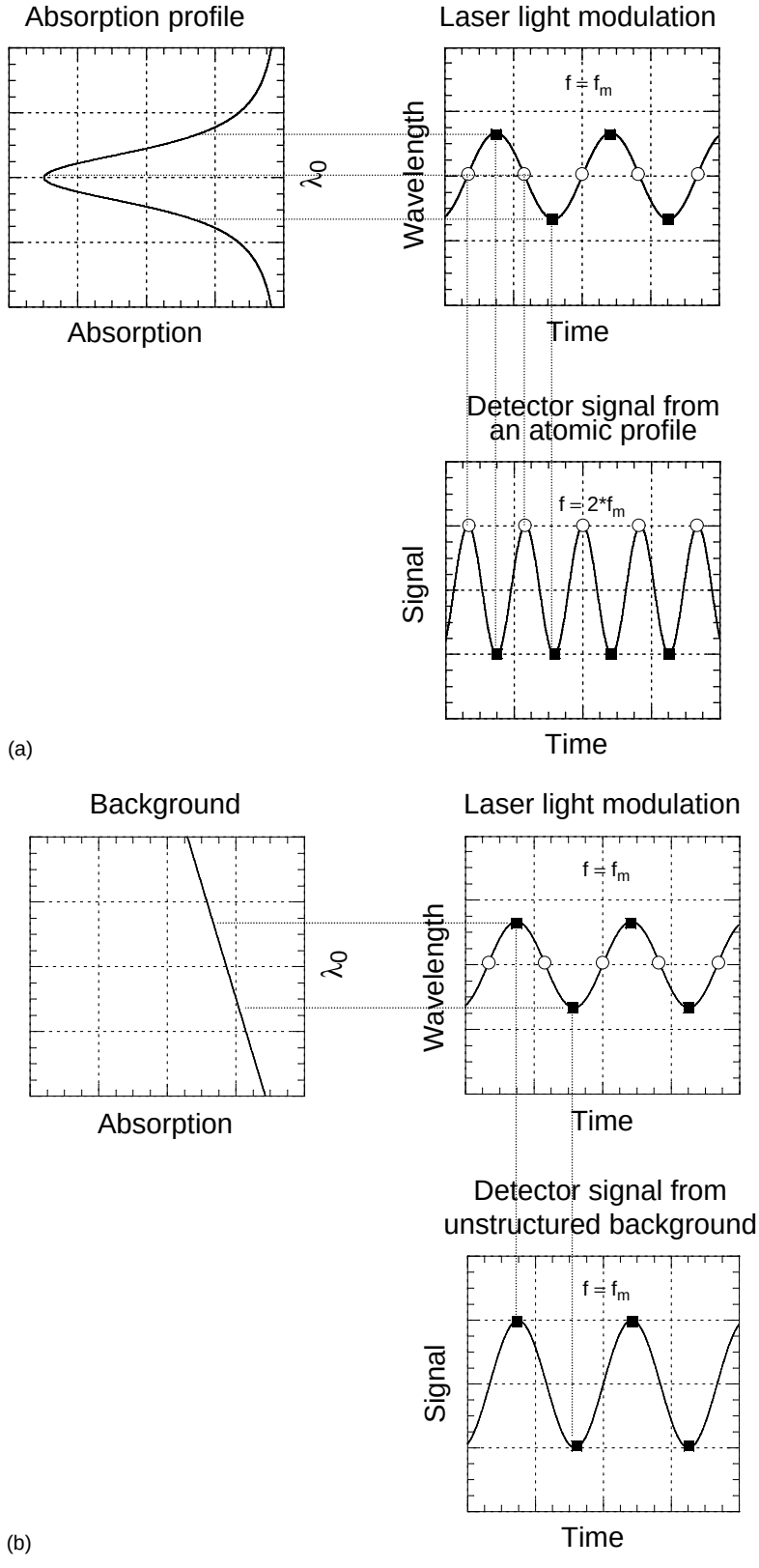

Figure 21 Absorptions, modulated laser light frequencies and resulting detector signals for the absorption profile (a) and for linear background absorption (b).

form of a single absorption transition has a Voigt form, originating from a convolution of a Lorentzian-broadened homogeneous profile (resulting from collision broadening) and a Gaussian-broadened inhomogeneous profile (resulting from Doppler broadening).

There are analytical expressions for the *nf* WM signal for only two limiting cases: when small modulation amplitudes are being used ($v_a \ll \Delta v_p$); and when homogeneous broadening dominates the inhomogeneous profite under optically thin conditions (so that the *nf* WM signal is directly proportional to the *n*th harmonics of a Lorentzian function). Equation (67) or (68) therefore has to be evaluated by numerical means for all other conditions (i.e. either by numerical integration[546] or by simulations[535–539]).

For sufficiently small modulation amplitudes (i.e. significantly smaller than the width of the absorption profile), the *nf* WM signal is directly proportional to the *n*th derivative of the line-shape function, as shown in Equation (70):

$$\lim \bar{S}_{nf}(v_c)|_{v_a \to 0} = -\kappa I_0(\rho L \sigma_0) \frac{(v_a)^n}{n! 2^{n-1}} \frac{\partial^n \bar{\chi}(v)}{\partial v^n}_{v=v_c} \quad (70)$$

This property has given the WM technique its alternative name of *derivative spectroscopy*. This description describes the shape of the WM spectra for small modulation amplitudes and also shows that the *nf* signal strength rapidly decreases with decreasing modulation amplitude (i.e. as $(v_a)^n$). This implies that it is not particularly applicable to the conditions under which the WM technique is used for analytical applications (which is to use a modulation amplitude that is a few times larger than the width of the absorption profile).

It is also possible to give an analytical description of the WM signal when the absorption profile is purely Lorentzian. In the description of the WM technique it has been found convenient to work with all frequency entities normalized with respect to the half-width at half-maximum of the atomic profile (i.e. $\bar{v}_d = v_d/\Delta v_p$ and $\bar{v}_a = v_a/\Delta v_p$, the latter often referred to as the modulation index, m). An analytical expression for the various *n*th harmonics of a peak-normalized Lorentzian-shaped line-shape function (expressed in normalized units) is given in Equation (71):

$$\bar{\chi}(t) = \frac{1}{1 + [\bar{v}_d + \bar{v}_a \sin(2\pi f_m t)]^2} \quad (71)$$

which is valid for arbitrary modulation amplitudes and has been derived previously by Arndt,[545] as shown in Equation (72):

$$\bar{\chi}_n(\bar{v}_d, \bar{v}_a) = \frac{2 - \delta_{n0}}{2} \frac{\left\{ \begin{matrix} \sqrt{[(1 - i\bar{v}_d)^2 + \bar{v}_a^2]} \\ -(1 - i\bar{v}_d) \end{matrix} \right\}^n}{\bar{v}_a^n \sqrt{[(1 - i\bar{v}_d)^2 + \bar{v}_a^2]}} i^n + c.c. \quad (72)$$

where i is the imaginary unit and *c.c.* represents the complex conjugate.

Although the above expression is exact, it is not always suitable for direct use because it is cast in terms of complex numbers. Kluczynski and Axner showed recently that this expression can be written more conveniently as shown in Equation (73):

$$\bar{\chi}_n(\bar{v}_d, \bar{v}_a) = \frac{A_n}{(\bar{v}_a)^n} \quad B_n + \frac{C_n S_+ + D_n S_-}{\sqrt{2}R} \quad (73)$$

where $S_+ = \sqrt{R + M}, S_- = \sqrt{R - M}, R = \sqrt{M^2 + 4(\bar{v}_d)^2}$, $M = 1 + (\bar{v}_a)^2 - (\bar{v}_d)^2$, and the four parameters A_n–D_n take different values (in terms of $\bar{v}_a$ and $\bar{v}_d$) depending on the order of the harmonics.[523] The components for the first five harmonics are given in Table 6.

Typical $1f$, $2f$, $4f$, and $6f$ WM spectra from a single-peak (Lorentzian-broadened) transition are exemplified in Figure 22. Each spectrum is evaluated for the modulation amplitude that maximizes that particular *nf* WM component (2.0, 2.2, 3.9 and 7.4, respectively; see section 6.3.3 below). As can be seen from the graph, the even harmonics have their maximum value at the position of the peak of the absorption profile, whereas the odd harmonics have it elsewhere, and the $2f$ signal gives rise to the largest signal of all even components.

Table 6 The A_n–D_n coefficients for the five lowest harmonics (i.e. $n = 0$–4) of the Fourier components of a Lorentzian absorption profile (Equation 73)

	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
A_n	1	−2	−2	2	2
B_n	0	0	−2	$-8\bar{v}_d$	$-4(1 + M - 5\bar{v}_d^2)$
C_n	1	$\bar{v}_d$	$1 + M - \bar{v}_d^2$	$[9 + 3M - \bar{v}_d^2]\bar{v}_d$	$(1 + 6M + M^2) - 2(17 + 3M)\bar{v}_d^2 + \bar{v}_d^4$
D_n	0	$\frac{-\bar{v}_d}{\lvert\bar{v}_d\rvert}$	$4\lvert\bar{v}_d\rvert$	$\frac{-(1 + 3M - 9\bar{v}_d^2)\bar{v}_d}{\lvert\bar{v}_d\rvert}$	$16(1 + M - \bar{v}_d^2)\lvert\bar{v}_d\rvert$

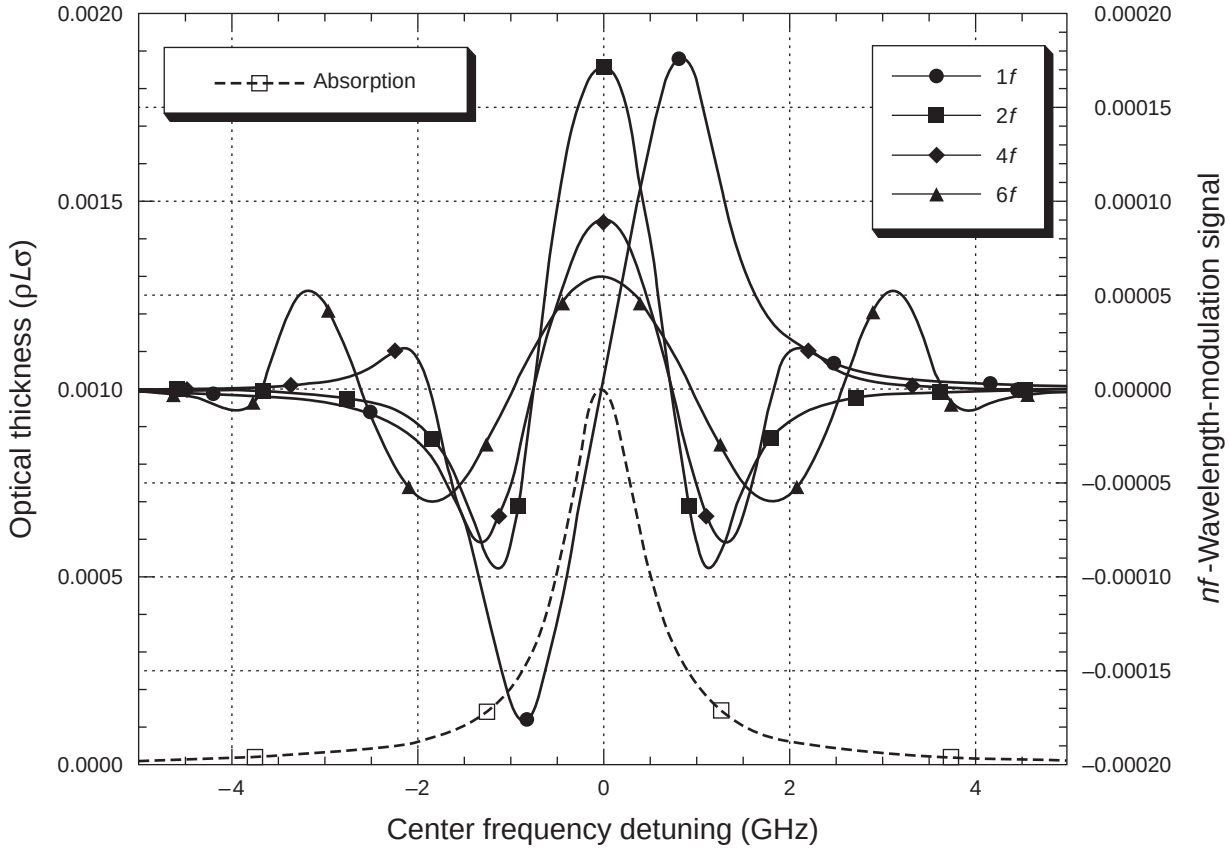

Figure 22 Schematic representation of a Lorentzian-shaped absorption profile (dashed line, left axis) and the corresponding $1f$, $2f$, $4f$ and $6f$ WM spectra (solid lines, right axis). The half-width at half-maximum of the absorption profile ($\Delta\nu_p$) has been taken as 0.5 GHz, for simplicity. The four WM spectra are evaluated for the FM amplitudes that maximize each nf-WM curve, which are 2.0, 2.2, 3.9, and 7.4 times $\Delta\nu_p$ for the four WM curves, respectively. The density of absorbers has been chosen to create an optical thickness of 10^{-3} at resonance.

6.3.3 Maximum Signal Conditions

The conditions for maximization of the WM signal have been scrutinized by a number of authors. Cassidy and Reid calculated relative peak heights for the 2nd, 4th, 6th and 8th harmonics of a pure Lorentzian profile.[467] Wilson used numerical integration to obtain the three first harmonics of both Gaussian and Lorentzian profiles.[546] Silver determined the optimum values of the modulation index (m) for both Gaussian and Lorentzian profiles for the 1st, 2nd, 4th and 6th harmonics.[42] He found the optimum modulation indices to be: 1.6 and 2.0; 2.1 and 2.2; 3.6 and 3.9; and 5.2 and 7.4, respectively. Rojas et al.[535] and Gustafsson et al.[536–539] investigated three different issues of the $2f$ WM technique: when atoms are detected under not necessarily optically thin conditions;[535,538,539] for atoms with complicated (multi-line) atomic structure (a special study was performed of the 780-nm transition in Rb, which constitutes two isotopes each with six individual close-lying transitions);[535–539] and the extent to which a low-pressure cell can be used as a wavelength reference for $2f$ WM experiments made under conditions of atmospheric pressure and high temperature (i.e. a GF).[536–538] The same authors also investigated the temperature dependence of the $2f$ WM technique.[539]

6.3.4 Signal-to-noise Ratios

The most extensive analysis of the S/N conditions for diode-laser-based modulation techniques has been given by Silver[42] and Bomse et al.[46] They discussed the concept of S/N for both WM and FM techniques and their conclusions are based upon experiments made with a liquid-nitrogen-cooled infrared lead–salt diode laser, working in the 8-μm wavelength range, used for detection of N_2O in a multipass cell. Although the conditions are not typical of those when diode-laser-based modulation techniques are used for analytical atomic spectrometry, some of their conclusions are of importance.

They defined an intrinsic S/N (power) ratio as the ratio of the square of the time-averaged analytical signal and the sum of the squares of the noise contribution from internal sources of noise, i.e. laser-induced detector

shot-noise, detector–preamplifier thermal noise, residual amplitude modulation (RAM)-induced noise, and laser excess noise. They analyzed their results, however, in terms of extrinsic sources of noise, in this case that originating from etalon background signals (see section 6.3.5).

In the absence of noise from RAM and background signals, the authors concluded that under their experimental conditions (a laser with 11 μW of output power) the WM signals ought to be dominated by laser excess noise at low modulation frequencies and by detector noise at higher frequencies. Shot-noise-dominated situations should, under optimum conditions (i.e. in the absence of RAM and background signals), be obtained for a combination of higher laser powers (in their case above 40 or 560 μW, depending on the detector used) and high modulation frequencies (in the megahertz range). At low or medium modulation frequencies the detectability of WM techniques should therefore (again under optimum conditions) be limited by laser excess noise (which is the dominating noise in the kilohertz range).[42]

6.3.5 Background Signals

In reality, there are background signals that can compete with the analytical signal when small enough quantities/concentrations of analyte are to be detected.

RAM (originating from an associated modulation of the intensity with the laser diode injection current) is a type of background signal frequently referred to. Although predicted by the theory presented by Silver,[42] and used by a number of authors since then for describing their background signals, it was first shown by Zhu and Cassidy[547] and then discussed further by Kluczynski and Axner[523] that there is no nf RAM signal for $n \geq 2$ (including $2f$ RAM) as long as the modulation of the intensity is linear with that of the injection current. The incorrect prediction of a $2f$ RAM signal from a linear intensity modulation originates from an uncritical application of a theoretical formalism developed for FM spectrometry (in which the intensity modulation is described as a modulation of the electrical field strength of the laser light instead of the intensity) to the WM technique.

The dominating cause of background signals is instead multiple reflections between various optical surfaces (which thus act as etalons). Because the transmission through an etalon is strongly wavelength dependent, the detector will experience a wavelength-dependent modulation of the intensity that will contain the same higher harmonics as the analytical signal.[523] The drift and fluctuations of these background signals are often the limiting factors to the WM technique. Therefore, unless any special concerns are taken, the detection power of modulation techniques is often limited by the noise and drift in the background signal originating from multiple reflections in the system rather than any of the intrinsic sources of noise, giving rise to detection limits (fractional absorbencies) in the 10^{-5} range.[43,46,467,495,498,513,517,530,544,548]

6.3.6 Means to Reach Optimum Conditions

Remedies for reducing etalon effects comprise: eliminating the use of optical surfaces that are perpendicular to the direction of the laser beam; antireflection coating; replacing parallel plates with wedges; and modulating either the position of the surfaces[499,520,549–551] or the atomization at a frequency different from that modulating the laser wavelength.[498,550]

6.3.6.1 Dithering of Optical Surfaces Bomse et al.[46] concluded that with piezocrystal dithering of appropriate optical components the optimum conditions for $1f$ and $2f$ WM spectrometry comprise modulation of the frequency at a rate between 100 kHz (which were sufficient for bringing the $1/f$ noise down to levels below that of the background signals in their experiment) and 20 MHz (at which instrumentation difficulties arose). They also concluded that neither of the FM techniques investigated could outperform the high-frequency WM techniques (although their theoretical treatment predicted that the one-tone FM technique should provide an almost fivefold improvement over high-frequency WM spectrometry). They attributed this to instrumental deficiencies (low detector sensitivity and inefficient radiofrequency coupling of the modulation current to the laser). The authors concluded finally that, owing to the high sensitivity of the etalon background signals to small changes in optical alignment, the reproducibility of the final detection power of the WM techniques ($<10^{-6}$) depended strongly on the details and stability of the optical alignment and variations in the effectiveness of the etalon reduction scheme.

It is important to point out, however, that dithering the position of various surfaces has the drawback that it can only eliminate background signals from etalons created *between* two different optical components. Any possible etalon created *inside* a component (i.e. between the front and back surface of the same optical component) cannot be eliminated by that technique.[523] Such etalons can therefore set the ultimate level for the background signal in conventional WM spectrometry and thereby also the S/N level.[523]

6.3.6.2 Modulation of the Atomization An efficient way around these residual problems is to modulate also the atomization process and detect the signal at a combination of the ordinary nf frequency and that of the atomization. Although modulation of an optical surface is rather straightforward to achieve, modulation

of the atomization process is more difficult in thermal atomizers such as flames and furnaces, although it can be used with advantage using various types of plasmas. The most impressive demonstration of a double-modulation technique that has been performed so far is the detection of Cl atoms in a modulated low-pressure microwave-induced plasma by Liger et al.[499] The authors showed that spurious etalon effects, background absorption, residual diode-laser-amplitude modulation and the noise that accompanies these entities could be suppressed by the use of a double-modulation technique in which both the diode laser wavelength and the plasma were modulated (although at different frequencies). This double-modulation technique enabled a detection limit of 1×10^{-6} AU to be obtained for Cl atoms in a modulated low-pressure microwave-induced plasma (with a time constant of 1 s).

6.3.6.3 Double-modulation, Dual-beam, and Logarithmic Detection In order to eliminate laser excess noise and signal variations due to changes/drifts of the optical transmittance, a double-beam arrangement with logarithmic subtraction of sample and reference detector currents was developed by the same set of authors.[499] As is exemplified in Figure 23, the logarithmic amplification not only makes the system independent of variations in the radiation intensity but it also enables suppression of the noise by decoupling the modulations of the intensity and wavelength. A detection limit of 2×10^{-7} AU limited by shot-noise could finally be established. The authors conclude finally that preliminary experiments showed that shot-noise was the dominating source of noise up to time constants of about 1000 s in their experiments and that as a consequence it should be possible to reach detection limits as low as 6×10^{-9} AU. If such detection powers can be realized, WM techniques would surely find a much broader applicability in spectrochemical analysis than they have today.

6.4 Conclusions and Outlook

Frequency- and wavelength-modulation diode laser absorption spectrometry (FM/WMDLAS) has the great potential to outperform conventional AAS in the future. There is no doubt that the technique has better sensitivity, LDR and selectivity than the AAS technique. The most significant drawback of FM/WMDLAS today is the lack of diode lasers producing light in the visible and UV regions.

7 CONCLUSIONS

It is clear that laser light, with its narrow bandwidth and high spectral irradiance, is ideal for selective

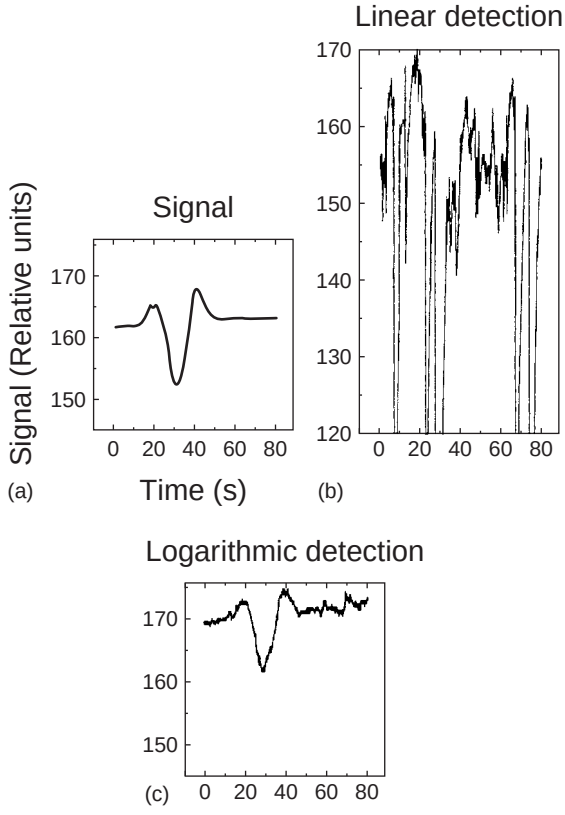

Figure 23 The 2*f* WM spectra from the 833-nm line in Cl measured in a directly coupled plasma: (a) single-beam absorption and linear amplification under optimum conditions; (b) the same instrumentation as in (a) but with turbulent smoke in the absorption path; (c) double-beam absorption with logarithmic conversion measured simultaneously with the curve displayed in (b). (Reproduced from V. Liger, A. Zybin, Y. Kuritsyn, K. Niemax, *Spectrochimica Acta*, **52B**, 1125–1138, Copyright 1997, with permission from Elsevier Science.)

interactions with atomic and molecular species. Extraordinary results have been achieved in a number of application areas,[47–52] of which the cooling and trapping of single atoms or ions and the creation of Bose–Einstein condensates[552] in the fields of atomic and molecular spectroscopy and quantum optics are just a few impressive examples. Despite many impressive demonstrations in the field of analytical spectrometry,[21,48] laser-based spectrometric techniques so far have been techniques only for the physical or chemical laboratories around the world. The reasons for this are that: the production of laser light has not been as carefree as first anticipated; the cost of tunable laser systems has been too high; and laser-based techniques are basically single-element detection techniques.

With the advent of single-mode, tunable diode lasers (and maybe also new OPO-based systems), however, and the new application areas that these handy types of lasers now start to pervade, the prospects for a more regular use of laser-based spectroscopic techniques

for analytical spectrometry seems to be better than ever before. If such techniques can be developed and commercialized in a proper manner in the future, then the regular analytical laboratory will be able to benefit from the extraordinary properties of laser-based spectroscopic techniques regarding sensitivity and selectivity (including various types of powerful means of background correction).

Even if laser-based spectroscopic techniques do not find their way onto the commercial market as regular techniques for trace species determinations, they will always have their role in certain types of applications. The nonintrusiveness of the techniques, for example, makes them attractive in certain situations, e.g. when samples with long-lived but toxic radioactive species are to be analyzed. In addition, as has been demonstrated repeatedly, their capabilities as diagnostic tools are indeed unmatched. Their applicability for remote-sensing applications is also unparalleled.

Looking back to what the joint community of laser spectroscopists has achieved during the last 20 years, it is difficult not to be impressed. Even in such a perspective, it is more difficult than ever to try to prophesy what the future will bring in this area: where will this field of science stand 20 years from now? Can we then, maybe, really count the atoms in an unknown sample?

ACKNOWLEDGMENTS

The author is indebted to all his colleagues in the Laser Physics Group at the Department of Experimental Physics at Umeå University for their continuous support during the writing of this review article. The author would also like to thank the Swedish Natural Science Research Council, the Swedish Research Council for Engineering Sciences (under project nos 251-97-733 and 288-98-40) and the Faculty of Science and Technology at Umeå University for various types of economic support, without which this review would never have become a reality.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
ASE	Amplified Spontaneous Emission
CCD	Charge-coupled Device
DC	Direct Current
EIE	Easily Ionized Element
ETA	Electrothermal Atomizer
FI/LEI	Flow Injection/Laser-enhanced Ionization
FM	Frequency-modulation
FM/WMDLAS	Frequency- and Wavelength-modulation Diode Laser Absorption Spectrometry
GF	Graphite Furnace
HPLC	High-performance Liquid Chromatography
ICCD	Intensified Charge-coupled Device
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IUPAC	International Union of Pure and Applied Chemistry
LDR	Linear Dynamic Range
LEAFS	Laser-excited Atomic Fluorescence Spectrometry
LEI	Laser-enhanced Ionization
LEMOFS	Laser-excited Molecular Fluorescence Spectrometry
LIF	Laser-induced Fluorescence
LIF/LEAFS	Laser-induced Fluorescence/ Laser-excited Atomic Fluorescence Spectrometry
LOD	Limits of Detection
NBS	National Bureau of Standards
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
NIST	National Institute of Standards and Technology
OPO	Optical Parametric Oscillator
PMT	Photomultiplier Tube
RAM	Residual Amplitude Modulation
RC	Resistance–Capacitance
RIMS	Resonance Ionization Mass Spectrometry
RIS	Resonance Ionization Spectrometry
RIS/RIMS	Resonance Ionization Spectrometry/Resonance Ionization Mass Spectrometry
RSD	Relative Standard Deviation
SAD	Single Atom Detection
S/N	Signal-to-noise
SONRES	Saturated Optical Nonresonant-emission Spectrometry
UV	Ultraviolet
WM	Wavelength-modulation
WM-DLAS	Wavelength-modulation Diode Laser Absorption Spectrometry
ZETA/LEAFS	Zeeman Electrothermal Atomization/Laser-excited Atomic Fluorescence Spectrometry

RELATED ARTICLES

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Automotive Emissions Analysis with Spectroscopic Techniques • Differential Optical Absorption Spectroscopy, Air Monitoring by • Diode Laser Spectroscopic Monitoring of Trace Gases • Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis • Infrared LIDAR Applications in Atmospheric Monitoring • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Laser-induced Breakdown Spectroscopy, Elemental Analysis • Laser Mass Spectrometry in Trace Analysis • Laser-based Combustion Diagnostics • Laser- and Optical-based Techniques for the Detection of Explosives • Optical Gas Sensors in Analytical Chemistry: Applications, Trends and General Comments • Photoacoustic Spectroscopy in Trace Gas Monitoring • Ultraviolet/Visible Light Detection and Ranging Applications in Air Monitoring

Environment: Water and Waste **(Volume 3)**
Atomic Fluorescence in Environmental Analysis • Flame and Graphite Furnace Atomic Absorption Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Food **(Volume 5)**
Atomic Spectroscopy in Food Analysis • Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science • X-ray Fluorescence in Forensic Science

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Atomic Spectroscopy **(Volume 11)**
Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Glow Discharge Optical Spectroscopy and Mass Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Ablation in Atomic Spectroscopy • Laser-induced Breakdown Spectroscopy

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy • Fluorescence Lifetime Measurements, Applications of

Infrared Spectroscopy **(Volume 12)**
Cavity Ringdown Laser Absorption Spectroscopy

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Absorption Techniques in X-ray Spectrometry • Energy Dispersive, X-ray Fluorescence Analysis • Sample Preparation for X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Ultrafast Diffraction Techniques • Wavelength-dispersive X-ray Fluorescence Analysis

General Articles **(Volume 15)**
Ultrafast Laser Technology and Spectroscopy

REFERENCES

1. K.W. Jackson, S. Lu, 'Atomic Absorption, Atomic Emission, and Flame Emission Spectrometry', *Anal. Chem.*, **70**(12), 363R–383R (1998).
2. J. Marshall, E.H. Evans, A. Fisher, S. Chenery, 'Atomic Spectrometry Update – Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **12**(6), 263R–290R (1997).
3. J. Marshall, S. Chenery, E.H. Evans, A. Fisher, 'Atomic Spectrometry Update – Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **13**(6), 107R–128R (1998).
4. E.H. Evans, S. Chenery, A. Fisher, J. Marshall, K. Sutton, 'Atomic Spectrometry Update – Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **14**(6), 977–1004 (1999).
5. S.J. Hill, J.B. Dawson, W.J. Price, I.L. Shuttler, C.M.M. Smith, J.F. Tyson, 'Atomic Spectrometry Update – Advances in Atomic Absorption and Fluorescence Spectrometry and Related Techniques', *J. Anal. At. Spectrom.*, **12**(8), R327–R379 (1997).

6. S.J. Hill, J.B. Dawson, W.J. Price, I.L. Shuttler, C.M.M. Smith, J.F. Tyson, 'Atomic Spectrometry Update – Advances in Atomic Absorption and Fluorescence Spectrometry and Related Techniques', *J. Anal. At. Spectrom.*, **13**(8), 131R–170R (1998).
7. S.J. Hill, J.B. Dawson, W.J. Price, I.L. Shuttler, C.M.M. Smith, J.F. Tyson, 'Atomic Spectrometry Update – Advances in Atomic Absorption and Fluorescence Spectrometry and Related Techniques', *J. Anal. At. Spectrom.*, **14**(8), 1245–1285 (1999).
8. T.H. Maiman, 'Stimulated Optical Radiation in Ruby', *Nature*, **187**, 493–494 (1960).
9. G.S. Hurst, M.H. Nayfeh, J.P. Young, 'A Demonstration of One-atom Detection', *Appl. Phys. Lett.*, **30**(5), 229–231 (1977).
10. G.S. Hurst, M.H. Nayfeh, J.P. Young, 'One-atom Detection Using Resonance Ionization Spectroscopy', *Phys. Rev. A*, **15**(6), 2283–2292 (1977).
11. G.S. Hurst, M.G. Payne, S.D. Kramer, J.P. Young, 'Resonance Ionization Spectrometry and One-atom Detection', *Rev. Mod. Phys.*, **51**, 767–819 (1979).
12. C.T.J. Alkemade, 'Atomic Physics and Atomic Spectroscopy: Mother and Daughter', *Spectrochim. Acta*, **38B**, 1395–1409 (1983).
13. A.L. Schawlow, 'Laser Spectroscopy of Atoms and Molecules', *Science*, **202**, 141–147 (1978).
14. W. Demtröder, *Laser Spectroscopy – Basic Concepts and Instrumentation*, 2nd edition, Springer-Verlag, Berlin, 1996.
15. L.J. Radziemski, R.W. Solarz, J.A. Paisner, *Laser Spectroscopy and its Applications*, Marcel Dekker, New York, 1987.
16. N. Omenetto (ed.), *Analytical Laser Spectroscopy*, Wiley, New York, 1979.
17. V.S. Letokhov (ed.), *Laser Analytical Spectrochemistry*, *Optics and Optoelectronics*, eds. E.R. Pike, W.T. Welford, Adam Hilger, Bristol, 1985.
18. V.S. Letokhov, *Laser Photoionization Spectroscopy*, Academic Press, Orlando, 1987.
19. G.S. Hurst, M.G. Payne, *Principles and Applications of Resonance Ionization Spectroscopy*, Adam Hilger, Bristol, 1988.
20. D.L. Andrews (ed.), *Applied Laser Spectroscopy*, VCH Publishers, New York, 1992.
21. J. Sneddon, T.L. Thiem, Y.-I. Lee (eds.), *Lasers in Analytical Atomic Spectroscopy*, VCH Publishers, New York, 1997.
22. L.A. Lyon, C.D. Keating, A.P. Fox, B.E. Baker, L. He, S.R. Nicewarner, S.P. Mulvaney, M.J. Natan, 'Raman Spectroscopy', *Anal. Chem.*, **70**(12), 341R–361R (1998).
23. A.G. Abroskin, T.V. Belyaeva, V.A. Filichkina, E.K. Ivanova, M.A. Proscurnin, V.M. Savostina, Y.A. Barbalat, 'Thermal Lensing Spectrometry in Trace Metal Analysis', *Analyst*, **117**, 1957–1962 (1992).
24. C. Moulin, N. Delorme, T. Berthoud, P. Mauchien, 'Double Beam Thermal Lens Spectroscopy for Actinides Detection and Speciation', *Radiochim. Acta*, **44/45**, 103–106 (1988).
25. C.D. Tran, 'Helium–Neon Laser Intracavity Photothermal Beam Deflection Spectrometry', *Anal. Chem.*, **58**, 1714–1716 (1986).
26. N. Omenetto, P. Cavalli, G. Rossi, G. Bidoglio, G.C. Turk, 'Thermal Lensing Spectrophotometry of Uranium(VI) with Pulsed Laser Excitation', *J. Anal. At. Spectrom.*, **2**, 579–583 (1987).
27. T. Berthoud, N. Delorme, 'Differential Dual-beam Thermal Lensing Spectrometry: Determination of Lanthanides', *Appl. Spectrosc.*, **41**(1), 15–19 (1987).
28. C.M. Phillips, S.R. Crouch, G.E. Leroi, 'Matrix Effects in Thermal Lensing Spectrometry: Determination of Phosphate in Saline Solutions', *Anal. Chem.*, **58**, 1710–1714 (1986).
29. T. Berthoud, P. Decambox, B. Kirsch, P. Mauchien, C. Moulin, 'Direct Uranium Trace Analysis in Plutonium Solutions by Time-resolved Laser-induced Spectrofluorometry', *Anal. Chem.*, **60**, 1296–1299 (1988).
30. T. Berthoud, P. Decambox, B. Kirsch, P. Mauchien, C. Moulin, 'Direct Determination of Traces of Lanthanide Ions in Aqueous Solutions by Laser-induced Time-resolved Spectrofluorimetry', *Anal. Chim. Acta*, **220**, 235–241 (1989).
31. P. Decambox, P. Mauchien, C. Moulin, 'Direct Trace Determination of Curium by Laser-induced Time-resolved Spectrofluorometry', *Radiochim. Acta*, **48**, 23–28 (1989).
32. G.P. Smith, M.J. Dyer, D.R. Crosley, 'Pulsed Laser Optoacoustic Detection of Flame Species', *Appl. Opt.*, **22**(24), 3995–4003 (1983).
33. W.G. Tong, J.M. Andrews, Z. Wu, 'Laser Spectrometry Based on Phase Conjugation by Resonant Degenerate Four-wave Mixing in an Analytical Flame', *Anal. Chem.*, **59**, 896–899 (1987).
34. K. Song, Y.-I. Lee, J. Sneddon, 'Applications of Laser-induced Breakdown Spectrometry', *Appl. Spectrosc. Rev.*, **32**, 183–235 (1997).
35. N. Omenetto, J.D. Winefordner, 'Atomic Fluorescence Spectrometry Basic Principles and Applications', *Prog. Anal. Atom. Spectrosc.*, **2**, 1–183 (1979).
36. J.D. Winefordner, 'Principles, Methodologies, and Applications of Atomic Fluorescence Spectrometry', *J. Chem. Educ.*, **55**(2), 72–78 (1978).
37. D.J. Butcher, J.P. Dougherty, F.R. Preli, A.P. Walton, G.-T. Wei, R.L. Irwin, R.G. Michel, 'Laser-excited Atomic Fluorescence Spectrometry in Flames, Plasmas and Electrothermal Atomisers, a Review', *J. Anal. At. Spectrom.*, **3**, 1059–1078 (1988).
38. X. Hou, S.-J.J. Tsai, J.X. Zhou, K.X. Yang, R.F. Lonardo, R.G. Michel, 'Laser-excited Atomic Fluorescence Spectrometry: Principles, Instrumentation and

Applications', in *Lasers in Analytical Spectroscopy*, eds. J. Sneddon, T.L. Thiem, Y.-I. Lee, VCH Publishers, New York, 83–123, 1997.
39. J.C. Travis, G.C. Turk, R.B. Green, 'Laser-enhanced Ionization Spectrometry', *Anal. Chem.*, **54**, 1006A–1018A (1982).
40. J.C. Travis, G.C. Turk (eds.), *Laser-enhanced Ionization Spectrometry. Analytical Chemistry and its Application*, ed. J.D. Winefordner, John Wiley & Sons, New York, Vol. 136, 1996.
41. G.S. Hurst, 'Resonance Ionization Spectroscopy', *Anal. Chem.*, **53**(13), 1448A–1456A (1981).
42. J.A. Silver, 'Frequency-modulation Spectroscopy for Trace Species Detection: Theory and Comparison Among Experimental Methods', *Appl. Opt.*, **31**(6), 707–717 (1992).
43. K. Niemax, H. Groll, C. Schnürer-Patschan, 'Element Analysis by Diode Laser Spectroscopy', *Spectrochim. Acta Rev.*, **15**(5), 349–377 (1993).
44. J. Sneddon, Y.-I. Lee, X. Hou, J.X. Zhou, R.G. Michel, 'Lasers', in *Lasers in Analytical Spectroscopy*, eds. J. Sneddon, T.L. Thiem, Y.-I. Lee, VCH Publishers, New York, 41–81, 1997.
45. J. Franzke, A. Schnell, K. Niemax, 'Spectroscopic Properties of Commercial Laser Diodes', *Spectrochim. Acta Rev.*, **15**, 379–395 (1993).
46. D.S. Bomse, A.C. Stanton, J.A. Silver, 'Frequency Modulation and Wavelength Modulation Spectroscopies: Comparison of Experimental Methods Using a Lead–Salt Diode Laser', *Appl. Opt.*, **31**(6), 718–731 (1992).
47. V.S. Letokhov, 'Laser-induced Chemistry – Basic Nonlinear Processes and Applications', *Appl. Phys. B*, **46**, 237–251 (1988).
48. N. Omenetto, 'The Impact of Several Atomic and Molecular Laser Spectroscopic Techniques for Chemical Analysis', *Appl. Phys. B*, **46**, 209–220 (1988).
49. J. Wolfrum, 'Laser Spectroscopy for Studying Chemical Processes', *Appl. Phys. B*, **46**, 221–236 (1988).
50. J.A. Paisner, 'Atomic Vapor Laser Isotope Separation', *Appl. Phys. B*, **46**, 253–260 (1988).
51. D. Bäuerle, 'Chemical Processing with Lasers: Recent Developments', *Appl. Phys. B*, **46**, 261–270 (1988).
52. S. Svanberg, *Atomic and Molecular Spectroscopy – Basic Aspects and Practical Applications*, 2nd edition, eds. G. Ecker, P. Lambropoulus, I.I. Sobelman, H. Walther, Springer-Verlag, Berlin, Vol. 6, 1992.
53. J.D. Winefordner, B.W. Smith, N. Omenetto, 'Theoretical Considerations of Laser Induced Fluorescence and Ionization Spectrometry: How Close to Single Atom Detection', *Spectrochim. Acta*, **44B**(12), 1397–1403 (1989).
54. J.D. Winefordner, C.L. Stevenson, 'Linking Principles with Absolute Detection Power in Atomic Spectrometry: How Far Can We Go?', *Spectrochim. Acta*, **48B**, 757–767 (1993).
55. J.D. Winefordner, G.A. Petrucci, C.L. Stevenson, B.W. Smith, 'Theoretical and Practical Limits in Atomic Spectroscopy Plenary Lecture', *J. Anal. At. Spectrom.*, **9**(3), 131–143 (1994).
56. G.N. Fowler, T.W. Preist, 'Ionization Cross-sections in Flames', *J. Chem. Phys.*, **56**, 1601–1605 (1972).
57. O. Axner, H. Rubinsztein-Dunlop, 'Fundamental Mechanisms of Laser-enhanced Ionization: The Production of Ions', in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, John Wiley & Sons, New York, 1–98, 1996.
58. R. Kelly, P.J. Padley, 'Measurement of Collisional Ionization Cross-sections for Metal Atoms in Flames', *Proc. R. Soc. London, A*, **327**, 345–366 (1972).
59. G. Zizak, J.D. Bradshaw, J.D. Winefordner, 'Rate Equation Solution for the Temporal Behavior of a Three-level System', *Appl. Opt.*, **19**(21), 3631–3639 (1980).
60. N. Omenetto, 'Analytical and Diagnostic Applications of Laser-induced Fluorescence in Flames and Plasmas', in *Analytical Laser Spectroscopy*, eds. S. Martellucci, A.N. Chester, Plenum Press, New York, 131–146, 1985.
61. N. Omenetto, C.A.v. Dijk, J.D. Winefordner, 'Some Considerations on the Saturation Parameter for 2- and 3-Level Systems in Laser-excited Fluorescence', *Spectrochim. Acta*, **37B**, 703–711 (1982).
62. C.T.J. Alkemade, 'Anomalous Saturation Curves in Laser-induced Fluorescence', *Spectrochim. Acta*, **40B**, 1331–1368 (1985).
63. D.R. de Olivares, G.M. Hieftje, 'Saturation of Energy Levels in Analytical Atomic Fluorescence Spectrometry – I. Theory', *Spectrochim. Acta*, **33B**, 79–99 (1978).
64. D.R. de Olivares, G.M. Hiefte, 'Saturation of Energy Levels in Analytical Atomic Fluorescence Spectrometry – II. Experimental', *Spectrochim. Acta*, **36B**(11), 1059–1079 (1981).
65. N. Omenetto, J. Bower, J. Bradshaw, C.A.v. Dijk, J.D. Winefordner, 'A Theoretical and Experimental Approach to Laser Saturation Broadening in Flames', *J. Quant. Spectrosc. Radiat. Transfer*, **24**, 147–158 (1980).
66. N. Omenetto, H.G.C. Human, 'Laser Excited Analytical Atomic and Ionic Fluorescence in Flame, Furnaces and Inductively Coupled Plasmas: I. – General Considerations', *Spectrochim. Acta*, **39B**, 1333–1343 (1984).
67. P.E. Walters, J. Lanauze, J.D. Winefordner, 'Spatially Resolved Concentration Studies of Ground State Atoms in a Flame: Saturated Absorption Spectroscopic Method', *Spectrochim. Acta*, **39B**(1), 125–129 (1984).
68. J.D. Winefordner, N. Omenetto, 'Laser-excited Atomic and Ionic Fluorescence in Flames and Plasmas', in *Analytical Applications of Lasers*, ed. E.H. Piepmeier, John Wiley & Sons, New York, 31–73, 1986.
69. P. Hannaford, 'Time-resolved Atomic Fluorescence in Flames', in *21st Colloquium Spectroscopicum Interna-*

tionale and 8th International Conference on Atomic Physics, Heyden, Cambridge, 250–261, 1979.
70. N. Omenetto, L.P. Hart, P. Benetti, J.D. Winefordner, 'On the Shape of Atomic Fluorescence Analytical Curves with a Laser Excitation Source', *Spectrochim. Acta*, **28B**, 301–307 (1973).
71. J.W. Daily, 'Saturation of Fluorescence in Flames with a Gaussian Laser Beam', *Appl. Opt.*, **17**(2), 225–229 (1978).
72. N. Omenetto, B.W. Smith, L.P. Hart, 'Laser Induced Fluorescence and Ionization Spectroscopy: Theoretical and Analytical Considerations for Pulsed Sources', *Fresenius' Z. Anal. Chem.*, **324**, 683–697 (1986).
73. O. Axner, P. Ljungberg, 'A Tutorial Review of the Rate-equation and Density-matrix Formalisms for Two- and Three-level Atomic and Molecular Systems in High Collisional Media Exposed to Pulsed Laser Light with Arbitrary Bandwidths', *Spectrochim. Acta Rev.*, **15**(4), 181–287 (1993).
74. O. Axner, T. Berglind, J.L. Heully, I. Lindgren, H. Rubinsztein-Dunlop, 'Improved Theory of Laser-enhanced Ionization in Flames: Comparison with Experiment', *J. Appl. Phys.*, **55**(9), 3215–3225 (1984).
75. M.A. Bolshov, A.V. Zybin, V.G. Koloshnikov, K.N. Koshelev, 'Some Characteristics of Laser-excited Atomic Fluorescence in a Three-level Scheme', *Spectrochim. Acta*, **32B**, 279–286 (1977).
76. D. Kleppner, M.G. Littman, M.L. Zimmerman, 'Highly Excited Atoms', in *Sci. Am.*, **245**, 130–149 (1981).
77. A.W. Miziolek, R.J. Willis, 'Saturated Double-resonance Emission Spectroscopy of Lead for Sensitive Atomic Analysis', *Opt. Lett.*, **6**, 528–530 (1981).
78. C.A. van der Wijngaart, P. Kuik, H.A. Dijkerman, T. Hollander, C.T.J. Alkemade, 'Collisional Repopulation of Excited Na-states upon Laser-pumping in an H_2-O_2-Ar Flame', *J. Quant. Spectrosc. Radiat. Transfer*, **37**(3), 267–281 (1987).
79. J.A. Vera, C.L. Stevenson, B.W. Smith, N. Omenetto, J.D. Winefordner, 'Laser-excited Atomic Fluorescence Spectrometry Using Graphite Tube Electrothermal Atomization and Double-resonance Excitation', *J. Anal. At. Spectrom.*, **4**, 619–623 (1989).
80. P. Ljung, E. Nyström, J. Enger, P. Ljungberg, O. Axner, 'Detection of Titanium in Electrothermal Atomizers by Laser-induced Fluorescence: Part 1 – Determination of Optimum Excitation and Detection Wavelengths', *Spectrochim. Acta*, **52B**(6), 675–701 (1997).
81. N. Omenetto, T. Berthoud, P. Cavalli, G. Rossi, 'Lifetime Measurements of Metastable Levels of Thallium and Lead in the Air–Acetylene Flame by Laser-enhanced Ionization Spectrometry', *Appl. Spectrosc.*, **39**, 500–503 (1985).
82. O. Axner, P. Ljungberg, Y. Malmsten, 'Lifetime Measurements of Metastable States of Mg, Sr, Cd, Au, Pb and Bi in an Acetylene–Air Flame by Laser-enhanced Ionization', *Appl. Phys. B*, **54**, 144–155 (1992).
83. O. Axner, P. Ljungberg, Y. Malmsten, 'Monitoring of the Lifetimes of the Metastable States in Au in an Air–Acetylene Flame as a Probe for Local Stoichiometric Conditions by the Laser-enhanced Ionization Technique', *Appl. Phys. B*, **56**, 355–362 (1993).
84. O. Axner, P. Ljungberg, Y. Malmsten, 'Laser-based Spectroscopic Technique for Fast and Reliable Measurements of Lifetimes of Atomic Metastable States in Combustive Situations', *Appl. Opt.*, **32**(6), 899–906 (1993).
85. M.A. Bolshov, A.V. Zybin, L.A. Zybina, V.G. Koloshnikov, I.A. Majorov, 'The Use of a Dye-laser for the Detection of Sub-picogram Amounts of Lead and Iron by Atomic Fluorescence Spectrometry', *Spectrochim. Acta*, **31B**, 493–500 (1976).
86. N. Omenetto, M.S. Epstein, J.D. Bradshaw, S. Bayer, J.J. Horvath, J.D. Winefordner, 'Fluorescence Ratio of the Two D Sodium Lines in Flames for D_1 and D_2 Excitation', *J. Quant. Spectrosc. Radiat. Transfer*, **22**, 287–291 (1979).
87. C.T.J. Alkemade, T. Hollander, W. Snelleman, P.J.T. Zeegers, *Metal Vapors in Flames*, Pergamon Press, Oxford, 1982.
88. K.C. Smyth, P.K. Schenck, W.G. Mallard, 'What Really Does Happen to Electronically Excited Atoms in Flames?', in *Laser Probes of Chemistry*, Chapter 12, American Chemical Society, Washington, DC, 1980.
89. O. Axner, 'Determination of Optimal Conditions for Laser-enhanced Ionization Spectrometry in Flames. Part 1. One-step Signal Strength versus Excitation Transition', *Spectrochim. Acta*, **45B**, 561–579 (1990).
90. O. Axner, T. Berglind, 'Determination of Ionization Efficiencies of Excited Atoms in a Flame by Laser-enhanced Ionization', *Appl. Spectrosc.*, **43**, 940–952 (1989).
91. G.C. Turk, W.G. Mallard, P.K. Schenck, K.C. Smyth, 'Improved Sensitivity for Laser-enhanced Ionization Spectrometry in Flames by Stepwise Excitation', *Anal. Chem.*, **51**, 2408–2410 (1979).
92. J.R. Ackerhalt, J.H. Eberly, 'Coherence versus Incoherence in Stepwise Laser Excitation of Atoms and Molecules', *Phys. Rev. A*, **14**(5), 1705–1710 (1976).
93. R.M. Measures, P.G. Cardinal, 'Laser Ionization Based on Resonance Saturation – a Simple Model Description', *Phys. Rev. A*, **23**(2), 804–815 (1981).
94. N. Omenetto, B.W. Smith, L.P. Hart, P. Cavalli, G. Rossi, 'Laser-induced Double-resonance Ionic Fluorescence in an Inductively Coupled Plasma', *Spectrochim. Acta*, **40B**, 1411–1422 (1985).
95. O. Axner, M. Norberg, M. Persson, H.R. Dunlop, 'Reduction of Spectral Interferences from Na in Laser-enhanced Ionization Spectrometry by Laser-preionization', *Appl. Spectrosc.*, **44**, 1117–1123 (1990).
96. B.W. Smith, L.P. Hart, N. Omenetto, 'Measurement of the Laser-induced Ionization Yield for Lithium in an Air Acetylene Flame', *Anal. Chem.*, **58**, 2147–2151 (1986).

97. N. Omenetto, G.C. Turk, M. Rutledge, J.D. Winefordner, 'Atomic and Ionic Fluorescence Dip Spectroscopy as a Tool for Flame and Plasma Diagnostics', *Spectrochim. Acta*, **42B**(6), 807–817 (1987).
98. N. Omenetto, M. Leong, C. Stevensson, J. Vera, B.W. Smith, J.D. Winefordner, 'Fluorescence Dip Spectroscopy of Sodium Atoms in an Inductively Coupled Plasma', *Spectrochim. Acta*, **43B**, 1093–1100 (1988).
99. O. Axner, M. Norberg, H. Rubinsztein-Dunlop, 'Simultaneous Measurements of Laser-induced Fluorescence (LIF Dip) and Laser-enhanced Ionization (LEI Enhancement) in Flame – Investigation of Physical Properties', *Spectrochim. Acta*, **44B**, 693–712 (1989).
100. N. Omenetto, B.W. Smith, B.T. Jones, J.D. Winefordner, 'Considerations on the Simultaneous Behavior of the Resonance Fluorescence and Ionization Signals versus Laser Intensity in Flames', *Appl. Spectrosc.*, **43**, 595–598 (1989).
101. M. Sargent, III, P. Horwitz, 'Three-level Rabi Flopping', *Phys. Rev. A*, **13**(5), 1962–1964 (1976).
102. J.W. Daily, 'Coherent Optical Transient Spectroscopy in Flames', *Appl. Opt.*, **18**(3), 360–367 (1979).
103. J.W. Daily, 'Use of Rate Equations to Describe Laser Excitation in Flames', *Appl. Opt.*, **16**(8), 2322–2327 (1977).
104. R. Salomaa, S. Stenholm, 'Two-photon Spectroscopy: Effects of a Resonant Intermediate State', *J. Phys. B*, **8**(11), 1795–1805 (1975).
105. R. Salomaa, S. Stenholm, 'Two-photon Spectroscopy II. Effects of Residual Doppler Broadening', *J. Phys. B*, **9**(8), 1221–1235 (1976).
106. R. Salomaa, 'Effects of Finite Laser Bandwidths in Two-photon Spectroscopy', *J. Phys. B*, **11**, 3745–3755 (1978).
107. O. Axner, P. Ljungberg, 'A Density Matrix Treatment of Step-wise Laser Excitations of Atoms with Degenerate States in High Collisional Media', *J. Quant. Spectrosc. Radiat. Transfer*, **50**(3), 277–292 (1993).
108. P. Ljungberg, D. Boudreau, O. Axner, 'Full Temporal Density Matrix Treatment of Dual-wavelength, Arbitrary Bandwidth, Pulsed-laser Excitation of Atoms with Degenerate States in High Collisional Media', *Spectrochim. Acta*, **49B**(12–14), 1491–1505 (1994).
109. O. Axner, S. Sjöström, 'Experimental and Theoretical Investigation of Pulsed Step-wise Laser Excitations of Atoms in Flames – The Role of Two-step versus Two-photon Excitations and Dynamical Stark Effects', *Spectrochim. Acta*, **47B**(2), 245–273 (1992).
110. R.P. Lucht, D.W. Sweeney, N.M. Laurendeau, 'Balanced Cross-rate Model for Saturated Molecular Fluorescence in Flames Using a Nanosecond Pulse Length Laser', *Appl. Opt.*, **19**(19), 3295–3300 (1980).
111. R.P. Lucht, N.M. Laurendeau, 'Two-level Model for Near Saturated Fluorescence in Diatomic Molecules', *Appl. Opt.*, **18**(9), 856–861 (1979).
112. W.G. Mallard, J.H. Miller, K.C. Smyth, 'Resonantly Enhanced Two-photon Photoionization of NO in an Atmospheric Flame', *J. Chem. Phys.*, **76**, 3483–3492 (1982).
113. D. Boudreau, P. Ljungberg, O. Axner, 'DENSMAT: Fully Time Resolved Simulation of Two-step Pulsed Laser Excitation of Atoms in Highly Collisional Media', *Spectrochim. Acta*, **51B**, 413–428 (1996).
114. R. Salomaa, 'Two-photon Spectroscopy III. General Strong-field Aspects', *J. Phys. B*, **10**(15), 3005–3021 (1977).
115. J.W. Daily, 'Laser-induced Fluorescence Spectroscopy in Flames', in *Laser Probes for Combustion Chemistry, ACS Symp. Ser.*, **134**, 61–83, 1980.
116. J.D. Winefordner, T.J. Vickers, 'Atomic Fluorescence as a Means of Chemical Analysis', *Anal. Chem.*, **36**(1), 161–165 (1964).
117. J.D. Winefordner, R.A. Staab, 'Determination of Zinc, Cadmium, and Mercury by Atomic Fluorescence Flame Spectrometry', *Anal. Chem.*, **36**(1), 165–168 (1964).
118. L.M. Fraser, J.D. Winefordner, 'Laser-excited Atomic Fluorescence Flame Spectrometry', *Anal. Chem.*, **43**(12), 1693–1696 (1971).
119. L.M. Fraser, J.D. Winefordner, 'Laser-excited Atomic Fluorescence Flame Spectroscopy as an Analytical Method', *Anal. Chem.*, **44**(8), 1444–1451 (1972).
120. J. Kuhl, H. Spitschan, 'Flame-fluorescence Detection of Mg, Ni, and Pb with a Frequency-doubled Dye Laser as Excitation Source', *Opt. Comm.*, **7**(3), 256–259 (1973).
121. K. Dittrich, H.-J. Stark, 'Laser-excited Molecular Fluorescence Spectrometry for the Determination of Traces on Nonmetals. Part 1. Determination of Traces of Fluoride, Chloride and Bromide Based on Diatomic Molecules in a Graphite Furnace', *Anal. Chim. Acta*, **200**, 581–591 (1987).
122. G.M. Murray, S.J. Weeks, M.C. Edelson, 'Determination of Uranium Isotopes in a Complex Matrix by Optical Spectroscopy', *J. Alloys Compd.*, **181**, 57–62 (1992).
123. D.J. Butcher, J.P. Dougherty, J.T. McCaffrey, F.R. Preli, A.P. Walton, R.G. Michel, 'Conventional Source Excited Atomic Fluorescence Spectrometry', *Prog. Anal. Atom. Spectrosc.*, **10**, 359–506 (1987).
124. B.W. Smith, K.N. Spears, J.D. Winefordner, M.R. Glick, 'A Comprehensive Table of Atomic Fluorescence Detection Limits and Experimental Conditions', *Appl. Spectrosc.*, **43**(3), 376–414 (1989).
125. N. Omenetto, J.D. Winefordner, 'Atomic Fluorescence Spectroscopy with Laser Excitation', in *Analytical Laser Spectroscopy*, ed. N. Omenetto, Wiley, New York, Chapter 4, 1979.
126. N. Omenetto, J.D. Winefordner, 'Lasers in Analytical Spectroscopy', *Crit. Rev. Anal. Chem.*, **13**(1), 59–115 (1981).
127. N. Omenetto, 'Laser-induced Atomic Fluorescence Spectroscopy: A Personal Viewpoint on its Status,

Needs, and Perspectives', *Spectrochim. Acta*, **44B**(2), 131–146 (1989).
128. N. Omenetto, 'Laser-induced Fluorescence in a Furnace: A Viable Approach to Absolute Analysis?', *Mikrochim. Acta*, **II**, 277–285 (1991).
129. N. Omenetto, B.W. Smith, J.D. Winefordner, 'Laser-induced Fluorescence and Electrothermal Atomization: How Far from the Intrinsic Level of Detection?', *Spectrochim. Acta*, **43B**, 1111–1118 (1988).
130. S. Sjöström, 'Laser-excited Atomic Fluorescence Spectrometry in Graphite Furnace Electrothermal Atomizer', *Spectrochim. Acta Rev.*, **13**, 407–465 (1990).
131. S. Sjöström, P. Mauchien, 'Laser Atomic Spectroscopic Techniques – The Analytical Performance for Trace Element Analysis of Solid and Liquid Samples', *Spectrochim. Acta Rev.*, **15**(3), 153–180 (1993).
132. H. Falk, 'Analytical Capabilities at Atomic Spectrometric Methods Using Tunable Lasers: A Theoretical Approach', *Prog. Anal. Atom. Spectrosc.*, **3**, 181–208 (1980).
133. H. Falk, 'Methods for the Detection of Single Atoms Using Optical and Mass-spectrometry', *J. Anal. At. Spectrom.*, **7**, 255–260 (1992).
134. N. Omenetto, 'Role of Lasers in Analytical Atomic Spectroscopy: Where, When and Why – Plenary Lecture', *J. Anal. At. Spectrom.*, **13**(5), 385–399 (1998).
135. C.L. Stevenson, J.D. Winefordner, 'Estimating Detection Limits in Ultra Trace Analysis. 3. Monitoring Atoms and Molecules with Laser-induced Fluorescence', *Appl. Spectrosc.*, **46**(5), 715–724 (1992).
136. N. Omenetto, N.N. Hatch, L.M. Fraser, J.D. Winefordner, 'Laser-excited Atomic Fluorescence of Some Transition Elements in the Nitrous Oxide–Acetylene Flame', *Anal. Chem.*, **45**(1), 65–78 (1973).
137. N. Omenetto, P. Benetti, L.P. Hart, J.D. Winefordner, C.T.J. Alkemade, 'Nonlinear Optical Behavior in Atomic Fluorescence Flame Spectrometry', *Spectrochim. Acta*, **28B**, 289–300 (1973).
138. N. Omenetto, N.N. Hatch, L.M. Fraser, J.D. Winefordner, 'Laser-excited Atomic and Ionic Fluorescence of the Rare Earths in the Nitrous Oxide–Acetylene Flame', *Anal. Chem.*, **45**(1), 195–197 (1973).
139. S.J. Weeks, H. Haraguchi, J.D. Winefordner, 'Laser-excited Molecular Fluorescence of CaOH in an Air–Acetylene Flame', *J. Quant. Spectrosc. Radiat. Transfer*, **19**, 633–640 (1978).
140. K. Fujiwara, N. Omenetto, J.B. Bradshaw, J.N. Bower, S. Nikdel, J.D. Winefordner, 'Laser-induced Molecular Background Fluorescence in Flames', *Spectrochim. Acta*, **34B**, 317–329 (1979).
141. M.E. Blackburn, J.-M. Mermet, J.D. Winefordner, 'CW-Laser-excited Molecular Fluorescence of Species in Flames', *Spectrochim. Acta*, **34A**, 847–852 (1978).
142. R.B. Green, J.C. Travis, R.A. Keller, 'Resonance Flame Atomic Fluorescence Spectrometry with Continuous Wave Dye Laser', *Anal. Chem.*, **48**(13), 1954–1959 (1976).
143. D.A. Goff, E.S. Yeung, 'Atomic Fluorescence Spectrometry with a Wavelength-modulated Continuous Wave Dye Laser', *Anal. Chem.*, **50**(4), 625–627 (1978).
144. B.W. Smith, M.B. Blackburn, J.D. Winefordner, 'Atomic Fluorescence Flame Spectroscopy with a Continuous Wave Dye Laser', *Can. J. Spectrosc.*, **22**, 57–61 (1977).
145. S.J. Weeks, H. Haraguchi, J.D. Winefordner, 'Improvement of Detection Limits in Laser-excited Atomic Fluorescence Flame Spectrometry', *Anal. Chem.*, **50**, 360–368 (1978).
146. C.A.v. Dijk, 'Excitation Profiles of the Sodium D-doublet in an H_2-O_2-Ar Flame', *Opt. Commun.*, **22**(3), 343–345 (1977).
147. R.A. van Calcar, M.J. van de Ven, B.K. van Uitert, K.J. Biewenga, T. Hollander, C.H. Alkemade, 'Saturation of Sodium Fluorescence in a Flame Irradiated with a Pulsed Tunable Dye Laser', *J. Quant. Spectrosc. Radiat. Transfer*, **21**, 11–18 (1979).
148. J.J. Horvath, J.D. Bradshaw, J.N. Bower, M.S. Epstein, J.D. Winefordner, 'Comparison of Nebulizer–Burner Systems for Laser-excited Atomic Fluorescence Flame Spectrometry', *Anal. Chem.*, **53**, 6–9 (1981).
149. M.S. Epstein, S. Bayer, J. Bradshaw, E. Voigtman, J.D. Winefordner, 'Application of Laser-excited Fluorescence Spectrometry to the Determination of Iron', *Spectrochim. Acta*, **35B**, 233–237 (1980).
150. M.S. Epstein, S. Bayer, J. Bradshaw, E. Voigtman, J.D. Winefordner, 'Application of Laser-excited Fluorescence Spectrometry to the Determination of Nickel and Tin', *Appl. Spectrosc.*, **34**, 372–376 (1980).
151. F.E. Hovis, J.A. Gelbwachs, 'Determination of Barium at Trace Levels by Laser-induced Ionic Fluorescence Spectrometry', *Anal. Chem.*, **56**, 1392–1394 (1984).
152. S.V. Kachin, B.W. Smith, J.D. Winefordner, 'Laser-excited Atomic Fluorescence of some Precious Metals in the Air/Acetylene Flame', *Appl. Spectrosc.*, **39**(4), 587–590 (1985).
153. M.D. Seltzer, M.S. Hendrick, R.G. Michel, 'Photomultiplier Gating for Improved Detection in Laser-excited Atomic Fluorescence Spectrometry', *Anal. Chem.*, **57**, 1096–1100 (1985).
154. J.X. Zhou, X.D. Hou, S.J.J. Tsai, K.X. Yang, R.G. Michel, 'Characterization of a Tunable Optical Parametric Oscillator Laser System for Multielement Flame Laser-excited Atomic Fluorescence Spectrometry of Cobalt, Copper, Lead, Manganese, and Thallium in Buffalo River Sediment', *Anal. Chem.*, **69**(3), 490–499 (1997).
155. N. Omenetto, H.G.C. Human, P. Cavalli, G. Rossi, 'Direct Determination of Lead in Blood by Laser-excited Flame Atomic Fluorescence Spectrometry', *Analyst*, **109**, 1067–1070 (1984).

156. A.P. Walton, G.-T. Wei, Z. Liang, R.G. Michel, 'Laser-excited Atomic Fluorescence in a Flame as a High Sensitivity Detector for Organomanganese and Organotin Compounds Following Separation by High-performance Liquid Chromatography', *Anal. Chem.*, **63**, 232–240 (1991).
157. H.G.C. Human, N. Omenetto, P. Cavalli, G. Rossi, 'Laser-excited Analytical Atomic and Ionic Fluorescence in Flames, Furnaces and Inductively Coupled Plasmas – II. Fluorescence Characteristics and Detection Limits for Fourteen Elements', *Spectrochim. Acta*, **39B**, 1345–1363 (1984).
158. G.A. Petrucci, H. Beisler, O. Matveev, P. Cavalli, N. Omenetto, 'Analytical and Spectroscopic Characterization of Double-resonance Laser-induced Fluorescence of Gold Atoms in a Graphite Furnace and in a Flame', *J. Anal. At. Spectrom.*, **10**(10), 885–890 (1995).
159. J.P. Dougherty, F.R. Preli, R.G. Michel, 'Laser-excited Atomic Fluorescence Spectrometry in an Atomic Absorption Graphite Tube Furnace', *J. Anal. At. Spectrom.*, **2**, 429–434 (1987).
160. F.R. Preli, J.P. Dougherty, R.G. Michel, 'Laser-excited Atomic Fluorescence Spectrometry with a Laboratory Constructed Tube Electrothermal Atomizer', *Anal. Chem.*, **59**, 1784–1789 (1987).
161. J.P. Dougherty, F.R. Preli, Jr, J.T. McCaffrey, M.D. Seltzer, R.G. Michel, 'Instrumentation for Zeeman Electrothermal Atomizer Laser-excited Fluorescence Spectrometry', *Anal. Chem.*, **59**, 1112–1119 (1987).
162. J.P. Dougherty, J.A. Costello, R.G. Michel, 'Determination of Tl in Bovine Liver and Mouse Brains by ETA LEAFS', *Anal. Chem.*, **60**, 336–340 (1988).
163. F.R. Preli, J.P. Dougherty, R.G. Michel, 'Diagnostic Studies of the Zeeman Effect for Laser-excited Atomic Fluorescence Spectrometry in an Electrothermal Atomizer', *Spectrochim. Acta*, **43B**, 501–517 (1988).
164. J.P. Dougherty, F.R. Preli, G.-T. Wei, R.G. Michel, 'Nonlinearity of Calibration Graphs for Laser-excited Atomic Fluorescence in Graphite-tube Atomizers', *Appl. Spectrosc.*, **44**(6), 934–944 (1990).
165. V.M. Apatin, B.V. Arkhangel'skii, M.A. Bol'shov, V.V. Ermolov, V.G. Koloshnikov, O.N. Kompanetz, N.I. Kuznetsov, E.L. Mikhailov, V.S. Shishkovskii, 'Automated Laser-excited Atomic Fluorescence Spectrometer for Determination of Trace Concentration of Elements', *Spectrochim. Acta*, **44B**(3), 253–262 (1989).
166. M.A. Bolshov, V.G. Kolosnikov, S.N. Rudnev, C.F. Boutron, U. Görlach, C.C. Patterson, 'Detection of Trace Amounts of Toxic Metals in Environmental Samples by Laser-excited Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **7**, 99–104 (1992).
167. C.L. Boutron, M.A. Bolshov, V.G. Koloshnikov, C.C. Patterson, N.I. Barkov, 'Direct Determination of Lead in Vostok Antarctic Ancient Ice by Laser-excited Atomic Fluorescence Spectrometry', *Atmos. Environ.*, **24A**(7), 1797–1800 (1990).
168. M.A. Bolshov, C.F. Boutron, A.V. Zybin, 'Determination of Lead in Antarctic Ice at the Picogram-per-gram Level by Laser Atomic Fluorescence Spectrometry', *Anal. Chem.*, **61**, 1758–1762 (1989).
169. M.A. Bolshov, C.F. Boutron, F.M. Ducroz, U. Görlach, O.N. Kompanetz, S.N. Rudniev, B. Hutch, 'Direct Ultra-trace Determination of Cadmium in Antarctic and Greenland Snow and Ice by Laser Atomic Fluorescence Spectrometry', *Anal. Chim. Acta*, **251**, 169–175 (1991).
170. M.A. Bolshov, S.N. Rudnev, J. Brust, 'Analytical Characterization of Laser-excited Atomic Fluorescence of Bismuth', *Spectrochim. Acta*, **49B**(12–14), 1437–1444 (1994).
171. J.P. Dougherty, F.R. Preli, R.G. Michel, 'Laser-excited Atomic-fluorescence Spectrometry in an Electrothermal Atomizer with Zeeman Background Correction', *Talanta*, **36**(1/2), 151–159 (1989).
172. Z. Liang, G.-T. Wei, R.L. Irwin, A.P. Walton, R.G. Michel, J. Sneddon, 'Determination of Subnanogram per Cubic Meter Concentrations of Metals in the Air of a Trace Metal Clean Room by Impaction Graphite Furnace Atomic Absorption and Laser-excited Atomic Fluorescence Spectrometry', *Anal. Chem.*, **62**, 1452–1457 (1990).
173. O. Axner, H. Rubinsztein-Dunlop, 'Detection of Trace Amounts of Cr by Two Laser-based Spectroscopic Techniques: Laser-enhanced Ionization in Flames and Laser-induced Fluorescence in Graphite Furnace', *Appl. Opt.*, **32**(6), 867–884 (1993).
174. G.-T. Wei, J.P. Dougherty, F.R. Preli, R.G. Michel, 'Signal and Noise Considerations of Nondispersive Laser-excited Atomic Fluorescence in a Graphite Tube Atomizer with Front-surface Illumination', *J. Anal. At. Spectrom.*, **5**, 249–259 (1990).
175. D.J. Butcher, R.L. Irwin, J. Takahashi, G. Su, G.-T. Wei, R.G. Michel, 'Determination of Thallium, Manganese, Lead in Food, Agricultural Standard Reference Materials by Electrothermal Atomizer Laser-excited Atomic Fluorescence, Atomic Absorption Spectrometry with Slurry Sampling', *Appl. Spectrosc.*, **44**(9), 1521–1533 (1990).
176. N. Omenetto, P. Cavalli, M. Broglia, P. Qi, G. Rossi, 'Laser-induced Single Resonance and Double Resonance Atomic Fluorescence Spectrometry in a Graphite Tube Atomizer', *J. Anal. At. Spectrom.*, **3**, 231–235 (1988).
177. D.J. Butcher, R.L. Irwin, J. Takahashi, R.G. Michel, 'Determination of Fluorine in Urine and Tap Water by Laser-excited Molecular Fluorescence Spectrometry in a Graphite Tube Furnace with Front-surface Illumination', *J. Anal. At. Spectrom.*, **6**, 9–18 (1991).
178. W. Resto, R.G. Badini, B.W. Smith, C.L. Stevenson, J.D. Winefordner, 'Two-step Laser-excited Atomic Fluorescence Spectrometry Determination of Mercury', *Spectrochim. Acta*, **48B**(5), 627–632 (1993).

179. E. Masera, P. Mauchien, Y. Lerat, 'Electrothermal Atomization – Laser-induced Fluorescence Determination of Iridium, Rhodium, Palladium, Platinum and Gold at the ng/l Level in Pure Water', *Spectrochim. Acta*, **51B**, 543–548 (1996).
180. N. Chekalin, A. Marunkov, O. Axner, 'Laser-induced Fluorescence in Graphite Furnaces under Low Pressure Conditions as a Powerful Technique for Studies of Atomization Mechanisms: Investigation of Ag', *Spectrochim. Acta*, **49B**(12–14), 1411–1435 (1994).
181. O. Axner, N. Chekalin, P. Ljungberg, Y. Malmsten, 'Direct Determination of Thallium in Natural Waters by Laser-induced Fluorescence in a Graphite Furnace', *Int. J. Environ. Anal. Chem.*, **53**, 185–193 (1993).
182. R.L. Irwin, D.J. Butcher, J. Takahashi, G.-T. Wei, R.G. Michel, 'Determination of Thallium and Lead in Nickel-based Alloys by Direct Solid Sampling with Graphite Furnace Laser-excited Atomic Fluorescence', *J. Anal. At. Spectrom.*, **5**, 603–610 (1990).
183. Z. Liang, R.F. Lonardo, R.G. Michel, 'Determination of Tellurium and Antimony in Nickel Alloys by Laser-excited Atomic Fluorescence Spectrometry in a Graphite Furnace', *Spectrochim. Acta*, **48B**(1), 7–23 (1993).
184. E. Masera, P. Mauchien, B. Remy, Y. Lerat, 'Characterization and Reduction of Silver Matrix Induced Effects in the Determination of Gold, Iridium, Palladium, Platinum and Rhodium by Graphite Furnace Laser-induced Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **11**(3), 213–223 (1996).
185. E. Masera, P. Mauchien, Y. Lerat, 'Silver Matrix Effects on Gold Atomization in a Graphite-furnace Investigated by 2-Dimensional Laser Imaging with a Gated Charge-coupled-device Camera', *J. Anal. At. Spectrom.*, **10**(2), 137–144 (1995).
186. J. Enger, Y. Malmsten, P. Ljungberg, O. Axner, 'Laser-induced Fluorescence in a Graphite Furnace as a Sensitive Technique for Assessment of Traces in North Arctic Atmospheric Aerosol Samples', *Analyst*, **120**, 635–641 (1995).
187. A. Marunkov, N. Chekalin, J. Enger, O. Axner, 'Detection of Trace Amounts of Ni by Laser-induced Fluorescence in Graphite Furnace with ICCD Detection', *Spectrochim. Acta*, **49B**(12–14), 1385–1410 (1994).
188. P. Ljung, E. Nyström, O. Axner, W. Frech, 'Detection of Titanium in Electrothermal Atomizers by Laser-induced Fluorescence: Part 2 – Investigation of Various Types of Atomizers', *Spectrochim. Acta*, **52B**(6), 703–716 (1997).
189. J. Enger, A. Marunkov, N. Chekalin, O. Axner, 'Direct Detection of Sb in Environmental and Biological Samples at pg/ml Concentrations by Laser-induced Fluorescence in Graphite Furnace with Intensified Charge-coupled Device', *J. Anal. At. Spectrom.*, **10**, 539–549 (1995).
190. U. Heitmann, T. Sy, A. Hese, G. Schoknecht, 'High-sensitivity Detection of Selenium and Arsenic by Laser-excited Atomic Fluorescence Spectrometry Using Electrothermal Atomization', *J. Anal. At. Spectrom.*, **9**, 437–442 (1994).
191. A.I. Yuzefovsky, R.F. Lonardo, M. Wang, R.G. Michel, 'Determination of Ultra-trace Amounts of Cobalt in Ocean Water by Laser-excited Atomic Fluorescence Spectrometry in a Graphite Electrothermal Atomizer with Semi On-line Flow Injection Preconcentration', *J. Anal. At. Spectrom.*, **9**, 1195–1202 (1994).
192. A.I. Yuzefovsky, R.G. Michel, 'Role of Barium Chemical Modifier in the Determination of Fluoride by Laser-excited Molecular Fluorescence of Magnesium Fluoride in a Graphite Tube Furnace', *J. Anal. At. Spectrom.*, **9**, 1203–1207 (1994).
193. R.L. Irwin, G.T. Wei, D.J. Butcher, Z. Liang, E.G. Su, J. Takahashi, A.P. Walton, R.G. Michel, 'Transverse Zeeman Background Correction for Graphite-furnace Laser-excited Atomic Fluorescence Spectrometry – Determination of Lead and Cobalt in Standard Reference Materials', *Spectrochim. Acta*, **47B**(13), 1497–1515 (1992).
194. S. Sjöström, O. Axner, M. Norberg, 'Detection of Vanadium by Laser-excited Atomic Fluorescence Spectrometry in a Side-heated Graphite Furnace', *J. Anal. At. Spectrom.*, **8**(2), 375–378 (1993).
195. Z.W. Liang, R.F. Lonardo, J. Takahashi, R.G. Michel, F.R. Preli, 'Laser-excited Fluorescence Spectrometry of Phosphorus Monoxide and Phosphorus in an Electrothermal Atomizer – Determination of Phosphorus in Plant and Biological Reference Materials and Nickel-alloys', *J. Anal. At. Spectrom.*, **7**(6), 1019–1028 (1992).
196. B.W. Smith, P.B. Farnsworth, P. Cavalli, N. Omenetto, 'Optimization of Laser-excited Atomic Fluorescence in a Graphite Furnace for the Determination of Thallium', *Spectrochim. Acta*, **45B**(12), 1369–1373 (1990).
197. S. Sjöström, 'Multi-channel Background Correction Technique for Pulsed Laser-excited Atomic Fluorescence with a Graphite Furnace', *J. Anal. At. Spectrom.*, **5**, 261–267 (1990).
198. M.A. Bolshov, S.N. Rudnev, J.-P. Candelone, C.F. Boutron, S. Hong, 'Ultratrace Determination of Bi in Greenland Snow by Laser-excited Atomic Fluorescence Spectrometry', *Spectrochim. Acta*, **49B**(12–14), 1445–1452 (1994).
199. R.F. Lonardo, A.I. Yuzefovsky, K.X. Yang, R.G. Michel, E.S. Frame, J. Barren, 'Electrothermal Atomizer Laser-excited Atomic Fluorescence Spectrometry for the Determination of Phosphorus in Polymers by Direct Solid Analysis and Dissolution', *J. Anal. At. Spectrom.*, **11**(4), 279–285 (1996).
200. M. Leong, J. Vera, B.W. Smith, N. Omenetto, J.D. Winefordner, 'Laser-induced Double Resonance Flu-

orescence of Lead with Graphite Tube Atomization', *Anal. Chem.*, **60**, 1605–1610 (1988).
201. B. Remy, I. Verhaeghe, P. Mauchien, 'Real Sample Analysis by ETA/LEAFS with Background Correction: Application to Gold Determination in River Water', *Appl. Spectrosc.*, **44**, 1633–1638 (1990).
202. J.A. Vera, M.B. Leong, C.L. Stevenson, G. Petrucci, J.D. Winefordner, 'Laser-excited Atomic-fluorescence Spectrometry with Electrothermal Tube Atomization', *Talanta*, **36**(12), 1291–1293 (1989).
203. M.A. Bolshov, A.Z. Zybin, I.I. Smirenkina, 'Atomic Fluorescence Spectrometry with Laser Excitation', *Spectrochim. Acta*, **36B**, 1143–1152 (1981).
204. M.A. Bolshov, A.V. Zybin, V.G. Koloshnikov, I.I. Smirenka, 'Analytical Applications of a LEAFS ETA Method', *Spectrochim. Acta*, **43B**, 519–528 (1988).
205. L.K. Denisov, A.F. Loshin, N.A. Kozlov, V.G. Nikiforov, 'Analysis of Na and Ba in an Atomic Fluorescence Method with Excitation by a Pulsed Dye Laser with Lamp Pumping', *Zh. Prikl. Spektrosk.*, **43**(4), 1092–1096 (1985).
206. M.A. Bolshov, A.V. Zybin, V.G. Koloshnikov, A.V. Pisarskii, A.N. Smirnov, 'Atom-fluorescence Analysis of Pt, Ir, Eu under Excitation by Pulsed Dye Lasers', *Zh. Prikl. Spektrosk.*, **28**, 45–49 (1978).
207. M.A. Bolshov, A.V. Zybin, V.G. Koloshnikov, M.V. Vasnetsov, 'Detection of Extremely Low Lead Concentrations by Laser Atomic Fluorescence Spectrometry', *Spectrochim. Acta*, **36B**(4), 345–350 (1981).
208. M.A. Bolshov, A.V. Zybin, V.G. Koloshnikov, I.A. Mayorov, I.I. Smirenkina, 'Laser-excited Fluorescence Analysis with Electrothermal Sample Atomization in Vacuum', *Spectrochim. Acta*, **41B**, 487–492 (1986).
209. J.A. Vera, M.B. Leong, N. Omenetto, B.W. Smith, B. Womack, J.D. Winefordner, 'Evaluation of Three Different Laser Systems for the Determination of Lead by Laser-excited Atomic Fluorescence Spectrometry with Graphite Tube Atomizers', *Spectrochim. Acta*, **44B**(10), 939–948 (1989).
210. V. Cheam, J. Lechner, I. Sekerka, J. Nriagu, G. Lawson, 'Development of a Laser-excited Atomic Fluorescence Spectrometer and a Method for the Direct Determination of Lead in Great Lakes Waters', *Anal. Chim. Acta*, **269**, 129–136 (1992).
211. V. Cheam, J. Lechner, I. Sekerka, R. Desrosiers, 'Direct Determination of Lead in Seawaters by Laser-excited Atomic Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **9**, 315–320 (1994).
212. V. Cheam, J. Lechner, R. Desrosiers, I. Sekerka, 'Direct Determination of Dissolved and Total Thallium in Lake Waters by Laser-excited Atomic Fluorescence Spectrometry', *Int. J. Environ. Anal. Chem.*, **63**(2), 153–165 (1996).
213. J.B. Womack, C.A. Ricard, B.W. Smith, J.D. Winefordner, 'Evaluation of a Continuous-flow Furnace Atomizer for Laser-excited Atomic Fluorescence Spectrometry', *Spectrosc. Lett.*, **22**(10), 1333–1345 (1989).
214. H. Falk, J. Tilch, 'Atomization Efficiency and Over-all Performance of Electrothermal Atomizers in Atomic Absorption, Furnace Atomization Non Thermal Excitation and Laser-excited Atomic Fluorescence Spectroscopy', *J. Anal. At. Spectrom.*, **2**, 527–531 (1987).
215. H. Falk, H.-J. Paetzold, K.P. Schmidt, J. Tilch, 'Analytical Application of Laser-excited Atomic Fluorescence Using a Graphite Cup Atomizer', *Spectrochim. Acta*, **43B**, 1101–1109 (1988).
216. S. Neumann, M. Kriese, 'Sub-picogram Detection of Lead by Nonflame Atomic Fluorescence Spectrometry with Dye Laser Excitation', *Spectrochim. Acta*, **29**, 127–137 (1974).
217. J.P. Hohimer, P.J. Hargis, Jr, 'Automatic Fluorescence Spectrometry of Thallium with a Frequency-doubled Dye-laser and Vitreous Carbon Atomizer', *Anal. Chim. Acta*, **97**, 43–49 (1978).
218. K. Dittrich, H.-J. Stärk, 'Laser-excited Atomic Fluorescence Spectrometry as a Practical Analytical Method: Part 1 Design of a Graphite Tube Atomizer for the Determination of Trace Amounts of Lead', *J. Anal. At. Spectrom.*, **1**, 237–241 (1986).
219. J.P. Hohimer, P.I. Hargis, Jr, 'Picogram Detection of Cesium in Aqueous Solution by Nonflame Atomic Fluorescence Spectroscopy with Dye Laser Excitation', *Appl. Phys. Lett.*, **30**(7), 344–346 (1977).
220. K. Dittrich, H.-J. Stärk, 'Laser-excited Atomic Fluorescence Spectrometry as a Practical Analytical Method: Part 2 Evaluation of a Graphite Tube Atomizer for the Determination of Trace Amounts of Indium, Gallium, Aluminum, Vanadium and Iridium by LEAFS', *J. Anal. At. Spectrom.*, **2**, 63–66 (1987).
221. D. Goforth, J.D. Winefordner, 'Laser-excited Atomic Fluorescence of Atoms Produced in a Graphite Furnace', *Anal. Chem.*, **58**, 2598–2602 (1986).
222. P. Wittman, J.D. Winefordner, 'Laser-excited Atomic Fluorescence Spectrometry with Graphite Filament Atomization', *Can. J. Spectrosc.*, **29**(3), 75–78 (1983).
223. J. Anwar, J.M. Anzano, G. Petrucci, J.D. Winefordner, 'Determination of Chloride at Picogram Levels by Molecular Fluorescence in a Graphite Furnace', *Analyst*, **116**(10), 1025–1028 (1991).
224. D. Goforth, J.D. Winefordner, 'A Graphite Tube Furnace for Use in Laser-excited Atomic-fluorescence Spectrometry', *Talanta*, **34**(2), 290–292 (1987).
225. A. Zybin, C. Schnürer-Patschan, K. Niemax, 'Simultaneous Multi-element Analysis in a Commercial Graphite Furnace by Diode Laser-induced Fluorescence', *Spectrochim. Acta*, **47B**(14), 1519–1524 (1992).
226. A. Mellone, J.D. Winefordner, 'Graphite-furnace Vaporization of Polycyclic Aromatic-compounds with Laser-induced Fluorescence of Vapors to Fingerprint Complex Environmental Materials', *Microchem. J.*, **42**(1), 126–137 (1990).

227. J.X. Zhou, X.D. Hou, K.X. Yang, R.G. Michel, 'Laser-excited Atomic Fluorescence Spectrometry in a Graphite Furnace with an Optical Parametric Oscillator Laser for Sequential Multi-element Determination of Cadmium, Cobalt, Lead, Manganese and Thallium in Buffalo River Sediment', *J. Anal. At. Spectrom.*, **13**(1), 41–47 (1998).
228. N. Omenetto, S. Nikdel, J.D. Bradshaw, M.S. Epstein, R.D. Reeves, J.D. Winefordner, 'Diagnostic and Analytical Studies of the Inductively Coupled Plasma by Atomic Fluorescence Spectrometry', *Anal. Chem.*, **51**(9), 1521–1525 (1979).
229. X. Huang, J. Lanauze, J.D. Winefordner, 'Laser-excited Atomic Fluorescence of Some Precious Metals and Refractory Elements in the Inductively Coupled Plasma', *Appl. Spectrosc.*, **39**(6), 1042–1047 (1985).
230. J.B. Simeonsson, K.C. Ng, J.D. Winefordner, 'Single-resonance and Double-resonance Atomic Fluorescence Spectrometry with Inductively Coupled Plasma Atomization and Laser Excitation', *Appl Spectrosc.*, **45**(9), 1456–1462 (1991).
231. M. Glick, B.W. Smith, J.D. Winefordner, 'Laser-excited Atomic Fluorescence in a Pulsed Hollow-cathode Glow-discharge', *Anal. Chem.*, **62**(2), 157–161 (1990).
232. B.W. Smith, J.B. Womack, J.D. Winefordner, N. Omenetto, 'Approaching Single Atom Detection with Atomic Fluorescence in a Glow-discharge Atom Reservoir', *Appl. Spectrosc.*, **43**(5), 873–876 (1989).
233. O.S. Lunyov, S.V. Oshemkov, 'Determination of Trace Metals by Laser-excited Fluorescence with Hot Hollow Cathode Atomization', *Spectrochim. Acta*, **47B**(1), 71–81 (1992).
234. J.C. Travis, G.C. Turk, R.L. Watters, Jr, L.-J. Yu, J.L. Blue, 'Trace Detection in Conducting Solids Using Laser-induced Fluorescence in a Cathodic Sputtering Cell', *J. Anal. At. Spectrom.*, **6**, 261–271 (1991).
235. G.I. Romanovska, V.I. Pogonin, A.K. Chibisov, 'Fluorescence Determination of Trace Amounts of Uranium(VI) in Various Materials by a Repetitive Laser Technique', *Talanta*, **34**(1), 207–210 (1987).
236. J.X. Zhou, X.D. Hou, K.X. Yang, S.-J.J. Tsai, R.G. Michel, 'Lasers Based on Optical Parametric Devices: Wavelength Tunability Empowers Laser-based Techniques in the UV, VIS, and near-IR', *Appl. Spectrosc.*, **52**, 176A–189A (1998).
237. G.C. Turk, 'Analytical Performance of Laser-enhanced Ionization in Flames', in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, John Wiley & Sons, New York, 161–211, 1996.
238. M.J. Rutledge, M.E. Tremblay, J.D. Winefordner, 'Measurement Methods for Atomic Fluorescence and Laser-enhanced Ionization Spectrometries with a Copper-vapor Pumped Dye Laser', *Appl. Spectrosc.*, **41**(1), 5–9 (1987).
239. A.I. Yuzefovsky, R.F. Lonardo, R.G. Michel, 'Spatial Discrimination Against Background with Different Optical-systems for Collection of Fluorescence in Laser-excited Atomic Fluorescence Spectrometry with a Graphite Tube Electrothermal Atomizer', *Anal. Chem.*, **67**(13), 2246–2255 (1995).
240. P.B. Farnsworth, B.W. Smith, N. Omenetto, 'Computer Modeling of Collection Efficiency of Laser-excited Fluorescence from a Graphite Furnace', *Spectrochim. Acta*, **45B**, 1151–1166 (1990).
241. N. Omenetto, J.D. Winefordner, 'Types of Fluorescence Transitions in Atomic Fluorescence Spectrometry', *Appl. Spectrosc.*, **26**, 555–557 (1972).
242. N. Omenetto, J.M. Mermet, G.C. Turk, D.S. Moore, 'Nomenclature, Symbols, Units and their Usage in Spectrochemical Analysis. XIV. Laser-based Atomic Spectroscopy: A New Notation for Spectrochemical Processes (IUPAC recommendations 1997)', *Pure Appl. Chem.*, **70**(2), 517–526 (1998).
243. C.T.J. Alkemade, 'Detection of Small Numbers of Atoms and Molecules', in *Analytical Applications of Lasers*, ed. E.H. Piepmeier, John Wiley & Sons, New York, 107–162, 1986.
244. C.T.J. Alkemade, 'Single-atom Detection', *Appl. Spectrosc.*, **35**, 1–14 (1981).
245. M.S. Epstein, J.D. Winefordner, 'Summary of the Usefulness of Signal-to-noise Treatment in Analytical Spectrometry', *Prog. Anal. Atom. Spectrosc.*, **7**, 67–137 (1984).
246. P.A. Bonczyk, J.A. Shirley, 'Measurement of CH and CN Concentration in Flames by Laser-induced Saturated Fluorescence', *Combust. Flame*, **34**, 253–264 (1979).
247. H. Haraguchi, S.J. Weeks, J.D. Winefordner, 'Selective Excitation of Molecular Species in Flames by Laser-excited Molecular Fluorescence', *Spectrochim. Acta*, **35A**, 391–399 (1979).
248. E.G. Su, R.L. Irwin, Z.W. Liang, R.G. Michel, 'Background Correction by Wavelength Modulation for Pulsed-laser-excited Atomic Fluorescence Spectrometry', *Anal. Chem.*, **64**(15), 1710–1720 (1992).
249. R.D. Reeves, B.M. Patel, C.T. Molnar, 'Decay of Atom Population Following Graphite Rod Atomization in Atomic Absorption Spectrometry', *Anal. Chem.*, **45**, 246–249 (1973).
250. J. Aggett, T.S. West, 'Atomic Absorption and Fluorescence Spectroscopy with a Carbon Filament Atom Reservoir', *Anal. Chim. Acta*, **55**, 349–357 (1971).
251. K. Dittrich, B. Hanisch, H.-J. Stärk, 'Molecule Formation in Electrothermal Atomizers: Interferences and Analytical Possibilities by Absorption, Emission and Fluorescence', *Fresenius' Z Anal. Chem.*, **324**, 497–506 (1986).
252. G. Zizak, 'Laser Diagnostics in Flames by Fluorescence Techniques', in *Analytical Laser Spectroscopy*, eds. S. Martellucci, A.N. Chester, Plenum Press, New York, 147–158, 1985.
253. N. Omenetto, P. Benetti, G. Rossi, 'Flame Temperature Measurements by Means of Atomic Fluorescence Spectrometry', *Spectrochim. Acta*, **27B**, 453–461 (1972).

254. J.D. Bradshaw, N. Omenetto, G. Zizak, J.N. Bower, J.D. Winefordner, 'Five Laser-excited Fluorescence Methods for Measuring Spatial Flame Temperatures. 1: Theoretical Basis', *Appl. Opt.*, **19**(16), 2709–2716 (1980).
255. G. Zizak, J.D. Winefordner, 'Application of the Thermally Assisted Atomic Fluorescence Technique to the Temperature Measurement in a Gasoline–Air Flame', *Combust. Flame*, **44**, 35–41 (1982).
256. R.G. Joklik, J.W. Daily, 'Two-line Atomic Fluorescence Temperature Measurement in Flames: An Experimental Study', *Appl. Opt.*, **21**(22), 4158–4162 (1982).
257. D.R. Crosley, 'Laser Probes for Combustion Chemistry', *ACS Symp. Ser.*, **134**, 3–18 (1980).
258. J.W. Daily, 'Pulsed Resonance Spectroscopy Applied to Turbulent Combustion Flows', *Appl. Opt.*, **15**, 955–960 (1976).
259. J.W. Daily, 'Saturation Effects in Laser-induced Fluorescence Spectroscopy', *Appl. Opt.*, **16**(3), 568–572 (1977).
260. H. Uchida, M.A. Kosinski, J.D. Winefordner, 'Laser-excited Atomic and Ionic Fluorescence in an Inductively Coupled Plasma', *Spectrochim. Acta*, **38B**, 5–13 (1983).
261. N.S. Ham, P. Hannaford, 'Direct Observation of Lifetimes of Excited Atoms in a Flame at Atmospheric Pressure', *J. Phys. B*, **12**, L199–L204 (1979).
262. C.H. Muller, III, K. Schofield, M. Steinberg, 'Laser-induced Flame Chemistry of Li ($2\ ^2P_{1/2,3/2}$) and Na ($3\ ^2P_{1/2,3/2}$). Implications for Other Saturated Mode Measurements', *J. Chem. Phys.*, **72**(12), 6620–6631 (1980).
263. C.H. Muller, III, K. Schofield, M. Steinberg, 'Near Saturation Laser-induced Chemical Reactions of Na ($3\ ^2P_{1/2,3/2}$) in $H_2/O_2/N_2$ Flames', *Chem. Phys. Lett.*, **72**(12), 6620–6631 (1980).
264. D.R. Crosley, 'Collisional Effects on Laser-induced Fluorescence Flame Measurements', *Opt. Eng.*, **20**, 511–521 (1981).
265. M. Iino, H. Yano, Y. Takubo, M. Shimazu, 'Saturation Characteristics of Laser-induced Na Fluorescence in a Propane–Air Flame: The Role of Chemical Reaction', *J. Appl. Phys.*, **52**, 6025–6031 (1981).
266. J.B. Simeonsson, B.W. Smith, J.D. Winefordner, N. Omenetto, 'On the Possibility of Observing Negative Fluorescence Dips in Laser-induced Fluorescence Experiments', *Appl. Spectrosc.*, **45**(4), 521–523 (1991).
267. J.B. Simeonsson, K.C. Ng, J.D. Winefordner, 'Fluorescence-dip Spectroscopy as a Method for Measuring the Quantum Efficiency of Fluorescence', *J. Quant. Spectrosc. Radiat. Transfer*, **48**(2), 131–139 (1992).
268. L. Xu, Y.Y. Zhao, G.Y. Wang, M.Q. He, Z.Y. Wang, 'Observation of Laser-enhanced Radiative-transfer of Barium $6s\,6p\ ^1P_1$ to $6s\,5d\ ^1D_2$ by Means of Fluorescence-dip Detection', *J. Opt. Soc. Am. B*, **9**(7), 1017–1019 (1992).
269. G.A. Petrucci, C.L. Stevenson, B.W. Smith, J.D. Winefordner, N. Omenetto, 'Experimental Evaluation of the Autoionization Cross-section of the Magnesium Transition at 300.9 nm by Laser-induced Fluorescence', *Spectrochim. Acta*, **46B**(6/7), 975–981 (1991).
270. N. Omenetto, O.I. Matveev, 'Time-resolved Fluorescence as a Direct Experimental Approach to the Study of Excitation and Ionization Processes in Different Atom Reservoirs', *Spectrochim. Acta*, **49B**(12–14), 1519–1535 (1994).
271. M.H. Nayfeh, 'Laser Detection of Single Atoms', *Am. Sci.*, **67**(2), 204–213 (1979).
272. G.S. Hurst, M.G. Payne, S.D. Kramer, C.H. Chen, 'Counting the Atoms', *Phys. Today*, **24 Sept.**, 24–29 (1980).
273. G.I. Bekov, V.S. Lethokhov, 'Laser Atomic Photoionization Spectral Analysis of Element Traces', *Appl. Phys. B*, **30**, 161–176 (1983).
274. J.D. Fassett, L.J. Moore, J.C. Travis, J.R. DeVoe, 'Laser Resonance Ionization Mass Spectrometry', *Science*, **230**, 263–267 (1985).
275. T. Whitaker, 'Isotopically Selective Laser Measurements', *Laser Appl.*, **5**(8), 67–73 (1986).
276. G.S. Hurst, M.G. Payne, 'Elemental Analysis Using Resonance Ionization Spectroscopy', *Spectrochim. Acta*, **42B**, 715–726 (1987).
277. J.D. Fassett, R.J. Walker, J.C. Travis, F.C. Ruegg, 'Measurement of Low Abundance Isotopes by Laser Resonance Ionization Mass Spectrometry (RIMS)', *Anal. Instrum.*, **17**(1/2), 69–86 (1988).
278. J.D. Fassett, J.C. Travis, 'Analytical Applications of Resonance Ionization Mass Spectrometry (RIMS)', *Spectrochim. Acta*, **43B**, 1409–1422 (1988).
279. J.P. Young, R.W. Shaw, D.H. Smith, 'Resonance Ionization Mass Spectrometry', *Anal. Chem.*, **61**(22), 1271A–1279A (1989).
280. G.I. Bekov, V.N. Radayev, V.S. Lethokhov, 'Laser Photoionization Spectroscopy of Atomic Traces at Part per Trillion Levels', *Spectrochim. Acta*, **43B**, 491–499 (1988).
281. T.B. Lucatorto, J.E. Parks (eds.), Proceedings of the 4th International Symposium on Resonance Ionization Spectroscopy and its Applications: *Resonance Ionization Spectroscopy, 1988*, Institute of Physics Conference Series Number 94, National Bureau of Standards, Gaithersburg, MD, April 10–15, 1988.
282. J.E. Parks, N. Omenetto (eds.), Proceedings of the 5th International Symposium on Resonance Ionization Spectroscopy and its Applications: *Resonance Ionization Spectroscopy, 1990*, Institute of Physics Conference Series Number 114, Congress Center Villa Ponti, Varese, Italy, Sept. 16–21, 1990.
283. C.M. Miller, J.E. Parks (eds.), Proceedings of the 6th International Symposium on Resonance Ionization Spectroscopy and its Applications: *Resonance Ionization Spectroscopy, 1992*, Institute of Physics Conference

Series Number 128, Los Alamos National Laboratory, Santa Fe, NM, May 24–29, 1992.
284. H.-J. Kluge, J.E. Parks, K. Wendt (eds.), Proceedings of the 7th International Symposium on Resonance Ionization Spectroscopy and its Applications: *Resonance Ionization Spectroscopy, 1994*, American Institute of Physics Conference Proceeding 329, Bernkasel-Kues, Germany, July 3–8, 1994.
285. N. Winograd, J.E. Parks (eds.), Proceedings of the 8th International Symposium on Resonance Ionization Spectroscopy and its Applications: *Resonance Ionization Spectroscopy, 1996*, American Institute of Physics Conference Proceeding 338, State College, PA, June 30–July 5, 1996.
286. R.B. Green, R.A. Keller, P.K. Schenck, J.C. Travis, G.G. Luther, 'Opto-galvanic Detection of Species in Flames', *J. Am. Chem. Soc.*, **98**, 8517–8518 (1976).
287. R.A. Keller, R. Engleman, Jr, E.F. Zalewski, 'Opto-galvanic Spectroscopy in a Uranium Hollow Cathode Discharge', *J. Opt. Soc. Am. B*, **69**(5), 738–742 (1979).
288. A. Rosenfeld, S. Mory, R. König, 'Observation of the Opto-galvanic Effect in Neon and Argon Using a Nanosecond Dye Laser', *Opt. Commun.*, **30**(3), 394–396 (1979).
289. D.S. King, P.K. Schenck, 'Optogalvanic Spectroscopy', *Laser Focus*, **14**, 50–57 (1978).
290. P.K. Schenck, W.C. Mallard, J.C. Travis, K.C. Smyth, 'Absorption Spectra of Metal Oxides Using Optogalvanic Spectroscopy', *J. Chem. Phys.*, **69**(11), 5147–5150 (1978).
291. D.S. King, P.K. Schenck, K.C. Smyth, J.C. Travis, 'Direct Calibration of Laser Wavelength and Bandwidth Using the Optogalvanic Effect', *Appl. Opt.*, **16**(10), 2617–2619 (1977).
292. J.C. Travis, P.K. Schenck, G.C. Turk, W.G. Mallard, 'Effect of Selective Laser Excitation on the Ionization of Atomic Species in Flames', *Anal. Chem.*, **51**(9), 1516–1520 (1979).
293. G.C. Turk, J.C. Travis, J.R. DeVoe, T.C. O'Haver, 'Laser-enhanced Ionization Spectrometry in Analytical Flames', *Anal. Chem.*, **51**, 1890–1896 (1979).
294. A.S. Gonchakov, N.B. Zorov, Y.Y. Kuzyakov, O.I. Matveev, 'Determination of Picogram Concentrations of Sodium in Flame by Stepwise Photoionization of Atoms', *Anal. Lett.*, **12**, 1037–1048 (1979).
295. F.M. Curran, K.C. Lin, G.E. Leroi, P.M. Hunt, S.R. Crouch, 'Energy Considerations in Dual Laser Ionization Processes in Flames', *Anal. Chem.*, **55**, 2382–2387 (1983).
296. F.M. Curran, C.A.v. Dijk, S.R. Crouch, 'Dual Laser Ionization in Flames: a Search for Electrical Interferences', *Appl. Spectrosc.*, **37**, 385–389 (1983).
297. Y.-Y.J. Wu, P.M. Hunt, G.E. Leroi, S.R. Crouch, 'Dual Laser Ionization (DLI) Measurements of Ion Mobilities in a H_2-O_2-Ar Flame', *Chem. Phys. Lett.*, **155**(1), 69–76 (1989).
298. C.A.v. Dijk, F.M. Curran, K.C. Lin, S.R. Crouch, 'Two-step Laser-assisted Ionization of Sodium in a Hydrogen–Oxygen–Argon flame', *Anal. Chem.*, **53**, 1275–1279 (1981).
299. M.F. Hineman, S.R. Crouch, 'Modeling Signals, Background and Noise in Flame Laser-induced Ionization Spectroscopy', *Spectrochim. Acta*, **43B**(9–11), 1119–1131 (1988).
300. M.E. Churchwell, T. Beeler, J.D. Messman, R.B. Green, 'Laser-induced Ionization in an Atmospheric-pressure Microarc-induced Plasma', *Spectrosc. Lett.*, **18**(9), 679–693 (1985).
301. A.S. Gonchakov, N.B. Zorov, Y.Y. Kuzyakov, O.I. Matveev, 'Detection of Subpicograms of Sodium by Laser Step-photoionization of Atoms and Laser Atomic Fluorescence with Electrothermal Atomization', *J. Anal. Chem. (USSR)*, **34**, 1792–1795 (1980).
302. J.C. Travis, G.C. Turk, 'Preface', in *Laser-enhanced Ionization Spectrometry*, eds. J.C. Travis, G.C. Turk, John Wiley & Sons, New York, xv–xvi, 1996.
303. G.A. Petrucci, J.D. Winefordner, 'The Double-resonance Optogalvanic Effect of Neon as a Sensitive Photon Detector', *Spectrochim. Acta*, **47B**(3), 437–447 (1992).
304. C.R. Webster, C.T. Rettner, 'Laser Optogalvanic Spectroscopy of Molecules', *Laser Focus*, **19**, 41–53 (1983).
305. P.K. Schenck, J.W. Hastie, 'Optogalvanic Spectroscopy – Application to Combustion Systems', *Opt. Eng.*, **20**, 522–528 (1981).
306. J.C. Travis, G.C. Turk, R.B. Green, 'Laser-enhanced Ionization for Trace Metal Analysis in Flames', *ACS Symp. Ser.*, **85**, 91–101 (1978).
307. J.C. Travis, G.C. Turk, J.R. DeVoe, P.K. Schenck, C.A.v. Dijk, 'Principles of Laser-enhanced Ionization Spectrometry in Flames', *Prog. Anal. At. Spectrosc.*, **7**, 199–241 (1984).
308. O. Axner, 'Laser-enhanced Ionization Spectrometry and its Principles – a Powerful Tool for Ultra-sensitive Trace-element Analysis', PhD Thesis, Department of Physics, Chalmers University of Technology, Göteborg, Sweden, 1987.
309. O. Axner, H. Rubinsztein-Dunlop, 'Laser-enhanced Ionization Spectrometry in Flames – a Powerful and Versatile Technique for Ultra-sensitive Trace Element Analysis', *Spectrochim. Acta*, **44B**, 835–866 (1989).
310. O. Axner, S. Sjöström, H. Rubinsztein-Dunlop, 'Analytical Applications of LEI Spectrometry in Flames and Furnaces', *Mikrochim. Acta*, **III**, 197–214 (1989).
311. K. Niemax, 'LIF, LEI, RIMS etc. – New Promising Techniques in Elemental Micro- and Trace Analysis', *Fresenius' J. Anal. Chem.*, **337**, 551–556 (1990).
312. G.C. Turk, 'Laser-enhanced Ionization Spectroscopy in Flames and Plasmas – Plenary Lecture', *J. Anal. At. Spectrom.*, **2**, 573–577 (1987).
313. Y.Y. Kuzyakov, N.B. Zorov, 'Atomic Ionization Spectrometry: Prospects and Results', *Crit. Rev. Anal. Chem.*, **20**(4), 221–290 (1988).

314. R.B. Green, M.D. Seltzer, 'Laser-induced Ionization Spectrometry', in *Advances in Atomic Spectrometry*, ed. J. Sneddon, JAI Press, Greenwich, CT, 37–79, 1992.
315. D.J. Butcher, 'Laser-enhanced Ionization Spectrometry', in *Lasers in Analytical Spectroscopy*, eds. J. Sneddon, T.L. Thiem, Y.-I. Lee, VCH Publishers, New York, 237–272, 1997.
316. J.C. Travis, G.C. Turk, 'Fundamental Mechanisms of Laser-enhanced Ionization: Signal Detection', in *Laser-enhanced Ionization Spectrometry*, eds. J.C. Travis, G.C. Turk, John Wiley & Sons, New York, 99–160, 1996.
317. R.B. Green, 'Applications of Laser-enhanced Ionization', in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, John Wiley & Sons, New York, 213–232, 1996.
318. N.B. Zorov, 'Nonflame Reservoirs for Laser-enhanced Ionization Spectrometry', in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, John Wiley & Sons, New York, 233–264, 1996.
319. N. Omenetto, P.B. Farnsworth, 'Ions and Photons: Interplay of Laser-induced Ionization and Fluorescence Techniques in Different Atomic and Molecular Reservoirs', in *Laser-enhanced Ionization Spectrometry*, eds. J. Travis, G. Turk, John Wiley & Sons, New York, 265–326, 1996.
320. R.B. Green, 'Laser-enhanced Ionization in Flames', in *Analytical Applications of Lasers*, ed. E.H. Piepmeier, John Wiley & Sons, New York, 75–105, 1986.
321. J.D. Messman, N.E. Schmidt, J.D. Parli, R.B. Green, 'Laser-enhanced Ionization of Refractory Elements in a Nitrous Oxide–Acetylene Flame', *Appl. Spectrosc.*, **39**(3), 504–507 (1985).
322. V.I. Chaplygin, Y.Y. Kuzyakov, O.A. Novodvorsky, N.B. Zorov, 'Determination of Alkali Metals by Laser-induced Atomic-ionization in Flames', *Talanta*, **34**(1), 191–196 (1987).
323. G.C. Turk, R.L. Watters, Jr, 'Resonant Laser-induced Ionization of Atoms in an Inductively Coupled Plasma', *Anal. Chem.*, **57**, 1979–1983 (1985).
324. G.C. Turk, L.J. Yu, R.L. Watters, J.C. Travis, 'Laser-induced Ionization of Atoms in a Power-modulated Inductively Coupled Plasma', *Appl. Spectrosc.*, **46**(8), 1217–1222 (1992).
325. L.J. Yu, S.R. Koirtyohann, G.C. Turk, M.L. Salit, 'Selective Laser-induced Ionization in Inductively-coupled Plasma-mass Spectrometry', *J. Anal. Atom. Spectrom.*, **9**(9), 997–1000 (1994).
326. M.D. Seltzer, R.B. Green, 'An Active Nitrogen Plasma Atom Reservoir for Laser-induced Ionization Spectrometry', *Spectrosc. Lett.*, **20**, 601–617 (1987).
327. G.C. Turk, O. Axner, N. Omenetto, 'Optical Detection of Laser-induced Ionization in the Inductively Coupled Plasma for the Study of Ion–Electron Recombination and Ionization Equilibrium', *Spectrochim. Acta*, **42B**, 873–881 (1987).
328. I.V. Bykov, A.B. Skvortsov, T.G. Tatsii, N.V. Chekalin, 'Metal Trace Analysis by Flame/Graphite Furnace OG Spectroscopy', *J. Phys. Colloq.*, **44**(C7), 345–352 (1983).
329. I. Magnusson, O. Axner, I. Lindgren, H. Rubinsztein-Dunlop, 'Laser-enhanced Ionization Detection of Trace Elements in a Graphite Furnace', *Appl. Spectrosc.*, **40**(7), 968–971 (1986).
330. I. Magnusson, S. Sjöström, M. Lejon, H. Rubinsztein-Dunlop, 'Trace Element Analysis by Two-color Laser-enhanced Ionization Spectrometry in a Graphite Furnace', *Spectrochim. Acta*, **42B**(5), 713–718 (1986).
331. I. Magnusson, 'The Applicability to Trace Element Analysis of Laser-enhanced Ionization Spectroscopy in a Graphite Furnace', *Spectrochim. Acta*, **43B**, 727–735 (1988).
332. S. Sjöström, I. Magnusson, M. Lejon, H. Rubinsztein-Dunlop, 'Laser-enhanced Ionization Spectrometry in a T-furnace', *Anal. Chem.*, **60**(15), 1629–1631 (1988).
333. N.V. Chekalin, I.I. Vlasov, 'Direct Analysis of Liquid and Solid Samples without Sample Preparation Using Laser-enhanced Ionization', *J. Anal. At. Spectrom.*, **7**(2), 225–228 (1992).
334. D.J. Butcher, R.L. Irwin, S. Sjöström, A.P. Walton, R.G. Michel, 'Probe Atomization for Laser-enhanced Ionization in a Graphite Tube Furnace', *Spectrochim. Acta*, **46B**(1), 9–33 (1991).
335. N.V. Chekalin, V.I. Pavlutskaya, I.I. Vlasov, 'A ''Rod Flame'' System in Direct Atomic-ionization Analysis of High-purity Substances', *Inst. Phys. Conf. Ser.*, **114**, 283–288 (1990).
336. N.V. Chekalin, V.I. Pavlutskaya, I.I. Vlasov, 'A ''Rod Flame'' System in Direct Laser-enhanced Ionization Analysis of High-purity Substances', *Spectrochim. Acta*, **46B**(13), 1701–1709 (1991).
337. N.V. Chekalin, A.G. Marunkov, I.I. Vlasov, A.T. Khalmanov, 'Multipurpose Atomic-ionization Spectrometer – Analysis of High-purity Substances', *High Energy Chem.*, **28**(5), 412–416 (1994).
338. N.V. Chekalin, A. Khalmanov, A.G. Marunkov, I.I. Vlasov, Y. Malmsten, O. Axner, V.S. Dorofeev, E. Glukhan, 'Determination of Co, Cr, Mn and Ni Traces in Fluorine-containing Materials for Optical Fibers Using Laser-enhanced Ionization Techniques with Flame and Rod–Flame Atomizers', *Spectrochim. Acta*, **50B**(8), 753–761 (1995).
339. B.W. Smith, G.A. Petrucci, R.G. Badini, J.D. Winefordner, 'Graphite-furnace Vaporization with Laser-enhanced Ionization Detection', *Anal. Chem.*, **65**(2), 118–122 (1993).
340. K.L. Riter, W.L. Clevenger, L.S. Mordoh, B.W. Smith, O.I. Matveev, J.D. Winefordner, 'Laser-enhanced Ionization Detection of Magnesium Atoms by a Combination of Electrothermal Vaporization and Flame Atomization', *J. Anal. At. Spectrosc.*, **11**(6), 393–399 (1996).
341. N. Omenetto, O.I. Matveev, W. Resto, R. Badini, B.W. Smith, J.D. Winefordner, 'Nonlinear Behavior of Atomic

Fluorescence in Mercury Vapors Following Double-resonance Laser Excitation', *Appl. Phys. B*, **58**(4), 303–307 (1994).
342. O.I. Matveev, B.W. Smith, N. Omenetto, J.D. Winefordner, 'Single Photo-electron and Photon Detection in a Mercury Resonance Ionization Photon Detector (RID)', *Spectrochim. Acta*, **51B**(6), 563–567 (1996).
343. W.L. Clevenger, L.S. Mordoh, O.I. Matveev, N. Omenetto, B.W. Smith, J.D. Winefordner, 'Analytical Time-resolved Laser-enhanced Ionization Spectroscopy. 1. Collisional Ionization and Photoionization of the Hg Rydberg States in a Low Pressure Gas', *Spectrochim. Acta*, **52B**(3), 295–304 (1997).
344. W.L. Clevenger, O.I. Matveev, N. Omenetto, B.W. Smith, J.D. Winefordner, 'Laser-enhanced Ionization Spectroscopy of Mercury Rydberg States', *Spectrochim. Acta*, **52B**(8), 1139–1149 (1997).
345. W.L. Clevenger, O.I. Matveev, S. Cabredo, N. Omenetto, B.W. Smith, J.D. Winefordner, 'Laser-enhanced Ionization of Mercury Atoms in an Inert Atmosphere with Avalanche Amplification of the Signal', *Anal. Chem.*, **69**(13), 2232–2237 (1997).
346. W.L. Clevenger, B.W. Smith, J.D. Winefordner, 'Trace Determination of Mercury: a Review', *Crit. Rev. Anal. Chem.*, **27**(1), 1–26 (1997).
347. K. Niemax, 'Spectroscopy Using Thermionic Diode Detectors', *Appl. Phys. B*, **38**, 147–157 (1985).
348. K. Niemax, C.J. Lorenzen, 'The Thermionic Double-diode, an Efficient Detector in Excited-state Laser Spectroscopy', *Opt. Commun.*, **44**(3), 165–169 (1983).
349. K. Niemax, J. Lawrenz, A. Obrebski, K.-H. Weber, 'Isotope-selective Trace-element Detection with the Thermionic Diode', *Anal. Chem.*, **58**, 1566–1571 (1986).
350. K. Niemax, 'Investigations on the Thermionic Diode: The Ionization Probability of Rb ^2P_J Atoms by Noble Gas Collisions', *Appl. Phys. B*, **32**, 59–62 (1983).
351. A. Obrebski, R. Hergenröder, K. Niemax, 'Narrowing of Spectral Lines by Energy Pooling Collisions in Laser-enhanced Ionization Spectroscopy', *Z. Phys. D*, **14**, 289–292 (1989).
352. J. Lawrenz, A. Obreski, K. Niemax, 'Measurement of Isotope Ratio by Doppler-free Laser Spectroscopy Applying Semiconductor Diode Laser and Thermionic Diode Detection', *Anal. Chem.*, **59**, 1232–1236 (1987).
353. J. Brust, D. Veza, M. Movre, K. Niemax, 'Collisional Excitation Transfer between Lithium Isotopes', *Z. Phys. D*, **32**(4), 305–309 (1995).
354. J. Franzke, D. Veza, K. Niemax, 'An Improved Thermionic Diode Detector for Analytical Laser Spectroscopy', *Spectrochim. Acta*, **47B**(5), 593–599 (1992).
355. V. Horvatic, D. Veza, M. Movre, K. Niemax, C. Vadla, 'Collision Cross-sections for Excitation-energy Transfer in $Na^*(^3P_{1/2}) + K(^4S_{1/2})$ to $Na^*(^3P_{3/2})$', *Z. Phys. D*, **34**(3), 163–170 (1995).
356. V. Horvatic, C. Vadla, M. Movre, K. Niemax, 'The Collision Cross-sections for the Fine-structure Mixing of Caesium 6P Levels Induced by Collisions with Potassium Atoms', *Z. Phys. D*, **36**(2), 101–104 (1996).
357. C. Vadla, D. Veza, M. Movre, K. Niemax, 'Fine-structure Excitation Transfer between the Lithium d-lines by Collisions with Cesium Atoms', *Z. Phys. D*, **22**(3), 591–595 (1992).
358. D. Veza, C. Vadla, K. Niemax, 'Excitation-energy Transfer in the Li–Cs Collision $Li^*(2p) + Cs^-(6s) \longrightarrow Li(2s) + Cs^{-*}(5d)$', *Z. Phys. D*, **22**(3), 597–601 (1992).
359. H.D. Wizemann, K. Niemax, 'Isotope Selective Element Analysis by Diode Laser Atomic Absorption Spectrometry', *Mikrochim. Acta*, **129**(3/4), 209–216 (1998).
360. C. Vadla, K. Niemax, V. Horvatic, R. Beuc, 'Population and Deactivation of Lowest Lying Barium Levels by Collisions with He, Ar, Xe and Pa Ground-state Atoms', *Z. Phys. D*, **34**(3), 171–184 (1995).
361. C. Vadla, K. Niemax, J. Brust, 'Energy Pooling in Cesium Vapor', *Z. Phys. D*, **37**(3), 241–247 (1996).
362. C. Vadla, K. Niemax, V. Horvatic, 'Energy Pooling to the Ba 6s 6p $^1P_1^0$ Level Arising from Collisions between Pairs of Metastable Ba 6s 5d ^3D_J Atoms', *Z. Phys. D*, **1**(2), 139–147 (1998).
363. J. Brust, M. Movre, K. Niemax, 'Measurement and Calculation of the Fine-structure Changing Collision Cross-sections in the Mg^+ and Ca^+ Resonance States by Helium', *Z. Phys. D*, **27**(3), 243–248 (1993).
364. O. Axner, I. Magnusson, J. Petersson, S. Sjöström, 'Investigation of the Multi-element Capability of Laser-enhanced Ionization Spectrometry in Flames for Analysis of Trace Elements in Water Solutions', *Appl. Spectrosc.*, **41**, 19–26 (1987).
365. J.E. Hall, R.B. Green, 'Laser-enhanced Ionization Spectrometry with a Total Consumption Burner', *Anal. Chem.*, **55**, 1811–1814 (1983).
366. O. Axner, I. Lindgren, I. Magnusson, H. Rubinsztein-Dunlop, 'Trace Element Determination in Flames by Laser-enhanced Ionization Spectrometry', *Anal. Chem.*, **57**, 773–776 (1985).
367. O.I. Matveev, P. Cavalli, N. Omenetto, 'Three-step Laser-induced Ionization of Ir and Hg Atoms in an Air–Acetylene Flame and a Gas Cell', in *Resonance Ionization Spectroscopy 1994*, American Institute of Physics, New York, 269–272, 1994.
368. D.J. Ehrlich, R.M. Osgood, Jr, G.C. Turk, J.C. Travis, 'Atomic Resonance-line Lasers for Atomic Spectrometry', *Anal. Chem.*, **52**, 1354–1356 (1980).
369. L.P. Hart, B.W. Smith, N. Omenetto, 'Laser-induced Stepwise and Two-photon Ionization Studies of Strontium in the Air Acetylene Flame', *Spectrochim. Acta*, **40B**, 1637–1649 (1985).
370. N. Omenetto, T. Berthoud, P. Cavalli, G. Rossi, 'Analytical Laser-enhanced Ionization Studies of Thallium in the Air–Acetylene Flame', *Anal. Chem.*, **57**, 1256–1261 (1985).

371. O. Axner, I. Magnusson, 'Determination of Trace Elements in Water Solution by Laser-enhanced Ionization Using Coumarin 47', *Phys. Scr.*, **31**, 587–591 (1985).
372. O. Axner, T. Berglind, S. Sjöström, 'Laser-enhanced Ionization Spectroscopy Around the Ionization Limit', *Phys. Scr.*, **34**, 18–23 (1986).
373. O. Axner, T. Berglind, 'Stark Structure Observation in Rydberg States of Li in Flames by Laser-enhanced Ionization – a New Method for Probing Local Electrical Fields in Flames', *Appl. Spectrosc.*, **40**, 1224–1231 (1986).
374. O. Axner, M. Lejon, I. Magnusson, H. Rubinsztein-Dunlop, S. Sjöström, 'Detection of Traces in Semiconductor Materials by Two-color Laser-enhanced Ionization Spectroscopy in Flames', *Appl. Opt.*, **26**, 3521–3525 (1987).
375. O. Axner, S. Sjöström, 'Anomalous Contributions to the Signal in Laser-enhanced Ionization Spectrometry from ''Scattered'' Laser Light', *Appl. Spectrosc.*, **44**(5), 864–870 (1990).
376. O. Axner, S. Sjöström, 'Anomalous Lineshapes and Signals in Laser-enhanced Ionization Spectrometry in Flames – The Role of ''Scattered'' Light, Dynamic Stark Effects and Two-step versus Two-photon Excitations', *Inst. Phys. Conf. Ser.*, **114**(6), 31–36 (1990).
377. O. Axner, M. Norberg, H.R. Dunlop, 'Investigation of Background Signals from Na as a Source of Interference in Laser-enhanced Ionization Spectrometry in Flames', *Appl. Spectrosc.*, **44**, 1124–1133 (1990).
378. P. Ljungberg, Y. Malmsten, O. Axner, 'Monitoring of Atomic Metastable State Lifetimes by the Laser-enhanced Ionization Technique – Development of a Method for Probing Local Stoichiometric Combustive Conditions', in *Resonance Ionization Spectroscopy 1994*, American Institute of Physics, New York, 273–276, 1994.
379. I. Magnusson, O. Axner, H. Rubinsztein-Dunlop, 'Elimination of Spectral Interferences Using Two-step Excitation Laser-enhanced Ionization', *Phys. Scr.*, **33**, 429–433 (1986).
380. G. Nave, S. Johansson, O. Axner, P. Ljungberg, Y. Malmsten, B. Baschek, 'Analysis of the $3d^6$ 4s (^{6}D) 6d Subconfiguration of Fe I by Laser-enhanced Ionization and Grating Spectroscopy', *Phys. Scr.*, **49**, 581–587 (1994).
381. P.M. Paquet, J.-F. Gravel, P. Norbert, D. Boudreau, 'Speciation of Chromium by Ion Chromatography and Laser-enhanced Ionization: Optimization of the Excitation–Ionization Scheme', *Spectrochim. Acta*, **53**(12), 1907–1917 (1998).
382. G.J. Havrilla, C.C. Carter, 'Laser-enhanced Ionization Detection of Trace Copper in High Salt Matrices', *Appl. Opt.*, **17**, 3511–3515 (1987).
383. G.C. Turk, J.R. DeVoe, J.C. Travis, 'Stepwise Excitation Laser-enhanced Ionization Spectrometry', *Anal. Chem.*, **54**, 643–645 (1982).
384. G.C. Turk, J.C. Travis, J.R. DeVoe, 'Laser-enhanced Ionization Spectrometry for Trace Metal Analysis', *J. Phys. Colloq.*, **44**(11 : C7), 301–309 (1983).
385. G.J. Havrilla, K.-J. Choi, 'Detection and Spectroscopic Study of Zinc by Laser-enhanced Ionization Spectrometry', *Anal. Chem.*, **58**, 3095–3100 (1986).
386. M.D. Seltzer, R.B. Green, 'Direct Laser Ionization in Analytically Useful Flames', *Appl. Spectrosc.*, **43**, 257–263 (1989).
387. Y.Y. Kuzyakov, N.B. Zorov, V.I. Chaplygin, A.A. Gorbatenko, 'The Role of Resonance Photoionization of Atoms in Seeded Flames', in *Resonance Ionization Spectroscopy 1988*, IOP Publishing, Bristol, 179–182, 1988.
388. V.I. Chaplygin, N.B. Zorov, Y.Y. Kuzyakov, 'Laser Atomic-ionization Determination of Caesium in Flames', *Talanta*, **30**, 505–508 (1983).
389. G.C. Turk, J.C. Travis, J.R. DeVoe, T.C. O'Haver, 'Analytical Flame Spectrometry with Laser-enhanced Ionization', *Anal. Chem.*, **50**, 817–820 (1978).
390. O. Axner, T. Berglind, J.L. Heully, I. Lindgren, H. Rubinsztein-Dunlop, 'Theory of Laser-enhanced Ionization in Flames – Comparison with Experiments', *J. Phys. Colloq.*, **44**(C7), 311–317 (1983).
391. R.W. Fox, C.S. Weimer, L. Hollberg, G.C. Turk, 'The Diode-laser as a Spectroscopic Tool', *Spectrochim. Acta Rev.*, **15**(5), 291–299 (1993).
392. G.J. Havrilla, S.J. Weeks, J.C. Travis, 'Continuous Wave Excitation in Laser-enhanced Ionization Spectrometry', *Anal. Chem.*, **54**, 2566–2570 (1982).
393. P.K. Schenck, J.C. Travis, G.C. Turk, 'Studies of Physical Mechanisms in Laser-enhanced Ionization in Flames', *J. Phys. (Paris) Colloq.*, **C7**, 75–84 (1983).
394. O. Axner, S. Sjöström, 'Anomalous Contributions to the Signal in Laser-enhanced Ionization Spectrometry from ''Scattered'' Laser Light', *Appl. Spectrosc.*, **44**, 864–870 (1990).
395. G.C. Turk, J.C. Travis, 'Simultaneous Detection of Laser-enhanced Ionization and Laser-induced Fluorescence in Flames: Noise Correlation Studies', *Spectrochim. Acta*, **45B**, 409–419 (1990).
396. A.A. Gorbatenko, N.B. Zorov, S.Y. Karopova, Y.Y. Kuzyakov, V.I. Chapplygin, 'Determination of Trace Amounts of Calcium by Laser-enhanced Ionization Spectrometry in Flames', *J. Anal. At. Spectrom.*, **3**, 527–530 (1988).
397. N.B. Zorov, Y.Y. Kuzyakov, O.I. Matveev, 'Atomic Ionization Analysis by Tunable Lasers', *Zh. Anal. Khim.*, **37**(3), 520–533 (1982).
398. A.G. Marunkov, N.V. Chekalin, 'The Limiting Performance of a Flame Atomic-ionization Spectrometer', *J. Anal. Chem. USSR (Engl. Transl.)*, **42**, 506–508 (1987).
399. L.C. Chandola, P.P. Khanna, M.A.N. Razvi, 'Sub-picogram Detection of Sodium by One-step Laser-enhanced Ionization (LEI) Spectrometry Using Two-photon Transition', *Anal. Lett.*, **24**(9), 1685–1693 (1991).
400. K.L. Riter, O.I. Matveev, B.W. Smith, J.D. Winefordner, 'The Determination of Lead in Whole Blood by Laser-enhanced Ionization Using a Combination of

Electrothermal Vaporizer and Flame', *Anal. Chim. Acta*, **333**(1/2), 187–192 (1996).
401. I. Magnusson, 'On the Signal Collection in Laser-enhanced Ionization Spectrometry', *Spectrochim. Acta*, **42B**, 1113–1124 (1987).
402. S.C. Wang, K.C. Lin, 'Automatic-determination of Optimum Dilution Levels for Laser-enhanced Ionization Detection of Matrix-interfered Sample by Flow-injection', *Analyst*, **120**(10), 2593–2599 (1995).
403. S.C. Wang, K.C. Lin, 'Novel Technique to Reduce Electrical Interference Inherent in Laser-enhanced Ionization Detection by Using Flow-injection Analysis', *Anal. Chem.*, **66**(13), 2180–2186 (1994).
404. N.V. Chekalin, A.G. Marunkov, V.I. Pavlutskaya, S.V. Bachin, 'Determination of Traces of Cs, Li and Rb in Rocks by Laser-enhanced Ionization Spectrometry without Preconcentration', *Spectrochim. Acta*, **46B**, 551–558 (1991).
405. K.S. Epler, T.C. O'Haver, G.C. Turk, 'Liquid-chromatography Laser-enhanced Ionization Spectrometry for the Speciation of Organolead Compounds', *J. Anal. At. Spectrom.*, **9**(2), 79–82 (1994).
406. G.A. Petrucci, R.G. Badini, J.D. Winefordner, 'Photon Detection based on Pulsed Laser-enhanced Ionization and Photoionization of Magnesium Vapor – Experimental Characterization', *J. Anal. At. Spectrom.*, **7**(3), 481–491 (1992).
407. G.A. Petrucci, R.G. Badini, J.D. Winefordner, 'Resonance Detection of Photons by Atomic Ionization', *Inst. Phys. Conf. Ser.*, **128**, 333–336 (1992).
408. G.A. Petrucci, D. Imbroisi, B.W. Smith, J.D. Winefordner, 'Detection of OH in an Atmospheric-pressure Flame via Laser-enhanced Ionization of Indium', *Spectrochim. Acta*, **49B**(12–14), 1569–1578 (1994).
409. G.A. Petrucci, D. Imbroisi, R.D. Guenard, B.W. Smith, J.D. Winefordner, 'High-spatial-resolution OH Rotational Temperature-measurements in an Atmospheric-pressure Flame Using an Indium-based Resonance Ionization Detector', *Appl. Spectrosc.*, **49**(5), 655–659 (1995).
410. G.A. Petrucci, J.D. Winefordner, N. Omenetto, 'The Relaxation Oscillator as a Resonance Photon Detector', *Appl. Phys. B*, **62**(5), 457–464 (1996).
411. B.W. Smith, N. Omenetto, J.D. Winefordner, 'A New, Sensitive, High Resolution Raman Detector Based on Ionization', in *Future Trends in Spectroscopy*, Pergamon Press, Oxford, and Pontificia Academica Scientiarum, The Vatican, 101–111, 1989.
412. N. Omenetto, B.W. Smith, J.D. Winefordner, 'Theoretical Considerations on the Optogalvanic Detection of Laser-induced Fluorescence in Atmospheric Pressure Atomizers', in *Future Trends in Spectroscopy*, Pergamon Press Oxford and Pontificia Academica Scientiarum, The Vatican, 91–99, 1989.
413. D.L. Monts, Abhilasha, S.C. Qian, D. Kumar, X. Yao, S.P. McGlynn, 'Comparison of Atomization Sources for a Field-deployable Laser Optogalvanic Spectrometry System', *J. Thermophys. Heat Transfer*, **12**(1), 66–72 (1998).
414a. P.B. Farnsworth, N. Omenetto, 'The Kinetics of Charge-exchange between Argon and Magnesium in the Inductively-coupled Plasma', *Spectrochim. Acta*, **48B**(6/7), 809–816 (1993).
414b. G.C. Turk, 'Imaging the Active Flame Volume for Pulsed Laser-enhanced Ionization Spectroscopy', *Anal. Chem.*, **64**(17), 1836–1839 (1992).
414c. G.C. Turk, 'Spatially Selective Detection of Anomalous Contributions to Laser-enhanced Ionization in Flames', *Inst. Phys. Conf. Ser.*, **128**(4), 157–160 (1992).
415. I.I. Vlasov, N.V. Chekalin, 'A New Approach to the Determination of the Ionization Yield of Atoms by Laser-enhanced Ionization', *Spectrochim. Acta*, **48B**(4), 597–603 (1993).
416. K.C. Lin, 'Applications of Laser-enhanced Ionization in Analytical Chemistry', *J. Chin. Chem. Soc.*, **41**(3), 293–308 (1994).
417. K.D. Su, K.C. Lin, W.T. Luh, 'Application of Laser-enhanced Ionization–Atomization Efficiency Determination', *Appl. Spectrosc.*, **46**(9), 1370–1375 (1992).
418. O.I. Matveev, N. Omenetto, 'On the Possibility of Direct Evaluation of Ionization Rates Induced by Laser-radiation in Flames: Part 1 – Time-resolved Ionization Measurements', *Spectrochim. Acta*, **49B**(7), 691–702 (1994).
419. C.B. Ke, K.C. Lin, 'Spatially Resolved Temperature Determination of an Air/Acetylene Flame Using the Two-step Laser-enhanced Ionization Technique', *Appl. Spectrosc.*, **52**(2), 187–194 (1998).
420. P. Barker, A. Thomas, H. Rubinsztein-Dunlop, P. Ljungberg, 'Velocity Measurements by Flow Tagging Employing Laser-enhanced Ionization and Laser-induced Fluorescence', *Spectrochim. Acta*, **50B**(11), 1301–1310 (1995).
421. P. Barker, H. Rubinsztein-Dunlop, 'Measurements of Neutral Atom Diffusion and Electron–Ion Recombination by Laser-enhanced Ionization and Planar Laser-induced Fluorescence in an Air–Acetylene Flame', *Spectrochim. Acta*, **52B**(4), 459–469 (1997).
422. P. Barker, A. Bishop, H. Rubinsztein-Dunlop, 'Supersonic Velocimetry in a Shock Tube Using Laser-enhanced Ionization and Planar Laser-induced Fluorescence', *Appl. Phys. B*, **64**(3), 369–376 (1997).
423. P.F. Barker, A.M. Thomas, T.J. McIntyre, H. Rubinsztein-Dunlop, 'Velocimetry and Thermometry of Supersonic Flow around a Cylindrical Body', *AIAA J.*, **36**(6), 1055–1060 (1998).
424. G.C. Turk, L.J. Yu, S.R. Koirtyohann, 'Laser-enhanced Ionization Spectroscopy of Sodium Atoms in an Air Hydrogen Flame with Mass-spectrometric Detection', *Spectrochim. Acta*, **49B**(12–14), 1537–1543 (1994).

425. N. Trautmann, 'Ultratrace Analysis of Long-lived Radioisotopes in the Environment', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 209–214, 1994.
426. K. Wendt, B.A. Bushaw, G. Bhowmick, V.A. Bystrow, N. Kotovski, J.V. Kratz, J. Lantzsch, P. Muller, W. Nörtershäuser, E.W. Otten, A. Seibert, N. Trautmann, A. Waldek, Y. Yushkevich, '^{89,90}Sr-determination in Various Environmental Samples by Collinear Resonance Ionization Spectroscopy', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 361–364, 1996.
427. B.A. Bushaw, F. Juston, W. Nörtershäuser, N. Trautmann, P. Voss-de-Haan, K. Wendt, 'Multiple Resonance RIMS Measurements of Calcium Isotopes Using Diode Lasers', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 115–118, 1996.
428. M.I.K. Santala, H.M. Lauranto, R.R.E. Salomaa, 'RIS and Laser-induced Fluorescence of Nb from Laser-ablated Metal Surfaces', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 191–194, 1996.
429. J.E. Anderson, T.M. Allen, A.W. Garret, C.G. Gill, P.H. Hemberg, P.B. Kelly, N.S. Nogar, 'Resonant Laser Ablation: Mechanisms and Applications', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 195–198, 1996.
430. H.F. Arlinghaus, X.Q. Guo, T.J. Whitaker, M.N. Kwoka, 'Fish & Chips: Analytical Applications of Resonance Ionization Mass Spectrometry', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 139–144, 1996.
431. G.I. Bekov, 'Resonance Ionization Spectroscopy', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 9–14, 1994.
432. G. Passler, 'Application of Resonance Ionization Mass Spectrometry for Trace Analysis and in Fundamental Research', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 209–214, 1996.
433. B.A. Bushaw, H.-J. Kluge, J. Lantsch, R. Schwalbach, M. Schwarz, J. Stenner, H. Stevens, K. Wendt, K. Zimmer, 'Hyperfine Structure of ^{87,89}Sr 5s 4d ^{3}D–5s *n*f Transitions in Collinear Fast Beam RIMS', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 9–14, 1994.
434. S. Köhler, F. Albus, R. Deissenberger, N. Erdmann, H. Funk, H.-U. Hasse, G. Herrmann, G. Huber, H.-J.N. Kluge, G. Passler, P.M. Rao, J. Riegel, N. Trautmann, F.-J. Urban, 'Determination of the First Ionization Potential of Actinides by Resonance Ionization Mass Spectrometry', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 377–380, 1994.
435. S. Cavalieri, R. Eramo, L. Fini, M. Matera, 'Experimental Studies of Laser-induced Continuum Structure in the Atomic Ionization Continuum', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 165–170, 1994.
436. S.V. Bobashev, N.A. Cherepkov, A.Y. Elizarov, V.K. Prilipko, V.V. Korshunov, 'Linear and Circular Dichroism in Two-step Photoionization of Barium Atoms', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 399–402, 1994.
437. V.V. Petrunin, H.H. Andersen, P. Balling, P. Kristensen, T. Andersen, 'Resonance Ionization Spectroscopy of Negative Ions', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 103–108, 1996.
438. D. Hanstorp, U. Ljungblad, U. Berzinsh, D.J. Pegg, 'Observation of Doubly Excited States in Negative Ions Using Resonance Ionization Spectroscopy', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 97–102, 1996.
439. S.D. Bergeson, A. Balakrishnan, K.G.H. Baldwin, T.B. Lucatorto, J.P. Marangos, T.J. McIlrath, T.R. O'Brian, S.L. Rolston, N. Vansteenkiste, 'Doppler-free Resonance Ionization Spectroscopy of the He $1s^2\,{}^1S-1s\,2s\,{}^1S$ Transition at 120.3 nm', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 109–112, 1996.
440. H. Walther, 'Laser-experiments with Single Atoms and the Test of Basic Quantum Phenomena', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 3–12, 1996.
441. D.-Y. Jeong, K.S. Lee, A.S. Choe, J. Lee, B.K. Rhee, 'Dynamic Stark Shift in a 3-Level Atomic System', in *8th International Symposium on Resonance Ionization*

Spectroscopy and its Applications, American Institute of Physics Conference Proceeding 388, State College, PA, 89–92, 1996.

442. M.G. Payne, L. Deng, J.Y. Zhang, W.R. Garrett, 'Effect of Pressure Dependent Quantum Interference on the AC Stark Shifting of a Four-photon Resonance', in *7th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 329, Bernkastel-Kues, Germany, 373–376, 1994.
443. M. Shapiro, P. Brumer, 'Quantum Control of Dynamics', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 69–88, 1996.
444. B.A. Bushaw, 'Attogram Measurements of Rare Isotopes by CW Resonance Ionization Mass Spectrometry', in *6th International Symposium on Resonance Ionization Spectroscopy and its Applications*, Institute of Physics Conference Series Number 128, Los Alamos National Laboratory, Santa Fe, NM, 31–36, 1992.
445. B.A. Bushaw, J.T. Munely, 'Isotopically Selective RIMS of Rare Radionuclides by Double-resonance Excitation with Single-frequency CW Lasers', in *5th International Symposium on Resonance Ionization Spectroscopy and its Applications*, Institute of Physics Conference Series Number 114, Congress Center Villa Ponti, Varese, Italy, 387–392, 1990.
446. W.M. Fairbank, Jr, 'A Seat-of-the pants Approach to RIS Theory', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 457–464, 1996.
447. X. Xiong, J.M.R. Hutchinson, J. Fassett, W.M. Fairbank, Jr, 'Measurements of the Odd–Even Effect in the Resonance Ionization of Tin as a Function of Laser Intensity', in *6th International Symposium on Resonance Ionization Spectroscopy and its Applications*, Institute of Physics Conference Series Number 128, Los Alamos National Laboratory, Santa Fe, NM, 123–126, 1992.
448. N. Erdmann, G. Herrmann, G. Huber, S. Köhler, J.V. Kratz, A. Mansel, M. Nunnemann, G. Passler, N. Trautmann, A. Waldek, 'Trace Analysis of Plutonium in Environmental Samples by Resonance Ionization Mass Spectrometry (RIMS)', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 205–208, 1996.
449. W. Nörtershäuser, N. Trautmann, K. Wendt, B.A. Bushaw, 'Isotope Shifts and Hyperfine Structure in the $4s^2\,{}^1S_0 - 4s\,4p\,{}^1P_1 - 4s\,4d^1\,D_2$ Transitions of Stable Calcium Isotopes and Calcium-41', *Spectrochim. Acta*, **53B**(5), 709–721 (1998).
450. B.A. Bushaw, W. Nörtershäuser, K. Wendt, 'Lineshapes and Optical Selectivity in High-resolution Double-resonance Ionization Mass Spectrometry', *Spectrochim. Acta*, **54B**(2), 321–332 (1999).
451. C.S. Hansen, W.F. Calaway, M.J. Pellin, R.C. Wiens, D.S. Burnett, 'Three-color Resonance Ionization Spectroscopy of Zr in Si', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 215–218, 1996.
452. O.R. Jones, C.J. Abraham, H.H. Telle, A.E. Oakley, 'The Feasibility of RIMS for the Analysis of Potentially Toxic Elements Accumulation in Neural Tissue', in *8th International Symposium on Resonance Ionization Spectroscopy and its Applications*, American Institute of Physics Conference Proceeding 388, State College, PA, 289–292, 1996.
453. R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson, 'Coherent Light Emission from GaAs Junctions', *Phys. Rev. Lett.*, **9**, 366–368 (1962).
454. M.I. Nathan, W.P. Dumke, G. Burns, F.H. Dill, Jr, G. Lasher, 'Stimulated Emission of Radiation from GaAs p–n Junctions', *Appl. Phys. Lett.*, **1**(3), 62–64 (1962).
455. N. Holonyak, S.F. Bevacqua, 'Coherent (visible) Light Emission from $Ga(As_{1-x}P_x)$ Junctions', *Appl. Phys. Lett.*, **1**(4), 82–83 (1962).
456. T.M. Quist, R.H. Rediker, R.J. Keyes, W.E. Krag, B. Lax, A.L. McWhorter, H.J. Zeigler, 'Semiconductor Maser of GaAs', *Appl. Phys. Lett.*, **1**(4), 91–92 (1962).
457. A.W. Mantz, 'A Review of Spectroscopic Applications of Tunable Semiconductor Lasers', *Spectrochim. Acta*, **51A**(13), 2211–2236 (1995).
458. J.C. Camparo, C.M. Klimack, 'Laser Spectroscopy on a ''Shoestring'' ', *Am. J. Phys.*, **51**(12), 1077–1081 (1983).
459. J.C. Camparo, 'The Diode Laser in Atomic Physics', *Contemp. Phys.*, **26**(5), 443–477 (1985).
460. K.-H. Weber, J. Lawrenz, A. Obreski, K. Niemax, 'High-resolution Laser Spectroscopy of Aluminum, Gallium and Thallium', *Phys. Scr.*, **35**, 309–312 (1987).
461. C. Vadla, A. Obreski, K. Niemax, 'Isotope Shift of the $3s\,{}^2S_{1/2}$ and $3p\,{}^2P_J$ Levels in $^{6,7}Li$', *Opt. Commun.*, **63**(5), 288–292 (1987).
462. C. Vadla, J. Lawrenz, K. Niemax, 'Measurement of Velocity Dependence of Collisional Excitation Transfer Applying Resonant Doppler-free 2-photon Laser Spectroscopy', *Opt. Commun.*, **63**(5), 293–297 (1987).
463. D. Veza, J. Lawrenz, K. Niemax, 'Velocity Dependence of Impact Line-broadening Studied by Resonant Doppler-free Two-photon Laser Spectroscopy', *Z. Phys. D*, **9**, 135–141 (1988).
464. A.S. Zibrov, R.W. Fox, R. Ellingsen, C.S. Weimer, V.L. Velichansky, G.M. Tino, L. Hollberg, 'High-resolution Diode-laser Spectroscopy of Calcium', *Appl. Phys. B*, **59**, 327–331 (1994).
465. F.S. Pavone, 'Diode Lasers and their Applications in Spectroscopy', *Riv. Nuovo Cimento*, **19**(9), 1–42 (1996).

466. C.E. Wieman, L. Hollberg, 'Using Diode Laser for Atomic Physics', *Rev. Sci. Instrum.*, **62**(1), 1–20 (1991).
467. D.T. Cassidy, J. Reid, 'Atmospheric Pressure Monitoring of Trace Gases Using Tunable Diode Lasers', *Appl. Opt.*, **21**(7), 1185–1190 (1982).
468. D.T. Cassidy, J. Reid, 'High-sensitivity Detection of Trace Gases Using Sweep Integration and Tunable Diode Lasers', *Appl. Opt.*, **21**(14), 2527–2530 (1982).
469. P. Werle, R. Mucke, F. Slemr, 'The Limits of Signal Averaging in Atmospheric Trace-gas Monitoring by Tunable Diode-laser Absorption-spectroscopy (TDLAS)', *Appl. Phys. B*, **57**(2), 131–139 (1993).
470. P. Werle, B. Scheumann, J. Schandl, 'Real-time Signal-processing Concepts for Trace-gas Analysis by Diode-laser Spectroscopy', *Opt. Eng.*, **33**(9), 3093–3105 (1994).
471. P. Werle, 'Tunable Diode Laser Absorption Spectroscopy: Recent Findings and Novel Approaches', *Infrared Phys. Technol.*, **37**(1), 59–66 (1996).
472. P. Werle, 'Analytical Applications of Infrared Semiconductor-lasers in Atmospheric Trace Gas Monitoring', *J. Phys. IV*, **4**(C4), 9–12 (1994).
473. P. Werle, 'Spectroscopic Trace Gas Analysis Using Semiconductor Diode Lasers', *Spectrochim. Acta*, **52A**(8), 805–822 (1996).
474. P. Werle, 'A Review of Recent Advances in Semiconductor Laser Based Gas Monitors', *Spectrochim. Acta*, **54A**(2), 197–236 (1998).
475. R. Schieder, 'High Resolution Diode Laser and Heterodyne Spectroscopy with Applications toward Remote Sensing', *Infrared Phys. Technol.*, **35**(2/3), 477–486 (1994).
476. K. Niemax, 'Hochauflösende und hochempfindliche Laserspektroskopie in der Elementanalytik', *Naturwissenschaften*, **74**, 474–481 (1987).
477. R. Hegenröder, K. Niemax, 'Laser Atomic Absorption Spectroscopy Applying Semiconductor Diode Laser', *Spectrochim. Acta*, **43B**(12), 1443–1449 (1988).
478. R. Hegenröder, K. Niemax, 'Atomic Absorption Spectroscopy with Tunable Semiconductor Diode Lasers', *Trends Anal. Chem.*, **8**(9), 333–335 (1989).
479. A. Obreski, J. Lawrenz, K. Niemax, 'On the Potential and Limitations of Spectroscopic Isotope Ratio Measurements by Resonant Doppler-free Two-photon Laser-enhanced Ionization Spectroscopy', *Spectrochim. Acta*, **45B**(1/2), 15–36 (1990).
480. T. Imasaka, 'Analytical Molecular Spectroscopy with Diode Lasers', *Spectrochim. Acta Rev.*, **15**(5), 329–348 (1993).
481. T.E. Barber, P.E. Walters, J.D. Winefordner, N. Omenetto, 'Evaluation of Absolute Number Densities by Diode Laser Atomic Absorption Spectroscopy', *Appl. Spectrosc.*, **45**(4), 524–526 (1991).
482. P.W. Milonni, J.H. Eberly, *Lasers*, John Wiley & Sons, New York, 1988.
483. J. Hecht, 'Long-wavelength Diode Lasers are Tailored for Fiber optics', *Laser Focus World*, **Aug.**, 79–89 (1992).
484. J. Hecht, 'Diode-laser Performance Rises as Structures Shrink', *Laser Focus World*, **May**, 127–143 (1992).
485. W.R. Trutna, P. Zorabedian, 'Research on External-cavity Lasers', *Hewlett-Packard J.*, 35–38 (1993).
486. J. Hecht, 'Semiconductor Diode Lasers Span the Rainbow', *Laser Focus World*, **April**, 199–211 (1993).
487. J. Hecht, 'Gallium Arsenide Lasers offer an Array of Options', *Laser Focus World*, **July**, 83–92 (1993).
488. D.W. Nam, R.G. Waarts, 'Advanced Laser Diodes bring Compact Blue-green Sources to Light', *Laser Focus World*, **Aug.**, 49–55 (1994).
489. T.V. Higgins, 'The Smaller, Cheaper, Faster World of the Laser Diode', *Laser Focus World*, **April**, 65–76 (1995).
490. J. Lawrenz, K. Niemax, 'A Semiconductor Diode Laser Spectrometer for Laser Spectrochemistry', *Spectrochim. Acta*, **44B**(2), 155–164 (1989).
491. H.R. Telle, 'Stabilization and Modulation Schemes of Laser Diodes for Applied Spectroscopy', *Spectrochim. Acta Rev.*, **15**(5), 301–327 (1993).
492. A.W. Mantz, 'A Review of the Applicability of Tunable Diode-laser Spectroscopy at High-sensitivity', *Microchem. J.*, **50**(3), 351–364 (1994).
493. R.U. Martinelli, 'Mid-infrared Wavelengths Enhance Trace-gas Sensing', *Laser Focus World*, **March**, 77–81 (1996).
494. H. Groll, K. Niemax, 'Multielement Diode Laser Atomic Absorption Spectrometry in Graphite Tube Furnace and Analytical Flames', *Spectrochim. Acta*, **48B**(5), 633–641 (1993).
495. C. Schnürer-Patschan, A. Zybin, H. Groll, K. Niemax, 'Improvement in Detection Limit in Graphite Furnace Diode Laser Atomic Absorption Spectrometry by Wavelength Modulation Technique', *J. Anal. At. Spectrom.*, **8**, 1103–1107 (1993).
496. A. Zybin, C. Schnürer-Patschan, K. Niemax, 'Measurements of $C_2F_4Cl_2$, CCl_4, CHF_3 and O_2 by Wavelength Modulated Laser Absorption Spectroscopy of Excited Cl, F and O in a DC Discharge Applying Semiconductor Diode Lasers', *Spectrochim. Acta*, **48B**(14), 1713–1718 (1993).
497. C. Schnürer-Patschan, K. Niemax, 'Elemental Selective Detection of Chlorine in Capillary Gas Chromatography by Wavelength Modulation Diode Laser Atomic Absorption Spectrometry in a Microwave Induced Plasma', *Spectrochim. Acta*, **50**, 963–969 (1995).
498. A. Zybin, C. Schnürer-Patschan, K. Niemax, 'Wavelength Modulation Diode Laser Atomic Spectrometry in Modulated Low-pressure Helium Plasmas for Element-selective Detection in Gas Chromatography', *J. Anal. At. Spectrom.*, **10**, 563–567 (1995).
499. V. Liger, A. Zybin, Y. Kuritsyn, K. Niemax, 'Diode-laser Atomic-absorption Spectrometry by the Double-beam–Double-modulation Technique', *Spectrochim. Acta*, **52B**(8), 1125–1138 (1997).

500. H.D. Wizemann, K. Niemax, 'Cancellation of Matrix Effects and Calibration by Isotope Dilution in Isotope-selective Diode Laser Atomic Absorption Spectrometry', *Anal. Chem.*, **69**(20), 4291–4293 (1997).
501. A. Zybin, K. Niemax, 'GC Analysis of Chlorinated Hydrocarbons in Oil and Chlorophenols in Pliant Extracts Applying Element-selective Diode Laser Plasma Detection', *Anal. Chem.*, **69**(4), 755–757 (1997).
502. A. Zybin, K. Niemax, 'Improvement of the Wavelength Tunability of Etalon-type Laser Diodes and Mode Recognition and Stabilization in Diode Laser Spectrometers', *Spectrochim. Acta*, **52B**(8), 1215–1221 (1997).
503. A. Zybin, C. Schnürer-Patschan, M.A. Bolshov, K. Niemax, 'Elemental Analysis by Diode Laser Spectroscopy', *Trends Anal. Chem.*, **17**(8–9), 513–520 (1998).
504. P.A. Johnson, J.A. Vera, B.W. Smith, J.D. Winefordner, 'Determination of Rubidium Using Laser Diode Excited Atomic Fluorescence', *Spectrosc. Lett.*, **21**(7), 607–612 (1988).
505. K.C. Ng, A.H. Ali, T.E. Barber, J.D. Winefordner, 'Multiple Mode Semiconductor Diode Laser as a Spectral Line Source for Graphite Furnace Atomic Absorption Spectroscopy', *Anal. Chem.*, **62**, 1893–1895 (1990).
506. K.C. Ng, A.H. Ali, T.E. Barber, J.D. Winefordner, 'Flame Atomic Absorption Spectroscopy Using a Single-mode Laser Diode as the Line Source', *Appl. Spectrosc.*, **44**(6), 1094–1096 (1990).
507. K.C. Ng, A.H. Ali, T.E. Barber, J.D. Winefordner, 'The Applicability of a Multiple-mode Diode Laser in Flame Atomic Absorption Spectroscopy', *Appl. Spectrosc.*, **44**(5), 849–852 (1990).
508. H. Groll, C. Schnürer-Patschan, Y. Kuritsyn, K. Niemax, 'Wavelength Modulation Diode Laser Atomic Absorption Spectrometry in Analytical Flames', *Spectrochim. Acta*, **49B**(12–14), 1463–1472 (1994).
509. R. Hegenröder, D. Veza, K. Niemax, 'Detection Limit and Selectivity for Lithium Isotopes in Continuous Wave Field Ionization Laser Spectroscopy', *Spectrochim. Acta*, **48B**(4), 589–596 (1993).
510. P.E. Walters, T.E. Barber, M.W. Wensing, J.D. Winefordner, 'A Diode Laser Wavelength Reference System Applied to the Determination of Rb in Atomic Fluorescence Spectroscopy', *Spectrochim. Acta*, **46B**(6/7), 1015–1020 (1991).
511. D.H. McIntyre, C.E. Fairchild, J. Cooper, R. Walser, 'Diode-laser Noise Spectroscopy of Rubidium', *Opt. Lett.*, **18**(21), 1816–1818 (1993).
512. T.E. Barber, P.E. Walters, M.W. Wensing, J.D. Winefordner, 'Diode Laser Atomic Absorption Using a New Reference Method', *Spectrochim. Acta*, **46B**(6/7), 1009–1014 (1991).
513. P. Ljung, O. Axner, 'Measurements of Rubidium in Standard Reference Samples by Wavelength-modulation Diode Laser Absorption Spectrometry in a Graphite Furnace', *Spectrochim. Acta*, **52B**(3), 305–319 (1997).
514. K. Hayasaka, M. Watanabe, H. Imajo, R. Ohmukai, S. Urabe, 'Tunable 397-nm Light Source for Spectroscopy Obtained by Frequency Doubling of a Diode Laser', *Appl. Opt.*, **33**(12), 2290–2293 (1994).
515. W.P. Risk, W.J. Kozlovsky, S.D. Lau, 'Generation of 425-nm Light by Waveguide Frequency Doubling of a GaAlAs Laser Diode in an Extended-cavity Configuration', *Appl. Phys. Lett.*, **63**(23), 3134–3136 (1993).
516. J. Franzke, J. Brust, C. Vadla, H.D. Wizemann, K. Niemax, 'Second Harmonic Generation by Applying Injection-locked Radiation from a High-power, Broad-stripe Laser Diode', *Spectrochim. Acta*, **53B**(5), 763–768 (1998).
517. F. Slemr, G.W. Harris, D.R. Hastie, G.I. Mackay, H.I. Schiff, 'Measurement of Gas Phase Hydrogen Peroxide in Air by Tunable Diode Laser Absorption Spectrometry', *J. Geophys. Res.*, **91**, 5371–5378 (1986).
518. A. Stanton, C. Hovde, 'Near-infrared Diode Lasers Measure Greenhouse Gases', *Laser Focus World*, **Aug.**, 117–120 (1992).
519. L.-G. Wang, H. Riris, C.B. Carlisle, T.F. Gallagher, 'Comparison of Approaches to Modulation Spectroscopy with GaAlAs Semiconductor Lasers: Application to Water Vapor', *Appl. Opt.*, **27**(10), 2071–2077 (1988).
520. L.-G. Wang, D.A. Tate, H. Riris, T.F. Gallagher, 'High-sensitivity Frequency-modulation Spectroscopy with a GaAlAs Diode Laser', *J. Opt. Soc. Am. B*, **6**(5), 871–876 (1989).
521. G.S. Hurst, M.G. Payne, 'Elemental Analysis Using Resonance Ionization Spectroscopy', *Spectrochim. Acta*, **43B**, 715–726 (1988).
522. H. Groll, G. Schaldach, H. Berndt, K. Niemax, 'Measurement of Cr(III)/Cr(VI) Species by Wavelength Modulation Diode Laser Flame Atomic Absorption Spectrometry', *Spectrochim. Acta*, **50B**, 1293–1298 (1995).
523. P. Kluczynski, O. Axner, 'Theoretical Description of Wavelength Modulation Spectrometry in Terms of Analytical and Background Signals based upon Fourier Analysis', *Appl. Opt.*, **38**, 5803–5815 (1999).
524. G.C. Bjorklund, 'Frequency-modulation Spectroscopy: A New Method for Measuring Weak Absorption and Dispersion Lineshapes', *Opt. Lett.*, **5**(1), 15–17 (1980).
525. J.L. Hall, L. Hollberg, T. Baer, H.G. Robinson, 'Optical Heterodyne Saturation Spectroscopy', *Appl. Phys. Lett.*, **39**, 680–682 (1981).
526. W. Lenth, 'High Frequency Heterodyne Spectroscopy with Current-modulated Diode Lasers', *IEEE J. Quantum Electron.*, **QE-20**(9), 1045–1050 (1984).
527. M. Gehrtz, W. Lenth, A.T. Young, H.S. Johnston, 'High-frequency-modulation Spectroscopy with a Lead-salt Diode Laser', *Opt. Lett.*, **11**(3), 132–134 (1986).
528. M. Gehrtz, G.C. Bjorklund, E.A. Whittaker, 'Quantum-limited Laser Frequency-modulation Spectroscopy', *J. Opt. Soc. Am. B*, **2**(9), 1510–1526 (1985).

529. C.B. Carlisle, D.E. Cooper, H. Preier, 'Quantum Noise-limited FM Spectroscopy with a Lead-salt Diode Laser', *Appl. Opt.*, **28**(13), 2567–2576 (1989).
530. P. Werle, F. Slemr, M. Gehrtz, C. Bräuchle, 'Quantum-limited FM-spectroscopy with a Lead-salt Diode Laser', *Appl. Phys. B*, **49**, 99–108 (1989).
531. J.M. Supplee, E.A. Whittaker, W. Lenth, 'Theoretical Description of Frequency Modulation and Wavelength Modulation Spectroscopy', *Appl. Opt.*, **33**(27), 6294–6302 (1994).
532. G.R. Janik, C.B. Carlisle, T.F. Gallagher, 'Two-tone Frequency-modulation Spectroscopy', *J. Opt. Soc. Am. B*, **3**(8), 1070–1074 (1986).
533. G.C. Bjorklund, M.D. Levenson, W. Lenth, C. Ortiz, 'Frequency Modulation (FM) Spectroscopy – Theory of Lineshapes and Signal-to-noise Analysis', *Appl. Phys. B*, **32**, 145–152 (1983).
534. N. Hadgu, J. Gustafsson, W. Frech, O. Axner, 'Rubidium Atom Distribution and Nonspectral Interference Effects in Transversely Heated Graphite Atomizers Evaluated by Wavelength Modulated Diode Laser Absorption Spectrometry', *Spectrochim. Acta*, **53B**, 923–943 (1998).
535. D. Rojas, P. Ljung, O. Axner, 'An Investigation of the 2*f*-wavelength Modulation Technique for Detection of Atoms under Optically Thin as well as Thick Conditions', *Spectrochim. Acta*, **52B**(11), 1663–1686 (1997).
536. J. Gustafsson, D. Rojas, O. Axner, 'The Influence of Hyperfine Structure and Isotope Shift on the Detection of Rb Atoms in Atmospheric Pressure Atomizers by the 2*f*-wavelength Modulation Technique', *Spectrochim. Acta*, **52B**, 1937–1953 (1997).
537. J. Gustafsson, O. Axner, 'The Influence of Hyperfine Structure and Isotope Shift on the Detection of Rb by 2*f*-wavelength Modulation Diode Laser Absorption Spectrometry – Experimental Verification of Simulations', *Spectrochim. Acta*, **53B**, 1895–1905 (1998).
538. J. Gustafsson, N. Chekalin, D. Rojas, O. Axner, 'Extension of the Dynamic Range of the 2*f*-wavelength Modulated Diode Laser Absorption Spectrometry Technique – Detection of Atoms under Optically Thick Conditions', *Spectrochim. Acta*, **55B**, 237–262 (2000).
539. J. Gustafsson, O. Axner, 'Theoretical Investigation of the Temperature Dependence of the 2*f*-wavelength Modulation Diode-laser Absorption Signal', *Spectrochim. Acta*, **53B**, 1827–1846 (1998).
540. D.E. Cooper, R.E. Warren, 'Frequency Modulation Spectroscopy with Lead-salt Diode Lasers: A Comparison of Single-tone and Two-tone Techniques', *Appl. Opt.*, **26**, 3726–3732 (1987).
541. J. Reid, D. Labrie, 'Second-harmonic Detection with Tunable Diode Lasers – Comparison of Experiment and Theory', *Appl. Phys. B*, **26**, 203–210 (1981).
542. M.L. Olsen, D.L. Grieble, P.R. Griffiths, 'Second Derivative Tunable Diode Laser Spectrometry for Line Profile Determination. I. Theory', *Appl. Spectrosc.*, **34**(1), 50–56 (1980).
543. L.C. Philippe, R.K. Hanson, 'Laser Diode Wavelength-modulation Spectroscopy for Simultaneous Measurement of Temperature, Pressure, and Velocity in Shock-heated Oxygen Flows', *Appl. Opt.*, **32**(30), 6090–6103 (1993).
544. E.I. Moses, C.L. Tang, 'High-sensitivity Laser Wavelength-modulation Spectroscopy', *Opt. Lett.*, **1**, 115–117 (1977).
545. R. Arndt, 'Analytical Line Shapes for Lorentzian Signals Broadened by Modulation', *J. Appl. Phys.*, **36**(8), 2522–2524 (1965).
546. G.V.H. Wilson, 'Modulation Broadening of NMR and ESR Line Shapes', *J. Appl. Phys.*, **34**(11), 3276–3285 (1963).
547. X. Zhu, D.T. Cassidy, 'Modulation Spectroscopy with a Semiconductor Diode Laser by Injection-current Modulation', *J. Opt. Soc. Am. B*, **14**(8), 1945–1950 (1997).
548. D.M. Bruce, D.T. Cassidy, 'Detection of Oxygen Using Short-extended-cavity GaAs Semiconductor Diode Laser', *Appl. Opt.*, **29**, 1327–1332 (1990).
549. J.A. Silver, A.C. Stanton, 'Optical Interference Fringe Reduction in Laser Absorption Experiments', *Appl. Opt.*, **27**(10), 1914–1916 (1988).
550. C.R. Webster, 'Brewster-plate Spoiler: A Novel Method for Reducing the Amplitude of Interference Fringes that Limit Tunable-laser Absorption Sensitivities', *J. Opt. Soc. Am. B*, **2**(9), 1464–1470 (1985).
551. H.C. Sun, E.A. Whittaker, 'Novel Etalon Fringe Rejection Technique for Laser Absorption Spectroscopy', *Appl. Opt.*, **31**(24), 4998–5002 (1992).
552. M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, E.A. Cornell, 'Observation of Bose–Einstein Condensation in a Dilute Atomic Vapor', *Science*, **269**, 198–199 (1995).

Laser-induced Breakdown Spectroscopy

David A. Cremers and Andrew K. Knight
Los Alamos National Laboratory, Los Alamos, USA

1 Introduction 1
2 Theory 2
2.1 Fundamentals of Laser-induced Breakdown 2
2.2 Post-breakdown Phenomena 2
2.3 Dynamics of Laser Ablation 3
3 Laser-induced Breakdown Spectroscopy Instrumentation 4
3.1 General 4
3.2 Laser Systems 4
3.3 Methods of Spectral Resolution 4
3.4 Detectors 6
3.5 Examples of Instruments 7
4 Characteristics of the Laser Plasma 8
4.1 On Solids 8
4.2 In Gases 8
4.3 In Liquids 9
5 Analytical Capabilities 9
5.1 Advantages 9
5.2 Considerations in the Use of Laser-induced Breakdown Spectroscopy 11
5.3 Representative Figures-of-merit 12
6 Hybrid Techniques 13
6.1 Double-pulse Techniques 13
6.2 Combined with Other Methods 14
6.3 Continuous Optical Discharge 15
7 Perspective and Future Developments 16
Acknowledgments 17
Abbreviations and Acronyms 17
Related Articles 17
References 17

Laser-induced breakdown spectroscopy (LIBS) is a novel method of elemental analysis based on a laser-generated plasma. Pulses from a laser are focused on a sample to atomize a small amount of material resulting in the formation of a microplasma. Because of the high plasma temperature, the resulting atoms are electronically excited to emit light. The plasma light is spectrally resolved and detected to determine the elemental composition of the sample based on the unique emission spectrum of each element. Because of the simplicity of the method, it is suited for analyses that cannot be carried out using conventional methods of atomic emission spectroscopy (AES). This is particularly true for measurements that must be conducted outside of an analytical laboratory. A particular advantage of LIBS is the ability to analyze most types of samples without any preparation. This means that samples can be interrogated in situ, providing rapid measurement capability and permitting the method to be used in the analysis of gases, liquids, and solids in a variety of different sampling configurations. Although LIBS provides sensitive detection for many elements, it is not an ultrasensitive detection technique. In addition, under field conditions, the method typically does not provide the high accuracy and precision offered by laboratory-based methods of AES.

1 INTRODUCTION

LIBS is a method of elemental analysis based on AES.[1–3] Three basic steps are involved in AES measurements. These are (1) vaporization of the sample to produce free atoms, (2) electronic excitation of the atoms to induce optical emissions at discrete wavelengths indicative of the atoms, and (3) collection, recording and analysis of the optical emissions. Atoms in the sample are identified by their unique spectral emissions or 'fingerprint' spectra. In conventional AES methods, a plasma source is used for steps (1) and (2). Conventional plasma sources include arcs and sparks and the inductively coupled plasma (ICP). These sources all require the use of some physical device to deliver electrical energy to form and sustain the plasma such as metal electrodes (for sparks and arcs) and a metal coil (ICP). In LIBS, however, the plasma or laser spark is formed by intense optical radiation focused on the sample to produce dielectric breakdown leading to plasma formation.

The basic steps involved in a LIBS measurement are shown in Figure 1. Focusing a laser pulse on the sample produces local heating of material in the focal volume resulting in the ejection of a small mass of material in the form of solid particles, liquid drops, and atomic vapor. Via interactions between the incident laser pulse, electrons, and free atoms, a hot plasma is formed above the target surface. Through collisions between the different species in the plasma the atoms are electronically excited to emit light at discrete frequencies unique to each species. These unique spectral 'fingerprints' permit identification of the species through collection and analysis of the plasma light. Because the laser plasma is formed by a laser pulse of short duration (typically 5–20 ns), the plasma decays and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

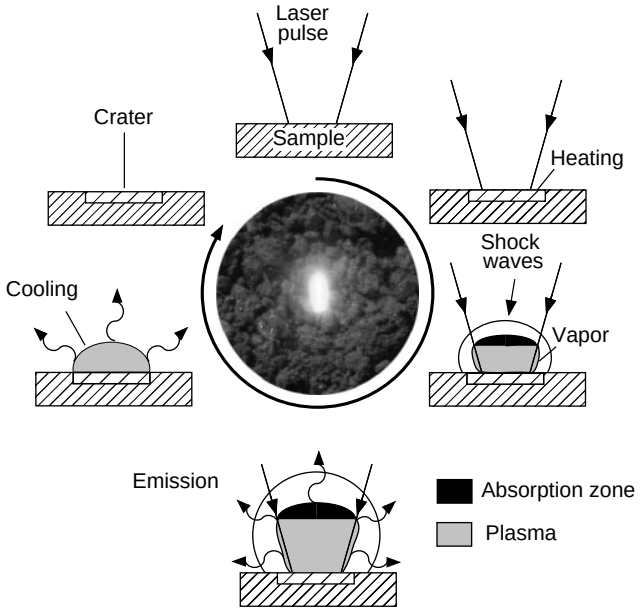

Figure 1 Diagram showing the main events in the interrogation of a solid surface by a laser pulse during a LIBS measurement.

cools rapidly, usually within tens of microseconds after formation. The ejected material results in a small crater formed on the sample surface.

Because the LIBS plasma is formed by light, it represents a unique method of elemental analysis having many advantages over conventional methods of AES. These advantages include (1) in situ analysis (only optical access to the sample is required), (2) little or no sample preparation, and (3) the ability to sample gases, liquids, and solids directly. On the other hand, like other plasma-based methods of AES, it provides for (1) simultaneous multi-element detection and (2) exhibits good sensitivity for many elements. The main limitations of LIBS relate to the sampling characteristics of the technique. Specifically, that the amount of sample vaporized and the plasma characteristics (e.g. temperature, electron density) are strongly dependent on the parameters of the laser pulse, the physical characteristics of the target sample, and the sampling geometry. These limitations are manifested in terms of limited accuracy and precision compared to more conventional analytical methods.

2 THEORY

2.1 Fundamentals of Laser-induced Breakdown

The LIBS method is based on the formation of a plasma, or highly ionized gas, within the focus of the laser pulse directed either into a material (gas or liquid) or onto the surface of a solid target. Plasma formation is accompanied by a bright flash of light and a loud 'snap' emanating from the focal region. Detailed analysis of the physics of laser-induced breakdown can be found elsewhere.[4,5] Briefly, the plasma is formed by the transfer of optical energy from the laser pulse to atoms, ions, and electrons resulting from deposition of energy into the target. This energy transfer occurs through the process of inverse bremmstrahlung absorption involving interactions between free electrons and atoms and ions. In order for the inverse bremmstrahlung process to be of sufficient strength to produce breakdown, a high density of electrons (and ions) must be present. Two mechanisms accounting for electron (e^-) production are described by two processes involving atoms (M) (Equations 1 and 2):

$$e^- + M \longrightarrow 2e^- + M^+ \quad (1)$$

$$M + nh\nu \longrightarrow M^+ + e^- \quad (2)$$

Mechanism (1) involves absorption of laser radiation by an electron during a collision with a neutral atom. If sufficient energy is absorbed to ionize the atom then two free electrons result. If this process continues generating two electrons for each one electron involved in the collision then what is termed cascade electron growth results. Mechanism (2) is termed multiphoton ionization. In this process a number of photons (n) each of energy $h\nu$ are absorbed simultaneously by an atom resulting in a free electron and an ionized species. In order for this process to occur, it must be that $nh\nu > Ip$, where Ip is the ionization potential of the atom.

2.2 Post-breakdown Phenomena

Following formation of the initial plasma by the laser pulse, the plasma passes through several different phases. Detailed descriptions of these can be found in the literature.[4,6] Here a brief overview is presented. Once free electrons have been generated in the focal volume, the plasma grows through the process of inverse bremsstrahlung (IB) of which two types can be identified, electron–neutral and electron–ion, each being important during different stages of plasma growth. When few electrons are present, electron–neutral IB (IB_{e-n}) predominates but as plasma formation continues to the point at which about 1% ionization occurs, electron–ion IB (IB_{e-i}) becomes the chief absorption process. The absorption coefficients (α) for each type are given in Equations (3) and (4):

$$\alpha(IB_{e-n}) = 1 - \exp\left(\frac{-hc}{\lambda kT}\right) \sum Q_j n_e n_j \quad (3)$$

$$\alpha(IB_{e-i}) = 1 - \exp\left(\frac{-hc}{\lambda kT}\right) \left(\frac{4e^6\lambda^3}{3hc^4 m_e}\right) \times \left(\frac{2\pi}{3m_e kT}\right)^{1/2} n_e \sum z_i^2 n_i g_i \quad (4)$$

where:

Q_j = average cross-section for absorption of a photon (wavelength λ) as a result of collisions with the jth species
n_e = electron number density
n_i and n_j = number density of ith and jth species
z_i = charge of the ith species
g_i = Gaunt factor.

The temperature regime over which the change from the two types of IB absorption occurs is between 6000 and 10 000 K, which is also the temperature regime of a LIBS plasma. Both absorption coefficients decrease with increased temperature with the coefficient $\alpha(IB_{e-i})$ having the stronger dependence. Because this coefficient depends on the square of the electron density compared to the linear dependence for $\alpha(IB_{e-n})$, electron–ion IB predominates as the temperature and hence the electron density increase.

Following the formation of an absorbing plasma within the focal zone, the plasma will continue to grow away from the surface towards the direction of the incoming laser pulse. This growth occurs because the laser pulse feeds energy into the plasma with most of the absorption occurring in the upper boundary region of the plasma as shown in Figure 1. This process accounts for the inverted cone shape of the plasma observed in the focal volume. The exact character of the evolving plasma will depend on parameters of the laser pulse (irradiance, spot size, wavelength) and the surrounding atmosphere. Three distinct evolutionary paths can be described in terms of the characteristics of absorption and shock or pressure waves emanating from the focal zone. The three types are classed as either laser-supported combustion (LSC), laser-supported detonation (LSD) or laser-supported radiation (LSR) waves. The interested reader is directed to the literature for a discussion of the characteristics of each type.[4,6]

2.3 Dynamics of Laser Ablation

The interaction between a laser pulse of high power density and a material is a complicated process. A thorough consideration of the interaction depends on characteristics of the laser pulse and the material. Several authors have considered the interaction in detail.[7,8] Here we present a brief description of the main processes involved.

The laser pulse incident on the surface interacts mainly with the electrons in the bulk material. These interactions involve absorption of energy from the pulse by electrons that in turn collide with other electrons and with other constituents of the matrix. These collisions result in the transfer of energy from the laser pulse to the material in the focal volume resulting in heating. This heating occurs rapidly and results in melting of a thin surface layer with some of the heat being conducted away into the bulk matrix. Because of the high power densities on the surface however, the melted material continues to be heated and when the deposited energy exceeds the latent heat of vaporization, atomized material is ejected from the surface. Some solid and liquid particles may also be ejected via the strong shock/pressure waves generated at the surface.

An estimate of the minimum laser pulse-power density required to produce vaporization (P_{min}) can be obtained from Equation (5):

$$P_{min} = \rho L_v \kappa^{1/2} \Delta t^{-1/2} \quad (5)$$

where:

ρ = mass density of target
L_v = latent heat of vaporization
κ = thermal diffusivity of target
Δt = pulse length.

For pure aluminum metal, $P_{min} \approx 1.75 \times 10^8\ W\,cm^{-2}$. Note that this calculation assumes all incident pulse energy goes toward heating therefore representing a lower bound on the minimum power density required for vaporization.

Although ejection of material can proceed over a wide range of power densities, two distinct operating regimes can be identified, one most favorable for quantitative analysis. For low power densities ($<10^8\ W\,cm^{-2}$), the ejection of positive ions of volatile elements such as Na and K is readily observed from the surface whereas more refractory elements (Mo, W, etc.) remain in the bulk material. In this case the ejected material will not be representative of the bulk composition and therefore operation at these low power densities is not appropriate for quantitative analysis.

At higher power densities ($>10^9\ W\,cm^{-2}$) significantly more laser energy is transferred into the material resulting in a greater degree of vaporization and atomization of a more representative distribution of elements from the molten pool. In addition, because of the greater mass of ablated material, the incident laser pulse also interacts with electrons and ions in the plume above the surface resulting in increased excitation of atomized species. Because laser ablation is more pronounced at higher power densities, a visible depression or crater is left on the target surface in the interaction zone. Typically, the mass of material ejected from the surface is very small. Assuming again that all energy goes toward the ablation process, a simple calculation shows that the upper limit to the amount of material that can be vaporized by a 100 mJ pulse is about 6.6 μg. Because of reflection losses

at the surface and the conversion of some of the incident energy into light and sound, the amount of material actually vaporized will be significantly less. Measurements indicate, for example, that about 10 ng are ablated from aluminum metal using a laser pulse of 175 mJ.

3 LASER-INDUCED BREAKDOWN SPECTROSCOPY INSTRUMENTATION

3.1 General

A diagram of a generalized LIBS instrument is shown in Figure 2. The main components include a laser, a method of spectrally resolving or of spectrally selecting a certain narrow region of the spectrum to monitor, and a method of detecting the spectrally selected light. The specifications of each component as well as the method of sampling used will depend on the application. Factors to consider include (1) the elements to be monitored (number and type), (2) the characteristics of the sample (complexity, homogeneity, etc.), (3) the type of analysis (e.g. a qualitative versus quantitative measurement), and (4) the state of the sample (e.g. gas, liquid, or solid).

3.2 Laser Systems

The parameters important in the specification of the laser to be used for LIBS include (1) pulse energy, (2) pulse repetition rate, (3) beam mode quality, (4) size/weight, and (5) cooling and electrical power requirements. The wavelength of the laser beam is not an important factor in most cases.

Solid state lasers, in particular pulsed and Q-switched Nd : YAG lasers having pulsewidths less than 15 ns, are typically used for LIBS measurements because these lasers are a compact and convenient source of the powerful pulses needed to generate the laser plasmas.[9] The fundamental wavelength of the Nd : YAG laser (1064 nm) can easily be converted to shorter wavelengths (532, 355, and 266 nm) via passive harmonic generation techniques which may have certain advantages in terms of increased energy coupling into a particular sample, but typically the 1064 nm wavelength is used because this provides the highest power density. Other types of lasers, most notably the pulsed CO_2 laser (10 600 nm wavelength) and the excimer laser (typical wavelengths of 193, 248, 308 nm) have been used for LIBS. In comparison to solid state lasers, however, these lasers require more maintenance (e.g. change in gases) and special optical materials because their wavelengths lie in the far-infrared and ultraviolet spectral regions, respectively. For this reason these lasers are not widely used.

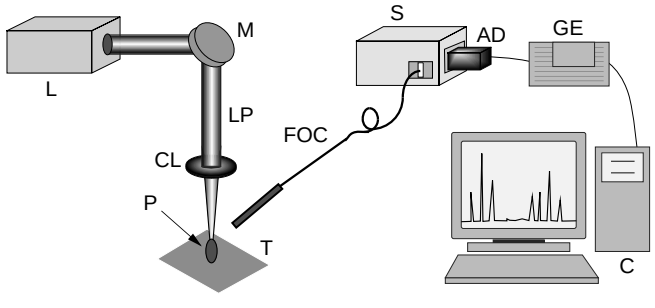

Figure 2 Diagram of a typical LIBS apparatus. L = laser; M = mirror; LP = laser pulse; CL = converging lens; P = plasma; T = target; FOC = fiber optic cable; S = spectrograph; AD = array detector; GE = gating electronics; C = computer.

Another advantage of the Nd : YAG laser is the variety of different sizes available commercially. These range from laboratory-based models which can output a Joule or more of pulse energy at repetition rates between 10 and 50 Hz to small hand-held versions with a repetition rate of 1 Hz and a pulse energy of about 17 mJ. The laboratory models require 208 VAC electrical services of at least 20 A and may require external water cooling or at least a heat exchanger. The hand-held versions are air cooled, and can be operated from batteries or low voltage direct current sources.

3.3 Methods of Spectral Resolution

The basis of a LIBS measurement is the collection and analysis of an emission spectrum. The emission lines of the elements are tabulated in various sources.[10,11] Examples of emission spectra from three different samples are shown in Figure 3.

All samples contained the element Si having a strong emission at 288.1 nm but each sample differed in the number of other elements present as major and minor species. In general, the greater the number of elements the more complicated the spectrum. The simplest spectrum is that of Si in water [Figure 3a]. Here the Si line appears alone without interferences from either H or O which have few emission lines making the emission spectrum particularly simple. Silicon in aluminum metal is readily observed in Figure 3(b) although lines due to Mg, Al, and Fe are also present. Steel represents a complex matrix because of the large number of Fe lines and steels typically contain a large number of other elements. The Si line from a steel sample is apparent in Figure 3(c) but because it is adjacent to an Fe line it is not completely resolved in this spectrum. Using a spectral resolution method having greater resolving power, however, the Si can be separated to some extent from the adjacent Fe line as shown in Figure 3(d). The spectra shown in Figure 3 demonstrate that the complexity of the sample will determine the method of spectral resolution needed to monitor the

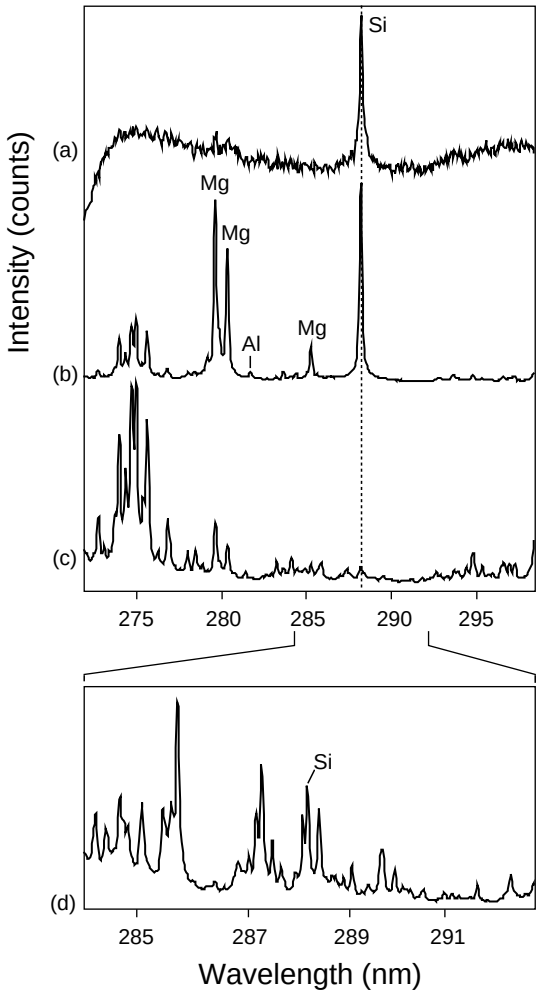

Figure 3 Emission line of Si (288.1 nm) observed from different matrices using LIBS. Spectra (a), (b), and (c) were obtained using a spectrograph of moderate resolution. (a) Si in water. (b) Si in aluminum alloy. (c) Si in steel. (d) Si in steel obtained using 3× greater spectra resolution than shown in (c).

element(s) of interest. Methods of providing spectral discrimination of the plasma light are listed below.[12]

- Narrow bandpass (<1 nm) fixed wavelength line filter.
- An acousto-optic tunable filter (AOTF) consisting of a crystalline material (e.g. TeO_2) to which a radio-frequency wave is applied.[13] By adjusting the frequency of the wave, the bandpass of the AOTF can be varied continuously over a certain range.
- A monochromator that is tuned to monitor a single wavelength and this wavelength is presented at the exit slit of the device for detection.
- A spectrograph is similar to a monochromator except it has an exit plane at which a continuous range of wavelengths is presented for detection using some type of AD or a series of single wavelength detectors positioned behind individual slits.

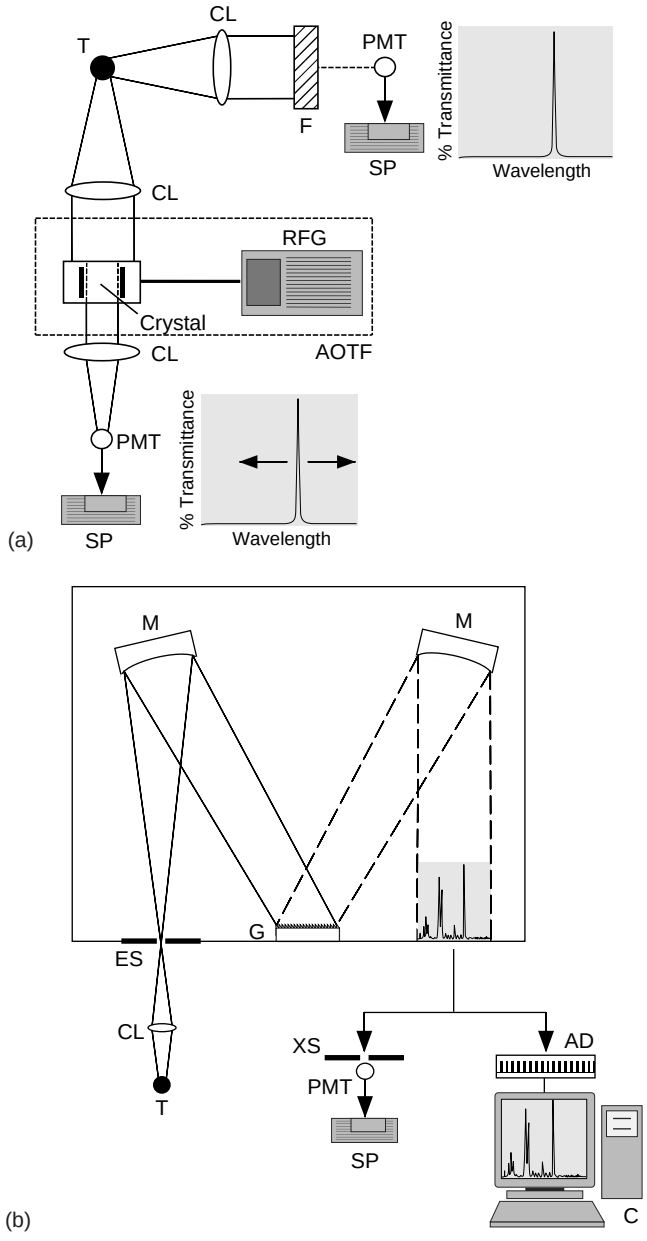

Figure 4 Methods of spectral resolution of the plasma light using (a) filter-based methods and (b) spectrometer-based methods. T = target; CL = converging lens; F = filter; photomultiplier tube (PMT); SP = signal processing; RFG = radio-frequency generator; AOTF = acousto-optic tunable filter; M = mirror; G = grating; ES = entrance slit; XS = exit slit; AD = array detector; C = computer.

Diagrams of methods (1) and (2) are shown in Figure 4(a). In the case of the narrow bandpass filter or AOTF, only a single narrow wavelength band is passed through the wavelength selective element. The

transmitted light is then detected using some sort of light detector (section 3.4). The advantage of the fixed wavelength filter is very small size and low weight and cost. The AOTF, on the other hand, is somewhat larger and requires a power supply, but it can be tuned to monitor different wavelength regions. The use of a monochromator or spectrograph is diagrammed in Figure 4(b). The monochromator can be considered a spectrograph with a slit at the exit focal plane so that only a single narrow wavelength region is monitored at one time. The spectrograph, on the other hand, provides simultaneous detection over a wide spectral range because the single exit slit is removed and an AD is used to record the spectrum or a series of exit slits each with its own detector is placed along the focal plane.

The most suitable method of spectral resolution depends mainly on the analysis requirements. Factors that must be considered include (1) the complexity of the sample (i.e. how many elements are in the sample and do these elements have many emission lines or only a few strong lines?), (2) the number of elements to be monitored, (3) whether the elements are to be monitored simultaneously or is sequential detection adequate, and (4) the location of the emission lines in the spectrum.

Although not all possible combinations of the spectral resolution methods can be considered here because of the wide range of factors that will influence the final choice, Table 1 summarizes the capabilities of each method. It should be noted that in the case of the fixed wavelength methods a second, third, etc. fixed wavelength system may be added to provide simultaneous multiple wavelength detection capabilities.

If only a single element or a few elements are to be monitored then narrow bandpass fixed wavelength filters may be the method of choice in terms of simplicity and small size. These filters can be manufactured having a narrow bandpass (down to 0.1 nm) to selectively monitor an element line in complex spectra as shown in Figures 3(c) and 3(d). If, on the other hand, multiple wavelengths are to be monitored using a spectrograph, then the spectral resolution and spectral coverage provided by the instrument must be considered. As spectral resolution increases, spectral coverage decreases. For example, in Figure 3(c) a fairly wide spectral range is shown. However, if Si is the element of interest, there appears to be an interference with an adjacent line. Using a spectrograph having resolution three times greater, results in the spectrum shown in Figure 3(d) with the degree of overlap between the Si line and the adjacent line reduced. There is, however, a loss in the spectral coverage so that element lines at both extremes of the spectrum shown in Figure 3(c) will not be recorded with the instrument of higher resolution. In the case of the monochromator and spectrograph, the spectral resolution or ability to resolve adjacent emission lines is determined by the focal length of the device and the number of lines per millimeter on the grating.

3.4 Detectors

The method of spectral selection determines the type of detector used for LIBS measurements.[14,15] The simplest detector consists of a photosensitive material that generates a signal proportional to the amount of light incident on the device. These PDs include PMTs and photodiodes. These devices are used with spectral selection methods such as fixed filters, AOTFs, and monochromators. By placing small photosensitive elements (pixels) in either a linear or 2-dimensional array, an AD is produced that provides spatial information concerning the light pattern incident on the array. Common examples of these ADs include photodiode arrays (PDA), charge coupled devices (CCD) and charge injection devices (CID). ADs are used with spectrographs to record the continuous spectrum presented at the focal plane of the instrument. The spectra shown in Figure 3 were obtained using a CCD detector. Each firing of the laser produced a spectrum.

The type of detector determines the method of signal processing. The signal from the PMT, for example, consists of a current that is converted to a voltage by the recording device. The response of a PMT is typically very fast, less than a few nanoseconds, so it can be used to record the temporal variation of the plasma light at the selected wavelength. If multiple wavelengths must be monitored simultaneously, several slit/PMT assemblies

Table 1 Comparison of methods of spectral resolution

Method	Wavelength region		Simultaneous detection of multiple wavelengths	Detector
	Fixed	Tunable		
Filter	X		X (using multiple systems)	PD
AOTF	X	X	X (using multiple systems)	PD
Monochromator	X	X		PD
Spectrograph	X	X	X	AD or multiple PDs

PD, photodetector.

can be placed in the focal plane of a spectrograph with the slit positions aligned to the emission peaks of the elements of interest. Time-resolved detection of an element signal is obtained by electronic processing of the PMT signal using a sample-and-hold circuit.

The spectral coverage provided by an AD is determined by the physical size of the array and the characteristics of the spectrograph. The PDA is a one-dimensional arrangement of diodes that provides spatial intensity information in one dimension. Typical spacing between individual photodiodes is 25 microns so an array of 1024 pixels has a physical length of about 25 mm. The CCD and CID, on the other hand, are two-dimensional arrays of photodiodes that can provide intensity information along two axes. Typical pixel sizes range from 9×9 microns up to 24×24 microns and array formats range from 576×384 pixels up to 3072×2048 pixels with a large number of other formats in between these extremes.

A generalized method of processing the plasma light using a detector is shown in Figure 5. The short duration laser pulse generates the plasma that typically lives for many microseconds after formation. Using either a PD or AD, the plasma light can be integrated over the entire emission period of the plasma. The resultant signal is shown in Figure 5.

In the case of a PMT, the detector current is stored on a capacitor that charges up to a certain voltage as light strikes the detector. In the case of an AD, the charge is stored on the pixel and then is read out at a later time. The magnitude of the voltage or stored charge is proportional to the number of photons incident on each detector. At the end of the period of light collection, the voltage or charge is read out and the signal levels are reset to zero for the next pulse. Instead of integrating the plasma light over the entire plasma lifetime, it may be useful to only monitor a certain period of light emission as the plasma decays. An example would be to discriminate against the intense white light that emanates from the plasma during the first 1–2 μs. Generally, a gate pulse is applied to the detection electronics that determines the time after plasma formation at which recording of the plasma light is to begin (T_d) and the length of time over which the light is collected (T_b). In the case of a PMT, the signal is processed by a simple circuit such as a gated integrator that only stores the current from the PMT according to the timing designated by an applied gate pulse. When the gate pulse is high, the current is stored. In the case of an AD, the gate pulse is directed to a microchannel plate image intensifier that only transmits light through to the photosensitive area of the PDA, CCD, or CID when the gate pulse is present.

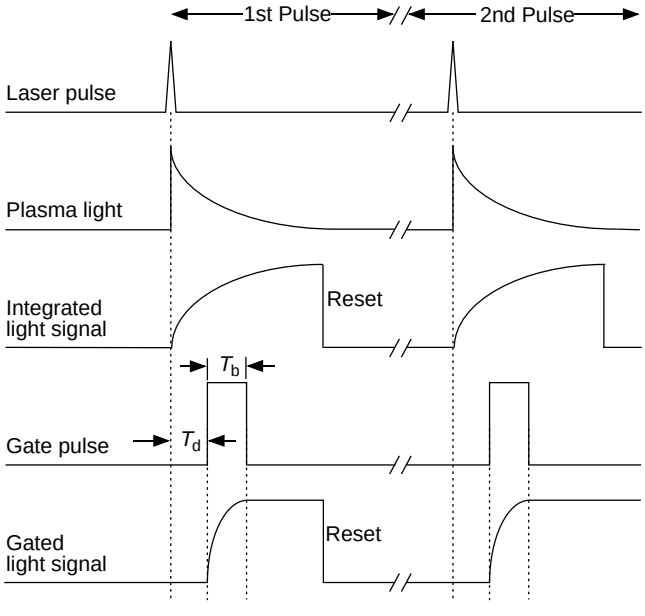

Figure 5 Methods of recording the plasma light during a LIBS measurement. The plasma light decays slowly after the laser pulse. The integrated light signal is the total light recorded by the detector. The gated signal only monitors the light during the gate pulse. After recording the light signals the signal levels are reset to zero to await the signal from the next pulse.

3.5 Examples of Instruments

3.5.1 Laser Microprobe

The first instrument developed for the commercial market based on the laser spark was the laser microprobe.[16] This device combined laser ablation of a solid sample with the spark produced by conventional metal electrodes (Figure 6). In this laboratory-based instrument, the ablating laser pulse produced an aerosol of material above the sample surface that was then interrogated by the spark

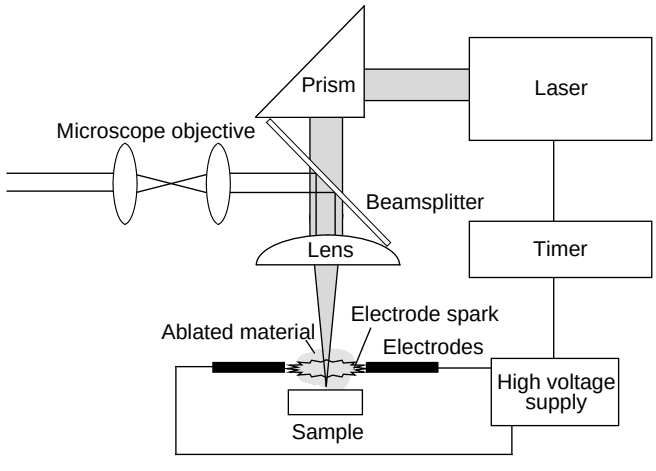

Figure 6 Laser microprobe instrument in which the laser pulse is used to ablate material from a surface that is subsequently analyzed using the spark produced by a pair of electrodes.

produced by the electrodes. These types of instruments were made by several manufacturers and were on the market for several years. Detailed reviews of applications of these instruments can be found in the literature.[16] Some additional information on the capabilities of the technique can be found in section 6.2.

3.5.2 Field-portable Instrument

Because of recent developments in miniaturization of components used for LIBS, it has been possible to fabricate truly portable LIBS-based instruments.[17] An example is shown in Figure 7. This device consists of a main analysis unit connected to a hand-held probe. The probe contains the laser, focusing optics, and a fiber optic cable to collect the plasma light. The light is transported back to the analysis unit that houses the small spectrograph, detector, and laser power supply. The instrument is operated through the use of a micro or laptop computer. An analysis is carried out by placing the probe on the sample and then firing the laser. The laser can be repetitively fired to average the spectra from many shots and increase measurement accuracy and precision.

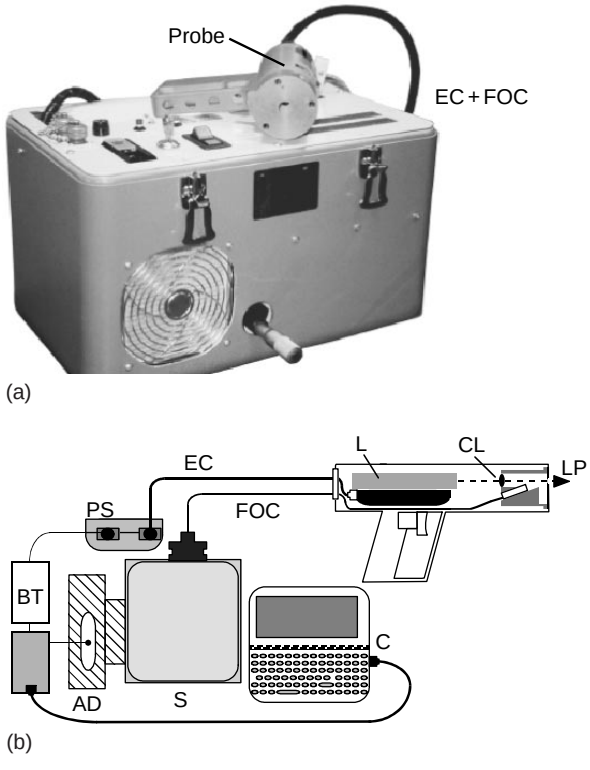

Figure 7 (a) Photograph of field-portable LIBS analyzer and (b) diagram showing main components of device. PS = power supply; EC = electrical cable; L = laser; CL = converging lens; LP = laser pulse; BT = battery; AD = array detector; S = spectrograph; FOC = fiber optic cable; C = computer. (Photograph supplied courtesy of CMEI, Inc.)

The resulting spectrum is then analyzed via software to determine the element signals. Quantitative analysis is possible through calibration of the instrument using calibration standards of known composition.

4 CHARACTERISTICS OF THE LASER PLASMA

The characteristics of the LIBS plasma (e.g. temperature, electron density, size) depend on the state of the sample (gas, liquid, solid) and other parameters such as the pressure and composition of the surrounding atmosphere. Here we present a brief overview of the properties of the laser plasma for different samples.

4.1 On Solids

Solid samples were among the first to be quantitatively analyzed with LIBS. Because most solid targets are opaque at the laser wavelength, the laser plasma is formed on the surface of the solid even though the distance between the target and the focusing lens may be different from the focal length of the lens. As long as the power density is above a threshold value, a laser plasma will be formed. This situation is different from that which occurs in the analysis of a transparent gas or liquid. In these cases, the laser plasma is always formed in the focal volume, the position of which is determined by the focal length of the lens. Because there are a range of lens-to-sample distances (LTSD) over which the plasma can be formed on a solid, the plasma parameters such as temperature and electron density and the amount of material ablated by the pulse have a range of values determined by the LTSD.[18]

For a Nd:YAG laser pulse (186 mJ), typical plasma temperatures range from 6000 to 9500 K with electron densities in the range of 10^{17} to 10^{19} cm^{-3} immediately after plasma formation. The mass of material ablated from aluminum metal ranges from 5 to 80 ng pulse^{-1}. These parameters vary within the specified limits and depend on factors such as ambient pressure and LTSD. As the plasma cools, the temperature and electron density both decrease. Using the white light from the plasma as an indicator, the temporal duration of the plasma ranges out to about 10–20 μs if formed on a metal in air at atmospheric pressure.

4.2 In Gases

The laser spark phenomenon was first discovered by the breakdown of gases shortly after the discovery of the laser. However, it was not until time-resolution was applied that quantitative measurements could be carried out. The LIBS technique allows for the analysis of gases and

combustion products as well as the analysis of aerosols and suspended particles in a gas. The typical precision for the analysis of gases is around 5–8% RSD (relative standard deviation). This value is considerably elevated in the analysis of aerosols and airborne particles to around 30% and is probably due to shot-to-shot differences in sampling particles of different sizes and having different locations in the plasma volume.

The spark formed in a gas has a much higher temperature than that formed on a solid. Although the temperature is dependent upon the ambient gas, values up to 20 000 K have been measured as the initial temperature of the plasma. The plasma cools with time to values as low as 12 000 K at 10 μs after plasma formation. The electron density of this high temperature plasma is about the same as measured for solid samples.

4.3 In Liquids

The spark produced in a bulk liquid is significantly different from that produced in a gas or on a solid.[19] The plasma is visibly much smaller and the temporal decay of light emission occurs more rapidly such that useful spectral emissions are absent for times >2 μs. In addition, there is significant temporal overlap between emissions from neutral and ionized species and also simple molecules formed in the recombining plasma so that time resolved detection is less useful for the spark in a liquid.

The electron density of the plasma formed in water is on the order of 10^{18} to 10^{19} cm^{-3} with temperatures in the range of 7000–12 000 K.

5 ANALYTICAL CAPABILITIES

5.1 Advantages

The analytical capabilities of a method are usually specified by a set of measurable parameters called figures-of-merit. The parameters we use here are (1) the limit of detection (LOD) for an element in a specified matrix and measurement, (2) precision, and (3) accuracy. Methods of computing the LOD can be found in the literature.[20,21] Here we use LOD = 3(SD)/m where SD is the standard deviation of replicate measurements of the element signal using the same sample and m is the slope of the linear calibration curve of element signal or referenced element signal (see section 5.2.2) vs. element concentration. Precision refers to how reproducibly an element signal or concentration can be measured and it is specified in terms of % RSD determined by repeating a measurement several times (preferably > 6), determining the average of the element signals (S_{ave}), and then computing %RSD = (SD/S_{ave}) × 100%. Accuracy refers to how close the predicted concentration of an element in a sample (C_{pred}) is to the actual concentration (C_{act}). The figure-of-merit here is %Accuracy where

$$\%\text{Accuracy} = \frac{100\%|C_{act} - C_{pred}|}{C_{act}}.$$

5.1.1 Range of Analysis Scenarios

Because the laser spark is formed by focused optical radiation, in addition to the conventional LIBS apparatus shown in Figure 2, several different analysis scenarios can be implemented. Some of these are shown in Figure 8 and have been incorporated into instruments.

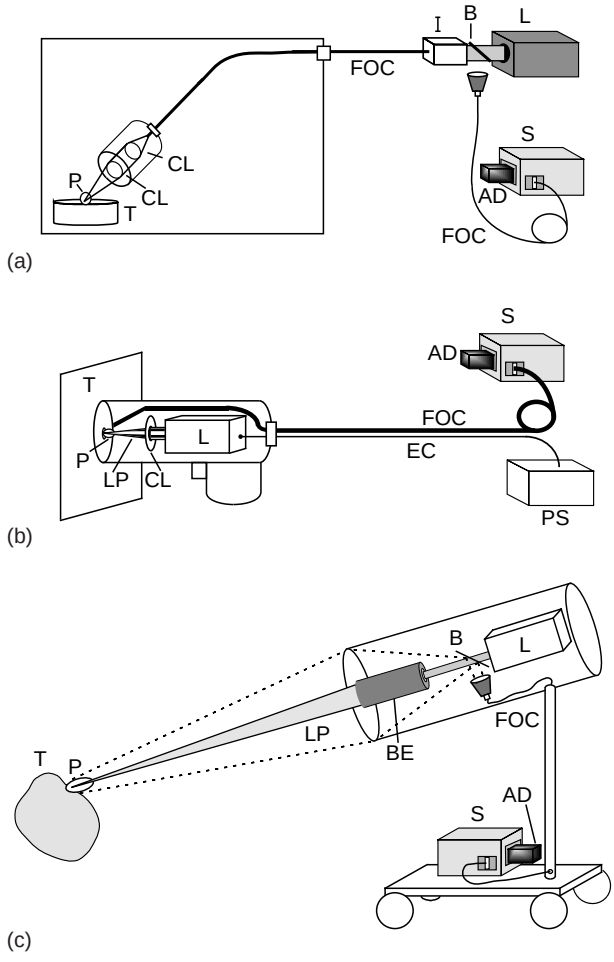

Figure 8 LIBS analysis of materials using different sampling methods. (a) Transport of the laser pulse to the remotely located sample using a fiber optic cable. (b) Incorporation of a small laser into a probe connected to the main analysis unit of a LIBS instrument through cabling. (c) Stand-off analysis in which the laser pulses are focused at a distance onto the remotely located sample. L = laser; B = beamsplitter; I = injector; FOC = fiber optic cable; CL = converging lens; P = plasma; T = target; S = spectrograph; AD = array detector; EC = electrical cable; LP = laser pulse; PS = power supply; BE = beam expander.

5.1.1.1 Direct Analysis This configuration, shown in Figure 2, is used for most LIBS measurements. Here a short focal length lens is used to focus the laser pulses onto the sample (solid, liquid) or into a liquid or gas to form the plasma. The plasma light can be collected using a lens which focuses the light onto a spectrograph slit or collimates the light to pass it directly through a bandpass filter or other frequency selective device (e.g. AOTF). As an alternative, a fiber optic cable can collect and transport the light to a remotely located frequency selective device, typically a spectrograph. This analysis method represents the simplest embodiment of a LIBS apparatus.

5.1.1.2 Fiber Optic Delivery With the development of improved fiber optic materials it is now possible to focus power densities on the order of MW cm^{-2} onto the end of a fiber, without damage, and inject and then transport tens of millijoules of energy through the fiber [Figure 8a].[22] By placing a lens system at the distal end of the fiber, the laser pulse can be focused to produce a spark. Transport of laser pulses over distances up to 100 meters has been demonstrated. Depending on the analysis requirements (e.g. the elements to be monitored, concentrations, etc.) the plasma light can be collected and transported back to the detection system using either the same fiber or a second fiber optic cable.

5.1.1.3 Compact Probe By combining a very compact laser with fiber optics it is possible to construct a small probe to use for remote LIBS measurements [Figure 8b].[17] The laser power supply and detection system can be located remotely from the probe, connected by an umbilical cable containing electrical cables for the laser and a fiber optic cable to transport the plasma light. This configuration is used in at least one commercial LIBS unit. Compared to fiber optic delivery, this method has the advantage that spot sizes of small diameter and hence greater power density can be delivered to the target enhancing the element signal.

5.1.1.4 Stand-off Analysis Because the laser plasma is formed by focused light, it is possible to generate the laser spark at a distance on a remotely located sample [Figure 8c].[23] The distances that can be achieved are a function of many parameters including the laser pulse energy and power, the beam divergence, and the optical system used to focus the pulses at a distance. With good quality components, for example, a laser plasma can be formed on soil at a distance of 19 meters using only 35 mJ $pulse^{-1}$. In such a system, efficient collection of the plasma light is critical to obtain useful signals. Unlike the other analysis scenarios described above in which a bare fiber can be aimed at the plasma to collect sufficient light, in the case of stand-off analysis, a lens is required to increase the solid angle over which the plasma light is collected and then directed into the fiber. The light collection system can be either adjacent to or collinear with the optical axis of the system used to focus the pulses on the sample. The latter configuration is shown in Figure 8(c).

5.1.2 Minimal Sample Preparation

Because of the high power densities used to form the laser spark, all types of materials can be atomized and a plasma formed in the focal region. Only optical access to the sample is required. Within the laser plasma additional atomization of ablated material continues and the resulting atoms are excited to emit light. For this reason, in general, no sample preparation is required for a LIBS measurement which is a distinct advantage compared to most other forms of analysis. For example, analysis using the ICP generally requires that the sample be chemically ashed to produce a solution that is then nebulized into the ICP plasma. Chemical preparation is a time-consuming and sometimes labor intensive process that precludes rapid sample analysis. Using LIBS however, a measurement can be carried out immediately merely by focusing the laser pulses directly on the sample.

5.1.3 Speed of Analysis

LIBS measurements can be carried out in what is considered real-time because of the lack of sample preparation and the simplicity of the method which make it amenable to deployment in the field. Due to the short lifetime of the plasma (few tens of microseconds), the time required to record a spectrum is less than 100 milliseconds and using automated instrumentation driven by computer software, the analysis of the resulting spectrum is immediate. In this case, the main factor determining the speed of a measurement is the number of laser shots required to obtain a representative spectrum. Many shots may be needed to:

(1) obtain an average composition reading due to compositional inhomogeneity of the sample;

(2) ablate away an overlying surface layer having a composition that may not be representative of the underlying bulk material;

(3) average out shot-to-shot variations in the plasma characteristics.

The use of multiple pulses to overcome sampling problems associated with (1) and (2) above is obvious and is discussed in section 5.2.1. Advantages to be gained from using many laser shots in the analysis of even a uniform sample when plasma characteristics are changing shot-to-shot can be demonstrated by performing

replicate measurements, each consisting of 50, 200, and 1600 laser pulses, and then computing the precision of the measurements. For uranium in solution, the results obtained in one set of measurements were 13.3, 7.2, and 1.8% RSD, respectively.[24] These measurements were carried out with repetitive laser plasmas formed on the surface of a liquid in which the lens-to sample distance and hence plasma characteristics changed on each shot due to strong pressure waves generated at the liquid surface. The strong dependence of precision on the number of averaged laser pulses demonstrates the advantage of repetitive measurements.

LIBS measurements made using uniform geological samples (i.e. certified reference materials such as those available from the National Institute of Standards and Technology), metals, and liquid samples indicates that 100 laser shots can produce measurement precision on the order of 10%.

5.2 Considerations in the Use of Laser-induced Breakdown Spectroscopy

5.2.1 Sample Homogeneity

One of the more appealing aspects of LIBS is the ability to analyze samples with little or no preparation. For samples such as mixed gases and liquids containing well mixed and dissolved materials, sample homogeneity may be assumed so every plasma interrogates a small volume having a composition representative of the bulk sample. In this case, the number of laser shots to be averaged for a measurement is determined by factors such as the method of sampling and perturbations of the laser plasma. Other samples, solids in particular, cannot always be assumed to be homogeneous and, in fact, except for metals and plastics, etc. inhomogeneity should be assumed. Two types of inhomogeneity likely to be encountered are listed below.

5.2.1.1 Bulk Non-uniformity The small area interrogated by the laser pulse represents point detection of an area of 0.1 to 1 mm diameter involving a very small mass of material (i.e. tens of nanograms, section 2.3). Surface features on rocks, for example, may display visual irregularities in the distribution of materials that are on the order of the area sampled on each shot. These non-uniformities may be averaged out using a number of laser plasmas to repetitively interrogate a sample with the results then averaged.

5.2.1.2 Non-representative Surface Composition Some samples, such as metal alloys, and rocks exposed to the elements may have a surface layer composition that is not representative of the underlying bulk composition. For example, weathered rocks usually have a desert varnish layer ranging from 30 to 100 microns thick that has a composition different from the bulk rock matrix. Depending on the laser parameters, each interrogation by the laser pulse will produce a sampling depth ranging from a few microns up to perhaps 10 to 20 microns. Therefore, to obtain a more representative analysis, repetitive sampling by the laser spark at the same location on the sample can be used to ablate away the outer layers revealing the true bulk composition underneath.

For samples that are significantly inhomogeneous either as a result of bulk or surface non-uniformity, it may be advisable to grind the samples and then press the resulting particles to produce a flat surface for analysis. Although this procedure eliminates the real-time and in-situ analysis advantages of LIBS, it still preserves analysis capability without the need for chemical ashing of the sample.

5.2.2 Matrix Effects

LIBS, as other analytical methods, displays so-called matrix effects. That is, the composition of the sample affects the element signal such that changes in concentration of one or more of the elements forming the matrix alter an element signal even though the element concentration remains constant. For example, the signals from Si in water, in steel, and in soil appear much different even though the concentration of the element is the same in all three matrices. Even for samples that are more closely allied in matrix composition such as soils and stream sediments, significant differences in signal levels are observed for elements at identical concentrations in these materials. Because the laser spark both ablates and excites the sample, these effects can be more pronounced than in other methods that require sample preparation. Matrix effects can be divided into two kinds, physical and chemical.

Physical matrix effects depend on the physical properties of the sample and generally relate to the ablation step of LIBS. That is, differences between the specific heat, latent heat of vaporization, thermal conductivity, absorption, etc. of different matrices can change the amount of an element ablated from one matrix compared to another matrix even though the properties of the ablation laser pulse remain constant. Changes in the amount of material ablated can often be corrected for by computing the ratio of the element emission signal to some reference element known to be in the matrix at a fixed or known concentration. In this case, it is assumed that the relative ablated masses of the element and reference elements remain constant although the total mass of ablated material may change on a shot-to-shot basis. In this case, calibration is provided by using the ratio of the element signal to the signal produced by the reference element.

Chemical matrix effects occur when the presence of one element affects the emission characteristics of another element. This can complicate calibration of the technique and hence the ability to obtain quantitative results. These effects can be calibrated out of the analysis if the concentration and effect of the interfering element(s) are known but changes in the concentration of the interfering species from sample-to-sample can be a difficult correction procedure. An example of a chemical matrix effect is the reduction in emission intensity of an ionized species [e.g. Ba(II)] upon a significant increase in the concentration of an easily ionizable species in the matrix. The easily ionizable species perturbs the electron density thereby decreasing the concentration of Ba(II).[25] In addition, there are indications that the compound form of an element [e.g. $PbNO_3$, $PbCl_2$, etc.] may result in different emission signal strengths for the same element concentrations.

Extensive work remains to be done to characterize chemical and physical matrix effects for all types of samples and to develop methods to correct for their effects and increase the quantitative ability of LIBS.

5.2.3 Sampling Geometry

In the analysis of a solid, a plasma will be formed on the surface if the power density is sufficiently high even though the distance between the sample and the lens may be different from the focal length of the lens. These changes in the LTSD can result in changes in the mass ablated as well as changes in the temperature and electron density of the plasma which in turn affect the element emission signals.[18] Keeping the sampling geometry constant is important to achieve the best analytical results. In the interrogation of some samples, such as soil or rocks on a conveyor belt, for example, maintaining the LTSD constant may not be possible to a high degree. This can be dealt with in several ways including the use of a lens of long focal length to focus the pulses on the sample so that relative changes in the LTSD are less important or developing an active feedback system to automatically change the lens position to keep the LTSD constant. In addition, it is often possible to compute the ratio of the element emission to the emission of some reference element (e.g. Fe in steel or in soil) known to be in the sample at a fixed concentration. In this case, relative changes in both signals are the same and the ratio remains constant as described in section 5.2.2.

5.2.4 Safety

There are three areas of safety that must be considered in the use of LIBS. These are (1) the ocular hazard posed by the laser pulse, (2) the potentially lethal high voltage circuits used by the laser, and (3) the explosive potential of the laser spark for certain materials. The first two hazards are discussed at length in the industry safety standard ANSI Z136.2 to which one is referred for further information.[26]

The lasers used for LIBS are almost exclusively typed as Class IV indicating that both eye and skin exposure should be avoided. Established safety procedures require that precautions be taken to ensure that personnel are not exposed to the laser radiation through proper administrative and engineering controls. Engineering controls include interlocks connected to an enclosure that if opened cause immediate deactivation of the laser. The safest procedure is to completely enclose the laser radiation so that operations can be carried out as if the LIBS system contained a Class I laser in which case no extraordinary safety measures are needed outside the enclosure. In most cases, it should be possible to completely enclose the laser to protect the operator against the laser light. An exception may be the stand-off sampling geometry shown in Figure 8(c). For Nd : YAG lasers of the powers normally used for LIBS measurements, the main hazard is eye exposure. Exposure of the skin should be avoided but rarely results in significant damage at the pulse energies used for LIBS. On the other hand, the use of lasers producing radiation in the ultraviolet spectral region, such as the excimer laser, can pose a significant hazard to the skin. If interlocks must be defeated to work on an operating system or for alignment purposes, eye protection is easily provided for by the use of goggles appropriate for the laser wavelength in use.

The high voltages used by all lasers represent a lethal hazard that is most easily avoided by operating the laser with all interlocks to electrical circuits maintained in place. Only qualified personnel should be permitted to work on the high voltage power supplies in either an energized or non-energized configuration. Industry procedures such as lockout/tagout are called for in many cases involving work on electrical systems of this magnitude.

The laser plasma, as any other type of plasma source, has the potential to ignite flammable gases or to cause detonation of explosives. For this reason, the environment in which LIBS measurements will be carried out must be evaluated regarding this risk.

5.3 Representative Figures-of-merit

Because LIBS analyzes a sample directly with little or no preparation, the analytical figures-of-merit of the method are strongly dependent on sample characteristics. Listed in Table 2 are some representative LOD values for elements in different matrices. For gases, liquids, and solids the detection limits are in units of ppm

Table 2 Representative detection limits for elements in selected matrices

Element	Gas (ppm)	Liquid (ppm)	Filter (ng cm^{-2})	Solid/Matrix (ppm)
Ag		20	17	
Al				130 (Fe ore)
As	0.5 (aerosol)		44	
B		80		
Ba		6.8; 130	1	3 (soil)
Be	0.0006 (aerosol)	10	1	0.5 (soil)
Ca		0.13		1300 (organic material)
				300 (Fe ore)
Cd	0.018 (aerosol)	0.8; 500	300	60; 90 (soil)
				20 (silica gel)
Cl	8 (freon)			
Cr			160	10; 150 (steel alloy)
				30 (soil)
				40 (silica gel)
Cs		1		
Cu			11	10; 38 (soil)
				7; 100 (steel alloy)
F	38 (freon)			
Fe		0.16		
Hg	0.5 (aerosol)		15	300 (soil)
K	1.5 (aerosol)	1.2		
Li		0.006; 0.013		2.5 (soil)
Mg		25; 100		230 (Fe ore)
Mn			115	160; 210 (steel alloy)
Mo				120; 200 (steel alloy)
Na	0.006 (aerosol)	0.0075; 0.014		
Ni			185	12; 20 (soil)
				64; 150 (steel alloy)
				310 (silica gel)
P	1.2 (aerosol)			
Pb	0.2 (aerosol)	12.5	450	17; 40 (soil)
				96 (leaded paint)
Rb		0.2		
S	200 (aerosol)			
Sb				73 (soil)
Si				1190 (Al alloy)
				380 (steel alloy)
				1500 (Fe ore)
Sn				26 (soil)
Sr			5	1.2 (soil)
Ti				44 (Al alloy)
				230 (Fe ore)
Tl			40	
U		100		1000 (soil)
V				200 (steel alloy)
Zn	0.23 (aerosol)		135	250 (soil)

(parts-per-million wt wt^{-1}) whereas for materials on filters, the LOD values refer to a surface detection limit with units of mass per surface area.

6 HYBRID TECHNIQUES

6.1 Double-pulse Techniques

A typical LIBS measurement is made by repetitively sampling a surface with a specified number of laser pulses and then averaging the element signals together. The time separation between the repetitive pulses, determined by the maximum repetition rate of the laser, is large compared to the lifetime of the plasma. This type of measurement is carried out using the RSS (repetitive single sparks) method, defined as such when the time separation between laser pulses is so large (e.g. of 1 ms) that interpulse effects are negligible. The RSS method is used most widely for both field and laboratory LIBS analyses. However, in some cases, the element signal

can be strongly enhanced by the use of a technique termed RSP (repetitive spark pair).[19] The RSP method is defined as interrogation of a target by dual laser pulses having a time separation on the order of tens of microseconds or less so that some interpulse effects are present. An interpulse effect is defined as some effect induced by the first pulse of the pulse pair that still resides in the focal volume interrogated by the second pulse. An example is ablated material from the first pulse still residing in the focal volume during formation of a plasma by the second pulse of the pulse pair. In general, the dual pulses of the RSP are focused by the same optical elements so the focal volumes of the two pulses are coincident.

The timing between the laser pulse(s) and the gate pulse to the detector for the RSS and the RSP techniques is shown in Figure 9. Here ΔT is the time between pulses of the RSP. In the analysis of metals, the greatest enhancements were observed for ΔT of about 5 μs.[27] In the RSS technique, the time delay T_d is referenced to the single pulse used to excite the sample at the laser repetition rate. In an RSP measurement, the parameters T_d and T_b are referenced to the second pulse of the pulse pair as shown. For soil and metal samples at atmospheric pressure, when identical timing parameters T_d and T_b were used to compare the RSS to the RSP, enhancements in element signals up to factor of 15 were obtained. The mechanism responsible for the signal enhancement seen with the RSP in these cases is that the first laser pulse ablates particles into the focal volume of the second pulse that can then more efficiently excite the ablated material.

Another use of the RSP is in the analysis of a bulk metal surface under water.[28] Using the RSS, no element signal is obtained as shown in Figure 9(b). However, interrogation of the metal using the RSP results in the appearance of a strong spectrum shown in Figure 9(d). The reason for the enhancement is that a large bubble is formed on the metal surface by the first pulse allowing the second pulse to form inside the bubble creating conditions similar to an analysis conducted in air. The cavity created from the first pulse depends upon laser energy but is typically about 8 mm in diameter.

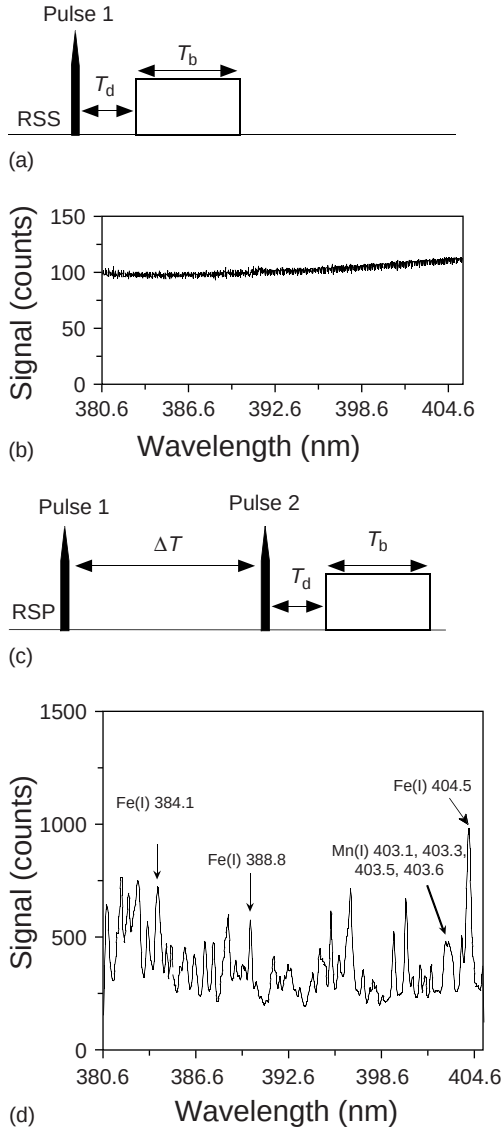

Figure 9 Use of double-pulse excitation to enhance detection capabilities of LIBS. Timing used in (a) RSS and (c) RSP measurements. Spectra obtained using (b) RSS and (d) RSP excitation of metal located under water. Here ΔT is the time between dual pulses, T_d is the delay time between a pulse and the start of the gate pulse and T_b is the duration of the gate pulse.

6.2 Combined with Other Methods

The use of the laser spark alone as an analytical technique has been discussed above. The ability of the laser pulse to ablate a wide range of materials resulting in atomized material has led to its combining with other analytical techniques. In most combined techniques, the laser spark serves to introduce the sample to the other technique in a vaporized form. This introduction of the vaporized sample eliminates the time-consuming sample preparation step allowing for rapid analysis. Another useful characteristic is that the vaporized sample material can be transported in a gas stream to a desired location. In this case, analysis of a large specimen can be completed without having to introduce the large sample into the analytical equipment since the laser spark ablates only a few nanograms of material. Furthermore, all types of solid materials can be introduced.

The first of the techniques discussed here is that of the laser microprobe with cross excitation referred to earlier (section 3.5.1).[16] This technique was developed

in 1962 and is the oldest laser plasma analytical device. A typical microprobe apparatus is shown in Figure 6. A laser pulse is focused onto the sample surface using a microscope objective. A collinear visual imaging system is incorporated into the instrument design to allow for accurate sample placement under the laser pulse. Directly above the sample surface is a pair of electrodes used to excite the ablated material. The light from the electrical discharge is transmitted to the spectrograph.

While the light from the laser plasma alone can be used to make analytical determinations in the microprobe devices, the combination of the electrode spark offers some spectrochemical advantages. The spark produced between the electrodes is more intense than the laser spark by a factor of 100 to 1000 times. This results in greater element emission line intensities resulting in greater sensitivity over the excitation provided by the laser spark alone. Also, because the electron density of the electrode spark is lower than that of the laser spark, the line widths are narrower and are less likely to be self-reversed. A third advantage of using the coupled technique is that the electrode spark does not contain as strong a broadband background continuum light.

The laser microprobe with cross excitation has been applied to a greater variety of materials than any other laser plasma-based method. The list of samples analyzed includes minerals, ceramics, glasses, biological materials, oils, crystals, alloys, powders, and geologic materials. The absolute detection capabilities depend on the element and the sample matrix. Typical mass detection limits for many elements lie in the range of 0.1 pg to 500 pg and concentration LOD values range from 10 ppm to 500 ppm for most elements.[16]

The laser microprobe with cross excitation is a well-researched technique and several disadvantages have become evident when compared to other analytical techniques. Since the method uses an electrode spark, the disadvantages associated with electrodes are present such as possible interference from the electrode material and wear of electrodes with use. However, the most important drawback of this technique is poor measurement precision. Typically around 25–40% RSD. Although creative attempts have been made to increase precision, the best achievable precision lies in the range of 3.6 to 16% RSD.

Laser-induced fluorescence (LIF) has been used to probe the atomic constituents of the laser plume.[29,30] In LIF, the free atoms of interest are electronically excited using a dye laser pulse tuned to a specific atomic transition. The relaxation of the excited species produces fluorescence. The geometry of a typical laser ablation/LIF set-up is shown in Figure 10. Some advantages of this method compared to monitoring the plasma light directly include (1) greater concentration and absolute mass detection sensitivity, (2) fewer matrix effects, (3) wider dynamic range, and (4) isotope selectivity.

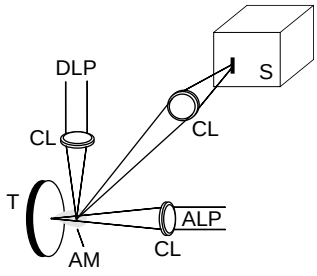

Figure 10 Diagram of the geometry used for LIF detection of laser-ablated material. DLP = dye laser pulse; CL = converging lens; T = target; AM = ablated material; ALP = ablating laser pulse; S = spectrograph.

In one study, Cr was analyzed from steel, flour, and skim milk powder.[29] The calibration curves were linear over 10 ppm to 1000 ppm and no chemical matrix effects were observed. The detection limit for Cr was determined to be 1 ppm corresponding to 0.1 pg of atomized Cr. The uncertainty in the measurements was 20–30% but was attributed to correctable experimental difficulties. Foremost was Mie scattering of the laser radiation from ablated particles in the laser plasma. This can be corrected by using filters to block the laser wavelength if the fluorescence is shifted from the laser wavelength or in the case of resonance fluorescence, by time gating the fluorescence detection. The effects of atmosphere, pressure, and time delay all have an effect on the analytical capabilities of this technique. At lower pressures, the longer lifetimes of the free atoms allow for greater resolution in time gating when resonance fluorescence is being detected.

6.3 Continuous Optical Discharge

The plasma formed by a laser pulse has a relatively low duty cycle (10^{-6}) given that the plasma lifetime is only about 20 μs and a typical laser repetition rate is 20 Hz. It is possible to generate a continuously operating laser plasma, called the continuous optical discharge (COD), by focusing a cw-CO_2 laser beam into air or another gas.[31] Whereas typically megawatts of pulsed energy are required to form the pulsed laser plasma, only hundreds or a few thousand watts are required to form the COD. The main reason for this is that absorption of the incoming laser energy by the plasma is proportional to the square of the laser wavelength making the power requirements of the long wavelength CO_2 laser (10 600 nm) a factor of 100 less then would be required using the 1064 nm wavelength of the Nd : YAG laser. Even though the CO_2 laser provides sufficient power to sustain the COD, a pulsed laser or other method of introducing a spark into

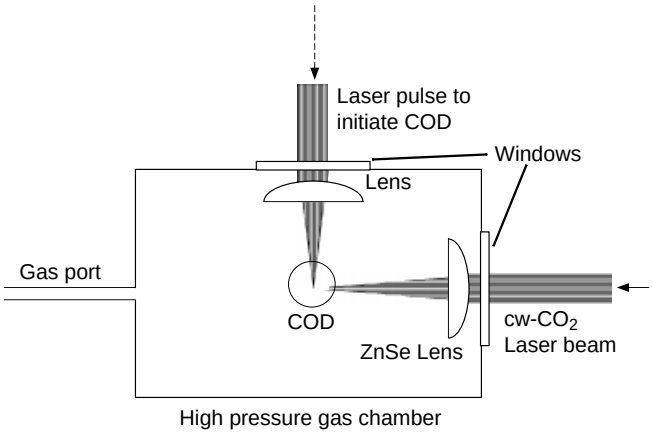

Figure 11 Diagram of the method used to generate the COD in a high-pressure cell.

the focal volume of the CO_2 laser is needed to initiate the COD. A diagram of a cell used to generate the COD is shown in Figure 11.

The laser power required to sustain the COD depends on several factors including type of gas, pressure, and quality of the laser beam. For Xe gas, which has a comparatively low ionization potential, 20 W are needed to sustain the COD at a pressure of 10 atm. While in air, which contains molecules and hard to ionize species, 2 kW are needed to maintain the plasma. For a given laser power, the COD can only be sustained over a certain pressure range for a specific gas. For instance, using a 45 W CO_2 beam in Xe, the operating pressures were between 1.8 atm and 4.2 atm.

The plasma temperature and electron density depend upon the operating parameters and the gas used. For a CO_2 laser power of 45 W, the temperature of the COD sustained in Ar at 16 atm was 13 000 K. The temperature of the COD increased to 18 000 K when a pressure of 2 atm in Ar was used. The type of gas greatly affects the plasma temperature as temperatures as high as 30 000 K have been measured in 5 atm of He. While the temperature is highly depend upon the gas and the pressure, the electron density is somewhat less variant and is typically in the range of 10^{17} to 10^{18} cm^{-3}.

The analytical capabilities of the COD are most easy to evaluate for analytical gas samples because these are easily introduced into a high-pressure gas cell. A linear relationship of the calibration curves for many gases has been demonstrated over several orders of magnitude. However, there seems to be an upper limit on the concentration of the analytical gas that the COD can tolerate without extinguishing. Also, some analytical gases absorb at the wavelength of the CO_2 beam resulting in a significant change in the excitation parameters of the COD. Another problem is that some gases have decomposition products that coat the optical elements used to form the COD. To compensate for these problems, the COD should operate as a plasmatron in which the gases are flowed through the laser beam focus and carried away without effect on the COD. Although the COD may be best suited to analyze gases, solids can be introduced into the COD via laser ablation. In conjunction with the gas stream formerly mentioned, solid samples can be ablated and carried into the COD for analysis. This process has yielded semi-quantitative analytical results using thin coatings of solid samples on a metal substrate ablated with a Nd:YAG laser pulse. Although spectrochemical analysis with the COD is possible, further development is needed to devise a method to reproducibly introduce samples into the COD without perturbing the plasma characteristics.

7 PERSPECTIVE AND FUTURE DEVELOPMENTS

The LIBS technique has been under investigation for almost four decades. Practical application of the method has been concurrent with much fundamental work aimed at characterizing and understanding the laser plasma and laser ablation. A variety of systems have been investigated and in some cases the processes are well understood. In other areas, much work remains. The realization of LIBS instrumentation has a less successful history. Although the laser microprobe as a laboratory instrument was offered commercially, the technology has fallen into disuse being displaced by more capable analytical methods meeting the increasing needs for better analytical results including highly sensitive detection. In the setting of an analytical laboratory, where sample preparation and the size of an instrument are not of significant concern, LIBS does not appear to have a viable role. In recent years there has been a resurgence of interest in the technology, however, for a variety of reasons the most notable of which has been environmental regulation. Often such measurements must be performed outside the laboratory in less than ideal conditions not suitable for more conventional instrumentation. In addition, often such measurements do not require the high accuracy and precision characteristic of conventional methods. In these cases, LIBS may have something to offer. Because of the simplicity of the method, the lack of sample preparation, and the non-invasive nature of the sampling step, LIBS may be the ideal technique. In addition, recent technological developments in computers, lasers, spectrographs, and detectors, resulting in rugged and compact field-portable LIBS devices, have given rise to renewed interest in the method. The activity in the

development of commercial LIBS devices over the past five years is probably greater than the combined activity over the first thirty years of the method. Whether this interest will continue and is realized on a large scale remains to be seen. What is certain, however, is that the method can uniquely meet some analysis requirements and does have a place, perhaps limited, in the arsenal of modern analytical techniques. The key to increasing these applications is a more thorough investigation and hopefully more thorough understanding of the method.

ACKNOWLEDGMENTS

This work was done under the auspices of the US Department of Energy. Los Alamos National Laboratory is operated by the University of California for the US DOE.

ABBREVIATIONS AND ACRONYMS

AOTF	Acousto-optic Tunable Filter
AD	Array Detector
AES	Atomic Emission Spectroscopy
CCD	Charge Coupled Devices
CID	Charge Injection Devices
COD	Continuous Optical Discharge
ICP	Inductively Coupled Plasma
IB	Inverse Bremsstrahlung
LIBS	Laser-induced Breakdown Spectroscopy
LIF	Laser-induced Fluorescence
LSC	Laser-supported Combustion
LSD	Laser-supported Detonation
LSR	Laser-supported Radiation
LTSD	Lens-to-sample Distances
LOD	Limit of Detection
PD	Photodetector
PDA	Photodiode Array
PMT	Photomultiplier Tube
RSD	Relative Standard Deviation
RSS	Repetitive Single Sparks
RSP	Repetitive Spark Pair
SD	Standard Deviation

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Laser-induced Breakdown Spectroscopy, Elemental Analysis

Environment: Water and Waste **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Optical Emission Inductively Coupled Plasma in Environmental Analysis

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Laser Ablation in Atomic Spectroscopy • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

REFERENCES

1. *Laser-induced Plasmas and Applications*, eds. L.J. Radziemski, D.A. Cremers, Marcel Dekker, New York, 1989.
2. L. Moenke-Blankenburg, *Laser Microanalysis*, John Wiley, New York, 1989.
3. D.A. Rusak, B.C. Castle, B.W. Smith, J.D. Winefordner, 'Fundamentals and Applications of Laser-induced Breakdown Spectroscopy', *Crit. Rev. Anal. Chem.*, **27**(4), 257–290 (1997).
4. T.P. Hughes, *Plasmas and Laser Light*, John Wiley and Sons, New York, 1975.
5. G.M. Weyl, 'Physics of Laser-induced Breakdown: An Update', in *Laser-induced Plasmas and Applications*, Marcel Dekker, New York, 1989.
6. R.G. Root, 'Modeling of Post-breakdown Phenomena', in *Laser-induced Plasmas and Applications*, Marcel Dekker, New York, 1989.
7. J.F. Ready, *Effects of High Power Laser Radiation*, Academic Press, New York, 1971.
8. C.R. Phipps, R.W. Dreyfus, 'The High Laser Irradiance Regime', in *Laser Ionization Mass Analysis*, John Wiley and Sons, New York, 1993.
9. O. Svelto, *Principles of Lasers*, Plenum Press, New York, 1976.
10. *Wavelengths and Transition Probabilities for Atoms and Atomic Ions Part II. Transition Probabilities*, NSRDS-NSB 68, U.S. Government Printing Office, Washington, DC, 1980.
11. A.R. Striganov, N.S. Sventitskii, *Tables of Spectral Lines of Neutral and Ionized Atoms*, IFI/Plenum, New York, 1968.
12. K.I. Tarasov, *The Spectroscope*, John Wiley & Sons, Inc., New York, 1974.
13. I.C. Chang, 'Acousto-optic Tunable Filters', *Opt. Eng.*, **20**, 824–829 (1981).
14. *The Photonics Design and Applications Handbook, Book 3*, Laurin Publishing Company, Inc., Pittsfield, MA, 1997.
15. Y. Talmi, (ed.), *Multichannel Image Detectors*, ACS Symp. Series No. 102, ACS, Washington, DC, 1979.
16. H. Moenke, L. Moenke-Blankenburg, *Laser Micro-spectrochemical Analysis*, Crane, Russak and Company, Inc., New York, 1973.

17. K.Y. Yamamoto, D.A. Cremers, L.E. Foster, M.J. Ferris, 'Detection of Metals in the Environment Using a Portable Laser-induced Breakdown Spectroscopy (LIBS) Instrument', *Appl. Spectrosc.*, **50**, 222–233 (1996).
18. R.A. Multari, L.E. Foster, D.A. Cremers, M.J. Ferris, 'The Effects of Sampling Geometry on Elemental Emissions in Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**, 1483–1499 (1996).
19. D.A. Cremers, L.J. Radziemski, T.R. Loree, 'Spectrochemical Analysis of Liquids Using the Laser Spark', *Appl. Spectrosc.*, **38**, 721–729 (1984).
20. C.Th.J. Alkemade, W. Snelleman, G.D. Boutilier, B.D. Pollard, J.D. Winefordner, T.L. Chester, N. Omenetto, 'Review and Tutorial Discussion of Noise and Signal-to-noise Ratios in Analytical Spectrometry. I. Fundamental Principles of Signal-to-noise Ratios', *Spectrochim. Acta*, **33B**, 383–399 (1978).
21. G.D. Boutilier, B.D. Pollard, J.D. Winefordner, T.L. Chester, N. Omenetto, 'Review and Tutorial Discussion of Noise and Signal-to-noise Ratios in Analytical Spectrometry. II. Fundamental Principles of Signal-to-noise Ratios', *Spectrochim. Acta*, **33B**, 401–416 (1978).
22. C.M. Davies, H.H. Telle, D.J. Montgomery, R.E. Corbett, 'Quantitative Analysis Using Remote Laser-induced Breakdown Spectroscopy (LIBS)', *Spectrochim. Acta*, **50B**, 1059–1075 (1995).
23. J.D. Blacic, D.R. Pettit, D.A. Cremers, 'Laser-induced Breakdown Spectroscopy for Remote Elemental Analysis of Planetary Surfaces', in *Proc. Int. Symp.* Spectral Sensing Research, Maui, HI, November, 15–20, 1992.
24. J.R. Wachter, D.A. Cremers, 'Determination of Uranium in Solution Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **41**, 1042–1048 (1987).
25. A.S. Eppler, D.A. Cremers, D.D. Hickmott, M.J. Ferris, A.C. Koskelo, 'Matrix Effects in the Determination of Pb and Ba in Soils Using Laser-induced Breakdown Spectroscopy', *Appl. Spectrosc.*, **50**(9), 1175–1181 (1996).
26. *American National Standard for the Safe Use of Lasers ANSI standard ANSI Z136.2*, American National Standards Institute (most recent issue).
27. R. Sattman, V. Sturm, R. Noll, J., 'Laser-induced Breakdown Spectroscopy of Steel Samples Using Multiple Q-switched Nd : YAG Laser Pulses', *Appl. Phys. D*, **28**, 2181–2187 (1995).
28. A.E. Pichahchy, D.A. Cremers, M.J. Ferris, 'Elemental Analysis of Metals Under Water Using Laser-induced Breakdown Spectroscopy', *Spectrochim. Acta*, **52B**, 25–39 (1997).
29. H.S. Kwong, R.M. Measures, 'Trace Element Laser Microanalyzer with Freedom from Chemical Matrix Effect', *Anal. Chem.*, **51**, 428–432 (1979).
30. I. Gobernado-Mitre, A.C. Prieto, V. Zafiropulos, Y. Spetsidou, C. Fotakis, 'On-line Monitoring of Laser Cleaning of Limestone by Laser-induced Breakdown Spectroscopy and Laser-induced Fluorescence', *Appl. Spectrosc.*, **51**, 1125–1129 (1997).
31. D.R. Keefer, 'Laser-sustained Plasmas', in *Laser-induced Plasmas and Applications*, Marcel Dekker, New York, 1989.

Microwave-induced Plasma Systems in Atomic Spectroscopy

José A.C. Broekaert
University of Leipzig, Leipzig, Germany

Ulrich Engel
University of Dortmund, Dortmund, Germany

1 Introduction **1**

2 Devices for the Generation of Microwave Plasmas **2**
2.1 Microwave Generators 3
2.2 Cavities and Resonant Structures 3
2.3 Plasma Torches 5

3 Diagnostics of Microwave Plasmas **7**
3.1 Low-power Microwave-induced Plasmas 9
3.2 Single-electrode Plasmas 9
3.3 Plasma Torch Discharges 10

4 Sample Introduction **11**
4.1 Techniques for Sample Introduction 11
4.2 Sample Introduction in Low-power Microwave Plasmas 11
4.3 Sample Introduction in Medium- and High-power Microwave Plasmas 14
4.4 Element-specific Detection in Chromatography 16
4.5 Glow Discharges 19

5 Optical Atomic Spectroscopy with Microwave Plasmas **20**
5.1 Atomic Emission Spectroscopy with Low-power Microwave-induced Plasmas 21
5.2 Single-electrode Microwave Plasmas 22
5.3 Plasma Torch Sources 22
5.4 Stabilized Capacitively Coupled Plasma 23
5.5 Microwave-induced Plasmas as Atom Reservoirs for Atomic Absorption Spectroscopy, Atomic Fluorescence Spectroscopy and Laser-enhanced Ionization 23

6 Mass Spectrometry with Microwave Plasmas **24**
6.1 Low-power Microwave Plasmas 24
6.2 High-power Microwave Plasmas 26

7 Applications of Microwave Plasma Atomic Spectroscopy **26**
7.1 Analysis of Biological Samples 26
7.2 Analysis of Environmental Samples 27
7.3 On-line Monitoring 28
7.4 Direct Solids Analysis 29

8 Comparison with Other Methods and Outlook **31**

Abbreviations and Acronyms **32**

Related Articles **33**

References **33**

The use of microwave plasmas as radiation sources for optical atomic emission (AES), absorption (AAS) and fluorescence (AFS) spectroscopy and for laser ionization spectroscopy is treated and reference is also made to the use of microwave-induced plasmas (MIPs) as ion sources for mass spectrometry (MS). Devices for producing both single-electrode and electrodeless microwave plasmas are treated, in addition to methods for their diagnostics, and results for the analytically relevant plasma parameters are presented. Methods for sample introduction are discussed. They include dry aerosol generation techniques (hydride generation (HG), electrothermal evaporation, spark and laser ablation (LA)) and possibilities for the uptake of wet aerosols. Special reference is made to coupling with gas chromatography (GC) and also to the potential for coupling with high-performance liquid chromatography (HPLC) Further, the use of microwave plasmas for cross-excitation in the case of glow discharges (GDs) is treated. The analytical figures of merit in the case of AES with low-power microwave plasmas, single-electrode microwave plasmas, plasma torch sources and stabilized capacitively coupled plasmas are given also in the case of atomic absorption, fluorescence and laser ionization with these sources. The developments in MS in the case of both low-power and high-power microwave plasmas and in the case of various types of sample introduction are discussed.

Applications of MIP atomic spectroscopy are in the fields of biological samples with special reference to microanalysis, and of environmental samples with special emphasis on metal speciation, on-line monitoring and direct solids analysis. A critical comparison of the methodology with other methods for the determination of the elements and their species is given.

1 INTRODUCTION

For elemental determinations at widely differing concentration levels, since the 1950s plasma sources have

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

played an important role as radiation sources for atomic spectrometry. The latter methodology, especially in the case of optical emission spectroscopy (OES) and MS, enables many elements to be determined simultaneously in very different types of sample materials. The use of plasma sources started with the classical arc and spark sources, as known from direct solids analysis. Experiments were particularly designed aiming at a constant supply of analyte in these sources, be it for the case of liquids, finely dispersed solids or gases. This led to procedures dealing with dc plasma jets, which are still in use. Plasma atomic spectrometry in this sense, however, was hampered by the need to use electrodes for energy supply, causing blanks as a result of electrode burn-off and poor long-term stability. Therefore, there was intensive research on the availability of electrodeless discharges. Work at atmospheric pressure was found to have an important advantage, as then a number of techniques such as solution nebulization could be used easily and successfully for sample introduction into the sources. This explains the considerable attention which single-electrode microwave plasmas attracted towards the end of the 1950s, despite the fact that they were not electrodeless, as shown by papers by Mavrodineanu and Hughes,[1] Jecht and Kessler[2] and Tappe and Van Calker.[3] Here plasmas were mostly produced in noble gases by the interaction of microwave energy at gigahertz frequencies with gas flows. The hot plasmas permitted efficient drying of wet aerosols, evaporation of solid particles, atomization, excitation and ionization. They were operated at a metal electrode (W, Au-coated Cu, brass) and they could be operated with Ar, He and even with air at an electrical power ranging from 200 to 600 W. The discharge has a bush form and the hottest zones are in the center of the plasma. In the case of high concentrations of alkali, however, the plasma geometry drastically changed, which led to large concomitant effects as compared with the inductively coupled plasma (ICP).[4] Accordingly, these sources could be used successfully as radiation sources in atomic emission and, provided that ways are found to enter the ions in a mass spectrometer, also as ion sources for MS.

The trends in the development of plasma spectroscopy in general were reflected in regular surveys in the literature. They started with the conference proceedings of the first Winter Conference on Plasma Spectrochemistry,[5] and reviews on commercial instrumentation (see e.g. Broekaert[6]) and on the trends in developments in plasma spectrometry (see e.g. Broekaert[7]) regularly occur. Also, the *Journal of Analytical Atomic Spectrometry* updates are very informative to the field (see e.g. Sharp et al.[8]).

The success of plasma atomic spectroscopy first started with the availability of the ICP, first described by Greenfield et al.[9] and by Wendt and Fassel.[10] Here a toroidal plasma with a diameter of up to 20 mm and a height of up to 50 mm could be realized. Further, both wet and dry aerosols containing even large particulates could be entered, guaranteeing residence times of several milliseconds at temperatures above 4000 K. These sources, however, also necessitate working with gas flows higher than 10 L min^{-1} and at an electrical power in the kilowatt range, both of which make the operating costs high. Apart from the single-electrode microwave plasmas, considerable effort was put into low-pressure microwave plasmas, as they became available towards the end of the 1950s also (see e.g. Uden[11]). Here excitation in sealed tubes could be used for drinking water analysis, but the analysis times were long.[12]

The breakthrough in microwave plasmas first took place when electrodeless plasma discharges, which can be operated at atmospheric pressure, became available. This was realized in 1976 by the work of Beenakker,[13] who described a cavity in which both He and Ar microwave discharges could be produced with a power of up to ca. 100 W with a gas consumption of below 1 L min^{-1}. These sources became very popular for the excitation of dry analyte vapors, as obtained in the electrothermal vaporization (ETV) of dry solution residues and also in GC. For the latter case, low-power electrodeless microwave discharges are now widely used.

To bridge the gap towards high-power plasmas with and without electrodes, in which wet aerosols can be taken up and where air operation is also possible, intensive research was carried out. This resulted in a wide variety of devices such as the surfatron,[14] which will be discussed further, and the recent microwave plasma torch (MPT) developed by Jin et al.,[15] and microwave plasma atomic spectroscopy became a powerful tool for elemental analysis.

The special features of the microwave plasma itself have been the subject of several review and tutorial papers by experts in the field. Jin et al.[16] reviewed work up to 1997 with 255 references. Winefordner et al.[17] surveyed microwave plasmas and GDs with respect to their typical lines of development. Reviews on microwave plasmas have been regularly published, e.g. for the work before 1980 by Zander and Hieftje,[18] in 1988 by Abdillahi,[19] in 1995 by Culp and Ng,[20] in 1984 by Matousek et al.[21] and in 1990 and 1996 by Broekaert.[22,23]

2 DEVICES FOR THE GENERATION OF MICROWAVE PLASMAS

Microwave generators operated at the allowed frequency of 2.45 GHz make use of standard technology in fields such as microwave heating and microwave sputtering

devices, whereas the development of microwave cavities and resonators and properly designed torches was and still is a field of continuous innovation, as it can be seen from reviews such as that by Goode and Baughman.[24]

2.1 Microwave Generators

For the production of microwave energy, use is made of clystrons and magnetrons, the latter being the more important when a power level above ca. 100 W is required. In a magnetron (Figure 1), a filament current is amplified and flows to an annular anode, which contains a certain number of cavity resonators.[26] Perpendicular to this plane, a very strong stationary magnetic field is applied. Through the action of the secondary cavities, an ultrahigh frequency (UHF) current with a characteristic frequency is produced, which can be coupled out with a loop and transported through a coaxial connector to a waveguide end or to a resonator. The power output of the magnetron can be regulated and must be stabilized and smoothed so as to allow, at any level, stable operation of the plasmas produced. Means for converting pulsed into continuous-wave operation in generators for microwave plasma production were described by Brandl et al.[27] Microwave generators operating at power levels from a few watts to 1 kW with sufficient stability (ripple and drift $<1\%$) are now commercially available. Miniaturization of microwave generators, especially with the aim of realizing small size and low power as required when generating smaller plasmas is possible through the use of transistor technology.

When transporting microwave energy into suitable structures, only part of the energy will be taken up in the device, part being reflected to the generator. The forward and reflected power can be measured with the aid of a bidirectional coupling loop. When transporting microwaves to a coaxial or cavity waveguide, the dimensions of the latter must be tuned to the wavelength of the microwave. Cavity conductors are rigid metal tubes with a rectangular or circular cross-section in which the microwaves propagate and their dimensions are standardized, e.g. for the case of 2.54 GHz, 91 mm internal width and 42 mm internal height. Individual waveguides must be tightly connected through suitable flanges to prevent microwave leakage. Each slit acts as a secondary source. Hence holes and slits for observation must be provided with a chimney so as to prevent the sorting of microwave radiation by damping. The resonance conditions required to optimize the ratio of forward to reflected power can be realized by fine tuning with stubs or displaceable walls and with screws perturbing the microwave field.

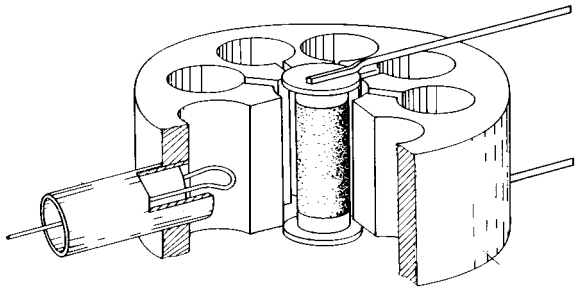

Figure 1 Schematic diagram of a magnetron.[25] (Reproduced by permission from O. Zinke, H.L. Hartnagel, *Lehrbuch der Hochfrequenztechnik*, Springer, Berlin, Band 2, 1987. Copyright Springer Verlag, 1987.)

2.2 Cavities and Resonant Structures

A plasma can be operated at the top of the internal conductor of a waveguide. To couple microwave energy inductively into a gas, a microwave cavity or resonator is used. The waveguide then should end at a length where the microwave amplitude maximum is located, whereas a cavity consists of a closed metal tube of rectangular or circular cross-section and its internal dimensions allow the formation of standing waves. Tapered rectangular cavities, Evenson quarter-wave cavities and Broida quarter- or three-quarter-wave cavities are commonly used. Anyhow, as the plasma formed disturbs the field, fine tuning with screws and stubs is required.

At reduced pressure many cavities could be used to obtain a microwave discharge, even in the case of gas mixtures (see e.g. McKenna et al.[28]). They include Evanson cavities and other devices. They played an important role in the excitation of analytes enclosed in quartz ampules, which was developed as a technique for dry solution residue analyses for volatile elements by OES[29–31] and especially also to element-specific detection in GC (see section 4.4).

Microwave plasmas operated at the tip of a metal electrode connected with the internal conductor were the oldest microwave structures used for plasma spectrochemical analysis. Here the analyte flows around the tip of the electrode and enters the plasma axially (Figure 2). Although this method is not optimal as the plasma has a bush form and maximum temperatures in the center lead to poorer signal-to-noise ratios (S/Ns), it found widespread use. Indeed, it is stable, can be operated with both noble and molecular gases and it even accepts wet aerosols. Attempts were made to improve the sampling capacity further by using a gas flow around the tip of the electrode to stabilize and to concentrate the analyte flow.[33] It could be further decreased in size and in this form it still attracts interest (for the analytical figures of merit, see e.g. Wünsch et al.[34]).

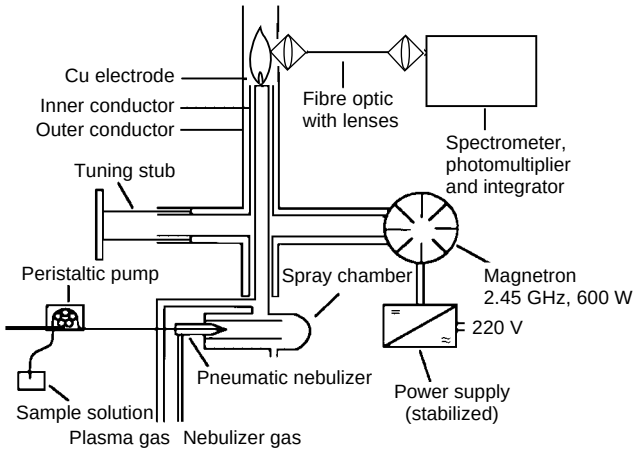

Figure 2 Schematic diagram of a set-up for capacitively coupled microwave plasma optical emission spectroscopy (CMPOES).[32]

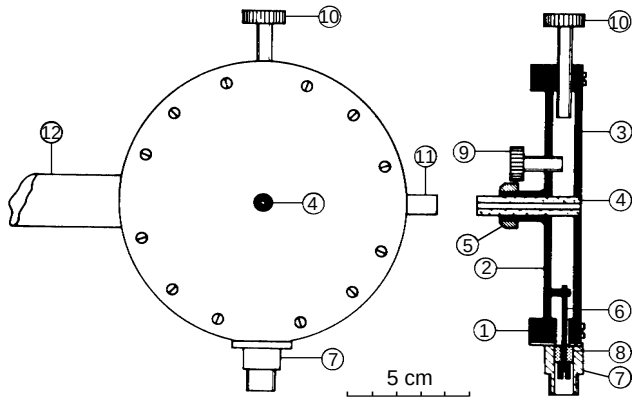

Figure 3 Schematic diagram of the Beenakker cavity.[38] (1) Cylindrical wall with (2) a fixed bottom and (3) a removable lid; (4) fused-silica discharge tube; (5) tube holder; (6) coupling loop; (7) connector; (8) sealing kit; (9,10) tuning screws; (11) air cooling; (12) cavity holder. (Reprinted from C.I.M. Beenakker, *Spectrochim. Acta, Part B*, **31**(7), 931–937 (1976), Copyright 1976, with permission from Elsevier Science.)

Microwave discharges first became popular when it became possible to produce them at atmospheric pressure and low power, as was first possible with the TM_{010} resonator described by Beenakker and was improved especially with respect to coupling.[35–37] As shown in Figure 3, in this device a standing wave is produced in a circular cavity into which the microwave energy is entered with the aid of a loop or with the aid of the end of the inner conductor acting as an antenna. Coarse tuning can be accomplished through the positioning of the antenna and for fine tuning a ceramic stub is provided, with which distortions of the field in the cavity can be produced. The latter was found to be no longer required in the case of delivering the microwave power with the aid of a loop.[39] In this cavity the power may vary from a few tens of watts to more than 300 W, depending on a cooling of the cavity, the burner and the respective gas flows. These parameters also determine the ways of sample introduction possible and sources operated at moderate power especially were designed with the aim of taking up wet aerosols. (see e.g. Haas et al.[40]).

At the end of the 1970s the surfatron was proposed as a source for analytical atomic spectroscopy. This microwave plasma, as shown in Figure 4, is produced in a microwave structure with varying depth and side-on coupling of the microwave power through the internal conductor.[41] Here, a microwave expands through the slit between the front plate and the plasma, through which a slightly hollow plasma with improved sampling capacity is realized. Selby and Hieftje[42] optimized the construction of the surfatron as a device for analytical work by making it of the same dimensions as the TM_{010} resonator according to Beenakker. The working power and pressure then become very similar to those of the resonator according to Beenakker. However, the plasma in a surfatron seemed to be more robust, e.g. when entering gases such as H_2 accompanying the volatile hydrides in HG work.[43]

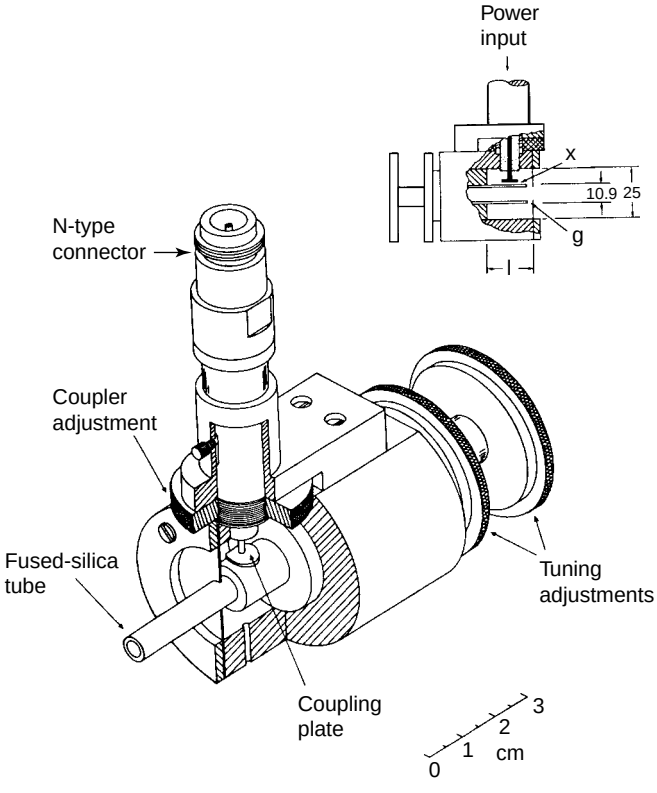

Figure 4 Schematic diagram of a surfatron set-up.[42] (Reprinted from M. Selby, G.M. Hieftje, *Spectrochim. Acta, Part B*, **42**(1–2), 285–298 (1987), Copyright 1987, with permission from Elsevier Science.)

Rectangular cavities, as described by Matusiewicz,[44] offer good possibilities for spectrochemical analysis. They can be operated from 100 W to several hundred watts. Even solvent-loaded aerosols can be entered in these sources after careful optimization. Further, the slab-line cavity described by Estes et al.[45] is also of use, but for the excitation of gaseous substances only. The different devices described in the literature were often compared when using one sample introduction technique (in the case of HG, see e.g. Mulligan et al.[46]) or with other plasmas[47] also for the case of solution analysis.

2.3 Plasma Torches

The analytical features of different types of microwave discharges mainly depend on the type of torch in which they are produced. The latter especially determines the admissible power and whether a filament-type, a delocalized or a toroidal plasma can be obtained.

In Ar or He microwave discharges, up to about 80 W can easily be produced in a quartz capillary with an internal diameter of up to 1 mm, as described in the original version according to Beenakker. However, they also can be operated with cooling by an outer gas flow. Through observation of the Si lines, it could be shown that the erosion of the quartz tube can be considerable and it could be substantially decreased by using a torch where the discharge capillary is liquid cooled.[48] This tube erosion could even be used as a technique of sample introduction, namely by supplying a liquid aliquot in the capillary of the MIP prior to initiating the discharge and ablating the dry solution residue with the aid of the discharge.[49]

Extensive efforts have been made to produce toroidal MIP plasma sources for OES. Axial observation then enables it to obtain high signal-to-background ratios and accordingly low detection limits. Success in this respect was achieved by Kollotzek et al.[50,51] by using wetted Ar and carefully centering a quartz capillary of 4 mm internal diameter in the resonator according to Beenakker. Later on it was found that such a plasma (Figure 5a) at a power of about 100 W could be obtained when the working gas only was led through a water-filled washing flask.[52] With dry He a diffuse delocalized discharge, which homogeneously fills a quartz tube of 4 mm internal diameter, could be obtained at 100–300 W (Figure 5b). However, here the cooling of the discharge tube is much more critical.

Another approach especially makes use of special torch constructions. Here Bollo-Kamara and Codding[53] were successful in using a dual-tube arrangement with a threaded inset, by means of which a tangential gas flow could be realized. More work on tangential flow torches included extensive diagnostic

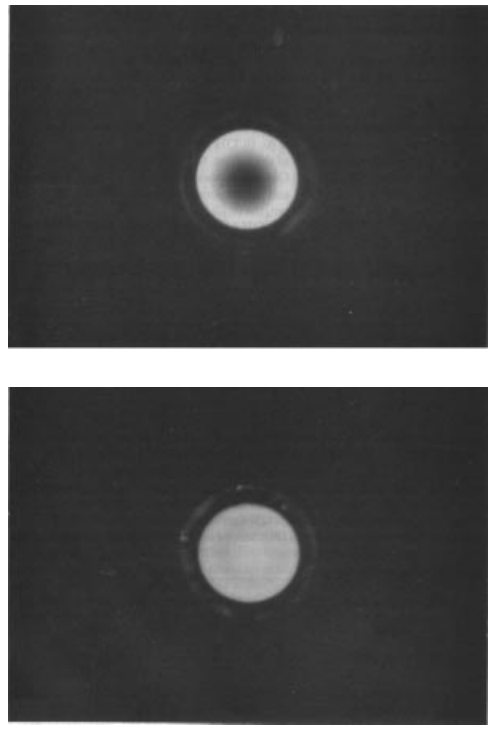

Figure 5 Pictures of (a) toroidal Ar MIP and (b) diffuse He MIP. Quartz torch tube: internal diameter 4 mm, external diameter 6 mm. (a) Pneumatic nebulization (PN), 1.2 L min^{-1} argon, −110 W; (b) without aerosol introduction, 0.2 L min^{-1} He, −130 W.[52]

studies of hydrodynamic flows and spatially resolved temperature measurements[54] and studies on self-centering plasmas[55] and their use in GC detection.[56] For the study of different arrangements, demountable torches in a polytetrafluoroethylene (PTFE) socket are very useful, as then the induividual tubes can easily be changed in position or exchanged after attack by the discharge.[57] A detailed study and an improvement of the torch proposed by Bollo-Kamara and Codding was realized with a such demountable torch; here a PTFE gas-swirl modifier was used and also a PTFE socket for the tubes, making glass-blowing superfluous.[58]

Further work was also done on laminar-flow plasmas with the aim of obtaining a plasma with a high discharge stability and robustness with respect to incoming analyte clouds. This is very important so as to keep baseline changes in GC detection to the minimum.[59,60] It was found that because of improved stability the detection limits obtainable with laminar flow torches were considerably lower than those with tangential flow torches. It was reported that stabilities even better than 2% over 2 days could be realized.[61]

A significant innovation in MIP work with respect to burner design was made by Jin et al.[15] They used a concentric arrangement of two metal tubes mounted in

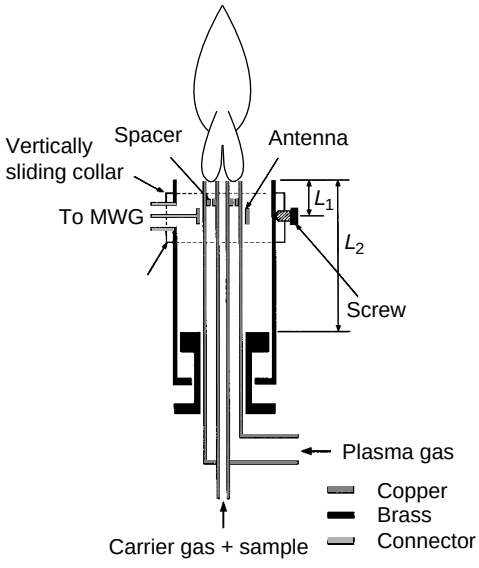

Figure 6 Schematic diagram of the MPT.[15] (Reprinted from Q. Jin, C. Zhu, W. Borer, G.M. Hieftje, *Spectrochim. Acta, Part B*, **46**(3), 417–430 (1991), Copyright 1991, with permission from Elsevier Science.)

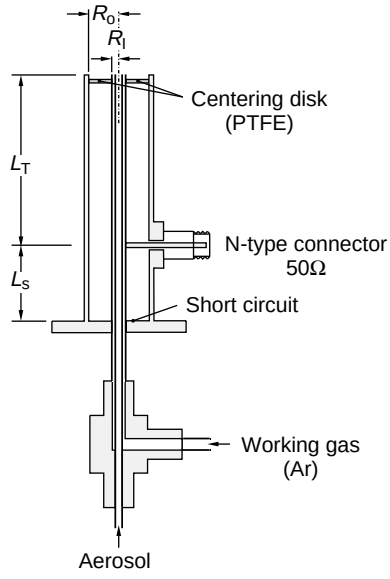

L_T	L_s	R_o	R_i	d_i	d_{ao}	d_{ai}
67	34	11	3	4	3	2

Figure 7 Modified MPT set-up for atomic spectroscopy.[63] Dimensions of the MPT are in millimeters; inner diameter of center conductor d_i and outer diameter d_{ao} together with inner diameter of analyte introduction tube.

an enclosure and the power was coupled in by means of the intermediate tube (Figure 6). Power coupling initially was realized with the aid of a gliding ring moving over the intermediate tube. Effective power coupling could be achieved by varying the depth of the outer enclosure and by changing the coupler position. It was found that with both Ar and He stable discharges could be realized at a gas consumption below 1 L min^{-1} and with a power of up to 200 W depending on the efficiency of the cooling. As a wide variety of sample introduction techniques could be used, this source is now one of the prominent microwave structures.

The design of the MPT was considerably improved by Pack and Hieftje.[62] First, it was possible to close the adjustment slit completely, so that air no longer enters between the intermediate and the outer tubes and microwave leakages also are minimized. Further, it was possible to use a quartz instead of a metal tube as the internal tube. This considerably lowers the risks for contamination when analyzing real samples. It was also found that sheathing of the plasma with N_2 greatly improved its stability. Further improvements of the torch design resulted from modeling of the electromagnetic fields without and with plasma. As a result, it was possible to design a torch which no longer requires tuning and with which ignition and stable operation of the MPT was also possible (Figure 7).[63] This significantly improves the reproducibility of the device with respect to ignition and operation and makes the MPT a robust spectrochemical source. Here the internal conductor of the coaxial cable is directly connected to the coupler, which is rigidly fixed to the intermediate tube. Further, it was no longer required to use a sheating gas in the metal part of the torch. A quartz enclosure around the plasma itself, however, was found to stabilize it considerably against air turbulence in the laboratory. Further developments include pulsed operation of a microwave discharge at higher power (500 W) and its coupling with spark ablation (SA) sampling devices.[64]

High-power inductively coupled microwave plasmas could be operated in various MIP resonant structures. This first is possible in the TM_{010} resonator according to Beenakker where, at powers of up to 500 W, Ar, He and air plasmas can be obtained, again by using modular

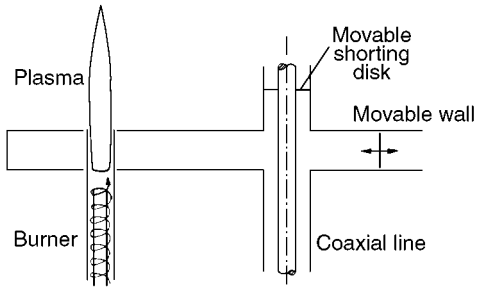

Figure 8 Resonant rectangular TE_{102} resonator and plasma torch for the 800 W MIP.[67]

cooled torches.[65] Such devices were often compared in the case of well-defined sampling techniques such as GC detection.[66]

Another possibility is the use of ICP-like torches in a rectangular cavity (Figure 8). Here a power up to the kilowatt range can be used and at several liters per minute gas consumption a very stable plasma can be obtained with Ar and also with air. This MIP is a powerful source for AES and for MS.[68]

3 DIAGNOSTICS OF MICROWAVE PLASMAS

The diagnostics of microwave plasmas deal with both the measurements of the plasma physical properties, including the flow dynamics, temperatures and particle number densities, and their spectroscopic properties with respect to optical emission, absorption and fluorescence, their properties as atom reservoirs and their properties as ion sources. The results obtained here differ for the different types of microwave discharges used for spectrochemical purposes; however, some trends are common.

As long as MIPs at atmospheric pressure are considered, the plasma is near to thermal equilibrium (TE); however, there might already be much greater differences between the different temperatures than in the case of the ICP. The main processes taking place in the plasma are as follows:

- electron impact for the excitation of working gas species (Equation 1):

$$A + e \longrightarrow A^* \tag{1}$$

- electron impact for the ionization of working gas species (Equation 2):

$$A + e \longrightarrow A^+ \tag{2}$$

- electron impact for the excitation of analyte species (Equation 3):

$$S + e \longrightarrow S^* \tag{3}$$

- electron impact for the direct ionization of the analyte species (Equation 4):

$$S + e \longrightarrow S^+ \tag{4}$$

- radiative recombination (Equation 5):

$$A^* \text{ or } S^* + e \longrightarrow A \text{ or } S + h\nu \tag{5}$$

- excitation of analyte through collisions with excited working gas atoms (Equation 6):

$$S + A^* \longrightarrow S^* + A \tag{6}$$

- ionization of analyte through collisions with excited working gas atoms (Equation 7):

$$S + A^* \longrightarrow S^+ + A \tag{7}$$

A special case is excitation through collisions with metastable noble gas species (Ar^m). These are species for which a decay through the emission of radiation is forbidden as the multiplicity of the lower states differs from that of the excited states. As the energy of these species is high (e.g. Ar^m at 11.7 eV, He_2^m at 14.6–17.4 eV and $He_2{}^{+m}$ at 18.8–21.6 eV[69]), this process may be relevant for the one-step excitation of ion lines and it is highly selective.

- Charge-transfer processes (Equation 8):

$$A^{+(*)} + S \longrightarrow A + S^{+(*)} \tag{8}$$

- radiation trapping (Equation 9):

$$A + h\nu \longrightarrow A^* \tag{9}$$

may be important for energy exchange and excitation as well.

With respect to diagnostics, many papers have described the determination of excitation temperatures from the relative intensities of atomic emission lines originating from the same stage of ionization. The intensity of an atomic emission line I_{qp}, corresponding with a transition from the higher level q to the lower level p, is proportional to the population of the atoms a in level q (n_{aq}) and is given by Equation (10):

$$I_{qp} = A_{qp} n_{aq} h\nu_{qp} \tag{10}$$

After substitution of the population of the excited level according to Boltzmann's equation (Equation 11):

$$\frac{n_{aq}}{n_a} = \left(\frac{g_q}{Z_a} \quad \exp\left(\frac{-E_q}{kT} \right. \right. \tag{11}$$

where n_a is the total density of atoms a over all levels, g_q the statistical weight of the excited level q, Z_a the partition function for the atoms (being a function of the temperature), E_q the excitation energy for the excited level q and T the excitation temperature (often denoted T_{ex}), one obtains Equation (12):

$$I_{qp} = A_{qp} h\nu_{qp} n_a \left(\frac{g_q}{Z_a} \quad \exp\left(\frac{-E_q}{kT} \right. \right. \tag{12}$$

T then can be determined from the intensity ratio for two lines (a and b) with wavelength λ_a and λ_b of the same ionization state of an element according to Equation (13):

$$T = \frac{5040(V_a - V_b)}{\log[(gA)_a/(gA)_b] - \log(\lambda_a/\lambda_b) - \log(I_a/I_b)} \quad (13)$$

Often the line pair Zn 307.206/Zn 307.59 nm is used. This line pair is very suitable because ionization of zinc is low as a result of its relatively high ionization energy. Also, the wavelengths are close to each other, which minimizes errors introduced by changes in the spectral response of the detector, and the ratio of the gA values is well known.

The excitation temperatures can also be determined from the slope of the plot of $\ln[I_{qp}/(g_q A_{qp} \nu_{qp})]$ or $\ln(I_{qp}\lambda/gA_{qp})$ versus E_q, which is $-1/kT$. The λ/gA values for a large number of elements and lines are available. Spectroscopic measurements of temperatures from line intensity ratios may be hindered by deviations from thermodynamic behavior in real radiation sources and by inaccuracies in the transition probabilities. The determination of excitation temperatures in spatially inhomogeneous plasmas, which have a cylindrical symmetry, can be performed with the aid of Abel inversion, as treated extensively in classical textbooks on atomic spectroscopy (see e.g. the book by Boumans[70]).

For the measurement of the gas kinetic temperatures, both measurements of the rotational temperatures from the band spectra of species such as the OH radical or the N_2^+ ion and measurements of Doppler widths have been proposed. In the first case one makes use of the rotational fine structure of the vibration bands.

Molecules or radicals have various electronic energy levels ($^1\Sigma$, $^2\Sigma$, $^2\Pi$, etc.) which have a vibrational fine structure ($v = 0, 1, 2, 3, \ldots$), and these levels, in turn, have a rotational hyperfine structure ($J = 0, 1, 2, 3, \ldots$). The total energy of a state may be written as Equation (14):

$$E_i = E_{el} + E_{vib} + E_{rot} \quad (14)$$

E_{el} is of the order of 1–10 eV, the energy difference between two vibrational levels of the same electronic state is of the order of 0.25 eV and the separation of rotational levels is of the order of 0.005 eV. When the rotational levels considered belong to the same electronic level, the emitted radiation is in the infrared (IR) region. When they belong to different electronic levels, they occur in the ultraviolet (UV) or visible (VIS) region. Transitions are characterized by the three quantum numbers of the states involved: n', v', j' and n'', v'', j''. All lines which originate from transitions between rotational levels belonging to different vibrational levels of two electronic states form the band: $n', v' \rightarrow n'', v''$. For these band spectra, the selection rule is $\Delta j = j' - j'' = \pm 1, 0$. Transitions for which $J'' = j' + 1$ give rise to the P-branch, $j'' = j' - 1$ to the R-branch and $j' = j''$ to the Q-branch of the band. The line corresponding to $j' = j'' = 0$ is the zero line of the band. When $v' = v'' = 0$ it is also the zero line of the system. The difference between the wavenumber of a rotational line and the wavenumber of the zero line in the case of the P- and the R-branch is a function of the rotational quantum number j and the rotation constant B_v, for which (Equation 15):

$$\frac{E_j}{hc} = B_v j(j+1) \quad (15)$$

The functional relation is quadratic and known as the Fortrat parabola.

As in the case of atomic spectral lines, the intensity of a rotational line can be written as Equation (16):

$$I_{nm} = \frac{N_m A_{nm} h\nu_{nm}}{2\pi} \quad (16)$$

where N_m is the population of the excited level and ν_{nm} the frequency of the emitted radiation. The transition probability for dipole radiation is given by Equation (17):

$$A_{nm} = \frac{64\pi^4 \nu_{nm}^3}{3k} \frac{1}{g_m} \sum |Rn_i m_k|^2 \quad (17)$$

where i and k are the degenerancies of the upper (m) and the lower state (n). $Rn_i m_k$ is a matrix element of the electrical dipole moment and g_m is the statistical weight of the upper state; N_m is given by the Boltzmann equation (see Equation 11) where the energy to be included is E_j (see Equation 15) with $j = J'$, being the rotational term of the upper level. For a $^2\Sigma_g - ^2\Sigma_u$ transition, between a so-called "gerade" (g) and "ungerade" (u) level, the term $\Sigma|Rn_i m_k|^2 = J' + J'' + 1$, where J' and J'' are the rotational quantum numbers of the upper and lower state. Accordingly (Equation 18):

$$I_{nm} = \frac{16\pi^3 cN\nu_{nm}^4}{3Z(T)} (J' + J'' + 1) \times \exp \frac{-hcB_v J'(J'+1)}{kT} \quad (18)$$

or

$$\ln \frac{I_{nm}}{J' + J'' + 1} = \ln \frac{16\pi^3 cN\nu_{nm}^4}{3Z(T)} - \frac{hcB_v J'(J'+1)}{kT} \quad (19)$$

By plotting $\ln[I_{nm}/(J' + J'' + 1)]$ versus $J'(J' + 1)$ for a series of rotational lines, a so-called rotational temperature can be determined from the slope. It reflects the kinetic energy of neutrals and ions in the plasma. For the determination of the gas kinetic temperatures, both measurements of the rotational temperatures from the band spectra of species such as the OH radical or the N_2^+

ion as well as measurements of Doppler widths have been used. For the measurements of electron temperatures Thomson scattering has also been used.[71] The basics of these methods are described in classical textbooks[70,72] and papers.[73] Also, efforts have been made to describe the flow dynamics of microwave discharges, which especially were found to be very useful for improving the design of microwave structures used for spectrochemical analysis.[63]

3.1 Low-power Microwave-induced Plasmas

In early MIP work, Beenakker MIP[69] mentioned that the electron temperatures in an MIP operated in a TM_{010} cavity should be high whereas rotational temperatures, being a good approximation to the gas kinetic temperatures, as described by Heltai et al.,[52] are only of the order of 1500 –2400 K (Figure 9a and b) both for He and Ar MIPs. Similar measurements were made in the surfatron plasma,[74] in the modulated surfatron[75] and also in MIPs operated in CO_2 and in He.[76] For the case of a surfatron the electromagnetic properties were studied[77] in addition and also their fundamental parameters.[78] In the case of the MIP the importance of charge transfer was shown by Brandl and Carnahan.[79] The diagnostic data are important for the dissociation of molecular species, being especially studied with respect to combinations of chromatography and MIP excitation, as described for hydrocarbons.[80]

Diagnostics have been performed not only for some widely used types of MIPs but also for a low-power Ar MIP used for the analysis of solutions[81] and for an MIP operated in a liquid-cooled discharge tube.[82] Studies on the energetic balance in low-pressure MIPs have also been made.[83] With special plasmas, such as a low-density He electron cyclotron resonance microwave plasma, techniques for temperature measurements, e.g. with line broadening, have been developed.[84]

In addition to plasma parameter measurements, noise analysis of the reflected power of MIPs,[85] near-infrared (NIR) characteristics of MIPs with respect to the determination of nonmetals[86] and studies on MS of MIPs[87] were also performed.

3.2 Single-electrode Plasmas

For single-electrode plasmas, electron number densities and excitation temperatures were determined by Kirsch et al.,[88] who had already extensively used this plasma for analytical purposes (see e.g. Hanamura et al.[89]), as did Boumans et al. (see e.g. Boumans et al.[4]). Here excitation temperatures of 5000–7000 K were reported for a 500 W capacitively coupled microwave plasma (CMP) operated in Ar or He with N_2 as wall-stabilizing

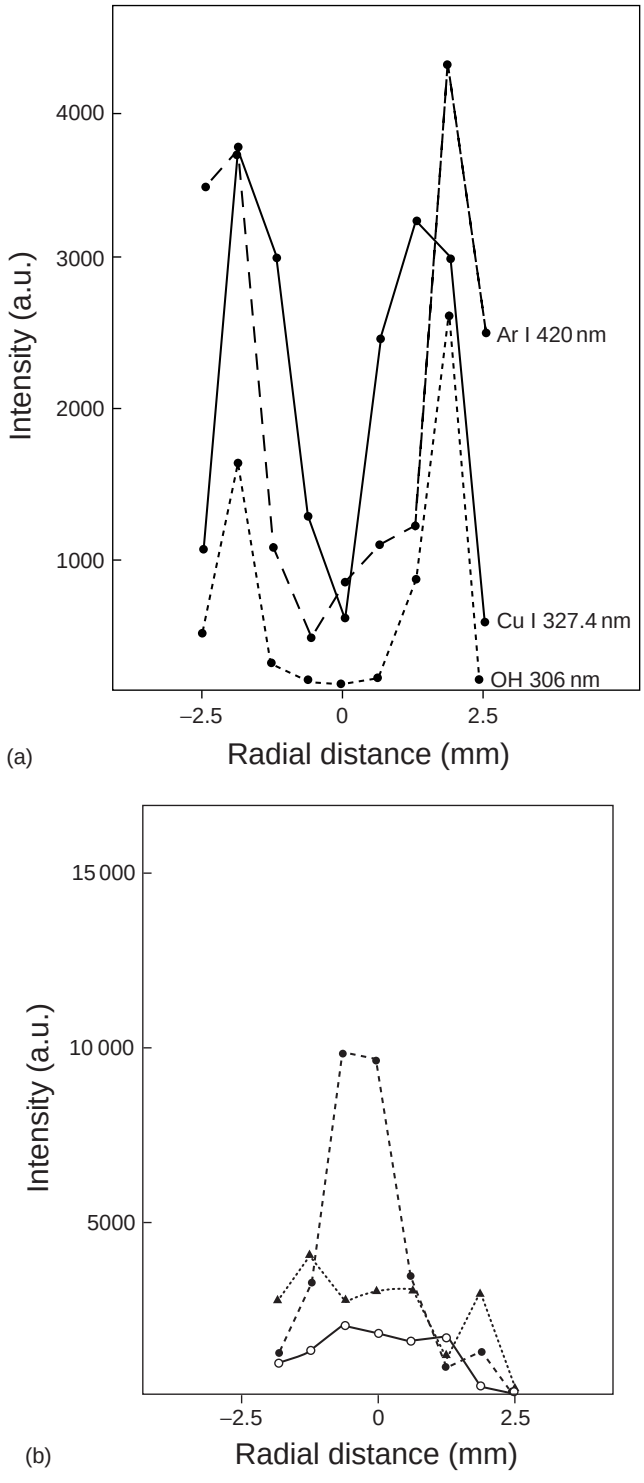

Figure 9 (a) Radial intensity profiles of a toroidal Ar MIP. Discharge tube: external diameter 7 mm, internal diameter 5 mm; PN of H_2O or of a 100 µg mL^{-1} solution of copper, −100 W. (b) Radial intensity profiles measured for a cylindrical He MIP. Discharge tube: external diameter 7 mm, internal diameter 5 mm; 0.2 L min^{-1} helium; without aerosol introduction, −150 W. (○) He 388.7 nm; (●) (0, 0) N_2^+ band head at 390 nm; ( ) (0, 0) OH band head at 306 nm.[52]

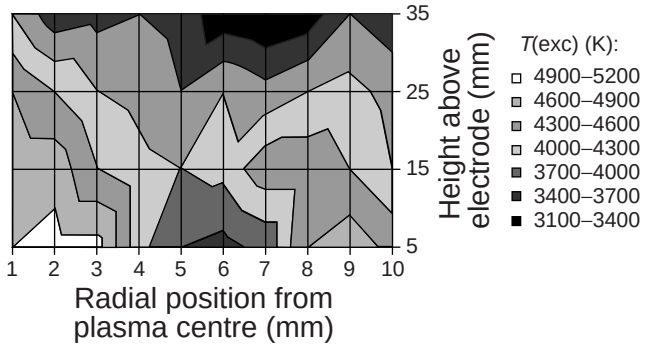

Figure 10 Mapping for excitation temperature (T_{exc}) in an air CMP.[32]

Table 1 Ranges of T_{exc}, T_{rot} and n_e obtained for 600 W N_2, air and Ar CMPs[32]

Working gas	T_{exc} (K)	T_{rot} (K)	n_e (cm^{-3})
N_2	5500–3400	4300–2500	10^{11}–10^{14}
Air	4600–3400	3400–2800	10^{10}–10^{14}
Ar	4900–3100	4300–2800	10^{9}–10^{14}
Standard deviation	300	300	10^{7}–10^{12}

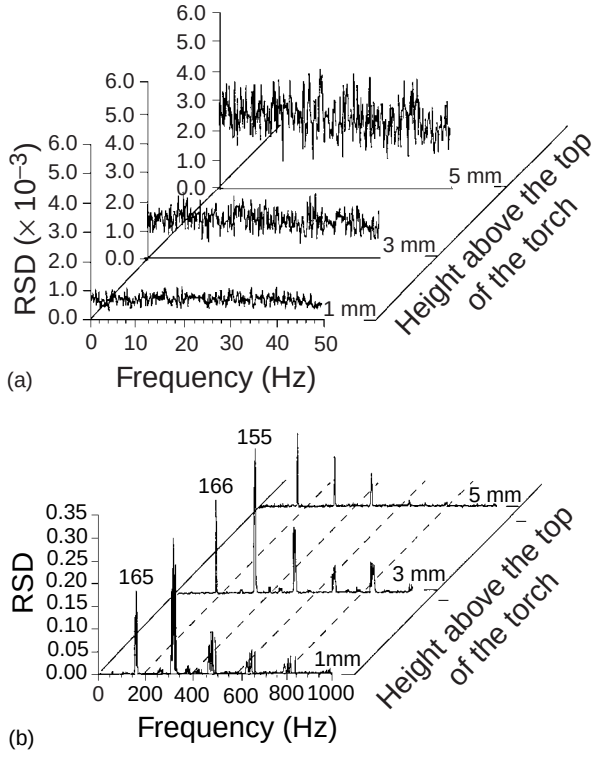

Figure 11 Effect of H_2 addition (13.5 mL min^{-1}) on the noise spectra of the He I 501.56 nm line at different heights above the top of the torch.[92]

gas. The flow pattern changes as the result of the addition of the alkali metals, as reflected in the tomographic studies reported by Bings and Broekaert.[90] Here it could be shown that the plasma blows up as a result of ambipolar diffusion and becomes broader when Cs or Na is added (Figure 10). Ethanol addition could also be shown to thermalize the plasma, as excitation and rotational temperatures measured then came nearer to each other. The plasma temperatures for CMPs operated with air only could be measured and tomographic studies performed. As shown in Table 1, these temperatures did not differ much from these obtained by Kirsch et al.[88] with noble gases and neither did the electron number densities or the rotational temperatures. Anyhow, it could be shown that CMPs, despite the fact that they are often operated at a power of 500 W which is near to the power of low-power analytical ICPs, have a larger deviation from local thermal equilibrium (LTE) than the ICP. This can be understood from the higher operational frequencies of microwave discharges, which even more hamper energy exchange as compared with the ICP.

3.3 Plasma Torch Discharges

Since the first publications by Jin et al.[15,91] on the MPT, it became clear that a microwave plasma with a toroidal structure can be operated in an ICP-like torch and that the moderate power and the possibility of using both He and Ar as working gases made it a very powerful atomic spectrometric source. This MIP could be well operated in combination with flow-cell HG. There was no need to remove the excess of H_2 produced before leading the reaction gases, including H_2 and the hydrides, into the plasma, which was found to be impossible with an MIP according to Beenakker. From noise power spectra obtained with the MPT in the case of HG, as shown in Figure 11(a) and (b), it was found that even with an excess of H_2 the frequency component stemming from the plasma swirl still is the important one, as it is in the case of the ICP. Further, however, flicker noise was found to occur at frequencies of 0.2–0.5 Hz and white noise was found to dominate below 100 Hz.[93] Temperatures were also measured in the microwave plasma operated in the MPT. For electron and gas kinetic temperatures, values at different heights in the plasma were obtained by Thomson scattering.[94] A complete temperature and electron number density mapping in the case of a modified MPT was performed recently.[95] It showed that in the case of wet plasmas the electron temperature increased similarly as in the case of the ICP. Further, mappings of the rotational temperatures, as possible with a charge-coupled device (CCD) spectrometer, were performed by Engel et al.[96] They showed that the rotational temperature in the

MPT increases with the water uptake, which might even improve the robustness of the source. This may explain why it easily seemed to be possible to take up water- and even acetonitrile-loaded aerosols into the MPT, as described by Prokisch and Broekaert.[97]

Also for other types of MIPs, spatial mappings of analyte intensities and plasma parameters were made as an aid in optimization studies.[98] Further, studies were not confined to noble gas plasmas, but also plasma torches obtained in the case of N_2 at a higher power were operated in a TM_{010} resonator.[99] These devices turned out to be of use for solution analysis and could use an ICP-like torch. This discharge, as well as that referred to by Ohata and Furuta[100] and later proposed as a source for MIP/MS, is similar to that described by Leis and Broekaert.[67] It is operated in a rectangular cavity and is very stable; however, it is much more prone to easily ionized element interferences than the ICP.

Thorough diagnostic studies also have been performed with the so-called ''torche à injection axiale'' (TIA) by the group of Van der Mullen as described by Jonkers et al.[71,101] recently. This plasma has a toroidal shape, it can take up water-loaded aerosols and it can be operated with both molecular and noble gases.

For this plasma it has been reported that the Saha equation applies. This equation describes the ionization in a thermal plasma as (Equation 20):

$$S_{pj}(T) = \frac{p_{ij}p_e}{p_{aj}} = \frac{(2\pi m)^{\frac{3}{2}}(kT)^{\frac{5}{2}}}{h^3(2Z_{ij}/Z_{aj})[\exp(-E_{ij}/kT)]} \qquad (20)$$

where S_{pj} is the Saha function expressed in terms of the partial pressures for the component j, p_i, p_e and p_a are the partial pressures for ions, electrons and atoms, respectively, the factor 2 is the statistical weight of the free electron (two spin orientations), m is the mass of the electron (9.11×10^{-28} g), k is the Boltzmann constant ($k = 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$), T is the absolute temperature (K), h is Planck's constant ($h = 6.62 \times 10^{-34}$ J s), Z_{ij} and Z_{aj} are the partition functions of the ions and the atoms of the element j, respectively, and E_{ij} is the ionization energy of the element j. As $1\,\mathrm{eV} = 1.6 \times 10^{-19}$ J, this leads to Equation (21):

$$\log S_{pj} = \frac{5}{2}\log\ \frac{T-5040}{T}V_{ij}\ + \log\left(\frac{Z_{ij}}{Z_{aj}}\right. - 6.18 \qquad (21)$$

where V_{ij} is the ionization energy (eV). It should be mentioned, however, that the Saha equation is only valid for plasmas in LTE, which strictly does not apply in the case of many MIPs. The latter is understandable from the fact that the light electrons can easily take up energy at microwave frequencies, but heavy ions much less so. This results in different kinetic energies and different temperatures for the light electrons and the heavy ions and, accordingly, in departures from LTE. By Thomson scattering, the electron temperature can be determined irrespective of whether the plasma is in LTE or not and for the TIA, for instance, electron temperatures in the case of Ar were reported to be of the order of >8700 K.[71] This was the case even for a He plasma operated in air.[102] Atom state densities for this plasma have also been calculated.[101]

4 SAMPLE INTRODUCTION

As in any spectrochemical source, the analyte has to be brought as an atomic cloud in the zones where excitation and/or ionization takes place. Because of the relatively low gas kinetic temperatures in the microwave discharges, their atomization capacity is not as high as in the ICP, which is generally operated at higher power than microwave discharges. Accordingly, the analyte must be atomized more efficiently or at least it must be present in a very fine form before it enters the microwave plasma.

4.1 Techniques for Sample Introduction

Sample introduction is the ''Achilles' heel'' of plasma spectrochemical analysis[103,104] and many techniques which can be used in the case of microwave plasmas have been described in the plasma spectrometry literature, as shown in a monograph by Sneddon,[105] a special issue of *Spectrochimica Acta, Part B,*[106] and various reviews treating aspects such as flow injection,[107] coupling techniques,[108] on-line separation[109] and sample introduction in atmospheric pressure MIPs[110] or MIPs in general.[111,112] These techniques include PN, ultrasonic nebulization (USN), ETV of microsamples, HG for the elements having volatile hydrides and direct solids sampling methods such as slurry nebulization, SA and LA and direct sample insertion, as are well known from the ICP literature (Figure 12). In the case of microwave plasmas at low power, especially techniques where a dry atom cloud is formed will be favorable. This explains, why tandem systems using an MIP for the excitation of the exhaust aerosols of an ICP were of use,[114] in which the MIP even could be operated at reduced pressure.[115]

4.2 Sample Introduction in Low-power Microwave Plasmas

In the case of very weak discharges such as the MIP according to Beenakker, especially ETV and HG[116] have been used for sample introduction. In the case of ETV, as reviewed by Matusiewicz,[117,118] different types of electrothermal devices can be used. They include

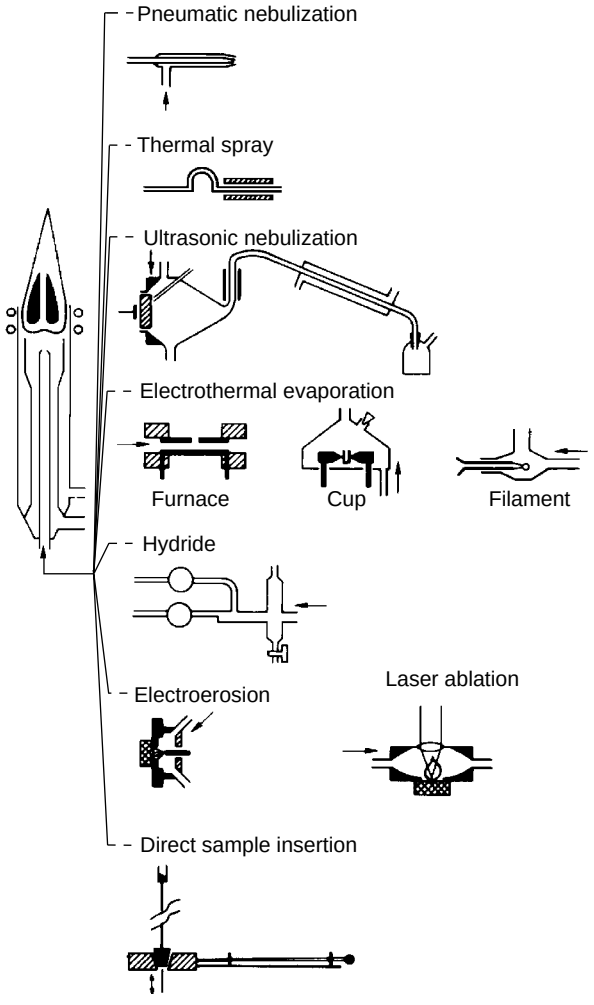

Figure 12 Different techniques for sample introduction on inductively coupled plasma optical emission spectroscopy (ICPOES).[113]

a commercial graphite furnace (GF) as known from AAS[119–121] as well as a graphite rod atomizer,[122] even for the case of a low-pressure MIP,[123] wire-loop devices entered into a microtube[124,125] and Ta boat[126] and Ta strip devices.[127] In the case of the resonator according to Beenakker, it could be shown that the detection limits with GF sample evaporation can be considerably improved by using a toroidal instead of a filament discharge.[120] Special attention was also given to the acquisition of the transient signals, for which a rapid scanning monochromator can be very helpful.[128] The technique could be used for the determination of I in HCl,[129] for the determination of nonmetals in the case of an He discharge,[130,131] for the determination of S,[132,133] of halogenated compounds[134] and of Zn[135] and for the determination of Ni and Pb in bones.[136] A useful way to increase further the detection power of ETV coupled to OES with the MIP consists of preconcentration of the trace elements to be determined by electrolysis on a graphite tube filled with vitreous C, as shown for the case of Pb.[137] Chromatography on a vitrous C-filled column can also be used for this purpose.[138] Many phenomena, however, still have to be studied, such as the aerosol transport mechanisms[139] and also vaporization at low pressure. The latter, for example, has been studied for the case of chlorides, sulfates and nitrates.[140] Special ETV techniques may also make use of moving bands.[141]

The generation of hydrides and of other volatile compounds has found widespread use in MIP work. In a number of cases, however, the vapors generated have to be freed from an excess of molecular gases such as H_2 or even water vapor, as these species may change the plasma impedance so much that the circuit leaves the resonance frequency and the plasma is extinguished. In the case of a surfatron MIP, the excess of H_2 envolved in continuous HG was found not to hamper the discharge stability,[142] whereas in the case of the MIP in a Beenakker resonator the excess of H_2 is mostly removed by fixing the hydrides in a cold trap and releasing them at once into the MIP.[143] In this way, detection limits at the sub nanograms per milliliter level can be obtained (Tables 2 and 3), and also by hot-trapping in a preheated GF.[143,152,153] Also a Nafion® membrane can be used for separating moisture or H_2.[144] Further, one can work with microtechniques, where microaliquots are deposited on an $NaBH_4$ pellet.[154,155] At well selected operating conditions, continuous HG could even be directly coupled with microwave plasmas, also for the case of real samples, such as for the determination of Se in soils.[156] With low-pressure microwave plasmas, the introduction of an excess of H_2 was found to hamper the discharge stability less.[157] Also, the determination of As and Se subsequent to microwave sample dissolution and hot hydride trapping has been described.[158] Further, miniature hydride systems for the determination of As, Sb, Pb and Sn[155] and the determination of Pb by HG[159] have been described. For their optimization, spatially

Table 2 Comparison of the concentration detection limits in microwave-induced plasma optical emission spectroscopy (MIPOES) for As, Sb and Se obtained with different sample introduction techniques[a] [143]

Procedure[b]	As (ng mL^{-1})	Sb (ng mL^{-1})	Se (ng mL^{-1})
PN/T/MIP	500 (4.8)	250 (4.0)	540 (4.2)
GF/3F/MIP	50 (1.8)	20 (2.0)	46 (1.5)
HG/CT/3F/MIP	0.8 (6.7)	0.4 (6.8)	0.5 (7.2)
HG/GFT/3F/MIP	0.4 (4.5)	0.35 (4.8)	0.25 (4.6)

[a] At 10 × detection limit; data in parentheses are RSD, relative standard deviation (%).
[b] T/MIP, toroidal microwave-induced plasma; 3F/MIP, three-filament/microwave-induced plasma; GFT, graphite furnace trapping. CT, cold trapping.

Table 3 Comparison of literature values for the detection limits (ng mL^{-1}) for As, Sb and Se in AAS, AES and MS[143]

Technique[a]	Reference	As	Sb	Se
PN/ICP	145	50	32	75
PN/MIP	51	300	–	–
HG/ICP	146	1.0	2.4	1.3
HG/ICP	147	1.0	–	–
HG/ICP	148	0.06	0.18	–
HG/MIP	144	0.32	6.1	–
HG/GF/MIP	149	0.12	–	–
PN/ICP/MS	150	0.14	0.019	1.5
HG/ICP/MS	150	0.017	0.031	0.17
HG/AAS	151	0.16	0.08	0.18

[a] PN/ICP, pneumatic nebulization/inductively coupled plasma; PN/MIP, pneumatic nebulization/microwave-induced plasma; HG/ICP, hydride generation/inductively coupled plasma; HG/MIP, hydride generation/microwave-induced plasma; HG/GF/MIP, hydride generation/graphite furnace/microwave-induced plasma; PN/ICPMS, pneumatic nebulization/inductively coupled plasma mass spectrometry; HG/ICPMS, hydride generation/inductively coupled plasma mass spectrometry; HG/AAS, hydride generation atomic absorption spectroscopy.

resolved studies of Ga, In, Se, Te, As and Sb lines[160] were very useful. Not only the hydrides but also further volatile compound-forming elements can be easily determined with the aid of an MIP at a very high power of detection. This first applies to the determination of Hg with the aid of the cold vapor (CV) technique, known from AAS work. This can be done at low pressure[161] or at atmospheric pressure[162] and also down to ultratrace levels, as already shown in early work.[163] By this method I also can be determined, simply by a ''reverse titration'',[164] as already shown for the case of seawater.[165] Much work has also been done on the determination of the halogens after volatilization into an He MIP.[166]

Cl could be determined as HCl,[167,168] all halogens could be determined with the aid of detected ion with a surfatron MIP[169] and also Br[170–173] and I[174,175] could be determined using an MIP. In the case of I, methods for differentiation between iodate and iodide in brines have been described.[176] The influences of N_2 impurities on the F and Cl signals were also studied.[177] Cl and Br could be volatilized chemically through the use of strong oxidants such as a solution of $K_2Cr_2O_7$ in highly concentrated sulfuric acid.[178] Also further nonmetals can be determined by volatilization methods,[179] including S after H_2S/SO_2 generation.[180] Generation of volatile species can be applied for the determination of Ni by using the formation of volatile carbonyls.[181,182] The introduction of solid halocarbons,[183] the determination of C, H, N and O for elemental analysis purposes with the aid of a diode-array spectrometer and atomic emission lines in the IR region[184] and organic microanalysis have further been reported.[185] Special attention was also given to the determination of the N species in the MIP[186] and to the emission of H for radiation standardization purposes.[187]

Wet aerosols are mostly generated by PN, USN or by newer techniques such as hydraulic high-pressure nebulization (HHPN). In PN with cross-flow, concentric, grid or Babington nebulizers (Figure 13a–d), the aerosol is generated by splitting off small droplets from a liquid surface by viscosity drag forces and a good description

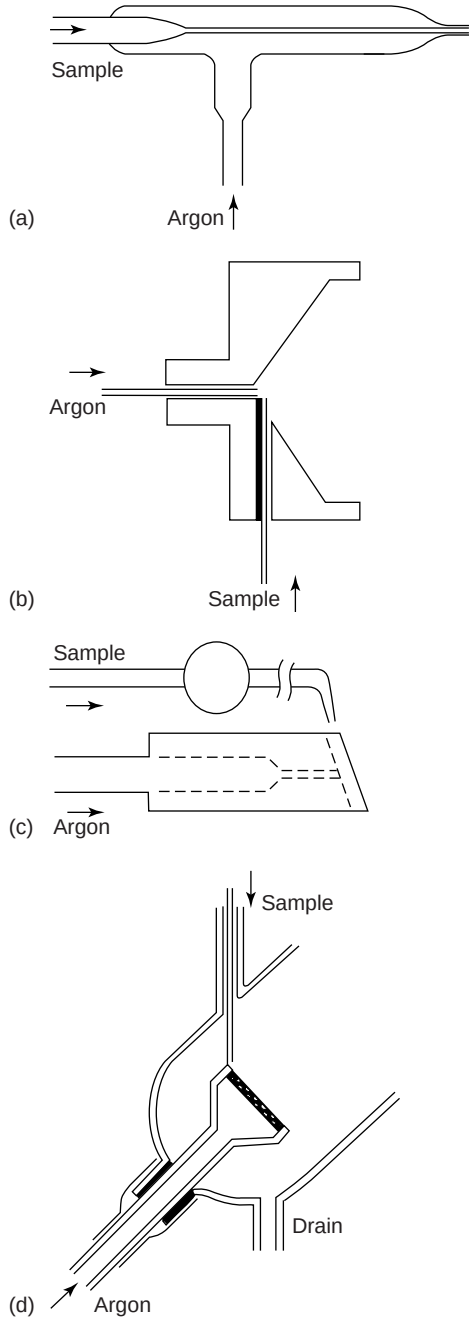

Figure 13 Pneumatic nebulizers for plasma atomic spectrometry: (a) concentric glass nebulizer; (b) cross-flow nebulizer; (c) Babington-type nebulizer; (d) fritted-disc nebulizer.[188]

of the relation between the gas flow nebulizing a liquid flow and the pressure decay at the nebulizer nozzle is given by the Nukuyama–Tanasawa equation. The Sauter diameter, being the diameter of the droplets for which the volume-to-surface ratio equals that of the complete aerosol, is given by Equation (22):

$$d_0 = \left(\frac{C}{v_G}\right)^{C''} \left(\frac{\sigma}{\rho}^{\frac{1}{2}} + C' \frac{\eta}{(\sigma\rho)^{\frac{1}{2}}}\right)^{C'''} 1000 \left(\frac{Q_L}{Q_G}\right)^{C'''} \quad (22)$$

where v_G is the gas velocity, Q_G the gas flow, Q_L the liquid flow, η the viscosity, ρ the density, σ the surface tension of the liquid, and C, C', C'' and C''' are constants. When the nebulizer gas flow increases d_0 becomes smaller, the sample introduction efficiency increases and so do the signals. However, as more gas is blown through the plasma, it is cooled and the residence time of the droplets decreases so that atomization, excitation and ionization also decrease. These facts counteract the increase in the signals as a result of the improved sampling. Maximum signal intensity and power of detection are thus achieved at a compromise gas flow. The size of the aerosol droplets obtained is in the low micrometer range and their diameter depends on the physical properties of the liquid as described above. Apart from PN, HHPN, as developed by Berndt[189] and based on the expansion of a liquid jet, can also be used subsequent to desolvation.[190] This, of course, also applies for USN, of which the state-of-the-art was described by Fassel and Bear.[191] The introduction of wet aerosols into the MIP requires the use of higher power and special geometries.

Some papers have described the use of PN subsequent to aerosol desolvation, e.g. in the case of a glass frit nebulizer and an He MIP.[192] At moderate microwave power, desolvation even might be superfluous, as shown in the case of a Hildebrand grid nebulizer.[193] With a Babington nebulizer it was even possible to analyze slurries directly using an MIP.[194] Toroidal plasmas were shown to well accept wet aerosols[51] and they are suitable for direct coupling with liquid chromatography (LC) (Figure 14a and b). In this field, refinement in HPLC and capillary zone electrophoresis (CZE) is still needed. Also, here a plasma with moderate power is advantageous.[196,197] With the MPT the introduction of wet aerosols generated from aqueous and even acetonitrile-containing solutions without desolvation is possible, but with lower detection limits compared with dry aerosols (Table 4). The concomitant effects were found to be low[199] and the system could be used for the determination of noble metals.[200] The use of USN in MIP work was described e.g. by Michlewicz and Carnahan.[201]

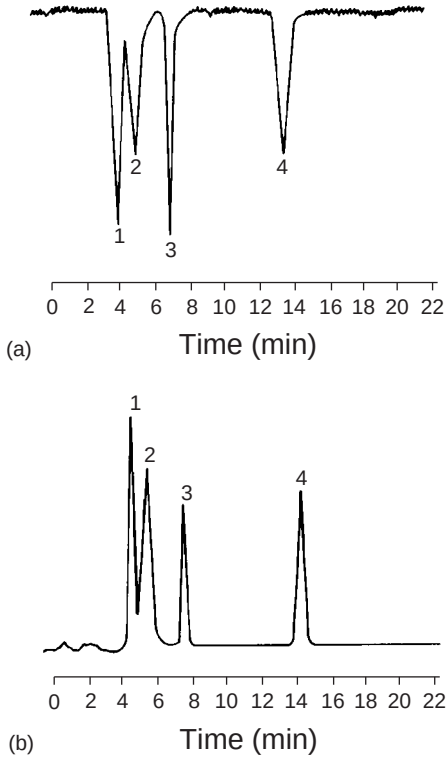

Figure 14 Ultraviolet/visible (UV/VIS) and MIP detection of different Hg species after their isocratic separation as 2-mercaptoethanol complexes by reversed-phase HPLC. Peaks: 1 = Hg^{2+}; 2 = CH_3Hg^+; 3 = $C_2H_5Hg^+$; 4 = phenyl-Hg^+; 300 ng Hg in each case. Column: RP-6, 5 µm film thickness, 250 µm × 4.6 mm i.d. Mobile phase: methanol–water (35 : 65, v/v). (a) MIP detection: Hg I 253.6 nm. (b) UV/VIS detection: absorption at 250 nm, reference wavelength 430 nm.[195] (Reproduced by permission from D. Kollotzek, D. Oechsle, G. Kaiser, P. Tschöpel, G. Tölg, *Fresenius' Z. Anal. Chem.*, **318**(7), 485–489 (1984). Copyright Springer Verlag, 1984.)

4.3 Sample Introduction in Medium- and High-power Microwave Plasmas

In microwave discharges operated at a power level of at least 150 W, many more techniques for sample introduction can be easily applied. This applies both to electrodeless MIPs and to single-electrode plasmas.

In the case of the MPT, it is possible to apply flow-cell HG, where the excess of the H_2 produced is also led into the discharge (Figure 15). The detection limits for the volatile hydride-forming elements As, Se, etc. are in the nanograms per milliliter range and in the case of hot trapping the absolute detection limits of Hg, As, Se and Sb with both He and Ar discharges are 0.1–0.2 ng.[202] Medium-power plasmas were also found to be useful for trace determinations in gases.[203] Also for element-specific detection in supercritical fluid chromatography (SFC), it was found that the high-power MIP can well cope with the excess of CO_2 produced.[204]

Table 4 Limits of detection (c_L) and upper limit of linear dynamic range (*ldr*) obtained with OES using an MPT for different sample solutions[97]

Element/line	Wavelength (nm)	Sample[a]	Without desolvation (this work)[97]		Without desolvation[198]	With desolvation[198]
			c_L (ng mL^{-1})	*ldr* (ng mL^{-1})	c_L (ng mL^{-1})	c_L (ng mL^{-1})
Li I	670.78	Water	20	200 000	–	0.99
		100 µg mL^{-1} Cs solution	20	200 000	–	–
		30% AcN	4	10 000	–	–
Cd I	228.80	Water	400	50 000	100	18
		100 µg mL^{-1} Cs solution	300	50 000	–	–
		30% AcN	2000	100 000	–	–
Cr I	359.35	Water	200	1 000 000	100	9.2
		100 µg mL^{-1} Cs solution	300	600 000	–	–
		30% AcN	300	100 000	–	–
Pb I	368.35	Water	500	500 000	800	27
		100 µg mL^{-1} Cs solution	600	500 000	–	–
		30% AcN	2000	500 000	–	–

[a] AcN = acetonitrile.

In the case of ETV the switching of gas lines for plasmas at medium and high power becomes less critical than with low-power MIPs. This has been shown for the case of the MPT, where detection limits of 3–100 pg for elements such as Ag, Au, Ge, Pb, S and Te were reported.[205] In the case of the CMP, the detection limits for direct volatilization of microsamples from a W cup are in the same range.[206,207]

With PN of solutions desolvation is often still necessary, but for the MPT[208] it becomes superfluous at powers higher than ca. 100 W.[209–211] This allows the use of direct injection nebulization (DIN)[212] and of flow injection analysis (FIA)[213] also for on-line preconcentration.[214] USN, because of the small droplet size, is very suitable, even for the the determination of nonmetals such as S[215] and C, S and P at powers above 1 kW.[216] Inexpensive air humidifiers can also be used for this purpose.[217] Even for the determination of F in solutions an He MIP was found suitable.[218] The high-power plasma torch (Figure 8) in a rectangular cavity operated with air was found to allow

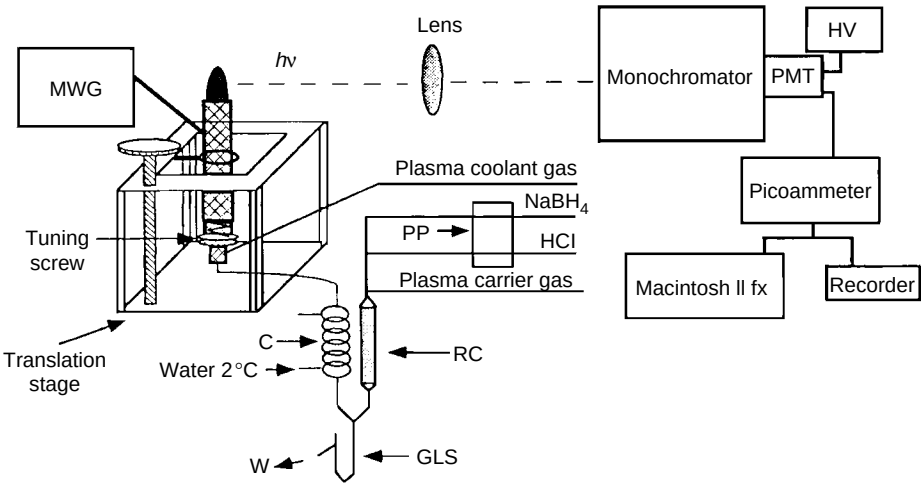

Figure 15 Schematic diagram of instrumentation for OES with an MPT and HG. PP = peristaltic pump; C = cooler; RC = reaction column (25 cm × 1.5 cm i.d.); GLS = gas–liquid separator (25 cm × 0.5 cm i.d.); W = waste; MWG = microwave generator; PMT = photomultiplier; HV = high-voltage power supply.[92]

the direct introduction of wet aerosols, but with a lower power of detection and higher concomitant effects than the ICP.[67]

4.4 Element-specific Detection in Chromatography

Microwave plasmas, owing to their compactness and the low operating costs of the plasma source and also owing to the possibility of constructing generators with lower costs than ICP generators, are valuable sources for element-specific detection in chromatography. This applies both when they are used as radiation sources for AES and as ion sources for elemental MS. A further advance lies in the dimensions of these sources, which are very small (some a few cubic millimeters) compared with ICPs (in the cubic centimeter range). The latter is advantageous both for preserving the obtained chromatographic resolution and for realizing a high absolute power of detection.

The features of MIPs both for LC and GC were outlined in reviews such as that by Jia et al.[219] and for the He plasmas especially that by Long et al.[220] In GC, He is often used as the working gas, which then is very suitable for entry into microwave plasmas as they also can be operated in He and then allow the detection of nonmetals including the halogens with a high absolute power of detection. The latter is due to the possible formation of He_2^{+m}, having energies of 18.8–21.6 eV.[221] In the case of LC, nebulization of the effluent from the chromatographic system has to be applied before excitation with microwave plasmas is possible. This is complicated by the fact that wet aerosols, at the present state-of-the-art, cannot be entered in small microwave plasma sources without significantly influencing their stability and excitation conditions. Here research efforts have to be made, either by the design of suitable nebulization systems, working with minimal effluent flows, as is now possible with CZE, or by using alternative microwave sources still to be developed. For SFC, which is becoming of increasing interest for the detection of metal chelates,[222,223] tolerance against the introduction of CO_2 must be ensured.

The possibilities of using MIPs for element-specific detection in GC have been described in several reviews[224–227] and their features as compared with AAS and MS have been outlined.[228] Many papers have described the use of microwave plasmas operated at reduced pressure for element-specific GC detection [see e.g. Hobbs[229]]. The commercially available microwave plasma detector (MPD) at reduced pressure in combination with a conventional polychromator was very useful. More recently, the MIP according to Beenakker was also used at reduced pressure for C, H and S detection.[230] However, the breakthrough first occurred with the availability of low-power microwave plasmas at atmospheric pressure coupled with GC[231] and of commercial equipment combining a gas chromatograph through a heated transfer line with an MIP and a suitable diode-array spectrophotometer, such as available from Hewlett-Packard (Figure 16).[233,234] Herewith, a system using an MIP for element-specific detection also of organometallic compounds at the trace level after GC is available.[235–237] However, this system makes use of a plasma as described by Beenakker, which is not the only possible way, as shown by work using a surfatron MIP[238] with the same spectrometer or by systems constructed from different components.[239] Here both glass capillaries[240] and fused-silica capillaries[45] can be used.

Comparisons were made between MIPs operated with He only and those operated with He–Ar mixtures.[241] With respect to the types of MIPs used, the surfatron,[242] the MIP according to Beenakker and other types were compared.[243] Such studies also include noise studies, as was done in the case of the surfatron[244] and the MPT.[245] With respect to torch design, both a tangential flow torch and a capillary water-cooled MIP torch can be used, where the latter was found to result in the best sensitivity.[246]

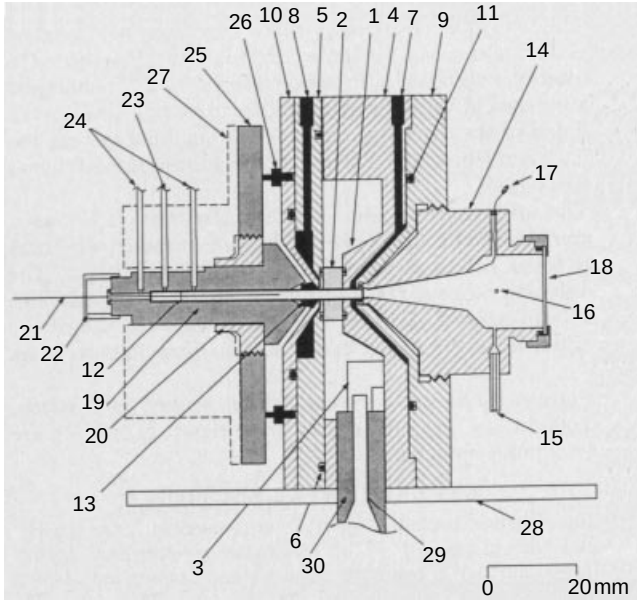

Figure 16 Re-entrant cavity for gas chromatography/microwave-induced plasma atomic emission detector (GC/MIPAED) system. (1) Pedestal; (2) quartz jacket; (3) coupling loop; (4) main cavity body; (5) cavity cover plate; (6) gasket; (7,8) cooling water inlet and outlet; (9,10) water plates; (11) O-ring; (12) silica discharge tube; (13) polyimide ferrule; (14) exit chamber; (15,16) window purge inlet and outlet; (17) sparker wire; (18) window; (19) gas union; (20) threaded collar; (21) column; (22) capillary column fitting; (23) makeup and reagent gas inlet; (24) purge flow outlets; (25) stainless steel plate; (26) standoff; (27) heater block; (28) mounting flange; (29) brass center conductor; (30) PTFE coaxial insulator.[232] (Reprinted with permission from B.D. Quimby, J.J. Sullivan, *Anal. Chem.*, **62**(10), 1027–1034 (1990). Copyright 1990, American Chemical Society.)

Further, different torch tube materials can be used, such as BN and quartz. The former material is brittle but is more resistant to etching by chromatographic effluents such as acetone and acetonitrile. Laminar-flow torches were found to give better results in the determination of empirical formulas than the other types, as shown in the case of dioxins and pyrethroids.[247] Apart from the already mentioned types of MIPs, microwave plasmas generated at about 15 W[248] between the tip of a hollow metal electrode (e.g. the GC capillary end) and an annular electrode surrounding the tip have also been used.[249] Power modulation of the plasma at 10 Hz has also been applied and found useful to maximize the excitation efficiency of the MIP.[250] For the detection of dioxins the use of a tangential torch, of low flow rates and of low power, was found to give the best sensitivity.[251] It was even found possible to use the reflected power of the MIP for registration of the chromatogram.[252] Also the VIS wavelength range often is used in the case of nonmetals.[253,254] Here Fourier transform spectrometry is very useful and allows multiwavelength measurements.[255] Apart from rather expensive dispersive spectrometers using a diode array, interference filtering[256] can also be used. Attention must also be given to the acquisition of the spectral background, which may be influenced by the effluent. This can be done by using rapid scanning techniques, as has long been known.[257] Multicapillary GC is very useful for obtaining high throughput at high resolution, as shown for hydrocarbons and organotin compound separations.[258] On-column detection is also useful in elemental ratio detection,[259] as then the chromatographic resolution is maintained. This has been described for the characterization of an organomercury reference material held on microcolumns.[260] Also the use of zero-dead-volume cross and PTFE transfer lines in headspace work for organomercury detection has been shown to be useful in minimizing memory effects.[261] Further, provisions for solvent venting have to be realized in the coupling between GC and MIP, as already known from low-pressure MPD.[262] Here heated Al tubes are used and near to the plasma one employs all-glass open-split connections[263] and zero-transfer line systems.[264] The He MIP can even be operated as a very soft electron source and used as an alternative to a β-emitter in electron-capture detection.[265]

Many papers have dealt with the optimization and use of the gas chromatography/atomic emission detector (GC/AED) system for element-specific detection. It has been increasingly used for speciation in environmental samples for both metals and nonmetals [for a state-of-the-art report, see e.g. Lobinski and Adams.[266]]. During optimization special attention has to be paid to memory effects arising from deposits and also from the detector response, which could become of relevance in high-speed GC.[267] For the optimization factorial analysis was applied and response surfaces were studied.[268] Influences of the compound structures on the elemental response obtained were also investigated.[269] Further, noise studies were performed and showed some flicker noise contributions of the plasma depending on the concentration range of the analytes.[270] Detection with a microwave-induced plasma/atomic emission detector (MIP/AED) was found to facilitate considerably the structure of gas chromatograms as compared with those obtained with conventional detection, as shown for polychlorinated biphenyl (PCB) mixtures[271] and also for amines[272] and carboxylic acids[273] both after derivatization. As one works with an He plasma nonmetals can be detected with low detection limits.[274]

In view of the possibility of elemental analysis of organic compounds, efforts were made to determine O selectively. This was found to be difficult,[275] because of the need to exclude contributions from the atmosphere. This necessitates purification of the scavenger and carrier gas[276] but also optimization of the viewing position in the plasma,[277] of the gas flow rates[278] and of the addition of Cu as catalyst and of I_2 to reduce the spectral background for O.[279] Also the detection of H was optimized.[280] As the signals of deuterated and nondeuterated compounds can often be separated by GC/AED this is of interest for the control of deuteration experiments. This can be performed both in solution and in the gas phase with suitable provisions,[281] but unfortunately the spectral lines of ^1H and ^2H partially overlap.[282] The method was shown to be suitable in the separation of the deuteroisotopomers of caffeine.[283] The detection of C has also been carefully studied. After optimization ^{13}C can be detected down to the sub-picograms per second level.[284] Owing to the experience with the detection of C and H, alkanes could be unambiguously identified in gas chromatograms, based on the C/H ratios[285] and empirical formula determinations can also be performed.[286] This has been shown, for example, for polymer pyrolyzates.[287,288] As N can also be well determined,[289] the GC/AED system could be applied to a wide diversity of organics.

The halogens certainly can be detected with an He MIP, as shown with a system using an oscillating filter-based detection system, developed by Mueller and Cammann.[290] Accordingly, empirical formulas can also be determined for chlorinated hydrocarbons[291] even including the fluorinated compounds[292,293] and quantitative determinations of dioxins and dibenzofurans were shown to be possible in the micrograms per milliliter range.[294] The results agreed with those of MS detection and also in the case of determinations of chlorophenols.[295] However, it was shown

that the structure of the halogenated compounds can influence the elemental signals obtained.[296] Further, brominated compounds can also be detected, as shown for hydrocarbons[297] or after derivatization with phenylboronic acid.[298] For S the power of detection of the MIP was shown to be better than that with chemiluminescence detection.[299] With the S 525.45-nm line a high power of detection and linearity were reported.[300] The element-specific detection of B was shown to permit the determination of organoboron compounds in motor and lubricating oils by GC/AED.[301] Detection of Si down to a few picograms as required for the determination of organosilicon compounds was shown to necessitate the use of plasma discharge tubes made of BN or alumina.[302]

Organoarsenicals can also be determined by GC/AED. This could be shown for the determination of the methylated arsines together with organomercurials in oil shale[198] and for the analysis of complex organoarsenical mixtures,[303] also after derivatization.[304] Futher interesting applications of GC/AED analysis are in the analysis of petroleum-related products for dioxane,[305] the analysis of gasoline for oxygenated compounds,[306] the determination of the halogenated products of humic substances in water,[307] pesticide residue analysis[308] and dioxin detection.[309] The use of a GC/MIPAED for the determination of molecular impurities such as CH_4 in Ar[310] or for the determination of organomercury compounds in air, after trapping them on Chromosorb®, has been described.[311]

Apart from nonmetals, metal chelates can also be determined with GC/MIPAED and its application to metal speciation has been widely studied. The latter field especially developed through coupling of chromatography and atomic spectroscopy (for a review see Lobinski[223]). Such work was initiated in 1980[312] and has now been performed for Fe, Ge, Hg, Ni, Pb, Se, Sn and V species, as discussed by Lobinski and Adams.[313] Organotin determinations in water samples and sediments were performed, including a thorough optimization of the sample preparation and the measurement conditions.[314] Also a derivatization by a Grignard reaction often is applied [see e.g. Tutschku et al.[315]]. The determination of alkyllead compounds was described very early.[316,317] A review on the optimization of microwave-induced plasma/gas chromatography (MIP/GC) determinations of organolead and organomercury compounds with the aid of simplex methods was given by Greenway and Barnett.[318] For the speciation of Hg, however, cold vapor atomic absorption spectroscopy (CVAAS) and AFS are very powerful. The determination of methylmercury in biological samples has been performed with headspace techniques.[261,319–321] After derivatization, methylmercury could be determined in fish subsequent to microwave-assisted extraction.[322] In addition to speciation, pre-enrichment for inorganic analysis can also be performed,[323] as shown by the preconcentration of Be from natural waters with the aid of acetylacetonates.[324] The use of acetylacetonates could also be applied in the determination of Ni and Cu in ores.[325] Further fluorinated and non-fluorinated tetradentate β-ketoaminate ligands could be used in the determination of Pd.[326] Se(V) could be determined by volatilization after ethylation with tetraethylborate[327] and volatile Ni, V and Fe porphyrins could be determined directly in crude oils.[328]

Also for element-specific detection in LC, carefully optimized types of MIPs already have some prospects. This is certainly true when using desolvated aerosols, as shown by Billiet et al.[329] The latter especially makes sense with USN, after which halides and oxyhalide salts subsequent to anion exchange could be determined.[330] Further, dry aerosol generation with a moving-wheel system is also possible.[331]

PN without desolvation also could be shown to be of use. Kollotzek et al.[195] used a toroidal mixed-gas MIP in a Beenakker cavity to obtain element-specific signals for Hg. Also the modified MPT can take up wet aerosols without applying desolvation. This is even possible for acetonitrile-containing solutions, but with deterioration in the power of detection, as shown e.g. also for Cr.[97] In the case of Hg the problems can be solved by an on-line CV technique between the exit of the HPLC system and the MIP. However, care must be taken to minimize dead volumes so as not to deteriorate the chromatographic resolution.[332] Wet aerosols produced with a frit nebulizer in the case of microcolumn HPLC with some restrictions can be led into MIPs.[333] In CZE work the effluent rates are low. When using an ion-exchange membrane as an electrical junction in front of the capillary, dead volumes can be minimized and with high-efficiency nebulizers sensitive detection becomes possible.[334] Also for element-specific detection in ion chromatography (IC)[335] the use of MIPOES has been described.

For element-specific detection in SFC,[336] the gases used are very important. Mostly CO_2 is used and it was shown that in the case of an He MIP a sensitive detection of S, Cl and P is possible.[337] The use of Xe both for atmospheric- and low-pressure plasma sources cannot be justified.[338] Also the use of N_2O as mobile phase was reported to allow the sensitive detection of S.[339] Both CO_2 and H_2 addition, however, were shown to decrease the emission intensities of Cl, Br, I, S and P considerably.[340] Binary mobile phases such as CO_2 with methanol addition also have prospects especially for the extraction of metal chelates and these mixed phases were shown to be tolerated by the MIP.[341] Further, the use of medium-power MIPs was shown to be advantageous so as to reduce the influence of the mobile phase of the stability

of the MIP.[342] Jin et al. showed that the MPT was very useful in the case of SFC and in OES they reported a detection limit for Cl of 300 pg.[343] A surfatron can also be used for element-specific detection in SFC for the determination of S-containing aromatics.[344] The NIR lines can also be used, as shown for the detection of S and Cl.[345] With CO_2 as mobile phase and an He MIP, Zhang et al. reported a detection limit for Cl of $40\,pg\,s^{-1}$.[346] Webster and Carnahan also proposed the use of internal standards to cope with the influence of the effluent gases on the plasma in the case of nonmetal detection.[347] With the example of ferrocene, it could be shown that the approach is also viable for the isolation and determination of organometallics.[348] In the case of packed columns, the detection limit of supercritical fluid chromatography/microwave-induced plasma (SFC/MIP) detection of Fe was reported to be 30 pg.[349]

4.5 Glow Discharges

Since the work of Paschen on hollow cathodes,[350] the GD has become well known as a source for atomic spectroscopic analyses, especially in the case of solid samples. GDs are operated at low pressure. In a classical GD there are different zones, where the formation of excited species takes place. Near the cathode positive ions are accelerated and impact on the cathode practically without any energy exchange with gas ions (dark zone), whereas in the negative glow the free electron concentrations are high and considerable ionization of the gas takes place. Near the anode the energy conversion is again very low. Accordingly, volatilization, excitation and ionization processes differ widely from those in plasma sources operated at atmospheric pressure. As outlined in state-of-the-art reviews,[351,352] the samples to be analyzed mostly constitute or are placed in one of the electrodes, preferably the cathode. From here they can be volatilzed by cathodic sputtering resulting from the impact of heavy gas ions formed in the negative glow of the discharge. Often heating of the cathode also occurs and this still enhances sample volatilization and in some cases it is the process mainly responsible for material volatilization. Excitation and ionization mainly result from electron impact; however, collisions with metastables, which have a long lifetime in a discharge at reduced pressure, and charge transfer[353] may also play a role. GDs may be produced by dc and also with rf energy, as proposed for the spectrochemical analysis especially of non-conducting samples by Winchester et al.[354] and other groups. Indeed, as the electrons can take up energy in the high-frequency field much better than the heavy positive ions do, a bias potential is built up in front of the sample, through which the field is coupled into the working gas. Accordingly, sputtering through the presence of this field can occur as in a dc field, but also for electrically nonconducting samples. Although GDs are now routinely used for solids analysis, especially when it comes to depth profiling as is now required for new surface-improved materials (for a thorough review see Bengtson[355]), its potential is not fully used. Indeed, much of the sputtered material is not excited and certainly not ionized in the optically thin low-pressure discharge. Therefore, already in early work endeavors to achieve cross-excitation were made. Leis et al.[356] successfully designed a GD incorporating cross-excitation of the sputtered material in an MIP produced at low pressure in a resonator according to Beenakker, which is constructed just in front of the GD source (Figure 17). The system was optimized and it could be shown that owing to the increased excitation efficiencies the line intensities in OES were considerably improved, but also the spectral background intensities. It could be shown by Fourier transform high-resolution studies of line profiles[357,358] that the self-reversal in the case of microwave boosting diminishes. This proves the increase in the excitation and ionization efficiency in the plasma especially in front of the sample. For steels a gain in power of detection by up to an order of magnitude could be realized (Table 5) and the system also could be used successfully for the analysis of Al, Cu and Pb samples.[359] Similar results were also reported by other groups.[360] The analyte signal enhancements due to microwave boosting seemed to be at least as strong for a microsecond-pulsed GD.[361]

Hollow cathodes have also been widely applied as atomic emission spectrochemical sources. Here the

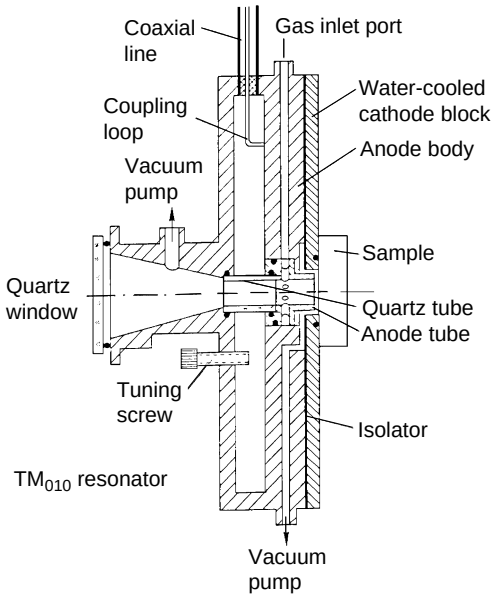

Figure 17 Schematic representation of GD lamp with integrated microwave cavity.[356]

Table 5 Detection limits for some elements in steel obtained with a microwave-supported GD lamp, calculated with an RSD of the background signal equalized to 1%; for comparison the values obtained with a conventional GD lamp are also presented[356]

Element/ line	Wavelength (nm)	Excitation energy (eV)	Detection limits ($\mu g\,g^{-1}$)	
			Without microwaves	With microwaves
Al I	396.2	3.1	0.4	0.1
B I	209.1	5.9	0.8	0.3
B I	208.9	5.9	0.9	0.4
Cr I	425.4	2.9	0.2	0.05
Cu I	327.4	3.8	1.5	0.3
Cu I	324.8	3.8	2	0.9
Mg I	285.2	4.3	2	1.5
Mg II	279.6	4.4	1.3	0.9
Mn I	403.1	3.1	1	0.2
Mo I	386.4	3.2	1.5	0.8
Nb I	405.9	3.2	4	0.6
Ni I	232.0	5.3	0.5	0.1
Si I	288.2	5.1	3	0.4
Ti I	364.3	3.4	3	0.6
V I	318.4	3.9	3	1
Zr I	360.1	3.6	8	1.5

analyte residence time in the plasma is very high, leading to high excitation and ionization efficiencies. Also here microwave boosting was found to increase the excitation and ionization efficiencies and accordingly the analytical signals.[362] These sources have been found especially useful for dry solution residue analysis; however, they also might have potential for elemental detection in GC, because of the low gas consumption and simple power supply systems.

5 OPTICAL ATOMIC SPECTROSCOPY WITH MICROWAVE PLASMAS

Many method developments in optical atomic spectrochemical procedures making use of MIPs have been described. For this purpose conventional grating spectrometers using gratings with constants down to 1/3600 mm and focal lengths of 0.5–1 m with a Czerny–Turner or Ebert set-up are used and sequential systems or simultaneous spectrometers including polychromators mainly use a Paschen–Runge mounting (Figure 18a–d). Radiation measurements are made with photomultipliers but also more and more with diode-array detection, as then line and background intensities for the atomic spectral lines can be recorded simultaneously. The latter systems also are very useful for measurements of band intensities, as shown in the case of the OH bands[96] (Figure 19a and b), which are required for the determination of rotational temperatures. Here illumination of the entrance slit with a 90° rotated image of the plasma and measurements at different heights along each rotational line allow tomography of the plasma for the rotational temperature, as shown as an example in Figure 20.

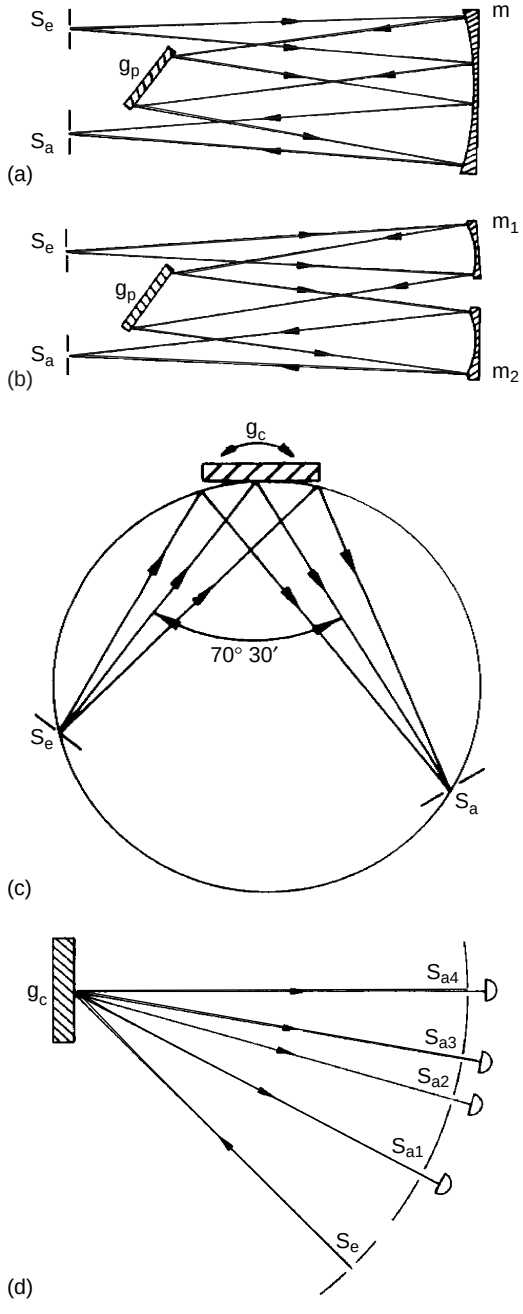

Figure 18 Optical mountings for optical spectrometers with a plane grating.[363] (a) Ebert; (b) Czerny–Turner; (c) Seya–Namioka; (d) Paschen–Runge. S_e = Entrance slit; S_a = exit slit; g_p = plane grating; g_c = concave grating; m = mirror.

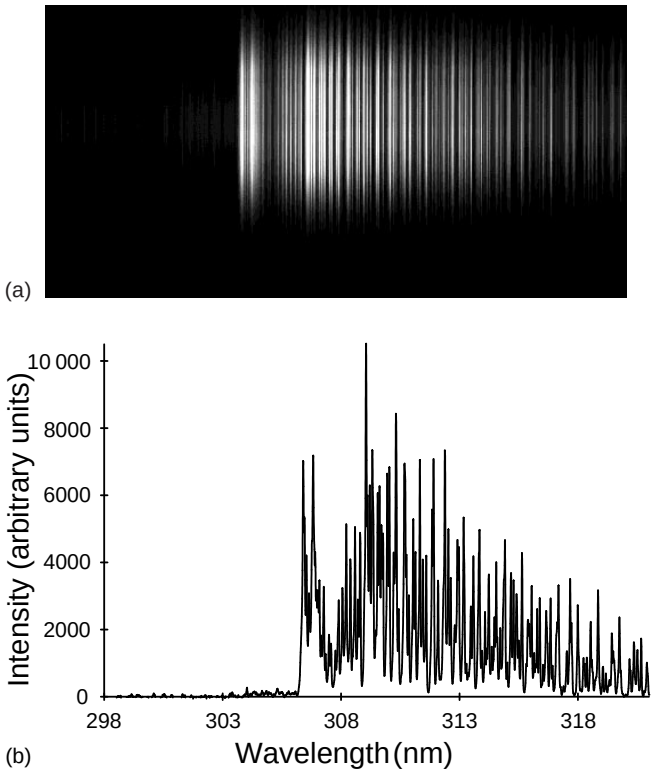

Figure 19 OH band at 306.4 nm. (a) CCD image; (b) signal of one row of the CCD image.[96]

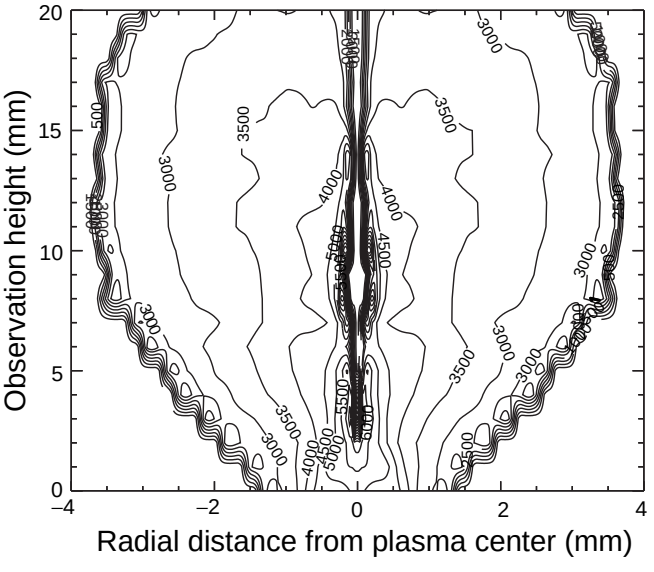

Figure 20 Radially resolved rotational temperatures in an Ar MPT (forward power 100 W, 150 mL min^{-1} working gas, 450 mL min^{-1} carrier gas.)[96]

More advanced spectrometers can also be used, such as those including an interferometer and a photodiode-array detector[364] or a Fourier transform spectrometer (for a tutorial discussion, see Faires[365]).

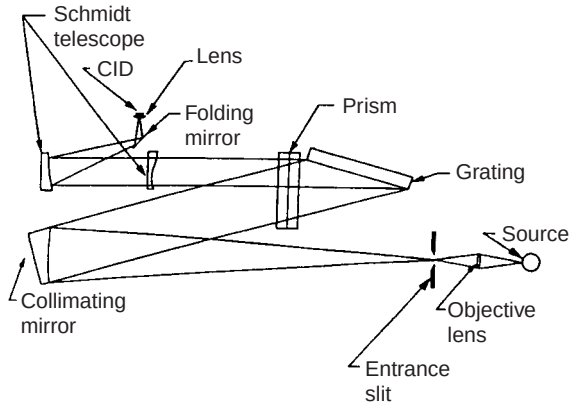

Figure 21 Échelle spectrometer with crossed dispersion and charge injection device (CID) detector for plasma spectrometery (courtesy of Perkin-Elmer Corporation).

Particularly interesting also is the use of échelle spectrometers with crossed dispersion (see Figure 21). Here an échelle grating with a small number of lines (often down to 50 lines mm^{-1}), a high order (30 and more) and a prism is employed, which has its dispersion direction perpendicular to that of the grating so as to separate the different spectral orders. Accordingly, a two-dimensional spectrum is obtained. As échelle systems require the use of a small slit height, high radiation density sources are required so as not to come into shot-noise limitations with the detector. As microwave plasmas also at low power fulfill this condition, they easily can be used together with échelle spectrometers realizing both high resolution with short focal length spectrometers, abscence from shot-noise limitations and high resolving power. This combination is particularly promising as two-dimensional array detectors then also allow simultaneous multielement detection with true background correction.

Systems with N_2-purged optics have also been used, with which the lines at vacuum-ultraviolet (VUV) wavelengths for I become accessible.[366] Method development depends, of course, on the type of microwave plasma used in particular and it will be outlined for the low- and medium-power MIPs, the single-electrode plasmas and the related stabilized capacitively coupled high-frequency plasma (SCP), as well as for atomic emission, absorption, fluorescence and laser-induced ionization.

5.1 Atomic Emission Spectroscopy with Low-power Microwave-induced Plasmas

The special features of the low-power MIPs are their low instrument and operating costs, but they are restricted in their use to those techniques where gaseous and dry analyte clouds are entered into the plasma. This has been discussed for ETV and HG and also includes the use of

volatile compound formation, as developed, for instance, for the determination of I.[367] Here detection limits below 100 ng mL^{-1} can easily be obtained. The determination of Hg can also be performed with extremely low detection limits, as here the small volume of the source, the absence of an excess of H_2 when reducing with $SnCl_2$ and the possibility of using He are very favorable conditions. Here the use of a surfatron source was shown to allow the highest sensitivity.[368] For the determination of N in natural gas, a low-pressure microwave plasma was found to permit detection limits down to 0.01 ppm by using the band heads at 336 and 337 nm[369] and also O could be determined from the OH band emission.[370] For the low-power microwave plasma sources dopants were found to considerably influence the atomic line intensities, as shown for the case of Sn analyte lines in the GC of alkylated Sn compounds.[371] Low-power microwave plasmas in the case of Ar as working gas further seemed particularly useful for the excitation of elements with low excitation potentials such as alkali metals.[372]

5.2 Single-electrode Microwave Plasmas

Single-electrode CMPs are very stable in the case of many techniques for sample introduction. They can be operated between about 200 and 600 W with different gases. It might be problematic, however, to enter the analytes in an efficient way into the hot plasma center. For this purpose the aerosols are often led into the CMP through holes near the plasma tip and wall stabilization with the aid of a mantle gas flow is used, as described by Patel et al.[373] In the case of a high-power plasma a graphite cup which contains the sample can even be an integrated part of the torch.[374] At a sufficiently high power the CMP can easily be operated with air only and, as shown by Bings et al.,[32] the excitation and rotational temperatures in this source are still in the 5000 and 4000 K range, respectively. As shown already for many elements by Zhang et al.,[375] the detection limits in aqueous solutions with this source are still in the sub-micrograms per liter range. A CMP can be operated also in He/H_2 at the tip of a graphite electrode[376] with a power of up to 800 W. The addition of H_2 was found to remove typical volatilization interferences such as the Ca^{2+}–PO_4^{3-} interference. With an He CMP Cl and Br can also be determined down to detection limits of 1 and 0.4 μg mL^{-1}, respectively, as required in the case of organohalogenated compounds.[377] With ETV Si can be determined with a detection limit as low as 0.03 μg mL^{-1} and in the case of PN 0.3 μg mL^{-1}.[378] With an He CMP organotin compounds can determined in GC effluents with a detection limit of 1 ng.[379]

When combining OES using the CMP with HG, As can be determined with absolute detection limits of down to 60 pg in the case of CT.[380] In GC, element-specific detection of the halogens can be performed.[381,382] H and O can be determined in metals by heating the samples and bringing the released gases into the CMP.[383] The plasma even is so stable that insertion of samples in microcups placed on the tip of the electrode is possible. Then extremely low absolute detection limits are obtained, e.g. 8 pg in the case of Cd.[384,385] From an external GF several heavy metals in National Institute of Science and Technology (NIST) fly ash and tomato leaf samples could be evaporated and determinend with the CMP.[386] For Hg, thermal vaporization with steam led to a detection limit of 1 μg L^{-1} with the CMP.[387] In the case of wet aerosols desolvation can be applied and just as in ICPOES it may improve the detection limits.[388] However, wet aerosols can easily be entered directly into the plasma, especially when using a tubular electrode of Ta or Al.[33,389] The plasma can also take up solvent-loaded aerosols obtained by the PN of organic solutions, which made it possible to use the CMP to perform analysis in oils eventually after dilution with the appropriate organic solvent[390] (Table 6).

As in the plasma the ground-state atom population is still considerable, the CMP can also be used as an atom reservoir for atomic absorption work.[34]

5.3 Plasma Torch Sources

In a review, Blades[391] showed the state of development of He MPTs and other microwave plasmas and compared them with the ICP. The so-called microwave-induced nitrogen discharge at atmospheric pressure (MINDAP) source described by Deutsch et al.[392,393] operates at high power and it is shown that also in a Beenakker cavity an

Table 6 Analysis of a waste motor oil, a standard oil and a fresh motor oil by ICPOES subsequent to digestion and by CMPOES (± standard deviations from five replicate analyses)[390]

Method	Sample	Co (μg g^{-1})	Cr (μg g^{-1})	Fe (μg g^{-1})	Ni (μg g^{-1})
ICPOES	Standard oil	<0.1	1.2 ± 0.1	2.1 ± 0.3	0.5 ± 0.1
	Fresh motor oil	0.6 ± 0.1	0.8 ± 0.1	2.6 ± 0.1	1.1 ± 0.1
	Waste motor oil	1.0 ± 0.1	7.1 ± 0.1	52.0 ± 2.2	4.0 ± 0.3
CMPOES	Waste motor oil	–	6.0 ± 0.9	51.4 ± 1.0	3.9 ± 0.5

N_2 discharge can accept wet aerosols with low detection limits, especially for the alkali metals. The MPT, however, became a well-studied source. A comparison with the surfatron showed that for Hg determinations the two sources have similar performances.[394] However, in the case of HG the MPT allows it to enter the excess of H_2 produced in chemical HG together with the hydrides into the plasma,[92] which also was found to be possible with a TE_{010} cavity MIP in the case of electrochemical HG.[395] In the case of the MPT, volatilization also was used for the determination of the halogens in aqueous solutions.[396] For the nonmetals also the sensitive lines in the IR region are accessible with Fourier transform spectroscopy, as shown for a high-power MIP.[397] Flow injection can be coupled on-line with OES using an MPT in the case of USN including desolvation[398] or of PN and desolvation and Cd, Cu, Mn and Zn can be determined at concentrations down to some nanograms per milliliter.[399] The necessity for desolvation, however, is not always given.[400] This could be confirmed by measurements with a Légère nebulizer, where desolvation proved to be not necessary at all.[97] Also, ion exchange can be very efficiently coupled on-line with an MPT, as shown by the determination of B using cation exchange[401] and also for metals forming chloro complexes in the case of anion exchange.[402] Ion exchange could be applied for a removal of the matrix element Fe in the determination of Si in steel.[403] The analytical performance of the MPT was also demonstrated by the determination of the rare earths in solutions using USN.[404] It was shown that with He as discharge gas, the halogens, P and S could be determined at the 100 ng mL^{-1} level.[405] At high power (>500 W), it was shown by Urh and Carnahan[406] that MPT discharges can take up wet aerosols without any problem. Then also Cl[407] could be determined with an ICP like torch positioned in a rectangular resonator. Leis and Broekaert[67] showed that with wet aerosols sub-μg mL^{-1} detection limits can be obtained but that the interferences by easily ionized elements are less favorable than in ICPOES.

5.4 Stabilized Capacitively Coupled Plasma

The SCP described by Gross et al.,[408] through its low operating power, is very similar in its features to low-power microwave plasmas. It is operated between two ring-shaped electrodes at 200 W and has a diameter of 1 mm, defined by the water-cooled plasma capillary (Figure 22). Originally designed for element-specific detection in GC, it also can be well used for the detection of nonmetals after volatilization[409] and in combination with all dry vapor generation devices, including HG without removal of the H_2, ETV and PN using desolvation.[410]

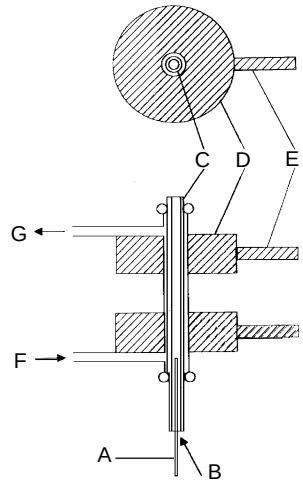

Figure 22 Cross-section of torch unit for SCP. (A) chromatographic capillary column; (B) plasma gas (He); (C) fused-silica discharge tube; (D) annular electrode (gold plated brass); (E) rf connector; (F) cooling water in; (G) cooling water out.[408] (Reprinted from R. Gross, B. Platzer, E. Leitner, A. Schalk, H. Sinabell, H. Zach, G. Knapp, *Spectrochim. Acta, Part B*, **47**(1), 95–106 (1992), Copyright 1992, with permission from Elsevier Science.)

5.5 Microwave-induced Plasmas as Atom Reservoirs for Atomic Absorption Spectroscopy, Atomic Fluorescence Spectroscopy and Laser-enhanced Ionization

As the ground-state atom population densities in the case of microwave plasma discharges are still high and the plasma volumes small, microwave discharges in a number of cases could be very well used as atom reservoirs. This was shown for the case of the MPT in which samples were introduced by USN. Here with AAS lower detection limits could be obtained than in the case of the MIP.[411] For sample introduction GF evaporation could also be combined with the MPT[412] and in the case of a Beenakker MIP PN with a glass frit nebulizer also was applied for AAS measurements.[413] The performance in the case of desolvation is improved,[414] as also for an MPT as described by Duan et al.[415] Because of the longer pathlength and decreases in matrix effects in the latter case, axial measurements are worth considering.[416]

Apart from hollow-cathode lamps, diode lasers can also be used as primary radiation sources for AAS. For the halogens a low-pressure He MIP can be used as atom reservoir, thus allowing the detection of Cl compounds through the metastable Cl levels in GC effluents.[417–419] Applications of microwave plasma torch atomic absorption spectrometry (MPTAAS) include the determination of Cd.[420]

For atomic fluorescence both line (hollow-cathode lamps) and continuous sources (Xe arc) were applied

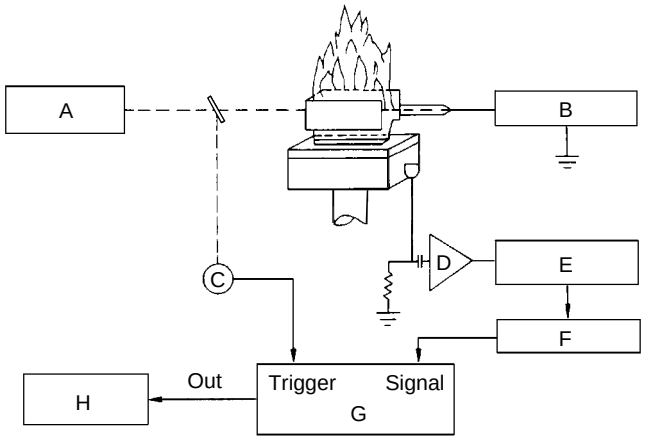

Figure 23 Flame LEI spectroscopy.[428] (A) Flashlamp/dye laser; (B) high voltage; (C) trigger photodiode; (D) pre-amplifier; (E) pulse amplifier; (F) active filter; (G) boxcar averager; (H) chart recorder. (Reprinted with permission from G.C. Turk, J.C. Travis, J.R. DeVoe, T.C. O'Haver, *Anal. Chem.*, **51**(12), 1890–1896 (1979). Copyright 1979, American Chemical Society.)

in the case of a high-power Ar MIP according to Beenakker,[421] but the detection limits obtained were poorer than those in AAS and OES. In the case of an He MIP nonmetals could also be determined[422] and detection limits of 0.25 ng mL^{-1} Cd and 120 ng mL^{-1} As can be obtained with AFS[423] and 3 ng mL^{-1} for Hg.[424] For the determination of down to 10 μg L^{-1} Na, ETV from a W filament and AFS in an MIP could be successfully applied.[425] For the determination of Fe, Ni and Co in pure water, introduction of their volatile carbonyls could be applied.[426] The full power of detection of AFS for the analysis of pure waters can only be obtained with laser excitation; only then can saturation of the excited levels be obtained, as shown in work using ETV from a W filament and an MIP.[427]

LEI (laser-enhanced ionization) can also be performed with an MIP as atom reservoir. Here the ion currents resulting from increases in ionization when populating the excited states of analytes with the aid of laser radiation of a suitable wavelength are measured and the technique is well known from flame work.[428] As the method needs no spectrometric equipment at all (Figure 23), its instrumentation costs are low but nevertheless the method is very sensitive and even isotope-selective determinations are possible. The features of a set-up using a nitrogen MIP and an Ar ion laser were shown in the case of Na, for which a detection limit of a few nanograms per milliliter was obtained.[429] Especially, the electrical characteristics of the plasma and their effect on ion detection have to be studied.[430] With a microarc atomizer and LEI in a nitrogen plasma using a flashlamp-pumped dye laser, In could be determined down to the picogram level.[431]

Comparative studies of AAS, AFS and OES with an MIP for the case of Hg revealed differences in detection limits of less than a factor of two.[432] The plasma afterglow can also be used for atomization.[433]

6 MASS SPECTROMETRY WITH MICROWAVE PLASMAS

Also for plasma MS, MIPs have been used as ion source.[434,435] They even were used earlier than the ICP. The latter became more succesful, as with aqueous solution analysis desolvation was not required. However, MIPs in their whole variety at low and high power and at atmospheric and reduced pressure, just as the ICP, are of use as ion sources for MS. This could be shown for the detection of metals and nonmetals in aqueous solutions.[436] Microwave-induced plasma mass spectrometry (MIPMS) especially found use for element-specific detection in chromatography.[437–440] Compared with OES, MS not only has the advantage of increased power of detection but it also permits the determination of isotope ratios, with a precision of better than 3% in the case of comparable abundances.[441]

6.1 Low-power Microwave Plasmas

Especially at low pressure and in the case of He as discharge gas, MIPs have been shown to be very useful ionization sources for MS and both dry vapors and as solvent-loaded aerosols[442] and organics[443] can be introduced. These working conditions especially promote microwave discharges to soft ionization sources, yielding molecular and radical species, as does electron impact ionization. Their use facilitates the identification of chromatographic peaks[444] and in this respect MIPs have similar potential to low-pressure rf plasmas.[445]

At atmospheric pressure the use of a nitrogen MIP and USN including desolvation has been described. In this case, however, contributions of N species to the background were found to be considerable.[446] With a high-power MIP operated in a surfatron, desolvation even in the case of USN was found not to be necessary.[447] Particularly with low-power MIPs in a TM_{010} cavity, fragmentation spectra could be obtained, when working under reduced pressure and at the lowest possible power.[448]

Especially in the case of He the MPT also became of use as ion source for MS. When operated at 150 W and using USN and subsequent desolvation, the nonmetals could be determined with detection limits below 1 μg mL^{-1}.[449]

Especially for GC detection, MIP/MS became recognized as very powerful.[450] This is due to the small

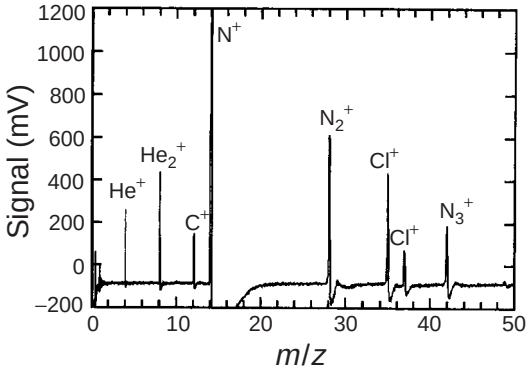

Figure 24 TOF (time-of-flight mass spectrum of CCl_4 obtained with the MPT and a flow cell sample introduction technique with the use of a digital oscilloscope for data acquisition. Peaks arising from air entrainment have been eliminated by a flowing sheath of N_2. Peak undershoot is caused by the high-speed amplifier used in this study.[453]

volume of the MIP, which keeps the deterioration of the high resolution in GC work to the minimum. The He MPT became of use for the detection of many elements by GC, including the halogens,[451] O and H.[452] For this application microwave plasma torch time-of-flight mass spectrometry (MPTTOFMS) is very useful. Here the chromatographic speed and resolution can remain high while acquiring all element and radical signals for a peak simultaneously. Accordingly, complete mass spectra, except for the range around the He ion and the N species stemming from the surrounding atmosphere and which cannot be completely removed by deflection (Figure 24), can be recorded. As the mass spectrometer ion conductance remains high compared with quadrupoles, one can achieve very low absolute detection limits, as found with halogenated hydrocarbons (see Table 7). Elemental ratio determinations and isotopic dilution in principle can also be performed with high precision compared with sequentially operating quadrupoles and sector feld instruments.

Many studies have been devoted to the detection of the halogens, because of the environmental interest in the detection of halogenated hydrocarbons. The detection of Br, Cl and I down to the picogram level is possible with a low-pressure MIP in a resonator according to Beenakker[455] when working in the positive ion mode.[456] This value is still lower in the case of MPT-TOFMS as discussed.[453] The detection of P and S, which is especially interesting from the point of view of pesticide residue analysis, could be perfomed with a water-cooled low-pressure MIP.[457] The atmospheric pressure plasma in this case was found to give more background species.[458] For such applications a tangential flow torch was often found advantageous.[454] The mass spectra for organics in the case of plasmas in O_2 or N_2 were also studied[459] and clusters identified.[460] The system can accordingly be very useful for elemental analysis.[434] However, organometallics can also be well determined, as impressively shown for the case of the alkylated Sn compounds.[461]

GC often has the disadvantage of requiring derivatization of the analytes so as to obtain volatile compounds. This is not necessary when applying SFC. In the case of halogenated compounds this could be well applied in combination with element-specific detection by MIPMS when using CO_2 and a low-power plasma.[462] HPLC could be applied for the case of halogenated compounds and also organoarsenicals, but requires the use of a 300 W MIP.[463]

Apart from its use in combination with chromatography, MIPMS in the case of vapor generation can also be applied for the determiantion of nonmetals. Here Cl, Br, I and S compounds[464,465] can be determined down to the nanograms per milliliter level. Se in water can also

Table 7 Detection limits for halogenated compounds[453] using MPTTOFMS GC

Compound	MPTTOFMS					Detection limit as total compound (Mohamad et al.[454]) (fg)
	Amount injected (pg)	S/N	Detection limit[a]			
			As halogen (fg)	As total compound (fg)	As total compound (fg)	
Chlorobenzene	10.1	118.6	75.4	2.21	260	9200
1-Chloropentane	8.8	91.6	95.2	2.72	290	–[b]
p-Chlorotoluene	12.7	114.9	91.2	2.61	330	11000
Bromobenzene	14.9	287.0	74.0	0.93	150	–[b]
Bromoform	28.0	320.7	254	1.01	270	–[b]
Iodobenzene	18.2	199.7	168	1.32	270	1500
1-Iodobutane	15.0	272.0	110	0.87	160	–[b]

[a] Detection limits were calculated at 3σ and were based on peak heights.
[b] Detection limits for these compounds were not reported.

be determined after its volatilization as a complex into MIPMS[466] and halogens and Br-containing species after volatilizing their derivatives.[467]

6.2 High-power Microwave Plasmas

The introduction of high-power microwave plasmas as ion sources for MS is aimed at the use of conventional nebulization of solutions without need for desolvation. This is possible at 150 W when using a micronebulizer: for Pd, Ru and Rh detection limits down to 0.04 $\mu g\,mL^{-1}$ could be achieved in this way with AES[468] and with an He MIP at 350 W detection limits for Se and Cd of 19 and 0.9 $ng\,mL^{-1}$, respectively, could be obtained with MIPMS, when applying direct solution analysis using DIN.[469] At a moderate power of 500 W, an N_2-operated MIP can easily accept nondesolvated aerosols. When using a water-cooled torch[470] or an Al_2O_3 tube,[471] the detection limits for Ca and K were shown to be lower than those with the ICP. The use of an annular nitrogen microwave plasma at 1.3 kW removes all restrictions for solution work and in MS detection limits down to the picograms per milliliter level have been reported.[472] The source was used for the analysis of biological samples using calibration by isotopic dilution[473] and even work with high salt concentrations was found to be possible.[474] The source also can easily cope with organics when using O_2 as the working gas, as shown in the analysis of photoresists.[475]

7 APPLICATIONS OF MICROWAVE PLASMA ATOMIC SPECTROSCOPY

The applications of microwave plasma atomic spectroscopy narrowly hang together with the types of available instrumentation. Therefore, element-specific detection in GC for residue analysis and now also for metal speciation is a major field of study. A review dealing with topics from the coupling of AAS to GC to the use of the MIP illustrates this development[476] and the importance of coupled techniques for speciation in general is now obvious.[477] As shown e.g. by the determination of organomercurials in biological materials,[478] the coupling of the MIP to GC is now well developed. There are, however, precautions to be taken in calibrating for organometallic compounds as a result of their stability [for organomercury compounds, see e.g. Lansens et al.[479]] and there is certainly room for advanced spectroscopic concepts[480] making the system cheaper.

7.1 Analysis of Biological Samples

As shown in a review,[481] most MIP work for the analysis of biological samples makes use of volatile compound formation using hydrides or GC subsequent to derivatization [see Krull et al.[482,483]] or ETV. For biological studies ^{15}N detection is very important, as discussed by Deruaz et al.,[484] and makes thorough optimization of column separation and MIP detection necessary.

With respect to plants, volatile elements have been determined in tobacco using sampling of fumes on Tenax® tubes and thermal desorption with subsequent gas chromatography/mass spectrometry (GC/MS) or GC/AED.[485] Pesticide residue analyses,[486] e.g. in the case of fruit,[487] also are of crucial importance and may make use of Cl-specific MIP detection. Further, the volatile hydride-forming elements As, Sb and Se have been determined in Chinese tea, and electrochemical HG was found to give detection limits as low as the chemical conventional method using $NaBH_4$.[395] In the case of the speciation of Se, GC has to be included, as shown for the determination of selenomethionine in wheat.[488] The power of detection of these procedures is well illustrated by the analysis of human garlic breath for organoselenium compounds.[489] Also in beer, S compounds and coffee can be determined with gas chromatography/microwave-induced plasma optical emission spectrometry (GC/MIPOES) subsequent to suitable trapping.[490] With respect to organometallic species, especially the determination of organolead in wine has been studied[491] and for very volatile species such as the organomercurials headspace techniques were optimized.[492] ETV from a GF, as already shown by Aziz[119] (Figure 25), can be very well combined even with a low-power MIP and used for determinations in plant materials. This also applies to the direct determination of very volatile elements in solids such as As[493] in leaves and for Pd, Cd. With selective volatilization speciation also can be realized, as shown for Hg and As compounds.[494]

In blood[495] and in water and urine,[496] F-containing compounds could be determined with GC coupled to MIPOES. As metals, and especially total Hg concentrations[497] and organomercurials,[498] and also Pb have been determined. For the determination of Pb in blood, the CMP could be used very successfully when applying dry residue analysis in a W cup placed on top of the electrode.[499–501] In fish tissue especially organomercury compounds have to be determined[502,503] and also the element Se [see e.g. Tsunoda et al.[504]]. As has been determined in human air, pig kidney and mussel tissue by electrochemical HG combined with hot trapping and CT.[505] For the determination of organotin compounds in tissue, microwave-assisted leaching procedures have been shown to be very helpful,[506] but clean-up of the extracts and derivatization of the organotin compounds are required.[507]

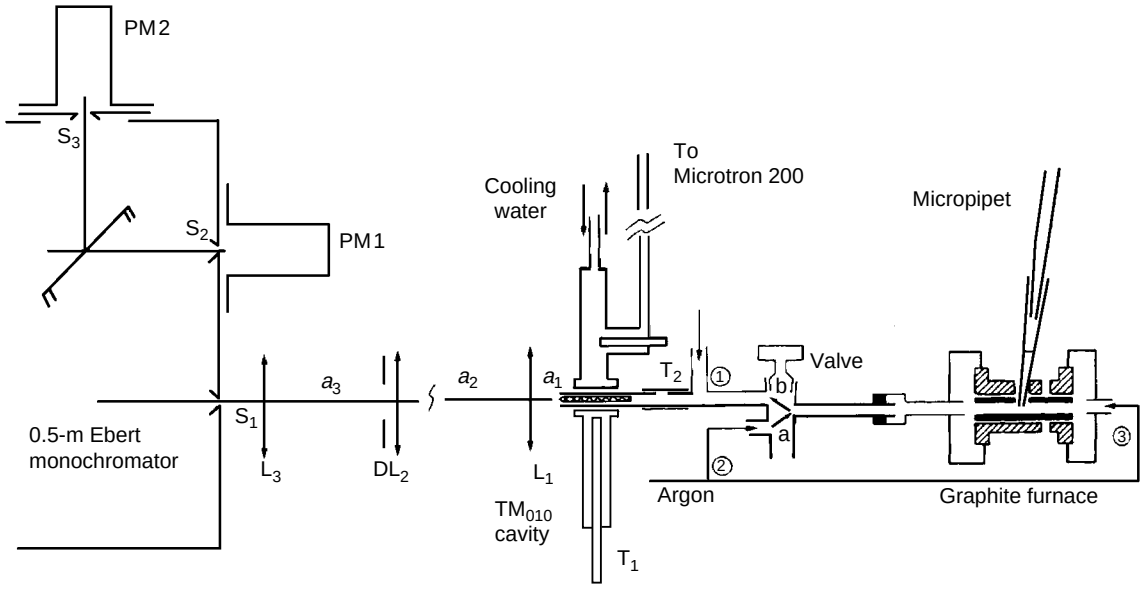

Figure 25 GF MIP instrumentation. Ar gas flows: (1) plasma gas flow; (2) for exhaust of solvent vapor; (3) carrier gas flow. (T_1) Coupling rod; (T_2) fine tuning rod. Spherical quartz lenses (f = focal length, d = diameter): (L_1) $f = 35$ mm, $d = 19$ mm, $a_1 = 41$ mm; (L_2) $f = 66$ mm, $d = 19$ mm, $a_2 = 249$ mm; (L_3) $f = 82$ mm, $d = 19$ mm, $a_3 = 90$ mm. (S_1) Entrance slit; (S_2) fixed exit slit; (S_3) movable exit slit (wavelength coverage: $\lambda_a \pm 2$ nm); (PM_1, PM_2) photomultipliers; (D) diaphragm (6×6 mm).[119]

7.2 Analysis of Environmental Samples

MIPs have been used both for the determination of the elements and for the determination of the species in water, soil and air. For this purpose hyphenated techniques, coupling chromatography to atomic spectroscopic methods, are often used so as to realize on-line pre-enrichment or separation and the determination of elemental species. For this purpose, GC, LC and SFC have been used (for a review see Dai and Jia[508]). With this approach organics, metals and organometallic species have been determined in various types of water samples.

For organics, GC coupled to MS and to MIP AES have been shown to be complementary for screening tests. This has been shown e.g. for the case of atrazine, where a peak identification was possible from the elemental signal ratios for C, H, Cl, Br, N and S on the one hand and from library mass spectra on the other.[509] Further, the power of detection of the GC/MIPAED technique for organics using empirical formula determinations, as required for screening in river water, could be considerably increased by preconcentration on LiChrolut EN® microcartridges.[510] Especially organohalogens have long been determined in waters by MIPOES.[511] Cl-containing organics, such as trihalomethane[512] and trichloroacetic acid,[134] have long been determined by precolumn enrichment and microwave plasma detection. ''Purge-and-trap'' prior to pyrolysis with subsequent GC using MIP detection was used for the determination of 20 halogenated hydrocarbons in drinking water.[513] For pesticide residue analysis, element-specific detection for C, H, N, O, S, P, F, Cl and Br with a commercially available GC/MIPAED system (Hewlett-Packard) was shown to be very useful.[514] Pyrolysis, known from GC/MS work, could be shown to be useful also in the case of MIP/AED for improved elemental detection both for chlorinated[515] and for S-containing species.[516] In the latter case proteins and S-containing amino acids could be detected very sensitively after converting S into H_2S and trapping the latter by freezing.

In synthetic ocean water, metal determinations (Cr, Mn, In, V, Pb and Sr) could be performed by the direct introduction of aerosols produced by a MAK pneumatic nebulizer without the need for desolvation.[517] For elements such as Pb a pre-enrichment as dithiocarbamate and direct GC/MIPAED were found to be very useful in improving the power of detection as required for the analysis of rain and tap water.[518] For volatile and volatile compound-forming elements very sensitive detection also is possible. This has been shown for the case of I, which chemically could be volatilized from brines or seawater,[519] for Hg, which could be determined with the CV technique[520] and after Au amalgamation down to the nanograms per liter level in lake water,[521] and for As, which could be determined in sewage sludge when using a CMP and HG.[522]

Especially more recently much work has been done on the determination of organometallic species in water samples. Especially for organotin, organolead

and organomercury compounds, derivatization was found very useful.[523] Both ethylation and phenylation with sodium tetraethylborate or tetraphenylborate, respecively, and Grignard reactions are useful for converting the analytes into thermally better stable and more volatile compounds. Organolead species in this way could be determined in River Danube water down to the nanograms per milliliter level[524] and organotin compounds in water from the Thermaikos Gulf.[525] In the case of Sn, derivatization with tetraethyl borate could be performed on-line for the case of river water.[526] For drinking water an extraction of 30 mL with hexane was found to be sufficient to allow the detection of organo-lead and -tin compounds, the latter down to the $40\,ng\,L^{-1}$ level.[527] Much work has been done on the detection of the organomercurials in water samples, as already mentioned for the case of seawater in early MIP work.[528] Owing to the labile character of many organomercurials and organolead compounds, filtration through a dithiocarbamate-containing cartridge has been proposed and tested for pre-enrichment in the case of natural waters[529] and seawater;[530] however, it was found to be problematic for long storage. For integrated sample preparation and speciation, purge-and-trap injection GC is very useful. Here sodium tetraethylborate is added to the water sample and the mixture is purged with He while trapping the analytes at $-100\,^{\circ}C$ and applying subsequent flash desorption and GC analysis. Accordingly, organo-tin, -lead and -mercury compounds could be determined in estuarine waters down to the nanograms per liter range.[531] For organomercury determinations in water samples, adsorption of the derivatives on a nonpolar microcolumn also was applied.[532] For the case of organoselenium compounds in water samples, purge-and-trap procedures were also found to be very useful.[533]

The determination of organolead species in ice samples was also found to be interesting as it delivers archive data on the nature of Pb pollution. When using derivatization and solvent venting procedures, the analysis of Greenland ice was shown to be possible down to the femtograms per gram level.[534]

The determination of traces of organometallic compounds in soils and sediments puts high requirements on the sample preparation. The determination of Pd in spent catalysts in waste from a production plant was found to be possible by acid extraction and analysis of the extracts by MIPOES using USN. In the extraction of soils with organic solvents, recoveries of organotin species were determined[535] and the organo-lead, -tin and -mercury analytes were derivatized with a Grignard reaction. As a soft extraction method for methylmercury determinations in sediments, supercritical fluid extraction with CO_2 was found to be promising.[536] ^{15}N is an important tracer for studies on the N cycle in agriculture and it was shown that both ^{14}N and ^{15}N can be determined separately in soils by MIPOES on converting NH_3 nitrogen into N_2 with NaOBr and measuring the intensities of the N_2 bandheads.[537] Volatile arsenic compounds could be determined in soils after their thermal volatilization and MIPOES.[538]

With respect to gas analysis, organobromine compounds and hydrogen bromide could be determined in car exhaust gases after trapping them on Tenax® and applying thermal desorption with subsequent GC/MIPAED.[539] Also low-molecular-weight S compounds could be determined in air with MIPOES.[540] A direct determination of gaseous and particulate Pb in air could be performed with an air CMP[541] and a detection limit of $5\,\mu g\,m^{-3}$ was reported.

For the analysis of petroleum- and natural gas-related products, MIPOES is also of use.[542] Organomercury species, for example, could be determined in natural gas using on-line amalgamation for trapping or solid-phase microextraction prior to GC/MIPAED,[543] and detection limits at the $0.2\,\mu g\,L^{-1}$ level could be obtained for the various species.

7.3 On-line Monitoring

Because of the low capital and operating costs of microwave plasma discharges compared with ICPs, they are very useful for on-line monitoring purposes where metal concentrations in gas flows or in flowing liquids have to be monitored continuously. This growing need is introduced by the requirements for process control so as to optimize industrial processes with respect to product quality, costs and environmentally relevant emissions. Especially the cost effectiveness but also the fact that microwave plasmas in a number of cases can be operated with air makes them very attractive for on-line monitoring.

Microwave discharges have long been proposed for air quality monitoring.[544] Their possibilities for the determination of Hg in molecular gases were investigated in the case of an MIP operated in a TM_{010} cavity according to Beenakker.[545] When using a tangential flow torch, N_2 loaded with Hg can be entered as outer gas and detection limits in the case of OES are in the $10\,\mu g\,m^{-3}$ range. It was found that in view of possible applications of such a system for the on-line analysis of flue gases, considerable deterioration of the analytical signals occurred as a consequence of the presence of water vapor or molecular gases such as CO_2 and SO_2. It also could be shown that on-line speciation for Hg is possible. Indeed, by the use of a KCl-coated denuder (for $HgCl_2$) and an Au- or Ag-coated denuder (for metallic and oxidized Hg) placed in-line, the most relevant species can be separated.

MIPOES, however, must still be optimized further so as to reach the detection limits of filter-based methods using determinations in decomposed filter residues by CVAAS, where a detection limit of $0.1\,\mu g\,m^{-3}$ was obtained for Hg.[546] Also for the determination of Hg in wastewater, MIPOES subsequent to the volatilization of metallic Hg can be used and here a detection limit of $0.31\,\mu g\,L^{-1}$ has been reported.[547]

The CMP, owing its stability, is much more appropriate for on-line measurements, as shown by lead determinations in air.[541] This plasma can be operated with air only at a power of some ca. 100 W and the gas flows can be varied from 0.2 to several liters per minute, allowing for flexibility in sampling, which should be performed isokinetically with the flowing medium to be monitored. The air CMP also tolerated the uptake of large amounts of organics and with both wet and dry aerosols the excitation temperatures were above 5000 K. Rotational temperatures were about 4000 K,[32] while the electron number densities were at the $10^{14}\,cm^{-3}$ level. With an air CMP operated at 600 W, detection limits for Cr, Cd, Pb, Co, Fe and Ni are at the $0.03–10\,\mu g\,m^{-3}$ level but they increase by up to 10-fold when the water loading increases above $100\,g\,m^{-3}$ or the CO_2 concentration increases to 20% (Table 8). As the air sampling certainly can still be improved, this especially shows that the CMP is a strong tool for on-line monitoring not only in complex gas mixtures such as flue gases but also for liquids, as the uptake of wet aerosols without the need for desolvation is possible.

7.4 Direct Solids Analysis

Different direct solid sampling techniques known from ICP work can also be directly used in the case of microwave plasmas. As the relevant techniques, such as SA, LA and ETV, deliver dry analyte vapors or aerosols, they do not even require any precautions in the case of low-power microwave discharges.

Direct sample insertion requires a more voluminous plasma to accommodate nonmicrosamples. For this technique the CMP could be used very successfully. In biological samples such as soils, tomato leaves and bovine liver, Cd could be determined by placing the solid samples in a W cup loaded in a graphite holder in the center tube of a quartz torch with two concentric tubes. With suitable temperature programming, absolute detection limits of 40 pg for Cd could be obtained.[549] Steel samples could also be analyzed with the same set-up and Cr, Mn, Sn and Pb could be determined, Sn and Pb with detection limits of 0.08 and $5\,\mu g\,g^{-1}$, respectively. The water content of solid samples could also be determined with the aid of OES using a CMP.[550] Here the solid samples are heated in a quartz crucible and the water evolved is carried with a He flow into the CMP, while measuring the line intensities of O and H. The technique has been applied to water determinations in biological samples (tuna fish, freeze-concentrated and dried orange juice), coal and $CuSO_4 \cdot 5H_2O$.

Table 8 Detection limits ($\mu g\,m^{-3}$) using OES with an air CMP and different CO_2 concentrations; total gas flow = $1.2\,L\,min^{-1}$[548]

Element	CO_2 (%)				
	0	3	5	11	22
Cd	0.03	0.1	0.1	0.2	0.2
Cr	1.1	0.8	1.0	1.8	3.2
Pb	9.6	15	15	22	84

Because of its small volume and low power, the MIP is only capable of evaporating very small particles (micrometer range) or of exciting vapors. This capacity can even be used for sizing particles in aerosol flows.[551] Here particles are first sampled on a filter and then aspirated into an He MIP with the aid of a scanning nozzle. The emission signals are proportional to the mass of the particles and the wavelengths present in the spectrum provide information on their composition. As the particles enter the MIP one by one, the number of signals counted provides information on their number. The apparatus is commercially available and a number of industrial and environmental applications have been reported.

ETV can easily be used for the analysis of powdered biological samples with MIPOES. Here a conventional GF can be used for sample volatilization and the lead vapor without overpressures suddenly occurring into the MIP.[552] For sample dispensing, a powder pipet can be used with which samples of down to 2 mg can be dispensed reproducibly and a calibration can be performed by standard addition with solutions. After performing a simultaneous and time-resolved measurement of line and background intensities, Mn could be determined reliably in the $10–150\,\mu g\,g^{-1}$ range and detection limits of $0.07\,\mu g\,g^{-1}$ Cu, $0.5\,\mu g\,g^{-1}$ Fe and $0.2\,\mu g\,g^{-1}$ Mn could be obtained. A real-time background correction was found necessary as small sample amounts entering the plasma could be observed visually to change its form and excitation conditions.

SA is known to produce aerosols with particle sizes in the micrometer range, as described by Raeymaekers et al.[553] The concentrations of analytes in the aerosol as a function of the sparking conditions are well representative of the composition of the solids, as investigated for steels, brass and aluminum samples. However, as shown by single-particle analysis with an electron microprobe, a redistribution of the elements over the different particle classes may occur. This may lead to matrix effects

as shown in early spark ablation/inductively coupled plasma optical emission spectroscopy (SA/ICPOES) work on aluminum samples.[554] Referring to Helmer and Walters,[555] the production of particles in SA is very complex, as both large particles, which are deposited immediately besides the spark gap, and small transportable particles are formed in a spark. The formation results both from direct ablation and melting in the spark plasma and from vapor condensation. With an Ar carrier gas flow the aerosols obtained by SA can be transported reproducibly and with low memory effects into an MIP discharge. With the aid of vidicon detection, spectra from the material produced by microwave plasma discharge excitation could be observed, which were less complex than the spark spectra.[556] In the case of a moderate-power MIP, the technique could be shown to be useful for the determination of Cr, Ni, Mn, Cu and Si in steel and of Fe, Si, Mg, Mn, Cu and Zn in aluminum alloys.[557] The spark-ablated aerosols also could be led into an MPT, which is viewed axially or laterally with an optical spectrometer. After optimizing the gas flows in the ablation chamber and the MIP, detection limits were obtained which were higher than in the case of SA/ICPOES using a sequential spectrometer. However, probably the difference by one order of magnitude is due to the differences in the spectrometers used. Further, it could be shown that the precision and the linearity of the calibration curves could be considerably improved by using a matrix line as reference.[558] A pulsed MIP could also be operated very successfully for the analysis of steels.[559]

As a direct solids sampling technique, LA has the advantage that also for electrically nonconducting samples aerosols can be produced and that at the same time laterally resolved information can be obtained. Because of the availability of very robust Nd : YAG lasers with a reproducible power output from one laser pulse to another and the growing need for the direct characterization of electrically nonconducting samples such as plastics and ceramics, the technique has been investigated in detail and combined with various radiation and ion sources including microwave plasma discharges, as described as early as 1980.[560] In the case of brass, it must be mentioned, however, that without carefully optimizing the working pressure and the laser parameters, selective volatilization can easily occur, as reflected by electron probe microanalysis in the crater, at the crater wall and of the ablated particles. As shown in Table 9, this contrasts with steel samples.[561]

LA has developed into a mature procedure. When combining it with an MIP operated in a TM_{010} resonator and when using a simultaneous échelle emission spectrometer with crossed dispersion and accordingly high resolving power, Be, Mg, Cr and Al could be determined in Cu, Ni and borax glass. High precision could be obtained when using internal standardization. Even from a single laser shot Na and Li could be determined in quartz with detection limits of 35 and 70 ng g^{-1}, respectively.[562] Fluctuations in LA resulting from power changes or surface effects could be corrected for by applying scattering light measurements and on comparing an Ar ICP with an He MIP the latter was found to give the lowest background intensities.[563] As for a few microseconds the plasma is briefly perturbed as a result of the laser pulse whereas the emission lasts for milliseconds in the case of an MIP operated at reduced pressure, it was proposed to perform time-resolved measurements.[564] When applying time-gated detection of the emission spectra combined with internal standardization of the line intensities for both the main and trace elements in steel, Cu, Al and borax glass, a matrix-independent analysis was found to be possible.[565] With LA coupled to MIPOES using a low-pressure plasma, traces of Mg, Al, Si and Fe also could be determined in high-temperature superconductors ($YBa_2Cu_3O_7$) and Na could be determined in GeO_2-doped quartz tubes down to 40 ng g^{-1}. Care must be taken to ensure an air-leak-free construction of the MIP source.[566]

Table 9 Selective ablation by LA and SA for brass and steel[561]

LA					SA			
Sample	Type	Concentration (%)			Type	Concentration (%)		
		Cu	Zn			Cu	Zn	
Brass (S 40/2)	Bulk	58 ± 1	42 ± 1		Bulk	30 ± 3	40 ± 3	
	Crater wall	66 ± 1	34 ± 1		Burning spot	71 ± 3	29 ± 2	
	Droplet	65 ± 1	35 ± 1		Particles	51 ± 6	49 ± 7	
		Concentration (%)				Concentration (%)		
		Fe	Ni	Cr		Fe	Ni	Cr
Steel (E-106)	Bulk	51 ± 2	20 ± 1	28 ± 1	Bulk	69 ± 2	18 ± 1	13.2 ± 0.6
	Crater wall	50 ± 2	20 ± 1	30 ± 1	Burning spot	70 ± 2	17 ± 1	13.4 ± 0.7

When using MIPs at higher power, direct powder injection could also be applied succesfully, as shown by the example of coal analysis.[567] C/H elemental ratios could be determined, but not Cl and S concentrations, when flushing the vials containing ground coal with He.

8 COMPARISON WITH OTHER METHODS AND OUTLOOK

Compared with other plasma sources for chemical analysis, such as plasma arcs and ICPs, MIPs have some differences which make their analytical features understandable. First they are operated at frequencies in the gigahertz range. This makes it understandable that deviations from TE are more pronounced than in high-frequency or even in arc sources. Indeed, the electrons can easily take up the microwave energy rather than giving it to the much slower gas atoms and ions. These deviations from LTE are known to lead to higher line to background intensity ratios as described in pertinent discussions in the literature. This is responsible for their high absolute power of detection. Further, microwave plasmas can be small and can often be contained in capillaries. This leads to high analyte number densities. This, however, is not always most favorable in AES as the background intensities in the analytical zone are then relatively high. Therefore, the development of toroidal plasmas, such as that realized in a resonator according to Beenakker with quartz tubes having a 4 mm internal diameter, and the use of the MPT are very important, as in their sources the analytical zone has a lower temperature than the plasma-generating zones which are outside. Microwave discharges have been shown to be very useful atom reservoirs for AAS and AFS or LEI, because of the high analyte number densities. Indeed, these ''zero-background'' types of spectroscopy are not hampered by spectral background emission in the analytical zones. In the case of MS analyses, the very small plasmas may not suffer from this drawback that limits OES, but also here larger plasmas with a better stabilized analytical zone may be advantageous so as to keep matrix influences resulting from changes in the plasma geometry low. Microwave plasmas also have the advantage over ICPs that they can easily be operated with He, which makes economic sense because of their low gas consumption. This makes microwave plasmas very useful for the determination and detection of nonmetals, which explains their success in speciation and in empirical formula determinations for organics. The fact that MIPs and especially the CMP can easily be operated on an air base only makes it very useful for on-line measurements. Indeed, here the operation costs really become low and as both wet and dry aerosols and gases can be sampled, both in process flows and exhaust gas flows, and also in wastewater, metals and nonmetals can be determined on-line. This makes microwave discharges very promising for environmental monitoring, e.g. in flue gases from waste incineration.

Microwave plasmas still have considerable potential for further development. Especially the variety of cavities which can be used for the production of different types of discharges with varying properties has hardly been exploited up to now. Here, both the demand for robust discharges with good sample uptake capacities also in the case of solvent-loaded aerosols, operation with both noble and molecular gases, the costs of operation and the ease of handling are important criteria. With respect to the last point, it was interesting how the need for retuning could be made superfluous in the case of the MPT by redesigning the device with respect to the coupler and dimensions as revealed by modeling.[63]

A further trend in development lies in the miniaturization of MIP devices together with their use in combination small spectrometers. Ideally, both the source and the spectrometer than could become plug-in accessories to be housed in computer racks. This point is of key importance for the wider use of microwave plasma detection in chromatography. Then the MIP surely could replace the conventional flame ionization detector (FID) and electron capture detector (ECD) much more easily and their capabilities for element-specific detection making the interpretation of chromatograms much easier could be more widely used.

The microwave discharges certainly should be improved with respect to their uptake of wet aerosols. Here, however, more powerful aerosol generation devices, which produce aerosols with a smaller droplet size, a narrower droplet size distribution and a better independence from the physical properties of the liquids are desirable, in addition to optimal adaption to the microwave plasmas. In this respect nebulization systems such as the DIN, HHPN, as introduced by Berndt,[189] and also the straightforward design of nebulization chambers to be used are important. Indeed, these are prerogatives for making use of the small volume and thus the less disturbing properties of MIPs in LC detection as compared with the ICP.

Finally, as ionization sources for MS, microwave plasmas also have great potential, as here He can easily be used. This reduces the problems of interferences, especially in the lower mass range as compared with inductively coupled plasma mass spectrometry (ICPMS). With molecular gases toroidal plasmas can successfully be operated, as shown by Leis and Broekaert.[67] In atomic emission they suffer from a number of limitations as a result of background intensities, but this is not the case to the same extent when using them as ion sources for MS.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AES	Atomic Emission Spectroscopy
AFS	Atomic Fluorescence Spectroscopy
CCD	Charge-coupled Device
CID	Charge Injection Device
CMP	Capacitively Coupled Microwave Plasma
CMPOES	Capacitively Coupled Microwave Plasma Optical Emission Spectroscopy
CT	Cold Trapping
CV	Cold Vapor
CVAAS	Cold Vapor Atomic Absorption Spectroscopy
CZE	Capillary Zone Electrophoresis
DIN	Direct Injection Nebulization
ECD	Electron Capture Detector
ETV	Electrothermal Vaporization
FIA	Flow Injection Analysis
FID	Flame Ionization Detector
GC	Gas Chromatography
GC/AED	Gas Chromatography/Atomic Emission Detector
GC/MIPAED	Gas Chromatography/Microwave-induced Plasma Atomic Emission Detector
GC/MIPOES	Gas Chromatography/Microwave-induced Plasma Optical Emission Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
GD	Glow Discharge
GF	Graphite Furnace
GFT	Graphite Furnace Trapping
HG	Hydride Generation
HG/AAS	Hydride Generation Atomic Absorption Spectroscopy
HG/GF/MIP	Hydride Generation/Graphite Furnace/Microwave-induced Plasma
HG/ICP	Hydride Generation/Inductively Coupled Plasma
HG/ICPMS	Hydride Generation/Inductively Coupled Plasma Mass Spectrometry
HG/MIP	Hydride Generation/Microwave-induced Plasma
HHPN	Hydraulic High-pressure Nebulization
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICPOES	Inductively Coupled Plasma Optical Emission Spectroscopy
IR	Infrared
LA	Laser Ablation
LC	Liquid Chromatography
LEI	Laser-enhanced Ionization
LTE	Local Thermal Equilibrium
MINDAP	Microwave-induced Nitrogen Discharge at Atmospheric Pressure
MIP	Microwave-induced Plasma
MIP/AED	Microwave-induced Plasma/Atomic Emission Detector
MIP/GC	Microwave-induced Plasma/Gas Chromatography
MIPMS	Microwave-induced Plasma Mass Spectrometry
MIPOES	Microwave-induced Plasma Optical Emission Spectroscopy
MPD	Microwave Plasma Detector
MPT	Microwave Plasma Torch
MPTAAS	Microwave Plasma Torch Atomic Absorption Spectrometry
MPTTOFMS	Microwave Plasma Torch Time-of-flight Mass Spectrometry
MS	Mass Spectrometry
NIR	Near-infrared
NIST	National Institute of Science and Technology
OES	Optical Emission Spectroscopy
PCB	Polychlorinated Biphenyl
PN	Pneumatic Nebulization
PN/ICP	Pneumatic Nebulization/Inductively Coupled Plasma
PN/ICPMS	Pneumatic Nebulization/Inductively Coupled Plasma Mass Spectrometry
PN/MIP	Pneumatic Nebulization/Microwave-induced Plasma
PTFE	Polytetrafluoroethylene
RSD	Relative Standard Deviation
SA	Spark Ablation
SA/ICPOES	Spark Ablation/Inductively Coupled Plasma Optical Emission Spectroscopy
SCP	Stabilized Capacitively Coupled High-frequency Plasma
SFC	Supercritical Fluid Chromatography
SFC/MIP	Supercritical Fluid Chromatography/ Microwave-induced Plasma
S/N	Signal-to-noise Ratio
TE	Thermal Equilibrium
TIA	Torche à Injection Axiale
T/MIP	Toroidal Microwave-induced Plasma
TOF	Time-of-flight
UHF	Ultrahigh Frequency
USN	Ultrasonic Nebulization

UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible
VUV	Vacuum-ultraviolet
3F/MIP	Three-filament/Microwave-induced Plasma

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Atomic Fluorescence in Environmental Analysis • Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Hydride Generation Sample Introduction for Spectroscopic Analysis in Environmental Samples • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Atomic Spectroscopy **(Volume 11)**
Atomic Spectroscopy: Introduction • Background Correction Methods in Atomic Absorption Spectroscopy • Flame and Vapor Generation Atomic Absorption Spectrometry • Flow Injection Analysis Techniques in Atomic Spectroscopy • Glow Discharge Optical Spectroscopy and Mass Spectrometry • Graphite Furnace Atomic Absorption Spectrometry • Inductively Coupled Plasma/Optical Emission Spectrometry • Laser Spectrometric Techniques in Analytical Atomic Spectrometry

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Instrumentation of Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

REFERENCES

1. R. Mavrodineanu, R.C. Hughes, 'Excitation in Radio Frequency Discharges', *Spectrochim. Acta, Part B*, **19**(8), 1309–1317 (1963).
2. U. Jecht, W. Kessler, 'Uber den Anregungsmechanismus einer HF-Fackelentladung bei 2400 MHz', *Z. Phys.*, **178**, 133–145 (1964).
3. W. Tappe, J. van Calker, 'Quantitative Spektrochemische Untersuchungen mit hochfrequenten Plasmaflammen', *Z. Anal. Chem.*, **198**, 13–20 (1963).
4. P.W.J.M. Boumans, F.J. de Boer, F.J. Dahmen, H. Hoelzel, A. Meier, 'A Comparative Investigation of Some Analytical Performance Characteristics of an ICP and a CMP for Solution Analysis by Emission Spectrometry', *Spectrochim. Acta, Part B*, **30**(10–11), 449–469 (1975).
5. R.M. Barnes, *Developments in Atomic Plasma Spectrochemical Analysis*, Heyden, London, 1982.
6. J.A.C. Broekaert, 'Plasma Atomic Spectrometry', *Nachr. Chem. Tech. Lab.*, **37**, 57–80 (1989).
7. J.A.C. Broekaert, 'Trends in Optical Spectrochemical Trace Analysis with Plasma Sources', *Anal. Chim. Acta*, **196**, 1–21 (1987).
8. B.L. Sharp, S. Chenery, R. Jowitt, S.T. Sparkes, A. Fisher, 'Atomic Spectrometry Update – Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **9**(6), 171R–188R (1994).
9. S. Greenfield, I.L. Jones, C.T. Berry, 'High-pressure Plasmas as Spectroscopic Emission Sources', *Analyst*, **89**(1064), 713–720 (1964).
10. R.H. Wendt, V.A. Fassel, 'Induction-coupled Plasma Spectrometric Excitation Source', *Anal. Chem.*, **37**(7), 920–922 (1965).
11. P.C. Uden, *Element-specific Chromatographic Detection by Atomic Emission Spectroscopy*, American Chemical Society, Washington, DC, 1992.
12. D.W. Holman, T.J. Vickers, 'Optical Emission Spectroscopy with a Microwave-induced Plasma in a Sealed Micro-tube', *Talanta*, **29**(5), 419–421 (1982).
13. C.I.M. Beenakker, 'A Cavity for Microwave-induced Plasmas Operated in Helium and Argon at Atmospheric Pressure', *Spectrochim. Acta, Part B*, **31**(8–9), 483–486 (1976).
14. J. Hubert, M. Moisan, A. Ricard, 'A New Microwave Plasma at Atmospheric Pressure', *Spectrochim. Acta, Part B*, **34**(1), 1–10 (1979).
15. Q. Jin, C. Zhu, W. Borer, G.M. Hieftje, 'Microwave-plasma Torch Assembly for Atomic-emission Spectrometry', *Spectrochim. Acta, Part B*, **46**(3), 417–430 (1991).
16. Q.H. Jin, Y.X. Duan, J.A. Olivares, 'Development and Investigation of Microwave Plasma Techniques in Analytical Atomic Spectrometry', *Spectrochim. Acta, Part B*, **52**(2), 131–161 (1997).
17. J.D. Winefordner, E.P. Wagner, B.W. Smith, 'Status of and Perspectives on Microwave and Glow Discharges for Spectrochemical Analysis. Plenary Lecture', *J. Anal. At. Spectrom.*, **11**(9), 689–702 (1996).
18. A.T. Zander, G.M. Hieftje, 'Microwave-supported Discharges', *Appl. Spectrosc.*, **35**(4), 357–371 (1981).
19. M.M. Abdillahi, 'Microwave-induced Plasmas for Trace Elemental Analysis', *Int. Lab.*, **18**(5), 16–24 (1988).
20. R.C. Culp, K.C. Ng, 'Recent Developments in Analytical Microwave-induced Plasmas', in *Advances in Atomic*

Spectrometry, ed. J. Sneddon, JAI Press, Greenwich, CT, 215–283, Vol. 2, 1995.
21. J.P. Matousek, B.J. Orr, M. Selby, 'Microwave-induced Plasmas: Implementation and Application', *Prog. Anal. At. Spectrosc.*, **7**(3), 275–314 (1984).
22. J.A.C. Broekaert, 'Current Developments in Optical-emission Spectrometry with High-frequency and Microwave Induced Plasmas', *J. Chin. Chem. Soc.*, **37**(1), 1–9 (1990).
23. J.A.C. Broekaert, 'Microwave Plasmas for Atomic Spectrometry – State of the Art and Analytical Applications', *GIT Fachz. Lab.*, **40**(4), 323–327 (1996).
24. S.R. Goode, K.W. Baughman, 'Review of Instrumentation Used to Generate Microwave-induced Plasmas', *Appl. Spectrosc.*, **38**(6), 755–763 (1984).
25. O. Zinke, H.L. Hartnagel, *Lehrbuch der Hochfrequenztechnik*, Springer, Berlin, Band 2, 1987.
26. A.J. Baden Fuller, *Mikrowellen*, Vieweg & Sohn, Braunschweig, 1974.
27. P.G. Brandl, C.L. Arnow, C.L. Caldwell, J.W. Carnahan, 'Conversion of a Pulsed 120 Hz, 3 kW, 2.45 GHz Generator for Continuous-wave Output for Microwave-induced Plasma Maintenance', *Appl. Spectrosc.*, **51**(2), 280–283 (1997).
28. M. McKenna, I.L. Marr, M.S. Cresser, E. Lam, 'Analysis of Some Diving Gas Mixtures by Microwave-induced Plasma Optical Emission Spectroscopy', *Spectrochim. Acta, Part B*, **41**(7), 669–676 (1986).
29. P. Brassem, F.J.M.J. Maessen, 'The Influence of Flow Rate and Pressure on the Excitation Conditions in Low-pressure Microwave-induced Plasmas', *Spectrochim. Acta, Part B*, **30**(12), 547–556 (1975).
30. P. Brassem, F.J.M.J. Maessen, L.de Galan, 'Excitation Conditions in a Mixed-gas Low-pressure Microwave-induced Plasma', *Spectrochim. Acta, Part B*, **31**(10–12), 537–545 (1976).
31. P. Brassem, F.J.M.J. Maessen, 'Electron Temperatures and Electron Concentrations at Low Pressure in Microwave-induced Plasmas', *Spectrochim. Acta, Part B*, **29**(7–8), 203–210 (1974).
32. N.H. Bings, M. Olschewski, J.A.C. Broekaert, 'Two-dimensional Spatially Resolved Excitation and Rotational Temperatures as well as Electron Number Density Measurements in Capacitively Coupled Microwave Plasmas Using Argon, Nitrogen and Air as Working Gases by Spectroscopic Methods', *Spectrochim. Acta, Part B*, **52**(13), 1965–1981 (1997).
33. A. Disam, P. Tschöpel, G. Tölg, 'Emission-spectrometric Determination of Trace Elements in Aqueous Solutions by Means of a Sheath-gas-stabilized Capacitatively Coupled Microwave Plasma (CMP)', *Fresenius' Z. Anal. Chem.*, **310**(1–2), 131–143 (1982).
34. G. Wünsch, G. Hegenberg, N. Czech, 'Capacitively Coupled Microwave Plasma (CMP) as Source of Atomization in Atomic-absorption Spectroscopy', *Spectrochim. Acta, Part B*, **38**(8), 1135–1141 (1983).
35. C.I.M. Beenakker, P.W.J.M. Boumans, P.J. Rommers, 'Microwave-induced Plasma as Excitation Source for Atomic-emission Spectrometry', *Philips Tech. Rev.*, **39**(3–4), 65–77 (1980).
36. C.I.M. Beenakker, P.W.J.M. Boumans, P.J. Rommers, 'Microwave-induced Plasma as Excitation Source for Atomic-emission Spectrometry. II', *GIT Fachz. Lab.*, **25**(3), 179–189 (1981).
37. C.I.M. Beenakker, P.W.J.M. Boumans, P.J. Rommers, 'Microwave-induced Plasma as Excitation Source for Atomic-emission Spectrometry', *GIT Fachz. Lab.*, **25**(2), 82–87 (1981).
38. C.I.M. Beenakker, 'A Cavity for Microwave-induced Plasmas Operated in Helium and Argon at Atmospheric Pressure', *Spectrochim. Acta, Part B*, **31**(7), 483–486 (1976).
39. N. Rait, D.W. Golightly, C.J. Massoni, 'Improved Beenakker-type Cavity for Microwave-induced Plasma Spectrometry', *Spectrochim. Acta, Part B*, **39**(7), 931–937 (1984).
40. D.L. Haas, J.W. Carnahan, J.A. Caruso, 'Internally Tuned TM_{010} Microwave Resonant Cavity for Moderate-power Microwave-induced Plasmas', *Appl. Spectrosc.*, **37**(1), 82–85 (1983).
41. J. Hubert, S. Bordeleau, K.C. Tran, S. Michaud, B. Milette, R. Sing, J. Jalbert, D. Boudreau, M. Moisan, J. Margot, 'Atomic Spectroscopy with Surface-wave Plasmas', *Fresenius' J. Anal. Chem.*, **355**(5–6), 494–500 (1996).
42. M. Selby, G.M. Hieftje, 'Taming the Surfatron', *Spectrochim. Acta, Part B*, **42**(1–2), 285–298 (1987).
43. S. Luge, J.A.C. Broekaert, 'Use of Optical Emission Spectrometry with Microwave-induced Plasma (MIP) Discharges in a Surfatron Combined to Different Types of Hydride Generation for the Determination of Arsenic', *Mikrochim. Acta*, **113**(3–6), 277–286 (1994).
44. H. Matusiewicz, 'Novel Microwave Plasma Cavity Assembly for Atomic-emission Spectrometry', *Spectrochim. Acta, Part B*, **47**(10), 1221–1227 (1992).
45. S.A. Estes, P.C. Uden, R.M. Barnes, 'Microwave-Excited Atmospheric-pressure Helium-plasma Emission Detection Characteristics in Fused-silica Capillary Gas Chromatography', *Anal. Chem.*, **53**(12), 1829–1837 (1981).
46. K.J. Mulligan, M.H. Hahn, J.A. Caruso, F.L. Fricke, 'Comparison of Several Microwave Cavities for Simultaneous Determination of Arsenic, Germanium, Antimony and Tin by Plasma Emission Spectrometry', *Anal. Chem.*, **51**(12), 1935–1938 (1979).
47. A.T. Zander, 'Atomic-emission Sources for Solution Spectrochemistry', *Anal. Chem.*, **58**(11), 1139A–1149A (1986).
48. H. Matusiewicz, R.E. Sturgeon, 'Liquid Cooling of a Torch for Microwave-induced Plasma Spectrometry', *Spectrochim. Acta, Part B*, **48**(4), 515–519 (1993).

49. R.F. Wandro, H.B. Friedrich, 'Investigation of the Interactions of a Microwave-induced Plasma with a Quartz Capillary Wall', *Anal. Chem.*, **56**(14), 2727–2731 (1984).
50. D. Kollotzek, P. Tschöpel, G. Tölg, 'Lösungsemissionsspektrometrie mit mikrowelleninduzierten Mehrfaden- und Hohlzylinderplasmen – I: Die radiale Verteilung von Signal- und Untergrundintensitäten', *Spectrochim. Acta, Part B*, **37**, 91–96 (1982).
51. D. Kollotzek, P. Tschöpel, G. Tölg, 'Three-filament and Toroidal Microwave-induced Plasmas as Radiation Sources for Emission-spectrometric Analysis of Solutions and Gaseous Samples. II. Analytical Performance', *Spectrochim. Acta, Part B*, **39**(5), 625–636 (1984).
52. G. Heltai, J.A.C. Broekaert, F. Leis, G. Tölg, 'Study of a Toroidal Argon and a Cylindrical Helium Microwave-induced Plasma for Analytical Atomic-emission Spectrometry. I. Configurations and Spectroscopic Properties', *Spectrochim. Acta, Part B*, **45**(3), 301–311 (1990).
53. A. Bollo-Kamara, E.G. Codding, 'Considerations in the Design of a Microwave-induced Plasma Utilizing the TM_{010} Cavity for Optical Emission Spectroscopy', *Spectrochim. Acta, Part B*, **36**(10), 973–982 (1981).
54. G.K. Webster, J.W. Carnahan, 'Improved Supercritical-fluid Chromatography Mobile-phase Tolerance with a Moderate-power Helium Microwave-induced Plasma', *Appl. Spectrosc.*, **45**(8), 1285–1290 (1991).
55. S.R. Goode, B. Chambers, N.P. Buddin, 'Use of a Tangential-flow Torch with a Microwave-induced Plasma Emission Detector for Gas Chromatography', *Spectrochim. Acta, Part B*, **40**(1–2), 329–333 (1985).
56. S.R. Goode, B. Chambers, N.P. Buddin, 'Critical Evaluation of the Tangential-flow Torch Microwave-induced Plasma Detector for Gas Chromatography', *Appl. Spectrosc.*, **37**(5), 439–442 (1983).
57. K.C. Ng, M.J. Brechmann, 'Demountable Torch for Combination-gas Microwave-induced Plasmas', *Spectroscopy*, **2**(1), 23–25 (1987).
58. G.S. Sobering, T.D. Bailey, T.C. Farrar, 'Simple Torch Design for a Microwave-induced Plasma', *Appl. Spectrosc.*, **42**(6), 1023–1025 (1998).
59. P.R. Fielden, M. Jiang, R.D. Snook, 'Laminar-flow Microwave-plasma Detector for Gas Chromatography', *Appl. Spectrosc.*, **43**(8), 1444–1449 (1989).
60. M.R. Jiang, P.R. Fielden, R.D. Snook, 'Performance of a Laminar-flow Torch – Microwave Plasma Detector for Gas Chromatography', *Appl. Spectrosc.*, **45**(2), 227–230 (1991).
61. M.L. Bruce, J.M. Workman, J.A. Caruso, D.J. Lahti, 'Low-flow Laminar-flow Torch for Microwave-induced Plasma Emission Spectrometry', *Appl. Spectrosc.*, **39**(6), 935–942 (1985).
62. B.W. Pack, G.M. Hieftje, 'An Improved Microwave Plasma Torch for Atomic Spectrometry', *Spectrochim. Acta, Part B*, **52**(14), 2163–2168 (1997).
63. A.M. Bilgic, C. Prokisch, J.A.C. Broekaert, E. Voges, 'Design and Modelling of a Modified 2.45 GHz Coaxial Plasma Torch for Atomic Spectrometry', *Spectrochim. Acta, Part B*, **53**(5), 773–777 (1998).
64. M.M. Mohamed, T. Uchida, S. Minami, 'Direct Sample Introduction of Solid Material into a Pulse-operated MIP [Microwave-induced Plasma]', *Appl. Spectrosc.*, **43**(5), 794–800 (1989).
65. K.G. Michlewicz, J.J. Urh, J.W. Carnahan, 'Microwave-induced Plasma System for Maintenance of Moderate-power Plasmas of Helium, Argon, Nitrogen and Air', *Spectrochim. Acta, Part B*, **40**(3), 493–499 (1985).
66. H.A. Dingjan, H.J. De Jong, 'Comparative Study of Two Cavities for Generating a Microwave-induced Plasma in Helium or Argon as a Detector in Gas Chromatography', *Spectrochim. Acta, Part B*, **36**(4), 325–331 (1981).
67. F. Leis, J.A.C. Broekaert, 'High-power Microwave-induced Plasma for Analysis of Solutions', *Spectrochim. Acta, Part B*, **39**(9–11), 1459–1463 (1984).
68. M. Ohata, N. Furuta, 'Evaluation of the Detection Capability of a High Power Nitrogen Microwave-induced Plasma for Both Atomic Emission and Mass Spectrometry', *J. Anal. At. Spectrom.*, **5**(13), 447–453 (1998).
69. C.I.M. Beenakker, 'Evaluation of a Microwave-induced Plasma in Helium at Atmospheric Pressure as an Element-selective Detector for Gas Chromatography', *Spectrochim. Acta, Part B*, **32**(3–4), 173–187 (1977).
70. P.W.J.M. Boumans, *Theory of Spectrochemical Excitation*, Hilger and Watts, London, 1966.
71. J. Jonkers, J.M. De Regt, J.A.M. Van der Mullen, H.P.C. Vos, F.P.J. De Groote, E.A.H. Timmermans, 'On the Electron Temperatures and Densities in Plasmas Produced by the "torche à Injection Axiale"', *Spectrochim. Acta, Part B*, **51**(11), 1385–1392 (1996).
72. H.R. Griem, *Plasma Spectroscopy*, McGraw-Hill, New York, 1964.
73. D. Robinson, P.D. Lenn, 'Plasma Diagnostics by Spectroscopic Methods', *Appl. Opt.*, **6**, 983–1000 (1967).
74. A. Garnier, E. Bloyet, P. Leprince, J. Marec, 'Microwave Plasma in Argon Produced by a Surface Wave: Study of the Effect of Pressure on the Optical Emission and the Potentials for Analysis of Gaseous Samples', *Spectrochim. Acta, Part B*, **43**(8), 963–970 (1988).
75. B. Riviere, J.M. Mermet, D. Deruaz, 'Behaviour and Analytical Applications of a Modulated-power Microwave-induced Plasma (Surfatron)', *J. Anal. At. Spectrom.*, **4**(6), 519–523 (1989).
76. L.K. Olson, W.C. Story, J.T. Creed, W.L. Shen, J.A. Caruso, 'Fragmentation of Organic Compounds Using Low-pressure Microwave-induced Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **5**(6), 471–475 (1990).
77. M. Moisan, R. Grenier, Z. Zahrzewski, 'The Electromagnetic Performance of a Surfatron-based Coaxial Microwave Plasma Torch', *Spectrochim. Acta, Part B*, **50**(8), 781–789 (1995).

78. J. Cotrino, M. Saez, M.C. Quinero, A. Menendez, E. Sanchez-Uria, A. Sanz-Medel, 'Spectroscopic Determination of Fundamental Parameters in an Argon Microwave-induced Plasma (Surfatron) at Atmospheric Pressure', *Spectrochim. Acta, Part B*, **47**(3), 425–435 (1992).
79. P.G. Brandl, J.W. Carnahan, 'Charge Transfer in Analytical Helium Plasmas', *Spectrochim. Acta, Part B*, **49**(1), 105–115 (1994).
80. M.P. Dziewatkoski, C.B. Boss, 'Study of Hydrocarbon Dissociation in a Microwave-induced Plasma Sustained in Ar Gas', *Spectrochim. Acta, Part B*, **49**(2), 117–135 (1994).
81. W. Zyrnicki, 'Spectroscopic Study of Argon MIP (Microwave-induced Plasma) Discharge Used for Analysis of Solutions', *Microchem. J.*, **46**(3), 356–351 (1992).
82. J. Mierzwa, R. Brandt, J.A.C. Broekaert, P. Tschöpel, G. Tölg, 'Performance of a Microwave-induced Plasma (MIP) Operated in a Liquid-cooled Discharge Tube for Atomic-emission Spectrometry', *Spectrochim. Acta, Part B*, **51**(1), 117–126 (1996).
83. J. Roepcke, A. Ohl, M. Schmidt, 'Comparison of Optical-emission Spectrometric Measurements of the Concentration and Energy of Species in Low-pressure Microwave and Radio-frequency Plasma Sources', *J. Anal. At. Spectrom.*, **8**(6), 803–808 (1983).
84. H. Jimenez-Dominguez, S. Cruz-Jimenez, A. Cabral-Prieto, 'Spectroscopic Applications of the Plasma Dispersion Function', *Spectrochim. Acta, Part B*, **51**(1), 165–174 (1996).
85. R.M. Alvarez-Bolainez, C.B. Boss, 'Noise Analysis of the Reflected Power Signal from a Microwave-induced Plasma Gas-chromatographic Detector', *Anal. Chim. Acta*, **269**(1), 89–97 (1992).
86. J.E. Freeman, G.M. Hieftje, 'Analytical Characteristics of Near-infrared Non-metal Atomic Emission from a Helium Microwave-induced Plasma', *Spectrochim. Acta, Part B*, **40**(3), 475–492 (1985).
87. D.M. Chambers, J.W. Carnahan, Q. Jin, G.M. Hieftje, 'Fundamental Studies of the Sampling Process in an Inductively Coupled Plasma Mass Spectrometer. IV. Replacement of the Inductively Coupled Plasma with a Helium Microwave-induced Plasma', *Spectrochim. Acta, Part B*, **46**(13), 1745–1765 (1991).
88. B. Kirsch, B.S. Hanamura, J.D. Winefordner, 'Diagnostical Measurements in a Single-electrode, Atmospheric-pressure, Microwave Plasma', *Spectrochim. Acta, Part B*, **39**(8), 955–963 (1984).
89. S. Hanamura, B.W. Smith, J.D. Winefordner, 'Single-electrode Atmospheric-pressure Microwave Discharge System for Elemental Analysis', *Can. J. Spectrosc.*, **29**(1), 13–18 (1984).
90. N.H. Bings, J.A.C. Broekaert, 'The Use of Different Plasma Gases (Argon, Nitrogen and Air) for Capacitively Coupled Microwave-plasma Optical Emission Spectrophotometry (CMP OES): Figures of Merit and Temperature Measurements', *Fresenius' J. Anal. Chem.*, **355**(3–4), 242–243 (1996).
91. Q. Jin, G. Yang, A. Yu, J. Liu, H. Zhang, Y. Ben, 'A Novel Plasma Emission Source', in *Abstracts Book of Pittcon '85*, Abstract No. 1171, 1985.
92. R. Pereiro, M. Wu, J.A.C. Broekaert, G.M. Hieftje, 'Direct Coupling of Continuous Hydride Generation with Microwave Plasma Torch Atomic-emission Spectrometry for the Determination of Arsenic, Antimony and Tin', *Spectrochim. Acta, Part B*, **49**(1), 59–73 (1994).
93. Y. Madrid, M.W. Borer, C. Zhu, Q.H. Jin, G.M. Hieftje, 'Noise Characterization of the Microwave Plasma Torch (MPT) Source', *Appl. Spectrosc.*, **48**(8), 994–1002 (1994).
94. M. Huang, D.S. Hanselman, Q. Jin, G.M. Hieftje, 'Non-thermal Features of Atmospheric-pressure Argon and Helium Microwave-induced Plasmas Observed by Laser-light Thomson Scattering and Rayleigh Scattering', *Spectrochim. Acta, Part B*, **45**(12), 1339–1352 (1990).
95. C. Prokisch, A.M. Bilgic, E. Voges, J.A.C. Broekaert, J. Jonkers, M. van Sande, J.A.M. van der Mullen, 'Photographic Plasma Images and Electron Number Density as well as Electron Temperature Mappings of a Plasma in a Modified Argon Microwave Plasma Torch (MPT) Measured by Spatially Resolved Thomson Scattering', *Spectrochim. Acta, Part B*, **54**(9), 1253–1266 (1999).
96. U. Engel, C. Prokisch, E. Voges, G.M. Hieftje, J.A.C. Broekaert, 'Spatially Resolved Measurements and Plasma Tomography with Respect to the Rotational Temperatures for a Microwave Plasma Torch', *J. Anal. At. Spectrom.*, **13**(9), 955–961 (1998).
97. C. Prokisch, J.A.C. Broekaert, 'Element Determination in Aqueous and Acetonitrile Containing Solutions by Atomic Emission Spectrometry Using a Microwave Plasma Torch', *Spectrochim. Acta, Part B*, **52**(6–8), 1909–1119 (1998).
98. S.R. Goode, J.N. Emily, 'Measuring the Spatial Distribution of Properties and Species in Microwave-induced Helium Plasmas', *Spectrochim. Acta, Part B*, **49**(1), 31–45 (1998).
99. Y. Okamoto, 'Annular-shaped Microwave-induced Nitrogen Plasma at Atmospheric Pressure for Emission Spectrometry of Solutions', *Anal. Sci.*, **7**(2), 283–288 (1991).
100. M. Ohata, N. Furuta, 'Spatial Characterization of Emission Intensities and Temperatures of a High-power Nitrogen Microwave-induced Plasma', *J. Anal. At. Spectrom.*, **12**(3), 341–347 (1997).
101. J. Jonkers, H.P.C. Vos, J.A.M. Van der Mullen, E.A.H. Timmermans, 'On the Atomic State Densities of Plasmas Produced by the ''Torche à Injection Axiale'' ', *Spectrochim. Acta, Part B*, **51**(5), 457–465 (1996).
102. A. Rodero, M.C. Quintero, A. Sola, A. Gamero, 'Preliminary Spectroscopic Experiments with Helium Microwave-induced Plasma Produced in Air by Use of a New

Structure: The Axial Injection Torch', *Spectrochim. Acta, Part B*, **51**(5), 467–479 (1996).
103. R.F. Browner, A.W. Boorn, 'Sample Introduction: The Achilles' Heel of Atomic Spectroscopy?', *Anal. Chem.*, **56**(7), 786A–798A (1984).
104. R.F. Browner, A.W. Boorn, 'Sample-introduction Techniques for Atomic Spectroscopy.', *Anal. Chem.*, **56**(7), 875A–888A (1984).
105. J. Sneddon, *Sample Introduction in Atomic Spectroscopy*, Elsevier, Amsterdam, 1990.
106. Special Issue: 'Sample Introduction in Atomic Spectrometry', *Spectrochim. Acta, Part B*, **50**(4–7), 271–654 (1995).
107. Special Issue: 'Flow Injection Analysis – State of the Art Applied to Atomic Spectroscopy', *Spectrochim. Acta, Part B*, **51**(14), 1733–1941 (1996).
108. L. Ebdon, 'The Joys of Coupling', *Lab. Pract.*, **38**(3), 15–20 (1989).
109. A. Sanz-Medel, A. Menendez, M.L. Fernandez, E. Sanchez-Uria, 'Tandem Online Separations: An Alternative Sample Presentation in Atomic Spectrometry for Ultratrace Analysis', *Acta. Chim. Hung.*, **128**(4–5), 551–558 (1991).
110. G. Heltai, J.A.C. Broekaert, 'Sample Introduction Problems of Atmospheric MIPs (Microwave-induced Plasmas)', *Acta. Chim. Hung.*, **128**(4–5), 599–611 (1991).
111. J.A. Caruso, 'Microwave-induced Plasmas as Sources for Atomic Spectrometry', *J. Res. Natl. Bur. Stand. (US)*, **93**(3), 447–449 (1988).
112. E.I. Brooks, K.J. Timmins, 'Sample-introduction Device for Use with a Microwave-induced Plasma', *Analyst*, **110**(5), 557–558 (1985).
113. J.A.C. Broekaert, G. Tölg, 'Recent Developments in Atomic Spectrometry Methods for Elemental Trace Determinations', *Fresenius' Z. Anal. Chem.*, **326**(6), 495–509 (1987).
114. W.M. Borer, G.M. Hieftje, 'Inductively Coupled Plasma–Microwave-induced Plasma Tandem Source for Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **8**(2), 339–348 (1993).
115. W.M. Borer, G.M. Hieftje, 'Design Considerations for a Pressure-differential Tandem Source for Use in Atomic Spectrometry', *J. Anal. At. Spectrom.*, **8**(2), 333–338 (1993).
116. N.W. Barnett, 'Some Experience with Various Sample Introduction Techniques for Microwave-induced Plasma Atomic-emission Spectrometry [MIP/AES]', *Vestn. Slov. Kem. Drus.*, **36**(1), 1–23 (1989).
117. H. Matusiewicz, 'A Microwave Plasma Cavity Assembly for Atomic-emission Spectrometry', *Fresenius' J. Anal. Chem.*, **355**(5–6), 623–625 (1996).
118. H. Matusiewicz, 'Thermal Vaporization for Sample Introduction in Microwave-induced Plasma Analytical Emission Spectrometry', *Spectrochim. Acta Rev.*, **13**(1), 47–68 (1990).
119. A. Aziz, J.A.C. Broekaert, F. Leis, 'Contribution to the Analysis of Microamounts of Biological Samples Using a Combination of Graphite-furnace and Microwave-induced Plasma Atomic-emission Spectroscopy', *Spectrochim. Acta, Part B*, **37**(5), 381–389 (1982).
120. G. Heltai, J.A.C. Broekaert, P. Burba, F. Leis, P. Tschöpel, G. Tölg, 'Study of a Toroidal Argon MIP (Microwave-induced Plasma) and a Cylindrical Helium MIP for Atomic-emission Spectrometry. II. Combination with Graphite-furnace Vaporization and Use for Analysis of Biological Samples', *Spectrochim. Acta, Part B*, **45**(8), 857–866 (1990).
121. J.F. Yang, J.Y. Zhang, C. Schickling, J.A.C. Broekaert, 'Study of a Microwave-induced Argon Plasma Sustained in a TE_{101} Cavity as Spectrochemical Emission Source Coupled with Graphite-furnace Evaporation', *Spectrochim. Acta, Part B*, **51**(6), 551–562 (1996).
122. M.M. Abdillahi, 'Analytical Feasibility of Graphite-rod Vaporization–Helium Microwave-induced Plasma (GRV–He-MIP) for Some Non-metal Determinations', *Appl. Spectrosc.*, **47**(3), 366–374 (1993).
123. J.F. Alder, M.T.C. Da Cunha, 'Evaluation of a Low-pressure Microwave-induced Argon Plasma with Carbon-rod Sample Introduction for Trace-metal Analysis', *Can. J. Spectrosc.*, **25**(2), 32–38 (1980).
124. C.E. Pfluger, T. Nessel, 'Wire-loop Microfurnace–Power Supply Controller System for the Miniature Helium Direct-current Discharge Tube Emission Method', *Analyst*, **109**(5), 593–596 (1984).
125. R.G. Stahl, L. Brett, K.J. Timmins, 'Automation of a Wire-filament Atomizer Helium Microwave-induced Plasma System', *J. Anal. At. Spectrom.*, **4**(4), 337–340 (1989).
126. K. Chiba, M. Kurosawa, K. Tanabe, H. Haraguchi, 'Micro-sampling Technique Utilizing an Electrothermal Tungsten-boat Vaporization Device for Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry', *Chem. Lett.*, 75–78 (1984).
127. W.H. Evans, J.A. Caruso, R.D. Satzger, 'Evaluation of a Tantalum-tip Electrothermal-vaporization Sample-introduction Device for Microwave-induced Plasma Mass Spectrometry and Atomic-emission Spectrometry', *Appl. Spectrosc.*, **45**(9), 1478–1484 (1991).
128. M. Zerezghi, K.J. Mulligan, J.A. Caruso, 'Simultaneous Multi-element Determination in Microlitre Samples by Rapid-scanning Spectrometry Coupled to a Microwave-induced Plasma', *Anal. Chim. Acta*, **154**, 219–226 (1983).
129. N.W. Barnett, G.F. Kirkbright, 'Electrothermal-vaporization Sample Introduction into an Atmospheric-pressure Helium Microwave-induced Plasma for the Determination of Iodine in Hydrochloric Acid', *J. Anal. At. Spectrom.*, **1**(5), 337–342 (1986).
130. J.P. Matousek, B.J. Orr, M. Selby, 'Spectrometric Analysis of Nonmetals Introduced from a Graphite Furnace into a Microwave-induced Plasma', *Talanta*, **33**(11), 875–882 (1986).

131. M. Wu, J.W. Carnahan, 'Trace Determination of Cadmium, Copper, Bromine and Chlorine with Electrothermal Vaporization into a Helium Microwave-induced Plasma', *Appl. Spectrosc.*, **44**(4), 673–678 (1990).
132. K. Dittrich, H. Fuchs, H. Berndt, J.A.C. Broekaert, G. Schaldach, 'Determination of Traces of Sulfur by Electrothermal Evaporation and Non-thermal Excitation of Sulfur-containing Species in a Hollow-cathode Discharge (FANES/MONES) [Furnace Non-thermal Excitation Spectrometry/molecular Non-thermal Excitation Spectrometry] and in a Microwave-induced Plasma (MIP)', *Fresenius' J. Anal. Chem.*, **336**(4), 303–310 (1990).
133. H.P.J. Van Dalen, B.G. Kwee, L. De Galan, 'Selective Determination of Halogens and Sulfur in Solution by Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry Coupled to an Electrothermal Introduction System', *Anal. Chim. Acta*, **142**, 159–171 (1982).
134. J.W. Carnahan, J.A. Caruso, 'Determination of Halogenated Organic Compounds by Electrothermal Vaporization into a Helium Microwave-induced Plasma at Atmospheric Pressure', *Anal. Chim. Acta*, **136**, 261–267 (1982).
135. T. Kumamaru, J.F. Riordan, B.L. Vallee, 'Low-pressure Microwave-induced Helium Plasma Emission Spectrometry: Determination of Sub-nanogram Quantities of Zinc by Use of a Tungsten-filament Vaporization System', *Anal. Biochem.*, **126**(1), 208–213 (1982).
136. N.W. Barnett, 'Determination of Lead and Nickel in Animal Bone by Microwave-induced Plasma Atomic-emission Spectrometry with Sample Introduction by Electrothermal Vaporization', *Anal. Chim. Acta*, **198**, 309–314 (1987).
137. E. Beinrohr, E. Bulska, P. Tschöpel, G. Tölg, 'Determination of Lead by Electrothermal-vaporization Microwave-induced-plasma Atomic-emission Spectrometry After Flow-through Electrolytic Deposition in a Graphite Tube Packed with Reticulated Vitreous Carbon', *J. Anal. At. Spectrom.*, **8**(7), 965–968 (1993).
138. K. Kitagawa, A. Mizutani, M. Yanagisawa, 'Separative Column Atomizer Attached to Atmospheric-pressure Helium Microwave-induced Plasma for Direct Trace-element Analysis by Atomic-emission Spectrometry', *Anal. Sci.*, **5**(5), 539–544 (1989).
139. H. Matusiewicz, I.A. Brovko, R.E. Sturgeon, V.T. Luong, 'Aerosol Transport Interface for Electrothermal Vaporization–Microwave-induced Plasma Emission Spectrometry', *Appl. Spectrosc.*, **44**(4), 736–739 (1990).
140. M. Yanagisawa, H. Kawaguchi, B.L. Vallee, 'Low-pressure Microwave-induced Helium-plasma Emission Spectrometry: Vaporization Characteristerics of Metal Chlorides, Nitrates and Sulfates', *Anal. Biochem.*, **95**(1), 8–13 (1979).
141. P.B. Mason, L. Zhang, J.W. Carnahan, R.E. Winans, 'Helium Microwave-induced Plasma Atomic-emission Detection for Liquid Chromatography Utilizing a Moving Band Interface', *Anal. Chem.*, **65**(19), 2596–2600 (1993).
142. S. Luge, J.A.C. Broekaert, 'Use of Optical Emission Spectrometry with Microwave-induced Plasma (MIP) Discharges in a Surfatron Combined to Different Types of Hydride Generation for the Determination of Arsenic', *Mikrochim. Acta*, **113**(3–6), 277–286 (1994).
143. E. Bulska, P. Tschöpel, J.A.C. Broekaert, G. Tölg, 'Different Sample Introduction Systems for the Simultaneous Determination of Arsenic, Antimony and Selenium by Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **271**(1), 171–181 (1993).
144. H. Tao, A. Miyazaki, 'Determination of Germanium, Arsenic, Antimony, Tin and Mercury at Trace Levels by Continuous Hydride-generation Helium Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Sci.*, **7**(1), 55–59 (1991).
145. A. Montaser, D.W. Golightly, *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*, Verlag Chemie, Weinheim, 1987.
146. G.S. Pyen, R.F. Browner, 'Comparison of Flow Injection and Continuous Sample Introduction for Automatic Simultaneous Determination of Arsenic, Antimony and Selenium by Hydride Generation and Inductively Coupled Plasma Optical Emission Spectrometry', *Appl. Spectrosc.*, **42**(3), 508–512 (1988).
147. J.D. Hwang, H.P. Huxley, J.P. Diomiguardi, W.J. Vaughn, 'Determination of Arsenic in Environmental Samples by Inductively Coupled Plasma Atomic-emission Spectrometry with an In Situ Nebulizer/Hydride Generator', *Appl. Spectrosc.*, **44**(3), 491–496 (1990).
148. H. Tao, A. Miyazaki, K. Bansho, 'Hydride-generation System with a Hydrogen Separation Membrane for Low-power Inductively Coupled Plasma Emission Spectrometry', *Anal. Sci.*, **6**(2), 195–199 (1990).
149. H. Matusiewicz, R.E. Sturgeon, S.S. Bermann, 'In Situ Hydride Generation Preconcentration of Arsenic in a Graphite Furnace with Sample Vaporization into a Microwave-induced Plasma for Emission Spectrometry', *Spectrochim. Acta, Part B*, **45**(1–2), 209–214 (1990).
150. D.T. Heitkemper, J.A. Caruso, 'Continuous Hydride Generation for Simultaneous Multi-element Detection with Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **44**(2), 228–234 (1990).
151. K. Dittrich, R. Mandry, 'Investigations into Improvement of Analytical Application of the Hydride Technique in Atomic-absorption Spectrometry by Matrix Modification and Graphite-furnace Atomization. I. Analytical Results', *Analyst*, **111**(3), 269–275 (1986).
152. E. Bulska, E. Beinrohr, P. Tschöpel, J.A.C. Broekaert, G. Tölg, 'Determination of Arsenic, Antimony and Selenium After Preconcentration onto the Reticulated

Vitreous Carbon with MIP AES Detection', *Chem. Anal. (Warsaw)*, **41**(4), 615–623 (1996).

153. H. Matusiewicz, R.E. Sturgeon, S.S. Berman, 'In Situ Hydride-generation Pre-concentration of Arsenic in a Graphite Furnace with Sample Vaporization into a Microwave-induced Plasma for Emission Spectrometry', *Spectrochim. Acta, Part B*, **45**(1–2), 209–214 (1990).
154. N.W. Barnett, L.S. Chen, G.F. Kirkbright, 'Rapid Determination of Arsenic by Optical Emission Spectroscopy Using a Microwave-induced Plasma Source and a Miniature Hydride-generation Device', *Spectrochim. Acta, Part B*, **39**(9–11), 1141–1147 (1984).
155. N.W. Barnett, 'Further Experience with a Miniature Hydride Generation Device Used in Conjunction with Microwave-induced Plasma Atomic-emission Spectrometry (MIP/AES) for the Determination of Antimony, Arsenic, Lead and Tin', *Spectrochim. Acta, Part B*, **42**(6), 859–864 (1987).
156. K.C. Ng, X. Xu, M.J. Brechmann, 'Direct, Continuous Hydride Generation Coupled with Microwave-induced Plasma Atomic-emission Spectrometry for the Determination of Selenium', *Spectrosc. Lett.*, **22**(9), 1251–1262 (1989).
157. F. Lunzer, R. Pereiro-Garcia, N. Bordel-Garcia, A. Sanz-Medel, 'Continuous Hydride Generation Low-pressure Microwave-induced Plasma Atomic Emission Spectrometry for the Determination of Arsenic, Antimony and Selenium', *J. Anal. At. Spectrom.*, **10**(4), 311–315 (1995).
158. H. Matusiewicz, Z. Kurzawa, 'Electrothermal-vaporization Microwave-induced Plasma Atomic-emission Spectrometric Determination of Arsenic and Selenium in Sulfur Following Teflon (PTFE) Bomb Microwave Acid Dissolution and In Situ Hydride Generation', *Acta Chim. Hung.*, **128**(3), 401–410 (1991).
159. K.C. Ng, W. Shen, 'Direct, Continuous Lead Hydride Generation into a Low-power Helium–Hydrogen Plasma for Atomic-emission Spectrometry', *Spectroscopy*, **2**(1), 50–53 (1987).
160. K.J. Timmins, 'Excitation of Gallium, Indium, Selenium, Tellurium, Arsenic and Antimony Microwave-induced Plasma', *J. Anal. At. Spectrom.*, **2**(2), 251–252 (1987).
161. J.M. Costa-Fernandez, R. Pereiro-Garcia, A. Sanz-Medel, N. Bordel-Garcia, 'Effect of Plasma Pressure on the Determination of Mercury by Microwave-induced Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **10**(9), 649–653 (1995).
162. T. Nakahara, K. Nakanishi, T. Walsa, 'Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry for Determination of Mercury', *Chem. Express*, **1**(3), 149–152 (1986).
163. K. Tanabe, K. Chiba, H. Haraguchi, K. Fuwa, 'Determination of Mercury at the Ultra-trace Level by Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Chem.*, **53**(9), 1450–1453 (1981).
164. T. Nakahara, T. Wasa, 'Feasibility Study of Indirect Determination of Iodine by Helium Microwave-induced Plasma Atomic-emission Spectrometry of Mercury', *Chem. Express*, **4**(8), 495–498 (1989).
165. T. Nakahara, T. Walsa, 'Indirect Determination of Iodine in Seawater and Brine by Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry Using Continuous-flow Cold-vapor Generation of Mercury', *Microchem. J.*, **41**(2), 148–155 (1990).
166. F. Camuna, J.E. Sanchez-Uria, A. Sanz-Medel, 'Continuous-flow and Flow Injection Halogen Generation for Chloride, Bromide and Iodide Determinations by Microwave-induced Plasma Atomic-emission Spectroscopy', *Spectrochim. Acta, Part B*, **48**(9), 1115–1125 (1993).
167. J.F. Alder, Q. Jin, R.D. Snook, 'Determination of Chloride by Microwave Helium-plasma Emission Spectrometry Using a Hydrogen Chloride-generation Technique', *Anal. Chim. Acta*, **123**, 329–333 (1981).
168. J.F. Alder, Q. Jin, R.D. Snook, 'Determination of Traces of Chloride in Solution by Microwave-induced Plasma Emission Spectrometry Using Chlorine Generation', *Anal. Chim. Acta*, **120**, 147–154 (1980).
169. M.D. Calzada, M.C. Quintero, A. Gamero, M. Gallego, 'Chemical Generation of Chlorine, Bromine and Iodine for Sample Introduction into a Surfatron-generated Argon Microwave-induced Plasma' *Anal. Chem.*, **64**(13), 1374–1378 (1992).
170. T. Nakahara, S. Morimoto, T. Wasa, 'Analyte Volatilization Procedure for Continuous-flow Determination of Bromine by Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **7**(2), 211–218 (1992).
171. M.D. Calzada, M.C. Quintero, A. Gamero, J. Cotrino, J.E. Sanchez-Uria, A. Sanz-Medel, 'Determination of Bromide by Low Power Surfatron Microwave-induced Plasma After Bromine Continuous Generation', *Talanta*, **39**(4), 341–347 (1992).
172. M.M. Abdillahi, R.D. Snook, 'Determination of Bromide Using a Helium Microwave-induced Plasma with Bromine Generation and Electrothermal Vaporization for Sample Introduction', *Analyst*, **111**(3), 265–267 (1986).
173. M.M. Abdillahi, W. Tschanen, R.D. Snook, 'Microwave-induced Plasma Emission-spectrometric Determination of Bromide', *Anal. Chim. Acta*, **172**, 139–145 (1985).
174. M.C. Quintero-Ortega, J. Cotrino-Bautista, M. Saez, A. Menendez-Garcia, J.E. Sanchez-Uria, A. Sanz-Medel, 'Determination of Iodide by Low Power Surfatron Microwave-induced Plasma After Iodine

Continuous Generation', *Spectrochim. Acta, Part B*, **47**(1), 79–87 (1992).
175. T. Nakahara, T. Wasa, 'Effect of Acidity and Reductant Concentration on the Indirect Determination of Iodine by Helium Microwave-induced Plasma Atomic-emission Spectrometry', *Chem. Express*, **5**(3), 121–124 (1990).
176. T. Nakahara, S. Yamada, T. Wasa, 'Selective Determination of Iodide and Iodate in Brine Waters by Atmospheric Pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry with Continuous-flow Generation of Volatile Iodine', *Chem. Express*, **6**(1), 5–8 (1991).
177. S.R. Koirtyohann, 'Effect of Nitrogen Impurity on Fluorine and Chlorine Emission from an Atmospheric-pressure Helium Microwave Plasma', *Anal. Chem.*, **55**(2), 374–376 (1983).
178. N.W. Barnett, 'Improvements in the Chemical Generation of Chlorine and Bromine, and their Respective Hydrides, as a Means of Sample Introduction into an Atmospheric-pressure Helium Microwave-induced Plasma', *J. Anal. At. Spectrom.*, **3**(7), 969–972 (1988).
179. J. Alvarado, J.W. Carnahan, 'Direct Detection of Vacuum-ultra-violet Radiation for Non-metal Determinations with a Helium Microwave-induced Plasma', *Appl. Spectrosc.*, **47**(12), 2036–2043 (1993).
180. T. Nakahara, T. Mori, S. Morimoto, H. Ishikawa, 'Continuous-flow Determination of Aqueous Sulfur by Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry with Gas-phase Sample Introduction', *Spectrochim. Acta, Part B*, **50**(4–7), 393–403 (1995).
181. W. Drews, G. Weber, G. Tölg, 'Trace Determination of Nickel by Microwave-induced Plasma Atomic-emission Spectrometry After Pre-concentration of Nickel Tetracarbonyl on Chromosorb®', *Anal. Chim. Acta*, **231**(2), 265–271 (1990).
182. W. Drews, G. Weber, G. Tölg, 'Flow Injection System for the Determination of Nickel by Means of MIPOES [Microwave-induced Plasma Optical Emission Spectrometry] After Conversion to Nickel Tetracarbonyl', *Fresenius' J. Anal. Chem.*, **332**(8), 862–865 (1989).
183. G.K. Webster, J.W. Carnahan, 'Non-metal Analyte Introduction Device for Atomic Spectrometry', *Anal. Chem.*, **61**(1), 790–793 (1989).
184. J.M. Keane, D.C. Brown, R.C. Fry, 'Red and Near-infrared Photodiode Array Atomic-emission Spectrograph for the Simultaneous Determination of Carbon, Hydrogen, Nitrogen and Oxygen', *Anal. Chem.*, **57**(13), 2526–2533 (1985).
185. D.B. Hooker, J. DeZwaan, 'Organic Micro-analysis of Sub-microgram Samples', *J. Res. Natl. Bur. Stand.*, **93**(3), 245–249 (1988).
186. K. Tanabe, K. Matsumoto, H. Haraguchi, K. Fuwa, 'Determination of Ultra-trace Ammonium-, Nitrite- and Nitrate-nitrogen by Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry with Gas-generation Technique', *Anal. Chem.*, **52**(14), 2361–2365 (1980).
187. F.P. Schwarz, 'Characterization of Emission from Atomic Hydrogen in a Microwave-induced Plasma', *Anal. Chem.*, **51**(9), 1508–1512 (1979).
188. J.A.C. Broekaert, P.W.J.M. Boumans, in *Inductively Coupled Plasma Emission Spectroscopy*, ed. P.W.J.M. Boumans, Wiley, New York, 296–357, Chapter 6, Vol. 1, 1987.
189. H. Berndt, 'High-pressure Nebulization: A New Way of Sample Introduction for Atomic Spectroscopy', *Fresenius' Z. Anal. Chem.*, **331**(3–4), 321–323 (1988).
190. G. Heltai, T. Jozsa, K. Percsich, 'A Possibility of Element-specific Detection in HPLC by Means of MIP/AES Coupled with Hydraulic High-pressure Nebulization', *Fresenius' J. Anal. Chem.*, **355**(5–6), 638–641 (1996).
191. V.A. Fassel, B.R. Bear, 'Ultrasonic Nebulization of Liquid Samples for Analytical Inductively Coupled Plasma–Atomic Spectroscopy: An Update', *Spectrochim. Acta, Part B*, **41**(10), 1089–1113 (1986).
192. R.G. Stahl, K.J. Timmins, 'Use of a Glass-frit Nebulizer with a Helium Microwave-induced Plasma', *J. Anal. At. Spectrom.*, **2**(6), 557–559 (1987).
193. H. Matusiewicz, 'Use of the Hildebrand Grid Nebulizer as a Sample-introduction System for Microwave-induced Plasma Spectrometry', *J. Anal. At. Spectrom.*, **8**(7), 961–964 (1993).
194. H. Matusiewicz, R.W. Sturgeon, 'Slurry Sample Introduction with Microwave-induced Plasma Atomic-emission Spectrometry', *Spectrochim. Acta, Part B*, **48**(5), 723–727 (1993).
195. D. Kollotzek, D. Oechsle, G. Kaiser, P. Tschöpel, G. Tölg, 'Application of a Mixed-gas Microwave-induced Plasma as an On-line Element-specific Detector in High-performance Liquid Chromatography', *Fresenius' Z. Anal. Chem.*, **318**(7), 485–489 (1984).
196. G.L. Long, L.D. Perkins, 'Direct Introduction of Aqueous Samples into a Low-powered Microwave-induced Plasma for Atomic-emission Spectrometry', *Appl. Spectrosc.*, **41**(6), 980–985 (1987).
197. P.G. Brown, D.L. Haas, J.M. Workman, J.A. Caruso, F.L. Fricke, 'Moderate-power Argon Microwave-induced Plasma for the Detection of Metal Ions in Aqueous Samples of Complex Matrix', *Anal. Chem.*, **59**(10), 1433–1436 (1987).
198. K.B. Olsen, D.S. Sklarew, J.C. Evans, 'Detection of Organomercury, Selenium and Arsenic Compounds by a Capillary-column Gas Chromatography–Microwave Plasma Detector System', *Spectrochim. Acta, Part B*, **40**(1–2), 357–365 (1985).
199. V.P. Baluda, L.M. Filimonov, 'Effect of Accompanying Components on Spectrographic Analyses with Use of

the PVS-1 Microwave Plasmatron', *Zh. Anal. Khim.*, **35**(6), 1061–1073 (1980).
200. F. Liang, D.X. Zhang, Y.H. Lei, H.Q. Zhang, Q.H. Jin, 'Determination of Selected Noble Metals by MPT/AES Using a Pneumatic Nebulizer', *Microchem. J.*, **52**(2), 181–187 (1995).
201. K.G. Michlewicz, J.W. Carnahan, 'Determination of Aqueous Bromide, Iodide and Chloride with Pneumatic and Ultrasonic Nebulization into a Helium Microwave-induced Plasma', *Anal. Chem.*, **58**(14), 3122–3125 (1986).
202. E. Bulska, J.A.C. Broekaert, P. Tschöpel, G. Tölg, 'Comparative Study of Argon and Helium Plasmas in a TM_{010} Cavity and a Surfatron and Their Use for Hydride-generation Microwave-induced-plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **276**(2), 377–384 (1993).
203. S. Kirschner, A. Golloch, U. Teigheder, 'First Investigations for the Development of a Microwave-induced Plasma Atomic-emission Spectrometry System to Determine Trace Metals in Gases', *J. Anal. At. Spectrom.*, **9**(9), 971–974 (1994).
204. G.K. Webster, J.W. Carnahan, 'Improved Supercritical-fluid Chromatography Mobile-phase Tolerance with a Moderate-power Helium Microwave-induced Plasma', *Appl. Spectrosc.*, **45**(8), 1285–1290 (1991).
205. Q.H. Jin, H.Q. Zhang, W.J. Yang, Q. Jin, Y.H. Shi, 'Determination of Trace Silver, Gold, Germanium, Lead, Tin and Tellurium by Microwave Plasma Torch Atomic Emission Spectrometry Coupled with an Electrothermal Vaporization Sample Introduction System', *Talanta*, **44**(9), 1605–1614 (1997).
206. A.H. Ali, J.D. Winefordner, 'Micro-sample Introduction by Tungsten-filament Electrode into Capacitively Coupled Microwave Plasma for Atomic-emission Spectroscopy: Analytical Figures of Merit', *Anal. Chim. Acta*, **264**(2), 327–332 (1992).
207. A.H. Ali, J.D. Winefordner, 'Micro-sample Introduction by Tungsten-filament Electrode into Capacitively Coupled Microwave Plasma for Atomic-emission Spectroscopy: Diagnostics', *Anal. Chim. Acta*, **262**(2), 319–325 (1992).
208. Q.H. Jin, H.Q. Zhang, Y. Wang, X.L. Yuan, W.J. Yang, 'Study of Analytical Performance of a Low-powered Microwave Plasma Torch in Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **9**(8), 851–856 (1994).
209. Y.N. Pak, S.R. Koirtyohann, 'Fundamental Characterization of the Moderate-power Helium Microwave-induced Plasma', *Spectrochim. Acta, Part B*, **49**(6), 593–606 (1994).
210. Y. Okamoto, M. Yasuda, S. Murayama, 'High-power Microwave-induced Plasma Source for Trace-element Analysis', *Jpn. J. Appl. Phys., Part 2*, **29**(4), L670–L672 (1990).
211. D.L. Haas, J.A. Caruso, 'Characterization of a Moderate-power Microwave-induced Plasma for Direct Solution Nebulization of Metal Ions', *Anal. Chem.*, **56**(12), 2014–2019 (1984).
212. K.C. Ng, R.C. Culp, 'Direct Injection Nebulizer-coupled Microwave-induced Plasma for Atomic-emission Spectroscopy', *Appl. Spectrosc.*, **51**(10), 1447–1452 (1997).
213. M. Wu, Y. Madrid, J.A. Auxier, G.M. Hieftje, 'New Spray Chamber for Use in Flow Injection Plasma Emission Spectrometry', *Anal. Chim. Acta*, **286**(2), 155–167 (1994).
214. H.Q. Zhang, X.L. Yuan, X.J. Zhao, Q.H. Jin, 'On-line Preconcentration with Activated Carbon for Microwave Plasma Torch Atomic-emission Spectrometry', *Talanta*, **44**(9), 1615–1623 (1997).
215. J. Alvarado, M.G. Wu, J.W. Carnahan, 'Electrothermal Vaporization and Ultrasonic Nebulization for the Determination of Aqueous Sulfur Using a Kilowatt-plus Helium Microwave-induced Plasma', *J. Anal. At. Spectrom.*, **7**(8), 1253–1256 (1992).
216. M. Wu, J.W. Carnahan, 'Direct Determination of Aqueous Carbon, Phosphorus and Sulfur Using a Kilowatt-plus Helium Microwave-induced Plasma System with Ultrasonic Nebulization', *J. Anal. At. Spectrom.*, **7**(8), 1249–1252 (1992).
217. Q. Jin, C. Zhu, K. Brushwyler, G.M. Hieftje, 'Efficient and Inexpensive Ultrasonic Nebulizer for Atomic Spectrometry', *Appl. Spectrosc.*, **44**(2), 183–186 (1990).
218. J.M. Gehlhausen, J.W. Carnahan, 'Determination of Aqueous Fluoride with a Helium Microwave-induced Plasma and Flow Injection Analysis', *Anal. Chem.*, **61**(7), 674–677 (1989).
219. C.R. Jia, X.H. Wang, Q.Y. Ou, 'Microwave-induced Plasma Atomic-emission Spectrometer Used as the Detector for Online Coupled High-performance Liquid Chromatography–Gas Chromatography', *Anal. Commun.*, **34**(2), 53–56 (1997).
220. G.L. Long, G.R. Ducatte, E.D. Lancaster, 'Helium Microwave-induced Plasmas for Element Specific Detection in Chromatography', *Spectrochim. Acta, Part B*, **49**(1), 75–87 (1994).
221. J.P.J. van Dalen, P.A. de Lezenne-Coulander, L. de Galan, 'Optimization of the Microwave-induced Plasma as an Element-selective Detector for Non-metals', *Anal. Chim. Acta*, **94**(1), 1–19 (1977).
222. N.P. Vela, L.K. Olson, J.A. Caruso, 'Elemental Speciation with Plasma Mass Spectrometry', *Anal. Chem.*, **65**(13), 585A–597A (1993).
223. R. Lobinski, 'Elemental Speciation and Coupled Techniques', *Appl. Spectrosc.*, **51**(7), 260A–278A (1997).
224. D. Deruaz, J.M. Mermet, 'Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *Analusis*, **14**(3), 107–118 (1986).
225. P.C. Uden, 'Element-selective Chromatographic Detection by Plasma Atomic-emission Spectroscopy', *Trends Anal. Chem.*, **6**(9), 238–246 (1987).

226. P.C. Uden, Y. Yoo, T. Wang, Z. Cheng, 'Element-selective Gas-chromatographic Detection by Atomic Plasma Emission Spectroscopy. Review and Developments', *J. Chromatogr.*, **468**, 319–328 (1989).
227. W. Yu, 'Development and Application of a Microwave-induced Plasma Emission-spectrometric Detector for Gas Chromatography in China', *J. Anal. At. Spectrom.*, **3**(6), 893–900 (1988).
228. P.C. Uden, 'Element-specific Chromatographic Detection by Atomic Absorption, Plasma Atomic Emission and Plasma Mass Spectrometry', *J. Chromatogr. A*, **703**(1–2), 393–416 (1995).
229. J. Hobbs, 'Microwave Plasma Detection [in Gas Chromatography] with the MPD 80', *Eur. Spectrosc. News*, **26**, 40 (1979).
230. J.C. Evans, K.B. Olsen, D.S. Sklarew, 'Low-pressure Beenakker-type Microwave-induced Helium Plasma Source as a Simultaneous Multi-element Gas-chromatographic Detector', *Anal. Chim. Acta*, **194**, 247–260 (1987).
231. K.J. Slatkavitz, P.C. Uden, L.D. Hoey, R.M. Barnes, 'Atmospheric-pressure Microwave-induced Helium Plasma Spectroscopy for Simultaneous Multi-element Gas-chromatographic Detection', *J. Chromatogr.*, **302**, 277–287 (1984).
232. B.D. Quimby, J.J. Sullivan, 'Evaluation of a Microwave Cavity, Discharge Tube and Gas Flow System for Combined Gas Chromatography–Atomic Emission Detection', *Anal. Chem.*, **62**(10), 1027–1034 (1990).
233. T. Bandemer, 'Element-selective GC Detection by AES', *Int. Lab.*, **20**(8), 28–35 (1990).
234. R.L. Firor, 'Multi-element Detection Using GC–Atomic-emission Spectroscopy', *Int. Lab.*, **19**(7), 44–52 (1989).
235. K. Tanabe, H. Haraguchi, K. Fuwa, 'Application of an Atmospheric-pressure Helium Microwave-induced Plasma as an Element-selective Detector for Gas Chromatography', *Spectrochim. Acta, Part B*, **36**(7), 633–639 (1981).
236. C.F. Bauer, D.F.S. Natusch, 'Speciation at Trace Levels by Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Chem.*, **53**(13), 2020–2027 (1981).
237. E. Bulska, 'Microwave-induced Plasma as an Element-specific Detector for Speciation Studies at the Trace Level', *J. Anal. At. Spectrom.*, **7**(2), 201–210 (1992).
238. Y. Takigawa, T. Hanai, J. Hubert, 'Microwave-induced Plasma Emission Spectrophotometer Combined with a Photodiode-array Monitor for Capillary-column Gas Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **9**(11), 698–702 (1986).
239. C.S. Cerbus, S.J. Gluck, 'Design, Optimization and Utilization of a Microwave-induced Helium-plasma Gas-chromatographic Detector in an Industrial Laboratory', *Spectrochim. Acta, Part B*, **38**(1–2), 387–397 (1983).
240. Q.Y. Ou, G.C. Wang, K.W. Zeng, W.L. Yu, 'Investigation of a Microwave-plasma Emission Spectrometer as a Quantitative Detector for Glass-capillary Gas Chromatography', *Spectrochim. Acta, Part B*, **38**(1–2), 419–425 (1983).
241. A.L. Pires-Valente, P.C. Uden, 'Comparison of a Combined Helium–Argon Plasma with Pure Helium Plasmas for Gas Chromatography with Atomic-emission Detection', *Analyst*, **120**(2), 419–421 (1995).
242. G. Chevrier, T. Hanai, K.C. Tran, J. Hubert, 'Development of a Microwave Plasma Detector for Gas Chromatography', *Can. J. Chem.*, **60**(7), 898–903 (1982).
243. K.A. Forbes, E.E. Reszke, P.C. Uden, R.M. Barnes, 'Comparison of Microwave-induced Plasma Sources', *J. Anal. At. Spectrom.*, **6**(1), 57–71 (1991).
244. R.L.A. Sing, J. Hubert, 'Noise Characterization of a 'Surfatron' Microwave-induced Plasma and the Implications for Fourier-transform-based Detection in GC–Microwave-induced Plasma AES', *Appl. Spectrosc.*, **44**(10), 1605–1612 (1990).
245. J.F. Camuna-Aguilar, R. Pereiro-Garcia, J.E. Sanchez-Uria, A. Sanz-Medel, 'Comparative Study of Three Microwave-induced Plasma Sources for Atomic-emission Spectrometry. II. Evaluation of Their Atomization/Excitation Capabilities for Chlorinated Hydrocarbons', *Spectrochim. Acta, Part B*, **49**(6), 545–554 (1994).
246. R.M. Alvarez-Bolainez, M.P. Dziewatkoski, C.B. Boss, 'Sensitivity Comparison in a Microwave-induced Plasma Gas-chromatographic Detector: Effect of Plasma Torch Design', *Anal. Chem.*, **64**(5), 541–544 (1992).
247. M.L. Bruce, J.A. Caruso, 'Laminar-flow Torch for Gas-chromatographic Helium Microwave Plasma Detection of Pyrethroids and Dioxins', *Appl. Spectrosc.*, **39**(6), 942–949 (1985).
248. G.W. Jansen, F.A. Huf, H.J. De Jong, 'Low-power Microwave Plasma Source for Chromatography Detection', *Spectrochim. Acta, Part B*, **40**(1–2), 307–316 (1985).
249. F.A. Huf, G.W. Jansen, 'Microwave-induced Plasma Between Electrodes as a Detector for Gas and Liquid Chromatography', *Spectrochim. Acta, Part B*, **38**(7), 1061–1064 (1983).
250. K. Cammann, L. Lendero, H. Feuerbacher, K. Ballschmiter, 'Power-modulated Microwave-induced Plasma with Enhanced Sensitivity and Practicability as an Element-specific GC Detector', *Fresenius' Z. Anal. Chem.*, **316**(2), 194–200 (1983).
251. D.L. Haas, J.A. Caruso, 'Moderate-power Helium Plasma as an Element-selective Detector for Gas Chromatography of Dioxins and Other Halogenated Compounds', *Anal. Chem.*, **57**(3), 846–851 (1985).

252. R.M. Alvarez-Bolainez, C.B. Boss, 'Microwave-induced-plasma Reflected-power Detector for Gas Chromatography', *Anal. Chem.*, **63**(2), 159–163 (1991).
253. J.E. Freeman, G.M. Hieftje, 'Near-infrared Nonmetal Atomic Emission from a Helium Microwave-induced Plasma: Element Ratio Determinations', *Spectrochim. Acta, Part B*, **40**(4), 653–664 (1985).
254. D.E. Pivonka, W.G. Fateley, R.C. Fry, 'Simultaneous Determination of Carbon, Hydrogen, Nitrogen, Oxygen, Fluorine, Chlorine, Bromine and Sulphur in Gas-chromatographic Effluent by Fourier-transform Red–Near-infrared Atomic-emission Spectroscopy', *Appl. Spectrosc.*, **40**(3), 291–297 (1986).
255. C. Lauzon, K.C. Tran, J. Hubert, 'Multi-wavelength Detection in Gas Chromatography with Microwave-induced Plasma Atomic-emission Fourier Transform Spectrometry', *J. Anal. At. Spectrom.*, **3**(6), 901–905 (1988).
256. D. Rieping, J. Bettmer, W. Buscher, K. Cammann, 'The Development of a Simple Gas Chromatography Detector Based on Atomic-emission Spectrophotometry', *GIT Fachz. Lab.*, **39**(2), 95–98 (1995).
257. S.R. Koirtyohann, 'Effect of Nitrogen Impurity on Fluorine and Chlorine Emission from an Atmospheric-pressure Helium Microwave Plasma', *Anal. Chem.*, **55**(2), 374–376 (1983).
258. V.O. Schmitt, I. Rodriguez-Pereiro, R. Lobinski, 'Flash Species-selective Analysis of Multicapillary Gas Chromatography with Microwave-induced-plasma Atomic-spectrometric Detection', *Anal. Commun.*, **34**(5), 141–143 (1997).
259. T.N. Asp, S. Pedersen Bjergaard, T. Greibrokk, 'Calculation of Elemental Ratios by On-column Radio-frequency Plasma Atomic-emission Detection Coupled with Capillary Gas Chromatography', *J. Chromatogr. A*, **736**(1–2), 157–164 (1996).
260. M.L. Mena, C.W. McLeod, 'Mercury Species Immobilized on Sulphydryl Cotton: A New Candidate Reference Material for Mercury Speciation', *Mikrochim. Acta*, **123**(1–4), 103–108 (1996).
261. P. Lansens, W. Baeyens, M. Termonia, 'Modification of the Perkin-Elmer HS-6 Headspace Sampler for the Determination of Organomercury Compounds', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **12**(3), 132–133 (1989).
262. M.E. Birch, 'Solvent-venting Technique for Gas Chromatography with Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **282**(2), 451–458 (1993).
263. S.W. Jordan, B.L. Karger, I.S. Krull, S.B. Smith Jr, 'All-glass-lined, Open-split, Solvent-venting Interface for Gas Chromatography–Microwave-induced Plasma Emission Spectroscopy (GC/MIP)', *Chem. Biomed. Environ. Instrum.*, **12**(4), 263–274 (1983).
264. L. Zhang, J.W. Carnahan, R.E. Winans, P.H. Neill, 'Solvent-venting Interface for Capillary Gas Chromatography and a Microwave-induced Plasma', *Anal. Chim. Acta*, **233**(1), 149–154 (1990).
265. W.E. Wentworth, T. Limero, C.F. Batten, E.C.M. Chen, 'Non-radioactive Electron-capture Detector for Gas Chromatography', *J. Chromatgr.*, **468**(5), 215–224 (1989).
266. R. Lobinski, F.C. Adams, 'Application of Capillary Gas Chromatography with Atomic Emission Detection (CGC–AED) in Ultratrace Speciation Analysis of Organometals and Metalloids', *ICP Inf. Newsl.*, **19**(3), 164–166 (1993).
267. A.L.P. Valente, P.C. Uden, 'Memory Effects in Gas-chromatographic Detection with a Microwave-induced Plasma', *J. High Resolut. Chromatogr.*, **16**(5), 275–278 (1993).
268. M. Caetano, R.E. Golding, E.A. Key, 'Factorial Analysis and Response Surface of a Gas Chromatography Microwave-induced Plasma System for the Determination of Halogenated Compounds', *J. Anal. At. Spectrom.*, **7**(6), 1007–1011 (1992).
269. H. Yieru, O. Qingyu, Y. Weile, 'Effect of Compound Structure on the Elemental Responses in Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *J. Chromatogr. Sci.*, **28**(11), 584–588 (1990).
270. S.R. Goode, L.K. Kimbrough, 'Experimental Study of the Signal-to-noise Ratio in the Microwave-induced Plasma Gas-chromatographic Detector', *Spectrochim. Acta, Part B*, **42**(1–2), 309–322 (1987).
271. K. Cammann, H. Mueller, 'Reduced Calibration Efforts in Gas Chromatography by the Use of an Element-selective Plasma-emission Detector', *Fresenius' Z. Anal. Chem.*, **331**(3–4), 336–341 (1988).
272. D.F. Hagen, J.S. Marhevka, L.C. Haddad, 'Multi-element Taggant and Dual Homolog Derivatization Concepts in G.C.–MED [Microwave-sustained Helium Emission Detector]', *Spectrochim. Acta, Part B*, **40**(1–2), 335–347 (1985).
273. J.H. Brill, B.A. Narayanan, J.P. McCormick, 'Selective Determination of Pentafluorobenzyl Ester Derivatives of Carboxylic Acids by GC Using Microwave Plasma and Mass Selective Detection', *Appl. Spectrosc.*, **45**(10), 1617–1620 (1991).
274. D. Deruaz, J.L. Brazier, 'Applications of Gas Chromatography–Helium Microwave Plasma Detection', *Spectra 2000*, **161**, 8–13 (1991).
275. C. Webster, M. Cooke, 'Use of Microwave-induced Plasma Atomic-emission Detection for the Quantification of Oxygen Containing Compounds', *Anal. Proc.*, **31**(8), 237–240 (1994).
276. K. Zeng, Q. Ou, G. Wang, W. Yu, 'Analysis of Oxygenated Compounds by Gas Chromatography–Microwave

Plasma Emission Spectrometry', *Spectrochim. Acta, Part B*, **40**(1–2), 349–356 (1985).
277. S.R. Goode, C.L. Thomas, 'Characterizing the Factors that Influence Oxygen Selectivity in Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **9**(9), 965–970 (1994).
278. S.R. Goode, L.K. Kimbrough, 'Study of the Factors Influencing the Detection of Oxygen [in Organic Compounds] with Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **3**(6), 915–918 (1988).
279. K.J. Slatkavitz, P.C. Uden, R.M. Barnes, 'Consideration of an Atmospheric-pressure Microwave-induced Helium Plasma as an Oxygen-selective Gas-chromatographic Detector', *J. Chromatogr.*, **355**(1), 117–126 (1986).
280. P. Stilkenboehmer, K. Cammann, 'Plasma Emission GC Detector for Selective Determination of Hydrogen-containing Organic Compounds', *Fresenius' Z. Anal. Chem.*, **335**(7), 764–768 (1989).
281. D.F. Hagen, L.C. Haddad, J.S. Marhevka, 'Solution and Gas-phase Deuterium Derivatization Reactions for Gas Chromatography–Microwave Emission Detection', *Spectrochim. Acta, Part B*, **42**(1–2), 253–267 (1987).
282. S.R. Goode, J.J. Gemmill, B.E. Watt, 'Determination of Deuterium by Gas Chromatography with a Microwave-induced Plasma Emission Detector', *J. Anal. At. Spectrom.*, **5**(6), 483–486 (1990).
283. A. Bannier, D. Deruz, C. Weber, J.L. Brazier, 'Analysis of Caffeine Deutero-isotopomers by Gas Chromatography and Atomic-emission Detection', *Anal. Lett.*, **25**(6), 1073–1085 (1992).
284. F. Leclerc, D. Deruaz, A. Bannier, J.L. Brazier, 'Limit of Detection of Carbon-13 Using an Atomic-emission Detector (MIP) Coupled to Gas Chromatography', *Anal. Lett.*, **27**(7), 1325–1338 (1994).
285. J.T. Jelink, A. Venema, 'Investigations into the Use of Capillary GC with Atomic-emission Detection', *J. High Resolut. Chromatogr.*, **13**(6), 447–450 (1990).
286. H.A. Dingjan, H.J. De Jong, 'Determination of Ratio Formulae for Organic Compounds Using a Microwave-induced Plasma', *Spectrochim. Acta, Part B*, **38**(5–6), 777–781 (1983).
287. H.J. Perpall, P.C. Uden, R.L. Deming, 'Empirical and Molecular Formula Determination of Polymer Pyrolysates by Multi-referencing GC–Microwave-induced Plasma Spectroscopy', *Spectrochim. Acta, Part B*, **42**(1–2), 243–251 (1987).
288. L.G. Sarto Jr, S.A. Estes, P.C. Uden, S. Siggia, R.M. Barnes, 'Fused-silica Capillary GC with Boron-specific Microwave-excited Plasma Detection in Derivatization and Polymer Pyrolysis', *Anal. Lett.*, **14**(3), 205–218 (1981).
289. J.J. Sullivan, B.D. Quimby, 'Detection of Carbon, Hydrogen, Nitrogen and Oxygen by Atomic-emission in Capillary Gas Chromatography', *Chem. Labor. Betr.*, **41**(4), 200–207 (1990).
290. H. Mueller, K. Cammann, 'Development of a Plasma Emission Gas-chromatographic Detector for the Simultaneous Selective Detection of Chlorine, Bromine and Carbon', *J. Anal. At. Spectrom.*, **3**(6), 907–913 (1988).
291. P.C. Uden, K.J. Slatkavitz, R.M. Barnes, R.L. Deming, 'Empirical and Molecular Formula Determination by Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **180**, 401–416 (1986).
292. N. Kovacic, T.L. Ramus, 'Application of a Microwave-induced Plasma Atomic-emission Detector for Quantification of Halogenated Compounds by Gas Chromatography', *J. Anal. At. Spectrom.*, **7**(6), 999–1005 (1992).
293. J.H. Brill, B.A. Narayanan, J.P. Doom, J.P. McCormick, 'Selective Determination of Organofluorine Compounds by Capillary Column Gas Chromatography with an Atmospheric Pressure Helium Microwave-injected Plasma Detector', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**(5), 368–374 (1988).
294. H. Schimmel, B. Schmidt, R. Bacher, K. Ballschmiter, 'Molar Response of Polychlorinated Dibenzo-*p*-dioxins and Dibenzofurans by the Mass-spectrometric Detector', *Anal. Chem.*, **65**(5), 640–644 (1993).
295. C. Webster, M.J. Cooke, 'Use of an Atomic-emission Detector to Study the Variation in Elemental Response for Chlorine, Carbon and Oxygen in Phenols', *J. High Resolut. Chromatogr.*, **18**(5), 319–322 (1995).
296. Y. Huang, Q. Ou, W. Yu, 'Study of gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry. I. Effect of the Structure of a Compound on the Determination of its Empirical Formula', *J. Anal. At. Spectrom.*, **5**(2), 115–120 (1990).
297. M.M. Abdillahi, 'Gas Chromatography–Microwave-induced Plasma for the Determination of Halogenated Hydrocarbons', *J. Chromatogr. Sci.*, **28**(12), 613–616 (1990).
298. A. Sarafraz-Yazdi, M.Y. Khuhawar, P.C. Uden, 'Gas Chromatography of Chloride and Bromide as Phenylboronic Acid–Mercuric Nitrate Derivatives with Microwave-induced Plasma Atomic-emission Detection', *J. Chromatogr.*, **594**(1–2), 395–399 (1992).
299. S.E. Eckert-Tilotta, S.B. Hawthorne, D.J. Miller, 'Comparison of Commercially Available Atomic-emission and Chemiluminescence Detectors for Sulphur-selective Gas Chromatographic Detection', *J. Chromatogr.*, **591**(1–2), 313–323 (1992).
300. B. Schmidt, J.T. Anderson, K. Ballschmiter, 'Sulfur-selective Detection in Gas Chromatography with a Microwave-induced Plasma Detector', *Fresenius' Z. Anal. Chem.*, **333**(7), 765–766 (1989).
301. J.A. Seeley, P.C. Uden, 'Element-selective Gas-chromatographic Detection and Determination of Organoboron Compounds (in Motor and Lubricating Oils)', *Analyst*, **116**(2), 1321–1326 (1991).

302. K.J. Slatkavitz, L.D. Hoey, P.C. Uden, R.M. Barnes, 'Element-specific Detection of Organosilicon Compounds by Gas Chromatography–Atmospheric-pressure Microwave-induced Helium Plasma Spectrometry', *Anal. Chem.*, **57**(9), 1846–1853 (1985).
303. G.B. Limentani, P.C. Uden, 'High-resolution Gas and Liquid Chromatography of Organoarsenic Compounds', *J. Chromatogr.*, **325**(1), 53–60 (1985).
304. H. Haraguchi, A. Takatsu, 'Derivative Gas Chromatography of Methylarsenic Compounds with Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometric Detection', *Spectrochim. Acta, Part B*, **42**(1–2), 235–241 (1987).
305. C. Bradley, J.W. Carnahan, 'Oxygen-selective Microwave-induced Plasma Gas Chromatography Detector for Petroleum-related Samples', *Anal. Chem.*, **60**(9), 858–863 (1988).
306. S.R. Goods, C.L. Thomas, 'Determination of Oxygen-containing Additives in Gasoline by Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **9**(2), 73–78 (1994).
307. M.P. Italia, P.C. Uden, 'Multiple-element Emission-spectral-detection Gas-chromatographic Profiles of Halogenated Products from Chlorination of Humic Acid and Drinking Water', *J. Chromatogr.*, **438**(1), 35–43 (1988).
308. B. Riviere, J.M. Mermet, D. Deruaz, 'Study of a Microwave-induced Plasma (Surfatron) as a Detector in Capillary-column Gas Chromatography with Reference to Pesticides', *J. Anal. At. Spectrom.*, **2**(7), 705–709 (1987).
309. A.H. Mohamad, M. Zerezghi, J.A. Caruso, 'Determination of Polychlorinated Dibenzo-*p*-dioxins Using Capillary Gas Chromatography with Microwave-induced Plasma Detection', *Anal. Chem.*, **58**(2), 469–471 (1986).
310. M.A. George, J.P. Hessler, J.W. Carnahan, 'Determination of Impurities in Argon by Gas Chromatography with a Microwave-induced Plasma Detector', *J. Anal. At. Spectrom.*, **4**(1), 51–54 (1989).
311. D.S. Ballantine Jr, W.H. Zoller, 'Collection and Determination of Volatile Organic Mercury Compounds in the Atmosphere by Gas Chromatography with Microwave Plasma Detection', *Anal. Chem.*, **56**(8), 1288–1293 (1984).
312. S.A. Estes, P.C. Uden, M.D. Rausch, R.M. Barnes, 'Fused-silica Capillary GC Separation and Element-selective Microwave Plasma-emission Detection of Volatile Organometallics', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **3**(9), 471–472 (1980).
313. R. Lobinski, F.C. Adams, 'Recent Advances in Speciation Analysis by Capillary Gas Chromatography–Microwave-induced-plasma Atomic-emission Spectrometry', *Trends Anal. Chem.*, **12**(2), 41–49 (1993).
314. R. Lobinski, W.M.R. Dirkx, M. Ceulemans, F.C. Adams, 'Optimization of Comprehensive Speciation of Organotin Compounds in Environmental Samples by Capillary Gas Chromatography–Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Chem.*, **64**(2), 159–165 (1992).
315. S. Tutschku, S. Mothes, K. Dittrich, 'Determination and Speciation of Organotin Compounds by Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *J. Chromatogr. A.*, **683**(1), 269–276 (1994).
316. S.A. Estes, P.C. Uden, R.M. Barnes, 'Determination of *n*-butylated Trialkyl-lead Compounds by Gas Chromatography with Microwave Plasma-emission Detection', *Anal. Chem.*, **54**(14), 2402–2405 (1982).
317. S.A. Estes, P.C. Uden, R.M. Barnes, 'High-resolution Gas Chromatography of Trialkyl-lead Chlorides with an Inert-solvent-venting Interface for Microwave-excited Helium-plasma Detection', *Anal. Chem.*, **53**(9), 1336–1340 (1981).
318. G.M. Greenway, N.W. Barnett, 'Optimization of an Atmospheric-pressure Helium Microwave-induced Plasma Coupled with Capillary Gas Chromatography for the Determination of Alkyl-lead and Alkylmercury Compounds', *J. Anal. At. Spectrom.*, **4**(8), 783–787 (1989).
319. G. Decadt, W. Baeyens, D. Bradley, L. Goeyens, 'Determination of Methylmercury in Biological Samples by Semi-automated Headspace Analysis', *Anal. Chem.*, **57**(14), 2788–2791 (1985).
320. P. Lansens, C. Meuleman, M. Leermakers, W. Baeyens, 'Determination of Methylmercury in Natural Waters by Headspace Gas Chromatography with Microwave-induced Plasma Detection After Pre-concentration on a Resin Containing Dithiocarbamate Groups', *Anal. Chim. Acta*, **234**(2), 417–424 (1990).
321. P. Lansens, W. Baeyens, 'Improvement of the Semi-automated Headspace Analysis Method for the Determination of Methylmercury in Biological Samples', *Anal. Chim. Acta*, **228**(1), 93–99 (1990).
322. E. Bulska, D.C. Baxter, W. Frech, 'Capillary Column Gas Chromatography for Mercury Speciation (in Environmental and Biological Materials)', *Anal. Chim. Acta*, **249**, 545–554 (1991).
323. P.C. Uden, 'Element Specific Chromatographic Detection for Trace Inorganic Analysis', *Anal. Proc.*, **30**(10), 405–408 (1993).
324. H. Tao, A. Miyazaki, K. Bansho, 'Determination of Trace Beryllium in Natural Waters by Gas Chromatography–Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Sci.*, **4**(3), 299–302 (1988).
325. M.Y. Khuhawar, A. Sarafraz-Yazdi, P.C. Uden, 'Capillary Gas-chromatographic Determination of Copper and Nickel Using Microwave-induced Plasma Atomic-emission Detection.', *J. Chromatogr.*, **636**(2), 271–276 (1993).
326. M.Y. Khuhawar, A. Sarafraz-Yazdi, Y. Zeng, P.C. Uden, 'Capillary Gas-chromatographic Determination of Palladium Chelates Using Microwave-induced Plasma

Atomic-emission Detection', *Chem. Anal. (Warsaw)*, **40**(3), 271–279 (1995).
327. M.B. De la Calle Guntinas, R. Lobinski, F.C. Adams, 'Interference-free Determination of Selenium(IV) by Capillary Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry After Volatilization with Sodium Tetraethylborate', *J. Anal. At. Spectrom.*, **10**(2), 111–115 (1995).
328. B.D. Quimby, P.C. Dryden, J.J. Sullivan, 'Selective Detection of Volatile Nickel, Vanadium and Iron Porphyrins in Crude Oils by Gas Chromatography–Atomic-emission Spectroscopy', *J. High Resolut. Chromatogr.*, **14**(2), 110–116 (1991).
329. H.A.H. Billiet, J.P.J. Van Dalen, P.J. Schoenmakers, L. De Galan, 'Measurement of (Column Volume from) Deuterium Oxide Elution Data in Reversed-phase Liquid Chromatography with Microwave-induced-plasma Detection', *Anal. Chem.*, **55**(6), 847–851 (1983).
330. K.G. Michlewicz, J.W. Carnahan, 'Directly Coupled Microwave-induced Plasma Atomic-emission Liquid Chromatography Detector for Non-metals: Preliminary Characterization with Halides and Oxyhalogen Salts', *Anal. Lett.*, **20**(8), 1193–1205 (1987).
331. L. Zhang, J.W. Carnahan, R.E. Winans, P.H. Neill, 'Moving-wheel Liquid Chromatography–Helium Microwave-induced Plasma Interface', *Anal. Chem.*, **61**(8), 895–897 (1989).
332. J. Costa-Fernandez, F. Lunzer, R. Pereiro-Garcia, A. Sanz-Medel, N. Bordel-Garcia, 'Direct Coupling of High-performance Liquid Chromatography to Microwave-induced Plasma Atomic-emission Spectrometry Via Volatile-species Generation and its Application to Mercury and Arsenic Speciation', *J. Anal. At. Spectrom.*, **10**(11), 1019–1025 (1995).
333. M. Ibrahim, W. Nisamaneepong, J. Caruso, 'Micro-column High-pressure Liquid Chromatography with a Glass-frit Nebulizer Interface for Plasma Emission Detection', *J. Chromatogr. Sci.*, **23**(4), 144–150 (1985).
334. Y. Liu, V. Lopez-Avila, 'Online Microwave-induced Helium Plasma Atomic-emission Detection for Capillary Zone Electrophoresis', *J. High Resolut. Chromatogr.*, **16**(12), 717–720 (1993).
335. L.J. Galante, D.A. Wilson, G.M. Hieftje, 'Detection of Ions by Replacement-ion Chromatography Coupled to a Microwave-induced Nitrogen Discharge at Atmospheric Pressure', *Anal. Chim. Acta*, **215**(1–2), 99–109 (1988).
336. J.M. Carey, J.A. Caruso, 'Plasma-spectrometric Detection for Supercritical-fluid Chromatography', *Trends Anal. Chem.*, **11**(8), 287–293 (1992).
337. C.B. Motley, G.L. Long, 'Examination of a Helium Highly Efficient Microwave-induced Plasma as an Element-selective Detector for Supercritical-fluid Chromatography', *J. Anal. At. Spectrom.*, **5**(6), 477–482 (1990).
338. M. Montes-Bayon, F. Camuna-Aguilar, R. Pereiro, J.E. Sanchez-Uria, A. Sanz-Medel, 'The Effect of Two Gases Forming Supercritical Fluids (Xenon and Carbon Dioxide) on the Spectral Characteristics and Analytical Capabilities of Microwave-induced Plasmas', *Spectrochim. Acta, Part B*, **51**(7), 685–695 (1996).
339. L.J. Galante, M. Selby, D.R. Luffer, G.M. Hieftje, M. Novotny, 'Characterization of Microwave-induced Plasma as a Detector for Supercritical-fluid Chromatography', *Anal. Chem.*, **60**(14), 1370–1376 (1988).
340. G.R. Ducatte, G.L. Long, 'Effect of Carbon Dioxide and Hydrogen on Non-metal Emission Intensities in a Helium Microwave-induced Plasma', *Appl. Spectrosc.*, **48**(4), 493–501 (1994).
341. Y. Wang, J.W. Carnahan, 'Binary Mobile Phases for Supercritical-fluid Chromatography with Helium Microwave-induced Plasma Detection', *Anal. Chem.*, **65**(22), 3290–3294 (1993).
342. G.K. Webster, J.W. Carnahan, 'Atomic-emission Detection for Supercritical-fluid Chromatography Using a Moderate-power Helium Microwave-induced Plasma', *Anal. Chem.*, **64**(1), 50–55 (1992).
343. Q. Jin, F. Wang, C. Zhu, D.M. Chambers, G.M. Hieftje, 'Atomic-emission Detector for Gas Chromatography and Supercritical-fluid Chromatography', *J. Anal. At. Spectrom.*, **5**(6), 487–494 (1990).
344. D.R. Luffer, L.J. Galante, P.A. David, M. Novotny, G.M. Hieftje, 'Evaluation of a Supercritical-fluid Chromatograph Coupled to a Surface-wave-sustained Microwave-induced Plasma Detector', *Anal. Chem.*, **60**(14), 1365–1369 (1988).
345. D.R. Luffer, M. Novotny, 'Element-selective Detection After Supercritical-fluid Chromatography by Means of a Surfatron Plasma in the Near-infrared Spectral Region', *J. Chromatogr.*, **517**, 477–489 (1990).
346. L. Zhang, J.W. Carnahan, R.E. Winans, P.H. Neill, 'Supercritical-fluid Chromatography with a Helium Microwave-induced Plasma for Chlorine-selective Detection', *Anal. Chem.*, **63**(3), 212–216 (1991).
347. G.K. Webster, J.W. Carnahan, 'Simulated Supercritical-fluid Chromatography Mobile Phase Introduction to a Helium Microwave-induced Plasma: Effects on Non-metal Emission and Spectral Behaviour', *Appl. Spectrosc.*, **44**(6), 1020–1027 (1990).
348. C.B. Motley, M. Ashraf-Khorassani, G.L. Long, 'Microwave-induced Plasma as an Elemental Detector for Packed-column Supercritical-fluid Chromatography', *Appl. Spectrosc.*, **43**(5), 737–741 (1989).
349. C.B. Motley, G.L. Long, 'Evaluation of Sample Introduction Techniques of Packed-column SFC into an MIP [Microwave-induced Plasma]', *Appl. Spectrosc.*, **44**(4), 667–672 (1990).
350. F. Paschen, 'Bohrs Heliumlinien', *Ann. Phys.*, **50**(14), 901–940 (1916).
351. R.K. Marcus, *Glow Discharge Spectroscopies*, Plenum Press, New York, 1993.

352. J.A.C. Broekaert, 'State of the Art of Glow Discharges Lamp Spectrometry–Plenary Lecture', *J. Anal. At. Spectrom.*, **2**(6), 537–542 (1987).
353. E.B.M. Steers, F. Leis, 'Observations on the Use of the Microwave-boosted Glow-discharge Lamp and the Relevant Excitation Processes', *J. Anal. At. Spectrom.*, **4**(2), 199–204 (1989).
354. M.R. Winchester, C. Lazik, R.K. Marcus, 'Characterization of a Radio Frequency Glow Discharge Emission Source', *Spectrochim. Acta, Part B*, **46**(4), 483–499 (1991).
355. A. Bengtson, 'Quantitative Depth-profile Analysis by Glow Discharge', *Spectrochim. Acta, Part B*, **49**(4), 411–429 (1994).
356. F. Leis, J.A.C. Broekaert, K. Laqua, 'Design and Properties of a Microwave Boosted Glow-discharge Lamp', *Spectrochim. Acta, Part B*, **42**(11–12), 1169–1176 (1987).
357. E.B.M. Steers, A.P. Thorne, 'Application of High-resolution Fourier-transform Spectrometry to the Study of Glow-discharge Sources. I. Excitation of Iron and Chromium Spectra in a Microwave-boosted Glow-discharge Source', *J. Anal. At. Spectrom.*, **8**(2), 309–315 (1993).
358. F. Leis, E.B.M. Steers, 'Boosted Glow Discharges for Atomic Spectroscopy–Analytical and Fundamental Properties', *Spectrochim. Acta, Part B*, **49**(3), 289–325 (1994).
359. F. Leis, J.A.C. Broekaert, 'Optical-emission Spectroscopic Measurements on Aluminium, Copper and Lead Samples with a Microwave-boosted Glow-discharge Lamp', *Spectrochim. Acta, Part B*, **46**(2), 243–251 (1991).
360. Y.X. Duan, Y.M. Li, Z.H. Du, Q.H. Jin, J.A. Olivares, 'Instrumentation and Fundamental Studies on Glow-discharge–Microwave-induced Plasma (GD/MIP) Tandem Source for Optical Emission Spectrometry', *Appl. Spectrosc.*, **50**(8), 977–984 (1996).
361. Y.X. Su, P.Y. Yang, D.Y. Chen, Z.G. Zhang, Z. Zhou, X.R. Wang, B.L. Huang, 'Microwave-induced-plasma-boosted Microsecond-pulse Glow-discharge Optical Emission Spectrometry', *J. Anal. At. Spectrom.*, **12**(8), 817–822 (1997).
362. S. Caroli, O. Senofonte, N. Violante, L. Di Simone, 'Novel Version of the Microwave-coupled Hollow-cathode Lamp for Atomic-emission Spectrometry', *Appl. Spectrosc.*, **41**(4), 579–583 (1987).
363. J.A.C. Broekaert, 'Atomic Spectroscopy', in *Ullmann's Encyclopedia of Industrial Chemistry*, eds. B. Elvers, S. Hawkins, W. Russey (Volume ed. H. Günzler), Verlag Chemie, Weinheim, 559–652, Vol. B5, 1994.
364. J.J. Sullivan, B.D. Quimby, 'Detection of Carbon, Hydrogen, Nitrogen and Oxygen by Atomic-emission in Capillary Gas Chromatography', *Chem. Labor. Betr.*, **41**(4), 200–207 (1990).
365. L.M. Faires, 'Fourier Transforms for Analytical Atomic Spectroscopy', *Anal. Chem.*, **58**(9), 1023A–1032A (1986).
366. T. Nakahara, S. Yamada, T. Wasa, 'Simple Nitrogen-purged Optical System for Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry in the Vacuum Ultra-violet Spectral Region', *Bunko Kenkyu*, **38**(3), 205–208 (1989).
367. T. Nakahara, S. Yamada, T. Wasa, 'Continuous-flow Determination of Trace Iodine by Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry Using Generation of Volatile Iodine from Iodide', *Appl. Spectrosc.*, **44**(10), 1673–1678 (1990).
368. J.F. Camuna-Aguilar, R. Pereiro-Garcia, J.E. Sanchez-Uria, A. Sanz-Medel, 'Comparative Study of Three Microwave-induced Plasma Sources for Atomic-emission Spectrometry. I. Excitation of Mercury and its Determination After On-line Continuous Cold Vapor Generation', *Spectrochim. Acta, Part B*, **49**(5), 475–484 (1994).
369. E. Denkhaus, A. Golloch, H.M. Kuss, 'Determination of Gaseous Nitrogen in Gas Mixtures Using Low Pressure Microwave-induced Plasma Emission Spectrometry', *Fresenius' J. Anal. Chem.*, **353**(2), 156–161 (1995).
370. K. Dittrich, H. Fuchs, J.M. Mermet, B. Riviere, 'Determination of Nitrogen and Oxygen and Species Containing Nitrogen by Molecular Non-thermal Excitation Spectrometry (MONES) Using Microwave-induced Plasma (MIP) and Furnace Atomization Non-thermal Excitation Spectrometry (FANES) Sources', *J. Anal. At. Spectrom.*, **6**(4), 313–316 (1991).
371. A. Besner, J. Hubert, 'Effects of Dopants on Tin Emission in a Helium Microwave-induced Plasma', *J. Anal. At. Spectrom.*, **3**(2), 381–385 (1988).
372. D.F. Hagen, L.C. Haddad, J.S. Marhevka, 'Solution and Gas-phase Deuterium Derivatization Reactions for Gas Chromatography–Microwave Emission Detection', *Spectrochim. Acta, Part B*, **42**(1–2), 253–267 (1987).
373. B.M. Patel, E. Heithmar, J.D. Winefordner, 'Tubular Electrode Torch for Capacitatively Coupled Helium Microwave Plasma as a Spectrochemical Excitation Source', *Anal. Chem.*, **59**(19), 2374–2377 (1987).
374. J.D. Hwang, W. Masamba, B.W. Smith, J.D. Winefordner, 'Development of High-power, Capacitatively Coupled Microwave Helium Plasma for Atomic-emission Spectroscopy (CMP/AES)', *Can. J. Spectrosc.*, **33**(6), 156–160 (1988).
375. Y.K. Zhang, S. Hanamura, J.D. Winefordner, 'Evaluation of a Microwave-induced Air Plasma as an Excitation Source', *Appl. Spectrosc.*, **39**(2), 226–230 (1985).
376. W.R.L. Masamba, J.D. Winefordner, 'Analytical Characteristics of a Helium–Hydrogen Capacitively Coupled Microwave Plasma', *Spectrochim. Acta, Part B*, **48**(4), 521–529 (1993).
377. B.M. Spencer, A.R. Raghani, J.D. Winefordner, 'Investigation of Halogen Determination in a Helium Capacitively-coupled Microwave Plasma Atomic-emission Spectrometer', *Appl. Spectrosc.*, **48**(5), 643–646 (1994).

378. B.M. Spencer, J.D. Winefordner, 'Determination of Silicon in Organic Samples by Atomic-emission Spectrometry by Using a Capacitively-coupled Microwave Plasma', *Can. J. Spectrosc.*, **39**(2), 43–53 (1994).
379. H. Uchida, P.A. Johnson, J.D. Winefordner, 'Evaluation of the Capacitatively Coupled Helium Microwave Plasma as an Excitation Source for the Determination of Inorganic and Organic Tin', *J. Anal. At. Spectrom.*, **5**(1), 81–85 (1990).
380. D.M. Hueber, W.R.L. Masamba, B.M. Spencer, J.D. Winefordner, 'Application of Hydride Generation to the Determination of Trace Concentrations of Arsenic by Capacitively Coupled Microwave Plasma', *Anal. Chim. Acta*, **278**(2), 279–285 (1993).
381. H. Uchida, A. Berthod, J.D. Winefordner, 'Determination of Non-metallic Elements by Capacitively Coupled Helium Microwave Plasma Atomic-emission Spectrometry with Capillary Gas Chromatography', *Analyst*, **115**(7), 933–937 (1990).
382. D. Huang, D.C. Liang, M.W. Blades, 'Capacitatively Coupled Plasma Detector for Gas Chromatography', *J. Anal. At. Spectrom.*, **4**(8), 789–791 (1989).
383. S. Hanamura, W.J. Wang, J.D. Winefordner, 'Determination of Hydrogen and Oxygen in Metals with the Aid of a Helium Single-electrode Microwave-plasma Emission Technique', *Can. J. Spectrosc.*, **30**(2), 46–49 (1985).
384. A.M. Pless, B.W. Smith, M.A. Bolshov, J.D. Winefordner, 'A Capacitively Coupled Microwave-plasma Atomic-emission Spectrometer for the Determination of Trace Metals in Micro Samples', *Spectrochim. Acta, Part B*, **51**(1), 55–64 (1996).
385. A.H. Ali, K.C. Ng, J.D. Winefordner, 'Micro-sampling in Graphite-cup Capacitively Coupled Microwave-plasma Atomic-emission Spectroscopy', *Spectrochim. Acta, Part B*, **46**(8), 1207–1214 (1991).
386. A.H. Ali, K.C. Ng, J.D. Winefordner, 'Direct Solid Sampling in Capacitively Coupled Microwave-plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **6**(3), 211–213 (1991).
387. K. Kitagawa, N. Nishimoto, 'Thermal Vaporizer–Capacitively Coupled Microwave Plasma System for Trace Mercury Determination', *Bunko Kenkyu*, **38**(4), 282–287 (1989).
388. H. Uchida, W.R. Masamba, T. Uchida, B.W. Smith, J.D. Winefordner, 'New Desolvation System for Use with Capacitatively Coupled Microwave Plasma and Inductively Coupled Plasma Atomic-emission Spectrometry', *Appl. Spectrosc.*, **43**(3), 425–430 (1989).
389. B.M. Patel, J.P. Deavor, J.D. Winefordner, 'Solution Nebulization of Aqueous Samples into the Tubular-electrode Torch Capacitatively Coupled Microwave Plasma', *Talanta*, **35**(8), 641–645 (1988).
390. M. Seelig, N.H. Bings, J.A.C. Broekaert, 'Use of a Capacitively Coupled Microwave Plasma (CMP) with Ar, N_2 and Air as Working Gases for Atomic Spectrometric Elemental Determinations in Aqueous Solutions and Oils', *Fresenius' J. Anal. Chem.*, **360**(2), 161–166 (1998).
391. M.W. Blades, 'Atmospheric-pressure, Radio-frequency, Capacitively Coupled Helium Plasmas', *Spectrochim. Acta, Part B*, **49**(1), 47–57 (1994).
392. R.D. Deutsch, J.P. Keilson, G.M. Hieftje, 'Analytical Characteristics of the Microwave-induced Nitrogen Discharge at Atmospheric Pressure (MINDAP)', *Appl. Spectrosc.*, **39**(3), 531–534 (1985).
393. R.D. Deutsch, G.M. Hieftje, 'Development of a Microwave-induced Nitrogen Discharge at Atmospheric Pressure (MINDAP)', *Appl. Spectrosc.*, **39**(2), 214–222 (1985).
394. Y.X. Duan, X.G. Du, Q.H. Jin, 'Comparative Studies of Surfatron and Microwave Plasma Torch Sources for Determination of Mercury by Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **9**(5), 629–633 (1994).
395. J.F. Yang, C. Schickling, J.A.C. Broekaert, P. Tschöpel, G. Tölg, 'Evaluation of Continuous Hydride Generation Combined with Helium and Argon Microwave-induced Plasmas Using a Surfatron for Atomic-emission Spectrometric Determination of Arsenic, Antimony and Selenium', *Spectrochim. Acta, Part B*, **50**(11), 1351–1363 (1995).
396. K.G. Michlewicz, J.W. Carnahan, 'Quantitation of Trace Aqueous Halides by Volatilization into a Microwave-induced Helium Plasma', *Anal. Chim. Acta*, **183**, 275–280 (1986).
397. J.E. Freeman, G.M. Hieftje, 'Interferometric Detection of Near-infrared Non-metal Atomic Emission from a Microwave-induced Plasma', *Appl. Spectrosc.*, **39**(2), 211–214 (1985).
398. Y. Madrid, M. Wu, Q.H. Jin, G.M. Hieftje, 'Evaluation of Flow Injection Techniques for Microwave Plasma Torch Atomic-emission Spectrometry', *Anal. Chim. Acta*, **277**(1), 1–8 (1993).
399. D.M. Ye, H.Q. Zhang, Q.H. Jin, 'Flow Injection Online Column Preconcentration for Low Powered Microwave Plasma Torch Atomic-emission Spectrometry', *Talanta*, **43**(4), 535–544 (1996).
400. H.Q. Zhang, D.M. Ye, J.H. Zhao, J.L. Yu, R.Z. Men, Q.H. Jin, D.F. Dong, 'Flow Injection Sample Introduction into Microwave Plasma', *Microchem. J.*, **53**(1), 69–78 (1996).
401. Q. Jin, H.Q. Zhang, F. Liang, W.J. Yang, Q.H. Jin, 'Determination of Trace Amounts of Boron by Microwave Plasma Torch Atomic-emission Spectrometry Using an Online Separation and Preconcentration Technique', *J. Anal. At. Spectrom.*, **11**(5), 331–337 (1996).
402. Q.H. Jin, H.Q. Zhang, F. Liang, Q. Jin, 'Some Observations on the On-line Anion-exchange Preconcentration for Microwave Plasma-torch Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **10**(10), 875–879 (1995).

403. F. Liang, H.Q. Zhang, Q. Jin, D.X. Zhang, Y.H. Lei, Q.H. Jin, 'Use of Microwave Plasma Torch Atomic Emission Spectrometry for the Determination of Silicon', *Fresenius' J. Anal. Chem.*, **357**(4), 384–388 (1997).
404. Y. Duan, Y. Li, X. Tian, H. Zhang, Q. Jin, 'Analytical Performance of the Microwave Plasma Torch in the Determination of Rare-earth Elements with Optical-emission Spectrometry', *Anal. Chim. Acta*, **295**(3), 315–324 (1994).
405. Q.H. Jin, H.Q. Zhang, D.M. Ye, J.S. Zhang, 'Study of a Low-powered Helium Microwave-induced Plasma for Determination of Phosphorus, Chlorine, Bromine and Iodine by Atomic-emission Spectrometry', *Microchem. J.*, **47**(3), 278–286 (1993).
406. J.J. Urh, J.W. Carnahan, 'Determination of Metals in Aqueous Solution by Direct Nebulization into an Air Microwave-induced Plasma', *Anal. Chem.*, **57**(7), 1253–1255 (1985).
407. K.G. Michlewicz, J.W. Carnahan, 'Determination of Aqueous Chloride by Direct Nebulization into a Helium Microwave-induced Plasma', *Anal. Chem.*, **57**(6), 1092–1095 (1985).
408. R. Gross, B. Platzer, E. Leitner, A. Schalk, H. Sinabell, H. Zach, G. Knapp, 'Atomic-emission Gas-chromatographic Detection – Chemical and Spectral Interferences in the Stabilized Capacitive Plasma (SCP)', *Spectrochim. Acta, Part B*, **47**(1), 95–106 (1992).
409. J.F. Camuna, M. Montes, R. Pereiro, A. Sanz-Medel, C. Katschthaler, R. Gross, G. Knapp, 'Determination of Halides by Microwave-induced Plasma and Stabilized Capacitive Plasma Atomic-emission Spectrometry After Online Continuous Halogen Generation', *Talanta*, **44**(4), 535–544 (1997).
410. S. Luge, J.A.C. Broekaert, A. Schalk, H. Zach, 'The Use of Different Sample-introduction Techniques in Combination with the Low-power Stabilized Capacitive Plasma (SCP) as a Radiation Source for Atomic-emission Spectrometry', *Spectrochim. Acta, Part B*, **50**(4–7), 441–452 (1995).
411. Y.X. Duan, M.Y. Hou, Z.H. Du, Q.H. Jin, 'Evaluation of the Performance of Microwave-induced Plasma Atomic-absorption Spectrometry (MIP/AAS)', *Appl. Spectrosc.*, **47**(11), 1871–1879 (1993).
412. Y.X. Duan, X.Y. Li, Q.H. Jin, 'Electrothermal Vaporization for Sample Introduction in Microwave-induced-plasma Atomic-absorption Spectrometry', *J. Anal. At. Spectrom.*, **8**(8), 1091–1096 (1992).
413. K.C. Ng, R.S. Jensen, M.J. Brechmann, W.C. Santos, 'Microwave-induced Plasma Atomic-absorption Spectrometry with Solution Nebulization', *Anal. Chem.*, **60**(24), 2818–2821 (1988).
414. K.C. Ng, T.J. Garner, 'Microwave-induced Plasma Atomic-absorption Spectrometry with Solution Nebulization and Desolvation–Condensation', *Appl. Spectrosc.*, **47**(2), 241–242 (1993).
415. Y.X. Duan, H.Q. Zhang, M.Y. Huo, Q.H. Jin, 'Desolvation Effect on the Analytical Performance of Microwave-induced-plasma Atomic-absorption Spectrometry (MIP/AAS)', *Spectrochim. Acta, Part B*, **49**(6), 583–592 (1994).
416. Y.X. Duan, M.Y. Huo, J. Liu, Q.H. Jin, 'Study on Microwave-induced Plasma Atomic-absorption Spectrometry (MIP/AAS) for the Determination of Calcium in Axial Viewing Mode', *Fresenius' J. Anal. Chem.*, **349**(4), 277–282 (1994).
417. A. Zybin, K. Niemax, 'GC Analysis of Chlorinated Hydrocarbons in Oil and Chlorophenols in Plant Extracts Applying Element-selective Diode Laser Plasma Detection', *Anal. Chem.*, **69**(4), 755–757 (1997).
418. C. Schnuerer-Patschan, K. Niemax, 'Element-selective Detection of Chlorine in Capillary Gas Chromatography by Wavelength-modulation Diode-laser Atomic-absorption Spectrometry in a Microwave-induced Plasma', *Spectrochim. Acta, Part B*, **50**(9), 963–969 (1995).
419. A. Zybin, C. Schnuerer-Patschan, K. Niemax, 'Wavelength Modulation Diode Laser Atomic-absorption Spectrometry in Modulated Low-pressure Helium Plasmas for Element-selective Detection in Gas Chromatography', *J. Anal. At. Spectrom.*, **10**(9), 563–567 (1995).
420. H. Lu, Y.L. Ren, H.Q. Zhang, Q. Jin, 'Quantitation of Cadmium by Microwave-induced-plasma Atomic-absorption Spectrometry', *Microchem. J.*, **44**(1), 86–92 (1991).
421. L.D. Perkins, G.L. Long, 'Evaluation of Line and Continuum Sources for Atomic-fluorescence Spectrometry Using a Low-powered Argon Microwave-induced Plasma', *Appl. Spectrosc.*, **42**(7), 1285–1289 (1988).
422. L.D. Perkins, G.L. Long, 'Characterization of a High-efficiency Helium Microwave-induced Plasma as an Atomization Source for Atomic-spectrometric Analysis', *Appl. Spectrosc.*, **43**(3), 499–504 (1989).
423. Y.X. Duan, X.G. Du, Y.M. Li, Q.H. Jin, 'Characterization of a Modified, Low-power Argon Microwave Plasma Torch (MPT) as an Atomization Cell for Atomic-fluorescence Spectrometry', *Appl. Spectrosc.*, **49**(8), 1079–1085 (1995).
424. Y. Duan, X. Kong, H. Zhang, J. Liu, Q. Jin, 'Evaluation of a Low-powered Argon Microwave Plasma Discharge as an Atomizer for the Determination of Mercury by Atomic-fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **7**(1), 7–10 (1992).
425. Y. Oki, H. Uda, C. Honda, M. Maeda, J. Izumi, T. Morimoto, M. Tanoura, 'Laser-induced-fluorescence Detection of Sodium Atomized by a Microwave-induced Plasma with Tungsten-filament Vaporization', *Anal. Chem.*, **62**(7), 680–683 (1990).
426. V. Rigin, 'Simultaneous Atomic-fluorescence-spectrometric Determination of Traces of Iron, Cobalt and Nickel After Conversion to Their Carbonyls and Gas-phase Atomization by Microwave-induced Plasma', *Anal. Chim. Acta*, **283**(2), 895–901 (1993).

427. Y. Oki, E. Tashiro, M. Maeda, C. Honda, Y. Hasegawa, H. Futami, J. Izumi, K. Matsuda, 'Sensitive Detection of Trace Elements in Pure Water by Laser-induced Atomic-fluorescence Spectroscopy in Microwave Discharge Atomization', *Anal. Chem.*, **65**(15), 2096–2101 (1993).
428. G.C. Turk, J.C. Travis, J.R. DeVoe, T.C. O'Haver, 'Laser-enhanced Ionization Spectrometry in Analytical Flames', *Anal. Chem.*, **51**(12), 1890–1896 (1979).
429. R.S. Lysakowski Jr, R.E. Dessey, G.L. Long, 'Laser-enhanced Ionization in Microwave-induced Plasmas', *Appl. Spectrosc.*, **43**(7), 1139–1145 (1989).
430. M.D. Seltzer, R.B. Green, 'Electrical Characteristics of Microwave-induced Plasmas for Laser-induced-ionization Spectrometry', *Spectrosc. Lett.*, **22**(4), 461–470 (1989).
431. M.D. Seltzer, R.B. Green, 'Active Nitrogen Plasma Atom Reservoir for Laser-induced Ionization Spectrometry', *Spectrosc. Lett.*, **20**(8), 601–617 (1987).
432. M. Okumura, K. Fukushi, S.N. Willie, R.E. Sturgeon, 'Evaluation of Atomic Fluorescence, Absorption and Emission Techniques for the Determination of Mercury', *Fresenius' J. Anal. Chem.*, **345**(8–9), 570–574 (1993).
433. H. Brouwers, S. De Jaegere, 'Plasma Afterglow Atomization in Electrothermal Atomic-absorption Spectrometry', *Spectrochim. Acta, Part B*, **43**(8), 901–915 (1988).
434. D.J. Douglas, J.B. French, 'Elemental Analysis with a Microwave-induced Plasma/Quadrupole Mass Spectrometer System', *Anal. Chem.*, **53**(1), 37–41 (1981).
435. D.J. Douglas, E.S.K. Quan, R.G. Smith, 'Elemental Analysis with an Atmospheric-pressure Plasma (Microwave-induced Plasma, ICP)/Quadrupole Mass Spectrometer System', *Spectrochim. Acta, Part B*, **38**(1–2), 39–48 (1983).
436. J.T. Creed, T.M. Davidson, W.L. Shen, P.G. Brown, J.A. Caruso, 'Helium Microwave-induced Plasma Mass Spectrometry for Detection of Metals and Non-metals in Aqueous Solutions', *Spectrochim. Acta, Part B*, **44**(9), 909–924 (1989).
437. E.H. Evans, J.J. Giglio, T.M. Castillano, J.A. Caruso, N.W. Barnett, *Inductively Coupled and Microwave-induced Plasma Sources for Mass Spectrometry*, Royal Society of Chemistry, Cambridge, 1995.
438. F.A. Byrdy, J.A. Caruso, 'Elemental Analysis of Environmental Samples. Using Chromatography Coupled with Plasma Mass Spectrometry', *Environ. Sci. Technol.*, **28**(12), 528A–534A (1994).
439. B.S. Sheppard, J.A. Caruso, 'Plasma Mass Spectrometry: Consider the Source', *J. Anal. At. Spectrom.*, **9**(3), 145–149 (1994).
440. N.P. Vela, L.K. Olson, J.A. Caruso, 'Elemental Speciation with Plasma Mass Spectrometry', *Anal. Chem.*, **65**(13), 585A–597A (1993).
441. C.J. Park, Y.N. Pak, K.W. Lee, 'Isotope-ratio Measurements of Iron, Bromine and Selenium in Aqueous Solutions by Helium Microwave-induced-plasma Mass Spectrometry', *Anal. Sci.*, **8**(3), 443–448 (1992).
442. L.K. Olson, J.A. Caruso, 'Helium Microwave-induced Plasma: An Alternative Ion Source for Plasma Mass Spectrometry', *Spectrochim. Acta, Part B*, **49**(1), 7–30 (1994).
443. W.L. Shen, R.D. Satzger, 'Development of a Low-power Microwave Atmospheric-pressure Molecular Ionization Source for Mass Spectrometry with Direct Introduction of Gaseous and Liquid Organic Samples', *Anal. Chem.*, **63**(18), 1960–1964 (1991).
444. E. Poussel, J.M. Mermet, D. Deruaz, C. Beaugrand, 'Evaluation of a Microwave-induced Plasma as a Soft Ionization Source in Mass Spectrometry', *Anal. Chem.*, **60**(9), 923–927 (1988).
445. A.G. Kouzmin, L.N. Gall, M.Z. Muradimov, 'About Some Peculiarities of Ionization in Radio-frequency Glow-discharge (r.f. GD)', *Fresenius' J. Anal. Chem.*, **355**(7–8), 866–867 (1996).
446. D.A. Wilson, G.H. Vickers, G.M. Hieftje, 'Use of Microwave-induced Nitrogen Discharge at Atmospheric Pressure as an Ion Source for Elemental Mass Spectrometry', *Anal. Chem.*, **59**(13), 1664–1670 (1987).
447. D. Boudreau, J. Hubert, 'Atmospheric-pressure Argon Surface-wave Plasma (SWP) as an Ion Source in Elemental Mass Spectrometry', *Appl. Spectrosc.*, **47**(5), 609–614 (1993).
448. N.P. Vela, J.A. Caruso, R.D. Satzger, 'Potential for an Atmospheric-pressure Low-power Microwave-induced Plasma Ionization Source for Mass Spectrometry', *Appl. Spectrosc.*, **51**(10), 1500–1503 (1997).
449. M. Wu, Y.X. Duan, Q.H. Jin, G.M. Hieftje, 'Elemental Mass Spectrometry Using a Helium Microwave Plasma Torch as an Ion Source', *Spectrochim. Acta, Part B*, **49**(2), 137–148 (1994).
450. J.T. Creed, A.H. Mohamad, T.M. Davidson, G. Ataman, J.A. Caruso, 'Helium Source Microwave-induced Plasma Mass-spectrometric Detection in the Analysis of Gas-chromatographic Eluates', *J. Anal. At. Spectrom.*, **3**(6), 923–926 (1988).
451. P.G. Brown, T.M. Davidson, J.A. Caruso, 'Application of Helium Microwave-induced Plasma Mass Spectrometry to the Detection of High Ionization Potential Gas-phase Species', *J. Anal. At. Spectrom.*, **3**(6), 763–769 (1988).
452. Q. Jin, G. Yang, Z. Guo, A. Yu, J. Liu, 'Some Observations on the Development of a Gas-chromatographic Microwave-plasma Ionization Detector', *Microchem. J.*, **35**(3), 281–287 (1987).
453. B. Pack, J.A.C. Broekaert, J. Guzowski, J. Poehlman, G.M. Hieftje, 'Determination of Halogenated Hydrocarbons by Helium Microwave Plasma Torch Time-of-flight Mass Spectrometry Coupled to Gas Chromatography', *Anal. Chem.*, **70**(18), 3957–3963 (1998).
454. A.H. Mohamad, J.T. Creed, T.M. Davidson, J.A. Caruso, 'Detection of Halogenated Compounds by Capillary Gas Chromatography with Helium-plasma Mass

Spectrometry Detection', *Appl. Spectrosc.*, **43**(7), 1127–1131 (1989).
455. J.T. Jelink, A. Venema, 'Investigations into the Use of Capillary GC with Atomic-emission Detection', *J. High Resolut. Chromatogr.*, **13**(6), 447–450 (1990).
456. R.D. Satzger, F.L. Fricke, P.G. Brown, J.A. Caruso, 'Detection of Halogens as Positive Ions Using a He Microwave-induced Plasma as an Ion Source for Mass Spectrometry', *Spectrochim. Acta, Part B*, **42**(5), 705–712 (1987).
457. W.C. Story, J.A. Caruso, 'Gas-chromatographic Determination of Phosphorus, Sulfur and Halogens (in Pesticides) Using a Water-cooled Torch with Reduced-pressure Helium Microwave-induced Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **8**(4), 571–575 (1993).
458. W.C. Story, L.K. Olson, W.L. Shen, J.T. Creed, J.A. Caruso, 'Reduced-pressure Microwave-induced Plasma Mass Spectrometric Detection of Phosphorus and Sulphur in Gas Chromatographic Eluates', *J. Anal. At. Spectrom.*, **5**(6), 467–470 (1990).
459. T. Shirasaki, K. Yasuda, 'Mass Spectra of Oxygen–Nitrogen Plasma by Introduction of Organic Solvent in a Microwave-induced-plasma Mass Spectrometer', *Anal. Sci.*, **8**(3), 375–376 (1992).
460. R.A. Heppner, 'Elemental Detection with a Microwave-induced Plasma–Gas Chromatograph–Mass Spectrometer System', *Anal. Chem.*, **55**(13), 2170–2174 (1983).
461. H. Suyani, J. Creed, J. Caruso, R.D. Satzger, 'Helium Microwave-induced Plasma Mass Spectrometry for Capillary Gas Chromatographic Detection: Speciation of Organotin Compounds', *J. Anal. At. Spectrom.*, **4**(8), 777–782 (1989).
462. L.K. Olson, J.A. Caruso, 'Determination of Halogenated Compounds with Supercritical-fluid Chromatography–Microwave-induced Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **7**(6), 993–998 (1992).
463. D. Heitkemper, J. Creed, J.A. Caruso, 'Helium Microwave-induced Plasma Mass-spectrometric Detection for Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr. Sci.*, **28**(4), 175–181 (1990).
464. Y.X. Duan, M. Wu, Q.H. Jin, G.M. Hieftje, 'Vapor Generation of Non-metals Coupled to Microwave Plasma-torch Mass Spectrometry', *Spectrochim. Acta, Part B*, **50**(9), 1095–1108 (1995).
465. Y.X. Duan, M. Wu, Q.H. Jin, G.M. Hieftje, 'Vapor-generation-assisted Nebulization for Non-metal Determination', *Spectrochim. Acta, Part B*, **50**(9), 971–974 (1995).
466. M. Moini, G. Li, F. Perez, F.E. Ibarra, D. Sandoval, 'Selective Detection of Selenium in Water Utilizing Chemical Reaction Interface Mass Spectrometry', *J. Mass Spectrom.*, **32**(4), 420–424 (1997).
467. J.T. Morre, M. Moini, 'Selective Detection and Characterization of Chlorine- and Bromine-containing Compounds in Complex Mixtures Using Microwave-induced Plasma–Chemical-reaction-interface Mass Spectrometry', *Biol. Mass. Spectrom.*, **21**(12), 693–699 (1992).
468. K. Jankowski, 'Application of the Microwave-induced Plasma Atomic-emission Spectrometry to Platinum-group Metals Analysis', *Chem. Anal. (Warsaw)*, **39**(3), 381–387 (1994).
469. J.J. Giglio, J.S. Wang, J.A. Caruso, 'Evaluation of a Direct-injection Nebulizer (DIN) for Liquid Sample Introduction in Helium Microwave-induced Plasma Mass Spectrometry (MIP/MS)', *Appl. Spectrosc.*, **49**(3), 314–319 (1995).
470. W.S. Shen, T.M. Davidson, J.T. Creed, J.A. Caruso, 'Moderate-power Nitrogen Microwave-induced Plasma as an Alternative Ion Source for Mass Spectrometry', *Appl. Spectrosc.*, **44**(6), 1003–1010 (1990).
471. R.D. Satzger, F.L. Fricke, J.A. Caruso, 'Elemental Mass Spectrometry Using a Moderate-power Microwave-induced Plasma as an Ion Source', *J. Anal. At. Spectrom.*, **3**(2), 319–323 (1988).
472. Y. Okamoto, 'High-sensitivity Microwave-induced Plasma Mass Spectrometry for Trace Element Analysis', *J. Anal. At. Spectrom.*, **9**(7), 745–749 (1994).
473. J. Yoshinaga, T. Shirasaki, K. Oishi, M. Morita, 'Isotope Dilution Analysis of Selenium in Biological Materials by Nitrogen Microwave-induced Plasma Mass Spectrometry', *Anal. Chem.*, **67**(9), 1568–1574 (1995).
474. P.A. Fecher, A. Nagengast, 'Trace Analysis in High Matrix Aqueous Solutions Using Helium Microwave-induced Plasma Mass Spectrometry', *J. Anal. At. Spectrom.*, **9**(9), 1021–1027 (1994).
475. T. Shirasaki, K. Hiraki, 'Determination of Trace Elements in a Photoresist Solution by Microwave-induced Plasma Mass Spectrometry', *Bunseki Kagaku*, **43**(1), 25–29 (1994).
476. M. Ceulemans, F.C. Adams, 'GC-based Hyphenated Techniques Applied to the Speciation Analysis of Organometals', *LC-GC Int.*, **7**(12), 694–697 (1994).
477. S.J. Hill, 'Advances in Coupled Techniques for Speciation Studies', *Anal. Proc.*, **29**(9), 399–401 (1992).
478. A.M. Carro-Diaz, R.A. Lorenzo-Ferreira, R. Cela-Torrijos, 'Speciation of Organomercurials in Biological and Environmental Samples by Gas Chromatography with Microwave-induced Plasma Atomic-emission Detection', *J. Chromatogr. A*, **683**(1), 245–252 (1994).
479. P. Lansens, C. Meuleman, W. Baeyens, 'Long-term Stability of Methylmercury Standard Solutions in Distilled, Deionized Water', *Anal. Chim. Acta*, **229**(2), 281–285 (1990).
480. B. Rosenkranz, C.B. Breer, W. Buscher, J. Bettmer, K. Cammann, 'The Plasma Emission Detector – a Suitable Detector for Speciation and Sum Parameter Analysis', *J. Anal. At. Spectrom.*, **12**(9), 993–996 (1997).
481. A. Sanz-Medel, 'Atomic Spectrometry in the Analysis and Research of Biological Systems', *Egypt. J. Anal. Chem.*, **3**(1), 60–67 (1994).

482. I.S. Krull, S.W. Jordan, S. Kahl, S.B. Smith, Jr, 'Trace Analysis for Steroidal Carboranes Via Gas Chromatography–Microwave-induced Plasma Emission-spectroscopic Detection', *J. Chromatogr. Sci.*, **20**(11), 489–498 (1982).
483. S.W. Jordan, I.S. Krull, S.B. Smith Jr, 'Trace Analysis for Catechol Derivatives Via Boronate Ester Formation and Gas Chromatographic–Microwave-induced Plasma-emission Spectroscopic Detection (GC/MIP)', *Anal. Lett.*, **15**(A14), 1131–1148 (1982).
484. D. Deruaz, A. Bannier, C. Pionchon, M.S. Boukraa, W. Elbast, J.L. Brazier, 'Limit of Detection of Nitrogen-15 by Gas-chromatography Atomic-emission Detection. Optimization Using an Experimental Design', *Anal. Lett.*, **28**(11), 2095–2113 (1995).
485. P. Clarkson, M. Cooke, 'The Identification of an Unusual Volatile Component in Processed Tobacco by Gas Chromatography with Mass Spectrometry and Atomic-emission Detection', *Anal. Chim. Acta*, **335**(3), 253–259 (1996).
486. J. Hajslova, P. Cuhra, M. Kempny, J. Poustka, K. Holadova, V. Kocourek, 'Determination of Polychlorinated Biphenyls in Biotic Matrices Using Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *J. Chromatogr. A*, **699**(1–2), 231–239 (1995).
487. K.C. Ting, P. Kho, 'GC–Microwave-induced Plasma Atomic-emission Detection Method for Pesticide Residue Determination in Fruits and Vegetables', *J. Assoc. Off. Anal. Chem.*, **74**(6), 991–998 (1991).
488. M.B. De La Calle Guntinas, C. Brunori, R. Scerbo, S. Chiavarini, P. Quevauviller, F.C. Adams, R. Morabito, 'Determination of Selenomethionine in Wheat Samples: Comparison of Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry, Gas Chromatography–Flame-photometric Detection and Gas Chromatography–Mass Spectrometry', *J. Anal. At. Spectrom.*, **12**(9), 1041–1046 (1997).
489. E. Block, X.J. Cai, P.C. Uden, X. Zhang, B.D. Quimby, J.J. Sullivan, 'Allium Chemistry: Natural Abundance of Organoselenium Compounds from Garlic, Onion and Related Plants and in Human Garlic Breath', *Pure Appl. Chem.*, **68**(4), 937–944 (1996).
490. C. Gerbersmann, R. Lobinski, F.C. Adams, 'Determination of Volatile Sulfur Compounds in Water Samples, Beer and Coffee with Purge and Trap Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **316**(1), 93–104 (1995).
491. R. Lobinski, J. Szpunar Lobinska, F.C. Adams, P.L. Teissedre, J.C. Cabanis, 'Speciation Analysis of Organolead Compounds in Wine by Capillary Gas Chromatography Microwave-induced Plasma Atomic-emission Spectrometry', *J. AOAC Int.*, **76**(6), 1262–1267 (1993).
492. P. Lansens, C. Casais Laino, C. Meuleman, W. Baeyens, 'Evaluation of Gas-chromatographic Columns for the Determination of Methylmercury in Aqueous Headspace Extracts from Biological Samples', *J. Chromatogr.*, **586**(2), 329–340 (1991).
493. B. Sarx, K. Baechmann, 'Speciation of Arsenic Compounds by Volatilization from Solid Samples', *Fresenius' Z. Anal. Chem.*, **316**(6), 621–626 (1983).
494. S. Hanamura, B.W. Smith, J.D. Winefordner, 'Speciation of Inorganic and Organometallic Compounds in Solid Biological Samples by Thermal Vaporization and Plasma Emission Spectrometry', *Anal. Chem.*, **55**(13), 2026–2032 (1983).
495. D.F. Hagen, J. Belisle, J.S. Marhevka, 'Capillary GC Helium Microwave Emission Detector Characterization of Fluorine-containing Metabolites in Blood Plasma', *Spectrochim. Acta, Part B*, **38**(1–2), 377–385 (1983).
496. K. Chiba, K. Yoshida, K. Tanabe, M. Ozaki, H. Haraguchi, J.D. Winefordner, 'Determination of Ultra-trace Levels of Fluorine in Water and Urine Samples by a Gas-chromatographic–Atmospheric-pressure Helium Microwave-induced Plasma Emission-spectrometric System', *Anal. Chem.*, **54**(4), 761–764 (1982).
497. B. Lind, R. Body, L. Friberg, 'Mercury Speciation in Blood and Brain Tissue from Monkeys. Inter-laboratory Comparison of Magos' Method with Other Spectroscopic Methods, Using Alkylation and Gas Chromatography Separation as well as RNAA (Radiochemical NAA) in Combination with Westoo's Extraction Methods', *Fresenius' J. Anal. Chem.*, **345**(2–4), 314–317 (1993).
498. E. Bulska, H. Emteborg, D.C. Baxter, W. Frech, D. Ellingsen, Y. Thomassen, 'Speciation of Mercury in Human Whole Blood by Capillary Gas Chromatography with a Microwave-induced Plasma Emission Detector System Following Complexometric Extraction and Butylation', *Analyst*, **117**(3), 657–663 (1992).
499. A.D. Besteman, N. Lau, D.Y. Liu, B.W. Smith, J.D. Winefordner, 'Determination of Lead in Whole Blood by Capacitively-coupled Microwave Plasma Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **11**(7), 479–481 (1996).
500. M.W. Wensing, D.Y. Liu, B.W. Smith, J.D. Winefordner, 'Determination of Lead in Whole Blood Using a Capacitively Coupled Microwave Plasma Atomic-emission Spectrometer', *Anal. Chim. Acta*, **299**(1), 1–7 (1994).
501. M.W. Wensing, B.W. Smith, J.D. Winefordner, 'Capacitively Coupled Microwave Plasma Atomic-emission Spectrometer for the Determination of Lead in Whole Blood', *Anal. Chem.*, **66**(4), 531–535 (1994).
502. S.C. Hight, M.T. Corcoran, 'Rapid Determination of Methylmercury in Fish and Shellfish: Method Development', *J. Assoc. Off. Anal. Chem.*, **70**(1), 24–30 (1987).
503. K. Fukushi, S.N. Willie, R.E. Sturgeon, 'Subnanogram Determination of Inorganic and Organic Mercury by Helium-microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Lett.*, **26**(2), 325–340 (1993).

504. A. Tsunoda, K. Matsumoto, H. Haraguchi, K. Fuwa, 'Application of Microwave-induced Helium Plasma Atomic-emission Spectrometry in Gas-chromatographic Detection of Dimethyl Selenide in Biological Samples', *Anal. Sci.*, **2**(1), 99–100 (1986).
505. C. Schickling, J. Yang, J.A.C. Broekaert, 'Optimization of Electrochemical Hydride Generation Coupled to Microwave-induced Plasma Atomic-emission Spectrometry for the Determination of Arsenic and its Use for the Analysis of Biological Tissue Samples', *J. Anal. At. Spectrom.*, **11**(9), 739–745 (1996).
506. J. Szpunar, V.O. Schmitt, R. Lobinski, J.L. Monod, 'Rapid Speciation of Butyltin Compounds in Sediments and Biomaterials by Capillary Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry After Microwave-assisted Leaching-digestion', *J. Anal. At. Spectrom.*, **11**(3), 193–199 (1996).
507. T. Suzuki, R. Matsuda, Y. Saito, H. Yamada, 'Application of Helium Microwave-induced Plasma Emission Detection System to Analysis of Organotin Compounds in Biological Samples', *J. Agric. Food Chem.*, **42**(1), 216–220 (1994).
508. S.G. Dai, C.R. Jia, 'Applications of Hyphenated Techniques in Environmental Analysis', *Anal. Sci.*, **12**(2), 355–361 (1996).
509. H. Frischenschlager, C. Mittermayr, M. Peck, E. Rosenberg, M. Grasserbauer, 'The Potential of Gas Chromatography with Microwave-induced Plasma Atomic-emission Detection (GC/MIP/AED) as a Complementary Analytical Technique in Environmental Screening Analysis of Aqueous Samples', *Fresenius' J. Anal. Chem.*, **359**(3), 213–221 (1997).
510. H. Frischenschlager, M. Peck, C. Mittermayr, E. Rosenberg, M. Grasserbauer, 'Improved Screening Analysis of Organic Pollutants in River Water Samples by Gas Chromatography with Atomic Emission Detection (GC/MIPAED)', *Fresenius' J. Anal. Chem.*, **357**(8), 1133–1141 (1997).
511. R. Roehl, H.J. Hoffmann, 'Determination of Organic Halogen Compounds by MIP/AES [Microwave-induced Plasma–a.e.s.]', *Vom Wasser*, **63**, 225–237 (1984).
512. B.D. Quimby, M.F. Delaney, P.C. Uden, R.M. Barnes, 'Determination of Trihalomethanes in Drinking Water by Gas Chromatography with a Microwave-plasma Emission Detector', *Anal. Chem.*, **51**(7), 875–880 (1979).
513. K. Chiba, H. Haraguchi, 'Determination of Halogenated Organic Compounds in Water by Gas Chromatography–Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry with a Heated Discharge Tube for Pyrolysis', *Anal. Chem.*, **55**(9), 1504–1508 (1983).
514. P.L. Wylie, R. Oguchi, 'Pesticide Analysis by Gas Chromatography with a Novel Atomic-emission Detector', *J. Chromatogr.*, **517**, 131–142 (1990).
515. C. Flodin, M. Ekelund, H. Boren, A. Grimvall, 'Pyrolysis–GC/AED and Pyrolysis–GC/MS Analysis of Chlorinated Structures in Aquatic Fulvic Acids and Chlorolignins', *Chemosphere*, **34**(11), 2319–2328 (1997).
516. J.S. Alvarado, J.W. Carnahan, 'Reductive Pyrolysis for the Determination of Aqueous Sulfur Compounds with a Helium Microwave-induced Plasma', *Anal. Chem.*, **65**(22), 3295–3298 (1993).
517. K.C. Ng, W. Shen, 'Solution Nebulization into a Low-power Argon Microwave-induced Plasma for Atomic-emission Spectrometry Study of Synthetic Ocean Water', *Anal. Chem.*, **58**(9), 2084–2087 (1986).
518. R. Lobinski, F.C. Adams, 'Sensitive Speciation Analysis of Lead in Environmental Waters by Capillary Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **262**(2), 285–297 (1992).
519. T. Nakahara, S. Yamada, T. Wasa, 'Determination of Total Iodine in Brines and Sea-waters by Atmospheric-pressure Helium Microwave-induced Plasma Atomic-emission Spectrometry with Continuous-flow Gas-phase Sample Introduction', *Appl. Spectrosc.*, **45**(9), 1561–1563 (1991).
520. S. Natarajan, 'Determination of Parts-per-trillion Levels of Mercury with Low-power Microwave-induced Argon-plasma Emission Spectrometry', *At. Spectrosc.*, **9**(2), 59–62 (1988).
521. Y. Nojiri, A. Otsuki, K. Fuwa, 'Determination of Sub-nanogram-per-litre Levels of Mercury in Lake Water with Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Chem.*, **58**(3), 544–547 (1986).
522. I. Atsuya, K. Akatsuka, 'Determination of Trace Amounts of Arsenic Using a Capacitively Coupled Microwave Plasma Combined with Hydride Generation: Application to Sewage Sludge, Iron and Steels', *Spectrochim. Acta, Part B*, **36**(7), 747–755 (1981).
523. V. Minganti, R. Capelli, R. De Pellegrini, 'Evaluation of Different Derivatization Methods for the Multi-element Detection of Mercury, Lead and Tin Compounds by Gas Chromatography–Microwave-induced Plasma–Atomic Emission Spectrometry in Environmental Samples', *Fresenius' J. Anal. Chem.*, **351**(4–5), 471–477 (1995).
524. M. Paneli, E. Rosenberg, M. Grasserbauer, M. Ceulemans, F. Adams, 'Assessment of Organolead Species in the Austrian Danube Basin Using GC/MIPAED', *Fresenius' J. Anal. Chem.*, **357**(6), 756–762 (1997).
525. S. Girousi, E. Rosenberg, A. Voulgaropoulos, M. Grasserbauer, 'Speciation Analysis of Organotin Compounds in Thermaikos Gulf by GC/MIPAED', *Fresenius' J. Anal. Chem.*, **358**(7–8), 828–832 (1997).
526. J. Szpunar Lobinska, M. Ceulemans, R. Lobinski, F.C. Adams, 'Flow Injection Sample Preparation for Organotin Speciation Analysis of Water by Capillary Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *Anal. Chim. Acta*, **278**(1), 99–113 (1993).

527. A. Sadiki, D.T. Williams, 'Speciation of Organotin and Organolead Compounds in Drinking Water by Gas Chromatography–Atomic-emission Spectrometry', *Chemosphere*, **32**(10), 1983–1992 (1996).
528. K. Chiba, K. Yoshida, K. Tanabe, H. Haraguchi, K. Fuwa, 'Determination of Alkylmercury in Sea-water at the Nanogram per Litre Level by Gas Chromatography–Atmospheric-pressure Helium Microwave-induced Plasma Emission Spectrometry', *Anal. Chem.*, **55**(3), 450–453 (1983).
529. H. Emteborg, D.C. Baxter, W. Frech, 'Speciation of Mercury in Natural Waters by Capillary Gas Chromatography with a Microwave-induced Plasma Emission Detector Following Pre-concentration Using a Dithiocarbamate Resin Micro-column Installed in a Closed Flow Injection System', *Analyst*, **118**(8), 1007–1013 (1993).
530. M. Johansson, H. Emteborg, B. Glad, F. Reinholdsson, D.C. Baxter, 'Preliminary Appraisal of a Novel Sampling and Storage Technique for the Speciation Analysis of Lead and Mercury in Seawater', *Fresenius' J. Anal. Chem.*, **351**(4–5), 461–466 (1995).
531. M. Ceulemans, F.C. Adams, 'Integrated Sample Preparation and Speciation Analysis for the Simultaneous Determination of Methylated Species of Tin, Lead and Mercury in Water by Purge-and-trap Injection Capillary Gas Chromatography–Atomic-emission Spectrometry', *J. Anal. At. Spectrom.*, **11**(3), 201–206 (1996).
532. M.L. Mena, C.W. McLeod, P. Jones, A. Withers, V. Minganti, R. Capelli, P. Quevauviller, 'Microcolumn Preconcentration and Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry (GC/MIP/AES) for Mercury Speciation in Waters', *Fresenius' J. Anal. Chem.*, **351**(4–5), 456–460 (1995).
533. M.B. De La Calle Guntinas, M. Ceulemans, C. Witte, R. Lobinski, F.C. Adams, 'Evaluation of a Purge-and-trap Injection System for Capillary Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry for the Determination of Volatile Selenium Compounds in Water', *Mikrochim. Acta*, **120**(1–4), 73–82 (1995).
534. R. Lobinski, J. Szpunar Lobinska, F.C. Adams, 'Ultratrace Analysis for Organolead Compounds in Greenland Snow', *Analusis*, **22**(7), M54–M57 (1994).
535. Y. Liu, V. Lopez-Avila, M. Alcaraz, W.F. Beckert, 'Simultaneous Determination of Organotin, Organolead and Organomercury Compounds in Environmental Samples Using Capillary Gas Chromatography with Atomic-emission Detection', *J. High Resolut. Chromatogr.*, **17**(7), 527–536 (1994).
536. H. Emteborg, E. Bjorklund, F. Odman, L. Karlsson, L. Mathiasson, W. Frech, 'Determination of Methylmercury in Sediments Using Supercritical Fluid Extraction and Gas Chromatography Coupled with Microwave-induced Plasma Atomic Emission Spectrometry', *Analyst*, **121**(1), 19–29 (1996).
537. G. Heltai, T. Jozsa, 'Nitrogen-15-tracer Technique with the MIPOES Detection of Stable Nitrogen-isotopes for Soil Ecological Studies', *Microchem. J.*, **51**(1–2), 245–255 (1995).
538. B. Pohl, K. Baechmann, 'Comparative Investigations of the Concentrations of Dimethylarsinic Acid (Hydroxydimethylarsine Oxide), Arsenic(III) and Arsenic(V) in Soil and Air-borne Dust Samples', *Fresenius' Z. Anal. Chem.*, **323**(8), 859–864 (1986).
539. H. Baumann, K.G. Heumann, 'Analysis of Organobromine Compounds and Hydrogen Bromide in Motor Car Exhaust Gases with a GC–Microwave Plasma System', *Fresenius' Z. Anal. Chem.*, **327**(2), 186–192 (1987).
540. J.L. Genna, W.D. McAninch, R.A. Reich, 'Atmospheric Microwave-induced Plasma Detector for the Gas-chromatographic Analysis of Low-molecular-weight Sulfur Gases [in Air]', *J. Chromatogr.*, **238**(1), 103–112 (1982).
541. H. Vermaak, O. Kujirai, S. Hanamura, J.D. Winefordner, 'Potential Method for Determination of Gaseous and Particulate Lead in Exhaust Gas by Microwave-induced Air-plasma Emission Spectrometry and Zeeman Furnace Atomic-absorption Spectrometry', *Can. J. Spectrosc.*, **31**(4), 95–99 (1986).
542. E.R. Adlard, *Chromatography in the Petroleum Industry*, Elsevier, Amsterdam, 1995.
543. J.P. Snell, W. Frech, Y. Thomassen, 'Performance Improvements in the Determination of Mercury Species in Natural Gas Condensate Using Online Amalgamation Trap or Solid Phase Microextraction with Capillary Gas Chromatography–Microwave-induced Plasma Atomic-emission Spectrometry', *Analyst*, **121**(8), 1055–1060 (1996).
544. *Proceedings [of the] 76th APCA [Air Pollution Control Association] Annual Meeting, June 19–24, 1983, Atlanta, Georgia*, Air Pollution Control Association, Pittsburgh, PA, 1983.
545. V. Siemens, T. Harju, T. Laitinen, K. Larjava, J.A.C. Broekaert, 'Applicability of Microwave-induced Plasma Optical Emission Spectrometry (MIPOES) for Continuous Monitoring of Mercury in Flue Gases', *Fresenius' J. Anal. Chem.*, **351**(1), 11–18 (1995).
546. K. Larjava, T. Laitinen, T. Vahlman, S. Artmann, V. Siemens, J.A.C. Broekaert, D. Klockow, 'Measurement and Control of Mercury Species in Flue Gases from Liquid Waste Incineration', *Int. J. Environ. Anal. Chem.*, **49**(1–2), 73–85 (1992).
547. T. Nakahara, K. Kawakami, T. Wasa, 'Continuous Determination of Low Concentrations of Mercury by Atomic-emission Spectrometry with Helium Microwave-induced Plasma', *Chem. Express.*, **3**(11), 651–654 (1988).
548. M. Seelig, J.A.C. Broekaert, Unpublished Work.
549. A.M. Pless, A. Croslyn, M.J. Gordon, B.W. Smith, J.D. Winefordner, 'Direct Determination of Cadmium in Solids Using a Capacitively Coupled Microwave Plasma

Atomic-emission Spectrometer', *Talanta*, **44**(1), 39–46 (1997).
550. S. Hanamura, B. Kirsch, J.D. Winefordner, 'Determination of Trace Levels of Water in Solid Samples by Evolved Gas Analysis–Helium Microwave Plasma Emission Spectrometry', *Anal. Chem.*, **57**(1), 9–13 (1985).
551. E.M. Skelly Frame, Y. Takamatsu, T. Suzuki, 'Characterization of Solid Particles by Helium Microwave-induced Plasma Atomic-emission Spectrometry', *Spectroscopy*, **11**(1), 17–22 (1996).
552. J.A.C. Broekaert, F. Leis, 'Application of Electrothermal Evaporation Using Direct Solids Sampling Coupled with Microwave-induced Plasma Optical Emission Spectroscopy to Elemental Determinations in Biological Matrices', *Mikrochim. Acta*, **II**(3–4), 261–272 (1985).
553. B. Raeymaekers, P. Van Espen, F. Adams, J.A.C. Broekaert, 'A Characterization of Spark-produced Aerosols by Automated Electron-probe Micro-analysis', *Appl. Spectrosc.*, **42**(1), 142–150 (1988).
554. A. Aziz, J.A.C. Broekaert, K. Laqua, F. Leis, 'A Study of Direct Analysis of Solid Samples Using Spark Ablation Combined with Excitation in an Inductively Coupled Plasma', *Spectrochim. Acta, Part B*, **39**(9–11), 1091–1103 (1984).
555. D.J.C. Helmer, J.P. Walters, 'Microscopic Investigations of the Material Eroded by a High-voltage Spark Discharge', *Appl. Spectrosc.*, **38**(3), 399–405 (1984).
556. D.J.C. Helmer, J.P. Walters, 'Analysis of the Effluent from a High-voltage Spark Discharge with a Microwave-induced Plasma', *Appl. Spectrosc.*, **38**(3), 392–398 (1984).
557. Yong Nam Pak, S.R. Koirtyohann, 'Direct Solid Sample Analysis in a Moderate-power Argon Microwave-induced Plasma with Spark Generation', *J. Anal. At. Spectrom.*, **9**(11), 1305–1310 (1994).
558. U. Engel, A. Kehden, E. Voges, J.A.C. Broekaert, 'Direct Solid Atomic Emission Spectrometric Analysis of Metal Samples by an Argon Microwave Plasma Torch Coupled to Spark Ablation', *Spectrochim. Acta, Part B*, **54**(9), 1279–1289 (1999).
559. M.M. Mohamed, T. Uchida, S. Minami, 'Direct Sample Introduction of Solid Material into a Pulse-operated MIP [Microwave-induced Plasma]', *Appl. Spectrosc.*, **43**(5), 794–800 (1989).
560. T. Ishizuka, Y. Uwamino, 'Atomic-emission Spectrometry of Solid Samples with Laser Vaporization–Microwave-induced Plasma System', *Anal. Chem.*, **52**(1), 125–129 (1980).
561. A. Kehden, J.A.C. Broekaert, Unpublished Work.
562. L. Hiddemann, J. Uebbing, A. Ciocan, O. Dessenne, K. Niemax, 'Simultaneous Multi-element Analysis of Solid Samples by Laser Ablation–Microwave-induced-plasma Optical Emission Spectrometry', *Anal. Chim. Acta*, **283**(1), 152–159 (1993).
563. P. Richner, M.W. Borer, K.R. Brushwyler, G.M. Hieftje, 'Comparison of Different Excitation Sources and Normalization Techniques in Laser-ablation AES Using a Photodiode-based Spectrometer', *Appl. Spectrosc.*, **44**(8), 1290–1296 (1990).
564. J. Uebbing, A. Ciocan, K. Niemax, 'Optical Emission Spectrometry of a Microwave-induced Plasma Used with Laser Ablation of Solid Samples', *Spectrochim. Acta, Part B*, **47**(5), 601–610 (1992).
565. A. Ciocan, J. Uebbing, K. Niemax, 'Analytical Application of the Microwave-induced Plasma Used with Laser Ablation of Solid Samples', *Spectrochim. Acta, Part B*, **47**(5), 611–617 (1992).
566. A. Ciocan, L. Hiddemann, J. Uebbing, K. Niemax, 'Measurement of Trace Elements in Ceramic and Quartz by Laser Ablation Microwave-induced Plasma Atomic-emission Spectrometry. Plenary Lecture', *J. Anal. At. Spectrom.*, **8**(2), 273–278 (1993).
567. J.M. Gehlhausen, J.W. Carnahan, 'Simultaneous Determination of Elemental Ratios in Coal by Direct Powder Injection into a Helium Microwave-induced Plasma', *Anal. Chem.*, **63**(21), 2430–2434 (1991).

Chemometrics

Steven D. Brown
University of Delaware, Newark, USA

1 **Introduction** 1
2 **Soft Modeling in Latent Variables** 2
3 **Current Directions in Chemometrics** 4
4 **Multivariate (First-order) Calibration** 4
5 **Second- and Higher-order Calibration** 5
6 **Nonlinear Methods** 5
7 **Semi-quantitative Methods using Soft Modeling** 5
8 **Classification and Clustering of Chemical Data** 6
9 **Signal Processing Methods** 6
10 **Classical and Nonclassical Optimization Methods** 7
11 **Key Reviews of the Field of Chemometrics** 7
Abbreviations and Acronyms 7
References 7

1 INTRODUCTION

The mathematical and statistical analysis of chemical data is generally discussed collectively under the term ''chemometrics.'' Chemometrics is an interfacial field, lying between the more established fields of chemistry, chemical engineering, statistics, electrical engineering, and computer science. Like many interfaces, there is uncertainty in the exact location of the borders. Many of the borders between chemometrics and the other disciplines are unclear or even disputed, and there are even differences of opinion on the correct pronunciation of the word ''chemometrics,'' but most agree that the goals of chemometrics include the development of mathematical and statistical methods for extracting and representing chemically relevant information from chemical data. The most important applications of chemometrics come in the discovery of the quantitative relationships between chemical measurements and properties and the qualitative modeling and classification of the behavior of chemical systems. A sizeable fraction of the applications of chemometrics have been reported in analytical chemistry and in the closely related field of process analytical chemistry, where the main goal is often the quantitative estimation of chemical concentrations of mixtures from multivariate physical or chemical measurements obtained on those mixtures. Applications of chemometrics to problems in the fields of chemical and electrical engineering are far less numerous, and most of these have dealt with applications of chemometric methods in connection with process monitoring and control or for signal processing.

The field of chemometrics has a similarly uncertain beginning. Several prominent workers in chemometrics, when interviewed by Geladi and Esbensen,[1,2] reported a wide range of dates and people as being the first to do what is now called chemometrics. The dates cited by these workers ranged from about 1976 to as early as 1908. The early date cited here is important, because it focuses our attention on the fact that researchers have done chemometrics from the point when substantial chemical data began to become available. Gossett published his work for Guinness in 1908. Youden and Wernimont did data analysis and wrote articles and books on quality control in the 1940s and 1950s. There was some interest in their work in chemistry for a good while, but much of that died away by the early 1960s. Box, now a famous statistician, also started as a research chemist who got involved with analysis of plant data. Substantial support by chemists for analysis of chemical data as a sub-field of chemistry failed to occur during these times, and interest died out in chemistry. The individual workers doing such analyses mostly went in the direction of statistics because there they found numbers, interest and acceptance. Statisticians then were a relatively small group too, but they were vastly more numerous than the chemical data analysis community, and they had enough desire to broaden their field that these more applied people were welcomed.[3]

Only in the late 1960s and early 1970s, when automated data collection and computation had both gotten much easier, did the need in chemistry again arise for the field. Factors that led to the formation of the field were the press of too much data, the increasing availability of improved mainframe computers with new statistical software packages, the changing attitudes in the statistics and analytical chemistry communities, and the need for answers.[4] A few researchers realized very early that mathematical methods for finding the relations between measured variables and relations between measurements and latent variables could help in finding the information embedded in large amounts of data. These ideas were revolutionary, in that they offered completely new ways of examining the complex, multivariate data sets that were beginning to appear in chemical studies. Because these methods provided useful answers to questions posed by chemists studying large data sets, they attracted a great deal of interest. Very quickly, a sizeable number of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

researchers were exploring the limits of this new approach to data analysis, and the field now known as chemometrics came to be. The field's very name, the publication of the early Chemometrics Newsletter and the formation of the Society for Chemometrics were early indicators of the eagerness with which the field was accepted at first. Howery and Hirsch[3] have briefly summarized the early days of chemometrics.

Changes to the fields of analytical chemistry and statistics also helped shape the new sub-field's goals and make-up. The huge increase in the number and sophistication of chemical instruments made available the chemical data that drives chemometrics. It also changed analytical chemistry into an instrument-oriented field, with more emphasis on the physical chemistry of measurements and less on the classical elements of equilibrium theory. Analytical chemists went from being practitioners of the meticulous art of classical analysis, where the quirks of solution-phase chemical reactions were either used or avoided, to specialist workers focusing on the construction and use of a single, increasingly complex instrument that, in ideal cases, made measurement of some analyte species concentration simpler and more sensitive through the application of ideas from physical chemistry, as well as physics and engineering.[5] In that process, these chemists subscribed – perhaps unknowingly – to the idea of latent relationships in data. Finding these latent relationships in data is often the goal of a chemometric study.

The increased sophistication of the computer also changed statistics, though not as much as might have been expected. The field of statistics has always had an uneasy time with its close relationship with mathematics. The 1930s and 1940s saw rapid progress in the field of statistics. Mathematical statisticians made great progress by using properties of distributions to solve a wide range of problems concerned with classical modeling. The more applied statisticians tended to take a Bayesian approach to problems when developing their applications of these newer methods in the many areas where data analysis was applied. The analysis of mostly chemical data from the medical sciences was an area of particular focus because a large amount of univariate data was routinely available. Economics and psychology, two areas whose data sets also involve latent relationships, but with far fewer variables than were routinely being generated by the new chemical instruments, were, like chemometrics, left to those with an interest in nonclassical modeling methods.

2 SOFT MODELING IN LATENT VARIABLES

Many of the methods employed in chemometrics are based on the concept of soft modeling, a linear modeling method that originated in the field of multivariate statistical analysis but which has become synonymous with the term ''chemometrics.'' The focus of the soft modeling method on the properties of the signal rather than on the noise help to distinguish chemometrics from statistics, where the emphasis is usually on the structure and properties of the error term. Chemists often confuse the two fields. It is worthy of note that, while chemometrics brought many statistical concepts to chemical applications early in its history, in recent years the transfer of information and technology has been from chemometrics to other fields, including statistics. Soft modeling is an area where advances from chemometrics are now aiding a wide range of diverse fields, from remote sensing to psychology.

Because of the heavy emphasis on soft modeling in chemometrics, the field developed around an algorithmic rather than theoretical framework, an attribute that is only now beginning to change. Discussions of soft modeling in current literature are more likely to focus on the linear algebraic theory of the modeling than on the specific steps needed to form the model.

To understand the field's present status and direction, it is necessary to have an appreciation for some of the key approaches and assumptions of soft modeling, as these underlie the logic of many of the chemometric methods.[6] More details are available in the Encyclopedia article **Soft Modeling of Analytical Data** to follow, so it is appropriate to provide only an overview here.

Traditionally, modeling in chemistry and chemical engineering has been done using ''first-principles'' (hard) models. A hard model is one that describes the system in terms of mathematical relationships developed using the measurement variables as independent variables and the desired outputs as dependent variables. Because chemical systems studied are complex, the hard modeling used in chemistry has either been applied to simplified systems or has involved either limiting ''laws'' or other approximations and restrictions to the region of application of the hard model. Soft modeling sees the problem from an entirely different logical perspective: it presumes that the chemical system under study is complex, and that it is not possible or economically feasible to adequately describe the behavior of the system using a hard model. The soft model is based on variation and correlation in the data, as captured in a covariance matrix – which can be thought of as a measure of the overall fluctuation in each independent variable present in the data set, as well as the variable–variable interactions.

A set of orthogonal components made from linear combinations of the independent variables is created to describe independent sources of the observed variation in the covariance matrix created from the data set analyzed,

according to Equation (1)

$$\mathbf{X} = \mathbf{U}\mathbf{V}^{\mathrm{T}} \quad (1)$$

where matrix **V**, the loadings of **X**, contains the linear combinations of the original measurement variables that define the new variation-based coordinate system spanning the data in **X**, and matrix **U**, called the scores of **X**, contains the coordinates of the data **X** in that variation-based coordinate system.[7]

These linear combinations of the measured variables are called latent variables because they are derived rather than measured. The latent variables extracted to describe a data set are ordered in terms of the size of the independent sources of variation that they explain: the first latent variable explains the largest independent source of variance in the data, the second latent variable the second largest, and so on, until all variation in the data set is accounted for by one of the linear combinations of the measurements. Generally, re-expression of the data set **X** in terms of latent variables is not useful unless a decision is also made on the *number* of latent variables that are needed to adequately explain the systematic variation in the data **X**.

Any correlation between the dependent variable(s) and these latent variables is captured by means of a regression model. The regression step can be done after the creation of latent variables for the independent variables and the removal of sources of variation believed to be unrelated to the systematic effects under study. Building a regression model by first soft-modeling and truncating the independent (measured) variables is known as principal components regression (PCR) in statistics and chemometrics (or it can be done in concert with the extraction of the latent variables using a modeling method known as partial least squares (PLS) regression).

The strength of soft modeling with latent variables is that systematic sources of variation in the data do not go unmodeled, as they might under hard models developed with inadequate theory. Relationships in data can be modeled even when the existing theory is incomplete, wrong, or even missing. The ability to model linear relationships in the absence of suitable theory has made soft modeling an attractive approach for a wide range of problems in applied and analytical chemistry, including modeling of quantitative relationships in spectroscopy, quantitative structure–property relationships, and many other quantitative modeling scenarios. Soft modeling with latent variables also has uses in semi-quantitative modeling, where latent variables can be used to group or to classify samples by some latent property. Examination of the latent variables can often provide insights into the chemistry behind the observations. And, because latent variables are linear combinations of what may be many measurements, they offer a way to screen measurements for appropriateness and impact on the modeling. For this reason, soft modeling in latent variables is often used to discover relationships in data.

Soft modeling comes with several inherent defects, however. The most serious is that the latent variables created in the soft modeling are *local* to the data set analyzed. The soft model developed explains the variation seen in the data set used to create it. Unfortunately, quantitative modeling often requires a predictive model, and there is strong incentive to take a soft model developed on one set of data for use on an entirely different set. Any use of this model on new data is an extrapolation and, unless great care is taken in collecting and analyzing the data, the soft model may have little predictive use. Because the soft model defines a truncated mathematical basis for a series of measurements, it is possible to create a soft model on one set of data and to use it on a closely related set. To make this possible, though, the truncated basis defining the latent variables developed in the soft model must span the bases defining any new set of experimental data subjected to the model. The subtle and not-so-subtle consequences of this requirement force the user of soft models to consider the experimental design of data sets and the sensitivity of the models to individual samples to a degree that goes far beyond that done with more traditional modeling methods.

A second difficulty with soft modeling arises from the fact that the latent variables describe orthogonal mathematical effects, and not physical effects. Most physical effects show high correlation between measurements, especially between adjacent measurement channels (e.g. individual variable ''channels'' in spectra or chromatograms). For example, suppose that in a chemical calibration, the varying amounts of two different compounds with different spectra are responsible for the two different sources of variation in the analytical response. These two chemically distinct sources may not be *mathematically* independent, however, because of the redundant nature of most multichannel responses measured with chemical instrumentation. Described in terms of latent variables, though, the sources must be mathematically independent. The result is the mixing of separate, chemically significant effects in each of the latent variables and the spread of these effects over several latent variables. The latent variables describe *variation* in measurements rather than some easily understood measurement phenomenon, such as the absorbance at a particular wavelength. Direct interpretation of the chemical significance of these latent variables is difficult or impossible. Usually, conversion of mathematically significant effects as evidenced in the latent variables surviving after truncation of the soft model back to physically significant effects is needed to extract chemically relevant

information. In essence, the conversion needed is an oblique rotation of orthogonal basis defined by the latent variables into the nonorthogonal basis defined by the nature of the physically significant information desired. Without external constraints or some prior knowledge of the solution, constraining the large number of possible rotations is impossible in many cases. The conversion of a mathematically useful latent variable set to a chemically relevant basis is therefore a stumbling block in getting usable chemical information from the soft modeling. This fact more than anything limits the use of chemometrics to those secure enough in the mathematics to apply it without many of the usual ''safety lines'' associated with visual interpretation or simple statistical inference (confidence intervals, etc.).

Another defect in soft modeling is the lack of an easy route to distinguish the sources of variation attributable to systematic, chemical effects in the data from other sources of variation. Accidental correlations arising between systematic variation in the data and various kinds of noise present either in the measurements themselves or in the taking of the data can corrupt the latent variables. One consequence of this accidental correlation is the aliasing of measurement variables weighted heavily in the latent variable; it is possible to see increased or decreased weights given measurements in a latent variable as a consequence of noise-induced artifactual correlation with another measurement variable with large variation. These artifacts in the correlation and variation present in the data set frustrate the modeling effort in several ways. First, when present, accidental correlation induced by noise effects can make difficult the direct interpretation of the latent variables, an important part in extracting information from the soft model. Second, the presence of the incorrect weighting of the measurement variables in the latent variables can corrupt the rotation of latent variables back into chemically significant information presented in terms of the original measurement variables. Third, the noise-related artifacts complicate statistical decisions on the number of latent variables needed to describe a systematic effect in data, a key part of the development of the soft model.

3 CURRENT DIRECTIONS IN CHEMOMETRICS

In view of the importance of soft modeling to the study of complex chemical systems, most standard chemometric methods rely heavily on soft modeling in one form or another and make full use of its advantages. Advances in these chemometric methods often come in the development of new algorithms or new ways of using older methods that can either ameliorate or avoid the above-mentioned difficulties associated with soft modeling.

4 MULTIVARIATE (FIRST-ORDER) CALIBRATION

The best-known examples of the use of new algorithms to make use of soft modeling practical come in the area of multivariate calibration. This is the subject of the Encyclopedia article **Multivariate Calibration of Analytical Data**. An overview is provided here.

Here, the PCR algorithm and later the PLS algorithm were modified and put to use in improving the extraction of quantitative information such as concentrations from analytical responses. One modification came in that the soft modeling was done on an inverse relationship between responses and concentration, namely Equation (2)

$$\mathbf{C} = \mathbf{RB} \tag{2}$$

where $\mathbf{C}$ is an $m \times n$ matrix of concentrations, $\mathbf{R}$ is an $m \times p$ matrix of responses, and $\mathbf{B}$ is a $p \times n$ matrix describing the calibration relationship. Equation (2) is an inverse Beer's law relation, as it implies that concentration $\mathbf{C}$ carries the error and is dependent on the response variables in $\mathbf{R}$. By soft modeling the response $\mathbf{R}$ (and $\mathbf{C}$ with PLS), then truncating the soft model to decrease the variance unrelated to the calibration relationship, a useful calibration model results. This model has the benefit of being suited to predict one component in a calibrated mixture.[8]

The most widely known success of first-order calibration methods has come in near-infrared (NIR) spectrometry, a fairly general analytical measurement that is often useless for direct quantitative measurements because of the lack of specificity of NIR bands.[9,10] Efforts to enhance the specificity or selectivity of analytical instrumentation, especially spectroscopic and chromatographic instrumentation, continue to make up a sizable fraction of chemometric research. Multivariate calibration using PLS for the modeling of the inverse response–concentration relationship has become commonplace in the last 10 years, and these techniques have begun to be embraced by industrial chemists and chemical engineers.[11] Published applications of multivariate calibration abound, as is clear from the number of citations in the Fundamental Reviews reported over the last 10 years,[12] but it is the author's experience that many more applications of multivariate calibration are put into practice than are reported. A good, short introduction to PLS theory and coding is also available.[13,14]

The current trend in first-order calibration is to explain the calibration in terms of the net analyte signal.[15,16]

5 SECOND- AND HIGHER-ORDER CALIBRATION

Second-order calibration, where the analytical response gives rise to a matrix rather than a vector of data, is a very active area of research at present. This area is the subject of the Encyclopedia article **Second-order Calibration and Higher**, so again, only an overview is provided here. There are two reasons for the interest. One is practical: a large number of analytical measurements give rise to a matrix of data (e.g. a liquid chromatographic separation monitored with a diode-array detector). There is also a theoretical ''second-order advantage'' in using these data in a calibration, in that the mathematics permits analysis of a calibrated component in the presence of an uncalibrated interference.[17]

The analysis of higher-order chemical data – that is, data with dimension higher than one – involves a choice of the way that the data are unfolded to make matrices of data (where the rows are defined by the individual samples and the columns by the components of the one-dimensional measurements) for more conventional matrix-oriented analysis. Smilde has reviewed different unfolding methods, including the various Tucker unfolding schemes and parallel factor analysis (PARAFAC) modeling, and offered an explanation of the history and applications of higher-order analysis.[18] As in first-order data analysis, early chemometrics work on higher-order data has benefited from porting of the substantial work on three-way methods published in statistics, psychometrics, and related fields. But, as with analysis of first-order data, the maturation of chemometrics research has led to significant contributions by chemometricians to this approach to data analysis.

Like multivariate calibration of first-order data, higher-order analysis is also based on soft modeling. It is no surprise that it carries some of the defects inherent in soft modeling, including sensitivity to noise, difficulty in getting the correct model size (the rank here) and problems in assuring the match of bases. There are some important differences, however: the rank annihilation method deals with information in multiple measurement dimensions, and it is far less susceptible to the effect of an interference that shows in a prediction sample but not in any calibration sample. Higher-order calibration is very dependent on a good match of the calibration model with the response data to be determined.

6 NONLINEAR METHODS

The history of measurement chemistry is filled with studies of linear systems and of efforts to make those nonlinear systems studied linear enough for examination by linear methods. It is therefore not surprising that relatively little effort has been placed on the systematic study of nonlinear methods as applied to chemical calibrations. The large majority of studies of nonlinear systems have been done using nonlinear causal models, because methods of nonlinear soft modeling developed to date require the underlying presence of specific nonlinear functional forms (e.g. quadratic relationships for loadings) and have not proved useful in practice.

Recent interest in soft modeling nonlinear relationships has centered on the use of neural networks. Artificial neural networks (ANNs), a modeling method that does not presume any underlying relationship in the data, have received some attention for the calibration of nonlinear chemical systems and have been used for modeling a number of chemical processes with success. An ANN models the relationship between the input independent variables and the output dependent variables by using weighted sums of transforms usually based on a sigmoid function or the tanh function, a close relative. These transforms are capable of describing very complex relationships between the independent and the dependent variables. According to theory, a three-layer net with sigmoid transfer function in the hidden layer and linear transfer function in the output layer can model any continuous function.[19] The modeling is done by first establishing the topology of the network model, i.e. number of layers of nodes to be used, the number of nodes in each layer, and the transform operation used in each layer. Finding an appropriate topology is a fairly straightforward but time-consuming process. Generally, a number of possible candidate topologies are tried – these can be found by simple enumeration or by more sophisticated methods such as genetic algorithms – and the topologies giving the best results are kept for further analysis and optimization.

7 SEMI-QUANTITATIVE METHODS USING SOFT MODELING

In many studies, the goal is not quantitative, indirect measurement of one or more known, calibrated chemical species in a series of mixtures. The goal may be discovery of what chemical species are involved in a dynamic system, or possibly discovery of the number of species changing in that system. It may just be an objective measure that the dynamic system has not changed in its overall make-up. Chemometric methods based on soft-modeling of the

multivariate sensor measurements taken on these systems are well suited to obtaining information on the identity and number of chemical species.

When less is known about a data set, self-modeling can be attempted. The loadings used to describe a set of data are rotated to find another set of axes that has physical significance. In this case, less may be known about the composition of the system under study, and the number of possible rotations is infinite. External constraints such as all positive spectra and all positive amounts of species are needed to help limit the search to the space between the convex hulls defined by the different constraints. It is now known that the solutions are found near the apexes of the convex hulls, but there still is no systematic means to map out convex hulls and locate their apexes in the presence of noise and confounding effects. SIMPLISMA, one popular method for self-modeling, uses a visual examination to help find suitable rotations of the soft model.[20] The solutions that result from this self-modeling are subject to uncertainty unless each species has a signature that is independent of the others at one or more of the sensors used to collect the multivariate set. The more similar the species' true responses over the sensor set used, the larger the uncertainty in the self-modeled responses for these species will tend to be.

Often, self-modeling is needed because of a lack of solid information on the composition of the system, but the data are collected in such a way that ensures their serial correlation in both the sensor axis and another axis, often time. In these cases, the additional correlation offers an additional constraint that helps to remove ambiguity in the rotated solution. Now, it is possible to develop a series of soft models that evolve as the data evolves in time, in both the forward (increasing time) and backward (decreasing time) directions. Tracking of the number of significant principal components in the soft models over the time axis permits an estimate of the number of species that vary over the data set, as well as where it is in the data that the different species enter into the chemical system's response. Rotation of the soft model to simultaneously describe both the response and the time-dependent data axes, subject to constraints in spectra and amounts, gives semiquantitative estimates of pure responses and relative amounts. Malinowski reviewed early work showing that this modeling approach can be used to discover relationships in data.[21]

8 CLASSIFICATION AND CLUSTERING OF CHEMICAL DATA

It may be that there is grouping of the samples in the data according to some latent property. Often, a set of data is heterogeneous, in that the samples clump together rather than spread evenly over the multivariate space. Building a single PLS regression model on the full set of multivariate data is inappropriate because the variation captured in the modeling will be heavily influenced by variation between groups of data and will be less influenced by the variation between the independent and dependent variables. In these cases, a grouping step is needed before regression modeling to ensure that a homogeneous structure is present in some subset of the data prior to modeling this portion of the data. If a multivariate grouping structure present in the data can be guessed or determined, separate soft models can be developed for each group in the data.

Once grouping is decided, separate soft models can be built for each of the groups. Soft modeling can be used to develop class regression models for different values of the latent property. New data can be classified into the appropriate group on the basis of the distance of the new data to each of the class models. An F-test is used to decide on membership, again using residual variance in the fit and in the model. The SIMCA method focuses on modeling the classes rather than on finding an optimal classifier.[22,23] By examining the mathematics behind SIMCA, it has become apparent that there may be better classifiers, depending on the structure of the data.[24,25]

Current research in soft modeling for semiquantitative modeling has as an emphasis the removal of many of the problems associated with soft modeling that carry over to self-modeling. Making the evolving factor analysis methods work more reliably is a major focus of research, as are improved methods for getting self-modeling to work on many systems. Background and current research in classification methods are discussed in the Encyclopedia article **Clustering and Classification of Analytical Data**.

9 SIGNAL PROCESSING METHODS

Since the initial days of chemometrics, signal processing has long been a focus. Initially, the aim was noise reduction through smoothing off-line, but in recent times, the aim of signal processing has focused more on real-time data enhancement through filtering or on real-time control of a process through use of real-time data. Details are available in the Encyclopedia article **Signal Processing in Analytical Chemistry**. As this area has evolved, the interest of analytical chemists has waned while interest from chemical engineers and process chemists has increased substantially.[26,27]

10 CLASSICAL AND NONCLASSICAL OPTIMIZATION METHODS

Chemometrics is mostly focused on data analysis but one aim has been the improvement of a process or system producing the data. Chemometrics has long been involved in process optimization, from the evolutionary optimization schemes of Box to the Simplex optimization methods perfected by Deming. Newer work on genetic algorithms and simulated annealing continue the advancement of optimization methods. The Encyclopedia article **Classical and Nonclassical Optimization Methods** discusses these optimizers.

11 KEY REVIEWS OF THE FIELD OF CHEMOMETRICS

Because of its early, close association with the field of analytical chemistry, the field of chemometrics has received frequent and thorough review in that field's literature. Beginning in 1976, a detailed review of the methodology and practice of data analysis in chemistry has appeared in the biennial Fundamental Reviews issue of the journal *Analytical Chemistry* (the most recent as of this writing is that of Lavine[12]). These reviews provide a comprehensive analysis of the state of chemometrics and, when examined together, also provide a glimpse at the evolution of concepts in chemometrics over the past 10 years. Along with the reviews of basic research in chemometrics, three major reviews on chemometrics applied to problems arising in process analytical chemistry also have appeared in the Applications Reviews issue of *Analytical Chemistry*.

ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Network
NIR	Near-infrared
PCR	Principal Components Regression
PLS	Partial Least Squares

REFERENCES

1. P. Geladi, K. Esbensen, 'The Start and Early History of Chemometrics: Selected Interviews', *J. Chemometrics*, **4**, 337–354 (1990).
2. P. Geladi, K. Esbensen, 'The Start and Early History of Chemometrics: Selected Interviews. Part 2', *J. Chemometrics*, **4**, 389–412 (1990).
3. D.G. Howery, R.L. Hirsch, 'Teaching Chemometrics', *J. Chem. Educ.*, **60**, 656–659 (1983).
4. S.A. Borman, 'Math is Cheaper than Physics', *Anal. Chem.*, **54**, 1379A–1380A (1982).
5. B.G.M. Vandeginste, 'Teaching Chemometrics', *Anal. Chim. Acta*, **150**, 199–206 (1983).
6. S. Wold, 'Chemometrics: What do we mean with it, and what do we want from it?', *Chemom. Intell. Lab. Syst.*, **30**, 109–115 (1995).
7. J.E. Jackson, *A User's Guide to Principal Components*, Wiley-Interscience, New York, 1991.
8. P.J. Brown, 'Multivariate Calibration (with Discussion)', *J. Roy Stat. Soc., Ser. B*, **44**, 287–321 (1982).
9. H. Martens, T. Karstang, T. Næs, 'Improved Selectivity in Spectroscopy by Multivariate Calibration', *J. Chemometrics*, **1**, 201–219 (1987).
10. T. Næs, H. Martens, 'Principal Component Regression in NIR Analysis: Viewpoints, Background Details and Selection of Components', *J. Chemometrics*, **2**, 155–168 (1988).
11. H. Martens, T. Næs, *Multivariate Calibration*, John Wiley & Sons, Chichester, 1989.
12. B.K. Lavine, 'Chemometrics: Fundamental Review', *Anal. Chem.*, **70**, 209R–218R (1998).
13. P. Geladi, B.R. Kowalski, 'Partial Least Squares Regression: A Tutorial', *Anal. Chim. Acta*, **185**, 1–17 (1986).
14. P. Geladi, B.R. Kowalski, 'An example of 2-Block Predictive Partial Least Squares Regression with Simulated Data', *Anal. Chim. Acta*, **185**, 19–32 (1986).
15. K.S. Booksh, B.R. Kowalski, 'Theory of Analytical Chemistry', *Anal. Chem.*, **66**, 782A–791A (1994).
16. K. Faber, A. Lorber, B.R. Kowalski, 'Net Analyte Signal Calculation in Multivariate Calibration', *Anal. Chem.*, **69**, 1620–1626 (1997).
17. B.R. Kowalski, M.B. Seasholz, 'Recent Developments in Multivariate Calibration', *J. Chemometrics*, **5**, 129–146 (1992).
18. A.K. Smilde, 'Three-way Analyses: Problems and Prospects', *Chemom. Intell. Lab. Syst.*, **15**, 143–157 (1992).
19. G. Cybenko, 'Approximation by Superpositions of a Sigmoidal Function', *Math. Control Signals Systems*, **2**, 303–314 (1989).
20. W.Windig, 'Mixture Analysis of Spectral Data by Multivariate Methods', *Chemom. Intell. Lab. Syst.*, **4**, 201–213 (1988).
21. E.R. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, Wiley-Interscience, New York, 1991.
22. S. Wold, 'Pattern Recognition by Means of Disjoint Principal Component Models', *Patt. Recog.*, **8**, 127–139 (1976).
23. S. Wold, C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjöström, 'Multivariate Data Analysis in Chemistry', in *Chemometrics: Mathematics and Statistics in Chemistry*, ed. B.R. Kowalski, NATO ASI Series, Reidel, Dordrecht, 17–95, Vol. 138, 1984.

24. I.E. Frank, J.H. Friedman, 'Classification: Oldtimers and Newcomers', *J. Chemometrics*, **3**, 463–475 (1989).
25. G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley-Interscience, New York, 1992.
26. J.F. MacGregor, 'Statistical Process Control of Multivariable Processes', *Control Eng. Practice*, **3**, 403–414 (1995).
27. C.E. Miller, 'The Use of Chemometric Techniques in Process Analytical Method Development and Operation', *Chemom. Intell. Lab. Syst.*, **30**, 11–22 (1995).

Classical and Nonclassical Optimization Methods

Ron Wehrens and Lutgarde M.C. Buydens
University of Nijmegen, The Netherlands

1 Introduction **1**
1.1 Local and Global Optimality 2
1.2 Problem Types 2
1.3 Example Problem: Fitting Laser-induced Fluorescence Spectra 3
1.4 Criteria for Optimization 4
1.5 Multicriteria Optimization 4
2 Strong Methods **4**
2.1 Gradient-based Optimization 5
2.2 Response Surface Methods 5
2.3 Simplex Methods 6
3 Weak Methods **7**
3.1 Random Search 7
3.2 Sampling Methods 7
4 Intermediate Methods **8**
4.1 Monte Carlo Simulated Annealing 8
4.2 Tabu Search 9
4.3 Evolutionary Optimization 9
4.4 Hybrid Methods 10
4.5 Comparison of Methods 10
5 Conclusions and Outlook **11**
Abbreviations and Acronyms **11**
Related Articles **11**
References **11**

Optimization problems are abundant in analytical chemistry, examples being the determination of optimal conditions for experiments or optimal settings for instruments. In general, all the required information should be obtained from as few experiments as possible. Classical techniques such as response surface models or simplex optimization are often used. These techniques, which can be very efficient in cases where the underlying assumptions are fulfilled, are called ''strong'' methods.

With the advent of the computer in the laboratory, a new class of optimization problems arose which could not be tackled with the standard methodologies. For these search-type problems, new strategies such as simulated annealing (SA) and genetic algorithms (GA) are applied. Although these are not guaranteed to give the optimal result, in almost all cases they are able to find very good solutions where other techniques fail completely. These methods find themselves in an intermediate position between the strong methods, mentioned above, and weak methods, where hardly any assumptions are made.

This article provides an overview of classical and nonclassical optimization techniques, and stresses the differences in their areas of application. The key ideas are highlighted and references to important publications are given.

1 INTRODUCTION

As in all empirical sciences, optimization problems are abundant in chemistry, and have been since the very first alchemist experiments. In the synthesis of gold from lesser substances, every conceivable variation of ingredients and experimental conditions was tried and eventually rejected. Sometimes these experiments were conducted in a very systematic way, sometimes random, sometimes somewhere in between. Although the main goal was never achieved, a lot was learned from these early experiments.

In the modern laboratory, the advent of complicated instruments controlled by computers has changed the nature of the problems faced by the chemist. Whereas the actual doing of the experiment used to be a major limiting factor in obtaining relevant information, and therefore warranted a large investment in time and energy to be designed and executed, in the modern laboratory experiments can be performed in a fast, reproducible way. For the sake of ever higher performance requirements, dictated by economic reasons, systematic optimizations are performed in all chemical laboratories, often consisting of hundreds or thousands of experiments.

Optimization is the search for a maximum or minimum in the value of a certain response function. For example, the yield of a chemical reaction is a function of several variables, such as concentrations of components in the mixture and physical characteristics such as temperature and pressure. The relationship between these variables and the response function (the yield) is unknown, but it can be sampled by performing experiments. The conditions in these experiments can be carefully selected by using knowledge (section 2.2) and, depending on the outcome, new experiments may be performed.

Such optimizations are typically performed in an iterative fashion (Figure 1). The search is started either from one or more random positions or from a set of points, picked according to some criterion. In the evaluation stage, the quality of the current point(s) is assessed by experiments. Several criteria may be used to stop the optimization, such as the number of experiments, or the quality of the solutions found so far. If no stopping

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

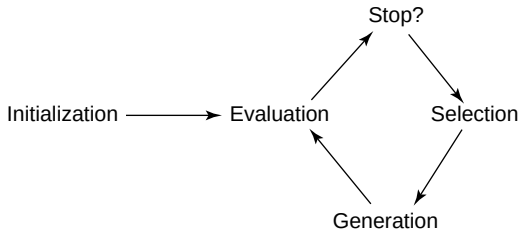

Figure 1 The basic iterative optimization cycle.

criterion applies, the next stage is to accept the solution(s) as a starting point for new solutions, or to reject and to proceed from another solution. This process is called selection. Finally, the optimization method generates one or more new candidate solutions. These in turn are evaluated and the cycle enters its next iteration.

Many different optimization methods exist and they can be classified in a number of ways. Among the classical methods[1] are methods relying on a model of the response surface. The sampling of the space of possible solutions is meant either to describe this surface or to define the direction in which the largest improvement will be found. Typical examples are steepest-descent and response-surface methods. Many assumptions about the relationship between problem parameters and response are made, but if they are correct then the performance of these methods is very good. These ''strong'' methods are treated in section 2. The other category of classical optimization methods consists of the weak methods. These methods (section 3) make no assumptions about the response surface, and include the well-known but rarely used random search technique.

The newest optimization techniques, i.e. the nonclassical techniques, place themselves between these two extremes. Whereas weak methods rely on chance to find the optimum and strong methods rely on problem-specific assumptions, these ''intermediate'' methods seek a balance between chance and information. As they typically require a large number of experiments, they are often applied in computational optimization problems. The most well-known method is SA.[2]

The differences between the methods discussed here and the problems for which they are applied can be found at all levels of the basic cycle depicted in Figure 1. Whereas weak and intermediate methods typically rely on random choices for the starting points, strong methods utilize knowledge about the problem domain to identify promising regions. Some methods stop as soon as no improvement is found in the last step, whereas others just proceed with the best move available. Almost all methods use a quality-weighted selection – bad solutions are more often rejected than good solutions, but traditionally each optimization method uses its own particular selection scheme. Finally, in the generation of new trial solutions, strong methods use moves defined a priori by the user; weak methods randomly generate the next trial set; and intermediate methods usually apply a method-specific and/or problem-specific mechanism.

In cases where the response is an error value, as in curve-fitting applications, the optimal value is zero and all other response values are larger than zero. In many other applications the response variable is maximized, such as the yield of a chemical reaction. This article does not distinguish between these two cases. The examples given always relate to minimization problems without loss of generality.

1.1 Local and Global Optimality

To clarify the concepts of local and global optimality, the map of a mountain landscape, in which lines indicate regions of similar response, is often used as a metaphor. The peak of a mountain always is a local optimum, because in its direct neighborhood there are no higher places. However, only the peak of the highest mountain is the global optimum. In many optimization problems, finding the global optimum is the challenge, and the presence of many local optima complicates the problem significantly. In the same vein, local optimizers are methods that always find the (usually local) optimum near the starting position, and global optimizers are methods that end up at the highest peak, no matter where the search started from.

Many methods utilize a sense of direction in which to proceed. Steepest ascent/descent methods (discussed below) choose the direction in which most short-term gains can be made. These invariably end up at the nearest local optimum, and are therefore very bad global optimizers. However, techniques that have a better search behavior can also be led astray. Problems in which the global optimum is located in another area, as suggested by the slope of the landscape, are described as deceptive, and much research has been devoted to these. It should be noted that no method guarantees finding the global solution in finite time.

1.2 Problem Types

Several different problem types can be distinguished. By far the most common in analytical chemistry is the numerical problem, where for a number of parameters optimal values must be found. These parameters might be settings for a chromatograph or another instrument, or coefficients in a nonlinear equation describing a curve. The parameters may interact, or may be constrained to certain regions, and the resulting problem landscape may therefore be quite complex.

The second type of problem is formed by combinatorial problems. The complexity of this kind of problem often

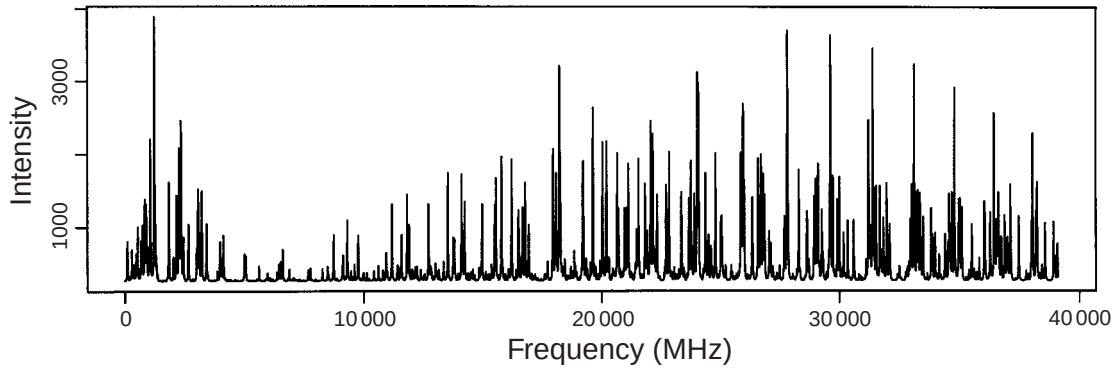

Figure 2 Part of the LIF spectrum of benzimidazole (22 797 data points).

is huge; most are nonpolynomial (NP)-complete, an indication meaning that when the size of the problem grows, the time needed to evaluate all possible solutions increases faster than the polynomial. In practice this means that even for small problems evaluating all possibilities is completely impossible. This also implies that only in special cases it is possible to know whether an optimum is truly a global optimum. Two basic forms can be distinguished: nonordered and ordered problems. A well-known example of nonordered problems is subset selection, which is a common issue in calibration and classification procedures. The problem relates to which subset of all available variables should be included in the statistical model. Not only does a small number of variables in model building lead to more robust and parsimonious models, it also may decrease experimental effort. Examples include wavelength selection from IR (infrared) spectra and variable selection in quantitative structure–activity relationship (QSAR) modeling, where a small set of relevant descriptors is to be selected from a potentially very large list.

Ordered combinatorial problems are also called sequential or scheduling problems. They are not so common in (analytical) chemistry but of tremendous importance in operations research, for example. The question is to put a set of objects in the optimal order, or in more than one dimension, at the optimal location. The traveling salesperson problem is a well-known example; microchip design and scheduling of analyses in a routine laboratory belong to this category. Just as problems from the previous class, sequential problems are more often than not NP-complete.

1.3 Example Problem: Fitting Laser-induced Fluorescence Spectra

An example used throughout this article concerns the description of laser-induced fluorescence (LIF) spectra of organic molecules.[3] This kind of spectroscopy can provide much information on the structure and dynamics of molecules and molecular complexes in ground states as well as excited states. The experimental spectrum of benzimidazole is shown in Figure 2.

Hundreds of well-resolved peaks can be seen, whose positions and intensities depend on a small number of parameters. Some of these parameters are related to experimental conditions such as temperature; others are related to quantum-chemical properties such as rotational constants. For small molecules, the spectra can be interpreted and in this way a set of valid parameters is found. However, for larger molecules the number of possibilities quickly becomes too large and one is forced to resort to search methods. For a given set of parameters the theoretical spectrum for a molecule, can be calculated, and so the agreement between theory and experiment can be used to assess the validity of the parameter set. The

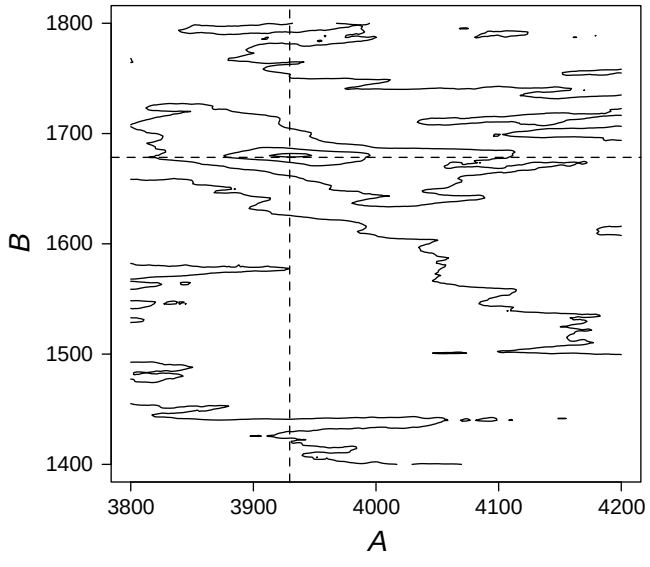

Figure 3 True response surface showing the RMS error between experimental and simulated LIF spectra of benzimidazole, depending on two parameters A and B. The true values for A and B are indicated by the dotted lines.

goal of optimization is to find parameters that yield a good approximation to the experimental spectrum. The difference is expressed as the root mean square (RMS) error.

In Figure 3 the RMS response surface is plotted if two parameters, the rotational constants around the two main axes, A and B, are systematically varied. All other parameters are kept constant at their true levels. This response surface is used throughout the article. Note that the set-up where only A and B are optimized is only a toy problem, although the contour lines are not very smooth. The real, high-dimensional optimization problem is much more difficult.

1.4 Criteria for Optimization

In an optimization problem a minimum or a maximum value is sought for a response variable as a function of problem parameters that can be controlled by the user. The extent to which these parameters can be controlled determines the precision of the outcome. It is of course senseless to try and optimize beyond experimental error; nevertheless, it is easy to forget the obvious.

Whereas the previous considerations concerned noise in the response value, there is also the case where the parameters can be influenced by noise. Moreover, if a small change in optimal parameter settings (e.g. noise) leads to a drastic decrease in product quality, it may be useful to look not for the optimal solution but for solutions which are acceptable and not so sensitive to small deviations from optimality. In an industrial context robust and good solutions are often of greater interest than the very best solution, which may be unstable.

In other cases solutions proposed by an optimization algorithm may be physically impossible or economically undesirable. These constraints are not always easy to define. If that is the case, one alternative is to generate several solutions and pick from those the ones that least violate the boundary conditions.

Good optimization procedures should yield the desired information in a fast, reproducible, and reliable way. This means that repeated application should yield more or less the same answers or answers of the same quality. In some cases it may be important that all acceptable solutions are found, in other cases it is more important that the global optimum is found or that at least one good solution is found within time or cost constraints. These aspects are to some extent determined by the problem at hand.

1.5 Multicriteria Optimization

If additional constraints can be defined, often in terms of costs, it may be a good idea to include them in the response from the experiments. There are several ways of doing this, the simplest of which is a weighted sum of several response variables. However, the weights need to be set to realistic values and in some cases this requires quite a bit of fine tuning. Another approach is to consider several response variables at a time. One solution then dominates another if it is better at one or more of the response variables while not being worse for all other criteria. This is called Pareto optimality, after the Italian economist Vilfredo Pareto. The result is a set of solutions that dominate other solutions at least for a number of variables. The user than has to make the trade-off between optimality in one variable versus optimality in the other. Especially with a small number of responses it may be that a useful Pareto-optimal set of solutions can be found.

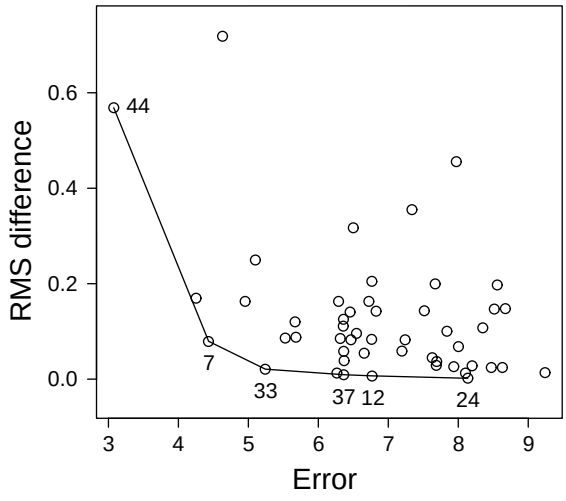

Figure 4 Pareto-optimality. Solutions 44, 7, 33, 37, 12, and 24 are on the Pareto-optimal front, where attention shifts from a small error to a smaller difference with neighboring points.

Optimization problems where the robustness of the solution is important form a good example of the possibilities of multicriteria optimization. Then one criterion is the quality of the optimum and the other the sensitivity of the response to small changes in the parameters (Figure 4). Fifty random samples were taken from the contour surface of Figure 3, and the RMS error (x-axis) is plotted against the difference in RMS value with the surrounding points (y-axis). The smaller this RMS difference, the more robust the solution. A front of Pareto-optimal solutions delimits a cloud of possible solutions. Note that the point with the lowest error (number 44) is not very robust, so in cases where robustness is important solution 7 might be a better alternative.

2 STRONG METHODS

Strong methods assume a certain structure in the solution space, such as a quadratic response surface or only

one optimum. If the assumptions are correct, these methods are fast and reliable. The main problem is that if the assumptions are not correct, these methods will not find the global optimum even if the number of experiments is increased drastically. In difficult optimization problems they are often used for the last part of the optimization, where the approximate global optimum has been identified and the exact location still must be found. Especially in low-dimensional optimization problems these methods are often very suitable. Examples include optimization in organic synthesis[4,5] and method development in high-performance liquid chromatography (HPLC).[6,7]

2.1 Gradient-based Optimization

The most intuitive methods of this class rely on an accurate estimate of the local gradient, and proceeding in the steepest direction. Depending on whether we are dealing with a minimization or maximization problem, the method is called "steepest descent" or "steepest ascent", respectively.

The simplest and rather naive form is to assess the quality of neighboring points and, if there are better points, to proceed towards the best of these. Then, new neighboring points are evaluated and this process continues until there is no neighbor better than the current point. This process is depicted in Figure 5, using the error surface from Figure 3. Clearly, only from one of the 16 starting points the global optimum is located. The other starting points lead to (sometimes even quite bad) local optima. Each optimum has its own basin of attraction, indicating that any steepest descent search starting from a point within that basin will end up at that particular optimum. This, of course, is the definition of a local search method.

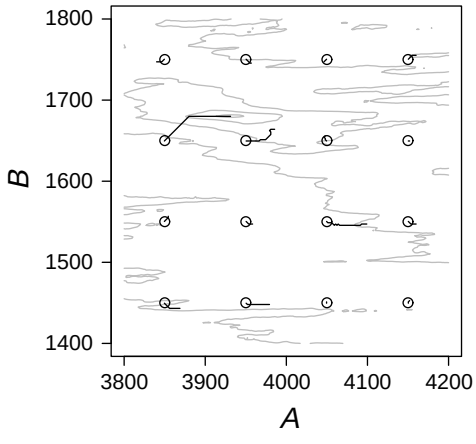

Figure 5 Gradient optimization (naive form) starting from 16 points. The global optimum was reached in 52 steps from the starting point at (3850, 1650). The mean number of steps before the gradient optimizations stopped was just above 12.

In some cases, especially those where the landscape being searched looks like a long narrow valley (in the case of minimization), steepest descent does very poorly. Unless the valley is approached at a perfect right-angle, it takes many small steps to reach the optimum. The more sophisticated conjugate gradient (CG) methods[8] determine the new direction based not solely on the current gradient but on old gradients as well. This leads to a much better search behavior.

Inclusion of the second derivative (the Hessian matrix) is often used in curve-fit problems, or nonlinear least-squares routines. The quantity that is minimized must follow a χ^2 distribution, which is often the case with errors from a fit. Calculation of the Hessian is possible precisely because of the assumption of the χ^2 distribution. The standard method, which combines steepest descent and Hessian approaches, is the Marquardt–Levenberg method,[8] using the former far from the optimum and the latter in the neighborhood of the optimum.

2.2 Response Surface Methods

Response surface methods stem from the area of experimental design,[4,9] where in as few experiments as possible the maximum amount of information is extracted from a system. It is assumed that the response surface can be parameterized by a simple function with one optimum. The aim is to estimate the response function surface by choosing the experimental conditions intelligently. An often-used set-up is central composite design (CCD) depicted in Figure 6(a). Four points, arranged in a square, are combined with five points in the form of a star. In cases where the response is subject to error, one or more points (often the central point) may be replicated to obtain an estimate of the standard deviation. The fitted surface, using all terms to second degree, is given by Equation (1):

$$z = 433.7 - 0.061x - 0.371y - 1.1210^{-5}x^2 - 3.8810^{-6}y^2 + 9.4010^{-5}xy \quad (1)$$

where z is the error, x is A and y is B. The resulting surface is depicted in Figure 6(b).

This figure also clearly demonstrates the danger of extrapolating. According to the model, the optimal values are to be found in the upper left or lower right corners. One should only use these designs in the neighborhood of an optimum where one can be reasonably sure that the surface behaves as expected. Even very close to the global optimum this may not be the case.

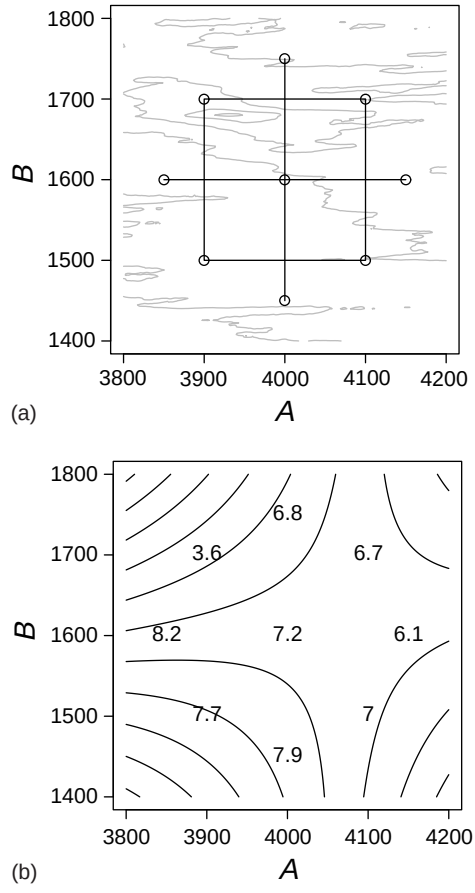

Figure 6 (a) Application of a CCD to the response surface of Figure 3. (b) Contour lines of the fitted surface; error values at the experimental points are indicated. Optimal values according to the model are the upper left and lower right corners.

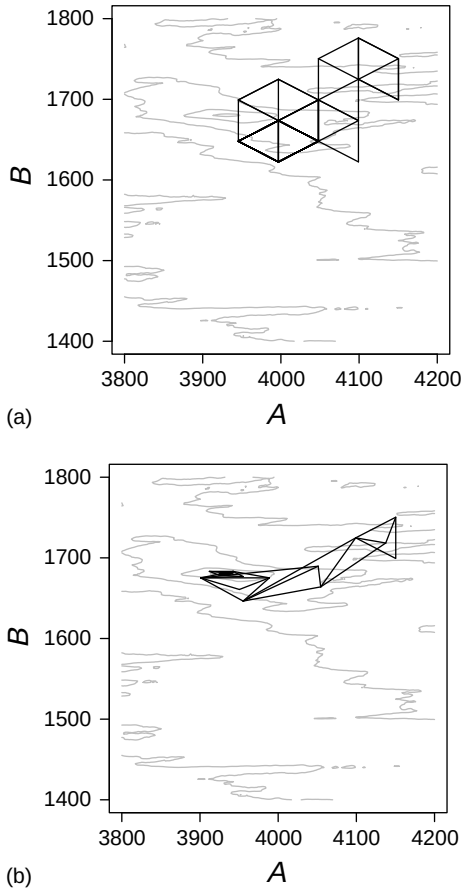

Figure 7 (a) Simplex method and (b) modified simplex method.

2.3 Simplex Methods

Repeated application of small experimental designs can lead to the accurate and precise estimate of the optimum. Although sampling points can be reused, significantly decreasing the experimental effort, most response surface methods become impractical in spaces with more than three or four dimensions.

A simple extension that is more practical in spaces of moderate dimension is the simplex method.[10] Instead of picking points in the search space according to an experimental design, one starts with the smallest number of experiments defining a surface in space; for example, for a two-dimensional space three points are chosen or, for a three-dimensional space, four points are chosen. This set of experiments, or points, is called the initial simplex. Next, one iterates by repeatedly selecting the worst point of the simplex and replacing it with a new point, obtained by mirroring the worst point in the side defined by the remaining points.

In two dimensions, this yields a series of adjoining triangles, as depicted in Figure 7(a), where the simplex method is applied to the error surface of Figure 3. The simplex is able to locate the global optimum approximately. The size of the simplex remains constant. This prohibits the method from finding an optimum with a better accuracy, and leads to an inefficient search behavior in smooth and monotonous error surfaces. An extension is the so-called modified simplex method,[11] where the size of the simplex can be increased for more rapid convergence, or decreased for a more precise search. The results of this method are depicted in Figure 7(b). However, the optimum is found much more precisely, and in several cases the method is able to take big steps in the right direction, thus speeding up the optimization considerably.

In conclusion, the simplex method (and especially the modified simplex method) is an attractive optimization method in problems where experimentation is difficult or expensive, and where a number of other constraints are satisfied. The most important of these is the requirement that there is only one optimum, because otherwise

the simplex method may easily get stuck in a local optimum. This also implies that there should not be too much noise on the response variable, because this may lead to a projection of the wrong simplex vertex. In some cases this merely decreases the efficiency of the algorithm, in other cases it leads to completely wrong answers.[12]

Although a number of experiments are performed, the results cannot readily be used to calculated response surfaces, because the experiments may be concentrated in one area of the search space. However, response-surface methods explicitly cover the relevant space. Simplex methods are very useful if improvement rather than an optimal solution is sought.

3 WEAK METHODS

Weak methods make almost no assumptions, but at a certain cost. In many cases they are not very effective, and they should only be used in very specific circumstances. The ultimate weak technique, evaluating all possible solutions, is practically infeasible for all but the most simple problems. Therefore, weak techniques sample the solution space in order to locate areas of good solutions. A strong random component is used in either selecting the sampling locations or defining the orientation of a group of samples. Repeated application of these methods may lead to quite different results, so the reproducibility of these searches is low.

3.1 Random Search

Random search is not a strategy many people would use, unless there is no other alternative. Yet in some molecular mechanics software packages it is implemented to serve as a reference point for other optimization procedures. The strategy is simple: just keep on trying new candidate solutions and keep the best one(s) until the time is up. The only situation in which this strategy is expected to be not worse than other methods is for ''needle-in-a-haystack'' problems – the landscape is completely flat and does not give any indication that one is near the optimum until the optimum is actually found. In such cases an exhaustive search is the only method guaranteed to find a solution, but is often not feasible. A typical example is molecular recognition – seemingly small changes to an active molecule can lead to a drastic decrease in activity. It is therefore not surprising that, especially in combinatorial chemistry, the random search strategy has led to the discovery of several interesting and completely new drug leads, although also in this field more sophisticated techniques are being investigated.

3.2 Sampling Methods

A more systematic way of sampling a landscape is to use a grid of a predefined size. Additional grids with smaller spacings may be placed in promising areas. An example of this is depicted in Figure 8(a), where the grid of 16 points used in the gradient search is followed by two other sampling steps. A disadvantage of this method is that phenomena smaller than the grid spacing are easily overlooked. Moreover, in search spaces with more dimensions the number of samples required increases drastically. In some well-understood problems of low dimension, however, this method can be used very well. The main advantage is its versatility – no attempt is made to parameterize the response surface and, if more than one optimum is found, several grids can be evaluated in parallel.

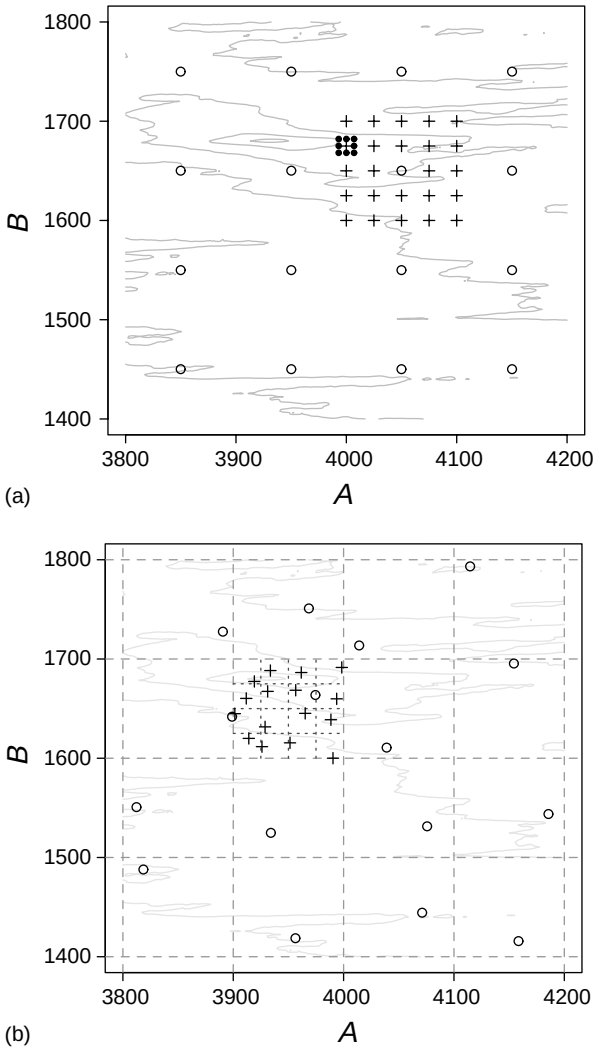

Figure 8 (a) Grid sampling and (b) stratified sampling. The spacing in the grids is 100, 25, and 7 units, respectively. The strata sizes are 100×100 and 25×25 units.

The stratified random search combines grid search with random search, by dividing the search space into strata and taking a random sample from each stratum. The problem that phenomena with a size smaller than the grid spacing may be overlooked is somewhat less acute in this approach. However, especially with large strata, some samples may be very close to each other leaving large parts of the search space empty. The number of samples per stratum may be increased if this is a problem. Again, a sequential approach may be used in which larger strata are followed by smaller ones. An example is indicated in Figure 8(b). There, the square in the original grid with the best sample value is further subdivided into smaller strata. Alternatively, a new set of strata can be centered at the location of the best sample.

Although sampling methods are grouped with the class of weak optimization methods, for a repeated application with decreasing intersample distance one must assume a more or less continuous form of response surface. This is not a very demanding assumption, however. The main problem with sampling approaches is the chance of missing important optima with a basin of attraction smaller than the intersample distance. In more than two or three dimensions this will frequently occur, because the number of sampling points must be kept manageable.

4 INTERMEDIATE METHODS

Stochastic optimization methods have been developed to counter the weaknesses of classical optimization methods in high-dimensional search problems. Instead of following a fixed path that is only determined by the choice of the starting point, they feature a strong random component, mainly to avoid getting trapped in local optima. Although finding the optimal solution is not guaranteed for all methods, in almost all cases these methods will find very good solutions. The most important methods are Monte Carlo simulated annealing (MCSA, or SA for short), GA and evolutionary strategies (ES), and tabu search (TS). The principles behind these methods are analogous to the physical cooling of a liquid (SA), Darwin's theories of evolution (GA, ES) and learning processes (TS), respectively.

Although the roots of these methods lie in the 1960s, it was not until the mid-1980s (SA and GA) or later (TS) that these methods were applied in chemistry. The results clearly indicated that the methods offered solutions not achievable with the classical optimization methods.

4.1 Monte Carlo Simulated Annealing

MCSA, also known as the Metropolis algorithm, is probably the most used optimization method in chemistry today. Although the principle of the method was used in 1953 by Metropolis et al.,[13] it only became popular after 1983 when Kirkpatrick et al.[2] described its use as a general optimization scheme. The method is implemented in a large number of software packages and is the de facto standard in structure optimization problems.[14]

The principle of the method is very simple indeed. A random walk is performed in the search space, accepting all moves that lead to a better solution, and accepting moves that lead to a worse solution with a probability $e^{-\Delta E/T}$. This is analogous to the well-known Boltzmann distribution. If T, a control parameter often called the temperature is large, the chance of accepting a bad move is large as well and the search truly resembles a random walk. The search characteristics of SA are obtained by gradually lowering the temperature so that bad moves are more and more often rejected. It is clear that the cooling scheme is of prime importance for the success of the method. Indeed, proofs have been published guaranteeing convergence to the global optimum provided the cooling is done slowly enough. In practice, this proof is not very useful because adequate cooling may be infinitely slow. Many variants exist, including schemes that take into account the distance to the optimal evaluation function value (e.g. error = 0.0; this particular form of SA is called generalized simulated annealing (GSA)) or algorithms that use a Cauchy rather than a Boltzmann distribution.

The only parts that must be provided by the user are an evaluation function, a procedure to generate a new trial state and a cooling scheme. In particular the latter two have a profound influence on the performance of the algorithm and in some cases may be difficult to define. The algorithm is stopped when either the optimal solution has been found, a predefined number of evaluations is reached, or no improvement has been obtained for a specified number of evaluations.

A simple example using the error surface of Figure 3 is depicted in Figure 9(a). The SA is started from the center of the permissible region. A step consists of moving to one of the 24 closest points in the neighborhood. The temperature factor was set to 575 energy units at the beginning and gradually lowered to 525 units at the last of the 300 iterations. Nearly 40% of all proposed moves was accepted. As can be seen in the right plot in Figure 9(b), the algorithms wander around aimlessly for several periods of time, whereas in other periods rapid improvements are made.

Although in this case the SA finds the optimum without much trouble, in real problems a fair amount of fine tuning is usually necessary, especially in the move-generating and cooling procedures. Many practitioners use cooling schedules that keep the ratio of rejected versus accepted moves at around 1. Because of the stochastic

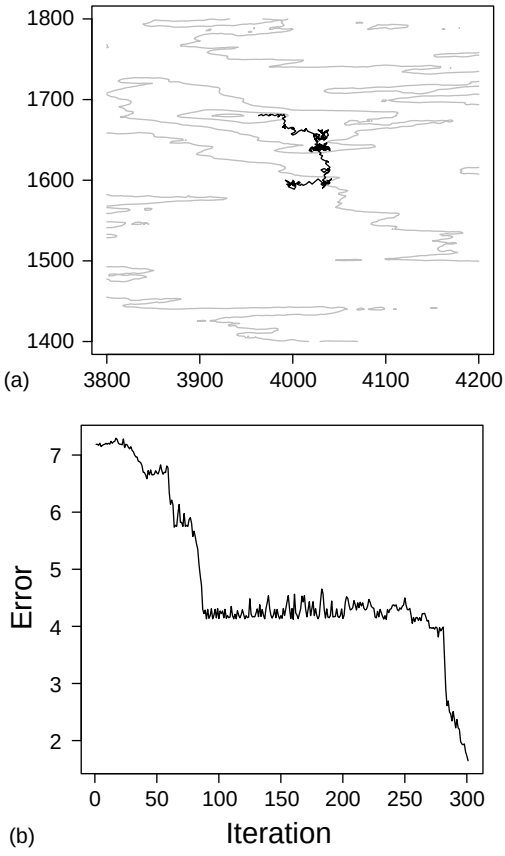

Figure 9 (a) SA to locate the optimum. Starting from the center of the admissible region, 300 iterations are performed. (b) The development of the error is depicted. Note the long period between iterations 80 and 260 where no improvement is made.

nature of SA, it is advisable to perform several runs from different starting points. There is no guarantee of finding the global optimum, especially in more difficult problems.

4.2 Tabu Search

Whereas SA performs an optimization by randomly generating a new state and applying the Boltzmann rule to decide whether to accept or reject, TS uses sophisticated history mechanisms to avoid evaluating solutions that are already known.[15] In its simplest form, TS generates a new state by selecting the best available move, excluding those states that are members of the tabu list T. To prevent memory problems, however, a tabu list usually contains moves, or components of moves, instead of states. This list of tabu moves is continually updated and consists of those moves (or move components) that would undo the most recent iterations in the search process. The net effect of the tabu list is that cycles with a length smaller than the length of the tabu list are avoided.

The TS paradigm is made more sophisticated by inclusions of aspiration criteria. These play the opposite role of the tabu moves – a move is regarded as admissible if its aspiration criteria are satisfied. Examples include situations where a move would lead to a better solution than all previously visited solutions, or where a move (component) has not been executed for a long time. Furthermore, intermediate- and long-term memory functions are used to identify attribute values that are clearly important for good solutions, and to diversify the search, respectively. Thus, the search can be focused to particular regions in space or the algorithm can escape from local optima.

The separate components of TS methods are much more problem dependent than is usually the case with other optimization methods. However, they are intuitively appealing and with expert knowledge about the problem not too difficult to define. The number of applications in chemistry is still limited, mainly because the method has been used primarily for sequential problems. Recently, several applications have been reported on structure optimization problems,[15] assessing chemical similarity,[16] and molecular docking.[17]

4.3 Evolutionary Optimization

A large number of different techniques fall under the header of evolutionary optimization, most notably evolutionary algorithms (EA)[18] and GAs.[19,20] Their main characteristic is that they use a *population* of trial solutions, instead of proceeding from one trial solution to the next. This may have several advantages. For example, in cases where one is interested not only in the global optimum but also in other good solutions, population-based methods may prevent the same optimum being found continuously. Thus, in principle, a more diverse set is found. Examples where this is very important are abundant – in chemistry, molecular structure optimization is used to find not only the global energy minimum but also all other structures that are feasible at room temperature.

The second important aspect of the use of a population is that there is no need to define how new solutions are obtained. In methods such as TS or SA one has to define how to proceed from the current solution to the next, but in EA this is typically performed by standard operators such as cross-over and mutation. Where the latter in most applications is a more or less random disruption of a member of the population, the former is specific for evolutionary optimization. In analogy with nature, two (or more) members of the population mate and exchange genetic material to produce children. The genetic material consists of parts of the solutions, and is coded either in binary form or as real numbers. Cross-over is also governed by chance, so no attempt is made

to identify ''good'' parts of a solution for reproduction. The best genetic material is more or less automatically preserved because solutions of a high quality will be selected for parenthood more often than bad solutions. The ''genes'' of the latter will gradually be removed from the population as the evolution goes on for a number of generations.

The success of EAs depends on the combination of representation and evolutionary operators. In general, evolutionary operators that yield invalid solutions (such as parameter values outside a permissible range, or impossible chemical structures in a structure-manipulation application) are to be avoided. Many applications have been described in molecular structure optimization, spectrum interpretation, subset selection for QSAR, and multivariate calibration and other fields.[21,22]

4.4 Hybrid Methods

Hybrid methods combine the characteristics of different optimization methods. Typically, the basic optimization cycle of Figure 1 is composed of elements from the individual methods. Of course, many methods differ only in one aspect. The main difference between TS and steepest descent methods, for example, is in the generation stage – some solutions are forbidden by the tabu list or explicitly allowed, but each method selects the best of the admissible moves. One further difference is that TS may accept a worse solution if that is the best one available, whereas steepest descent then stops.

Of course, hybrid methods are most useful when they combine the best features of their components. One approach which has been used in practice is a hybrid form of a GA and SA. The GA is used to maintain a pool of trial solutions and to generate offspring; the SA governs the selection part by accepting all improving solutions and accepting worse solutions according to the Boltzmann distribution. The advantages of the GA are that inadmissible regions of the search space are handled gracefully and do not constitute a barrier that a vintage SA may find difficult to cross. The advantage of the SA is the added control that the user may exert through the cooling scheme. For the selection and generation stages, typical strategies are summarized in Table 1. Other possible hybrids are GA/TS (a pool of solutions is maintained and the random reproduction mechanism is supervised by TS methods) and SA/TS (probabilistic TS, where the chance of accepting a new solution is governed by the SA Boltzmann distribution).

Table 1 Selection and generation characteristics for SA, TS, and GA

Method	Characteristic
Selection	
SA	Accept new solution if better, else accept with probability $e^{-\Delta E/T}$
TS	Accept best of all admissible new solutions, even if it is worse than the previous solution
GA	Accept solutions according to their relative quality in the population
Generation	
SA	Randomly pick one solution out of the neighborhood
TS	Generate all solutions in neighborhood
GA	Use random-based procedures such as cross-over and mutation to generate new solutions

The main strengths of all these methods – their versatility and their ability to adapt to all kinds of problems – are also their weakness. No generally accepted strategies for selecting, adapting and fine tuning these global search methods exist. If one is lucky, the methods work straight away. However, one may not be lucky, and although there are a million ways to alter and tweak the algorithms, no guidelines exist as to which one is best. Fortunately, in many cases suitable starting points and standard settings can be found in the literature for analogous problems.

4.5 Comparison of Methods

The question as to which of the intermediate methods mentioned so far is most suited for a particular problem is not easily answered, for several reasons. Although they are often called ''global'' optimization methods, in many practical problems there is no guarantee that the true global optimum will be found. Replicate runs may find radically different solutions, either because of the stochastic nature of the search performed (all but TS) or because of the influence of the starting position (all methods). This means that for a thorough comparison many replicate runs should be performed on a representative set of example problems. As these problems usually take a fair amount of computing time, extensive comparisons between algorithms are rarely performed.

A second problem is that each algorithm comes in many variants, each with a set of optimization parameters that can be fine tuned. This tuning may be critical to the performance of the algorithm and constitutes a meta-optimization problem. It is also difficult to give each of the algorithms tested the same level of attention – often, one of the techniques is already known to the implementors and this may give it a decisive advantage.

Finally, there is no clear-cut criterion with which to judge the performance of the optimization methods. Simply looking at the quality of the best solution clearly does not provide all the information. Therefore, other criteria are needed.[17,23] If these are not explicitly based

on the value of the evaluation function, the form of this function can be tuned as well. For a successful application of these intermediate search methods, the ensemble of representation, the search operations, and the evaluation function must be adjusted to each other.

5 CONCLUSIONS AND OUTLOOK

The field of optimization is broad and has applications in all areas of chemistry. Naturally, many different methods have been used and in some cases even developed by chemists. In this article, an overview is given where the methods have been classified as ''classical'' or ''nonclassical''. The former category comprises methods that are very good in dealing with low-dimensional problems. An example is the optimization of a chromatographic experiment, where optimal separation conditions should be found in as few experiments as possible. These methods are said to be strong because they incorporate much knowledge from the user about the problem at hand. The classical category also embraces weak methods where knowledge is not a prerequisite. As the efficiency of these methods is in many cases very low, they are not often used in chemistry.

The nonclassical optimization techniques described in this article have been specifically developed for those cases where the classical techniques were not suitable – high-dimensional search problems with many local optima. Because the number of evaluations may be quite high they usually are applied in connection with computer experiments rather than with laboratory experiments. A notable exception is found in the field of combinatorial chemistry, where GAs have been used to guide the synthesis of new promising lead molecules.[24]

The classical and nonclassical methods have complementary roles in chemistry. With the increasing role of computational chemistry, the nonclassical methods will continue to flourish. However, due to their complicated nature and great versatility, they will remain the playground of experts in the near future. With classical methods the situation is different. Many standard computer software packages are available for experimental design, and are sometimes even included in laboratory instruments. Although the principles of the methods should still be understood by the user, no programming is required, thereby significantly lowering the threshold for their use.

ABBREVIATIONS AND ACRONYMS

CCD	Central Composite Design
CG	Conjugate Gradient
EA	Evolutionary Algorithms
ES	Evolutionary Strategies
GA	Genetic Algorithms
GSA	Generalized Simulated Annealing
HPLC	High-performance Liquid Chromatography
IR	Infrared
LIF	Laser-induced Fluorescence
MCSA	Monte Carlo Simulated Annealing
NP	Nonpolynomial
QSAR	Quantitative Structure–Activity Relationship
RMS	Root Mean Square
SA	Simulated Annealing
TS	Tabu Search

RELATED ARTICLES

Pharmaceuticals and Drugs **(Volume 8)**
Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher • Signal Processing in Analytical Chemistry

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

REFERENCES

1. T. Schlick, 'Optimization Methods in Computational Chemistry', in eds. K.B. Lipkowitz, D.B. Boyd, *Reviews in Computational Chemistry*, VCH, New York, Chapter 1, Vol. 3, 1992.
2. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, 'Optimization by Simulated Annealing', *Science*, **220**, 671–680 (1983).
3. G. Berden, W. Leo Meerts, E. Jalviste, 'Rotationally Resolved UV Spectroscopy of Indole, Indazole and Benzimidazole, Inertial Axis Reorientation in the $S_1(^1L_b) \leftarrow S_0$ Transitions', *J. Chem. Phys.*, **103**, 9596–9606 (1995).
4. G.E.P. Box, W. Hunter, J. Hunter, *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*, John Wiley & Sons, New York, 1978.
5. R. Carlson, *Design and Optimization in Organic Synthesis*, Elsevier, Amsterdam, 1992.

6. J.C. Berridge, *Techniques for the Automated Optimization of HPLC Separations*, Wiley, New York, 1985.
7. J.L. Glajch, L.R. Snyder, *Computer-assisted Method Development for HPLC*, Elsevier, Amsterdam, 1990.
8. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1988.
9. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke (eds.), *Handbook of Chemometrics and Qualimetrics: Part A*, *Data Handling in Science and Technology*, Elsevier Science, Vol. 20A, 1998.
10. W. Spendley, G.R. Hext, F.R. Himsworth, 'Sequential Application of Simplex Designs in Optimization and Evolutionary Operations', *Technometrics*, **4**, 441–461 (1962).
11. J.A. Nelder, R. Mead, 'A Simplex Method for Function Optimization', *Computer J.*, **7**, 308–313 (1965).
12. F.H. Walter, L.R. Parker, S.L. Morgan, S.N. Deming, *Sequential Simplex Optimization*, CRC Press, Boca Raton, 1991.
13. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, 'Equation for State Calculation by Fast Computing Machines', *J. Chem. Phys.*, **21**, 1087–1091 (1953).
14. J.H. Kalivas (ed.), *Adaption of Simulated Annealing to Chemical Optimization Problems: Data Handling in Science and Technology*, Elsevier Science, Vol. 15, 1995.
15. F. Glover, M. Laguna, *Tabu Search*, Kluwer, Boston, 1997.
16. V. Kvasnicka, J. Pospichal, 'Fast Evaluation of Chemical Distance by Tabu Search Algorithm', *J. Chem. Inf. Comput. Sci.*, **34**, 1109–1112 (1994).
17. D.R. Westhead, D.E. Clark, C.W. Murray, 'A Comparison of Heuristic Search Algorithms for Molecular Docking', *J. Comput.-Aided Mol. Des.*, **11**, 209–228 (1997).
18. H.P. Schwefel, *Numerical Optimization of Computer Models*, Wiley, Chichester, 1981.
19. D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, 1989.
20. L. Davis (ed.), *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, 1991.
21. J. DeVillers (ed.), *Genetic Algorithms in Molecular Modeling, Principles of QSAR and Drug Design*, Academic Press, 1996.
22. R.S. Judson, 'Genetic Algorithms and Their use in Chemistry', in eds. K.B. Lipkowitz, D.B. Boyd, *Reviews in Computational Chemistry*, VCH, 1–73, 1997.
23. R. Wehrens, E. Pretsch, L.M.C. Buydens, 'Quality Criteria of Genetic Algorithms for Structure Optimization', *J. Chem. Inf. Comput. Sci.*, **38**, 151–157 (1998).
24. J.L. Fauchere, J.A. Boutin, J.M. Henlin, N. Kucharczyk, J.C. Ortuno, 'Combinatorial Chemistry for the Generation of Molecular Diversity and the Discovery of Bioactive Leads', *Chemometr. Intell. Lab. Systems*, **43**, 43–68 (1998).

Clustering and Classification of Analytical Data

Barry K. Lavine
Clarkson University, Potsdam, USA

1 Introduction **1**
2 Principal Component Analysis **2**
2.1 Variance-based Coordinate System 3
2.2 Information Content of Principal Components 3
2.3 Case Studies 4
3 Cluster Analysis **5**
3.1 Hierarchical Clustering 6
3.2 Practical Considerations 8
3.3 Case Studies 8
4 Pattern Recognition **9**
4.1 *k*-Nearest Neighbor 12
4.2 Soft Independent Modeling by Class Analogy 12
4.3 Feature Selection 13
4.4 Case Studies 13
5 Software **19**
6 Conclusion **19**
Abbreviations and Acronyms **19**
Related Articles **19**
References **20**

Clustering and classification are the major subdivisions of pattern recognition techniques. Using these techniques, samples can be classified according to a specific property by measurements indirectly related to the property of interest (such as the type of fuel responsible for an underground spill). An empirical relationship or classification rule can be developed from a set of samples for which the property of interest and the measurements are known. The classification rule can then be used to predict the property in samples that are not part of the original training set. The set of samples for which the property of interest and measurements is known is called the training set. The set of measurements that describe each sample in the data set is called a pattern. The determination of the property of interest by assigning a sample to its respective category is called recognition, hence the term pattern recognition.

For pattern recognition analysis, each sample is represented as a data vector $\boldsymbol{x} = (x_1, x_2, x_3, x_j, \ldots, x_n)$*, where component* x_j *is a measurement, e.g. the area* a *of the* j*th peak in a chromatogram. Thus, each sample is considered as a point in an* n*-dimensional measurement space. The dimensionality of the space corresponds to the number of measurements that are available for each sample. A basic assumption is that the distance between pairs of points in this measurement space is inversely related to the degree of similarity between the corresponding samples. Points representing samples from one class will cluster in a limited region of the measurement space distant from the points corresponding to the other class. Pattern recognition (i.e. clustering and classification) is a set of methods for investigating data represented in this manner, in order to assess its overall structure, which is defined as the overall relationship of each sample to every other in the data set.*

1 INTRODUCTION

Since the early 1980s, a major effort has been made to substantially improve the analytical methodology applied to the study of environmental samples. Instrumental techniques such as gas chromatography, high-performance liquid chromatography (HPLC) and X-ray fluorescence spectroscopy have dramatically increased the number of organic and inorganic compounds that can be identified and quantified, even at trace levels, in the environment. This capability, in turn, has allowed scientists to attack ever more complex problems, such as oil and fuel spill identification, but has also led to an information-handling problem.[1]

The reason for this problem is that in any monitoring effort it is necessary to analyze a large number of samples in order to assess the wide variation in composition that an environmental system may possess. The large number of samples that must be analyzed and the number of constituents that must be measured per sample give rise to data sets of enormous size and complexity. Often, important relationships in these data sets cannot be uncovered, when the data are examined one variable at a time, because of correlations between measurement variables, which tend to dominate the data and prevent information from being extracted.

Furthermore, the relationships sought in the data often cannot be expressed in quantitative terms, such as the source of a pollutant in the environment. These relationships are better expressed in terms of similarity or dissimilarity among groups of multivariate data. The task that confronts the scientist when investigating these sorts of relationships in multivariate data, is twofold:

- Can a useful structure based on distinct sample groups be discerned?
- Can a sample be classified into one of these groups for the prediction of some property?

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The first question is addressed using principal component analysis (PCA)[2] or cluster analysis,[3] whereas the second question is addressed using pattern recognition methods.[4]

PCA is the most widely used multivariate analysis technique in science and engineering.[5] It is a method for transforming the original measurement variables into new variables called principal components. Each principal component is a linear combination of the original measurement variables. Often, only two or three principal components are necessary to explain all of the information present in the data. By plotting the data in a coordinate system defined by the two or three largest principal components, it is possible to identify key relationships in the data, that is, find similarities and differences among objects (such as chromatograms or spectra) in a data set.

Cluster analysis[6] is the name given to a set of techniques that seek to determine the structural characteristics of a data set by dividing the data into groups, clusters, or hierarchies. Samples within the same group are more similar to each other than samples in different groups. Cluster analysis is an exploratory data analysis procedure. Hence, it is usually applied to data sets for which there is no a priori knowledge concerning the class membership of the samples.

Pattern recognition[7] is a name given to a set of techniques developed to solve the class-membership problem. In a typical pattern recognition study, samples are classified according to a specific property using measurements that are indirectly related to that property. An empirical relationship or classification rule is developed from a set of samples for which the property of interest and the measurements are known. The classification rule is then used to predict this property in samples that are not part of the original training set. The property in question may be the type of fuel responsible for a spill, and the measurements are the areas of selected gas chromatographic (GC) peaks. Classification is synonymous with pattern recognition, and scientists have turned to it and PCA and cluster analysis to analyze the large data sets typically generated in monitoring studies that employ computerized instrumentation.

This article explores the techniques of PCA, cluster analysis, and classification. The procedures that must be implemented to apply these techniques to real problems are also enumerated. Special emphasis is placed on the application of these techniques to problems in environmental analysis.

2 PRINCIPAL COMPONENT ANALYSIS

PCA is probably the oldest and best known of the techniques used for multivariate analysis. The overall goal of PCA is to reduce the dimensionality of a data set, while simultaneously retaining the information present in the data. Dimensionality reduction or data compression is possible with PCA because chemical data sets are often redundant. That is, chemical data sets are not information rich. Consider a gas chromatogram of a JP-4 fuel (Figure 1), which is a mixture of alkanes, alkenes, and aromatics. The gas chromatogram of a JP-4 fuel is characterized by a large number of early-eluting peaks, which are large in size. There are a few late-eluting peaks, but their size is small. Clearly, there is a strong negative correlation between the early- and late-eluting peaks of the JP-4 fuel. Furthermore, many of the alkane and alkene peaks are correlated, which should not come as a surprise as alkenes are not constituents of crude oil but instead are formed from alkanes during the refining process. In addition, the property of a fuel most likely to be reflected in a high resolution gas chromatogram is its distillation curve, which does not require all 85 peaks for characterization.

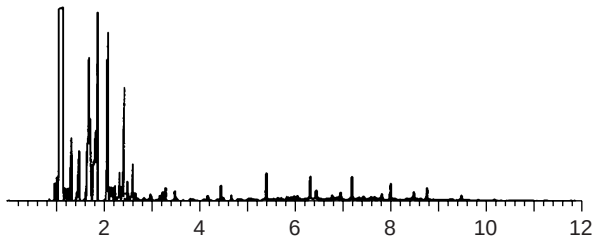

Figure 1 A high-resolution capillary column gas chromatogram of a JP-4 fuel.

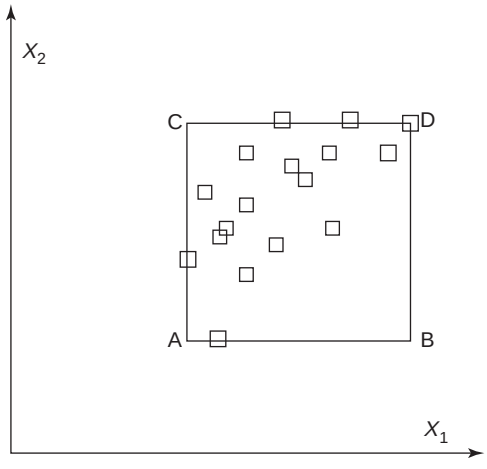

Figure 2 Seventeen hypothetical samples projected onto a two-dimensional measurement space defined by the measurement variables X_1 and X_2. The vertices, A, B, C, and D, of the rectangle represent the smallest and largest values of X_1 and X_2. (Adapted from Mandel.[8])

Redundancy in data is due to collinearity (i.e. correlations) among the measurement variables. Collinearity diminishes the information content of the data. Consider a set of samples characterized by two measurements, X_1 and X_2. Figure 2 shows a plot of these data in a two-dimensional measurement space, where the coordinate axes (or basis vectors) of this measurement space are the variables X_1 and X_2. There appears to be a relationship between these two measurement variables, which suggests that X_1 and X_2 are correlated, because fixing the value of X_1 limits the range of values possible for X_2. If the two measurement variables were uncorrelated, the enclosed rectangle in Figure 2 would be fully populated by the data points. Because information is defined as the scatter of points in a measurement space, it is evident that correlations between the measurement variables decrease the information content of this space. The data points, which are restricted to a small region of the measurement space due to correlations among the variables, could even reside in a subspace if the measurement variables are highly correlated. This is shown in Figure 3. Here X_3 is perfectly correlated with X_1 and X_2 because X_1 plus X_2 equals X_3. Hence, the seven sample points lie in a plane even though each data point has three measurements associated with it.

2.1 Variance-based Coordinate System

Variables that have a great deal of redundancy or are highly correlated are said to be collinear. High collinearity between variables is a strong indication that a new set of basis vectors can be found that will be better at conveying the information content present in data than axes defined by the original measurement variables. The new basis set that is linked to variation in the data can be used to develop a new coordinate system for displaying the data. The principal components of the data define the variance-based axes of this new coordinate system.

$$X = \begin{bmatrix} X_1 & X_2 & X_3 \\ 0.5 & 0.5 & 1.0 \\ 1.9 & 0.7 & 2.6 \\ 2.0 & 2.0 & 4.0 \\ 0.3 & 1.8 & 2.1 \\ 1.9 & 1.7 & 3.6 \\ 1.2 & 0.2 & 1.4 \\ 1.9 & 0.9 & 2.8 \end{bmatrix}$$

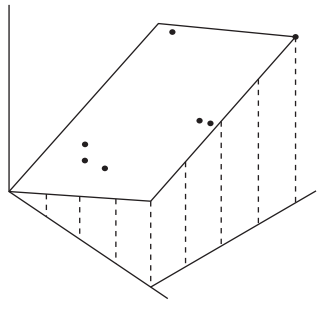

Figure 3 In the case of strongly correlated variables, the data points may reside in a subspace of the original measurement space. (Adapted from *Multivariate Pattern Recognition in Chemometrics*.[4] Copyright 1992, with permission from Elsevier Science.)

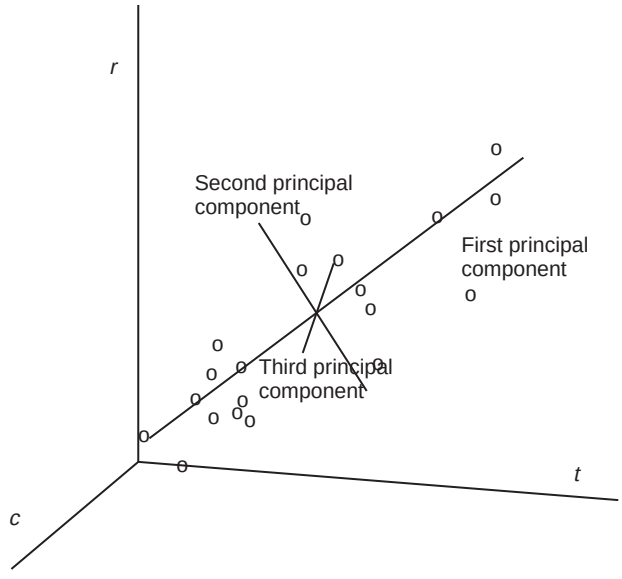

Figure 4 Principal component axes developed from the measurement variables *r*, *c*, and *t*. (Reproduced by permission from Brown[5] courtesy of Society of Applied Spectroscopy.)

The largest or first principal component is formed by determining the direction of largest variation in the original measurement space and modeling it with a line fitted by linear least squares (Figure 4) that passes through the center of the data. The second largest principal component lies in the direction of next largest variation – it passes through the center of the data and is orthogonal to the first principal component. The third largest principal component lies in the direction of next largest variation – it also passes through the center of the data, it is orthogonal to the first and second principal components, and so forth. Each principal component describes a different source of information because each defines a different direction of scatter or variance in the data. (The scatter of the data points in the measurement space is a direct measure of the data's variance.) Hence, the orthogonality constraint imposed by the mathematics of PCA ensures that each variance-based axis will be independent.

2.2 Information Content of Principal Components

One measure of the amount of information conveyed by each principal component is the variance of the data explained by the principal component. The variance explained by each principal component is expressed in terms of its eigenvalue. For this reason, principal components are usually arranged in order of decreasing eigenvalues or waning information content. The most informative principal component is the first and the least informative is the last. The maximum number of

principal components that can be extracted from the data is the smaller of either the number of samples or number of measurements in the data set, as this number defines the largest number of independent variables in the data.

If the data are collected with due care, one would expect that only the first few principal components would convey information about the signal, as most of the information in the data should be about the effect or property of interest being studied. However, the situation is not always this straightforward. Each principal component describes some amount of signal and some amount of noise in the data because of accidental correlation between signal and noise. The larger principal components primarily describe signal variation, whereas the smaller principal components essentially describe noise. When smaller principal components are deleted, noise is being discarded from the data, but so is a small amount of signal. However, the reduction in noise more than compensates for the biased representation of the signal that results from discarding principal components that contain a small amount of signal but a large amount of noise. Plotting the data in a coordinate system defined by the two or three largest principal components often provides more than enough information about the overall structure of the data. This approach to describing a data set in terms of important and unimportant variation is known as soft modeling in latent variables.

PCA takes advantage of the fact that a large amount of data is usually generated in monitoring studies when sophisticated chemical instrumentation, which is commonly under computer control, is used. The data have a great deal of redundancy and therefore a great deal of collinearity. Because the measurement variables are correlated, 85 peak gas chromatograms do not necessarily require 85 independent axes to define the position of the sample points. Utilizing PCA, the original measurement variables that constitute a correlated axis system can be converted into a system that removes correlation by forcing the new axes to be independent and orthogonal. This requirement greatly simplifies the data because the correlations present in the data often allow us to use fewer axes to describe the sample points. Hence, the gas chromatograms of a set of JP-4 and Jet-A fuel samples may reside in a subspace of the 85-dimensional measurement space. A plot of the two or three largest principal components of the data can help us to visualize the relative position of the Jet-A and JP-4 fuel samples in this subspace.

2.3 Case Studies

With PCA, we are able to plot the data in a new coordinate system based on variance. The origin of the new coordinate system is the center of the data, and the coordinate axes of the new system are the principal components of the data. Employing this new coordinate system, we can uncover relationships present in the data, that is, find distinct samples subgroups within the data. This section shows, by way of two published studies, how principal components can be used to discern similarities and differences among sample within a data set.

2.3.1 Troodos Data Set

In the first study, 143 rock samples collected in the Troodos region of Cyprus were analyzed by X-ray fluorescence spectroscopy for 10 metal oxides, which contained information about the formation of these rocks. If the formation of the entire Troodos region occurred at the same time, one would expect all of the rocks to be similar in composition. However, if there are distinct subgroups in the data, other conclusions may have to be drawn about the formation of the Troodos region. This study[9] was initiated to settle a controversy about the geological history of Cyprus.

Figure 5 shows a plot of the two largest principal components of the 143 rock samples. (The original Troodos data set was modified for the purpose of this principal component mapping exercise.) Samples 65 and 66 appear to be outliers in the plot as they are distant from the other samples. As a general rule, outliers should deleted because of the least-squares property of principal components. In other words, a sample that is distant from the other points in the measurement space can pull the principal components towards it and away from the direction of maximum variance. Figure 6 shows the results of a principal component mapping experiment

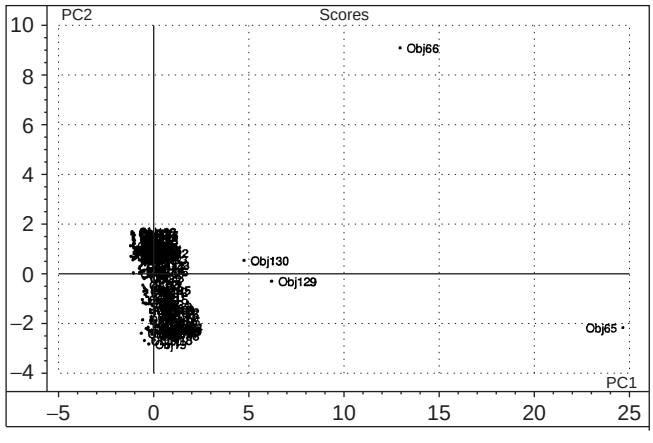

Figure 5 A principal component map of the 143 rock samples. Samples 65 and 66 are outliers. (The original data set was modified for this principal component mapping exercise.) The principal component map was generated using the program UNSCRAMBLER.

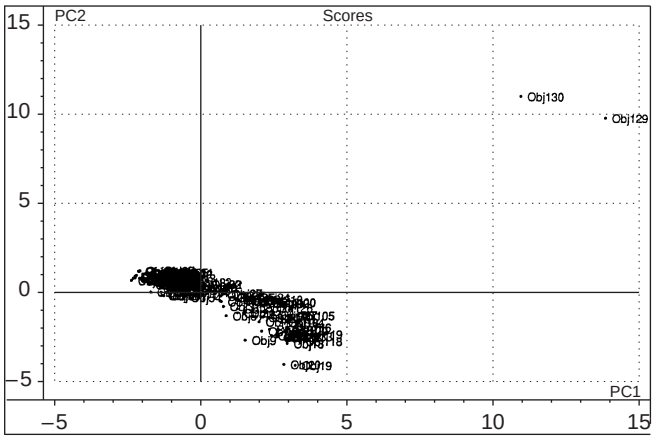

Figure 6 A principal component map of the Troodos rock samples with samples 65 and 66 removed. Samples 129 and 130 appear as outliers in the plot. (The original data set was modified for this principal component mapping exercise.) The principal component map was generated using the program UNSCRAMBLER.

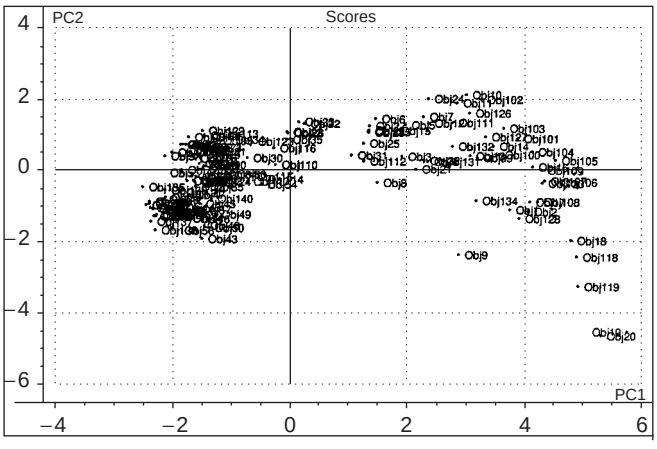

Figure 7 A principal component map of the Troodos rock samples with samples 65, 66, 129, and 130 removed. The principal component map was generated using the program UNSCRAMBLER.

with samples 65 and 66 removed from the data. It is evident from the plot that samples 129 and 130 are also outliers. Figure 7 summarizes the results of a principal component mapping experiment with samples 65, 66, 129, and 130 removed. Although samples 19 and 20 are probably outliers and are also candidates for removal, it is evident from the principal component plot that the rock samples can be divided into two groups, which would suggest that other conclusions should be drawn about the geological history of the Troodos region. The clustering of the rocks samples into two distinct groups was not apparent until the four outliers were removed from the data.

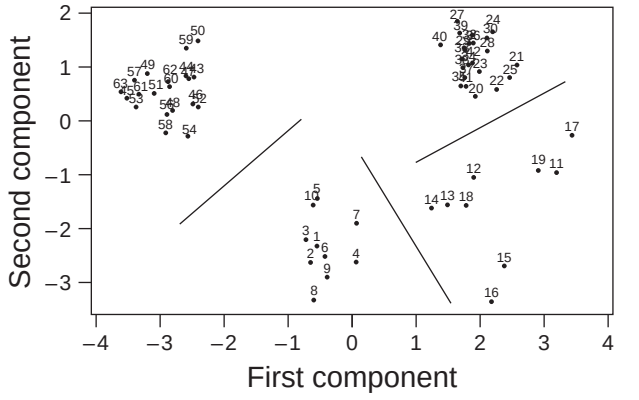

Figure 8 A plot of the two largest principal components for the 63 Indian artifacts developed from the concentration data of 10 metals. The principal component map was generated using the program SCAN.

2.3.2 Obsidian Data Set

The second study[10] also involves X-ray fluorescence data. Sixty-three Indian artifacts (such as jewelry, weapons, and tools) made from volcanic glass, were collected from four quarries in the San Francisco Bay area. (Samples 1–10 are from quarry 1, samples 11–19 are from quarry 2, samples 20–42 are from quarry 3, and samples 43–63 are from quarry 4.) Because the composition of volcanic glass is characteristic of the site and tends to be homogeneous, it is reasonable to assume that it should be possible to trace these artifacts to their original source material. In this study, the investigators attempted to do this by analyzing the 63 glass samples for 10 elements: Fe, Ti, Ba, Ca, K, Mn, Rb, Sr, Y, and Zn. Next, a PCA was performed on the data (63 artifacts with 10 features per artifact). The goal was to identify the overall trends present in the data. Figure 8 shows a plot of the two largest principal components of the data. From the principal component map, it is evident that the 63 Indian artifacts can be divided into four groups, which correspond to the quarry sites from which the artifacts were collected. Evidently, the artifacts in each quarry were made from the same source material. This result is significant because it provides the archaeologists with important information about the migration patterns and trading routes of the Indians in this region. Further details about the obsidian data can be found elsewhere.[10]

3 CLUSTER ANALYSIS

Cluster analysis is a popular technique whose basic objective is to discover sample groupings within data. The technique is encountered in many fields, such as biology, geology, and geochemistry, under such

names as unsupervised pattern recognition and numerical taxonomy. Clustering methods are divided into three categories, hierarchical, object-functional, and graph theoretical. The focus here is on hierarchical methods, as they are the most popular.

For cluster analysis, each sample is treated as a point in an n-dimensional measurement space. The coordinate axes of this space are defined by the measurements used to characterize the samples. Cluster analysis assesses the similarity between samples by measuring the distances between the points in the measurement space. Samples that are similar will lie close to one another, whereas dissimilar samples are distant from each other. The choice of the distance metric to express similarity between samples in a data set depends on the type of measurement variables used.

Typically, three types of variables – categorical, ordinal, and continuous – are used to characterize chemical samples. Categorical variables denote the assignment of a sample to a specific category. Each category is represented by a number, such as 1, 2, 3, etc. Ordinal variables are categorical variables, in which the categories follow a logical progression or order, such as 1, 2, and 3 denoting low, middle, and high, respectively. However, continuous variables are quantitative. The difference between two values for a continuous variable has a precise meaning. If a continuous variable assumes the values 1, 2, and 3, the difference between the values 3 and 2 will have the same meaning as the difference between the values 2 and 1, because they are equal.

Measurement variables are usually continuous. For continuous variables, the Euclidean distance is the best choice for the distance metric, because interpoint distances between the samples can be computed directly

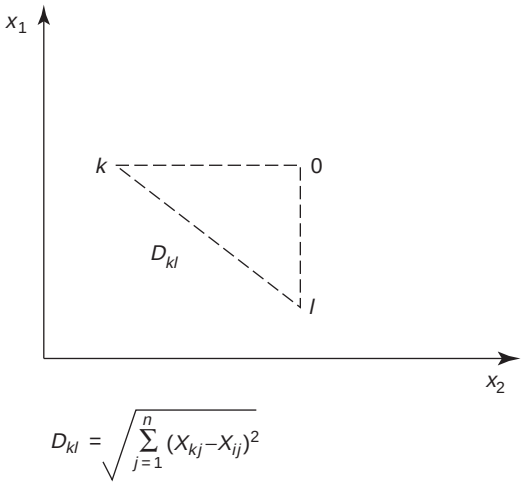

Figure 9 Euclidean distance between two data points in a two-dimensional measurement space defined by the measurement variables x_1 and x_2. (Reprinted from ref. 3.)

(Figure 9). However, there is a problem with using the Euclidean distance, which is the so-called scaling effect. It arises from inadvertent weighting of the variables in the analysis that can occur due to differences in magnitude among the measurement variables. For example, consider a data set where each sample is described by two variables: the concentration of Na and the concentration of K as measured by atomic flame emission spectroscopy. The concentration of Na varies from 50 to 500 ppm, whereas the concentration of K in the same samples varies from 5 to 50 ppm. A 10% change in the Na concentration will have a greater effect on Euclidean distance than a 10% change in K concentration. The influence of variable scaling on the Euclidean distance can be mitigated by autoscaling the data, which involves standardizing the measurement variables, so that each variable has a mean of zero and a standard deviation of 1 (Equation 1):

$$x_{i,\text{standardized}} = \frac{x_{i,\text{orig}} - m_{i,\text{orig}}}{s_{i,\text{orig}}} \tag{1}$$

where $x_{i,\text{orig}}$ is the original measurement variable i, $m_{i,\text{orig}}$ is the mean of the original measurement variable i, and $s_{i,\text{orig}}$ is the standard deviation of the original measurement variable i. Thus, a 10% change in K concentration has the same effect on the Euclidean distance as a 10% change in Na concentration when the data is autoscaled. Clearly, autoscaling ensures that each measurement variable has an equal weight in the analysis. For cluster analysis, it is best to autoscale the data, because similarity is directly determined by a majority vote of the measurement variables.

3.1 Hierarchical Clustering

Clustering methods attempt to find clusters of patterns (i.e. data points) in the measurement space, hence the term cluster analysis. Although several clustering algorithms exist, e.g. K-means, K-median, Patrick-Jarvis, FCV (fuzzy clustering varieties), hierarchical clustering is by far the most widely used clustering method. The starting point for a hierarchical clustering experiment is the similarity matrix which is formed by first computing the distances between all pairs of points in the data set. Each distance is then converted into a similarity value (Equation 2):

$$s_{ik} = 1 - \frac{d_{ik}}{d_{\max}} \tag{2}$$

where s_{ik} (which varies from 0 to 1) is the similarity between samples i and k, d_{ik} is the Euclidean distance between samples i and k, and $d_{\max}$ is the distance between the two most dissimilar samples (i.e. the largest distance) in the data set. The similarity values are organized in the form of a table or matrix. The similarity matrix is

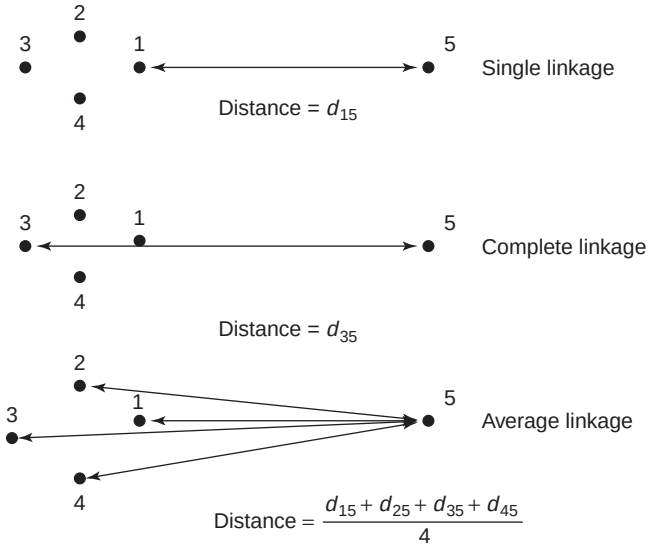

Figure 10 The distance between a data cluster and a point using single linkage, complete linkage, and average linkage. (Reproduced by permission from Lavine[20] courtesy of Marcel Dekker, Inc.)

then scanned for the largest value, which corresponds to the most similar point pair. The two samples constituting the point pair are combined to form a new point, which is located midway between the two original points. The rows and columns corresponding to the old data points are then removed from the matrix. The similarity matrix for the data set is then recomputed. In other words, the matrix is updated to include information about the similarity between the new point and every other point in the data set. The new nearest point pair is identified, and combined to form a single point. This process is repeated until all points have been linked.

There are a variety of ways to compute the distances between data points and clusters in hierarchical clustering (Figure 10). The single-linkage method assesses similarity between a point and a cluster of points by measuring the distance to the closest point in the cluster. The complete linkage method assesses similarity by measuring the distance to the farthest point in the cluster. Average linkage assesses the similarity by computing the distances between all point pairs where a member of each pair belongs to the cluster. The average of these distances is used to compute the similarity between the data point and the cluster.

Table 1 HPLC data set

Column	k_1	k_2	k_3
A	0.31	17.8	3
B	0.10	9.30	3
C	0.11	21.5	1
D	0.58	22.0	2
E	0.50	16.0	1

(Reprinted from *Multivariate Pattern Recognition in Chemometrics*.[4] Copyright 1992, by permission of Elsevier Science.)

Table 2 Similarity matrix

Columns	A	B	C	D	E
A	1.00	0.79	0.58	0.69	0.61
B	0.79	1.00	0.36	0.17	0.34
C	0.58	0.36	1.00	0.51	0.72
D	0.69	0.17	0.51	1.00	0.75
E	0.61	0.34	0.72	0.75	1.00

To illustrate hierarchical clustering, consider the data shown in Table 1. (The example shown here is an adaptation of the exercise described in Chapter 6 of reference 4.) Five HPLC columns were characterized by the capacity factor values obtained from three substances, which served as retention probes. To perform single-linkage hierarchical clustering on this chromatographic data, it is necessary to first compute the similarity matrix for the data, given as Table 2.

The similarity matrix (Table 2) is then scanned for the largest value, which corresponds to the two HPLC columns that are most similar. An examination of the similarity matrix suggests that chromatographic columns A and B with a score of 0.79 are the most similar. Hence, chromatographic columns A and B should be combined to form a new point. The rows and columns corresponding to the two original points (A and B) are removed from the similarity matrix. The similarity matrix for the data set is then updated to include information about the similarity between the new point and every other point (C, D, and E) in the data set. In this study, the investigators chose to use the single-linkage criterion for assessing the similarity between a data point and a point cluster, see Table 3. (Using the single-linkage method, the similarity between point D and the cluster consisting of columns A and B is the larger of the values of 0.69

Table 3 Updated similarity matrix

Columns	A, B	C	D	E
A, B	1.00	0.58	0.69	0.61
C	0.58	1.00	0.51	0.72
D	0.69	0.51	1.00	0.75
E	0.61	0.72	0.75	1.00

Table 4 Updated similarity matrix

Columns	A, B	C	D, E
A, B	1.00	0.58	0.61
C	0.58	1.00	0.72
D, E	0.61	0.72	1.00

and 0.17, see Table 2. For complete linkage, the similarity between this cluster and point D is the smaller of the two values.)

The updated similarity matrix is then scanned for the largest value; the new nearest pair is combined to form a single point, which is points D and E. The rows and columns corresponding to points D and E are deleted from the similarity matrix. The similarity matrix for the data set is then updated (Table 4) to include information about the similarity between the new point (D + E) and every other point in the data set. This process is repeated (Table 5) until all points are merged into a single cluster.

The results of a hierarchical clustering study are usually displayed as a dendogram, which is a tree-shaped map of the intersample distances in the data set. The dendogram shows the merging of samples into clusters at various stages of the analysis and the similarities at which the clusters merge, with the clustering displayed hierarchically. The dendogram for the single-linkage analysis of the HPLC data is shown in Figure 11. Interpretation of the results is intuitive, which is the major reason for the popularity of these methods.

3.2 Practical Considerations

A major problem in hierarchical clustering (or cluster analysis for that matter) is defining a cluster. Contrary to many published reports, there is no cluster validity measure that can serve as an indicator of the quality of a proposed partitioning of the data. Hence, clusters are defined intuitively, depending on the context of the problem, not mathematically, which limits the utility of this technique. Clearly, prior knowledge about the problem is essential when using these methods. The criterion for determining the threshold value for similarity is often subjective and depends to a large degree on the nature of the problem investigated – for example, the goals of the study, the number of clusters sought, previous experience, and common sense.

Table 5 Updated similarity matrix

Columns	A, B	C, D, E
A, B	1.00	0.69
C, D, E	0.69	1.00

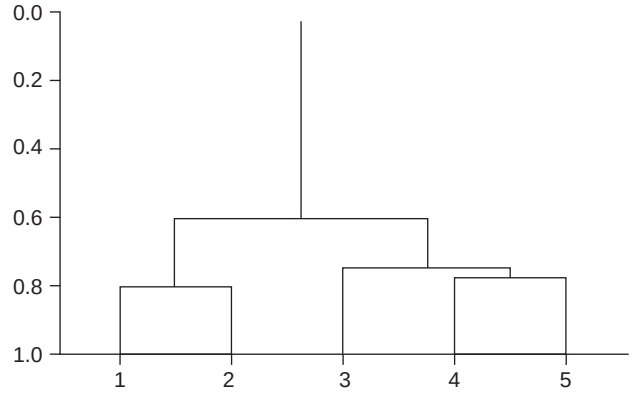

Figure 11 Single-linkage dendogram of the HPLC data set. There are two groups of HPLC columns: 1, 2 and 3, 4, 5. (Reprinted from *Multivariate Pattern Recognition in Chemometrics*.[4] Copyright 1992, with permission from Elsevier Science.)

All clustering procedures yield the same results for data sets with well-separated clusters. However, the results will differ when the clusters overlap. That is why it is a good idea to use at least two different clustering algorithms, such as single and complete linkage, when studying a data set. If the dendograms are in agreement, then a strong case can be made for partitioning the data into distinct groups as suggested by the dendograms. If the cluster memberships differ, the data should be further investigated using average linkage or PCA. The results from average linkage or PCA can be used to gauge whether the single or farthest linkage solution is the better one.

All hierarchical clustering techniques suffer from so-called space distorting effects. For example, single-linkage favors the formation of large linear clusters instead of the usual elliptical or spherical clusters. As a result, poorly separated clusters are often chained together. However, complete linkage favors the formation of small spherical clusters. Because of these space-distorting effects, hierarchical clustering methods should be used in tandem with PCA to detect clusters in multivariate data sets.

All hierarchical methods will always partition data, even randomly generated data, into distinct groups or clusters. Hence, it is important to ascertain the significance level of the similarity value selected by the user. For this task, a simple three-step procedure is proposed. First, a random data set is generated with the same correlation structure, the same number of samples, and the same number of measurements as the real data set that is currently being investigated. Second, the same clustering technique(s) is applied to the random data. Third, the similarity value, which generates the same number of clusters as identified in the real data set, is determined from the dendogram of the random data. If the similarity value is substantially larger for the real data set, the likelihood of having inadvertently exploited random variation in the data to achieve clustering is probably insignificant.

3.3 Case Studies

Hierarchical clustering methods attempt to uncover the intrinsic structure of a multivariate data set without making a priori assumptions about the data. This section,

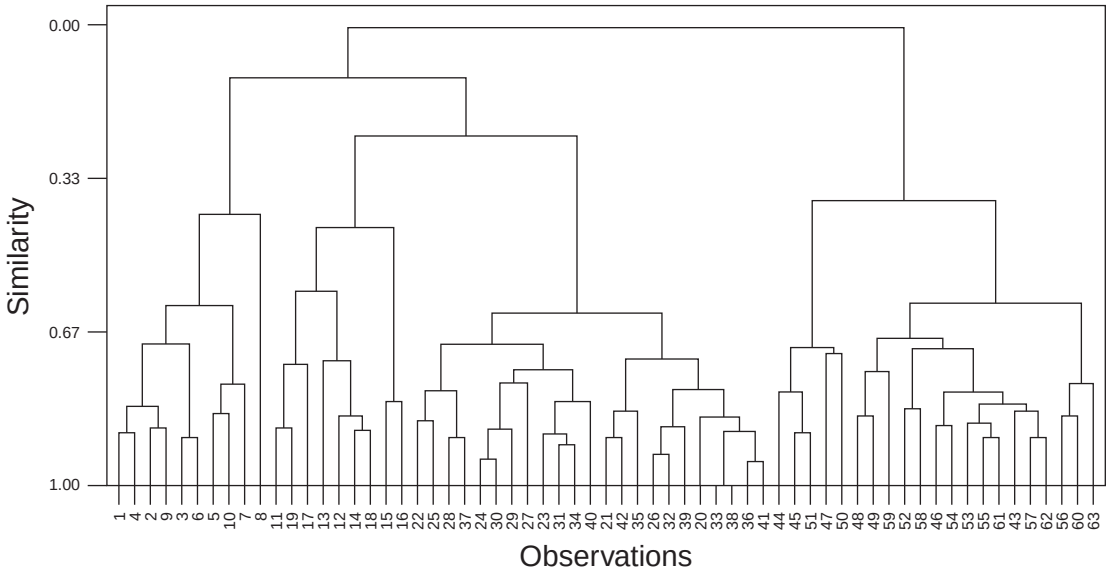

Figure 12 A single-linkage dendogram of the obsidian data set. The dendogram was generated using the program SCAN.

shows, by way of two published studies, how clustering methods can be used to find clusters of points in data.

3.3.1 Obsidian Data Set

The first study is the obsidian data set, discussed in section 2.3.2. A principal component map of the data (Figure 8) revealed four distinct clusters, which correspond to the sites from which these artifacts were obtained. To confirm the four-cluster hypothesis, the investigators also analyzed their data using single-linkage analysis. The resulting dendogram (Figure 12) indicated that it is reasonable to divide the glass samples into four categories based on the quarry sites from which these artifacts were obtained. (At similarity 0.40, samples 1–10 form a cluster, samples 11–19 form another cluster, samples 20–42 form a third cluster, and samples 43–63 form the fourth cluster.) Because the dendogram and principal component map of the data are in strong agreement, the investigators decided to partition their data into four groups based on the quarry labels of the samples.

3.3.2 Artificial Nose Data Set

The second study involves data from an artificial nose.[11] A salesman claims that an electronic nose can successfully sniff odor. The salesman obtained data from the literature to support his claim. Of the 41 compounds in the data set, compounds 1–21 are etheral, compounds 22–32 are pungent, and compounds 33–41 are minty. Each compound was characterized by six electronic measurements. Using a back-propagation neural network algorithm, the salesman was able to correctly classify all of the compounds in the data set. Should we then accept the salesman's claim that an electronic nose can sniff odor?

In order to validate the salesman's claim, the odor data was analyzed using single and complete linkage hierarchical clustering. Figures 13 and 14 show dendograms of the odor data. It is evident from the dendograms that dividing the 41 compounds into three categories based on odor type cannot be justified. (Although the data can be divided into three clusters by complete linkage at a 0.56 similarity, the cluster memberships cannot be correlated to compound odor – why? Cluster 1 consists of samples 41, 21, 19, 16, 13, 39, 31, 36, 29, 8, 34, and 32. Cluster 2 consists of samples 38, 5, 26, 14, 24, 15, 37, 6, 33, 35, 20, 30, and 7. Cluster 3 consists of samples 40, 17, 4, 22, 2, 25, 10, 28, 18, 27, 23, 12, 11, 1, 9, and 3.) Evidently, the six electronic measurements from the nose do not contain sufficient discriminatory information to force the compounds to cluster on the basis of odor type. Therefore, the salesman's claim about the efficacy of the proposed artificial nose should not be accepted at face value.

4 PATTERN RECOGNITION

So far, only exploratory data analysis techniques, i.e. cluster analysis and PCA, have been discussed. These techniques attempt to analyze data without directly using information about the class assignment of the samples. Although cluster analysis and PCA are powerful methods for uncovering relationships in large multivariate data sets, they are not sufficient for developing a classification rule that can accurately predict the class-membership of

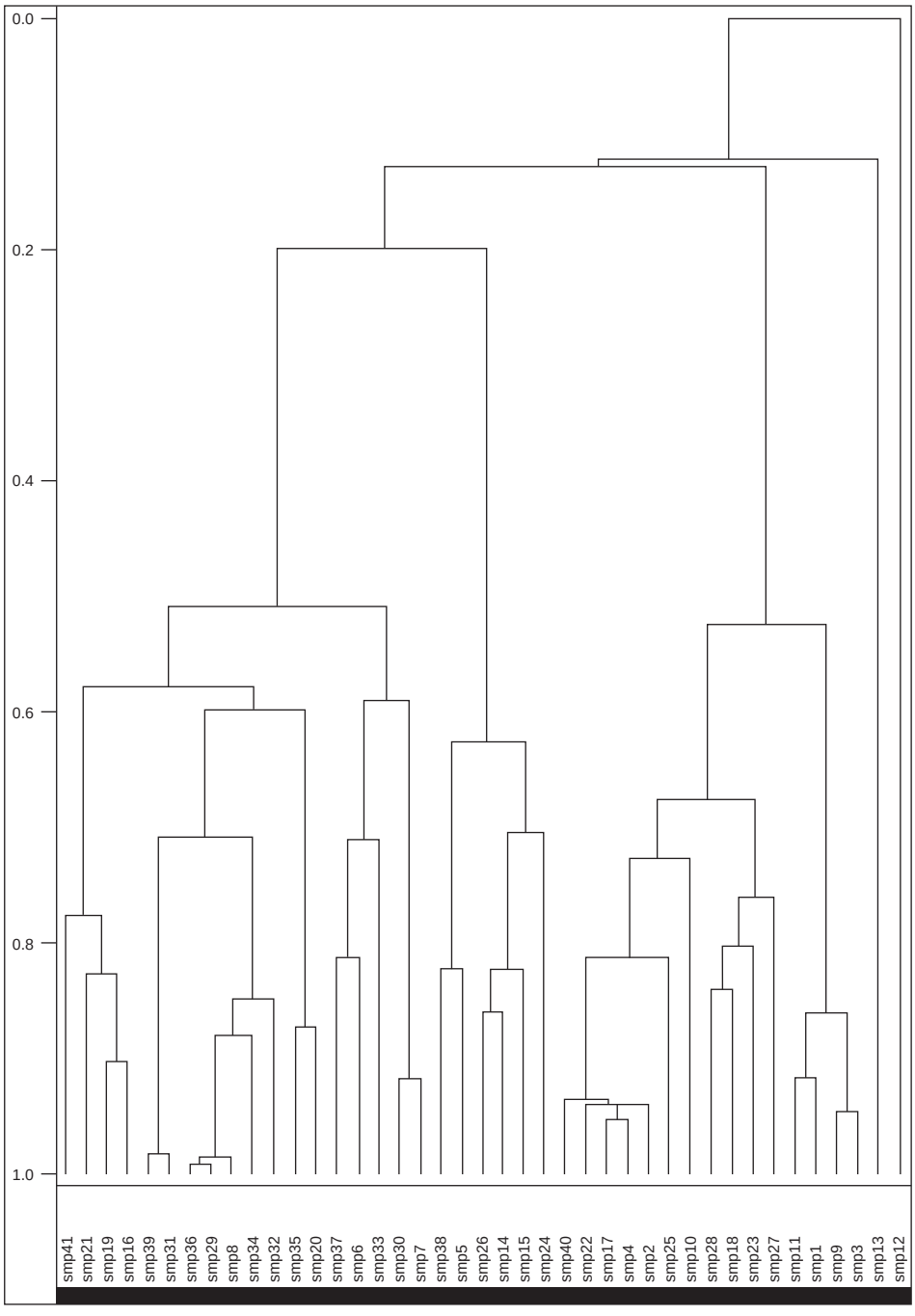

Figure 13 A single-linkage dendogram of the nose data set. The dendogram was generated using the program Pirouette.

an unknown sample. In this section, pattern recognition techniques will be discussed. These techniques were originally developed to categorize a sample on the basis of regularities in observed data. The first applications of pattern recognition to chemistry were studies involving low-resolution mass spectrometry.[12] Since then, pattern recognition techniques have been applied to a wide variety of chemical problems, such as chromatographic fingerprinting,[13–15] spectroscopic imaging,[16–18] and data interpretation.[19–21]

Pattern recognition techniques fall into one of two categories: nonparametric discriminants, and similarity-based classifiers. Nonparametric discriminants,[22–24] such as neural networks, attempt to divide a data space into different regions. In the simplest case, that of a binary classifier, the data space is divided into two regions.

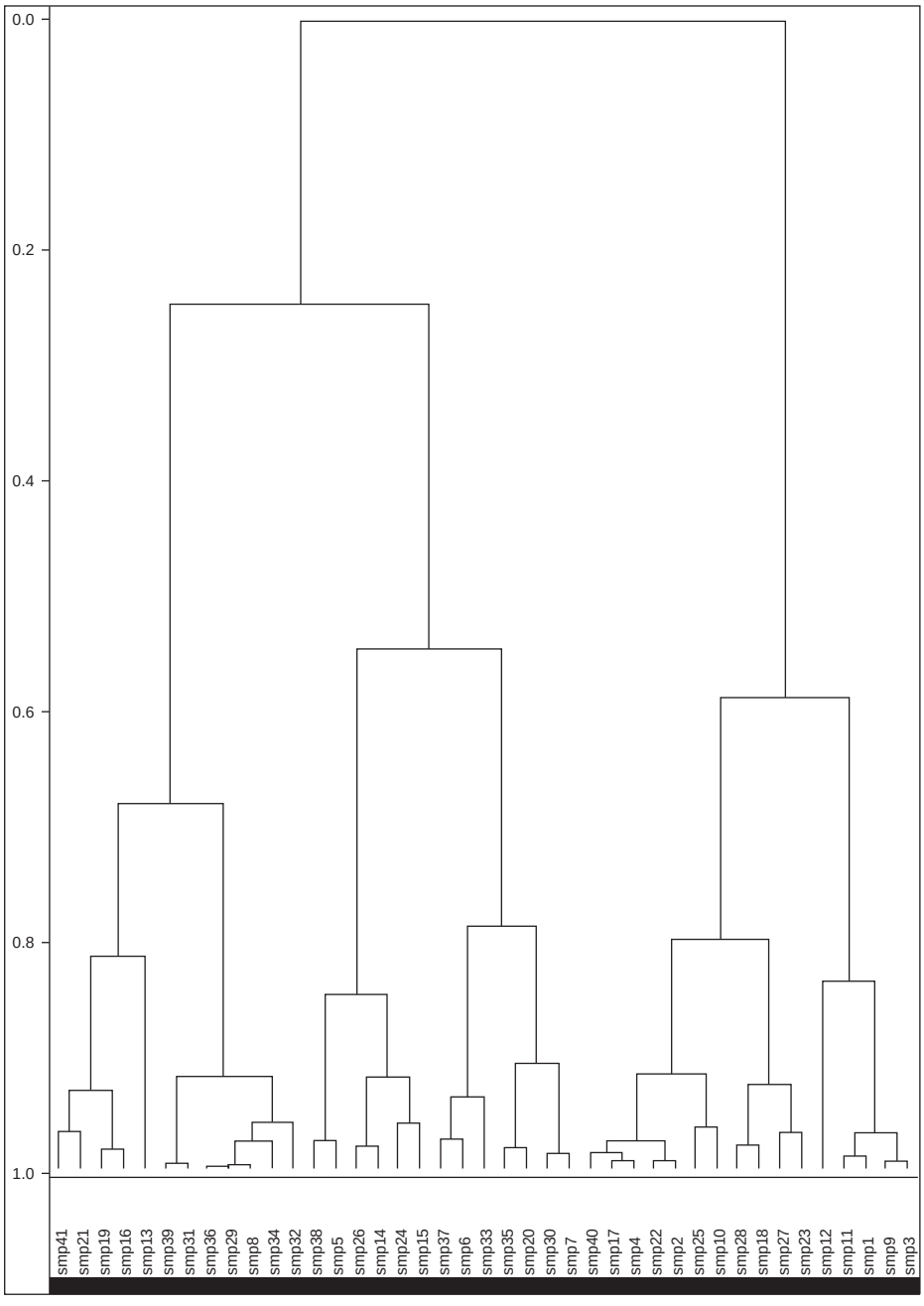

Figure 14 A complete-linkage dendogram of the nose data set. The dendogram was generated using the program Pirouette.

Samples that share a common property (such as fuel type) will be found on one side of the decision surface, whereas those samples comprising the other category will be found on the other side. Nonparametric discriminants have provided insight into relationships contained within sets of chemical measurements. However, classification based on random or chance separation[24] can be a serious problem if the data set is not sample rich. Because chemical data sets usually contain more variables than samples, similarity-based classifiers are generally preferred.

Similarity-based classifiers, e.g. *k*-nearest neighbor (KNN)[25] and soft independent modeling by class analogy (SIMCA),[26–30] treat each chromatogram or spectrum as a data vector $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_j, \ldots, x_p)$ where component x_j is the area of the jth peak or the absorbance value of the jth wavelength. Such a vector can also be viewed as a point in a high-dimensional measurement space. A basic assumption is that distances between points in the measurement space will be inversely related to their degree of similarity.

Using a similarity-based classifier we can determine the class-membership of a sample by examining the class label of the data point closest to it or from the principal component model of the class, which lies closest to the sample in the measurement space. In chemistry, similarity-based classification rules are implemented using either KNN or SIMCA.

4.1 *k*-Nearest Neighbor

For its simplicity, KNN is a powerful classification technique. A sample is classified according to the majority vote of its KNNs, where k is an odd integer (one, three, or five). For a given sample, Euclidean distances are first computed from the sample to every other point in the data set. These distances arranged from smallest to the largest are used to define the sample's KNNs. A poll is then taken by examining the class identities among the point's KNNs. Based on the class identity of the majority of its KNNs, the sample is assigned to a class in the data set. If the assigned class and the actual class of the sample match, the test is considered a success. The overall classification success rate, calculated over the entire set of points, is a measure of the degree of clustering in the set of data. Clearly, a majority vote of the KNNs can only occur if the majority of the measurement variables concur, because the data is usually autoscaled.

KNN cannot furnish a statement about the reliability of a classification. However, its classification risk is bounded. In other words, the Bayes classifier will generate the optimal classification rule for the data, and 1-nearest neighbor has an error rate which is twice as large as the Bayes classifier. (To implement the Bayes classifier, one must have knowledge about all the statistics of the data set including the underlying probability distribution function for each class. Usually, this knowledge is not available.) Hence, any other classification method, no matter how sophisticated, can at best only improve on the performance of KNN by a factor of two.

4.2 Soft Independent Modeling by Class Analogy

In recent years, modeling approaches have become popular in analytical chemistry for developing classification rules because of the problems with nonparametric discriminants. Although there are a number of approaches to modeling classes, the SIMCA method, based on PCA, has been developed by Wold for isolating groups of multivariate data or classes in a data set. In SIMCA, a PCA is performed on each class in the data set, and a sufficient number of principal components are retained to account for most of the variation within each class. Hence, a principal component model is used to represent each class in the data set. The number of principal components retained for each class is usually different. Deciding on the number of principal components that should be retained for each class is important, as retention of too few components can distort the signal or information content contained in the model about the class, whereas retention of too many principal components diminishes the signal-to-noise. A procedure called cross-validation[31] ensures that the model size can be determined directly from the data. To perform cross-validation, segments of the data are omitted during the PCA. Using one, two, three, etc., principal components, omitted data are predicted and compared to the actual values. This procedure is repeated until every data element has been kept out once. The principal component model that yields the minimum prediction error for the omitted data is retained. Hence, cross-validation can be used to find the number of principal components necessary to describe the signal in the data while ensuring high signal-to-noise by not including the so-called secondary or noise-laden principal components in the class model.

The variance that is explained by the class model is called the modeled variance, which describes the signal, whereas the noise in the data is described by the residual variance or the variance not accounted for by the model. (The residual variance is explained by the secondary principal components, which have been truncated or omitted from the principal component model.) By comparing the residual variance of an unknown to the average residual variance of those samples that make up the class, it is possible to obtain a direct measure of the similarity of the unknown to the class. This comparison is also a measure of the goodness of fit of the sample to a particular principal component model. Often, the F-statistic is used to compare the residual variance of a sample with the mean residual variance of the class.[32] Employing the F-statistic, an upper limit for the residual variance can be calculated for those samples belonging to the class. The final result is a set of probabilities of class-membership for each sample.

An attractive feature of SIMCA is that a principal component mapping of the data has occurred. Hence, samples that may be described by spectra or chromatograms are mapped onto a much lower dimensional subspace for classification. If a sample is similar to the other samples in the class, it will lie near them in the principal component map defined by the samples representing that class. Another advantage of SIMCA is that an unknown is only assigned to the class for which it has a high probability. If the residual variance of a sample exceeds the upper limit for every modeled class in the data set, the sample would not be assigned to any of the classes because it is either an outlier or comes from a class that is not represented in the data set. Finally, SIMCA is sensitive to the quality of the

data used to generate the principal component models. As a result, there are diagnostics to assess the quality of the data, such as the modeling power[33] and the discriminatory power.[34] The modeling power describes how well a variable helps the principal components to model variation, and discriminatory power describes how well the variable helps the principal components to classify the samples in the data set. Variables with low modeling power and low discriminatory power are usually deleted from the data because they contribute only noise to the principal component models.

SIMCA can work with as few as 10 samples per class, and there is no restriction on the number of measurement variables, which is an important consideration, because the number of measurement variables often exceeds the number of samples in chemical studies. Most standard discrimination techniques would break down in these situations because of problems arising from collinearity and chance classification.[24]

4.3 Feature Selection

Feature selection is a crucial step in KNN or SIMCA, because it is important to delete features or measurements that contain information about experimental artifacts or other systematic variations in the data not related to legitimate chemical differences between classes in a data set. For profiling experiments of the type that are being considered (see section 4.4) it is inevitable that relationships may exist among sets of conditions used to generate the data and the patterns that result. One must realize this in advance when approaching the task of analyzing such data. Therefore, the problem is utilizing information contained in the data characteristic of the class without being swamped by the large amount of qualitative and quantitative information contained in the chromatograms or spectra about the experimental conditions used to generate the data. If the basis of classification for samples in the training set is other than desired group differences, unfavorable classification results for the prediction set will be obtained despite a linearly separable training set. The existence of these confounding relationships is an inherent part of profiling data. Hence, the goal of feature selection is to increase the signal-to-noise ratio of the data by discarding measurements on chemical components that are not characteristic of the source profile of the classes in the data set. Feature selection in the context of pattern recognition is described in greater detail in the next section by way of the two worked examples.

4.4 Case Studies

Pattern recognition is about reasoning, using the available information about the problem to uncover information contained within the data. Autoscaling, feature selection, and classification are an integral part of this reasoning process. Each plays a role in uncovering information contained within the data.

Pattern recognition analyses are usually implemented in four distinct steps: data preprocessing, feature selection, classification, and mapping and display. However, the process is iterative, with the results of a classification or display often determining a further preprocessing step and reanalysis of the data. Although the procedures selected for a given problem are highly dependent upon the nature of the problem, it is still possible to develop a general set of guidelines for applying pattern recognition techniques to real data sets. In this section, a framework for solving the class-membership problem is presented by way of two recently published studies on chromatographic fingerprinting of complex biological and environmental samples.

4.4.1 Fuel Spill Identification

The first study[35] involves the application of gas chromatographic and pattern recognition (GC/PR) methods to the problem of typing jet fuels, so that a spill sample in the environment can be traced to its source. The test data consisted of 228 gas chromatograms of neat jet fuel samples representing the major aviation fuels (JP-4, Jet-A, JP-7, JPTS, and JP-5) found in the USA. The neat jet fuel samples used in this study were obtained from Wright Patterson Air Force Base or Mulkiteo Energy Management Laboratory (Table 6). They were splits from regular quality control standards, which were purchased by the United States Air Force (USAF) to verify the authenticity of the manufacturer's claims.

The prediction set consisted of 25 gas chromatograms of weathered jet fuels (Table 7). Eleven of the 25 weathered fuels were collected from sampling wells as a neat oily phase found floating on top of the well water. Eleven of the 25 weathered fuel samples were extracted from the soil near various fuel spills. The other three fuel samples had been subjected to weathering in a laboratory.

The neat jet fuel samples were stored in sealed containers at $-20\,^{\circ}C$. Prior to chromatographic analysis, each fuel sample was diluted with methylene chloride

Table 6 Training set

Number of samples	Fuel type
54	JP-4 (fuel used by USAF fighters)
70	Jet-A (fuel used by civilian airliners)
32	JP-7 (fuel used by SR-71 reconnaissance plane)
29	JPTS (fuel used by TR-1 and U-2 aircraft)
43	JP-5 (fuel used by Navy jets)

Table 7 Prediction set

Sample	Identity	Source	Sample	Identity	Source
PF007	JP-4	A[a]	MIX1	JP-4	C[c]
PF008	JP-4	A[a]	MIX2	JP-4	C[c]
PF009	JP-4	A[a]	MIX3	JP-4	C[c]
PF010	JP-4	A[a]	MIX4	JP-4	C[c]
PF011	JP-4	A[a]	STALE-1	JP-4	D[d]
PF012	JP-4	A[a]	STALE-2	JP-4	D[d]
PF013	JP-4	A[a]	STALE-3	JP-4	D[d]
KSE1M2	JP-4	B[b]	PIT1UNK	JP-5	E[e]
KSE2M2	JP-4	B[b]	PIT1UNK	JP-5	E[e]
KSE3M2	JP-4	B[b]	PIT2UNK	JP-5	E[e]
KSE4M2	JP-4	B[b]	PIT2UNK	JP-5	E[e]
KSE5M2	JP-4	B[b]			
KSE6M2	JP-4	B[b]			
KSE7M2	JP-4	B[b]			

[a] Sampling well at Tyndall AFB: the sampling well was near a previously functioning storage depot. Each well sample was collected on a different day.
[b] Soil extract near a sampling well: dug with a hand auger at various depths. Distance between sampling well Tyndall and soil extract was approximately 80 yards.
[c] Weathered fuel added to sand.
[d] Old JP-4 fuel samples that had undergone weathering in a laboratory refrigerator.
[e] Sampling pit at Keywest Air Station: two pits were dug near a seawall to investigate a suspected JP-5 fuel leak.

and injected onto a fused silica capillary column (10 m × 0.10 mm) using a split injection technique. The fused silica capillary column was temperature programmed from 60 °C to 270 °C at 18 °C min^{-1}. High-speed gas chromatograms representative of the five fuel types (JP-4, Jet-A, JP-7, JPTS, and JP-5) are shown in Figure 15.

The gas chromatograms were peak matched using a computer program[36] that correctly assigned peaks by first computing the Kovats retention index (KI)[37] for compounds eluting off the GC column. Because the *n*-alkane peaks are the most prominent features present in the gas chromatograms of these fuels, it is a simple matter to compute KI values. The peak-matching program then developed a template of peaks by examining integration reports and adding peaks to the template that did not match the retention indices of previously observed peaks. By matching the retention indices of the peaks in each chromatogram with the retention indices of the features in the template, it was possible to produce a data vector for each gas chromatogram. Each feature in a peak-matched chromatogram was assigned a value corresponding to the normalized area of the peak in the chromatogram. (If the peak was present, its normalized area from the integration report was assigned to the corresponding element of the vector. If the peak was not present, the corresponding feature was assigned a value of zero.) The number of times a particular peak was found to have a nonzero value was also computed, and features below a user-specified number of nonzero occurrences (which was set equal to 5% of the total number of fuel samples in the training set) were deleted from the data set. This peak-matching procedure yielded a final cumulative reference file containing 85 features, though not all peaks were present in all chromatograms.

Because outliers have the potential to adversely influence the performance of pattern recognition methods, outlier analysis was performed on each fuel class in the training set prior to pattern recognition analysis. The generalized distance test[38] at the 0.01 significance level was implemented via SCOUT[39] to identify discordant observations in the data. Three Jet-A and four JP-7 fuel samples were found to be outliers and were subsequently removed from the data set. The training set, comprising 221 gas chromatograms of 85 peaks each, was analyzed using pattern recognition methods. Prior to pattern recognition analysis, the data were autoscaled so that each feature had a mean of zero and standard deviation of one within the set of 221 gas chromatograms. Hence, each gas chromatogram was initially represented as an 85-dimensional data vector, $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_j, \ldots, x_{85})$, where x_j is the normalized area of the jth peak.

The first step in the study was to apply PCA to the training set data. Figure 16 shows a plot of the two largest principal components of the 85 GC peaks obtained from the 221 neat jet fuel samples. Each fuel sample or gas chromatogram is represented as a point in the principal component map of the data. The JP-7 and JPTS fuel samples are well separated from one another and from the gas chromatograms of the JP-4, Jet-A, and JP-5 fuel samples, suggesting that information about fuel type is

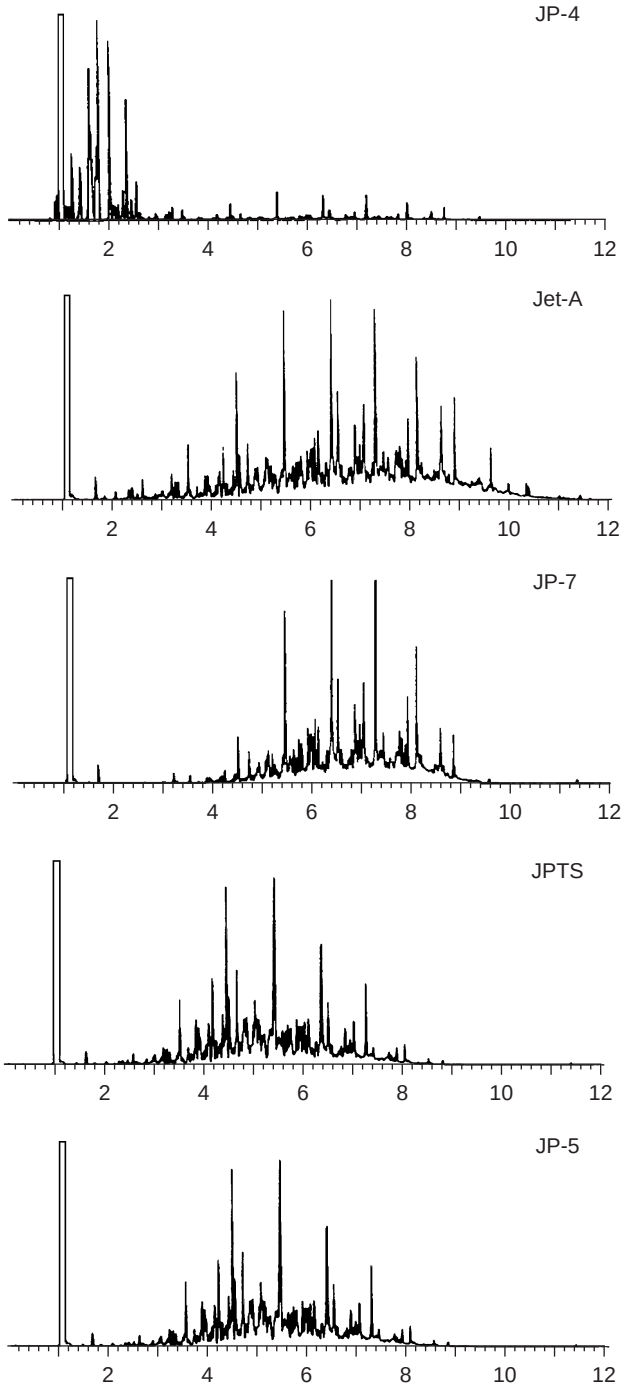

Figure 15 High speed gas chromatograms of JP-4, Jet-A, JP-7, JPTS, and JP-5. (Reproduced with permission from Lavine et al.[35] Copyright 1995 American Chemical Society.)

present in the high speed gas chromatograms of the neat jet fuels. However, the overlap of the JP-5 and Jet-A fuels in the principal component map suggests that gas chromatograms of these two fuels share a common set of attributes, which is not surprising in view of their

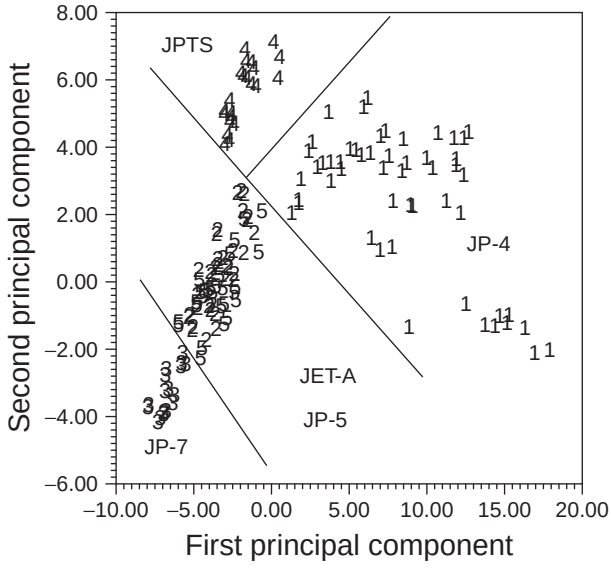

Figure 16 A plot of the two largest principal components of the 85 GC peaks for the 221 neat jet fuel samples. The map explains 72.3% of the total cumulative variance: 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5. (Reproduced with permission from Lavine et al.[35] Copyright 1995 American Chemical Society.)

similar physical and chemical properties.[40] Mayfield and Henley[41] have also reported that gas chromatograms of Jet-A and JP-5 fuels were more difficult to classify than gas chromatograms of other types of jet fuels. Nevertheless, they concluded that fingerprint patterns exist within GC profiles of Jet-A and JP-5 fuels characteristic of fuel type, which is consistent with a plot of the second and third largest principal components of the training set data. The plot in Figure 17 indicates that differences do indeed exist between the GC profiles of Jet-A and JP-5 fuels. However, the second and third largest principal components do not represent the direction of maximum variance in the data. (In fact, they only represent 23.1% of the total cumulative variance or information content of the data.) Hence, it must be concluded that the bulk of the information contained within the 85 GC peaks is not about differences between the GC profiles of Jet-A and JP-5 fuels.

To better understand the problems associated with classifying the gas chromatograms of Jet-A and JP-5 fuels, it was necessary to reexamine this classification problem in greater detail. Figure 18 shows a plot of the two largest principal components of the 85 GC peaks of the 110 Jet-A and JP-5 neat fuel samples. It is evident from an examination of the principal component map that Jet-A and JP-5 fuel samples lie in different regions, suggesting that Jet-A and JP-5 fuels can be differentiated from each other on the basis of their GC profiles. However, the points representing the JP-5 fuels form

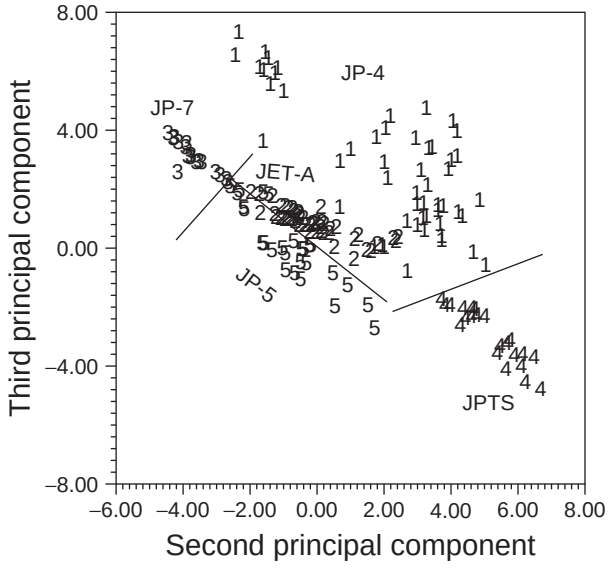

Figure 17 A plot of the second and third largest principal components of the 85 GC peaks for the 221 neat jet fuel samples. The map explains 23.1% of the total cumulative variance: 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5. (Reproduced with permission from Lavine et al.[35] Copyright 1995 American Chemical Society.)

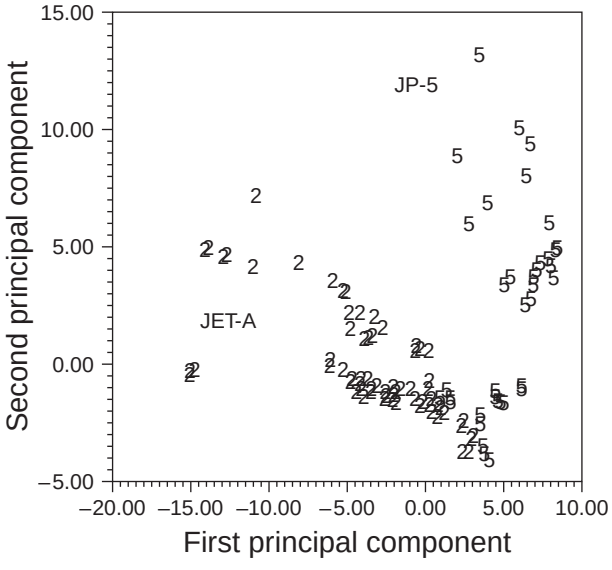

Figure 18 A principal component map of the 110 neat Jet-A and JP-5 fuel samples developed from the 85 GC peaks. The JP-5 fuel samples can be divided into two distinct groups: fuel samples that lie close to the Jet-A fuels and fuel samples distant from the Jet-A fuels. (Reproduced with permission from Lavine et al.[35] Copyright 1995 American Chemical Society.)

two distinct subgroups in the map, which could pose a problem as an important requirement for a successful pattern recognition study is that each class is represented by a collection of samples that are in some way similar. The subclustering suggests a lack of homogeneity among the samples representing the JP-5 fuels. Therefore, it is important to identify and delete the GC peaks responsible for the subclustering of the JP-5 fuels.

The following procedure was used to identify the GC peaks strongly correlated with the subclustering. First, the JP-5 fuel samples were divided into two categories on the basis of the observed subclustering. Next, the ability of each GC peak alone to discriminate between the gas chromatograms from the two JP-5 subclusters was assessed. The dichotomization power of each of the 85 GC peaks was also computed for the following category pairs: JP-5 versus JP-4, JP-5 versus Jet-A, JP-5 versus JP-7, and JP-5 versus JPTS. A GC peak was retained for further analysis only if its dichotomization power for the subclustering dichotomy was lower than for any of the other category pairs. Twenty-seven GC peaks that produced the best classification results when the chromatograms were classified as Jet-A, JPTS, JP-7, or JP-5 were retained for further study.

Figure 19 shows a plot of the two largest principal components of the 27 GC peaks obtained from the 221 neat jet fuel samples. It is evident from the principal component map of the 27 features that the five fuel classes are well separated. Furthermore, the principal component map of the data does not reveal subclustering within any class. This indicates that each fuel class is represented by a collection of samples that are in some way similar when the 27 GC peaks are used as features.

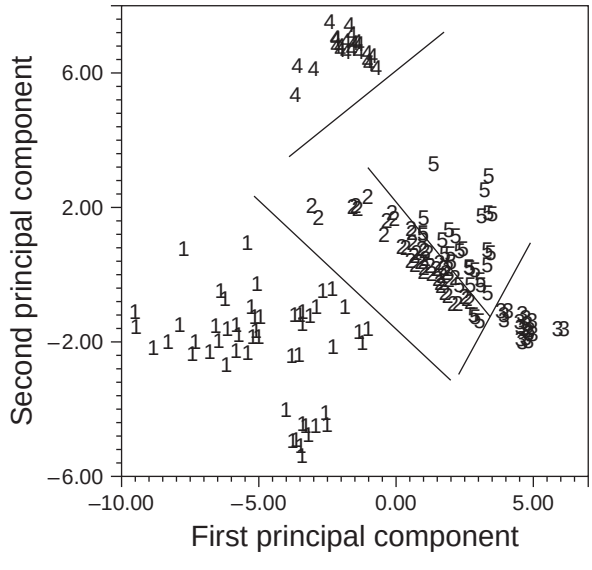

Figure 19 A principal component map of the 27 GC peaks: 1 = JP-4, 2 = Jet-A, 3 = JP-7, 4 = JPTS, 5 = JP-5. (Reproduced with permission from Lavine et al.[35] Copyright 1995 American Chemical Society.)

A five-way classification study involving JP-4, Jet-A, JP-7, JPTS, and JP-5 fuels was also undertaken using SIMCA. A principal component model for each fuel class in the training set was developed from the 27 GC peaks. The complexity of the principal component model was determined directly from the data using the technique of cross-validation. For each class, a single principal component was used to model the data. The gas chromatograms in the training set were then fitted to these models, and the residual – that is the sum of the squares difference between the original gas chromatogram and the chromatogram reproduced by the model – was computed for each gas chromatogram.

Each gas chromatogram in the training set was then classified on the basis of its goodness of fit. The probability of a gas chromatogram belonging to a particular fuel class was determined from its residual for the corresponding principal component model by way of an F-test. A gas chromatogram was assigned to the fuel class for which it had the lowest variance ratio. However, if the variance ratio exceeded the critical F-value for that class, then the sample would not be assigned to it. Results from the five-way classification study involving the training set samples are summarized in Table 8. The recognition rate for JP-4, Jet-A, JP-7, and JPTS fuels is very high. However, Jet-A is more difficult to recognize because of its similarity to JP-5, which is undoubtedly the reason for SIMCA classifying 16 Jet-A fuel samples as JP-5.

The ability of the principal component models to predict the class of an unknown fuel was first tested using a method called internal validation. The training set of 221 gas chromatograms was subdivided into 13 training set – prediction set pairs. Each training set had 204 gas chromatograms and each prediction set had 17 gas chromatograms. The members of the sets were chosen randomly. Furthermore, a particular chromatogram was present in only 1 of the 13 prediction sets generated. Principal component models were developed for each of the training sets and tested on the corresponding prediction set. The mean classification success rate for these so-called prediction sets was 90.5%.

To further test the predictive ability of the 27 GC peaks and the classification models associated with them, an external prediction set of 25 gas chromatograms was employed. The gas chromatograms in the prediction set were run a few months before the neat jet fuel gas chromatograms were run. The results of this study are shown in Table 9. All the weathered fuel samples were correctly classified. This is an important result, as the changes in composition that occur after a jet fuel is released into the environment constitute a major problem in fuel spill identification. These changes may arise from evaporation of lower-molecular-weight alkanes,

Table 8 SIMCA training set results

Class	F-criterion[a]			
	Principal components	Number in class	Correct	Percentage
JP-4	1	54	54	100
Jet-A[b]	1	67	51	76.1
JP-7[b]	1	28	27	96.4
JPTS	1	29	29	100
JP-5	1	43	43	100
Total		221	204	92.3

[a] Classifications were made on the basis of the variance ratio $F = [s_p/S_o]^2[N_q - NC - 1]$, where s_p^2 is the residual of sample p for class i, S_o^2 is the variance of class i, N_q is the number of samples in the class, and NC is the number of principal components used to model the class. A sample is assigned to the class for which it has the lowest variance ratio. However, if the sample's variance ratio exceeds the critical F-value for the class, then the sample cannot be assigned to the class. The critical F-value for each training set sample is $F_\alpha = 0.975$ $[(M - NC),(M - NC)(N_q - NC - 1)]$ where M is the number of measurement variables or GC peaks used to develop the principal component model.

[b] Misclassified Jet-A and JP-5 fuel samples were categorized as JP-5.

Table 9 Prediction set results

Samples	F-values[a]				
	JP-4	JET-A	JP-7	JPTS	JP-5
PF007	3.82	10.04	58.9	12.4	7.43
PF008	3.69	9.62	57.6	12.5	7.14
PF009	3.71	9.84	57.6	12.6	7.32
PF010	3.30	16.7	73.7	11.8	10.8
PF011	3.57	9.64	58.9	12.8	7.39
PF012	4.11	7.74	78.2	13.5	12.04
PF013	4.33	8.19	79.8	12.6	12.3
KSE1M2	2.83	24.4	63.9	30.4	11.21
KSE2M2	2.25	16.2	70.8	21.6	11.09
KSE3M2	2.51	9.41	71.0	17.3	10.2
KSE4M2	2.40	10.11	71.3	17.83	10.4
KSE5M2	2.33	7.76	56.4	17.9	7.61
KSE6M2	1.87	13.4	69.3	20.8	10.4
KSE7M2	2.21	9.85	67.3	18.3	9.78
MIX1	1.33	34.9	71.3	38.2	13.3
MIX2	1.33	11.93	53.3	20.9	7.37
MIX3	1.44	12.3	55.2	20.6	7.71
MIX4	1.59	9.51	48.6	19.9	6.27
STALE-1	1.72	73.7	151.9	54.7	31.5
STALE-2	0.58	28.7	123.8	30.9	22.6
STALE-3	0.541	28.7	127.3	29.9	22.6
PIT1UNK	6.62	1.19	6.106	33.02	0.504
PIT1UNK	6.57	1.15	6.03	32.9	0.496
PIT2UNK	6.51	1.14	6.14	32.8	0.479
PIT2UNK	6.51	1.14	6.27	32.7	0.471

[a] An object is assigned to the class for which it has the lowest variance ratio. However, if the variance ratio exceeds the critical F-value for that class, then the object cannot be assigned to it. Critical F-values of prediction set samples are obtained using one degree of freedom for the numerator and $N_q - NC - 1$ degrees of freedom for the denominator.[32] For JP-4, the critical F-value at $\alpha = 0.975$ is $F(1, 52) = 5.35$, and it is $F(1, 41) = 5.47$ for JP-5.

microbial degradation, and the loss of water-soluble compounds due to dissolution.[42] Because the weathered fuel samples used in this study were recovered from a subsurface environment, loss of lower alkanes due to evaporation will be severely retarded. Furthermore, dissolution of water-soluble components should not pose a serious problem as only a small fraction of the fuel's components is soluble in water.[43] Hence, the predominant weathering factor in subsurface fuel spills is probably biodegradation, which does not appear to have a pronounced effect on the overall GC profile of the fuels. Clearly, weathering of aviation turbine fuels in a subsurface environment will be greatly retarded compared to surface spills, thereby preserving the fuel's identity for a longer period of time.

4.4.2 Africanized Honeybees

GC/PR has also been used to develop a potential method for differentiating Africanized honeybees from European honeybees.[44–47] The test data consisted of 109 gas chromatograms (49 Africanized and 60 European) of cuticular hydrocarbons obtained from bodies of Africanized and European honeybees. Cuticular hydrocarbons were obtained by rinsing the dry or pinned bee specimens in hexane for approximately 15 min. The cuticular hydrocarbon fraction analyzed by gas chromatography was isolated from the concentrated washings by means of a silicic acid column. Hexane was used as the eluent. The extracted hydrocarbons (equivalent to 4% of a bee) were coinjected with authentic *n*-alkane standards. KIs were assigned to compounds eluting off the column. These indices were used for peak identification.

Each gas chromatogram contained 40 peaks corresponding to a set of standardized retention time windows. A typical GC trace of the cuticular hydrocarbons from an Africanized honeybee sample is shown in Figure 20. The GC column had about 5000 plates. The hydrocarbon extract was analyzed on a glass column (1.8 m × 2 mm) packed with 3% OV-17 on Chromosorb® WAW DMCS packing (120–140 mesh).

The gas chromatograms were translated into data vectors by measuring the area of the 40 GC peaks. However, only 10 of the GC peaks were considered for pattern recognition analysis. Compounds comprising these peaks were found in the wax produced by nest bees, and the concentration pattern of the wax constituents is believed to convey genetic information about the honeybee colony. Because the feature selection process was carried out on the basis of a priori considerations, the probability of inadvertently exploiting random variation in the data was minimized.

Each gas chromatogram was normalized to constant sum using the total integrated area of the 40 GC peaks.

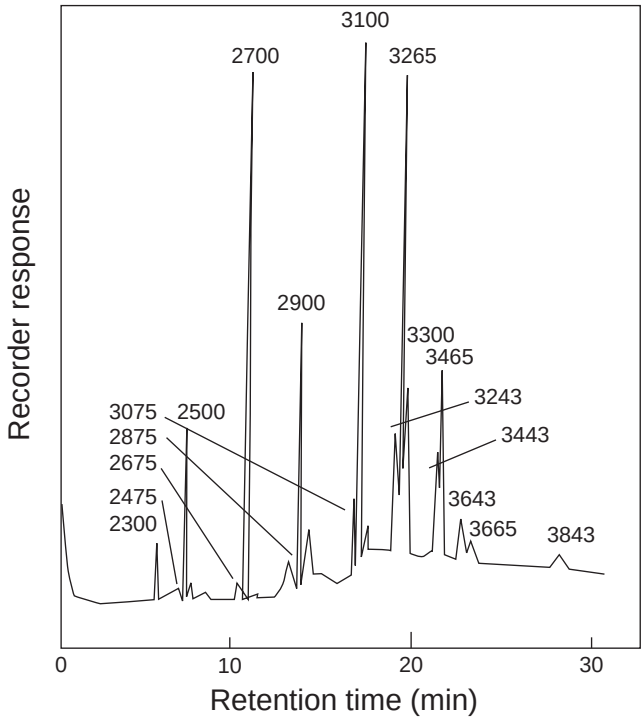

Figure 20 A gas chromatogram of the cuticular hydrocarbons from an Africanized honey bee. The KI values for the peaks used in the pattern recognition study were 2300, 2675, 2875, 3075, 3100, 3243, 3265, 3300, 3443, and 3465. (Reproduced with permission from Lavine and Carlson.[15] Copyright 1987 American Chemical Society.)

Also, the training set data were autoscaled to ensure that each feature had equal weight in the analysis. The normalized and autoscaled data were then analyzed using KNN, which classifies the data vectors in the training set according to a majority vote of its KNNs. Hence, a sample will be classified as an Africanized or European bee only if the majority of its KNNs in the measurement space are Africanized bees. When the 1-nearest neighbor classification rule was applied to the 10 GC peaks, it could correctly classify every chromatogram in the training set. This result indicates that Africanized and European bee specimens are well separated from each other in the feature space defined by the 10 GC peaks.

To test the predictive ability of these descriptors and the classifier associated with them, a prediction set of 55 gas chromatograms (15 Africanized and 40 European) was employed. The distances between the prediction set samples and the samples in the training set were calculated, with class assignments computed in the same manner as in the training phase. Using the 1-nearest neighbor classification rule, a classification success rate of 100% was achieved for the gas chromatograms in the prediction set. This result

is important because it demonstrates that information derived solely from cuticular hydrocarbons can categorize bees as to subspecies. This suggests a direct relationship between the concentration pattern of these compounds and the identity of the subspecies (Africanized or European). Clearly, these results imply that GC/PR can be used to identify the presence of the African genotype in honeybees.

5 SOFTWARE

There are a number of Windows 95/98 software packages sold by commercial vendors that can be used for clustering and classification. UNSCRAMBLER (Camo A/S, Olav Tryggvasonsgt. 24, N-7011 Trondheim, Norway) offers data preprocessing, PCA, SIMCA classification, and graphics in a flexible package. Pirouette (Infometrix Inc., P.O. Box 1528, 17270 Woodinville-Redmond Road NE, Suite 777, Woodinville, WA 98072-1528) has a nice user interface, with good quality graphics. The package has data preprocessing, hierarchical clustering, PCA, KNN, and SIMCA classification. Pirouette, which has been validated according to the United States Food and Drug Administration Standard Operating Procedure, is a good introductory package because of its broad functionality.

SCAN (Minitab Inc., 3081 Enterprise Drive, State College, PA 16801-3008) has PCA, hierarchical clustering, KNN, SIMCA, and discriminant analysis (quadratic, linear, regularized, and DASCO, which is an advanced classification method in chemometrics). The user interface, which is similar to the popular Minitab statistics package, has many advanced editing features, such as brushing. The package is a good mix of statistical and pattern recognition methods. SIRIUS (Pattern Recognition Associates, P.O. Box 9280, The Woodlands, TX 77387-9280) is a graphics oriented package intended for modeling and exploratory data analysis, such as SIMCA and PCA. The PLS TOOLBOX (Eigenvector Technologies, P.O. Box 483, 196 Hyacinth, Manson, WA 89931) is for Matlab and contains routines for PCA, discriminant analysis, and cluster analysis.

6 CONCLUSION

In this article, a basic methodology for analyzing large multivariate chemical data sets is described. A chromatogram or spectrum is represented as a point in a high-dimensional measurement space. Exploratory data analysis techniques (PCA and hierarchical clustering) are then used to investigate the properties of this measurement space. These methods can provide information about trends present in the data. Classification methods can then be used to further quantify these relationships. The techniques, which have been found to be most useful, are nonparametric in nature. As such, they do not attempt to fit the data to an exact functional form; rather, they use the data to suggest an appropriate mathematical model, which can identify structure within the data. Hence, the approach described in this article relies heavily on graphics for the presentation of results, because clustering and classification methods should be used to extend the ability of human pattern recognition to uncover structure in multivariate data. Although the computer can assimilate more data at any given time than can the chemist, it is the chemist, in the end, who must make the necessary decisions and judgements about their data.

ABBREVIATIONS AND ACRONYMS

FCV	Fuzzy Clustering Varieties
GC	Gas Chromatographic
GC/PR	Gas Chromatographic and Pattern Recognition
HPLC	High-performance Liquid Chromatography
KI	Kovats Retention Index
KNN	*k*-Nearest Neighbor
PCA	Principal Component Analysis
SIMCA	Soft Independent Modeling by Class Analogy
USAF	United States Air Force

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis • Underground Fuel Spills, Source Identification

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Chemometrics **(Volume 11)**
Chemometrics

Gas Chromatography **(Volume 12)**
Data Reduction in Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

REFERENCES

1. B.R. Kowalski, 'Analytical Chemistry as an Information Science', *TRACS*, **1**, 71–74 (1988).
2. I.T. Jolliffee, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
3. D.L. Massart, L. Kaufman, *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, John Wiley & Sons, New York, 1983.
4. R.G. Brereton (ed.), *Multivariate Pattern Recognition in Chemometrics*, Elsevier, Amsterdam, 1992.
5. S.D. Brown, 'Chemical Systems Under Indirect Observation: Latent Properties and Chemometrics', *Appl. Spectrosc.*, **49**(12), 14A–31A (1995).
6. L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, New York, 1990.
7. G.L. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York, 1992.
8. J. Mandel, 'The Regression Analysis of Collinear Data', *J. Res. Natl. Bur. Stand.*, **90**(6), 465–476 (1985).
9. P. Thy, K. Esbensen, 'Seafloor Spreading and the Ophiolitic Sequence of the Troodos Complex: A Principal Component Analysis of Lava and Dike Compositions', *J. Geophysical Research*, **98**(B7), 11799–11805 (1993).
10. B.R. Kowalski, T.F. Schatzki, F.H. Stross, 'Classification of Archaeological Artifacts by Applying Pattern Recognition Techniques to Trace Element Data', *Anal. Chem.*, **44**, 2176–2180 (1972).
11. H. Abe, S. Kanaya, Y. Takahashi, S. Sasaki, 'Extended Studies of the Automated Odor-sensing Systems Based on Plural-semiconductor Gas Sensors with Computerized Pattern Recognition Techniques', *Anal. Chim. Acta*, **215**, 155–168 (1988).
12. P.C. Jurs, B.R. Kowalski, T.L. Isenhour, C.N. Reilley, 'Computerized Learning Machines Applied to Chemical Problems: Molecular Structural Parameters from Low Resolution Mass Spectrometry', *Anal. Chem.*, **42**(12), 1387–1394 (1970).
13. J.A. Pino, J.E. McMurry, P.C. Jurs, B.K. Lavine, A.M. Harper, 'Application of Pyrolysis/Gas Chromatography/Pattern Recognition to the Detection of Cystic Fibrosis Heterozygotes', *Anal. Chem.*, **57**(1), 295–302 (1985).
14. A.B. Smith, A.M. Belcher, G. Epple, P.C. Jurs, B.K. Lavine, 'Computerized Pattern Recognition: A New Technique for the Analysis of Chemical Communication', *Science*, **228**(4696), 175–177 (1985).
15. B.K. Lavine, D. Carlson, 'European Bee or Africanized Bee? Species Identification Through Chemical Analysis', *Anal. Chem.*, **59**(6), 468A–470A (1987).
16. P. Geladi, H. Grahn, *Multivariate Image Analysis*, John Wiley and Sons, New York, 1996.
17. W.H.A. van den Broek, D. Wienke, W.J. Melssen, R. Feldhoff, T. Huth-Fehre, T. Kantimm, L.M.C. Buydens, 'Application of a Spectroscopic Infrared Focal Plane Array Sensor for On-line Identification of Plastic Waste', *Appl. Spectrosc.*, **51**(6), 856–865 (1997).
18. P. Robert, D. Bertrand, M.F. Devaux, A. Sire, 'Identification of Chemical Constituents by Multivariate Near-infrared Spectral Imaging', *Anal. Chem.*, **64**, 664–667 (1992).
19. D.D. Coomans, D.I. Broeckaert, *Potential Pattern Recognition in Chemical and Medical Decision-making*, Research Studies Press Ltd, Letchworth, England, 1986.
20. B.K. Lavine, 'Signal Processing and Data Analysis', in *Practical Guide to Chemometrics*, ed. S.J. Haswell, Marcel Dekker, New York, 1992.
21. F.W. Pjipers, 'Failures and Successes with Pattern Recognition for Solving Problems in Analytical Chemistry', *Analyst*, **109**(3), 299–303 (1984).
22. P.D. Wasserman, *Neural Computing*, Van Nostrand Reinhold, New York, 1989.
23. J. Zupan, J. Gasteiger, *Neural Networks for Chemists*, VCH Publishers, New York, 1993.
24. B.K. Lavine, P.C. Jurs, D.R. Henry, 'Chance Classifications by Nonparametric Linear Discriminant Functions', *J. Chemom.*, **2**(1), 1–10 (1988).
25. B.R. Kowalski, C.F. Bender, 'Pattern Recognition. A Powerful Approach to Interpreting Chemical Data', *J. Am. Chem. Soc.*, **94**, 5632–5639 (1972).
26. B.R. Kowalski, S. Wold, 'Pattern Recognition in Chemistry', in *Classification, Pattern Recognition and Reduction of Dimensionality*, eds. P.R. Krishnaiah, L.N. Kanal, North Holland, Amsterdam, 1982.
27. M. Sjostrom, B.R. Kowalski, 'A Comparison of Five Pattern Recognition Methods based on the Classification Results from Six Real Data Bases', *Anal. Chim. Acta*, **112**, 11–30 (1979).
28. B. Soderstrom, S. Wold, G. Blomqvist, 'Pyrolysis Gas Chromatography Combined with SIMCA Pattern Recognition for Classification of Fruit-bodies of Some Ectomycorrhizal Suillus Species', *J. Gen. Microbiol.*, **128**, 1783–1794 (1982).
29. S. Wold, 'Pattern Recognition: Finding and Using Regularities in Multivariate Data', in *Food Research and Data Analysis*, eds. H. Martens, H. Russwurm, Applied Science, Essex, England, 1983.
30. S. Wold, 'Pattern Recognition by Means of Disjoint Principal Component Models', *Patt. Recog.*, **8**, 127–139 (1976).
31. S. Wold, 'Cross-validatory Estimation of the Number of Components in Factor and Principal Components Models', *Technometrics*, **20**, 397–406 (1978).
32. S. Wold, C. Albano, U. Edlund, K. Esbensen, S. Hellberg, E. Johansson, W. Lindberg, M. Sjostrom, 'Pattern Recognition by Means of Disjoint Principal Component Models (SIMCA). Philosophy and Methods', in *Proc. Symp. Applied Statistics, NEUCC*, eds. A. Hoskuldsson, K. Esbensen, RECAU and RECKU, Copenhagen, 183–218, 1981.

33. S. Wold, C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, S. Hellberg, E. Johansson, W. Lindberg, M. Sjostrom, 'Multivariate Data Analysis in Chemistry', in *Chemometrics, Mathematics and Statistics in Chemistry*, eds. B.R. Kowalski, D. Reidel Publishing Company, Dordrecht, 1984.
34. S. Wold, M. Sjostrom, 'SIMCA, a Method for Analyzing Chemical Data in Terms of Similarity and Analogy', in *Chemometrics, Theory and Application*, ed. B.R. Kowalski, American Chemical Society, Washington, DC, 1977.
35. B.K. Lavine, H. Mayfield, P.R. Kroman, A. Faruque, 'Source Identification of Underground Fuel Spills by Pattern Recognition Analysis of High-speed Gas Chromatograms', *Anal. Chem.*, **67**, 3846–3852 (1995).
36. H.T. Mayfield, W. Bertsch, 'An Algorithm for Rapidly Organizing Gas Chromatographic Data into Data Sets for Chemometric Analysis', *Comput. Appl. Lab.*, **1**, 13–137 (1983).
37. E. Kovats, in *Advances in Chromatography*, eds. J.C. Giddings, R.A. Keller, Marcel Dekker, New York, Vol. 1, 230, 1965.
38. R.A. Johnson, D.W. Wichern, 'Applied Multivariate Statistical Analysis', Prentice Hall, Englewood Cliffs, New Jersey, 1982.
39. M.A. Stapanian, F.C. Garner, K.E. Fitzgerald, G.T. Flatman, J.M. Nocerino, 'Finding Suspected Causes of Measurement Error in Multivariate Environmental Data', *J. Chemom.*, **7**, 165–176 (1993).
40. *Handbook of Aviation Fuel Properties*, Coordinating Research Council, Inc., Atlanta, GA, 1983.
41. H.T. Mayfield, M. Henley, in *Monitoring Water in the 1990's: Meeting New Challenges*, eds. J.R. Hall, G.D. Glayson, American Chemical Society for Testing of Materials, Philadelphia, PA, 1991.
42. J.C. Spain, C.C. Sommerville, L.C. Butler, T.J. Lee, A.W. Bourquin, 'Degradation of Jet Fuel Hydrocarbons in Aquatic Communities', USAF Report ESL-TR-83-26, AFESC, Tyndall AFB, 1983.
43. W.E. Coleman, J.W. Munch, R.P. Streicher, H.P. Ringhand, W.F. Kopfler, 'Optimization of Purging Efficiency and Quantification of Organic Contaminants from Water using a 1-L Closed Looped Stripping Apparatus and Computerized Capillary Columns', *Arch. Environ. Contam. Toxicol.*, **13**, 171–180 (1984).
44. B.K. Lavine, D.A. Carlson, D.R. Henry, P.C. Jurs, 'Taxonomy Based on Chemical Constitution: Differentiation of Africanized Honeybees from European Honeybees', *J. Chemomet.*, **2**(1), 29–38 (1988).
45. B.K. Lavine, A.J.I. Ward, R.K. Smith, O.R. Taylor, 'Application of Gas Chromatography/Pattern Recognition Techniques to the Problem of Identifying Africanized Honeybees', *Microchem. J.*, **39**, 308–316 (1989).
46. B.K. Lavine, D.A. Carlson, 'Chemical Fingerprinting of Africanized Honeybees by Gas Chromatography/Pattern Recognition Techniques', *Microchem. J.*, **42**, 121–125 (1990).
47. B.K. Lavine, A.J.I. Ward, J.H. Han, R.K. Smith, O.R. Taylor, 'Taxonomy Based on Chemical Constitution: Differentiation of Heavily Africanized Honeybees from Moderately Africanized Honeybees', *Chemolab.*, **8**, 239–243 (1990).

Multivariate Calibration of Analytical Data

Svante Wold
Research Group for Chemometrics, Umeå University, Sweden

Mats Josefson
AstraZeneca R&D, Mölndal, Sweden

1 Introduction **2**
1.1 Univariate Calibration 3
1.2 Standard Curve, Predictions, Measures of Uncertainty 3
1.3 Multivariate Calibration; Many Signals 3
1.4 Interferences 3
1.5 Direct and Indirect Calibration 4
1.6 The Concentrations of Two Drug Compounds Determined by Ultraviolet Spectroscopy 4
1.7 Mathematical and Statistical Formulation 5

2 Notation and Abbreviations Used in Multivariate Calibration **5**
2.1 Mathematical Notation 5

3 Problem Formulation and Choice of Analytical Method **5**
3.1 How Much of a Known Compound is Present in the Given Sample? 5
3.2 Prepared Reference Samples, or Use of a Reference Method 6
3.3 Multivariate Calibration for Explorative Work 6
3.4 Multivariate Classification of Samples in Narrow Concentration Intervals 6
3.5 One or Several Analytes 6
3.6 The Chemical Analytical Method 7
3.7 Spectral or Time Domain 8
3.8 Chromatography 8
3.9 Combination Methods 8
3.10 Sensor Arrays (Electronic Noses, Crystal Sensors) 9
3.11 Kinetic Curves 9
3.12 Images and Other Measured Data 9

4 Selection of the Calibration Set **9**
4.1 Validation Set (Test Set) 10
4.2 Prediction Set 10

5 Making the Measurements **10**
5.1 Uncontrolled Variables 10
5.2 Ambiguous Spectral Regions (Variables) 10
5.3 Randomization of Measuring Order 10
5.4 Data Collection and Administration 10

6 Initial Quality Control of the Data **11**
6.1 Deviations of Individual Spectra from the Average 11
6.2 Multivariate Quality Control of the Data by Principal Components Analysis 11
6.3 Outliers in X and Unbalanced X-data 11
6.4 Outliers in Y 12

7 Data Preprocessing **12**
7.1 Choice of Representation (Time or Frequency Domain) 12
7.2 Logarithms and Other Stabilizing Transformations 13
7.3 Spectral Correction Methods 13

8 The Analysis of the Training Set Data; the Calibration Model Development **15**
8.1 The Mathematical Formulation of Multivariate Calibration 15
8.2 Linear or Nonlinear Calibration Model 16
8.3 One or Several Analytes 16
8.4 The Calibration Model 16

9 Method Performance and Validation **20**
9.1 Measures of Model Performance 20
9.2 Selectivity 20
9.3 Accuracy 20
9.4 Precision 21
9.5 Sensitivity 21
9.6 Model Fit or Linearity 21
9.7 Range 22
9.8 Robustness 22
9.9 System Suitability Testing 22
9.10 Validating the Calibration Model (Including its Robustness) 22

10 Calibration Transfer (Between Instruments, Labs, and Over Time) **23**

11 Summary and Conclusions **24**

List of Symbols **24**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Abbreviations and Acronyms 24
Related Articles 25
References 25

Multivariate calibration (MVC) is a methodology for using multiple signals, for instance a digitized spectrum, to determine the levels of concentrations of chemical compounds in analytical samples. MVC can also be used to determine other properties of interest, for instance viscosity, particle size distribution, energy content, or taste. MVC is made in two phases. In the first, the ''training'' or ''calibration'' phase, samples with known concentration (property) values and their signal profiles are used to develop a model of their relationship, a multivariate standard curve. In the second phase, this model is used with new samples to determine their concentration (property) values from their signal profiles.

1 INTRODUCTION

The calibration of an analytical instrument means the construction of a quantitative relationship between, on the one hand the instrument ''signals'' as they are measured on analytical samples, and on the other hand one or several properties of the samples, usually concentrations of analytes in the samples.

In univariate calibration (UVC) the amplitude of one signal, e.g. the absorption of light at a certain wavelength, is related to the concentration of one analyte. In MVC, several signals, for instance from a whole spectrum digitized at regularly spaced wavelengths, are used to derive a multivariate model (a generalized standard curve).[1] This model simultaneously relates the amplitudes of *all* the signals, or a substantial part of them, to the concentrations of one or several analytes in the samples. The multivariate model is then used to estimate the analyte concentrations in new samples from the multiple signals measured on these. The use of MVC and spectral profiles to ''measure'' concentrations or other properties is often referred to as ''indirect measurement'' in contrast to the direct measurement of, say, concentration by means of titration, or precipitation and weighing.

In both UVC and MVC, the deviations between the model and the data are used to derive statistical measures of uncertainty of the model and of the estimated concentrations of the new samples. In MVC, the multivariate information can be used for additional statistical diagnostics such as the similarity between a new sample and the calibration set (CS), the presence of clusters in the data, and the relative information content of the ''signals'' (predictor variables).

The relation between spectroscopic signals and concentrations is fairly linear as long as the variation of concentration is moderate. This is often referred to as Lambert–Beer's law. Most calibrations are limited to concentration intervals where a linear standard curve is adequate, but the extension to mildly nonlinear curves are straightforward using, for instance, quadratic polynomials as the functional basis.

MVC has six steps, the first five of which comprise the training phase, and the sixth is the prediction phase. These steps are:

1. Specification of the analytes with concentration ranges. Selection of the instrumental method, including the range of wavelengths, reflectance or transmission mode, etc.
2. Selection of a representative set of calibration samples – the training set (TS) or CS. This CS should span the range of analyte concentrations and also the concentration ranges of interferents.
3. The multivariate signals (typically the digitized spectra) are recorded for the calibration samples and stored in an appropriate data base. The analyte concentrations are measured by a reference method.
4. The data are investigated for the presence of outliers and other anomalies. Thereafter, the data are preprocessed and transformed to a form suitable for the subsequent data analysis.
5. The calibration model is developed and optimized. This includes checking for linearity, and the determination of selectivity, detection limits, precision, accuracy, and other measures of performance. Statistical measures of uncertainty are calculated and used to construct confidence intervals for predicted values. Also, the model is interpreted chemically, important variables (wavelength regions) looked at, interferences are identified, etc.
6. The model is used to estimate the analyte concentrations in new samples (prediction set), including confidence intervals. Diagnostics for dissimilarity (outliers) are checked.

MVC has a much improved precision and selectivity compared to UVC. Also MVC can handle complicated samples with unknown interfering compounds, and it works even when there is no selective wavelength region for the analyte. This has made MVC particularly useful with nonselective spectral methods, such as near-infrared spectroscopy (NIR). In a slightly generalized sense, MVC can be applied to the relation between any instrumental ''profile'', including chromatograms, kinetic curves, thermo-gravimetric curves, sound spectra, etc. and any properties of the analyzed sample. The

properties can be other than concentrations, for instance viscosity and molecular weight of polymer samples, the energy content of oil, gasoline, coal, or peat, or the taste of cheese, wine, or beer. Hence the MVC methodology provides exciting possibilities for the indirect measurement of complicated properties of complicated samples.

1.1 Univariate Calibration

The idea of calibration is simple and powerful. The instrument is given a training or CS of samples with known concentration (c_i) of the analyte, and the amplitude of the instrument signal (z_i) is recorded for each sample, i. The resulting data ($z_i, c_i, i = 1, 2, \ldots, n$) are then used to construct a standard curve (usually a straight line), which is used for new samples to predict the concentration for new samples from the amplitude of the signal measured on these samples (Figure 1).

A prerequisite for calibration to work is that the signal amplitude really changes in a reproducible way with the analyte concentration. The greater this change is for a unit change of concentration, the greater is the sensitivity of the instrument with respect to the analyte. Moreover, the signal must not be much affected by other components in the sample, it must be fairly selective.

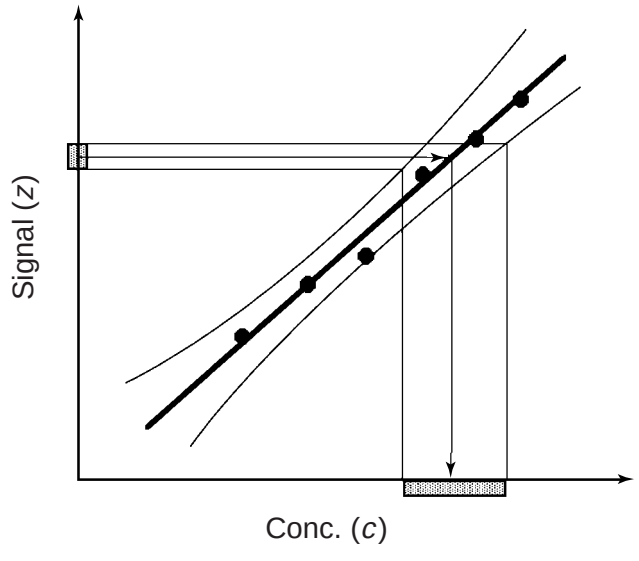

Figure 1 The CS (filled circles) of samples with known concentrations are used to derive a standard curve (here straight line). For new samples, the concentrations can be predicted from the amplitude of the signal measured on these samples by moving horizontally from the signal axis to the standard curve and then vertically down to the concentration axis. Owing to the noise in the data, the standard curve has some uncertainty in slope and intercept, which together with the imprecision of the measured signal, causes an uncertainty in the predicted concentration value.

1.2 Standard Curve, Predictions, Measures of Uncertainty

Normally in UVC, the signal amplitude (z) is used as the y-variable and the concentration (c) used as $\boldsymbol{x}$ in the calculation of the standard curve with intercept a and slope b (Equation 1).

$$z_i = a + bc_i + e_i \tag{1}$$

Least squares (linear regression) is normally used for this calculation.[2–4] The residuals e_i are calculated, with their variance (s^2) and standard deviation (SD), s, (Equations 2 and 3).

$$e_i = z_{i_{\text{observed}}} - z_{i_{\text{model}}} \tag{2}$$

$$s = \sqrt{\sum_i \frac{e_i^2}{n-1}} \tag{3}$$

This residual standard deviation (RSD) should be the same as the precision of measurement within statistical uncertainty. Regression[2–4] is then used to calculate standard errors and confidence intervals for the parameters, as well as for the concentrations of new samples (calculated from their signal amplitudes and the standard curve). If the RSD is substantially different at different concentrations, a weighted regression may be needed for an efficient data analysis. It should be noted that the estimation of the parameters (a and b in Equation 1) by regression, is not based on the assumption of constant and normally distributed noise. The sizes of their confidence intervals depend, however, on the distribution of the noise.

1.3 Multivariate Calibration; Many Signals

In MVC the *whole* spectral profile is used for the calibration instead of just the signal at one single wavelength.[1,5] The signal data now consists of a matrix, $\mathbf{Z}$ (dimension $n \times K$). In MVC, one can calibrate on a single analyte, i.e. $\mathbf{y}$ is an ($n \times 1$) vector, or one can simultaneously calibrate on several (M) analytes. In the latter case, the concentrations of the CS form an ($n \times M$) matrix, $\mathbf{Y}$. The latter is discussed in sections 3.5 and 8.3.

1.4 Interferences

In most calibrations, the samples contain compounds other than the actual analyte, so-called interferents. These other compounds will most likely also affect the signal, and hence the measured signal will have a systematic bias, an interference. UVC thus works well only when the signal is close to selective, and is affected only by the change of the analyte concentration. This often necessitates the use of chromatography or some other separation method

to clean up the sample from interfering compounds. MVC is affected much less by the presence of other compounds due to its ability to find nearly selective combinations of the multiple signals that model and predict the analyte concentration(s). A preprocessing of the data to enhance this selectivity may still be needed (see section 7).

A second problem is that any calibration model is valid only for samples that are similar to the calibration samples. Thus for example, a model developed for gold in seawater samples, does not give accurate results for fresh water samples, and even less well with samples of orange juice. Hence one should check that each new sample is indeed similar to the calibration samples. Unfortunately, there is nothing in the single signal employed in UVC that is helpful for this matter, and the similarity must be checked in some other way, or accepted by faith. In MVC, however, the profile of the multiple signals contains information also about the similarity between the samples. Hence, MVC provides an ''autodiagnosis'' for the similarity between the new samples and the calibration samples, which greatly improves the reliability of the results.[1]

1.5 Direct and Indirect Calibration

If the spectra of all the pure constituents are available (denoted by r_{jk} for the spectral amplitude of constituent j at wavelength k), the concentrations c_{ij} in sample i can be estimated from the spectrum z_i' of a mixture of the analytes by means of multiple regression using the linear model (Equation 4):

$$z_i = \sum_j r_j c_{ij} + e_j \tag{4}$$

This is often called direct calibration because the causal physical model (concentration → spectrum) and the statistical model ($z = \mathrm{f}(c)$) coincide.[1] This direct approach works when (a) *all* constituents in a given sample are known together with their spectra, and (b) the spectrum of any constituent is the same when it is pure and when it is present in the mixture (sample). This makes direct calibration work for simple and rather dilute gas samples and some simple dilute liquid samples, but less well or not at all with complicated samples where many constituents are unknown (interferents) and where the spectra change with concentration. This is exemplified by samples of grain, blood, tissue, lake sediment, polymer formulations, pharmaceutical formulations, wood, pulp, etc. where indirect MVC is the only feasible approach.

The MVC model is formulated inversely to the direct model above, with the explicit objective to predict the concentrations (**Y**) from the digitized and preprocessed spectra (**X**), (Equation 5):

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F} \tag{5}$$

Here the information flow (from **X** to **Y**) is the opposite to the causal model (**Y** → **X**), which made this be called the ''indirect calibration'' model. This inverse formulation of the calibration model was first proposed by Krutchoff,[6] causing much initial consternation and controversy. With the work of Hoadley,[7] Naes,[8] and others, however, we have learnt to distinguish between the direction of the information flow and the direction of causality, and (indirect) MVC is today established and accepted as a good working approach.

The mechanism of inverse MVC is, quite simply, that as long as the correlation structure between **X** (signals) and **Y** (concentrations, properties) remains, then the predictions can be made in either direction **X** → **Y**, or **Y** → **X**. The principle of MVC is to utilize this fact in the direction **X** → **Y** (signals predict concentrations or properties). The correlation structure of (**X**, **Y**) is constant as long as the samples are similar to each other, but it changes when very different samples are brought in. Hence the CS must be representative for the future samples for which the concentrations or other properties are to be determined, see section 4.

1.6 The Concentrations of Two Drug Compounds Determined by Ultraviolet Spectroscopy

To illustrate the calculations and the interpretation of resulting parameters, we shall use a small example with ultraviolet (UV) spectra of samples containing two drug substances. There is a set of $n = 10$ spectra of mixed standards, where all combinations of three levels of two substances (felodipine and metoprolol) are included. One of the samples is included in duplicate (#6 and 7). Three additional samples are included as a prediction set. In addition, the spectra of the two pure substances are included, showing overlap with no selective region for metoprolol, but a selective region of felodipine with a maximum at around variable 58, corresponding to around 363 nm (Figure 2).

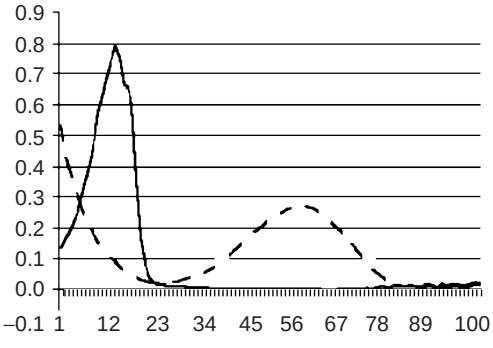

Figure 2 UV Spectra of felodipine (- - - -) and metoprolol (——) (pure samples 11 and 12).

UV spectra in the wavelength range 250 to 450 nm have been recorded, and digitized at every 2 nm giving $K = 101$ spectral variables per sample. The concentration intervals are 20–29 $\mu g\,L^{-1}$ for the first substance (felodipine) and 191–239 $\mu g\,L^{-1}$ for the second (metoprolol).

1.7 Mathematical and Statistical Formulation

Assuming that the signal amplitudes change proportionally to the concentrations (Lambert–Beer's law), the calibration model will be linear, with **B** a matrix of parameters to be determined in the calibration stage.

$$\mathbf{Y} = \mathbf{ZB} + \mathbf{F} \tag{6}$$

The model (Equation 6) works as a multivariate standard curve, where the parameters **B** define directions in the Z-space (the space of the spectral variables) which are linearly related to the **Y**-variables, i.e. the analyte concentrations or other properties of interest.

The **Y**-residual vector or matrix **F** contains the deviations between model and the concentration/property values (**Y**), and should have small elements everywhere. The size of the elements in **F** is determined by the precision of the instrument, the experimental control, the adequacy of the linear model and the method of data analysis. These residuals can be inspected for large values (outliers) and systematic patterns, which indicate model inadequacies. The SD of these **Y**-residuals should not greatly exceed the known precision of the used method and instrument.

In the case where Equation (6) is estimated by a projection model (section 8) this allows also the estimation of residuals of the (preprocessed) signals, **X**. These residuals, denoted **E**, provide indications of spectral areas with large and small information content.

2 NOTATION AND ABBREVIATIONS USED IN MULTIVARIATE CALIBRATION

2.1 Mathematical Notation

Vectors are denoted by bold lower-case letters, e.g. **t**, and matrices by bold capital letters, e.g. **X**. Vectors are assumed to be column vectors unless indicated by a prime (transpose), e.g. **p′**. The length (norm) of a vector or a matrix is denoted by $\|\mathbf{v}\|$ or $\|\mathbf{X}\|$ respectively. This length divided by the square root of the number of elements, often corrected for the degrees of freedom, is the SD of the vector or the matrix.

For notation see the list of symbols.

3 PROBLEM FORMULATION AND CHOICE OF ANALYTICAL METHOD

MVC provides an improved set of tools for analytical chemistry compared to UVC. There is no longer a need to find the ''ultimate variable'' with as little interference as possible. One or a few analytical techniques yielding several variables that carry relevant information for the analysis are preferred. Interferences can be compensated for, a measure of the appropriateness of using an MVC for a given sample is available, and the correlation patterns among the variables are a starting point for new knowledge.

By treating analytical data from a given sample as a profile of variables, where the profile is used as a total signature of the sample, there is less need for special optimization regarding the protection of the single variable from interfering compounds and processes. For efficient use, the focus should be on analytical methods that give several variables with a short measurement time. The measurements should be made using best analytical chemistry practice.

A number of questions are closely related to the MVC problem, but we will here only discuss the first, namely (1) How much of a known compound is present in the given sample? Another, more demanding question is: (2) Which compounds are present in the given sample? An even more elusive type of questions is: (3) Will this intermediate product yield a good-quality end product?; e.g. Will this wine be better with storage? Finally, there is a related question of classification that in some cases is easier to answer if the amount of reference samples is limited: (4) Is the analytical profile in the present sample similar to that of the previous samples?

3.1 How Much of a Known Compound is Present in the Given Sample?

The type 1 question has traditionally been answered with UVC. The use of UVC in this case implies that the signal is selective, i.e. the interfering compounds in the sample matrix must have a low influence on the magnitude of the signal. This is today often accomplished by a chromatographic separation step. By using MVC it is possible to omit the chromatographic separation before a spectrometric detection. This will save run time for a long series of determinations and even allow for direct measurements on solid materials, thus avoiding the time spent for sample preparation. The recognition of aberrant samples usually attained by visual inspection of chromatograms, can be automated in MVC. The spectral residual profile provides this ability.

3.2 Prepared Reference Samples, or Use of a Reference Method

We can distinguish two calibration situations, the first where reference samples can be prepared with desired levels of analytes and interferents, and the second where samples are collected from nature or some process of interest, and analyzed by a reference method. In statistics this is referred to as the concentrations being fixed or random, respectively.

The first situation is illustrated by the example (section 1.6), where two known substances are weighed and dissolved to make reference samples of known composition. The concentration levels in these samples should then be laid out according to a statistical design, typically a full or fractional grid design with five or more levels. The second situation is typical for complicated samples, such as samples of grain where one wishes to determine the concentrations of protein, fat, and water. Then many samples are needed in the TS to ensure the proper spanning of all important directions in the concentration space, and to compensate for the imprecision of the reference method.

The reference method should have good accuracy and precision, providing accurate reference values for the TS and the validation set. Effectively the second type of MVC involves an inter-calibration of two different analytical methods that puts both the MVC and the reference method in a critical light. This is not always in favor of the reference method and it may need improvement in order to yield high-quality data.

3.3 Multivariate Calibration for Explorative Work

Often the objective is not to build a precise calibration but rather to explore the correlations between variables investigated in development work. In this case the MVC is used in an inductive way to make steps forward in the building of knowledge during explorative work. Here the resulting information is the structure of the calibration model rather than the specific amount of a compound. This use of MVC is less obvious but very important in industrial development work.

A much more rough calibration can then be made based on a minimal number of samples, just to give an idea about the feasibility of the methodology. Here the minimum number of samples is five, and these samples must also be well distributed over the range of interest. With so few samples and with many spectral variables, no extensive preprocessing can be afforded, and the partial least squares (PLS) projection to latent structures approach is preferred.

3.4 Multivariate Classification of Samples in Narrow Concentration Intervals

In some cases it less practical to make a calibration, for instance, when the properties of an industrial product are going to be monitored. The product is produced according to a specification aiming at a constant set of properties. Then a large set of samples is available with concentrations close to the specification in a very narrow distribution. Samples outside this narrow distribution may not be available from the production. Then calibration samples outside the specification have to be manufactured at lab scale in order to span a variation of the concentration that is needed for the TS. This may be feasible for liquid solutions with few components, but it is more involved for solid samples such as powders and pharmaceutical tablets.

An alternative approach in this case is to use multivariate classification.[9] Then the classification is used as a positive verification of the presence of the correct concentration in a certain concentration interval. In the method validation step, concentrations outside this interval have to be tested and should fail a positive identity. Analogously to the residual variance in MVC, the classification residuals indicate aberrant spectra resulting from deviations from the specified composition profile.[10,11]

Using multivariate classification in this way may save a lot of development time when it is costly to create concentration variations for the calibration samples, while it is still possible to get the benefit of using spectral residuals to find the sources for variations when they occur.

3.5 One or Several Analytes

When PLS is used in the MVC, it is possible to have more than one response from the same calibration model. This is useful in cases where the responses are correlated. When the reference method for one or a few correlated responses is performing less well, the use of a joint model can support the prediction of these responses. This is useful in explorative work when the most information has to be extracted from limited amounts of data.

When uncorrelated responses are used in the same model, this will need more PLS components than the corresponding single-response calibration models. Hence it is better in this case to develop separate PLS models for each response. Typical situations with correlated responses are examplified by isocratic chromatography where the peak widths are correlated with the peak retention time, by environmental analysis where pollutants often vary together, and by kinetic measurements where the amount of reacted compound at time j is also correlated with the amount of reacted compound at time $j-1, j-2$, etc.

3.6 The Chemical Analytical Method

Analyte concentrations can be determined in various types of phases, i.e. transparent and turbid liquids, solids, gases, and plasma. Traditionally, most analyses have been performed in the liquid state and in some cases in the gas phase. If the sample came as a solid, it would be dissolved in an appropriate solvent and then analyzed. In chromatographic and electrophoretic separations the sample is also transported in solution or in the gas phase.

The technology for measuring absorbance by light passing through a liquid or gas sample cell is well established and yields close to linear responses in the appropriate concentration range. These optical transmission techniques are applicable for ultraviolet/visible (UV/VIS), NIR, infrared (IR), and their combinations with separation methods.

Transmission spectroscopy is limited by the amount of light that is coming through the sample. If the sample in the sample cell is absorbing too much light in some wavelength regions, the measurement at those wavelengths will be of little analytical value since stray light in the instrument will be the dominating feature. This may be adjusted by the cell path length. Technically it is more difficult to get a close to linear response at short wavelengths in UV below 210 nm due to absorption of glass and Rayleigh scattering due to dust. However, this is a range that is commonly used for amino acids and proteins.

When water is a solvent, IR spectrometry has to be done with very short path lengths in attenuated total reflection (ATR). For NIR transmission in water solutions, the proportion between the $\approx$1440 nm and the $\approx$1930 nm water peaks may be used as an indicator of nonlinearity. In many cases the longer wavelengths including the 1900 nm peak have to be discarded to get a close to linear response.

For trace amounts of substances it is advantageous to employ fluorescence measurements to get a high sensitivity. Fluorescence data with both excitation and emission wavelengths collected is also a case where two-dimensional (2-D) MVC is working well.

With MVC it is also possible to use UV/VIS spectrometry in turbid solutions.[12] For NIR the turbidity is less of a problem due to the smaller spectral response for turbidity at longer wavelengths. If larger particles are present in the sample, the optics of the instrument has to have a wide enough beam to pass beside the particles. For optical fibers this means that when larger particles are present it is better with fiber bundles than single fibers. When the sample becomes too turbid it is advantageous to change from transmission to diffuse reflectance.

Solid samples also have polymorphic properties. These are destroyed if the sample is dissolved. Hence it is often of value to measure directly on solid samples. This can be done by IR, NIR or UV/VIS diffuse reflectance, NIR transmission, and Raman. X-ray diffraction is a possible reference method for MVC with polymorphic samples. For most complex solid samples measured with NIR, the MVC is the only way to get a calibration. NIR is working well with most organic solids but an exception is black plastics, which are better characterized with diffuse reflectance IR. Raman is also working for some inorganic solids.

Color measurement is a special form of UV/VIS diffuse reflectance with the colors as a standardized set of responses in a fixed instrument calibration.

Measurements in the gas phase are theoretically simpler, since there are less interactions between molecules than in the solid and liquid state. However, the equipment may become bulkier because of the longer cell paths needed to get sufficient sensitivity.

Measurements in plasma are done for example in inductively coupled plasma (ICP) and inductively coupled plasma/mass spectrometry (ICP/MS) where metal salts in solution are burnt in a plasma and the emission spectrum is measured, or the plasma is subsequently put into a mass spectrometer. The plasma is used to eliminate molecules and only keep the elemental responses from the excited atoms. Spectra are obtained in parallel with many elements. A potential for MVC exists with ICP and ICP/MS.

Spectroscopy is ideal for creating variables for MVC. One reason is the stability of the wavelength scale. MVC is dependent upon having the same information sorted into the same variables (wavelengths) over all measured samples. This generally holds when a single instrument is used for both calibration and prediction. For dispersive spectrometers such as UV/VIS, NIR, IR, and fluorescence and Raman, there are several levels of methods adapted for wavelength-scale verification and correction, all the way from using, for instance, a rare-earth oxide or polystyrene transmission or reflection standard manually during maintenance, to an automatic wavelength correction preceding each obtained spectrum. A frequent and automatic wavelength correction is preferable.

The photo diode array spectrometer (PDA) is a special case of the dispersive spectrometers where the grating is mounted in a fixed position. Due to the nonmoving parts, the PDA is suitable for fast data acquisition but is in general limited to narrower wavelength ranges and is less sensitive than the corresponding nonarray detector. For use with NIR the InGaAs array is presently emerging as a sensitive detector with high speed data acquisition in the 800–1700 nm range, and also in extended versions up to 2300 nm.

For nuclear magnetic resonance (NMR) it is possible to use compounds with known resonating frequencies as internal standards for the frequency scale. In mass

spectrometry (MS), mass standards also have to be employed in order to verify and adjust the mass scale.

The wavelength scale may also be controlled by fixed or variable filters. Arrays of fixed filters have been designed for specific measurements, such as protein in wheat by NIR diffuse reflectance. In general, fixed filter instruments are considered to be less repeatable between instruments and filters tend to age faster than gratings. Another line of filter instruments uses acousto-optical tuneable filters (AOTF). This type of instrument is suitable when fast spectral acquisition is needed in process applications.

Even if MVC can cope with mild instrumental non-linearities, it is preferable to obtain spectra of good quality from analytical instruments with good specifications. If a certain variation is not present in the spectra, it will not be caught by the MVC either. However, the presence of a variation need not be visible for the human eye in, e.g. a set of NIR spectra. MVC performs better than the human eye in catching small systematic variations that occur in combinations over a large number of wavelengths.

3.7 Spectral or Time Domain

The IR, NIR, NMR, and Raman spectrometers have developed to a state where spectra are often obtained as interferograms. Then Fourier transform (FT) is applied to translate the measured signals into spectra. For IR, NIR, and Raman the raw data for the wavelength or frequency scale is generated by a laser beam that passes through the same path as the analytical light in the spectrometer. By the use of interference fringes between the laser beams passing through the two different pathways in the spectrometer, it is possible to accurately define the relative position of the instrument mirrors and thus the frequency scale in the calculated spectrum. The precision of the wavelength scale from this type of instrument is generally high.

From an MVC point of view there may be alternative ways to treat the interferograms. They could be used directly as variables in an MVC or be preprocessed by, for instance, a wavelet transform.[13,14] The main reason for the use of FT methods is that the spectra should be convenient to inspect one by one for the human eye. To accomplish this, the interferogram is usually premultiplied by a filter that has a slight smoothing effect before the FT is applied. Since MVC works also for patterns that are less interpretable by the human eye, this smoothing may not be necessary and may even remove some of the important spectral information.

When using MVC based on FT generated spectra, it is important to keep track of the filter used before the FT calculation. This filter should be the same for all spectra in a calibration. With a less smoothing filter, wiggles may occur on the spectral baseline. These may not show up in the spectra but will still be visible in PLS loading vectors.

3.8 Chromatography

Chromatographic and electrophoretic methods also yield a vector (i.e. a chromatogram) for each sample with peaks representing different compounds separated in time. Such vectors are useful for characterizing biological samples and their relation to properties such as flavor, taste, and degree of disease.[15–17] Examples are gas chromatography (GC), liquid chromatography (LC) and thin-layer chromatography (TLC).

Just treating the chromatogram as a spectrum will not work well since the retention time axis of chromatograms varies with e.g. the column temperature, the mobile phase, the column age, and the column to column variation. Thus it is not straightforward to align chromatographic peaks. When there are few and known peaks, the peak areas or peak heights can be used provided that they are assigned correctly. Internal standards may be used as retention time reference points and also as a basis for the magnitude scale. When the whole chromatographic profile is important, such as for characterization of natural materials, e.g. oil, peat, and wine, it is more appropriate to use the entire chromatogram as a set of variables. Successful approaches for aligning the retention times have been reported.[18] Alternatively, Fourier or wavelet transformation of the chromatographic profile may give a representation less sensitive to shifts in the chromatogram.

When quantitative measurements are based on image scanned TLC plates or electrophoretic gels it is important to recognize the limited resolution in magnitude which may cause a nonlinear response.

Often a separation is optimized to give an even separation of the peaks. This is not the optimal way to run a separation method when molecular properties are wanted. Instead the conditions should be selected to show interactions of different kinds such as hydrogen bonding, $\eta-\eta$ interaction, acid–base properties, etc. The interactions of a compound with a set of mobile and stationary phases may be used in an MVC for molecular properties, such as lipophilicity, polarity and other related chemical properties for a class of chemical compounds.

3.9 Combination Methods

The combination methods often use one method to separate the sample into purer fractions, followed by a spectrometric method to give as much information as possible about the separated fractions. The fractions may not always be pure compounds, they can also be combinations of related compounds. GC may be

combined with MS or IR; LC may be combined with MS or UV; tandem MS may be used.

In these cases spectral data may be directly used in MVC. Automated multivariate classification is a good primary step in order to align retention times. This will now be easier since each chromatographic peak also has a spectral profile.

Hyphenated methods also lend themselves to the next level of MVC, working with variables as 2-D arrays instead of vectors. Special MVC methods have been developed for this situation.[19,20] However, when a chromatographic or electrophoretic separation method is involved, the time alignment problem in section 3.8.4 still persists.

3.10 Sensor Arrays (Electronic Noses, Crystal Sensors)

Sensor arrays are often constructed with silicon chip technology as resonators with a layer of substrate that is catching target molecules. The idea behind sensor arrays is that the individual sensor does not need to be totally selective, but through a combination of slightly different substrates, MVC is given the possibility to distinguish between different compounds. There are no alignment problems, but memory effects may be a problem, as the substrate may age and become chemically modified during harsh chemical conditions. When used for MVC these type of calibrations should have some means for shutting down individual sensors without too much degradation of the calibration performance.

3.11 Kinetic Curves

Kinetic measurements give a profile of variables that changes with time. Kinetic raw variables can be used in MVC as they are.

Another approach is to fit a physical kinetic model to the time traces and then use the coefficients of this model as variables in an MVC. In this case it is important that these coefficients really reflect a measured property and not just constitute an extrapolation. For example, an LD50 value will be an extrapolation in a toxicological study if the substance had such a low toxicity that the LD50 conditions were never approached in the study. In this case the LD10 or LD20 values are bracketed by measured values, and hence are more suitable for MVC.

3.12 Images and Other Measured Data

Two- or three-dimensional images from microscopes, digital cameras, tomography, etc. can all be transformed to data vectors and hence be used as data for MVC. Special preprocessing may be needed to make the data comparable between samples, such as 2-D FT or 2-D wavelet transforms. Actually, ordinary spectra UV, IR, etc.) can be considered as one-dimensional images, and quantitative image analysis can be considered as just a special case of MVC.[21]

Any quantitative multivariate data measured on analytical samples can be used as the basis for MVC provided that they contain relevant information. Recently, profiles of sound and other vibrations have been shown to contain interesting information about concentrations, the degree of mixing and other properties of complex material flows.[22]

4 SELECTION OF THE CALIBRATION SET

The calibration results in a model that relates the signal amplitudes to concentration, the so-called standard curve. All models have a domain of validity which depends on the mathematical stability of the model and the range of the data on which the model is based. The uncertainty of the predicted concentrations for a given signal amplitude increase rapidly outside the concentration range defined by the calibration samples (see Figure 1). Hence the calibration model (the standard curve) should not be applied far outside this range. Consequently, if one knows the future operating range of the analytical method, the calibration samples must span this range. The calibration sample concentrations should also be distributed fairly uniformly over the range of interest.

The precision of the standard curve improves with an increasing number of calibration samples, n.[2–4] With large n, the uncertainty of the predicted concentrations is dominated by the uncertainty of the signal amplitude of the new samples. In the situation where reference samples are prepared from pure substances, there is therefore little gain in increasing n above around $M \times 10$ (M is the number of analytes), except for validation of linearity and for the estimation of the precision of measurement in different parts of the concentration interval.

In the situation where the concentration levels in the reference samples cannot be directly controlled, however, one needs a fairly large number of such samples for the calibration of the model to ensure representativity, typically thirty to several hundreds.[1,5] In the subsequent data analysis one must also confirm that indeed the concentration space is appropriately represented.

The reference samples provide a TS of data where both the ''signal matrix'' $\mathbf{X}$ – usually consisting of a set of digitized spectra from NMR, NIR, FTIR, ... – and the property matrix $\mathbf{Y}$ for the same samples (the

concentrations or other property values) are defined. These data provide the basic structure of the model, i.e. how the signals (digitized spectra) co-vary with the concentrations (or other properties) in the **Y** matrix.

4.1 Validation Set (Test Set)

To ensure that the calibration model works well in the desired concentration range(s), it is essential to validate the model by using it on a representative set of **new** samples – the validation set, often also called the test set. This is further discussed in section 9.

4.2 Prediction Set

The samples on which the developed model finally is used are often collectively called the prediction set. This set is often open ended and in principle infinite, and is usually not known or available at the time of the model development, the calibration.

5 MAKING THE MEASUREMENTS

This part is of course a most important part of the calibration. MVC is by no way a substitute for sloppy measurements. On the contrary, MVC shows the weakness of the data, and is negatively affected by imprecision, bad sampling, etc. just as UVC. Hence, the multivariate measurements must be made in a reproducible and precise way. This means, in turn, that the measuring conditions should be controlled to be as constant as possible, or at least carefully recorded, solvents should be checked, etc. i.e. normal analytical precautions should be taken also with MVC. See section 5.1.

5.1 Uncontrolled Variables

Uncontrolled variables such as temperature, humidity, time order, powder packing density, cell path length, etc. may affect the spectral response and in some cases the chromatographic retention. MVC can often compensate for variations in uncontrolled variables. This is achieved if the uncontrolled variable is allowed to vary in the calibration TS. The best form of such variation is according to a statistical experimental design. The variations could also be introduced as random variations. It is bad practice to try to keep all influencing variables constant in the calibration training and test sets, and thereafter relax the conditions in routine use. On the contrary, the calibration should contain slightly larger variations than expected in routine use in order to enter compensations for the interfering variations in the calibration model. If, for example, means to control the climate and precise powder packing devices are available during calibration time for a diffuse reflectance NIR method for powders, these should be used to induce known designed or random variations rather than keeping the conditions unrealistically constant.

5.2 Ambiguous Spectral Regions (Variables)

Together with high-performing variables, in general analytical instruments produce variables with bad performance in spectral regions that are technically available but of little analytical value. One example is when optical fibers are used together with a spectrometer. The original design of the spectrometer usually allows for measurements at wavelengths that are not coming through the optical fiber. Rather than resorting to automatic variable selection schemes, the analyst should take decisions in these cases and avoid using variables that are out of the analytical range.

5.3 Randomization of Measuring Order

It is good practice to make the measurements in random order when the training and test sets are made. It is not advisable to order the samples in the same or reversed order as the magnitude of the *y*-variable to be calibrated for. If the samples are ordered, changes that occur with time, such as instrument warming, operator learning, etc. will introduce a bias in the calibration.

When trace amounts of compounds are to be measured, this randomized measuring order is not always practical. In these cases the samples may have to be grouped to reduce carry over. However, extra diagnostics must be applied to verify that the measurements are not correlated with time. This can be done with an explorative MVC with the actual time or the run order for the measurement as the response to the measured variables.

5.4 Data Collection and Administration

Data suitable for import into for MVC software are usually organized as tables or matrices where each row holds the measured values for one sample. These row vectors are also commonly called objects or observations. Each column in the table holds the measurements of a specific variable. Each object and variable should be identified by a name (label) in ASCII text format. When the objects are spectra, it is useful to attach the spectral scale to the data table. This may be done in the name fields of the variables or as a separate object.

6 INITIAL QUALITY CONTROL OF THE DATA

In any data set there is a risk of bad measurements, fouled samples, etc. giving outlying observations. Unless these problems are identified and properly dealt with, they will badly affect the calibration model.

6.1 Deviations of Individual Spectra from the Average

During work with small changes in spectra, it is beneficial to inspect the spectral shapes after subtraction of the mean spectrum. An example is shown in Figures (3a–c) where the diffuse reflectance NIR spectra from three homogeneous metoprolol tablets are shown, first (a) as original spectra, then (b) with the mean from the three tablets subtracted, and finally (c) a spectrum from the pure metoprolol tartrate. From Figure 3(a) it is not possible to see the metoprolol variation. However, in Figure 3(b) the same main patterns are visible as in the pure substance spectrum in Figure 3(c). Also note that the scale in Figure 3(b) covers 0.05 $\log(1/R)$ units while the scale in Figure 3(a) covers 0.6 units. This inspection serves well to give a first impression of the main spectral variation. The spectral interpretation aids in the confidence of the method. When the TS is generated according to a design there may be more than one distinct spectral variation. In that case this inspection may be complemented by inspection of the principal components analysis (PCA) spectral loadings as described below. This type of inspection is a good alternative to the use of derivatives as inspection tool. The original shape of the spectral variation is preserved and no information is removed.

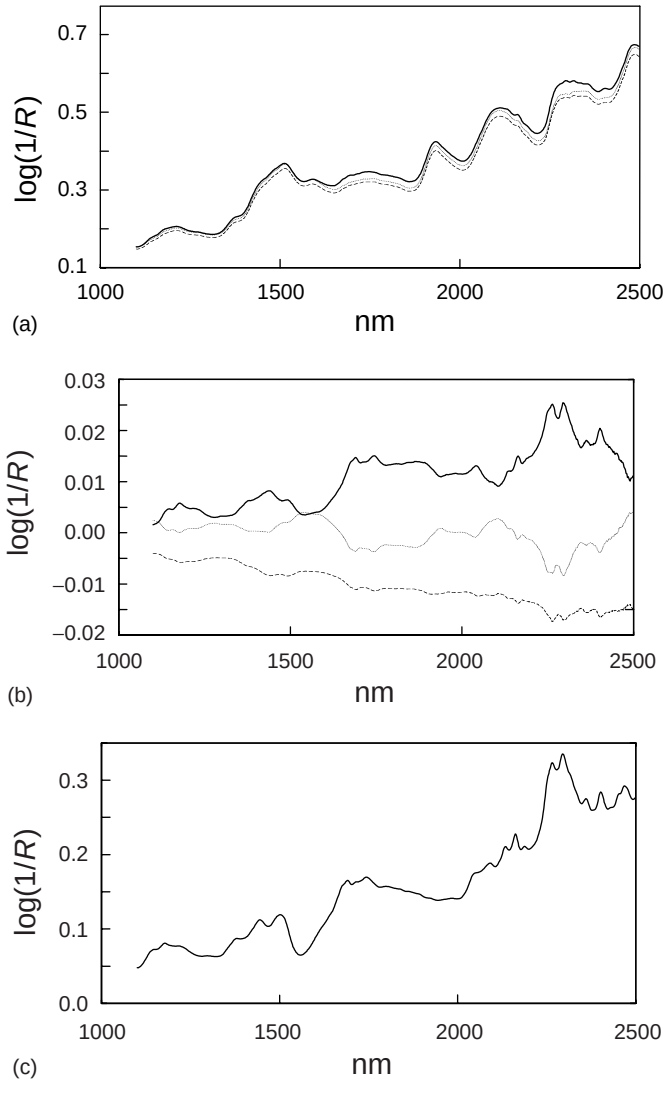

Figure 3 NIR spectra from three homogeneous metoprolol tablets: (a) original spectra, (b) with the average spectrum subtracted, and (c) a spectrum from the pure metoprolol tartrate.

6.2 Multivariate Quality Control of the Data by Principal Components Analysis

The correlation structure in multivariate data (here digitized spectra) allows the straightforward identification of outliers by means of PCA. Also "univariate outliers", deviating strongly in only a single variable, are seen in the PCA, particularly in the RSDs, s_i. The same analysis provides indications also about the balance and representativity of the signal matrix, **X**, as well as about the rank of **X**, i.e. the inherent dimensionality (complexity) of **X**. The latter is closely related to the number of detectable constituents in the investigated set of samples.

In the example, a PCA of the training data (centered but unscaled, see section 7.3.4) gives three significant components indicating the presence of three constituents. However, only two components are expected with two analytes. The third, very small component ($R^2 = 1\%$) indicates some kind of deviation from Lambert–Beer behavior. The loading vector of the third component (p_3) looks similar to a linear combination of the first two loadings (Figure 4), but the third score vector (t3) does not have any clear relation neither to felodipine, nor to metoprolol (Figure 5b).

6.3 Outliers in X and Unbalanced X-data

The t1, t2 score plot of the TS (Figure 5a) does not show any strongly deviating observations, but displays a pattern corresponding to the concentration design matrix. This indicates that the dominating systematic part of the spectroscopic data is related to the concentrations, **Y**. The RSDs, s_i (section 8.4.2) also are within the normal limits.

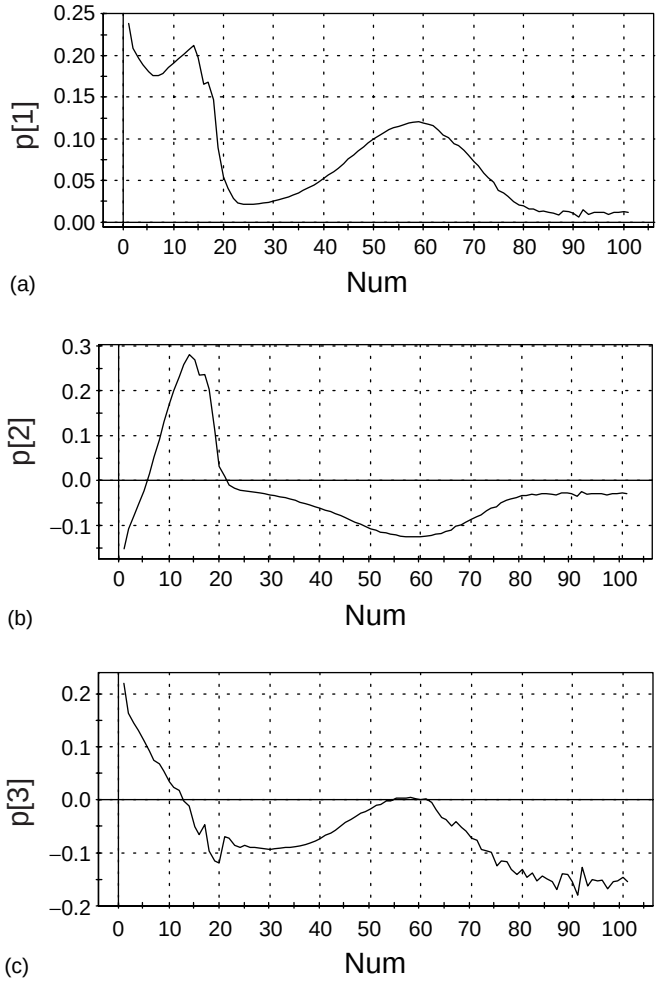

Figure 4 The three PC loading vectors (p1 to p3) of ex.1 plotted vs variable number.

The t1, t2 score plot also shows a well balanced TS without any empty regions or strong clusters. This further corroborates the good distribution of the data, and hence these training data sets have passed the first ''quality control''.

6.4 Outliers in Y

With many **Y**-variables, a PCA of **Y** shows that, analogously to above, the balance of the **Y**-data, the presence/absence of outliers and clusters, etc. This analysis is relevant when the data has more than four or five *y*'s. For a partial least-squares multivariate calibration (PLS/MVC) the **Y**-outliers will also be visible in *t* vs. *u* plots for the significant components. If several outliers are present the most deviating will show up first. When these are removed additional milder outliers may show up. A common cause for **Y**-outliers is that the value from the reference method is wrong.

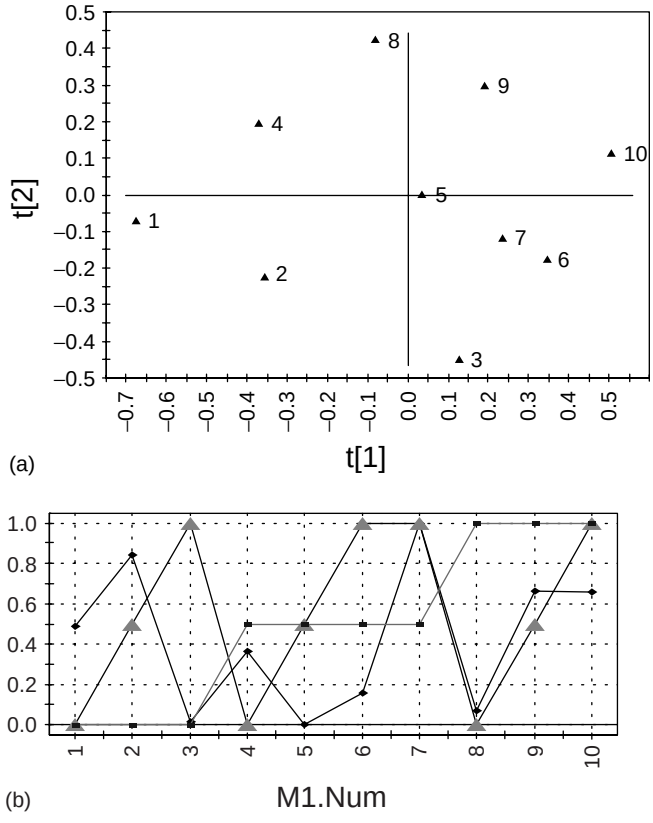

Figure 5 (a) The t1, t2 score plot of a PCA of the TS of ex.1. (b) The third scores (t3) and Y1 and Y2 vs obs number M1.t(3); ▲ DS1. Felodipi (DS); ■ DS1. Metoprol (DS).

7 DATA PREPROCESSING

The raw data coming out of the measuring instrument (typically the spectrometer) are not necessarily expressed in the form that is most suitable for the mathematical calibration. Hence there is often a motivation for modifying the raw signals, both by simple transformations such as logarithmic, or by complex modifications such as Fourier transformation, filtering, variable selection, baseline correction, and differentiation.

The preprocessing of the TS and of new samples (the prediction set) should be the same. When the preprocessing contains adjustable parameters, e.g. averages, SDs, orthogonal signal correction (OSC) components, etc. these parameters are based on the TS, and the same parameter values are then used for the new samples.

7.1 Choice of Representation (Time or Frequency Domain)

As discussed in section 3.7, multivariate signal profiles come out of the instruments usually either in the form of

a spectrum (light absorption at different wavelengths or frequencies) or, with a FT spectrometer, as an interferogram (spectral amplitude as a function of mirror distance or time). As humans we usually prefer the first representation, but nothing indicates that either is more useful for MVC. Hence, with FT spectroscopy it may sometimes be preferable to use the digitized interferograms directly as objects, and then back-transform the resulting model coefficient profiles to a spectral representation for the interpretation of the results. Our experience is still not sufficient to provide any guidelines.

7.2 Logarithms and Other Stabilizing Transformations

If the signals are exponentially distributed, taking their logarithms is recommended before the data analysis. This may also stabilize the influence of the noise on the model if the size of the noise is approximately proportional to the concentration. This is common when the latter varies over several magnitude of ten. If the value zero occurs frequently in the signal matrix, the fourth root is an alternative transformation which has a similar ''compressing'' effect as the logarithm on large values, but it does not inflate very small values, and the fourth root of zero remains zero.

If a linear relation is expected between the original signals ($\mathbf{Z}$) and the concentrations ($\mathbf{Y}$), then both $\mathbf{Z}$ and $\mathbf{Y}$ should be transformed in the same way. Hence, if $\log(\mathbf{Z})$ is used, this implies that also $\log(\mathbf{Y})$ should be used in the calibration.

7.3 Spectral Correction Methods

Particularly with diffuse reflectance spectroscopy, there is often a need of a spectral preprocessing to remove baseline variation and other systematic variations in the data that are not related to the analyte concentrations or other properties of interest. Then so-called signal correction methods are applied, or, alternatively, first or second derivatives of the spectra along the wavelength axis are calculated (spectral differentiation). We emphasize that there is presently not sufficient experience with data preprocessing and MVC to select the appropriate approach a priori, and a comparison between the results of different approaches including the use of unmodified data should always be made.

7.3.1 Additive and Multiplicative Signal Corrections

Large interfering additive and multiplicative variations are often present in spectra along with the desired information. The additive effects are visible in spectra as offsets. The multiplicative effects are visible as a scaling factor where proportions between spectral peaks are preserved while the spectral magnitude varies.

It is possible to include these variations in the MVC, but at the cost of an increased complexity of the calibration model. With principal components regression (PCR) and partial least-squares regression (PLSR) several extra components may be needed. This often affects the predictive precision negatively, and also makes the model interpretation more difficult.

This type of variation is common with diffuse reflectance methods, since the effective cell path is not exactly defined, as in transmission spectrometry. In NIR diffuse reflectance the cell path variations will cause multiplicative effects and e.g. a smaller particle size will give a lower baseline of a $\log(1/R)$ spectrum, i.e. the spectrum will appear ''lighter'' for the spectrophotometer.

In diffuse reflectance NIR spectrometry these effects are often the most dominating, which means that they are modeled in the first component if PCR or PLSR is used. When the objective is to model chemical properties with a simple MVC model, the corrections can be used. As a consequence of the simple model, the scores and loadings will in many cases be easier to interpret. The correction will, however, not cure bad data.

In cases when the MVC is aimed at physical properties such as particle size, no correction should usually be made, however. This because these physical properties may be reflected in light scatter, etc. i.e. components little related to concentration. As an alternative, the sizes of the additive and multiplicative corrections may be added as two appropriately scaled extra variables in the calibration.[1]

Multiplicative signal correction (MSC), or mean centering of the objects followed by a scaling of each row to unit SD, are two variants of the same correction except for a scaling factor.[23,24] The object centering and auto-scaling has also been named the standard normal variate (SNV) transform.[25] Both transforms remove the additive and multiplicative effects. Spectra may also be detrended by fitting and subtraction of a low order polynomial. Dhanoa et al. recommends that this should be done before the SNV correction.[24]

The MSC algorithm uses univariate linear regression with a plot of the TS mean spectrum against each individual spectrum. The resulting intercept is used for the additive correction and the slope is used as the multiplicative correction of each spectrum in the TS. Further measured samples, in the test and prediction sets, have to be corrected with the saved TS mean spectrum.

Denoting the row vector of the average spectrum of the TS by $\mathbf{m}'$, each training and prediction set observation row vector, $\mathbf{z}_\mathbf{i}'$ is corrected as follows. First the slope, a_i, and the intercept b_i, are calculated by least squares ($\mathbf{e}$ is a residual vector), Equation (7):

$$\mathbf{z_i} = a_\mathrm{i} + b_\mathrm{i}\mathbf{m} + \mathbf{e} \tag{7}$$

Then, the ''corrected'' vector, $\mathbf{x_i}$, is calculated as Equation (8):

$$\mathbf{x_i} = \frac{\mathbf{z_i} - a_i}{b_i} \tag{8}$$

The relation to a mean spectrum is a disadvantage with MSC since the mean spectrum has to be recalculated for each selected TS. That is avoided with SNV.

The MSC has been applied both to the entire spectral range present in the MVC and to parts thereof. One way to correct measured variables that contains an internal standard, would be to correct the whole spectrum with an MSC that get its slope and intercept only at the variable interval where an internal standard is present.[1]

7.3.2 Optimal Scaling

Optimal scaling is preprocessing that corrects for the nonlinearity created when spectra are normalized before the calibration.[26] Optimal scaling can also correct for multiplicative effects. This is done by the addition of a multiplier m_i for each sample in $\mathbf{Y}$. Then Equation (9) will be complemented by a diagonal multiplier matrix $\mathbf{M}$:

$$\mathbf{MY} = \mathbf{XB} + \mathbf{F} \tag{9}$$

The MVC is then constituted as a solution to Equation (9) including the optimized $m_i : s$. The regression coefficients B may be solved as

$$\mathbf{B} = \mathbf{X}^{+}\mathbf{MY} \tag{10}$$

where $\mathbf{X}^{+}$ is the generalized inverse of $\mathbf{X}^{T}\mathbf{X}$ for the cases when $\mathbf{X}$ is singular.

Optimal scaling improves PLS calibration on normalized X-ray diffraction spectra.[26] This scaling is, however, not appropriate for the correction of additive effects. Raw data with additive anomalies should be corrected with another method before the application of optimal scaling.

7.3.3 Orthogonal Signal Correction

OSC[27,28] is a recently published correction method. The main idea is to remove systematic variation in the $\mathbf{X}$-variables that is not correlated to the variation in the y-variables. This is another way to remove variations introduced by variations in the apparent cell pathlength. In contrast to the MSC and SNV, OSC needs the y-variable to guide the correction. It is also possible to use OSC to reduce the size of interferences such as the switch of instruments.[28,29]

The OSC operates as a reversed PLS calibration. A multivariate model of the TS is built, but the objective of the model is to describe variations that are orthogonal to the Y-variable(s). The model is composed of OSC-components, just as a PLS-model is composed of PLS-components, (Equation 11):

$$\mathbf{X} = \mathbf{T_{OSC}}\mathbf{P'_{OSC}} + \mathbf{E} \tag{11}$$

The residual matrix $\mathbf{E}$ contains the OSC-filtered data, which are thereafter used as the TS signal matrix in the MVC. Just like ordinary PLS, OSC also gives a weight matrix, $\mathbf{W^{*}_{OSC}}$ such that (Equation 12):

$$\mathbf{T_{OSC}} = \mathbf{XW^{*}_{OSC}} \tag{12}$$

Spectra of new samples (centered and scaled) are then filtered using the same OSC model before they are entered the calibration model to predict the values of $\mathbf{Y}$ (the concentrations). First their ''OSC-score'' values are calculated, Equation (13):

$$\mathbf{t'_{OSC}} = \mathbf{x'_{new}}\mathbf{W^{*}_{OSC}} \tag{13}$$

and thereafter the ''filtered'' values are calculated as Equation (14):

$$\mathbf{x'_{new,\ filtered}} = \mathbf{x'_{new}} - \mathbf{t'_{OSC}}\mathbf{P_{OSC}} \tag{14}$$

The limited experience is that one or two OSC components usually are adequate. The score plots of both the OSC model and of the final PLS/MVC should be used as diagnostic tools. When the score-pattern that is being corrected for disappears from the final MVC, no further OSC components should be used. As always with a multivariate modelling, the OSC may give overfitting if too many components are used.

7.3.4 Centering and Scaling

After transformations (log, etc.), the variables are usually centered by subtracting the average values (a_k, see below) from each column in $\mathbf{X}$ and $\mathbf{Y}$. Thereafter, one can optionally scale the variables by multiplying each by a scaling weight, v_k or v_m. The most common is to scale both the $\mathbf{X}$ and $\mathbf{Y}$ variables to unit variance. This is done by calculating the column averages (a_k) and column SDs (s_k), and then make the linear transformation, Equation (15):

$$x_{ik} = \frac{z_{ik} - a_k}{s_k} \tag{15}$$

This corresponds to using the inverse SDs as scaling weights (i.e. $v_k = 1/s_k$). The result is that each variable gets the same initial importance in the calculation of the calibration model.

When the variables comprise digitized spectra and all the variables are in the same unit, another common ''scaling'' is to keep the spectrum as it is, except for centering. With a projection method such as PCR and

PLSR this means that the variables get the importance for the calibration proportional to their variation. Variables should still be centered, and it is the variation after centering that gives the weight (see Figure 3b). This is sometimes a good approach for spectra if large sections of spectra are only reflecting baseline noise. If scaling to unit variance is applied to such data, it will magnify the amount of noise that is entered into the calibration model. However, sometimes the variations in the baseline may be significant for the analysis. This may occur with e.g. fluorescence in Raman spectrometry and in gradient LC. A compromise is to weight the variables proportionally to the inverse of the square root of their SD, often called Pareto scaling. Thus the large variations are still dominating, but the small variations are not scaled down as much.

When blocks of variables give less than optimal response for the instrument type, but these blocks still contain information that is important for the calibration, an option is to down-weight these blocks. This will enhance the detection of aberrant measurements and the calibration performance may increase.

When the **X**-variables comprise combinations of spectra from several analytical instruments, the blocks of variables from the different instruments may be combined with weights that let the instruments have ''equal votes'' in the calibration (block scaling).

It may be also possible to automatically optimize the individual variable weights in a calibration. This is in general not a good practice, since the risk of over-fitting will increase. This method is related to variable selection methods where a large set of selections are tried and the ''best'' is chosen. This greatly increases the risk of producing spurious correlations.

7.3.5 First and Second Derivatives

When the *x*-variables for an MVC have the same unit and are arranged as a continuous function (e.g. in spectra), the first or second derivatives may replace the original variables as **X**-variables for the MVC. Using first derivatives removes the linear slope of the baseline and changes all peak maxima to zero-crossings. Using second derivatives reduces the background further and leaves narrower negative peaks at the original peak positions. However, the differentiation also removes all variation in the variables that is not carried by differences between adjacent variables. In the rare case where the majority of the variation in the **X**-variables remains after the differentiation, this may be a good pretreatment.

Differentiation removes a sizable amount of variation that may be important for the calibration. Noise will then be a larger proportion of the signal, and will be more visible in spectra. This has made it possible for most derivatives in instrument software to be calculated with smoothing, for instance using Savitsky–Golay approach,[30] or by subsequent differences between moving averages.[31]

Care should be taken not to smooth away information. When digital spectra are collected, there is often a digital over-sampling in relation to the bandwidth of the spectrometer. A good rule of thumb is to keep the smoothing parameters in a range that is not degrading the efficient bandwidth of the instrument significantly.

A common misuse of derivatives to use it as a replacement for mean-centering of the variables. MVC models with and without differentiated **X**-variables should always be calculated with mean centered **X**-variables before comparison.

8 THE ANALYSIS OF THE TRAINING SET DATA; THE CALIBRATION MODEL DEVELOPMENT

With many spectral variables (more than one), a computerized analysis is necessary to use the data efficiently. The basis for this analysis is to develop a model, a mathematical description, of the relationship between the raw or preprocessed spectral variables (**Z** or **X**) and the concentrations (**Y**). There are two basic approaches to this data analysis, multiple linear regression (MLR), or a multivariate projection. The most commonly used projection methods are PCR and PLSR.

8.1 The Mathematical Formulation of Multivariate Calibration

The general form of the calibration model can be derived from Lambert–Beer's law, which states that at any wavelength the light absorption is linearly related to the concentrations of constituents that absorb at that wavelength.

Hence, according to Lambert–Beer's law, spectral data z_{ik} – the spectral absorbance of sample i at the kth wavelength – is a sum of contributions from chemical constituents with the concentrations c_{im} and the spectral unit absorbance d_{km}, plus noise, e_{ik} (Equation 16):

$$z_{ik} = \sum_{m} c_{im} d_{km} + e_{ik} \tag{16}$$

In matrix form, this is Equation (17):

$$\mathbf{Z} = \mathbf{C}\mathbf{D}' + \mathbf{E} \tag{17}$$

After monotonic transformations, centering, scaling, and preprocessing (section 7), giving the signal data matrix **X**, this matrix is still decomposable in the same

way (the resulting residual matrix, **E**, is not the same as above, however), Equation (18):

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \tag{18}$$

An important conclusion is that the signal matrix **X** can be modeled by this PC-like model, also called a bilinear model.

The data analytical problem is now basically to find the parameter matrices **T** and **P′** which are related to the concentrations of the analytes of interest. There are several complications, however. First, the number of constituents in a complicated sample is often large, meaning that the rank of **T** and **P** is potentially large. Second, there is a rotation problem, which makes **T** and **P** nonunique. Inserting any nonsingular matrix **R** and its inverse $\mathbf{R}^{-1}$ between **T** and **P′** in Equation (18) gives Equation (19) (remembering that $\mathbf{R}^{-1}\mathbf{R} = \mathbf{1}$):

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} = \mathbf{TR}^{-1}\mathbf{RP}' + \mathbf{E} = \mathbf{VQ}' + \mathbf{E} \tag{19}$$

Finally, the presence of the noise, **E**, presents further problems. The rank of **X** cannot be determined with full certainty, and any parameter estimation based on **X** will also result in noise in the parameters.

Hence it is not sufficient to just decompose **X** by PCA because the resulting score matrix **T** cannot be identified with the desired analyte concentrations **Y**. Instead one has to find a way to find the loadings **P** so that at least the resulting scores (**T** = **XP**) are linearly related to the concentrations **Y** (PCR and PLSR, below) by coefficients **Q**, Equation (20):

$$\mathbf{Y} = \mathbf{TQ}' + \mathbf{F} \tag{20}$$

Equation (17) can be multiplied from the right with $\mathbf{D}'^{-1}$, giving Equation (21), after rearrangement:

$$\mathbf{C}(\text{or } \mathbf{Y}) = \mathbf{ZD}'^{-1} + \mathbf{F} = \mathbf{ZB} + \mathbf{F} \tag{21}$$

Hence, we can, in principle, find coefficients **B** that directly relate the signal matrix to the concentrations by a MLR model (Equation 5). This might seem more attractive than going via the bilinear decomposition of **X**, but actually leads to a loss of information since no modelling is made of the signal matrix **X** and its correlation structure. This information loss makes both the interpretation of the model and the recognition of anomalous samples model difficult. Also, usually the variables in **X** are strongly correlated, which makes the MLR solution impossible to compute without an extensive variable reduction. This can be made in simple cases, but is difficult with complicated samples without distinct spectral regions. The different approaches – PCR, PLSR, and MLR – are discussed below (section 8.4).

8.2 Linear or Nonlinear Calibration Model

Relationships between measured data are, in principle, nonlinear. However, over limited intervals these relationships are usually very close to linear, and can hence be adequately approximated by a linear model. However, if the relationship appears to be curved, the linear models can easily be extended to quadratic or cubic models, handling polynomial (mild) nonlinearities (section 8.4). These nonlinear models work as long as there is a monotonic relationship between signals (**X**) and concentrations (**Y**), i.e. the spectral variables always increase (or decrease) when the concentrations or other calibration properties increase in value. Stronger nonlinearities are impossible to deal with even in principle, since then the calibration model becomes noninvertable and a given signal profile will correspond to several different concentrations.

8.3 One or Several Analytes

With several analytes and a multivariate **Y**-matrix of concentrations, there are two principally different ways to perform the MVC. Either one can develop a separate model for each analyte, or one can develop a single model for all analytes. Each approach has its merits. The separate models are often easier to understand and use, and often give slightly better predictions, especially if the *y*-variables are close to orthogonal. This situation typically arises when calibration samples are prepared from pure substances and their mixtures varied according to a statistical design as discussed in section 3.2 above.

The multi-analyte model is preferable when the number of analytes is large and the concentrations (or other properties used as **Y**) are correlated. This is typical for noncontrolled samples of natural origin, i.e. the second case discussed in section 3.2. Then the simplification of a single model is strongly desirable, and possible small gains in predictive precision in one or a few of the analytes by using separate calibration models are offset by the ensuing increase in complexity of developing and maintaining a large number of different calibration models.

With noncontrolled samples, it often happens that the analytes are correlated groupwise. This grouping is usually seen in the loading plots of a PCA of the concentration matrix, **Y**. Making a separate model for each group of correlated analytes is then a natural approach.

8.4 The Calibration Model

The choice of the calibration model depends mainly on the complexity of the calibration problem and the corresponding data structure. One of the great advantages

with projection methods is that they can handle data sets with very many and correlated **X**-variables, even many more than the number of calibration samples, *N*. For such data, the ordinary MLR cannot even be computed. With many **X**-variables, K, and most of them containing information about **Y**, projection methods are hence the choice, while for very simple problems where a few wavelengths can be selected to give an adequate calibration model, MLR is often used. As indicated in the end of section 8.1, however, projection methods may still be preferable also in simpler problems, because they give additional information beyond the results of MLR.

8.4.1 Projection Methods

As discussed in section 8.1, we expect that the signal matrix **X** can be decomposed as a product of a matrix **T** (the score matrix) and a matrix $\mathbf{P}'$ (the loading matrix), Equation (18). Moreover, the scores (the columns of **T**) can be used as predictor variables of the concentrations, **Y** (Equation 20). This is the basis for the two most commonly used MVC models, namely PCR[32] and PLSR.[33,34]

The former PCR is based on first making a PCA of the signal matrix **X** according to Equation (18), and then developing a second model of the relation between the score matrix (**T**) and the concentration matrix **Y** (Equation 20). Each PC consists of one score vector ($\mathbf{t_a}$) and one loading vector ($\mathbf{p_a}$), and the component index (a) runs between one and A (the number of components of the model). The components, i.e. the columns of **T** and **P**, are mutually orthogonal, and the latter (the **P** columns) are normalized to length one. Moreover, the score vectors ($\mathbf{t_a}$, i.e. the columns of **T**) are sorted in order of the importance for **X**, i.e. $\mathbf{t_1 p_1'}$ describes more of **X** (its sum of squares (SS)) than does $\mathbf{t_2 p_2'}$, etc. The PC loadings ($\mathbf{p_a}$) and score vectors ($\mathbf{t_a}$) are eigenvectors to the variance covariance matrix ($\mathbf{X'X}$) and the association matrix ($\mathbf{XX'}$), respectively.

Due to the orthonormality of the PCA loading matrix **P**, and the orthogonality between $\mathbf{P}'$ and the residual matrix **E**, Equation (18) multiplied from the right by $\mathbf{P}'$ gives Equation (22):

$$\mathbf{T} = \mathbf{XP}' \tag{22}$$

For an additional sample (in the prediction set) with the preprocessed data vector $\mathbf{x'_{new}}$, the predicted PC score values are calculated from this relation, which when inserted in the calibration model (Equation 20) gives the predicted values of **Y**, Equations (23) and (24):

$$\mathbf{t_{new}} = \mathbf{x'_{new}P} \tag{23}$$

$$\mathbf{y'_{new}} = \mathbf{x'_{new}PQ'} \tag{24}$$

Standard errors ($\mathrm{serr_y}$) and confidence intervals (approximately 2 or 3 times the standard errors, according to the t-distribution) can be calculated in many ways for the predicted *y*-values, the simplest being based on the regression formalism and the fairly accurate assumption of a fixed (exact) X-score matrix, **T**. The square of the standard error (error variance) including the variability of the new measure *y*-values, and assuming a variance of the measurement errors of σ^2, is Equation (25):[35]

$$\mathrm{serr_y} = \left(1 + \frac{1}{\mathrm{N}} + \mathbf{x'_{new}}(\mathbf{T'T})^{-1}\mathbf{x_{new}} \quad \sigma^2 \right. \tag{25}$$

PCR presents some complications, however. In particular in complicated samples, there is often a strong systematic variation in **X** that is not related to **Y**. Also, especially in reflectance spectroscopy, there is often a baseline variation and a multiplicative effect due to light scatter and varying light path length. This makes the first few PCs little related to **Y**, with the ensuing loss of information. This may be remedied by appropriate data preprocessing (section 7). However, the fact that the components of **X** are derived without using the information in **Y**, makes PCR often need many components to capture the parts of **X** that are relevant to **Y**, with a resulting loss of predictive precision.

PLS projection to latent structure PLS[33–36] is a method where the projection of **X** is made to enhance the information content about **Y** in the score matrix, **T**. This decreases the risk for losing information about **Y** in the projection of **X**, and with complicated samples PLSR usually needs fewer components than PCR to achieve the same explanation of **Y**. With few calibration samples and many **X**-variables, this is often an advantage, and PLS usually gives a smaller prediction error than PCR. In many applications, however, where the information in **X** about **Y** sits in the first few components like in the present example, the differences are miniscule and chemically and statistically insignificant.

The PLS model is developed with two objectives, namely to model **X**, and to predict (model) **Y** from **X**. The former is accomplished by finding new ''variables'', the scores ($\mathbf{t_a}$), that well summarize **X**, and the latter objective is achieved by focusing the projection on the parts of **X** that are related to **Y**.

The PLSR model has identical form as the PCR model (Equations 18 to 20), but the actual values of the parameters (**T** and **P**) are different, as well as the dimensionality (A). The computation of the PLS parameters is made in such a way that the PLS score matrix, **T**, is an optimal compromise between modelling **X** and predicting **Y**. This compromise is sometimes called the ''H-principle'' which points to the analogy with Heisenberg's principle of uncertainty.[36] This is achieved by having the first PLS score vector ($\mathbf{t_1}$) be the largest

eigenvector of the extended association matrix, $\mathbf{XX'YY'}$ in contrast to the first PC score vector which is the largest eigenvector to just $\mathbf{XX'}$. The subsequent PLS score vectors ($\mathbf{t_a}$) are eigenvectors of $\mathbf{E_{a-1}E_{a-1}YY'}$, where $\mathbf{E_{a-1}}$ is the matrix of X-residuals ($\mathbf{X - T_{a-1}P'_{a-1}}$) after a − 1 components. The PLS model of $\mathbf{Y}$ is formulated as a linear model between the $\mathbf{X}$-scores and the $\mathbf{Y}$-variables, with a coefficient matrix $\mathbf{C}$, Equation (26):

$$\mathbf{Y = TC' + F = XW^*C' + F = XB + F} \tag{26}$$

In PLS there is a second matrix of ''loading vectors'', denoted by $\mathbf{W}$, with one column per PLS component. A variant of $\mathbf{W}$ denoted by $\mathbf{W^*}$ contains the weights that directly combine the $\mathbf{X}$-variables to form the scores, $\mathbf{t_a}$. Thus, analogously to Equation (22) we obtain Equations (27) and (28):

$$\mathbf{t_a = Xw_a^*} \tag{27}$$

$$\mathbf{T = XW^*} \tag{28}$$

This is used for calculating the predicted values of $\mathbf{Y}$ for new samples with X-data $\mathbf{x'_{new}}$ according to Equation (29):

$$\mathbf{y'_{new} = x'_{new}W^*C'} \tag{29}$$

The standard errors and confidence intervals for these predicted values are calculated in the same way as for PCR, i.e. by Equation (25).

8.4.2 Diagnostics Common to Both Regression Methods

The diagnostics and the interpretation of PLSR is the same as for PCR, as well as the way to calculate predicted y-values (e.g. concentrations) for new samples. The only difference is that in PCR, the X-scores for new samples are calculated by means of the loadings, $\mathbf{P}$, while in PLS, these scores are calculated by means of the weight matrix $\mathbf{W^*}$ (Equations 22 and 28 respectively). We note that the X-residuals are calculated in the same way for both PCR and PLS (from Equation 18) by Equation (30):

$$\mathbf{E = X - TP'} \tag{30}$$

The X-residuals, $\mathbf{E}$, express the deviations between the data ($\mathbf{X}$) and the model ($\mathbf{TP'}$). For new observations (samples), the X-residuals ($\mathbf{e'_{new}}$) are calculated in the same way as the preprocessed sample x-vector minus the calculated scores times the loading matrix of the model, Equation (31):

$$\mathbf{e'_{new} = x'_{new} - t_{new}P'} \tag{31}$$

A great advantage of projection methods is that data of new samples in the prediction set can be ''quality controlled'' by inspection of their RSD values (s_{new}) (Equation 32). The RSD of observation i shows how far this observation is from the model, however dissimilar it is to the model.

$$s_{new} = \frac{\mathbf{e'_{new}}}{K - A} \tag{32}$$

If the value of this RSD for a new sample is larger than a certain limit (d below), the sample is an outlier not being similar to the TS. This, in turn, indicates that the predictions of the concentrations ($\mathbf{y}$) are unreliable for the sample. The limit (d) for the RSD is calculated from the RSD of the whole training matrix (s_0) and a critical value of the F-distribution (Equations 33 and 34).

$$s_0 = \text{sqrt} \sum_i \sum_k \frac{e_{ik}^2}{(N - A - 1)(K - A)} \tag{33}$$

$$d = s_0 \times \text{sqrt}(F_{crit}) \tag{34}$$

8.4.3 The Number of Model Components (A)

The number of components used in the PCA of $\mathbf{X}$ (A) is determined either by cross-validation (CV), or by using components with eigenvalues significantly larger than some limit, usually between 1.0 and 2.0. The eigenvalue of a component is the fraction of the SS of the signal matrix $\mathbf{X}$ that is explained by the component, multiplied by the number of variables (K). A component with an eigenvalue larger than one explains the worth of at least one variable. One further alternative (not recommended) is to sufficiently use many components to make the SS of the residuals $\mathbf{E}$ smaller than a fraction of the SS of the signal matrix $\mathbf{X}$.

With CV one estimates the number of components (A) by dividing the elements in the data matrix $\mathbf{X}$ in a number of groups, usually seven or nine. One then makes a number of separate analyses, each time keeping out one of the groups of the data (setting them as ''missing''). The model is then used to predict the deleted data, and the ensuing prediction errors (predicted value minus actual values) are calculated. The resulting SS of these prediction errors is called predictive residuals sum of squares (PRESS), and one selects the model dimensionality (A), which gives the smallest PRESS. This approach is called full CV. Alternatively, one calculates one PC after another with a renewed CV for each component, until PRESS stops to decrease in size, so-called partial CV. Full and partial CV usually indicate the same number of components or very close.

Like with PCR, the number of significant PLS components is determined by CV. Since there is a $\mathbf{Y}$-matrix in PLS, the CV of PLSR is different from that of PCA, and the PRESS relates to the predictive residuals of the $\mathbf{Y}$-matrix. In CV of PLS, the deleted groups of data contain

whole objects (**X** and **Y**-data). A PLS model is developed without this group, and the predicted Y-values are then calculated for the deleted objects. These Y-values are compared with the actual values, and PRESS equals the SS of the difference between observed and predicted **Y**, the predictive residuals. Hence the CV in PLS relates directly to the objective of predicting **Y**, while the CV of PCA tells us only about the significant structure of **X**. Also for PLS, CV can be made in a full or partial way, giving similar results in most practical cases.

8.4.4 Computational Algorithms for Both Regression Methods

PCA is mathematically equivalent to singular value decomposition (SVD) as it is usually called in numerical analysis and computation science, and to principal factor analysis, Karhunen-Loève decomposition, and eigenvector analysis as this is often called in physics and engineering. Hence, PCR starts with a PCA of **X**, followed by a simple MLR using the first few PC score vectors, $\mathbf{t_a}$, and predictor variables.

Many different algorithms exist for the calculation of the principal components parameters, where, historically, the first were based on the diagonalization of the variance covariance matrix $\mathbf{X'X}$. The resulting eigenvectors equal the loadings, $\mathbf{p_a}$. Today, however, the most commonly used algorithms are based on SVD,[37] or on the NIPALS method.[38] The latter is very fast if just a few PCs are needed, and also works for incomplete data matrices.

Like for PCR, many algorithms exist for calculating the PLS parameters. The most common is based on the NIPALS algorithm[39] which has the advantage of also operating with incomplete **X** and **Y** matrices, and to extract only as many components as actually are needed in the model. Other algorithms are based on the calculation of eigenvectors of the matrices $\mathbf{XX'YY'}$ or $\mathbf{X'YY'X}$.[40–42] These two matrices have the dimensions $(n \times n)$ and $(K \times K)$. The first is small for few samples, and hence advantageous for fast computations.

8.4.5 The Example

In the example, a PCA of the first 10 samples (the TS) gives three components as shown above in section 6. The two PCR calibration models for y1 = felodipine and y2 = metoprolol based on the corresponding score vectors explain the two Y-variables well ($R^2 = 0.9996$ and 0.9991, and $Q^2 = 0.9817$ and 0.9153 respectively). The fitted and predicted values for the second model are shown in Figure 6, as well as in Table 1.

8.4.6 Multiple Linear Regression

The first method used in MVC was MLR used with a small selected subset of variables. These variables are usually

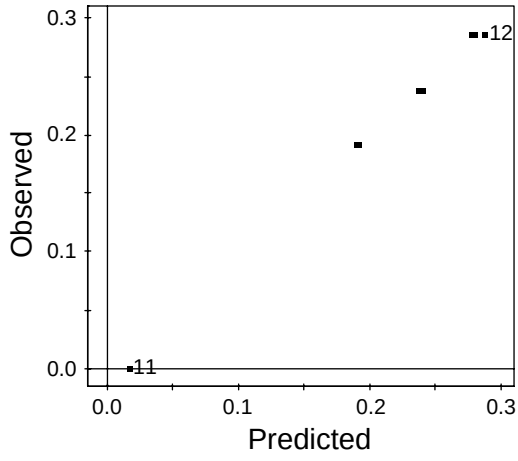

Figure 6 The observed versus fitted or predicted values for y2 = metoprolol in the example. The calibration model was one of PCR (A = 3 components) based on samples 1 to 10 (not labeled). Samples 11 and 12 are prediction samples.

Table 1 Resulting R^2 and Q^2 values of the PCA of the spectral data of the example. Symbols explained in 10.1

A	R^2X	M6 R^2X(cum)	$N = 15$ Q^2	Q^2(cum)
1	0.659	0.659	0.625	0.625
2	0.331	0.989	0.957	0.984
3	0.010	1.000	0.956	0.999

selected by inspecting spectra of fairly pure constituents. The number of variables should be at least two per constituent, i.e. four in the example. In the absence of impurities, the 4 wavelengths 11, 2, 39, and 57 seem like a good selection in this example. A number of automatic variable selection procedures exist for MLR, but the experience shows that a selection based on chemical and spectroscopic knowledge is preferable. In the case where such knowledge is weak or absent, full spectrum methods such as PCR and PLSR should be used at least as starting points to identify important regions in the spectrum.

MLR starts directly with the model in the form of Equation (5). The coefficients **B** that minimize the SS of the residuals in **F** are given by Equation (35):

$$\mathbf{B} = (\mathbf{X'X})^{-1}\mathbf{X'Y} \tag{35}$$

Predicted values of **Y** for new samples are obtained as Equation (36):

$$\mathbf{y'_{new}} = \mathbf{x'_{new}B} \tag{36}$$

The standard errors and ensuing confidence intervals are given in Equation (37) which is analogous to Equation (25), but with **X** instead of **T**, because in MLR all

X-variables are assumed to be independent:

$$\mathrm{serr_y} = \left(1 + \frac{1}{\mathrm{N}} + \mathbf{x'_{new}}(\mathbf{X'X})^{-1}\mathbf{x_{new}} \quad \sigma^2 \right. \qquad (37)$$

9 METHOD PERFORMANCE AND VALIDATION

The use of analytical method validation is crucial for the operation of many regulated businesses such as the pharmaceutical industry, and for quality-rated analytical laboratories. Method validation is an activity that is meant to prove that an analytical method is indeed doing what it is supposed to do. The requirements on the analytical method should correspond to the needs of the operation. For process analysis, speed of analysis may be more important than the lowest possible error in each analysis. When product bulk properties are measured, other requirements are more important than when environmental trace analysis is performed. This means that the first step in method validation is to determine what we want and what requirements that will lead to.

The validation is usually broken down into a set of performance measures such as selectivity, sensitivity, limit of detection, linearity or model fit, repeatability, reproducibility, accuracy and robustness. The possible lowest limits for these performance measures vary with the analytical method. The validation task for a MVC is often to prove that the MVC performs on the same level as the reference method. It is important to notice that the model is not the method. The analytical method has to be a description of the development of calibration models.

9.1 Measures of Model Performance

Today the pharmaceutical industry has a well regulated set of performance measures that has to be reported for an analytical method. Hence these can be used as a framework for MVC. A set of performance measures in pharmaceutical analytical method validation mainly taken from the suggestion list of the International Harmonization Committee (IHC) (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use IHC secretariat c/o IFPMA, 30 rue de St-Jean, PO Box 9, CH-1211 Geneva 18, Switzerland.)[43] will be covered. These measures are mostly developed with UVCs in mind. Their application for MVC will be discussed.

The calibration model can be evaluated according to very many criteria. According to our present knowledge of analytical chemistry, the following set of criteria cover most practical needs.

9.2 Selectivity

The selectivity of an analytical method is defined as the extent to which it can determine particular analytes in a complex mixture without interference from other components in the mixture.

MVC is able to make a selective analysis by the use of mathematics. The requirements for this to work is that the other compounds in a mixture have different magnitude relations for the same spectral variables and that the amounts of the interfering compounds are varied in the TS independently of the analyte amount. This can be tested by the variation of physical properties, matrix components, and interfering compounds to samples in the test set. Then the selectivity as a deviation in the analytical result can be estimated as Q_{test} for this test set, by Equation (38):

$$Q_{\mathrm{test}} = \quad 1 - \frac{S^2_{F,\mathrm{test}}}{S^2_{Y,\mathrm{test}}} \qquad (38)$$

Here $S_{F,\mathrm{test}}$ is the SD of the y-residual for the selected test set and $S_{Y,\mathrm{test}}$ is the SD for the total variation in the y-variables.

When the amounts of interfering compounds are outside the scope of the model, the residual error for the x-variables (**E**) should be large enough to rise beyond a decided limit of acceptance. The verification of the residual diagnostics should be included in the method validation.

Selectivity is related to the concept of ''net analyte signal'' (NAS) of Lorber et al.[44] It is claimed that NAS can be used as a basis for a broad range of MVA diagnostics. In its present form, however, the NAS estimation needs the pure spectra of the constituents of interest, making NAS work only in simple concentration calibrations. Also, the experience with the NAS concept in MVC is too short to allow an evaluation.

9.3 Accuracy

The accuracy of a method can be defined as the closeness between the analytical result and an accepted reference value, sometimes termed the true value.

The estimation of the accuracy should be applied across the range of the analytical method. Reference materials, or a separate well-characterized procedure with known accuracy, may be used for this purpose. The reference method for the MVC is a valid alternative for the comparison if it is well characterized. A report of the accuracy may include a root mean square error of prediction (RMSEP) for the test set in Equation (39):

$$\mathrm{RMSEP} = \frac{\sum_{\mathrm{i}=1}^{n}(y_{\mathrm{i,ref}} - y_{\mathrm{i,pred}})^2}{n} \qquad (39)$$

where $y_{i,ref}$ denotes the values obtained from the reference method and $y_{i,pred}$ the values predicted by the MVC. This may be complemented by an estimate of the average bias between the reference method and the MVC results for the test set, in Equation (40):

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(y_{i,ref} - y_{i,pred}) \quad (40)$$

For pharmaceutical-method validation with one y-variable, a minimum of nine determinations is recommended by IHC over a minimum of three concentration levels. For the MVC with more than one y-variable this may be extended to minimum three levels in each y-variable.

9.4 Precision

The precision of an analytical method can be defined as the closeness between repeated measurement of subsamples taken from the same homogeneous sample. The sample should be authentic. Three levels of precision measurements are defined by IHC: repeatability, intermediate precision, and reproducibility. The precision is often reported as the variance, SD, or coefficient of variation of a series.

9.4.1 Repeatability

The repeatability expresses the variation during a short interval of time with the same operating conditions. For pharmaceutical use, a minimum is six determinations at a 100% level of the analyte, or nine determinations covering the analytical range.

9.4.2 Intermediate Precision

The intermediate precision should include possible variations within a single laboratory, such as different days, different analysts, and different equipment. It is not necessary to study these effects one by one. For pharmaceutical use, the experimental design approach is encouraged. Then the sizes of e.g. the day, analyst and equipment effects are quantified as coefficients in a regression for evaluation of the experimental design. See section 9.10 below.

9.4.3 Reproducibility

The reproducibility can be defined as a case when both within and between laboratory variations are included in the evaluation of the precision. This involves an inter-laboratory trial. For certain types of MVC, e.g. NIR calibrations, this level of precision involves the topic of calibration transfer or the use of well-accepted standards. As with intermediate precision, the results can be evaluated using an experimental design.

9.5 Sensitivity

Sensitivity, or how low are the concentrations that an instrument can detect, can be defined in many ways, where the most common are detection and quantitation limits.

9.5.1 Detection Limit

The detection limit can be defined as the lowest amount of analyte that is still detectable by the method but not necessarily quantifiable with good precision.

The detection limit may be estimated in ways such as visual evaluation, signal-to-noise ratio (S/N) (often 3 : 1 or 2 : 1), SD of the blank, a calibration curve on the same level as the detection limit, or an extrapolation of the calibration curve. When extrapolation is used it should be justified with test samples.

The multivariate spectral S/N can be defined (Equation 41) as the net signal ($S_X^2 - S_E^2$) divided by the noise:

$$\text{S/N} = \frac{\sqrt{S_X^2 - S_E^2}}{\sqrt{S_E^2}} \quad (41)$$

where S_X is the SD of the gross signal i.e. the SD over all measured samples and variables, and S_E is the SD of the noise i.e. the spectral x-residuals.

The MVC detection limit is, however, more directly dependent of the S_F. which is the SD of the y-residuals. Hence a multivariate detection limit can be estimated as $3S_F$.

For pharmaceutical use this estimation is recommended for nonquantitative limit tests, i.e. classification methods, while the quantitation limit is recommended in use of quantitative calibrations at low analyte levels.

9.5.2 Quantitation Limit

The quantitation limit can be defined as the lowest level of the analyte where the calibration is still working with a suitable accuracy and precision. The same methods as for the detection limit are applicable but a recommended S/N is now increased to 10 : 1. Analogously the multivariate quantitation limit can be estimated as $10 \cdot S_F$.

For pharmaceutical use, this estimation is recommended for low analyte levels, but it is not necessary if the calibration works in a range well above the low levels for the given analytical method.

9.6 Model Fit or Linearity

The linearity of an analytical procedure may be defined as the ability to give results that are proportional to the

analyte concentration in the sample within a given range. This works well with univariate linear regression. For MVC there are additional measures of performance.

The model fit may be based on diagnostics such as percent explained variation, CV, t vs. u score plots, calibration y-residuals vs. y-size, multivariate correlation coefficient, etc.

9.6.1 Degree of Fit

The size of the ''leftovers'' after modelling, i.e. the residuals, is useful as a measure of how well the model fits the data. This is often re-expressed as the amount of variation explained, and denoted as R^2. This can, in projection models, be calculated both for the Y and the X parts of the data, Equations (42) and (43):

$$R^2_{\mathrm{Y}} = \overline{1 - \frac{S^2_F}{S^2_Y}} \tag{42}$$

$$R^2_{\mathrm{X}} = \overline{1 - \frac{S^2_E}{S^2_X}} \tag{43}$$

Here the residual variances, S^2_F and S^2_E, as well as the variances of the data, S^2_Y and S^2_X, are uncorrected for degrees of freedom, i.e. the sums of squares divided by $\mathrm{N} \times \mathrm{M}$ and $\mathrm{N} \times \mathrm{K}$, respectively. If one uses the variance corrected for degrees of freedom, which for the residuals **F** and **E** are $(\mathrm{N} - \mathrm{A} - 1) \times (\mathrm{M} - \mathrm{A})$ and $(\mathrm{N} - \mathrm{A} - 1) \times (\mathrm{K} - \mathrm{A})$, respectively, the R^2 s are called ''adjusted''.

The square roots of these variances, then usually corrected for the degrees of freedom, are called the RSDs, of Y and X, as given above in Equation (33).

9.6.2 Cross-validation

If the *predictive* residuals from CV (8.4.3) are used instead of the ''fitted'' residuals in Equations (42) and (43) we get the cross-validated R^2 s. These are often denoted by Q^2_{Y} and Q^2_{X}. Here Q^2_{Y} is usually the only one of interest, and it is often denoted just by Q^2. The corresponding RSD based on predictive residuals is often called SECV (standard error of cross-validation). One should note that the denominator in this is always N, since there is no loss of degrees of freedom in the CV process.

9.6.3 Prediction of Validation or Prediction Set

Analogously, if the residuals from a real prediction with a new prediction or test set are available, they can be used to calculate a predictive Q^2, denoted by Q^2_{pred}, with the corresponding predictive RSD, denoted by SEP (standard error of prediction).

9.7 Range

The range of an analytical method can be defined as the interval between the upper and the lower levels of analyte present in the samples, where the precision, accuracy and model fit are suitable for the analysis.

Since MVC involves the use of a local model, extrapolations will not be accurate. This means that the range will be limited approximately to the range of levels entered in the CS used during method validation. Since samples outside the validated range may occur, it is advisable to include a path for extension of the range in the analytical MVC method together with its validation e.g. using the diagnostic tools for model fit above.

9.8 Robustness

The robustness can be defined as the measure of the ability of the analytical method to remain unaffected by small, but purposely made disturbances in method parameters. These are: measurement errors during sample pretreatment, temperature, humidity, variations in grinding, the stability of analytical solutions, extraction times, influence of chromatographic mobile-phase composition, flow rate, different lots of columns, etc. The evaluation of the robustness indicates the level of reliability of the method. Individual factors that may harm the robustness can be identified and compensated for, if the robustness test is made according to an experimental design.

9.9 System Suitability Testing

The system suitability testing can be made as an integral part of an analytical method. The system suitability test is aimed to make an overall test of the entire function of the analytical equipment used.

During the use of a MVC model for prediction, the RSD from the calibration x-variables is useful as a system suitability test.

9.10 Validating the Calibration Model (Including its Robustness)

The investigation of all points above is very time-consuming unless an experimental design is made taking into factors that may influence the results of the MVC. The ranges of the design factors should be small and correspond to around three times the range of experimental control of these factors. The center of the design should be the conditions specified for the analytical method. Since the variations are small, a reduced design aimed at a linear model for evaluation is expected.

If these design factors include ''day of analysis'', ''analyst'', ''instrument'' (equipment), and ''laboratory'', the results will be informative both for intermediate

precision and robustness. If the center point of the design is repeated six times the results also indicate the repeatability of the method.

We must remember that this validation (and robustness study) indicates the performance of the whole analytical process, from sample preparation to the stability of the instrument, and that the MVC is just a part of this process. A common and good result from robustness studies will be that none or few factors are affecting the analytical result. This is reflected in the evaluation of the design as no model present or only a few significant factors.

10 CALIBRATION TRANSFER (BETWEEN INSTRUMENTS, LABS, AND OVER TIME)

Calibration transfer is a process where a calibration model developed for one or more instruments is going to be applied to data from additional instruments not included in the primary calibration model. If the variations are large and the instrument is linear, external standards may be applied. This often holds for UV/VIS spectrometry in liquids and LC. In e.g. NIR diffuse reflectance measurements the magnitude range that is actually used is much smaller (see Figure 3b). For these types of cases, a very local calibration in one instrument has to be moved to another instrument. To preserve the accuracy of the analysis, the local magnitude range of the secondary instrument has to be accurately mapped. This is best done with the same type of samples as the ones present in the MVC.[45]

When the calibration samples are stable with time, the most straightforward solution is to build a new calibration for the secondary instrument but with the same samples and reference analysis results.

Another approach that works for e.g. the same brand and type of NIR instruments even with non-stable samples, is to complement a previous PLS-calibration model by adding samples measured with the new instrument. When the calibration for the primary instrument is well developed also for the interfering variations, it may be sufficient with 10–20% of the original number of samples to span the calibration model also for the secondary instrument. This updating scheme will also work for the same instrument over time.

A variation of this is to make deliberate variations of the instrument parameters and develop the MVC for a set of instruments of the same type from the beginning.[46] This is common in the agricultural sector with calibrations for e.g. protein in wheat. It is also common to adjust this type of calibration for bias and slope on the y-side for the individual instruments.

Another approach is to make the instruments as similar as possible. This route has been taken e.g. by Perkin Elmer for an oil analysis NIR spectrometer.[47] Here the optical bandwidth is artificially degraded to a preset level in order to be constant for all instruments. In addition to this, corrections are made to the wavelength scale and the absorbance scale.

When working in diffuse reflectance mode, the blank spectrum is usually defined by a material that has a flat NIR spectrum. This material should be stable with time. Often a PTFE standard is used.[48] Ceramic is another common material. The batch to batch variations in these materials may lead to different zero levels for the obtained spectra. In some cases is sufficient to bias correct all variables in the spectra[28,29] in such a way that they appear to have been measured with the primary instrument, and then to use the primary calibration model for the adjusted spectra. The relative differences of the zero levels may be established by a set of standards that is measured with both instruments.

Shenk has developed another transfer method.[49,50] Here the wavelength scale is scaled by a wavelength standard, then the intensity scale is scaled in an univariate way for each wavelength using approximately 30 standards.

Direct standardization (DS) involves the use of a normal MVC with spectra from standards measured on the secondary instrument as x-variables and spectra from the same standards measured with the primary instrument as y-variables. This technique has been shown to work by Hanssen.[51] One problem here is that the random variations in the translated spectra are lost. Thus the variations that result in large residual errors for the primary MVC will be masked in the translation step. To compensate for this, the residual error should be monitored with the MVC for translation.

Piecewise direct standardization (PDS) is another spectrum modifying technique.[52] Adjacent x-variables have to be continuous and correlated, as they are e.g. in spectra or chromatograms. Here the main idea is to correct both for magnitude and wavelength differences at the same time. This is done by measuring a set of transfer standards that are similar to the measured samples. Then a sliding window for calculations is used along with the standards to build a series of PCR or PLS calibration models with the window range in the secondary instrument as x-variables to the window center of the primary instrument as y-variable. The transfer mechanism is then to apply the array of models to translate spectra from e.g. the secondary spectrometer to spectra as measured by the primary spectrometer.

With PDS the transfer models for some variables are likely to perform less well. This is visible as discontinuities in the translated spectra. In many cases these discontinuities are sufficiently small to be ignored but in other cases, certain regions of the spectra may have to be removed. This may occur for grating spectrometers at the

wavelengths where the Wood's anomaly occurs,[53] and for sections of the spectra that have less dynamic range.

It is also possible to use the OSC[27] for the removal of spectral instrument differences.[28,29] In this case an OSC correction model is developed by measuring a set of samples of the same type as in the primary calibration. They are measured on both instruments in the same manner as for PDS. Spectra from both instruments are then put together in the OSC calculation with the corresponding reference method y-values. Since the instrumental variation is uncorrelated with the y-value variation, the OSC will be able to correct for the instrument effects.

11 SUMMARY AND CONCLUSIONS

MVC is a recently developed approach to use the multidimensional data emerging from spectrometers and other analytical instruments to indirectly estimate properties of interest in chemical, biological, environmental, and other complicated samples. These properties often are concentrations, but, particularly in process analytical chemistry, MVC is often used to calibrate for other qualities such as strength, film thickness, and viscosity. Due to its novelty, the approach is still not fully developed, and interesting research remains regarding both the applicability of MVC in various areas, and the further development of the methodology toolbox. However, MVC is starting to be used extensively in the chemical industry, which is seen in the appearance of introductory books for the industrial analytical chemist.[5]

The idea of using ''all'' signals instead of one selected frequency or wavelength is simple but powerful. Simple mathematics is used to combine the signals so that the resulting combinations are much more selective, and usually also much more precise, than the original individual signals. Provided that the instrumental data, the signals X, are *relevant* for the actual measurement problem, and provided that the TS is representative, and the TS Y-values are sufficiently accurate, MVC works very well.

MVC needs a computer for the computations, which is a major reason why this approach was not used before about 1970. Today, when most analytical instruments come integrated with a computer, MVC is the natural choice for most calibrations.

LIST OF SYMBOLS

a	component index ($a = 1, \ldots, A$)
$\mathbf{b}$	vector of coefficients ($K \times 1$) that combine the $\mathbf{X}$ columns to map $\mathbf{y}$
$\mathbf{B}$	matrix of coefficients ($K \times M$) that combine the $\mathbf{X}$ columns to map $\mathbf{Y}$
$\mathbf{c}$, $\mathbf{C}$	concentration vector and matrix
$\mathbf{E}$	matrix of $\mathbf{X}$-residuals; $\mathbf{E_a}$ denotes the $\mathbf{X}$-residuals after a components
$\mathbf{F}$	matrix of $\mathbf{Y}$-residuals
$\mathbf{H}$	a general mathematical function, i.e. $\mathbf{Y} = \mathbf{H(X, B)} + \mathbf{F}$
i	sample index ($i = 1, 2, \ldots, n$)
k	index of $\mathbf{X}$-variables, columns ($k = 1, 2, \ldots, K$)
K	number of $\mathbf{X}$-variables
m	index of $\mathbf{Y}$-variables, columns ($m = 1, 2, \ldots, M$)
M	number of $\mathbf{Y}$-variables
n, M	number of calibration (CS/TS) samples
$\mathbf{p_a}$	$\mathbf{X}$-loading vector ($K \times 1$)
$\mathbf{P}$	$\mathbf{X}$-loading matrix ($K \times A$)
$\mathbf{s_i}$	residual SD of one row in $\mathbf{E} = \|\mathbf{e_i}/(\mathrm{K} - \mathrm{A})\|$
$\mathbf{t_a}$	$\mathbf{X}$-score vector ($n \times 1$) of component a
$\mathbf{T}$	matrix of $\mathbf{X}$-score vectors ($n \times A$)
$\| \mathbf{v} \|$	the length (norm) of the vector $\mathbf{v}$
$\mathbf{w_a}$	implicit $\mathbf{X}$-weight vector in PLS models ($K \times 1$): $\mathbf{t_a} = \mathbf{E_{a-1}w_a}$
$\mathbf{w_a^*}$	explicit $\mathbf{X}$-weight vector in PLS models ($K \times 1$); $\mathbf{t_a} = \mathbf{Xw_a^*}$
$\mathbf{W}$	implicit $\mathbf{X}$-weight matrix in PLS models ($K \times A$)
$\mathbf{W^*}$	explicit $\mathbf{X}$-weight matrix in PLS models ($K \times A$)
$\mathbf{x_i'}$	row vector ($1 \times K$) of one observation in $\mathbf{X}$ (e.g. one digitized spectrum)
$\mathbf{X}$	TS matrix ($n \times K$) of preprocessed spectra (or other signal profiles)
$\mathbf{y}$	TS column vector ($n \times 1$) of concentrations or other property
$\mathbf{Y}$	TS matrix of concentrations ($n \times M$) or other property
$\mathbf{z}$, $\mathbf{Z}$	raw data of digitized spectra (before transformation and preprocessing)
$'$	denotes the transpose of a vector or a matrix, e.g. $\mathbf{x}'$

ABBREVIATIONS AND ACRONYMS

AOTF	Acousto-optical Tuneable Filters
ATR	Attenuated Total Reflection
CS	Calibration Set
CV	Cross-validation
DS	Direct Standardization
FT	Fourier Transform
GC	Gas Chromatography
ICP	Inductively Coupled Plasma
ICP/MS	Inductively Coupled Plasma/Mass Spectrometry

IHC	International Harmonization Committee
IR	Infrared
LC	Liquid Chromatography
MLR	Multiple Linear Regression
MS	Mass Spectrometry
MSC	Multiplicative Signal Correction
MVC	Multivariate Calibration
NAS	Net Analyte Signal
NIR	Near-infrared Spectroscopy
NMR	Nuclear Magnetic Resonance
OSC	Orthogonal Signal Correction
PCA	Principal Components Analysis
PCR	Principal Components Regression
PDA	Photo Diode Array Spectrometer
PDS	Piecewise Direct Standardization
PLS	Partial Least Squares
PLS/MVC	Partial Least-squares Multivariate Calibration
PLSR	Partial Least-squares Regression
PRESS	Predictive Residuals Sum of Squares
RMSEP	Root Mean Square Error of Prediction
RSD	Residual Standard Deviation
SD	Standard Deviation
SECV	Standard Error of Cross-validation
SEP	Standard Error of Prediction
S/N	Signal-to-noise Ratio
SNV	Standard Normal Variate
SS	Sum of Squares
SVD	Singular Value Decomposition
TLC	Thin-layer Chromatography
TS	Training Set
UV	Ultraviolet
UVC	Univariate Calibration
UV/VIS	Ultraviolet/Visible
2-D	Two-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance, General Medical • Multinuclear Magnetic Resonance Spectroscopic Imaging

Food **(Volume 5)**
Flavor Analysis in Food • Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Near-infrared Spectroscopy in Food Analysis • Water Determination in Food

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels • Petroleum Residues, Characterization of

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis • Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships

Process Instrumental Methods **(Volume 9)**
Process Analysis: Introduction • Chemometric Methods in Process Analysis • Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Clustering and Classification of Analytical Data • Second-order Calibration and Higher • Signal Processing in Analytical Chemistry • Soft Modeling of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. H. Martens, T. Naes, *Multivariate Calibration*, Wiley, New York, 1989.
2. N.R. Draper, H. Smith, *Applied Regression Analysis*, 2nd edition, Wiley, New York, 1981.
3. O.L. Davies, P.L. Goldsmith (eds.), *Statistical Methods in Research and Production*, 4th edition, Longman, London, 1980.
4. J.C. Miller, J.N. Miller, *Statistics for Analytical Chemistry*, 3rd edition, Ellis Horwood, New York, 1993.
5. K.R. Beebe, R.J. Pell, M.B. Seasholtz, *Chemometrics, a Practical Guide*, Wiley, New York, 1998.
6. R.G. Krutchoff, 'Classical and Inverse Regression Models of Calibration', *Technometrics*, **9**, 425–439 (1967).
7. B. Hoadley, 'A Bayesean Look at Inverse Linear Regression', *J. Am. Stat. Assoc.*, **65**, 356–369 (1970).
8. T. Naes, 'Comparison of Approaches of Multivariate Linear Calibration', *Biometrical J.*, **27**, 267–275 (1985).

9. W. Plugge, C. van der Vlies, 'Near-infrared Spectroscopy as an Alternative to Assess Compliance of Ampicillin Trihydrate with Compendial Specifications', *J. Pharm. Biomed. Anal.*, **11**(6), 435–442 (1993).
10. S. Wold, 'Pattern Recognition by Means of Disjoint Principal Components Models', *Pattern Recognition*, **8**, 127–139 (1976).
11. J.V. Kresta, J.F. MacGregor, T.E. Marlin, 'Multivariate Statistical Monitoring of Process Operating Performance', *Can. J. Chem. Eng.*, **69**, 35–47 (1991).
12. M. Josefson, E. Johanson, A. Torstensson, 'Optical Fiber Spectroscopy in Turbid Solutions by Multivariate Calibration Applied to Tablet Dissolution Testing', *Anal. Chem.*, **60**, 2666–2671 (1988).
13. B. Walczak, B. Bogaert, D.L. Massart, 'Application of Wavelet Packet Transform in Pattern Recognition of Near-IR Data', *Anal. Chem.*, **68**, 1742–1747 (1996).
14. J. Trygg, S. Wold, 'PLS Regression on Wavelet Compressed NIR Spectra', *Chemom. Intell. Lab. Syst.*, **42**, 209–220 (1998).
15. E. Jellum, I. Björnson, R. Nesbakken, E. Johansson, S. Wold, 'Classification of Human Cancer Cells by Means of Capillary Gas Chromatography and Pattern Recognition Analysis', *J. Chromatogr.*, **217**, 231–237 (1981).
16. J.S. Swan, D. Howie, 'Correlation of Sensory and Analytical Data in Flavor Studies into Scotch Malt Whisky', *Found. Biotech. Industr. Fermentation*, **3**, 291–309 (1984).
17. M. Forina, G. Drava, 'Chemometrics for Wine Applications', *Analusis*, **25**(3), 38–42 (1997).
18. G. Malmquist, R. Danielsson, 'Alignment of Chromatographic Profiles for Principal Component Analysis: a Prerequisite for Fingerprinting Methods', *J. Chromatogr. A*, **687**, 71–88 (1994).
19. L. Ståhle, 'Three-way PLS Regression', *Chemom. Intell. Lab. Syst.*, **7**, 95–100 (1989).
20. R. Bro, 'Multiway Calibration, Multi-linear PLS', *J. Chemometrics*, **10**, 47–62 (1996).
21. P. Geladi, H. Grahn, *Multivariate Image Analysis*, John Wiley & Sons, Chichester, UK, 1996.
22. K.H. Esbensen, M. Halstensen, et al. 'Acoustic Chemometrics – from Noise to Information', *Chemom. Intell. Lab. Syst.*, **44**, 61–76 (1998).
23. I.S. Helland, T. Naes, T. Isaksson, 'Related Versions of the Multiplicative Scatter Correction Method for Preprocessing Spectroscopic Data', *Chemom. Intell. Lab. Syst.*, **29**, 233–241 (1995).
24. M.S. Dhanoa, S.J. Lister, R. Sanderson, R.J. Barnes, 'The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformation of NIR Spectra', *J. Near Infrared Spectrosc.*, **2**, 43–47 (1994).
25. R.J. Barnes, M.S. Dhanoa, S.J. Lister, 'Standard Normal Variate Transformation and De-trending of Near Infrared Diffuse Reflectance Spectra', *Appl. Spectrosc.*, **43**, 772–777 (1989).
26. T.V. Karstang, R. Manne, 'Optimized Scaling. A Novel Approach to Linear Calibration with Closed Data Sets', *Chemom. Intell. Lab. Syst.*, **14**, 165–173 (1992).
27. S. Wold, H. Antti, F. Lindgren, J. Öhman, 'Orthogonal Signal Correction of Near-infrared Spectra', *Chemom. Intell. Lab. Syst.*, **44**, 175–185 (1998).
28. J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold, 'An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near-infrared Spectra', *Chemom. Intell. Lab. Syst.*, **44**, 229–244 (1998).
29. P. Geladi, H. Barring, E. Dabakk, J. Trygg, H. Antti, S. Wold, A. Karlberg, 'Calibration Transfer for pH Prediction in Lake Sediment Samples', *J. NIR Spectrosc.*, (1999) (in press).
30. A. Savitzky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Anal. Chem.*, **36**, 1627–1639 (1964).
31. K.H. Norris, P.C. Williams, *Cereal Chem.*, **61**, 158 (1984).
32. Ref. 2, 327.
33. S. Wold, H. Martens, H. Wold, 'The Multivariate Calibration Method in Chemistry Solved by the PLS Method', in *Proc. Conf. Matrix Pencils*, eds. A. Ruhe, B. Kågström, Lecture Notes in Mathematics, Springer Verlag, Heidelberg, 286–293, 1983.
34. W. Lindberg, J.-Å. Persson, S. Wold, 'Partial Least-squares Method for Spectro-fluorimetric Analysis of Mixtures of Humic Acid and Ligninsulfonate', *Anal. Chem.*, **55**, 643–648 (1983).
35. A. Höskuldsson, 'PLS Regression Methods', *J. Chemometrics*, **2**, 211–228 (1988).
36. A. Höskuldsson, *Prediction Methods in Science and Technology*, Thor Publishing, Copenhagen, Vol. 1, 1996.
37. G.H. Golub, W. Kahan, 'Calculating the Singular Values and Pseudoinverse of a Matrix', *SIAM J. Numer. Anal.*, **B2**, 205–224 (1965).
38. H. Wold, 'Nonlinear Estimation by Iterative Least Squares Procedures', in *Research Papers in Statistics*, ed. F. David, John Wiley & Sons, New York, 411–444, 1966.
39. S. Wold, A. Ruhe, H. Wold, W.J. Dunn III, 'The Collinearity Problem in Linear Regression. The Partial Least Squares Approach to Generalized Inverses', *SIAM J. Sci. Stat. Comput.*, **5**, 735–743 (1984).
40. F. Lindgren, P. Geladi, S. Wold, 'The Kernel Algorithm for PLS', *J. Chemometrics*, **7**, 45–59 (1993).
41. S. Rännar, F. Lindgren, P. Geladi, S. Wold, 'A PLS Kernel Algorithm for Data Sets with Many Variables and Less Objects. Part I. Theory and Algorithms', *J. Chemometrics*, **8**, 111–125 (1994).
42. B.S. Dayal, J.F. MacGregor, 'Improved PLS Algorithms', *J. Chemometrics*, **11**, 73–85 (1997).
43. The International Harmonization Committee (IHC) Q2: Validation of Analytical Procedures A: Definitions and Terminology http://www.ifpma.org/pdfifpma/q2a.pdf Q2: Validation of Analytical Procedures B: Methodology http://www.ifpma.org/pdfifpma/q2b.pdf

44. A. Lorber, K. Faaber, B.R. Kowalski, 'Net Analyte Signal Calculation in Multivariate Calibration', *Anal. Chem.*, **69**, 1620–1626 (1997).
45. O. de Nord, 'Multivariate Calibration Standardization', *Chemom. Intell. Lab. Syst.*, **25**, 85–97 (1994).
46. N.B. Büchmann, S. Runfors, 'The Standardisation of Infratech 1221 Near Infrared Transmission Instruments in the Danish Network used for the Determination of Protein and Moisture in Grains', *J. Near Infrared Spectrosc.*, **3**, 35–42 (1995).
47. J. Workman, J. Coates, 'Multivariate Calibration Transfer. The Importance of Standardizing Instrumentation', *J. Spectrosc.*, **8**(9), 36–42 (1993).
48. S.V. Hammond, 'Networking Qualitative Databases of Pharmaceutical Raw Materials', Lecture at International Conference on Pharmaceutical Applications of NIR Spectroscopy, Stockholm, Sweden, June 13–15, 1996.
49. J.S. Shenk, M.O. Westerhaus, US Patent No. 4866644, Sept 12, 1989.
50. Y. Wang, D.J. Veltkamp, B.R. Kowalski, 'Multivariate Instrument Standardization', *Anal. Chem.*, **63**, 2750–2756 (1991).
51. W.G. Hansen, 'Production Control by Near Infrared Spectroscopy', in *Leaping Ahead Near Infrared Spectrosc.*, [Proc. Int. Conf. Near Infrared Spectrosc.], 6th Int. Conf., NIR, 376–81, 1994.
52. Y. Wang, D.J. Veltkamp, B.R. Kowalski, 'Multivariate Instrument Standardization', *Anal. Chem.*, **63**, 2750–2756 (1991).
53. R.W. Wood, 'On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum', *Philos. Mag.*, **4**, 396–402 (1902).

Second-order Calibration and Higher

Cliona M. Fleming and Bruce R. Kowalski
University of Washington, Seattle, USA

1 Introduction **1**
2 Three-way Models **4**
3 Calibration Methods **5**
3.1 Alternating Least-squares Techniques 6
3.2 Rank Annihilation Methods 12
3.3 Multiway Partial Least Squares 19
4 Calibration Transfer/Standardization **20**
5 Standard Addition **22**
6 Figures of Merit **24**
7 Future Work **26**
Acknowledgments **26**
Abbreviations and Acronyms **26**
Related Articles **26**
References **27**

Most analytical chemists recognize calibration as the quantitation of analytes by relating the analyte signal and analyte concentration, although it is also possible to obtain information about other characteristics of the sample through calibration. In the calibration process, a number of samples that have known values of the characteristic to be determined, e.g. concentration, are measured. These are reference measurements, and are known as calibration standards. Traditionally, the measurements have consisted of a single number (a univariate or scalar measurement); they are plotted against the known concentration, to generate a ''calibration curve''. (An example of a scalar measurement is a pH value or the area under the curve of a chromatographic peak.) When unknown (test) samples are measured under the same experimental conditions as the standards, their concentrations may be found by regressing their measured signals against the standard calibration curve. However, a significant disadvantage of univariate calibration is that a sensor must be fully selective for the analyte of interest, since any interfering species that may be present cannot be detected when measurements are scalar. This has led to the development of ''first-order'' calibration methods. These techniques use the relationship between an array of measurements (such as a chromatogram or spectrum) and the analyte concentration to develop a calibration model from which analyte concentrations in test samples can be determined. The array of measurements is a vector, and is known as a first-order tensor, while the calibration model is equivalent to the calibration curve. Since first-order calibration uses many measurements per sample to generate a model, rather than just a single measurement as in the univariate case, it has a number of advantages. For example, multianalyte analysis is made possible as long as there is a standard available for each analyte present in a sample, and outlier detection is also feasible. These types of methods have become popular in recent years, and include partial least squares (PLS) and principal component regression (PCR). It makes sense that when more data are available per sample, more information may be extracted, and this is the source of the success of first-order methods. It therefore follows that, if an analyte signal consists of a matrix of data (also known as a second-order tensor), even more advantages exist. [Examples of instruments that produce matrices of data for each sample analyzed include liquid chromatography/ultraviolet detection (LC/UV) and gas chromatography/mass spectrometry (GC/MS).] This situation allows powerful calibration techniques, known as second-order methods, to be used. The primary advantage of these methods is known as the ''second-order advantage''; it allows analytes to be quantitated even in the presence of unmodeled interferents. In other words, the calibration standards do not have to contain any information about the interfering species in the sample. This is of particular value in situations where complex samples are being analyzed, as they do not have to be fully characterized for quantitation of a single analyte. The disadvantage of these methods, however, is their complexity: there are often decisions to be made, such as the values to assign to various parameters, and each method has analysis situations which are more favorable than others. The most common second-order analysis methods are alternating least-squares (ALS) methods, such as parallel factor analysis (PARAFAC) and multivariate curve resolution (MCR), and eigenvalue–eigenvector based methods, such as the generalized rank annihilation method (GRAM) and direct trilinear decomposition (DTD).

1 INTRODUCTION

Calibration is the mathematical and statistical process of extracting information, usually analyte concentration, from an instrument signal. The analytical instruments can be classified according to the tensor order of the data they are capable of acquiring.[1–3] A single sensor (e.g. ion-selective electrode) generates a single number,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

a scalar or zero-order tensor, per sample. Calibration of zero-order data requires the use of univariate methods, where a calibration curve is generated by plotting standard concentrations versus the instrument response. However, the sensor must be fully selective, since if an interferent is present it cannot be detected. An array of sensors, or a spectrometer, generates an array of measurements, a vector or first-order tensor, per sample. In this case, the existence of many measurements per sample makes multivariate analysis and outlier detection possible. Interferents are not a problem as long as they are included in the calibration design. The calibration methods that are suited to this type of data include PLS and PCR, which have become common analysis tools in today's laboratories.

Recent years have seen the increased use of analytical instruments that produce so-called multiway data, classed as having a tensor order of two (a matrix) or higher.[3] ''Multiway'' data consist of at least a matrix of numbers per sample analyzed. Powerful calibration methods exist specifically for the analysis of such data, and these methods produce the so-called ''second-order advantage''.[4,5] This advantage allows analytes to be quantified even in the presence of unknown interferents, along with the resolution of the analyte signal in each order. Another advantage of higher-order tensor analysis is the possibility of automatic correction of instrument malfunction and/or drift or baseline changes.

When a matrix of data is produced per sample, the data are known as ''second-order'' data, because it is produced by a second-order instrument. Some examples of techniques which produce second-order data are the GC/MS, liquid chromatography/mass spectrometry (LC/MS), tandem mass spectrometry (MS/MS) and LC/UV. Fluorescence excitation/emission is also an example of such an instrument. Figure 1(a) and (b) show the three-dimensional image of a sample that was measured by comprehensive two-dimensional gas chromatography (GC × GC).

When multiple samples are analyzed simultaneously, the matrices are combined to form a block of data, often referred to as a three-way data set. For LC/MS data, use of a second-order calibration method would result in the resolution of the chromatographic profile for each analyte, along with the appropriate mass spectrum. With knowledge of the analyte concentrations in the calibration sample, their quantitation in the unknown samples is also possible.

Instruments also exist that produce even higher order data, although they are less common. Liquid chromatography/tandem mass spectrometry (LC/MS/MS) is an example of the type of data one might obtain from such an instrument. The resulting data from this instrument are in the form of a block of data per sample, creating a

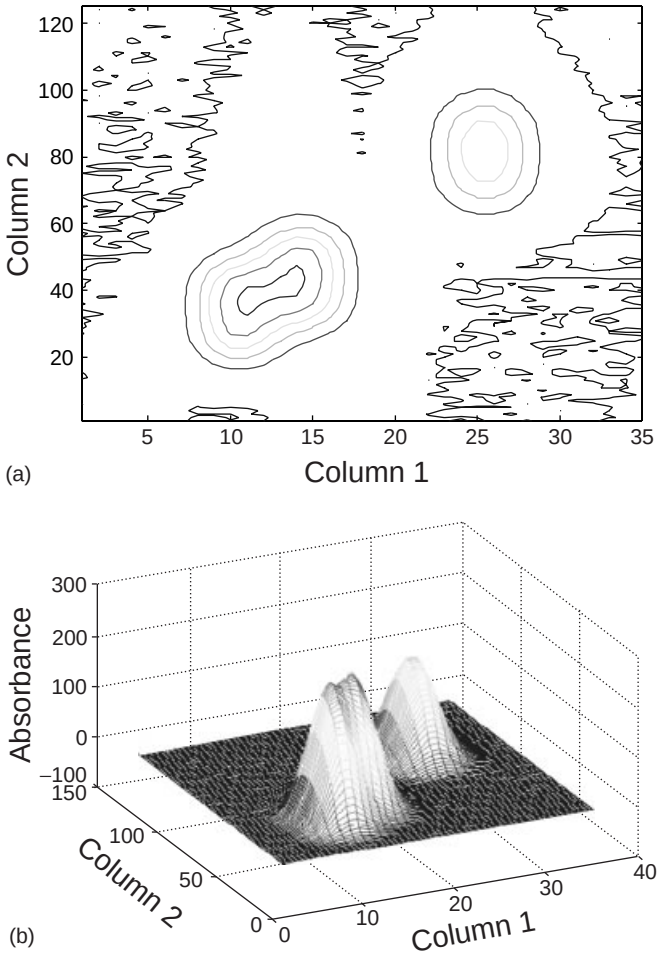

Figure 1 GC × GC instrument response for a sample containing three analytes. (a) Contour plot, where the x- and y-axes represent the separations performed by the first and second columns, respectively; (b) three-dimensional view, showing the intensity of the responses on the z-axis.

third-order tensor, and therefore a collection of samples analyzed together will produce a fourth-order tensor. Some of the same calibration methods can be used to analyze both second- and higher-order data. The methods used, however, depend on the underlying structure of the data.

A second-order data matrix may be classed as bilinear or nonbilinear. Data can be described as being bilinear if the pseudo-rank of the matrix (i.e. rank in the absence of noise) equals the number of chemical components in the sample. This type of data results when one order modulates the other, such as in techniques that combine separation with spectrometry. In GC/MS data, for example, the mass spectrum of an analyte will not change as it elutes except in magnitude. Equation (1) represents the nth analyte in a bilinear matrix:

$$\mathbf{M}_{b,n} = \mathbf{x}_n c_n \mathbf{y}_n^T + \mathbf{E} \tag{1}$$

where $\mathbf{M}_{b,n}$ is the bilinear data matrix for the nth analyte, $\mathbf{x}$ is the analyte profile in the first-order (e.g. pure elution profile), $\mathbf{y}$ is the analyte profile in the second-order (e.g. pure spectrum), c is the concentration of the analyte and $\mathbf{E}$ represents the unmodeled error. The superscript T represents the transpose operator. A mixture of chemical components is therefore represented by Equation (2):

$$\mathbf{M}_b = \sum_{n=1}^{N} \mathbf{M}_{b,n} + \mathbf{E} = \sum_{n=1}^{N} \mathbf{x}_n c_n \mathbf{y}_n^T \tag{2}$$

where N is the number of chemical components present in the mixture. The decomposition of a bilinear data matrix is depicted in Figure 2.

Nonbilinear data, on the other hand, may be represented by $\mathbf{M}_{nb}$ (Equation 3):

$$\mathbf{M}_{nb} = \sum_{p=1}^{P} \mathbf{q}_p r_p \mathbf{s}_p^T + \mathbf{E} \tag{3}$$

which describes a matrix that contains a single pure component,[6] where P is the rank of $\mathbf{M}_{nb}$, $(\mathbf{q}, \mathbf{s})$ is a set of vectors which span the column and row spaces of $\mathbf{M}_{nb}$ and each r is a scaling factor that normalizes the corresponding $(\mathbf{q}, \mathbf{s})$ vectors. In other words, a pure component response has a pseudo-rank >1 when the data are nonbilinear.

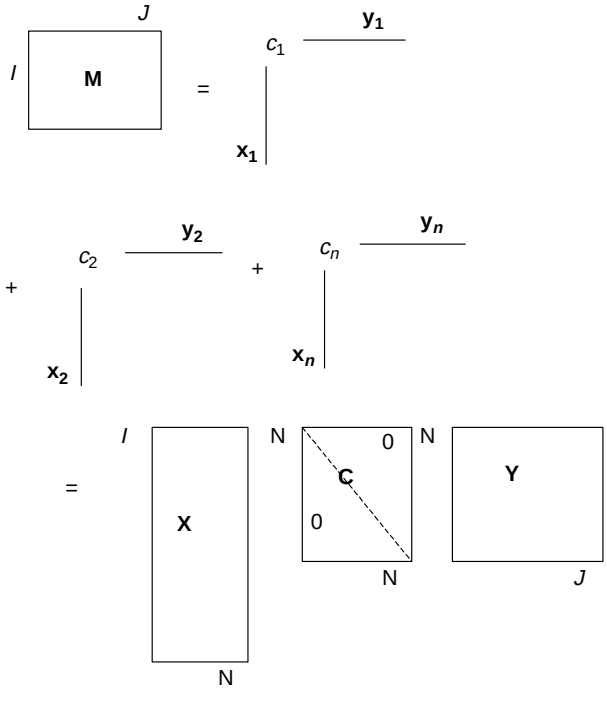

Figure 2 Decomposition of a bilinear data matrix: one dyad [$(\mathbf{x}, \mathbf{y})$ pair] exists for each component/chemical species.

The reason for this difference lies in the way the data are constructed; for bilinear data, one order modulates the other, whereas nonbilinear data do not have this characteristic. In the case of MS/MS data, the mass spectra in the second-order change independently over the mass to charge (m/z) range of the first-order. Consequently, there is no correlation between the mass spectrum at m/z i (in the first-order) and the mass spectrum at m/z $i+1$, so that a pure component cannot be represented by the outer product of two mass spectra.

The different characteristics of bilinear and nonbilinear data call for different calibration methods. There have been some relatively successful attempts to develop calibration methods for nonbilinear data, such as the nonbilinear rank annihilation method (NBRA) which is based on an eigenvalue decomposition.[6] However, analysis methods are far more common for bilinear data, for a number of reasons. First, bilinear data are more commonly produced by instruments in analytical chemistry, particularly through the combination of separation and spectrometric techniques such as LC/UV. Second, the bilinear model is more amenable to mathematical solutions using linear models.

Bilinear data analysis methods can calibrate for analytes with as few as two response matrices, where one is acquired from a standard sample with known analyte concentrations. Of course, owing to the second-order advantage that has been previously mentioned, either the standard or the sample may contain unknown interferents without adversely affecting the calibration. The particular method that operates on two matrices is named the GRAM, and is based on an eigenvalue–eigenvector decomposition.[4]

In cases where multiple standards and/or multiple test samples of second-order bilinear data are available, blocks of data, or data cubes, can be created. This is commonly referred to as three-way data, and the calibration methods that are utilized may be termed either second-order methods or three-way methods. These techniques decompose the blocks into three ''ways'', where one way corresponds to the sample concentration information. This allows calibration when a standard is available as part of the data set. One such method is an iterative procedure known as PARAFAC, which was developed in the discipline of psychometrics in the early 1970s for decomposition of three-way data.[7] This method uses alternating least squares, and is suitable for data that have low signal-to-noise ratios (S/N). Since its development, other methods have been reported that also utilize ALS algorithms, and they have a similar advantage with low S/Ns.[8,9] A second group of methods, which includes DTD, uses the generalized eigenvalue problem to decompose three-way data.[5,10] These methods are

non-iterative, and are especially effective when the data have a high S/N.

In some cases, the accuracy of the calibration may be threatened when unknown components in the sample interfere with the analyte response so that the analyte signal actually depends on the interferent concentration. In such situations, many of the methods mentioned above will be unable to predict the analyte concentrations reliably; this requires the use of a method for standard additions. Such a method exists that is analogous to the standard addition method commonly used with zero-order data: it is termed the second-order standard addition method (SOSAM),[11] and has been shown to increase the reliability of the calibration when matrix effects exist in the sample.

Another tool that can improve calibration accuracy is data pretreatment. In some cases, where the raw data do not exactly fit the model upon which a particular method is based, a preprocessing technique may be of some assistance in improving the fit. For example, in cases where instrumental variations cause minute changes in analyte response from sample to sample, methods that rely on identical responses may fail. This is especially true in cases involving time series data, as the response in the time domain may not be exactly the same from run to run. One approach to solving this problem is to standardize the responses so that the instrumental variations are reduced. Second-order standardization methods are available that perform such an operation.[12,13]

Other work that has been carried out in the second-order area and is related to calibration involves the figures of merit.[14] These include net analyte signal, selectivity, sensitivity, limit of detection and S/N, and they can be used to aid in the development of better second-order instruments which will ultimately result in improved calibration due to improved data.

2 THREE-WAY MODELS

Calibration methods are based upon different models that describe the structure of the data. In calibration, the most commonly used model for second-order (three-way) data is the trilinear model.[15] However, the trilinear model is actually one specific case of a broader model, the Tucker3 model. Tucker proposed his model for use in multiway psychology, in cases where the data contained three modes.[16] His example was of individuals × traits × raters, and the aim of the analysis was to determine the underlying trends in each mode. The Tucker3 decomposition can be written as Equation (4):

$$r_{ijk} = \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} x_{ia}\, y_{jb}\, z_{kc}\, g_{abc} + e_{ijk} \tag{4}$$

where r_{ijk} is a single data point in the three way data block under analysis, x_{ia} is an element of the loading matrix from the first mode, y_{jb} is an element of the loading matrix from the second mode and z_{kc} is an element of the loading matrix from the third mode; g_{abc} is an element of Tucker's core matrix, and represents the magnitude of any interactions between the modes, while e_{ijk} is an element of a matrix of residuals. This equation is written in matrix form as Equation (5):

$$\underline{\mathbf{R}} = \mathbf{X}\underline{\mathbf{G}}\mathbf{Y}\mathbf{Z} \tag{5}$$

Another form (Equation 6) was suggested by Geladi to eliminate confusion in the placement of the matrices:[17]

$$\mathbf{Y}$$

$$\underline{\mathbf{R}} = \mathbf{X}\underline{\mathbf{G}}\mathbf{Y} + \mathbf{E} \tag{6}$$

The general decomposition described by these equations is visually represented in Figure 3. It can be seen that this decomposition allows a different number of factors to be present in each mode (a, b and c), a fact that can complicate analysis. The core models interactions between factors by permitting nonzero elements on the off-diagonal. A unique solution is not obtained by this decomposition, however, as rotations can provide different solutions.

When the core is constrained to have nonzero elements only on the superdiagonal, a specific case of the Tucker3 model results. This model, the trilinear model, is much more restricted and can provide unique solutions. The calibration methods described later in this paper (e.g. DTD and PARAFAC) will generally use this trilinear model as in Equation (7):

$$r_{ijk} = \sum_{n=1}^{N} x_{in}\, y_{jn}\, z_{kn} + e_{ijk} \tag{7}$$

where N is the number of underlying factors in the data, x_{in} is an element of the loading matrix from the first

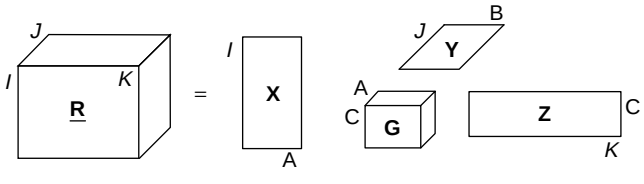

Figure 3 The Tucker3 decomposition: a core matrix, $\underline{\mathbf{G}}$, models the interactions between different factors.

mode, y_{jn} is an element of the loading matrix from the second mode and z_{kn} is an element of the loading matrix from the third mode. This is also shown in matrix form in Equation (8a):

$$\underline{\mathbf{R}} = \mathbf{XYZ} + \underline{\mathbf{E}} \tag{8a}$$

or Equation (8b):

$$\underline{\mathbf{R}} = \mathbf{X}\underline{\mathbf{I}}\mathbf{YZ} + \underline{\mathbf{E}} \tag{8b}$$

The visual representation of this model is shown in Figure 4. Because the core matrix now just contains ''ones'' on the superdiagonal, and is, in effect an identity matrix, this is a trilinear model that does not allow interactions between modes. Another fact that simplifies analysis is that the number of factors in each mode must be the same: this requires a single rank estimation rather than three, which was the case with the Tucker3 model. In addition, the decomposition does not require the loading vectors **x**, **y** and **z** to be orthogonal, unlike the Tucker3 model, allowing the solutions to approximate more closely the true solutions found in analytical data which may not necessarily be orthogonal. The trilinear model is considered to be the most restricted model of all three-way models, and is the only one that allows a unique solution, provided the following condition is met: if $\underline{\mathbf{R}}$ is decomposed into factor matrices **X**, **Y** and **Z**, and every I column of **X** is linearly independent, every J column of **Y** is linearly independent and every K column of **Z** is also independent, then $I + J + K \leq 2M + 2$, where M is the number of columns in each of A, B and C.

There are other special cases of the Tucker model, one of which is the basis of unfold principal component analysis (Unfold-PCA) and unfold partial least squares (Unfold-PLS) techniques. In this special case, the core matrix is not calculated, but is multiplied with two of the loading matrices. Only one of the loading matrices

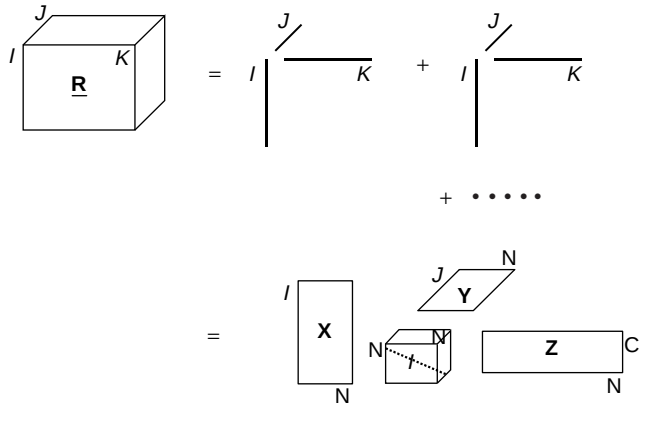

Figure 4 The trilinear model: **X**, **Y** and **Z** are the loading matrices for each mode.

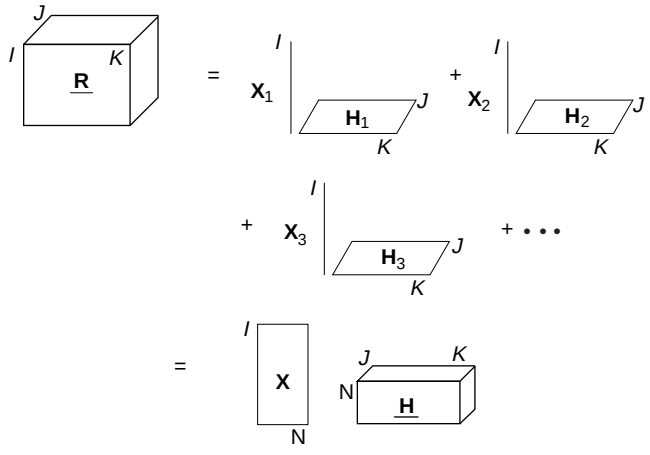

Figure 5 Unfold-PCA model: only one of the loading matrices (**X**) is explicitly calculated.

is explicitly calculated.[18] This is performed by unfolding the data block along one of its ways, and placing each of the resulting ''slices'' side-by-side to form a new matrix. There are three possible ways to unfold the cube, but usually if two of the modes are linked or associated in some way, they are kept together. The possible models that underlie this procedure are specified in Equation (9a–c):

$$r_{ijk} = \sum_{n=1}^{N} x_{in}\, h_{jkn} + e_{ijk} \tag{9a}$$

$$r_{ijk} = \sum_{n=1}^{N} y_{jn}\, h_{ikn} + e_{ijk} \tag{9b}$$

$$r_{ijk} = \sum_{n=1}^{N} z_{kn}\, h_{ijn} + e_{ijk} \tag{9c}$$

where x_{in} is an element of the (I × N) score matrix **X** and h_{jkn} is an element of the loading matrix $\mathbf{H}_n$.

The loading matrices $\mathbf{H}_1$ to $\mathbf{H}_n$ can be stacked and the three-way loadings matrix $\underline{\mathbf{H}}$ is obtained. Figure 5 shows one form of this unfolded model, corresponding to Equation (9a). This model does not produce a unique solution, as rotations will produce different solutions. In addition, there will more than likely be different residual matrices depending on which way the data block was unfolded.

3 CALIBRATION METHODS

There are many methods that have been published on solving three-way decompositions, which then allow quantitative analysis. Each model requires different techniques, and choosing the correct calibration method

therefore requires that the user know the type of data structure present. Booksh et al. discussed ways of choosing the calibration method, based on the theoretical instrument response function.[19] If this is known, it may be compared with the models described below, and then the appropriate method may be used.

3.1 Alternating Least-squares Techniques

ALS methods have been used for second-order analysis for some years. PARAFAC is one such method that solves the trilinear model that is shown in Equation (4). This algorithm was first introduced in the psychometric literature in 1970.[7] At the same time, an equivalent algorithm was introduced by a separate research group, who named their approach canonical decomposition (CANDECOMP).[20] The approaches are equivalent, and use an ALS algorithm in their solution. The first application of this procedure in the chemical literature was by Appeloff and Davidson in 1980.[21]

The first step in PARAFAC is to obtain initial estimates for two of its ways, for example, **X** and **Y** (Step 0 below). These guesses are then used to estimate the third way, $\widehat{\mathbf{Z}}$. In the next step, $\widehat{\mathbf{Z}}$ and **X** are used to estimate $\widehat{\mathbf{Y}}$, and then $\widehat{\mathbf{Z}}$ and $\widehat{\mathbf{Y}}$ are used to estimate $\widehat{\mathbf{X}}$. The process continues in an iterative fashion until a specified ending criterion is met. The algorithm is given in Table 1.

3.1.1 Step 0: Choosing the Starting Estimates

Choosing a starting estimate for two of the ways is the initialization step in the algorithm. A number of approaches are possible, but it must be noted that this choice will greatly affect how the algorithm proceeds, since the PARAFAC algorithm can become trapped in local minima so that convergence is slow and erroneous solutions can be produced.[22] If there is a priori knowledge of the analyte profiles, these may be used. A random number generator is also a frequent choice.[23] In that case in particular, a wise decision is to consider many starting values, although this will considerably delay the process. If a number of different starting values give the same solution, then confidence in the accuracy of the result is greater. The closer the starting guesses are to the true solution, the faster the algorithm will proceed, and the probability of encountering local minima is reduced. Speed is a big consideration when using PARAFAC analyses, as convergence is usually very slow. Results obtained from DTD provide another option for starting estimates: when these results are not complex, this choice can speed up the convergence process.[24]

Table 1 The PARAFAC algorithm

0	Initialize X and Y
1	$\widehat{\mathbf{Z}} = \mathbf{R}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}$
2	$\widehat{\mathbf{Y}} = \mathbf{R}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}$
3	$\widehat{\mathbf{X}} = \mathbf{R}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}$
4	Repeat steps 1–3 until convergence criterion is satisfied

3.1.2 Step 1: Estimating the First Set of Profiles

In each of steps 1–3, the **R** and **H** matrices are working matrices, redefined at each step. **R** is the data block, $\underline{\mathbf{R}}$, rearranged in matrix form, and **H** is created from the parameters not being estimated in that step. When calculating the **Z** profiles, therefore, the current **X** and **Y** information is employed to define **H**, as in Equation (10):

$$\mathbf{H}^T\mathbf{H} = (\mathbf{X}^T\mathbf{X}) \otimes (\mathbf{Y}^T\mathbf{Y}) \tag{10}$$

and Equation (11):

$$\mathbf{R}^T\mathbf{H} = \begin{matrix} diag(\mathbf{X}^T\mathbf{R}_1\mathbf{Y})^T \\ diag(\mathbf{X}^T\mathbf{R}_2\mathbf{Y})^T \\ \vdots \\ diag(\mathbf{X}^T\mathbf{R}_k\mathbf{Y})^T \end{matrix} \tag{11}$$

where $\otimes$ represents the Kronecker operator (element-wise multiplication), **R** is the data block unfolded in the K direction, so that $\mathbf{R}_1$ is the $I \times J$ submatrix of $\underline{\mathbf{R}}$, and $diag(\mathbf{X}^T\mathbf{R}_1\mathbf{Y})^T$ is a column vector containing the diagonal of $\mathbf{X}^T\mathbf{R}_1\mathbf{Y}$. At this step, Equation (12) is being minimized as

$$\min_z \ \left\| \mathbf{R}^T - \mathbf{H}\mathbf{Z}^T \right\|_2^2 \tag{12}$$

3.1.3 Steps 2 and 3: Estimating the Second and Third Sets of Profiles

The procedures for estimating **X** and **Y** are similar to that for determining **Z**, but in each case the **H** and **R** matrices are newly defined, using the parameters not being estimated.

3.1.4 Step 4: Convergence

There are two main approaches that are used to determine when the solution has been reached and the iterations should be stopped. One of these is to compare changes in the fit after each iteration, stopping when the change in the residuals is less than a specified tolerance, generally close to 10^{-6}. The residual change is calculated using the root of squared residuals. An alternative is to compare the loadings obtained after each iteration, stopping when they do not change more than a specified amount. This comparison is performed by calculating the angle between the predicted profile from successive iterations, which is

found in each case by unfolding the **X**, **Y** and **Z** matrices into column vectors and then calculating (Equation 13):

$$\cos\theta_x \cos\theta_y \cos\theta_z = \frac{\mathbf{x}_i\mathbf{x}_{i+1}}{\sqrt{\mathbf{x}_i^2\mathbf{x}_{i+1}^2}} \quad \frac{\mathbf{y}_i\mathbf{y}_{i+1}}{\sqrt{\mathbf{y}_i^2\mathbf{y}_{i+1}^2}} \times \frac{\mathbf{z}_i\mathbf{z}_{i+1}}{\sqrt{\mathbf{z}_i^2\mathbf{z}_{i+1}^2}} \tag{13}$$

where $\mathbf{x}_i$ is from the ith iteration, and $\mathbf{x}_{i+1}$ is from the following iteration. When this term is less than a specified criterion, for example 10^{-6} once again, convergence has been reached.

A difficulty that exists with PARAFAC is the occasional occurrence of two-factor degeneracies.[25] This is signified when the estimated profiles for two components are highly correlated in all three modes. The uncorrelated correction coefficient (UCC) has been introduced to detect such a situation. The UCC for two components, A and B, is (Equation 14):

$$\mathrm{UCC} = \frac{\mathbf{x}_\mathrm{A}\mathbf{x}_\mathrm{B}}{\sqrt{\mathbf{x}_\mathrm{A}^2\mathbf{x}_\mathrm{B}^2}} \quad \frac{\mathbf{y}_\mathrm{A}\mathbf{y}_\mathrm{B}}{\sqrt{\mathbf{y}_\mathrm{A}^2\mathbf{y}_\mathrm{B}^2}} \quad \frac{\mathbf{z}_\mathrm{A}\mathbf{z}_\mathrm{B}}{\sqrt{\mathbf{z}_\mathrm{A}^2\mathbf{z}_\mathrm{B}^2}} \tag{14}$$

where $\mathbf{x}_\mathrm{A}, \mathbf{x}_\mathrm{B}$ are the **x** profiles, $\mathbf{y}_\mathrm{A}, \mathbf{y}_\mathrm{B}$ are the **y** profiles and $\mathbf{z}_\mathrm{A}$, $\mathbf{z}_\mathrm{B}$ are the **z** profiles of the two components. When the UCC is close to -1, a two-factor degeneracy has occurred. In that case, the profiles will be either positively correlated in two or more of the modes or negatively correlated in one or all three of the modes. Effectively, the results suggest that there is one factor fewer present than is actually the case. Sometimes it happens that a solution is close to degeneracy; this is indicated when the UCC is close to -0.85. This is not considered to be a true degeneracy, and further iterations may result in an improvement in the UCC as the solution moves closer to the real one. The temporary degeneracy can then be considered to be due to a local minimum, most likely due to a poor choice of initial estimates, and is known as a ''swamp''. A true degeneracy may arise if the algorithm has been asked to calculate more factors than exist in the data. It may also occur if the wrong model has been chosen, i.e. if the trilinear model does not fit the data.

The existence of swamps is a source of difficulty, as extremely slow convergences may result.[22] These are areas where the convergence slows dramatically, making it appear as though the solution has been reached. However, if the iterations are allowed to proceed, the slow convergence will once again become faster and the true solution will emerge. While the algorithm is in a swamp, the residuals will not change dramatically, whereas the loadings will. If a swamp is mistakenly thought to indicate convergence, therefore, the solution provided will not be accurate. Mitchell and Burdick[24] have proposed that swamps tend to occur simultaneously with two-factor degeneracies, and caution must be exercised when these play a role in convergence. A solution that uses a stabilization method has been proposed, so that the number of iterations spent in a swamp is greatly reduced.[25] It is suggested that a high positive UCC may be indicative of serious problems due to instabilities in the estimation steps.

Constraints are generally imposed on the parameters to be estimated in order to increase interpretability. As was mentioned previously, PARAFAC was used in the psychometric field for some years before its applicability to chemical data was recognized. Psychometricians have employed a number of constraints, including orthogonal constraints, to overcome problems with unstable solutions to help convergence to the correct solution. (An orthogonal constraint helps with scaling: although scaling does not affect the uniqueness of the solution, it can be an issue in determining how long it takes to converge.) In chemistry, constraints are chosen based on a priori knowledge of the data being analyzed. For example, knowing that mass spectra, chromatographic profiles and concentrations are always positive, non-negativity constraints can be applied to all modes of a set of LC/MS samples. In addition, it is known that chromatographic profiles should be unimodal, allowing the imposition of a unimodality constraint. While non-negativity constrained least-squares techniques are common, however, the unimodality constraint is more difficult to impose.

The results that are obtained when PARAFAC is applied to excitation–emission data demonstrate the improvement that can occur when constraints are imposed. Samples that contained tyrosine, tryptophan and phenylalanine were measured by fluorescence, and unconstrained PARAFAC was used in an attempt to resolve the excitation and emission spectra.[15] Figure 6(a) shows the emission spectra that were obtained. It is obvious that the profiles are not correct, as the spectra should be non-negative; however, when the concentrations and spectra are constrained to be positive, PARAFAC produces emission profiles (Figure 6b) that are very close to the true spectra (Figure 6c).

The only deviation from the true spectra is a slight hump below 300 nm, and this is reported to be due to non-multilinear Rayleigh scatter. Using the knowledge that the spectra should be positive enabled the analysts to use PARAFAC in such a way (i.e. with a non-negativity constraint) that the difficulties encountered could be overcome. These problems were most likely due to the small number of samples used, which in this case was just two. A related area of research in PARAFAC analysis is in increasing the algorithm's efficiency by

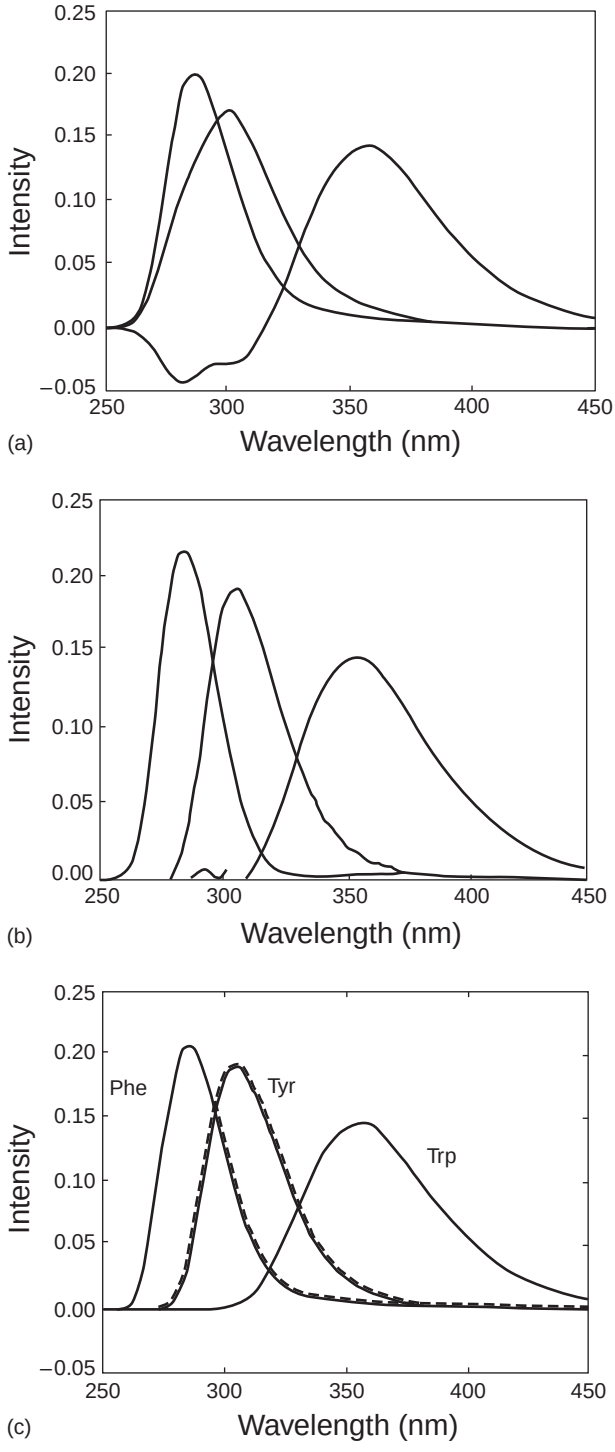

Figure 6 Emission spectra from (a) unconstrained PARAFAC, (b) PARAFAC with non-negativity constraints and (c) true emission profiles. (Reprinted with permission from R. Bro, *Chemom. Intell. Lab. Syst.*, **38**, 149–171 (1997). Copyright 1997 Elsevier Science.)

reducing the computations required;[26–29] constrained algorithm's tend to reach convergence more quickly than the traditional versions, thereby resulting in research that combines the two.[30]

PARAFAC has been applied to various analytical problems, and has resulted in the successful calibration of naphthalene and styrene in ocean water.[31] In that case, an excitation–emission matrix imaging spectrofluorimeter was used, and the data contained a complex background signal in addition to overlapping spectral interferents that were not included in the calibration set. It was found that a two-factor PARAFAC model could be used to decompose the data so as to produce estimates of the excitation and emission profiles of naphthalene and background intensities in the samples. The factor that corresponded to naphthalene had a 2.6 ppb standard deviation of the blank, which equaled a limit of determination of 8 ppb, so that the measured linear dynamic range for naphthalene extended from 8 to 2000 ppb. When gasoline/naphthalene/ocean water samples were included in the decomposition, one extra factor was needed to account for the fluorescence spectrum of the gasoline; the question to be answered was whether or not the quantitation would be affected by the inclusion of these samples. Calibration curves were generated for both the two- and three-factor decompositions, where standard additions of naphthalene had been made to generate the standards. It was found that the quality of the calibration curve formed by the naphthalene/ocean water standards was not degraded by the presence of the gasoline as long as

Table 2 Prediction errors obtained from PARAFAC quantitation of naphthalene: (i) two-factor model (calibration set contains naphthalene/ocean water standards only) and (ii) three-factor model (calibration set includes naphthalene/gasoline/ocean water). (Reprinted with permission from K.S. Booksh, A.R. Muroski, M.L. Myrick *Anal. Chem.*, **68**, 3539–3544 (1996). Copyright 1996 American Chemical Society.)

Naphthalene added (ppb)	Prediction error	
	(i) 2-factor	(ii) 3-factor
28	−9.5(4.0)[a]	−10.8(2.6)[a]
28	−13.3(−0.1)	−14.3(−1.0)
56	−6.7(7.4)	−4.1(10.0)
84	−11.5(3.1)	−12.8(1.8)
110	−18.3(−3.2)	−17.5(−2.5)
286	−14.0(4.7)	−16.0(2.7)
571	−30.2(−5.9)	−31.9(−7.6)
666	78.3	77.7
800	113.0	112.2
1000	40.8	42.7
1333	−90.3(−51.5)	−91.0(−52.0)
2000	−18.9(35.1)	−18.8(36.2)
RMSE[b]	47.2(19.2)	47.3(19.7)

[a] Errors derived after discarding the 666, 800 and 1000 ppb naphthalene/ocean water standards.
[b] Root mean square error, including blanks.

Table 3 Comparison of PARAFAC and ATLD quantitation of overlapped chlorinated aromatic hydrocarbons. (Reprinted with permission from H.-L. Wu, M. Shibukawa, K. Oguma, *J. Chemom.*, **12**, 1–26 (1998). Copyright 1998 John Wiley & Sons Limited.)

Unknown sample no.	*p*-Chlorotoluene (μg ml^{-1})					*o*-Chlorotoluene (μg ml^{-1})				
	Known	PARAFAC		ATLD		Known	PARAFAC		ATLD	
		Predicted	Recovery (%)	Predicted	Recovery (%)		Predicted	Recovery (%)	Predicted	Recovery (%)
7	25.2	24.2 ± 3.9	95.8 ± 15.5	25.2 ± 0.2	100.0 ± 0.7	91.2	88.1 ± 18.2	108.7 ± 19.9	90.8 ± 0.9	99.6 ± 1.0
8	50.4	54.9 ± 2.8	109.3 ± 5.5	53.7 ± 0.5	106.6 ± 1.0	30.4	30.7 ± 12.0	101.1 ± 39.5	29.5 ± 0.8	97.2 ± 2.7
9	75.6	75.3 ± 6.1	99.6 ± 8.1	78.2 ± 0.1	103.4 ± 0.1	60.8	67.7 ± 9.9	111.3 ± 16.3	64.7 ± 0.9	106.4 ± 1.4

an extra factor was included to account for its variation. Table 2 displays the prediction errors obtained in each calibration, and it is apparent that they are very similar, each having a root mean square error of prediction (RMSEP) of 47 ppb. However, three of the naphthalene standards deviate from linearity on the calibration curve: when these standards are excluded from the calibration, the RMSEP drops to <20 ppb in each case. The important conclusion that can be drawn from this analysis is that when second-order calibration techniques are employed, accurate analyte quantitation is possible even in the presence of large background signals and interfering chemical species.

An accurate calibration of a model system using flow-injection analysis/diode-array data has also been reported.[32] Yet another investigation used PARAFAC to predict enzymatic activity and substrate consumption during the enzymatic browning of vegetables,[33] while kinetic studies have also been investigated using this technique.[34]

A recently published algorithm, the alternating trilinear decomposition (ATLD) algorithm, also solves the trilinear model with the aid of ALS, and is in fact an improvement of the traditional PARAFAC algorithm without constraints.[8] While it does not involve the use of constraints, it still converges faster than the original PARAFAC algorithm (ATLD is reported as converging in less than ten iterations), and its iterative procedure uses the Moore–Penrose generalized inverse with singular value decomposition (SVD) in the trilinear sense. As mentioned previously, some of the main drawbacks to the use of PARAFAC are (a) the occurrence of two-factor degeneracies and (b) the presence of swamps. ATLD was developed with the aim of regularizing the procedure, thus avoiding these traps. The loss function to be minimized in this case is (Equation 15):

$$\sigma = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(r_{ijk} - \sum_{n=1}^{N} x_{in}\, y_{jn}\, z_{kn}\right)^2 \qquad (15)$$

where the symbols have the same meanings as before. In order to solve the loss function, the algorithm alternates over **Y** for fixed **X** and **Z**, over **X** for fixed **Y** and **Z**, and over **Z** for fixed **X** and **Y**. The equations that are solved at each step are (Equations 16a–c)

$$\mathbf{x}_i^T = diag\left(\mathbf{Y}^+\mathbf{R}_{i..}(\mathbf{Z}^T)^+\right), \quad i = 1, \ldots, I \qquad (16a)$$

$$\mathbf{y}_j^T = diag\left(\mathbf{Z}^+\mathbf{R}_{.j.}(\mathbf{X}^T)^+\right), \quad j = 1, \ldots, J \qquad (16b)$$

$$\mathbf{z}_k^T = diag\left(\mathbf{X}^+\mathbf{R}_{..k}(\mathbf{Y}^T)^+\right), \quad k = 1, \ldots, K \qquad (16c)$$

where $\mathbf{R}_{i..}$ is the data cube $\underline{\mathbf{R}}$ unfolded along the $J \times K$ axis. These equations involve the use of the Moore–Penrose generalized inverse, denoted by the + superscript, whereas the original PARAFAC algorithm did not (see Equation 11). Because this computation is based on the SVD, singular values less than a certain tolerance are treated as zero. This makes it possible to perform the calculations even when N is greater than the number of chemical species present in the sample, reducing the rank deficiency problem that exists with PARAFAC and that can cause two-factor degeneracies. The importance of this result stems from the fact that rank determination is not always straightforward, especially in the presence of noise, and an overestimation of the number of chemical species in the sample may sometimes occur.

A comparison of the ATLD and traditional PARAFAC algorithm has been carried out on liquid chromatography/diode-array detection (LC/DAD) data for overlapped chlorinated aromatic hydrocarbons. The analytes were *p*-chlorotoluene and *o*-chlorotoluene, with *o*-dichlorobenzene being an interfering species (i.e. not included in the calibration set). Four samples were included in the calibration set, while there were three test samples. Table 3 shows the true concentrations of the analytes in each sample, along with the predicted concentrations that were given by both the traditional PARAFAC algorithm and the ATLD algorithm.

The concentration estimates are, in general, closer to the true values than the PARAFAC predictions; moreover, convergence was reached much more quickly with the ATLD algorithm. It must also be noted that the standard deviations of the ATLD predictions are much smaller than those of the PARAFAC predictions. However, it is known that PARAFAC performance can

be improved through the use of constraints, and such an algorithm may have performed better.

Another method that uses an ALS algorithm is MCR. As the name suggests, this technique was originally developed for resolution of the profiles in second-order data.[9,35,36] However, further development of the method has allowed the calibration of three-way data in addition to profile resolution.[37]

There are three main steps that are carried out during this procedure. First, the chemical rank of the data, N, must be determined. Generally, SVD may be used. By examining the singular values in regions where no chemical components exist, variance due to noise can be estimated. Once the noise estimate is known, singular values due to the chemical components can be determined in different regions of the data, thereby allowing pseudo-rank estimation.

As with other techniques that employ ALS, initial estimates of the profiles are required. In this case, however, only a concentration profile estimate is needed. Evolving factor analysis (EFA) is used,[38] and for a specific reason: it is known that, in order to use curve-resolution methods, assumptions must be made about the signal. Common assumptions include bilinearity, non-negativity, unimodality and closure. However, even with such constraints, a unique solution (for the analysis of a single sample) is not guaranteed, as rotational and intensity ambiguities may still exist.[9] The use of EFA allows curve resolution to overcome, in part, the rotational ambiguity, as it provides information regarding the windows of existence of components in the data. Of course, when multiple samples are analyzed simultaneously, the rotational ambiguity no longer exists, but EFA remains a good choice for initial estimates. The third step to be taken when performing MCR is to calculate iteratively new profiles using constrained ALS. When a single sample is being analyzed (Equation 17):

$$\mathbf{D} = \mathbf{CS} \tag{17}$$

where $\mathbf{D}$ is a bilinear data matrix, dimensioned $m \times n$, $\mathbf{C}$ is a matrix of concentration profiles, dimensioned $m \times N$, and $\mathbf{S}$ is a matrix of spectral profiles, dimensioned $N \times n$. It can be seen that this is not the trilinear model used by PARAFAC and the eigenvalue–eigenvector-based methods, as the concentration information is contained within the concentration profiles rather than being separately calculated.

The initial estimates of the concentration profiles in $\mathbf{C}$, obtained by EFA, are then used to calculate the corresponding spectra (Equation 18):

$$\widehat{\mathbf{S}} = \mathbf{C}^{+}\overline{\mathbf{D}} \tag{18}$$

where $\overline{\mathbf{D}}$ has been truncated to rank using SVD. The following step involves a new estimation of $\mathbf{C}$, using the new spectral profiles, $\widehat{\mathbf{S}}$ (Equation 19):

$$\widehat{\mathbf{C}} = \overline{\mathbf{D}}\,\widehat{\mathbf{S}}^{+} \tag{19}$$

The algorithm iterates back and forth between Equations (18) and (19) until the stopping criterion is met. This criterion is similar to PARAFAC ending criteria.

There are a number of constraints that may be used, based on the previously mentioned assumptions. Constraining the spectra and concentration profiles to have non-negative values is common, while the shape of the signal may also be a constraint. For example, a unimodality constraint is frequently imposed upon chromatographic profiles. As mentioned already, however, the use of such constraints will not always remove the ambiguity problems. Analyzing two or more samples simultaneously, on the other hand, often allows a unique solution. In addition, quantitation in the presence of unknown analytes becomes possible: the second-order advantage.[39,40]

In order to use MCR for calibration, the data matrices to be analyzed must be arranged in one matrix. They may be concatenated in three ways: row-wise, column-wise or along the tubes. Samples are concatenated so that the columns, or spectra, are in common (Equation 20a):

$$\mathbf{D}_s = \mathbf{C}_s\mathbf{S} \tag{20a}$$

or (Equation 20b)

$$\begin{matrix}\mathbf{D}_1\\ \mathbf{D}_2\\ \vdots\\ \mathbf{D}_s\end{matrix} = \begin{matrix}\mathbf{C}_1\\ \mathbf{C}_2\\ \vdots\\ \mathbf{C}_s\end{matrix} \mathrm{S} \tag{20b}$$

$\mathbf{D}_s$ is dimensioned $m^*s \times n$, $\mathbf{C}_s$ is $m^*s \times N$ and $\mathbf{S}$ is $N \times s$, where s is the number of samples, N is the number of principal components required to describe the data and $m \times n$ are the dimensions of each individual $\mathbf{D}$ matrix. It can be seen that $\mathbf{S}$, which contains the spectral profiles, is common to each individual sample, while the concentration profiles vary.

The quantitative results are obtained through the resolved concentration profiles in $\mathbf{C}$. Since the concentration information is embedded in the profiles, the area under the curves can be related to concentration. When calibration standards are included in the MCR analysis, the unknown analyte concentrations can be found via Equation (21):

$$c_{\mathrm{un}} = \frac{A_{\mathrm{un}}}{A_{\mathrm{std}}}c_{\mathrm{std}} \tag{21}$$

where c_{un} and c_{std} are the unknown and standard concentrations for a particular analyte, and A_{un}, A_{std}

are the areas under the resolved profiles for that analyte in the unknown and standard samples.

With MCR analysis, it is possible to analyze data matrices that do not strictly follow a trilinear data structure when concatenated. Although each individual matrix must be bilinear, the profiles of one mode may vary. For example, if the data are chromatographic/spectrometric, the chromatographic profiles may vary in retention time and/or shape from one sample to the next. Such data do not have a trilinear data structure, and methods that rely on a trilinear model will fail. By constraining the spectra to be equal, but not the concentration profiles, MCR can resolve the profiles and also obtain adequate quantitative results.[37]

For example, quantitative analysis using second-order MCR has been applied to mixtures of amino acids that were measured with flow-injection analysis and diode-array multiwavelength detection.[39] The system monitored the reaction of lysine and proline with 1,2-naphthoquinone-4-sulfonate (NQS). Four samples were analyzed, water, lysine, proline and a mixture of lysine and proline, all of which were injected into a system containing NQS. This data is not trilinear; whereas the concentration profiles of the amino acid derivatives retain their shapes over the course of the reaction, the shapes of unreacted NQS species vary depending on the amino acid concentration in the mixture. The three-dimensional data from the mixture of lysine and proline are shown in Figure 7.

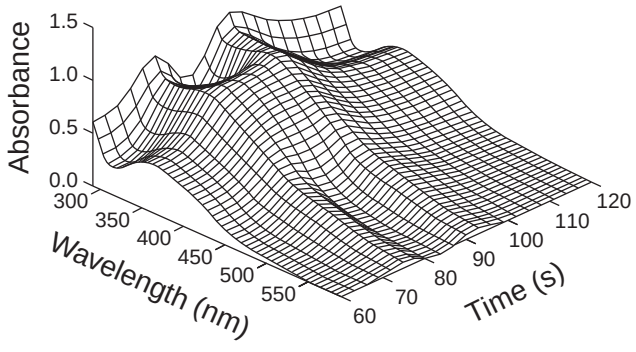

Figure 7 Flow-injection analysis/diode-array detector response for a mixture of lysine and proline. (Reprinted with permission from J. Saurina, S. Hernandez-Cassou, R. Tauler, *Anal. Chim. Acta*, **335**, 41–49 (1996). Copyright 1996 Elsevier Science.)

The sample data matrices were concatenated and analyzed as described above, and Table 4 shows the predicted concentrations that were obtained under a number of different constraints (options a–d). The fitting error of the resolved profiles is also reported.

Because of the nontrilinear nature of the data, option c (which forces both concentration and spectral profiles to be identical and therefore assumes trilinearity) does not give the best quantitative results, producing a quantitation error of 3.34%. Conversely, option d constrains only the concentration profiles of the amino acid derivatives to be the same, and this more flexible option produces much better quantitative results with a quantitation error of 0.36%.

In contrast, when data with a trilinear structure are analyzed, the concentration profiles can be constrained to be identical in addition to the spectral profiles. In that case, the additional constraint improves both quantitative and qualitative results. It has been found that when data follow the trilinear data structure, the predictive error is

Table 4 Predicted concentrations for lysine and proline obtained from MCR under four different constraints. (Reprinted with permission from J. Saurina, S. Hernandez-Cassou, R. Tauler, *Anal. Chim. Acta*, **335**, 41–49 (1996). Copyright 1996 Elsevier Science.)

Option[a]	Recovered Lys concentration ($M \times 10^4$)	Recovered Pro concentration ($M \times 10^4$)	Fitting error(%)[b]	Quantitation error(%)[c]
a	2.04	1.89	3.75	4.21
b	1.97	1.93	4.77	2.76
c	1.93	1.94	6.08	3.34
d	2.01	2.00	5.36	0.36

[a] Applied constraints: (a) no closure or equal shape constraint; (b) closure constraint with respect to NQS; (c) closure constraint with respect to NQS and equal shape constraint to all concentration profiles; (d) closure constraint with respect to NQS and equal shape constraint to only the derivative concentration profiles.

[b] Fitting error (%) = $(A_{ij,\text{experimental}} - A_{ij,\text{reproduced}})^2{}^{1/2}$ ($A_{ij,\text{experimental}})^2{}^{1/2} \times 100$, where $A_{ij,\text{experimental}}$ is the experimental absorbance at wavelength i and time j and $A_{ij,\text{reproduced}}$ is the reproduced absorbance obtained by MCR.

[c] Quantitation error (%) = $(C_{\text{known}} - C_{\text{calc}})^2{}^{1/2}$ ($C_{\text{known}}{}^2)^{1/2} \times 100$, where C_{known} is the known concentration of each analyte, Lys and Pro, in the unknown sample and C_{calc} is the calculated concentration of each analyte recovered from MCR.

within the level of measurement noise.[41] In cases where only one order is identical from one sample to another, the error is higher, but still within acceptable limits, provided that there is some selectivity present for the analyte in question. If complete overlap exists, then the results will be ambiguous.[40] Comparison studies have been carried out between MCR and DTD, an eigenproblem-based method that operates specifically on trilinear data, and the results indicate that each method's performance is comparable for trilinear data.[41] It has also been suggested that a simple method for determining the presence or absence of trilinearity might be to compare the lack of fit obtained with DTD and MCR. This is because data that fit the trilinear model will be fitted well with DTD, whereas deviations from trilinearity will result in a better fit with MCR.[42]

3.2 Rank Annihilation Methods

Calibration methods that use rank annihilation have been used in the chemical field for almost 20 years, and they have undergone many changes and improvements in that time. The first form of rank annihilation factor analysis (RAFA) was introduced by Ho et al. in 1980.[43–45] That method suffered from being very slow, as it employed iterative methods for its solution. When an improved version of RAFA was proposed by Lorber 4 years later, the problem was transformed into an eigenvalue problem that could be directly solved in a very short time.[46] In this method, two samples were required for calibration. One was the unknown or test matrix that was to be analyzed, **M**, while the second was the calibration sample, **N**, that could contain just one of the analytes present in **M**. Both **M** and **N** must follow the bilinear data format, where the mixture matrix, **M**, can be represented by Equation (2), and the pure standard matrix, **N**, can be represented by Equation (1). The kth component in **M** is common to both samples.

These equations can be rearranged to form a generalized eigenvalue problem (Equation 22):

$$\mathbf{N}_k\mathbf{z}_k = \lambda_k\mathbf{M}\mathbf{z}_k \tag{22}$$

where λ_k is an eigenvalue that is equivalent to the concentration of the kth analyte in **N** divided by its concentration in **M**. $\mathbf{z}_k$ is an eigenvector that was later shown by Sanchez and Kowalski to be related to the response profile of the analyte.[4] However, because **M** and **N** are not necessarily square, the problem as written may not be solved. **M** is therefore truncated using SVD (Equation 23):

$$\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{23}$$

Equation (22) is rewritten as Equation (24):

$$\mathbf{x}_k\,\mathbf{y}_k^T\,\mathbf{z}_k = \lambda_k\,\overline{\mathbf{U}\mathbf{S}\mathbf{V}}^T\mathbf{z}_k \tag{24}$$

and an eigenvalue problem is formed (Equation 25):

$$\overline{\mathbf{U}}^T\mathbf{x}_k\,\mathbf{y}_k^T\,\overline{\mathbf{V}\mathbf{S}}^{-1}\mathbf{z}' = \lambda_k\,\mathbf{z}' \tag{25}$$

where $\overline{\mathbf{U}}$, $\overline{\mathbf{S}}$ and $\overline{\mathbf{V}}$ are the SVD factors truncated to the number of principal components in **M**, $\mathbf{z}' = \overline{\mathbf{S}\mathbf{V}}^T\mathbf{z}_k$, $\overline{\mathbf{U}}^T\mathbf{x}_k\,\mathbf{y}_k^T\,\overline{\mathbf{V}\mathbf{S}}^{-1}$ is square, and $\mathbf{z}'$ is the eigenvector to be found. The column and row profiles, $\mathbf{y}_k$ and $\mathbf{x}_k$, may be found (Equation 26):

$$\mathbf{y}_k = (\overline{\mathbf{V}\mathbf{S}}^{-1}\mathbf{z}')^{T+}; \qquad \mathbf{x}_k = \mathbf{N}\mathbf{z}_k\,c_{n,k} \tag{26}$$

This formulation of RAFA (minus the use of Equation 26), when introduced, decreased computation time, and it also overcame a problem that had existed with the original version. In using that particular method, it was necessary to decide which eigenvalues should be examined to find the minimum which indicates rank annihilation; the use of an eigenvalue problem eliminates that need.

The form of RAFA that is generally applied today is called generalized RAFA, also known as GRAM.[4,47] This technique allows the simultaneous quantitation of multiple analytes, whereas the earlier RAFA formulations permit the quantitation of just one analyte at a time. Again, one standard matrix, **N**, is required for the calibration of one sample matrix, **M**. However, in this case, the standard and the test samples may contain unknown interferents that will not affect the calibration; this is the second-order advantage previously mentioned.

The first step that must be taken in a GRAM analysis is to find joint basis sets for the analyte signals in **M** and **N**. The fact that interferents may be present in either **M** or **N** means that projection of one matrix on to the principal components of the other (as in Lorber's method, Equation 25) will change the information present and result in an inaccurate calibration. GRAM overcomes this by finding the principal components from both matrices. There have been a few suggestions on the best way to calculate the principal components: Sanchez and Kowalski proposed calculating them from the sum of **M** and **N**,[4] thereby producing (Equation 27):

$$\mathbf{W} = \mathbf{M} + \mathbf{N} = \mathbf{U}_W\mathbf{S}_W\mathbf{V}_W^T \tag{27}$$

where $\mathbf{U}_W$ and $\mathbf{V}_W$ are the principal components of the joint row and column spaces. Wilson et al. found the principal components as[47] (Equation 28):

$$\mathbf{M}|\mathbf{N} = \mathbf{Q}\mathbf{S}\mathbf{V}^T \tag{28a}$$

$$\frac{\mathbf{M}}{\mathbf{N}} = \mathbf{P}\mathbf{S}\mathbf{V}^T \tag{28b}$$

M|**N** is obtained by concatenating the matrices row-wise, whereas **M**/**N** is obtained through the column-wise

concatenation of **M** and **N**. **Q** and **P** are the joint row and column spaces found by this technique.

Following the principal component calculation, the eigenvalue problem must be formulated. There are a number of ways to calculate the eigenvalues and eigenvectors from which the concentration ratios and row and column profiles are determined. Either a standard eigenproblem or a generalized eigenproblem may be solved. The standard eigenproblem is similar to Lorber's formulation for a single component, as it involves projecting the test sample matrix onto the principal components of the calibration matrix, and is shown using the singular vectors from Equation (27) (Equation 29):

$$(\overline{\mathbf{U}}_W^T \mathbf{M} \overline{\mathbf{V}}_W \overline{\mathbf{S}}_W^{-1}) \mathbf{Z}'_W = \mathbf{Z}'_W \mathbf{\Lambda} \tag{29}$$

where **U**, **S** and **V** calculated in Equation (27) have been truncated to the number of principal components present in **W** and $\mathbf{Z}'_W = \overline{\mathbf{S}}_W \overline{\mathbf{V}}_W^T \mathbf{Z}$ and **Z** is a matrix of eigenvectors. In this case, **Λ** is a diagonal matrix that consists of the concentration ratios for all of the components common to **M** and **N**. It should be noted that the ratios that result from this formulation are (Equation 30):

$$\lambda_k = \frac{c_{m,k}}{c_{n,k} + c_{m,k}} \tag{30}$$

where $c_{n,k}$ and $c_{m,k}$ are the concentrations of analyte k in **N** and **M**, respectively. This means that if a component is not present in the calibration sample (**N**), $\lambda_k = 0$, while $\lambda_k = 1$ if a component is not present in the unknown sample (**M**). From this fact, the analyst can determine which components are present in either or both samples. The row and column profiles of each analyte common to both matrices may also be found (Equation 31):

$$\mathbf{Y}^T = (\overline{\mathbf{V}}_W \overline{\mathbf{S}}_W^{-1} \mathbf{Z}'_W)^+; \qquad \mathbf{X} = \overline{\mathbf{U}}_W \mathbf{Z}'_W \mathbf{C}_m^{-1} \tag{31}$$

An alternative formulation of GRAM uses the generalized eigenvalue problem.[47] It has a similar form to Equation (22), but in this case there are multiple analytes in both calibration matrices (Equation 32):

$$\mathbf{MZ} = \mathbf{NZ\Lambda} \tag{32}$$

where **Z** is a matrix of eigenvectors and **Λ** is a diagonal matrix of eigenvalues. If **M** and **N** are square, this equation can be solved directly using the QZ algorithm. However, **M** and **N** are generally rectangular matrices, and therefore a projection operation is needed to make them square. The principal components for the joint row and column spaces from either Equation (27) or (28) may be used, and this then leads to Equations (33a) and (33b):

$$\overline{\mathbf{M}}_{PQ} = \overline{\mathbf{P}}^T \mathbf{M} \overline{\mathbf{Q}} \tag{33a}$$

$$\overline{\mathbf{N}}_{PQ} = \overline{\mathbf{P}}^T \mathbf{N} \overline{\mathbf{Q}} \tag{33b}$$

where $\overline{\mathbf{M}}_{PQ}$ and $\overline{\mathbf{N}}_{PQ}$ are square and $\overline{\mathbf{P}}$ and $\overline{\mathbf{Q}}$ are truncated principal components from Equation (28). The eigenproblem can then be rewritten as Equation (34):

$$\overline{\mathbf{M}}_{PQ} \mathbf{Z}_Q = \overline{\mathbf{N}}_{PQ} \mathbf{Z}_Q \mathbf{\Lambda} \tag{34}$$

where $\mathbf{Z}_Q = (\mathbf{Y}_Q^T)^+$. **Λ** is a diagonal matrix that contains eigenvalues of the form $c_{k,m}/c_{k,n}$. Use of the QZ algorithm has the advantage that it yields two matrices α and β where $\Lambda = \alpha/\beta$, rather than calculating **Λ** directly. This means that there is no danger of running into difficulties when the calibration matrix, **N**, is missing an analyte that is present in **M**. With the original method (Equation 29), this scenario would have led to a calculation of 1 divided by 0. The analyte concentrations in the test matrix, **M**, may be calculated from this relation if the concentrations in **N** are known, and vice versa. As with the previous description of GRAM above, the row and column profiles can then be found (Equation 35):

$$\mathbf{Y} = \overline{\mathbf{Q}}(\mathbf{Z}^+)^T; \qquad \mathbf{X} = \overline{\mathbf{P}}(\mathbf{N}_{PQ} + \mathbf{M}_{PQ})\mathbf{Z} \tag{35}$$

The use of the generalized eigenvalue problem and QZ algorithm is preferable in situations where construction of the factor space has not been performed with maximum precision.[48]

There is another modification to the GRAM algorithm that in some instances can prevent the occurrence of complex solutions. It applies to the GRAM formulation that uses the QZ algorithm, and simply involves projecting the sum of **M** and **N** (i.e. **W**) rather than **M** alone when formulating the eigenproblem.[49] This makes the QZ formulation more similar to the original Sanchez–Kowalski method. In the case of the generalized eigenvalue problem, Equation (33a) becomes Equation (36):

$$\overline{\mathbf{W}}_{PQ} = \overline{\mathbf{P}}^T \mathbf{W} \overline{\mathbf{Q}} \tag{36}$$

and Equation (34) is rewritten as Equation (37):

$$\overline{\mathbf{W}}_{PQ} \mathbf{Z}_Q = \overline{\mathbf{N}}_{PQ} \mathbf{Z}_Q \mathbf{\Lambda} \tag{37}$$

such that **Λ** now contains eigenvalues as per Equation (30) rather than the $c_{k,m}/c_{k,n}$ ratios that are obtained from the eigenproblem in Equation (34). The advantage of this variant is that it is numerically more stable, because the eigenvectors that are calculated are common to both of the matrices under analysis, **W** and **N**. In cases where the analyte profiles of the common components in **M** and **N** vary, such as in situations where retention time shifts are present, the use of this version avoids the complex solutions that can result from dissimilarities in the matrices analyzed.

The utility of GRAM is demonstrated in an application that quantitates ethylbenzene and *m*-xylene, in a white gas mixture that was analyzed by GC × GC.[50] This

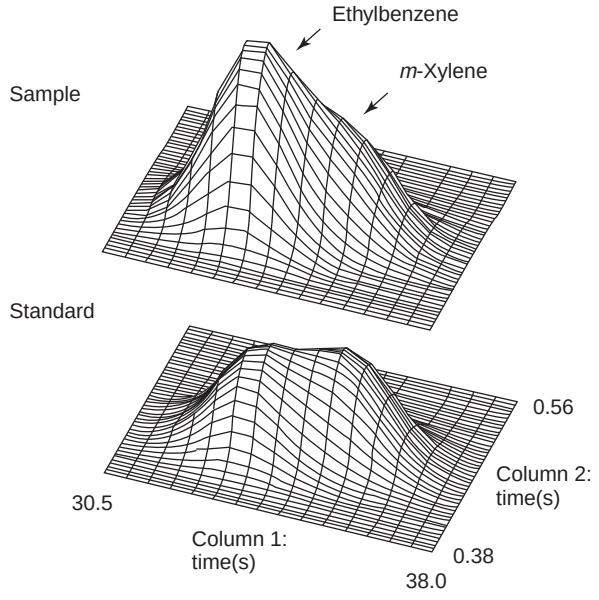

Figure 8 GC × GC responses for two mixtures of ethylbenzene and *m*-xylene. The sample is set no. 4 and the standard is set no. 6 as defined in Table 5. (Reprinted with permission from C.A. Bruckner, B.J. Prazen, R.E. Synovec, *Anal. Chem.*, **70**, 2796–2804 (1998). Copyright 1998 American Chemical Society.)

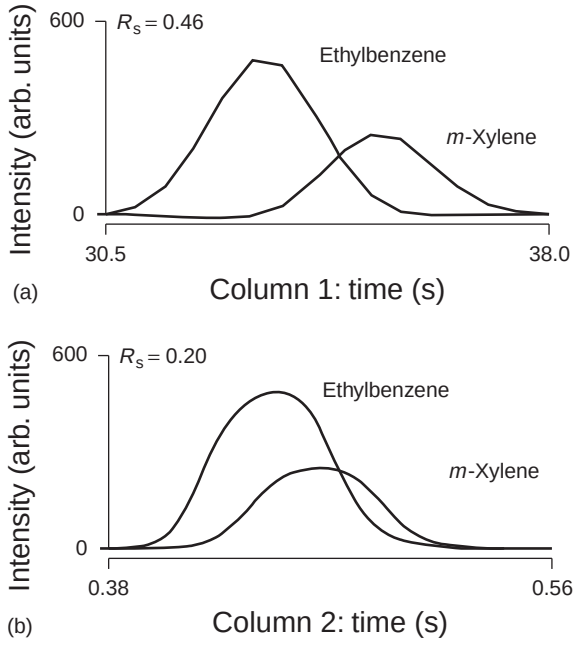

Figure 9 Resolved profiles of ethylbenzene and *m*-xylene on (a) the first GC column and (b) the second GC column after application of GRAM. (Reprinted with permission from C.A. Bruckner, B.J. Prazen, R.E. Synovec, *Anal. Chem.*, **70**, 2796–2804 (1998). Copyright 1998 American Chemical Society.)

technique produces bilinear data matrices, and the samples analyzed together give a trilinear structure, allowing GRAM to be used. Data analysis was carried out only on the section of data that contained the analytes of interest; the ethylbenzene and *m*-xylene responses were severely overlapped, with a resolution of 0.46 on the first axis of the gas chromatogram and 0.2 on the second axis: these data can be seen in Figure 8.

Five samples with varying concentrations of the analytes were used as test samples in the analysis, and a sixth sample was available as a standard. Despite the low resolution of the analytes, use of GRAM produced the pure compound elution profiles shown in Figure 9(a) and (b), in addition to the quantitative results shown in Table 5.

It can be seen that, for most of the samples, the GRAM results compare favorably with those of the reference method (where the gasoline samples were run on a conventional one-dimensional gas chromatograph so that all components were fully resolved). For the samples with low concentrations, however, the prediction error is fairly high. For ethylbenzene in sample 1, for example, the prediction is 22% lower than that of the reference method. The reason for this may be due to the large difference in ethylbenzene concentration between the sample and standard, as the level of ethylbenzene in the standard was 35 times greater than that in the sample, implying that analyte concentrations in the sample and standard should be similar in order to obtain accurate and precise quantitation with GRAM.

Table 5 GRAM concentration estimates for ethylbenzene and *m*-xylene, along with the relative standard deviation in each case. (Reprinted with permission from C.A. Bruckner, B.J. Prazen, R.E. Synovec, *Anal. Chem.*, **70**, 2796–2804 (1998). Copyright 1998 American Chemical Society.)

Set no.	Ethylbenzene (%, w/w)		*m*-Xylene (%, w/w)	
	Ref.[a]	GRAM[b] (RSD, %)[c]	Ref.[a]	GRAM[b] (RSD, %)[c]
1	0.031	0.024(7.8)	0.090	0.094(2.9)
2	1.10	1.09 (4.1)	0.090	0.097(5.0)
3	0.031	0.016(42)	1.11	1.08 (2.7)
4	2.20	2.19 (5.9)	1.11	1.07 (3.0)
5	1.12	1.06 (5.9)	2.13	2.20 (4.7)
6	1.11[d]		1.12[d]	

[a] Quantifying resolved peaks by peak area using a reference single-column gas chromatography (GC) method.
[b] Quantifying incompletely resolved components in GC × GC by GRAM.
[c] RSD is the relative standard deviation of predictions, expressed as a percentage.
[d] Set 6 used as standard. Compound concentrations were determined gravimetrically during sample preparation.

As was mentioned previously, complex results from GRAM are occasionally obtained. In fact, the eigenvalues and eigenvectors from the generalized eigenvalue

problem are guaranteed to be real only when **M** and **N** are positive definite and nonsingular. As they are generally real, unsymmetrical rectangular matrices, complex solutions will result at times. In particular, when the shape and retention time of any analyte vary between **M** and **N**, perhaps owing to the presence of a high-noise signal, or at other times to variations in instrumental conditions, complex results are produced. In such cases, the imaginary part of the estimated row and column profiles is too large to be attributed to round-off error. However, when such a problem arises, there is a solution in the form of a similarity transform.[51]

The success of the similarity transform is due to the fact that it can rotate the imaginary part of the complex solution to the real plane, and this can be achieved when complex eigenvalues and their corresponding eigenvectors occur in conjugate pairs. However, because eigenvalues and eigenvectors from the QZ algorithm do not always occur in this fashion, a preliminary step is required before the transform is applied. This first step following the solution of the eigenproblem is to order the eigenvalues and eigenvectors in descending order: the complex conjugate pairs of eigenvalues are grouped together by this operation. A preliminary transform, $\mathbf{T}_1$, is then carried out to group real eigenvalues with their corresponding eigenvectors. The direction of a complex number in the complex plane can be changed by multiplication with $e^{i\theta}$ without changing its magnitude. When eigenvalues occur in complex conjugate pairs but eigenvectors do not, it is therefore possible to find two arguments α and β that transform the eigenvectors simultaneously to complex conjugate pairs. The elements of $\mathbf{T}_1$ that perform this operation are $e^{i\alpha}$ and $e^{i\beta}$. The operation on the eigenvalues and eigenvectors is (Equation 38a and 38b):

$$\mathbf{Z}^* = \mathbf{Z}(\mathbf{T}_1)^{-1} \tag{38a}$$

$$\mathbf{\Lambda}^* = \mathbf{T}_1\mathbf{\Lambda}(\mathbf{T}_1)^{-1} \tag{38b}$$

Table 6 Eigenvectors, **V**, and eigenvalues, **Λ**, obtained from GRAM before application of the similarity transform. (Reprinted with permission from S. Li, J.C. Hamilton, P.J. Gemperline, *Anal. Chem.*, **64**, 599–608 (1992). Copyright 1992 American Chemical Society.)

Parameter	Value					
V	0.32051	$0.03541 + 0.01033i$	$0.01033 + 0.03541i$	−0.08109	−0.00536	−0.04901
	0.53393	$-0.01334 + 0.01781i$	$0.01781 - 0.01334i$	0.04569	0.02561	0.34098
	−0.62034	$0.03839 + 0.04811i$	$-0.04811 + 0.03839i$	−0.26059	0.03452	0.76677
	−0.03676	$0.43400 - 0.37987i$	$-0.37987 + 0.43400i$	0.63946	0.02958	0.47519
	0.44596	$0.22458 + 0.74070i$	$0.74070 + 0.22458i$	−0.28228	−0.23033	−0.22309
	−0.16473	$-0.07732 - 0.23799i$	$-0.23799 - 0.07732i$	0.65942	−0.97170	−0.13352
Λ	490.86	0	0	0	0	0
	0	$16.9 - 18.54i$	0	0	0	0
	0	0	$16.90 + 18.54i$	0	0	0
	0	0	0	−4.30	0	0
	0	0	0	0	0.81	0
	0	0	0	0	0	0.35

Table 7 After using the similarity transform, the eigenvectors, $\mathbf{V}^{**}$, and the eigenvalues, $\mathbf{\Lambda}^{**}$, no longer have imaginary parts. (Reprinted with permission from S. Li, J.C. Hamilton, P.J. Gemperline, *Anal. Chem.*, **64**, 599–608 (1992). Copyright 1992 American Chemical Society.)

Parameter	Value					
$\mathbf{V}^{**}$	0.32051	−0.01864	−0.03183	−0.08109	−0.00536	−0.04901
	0.53393	−0.01402	−0.01728	0.04569	0.025661	0.34098
	−0.62034	0.03731	−0.04895	−0.26059	0.03452	0.76677
	−0.03676	0.26265	−0.51349	0.63946	0.02958	0.47519
	0.44596	−0.77310	−0.03731	−0.28228	−0.23033	−0.22309
	−0.16473	0.24966	0.01699	0.65942	−0.97170	−0.13352
$\mathbf{\Lambda}^{**}$	490.86	0	0	0	0	0
	0	16.90	18.54	0	0	0
	0	−18.54	16.90	0	0	0
	0	0	0	−4.3	0	0
	0	0	0	0	0.81	0
	0	0	0	0	0	0.35

Finally, a second transform (the similarity transform, $\mathbf{T}_2$) is found which transforms the complex eigenvalues and eigenvectors to real eigenvalues and eigenvectors. The elements of $\mathbf{T}_2$ for a complex conjugate pair are (Equation 39):

$$\begin{matrix} t_{2_{j,j}} & t_{2_{j,j+1}} \\ t_{2_{j+1,j}} & t_{2_{j+1,j+1}} \end{matrix} = \begin{matrix} 1 & 1 \\ 1 & -1 \end{matrix} \quad (39)$$

The $\mathbf{T}_2$ matrix is applied to the eigenvalues and eigenvectors that are produced by the first operation displayed in Equation (38), to give Equations (40a) and (40b):

$$\mathbf{Z}^{**} = \mathbf{Z}^{*}(\mathbf{T}_2)^{-1} \quad (40a)$$

$$\mathbf{\Lambda}^{**} = \mathbf{T}_2\mathbf{\Lambda}^{*}(\mathbf{T}_2)^{-1} \quad (40b)$$

The effectiveness of this transform has been shown in the determination of unbound hydrocortisone in urine. LC/DAD was used to measure a real urine sample in addition to a standard hydrocortisone sample. When these data were analyzed by GRAM, complex values appeared in both the eigenvectors, $\mathbf{V}$, and the eigenvalues, $\mathbf{\Lambda}$ (see Table 6) thereby requiring the use of the transform.

Following this procedure, the imaginary parts of the eigenvalues and the eigenvectors had disappeared, as can be seen in Table 7.

The validity of the results is evident upon examination of the true hydrocortisone concentration ratio (33.3%) versus that produced by the modified GRAM algorithm (34.6%). The source of the imaginary numbers in this particular example may be due in part to the large rank of the system; owing to the complicated nature of the urine sample, it was found that six principal components were required to describe the data. As mentioned previously, slight retention time shifts and the presence of noise can cause complex results to be produced by GRAM, and in a case such as this where the samples contain many components, the adverse effects of shifts can be exacerbated. This is especially true if the shifts are nonlinear.

GRAM theory has been discussed at length in various publications since its introduction.[48,52,53] Degenerate solutions, for example, can arise when two components have the same eigenvalues and, in such a case, the corresponding eigenvectors do not have a fixed direction.[52] In calibration, this case can occur when the ratio of analyte A (between samples **M** and **N**) is the same as the concentration ratio of analyte B. In such a situation, the profiles determined by GRAM are not unique. If each of the signals from A and B are unimodal chromatographic peaks, both profiles will appear to be bimodal. Closer inspection will reveal that both profiles display maxima/minima at the retention times of A and B, as they are in fact linear combinations of these analytes. However, quantitation in this situation is not affected, as the ratios of the analytes are still valid. This has been demonstrated through the use of both GRAM and trilinear decomposition (TLD) on simulated emission–excitation data.[5]

Effects of model errors in GRAM have also been investigated.[53–55] Matrix effects can be minimized through the use of standard addition,[11] while interaction effects cannot (because a cross-term between concentrations is present in such a case). GRAM is very sensitive to retention time shifts, as stated above, thereby producing a bias in the estimated concentration ratios. A theoretical study of bias and variance in the estimated eigenvalues in a case where heteroscedastic noise exists found that a negative bias is generally obtained,[54] and these results have also been shown with realistic simulations.[56] A similar study in the case of homoscedastic noise, however, found a positive error to occur.[53] The agreement of the study with heteroscedastic noise to the simulation lends credence to the theory that the presence of homoscedastic noise is an unrealistic assumption, and does not adequately describe real systems.

GRAM can be extended to allow a TLD, also known as DTD.[5,10] GRAM is actually the simplest case of a third-order tensor, with only two slices in the third-order. When more than two slices are present, DTD can be applied to project all slices to just two, followed by SVD which finds a joint row, column and sample space.

The input to the algorithm is an $I \times J \times K$ block of data. The data are unfolded along each order, and in each case the resulting matrices are decomposed by SVD, as mentioned above. The left singular vectors for the first n principal components are retained from the decomposition of the joint row and column spaces, but in the case of the joint sample space, only the first two singular vectors are retained, to give Equations (41a–c):

$$[\mathbf{R}_1|\mathbf{R}_2|\mathbf{R}_3\cdots|\mathbf{R}_K] = \mathbf{U}_r\mathbf{S}_r\mathbf{V}_r^T \qquad \mathbf{U} = \mathbf{U}_r(I \times N) \quad (41a)$$

$$[\mathbf{R}_1|\mathbf{R}_2|\mathbf{R}_3\cdots|\mathbf{R}_I] = \mathbf{U}_c\mathbf{S}_c\mathbf{V}_c^T \qquad \mathbf{V} = \mathbf{U}_c(J \times N) \quad (41b)$$

$$[\mathbf{R}_1|\mathbf{R}_2|\mathbf{R}_3\cdots|\mathbf{R}_J] = \mathbf{U}_w\mathbf{S}_w\mathbf{V}_w^T \qquad \mathbf{W} = \mathbf{U}_w(K \times 2) \quad (41c)$$

Two representative matrices are then generated using Equations (42a) and (42b):

$$\mathbf{G}_1 = \sum_{k=1}^{K} w_{k,2}\mathbf{U}^T\mathbf{R}_k\mathbf{V} \quad (42a)$$

$$\mathbf{G}_2 = \sum_{k=1}^{K} w_{k,1}\mathbf{U}^T\mathbf{R}_k\mathbf{V} \quad (42b)$$

The first two columns of $\mathbf{W}$ are suitable for calculating representative matrices because they are the ''scores'' of the first two principal components, and are therefore likely to contain contributions from all components. GRAM is then carried out on $\mathbf{G}_1$ and $\mathbf{G}_2$ to find $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, where $\mathbf{Z}$ is generally the concentration order. The generalized

Table 8 Concentration estimates from GRAM and various formulations of TLD. (Reprinted with permission from K.S. Booksh, Z. Lin, Z. Wang, B.R. Kowalski, *Anal. Chem.*, **66**, 2561–2569 (1994). Copyright 1994 American Chemical Society.)

True	GRAM[a]	TLD[b]	TLD [Equation 49]	TLD [Equation 50]
1.0000	1.0009	0.9802	1.009	1.009

[a] From Li et al.[51]
[b] Using the algorithm from Sanchez and Kowalski.[5]

eigenproblem that is solved is[10] (Equation 43):

$$\mathbf{G}_1\boldsymbol{\Psi} = \mathbf{G}_2\boldsymbol{\Psi}\boldsymbol{\Lambda} \tag{43}$$

X and **Y** are then determined from the calculated eigenvectors (Equation 44):

$$\widehat{\mathbf{Y}} = \mathbf{V}(\boldsymbol{\Psi}^{\mathbf{T}})^{-1} \qquad \widehat{\mathbf{X}} = \mathbf{U}\mathbf{G}_2\boldsymbol{\Psi} \tag{44}$$

The concentrations in **Z** are determined by least squares. For each slice in the instrument response, $\mathbf{R}_k$, **Z** is calculated to minimize the error term (Equation 45):

$$\mathbf{R}_k = \sum_{n=1}^{N} \mathbf{z}_n \mathbf{x}_n \mathbf{y}_n^T + \mathbf{E}_k \tag{45}$$

The full $\widehat{\mathbf{Z}}$ $(K \times N)$ matrix can be calculated simultaneously by Equation (46):

$$\widehat{\mathbf{Z}} = \mathbf{P}\mathbf{Q}^{-1} \tag{46}$$

where (Equation 47):

$$p_{kn} = \sum_{i=1}^{I}\sum_{j=1}^{J} r_{ijk}\, \hat{x}_{in}\, \hat{y}_{jn} \tag{47}$$

and (Equation 48):

$$\mathbf{Q} = (\widehat{\mathbf{X}}\widehat{\mathbf{X}}^T) \cdot (\widehat{\mathbf{X}}\widehat{\mathbf{X}}^T) \tag{48}$$

where the $\cdot$ symbol signifies element-by-element multiplication.

The original DTD algorithm was slightly different, in that there were two eigenproblems to be solved,[10] where $\widehat{\mathbf{Y}}$ was calculated from Equation (49):

$$\mathbf{G}_1\boldsymbol{\Psi}_a = \mathbf{G}_2\boldsymbol{\Psi}_a\boldsymbol{\Lambda}_a \tag{49}$$

and $\widehat{\mathbf{X}}$ from Equation (50):

$$\mathbf{G}_1^T\boldsymbol{\Psi}_b = \mathbf{G}_2^T\boldsymbol{\Psi}_b\boldsymbol{\Lambda}_b \tag{50}$$

However, the resulting **X** and **Y** are often mismatched when this version is implemented, and therefore it is advised to calculate these parameters using a single eigenproblem. In other words, the *n*th eigenvector in Ψ_a does not always correspond to the *n*th eigenvector in Ψ_b. This has been demonstrated with two samples of simulated LC/DAD data, where the test sample contained the analyte of interest plus four interfering species, and the calibration standard contained only the analyte of interest.[10] The true relative analyte concentration in this case was 1.0000, and Table 8 shows the results that were obtained with different formulations of TLD in addition to the result given by GRAM. The prediction given by the original TLD algorithm (Equations 49 and 50) is low by 2%, owing to mismatched eigenvalues and eigenvectors; however, when they are rearranged so that they correspond, the error, although improved, is still higher than that given by GRAM plus the similarity transform. This is because, even though the pair of eigenvectors ($\boldsymbol{\psi}_{a1}$ and $\boldsymbol{\psi}_{a2}$) calculated from Equation (49) span the same space as the pair of eigenvectors ($\boldsymbol{\psi}_{b1}$ and $\boldsymbol{\psi}_{b2}$) calculated from Equation (50), different linear combinations of the eigenvectors yield different intrinsic profile estimates. In other words (Equation 51):

$$\hat{\mathbf{x}}_a \neq \hat{\mathbf{x}}_b \quad \text{and} \quad \hat{\mathbf{y}}_a \neq \hat{\mathbf{y}}_b \tag{51}$$

Therefore, Equation (52):

$$\mathbf{R}_1 + \mathbf{R}_2 \neq \hat{\mathbf{x}}_{b1}\hat{\mathbf{y}}_{a1} + \hat{\mathbf{x}}_{b2}\hat{\mathbf{y}}_{a2}^T \tag{52}$$

On the other hand, when Equations (43) and (44) are used, it is found that the quantitation results are identical with GRAM, and the prediction error is only 0.09%.

Similarly to GRAM, DTD occasionally has problems with complex solutions. This has been addressed with the use of similarity transforms, as with GRAM.[57] In this case, the transform is carried out on $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$ from the eigenvalue problem. However, the similarity transform, when applied to DTD decomposition factors, contains a rotational ambiguity. This means that when eigenvectors have occurred in complex pairs, the resulting chromatograms and spectra are not uniquely identified. The use of similarity transforms to eliminate the problem of complex solutions is therefore not as widely applicable to DTD data as to GRAM.

The rank annihilation methods have been applied in many different situations. GRAM has been successfully applied to LC/UV data, and also to bimodal chromatography data.[58,59] A unique application involved

nuclear magnetic resonance (NMR) data, where a single multicomponent pulsed-gradient spin-echo nuclear magnetic resonance (PGSENMR) data set was analyzed by GRAM.[60] This was possible because PGSENMR data consist of exponential decaying profiles, and the authors were able to "create" two data sets from a single one by using spectra 1 to $n-1$ for the first set and spectra 2 to n for the second set. Despite the highly overlapped nature of the data, GRAM could resolve the spectra of the two components present. Another extension or application of DTD is direct trilinear decomposition with matrix reconstruction (DTDMR).[61] This method is suitable for cases where one of the responses is actually a matrix. GRAM and DTD have both been used to predict heavy metal concentrations, from data obtained from a second-order fiber-optic heavy metal sensor.[62] The signals in each order were time and wavelength, respectively, and unknown interferents were present. The GRAM and DTD predictions both gave improvements over a zero-order analysis, which gave a positive bias. It was found, however, that when the sensor response deviated from linearity, the prediction errors increased, as is to be expected from these methods when the trilinear model is not held. In a separate study, DTD was used to resolve mathematically porphyrins from emission–excitation fluorescent spectra of canine and feline dental calculus deposits.[63] The small quantity of samples in that study had made physical separation and analysis unfeasible.

Although the trilinear model is the most commonly employed model for second-order calibration, there are cases where data do not fit this model. For example, when an instrument response is such that a pure analyte has a rank of >1, the data are nonbilinear. In such a case, application of any of the methods that specifically solve that type of data will not be of benefit, and may in fact result in an unreliable calibration. NBRA is a method that enables calibration of nonbilinear data when the response matrix of the pure analyte is available as a calibration standard.[6,64] The method is based on the model in Equation (53):

$$\mathbf{N} = \sum_{p=1}^{P} \mathbf{x}_p c_p \mathbf{y}_p^T \tag{53}$$

where $\mathbf{N}$ is a data matrix of a single pure component from a second-order instrument that produces nonbilinear data, P is the total number of detected components in the mixture, $\mathbf{x}$ and $\mathbf{y}$ are factors relating to the response profiles and c is related to concentration. The factors can be approximated using any three-way decomposition method, such as an eigenvalue–eigenvector decomposition, but it must be noted that the bilinear components from the decomposition do not contain the chemical and physical meaning that they do for bilinear data. (Approximating nonbilinear data using an eigenvalue–eigenvector decomposition is analogous to using PLS or PCR to approximate nonlinear first-order data with extra factors.)

It can be seen, therefore, that the greatest difference between nonbilinear and bilinear data lies in P, the number of factors in the data. For a pure component in a bilinear response matrix, there is one detectable factor present, as previously explained and shown in Equation (1). In other words, the rank of the matrix in the absence of noise is 1. On the other hand, the rank in a nonbilinear response matrix for a pure component is >1, so that when a second-order decomposition is carried out more than one concentration estimate is obtained.

A figure of merit that may be calculated to determine the reliability of the calibration on a particular data set is the rank linear additivity (RLA). It can be said that RLA holds if the rank of a mixture, $\mathbf{M}$, is equal to the sum of the ranks of individual components in the mixture. In that case, different components do not overlap with respect to their rank. For example, if $\mathbf{M}$ consists two components, a and b, whose instrument responses can be represented as in Equation (54a):

$$\mathbf{M}_a = c_a(\mathbf{x}_1\mathbf{y}_1^T + \mathbf{x}_2\mathbf{y}_2^T) \tag{54a}$$

and Equation (54b):

$$\mathbf{M}_b = c_b(\mathbf{x}_3\mathbf{y}_3^T + \mathbf{x}_4\mathbf{y}_4^T) \tag{54b}$$

where c_a and c_b are the concentrations of the components, and (Equation 55):

$$\mathbf{M} = \mathbf{M}_a + \mathbf{M}_b \tag{55}$$

then the calibration is similar to the case of bilinear components with identical concentrations. In such a case, the response profiles are not resolved because they are linearly combined, but the concentration prediction is accurate.[10] In this case, $(\mathbf{x}_1, \mathbf{x}_2)$ and $(\mathbf{y}_1, \mathbf{y}_2)$ would be linear combinations of one another, as would $(\mathbf{x}_3, \mathbf{x}_3)$ and $(\mathbf{y}_4, \mathbf{y}_4)$, but the four concentration estimates will simply be c_a repeated twice, and similarly for c_b.

If RLA does not hold, then the calibration becomes more complicated, because the components are now overlapping in rank. For example, if (Equation 56):

$$\mathbf{M} = c_a\mathbf{x}_1\mathbf{y}_1^T + c_b\mathbf{x}_3\mathbf{y}_3^T + \mathbf{x}_2(c_a\mathbf{y}_2^T + c_b\mathbf{y}_4^T) \tag{56}$$

then the rank of $\mathbf{M}$ equals 3 rather than 4.

Here it can be seen that $\mathbf{x}_2 = \mathbf{x}_4$, so that these factors, from different components, are not independent. What will happen when a decomposition is carried out? Will it be possible to calculate the concentrations of these

analytes? In a case such as this, it is important to calculate the net analyte rank (NAR).

NAR may be defined as (Equation 57):

$$\mathrm{NAR}(\mathbf{N}) = rank(\mathbf{M}) - rank(\mathbf{M}|\overline{\mathbf{N}}) \tag{57}$$

where $rank(\mathbf{M}|\overline{\mathbf{N}})$ is the rank of mixture **M** in the absence of the pure analyte **N**. If the NAR is equal to at least one, then calibration is still possible: it has been determined that NBRA retains the second-order advantage as long as the "NAR" of the data is ≥ 1.[65] In this case, shown in Equation (54), if the NAR is calculated for component a, then $\mathrm{NAR} = 3 - 2 = 1$. When the decomposition is carried out, therefore, the concentration estimate corresponding to the $\mathbf{x}_1\mathbf{y}_1$ factor will be c_a. The same situation holds for analyte b in this case. There is another way to determine the correct concentration estimate when there is just one analyte in the calibration sample, **N**. This procedure requires NBRA to be repeated using different multiples of N in place of N ($\varepsilon \times N$), and the resulting concentrations which are proportional to $1/\varepsilon$ are chosen as the correct concentration estimates. The corresponding $\mathbf{x}_i$, $\mathbf{y}_i$ pairs then comprise the net analyte signal.

Whether or not the data can be calibrated depends considerably on the instrument on which the data is collected. If the instrument produces data whose nonlinearities are polynomial in nature, then the data in general can be described by a low rank, and the NAR will generally be >1. If the nonlinearities are more complex, the rank will be higher and the bilinear model is less likely to be able to model the data. In such cases, the rank of a single component will be very high, and it is likely that there will be no NAR, thereby eliminating the ability to perform calibration. It has been found that MS/MS data generally fall into the first category, so that calibration of MS/MS data is possible with NBRA.

The NBRA method has been applied to two-dimensional NMR and to MS/MS data.[66] It was found that whereas NBRA tends to have a negative bias, the quantitative results were better than those provided by a first-order technique, multiple linear regression (MLR). The NMR data were for a six-component sugar mixture, while the MS/MS data were for warfarin, 3′-, 4′-, 5-, 6- and 7-hydroxywarfarin plus phenylbutazone. The average quantitation error over the test sets was 2.5%.

3.3 Multiway Partial Least Squares

PLS is a calibration technique that is widely used on first-order data. It is a method for building regression models between independent and dependent variables. Because of its success with first-order data, there has been interest in extending this technique to the calibration of second-order data.

The most obvious procedure might be to take the three-way data block, unfold it into a matrix and continue with regular PLS. However, this process has no advantage over two-way methods as far as prediction errors are concerned.[67] A second approach that may be taken is to divide the data block into two sets, one of which is the predictor set, the other being the test set. The model being estimated by Unfold-PLS is that represented by Equation (9a–c). However, because two sets of data are used, both must be modeled as Equation (58a) and (58b):

$$r_{ijk} = \sum_{n=1}^{N} x_{in}\, h_{njk} + e_{r,ijk} \tag{58a}$$

$$s_{ijk} = \sum_{n=1}^{N} x_{in}\, c_{njk} + e_{s,ijk} \tag{58b}$$

where s_{ijk} is an element of the dependent variable data set, $\underline{\mathbf{S}}$, and r_{ijk} is an element of the independent variable set, $\underline{\mathbf{R}}$. The model estimates the x_{in}, h_{njk} and c_{njk} parameters such that a compromise is made between minimizing $e^2_{r,ijk}$ and $e^2_{s,ijk}$, by maximizing the covariance of $\mathbf{x}_n$ and a linear combination of the variables in $\underline{\mathbf{S}}$.[68]

A third PLS method is named n-PLS, so-called because it is applicable to second-order and higher data.[69,70] In the case of three-way data, it is termed tri-PLS. PLS actually consists of two steps. In the first, the data array is decomposed, and in the second, a relationship between the independent ($\underline{\mathbf{R}}$) and dependent ($\mathbf{y}$) variables is established. While the array decomposition is trilinear, the part of $\underline{\mathbf{R}}$ relevant for describing $\mathbf{y}$ does not have to be of rank one. This enables calibration in the presence of matrix and other nonlinear effects. The tri-PLS algorithm is based on a decomposition of the data block such that the successively computed score factors have the property of maximum paired covariance with the unexplained part of the dependent variable; this constraint is not met in a trilinear sense with the other Unfold-PLS techniques. In this case, it is assumed that the dependent variables are in the form of a vector, **s**, rather than a block of data. This tri-PLS model boils down to the problem of finding the weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ that satisfy Equation (59):

$$\max_{\mathbf{w}_1\mathbf{w}_2} \quad \mathrm{cov}(\mathbf{x},\mathbf{s}) \;\; \min \;\; \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(r_{ijk} - x_i w_{1,j} w_{2,k})^2 \tag{59}$$

where **x**, $\mathbf{w}_1$ and $\mathbf{w}_2$ correspond to the factors for each mode in the decomposition. The least-squares solution to the tri-PLS model is given by Equation (60):

$$\max_{\mathbf{w}_1,\mathbf{w}_2} \; (\mathbf{w}_1)^T \mathbf{Z} \mathbf{w}_2 \tag{60}$$

where $\mathbf{Z}$ is a $J \times K$ matrix formed by $\underline{\mathbf{R}}$ and $\mathbf{s}$. The $\mathbf{w}_1$ and $\mathbf{w}_2$ vectors can be found by performing an SVD on $\mathbf{Z}$, and then the scores matrix, $\mathbf{X}$, can also be determined. The regression coefficients are calculated according to Equation (61):

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S} \tag{61}$$

This method can be extended to higher order arrays with an additional weight vector introduced for each extra way.

The applicability of Unfold-PLS and *n*-PLS can be seen in an example that determines the ash content of sugar.[69] In the analysis, the fluorescence of 67 samples of sugar was measured at four excitation and 63 emission wavelengths – this produced the $\underline{\mathbf{R}}$ block of data. The $\mathbf{y}$ block contained data regarding the ash content of each sample, as determined by a standard European Union method (no. 1265/09). The $\underline{\mathbf{R}}$ and $\mathbf{y}$ blocks were divided into a calibration and test set, and mean centering was carried out. The unfolded $\underline{\mathbf{R}}$ block and the $\mathbf{y}$ block are shown in Figure 10(a) and (b).

During the analysis, both Unfold-PLS and *n*-PLS used two components to describe the variations in $\mathbf{y}$, and the predictions of ash in the test sets were practically identical for both methods (see Table 9) This is due to the fact that the number of samples was large compared

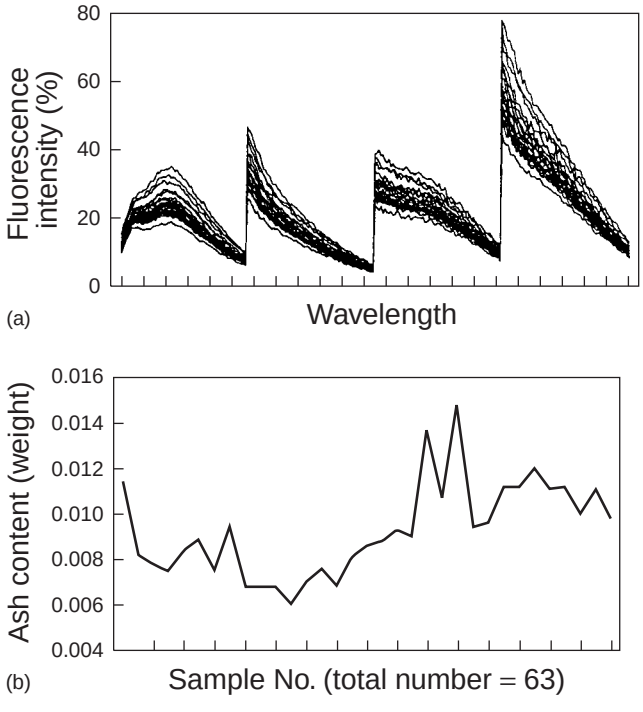

Figure 10 (a) Unfolded excitation–emission spectra from sugar samples and (b) the ash content in each sample. (Reprinted with permission from R. Bro, *J. Chemom.*, **10**, 47–61 (1996). Copyright 1996 John Wiley & Sons Limited.)

Table 9 Results from (i) Unfold-PLS and (ii) tri-PLS. (Reprinted with permission from R. Bro, *J. Chemom.*, **10**, 47–61 (1996). Copyright 1996 John Wiley & Sons Limited.)

Method	R^{2a}	RSD[b]
Tri-PLS	0.9396	6.95%
Unfold-PLS	0.9401	7.01%

[a] Correlation between concentrations and predictions.
[b] Relative standard deviation.

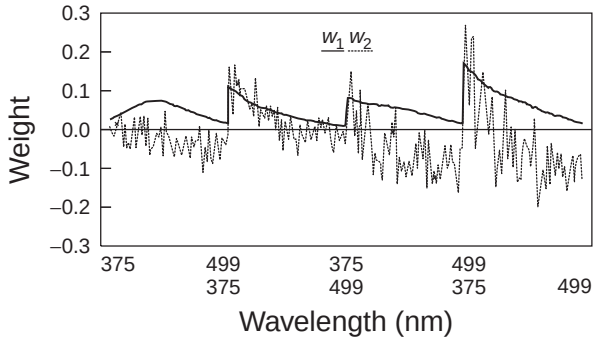

Figure 11 Weight vectors from Unfold-PLS. (Reprinted with permission from R. Bro, *J. Chemom.*, **10**, 47–61 (1996). Copyright 1996 John Wiley & Sons Limited.)

with the number of components. However, the difference between the methods is seen in the weight vectors, which show that the solution is much more stabilized in the case of *n*-PLS. (see Figures 11 and 12a and b). The Unfold-PLS weight vectors, especially that of the second component, have incorporated a lot of noise, unlike the tri-PLS vectors, which had fewer parameters to calculate.

4 CALIBRATION TRANSFER/ STANDARDIZATION

Standardization and calibration transfer are important topics that must often be addressed in order to obtain accurate calibration results.

Many of the multivariate calibration methods that have been discussed have strict requirements when it comes to the characteristics of the data employed and, in particular, these methods perform better when the signal from all analytes in different samples remain constant. Often, however, this is not the case. Consider, for example, an analyte that has been analyzed by LC/MS. If that analyte is analyzed again at a later time, its profile in the new

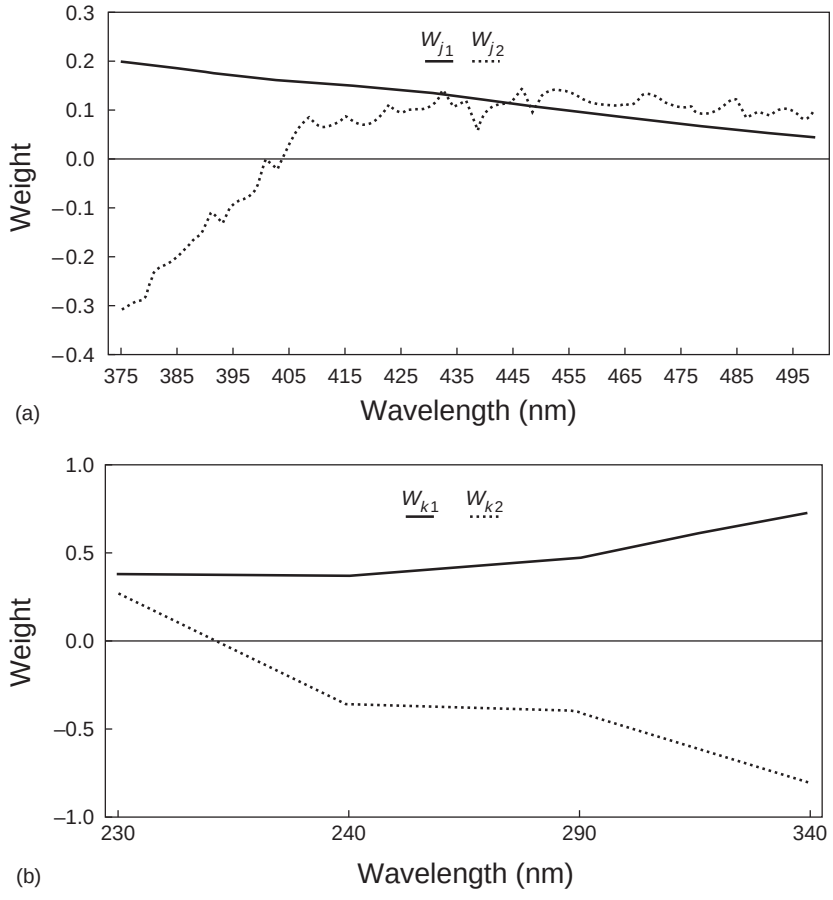

Figure 12 Weight vectors from trilinear-PLS in order of (a) excitation and (b) emission. (Reprinted with permission from R. Bro, *J. Chemom.*, **10**, 47–61 (1996). Copyright 1996 John Wiley & Sons Limited.)

data will not be identical with that in the earlier sample. Its retention time on the chromatographic axis may have changed, its peak shape may have changed or there may be a combination of both. In order for the second-order calibration to be effective, the differences between the analyte signals in the samples must first be minimized.

The first solution to this problem was proposed by Wang and Kowalski, and was named ''second-order standardization''.[12] This method standardizes two-dimensional responses measured on multiple instruments, or on a single instrument under different operational conditions. This technique does not, however, address the case where analyte response is constantly changing from sample to sample.

Second-order standardization is related to a method that exists for first-order (two-way) data, namely piecewise direct standardization (PDS).[71] PDS relates responses of samples measured on different instruments through the use of a banded diagonal matrix, **F** (Equation 62):

$$\overline{\mathbf{R}}_1 = \overline{\mathbf{R}}_2\mathbf{F} \tag{62}$$

where $\overline{\mathbf{R}}_1$ and $\overline{\mathbf{R}}_2$ are a small transfer set of samples measured on instruments 1 and 2 ($\overline{\mathbf{R}}$ refers to a truncated data set). This method assumes that the response at each channel is related to the response in a small window surrounding that channel on the second instrument. However, in second-order data, an analyte response may vary in both orders from instrument 1 to instrument 2, requiring a second banded matrix to account for changes in the second order. The placement of the banded matrices is (Equation 63):

$$\mathbf{R}_{1,\alpha} = \mathbf{A}\mathbf{R}_{2,\alpha}\mathbf{B} \tag{63}$$

where $\mathbf{R}_{1,\alpha}$ is the response from sample α on instrument 1 and $\mathbf{R}_{2,\alpha}$ is the response on instrument 2. **A** and **B** are banded left and right transformation matrices that correct for differences between the responses: these are estimated from a set of simultaneous nonlinear equations via the Gauss–Newton method. It must be noted that the left and right transformation matrices may not be found separately, as their solutions depend upon one another.

Table 10 Quantitative results obtained from GRAM for two different MVK–DNPH standards, before and after standardization. (Reprinted with permission from B.J. Prazen, R.E. Synovec, B.R. Kowalski, *Anal. Chem.*, **70**, 218–226 (1998). Copyright 1998 American Chemical Society.)

Experiment	Volumetric ratio of concentrations	GRAM ratio of concentrations without standardization	GRAM ratio of concentrations with standardization	Error without standardization (%)	Error with standardization (%)
MVK–DNPH standard 1	1.44	2.14	1.67	48	15
MVK–DNPH standard 2	0.722	0.871	0.0713	21	1.3

Following the estimation of **A** and **B**, they may then be used to transfer other samples that have also been run on the second instrument, so that it appears as if they have been run on the first, as (Equation 64):

$$\mathbf{R}_{1,\beta} = \mathbf{A}\mathbf{R}_{2,\beta}\mathbf{B} \tag{64}$$

where $\mathbf{R}_{1,\beta}$ and $\mathbf{R}_{2,\beta}$ are the responses of sample β run on the first and second instruments respectively.

In this way, samples run on instrument 2 can be calibrated using samples run on instrument 1, allowing a laboratory to calibrate their samples using a calibration set from a second laboratory, for example. However, one sample must be run on both instruments, and this transfer sample must have information in all channels, or the transfer will not be optimal. This is because the banded transformation matrices must contain information at all channels that are in new samples to be transferred. An additive correction matrix may be required in addition to Equations (62–64) to correct for additive differences between $\mathbf{R}_1$ and $\mathbf{R}_2$. However, a method to find the additive matrix does not currently exist in the literature.

This technique has been tested on LC/UV data, and variation from the ultraviolet (UV) data was found to be reduced from 0.2 AU before the transfer to 0.03 AU after the standardization.

A second standardization method that has been reported for second-order data is ''second-order chromatographic standardization''.[13] This technique, introduced by Prazen et al., is specifically directed toward data in which one of the orders is a chromatographic separation, while the other order is not chromatographic, e.g. GC/MS. The method corrects for retention shifts between a standard sample and a calibration sample, which allows calibration then to be performed.

In the first step of this method, the pseudo-rank of the sample matrix is estimated, and the two data sets are then augmented. SVD is carried out on the augmented matrix. The remaining eigenvalues beyond the pseudo-rank of the sample are used to calculate the percentage residual variance (%RV). The retention time of the samples is shifted and the process is repeated until the point is identified where the residual variance is at a minimum. The residual variance calculation is (Equation 65):

$$\%RV = 100 \left(\frac{\sum_{j=n+1}^{c} \lambda_j^0}{\sum_{j=1}^{c} \lambda_j} \right) \left(\frac{c \times r}{(c-n)(r-n)} \right) \tag{65}$$

where n is the number of chemical components, λ is an eigenvalue and c and r are the dimensions of the augmented matrix.

In order to use this standardization method effectively, the pseudo-rank of the sample matrix must be determined, which is not always a straightforward matter. In addition, the test sample must contain the same analytes as the standard sample, which reduces the generality of the procedure. However, the correct use of this technique allows calibration to be carried out with greater accuracy, as was demonstrated on LC/DAD data: methyl vinyl ketone–dinitrophenylhydrazine (MVK–DNPH) in a mixture of five DNPH derivatives was measured, along with two calibration standards, each of which contained only the MVK–DNPH analyte at different concentrations. When GRAM was carried out on the mixture with each of the standards, quantitation was inaccurate (see Table 10). Upon calculation of the %RV at different time shifts, it was found that one of the analyses required a shift of 0.2 s in the retention time, while the other required a 0.3 s shift, as can be seen in Figure 13.

After the correction, GRAM was performed once again, and the new concentration estimates were drastically improved, as is evident in Table 10. This shows the importance of data quality for accurate GRAM quantitation and the value of standardization in this type of case.

5 STANDARD ADDITION

Standard addition is a method that is commonly employed in zero- and first-order analyses to reduce the effects of

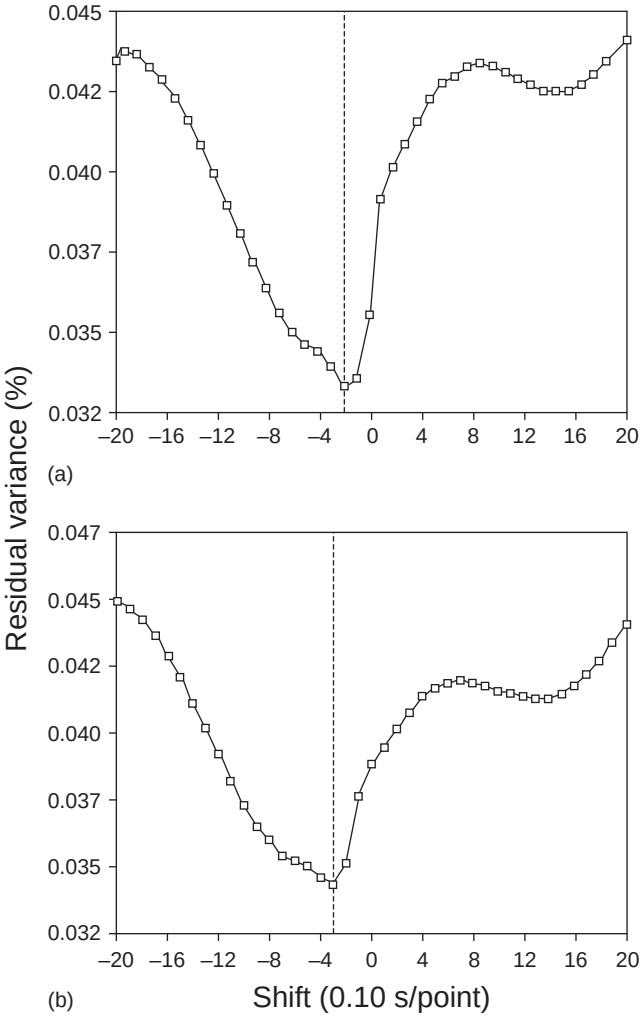

Figure 13 Standardization plots for two different MVK–DNPH analyses. The minimum shows the shift needed to align the matrices for GRAM analysis. (Reprinted with permission from B.J. Prazen, R.E. Synovec, B.R. Kowalski, *Anal. Chem.*, **70**, 218–226 (1998). Copyright 1998 American Chemical Society.)

matrix effects on calibration. In order for this method to be effective for zero-order data, there are two assumptions that must be fulfilled. First, the instrument response must vary linearly with analyte concentration. Second, for an analyte response of zero, the instrument response must be zero. The extension to first-order data, generalized standard addition, involves sequential spiking of the sample with the analyte plus each interferent that is affecting the analyte signal. In other words, each source of instrumental signal must be modeled and calibrated. When the data are from an instrument that produces second-order data, however, second-order methods can be used that eliminate this constraint. While the second-order advantage allows calibration in the presence of unmodeled interferents, this is only true when the data are completely bilinear. When there are interferents present in the sample that change the way in which the instrument responds to the analyte, then the data are no longer bilinear. For example, the analyte signal may change in scale or shape depending on the other components that are present in the sample, making traditional calibration unreliable. (Traditional in this sense refers to the second-order calibration methods that rely on the trilinear model.)

A SOSAM that mathematically separates the instrument response of the analyte from the instrument response of any interfering species was proposed by Booksh et al. in order to solve this problem.[11] While trilinear data follow the trilinear model (see Equation 7) which depends on the number of chemical species in the sample (N), ''fragmented'' data follow the model (Equation 66):

$$R_{ijk} = \sum_{m=1}^{M} X_{im}\, Y_{jm}\, Z_{km} + E_{ijk} \tag{66}$$

where M is the total number of detectable species formed by the N chemical components in the sample. This model is applicable in cases where two or more components in a sample mixture combine to form an additional component that can be detected. Matrix effects, on the other hand, occur when two or more sample components interact such that the profiles of the analyte depend on the concentration of the interferent in the sample. Both scenarios result in the data requiring standard addition for reliable calibration.

There are three steps to the SOSAM. The data block, $\underline{\mathbf{R}}$, must first be decomposed. This can be achieved using any of the second-order methods already mentioned, such as DTD or PARAFAC. Following this step, the N columns in the sample space of $\underline{\mathbf{R}}$ (i.e. **Z)** must be regressed against the standard additions to obtain N concentration estimates. For zero-order standard addition, a least-squares model is used (Equation 67):

$$r_k = \delta_k m + b \tag{67}$$

where r_k is the instrument response for the kth sample, δ_k is the change in analyte concentration from the first to the kth sample (from the standard additions) and m and b are the slope and intercept of the regression line, respectively. Using this model, the estimated analyte concentration in the original (first) sample is given by Equation (68):

$$\hat{c} = \frac{b}{m} \tag{68}$$

For the extension to second-order data, $\mathbf{Z}$ is the equivalent of r_k in Equation (65), because whether the data are bilinear or fragmented, there will be one column of $\mathbf{Z}$ that

is linearly related to the analyte of interest. Equation (69) shows how the slope and intercept may be calculated in the second-order case, so that Equation (68) may then be used to calculate the analyte concentration:

$$\hat{\mathbf{z}}_n = \delta m + b \tag{69}$$

Following the regression, a decision may have to be made as to which concentration estimate is correct. This will be the case when the data are fragmented, as there will be a number of concentration estimates, each of which will relate to a different factor. For example, there will be unique factors for both the interferent and analyte, but there may also be factors that relate to the interactions between them, in addition to factors that describe nonlinearities. The analyte concentration estimate may be determined as the smallest estimate that does not change when additional standard additions are included in $\underline{\mathbf{R}}$.

When applied to bilinear data, this algorithm is robust to shifting errors, and the bias in concentration estimates is low. Calibration in the case of fragmented data results in a low bias when the data are nondegenerate. When the algorithm is applied to degenerate data, however, the effect of random errors is increased. This may occur, for example, when two components have a common profile in one of the orders and therefore cannot be resolved.

This technique has been applied to a data set that resulted from the collection of UV diode-array spectra over a 30 min time period, over the course of a Fujiwara reaction. Trichloroethylene (TCE) in the presence of chloroform ($CHCl_3$) was quantified, and six standard additions of TCE were made. It had previously been discovered that TLD fails to predict the concentration of TCE in the presence of large quantities of $CHCl_3$ owing to an interaction between the two reagents.[72] In that case, the prediction error was found to be 29.5%. However, the data were determined to be fragmented and degenerate, therefore the model in Equation (64) applies. The prediction error was 10.8% when all six standard additions were included in the calculation, which is an improvement over the TLD error of 29.5%. Furthermore, when only the four most linear standard additions were used in the analysis, the error decreased to 2.6%, while TLD produced an error of 30.0% for the same standards.

6 FIGURES OF MERIT

Figures of merit provide a means of determining how well analytical determinations, including calibrations, perform. These performance characteristics include sensitivity, selectivity and S/N, in addition to net analyte signal. For good calibration, a high sensitivity and selectivity are desirable, as is a high S/N. A high sensitivity indicates that the analyte response changes significantly with concentration, which is important since small concentration changes may then be detected. Selectivity, on the other hand, is a measure of how well the method can distinguish the analyte in question from other analytes. This characteristic is related to the net analyte signal, which calculates the proportion of the signal unique to the analyte. A high net analyte signal translates into a high selectivity. Taken together, these characteristics can be used as criteria to decide whether or not a given instrument is suitable for attacking a particular analytical problem, or they can be used to determine the optimal set of analytical conditions for the problem.

The figures of merit mentioned above are well defined for zero-order data: the net analyte signal can be obtained simply be performing a background subtraction, while the sensitivity is determined from the slope of the calibration curve. Work by Lorber allowed them to be calculated for first-order data,[73,74] and there has also been research in the area of second-order figures of merit.[14,75] The derivation of the second-order figures of merit is much more complicated than the zero- and first-order versions, owing to the added complexity of the data. To date, there have been derivations only for bilinear data matrices.

Most figures of merit are related to the net analyte response, $\mathbf{R}_k^*$, which for the kth analyte in a second-order data matrix, $\mathbf{R}$, can be defined as (Equation 70):

$$\mathbf{R}_k^* = \mathbf{P}_{k,x}\mathbf{R}\mathbf{P}_{k,y} \tag{70}$$

where (Equations 71a and 71b):

$$\mathbf{P}_{k,x} = \mathbf{I} - \mathbf{X}_{-k}\mathbf{X}_{-k}^* \tag{71a}$$

$$\mathbf{P}_{k,y} = \mathbf{I} - \mathbf{Y}_{-k}\mathbf{Y}_{-k}^* \tag{71b}$$

While the net analyte response is a matrix that contains the profile of that part of the analyte signal that is orthogonal to every other component in the matrix, the net analyte signal, r_k^*, is a scalar that is derived from the net analyte response (Equation 72):

$$r_k^* = \| vec\mathbf{R}_k^* \| \tag{72}$$

where $\| \|$ refers to the Euclidean norm, and the *vec* operator corresponds to "vectorizing" the matrix. Selectivity, l_k, and sensitivity, s_k, are both found using the net analyte signal as (Equations 73 and 74):

$$l_k = \frac{r_k^*}{\| vec\underline{\mathbf{R}}_k \|} \tag{73}$$

$$s_k = \frac{r_k^*}{c_k} \tag{74}$$

A selectivity of zero implies that an analyte signal cannot be distinguished from other analytes – in other words, the net analyte signal is completely overlapped by other analyte signals. A selectivity of 1, on the other hand, means that the net analyte signal is not at all overlapped by other analyte signals.

These equations can be extended to *n*th-order data by using the *vec* operator to string out the data. Carrying this out on second-order data (e.g. Equation 2) produces Equation (75):

$$vec\mathbf{R} = \sum \mathbf{y}_k \otimes \mathbf{x}_k \tag{75}$$

where $\otimes$ is the N-fold Kronecker product.[14] Similarly, the *vec* operator can be applied to three-way arrays such that (Equation 76):

$$vec\underline{R} = \sum_{k=1}^{K} \overset{N}{\underset{n=1}{\otimes}} \mathbf{a}_{k,n} \tag{76}$$

where $\mathbf{A}_n$ $(n = 1, \ldots, N)$ is a factor matrix of $\underline{\mathbf{R}}$ instead of $\mathbf{X}$ and $\mathbf{Y}$. The net analyte response may then be denoted by Equation (77):

$$vec\underline{\mathbf{R}}_k^* = \left(\overset{N}{\underset{n=1}{\otimes}} \mathbf{P}_{k,n} \right) \quad vec\underline{\mathbf{R}} = \mathbf{P}_k vec\underline{\mathbf{R}} \tag{77}$$

As is the case with the second-order data, all of the figures of merit may be calculated from the net analyte response.

The second-order selectivity is one of the most important figures of merit for calibration, as a low value means that quantitation will be very difficult in the presence of interfering species. Selectivities of analytes in LC/DAD spectrochromatograms have been calculated, and serve to demonstrate typical values based on overlapped chromatograms and spectra in a data matrix. The analytes are toluene, naphthalene and *m*-xylene, all of which are spectrally similar, and biphenyl, which is spectrally dissimilar. The pure spectra and chromatograms are shown in Figure 14(a) and (b).

Upon calculation with Equation (73) for each analyte, the values in Table 11 were obtained. The first-order selectivities for the spectra of these analytes are shown in Table 12 for comparison.

The second-order selectivities can be explained based on the signals in the two orders (chromatographic and spectroscopic). For example, the largest value is for the comparison of naphthalene and biphenyl. When the first-order selectivity for this combination is examined, it can be seen that it is high, meaning that these analytes are the most dissimilar spectrally; moreover, the chromatographic profiles show that they are also the most dissimilar in that axis, resulting in the high second-order selectivity. On the other hand, *m*-xylene and biphenyl have the lowest selectivity, implying that

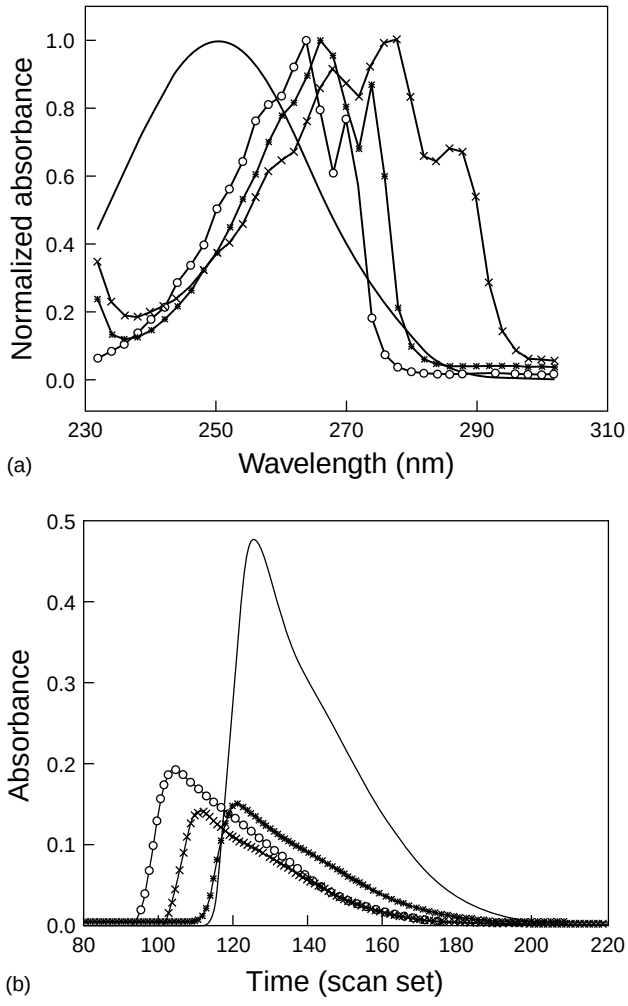

Figure 14 (a) Pure spectra and (b) chromatograms of toluene (○), naphthalene (×), *m*-xylene (∗) and biphenyl (no symbol). (Reprinted with permission from N.J. Messick, J.H. Kalivas, P.M. Lang, *Anal. Chem.*, **68**, 1572–1579 (1996). Copyright 1996 American Chemical Society.)

Table 11 Second-order selectivities for LC/DAD spectra for each pair of analytes. (Reprinted with permission from N.J. Messick, J.H. Kalivas, P.M. Lang, *Anal. Chem.*, **68**, 1572–1579 (1996). Copyright 1996 American Chemical Society.)

Component	Compared component	Second-order selectivity
Toluene	Naphthalene	0.7758
Toluene	*m*-Xylene	0.7893
Toluene	Biphenyl	0.8873
Naphthalene	*m*-Xylene	0.7272
Naphthalene	Biphenyl	0.8920
m-Xylene	Biphenyl	0.6916

Table 12 Spectral first-order selectivities for each pair of analytes. (Reprinted with permission from N.J. Messick, J.H. Kalivas, P.M. Lang, *Anal. Chem.*, **68**, 1572–1579 (1996). Copyright 1996 American Chemical Society.)

Component	Compared component	First-order selectivity
Toluene	Naphthalene	0.6877
Toluene	*m*-Xylene	0.3727
Toluene	Biphenyl	0.5530
Naphthalene	*m*-Xylene	0.5287
Naphthalene	Biphenyl	0.7684
m-Xylene	Biphenyl	0.6514

the signals from these analytes are more similar than any of the others. While their spectral similarity is not the highest (as seen by the relatively low first-order selectivity), their chromatographic profiles are very much overlapped, and this is what leads to their low second-order selectivity.

7 FUTURE WORK

While there have been tremendous advances in the area of second-order calibration, much work remains. Some of that work will pertain to improving the robustness and speed of the algorithms, while another part of the work should focus on increasing the awareness of the analytical community as to the value of such tools. Generally, while two-dimensional instruments are commonly used in today's laboratories, analysts often reduce the huge amount of data produced to one-dimension, and consequently they also reduce the quantity of information present. For example, when LC/MS is used, a two-dimensional data array is produced, but these data are often only viewed as a total ion chromatogram plot, thereby losing the mass spectral information. Often, the analysts are not aware of the existence of second-order methods. However, when this is not the case, the techniques are still not used owing to the perceived difficulties in their application. The great difficulty, therefore, is with the transfer of technology from the research arena to industry. Once this occurs, the use of second-order tools, including calibration, will become common place.

ACKNOWLEDGMENTS

C.M.F. acknowledges the financial support of the Endowed Analytical Professorship.

ABBREVIATIONS AND ACRONYMS

ALS	Alternating Least-squares
ATLD	Alternating Trilinear Decomposition
CANDECOMP	Canonical Decomposition
DTD	Direct Trilinear Decomposition
DTDMR	Direct Trilinear Decomposition with Matrix Reconstruction
EFA	Evolving Factor Analysis
GC	Gas Chromatography
GC × GC	Comprehensive Two-dimensional Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GRAM	Generalized Rank Annihilation Method
LC/DAD	Liquid Chromatography/Diode-array Detection
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LC/UV	Liquid Chromatography/ Ultraviolet Detection
MCR	Multivariate Curve Resolution
MLR	Multiple Linear Regression
MS/MS	Tandem Mass Spectrometry
NAR	Net Analyte Rank
NBRA	Nonbilinear Rank Annihilation
NMR	Nuclear Magnetic Resonance
NQS	1,2-Naphthoquinone-4-sulfonate
PARAFAC	Parallel Factor Analysis
PCR	Principal Component Regression
PDS	Piecewise Direct Standardization
PGSENMR	Pulsed-gradient Spin-echo Nuclear Magnetic Resonance
PLS	Partial Least Squares
RAFA	Rank Annihilation Factor Analysis
RLA	Rank Linear Additivity
RMSEP	Root Mean Square Error of Prediction
S/N	Signal-to-noise Ratio
SOSAM	Second-order Standard Addition Method
SVD	Singular Value Decomposition
TCE	Trichloroethylene
TLD	Trilinear Decomposition
UCC	Uncorrelated Correction Coefficient
Unfold-PCA	Unfold Principal Component Analysis
Unfold-PLS	Unfold Partial Least Squares
UV	Ultraviolet

RELATED ARTICLES

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Signal Processing in Analytical Chemistry • Soft Modeling of Analytical Data

Gas Chromatography **(Volume 12)**
Multidimensional Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy • Liquid Chromatography/Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

REFERENCES

1. K.S. Booksh, B.R. Kowalski, 'Theory of Analytical Chemistry', *Anal. Chem.*, **66**(15), 782A–791A (1994).
2. E. Sanchez, B.R. Kowalski, 'Tensorial Calibration: I. First-order Calibration', *J. Chemom.*, **2**(4), 247–264 (1988).
3. E. Sanchez, B.R. Kowalski, 'Tensorial Calibration: II. Second-order Calibration', *J. Chemom.*, **2**(4), 265–280 (1988).
4. E. Sanchez, B.R. Kowalski, 'Generalized Rank Annihilation Factor Analysis', *Anal. Chem.*, **58**(2), 496–499 (1986).
5. E. Sanchez, B.R. Kowalski, 'Tensorial Resolution: a Direct Trilinear Decomposition', *J. Chemom.*, **4**(1), 29–45 (1990).
6. B.E. Wilson, W. Lindberg, B.R. Kowalski, 'Multicomponent Quantitative Analysis Using Second-order Nonbilinear Data: Theory and Simulations', *J. Am. Chem. Soc.*, **111**(11), 3797–3804 (1989).
7. R.A. Harshman, 'Foundations of the PARAFAC Procedure: Models and Methods for an ''Explanatory'' Multi-Mode Factor Analysis', *UCLA Working Papers Phonetics*, **16**, 1–84, 1970.
8. H.-L. Wu, M. Shibukawa, K. Oguma, 'An Alternating Trilinear Decomposition Algorithm with Application to Calibration of HPLC–DAD for Simultaneous Determination of Overlapped Chlorinated Aromatic Hydrocarbons', *J. Chemom.*, **12**(1), 1–26 (1998).
9. R. Tauler, B.R. Kowalski, S. Fleming, 'Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial Process', *Anal. Chem.*, **65**(15), 2040–2047 (1993).
10. K.S. Booksh, Z. Lin, Z. Wang, B.R. Kowalski, 'Extension of Trilinear Decomposition Method with an Application to the Flow Probe Sensor', *Anal. Chem.*, **66**(15), 2561–2569 (1994).
11. K. Booksh, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, 'A Second-order Standard Addition Method with Application to Calibration of a Kinetics–Spectroscopic Sensor for Quantitation of Trichloroethylene', *J. Chemom.*, **9**(4), 263–282 (1995).
12. Y. Wang, B.R. Kowalski, 'Standardization of Second Order Instruments', *Anal. Chem.*, **65**(9), 1174–1180 (1993).
13. B.J. Prazen, R.E. Synovec, B.R. Kowalski, 'Standardization of Second-order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis', *Anal. Chem.*, **70**(2), 218–226 (1998).
14. K. Faber, A. Lorber, B.R. Kowalski, 'Analytical Figures of Merit for Tensorial Calibration', *J. Chemom.*, **11**(5), 419–461 (1997).
15. R. Bro, 'PARAFAC. Tutorials and Applications', *Chemom. Intell. Lab. Syst.*, **38**(2), 149–171 (1997).
16. L.R. Tucker, 'Relations between Multidimensional Scaling and Three-mode Factor Analysis', *Psychometrika*, **37**(1), 3–27 (1972).
17. P. Geladi, 'Analysis of Multi-way (Multi-mode) Data', *Chemom. Intell. Lab. Syst.*, **7**(1–2), 11–30 (1989).
18. A.K. Smilde, 'Three-way Analyses. Problems and Aspects', *Chemom. Intell. Lab. Syst.*, **15**(2–3), 143–157 (1992).
19. K.S. Booksh, B.R. Kowalski, 'Calibration Method Choice by Comparison of Model Basis Functions to the Theoretical Instrument Response Function', *Anal. Chim. Acta*, **348**(1–3), 1–9 (1997).
20. J.D. Carroll, J.J. Chang, 'Analysis of Individual Differences in Multidimensional Scaling via an *N*-way Generalization of ''Eckart–Young'' Decomposition', *Psychometrika*, **35**(3), 283–319 (1970).
21. C.J. Appellof, E.R. Davidson, 'Strategies for Analyzing Data from Video Fluorometric Monitoring of Liquid Chromatographic Effluents', *Anal. Chem.*, **53**(13), 2053–2056 (1981).
22. B.C. Mitchell, D.S. Burdick, 'Slowly Converging PARAFAC Sequences: Swamps and Two-factor Degeneracies', *J. Chemom.*, **8**(2), 155–168 (1994).
23. R.A. Harshman, M.E. Lundy, 'The PARAFAC Model for Three-way Factor Analysis and Multidimensional Scaling', in *Research Methods for Multimode Data Analysis*, eds. H.G. Law, C.W. Snyder Jr, J.A. Hattie, R.P. McDonald, Praeger, New York, 122–215, 1984.
24. B.C. Mitchell, D.S. Burdick, 'An Empirical Comparison of Resolution Methods for Three-way Arrays', *Chemom. Intell. Lab. Syst.*, **20**(2), 149–161 (1993).
25. W.S. Rayens, B.C. Mitchell, 'Two-factor Degeneracies and a Stabilization of PARAFAC', *Chemom. Intell. Lab. Syst.*, **38**(2), 173–181 (1997).

26. C.A. Andersson, R. Bro, 'Improving the Speed of Multiway Algorithms. Part I: Tucker3', *Chemom. Intell. Lab. Syst.*, **42**(2), 93–103 (1998).
27. R. Bro, C.A. Andersson, 'Improving the Speed of Multiway Algorithms. Part II: Compression', *Chemom. Intell. Lab. Syst.*, **42**(2), 105–113 (1998).
28. H.A.L. Kiers, W.P. Krijnen, 'An Efficient Algorithm for PARAFAC of Three-way Data with Large Numbers of Observation Units', *Psychometrika*, **56**(1), 147–152 (1991).
29. R. Bro, S. de Jong, 'A Fast Non-negativity Constrained Least Squares Algorithm', *J. Chemom.*, **11**(5), 393–401 (1997).
30. P. Paatero, 'A Weighted Non-negative Least Squares Algorithm for Three-way 'PARAFAC' Factor Analysis', *Chemom. Intell. Lab. Syst.*, **38**(2), 223–242 (1997).
31. K.S. Booksh, A.R. Muroski, M.L. Myrick, 'Single-measurement Excitation/Emission Matrix Spectrofluorometer for Determination of Hydrocarbons in Ocean Water. 2. Calibration of Naphthalene and Styrene', *Anal. Chem.*, **68**(20), 3539–3544 (1996).
32. I.E. Bechmann, 'Second-order Data by Flow Injection Analysis with Spectrophotometric Diode-array Detection and Incorporated Gel-filtration Chromatographic Column', *Talanta*, **44**(4), 585–591 (1997).
33. R. Bro, H. Heimdal, 'Enzymatic Browning of Vegetables. Calibration and Analysis of Variance by Multiway Methods', *Chemom. Intell. Lab. Syst.*, **34**(1), 85–102 (1996).
34. P. Hindmarch, K. Kavianpour, R.G. Brereton, 'Evaluation of Parallel Factor Analysis for the Resolution of Kinetic Data by Diode-array High Performance Liquid Chromatography', *Analyst*, **122**(9), 871–877 (1997).
35. R. Tauler, A. Izquierdo-Ridorsa, E. Casassas, 'Simultaneous Analysis of Several Spectroscopic Titrations with Self-modeling Curve Resolution', *Chemom. Intell. Lab. Syst.*, **18**(3), 293–300 (1993).
36. R. Tauler, S. Lacorte, D. Barcelo, 'Application of Multivariate Self-modeling Curve Resolution to the Quantitation of Trace Levels of Organophosphorus Pesticides in the Natural Waters from Interlaboratory Studies', *J. Chromatogr. A*, **730**(1–2), 117–183 (1996).
37. R. Tauler, 'Multivariate Curve Resolution Applied to Second-order Data', *Chemom. Intell. Lab. Syst.*, **30**(1), 133–146 (1995).
38. M. Maeder, 'Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks', *Anal. Chem.*, **59**(3), 527–530 (1997).
39. J. Saurina, S. Hernandez-Cassou, R. Tauler, 'Second-order Curve Resolution Applied to the Flow Injection Analysis of Mixtures of Amino Acids', *Anal. Chim. Acta*, **335**(1–2), 41–49 (1996).
40. R. Tauler, A. Smilde, B. Kowalski, 'Selectivity, Local Rank, Three-way Data Analysis and Ambiguity in Multivariate Curve Resolution', *J. Chemom.*, **9**(1), 31–58 (1995).
41. J. Saurina, S. Hernandez-Cassou, R. Tauler, 'Multivariate Curve Resolution and Trilinear Decomposition Methods in the Analysis of Stopped-flow Kinetic Data for Binary Amino Acid Mixtures', *Anal. Chem.*, **69**(13), 2329–2336 (1997).
42. A. de Juan, S.C. Rutan, R. Tauler, D.L. Massart, 'Comparison Between the Direct Trilinear Decomposition and the Multivariate Curve Resolution–Alternating Least Squares Methods for the Resolution of Three-way Data Sets', *Chemom. Intell. Lab. Syst.*, **40**(1), 19–32 (1998).
43. C.-N. Ho, G.D. Christian, E.R. Davidson, 'Application of the Method of Rank Annihilation to Quantitative Analysis of Multicomponent Fluorescence Data from the Video Fluorometer', *Anal. Chem.*, **50**(8), 1108–1113 (1978).
44. C.-N. Ho, G.D. Christian, E.R. Davidson, 'Application of the Method of Rank Annihilation to Fluorescent Multicomponent Mixtures of Polynuclear Aromatic Hydrocarbons', *Anal. Chem.*, **52**(7), 1071–1079 (1980).
45. C.-N. Ho, G.D. Christian, E.R. Davidson, 'Simultaneous Multicomponent Rank Annihilation and Applications to Multicomponent Fluorescent Data Acquired by the Video Fluorometer', *Anal. Chem.*, **53**(1), 92–98 (1981).
46. A. Lorber, 'Quantifying Chemical Composition from Two-dimensional Data Arrays', *Anal. Chim. Acta*, **164**, 293–297 (1984).
47. B.E. Wilson, E. Sanchez, B.R. Kowalski, 'An Improved Algorithm for the Generalized Rank Annihilation Method', *J. Chemom.*, **3**(3), 493–498 (1988).
48. N.M. Faber, L.M.C. Buydens, G. Kateman, 'Generalized Rank Annihilation Method. III: Practical Implementation', *J. Chemom.*, **8**(4), 273–285 (1994).
49. C.M. Fleming, B.R. Kowalski, unpublished work, 1998.
50. C.A. Bruckner, P.J. Prazen, R.E. Synovec, 'Comprehensive Two-dimensional High-speed Gas Chromatography with Chemometric Analysis', *Anal. Chem.*, **70**(14), 2796–2804 (1998).
51. S. Li, J.C. Hamilton, P.J. Gemperline, 'Generalized Rank Annihilation Method using Similarity Transformations', *Anal. Chem.*, **64**(6), 599–607 (1992).
52. N.M. Faber, L.M.C. Buydens, G. Kateman, 'Generalized Rank Annihilation Method. I: Derivation of Eigenvalue Problems', *J. Chemom.*, **8**(2), 147–154 (1994).
53. N.M. Faber, L.M.C. Buydens, G. Kateman, 'Generalized Rank Annihilation Method. II: Bias and Variance in the Estimated Eigenvalues', *J. Chemom.*, **8**(3), 181–203 (1994).
54. K. Faber, A. Lorber, B.R. Kowalski, 'Generalized Rank Annihilation Methods: Standard Errors in the Estimated Eigenvalues if the Instrumental Errors are Heteroscedastic and Correlated', *J. Chemom.*, **11**(2), 95–109 (1997).
55. K. Booksh, B.R. Kowalski, 'Error Analysis of the Generalized Rank Annihilation Method', *J. Chemom.*, **8**(1), 45–63 (1994).

56. R.B. Poe, S.C. Rutan, 'Effects of Resolution, Peak Ratios and Sampling Frequency in Diode-array Fluorescence Detection in Liquid Chromatography', *Anal. Chim. Acta*, **283**(2), 845–853 (1993).
57. S.S. Li, P.J. Gemperline, 'Eliminating Complex Eigenvectors and Eigenvalues in Multiway Analyses Using the Direct Trilinear Decomposition Method', *J. Chemom.*, **7**(2), 77–88 (1993).
58. E. Sanchez, L.S. Ramos, B.R. Kowalski, 'Generalized Rank Annihilation Method. I. Application to Liquid Chromatography–Diode Array Ultraviolet Detection Data', *J. Chromatogr.*, **385**, 151–164 (1987).
59. L.S. Ramos, E. Sanchez, B.R. Kowalski, 'Generalized Rank Annihilation Method. II. Analysis of Bimodal Chromatographic Data', *J. Chromatogr.*, **385**, 165–180 (1987).
60. W. Windig, B. Antalek, 'Direct Exponential Curve Resolution Algorithm (DECRA): A novel Application of the Generalized Rank Annihilation Method for a Single Spectral Mixture Data Set with Exponentially Decaying Contribution Profiles', *Chemom. Intell. Lab. Syst.*, **37**(2), 241–254 (1997).
61. Y.S. Zeng, P.K. Hopke, 'A New Receptor Model: A Direct Trilinear Decomposition Followed by a Matrix Reconstruction', *J. Chemom.*, **6**(2), 65–83 (1992).
62. Z.H. Lin, K.S. Booksh, L.W. Burgess, B.R. Kowalski, 'Second-order Fiber Optic Heavy Metal Sensor Employing Second-order Tensorial Calibration', *Anal. Chem.*, **66**(15), 2552–2560 (1994).
63. M.M.C. Ferreira, M.L. Brandes, I.M.C. Ferreira, K.S. Booksh, W.C. Dolowy, M. Gouterman, B.R. Kowalski, 'Chemometric Study of the Fluorescence of Dental Calculus by Trilinear Decomposition', *Appl. Spectrosc.*, **49**(9), 1317–1325 (1995).
64. A. Lorber, 'Features of Quantifying Chemical Composition from Two-dimensional Data Array by the Rank Annihilation Factor Analysis Method', *Anal. Chem.*, **57**(12), 2395–2397 (1985).
65. Y.D. Wang, O.S. Borgen, B.R. Kowalski, M. Gu, F. Turecek, 'Advances in Second-order Calibration', *J. Chemom.*, **7**(2), 117–130 (1993).
66. B.E. Wilson, B.R. Kowalski, 'Quantitative Analysis in the Presence of Spectral Interferents using Second-order Non-bilinear Data', *Anal. Chem.*, **61**(20), 2277–2284 (1989).
67. C. Demir, R.G. Brereton, 'Calibration of Gas Chromatography–Mass Spectrometry of Two-component Mixtures using Univariate Regression and Two- and Three-way Partial Least Squares', *Analyst*, **122**(7), 631–638 (1997).
68. A.K. Smilde, D.A. Doornbos, 'Simple Validatory Tools for Judging the Predictive Performance of PARAFAC and Three-way PLS', *J. Chemom.*, **6**(1), 11–28 (1992).
69. R. Bro, 'Multiway Calibration. Multilinear PLS', *J. Chemom.*, **10**(1), 47–61 (1996).
70. A.K. Smilde, 'Comments on Multilinear PLS', *J. Chemom.*, **11**(5), 367–377 (1997).
71. Y.D. Wang, B.R. Kowalski, 'Calibration Transfer and Measurement Stability of Near-infrared Spectrometers', *Appl. Spectrosc.*, **46**(5), 764–771 (1992).
72. J.M. Henshaw, 'Reaction-based Chemical Sensing for the Multicomponent Determination of Chlorinated Hydrocarbons', Doctoral Dissertation, University of Washington, 1993.
73. A. Lorber, 'Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations', *Anal. Chem.*, **58**(6), 1167–1172 (1986).
74. A. Lorber, A. Harel, Z. Goldbart, I.B. Brenner, 'Curve Resolution and Figures of Merit Estimation for Determination of Trace Elements in Geological Materials by Inductively Coupled Plasma Atomic Emission Spectrometry', *Anal. Chem.*, **59**(9), 1260–1266 (1987).
75. N.J. Messick, J.H. Kalivas, P.M. Lang, 'Selectivity and Related Measures for *n*th Order Data', *Anal. Chem.*, **68**(9), 1572–1579 (1996).

Signal Processing in Analytical Chemistry

Peter D. Wentzell and Christopher D. Brown
Dalhousie University, Halifax, Canada

1 Introduction **1**
2 Overview **2**
2.1 History 2
2.2 Definitions and Notation 2
3 Signals and Noise **2**
3.1 Signals and Signal Domains 2
3.2 Noise 3
3.3 Signal Averaging 5
4 Analog Signal Processing **6**
4.1 Overview 6
4.2 Analog Filters 7
4.3 Analog-to-digital Conversion 10
5 Digital Filtering **12**
5.1 Introduction 12
5.2 Filter Types 12
5.3 Polynomial Least-squares Smoothing Filters 14
5.4 Derivative Filters 17
5.5 Kalman Filters 18
5.6 Other Filters 21
6 Domain Transformations **21**
6.1 Introduction 21
6.2 Fourier Transforms 21
6.3 Wavelet Transforms 27
6.4 Hadamard Transforms 30
7 Higher-order Signal Processing **31**
8 Conclusions **33**
Acknowledgments **34**
Abbreviations and Acronyms **34**
Related Articles **34**
References **34**

Signal processing refers to a variety of operations that can be carried out on a continuous (analog) or discrete (digital) sequence of measurements in order to enhance the quality of information it is intended to convey. In the analog domain, electronic signal processing can encompass such operations as amplification, filtering, integration, differentiation, modulation/demodulation, peak detection, and analog-to-digital (A/D) conversion. Digital signal processing can include a variety of filtering methods (e.g. polynomial least-squares smoothing, differentiation, median smoothing, matched filtering, boxcar averaging, interpolation, decimation, and Kalman filtering) and domain transformations (e.g. Fourier transform (FT), Hadamard transform (HT), and wavelet transform (WT)). Generally the objective is to separate the useful part of the signal from the part that contains no useful information (the noise) using either explicit or implicit models that distinguish these two components. Signal processing at various stages has become an integral part of most modern analytical measurement systems and plays a critical role in ensuring the quality of those measurements.

1 INTRODUCTION

The reliability of analytical results is vitally dependent on the quality of the measurements leading to their determination. Signal processing refers to a variety of operations that can be carried out on a continuous or discrete sequence of measurements in order to enhance the quality of information they are intended to convey. The term 'signal' is ordinarily applied to a sequence of measurements that are related by some ordinal variable, such as time or wavelength, and usually obtained via electrical transduction. Operations can be carried out on a continuous electrical signal (analog signal processing) or on discretely sampled numerical values (digital signal processing). Generally, the objective is to separate the desired part of the signal (the pure signal, which is correlated to some physical or chemical property of interest) from the unwanted part of the signal, or noise. Often, the term 'signal processing' implies that operations are carried out in real time, or as the data are acquired, but this is not a requirement. The distinction between signal processing and other forms of data analysis is often open to interpretation, but usually signal processing emphasizes alternative representations of the sequence of measurements as opposed to the direct extraction of secondary information such as analyte concentrations or chemical structures.

The goal of this article is to provide an overview of signal processing methods used in analytical chemistry with an emphasis on their capabilities, weaknesses, and practical implementation. Although both analog and digital signal processing are discussed, a much greater emphasis is placed on the latter because of its greater relevance to the practicing analytical chemist. Owing to the scope of the subject area, some topics will no doubt be neglected or underemphasized, but an attempt has been made to balance coverage with the importance to the field.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 OVERVIEW

2.1 History

Historically, signal processing in analytical chemistry can be regarded as originating with the first quantitative analytical measurements, such as the end-point of a titration, the potential of an electrochemical cell, or the absorbance of a solution at a single wavelength. Such scalar quantities are now sometimes referred to as zero-order measurements, with reference to the fact that a scalar is a zero-order tensor. Rudimentary signal processing consisted of averaging replicate measurements or damping the response of electrical signals, but these were rather simplistic approaches when compared to today's more sophisticated methods.

The evolution of modern signal processing in analytical chemistry can be traced to three parallel developments. The first was the emergence of analytical instruments capable of producing first-order data, or a vector (first-order tensor) of measurements. The development of instruments such as chromatographs and scanning spectrometers meant that the signal from an instrument could no longer be considered static, but rather changed in a regular fashion with some variable such as time or wavelength, thereby requiring more flexible signal processing methods that could remove the noise without distorting the pure signal. A second important influence was the appearance of the analog-to-digital converter (ADC) and the computers that drove them. This opened the door to more versatile digital signal processing methods. Finally, there was the parallel development of efficient digital signal processing algorithms, such as the Savitzky–Golay (SG) implementation of polynomial least-squares filters,[1] and the Cooley–Tukey algorithm for the fast Fourier transform (FFT),[2] which are still among the most highly cited papers in the literature. Such algorithms made the practical implementation of signal processing methods a reality.

Although first-order instruments are still the mainstay of analytical chemistry, second-order instruments which provide a matrix of data, such as chromatographs with multichannel detection and tandem mass spectrometers, are now routinely employed, and higher-order instruments are commonplace. These, combined with ever more powerful computational platforms, have advanced modern signal processing to yet another level.

2.2 Definitions and Notation

Throughout this article, the term 'signal' will be used to refer to either a continuous or discrete measurement sequence which consists of a pure or undistorted signal corrupted by noise. The signal is implicitly measured as a function of some other variable which will be referred to as the ordinal variable because it correlates directly with the sequence order. Traditionally, this variable is time, but other variables such as wavelength or applied voltage (which may or may not be correlated with time), can be employed without loss of generality. References to representation of the signal in the 'time domain' will therefore refer to this the original measurement sequence even if the ordinal variable is not time. Likewise, references to 'frequency domain' representations (section 3) will refer to a FT (section 6) into the inverse domain of the ordinal variable.

The term 'noise' is used to refer to unwanted fluctuations from the pure signal that obscure its measurement. This definition is quite general and means that what is considered noise can vary with the situation. If a signal consists of contributions from two sources, A and B, then B is considered noise if one is looking for A, and vice versa. Noise, especially random noise, is often characterized by its root-mean-square (rms) amplitude (continuous signals) or by its standard deviation (discrete signals).

Most of this article concerns digital signal processing methods where the signal is a vector of discrete measurements, usually made on equal intervals of the ordinal variable referred to as the sampling interval, t_s. The reciprocal of this interval will be referred to as the sampling frequency, f_s. (Again, analogous definitions hold when the ordinal variable is not time.) In vector notation, the signal $\mathbf{x}$ is the result of combining the pure signal vector $\mathbf{x^0}$ with the noise vector $\mathbf{e}$ (Equation 1):

$$\mathbf{x} = \mathbf{x^0} + \mathbf{e} \tag{1}$$

Throughout this article, boldface lower case letters are used to represent vectors (assumed to be column vectors unless otherwise indicated). Scalar quantities will be represented by lower case italic letters and matrices by boldface upper case letters. Transposes of vectors or matrices will be indicated with a superscript 'T' and inverses of matrices with a superscript '-1'. The identity matrix will be designated '$\mathbf{I}$'.

3 SIGNALS AND NOISE

3.1 Signals and Signal Domains

The features that distinguish a signal from a random series of measurements are: the measurements have a definite order, and the measurements are generally correlated in the time domain. For a discrete signal undistorted by noise, $\mathbf{x^0} = [x_1^0\, x_2^0\, x_3^0 \ldots]$, the correlation of two measurements is described by their covariance (Equation 2),

$$\mathrm{cov}(x_1^0, x_2^0) = \mathrm{E}(x_1^0 x_2^0) \tag{2}$$

where 'E' denotes the expectation value. To say that two measurements are correlated means that knowing one allows us to say something about the other. Although the covariance among measurements in a sequence is not generally known, some knowledge of its characteristics may be known, and it is these characteristics that are often used to distinguish the pure signal from the noise. For example, if the pure signal changes relatively slowly, it exhibits long-range correlations which may not be present in the noise.

Although signals are usually presented in the time domain, the same information can be conveyed by transforming them in alternate domains. The most useful of these is the frequency domain. For a continuous signal, this transformation can be accomplished by using a spectrum analyzer, a device that consists of a continuous series of electronic bandpass filters. The signal, which must be repetitively applied to the input of the device, is filtered by the spectrum analyzer and the rms signal at each frequency is determined. In essence, the power of the signal is plotted as a function of frequency to give a power spectrum. In the age of discrete signals, this approach is rarely used anymore, instead being replaced by the discrete FT, which is the digital counterpart of the analog spectrum analyzer. FTs are discussed in more detail in section 6 but, because of their importance to signal processing, they are introduced here. All of the information conveyed in the original signal is carried in the FT, but it is represented in the frequency domain rather than the time domain. This allows the composition of the signal in terms of sinusoidal frequencies to be analyzed, providing more direct information about the correlations in the time domain. Slowly varying signals in the time domain will have significant low-frequency components, and the shape of the power spectrum conveys important information about optimal signal processing methods. Figure 1 shows some simple signals and their FTs. The first example shows that a pure sinusoid gives rise to a single peak in the frequency domain. In contrast, the sharp edges of the square wave in the second example lead to high-frequency components in the FT. The last two examples demonstrate that more slowly varying signals have fewer high frequency components.

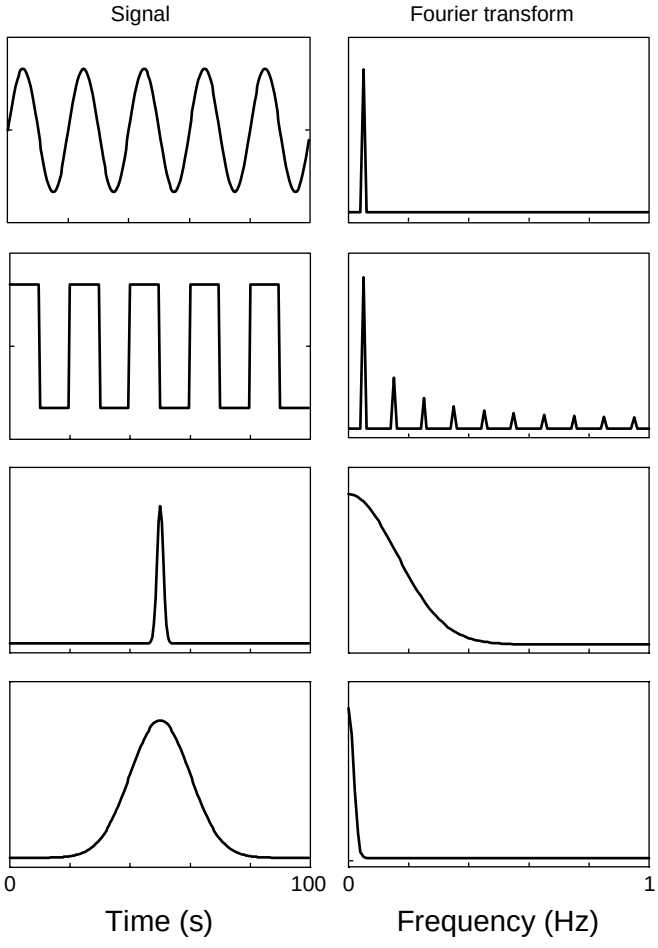

Figure 1 FTs of some simple signals.

3.2 Noise

The noise in an analytical signal can be classified in a number of ways, including (a) its distribution, (b) its source, (c) its characteristics in the time domain, and (d) its characteristics in the frequency domain. Because different classifications are used in different situations and they are not all mutually exclusive, it is necessary to understand the different cases and how they relate to signal processing. In doing so, it is helpful to imagine a signal from which we could subtract the pure signal component leaving only the noise or, alternatively, a situation in which the pure signal is zero so that we are measuring only noise, as shown in Figure 2(a).

If we were to plot a histogram of the magnitude of the noise for a large number of measurements, we might obtain a distribution such as that shown in Figure 2(b). By far the most common noise distribution (assumed or measured) for analytical measurements is the normal, or Gaussian, distribution, shown by the solid line in the figure. The reason for this is the central limit theorem which, simply put, states that if a measurement is the sum of a series of values drawn from arbitrary distributions, the distribution of the measurement will approach a normal distribution as the length of the series approaches infinity. As, in an analytical instrument, the observed noise is a consequence of many smaller random events, the central limit theorem can be rationalized to hold. Other noise distributions (e.g. uniform, log-normal) are also observed but are much less common. One other type which is common, however, is the Poisson distribution, which is observed in cases where the signal arises from

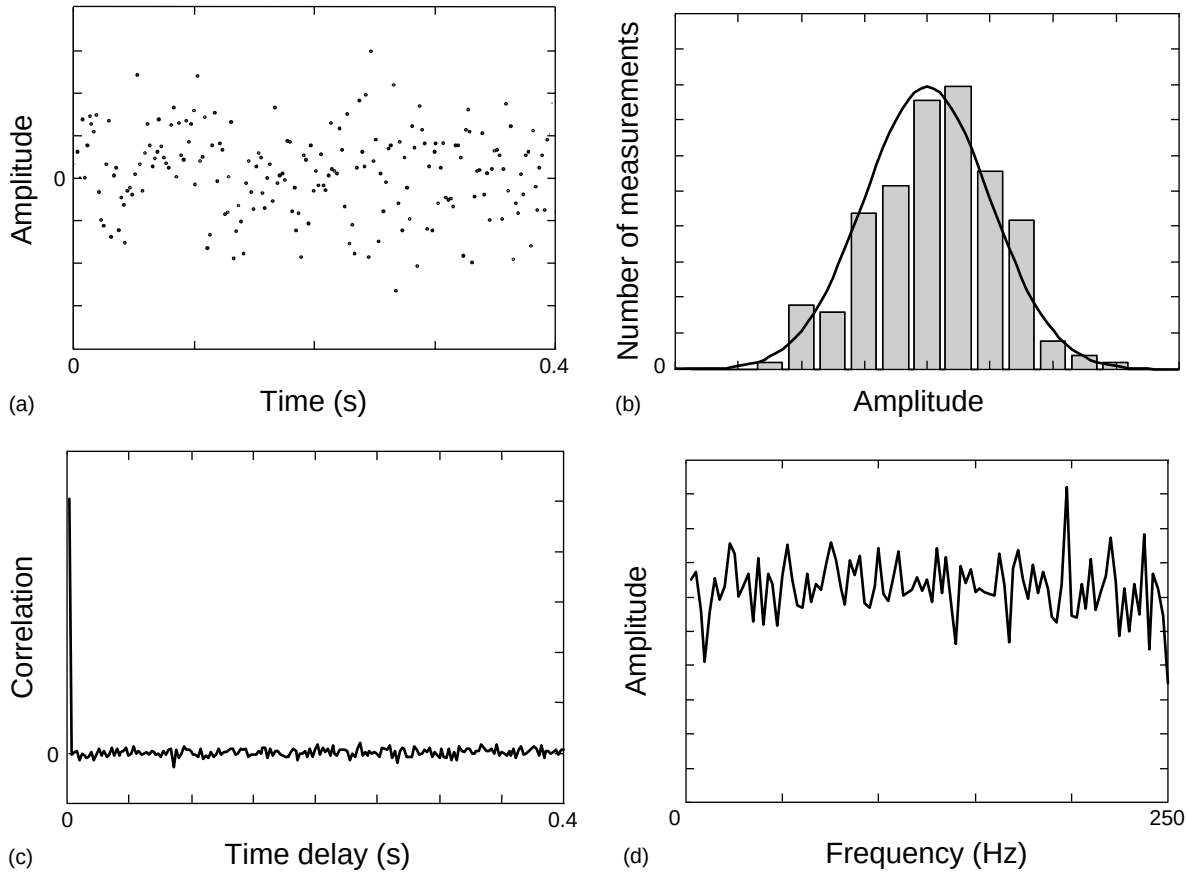

Figure 2 Representations of white noise: (a) noise sequence in the time domain, (b) distribution, (c) autocorrelation function, (d) FT.

a collection of discrete events, such as photons striking a photomultiplier tube. However, in reality a histogram constructed from Poisson noise would look essentially the same as the normal distribution in most cases. The distinction is in how the magnitude of the noise (i.e. its standard deviation) changes with signal intensity. The standard deviation of Poisson noise will increase with the square root of the signal intensity. To say that noise has a normal distribution, however, does not imply anything about how its magnitude changes with signal intensity. In this sense, the Poisson distribution can be regarded as a special case of the normal distribution.

In some cases noise is classified according to its dominant source. Such classifications often imply information about the distribution or temporal characteristics of the noise. For example shot noise, also called Schottky noise or quantum noise, arises in detectors based on discrete events, such as photons striking a photomultiplier, and exhibits a Poisson distribution. Shot noise is a type of fundamental noise because it originates from the random statistical nature of the events themselves and not from any deficiencies of the instrument. This type of noise dominates in cases where the number of events is relatively small, such as in fluorescence measurements. Johnson noise is another type of fundamental noise that arises from the random thermal motion of electrons in resistors. Flicker noise is considered to be a type of non-fundamental or excess noise in which the magnitude of the noise is directly proportional to the signal amplitude, hence it is often referred to as proportional or multiplicative noise. Flicker noise is often associated with variations in source intensity in absorption spectroscopy and can have distinctive frequency characteristics (see below). Other types of noise include interference noise (electrical, optical or other interferences that arise at specific frequencies, such as 60 Hz line interference), detector noise (a general term referring to instruments in which the limiting noise, such as shot noise or thermal noise, occurs at the detector), amplifier-readout noise (the noise observed for the readout circuitry when the input signal is zero) and quantization noise (observed when measurement precision is limited by the A/D conversion step).

In the time domain, noise can be classified in two ways: correlated and uncorrelated. Uncorrelated, or independent, noise implies that the noise observed at one point in the series of measurements is not related

in any way to that at other points. If we consider the sequence of noise to be represented by the vector [e_1 e_2 e_3 ...], then the covariance between the first two noise elements can be represented by Equation (3),

$$\sigma_{12} = \mathrm{E}(e_1 e_2) \tag{3}$$

and will be zero for uncorrelated noise. In many applications, independent and identically distributed noise with a normal distribution, or *iid normal* noise, is assumed, implying uncorrelated noise with a Gaussian distribution and equal variances at all channels. However, in reality correlated noise is commonplace, arising from varied sources such as temporal variations in spectroscopic source intensities, spatial correlations (cross-talk) in array detectors, thermal variations and electronic filtering. Signal processing itself can turn uncorrelated noise into correlated noise, a fact that can be important for subsequent data analysis methods. The correlation of a noise sequence can be examined through its autocorrelation function. This is obtained through an element-by-element multiplication of the noise sequence by a time shifted version of itself and averaging the results of the n multiplications. This is repeated for each time shift. Uncorrelated noise should give a single spike at zero time delay, with the products averaging to zero everywhere else. This is demonstrated in Figure 2(c).

The complete characterization of the correlation among elements of a noise vector of length n is given by the error covariance matrix Σ (Equation 4):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix} \tag{4}$$

where the diagonal elements represent the noise variances and the off-diagonal elements (zero for uncorrelated noise) represent the covariance. If a noise sequence is a stationary process, then its statistical properties remain constant throughout the sequence, which means $\sigma_{12} = \sigma_{23} = \sigma_{34}$ and so on. The terms 'homoscedastic' and 'heteroscedastic' are also used to indicate whether the variance of the noise remains constant or changes with the position in the series, respectively.

As with the signal, the characteristics of the noise can be examined in the frequency domain by using a spectrum analyzer or the FT. The result is called the noise power spectrum (NPS) and it conveys important information about the time domain correlations of the noise. Uncorrelated noise is referred to as white noise and, analogous to white light, it contains equal contributions at all frequencies. This is the type of noise that is often assumed or hoped for when designing or implementing signal processing methods, but exceptions are very common. The NPS for the white noise in Figure 2(a) is shown in Figure 2(d). Note that because of the stochastic nature of noise, the NPS is not perfectly flat.

Often purely white noise is corrupted by pink noise or $1/f$ noise. This noise, also known as drift in the time domain and arising from flicker noise, has a NPS which varies as the reciprocal of the intensity. The dominance at low frequencies is indicative of correlated noise in the time domain. The third type of noise readily identified in the NPS is interference noise, which appears as spikes at the corresponding frequencies and often higher harmonics. Figure 3 illustrates a typical correlated noise sequence and its representative autocorrelation function and NPS. Note the slow decay of the autocorrelation function indicating that the noise is not white. Both $1/f$ noise and interference noise are apparent in the NPS. It should be noted that, in order to obtain a clear representation of the autocorrelation function and the NPS in this case, it was necessary to average results from 20 noise sequences, because a single noise sampling does not usually give a clear indication of its characteristics.

3.3 Signal Averaging

One of the most effective ways to separate the pure signal from the noise is through the use of signal averaging. True signal averaging, as opposed to boxcar averaging or smoothing (section 5), requires a repetitive signal sequence. A common example is the fluorescence or phosphorescence decay curve generated by pulsing a laser or flash lamp repetitively. Each experiment (pulse) results in a signal that can be added to the next, and the total sum can be averaged over the number of experiments. Unlike most of the methods described here, which take advantage of signal correlation within a single experiment, signal averaging exploits signal correlation at a given time channel for repeated experiments. For this reason, it is discussed here, separately from the other techniques.

In order for signal averaging to be effective and useful, a number of conditions need to be met. First, the noise should be uncorrelated between corresponding time channels for successive experiments or else it will not be effectively removed. (Note that correlation of noise within an experiment is not important.) A second requirement is that the shape of the signal needs to be truly repetitive in the time domain. Translation of the signal or changes in its profile (other than scaling) due to poor synchronization or changes in the process will lead to an unrepresentative average, although noise should be reduced. Finally, the duration of the experiment needs to be short enough to make signal averaging practical.

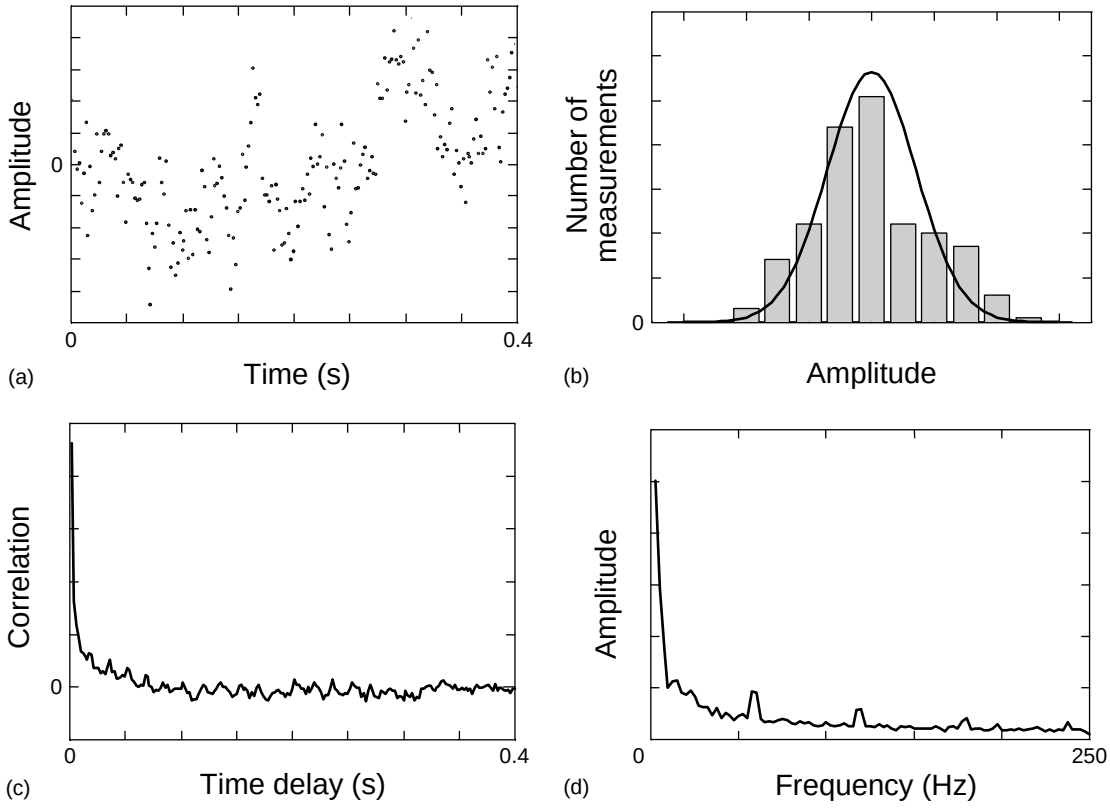

Figure 3 Representations of correlated noise: (a) noise sequence in the time domain, (b) distribution, (c) autocorrelation function, (d) FT.

When these conditions are met, signal averaging can be particularly advantageous because: (a) the signal-to-noise ratio (SNR) improves by a factor of $\sqrt{n}$, where n is the number of repetitions, and (b) there is no distortion in the profile of the signal. In contrast, most of the methods discussed in the sections that follow have the potential to distort the shape of the signal.

4 ANALOG SIGNAL PROCESSING

4.1 Overview

As virtually all instrumental methods involve some form of electrical transduction of a particular phenomenon into a continuous signal, analog signal processing is as universal as it is diverse. Such processing begins the moment the quantity being measured is converted into some electrical property such as current or voltage (assuming that it did not originate in that form) and continues until the final measurements are recorded in digital or analog form. This obviously opens up a tremendous range of topics, the detailed coverage of which is beyond the scope of this article. However, there has been a gradual shift over the years which has placed an increasing emphasis on digital signal processing over analog signal processing. One reason for this is that access to the analog signal in modern instruments has become more restricted and more often the chemist is presented with data that have already been digitized, as evidenced by the demise of the chart recorder from most analytical labs. A second reason is that improvements in the speed and storage capacity of digital components have removed many of the limitations of early devices. Finally, digital processing of results has the advantage that it can be carried out any time after the data are required, whereas this is not true for analog signal processing.

On this basis, it could be argued that a detailed comprehension of analog signal processing is less important now than it once was, although it is still essential to understand the basic capabilities and limitations of this stage of processing, because it will always precede the generation of digital information and can be the 'weak link in the chain' if care is not taken. For this reason, a brief coverage of electronic signal processing methods is presented here, with a special emphasis on two aspects which are closely related to digital processing, analog filters and A/D conversion.

The electronic manipulation of signals can be divided into methods which are based on passive or active

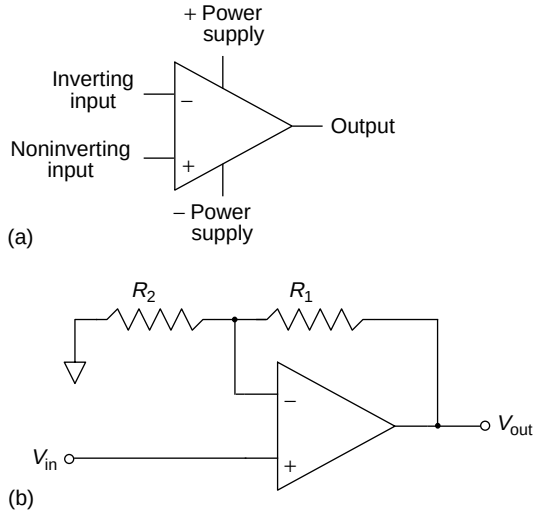

Figure 4 (a) Symbolic representation of an operational amplifier. (b) A noninverting amplifier with gain.

circuits. Passive circuits (those consisting only of simple components such as resistors, capacitors or diodes) do not require an external power source (other than the signal itself) whereas active circuits do. Although passive circuits are normally much simpler than active circuits, their capabilities are severely limited by comparison. Most complex manipulations of electronic signals are based on operational amplifiers, active circuit elements employed for a wide variety of linear and nonlinear operations. A basic operational amplifier, represented symbolically in Figure 4(a), is essentially a high gain (typically 10^5) differential amplifier. By taking advantage of the high gain in negative feedback, the two inputs (referred to as the inverting (−) and noninverting (+) inputs) and the output can be configured into a wide range of useful circuits using passive components. As an example, the configuration for a simple fixed-gain voltage amplifier (gain $= V_{out}/V_{in} = (R_1 + R_2)/R_2$) is shown in Figure 4(b).

The number of electronic operations that can be carried out on electrical signals is too extensive to cover here, but a brief summary of some important operations is given in Table 1. More details on analog signal processing can be found in appropriate references on the subject.[3–5] Analog filters and A/D conversion are also covered in more detail in sections 4.2 and 4.3.

4.2 Analog Filters

One of the simplest operations that can be carried out to improve signal quality is the application of an electronic filter. As the object of signal processing is to distinguish the pure signal from the noise, a means

Table 1 Summary of some common analog signal processing applications

Operation or circuit	Description
Domain conversion	Conversion of an analog signal between domains, such as current-to-voltage, resistance-to-voltage, time-to-amplitude, A/D.
Amplification	Multiplication of an analog signal by a constant factor called the gain. Sometimes coupled with domain conversion.
Inversion	Changes the sign of an analog signal.
Addition/subtraction	Two or more signals are added/subtracted.
Multiplication/division	Two signals are multiplied/divided by one another. Often used with modulation (see below).
Other mathematical operations	Includes logarithm, antilogarithm, absolute value, reciprocal, etc.
Integration/differentiation	Integrals or derivatives of analog signals.
Modulation	Process by which the property (e.g. amplitude or frequency) of a carrier wave, typically a high frequency sinusoid or square wave, is modified to convey information about an analog signal of interest. Demodulation is used to recover the original analog signal.
Comparator	Circuit that compares two voltages and produces one of two outputs, depending on which signal is larger.
Pulse height discriminator	Circuit that detects the presence of pulses with a peak amplitude within a certain threshold region.
Peak detector	Circuit that follows an analog signal until it reaches a peak value and then holds that value for a certain time period.
Sample-and-hold amplifier	Circuit that samples an analog voltage at a particular point in time and holds the value until further processing, normally A/D conversion, is completed.
Boxcar integrator	Circuit used to measure a rapidly changing but repetitive signal by sampling it after a particular delay time and holding the sampled measurement. Often the delay time is scanned to obtain the time profile of the signal.
Lock-in amplifier	One type of circuit used to demodulate analog signals (see above).

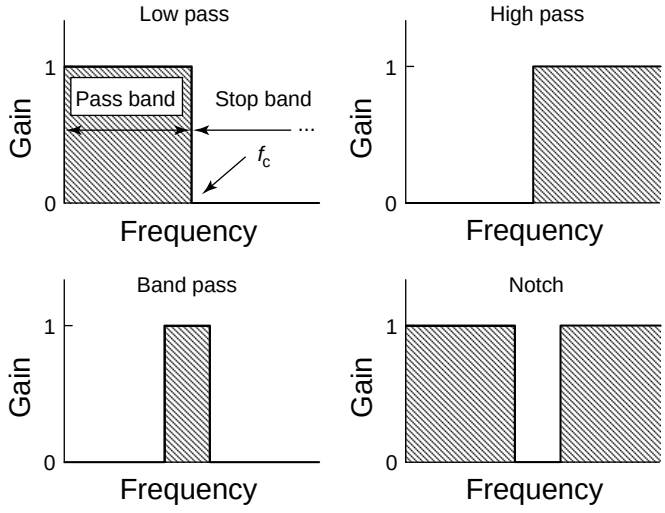

Figure 5 Transfer functions for ideal filters of various types.

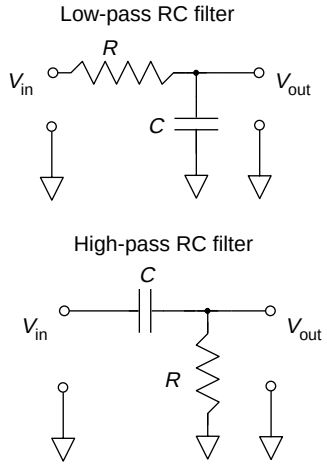

Figure 6 Simple passive RC low-pass and high-pass filters.

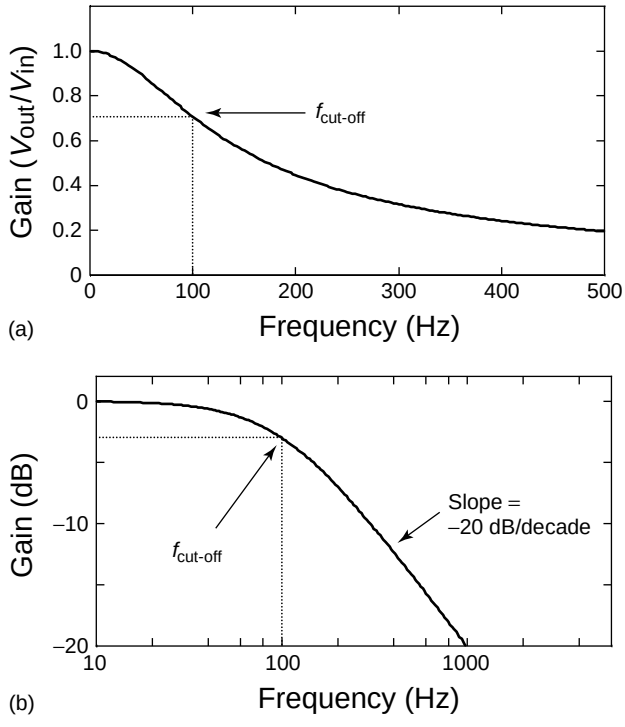

Figure 7 (a) Transfer function and (b) Bode plot for a simple low-pass RC filter with a cut-off frequency of 100 Hz.

is generally sought to distinguish the two. As noted in section 3, this distinction can often be made in the frequency domain. Typically, white noise has a flat NPS, whereas a slowly varying signal will have most of the information at low frequencies. Therefore, a filter that removes high-frequency components from a noisy signal, called a low-pass filter, will retain most of the information about the pure signal while eliminating much of the noise. However, drift and offset noise (low frequency) may be dominant features in the noise, and a high-pass filter may be more effective, provided there are sufficient high-frequency components in the signal. In another scenario, the signal may be modulated at a particular frequency and it may be necessary to use a band pass filter to isolate the components of interest. Finally, if interference noise, such as 60 Hz noise, is a problem, it may be removed using a notch filter. The ideal transfer functions for each of these types of filter is shown in Figure 5. The transfer function gives the amplification or gain of the filter as a function of the frequency of the input and ideally should be unity within the pass band and zero in the stop band.

As shown in section 5, analog filters bear many similarities to digital filters, but it is important to understand the former for several reasons. First, analog filters evolved before digital filters and there is substantial overlap of terminology. Second, although digital filters are becoming more widely used, analog filters are often a more effective way to eliminate noise near the source and are essential to limit the noise bandwidth in any digital data acquisition system. Finally, although they share similarities in their characteristics, transfer functions for digital and analog filters are significantly different.

Analog filters can be classified as either active or passive. Passive filters use only resistive (i.e. resistors) and reactive (i.e. capacitors and inductors) components, whereas active filters employ operational amplifiers as well. Figure 6 shows examples of simple low-pass and high-pass resistor–capacitor (RC) filters. In principle, filters could also be constructed using resistors and inductors, but inductors tend to be more bulky, expensive and less ideal, so in practice capacitors are more commonly used. The transfer function for a simple low-pass filter is shown in Figure 7(a). The gain is given by the ratio of the capacitive reactance to the total impedance

(Equation 5):

$$\text{gain} = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{X_{\text{C}}}{Z} = \frac{1/(2\pi fC)}{\sqrt{R^2 + 1/(2\pi fC)^2}}$$
$$= \frac{1}{\sqrt{(2\pi fRC)^2 + 1}} \quad (5)$$

Here X_{C} is the capacitive reactance, Z is the total impedance of the circuit, f is the frequency in hertz, R is the resistance in ohms, and C is the capacitance in farads. Note that the voltages referred to in this equation are not instantaneous voltages, but rather peak or rms values for inputs of a fixed frequency, as analog filters usually impose a phase shift on the original signal. More commonly, the frequency response of a filter is shown with a Bode plot, in which logarithmic scales are used on both axes. The voltage gain on the vertical axis is normally expressed in decibels (dB), given by Equation (6):

$$\text{gain(dB)} = 20 \log \left(\frac{V_{\text{out}}}{V_{\text{in}}} \right) \quad (6)$$

A Bode plot for a simple low-pass filter is shown in Figure 7(b). The frequency cut-off is usually taken to be the point at which the capacitive reactance equals the resistance, i.e. $V_{\text{out}}/V_{\text{in}} = 1/\sqrt{2}$. This corresponds to a gain of -3.01 dB, and so the operational cut-off frequency is usually referred to as the '3 dB point'. The cut-off frequency depends on the product RC and, for the simple low pass filter shown in Figure 6, this is given by Equation (7):

$$f_{\text{cut-off}} = \frac{1}{2\pi RC} \quad (7)$$

After the cut-off frequency, the Bode plot shows a linear region which has a slope of -20 dB per decade, more gradual than the ideal case shown in Figure 5. In order to provide a steeper slope, the order of a filter has to be increased. The filter order, also referred to as the number of poles, is the number of reactive components required for each cut-off frequency. Therefore, a first-order low- or high-pass filter such as the ones above requires only one capacitor, whereas a band-pass filter would require two. The higher the order of the filter, the more one can approach the ideal filter with a sharp transition between the pass band and the stop band. Practical limitations related to loading and other factors restrict the number of poles that can be used with passive filters, however, and active filters are normally used when more sophisticated filters are required.

The use of operational amplifiers in active filters allows greater flexibility in filter design. The design of active filters is well beyond the scope of this article, but fortunately, in those cases where a chemist may wish to incorporate an active filter, standard designs have been developed.[3–7] An active filter is classified not only according to its order, but also according to the way it optimizes a variety of other parameters, such as the steepness of the transition region, the flatness of the pass band, and its phase characteristics. Two of the most common types are the Butterworth, or maximally flat, filter and the Chebyshev, or equal ripple, filter. As shown in Figure 8, the frequency response of the Butterworth filter is very flat in the pass band, but its transition region is not as steep as the Chebyshev filter, which exhibits ripples in the pass band. The design for this active filter is shown in Figure 9. Note that the difference between the two filter types lies simply in the selection of the components and that the ripple can always be decreased at the expense of sharpness. The implementation of these filters is made even simpler through the use of integrated circuits incorporating switched capacitor filters that allow the cut-off frequency to be set simply with a reference clock signal.

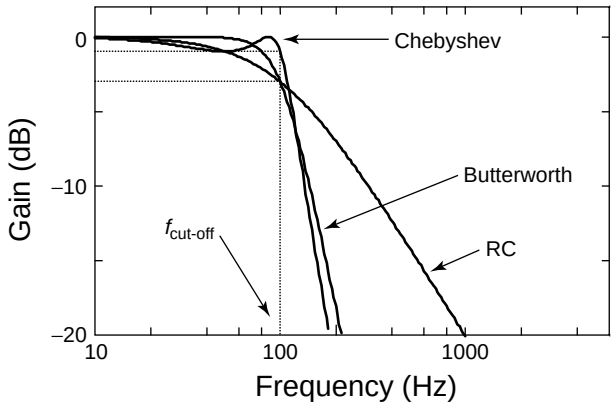

Figure 8 Frequency response (Bode plot) for passive (RC) and active (Butterworth and Chebyshev) filters.

From a digital signal acquisition perspective, perhaps the most important kind of filter is the anti-aliasing filter. When a signal is converted from the analog domain into the digital domain, it must be sampled at discrete and normally equally spaced time intervals, Δt. The sampling frequency is given by $f_{\text{s}} = 1/\Delta t$, and restricts the

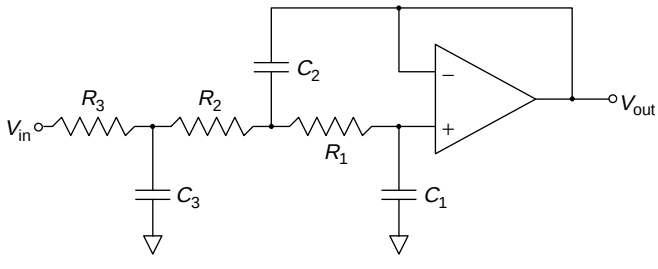

Figure 9 A third-order active filter.

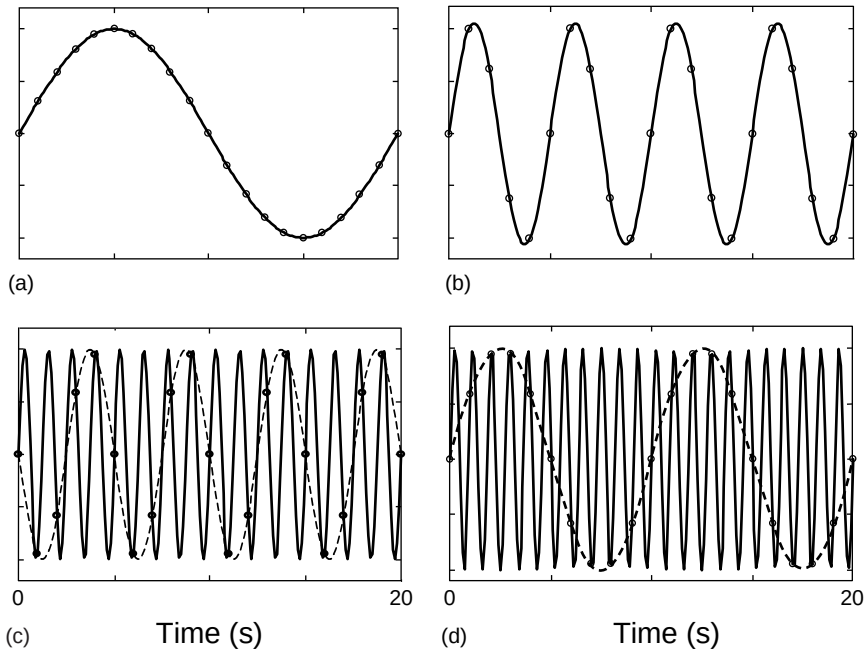

Figure 10 An illustration of aliasing in which open circles indicate sampled points. Parts (a) and (b) represent signals below the Nyquist frequency which are not aliased. The solid lines in (c) and (d) represent signals above the Nyquist frequency which are aliased to lower frequencies (dashed lines).

upper limit of signal frequencies that can be accurately represented in the transition from A/D information. More specifically, any components of the signal with frequencies above the Nyquist frequency will have their components aliased to lower frequencies. The Nyquist frequency, f_N, is given by Equation (8):

$$f_N = \frac{f_s}{2} = \frac{1}{2\Delta t} \tag{8}$$

The phenomenon of aliasing is shown in Figure 10, where four different sinusoids are sampled at a frequency of 1 Hz (i.e. $f_N = 0.5$ Hz). The first two signals shown are below the Nyquist frequency and the sampling reflects their variations accurately. However, the last two signals are above the Nyquist frequency and aliased back to lower frequencies (Figure 10).

The importance of an anti-aliasing filter, which is simply a low-pass filter with a cut-off near the Nyquist frequency, has to do with the noise bandwidth of the analog signal. If noise components are present above the Nyquist frequency and no filtering is applied, this noise will be aliased back to lower frequencies and appear in the acquired signal. As no information at frequencies higher than the Nyquist frequency can be accurately extracted anyway, it is wise to always use an anti-aliasing filter to reduce noise in a digitally acquired signal. Also note that this noise can only be removed in the analog domain and digital filtering cannot help once the noise is aliased to lower frequencies.

4.3 Analog-to-digital Conversion

A/D conversion refers to the process by which a continuously variable analog signal, usually a voltage, is converted into a discrete numerical value with a fixed precision. In most modern analytical instruments, A/D conversion is a key step in the signal processing sequence. Although this process is generally transparent to the user, an understanding of the principles involved can be useful in practice.

As digital logic circuits are based on binary states, the digital representation of a measurement is made in the binary, or base 2, number system, consisting of a series of binary digits, or bits. Thus the number 27 in base 10 would be represented by the following 8 bits, or byte, in base 2:

$$\begin{aligned} 27_{10} &= 0001\,1011_2 \\ &= 1(2^4) + 1(2^3) + 1(2^1) + 1(2^0) \end{aligned}$$

Alternatively, binary coded decimal (BCD) can be used in which each series of 4 bits (sometimes called a nibble) represents a decimal digit:

$$27_{10} = 0010\,0111\ (\text{BCD})$$

For n bits, a binary representation gives the numbers from 0 to $2^n - 1$, whereas a BCD representation gives a smaller range, from 0 to $10^{n/4} - 1$. To include the sign with a binary number several strategies can be employed.

In the offset binary notation, a value of 2^{n-1} is subtracted from the binary number to give the signed result. In the sign magnitude notation, the most significant bit (MSB) (the left-most bit) is used to represent the sign (0 = positive, 1 = negative). Finally, the most practical representation from a mathematical point of view is the 2's complement representation. In this case, a number is negated by inverting the original bits and adding 1. The 8-bit representations of the number −27 using each of these notations is given below.

$$-27_{10} = 0110\,0101 \text{ (offset binary)}$$
$$= 1001\,1011 \text{ (sign magnitude binary)}$$
$$= 1110\,0101 \text{ (2's complement binary)}$$

A/D conversion is facilitated using a largely self-contained circuit, the ADC. In addition to binary coding issues, important ADC parameters from a signal-processing perspective include precision, accuracy, linearity, monotonicity, and speed. The precision of an ADC is directly related to the number of bits in the digital output, but increased precision also means decreased conversion speed for a given type of ADC. Typically, 8-bit converters are used in applications where speed is more critical than precision, whereas precision applications can use as many as 16 or 20 bits. For many scientific applications, a 12-bit ADC is a good compromise, with a precision of 1 part in 4095, or 0.02%. It is important that the precision of the ADC is better than the standard deviation of the noise in the measurement, or else the dominant source of noise will be the quantization (or digitization) noise arising from rounding of the result. As the quantization noise is always fixed at a value corresponding to the least significant bit (LSB), its relative contribution increases for small values. In some cases, this problem can be addressed through autoranging, in which small signals are amplified in the analog domain prior to conversion. This is most important when the amplitude of the noise increases with the signal.

The accuracy of an ADC refers to the closeness of the converted value to the expected value based on the range of the ADC and the reference voltage. Linearity is an indication of the constancy of the proportionality of the digital output to the analog voltage over the full range of conversion, and in some cases may be more important than the actual accuracy. In a plot of the digital output versus the analog input, linearity is often specified as the maximum deviation from the straight line drawn between zero and the full scale output, or alternatively the best-fit straight line. Monotonicity is a specification that requires an increasing analog input to give an increasing digital output with no missing codes over the full range, and likewise for decreasing inputs. These specifications (accuracy, linearity, and monotonicity) are primarily a function of the ADC design and the quality of the components used.

In addition to being discrete in the measurement domain, analog signals are also discrete in the time domain, and the required conversion speed, or sampling frequency, will depend on the application. It is essential that all of the important signal characteristics in the frequency domain fall below the Nyquist frequency, f_N (section 4.2). The speed of an ADC is primarily a function of the type of converter used, although it is also dependent on factors such as the number of bits and clock frequency. Although the design of new ADCs is an ongoing process driven by consumer electronics, most are variations on five basic types: (1) parallel, (2) tracking, (3) successive approximation, (4) integration, and (5) voltage-to-frequency (V/F) conversion.

The fastest and conceptually simplest ADC is the parallel ADC, often referred to as a flash converter. In this circuit, the input voltage is simultaneously compared with 2^n reference values and logic circuits use the closest match to produce the digital output. This brute-force approach can provide conversion speeds under 10 ns, but is not component efficient, requiring 2^n voltage comparators. As a consequence, this type of ADC is usually expensive and limited to 8 bits, although hybrid circuits referred to as half-flash converters, can increase the precision.

Other relatively fast circuits are based on the use of a digital-to-analog converter (DAC) to provide an analog voltage for comparison with the input voltage. The digital value passed to the DAC is systematically changed until its analog output matches as closely as possible the input voltage. At that point, the digital input to the DAC is taken to be the digital output of the ADC. Differences exist in the way that these devices change the digital values for comparison. In a tracking ADC, a simple counter is used to increment or decrement the digital value until a match is found. A block diagram for a simple tracking ADC is shown in Figure 11. Depending on the change in the voltage, the conversion time for this type of ADC could range from 0 to 2^n clock cycles for an n-bit converter. The successive approximation ADC operates on the same basic principle as the tracking ADC, but has a fixed conversion time of only n clock cycles. This is accomplished by replacing the counter with circuitry that uses an efficient binary search algorithm to split the digital range in two for each comparison. Successive approximation ADCs have typical conversion times ranging from 1 to 100 μs and are among the most common converters in use.

The remaining two types of ADCs are relatively slow by comparison. Integrating ADCs are based on the use of a

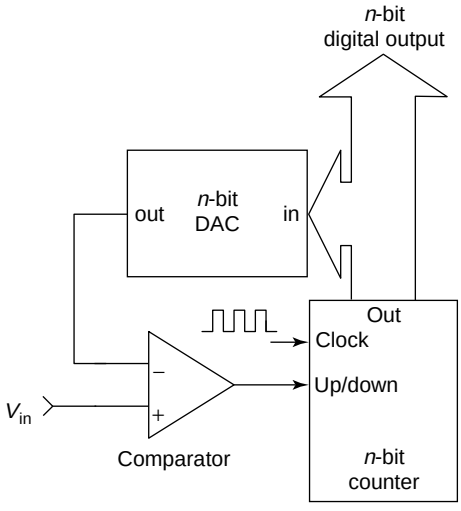

Figure 11 Block diagram of a simple tracking ADC.

fixed current to charge a capacitor through a resistor. The digital result is obtained by timing the period necessary for the capacitor to reach the input voltage. This basic strategy, termed a single slope integrating ADC, suffers from a dependence on accurate and stable circuit components and has been replaced by the dual slope (or even quad slope) integrating ADC. The dual slope ADC is also based on charge integration, but uses a charge–discharge cycle to cancel the effects of component variations. This type of ADC is known for its accuracy, stability, monotonicity, low cost and noise rejection characteristics and has been used in a large number of precision applications where speed is not critical. Typical conversion rates are around 10 conversions per second. In many applications, the dual slope ADC has been replaced by the V/F converter. This device simply produces a series of pulses whose frequency is proportional to the input voltage applied, with typical maximum frequencies in the range from 10 kHz to 1 MHz. To complete the conversion to a binary number, a counter is attached to the pulsed output and counted for a fixed period of time. Precision can be improved by counting for a longer period. In addition to low cost, simplicity, and good linearity, the V/F converter has the advantage that its output can be transmitted in serial over a single line, simplifying remote data acquisition. Like the integrating ADCs, however, this type of ADC is relatively slow.

Although the analyst typically has very little control over the type of A/D conversion that is used in a particular instrument, this is the first stage of all subsequent digital signal processing, so it is important to recognize the strengths and weaknesses of these devices. Further details on the design and application can be found in a number of references.[3–5,7–8]

5 DIGITAL FILTERING

5.1 Introduction

Because digital filters are among the most widely used methods for signal processing in analytical chemistry, much of this article is dedicated to describing their implementation and operation. The term 'filter' is a reference to the similarities they share with their electronic counterparts. In both cases, the data are presented to the filter in a sequential fashion and the distinction between pure signal and noise is often made on the basis of differences in power spectra. Strictly speaking, however, a filter uses only information from the past up to and including the current point in obtaining an estimate of the current point. Although this is true for electronic filters, it does not hold for most digital filters, which use points before and after the measurement of interest to form an estimate. Thus, they may be more properly classified as smoothers or smoothing filters, but all three terms are used in the literature. The term 'filter' is used in a general way throughout this section, incorporating both smoothers and other types of filters.

Digital filtering can be performed either in real-time or in a post-acquisition mode. The advantage of the former is that it can be transparent to the user and optimized at the time of instrument design. Fast digital signal processors (DSPs) are available for performing common signal processing operations and can be built into the instrument itself. A disadvantage of this approach is that it removes some flexibility and may obscure some of the features of the original data. For these reasons, real-time digital filtering is often kept to a minimum and much signal processing is carried out in post-acquisition mode. Therefore, it is important to appreciate the advantages and limitations of different types of digital filters. A complete coverage of the subject of digital filters is well beyond the scope of this article. The objective here is to present some of the terminology and describe some of the digital filters commonly used in analytical chemistry. For readers seeking more information on the subject, there is abundant literature available.[9–19]

5.2 Filter Types

In general terms, a digital filter could be defined simply as an operation that is carried out on a contiguous subset of the original signal sequence to produce an estimate of a value in the filtered signal sequence. This is illustrated in Figure 12. For conventional digital filters, the filtering operation consists of the convolution of a series of filter coefficients with the signal, but this approach is by no means universal. Digital filters are therefore classified

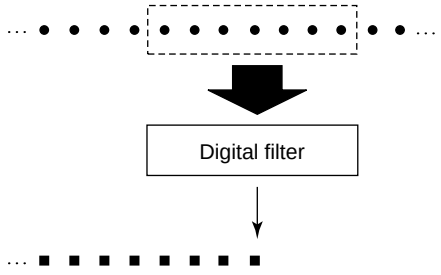

Figure 12 General operation of a digital filter.

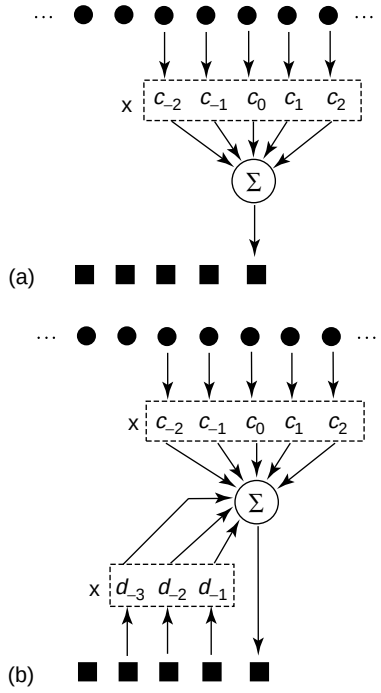

Figure 13 Operation of (a) nonrecursive and (b) recursive digital filters.

in a number of ways according to the manner of their operation.

The most common type of digital filter employed in analytical chemistry is the nonrecursive filter, also referred to as the finite impulse response (FIR) filter because its response to an impulse (delta) function will always fall to zero at some point in time. As illustrated in Figure 13(a), nonrecursive filters use the conventional approach of convoluting a set of filter coefficients with the sequence of measurements to produce the filtered signal. If z_i represents the filtered value for measurement i, it is determined mathematically by Equation (9):

$$z_i = \sum_{j=p}^{q} c_j y_{i+j} \tag{9}$$

where the c terms are the filter coefficients and the y terms are the original measurements. Most often, the filter coefficients are arranged symmetrically around the point to be estimated so that $p = -q$, but this is not a requirement. Nonrecursive filters have good stability and are relatively easy to design. A simple example of a nonrecursive filter is a five-point moving average filter, which averages the five points around a central value (two on either side plus the point itself) to obtain its estimate. The coefficients in this case would be $c_{-2} = c_{-1} = c_0 = c_1 = c_2 = 0.2$.

Recursive filters differ from nonrecursive filters in that they make use of previously filtered values to estimate the current measurement, as shown in Figure 13(b). Mathematically, this can be represented by Equation (10),

$$z_i = \sum_{j=p}^{q} c_j y_{i+j} + \sum_{k=r}^{s} d_k z_{i+k} \tag{10}$$

where z represents the filtered measurement, y is the unfiltered measurement, and c and d are the filter coefficients. Note that the indices r and s must be less than zero, as the filter coefficients d_k can only be applied to previously filtered values. In practice, many nonrecursive filters are designed to function in real time and so also limit the maximum value of q to be zero to make them physically realizable (i.e. they do not make use of future values). A simple example of a recursive digital filter is integration using the trapezoid rule, which gives (Equation 11):

$$z_i = \tfrac{1}{2}(y_{i-1} + y_i) + z_{i-1} \tag{11}$$

This corresponds to coefficients $c_{-1} = 0.5$, $c_0 = 0.5$ and $d_{-1} = 1$. Recursive digital filters are also known as infinite impulse response (IIR) filters because it is possible for a single impulse input to influence filter output values indefinitely, as illustrated in this example. These filters can be used for normal smoothing operations in addition to integration, and have the advantage that they provide more efficient filters with fewer coefficients than nonrecursive filters. However, recursive filters are more difficult to design and have more complex properties than their nonrecursive counterparts, so they have not been widely used in analytical chemistry.

It is not required that digital filter coefficients remain constant throughout a sequence of measurements. In some cases, the coefficients may change in response to variations in another external input or the signal itself. These kinds of filters are called adaptive filters. Such filters are used, for example, in removing noise from audio signals using external measurements of noise characteristics. Chemical applications of this type of filter are rare, although digital processing of modulated signals, such as modulated sources in atomic absorption

spectroscopy, could be considered to be a simple form of adaptive filtering.

Filters can also be classified according to the types of results they are intended to produce. Smoothing filters are intended to reproduce the pure signal and suppress the noise, whereas derivative filters are intended to estimate the derivative (first, second, or higher) of the pure signal. In other cases, one may wish to increase the apparent sampling frequency of a set of data, filling in points for esthetic or practical reasons (e.g. locating a peak maximum). This can be done using an interpolation filter, although care should be taken when applying such approaches. In contrast, in cases where the signal is highly oversampled, one may wish to reduce the volume of data while improving its quality by using a decimation filter.

Some types of digital filters do not follow the usual pattern of smooth convolution of a set of filter coefficients with the data, but are considered digital filters nevertheless. In boxcar averaging, for example, the signal is divided into subsets of n measurements which are averaged to produce a single result, as shown in Figure 14. As this reduces the total number of points, it is one type of decimation filter. Another rarely used but very useful filter is the median filter. Although this filter should be used with care because of its nonlinear transformation of the data, it is particularly effective at removing spikes or outliers from a measurement sequence. These outliers may arise, for example, from cosmic rays striking a photodetector or bubbles passing through a detector flow-cell. The median filter works by sorting the data within a window of length n and choosing the median value as the filtered estimate, thus automatically eliminating outliers unless they occur in clusters. An example of the application of the median filter is shown in Figure 15.

Because of their dominance in chemical applications the main emphasis of this article is on nonrecursive digital filters, particularly polynomial least-squares filters, although some discussion of Kalman filters is also presented.

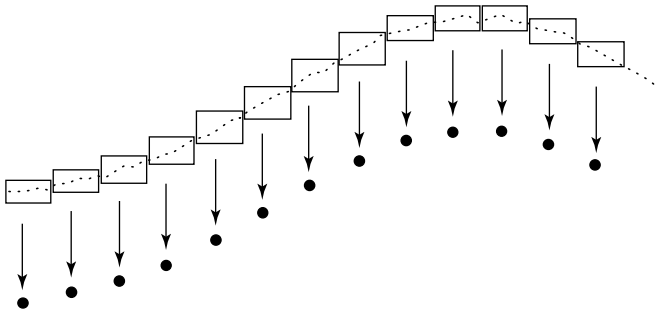

Figure 14 Illustration of boxcar averaging.

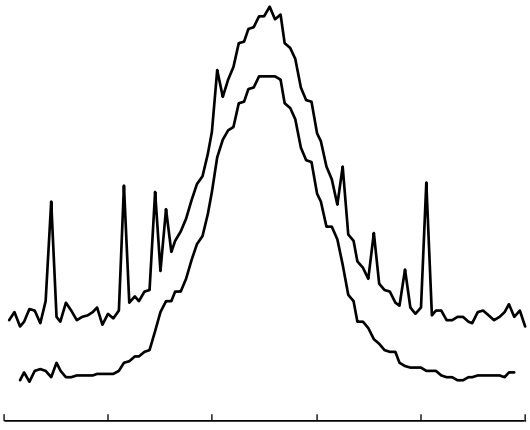

Figure 15 A noisy peak (upper curve) before and (lower curve) after filtering with a median filter. Note the removal of spikes.

5.3 Polynomial Least-squares Smoothing Filters

By far the most widely applied digital filter in analytical chemistry is the polynomial least-squares smoothing filter. These filters are more commonly known to analytical chemists as SG filters, a reference to their introduction into the analytical chemistry literature by Savitzky and Golay in 1964.[1] Although these filters were known in the field of signal processing prior to this, the SG paper made their utility known to chemists at a time when the digital acquisition of signals was becoming more commonplace. In addition, the paper presented tables of precalculated coefficients for different types of filters. At the time, computational efficiency was poor, so the authors presented the coefficients as integers with a normalization factor rather than as a series of floating point numbers.

Among the advantages of the SG filters are their simplicity and versatility. Application of these filters assumes that a local region of the data set (i.e. the filter window) can be fit to a low-order polynomial, and the central point within that window is estimated by performing such a fit. This is illustrated in Figure 16 for first- and second-order polynomials. Fitting the data in this way should model the correlations in the pure signal while reducing the influence of random noise fluctuations. Furthermore, when the measurements are evenly spaced in the time domain, the fitted estimate of the central point can be obtained simply by multiplying the points in the window by a set of precalculated coefficients, thus making the fit equivalent to a nonrecursive digital filter.

The original tables published by SG had a number of errors that were later corrected in the literature,[20] but it is now just as simple to calculate the coefficients for a given application, so the tables are seldom used. To illustrate how this is done, consider the design of a second-order polynomial smoothing filter using a five-point window.

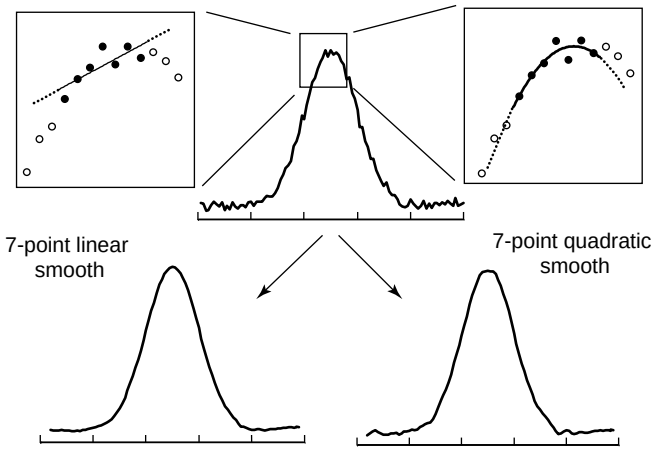

Figure 16 Illustration of polynomial smoothing as a least-squares fitting procedure.

The model to be fit is Equation (12):

$$y = b_0 + b_1 x + b_2 x^2 \tag{12}$$

Equation (13) is the equivalent matrix form:

$$\begin{matrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{matrix} = \begin{matrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{matrix} \begin{matrix} \\ \\ b_0 \\ b_1 \\ b_2 \end{matrix} \tag{13}$$

which can be expressed as (Equation 14):

$$\mathbf{y = Xb} \tag{14}$$

In these expressions, x represents the time or other ordinal variable, while $\mathbf{X}$ is the matrix containing the basis functions for the polynomial fit. It is important to note that the fitted values obtained for y are independent of the scale of x and, if the time interval between each measurement is equal (as is usually the case), we can arbitrarily set $\mathbf{x} = [-2 \quad -1 \quad 0 \quad 1 \quad 2]$, giving Equation (15):

$$\mathbf{X} = \begin{matrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{matrix} \tag{15}$$

The least-squares solution for the vector of regression coefficients, $\mathbf{b}$, is well known from linear algebra to be (Equation 16):

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{A}\mathbf{y} \tag{16}$$

The matrix $\mathbf{A}$ is a 3×5 matrix which can be regarded as being composed of three row vectors, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$; as in Equation (17):

$$\mathbf{A} = \begin{matrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{matrix} = \begin{matrix} \leftarrow \mathbf{a}_1 \rightarrow \\ \leftarrow \mathbf{a}_2 \rightarrow \\ \leftarrow \mathbf{a}_3 \rightarrow \end{matrix} \tag{17}$$

Note that the intercept coefficient for the fit, b_0, is obtained from Equation (18),

$$b_0 = \mathbf{a}_1\mathbf{y} = a_{11}y_1 + a_{12}y_2 + \cdots + a_{15}y_5 \tag{18}$$

Also note that, as $x = 0$ for the central point in the five point sequence, Equation (19) holds:

$$\hat{y}_3 = b_0 + b_1(0) + b_2(0)^2 = b_0 = \mathbf{a}_1\mathbf{y} \tag{19}$$

Therefore, because of the way the problem has been set-up, the estimate of the central point in the sequence is obtained simply by multiplying each measurement by the corresponding element in $\mathbf{a}_1$. In other words, the digital filter coefficients are simply the first row of the matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$, i.e. $\mathbf{c} = \mathbf{a}_1$.

The above reasoning holds for polynomial smoothing filters of any length and any order. All that is required to determine the filter coefficients is to set up the matrix of basis functions, $\mathbf{X}$, perform the calculation in Equation (16) using a spreadsheet or other software, and extract the first row of the resulting matrix. Polynomial smoothing filters are convenient for improving the appearance and SNR of many signals and have the added advantage over electronic filters that different types of filters can be applied after the signal has been recorded.

One of the drawbacks of polynomial smoothing filters is sometimes referred to as the edge effect. As the filters are designed to obtain an estimate of the central point in a window, there will be points at the beginning and end of a measurement vector that cannot be estimated with the symmetric filter. For example, with the five-point filter described above, two points could not be filtered at each end of the data sequence. Several options are available to deal with this problem. The simplest is either to drop these points from the data set, or leave them in the data set unfiltered. Another possibility, if only baseline data occurs at the limits of the measurement vector, is to use the points at one end of the data set to filter those at the other. For example, to estimate y_1 with the five-point filter, we could use the sequence (y_{n-2}, y_{n-1}, y_1, y_2, y_3). Finally, we could employ what are sometimes referred to as initial point filters or extended sliding window filters,[21,22] designed to estimate values other than the central point of a sequence. The coefficients for these filters are easily obtained by simply shifting the $\mathbf{X}$ matrix accordingly. For example, to obtain coefficients to estimate the first point of a five point sequence with a

second-order smooth, we would use Equation (20):

$$\mathbf{X} = \begin{matrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{matrix} \tag{20}$$

All other aspects of the problem are the same. Although this is an elegant way to solve the problem, it requires several sets of filter coefficients to handle the points at the edges of the data and the noise rejection and signal distortion characteristics of these filters are not identical to the symmetric filters.[23]

The selection of smoothing filter parameters (order, number of points) for a given application is often a matter of trial and error and intuition. Obviously, one would like to obtain the maximum noise reduction with the minimum amount of distortion. Although the best noise reduction occurs with wider filters (more points), wider filters also limit the ability of the chosen function to obtain a good local model for a changing signal. In general, noise rejection improves and signal distortion increases as the width of the signal increases and the order of the filter decreases. It should be noted, however, that as a consequence of the mathematics, smoothing filters for orders 0 and 1 are identical, as are those for orders 2 and 3, and so on. To determine the amount of noise reduction a filter will provide, there is a very simple relationship (Equation 21):

$$\frac{\sigma^2_{\text{filtered}}}{\sigma^2_{\text{unfiltered}}} = \sum_{j=p}^{q} c_j^2 \tag{21}$$

Thus the ratio of the variance of the noise in the filtered signal to that in the unfiltered signal is simply the sum of the squared filter coefficients. This equation applies to any type of nonrecursive digital filter, but is only valid in cases where the signal exhibits white noise.

Although Equation (21) is useful for describing the amount of noise reduction, it provides no information about the extent of signal distortion. To this end, it is often more useful to examine the response of a digital filter in the frequency domain in much the same way as for an electronic filter. For a symmetric smoothing filter, the frequency response is given by Equation (22),

$$H(f) = \sum_{k=-m}^{m} c_k \cos\left(\frac{k\pi f}{f_{\text{N}}}\right. \tag{22}$$

where $H(f)$ is the amplitude gain of the filter at frequency f, c_k represents the filter coefficients, and f_{N} is the Nyquist frequency. Note that this does not apply to unsymmetric smoothing filters, such as the initial point filters, because these filters also involve a phase shift,

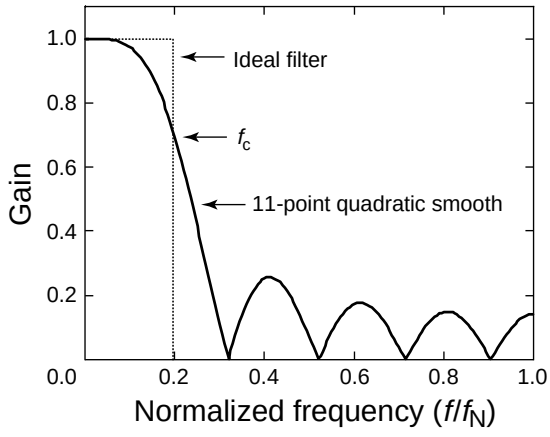

Figure 17 Frequency response (transfer function) for an 11-point quadratic smoothing filter.

but alternative expressions are available.[23] Figure 17 shows the frequency response for an 11-point quadratic smoothing filter, plotted as a function of f/f_{N} to make it universal. As expected, the amplitude gain is unity at low frequencies and drops off at higher frequencies. Unlike simple electronic filters, however, the gain of these filters does not smoothly approach zero, but instead oscillates around a number of nodes that are related to the size and order of the filter. Furthermore, if the frequency response were plotted beyond the Nyquist frequency, the function would simply reflect itself as signals were aliased to lower frequencies, reaching a gain of unity once again at the sampling frequency. If one has some idea of the amplitude spectrum of the signal to be filtered, plots such as Figure 17 can be very useful in assessing the degree of signal distortion that will result from filter application. The amount of noise rejected can be also ascertained from the ratio of the area under the NPS before and after multiplication by the filter frequency response. Clearly, this type of filter will be effective when white noise is present, but will be less effective for $1/f$ (drift) noise, as the noise exists predominantly at low frequencies. It should also be noted that even if white noise was present before filtering, measurement noise will become correlated after filtering.

Of course, whenever a digital filter is applied to experimental data, there will be changes in the shape of the signal. In certain cases where parameters of the signal such as peak height, area or width are important in themselves, consideration must be given to the consequences of applying a digital filter. Generally, for signals which exhibit the same shape, the effects in the time domain (e.g. width at half-maximum) will be the same, whereas effects on the amplitude (e.g. peak height/area) will be linear with the magnitude of

the signal. The reader is referred to a useful, although somewhat empirical, study by Enke and Nieman for more details on these effects.[16]

5.4 Derivative Filters

The numerical differentiation of signals with respect to time (or other ordinal variable) is a common practice in analytical chemistry. This procedure can be used, for example, to locate the position of a peak maximum, to determine the end-point of a titration, or to highlight poorly defined features in a signal sequence (e.g. the shoulder on a peak). These can be regarded as qualitative applications in the sense that one is looking for the location of specific points in the signal derivative (e.g. the maximum or zero values) but not using the derivative sequence for further calculations. Quantitative applications, in which the differentiated signal is used for purposes such as calibration, have become more common in recent years. Because the derivative of a function is unaffected by the addition of a constant, differentiation proves valuable for methods which exhibit a baseline shift (offset) between samples. Likewise, if sample measurements are plagued with a baseline that changes linearly with the ordinal variable (drift), calculation of the second derivative can solve the problem.

One serious problem with signal differentiation is the selective amplification of high-frequency noise. Because the derivative of a signal is, by definition, its rate of change, the more rapidly varying components of a signal, including noise, are amplified to a greater extent than the more slowly changing features typically associated with the pure signal. Because of this, derivative filters (first, second or higher) are most useful for signals which exhibit relatively small amounts of high-frequency noise compared to the low-frequency contributions, i.e. cases where $1/f$ or drift noise dominates. A classic example of this is near infrared (NIR) spectroscopy, where second derivatives are routinely calculated prior to quantitative analysis. Although NIR measurements are widely characterized in the literature as having very high SNRs, these measurements suffer from serious noise problems in the form of baseline offset and drift, but traditional SNR calculations normally do not incorporate these components. Fortunately, the characteristics of NIR spectra make them almost ideal benefactors of derivative filtering.

The noise amplification characteristics of derivative filters can be better understood by examining the transfer function in the frequency domain. As shown in Figure 17, the gain of a smoothing filter is typically unity at low frequencies and falls off at high frequencies. In contrast, Figure 18 shows that the gain of a true derivative filter increases linearly with frequency. This can be easily

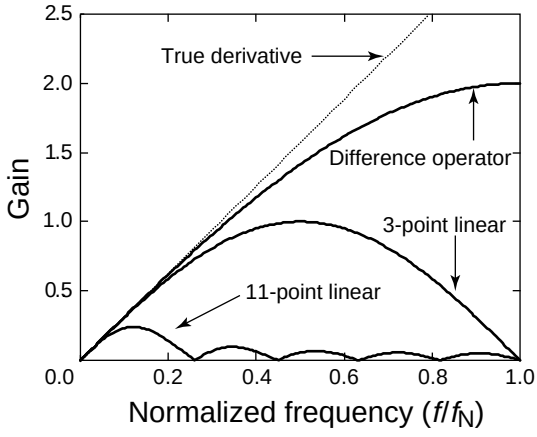

Figure 18 Frequency response (transfer function) for various types of derivative filters.

confirmed by recognizing that (Equation 23),

$$\frac{\mathrm{d}(\sin wt)}{\mathrm{d}t} = w\cos wt \tag{23}$$

where $w = 2\pi f$ is the angular frequency. As a signal in the time domain can be represented as the sum of a series of sines and cosines (section 6.2) it is clear the calculation of the derivative amplifies a signal component by a factor of w and also gives rise to a 90° phase shift. The consequence of this is that the excessive amplification of high-frequency noise components generally makes the calculation of the true derivative for practical measurements useless (and is the reason why the term 'true' is used instead of 'ideal'). In practice, most derivative filters implicitly combine the derivative calculation with a low-pass filter to reduce the contribution of high frequency components.

The simplest method for calculating a signal derivative, and one which can be described as a simple digital filter, is the difference operator, defined by Equation (24),

$$\hat{y}'_i = \frac{y_{i+1} - y_i}{\Delta t} \tag{24}$$

where $\hat{y}'_i$ represents the estimate of the derivative of the function at point i, and Δt is the sampling interval. The transfer function for this type of filter is shown in Figure 18. It is clear that the frequency response for a simple difference operator matches that of the true derivative closely except at very high frequencies. This type of filter provides no low-pass filtering, however.

An alternative way of calculating derivatives is to use a slight modification of the polynomial least-squares filters described in section 5.2. If we carry out the least-squares fit in the same manner as the previous example for a five-point quadratic model, the estimate of the derivative for the central point is given by

Equation (25),

$$\hat{y}'_0 = \frac{\mathrm{d}}{\mathrm{d}x}(b_0 + b_1 x_0 + b_2 x_0^2) = b_1 + 2b_2 x_0$$
$$= b_1 + 2b_2(0) = b_1 \qquad (25)$$

Therefore, the derivative estimate of the central point is simply the first-order coefficient of the fit. In a manner analogous to Equation (18), this is obtained by simply multiplying the second row of the **A** matrix by the windowed measurement vector **y**. This means that the coefficients of the derivative filter are given by the second row of **A** ($\mathbf{c} = \mathbf{a}_2$) as opposed to the first row of **A** for the smoothing filter. Likewise, the filter coefficients for the second derivative are given by the third row of **A**, and so on. It is clear then, that there is a simple, common path to the calculation of polynomial filters of various types.

A number of characteristics of derivative filters calculated in this way should be noted. First, unlike smoothing filters, the calculation of numerically correct derivatives requires consideration of the sampling interval. The adjusted coefficients necessary to obtain the correct scale are given by Equation (26),

$$\mathbf{c}_{\text{adjusted}} = \frac{\mathbf{c}_{\text{original}}\, p!}{(\Delta t)^p} \qquad (26)$$

where p is the order of the derivative. In many cases, this scaling is ignored, because it is only the relative changes in the derivative that are important. A second characteristic of these derivative filters that should be apparent from the mathematics is that the determination of coefficients for a pth order derivative requires at least a pth order polynomial. Also, as with smoothing filters, there is a duplication of coefficients for adjacent polynomial orders, although the pairing shifts with each higher derivative. For example, for the first derivative, the filter coefficients for the linear and quadratic polynomials are the same, whereas the quadratic and cubic coefficients are the same for the second derivative.

The polynomial filters described here are symmetric in the sense that there are an equal number of coefficients on either side of the central point (in fact, a first-derivative filter is better described as antisymmetric, as $c_{-i} = -c_i$). As with smoothing filters, it is possible to develop derivative filters for the edges of a window, but the usual precautions regarding the quality of estimates apply. For symmetric first-derivative filters, the frequency response is given by,

$$H(f) = \sum_{k=-m}^{m} c_k \sin\left(\frac{k\pi f}{f_{\mathrm{N}}}\right. \qquad (27)$$

Equation (27) is a slight modification of Equation (22), where the substitution of sine for cosine results from the 90° phase shift brought about by the derivative filter. The frequency responses for 3-point and 11-point first-order derivative filters are shown in Figure 18 for comparison with the true derivative and difference operator. Note that although the transfer functions for the polynomial filters match the true derivative at low frequencies, there is significant attenuation at high frequencies due to the low-pass filtering. The effects of this low-pass filtering are clearly seen in Figure 19, which shows the application of three types of derivative filters to a noisy signal. Because of the effect of high-frequency noise on derivative filter response, it is a common practice by some to first apply a smoothing filter to the data, but Figures 18 and 19 demonstrate that if the derivative filter is properly designed, this practice is redundant.

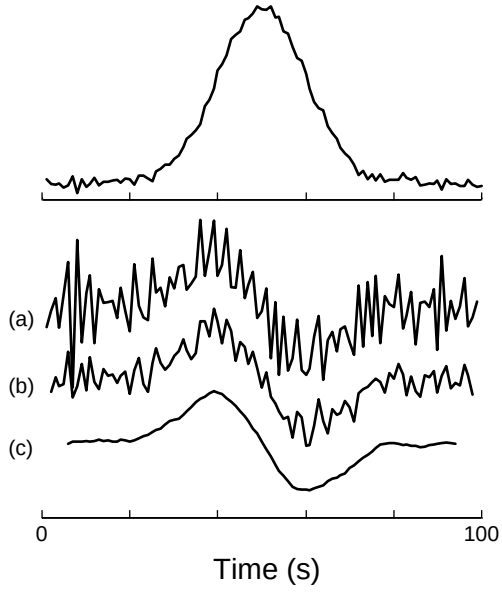

Figure 19 Result of the application of various types of derivative filters to a noisy peak, shown above: (a) difference operator, (b) 3-point linear derivative filter, (c) 11-point linear derivative filter.

5.5 Kalman Filters

The Kalman filter is a recursive linear least-squares estimator with the capability of estimating the parameters associated with a system model in real time.[17–19] It is not so much a filter in the conventional sense as it is a means for carrying out linear least-squares in a recursive fashion. The estimates it provides are not the smoothed measurements, but rather the parameters associated with the linear model, or the state parameters. These parameters are sometimes considered to represent a state vector in an n-dimensional state space (n = number of parameters). Once the state parameters have been estimated, it is possible to generate a smooth curve for

the measurements from the model, but this is usually a secondary objective.

A simple example of recursive estimation in a manner similar to the Kalman filter is the calculation of a mean from a series of measurements as the measurements are being acquired. It is usual to start with an estimate of the mean equal to the first measurement, i.e. $\hat{m}_1 = x_1$. Once a second measurement was acquired, the estimate could be improved by using $\hat{m}_2 = 1/2\hat{m}_1 + 1/2x_2$. In general, after the ith measurement, Equation (28) would hold:

$$\hat{m}_i = \frac{i-1}{i}\hat{m}_{i-1} + \frac{1}{i}x_i \tag{28}$$

As each new measurement is assimilated into the estimation, the quality of the parameter estimate improves. It is apparent that Equation (28) has the form of a recursive filter whose coefficients are changing with each measurement. An advantage of recursive estimation is that continuous updates of the parameter(s) of interest are obtained with each new measurement. Although this could also be done in batch mode, the recursive formulation is computationally more efficient.

The general model which is covered by the Kalman filter can be described by Equations (29) and (30):

$$\mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{w}_k \tag{29}$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \tag{30}$$

In these equations, $\mathbf{x}_k$ represents the $n \times 1$ vector of parameters to be estimated (the state vector) at measurement interval k and $\mathbf{z}_k$ represents the $m \times 1$ vector of measurements at interval k. The first equation describes how the state vector is expected to change from one measurement interval to the next and contains both systematic and stochastic terms. The $n \times n$ state transition matrix, $\mathbf{F}$, describes the systematic linear transformation, whereas the vector of random variables, $\mathbf{w}$, represents the stochastic change. Each element of $\mathbf{w}$ is assumed to be derived from a zero-mean white-noise sequence and $\mathbf{w}$ is characterized by an $n \times n$ covariance matrix $\mathbf{Q}$.

Equation (30) describes how the state parameters are translated into a measurement or observation vector. The linear relationship is described by the $m \times n$ observation matrix, $\mathbf{H}$. There is also a random noise component assumed for the observations, represented by the vector $\mathbf{v}$. The elements of this vector are also assumed to comprise a white-noise sequence and the covariance of noise in the measurement vector is described by the $m \times m$ covariance matrix, $\mathbf{R}$.

The basic algorithm for the Kalman filter is shown in Figure 20, although several variants exist. The application of this algorithm is best described through a simple example. Suppose, for the purposes of illustration, we are using an absorption spectrometer to monitor a reaction

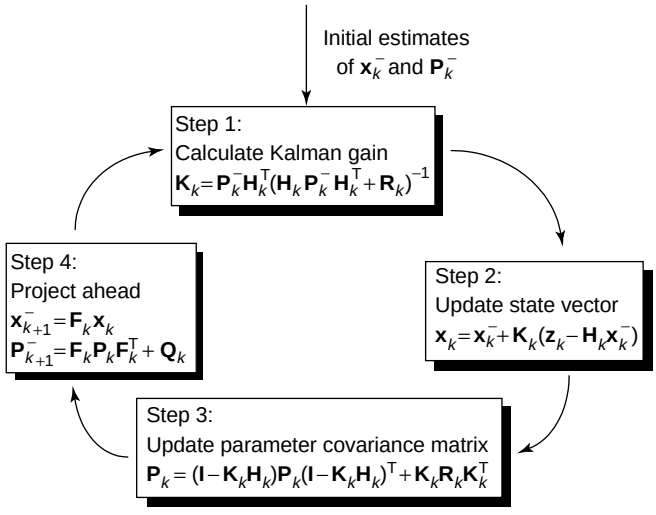

Figure 20 Basic algorithm for the Kalman filter.

in which two absorbing species, A and B, are reacting independently to form nonabsorbing products by first-order kinetics. The defining Equation (31) is:

$$\mathbf{z}(t) = A(t) = C_A\varepsilon_A e^{-k_A t} + C_B\varepsilon_B e^{-k_B t} + v(t) \tag{31}$$

where $A(t)$ is the absorbance at time t, C_A and C_B represent the initial concentrations of the two species, ε_A and ε_B are their molar absorptivities, and k_A and k_B are their first-order decay constants. Assuming that all quantities are known except the initial concentrations which are to be estimated, this is a linear problem (Equations 32):

$$\mathbf{x}(t) = \begin{matrix} C_A \\ C_B \end{matrix} \quad ; \quad \mathbf{H}(t) = \begin{bmatrix} \varepsilon_A e^{-k_A t} & \varepsilon_B e^{-k_B t} \end{bmatrix} \tag{32}$$

A simulated data set was generated using $C_A = 0.2\,\text{mM}$, $C_B = 0.5\,\text{mM}$, $\varepsilon_A = 1000\,\text{M}^{-1}\,\text{cm}^{-1}$, $\varepsilon_B = 1500\,\text{M}^{-1}\,\text{cm}^{-1}$, $k_A = 0.1\,\text{s}^{-1}$, $k_B = 0.3\,\text{s}^{-1}$, and a noise level of 0.02 absorbance units (AU). These data, showing the measured points and the curve with no error, are presented in Figure 21(a). The objective of the Kalman filter is to provide estimates of C_A and C_B as each new measurement is made.

To initiate the Kalman filter prior to step 1, we need an estimate of the state parameters and the error covariance matrix, $\mathbf{P}$, describing the uncertainty in those parameters. Assuming we have no prior knowledge, we use Equations (33):

$$\mathbf{x}_1^- = \begin{matrix} 0 \\ 0 \end{matrix} \quad ; \quad \mathbf{P}_1^- = \begin{matrix} 10^{10} & 0 \\ 0 & 10^{10} \end{matrix} \tag{33}$$

The superscript '$-$' indicates that these are the estimates before we have assimilated the first measurement. As we have no prior knowledge of the parameters, the diagonal

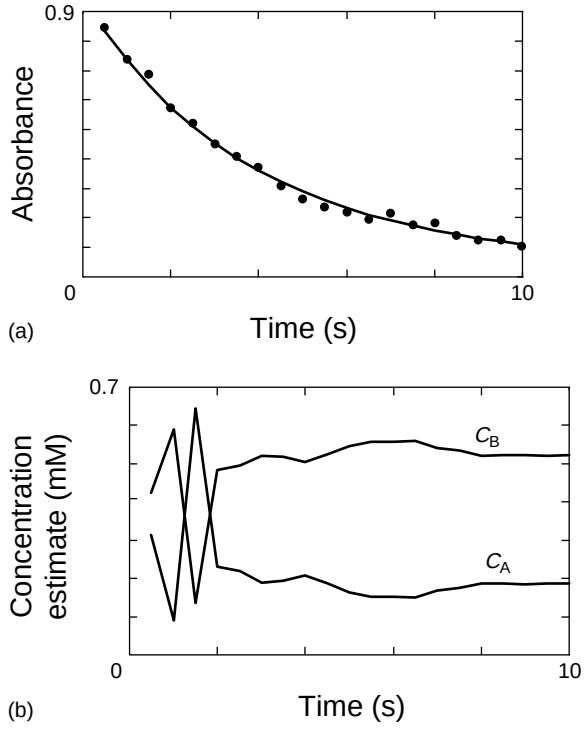

Figure 21 Results from the application of the Kalman filter to the first-order kinetic data described in the text. (a) The solid line shows the true absorbance decay curve and the points indicate the noisy measurements made. (b) The evolution of estimates for the initial concentrations of components A and B by the Kalman filter.

elements (variances) of the covariance matrix are set to very large values. In order to estimate the Kalman gain, **K**, in step 1, the observation matrix, **H**, can be calculated from Equation (32). As the measurement, **z**, is a scalar in this case, the measurement error covariance matrix, **R**, is simply equal to the variance of the measurements. Thus, at $t = 0.5$ s, Equations (34) hold:

$$\mathbf{H}_1 = [951.2 \quad 1291.1]; \quad \mathbf{R}_1 = (0.02)^2 = 0.0004 \qquad (34)$$

With these values, the $n \times 1$ Kalman gain vector for this iteration is Equation (35):

$$\mathbf{K}_1 = \begin{matrix} 3.699 \times 10^{-4} \\ 5.020 \times 10^{-4} \end{matrix} \qquad (35)$$

In step 2 of the algorithm, the difference between the actual observation, $\mathbf{z}_k$, and the observation predicted by the current state parameters, $\mathbf{H}_k\mathbf{x}_k$, is calculated. This difference is sometimes called the innovation and is like an ordinary residual except that it is calculated using the current rather than the final parameter estimates. The Kalman gain vector, **K**, determines how much the innovation is weighted in updating the state parameter estimates. It is also used in the third step of the algorithm to update the error parameter error covariance matrix, **P**, following integration of the new measurement. Using the first observation of 0.846 AU gives Equations (36):

$$\mathbf{x}_1 = \begin{matrix} 0.313 \\ 0.425 \end{matrix} \times 10^{-3}; \quad \mathbf{P}_1 = \begin{matrix} 6.5 & -4.8 \\ -4.8 & 3.5 \end{matrix} \times 10^{9} \qquad (36)$$

Note that because only one measurement has been processed and there are two parameters to be estimated, neither **x** nor **P** can be regarded as reliable at this point.

In step 4, the state vector and its covariance matrix are projected ahead to the next measurement interval. This requires a knowledge of **F** and **Q**, which are trivial in this example. As the state parameters here are static (i.e. $\mathbf{x} \neq \mathbf{f}(t)$), the state transition matrix, **F**, is simply the identity matrix. This would not be the case, for example, if the state parameters were the concentrations at time t, rather than the initial concentrations, but the modifications to **H** and **F** would be straightforward in that case. Likewise, we are assuming no random variation in the initial concentrations, so the state vector covariance matrix, **Q**, is equal to zeros. The role of **Q** in a more complex application is to allow for random variation in the state parameters over time. As an example, suppose that instead of absorbance, we were measuring total pressure in a gas phase reaction which was subject to random temperature fluctuations between measurements. This would effectively change the initial pressures we were trying to estimate.

The iterations of the Kalman filter continue in this way until all of the measurements have been processed. For the example presented here, Figure 21(b) shows the concentration estimates as a function of time with the final estimates:

$$C_A = 0.186 \pm 0.019\,\text{mM}$$

$$C_B = 0.522 \pm 0.023\,\text{mM}$$

The uncertainties are the standard deviations of the parameters from **P**. Note that these converge to values close to the true concentrations. It should also be noted that these are essentially the same estimates that would have been obtained by linear least-squares implemented in batch mode.

Although Kalman first introduced this filter in 1960,[24] applications in chemistry were not abundant until the late 1970s and early 1980s. However, many of these applications employed the Kalman filter mainly as a recursive implementation of simple least-squares, such as the example above, and did not exploit its full capabilities. For instance, given spectra of mixture components, the Kalman filter can be used to estimate component concentrations as a spectrum was being scanned. Although this offered certain advantages such as speed and the ability to terminate an experiment

when the desired precision was achieved, developments in instrumental and computational efficiency have made these benefits less significant. The potential exists for more effective utilization of the algorithm, however.

At least two modifications of the basic Kalman filter have also appeared in the analytical chemistry literature. The extended Kalman filter[25] has been used to model nonlinear systems (e.g. the estimation of rate constants in the example above) through a linearization of the equations but, like most nonlinear methods, convergence can be slow and subject to initial estimates. Often, several passes are needed, defeating the advantages of recursion. A more successful application has been the adaptive Kalman filter,[26] which examines the innovations sequence to detect model errors and effectively turns the filter off in those regions by using an inflated measurement variance estimate. This allows the filter to be applied in situations where strict adherence to the model is not a certainty.

5.6 Other Filters

This section has only scratched the surface of digital filter design, focusing on those filters which are most commonly implemented in analytical chemistry. The reader should be aware that nonrecursive filters with more desirable transfer characteristics, such as a flatter stop band, can be designed with relatively little additional effort, and is referred to appropriate texts on the subject.[9–11] Even more flexibility can be achieved with recursive filters, with characteristics analogous to the Butterworth and Chebyshev designs described earlier for analog signal processing. The popularity of polynomial least-squares filters appears to be a consequence of their intuitive simplicity and the fact that, although not necessarily optimal, they are sufficient for many applications.

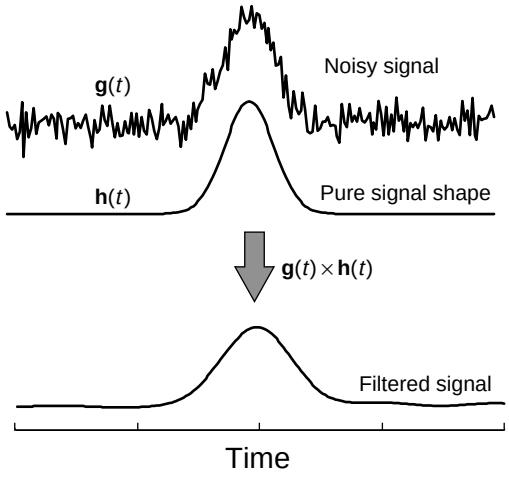

Figure 22 Illustration of the application of a matched filter ($\mathbf{h}(t)$) to a noisy signal ($\mathbf{g}(t)$).

The subject of optimal filtering is revisited in the next section with the Weiner filter in the Fourier domain. In terms of optimal filtering in the time domain, however, one additional filter, the matched filter, deserves mention because it often appears in the analytical literature. With a matched filter, the filter coefficients are obtained simply by normalizing the shape of the pure signal. This is illustrated in Figure 22 with a noisy Gaussian. For white noise, the matched filter is optimal in the sense that it produces the largest SNR, interpreted as the maximum value divided by the baseline noise. Unfortunately, it requires an advance knowledge of the signal shape and has the undesirable consequence of broadening the peak. The optimality of the matched filter derives from its connection to regression. This connection, as well as the relationship between Kalman filtering and regression, has been described by Erickson et al.[27]

6 DOMAIN TRANSFORMATIONS

6.1 Introduction

In the context of signal processing, a domain transformation can be defined as a mathematical or physical process that converts a sequence of measurements into an alternative representation which retains all of the information in the original sequence. A domain transformation is distinguished from a simple domain conversion, such as scaling or current-to-voltage conversion, in that it involves a redefinition of the ordinal variable. As such, the procedures are comparatively complex.

There are two principal reasons why domain transformations are used in chemistry. The first is so that information can be represented in a form commonly used for interpretation. A familiar example is Fourier transform infrared (FTIR) spectroscopy in which the signal, collected by means of an interferometer in the time domain, must undergo a transformation in order to represent it as the familiar plot of transmittance versus wavenumber. The second use of domain transformations is to allow certain operations to be carried out on signals with greater ease. As the objective of signal processing is to separate the pure signal from the noise, transformations which provide a better distinction between these two elements of the signal are useful.

Although there are a large number of possible domain transformations that can be employed, this section will focus on three which have been particularly useful in analytical chemistry: the FT, the WT and the HT.

6.2 Fourier Transforms

Without a doubt, the most widely encountered domain transformation in chemistry is the FT. In addition to being

a useful stand-alone signal processing tool, the FT has become an integral part of many instrumental methods (FTIR, FT/NMR (nuclear magnetic resonance), FTRS (Fourier transform Raman Spectroscopy) and FTMS (Fourier transform mass spectrometry)).[28,29] Although the FT can be applied to both continuous and discretely sampled functions, it is the latter which dominates instrumental applications and will be the focus of this section. The section begins with a basic description and simple illustration of the principles of the FT and concludes with some examples of its application to signal processing. Abundant supplementary information can be found in the literature.[13,14,28–34]

The fundamental principle behind the discrete FT is that any signal sampled at equal intervals in the time domain can have the sampled points reproduced by the addition of a finite number of sinusoids at defined frequency intervals with variable amplitude and phase. This is illustrated in Figure 23 with the simple example of a sampled square wave. Figure 23(a) shows the square wave with the sampled points and the reconstruction using the combination of sinusoids. Although the reconstruction does not match the square wave exactly, it does reproduce the sampled points exactly, which is its only requirement. If the square wave were sampled more frequently, a larger number of sinusoids would be required for reconstruction. Figure 23(b) shows the individual sinusoids added to give the reconstruction, including the DC (direct current) offset (sine wave with frequency of zero).

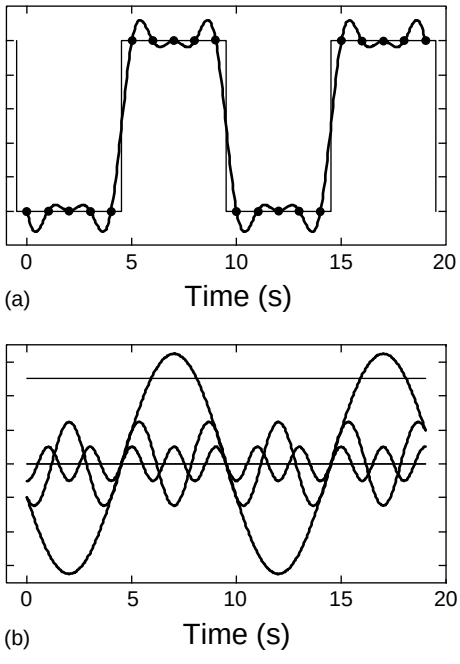

Figure 23 (a) Illustration of the reconstruction of a sampled square wave with a sum of sinusoids. (b) The individual sinusoids used for the reconstruction.

Although two cycles of the square wave are shown in Figure 23 for clarity, the FT is based on a single cycle of 10 points. The reconstruction of the sampled points from the sinusoid terms can be represented as

$$h(t) = \sum_{n=0}^{N/2} C_n \cos\left(\frac{2\pi n f_s t}{N} + \phi_n\right) \qquad (37)$$

In Equation (37), $h(t)$ represents the reconstructed signal at time t, f_s is the sampling frequency, N is the number of points sampled, and C_n and ϕ_n represent the amplitude and phase of the nth sinusoid, which has a frequency of $f_n = (n/N)f_s$. Note that an equivalent representation using sines rather than cosines could have been written simply by adding 90° to ϕ_n, but a cosine expansion is more consistent with FT calculations. For the example in Figure 23, the coefficients C_n and phase angles ϕ_n are given in Table 2 (a sampling interval of 1 s was assumed). Several points should be noted here. First, two sets of amplitudes and angles are given in the table to illustrate an ambiguity in this type of representation – the same result can be obtained by changing the sign of any of the coefficients and shifting the corresponding phase angle by 180°. A second important point is that, whichever set of values is used, there are 12 parameters provided to describe the sinusoids (6 amplitudes and 6 phase angles). Given that we are representing 10 points in the time domain, it would seem that an excessive number of parameters is needed to describe the signal in the Fourier (frequency) domain. However, this is misleading, because the mathematical restrictions can always fix the phase angles for 0 Hz and the Nyquist frequency (0.5 Hz in this case) to be 0°. Therefore, an equal number of values can be used to represent the signal in both domains. Finally, it should be noted that the periodic nature of the sinusoidal basis functions will give rise to a periodic reconstruction even if the original signal is not periodic. This does not mean that nonperiodic signals cannot be transformed, but it should be kept in mind that the FT will treat them as if they are periodic. Discontinuities in amplitude between the beginning and the end of a signal sequence will be reflected in the high frequency components of the FT.

The ambiguity which arose in the amplitude/phase representation of the FT can be resolved by exploiting

Table 2 Amplitudes and phase angles for simple FT example

n	0	1	2	3	4	5
f_n (Hz)	0	0.1	0.2	0.3	0.4	0.5
C_n	0.5	−0.647	0	−0.247	0	−0.1
ϕ_n (°)	0	−72	0	−36	0	0
C_n	0.5	0.647	0	0.247	0	0.1
ϕ_n (°)	0	108	0	144	0	180
A_n	0.5	−0.2	0	−0.2	0	−0.1
B_n	0	0.616	0	0.145	0	0

the fact that a phase shifted sinusoid can be represented as a linear combination of sine and cosine terms. Therefore, an equivalent form of Equation (37) is Equation (38):

$$h(t) = \sum_{n=0}^{N/2} A_n \cos\left(\frac{2\pi n f_s t}{N}\right. + B_n \sin\left(\frac{2\pi n f_s t}{N}\right. \qquad (38)$$

For the square wave example, the coefficients A_n and B_n are also given in Table 2. Although the phase angle has been removed, the same number of parameters as before is required to describe the signal, but there is no ambiguity. The basic objective of the FT is to obtain the coefficients A_n and B_n. Mathematically, this is done by separating the sine and cosine terms through complex arithmetic, recalling Euler's relationship (Equation 39):

$$e^{-i\theta} = \cos\theta - i\sin\theta \qquad (39)$$

where $i = \sqrt{-1}$.

The discrete FT is mathematically defined as follows. Given a series of N measurements in the time domain, h_k, where $k = 0 \ldots N-1$, the N complex coefficients of the FT, H_n, where $n = 0 \ldots N-1$, are given by Equation (40):

$$H_n = \sum_{k=0}^{N-1} h_k e^{2\pi i k n/N} \qquad (40)$$

This calculation results in N real coefficients and N imaginary coefficients which are related to the coefficients A_n and B_n in Equation (38). However, the total number of coefficients here is $2N$, whereas the total number in Equation (38) was $N+2$, which suggests that there is some redundancy. This arises because the FT produces coefficients at both positive and negative frequencies. This is illustrated in Figure 24 which shows the actual FT of the sampled square wave in Figure 23. Both the real and imaginary parts of the FT are shown and the labels on the x axis indicate the correspondence between the coefficient number, n, and the frequency. Note that 0 Hz (DC) and the Nyquist frequency (f_N) are represented only once in the mapping, whereas both positive and negative values are shown for other frequencies. Furthermore, there is a symmetry between the positive and negative frequencies such that Equation (41) holds:

$$H(-f) = H(f)^* \qquad (41)$$

where the asterisk indicates the complex conjugate. This symmetry arises from the fact that h_k is a real function and has no imaginary components.

The representation of the real and imaginary parts of the FT as shown in Figure 24 is an unambiguous presentation of the transform and is the one used in calculations, but it is not normally the one shown in practice. Typically, figures show an amplitude spectrum or a power spectrum and (less frequently) a phase spectrum. Unfortunately, there is a great variability in the scaling and presentation of these spectra, so caution needs to be employed in their interpretation. One way to calculate the amplitude spectrum is to calculate the modulus of H directly (Equation 42):

$$\begin{aligned} \mathrm{Amp}(f) &= \sqrt{\mathrm{real}(H(f))^2 + \mathrm{imag}(H(f))^2} \\ &= \sqrt{H(f)H(f)^*} = |H(f)| \end{aligned} \qquad (42)$$

The results of this calculation for the square wave example are shown in Figure 25(a), where the frequency values have also been properly ordered. Because of the symmetry of the figure for real data, the amplitude spectrum is often represented as simply the right-hand side, excluding negative frequencies. If this were done here, it is clear that the amplitude spectrum would not be consistent with the values given in Table 2. There are three reasons for this. First, the amplitude spectrum calculated in this way will always give positive values because of the ambiguity in the sign of the square root. Second, there is a scaling factor of $1/N$ needed to go between Figure 25(a) and Table 2. This scaling factor normally appears in the *inverse Fourier transform* (IFT), defined by Equation (43):

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-2\pi i k n/N} \qquad (43)$$

(Note that the IFT is essentially the same as the forward FT except for the sign change and the scaling factor.)

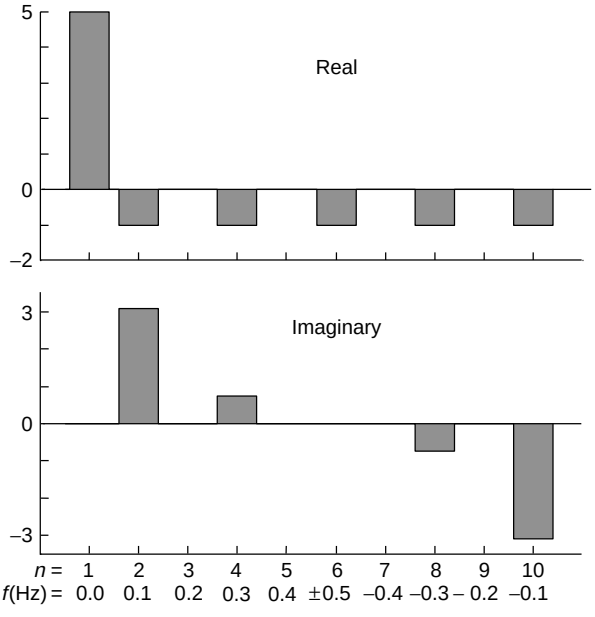

Figure 24 FT of the sampled square wave in Figure 23(a) (one cycle of 10 points only) showing the real and imaginary components and the mapping of points to the frequency domain.

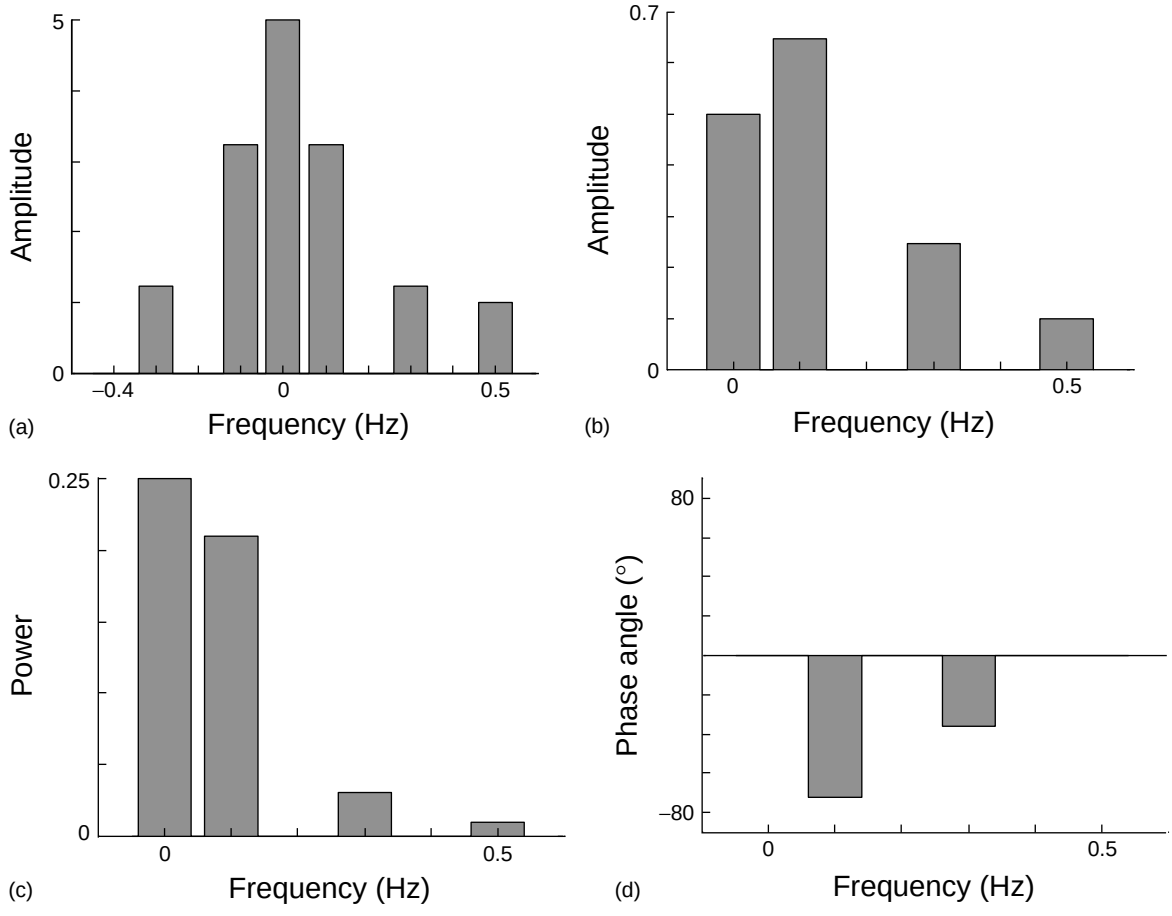

Figure 25 Alternative representations of the FT of the sampled square wave in Figure 23(a). Parts (a) and (b) are different representations of the amplitude spectrum sometimes used; (c) is the power spectrum and (d) is the phase spectrum.

Finally, in order to arrive at an amplitude consistent with Table 2, it is necessary to combine positive and negative frequencies (except for DC and f_N). Thus an alternative definition of the amplitude spectrum (nonnegative frequencies only) is Equations (44):

$$\begin{aligned} \text{Amp}(0) &= \frac{1}{N}|H(0)| \\ \text{Amp}(f) &= \frac{1}{N}(|H(f)| + |H(-f)|) = \frac{2}{N}|H(f)| \\ \text{Amp}(f_N) &= \frac{1}{N}|H(f_N)| \end{aligned} \tag{44}$$

This representation of the amplitude spectrum is shown in Figure 25(b) and, except for the signs, is consistent with the data in Table 2.

The power spectrum (or power spectral density function) is often used in place of the amplitude spectrum. As its calculation involves squaring the amplitudes, the same scaling inconsistencies exist here as for the amplitudes, so care should be taken. The power of the signal in the two domains is related through Parseval's theorem (Equation 45):

$$\sum_{k=0}^{N-1} |h_k|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |H_n|^2 \tag{45}$$

One method to arrive at a valid power spectrum is to use a set of equations similar to those given in Equation (44) for the amplitude, replacing amplitude with power by squaring each of the modulus terms and each N. This gives the power spectrum in Figure 25(c). It can be verified that the sum of the elements is equal to 0.5, which is equal to the mean squared value of the signal in the time domain. There are other aspects to the calculation of power spectra, such as windowing methods to prevent leakage among frequencies and improve the quality of the spectral estimation, but these will not be described here.

Although there are some variations in the manner of calculation of amplitude and power spectra, these are not especially serious when the FT is used for descriptive purposes. The phase spectrum, which plots the phase angle as a function of frequency, is less useful than the amplitude spectrum in most instances.

The phase spectrum for the current example is shown in Figure 25(d). As the negative frequencies contain redundant information by symmetry, only the positive half of the phase spectrum is shown. The phase angles are calculated from Equation (46):

$$\phi(f) = \tan^{-1}\left(\frac{\text{imag}(H(f))}{\text{real}(H(f))}\right) = \arg(H(f)) \qquad (46)$$

As in the amplitude calculation, this equation will have an ambiguity due to the fact that the angle calculated will always be between -90° and $+90^\circ$. Table 2 indicates that, in this case, positive amplitudes should produce phase angles outside this range, but this is clearly not reflected in the phase spectrum. Therefore, although the amplitude/power and phase spectra are the most common descriptive forms of the FT, the indeterminacy of the resultant parameters means that they cannot actually be used to regenerate the original signal in most cases. It should also be noted that Equation (46) involves the ratio of two numbers and this can cause problems in the phase angle calculation when both terms are very close to zero, since round-off error leads to an arbitrary phase angle. Although this is of no importance in the final result (the amplitude of the frequency component is zero), it can complicate the interpretation of the phase spectrum.

As already noted, the calculation of the discrete FT can be carried out using Equation (40), but for most real applications involving a substantial number of data points, the application of this equation is impractical due to the large number of operations required (on the order of N^2). For this reason, few applications employed the FT until the mid-1960s, when the FFT algorithm was popularized by Cooley and Tukey.[2] The FFT greatly reduced the number of operations required (of the order of $N \log N$) and made transformations practical for a wide variety of problems. Although beautifully elegant in its partitioning of the problem, a somewhat annoying requirement of the original algorithm was that it required the number of points to be equal to a power of two. Improvements on the original algorithm have largely removed this restriction (although they are not quite as efficient), and FTs are now calculated with ease for most signals of moderate length using a variety of software packages.

One of the principal applications of the FT is the conversion of data recorded by instruments such as FTIR spectrometers, where it is needed to transform the measurements from the time domain to the frequency domain (or vice versa) before it can be interpreted in the conventional way. In addition to this, the FT is used for a great many signal manipulation purposes, such as smoothing, deconvolution, and interpolation. Some of these applications are now illustrated.

A great deal of the utility of the FT in signal processing derives from the convolution theorem, which states that the convolution of two signals in the time domain is equivalent to the element-by-element multiplication of the functions in the frequency domain. Mathematically, if g and h are functions in the time domain and G and H are the corresponding functions in the frequency domain, Equation (47) holds:

$$g(t) * h(t) = \text{IFT}[G(f)H(f)] \qquad (47)$$

where the asterisk indicates the convolution of the two functions and IFT indicates the inverse transform. As digital filtering is the convolution of filter coefficients with a noisy signal, this immediately leads to an application in Fourier smoothing. The difference here is that we can specify the transfer function of the filter exactly through its FT. This is illustrated in Figure 26 with the smoothing of a noisy Gaussian. The FT of the noisy signal is first calculated and then both the real and imaginary parts are multiplied by the ideal transfer function which sets all of the high frequencies where there is no significant signal

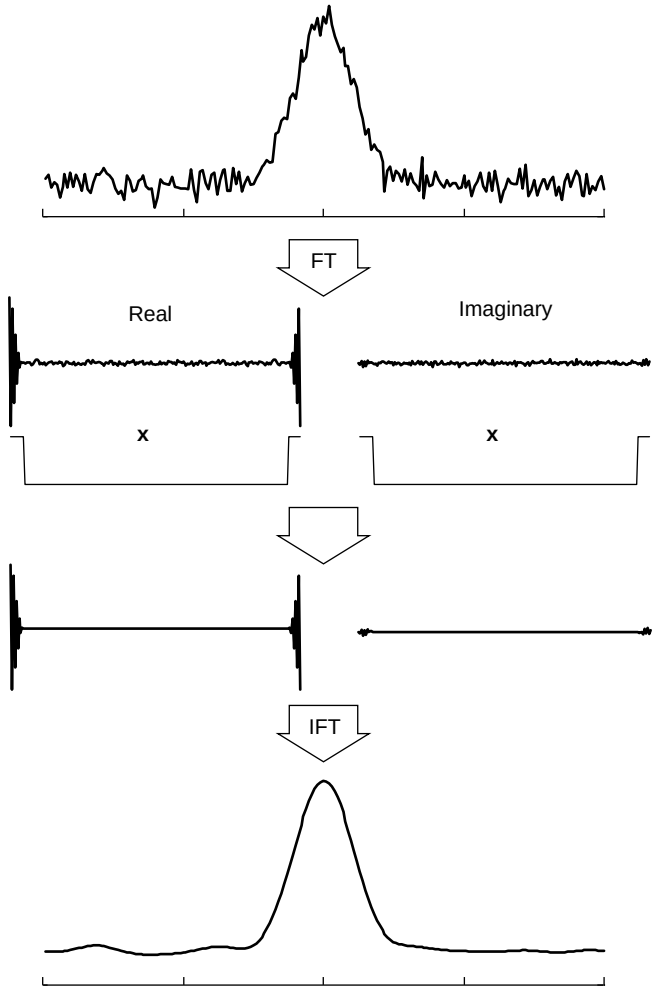

Figure 26 Fourier filtering of a noisy signal.

contribution to zero. Note that both positive and negative frequencies must be included in this multiplication, which is why the transfer function looks somewhat different than that shown earlier. After the multiplication is carried out, an IFT is applied to the result to give the smoothed signal in the time domain.

Although this procedure works very well, it has some drawbacks. First, it is slower than a digital filter and cannot be done in real time because the entire signal is required. Second, artifacts such as the oscillations near the tails of the peak are often observed due to the sharp transition of the transfer function. More severe distortion can result if the cut-off frequency is moved closer to the signal components, but there will be less noise reduction if it is moved to higher frequencies. To avoid this characteristic of the ideal filter, a more gradual decrease in the transfer function is often employed. Such a function is sometimes referred to as an apodization function. If the FT of the pure signal is designated as S and that of the noise as N, it can be shown that the transfer function of the optimal, or Weiner, filter Φ is given by Equation (48):

$$\Phi(f) = \frac{|S(f)|^2}{|S(f)|^2 + |N(f)|^2} \tag{48}$$

This is the transfer function that will give the optimal reproduction of the true signal in the least-squares sense. The difficulty with applying this filter is in the estimation of $S(f)$ and $N(f)$ for the pure signal and noise, but that is beyond the scope of this article.

In addition to convolution, the FT can aid in the deconvolution of two signals. If h represents the convolution of two signals, f and g in the time domain ($h = f^*g$), where g is known, the deconvoluted signal f can be obtained through an element-by-element division in the frequency domain (Equation 49):

$$f(t) = \text{IFT}\ \frac{H(f)}{G(f)} \tag{49}$$

This is illustrated in Figure 27 where a simulated spectral doublet has been convoluted with the slit function of the spectrometer which smears the two peaks. Through Fourier deconvolution, it is seen that the original line shape can be recovered. Practically speaking, there are a number of difficulties associated with this procedure. First, one of the convolution functions needs to be known in advance. Second, the division presents problems when the denominator in Equation (49) is close to zero and adjustments need to be made in this case. Depending on how this is done, artifacts such as those apparent in the baseline of the deconvoluted spectrum of Figure 27 can result. Finally, the presence of noise can lead to additional complications in the deconvolution. In a best-case scenario, however, this type of deconvolution can be

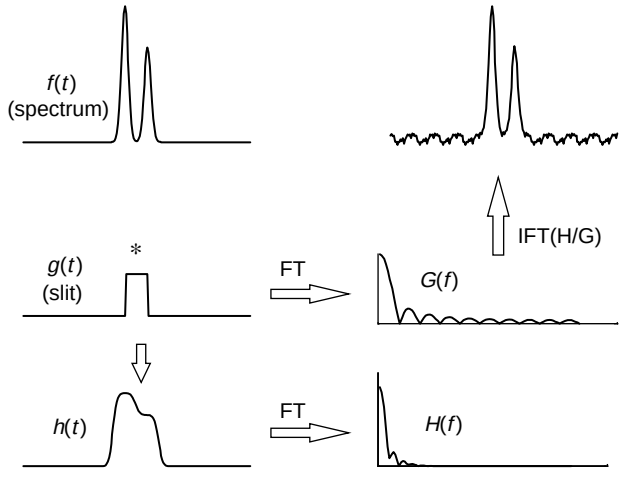

Figure 27 Convolution of a spectral doublet with a slit function followed by Fourier deconvolution to recover the original line shapes.

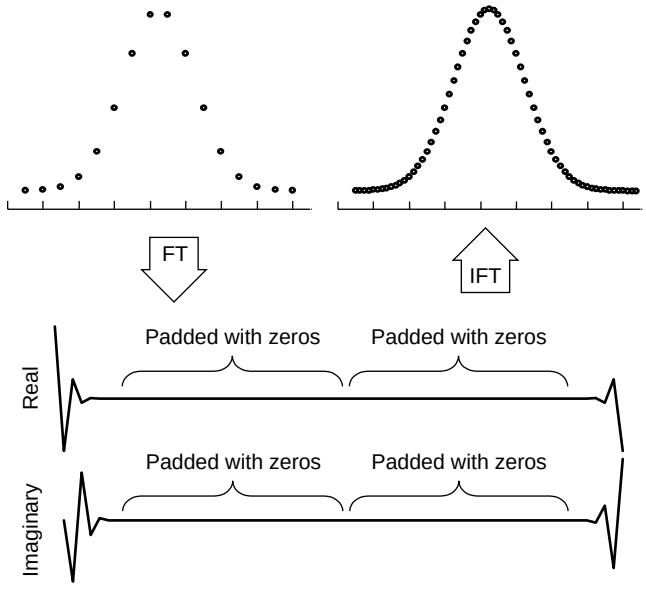

Figure 28 Fourier interpolation of an undersampled signal by a factor of four. Note that the FT has been padded with 24 zeros in both the positive and negative frequency regions to give a total of 64 points.

used to improve the resolution of the instrument after the measurements have been obtained.

As a final example of the use of the FT in signal processing, Figure 28 shows an example of function interpolation. This procedure, which has been referred to as the zoom FT, is based on the fact that the Nyquist frequency is directly related to the sampling frequency. If the FT of a signal is padded with zeros at the frequency limit where there is little signal contribution, the Nyquist frequency can be increased,

and consequently the sampling interval decreased. This essentially interpolates the function between existing measurements. Although this approach has been used in certain applications, such as locating the peak maxima in undersampled mass spectrometry peaks, it should be used with caution because the zero-padding makes implicit assumptions about the form of the function between the original points which may be erroneous.

6.3 Wavelet Transforms

Although a more recent development than the FT, the WT approach is gaining increased acceptance as a signal processing tool for the analytical chemist. The first applications in the chemical literature of the WT as a denoising, smoothing and data compression procedure appeared in the early 1990s, and their frequency of mention has increased steadily. The utility of the WT for noise reduction purposes rests largely on its decomposition of the signal into successive levels of high- and low-frequency components. In data compression applications the decimation filter property of the WT is useful, effectively reducing the number of elements needed to represent the signal with minimal loss of information. Numerous algorithms for the WT have been devised, with the most popular being the recursive form of the discrete wavelet transform (DWT) attributable to Mallat,[35] the generalization of which is known as the wavelet packet transform (WPT).[30,36] An increasing number of software packages are now available for performing DWTs and WPTs as well as related functions in a relatively straightforward manner.

Like the FT, the WT converts the data into a more useful domain for signal processing by projecting the observed signal onto a set of orthogonal basis functions. In the FT, the signal is projected into the frequency domain using sinusoids as the basis functions. In the frequency domain, the basis functions are localized, but when transformed to the time domain the functions extend globally along the time axis. In contrast, the WT uses basis functions that are both localized in the frequency and time domains to project the data into the wavelet domain. The WT, therefore, has perhaps a more intuitive appeal for some who routinely deal with signals that are time localized, such as chromatographic or spectroscopic measurements. A very readable introduction to denoising and compression of chemical signals using wavelets has been written in tutorial fashion by Walczak and Massart.[37]

In the popular DWT pyramid algorithm of Mallat,[35] a recursive decomposition of the signal is performed using both high- (**H**) and low-pass (**L**) filter matrices which are rectangular (each $n/2 \times n$, if the observed signal vector **x** has n elements). The coefficients of these decimation filters depend on the family of wavelets that are used; the Daubechies family appears to be the most popular in chemical applications. The portion of the signal that passes through the low-pass filter is typically called the mth-level approximation to **x**, $\mathbf{a}_m$, and the portion of the signal that is rejected by **L** (passes through the high-pass filter) is referred to as the mth-level detail in **x**, $\mathbf{d}_m$ ($\mathbf{d}_m$ can be considered to be the information in $\mathbf{a}_{m-1}$ not included in the approximation $\mathbf{a}_m$). As **H** and **L** are decimation matrices with a down-sampling rate of two, the number of elements in both **a** and **d** drops by a half at each level of approximation (an effective doubling of the sampling interval). The algorithm requires the number of measured channels in the signal vector to be a power of two, although this requirement can be side-stepped with zero-padding. Edge effects can often result from the filter convolutions; however, signals that are not of a length 2^J (J is an integer) can be zero-padded on both ends of the signal vector to sterilize the distortion.

The DWT algorithm is logistically summarized in Figure 29(a). Clearly, the detail at a given level is not used in subsequent approximation steps of the pyramid algorithm, because $\mathbf{a}_m$ is only approximated from $\mathbf{a}_{m-1}$, and the possibility exists that important information was rejected by the low-pass filter and resides in the detail vector. When the detail vectors are also incorporated into the decomposition, the WPT algorithm results (illustrated in Figure 29b). The structure of the **L** and **H** filters used in the decomposition is shown in Figure 30 for the Daubechies-4 wavelet (four coefficients). Note that the coefficients for **H** are the same as for **L** except that the order is reversed and alternate signs are changed. Normally, the **L** and **H** matrices are concatenated and multiplied by the concatenated basis vectors at each level. With this approach, the arrangement of **L** and **H** matrices will change at each level, as shown in Figure 31.

With the decomposition of the original signal accomplished, it is possible to examine the coefficients of the approximation and detail levels. This is analogous to examining the coefficients of a Fourier decomposition, except that the wavelet decomposition is arranged in two dimensions – time and frequency. As in Fourier denoising applications, it is desirable to eliminate or reduce coefficients believed to be associated with noise, and retain coefficients reflecting information in the original signal. In the WPT, however, not only must coefficients be selected in the frequency realm, but also in the level of approximation. In order to determine the appropriate basis vectors for reconstruction of the signal, it is useful to examine the approximation and detail vectors resulting from the transform. A simple example of denoising using the WPT is presented in Figure 32. In this example, the WPT has only been carried out to two levels for purposes of illustration. At level 2, it is apparent that three of the

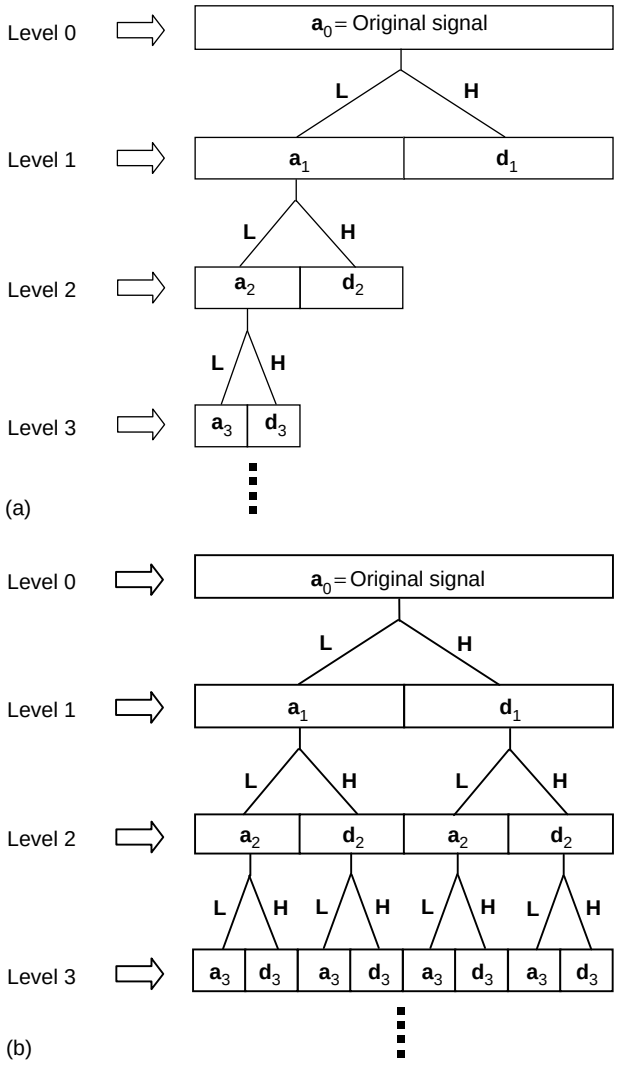

Figure 29 Illustration of (a) the DWT and (b) the WPT.

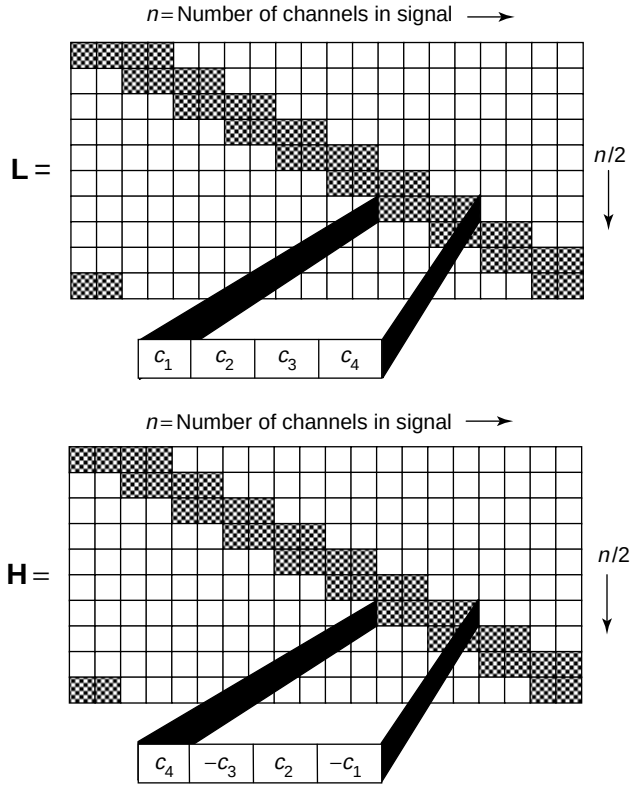

Figure 30 Illustration of the **L** and **H** matrices (four coefficients) used in the WPT to generate the approximation (**a**) and detail (**d**) vectors, respectively, at successive levels of resolution. For the Daubechies-4 wavelet, the coefficients are $c_1 = 0.4830$, $c_2 = 0.8365$, $c_3 = 0.2241$, and $c_4 = -0.1294$. Note that the coefficients will not change for different levels of resolution, but the size of the filter matrix will depend on n, the number of channels in the signal at the previous level of resolution.

four vectors contain little useful information, so these are set to zero before the inverse wavelet transform (IWT) is carried out, resulting in a reduction in the noise. It should be noted that no attempt was made to optimize the denoising in this example, and decompositions to additional levels may have allowed further improvement. Denoising with the WT involves more options than that for the FT, such as the selection of a set of basis functions, the level of decomposition used, and the choice of basis vectors to set to zero.

Because of the wide range of possibilities, algorithmic methods of basis selection have been proposed, with the most intuitive and successful using the minimum entropy (or maximum information) condition of Coifman and Wickerhauser.[38] Best-basis selection according to the minimum entropy condition proceeds on the principle that the most useful basis vectors will be those that contain the most information, and informative vectors will tend to be those which have some large coefficients and some small ones. The coefficients of uninformative basis vectors will largely be the same. The minimum entropy condition in best-basis selection is typically applied by seeking the basis vectors which have the greatest number of coefficients above a preset threshold value.[39] Another method of selecting the best basis vectors is the minimum description length (MDL) method.[40,41] The MDL is primarily used when data compression is desired, and proceeds on a version of the principle of parsimony, seeking the basis vectors which contain the most information in the fewest coefficients.

With the best set of basis vectors selected from the full decomposition, additional signal adjustment can be made by using hard or soft thresholding measures on the remaining coefficients. In hard thresholding, coefficients above a preset threshold are retained, whereas coefficients below this level are discarded. Soft thresholding can also entail zeroing of coefficients below the threshold,

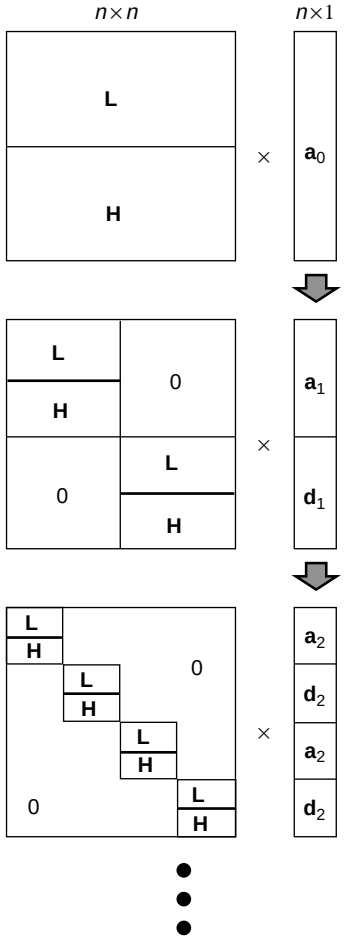

Figure 31 Arrangement of the **L** and **H** filter matrices through the first three levels of the WPT.

but coefficients above the threshold are also typically shrunk towards zero by an amount inversely proportional to their magnitude. Evidently the selection of the threshold value is crucial in these procedures, and several methods exist for estimating the optimal threshold value, including estimating the threshold based on the level of noise,[42] and setting the threshold as a percentage of the largest coefficient.[43] Typical hard, and soft thresholding functions are shown in Equations (50) and (51):

hard thresholding:

$$c_{\text{new}} = \begin{cases} 0, & \text{if } |c_{\text{old}}| < T \\ c_{\text{old}}, & \text{if } |c_{\text{old}}| \geq T \end{cases} \tag{50}$$

soft thresholding:

$$c_{\text{new}} = \begin{cases} 0, & \text{if } |c_{\text{old}}| \leq T \\ \text{sign}(c_{\text{old}})(|c_{\text{old}}| - T) & \text{if } |c_{\text{old}}| > T \end{cases} \tag{51}$$

As an alternative to using threshold values to select relevant coefficients, wavelet smoothing can be achieved by simply discarding detail vectors and performing the inverse WPT from the desired approximation vectors. Although this technique has the potential to achieve greater compression ratios, it is a perilous operation when one lacks knowledge of the location of relevant information in the wavelet decomposition – some useful information may well be contained in some detail vectors. With this possibility looming, it is generally recommended that wavelet smoothing by discarding detail vectors be reserved for situations in which extensive knowledge of the signal allows for educated detail removal.

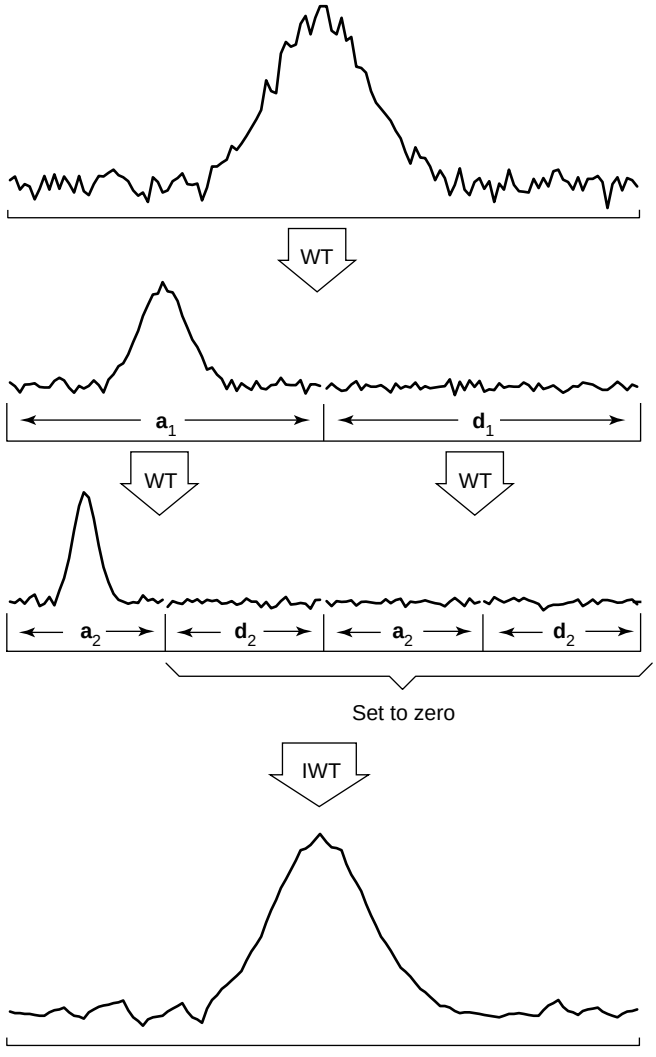

Figure 32 Simple illustration of signal denoising using the WPT.

With basis selection, and coefficient adjustment complete, it is possible to approximate the original signal in the original resolution. This domain is revisited by passing the selected basis vectors back through the high- and low-pass filters. To make the filters interpolation rather than decimation filters, the conjugates of **H**, and **L** are employed. Based on the orthonormality of the

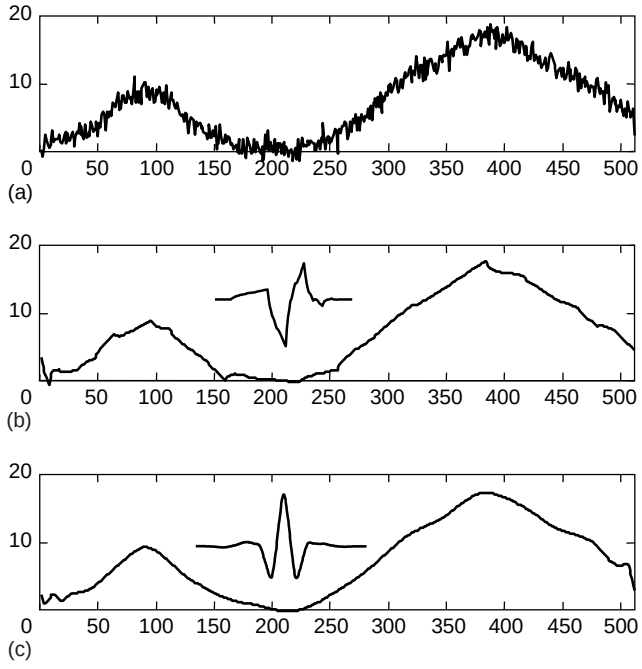

Figure 33 Illustration of signal denoising using Coifman's best-basis selection algorithm and coefficient adjustments using hard thresholding: (a) original simulated noisy spectrum, (b) denoised spectrum using the Daubechies-4 wavelet (inset), and (c) denoised spectrum using the Coiflet-3 wavelet (inset).

two matrices, the conjugates of **H** and **L** are equivalent to the transposes. Therefore the IWT proceeds straightforwardly through the up-sampling filters $\mathbf{H}^{\mathrm{T}}$ and $\mathbf{L}^{\mathrm{T}}$ (transposes of the matrices in Figure 31) until the desired resolution is achieved, with the denoised signal resulting.

Figure 33 illustrates the utility of the WT in signal denoising using a simulated spectrum. White noise was added at a level corresponding to $\sigma_{\mathrm{noise}} = 0.1$, giving the noisy spectrum in Figure 33(a). The first wavelet chosen for the transform was the Daubechies-4 ($c_1 = 0.4829$, $c_2 = 0.8365$, $c_3 = 0.2241$, $c_4 = -0.1294$). Best-basis selection was performed using the minimum entropy procedure of Coifman, and hard thresholding was used on the chosen basis vectors. The results of the overall wavelet denoising with the Daubechies-4 wavelet are shown in Figure 33(b). To illustrate the effect that different wavelet families can have, the Coiflet-3 wavelet was also used, and the results shown in Figure 33(c). In practice, a comparison would more probably be made between different members of a family to find the best result.

6.4 Hadamard Transforms

Like other transform methods, the HT can be thought of as a transformation from one space to another, with a Hadamard matrix acting as the transformation matrix.[44–47] HTs are one method of gaining the Felgett advantage, or multiplex advantage, as it is often called. The multiplex advantage is a statistical gain in SNR as a result of simultaneously measuring multiple spectral resolution elements. In contrast to dispersive methods, in which a single spectral element is measured at a time, multiplex methods measure several coincident spectral elements simultaneously. In order for the multiplex design to prove beneficial from an SNR perspective, the noise in the signal must be considered to be independent of the strength of the incident radiation (i.e. detector noise is the overwhelming noise source). If this condition is met, simple propagation of error reveals that the SNR of a multiplex instrument, relative to a dispersive instrument is $\sqrt{N}$, where N is the number of spectral elements that impinge on the detector at any one time.

Although FT spectrometers are perhaps the best known instruments to utilize the multiplex advantage, the HT spectrometer is also a valuable option. The principles of the HT are based on the concept of Hadamard matrices. As noted above, the benefit of the HT stems from propagation of measurement error into the estimated spectral values. If one observation is made with an inherent detector error of e, then the error in the estimated value is e. However when we wish to estimate several unknowns we can reduce the error associated with a particular estimate by measuring groups of unknowns together in a well-designed fashion. We can subsequently use systems of linear equations to solve for the estimates, and decrease the error in those estimates in the process. The classic analogy is to a weighing scheme for several unknown objects. In the example sketched here, four objects are weighed in experiment A one at a time, with a detector noise level of 0.1. (Experiment A is analogous to a dispersive spectrometer.) In experiment B, two or three objects are weighed together at any one time, although we still only have four total measurements to estimate each individual object's mass. (Experiment B is analogous to a single detector HT spectrometer.)

Experiment A

$$\begin{matrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{matrix}\begin{matrix}m_1\\m_2\\m_3\\m_4\end{matrix}=\begin{matrix}x_1\\x_2\\x_3\\x_4\end{matrix}+\begin{matrix}e_1\\e_2\\e_3\\e_4\end{matrix}$$

$$\mathbf{W}_{\mathrm{A}}\mathbf{m} = \mathbf{x} + \mathbf{e}$$

$$\mathbf{m} = \mathbf{W}_{\mathrm{A}}^{-1}\mathbf{x} + \mathbf{W}_{\mathrm{A}}^{-1}\mathbf{e}$$

Experiment B

$$\begin{matrix}0&1&1&1\\1&1&0&0\\1&0&1&0\\1&0&0&1\end{matrix}\begin{matrix}m_1\\m_2\\m_3\\m_4\end{matrix}=\begin{matrix}x_1\\x_2\\x_3\\x_4\end{matrix}+\begin{matrix}e_1\\e_2\\e_3\\e_4\end{matrix}$$

$$\mathbf{W}_{\mathrm{B}}\mathbf{m} = \mathbf{x} + \mathbf{e}$$

$$\mathbf{m} = \mathbf{W}_{\mathrm{B}}^{-1}\mathbf{x} + \mathbf{W}_{\mathrm{B}}^{-1}\mathbf{e}$$

Here the x_i terms represent the observed reading on the scales, and the m_i terms represent the estimated mass of the ith object. Through propagation of error, it is relatively easy to show that if the measurement uncertainties for the x_i terms are independent and given by σ_x, then the uncertainties in the masses will be given by Equation (52),

$$\begin{matrix}\sigma^2_{m_1}\\ \sigma^2_{m_2}\\ \vdots\end{matrix} = \sigma^2_x \text{diag}\,[(\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1}] \qquad (52)$$

where 'diag' indicates extraction of the diagonal elements. Solution of this equation using $\sigma_x = 0.1$ gives $\sigma_m = 0.1$ for all masses in experiment A, whereas the values for experiment B are 0.067, 0.088, 0.088, 0.088 for m_1–m_4, respectively. Clearly, noise reduction in the estimates has occurred via the multiplex advantage.

In HT spectrometers, the weighing design matrix as shown above is embodied by a mask (Hadamard mask) that physically impedes the incidence of some spectral elements while letting others pass through to the detector. Whereas early HT instruments used a moving mask, the current inclination is toward stationary masks whose codes are changed using electrooptical devices. In true HT spectrometers, light is not only blocked from the coagulating detector, but it is also reflected back to a subtracting detector, such that the measured total intensity is the difference of the adding and subtracting detectors. The weighing matrix in these scenarios, **H**, is a series of 1 and (-1) values representing which elements are subtracted and which are added. These matrices are designed based on Hadamard mathematics. When this arrangement is used, the SNR enhancements observed in FT instruments can be achieved. In practice the HT instruments are difficult to construct to the required specifications and thus single-detector instruments are principally used. The weighing matrix used in these systems is the **S** matrix, and the elements are similar to the weighing matrices shown above (zeros and ones). **S** matrices can be easily constructed from Hadamard matrices by removing the first row and column of **H** and changing all -1 elements in **H** to zeros in the **S** matrix. Although closely related to the Hadamard matrices, **S**-matrix methods do not afford the same enhancement in the SNR as **H**-matrix methods because, with N spectral elements, only $(N+1)/2$ may be measured at any one time.

Like the interferogram resulting from the Michelson interferometer, the encodegram is the resulting signal output from a Hadamard mask experiment. The encodegram relates the radiative flux reaching the detector with the position of the Hadamard mask. To convert this signal in the Hadamard domain to the desired frequency domain the inverse HT is used. Given the properties of **S** (orthonormal rows/columns and square) this is easily accomplished by convolution of the encodegram with the inverse of **S**, i.e. $\mathbf{S}^{-1} = \mathbf{S}^{\mathrm{T}}$.

With the use of electrooptic Hadamard masks come new problems with the standard HT. Although these stationary masks remove the problem associated with the continuously moving parts of the FT instruments, nonidealities in the opacity or transmissiveness of the mask require adjustments to the weighing matrices.[47]

When noise is independent of the signal intensity, as is the case when detector noise dominates, Hadamard multiplexing can prove a useful method of improving the SNR of the spectral estimates. In true Hadamard multiplexing noise reduction follows the general formula of Equation (53):

$$\sigma_{\mathrm{HT}} = \frac{\sigma}{\sqrt{N}} \qquad (53)$$

where σ_{HT} is the standard deviation of the estimated elements using HT methods, σ is the standard deviation of the detector output (equivalent to the noise level in the same experiment using a monochromator), and N is the number of spectral elements to be estimated. However, most HT instruments employ **S** matrix methods which, at best, allow reduction of the uncertainty of the estimate according to Equation (54):

$$\sigma_{\mathrm{HT}} = \frac{2\sqrt{N}}{N+1}\sigma \approx \frac{2\sigma}{\sqrt{N}} \qquad (54)$$

Although the HT has found some utility in analytical applications and is likely to continue to do so in situations for which multiple channel detection or FT methods are unfeasible, its implementation has not been extensive. With the increasing prominence and quality of multichannel detection systems, the multiplex approach of the HT to signal processing is likely to have limited future utility.

7 HIGHER-ORDER SIGNAL PROCESSING

The bulk of this article centers on signal processing methods for first-order data sets – those cases where the signal can be represented as a vector of measurements. In recent years, however, there has been an increased emphasis on the use of higher order data in analytical chemistry. This is particularly true for second-order data sets (matrices of measurements), but increased use of third-order data is also apparent in the literature. This phenomenon can be attributed to three main factors: (a) the demand for new kinds of analytical information and more efficient analytical methods; (b) the increased availability of multichannel detectors, such as photodiode arrays and charge-coupled

devices, as well as rapid scanning instruments; and (c) the development of chemometric methods capable of dealing with multidimensional data. Techniques such as multivariate calibration and pattern recognition are now used routinely, and their application has led to the increased need for signal processing for higher-order data.

In discussing signal processing for higher-order data, it is necessary to make the distinction between the order of the data and the order of the signal. A defining characteristic of a signal is that it exhibits correlation in the ordinal variable for some domain, so a higher-order signal should exhibit correlation in the ordinal variable for each dimension. For example, a collection of spectra from different samples for a multivariate calibration or pattern recognition study would not normally exhibit correlations among the samples, and so this second-order data set can be regarded as a collection of first-order signals. In contrast, spectra obtained during a chromatography or kinetics experiment would result in second-order signals in a second-order data set, because there would be a relationship among the spectra in the time domain. Other combinations are also possible. For instance, if fluorescence emission–excitation spectra were collected for an arbitrary series of samples, we would have second-order signals composing a third-order data set.

For data sets that are composed of first-order signals, signal processing is generally restricted to first-order methods such as those already described. Nevertheless, such signal processing can still have effects across multiple orders and for that reason may be regarded as even more important for higher-order data than for the first-order case. For example, the presence of a variable baseline offset or drift between sample spectra can be detrimental to multivariate calibration methods, but this effect can be minimized by derivative filtering in the spectral domain. The application of such techniques prior to data analysis falls under the subject area of data preprocessing, and includes such methods as mean-centering, baseline subtraction, scaling, smoothing, differentiation, and domain transformation. Choice of an appropriate preprocessing method can be critical and often determines the success or failure of a multivariate analysis application. A complete discussion of these methods is beyond the scope of this article; however, many appropriate texts on chemometrics give more information.[13,48–50]

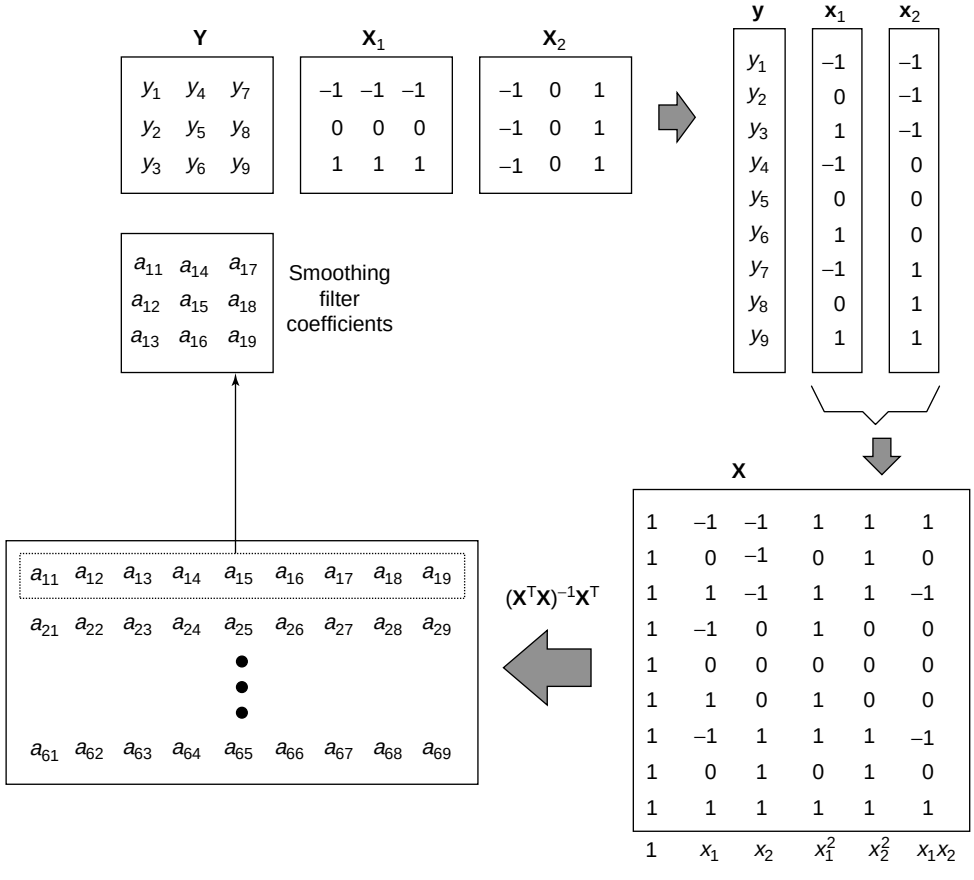

Figure 34 Illustration of the calculation of filter coefficients for a 3 × 3 quadratic filter with an interaction term.

For signals that are truly higher order, first-order signal processing methods can still be used, but other options are also available. In large part, these are extensions of the first-order methods which have already been discussed. In the case of second-order signals, for example, there are two-dimensional (2D) smoothing methods, 2D FTs and 2D WTs. The application of these techniques can offer greater power and flexibility since the characteristics of the signals in both dimensions can be exploited. For example, in the case of spectra collected during a chromatography experiment (a spectrochromatogram), filtering using a nine-point moving average filter in either the time or spectral dimension requires convoluting each signal vector with a 1×9 smoothing vector. However, 2D smoothing could use a convolution of the full matrix with a 3×3 smoothing matrix and the same level of noise reduction would be achieved (in the case of a moving average filter) with less distortion. Understandably, the use of 2D techniques introduces greater complexities in terms of computation, implementation, interpretation and optimization than their one-dimensional (1D) counterparts, but these can be overcome.

As most higher-order signal processing methods are extensions of their 1D counterparts, a detailed discussion is not presented here. However, one example of 2D smoothing is presented as an illustration. In this example, a 3×3 polynomial smoothing filter is used. To demonstrate the design, a quadratic filter with an interaction term was chosen, with the corresponding Equation (55):

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2 + b_5 x_1 x_2 \quad (55)$$

where x_1 and x_2 represent the two ordinal variables. The generation of filter coefficients requires the unfolding of matrices representing the ordinal variables and the process is illustrated in Figure 34. The resulting smoothing coefficients are (Equation 56)

$$\mathbf{C} = \begin{matrix} -0.111 & 0.222 & -0.111 \\ 0.222 & 0.556 & 0.222 \\ -0.111 & 0.222 & -0.111 \end{matrix} \quad (56)$$

The result of the application of this filter to the noisy fluorochromatogram of a mixture of pyrenes in Figure 35(a) is shown in Figure 35(b). Although some noise reduction results, it is not as great as that obtained with a simple 3×3 moving average filter, as shown in Figure 35(c). It is clear that the 2D filter involves the same trade-off between noise reduction and distortion as the 1D filters, but the optimization in the 2D case involves a greater number of options, such as the size and order in each dimension and the inclusion of interaction terms. So far, unlike first-order methods, there have not been extensive studies on the relationship between second-order signal processing methods and the signals they are applied to in chemistry, but this is likely to change as higher-order data become more prevalent.

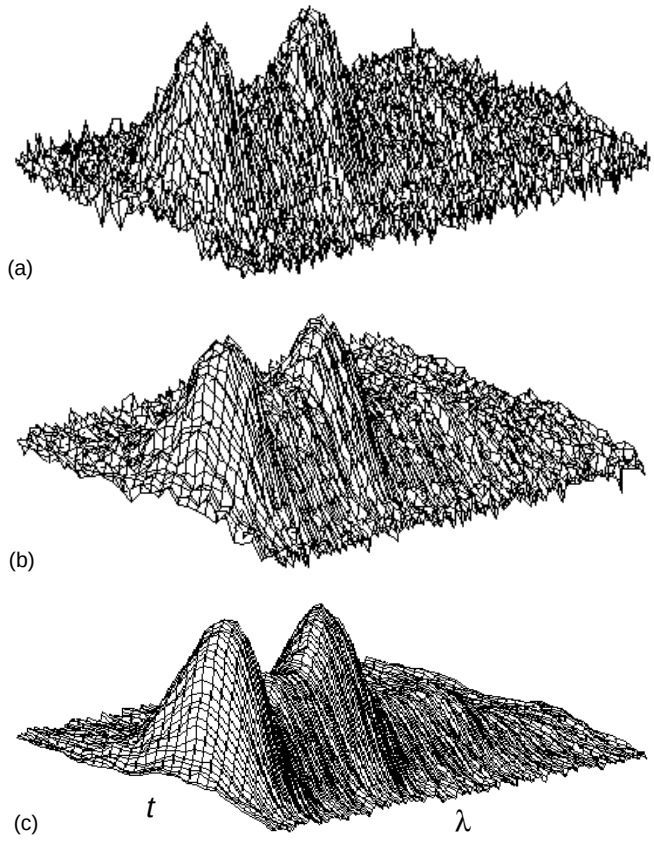

Figure 35 Application of a 2D smoothing filter to noisy data from a chromatogram with multiwavelength fluorescence detection: (a) original data, (b) data filtered with a 3×3 quadratic filter, (c) data filtered with a 3×3 moving average filter.

8 CONCLUSIONS

This article provides a general description of some of the signal processing tools commonly employed in analytical chemistry. As a general principle, it is apparent that all signal processing methods make assumptions about the models for signals and for noise in order to distinguish the two. The power of a particular method in a given application depends on the nature of the assumptions made (very general or very restrictive) and the extent to which they are valid. It is also true that the use of signal processing methods is a double-edged sword. Although the quality of information may be enhanced, it is also possible to distort the signal to the point where results become unreliable. Clearly, a knowledge of the nature of signals, noise and

the capabilities of signal processing methods is essential. For this reason, a significant portion of this article is dedicated to the practical aspects of implementing different methods and their effects on signals.

Developments in signal processing applications to analytical measurements will no doubt continue, particularly for digital signals. Although some methods, such as polynomial smoothing and FT-related techniques, will continue to permeate all areas of analytical chemistry, other methods, such as Kalman filtering and HTs, have found more specialized niches. The impact of WTs is evidence of the ongoing research in signal processing applications. Undoubtedly, future developments will exploit greater computational abilities and present new challenges in application and interpretation. As noted in the previous section, applications to higher-order methods will be a focus of research. In any case, it is apparent that, whether it is the enhancement of fuzzy images from atomic microscopes or the removal of background signals in remote sensing from space, signal processing methods will continue to play a key role in all aspects of analytical chemistry.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the research support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Dow Chemical Company. Prof. L. Ramaley is thanked for his helpful comments.

ABBREVIATIONS AND ACRONYMS

A/D	Analog-to-digital
ADC	Analog-to-digital Converter
AU	Absorbance Units
BCD	Binary Coded Decimal
DAC	Digital-to-analog Converter
DC	Direct Current
DSP	Digital Signal Processor
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FT	Fourier Transform
FTIR	Fourier Transform Infrared
FTMS	Fourier Transform Mass Spectrometry
FTRS	Fourier Transform Raman Spectroscopy
HT	Hadamard Transform
IFT	Inverse Fourier Transform
IIR	Infinite Impulse Response
IWT	Inverse Wavelet Transform
LSB	Least Significant Bit
MDL	Minimum Description Length
MSB	Most Significant Bit
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
NPS	Noise Power Spectrum
RC	Resistor–Capacitor
rms	root-mean-square
SG	Savitzky–Golay
SNR	Signal-to-noise Ratio
V/F	Voltage-to-frequency
WPT	Wavelet Packet Transform
WT	Wavelet Transform
1D	One-dimensional
2D	Two-dimensional

RELATED ARTICLES

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher

Electronic Absorption and Luminescence **(Volume 12)**
Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

Gas Chromatography **(Volume 12)**
Data Reduction in Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

Kinetic Determinations **(Volume 12)**
Data Treatment and Error Analysis in Kinetics

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. A. Savitzky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Square Procedures', *Anal. Chem.*, **36**, 1627–1639 (1964).
2. J.W. Cooley, J.W. Tukey, 'An Algorithm for the Machine Calculation of Complex Fourier Series', *Math. Comput.*, **19**, 297–301 (1965).

3. A.J. Diefenderfer, B.E. Holton, *Principles of Electronic Instrumentation*, 3rd edition, Saunder's College Publishing, Philadelphia, 1994.
4. P. Horowitz, W. Hill, *The Art of Electronics*, 2nd edition, Cambridge University Press, New York, 1989.
5. H.V. Malmstadt, C.G. Enke, S.R. Crouch, *Electronics and Instrumentation for Scientists*, Benjamin/Cummings, Menlo Park, CA, 1981.
6. A.B. Williams, *Active Filter Design Handbook: LC, Active and Digital Filters*, 2nd edition, McGraw-Hill, New York, 1988.
7. Z.H. Meiksin, P.C. Thackray, *Electronic Design with Off-the-shelf Integrated Circuits*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1984.
8. K.M. Daugherty, *Analog-to-digital Conversion: A Practical Approach*, McGraw-Hill, New York, 1995.
9. R.W. Hamming, *Digital Filters*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1983.
10. T.J. Terrell, *Introduction to Digital Filters*, 2nd edition, Wiley, New York, 1988.
11. L.R. Rabiner, B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
12. A.V. Oppenheim, R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
13. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1988.
14. S.D. Brown, 'Signal Processing and Data Enhancement', in *Practical Guide to Chemometrics*, ed. S.J. Haswell, Marcel Dekker, New York, 1992.
15. M.U.A. Bromba, H. Ziegler, 'Application Hints for Savitzky–Golay Digital Smoothing Filters', *Anal. Chem.*, **53**, 1583–1586 (1981).
16. C.G. Enke, T.A. Nieman, 'Signal-to-noise Ratio Enhancement by Least-squares Polynomial Soothing', *Anal. Chem.*, **48**, 705A–712A (1976).
17. R.G. Brown, *Introduction to Random Signal Analysis and Kalman Filtering*, Wiley, New York, 1983.
18. S.D. Brown, 'The Kalman Filter in Analytical Chemistry', *Anal. Chim. Acta*, **181**, 1–26 (1986).
19. H.N.J. Poulisse, 'Multicomponent Analysis Computations Based on Kalman Filtering', *Anal. Chim. Acta*, 361–374 (1979).
20. J. Steinier, Y. Termonia, J. Deltour, 'Comments on Smoothing and Differentiation of Data by Simplified Least Square Procedures', *Anal. Chem.*, **44**, 1906–1909 (1972).
21. A. Proctor, P.M.A. Sherwood, 'Smoothing of Digital X-ray Photoelectron Spectra by an Extended Sliding Least-squares Approach', *Anal. Chem.*, **52**, 2315–2321 (1980).
22. R.A. Leach, C.A. Carter, J.M. Harris, 'Least-squares Polynomial Filters for Initial Point and Slope Estimation', *Anal. Chem.*, **56**, 2304–2307 (1984).
23. P.D. Wentzell, T.P. Doherty, S.R. Crouch, 'Frequency Response of Initial Point Least Squares Polynomial Filters', *Anal. Chem.*, **59**, 367–371 (1987).
24. R.E. Kalman, 'A New Approach to Linear Filtering and Prediction Problems', *Trans. ASME, Ser. D*, **82**(Mar.), 35–45 (1960).
25. S.C. Rutan, S.D. Brown, 'Estimation of First-order Kinetic Parameters by Using the Extended Kalman Filter', *Anal. Chim. Acta*, **167**, 23–37 (1985).
26. S.C. Rutan, S.D. Brown, 'Adaptive Kalman Filtering Used to Compensate for Model Errors in Multicomponent Methods', *Anal. Chim. Acta*, **160**, 99–119 (1984).
27. C.L. Erickson, M.J. Lysaght, J.B. Callis, 'Relationship Between Digital Filtering and Multivariate Regression in Quantitative Analysis', *Anal. Chem.*, **64**, 1155A–1163A (1992).
28. A.G. Marshall, F.R. Verdun, *Fourier Transforms in NMR, Optical and Mass Spectrometry: A User's Handbook*, Elsevier, Amsterdam, 1990.
29. M. Cartwright, *Fourier Methods for Mathematicians, Scientists and Engineers*, Ellis Horwood Ltd, Chichester, 1990.
30. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, New York, 1992.
31. R.P. Wayne, 'Fourier Transformed', *Chem. Britain*, **23**, 440–446 (1987).
32. R.B. Lam, R.C. Wieboldt, T.L. Isenhour, 'Practical Computations with Fourier Transforms for Data Analysis', *Anal. Chem.*, **53**, 889A–901A (1981).
33. G. Horlick, 'Digital Data Handling of Spectra Utilizing Fourier Transforms', *Anal. Chem.*, **44**, 943–947 (1972).
34. G. Horlick, 'Fourier Transform Approaches to Spectroscopy', *Anal. Chem.*, 61A–66A (1971).
35. S. Mallat, 'A Theory for Multiresolution Signal Decomposition: The Wavelet Representation', *IEEE Trans. Pattern Anal. Machine Intell.*, **11**, 674–693 (1989).
36. C.K. Chui, *Introduction to Wavelets*, Academic Press, Boston, 1991.
37. B. Walczak, D.L. Massart, 'Noise Suppression and Signal Compression Using the Wavelet Packet Transform', *Chemom. Intell. Lab. Syst.*, **36**, 81–94 (1997).
38. R.R. Coifman, M.V. Wickerhauser, 'Entropy-based Algorithms for Best Basis Selection', *IEEE Trans. Inform. Theory*, **38**, 713–719 (1992).
39. M.V. Wickenhauser, *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters, Ltd, Wellesley, MA, 1994.
40. J. Rissanen, 'A Universal Prior for Integers and Estimation by Minimum Description Length', *Ann. Stat.*, **11**, 416–431 (1983).
41. J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific Publishing, Singapore, 1989.
42. D.L. Donoho, 'De-noising by Soft Thresholding', *IEEE Trans. Inform. Theory*, **41**, 613–627 (1995).

43. V.J. Barclay, R.F. Bonner, I.P. Hamilton, 'Application of Wavelet Transforms to Experimental Spectra: Smoothing, Denoising, and Data Set Compression', *Anal. Chem.*, **69**, 78–90 (1997).
44. A.G. Marshall, *Fourier, Hadamard, and Hilbert Transforms in Chemistry*, Plenum Press, New York, 1982.
45. D.K. Graff, 'Fourier and Hadamard: Transforms in Spectroscopy', *J. Chem. Ed.*, **73**, 304–309 (1995).
46. P.J. Treado, M.D. Morris, 'A Thousand Points of Light: The Hadamard Transform in Chemical Analysis and Instrumentation', *Anal. Chem.*, **61**, 723A–734A (1989).
47. M. Harwit, N.J.A. Sloane, *Hadamard Transform Optics*, Academic Press, New York, 1979.
48. K.R. Beebe, R.J. Pell, M.B. Seasholtz, *Chemometrics: A Practical Guide*, Wiley, New York, 1998.
49. H. Martens, T. Næs, *Multivariate Calibration*, Wiley, New York, 1989.
50. M.A. Sharaf, D.L. Illman, B.R. Kowalski, *Chemometrics*, Wiley, New York, 1986.

Soft Modeling of Analytical Data

A. de Juan, E. Casassas, and R. Tauler
University of Barcelona, Barcelona, Spain

1 Introduction 1
2 Soft-modeling Data Analysis Based on Factor Analysis 1
3 Multivariate Curve Resolution: General Background 9
4 Noniterative and Iterative Curve Resolution Techniques 11
5 Three-way Resolution Methods 17
6 Examples of Application 21
6.1 Example 1: Chromatographic Co-elution Using Liquid Chromatography/Diode-array Detection 21
6.2 Example 2: Infrared Spectral Data Analysis from Multiple Runs of an Industrial Process 24
6.3 Example 3: Interpretation of Thermodynamic and Conformational Transitions of Polynucleotides 27
Abbreviations and Acronyms 35
Related Articles 35
References 35

In this contribution different methods for soft modeling data analysis are reviewed. Soft modeling approaches attempt the description of a system without the need of an a priori model postulation, physical and/or chemical. The goal of these methods is the explanation of data variance using the minimal or softer assumptions about data. Most of these soft modeling approaches are based on factor analysis (FA) decompositions of experimental data matrices. These decompositions are done by pure mathematical means and allow the identification of the number of data variance sources, their qualitative and, eventually, quantitative estimation. Results of soft modeling data analysis are useful to validate hard modeling results and also for investigation of complex chemical systems. In this contribution, the soft modeling data analysis methods described can be applied to one data matrix or to several data matrices (three-way data sets). The purpose of these methods are mainly exploratory analysis and resolution of mixture data sets. Within the latter group, special attention is devoted to multivariate curve resolution (MCR) techniques and their extension to three-way data analysis. Three examples of application are given covering the chromatographic coelution of mixtures of pesticides using liquid chromatography/diode-array detection (LC/DAD), the infrared (IR) spectral data analysis from multiple runs of an industrial process and the interpretation of thermodynamic and conformational transitions of polynucleotides using spectrometric titrations.

1 INTRODUCTION

Data modeling and data fitting in chemical sciences has been traditionally done by hard modeling techniques, i.e. data are tested against a model based on physical and chemical laws and the parameters of this model are obtained by least-squares curve fitting optimization techniques. This approach is valid for well-known phenomena and laboratory data, where the variables of the model are under control during the experiments and only the phenomena under study affect the data. However, this ideal situation is not obeyed in many circumstances in chemistry, specially in analytical chemistry, when natural samples or unknown processes are investigated. Complex phenomena like those involving macromolecular compounds or industrial processes, where physical parameters cannot be appropriately fixed, are typical examples not solved by the traditional model-based data treatments. Alternative approaches to hard modeling have been proposed. In particular, soft modeling approaches attempt the description of a system without the need of an a priori model postulation, physical and/or chemical. The goal of these methods is the explanation of data variance using the minimal or softer assumptions about data. Most of these soft modeling approaches are based on FA decompositions of experimental data. These decompositions are done by pure mathematical means and allow the identification of the number of data variance sources and often their qualitative and, eventually, quantitative estimation. Results of soft modeling data analysis are useful to validate hard modeling results and also for investigation of complex chemical systems. In this contribution, soft modeling data analysis based on FA techniques and, especially, MCR techniques are reviewed and some examples of application are given at the end of the chapter.

2 SOFT-MODELING DATA ANALYSIS BASED ON FACTOR ANALYSIS

Dealing with huge data sets is a common situation in analytical chemistry. Indeed, the characterization of a

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

sample involves usually the measurement of several independent parameters (e.g. concentrations of constituents, pH, temperature, etc.) or the recording of a multivariate instrumental response (e.g. a spectrum). Modern instrumentation able to acquire and store data easily and powerful computers that can handle efficiently large amounts of information have promoted the transition from the univariate to the multivariate domain. However, enormous amounts of raw experimental measurements reveal little about a chemical problem unless they are treated with the appropriate data analysis method.

In this and subsequent sections, the explanations are focused on data sets that can be organized in a table or, in mathematical words, in a data matrix. These data matrices can be the result of organizing in rows the multivariate response of a series of samples or the responses recorded at different stages of a chemical reaction (see Figure 1). The general terminology adopted calls *objects* the parameters that define the rows and *variables* the parameters that define the columns in a data matrix.

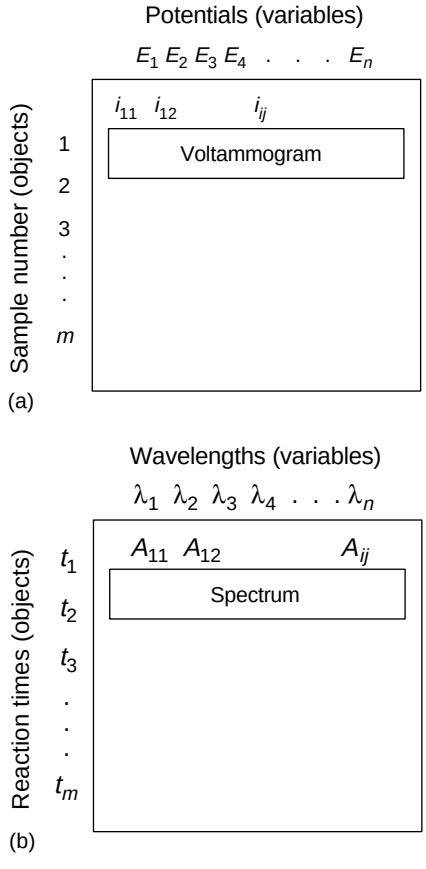

Figure 1 Examples of multivariate data sets. (a) Matrix of samples analyzed voltammetrically. i_{ij} is an intensity measurement collected for the ith sample at the jth potential. (b) Spectrometric monitoring of a kinetic process. A_{ij} is an absorbance reading recorded at the ith reaction time for the jth wavelength.

FA started as a method to find out the underlying structure in these data matrices, i.e. to know how many sources of variation (*factors*) are needed to describe the data sets and, if possible, to identify the chemical nature of these factors.[1] The first step in this process is called abstract FA or principal component analysis (PCA) and establishes a bilinear model of variation in the data set by using abstract variables. The second and more important step links these abstract factors with chemical sources of variation and can be done using several procedures that will be described below.

The first problem associated with large data sets is "how to look at them". The observation of hundreds or even thousands of numerical values does not provide any straightforward information, but displaying the raw data in plots does not seem possible due to the size of the data space (e.g. displaying samples whose spectra have 100 absorbance readings would imply drawing a plot with 100 axes, when the human eye can cope with, at most, three-dimensional figures). Trying to find out methods to visualize multidimensional data spaces, the next question is if all the variables are necessary to describe properly the variation in our data set. There are often many original variables that are clearly correlated and, as a consequence, most of the information that they provide is redundant. In view of this fact, a plot of the data set could be drawn either using the most representative and uncorrelated original variables[2] or replacing the pool of original variables by a reduced set of efficient abstract variables, calculated as a linear combination of the original ones. These new variables, which are completely uncorrelated and describe the main directions of variation of the data set, are the so-called *principal components*.

Figure 2 shows graphically how principal components are designed with a simple example. The data set used consists of a series of spectra for solutions of a pure substance with different concentrations recorded at three different wavelengths. All the points plotted lie in a straight line because spectra of the same pure substance differ only in intensity and not in shape. It can be seen that the information contained in the plot of the spectra in the original space could be kept using only one axis drawn in the direction of maximum variation of the data set. This axis is the first principal component (PC_1) and allows the reduction of the dimensions of the data space from three to one (note that if the spectra for the pure substance had been recorded at m wavelengths, the reduction in dimensions would have gone from m to one). In general, for complex systems, more than one principal component is needed, but the reduction of dimensions from the original space of variables to the space of principal components is always drastic.

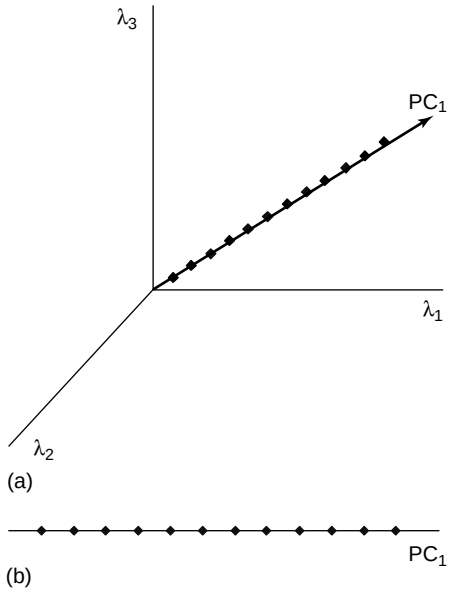

Figure 2 Graphical representation of a data set formed by solution spectra of a pure substance recorded at three wavelengths. (a) Plot in the original data space. (b) Plot in the space of principal components.

The principal components are axes (vectors) which are orthogonal to each other, i.e. completely uncorrelated (eigenvectors). Each new principal component is calculated so as to describe the largest direction of variation in the data set that has not yet been described by other principal components. This means that they decrease gradually in importance because every time the remaining variation in the data set is less. Given a data set with m rows and n columns ($m \times n$), the number of principal components that can be calculated is m or n, whichever is the smallest; however, only a few of them are needed to obtain a good picture of the variation in the original matrix. Thus, for any data matrix, the principal components can be divided into two groups: those which account for the chemical variations in the data set, which are calculated first and are usually few; and those which are related to the noise description. Only the first group will be used to model the data matrix.

The concepts that relate the objects and the variables of the original data matrix to the space of principal components are the *scores* and the *loadings*, respectively. The scores are the coordinates of each object in the principal component axes. The loadings express how large the contribution of each original variable is to the principal components. As mentioned above, principal components are obtained as weighted sums of the original variables. These weights are the loadings and their magnitude (in absolute value) is proportional to the importance they have in the description of the principal component. A variable with large loadings for a certain principal component is significantly correlated to this abstract variable, i.e. their directions in the data space are similar.

Figure 3 shows the PCA results of a real example related to an environmental problem.[3] The data set in this case is formed by 22 sediment samples collected in the northwestern Mediterranean Sea (objects) for which the concentration of 96 compounds of anthropogenic and biogenic origin (variables) has been determined. The first two principal components explain around 70% of the total variation in the data set. Samples close to each other in the *score plot* of PC_2 versus PC_1 are similar. The PCA allowed the identification of several groups of samples related, in this case, to diverse sampling areas with clearly different compound apportionment. The *loading plot* linked principally the concentrations of compounds coming from land-based pollution sources to the first principal component, whereas those due to atmospheric deposition were mainly related to the second principal component. Correlated variables were close to each other or placed along the same straight line through the origin in the loading plot. The information in the scores plot and the loadings plot can be put together in a *biplot*, which relates the objects and variables of a data set. A biplot helps to characterize the different clusters of objects found in the data set. Thus, in this example, the cluster of samples collected in the Ebro prodelta (region 3 in Figure 3), which extends primarily along the PC_1, is mostly related to the variables associated with this first principal component, namely those compounds coming from land-based pollution sources. The other two groups of samples, collected in the open sea, have an expected larger contribution of compounds coming from the deposition of airborne particles.

Mathematically, the relationship between the original data space and the principal component space is expressed by Equation (1):

$$\mathbf{D} = \mathbf{T}\mathbf{P}^{\mathrm{T}} \tag{1}$$

where $\mathbf{D}$ ($m \times n$) is the original data matrix, $\mathbf{T}$ ($m \times npc$) is the matrix of scores and $\mathbf{P}^{\mathrm{T}}$ ($npc \times n$) is the loading matrix. npc is the number of principal components related to chemical sources of variation in the data set. All columns in the $\mathbf{T}$ matrix are orthogonal to each other and so are all rows in $\mathbf{P}^{\mathrm{T}}$. This is the data decomposition obtained using the nonlinear iterative partial least squares (NIPALS) algorithm,[4] by far the most used when FA was introduced in chemistry (it is important to note that the rows of the $\mathbf{P}^{\mathrm{T}}$ matrix are the eigenvectors obtained from the diagonalization of the covariance matrix $\mathbf{Z} = \mathbf{D}\mathbf{D}^{\mathrm{T}}$). Nowadays, the most popular algorithm is singular value decomposition (SVD),[5] which provides a matrix decomposition closely related to Equation (1), though formally

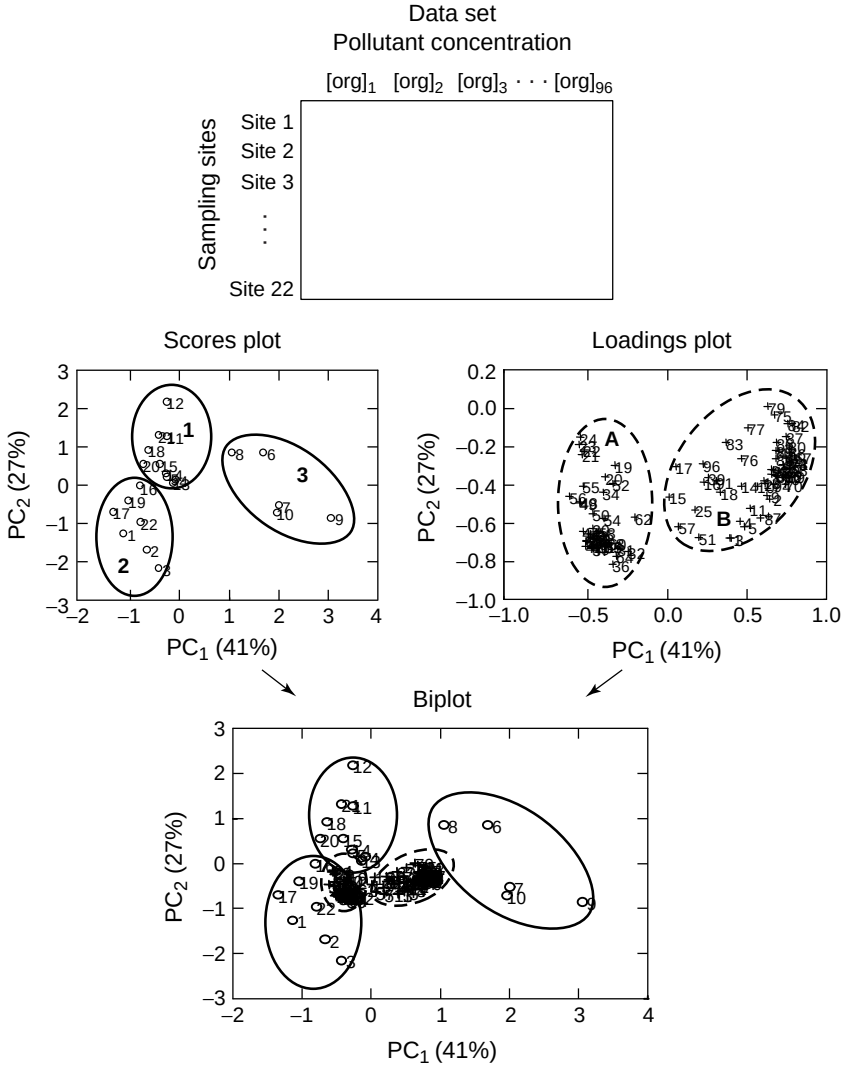

Figure 3 Scores plot, loadings plot and biplot of an environmental data set. Each row in the original data matrix refers to a sampling site (○) and each column to the concentration of an organic compound (+). Figures in the scores plot indicate different geographical zones (1 and 2 are open sea zones, whereas 3 is a river delta). Letters in the loadings plot indicate the origin of the contaminants (A refers mainly to deposited atmospheric pollutants and B mainly to anthropogenic or biogenic pollutants).

different. Equation (2) is provided by SVD:

$$\mathbf{D} = \mathbf{USV}^{\mathrm{T}} \tag{2}$$

where the columns in $\mathbf{U}$ and the rows in $\mathbf{V}^{\mathrm{T}}$ are orthonormal vectors, i.e. unit length orthogonal vectors. The information contained in $\mathbf{U}$ and in $\mathbf{V}^{\mathrm{T}}$ is related to the scores matrix $\mathbf{T}$ and the loading matrix $\mathbf{P}^{\mathrm{T}}$ respectively. $\mathbf{S}$ is a diagonal matrix, whose diagonal elements are the singular values of $\mathbf{D}$, i.e. the square root of the eigenvalues. The mathematical expressions that are generally used to link the NIPALS and SVD results are given in Equations (3) and (4):

$$\mathbf{T} = \mathbf{US} \tag{3}$$

$$\mathbf{V} = \mathbf{P} \tag{4}$$

though some other combinations have also been reported.[6]

The singular values are ranked in descending order of magnitude along the diagonal of the $\mathbf{S}$ matrix and their numerical values are related to the importance of their associated principal component. When the singular values (or the logarithm of the eigenvalues) of a matrix are plotted vs the principal component number, there is often a clear cut-off point between the eigenvalues connected with chemically meaningful factors and those describing noise contributions. This kind of plot is frequently used to find out how many principal components (*factors*) are needed to describe the sources of variation in a data set.

Deduction of this number of principal components is sometimes not an easy task because of the presence

of nonrandom experimental noise. Besides this visual method, several methods have been proposed:[1] methods based on the previous knowledge of experimental error; and approximate methods which do not require previous knowledge of experimental noise. Obviously, the first type of method is preferred when the experimental error is known. However, this knowledge is not available in many circumstances and the methods of the second group should be used. One of the methods which has been found to be useful and does not require previous knowledge of the experimental error is the indicator function (IND) proposed by Malinowski.[7] This function reaches a minimum when the correct number of components has been chosen. It is more sensitive than other functions proposed for the same purpose. There are also other statistically sounder methods, like cross-validation[8] methods. These methods consist of the elimination of a reduced data subset of the whole data set, for instance one row (or several rows) of the data matrix, and the determination of the eigenvalues and eigenvectors of the reduced data matrix which does not have the eliminated rows. All these eliminated rows are then estimated using different numbers of components and the differences between the reproduced values and those which were eliminated are calculated and expressed in a table. The process is repeated for all the rows of the matrix and the prediction errors are summed, giving the function PRESS (prediction error sum of squares). Once this function has been calculated, different approaches can be applied; the simplest one is just plotting it for different numbers of components and looking for a minimum or for a nonstatistical difference between two consecutive numbers of components. Malinowski[9] has also proposed other methods based on the concept of reduced eigenvalue, which theoretically should follow a normal distribution. The significant reduced eigenvalues will be statistically larger than the average variance associated with the error eigenvalues, and therefore, distinguishable from them. Malinowski[1] gives more details about these methods.

The value of PCA for exploratory data analysis has been shown in this section and in many examples published along the history of chemometrics; however, the abstract model provided by PCA is useful for other purposes. Figure 4 shows the bilinear PCA model in detail.

According to Figure 4, the PCA decomposition of a matrix can be written as in Equation (1) or as the sum of bilinear terms as shown in Equation (5):

$$\mathbf{D} = \mathbf{t_1}\mathbf{p_1}^{\mathrm{T}} + \mathbf{t_2}\mathbf{p_2}^{\mathrm{T}} + \cdots + \mathbf{t}_n\mathbf{p}_n^{\mathrm{T}} \qquad (5)$$

In using PCA as the starting point to find the underlying model of a chemical data set, the most important

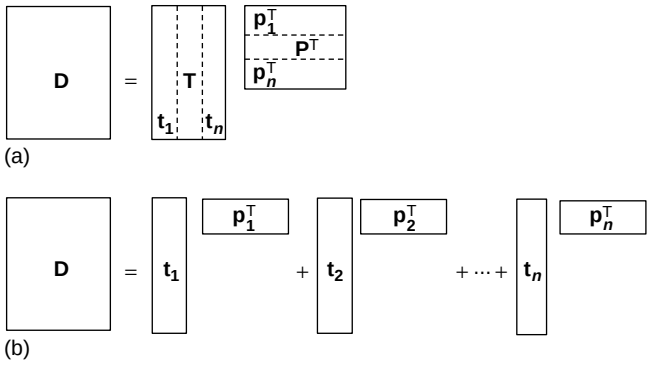

Figure 4 PCA decomposition of a data matrix written as: (a) the product of the matrix of scores (**T**) and the loadings matrix ($\mathbf{P}^{\mathrm{T}}$); and (b) a bilinear model formed by terms described as the outer product of each score vector by its related loading.

requirement is that this set should be intrinsically bilinear, i.e. should have the model form of Equation (5). Happily, many chemical models follow this pattern. A known example is a spectrometric data set, which behaves according to the Beer–Lambert law as shown in Equation (6):

$$\mathbf{D} = \boldsymbol{\varepsilon}_1\mathbf{c}_1 + \boldsymbol{\varepsilon}_2\mathbf{c}_2 + \cdots + \boldsymbol{\varepsilon}_i\mathbf{c}_i \qquad (6)$$

where **D** is a matrix of composite spectra, $\mathbf{c}_i$ is the vector that contains the concentrations of the *i*th component in the different samples and $\boldsymbol{\varepsilon}_i$ is the vector of absorptivities for the *i*th component. As can be seen, Equations (5) and (6) are formally identical. Such a similarity suggests that PCA can be useful to decide which size the simplest bilinear model to describe a data set should have, but once this information is obtained, the chemist may be more interested in a model equally sized where the scores and loadings be replaced by terms with a chemical meaning.

At the beginning of this section, it was pointed out that principal components are orthogonal axes that describe the directions of largest variation in a data set. These directions, however, are not necessarily the most informative. In these cases, the best option is performing a *rotation* of the original axes so that the new axes span the same data space, this being more easily interpretable. For example, for a system with two principal components, these components can be replaced by any other couple of axes which lie on the same plane defined by the two principal components. Some rotation procedures are based on mathematical criteria,[10] such as the VARIMAX rotation, which replaces the principal components by the set of orthogonal axes of maximal simplicity.[11] When two axes are compared, e.g. [0.3 0.3 0.3] and [0 0 1], the simplest one, [0 0 1], has its nonnull loadings grouped over a smaller number of variables. In chemistry, these simple factors

are often closer to the real ones and, when not, they help to simplify the interpretation of the data set.

Despite the benefits provided by VARIMAX and other abstract rotations, the rotated axes can seldom be identified as chemical factors. Whenever this is possible, the use of rotation procedures which take into account chemical information is preferred. This is the basis of target factor analysis (TFA), which replaces the principal component axes by real factors with chemical meaning.[1] This is done after testing individually a pool of potential real factors (*targets*) which may be useful to build a chemical model for the data set.

TFA works by following the steps below:

1. determination of the number of factors of the data set (PCA);
2. selection of potential real factors (targets);
3. target testing;
4. data reconstruction with the targets accepted.

A crucial step in this method is the selection of the targets. The appropriate choice of these vectors and, consequently, the quality of the final model obtained, would largely depend on the chemical knowledge of the data set. The criterion to choose a target can be theoretical or empirical, or simply rely on previous experience but, in all cases, the target selected should potentially justify part of the variation in the data set. There is no limitation on the number of targets to be tested.

Figure 5 shows two real examples of data sets with their associated targets. The example proposed in (a) is a matrix whose columns are mixture spectra. The targets,

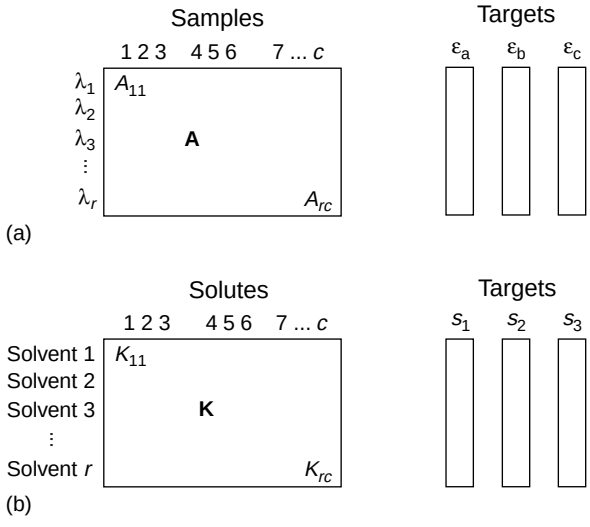

Figure 5 Examples of data sets with their related targets. (a) Matrix of mixed spectra. Targets: pure spectra of potential compounds in the mixture. (b) Matrix of solute properties measured in different solvents. Targets: vectors of solvent properties.

which are proposed to replace the score vectors, would be pure spectra of compounds that may be present in the mixtures. TFA would link the models in Equations (5) and (6).[12] The example in (b) belongs to the field of linear solvation energy relationships (LSER).[13] LSER are linear models that explain solvent-dependent variations in solute properties as a function of changes in certain solvent properties (see Equation 7).

$$\mathbf{XYZ} = (\mathbf{XYZ})_o + \mathbf{s_1a_1} + \mathbf{s_2a_2} + \cdots + \mathbf{s_ia_i} \qquad (7)$$

XYZ is the solvent-dependent solute property, $(\mathbf{XYZ})_o$ is the value of the **XYZ** property in a hypothetical solvent with null interaction with the solute, $\mathbf{s_i}$ is a solvent property responsible for the variation of the solute property and $\mathbf{a_i}$ is the weight coefficient related to the contribution of $\mathbf{s_i}$ in the variation of the solute property **XYZ**. Each column in the data matrix of Figure 5(b) contains the equilibrium constants of a solute determined in different solvents. The targets are vectors of solvent properties that may cause the variation in the equilibrium constants of solutes. TFA links the model in Equation (5) with the model in Equation (7). Apart from the targets with chemical meaning, the *unity target*, which is a vector whose elements are all equal to one, is always tested to account for the possible presence of an offset in the bilinear model.

Once the selection of targets is finished, the next step is assessing if they can be included in the model that describes the variation of the data set. Till this point in the modeling process, PCA has indicated how many real factors should be present in the model and has perfectly defined the data space by means of the score vectors. Any real factor used to replace the scores must belong to the space defined by these abstract vectors. Therefore, testing the goodness of a target is simply finding out how far the vector related to this potential factor is from the space defined by the score vectors.

From a geometrical point of view, target testing consists of projecting the target selected, called hereafter *input target* ($\mathbf{x}_i$), onto the scores space. The projected target, the so-called *output target* ($\mathbf{x}_o$), lies on the scores space. Figure 6 shows the process of target testing for a data set described with two factors. The two principal components define, in this case, a plane.

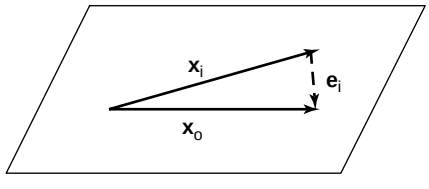

Figure 6 Geometrical representation of the target testing process for a data set described by two factors.

The perfect target would give $\mathbf{x}_i = \mathbf{x}_o$. However, this equality is never fulfilled in practice due to the experimental error in both the data matrix and the target vector. As a consequence, a target is accepted when the input target is close enough to the output target, i.e. when the length of the vector $\mathbf{e}_i$ (*apparent error in the target*) is small enough.[1] The difference between analogous elements of $\mathbf{x}_i$ and $\mathbf{x}_o$ is related to the quality of the target; thus, the smaller the difference the better the target. Nevertheless, this observation alone is not reliable enough to decide whether a target must be accepted. Empirical methods, such as the SPOIL function,[14] and statistical tests, such as an *F*-test that compares the error in the target with the error in the data matrix,[9] have been proposed to decide more soundly the acceptance or rejection of a target.

After the target testing step, several targets have been confirmed as suitable candidates to build the real model of variation of the data set. From all the targets that passed this test, a number equal to the number of factors determined by PCA should be chosen. When the number of targets accepted exceeds the size of the model, the best among them are selected.

In the definite model, the targets accepted replace the scores. To explain how to replace the loadings ($\mathbf{P}^T$) by a matrix of weight coefficients associated with the real targets ($\mathbf{C}^T$ for the model in Equation 6 and $\mathbf{A}^T$ for the model in Equation 7), the close relationship between a PCA model and a real model should be recalled. Taking the LSER example in Equation (7) yields Equation (8)

$$\mathbf{TP}^T = \mathbf{SA}^T \tag{8}$$

where $\mathbf{S}$ is the matrix whose columns are the solvent properties ($\mathbf{s_i}$) accepted in the target testing step. This equation can also be written as Equation (9):

$$\mathbf{TY}(\mathbf{Y}^{-1}\mathbf{P}^T) = \mathbf{SA}^T \tag{9}$$

where

$$\mathbf{TY} = \mathbf{S} \tag{10}$$

$$\mathbf{Y}^{-1}\mathbf{P}^T = \mathbf{A}^T \tag{11}$$

From Equation (10), where $\mathbf{T}$ and $\mathbf{S}$ are known, $\mathbf{Y}$ can be calculated. Once $\mathbf{Y}$ is known, $\mathbf{A}^T$ can be calculated using Equation (11) and the real model is completed. The reproduction of the original data set, $\mathbf{D}$, would be as shown in Equation (12):

$$\mathbf{D_{TFA}} = \mathbf{SA}^T \tag{12}$$

If $\mathbf{D_{TFA}}$ and $\mathbf{D}$ are sufficiently similar, i.e. the differences can be considered as due to experimental error, the real model proposed is suitable to describe the data set. When this does not happen, some targets in $\mathbf{S}$ can be correlated and, then, new models with other accepted targets should be tried.

All the data sets that can be organized in matrices do not have the same features. This is clearly seen when the two data sets in Figure 1 are compared. The data set in Figure 1(a) is formed by voltammograms related to different samples. There is no loss of information if the order of the rows (samples) is exchanged. Figure 1(b) contains the data set obtained in the spectrometric monitoring of a kinetic process. If the rows of the data matrix, which are related to each reaction time, are organized differently, the information about the kinetic evolution is lost. Data sets like the latter describe dynamic systems and are often found in chemistry, e.g. high-performance liquid chromatography/diode-array detector (HPLC/DAD) data, or spectrometric monitoring of thermodynamic, thermal or any other parameter-dependent reaction. For these cases, there are some FA-based techniques, namely evolving factor analysis (EFA) and fixed-size moving window/evolving factor analysis (FSMW/EFA), which are designed to model the evolution of the factors in the data set. The evolution of these abstract factors is connected with the development of the chemical process.

EFA was first proposed by Maeder and co-workers[15–17] to treat spectrometrically monitored equilibria and was afterwards applied to any other kind of data set coming from dynamic chemical processes,[18] such as HPLC/DAD, kinetic transformations, or structural transitions.

The EFA method calculates the eigenvalues obtained from the PCA of gradually growing submatrices of the original data set, $\mathbf{D}$ ($m \times n$). PCA is performed repeatedly in submatrices generated by increasing the size of the previous submatrix by one row. This size increase is performed from the top to the bottom of the original data matrix $\mathbf{D}$ (*forward EFA*) and also in the opposite sense (*backward EFA*). Thus, forward EFA first carries out PCA on rows 1 and 2 of the data matrix, then on rows 1, 2, and 3 and so on until the final PCA on the m rows of the whole data matrix is performed. Backward EFA starts by performing PCA on the rows m and $(m-1)$, then adds the row $(m-2)$ to the analysis and so forth until the final PCA is carried out with the whole data matrix.

The EFA plot displays the evolution of the eigenvalues as the submatrix analyzed grows, i.e. as the evolutionary process in the data matrix goes from the beginning to the end (forward EFA) or vice versa (backward EFA). The log(eigenvalues) are plotted versus the row number of the last row included in the PCA or, better, versus the value of the variable responsible for the evolutionary process related to this row (e.g. a pH value in an acid–base equilibrium or a retention time in an HPLC/DAD

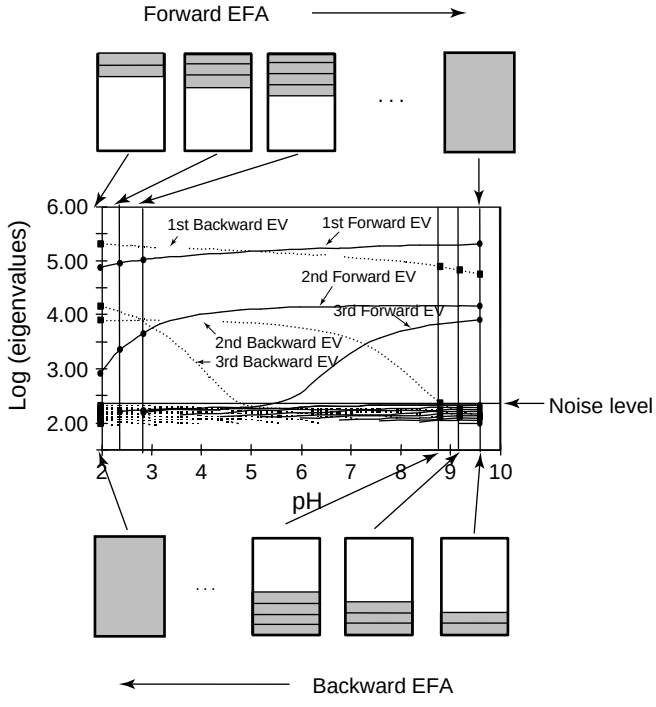

Figure 7 Construction of an EFA plot related to the spectrometric titration of a diprotic acid (simulated data). The continuous lines join the eigenvalues (circles) obtained in each PCA of forward EFA and the dashed lines connect the eigenvalues (squares) obtained in each PCA of backward EFA. EV stands for eigenvalue.

data set). Figure 7 shows the EFA plot related to a matrix formed by data coming from the spectrometric monitoring of the acid–base equilibria of a diprotic acid. Each row in the matrix is a spectrum collected at a certain pH value.

As PCA does, an EFA plot indicates the total number of components in the data set. The eigenvalues related to chemical components are clearly higher than the eigenvalues associated with noise. At the bottom of the plot, the many overlapping noise eigenvalues define a graphical threshold that can be considered the noise level of the system. All the eigenvalues above this limit are chemically significant. The total number of chemical components can be determined from the number of significant eigenvalues obtained in the last PCA performed in both forward and backward data analyses (i.e. with the results in the right and left extremes of the forward and backward EFA plot respectively). In the example of Figure 7, three chemical components, related to each of the species of a diprotic acid, are detected.

In contrast to PCA, an EFA plot also marks the zones of appearance and disappearance of chemical components. The appearance of a new species is detected in the forward EFA through the emergence of a new significant eigenvalue. Thus, when the row added to a certain submatrix includes a new chemical component, the different information introduced by this component will cause a significant increase in the 'until then' first noise eigenvalue. The chemical variable (e.g. pH, time, etc.) corresponding to the new row included in the analysis indicates the chemical conditions in which the new component appears. When performing backward EFA, the process is followed from the end to the beginning. Therefore, the presence of a new significant eigenvalue indicates the row, and consequently the chemical conditions, in which a component disappears. According to the forward EFA results in Figure 7, the two most protonated forms of the diprotic acid are present at pH 2, whereas the third one appears at a pH close to 5.5. Looking at the backward EFA plot, the most deprotonated form is still present at pH 10, the next one disappears at a pH close to 9 and the most protonated at a pH around 5.

The evolution of the eigenvalues in the forward and backward EFA is connected with the formation and decay curves of the chemical components present in the data matrix. When the appearance and disappearance of the chemical components are sequential, as in the example of Figure 7 and other reaction processes, or as in most of the chromatographic elutions, the suitable connections between the lines of the forward EFA plot and the backward EFA plot can provide abstract profiles related to the evolution of each of the chemical components in the data matrix. Thus, for a system with n components, the profile of the first component appearing is built connecting the line of the first eigenvalue of the EFA forward plot with the line of the nth significant eigenvalue of the EFA backward plot (related to the first compound disappearing); the profile of the second component connects the second eigenvalue in the EFA forward plot with the $(n-1)$th eigenvalue of the EFA backward plot and so on. The zone of the matrix (rows) where the concentration profile of one compound shows values different from zero is the *concentration window* related to this compound (see Figure 8). The abstract profiles obtained with EFA are often used as a starting point in the iterative resolution methods described in sections 3 and 4 of this chapter.

The EFA method can be applied to both directions in the data matrix (i.e. the successive PCA can be performed in row- or column-wise growing submatrices of the original data matrix). In any case, the most useful results will always be obtained in the direction where the overlap of the different chemical components is less.

FSMW/EFA was proposed by Keller and Massart[18] as a derivation of EFA. In contrast to EFA, FSMW/EFA

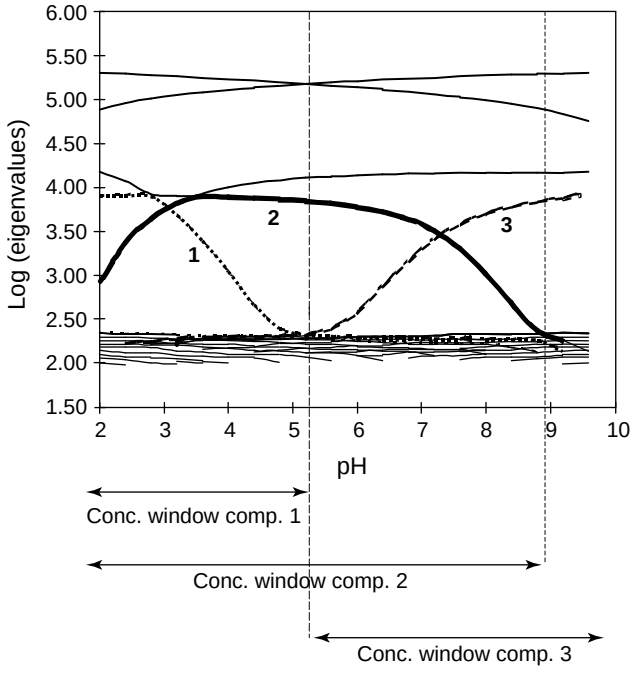

Figure 8 Abstract concentration profiles of a diprotic acid obtained using EFA, with their respective concentration windows.

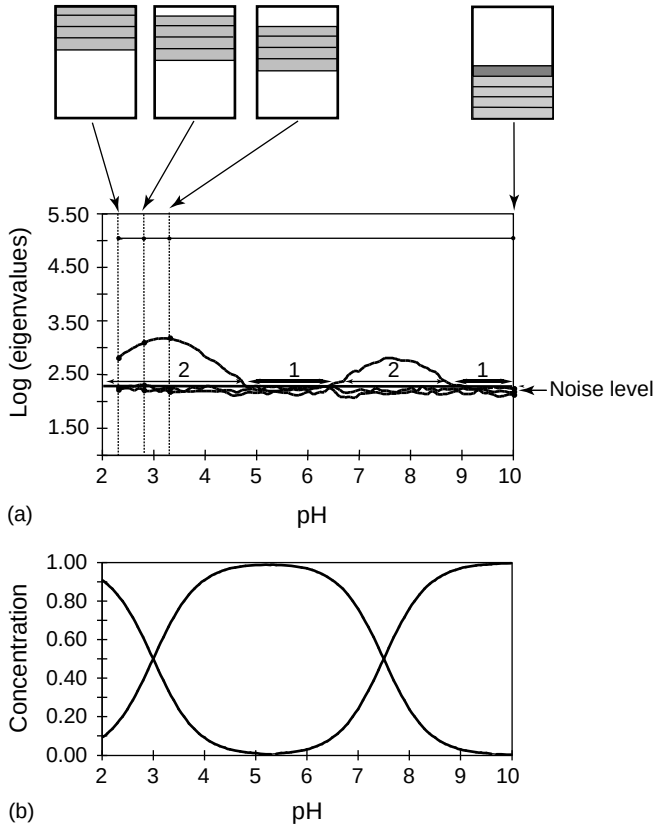

Figure 9 (a) FSMW/EFA plot related to the spectrometric titration of a diprotic acid. The window size is four rows. The figures within the plot indicate the number of species in the zones defined by the lines. The thick lines mark the selective zones. Circles are the EV obtained in the PCA of each submatrix. (b) True distribution plot of the diprotic acid.

scans the original data matrix from top to bottom by performing PCA on ''equally sized'' submatrices. A new submatrix is built by removing the first row of the previous submatrix and adding the following row in the original data matrix, i.e. moving a window of a fixed number of rows one row downwards. The size of the window usually exceeds the number of chemical components by one, though more information can be obtained if FSMW/EFA is applied several times using different window sizes.[19]

A local rank map, i.e. a graph with information about the number of significant principal components (chemical components) in each zone of the data matrix, is obtained by plotting the eigenvalues obtained in each PCA vs the variable responsible for the evolution of the process, as shown in Figure 9. In this plot, the noise level is also defined graphically by the zone in which the noise eigenvalues appear together. The emergence of one significant eigenvalue from the noise level zone indicates the presence of only one species in that zone; the presence of two significant eigenvalues defines the zone where two species coexist, three significant eigenvalues would detect the overlap of three species, and so on.

The local rank map obtained through the application of FSMW/EFA allows the *detection of selective zones*, i.e. zones where only one chemical component is present. The zones where only one significant eigenvalue is present are the selective zones of the species (the detection and location of these regions is essential in the resolution of a data matrix). In Figure 9, two selective zones are detected, one for the most deprotonated form of the diprotic acid and another for the intermediate form of this compound. FSMW/EFA also contributes to the *detection of minor constituents*. Actually, FSMW/EFA was initially proposed to deal with peak purity problems in chromatography. The local analysis of the data matrix and the clear graphical information help in the detection of minor species, although they are embedded under major constituents.

3 MULTIVARIATE CURVE RESOLUTION: GENERAL BACKGROUND

All the resolution methods were born as a tool to analyze multivariate experimental data coming from multicomponent dynamic systems.[20–27] The common goal for all these methods is to mathematically decompose the global

instrumental response into the pure contributions due to each of the components in the system. The use of such methods has become valuable when obtaining individual signals experimentally is not possible or when this process is too complex or too time-consuming.

The multivariate output of an experiment monitoring a dynamic process is organized in a double-structured matrix **D** containing mixed information about the evolution of all the components present in the successive stages of the chemical process. The ultimate goal of the curve resolution (CR) methods is the decomposition of the initial mixture data matrix **D** into the product of two data matrices **C** and **S**, each of them including the pure response profiles of the *n* mixture components associated with one of the directions of the initial data matrix (see Figure 10).

In matrix notation, Equation (13) gives the general expression valid for all CR procedures:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \tag{13}$$

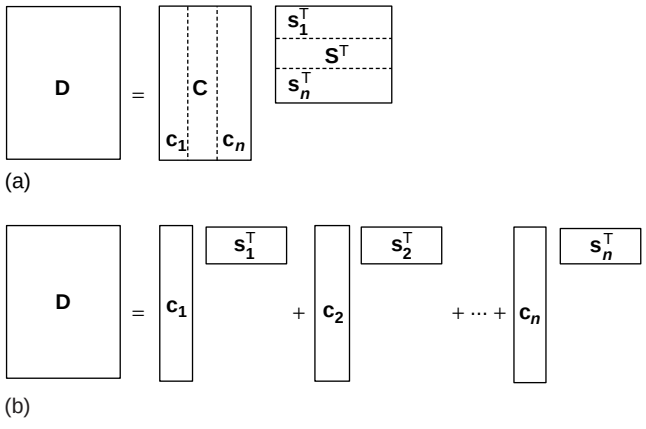

Figure 10 Two graphical views of the decomposition of a multicomponent data matrix: (a) as the product of the matrices including the pure response profiles; (b) as the sum of the products related to each pure contribution of the mixture components.

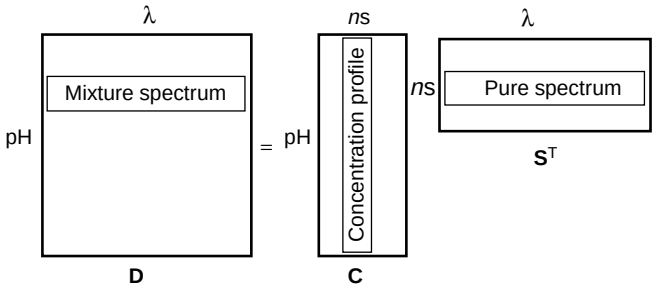

Figure 11 Results obtained after the application of a CR method on a data matrix coming from a pH-dependent process monitored spectrometrically. *n*s means number of absorbing species.

where **D** ($r \times c$) is the original data matrix, **C** ($r \times n$) and $\mathbf{S}^{\mathrm{T}}$($n \times c$) are the matrices containing the pure response profiles related to the data variation in the row direction and in the column direction respectively, and **E** ($r \times c$) is the error matrix, i.e. the residual variation of the data set that is not related to any chemical contribution. *n* is the number of chemical components in matrix **D**. Taking as an example a pH-dependent process monitored spectrometrically, the **C** matrix would contain the pure concentration profiles of all the absorbing species and the $\mathbf{S}^{\mathrm{T}}$ matrix would be formed by their related pure spectra (see Figure 11). For the sake of simplicity, **C** and $\mathbf{S}^{\mathrm{T}}$ are referred to as concentration profile matrix and spectra matrix, though this does not mean that the applicability of CR methods is restricted to this kind of chemical data.

The mathematical decomposition of a single data matrix is inherently affected by two sources of ambiguity: rotational ambiguity and intensity ambiguity. Whereas the former accounts for the possibility of reproducing correctly the original data matrix by using **C** and **S** matrices containing linear combinations of the true profiles, the latter warns about the possibility of having profiles equal in shape to the true ones, though different in magnitude. In other words, the correct reproduction of the original data matrix can be achieved by using response profiles differing in shape (rotational ambiguity) or in magnitude (intensity ambiguity) from the true ones.

The explanation of these two ambiguities is simple. The basic equation associated with CR methods, $\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}}$, can be easily transformed as shown in Equations (14–16):

$$\mathbf{D} = \mathbf{C}(\mathbf{T}\mathbf{T}^{-1})\mathbf{S}^{\mathrm{T}} \tag{14}$$

$$\mathbf{D} = (\mathbf{C}\mathbf{T})(\mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}}) \tag{15}$$

$$\mathbf{D} = \mathbf{C}'\mathbf{S}'^{\mathrm{T}} \tag{16}$$

where $\mathbf{C}' = \mathbf{C}\mathbf{T}$ and $\mathbf{S}'^{\mathrm{T}} = (\mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}})$ describe the **D** matrix as correctly as the true **C** and $\mathbf{S}^{\mathrm{T}}$ matrices do, though $\mathbf{C}'$ and $\mathbf{S}'^{\mathrm{T}}$ lack chemical sense. On the basis of the transformation shown in these equations, the mathematical formulation of the rotational ambiguity problem indicates that the possible solutions of a resolution method are as numerous as the **T** matrices can be, i.e infinite. However, the inclusion of information related to the internal structure of the data (e.g. the presence of selective zones) and to their chemical properties in the resolution process often allows the suppression of this ambiguity or, at least, a large decrease in the number of feasible solutions.

When a system lacking rotational ambiguity is considered, the basic CR equation can still be rewritten as shown in Equations (17) and (18):

$$\mathbf{D} = \left(\frac{1}{k}\right. \quad \mathbf{C}k\mathbf{S}^{\mathrm{T}} \tag{17}$$

$$\mathbf{D} = \mathbf{C}'\mathbf{S}'^{\mathrm{T}} \tag{18}$$

where k is a scalar. The concentration profiles of the new $\mathbf{C}' = (1/k)\mathbf{C}$ matrix have the same shape as the real ones, but are k times smaller, whereas the spectra of the new $\mathbf{S}' = k\mathbf{S}$ matrix are shaped like the $\mathbf{S}$ spectra, though k times more intense. This ambiguity cannot be solved unless external information is introduced in the resolution process. Both rotational and intensity ambiguities are drastically diminished when several matrices (*three-way data sets*) are analyzed together.

The correct performance of the CR methods depends strongly on the internal features of the data set being analyzed, specially on selectivity and local rank.[28] Regardless of the quality of each CR method, the two following conditions must be fulfilled if the true concentration profile and spectrum of each compound in the data matrix are to be recovered:[29]

- The true concentration profile of a compound can be recovered when all the compounds inside its concentration window are also present outside.
- The true spectrum of a compound can also be recovered if its concentration window is not completely embedded inside the concentration window of a different compound.

The same formulation written above holds when, instead of looking at the concentration windows (windows of rows), the 'spectral' windows (windows of columns) are considered. The content of these theorems supports the fact that the goodness of the resolution result depends more strongly on the features of the data set (particularly those related to selectivity and local rank) than on the mathematical background of the CR method. Therefore, a good knowledge of the properties of the data sets before carrying out a resolution process provides a clear idea about the quality of the results that can be expected.

Seeing how close the PCA and the CR decomposition of a matrix are formally, it is not surprising that some CR methods, such as window factor analysis (WFA)[30] or heuristic evolving latent projections (HELP),[31] work with the abstract variables obtained by PCA in the resolution process. Some other methods transform initial estimates into real solutions projecting them iteratively onto the scores space, such as iterative transformation target factor analysis (ITTFA).[24,32] Methods like multivariate curve resolution/alternating least squares (MCR/ALS)[33] are linked to PCA because they often use initial estimates obtained by PCA-based techniques, such as EFA. Initial estimates can be also obtained by some other methods that use real variables, such as the simple-to-use interactive self-modeling analysis (SIMPLISMA)[26] or the orthogonal projection approach (OPA).[34] The CR methods can also be classified according to the data sets they can deal with. Thus, some methods are applied to the resolution of single matrices, such as HELP, WFA and ITTFA, whereas others can deal with two (generalized rank annihilation method, GRAM)[35] or more matrices together (direct trilinear decomposition (DTD),[36] parallel factor analysis (PARAFAC),[37–39] and restricted Tucker models).[40] Methods like MCR/ALS are adapted to one or more matrices.[41–43]

4 NONITERATIVE AND ITERATIVE CURVE RESOLUTION TECHNIQUES

As mentioned in section 3, the goal of all CR is the decomposition of a mixture data matrix $\mathbf{D}$ into the product of the small matrices, $\mathbf{C}$ and $\mathbf{S}^{\mathrm{T}}$, which contain profiles related to the evolution along the rows and along the columns of each pure component in $\mathbf{D}$. The noniterative resolution methods obtain $\mathbf{C}$ and $\mathbf{S}^{\mathrm{T}}$ in one calculation step, whereas iterative resolution methods refine the profiles in $\mathbf{C}$, in $\mathbf{S}^{\mathrm{T}}$ or in both matrices at each iterative cycle till an optimal solution is attained.

Noniterative resolution methods are the fastest, but they often require that the data set has certain features for them to be applied (e.g. WFA is meant to work with evolutionary data where each chemical compound has a unique maximum in its concentration profile). The user intervention is less, but usually more critical; wrong decisions cannot be changed in a one-step calculation method. Nevertheless, when the data sets fulfill the conditions needed by the CR method, these resolution procedures are a good alternative. The information provided by the global and local application of PCA to the data set is essential in most of these noniterative methods; thus, the quality of the final solutions depends basically on the correct determination of the total number of chemical components in the data set and on the construction of a reliable local rank map.

Within the group of noniterative resolution methods, HELP[31] and WFA[30] are probably the most well known. The HELP method is based on the local rank analysis of the data set and focuses on finding selective concentration and/or spectral windows. When these selective zones exist, the resolution of the system is clear. Thus, for a data set related to the spectrometric monitoring of a kinetic process (see Figure 1b), the row related to a selective reaction time directly provides the shape of the spectrum of the only component present at that stage of the reaction and the column related to a selective wavelength directly provides the kinetic concentration profile of the only absorbing compound at that wavelength. The main contributions of HELP have been offering a sophisticated graphical tool (the so-called datascope) to visually detect potential selective zones in the score plot of the data matrix and a statistical method to confirm the presence of

selectivity in the concentration and/or spectral windows graphically chosen. The statistical method is based on the use of an *F*-test to compare the magnitude of eigenvalues related to selective zones of the data set with eigenvalues related to noise zones of the data matrix, i.e. those where no chemical components are supposed to be present. The confirmation of a selective zone in the data set, which is actually a rank-one window in the data matrix, will then be obtained when no significant differences are found among the first eigenvalue of a noise-related zone of the data matrix and the second eigenvalue of the potential selective zone. The statistical part of the method can be negatively affected by the presence of nonchemical significant variations in the data, such as instrumental drift or nonrandom noise. It should be also noted that the application of this method requires that the data set has zero-component zones, which is easy in an HPLC/DAD data set (time windows with no compounds eluting are commonly found) but not so evident when a reaction monitoring is carried out (there is always some chemical compound in the reaction vessel), for instance.

WFA seems to have a wider field of applications than HELP and improved versions of the algorithm have recently been published.[44] This method inspired in the rank annihilation evolving factor analysis (RAEFA)[25] was proposed by Malinowski.[30,45,46] WFA is applicable to data sets where the formation and decay of the compound profiles in one of the directions of the matrix (usually the concentration direction) is sequential, i.e. the compound appearing in the *n*th position disappears in the *n*th position. As a consequence, out of the concentration window of a certain compound all the other compounds in the data set are present.

WFA recovers each true concentration profile as follows (see Figure 12):

1. PCA of the original data matrix, **D**.
2. Determination of the concentration window of each component.
3. PCA of a matrix, $\mathbf{D^o}$, where the rows related to the concentration window of the *n*th compound have been removed.
4. Calculation of the vector, $\mathbf{p}_n^{oT}$, which is the part of the spectrum of the *n*th component orthogonal to the spectra of the other components.
5. Recovery of the true concentration profile of the *n*th component using $\mathbf{p}_n^{oT}$ and **D**.
6. Calculation of the spectra matrix $\mathbf{S}^T$ by least squares using **D** and **C**.

The mathematical formulation associated with this method is not trivial and only a rough intuitive explanation will be given in this section. A detailed description can be found in the literature.[30]

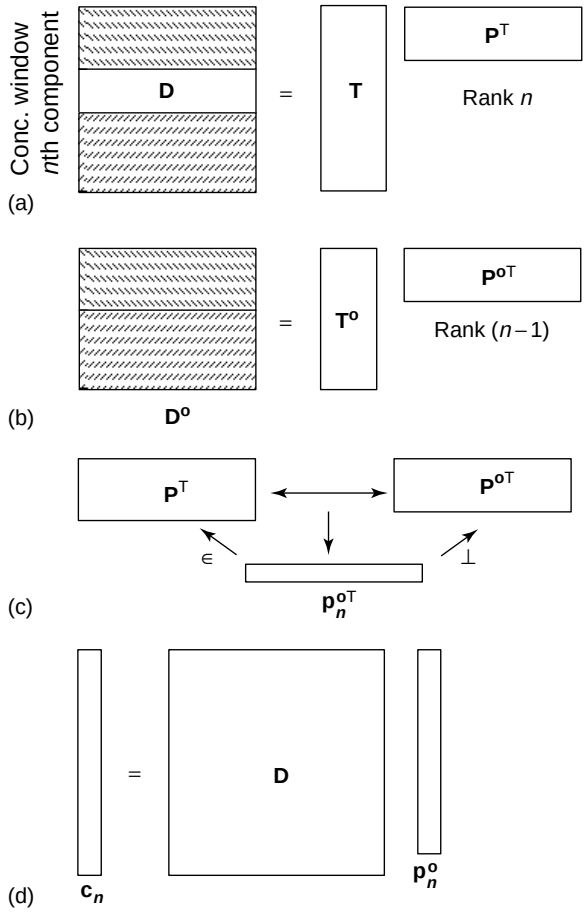

Figure 12 Recovery of the concentration profile of the *n*th compound by WFA. (a) PCA of the raw data matrix, **D**; (b) PCA of the matrix formed by suppression of the concentration window of the *n*th component, $\mathbf{D^o}$; (c) recovery of the component of the spectrum of the *n*th component orthogonal to all the spectra in $\mathbf{D^o}$, $\mathbf{p}_n^{oT}$; and (d) recovery of the concentration profile of the *n*th component.

WFA starts with the PCA decomposition of the **D** matrix into the product $\mathbf{TP}^T$ (remember Equation 1). In a general case, the **D** matrix can have *n* components, i.e. rank *n*. The definition of the concentration windows for each component is carried out later using EFA (see Figure 8) or other methods. Steps 3 to 5 are the core of the WFA method and should be performed as many times as compounds are present in matrix **D** to recover every time one concentration profile of the **C** matrix.

For each component, a $\mathbf{D^o}$ matrix is built removing the rows related to its concentration window. Then, a PCA is performed and the product $\mathbf{T^oP^{oT}}$ is obtained. Note that the $\mathbf{D^o}$ matrix has rank $n-1$ because the variation due to one of the components in the data set disappears when the related rows of the data matrix **D** are deleted. The loading matrices, $\mathbf{P}^T$ and $\mathbf{P^{oT}}$, describe the space of the n pure spectra in **D** and the $(n-1)$

pure spectra in $\mathbf{D^o}$ respectively. The rows in these loading matrices are actually ''abstract spectra'' and the real spectra can be expressed as a linear combination of them. The vectors either in $\mathbf{P}^T$ or in $\mathbf{P^{oT}}$ could be interchangeably used to build these linear combinations if the information related to the removed component was present in $\mathbf{P^{oT}}$. Using these two loading matrices, it is possible to calculate a vector $\mathbf{p}_n^{oT}$, which is orthogonal to the $(n-1)$ $\mathbf{p}_i^{oT}$ vectors and which belongs to the space defined by $\mathbf{P}^T$. This vector completes the set of vectors in $\mathbf{P^{oT}}$ and contains the part of the spectra of the removed component which is uncorrelated to the spectra of the other $(n-1)$ components in the data matrix. Using this vector with information exclusively related to the removed component, the true concentration profile of this compound can be calculated as shown in Equation (19):

$$\mathbf{D}\mathbf{p}_n^o = \mathbf{c_n} \tag{19}$$

The complete $\mathbf{C}$ matrix is then obtained by appending row-wise the concentration profiles found for each compound in the $\mathbf{D}$ matrix. Using the basic equation of CR methods, $\mathbf{D} = \mathbf{CS}^T$, the matrix of spectra, $\mathbf{S}^T$, is obtained by least squares using the $\mathbf{D}$ and $\mathbf{C}$ matrices as shown in Equation (20):

$$\mathbf{S}^T = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{D} \tag{20}$$

The main drawbacks of WFA are the impossibility of solving data sets with nonsequential profiles (e.g. data sets with embedded profiles, for instance) and the dangerous effects of a bad definition of the concentration windows. Tackling this last point has been the main goal of recent modifications of this algorithm.[44]

Despite the limitations associated with the noniterative methods, the information provided by local PCA performed in selected windows of the data matrix still inspires new algorithms for resolution, such as the very recent subwindow factor analysis (SFA), proposed by Manne et al.[47] based on the comparison of matrix windows sharing one compound in common.

Iterative resolution methods are in general more versatile than noniterative methods. They apply to more diverse data (e.g. with sequential and nonsequential profiles, with different degrees of selectivity, etc.) and the previous knowledge about the data set (chemical or related to mathematical features) can be used in the optimization process. The main complaint about iterative resolution methods has often been the larger calculation times required to obtain the optimal results; however, improved fast algorithms[48,49] and more powerful personal computers have overcome this historical limitation.

All iterative resolution methods refine the profiles in $\mathbf{C}$, in $\mathbf{S}^T$ or in both matrices at each step of the optimization process. The profiles in $\mathbf{C}$ and/or $\mathbf{S}^T$ are ''tailored'' according to the chemical properties and the mathematical features of each particular data set. The iterative process stops when a convergence criterion is fulfilled (e.g. a preset number of iterative cycles is exceeded or the lack of fit goes below a certain value).

Starting the iterative optimization of the profiles in $\mathbf{C}$ or $\mathbf{S}^T$ requires a matrix sized as $\mathbf{C}$ or as $\mathbf{S}^T$ with more or less rough approximations of the concentration profiles or spectra that will be obtained as the final results. This matrix contains the *initial estimates* of the resolution process. Though some authors are in favor of using random initial estimates,[50] the most generalized opinion tends to recommend the use of nonrandom estimates to shorten the iterative optimization and to avoid the convergence to local optima different from the solution searched. If the initial estimates are either a $\mathbf{C}$-type or an $\mathbf{S}^T$-type matrix can depend on which kind of profiles are less overlapped, on which direction of the matrix (rows or columns) has more available information or simply on the will of the chemist. There are many chemometric methods to build initial estimates: some are especially suitable when the data consist of evolutionary profiles, such as EFA (see Figure 8), whereas some others mathematically select the purest rows or the purest columns of the data matrix as initial profiles. Within the latter, key set factor analysis (KSFA)[51] works in the FA abstract domain and some other procedures, such as SIMPLISMA[26] and OPA,[34] work with the real variables in the data set and select the rows or the columns most representative and most dissimilar to each other. Besides the use of chemometric methods, a matrix of initial estimates can always be formed including the rows or columns of the data set that the researcher considers most representative for chemical reasons.

The general process in an iterative resolution method goes on with the optimization of the initial profiles with the help of some selected constraints. A constraint can be defined as any systematic feature in the data set (mathematical or chemical) translated into mathematical language. Therefore, the effectiveness of a constraint can change depending on the way it has been implemented and active research is in progress to optimize this point.[52–55] In any profile, a constraint works by updating the elements that do not allow the fulfillment of a certain condition by some others that do. The use of constraints is optional and should be adapted to each particular data set. Small departures from the conditions imposed by some constraints are often allowed.

The most essential constraint in the resolution process is *selectivity*.[28] The selective zones in a data matrix are those regions (row windows or column windows) where only one species is present, i.e. those with rank one; therefore, this is a constraint related to a mathematical feature. As mentioned in the brief comment about the

HELP method (see this section, above), the presence of selective zones for all the species in a data matrix eliminates the rotational ambiguity and ensures the recovery of the real response profiles of the chemical system. Local rank analysis methods, such as EFA and FSMW/EFA, are the most suitable techniques to detect and locate the selective zones of a system, as shown in section 2. The selectivity for one compound can be forced by setting to zero the elements of all the other response profiles in the selective regions of this compound.

The most frequent constraints related to chemical features of the profiles are described below:

1. Non-negativity. This constraint forces the values in a profile to be equal to or greater than zero and is applied to all the concentration profiles and to some experimental responses, such as ultraviolet (UV) absorbances.
2. Unimodality. This constraint allows the presence of only one maximum per response profile. It is applied to chromatographic peaks, to the concentration profiles of some chemical reactions and to some peak-shaped instrumental responses, such as voltammograms.
3. Closure. This constraint is applied to closed reaction systems for which the sum of the concentrations of all the species involved in the reaction or the sum of some of them is forced to be constant at each stage of the reaction. Actually, closure is a mass balance constraint.

A possible modification of some simulated profiles after the application of the constraints above is graphically shown in Figure 13.

Two of the most representative iterative resolution methods are ITTFA[24,32,56] and MCR/ALS.[27,28,41–43] Whereas ITTFA optimizes either the **C** profiles or the $\mathbf{S}^T$ profiles in each iterative cycle, MCR/ALS modifies the profiles in both matrices in each step of the optimization. An additional difference is that ITTFA works by optimizing one profile at a time, whereas MCR/ALS refines together all the profiles in the constrained data matrices.

As the name suggests, ITTFA is based on TFA (see section 2). In TFA, some vectors (targets) are tested to see if they can be used to describe real sources of variation in the data set. These vectors have a chemical meaning and are perfectly characterized. In facing a resolution problem, the appropriate targets should be either potential concentration profiles or instrumental responses (spectra). However, in practice, the user can, at most, know exactly some of these real profiles and, in some situations, none of them. Therefore, the straightforward application of TFA to solve completely a resolution problem is not possible. ITTFA borrows two main ideas from TFA: the fact that the space of either the concentration profiles or the spectra can be perfectly known and the initial use of a target to finally obtain a true profile of the **C** or the $\mathbf{S}^T$ matrix. Essentially, what ITTFA does is to modify a target until it lies on the real space of concentrations or spectra and fulfills the appropriate constraints. This process is repeated with as many targets as the data set has components.

It has been pointed out that ITTFA works by optimizing one at a time the profiles in either the **C** matrix or the $\mathbf{S}^T$ matrix. The direction of optimization will depend on which information is considered most important by the chemist: e.g. in an HPLC/DAD run, getting the chromatographic peaks of each compound may be the priority; whereas in a process, monitoring the knowledge of the identity of the compounds involved (spectra) can be the main goal. The explanation given below holds for a data set where ITTFA is applied to obtain the profiles in the **C** matrix. Transposing the original data matrix, this process would be useful to get the profiles in $\mathbf{S}^T$. ITTFA gets each concentration profile following the steps below:

1. Calculation of the score matrix by PCA.
2. Use of an estimated concentration profile as initial target.
3. Projection of the target onto the score space.
4. Constraint of the target projected.
5. Projection of the constrained target.
6. Return to step 4 until convergence is achieved.

ITTFA starts performing PCA in the original data matrix, **D**. There is a formal analogy between the PCA decomposition, i.e. $\mathbf{D} = \mathbf{TP}^T$, and the CR decomposition, i.e. $\mathbf{D} = \mathbf{CS}^T$, of a data matrix. The scores matrix, **T**, and the loadings matrix, $\mathbf{P}^T$, span the same data space as the **C** and the $\mathbf{S}^T$ matrix do and their profiles can be described as abstract concentration profiles and abstract spectra respectively. This means that any real concentration profile of **C** belongs to the score space and can be described as a linear combination of the abstract concentration profiles in the **T** matrix.

The next step would be choosing approximate concentration profiles as targets. There are many ways to select these initial vectors; actually, any method used to provide initial CR estimates can be useful for this purpose. Historically, the vectors obtained after performing VARIMAX rotation onto the scores were used[24] and also the needle targets (i.e. vectors with only one nonnull element equal to 1), which are the simplest representation of a peak-shaped profile.[32,57]

The next step is the projection of the initial target ($\mathbf{x}_{1\,in}$) onto the score space. The projected target ($\mathbf{x}_{1\,out}$) belongs to the space of the real concentration profiles and, from

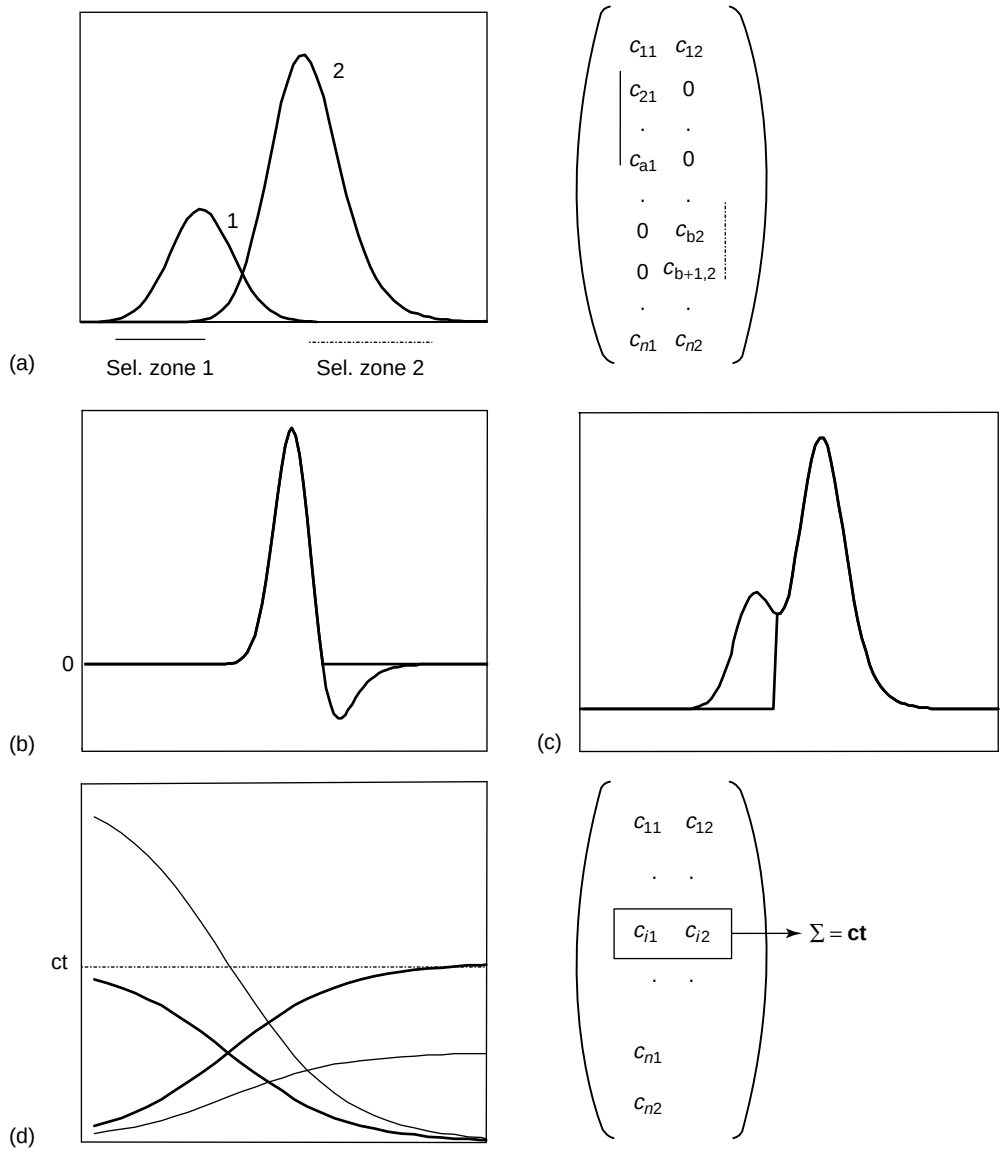

Figure 13 Performance of the constraints in the MCR/ALS method. (a) Selectivity. The solid line and the dashed line indicate the selective zones for compounds 1 and 2, respectively. On the right, the constrained matrix is shown. (b) Non-negativity, (c) unimodality and (d) closure: the normal line and the thick line show a response profile before and after the application of the constraint, respectively. (d) also presents the constrained concentration matrix.

a pure mathematical point of view, could be accepted to describe the data set. However, when this profile is plotted, the chemist may not like some of the features that present (e.g. negative parts, secondary maxima, etc.). When this happens, the projected target $\mathbf{x}_{1\ out}$ is modified by using the appropriate constraints. The application of constraints satisfies the chemical features demanded by the profile but, as a consequence of the modification of its elements, lifts the target from the score plane. The constrained target, $\mathbf{x}_2$, is projected again onto the score space and the new projected target, $\mathbf{x}_{2\ out}$, is constrained if necessary. The process goes on until the projected target makes sense from both mathematical and chemical points of view, i.e. until the constrained profile belongs to the score space or until it is very close to it. Figure 14 shows the ITTFA optimization of a chromatographic profile for a system with two compounds.

Once all the concentration profiles obtained are appended to form the **C** matrix, the $\mathbf{S}^T$ matrix can be calculated by least squares from **D** and **C** (see Equation 13).

MCR/ALS uses an alternative approach to find iteratively the matrices of concentration profiles and instrumental responses. In this method, neither the **C** nor the

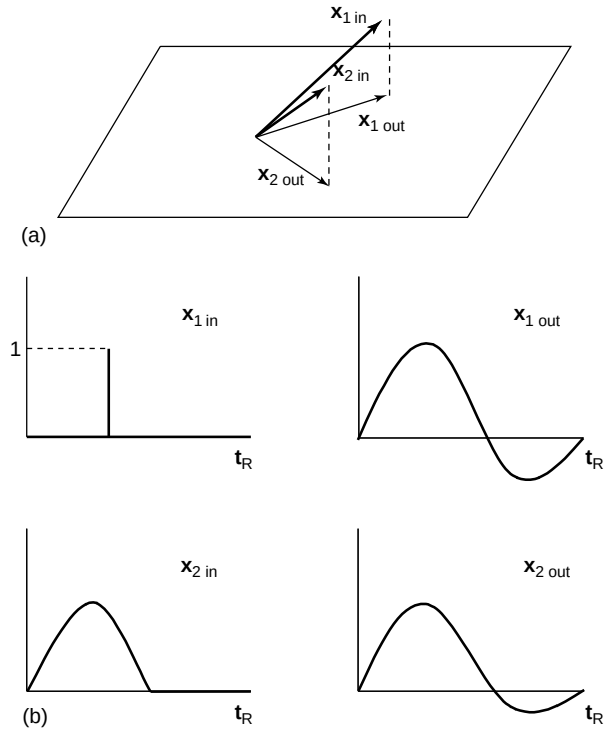

Figure 14 (a) Geometrical representation of the optimization of a chromatographic profile by ITTFA from an initial needle target. The example represents a two-compound data set. Thick lines represent targets out of the score plane; thinner lines are targets on the score plane. (b) Evolution in the shape of the chromatographic profile through the ITTFA process.

$\mathbf{S}^T$ matrix have priority over each other and both are optimized at each iterative cycle. The general operating procedure of MCR/ALS includes:

1. Determination of the number of compounds in **D**.
2. Calculation of initial estimates (e.g. **C**-type matrix).
3. Calculation and constraint of the $\mathbf{S}^T$ matrix.
4. Calculation and constraint of the **C** matrix.
5. Return to step 3 until convergence is achieved.

The number of compounds in **D** can be determined using PCA or can be known beforehand. In any case, the number obtained must not be considered a fixed parameter and resolution of the system considering different numbers of components is a usual and recommendable practice. In contrast to ITTFA, MCR/ALS uses complete **C**- or $\mathbf{S}^T$-type matrices during the alternating least squares (ALS) optimization. The two least-squares problems shown in Equations (21) and (22) are solved under restrictions such as non-negativity, unimodality or closure.

$$\mathbf{min}_{\widehat{\mathbf{C}}} \|\widehat{\mathbf{D}}_{\mathbf{PCA}} - \widehat{\mathbf{C}}\widehat{\mathbf{S}}^T\| \tag{21}$$

and

$$\mathbf{min}_{\widehat{\mathbf{S}}^T} \|\widehat{\mathbf{D}}_{\mathbf{PCA}} - \widehat{\mathbf{C}}\widehat{\mathbf{S}}^T\| \tag{22}$$

In these two equations the norm of the residuals between the PCA reproduced data, $\widehat{\mathbf{D}}_{\mathbf{PCA}}$, using the selected number of components and the ALS reproduced data using the least-squares estimates of **C** and $\mathbf{S}^T$ matrices, $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{S}}^T$, is alternatively minimized keeping constant $\widehat{\mathbf{C}}$ (Equation 21) or $\widehat{\mathbf{S}}^T$ (Equation 22). This is equivalent to the minimization of the least squares function $(d_{ij} - \hat{c}_{ik}\hat{s}_{kj})^2$, where d_{ij} are the experimental data and $\hat{c}_{ik}$ and $\hat{s}_{kj}$ are the current ALS estimations of the concentrations and spectra values (spectra i, wavelength j and species k). The unconstrained least-squares solution of Equation (21) is given by Equation (23)

$$\widehat{\mathbf{S}}^T = \widehat{\mathbf{C}}^+\mathbf{D}_{\mathbf{PCA}} \tag{23}$$

where $\widehat{\mathbf{C}}^+$ is the pseudoinverse of the concentration matrix which for a full rank matrix gives Equation (24)

$$\widehat{\mathbf{S}}^T = (\widehat{\mathbf{C}}^T\widehat{\mathbf{C}})^{-1}\widehat{\mathbf{C}}^T\mathbf{D}_{\mathbf{PCA}} \tag{24}$$

and the unconstrained least-squares solution of Equation (22) is given by Equation (25):

$$\widehat{\mathbf{C}} = \mathbf{D}_{\mathbf{PCA}}(\widehat{\mathbf{S}}^T)^+ \tag{25}$$

where $(\widehat{\mathbf{S}}^T)^+$ is the pseudoinverse of the spectra matrix which for a full rank matrix gives Equation (26):

$$\widehat{\mathbf{C}} = \mathbf{D}_{\mathbf{PCA}}\widehat{\mathbf{S}}(\widehat{\mathbf{S}}^T\widehat{\mathbf{S}})^{-1} \tag{26}$$

Equations (23) and (24) are solved sequentially, i.e. from a given estimation of the concentration matrix **C** obtained in the previous cycle of the ALS optimization a new estimation of the spectra matrix $\mathbf{S}^T$ (Equation 24) is calculated, and from this a new estimation of the concentration matrix **C** is then calculated (Equation 26). These solutions may be constrained to fulfill particular requirements like non-negativity, unimodality and closure immediately after they are calculated using Equations (23) and (25) (i.e. before the next equation is applied). In this way, at each iteration of the ALS procedure, the solutions are improved not only from a least-squares sense but also to fulfill a particular set of constraints. Direct non-negative least squares solutions of Equations (21) and (22) can also be found using either the Lawson and the Hanson algorithm[58] or recent faster modifications.[54] Likewise, special algorithms have been proposed for the implementation of unimodality constraints.[55] Initial estimates to start the ALS optimization can be obtained for **C** or for $\mathbf{S}^T$ matrices, using either pure variable detection methods, EFA-derived methods[59] or previously known profiles. In the current implementations of the MCR/ALS method,[60] different constraints may be selected for the **C** and the $\mathbf{S}^T$ matrix and, within each of those matrices, all or some of the profiles can be constrained.

The convergence criterion in the ALS optimization is based on the comparison of the fit obtained in two consecutive iterations. When the relative difference in fit is below a threshold value, the optimization is finished. Sometimes a maximum number of iterative cycles is used as the stop criterion. This method is very flexible and may be adapted to very diverse real examples, as shown in section 6.

5 THREE-WAY RESOLUTION METHODS

The methods presented in previous sections are suitable to work with a data matrix and give results related to the two different directions of the data matrix, i.e. profiles related to the variation along the rows and along the columns of the data matrix. This is the reason why a data matrix is also called a two-way data set.

A data matrix is not the most complex data set that can be found in chemistry. Let us consider a kinetic process monitored fluorimetrically; at each reaction time, a series of emission spectra recorded at different excitation wavelengths are obtained. This means that we collect a data matrix at each stage of the reaction and if the goal is getting a picture of the global kinetic process, the matrices should be considered altogether. The information about the whole kinetic process should now be organized in a cube of data (tensor) with three informative directions, i.e. in a three-way data set. Another usual example is coupling data matrices from different samples that share all or some of their compounds, e.g. several HPLC/DAD runs. In this case, the third direction of the data set accounts for the quantitative differences among samples. Figure 15 shows both examples mentioned.

Though there is a clear gain in quality and quantity of information when going from two- to three-way data sets, the mathematical complexity associated with the treatment of three-way data sets can seem, at first sight, a drawback. To overcome this problem, most of the three-way data analysis methods transform the original cube of data into a stack of matrices, where simpler mathematical methods can be applied. This process is often known as *unfolding*. A cube of data sized $(m \times n \times p)$ can be unfolded in three different directions: along the row space, along the column space and along the third direction of the cube, also called the tube space. The three unfolding procedures give a row-wise augmented matrix $\mathbf{D_r}(m \times np)$, a column-wise augmented matrix $\mathbf{D_c}(n \times mp)$ and a tube-wise augmented matrix $\mathbf{D_t}(p \times mn)$, respectively (see Figure 16). When the rank analysis of the three augmented matrices is carried out, the number of components obtained for the three different directions (*modes*) of the data set may be the same or not. When $\mathbf{D_r}$, $\mathbf{D_c}$ and $\mathbf{D_t}$ have the same rank, the three-way data set is said to be *trilinear* and when their ranks are different from each other, the data set is *nontrilinear*. (Please note that this definition holds for by far most of the chemical data sets, except those for which phenomena of rank deficiency or rank overlap are present).[61,62] The resolution of a three-way data set into the matrices **X**, **Y** and **Z**, which contain the pure profiles related to each of the directions of the three-way data set, changes for trilinear and nontrilinear systems, as can be seen in Figure 17. For trilinear systems, **X**, **Y** and **Z** have the same number of profiles (*nc*) and the three-way core, **C**, is an identity cube $(nc \times nc \times nc)$ whose unity elements are placed in the superdiagonal. In this case, the three-way core is often omitted because it does not modify numerically the reproduction of the original tensor. Each element in the original three-way data set can be reproduced as follows:

$$d_{ijk} = \sum_{f=1}^{nc} x_{if}\, y_{if}\, z_{kf} \tag{27}$$

Equation (27) is the fundamental expression of the PARAFAC model,[37–39] which is used to describe the decomposition of trilinear data sets.

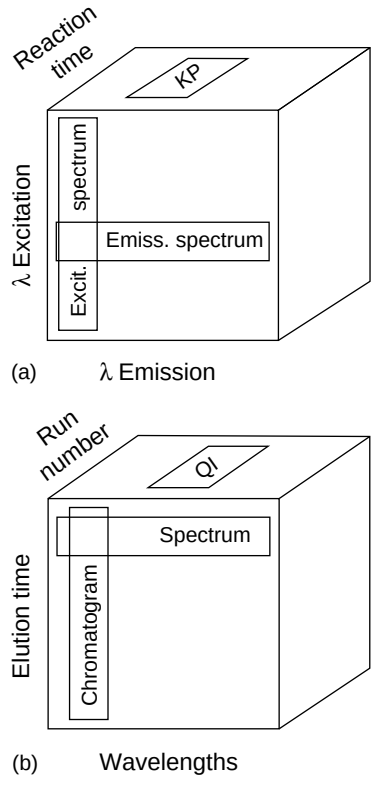

Figure 15 Examples of three-way data sets. (a) Fluorimetric monitoring of a kinetic process. KP, kinetic profile. (b) Coupling of several HPLC/DAD runs. QI, quantitative information.

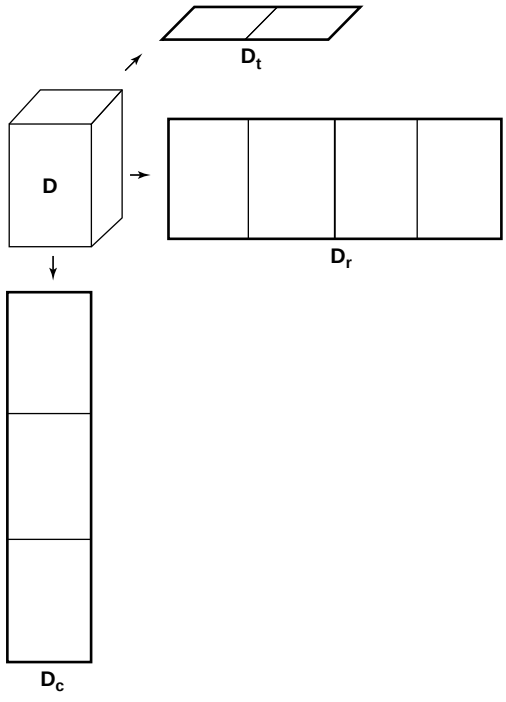

Figure 16 Unfolding of a three-way data set into a row-wise data matrix, $\mathbf{D_r}$, a column-wise data matrix, $\mathbf{D_c}$, and a tube-wise data matrix, $\mathbf{D_t}$.

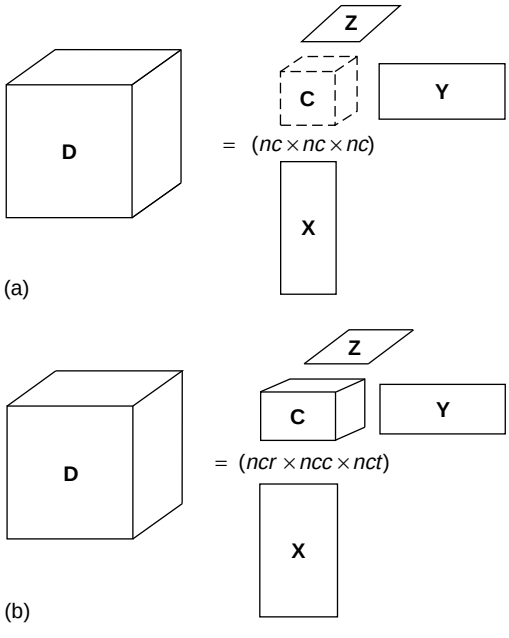

Figure 17 Decomposition of three-way data sets. (a) With trilinear structure. The dashed core can be omitted and is a regular identity cube. (b) With nontrilinear structure. The core is necessary and has a different number of components in each direction.

For nontrilinear systems, the core **C** is no longer a regular cube ($ncr \times ncc \times nct$) and the nonnull elements are spread out in different ways depending on each particular data set. *ncr*, *ncc* and *nct* hold for the rank in the row-wise, column-wise and tube-wise augmented data matrices respectively. Each element in the original data set can now be obtained as shown in Equation (28):

$$d_{ijk} = \sum_{f=1}^{ncr} \sum_{g=1}^{ncc} \sum_{h=1}^{nct} x_{if}\, y_{jg}\, z_{kh}\, c_{fgh} \qquad (28)$$

Equation (28) defines the decomposition of nontrilinear data sets and is the underlying expression of the Tucker3 model.[40]

Decompositions of three-way arrays into these two different models require different data analysis methods; therefore, finding out if the internal structure of a three-way data set is trilinear or nontrilinear is essential to ensure the selection of the suitable chemometric treatment.

In this section, the concept of trilinearity has been tackled as an exclusively mathematical matter. However, the chemical information is often enough to determine if a three-way data set presents this feature. How to link the chemical knowledge with the mathematical structure of a three-way data set can be easily seen with real examples. Let us consider the three-way data sets in Figure 15. If a data set is trilinear, **X**, **Y** and **Z** will have as many profiles as chemical compounds in the original data set and this number will be equal to the rank of the data set. For each chemical compound, there will be only one profile in **X**, in **Y** and in **Z** common to all the appended matrices in the original data set. In the HPLC/DAD example, the decomposition of the three-way array gives an **X** matrix with chromatographic profiles, a **Y** matrix with pure spectra and a **Z** matrix with the quantitative information about each compound in the different chromatographic runs. In this case, a trilinear structure would imply that the pure spectrum and the pure chromatogram of a compound remain invariant in the different chromatographic runs. If the experimental conditions in the runs analyzed are similar enough, the UV spectrum of a pure compound should not change; however, run-to-run differences in peak shape and position are commonly found in practice. Assuming that the elution process of the same compound in the different runs yields always the same chromatographic profile does not make sense from a chemical point of view and, therefore, the data set should be considered nontrilinear. In the example related to the fluorimetric monitoring of a kinetic process, the decomposition of the original data set gives a matrix **X** with pure excitation spectra, a matrix **Y** with pure emission spectra and a matrix **Z** with the kinetic profiles of the process. A trilinear structure would indicate that the shape of the excitation spectrum and the emission spectrum of a

compound does not change at the different reaction times of the kinetic process. This invariability of the spectra is an acceptable statement if the experimental conditions during the process are not modified. Therefore, this data set may be considered trilinear.

In practice, most of the systems are nontrilinear due either to the underlying chemical process (e.g. UV reaction monitoring coupling experiments with different reagent ratios) or to the instrumental lack of reproducibility in the response profiles (e.g. chromatographic profiles in different HPLC/DAD runs). Therefore, section 6 is mainly focused on the study of real examples lacking the trilinear structure.

Despite the higher abundance of nontrilinear data sets, many of the algorithms proposed to study three-way arrays rely on the assumption of trilinear structure. This is the case of GRAM,[35] designed to work with two matrices, or its natural extension, DTD,[36,63] which can handle larger data sets with more appended matrices. Both GRAM and DTD are noniterative methods and use latent variables to resolve the profiles in **X**, **Y** and **Z**. When these methods are applied to nontrilinear data sets, the profiles obtained often belong to the imaginary domain. Iterative methods are also used in three-way data and the scheme followed in their application is the same as for a single data matrix, i.e. determination of number of components, use of initial estimates, application of constraints and iterative optimization until convergence. Most of the iterative algorithms are based on least-squares calculations. As for two-way data sets, three-way iterative methods are more flexible and can deal with more diverse data sets.

It has been commented that three-way resolution methods generally work with the unfolded matrices. Depending on the algorithm used, the three unfolded matrices are used or only some of them. As examples, the PARAFAC decomposition of a trilinear data set using the three unfolded data matrices and the resolution of a nontrilinear data set by applying the MCR/ALS method to only one of the unfolded matrices will be described.

The PARAFAC decomposition of a trilinear data set can be obtained by using a constrained least-squares algorithm.[37–39] The steps of the method are listed below:

1. Determination of the number of chemical compounds (nc) in the original three-way array.
2. Calculation of initial estimates for **X** and **Y**.
3. Estimation of **Z**, given $\mathbf{D_t}$, **X** and **Y**.
4. Estimation of **X**, given $\mathbf{D_r}$, **Y** and **Z**.
5. Estimation of **Y**, given $\mathbf{D_c}$, **X** and **Z**.
6. Return to step 3 until convergence is achieved.

The number of chemical compounds (rank) in the three-way array is estimated for the unfolded matrices in the three directions of the data set ($\mathbf{D_t}$, $\mathbf{D_r}$ and $\mathbf{D_c}$). If the data set is really trilinear, the rank values should coincide unless situations of rank deficiency or rank overlap occur.

Since the original data set is decomposed into three matrices, **X**, **Y** and **Z**, the initial estimates should be a pair of matrices sized as two of the matrices resulting from the decomposition. Any combination, e.g. **X**- and **Y**-type matrices, and **X**- and **Z**-type matrices, is accepted. The initial estimates can be calculated using any of the methods explained for resolution of two-way data sets.

From the original data set and the two matrices proposed as initial estimates, e.g. **X** ($m \times nc$) and **Y** ($n \times nc$), the third matrix of the PARAFAC model, **Z** ($p \times nc$), may be calculated. To do so, the unfolded matrix in the direction of the matrix to be calculated, $\mathbf{D_t}(p \times mn)$, is used (mathematical details related to this step are in the literature[39]). In each iterative cycle, this operation is performed analogously to obtain **X** and **Y** and every time the suitable unfolded matrix, $\mathbf{D_r}(m \times np)$ or $\mathbf{D_c}(n \times mp)$, participates in the calculation. The constraints explained in section 4 may be optionally applied to the profiles in **X**, **Y** and **Z**. The original three-way array is reproduced according to Equation (27) and the end of the iterative process arrives when changes in the fit among two consecutive iterations are small enough.

Three-way data sets have been presented as cubes of data formed by appending several matrices together. This means implicitly that all the data matrices in a tensor should be equally sized; otherwise, the cube cannot be constructed. Besides, the information in the rows, in the columns and in the third direction of the array must be synchronized for each of the layers of the cube, e.g. in the HPLC/DAD example, if the columns are wavelengths, all the runs should span the same wavelength range and if the rows are retention times, the elution time range should also coincide. In experimental measurements, it may not be easy or convenient to fulfill these two requirements. Actually, synchronization can be difficult when the parameter that changes in one of the directions of the array cannot be controlled in a simple manner and getting matrices equally sized may also be inconvenient if this condition forces the inclusion of irrelevant information in some of the two-way appended arrays. An example of difficult synchronization is the coupling of experiments related to UV monitoring of pH-dependent processes since pH variations may not be easily reproducible among experiments. The inconvenience of appending equally sized two-way arrays is evident when HPLC/DAD runs related to a mixture and to matrices of single standards are treated together; if the runs of standards should cover the same elution time range as the run of the mixture, most of the information in these standard matrices will be formed by baseline spectra which are not relevant for the resolution of the mixture.

When building a typical three-way data set is not possible, there is no need to give up the simultaneous analysis of a group of matrices that have something in common. Some methods, such as MCR/ALS, are designed to work with only one of the three possible unfolded matrices. This operating procedure greatly relaxes the demands to the two-way arrays that are treated together. Indeed, MCR/ALS requires only one common direction in all the matrices analyzed, e.g. in the previous examples, the wavelength range of the spectra collected (see Figure 18).

The MCR/ALS decomposition of the three-way data set shows the ability of this method to deal with nontrilinear systems. Whereas the profile of each compound related to the common order of the column-wise augmented matrix (spectra) is considered to be invariant for all the matrices, the unfolded **C** matrix allows the profile of each compound in the concentration direction to be different for each appended data matrix. This freedom in the shape of the **C** profiles agrees with the nontrilinear structure of the data array. The least-squares problems solved by MCR/ALS when applied to a three-way data set are the same as those in Equations (23) and (25); the only difference is that **D** and **C** are now augmented matrices. Though the operating procedure of the method has been shown in section 5, some particularities connected with the treatment of three-way data sets deserve a comment.

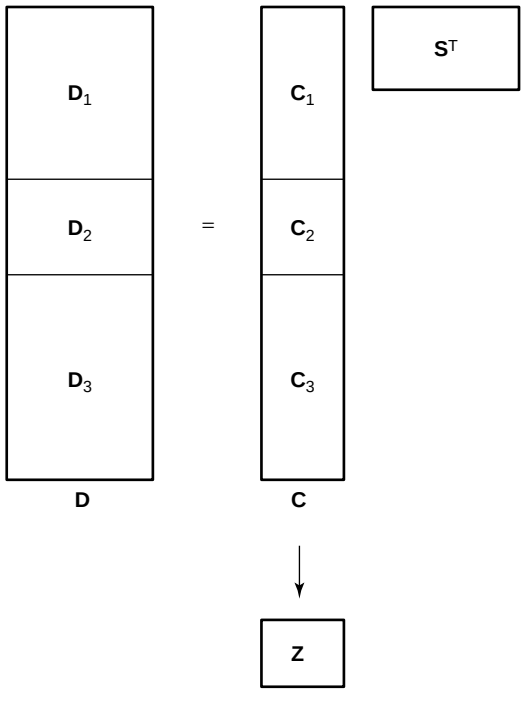

Figure 18 MCR/ALS decomposition of a three-way data set formed by three appended matrices.

In the resolution of a column-wise augmented data matrix, the initial estimates can be either a single $\mathbf{S}^T$ matrix or a column-wise augmented **C** matrix. The column-wise concentration matrix is built by placing the initial **C**-type estimates obtained for each data matrix in the three-way data set one on top of the other. The appended initial estimates must be sorted as the initial data matrices are in **D** and must keep a correct correspondence of species, i.e. each column in the augmented **C** matrix must be formed by appended concentration profiles related to the same chemical compound. When no prior information about the identity of the compounds in the different data matrices is available, the correct correspondence of species can be known from the resolution results of each single matrix.

The same constraints used in the resolution of a data matrix can be applied to three-way data sets. Selectivity and non-negativity affect the spectrum and the augmented concentration profile of each species, whereas unimodality is applied separately to each of the profiles appended to form the augmented concentration profile. The closure constraint operates by applying the corresponding closure constant to each of the single matrices in the column-wise concentration matrix. Another constraint specific of three-way data sets is the so-called *correspondence among species*. In each single matrix of a three-way data set, the concentration profiles of absent compounds are set equal to zero after each iterative cycle.

Though MCR/ALS is specially relevant to cope with nontrilinear data sets formed by matrices of varying sizes, it can also work with trilinear data sets. Because of the inherent freedom in the modeling of the profiles of the augmented **C** matrix, the trilinear feature is introduced in the MCR/ALS method as an optional constraint.[41] The application of this constraint is performed separately on the concentration profile of each species. Thus, the profiles appended to form the augmented concentration profile of a certain species are placed one beside the other to form a new data matrix and PCA is performed on it. If the system is trilinear, the score vector related to the first principal component must show the real shape of the concentration profile and the rest of principal components must be related to noise contributions. The loadings related to the first principal component are scaling factors accounting for the species concentration level in the different appended matrices. Therefore, the new single profiles will be calculated as the product of the score vector and their corresponding scaling factor. The constrained single profiles are finally appended to form the new augmented concentration profile. All of this process is shown graphically in Figure 19. In contrast to some other three-way resolution methods specially designed to work with trilinear systems, the

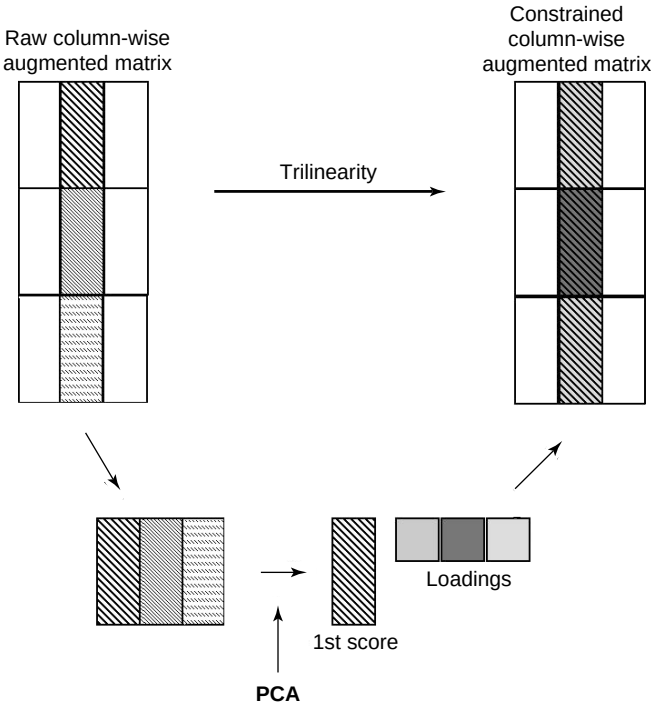

Figure 19 Application of the trilinearity constraint in the MCR/ALS method.

implementation of this constraint in MCR/ALS should not necessarily be complete, i.e. all or some of the compounds can be forced to have common profiles in the **C** matrix. This flexibility allows a more representative modeling of some real situations, like those of systems with trilinear profiles related to the evolution of chemical compounds and a free modeled profile related to an important background contribution.

The information in the third direction of the array, i.e. the **Z** matrix, is directly extracted from the augmented matrix **C** in MCR/ALS.[41] This dimension of the data set is usually the smallest in size and is connected with scaling differences among the matrices appended. Since the $\mathbf{S}^T$ profile of one compound is common to all the appended data matrices, the area of the concentration profiles of this compound is scaled according to the concentration level of the species in each single data matrix. Thus, the profile of a compound in the **Z** matrix accounts for the relative concentration of a particular compound in each of the appended matrices and can be obtained from the ratio between the area of its concentration profile in a given matrix and the area related to the concentration profile of the same compound in a matrix taken as reference.

Since MCR/ALS has been proven to be a very versatile method to deal with any kind of three-way data set, this method has been used to work with all the diverse examples shown in section 6.

6 EXAMPLES OF APPLICATION

Different examples have been selected showing the possibilities of soft modeling methods in the analysis of chemical data. The first example refers to a chromatographic co-elution using LC/DAD. This is a typical example of analytical chemistry where the motivation of the data analysis covers most of the aspects analyzed in previous sections: estimation of the number of components coeluting in a chromatographic peak using FA-derived methods; estimation of the elution windows of each of these co-eluting components using EFA-derived methods; resolution of these components by means of multivariate resolution methods; and finally the eventual quantitation of these components. The second example will be the analysis of IR spectral data from multiple runs of an industrial process. In this case the goal of the data analysis is the development of a soft model explaining how the industrial process evolves, resolving the different components of the system and estimating the relative concentrations of the different constituents at any stage of the process from its time spectrum. The third example refers to the study and interpretation of thermodynamic and conformational transitions of polynucleotides using UV absorption and circular dichroism (CD) spectrometric methods.

6.1 Example 1: Chromatographic Co-elution Using Liquid Chromatography/Diode-array Detection[41]

Soft modeling FA-based methods have been proposed for the analysis of unresolved peaks in LC/DAD and particular attention has been focused in the use of FA and MCR methods.[23] CR methods, like ITTFA,[24,32] EFA,[15] WFA,[30] SIMPLISMA[26] or HELP,[31] have been used as qualitative tools for resolution of co-eluted peaks and peak purity, and they are not intended to provide fully quantitative information about the overlapped components. However if several chromatographic runs are simultaneously analyzed using MCR, relative quantitative information may be obtained similarly to three-way data analysis and second-order calibration methods.[36] The key aspects for this quantitation to be possible are that the spectrum of the same chemical component in the different chromatographic runs does not change and, eventually, that the elution profile of a component in the different chromatographic runs also has the same shape and is synchronized. In case this second condition be not totally obeyed, resolution and quantitation of the co-eluted compounds is still possible, but the accuracy of the obtained solutions will depend on local rank and resolution conditions.

As an example of chromatographic co-elution, Figure 20 shows a three-dimensional plot of the data

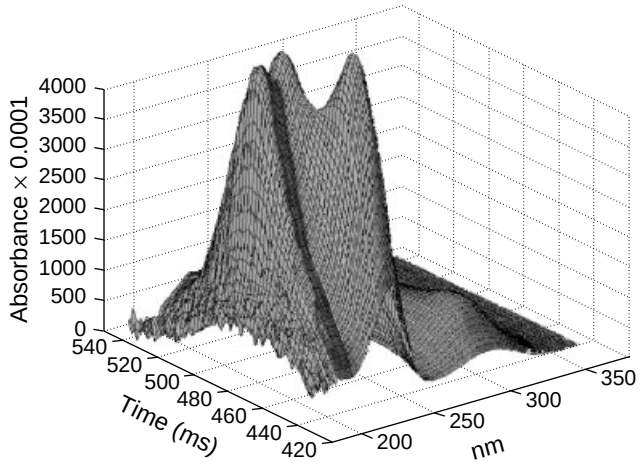

Figure 20 Three-dimensional plot of the LC/DAD peak of the unresolved mixture naphthol–pirimicarb.

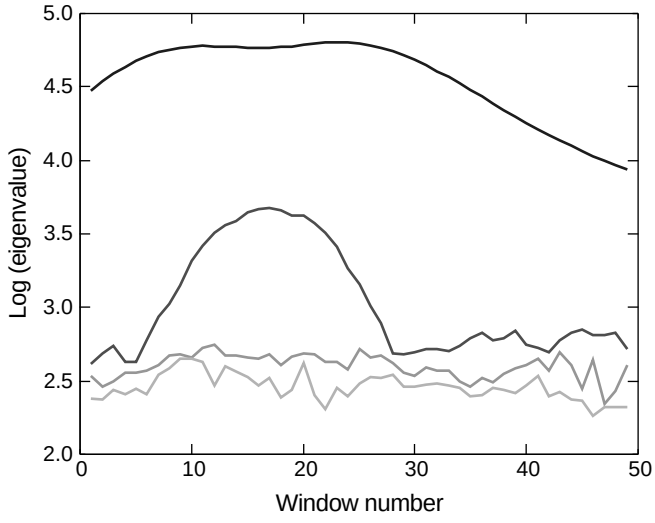

Figure 21 FSMW/EFA of the data given in Figure 20. During the experiment two distinctive components are clearly distinguished with an overlapped range between 450×10^{-3} and 510×10^{-3} min.

matrix obtained in the co-elution of two pesticides, pirimicarb and 1-naphthol, using LC/DAD. Apparatus, reagents, chromatographic conditions and other experimental details are given in Tauler et al.[64] Figure 21 shows the results of EFA using a fixed-size moving window of the data given in Figure 20. Along the experiment, two components were detected and distinguished from the noise level. The FSMW/EFA results suggested that the system might be resolved with few ambiguities since selectivity and local rank conditions were present. This is a common situation in chromatography, except for embedded peaks (one peak inside another[29]). Embedded unresolved peaks can still be resolved when several chromatographic runs are simultaneously analyzed and at least in one of the runs, the local rank and resolution conditions are also achieved for that embedded component.[65,66]

Application of MCR/ALS to the data of Figure 20, together with the data from the individual analyses of the two pesticides, gave the resolved pure spectra and elution profiles of Figures 22 and 23. Component 1 refers to pirimicarb, and component 2 refers to naphthol which elutes second. This resolution was obtained using non-negativity constraints for elution and spectral profiles and using unimodality constraint for the elution profiles. Each component is characterized by a pure spectrum equal in all the data matrices where this component is present. The pure spectra recovered (Figure 23) in the simultaneous numerical treatment were found to be equal to the pure spectra obtained in the individual chromatographic analysis of the pure analyte samples. In Figure 22, the two elution (concentration) profiles for each of the two components are given, one for the component in the pure analyte sample and the other for the same component in the unknown mixture.

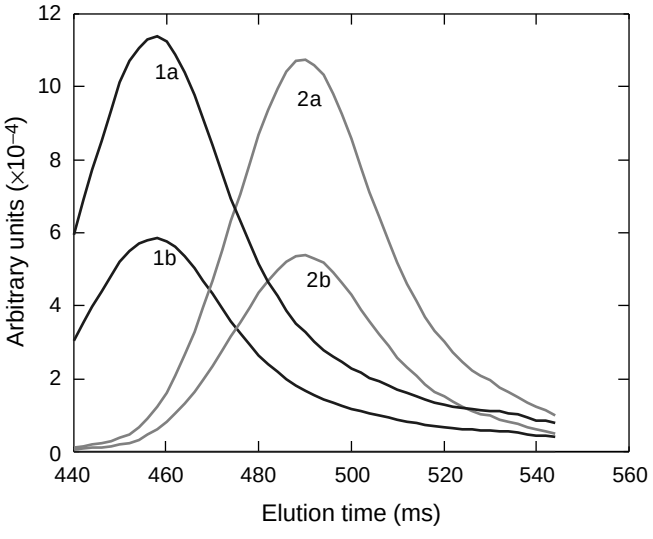

Figure 22 Resolved elution–concentration profiles of the components in the three samples simultaneously analyzed: (1a) pirimicarb in the pure analyte sample; (1b) pirimicarb in the mixture; (2a) naphthol in the pure analyte sample; (2b) naphthol in the mixture.

The resolution of the two co-eluted components in the unknown mixture was equal to 0.489.

Once the concentration profiles of the different components in the mixture are recovered using the proposed method, the concentration of the analytes in the samples can be estimated. If the conditions of linearity hold, the area under the concentration profile of a certain component is proportional to the concentration of this analyte. The ratio between these areas for a particular component

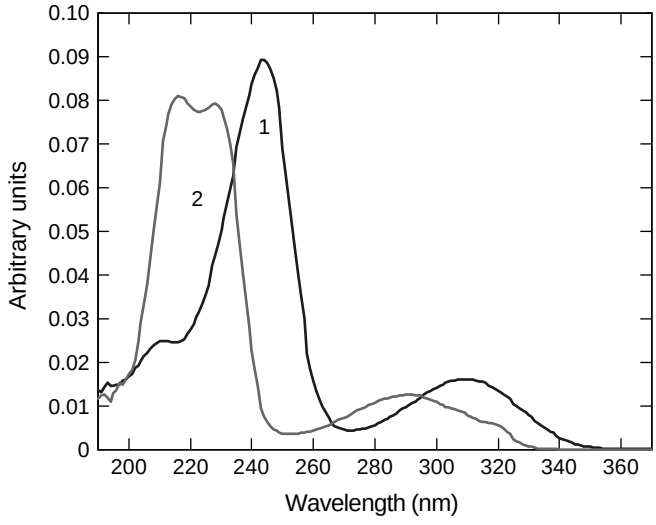

Figure 23 Resolved pure spectra of the two overlapped components. Spectrum 1 corresponds to pirimicarb and spectrum 2 corresponds to naphthol.

gives the ratio between the concentrations of that particular analyte in the different samples, recovering directly the relative quantitative information. If, in addition, in one or some of the experiments the concentration of the analyte is known, the concentration of the analytes in the unknown samples can be calculated also in absolute values. In this example, the method of quantitation is improved when the constraint of equal shape over the concentration profiles is applied (trilinear data).

In Table 1 a summary of the results obtained when different methods of quantitation are applied to the unresolved mixture of pirimicarb and naphthol (Figure 20) are given. For comparison, the results given first (method 1) correspond to those obtained by classical least-squares regression when the pure unit spectra of all the eluted components are given in the input. The analysis is performed only over the unresolved mixture data matrix. Whereas the results for pirimicarb are rather good, those for naphthol are poorer. The results of the simultaneous analysis of the data matrices obtained in the chromatographic analysis of an unknown mixture together with one or two pure analyte samples are also given. When the unknown mixture is analyzed together with these two samples, the results (method 2a) are better than before, although still with some error in the estimation of the pirimicarb concentration. When one of the two co-eluted components is considered an interferent and the unknown mixture is analyzed together with a pure analyte sample containing the other co-eluted component, the results (methods 2b and 2c) are worse, specially for pirimicarb with an error of 26% (method 2c). Conversely, when the constraint of equal shape in the elution profiles is added (method 3), the results of the quantitation are good in

Table 1 Results of the quantitation of the unresolved LC/DAD peak mixture of pirimicarb and naphthol[a]

Method[b]	Pirimicarb	Naphthol	N[c]
1	0.267 (4%)[d]	0.039 (17%)	1
2a	0.270 (6%)	0.045 (3%)	3
2b	–	0.043 (8%)	2
2c	0.189 (26%)	–	2
3a	0.263 (3%)	0.047 (1%)	3
3b	–	0.049 (5%)	2
3c	0.252 (1%)	–	2
4a	–	0.050 (2%)	2
4b	0.256 (0%)	–	2

[a] Quantitation of the unresolved pirimicarb–naphthol mixture shown in Figure 20. Correct concentrations are, respectively (in ppm), pirimicarb 0.256 and naphthol 0.047.
[b] Method used in the quantitation of the unresolved mixture: (1) least-squares estimation using known pure spectra of the two unresolved components; (2) simultaneous analysis of multiple chromatographic runs using MCR and nonnegative constraint; (2a) mixture sample and two pure analyte samples; (2b) mixture sample and naphthol pure analyte samples; (2c) mixture sample and pure pirimicarb sample; (3) As in (2) but with the additional constraint of equal shape elution profiles of common components between runs (trilinearity constraint, Figure 19); (4) GRAM;[35] (4a) mixture sample and pure naphthol sample; (4b) mixture sample and pure pirimicarb sample.
[c] Number of data matrices used in the analysis.
[d] Error percentage.

all the cases. When naphthol is considered an interferent and only the pure pirimicarb sample is included in the simultaneous analysis, the quantitation of pirimicarb is still very good (1% error). Although method 2 has the advantage that it did not require synchronization in the time order, it has the disadvantage that it requires more knowledge about the interferents to resolve the overlapped components in quantitative terms. Method 3 is the best choice for quantitation when interferents are present and it is closely related to second-order calibration and trilinear resolution methods.[35,36] However, it requires the data to be trilinear, which is rather unusual in real chromatographic co-elution conditions.

The results of Table 1 show that MCR can be adapted to handle LC/DAD data of different complexity. For those cases where the elution of the components is not completely reproducible because of experimental or instrumental limitations, the method can still be used for quantitation, although now the error caused by the presence of unknown interferents not present in the known analyte samples will be higher. Other recent examples of application of the MCR method to chromatographic co-elution problems have proved the utility of the proposed method and have been extended to more difficult resolution problems. Examples are the resolution of mixtures of traces of herbicides and pesticides,[66] in interlaboratory studies[67] in the resolution of liquid chromatography/mass spectrometry (LC/MS) co-eluted peaks with common mass spectrometric ions,[68] as well as in

the validation of the MCR/ALS method for peak purity and quantitation in chromatographic analysis of mixtures of unresolved hormones.[65]

6.2 Example 2: Infrared Spectral Data Analysis from Multiple Runs of an Industrial Process[69]

Recent advances in process instrumentation and in data collection techniques have resulted in a rapid increase in the amount of data that can be acquired from chemical processes. Extracting the significant information from the data produced by modern instrumentation is in many circumstances a nontrivial task. The description and modeling of the evolution of a chemical process is important for both practical and economic reasons. In many of these cases multivariate calibration and regression methods[70] cannot be applied because there is no previous information available to perform the calibration of the system. Examples of this situation in process analysis are abundant and include monitoring the evolution of chemical processes where one or more parameters are changed, like time, temperature, pH, the concentration of a reagent or any other parameter, and there is no previous quantitative information about the evolution of the process. The multivariate data acquired with spectroscopic probes produce continuous data which can be arranged in an ordered data matrix according to the variation of the parameter changed during the process. One way to address this problem is by CR methods.

In the case of process analytical chemistry[71] the final goal is the estimation of the concentration of the constituents in the mixture change along the process and simultaneously the estimation of the unit responses (pure spectra) of those constituents. To have a meaningful solution from the CR decompositions it is necessary to make some assumptions about the signals obtained apart from their bilinearity, such as non-negativity, unimodality or closure. Taking advantage of the ordered structure of the process analysis data, EFA provides very valuable information concerning the windows of existence and relative importance of every component in the unknown mixtures existing at any time during the evolution of the process.

In addition, if several runs of the same industrial process are available and as usually at least one of the two orders is common between them (e.g. the scanned spectral range), the intensity ambiguities associated with the analysis of a single process run can be resolved in relative terms. Assuming that the same component in the mixture has the same pure spectrum in the different process runs, the simultaneous analysis of different process runs can give the correct relative amounts of the common components in the different mixtures analyzed in the different process runs at any particular stage of the process.

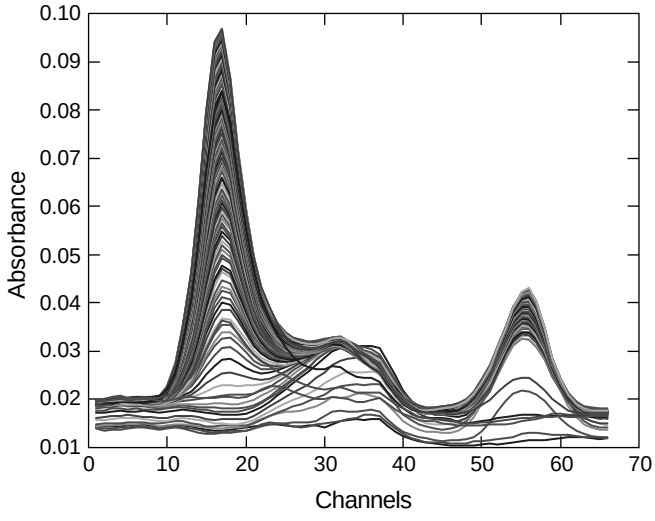

Figure 24 Example of the spectra acquired in one process run.

Data for the present example are spectra obtained from successive runs of an industrial chemical process. The spectra are evenly spaced in time, but may represent different elapsed times from beginning to end. The spectral intensities are stored in a matrix **D** with one spectrum per row. The dimensions of **D** are the number of spectra by the number of spectral channels (e.g. wavelength). The object is to determine the number of unique chemical components included during the full process run along with their pure spectra and concentration/time profiles. Eight different runs of the same industrial chemical process at different days of production were analyzed. Every run contained between 75 and 125 spectra, 795 in total, measured along a broad IR spectral range of 66 channels. An example of the data collected in one of the runs of the process is given in Figure 24. As seen from Figure 24, the spectra change with time, starting from a very weak and flat background absorption, increasing the absorption giving two main absorption bands at channel numbers 10–20 and 50–60 and a broader absorption band around channel number 30, and finally decreasing the absorption very fast on all those bands when the process is terminated. All the runs show a similar pattern but with slight differences between them in the timing and in the position of the maxima (shifting), which show that there should be also some differences in the chemistry of the different runs of the process.

As also seen in Figure 24, the background absorption and baseline can also change during the process and are different among different runs. Therefore some data pretreatment is needed. First, to account for the differences in the initial baseline absorption among different process runs, subtraction of the first spectrum

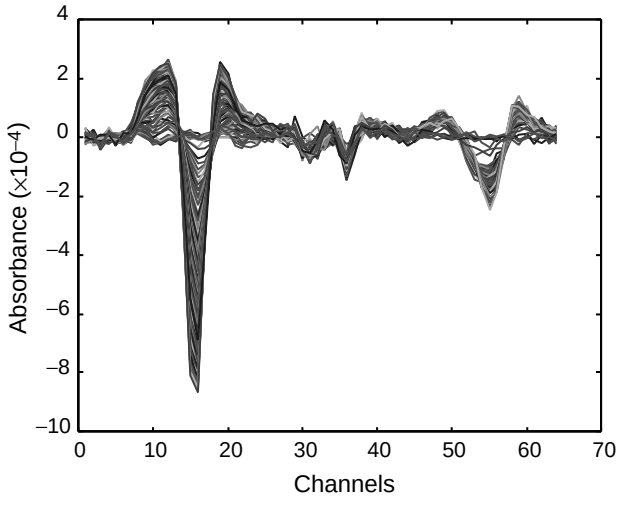

Figure 25 Second derivative spectra of one process run.

of each run from the following spectra in the same run removes these differences. This is true because in the first spectrum only the baseline or initial background absorption is present and the spectral bands of interest for the process itself have not appeared yet. With this treatment the first spectrum in each run always will be zero. Second, to account for the changes in the baseline or background during a particular run, the first and second derivatives of the raw spectra are calculated. That pretreatment allows the minimization of the contributions which are constant along a particular spectrum (first derivative) as well as those contributions which produce constant slopes within every spectrum (second derivative). In Figure 25, the second derivative spectra of the data set of Figure 24 are given. Most of the baseline changes are now removed. Pretreatment, second derivative and subtraction of the first spectrum were all analyzed and also compared with the analysis of the raw experimental spectra without any pretreatment.

Table 2 Comparison of the reduced singular values[9] obtained in the analysis of the complete second derivative data set to the reduced singular values obtained in the analysis of the spectral regions where there is no contribution of the components of interest

795 spectra 64 channels	795 spectra 5 blank channels
1.5381559×10^{-6}	1.4242273×10^{-7}
4.4253437×10^{-7}	1.1659144×10^{-7}
1.3784506×10^{-7}	1.3244814×10^{-7}
9.2429634×10^{-8}	1.6341295×10^{-7}
4.8231603×10^{-8}	2.0548819×10^{-7}

The estimation of the singular values related with noise is performed using the first channels of the second derivative spectra where no band is present (Figure 25). As mentioned before, in second derivative spectra, the background and baseline contributions to the data variance are considerably diminished. The singular values obtained in this narrow spectral range are estimated for all the process runs together, to include the variations between process runs. At the same time, the singular values of the complete data set comprising the 64 channels of measurement, the process runs individually, and all together, were also calculated. In the comparison (Table 2) it was taken into account that the dimensions of the data matrix in one case and another were not the same and therefore the reduced singular values[9] were used. When the complete set of runs and spectra were analyzed, the maximum number of contributions associated with different chemical contributions was estimated as two or three, since the value of the third singular value is similar to the first singular value associated with the noise in the nonabsorbing parts of the spectra. When the analysis was performed over the individual process runs, it was found also that the number of chemical contributions was between two and three. The number of components obtained in this way was only a first estimation to start the resolution procedure. The number of three components was finally confirmed from the results of EFA and of the ALS optimization procedures (see below).

Figure 26 shows the spectra shown in Figure 25 recalculated using PCA. The principal components used in that reproduction are the three more significant found by the PCA of the whole augmented data matrix. As seen by comparison of Figures 25 and 26, most of the dominant spectral features are described by the three principal components (albeit obtained in the analysis of the 795 spectra

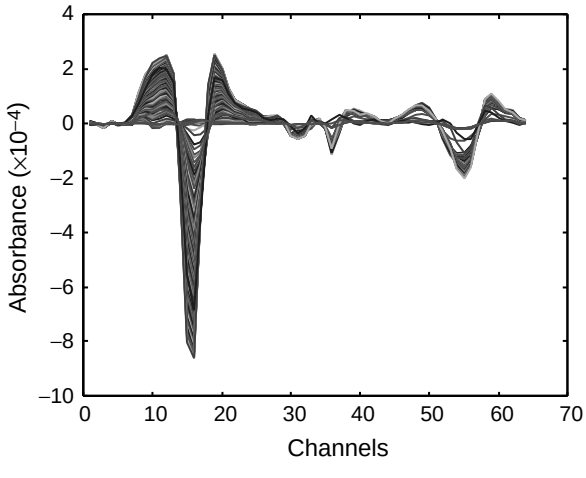

Figure 26 Reproduced second derivative spectra of the process run given in Figure 24 using three principal components.

Table 3 Standard deviation of the residuals obtained in the data analysis[a]

	Two components	Three components	Four components
exp–PCA[b]	2.82×10^{-3}	1.18×10^{-3}	1.16×10^{-3}
PCA–calc[c]	5.50×10^{-3}	3.30×10^{-3}	6.76×10^{-3}
exp–calc[d]	6.19×10^{-3}	3.80×10^{-3}	6.86×10^{-3}

[a] Results are given for two, three and four components in the simultaneous analysis of the eight runs of the process.
[b] Standard deviation of the residuals between the experimental data matrix and the PCA reproduced data matrix.
[c] Standard deviation of the residuals between the PCA reproduced data matrix and the calculated data matrix using the MCR/ALS optimized set of pure spectra and concentration profiles given in Figures 28 and 29.
[d] Standard deviation of the residuals between the experimental data matrix and the calculated data matrix using the MCR/ALS optimized set of pure spectra and concentration profiles given in Figures 28 and 29.

of the eight different runs of the process, each one with 66 channels). Noise filtering is also achieved in this way and very little information is lost by considering only these three components. If only two components are considered a poorer reproduction of the original data is observed and if four components are considered very little improvement is achieved. The standard deviations of the residuals between the eight-run experimental matrix and the reproduced matrix considering two, three and four principal components are given in Table 3.

In Figure 27 an example of the plot of the concentration profiles obtained by EFA when applied to one run of the process is given. From EFA, three components were detected and differentiated from the other contributions. The fourth and fifth components emerged significantly from the error contributions only at the very end of the process when the reaction was terminated and therefore not of interest for the present study. Similarly, the EFA of each of the eight other process runs provides an initial estimation of the concentration profiles of the components in each process run. These concentration profiles were used as initial values in the ALS optimization. They can be used in the individual analysis of each set of data, or better, to build up the initial estimation of the augmented concentration matrix to be used in the simultaneous analysis of the eight runs of the process.

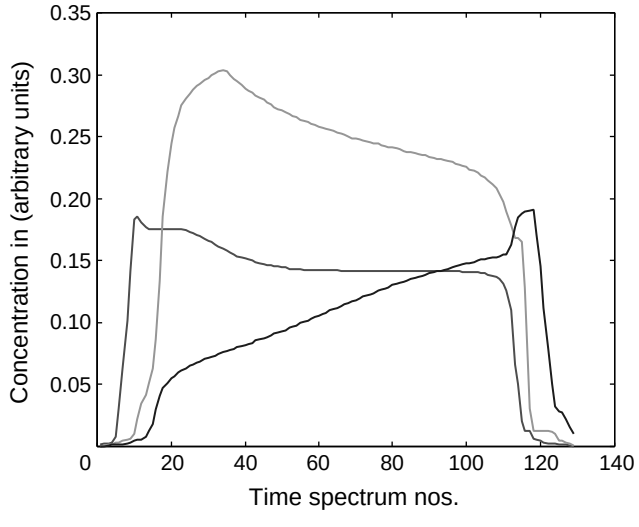

Figure 27 Initial concentration changes estimated using EFA of a single process run.

Although each process run was extensively analyzed individually, the results presented here refer only to those obtained in the simultaneous analysis. These solutions were improved by the added constraint of equality of spectra between process runs. The alternating and constrained least-squares optimization method was applied to the augmented data matrix containing the eight runs of the process arranged in the following three forms: without any pretreatment; with the first spectrum of each run subtracted; and with the second derivative augmented data matrix. The concentration matrix used initially in the optimization was the augmented concentration matrix containing the concentrations obtained in the individual EFA of each run. Of the three arrangements of the data matrix, the one which gave better results is the second case where the spectra of each run were corrected by subtracting the first spectrum of the same run. The reason for the better results in this case is that the subtraction of the first spectrum of each run removes the arbitrary offset between data from different process runs and that the non-negativity constraint can also be applied over the unit pure spectra. Conversely, when the optimization is performed over the second derivative spectra the non-negativity constraint is lost and it cannot be applied over the unit spectra. While the results are still in agreement with those obtained with the subtracted data matrix, the shapes of the recovered concentration profiles and pure spectra are less reliable. Of the three cases, the poorest results were obtained when no pretreatment was performed. The reason for this degradation of the resolution is because in that case the effects of the baseline (background absorption) between process runs and within process runs is rather high and the optimization is more difficult. In order to summarize the large amount of calculations performed during the work, only the results obtained in the simultaneous analysis of the eight runs of the process will be given. In the ALS optimization of the complete data set, the number of three components was again reconfirmed. If another number of components is considered, not only is the fit worse (see Table 3) but also the shapes of the recovered unit spectra and concentration profiles do not make chemical sense.

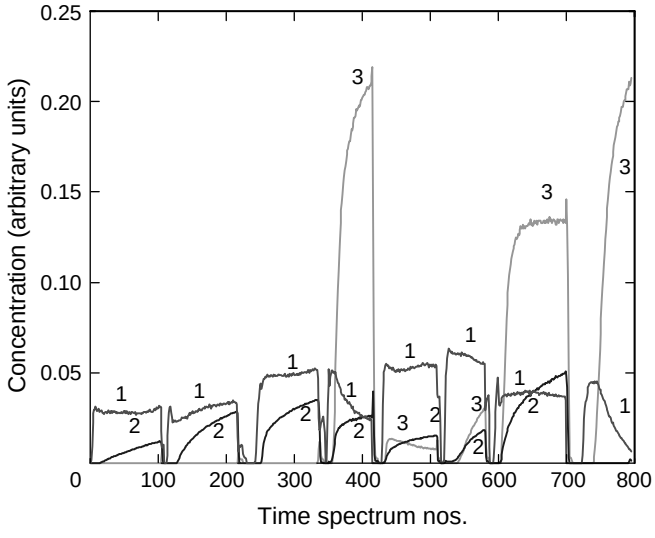

Figure 28 Concentration changes calculated by constrained ALS of eight process runs. The *x*-axis is a time axis and gives an indication of the spectrum number within the total number of spectra analyzed (795 in total).

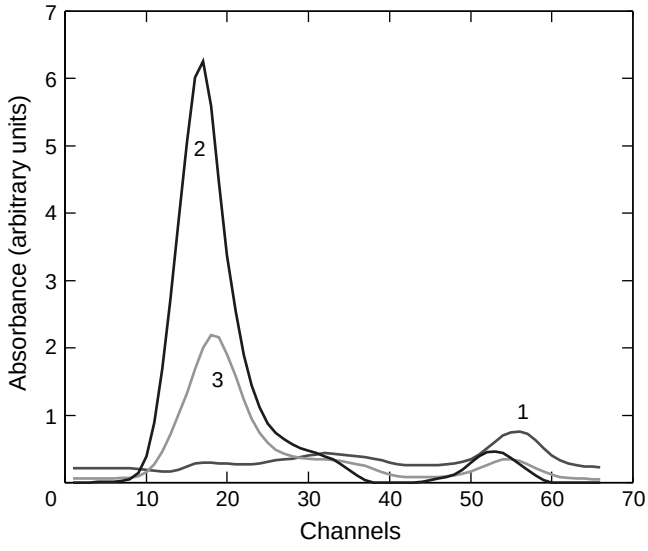

Figure 29 Pure spectra calculated by constrained ALS of eight process runs.

Figure 28 shows the concentration profiles obtained in the analysis of the eight runs of the process after applying the constrained ALS optimization. Component one rises very fast at the beginning of the process, decreases a little and then keeps stable until the process is terminated. Component two is present except in process run 8, and always shows the same pattern of growth. The reason why this component does not appear in process run 8 is because the data analyzed pertain only to the first part of the process before it became present in appreciable concentrations. Component three does not appear at appreciable concentrations in all the runs. It is nearly nonexistent in process runs 1, 2 and 3, but it is the dominant contribution in process runs 4, 7 and 8.

The two constraints applied on the shape of the concentration profiles, non-negativity and unimodality, have an important role during the optimization. It was a little more difficult to handle the unimodality constraint because it has to account also for the small random oscillations and changes in the concentration of the components during the process; this means that although the global shape has to be unimodal and smooth, locally some small oscillations of the unimodality condition have to be allowed.

Figure 29 gives the pure spectra of the three common components deduced from the analysis of the eight runs of the process with the ALS optimization procedure applied to the augmented data matrix describing the eight runs of the process. The shapes obtained for these three unit spectra explain very well the changes observed in the shapes of the raw process experimental spectra of every run. For instance the first component only has the band around channels 50–60; this is in agreement with the first experimental spectra of every run which only show that band. Conversely, the second and third components have two bands approximately in the same locations but shifted between them. The strong absorption around channels 10–20 is more important for the third component than for the second, but the latter becomes the dominant contribution in some experiments. This is in agreement with what is observed from the detailed comparison of the spectra of the different runs. Moreover, the recovered pure spectra show a different baseline as a consequence of the observed baseline changes in the experimental spectra.

When the analysis is performed using the second derivative spectra the results are very similar to the ones given in Figures 28 and 29 but with a poorer description of the concentration profiles in some parts of the process, especially for component 2 in the first three runs. The reason for this degradation of the resolution is that the non-negativity constraint cannot be applied over the second derivative unit spectra. The non-negativity constraints applied over both the concentrations and unit spectra imply that the raw spectra contain only positive values, which is not always the case; therefore in the presence of negative absorbance the model will not be correct.

6.3 Example 3: Interpretation of Thermodynamic and Conformational Transitions of Polynucleotides[72]

Nowadays, the experimental monitoring of biochemical processes is relatively straightforward due to the instrumental techniques and data acquisition systems available to the researcher. The typical output of an instrument used in monitoring the evolution of a process according

to the variation of a certain chemical variable consists of arrays of data (e.g. spectra) recorded at certain stages during the reaction (e.g. pH values, temperature, solvent polarity, etc.). These data can be organized in a data matrix, where the rows contain the instrumental responses and the columns reflect the relationship between the variation of the chemical variable and the evolution of the concentration of the species in the process. Despite the availability of such experimental data, univariate monitoring is still widely used in many chemical fields, such as biochemistry. Traditional biochemical studies tend to focus either on obtaining structural information under fixed conditions (e.g. physiological conditions) or on studying dynamic processes by recording univariate measurements (e.g. melting studies using single-wavelength absorbance readings).

Experiments conducted with small molecules have been traditionally interpreted by applying classical iterative least-squares methods based on the refinement of a postulated chemical model to obtain the optimal fit to the experimental data.[73] This approach is often applied to the data coming from the spectrometric monitoring of chemical equilibria and is also known under the name of global analysis.[74] The clear understanding of these simple equilibria allows the following assumptions in the process of model building: the fulfillment of the mass action law (i.e. the validity of a fixed equilibrium constant throughout the reaction); and the one-to-one correspondence between instrumental response and chemical species.

Nevertheless, most of these procedures are unable to interpret many biochemical processes due to the macromolecular nature of many biomolecules (e.g. polynucleotides, proteins, etc.) that causes a more complex evolution of the processes. The inapplicability of the classical methods can be explained by:

1. The existence of polyelectrolyte effects. The mass action law is no longer valid when either important changes in the electric field on the surface of the macromolecule or other effects related to conformational transitions modify the tendency of analogous sites to react. If this happens, Equation (29) gives

$$\log K = f(\alpha) \tag{29}$$

   where α denotes the extent of the reaction. Mathematical expressions which might model this effect cannot be devised since no information concerning its existence and its pattern (i.e. linear or nonlinear) for an unknown macromolecular process is available beforehand.
2. The existence of conformational transitions. Macromolecular biomolecules can show conformational transitions associated with a chemical reaction or with a spatial rearrangement of the molecule due to physical reasons. If the latter phenomenon occurs, the number of structural species will exceed the number of chemical species and the assumption of a one-to-one correspondence between chemical species and spatial configurations will be erroneous.

Of these techniques available for analyzing multivariate data sets, the self-modeling CR methods are the most appropriate for the analysis of evolutionary chemical processes when neither models nor prior information about the number and identity of the species involved are available, as is the case with biomacromolecular equilibria.

To illustrate the usefulness of the soft-modeling CR methods in the interpretation of biomacromolecular processes, the spectrometric (UV and CD) study of the acid–base behavior of the homopolynucleotides poly(uridylic acid) (polyU),[75] poly(cytidylic acid) (polyC)[76] and poly(adenylic acid) (polyA) in dioxane–water (30% v/v) is shown by way of example. Information about the identification and evolution of the chemical species, the existence and pattern of the polyelectrolyte effect and the conformational transitions associated with the studied reactions is presented. The solvent effect on these equilibria can be determined by comparing the results obtained in the hydroorganic mixture with those previously reported in aqueous solution using the same resolution method.

All the experiments were carried out in a dioxane–water (30% v/v) mixture. This hydroorganic mixture is often employed in biocoordination studies to emulate low-polar biological microenvironments.[75,77,78] See the references for experimental details of these studies.

6.3.1 Concentration Profiles and Pure Unit Spectra of the Individual Species Involved in Acid–Base Equilibria

The number of species involved in each acid–base equilibrium and the initial estimates were determined using EFA. The acid–base equilibria of polyC and polyU were explained with two species, related to the protonated and the deprotonated species, whereas the polyA protonation process needed three species to be described. Figures 30 and 31 show the concentration profiles and spectra obtained in the MCR/ALS optimization for the three polynucleotides from the UV and CD titrations, respectively. The agreement between the distribution plots obtained with both spectrometric techniques for the same polynucleotide protonation process reflects the reproducibility of the experimental work and the good performance of the MCR/ALS method on this kind of biochemical data.

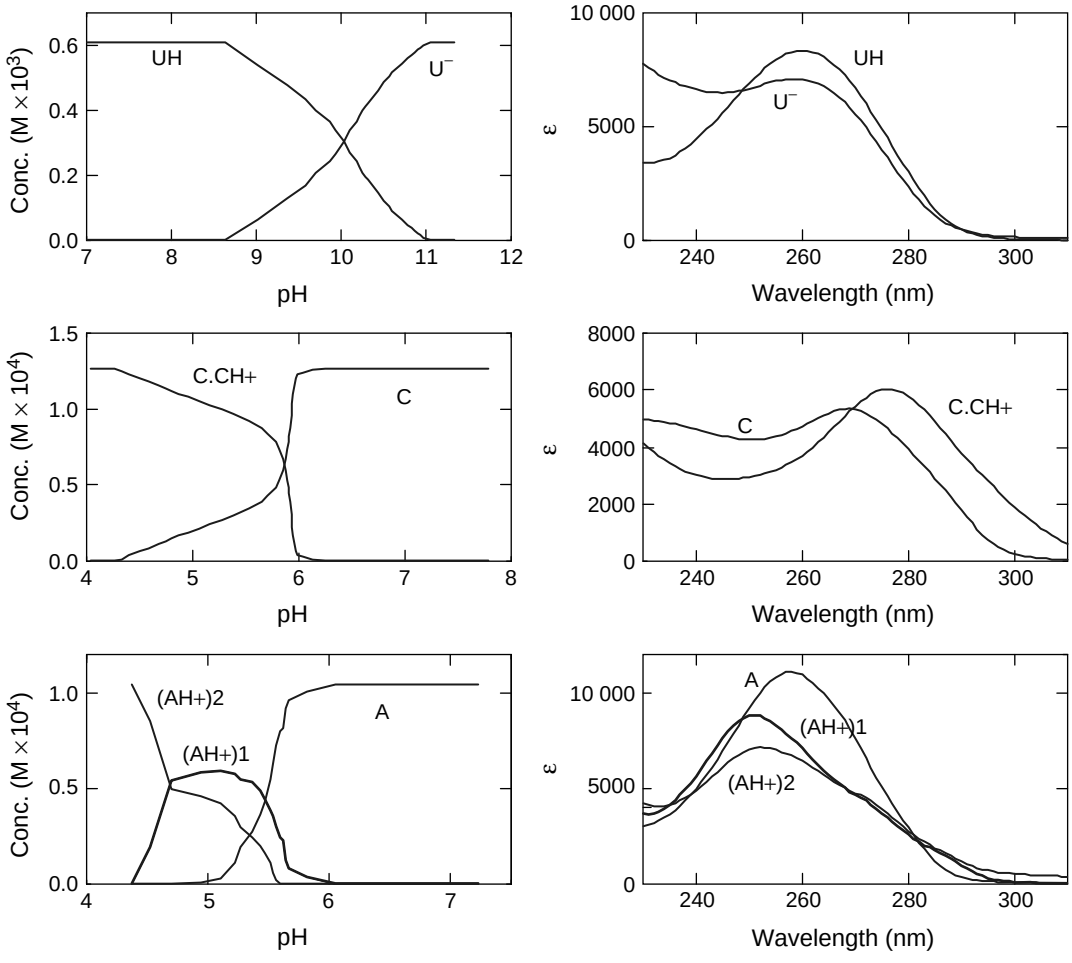

Figure 30 MCR/ALS results from UV titrations in dioxane–water (30% v/v): concentration profiles and UV spectra related to the polyU–H, polyC–H and polyA–H systems. ε values are molar absorptivities.

6.3.2 Detection and Description of Conformational Transitions

The concentration profiles diagram obtained with ALS provides complete information about the evolution of all the species present in the acid–base equilibria. This includes transitions between chemical species which do not involve changes in the spatial structure of the molecule, conformational changes associated with the proton uptake process and changes in the spatial configuration of a single chemical species which do not alter the protonation state of the molecule. Identifying which kind of transition takes place and which conformations are involved depends on the chemical knowledge about the process being analyzed. polyU, polyC and polyA are good examples of the three transitions mentioned above.

Chemical literature on polyU generally accepts the existence of random coil structures related to both protonated and deprotonated species.[79,80] This is confirmed by the low ellipticities in polyU experiments compared with the ellipticities obtained for polyC and polyA experiments (Figure 31). The similar shape of the UV and CD spectra related to poly(UH) and poly(U^-), where the shift between their absorption maxima is the main difference, is in agreement with the hypothesis of a protonation process without conformational changes.

Literature studies of the polyC structure in aqueous solution suggested three different chemical species and conformations depending on the protonation form of the molecule, namely a single-helical deprotonated polyC, a double-stranded helical half-protonated [poly(C) · poly(CH^+)] and a fully protonated random coiled poly(CH^+).[79] Only two species were detected when this equilibrium took place in the hydroorganic mixture because of the narrower pH working range, limited by the precipitation of the charged polynucleotide at more acidic pH values. The most plausible identification of the two existing species includes the presence of the deprotonated and half-protonated polyC species. Apart from the similar shape of the spectra and the similar pH

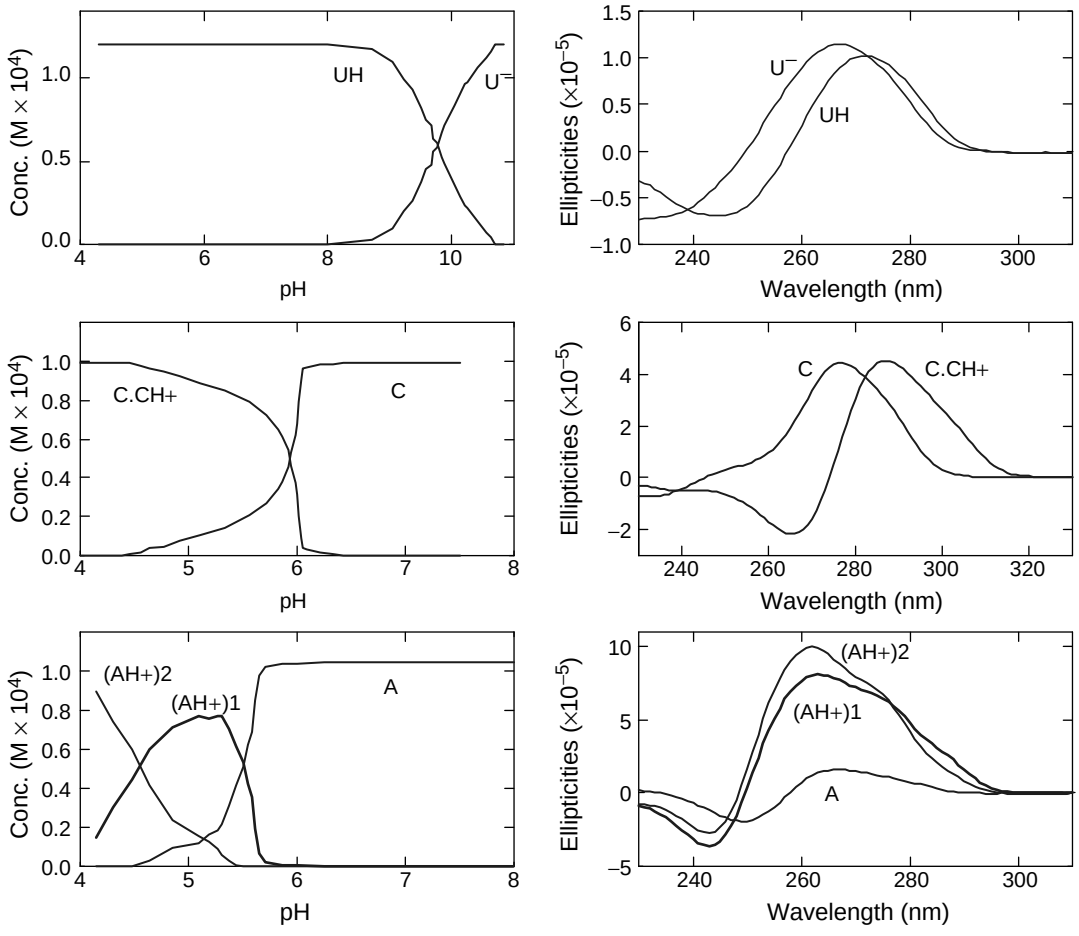

Figure 31 MCR/ALS results from CD titrations in dioxane–water (30% v/v): concentration profiles and CD spectra related to the polyU–H and polyA–H systems.

region of existence of the species presented in Figures 30 and 31 and those found in aqueous solution, the absence of the characteristic decrease in the CD ellipticities associated with a helix to random coil transition, and the more favorable situation of a charged species partially stabilized with the formation of an interstrand hydrogen bond $N-H^+\cdots N$ in comparison with a species with a net charge NH^+ in a low-polar solution, support the identification proposed. Thus, the polyC–H system is a good example of acid–base equilibrium with a conformational change associated with the proton uptake process.

PolyA behaves in a way which is somewhat more complex than that of the two previous polynucleotides. Though the single-helical conformation of deprotonated polyA and its first transition to double-stranded protonated polyA is widely accepted,[79,81–83] different conformational transitions between double-helical configurations of the protonated polynucleotide have also been proposed.[81–83] Figures 30 and 31 show three different species associated with the polyA–H system in the dioxane–water mixture. The similar shape of the two species occurring at more acidic pH values seem to identify them as different double-helical protonated configurations, whereas the species present at higher pH values can probably be attributed to the single-helical deprotonated polyA. The transition between the deprotonated and the first protonated species, poly(AH^+)1, is associated with the proton uptake process, and therefore very reproducible. The transformation between the two protonated forms, poly(AH^+)1 → poly(AH^+)2, presents more irregular concentration profiles owing to the lack of chemical reaction and to the probable time-dependency of the process.[83] Nevertheless, CD and UV spectra show coherent hyperchromism and hypochromism, respectively, when going from poly(AH^+)1 to poly(AH^+)2. Such a phenomenon could be explained by the gradual minimization of the electrostatic repulsion of neighboring phosphate groups due to the increase in the stabilizing electrostatic interactions between the negative charges of these groups and the protonated sites of the adenine bases as the protonation proceeds. Such a stabilizing effect would probably allow the formation of a more compact

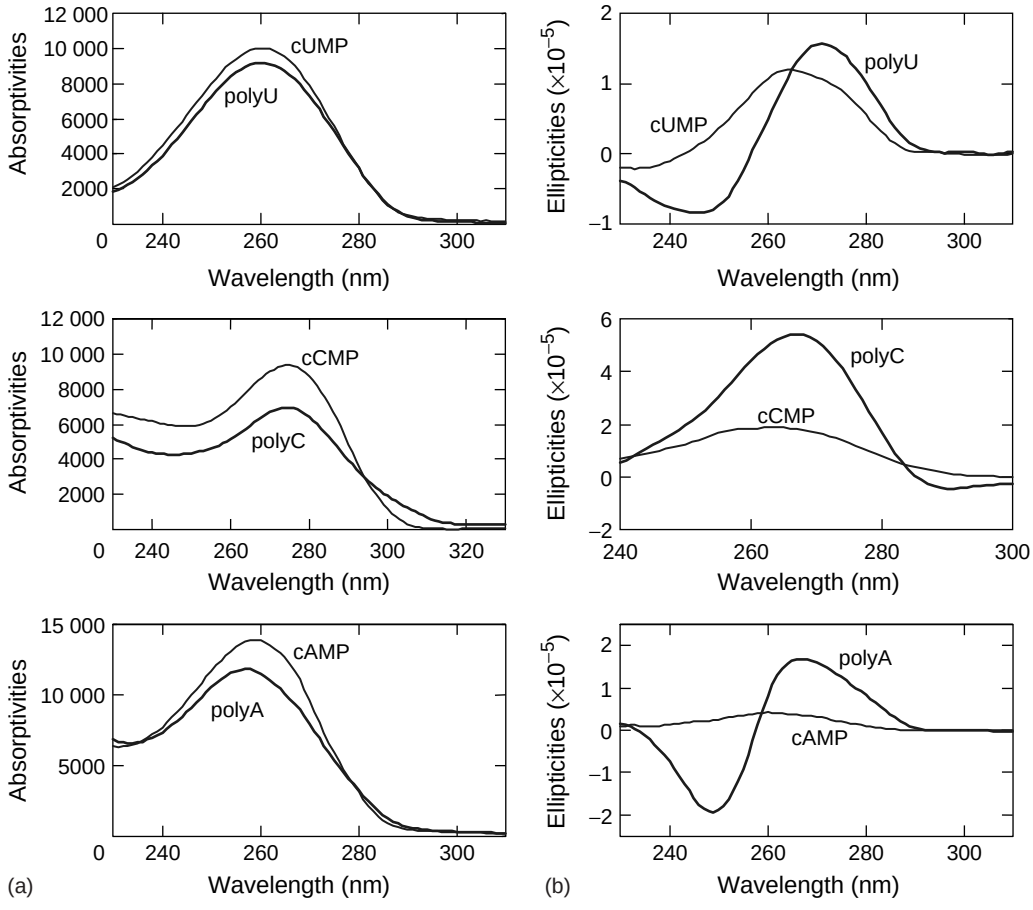

Figure 32 (a) UV and (b) CD spectra of the polymer and the respective cyclic nucleotide (polyU/cUMP, polyC/cCMP and polyA/cAMP) in dioxane–water (30% v/v).

and ordered structure with the phosphate groups occurring closer together and with a consequent stronger base stacking.

The presence of highly ordered polymeric structures could be confirmed by comparing the polynucleotide spectra with the spectra of their respective cyclic nucleotides, as shown in Figure 32. The marked hyperchromism in the CD spectra and hypochromism in the UV spectra of polyA and polyC with respect to cyclic adenosine monophosphate (cAMP) and cyclic cytidine monophosphate (cCMP) indicate a significant stacking in both polynucleotides and, therefore, the existence of ordered structures. The slight differences in terms of intensity between the polyU and cyclic uridine monophosphate (cUMP), and their respective low ellipticities compared with the other two previous cases, support the hypothesis of random structure for this polymer.

6.3.3 Existence and Pattern of Polyelectrolyte Effect

Once the species in the polynucleotide protonation process are identified, the equilibrium constant of this reaction can be properly evaluated. The distribution plot obtained by applying the MCR/ALS method provides an estimation of the concentration values of all the protonated and the deprotonated species for each pH measured. A $\log K$ value may be then calculated for each titration point, bearing in mind that a fixed $\log K$ value for each functional site of the polynucleotide during the whole protonation process cannot be obtained unless no polyelectrolyte effect exists. Plotting the $\log K$ values versus their corresponding protonation degrees (α) is a graphical way of studying both the existence and the pattern of a polyelectrolyte effect. If this effect is revealed, the so-called apparent constant (K_{app}) of analogous sites of the polynucleotide changes as the protonation process advances and the $\log K$ value usually given is not a thermodynamic constant, but an intrinsic constant (K_{int}) defined as the extrapolated K_{app} value for a protonation degree equal to zero, i.e. for the theoretical point where no effects of neighboring protonated sites are present.

Bearing in mind that the equilibrium process followed is the protonation of the site in the monomer unit of the macromolecule, the concentration of the polynucleotide

is expressed as moles of monomer per volume unit and, therefore, K is defined by Equation (30):

$$K = \frac{[\text{protonated monomer}]}{[\text{deprotonated monomer}]\,[H^+]} \tag{30}$$

and consequently, α is given by Equation (31):

$$\alpha = \frac{[\text{protonated monomer}]}{[\text{total monomer}]} \tag{31}$$

In the expressions below referring to the protonation constants of each polynucleotide, the names between brackets indicate the form of the macromolecule where the protonated or deprotonated monomers are placed.

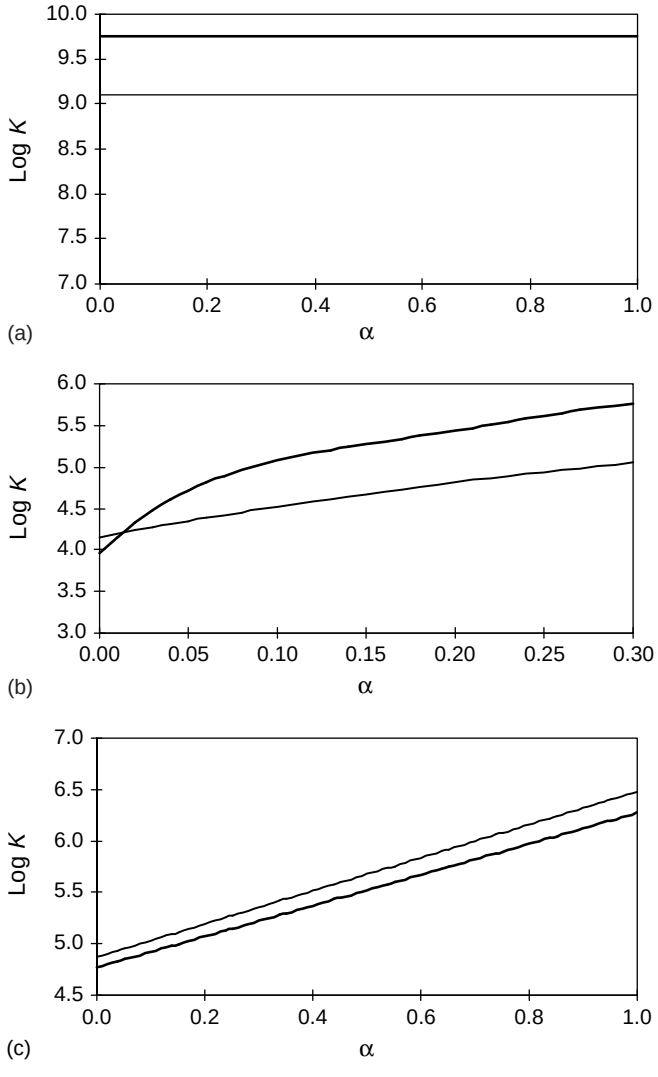

Figure 33 Theoretical models, $\log K = f(\alpha)$, of the polyelectrolyte effect related to the (a) polyU, (b) polyC and (c) polyA protonation processes in dioxane–water (30% v/v) and in aqueous solution (thin line).

Figure 33 shows in thick lines the $\log K$ versus α plots for polyU, polyC and polyA in dioxane–water (30% v/v). The plots include the theoretical lines obtained after the polynomial fitting of the experimental (α, $\log K$) values. The related equations $\log K = f(\alpha)$ can be found in Table 4.

According to the identification of species in the MCR/ALS resolved distribution plots, the polyU protonation constant may be determined as shown in Equation (32):

$$K = \frac{[\text{poly(U}-\text{H)}]}{[\text{poly(U}^-)]\,[H^+]} \tag{32}$$

No polyelectrolyte effect has been detected and, therefore, a thermodynamic protonation constant can be given for all the analogous protonation sites in the macromolecule. The absence of polyelectrolyte effect is probably due to previously detected random coiled structures associated with the protonated and the deprotonated species. These disordered structures are more flexible and allow spatial rearrangements of the macromolecule to minimize the between-sites effect during the protonation process.

The polyC protonation constant has been calculated using Equation (33):

$$K = \frac{[\text{poly(C)} \cdot \text{poly(CH}^+)]/2}{\{[\text{poly(C)}] + [\text{poly(C)} \cdot \text{poly(CH}^+)]/2\}\,[H^+]} \tag{33}$$

Please note that the $[\text{poly(C)} \cdot (\text{poly(CH}^+)]$ concentration is always divided by two when included as protonated and deprotonated forms due to the existence of one protonated base and one deprotonated base per base pair. There is a nonlinear pattern in the polyelectrolyte effect owing to the cooperative action between the protonation process and the formation of the double-stranded helical structure. Thus, the formation of this helix stabilizes the protonated base because of the interstrand hydrogen bond $\text{N}-\text{H}^+ \cdots \text{N}$ and, at the same time, this base pair arrangement is responsible for the growth and stabilization of the helical structure. As the protonation process advances, the intensity of the positive polyelectrolyte effect decreases, becoming negative for α_p values higher than 0.3[76] because the increase of charge density in the macromolecular structure means that the repulsive effect between the protonated sites is more important than the stabilization caused by the formation of the double-helical structure.

The equation related to the polyA protonation process is given by Equation (34):

$$K = \frac{([\text{poly(AH}^+)1] + [\text{poly(AH}^+)2])}{[\text{poly(A)}][H^+]} \tag{34}$$

Table 4 Results of the experiments performed

System	Solvent	Technique	Data treatment	log K	σ[a]	χ^{2}[a]	Lack of fit (%)[b]	Fit log $K = f(\alpha)$	
								Model	r^2
cUMP–H	Dioxane–water (30%)	Potentiometry	CV[c]	9.250(4)[d]	2.14	60.09	–	–	–
cCMP–H	Dioxane–water (30%)	Potentiometry	CV	3.766(3)	2.41	34.03	–	–	–
cAMP–H	Dioxane–water (30%)	Potentiometry	CV	3.195(8)	1.29	30.55	–	–	
polyU–H	Water[e]	Potentiometry	CV	9.364	–	–	–	$\log K = 9.364$	–
	Water	CD	ALS	9.1(2)	–	–	5.9	$\log K = 9.1$	–
	Dioxane–water (30%)	Potentiometry	CV	9.756(4)	1.40	80	–	$\log K = 9.756$	–
	Dioxane–water (30%)	UV	ALS	9.9(2)	–	–	6.2	$\log K = 9.9$	–
	Dioxane–water (30%)	CD	ALS	9.75(6)	–	–	5.8	$\log K = 9.75$	–
polyC–H[f]	Water[e]	UV	ALS	4.21(5)	–	–	1.3	$\log K = 4.21 + 4.6\alpha - 4.5\alpha^2$	0.980
	Water	CD	ALS	4.16(4)	–	–	3.6	$\log K = 4.16 + 3.9\alpha - 3.1\alpha^2$	0.985
	Dioxane–water (30%)	UV	ALS	4.04(2)	–	–	1.6	$\log K = 4.04 + 17\alpha - 110\alpha^2 + 350\alpha^3 - 400\alpha^4$	0.996
	Dioxane–water (30%)	CD	ALS	3.96(2)	–	–	3.5	$\log K = 3.96 + 21\alpha - 140\alpha^2 + 480\alpha^3 - 600\alpha^4$	0.996
polyA–H[f]	Water[e]	CD, UV	ALS	4.87(4)	–	–	2.0	$\log K = 1.61\alpha + 4.87$	0.94
	Dioxane–water (30%)	UV	ALS	4.74(6)	–	–	3.3	$\log K = 1.57\alpha + 4.74$	0.97
	Dioxane–water (30%)	CD	ALS	4.78(4)	–	–	4.9	$\log K = 1.46\alpha + 4.78$	0.99

[a] Figures of merit related to the curve fitting program SUPERQUAD:[85] σ, ratio between the root mean square of the weighted residuals and the estimated error in the working conditions, and the statistical parameter χ^2 based on the distribution of the weighted residuals.

[b] Figure of merit lack of fit related to ALS: lack of fit $= \sqrt{\sum_{ij}(d_{ij} - \widehat{d}_{ij})^2 \Big/ \sum_{ij} d_{ij}^2}$ where d_{ij} and $\widehat{d}_{ij}$ are, respectively, the experimental and ALS calculated data values.

[c] CV, classical least-squares curve fitting procedure (SUPERQUAD program[85]).

[d] Numbers in parentheses are the errors associated with the last numeral.

[e] Results in aqueous solution are taken from Casassas et al.[80] for polyU, Casassas et al.[84] for polyC, and Casassas et al.[81] for polyA.

[f] log K values for polyC and polyA are always intrinsic constants.

A linear positive polyelectrolyte effect caused by the stabilizing electrostatic interactions between the negatively charged phosphate groups and the protonated adenine sites is shown. The difference between the polyelectrolyte effect patterns of polyC and polyA depends on the role of the protonation sites in the formation of the double-stranded helix of the polynucleotide. PolyC protonated sites are responsible for the formation of one of the interstrand hydrogen bonds, whereas polyA protonated sites, though having a positive effect on the stabilization of the double-helical structure because of the minimization of the electrostatic repulsion of the phosphates, do not participate directly in the interstrand hydrogen bonding.[82] Thus, a nonlinear pattern of the polyelectrolyte effect can in all likelihood be attributed to the presence of a cooperative mechanism associated with the protonation process, as occurs in polyC, whereas the more common linear pattern would appear to be caused mainly by purely electrostatic interactions, as shown in polyA and in other simpler polyelectrolytes, such as polyacrylic acid.

A good agreement between the numerical values of polyU and of cUMP protonation constants and between the intrinsic protonation constant of polyC and the protonation constant of cCMP is shown, the slight differences between the values of K_{polyU} and K_{cUMP} being due to hindrance effects (see Table 4). No such agreement is noticed when the cAMP protonation constant and the polyA intrinsic protonation constant are compared. The significantly lower value of the cAMP protonation constant can probably be explained by the absence of the positive

influence of the surrounding phosphate groups, present on the protonation of all the adenine sites of polyA, on the protonation process of the cyclic nucleotide, cAMP.

6.3.4 *Solvent Effect on the Acid–Base Equilibria of Polynucleotides*

The acid–base equilibria of polyA, polyC and polyU were also studied in aqueous solution.[80,81,84] The comparison of the results obtained in water with those obtained in dioxane–water (30% v/v) allows the inference of solvent effects on the protonation processes of the polynucleotides. The system dioxane–water was used because, in comparison with some other hydroorganic mixtures, it provides a larger decrease of polarity with the inclusion of smaller amounts of organic cosolvent. No experiments were performed at dioxane proportions higher than 30% (v/v) because some of the polynucleotides studied presented precipitation phenomena.

When comparing the experiments mentioned above, a general observation is the greater instability of charged macromolecular structures in the hydroorganic medium. Indeed, some species detected in aqueous solutions are apparently not formed in the lower-polar medium because the macromolecule is precipitated at suitable pH values. This is the case for the fully protonated poly(CH^+), whose formation involves the appearance of two net charges in solution, one from the new protonated site and the other from the breaking of the interstrand hydrogen-bonding $N{-}H^+ \cdots N$. The precipitation of the macromolecule from dioxane–water solution also takes place for the protonated polyA at pH values lower than 4 and this prevents the formation of other postulated double-helical configurations at more acidic pH values. Disordered charged structures, such as deprotonated polyU, are more stable than ordered forms in dioxane–water solution due to the greater ability of the macromolecule to reach spatial arrangements suitable for accommodating the net charges.

All the species detected in dioxane–water (30% v/v) are present in water with the same spatial structure. The only difference is the formation of more relaxed configurations in the dioxane–water mixtures because of the weakening of the base stacking interactions caused by the low-polar solvents.

The polyelectrolyte effect, if absent or linear, maintains the same behavior in both solvents studied, as shown in Table 4 and in Figure 33. Thus, the polyU acid–base behavior does not present a polyelectrolyte effect either in water or in the working dioxane–water mixture and the only difference between the polyU equilibrium in both media is the higher value of the protonation constant in the hydroorganic mixture owing to the lower stabilizing effect of this low-polar solvent on the negatively charged deprotonated polynucleotide. Linear models with fairly similar slopes describe the polyelectrolyte effect associated with the polyA protonation process in hydroorganic and in aqueous solutions. Nevertheless, there is not enough information to determine whether the small difference between slopes is due to the intrinsic similarity of the polyelectrolyte effects in both media or whether the stabilizing effect of the inert salt counterions around the negatively charged phosphates conceals the real solvent effect on the polyA protonation process. To clarify this point, it would be necessary to work at lower ionic strengths.

There is a clear solvent influence on the pattern of the polyelectrolyte effect when a cooperative mechanism is involved in the polynucleotide protonation process, as can be seen in the polyC protonation process. Though not visible in Figure 33, where only the mathematically fitted positive polyelectrolyte effect is shown, there is a change in the sign of the polyelectrolyte effect associated with the polyC protonation process in both water and water–dioxane solutions.[76] Solvent effect is first noticed in the ranges of existence of the positive and the negative polyelectrolyte effects. Whereas the change of sign appears in α values around 0.3 in the hydroorganic solution, this negative behavior is not seen until α values higher than 0.5 in water. The greater stability of charged structures in water explains why the negative polyelectrolyte effect does not appear in this solvent until the breakdown of the stable double-stranded helical structure takes place, whereas an increase in the charge density of the double-stranded helix is enough to change the tendency of the polyelectrolyte effect in less polar media. The second difference concerns the pattern of the positive effect in both media: in the hydroorganic mixture, a fourth-order polynomial is needed to explain the variation of $\log K$ with α, whereas a second-order polynomial is enough to explain the same data in water, as shown in Table 4. A much steeper effect for low α values is detected in water–dioxane because of the easier formation of the interstrand hydrogen bond $N{-}H^+ \cdots N$ due to the weaker competition of the solvent molecules in the development of these interactions in a hydroorganic mixture less polar than water. As the protonation degree increases, this tendency is inverted and the polyelectrolyte effect in water–dioxane becomes less and less pronounced, being close to a plateau for α values next to 0.3. The smoothing in the evolution of the polyelectrolyte effect comes from the gradual balance between the favorable tendency to form the hydrogen bond $N{-}H^+ \cdots N$ and the destabilizing influence associated with the increase of charge density in the macromolecular structure.

ABBREVIATIONS AND ACRONYMS

ALS	Alternating Least Squares
cAMP	Cyclic Adenosine Monophosphate
cCMP	Cyclic Cytidine Monophosphate
CD	Circular Dichroism
CR	Curve Resolution
cUMP	Cyclic Uridine Monophosphate
CV	Classical Least-squares Curve Fitting Procedure
DTD	Direct Trilinear Decomposition
EFA	Evolving Factor Analysis
FA	Factor Analysis
FSMW/EFA	Fixed-size Moving Window/ Evolving Factor Analysis
GRAM	Generalized Rank Annihilation Method
HELP	Heuristic Evolving Latent Projections
HPLC/DAD	High-performance Liquid Chromatography/Diode-array Detector
IND	Indicator Function
IR	Infrared
ITTFA	Iterative Transformation Target Factor Analysis
KSFA	Key Set Factor Analysis
LC/DAD	Liquid Chromatography/Diode-array Detection
LC/MS	Liquid Chromatography/Mass Spectrometry
LSER	Linear Solvation Energy Relationships
MCR	Multivariate Curve Resolution
MCR/ALS	Multivariate Curve Resolution/ Alternating Least Squares
NIPALS	Nonlinear Iterative Partial Least Squares
OPA	Orthogonal Projection Approach
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
polyA	Poly(adenylic Acid)
polyC	Poly(cytidylic Acid)
polyU	Poly(uridylic Acid)
PRESS	Prediction Error Sum of Squares
RAEFA	Rank Annihilation Evolving Factor Analysis
SFA	Subwindow Factor Analysis
SIMPLISMA	Simple-to-use Interactive Self-modeling Analysis
SVD	Singular Value Decomposition
TFA	Target Factor Analysis
UV	Ultraviolet
WFA	Window Factor Analysis

RELATED ARTICLES

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher • Signal Processing in Analytical Chemistry

Infrared Spectroscopy **(Volume 12)**
Spectral Data, Modern Classification Methods for

Kinetic Determinations **(Volume 12)**
Data Treatment and Error Analysis in Kinetics

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. E.R. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, Wiley-Interscience, New York, 1992.
2. W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, F. Erni, 'Comparison of Regularized Discriminant Analysis, Linear Discriminant Analysis and Quadratic Discriminant Analysis, Applied to NIR Data', *Anal. Chim. Acta*, **329**, 257–265 (1996).
3. J.S. i Salau, R. Tauler, J.M. Bayona, I. Tolosa, 'Input Characterization of Sedimentary Organic Contaminants and Molecular Markers in the Northwestern Mediterranean Sea by Exploratory Data Analysis', *Environ. Sci. Technol.*, **37**, 3482–3490 (1997).
4. S. Wold, 'Soft-modeling by Latent Variables: the Nonlinear Iterative Partial Least Squares (NIPALS) Algorithm', in *Perspectives in Probability and Statistics*, ed. J. Garni, Academic Press, London, 117–142, 1975.
5. G.H. Golub, C. Reinsch, 'Singular Value Decomposition and Least Squares Solutions', *Numer. Math.*, **14**, 403–420 (1970).
6. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, 'Handbook of Chemometrics and Qualimetrics', in *Data Handling in Science and Technology*, Elsevier, Amsterdam, Vol. 20, 1997.
7. E.R. Malinowski, 'Theory of the Distribution of the Error Eigenvalues Resulting from Principal Component Analysis with Application to Spectroscopic Data', *J. Chemom.*, **1**, 33–40 (1987).
8. S. Wold, 'Cross-validatory Estimation of the Number of Components in Factor and Principal Component Models', *Technometrics*, **20**, 397–405 (1978).
9. E.R. Malinowski, 'Statistical F-tests for Abstract Factor Analysis and Target Testing', *J. Chemom.*, **3**, 49–60 (1988).

10. R.J. Rummel, *Applied Factor Analysis*, Northwestern University Press, Evanston, IL, 1970.
11. H.F. Kaiser, 'The Varimax Criterion for Analytic Rotation in Factor Analysis', *Psychometrika*, **23**, 187–200 (1958).
12. S.Q. Xia, P.L. Su, J.H. Sun, Z.X. Pan, M.S. Zhang, L.M. Shi, 'Determination of Six Aromatic Compounds by Using Target Factor Analysis UV Spectrometry', *Fresenius' J. Anal. Chem.*, **351**, 325–327 (1995).
13. E. Casassas, G. Fonrodona, A. de Juan, R. Tauler, 'Assessment of Solvent Parameters and their Correlation with Protonation Constants in Dioxane–Water Mixtures Using Factor Analysis', *Chemom. Intell. Lab. Syst.*, **12**, 29–38 (1991).
14. E.R. Malinowski, 'Theory of the Error for Target Factor Analysis with Applications to Mass Spectrometry and Nuclear Magnetic Resonance Spectrometry', *Anal. Chim. Acta*, **103**, 339–354 (1978).
15. H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbühler, 'Calculation of Equilibrium Constants from Multiwavelength Spectroscopic Data', III. Model-free Analysis of Spectrophotometric and ESR Titrations', *Talanta*, **32**, 1133–1139 (1985).
16. M. Maeder, 'Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks', *Anal. Chem.*, **59**, 527–530 (1987).
17. R. Tauler, E. Casassas, 'Principal Component Analysis Applied to the Study of Successive Complex Formation Data in the Cu(II) Ethanolamine Systems', *J. Chemom.*, **3**, 151–161 (1988).
18. H.R. Keller, D.L. Massart, 'Peak Purity Control in Liquid Chromatography with Photodiode Array Detection by Fixed Size Moving Window Evolving Factor Analysis', *Anal. Chim. Acta*, **246**, 379–390 (1991).
19. J. Toft, O.M. Kvalheim, 'Eigenstructure Tracking Analysis for Revealing Noise Patterns and Local Rank in Instrumental Profiles: Application to Transmittance and Absorbance IR Spectroscopy', *Chemom. Intell. Lab. Syst.*, **19**, 65–73 (1993).
20. W.H. Lawton, E.A. Sylvestre, 'Self Modeling Curve Resolution', *Technometrics*, **13**, 617–633 (1971).
21. E. Spjotvoll, H. Martens, R. Volden, 'Restricted Least Squares Estimation of the Spectra and Calibration of Two Unknown Constituents Available in Mixtures', *Technometrics*, **24**, 173–180 (1982).
22. O.S. Borgen, B.R. Kowalski, 'An Extension of the Multivariate Component-resolution Method to Three Components', *Anal. Chim. Acta*, **174**, 1–26 (1985).
23. J.C. Hamilton, P.J. Gemperline, 'Mixture Analysis Using Factor Analysis. II. Self-modeling Curve Resolution', *J. Chemom.*, **4**, 1–13 (1990).
24. B.G.M. Vandeginste, W. Derks, G. Kateman, 'Multicomponent Self-modeling Curve Resolution in High Performance Liquid Chromatography by Iterative Target Transformation Factor Analysis', *Anal. Chim. Acta.*, **173**, 253–264 (1985).
25. H. Gampp, M. Maeder, Ch.J. Meyer, A.D. Zuberbühler, 'Quantification of a Known Component in an Unknown Mixture', *Anal. Chim Acta*, **193**, 287–292 (1987).
26. W. Windig, J. Guilment, 'Interactive Self-modeling Mixture Analysis', *Anal. Chem.*, **63**, 1425–1432 (1991).
27. R. Tauler, E. Casassas, 'Application of Principal Component Analysis to the Study of Multiple Equilibria Systems. Study of Copper(II) Salicylate Monoethanolamine, Diethanolamine and Triethanolamine Systems', *Anal. Chim. Acta*, **223**, 257–268 (1989).
28. R. Tauler, A.K. Smilde, B.R. Kowalski, 'Selectivity, Local Rank, Three-way Data Analysis and Ambiguity in Multivariate Curve Resolution', *J. Chemom.*, **9**, 31–58 (1995).
29. R. Manne, 'On the Resolution Problem in Hyphenated Chromatography', *Chemom. Intell. Lab. Syst.*, **27**, 89–94 (1995).
30. E.R. Malinowski, 'Window Factor Analysis: Theoretical Derivation and Application to Flow Injection Analysis Data', *J. Chemom.*, **6**, 29–40 (1992).
31. O.M. Kvalheim, Y.Z. Liang, 'Heuristic Evolving Latent Projections–Resolving 2-way Multicomponent Data. 1. Selectivity, Latent Projective Graph, Datascope, Local Rank and Unique Resolution', *Anal. Chem.*, **64**, 936–946 (1992).
32. P.J. Gemperline, 'A Priori Estimates of the Elution Profiles of the Pure Components in Overlapped Liquid Chromatography Peaks Using Target Factor Analysis', *J. Chem. Inf. Comput. Sci.*, **24**, 206–212 (1984).
33. R. Tauler, 'Multivariate Curve Resolution Applied to Second Order Data', *Chemom. Intell. Lab. Syst.*, **30**, 133–146 (1995).
34. F.C. Sánchez, J. Toft, B. van den Bogaert, D.L. Massart, 'Orthogonal Projection Approach Applied to Peak Purity Assessment', *Anal. Chem.*, **68**, 79–85 (1996).
35. E. Sánchez, B.R. Kowalski, 'Generalized Rank Annihilation Method', *Anal. Chem.*, **58**, 496–499 (1986).
36. E. Sánchez, B.R. Kowalski, 'Tensorial Resolution: a Direct Trilinear Decomposition', *J. Chemom.*, **4**, 29–45 (1990).
37. R.A. Harshman, 'Foundations of the PARAFAC Procedure: Models and Conditions for an ''Explanatory'' Multi-modal Factor Analysis', *UCLA Working Papers in Phonetics*, 1–84 (1970).
38. J.D. Carroll, J. Chang, 'Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of ''Eckart-Young'' Decomposition', *Psychometrika*, 283–319 (1970).
39. R. Bro, 'PARAFAC: Tutorial and Applications', *Chemom. Intell. Lab. Syst.*, 149–171 (1997).
40. A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, 'Multicomponent Determination of Chlorinated Hydrocarbons Using a Reaction-based Chemical Sensor. 3. Medium-rank 2nd-order Calibration with Restricted Tucker Models', *Anal. Chem.*, **66**, 3345–3351 (1994).

41. R. Tauler, D. Barceló, 'Multivariate Curve Resolution Applied to Liquid Chromatography–Diode Array Detection', *Trends Anal. Chem.*, **12**, 319–327 (1993).
42. R. Gargallo, R. Tauler, A. Izquierdo-Ridorsa, 'Application of a Soft Modeling Curve Resolution Method to the Analysis of Second-order Melting Data of Synthetic and Natural Polynucleotides', *Anal. Chem.*, **69**, 1785–1792 (1997).
43. R. Tauler, A. Izquierdo-Ridorsa, E. Casassas, 'Simultaneous Analysis of Several Spectroscopic Titrations with Self-modeling Curve Resolution', *Chemom. Intell. Lab. Syst.*, **18**, 293–300 (1993).
44. E.R. Malinowski, 'Automatic Window Factor Analysis. A More Efficient Method for Determining Concentration Profiles from Evolutionary Spectra', *J. Chemom.*, **10**, 273–279 (1996).
45. W. Den, E.R. Malinowski, 'Investigation of Copper(II)–Ethylenediaminetetraacetate Complexation by Window Factor Analysis of Ultraviolet Spectra', *J. Chemom.*, **7**, 89–98 (1993).
46. M.M. Darj, E.R. Malinowski, 'Complexation Between Copper(II) and Glycine in Aqueous Acid Solutions by Window Factor Analysis of Visible Spectra', *Anal. Chem.*, **68**, 1593–1598 (1996).
47. R. Manne, H. Shen, Y. Liang, 'Subwindow Factor Analysis', *Chemom. Intell. Lab. Syst.*, **45**, 171–176 (1999).
48. C.A. Andersson, R. Bro, 'Improving the Speed of Multiway Algorithms: Part I. Tucker3', *Chemom. Intell. Lab. Syst.*, **42**, 93–103 (1998).
49. R. Bro, C.A. Andersson, 'Improving the Speed of Multiway Algorithms: Part II. Compression', *Chemom. Intell. Lab. Syst.*, **42**, 105–113 (1998).
50. E.J. Karjalainen, 'The Spectrum Reconstruction Problem. Use of Alternating Regression for Unexpected Spectral Components in Two-dimensional Spectroscopies', *Chemom. Intell. Lab. Syst.*, **7**, 31–38 (1989).
51. E.R. Malinowski, 'Obtaining the Key Set of Typical Vectors by Factor Analysis and Subsequent Isolation of Component Spectra', *Anal. Chim. Acta*, **134**, 129–137 (1982).
52. A. de Juan, Y. Vander Heyden, R. Tauler, D.L. Massart, 'Assessment of New Constraints Applied to the Alternating Least Squares Method', *Anal. Chim. Acta*, **346**, 307–318 (1997).
53. J. Mendieta, M.S. Díaz-Cruz, R. Tauler, M. Esteban, 'Application of Multivariate Curve Resolution to Voltammetric Data. Part 2: Study of Metal-binding Properties of the Peptides', *Anal. Biochem.*, **240**, 134–141 (1996).
54. R. Bro, S. de Jong, 'A Fast Nonnegativity-constrained Least Squares Algorithm', *J. Chemom.*, **11**, 393–401 (1997).
55. R. Bro, N.D. Sidiropoulos, 'Least Squares Algorithms Under Unimodality and Nonnegativity Constraints', *J. Chemom.*, **12**, 223–247 (1998).
56. P.K. Hopke, 'Target Transformation Factor Analysis', *Chemom. Intell. Lab. Syst.*, **6**, 7–19 (1989).
57. A. de Juan, B. van den Bogaert, F. Cuesta Sánchez, D.L. Massart, 'Application of the Needle Algorithm for Exploratory Analysis and Resolution of HPLC–DAD Data', *Chemom. Intell. Lab. Syst.*, **33**, 133–145 (1996).
58. R.J. Hanson, C.L. Lawson, *Solving Least-squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
59. H. Gampp, M. Maeder, Ch.J. Meyer, A. Zuberbühler, 'Calculation of Equilibrium Constants from Multiwavelength Spectroscopic Data IV. Model Free Least Squares Refinement by Use of Evolving Factor Analysis', *Talanta*, **33**, 943–951 (1986).
60. http://www.ub.es/gesq/eq1_eng.htm
61. A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou, R. Tauler, 'Second-order Multivariate Curve Resolution Applied to Rank Deficient Data Obtained from Acid–Base Spectrophotometric Titrations of Mixtures of Nucleic Bases', *Chemom. Intell. Lab. Syst.*, **38**, 183–196 (1997).
62. M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, 'On the Rank Deficiency and Rank Augmentation of the Spectral Measurement Matrix', *Chemom. Intell. Lab. Syst.*, **33**, 17–33 (1996).
63. M. Gui, S.C. Rutan, A. Agbodjan, 'Kinetic Detection of Overlapped Aminoacids in Thin-layer Chromatography with a Direct Trilinear Decomposition Method', *Anal. Chem.*, **67**, 3293–3299 (1995).
64. R. Tauler, G. Durand, D. Barceló, 'Deconvolution and Quantitation of Unresolved Mixtures in Liquid Chromatography–Diode Array Detection Using Evolving Factor Analysis', *Chromatographia*, **33**, 244–254 (1992).
65. R. Gargallo, F. Cuesta-Sànchez, D.L. Massart, R. Tauler, 'Validation of Alternating Least Squares Multivariate Curve Resolution for the Resolution and Quantitation of Overlapped Peaks Obtained in Liquid Chromatography with Diode Array Detection', *Trends Anal. Chem.*, **15**, 279–286 (1996).
66. S. Lacorte, D. Barceló, R. Tauler, 'Quantitation and Validation of Trace Herbicides Using On-line Solid-phase Extraction Followed by Liquid Chromatography Diode-array and Multivariate Curve Resolution', *J. Chromatogr. A*, **697**, 345–355 (1995).
67. R. Tauler, S. Lacorte, D. Barceló, 'Application of Multivariate Curve Self-modeling Curve Resolution for the Quantitation of Trace Levels of Organophosphorous Pesticides in Natural Waters from Interlaboratory Studies', *J. Chromatogr. A*, **730**, 177–183 (1996).
68. J.M. Salou, M. Honing, R. Tauler, D. Barceló, 'Resolution and Quantitative Determination of Coeluted Pesticide Mixtures in Mass Spectrometry Liquid Chromatography by Multivariate Curve Resolution', *J. Chromatogr. A*, **795**, 3–12 (1998).
69. R. Tauler, B.R. Kowalski, S. Fleming, 'Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial Process', *Anal. Chem.*, **65**, 2040–2047 (1993).

70. H. Martens, T. Naes, *Multivariate Calibration*, Wiley & Sons, New York, 1989.
71. F.C. McLennan, B.R. Kowalski (eds.), *Process Analytical Chemistry*, Chapman & Hall, London, 1995.
72. A. de Juan, A. Izquierdo-Ridorsa, R. Tauler, G. Fonrodona, E. Casassas, 'A Soft Modeling Approach to Interpret Thermodynamic and Conformational Transitions of Polynucleotides', *Biophys. J.*, **73**, 2937–2948 (1997).
73. D.J. Legget (ed.), *Computational Methods for the Determination of Formation Constants*, Modern Inorganic Chemistry Series, Plenum Press, New York, 1985.
74. R.M. Dyson, S. Kardeli, G.A. Lawrence, M. Maeder, A. Zuberbühler, 'Second Order Global Analysis: the Evaluation of Series of Spectrophotometric Titrations for Improved Determination of Equilibrium Constants', *Anal. Chim. Acta*, **353**, 381–393 (1997).
75. A. de Juan, G. Fonrodona, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler, E. Casassas, 'Application of a Self-modeling Curve Resolution Approach to the Study of Solvent Effects on the Acid–Base and Copper(II)-complexing Behaviour of Polyuridylic Acid', *J. Inorg. Biochem.*, **63**, 155–173 (1996).
76. A. de Juan, A. Izquierdo-Ridorsa, R. Gargallo, R. Tauler, G. Fonrodona, E. Casassas, 'Three-way Curve Resolution Applied to the Study of Solvent Effect on the Thermodynamic and Conformational Transitions Related to the Protonation of Polycytidylic Acid', *Anal. Biochem.*, **249**, 174–183 (1997).
77. H. Sigel, 'Interactions of Metal Ions with Nucleotides and Nucleic Acids and their Constituents', *Chem. Soc. Rev.*, **22**, 255–267 (1993).
78. H. Sigel, R.B. Martin, R. Tribolet, U.K. Häring, R. Malini-Balakrishnan, 'An Estimation of the Equivalent Solution Dielectric Constant in the Active-site Cavity of Metalloenzymes. Dependence of Carboxylate–Metal Ion Complex Stabilities on the Polarity of Mixed Aqueous/Organic Solvents', *Eur. J. Biochem.*, **152**, 187–193 (1985).
79. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1988.
80. E. Casassas, R. Gargallo, I. Giménez, A. Izquierdo-Ridorsa, R. Tauler, 'Application of an Evolving Factor Analysis-based Procedure to Speciation Analysis in the Copper(II)–Polyuridylic Acid System', *Anal. Chim. Acta*, **283**, 538–547 (1993).
81. E. Casassas, R. Tauler, I. Marqués, 'Interactions of H^+ and Cu(II) ions with Poly(adenylic) Acid: Study by Factor Analysis', *Macromolecules*, **27**, 1729–1737 (1994).
82. V.P. Antao, D.M. Gray, 'CD Spectral Comparisons of the Acid-induced Structures of poly[d(A)], poly[r(A)], poly[d(C)], and poly[r(C)]', *J. Biomol. Struct. Dyn.*, **10**, 819–838 (1993).
83. R. Maggini, F. Secco, M. Venturini, H. Diebler, 'Kinetic Study of Double-helix Formation and Double-helix Dissociation of Polyadenylic Acid', *J. Chem. Soc., Faraday Trans.*, **90**, 2359–2363 (1994).
84. E. Casassas, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler, 'Application of a Multivariate Curve Resolution Procedure for the Study of the Acid–Base and Copper(II) Complexation Equilibria of Polycytidylic Acid', *React. Polym.*, **27**, 1–14 (1995).
85. P. Gans, A. Sabatini, A. Vacca, 'SUPERQUAD: an Improved General Program for Computation of Formation Constants from Potentiometric Data', *J. Chem. Soc., Dalton Trans.*, 1195–1200 (1985).

Electroanalytical Methods: Introduction

Henry S. White
University of Utah, Salt Lake City, USA

Richard M. Crooks
Texas A&M University, College Station, USA

This section focuses on recent advances in the theory, instrumentation, and methods of electroanalytical chemistry. Electroanalytical methods are well-suited for analysis of a wide range of chemical systems and problems of both fundamental and technological interest. These methods tend to have excellent sensitivity, are versatile and relatively inexpensive, and in a growing number of recent examples, exhibit the high spatial resolution required for modern analysis in the arenas of biological chemistry and nanotechnology. For these reasons, electroanalysis remains at the forefront of modern analytical methods. Advances in sensitivity and range of electroanalytical methods during the past decade, including single-molecule and single-reaction event detection, high-speed electrochemical measurements for investigations of fast kinetics, and the growing role of electrochemistry in bioanalysis have had an especially high impact on the field.

The authors contributing to this section have produced a well-rounded perspective of electroanalytical chemistry at the turn of the century. For example, the importance of electroanalytical chemistry to the field of biology and biological chemistry is represented in the article by Ewing, **Neurotransmitters, Electrochemical Detection of**. Here, the focus of the methods is on very high spatial resolution and time-resolved detection of biological signaling events.

The theme of detection and chemical sensing is continued in the chapters by Bard, Peters, and Young. Bard pioneered the field of electrogenerated chemiluminescence, and over the years has played a major role in elucidating the theory and developing new analytical applications of this technique. The article **Chemiluminescence, Electrogenerated** focuses on analytical applications, and readers interested in the use of this high-sensitivity, zero-background method for bioassays are especially fortunate to have this review at their disposal. Peters has been working in the field of electroanalytical chemistry for many years, and in **Selective Electrode Coatings for Electroanalysis** he discusses recent advances and applications of electrochemistry to detection. As the title suggests, the focus of this chapter is on the use of electrodes that have been engineered to respond selectively to target analytes. This theme is continued in **Ion-selective Electrodes: Fundamentals** by Young. The remarkable attributes of electrodes modified with well-defined organic monolayers, and especially their application to electroanalytical chemistry, are described by Finklea in **Self-assembled Monolayers on Electrodes**. Such materials are useful for fundamental studies of electrode processes, but also for designing chemical detection systems.

Similarly to other analytical methods, the usefulness of electroanalytical methods are greatly enhanced when they are coupled to instrumental techniques. The articles **Ion-selective Electrodes: Fundamentals** and **Ultraviolet/Visible Spectroelectrochemistry** by Anderson and Scherson, respectively, provide excellent examples of how optical methods can be coupled with electroanalytical chemistry to learn more about the system under study than either technique could provide if applied independently. Similarly, in **Microbalance, Electrochemical Quartz Crystal** Ward clearly demonstrates that gravimetric methods are extremely powerful for studying both electroactive films sorbed to electrode surfaces and physical properties of electrolyte solutions. In addition to spectroscopic and gravimetric methods, scanning probe microscopies can provide detailed information about the structure and electronic properties of electrode surfaces. The article by Itaya, **Self-assembled Monolayers on Electrodes**, deals with this subject and shows that the structure of solid electrodes, as well as the molecular structure of adsorbed organic monolayers, can be elucidated as a function of electrode potential. Finally, the article by Herrero, **X-ray Methods for the Study of Electrode Interaction**, describes how powerful diffraction methods can be applied in situ for the study of electrode processes.

In-situ methods are, of course, most desirable for studying electrode processes, but often it is either impractical or impossible to use such technique. So it is with ultrahigh vacuum surface spectroscopy. However, in the article **Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques** Soriaga shows how well-defined electrodes, combined with carefully developed emersion procedures, provide detailed chemical and structural information about electrode surfaces.

The study of organic reaction mechanisms at electrode surfaces has been a central focus of electroanalytical chemistry for many years. In the article **Organic Electrochemical Mechanisms**, Andrieux provides an excellent update of the field. Of course not all electrochemical reactions occur at solid electrodes, and in recent years there has been much interest in electron transfer at liquid–liquid interfaces, for among other reasons the key role played by such interfaces in catalytic and biological reactions. This subject is covered in the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

article **Liquid/Liquid Interfaces, Electrochemistry at** by Girault.

Carter provides an excellent review on the many techniques that comprise the more traditional workhorse family of methods that fall under the category of Pulse Voltammetry. Much faster methods, based on the application of ultramicroelectrodes, are described by Forster in the article **Ultrafast Electrochemical Techniques**. These techniques offer great promise in improving the temporal aspects of electroanalytical methods in the coming years.

Taken together, this section of the *Encyclopedia of Analytical Chemistry* provides the reader with an insightful overview of modern electroanalytical chemistry and its application to real-world problems. The remarkable improvements in the sensitivity, spatial resolution, and speed of these methods is sure to expand the scope of their use in the years ahead.

Chemiluminescence, Electrogenerated

Allen J. Bard
University of Texas at Austin, Austin, USA
Jeff D. Debad, Jonathan K. Leland, George B. Sigal, James L. Wilbur, and Jacob N. Wohlstadter
IGEN International, Inc., Gaithersburg, USA

1 Introduction 1
2 Principles and Theory 1
2.1 Annihilation Electrogenerated Chemiluminescence 1
2.2 Coreactant Electrogenerated Chemiluminescence 2
3 Instrumentation 3
4 Analytical Applications 4
4.1 Assays for Electrogenerated Chemiluminescence Coreactants 4
4.2 Assays for Electrogenerated Chemiluminescence Emitters 4
5 Future Considerations 7
Abbreviations and Acronyms 7
Related Articles 7
References 7

Electrogenerated chemiluminescence (ECL) is the process in which electrogenerated species undergo electron transfer reactions to form excited states that emit light. Many molecules have the potential to produce ECL, however $Ru(bpy)_3^{2+}$ (bpy = 2,2′-bipyridine) is the most common emitter used for analytical applications. Application of a voltage to an electrode in the presence of an emitter induces light production and allows for the detection of the emitter at very low concentrations. Advantages over other analytical methods include low backgrounds, precise spatial and temporal control over the emission, and the possibility of signal amplification. Commercial systems exist that use ECL to detect numerous clinically relevant analytes with high sensitivity using a variety of assay formats.

1 INTRODUCTION

ECL (also called electrochemiluminescence) is a term commonly applied to systems in which energetic electron transfer reactions of electrogenerated species occur to produce excited states, usually with regeneration of luminescent species. An important distinction between ECL and chemiluminescence (CL) is control of the light generation process. CL processes are initiated and controlled by mixing and fluid flow, while ECL processes are initiated and controlled by switching an electrode voltage.

ECL reactions were first investigated in the early 1960s and were carried out with systems composed of aromatic compounds in highly purified and carefully deaerated aprotic solvents such as acetonitrile (MeCN). In 1972 ECL of luminescent inorganic species, mainly $Ru(bpy)_3^{2+}$, was introduced. Much of this early work has been reviewed.[1,2] The use of aqueous ECL in 1981 was made possible by the introduction of the concept of coreactants in 1977, and this led to the use of ECL labels in clinical diagnostics.[3–5]

ECL has been used for a wide range of analytical applications.[6] The most important commercial application to date for ECL is its use in diagnostic assays. ECL provides for highly sensitive and precise diagnostic assays and offers several advantages when compared with other methods. Unlike fluorescent methods, ECL does not require expensive and complicated light sources (i.e. lasers) to generate luminescence. Moreover, because ECL excitation is very selective, it does not suffer from the high background signals that are typical of fluorescent methods. ECL has the advantage over analytical CL applications in having high spatial and temporal resolution, since the ECL reaction is initiated at the surface of an electrode and only when the system is electrochemically activated. Moreover, unlike most CL systems, the luminescent species can be regenerated in ECL, allowing the possibility of amplification.

2 PRINCIPLES AND THEORY

2.1 Annihilation Electrogenerated Chemiluminescence

The ECL annihilation mechanism describes a process wherein a compound is electrochemically reduced and oxidized at an electrode (Equations 1 and 2, where E^0_{R,R^-} and $E^0_{R^+,R}$ are reduction and oxidation potentials, respectively). The resulting products can react (Equation 3, a ''radical annihilation reaction'') to produce excited species ($^1R^*$ in Equation 3) capable of emitting light.[7]

$$R + e \longrightarrow R^{\bullet -} \qquad E^0_{R,R^-} \qquad (1)$$

$$R - e \longrightarrow R^{\bullet +} \qquad E^0_{R^+,R} \qquad (2)$$

$$R^{\bullet +} + R^{\bullet -} \longrightarrow {}^1R^* + R \qquad (3)$$

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

If the enthalpy, ΔH^0, of the electron transfer reaction shown in Equation (3) is larger than the energy required to produce the excited singlet state from the ground state, E_S (Equation 4), then one of the products of the reaction is produced with excess energy ($^1R^*$). This excess energy can be emitted as light (Equation 5). This process is known as the ''S-route'', indicating that the luminescence is emitted by a species in an excited singlet state.

$$\Delta H^0 = \Delta G^0 + T\Delta S^0 = E^0_{R^+,R} - E^0_{R,R^-} - 0.1\,\text{eV} > E_S \quad (4)$$

$$^1R^* \longrightarrow R + h\nu \quad (5)$$

It is also possible to use two different precursors, A and D, in an ECL annihilation process (Equations 6–8):

$$A + e \longrightarrow A^{\bullet -} \quad (6)$$

$$D - e \longrightarrow D^{\bullet +} \quad (7)$$

$$D^{\bullet +} + A^{\bullet -} \longrightarrow {}^1A^* + D \quad \text{or} \quad {}^1D^* + A \quad (8)$$

Light may be emitted from either of the possible high-energy products $^1A^*$ or $^1D^*$.

Even if the energetic condition of Equation (4) is not satisfied, it is possible for light emission to occur via the formation of lower energy triplet states, followed by triplet–triplet annihilation (the T-route, Equations 9 and 10):

$$D^{\bullet +} + A^{\bullet -} \longrightarrow {}^3A^* + D \quad (9)$$

$${}^3A^* + {}^3A^* \longrightarrow {}^1A^* + A \quad (10)$$

Direct emission from the triplet state is usually not observed for organic species, because they have long radiative lifetimes and are quenched before emitting. Quenching of the triplets by the radical ions or other species in the system leads to lower emission yields for the T-route than the S-route systems.

The ECL mechanisms shown above generate oxidized and reduced species at a single electrode. It is also possible to obtain emission through the steady generation of the two reactants at two different electrodes that are sufficiently close to allow the reactants to interdiffuse and react.[8,9]

2.2 Coreactant Electrogenerated Chemiluminescence

Although analytical ECL is possible with annihilation systems,[10] most systems require impractical procedures like the use of aprotic solvents and rigorous deaeration of solvents. Annihilation ECL is generally not convenient in aqueous solutions, because the stability range for water oxidation and reduction is too small to generate both species needed for the radical ion annihilation reaction. Moreover, one or both of the precursors are usually unstable in water. It is possible, however, to obtain ECL in aqueous solution through the use of a coreactant.[3,11] ECL coreactants are species that, upon electrochemical oxidation or reduction, produce species that react with other compounds to produce excited species capable of emitting light. Oxalate ion, $C_2O_4^{2-}$, the first ECL coreactant discovered, represents a typical oxidative coreactant.[3] Consider an ECL system composed of species D and oxalate in an aqueous solution. An oxidizing potential at an electrode oxidizes D to produce the radical cation $D^{\bullet +}$. It also oxidizes the oxalate, which decomposes to produce the reducing agent $CO_2^{\bullet -}$ as shown in Equations (11) and (12):

$$C_2O_4^{2-} - e \longrightarrow C_2O_4^{-} \longrightarrow CO_2^{\bullet -} + CO_2 \quad (11)$$

$$D^{\bullet +} + CO_2^{\bullet -} \longrightarrow D^* + CO_2 \quad (12)$$

In ECL systems that use coreactants, the electrode typically only oxidizes or reduces the reagents; chemical transformations of the electrochemical oxidation or reduction products provide the other required species. For example, in the oxalate system the electrode only oxidizes oxalate and the ECL reactant D; the reductant, $CO_2^{\bullet -}$, is generated via Equation (11). This strategy is used in most analytical applications, with the reactant D being $Ru(bpy)_3^{2+}$ (oxidized to $Ru(bpy)_3^{3+}$).

A variety of different coreactants, most notably aliphatic amines, can be used with the $Ru(bpy)_3^{2+}$ system[6,12,13] and form the basis of many analytical schemes described in the following section. Commercial applications typically use $Ru(bpy)_3^{2+}$ as the emitter and tripropylamine (TPA) as a coreactant. In this ECL system, developed by IGEN International Inc.[14] in the 1980s, $Ru(bpy)_3^{2+}$ is excited repeatedly and emits a photon many times (Figure 1).[3,13] This process – often referred to as an ''amplification'' process – allows a single label to generate many photons and contributes to the remarkable sensitivity of ECL systems.

Other systems use coreactants that are reduced to generate reactive species. For example, ECL emission

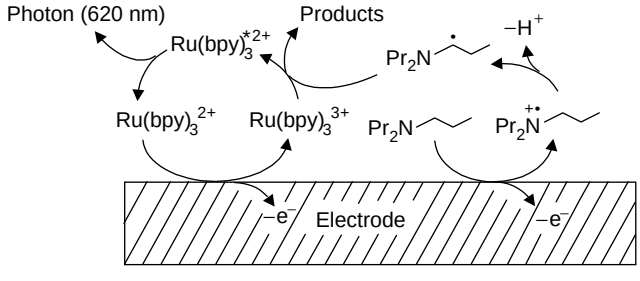

Figure 1 Proposed reaction mechanism for $Ru(bpy)_3^{2+}$/TPA ECL system. The coreactant forms a strong reducing agent upon oxidation that reacts with the oxidized form of the emitter ($Ru(bpy)_3^{3+}$) to generate light.

occurs on reduction of a mixture of $Ru(bpy)_3^{2+}$ and peroxydisulfate ($S_2O_8^{2-}$) via the sequence shown in Equations (13–15):[15]

$$Ru(bpy)_3^{2+} + e \longrightarrow Ru(bpy)_3^{+} \quad (13)$$

$$S_2O_8^{2-} + e \longrightarrow SO_4^{\bullet -} + SO_4^{2-} \quad (14)$$

$$Ru(bpy)_3^{+} + SO_4^{\bullet -} \longrightarrow Ru(bpy)_3^{2+*} + SO_4^{2-} \quad (15)$$

Here, reduction of $S_2O_8^{2-}$ generates the oxidant, $SO_4^{\bullet -}$, which reacts with the ECL label to form the excited state of the label.

3 INSTRUMENTATION

The design of an instrument based on ECL depends on the intended application and the mechanism used for light generation. Three major components are found in most devices: an electrochemical cell, a potentiostat, and a light detection system. Here we discuss these basic elements and other important aspects of ECL; examples of integrated instruments are presented in later sections.

The electrochemical cell is the center of the ECL process. A basic description of electrochemical cells can be found in any introductory electrochemistry text; Figure 2 shows a diagram of a typical electrochemical flow cell used for ECL. Generation of ECL takes place in the reaction chamber, which holds a solution containing the reagents required for ECL. This chamber also houses three electrodes that, during an ECL measurement, are in contact with the solution. The working and counter electrodes are essential for ECL, because they, upon passage of the current, generate the energetic reactants needed for ECL. A reference electrode controls the electrical potential of the working electrode and thus the reactions that occur there. The reference electrode is often separated from the reaction chamber by a porous

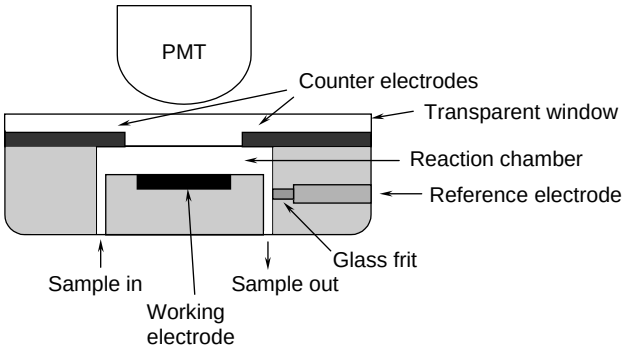

Figure 2 Schematic of a simple electrochemical flow cell used for ECL.

frit that prevents fluid flow between the chamber and the reference electrode compartment. A potentiostat controls and applies potentials in this three-electrode system. Computer controlled potentiostats are readily available that have the flexibility to deliver any waveform required. Solutions within the electrochemical cell must contain an electrolyte to support the passage of current. This requirement may be met in aqueous systems simply by the presence of a buffer, while quaternary ammonium salts are common electrolytes for nonaqueous solvents.

Electrodes used in electrochemical ECL cells are typically made of gold or platinum. Even these robust electrodes may require periodic cleaning, and many cells are designed for easy dismantling to aid in electrode replacement or mechanical polishing. Highly reproducible methods of cleaning the electrodes without removing them from the cell have also been introduced, for example by electrochemical oxidation of water at the electrode surface in basic media.

Light generated by ECL is typically detected with a photodetector such as a photomultiplier tube (PMT), a photodiode, or a charge-coupled device (CCD). The choice of detectors depends on several factors including size, cost, sensitivity, power requirements and the ability to image. The spectral sensitivity of the detector may also be important because different ECL emitters can have very different emission wavelengths. For example, $Ru(bpy)_3^{2+}$ emits red light at 610 nm, $Os(bpy)_3^{2+}$ emits infrared light (720 nm), while diphenylanthracene's emission is blue (420 nm).

Efficient light detection helps to maximize the sensitivity of ECL measurements. Part of the electrochemical cell is transparent so that the light generated at the electrode(s) can escape, and the detector is positioned to maximize collection of light from the electrochemical cell. The cell and photodetector are housed in a light-tight box to reduce background light.

Virtually all integrated systems that use ECL for diagnostic measurements use the ORIGEN™ technology developed and patented by IGEN International, Inc. The essential component of the ORIGEN™ system is the measurement module, which consists of a flow cell containing an electrode and a light detection means such as a PMT or a photodiode. The current design of the ORIGEN™ measurement module includes a resusable platinum or gold electrode to generate ECL. ORIGEN™ assays typically employ magnetic microparticles as a binding phase; these microparticles are concentrated at the surface of the electrode to improve the efficiency and selectivity of the ECL process. More detailed descriptions of certain aspects of the ORIGEN™ technology are described in subsequent sections.

4 ANALYTICAL APPLICATIONS

For convenience, we divide analytical measurements that use ECL into two broad categories: assays for ECL coreactants; and assays for species capable of ECL (emitters). In the first category, biologically or pharmacologically important compounds act as ECL coreactants and react with ECL labels to generate luminescent species. The intensity of ECL measured in the presence of these coreactants is a function of the concentration of the coreactant as well as that of the emitter. ECL measurements in the presence of high, predetermined concentrations of ECL emitters can be used, therefore, as a means to assay for compounds that act as coreactants. The second category of analytical measurements uses ECL emitters as labels in affinity binding assays that attach the ECL emitter to the analyte of interest. The coreactant, typically TPA, is present in high concentrations; the amount of luminescence is thus dependent on the amount of the ECL emitter present in the assay. Since the emitters are bound to the analyte of interest, the amount of luminescence can be correlated with the concentration of the analyte.

4.1 Assays for Electrogenerated Chemiluminescence Coreactants

ECL from $Ru(bpy)_3^{2+}$ has been used to measure the concentration of coreactants such as peroxydisulfate, oxalate, and a variety of amines.[6] ECL assays for amines find many applications because amine groups are prevalent in biologically and pharmacologically important compounds and because amines absorb weakly in the ultraviolet/visible spectrum. ECL assays have been reported for numerous amine-containing compounds including alkylamines, antibiotics, antihistamines, opiates, nicotinamide adenine dinucleotide (reduced form) (NADH) and β-blockers.[6,16] As a general rule, the ECL signal from alkylamine coreactants follows the order: $3^\circ > 2^\circ > 1^\circ$.[13]

In a typical solution-based assay, a sample is combined with a solution containing $Ru(bpy)_3^{2+}$; the resulting solution is mixed and introduced into the electrochemical cell (see Figure 2). An oxidizing potential is applied to the working electrode and the ECL generated at the working electrode is measured using a light detector coupled to the flow cell. Similar methods have been used as detectors for the chromatographic separation of amines and in high-performance liquid chromatography (HPLC) systems that separate and detect amino acids for amino acid analysis.[17] Alternatively, $Ru(bpy)_3^{2+}$ may be immobilized in a thin polymer film deposited directly on the working electrode, which eliminates the need for a constant stream of $Ru(bpy)_3^{2+}$.[18]

ECL has been used to monitor enzymatic reactions by coupling the enzymatic reaction to the generation or consumption of an ECL coreactant. The coenzyme NADH contains an amine moiety and acts as a coreactant for $Ru(bpy)_3^{2+}$, however its oxidized form (NAD^+) is not a coreactant.[18] Scheme 1 illustrates an ECL measurement used to assay for glucose.[19] Numerous NADH-dependent enzymes are known, allowing for the adaptation of this technique to assay for a variety of different analytes.

ECL has also been used to detect β-lactamase activity.[20] Penicillin does not act as a coreactant with $Ru(bpy)_3^{2+}$ to produce ECL but β-lactamase catalyzed hydrolysis of penicillin forms a molecule with a secondary amine that can act as a coreactant (Scheme 2). The efficiency of the ECL process can be increased by direct covalent attachment of the β-lactamase substrate to a $Ru(bpy)_3^{2+}$ derivative.[21]

The ECL detection of enzyme activity has also been used to develop highly sensitive assays in which an enzyme that generates or consumes an ECL coreactant is used as the detectable label on an antibody. The presence of the antibody is determined by measuring the activity of the enzyme label via ECL.

4.2 Assays for Electrogenerated Chemiluminescence Emitters

The most common analytical uses of ECL employ the emitter as a label in affinity binding assays.[5]

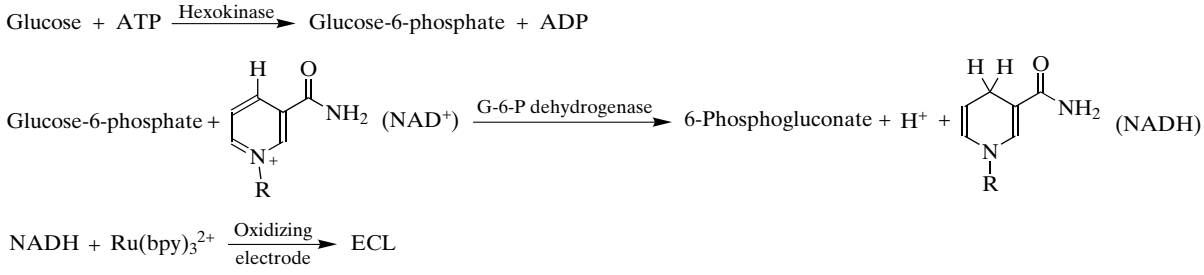

Scheme 1 Schematic of reactions used to measure glucose by ECL. NADH, which is produced enzymatically, is used as a coreactant for $Ru(bpy)_3^{2+}$ ECL.

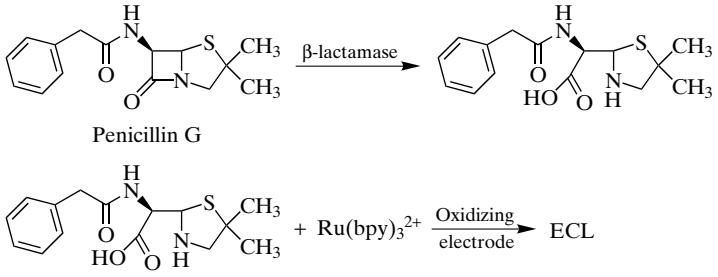

Scheme 2 ECL assay that measures the presence of β-lactamase and/or its activity. Enzymatic hydrolysis of penicillin by β-lactamase creates a coreactant that can be used to produce ECL from Ru(bpy)$_3^{2+}$.

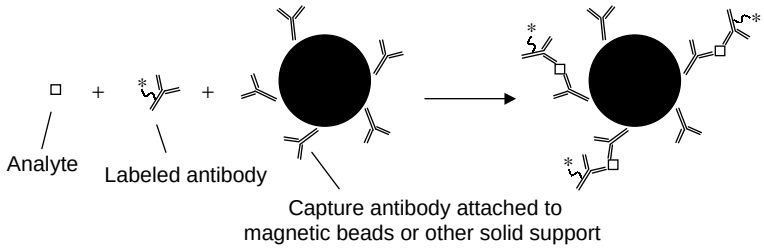

Figure 3 An affinity assay using an antibody–antigen–antibody ''sandwich'' assay. Once the analyte and labeled antibody are bound, the support is washed, and ECL is initiated after addition of a coreactant solution (the asterisk is an ECL label such as Ru(bpy)$_3^{2+}$, and the open square is an analyte of interest).

Affinity binding assays detect an analyte of interest by monitoring the binding of the analyte to one or more binding partners specific for the analyte. The label is physically linked to one of the binding partners in the assay and provides the means for detecting the coupling of the binding partner to the analyte. Several classes of binding partners are used in these assays: antibody/antigen, enzyme/inhibitor, carbohydrate/lectin, and nucleic acid/complementary nucleic acid are the most common.[22]

The ''sandwich assay'' format (Figure 3) is often used for affinity binding assays. In a sandwich immunoassay, a sample is combined with two antibodies specific for different regions of the analyte: one of these antibodies is immobilized on a solid phase support and the other is linked to a detectable label. The presence of analyte leads to the formation of ''sandwich'' complexes that link the detectable label to the solid phase; the concentration of the label on the solid phase is directly proportional to the concentration of analyte. In the ECL version of the sandwich assay, the number of ECL labels on the solid phase are determined through an ECL measurement.

ECL-based affinity assays typically employ coreactants for compatibility with the aqueous conditions used in biological assays. The concentrations of the coreactants are kept constant and high to maximize the sensitivity of the detection and to prevent fluctuations in the concentration of the coreactant from changing the ECL.

Commercial instruments that use ECL and affinity binding assays utilize Ru(bpy)$_3^{2+}$ as the emitter and TPA as coreactant. The coreactant TPA dissolves in buffers, does not interfere with binding reactions, and is very stable. Ru(bpy)$_3^{2+}$ is also very stable, and is a versatile label that can be readily linked to biological molecules using various linking groups, for example, activated *N*-hydroxysuccinamide esters.

Magnetic microparticles are used in most commercial ECL assays as supports for affinity binding assays. In a sandwich assay format, the microparticles are coated with an antibody specific for the analyte to be measured and mixed with a solution that contains both the sample and a second antibody that is linked to one or more ECL labels. Analytes present in the sample bind to the antibodies on the microparticles, and the labeled antibody binds to these analytes to complete the sandwich complex. After the binding reactions are complete, the magnetic microparticles, (and the associated sandwich complexes) are collected on the working electrode by positioning a magnet beneath the working electrode (Figure 4). The electrode is then used to initiate ECL and the intensity of the emitted light is measured. This intensity can be correlated to the amount of emitter, and hence analyte, that is bound to the magnetic beads, which in turn can be correlated with the amount of analyte that was present in the sample. The use of magnetic microparticles enables very sensitive assay measurements by concentrating analyte and labels bound to the particles on the electrode.

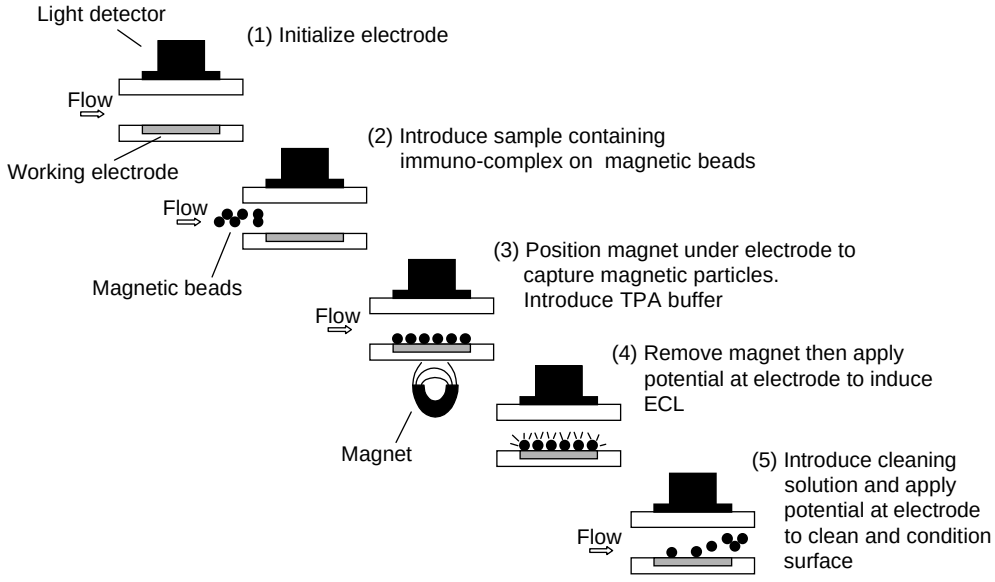

Figure 4 Schematic of process used for ECL assays employing magnetic microparticles. The microparticles are captured by positioning a magnet underneath the working electrode, and the ECL is initiated upon introduction of a coreactant solution. Once the electrode is electrochemically cleaned, another sample can be introduced.

Figure 5 The ORIGEN™ analyzer (IGEN International, Inc.) was the first commercial ECL instrument (a). The ELECSYS™ 2010 (Roche Diagnostics) was designed for use in centralized testing and reference laboratories (b).

The first commercial instrument that used ECL was the ORIGEN™ analyzer (IGEN International, Inc., Figure 5). This ECL analyzer provides highly sensitive and precise assays in an automated format. Typically, the assays use magnetic microparticles as a solid support and $Ru(bpy)_3^{2+}$/TPA as the label and coreactant, respectively. A personal computer controls the instrument and aids in the processing and storage of data. The sample and assay reagents are combined in plastic tubes; these tubes are placed in a carousel (located on the ORIGEN™ analyzer) that can accommodate 50 tubes at one time. The carousel agitates the tubes to mix the sample and reagents and to enhance the speed of the assay binding reactions. After the reactions are complete, the contents of the tubes are combined with other assay reagents and transported into the electrochemical cell. ECL is induced, the light is measured with a PMT, the contents of the cell are removed and the cell is reconditioned for the next measurement. A typical read cycle requires approximately 1 min. The ORIGEN™ analyzer is capable of conducting a broad menu of assays; different assays require only different assay reagents and, occasionally, modest modification of the instrumental procedures.

Roche Diagnostics, a licensee of IGEN International, Inc.'s technology, developed the ELECSYS™ instrument for conducting immunoassays in centralized hospital and reference laboratories. This highly automated instrument offers some of the most sensitive and precise assays available to the medical diagnostics market. (For example, the lower detection limit for the total concentration of the tumor marker prostate specific antigen (PSA) is less than 0.01 ng ml^{-1}.) Its fundamental technology and operation is similar, in principle, to the ORIGEN™ analyzer. The ELECSYS™ offers a menu of more than 40 assays, including assays for tumor markers, cardiac markers, and analytes relevant to infectious diseases, fertility therapies, thyroid disease, and many other clinical conditions. The instrument can operate in random access mode and has the capability to process STAT samples.

ECL has also been used to detect cryptosporidium in water, *Escherichia coli* in foods, and in the very precise quantitation of polymerase chain reaction (PCR) products for DNA-based diagnostic testing.[23] All

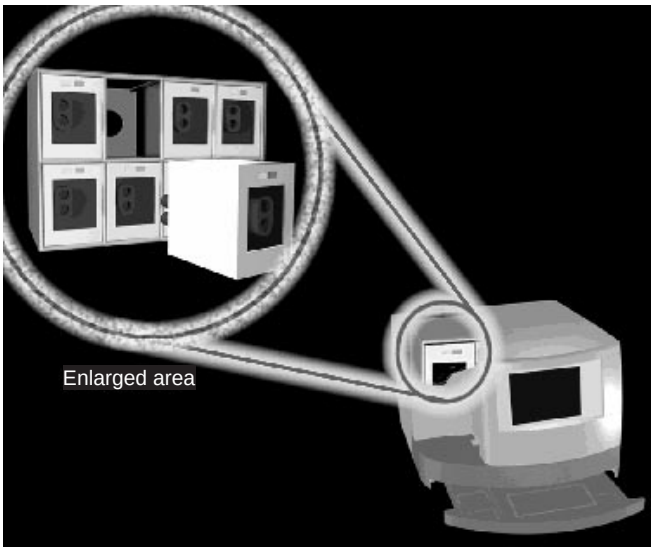

Figure 6 Schematic of an instrument based on the ECL module technology (IGEN International, Inc.). This instrument can be used for high-throughput diagnostic testing and drug discovery. Multiple self-contained, fully functional diagnostic modules (inset) are incorporated into an integrated device. The modules are interchangeable and easily removed and installed. Instruments intended for other applications (i.e. instruments that may have different requirements for operation, throughput and supporting infrastructure) also use ECL modules.

these detection schemes rely on affinity binding mechanisms, and typically use $Ru(bpy)_3^{2+}$/TPA as the ECL reagents.

Recently, IGEN International, Inc. developed a new generation of ECL-based instruments. These instruments are based on IGEN's ECL module, a self-contained, fully functional diagnostic operating system that approximates 1/20th the size of the first generation of ORIGEN™ operating systems and which has accuracy and sensitivity equal to the first generation. The ECL module forms the foundation for a wide variety of diagnostic instruments and allows for flexible design and development of ECL-based instrumentation. ECL module-based systems (Figure 6) should have a variety of applications in patient point-of-care, high-throughput drug screening, food, water and animal health testing.

5 FUTURE CONSIDERATIONS

ECL will continue as an important area of exploration for basic research. It provides a rich platform for understanding fundamental questions in chemistry, biology and physics. Applications for ECL will also continue to expand in scope and importance. Analytical and diagnostic technology based on ECL currently provides a route to simple, low-cost, highly sensitive and very precise measurements. New ECL-based systems will provide sensitive measurements in local doctor's offices, small clinics, emergency rooms, and even in the home. These new developments will expand the impact of ECL in medicine, drug discovery, analytical science, environmental testing and industrial diagnostics.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
CL	Chemiluminescence
ECL	Electrogenerated Chemiluminescence
HPLC	High-performance Liquid Chromatography
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
PCR	Polymerase Chain Reaction
PMT	Photomultiplier Tube
PSA	Prostate Specific Antigen
TPA	Tripropylamine

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Immunochemistry • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

REFERENCES

1. A.J. Bard, L.R. Faulkner, 'Techniques of Electrogenerated Chemiluminescence', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–95, Vol. 10, 1977.
2. R.S. Glass, L.R. Faulkner, 'Electrochemiluminescence', in *Chemical and Biological Generation of Excited States*, eds. W. Adam, G. Cilento, Academic Press, New York, Chapter 6, 1982.
3. I. Rubinstein, A.J. Bard, 'ECL 37. Aqueous ECL Systems Based on $Ru(bpy)_3^{2+}$ and Oxalate or Organic Acids', *J. Am. Chem. Soc.*, **103**, 512–516 (1981).
4. A.J. Bard, G.M. Whitesides, 'Luminescent Metal Chelate Labels and Means for Detection,' US Patent No. 5 221 605 (June 22, 1993).
5. G.F. Blackburn, H.P. Shah, J.H. Kenten, J.K. Leland, R.A. Kamin, J. Link, J. Peterman, M.J. Powell, A. Shah,

D.B. Talley, S.K. Tyagi, E. Wilkins, T.-G. Wu, R.J. Massey, 'Electrochemiluminescence Detection for Development of Immunoassays and DNA Probe Assays for Clinical Diagnostics', *Clin. Chem.*, **37**, 1534–1539 (1991).

6. A.W. Knight, G.M. Greenway, 'Occurrence, Mechanisms and Analytical Applications of Electrogenerated Chemiluminescence', *Analyst*, **119**, 879–890 (1994).
7. K.M. Maness, J.E. Bartelt, R.M. Wightman, 'Effects of Solvent and Ionic Strength on the Electrochemiluminescence of 9,10-Diphenylanthracene', *J. Phys. Chem.*, **98**, 3993–3998 (1994).
8. J.T. Maloy, K.B. Prater, A.J. Bard, 'Electrogenerated Chemiluminescence. V. The Rotating Ring-disk Electrode-digital Simulation and Experimental Evaluation', *J. Am. Chem. Soc.*, **93**, 5959 (1971).
9. G.H. Brilmyer, A.J. Bard, 'Electrogenerated Chemiluminescence. XXXVI. The Production of Steady Direct Current ECL in Thin Layer and Flow Cells', *J. Electrochem. Soc.*, **127**, 104 (1980).
10. S.A. Cruser, A.J. Bard, 'Concentration–Intensity Relationships in Electrogenerated Chemiluminescence', *Anal. Lett.*, **1**, 11–17 (1967).
11. M.-M. Chang, T. Saji, A.J. Bard, 'ECL 30. Electrochemical Oxidation of Oxalate in the Presence of Luminescers in Acetonitrile Solution', *J. Am. Chem. Soc.*, **99**, 5399–5403 (1977).
12. J.B. Noffsinger, N.D. Danielson, 'Generation of CL upon Reaction of Aliphatic Amines with $Ru(bpy)_3^{2+}$', *Anal. Chem.*, **59**, 865–868 (1987).
13. J.K. Leland, M.J. Powell, 'ECL: An Oxidative–Reduction Type ECL Reaction Sequence Using Tripropyl Amine', *J. Electrochem. Soc.*, **137**, 3127–3131 (1990).
14. See www.igen.com.
15. H.S. White, A.J. Bard, 'ECL 41. ECL and CL of the $Ru(bpy)_3^{2+}$–$S_2O_8^{2-}$ System in Acetonitrile–Water Solutions', *J. Am. Chem. Soc.*, **104**, 6891–6895 (1982).
16. A.W. Knight, G.M. Greenway, 'Relationship between Structural Attributes and Observed Electrogenerated Chemiluminescence (ECL) Activity of Tertiary Amines as Potential Analytes for the Tris(2,2′-Bipyridine)Ruthenium(II) ECL Reaction', *Analyst*, **121**, 101R–106R (1996).
17. W.A. Jackson, D.R. Bobbitt, 'Chemiluminescent Detection of Amino Acids Using In Situ Generated $Ru(bpy)_3^{3+}$', *Anal. Chem. Acta*, **285**, 309–320 (1994).
18. T.M. Downey, T.A. Nieman, 'Chemiluminescence Detection Using Regenerable Tris(2,2′-Bipyridyl)Ruthenium(II) Immobilized in Nafion', *Anal. Chem.*, **64**, 261–268 (1992).
19. F. Jameison, R.I. Sanchez, L. Dory, J.K. Leland, D. Yost, M.T. Martin, 'Electrochemiluminescence-based Quantitation of Classical Clinical Chemistry Analytes', *Anal. Chem.*, **68**, 1298–1302 (1996).
20. P. Liang, R.I. Sanches, M.T. Martin, 'Electrochemiluminescence-based Detection of β-Lactam Antibiotics and β-Lactamases', *Anal. Chem.*, **68**, 2426–2431 (1996).
21. P. Liang, L. Dong, M.T. Martin, 'Light Emission from Ruthenium-labeled Penicillins Signaling their Hydrolysis by β-Lactamase', *J. Am. Chem. Soc.*, **118**, 9198–9199 (1996).
22. D. Wild (ed.), *The Immunoassay Handbook*, Macmillan Press Ltd, 1994.
23. J. DiCesare, B. Grossman, E. Katz, E. Picozza, R. Ragusa, T. Woudenberg, 'A High-Sensitivity ECL Based Detection System for Automated PCR Product Quantitation', *Biotechniques*, **15**, 152–157 (1993).

Infrared Spectroelectrochemistry

Mark R. Anderson and C. Douglas Taylor
Virginia Polytechnic Institute and State University, Blacksburg, USA

1 Introduction **1**
1.1 Historical Background 1
2 Transmission Infrared Spectroelectrochemistry **4**
2.1 Optically Transparent Thin-layer Electrode Cell 4
2.2 Cell Characteristics 5
2.3 New Cell Designs 6
3 Internal Reflection Infrared Spectroelectrochemistry **7**
3.1 Evanescent Waves 9
3.2 Sampling in Internal Reflection Measurements 10
3.3 Dynamic Measurements 14
3.4 Other Sampling Configurations 15
4 Specular Reflection **15**
4.1 Spectroelectrochemical Cells for Specular Reflection Experiments 16
4.2 Potential Modulation 18
4.3 Polarization Effects 21
5 Summary **26**
Acknowledgments **26**
Abbreviations and Acronyms **26**
Related Articles **26**
References **26**

The combination of infrared (IR) spectroscopy and electrochemistry in a single measurement has been an area of active research interest since the late 1960s. IR spectroelectrochemical experiments provide insight about interfacial electrochemical phenomena and/or oxidation–reduction reaction mechanisms that cannot be obtained using either technique individually. The power of the measurement is in the complementary nature of the spectroscopic and electrochemical information available when obtained simultaneously. The combination of measurements, however, requires compromise between the IR spectroscopy and the electrochemistry because of their different sampling requirements. The need for compromise illustrates the unique strengths and weaknesses of the combined spectroelectrochemical measurement. Three spectral sampling methods are commonly used: transmission, internal reflection, and specular reflection sampling of the electrochemical experiment. Additional experimental perturbations, such as polarization and electrode potential modulation, are frequently applied to increase the spectral sensitivity toward changes confined to the electrode interfacial region. This article presents the instrumental requirements for, and a theoretical background of, different types of IR spectroelectrochemical measurements.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The combination of two or more unique measurements conducted simultaneously provides an opportunity to more completely characterize a chemical system without dramatically increasing the time of analysis. This concept has driven many instrumental developments in analytical chemistry over the last 30 years, and is a principal force behind the explosive growth of the field of spectroelectrochemistry. The combination of molecular structural/electronic information that is provided by spectroscopy with the analytical sensitivity and ability to control interfacial properties afforded by electrochemistry suggests the utility of spectroelectrochemical measurements. When combining these measurements, however, care must be taken to ensure that the efforts to conduct the simultaneous determinations do not dramatically compromise the quality of either the spectroscopic or the electrochemical data. This issue is particularly important to the field of IR spectroelectrochemistry because of the strict experimental requirements of both the IR and electrochemical measurements.

The article comprises four sections. Initially, background information about the field of spectroelectrochemistry is presented. This discussion provides a perspective on the issues of spectral sampling of the electrochemical experiment, and on the instrumental challenges faced by combined IR spectroscopy and electrochemical experiments. The focus then is on the three principal sampling methods commonly used in spectroelectrochemistry: transmission, internal reflection, and specular reflection. Here, discussion centers on the experimental requirements for acquiring spectral data during the course of an electrochemical experiment.

1.1 Historical Background

The field of spectroelectrochemistry has its origins in the mid-1960s.[1–4] The first publication that demonstrated the feasibility of combined spectroscopy and electrochemistry involved the spectral monitoring of the oxidation of *o*-toluidine using an optically transparent

tin oxide conducting glass electrode (CGE).[1] Use of the semiconducting SnO_2 compromised the electrochemistry slightly, but the transparency of the CGE to UV/VIS (ultraviolet/visible) radiation was crucial so that the incident visible radiation could pass through the working electrode and probe the electrolyte solution during the heterogeneous redox reaction. Although the spectral monitoring in this application was with visible radiation, this demonstration of spectroelectrochemical measurements illustrated the potential utility of combining electrochemistry with spectroscopic monitoring. This measurement also illustrated that instrumental compromises, such as the introduction of resistance by the use of the semiconducting electrode, are frequently required when combining these measurements.

Several other publications in which the method of spectral sampling of the electrogenerated/consumed material was altered followed this initial publication. Hansen et al. demonstrated the ability of the SnO_2-coated glass to serve as the electrode as well as an internal reflection element (IRE)/waveguide in a spectroelectrochemical measurement.[2] Here, due to the limited penetration depth of the evanescent field, they were able to restrict spectral monitoring to the region of solution adjacent to the electrode, effectively eliminating spectral response from the bulk of the solution. Murray et al. developed the optically transparent thin-layer electrode (OTTLE) by sandwiching a conducting grid, prepared from fine-mesh Au wires, between two optically flat glass or quartz microscope slides.[3] Unlike the transmission experiment by Kuwana et al.,[1] the thin-layer cell restricted the amount of solution monitored by the spectroscopy.

Although these initial efforts in the field of spectroelectrochemistry utilized the visible region of the spectrum, they address many of the fundamental issues central to all spectroelectrochemical measurements. The initial CGE configuration used by Kuwana indicated the need to compromise the requirements of the electrochemical measurement so that spectroscopy could be conducted simultaneously. The internal reflection measurements by Hansen addressed the issue of spectral selectivity for the electrode interfacial region. The OTTLE cell design by Murray considered minimizing the interference from a complex sample by eliminating most absorbances from species in the bulk solution that are not influenced by the applied potential. These issues, which are pivotal to the development of spectroelectrochemistry in general, are particularly important as the technique migrated to the IR region of the spectrum.

1.1.1 Infrared Spectroelectrochemistry

In 1966, Mark and Pons published the first application of spectroelectrochemistry to the IR region of the electromagnetic spectrum.[5] The experimental set-up for this measurement made use of a semiconducting Ge plate as both the working electrode and as an IRE for the IR spectroscopy. As with the UV/VIS measurements by Hansen, this IR spectroscopic measurement isolated the interfacial region by using the evanescent wave emanating from the semiconducting working electrode as the incident radiation. Minimizing the spectral sampling of the bulk solvent was a driving force to the internal reflection method. Unlike in the visible region of the spectrum, typical solvents used in electrochemistry experiments are strong absorbers of IR radiation and can easily dominate the IR spectrum. In the internal reflection experiment, the evanescent wave extends approximately one-tenth of a wavelength into the solution, effectively minimizing the amount of solvent that is probed by the IR radiation. As was the case in the application by Hansen et al., use of a semiconducting working electrode compromised the electrochemical response in favor of IR transparency. Despite the poor signal-to-noise, this result clearly demonstrated the possibility of extending spectroelectrochemical measurements to the IR region of the spectrum.

In 1968 Heineman, Burnett, and Murray designed an OTTLE cell for use in the IR region of the spectrum.[6] This cell used an Au-mesh minigrid electrode sandwiched between two IR transparent NaCl plates. This design parallels earlier OTTLE cells developed for use in the UV/VIS regions of the spectrum. As with the research by Mark and Pons, the OTTLE cell designed by Heineman et al. compromised the electrochemistry by increased resistance of the thin layer in order to accommodate the IR spectroscopic measurement. Of particular importance, both experimental configurations by Mark and by Heineman had low signal-to-noise, limiting the utility of these early IR spectroelectrochemistry measurements.

These initial demonstrations of IR spectroelectrochemistry illustrate the principal issues faced when conducting this combined measurement: (a) low spectral signal-to-noise for solutes, and (b) absorption by the background electrolyte. Several other papers appeared in the late 1960s and early 1970s that tried to address the experimental limitations with qualified success.[7,8] The potential of IR spectroelectrochemistry to obtain structural information about changes that accompany electrode reactions was intriguing and provided sufficient reason for the area to be revisited as instrumentation improved over the years.

1.1.2 Instrumental Developments of Importance to Infrared Spectroelectrochemistry

Two developments, independent of electrochemical measurements, in the mid-1960s and early 1970s made

significant contributions to the field of IR spectroelectrochemistry. In 1966, Greenler published the first of two papers that describe the IR spectroscopy of molecules adsorbed to metallic substrates using a specular reflection geometry.[9] Greenler showed that the absorbance of a thin film was nearly 5000 times greater at a glancing angle for IR light polarized parallel to the plane of reflection compared to the absorbance obtained with light reflecting from the surface with normal incidence. Further, he calculated that the absorbance from the glancing angle reflection experiment was 25 times larger than the absorbance for a transmission experiment with a free-standing film of similar dimensions. This result provided the theoretical basis for the infrared reflection–absorption spectroscopy (IRRAS) experiment that has become a routine measurement today. As was the case with the early efforts in the field of IR spectroelectrochemistry, the external reflection measurements of supported thin films showed promise but were ultimately limited by the poor spectral signal-to-noise obtainable at the time.[10]

Throughout the 1960s and 1970s, instruments for IR spectroscopy were traditional spectrometers which based wavelength selection on prism or grating dispersion and slits. These instruments suffered from low source throughput, slit width limited resolution, grating/prism dispersion, detector sensitivity, and slow detector-response time. Each of these instrumental properties contributed to the low signal-to-noise of the initial IR spectroelectrochemistry measurements and the initial specular reflection measurements. It was not until Fourier transform infrared (FTIR) spectroscopy and the associated accessories (e.g. detectors and electronics) became widely available that the field of IR spectroelectrochemistry became instrumentally feasible.

During the 1970s and early 1980s FTIR equipment began replacing instruments based on the dispersion of light for collecting IR spectra.[11] Instrumental advances that accompanied the development of FTIR addressed many of the signal-to-noise limitations found in the early IR spectroelectrochemistry measurements. The poor signal-to-noise of the initial spectroelectrochemical measurements may be attributed to the combination of low concentration of solutes that are being measured, and to the low intensity of the IR source that probes the sample at any given wavelength using the dispersive instrument. The issue of low signal-to-noise can be addressed by taking advantage of the high throughput of a Fourier transform instrument. Because of the design and use of an interferometer with a Fourier transform instrument, all frequencies are measured simultaneously. Compared to a dispersive instrument, therefore, the Fourier transform instrument may obtain the same resolution, but it does so in a fraction of the time. Consequently, multiple spectra may be obtained and co-added to increase the signal-to-noise in the same amount of time that it would take to monitor the wavelength region of interest with a dispersive instrument. This multiplex advantage leads to a substantial improvement in the signal-to-noise of the spectrum (the signal-to-noise increases by a factor of $\sqrt{N}$ where N is the number of co-added spectra). Similarly, use of the interferometer to encode the wavelength data, as opposed to separating the light by dispersion and slits, leads to greater source throughput. The dispersion of the source radiation and selection of individual wavelengths by slits is inefficient and lowers the total source intensity directed through the sample. The Fourier transform instrument directs most of the source intensity through the sample and separates the individual frequencies by mathematical transformation of the interferometer encoded source intensity. In this manner, less of the typically weak IR source is needlessly wasted. Each of these characteristics of the FTIR improves the signal-to-noise of the IR spectroscopy, providing the opportunity to acquire spectra of more challenging, less concentrated samples typically encountered in spectroelectrochemistry experiments.

An additional development in FTIR offers new opportunities for IR spectroelectrochemistry.[12–15] Step-scan FTIR instruments provide the same advantages as the continuous-scan instruments described previously; in addition, step-scan instruments introduce time as an accessible experimental parameter. In a step-scan instrument, the interferogram is built up point-by-point by moving the mirror of the interferometer in discrete steps. The timing of the experiment can be gated to the interferometer mirror steps, allowing acquisition of spectra as a function of time. With continuous-scan FTIR instruments, limited control of the time-frame is available through the interferometer scan speed and the overall instrumental resolution (e.g. the scan distance of the moving mirror). While individual scans of the interferometer are being co-added, the system being interrogated should be in some equilibrium condition. With the step-scan instrument, the time frame of the experiment is established by the dwell time at each step position of the interferometer. The dwell time may be set, providing an additional degree of flexibility for monitoring phenomena that occur at both short and long times. In this manner, one can easily monitor chemical systems at defined intervals after the interferometer mirror step, providing the opportunity to make measurements on dynamic systems whose processes occur over tens of microseconds. The utility of step-scan instruments for monitoring dynamic electrochemical processes is discussed later.

Although advances in FTIR instrumentation addressed some of the signal-to-noise limitations of IR spectroelectrochemistry measurements, these advances could not resolve all the fundamental issues faced when combining IR spectroscopy and electrochemical measurements:

- The general incompatibility of the solvent types required for these two measurements.
- The small quantity of material that must be detected during the experiment.

Most typical solvents for electrochemistry are poor solvents for IR spectroscopy (and vice versa). The polar solvents required to dissolve the supporting electrolyte generally have strong absorptions in the IR region of the spectrum such that the solvent would dominate the measured spectrum.

Two methods designed to reduce the solvent contribution to the IR spectrum have been developed:

- The use of thin solution layers to minimize the amount of solvent that the incident radiation samples.
- Modulating the interfacial properties and using phase-sensitive detection to selectively detect the interfacial region at the frequency of the modulation.

Advances in instrumentation over the years provide sufficient spectral sensitivity that IR spectroelectrochemical measurements have become common measurements despite the restrictions confronting them. In the following, the instrumental requirements for conducting IR spectroelectrochemical experiments are reviewed and discussed. The focus is on the strict conditions placed on sampling methods, and on how the experiment is used to obtain both electrochemical and spectroscopic data with good signal-to-noise. Where required, fundamental principles are addressed to illustrate why methods have certain restrictions, and why some experimental compromises must be made to accommodate the simultaneous spectroscopy and electrochemistry. The following sections are organized according to the main sampling geometries in common use today: transmission using the OTTLE cell, internal reflection, and external reflection. The discussion in each section focuses on the instrumentation requirements, the type of information available from the experiment, the limitations of the experiment, and the theory behind the measurements.

2 TRANSMISSION INFRARED SPECTROELECTROCHEMISTRY

Perhaps the most common applications of IR spectroelectrochemistry use the transmission mode for spectral sampling. In this case an OTTLE cell is used. Here the source radiation is directed through the transparent working electrode, sampling solution throughout the entire dimension of the thin layer. The relatively large surface area of the transparent minigrid electrode, combined with the small solution volume within the thin layer,

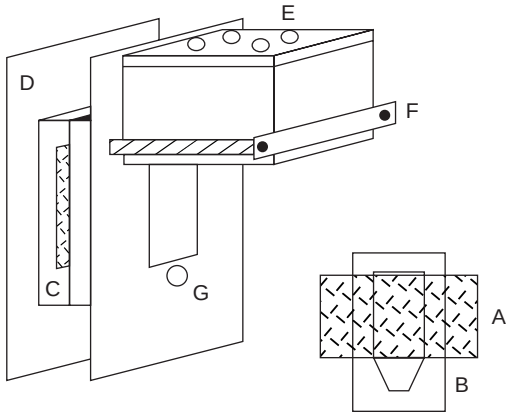

Figure 1 Prototype IR OTTLE spectroelectrochemical cell where: A is the gold minigrid electrode; B is a Teflon® spacer; C are NaCl plates; D is the steel frame; E are holes for reference and auxillary electrodes, and nitrogen inlet and outlet; F is a Teflon® cup and support bar; G is an outlet to draw solution from the cell. (Reprinted from W.R. Heineman, J.N. Burnett, R.W. Murray, 'Optically Transparent Thin-layer Electrodes: Ninhydrin Reduction in an Infrared Transparent Cell', *Anal. Chem.*, **40**(13), 1974–1978, copyright (1968) American Chemical Society.)

allow complete electrochemical conversion of the solute. This method is therefore a convenient way to spectroscopically probe the structure and identity of products and long-lived intermediates (e.g. those with lifetimes of the order of minutes) of electrode reactions. In addition, the short path length of the thin-layer cell helps to minimize the absorbance by the background electrolyte. Despite the amount of use that the OTTLE cell finds today, the basic design has not significantly changed from that published by Heineman in 1968 (Figure 1).[6]

2.1 Optically Transparent Thin-layer Electrode Cell

Using the concepts first illustrated with UV/VIS thin-layer cells, the OTTLE cell for use in the IR region of the electromagnetic spectrum sandwiches an Au minigrid working electrode between two NaCl plates. A Teflon® gasket of known thickness spaces the two IR-transparent NaCl plates from each other and defines the thickness of the thin layer. A hole near the top of the front NaCl plate provides a path for solution to enter a larger-volume reservoir. The secondary and reference electrodes are placed in this larger solution reservoir to complete the electrochemical circuit.

Heineman used what is now commonly called an optically thick thin-layer design by placing multiple Au minigrids parallel to and spaced from each other with Teflon® gaskets. Although not required, the optically thick thin-layer design helps in the spectral detection of weak IR absorptions by allowing a larger quantity of solute to experience the applied electrode potential.

The closely spaced minigrid electrodes allow efficient conversion of the electroactive material within the thin layer because the distance required for material to diffuse to the nearest electrode is diminished in this arrangement.

The principal advantage of the thin-layer design is the ability to quantitatively convert material within the thin layer between oxidation states. The depletion of electroactive species in the thin-layer cell avoids most bulk mass-transport effects. This occurs because the size of the thin layer is small compared to that of the diffusion layer thickness over the time-frame of the measurement (minutes). The acquired IR spectrum is therefore not complicated by the presence of unreacted analyte, nor is the response complicated by species diffusing into or out of the spectral observation window. The electrochemical behavior of thin-layer cells has been described previously.[16,17]

The open nature of the minigrid electrode allows IR source radiation to shine through the working electrode while maintaining electrochemical behavior that approximates a continuous conductor. Electrochemical behavior of the minigrid electrode can be understood using simple diffusion mass-transport arguments.[18] Consider the diffusion behavior of electroactive species at an isolated wire (Figure 2a). When a potential sufficient to oxidize or reduce electroactive material is applied, the species adjacent to the electrode are converted, establishing a concentration gradient in the adjacent solution. Under the influence of this gradient, electroactive reactant diffuses toward the electrode where it will then react, whereas the product of the electrode reaction will diffuse away from the electrode surface into the bulk solution. Consequently, the thickness of this concentration gradient expands in a radial direction away from the isolated wire with time after application of the potential.

Now consider two wires of the same dimension placed a small distance from each other. After application of the same potential to both wires the concentration gradient associated with the electrode reaction at each wire will similarly expand in time (Figure 2b). Ultimately, the diffusion layers from these adjacent wires will overlap. At that point, the diffusion approximates the behavior of a single electrode, not a pair of independent electrodes.

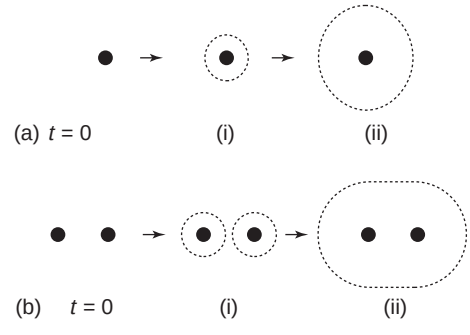

Figure 2 (a) Illustration of the diffusion layer generated at an isolated wire, cross-sectional view. (b) Model of diffusion at two isolated wires illustrating that, after some period of time, the individual diffusion layers overlap.

The minigrid electrodes commonly used in OTTLE cells have many closely spaced wires arranged parallel and perpendicular to each other. By analogy to the simple two-wire model given above, shortly after application of the potential to the minigrid, the diffusion layers associated with any given portion of the minigrid will overlap with that of its nearest neighbors. Consequently, the depletion of electroactive material due to oxidation or reduction at the working electrode will rapidly approximate linear diffusion to the macroscopic dimension of the minigrid. The electrochemical response, therefore, has the appearance of a reaction occurring at a planar electrode despite the open structure of the minigrid. Under these conditions, the electrochemical behavior of the thin-layer cell can be simplified to a one-dimensional model of linear diffusion over a finite distance to an infinitely large planar electrode.[18] Because the walls to the vessel are a finite distance from each other, depletion of the solute will extend to the dimensions of the thin layer, provided that the time-frame of the experiment is long compared to the time required for analytes to traverse the diffusion distance. Once this happens, the faradaic current due to the electrode reaction will return to the baseline level because all of the electroactive material within the thin layer has been converted into product. This thin-layer behavior has been described in detail previously.[16,17]

2.2 Cell Characteristics

An advantage of thin-layer electrochemical behavior is that experimental limitations due to mass transport are eliminated. This thin-layer configuration allows quantitative conversion between reactants, long-lived intermediates, and products for spectral monitoring. From an electrochemical point of view, this allows separation of the effects due to the heterogeneous reaction from the effects due to mass transport phenomena. Spectroscopic determination, therefore, provides the opportunity to acquire information regarding the mechanism of electrode reactions that is not available from the electrochemical data alone.

Although the response and planar electrode behavior of the OTTLE cell provide for a conceptually simple experiment, the thin-layer minigrid working electrode introduces some experimental complications. In electrochemical measurements, a three-electrode electrochemical cell is generally used to maintain accurate control of the working electrode potential. Accurate potential control, however, can be compromised by the OTTLE cell design due to the nonideal electrode arrangement.

Without accurate potential control, the experiment runs the risk of unevenly polarizing the working electrode (e.g. having an unequal potential distribution along the surface of the working electrode). Uneven working electrode polarization is generally avoided by maintaining the surface area of the secondary electrode large compared to that of the working electrode, and by the arrangement of the working, secondary, and reference electrodes. Typically, the surface area of the secondary electrode is 10–100 times larger than that of the working electrode. Similarly, the ideal arrangement of the working and secondary electrodes is opposite and parallel to each other with the reference electrode placed between these electrodes and close to the surface of the working electrode.[19] These conditions prevent the generation of current paths that differ significantly from each other between the secondary electrode and points along the surface of the working electrode. Under these conditions an equipotential distribution exists along the surface of the working electrode.

If these conditions are not met, the working electrode may have a distribution of potentials along the surface, introducing the risk that a distribution of phenomena will occur at the working electrode. This would complicate the electrochemistry and make spectral interpretation difficult. Clearly the gold minigrid electrode has a large surface area, which requires a secondary electrode of large area so as to avoid uneven electrode polarization. Unfortunately, the thin-layer cell prevents placement of the secondary electrode in the ideal geometry. It is therefore likely that the working electrode experiences some uneven polarization.

Placement of the reference electrode in the reservoir away from the working electrode may also have some deleterious effects. Under normal bulk-solution experimental conditions, the solution resistance that exists along the current path between the working and secondary electrodes may be partially eliminated by placing the reference electrode between these two electrodes and close to the working electrode. In an OTTLE thin-layer cell, it is difficult to place the reference electrode in this arrangement. Consequently, substantial solution resistance may exist within the thin layer. In addition, the total solution resistance experienced will vary along the surface of the working electrode, depending on the distance between points along the working electrode and the secondary electrode. In the thin-layer geometry, it is impossible to avoid different distances between points along the working electrode and the secondary electrode, leading to variable levels of resistance within the thin layer. This variable solution resistance leads to uneven electrode polarization and potential gradients along the working electrode within the thin layer. Murray et al. illustrated this when they were able to visually monitor a color change migrating along the length of the working electrode in an OTTLE cell following the application of potential.[3]

The presence of solution resistance and potential gradients principally influences the time response of the electrochemical measurement. Electrochemical experiments may be simply modeled as an equivalent circuit of resistors and capacitors. Because of the nature of the electrochemical experiments, the time response (e.g. how quickly the working electrode potential responds to the application of a new value) is an important parameter determined by values of the solution resistance and interfacial capacitance. Experiments in which the potential is rapidly altered in time need to have a fast response to accurately represent the electrochemical phenomena. With high resistance, OTTLE cells generally have a slow time response, resulting in distortion of the electrochemical data. For example, Heineman et al. obtained cyclic voltammetry data for the reduction of ninhydrin in their thick-layer OTTLE cell.[6] In this experiment there is a clear distortion of the current–potential response when measured in the thin layer compared to the current potential behavior in the bulk solution, suggestive of substantial resistance effects within the thin layer.

However, for most applications of transmission IR spectroelectrochemistry the slow time response of the OTTLE cell is not debilitating. Commonly, the working electrode potential is set to some value suitable for quantitatively oxidizing or reducing the analyte. When the current decays to the baseline level, indicating that quantitative transformation has occurred and equilibrium established, the sample is spectroscopically probed. In this manner, IR spectra of the reactants, products, or intermediates of the electrode reaction are obtained at their equilibrium concentrations determined by the electrode potential.

2.3 New Cell Designs

Although the basic design of the OTTLE cell for IR spectroelectrochemistry has not appreciably changed since the initial publication of Heineman, some alterations aimed at improving the time response, and at improving the compatibility of the cell with organic solvents, have appeared.

Hartl et al. prepared an OTTLE cell in which all three electrodes are placed within the solution thin layer. In this design a gold minigrid serves as the working electrode, a platinum mesh serves as the secondary electrode, and a silver wire serves as a pseudo-reference electrode (Figure 3).[20] These three electrodes are melt sealed into a 200-μm-thick polyethylene spacer to establish the small dimension of the thin-layer cell. In this arrangement, all three electrodes sample the same small portion of the solution. Although this

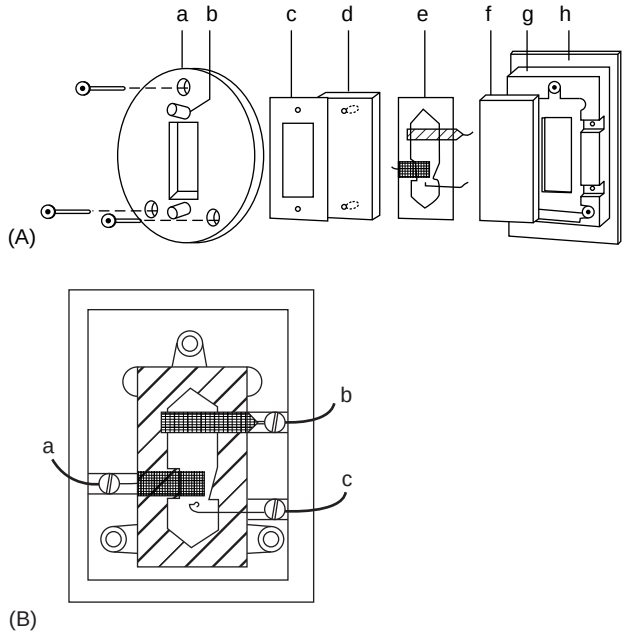

Figure 3 IR OTTLE cell in which the working, secondary and reference electrodes are all melt sealed into a polyethylene gasket. Part (A) shows: (a) steel pressure plate; (b) inlet; (c) Teflon® gasket; (d, f) front and back KBr windows; (e) polyethylene spacer with melt sealed electrodes; (g) Teflon® holder; (h) back plate. Part (B) shows: (a) Au minigrid working electrode; (b) Pt mesh auxillary electrodes; (c) Ag wire pseudo-reference electrode. (Reprinted from M. Krejcik, M. Danek, F. Hartl, 'Simple Construction of an Infrared Optically Transparent Thin-layer Electrochemical Cell: Applications to the Redox Reactions of Ferrocene, $Mn_2(CO)_{10}$ and $Mn(CO)_3$(3,5-di-*t*-butyl-catecholate)', *J. Electroanal. Chem.*, **317**, 179–187, copyright (1991) with permission from Elsevier Science.)

design is intended to minimize solution resistance and increase the slow current response, cyclic voltammetry of 0.018 M ferrocene (in CH_2Cl_2 containing 0.10 M tetra-*n*-butylammonium hexafluorophosphate (TBAF)) had substantial current–potential distortion.[20]

Placement of the working and secondary electrodes within the same volume of solution may introduce some problems. The electrode reaction that occurs at the working electrode has a corresponding but opposite reaction at the secondary electrode to maintain electroneutrality (e.g. if a reduction occurs at the working electrode, then an oxidation must occur at the secondary electrode). Typically, when the solution is completely electrolyzed, the working and the secondary electrodes are placed in separate compartments to prevent mixing of the products of the separate electrode reactions. Mixing of the products generated at these two electrodes can result in undesired homogeneous reactions occurring. Brisdon et al. suggest that mixing of the counter-electrode products in a thin-layer cell is negligible on the time-frame of the spectroelectrochemical measurement.[21] This is probably due to the relatively large distances that exist between the anode and cathode within the thin layer. However, repeated and prolonged use of this type of OTTLE cell can lead to the eventual build-up of residue at the edges of the working and counter electrodes. Permanently sealed cells therefore need to be discarded after several uses because of this residue build-up. Designs that rely on gaskets and pressure to seal the thin layer have also appeared.[21–23] These designs allow disassembly of the thin layer for cleaning and subsequent re-use.

The design described by Hartl allows maintenance of anaerobic conditions over the relatively long time-frame of the spectroelectrochemical experiment. The exclusion of oxygen from the electrolyte solution is particularly important when studying reductive electrochemistry. Oxygen is easily reduced and can interfere with the desired electrode reactions under these conditions. In most electrochemical measurements, the electrolyte solution is purged with an inert gas immediately prior to conducting the electrochemical reduction. However, with the OTTLE cell the solution within the thin layer cannot be easily purged. In this case, the analyte solution is purged prior to placement within the thin layer. Maintaining anaerobic conditions within the thin layer over the long time-frame of the transmission spectroelectrochemical measurement is required to avoid oxygen interference with the desired electrode reaction. As they demonstrate, this feature is particularly important for spectroscopically identifying oxygen-sensitive reaction intermediates frequently encountered in inorganic chemistry.[20]

The ease and overall utility of transmission spectroelectrochemical measurement makes it one of the more widely applied spectroelectrochemical measurements. Despite the many applications of this technique that have appeared in the literature, the basic sampling is largely the same as described by Heineman et al. in 1968.[6] The method is most useful for identifying bulk phase products and reactants of oxidation–reduction reactions. The thin layer of the OTTLE cell compromises the time response of electrochemical reactions. Consequently, transmission IR spectroelectrochemical measurements are best suited for identifying solution components after the solution has reached its equilibrium composition at a given applied potential. Recognizing the limitations allows one to study redox reaction mechanisms using transmission IR spectroelectrochemistry and the OTTLE cell.

3 INTERNAL REFLECTION INFRARED SPECTROELECTROCHEMISTRY

Although IR spectroelectrochemical applications using an OTTLE cell are numerous, the restrictions placed

on electrochemical measurement in this sampling configuration are significant. The resistance within the thin layer prevents rapid changes of the electrode potential; consequently, the observation of short-lived intermediates is virtually impossible. Mass transport into or out of the thin layer is also compromised by the geometry of the thin layer. This behavior is described as diffusionally decoupling the electrolyte which is trapped within the thin layer from the rest of the solution.[24] This may result in a number of deleterious effects including electrolyte concentration gradients, pH gradients, and slow diffusional replenishment of materials within the thin layer. A configuration of the spectroelectrochemical cell for improved electrochemical response would probably have to eliminate the thin layer.

The thin solution layer was required for these spectroelectrochemical measurements to minimize the absorbance contribution by the bulk electrolyte. Mark and Pons, in the first publication in the field of IR spectroelectrochemistry, showed in 1966 that absorption by the bulk electrolyte could also be reduced by sampling the interface with the evanescent wave that develops at the interface where IR light is internally reflected.[5] In this application, the evanescent wave emanates from a Ge plate configured as both the working electrode and as a waveguide for IR radiation (Figure 4). Although the spectroscopic signal-to-noise was poor, this measurement demonstrated the potential for using the evanescent wave for IR spectroelectrochemical measurements.

Hansen[25] and Harrick[26] demonstrated that light incident on an interface between a phase of high refractive index and a phase of lower refractive index reflects at the interface with a small portion of the incident light transmitted through the interface when the incident angle is greater than a critical value. The transmitted light, called the evanescent wave, decays exponentially into the lower-index medium with the depth of penetration being dependent on the wavelength of the incident light and on the angle of incidence. Typically, this decay length is approximated as one-tenth of the wavelength of the incident light. For IR spectroscopy, therefore, the spectral sampling by the evanescent wave will be of the order of hundreds of nanometers. The penetration of the evanescent wave is smaller than the typical diffusion layer thickness in electrochemical experiments, indicating that internal reflection spectroscopy will sample only the interfacial region during the IR spectroelectrochemical experiment. Because of the very small interfacial region probed by the internal reflection method, IR absorption by the bulk solvent is not significant and the need for a thin-layer configuration is eliminated. Eliminating the thin layer removes many of the electrochemical restrictions imposed on the experiment by the OTTLE cell (e.g. slow time response and resistance effects), and provides the opportunity to observe transient electrochemical phenomena on faster timescales.

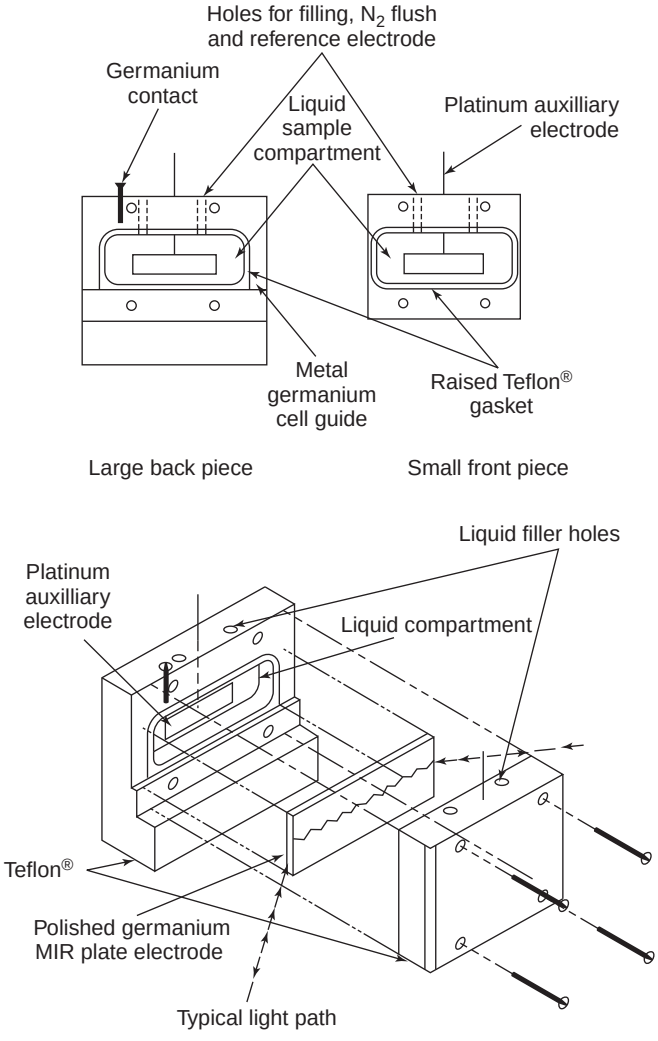

Figure 4 IR spectroelectrochemical cell with a Ge working electrode–waveguide combination for internal reflection spectroelectrochemistry. MIR = multiple internal reflections. (Reprinted by permission from H.B. Mark, B.S. Pons, 'An In Situ Spectrophotometric Method for Observing the Infrared Spectra of Species at the Electrode Surface during Electrolysis', *Anal. Chem.*, **38**(1), 119–121, copyright (1966) American Chemical Society.)

Unfortunately, the combination of the IRE and the working electrode into a single package introduces some new compromises to the combined spectroelectrochemical experiment. Mark and Pons,[5] as well as subsequent work by Tallant and Evans[7] and by Reed and Yeager,[8] used semiconducting *n*-doped Ge as both the working electrode and the IR-transparent IRE in

their spectroelectrochemical experiments. Because of the semiconducting nature of Ge, there was significant internal resistance across the Ge plates (100–2000 Ω as measured across a 5-cm dimension by Tallant and Evans[7]). Any resistance in the electrochemical measurement, whether it comes from the electrode or the solution, places restrictions on the time response of the measurement. Additional doping can decrease the resistance of the semiconducting Ge, but improvements in the electrochemical performance in this case are offset by the diminished IR transparency of the more highly doped Ge. Tallant and Evans suggested the use of metallic films deposited on the surface of the IR-transparent Ge as a way of increasing the conductivity of the working electrode.[7] In this case, the evanescent wave must penetrate through the metal film to sample the electrolyte solution. When using a metallic overlayer as the working electrode, the metal layer must be sufficiently thin to allow the evanescent wave to penetrate, because IR radiation is strongly absorbed by metallic films. To function as an electrode, however, the metallic film must be sufficiently thick to ensure a good, continuous conductor. As with the OTTLE cell, the different requirements of the spectroscopy and electrochemistry elements of the combined measurement require a compromise that ultimately may degrade the performance of both portions of the measurement.

It is evident that the strengths of the IRE spectroelectrochemical experiment are complementary to those of transmission IR spectroelectrochemistry with the OTTLE cell. Although the OTTLE cell provides a convenient method for acquiring spectra of reactants, intermediates, and products in the bulk solution, measurements that use IRE sampling isolate the interface and might be useful for identifying species at the electrode interface or within the electrochemical double layer during electrode reactions.

3.1 Evanescent Waves

A detailed discussion of the theory of internal reflection and evanescent wave properties is found elsewhere and is not reproduced here. Hansen[25,27] and Harrick[26] provide details for solving the boundary value problem for transmission and reflection of electromagnetic radiation at interfaces between media of different refractive indices using Fresnel's equations. Hansen in particular explored systems of relevance to spectroelectrochemical applications (e.g. two-layer and three-layer systems in which the middle layer is a thin metallic film).[25,27] He shows that the evanescent wave has a field strength that decays exponentially into the electrolyte medium. For a three-layer system (Figure 5), Equations (1–3) describe the field strength of the evanescent wave:

$$\langle E^2_{\parallel z}\rangle = \left|\frac{n_1 \sin\theta_1}{\hat{n}_3} t_{\parallel}\right|^2 \exp\left(\frac{4\pi}{\lambda} \operatorname{Im}\xi_3 (z-h)\right) \quad (1)$$

$$\langle E^2_{\parallel x}\rangle = \left|\frac{\xi_3}{\hat{n}_3} t_{\parallel}\right|^2 \exp\left(\frac{4\pi}{\lambda} \operatorname{Im}\xi_3 (z-h)\right) \quad (2)$$

$$\langle E^2_{\perp}\rangle = |t_{\perp}|^2 \exp\left(\frac{4\pi}{\lambda} \operatorname{Im}\xi_3 (z-h)\right) \quad (3)$$

In these equations, $\langle E^2_{\parallel z}\rangle$ and $\langle E^2_{\parallel x}\rangle$ are the mean electric field strengths in the z and x directions resulting from light polarized parallel to the plane of reflection, $\langle E^2_{\perp}\rangle$ is the mean electric field strength resulting from light polarized perpendicular to the plane of reflection, n_1 and $\hat{n}_3$ are the refractive index and complex refractive index of the incident and final phases, h is the thickness of the metallic layer, θ_1 is the incident angle, ξ_3 is the complex refractive index of the third phase multiplied by the cosine of the incident angle, $t_{\perp}$ and $t_{\parallel}$ are symbols representing Fresnel's equations for the transmitted light polarized perpendicular and parallel to the plane of reflection, and λ is the wavelength of the incident light. The Im function indicates that the imaginary part of the complex refractive index is taken, and z is defined in Figure 5.

Because the imaginary portion of ξ_3 is negative, the field strength decays exponentially into the lower index medium. It is estimated that the penetration depth, d_p, of the evanescent wave is equal to the distance away from the interface where the amplitude of the evanescent electric field strength decays to 1/e of its initial value. This is given by[28]

$$d_p = \frac{\lambda}{2\pi \left(\sin^2\theta - (n_2/n_1)^2\right)} \quad (4)$$

Equation (4), strictly applied, is for a two-layer system; however, it is frequently used to approximate the

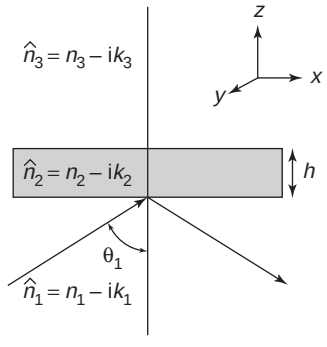

Figure 5 Diagram of the three-layer model of the interface between an IRE, a thin metallic layer and the electrolyte solution used to define the Cartesian axes of the boundary value problem for calculating the mean electric field strength of the evanescent wave.

penetration depth of three-layer systems where the assumption is made that the metallic film is sufficiently thin so as not to dramatically attenuate the evanescent wave. This is generally thought to be the case if the thickness of the layer is much smaller than the wavelength of the incident light. Equation (4) illustrates that the sampling depth of the evanescent wave in an internal reflection experiment depends explicitly on several factors:

- the refractive indices of the incident (e.g. the waveguide with high refractive index) and final layer (e.g. the electrolyte solution);
- the angle of incidence;
- the wavelength of the incident radiation.

Equation (4) illustrates that, over a broad range of angles, the sampling depth of the internal reflection experiment can probe different total dimensions of the interface. As the angle approaches the critical angle, the penetration depth increases, as expected. This behavior suggests that internal reflection methods may be used to spectroscopically probe interfacial electrolyte concentration gradients at the electrode interface by varying the angle of the incident radiation. Equation (4) also shows that over the mid-IR region of the electromagnetic spectrum, the sampling depth changes dramatically. For example, d_p is calculated to change from 600 nm at 1000 cm^{-1} to only 200 nm at 3000 cm^{-1}. This result suggests that quantitative spectroscopic analysis may be distorted if using spectral features having different wavelengths for analysis, a problem when using FTIR spectroscopy.

3.2 Sampling in Internal Reflection Measurements

Although not as widely applied as other IR spectroelectrochemical methods, several different sampling methods have been reported since the pioneering work of Mark and Pons.[5] The following examines these different methods.

3.2.1 Multiple Internal Reflections

The initial spectroelectrochemical experiments of Mark and Pons,[5] Tallant and Evans,[7] and Reed and Yeager[8] used a multiple reflection IRE for acquiring the IR spectrum. In this arrangement, the Ge waveguide is much longer than it is thick. As a consequence, the incident light reflects multiple times within the Ge working electrode as it propagates down the length of the waveguide. At each reflection of the incident light with the interface between the Ge plate and the electrolyte solution, attenuation of the evanescent wave due to absorbance by species on the electrolyte side of the interface is possible. Consequently, the multiple reflections can lead to greater spectroscopic signal-to-noise for species at the electrode interface.

Ashley et al. demonstrate the utility of multiple internal reflection Fourier transform infrared spectroscopy (MIRFTIRS) for studying material adsorbed onto a metallic interface.[29] Here they investigate the adsorption of SCN^- on silver and gold films by MIRFTIRS, and compare the results to other spectroelectrochemical experiments. Their spectroelectrochemical cell is similar in construction to Mark and Pons'; however, Ashley et al. used a 200-nm thin film of Au or Ag deposited on Si substrates as the working electrode. The sampling depth of the IR radiation in this cell was estimated by Ashley et al. to be approximately of 350 nm for 2000 cm^{-1} incident radiation. This estimate of the sampling depth apparently does not consider the attenuation of the incident radiation that may occur due to the presence of a relatively thick layer of metal between the Si IRE and the electrolyte solution. The thickness of the metal layer is approximately 5% of the wavelength of the incident light in this experiment; consequently, estimating the evanescent wave penetration depth with the expression for a two-layer system is not adequate for this system. The 200 nm metal thin film used by Ashley et al. is approximately 10 times larger than metal thin films used in other internal reflection experiments.[24,30–33] The presence of the thick metal film probably attenuates the penetration depth of the evanescent wave (considering the dependence on h, i.e. the metal film thickness, in the mean evanescent field strength given by Equations 1–3). This probably makes the actual sampling depth in this application somewhat less than this estimate of h. If the sampling depth is assumed to be much smaller than the initial estimate, the spectral results with the thick metal layer correlate with the observation that these measurements detect only SCN^- species adsorbed directly to the metal (e.g. those species within approximately 10 Å of the interface). In this case, the multiple internal reflections provide sufficient sampling of the interface to allow adequate signal-to-noise for observation of the adsorbed species. The evanescent wave attenuation by the thick metal film apparently prevents the observation of species in solution.

Sherson et al. also compared MIRFTIRS with other sampling techniques for IR spectroelectrochemistry.[34] Unlike Ashley et al., they prepared their metal film by patterning a thin metal section (thickness estimated to be approximately 4 nm) down the middle of the IRE surrounded by a thicker metal layer on the IRE periphery (Figure 6). A problem that has been observed with very thin metal layers on IREs is the formation of nonisotropic, nonideal interfacial structures (e.g. island formation) during the deposition step.[30,35] Island formation may result in electronic isolation of

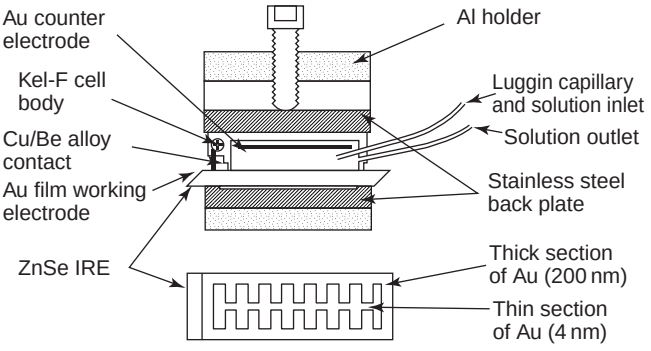

Figure 6 Patterned electrode and spectroelectrochemical cell used by Sherson in MIRFTIRS to study the electrochemical adsorption of SCN^-. (Reprinted by permission from I.T. Bae, M. Sandifer, Y.W. Lee, D.A. Tryk, C.N. Sukenik, D.A. Sherson, 'In Situ Fourier Transform Infrared Spectroscopy of Molecular Adsorbates at Electrode–Electrolyte Interfaces: A Comparison between Internal and External Reflection Modes', *Anal. Chem.*, **67**(24), 4508–4513, copyright (1995) American Chemical Society.)

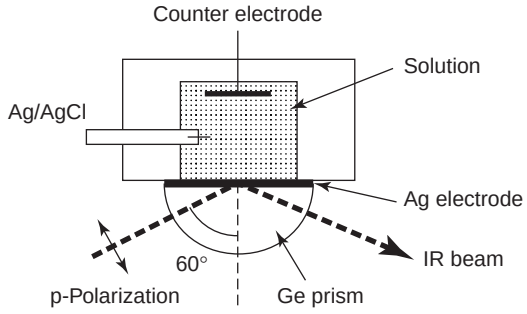

Figure 7 A Typical Kretschmann spectroelectrochemical cell. (Reprinted from M. Osawa, K. Ataka, K. Yoshii, T. Yotsuyanagi, 'Surface-enhanced Infrared ATR Spectroscopy for In Situ Studies of Electrode–Electrolyte Interfaces', *J. Electron Spectrosc. Relat. Phenom.*, **64/65**, 371–379, copyright (1993) with permission from Elsevier Science.)

portions of the thin metal layer, making electrochemical phenomena inhomogeneous across the surface of the metal. The intent of the patterned electrode was to maintain a thin metallic region for efficient evanescent wave penetration into the solution while also ensuring that the bulk metal film was sufficiently thick to maintain good conductivity for the electrochemistry. This cell demonstrates almost no electrochemical distortion due to resistance effects, and no obvious spectral irregularities from a nonisotropic surface. IR spectra with excellent signal-to-noise that are qualitatively equivalent to those obtained by IRRAS methods are obtained using this spectroelectrochemical cell.

Multiple internal reflection methods demonstrate excellent signal-to-noise characteristics. This is probably due to the many times that the interface is sampled as the incident light propagates down the waveguide. Ashley et al. showed that MIRFTIRS measurements can use thicker metallic films than others predict, probably due to the many times that the interface is sampled. Sherson demonstrated that patterned interfaces can provide continuous electrical contact across extremely thin metallic films. These results demonstrate the applicability of MIRFTIRS for studying interfacial electrochemical phenomena.

3.2.2 Single Reflection Interfaces

Single reflection interfaces have also been used as IREs for IR spectroelectrochemical measurements. A typical spectroelectrochemical cell for a single internal reflection experiment is shown in Figure 7.[36] In these applications, the IRE is a prism rather than a thin waveguide. Many different materials have been used for the prism (e.g. Ge, GaAs, Si, KRS-5, ZnSe, CdTe, BaF_2, CaF_2).[30] Dependent on the optical properties of the prism material and the adjacent solution, it is found that allowable dimensions of the metallic layer for adequate penetration of the evanescent wave through the metal film vary from 10 nm up to 40 nm.[30] This dependence of the allowable metal thickness on the optical properties of the prism is evident from Hansen's treatment of the three-layer system (Equations 1–3), yet it introduces complications when trying to compare results that are obtained with different IRE–metal overlayer combinations.

Johnson et al. attempted to rationalize behavioral differences observed based upon the prism–metal film combinations, particularly with regard to the inability of Hansen's treatment to make accurate quantitative predictions for some of these IRE–metal overlayer combinations.[30] They found that the optical properties of the metal films are different from those of the bulk metals, making predictions based on bulk metal properties invalid. They proposed that the origin of these property differences comes from two effects, namely the inhomogeneous nature of the metallic thin film and the interdiffusion of semiconducting metalloids and the metal overlayer. The three-layer model used for theoretical description of the internal reflection behavior at an IRE, as originally developed by Hansen,[25,27] assumes that all layers are isotropic. However, when depositing a thin metal layer for use as an electrode it is known that the metal film is not isotropic.[30,35] Instead, it forms initially as isolated islands that merge into a rough, continuous film as more material is deposited. Because of the inhomogeneities in the dimensions of the metal film, the isotropic model that is generally used does not accurately describe the system.

The role of interdiffusion is clearly demonstrated when comparing spectral results obtained with thin metal films deposited on an ionic reflection element

(such as BaF_2) and a semiconducting metal (such as Ge).[30,37] With Ge substrates, diffusion of the Ge into the metal overlayer may occur and change the optical properties of the metal overlayer. With the ionic crystal, interdiffusion of the substrate into the metal overlayer will not occur. When measurements were made with BaF_2 substrates coated with metallic thin films, close agreement between experimental results and Hansen's theory was obtained.[30,35] Agreement between experimental spectral data with the metal layer on a Ge substrate and theoretical predictions was not as close.

The single-reflection arrangement using a prism IRE with a thin metallic layer for IR spectral sampling is generally known as the Kretschmann arrangement, and has been widely applied for studying the structure of materials adsorbed on metallic films. Kretschmann IRE cells possess several characteristics that are potentially useful for spectroelectrochemical applications. The metallic thin film can be used as the working electrode in a spectroelectrochemical experiment, provided that it is thick enough to conduct. The film attenuates the evanescent wave penetration into the electrolyte, serving to further isolate the interface for spectral observation. As demonstrated in MIRFTIRS applications, this provides improved selectivity for adsorbed materials relative to bulk species. Finally, the extremely thin metallic films (generally 10–20 nm thick) typically used in this application are found to enhance the IR spectrum of species adjacent to the metallic layer.[31,32,38]

The enhancement of the IR spectrum of adsorbed material when using the Kretschmann arrangement is generally known as surface-enhanced infrared absorption spectroscopy (SEIRAS), and was first demonstrated by Hartstein et al. in 1980.[38] Like the related technique of surface-enhanced Raman spectroscopy (SERS), SEIRAS has been associated with the roughness of the thin metallic layer. However, unlike SERS, the phenomenon is observed using a variety of metallic substrates including Ag, Au, Cu, In, Pt, and Pd.[30,36]

Osawa et al. have studied applications and the theory of the SEIRAS technique.[35,36,39] They find that the enhancement is influenced by the optical properties of the IR-transparent substrate, the conditions used in depositing the metal film on the substrate, and the overall morphology of the film. Although each of these factors plays a role in the SEIRAS effect, the effects are not independent of each other. For example, when comparing BaF_2 and Ge IRE substrates ($n = 1.42$ and 4.0, respectively), Osawa found the SEIRAS intensity to be larger for the metal films deposited on BaF_2. This result can be partially understood from the optical properties of the materials and the field strength of the evanescent wave in a three-phase system (Equations 1–3). The closer the optical properties of the waveguide and the phase on the opposite side of the metal film, the larger will be the strength of the evanescent electric field. This result corresponds with those of Johnson.[37]

Different substrate–metal combinations generate different metal film morphologies when the film is vapor deposited. Wetting of the substrate surface during deposition refers to the ability of the deposited material to spread over the substrate surface and form an isotropic layer. The ability of a metal vapor to wet the substrate is related to the relative surface energies of the substrate and the deposition metal. In general, metal deposition initially occurs as isolated metal islands that eventually merge together as additional metal is deposited. The amount of metal required before the isolated islands merge into a continuous film depends on the efficiency with which the metal spreads across the substrate. Inefficient wetting of the substrate suggests that the metal will not deposit as a continuous film; rather, it will deposit and grow as discrete ellipsoids. These ellipsoids generate morphology/roughness of the metal film at the interface, a property that has been identified as an important contributor to the SEIRAS effect.[40,41] Different IRE materials have different surface energies; consequently, it is expected that there will be morphological differences between substrates resulting from the vapor deposition of thin metal films.

Osawa et al. report enhancement of the IR spectrum by a factor of 1000 when using a thin metal film made up of discontinuous islands of prolate ellipsoids.[39] In general, the deposition of metals on substrates begins by forming individual islands at very small coverage (5 nm films and smaller), and these islands merge into a continuous film as the thickness of the layer increases. The term ''thickness'' for a vapor-deposited metallic film is perhaps a misnomer because it is assumed that all the evaporated metal deposits on the substrate to form a uniformly thick film. The film thickness is calculated based upon the rate of thermal evaporation of the source metal and not on the actual thickness of the deposited film. When the individual metallic islands begin to merge, the amount of spectral enhancement decreases sharply (Figure 8).[40] This result is unfortunate for spectroelectrochemical applications because the electrochemistry needs to have a continuous metallic film to ensure uniform conductivity across the surface of the IRE. For spectroelectrochemical applications, therefore, the amount of spectral enhancement does not approach the factor of 1000 reported by Osawa et al. Tenfold sensitivity increases are observed in spectroelectrochemical applications where a continuous film is required (the sensitivity enhancement is calculated by comparison to spectra acquired by other IR methods).

Several groups have considered the mechanism of SEIRAS theoretically, with the most detailed descriptions given by Osawa et al.[35,36,39,40] and by

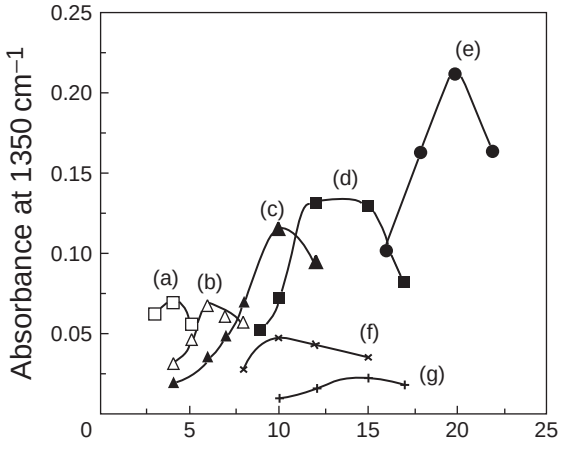

Figure 8 Plot of peak intensity of the 1350 cm^{-1} band as a function of Ag film thickness for different substrates. (a) Glow-discharge treated carbon-BaF_2; (b) untreated carbon-BaF_2; (c–g) bare BaF_2 prepared with different metal evaporation rates: (a–c, f) 0.10 nm s^{-1}, (e) 0.001 nm s^{-1}; and with different substrate temperatures: (a–e) 27 °C, (f, g) 100 °C. (Reprinted from Y. Nishikawa, T. Nagasawa, K. Fujiwara, M. Osawa, 'Silver Island Films for Surface-enhanced Infrared Absorption Spectroscopy: Effect of Island Morphology on the Absorption Enhancement', *Vib. Spectrosc.*, **6**, 43–53, copyright (1993) with permission from Elsevier Science.)

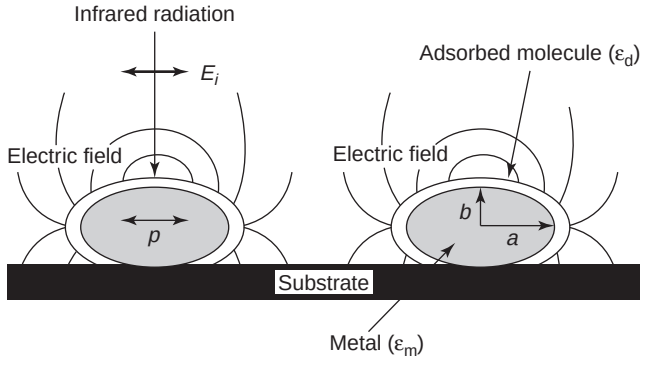

Figure 9 Diagram of the electromagnetic enhancement mechanism for SEIRAS experiments. (Reprinted from M. Osawa, K. Ataka, K. Yoshii, Y. Nishikawa, 'Surface-enhanced Infrared Spectroscopy: The Origin of the Absorption Enhancement and Band Selection Rule in the Infrared Spectra of Molecules Adsorbed on Fine Metal Particles', *Appl. Spectrosc.*, **47**(9), 1497–1502, copyright (1993) with permission from the Society for Applied Spectroscopy.)

Suetaka et al.[31,41,42] Osawa et al. have shown that the enhancement of the IR absorption is general for molecules near small metallic particles regardless of the spectral probe used (e.g. transmission, internal reflection, or specular reflection).[39] Regardless of the method of interrogating the sample, the general premise is that the oscillating field from the incident light induces an oscillating dipole in the metallic particle (illustrated in Figure 9).[39] The induced dipole of the particle generates an electric field with a principal direction perpendicular to the particle surface. Because of the small size of the particle (estimated to be approximately 10–20 nm along the major axis of the ellipsoid), the molecular vibrations of the adsorbed material modulate the polarizability of the metallic particle. Consequently, maximum field enhancement will occur when the system (incident radiation, particle dipole moment change, and vibrational motion of adsorbed material) is in resonance. Calculations by Osawa et al. using this model are in excellent agreement with experimental data.

Suetaka et al. investigated different contributions to the enhancement of the IR absorption.[31,35,41] They found that the thickness of the metal played a significant role in determining the type of enhancement mechanism operating. At low metal thickness (approximately 5 nm), both p- and s-polarized light (that is, light whose electric field is linearly polarized and aligned with the plane of reflection and perpendicular to the plane of reflection, respectively) produced an enhanced IR spectrum. This was attributed to an electromagnetic effect. At higher coverage (200 nm), only the p-polarized light was found to enhance the IR absorption. This observation is important to spectroelectrochemical applications because these generally use thick metal film electrodes. The p-polarized light introduces a directional aspect to the IR experiment that may be used to obtain information about the structure of the molecule adsorbed to the interface. Originally, the thick film behavior was attributed to an excitation of delocalized surface plasmons.[41] Osawa et al. showed later that the surface plasmons could not explain the broad range of incident angles that lead to the observed enhancement.[39] They assert that the polarization dependence of the enhancement with the thick films is likely a function of the continuous nature of the metal layer. That is, the p-polarized light can more efficiently induce polarization changes of the metal film in a direction normal to the macroscopic dimension of the film than can the s-polarized light.

For spectroelectrochemical applications, the field enhancement at the metal–particle interface has several benefits. The large magnitude of the electric field at the interface increases the absorbance by material adjacent to the metallic thin film. This leads to a spectral response that is approximately 10 times larger than that obtained by other IR spectroelectrochemical techniques. Hatta et al. showed that the SEIRAS enhancement was dependent upon the incident polarization as well as the thickness of the metallic layer.[41] For thicker metal layers (approximately 20 nm) where the islands have merged into a continuous film, only incident light that is p-polarized generated the enhanced field at the interface. Because the metallic film is continuous in electrochemical applications, and because the incident radiation probes

the interface by internal reflection, the field enhancement on the solution side of the interface has a principal direction largely normal to the macroscopic dimension of the IRE interface. The directional aspect of the induced electric field results in selective differentiation of the vibrational modes for molecules adsorbed on the surface. Vibrational modes whose transition dipole moment change is aligned with the enhanced field are observed, whereas those modes whose transition dipole moment change is perpendicular to the field enhancement are not observed. This behavior is generally known as the surface selection rule and provides information about the orientation that adsorbed molecules make with respect to the interface. Inspection of the equations for the evanescent wave field strength given by Hansen (Equations 1–3) shows that the evanescent wave has polarization components both perpendicular and parallel to the plane of reflection. For the internal reflection experiment, therefore, the surface selection rules must arise from the field enhancement due to the SEIRAS effect.

Taking advantage of this polarization dependence of the field enhancement, Hatta et al. demonstrated a polarization modulation method used in conjunction with the Kretschmann configuration.[43] In this method, alternating p- and s-polarized light is reflected from the prism interface. The reflected light is detected and demodulated with a phase-sensitive lock-in amplifier at the frequency of the polarization modulation. Because only the p-polarized light efficiently couples into the continuous metallic film to enhance the IR spectrum of adsorbed material, modulating the incident radiation between p- and s-polarization has the effect of introducing time dependence to the spectral signal. Consequently, lock-in detection improves the signal-to-noise ratio for the IR spectrum obtained in this fashion. Because this polarization effect is present only with continuous metallic films, the polarization modulation technique can easily be applied to electrochemical systems.

The polarization modulation used with the Kretschmann arrangement illustrates a method where modulating the analytical signal improves the signal-to-noise of the experiment. In this application, the polarization modulation relies on the properties of the interface to differentiate the effects measured.

3.3 Dynamic Measurements

Osawa et al. recently demonstrated how SEIRAS measurements using the Kretschmann arrangement could be used to characterize dynamic (e.g. short time-frame) electrochemical phenomena.[44] Typical IR spectroelectrochemical measurements yield spectra under equilibrium conditions; that is, conditions at some time after the potential alteration has occurred and the system current has returned to baseline levels. In making spectral measurements in this fashion, the dynamics of the electrode process are lost. This restriction is largely due to the incompatible time-frames of the electrochemical phenomena (milliseconds) and the IR spectral acquisition time when using FTIR (seconds to minutes).

The advent of step-scan FTIR instrumentation makes dynamic FTIR measurements possible. With this instrument, the moving mirror of the interferometer is not continuously scanned; rather, it is sequentially stepped and held at discrete positions along the scan dimension. In scanning interferometry, each scan of the moving mirror represents the acquisition of one spectrum, and the signal-to-noise of the final spectrum is improved by co-adding many individual spectra. Depending on the scan distance and the mirror velocity, acquiring one spectrum requires some defined time. In a step-scan instrument, the signal is integrated not as a function of the number of scans, but as a function of time at each step position. With this instrument, therefore, the interferogram is built up point-by-point with the signal-to-noise of the final spectrum being related to the dwell time at each mirror position. If used effectively, there is the opportunity to gate an experiment to the individual steps of the interferometer so that time-resolved spectral data might be acquired. In this manner, one may obtain IR spectra at a defined duration after application of some experimental perturbation. For example, if the interferometer mirror position step were synchronized with a change in the applied potential at the working electrode, then it would be possible to acquire IR spectra at some defined time after the potential change. Alternatively, one could use the interferometer mirror steps to initiate a linear potential scan of the working electrode. In this case, one could acquire spectra of the interface at discrete potentials (because time and potential are linearly related in a linear potential scan experiment) during the potential scan.

To conduct these synchronized experiments, the system being studied must exhibit reversible behavior with respect to the perturbation applied. This requirement is easily understood because the experiment must be repeated many times at each mirror position so that the corresponding interferogram can be built up point-by-point. Consider that each mirror position is providing only a portion of the total interferogram. Only after each portion of the interferogram (e.g. at each mirror position) is collected and assembled can the final spectrum be calculated. If the system being studied is not reversible with respect to the experimental perturbation, then the chemical system is different at each mirror position. To obtain a sensible spectrum, then the condition of the sample must be the same at every interval following the mirror position change, requiring that the associated chemistry be reversible and reproducible.

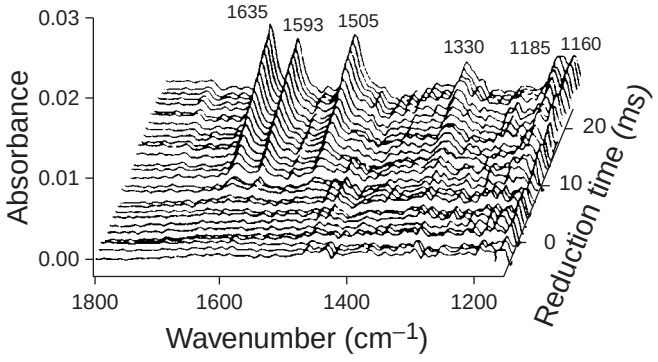

Figure 10 IR spectra acquired using a step-scan FTIR at 100-μs intervals after the application of a potential to the working electrode in a Kretschmann spectroelectrochemical cell. (Reprinted with permission from M. Osawa, K. Yoshii, K. Ataka, T. Yotsuyanagi, 'Real-time Monitoring of Electrochemical Dynamics by Submillisecond Time-resolved Surface-enhanced Infrared Attenuated-total-reflection Spectroscopy', *Langmuir*, **10**(3), 640–642, copyright (1994) American Chemical Society.)

Osawa et al. have demonstrated the utility of step-scan coupled with the Kretschmann arrangement for monitoring electrochemical processes at short times.[44] In this application, they monitor the deposition of heptaviologen radical cation as a function of time after the electrode is stepped from a value where no electrochemistry occurs (−0.2 V) to a value where the mass-transport-limited one-electron reduction of heptaviologen occurs (−0.55 V). After the potential step, spectral data are acquired at discrete times (every 100 μs) at each mirror position. From this spectral data, individual IR spectra of the interface corresponding to every 100 μs after the potential step are constructed (Figure 10). This provides a dynamic spectral picture of the interfacial electrochemical phenomena with time resolution acceptable for following dynamic electrode processes. The combination of step-scan FTIR and the rapid response, low resistance spectroelectrochemical cell used in the internal reflection sampling mode provide the instrumental characteristics necessary for this experiment.

3.4 Other Sampling Configurations

Internal reflection IR spectroelectrochemical measurements in which the working electrode is a thin metal film deposited on the IRE are shown to have excellent sensitivity for material adsorbed to the metal thin film. However, the technique is poor at monitoring species present only in the solution phase. Sherson et al. address this shortcoming by using a channel flow electrochemical cell (Figure 11).[34,45] With this cell, the electrochemical process and the spectral monitoring are separated in

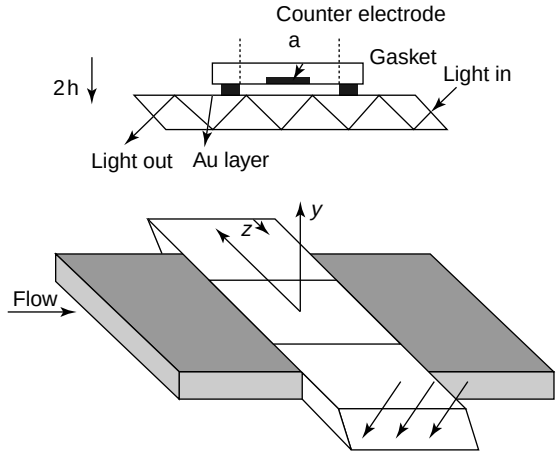

Figure 11 Schematic diagram of a channel-type spectroelectrochemical cell for ATR/FTIR measurements of solution phase electrogenerated species. (2h) is the full height of the cell and (a) is the width of the flow channel. (Reprinted with permission from R. Barbour, Z. Wang, I.T. Bae, Y.V. Tolmachev, D.A. Sherson, 'Channel Flow Cell for Attenuated Total Reflection Fourier Transform Infrared Spectroelectrochemistry', *Anal. Chem.*, **67**(21), 4024–4027, copyright (1995) American Chemical Society.)

space. Using a flowing solution through a thin channel, electroactive material undergoes an electrode reaction at the working electrode and then it is swept into the spectroscopic observation region by the flowing solution. The electrochemical response to the flowing stream follows standard steady-state electrochemical behavior, with the flowing stream maintaining the concentration of the electroactive material constant at the diffusion layer boundary.[16] Because the evanescent wave penetrates only a small distance into the solution phase, only species within the diffusion layer are observed spectroscopically. The concentration determined by the spectroscopic measurement is a function of only the down stream distance from the near edge of the working electrode. This method holds great promise for extending the internal reflection sampling geometry for studying solution phase reactants and products of electrode reactions.

4 SPECULAR REFLECTION

The strengths and weaknesses of transmission spectroelectrochemistry with an OTTLE cell and internal reflection spectroelectrochemical measurements are somewhat complementary to each other. The OTTLE cell allows spectral observation of solution phase products and intermediates of electrode reactions. However, this method does not provide any information about material at the solution–electrode interface. Because of the shallow

depth of penetration, internal reflection methods are most sensitive to the interface and provide almost no information about material in the bulk solution. Resistance effects limit transmission measurements to static, equilibrium conditions whereas, as demonstrated by Osawa et al. with step-scan FTIR,[44] sampling by internal reflection allows acquisition of spectra during dynamic electrochemical phenomena.

Ideally, a single IR spectroscopic sampling method would be able to provide information about both solution phase and adsorbed materials, as well as information about dynamic electrode processes. One sampling method that may be able to provide this information uses a specular, or external, reflection sampling geometry. In this geometry for spectral sampling, the IR source passes through a small thickness of the bulk solution before and after reflection, sampling the solution phase species. According to Greenler, the phase behavior of reflected light provides an enhanced sensitivity for species at the metallic surface.[9,10] The specular reflection sampling method therefore appears to be able to provide information about both solution phase and surface adsorbed material.

4.1 Spectroelectrochemical Cells for Specular Reflection Experiments

Although specular reflection has the potential to provide more spectral information than either transmission or internal reflection methods, it also has limitations. Primary among these limitations is the required use of a solution thin layer. The thin layer, as in transmission experiments, is required to improve the total throughput of the source radiation through the highly absorbing solvent. As with the OTTLE cell, use of the thin layer creates resistance effects and possible uneven polarization of the working electrode interface, limiting the opportunity for dynamic measurements. Designs of the specular reflection IR spectroelectrochemical cells provide additional flexibility, compared to OTTLE cells, which help to reduce resistance effects and improve the time response of the measurements.

The cell design used by Bewick et al. in the first demonstration of specular reflection IR spectroelectrochemistry is shown in Figure 12.[46] Here the working electrode was a Pt disk mounted on a syringe plunger. The cell body was fashioned from the syringe barrel with the front sealed with an IR-transparent window. When assembled, the working electrode is inserted through the syringe barrel until it is pressed against the IR-transparent window, forming the solution thin layer with a thickness estimated to be approximately 10–20 μm. To complete the electrochemical circuit, the secondary electrode is a Pt coil placed close to the IR-transparent window and surrounding the working electrode. The reference electrode tip is also placed close to the edge of the working electrode using a Luggin capillary. Although the electrochemistry is conducted on a thin layer of solution, in this specular reflection cell design the placement of the secondary and reference electrodes is more ideal than in the OTTLE cell, helping to reduce some of the resistance effects.

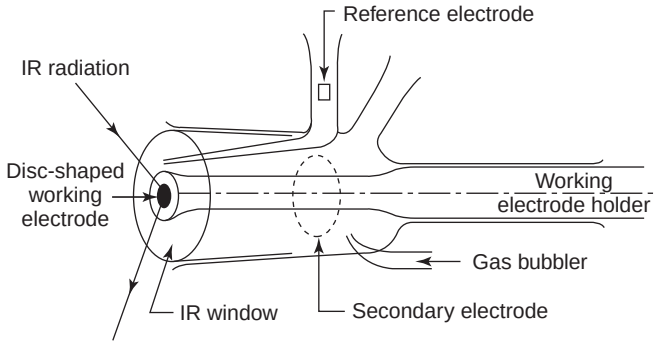

Figure 12 Diagram of the spectroelectrochemical cell used for specular reflection experiments. (Reprinted from A. Bewick, K. Kunimatsu, B.S. Pons, J.W. Russell, 'Electrochemically Modulated Infrared Spectroscopy (EMIRS): Experimental Details', *J. Electroanal. Chem.*, **160**, 47–61, copyright (1984) with permission from Elsevier Science.)

The cell design by Bewick et al. is similar to the full-edge current flow thin-layer spectroelectrochemical cell described by Lin and Kadish.[47] Although Lin and Kadish's cell is designed for use in the UV/VIS region of the spectrum, the placement of the secondary electrode surrounding the working electrode and just outside of the thin layer is like the design of Bewick et al's. Similarly, placement of the reference electrode tip just outside the thin layer is similar to the cell of Bewick et al. Lin and Kadish's thin-layer cell has excellent current–voltage response, suggestive of reduced resistance effects (compared to that of a traditional OTTLE cell). Indeed, equivalent circuit analysis suggests that the more ideal arrangement of the electrodes, particularly the reference electrode, make significant contributions toward reducing the resistance effects.[47]

In addition to possible thin-layer solution resistance effects, a problem introduced by this design that should be addressed is the need to carefully align the IR-transparent window with the working electrode. Unlike the OTTLE cell, where the boundaries of the thin layer are held apart and parallel to each other by a Teflon® gasket of known thickness, the specular reflection cell does not define the thin-layer dimension. Consequently, the thin-layer size is difficult to reproduce experiment to experiment. Defining the thin-layer dimensions with partial gaskets at the edge of the working electrode is possible; however, this may increase the resistance of the thin layer by impeding current paths. The IR-transparent window and the surface of the electrode must also be parallel to each

other. If they are not parallel, then the thickness of the thin layer will vary along the surface of the working electrode. The irreproducibility of the solution thin layer may result in variation in the spectral signal-to-noise from one experiment to another.

Although the basic design principles used by Bewick et al. in constructing their specular reflection cell have not changed, many alternative designs have appeared. These designs address the sampling issues identified above.

Mosier-Boss et al. describe a cell, similar to that of Bewick et al., that is machined from a glass ceramic material.[48] The arrangement of the working, secondary, and reference electrodes is similar to the glass cell of Bewick et al. By machining the cell body from a ceramic, higher tolerances for the electrode and window alignments are possible. These higher construction tolerances improve the reproducibility for generating the solution thin layer, for ensuring that the IR-transparent window and the working electrode are parallel to each other, and for decreasing the dimension of the thin layer. Using this cell with a 10-μm thin-layer thickness, Mosier-Boss et al. were able to obtain current versus potential responses in cyclic voltammetry experiments that are largely undistorted by resistance effects.[48] This particular current versus potential response is characteristic of semi-infinite diffusion, not thin-layer electrochemistry. At these potential scan rates, the 10-μm dimension of the thin layer is much larger than the diffusion layer thickness. This cell can respond quickly to changing potentials (e.g. potential scans of $100\,\mathrm{mV\,s^{-1}}$) and is not subject to high thin-layer resistance effects; consequently, it is not limited to static, equilibrium electrochemical measurements and could be used in dynamic electrochemical measurements. This result illustrates the effectiveness of using more ideal electrode placements on reducing resistance effects within the thin layer.

Reproducing the thin-layer thickness is a difficult, yet important, parameter to the specular reflection spectroelectrochemistry experiments. Frequently, one needs to collect spectral data, pull the electrode away from the window to allow replenishment of the thin layer, and then re-establish the thin layer for additional spectroscopic measurements. If the thin-layer thickness is not reproduced, spectra acquired before and after the thin-layer disruption cannot be reliably compared with each other. Several spectroelectrochemical cells have been designed in which the working electrode is mounted on a micrometer.[49–51] The micrometer allows exact and reproducible placement of the working electrode surface based on the micrometer setting.

Zhang and Lin sought to reduce even further the resistance within the thin layer by using an electrode microarray as their working electrode.[52] Here, rather than having a continuous disk, they use a 3×3 array

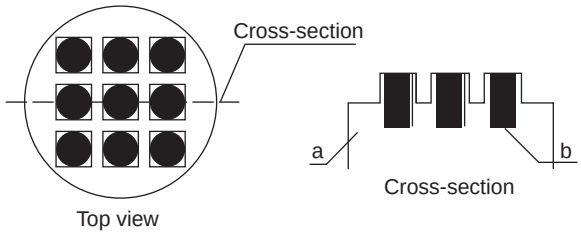

Figure 13 Diagram of the 3×3 microarray working electrode; (a) is a Teflon® rod and (b) are the embedded 1-mm Pt wires. (Reprinted from H.Q. Zhang, X.Q. Lin, 'In Situ FTIR Spectroelectrochemistry of a Microarray Electrode', *J. Electroanal. Chem.*, **434**, 55–59, copyright (1997) with permission from Elsevier Science.)

of 1-mm diameter disks (Figure 13) as their working electrode. In this design, nine closely spaced, full-edge current-flow thin-layer cells[47] with dimensions much smaller than that of a single disk electrode with an equivalent surface area act as a single working electrode. Using equivalent circuit analysis, they show that the resistance of this microarray is one-ninth the resistance of a single electrode with an equivalent surface area. Likewise, the time-constant of the thin-layer cell using the microarray working electrode will improve by a factor of nine compared to the single disk. These results suggest that a fast electrochemical response would be expected when using this microarray in a thin-layer cell. Indeed, when cyclic voltammetry experiments are conducted with the microarray in the thin-layer arrangement, negligible distortion of the current versus potential curve is observed. Importantly, peak current heights obtained with the microarray electrode in the thin-layer arrangement are the same as that obtained under bulk solution conditions. This result indicates that, under the thin-layer condition, the entire surface area of the microarray electrode is available for electrochemical reaction and is participating in the reaction. If the electrode were unevenly polarized, one would expect that only a portion of the electrode surface area would be involved in the electrochemical reaction, and the observed peak currents would be reduced or the current–voltage behavior distorted compared to the bulk electrochemical behavior. This electrochemical result is important because the reflection IR measurement samples the entire surface area of the electrode. To ensure that the spectroscopy results properly represent the electrochemical phenomena, the electrochemical process must be homogeneous across the entire spectral sampling area (e.g. generally the entire electrode surface area in a specular reflection experiment).

Li and Lin describe a reflection IR spectroelectrochemical cell that uses a 10-μm Pt disk as the working electrode.[53] The advantage of using a microelectrode is the small influence that resistance effects have on

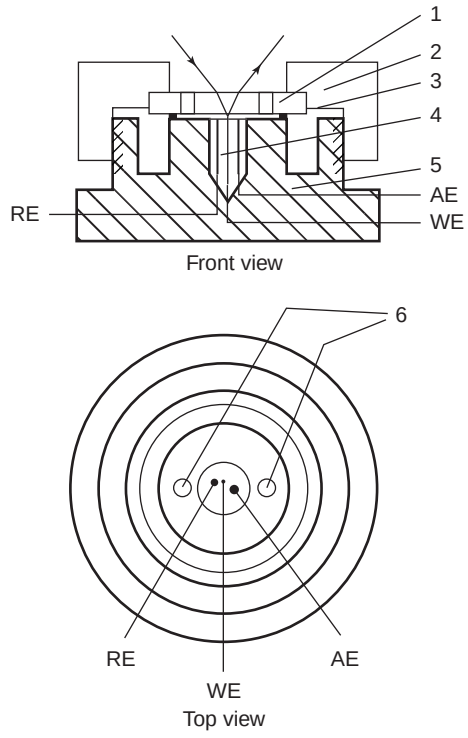

Figure 14 Diagram of the thin-layer microelectrode reflection cell where (1) is the IR optical window, (2) is the Teflon® cell top, (3) is a ring spacer, (4) is an epoxy adhesive, (5) is the Teflon® cell body, (6) are the inlet and outlet, WE is the working electrode, RE is the reference electrode and AE is the auxiliary electrode. (Reprinted from Z. Li, X. Lin, 'Electrochemistry and In Situ IR Microspectroelectrochemistry in a Versatile Thin Layer Reflection Fourier Transform IR Microspectroelectrochemical Cell', *J. Electroanal. Chem.*, **386**, 83–87, copyright (1995) with permission from Elsevier Science.)

microelectrode applications. The small surface area of the working electrode generates very small faradaic currents (nanoamperes to picoamperes). With small currents, even a large thin-layer resistance will have only small influence on potential polarization of the working electrode; consequently, microelectrodes are frequently used in high-resistance applications. This microelectrode cell, therefore, should demonstrate little uneven polarization of the working electrode and have a fast time response. As with the OTTLE thin-layer cells designed by Hartl et al.,[20,54] this microspectroelectrochemical cell places all three electrodes within the solution thin layer (Figure 14). In this arrangement, electrochemistry and spectroscopy measurements could be made on less than 20 μL of solution, making this a useful sampling configuration when the sample size is limited. Because of the small dimension of the working electrode, this cell required the tight focus of the IR beam offered by an IR microscope to conduct the spectroelectrochemical experiment.

Although each of these specular reflection cell designs requires the use of a thin layer of solution, all have demonstrated resistance characteristics that are much improved compared to that of a typical OTTLE thin-layer cell. This result suggests that the time response for specular reflection experiments will be sufficient for measurements of dynamic electrochemical phenomena. A rapid electrochemical response is important to specular reflection experiments because electrode potential modulation is one technique that is frequently utilized to improve the signal-to-noise ratio of the spectroelectrochemical measurement.

4.2 Potential Modulation

In specular reflection IR spectroelectrochemistry, one is trying to detect chemical species that are trapped within a solution thin layer, or are present at the electrode–electrolyte interface. Generally this is a very small amount of material, particularly in comparison with the amount of solvent and supporting electrolyte also present in the thin layer. Bewick demonstrated that the signal-to-noise for electrochemically active solutes could be substantially improved by introducing a time variation to the IR spectroelectrochemical experiment.[55] In this first application of specular reflection IR spectroelectrochemistry, a 10-Hz potential modulation was applied to the working electrode. The potential modulation created a continuous oscillation of the identity of the chemical species present within the solution thin layer that was monitored spectroscopically. In many respects, this procedure is analogous to using an optical chopper to introduce a time dependence on the electromagnetic radiation as it arrives at the detector. With the potential modulation, the radiation is being turned on and off at a defined frequency because the interaction between the sample (e.g. the solution thin layer) and the light changes in phase with the potential modulation. The output of the IR detector is directed to a lock-in amplifier that demodulates the detector signal at the frequency of the potential modulation. Consequently, the final output from the lock-in detection is representative of the changes in the IR spectrum of the interface that are in phase with the electrochemical perturbation.

With the potential modulation, the phase-sensitive detection will detect only spectral changes that occur at the frequency of the potential changes. This aspect of the measurement has several important implications to spectroelectrochemical measurements made in this manner. Because detection is at the frequency of the potential modulation, the acquired spectra result from the interfacial reflectance differences at the two potentials of the modulation. The magnitude of the reflectance

spectrum is normally given by Equation (5):

$$\frac{\Delta R}{R} = \frac{R_2(\nu) - R_1(\nu)}{R_1(\nu)} \tag{5}$$

where $R_1(\nu)$ is the interfacial reflectance at applied potential E_1 and $R_2(\nu)$ is the reflectance at potential E_2. The reflectance spectrum is related to the absorbance by Equation (6):

$$A = -\log\left(\frac{\Delta R}{R} - 1\right. \tag{6}$$

As a difference spectrum, the appearance is not the same as an absolute spectrum of the thin layer at any given potential. Only those vibrational features that undergo some change caused by the changing value of the applied potential have the same time variance as the potential modulation; therefore, they are the only features that are detected in the IR spectrum. As a result, the baseline of the measurement is at $\Delta R/R = 0$. This reflectance baseline suggests that, for most of the spectral region, there are no changes in the sample brought about by the potential alteration. Considering the spectral contributions from the solvent over most of the available potential window for a given solvent–electrolyte combination, the vibrational features of the solvent will be invariant with the value of the applied potential. Consequently, the solvent contributions to the IR spectrum are effectively cancelled out by measuring the difference spectrum.

A second implication of the potential modulation spectrum is that the final reflectance spectrum will have features with both positive and negative values. The positive and negative features of the difference spectrum arise from changes in the thin-layer reflectances brought about by the two potentials used in the modulation. To illustrate this concept, consider a potential modulation about the standard potential for the oxidation/reduction of some solute. At one potential (E_1), the oxidized form is favored within the thin layer. At the other potential (E_2), the reduced form is favored. From the $\Delta R/R$ expression (Equation 5), it can be seen that the difference spectrum will have negative-pointing features corresponding to vibrations characteristic of the oxidized form and positive features characteristic of the reduced form. Correlation of the observed spectral features with the potential allows diagnosis of mechanistic aspects of the electrochemical response of the system under investigation. Unlike transmission spectroelectrochemical measurements with the OTTLE cell, the potential changes do not have to convert all of the material within the thin layer. The potential modulation need only create enough of a time-dependent difference in the vibrational modes of species within the thin layer to be observed by the IR spectroscopy. How much of a change is enough depends on the absorptivities of the individual species.

Only those portions of the molecule whose vibrational features are altered by the potential changes are observed. For complex molecules, therefore, the potential modulation method offers the opportunity to identify the site of oxidation/reduction within the molecule (e.g. by observing which vibrational features are most influenced by the potential changes). Although redox conversion of electroactive solutes represents an obvious application of the potential modulation, other more subtle changes brought about by the potential changes may also be observed. Other example applications of processes associated with a changing applied potential include: studies of potential-induced adsorption or desorption of molecules;[56] orientational changes of molecules adsorbed to the electrode interface;[57,58] changes in the type or strength of adsorption interactions at the interface;[59] changes brought about by the changing electric field strength at the interface;[60] and ion migration into or out of the thin layer.[58,61]

4.2.1 Electrochemically Modulated Infrared Spectroscopy

The initial demonstration of specular reflection IR spectroelectrochemistry by Bewick et al. used a continuous 10 Hz potential modulation and a high-throughput dispersive-IR spectrometer.[55] In this case, the potential modulation was much faster than the rate of wavelength scanning, providing improved signal-to-noise for the measurement. They demonstrated the utility of this method by modulating the oxidation state of thianthrene, and by spectroscopically monitoring the potential-induced adsorption/desorption of indole.[55] The method, which they called electrochemically modulated infrared spectroscopy (EMIRS), provided improved signal-to-noise by having relatively long dwell times at each resolution element of the dispersive spectrometer, and by introducing the time dependence of the spectral signal.

Shortly after its initial demonstration, the EMIRS technique was used to study the electrosorption of methanol.[59] Modulating the electrode potential at 8.5 Hz, two IR features were observed. The first, at 1850 cm^{-1}, was attributed to an adsorbed CO at a bridging site, and a complex bipolar band centered at approximately 2070 cm^{-1} was attributed to adsorbed CO bonded to a linear site. The bipolar nature of the band at 2070 cm^{-1} is representative of the changing interfacial properties as the potential is altered. In this case, changes in the IR spectrum of the adsorbed CO with potential alterations provide a probe of the interfacial properties. The shift in the absorption position of the linearly

bonded CO illustrates the subtle changes that arise from the potential modulation that can be monitored by EMIRS.

The power of EMIRS measurements comes from the ability to measure subtle changes within the thin layer that are brought about by the potential change. Although other techniques use the EMIRS method with a dispersive spectrometer, FTIR instruments offer signal-to-noise advantages beyond those obtained by using a potential modulation. When using a scanning FTIR instrument, however, the method of spectral data collection must be modified from that used with EMIRS.

4.2.2 Potential Difference Infrared Spectroscopy

With scanning FTIR instruments, a continuous 10-Hz potential modulation is not useful because the sample must be at equilibrium during each interferometer scan. However, potential difference infrared spectroscopy (PDIRS) may still be acquired by synchronizing the potential alteration with the interferometer scans. Two general methods have been used.

In the first potential difference FTIR technique, a small number of interferometer scans are collected at one applied potential, then the potential is changed to a second value and a second set of interferometer data collected. The electrode potential is returned to the first value and this sequence is repeated until a sufficient total number of interferometer scans are collected and co-added at each potential value. The interferograms for each potential value are then Fourier transformed, and the spectrum corresponding to the first potential is ratioed against the spectrum obtained at the second potential and then normalized. In this sequence, the potential is modulated between two values at a frequency that depends on the number of interferometer scans and the individual interferometer scan times. The difference spectra acquired by this technique are equivalent to those obtained by the EMIRS method. As with EMIRS, because the potential is altered between two values repeatedly, the electrochemical response must be reversible with respect to those potential values.

If the electrode response is not reversible with respect to the potential, difference spectra can still be measured, but spectra at individual potentials must be acquired in sequence. Here, the potential is systematically altered and individual spectra are acquired at each applied potential before measuring spectra at the second, third, fourth, etc. potentials. The catalytic decomposition of small organic molecules on Pt has been extensively studied in this manner.[62–64] After Fourier transformation, any two of these spectra may be ratioed against each other to obtain spectral information regarding changes that occur in the system between these different applied potentials.

An alternative strategy, used by Weaver et al., is to collective spectra as a very slow potential ramp is applied to the working electrode.[63–65] In this manner, IR spectra are acquired during different time blocks along the potential ramp. Assuming that the spectral acquisition is fast enough (depending on resolution, and interferometer frequency) and/or the potential scan slow enough, during the individual blocks of time the electrode potential changes only slightly. Using this technique, one can obtain a measure of the interfacial dynamics during a continuous, albeit slow, potential scan.

With the exception of the slow potential ramp method used by Weaver et al., continuous-scan FTIR instrumentation requires that spectra be acquired with the chemical system at equilibrium. Consequently, spectral information about fast, dynamic electrochemical phenomena cannot be studied using PDIRS. As described for internal reflection, however, step-scan FTIR instruments provide the opportunity to study dynamic phenomena on relatively short time-frames.

4.2.3 Electrochemically Modulated Infrared Reflection Spectroscopy by Step-scan Fourier Transform Infrared

Griffiths et al. have demonstrated the utility of step-scan methods for obtaining spectral information about dynamic electrochemical processes using specular reflection sampling.[13–15] One application studied the potential dependence of the IR spectrum of adsorbed CO. CO adsorbed on Pt has been extensively studied, providing a wealth of literature for comparison. They found that the signal-to-noise in their step-scan measurements was comparable to data collected by other EMIRS and PDIRS techniques. In addition, the opportunity to conduct a double modulation experiment with the step-scan instrument allowed Griffiths et al. to identify and compensate for the slow time response of the thin-layer cell – effectively using the spectral data and computational methods to correct for the resistance within the thin layer. Although not a critical parameter in these measurements, the opportunity to use the step-scan instrument to determine the level of resistance effects on the electrochemical response is a unique application of the spectral data. As step-scan instrumentation becomes more widely available this may become an important method if quantitative dynamic information is to be extracted from the measurements (e.g. reaction kinetics).

Griffiths et al. also demonstrated that EMIRS experiments using step-scan FTIR can be used to differentiate between species in the bulk solution and those confined to the interface.[14,15] When demodulating the spectral response due to the continuous potential modulation, the signal may be separated into in-phase and quadrature components. From these components, a magnitude

and a phase spectrum are calculated. When compared, the magnitude spectrum indicates when absorption of the source radiation is occurring and the phase spectrum indicates whether the absorption arises from species in an isotropic (e.g. solution) or a nonisotropic (e.g. interface) environment. Knowing the phase behavior of solution species allows the separation of all of the solution vibrational bands into one of the detection channels (e.g. the quadrature channel), and all the other IR absorptions arising from interfacial species in the other detection channel.[14,15] This method provides an opportunity to determine both solution and interfacial species during the same spectroelectrochemical experiment using phase-selective detection. In this case, the contributions from the solution and the interface are separated from each other in the final spectrum.

The combination of the more ideal electrode geometry to reduce the thin-layer resistance, along with the ability to measure and compensate for electrode time delay, suggests that step-scan instrumentation with specular reflection sampling will be a powerful method for studying the dynamics of electrode processes. Although the Griffiths' method clearly distinguishes solution and interfacial species, alternative methods for separating the interfacial contributions from the solution contributions are also available.

4.3 Polarization Effects

Greenler in 1966 demonstrated that specular reflection IR spectroscopy could be used to obtain spectra of thin organic films on metallic substrates.[9] Although not initially applied to spectroelectrochemistry, this work established the fundamental principles that govern the experimental aspects of specular reflection IR spectroelectrochemistry today. As with internal reflection methods, the behavior of IR light on reflection from a metallic substrate is determined by solving the boundary value problem for reflection with Fresnel's equations. This calculation has been conducted for many different systems of relevance to spectroelectrochemical applications, and are not reproduced here.[66,67] Only the principal results are discussed, in terms of their influence on sampling and instrumentation in the spectroelectrochemical experiment.

Greenler showed that the mean feld strength for the incident radiation maximizes when the IR radiation is incident on the substrate at glancing angles.[9] Quantitatively, the mean field strength is calculated by Greenler to be nearly 5000 times larger when the incident light reflects from the surface at 88°, compared to the value obtained for a normal-incidence reflection. In addition, Greenler showed that the sensitivity of the reflection experiment is approximately 25 times larger than that for a transmission experiment using a free-standing film of the same dimension. These calculations were for a thin isotropic organic layer on top of a metallic substrate and exposed to an air atmosphere. Although these results illustrate the utility of the specular reflection technique, they do not realistically model electrochemical systems.

Faguy and Fawcett reproduced Greenler's calculations assuming a multilayer model that more closely resembles an electrochemical system.[66] Their calculations not only accounted for many of the electrochemical parameters, they also considered real instrumental limitations of the spectroelectrochemical experiment; that is, parameters such as the angular divergence of the focused IR beam and the beam focus size were included in the model. In this model, a thin organic layer exists on top of the metallic substrate. The ambient atmosphere is a thin solution layer of some defined thickness (6 μm of acetonitrile was used in their calculations). On top of the solution layer is an initial layer that represents the IR-transparent window (e.g. CaF_2). This four-layer model allows calculation of the mean electric field strength at the metal interface using conditions that closely approximate the electrochemical experiment.

A result of particular interest from this work was the optimum angle of incidence for maximum field strength at the interface calculated by Faguy and Fawcett for this electrochemical system.[66] They found that the optimum angle of incidence for maximum reflectance for these electrochemical systems was not as glancing as reported by Greenler (Figure 15).[66] When the angular divergence due to a focusing beam of incident IR radiation is considered, the optimum incident angle is

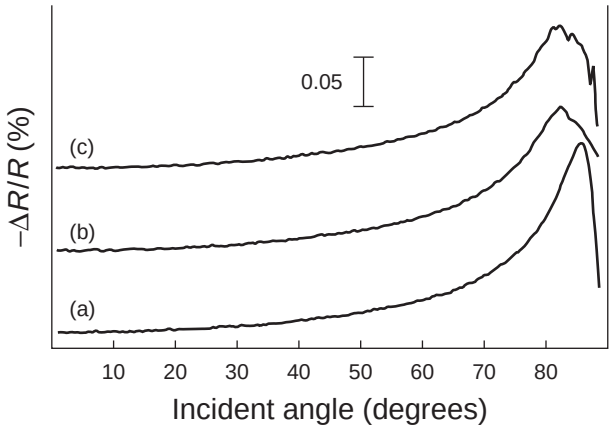

Figure 15 Calculated $-\Delta R/R$ versus incident angle at 2350 cm^{-1} for an adsorbed layer thickness of 0.5 mm (a) with noise injected, (b) as with (a) but with Gaussian averaging for angular divergence and (c) as with (b) but with beam attenuation. (Reprinted from P.W. Faguy, W.R. Fawcett, 'Infrared Reflection–Absorption Spectroscopy of the Electrode–Electrolyte Solution Interface: Optical Considerations', *Appl. Spectrosc.*, **44**(8), 1309–1316, copyright (1990) with permission from the Society for Applied Spectroscopy.)

reduced even further. Faguy and Fowcett's more realistic electrochemical model also found that the mean field strength is reduced at the interface, compared to values calculated by Greenler. This result indicates that the reflectance for these multilayer systems will be smaller than predicted by Greenler's model. These results are important because a glancing angle of incidence is a nearly impossible sampling geometry for spectroelectrochemical cells designed for external reflection experiments to attain. With a finite thickness of the solution thin-layer and refraction effects at the boundary between the window and the solution, experimental incident angles of 60–65° are more common. Although Faguy and Fawcett's calculations predict lower reflectances for real systems, they also show that the angle of incidence typically used in spectroelectrochemical measurements is not as debilitating as expected from Greenler's work.

In his calculations, Greenler also showed that on reflection from the metallic substrate the phase behavior of the incident light depended on the polarization of the incident light.[9] Light that is polarized perpendicular to the plane of reflection, commonly called s-polarized light, has a near 180° phase shift at the reflection plane for all angles of incidence. Consequently, the incident field strength from the reflected light with perpendicular polarization goes to zero at the interface. Light which is polarized parallel to the plane of reflection, commonly called p-polarized light, has a phase shift that depends on the angle of incidence. For glancing angles, this phase shift is nearly 90°, resulting in an electric field that oscillates normal to the substrate. The result of this phase behavior is that, on reflection, the incident radiation generates a field at the interface that is oriented normal to the metallic substrate. This polarization dependence at the interface generates what has commonly been referred to as the surface selection rules. That is, only those vibrational motions for confined species with a component of their transition dipole moment change that lies perpendicular to the substrate are observed. This is due to the electric field from the incident radiation being present only in a direction normal to the substrate. There are a number of important experimental consequences that come from this result.

Discrimination between solution-phase material and adsorbed species can be obtained by comparing spectra collected using polarized incident radiation. When the incident light is fixed so that its electric field is oriented parallel to the plane of incidence (p-polarized), the resultant spectrum contains information about species present in both the bulk solution and those confined to the interface. Repeating the experiment with the incident light polarized perpendicular to the plane of reflection results in spectra that contain information only about species in the bulk solution. These results are due to the directional aspect of the interfacial electric field. Species present in the solution are randomly oriented and can interact equally with the p- and s-polarized light. When confined to the interface, species can only interact with incident radiation that has p-polarization. Comparison of spectra, obtained under identical conditions with the exception of the polarization of incident light, allows identification of those vibrational modes that arise from confined versus solution-phase material. Foley et al. demonstrated this effect when investigating the potential induced adsorption of SCN^- by specular reflection PDIRS.[56]

For systems in which only species confined to the interface are studied, the spectral signal-to-noise ratio may be increased by a factor of two by using incident radiation that is only p-polarized. This is understood by considering the behavior of p- and s-polarized light at the interface. p-Polarized light is able to interact with the confined material; consequently, it will possess spectral information as well as random noise. Incident light that is s-polarized has no interaction with species confined to the electrochemical interface. Consequently, it will contain no spectral information about the confined species, but will carry random noise. By filtering the s-polarized light, 50% of the noise contribution has been eliminated without affecting the spectral signal.

Evaluation of the reflection IR spectrum of a material confined to the interface relative to its isotropic bulk spectrum can also provide information about the orientation/structure of the molecule at the interface. Although not an electrochemical application, Allara and Nuzzo illustrate the principle of this analysis when characterizing the structure of long-chain surfactants adsorbed on metallic substrates.[68] From the results of a Kramers–Kronig evaluation of an IR spectrum of a randomly distributed isotropic sample, a theoretical reflection spectrum of the sample can be obtained. The calculated reflection spectrum assumes that the adsorbed layer is homogeneous, isotropic and has the same structure as found in the bulk. If the interfacial structure is anisotropic, the experimental reflection spectrum will differ from the theoretical spectrum by the intensity of the vibrational modes. These intensity differences arise from the anisotropy of the sample and the directional aspect of the incident field strength at the interface. Comparison of this calculated, theoretical spectrum with the actual experimental reflection spectrum allows evaluation of the average orientation/structure of the molecules when they are confined to the interface.

4.3.1 Reflection Losses

Although polarization behavior on reflection from the electrode surface provides valuable information regarding surface-confined species, care must be taken to ensure

that the experiment does not introduce polarization artifacts. Just as the two polarizations have different reflectivities from the electrode, they also reflect with different efficiencies from other surfaces. It is possible, therefore, that the optical path of the instrument may introduce some differential polarization of the source radiation. For example, the different polarizations of light may reflect with different efficiencies from the front surface of the IR-transparent window of the spectroelectrochemical cell. For a flat disk-shaped window, the amount of light that is transmitted through the window depends on the incident polarization as well as the angle of incidence. When using a glancing angle, light polarized perpendicular to the reflection plane reflects from the window and is not as efficiently transmitted through to probe the electrochemical interface as is light polarized parallel to the reflection plane. Using a trapezoidal or hemicylindrical prism as the window may circumvent this problem. With prism windows, the IR light has normal incidence at the ambient–window interface and each polarization is transmitted with equal efficiency. Differential polarization induced by the instrument or sample cell is of particular concern when using a continuous polarization modulation for spectral sampling.

4.3.2 Polarization Modulation

Like potential modulation, one can use a continuous polarization modulation to effectively cancel out those portions of the spectrum that are invariant to the polarization of light. Unlike potential modulation, however, the incident light polarization may be modulated at a frequency high enough that this modulation method may be coupled to scanning FTIR spectral data acquisition. Conceptually, polarization modulation takes advantage of the surface selection rules to differentiate between the interphase and the bulk. Species in the bulk solution are equally likely to interact with p- and s-polarized light because their vibrational motions are randomly oriented in space and not confined to a defined direction over the time of the experiment. Due to the phase behavior at the reflection, interfacial material only interacts with the p-polarized light. If the reflected light is detected at the frequency of the polarization modulation, the resultant reflection spectrum is given by Equation (7),

$$-\frac{\Delta R}{R} = \frac{R_p - R_s}{R_p + R_s} \tag{7}$$

where R_p is the sample reflectance obtained with the p-polarized incident radiation and R_s is the sample reflectance obtained with the s-polarized radiation. It is obvious from this equation that species that have equal interaction with both p- and s-polarized light (e.g. bulk-phase material) will cancel in the final reflectance spectrum, whereas those with differential reflectivity will be observed. Consequently, the final reflectance spectrum will contain spectral information about only the interfacial region.

Typically the phasing behavior is thought to predominate in the region of space extending a distance approximately one-quarter of a wavelength away from the substrate into the solution. In the mid-IR region of the spectrum, therefore, the polarization modulation experiment samples approximately the first 1.5–2.5 μm into the solution away from the substrate. For a well-aligned spectroelectrochemical cell, this sampling dimension can be up to 50% of the total thin-layer thickness. Although sampling by polarization modulation does not discriminate entirely against solvent in the thin layer, it does isolate a smaller region of the interphase. This results in an overall enhancement of the spectral signal-to-noise. This provides the possibility, under the correct experimental conditions, of monitoring by IR spectroscopy the diffusion behavior at the interface. Combining polarization modulation with a potential modulation scheme, as Kunimatsu et al. demonstrate,[49] provides excellent signal-to-noise for monitoring the potential-dependent behavior of adsorbed molecules.

There have been several experimental demonstrations of the utility of polarization modulation to electrochemical systems. Kunimatsu et al. used the technique to study the adsorption of CO to Pt electrodes.[49] Anderson and Gatin investigated the structural behavior of self-assembled monolayers when potential is applied to the substrate.[57] Saez and Corn investigated the structure of polymer films deposited on electrodes as a function of electrode potential.[69]

4.3.3 Photoelastic Modulator

To conduct the polarization modulation experiment, the polarization of the incident radiation must be rapidly alternated between the linear p- and s-polarization extremes. The frequency of this polarization modulation must be fast relative to the frequency of the interferometer frequency in order to separate the signal modulation due to the polarization from that due to the scanning interferometer. This separation of the two time-dependent signals is possible if the light polarization is modulated at a high frequency relative to the frequency of the interferometer. An instrument called a photoelastic modulator provides this high-frequency dynamic change in the incident light polarization state.

Construction of the photoelastic modulator was first described in 1966.[70] Hipps and Crosby have described the principle of the dynamic polarization of light by photoelastic modulation.[71] The optically active portion of the photoelastic modulator is an isotropic crystal

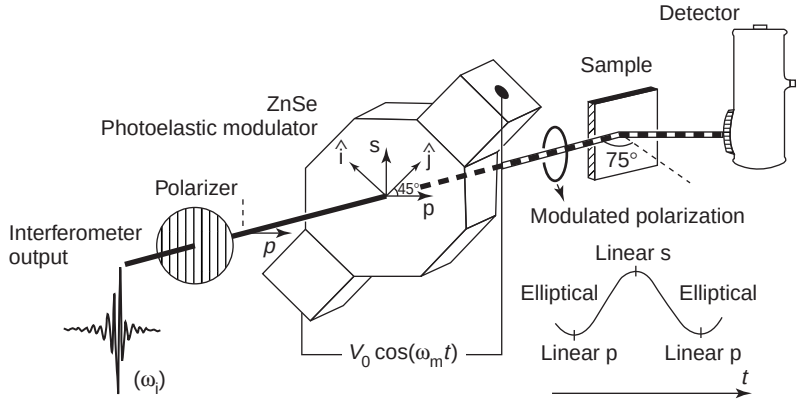

Figure 16 Diagram of the optical path used in the polarization modulation FTIR experiment. Illustration of the linear to elliptical polarization change as the linearly polarized light passes through the photoelastic modulator. (Reprinted from T. Buffeteau, B. Desbat, J.M. Turlet, 'Polarization Modulation FTIR Spectroscopy of Surfaces and Ultra-thin Films: Experimental Procedure and Quantitative Analysis', *Appl. Spectros.*, **45**(3), 380–389, copyright (1991) with permission from the Society for Applied Spectroscopy.)

of some material transparent in the spectral region of interest. For IR measurements, instruments that use CaF_2 and ZnSe crystals are commercially available. A periodic stress is applied to the isotropic crystal with a physical transducer, generally a piezoelectric material bonded to the stress axis of the IR-transparent crystal (Figure 16).[72] This mechanical stress induces a strain in the crystal along this axis. The strain in the crystal induces a variation of the optic ellipsoid of the crystal material at a resonance frequency.

As seen in Figure 16, the incident radiation is linearly polarized prior to passing through the photoelastic modulator. The electric field of the incident linear polarized light is aligned 45° away from the crystal strain axis. The strain is driven sinusoidally at the resonance frequency of the crystal, so the refractive index of the crystal along this axis varies sinusoidally as well. The refractive index change creates a time dependent phase shift of the incident light, which is given by Equation (8),

$$\phi = \phi_0 \cos(wt) \tag{8}$$

where ϕ_0 is the maximum phase shift (180°), w is the resonant frequency, and t is time. When linearly polarized light passes through the photoelastic modulator crystal, the light emerges elliptically polarized due to the strain-induced phase shift. Dependent on when during the stress–strain cycle the light emerges, the polarization of the emerging light changes between linear polarization and different degrees of elliptical polarization. If the maximum phase shift (ϕ_0) is 180°, then during a complete strain cycle the polarization of the light emerging from the modulator goes through two complete cycles of linear p- to linear s-polarized light (which are 90° out of phase from each other).

The intensity of the light that reaches the detector after passing through the photoelastic modulator and reflecting from the metallic substrate is given by Equation (9):[72]

$$I_d = C\frac{I_o(w)}{2}\{(R_p + R_s) - 2(R_p - R_s)J_2(\phi_0)\cos(2wt)\} \tag{9}$$

where C is a constant, $I_o(w)$ is the light intensity as it enters the photoelastic modulator, $J_2(\phi_0)$ is a second-order Bessel function, and all other terms have their normal meaning. The second-order Bessel function originates from the wavelength dependence of the peak efficiency from the polarization modulation. Strictly, the photoelastic modulator will only modulate the polarization of the light between the linear p- and s-polarized extremes for a single wavelength. This wavelength is determined by the potential used to drive the strain of the modulator crystal. That is, the maximum phase change, ϕ_0, is given by Equation (10):[71]

$$\phi_0 = \frac{GV}{\lambda} \tag{10}$$

where G is a constant value, V is the strain driving voltage, and λ is the wavelength of peak polarization retardation. For a given drive voltage, therefore, only one wavelength will have the maximum phase shift. For all other wavelengths, the polarization states during the mechanical strain cycle are elliptical.

Seki et al. investigated this wavelength dependence of the polarization state and its influence on distorting the measured IR spectrum.[49] Although the polarization modulation is only accurate for a single wavelength, they found that the elliptical light does not dramatically

distort the measured spectrum in the wavelength region surrounding modulation wavelength. For example, if the retardation of the photoelastic modulator is set at 2100 cm^{-1}, the multiplication factor from the Bessel function varies from 1 (at 2100 cm^{-1}) to 0.8 (at 1450 cm^{-1} and 2600 cm^{-1}) within the range 1450–2600 cm^{-1}. Over this wavenumber region with this limited distortion, the Bessel function can be approximated by a parabola, fit to the baseline, and then subtracted from the polarization modulation spectrum.

4.3.4 Detection of Polarization Modulation/Fourier Transform/Infrared Reflection–Absorption Spectroscopy

As stated before, the different polarizations of light will reflect from surfaces with different efficiencies. In the polarization modulation experiment, after the source radiation passes through the photoelastic modulator it should reflect from only one surface (e.g. the surface being characterized). This ensures that the spectral information encoded by the differential reflectivity at the interface of interest is not mixed with polarization effects at other reflections. Consequently, lenses are used to collect and focus the light after reflecting from the interface of interest in the polarization modulation experiment. In cases where mirrors are used after the sample reflection, a plate of some IR-transparent material (e.g. KBr) is placed into the optical path and at an angle relative to the direction of light propagation.[73] This plate offsets any polarization discrimination due to reflection differences from the optics of the instrument. Although this optical compensation works to balance polarization differentiation that occurs within the optics of the instrument, it is inefficient because it throws away some of the source radiation.

Polarization modulation reflection–absorption spectroscopy using FTIR instrumentation is a double modulation experiment. As such, the frequency components of these two modulations need to be sufficiently disparate so that they may be effectively separated by lock-in detection. A typical IR photoelastic modulator operates at a frequency of 74 kHz. The spectral information from an FTIR instrument operating under normal conditions contains frequencies ranging from a few hertz to tens of kilohertz. The upper range of the spectral frequencies is too close to the frequency of the polarization modulation by the photoelastic modulator for efficient separation by lock-in detection. Lowering the scan speed of the interferometer increases the difference between the spectral frequency and the polarization modulation frequency to allow lock-in detection. The ultimate decrease in the interferometer frequency is accomplished using a step-scan interferometer where the discrete steps of the interferometer moving mirror reduce the instrument frequency to essentially zero.[15] Highly sensitive detectors, such as HgCdTe and InSb, operate most efficiently when there are fast changes of the IR light intensity incident on them. Decreasing the interferometer scan speed lowers the efficiency of these detectors. Lowering the scan speed of the interferometer also lowers the stability of the interferometer and increases the level of noise in the measured spectrum. Although lowering the interferometer scan rate is often required for detection with a lock-in amplifier, it does come at the cost of lower spectral signal-to-noise. To efficiently collect the double modulation spectra with a continuous scan FTIR instrument, the rise time of the lock-in amplifier must be short, requiring a short time constant of the lock-in amplifier instrument (below 1 ms). This requirement places additional constraints on the instrumentation required for these measurements. Nevertheless, polarization modulation FTIR methods using lock-in detection have been demonstrated, and have been used in several spectroelectrochemical applications.[49,57,69]

Corn et al. developed electronics that allow real-time sampling for the acquisition of FTIR spectra during the double modulation experiment.[74,75] Unlike demodulating the frequency components of the experiment with a lock-in amplifier, these electronics do not require that the two modulation frequencies have a large difference. The instrument provides an accurate separation of the high-frequency component of the polarization modulation from the slower spectral frequency of the interferometer. Consequently, the interferometer can be operated using normal scan rates during the spectroelectrochemical experiment. This improves the overall signal-to-noise of the polarization modulation experiment because the interferometer is stable at the higher scan rates, and because the IR detectors operate more efficiently at the higher scan rates. Several spectroelectrochemical applications of the real-time sampling electronics have appeared. Richmond et al. acquired spectrochemical data for the adsorption of thiocyanate, imidazole, and glucose using real-time sampling electronics with the polarization modulation method as well as a static polarization technique. They found that the spectra were virtually identical, but that the real-time electronics provided better rejection of the solvent and better overall signal-to-noise.[76] Real-time sampling of polarization modulation FTIR experiments has also been used to monitor water and hydroxide vibrations associated with the interface.[77] This is an extremely challenging spectroscopic measurement because of the presence of water as the bulk solvent in the thin layer. The properties of the polarization modulation experiment provide the sensitivity and ability to discriminate against the bulk required to make these types of measurements.

5 SUMMARY

As illustrated throughout this article, the combination of IR spectroscopy and electrochemistry is an area of active research interest. The two experiments provide complementary information. Conducting the two measurements simultaneously provides insight about the chemical system that could not be obtained otherwise. The principal issues faced when conducting these simultaneous determinations are spectral sampling of the electrochemical system, and the overall signal-to-noise of the measurement. The three dominant sampling methods have unique strengths and weaknesses. Interestingly, these three methods tend to be complementary to each other, each having strengths in different areas. The type of information desired from the spectroelectrochemical experiment therefore dictates the sampling method used.

Sensitivity advances in the field of IR spectroelectrochemistry have followed improvements in the instrumentation available for the measurements, and this is likely to continue to drive the area. Modulation techniques introduce a time-variant signal that allows use of phase-sensitive detection methods to improve the sensitivity of the measurement. Polarization modulation methods help to isolate the interface, enhancing the sensitivity for double-layer phenomena. The availability of step-scan FTIR instrumentation allows use of a wider range of modulation frequencies that were not previously available for spectroelectrochemical experiments with traditional scanning FTIR instrumentation. These developments have made accessible for spectroelectrochemical measurement those phenomena that occur at the electrode interface and at short times after initiating the electrochemical experiment. New challenges continue to appear that can benefit from IR spectroelectrochemical measurements.

ACKNOWLEDGMENTS

M.R.A. would like to acknowledge Dr Marilyn Gatin, Dr John Roush, Dr Jimin Huang, Susanne Dana, Minhui Zhang, C. Douglas Taylor, Alison Grieshaber, Joshua Joseph, Mark Scalf, Cynthia Kraft, Richard Anderson, and Michaiah Parker for the many contributions that they have made to our research group over the years.

ABBREVIATIONS AND ACRONYMS

CGE	Conducting Glass Electrode
EMIRS	Electrochemically Modulated Infrared Spectroscopy
FTIR	Fourier Transform Infrared
IR	Infrared
IRE	Internal Reflection Element
IRRAS	Infrared Reflection–Absorption Spectroscopy
MIRFTIRS	Multiple Internal Reflection Fourier Transform Infrared Spectroscopy
OTTLE	Optically Transparent Thin-layer Electrode
PDIRS	Potential Difference Infrared Spectroscopy
SEIRAS	Surface-enhanced Infrared Absorption Spectroscopy
SERS	Surface-enhanced Raman Spectroscopy
TBAF	Tetra-*n*-butylammonium Hexafluorophosphate
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Microbalance, Electrochemical Quartz Crystal • Selective Electrode Coatings for Electroanalysis • Self-assembled Monolayers on Electrodes • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques • Ultraviolet/Visible Spectroelectrochemistry

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy

REFERENCES

1. R. Kuwana, R.K. Darlinton, D.W. Leedy, 'Electrochemical Studies Using Conducting Glass Indicator Electrodes', *Anal. Chem.*, **36**(10), 2023–2025 (1964).
2. W.N. Hansen, T. Kuwana, R.A. Osteryoung, 'Observation of Electrode–Solution Interface by Means of Internal Reflection Spectroscopy', *Anal. Chem.*, **38**(13), 1810–1821 (1966).
3. R.W. Murray, W.R. Heineman, G.W. O'Dom, 'An Optically Transparent Thin-layer Electrochemical Cell', *Anal. Chem.*, **39**(13), 1666–1668 (1967).
4. A. Yildiz, P.T. Kissinger, C.N. Reilley, 'Evaluation of an Improved Thin-layer Electrode', *Anal. Chem.*, **40**(7), 1018–1024 (1968).

5. H.B. Mark, B.S. Pons, 'An In Situ Spectrophotometric Method for Observing the Infrared Spectra of Species at the Electrode Surface During Electrolysis', *Anal. Chem.*, **38**(1), 119–121 (1966).
6. W.R. Heineman, J.N. Burnett, R.W. Murray, 'Optically Transparent Thin-layer Electrodes: Ninhydrin Reduction in an Infrared Transparent Cell', *Anal. Chem.*, **40**(13), 1974–1978 (1968).
7. D.R. Tallant, D.H. Evans, 'Application of Infrared Internal Reflection Spectrometry to Studies of the Electrochemical Reduction of Carbonyl Compounds', *Anal. Chem.*, **41**(6), 835–838 (1969).
8. A.H. Reed, E. Yeager, 'Infrared Internal Reflexion Studies of the Germanium/Electrolyte Interface', *Electrochim. Acta*, **15**, 1345–1354 (1970).
9. R.G. Greenler, 'Infrared Study of Adsorbed Molecules on Metal Surfaces by Reflection Techniques', *J. Chem. Phys.*, **44**(1), 310–315 (1966).
10. R.G. Greenler, 'Reflection Method for Obtaining the Infrared Spectrum of a Thin Layer on a Metal Surface', *J. Chem. Phys.*, **50**(5), 1963–1968 (1969).
11. S.A. Borman, 'Fourier Transform IR: Are the Older Grating Instruments Going the Way of the Dinosaurs?', *Anal. Chem.*, **55**(11), 1054A–1056A (1983).
12. R.A. Palmer, J.L. Chao, R.M. Dittmar, V.G. Gregoriou, S.E. Plunkett, 'Investigation of Time-dependent Phenomena by Use of Step-scan FTIR', *Appl. Spectrosc.*, **47**(9), 1297–1310 (1993).
13. B. Budevska, P.R. Griffiths, 'Step-scan FTIR External Reflection Spectrometry of the Electrode–Electrolyte Interface by Potential Modulation', *Anal. Chem.*, **65**(21), 2963–2971 (1993).
14. C.M. Pharr, P.R. Griffiths, 'Infrared Spectroelectrochemical Analysis of Adsorbed Hexacyanoferrate Species Formed During Potential Cycling in the Ferrocyanide/Ferricyanide Redox Couple', *Anal. Chem.*, **69**(22), 4673–4679 (1997).
15. C.M. Pharr, P.R. Griffiths, 'Step-scan FTIR Spectroelectrochemical Analysis of Surface and Solution Species in the Ferricyanide/Ferrocyanide Redox Couple', *Anal. Chem.*, **69**(22), 4665–4672 (1997).
16. A.J. Bard, L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, John Wiley and Sons, New York, 1980.
17. A.T. Hubbard, F.C. Anson, 'The Theory and Practice of Electrochemistry with Thin Layer Cells', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 129–214, Vol. 4, 1970.
18. W.R. Heineman, F.M. Hawkridge, H.N. Blount, 'Spectroelectrochemistry at Optically Transparent Electrodes. II. Electrodes Under Thin-layer and Semi-infinite Diffusion Conditions and Indirect Coulometric Titrations', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–113, Vol. 13, 1984.
19. F.M. Hawkridge, 'Electrochemical Cells', in *Laboratory Techniques in Electroanalytical Chemistry*, eds. P.T. Kissinger, W.R. Heineman, Marcel Dekker, New York, 267–291, 1996.
20. M. Krejcik, M. Danek, F. Hartl, 'Simple Construction of an Infrared Optically Transparent Thin-layer Electrochemical Cell: Applications to the Redox Reactions of Ferrocene, $Mn_2(CO)_{10}$ and $Mn(CO)_3$(3,5-di-*t*-butylcatecholate)', *J. Electroanal. Chem.*, **317**, 179–187 (1991).
21. B.J. Brisdon, S.K. Enger, M.J. Weaver, R.A. Walton, 'Infrared Spectroelectrochemistry of η^3-Allyl Dicarbonyl Complexes of Molybdenum(II) and Tungsten(II)', *Inorg. Chem.*, **26**, 3340–3344 (1987).
22. J.P. Bullock, D.C. Boyd, K.R. Mann, 'In Situ Infrared Spectroelectrochemistry Studies of Ferrocene, $[Rh_2(dimen)_2(dppm)_2](PF_6)_2$, and (mes)$Cr(CO)_3$(dimen = 1,8-Diisocyanomenthane; dppm = Bis(diphenylphosphinomethane); mes = Mesitylene). A Useful Technique for the Characterization of Electrochemically Generated Organometallic Species', *Inorg. Chem.*, **26**, 3084–3086 (1987).
23. W.A. Nevin, A.B.P. Lever, 'Reusable Thin-layer Spectroelectrochemical Cell for Nonaqueous Solvent Systems', *Anal. Chem.*, **60**, 727–730 (1988).
24. I.T. Bae, M. Sandifer, Y.W. Lee, D.A. Tryk, C.N. Sukenik, D.A. Sherson, 'In Situ Fourier Transform Infrared Spectroscopy of Molecular Adsorbates at Electrode–Electrolyte Interfaces: A Comparison between Internal and External Reflection Modes', *Anal. Chem.*, **67**(24), 4508–4513 (1995).
25. W.N. Hansen, 'Internal Reflection Spectroscopy in Electrochemistry', in *Advances in Electrochemistry and Electrochemical Engineering*, eds. P. Delahay, C.W. Tobias, John Wiley & Sons, New York, 1–60, Vol. 9, 1973.
26. N. Harrick, *Internal Reflection Spectroscopy*, Interscience Publishers, New York, 1967.
27. W.N. Hansen, 'Electric Fields Produced by the Propagation of Plane Coherent Electromagnetic Radiation in a Stratified Medium', *J. Opt. Soc. Am.*, **58**(3), 380–390 (1968).
28. B.D. MacCraith, 'Enhanced Evanescent Wave Sensors Based on Sol-gel-derived Porous Glass Coating', *Sens. Actuators, B*, **11**, 29–34 (1993).
29. D.B. Parry, J.M. Harris, K. Ashley, 'Multiple Internal Reflection Fourier Transform Infrared Spectroscopic Studies of Thiocyanate Adsorption on Silver and Gold', *Langmuir*, **6**(1), 209–217 (1990).
30. B.W. Johnson, J. Bauhofer, K. Doblhofer, B. Pettinger, 'Practical Considerations of the IR Attenuated Total Reflection (IR-ATR) Technique for Electrochemical Investigations', *Electrochim. Acta*, **37**(12), 2321–2329 (1992).
31. A. Hatta, T. Ohshima, W. Suëtaka, 'Observation of the Enhanced Infrared Absorption of *p*-Nitrobenzoate on Ag Island Films with an ATR Technique', *Appl. Phys. A*, **29**, 71–75 (1982).
32. A. Hatta, Y. Chiba, W. Suëtaka, 'Infrared Absorption Study of Adsorbed Species at Metal/Water Interfaces

by use of the Kretschmann Configuration', *Surf. Sci.*, **158**, 616–623 (1985).
33. T. Wadayama, O. Suzuki, Y. Suzuki, A. Hatta, 'Infrared Absorption Enhancement of *p*-Cyanobenzoic Acid on Silver Island Films Deposited on Oxidized and Hydrogen-terminated Si(100)', *Appl. Phys. A*, **64**(501), 506 (1997).
34. R. Barbour, Z. Wang, I.T. Bae, Y.V. Tolmachev, D.A. Sherson, 'Channel Flow Cell for Attenuated Total Reflection Fourier Transform Infrared Spectroelectrochemistry', *Anal. Chem.*, **67**(21), 4024–4027 (1995).
35. M. Osawa, M. Kuramitsu, A. Hatta, W. Suëtaka, 'Electromagnetic Effect in Enhanced Infrared Absorption of Adsorbed Molecules on Thin Metal Films', *Surf. Sci.*, **175**, L787–L793 (1986).
36. M. Osawa, K. Ataka, K. Yoshii, T. Yotsuyanagi, 'Surface-enhanced Infrared ATR Spectroscopy for In Situ Studies of Electrode–Electrolyte Interfaces', *J. Electron Spectrosc. Relat. Phenom.*, **64/65**, 371–379 (1993).
37. B.W. Johnson, K. Doblhofer, 'Barium Fluoride as an Internal Reflection Element of IR-ATR Spectroelectrochemistry', *Electrochim. Acta*, **38**(5), 695–701 (1993).
38. A. Hartstein, J.R. Kirtley, J.C. Tsang, 'Enhancement of the Infrared Absorption from Molecular Monolayers with Thin Metal Overlayers', *Phys. Rev. Lett.*, **45**(3), 201–204 (1980).
39. M. Osawa, K. Ataka, K. Yoshii, Y. Nishikawa, 'Surface-enhanced Infrared Spectroscopy: The Origin of the Absorption Enhancement and Band Selection Rule in the Infrared Spectra of Molecules Adsorbed on Fine Metal Particles', *Appl. Spectrosc.*, **47**(9), 1497–1502 (1993).
40. Y. Nishikawa, T. Nagasawa, K. Fujiwara, M. Osawa, 'Silver Island Films for Surface-enhanced Infrared Absorption Spectroscopy: Effect of Island Morphology on the Absorption Enhancement', *Vib. Spectrosc.*, **6**, 43–53 (1993).
41. A. Hatta, Y. Suzuki, W. Suëtaka, 'Infrared Absorption Enhancement of Monolayer Species on Thin Evaporated Ag Films by Use of a Kretschmann Configuration: Evidence for Two Types of Enhanced Surface Electric Fields', *Appl. Phys. A*, **35**, 135–140 (1984).
42. T. Wadayama, T. Sakurai, S. Ichikawa, W. Suëtaka, 'Charge-transfer Enhancement in Infrared Absorption of Thiocyanate Ions Adsorbed on a Gold Electrode in the Kretschmann ATR Configuration', *Surf. Sci.*, **198**, L359–L364 (1988).
43. A. Hatta, Y. Sasaki, W. Suëtaka, 'Polarization Modulation Infrared Spectroscopic Measurements of Thiocyanate and Cyanide at the Silver Electrode/Aqueous Electrolyte Interface by Means of Kretschmann's ATR Prism Configuration', *J. Electroanal. Chem.*, **215**, 93–102 (1986).
44. M. Osawa, K. Yoshii, K. Ataka, T. Yotsuyanagi, 'Real-time Monitoring of Electrochemical Dynamics by Submillisecond Time-resolved Surface-enhanced Infrared Attenuated Total Reflection Spectroscopy', *Langmuir*, **10**(3), 640–642 (1994).
45. Y.V. Tolmachev, Z. Wang, D.A. Sherson, 'Theoretical Aspects of Laminar Flow in a Channel-type Electrochemical Cell as Applied to In Situ Attenuated Total Reflection/Infrared Spectroscopy', *J. Electrochem. Soc.*, **143**(10), 3160–3166 (1996).
46. A. Bewick, K. Kunimatsu, B.S. Pons, J.W. Russell, 'Electrochemically Modulated Infrared Spectroscopy (EMIRS): Experimental Details', *J. Electroanal. Chem.*, **160**, 47–61 (1984).
47. X.Q. Lin, K.M. Kadish, 'Characteristics of a Full Edge Current Flow Thin-layer Electrochemical Cell that Uses Both Internal (Real) and External (Auxillary) Reference Points', *Anal. Chem.*, **58**(7), 1493–1497 (1986).
48. P.a. Mosier-Boss, R. Newbery, S. Szpak, S.H. Lieberman, 'Versatile, Low-volume, Thin-layer Cell for In Situ Spectroelectrochemistry', *Anal. Chem.*, **68**(18), 3277–3282 (1996).
49. H. Seki, K. Kunimatsu, W.G. Golden, 'A Thin-layer Electrochemical Cell for Infrared Spectroscopic Measurements of the Electrode–Electrolyte Interface', *Appl. Spectrosc.*, **39**(3), 437–443 (1985).
50. D.S. Bethune, A.C. Luntz, J.K. Sass, D.K. Roe, 'Optical Analysis of Thin-layer Electrochemical Cells for Infrared Spectroscopy of Adsorbates', *Surf. Sci.*, **197**, 44–66 (1988).
51. D.K. Roe, J.K. Sass, D.S. Bethune, A.C. Luntz, 'Prospects for Transient IR Reflection–Absorption Spectroscopy of Adsorbed Species on Electrode Surfaces: Cell Design for a Laser Source', *J. Electroanal. Chem.*, **216**, 293–301 (1987).
52. H.Q. Zhang, X.Q. Lin, 'In Situ FTIR Spectroelectrochemistry of a Microarray Electrode', *J. Electroanal. Chem.*, **434**, 55–59 (1997).
53. Z. Li, X. Lin, 'Electrochemistry and In-situ IR Microspectroelectrochemistry in a Versatile Thin Layer Reflection Fourier Transform IR Microspectroelectrochemical Cell', *J. Electroanal. Chem.*, **386**, 83–87 (1995).
54. F. Hartl, H. Luyten, H.A. Nieuwenhuis, G.C. Shoemaker, 'Versatile Cryostated Optically Transparent Thin-layer Electrochemical (OTTLE) Cell for Variable-temperature UV/VIS IR Spectroelectrochemical Studies', *Appl. Spectrosc.*, **48**(12), 1522–1528 (1994).
55. A. Bewick, K. Kunimatsu, B.S. Pons, 'Infrared Spectroscopy of the Electrode–Electrolyte Interphase', *Electrochim. Acta*, **25**, 465–468 (1980).
56. J.K. Foley, B.S. Pons, J.J. Smith, 'Fourier Transform Infrared Spectroelectrochemical Studies of Anodic Processes in Thiocyanate Solutions', *Langmuir*, **1**(6), 697–701 (1985).
57. M.R. Anderson, M. Gatin, 'Effects of Applied Potential upon the In Situ Structure of Self-assembled Monolayers on Gold Electrodes', *Langmuir*, **10**(6), 1638–1641 (1994).
58. F.C. Nart, T. Iwasita, 'On the Adsorption of $H_2PO_4^-$ and H_3PO_4 on Platinum: An In Situ Study', *Electrochim. Acta*, **37**(3), 385–391 (1992).
59. B. Beden, C. Lamy, A. Bewick, K. Kunimatsu, 'Electrosorption of Methanol on a Platinum Electrode. IR

Spectroscopic Evidence for Adsorbed CO Species', *J. Electroanal. Chem.*, **121**, 343–347 (1981).
60. M.R. Anderson, J. Huang, 'The Influence of Cation Size upon the Infrared Spectrum of Carbon Monoxide Adsorbed on Platinum Electrodes', *J. Electroanal. Chem.*, **318**, 335–347 (1991).
61. V.B. Paulissen, C. Korzeniewski, 'Vibrational Analysis of Interfacial Phosphate Equilibria', *J. Electroanal. Chem.*, **290**, 181–189 (1990).
62. T. Iwasita, F.C. Nart, B. Lopez, W. Vielstich, 'On the Study of Adsorbed Species at Platinum from Methanol, Formic Acid, and Reduced Carbon Dioxide via In Situ FTIR Spectroscopy', *Electrochim. Acta*, **37**(12), 2361–2367 (1992).
63. D.S. Corrigan, M.J. Weaver, 'Adsorption and Oxidation of Benzoic Acid, Benzoate, and Cyanate at Gold and Platinum Electrodes as Probed by Potential-difference Infrared Spectroscopy', *Langmuir*, **4**(3), 599–606 (1988).
64. D.S. Corrigan, M.J. Weaver, 'Mechanism of Formic Acid, Methanol, and Corbon Monoxide Electrooxidation at Platinum as Examined by Single Potential Alteration Infrared Spectroscopy', *J. Electroanal. Chem.*, **241**, 143–162 (1988).
65. L.H. Leung, M.J. Weaver, 'Reactant Structural Effects on the Formation and Electro-oxidation of Adsorbed Carbon Monoxide from Small Organic Molecules at Platinum as Studied by Time-resolved FTIR Spectroscopy', *J. Electroanal. Chem.*, **240**, 341–348 (1988).
66. P.W. Faguy, W.R. Fawcett, 'Infrared Reflection–Absorption Spectroscopy of the Electrode–Electrolyte Solution Interface: Optical Considerations', *Appl. Spectrosc.*, **44**(8), 1309–1316 (1990).
67. D.D. Popenoe, S.M. Stole, M.D. Porter, 'Optical Considerations for Infrared Reflection Spectroscopic Analysis in the C–H Stretching Region of Monolayer Films at an Aqueous/Metal Interface', *Appl. Spectrosc.*, **46**(1), 79–87 (1992).
68. D.L. Allara, R.G. Nuzzo, 'Spontaneously Organized Molecular Assemblies. 2. Quantitative Infrared Spectroscopic Determination of Equilibrium Structures of Solution-adsorbed *n*-Alkanoic Acids on an Oxidized Aluminum Surface', *Langmuir*, **1**, 52–66 (1985).
69. E.I. Saez, R.M. Corn, 'In Situ Polarization Modulation/Fourier Transform Infrared Spectroelectrochemistry of Phenazine and Phenothiazine Dye Films at Polycrystalline Gold Electrodes', *Electrochim. Acta*, **38**(12), 1619–1625 (1993).
70. S.N. Jasperson, S.E. Schnatterly, 'An Improved Method for High Reflectivity Ellipsometry Based on a New Polarization Modulation Technique', *Rev. Sci. Instrum.*, **40**(6), 761–767 (1969).
71. K.W. Hipps, G.A. Crosby, 'Applications of the Photoelastic Modulator to Polarization Spectroscopy', *J. Phys. Chem.*, **83**(5), 555–562 (1979).
72. T. Buffeteau, B. Desbat, J.M. Turlet, 'Polarization Modulation FTIR Spectroscopy of Surfaces and Ultra-thin Films: Experimental Procedure and Quantitative Analysis', *Appl. Spectrosc.*, **45**(3), 380–389 (1991).
73. W.G. Golden, D.S. Dunn, J. Overend, 'A Method for Measuring Infrared Reflection–Absorption Spectra of Molecules Adsorbed on Low-area Surfaces at Monolayer and Submonolayer Concentrations', *J. Catal.*, **71**, 395–404 (1981).
74. M.J. Green, B.J. Barner, R.M. Corn, 'Real-time Sampling Electronics for Double Modulation Experiments with Fourier Transform Infrared Spectrometers', *Rev. Sci. Instrum.*, **62**(6), 1426–1430 (1991).
75. B.J. Barner, M.J. Green, E. Sáez, R.M. Corn, 'Polarization Modulation Fourier Transform Infrared Reflectance Measurements of Thin Films and Monolayers at Metal Surfaces Utilizing Real-time Sampling Electronics', *Anal. Chem.*, **63**(1), 55–60 (1991).
76. W.N. Richmond, P.W. Faguy, R.S. Jackson, S.C. Weibel, 'Comparison Between Real-time Polarization Modulation and Static Linear Polarization for In Situ Infrared Spectroscopy at Electrode Surfaces', *Anal. Chem.*, **68**(4), 621–628 (1996).
77. P.W. Faguy, W.N. Richmond, 'Real-time Polarization Modulation Infrared Spectroscopy Applied to the Study of Water and Hydroxide Ions at Electrode Surfaces', *J. Electroanal. Chem.*, **410**, 109–113 (1996).

Ion-selective Electrodes: Fundamentals

Vaneica Young
University of Florida, Gainesville, USA

1 Introduction 1
2 Historical Development 3
3 Theory of Electrode Response 6
3.1 General 6
3.2 Specific 12
4 Instrumentation 21
5 Ion-selective Microelectrodes 23
6 Novel Developments in Ion-selective Electrode Research 24
List of Symbols 25
Abbreviations and Acronyms 26
Related Articles 26
References 26

This article begins with an introduction to ion-selective electrodes (ISEs). The classification protocol based on the most recent International Union of Pure and Applied Chemistry (IUPAC) recommendations is given. The historical development of ISEs is summarized, starting with the pH electrode. For the inorganic cations and anions, a periodic table ''time-line'' is given. Recent history has seen the rise of polymer membrane ISEs, which are important in clinical analysis. The historical section concludes with a brief look at these ISEs. The general theory of the potential generation process is explained, and then aspects of the potential generation process that are specific for the different classes of ISEs. A short review of the instrumentation needed to make the measurements is given, followed by an overview of the latest commercially available instrumentation. One aspect of ISE research is the development of selective probes for determining the analyte content of individual cells. This requires miniaturization, and this has been achieved by the development of microelectrode ISEs. An overview of this area is given. Finally, some recent developments that represent the directions of research in the ISE field are given. Included are the development of sensors for patient monitoring, the development of ISE sensor arrays, and the development of more sophisticated electronic and mathematical methods of data analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Potentiometry is a static, interfacial electroanalytical technique that has found acceptance across a wide range of disciplines. This is due in large part to the development of ISEs. In the most recent recommendations of IUPAC, an ISE is defined as follows:[1]

> This is an electrochemical sensor, based on thin films or selective membranes as recognition elements, and is an electrochemical half-cell equivalent to other half-cells of zeroth (inert metal in a redox electrolyte), 1st, 2nd and 3rd kinds. These devices are distinct from systems that involve redox reactions (electrodes of zeroth, 1st, 2nd and 3rd kinds), although they often contain a 2nd kind electrode as the 'inner' or 'internal' reference electrode. The potential difference response has, as its principal component, the Gibbs energy change associated with permselective mass transfer (by ion exchange, solvent extraction or some other mechanism) across a phase boundary.

As indicated by this definition, one surface of the thin film or membrane is in contact with a bulk phase whose composition may be varied. The ''heart'' of an ISE is its thin film or membrane. Thus, much of the research in this area has involved a search for or the fabrication of thin films or membranes that are highly selective for a single ion. The ideal thin film or membrane is one that is specific for a single ion. The actual thin film or membrane, designed to target a single ion, called the analyte, is selective for a small set of ions. The ions left in the set after the analyte ion is removed are called electrode interferences. For example, the lithia-based glass membrane pH electrode targets H^+, but is selective for $\{H^+, Li^+, Na^+\}$. Its electrode interferences are $\{Li^+, Na^+\}$. For the glass membrane pH electrodes, the electrode interferences are serious problems only when they are at fairly high concentrations. This is the exception, rather than the rule.

Electrode interferences are by no means the only interferences that one encounters in working with these devices. Chemical interferences may be of equal or even greater importance than electrode interferences. Chemical interferences are species that either tie up the analyte ion or that change the membrane so that it can no longer effectively interact with the analyte. These types of interferences are very dependent on the type of active material constituting the membrane. Thus a second area of research in the field of ISEs has involved identifying various interferences and quantifying their effects on an ISE's signal, since interferences affect how and when various ISEs may be used.

ISEs are now being exploited as detectors in flow injection analysis systems. In such systems, the dynamic

behavior of the ISE assumes great importance, since that will determine the signal shape and the sampling rate. Characterizing the dynamic behavior of ISEs constitutes a third area of research in the field.

For the most part, research in these three areas has unfolded in the order given. For example, when a new membrane was developed, the obvious electrode interferences were examined. However, many interferences were identified only after researchers tried to develop applications in various disciplines. Because of the way in which early analyses were done, the dynamic response of an ISE was not of extreme importance. The response rate had to be rapid enough that potentiometric titration curves would not be distorted. For direct analysis, one could simply wait for the signal to stabilize.

This field has just passed its golden anniversary. So many membranes have been developed that a classification scheme had to be recommended. The membranes will be introduced by means of the latest classification scheme. After a short delineation of the historical development, the theory of the electrode response will be presented. Both the equilibrium and the dynamic response behavior in the absence and presence of interferences will be discussed. Instrumentation has played a very large role in the ascendency of this field. The instrumentation requirements will be discussed, and then some of the most recent instrumentation will be presented. Physiological and biological applications of ISEs have been a goal since the inception of the field. Now, miniaturization in instrumentation seems to be a current paradigm. ISEs are part of this effort, as researchers have and continue to develop ion-selective microelectrodes for intracellular measurements of ion activities. The progress in this area will be summarized. Finally, some recent, novel developments in this field will be reported, particularly those which may find commercial applications within the next 10 years. In this article, only a very small part of a vast literature can be included. However, the journal *Analytical Chemistry* has published a series of review articles that include ISEs.[2–24] They are an excellent place to start a search for more detail on a specific topic. For the most part, physiological and biological applications of ISEs involve the use of carrier-based microelectrode ISEs. These have been comprehensively reviewed by Bakker et al.[25,26]

In the new classification scheme, the old scheme is retained as a single category, *primary ISEs*. Two new categories, *compound or multiple membrane ISEs* and *metal contact or all-solid-state ISEs*, have been added. None of these categories include the ISFET (ion-selective field effect transistor) or CHEMFET (chemical-sensing field effect transistor), because they are considered to be purely capacitatively coupled devices.[27] However, some of these devices are commercially available, and they are called ''electrodes'' by their vendors.[28] It should be noted that some researchers have argued that the glass membrane ISEs are capacitative devices.[29,30] The experimental results that are the basis of their conclusions have not been independently verified. If true, the glass membrane ISE would be a half-cell for an electrochemical capacitor. The development of electrochemical capacitors is a fairly new field; the United States Department of Energy has had such a program since 1992.[31] The electrochemical capacitor is just a third type of electrochemical cell. Should it be shown that at least one ISE is a half-cell for an electrochemical capacitor, then we might expect a future classification scheme to include ISFETs or CHEMFETs. Reproduced below is the classification scheme according to the IUPAC 1994 recommendations.[1]

A. Primary ISEs

1. *Crystalline electrodes*

   (a) Homogeneous membrane electrodes
   (b) Heterogeneous membrane electrodes

2. *Non-crystalline electrodes*

   (a) Rigid, self-supporting, matrix electrodes
   (b) Electrodes with mobile charged sites

   – Positively charged, hydrophobic cations
   – Negatively charged, hydrophobic anions
   – Uncharged carrier electrodes
   – Hydrophobic ion-pair electrodes

B. Compound or Multiple Membrane ISEs

1. *Gas-sensing electrodes*
2. *Enzyme substrate electrodes*

C. Metal Contact or All-solid-state ISEs

Examples of the constructions used for the three main categories are shown in Figures 1–3. Gas-sensing electrodes consist of a reference electrode and indicator electrode that are in electrolytic contact. Some researchers say that these devices should not be called electrodes, because they are really electrochemical cells.[32] An ordinary ISE is used by placing it in a solution containing the analyte ion and measuring the potential, using a high-impedance potentiometer circuit, with respect to a reference electrode, placed in the same solution. The reference electrode, analyte solution and ISE constitute an electrochemical cell.

Many ISEs may now be purchased as combination electrodes. A combination electrode consists of an ISE and a reference electrode in a single package. An example

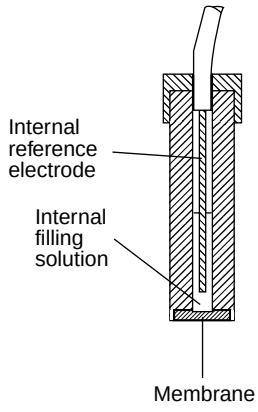

Figure 1 Schematic of a primary ISE.

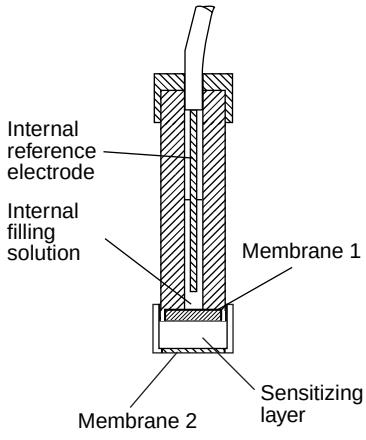

Figure 2 Schematic of a compound or multiple membrane ISE.

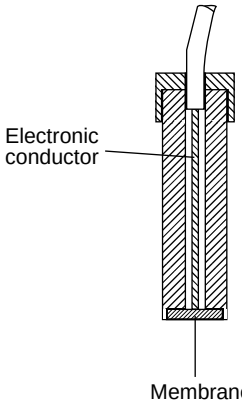

Figure 3 Schematic of a metal contact or all-solid-state ISE.

of a combination electrode, the Orion glass pH electrode, is shown in Figure 4. Some researchers would probably argue that a combination electrode is not really an electrode. If one immerses it in an analyte solution, then one has an electrochemical cell. An analogy is the

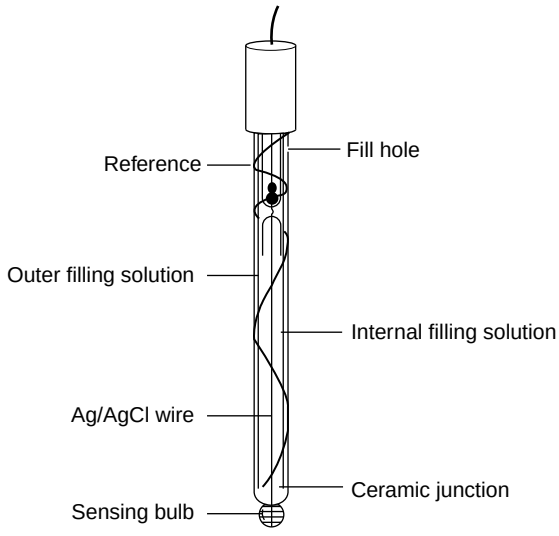

Figure 4 Orion combination glass pH electrode. (Reproduced from *A Concise Guide to Ion Analysis*,[28] by permission of Orion Research Inc.)

conventional car battery without the electrolyte. We do not regard it as an electrode just because the electrolyte is missing. For the foreseeable future, the use of the term combination electrode will undoubtedly continue. Technically correct and noncumbersome nomenclature for these devices does not exist at present.

2 HISTORICAL DEVELOPMENT

It is generally agreed that the progenitor of ISEs is the glass membrane acidity electrode. Cremer is credited with making the discovery that a large potential difference exists across a thin glass membrane in contact with solutions with large differences in acidity on opposite sides of the membrane.[33] It is fascinating to note that Cremer's work was actually the culmination of many studies on the electrical properties of glasses starting as far back as 1761![34] At that time, glasses were considered to be materials that consisted of silicon dioxide and various other oxides. The properties of the glasses were known to depend on which other oxides were added and on the stoichiometry of the resultant material. Thus, some glasses were shown to be conductors of electricity, while others were shown to be electrical insulators. Cremer, who was interested in the potential differences generated across certain biological tissues, recognized that the Warburg model of glass meant that it should be an ideal semipermeable membrane. He performed several experiments using a flask containing an electrolyte solution immersed in a beaker, also containing an electrolyte solution. A platinum wire was in contact with

each solution. He found that when the two electrolytes differed in their acidities, a large potential difference could be measured across the two platinum electrodes.

Systematic, quantitative studies on this phenomenon were made by Haber and Klemensiewicz on glass bulb electrodes.[35] In that paper, the theoretical model for the electrode response was developed by Haber, and the experimental confirmation of the model was performed by Klemensiewicz. For the theoretical model, Haber extended the work of H.W. Nernst.[36] Of particular relevance to the development of the mathematical form for the electrode potential of an ISE is Nernst's treatment of the potential difference across the interface between two electrolyte solutions, the well-known liquid junction potential, generated by nonFaradaic processes. He showed that the steady-state potential difference across an interface between a dilute solution of HCl (concentration c_1) and a more concentrated solution of HCl (concentration c_2) has the form of Equation (1):[36]

$$E = \frac{RT}{\Im}(t_+ - t_-)\ln\left(\frac{c_1}{c_2}\right) \tag{1}$$

where E is the potential of the interface on the side of the more concentrated solution relative to the more dilute solution, t_+ is the transport number for the cation, t_- is the transport number for the anion and $\Im$ is Faraday's constant. (Note that in this and other equations, the expressions for E have been written in modern form. In the original equations, R, the gas constant, is given in electrical dimensions. In modern usage, R is given in energy dimensions. R, in electrical units, is equal to R in energy dimensions divided by the Faraday constant $\Im$.) To verify his theory, Nernst studied concentration cells of the form in Equation (2):

$$\mathrm{M|MX}(c_1)\mathrm{|MX}(c_2)\mathrm{|M} \tag{2}$$

where MX is a strong electrolyte and $c_1 \neq c_2$. Here, Nernst proposed that the mathematical relationship for the potential difference across the metal electrolyte interface under conditions of zero current flow is given by Equation (3):[36]

$$E = \frac{RT}{\Im}\ln\left(\frac{C}{c}\right) \tag{3}$$

where C is the concentration of ion cores in the metal and c is the concentration of the electrolyte. According to Nernst, dissolution of the metal may occur, such that ions cross to the solution side and electrons remain in the metal, or metal ions from the electrolyte may specifically absorb on the metal surface. These processes are also nonFaradaic. Thus Nernst clearly proposed that his mathematical treatment was a general one valid for all electrified interfaces under conditions of equilibrium or steady-state charge transport. Haber extended Nernst's treatment to the case of a concentration cell involving a solid electrolyte MX and an aqueous electrolyte MX, with the cell notation as in Equation (4):

$$\mathrm{M|MX_{(s)}|MX_{(aq)}|M} \tag{4}$$

He showed that the potential at the interface between the aqueous and the solid electrolytes is given by Equation (5):

$$E = -\frac{RT}{\Im}\ln(C_{\mathrm{M}}) + K \tag{5}$$

where C_{M} is the concentration of the metal ion in the aqueous electrolyte solution and K is a constant. Haber then proposed that a hydrated glass membrane is a type of solid water phase. By considering the cell in Equation (6):

$$\mathrm{H_2|H_2O_{(s)}|H_2O_{(l)}|H_2} \tag{6}$$

he arrived at Equation (7) for the potential difference across the glass membrane – water interface:

$$E = -\frac{RT}{\Im}\ln(C_{\mathrm{H^+}}) + K \tag{7}$$

where $C_{\mathrm{H^+}}$ is the concentration of hydronium ions and K is a constant. Similarly, by considering the cell in Equation (8):

$$\mathrm{O_2|H_2O_{(s)}|H_2O_{(l)}|O_2} \tag{8}$$

he arrived at a second expression for the potential difference across the glass membrane – water interface (Equation 9):

$$E = +\frac{RT}{\Im}\ln(C_{\mathrm{OH^-}}) + K' \tag{9}$$

where $C_{\mathrm{OH^-}}$ is the concentration of hydroxide ions and K' is a constant. We see that Haber recognized that the glass membrane can sense either hydronium ions or hydroxide ions. He further indicated that Equation (9) could be written in terms of $C_{\mathrm{H^+}}$ by making use of the relationship $C_{\mathrm{OH^-}} = K_{\mathrm{w}}/C_{\mathrm{H^+}}$, giving Equation (10):

$$E = -\frac{RT}{\Im}\ln(C_{\mathrm{H^+}}) + \frac{RT}{\Im}\ln(K_{\mathrm{w}}) + K' \tag{10}$$

Haber then stated that Equations (7) and (10) are identical, which meant that K and K' are related by Equation (11):

$$K = \frac{RT}{\Im}\ln(K_{\mathrm{w}}) + K' \tag{11}$$

Haber's statement, given without proof, was widely accepted by subsequent analytical chemists. Later, in

describing the microscopic behavior giving rise to the electrode response, many analytical chemists only considered Equation (7), which has led to some misconceptions that are only now being addressed in the literature. Haber used his results to derive an expression for the potential difference across a glass membrane electrode which had an acidic solution on one side and an alkaline solution on the other. Thus, for the scheme in Equation (12):

$$\text{1. alkaline water } |H_2O_{(s)}| \qquad \text{2. acidic water} \tag{12}$$

he wrote Equation (13):

$$E = \frac{RT}{\Im} \ln \left(\frac{C_{H^+,2}}{C_{H^+,1}} \right) \tag{13}$$

where $C_{H^+,1}$ is the hydronium ion concentration in the alkaline water, $C_{H^+,2}$ is the hydronium ion concentration in the acidic water and E is the potential difference between the alkaline water – $H_2O_{(s)}$ interface and the acidic water – $H_2O_{(s)}$ interface.

Klemensiewicz performed an eloquent set of experiments that constituted the proof of Haber's theory. The electrochemical cells used resemble the type of set-up we use today, in that a reference electrode, a 1 N calomel electrode, and an indicator electrode, the glass membrane electrode, were used. The set-up is shown in Figure 5. The glass membrane electrodes were hand-blown from Thüringer glass rods to give a bulb approximately 2.5 cm in diameter and a thickness of 0.06–0.1 mm. Inside the glass membrane electrode, a dilute solution of HCl or KCl was used. Various strong acid – strong base titrations were performed, as follows: 0.1 N HCl with 1 N KOH, 0.1 N HCl with 1 N NaOH, 0.1 N H_2SO_4 with 0.1 N KOH, 0.1 N H_2SO_4 with 0.1 N NaOH, and 0.1 N KOH with 1 N HCl. The potential of the glass membrane electrode was measured with respect to the normal calomel electrode using a quadrant electrometer with reading telescope. Plots of the potential difference versus volume of titrant gave the typical sigmoidal curves that are well known to us today. The magnitude of the potential difference between the initial point of the titration and the final point of the titration, which corresponded to an extent of titration = 2, is expected to be given by Equation (13), with the final point corrected for dilution effects and the ratio of the concentrations of hydronium ion being that for the acidic solution to that for the basic solution. For titrations involving KOH, the experimental error for titrations performed at 20 °C averaged −7%, which is fairly good. When KOH was replaced with NaOH, much larger relative errors were obtained, −15 to −22%. This came to be known as the sodium error. Titration of a weak acid with a strong base was also performed; it showed the characteristic sigmoidal curve with the buffer region so familiar to us today.

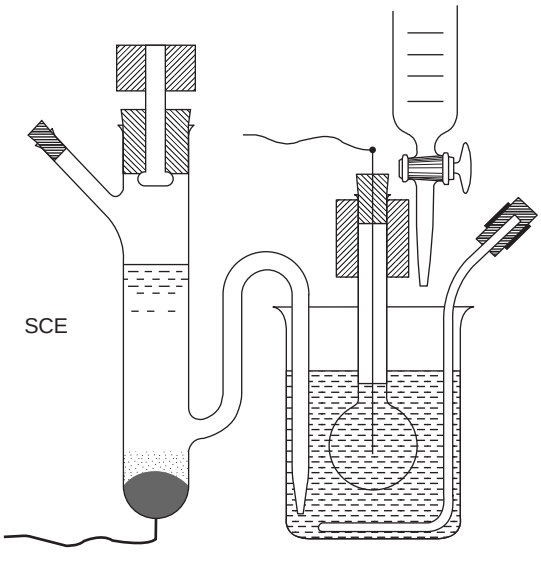

Figure 5 Experimental set-up for the pioneering studies of Klemensiewicz. SCE = saturated calomel electrode.

The effect of temperature was also studied, as was the effect of the glass on the potential generation. Two other easily fusible glasses, one a blue-colored, silicic acid-poor and sodium-rich glass and the other a colorless, silicic acid-rich and alkali-poor glass, were studied. The freshly blown blue-colored glass gave a relative error of −14%, but after steam treatment the error increased to −51%. The error for the colorless glass was even worse, being −57% for freshly blown and −71% for steam treated. When difficult to fuse Jena glass was used, no potential difference developed. Although the glass studies were qualitative at best, since the compositions were not measured, these results opened the way to what soon became an explosion in the development of ISEs.

Around this same time, Sørenson proposed the pH scale for acidity measurements[37] and the glass membrane acidity electrode ultimately came to be known as the pH electrode. From 1920 to 1935, quantitative studies on the effect of glass composition on the potential generation for the acidity electrode proliferated.[38] The compositions of Thuringian glass and the Jena glasses were determined.[39] This period was capped by the development of the first commercial pH glass manufactured by Corning as Corning 015 glass.[40] Because this glass composition was not ideal for the entire range of the pH scale, studies on glass composition for pH electrodes continued, with the results for over 500 different compositions appearing in the literature.[41] The sodium error results seem to have led some researchers to search for glass compositions that would give glass membranes sensitive to other ions. Thus, in 1934, Lengyel and Blum[42] reported on a glass

membrane electrode that gave a Nernstian response to sodium ions. Subsequent glass composition studies by Eisenman,[43] reported in 1957, led to commercial glass membrane electrodes for sodium and potassium ions.

It is interesting that materials other than glass were reported as active materials for potentiometric ion sensors in the period from 1930 to 1960. These included inorganic sparingly soluble salts, surface-modified collodion, and cation-exchange resins for ions such as Ca^{2+}, Na^+, K^+, NH_4^+, Mg^{2+}, Ag^+, Cl^-, Br^-, I^-, Ba^{2+}, ClO_4^-, NO_3^-, and SO_4^{2-}.[44] At the time, this research seems to have been ignored, largely because of irreproducible results and lack of a Nernstian response. Glass membrane studies continued, and additional glass membrane electrodes were developed for Li^+, Ag^+, NH_4^+, Rb^+ and Cs^+. In the 1960s, studies by Pungor et al.[45] led to the first nonglass ISEs to give thermodynamically reversible responses – sparingly soluble inorganic salts in silicone rubber. These were commercialized and marketed by Radelkis in Hungary in 1965. Up to this point, cation ISEs with acceptable responses existed only for univalent cations, so the next breakthrough was a very big one – the development in 1966 of the Ca^{2+} liquid membrane ISE by Ross.[45] As a result of this development and the development of the solid-state F^- ISE by Ross and Frant,[45] a small company in the USA, Orion Research, experienced tremendous growth.

Another development in 1966 was the introduction of the first neutral carrier-based membrane ISE by Stefanac and Simon.[46,47] The electrode was a K^+ ISE based on the macrotetrolide homologs of the antibiotic nonactin. The membrane supports studied were sintered-glass disks, Millipore filter disks, nylon mesh, polyethylene films and gel formers. In 1967, Shatkay described a neutral carrier-based polymeric ISE for calcium based on poly(vinyl chloride) (PVC), tributyl phosphate and theonyltrifluoroacetone.[48] This period of time is regarded by many as the incipient point for what Ruzicka[49] calls the ''golden age'' for ISEs, with obvious reference to ISEs which target inorganic cations and anions. In Figure 6, a timeline periodic table for ISE membranes for inorganic ions is given. The year of the first literature report is given, although many years often elapsed before a useful ISE was fabricated for an ion. Of the 79 ions in the table, membranes for 27 of them were reported from 1960 to 1969 and membranes for 24 of them were reported from 1970 to 1979.

Most research on membranes for inorganic ions now involves developing more selective membranes for ions already shown in Figure 6. This is reflected by the fact that only three ''brand new'' ISEs for inorganic ions were developed in the period from 1990 to 1999. ISEs for direct potentiometry are available commercially for only 24 of these ions, as shown in Figure 7. These are electrodes for which the analyte determinations have a high volume demand, and for which ISE potentiometry is competitive with other methods of analysis.

By contrast, the development of ISEs for organic and bio-organic analysis is now experiencing the kind of growth that was observed for inorganic ions from 1960 to 1979. This era seems to have commenced with the development in 1969 of the urea ISE by Guilbault and Montalvo,[50] the development in 1972 of coated-wire ISEs for amino acid anions and salicylate by James, Carmack and Freiser[51] and the development in 1972 of PVC matrix membrane ISEs for large onium ions by Scholer and Simon.[52] Electrodes have now been reported for glucose, L- and D-amino acids and their methyl esters, 1-phenylethylamine, 1-(1-naphthyl)ethylamine, ephedrine, norephedrine, pseudoephedrine, amphetamine, propanolol, α-amino-ε-caprolactam, amino acid amides, benzylamine, alkylamines, dopamine, procaine, prilocaine, lidocaine, bupivacaine, lignocaine, diquat, paraquat, acetylcholine, albumin, amygdalin, pencillin, protamine, catechol, cholesterol, creatinine, uric acid, 5′-AMP, glutamine, glutamate, pyruvate, vitamin B_6, guanine, guanidine, amprolium, heparin, methotrexate, mexiletine, metformin, phenformin, clotrimazole, bifonazole, human immunoglobulin G, organohosphate nerve agents, cationic surfactants, anionic surfactants, and organosulfonates.[25,50–63] Only the glucose, urea and creatinine electrodes are available commercially.

The first gas-sensing ISE was the carbon dioxide gas electrode, conceptualized in 1957 by Stow, Baer and Randall.[64] After a 10-year gap, the ammonia gas-sensing electrode was developed, and by 1976 electrodes had been developed additionally for NO_x, SO_2, H_2S and CH_3COOH.[64] As shown in Figure 7, 10 gas sensing ISEs are now available commercially. Further development in this area is not likely, because acoustic wave devices with higher sensitivities, faster responses, and simpler construction are likely to become the sensors of choice for gas analysis.[65]

3 THEORY OF ELECTRODE RESPONSE

3.1 General

All of the ISEs share in common that the component of the measured signal which is related to the activity of the analyte to be determined is the potential difference across interface 1 in Figure 8(a) or (b). Figure 8(a) is the configuration for the primary ISEs. Figure 8(b) is the configuration for the metal contact or all-solid-state ISEs. The compound or multiple membrane ISE has an

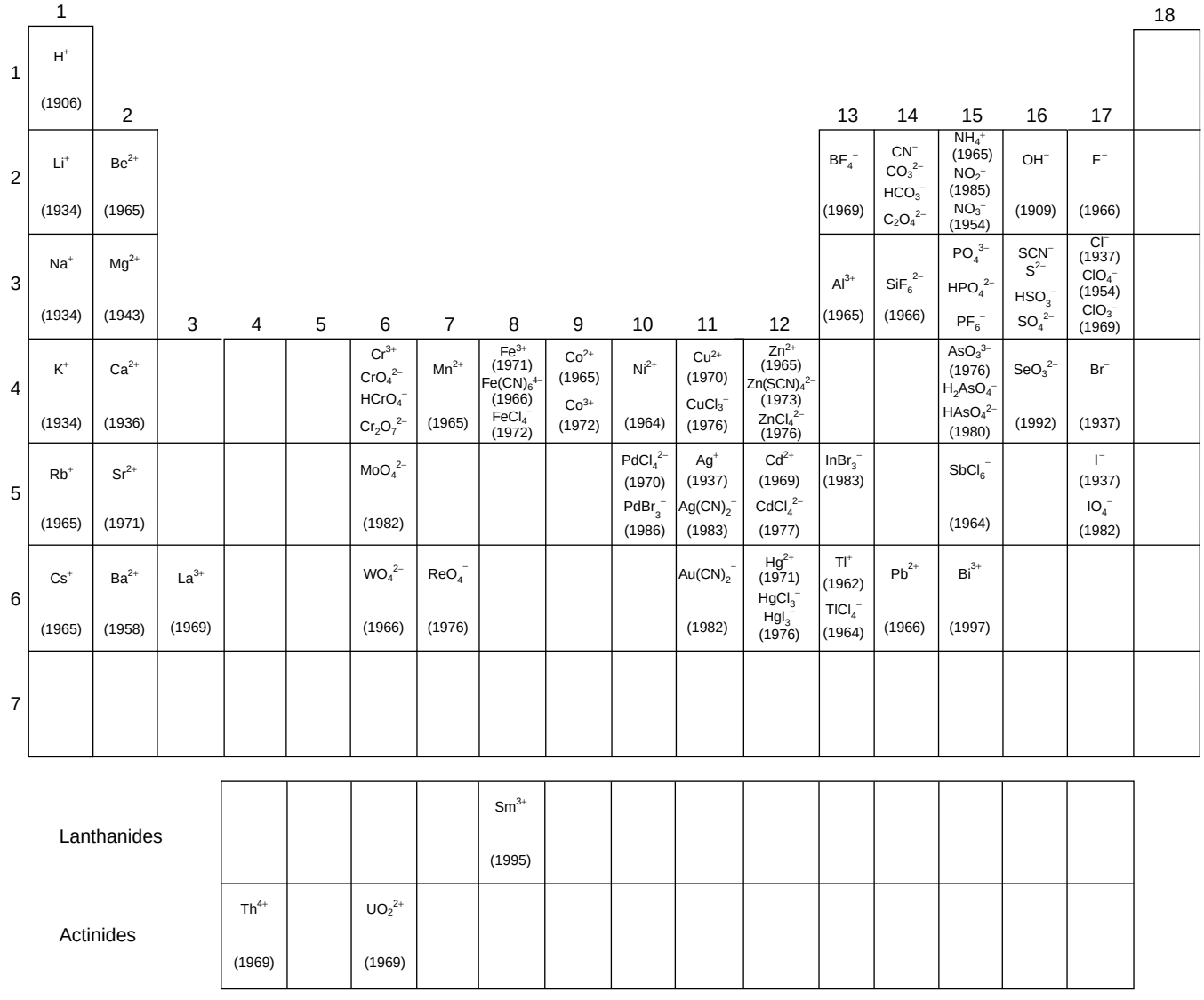

Figure 6 Periodic table profile for the historical development of ISE membranes for inorganic ions.

intermediate compartment which controls the activity of analyte in the external solution of an electrode of type shown in Figure 8.

The formation of a potential difference across the interface between two different phases is a general process. The potential difference can be written as the sum of two or more potential differences resulting from essentially independent processes occurring at the interface. Let us assume that the membrane is a perfectly ordered solid which fills all space (infinite three-dimensional lattice). If the solid is cut along the *yz* plane, leaving a semi-infinite three-dimensional solid with a face in contact with a vacuum, then surface states whose wave functions penetrate a short distance into the vacuum are created. The surface states are not empty. Since electron density is proportional to the product of a wave function and its complex conjugate, this means that some electrons have spilled across the solid surface and into the vacuum. However, these electrons belong to the solid, and not the vacuum (i.e. ionization has not occurred). Thus, all solids in a vacuum have a surface dipole barrier at the interface between the surface and the vacuum.[66] The magnitude of the surface dipole barrier would be largest for a free electron metal and smallest for an insulator. In the case of an ionic solid, surface defects are the source of the equilibrium space charge potential.[67] If a semi-infinite electrolyte solution can be placed in contact with a vacuum, even in this case a surface dipole barrier would be formed. It would be due to the reorientation of water molecules to give a net dipole moment at the surface of the solution.[68] One would expect that more water molecules

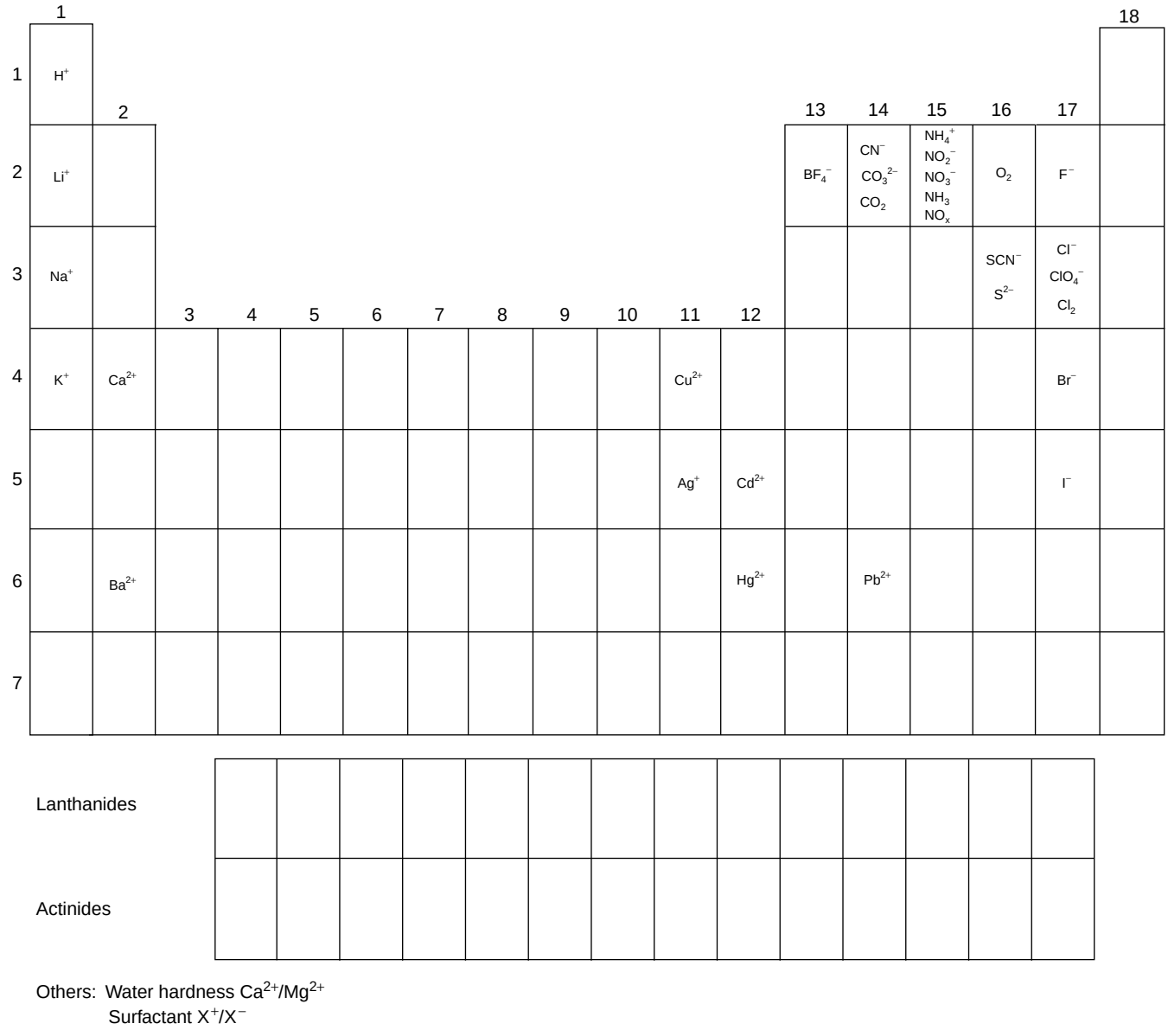

Figure 7 Periodic table diagrams for commercially available inorganic ions.

are orientated so that a lone pair on the oxygen points toward the vacuum than are oriented with the hydrogens pointed toward the vacuum. Thus the picture for the surface dipole barrier is analogous to that for the solid. These surface dipole potentials are often called the ''chi potentials''.[67]

Now, if the semi-infinite solid and the semi-infinite electrolyte solution are brought together and no transfer of charge occurs between them, the two phases remain neutral, but the interface between them is electrified owing to the difference between their surface potentials. The potential difference will be smaller than the numerical difference between the two vacuum referenced values, because there will be a readjustment of the electron density on the solid side, the distribution of defects, the orientation of water dipoles, etc. Nevertheless, there will be a finite potential difference across the interface. However, because this potential difference across the interface is not due to any potential-determining ion, it cannot provide a useful analytical signal. An analytically useful signal can be obtained only if a charge-transfer process can occur which results in a nonzero charge on each phase across the interface. Charge transfer may be caused by or influenced by concentration gradients, potential gradients, temperature gradients and/or pressure gradients.

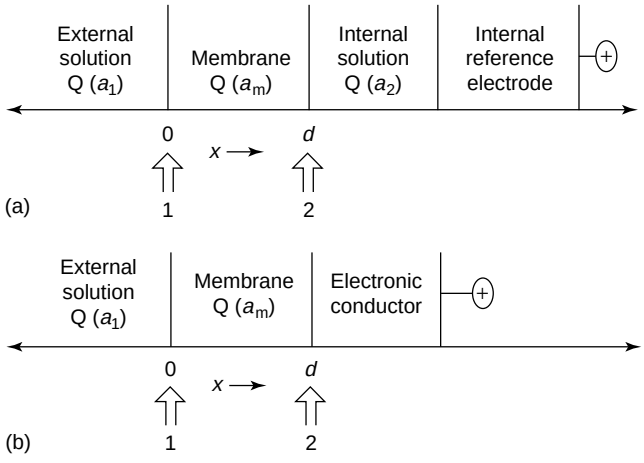

Figure 8 Simplest interface diagram for the physical interface locations of (a) primary ISEs and (b) metal contact or all-solid-state ISEs.

It is usual to classify charge-transfer processes as either Faradaic or nonFaradaic. A Faradaic process involves an oxidation or a reduction reaction. This type of process will alter the nature of certain ions or molecules on both sides of the interface. It has been shown that this type of process is frequently the basis of chemical interference for some type of membranes, but it does not usually lead to a potential difference across an interface that can be related to the activity of a potential-determining ion. One exception is the homogeneous Cu(II) membrane ISE, which will be discussed later. Thus, the relevant charge-transfer processes are mostly nonFaradaic. Examples of nonFaradaic chemical or physical processes that can lead to charge transfer are ion adsorption/desorption, complex formation, ion diffusion, ion extraction, and ion exchange.

A complete description of a membrane system involves the determination of various concentrations and potentials in both space and time. This requires that a set of partial differential equations for pertinent fluxes and a set of conservation equations be solved for some set of initial and boundary conditions. For the most general solution, only those initial and boundary conditions necessary to prevent the violation of physical laws may be specified. No such solution exists. Instead, solutions have been obtained for a number of special cases. In all of these special cases, temperature is kept constant.

We cannot examine all of these special cases here. Instead, it is instructive to examine the thermodynamic treatment for a number of simple cases in which the material properties of the nonaqueous phase are different. In every case, we consider that the aqueous phase and the nonaqueous phase have at least a univalent cation in common. Using in part the notation of Koryta,[69] we consider first the interface for zero current flow in Equation (14):

$$\overset{\alpha}{P^+X^-} \;\big|\; \overset{\beta}{P^+e^-} \tag{14}$$

where α is an aqueous solution of the strong electrolyte P^+X^- and β is a free electron metal P with ion cores P^+ and free electrons e^-. An example would be silver metal in contact with $AgNO_{3(aq)}$. Since X^- is not potential determining, on the α side of the interface, we have (Equation 15)[70]

$$\bar{\mu}_{P^+}(\alpha) = \mu_{P^+}(\alpha) + \quad \phi_\alpha \tag{15}$$

where $\bar{\mu}$ is the electrochemical potential, μ is the chemical potential and ϕ is the Galvani potential. On the metal side of the interface, we have (Equation 16)

$$\bar{\mu}_P = \bar{\mu}_{P^+}(\beta) + \bar{\mu}_{e^-}(\beta) = \mu_P \tag{16}$$

where

$$\bar{\mu}_{e^-}(\beta) = \mu_{e^-}(\beta) - \quad \phi_\beta \tag{17}$$

Substituting and rearranging gives Equation (18):

$$(\phi_\beta - \phi_\alpha) = \mu_{P^+}(\alpha) + \mu_{e^-}(\beta) - \mu_P \tag{18}$$

The chemical potentials may be written in terms of the standard chemical potential and the activity. For example (Equation 19):

$$\mu_{P^+}(\alpha) = \mu^0_{P^+}(\alpha) + RT\ln[a_{P^+}(\alpha)] \tag{19}$$

Making this substitution, and given that $a_{e^-}(\beta) = a_p = 1$, we obtain Equation (20):

$$\phi_\beta - \phi_\alpha = K + RT\ln[a_{P^+}(\alpha)] \tag{20}$$

where K is given by

$$K = \frac{[\mu^0_{P^+}(\alpha) + \mu^0_{e^-}(\beta) - \mu^0_P]}{} \tag{21}$$

Next, consider the case in Equation (22):

$$\overset{\alpha}{P^+X^-} \;\big|\; \overset{\beta}{P^+X^-} \tag{22}$$

where α is an aqueous phase containing the electrolyte P^+X^- and β is an immiscible, nonaqueous phase containing the same electrolyte. At equilibrium, we have (Equation 23):

$$(\phi_\beta - \phi_\alpha) = [\mu_{P^+}(\alpha) - \mu_{P^+}(\beta)] - [\mu_{X^-}(\alpha) - \mu_{X^-}(\beta)] \tag{23}$$

If there is differential transport of P^+ and X^-, then a Galvani potential difference results, but both ions are potential determining. How do we obtain an interface in which a single ion is potential determining? The result for

the aqueous solution/metal phase gives the clue. We see that X^- cannot cross into the metal side of the interface. Likewise, electrons cannot move from the metal into the solution phase. Therefore, if the negative ions in α are excluded from β, and the negative ions in β are excluded from α, only the cations will determine the Galvani potential difference. This would correspond to the case in Equation (24):

$$\overset{\alpha}{P^+X^-} \;\mid\; \overset{\beta}{P^+Y^-} \tag{24}$$

where X^- is hydrophilic and Y^- is lipophilic. This case represents the simplest model for the polymeric membrane-type ISEs. With the anion exclusions, Equation (23) becomes Equation (25):

$$(\phi_\beta - \phi_\alpha) = [\mu_{P^+}(\alpha) - \mu_{P^+}(\beta)] \tag{25}$$

$$(\phi_\beta - \phi_\alpha) = \mu^0_{P^+}(\alpha) - \mu^0_{P^+}(\beta) + RT \ln \frac{a_{P^+}(\alpha)}{a_{P^+}(\beta)} \tag{26}$$

If $a_{P^+}(\beta)$ is constant, an equation similar in form to Equation (20) is obtained, with K given by Equation (27):

$$K = \underline{\mu^0_{P^+}(\alpha) - \mu^0_{P^+}(\beta) + RT \ln[a_{P^+}(\beta)]} \tag{27}$$

Notice that the interface in Equation (24) could just as easily represent an aqueous phase containing the electrolyte P^+X^- in contact with a sparingly soluble salt P^+Y^-. Clearly, at equilibrium there will be some Y^- in α, but owing to the common ion effect, its activity will be very small over some range of activities for P^+. The interface can be approximated as one in which Y^- is essentially confined to β, and the equation for the Galvani potential difference across the interface has the same form as Equation (25). Again, Equation (20) gives the final form of the mathematical expression for the Galvani potential difference. Equation (20) is a form of the Nernst equation for an interface at equilibrium. The Galvani potential difference across an interface cannot be measured. What can be measured is the potential difference across a cell. By coupling the indicator electrode to a reference electrode, we obtain an electrochemical cell whose potential difference is the sum of the potential differences across many interfaces. If all interfaces except for the interface between the external membrane surface and the sample solution are constant, then $E_{cell} = (\phi_\beta - \phi_\alpha) + K'$. Furthermore, if we write $E_{cell} = E_{ind} - E_{ref}$, the potential difference between the indicator electrode (metal or ISE in the cases above) and the reference electrode, then the electrode potential for the indicator may be written in its usual form (Equation 28):

$$E_{IND} = K_{electrode} + \frac{RT}{} \ln[a_{P^+}(\alpha)] \tag{28}$$

where any liquid junction potentials in the cell are assumed to be negligible. Converting to common logarithms and generalizing to a potential-determining ion with charge z, sign and magnitude, Equation (28) becomes Equation (29):

$$E_{IND} = K_{electrode} + \frac{2.303RT}{z} \log[a_Q] \tag{29}$$

where Q is the target ion shown in Figure 8. At 25 °C, the magnitude of the multiplicand of the logarithm terms is 59.2, 29.6, and 19.7 mV for $z = \pm 1$, ± 2, and ± 3, respectively. In practice, these ideal values are frequently not observed for ISEs, so it is common practice to add a so-called efficiency factor, β, to that term and to treat the multiplicand as a constant which must be measured. Thus, most textbooks give Equation (30) for the electrode response of an ISE:

$$\begin{aligned} E_{ISE} &= K_{electrode} + \frac{\beta \times 2.303RT}{z} \log[a_Q] \\ &= K_{electrode} + S \log[a_Q] \end{aligned} \tag{30}$$

If Q can be transported through the membrane, i.e. the membrane is an ionic conductor for Q, then there may be a membrane diffusion potential difference across the bulk membrane.[71] Frequently, it is the case that the ion transported through the membrane is not the potential-determining ion responsible for the potentials generated at the boundaries. Even in that case, a membrane diffusion potential difference across the bulk membrane may be generated. The transport equations have been solved for several applicable cases.[71–73] The mathematical formulation is at best quasi-thermodynamic. Fortunately, it has been shown that the diffusion potential difference across the bulk membrane is negligible for most practical cases.[25,74,75]

The time-variable properties of ISEs determine the upper limit to the rate of sampling. They are particularly important in flow analysis applications or in monitoring fast processes. These properties are determined by subjecting an ISE to activity steps. The time-variable properties of ISEs are referred to as their dynamic characteristics. In Figure 9(a) and (b) are shown the responses of a Corning I^-/CN^- ISE to forward and reverse iodide steps.[76]

The characterization of the dynamic response behavior of ISEs seems to have commenced in the mid-1960s,[77] but progress has been slow owing to the complexity of the membrane response mechanisms. The figure of merit for the dynamic characteristics of an ISE is the response time. This has variously been defined in the literature as $t_{1/2}$, t_{90}, t_{95}, or $t_{99.5}$ (times to reach 50%, 90%, 95%, or 99.5% of the total potential change) for a 10-fold activity change. The IUPAC recommendation

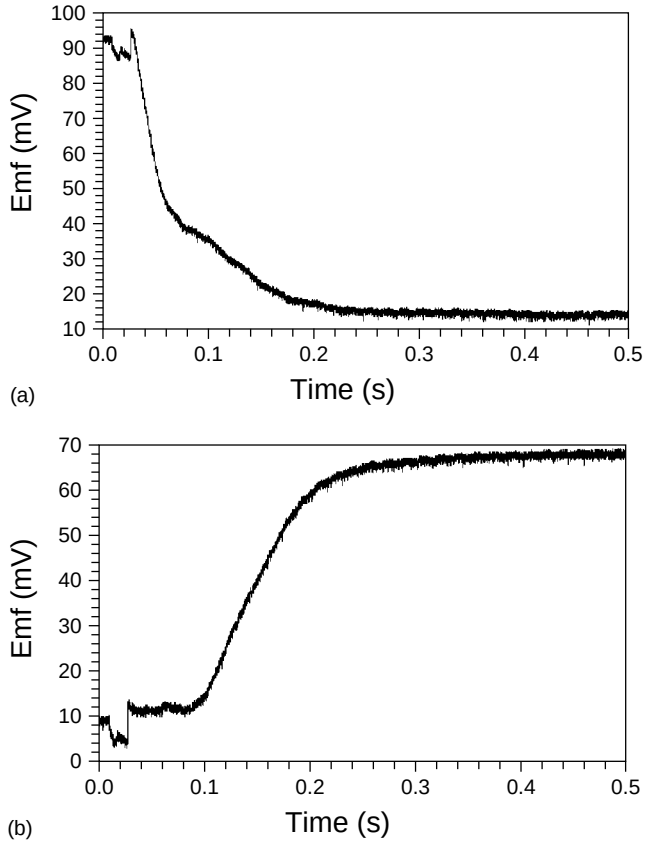

Figure 9 Dynamic responses of a Corning iodide/cyanide ISE to (a) forward and (b) reverse iodide steps. (a) 0.001 to 0.01 M I^-, ionic strength = 0.15 M; (b) 0.01 to 0.001 M I^{-1}, ionic strength = 0.15 M.

for the definition of this parameter is the following: "The time which elapses between the instant when an ISE and a reference electrode (ISE cell) are brought into contact with a sample solution (or at which the activity of the ion of interest in a solution is changed) and the first instance at which the electromotive force (emf)/time slope ($\Delta E/\Delta t$) becomes equal to a limiting value selected on the basis of the experimental conditions and/or requirements concerning the accuracy (e.g. 0.6 mV min^{-1}). . . . In clinical applications (the physiological activity range corresponds to a small emf span), a smaller slope, e.g. 0.1 mV min^{-1}, may be chosen, provided the standard deviation of the response is less than the required slope."[1] The response time defined as the time from which the activity step is initiated until the slope becomes zero is identical to t_{100}.

Both of these definitions are impractical, because many electrodes never reach a constant potential after the step. Unfortunately, it is not usually possible to convert from one definition to another, because the mathematical equation for the potential as a function of time must be known. There does not seem to be a single equation that adequately represents the time response for all ISEs. Berube et al.[78] examined eight different models from the literature for a study on the homogeneous, crystalline iodide ISE. If the measured time response is not corrected for drift, the determined response time will not be accurate. Drift is defined as "the slow non-random change with time in the emf of an ion-selective electrode cell assembly maintained in a solution of constant composition and temperature".[1] The drift may be measured on the front end before the activity step, and we have no choice but to assume that it does not change as a result of the step. If precise response times need to be known, then independent studies on drift as a function of the activity of the potential-determining ion should be performed. It is necessary to define the initial and final solution conditions for the step, since the response curves depend on these.

Finally, this subsection concludes with a discussion of the general treatment of electrode interferences. The 1994 IUPAC recommendations specify that the general expression for electrode response is the modified Nikolsky–Eisenman equation, given by Equation (31):[1]

$$E_{\mathrm{ISE}} = K_{\mathrm{electrode}} + \frac{2.303RT}{z}\log\ \left(a_{\mathrm{A}} + \sum K^{\mathrm{pot}}_{\mathrm{A,B}}(a_{\mathrm{B}})^{z_{\mathrm{B}}}\right) \tag{31}$$

where $K^{\mathrm{pot}}_{\mathrm{A,B}}$ are selectivity coefficients.

Over the years, several different methods have been proposed for the evaluation of selectivity coefficients. The IUPAC recommendation is that the fixed interference method is preferred and the separate solution method is acceptable, but less desirable. When the analyte ion and the electrode interferences have the same charge (sign and magnitude), and when the response is Nernstian for solutions of each type of ion, then the electrode interference is adequately described by the Nikolsky–Eisenman formalism. This formalism fails when either of these conditions is not met. Bakker et al.[79] examined critically the Nikolsky–Eisenman formalism for polymer membrane-based ISEs, demonstrated its general inadequacy and proposed a new formalism, which they showed corresponded to the matched potential method proposed earlier by Gadzekpo and Christian.[80] Umezawa et al.[81] critically evaluated the validity of Equation (28) for two crystalline membrane ISEs and three liquid membrane ISEs. They also recommended that selectivity coefficients be obtained using the matched potential method. More recently, Nägele et al.[82] developed a general formalism for electrode interferences for polymer membrane-based ISEs. Thus, the formalism for the response of polymeric membrane ISEs in the presence of electrode interferences has been completely and rigorously established. Since it is based on ion extraction/exchange equilibria, the formalism is not necessarily a general one applicable

to all ISEs. These results will be discussed in more detail under the specific heading of noncrystalline electrodes.

By using statistical mechanics and equations obtained from nonequilibrium thermodynamics, Hall[83] developed a general limiting expression valid for electrode interferences whose charges differ in both sign and magnitude. In this treatment, the total potential difference across the membrane is not partitioned into phase boundary potentials and a diffusion potential. With reference to Figure 8(a), Hall writes the potential difference across the membrane as $\phi^{(1)} - \phi^{(2)}$, which can be written as Equation (32):

$$\phi^{(1)} - \phi^{(2)} = \frac{\bar{\mu}_Q^{(1)} - \bar{\mu}_Q^{(2)}}{z} - \frac{\mu_Q^{(1)} - \mu_Q^{(2)}}{z} \qquad (32)$$

where z is the charge on Q. The first term in Equation (32) is the deviation of the membrane response from Nernstian behavior; it represents the contribution to the cell emf due to interference. In the equilibrium thermodynamic treatment, this term is zero. Considering both ionic and nonionic interferences, Hall developed a general expression for the interference term. Unfortunately, from a practical perspective, the mathematical expression is not useful. If it were possible to correlate these results with those obtained for the polymeric membranes, then it might be that modifications of those equations could be used for ISEs with other types of membranes.

3.2 Specific

The electrical properties of these membranes determine to a great extent the nature of their response to potential-determining ions. Both the electronic properties and the electrolytic properties must be considered. It is common knowledge that electronic materials may be classified as conductors, semiconductors, or insulators, based on the magnitude of their conductivities. It is not as widely appreciated that a similar classification scheme exists for electrolytic materials. They may be fast-ion conductors, electrolytic semiconductors, or electrolytic insulators. The boundaries between classes are determined by the magnitude of their ionic conductivities. In Table 1, we show these two classification schemes.[84]

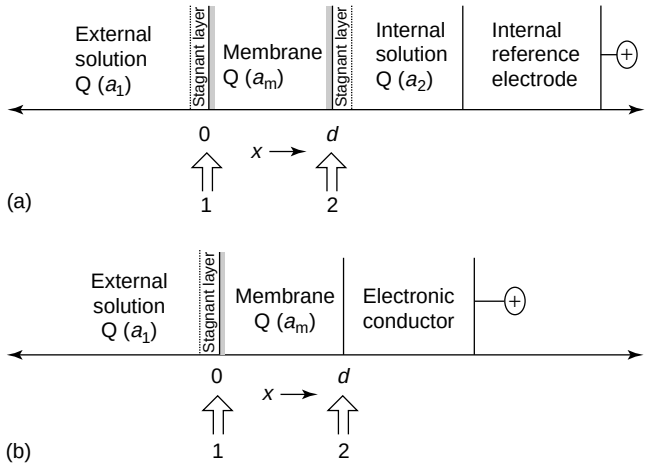

Figure 10 The segmented membrane model for (a) primary ISE and (b) a metal contact or all-solid-state ISE. In the latter case, the space-charge region in the membrane at the membrane/electronic conductor interface is not shown.

3.2.1 Primary Ion-selective Electrodes

3.2.1.1 Crystalline Electrodes The active materials in these electrodes are sparingly soluble, inorganic salts. They vary widely in their properties. Commercial electrodes with membranes in this category are the fluoride, chloride, bromide, iodide, sulfide, cyanide, thiocyanate, cadmium ion, cupric ion, lead ion, silver ion and mercuric ion ISEs. The fluoride ISE is the most important member of this group. Direct potentiometry for the in vitro determination of fluoride is the method of choice. Official methods for the determination of fluoride by use of the fluoride ISE appear in the *Official Methods of Analysis* Manual of the Association of Official Analytical Chemists, and also manuals of analysis published by the American Society for Testing and Materials and the Environmental Protection Agency. The active material of the fluoride ISE is lanthanum fluoride, a wide band gap electronic insulator, but an electrolytic semiconductor. Fluoride ion is the mobile charge carrier in this defect ionic conductor. Impedance spectroscopy[85] and X-ray photoelectron spectroscopy (XPS)[86] show that it forms a surface-hydrolyzed layer on contact with solution, hence a better model for this ISE is the segmented membrane model shown in Figure 10(a) and (b).

The solution in contact with the membrane may also be regarded as segmented into a bulk solution and a stagnant

Table 1 Electrical properties of materials

Type	Electronic	Electrolytic
Conductor	$10^2\ \Omega^{-1}\,cm^{-1} < \sigma_e$	$10^1\ \Omega^{-1}\,cm^{-1} < \sigma_I$
Semiconductor	$10^{-9}\ \Omega^{-1}\,cm^{-1} < \sigma_e < 10^2\ \Omega^{-1}\,cm^{-1}$	$10^2\ \Omega^{-1}\,cm^{-1} < \sigma_I < 10^1\ \Omega^{-1}\,cm^{-1}$
Insulator	$\sigma_e < 10^{-9}\ \Omega^{-1}\,cm^{-1}$	$\sigma_I < 10^{-9}\ \Omega^{-1}\,cm^{-1}$

layer in contact with the electrode. In a stirred solution, there is convection in the bulk, but no convection in the stagnant layer. Although a gel layer has been considered critical for the response of glass membrane electrodes, or the fluoride ISE, the gel layer may not be essential for the response, and is sometimes actually detrimental. The gel layer, described as $LaF_3 \cdot 3H_2O$, is produced by diffusion of water into the membrane; it reaches a thickness of 20 nm in less than 1 week. The dynamic response of the electrode is not significantly affected by the gel layer, but with increasing time, leaching of fluoride is observed, concomitant with the incorporation of hydroxide. The aged electrode shows a sluggish response to a decade fluoride activity step and the change in emf becomes significantly less than Nernstian. After 18 weeks of aging, the first 5 nm of the 20-nm-thick gel layer has the stoichiometry $LaF_{3-n}(OH)_n \cdot mH_2O$. Fortunately, the aging effects may be reverse by polishing. At room temperature, hydroxide is the only electrode interferent; the mechanism of interference has been shown to be ion exchange in the gel layer.[87] A very significant result of that study is that carbonate and phosphate, both of which form sparingly soluble salts with lanthanum, adsorbed on the membrane but did not affect the potential response at 20 °C, but at 60 °C the two ions did interfere with the potential response. At that temperature, they undergo ion exchange in the gel layer. There is some debate as to whether adsorption/desorption or ion exchange is the mechanism for potential generation of crystalline membrane ISEs. Experiments such as these may allow the debate to be resolved. The lanthanum fluoride membrane has not been used for the direct determination of lanthanum, so the membrane apparently does not give a useful potential response to this ion. The useful concentration range is from saturated to 10^{-6} M fluoride.

The other commercial crystalline membrane ISEs are based on sparingly soluble silver salts or a homogeneous mixture of silver sulfide and a sparingly soluble metal sulfide. The silver salts are mixed conductors, so that chemical interference from redox active species, in addition to electrode interferences, may adversely affect an analysis. For example, a chloride ISE may have silver chloride as its active material. Silver chloride is a mixed conductor; it is an electronic semiconductor and an electrolytic semiconductor (silver ion is the mobile ion). Redox interference is determined by the electronic band gap of the material. The band gap of the silver salts is large enough that redox interference is not usually a problem in practice.

The silver chloride membrane ISE differs from the fluoride ISE in other ways: the electrode does not have a gel layer and it gives a Nernstian response to both silver ions and chloride ions. An example of the response behavior is shown in Figure 11.

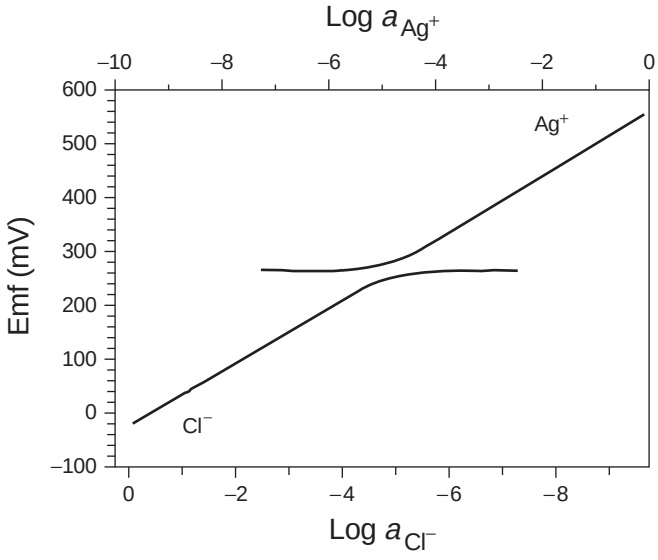

Figure 11 Response of an AgCl ISE to solutions of silver ions and solutions of chloride ions.

Note that the electrode gives a constant response whenever the activity of chloride or the activity of silver ion is less than the square root of the K_{sp} of silver chloride. By calling chloride ion a silver ion buffer and drawing the collinear line segments as a single straight line, the linear dynamic range for silver activity has been said to extend from 0.10 to 1.0×10^{-9} M. Mechanistically, these curves indicate that when silver ion is the dominant ion in solution, its interaction with the membrane determines the potential, and vice versa for chloride ion. It is well known that when solid silver chloride is present in a solution with excess silver ion (cation common ion effect), silver ion adsorbs on the surface of the solid. When the solid is in a solution with excess chloride ion (anion common ion effect), chloride ion adsorbs on the solid.

Adsorption/desorption equilibrium at a membrane–electrolyte interface will alter the distribution of defects in the space-charge layer of the membrane. For example, it has been shown that the space-charge potential of a silver chloride membrane in contact with an aqueous solution containing silver ion depends on pAg.[88,89] A diffusion-controlled transport model with $D \approx 5 \times 10^{-8}$ cm^2 s^{-1} gave a good fit to experimental conductance and capacitance data obtained on a frequency range of 30–10^5 Hz. Using data from Hoyen et al.,[89] the thickness of the space-charge layer is calculated to be 25 nm. An estimate of the time constant needed to achieve a new equilibrium in the space-charge region after changing the activity of silver ion or chloride ion at the membrane surface is given by Equation (33):

$$\tau = \frac{4\delta}{\pi^2 D} \tag{33}$$

for first-order diffusion transport, where δ is the space-charge layer thickness. A value of 64 µs is obtained, which corresponds to a t_{95} response time of 192 µs ($t_{95} = 3\tau$). Rangarajan and Rechnitz[90] found $t_{95} \approx$ 200 ms for an Orion-type 94-17-00 electrode subjected to a 10-fold concentration step-up in an electrode flow chamber. Pungor[91] reported a response time of 20 ms, which is probably τ, using a switched wall-jet measuring cell arrangement. The experimental response times include diffusion across the Nernst layer plus whatever process gives rise to the interfacial potential. Adsorption/desorption leading to modulation of the space-charge defect distribution in the space-charge layer is fast enough to account for these experimental results.

Buck[84] found that the impedance plots for silver chloride were independent of pCl. This would seem to suggest that chloride adsorption did not modulate the space-charge distribution, so that adsorption alone must be responsible for the interfacial potential at the membrane. In essence, the potential is changed only at the top monolayer of the membrane, which is equivalent to causing a jump discontinuity in the space-charge potential distribution. This is consistent with the coupled diffusion/adsorption model used by Berube et al.[78] to interprete dynamic response data for the iodide ISE perturbed by iodide steps. Here, diffusion refers to the transport that occurs through the Nernst layer. The bromide ISE based on silver bromide is in all respects similar to the chloride and iodide ISEs.

Silver sulfide has a much smaller band gap than the silver halides. At room temperature, the most stable crystallographic form is β-Ag_2S, which has a silver ion conductivity of $3.60 \times 10^{-4}\,\Omega^{-1}\,cm^{-1}$ and an electronic conductivity of only $4.00 \times 10^{-6}\,\Omega^{-1}\,cm^{-1}$. Hence, its electrolytic properties dominate, and it is this form that is useful as an ion sensor. Not only is it a sensor for silver ions and sulfide ions, but it is also a component of the homogeneous mixture membranes of the cadmium ISE (CdS/Ag_2S), the lead ISE (PbS/Ag_2S), the copper ISE (Cu_2S/Ag_2S and CuS/Ag_2S), the mercury(II) ISE (HgS/Ag_2S), and the iodide/cyanide ISE (AgI/Ag_2S). Other metal sulfide/silver sulfide homogeneous mixture membranes have been studied, but they have not been commercialized.

Like lanthanum fluoride, the surface of silver sulfide becomes hydrated on soaking in water.[92] However, the concentration profile for water shows an approximate exponential decay, $e^{-\lambda d}$, with depth d into the membrane. The reciprocal of λ is ~2 nm, and at $d \approx 8$ nm, the boundary with dry silver sulfide is reached. The interfacial potential is believed to be generated by a space-charge mechanism.[92] The divalent sulfides are electronic semiconductors but electrolytic insulators. In the homogeneous mixtures, they provide absorption centers for the divalent cations, while the silver sulfide provides the electrolytic conductivity. However, in an elegant experiment, Uosaki et al.[93] showed that a single crystal of CdS functions as an ISE. The configuration was one of all solid state, and the external surface was either an (0001) Cd face or an (0001) S face. The (0001) Cd face showed a Nernstian response to HS^- concentration (Na_2S solutions buffered at pH 9) and the (0001) S face showed a Nernstian response to Cd^{2+} concentration. The results have been plotted in Figure 12 using the same format as Figure 11.

The equilibrium expression allowing the concentrations of Cd^{2+} and HS^- to be related is given by Equation (34):

$$CdS_{(s)} + H_2O \rightleftharpoons Cd^{2+} + HS^- + OH^- \qquad (34)$$

with $K = 6.4 \times 10^{-28}$. Since pH = 9.0, $C_{cd} \times C_{HS} = 6.4 \times 10^{-19}$, where activity effects have been ignored. The plot is analogous to those obtained for silver iodide and silver sulfide.[94] Most significantly, the slopes of the emf versus log C plots were identical with the slopes of the flat band potential versus log C. This is, perhaps, the strongest evidence yet for an adsorption mechanism for interfacial potential generation. As discussed in that paper, a general equation for the variation of the flat band potential with activity of the adsorbate, assuming a Frumkin adsorption isotherm, reduces to the Nernst equation when there are a large number of adsorption sites and there is no interaction between adsorbates. Because repulsion will

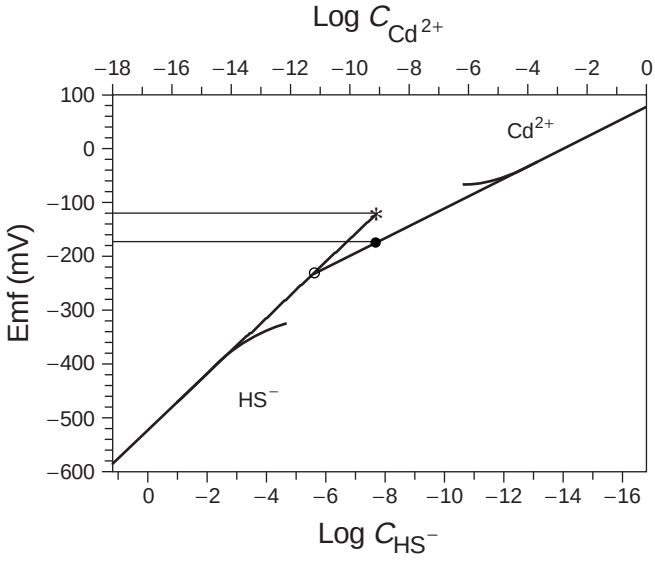

Figure 12 Response of a single-crystal cadmium sulfide ISE to solutions of cadmium ions and solutions of hydrogen sulfide ions. (○) Point of intersection of the extended linear regions; (∗) point on extended hydrogen sulfide line where the hydrogen sulfide concentration equals 8×10^{-10} M; (●) point on extended cadmium line where the cadmium concentration equals 8×10^{-10} M.

keep adsorbates from approaching close to each other on the surface, this approximation will almost always be valid.

The electrode interferences for these ISEs are easy to predict. Any cation or anion which forms a sparingly soluble salt with the counterion of the membrane is an electrode interference. A metathetical reaction, which alters the membrane surface composition and modifies the electrode response, will occur. The selectivity order may be predicted by calculating the equilibrium constant of the metathetical reaction using pertinent activity product constants. Selectivity will be of the order of the K values: the larger is K, the more serious is the interference by the ion. For quantitative work the selectivity coefficients should be measured.

These electrodes are also subject to numerous chemical interferences. The lead, cadmium and copper ISEs are subject to oxidation or reduction by sufficiently strong redox-active species. If the band gap is small, then photodegradation may also occur. Photodegradation may activate oxidation or reduction reactions which are energetically unfavorable in the absence of light. It can also catalyze the degradation of chemical interferences. All of the cations in these membranes form complex ions with some ligands. These are usually charged species, and therefore have a high solubility in aqueous media. The net result is that etching or leaching of the membrane will occur. Mild etching will cause so-called sub-Nernstian responses and longer response times. Severe etching may cause loss of response. In many cases, the responses of the electrodes may be restored by mechanical polishing. However, since membrane material is lost with each polish, the lifetime of the electrode will be significantly diminished. It is best to avoid exposing the electrodes to solutions containing known interferences. The interference problem has somewhat limited the use of these electrodes as detectors for real-time analysis of complex samples by methods such as flow injection analysis. Taken as a set, the quantitation detection limits of these electrodes range from 10^{-8} to 5×10^{-5} M, with 6×10^{-6} M as the median value. The upper quantitation detection limits range from 10^{-2} M to saturation, with 1.0 M as the median value.

3.2.1.2 Noncrystalline Electrodes The glass membrane electrodes and the electrodes with mobile charged sites, the various ionophore-containing polymer membranes, belong to this class. There are many types of inorganic glasses, but the silicate glasses are the only ones used to make commercial ion-sensing membranes. The most important ISEs of this type are the pH electrode and the sodium ion electrode. Silicate glass is formed from SiO_2, a network former, and one or more metal oxides, network modifiers. For example, in a sodium silicate glass, some of the oxygen atoms are bonded to two silicon atoms (bridging oxygen), other oxygen atoms are bonded to only one silicon atom [nonbridging oxygen (NBO) atoms, which carry a charge of -1], and sodium ions are either electrostatically bonded to NBOs or isolated in percolation channels. Percolation channels are tunnels through which alkali metal ions may move. Some network modifiers, e.g. aluminum oxide, give rise to substitutional impurities in the network itself, e.g. an aluminum substitutes for a silicon. When the valence of the substitutional impurity is not the same as that of silicon, a second network modifier, an alkali or alkaline earth metal oxide, must be used to provide for charge compensation.

Although silica glass membrane electrodes have been studied since the beginning of this century, it is easy to understand why their behavior is still being characterized. The properties of these materials are not monolithic; they depend strongly on the chemical composition of the dry glass, on the history of its preparation, and on its state of hydration during utilization. Depending on its composition, a dry silicate glass can be an electronic insulator and an electrolytic semiconductor or an electronic insulator and an electrolytic insulator. For example, at 27 °C NAS 25-25 glass has a conductivity of $1.0 \times 10^{-8}\ \Omega^{-1}\ cm^{-1}$ and is an electrolytic semiconductor, while Corning 015 glass has a conductivity of $8.6 \times 10^{-12}\ \Omega^{-1}\ cm^{-1}$ and is an electrolytic insulator. The former types might be expected to exhibit some of the behavior that we have observed for crystalline membrane ISEs. In the dry state, the latter types can only behave as capacitative devices. However, sufficient hydration of a dry glass can cause it to undergo a transition from an electrolytic insulator to an electrolytic semiconductor. Thus, conflicting reports on the behavior of glasses with the same nominal chemical composition can probably be traced to a failure to consider their preparation and utilization histories.

All glass membrane ISEs respond to monovalent cations and NH_4^+, and to a lesser extent to divalent cations. The dry glass chemical composition determines which ion gives the greatest response, i.e. it defines the target, and also defines the selectivity order for electrode interferences.

In Table 2, some important glass compositions which are used as membranes to target H^+, Na^+ and K^+ are shown. Note that a gel layer forms on Corning 015 glass when it comes in contact with water or aqueous solutions. A glass scientist would say that Corning 015 glass is subject to ''glass corrosion''. A glass subject to corrosion will begin to corrode immediately after water contact. As can be seen from Table 2, not all glasses are subject to significant glass corrosion.

Using infrared (IR) spectroscopy, Raman spectroscopy and magic-angle spinning (MAS) nuclear magnetic resonance (NMR) spectroscopy, Pandya et al.[95] examined

Table 2 Examples of glasses used to form membranes for H^+, Na^+ and K^+ (the numerical values give the composition in mol%)

Target	Li_2O	Na_2O	K_2O	CaO	BaO	Al_2O_3	SiO_2	Gel layer?	Comments
H^+	–	21.3	–	6.4	–	–	72.3	Yes	Corning 015
H^+	14.3	–	–	–	7.0	–	68.7		Lithia glass
–	30.0	–	–	–	–	–	70.0	No	
Na^+	–	11.0	–	–	–	18.0	70.0		NAS 11-18
–	–	45.0	–	–	–	–	55.0	Yes	
K^+	–	27.0	–	–	–	6.0	67.0		NAS 27-6
–	–	–	35.0	–	–	–	65.0	Yes	

water speciation in alkali silicate glasses with the compositions shown in rows 3, 5 and 7 of Table 2. These had been hydrated over a period of 6 years under ambient conditions, which apparently meant exposure to moisture-laden air at room temperature. The lithia glass had <0.01 wt% water, present as silanol, and no gel layer. The sodium glass had about 5.2 wt% water, present as monomeric, dimeric, and polymeric water. No silanol was observed. The potassium glass was essentially all gel, with 32 wt% water present as monomeric, dimeric, and polymeric water. Again, no silanol was observed. The absence of silanols in the latter cases was probably due to the conditions of hydration. Diffusion of water into glass initiates stage 1 corrosion, which involves the leaching of alkali metal ions from the glass, and which requires that the glass be in contact with bulk water. In stage 1 corrosion, alkali metal ions are leached as water reacts with $\equiv$SiONa groups to form silanol, $\equiv$SiOH, and release Na^+ and OH^- to water or to the solution. The process has been characterized for Corning 015 glass, both for pure water and for electrolyte solutions and for temperatures of 27, 46, and 57 °C.[96]

If pure water is used, the pH increases to $\geq$9, and then stage 2 corrosion begins. This involves hydration of the Si–O–Si groups, i.e. breakdown of the network. Stage 2 corrosion was not characterized for Corning 015 glass, but Zhang et al.[97] measured both sodium and silicon loss for densified and undensified $Na_2O \cdot 3SiO_2$ glasses in distilled water at 5 °C for both stage 1 and stage 2 corrosion. These researchers showed a scanning electron micrograph of the gel layer formed on the undensified glass. Similar behavior is expected for the Corning 015 glass, and other sodium- and potassium-containing glasses.

According to Buck,[98] gel layers with thicknesses between 60 and 1000 nm may form on pH-, sodium- and potassium-sensitive glasses. Since these gel layers may have diffusivities that are several orders of magnitude greater than dry glass,[99] the gel layers may exhibit electrolytic semiconductivity. These glass electrodes will exhibit a space-charge layer, and the mechanism of response should be similar to that of the crystalline membrane electrodes. The results on the lithia glass may indicate that the pH electrodes based on lithia glasses have no gel layer. In that case, the potential must be due to adsorption of hydronium ions or hydroxide ions limited to the surface layer. The adsorption model developed for CdS should apply. This is probably why pH electrodes based on lithia glasses have a quantitation concentration range from pH 0 to 14. The electrode probably has an upper quantitation detection limit of 1.0 M and a lower quantitation detection limit of 10^{-7} M for H^+ and also for OH^-. Hence the detection limits are similar to those observed for the crystalline membrane ISEs. For the sodium ISE, the upper quantitation detection limit is a saturated solution, and the lower quantitation detection limit is 10^{-6} M. More studies need to be carried out on these glasses, using techniques such as NMR, Raman spectroscopy, and scanning electron microscopy to characterize their behavior more fully.

The polymer-based liquid membrane ISEs come in three varieties: the neutral carrier-based ISE, the charged carrier-based ISE and the ion exchanger-based ISE.[25] The ion exchanger-based ISE membrane has an ion exchanger dissolved in an organic solvent, and the resultant solution is dispersed in an inert polymer support. Traditionally, long-chain alkylammonium salts or nonlabile metal complexes have been considered to be anion exchangers.[100] The electroactive carrier is dissolved in an organic solvent, and the organic solution formed is usually dispersed in an inert polymer support. For example, Nielsen et al.[101] constructed a nitrate ISE from tetraoctylammonium nitrate, dialkyl phthalate or dialkyl adipate, and PVC. When such a membrane is placed in contact with an aqueous phase, the tetraoctylammonium ion is confined to the membrane, but the nitrate ion is free to move into or out of the membrane. An example of a nonlabile metal complex is $Fe(II)(o\text{-phen})_3(ClO_4)_2$.[100] The term nonlabile means that the Fe(II) ion is not present as the free ion in the membrane. The anion, which is the counter ion, is free to move. Cation exchangers are typically long-chain metal sulfonate salts or labile metal ion complexes dissolved in an organic solvent, with this organic solution dispersed in an inert polymeric support. For example, the first Ross calcium ISE membrane was formed from calcium bis(2-ethylhexyl phosphate) dissolved in dioctyl phenylphosphonate and supported in a

porous, inert polymer.[45] The calcium bis(2-ethylhexyl phosphate) can ionize, and the free metal ion can move between the membrane and a contacting aqueous phase. However, the 2-ethylhexyl phosphate ligands, which carry a charge of −1, cannot leave the membrane.

Ion-exchanger ISE membranes are electronic insulators and electrolytic semiconductors. Potential generation for these membrane systems is similar to that for other electrolytic semiconductors. For example, Li and Harrison[102] characterized transport in a nitrite ion-selective membrane using spatial imaging photometry. The system consisted of bromo(pyridine)(5,10,15,20-tetraphenylporphyrinato)cobaltate in dioctyl adipate supported in PVC. They found that on exposure to water, hydration occurs, giving a gel layer that is about 50 μm thick. When the membrane was conditioned in a nitrite solution, the nitrite ion replaced the bromide ion. Nitrite ion diffused across the membrane, to give ultimately a nearly uniform distribution, but the diffusion coefficient was only about 5×10^{-9} cm^2 s^{-1}. The potential is established in a thin (~100 Å thick) region at the membrane/solution interface, i.e. a space-charge layer exists.

NMR studies have revealed that the uptake of water by polymer-based liquid membranes occurs in two stages.[103] The first stage has been attributed to water miscibility in the membrane phase. This is equivalent to the existence of monomeric water in the silicate glasses. In the second stage, droplets of water are formed. With the charged-carrier ISE membrane, the carrier is charged in its uncomplexed state, and forms a neutral complex with the target ion. Such membranes also contain ion-exchanger sites, which are needed to insure permselectivity for the target ion.[25]

An example of a charged-carrier anion ISE is the nitrite electrode based on the aquocyanocobalt(III) heptakis(phenylethyl)cobyrinate ion.[104] This membranes has the following composition: 1.0 wt% aquocyanocobalt(III) heptakis(phenylethyl)cobyrinate perchlorate, 37 mol% potassium tetrakis[3,5-bis(trifluoromethyl)phenyl]borate (KTFPB), 65 wt% 2-nitrophenyl octyl ether (NPOE) and 33 wt% PVC.[26] The ligand, confined to the membrane, carries a positive charge when not complexed to the target ion. Hence it excludes cations from the membrane. KTFPB is used to improve the selectivity of the membrane. It dissociates to give a large, lipophilic anion in the membrane which functions to exclude anions. Only an anion which forms a neutral complex with the carrier can enter the membrane.

A charged-carrier cation ISE will have the complementary membrane composition, i.e. a ligand which is negative when not complexed, a salt which contains a lipophilic cation, the solvent, and the polymer support. In their review, Bühlmann et al.[26] included examples of attempts to form such electrodes, but it seems that none of them have given adequate responses or selectivity. For example, Midgley[105] tried to make an ion-exchanger Fe(III) ISE based on the charged-carrier myobactin S. It failed to respond to Fe(III), but did respond to salicylate. The predominant mechanism responsible for potential generation for the charged-carrier ISEs is extraction by means of complex formation.

The third type of polymer-based liquid membrane is the neutral-carrier type. A neutral carrier is an uncharged, lipophilic complexing ligand. Some are large macrocycles with hydrophobic exteriors and internal cavities lined with polar groups. A metal ion which has a diameter such that a "lock and key" fit may occur between the metal and the ligand will form a strong complex with the ligand. Smaller or larger ions form weak or no complexes with the ligand. An example of a neutral-carrier macrocycle is shown in Figure 13(a). Many of the neutral carriers being used today are not macrocycles. Examples are the various Nile Blue oxazone derivatives (ETH *wxyz*, where *w*, *x*, *y*, and *z* are integers)[106] and *para*-substituted trifluoroacetyl benzenes. The former are chromoionophores originally developed for optodes, but have now been used as neutral carriers for several different cations.[26] The latter are neutral carriers used in commerical carbonate ISEs. An example of a neutral-carrier ISE membrane for a cation is the valinomycin potassium ion ISE. Valinomycin and potassium ion give a "lock and key" fit, as shown in Figure 13(b).

An example of a recipe for such a membrane is as follows: 1.0 wt% valinomycin, 0.01 wt% potassium tetraphenylborate ($KBPh_4$), 66.0 wt% dioctyl sebacate (DOS) and 33.0 wt% PVC. $KBPh_4$ is completely ionized in the organic solution, with the potassium ions all complexed by valinomycin and the tetraphenylborate ions constituting mobile negative ion sites confined to the membrane. A schematic of such a membrane is shown in Figure 14. The tetraphenylborate ions function to exclude sample anions from the membrane.

It has been shown that functional ISEs may be prepared without adding the ion-exchanger salt. Ye et al.[107] showed by XPS and time-of-flight (TOF)

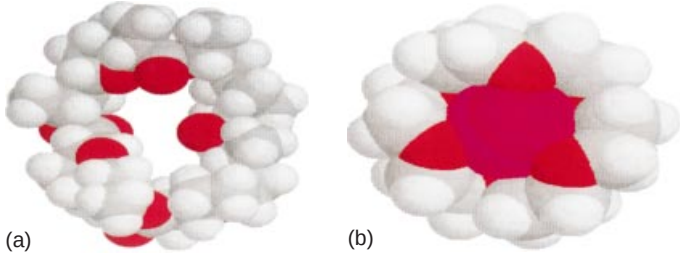

Figure 13 (a) A macrocyclic neutral ligand, nonactin. (b) The K^+–valinomycin complex ion. (Reproduced by permission of David Woodcock.)

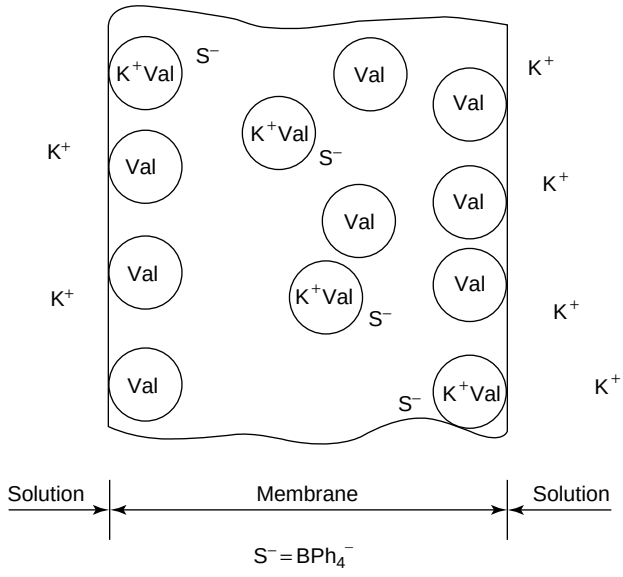

Figure 14 Schematic of a polymeric neutral-carrier liquid membrane.

secondary ion mass spectrometry (SIMS) that there are carboxylate and sulfate groups covalently attached to the polymer support. Where these are in contact with water in the membrane, they will form fixed charge sites in the membrane. The number of such sites, and their distribution, are unknown. The addition of ion-exchanger salts allows the number of sites to be controlled, and this leads to improvements in selectivity. The potassium ions from the $KBPh_4$ are complexed with the neutral carrier, giving lipophilic cations. These will be excluded from the membrane cations which do not complex with the neutral carrier. DOS is the solvent.

An example of a neutral-carrier anion ISE is one whose membrane contains *N*,*N*-dioctyl-4-trifluoroacetylbenzamide, a neutral carrier for carbonate. The membrane composition is as follows:[26] 3.4 wt% *N*,*N*-dioctyl-4-trifluoroacetylbenzamide, 41 mol% tridodecylmethylammonium chloride (TDDMACl), 54 wt% DOS and 41 wt% PVC. Here, TDDMACl is the ion-exchanger salt, which gives large, lipophilic cations in the membrane. With the neutral carrier-type ISEs, the mechanism for potential generation is ion extraction. An excellent discussion of the potential generation mechanism for the valinomycin potassium electrode has been published by Sandifer.[108]

While the polymeric membrane ISEs are being used to analyze for lithium, potassium, magnesium and carbonate in commercial clinical analyzers, the greatest potential for these electrodes lies in their ability to be tailored for the analysis of biologically important organic cations and anions. The key is to find carriers that selectively target an analyte present at a low concentration in a complex sample with high concentrations of other ions. The ability to achieve selectivity by design has been greatly advanced by the recent development of a precise selectivity model for these electrodes. It began with a demonstration of the inability of the Nicolsky–Eisenman formalism to give self-consistent results for a target ion with charge +1 and an ion interference with charge +2. If we consider the electrode response for an ISE immersed in series of mixed solutions with a fixed activity of the ion interference of charge +2 in which the activity of the target ion of charge +1 varies, we expect the response to be described by a single curve which increases linearly as a function of the logarithm of the target activity, a_Q, when a_Q is much larger than the activity of the ion interference and is constant when a_Q is much smaller than the activity of the ion interference. Thus, the latter part of the curve is described by a straight line of zero slope. In the intermediate region, a single curve must be obtained, because the electrode will give a single measured value for each solution. However, the Nicolsky–Eisenman formalism gives different curves, depending on whether the electrode is viewed as a sensor for Q or as a sensor for the ion interference.[79] By considering the response mechanisms for the three types of polymeric membrane electrodes and using appropriate electroneutrality conditions, complex formation constants and mass balances, the authors derived explicit electrode response functions for the case of $z_Q = +1$ and z ion interference $= +2$. This involved the introduction of a selectivity factor, k^{sel} = the ratio of the activity of the target to the activity of the ion interference which gives the same potential when measurements are made in solutions of the single ions. For the ion-exchanger polymeric membrane, k^{sel} can be expressed in terms of the so-called "single ion distribution coefficients". For the charged-carrier polymeric membrane, k^{sel} can be expressed in terms of respective single ion distribution coefficients and complex formation constants. For the neutral-carrier polymeric membrane, k^{sel} depends on the total neutral carrier, as well as the single ion distribution coefficients and complex formation constants. The explicit functions are obtained by expressing the conventional Nicolsky selectivity coefficient in terms of k^{sel}.

In a subsequent paper,[25] explicit response functions were given for the following coupled charge magnitudes: (1,2), (2,1), (1,3), (2,3), (3,2) and (3,1), where the first element of the pair corresponds to the target ion and the second to the ion interference. Most recently, a general implicit function valid for any number of sample ions with any type of charge has been developed.[82] Using E_B as the phase boundary potential difference (this is directly proportional to $\phi_\beta - \phi_\alpha$ in our treatment in the general section), the implicit function is given as

Equation (35):[82]

$$1 = \sum_i a_i(\text{org}) K_i \exp\left(\frac{-z_i E_B}{s}\right) \quad (35)$$

where i represents the ions, a is activity, z_i is the ion charge, $s = RT/\Im$ and K_i is a weighing factor, which combines a number of constants that depend on the type of polymeric membrane under consideration. For the ion-exchanger membrane, it is given by Equation (36):[82]

$$K_i = \frac{z_i k_i}{\gamma_i R_T} \quad (36)$$

where k_i is the single ion distribution coefficient, γ_i is the activity coefficient and R_T is the concentration of nonexchangeable ions in the membrane. For the ionophore-based membrane (neutral carrier or charged carrier), K_i is given by Equation (37):[82]

$$K_i = \frac{z_i k_i \beta_{i,n} c_L^0(\text{org})^n}{R_T \gamma_{i,n}} \quad (37)$$

where $\beta_{i,n}$ is the complex formation constant for ion i forming a $1:n$ complex with the ligand, $\gamma_{i,n}$ is the corresponding activity coefficient, c_L^0 (org) is the initial concentration of the uncomplexed carrier, and the other terms have been defined above. Potentiometric selectivity coefficients are expressed in terms of these weighing factors as Equation (38):[82]

$$K_{I,i}^{\text{pot}} = \frac{K_I}{K_i}^{z_I/z_i} \quad (38)$$

In turn, k_{Ii}^{sel} can be related to the potentiometric selectivity coefficients, and this leads to a simple expression for the boundary potential (Equation 39):[82]

$$E = E_I^0 + \frac{RT}{z_I} \ln \quad a_I(\text{aq}) + \sum_{i \neq I} k_{Ii}^{\text{sel}} a_i(\text{aq}) \quad (39)$$

The usefulness of the k^{sel} is obvious. First, by measuring single ion distribution coefficients and complex formation constants, and by determining the appropriate concentration of lipophilic ion sites needed, the selectivity of the membrane can be designed. Optimization techniques can be used to design membranes to give the best response for a set of specified ion interferences. This will allow the designer to generate quickly a set of membrane compositions likely to be suitable detectors for a given problem. The experimental work would then be limited to those compositions. Intelligent design allows time to be used more efficiently and minimizes waste of reagents.

Just as the interference order may be qualitatively predicted for crystalline membrane ISEs by using K_{ap} values, the selectivity order for charged-carrier ionophore anion ISEs often obeys the Hofmeister lyophilic series (Equation 40):[109]

$$R^- > ClO_4^- > I^- > NO_3^- > Br^- > Cl^- > F^- \quad (40)$$

where R^- is an organic anion. For charged-carrier ionophore cation ISEs, the Hofmeister series is [Equation 41][109]

$$R^+ > Cs^+ > Rb^+ > K^+ > Na^+ > Li^+ \quad (41)$$

where R^+ is an organic cation. There does not appear to be a recurring selectivity order for divalent cations, etc. For the neutral carrier-type liquid membranes, the size and shape of the ion cavity determine the selectivity order to a large extent. It is possible to use molecular modeling techniques to investigate selectivity orders for any given neutral carrier. Since the organic solvent also influences the selectivity order, these methods are a kind of zeroth-order approximation. Commercially available macroelectrode ISEs with plastic membranes have an average quantitation lower detection limit of about 4×10^{-6} M and an average quantitation upper detection limit of about 1.0 M. With crystalline membrane ISEs, the selectivity is determined by the properties of the solid. By buffering the sample solution, the lowest quantitation detection limits may be realized, but there can be no further improvement. With the polymeric membrane ISEs, the composition of the membrane can be changed to improve the selectivity, which, when coupled with buffering, can push the lower quantitation detection limit down. One way to improve membrane selectivity is to find more selective ionophores. This is why so much research in this area involves synthesizing and evaluating new ionophores, not only for new target ions, but also for ions that already have polymeric membrane ISE sensors. For example, neutral-carrier polymeric membrane ISEs for calcium and lead with lower quantitation detection limits of 10^{-9} M have been described.[110,111] In comparison with the commercial calcium polymeric membrane ISE, this respresents an improvement by a factor of 500, and in comparison with the commercial crystalline membrane lead ISE, an improvement by a factor of 1000.

3.2.2 *Compound or Multiple Membrane Ion-selective Electrodes*

These include gas-sensing electrodes (called Severinghaus-type potentiometric detectors) and enzyme electrodes. A gas-sensing electrode probes the concentration (activity) of a dissolved gas in an external solution as the result of admission of the dissolved gas into a compartment containing a solution which can interact with the gas to form the target ion of the internal ISE.

The compartment and the external solution are separated by a microporous, semipermeable membrane, so the gas must diffuse across this membrane. If the gas of interest is in the vapor phase, then its solution concentration (activity) will be determined by its solubility equilibrium. The process that occurs in the compartment is kinetically fast, so the response time of the electrode will be determined by the kinetics of the diffusion process and the kinetics of the ISE response. The former is always much slower. The measurement should be taken under steady-state conditions. The concentration (activity) of the gas to be measured can be related to the activity of the target ion by means of a mathematical expression based on the multiple equilibria which occur. No general expression is possible, but an example might be instructive. Suppose that we choose to monitor the partial pressure of ammonia in a stack gas using an ammonia gas ISE. The ammonia gas electrode contains a pH glass electrode, a compartment which contains either 10^{-3} M NH_4Cl or 0.1 M KNO_3, and a 0.1-mm micropore Teflon® membrane.[112] The electrochemical cell consists of the ammonia gas electrode, a reference electrode and an external solution in contact with the stack gas. The concentration of dissolved ammonia is related to the partial pressure of ammonia gas in the stack gas by means of Henry's law, $P_{gas} = k[NH_3]_{aq}$. Since a concentration gradient for dissolved ammonia exists across the membrane, ammonia diffuses into the compartment. Diffusion must occur until the concentration gradient is zero, so the concentration of dissolved ammonia in the compartment will be the same as that in the external solution at steady state. If the compartment contains 0.1 M KNO_3, then base hydrolysis of the ammonia fixes the hydroxide ion activity, which fixes the hydrogen ion activity. The ISE's electrode response is related to P_{gas} by Equation (42):

$$E_{ISE} = K_{electrode} + S\log\sqrt{kK_wK_a} - \frac{S}{2}\log(P_{gas}) \quad (42)$$

where S is the slope factor for the pH electrode and the relationship $K_b = K_w/K_a$ has been used. The electrode gives a linear response as long as the approximation $[NH_4^+] = [OH^-]$ is valid, which is so as long as the hydrolysis of water can be neglected. If 10^{-3} M NH_4Cl is in the compartment, then a buffer is formed in the compartment when ammonia diffuses across the membrane. In that case, the ISE's response is related to P_{gas} by Equation (43):

$$E_{ISE} = K_{electrode} + S\log(10^{-3}kK_a) - S\log(P_{gas}) \quad (43)$$

The electrode response is linear so long as the equilibrium concentrations of ammonia and ammonium ion are given by their analytical concentrations. In both cases, ionic strength effects must be adequately addressed. With these electrodes, the recovery times may be long (many minutes), so care must be taken to avoid memory effects in successive measurements. Interferences arise from other gases which can diffuse across the membrane and either control the activity of the target ion or form an ion which is an electrode interference for the internal ISE. For example, for the ammonia gas electrode, volatile amines interfere, since they can diffuse across the Teflon® membrane and they are weak bases, like ammonia. The quantitation lower detection limits for the commercial macroelectrode ISEs for NO_x and ammonia are 4×10^{-6} and 5×10^{-7} M, respectively. For the commercial carbon dioxide macroelectrode ISE, the quantitation lower detection limit is only 10^{-4} M. The detection ranges for the NO_x and for the carbon dioxide macroelectrode ISEs are fairly small; the quantitation upper detection limits are 5×10^{-3} and 10^{-2} M, respectively. By contrast, the ammonia macroelectrode ISE has a quantitation upper detection limit of 1.0 M. Recently, polymeric membrane-type ISEs have been used to assemble carbon dioxide and ammonia sensors that are analogs to the Severinghaus-type sensors. The carbon dioxide sensor uses a neutral-carrier hydrogen ion polymeric membrane,[26,113] whereas the ammonia electrode uses a neutral-carrier ammonium ion polymeric membrane.[26,114–116] The neutral-carrier ammonium ion polymeric membrane shows an improved quantitation lower detection limit of 25, relative to the macroelectrode ISE. A neutral-carrier polymeric membrane-based ISE has been developed for sulfur dioxide.[26] At present, there is no commercial macroelectrode ISE for this gas. Oxygen sensors have also been described,[26] but it is not clear that the response is due entirely to ion transport and not electron transfer. These electrodes may not be ISEs in the traditional sense.

An enzyme electrode has a gel layer containing an enzyme interposed between the indicator ISE membrane and an external solution. The enzyme catalyzes the conversion of a substrate to a set of products, one of which is a target of the ISE. Targets are frequently small molecules or ions, e.g. ammonia, carbon dioxide, hydronium ion, cyanide ion.[50] The selectivity of these electrodes is governed by the specificity of the enzyme. Because enzymes are stereoselective, these ISEs come the closest to being electrodes that give an activity-dependent response for a single type of molecule or ion, the substrate of the enzyme. The generation of the signal requires three steps: (1) the substrate must diffuse from the external solution into the enzyme gel layer, (2) the enzyme must catalysis the conversion to the set of products and (3) the product which is the ISE target must diffuse to the surface of the ISE membrane. In principle, the kinetics associated with step 2 can be complex. Also, while the concentration gradient is such

that the substrate diffuses into the gel, the products will diffuse both toward the membrane surface and toward the external solution. Practitioners have chosen enzyme systems which obey Michaelis–Menten kinetics, which is the simplest type. Even so, the transport equations are inherently nonlinear. In order to obtain an equation which relates the potential of the enzyme electrode to the substrate activity (concentration), the equations are solved for steady-state conditions. At the steady state, the diffusion processes may be ignored, and only step 2 is important. In such a case, a set of simultaneous first-order differential equations need to be solved. An example of such a treatment can be found in the book by Morf.[117] Brady and Carr[118] actually solved numerically the coupled set of nonlinear, second-order differential equations for a set of reasonable boundary conditions. Since an expression for the electrode potential as a function of substrate activity (concentration) cannot be obtained in analytical form, this method is not useful to the practitioner. However, it has allowed certain approximations, which constitute general guidelines for the use of such electrodes, to be validated. In general, a linear response region is observed, the lower bound being determined by the ionic background in the sample, and the upper bound being determined by the Michaelis constant for the enzyme system.

3.2.3 Metal Contact or All-solid-state Electrodes

The behavior of these electrodes is similar to that of the primary electrodes. One type of failure which we have seen for microelectrodes of this type is the decoration of the membrane surface with silver, when silver is used to make the internal contact. Silver atoms have fairly large diffusion constants, in comparison with other metal atoms, and thus even microcracks will result in its ultimate transport to the surface. When this happens, the electrode response gives sub-Nernstian slope factors.

4 INSTRUMENTATION

An ISE measurement involves the determination of the potential difference between an ISE and an external reference electrode. The instrumentation used to make this voltage measurement must not load the circuit. There are two types of instrumentation that may be used to make voltage measurements – a potentiometer or a voltmeter. A potentiometer is a null device, whose limitations depend on the current sensitivity of the null detector. The classical voltmeter is a direct measuring device, e.g. the D'Arsonval meter, in which a small current causes a meter deflection. For a given electrochemical cell, the absolute error in the voltage measurement will depend on both the magnitude of the actual voltage to be measured and on the meter resistance (impedance). The electrochemical cell resistance is determined essentially by the resistance of the ISE, which in turn depends on the geometry of the sensing membrane. In Table 3 are shown some typical values of ISE resistances for various types of electrodes.[119,120] The meter resistance needed to give an absolute error of no more than 1 mV for the measurement of pH values from 1 to 14 may be calculated using the extremes of the output potential range, +414 to −414 mV.[119] A value of $8.0 \times 10^{11}\ \Omega$ is calculated. Thus, a meter resistance (impedance) of 10^3–10^4 times that of the cell resistance is needed.

Classical voltmeters, so-called VOM instruments, cannot be used to measure the voltages of electrochemical cell using ISEs, because the meter resistances (impedances) are too small (typically $1 \times 10^5\ \Omega$). Thus, early pH meters made use of potentiometers, wherein the null detectors evolved from galvanometers to vacuum tube configurations. However, the classical instrumentation era culminated with the development of the vacuum tube voltmeter by A. O. Beckman. This interesting history has been discussed by Jaselskis et al.[121]

The semiconductor revolution introduced the next innovation, as vacuum tubes were replaced by semiconductors, giving rise to the electronic voltmeter. The early electronic ion meters were meter readout instruments, but today's electronic ion meters are digital readout instruments, which utilize either a light-emitting diode (LED) display or a liquid crystal display (LCD). These instruments can have meter resistances (impedances) in excess of $10^{13}\ \Omega$. The heart of a modern electronic ion meter is the operational amplifier. The operational amplifier is a small-scale integration device which has a high input impedance ($\geq 10^{12}\ \Omega$), a low output impedance (≤ 1–$10\ \Omega$), and a large open-loop gain (10^4–10^6).[122] Many circuit design are possible; in Figure 15 is shown a simple circuit design which uses a voltage follower configuration with manual adjustment for the slope factor. Slope factor adjustment is needed to adjust for measurements made at different temperatures.

Table 3 Typical ISE resistances

Type of ISE	$R\ (\Omega)$
Conventional glass membrane	5.0×10^7–2.0×10^8
Conventional crystalline membrane	2.0×10^5–3.0×10^7
Conventional liquid membrane	1.0×10^7–2.5×10^8
Conventional gas sensing	$\sim 5.0 \times 10^9$
Conventional enzyme electrode	1.0×10^7–2.0×10^9
Microelectrode	Up to 10^{12}

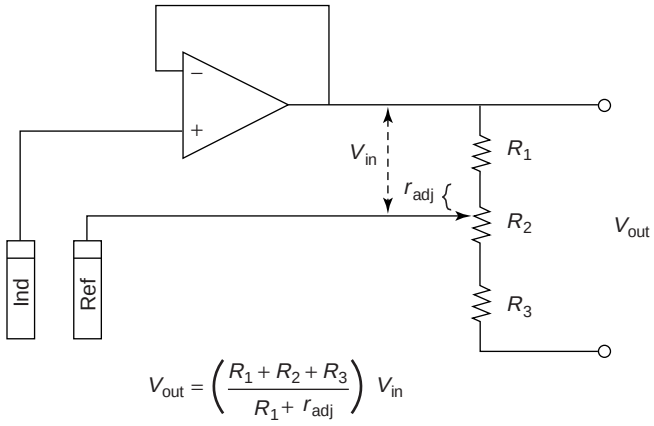

Figure 15 E_{cell} measurement with manual slope adjustment using a voltage follower operational amplifier circuit.

Modern digital instruments have many more features. Most are microprocessor control instruments with multiple measuring modes (e.g. millivolts, pX, concentration mode in various units), with temperature compensation probes, and connectors for interfacing the instrument with auxiliary instruments. In Figure 16 is shown a block diagram of the Fisher Accumet Model 750 selective ion analyzer. The important circuits are shown in blocks. For example, the temperature compensation may be done manually (by entering information into the system via the keyboard) or automatically by means of a probe. The instrument has an analog-to-digital converter, and the digital data are available in binary coded decimal at an external connector and in decimal form at an LED display. This meter is a desktop model, and the technology dates from about 1981.

The new benchtop Accumet research meters are microprocessor-based instruments with backlit LCD touchscreens, prompts and context-specific help screens. In both the pH mode and ion mode, they offer up to five points of calibration. There are also portable Accumet meters that offer the same calibration options, but these do not have the help screens.

In Figures 17–19 are shown examples of the most recent Orion technology. In Figure 17 is shown the Orion Model 710A ion meter, a basic benchtop ISE/pH meter. It also uses an LCD and a touchpad input module with a large number of functions. It covers a pX range of −2.000 to 19.999, but the resolution depends on the pX value. For example, with pX = 7.00, the resolution is two decimal places. The meter has auto-buffer recognition for five-point pH standardization. There are built-in buffer/temperature tables. By contrast, the older

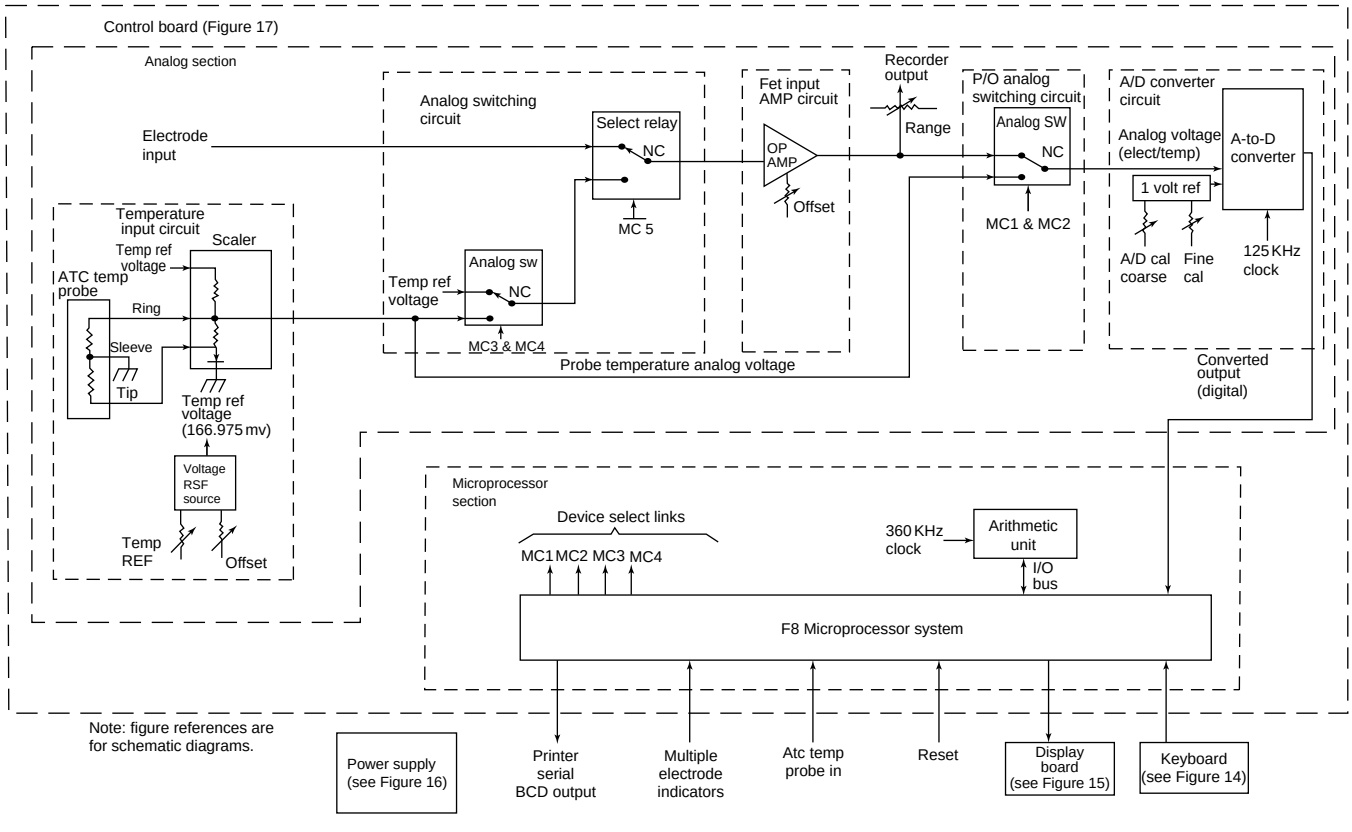

Figure 16 Block diagram for the Fisher Accumet Model 750 selective ion analyzer. (Reproduced by permission of Fisher Scientific.)

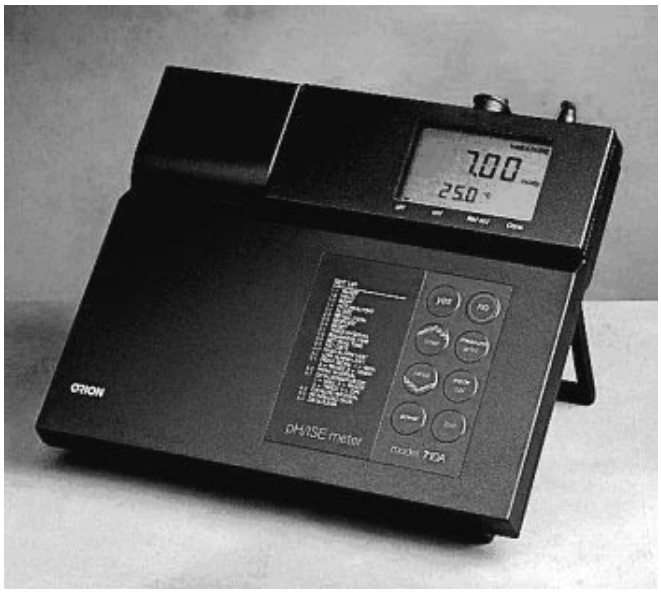

Figure 17 The Orion Model 710A pH/ISE meter. (Reproduced by permission of Orion Research, Inc.)

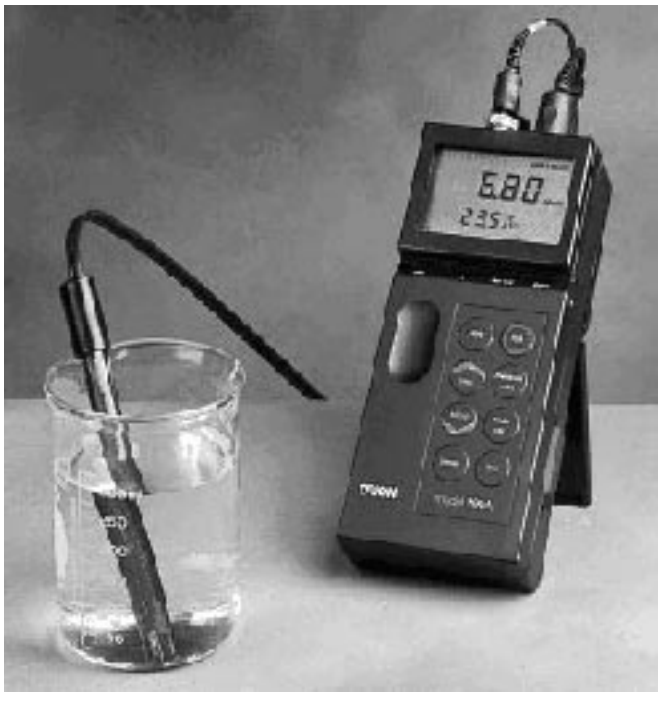

Figure 18 A portable pH/ISE meter, the Orion Model 290A. (Reproduced by permission of Orion Research, Inc.)

technology, e.g. the Accumet Model 750 meter mentioned above, uses two-point calibration, with the data entered into the system manually. In Figure 18, the Orion Model 290A meter is shown. This is an advanced, portable pH/ISE meter which is not much larger than a pocket calculator. Portable instruments are ideal for field measurements; the two-way RS232 communication and datalogging capabilities allow results to be downloaded later into a laboratory computer. In Figure 19 is shown one of the Orion SensorLink® systems. This is a PCMCIA card-based system, shown here interfaced with a PC; systems are also available for laptop computers. A standard Windows interface, using either Windows 95 or Windows 3.1, is used. The laptop systems are ideal for traveling scientific consultants or for Environmental Protection Agency compliance officers. Data can be obtained and analyzed and reports generated all within 1 day. The portable pH/ISE meters and the card-based systems are truly the wave of the future, being realized today.

Figure 19 The Orion SensorLink® system interfaced to a PC. (Reproduced by permission of Orion Research, Inc.)

5 ION-SELECTIVE MICROELECTRODES

As we have seen, the development of the pH electrode was initiated by researchers interested in making physiological measurements of acidity. Although the Beckman pH meter revolutionized the routine measurement of pH in wine, pickling baths, etc., the first pH meter for measuring the pH of blood was not introduced until 1954.[123] ISE technology has replaced many of the clinical analysis techniques based on colorimetric analysis, flame photometry, and atomic absorption. Flame photometry was once the method of choice for the determination of lithium, sodium and potassium in samples. However, the inorganic material had to be separated from its biological matrix, a time-consuming process. Using ISE analyzers, whole blood samples can be analyzed for these ions.[123] Atomic absorption used to be the method of choice for

the clinical determination of calcium and magnesium. ISE analyzers are now used for these determinations. The trend has been to replace multiple instruments with a single analyzer capable of providing analysis for multiple targets. For example, the Nova 16 Stat chemical analyzer, an all-electrode (but not all-ISE) instrument, provides for the analysis of blood urea nitrogen, glucose, creatinine, sodium, potassium, chloride, and total carbon dioxide in less than 1 mL of whole blood.

The development of ISEs for both in vivo and in vitro biological measurements and medical applications has always been a goal of researchers in the health-related fields and in the biological sciences. Since in vivo measurements must be made in very tiny compartments, very tiny probes – microelectrodes – have to be developed. Ammann[124] presented an excellent chart of the size ranges of various biological compartments, against which we may assess how far we have come and how much further we need to go in developing such electrodes. In Table 4, that information is shown using the median size of the biological compartments. In Figure 20 the three most useful geometries, in our opinion, for making in vivo measurement are shown. Coated-wire electrodes can be inserted into egg cells, but they are too large to be inserted into any of the other cells. The catheter electrodes are about the same size as the coated-wire electrodes. Just as catheters are useful for the continuous injection of drugs intravenously, catheter electrodes could be used for the continuous monitoring of molecules and ions that circulate in the bloodstream. However, only the microelectrode has been made small enough that the contents of cells and some of their organelles may, in principle, be measured. The smallest tip diameter of an enzyme microelectrode is 2 μm, which means that these electrodes may be used to probe the cytoplasmic regions of plant and animal cells, vacuoles and the nucleus. With the neutral carrier microelectrodes, with the smallest tip diameter of 300 nm, the above compartments, and also mitochondria, bacteria, and lysosomes, may be probed. For the ion-exchanger and charged-carrier-type polymeric membranes, microelectrodes with minimum tip diameters of 100 nm have been constructed. These can interrogate the same compartments as the neutral carrier-type microelectrodes, but are not small enough to insert into ribosomes or viruses.

Table 4 Median sizes (equivalent spherical diameter) of selective biological samples

Biological sample	Median size
Egg cell	820 μm
Animal cell	53 μm
Plant cell	53 μm
Vacuole	18 μm
Nucleus	6.9 μm
Mitochondrion	1.8 μm
Bacterium	770 nm
Lysosome	350 nm
Ribosome	46 nm
Virus	46 nm

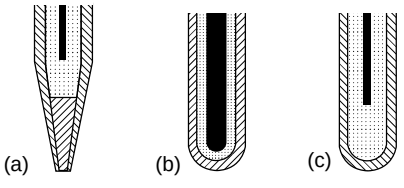

Figure 20 Some useful microelectrode geometries: (a) microelectrode; (b) coated-wire electrode; (c) catheter electrode.

The prospect of using enzyme microelectrodes for the in vivo, selective monitoring of biological molecules is extremely attractive. What is needed is another factor of 10 decrease in the smallest tip diameter of these microelectrodes. However, obtaining useful signals from these smaller microelectrodes would require concomitant improvements in instrumentation capabilities. At present, microelectrodes may have resistances of the order of $1 \times 10^{12}\ \Omega$. A decrease in diameter by a factor of 10 with no decrease in thickness would cause the resistance to increase by a factor of 100. This means that for comparable measurement errors, the meter resistances also need to increase by the same factor. There are other problems. The resistance of the cable cladding needs to be much larger than the resistance of the ISE, so that voltage drops between the cable core and the cladding are negligible compared with the magnitude of the signal. When the cladding resistance and the ISE resistance are similar in magnitude, electrode cable shielding must be performed in such a way as to reduce the voltage drop between the cable core and the cable cladding to virtually zero.[120] The cladding problem is already being encountered for ISE resistances of $1 \times 10^{12}\ \Omega$. Commercially available neutral carrier microelectrodes exist for the monitoring of lithium, sodium, potassium, magnesium, calcium, and hydrogen ions. Since the uptake or release of various of these ions often accompanies metabolic processes, these are useful for in vivo studies. Numerous examples of intracellular measurements of ions using neutral carrier microelectrodes exist.[125–129] Recently, a neutral-carrier polymeric ISE microelectrode was developed to study the uptake of cadmium in plant cells.[130]

6 NOVEL DEVELOPMENTS IN ION-SELECTIVE ELECTRODE RESEARCH

In this section, some recent developments which show particular promise in advancing the field of ISEs are

discussed. This section is not intended to be comprehensive, and may unintentionally reflect the biases of the author! The development of polymer membrane-type ISEs for clinical uses and for monitoring the condition of patients during surgical intervention or during their recuperation in critical care units is a very active area. Several specific reviews have appeared.[131–133] In particular, the development of polymeric ISEs for polyion analysis must be mentioned. One such polyion is heparin. Heparin is a glycosaminoglycan that occurs naturally in mast cells and cells that line the walls of arteries. It functions as an anticoagulant, and is used in hemodialysis, during heart surgery, and postsurgically. A heparin ISE has been developed by Meyerhoff et al. and is being improved continuously.[134–136]

Of course, the monitor itself must not induce thrombosis, so it is important that the polymeric membrane be biocompatible. Therefore, studies on modified PVC and other polymer matrices have been carried out.[137] A novel development is a heparin ISE which slowly releases NO, a platelet antiaggregation agent.[138] A disadvantage of these ISEs is that heparin must be removed from the membrane between measurements. An improved cleaning method has recently been developed.[139] The development of polyion-sensitive membrane ISEs for nonseparation immunoassays also seems promising.[63] The ability to analyze economically and quickly undiluted blood samples for illicit drugs would be useful for sports compliance monitoring and also for police monitoring for drug use.

Another development is ion imaging by means of scanning ion-selective potentiometric microscopy.[140] It was shown that this technique could be used to image ions over conducting and insulating targets. This comes very close to being a measuring instrument that does not perturb the system that is being measured. This instrument will make it possible to use microelectrodes in order to advance our understanding of the detailed behavior of macroelectrodes, which can be used as targets.

More developments in the area of simultaneous analyte monitoring are expected. This area began with the development of sensors for the simultaneous monitoring of two analytes, e.g. pH and carbon dioxide.[141,142] The further development of ISE arrays for simultaneous multitarget determinations will continue.[143,144] This also includes methods of microfabrication.[145,146] The development of data acquisition cards, which allow as many as 14 different channels to be interfaced to a computer, and sampling under computer control make these types of arrays attractive. The main problem with ISE arrays is that many of the ion mixtures that one might want to determine give rise to a complex electrode interference problem. The development of data acquisition designs and chemometric approaches to the analysis of such data are steps toward solving that problem.[147–149]

Table 5 Ion-selective sensors in the WCL

Sensor	Type	Analyte
pH	Polymer membrane	pH
pH	Iridium dioxide	pH
Silver/sulfide	Crystalline membrane	Silver or sulfide
Cadmium	Crystalline membrane	Cadmium
Chloride	Crystalline membrane	Chloride
Bromide	Crystalline membrane	Bromide
Iodide	Crystalline membrane	Iodide
Lithium	Polymer membrane	Used as reference
Sodium	Polymer membrane	Sodium
Potassium	Polymer membrane	Potassium
Magnesium	Polymer membrane	Magnesium
Calcium	Polymer membrane	Magnesium
Ammonium	Polymer membrane	Ammonium
Nitrate	Polymer membrane	Nitrate
Perchlorate	Polymer membrane	Perchloride
Carbon dioxide	Membrane gas sensor	Carbon dioxide/ hydrogencarbonate

Finally, ISEs are heading to Mars! A project called MECA (Mars Environmental Compatibility Assessment) is scheduled to fly aboard the Mars Surveyor 2001 Lander.[150] MECA contains a wet chemistry laboratory (WCL) that consists of four beakers, which contain sensor arrays that are mostly ISEs. Both solid-state and polymeric membrane electrodes are being used. In Table 5, the ISE sensors being used in the WCL are shown. Conspicuously absent are glass ISEs. Cracking due to thermal cycling and degradation of the insulation resistance led to catastrophic electrical leakage.

LIST OF SYMBOLS

a	activity
β	efficiency factor in electrode potential equation
$\beta_{i,n}$	complex formation constant
c, C	concentration, usually in mol L^{-1}
γ	activity coefficient
δ	space-charge layer thickness. SI units m, but usually expressed in μm, nm, etc.
D	diffusion coefficient. SI units $m^2\ s^{-1}$, but frequently given in $cm^2\ s^{-1}$ in the literature
E	cell emf, electrode emf, in V or mV; phase boundary potential difference
$\Im$	Faraday constant, 96 500 J C^{-1}
k	Henry's constant
k_i	single ion distribution coefficient
k^{sel}	selectivity coefficient
K	constant in the Nernst equation in V or mV
K_a	acid dissociation constant

K_b	base hydrolysis constant
K_i	weighing factor
K_{sp}	solubility product constant
K_w	autoprotolysis constant of water
$K^{pot}_{A,B}$	selectivity coefficient
μ	chemical potential, J mol^{-1}
p	osmotic pressure
P	partial pressure of a gas, atm
Φ, ϕ	Galvani potential, V or mV
R_T	concentration of nonexchangeable ions
S	empirical slope factor, V or mV
τ	time constant, s^{-1}
t_+, t_-	transport number
X, χ	chi potential, V or mV

ABBREVIATIONS AND ACRONYMS

CHEMFET	Chemical-sensing Field Effect Transistor
DOS	Dioctyl Sebacate
emf	Electromotive Force
IR	Infrared
ISE	Ion-selective Electrode
ISFET	Ion-selective Field Effect Transistor
IUPAC	International Union of Pure and Applied Chemistry
KTFPB	Potassium Tetrakis[3,5-bis(trifluoromethyl)phenyl]borate
LCD	Liquid Crystal Display
LED	Light-emitting Diode
MAS	Magic-angle Spinning
MECA	Mars Environmental Compatibility Assessment
NBO	Nonbridging Oxygen
NMR	Nuclear Magnetic Resonance
NPOE	2-Nitrophenyl Octyl Ether
PVC	Poly(vinyl chloride)
SIMS	Secondary Ion Mass Spectrometry
TDDMACl	Tridodecylmethylammonium Chloride
TOF	Time-of-flight
WCL	Wet Chemistry Laboratory
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • Automation in the Clinical Laboratory • Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Electrolytes, Blood Gases, and Blood pH • Laboratory Instruments in Clinical Chemistry, Principles of

Environment: Water and Waste **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Field-portable Instrumentation cont'd **(Volume 5)**
Solid-state Sensors for Field Measurements of Gases and Vapors

Remote Sensing **(Volume 10)**
Remote Sensing: Introduction

Surfaces **(Volume 10)**
X-ray Photoelectron Spectroscopy in Analysis of Surfaces

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Titrimetry

REFERENCES

1. R.P. Buck, E. Lindner, 'Recomendations for Nomenclature of Ion-selective Electrodes', *Pure Appl. Chem.*, **66**(12), 2527–2536 (1994).
2. N.H. Furman, 'Potentiometric Titrations', *Anal. Chem.*, **22**, 33–41 (1950).
3. N.H. Furman, 'Potentiometric Titrations', *Anal. Chem.*, **23**, 21–24 (1951).
4. N.H. Furman, 'Potentiometric Titrations', *Anal. Chem.*, **26**, 84–90 (1954).
5. C.N. Reilley, 'Potentiometric Titrations', *Anal. Chem.*, **28**, 671–678 (1956).
6. C.N. Reilley, 'Potentiometric Titrations', *Anal. Chem.*, **30**, 185R–193R (1960).
7. R.W. Murray, C.N. Reilley, 'Potentiometric Titrations', *Anal. Chem.*, **34**, 313R–380R (1962).
8. R.W. Murray, C.N. Reilley, 'Potentiometric Titrations', *Anal. Chem.*, **36**, 370R–380R (1964).
9. D.K. Roe, 'Potentiometric Titrations', *Anal. Chem.*, **38**, 461R–469R (1966).
10. E.C. Toren, Jr, 'Potentiometric Titrations', *Anal. Chem.*, **40**, 402R–412R (1968).
11. E.C. Torens, Jr, P.M. Gross, R.P. Buck, 'Potentiometric Titrations', *Anal. Chem.*, **42**, 284R–304R (1970).
12. R.P. Buck, 'Ion-selective Electrodes, Potentiometry, and Potentiometric Titrations', *Anal. Chem.*, **44**, 270R–295R (1972).

13. R.P. Buck, 'Ion-selective Electrodes, Potentiometry, and Potentiometric Titrations', *Anal. Chem.*, **46**, 28R–51R (1974).
14. R.P. Buck, 'Ion-selective Electrodes', *Anal. Chem.*, **48**, 23R–39R (1976).
15. R.P. Buck, 'Ion-selective Electrodes', *Anal. Chem.*, **50**, 17R–29R (1978).
16. G.H. Fricke, 'Ion-selective Electrodes', *Anal. Chem.*, **52**, 259R–275R (1980).
17. M.E. Meyerhoff, Y.M. Fraticelli, 'Ion-selective Electrodes', *Anal. Chem.*, **54**, 27R–44R (1982).
18. M.A. Arnold, M.E. Meyerhoff, 'Ion-selective Electrodes', *Anal. Chem.*, **56**, 20R–48R (1984).
19. M.A. Arnold, R.L. Solsky, 'Ion-selective Electrodes', *Anal. Chem.*, **58**, 84R–101R (1986).
20. R.L. Solsky, 'Ion-selective Electrodes', *Anal. Chem.*, **60**, 106R–113R (1988).
21. R.L. Solsky, 'Ion-selective Electrodes', *Anal. Chem.*, **62**, 21R–33R (1990).
22. J. Janata, 'Chemical Sensors', *Anal. Chem.*, **64**, 196R–219R (1992).
23. J. Janata, M. Josowicz, D.M. DeVaney, 'Chemical Sensors', *Anal. Chem.*, **66**, 207R–228R (1994).
24. J. Janata, M. Josowicz, P. Vanýsek, D.M. DeVaney, 'Chemical Sensors', *Anal. Chem.*, **70**, 179R–208R (1998).
25. E. Bakker, P. Bühlmann, E. Pretsch, 'Carrier-based Ion-selective Electrodes and Bulk Optodes. 1. General Characteristics', *Chem. Rev.*, **97**, 3083–3132 (1997).
26. P. Bühlmann, E. Pretsch, E. Bakker, 'Carrier-based Ion-selective Electrodes and Bulk Optodes. 2. Ionophores for Potentiometric and Optical Sensors', *Chem. Rev.*, **98**, 15933–1687 (1998).
27. R.P. Buck, 'Theory and Principles of Membrane Electrodes', in *Ion-selective Electrodes in Analytical Chemistry*, ed. H. Freiser, Plenum Press, New York, 1–141, Vol. 1, 1978.
28. *A Concise Guide to Ion Analysis*, Laboratory Products Catalog & Electrochemistry Handbook, Orion Research, Cambridge, MA, 85–92, 1998.
29. K.L. Cheng, 'pH Glass Electrode and Its Mechanism', in *Electrochemistry, Past and Present*, eds. J.S. Stock, M.V. Orna, American Chemical Society, Washington, DC, 286–302, 1989.
30. K.L. Cheng, 'Capacitor Theory for Nonfaradaic Potentiometry', *Microchem. J.*, **42**, 5–24 (1990).
31. A.F. Burke, T.C. Murphy, 'Materials Characteristics and the Performance of Electrochemical Capacitors for Electric/Hybrid Vehicle Applications', *Materials for Electrochemical Energy Storage and Conversion – Batteries, Capacitors and Fuel Cells*, Materials Research Society Symposium Proceedings, Materials Research Society, Pittsburg, PA, 375–395, Vol. 393, 1995.
32. D.A. Skoog, J.J. Leary, 'Potentiometric Methods', *Principles of Instrumental Analysis*, 4th edition, Saunders College Publishing, Fort Worth, TX, 503–506, 1992.
33. M. Cremer, 'Über die Ursache der Elektromotorischen Eigenschaften der Gewebe Zugleich ein Betrag zur Lehre von dem Polyphasichen Elekolytketten', *Z. Biol.*, **47**, 562–607 (1906).
34. C.E. Moore, B. Jaselskis, A. Smolinski, 'Development of the Glass Electrode', in *Electrochemistry, Past and Present*, eds. J.S. Stock, M.V. Orna, American Chemical Society, Washington, DC, 272–285, 1989.
35. F. Haber, Z. Klemensiewicz, 'Über Elektrische Phasengrenzkräfte', *Z. Phys. Chem.*, **67**, 385–431 (1909).
36. M.D. Archer, 'Genesis of the Nernst Equation', in *Electrochemistry, Past and Present*, eds. J.S. Stock, M.V. Orna, American Chemical Society, Washington, DC, 115–126, 1989.
37. S.P.L. Sørenson, 'Enzyme Studies. II. The Measurement and Importance of the Hydrogen Ion Concentration in Enzyme Reactions', *Biochem. Z.*, **21**, 131–200 (1909).
38. C.E. Moore, B. Jaselskis, A. Smolinski, 'Development of the Glass Electrode', in *Electrochemistry, Past and Present*, eds. J.S. Stock, M.V. Orna, American Chemical Society, Washington, DC, 272–285, 1989.
39. J.O. Isard, 'The Dependence of Glass-electrode Properties on Composition', in *Glass Electrodes for Hydrogen and Other Cations*, ed. G. Eisenman, Marcel Dekker, New York, 51–100, 1967.
40. G.A. Perley, 'Composition of pH-responsive Glasses', *Anal. Chem.*, **21**, 391–394 (1949).
41. G.A. Perley, 'Glasses for Measurement of pH', *Anal. Chem.*, **21**, 394–401 (1949).
42. B. Lengyel, E. Blum, 'The Behaviour of the Glass Electrode in Connection with Its Chemical Composition', *Trans. Faraday Soc.*, **30**, 461–471 (1934).
43. M.S. Frant, 'History of the Early Commercialization of Ion-selective Electrodes', *Analyst*, **119**, 2293–2301 (1994).
44. A.K. Covington, 'Introduction: Basic Electrode Types, Classification, and Selectivity Considerations', in *Ion-selective Electrode Methodology*, ed. A.K. Covington, CRC Press, Boca Raton, FL, 1–20, Vol. I, 1979.
45. M.S. Frant, 'Where Did Ion-selective Electrodes Come From?', *J. Chem. Educ.*, **74**, 159–166 (1997).
46. Z. Stefanac, W. Simon, 'In-vitro-verhalten von Makrotetroliden in Membranen als Grundlage für Hochselektive Kationenspezifische Elektrodensysteme', *Chimia*, **20**, 436 (1966).
47. Z. Stefanac, W. Simon, 'Ion Specific Electrochemical Behavior of Macrotetrolides in Membranes', *Microchem. J.*, **12**, 125–132 (1967).
48. A. Shatkay, 'Ion Specific Membranes as Electrodes in Determination of Activity of Calcium', *Anal. Chem.*, **39**, 1056–1065 (1967).
49. J. Ruzicka, 'The Seventies – Golden Age for Ion-selective Electrodes', *J. Chem. Educ.*, **74**, 167–170 (1997).

50. P. Vadgama, 'Enzyme Electrodes', in *Ion-selective Electrode Methodology*, ed. A.K. Covington, CRC Press, Boca Raton, FL, 23–40, Vol. II, 1979.
51. R.W. Cattrall, 'Heterogeneous-membrane, Carbon-supported, and Coated-wire Ion-selective Electrodes', in *Ion-selective Electrode Methodology*, ed. A.K. Covington, CRC Press, Boca Raton, FL, 131–1730, Vol. I, 1979.
52. G.J. Moody, J.D.R. Thomas, 'Polyvinyl Chloride Matrix Membrane Ion-selective Electrodes', in *Ion-selective Electrode Methodology*, ed. A.K. Covington, CRC Press, Boca Raton, FL, 111–130, Vol. I, 1979.
53. M.A. Arnold, G.A. Rechnitz, 'Optimization of a Tissue-based Membrane Electrode for Guanine', *Anal. Chem.*, **54**, 777–782 (1982).
54. P. Rolfe, M.J. Martin, 'Medical Sensors and Biosensors', *Chem. Br.*, **24**, 1026–1028 (1988).
55. S.M.S. Hassan, G.A. Rechnitz, 'Enzyme Amplification for Trace Level Determination of Pyridoxal 5′-Phosphate with a pCO_2 Electrode', *Anal. Chem.*, **53**, 512–515 (1981).
56. A. Mulchandani, P. Mulchandani, I. Kaneva, W. Chen, 'Biosensor for Direct Determination of Organophosphate Nerve Agents Using Recombinant *Escherichia coli* with Surface-expressed Organophosphorus Hydrolase. 1. Potentiometric Microbial Electrode', *Anal. Chem.*, **70**, 4140–4145 (1998).
57. P. Seegopaul, G.A. Rechnitz, 'Enzyme-amplified Determination of Methotrexate with a pCO_2 Membrane Electrode', *Anal. Chem.*, **56**, 852–854 (1994).
58. S.B. Amemiya, P. Buhlmann, Y. Umezawa, R.C. Jagessar, D.H. Burns, 'An Ion-selective Electrode for Acetate Based on a Urea-functionalized Porphyrin as a Hydrogen-bonding Ionophore', *Anal. Chem.*, **71**, 1049–1054 (1999).
59. T. Katsu, Y. Mori, K. Furuno, Y. Gomita, 'Mexiletine-sensitive Membrane Electrode for Medical Application', *J. Pharm. Biomed. Anal.*, **19**, 585–593 (1999).
60. O.V. Mushik, V.I. Tkach, N.I. Karandeeva, O.I. Glukhova, L.P. Tsyganok, 'Electrochemical and Analytical Properties of Solid-contact Ion-selective Electrodes Reversible to the Imidazole Derivatives Clotrimazole and Bifonazole', *J. Anal. Chem.*, **53**, 1110–1112 (1998).
61. R. Koncki, A. Owczarek, W. Dzwolak, S. Glab, 'Immunoenzymatic Sensitisation of Membrane Ion-selective Electrodes', *Sens. Actuators B*, **47**, 246–250 (1998).
62. Y.M. Issa, M.S. Rizk, A.F. Shoukry, E.M. Atia, 'Plastic Membrane Electrodes for Amprolium', *Mikrochim. Acta*, **129**, 195–200 (1998).
63. S. Dai, J.M. Esson, O. Lutze, N. Ramamurthy, V.C. Yang, M.E. Meyerhoff, 'Bioanalytical Applications of Polyion-sensitive Electrodes', *J. Pharm. Biomed. Anal.*, **19**, 1–14 (1999).
64. P.L. Bailey, *Analysis with Ion-selective Electrodes*, 2nd edition, Heyden, London, 158–181, 1980.
65. S.Z. Yao, X.L. Su, Y.J. Xu, 'New BAW Sensors for Dissolved Gases', *Instrum. Sci. Technol.*, **26**, 473–490 (1998).
66. A. Zangwill, *Physics at Surfaces*, Cambridge University Press, Cambridge, 54–109, 1988.
67. J.O'M. Bockris, A.K.N. Reddy, *Modern Electrochemistry*, 1st Edition, Plenum Press, New York, 667–669, Vol. 2, 1970.
68. R.P. Poeppel, J.M. Blakely, 'Origin of Equilibrium Space Charge Potentials in Ionic Crystals', *Surf. Sci.*, **15**, 507–523 (1969).
69. J. Koryta, 'Theoretical Treatment and Experimental Investigation of Ion Transfer Across Liquid–Liquid Interfaces', in *Ion-transfer Kinetics: Principles and Applications*, ed. J.R. Sandifer, VCH Publishers, New York, 1–17, 1995.
70. J.O'M. Bockris, A.K.N. Reddy, *Modern Electrochemistry*, 1st Edition, Plenum Press, New York, 897–901, Vol. 2, 1970.
71. W.E. Morf, 'Part A – Theory of Membrane Potentials and Membrane Transport', *The Principles of Ion-selective Electrodes and of Membrane Transport*, Elsevier, Amsterdam, 27–161, 1981.
72. R.P. Buck, 'Theory and Principles of Membrane Electrodes', in *Ion-selective Electrodes in Analytical Chemistry*, ed. H. Freiser, Plenum Press, New York, 1–141, Vol. 1, 1978.
73. G. Eisenman, 'The Origin of the Glass-electrode Potential', in *Glass Electrodes for Hydrogen and Other Cations*, ed. G. Eisenman, Marcel Dekker, New York, 133–173, 1967.
74. E. Pungor, 'Working Mechanism of Ion-selective Electrodes', *Pure Appl. Chem.*, **64**, 503–507 (1992).
75. E. Bakker, M. Nägele, U. Schaller, E. Pretsch, 'Applicability of the Phase Boundary Potential Model to the Mechanistic Understanding of Solvent Polymeric Membrane-based Ion-selective Electrodes', *Electroanalysis*, **7**, 817–822 (1995).
76. M.L. Clay, R.E. Pierce, V.Y. Young, S. Hoke, 'A Converging Wall-jet Instrument for Ion Selective Electrode Dynamic Response Studies', *Instrum. Sci. Technol.*, **26**, 461–472 (1998).
77. E. Lindner, K. Toth, E. Pungor, *Dynamic Characteristics of Ion-selective Electrodes*, CRC Press, Boca Raton, FL, 1988.
78. T.R. Berube, R.P. Buck, E. Lindner, K. Toth, E. Pungor, 'Coupled Diffusion/Adsorption Model for Response of Precipitate-based Iodide-selective Electrodes to Primary-ion Activity Steps', *Anal. Chem.*, **63**, 946–953 (1991).
79. E. Bakker, R.K. Meruva, E. Pretsch, M.E. Meyerhoff, 'Selectivity of Polymer Membrane-based Ion-selective Electrodes: Self-consistent Model Describing the Potentiometric Response in Mixed Ion Solutions of Different Charge', *Anal. Chem.*, **66**, 3021–3030 (1994).

80. V.P.Y. Gadzekpo, G.D. Christian, 'Determination of Selectivity Coefficients of Ion-selective Electrodes by a Matched-potential Method', *Anal. Chim. Acta*, **164**, 279–282 (1984).
81. Y. Umezawa, K. Umezawa, H. Sato, 'Selectivity Coefficients for Ion-selective Electrodes: Recommended Methods for Reporting $K^{pot}_{A,B}$ Values', *Pure Appl. Chem.*, **67**(3), 507–518 (1995).
82. M. Nägele, E. Bakker, E. Pretsch, 'General Description of the Simultaneous Response of Potentiometric Ionophore-based Sensors to Ions of Different Charge', *Anal. Chem.*, **71**, 1041–1048 (1999).
83. D.G. Hall, 'Ion-selective Membrane Electrodes: A General Limiting Treatment of Interference Effects', *J. Phys. Chem.*, **100**, 7230–7236 (1996).
84. R.P. Buck, 'The Impedance Method Applied to the Investigation of Ion-selective Electrodes', *Ion-Sel. Electrode Rev.*, **4**, 3–74 (1982).
85. M.J.D. Brand, G.A. Rechnitz, 'Mechanistic Studies on Crystal-membrane Ion-selective Electrodes', *Anal. Chem.*, **42**, 478–483 (1970).
86. R. De Marco, P.C. Hauser, R.W. Catrall, J. Liesegang, G.L. Nyberg, I.C. Hamilton, 'XPS Studies of the Fluoride Ion-selective Electrode Membrane LaF_3: Evidence for a Gel Layer on the Surface', *Surf. Interface Anal.*, **14**, 463–468 (1989).
87. R. De Marco, R.W. Catrall, J. Liesegang, G.L. Nyberg, I.C. Hamilton, 'XPS Studies of the Fluoride Ion-selective Electrode Membrane LaF_3: Ion Interferences', *Surf. Interface Anal.*, **14**, 457–462 (1989).
88. H.A. Hoyen, Jr, J.A. Strozier, Jr, C.-Y. Li, 'Evidence for Space Charge Limited Ionic Transport at the Silver Chloride–Aqueous Solution Interface', *Appl. Phys. Lett.*, **14**, 104–106 (1969).
89. H.A. Hoyen, Jr, J.A. Strozier, Jr, C.-Y. Li, 'Space Charge Limited AC Conduction with Non-blocking Electrodes', *Surf. Sci.*, **20**, 258–268 (1970).
90. R. Rangarajan, G.A. Rechnitz, 'Dynamic Response of Ion-selective Membrane Electrodes', *Anal. Chem.*, **47**, 324–326 (1975).
91. E. Pungor, 'The Theory of Ion-selective Electrodes', *Anal. Sci.*, **14**, 249–256 (1998).
92. R. De Marco, R.W. Cattrall, J. Liesegang, G.L. Nyberg, I.C. Hamilton, 'Surface Studies of the Silver Sulfide Ion-selective Electrode Membrane', *Anal. Chem.*, **62**, 2339–2346 (1990).
93. K. Uosaki, Y. Shigematsu, H. Kita, Y. Umezawa, 'Crystal-face Specific Response of a Single-crystal Cadmium Sulfide Based Ion-selective Electrode', *Anal. Chem.*, **61**, 1980–1983 (1989).
94. W.E. Morf, 'Part B – Ion-selective Electrodes', *The Principles of Ion-selective Electrodes and of Membrane Transport*, Elsevier, Amsterdam, 178–198, 1981.
95. N. Pandya, D.W. Muenow, S.K. Sharma, B.L. Sherriff, 'The Speciation of Water in Hydrated Alkali Silicate Glasses', *J. Non-Cryst. Solids*, **176**, 140–146 (1994).
96. H. Dunken, R.H. Doremus, 'Short Time Reactions of a $Na_2O-CaO-SiO_2$ Glass with Water and Salt Solutions', *J. Non-Cryst. Solids*, **92**, 61–72 (1987).
97. Z. Zhang, N. Soga, K. Hirao, 'Water Corrosion Behavior of Densified Glass. I. Silicate Glass', *J. Non-Cryst. Solids*, **135**, 55–61 (1991).
98. R.P. Buck, 'Transient Electrical Behavior of Glass Membranes. Part I. Theory of D.C. Pulse Processes', *J. Electroanal. Chem.*, **18**, 363–380 (1968).
99. W.M. Mularie, W.F. Furth, A.R.C. Westwood, 'Influence of Surface Potential on the Kinetics of Glass Reactions with Aqueous Solutions', *J. Mater. Sci.*, **14**, 2659–2664 (1979).
100. A.K. Covington, P. Davison, 'Liquid Exchanger Types', in *Ion-selective Electrode Methodology*, ed. A.K. Covington, CRC Press, Boca Raton, FL, 85–110, Vol. I, 1979.
101. I. Nielsen, H. Jorgen, E.H. Hansen, 'Nitrate Ion-selective Electrode', US Pat., 4 059 499, 1977.
102. X. Li, D.J. Harrison, 'Measurement of Concentration Profiles Inside a Nitrite Ion Selective Electrode Membrane', *Anal. Chem.*, **63**, 2168–2174 (1991).
103. A.D.C. Chan, D.J. Harrison, 'NMR Studies of the State of Water in Ion-selective Electrode Membranes', *Anal. Chem.*, **65**, 32–36 (1993).
104. R. Stepánek, B. Kräutler, P. Schulthess, B. Lindemann, D. Ammann, W. Simon, 'Aquocyanocobalt(III) Hepta(2-phenylethyl)cobyrinate as a Cationic Carrier for Nitrite-selective Liquid-membrane Electrodes', *Anal. Chim. Acta*, **182**, 83–90 (1986).
105. D. Midgley, 'Ion-selective Electrodes Based on Siderophores', *J. Chem. Soc., Faraday Trans. I*, **82**, 1187–1193 (1986).
106. E. Bakker, M. Lerchi, T. Rosatzin, B. Rusterholz, W. Simon, 'Synthesis and Characterization of Neutral Hydrogen Ion-selective Chromoionophores for Use in Bulk Optodes', *Anal. Chim. Acta*, **278**, 211–225 (1993).
107. Q. Ye, G. Horvai, A. Tóth, I. Bertóti, M. Botreau, T.M. Duc, 'Studies of Ion-selective Solvent Polymeric Membranes by X-ray Photoelectron Spectroscopy and Time-of-flight Static Secondary Ion Mass Spectroscopy', *Anal. Chem.*, **70**, 4241–4246 (1998).
108. J.R. Sandifer, 'Ion Transfer Into and Ion Transport within Plastic Ion-selective Membranes', in *Ion-transfer Kinetics: Principles and Applications*, ed. J.R. Sandifer, VCH Publishers, New York, 115–138, 1995.
109. W.E. Morf, 'Part B – Ion-selective Electrodes', *The Principles of Ion-selective Electrodes and of Membrane Transport*, Elsevier, Amsterdam, 211–218, 1981.
110. U. Schefer, D. Ammann, E. Pretsch, U. Oesch, W. Simon, 'Neutral Carrier Based Ca^{2+}-selective Electrode with Detection Limit in the Sub-nanomolar Range', *Anal. Chem.*, **58**, 2282–2285 (1986).
111. E. Bakker, M. Willer, E. Pretsch, 'Detection Limit of Ion-selective Bulk Optodes and Corresponding Electrodes', *Anal. Chim. Acta*, **282**, 265–271 (1993).

112. K. Cammann, 'Gas Sensors for CO_2, NH_3, SO_2, NO_2, HF, H_2S, HCN, etc.', *Working with Ion-selective Electrodes*, Springer, New York, 93–98, 1979.
113. U. Oesch, E. Malinowska, W. Simon, 'Bicarbonate-sensitive Electrode Based on Planar Thin Membrane Technology', *Anal. Chem.*, **59**, 2131–2135 (1987).
114. M.E. Meyerhoff, 'Polymer Membrane Electrode Based Potentiometric Ammonia Gas Sensor', *Anal. Chem.*, **52**, 1532–1534 (1980).
115. Y.M. Fraticelli, M.E. Meyerhoff, 'Selectivity Characteristics of Ammonia-gas Sensors Based on a Polymer Membrane Electrode', *Anal. Chem.*, **53**, 1857–1861 (1981).
116. D.M. Pranitis, M.E. Meyerhoff, 'Continuous Monitoring of Ambient Ammonia with a Membrane-electrode-based Detector', *Anal. Chem.*, **59**, 2345–2350 (1987).
117. W.E. Morf, 'Part B – Ion-selective Electrodes', *The Principles of Ion-selective Electrodes and of Membrane Transport*, Elsevier, Amsterdam, 406–413, 1981.
118. J.E. Brady, P.W. Carr, 'Theoretical Evaluation of the Steady-state Response of Potentiometric Enzyme Electrodes', *Anal. Chem.*, **52**, 977–980 (1980).
119. P.R. Burton, 'Instrumentation for Ion-selective Electrodes', in *Ion-Selective Electrode Methodology*, ed. A.K. Covington, CRC Press, Boca Raton, FL, 21–41, Vol. I, 1979.
120. K. Cammann, 'Measurements of Intracellular Ion Activities', *Working with Ion-selective Electrodes*, Springer, New York, 174–178, 1979.
121. B. Jaselskis, C.E. Moore, A. Smolinski, 'Development of the pH Meter', in *Electrochemistry, Past and Present*, eds. J.S. Stock, M.V. Orna, American Chemical Society, Washington, DC, 254–271, 1989.
122. D.A. Skoog, J.J. Leary, 'Operational Amplifiers in Chemical Instrumentation', *Principles of Instrumental Analysis*, 4th edition, Saunders College Publishing, Fort Worth, TX, 10–27, 1992.
123. C.C. Young, 'Evolution of Blood Chemistry Analyzers Based on Ion-selective Electrodes', *J. Chem. Educ.*, **74**, 177–182 (1997).
124. D. Ammann, 'Classification of Ion-selective Electrodes', *Ion-selective Microelectrodes*, Springer, Berlin, 3–8, 1986.
125. D. Ammann, 'General Aspects of Intracellular Measurements of Ions', *Ion-selective Microelectrodes*, Springer, Berlin, 154–193, 1986.
126. K. Venkova, J. Krier, 'Postjunctional α_1- and β-Adrenoceptor Effects of Noradrenaline on Electrical Slow Waves and Phasic Contractions of Cat Colon Circular Muscle', *Br. J. Pharmacol.*, **116**, 3265–3273 (1995).
127. Z. Hu, T. Buehrer, M. Mueller, B. Rusterholz, M. Rouilly, W. Simon, 'Intracellular Magnesium Ion-selective Microelectrode Based on a Neutral Carrier', *Anal. Chem.*, **61**, 574–576 (1989).
128. V. Lyall, T.L. Croxton, W.M. Armstrong, 'Measurement of Intracellular Chloride Activity in Mouse Liver Slices with Microelectrodes', *Biochim. Biophys. Acta*, **903**, 56–67 (1987).
129. C.O. Lee, 'Ionic Activities in Cardiac Muscle Cells and Application of Ion-selective Microelectrodes', *Am. J. Physiol.*, **241**, H459–H478 (1981).
130. M.A. Pineros, J.E. Shaff, L.V. Kochian, 'Development, Characterization, and Application of a Cadmium-selective Microelectrode for the Measurement of Cadmium Fluxes in Roots of Thlaspi Species and Wheat', *Plant Physiol.*, **116**, 1393–1401 (1998).
131. U. Oesch, D. Ammann, W. Simon, 'Ion-selective Membrane Electrodes for Clinical Use', *Clin. Chem.*, **32**, 1448–1459 (1986).
132. D.M. Pranitis, M. Telting-Diaz, M.E. Meyerhoff, 'Potentiometric Ion-, Gas-, and Bio-selective Membrane Electrodes', *Crit. Rev. Anal. Chem.*, **23**, 163–186 (1992).
133. M.E. Collison, M.E. Meyerhoff, 'Chemical Sensors for Bedside Monitoring of Critically Ill Patients', *Anal. Chem.*, **62**, 425A–437A (1990).
134. V.C. Yang, S.C. Ma, M.E. Meyerhoff, 'Thinning Blood Safely', *CHEMTECH*, **23**, 25–32 (1993).
135. M.E. Meyerhoff, V.C. Yang, J.A. Wahr, L.M. Lee, J.H. Yun, B. Fu, E. Bakker, 'Potentiometric Polyion Sensors: New Measurement Technology for Monitoring Blood Heparin Concentrations During Open Heart Surgery', *Clin. Chem.*, **41**, 1355–1356 (1995).
136. T.M. Ambrose, M.E. Meyerhoff, 'Photo-cross-linked Decyl Methacrylate Films for Electrochemical and Optical Polyion Probes', *Anal. Chem.*, **69**, 4092–4098 (1997).
137. C. Espadas-Torre, M.E. Meyerhoff, 'Thrombogenic Properties of Untreated and Poly(ethylene oxide)-modified Polymeric Matrixes Useful for Preparing Intraarterial Ion-selective Electrodes', *Anal. Chem.*, **67**, 3108–3114 (1995).
138. C. Espadas-Torre, V. Oklejas, K. Mowery, M.E. Meyerhoff, 'Thromboresistant Chemical Sensors Using Combined Nitric Oxide Release/Ion Sensing Polymeric Films', *J. Am. Chem. Soc.*, **119**, 2321–2322 (1997).
139. S. Mathison, E. Bakker, 'Renewable pH Cross-sensitive Potentiometric Heparin Sensors with Incorporated Electrically Charged H^+ Ionophores', *Anal. Chem.*, **71**, 4614–4621 (1999).
140. C. Wei, A.J. Bard, G. Nagy, K. Toth, 'Scanning Electrochemical Microscopy. 28. Ion-selective Neutral Carrier-based Microelectrode Potentiometry', *Anal. Chem.*, **67**, 1346–1356 (1995).
141. M.E. Collison, G.V. Aebli, J. Petty, M.E. Meyerhoff, 'Potentiometric Combination Ion–Carbon Dioxide Sensors for In Vitro and In Vivo Blood Measurements', *Anal. Chem.*, **61**, 2365–2372 (1989).
142. M. Telting-Diaz, M.E. Collison, M.E. Meyerhoff, 'Simplified Dual-lumen Catheter Design for Simultaneous Potentiometric Monitoring of Carbon Dioxide and pH', *Anal. Chem.*, **66**, 576–583 (1994).

143. K. Beebe, D. Uerz, J. Sandifer, B. Kowalski, 'Sparingly Selective Ion-selective Electrode Arrays for Multicomponent Analysis', *Anal. Chem.*, **60**, 66–71 (1988).
144. A. Lynch, D. Diamond, P. Lemoine, J. McLaughlin, M. Leader, 'Solid-state Ion-selective Arrays', *Electroanalysis*, **10**, 1096–1100 (1998).
145. R.S. Glass, R.G. Musket, K.C. Hong, 'Preparation of Solid Membrane Chloride Ion Selective Electrodes by Ion Implantation', *Anal. Chem.*, **63**, 2203–2206 (1991).
146. E. Lindner, V.V. Cosofret, T.M. Nahir, R.P. Buck, 'Characterization of Stability of Modified Poly(vinyl chloride) Membranes for Microfabricated Ion-selective Electrode Arrays in Biomedical Applications', *ACS Symp. Ser.*, **556**, 149–157 (1994).
147. K.R. Beebe, B.R. Kowalski, 'Nonlinear Calibration Using Projection Pursuit Regression: Application to an Array of Ion-selective Electrodes', *Anal. Chem.*, **60**, 2273–2278 (1988).
148. R.J. Forster, D. Diamond, 'Nonlinear Calibration of Ion-selective Electrode Arrays for Flow Injection Analysis', *Anal. Chem.*, **64**, 1721–1728 (1992).
149. A. Akhmetshin, V. Baranovsky, A. Akhmetshina, 'Use of Four-factorial Design in Ion-selective Potentiometry for Analysis of Multi-ionic Solutions', *Fresenius' J. Anal. Chem.*, **361**, 282–284 (1998).
150. S.J. West, M.S. Frant, X. Wen, R. Geis, J. Herdan, T. Gillette, M.H. Hecht, W. Schubert, S. Grannan, S.P. Kounaves, 'Electrochemistry on Mars', *Am. Lab.*, **31**(20), 48–54 (1999).

Liquid/Liquid Interfaces, Electrochemistry at

Frédéric Reymond and Hubert H. Girault
Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

1 Introduction 1
2 Theoretical Background 2
2.1 Thermodynamics for the Partition of Ions 2
2.2 Structure of the Interface and Potential Distribution 3
2.3 Charge-transfer Reactions 4
3 Analytical Methods 5
3.1 Cyclic Voltammetry 5
3.2 Dropping-electrolyte Electrode 6
3.3 Micro Interface Between Two Immiscible Electrolyte Solutions 6
3.4 Optical Techniques 7
4 Applications 9
4.1 Amperometric Ion Sensors and Detectors 9
4.2 Analytical Aspects of Metal Extraction 9
4.3 Lipophilicity of Ions 11
5 Conclusion 12
List of Symbols 12
Greek Letters 13
Superscripts 13
Subscripts 13
Abbreviations and Acronyms 13
Related Articles 13
References 13

This article outlines the electrochemical methodology at the interface between two electrolyte solutions (ITIES). The fundamental concepts of the thermodynamics in biphasic systems are presented in order to show how ions are distributed between the two adjacent phases, and hence how a Galvani potential difference is established at an ITIES. Polarizable and nonpolarizable ITIES are then characterized, and it is further evidenced that the classical electroanalytical methodology at a solid electrode can be directly transposed to the ITIES, thereby allowing reversible charge-transfer reactions to be easily monitored and interpreted.

This theoretical approach is completed by a review of the analytical methods used at ITIES, namely cyclic voltammetry, dropping electrolyte electrodes, optical techniques and microinterfaces (which are the biphasic analogs of microelectrodes). The last part deals with the practical applications that electrochemistry at ITIES has attracted during the last decade in the development of amperometric ion sensors and detectors, in the extraction of metal ions by interfacial formation of a complex and in the assessment of the lipophilicity of ionizable drugs.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Electrochemistry at the ITIES is concerned with three main types of charge-transfer reactions, namely:

- ion-transfer reactions from one phase to another
- ion-transfer reactions assisted by the presence of a complexing agent
- electron-transfer reactions between a hydrophilic and a hydrophobic redox couple.

So far, the electroanalytical applications of electrochemistry at liquid/liquid interfaces have been limited to simple or assisted ion transfers,[1] and this article will be restricted to this subject.

The amperometric study of ion-transfer reactions is a relatively recent area which was developed in the 1970s in France by Gavach et al.[2–5] and in Prague by Koryta et al.[6–10] who demonstrated that the ITIES were polarizable in the same way as the interface between a metallic electrode and an electrolyte solution. It ensued that techniques commonly used for the measurement of electron transfer at solid/liquid interfaces could be applied to study transfer processes through the ITIES, that is, cyclic voltammetry,[11] chronoamperometry,[12] polarography,[13] differential-pulse stripping voltammetry[14] and ac voltammetry.[15]

The route to modern studies of both the interfacial structure of the ITIES and of charge-transfer reactions started only with the introduction of the concept of the four-electrode potentiostat with ohmic drop compensation,[6] which permitted potentiostatic control of the interfacial potential difference. Most publications have reported mainly experimental results,[16–26] and few theoretical models of charge-transfer processes[27–35] and of the interfacial structure[36–43] have been proposed recently. Although the potential distribution at the ITIES is now well established,[44] the structure of the interface and the kinetics of transfer are still controversial subjects.[24,45]

Numerous efforts have been made to acquire a better understanding of the molecular mechanisms involved in

ionic motion in liquids,[1,46] and recent results suggest that the rate-limiting step in ion transport is the necessary interchange of the solvation shell from one liquid to the other. The roughness of the interface is likely to manifest itself as capillaries or fingers of one liquid protruding into another.[47,48] This ''fingering'' resulting from the long-range ion–dipole interactions plays a major role in the change of the solvation shell,[49] implying that ion transfer may be an activated process[50] This has been corroborated by molecular dynamics calculations[49,51] which tend to confirm that the interface is not a sharp but an extended region in which the two solvents mix. One of the greatest challenges in the theory of ITIES is to understand the spatial distribution of the driving forces. From an electroanalytical viewpoint, the majority of the charge-transfer reactions studied are reversible (i.e. kinetically fast) and the development of their applications is not hindered by these theoretical limitations.

This article is intended to describe the general theory of charge-transfer reactions at ITIES and to outline the various methodologies that electrochemistry at liquid/liquid interfaces offers for analytical purposes. It does not deal with interfacial structure and kinetics, but focuses on the various applications of ITIES and on recent advances in this field. Also, the very important field of the ion-selective electrode (ISE),[52–55] which is a significant application of electrochemistry at liquid/liquid interfaces is omitted, as it is a subject of its own.

2 THEORETICAL BACKGROUND

2.1 Thermodynamics for the Partition of Ions

2.1.1 *Gibbs Energy of Transfer and Nernst Equation at the Interface Between Two Immiscible Electrolyte Solutions*

The standard transfer Gibbs energy of a species from one phase (say water) to another phase (say the organic solvent), $\Delta G_{\mathrm{t}}^{o,\mathrm{w}\to\mathrm{o}}$, is by definition equal to the difference between the standard Gibbs energy of solvation, $\mu^{o,\mathrm{o}}$, and the standard Gibbs energy of hydration, $\mu^{o,\mathrm{w}}$ (Equation 1):

$$\Delta G_{\mathrm{t}}^{o,\mathrm{w}\to\mathrm{o}} = \mu^{o,\mathrm{o}} - \mu^{o,\mathrm{w}} \tag{1}$$

In the case of an ionic species i, we have to consider the electrochemical potentials which are equal at equilibrium. In developing this equality, we can write Equation (2):

$$\mu_{\mathrm{i}}^{o,\mathrm{w}} + RT\ln a_{\mathrm{i}}^{\mathrm{w}} + z_{\mathrm{i}}F\phi^{\mathrm{w}} = \mu_{\mathrm{i}}^{o,\mathrm{o}} + RT\ln a_{\mathrm{i}}^{\mathrm{o}} + z_{\mathrm{i}}F\phi^{\mathrm{o}} \tag{2}$$

from which we can express the Galvani potential difference between the two phases $\Delta_{\mathrm{o}}^{\mathrm{w}}\phi$ according to Equation (3):

$$\Delta_{\mathrm{o}}^{\mathrm{w}}\phi = \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{i}}^{o} + \frac{RT}{z_{\mathrm{i}}F}\ln\left(\frac{a_{\mathrm{i}}^{\mathrm{o}}}{a_{\mathrm{i}}^{\mathrm{w}}}\right. = \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{i}}^{o'} + \frac{RT}{z_{\mathrm{i}}F}\ln\left(\frac{c_{\mathrm{i}}^{\mathrm{o}}}{c_{\mathrm{i}}^{\mathrm{w}}}\right. \tag{3}$$

where a_{i} and c_{i} are the activity and the concentration of the ion, respectively, in both phases and $\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{i}}^{o}$ and $\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{i}}^{o'}$ are called the standard and the formal transfer potentials, respectively.

This equation is often called the Nernst equation for ion transfer at liquid/liquid interfaces, and the term $\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{i}}^{o}$ corresponds to the standard Gibbs energy of transfer expressed on a voltage scale (Equation 4):

$$\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{i}}^{o} = \frac{\Delta G_{\mathrm{t,i}}^{o,\mathrm{w}\to\mathrm{o}}}{z_{\mathrm{i}}F} \tag{4}$$

It is important to realize that although the Nernst equation for ion transfer resembles the classical Nernst equation for redox reactions on an electrode, there is no redox reaction involved in the definition of Equation (3).

If a ligand or an ionophore able to complex the transferring ion is present in the organic phase, then the complexation equilibrium can be taken into account. In the case of 1 : 1 stoichiometry, the association constant is simply given by Equation (5):

$$K_{\mathrm{a}} = \frac{a_{\mathrm{ML}}^{\mathrm{o}}}{a_{\mathrm{M}}^{\mathrm{o}}a_{\mathrm{L}}^{\mathrm{o}}} \tag{5}$$

where M refers to the transferring ion and L to the ligand. The Galvani potential difference now reads as Equation (6):

$$\begin{aligned}\Delta_{\mathrm{o}}^{\mathrm{w}}\phi &= \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{M}}^{o} + \frac{RT}{z_{\mathrm{M}}F}\ln\left(\frac{a_{\mathrm{M}}^{\mathrm{o}}}{a_{\mathrm{M}}^{\mathrm{W}}}\right.\\ &= \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{ML}}^{o} + \frac{RT}{z_{\mathrm{M}}F}\ln\left(\frac{a_{\mathrm{ML}}^{\mathrm{o}}}{a_{\mathrm{M}}^{\mathrm{W}}}\right.\end{aligned} \tag{6}$$

with the apparent standard transfer potential given by Equation (7):

$$\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{ML}}^{o} = \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{M}}^{o} - \frac{RT}{z_{\mathrm{M}}F}\ln\left(K_{\mathrm{a}}a_{\mathrm{L}}^{\mathrm{o}}\right) \tag{7}$$

This equation shows that the presence of an ionophore in the organic phase can shift the apparent solvation energy and therefore facilitate the transfer of hydrophilic species from water to oil. We shall call this type of ion-transfer reaction facilitated ion transfer.

2.1.2 *Polarizable and Non-polarizable Interface Between Two Immiscible Electrolyte Solutions*

If a salt such as tetrabutylammonium bromide is dissolved in two immiscible solvents in contact, the distribution

of the salt induces a polarization of the interface. The resulting Galvani potential difference is then called a distribution potential and is defined by applying Equation (3) to both the cation and the anion (Equation 8):

$$\Delta_o^w\phi = \Delta_o^w\phi_+^o + \frac{RT}{F}\ln\left(\frac{a_+^o}{a_+^w}\right. = \Delta_o^w\phi_-^o - \frac{RT}{F}\ln\left(\frac{a_-^o}{a_-^w}\right. \quad (8)$$

In the case of dilute solutions, this equation simplifies to Equation (9):

$$\Delta_o^w\phi = \tfrac{1}{2}(\Delta_o^w\phi_+^o + \Delta_o^w\phi_-^o) \quad (9)$$

This simple example illustrates that as soon as we partition salts between two adjacent phases, the interface becomes polarized at a fixed potential defined by the standard transfer potentials of the different ionic species. Because this polarization potential is fixed, we shall say that the interface is nonpolarizable in the sense that it is not possible to polarize the interface without modifying the chemical composition of the two phases.

In the case where a hydrophilic salt is dissolved in water and a hydrophobic salt is dissolved in the organic phase such that the concentration of the hydrophilic salt in the organic phase is negligible compared with that of the hydrophobic salt and, conversely, the concentration of the hydrophobic salt in water is negligible compared with that of the hydrophilic salt, then the interface will be called polarizable. This definition means that it is now possible to polarize the interface from an external potential source without modifying the chemical composition of the adjacent phases. In this way, there is what is called a potential window such that it is possible to polarize the interface up to a point where the applied Galvani potential difference is enough for an ion to transfer.

To illustrate the principle of the potential window, let us consider an interface between an aqueous solution of Li_2SO_4 and a solution of tetrabutylammonium (TBA^+) tetraphenylborate (TPB^-) in an organic solvent, e.g. 1,2-dichloroethane (1,2-DCE).

At the potential of zero charge, the two adjacent phases are by definition uncharged. If a positive Galvani potential difference (water vs oil) is applied from an external source, two back-to-back Gouy–Chapman diffuse layers will be established with an excess of cations in the aqueous phase and an excess of anions in the organic phase.[20,46,56] As shown in Figure 1, we can polarize the interface until the Galvani potential difference reaches the standard transfer potential of either Li^+ or TPB^-. As it happens, the standard transfer potential of TPB^- is less than that of Li^+ ($\Delta_o^w\phi_{TPB^-}^o = 340\,mV$ and $\Delta_o^w\phi_{Li^+}^o = 580\,mV$), and TPB^- then starts to transfer as soon as the Galvani potential difference approaches 200 mV. The chemical composition of the adjacent phases is then altered by the Faradaic current across the interface and the redox reactions at the two electrodes connected to the external source. If instead of using TPB^- as the organic anion we choose a more hydrophobic anion such as tetrakis(4-chlorophenyl)borate ($TPBCl^-$) for which the standard transfer potential is very large, then a positive polarization of the interface will result in the transfer of Li^+ from water to oil (see Figure 1). It can be concluded that $TPBCl^-$ is more hydrophobic than Li^+ is hydrophilic.

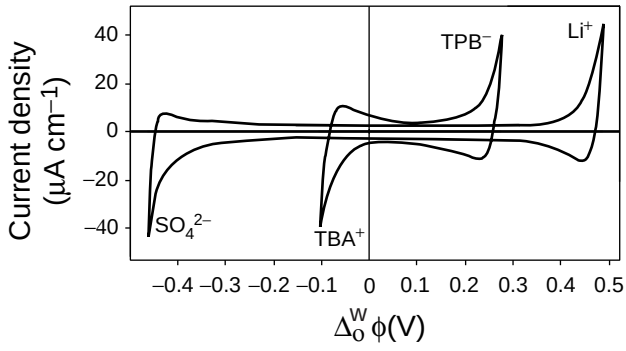

Figure 1 Potential window for 1,2-DCE/water systems containing 5 mM Li_2SO_4 as aqueous supporting electrolyte and 5 mM tetrabutylammonium tetraphenylborate (TBATPB) or 5 mM bis(triphenylphosphoranylidene) ammonium tetrakis(4-chlorophenyl borate) (BTTPATPBCl) as organic supporting electrolyte.

When a negative polarization is applied, then the interface is polarized until the Galvani potential difference reaches the standard transfer potential of either TBA^+ or SO_4^{2-} ($\Delta_o^w\phi_{TBA^+}^o = -230\,mV$ and $\Delta_o^w\phi_{SO_4^{2-}}^o < -600\,mV$, respectively). Since the standard transfer potential of SO_4^{2-} is more negative than that of TBA^+, the potential window is limited by the transfer of TBA^+. However, if we use in the organic phase a more hydrophobic cation such as bis(triphenylphosphoranylidene) ammonium ($BTPPA^+$), which has a very negative standard transfer potential, then the potential window is limited by the transfer of SO_4^{2-}. Again, it can be concluded that $BTPPA^+$ is more hydrophobic that sulfate is hydrophilic.

Standard Gibbs energies of ion transfer have been tabulated in different reviews.[22,57–59]

2.2 Structure of the Interface and Potential Distribution

A liquid/liquid interface is by definition a molecular interface between two condensed media. The solvent dynamics results in an interface fluctuating to a certain limit following the capillary wave theory. If we take a time average view of the interface, we can say that it is composed of a thin (i.e. about 1 nm) mixed solvent layer. Snapshots from molecular dynamics computer simulation show that the local structure of the interface is greatly influenced by the presence of ionic charges.[38,47,49,51,60,61]

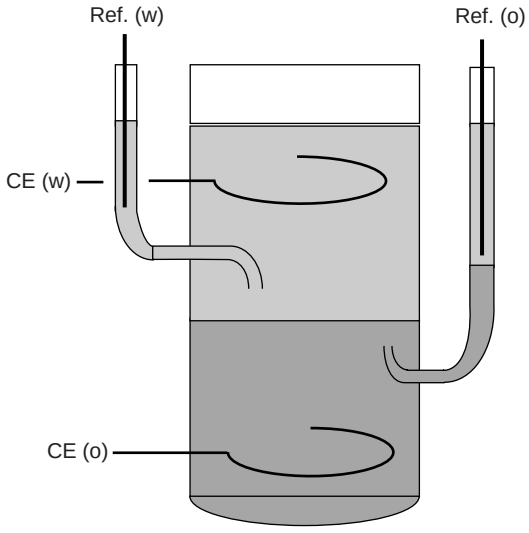

Figure 2 Schematic diagram of an electrochemical cell for the study of charge-transfer reactions at an ITIES. For the system in Figure 1, the aqueous reference electrode can be $Ag|Ag_2SO_4$. The organic reference electrode includes an unpolarized liquid junction using the organic cation or anion as the common ion. If the organic supporting electrolyte is TBATPB, the liquid junction can comprise either tetrabutyl ammonium chloride (TBACL) or sodium tetrabutylborate (NaTPB) so that the reference electrode can be $TBATPB_o|TBACL_w|AgCl|Ag$ or $TBATPB_o|NaTPB|AgTPB|Ag$.

As discussed above, the interface between two immiscible electrolytes can be polarized using an external source. From an experimental viewpoint, it is usual to operate with a four-electrode potentiostat comprising two reference electrodes and two counter electrodes to provide the current,[6,11] as illustrated in Figure 2.

At the interface, the polarization is distributed over the two back-to-back Gouy–Chapman diffuse layers. One of the key unresolved issues is the dependence of the interfacial electric field on the applied polarization. At the point of zero charge, it is usually accepted that there is no specific orientation of dipoles resulting from the molecular interactions between the solvent molecules. However, it is difficult to estimate if the applied polarization influences the orientation of the solvent molecules and to quantify the strength of the field. Some authors choose, as in bioelectrochemistry, a constant-field approach to model the polarized interface, as shown in Figure 3.

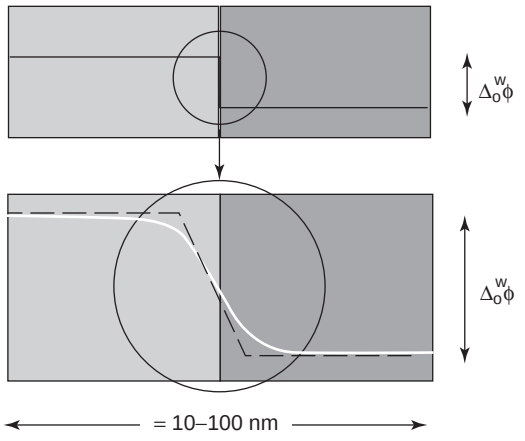

Figure 3 Potential distribution for a positive polarization (water vs oil). The top curve is a macroscopic representation, whereas the bottom curve illustrates the potential drop across the two back-to-back diffuse layers. The dashed line represents the constant-field model.

2.3 Charge-transfer Reactions

2.3.1 Simple Ion Transfer

From a practical viewpoint, the kinetics of ion transfer can be considered as very fast, such that it can be assumed that the surface concentrations always follow the Nernst Equation (4). In electrochemical nomenclature, it is then said that ion-transfer reactions are reversible. Similarly to a reversible redox reaction on an electrode that is limited by the mass transfer of the reactants to the electrode and by that of the products away from the electrode, an ion-transfer reaction is limited by the mass transfer of ions to the interface and away from it. Hence the mass transport differential equations and boundary conditions are similar in both cases, and all the electroanalytical methodology can therefore be transposed to the study of ion transfer reactions.

As in classical amperometry, the response of the system stems from the resolution of the diffusion equations of the ion in the two adjacent phases (Equation 10):

$$\frac{\partial c_i^w}{\partial t} = D_i^w \frac{\partial^2 c_i^w}{\partial x^2} \quad \text{and} \quad \frac{\partial c_i^o}{\partial t} = D_i^o \frac{\partial^2 c_i^o}{\partial x^2} \tag{10}$$

By taking the interface as the origin, the current is then simply given by the flux of i across the interface of area A (Equation 11):

$$I = z_i F A \left(\frac{\partial c_i^w}{\partial x} \right)_{x=0} \tag{11}$$

The boundary conditions are the Nernst Equation (4) and the equality of the fluxes (Equation 12):

$$D_i^w \left(\frac{\partial c_i^w}{\partial x} \right)_{x=0} + D_i^o \left(\frac{\partial c_i^o}{\partial x} \right)_{x=0} = 0 \tag{12}$$

2.3.2 Facilitated Ion Transfer

Assisted ion-transfer reactions are also very fast and can be considered reversible in most cases. However, mass

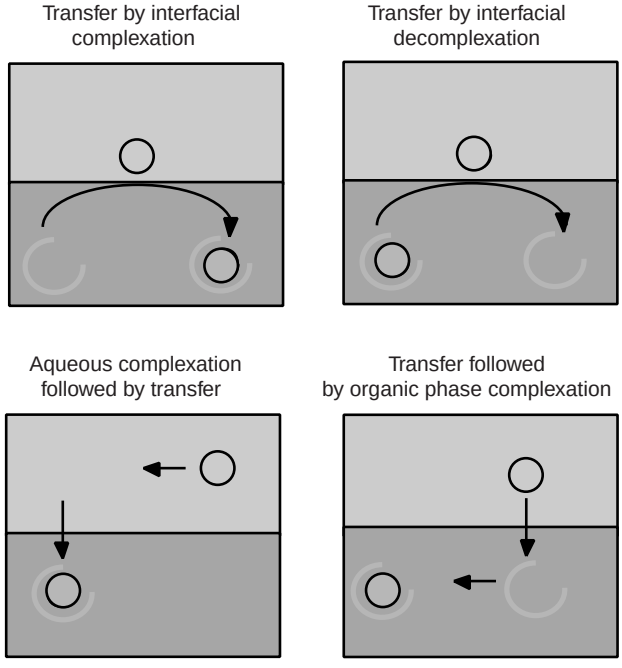

Figure 4 Schematic representation of assisted ion-transfer reactions.

transport is more complicated and, as shown schematically in Figure 4, we can distinguish four types of reactions:[62]

- aqueous complexation followed by transfer (ACT)
- transfer by interfacial complexation (TIC)
- transfer by interfacial decomplexation (TID)
- transfer followed by organic-phase complexation (TOC).

In each case, we have to consider the mass transport of the ions and of the ligands to and away from the interface. This leads to many specific cases, which have been treated in the literature (see section 4.2). More interesting from an electroanalytical aspect are the two limiting cases where either the ions or the ligands are in excess in their respective phases. In these two cases, the mass transport is limited by that of the ligand or that of the ions, respectively.

Assisted, or facilitated, ion transfer was first reported in 1979 by Koryta,[7] who observed that the transfer of potassium and sodium ions in the aqueous phase was facilitated by the formation of a complex in the organic phase with the synthetic polyether dibenzo-18-crown-6 (DB18C6) and with the natural antibiotic valinomycin, respectively. This work was a landmark in ITIES research, because it meant that both potential window-limiting species and neutral ionophore molecules were amenable to study. As a result, this field spread quickly to solvent extraction and purification, detection of trace ions, assisted transfer of proton and development of amperometric sensors (see also section 4).

3 ANALYTICAL METHODS

3.1 Cyclic Voltammetry

Cyclic voltammograms produced by reversible ion-transfer reactions are similar to those obtained for reversible electron-transfer reactions at a metal/electrolyte solution interface,[63] as shown in Figure 5. Thus, for reversible transfer reactions of an ion across a large (or planar) interface, the maximum forward peak current I_p^{FWD} may then be expressed by the Randles–Sevcik Equation (13):[64]

$$I_p^{FWD} = 0.4463\, z_i F A c_i^w \sqrt{\frac{z_i F D_i^w v}{RT}} \qquad (13)$$

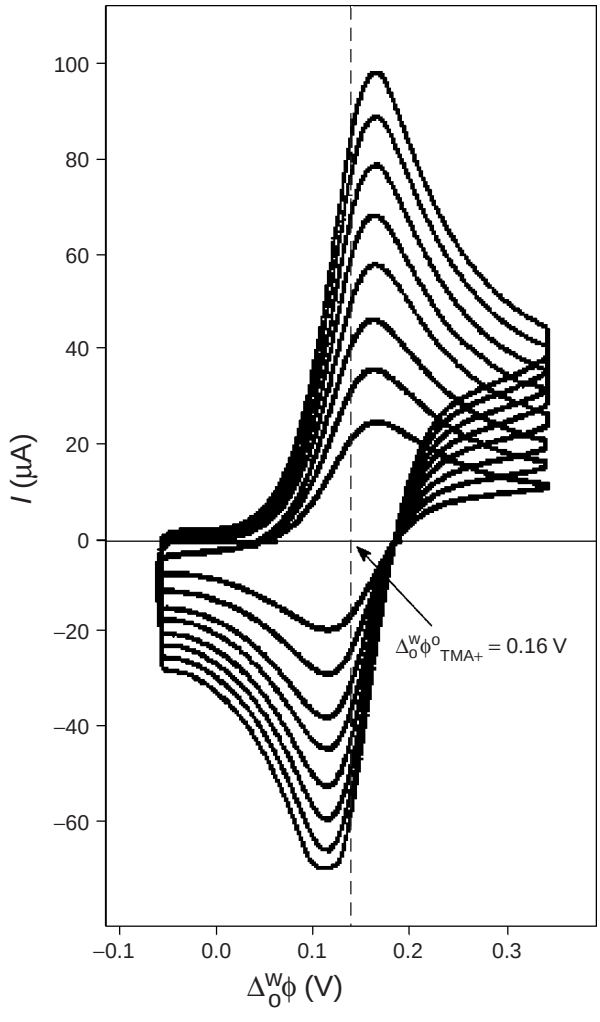

Figure 5 Cyclic voltammogram for the transfer of tetramethylammonium across the water/1,2-DCE interface at different sweep rates (the interfacial area is 1.13 cm^2).

where A is the interfacial area, v the rate of a potential sweep and c_i^w the aqueous bulk concentration of i.

Together with the evaluation of the diffusion coefficient of the transferring ion, the determination of the formal transfer potential of an ion (and thus of its Gibbs energy of transfer) is the most important application of cyclic voltammetry. For a reversible ion-transfer reaction at a large planar interface, $\Delta_o^w\phi_i^{o'}$ may be expressed in terms of the half-wave potential, $\Delta_o^w\phi_{i,1/2}$, by Equation (14):

$$\Delta_o^w\phi_{i,1/2} = \Delta_o^w\phi_i^{o'} + \frac{RT}{2z_iF}\ln\left(\frac{D_i^w}{D_i^o}\right) \quad (14)$$

Experimentally, $\Delta_o^w\phi_{i,1/2}$ is considered equal to the mid-peak potential, and is directly deduced from the voltammograms. However, as the diffusion coefficients in the organic phase are rarely known because of experimental difficulties, the ratio D_i^o/D_i^w in Equation (14) is usually approximated to the inverse ratio of the solvent viscosities, η, by application of Walden's rule[64] (Equation 15):

$$\frac{D_i^o}{D_i^w} = \frac{\eta^w}{\eta^o} \quad (15)$$

Finally, for a reversible charge transfer, the classical following conditions also apply:

- the peak potentials must be independent of the scan rate
- the peak-to-peak separation is $59/z_i$ mV at 25 °C.

3.2 Dropping-electrolyte Electrode

In the same way that the mercury-drop electrode permits very reproducible measurements owing to the continuous renewal of the interface for every drop, the dropping-electrolyte electrode allows the study of both ion-transfer and facilitated ion-transfer reactions.[13,65–77] This approach has been widely used to study facilitated ion transfer. The principle consists in either dropping down or dropping up an electrolyte solution and recording the current as in polarography. The methodology of polarography can then be transposed directly.

3.3 Micro Interface Between Two Immiscible Electrolyte Solutions

3.3.1 Micropipets and Microholes

Microelectrodes benefit from diffusion fields controlled by the geometry of the interface and from reduced ohmic loss. Because studies at ITIES always involve the use of an organic solvent, much effort has been dedicated to supporting micro liquid/liquid interfaces. Using the

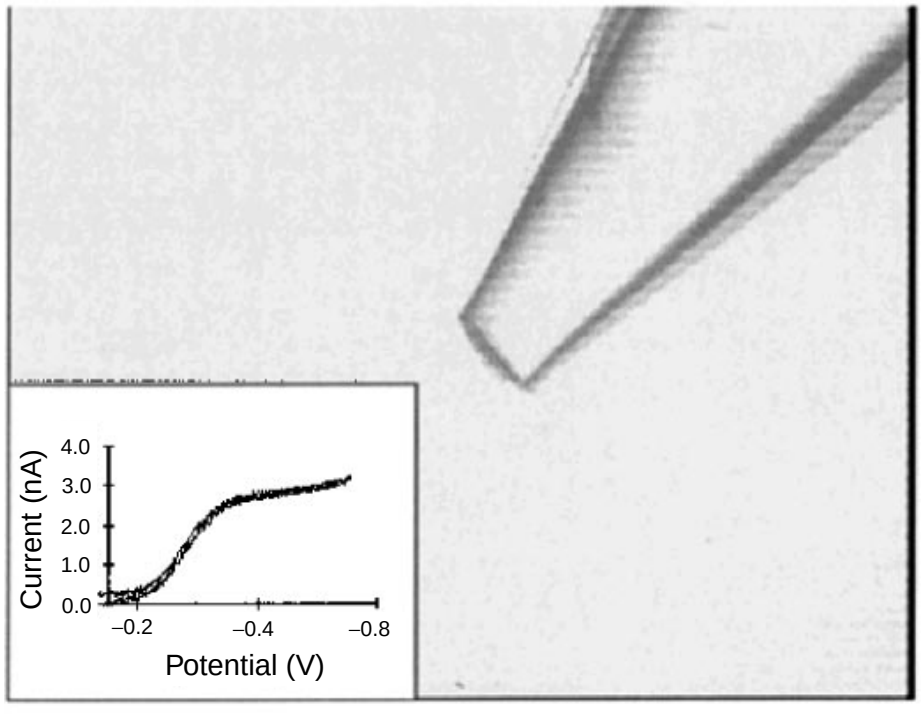

Figure 6 Video micrograph of a 15.5 μm radius micropipet filled with an aqueous KCl solution and immersed in a 1,2-DCE solution of DB18C6. No external pressure was applied, and the micro ITIES is flat. The insets show the corresponding steady-state voltammograms of facilitated transfer of potassium. (Reprinted with permission from Y. Shao, M.V. Mirkin, 'Voltammetry at Micropipet Electrodes', *Anal. Chem.*, **70**, 3155–3161 (1998). Copyright 1998 American Chemical Society.)

micropipet technology developed for electrophysiology, liquid/liquid interfaces were first supported at the tip of micropipets, the water phase being usually located inside the pipet.[62,78–87] Micropipets provide an asymmetry of diffusion fields. Ingress motion of ion occurs by pseudo-spherical diffusion to a microdisk interface whereas egress motion out of the pipet occurs by linear diffusion. This yields asymmetric voltammograms as shown in Figure 6. The ingress current reaches a steady-state value proportional to the radius, the diffusion coefficient and to a geometric factor.

Another approach is to support micro liquid/liquid interfaces in a microhole in a thin polymer film or in a silicon structure.[88–95] A technique often used to micromachine polymers is UV laser photo-ablation,[96,97] whereas that used to machine silicon is anisotropic etching. For microinterfaces supported in microholes, the thinner is the supporting film the more symmetric are the diffusion fields. For very thin films, the mass transport equations are similar to those for a microdisk electrode of radius r.[88,93] In this case, the half-wave potential is given by Equation (16):

$$\Delta_o^w\phi_{i,1/2} = \Delta_o^w\phi_i^{o'} + \frac{RT}{z_iF}\ln\left(\frac{D_i^w}{D_i^o}\right. \qquad (16)$$

and the steady-state current by Equation (17):

$$I_{SS} = 4z_iFD_i^wc_i^wr \qquad (17)$$

With thicker films, however, the microhole supporting the ITIES modifies the mass transport, and both the ingress and egress geometries in and out of the microhole interface must be introduced.[98,99] Following the shape given in Figure 7, the expression of the half-wave potential then becomes Equation (18):[100]

$$\Delta_o^w\phi_{i,1/2} = \Delta_o^w\phi_i^{o'} + \frac{RT}{z_iF}\ln\left(\frac{D_i^w}{D_i^o}\cdot\frac{h^w+\pi r/4}{h^o+\pi r/4}\right. \qquad (18)$$

The main advantage of working with microliquid/liquid interfaces stems from the low *IR* drop, which provides clean voltammetric responses.

3.3.2 Ionodes

The major barriers to the electroanalytical exploitation of amperometry at liquid/liquid interfaces have been the mechanical stability of the interface and the resistivity of the organic phase. To circumvent these difficulties, the first approach has been to ''gelify'' one of the phases (usually the organic phase) using the methodology developed to produce polymer membranes for the design of ISEs [see Figures 8 and 9a and b].

Recently, new composite membranes comprising an organic electrolyte gel supported on a thin polymer

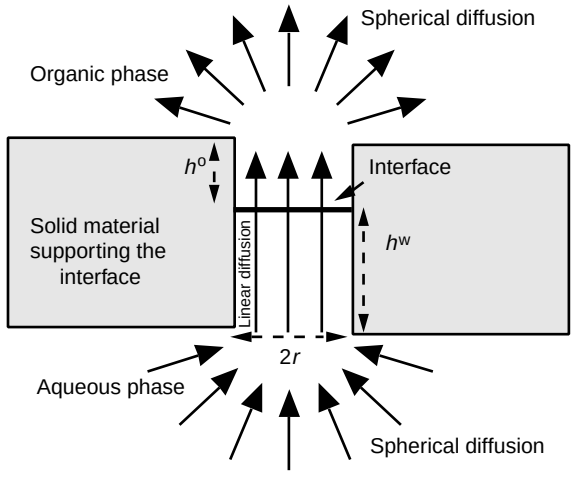

Figure 7 Schematic diagram showing the diffusion processes of an ion passing from water to the organic phase through a microhole supported liquid/liquid interface. r is the radius of the microhole, h^w the penetration depth of the aqueous phase and h^o that of the organic phase into the microhole.

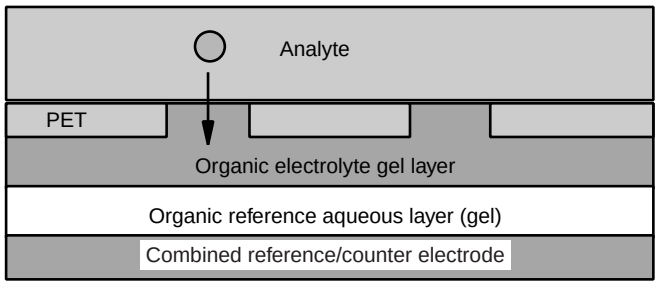

Figure 8 Simplified diagram showing the common composition of an ionode. The interface is supported in a microhole array pierced through poly(ethylene terephthalate) (PET) by laser photoablation.

micromachined so as to include a regular array of microholes have been developed and named ionodes for simplicity.[101,102]

3.4 Optical Techniques

Optical techniques are interesting alternatives to classical electrochemical methods, because they have the advantage of providing in situ spectroscopic measurements at liquid/liquid interfaces. Two main types of experiments (namely voltfluorimetry or voltabsorptometry on the one hand and sum-frequency generation (SFG) or second harmonic generation (SHG) on the other) are commonly conducted at ITIES depending on whether linear or, non-linear optics, respectively, are used. For simplicity, however, only the linear optical techniques are presented here, and readers interested in SHG or SFG techniques should consult reviews.[103–109]

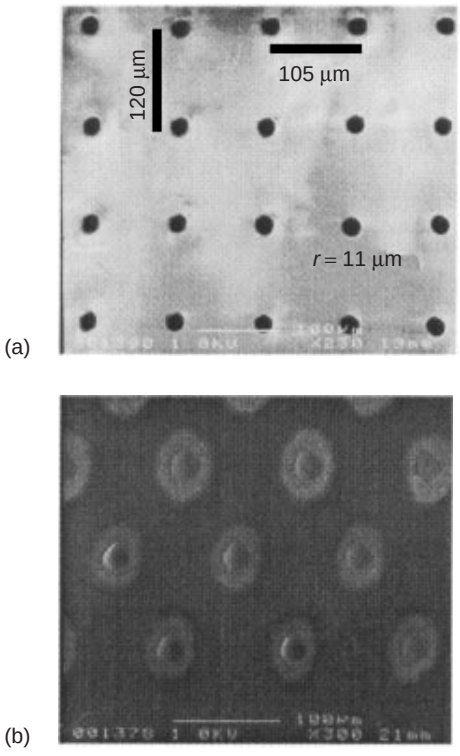

Figure 9 Scanning electron micrographs of the entrance-side holes in a PET film. (a) Empty holes; (b) array filled with poly(vinyl chloride) (PVC) at 70 °C. (Adapted with permission from H.J. Lee, P.D. Beattie, B.J. Seddon, M.D. Osborne, H.H. Girault, 'Amperometric Ion Sensors Based on Laser-patterned Composite Polymer Membranes', *J. Electroanal. Chem.*, **440**, 73–82 (1997). Copyright 1998 Elsevier Science.)

Ultraviolet/visible (UV/VIS) absorption measurements can be carried out when the incident light beam impinges on the interface from the phase of larger optical index in total internal reflection (TIR) geometry, i.e. with an incidence angle greater than the critical angle, so that the transmitted light wave in the adjacent phase cannot propagate. Voltabsorptometry measures then the changes of light intensity of the reflected beam due to absorption of charge-transfer products. Assuming that the absorbance A_{TIR} is proportional to the integral of the bulk concentration of the absorbing species, its time derivative is again proportional to the Faradaic current.[110] As an example, Figure 10(a) and (b) show the evolution of the absorbance in 1,2-DCE upon transfer of methyl orange. The evolution of the UV/VIS spectra follows the current change in the corresponding voltammogram. The band at 420 nm increases monotonically during the forward sweep owing to deprotonated methyl orange transfer from water to 1,2-DCE, and decreases only after the corresponding isobestic point after which the transfer is inverted.

This technique has also been used to study Cu(II) transfer assisted by 6,7-dimethyl-2,3-di(2-pyridyl)quinoxaline,

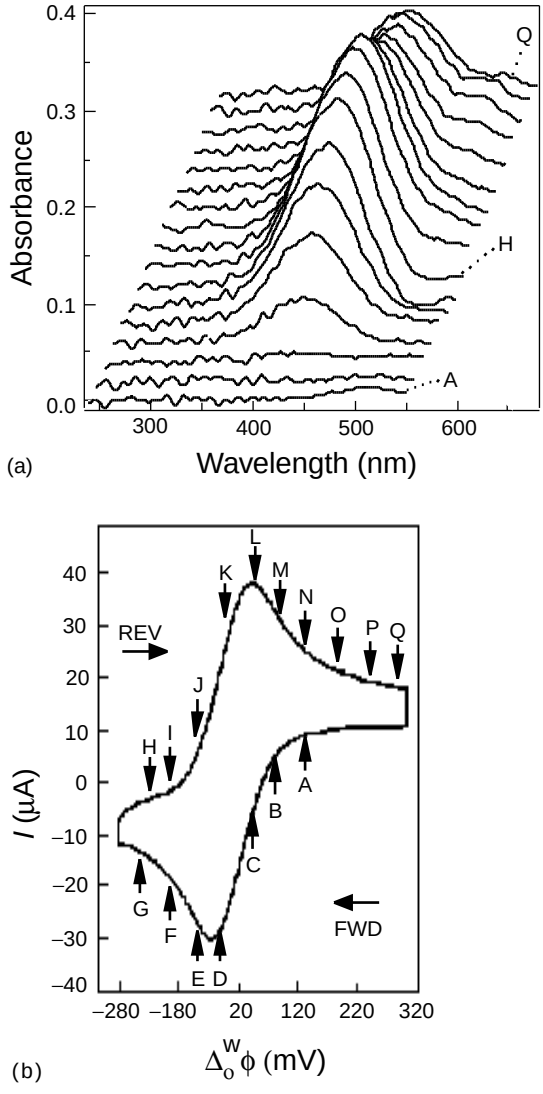

Figure 10 In situ UV/VIS spectra (a) and cyclic voltammogram (b) for the transfer of deprotonated methyl orange from water to 1,2-DCE at a scan rate of 36 mV s^{-1}. (Reprinted from Z. Ding, F. Reymond, P. Baumgartner, D. Fermin, P.-F. Brevet, P.-A. Carrupt, H.H. Girault, 'Mechanism and Dynamics of Methyl and Ethyl Orange Transfer across the Water/1,2-Dichloroethane Interface', *Electrochim. Acta*, **44**, 3–13 (1998). Copyright 1998, with permission from Elsevier Science.)

the absorption spectrum of which changes when it forms a complex. The transfer of rose bengal and eosin B has been studied by voltfluorometry,[111–114] where, similarly to UV/VIS experiments, the change in fluorescence intensity from an exciting beam is monitored as a function of applied interfacial potential.

The advantage of fluorescence and luminescence measurements is that the technique is extremely sensitive, and that small changes in interfacial concentration can be readily monitored. In contrast to cyclic voltammetry, these techniques are totally specific to the transferring

species and are not influenced by the transfer of undesired species, such as supporting electrolyte ions.[113,115] Furthermore, they are insensitive to double-layer charging and ohmic drop, so that chronoabsorptometry and chronofluorimetry[112,114] studies proved to be very attractive in measuring the kinetics of ion-transfer,[115–119] for which reliable data were previously difficult to obtain. Other spectroscopic techniques are also of interest in this field, such as time-resolved laser-induced fluorescence for monitoring the lifetimes of excited species[120] and potential-modulated reflectance spectroscopy,[117] where the transfer kinetics of a species are estimated by measuring its frequency-dependent absorption following an ac potential perturbation.

4 APPLICATIONS

4.1 Amperometric Ion Sensors and Detectors

Nonredox ionic species can be detected amperometrically by measuring the current associated with ion-transfer reactions across a polarized ITIES.[19,24] However, the major difficulties in designing transducers with liquid/liquid systems stem from the mechanical instability of the ITIES and from the resistive nature of the organic phase. As can be deduced from section 3.3, these two problems can be circumvented by using microinterfaces such as micropipets and microholes in thin polymer films or by gelifying one or two of the phases (although the diffusion coefficient of ions in a gel is much reduced).[121–123]

Plasticized polymers have been used to solidify the organic phase[121,124–128] and efficient sensors for sodium,[129–131] ammonium,[132] choline[133] and urea[134–136] have been obtained by immobilizing DB18C6 within the organic phase to facilitate the transfer of Na^+ and NH_4^+, which were detected amperometrically. Moreover, micro liquid/liquid interface arrays[93] have been developed for the assay of urea[89] and creatinine,[91] where the transducer relied again on the amperometric detection of ammonium.

Amperometric detection of choline using either direct ion-transfer[137] or stripping ion-transfer reactions[138] has been achieved with a water/*o*-nitrophenyl octyl ether (NPOE)–PVC gel interface supported on an array of microholes. Otherwise, the electrochemical behavior of NPOE gelified with 1,3 : 2,4-dibenzylidene sorbitol (DBS) has been characterized[139] and a NPOE–PVC gel has been used to manufacture an amperometric sensor for alkali metal ions by incorporation in the organic gel of ionophores such as DB18C6 and valinomycin,[140] which has been shown to facilitate discriminantly the transfer of lithium, sodium, ammonium and potassium.[101,102] This principle has also been used to detect lithium in samples such as blood serum that contains a large excess of sodium (dibenzyl-14-crown-4 and *o*-nitrophenyl phenyl ether were used as ionophore and organic phase, respectively).[141] These last two results are very promising, because they constitute a powerful alternative to conductimetric detection of nonredox species in ion-exchange chromatography,[142] which suffers from a lack of selectivity compared with optical or direct amperometric detectors (see Figure 11a and b).

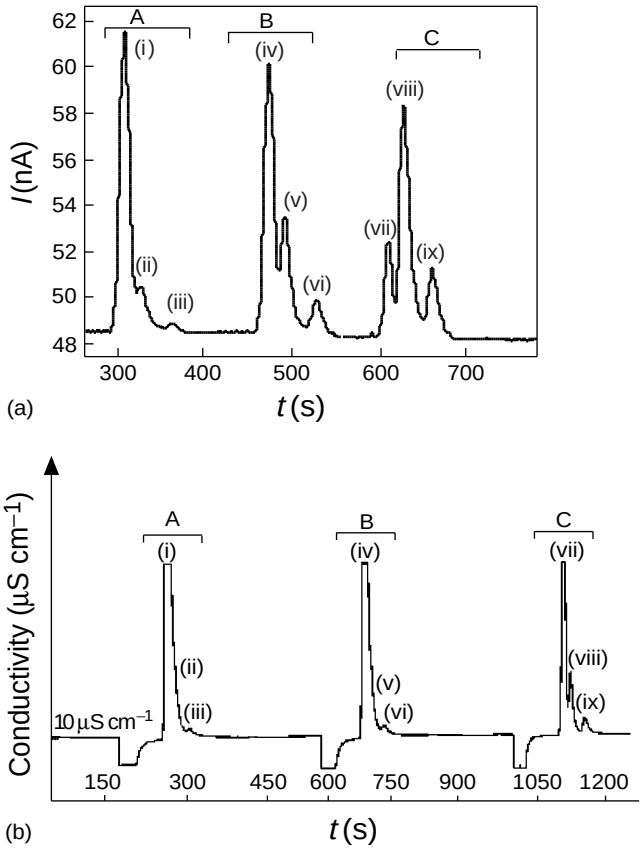

Figure 11 (a) Pulse amperograms recorded for monovalent cations facilitated by 1 mM valinomycin using a PVC–NPOE microgel membrane and tetrabutylammonium cation as organic supporting electrolyte. (b) Chromatogram obtained for the monovalent cations based on a conductimetric detector. The concentrations of the standard solutions are (A) (i) 100 ppm Na^+, (ii) 0.5 ppm NH_4^+ and (iii) 0.5 ppm K^+; (B) (iv) 100 ppm Na^+, (v) 1 ppm NH_4^+ and (vi) 1 ppm K^+; and (C) (vii) 20 ppm Na^+, (viii) 4 ppm NH_4^+ and (ix) 4 ppm K^+. Flow rate, 0.85 mL min^{-1}; 5 mM tartaric acid is used as the eluent. (Reprinted with permission from H.J. Lee, H.H. Girault, 'Amperometric Ion Detector for Ion Chromatography', *Anal. Chem.*, **70**, 4280–4285 (1998). Copyright 1998 American Chemical Society.)

4.2 Analytical Aspects of Metal Extraction

The chemical analysis of liquid solutions has for a long time been a very important domain of analytical chemistry.[143] Numerous applications in biology,

medicine and environmental studies require the sensitive and selective detection of chemical constituents in different media. Heavy metals are an example where the need for a precise analytical tool is of major importance owing to their high toxicity towards life in general and people in particular.

Electrochemistry at the ITIES has attracted much attention in this field during the last decade, because it provides a simple way of measuring the stoichiometry and the association constants of ion–ionophore complexes in organic solvents.[140,144,145] With the rapid development of coordination chemistry, numerous ligands with specific binding properties have become available, extending complexation reactions where the transfer of a monocharged metal ion is facilitated by the formation of a complex of 1 : 1 ion-to-ligand stoichiometry to more complicated systems.

Theoretical studies on facilitated ion-transfer reactions now provide a reliable framework to analyze experimental data and to determine the physicochemical parameters governing complexation reactions at ITIES. Matsuda et al.[146] published a general theoretical equation for the polarographic response of reversible facilitated ion-transfer reactions, leading to a prediction of the half-wave potential dependence on the initial concentrations of both the metal and the ionophore (denoted $c_{\mathrm{M}init}$ and $c_{\mathrm{L}init}$). In ligand excess, the transfer is limited by the diffusion of the free metal ions towards the interface, whereas it is limited by the diffusion of the complex away from the interface in metal excess. Between these two limiting cases, a mixed diffusion regime is established and, for 1 : 1 stoichiometry, the authors determined criteria separating these three regions in which the dependence of $\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{1/2}$ on the initial concentrations changes (see Figure 12a and b). They further described a simple method to calculate the association constants, and this work has been widely used to interpret both polarographic and cyclic voltammetric experiments.

Few studies have been carried out to model cyclic voltammetric experiments[147–150] and the approach followed by Matsuda et al. has recently been generalized to 1 : m ion-to-ligand stoichiometries, showing that variations of $c_{\mathrm{M}init}$ and $c_{\mathrm{L}init}$ do not lead to a similar evolution of $\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{1/2}$.[151,152] This is illustrated in Figure 12(a) and (b) and demonstrated by the relationships obtained for the TIC, TID and TOC mechanisms (Equations 19–21):

when $c_{\mathrm{L}init} \gg c_{\mathrm{M}init}$:

$$\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{1/2} = \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{M}^{z+},1/2} - \frac{RT}{zF}\ln \sum_{j=0}^{m} \beta_j^{\mathrm{o}} (c_{\mathrm{L}init})^j \qquad (19)$$

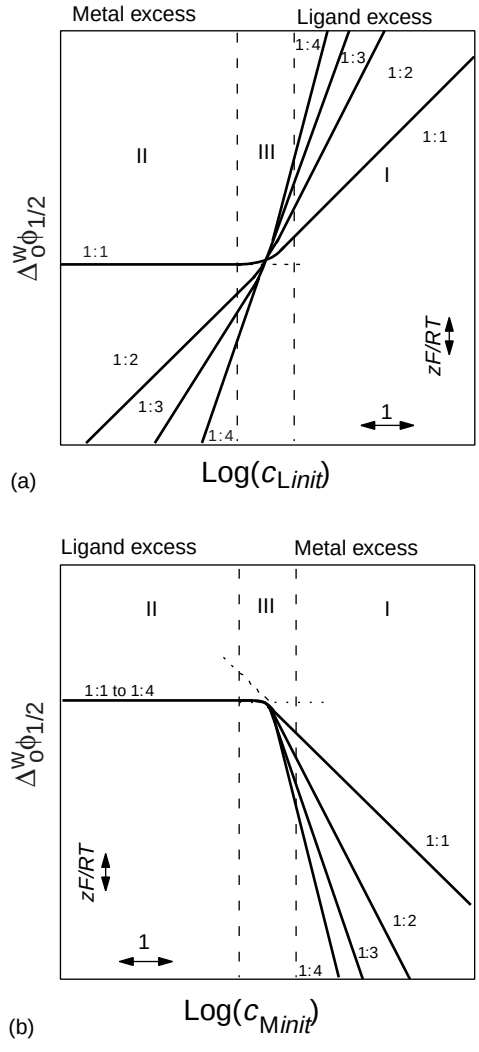

Figure 12 Schematic diagram showing the half-wave potential dependence on (a) the initial ligand concentration and (b) the initial metal concentration for TIC/TID facilitated ion-transfer reactions of 1 : 1 to 1 : 4 ion-to-ligand stoichiometry. An intersection point has been chosen arbitrarily to facilitate comparison between the various stoichiometries. These schemes show that these curves do not have equal slopes in metal or ligand excess and that a transition domain separates these two extreme cases [the roman numbers indicate the domains defined by Matsuda et al.,[146] who derived criteria defining these three regions].

when $c_{\mathrm{M}init} \gg c_{\mathrm{L}init}$:

$$\Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{1/2} = \Delta_{\mathrm{o}}^{\mathrm{w}}\phi_{\mathrm{M}^{z+}}^{o'} - \frac{RT}{zF}\ln \sum_{j=0}^{m} j\beta_j^{\mathrm{o}} c_{\mathrm{M}init} \left(\frac{c_{\mathrm{L}init}}{2}\right)^{(j-1)} \qquad (20)$$

where β_j^{o} are the reduced association constants, defined as

$$\beta_j^{\mathrm{o}} = \frac{c^{\mathrm{o}}_{\mathrm{ML}_j^{z+}}}{c^{\mathrm{o}}_{\mathrm{ML}_j^{z+}} (c_{\mathrm{L}}^{\mathrm{o}})^j} = \prod_{k=1}^{j} K_{\mathrm{a}k}^{\mathrm{o}} \qquad (21)$$

The transfer of alkali metal cations such as Li^+, Na^+, K^+, Rb^+, Cs^+ and Tl^+ is facilitated by DB18C6[71,131,153,154] and some of them by monensin, valinomycin, nigericin,[155] nonactin[156] and various crown ethers;[149,157] the transfer of alkaline earth metal ions such as Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+} is facilitated by various water-soluble and -insoluble crown ethers and by poly(oxyethylene)octylphenyl ethers[158] and calix[4]-arene.[159] The transfer of the alkali and alkaline earth metal ions is also facilitated by other synthetic neutral ionophores.[160,161] Various studies were carried out with 1,10-phenanthroline, which forms complexes with heavy metal ions such as Cd^{2+}, Zn^{2+}, Co^{2+}, Ni^{2+},[67] Mn^{2+}[72] and Cu^{2+},[76] and also with protons[68,162] and alkali and alkaline earth metal cations in various stoichiometries.[163] 8-Quinolinol (Ni^{2+}), hydroxy oximes (Cu^{2+}) and dithizones (Ni^{2+}, Zn^{2+} and Cd^{2+}) have been used for solvent extraction of these metals.[164,165] Thiocrown ligands such as 1,4,7,10-tetrathiacyclododecane facilitate the transfer of Cu^{2+},[166–168] Cu^+,[168,169] Hg^{2+},[168] Ag^+,[168,170] Ni^{2+}[171] and Pb^{2+} and Cd^{2+}[172] and extraction of lanthanide ions has been achieved with 1,4,10,13-tetrathia-7,6-diazacyclooctadecane.[173]

Likewise, ethylenediamines of different hydrophobicities have been shown to form complexes of various stoichiometries with Ni(II).[174] Complexes of 1:1 to 1:3 stoichiometries have been clearly identified for the transfer of Ni(II), Zn(II), Fe(II) and Fe(III) assisted by the bidentate ionophores *o*,*o*′-bipyridine and *o*-phenanthroline[175] and for that of Co^{2+}[176] and Cd^{2+}[177] assisted by 2,2′-bipyridine. Similar investigations were made with the terdentate ligands 2,2′,2″-terpyridine[178] and 2,2′,6′,2″-terpyridine[179] that form stable 1:1 and 1:2 complexes with transition metals[178] and with the tetradentate phosphorus–nitrogen ligand *N*,*N*′-bis[2-(diphenylphosphino)phenyl]propane-1,3-diamine for the complexation of Co^{2+}.[180] Dihydroxynaphthalene has also been used as a complexing agent to detect Mo[181] and cupferron to detect Al,[182] Cr,[183] U,[184] V[185] and Eu.[186]

The facilitation effect of drug molecules such as valinomycin,[7,140,187] monensin,[188] nigericin,[155] piroxicam,[189] creatinine,[91] terramycin,[190] rifamycin,[191] lidocaine and dicaine[192] has also been studied, as well as the transfer of U, Np and Pu ions facilitated by phosphine oxide.[193] Finally, phosphine ligands have been used successfully to separate simultaneously Pb^{2+}, Zn^{2+}, Co^{2+}, Fe^{2+}, Cd^{2+}, Mn^{2+} and Ca^{2+}[92]. Polarographic and molecular dynamics studies have also been introduced as methods for the calculation of complex association constants in polar solvents[194–198] and for the determination of the structural conformation of various crown ether complexes, respectively.[199–201] Recently, complex formation at ITIES has also been studied by SHG suggesting that a reorientation of the ionophore occurs before the complexation and the transfer of the metal ion take place.[202,203]

This listing shows the variety of the systems studied, and a complete review and discussion of the advances in assisted ion-transfer reactions at the ITIES up to 1993 have been published.[23,24] However, most of the above reactions follow a TIC or TID[62] mechanism, and only a few studies show evidence of an ACT mechanism,[158,204] because most ligands used in electrochemistry are poorly soluble in water, so that complexes may only be formed in the organic phase. Furthermore, the TOC mechanism has not been differentiated from the TIC mechanism, because this necessitates kinetic data that still remain unmeasurable, the diffusion being the factor limiting the transfer and not the energy required to allow the ions to cross the interface.

4.3 Lipophilicity of Ions

Because of their intrinsic nature, interfaces between two immiscible liquids may serve as simple artificial models of biological membranes. Therefore, studies on drug transfer characteristics and mechanisms in such systems are of great importance to understand better the behavior of drugs in their pharmacokinetic phase,[205,206] their distribution in vivo[207] and hence the delivery problems that limit their efficiency.[208–211] The transport of exogeneous chemicals (and hence of the majority of common drugs) is a passive process,[212] for which it is commonly assumed that ionizable compounds can only cross biological membranes in their neutral form.[213] However, recent studies suggest a significant passive transfer of ions.[214–217] As many drugs are organic compounds that are thus partly or largely ionized at physiological pH, membrane transport can be deeply affected by the lipophilicity of charged species.[218] The lipophilicity of a species is generally evaluated by its partition coefficient, defined as the logarithm of the ratio of the activity of a species in the organic phase to that in the aqueous phase and denoted $\log P$. This parameter is widely used in medicinal chemistry to relate the structure and the physicochemical properties of a drug to its biological activity,[219] which is the objective of all quantitative structure–activity relationship (QSAR) studies.[220]

Dealing only with ions for which $\log P$ is directly given by the formal transfer potential,[221] electrochemistry at the ITIES appears to be a method of choice for assessing the lipophilicity of ionizable drugs. The literature is still very scarce on this topic, and ion-transfer voltammetry has only been applied to determine the transfer potentials of a few compounds of biological interest

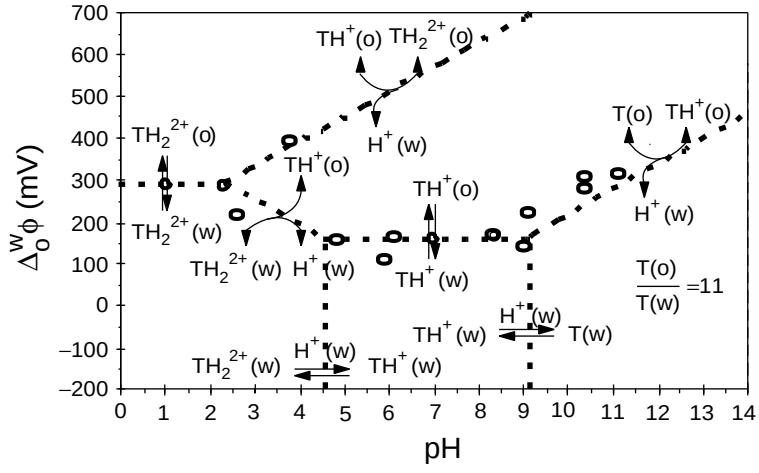

Figure 13 Ionic partition diagram of trimetazidine at 25 °C in water–1,2-DCE. The figure shows the formal transfer potentials obtained by cyclic voltammetry as a function of aqueous pH (bold circles), the equiconcentration lines between two adjacent species (bold dashed lines) and the corresponding transfer mechanisms (chemical equilibria). T, TH^+ and TH_2^{2+} stand for the neutral, the singly protonated and the doubly protonated forms of trimetazidine, respectively. (Adapted from Reymond et al.[247] by permission of Plenum Publishing Corporation.)

[1,10-phenanthroline,[68,222] acetylcholine,[125,223–225] various amines,[77,226,227] phosphorylation uncouplers,[228,229] pyrazolone derivatives,[230] picrate,[63,231–233] cinchonidine[75] and quinine[74,221,234,235]] and to investigate the transfer of several antibiotics[236–239] and of a series of hypnotic, anesthetic, cholinergic and adrenergic agents.[192,240–243] However, the recent introduction of ionic partition diagrams[234] (which are a transposition at liquid/liquid interfaces of Pourbaix's pH–potential diagrams[244] for metals in solution) has improved the understanding of the partition processes of ionizable compounds, since they allow reliable predictions and interpretations of their transfer mechanisms across ITIES, as exemplified by Figure 13 for the case of trimetazidine. Further insight into the influence of electronic structure on lipophilicity has revealed the importance of intramolecular charge delocalization to stabilize ions in the organic phase.[245–248] Although this effect has not yet been quantified, electrochemistry at ITIES is an easy methodology to assess the pH–lipophilicity profiles of ionizable drugs, and it should soon become increasingly popular.

It must finally be noted that octanol is not polarizable,[249] so that most electrochemical studies use 1,2-DCE or nitrobenzene as the organic phase. In order to interpret the lipophilicity of ionizable drugs in pharmacological terms, the thermodynamic parameters obtained with these solvents must be correlated to the octanol–water system commonly used in pharmacology. This correspondence has been established,[250] offering more relevance to the results presented above and opening ideal perspectives for extending electrochemical measurements to medicinal chemistry.

5 CONCLUSION

The increasing number of experiments conducted at liquid/liquid interfaces shows that electrochemical methodologies at such boundaries can be an efficient and versatile tool to probe ions. The great demand for innovative analytical techniques with accurate sensitivity and high selectivity offers a brilliant future for ITIES. Solvent extraction, separation, transfer-phase catalysis and biomembrane studies are other areas where ITIES are expected to develop rapidly.

Further, the study of charge-transfer processes gives an insight into the fundamental problems of physical chemistry in biphasic systems. Thanks to the improved theoretical understanding of interfacial structure, the transfer process of solvated ions from one phase to the other becomes clearer, but it is still very demanding because future applications of ITIES will depend on our theoretical knowledge of the motion of ions in biphasic systems.

Nevertheless, the above applications still suffer from a lack of experimental information compared with other techniques, and further investigations are required.

LIST OF SYMBOLS

a	Activity
A	Absorbance
A	Interfacial Area
c	Concentration
D	Diffusion Coefficient
F	Faraday Constant

h	Penetration Depth of a Phase into a Microhole
I	Current
I_p^{FWD}	Maximum Forward Peak Current
K_a	Association Constant
r	Radius of the Interface
R	Gas Constant
t	Time
T	Temperature
x	Position
z	Charge

Greek Letters

β_j	Reduced Association Constant of the ML_j Complex
$\Delta_o^w\phi$	Galvani Potential Difference Between the w and o Phases
$\Delta_o^w\phi_i^o$	Standard Transfer Potential of i
$\Delta_o^w\phi_i^{o'}$	Formal Transfer Potential of i
$\Delta_o^w\phi_{i,1/2}$	Half-wave Potential
$\Delta G_{t,i}^{o,w\rightarrow o}$	Standard Transfer Gibbs Energy of a Species i from Phase w to Phase o
η	Solvent Viscosity
μ^o	Standard Chemical Potential or Gibbs Energy of Solvation
v	Rate of a Potential Sweep

Superscripts

o	Organic Phase
w	Aqueous Phase
o	Standard Value
o'	Formal Value
FWD	Forward

Subscripts

i	Ion i
L	Ligand or Ionophore
init	Initial
M	Metal Ion
ML	Metal–Ionophore Complex
+	Positive Ion
−	Negative Ion
SS	Steady State
t	Transfer

ABBREVIATIONS AND ACRONYMS

ACT	Aqueous Complexation followed by Transfer
BTTPATPBCl	Bis(triphenylphosphoranylidene) Ammonium Tetrakis(4-chlorophenyl Borate)
DB18C6	Dibenzo-18-crown-6
DBS	1,3 : 2,4-Dibenzylidene Sorbitol
ISE	Ion-selective Electrode
ITIES	Interface Between Two Electrolyte Solutions
NaTPB	Sodium Tetrabutylborate
NPOE	*o*-Nitrophenyl Octyl Ether
PET	Poly(ethylene terephthalate)
PVC	Poly(vinyl chloride)
QSAR	Quantitative Structure–Activity Relationship
SFG	Sum-frequency Generation
SHG	Second Harmonic Generation
TBACL	Tetrabutyl Ammonium Chloride
TBATPB	Tetrabutylammonium Tetraphenylborate
TIC	Transfer by Interfacial Complexation
TID	Transfer by Interfacial Decomplexation
TIR	Total Internal Reflection
TOC	Transfer followed by Organic-phase Complexation
UV/VIS	Ultraviolet/Visible
1,2-DCE	1,2-Dichloroethane

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Quantitative Structure–Activity Relationships and Computational Methods in Drug Discovery

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Chemiluminescence, Electrogenerated • Ion-selective Electrodes: Fundamentals • Pulse Voltammetry

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction

REFERENCES

1. A.G. Volkov, D.W. Deamer (eds.), *Liquid–Liquid Interfaces: Theory and Methods*, CRC Press, Boca Raton, FL, 1996.

2. C. Gavach, 'Kinetics of Electroadsorption and Polarization at the Interface Between Certain Nonmiscible Ionic Solutions', *Experientia, Suppl.*, **18**, 321–331 (1971).
3. C. Gavach, A. Savajols, 'Biionic Potentials of Strongly Dissociated Liquid Membranes', *Electrochim. Acta*, **19**, 575–581 (1974).
4. C. Gavach, P. Seta, F. Henry, 'A Study of the Ionic Transfer Across an Aqueous Solution/Liquid Membrane Interface by Chronoamperometric and Impedance Measurements', *Bioelectrochem. Bioenerg.*, **1**, 329–342 (1974).
5. C. Gavach, F. Henry, 'Chronopotentiometric Investigation of the Diffusion Overvoltage at the Interface Between two Nonmiscible Solutions. Aqueous Solution-Tetrabutylammonium Ion Specific Liquid Membrane', *J. Electroanal. Chem.*, **54**, 361–370 (1974).
6. Z. Samec, V. Marecek, J. Koryta, M.W. Khalil, 'Investigation of Ion Transfer Across the Interface Between Two Immiscible Electrolyte Solutions by Cyclic Voltammetry', *J. Electroanal. Chem.*, **83**, 393–397 (1977).
7. J. Koryta, 'Electrochemical Polarization Phenomena at the Interface of Two Immiscible Electrolyte Solutions', *Electrochim. Acta*, **24**, 293–300 (1979).
8. D. Homolka, L.Q. Hung, A. Hofmanova, M.W. Khalil, J. Koryta, V. Marecek, Z. Samec, S.K. Sen, P. Vanysek, 'Faradaic Ion Transfer Across the Interface of Two Immiscible Electrolyte Solutions: Chronopotentiometry and Cyclic Voltammetry', *Anal. Chem.*, **52**, 1606–1610 (1980).
9. J. Koryta, M. Brezina, A. Hofmanova, D. Homolka, H. Le Quoc, W. Khalil, V. Marecek, Z. Samec, S.K. Sen, P. Vanysek, J. Weber, 'A New Model of Membrane Transport: Electrolysis at the Interface of Two Immiscible Electrolyte Solutions', *Bioelectrochem. Bioenerg.*, **7**, 61–68 (1980).
10. J. Koryta, 'Ion Transfer Across Water/Organic Phase Boundaries and Analytical Applications', *Ion-Sel. Electrode Rev.*, **5**, 131–164 (1983).
11. Z. Samec, J. Weber, V. Marecek, 'Charge Transfer Between Two Immiscible Electrolyte Solutions. Part II. The Investigation of Cs+ Ion Transfer Across the Nitrobenzene/Water Interface by Cyclic Voltammetry with IR Drop Compensation', *J. Electroanal. Chem.*, **100**, 841–852 (1979).
12. T. Kakutani, T. Osakai, M. Senda, 'A Potential-step Chronoamperometric Study of Ion Transfer at the Water/Nitrobenzene Interface', *Bull. Chem. Soc. Jpn.*, **56**, 991–996 (1983).
13. Z. Samec, V. Marecek, J. Weber, D. Homolka, 'Charge Transfer Between Two Immiscible Electrolyte Solutions. Advances in Method of Electrolysis with the Electrolyte Dropping Electrode (EDE)', *J. Electroanal. Chem.*, **99**, 385–389 (1979).
14. V. Marecek, Z. Samec, 'Determination of Calcium, Barium and Strontium Ions by Differential Pulse Stripping Voltammetry at a Hanging Electrolyte Drop Electrode', *Anal. Chim. Acta*, **151**, 265–269 (1983).
15. B. Hundhammer, T. Solomon, H. Alemu, 'Investigation of the Ion Transfer Across the Water/Nitrobenzene Interface by AC Cyclic Voltammetry', *J. Electroanal. Chem.*, **149**, 179–183 (1983).
16. P. Vanysek, *Electrochemistry on Liquid–Liquid Interfaces, Lecture Notes in Chemistry*, Springer-Verlag, Berlin, 1985.
17. H.H. Girault, 'Electrochemistry at the Interface Between Two Immiscible Electrolyte Solutions', *Electrochim. Acta*, **32**, 383–385 (1987).
18. J. Koryta, 'Electrochemical Polarization Phenomena at the Interface of Two Immiscible Electrolyte Solutions-III. Progress since 1983', *Electrochim. Acta*, **33**, 189–197 (1988).
19. V. Marecek, Z. Samec, J. Koryta, 'Charge Transfer Across the Interface of Two Immiscible Electrolyte Solutions', *Adv. Colloid Interface Sci.*, **29**, 1–78 (1988).
20. Z. Samec, 'Electrical Double Layer at the Interface Between Two Immiscible Electrolyte Solutions', *Chem. Rev.*, **88**, 617–632 (1988).
21. V.S. Markin, A.G. Volkov, 'The Gibbs Free Energy of Ion Transfer Between Two Immiscible Liquids', *Electrochim. Acta*, **34**, 93–107 (1989).
22. H.H. Girault, D.J. Schiffrin, 'Electrochemistry of Liquid–Liquid Interfaces', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–141, Vol. 15, 1989.
23. M. Senda, T. Kakiuchi, T. Osakai, 'Electrochemistry at the Interface Between Two Immiscible Electrolyte Solutions', *Electrochim. Acta*, **36**, 253–262 (1991).
24. H.H. Girault, 'Charge Transfer Across Liquid–Liquid Interfaces', in *Modern Aspects of Electrochemistry*, eds. J.O.M. Bockris, B. Conway, R. White, Plenum Press, New York, 1–62, Vol. 25, 1993.
25. P. Vanysek, 'Charge Transfer Processes on Liquid–Liquid Interfaces: The First Century', *Electrochim. Acta*, **40**, 2841–2847 (1995).
26. M. Senda, 'Voltammetry at Liquid–Liquid Interface', *Anal. Sci. Technol.*, **8**, 95A–100A (1995).
27. H.H. Girault, D.J. Schiffrin, 'Theory of the Kinetics of Ion Transfer Across Liquid–Liquid Interfaces', *J. Electroanal. Chem*, **195**, 213–227 (1985).
28. Y.Y. Gurevich, Y.I. Kharkats, 'Ion Transfer through a Phase Boundary: A Stochastic Approach', *J. Electroanal. Chem.*, **200**, 3–16 (1986).
29. Z. Samec, Y.I. Kharkats, Y.Y. Gurevich, 'Stochastic Approach to the Ion Transfer Kinetics across the Interface Between Two Immiscible Electrolyte Solutions. Comparison with Experimental Data', *J. Electroanal. Chem*, **204**, 257–266 (1986).
30. G.M. Torrie, J.P. Valleau, 'Double Layer Structure at the Interface Between Two Immiscible Electrolyte Solutions', *J. Electroanal. Chem.*, **206**, 69–79 (1986).

31. F. Milner, M.J. Weaver, 'A Simulation Analysis of Errors in the Measurement of Standard Electrochemical Rate Constants from Phase-selective Impedance Data', *J. Electroanal. Chem.*, **222**, 21–33 (1987).
32. Y.I. Kharkats, J. Ulstrup, 'The Electrostatic Gibbs Energy of Finite-size Ions Near a Planar Boundary Between Two Dielectric Media', *J. Electroanal. Chem.*, **308**, 17–26 (1991).
33. T. Kakiuchi, 'Current–Potential Characteristic of Ion Transfer Across the Interface Between Two Immiscible Electrolyte Solutions Based on the Nernst–Planck Equation', *J. Electroanal. Chem.*, **322**, 55–61 (1992).
34. A.V. Indenbom, 'Interpretation of Ion Transfer Reactions Across the Interface Between Two Immiscible Electrolyte Solutions by the Model of ''Elastic'' Inner Layer', *Electrochim. Acta*, **40**, 2985–2991 (1995).
35. W. Schmickler, 'A Unified Model for Electrochemical Electron and Ion Transfer Reactions', *Chem. Phys. Lett.*, **237**, 152–160 (1995).
36. H.H. Girault, D.J. Schiffrin, 'Thermodynamics of a Polarized Interface Between Two Immiscible Electrolyte Solutions', *J. Electroanal. Chem.*, **170**, 127–141 (1984).
37. I. Benjamin, 'Dynamics of Ion Transfer Across a Liquid–Liquid Interface: A Comparison Between Molecular Dynamics and Diffusion Model', *J. Chem. Phys.*, **96**, 577–585 (1992).
38. I. Benjamin, 'Theoretical Study of Water/1,2-Dichloroethane Interface: Structure, Dynamics, and Conformational Equilibria at the Liquid–Liquid Interface', *J. Chem. Phys.*, **97**, 1432–1445 (1992).
39. A. Vincze, G. Horvai, F.A.M. Leermakers, J.M.H.M. Scheutjens, 'Mathematical-modeling of the Interface of Two Immiscible Electrolyte-Solutions', *Sens. Actuators B*, **18**, 42–46 (1994).
40. T. Wandlowski, K. Holub, V. Marecek, Z. Samec, 'The Double-layer at the Interface Between 2 Immiscible Electrolyte-solutions. IV. Solvent Effect', *Electrochim. Acta*, **40**, 2887–2895 (1995).
41. S. Toxvaerd, J. Stecki, 'Density Profiles at a Planar Liquid–Liquid Interface', *J. Chem. Phys.*, **102**, 7163–7168 (1995).
42. I. Benjamin, 'Theory and Computer Simulations of Solvation and Chemical Reactions at Liquid Interfaces', *Acc. Chem. Res.*, **28**, 233–239 (1995).
43. D.J. Henderson, W. Schmickler, 'Simple Model for Liquid/Liquid Interfaces', *J. Chem. Soc., Faraday Trans.*, **92**, 3839–3842 (1996).
44. Y. Shao, J.A. Campbell, H.H. Girault, 'Kinetics of Transfer of Acetylcholine Across the Water/Nitrobenzene-Tetrachloromethane Interface. The Gibbs Energy of Transfer Dependence of the Standard Rate Constant', *J. Electroanal. Chem.*, **300**, 415–429 (1991).
45. T. Kakiuchi, J. Noguchi, M. Senda, 'Double-layer Effect on the Transfer of Some Monovalent Ions Across the Polarized Oil–Water Interface', *J. Electroanal. Chem.*, **336**, 137–152 (1992).
46. A.G. Volkov, D.W. Deamer, D.L. Tanelian, V.S. Markin, 'Electrical Double Layers at the Oil/Water Interface', *Prog. Surf. Sci.*, **53**, 1–134 (1996).
47. I. Benjamin, 'Mechanism and Dynamics of Ion Transfer Across a Liquid–Liquid Interface', *Science*, **261**, 1558–1560 (1993).
48. K.J. Schweighofer, I. Benjamin, 'Electric Field Effects on the Structure and Dynamics at a Liquid–Liquid Interface', *J. Electroanal. Chem.*, **391**, 1–10 (1995).
49. K.J. Schweighofer, I. Benjamin, 'Transfer of Small Ions Across the Water/1,2-Dichloroethane Interface', *J. Phys. Chem.*, **99**, 9974–9985 (1995).
50. H.H. Girault, 'The Potential Dependence of the Rate of Ion Transfer Reactions Across a Liquid/Liquid Interface', *J. Electroanal. Chem.*, **257**, 47–55 (1988).
51. W. Schmickler, 'A Model for Ion Transfer through Liquid/Liquid Interfaces', *J. Electroanal. Chem.*, **426**, 5–9 (1997).
52. J. Koryta, 'Theory and Applications of Ion-selective Electrodes', *Anal. Chim. Acta*, **183**, 1–46 (1986).
53. E. Wang, Z. Sun, 'Development of Electroanalytical Chemistry at the Liquid–Liquid Interface', *Trends Anal. Chem.*, **7**, 99–106 (1988).
54. M. Senda, H. Katano, M. Yamada, 'Amperometric Ion-selective Electrode. Voltammetric Theory and Analytical Applications at High Concentration and Trace Levels', *J. Electronal. Chem.*, **468**, 34–41 (1999).
55. M. Senda, Y. Yamamoto, 'Amperometric Ion-selective Electrode Sensors', in *Liquid–Liquid Interfaces. Theory and Methods*, eds. A.G. Volkov, D.W. Deamer, CRC Press, Boca Raton, FL, 277–293, 1996.
56. T. Kakiuchi, 'Partition Equilibrium of Ionic Components in Two Immiscible Electrolyte Solutions', in *Liquid–Liquid Interfaces, Theory and Methods*, eds. A.G. Volkov, D.W. Deamer, CRC Press, Boca Raton, FL, 1–18, 1996.
57. Y. Marcus, 'Single Ion Gibbs Free Energies of Transfer from Water to Organic and Mixed Solvents', *Anal. Chem.*, **5**, 53–137 (1980).
58. Y. Marcus, 'Thermodynamic Functions of Transfer of Single Ions from Water to Nonaqueous and Mixed Solvents: Part 1 – Gibbs Free Energies of Transfer to Nonaqueous Solvents', *Pure Appl. Chem.*, **55**, 977–1021 (1983).
59. Y. Marcus, 'Ion Solvation', in *Liquid–Liquid Interfaces. Theory and Methods*, eds. A.G. Volkov, D.W. Deamer, CRC Press, Boca Raton, FL, 39–61, 1996.
60. I. Benjamin, 'Chemical Reactions and Solvation at Liquid Interfaces: A Microscopic Perspective', *Chem. Rev.*, **96**, 1449–1475 (1996).
61. I. Benjamin, 'Molecular Structure and Dynamics at Liquid–Liquid Interfaces', *Annu. Rev. Phys. Chem.*, **48**, 407–451 (1997).

62. Y. Shao, M.D. Osborne, H.H. Girault, 'Assisted Ion Transfer at Micro-ITIES Supported at the Tip of Micropipettes', *J. Electroanal. Chem.*, **318**, 101–109 (1991).
63. D. Homolka, V. Marecek, 'Charge Transfer Between Two Immiscible Electrolyte Solutions. Part VI. Polarographic and Voltammetric Study of Picrate Ion Transfer Across the Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **112**, 91–96 (1980).
64. A.J. Bard, L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, John Wiley & Sons, New York, 1980.
65. J. Koryta, P. Vanysek, M. Brezina, 'Electrolysis with an Electrolyte Dropping Electrode', *J. Electroanal. Chem.*, **67**, 263–266 (1976).
66. J. Koryta, P. Vanysek, M. Brezina, 'Electrolysis with Electrolyte Dropping Electrode. II. Basic Properties of the System', *J. Electroanal. Chem.*, **75**, 211–228 (1977).
67. Z. Yoshida, H. Freiser, 'Ascending Water Electrode Studies of Metal Extractants. Role of Kinetic Factors in the Faradaic Ion Transfer of Metal–Phenanthroline Complex Ions Across an Aqueous–Organic Solvent Interface', *Inorg. Chem.*, **23**, 3931–3935 (1984).
68. Z. Yoshida, H. Freiser, 'Ascending Water Electrode Studies of Metal Extractants. Faradaic Ion Transfer of Protonated 1,10-Phenanthroline and Its Derivatives Across an Aqueous 1,2-Dichloroethane Interface', *J. Electroanal. Chem.*, **162**, 307–319 (1984).
69. S. Kihara, M. Suzuki, K. Maeda, K. Ogura, S. Umetani, M. Matsui, Z. Yoshida, 'Fundamental Factors in the Polarographic Measurements of Ion Transfer at the Aqueous/Organic Solution Interface', *Anal. Chem.*, **58**, 2954–2961 (1986).
70. L. Sinru, Z. Zaofan, H. Freiser, 'Potassium Ion Transport Processes Across the Interface of an Immiscible Liquid Pair in the Presence of Crown Ethers', *J. Electroanal. Chem.*, **210**, 137–146 (1986).
71. Z. Samec, P. Papoff, 'Electrolyte Dropping Electrode Polarographic Studies. Solvent Effect on Stability of Crown Ether Complexes of Alkali-metal Cations', *Anal. Chem.*, **62**, 1010–1015 (1990).
72. H. Doe, H. Freiser, 'Ion-transfer Current-scan Polarographic Studies of Metal Extractants with Ascending Water Electrode: Manganese(II)-1,10-Phenanthroline and Related Ligand Systems', *Anal. Sci.*, **7**, 303–311 (1991).
73. H. Doe, H. Freiser, 'Ion-transfer Current-scan Polarographic Studies of Metal Extractants with Ascending Water Electrode: Transfer of Divalent Metal Ion Complexes of Phenanthrolines Across a Liquid–Liquid Interface', *Anal. Sci.*, **7**, 313 (1991).
74. J. Xiao, L. Nie, S.-Z. Yao, 'Transfer Mechanism of Quinine Drug Across the Oil/Water Interface', *Sci. China (Ser. B)*, **34**, 42–53 (1991).
75. S. Lin, D. Shen, L. Nie, S. Yao, 'The Transfer Mechanism and Analytical Properties of the Variable-valency Drug Cinchonidine Across the Liquid/Liquid Interface', *Electrochim. Acta*, **38**, 207–213 (1993).
76. H. Doe, M. Hoshiyama, L. Jian, 'Cooperative Ion-transfer of Copper(II)-1,10-Phenanthroline-trifluoroacetylacetone Complex Across a Water/1,2-Dichloroethane Interface', *Electrochim. Acta*, **40**, 2947–2951 (1995).
77. Y. Kubota, H. Katano, K. Maeda, M. Senda, 'Voltammetric Study of the Partition of Amines Between Water and an Organic Solvent', *Electrochim. Acta*, **44**, 109–116 (1998).
78. G. Taylor, H.H. Girault, 'Ion Transfer Reaction across a Liquid–Liquid Interface Supported on a Micropipette Tip', *J. Electroanal. Chem.*, **208**, 179–183 (1986).
79. J.A. Campbell, H.H. Girault, 'Steady-state Current for Ion Transfer Reactions at a Micro Liquid/Liquid Interface', *J. Electroanal. Chem.*, **266**, 465–469 (1989).
80. J.A. Campbell, A.A. Stewart, H.H. Girault, 'Determination of the Kinetics of Facilitated Ion Transfer Reactions Across the Micro Interface Between Two Immiscible Electrolyte Solutions', *J. Chem. Soc., Faraday Trans. 1*, **85**, 843–853 (1989).
81. A.A. Stewart, G. Taylor, H.H. Girault, J. McAleer, 'Voltammetry at MicroITIES Supported at the Tip of a Micropipette. Part I. Linear Sweep Voltammetry', *J. Electroanal. Chem.*, **296**, 491–515 (1990).
82. A.A. Stewart, Y. Shao, C.M. Pereira, H.H. Girault, 'Micropipette as a Tool for the Determination of the Ionic Species Limiting the Potential Window at Liquid/Liquid Interfaces', *J. Electroanal. Chem.*, **305**, 135–139 (1991).
83. T. Ohkouchi, T. Kakutani, T. Osakai, M. Senda, 'Voltammetry with Ion Selective Microelectrode Based on Polarizable Oil/Water Interface', *Anal. Sci.*, **7**, 371–376 (1991).
84. P.D. Beattie, A. Delay, H.H. Girault, 'Investigation of the Kinetics of Ion and Assisted Ion Transfer by the Technique of ac Impedance of the Micro-ITIES', *Electrochim. Acta*, **40**, 2961–2969 (1995).
85. V. Cunnane, D.J. Schiffrin, D.E. Williams, 'Microcavity Electrode: A New Type of Liquid–Liquid Microelectrode', *Electrochim. Acta*, **40**, 2943–2946 (1995).
86. A.-K. Kontturi, K. Kontturi, L. Murtomäki, B. Quinn, V.J. Cunnane, 'Study of Ion Transfer across Phospholipid Monolayers Adsorbed at Micropipette ITIES', *J. Electroanal. Chem.*, **424**, 69–74 (1997).
87. Y. Shao, M.V. Mirkin, 'Voltammetry at Micropipet Electrodes', *Anal. Chem.*, **70**, 3155–3161 (1998).
88. M.C. Osborne, Y. Shao, C.M. Pereira, H.H. Girault, 'Micro-hole Interface for the Amperometric Determination of Ionic Species in Aqueous Solutions', *J. Electroanal. Chem.*, **364**, 155–161 (1994).
89. M.D. Osborne, H.H. Girault, 'The Liquid–Liquid Micro-Interface for the Amperometric Detection of Urea', *Electroanalysis*, **7**, 714–721 (1995).

90. C.M. Pereira, F. Silva, 'Square Wave Voltammetry with Arrays of Liquid/Liquid Microinterfaces', *Electroanalysis*, **6**, 1034–1039 (1994).
91. M.D. Osborne, H.H. Girault, 'The Micro Water/1,2-Dichloroethane Interface as a Transducer for Creatinine Assay', *Mikrochim. Acta*, **117**, 175–185 (1995).
92. S. Wilke, H. Wang, M. Muraczewska, H. Müller, 'Amperometric Detection of Heavy Metal Ions in Ion Pair Chromatography at an Array of Water/Nitrobenzene Micro Interfaces', *Fresenius' J. Anal. Chem.*, **356**, 233–236 (1996).
93. S. Wilke, M.D. Osborne, H.H. Girault, 'Electrochemical Characterization of Liquid/Liquid Microinterface Arrays', *J. Electroanal. Chem.*, **436**, 53–64 (1997).
94. C. Beriet, H.H. Girault, 'Electrochemical Studies of Ion Transfer at Micro-machined Supported Liquid Membranes', *J. Electroanal. Chem.*, **444**, 219–229 (1998).
95. W. Mickler, A. Mönner, E. Uhlemann, S. Wilke, H. Müller, 'Transfer of β-Diketone and 4-Acylpyrazolone Anions Across the Electrified Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **469**, 91–96 (1999).
96. B.J. Seddon, Y. Shao, J. Fost, H.H. Girault,'The Application of Excimer Laser Micromachining for the Fabrication of Disc Microelectrode', *Elecrochim. Acta*, **39**, 783–791 (1994).
97. M.A. Roberts, J.S. Rossier, P. Bercier, H. Girault, 'UV Laser Machined Polymer Substrates for the Development of Microdiagnostic Systems', *Anal. Chem.*, **69**, 2035–2042 (1997).
98. S. Wilke, T. Zerihun, 'Diffusion Effects at Microhole Supported Liquid/Liquid Interfaces', *Electrochim. Acta*, **44**, 15–22 (1998).
99. C. Beriet, H.H. Girault, 'Simulation of the Diffusion Process to a Regular Array of Micro-disc Electrodes', submitted.
100. J. Josserand, J. Morandini, H.J. Lee, R. Ferrigno, H.H. Girault, 'Finite Element Simulation of Ion Transfer Reactions at a Single Micro-Liquid/Liquid Interface Supported on a Thin Polymer Film, *J. Electroanal. Chem.*, **468**, 42–52 (1999).
101. H.J. Lee, C. Beriet, H.H. Girault, 'Amperometric Detection of Alkali Metal Ions on Micro-fabricated Composite Polymer Membranes', *J. Electroanal. Chem.*, **453**, 211–219 (1998).
102. H.J. Lee, H.H. Girault, 'Amperometric Ion Detector for Ion Chromatography', *Anal. Chem.*, **70**, 4280–4285 (1998).
103. P. Guyot-Sionnest, J.H. Hunt, Y.R. Shen, 'Sum-frequency Vibrational Spectroscopy of a Langmuir Film: Study of Molecular Orientation of a Two-dimensional System', *Phys. Rev. Lett.*, **59**, 1597–1600 (1987).
104. Y.R. Shen, 'Optical Second Harmonic Generation at Interfaces', *Annu. Rev. Phys. Chem.*, **40**, 327–350 (1989).
105. K. Wolfrum, H. Graener, A. Laubereau, 'Sum-frequency Vibrational Spectroscopy at the Liquid–Air Interface of Methanol. Water Solutions', *Chem. Phys. Lett.*, **213**, 41–46 (1993).
106. C.D. Stanners, Q. Du, R.P. Chin, P. Cremer, G.A. Somorjai, Y.R. Shen, 'Polar Ordering at the Liquid/Vapor Interface of *n*-Alcohol (C1–C8)', *Chem. Phys. Lett.*, **232**, 407–413 (1995).
107. K.B. Eisenthal, 'Liquid Interfaces Probed by Second-harmonic Generation and Sum-frequency Spectroscopy', *Chem. Rev.*, **96**, 1343–1360 (1996).
108. P.-F. Brevet, *Surface Second Harmonic Generation*, Presses Polytechniques et Universitaires Romandes, Lausanne, 1996.
109. P.F. Brevet, H.H. Girault, 'Second Harmonic Generation at Liquid/Liquid Interfaces', in *Liquid–Liquid Interfaces: Theory and Methods*, eds. A.G. Volkov, D.W. Deamer, CRC Press, Boca Raton, FL, 103–137, 1996.
110. Z. Ding, R.G. Wellington, P.F. Brevet, H.H. Girault, 'Spectroelectrochemical Studies of $Ru(bpy)_3^{2+}$' at the Water/1,2-Dichloethane Interface', *J. Phys. Chem.*, **100**, 10658–10663 (1996).
111. T. Kakiuchi, Y. Takasu, M. Senda, 'Voltage-scan Fluorometry of Rose Bengal Ion at the 1,2-Dichloroethane–Water Interface', *Anal. Chem.*, **64**, 3096–3100 (1992).
112. T. Kakiuchi, Y. Takasu, 'Differential Cyclic Voltfluorometry and Chronofluorometry of the Transfer of Fluorescent Ions Across the 1,2-Dichloroethane–Water Interface', *Anal. Chem.*, **66**, 1853–1859 (1994).
113. T. Kakiuchi, Y. Takasu, 'Ion Selectivity of Voltage-scan Fluorometry at the 1,2-Dichloroethane/Water Interface', *J. Electroanal. Chem.*, **365**, 293–297 (1994).
114. T. Kakiuchi, Y. Takasu, 'Potential-step Chronofluorometric Response of Fluorescent-ion Transfer Across a Liquid/Liquid Interface', *J. Electroanal. Chem.*, **381**, 5–9 (1995).
115. Z. Ding, F. Reymond, P. Baumgartner, D. Fermin, P.-F. Brevet, P.-A. Carrupt, H.H. Girault, 'Mechanism and Dynamics of Methyl and Ethyl Orange Transfer Across the Water/1,2-Dichloroethane Interface', *Electrochim. Acta*, **44**, 3–13 (1998).
116. H.D. Duong, P.-F. Brevet, H.H. Girault, 'Heterogeneous Electron Transfer Reactions at Liquid–Liquid Interfaces Studied by Time Resolved Absorption Spectroscopy', *J. Photochem. Photobiol. A*, **117**, 27–33 (1998).
117. D.J. Fermin, Z. Ding, P.-F. Brevet, H.H. Girault, 'Potential Modulated Reflectance Spectroscopy of the Methyl Orange Transfer Across the Water/1,2-Dichloroethane Interface', *J. Electroanal. Chem.*, **447**, 125–133 (1998).
118. Z. Ding, D.J. Fermin, P.-F. Brevet, H.H. Girault, 'Spectroelectrochemical Approaches to Heterogeneous Electron Transfer Reactions at the Polarized Water/1,2-Dichloroethane Interfaces', *J. Electroanal. Chem.*, **458**, 139–148 (1998).

119. L. Tomaszewski, Z. Ding, D.J. Fermin, H.M. Caçote, C.M. Pereira, F. Silva, H.H. Girault, 'Spectroelectrochemical Study of Copper(II) Transfer Assisted by 6,7-Dimethyl-2,3-di(2-pyridyl)quinoxaline at the Water/1,2-Dichloroethane Interface', *J. Electroanal. Chem.*, **453**, 171–177 (1998).
120. R.A.W. Dryfe, Z. Ding, R.G. Wellington, P.-F. Brevet, A.M. Kuznetzov, H.H. Girault, 'Time-resolved Laser-induced Fluorescence Study of Photoinduced Electron Transfer at the Water/1,2-Dichloroethane Interface', *J. Phys. Chem. A*, **101**, 2519–2524 (1997).
121. T. Osakai, T. Kakutani, M. Senda, 'Ion-transfer Voltammetry with the Interfaces Between Polymer–Electrolyte Gel and Electrolyte Solutions', *Bunseki Kagaku*, **33**, E371–E377 (1984).
122. O. Dvorak, V. Marecek, Z. Samec, 'Ion Transfer Across Polymer Gel/Liquid Boundaries. Electrochemical Kinetics by Faradaic Impedance', *J. Electroanal. Chem.*, **284**, 205–215 (1990).
123. V. Marecek, M.P. Colombini, 'Charge Transfer Across a Polymer Gel/Liquid Interface: The Polyvinyl Chloride+ Nitrobenzene Gel/Water Interface', *J. Electroanal. Chem.*, **241**, 133–141 (1988).
124. T. Kakutani, T. Ohkouchi, T. Osakai, T. Kakiuchi, M. Senda, 'Ion-transfer Voltammetry and Potentiometry of Acetylcholine with the Interface Between Polymer–Nitrobenzene Gel and Water', *Anal. Sci.*, **1**, 219–225 (1985).
125. V. Marecek, H. Janchenova, M.P. Colombini, P. Papoff, 'Charge Transfer Across a Polymer Gel/Liquid Interface. A Voltammetric Detector for a Flow System', *J. Electroanal. Chem.*, **217**, 213–219 (1987).
126. E. Wang, H. Ji, W. Hou, 'The Use of Chemically Modified Electrodes for Liquid Chromatography and Flow-injection Analysis', *Electroanalysis*, **3**, 1–11 (1991).
127. V. Marecek, H. Janchenova, M. Brezina, M. Betti, 'Charge Transfer Across a Polymer Gel/Liquid Interface: Determination of Ionophores', *Anal. Chim. Acta*, **244**, 15–19 (1991).
128. S. Wilke, H. Franzke, H. Müller, 'Simultaneous Determination of Nitrate and Chloride by Means of Flow-injection Amperometry at the Membrane-stabilized Water/Nitrobenzene Interface', *Anal. Chim. Acta*, **268**, 285–292 (1992).
129. T. Kakutani, Y. Nishiwaki, T. Osakai, M. Senda, 'On the Mechanism of Transfer of Sodium Ion Across the Nitrobenzene/Water Interface Facilitated by Dibenzo-18-crown-6', *Bull. Chem. Soc. Jpn.*, **59**, 781–788 (1986).
130. Y. Yamamoto, T. Osakai, M. Senda, 'Potassium and Sodium Ion Sensor Based on Amperometric Ion-selective Electrode', *Bunseki Kagaku*, **39**, 655 (1990).
131. B. Hundhammer, T. Solomon, T. Zerihun, M. Abegaz, A. Bekele, K. Graichen, 'Investigation of Ion Transfer Across the Membrane-stabilized Interface of Two Immiscible Electrolyte Solutions. Part III. Facilitated Ion Transfer', *J. Electroanal. Chem.*, **371**, 1–11 (1994).
132. T. Osakai, T. Kakutani, M. Senda, 'A Novel Amperometric Ammonia Sensor', *Anal. Sci.*, **3**, 521–526 (1987).
133. E. Wang, H. Ji, 'Ion Transfer Across Water/Solidified Nitrobenzene Interface as Amperometric Flow Detector', *Electroanalysis*, **1**, 75–80 (1989).
134. Y. Yamamoto, M. Senda, 'Amperometric Ammonium Ion Sensor and its Application to Biosensors', *Sens. Actuators B*, **13–14**, 57–60 (1993).
135. T. Osakai, T. Kakutani, M. Senda, 'A Novel Amperometric Urea Sensor', *Anal. Sci.*, **4**, 529–530 (1988).
136. M. Senda, Y. Yamamoto, 'Urea Biosensor Based on Amperometric Ammonium Ion Electrode', *Electroanalysis*, **5**, 775–779 (1993).
137. H.J. Lee, P.D. Beattie, B.J. Seddon, M.D. Osborne, H.H. Girault, 'Amperometric Ion Sensors Based on Laser-patterned Composite Polymer Membranes', *J. Electroanal. Chem.*, **440**, 73–82 (1997).
138. H.J. Lee, C. Beriet, H.H. Girault, 'Stripping Voltammetry Determination of Choline Based on Microfabricated Composite Membrane', *Anal. Sci.*, **14**, 71–77 (1998).
139. F. Silva, M.J. Sousa, C.M. Pereira, 'Electrochemical Study of Aqueous–Organic Gel Micro-interfaces', *Electrochim. Acta*, **42**, 3095–3103 (1997).
140. P. Vanysek, W. Ruth, J. Koryta, 'Valinomycin Mediated Transfer of Potassium Across the Water/Nitrobenzene Interface. A Study by Voltammetry at the Interface Between Two Immiscible Electrolyte Solutions', *J. Electroanal. Chem.*, **148**, 117–121 (1983).
141. S. Sawada, H. Torii, T. Osakai, T. Kimoto, 'Pulse Amperometric Detection of Lithium in Artificial Serum Using a Flow Injection System with a Liquid/Liquid-Type Ion-selective Electrode', *Anal. Chem.*, **70**, 4286–4290 (1998).
142. J. Weiss, *Ion Chromatography*, VCH, Weinheim, 1995.
143. B.A. Moyer, Y. Sun, 'Principles of Solvent Extraction of Alkali Metal Ions: Understanding Factors Leading to Cesium Selectivity in Extraction by Solvation, in *Ion Exchange and Solvent Extraction*, eds. J.A. Marinsky, Y. Marcus, Marcel Dekker, New York, 295–391, Vol. 13, 1997.
144. D. Guo, A. Hofmanova, J. Koryta, 'Voltammetric Study of Ion Transfer Across the Oil/Water Interface Facilitated by an Ionophore', *Yingyong Huaxue*, **2**, 29–33 (1985).
145. F. Heitz, C. Gavach, G. Spach, Y. Trudelle, 'Analysis of the Ion Transfer Through the Channel of 9,11,13,15-Phenylalanylgramicidin A', *Biophys. Chem.*, **24**, 143–148 (1986).
146. H. Matsuda, Y. Yamada, K. Kanamori, Y. Kudo, Y. Takeda, 'On the Facilitation Effect of Neutral Macrocyclic Ligands on the Ion Transfer Across the Interface Between Aqueous and Organic Solutions. I. Theoretical Equation of Ion-transfer–Polagraphic

Current–Potential Curves and Its Experimental Verification', *Bull. Chem. Soc. Jpn.*, **64**, 1497–1508 (1991).
147. D. Homolka, K. Holub, V. Marecek, 'Facilitated Ion Transfer Across the Water/Nitrobenzene Interface. Theory for Single-scan Voltammetry Applied to a Reversible System', *J. Electroanal. Chem.*, **138**, 29–36 (1982).
148. T. Kakiuchi, M. Senda, 'Current–Potential Curves for Facilitated Ion Transfer Across Oil/Water Interfaces in The Presence of Successive Complex Formation', *J. Electroanal. Chem.*, **300**, 431–445 (1991).
149. Y. Kudo, Y. Takeda, H. Matsuda, 'On the Facilitating Effect of Neutral Macrocyclic Ligands on Ion Transfer Across the Interface Between Aqueous and Organic Solutions. 2. Alkali-metal Ion Complexes with Hydrophilic Crown-ethers', *J. Electroanal. Chem.*, **396**, 333–338 (1995).
150. P.D. Beattie, R.G. Wellington, H.H. Girault, 'Cyclic Voltammetry for Assisted Ion Transfer at ITIES', *J. Electroanal. Chem.*, **396**, 317–323 (1995).
151. F. Reymond, P.-A. Carrupt, H.H. Girault, 'Facilitated Ion Transfer Reactions Across Oil/Water Interfaces. Part I. Algebraic Development and Calculation of Cyclic Voltammetry Experiments for Successive Complex Formation', *J. Electroanal. Chem.*, **449**, 49–65 (1998).
152. F. Reymond, G. Lagger, P.-A. Carrupt, H.H. Girault, 'Facilitated Ion Transfer Reactions Across Oil/Water Interfaces. Part II. Use of the Convoluted Current for the Calculation of the Association Constants and for an Amperometric Determination of the Stoichiometry of ML_j^{z+} Complexes', *J. Electroanal. Chem.*, **451**, 59–76 (1998).
153. G. Rounaghi, Z. Eshaghi, E. Ghiamati, 'Thermodynamic Study of Complex Formation between 18-Crown-6 and Potassium Ion in Some Binary Non-aqueous Solvents Using a Conductometric Method', *Talanta*, **44**, 275–282 (1997).
154. S.A. Dassie, L.M. Yudi, A.M. Baruzzi, 'Voltammetric Analysis of Cs^+-DB18C6 Complex Stoichiometry at the Water/1,2-Dichloroethane Interface', *J. Electroanal. Chem*, **464**, 54–60 (1999).
155. A. Sabela, J. Koryta, O. Valent, 'Ion Carrier Properties of Nigericin Studied by Voltammetry at the Interface of Two Immiscible Electrolyte Solutions', *J. Electroanal. Chem.*, **204**, 267–272 (1986).
156. M.D. Osborne, H.H. Girault, 'Amperometric Detection of the Ammonium Ion by Facilitated Ion Transfer Across the Interface Between Two Immiscible Electrolyte Solutions', *Electroanalysis*, **7**, 425–434 (1995).
157. T. Okada, T. Usui, 'Role of Anions in the Complex Formation of Crown Ethers with Ammonium Ions Chemically Bonded on Silica Gel', *J. Chem. Soc., Faraday Trans.*, **92**, 4977–4981 (1996).
158. Z. Yoshida, S. Kihara, 'The Role of Non-ionic Polyoxyethylene Ether Surfactants on Ion Transfer Across Aqueous/Organic Solutions Interfaces Studied by Polarography with the Electrolyte Dropping Electrode', *J. Electroanal. Chem.*, **227**, 171–181 (1987).
159. V. Marecek, A. Lhotsky, K. Holub, I. Stibor, 'Surface Complex Formation at the Water/1,2-Dichloroethane Interface', *Electrochim. Acta*, **44**, 155–159 (1998).
160. Z. Samec, D. Homolka, V. Marecek, 'Charge Transfer Between Two Immiscible Electrolyte Solutions. Part VIII. Transfer of Alkali and Alkaline Earth-metal Cations Across the Water/Nitrobenzene Interface Facilitated by Synthetic Neutral Carriers', *J. Electroanal. Chem.*, **135**, 265–283 (1982).
161. E. Wang, Z. Yu, D. Qi, C. Xu, 'Alkali and Alkaline Earth metal Ion Transfer Across the Liquid/Liquid Interface Facilitated by Ionophore', *Electroanalysis*, **5**, 149–154 (1993).
162. N. Ogawa, H. Freiser, 'Study of Ion Transfer at a Liquid–Liquid Interface by Current Linear Sweep Voltammetry. 1. The 1,10-Phenanthroline–Phenanthrolinium System', *Anal. Chem.*, **65**, 517–522 (1993).
163. S.A. Dassie, L.M. Yudi, A.M. Baruzzi, 'Comparative Analysis of the Transfer of Alkaline and Alkaline-earth Cations Across the Water/1,2-Dichloroethane Interface', *Electrochim. Acta*, **40**, 2953–2959 (1995).
164. H. Freiser, 'Metal Complexation at the Liquid–Liquid Interface', *Chem. Rev.*, **88**, 611–616 (1988).
165. W. Yu, H. Freiser, 'Electrochemical Study of the Mechanism of Cadmium Extraction with Dithizone', *Anal. Chem.*, **61**, 1621–1623 (1989).
166. T.E. Jones, L.L. Zimmer, L.L. Diaddario, D.B. Rorabacher, L.A. Ochrymowycz, 'Macrocyclic Ligand Ring Size Effects on Complex Stabilities and Kinetics. Copper(II) Complexes of Cyclic Polythiaethers', *J. Am. Chem. Soc.*, **97**, 7163–7165 (1975).
167. L.L. Diaddario, L.L. Zimmer, T.E. Jones, L.S.W.L. Sokol, R.B. Cruz, E.L. Yee, L.A. Ochrymowycz, D.B. Rorabacher, 'Macrocyclic Ligand Complexation Kinetics. Solvent, Ring Size, and Macrocyclic Effect on the Formation and Dissociation Reactions of Copper(II)-Cyclic Polythiaether Complexes', *J. Am. Chem. Soc.*, **101**, 3511–3520 (1979).
168. K. Saito, S. Murakami, A. Muromatsu, E. Sekido, 'Liquid–Liquid Extraction of Copper(II) with Cyclic Tetrathioethers', *Anal. Chim. Acta*, **294**, 329–335 (1994).
169. K. Saito, S. Murakami, A. Muromatsu, E. Sekido, 'Liquid–Liquid Extraction of Copper(I) by Cyclic tetrathio Ethers', *Anal. Chim. Acta*, **237**, 245–249 (1990).
170. S. Tanaka, H. Yoshida, 'Stripping Voltammetry of Silver(I) with a Carbon-paste Electrode Modified with Thiacrown Compounds', *Talanta*, **36**, 1044–1046 (1989).
171. W. Rosen, D.H. Busch, 'Octahedral Nickel(II) Complexes of some Cyclic Polyfunctional Thioethers', *Inorg. Chem.*, **9**, 262–265 (1970).
172. G. Lagger, L. Tomaszewski, M.D. Osborne, B.J. Seddon, H.H. Girault, 'Electrochemical Extraction of Heavy Metal Ions Assisted by Cyclic Thioether Ligands', *J. Electroanal. Chem.*, **451**, 29–37 (1998).

173. Y. Masuda, Y.W. Zhang, C.H. Yan, B.G. Li, 'Studies on the Extraction and Separation of Lanthanide Ions with a Synergistic Extraction System Combined with 1,4,10,13-Tetrathia-7,16-Diazacyclooctadecane and Lauric Acid', *Talanta*, **46**, 203–213 (1998).
174. A.M. Baruzzi, H. Wendt, 'Interfacial Phase Transfer of Amino-complexed Ni(II) Cations. A Model for Solvent Extraction', *J. Electroanal. Chem.*, **279**, 19–30 (1990).
175. D. Homolka, H. Wendt, 'Transfer of Multicomplexed Ions Across the Interface Between Two Immiscible Electrolyte Solutions; Fe(II,III), Ni(II) and Zn(II) Ions Complexed by Bidentate Nitrogen Bases', *Ber. Bunsenges. Phys. Chem.*, **89**, 1075–1082 (1985).
176. R. Fan, X. Wang, 'Semi-differential Cyclic Voltammetry of Cadmium Ion Transfer Across the Water/Nitrobenzene Interface Facilitated by 2,2′-Bipyridine', *J. Electroanal. Chem.*, **261**, 77–88 (1989).
177. E. Wang, Y. Liu, 'Electrochemistry of Cadmium Ion at the Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **214**, 465–472 (1986).
178. H. Alemu, B. Hundhammer, T. Solomon, 'Transfer of Transition Metal–terpyridine Complexes Across the Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **294**, 165–177 (1990).
179. I. Bustero, Y. Cheng, J.C. Mugica, T. Fernandez-Otero, F. Silva, D.J. Schiffrin, 'Electro-assisted Solvent Extraction of Cu^{2+}, Ni^{2+} and Cd^{2+}', *Electrochim. Acta*, **44**, 29–38 (1998).
180. Y. Cheng, D.J. Schiffrin, 'Copper(II) Ion Complexation by the Tetradentate Phosphorus–Nitrogen Ligand P_2N_2 at the Water/1,2-DCE Interface', *Inorg. Chem.*, **33**, 765–769 (1994).
181. H. Li, R.B. Smart, 'Square Wave Catalytic Stripping Voltammetry of Molybdenum Complexed with Dihydroxynaphthalene', *J. Electroanal. Chem.*, **429**, 169–174 (1997).
182. J. Wang, J. Lu, R. Setiadji, 'Adsorptive Stripping Measurements of Trace Aluminum in the Presence of Cupferron', *Talanta*, **40**, 351–354 (1993).
183. J. Wang, J. Lu, K. Olsen, 'Measurement of Ultratrace Levels of Chromium by Adsorptive-catalytic Stripping Voltammetry in the Presence of Cupferron', *Analyst*, **117**, 1913–1917 (1992).
184. J. Wang, R. Setiadji, 'Selective Determination of Trace Uranium by Stripping Voltammetry Following Adsorptive Accumulation of the Uranium–Cupferron Complex', *Anal. Chim. Acta*, **264**, 205–211 (1992).
185. J. Wang, B. Tian, J. Lu, 'Adsorptive-catalytic Stripping Measurements of Ultratrace Vanadium in the Presence of Cupferron and Bromate', *Talanta*, **39**, 1273–1276 (1992).
186. O. Abollino, M. Aceto, E. Mentasti, C. Sarzanini, C.M.G. van den Berg, 'Determination of Trace Europium by Adsorptive Cathodic Stripping Voltammetry after Complexation with Cupferron', *Electroanalysis*, **9**, 444–448 (1997).
187. J. Koryta, Y.N. Kozlov, M. Skalicky, 'Potassium Ion Transfer Across the Nitrobenzene/Water Interface Facilitated by Valinomycin', *J. Electroanal. Chem.*, **234**, 355–360 (1987).
188. G. Du, J. Koryta, W. Ruth, P. Vanysek, 'Diversity of Ion Carrier Functions of Monensin: A Study Using Voltammetry at the Interface of Two Immiscible Electrolyte Solutions', *J. Electroanal. Chem.*, **159**, 413–420 (1983).
189. F. Reymond, G. Steyaert, A. Pagliara, P.-A. Carrupt, B. Testa, H. Girault, 'Transfer Mechanism of Ionic Drugs: Piroxicam as an Agent Faciliting Proton Transfer', *Helv. Chim. Acta*, **79**, 1651–1669 (1996).
190. E. Wang, Y. Liu, 'Cyclic Voltammetry and Chronopotentiometry with Cyclic Linear Current Scanning of Terramycin at the Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **214**, 459–464 (1986).
191. S.M. Hassan, W.H. Mahmoud, M. Othman, 'A Novel Potassium Ion Membrane Sensor Based on Rifamycin Neutral Ionophore', *Talanta*, **44**, 1087–1094 (1997).
192. E. Wang, Z. Yu, N. Li, 'Anaesthetic Lidocaine and Dicaine Transfer across Liquid–Liquid Interfaces', *Electroanalysis*, **4**, 905–909 (1992).
193. Z. Yoshida, H. Aoyagi, Y. Meguro, Y. Kitatsuji, S. Kihara, 'Voltammetric Study on the Transfer of U, Np, and Pu Ions at the Aqueous–Organic Interface Facilitated by Phosphine Oxides', *J. Alloys Compd.*, **213/214**, 324–327 (1994).
194. D.P. Zollinger, E. Bulten, A. Christenhusz, M. Bos, W.E. Van Der Linden, 'Computerized Conductometric Determination of Stability Constants of Complexes of Crown Ethers with Alkali Metal Salts and with Neutral Molecules in Polar Solvents', *Anal. Chim. Acta*, **198**, 207–222 (1987).
195. R.D. Hancock, I. Cukrowski, J. Baloyi, J. Mashishi, 'The Affinity of Bismuth(III) for Nitrogen-donor Ligands', *J. Chem. Soc., Dalton Trans.*, 2895–2899 (1993).
196. A. Ohki, J.-P. Lu, R.A. Bartsch, 'Effect of Side-arm Variation in Dibenzo-16-crown-5 Compounds on the Potentiometric Selectivity for Sodium Ion', *Anal. Chem.*, **66**, 651–654 (1994).
197. I. Cukrowski, E. Cukrowska, R.D. Hancock, G. Anderegg, 'The Effect of Chelate Ring Size on Metal Ion Size-based Selectivity in Polyamine Ligands Containing Pyridyl and Saturated Nitrogen Donor Groups', *Anal. Chim. Acta*, **312**, 307–321 (1995).
198. I.Cukrowski, 'A Polarographic Method of Speciation of Labile Metal–Ligand Systems Based on Mass-balance Equations. A Differential Pulse Polarographic Study at Fixed Ligand to Metal Ratio and Varied pH', *Anal. Chim. Acta*, **336**, 23–36 (1996).
199. M.H. Mazor, J.A. McCammon, T.P. Lybrand, 'Molecular Recognition in Nonaqueous Solvents. 2. Structural and Thermodynamic Analysis of Cationic Selectivity of 18-Crown-6 in Methanol', *J. Am. Chem. Soc.*, **112**, 4411–4419 (1990).

200. L. Troxler, G. Wipff, 'Conformation and Dynamics of 18-Crown-6, Cryptand 222, and their Cation Complexes in Acetonitrile Studied by Molecular Dynamics Simulations', *J. Am. Chem. Soc.*, **116**, 1468–1480 (1994).
201. L.X. Dang, P.A. Kollman, 'Free Energy of Association of the K+:18-Crown-6 Complex in Water: A New Molecular Dynamics Study', *J. Phys. Chem.*, **99**, 55–58 (1995).
202. K. Tohda, Y. Umezawa, S. Yoshiyagawa, S. Hashimoto, M. Kawasaki, 'Cation Permselectivity at the Phase Boundary of Ionophore-incorporated Solvent Polymeric Membranes as Studied by Optical Second Harmonic Generation', *Anal. Chem.*, **67**, 570–577 (1995).
203. M.J. Crawford, J.G. Frey, T.J. VanderNoot, Y. Zhao, 'Investigation of Transport Across an Immiscible Liquid–Liquid Interface. Electrochemical and Second Harmonic Generation Studies', *J. Chem. Soc., Faraday Trans.*, **92**, 1369–1373 (1996).
204. S.N. Tan, H.H. Girault, 'Potassium Transfer Facilitated by Monoaza-18-crown-6 Across the Water–1,2-Dichloroethane Interface', *J. Electroanal. Chem.*, **332**, 101–112 (1992).
205. L.B. Kier, B. Testa, *Complexity and Emergence in Drug Research*, Academic Press, London, 1995.
206. L.Z. Benet, B.Y.T. Perotti, 'Drug Absorption, Distribution and Elimination', in: *Burger's Medicinal Chemistry and Drug Discovery*, ed. M.E. Wolff, John Wiley & Sons, New York, 113–128, Vol. 1, 1995.
207. R.A. Conradi, P.S. Burton, R.T. Borchardt, 'Physicochemical and Biological Factors that Influence a Drug's Cellular Permeability by Passive Diffusion', in: *Lipophilicity in Drug Action and Toxicology*, eds. V. Pliska, B. Testa, H. van de Waterbeemd, VCH, Weinheim, 233–252, Vol. 4, 1996.
208. J.B. Taylor, P.D. Kennewell, *Modern Medicinal Chemistry*, Ellis Horwood, New York, 1993.
209. V.H.L. Lee, *Peptide and Protein Delivery*, Marcel Dekker, New York, 1991.
210. K.L. Audus, T.J. Raub, *Biological Barriers to Protein Delivery*, Plenum Press, New York, 1993.
211. C.J.H. Porter, W.N. Charman, 'Uptake of Drugs into the Intestinal Lymphatics after Oral Administration', *Adv. Drug Del. Rev.*, **25**, 71–89 (1997).
212. P.D.Kennewell, 'The Architecture of the Cell', in *Comprehensive Medicinal Chemistry: The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds*, eds. C. Hansch, P.G. Sammes, J.B. Taylor, Pergamon Press, Oxford, 169–194, Vol. 1, 1990.
213. H. van de Waterbeemd, B. Testa, *The Parametrization of Lipophilicity and Other Structural Properties in Drug Research*, Academic Press, London, 1987.
214. M.G. Davis, C.N. Manners, D.W. Payling, D.A. Smith, C.A. Wilson, 'Gastrointestinal Absorption of the Strongly Acidic Drug Proxicromil', *J. Pharm. Sci.*, **73**, 949–953 (1984).
215. W.A. Banks, A.J. Kastin, 'Peptides and the Blood–Brain Barrier: Lipophilicity as a Predictor of Permeability', *Brain Res. Bull.*, **15**, 287–292 (1985).
216. D.A. Smith, K. Brown, M.G. Neale, 'Chromone-2-carboxylic Acids: Roles of Acidity and Lipophilicity in Drug Disposition', *Drug Metab. Rev.*, **16**, 365–388 (1986).
217. C.J. Alcorn, R.J. Simpson, D.E. Leahy, T.J. Peters, 'Partition and Distribution Coefficients of Solutes and Drugs in Brush Border Membrane Vesicles', *Biochem. Pharmacol.*, **45**, 1775–1782 (1993).
218. A. Pagliara, P.-A. Carrupt, G. Caron, P. Gaillard, B. Testa, 'Lipophilicity Profiles of Ampholytes', *Chem. Rev.*, **97**, 3385–3400 (1997).
219. V. Pliska, B. Testa, H. van de Waterbeemd, (eds.), *Lipophilicity in Drug Action and Toxicology. Methods and Principles in Medicinal Chemistry*, VCH, Weinheim, Vol. 4, 1996.
220. C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, 1995.
221. F. Reymond, G. Steyaert, P.-A. Carrupt, B. Testa, H.H. Girault, 'Mechanism of Transfer of a Basic Drug Across the Water/1,2-Dichloroethane Interface: The Case of Quinidine', *Helv. Chim. Acta*, **79**, 101–117 (1996).
222. H. Doe, K. Yoshioka, T. Kitagawa, 'Voltammetric Study of Protonated 1,10-Phenanthroline Cation Transfer Across the Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **324**, 69–78 (1992).
223. P. Vanysek, M. Behrendt, 'Investigation of Acetylcholine, Choline and Acetylcholinesterase at the Interface of the Two Immiscible Electrolyte Solutions', *J. Electroanal. Chem.*, **130**, 287–292 (1981).
224. V. Marecek, Z. Samec, 'Electrolysis at the Interface Between Two Immiscible Electrolyte Solutions: Determination of Acetylcholine by Differential Pulse Stripping Voltammetry', *Anal. Lett.*, **14**, 1241–1253 (1981).
225. Z. Samec, V. Marecek, D. Homolka, 'Charge Transfer Between Two Immiscible Electrolyte Solutions. Part IX. Kinetics of the Transfer of Choline and Acetylcholine Cations Across the Water/Nitrobenzene Interface', *J. Electroanal. Chem.*, **158**, 25–36 (1983).
226. D. Homolka, V. Marecek, Z. Samec, K. Base, W. Wendt, 'The Partition of Amines Between Water and an Organic Solvent Phase', *J. Electroanal. Chem.*, **163**, 159–170 (1984).
227. O. Dvorak, V. Marecek, Z. Samec, 'Selective Complexation of Biogenic Amines by Macrocyclic Polyethers at a Liquid/Liquid Interface', *J. Electroanal. Chem.*, **300**, 407–413 (1991).
228. T. Ohkouchi, T. Kakutani, M. Senda, 'Electrochemical Study of the Transfer of Uncouplers Across the Organic/Aqueous Interface', *Bioelectrochem. Bioenerg.*, **25**, 71–80 (1991).
229. T. Ohkouchi, T. Kakutani, M. Senda, 'Electrochemical Theory of the Transfer of Protons Across a Biological Membrane Facilitated by Weak Acid Uncouplers Added

to the Medium', *Bioelectrochem. Bioenerg.*, **25**, 81–89 (1991).
230. E. Wang, H. Wang, Z. Yu, 'Electrochemical Study of Pyrazolone Derivatives at the Liquid/Liquid Interface', *Electroanalysis*, **6**, 1020–1023 (1994).
231. T. Osakai, T. Kakutani, M. Senda, 'Kinetics of the Transfer of Picrate Ions at the Water/Nitrobenzene Interface', *Bull. Chem. Soc. Jpn.*, **58**, 2626–2633 (1985).
232. T. Wandlowski, V. Marecek, Z. Samec, 'Kinetic Analysis of the Picrate Ion Transfer Across the Interface Between Two Immiscible Electrolyte Solutions from Impedance Measurements at the Equilibrium Potential', *J. Electroanal. Chem.*, **242**, 291–302 (1988).
233. O. Shirai, Y. Yoshida, M. Matsui, K. Maeda, S. Kihara, 'Voltammetric Study on the Transport of Ions of Various Hydrophobicity Types through Bilayer Lipid Membranes Composed of Various Lipids', *Bull. Chem. Soc. Jpn.*, **69**, 3151–3162 (1996).
234. F. Reymond, G. Steyaert, P.-A. Carrupt, B. Testa, H.H. Girault, 'Ionic Partition Diagrams: A Potential-pH Representation', *J. Am. Chem. Soc.*, **118**, 11951–11957 (1996).
235. F. Reymond, P.F. Brevet, P.-A. Carrupt, H.H. Girault, 'Cyclic Voltammetry for the Transfer of Multiple Charged Ions at Large ITIES: General Computational Methodology and Application to Simple and Facilitated Ion Transfer Reactions', *J. Electroanal. Chem.*, **424**, 121–139 (1997).
236. Y.N. Kozlov, J. Koryta, 'Determination of Tetracycline Antibiotics by Voltammetry at the Interface of Two Immiscible Electrolyte Solutions', *Anal. Lett.* **B16**, 255–263 (1983).
237. L.M. Yudi, A.M. Baruzzi, V.M. Solis, 'Quantitative Determination of Erythromycin and its Hydrolysis Products by Cyclic Voltammetry at the Interface Between Water and 1,2-Dichloroethane', *J. Electroanal. Chem.*, **360**, 211–219 (1993).
238. L.M. Yudi, E. Santos, A.M. Baruzzi, V.M. Solis, 'Erythromycin Transfer across the Water/1,2-Dichloroethane Interface Modified by a Phospholipid Monolayer', *J. Electroanal. Chem.*, **379**, 151–158 (1994).
239. J. Di, X. Xu, J. Luo, 'Determination of Minocycline by Semi-differential Cyclic Voltammetry at a Liquid/Liquid Interface', *Anal. Lett.*, **29**, 2691–2700 (1996).
240. K. Arai, M. Ohsawa, F. Kusu, K. Takamura, 'Drug Ion Transfer Across an Oil/Water Interface and Pharmacological Activity', *Bioelectrochem. Bioenerg.*, **31**, 65–76 (1993).
241. K. Arai, F. Kusu, N. Tsuchiya, S. Fukuyama, 'Ion Transfer of Weak Electrolytes Across the Oil/Water Interface. Ion Transfer of Scopolamine and Lidocaine', *Denki Kagaku*, **62**, 840–845 (1994).
242. K. Arai, F. Kusu, K. Takamura, 'Electrochemical Behavior of Drugs at Oil/Water Interfaces', in *Liquid–Liquid Interfaces. Theory and Methods*, eds. A.G. Volkov, D.W. Deamer, CRC Press, Boca Raton, FL, 375–400, 1996.
243. Z. Samec, J. Langmaier, A. Trojanek, E. Samcova, J. Malek, 'Transfer of Protonated Anesthetics Across the Water/*o*-Nitrophenyl Octyl Ether Interface: Effect of the Ion Structure on the Transfer Kinetics and Pharmacological Activity', *Anal. Sci.*, **14**, 35–41 (1998).
244. M. Pourbaix, *Atlas d'Equilibres Electrochimiques*, Gauthier-Villars, Paris, 1963.
245. F. Reymond, P.-A. Carrupt, B. Testa, H.H. Girault, 'Charge and Delocalization Effects on the Lipophilicity of Protonable Drugs', *Chem. Eur. J.*, **5**, 39–47 (1999).
246. F. Reymond, V. Chopineaux-Courtois, G. Steyaert, G. Bouchard, P.-A. Carrupt, B. Testa, H.H. Girault, 'Ionic Partition Diagrams of Ionizable Solutes: pH Lipophilicity Profiles, Transfer Mechanisms and Charge Effects on Solvation', *J. Electroanal. Chem.*, **462**, 235–250 (1999).
247. F. Reymond, G. Steyaert, P.-A. Carrupt, D. Morin, J.-P. Tillement, H.H. Girault, B. Testa, 'The pH-Partition Profile of the Anti-ischemic Drug Trimetazidine May Explain Its Reduction of Intracellular Acidosis', *Pharm. Res.*, **16**, 616–624 (1999).
248. V. Chopineaux-Courtois, F. Reymond, G. Bouchard, P.-A. Carrupt, B. Testa, H.H. Girault, 'Effects of Charge and Intramolecular Structure on the Lipophilicity of Nitrophenols', *J. Am. Chem. Soc.*, **121**, 1743–1747 (1999).
249. K. Kontturi, L. Murtomaki, 'Electrochemical Determination of Partition Coefficients of Drugs', *J. Pharm. Sci.*, **81**, 970–975 (1992).
250. G. Steyaert, G. Lisa, P.-A. Carrupt, B. Testa, F. Reymond, H.H. Girault, 'Solvatochromic Analysis of Partition Coefficients in the Water/1,2-Dichloroethane System', *J. Chem. Soc., Faraday Trans.*, **93**, 401–406 (1997).

Microbalance, Electrochemical Quartz Crystal

Michael D. Ward
University of Minnesota, Minneapolis, USA

1 Introduction 1
2 Theory and Principles 2
3 Electrochemical Methods and Examples of Applications 6
3.1 Experimental Apparatus and Operation 6
3.2 Electrochemical Measurements of Crystal Resonant Properties in Fluids 9
3.3 Data Interpretation for Ideally Behaved Systems 10
3.4 Electrochemical Quartz Crystal Microbalance Investigations of Metal Electrodes and Films 12
3.5 Electrochemical Quartz Crystal Microbalance Investigations of Thin-film Growth 14
3.6 Electrochemically Active Polymer Films 16
3.7 Electrochemical Quartz Crystal Microbalance Investigations of Conducting Polymers 18
3.8 Electrochemical Quartz Crystal Microbalance Investigations of Biologically Relevant Films 20
4 Impedance Analysis Methods 21
5 Emerging Methods 23
6 Summary 24
Abbreviations and Acronyms 24
Related Articles 25
References 25

The need for direct measurement of interfacial events at solid surfaces under ambient conditions has provoked surface scientists and electrochemists to develop direct in situ methods. Such processes include electroless or electrochemical metal deposition, insertion of ions into polymer ion-exchange films, growth of oxide films on metals, or the loss of material from corrosion processes. An increasing interest in molecular films has prompted development of methods for measuring adsorption of molecules from vapor or liquid phases. Many of these processes share a common feature, namely they are accompanied by changes in mass at the solid surface.

During the 1990s a new analytical method for the in situ examination of interfacial electrode processes, the electrochemical quartz crystal microbalance (EQCM), has emerged that has substantially influenced electrochemical science. This method relies on a single crystal of quartz that has been cut into a thin wafer and coated with gold electrodes on both sides. These electrodes are used to provide an alternating electric field so that the crystal is vibrated at a specific resonant frequency, while one of the electrodes is simultaneously used as a working electrode in an electrochemical cell. The resonant frequency shifts upon changes in mass that occur on the working electrode during an electrochemical process, with sensitivity as high as 100 pg cm^{-2} of electrode area. Recent developments also have illustrated that the quartz crystal microbalance (QCM) is capable of measuring the viscosity and density of liquids near the QCM surface. The purpose of this article is to provide the reader with a fundamental understanding of the EQCM and several illustrative examples of applications that demonstrate its unique capabilities, as well as the technical details required to build this apparatus.

1 INTRODUCTION

In the 1980s and 1990s surface scientists and electrochemists have made considerable strides in sophisticated instrumental techniques for probing the structure and composition of solid interfaces and electrodes, motivated by the need to measure and understand interfacial processes. However, many of these techniques involve examination of solid surfaces outside the medium in which they are typically used. For example, ultrahigh vacuum studies of electrode and catalyst surfaces provide invaluable data concerning structure and composition. The relevance of data obtained from such ex situ methods to properties and applications under ambient conditions, particularly applications involving immersion of the surface in solution, can be questionable.

The need for direct measurement of interfacial events at solid surfaces under ambient conditions has provoked surface scientists and electrochemists to develop direct in situ methods. In addition to providing data that can be used for understanding fundamental mechanistic details, in situ approaches can be useful in sensor applications where measurement under specific, and sometimes harsh, conditions is required. Such processes include electroless or electrochemical metal deposition, insertion of ions into polymer ion-exchange films, growth of oxide films on metals, or the loss of material from corrosion processes.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

An increasing interest in molecular films has prompted development of methods for measuring adsorption of molecules from vapor or liquid phases. Many of these processes share a common feature, namely they are accompanied by changes in mass at the solid surface.

In the 1990s, a new analytical method for the in situ measurement of interfacial electrode processes, the EQCM, has emerged that has substantially influenced electrochemical science. This method relies on the piezoelectric properties of a single crystal of quartz that has been cut into a thin wafer. The word piezoelectric derives from the word *piezein*, meaning to press. Hence, the piezoelectric effect hinges on ''pressure electricity,'' a phenomenon first observed by Jacques and Pierre Curie when they discovered that mechanical stress applied to the surfaces of certain crystals, including quartz, resulted in an electrical potential across the crystal. Shortly afterwards, the converse piezoelectric effect – a mechanical strain produced by application of a electric potential across the crystal – was discovered. This effect is sometimes referred to as the converse piezoelectric effect. The motor generator properties have long been associated with underwater sound transducers (sonar), and electromechanical devices such as speakers, microphones, and phonograph pickups.

The QCM earns its name from its ability to measure the mass of thin films that have adhered to its surface. The QCM generally comprises a thin quartz wafer with a diameter of 0.25–1.0 in ($1\,\text{in} = 2.54 \times 10^{-2}\,\text{m}$), sandwiched between two metal electrodes which are used to establish an electric field across the crystal. If an alternating electric field and appropriate electronics are used, the crystal can be made to oscillate at its resonant frequency. Most crystals of current interest resonate between 5 and 30 MHz. The measured frequency is dependent upon the combined thickness of the quartz wafer, metal electrodes, and material deposited on the QCM surface. Because the resonance is very sharp, high-precision frequency measurements allow the detection of minute amounts of deposited material, as small as $100\,\text{pg}\,\text{cm}^{-2}$.

Early applications of the QCM involved the well-documented measurement of metal deposition in high-vacuum metal evaporators, which is still widely practiced. This allows for real-time rapid measurement of film thicknesses with ångström resolution ($1\,\text{Å} = 10^{-10}\,\text{m}$). Advances in QCM methodology in the 1990s allow for dynamic measurements of minute mass changes at surfaces, thin films, and electrode interfaces prepared on the quartz crystal, while the surface is immersed in liquid. The capability for direct real-time highly sensitive mass measurement in the liquid phase offers opportunities not available by other means. Developments also have illustrated that the QCM is capable of measuring the viscosity and density of liquids near the QCM surface. Combining this technique with electrochemical instrumentation allows simultaneous measurement of mass and electrochemical variables such as electrochemical potential, current, and charge.

The EQCM has emerged as a powerful technique capable of detecting very small mass changes at the electrode surface that accompany electrochemical processes. This relatively simple technique requires, in addition to conventional electrochemical equipment, an inexpensive radiofrequency (RF) oscillator, a frequency counter, and commercially available AT-cut quartz crystals. EQCM has evolved into a routine experimental method used in numerous electrochemical laboratories. The purpose of this article is to provide the reader with a fundamental understanding of the EQCM and several illustrative examples of applications that demonstrate its unique capabilities. It is hoped that this will enable readers to add the EQCM to their battery of electrochemical methods. Other review articles can be consulted for different emphasis or greater detail.[1–3]

2 THEORY AND PRINCIPLES

In order to understand the operation of the EQCM a fundamental understanding of the piezoelectric effect is required. In 1880, Jacques and Pierre Curie discovered that a mechanical stress applied to the surfaces of various crystals, including quartz, rochelle salt ($NaKC_4H_4O_6 \cdot 4H_2O$), and tourmaline, resulted in an electrical potential across the crystal whose magnitude was proportional to the applied stress.[4] This behavior is referred to as the piezoelectric effect which is derived from the Greek word *piezein* meaning ''to press.'' This property only exists in materials that are acentric, that is, those that crystallize into noncentrosymmetric space groups. A single crystal of an acentric material will possess a polar axis due to dipoles associated with the arrangement of atoms in the crystalline lattice. The charge generated in a quartz crystal under mechanical stress is a manifestation of a change in the net dipole moment because of the physical displacement of the atoms and a corresponding change in the net dipole moment. This results in a net change in electrical charge on the crystal faces, the magnitude and direction of which depends upon the relative orientation of the dipoles and the crystal faces. Following their initial discovery, the Curies discovered the converse piezoelectric effect, in which the application of a potential across these crystals resulted in a corresponding mechanical strain. It is this effect that is the operational basis of the EQCM.

The EQCM is actually the electrochemical version of the QCM, which has long been used for frequency control

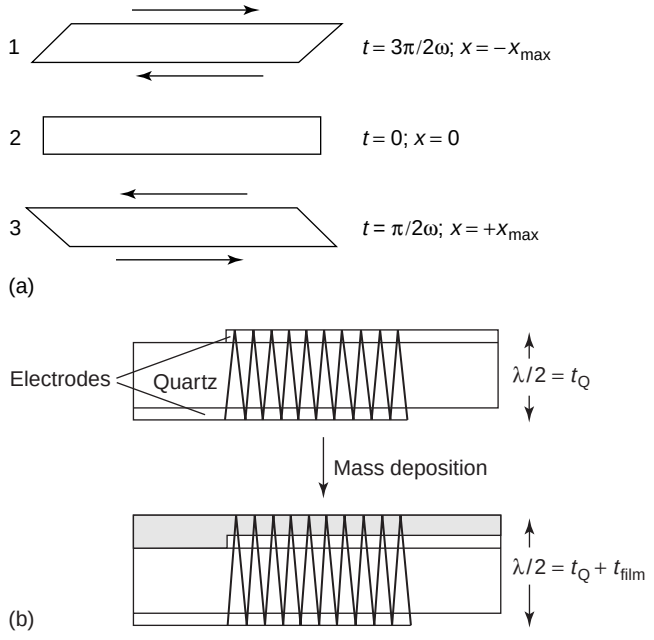

Figure 1 (a) Schematic representation of the shear vibration of an AT-cut quartz resonator. The time at which the crystal achieves maximum strain during oscillation is indicated. The crystal has maximum kinetic energy at $x = 0$, but maximum potential energy at $x = \pm x_{max}$, similar to classical oscillators. (b) Schematic representation of the transverse shear wave in a quartz crystal with excitation electrodes and a composite resonator comprising the quartz crystal, electrodes, and a thin layer of a foreign material. The acoustic wavelength is longer in the composite resonator because of the greater thickness, resulting in a lower frequency compared to the quartz crystal.

and mass sensing in vacuum and air. The QCM consists of a thin, AT-cut quartz crystal with very thin metal electrode ''pads'' on opposite sides of the crystal. The terminology ''AT'' simply refers to the orientation of the crystal with respect to its large faces; this particular crystal is fabricated by slicing through a quartz rod at an angle of approximately 35° with respect to the crystallographic x axis. The electrode pads overlap in the center of the crystal with tabs extending from each to the edge of the crystal where electrical contact is made. When an electrical potential is applied across the crystal using these electrodes, the AT-cut quartz crystal experiences a mechanical strain in the shear direction. Crystal symmetry dictates that the strain induced in a piezoelectric material by an applied potential of one polarity will be equal and opposite in direction to that resulting from the opposite polarity (Figure 1). Therefore, an alternating potential across the crystal causes vibrational motion of the quartz crystal with the vibrational amplitude parallel to the crystal surface and in the x-direction. This oscillatory behavior and the electromechanical ''motor generator'' properties are the basis of numerous applications, including the QCM, sonar transducers, speakers, microphones, phonograph pickups, and quartz digital watches. It is important to note that the direction of the crystal vibration is critical for liquid-phase applications. An AT-cut crystal vibrates in the shear mode, parallel to the crystal–liquid interface. Consequently, damping of the crystal vibration by a contacting fluid are minimized (see below).

The shear vibrational motion of the quartz crystal results in a transverse acoustic wave that propagates back and forth across the thickness of the crystal between the crystal faces. Accordingly, a standing wave condition is established in the quartz resonator when the acoustic wavelength is equal to twice the combined thickness of the crystal and electrodes. The frequency, f_0, of the acoustic wave fundamental mode is given by Equation (1), where v_{tr} is the transverse velocity of sound in AT-cut quartz ($3.34 \times 10^4\ m\,s^{-1}$) and t_Q is the resonator thickness. Some useful parameters and relationships for quartz resonators are provided in Table 1. An assumption is commonly made that the velocity of sound in quartz and the electrodes is identical. While this is not rigorously true, for small electrode thicknesses the error introduced by

Table 1 Some useful parameters and relationships for quartz resonators

Velocity of sound in quartz (v_Q)	$3340\ m\,s^{-1}$
Shear modulus of quartz (μ_Q)	$2.947 \times 10^{11}\ g\,cm^{-1}\,s^{-2}$
Density of quartz (ρ_Q)	$2.648\ g\,cm^{-3}$
Thickness–frequency relationship	$t_Q = v_Q/2f_0$
Amplitude of oscillation (thickness shear mode)	10–100 nm
Shear velocity	$v(z, t) = v_0 \exp(k_1 z) \times \cos(k_1 z - 2pf_0 t)$
Propagation constant	$k_1 = (\pi f_0 \rho_1/\eta_1)^{1/2}$
Decay length of shear wave	$1/k_1$
Decay length of 5-MHz shear wave in water	250 nm
Typical equivalent circuit values	
C_1 (energy stored during oscillation)	23×10^{-15} F
R_1 (energy lost during oscillation, air)	ca. 100 Ω
R_1 (energy lost during oscillation, water)	ca. 1000 Ω
L_1 (inertial component related to mass displaced during oscillation)	45×10^{-3} H
C_0 (dielectric capacitance of quartz)	10^{-12} F
Series resonance, f_s	$f_s = [2\pi(L_1C_1)^{1/2}]^{-1}$
Quality factor	$Q = (2pf_sC_1R_1)^{-1} = 2\pi f_s L_1/R_1$
Quality factor	Q = Peak width at half height of conductance/f_s

this approximation is negligible. The acoustic velocity is dependent upon the modulus and density of the crystal. The quartz crystal surface is at an antinode of the acoustic wave, and therefore the acoustic wave propagates across the interface between the crystal and a foreign layer on its surface. If it is assumed that the acoustic velocity in the foreign layer, and its density, are identical to those for quartz (cf. the assumption for the metal electrodes), a change in thickness of the foreign layer is tantamount to a change in the thickness of the quartz crystal. Under these conditions, a fractional change in thickness results in a fractional change in the resonant frequency; appropriate substitutions yield the well-known Sauerbrey equation, Equation (2), where Δf is the measured frequency change, f_0 the frequency of the quartz resonator prior to a mass change, Δm the mass change, A the piezoelectrically active area, ρ_Q the density of quartz ($2.648\,g\,cm^{-3}$) and μ_Q the shear modulus of AT-cut quartz ($2.947 \times 10^{11}\,dyn\,cm^{-2}$) ($1\,dyn = 10^{-5}\,N$).

$$f_0 = \frac{v_{tr}}{2t_Q} \tag{1}$$

$$\Delta f = \frac{-2f_0^2 \Delta m}{A\sqrt{\mu_Q \rho_Q}} \tag{2}$$

This equation is the primary basis of most QCM and EQCM measurements wherein mass changes occurring at the electrode interface are evaluated directly from the frequency changes of the quartz resonator. It is generally considered to be accurate as long as the thickness of the film added to the QCM is less than 2% of the quartz crystal thickness. With this constraint, the errors resulting from the discrepancy between the acoustic propagation characteristics in quartz and the film are minimal. Deviations from Equation (2) due to higher mass loadings may be compensated, however, by use of the ''Z-match'' method.[5] While this method has been used for vacuum applications, it has yet to be employed in EQCM applications. Typical operating frequencies of the EQCM of the EQCM lie within the range 5–10 MHz, although recently the operation of 30 MHz quartz crystals in EQCM applications has been achieved.[6] These operating frequencies provide for mass detection limits approaching $1\,ng\,cm^{-2}$.

The reader may better understand the mass sensing properties of the QCM by comparing the motion of the quartz crystal to other oscillating systems, such as a vibrating string, a pendulum, or a mass on a spring. In all cases the amplitude is defined by the initial energy input and the resonant frequencies are defined by characteristics of mass and length. The quartz crystal motion can be described as moving about the $x = 0$ rest point between limits of $-x_{max}$ and $+x_{max}$. The magnitude of x_{max} will depend upon the applied alternating voltage across the crystal. As with the more familiar oscillating systems, the potential energy of the crystal is at a maximum at $x = \pm x_{max}$, whereas the kinetic energy is at a maximum at $x = 0$. The effect of mass (or thickness) changes on the quartz resonant frequency can be understood by analogy to a classical system such as a vibrating string. Standing waves can exist in a vibrating string if their wavelengths are integral divisors of $2l$, where l is the length of the string. The fundamental frequency, f_0, is given by Equation (3), where S is the tension on the string and m_l is the mass per unit length.

$$f_0 = \frac{(S/m_l)^{1/2}}{2l} \tag{3}$$

An increase in the mass or length of the string therefore results in a decrease in f_0. This is identical to increasing the thickness of a quartz crystal, with a dimensional increase resulting in a longer standing wave propagation distance and a corresponding reduction in the frequency. The stress applied to the vibrating string is analogous to the modulus of the quartz crystal; an increase in either of these quantities increases the velocity of the standing wave.

The key distinguishing feature of quartz resonators is their negligible energy dissipation during oscillation. While a pendulum may lose considerable energy during oscillation because of friction, a quartz crystal loses only a minute amount due to phonon interactions that produce heat, vibrational damping by the mounting components, and acoustical losses to the environment. This property is generally characterized by the quality factor, Q, which is the ratio of the energy stored to energy lost during a single oscillation. For quartz crystals, this quantity can exceed 100 000. Low energy losses in oscillating systems are manifested as high accuracy. As a system loses more energy during oscillation the period of the oscillation becomes less well defined. This is the basis for the widespread use of quartz crystals in timepieces and frequency control elements. In fact, the frequency of a typical quartz crystal can be determined to an accuracy of 1 part in 10^8. It is this precision that makes the QCM and EQCM so useful. In liquid applications, including the EQCM, Q will generally have values of 1000–3000, indicative of energy damping by the fluid. Nevertheless, the quartz crystals will still perform acceptably at these levels.

The vibration of the quartz crystal parallel to the QCM–liquid interface results in the radiation of a shear wave into the liquid (Figure 2). The instantaneous shear wave velocity decays as an exponentially damped cosine function according to Equation (4), where k is the propagation constant, z the distance from the resonator surface, A the maximum amplitude of the shear wave, and w the angular frequency. The inverse of the propagation

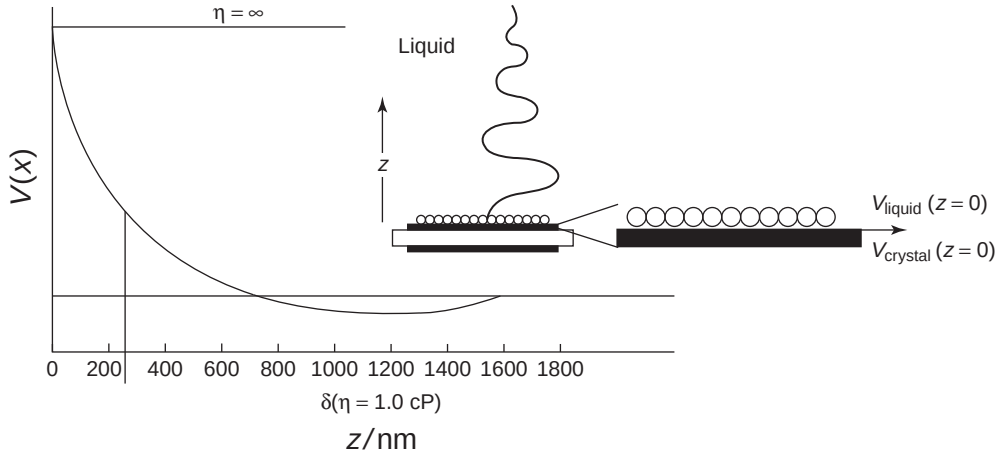

Figure 2 Description of the shear wave propagation in a Newtonian fluid in terms of the shear velocity in the *x*-direction (parallel to the resonator–liquid interface) as a function of distance from the resonator–liquid interface. If the viscosity of the film in contact with the resonator is infinitely large, the acoustic wave will propagate without loss. In water, the amplitude of the acoustic wave decays, with the decay length $\delta = 2500$ Å. The decay length represents the distance at which the amplitude of the shear wave decays to 1/e of its maximum value at the interface. The circles represent the layer of liquid molecules at the interface, at which a ''no-slip'' condition generally is assumed. That is, these molecules move with the same velocity and amplitude as the resonator ($V_{\text{liquid}} = V_{\text{crystal}}$).

constant k is the decay length, δ, which is given by Equation (5), where ρ_L and η_L are the liquid density and viscosity, respectively. This leads to Equation (6), which gives the dependence of the resonant frequency on $(\rho_L\eta_L)^{1/2}$.[7]

$$V_x(z, t) = Ae^{-kz}\cos(kz - wt) \tag{4}$$

$$\delta = \sqrt[z]{\frac{\eta_L}{\pi f_0 \rho_L}} \tag{5}$$

$$\Delta f = -f_0^{3/2}\left(\frac{\rho_L\eta_L}{\pi\rho_Q\mu_Q}\right)^{1/2} \tag{6}$$

This effect of a contacting fluid medium is evidenced by the approximately 750-Hz decrease in resonant frequency that occurs when a QCM is immersed in water. This decrease can be attributed to the effective mass of liquid contained in this decay length. It should be noted that Equations (4–6) rely on the assumption of a ''no-slip'' condition at the QCM–liquid interface. That is, the molecular layer of liquid directly in contact with the vibrating region of the QCM surface moves with the same velocity and amplitude as the vibrating region. The shear wave then decays in the liquid because of the velocity gradient along the direction normal to the surface. This is a consequence of the fluid viscosity. While there are some reports claiming deviations from the no-slip condition, these are sparse and not yet conclusive.[8,9,111]

Inspection of Equation (6) reveals that an increase in either the density or viscosity results in a decrease in the resonant frequency. This can be better understood by considering two extremes: in air, the quantity $(\rho_L\eta_L)^{1/2}$ is negligible, but if the liquid was actually a rigid solid, $(\rho_L\eta_L)^{1/2}$ would be substantial. In the case of a rigid solid with $\eta = \infty$, the acoustic wave will propagate indefinitely and the frequency change would reflect the true thickness of the solid. If the thickness of this rigid film was large, crystal oscillation would be difficult because of the significant damping by the large mass. Therefore, it is the decay of the shear wave in liquids that enables operation of the QCM and EQCM in liquid media.

As the viscosity of the fluid becomes larger, the accuracy and performance of the submerged quartz resonator diminishes. These effects can be especially important when the EQCM is modified with polymer films, which may undergo changes in $(\rho_L\eta_L)^{1/2}$ during measurements. In addition, polymer films may be viscoelastic, which can further complicate interpretation of EQCM data (see below). Indeed, there are several examples of significant QCM responses to these changes under conditions where mass changes are not operative.[10]

Another effect that needs to be considered when quartz resonators are immersed in liquids is the microscopic roughness of the resonator surface. The cavities on a rough surface can trap liquid which will be manifested as an additional mass on the surface. The amount of trapped liquid will depend upon the cavity geometry and size. This effect has been inferred in gold and copper surfaces on the EQCM, in which the electrode surfaces were roughened during electrochemical cycling through the oxide regions of the metal electrodes.[11–13] An extensive study of these gold surfaces involved comparison of the frequency shift with that expected for liquid trapped

in surface cavities whose dimensions were measured by scanning electron microscopy (SEM). These studies indicated that the frequency changes were smaller than expected based on the SEM measurements, suggesting that the trapped liquid did not behave as a rigid mass. Therefore, surface roughness effects can be very difficult to quantify even when the exact roughness is known. These effects may be significant in many published EQCM studies because of the wide use of unpolished quartz crystals, which are commonly used because of their ready availability and low cost. These are widely sold for vacuum thickness monitor applications where good adhesion of metal films is required, but liquid trapping is not a factor. Our laboratories have always used polished quartz crystals in order to minimize roughness effects. In any case, any published work employing the EQCM should contain a description of the quartz crystals and their roughness.

3 ELECTROCHEMICAL METHODS AND EXAMPLES OF APPLICATIONS

The EQCM has been used to examine a wide variety of electrochemical processes. Unfortunately, space does not permit here a comprehensive review of the area and the reader is referred to other reviews cited earlier. Rather, this section will highlight some illustrative examples of EQCM applications which are meant to provide the reader with a general understanding of the scope of this method with respect to its capability, phenomena that affect measurements, and the types of system that are amenable to examination.

3.1 Experimental Apparatus and Operation

Several versions of EQCM instrumentation have been described,[14–17] differing mostly in minor details. The system described below is one that has been developed in our laboratory (Figure 3). Commercial systems are now available (EG&G PAR, Elchema), but it has been our experience that the practitioner gains a better awareness of the strengths and limitations of the method by building the homemade system. EQCM quartz crystals commonly have diameters of 0.5 in and 1.0 in, with appropriately sized excitation electrodes. The Bechmann numbers[18–20] dictate that, for a crystal diameter that is 50 times the crystal thickness, the electrode diameter must be 18 times the crystal thickness in order to avoid interference from other acoustic modes. Accordingly, a reduction in crystal and electrode diameter must be accompanied by a corresponding reduction in the crystal thickness to maintain frequency stability.

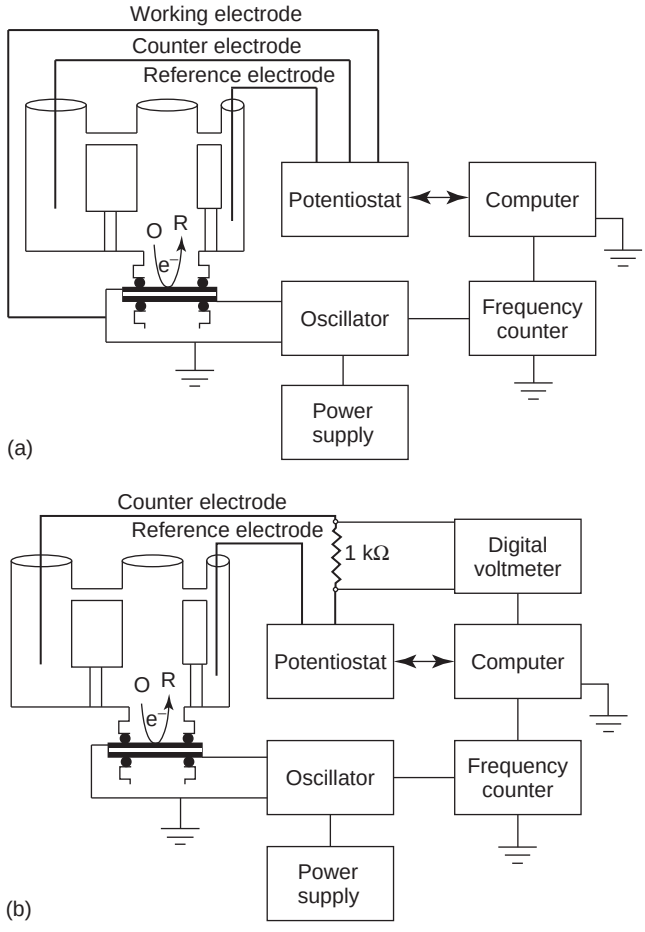

Figure 3 (a) Schematic representation of typical EQCM apparatus in which a Wenking potentiostat is employed when the working electrode is at hard ground. (b) An alternative EQCM in which a conventional commercially available potentiostat is used. In this arrangement, the working electrode lead of the potentiostat is not connected to the EQCM working electrode. Rather, the current is measured by the voltage drop across a 1 kW resistor in series with the counter electrode. This arrangement is required when using commercially available potentiostats because these potentiostats operate with the working electrode at virtual ground; connection of this virtual-ground to the hard-grounded EQCM working electrode can result in oscillator instabilities.

The crystals can be mounted at the bottom of a glass cylinder that assumes the role of the working electrode compartment of an otherwise conventional electrochemical cell. The crystal is mounted so that the excitation electrode that is to be employed as the working electrode is facing the solution; the opposite electrode is therefore facing air in order to avoid electrical shorting between the electrodes. If the quartz crystal is immersed under a column of the electrolyte solution, hydrostatic pressure results in a stress on the quartz crystal that affects the resonant frequency. Equation (7) describes a parabolic dependence of f_0 on the hydrostatic

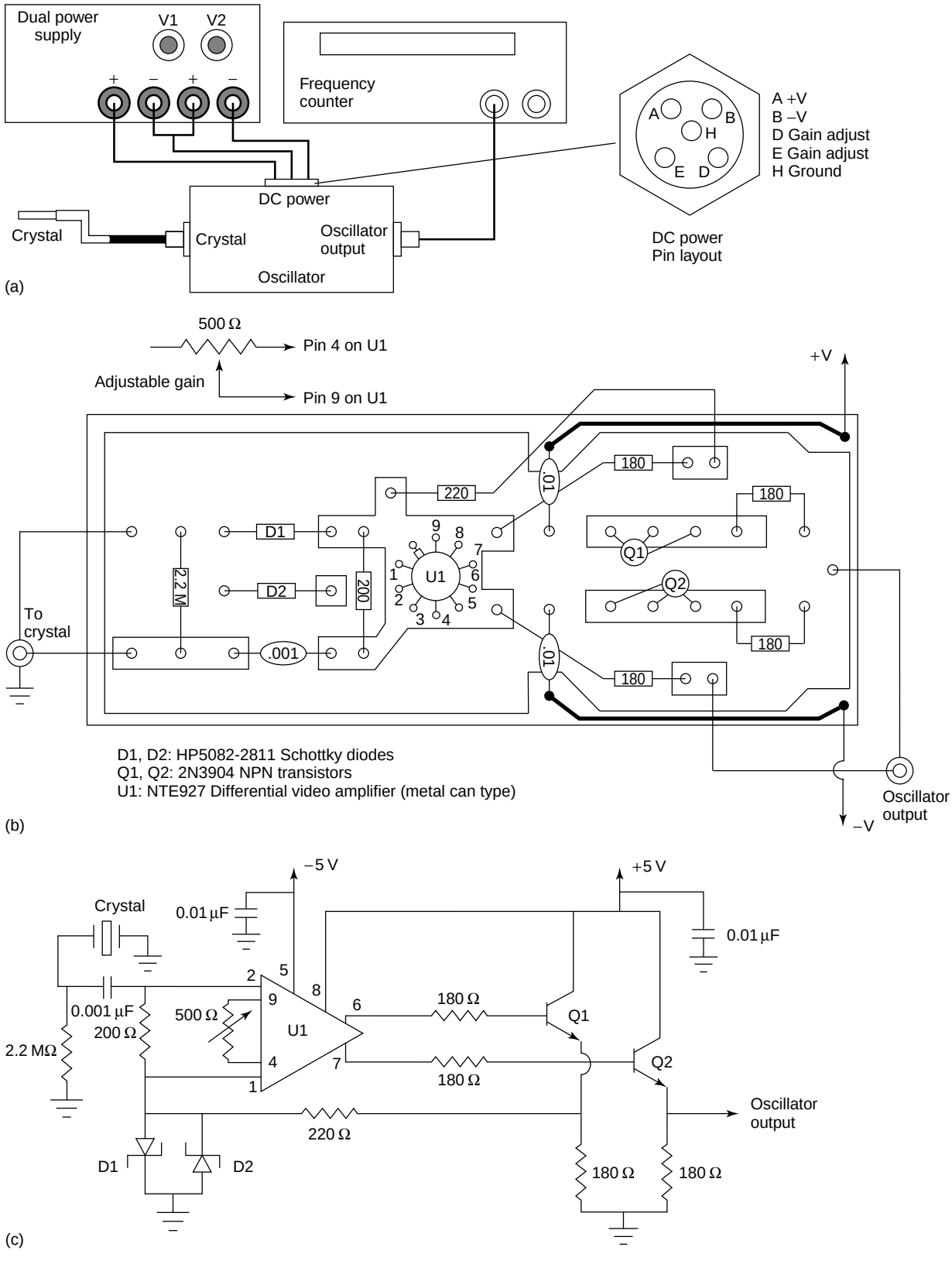

Figure 4 (a) Schematic representation of the hardware configuration for the EQCM, neglecting the electrochemical apparatus depicted in Figure 3. (b) Schematic of an oscillator circuit and related components. (c) Detailed schematic of the component layout for the oscillator circuit. The components are listed in Table 1.

pressure.[21,22]

$$f_0 - f_0^{\max} = A(p - p_{\max})^2 \quad (7)$$

However, the significance of this effect is probably minimal as the hydrostatic pressure is generally constant (barring evaporation) for most experiments. This problem can be avoided completely by configurations in which the crystal is simply mounted vertically. Crystals can be mounted between O-rings or with epoxy, the former being more convenient as the crystals can be easily demounted for reuse or further surface studies. Our laboratory has found that Teflon™-coated O-rings give the best performance.

The two excitation electrodes are electrically connected to an oscillator circuit that contains a broadband RF amplifier. Several oscillator designs are available, although standard commercial oscillators, such as those sold for metal evaporation control, typically need modification in order to supply the crystal with sufficient gain to sustain oscillation in liquids. Therefore, it is usually simpler to build the oscillator. The schematic for the oscillator circuit used in our laboratory is illustrated in Figure 4. For the convenience of the reader, a parts manifest is provided in Table 2. A commercially fabricated printed circuit board is also recommended in order to avoid noise pickup that accompanies circuit built on crude breadboards. The circuit is designed so that the crystal is in a feedback loop, therefore driving the crystal at a frequency at which the maximum current can be sustained in a zero-phase angle condition. The electrode facing solution is at hard ground. The output of the oscillator is connected to a conventional frequency meter for measurement. A critical feature of the EQCM is the type of potentiostat: the Wenking potentiostat functions with the working electrode at hard ground whereas most commercially available potentiostats generally function with the working electrode at virtual ground. Commercial potentiostats can only be used if the working electrode is not connected to the potentiostat and the potential difference between the reference and hard ground is used to control the working-electrode potential. Because the current is generally measured at the working-electrode side, this format requires that the current be measured by the voltage drop across a resistor in series with the counter-electrode connection. Because this equipment generally is available in most electrochemical laboratories, the EQCM practitioner need not build, or buy, a custom-made Wenking potentiostat. Finally, a computer is used to collect frequency and electrochemical data simultaneously, as well as control the waveform applied to the working electrode. This arrangement allows simultaneous measurement of the electrochemical charge, current, voltage, and EQCM frequency. The timescale

Table 2 Hardware and circuit components for a user-constructed EQCM

Hardware	Quantity	Vendor[a]	Part no.
Dual output power supply	1	1	6234A
Frequency counter	1	1(2)	5384A ()
IEEE interface board (data acquisition)	1	3	Any compatible version
Cable (oscillator to crystal)	1	3	RG 58 A/U (with two BNC connectors)
Cable (oscillator to power supply)	1	3	RG 174/U coaxial (with BNC connector)
Shielded IEEE cable	1	3	
Type 3302 component box	1	4	35F3515
H-90 trimmer tool (optional)	1	4	12F8866
5-pin jack	1	5	126-218
Printed circuit board	1	Custom	
Circuit components			
Cemet trimmer 500 Ω	1	4	12F9636
Panel adapter type 6-2-0	1	4	81F009
Differential video amplifier	1	6	NTE927
ULP-IC socket	1	4	65F1881
RN-55D type resistor, 182 Ω, 1/8 W	4	4	58F001
RN-55D type resistor, 200 Ω, 1/8 W	1	4	58F001
RN-55D type resistor, 221 Ω, 1/8 W	1	4	58F001
YX series polyester film capacitor, 0.001 μF	1	4	89F3228
YX series polyester film capacitors, 0.01 μF	2	4	89F3232
RC07 type CB resistor, 2.2 MΩ, 1/4 W		4	10F305
Schottky diodes	2	6	HP5082-2811
NPN transistors	2	3	2N3904

[a] Suggested vendors: (1) Hewlett-Packard, (2) Phillips Electronics, (3) any vendor, (4) Newark Electronics, (5) Wire-Pro, Inc., (6) Allied Electronics.

of the analysis is in a particularly useful domain for electrochemists. The time constant of a quartz resonator is fixed by $Q/\pi f_0$. The quality factor for a 5-MHz resonator is $Q \approx 10^3$, and therefore the minimum sampling time is in the millisecond range. Frequency counters are capable of sampling the frequency output of the oscillator at 100 ms intervals. This capability enables analysis of the kinetics of a wide range of electrochemical processes, including electrodeposition and dissolution, nucleation and growth, and ion/solvent insertion in redox polymer films. It is important to stress that this equipment is not prohibitively expensive and is well within the reach, economically and technically, of most electrochemical investigators.

3.2 Electrochemical Measurements of Crystal Resonant Properties in Fluids

There are other aspects of the EQCM that distinguish their operation in liquids from that of the QCM in vacuum or the gas phase, namely the effect of the liquid on the actual vibrating area of the quartz crystal. The Sauerbrey equation, Equation (2), assumes that the frequency shift resulting from a localized deposit is equivalent to the contribution of that deposit when it is a portion of a thin film of identical thickness distributed over the entire active EQCM area. However, a general expression that accounts for localized or nonuniform mass deposits covering the EQCM electrode to $r = r_e$ (where r_e is the ideal radius of the excitation electrodes, between which the electric field induces crystal motion) is given by Equation (8),

$$\Delta f = \left(\frac{1}{\pi r_e^2}\right) \int_0^{2\pi} \int_0^{r_e} S(r,\theta)\, m(r,\theta)\, r\, dr\, d\theta \qquad (8)$$

where $S(r,\theta)$ is the differential mass sensitivity (df/dm) and $m(r,\theta)$ is the mass distribution with respect to r, the distance of the deposit from the center of the crystal and θ, the angle in the crystal plane with respect to the x-axis. It has long been accepted, on the basis of theory and experiment, that $S(r,\theta)$ has a near-Gaussian form in which the crystal vibration is maximum in the center and negligible at the edges.[23–27] Indeed, Equation (8) and the Sauerbrey equation are based on the assumption that the vibrating area, and therefore the mass sensitive area, is limited to within the region where the excitation electrodes overlap. Any changes in $S(r,\theta)$ resulting from the liquid, or any crystal motion beyond the electrode edges, will cause deviations from the ideal response.

It has been shown in liquids that $S(r,\theta)$ and the actual vibrational area depend significantly upon the liquid and its properties, including its viscosity.[28–31] EQCM measurements performed in conjunction with SEM demonstrated that $S(r,\theta)$ was gaussian-like (Figure 5).

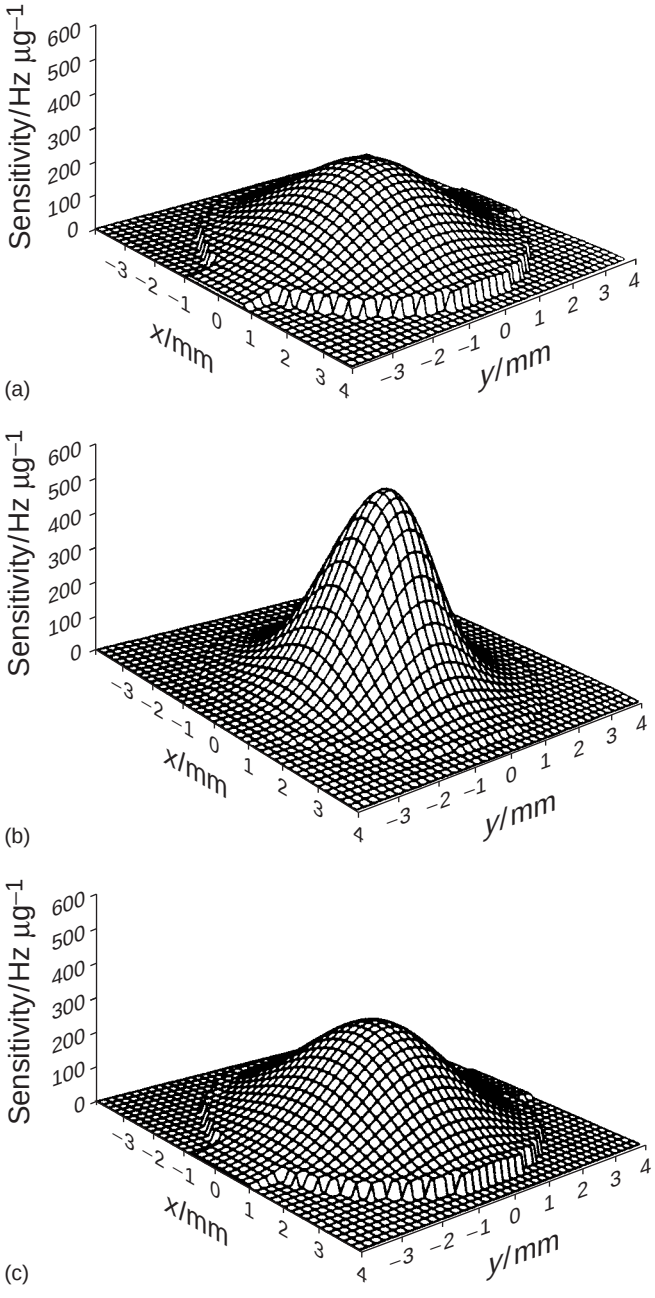

Figure 5 Sensitivity distributions $S(r,\theta)$ for a 5-MHz AT-cut quartz crystal determined by a SEM method in which 100-μm diameter circular copper features are electrodeposited at different values of r,θ while simultaneously measuring electrochemical charge and frequency. The amplitude of these plots at a given value of r,θ represents the sensitivity of the EQCM to mass changes at that location, and the area under the plot represents the total, or integral, sensitivity C_f. The $S(r,\theta)$ plots shown were measured on (a) plano–plano and (b) plano–convex quartz crystals in aqueous 20-mM $CuSO_4$ ($\eta_L = 0.8904 \times 10^{-2}\,g\,cm^{-1}\,s^{-1}$), and (c) plano–convex quartz crystals in aqueous 20-mM $CuSO_4$ containing 15% sucrose ($\eta_L = 1.469 \times 10^{-2}\,g\,cm^{-1}\,s^{-1}$). (Reproduced with permission from Hillier and Ward,[28] Copyright (1992) American Chemical Society.)

However, the shape and the maximum value in the center were dramatically affected by the viscosity of the liquid in contact with the EQCM. Increasing the viscosity resulted in a suppression of the maximum and an increase in the mass-sensitive area due to a phenomenon referred to as field fringing, in which the crystal vibrations extend beyond the electrode area. This results in an overall reduction in the sensitivity of the EQCM. The details of these effects will depend upon crystal contour, electrode geometry, and the liquid properties. Trapping of the crystal vibrations within the excitation electrode boundaries is much more efficient for plano–convex crystals, in which one side of the crystal is contoured by polishing such that its shape resembles a convex lens. The difference in mass between the center and outer regions of the crystal results in a focusing of the acoustic energy toward the center of the resonator, as is evident from an increase in the value of $S(r, \theta)$ at the center and negligible mass sensitivity at $r > r_e$. Therefore, there appear to be distinct advantages to using plano–convex crystals. Understanding of these effects is crucial if precise quantitative interpretation of EQCM data is required. At the very least, these effects should be appreciated in a qualitative way.

3.3 Data Interpretation for Ideally Behaved Systems

Interpretation of EQCM data is accomplished in a rather straightforward manner. Since the electrochemical charge represents the total number of electrons transferred in a given electrochemical process, it corresponds to mass changes occurring at the electrode surface. Accordingly, under ideal conditions, the frequency change measured with the EQCM will be proportional to the electrochemical charge, and will be related to the apparent molar mass by Equation (9), where MW is the apparent molar mass ($g\,mol^{-1}$), Q is the electrochemical charge, n is the number of electrons involved in the electrochemical process, F is the Faraday constant, and C_f ($Hz\,g^{-1}$) the sensitivity constant derived from the Sauerbrey relationship.

$$\Delta f = \frac{MWC_fQ}{nF} \tag{9}$$

Inspection of Equation (9) reveals that plots of Δf versus Q are particularly useful in the determination of MW/n, which represents the molar mass per electron transferred. This calculation, of course, depends upon knowledge of C_f. Because this value can differ for different crystal contours, electrode geometries, and solution conditions (see above), it is important to calibrate the EQCM with a well-behaved electrochemical reaction under conditions similar to those present during the experiments of interest. This is generally accomplished with copper or silver electrodeposition, for which all terms on the right side of Equation (9) are known except for C_f. The term apparent molar mass is stressed because in many cases the measured value of MW may not be that expected based on a simple stoichiometric relationship, but may involve solvent or co-adsorbed species that can reveal considerable insight into the electrochemical behavior. Nonlinearity in plots of Δf versus Q can be particularly useful for diagnosing nonideal behavior such as roughness and viscoelastic effects that may become evident over the range of frequency changes examined. An alternative approach to data analysis involves the relationship between the electrochemical current and the first derivative of the frequency change with respect to time, as given in Equation (10), where v is the scan rate in units $V\,s^{-1}$. This format is particularly useful for cyclic voltammetry experiments, as $d\Delta f/dt$ should appear similar in form to the voltammograms if the electrochemical events are accompanied by corresponding mass changes.

$$i = \frac{d(\Delta f)}{dE}\frac{nvF}{MWC_f} \tag{10}$$

The utility of the EQCM method stems from its capability of measuring electrochemical charge and current while simultaneously measuring mass changes with extraordinary sensitivity. A typical operating resonant frequency of 5 MHz provides a theoretical sensitivity of $0.5666\,Hz\,cm^{-2}\,ng^{-1}$. Since the frequency can generally be measured to within an accuracy of 1 Hz, the EQCM can detect approximately $10\,ng\,cm^{-2}$. This translates roughly into 10% of a monolayer of Pb atoms. Much higher sensitivity can be realized with quartz crystals that operate at a higher fundamental frequency, as the sensitivity increases with f_0^2. Alternatively, the quartz crystal can be driven at one of its odd harmonic modes with appropriate circuitry, which provides an n-fold increase in sensitivity, where n is the harmonic number. While the third harmonic of a 5-MHz crystal (i.e. an operating frequency of 15 MHz) has been employed successfully in the examination of underpotential deposition (UPD) on metal electrodes,[32] general use of this approach can be limited by the lower stability of the harmonic modes relative to the fundamental mode.

It is much more beneficial to employ crystals that can be operated at higher fundamental frequencies. However, higher fundamental frequencies can only be achieved by fabricating thinner crystals. The operating frequency range of the EQCM has been extended as far as 30 MHz. While these crystals are rather thin ($t_Q = 50\,\mu m$), they could be used conveniently by constructing the electrochemical cell with heat-shrinkable Teflon™ tubing constricted about the periphery of the crystal. In liquid media the frequency of a 30-MHz resonator responds to changes in the liquid viscosity and density according

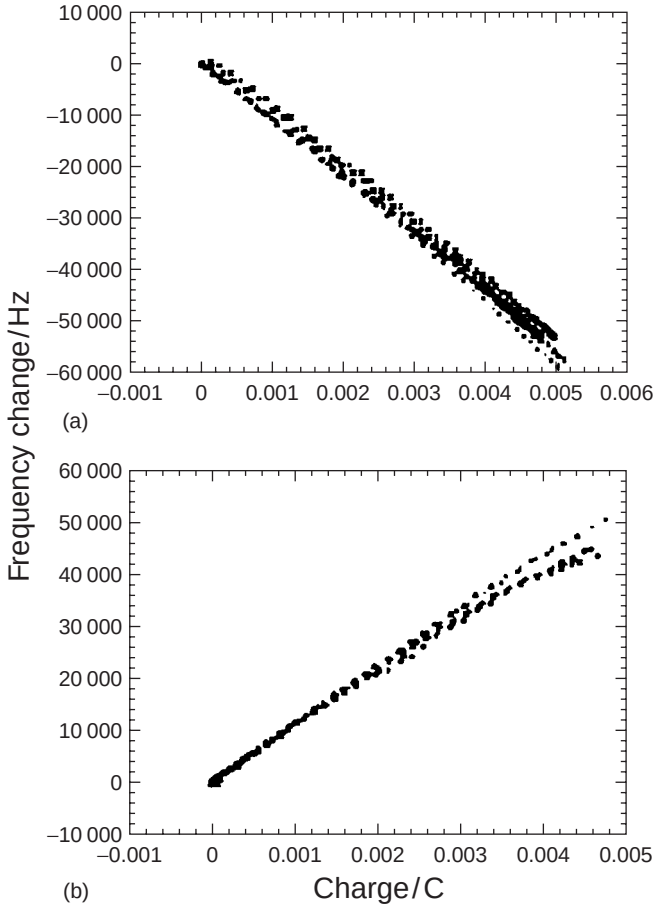

Figure 6 Frequency dependence of the charge consumed at the electrode of a 30-MHz EQCM during electrodeposition and removal of copper in 0.1-M $CuSO_4$ solution. (a) Electrodeposition of copper, slope $= 1.135 \pm 0.036 \times 10^7\,\mathrm{Hz\,C^{-1}}$. (b) Electrochemical removal of copper, slope $= 1.064 \pm 0.043 \times 10^7\,\mathrm{Hz\,C^{-1}}$. (Reproduced with permission from Lin et al.,[6] Copyright (1993) American Chemical Society.)

to Equation (6). That is, these resonators exhibit a linear relationship between Δf and $(\eta_L \rho_L)^{1/2}$, identical to behavior expected and observed for lower frequency resonators. Electrodeposition of copper on the EQCM surface (Figure 6) affords sensitivity constants in exact agreement with Equation (9), indicating that energy trapping of the fundamental mode is very efficient for the 30-MHz resonators (lower frequency resonators commonly exhibit sensitivity constants that are lower than expected owing to acoustic field fringing). The higher frequency resonators have many potential advantages, including greater sensitivity to mass, better signal-to-noise characteristics than other acoustic wave devices,[33] and less energy dissipation during oscillation. The origin of the improved stability is evident from admittance analysis of 30-MHz resonators (see below). Higher frequency EQCMs are feasible if chemically milled crystals are employed. Such a crystal is obtained by chemically etching the center of a quartz crystal to obtain a thin quartz "membrane" in the center of a thicker outer quartz ring, the outer ring providing improved mechanical stability.

Changes in the depletion layer that accompany redox processes also can influence EQCM measurements. Our laboratory demonstrated that cyclic voltammetry in solutions containing the $Fe(CN)_6^{3-/4-}$ or $Ru(NH_3)_6^{2+/3+}$ redox couples was accompanied by potential-dependent changes in the resonator frequency even though no mass changes occurred at the electrode surface (Figure 7).[34] This behavior can be attributed to changes in the density of the depletion layer that accompany the change in redox state. Modeling of the depletion-layer characteristics using known densities and viscosities of each redox species

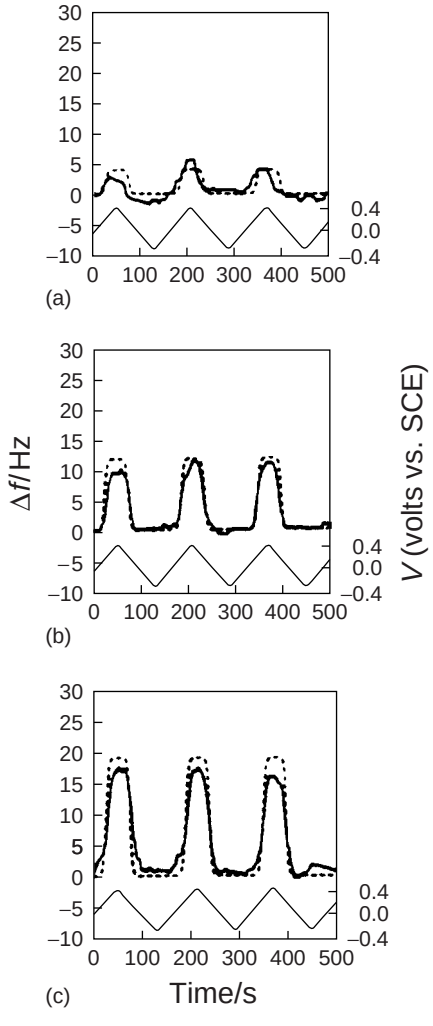

Figure 7 Comparison of measured (——) and simulated (- - - - -) frequency changes at a 5-MHz EQCM during potential cycling between −0.3 V and 0.5 V at $10\,\mathrm{mV\,s^{-1}}$ for solutions of $K_4Fe(CN)_6$ at concentrations of (a) 20, (b) 60, and (c) 100 mM in 1.0 M Na_2SO_4. (Reproduced with permission from Lee et al.,[34] Copyright (1993) American Chemical Society.)

have confirmed that the frequency changes during cyclic voltammetry behaved according to Equation (6).

3.4 Electrochemical Quartz Crystal Microbalance Investigations of Metal Electrodes and Films

As stated above, the most common technique for calibrating the EQCM involves simultaneous measurement of the electrochemical charge and resonant frequency during the electrochemical deposition or dissolution of a metal film. This capability was demonstrated by Bruckenstein and Shay in which the deposition of 10 layers of Ag metal on a 10-MHz EQCM gave frequency shifts that were within 3% of the amount expected based on the Sauerbrey equation.[35] Therefore, the EQCM can be an effective method for examining the Faradaic efficiencies of electroplating and dissolution processes such as corrosion.[36]

The high sensitivity of the EQCM has enabled examination of UPD processes in situ. For example, examination of the UPD of Pb on the Au working electrode of the EQCM revealed mass and electrochemical charge changes in the UPD region corresponding to a hexagonally closest packed monolayer. The ability to measure

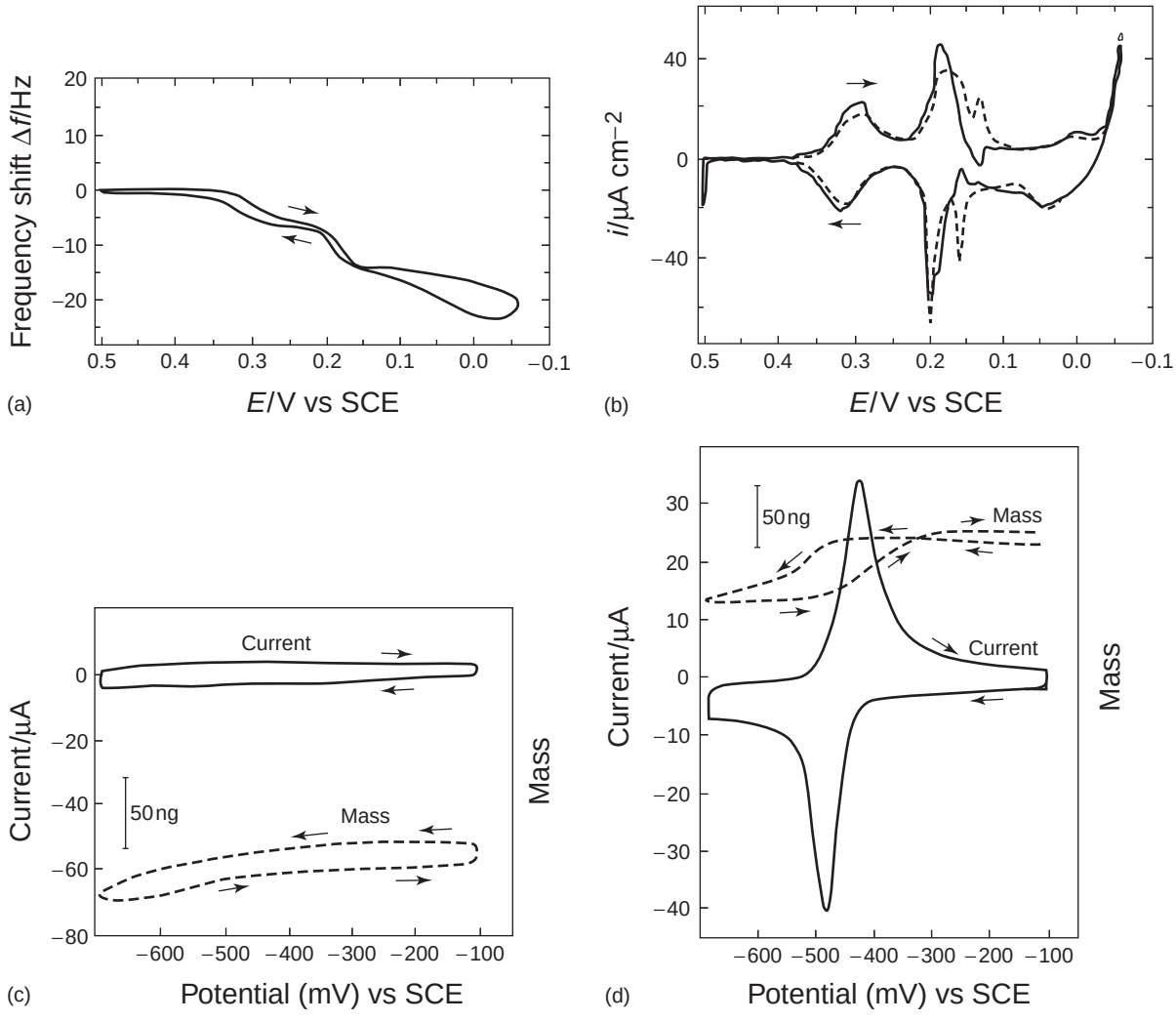

Figure 8 (a) Frequency shift of the EQCM during Bi UPD on a Au electrode (1.0-mM Bi in 0.1-M $HClO_4$). (b) Current response during Bi UPD on the Au EQCM electrode (- - - - -) and current response expected from the observed frequency response (———). (Reproduced from Deakin and Melroy, *J. Electroanal. Chem.*[37] Copyright (1998) with permission from Elsevier Science.) (c) Cyclic voltammetric scan showing the current and EQCM response obtained at an EQCM silver electrode in 0.1-M borate buffer at pH = 9.15; scan rate = 50 mV s^{-1}. The electrode was conditioned at −100 mV (vs SCE (saturated calomel electrode)) for 30 s before initiating the potential scan. (d) Cyclic voltammetric scan showing the current and EQCM response obtained at an EQCM silver electrode in 0.1-M borate buffer at pH = 9.15 containing 2.7×10^{-5} M Pb(II); scan rate = 50 mV s^{-1}; starting potential = −100 mV. (Reproduced from Hepel et al., *J. Electroanal. Chem.*[39] Copyright (1989) with permission from Elsevier Science.)

mass and charge simultaneously allowed determination of the electrovalency number for this process, which was found to be $\gamma = 2.08 \pm 0.10$. Similar experiments revealed that other UPD processes occurred with γ values lower than expected for complete electron transfer, with $\gamma_{Bi} = 2.7, \gamma_{Cu} = 1.4$.[37]

The EQCM has revealed several interesting features in UPD processes that otherwise would be difficult to detect. For example, in the aforementioned investigations with bismuth, it was discovered that a precipitous increase in electrochemical charge accompanied the third UPD peak, with a slight decrease in mass (Figure 8). This behavior may be associated with the loss of electrolyte ions that are adsorbed during the first two UPD events. More detailed studies of Pb UPD on Au and Ag revealed behavior that demonstrated the importance of the underlying substrate in UPD processes.[38–40] Whereas Pd UPD on Au electrodes in borate buffer was accompanied by a mass increase, UPD on Ag electrodes exhibited a mass decrease due to the desorption of BO_2^- ligands from a previously adsorbed anionic Pb(II) species. The adsorption of the latter was detected by the frequency decrease upon addition of Pb(II) to the electrolyte solution prior to electrochemical experiments. The data reflected the desorption of three BO_2^- ions per adsorbed Pb(II) species during the UPD process. These studies clearly reveal the value of the EQCM in probing electrode surface processes, providing information critical to complete understanding of rather complicated mechanistic schemes.

The EQCM has also enabled determination of the electrovalency numbers for anion adsorption. Simultaneous measurement of electrochemical charge and frequency during anodic adsorption of Br^- and I^- under conditions where monolayer coverage previously had been demonstrated revealed $\gamma_{Br^-} = -0.39 \pm 0.03$ and $\gamma_{I^-} = -1.01 \pm 0.03$.[41] These studies indicate that, while adsorption of iodide occurs with complete charge transfer, bromide ions partially retain their negative charge.

The capabilities of the EQCM provide a unique approach for examining electrochemical dissolution, which is important in processes such as corrosion and electrochemical machining. While EQCM studies of dissolution have been rather limited, these reports indicate that important mechanistic information can be obtained readily with this method. For example, EQCM investigations of the anodic dissolution of nickel films revealed two maxima in the Δf versus potential plots, indicating a potential-dependent dissolution of the α and β phases of NiH_x.[42] Analysis of the frequency changes and electrochemical charge in EQCM studies of the anodic dissolution of a nickel–phosphorus film revealed that two different Ni–P compositions were present in the film prior to dissolution. This was consistent with the known Ni–P phase diagram.[43] EQCM studies of the electrochemical dissolution of copper films in oxygenated sulfuric acid revealed that the dissolution rate was linearly dependent upon $[O_2]$ and $[H^+]$, enabling the authors to conclude that a heterogeneous surface reaction was operative.[44]

Several groups have demonstrated that the EQCM is also an ideal method for examining the electrochemically induced adsorption of hydrogen and deuterium in metal films such as palladium.[45–47] Such studies can have significant impact on understanding the commercially important isotopic separation of H and D. In addition

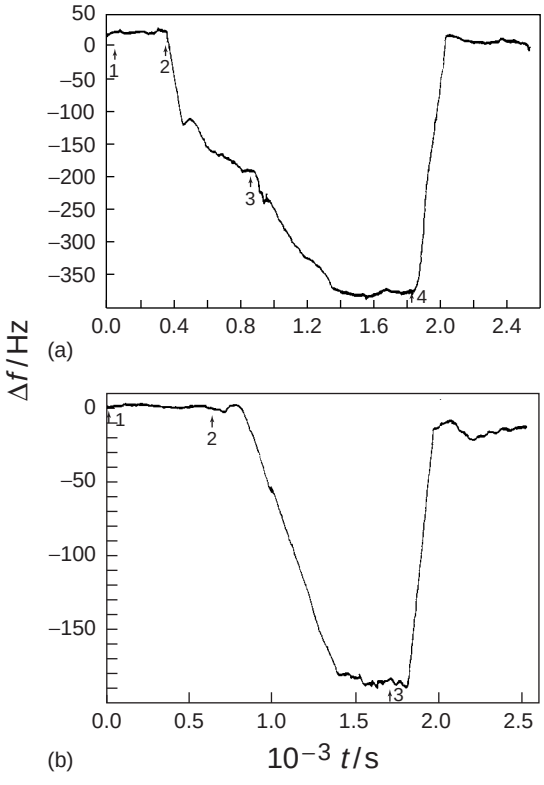

Figure 9 (a) Frequency shift versus time for a 366.6-nm thick Pd film on an AT-cut quartz resonator in 0.1-M $LiOH/H_2O$. The numbered arrows indicate the potential applied to the film: (1) 0.00 V; (2) E scanned to and held at −1.14 V; (3) E scanned to and held at −1.27 V; (4) E scanned to and held at 0.0 V. The frequency decrease is larger than that expected based on the mass of the cathodic hydrogen absorption process, and is attributed to compressive stress in the Pd film upon hydrogen absorption which decreases the frequency. (b) Frequency shift versus time for a 475-nm thick Pd film on an AT-cut quartz resonator in 0.1-M $LiOH/D_2O$. The numbered arrows indicate the potential applied to the film: (1) 0.00 V; (2) E scanned to and held at −1.20 V; (3) E scanned to and held at 0.0 V. The frequency decrease is smaller than that expected based on the mass of the cathodic deuterium absorption process, and is attributed to compressive stress in the Pd film upon deuterium absorption that increases the frequency. (Reproduced from Cheek and O'Grady, *J. Electroanal. Chem.*[45] Copyright (1990) with permission from Elsevier Science.)

to measurement of H/D adsorption, these EQCM studies have revealed the important role of stress in EQCM measurements. Cheek and O'Grady reported a rather novel approach to this issue by using a "double resonator" technique that previously had been reported for measuring stresses in Si films resulting from ion implantation (in vacuum).[48,49] This method involves the comparison of frequency responses from palladium-coated AT-cut and BT-cut quartz crystals during electrochemically induced ingress and egress of H or D (Figure 9). Both crystals have identical sensitivities to mass changes, but a compressive stress in a film on an AT-cut crystal results in a frequency decrease while an identical stress in a film on a BT-cut crystal results in an increase in frequency of similar magnitude. The amount of stress can be determined by Equation (11),

$$S = (K^{AT} - K^{BT}) \left(\frac{t_Q^{AT} \Delta f^{AT}}{f_0^{AT}} - \frac{t_Q^{BT} \Delta f^{BT}}{f_0^{BT}} \right) \qquad (11)$$

where K^{AT} and K^{BT} are the stress coefficients for the different resonators ($K^{AT} = 2.75 \times 10^{-12}\,cm^2\,dyn^{-1}$ and $K^{BT} = -2.65 \times 10^{-12}\,cm^2\,dyn^{-1}$). Stress in thin films is manifested in a stress in the quartz crystal, which is tantamount to a change in the elastic modulus, μ_q. Because the acoustic velocity depends upon this quantity, these stresses result in changes in the resonant frequency (this effect is similar to tightening a violin string to increase the frequency of vibration). Thus, hydrogen adsorption in a Pd film on an AT-cut crystal gave a frequency decrease that was much larger than expected, indicating compressive stresses in the Pd film upon H adsorption, −351 Hz for a 367-nm thick Pd film, compared to −185 Hz expected for the mass change due to H adsorption, However, the same experiment on a BT-cut crystal resulted in a very small increase in frequency, +20 Hz for a 475 nm thick Pd film.

3.5 Electrochemical Quartz Crystal Microbalance Investigations of Thin-film Growth

The EQCM provides an extremely useful approach to the in situ study of the nucleation and growth of a wide variety of thin films on electrodes, ranging from oxide films to molecular solids and monolayers. In principle, the ability to measure mass and electrochemical current simultaneously allows determination of chemical stoichiometries, Faradaic efficiencies, and reaction kinetics. The EQCM therefore complements current transient methods typically used to study nucleation and growth, providing information which otherwise cannot be obtained readily. The time constant for EQCM measurements (see above) is within the timescale of many electrochemical processes.

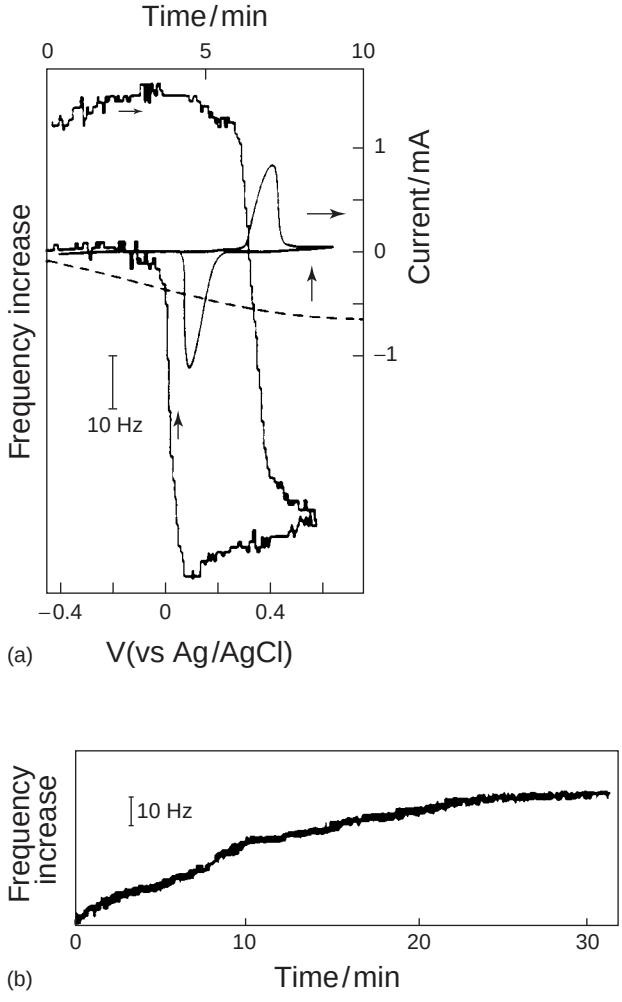

Figure 10 (a) Current versus potential (- - - - -) and frequency versus potential (jagged line) for a silver EQCM electrode during cyclic voltammetry from an initial potential of −0.4 V (vs Ag/AgCl). Scan rate = 20 mV s^{-1}, surface area = 0.33 cm^2. (———) Frequency versus time with the electrode held at −0.4 V at the end of the cycle. Note that the frequency at the end of the cycle is approximately 30 Hz less than at the beginning, and this difference increases to 45 Hz after 10 min at −0.4 V. This indicates a significant change in the morphology of the electrode. (b) Frequency versus time response for the silver electrode held at −0.4 V after the 10 min period depicted in (a). The frequency increase is attributed to a morphological relaxation of the silver electrode. (Reproduced from Schumacher et al.[12] Copyright (1987) with permission from Elsevier Science.)

The growth of metal-oxide films has been investigated at copper and gold electrodes, the EQCM enabling determination of the stoichiometry of the oxides formed.[11–13] In addition, these studies revealed frequency responses that were consistent with morphological changes of the electrode surface accompanying oxide formation. This resulted in much larger changes in the frequency during oxide growth that was attributed to surface roughening

and subsequent trapping of water in the cavities of the roughened surface. The surface roughness was retained initially after reduction of the oxide back to the metal, but the frequency gradually returned to its original value, indicating dynamic changes in the morphology of the roughened metal electrodes (Figure 10). In addition to providing valuable mechanistic insight into oxide growth, these studies illustrate the sensitivity of the EQCM to surface roughness effects.

EQCM studies of the electrochemical deposition of molecular films have also been studied. The adsorption of surfactants containing redox-active ferrocene (Fc) groups (referred to as C_{12} and C_{14}, which designates the length of the alkyl chain) could be induced electrochemically, because adsorption occurred when these species were in the reduced (Fc^0) state (Figure 11).[50,51] Upon oxidation to the oxidized form (Fc^+), the surfactants rapidly desorbed. Interestingly, the desorption rate was dependent upon the chain length of the surfactant. Whereas short-chain-length surfactants resulted in rapid desorption following electrochemical oxidation, the EQCM frequency changes indicated that long-chain surfactants desorbed much more slowly following oxidation. Mechanistic and thermodynamic information for these films was thereby attainable.[52] The nucleation, growth, and dissolution of thicker films of related species, namely diheptyl viologen, was also examined with the EQCM. Examination of these processes at different potentials provided insight into the nucleation behavior of this system.[53]

The EQCM has also been useful in the investigation of the nucleation and growth of ''electronic'' materials. For example, the deposition of thin semiconductor films was examined, and the rather complicated mechanism involved in the deposition of thin films of Te was determined by comparison of the EQCM frequency changes and electrochemical charge.[54] Similarly, the electrodeposition of B-doped β-PbO_2 thin films was studied with the EQCM and details of the film composition and

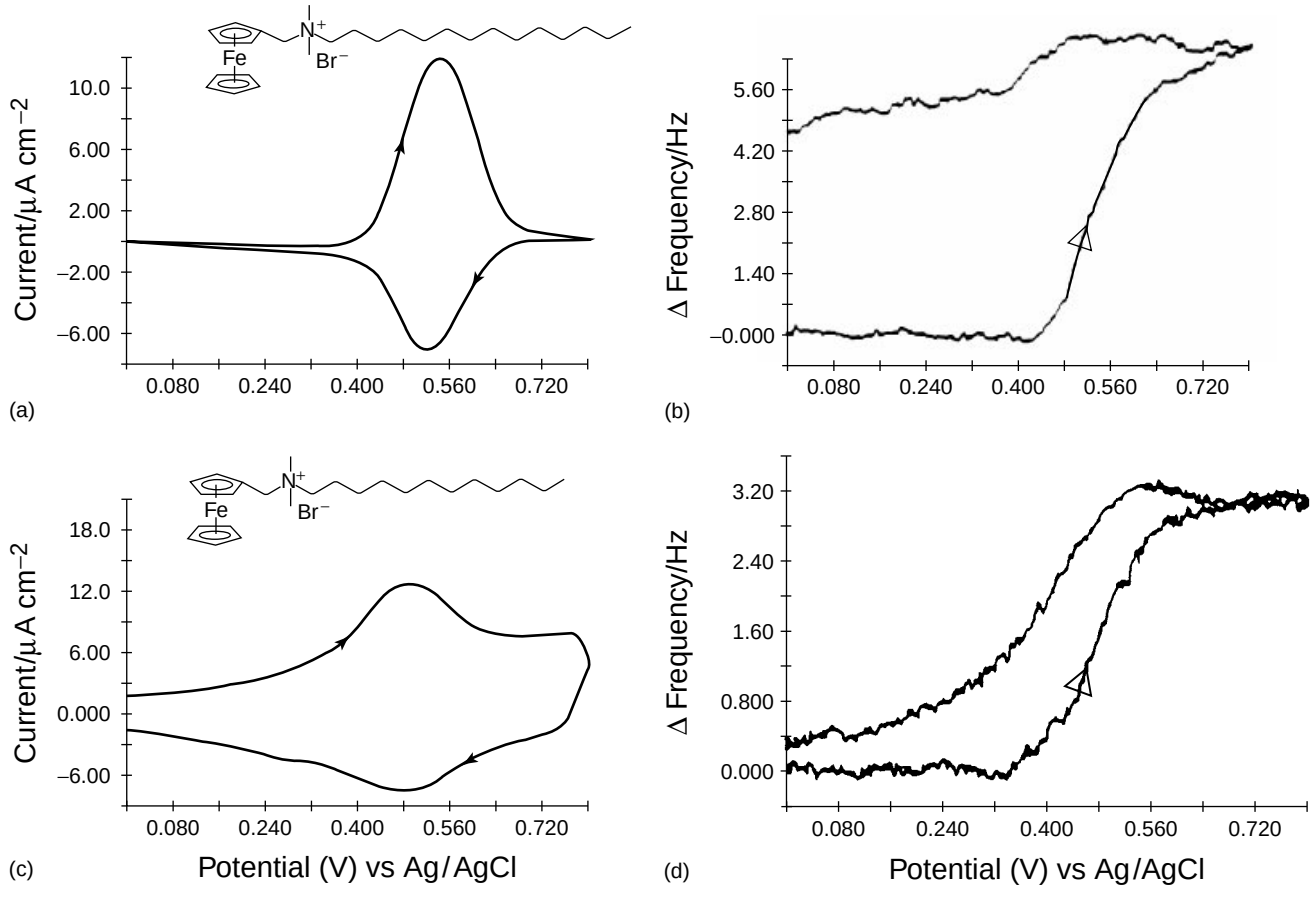

Figure 11 (a) Cyclic voltammetric scan for a 7-μM solution of C_{14} compound from 0.0 to 0.80 V (vs Ag/AgCl) in 1-M H_3PO_4; 50 mV s^{-1}. The current was digitally smoothed and corrected for background. (b) EQCM frequency response for (b). (c) Cyclic voltammetric scan for a 22-μM solution of C_{12} from 0.0 to 0.80 V in 1-M H_3PO_4; 50 mV s^{-1}. (d) EQCM frequency response for (c). Higher concentrations of the C_{12} compound were required to obtain the same coverage as for the C_{14}. Note that the frequency response in (d) is less than that in (b), consistent with the lower molecular weight of C_{12}. (Reproduced with permission from Donahue and Buttry,[50] Copyright (1989) American Chemical Society.)

catalytic activity toward oxygen atom transfer reactions were realized.[55]

The growth and redox chemistry of Prussian blue and related films has been examined by EQCM, with particular attention paid to the degree of solvation of these films and the ion transport during electrochemical cycling. These studies would be expected to have an impact on the use of these films in sensors and electrochromic displays. The initial EQCM study of Prussian blue films found that the frequency decrease during electrochemical deposition was consistent with a high degree of hydration.[56] Subsequent EQCM studies of these films revealed that cation transport accompanied changes in the redox state during potential cycling, although at low pH conditions proton transport was also involved. Ion and solvent transport in related nickel ferrocyanide films during potential cycling in 0.1-M CsCl solutions also has been examined. The results were consistent with transport of Cs^+ ions upon change in the redox state of the film.[57] The mass change calculated from the Sauerbrey equation, Equation (2), which assumes rigid-layer behavior, suggested that H_2O was expelled from the film as Cs^+ was incorporated during reduction. This was verified by a clever experiment in which the process was studied in H_2O and D_2O; the mass change associated with the solvent was found to increase by 10% in D_2O. This verified the participation of solvent and provided corroborative evidence of rigid-layer behavior.

The Prussian blue and nickel ferrocyanide films consist of a low density pseudocrystalline lattice in which the metal atoms are organized by cyanide ligand bridges between octahedral metal centers. As a result, metal ion intercalation into the interstices in the lattice is affected by the size and hydration of the cation. The different amount of work required for intercalation of the different ions is manifested as a dependence of the formal potential of the films on the identity and concentration of the metal ion in the supporting electrolyte. Analysis of EQCM data acquired during redox cycling of Prussian blue films in propylene carbonate containing different relative amounts of $NaClO_4$ and $LiClO_4$ revealed that intercalation of Na^+ was favored over Li^+ by a factor of 15.[58] This capability was demonstrated to be useful for the analytical identification of metal ions in a flow injection mode.[59] Simultaneous electrochemical and EQCM measurements were performed in which the Prussian blue film was held at a potential known to be selective for a given cation, and the charge and mass changes associated with ion transport were measured. Comparison of the charge and mass changes enabled determination of the molar mass and, therefore, identification of the cation.

The electrodeposition of polymer films has also been investigated with the EQCM, in particular the electrodeposition of poly(vinylferrocene) (PVF) films.[60] These studies indicated that in oxidative electrodeposition of PVF in CH_2Cl_2, initially more polymer was deposited than expected based on the deposition of one monomer unit per electron. It was suggested that during the initial stages partially oxidized polymer was deposited (i.e. large Faradaic efficiencies), possibly with solvent or electrolyte trapped in the polymer. However, the results indicated that the total frequency change was smaller than that expected for the amount of charge passed during the deposition. This apparent loss of mass sensitivity suggests, as with some of the examples above, that the deposited film is not ideally rigid.

3.6 Electrochemically Active Polymer Films

Without a doubt, the EQCM has found its most extensive use in the investigation of electrochemically active polymer films, including redox and conducting polymers. Indeed, to a large extent it was the first report of an EQCM study of polypyrrole films by Kaufman et al. that heightened interest in the EQCM method.[61] The benefits of the EQCM were immediately obvious, as it became clear that the identity of the counterions exchanging between the films and electrolyte could be determined. In addition, the amount of solvent accompanying ion exchange could be determined by calculating the mass in excess of that expected from the electrochemical charge. Since these initial studies, numerous polymer films have been investigated with the EQCM and many mechanistic and thermodynamic insights into their behavior have been realized. The practitioner should be cautious when using the EQCM with polymer films, as changes in the mechanical properties of the films are not uncommon when ion population and solvent swelling are involved. The corresponding changes in viscoelasticity can result in dramatic frequency changes which are unrelated to mass changes. In many cases, this warrants experimental verification that rigid-layer behavior is present over the thickness and composition ranges examined. This can be accomplished by performing experiments with different polymer thicknesses over a reasonable range and by using impedance analysis to elucidate the contribution from changes in mechanical properties. If these factors are taken into account, the EQCM can be an especially valuable tool in examining electrochemically active polymer films.

It is reasonable to claim that PVF is ubiquitous in the subdiscipline of polymer-modified electrodes. Similarly, it has been among the polymer films most extensively examined with the EQCM. One of the initial investigations involved measurement of the frequency changes associated with ion transport required for electroneutrality

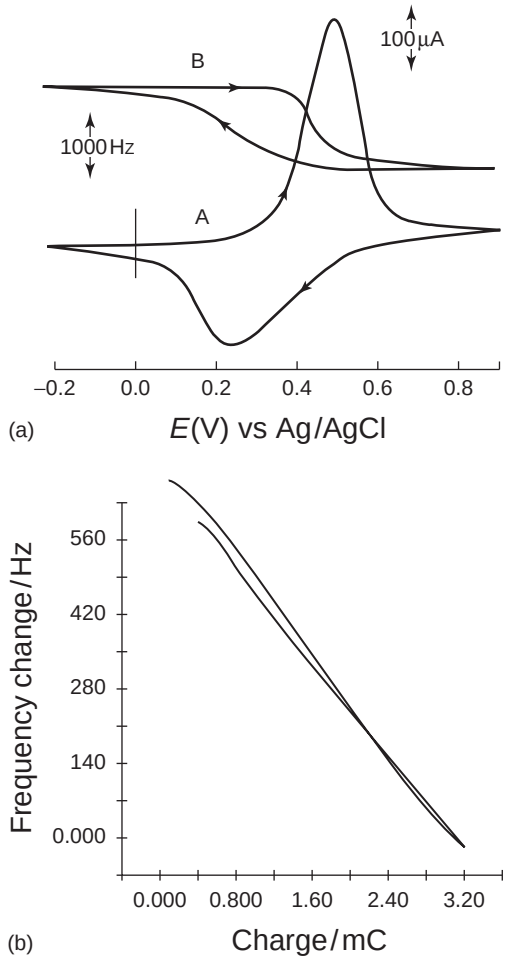

Figure 12 (a) (A) cyclic voltammogram of PVF on a gold EQCM electrode in 0.1-M KPF_6. Scan rate = 10 mV s^{-1}. (B) EQCM frequency response obtained simultaneously with (A). (b) Plot of frequency versus charge for a scan from 0.0 to 0.60 V and back for a PVF film in 0.1-M $NaClO_4$ + 0.1-M $HClO_4$. Scan rate = 25 mV s^{-1}. The linearity suggests ideal behavior and the slope gives the molar mass of the transport species. (Reproduced with permission from Varineau and Buttry,[62] Copyright (1987) American Chemical Society.)

and the amount of solvent accompanying ion transport in PVF films.[62] Based on a frequency decrease that accompanied oxidation of the PVF film, it was determined that ClO_4^- and PF_6^- were inserted without accompanying solvent, with one equivalent of anions inserted for each equivalent of electrons (Figure 12). The process was reversible, as evidenced by the frequency increase upon electrochemical reduction. Electrochemical oxidation of PVF films in the presence of other counterions, however, occurred with varying amounts of solvent incorporation, with the amount of incorporated water per ion decreasing in the order $Cl^- > IO_3^- \gg BrO_3^-(5) > ClO_3^-(1) \approx NO_3^-(1) > CH_3{-}C_6H_4{-}SO_3^-(0.5)$.[63] Indeed, cross-linking of the PVF film was necessary to prevent its dissolution upon oxidation. Unfortunately, determination of the amount of incorporated solvent is not feasible in the case of Cl^- or IO_3^- because of the dramatic changes in viscoelasticity that accompany the large degree of swelling when these ions are present. On the other hand, solvent incorporation for the other ions could be measured reliably (the amount of solvent incorporated per counterion are indicated above within the parentheses).

The EQCM has been employed to investigate transport in redox polymer films such as PVF and poly(thionine) of mobile neutral species such as water, particularly with regard to thermodynamic changes in solvent activity in the polymer film and transport kinetics.[64–66] These studies indicate that the number of solvent molecules per counterion need not be integral, and that it is feasible that neutral ion pairs are also involved in the transport. Comparison of EQCM frequency changes at different scan rates also revealed that mass changes and electrochemical charge changes do not always occur simultaneously.[67,68] That is, counterion motion to maintain electroneutrality during redox changes must always be established, but global equilibrium may lag behind. In particular, these studies revealed that electroneutrality would be achieved initially during redox by using counterions already present in the film, followed by transport of ions from the solution. Similar behavior was observed for redox-active poly(nitrostyrene) films.[69] The behavior has been attributed to potential gradients within the film that affect ion transport, and to the low dielectric constant of the films.[70–72] In the case of poly(thionine) experiments, comparison of current transients following a series of potential steps, suggested that transport of different species occurs on different timescales (Figure 13).[73] It was surmised that proton transport necessary to achieve electroneutrality was rapid, but that global equilibrium occurred more slowly by transport of the neutral ion pair $H_3O^+ClO_4^-$ and H_2O. These conclusions were supported by experiments performed in D_2O, which exhibited trends consistent with the difference in mass of H and D. These results have important consequences in the design and synthesis of polymer films in applications where charging and discharging rates are critical, for example in sensors or energy storage.

EQCM studies revealed that ion transport into PVF films upon oxidation was slower than their transport out of the film during reduction.[74] This is presumably because of the change in density accompanying the swelling of the oxidized film. It was estimated that in the first oxidation cycle of electrochemically deposited PVF films, water uptake began only after the film was approximately 40% oxidized and that 50% of the water was retained after subsequent reduction.[75] Water transport in following cycles was reversible, apparently

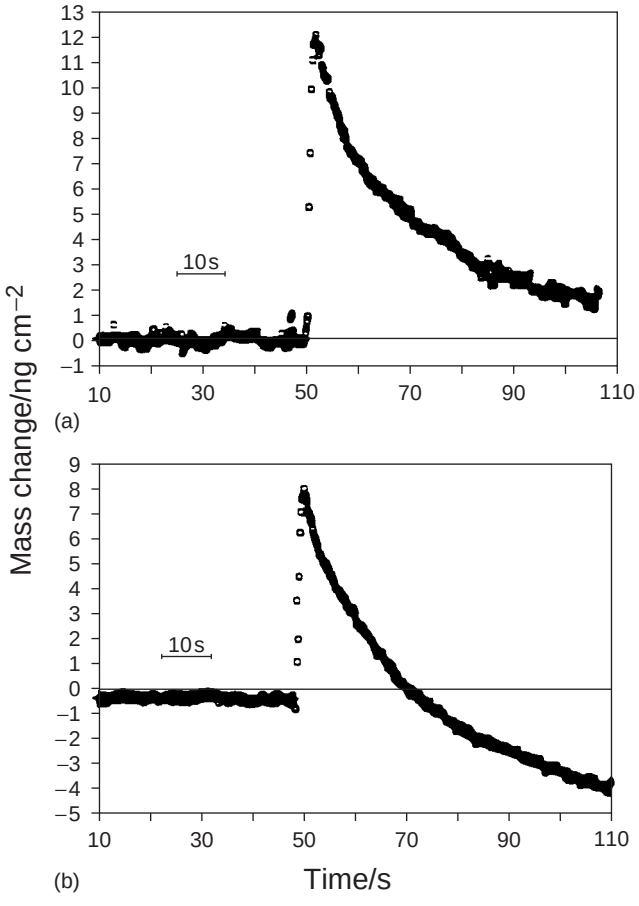

Figure 13 Transient mass changes following the application of a potential step from −0.1 V to +0.5 V (vs SCE) to a poly(thionine) film in $HClO_4$ solution. (a) Solvent in H_2O, pH = 1.6; (b) Solvent is D_2O, pD = 2.1. The potential step converts a fully reduced film to a fully oxidized film at long times. The pH and pD values were adjusted to achieve near zero net mass change at long times after switching. Mass increases upwards. The same polymer film was used in H_2O and D_2O. Note that the initial mass change is larger than that after global equilibrium is achieved at longer times. (Reproduced by permission of the Electrochemical Society, Inc. from Bruckenstein et al.[73])

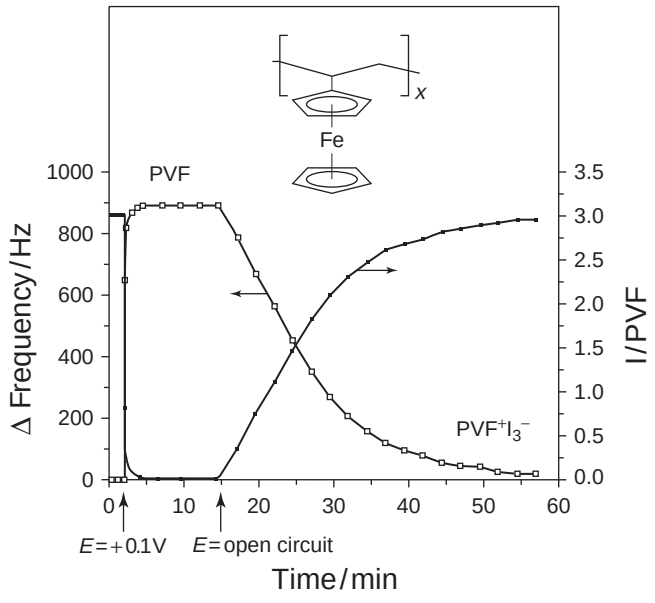

Figure 14 Frequency response of a PVF film in the presence of 5-mM KI_3 and 1.0-M KNO_3. The right hand ordinate refers to the number of equivalents of I incorporated into the film normalized to the amount of PVF coverage on the piezoelectrically active area (0.28 cm^2). $[\Gamma_{PVF}]_0 = 4.1 \times 10^{-8}$ equiv. cm^{-2}. (Reproduced with permission from Ward,[17] Copyright (1988) American Chemical Society.)

with some water always retained in the polymer film. Transport in PVF films can also be severely affected by the nature of the counterion. For example, EQCM studies indicated that ion exchange of ferro- or ferricyanide into PVF during electrochemical cycling resulted in irreversible incorporation of $Fe(CN)_6^{3-/4-}$ in the film.[76] This process led to a slow decrease in the electroactivity of the film, presumably due to an electrostatic "crosslinking" by the multiply charged ion that inhibited further transport of $Fe(CN)_6^{3-/4-}$. It should be noted that variability in these results may be expected in different laboratories because of differences in film preparation.

The EQCM can also be used to examine the kinetics of chemical reactions of solution species with redox polymer films. The chemical oxidation of a PVF film by KI_3 was monitored by measurement of the frequency decrease associated with insertion of the I_3^- counterion following oxidation (Figure 14). These experiments were possible because the PVF film could be held in its reduced form prior to the chemical reaction in the presence of the oxidant by holding the potential of the electrode negative of E^0 (PVF/PVF$^+$). This established the initial conditions for this measurement. The results yielded the stoichiometry of the reaction (1 equivalent I_3^- per PVF$^+$ formed) as well as the pseudo-first-order rate constant for the reaction.

3.7 Electrochemical Quartz Crystal Microbalance Investigations of Conducting Polymers

Interest in the EQCM method heightened considerably after it was reported in the investigation of polypyrrole films.[61] Since that time, numerous reports have appeared describing attempts to deconvolute the typically complicated behavior involved in the preparation of conducting polymer films and their subsequent doping behavior. The value of the EQCM was immediately evident in those initial polypyrrole studies in which it was determined that reduction of electrochemically (by oxidation of pyrrole in $LiClO_4$) prepared polypyrrole films resulted in Li^+ insertion to maintain electroneutrality rather than expulsion of ClO_4^-. This behavior was attributed to strong

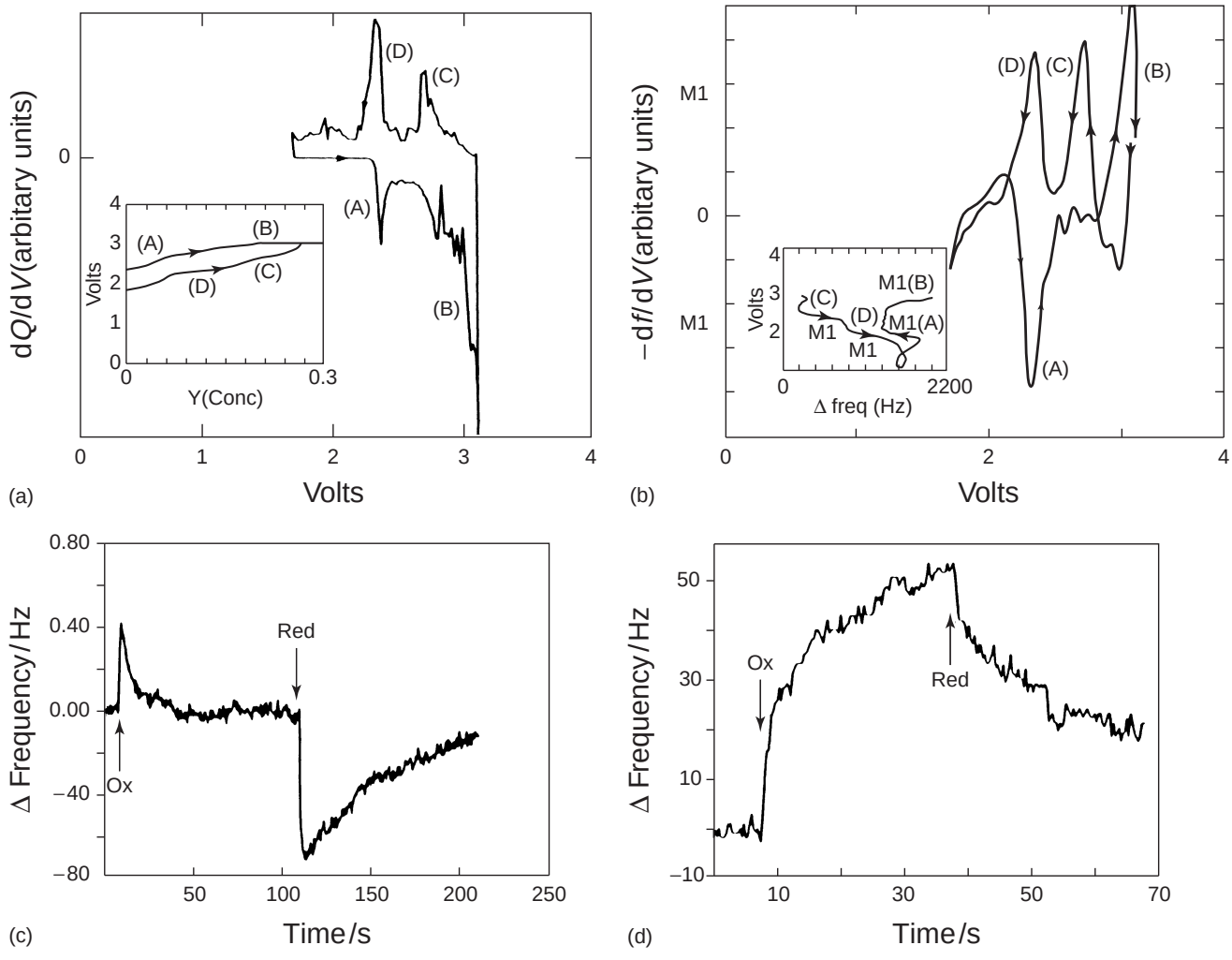

Figure 15 (a) Cyclic voltammetry data for a polypyrrole film immobilized on an EQCM in 1.0-M $LiClO_4$/THF, presented as dQ/dV versus V. Oxidation occurs at peaks A and B while reduction occurs at C and D. (inset shows V vs Q) (b) EQCM data presented as df/dV versus V showing differential mass changes associated with the oxidation and reduction features in (a). The up arrow (↑) indicates increasing mass and the down arrow indicates decreasing mass. (Reproduced with permission from Varineau and Buttry,[62] Copyright (1987) American Chemical Society.) (c) EQCM response upon oxidation of a polypyrrole film followed by reduction in tetraethylammonium tosylate electrolyte. The polypyrrole film was prepared by electrochemical oxidation of pyrrole. (d) EQCM response upon oxidation of a polypyrrole-co-[3-(pyrrol-1-yl)propanesulfonate] film followed by reduction in tetraethylammonium tosylate electrolyte. In both cases, the increase in frequency upon oxidation and decrease in frequency upon reduction are associated with transport of cations. (Reproduced with permission from Kaufman et al.[61] *Phys. Rev. Lett.*, Copyright (1984) by the American Physical Society.)

ionic pairing between cationic sites on the polymer and ClO_4^-. It was also discovered that subsequent cycling of these films in the presence of n-Bu_4N^+ p-$CH_3C_6H_4-SO_3^-$ resulted in frequency changes consistent with anion insertion during the oxidative doping step, with its expulsion during reduction (Figure 15). This behavior clearly revealed that the nature of ion transport in conducting polymers was strongly dependent upon the nature of the counterions. The impact of this behavior on the design of energy storage systems with regard to weight, power density, and energy density prompted several other EQCM investigations of conducting polymers, including poly(pyrrole), poly(aniline), and poly(thiophene). Decreasing the mass of the transport species would lead to higher energy densities, whereas higher power densities would be realized by faster transport rates.

Poly(pyrrole) film growth was also examined with the EQCM. It was concluded that film deposition involved the initial formation of soluble oligomers, which upon further polymerization precipitated on the electrode.[77,78] EQCM frequency changes that corresponded to the mass of deposited polypyrrole revealed

a second-order dependence of the electropolymerization rate on pyrrole concentration, suggesting the bimolecular coupling of oxidatively formed radical cations was involved in the rate determining step. Subsequent EQCM studies of poly(pyrrole) revealed more details of the ion transport and its dependence upon film preparation. Electropolymerization of pyrrole in the presence of large polymeric anions such as poly(4-styrenesulfonate)[79–82] and poly(vinylsulfonate) resulted in films whose transport properties differed significantly from films prepared in the presence of more conventional anions. In these studies conventionally prepared films exhibited frequency changes during doping–undoping cycles that indicated anion transport for ClO_4^-, BF_4^-, or PF_6^-. However, in electrolytes containing poly(4-styrenesulfonate) mixed transport of both poly(4-styrenesulfonate) anions and cations was observed in conventionally prepared films. If the films prepared in poly(4-styrenesulfonate) were used, then the transport during doping–undoping was completely dominated by the cation of the supporting electrolyte. Cation transport also dominated in the doping–undoping cycles of self-doped conducting copolymers of pyrrole and 3-(pyrrol-1-yl)propanesulfonate.

Poly(thiophene)-based polymers, conducting polymers closely related to poly(pyrrole) have also been examined with the EQCM.[83–85] These studies revealed the doping levels of poly(3-methylthiophene), as well as a nonlinear dependence of the charge on film thickness, the latter being inferred from the frequency change. Subsequent studies provided a thermodynamic model based on the EQCM response for transport of the ions, neutral ion pairs, and solvent based on the thermodynamic activity of the polymer film, similar to the model mentioned above. The transport of ions during doping–undoping cycles of poly(aniline), a conducting polymer film with doping behavior that is quite different from that of the aforementioned examples, has also been investigated with the EQCM.[86] These studies provided the Faradaic efficiencies during oxidative film growth by comparison of the coulometric charge and the frequency decrease. Interestingly, it was discovered that the Faradaic efficiencies for films prepared by cycling the potential were higher (ca. 40%) than those prepared at constant potential (ca. 10%). More importantly, mass changes during doping–undoping cycles revealed the extent of protonation of the polymer in both the doped and undoped states. Based on the EQCM results, it was concluded that each aniline ring contributed one electron to the conduction band of the polymer, in agreement with a proposed model.[87] Subsequent EQCM studies of poly(aniline) in nonaqueous solvents revealed that protons on the imine nitrogens of the polymer chains were retained during doping–undoping cycles, and that electroneutrality was maintained by transport of anions from the supporting electrolyte.[88]

3.8 Electrochemical Quartz Crystal Microbalance Investigations of Biologically Relevant Films

The EQCM has been used increasingly to examine the formation and properties of biologically relevant redox active films. One of the first such studies examined the electrochemically induced adsorption of the proteins human IgG and anti-IgG on Ag electrodes.[89] Upon oxidation of the Ag electrodes, the frequency of the EQCM decreased owing to formation of an insoluble layer of the proteins. These effects suggested that initial protein adsorption was related to interactions between oxidized Ag^+ ions on the electrode surface and functional groups on the protein.

Recent efforts have further demonstrated the utility of the EQCM in the characterization of electrode processes involving biologically relevant species. The adsorption of organosulfur reagents, followed by alternate electrostatic adsorption of ferrocene-modified poly(allylamine) and anionic glucose oxidase was monitored with the QCM in order to quantify the amount of redox-active material deposited on the gold electrode after each step.[90] Impedance analysis was employed to characterize viscoelastic contributions from the poly(allylamine) polymer solution in contact with the QCM. Viscoelastic contributions were also observed during the deposition of the glucose oxidase layer, although at long times these effects vanished so that the Sauerbrey approximation could be used reliably. The ability to measure the amount of material deposited in each step was important for elucidation of the electrochemical properties of these films. In a similar manner, the EQCM was employed to quantify the amount of electroactive and nonelectroactive protein in multilayer films prepared by alternate layer-by-layer deposition of cationic myoglobin or cationic cytochrome $P450_{cam}$ with either poly(styrenesulfonate) or poly(dimethyl diallyl) ammonium chloride.[91] The EQCM data reflected a linear increase of film mass with increasing number of layers.

The adsorption of flavin adenine dinucleotide (FAD) on the gold electrode of the EQCM revealed deposition of FAD upon reduction and its reversible desorption upon oxidation.[92] The voltammetric adsorption and desorption of cysteine at the gold electrode of an EQCM has been examined. Cysteine was adsorbed following cathodic reduction of the surface gold oxide, but the EQCM data indicated that the apparent surface mass upon reduction was less when cysteine was present than in solutions without cysteine. This was attributed to a decrease in surface hydration accompanying the loss of the oxide and adsorption of cysteine.[93]

The adsorption of green fibrous polymeric products formed by reaction of electrochemically generated dications of the carotenoid (7*E*,7′*Z*)-diphenyl-7,7′-diapocarotene was monitored with the EQCM. Coulometric measurements performed during cyclic voltammetry, combined with the EQCM measurement of the amount of polymer adsorbed, enabled determination of the average molar mass per electron (5400 g mol^{-1} electrons).[94]

4 IMPEDANCE ANALYSIS METHODS

Impedance analysis is being used increasingly to evaluate the contributions from viscous and viscoelastic effects associated with films on the quartz resonator. There is not sufficient space in this article to give a detailed explanation of this approach, but it can be described briefly as a technique in which a voltage within a specified range of frequencies is broadcast across the crystal and the current measured. The impedance or admittance is measured, and the Butterworth–van Dyke (BVD) equivalent circuit parameters determined by numerical fitting of the data. The mechanical properties corresponding to the electrical parameters can then be assessed. Commercially available impedance analyzers allow a complete set of parameters to be measured within 1 min, enabling dynamic measurements, albeit on a rather slow timescale.

Quartz crystals are electromechanical devices, and therefore their mechanical vibrations can be described in terms of electrical equivalents.[95] This also serves to enhance understanding of the EQCM, particularly the conditions under which the Sauerbrey equation, Equation (2), is valid. The quartz resonator can be described according to a mechanical model with elements of mass, compliance (the ability of an object to yield elastically under an applied force), and friction. The electrical equivalent of this system is an electrical circuit that has an inductor, a capacitor, and a resistor connected in series (Figure 16). In this equivalent circuit, the inductor, L_1, represents the mass displaced during oscillation, C_1 the energy stored during oscillation (the compliance is the inverse of the elastic, or Hooke's constant), and R_1 the energy dissipation due to losses that are tantamount to internal friction. In order to describe the quartz crystal behavior accurately, a parallel capacitance must also be included that represents the static capacitance of the quartz plate with its electrodes and any stray parasitic capacitances. The complete circuit is commonly referred to as the BVD circuit.[96,97] The series branch of the circuit is referred to as the motional branch since it reflects the vibrational behavior of the crystal.

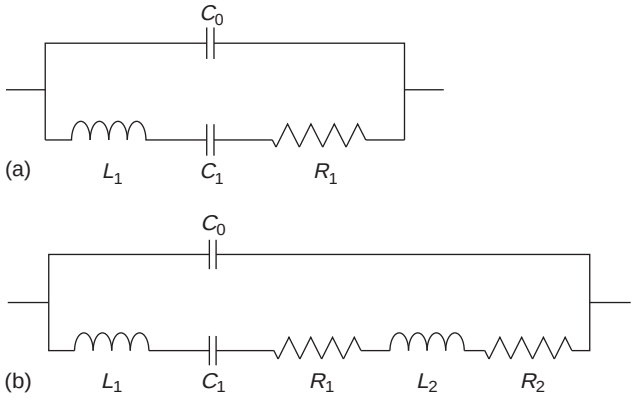

Figure 16 (a) BVD equivalent electrical circuit used to describe the mechanical properties of a quartz resonator. The components L_1, C_1 and R_1 in the motional branch of the circuit represent the inertial mass, compliance, and energy dissipation in the crystal, and C_0 represents the static capacitance of the quartz crystal. (b) Equivalent electrical circuit used to describe the mechanical properties of a quartz resonator immersed in a liquid. The inductance L_2 and resistance R_2 represent the mass and viscosity components of the liquid.

The relationship between this circuit and the quartz crystal is especially useful because the *LCR* branch is identical to a ''tank circuit'', in which oscillations can be sustained by cycling of current between the capacitor and the inductor. When the capacitor in this circuit discharges through the inductor, a magnetic field is established around the inductor as it opposes the current. When the capacitor discharge is complete and the current falls to zero, the electromotive force in the inductor creates a current in the direction opposite to the original current and the capacitor recharges. Repetition of this cycle results in electrical oscillation, with the oscillations dampened by an amount proportional to R. In the case of quartz crystals, the R values are rather small and sustained oscillations are favored.

As a result of the electromechanical relationship between quartz crystals and electrical circuits, the equations of harmonic motion of the quartz crystals are closely related in form to the expressions describing the properties of the *LCR* tank circuit. This has been reviewed elsewhere[3] and a detailed description is not given here. The equivalent electrical parameters in terms of crystal properties are given in Equations (12–15), along with typical experimental values for these parameters. In these relationships, D_Q is the dielectric constant of quartz, ε_0 the permittivity of free space, r a dissipation coefficient corresponding to the energy losses during oscillation, ε the piezoelectric stress constant and c the elastic constant. Note that while L_1, C_1, and R_1 depend upon ε, C_0 does not participate directly in piezoelectricity. It should also be noted that L_1 depends upon the density; in fact, the

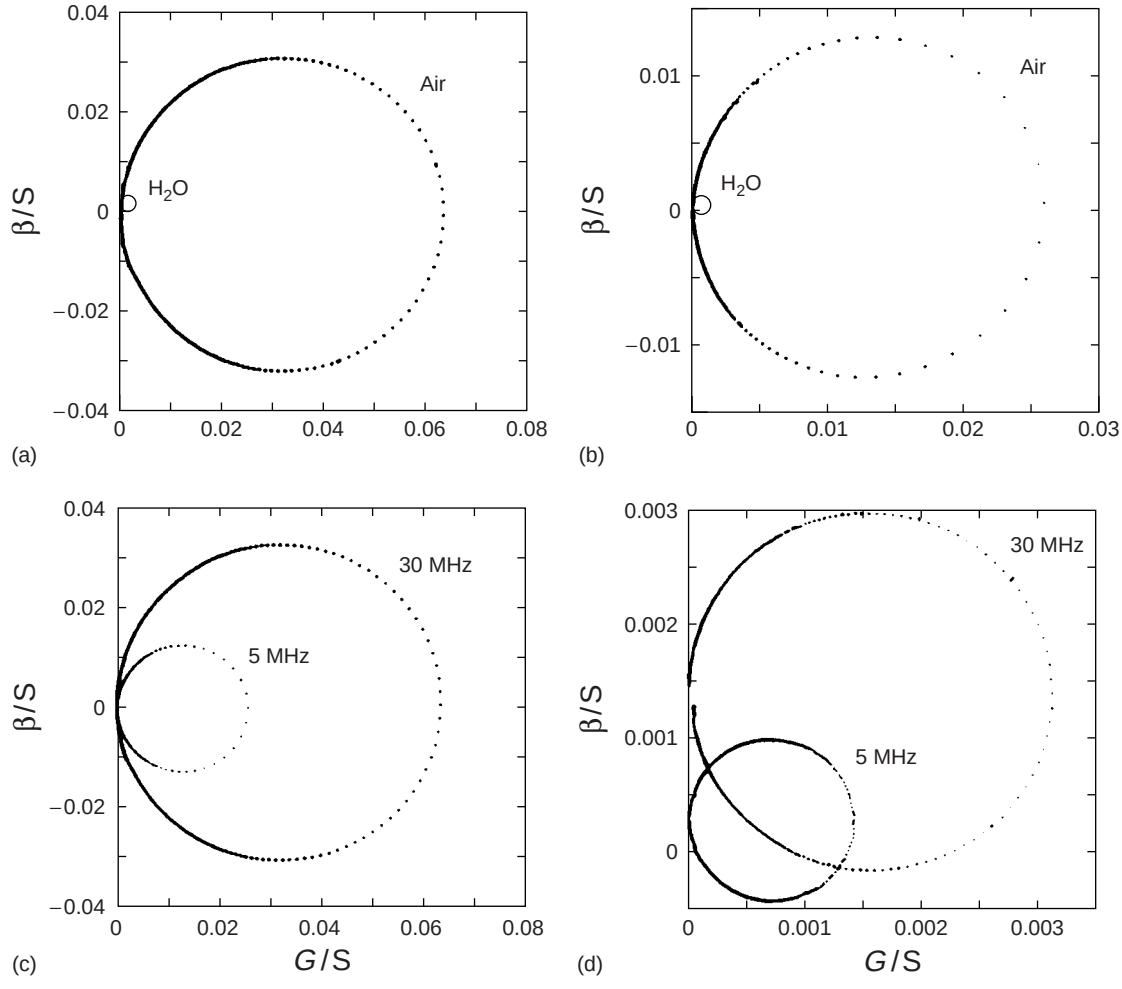

Figure 17 Admittance analysis data for 5-MHz and 30-MHz quartz resonators in aqueous media. The increased capacitance of the thinner 30-MHz resonator lifts the admittance circle upward along the imaginary susceptance (jB) axis, but the decreased resistance associated with the decreased crystal thickness increases the diameter of the admittance circle. The latter effect is critical as it compensates for the capacitance effect such that the admittance circle crosses the real axis, which is required for establishing resonance. The admittance data was collected with a Hewlett Packard Model 4194A Impedance Analyzer. (Reproduced with permission from Lin et al.,[6] Copyright (1993) American Chemical Society.)

quantity $t_Q^3\rho/A$ is equivalent to the mass per unit area in the Sauerbrey equation. These equivalent representations provide a quantitative approach to examining the properties of the EQCM, the role of the liquid environment and thin films on the resonant frequency response, and the design of quartz resonators.

$$C_0 = \frac{D_Q \varepsilon_0 A}{t_Q} \approx 10^{-12}\,\mathrm{F} \tag{12}$$

$$C_1 = \frac{8A\varepsilon^2}{\pi^2 t_Q c} \approx 10^{-14}\,\mathrm{F} \tag{13}$$

$$R_1 = \frac{t_Q^3 r}{8A\varepsilon^2} \approx 100\,\Omega \tag{14}$$

$$L_1 = \frac{t_Q^3 \rho}{8A\varepsilon^2} \approx 0.075\,\mathrm{H} \tag{15}$$

The use of impedance analysis, or its inverse admittance analysis, for characterization of crystal properties and performance can be demonstrated by the rather simple comparison of the admittance plots for 5-MHz and 30-MHz quartz crystals mentioned above (Figure 17). The smaller thickness of the 30-MHz quartz crystal leads to an increase in C_0 and the expected shift of the admittance locus upwards on the admittance plot (by an amount equal to wC_0). This increases the likelihood that the admittance locus will not cross the real axis. Under this condition, operation of the EQCM at its resonant frequency using a feedback-mode oscillator, which relies on maintaining a frequency at zero phase angle, can be difficult. The smaller area of the electrode overlap for the 30-MHz resonators somewhat diminishes this effect. More important, the 30-MHz resonators have smaller

values of R_1 and L_1, owing to the smaller thickness of these crystals compared to their lower frequency counterparts. The decrease in R_1 is evident from the 30-MHz crystal admittance locus, which is significantly larger than the 5-MHz locus. The admittance locus crosses the real axis for both resonators, but the conductance value of the series resonant frequency f_s, where the locus crosses the real axis, is larger for the 30-MHz resonator. Because EQCM experiments using a feedback-mode oscillator operate at f_s, the quartz crystals can be operated with lower loss in liquid media or in air than the more commonly used 5-MHz crystals.

While the BVD circuit accurately describes the operation of the QCM in vacuum or the gas phase, it is not a sufficient description when the QCM or EQCM is used in liquids where the liquid density and viscosity alter the resonator characteristics. The density of the liquid effectively adds to the mass of the resonator, while its viscosity provides additional energy damping. The effect of a Newtonian liquid can be described by adding an additional inductance and resistance L_2 and R_2 in series with the motional branch of the BVD circuit, in which L_2 and R_2 are related to the extra mass of the liquid and its viscosity, respectively (Figure 17).[98]

Our laboratory devised a method based on an electromechanical model that relates the electrical parameters to the mechanical characteristics of a composite resonator consisting of a AT-cut thickness shear mode quartz crystal and a viscoelastic polymer film. The theoretical component of this method is based on a previously reported relationship Equation (16),[99] where μ_F is the film shear modulus (or equivalently, the storage modulus G'), $\bar{\bar{\mu}}_F$ is the complex film shear modulus, η_F is the film viscosity (or equivalently, the loss modulus-frequency quotient G''/w), ρ_F is the film density, l_F is the film thickness, ρ_Q is the quartz density, w is the angular frequency ($w = 2\pi f$), l_Q is the quartz thickness, A_{ACT} is the piezoelectric active area, ε_{22} is the quartz dielectric constant, e_{26} is the quartz piezoelectric constant, c_{66} is the quartz shear modulus, $\bar{\bar{c}}_{66}$ is the complex quartz shear modulus, η_Q is the quartz viscosity, and Y is the complex resonant admittance.

$$Y = \frac{\dfrac{iwA_{ACT}\varepsilon_{22}}{l_Q}\; k_Q\bar{\bar{c}}_{66}\sin(k_Q l_Q) + k_F\bar{\bar{\mu}}_F\tan(k_F l_F)\cos(k_Q l_Q)}{k_Q\sin(k_Q l_Q) + k_F\bar{\bar{\mu}}_F\tan(k_F l_F)\cos(k_Q l_Q) - \dfrac{2e_{26}^2}{l_Q\varepsilon_{22}} \times \left[1 - \cos(k_Q l_Q) + \dfrac{k_F\bar{\bar{\mu}}_F}{2k_Q\bar{\bar{c}}_{66}}\tan(k_F l_F)\sin(k_Q l_Q)\right]} \tag{16}$$

where

$$k_Q = w\sqrt{\frac{\rho_Q}{\bar{\bar{c}}_{66}}} = w\sqrt{\frac{\rho_Q}{c_{66} + \dfrac{e_{26}^2}{\varepsilon_{22}} + iw\eta_Q}}$$

and

$$k_F = w\sqrt{\frac{\rho_F}{\bar{\bar{\mu}}_F}} = w\sqrt{\frac{\rho_F}{\mu_F + iw\eta_F}}$$

This electromechanical equation is derived from linear equations of motion relating the electrical admittance and the mechanical properties of a composite resonator. An algorithm based on this relationship was limited to measurement of the mechanical properties of polymer films containing various amounts of solvent that affected the viscoelastic properties. These measurements require somewhat cumbersome and time-consuming numerical analysis during measurement that inhibits the use of complete impedance analysis techniques in EQCM applications.[100–102] Even direct impedance analysis requires long analysis times because the measurement involves sweeping a range of frequencies to evaluate the resonator conductance and numerically fit the data to obtain the BVD parameters. Consequently, experimental strategies based on impedance analysis have not been used extensively in EQCM applications. However, changes in R can be rapidly inferred from changes in the amplitude of the voltage output of the oscillator during operation with a conventional feedback oscillator.

5 EMERGING METHODS

Several modifications of the quartz microbalance have appeared in which the oscillation is induced by electrodes that are near, but not in contact with the quartz surface.[103–107] These resonators, which can be referred to as electrode-separated quartz crystal resonators, allow examination of interfacial processes occurring directly on the quartz surface or on other films deposited on the quartz. The capabilities of these methods have been extended to a scanning configuration in which the position of a small conducting probe, serving as one of the excitation electrodes, is controlled by computer-activated positioning of the sample stage.[108] In principle, this procedure enables qualitative mapping of the mass distribution and viscoelastic properties of films on the quartz resonator. Two configurations have actually been demonstrated. In an ''overscanning'' mode the conducting probe is scanned over the upper surface, onto which a film of interest has been deposited, and the bottom surface has the conventional metal electrode

covering a large region of the crystal. An alternative ''underscanning'' mode relies on the conducting probe scanning the lower surface of the quartz crystal while the upper surface is coated conventionally with the large area excitation electrode. Both configurations allow introduction of an electrochemical working electrode. In the overscanning mode the working electrode is independent of the excitation electrodes, whereas in the underscanning mode the excitation and working electrode are the same.

The format of the EQCM is amenable to combination with other in situ techniques. This was illustrated for the SEM/EQCM experiments described earlier in this article. This capability is further demonstrated by ellipsometric studies of the nucleation and growth of poly(aniline) films performed in conjunction with the EQCM,[109,110] which allowed complete characterization of the polymer films (Figure 18). Whereas the EQCM provided the total mass of the deposit, ellipsometry provided the thickness. Accordingly, the combination of these techniques allows determination of the density, which is not realized by either technique alone. The major conclusions of these studies were that the kinetics for growth of the poly(aniline) films depended upon the electrochemical conditions, and that self-assembled monolayers of aniline derivatives on the electrode surface promoted nucleation and films with higher densities.

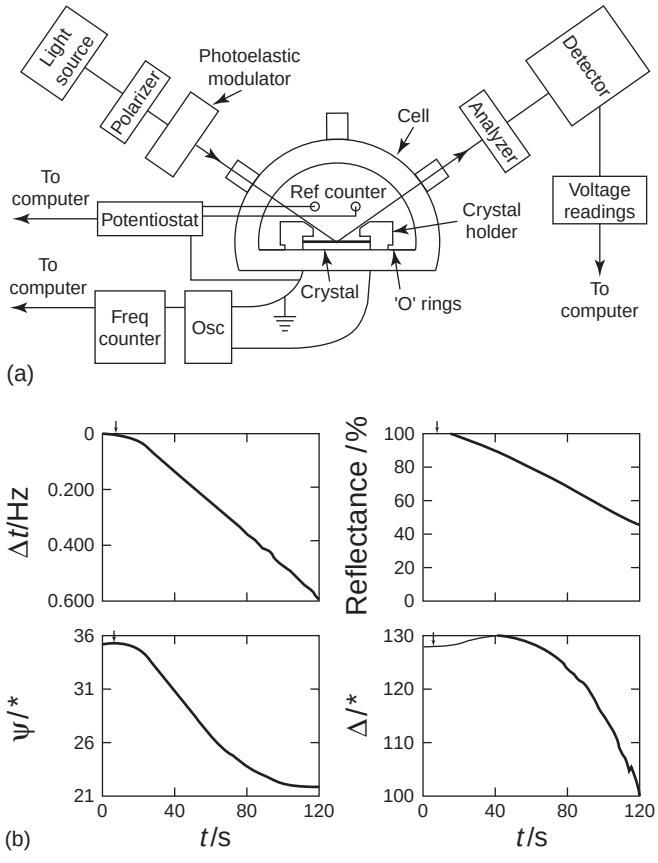

Figure 18 (a) Schematic representation of the experimental apparatus for simultaneous ellipsometry–EQCM measurements. (b) Optical and EQCM data for poly(aniline) growth on a Pt electrode in 1-M aniline containing 2-M HCl at 0.7 V (vs SCE). Wavelength = 550 nm, angle of incidence = 65°. Δ and ψ are the ellipsometric parameters derived from the 50 kHz, 100 kHz, and DC (direct current) components of the signal generated at the detector by the reflected beam. (Reproduced from Rishpon et al.[109] Copyright (1990) with permission from Elsevier Science.)

6 SUMMARY

The increasing use of the EQCM and the examples described in this article demonstrate the power of this method in elucidating the fundamental interfacial processes occurring at electrode surfaces. The rather simple concept and low cost of the equipment necessary to perform EQCM experiments should encourage physical, analytical, and electrochemists to use this technique routinely in the laboratory. Although numerous laboratories use the EQCM regularly, the contributions of interfacial slip, stress, surface roughness, viscosity, viscoelasticity, and influence of the liquid phase to propagation of the acoustic energy are difficult to quantify. The new practitioner (and even experienced ones) should expend some effort to grasp the basic principles of EQCM operation in order to avoid pitfalls and misinterpretations that can occur if these effects are ignored. Generally, verification of ideal behavior is possible through fairly simple experiments, as discussed in this article, or through more sophisticated impedance analysis techniques. It is anticipated that as the EQCM becomes more widely appreciated and used, many of these effects will be better understood. Nevertheless, the EQCM is well within the reach of most electrochemical laboratories and properly used can provide details of electrode processes that were previously unattainable.

ABBREVIATIONS AND ACRONYMS

BVD	Butterworth–van Dyke
DC	Direct Current
EQCM	Electrochemical Quartz Crystal Microbalance
FAD	Flavin Adenine Dinucleotide
PVF	Poly(vinylferrocene)
QCM	Quartz Crystal Microbalance
RF	Radiofrequency
SCE	Saturated Calomel Electrode

SEM Scanning Electron Microscopy
UPD Underpotential Deposition

RELATED ARTICLES

Coatings **(Volume 2)**
Rheology in Coatings, Principles and Methods

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

REFERENCES

1. M.D. Ward, D.A. Buttry, 'In Situ Interfacial Mass Detection with Piezoelectric Transducers', *Science*, **249**, 1000 (1990).
2. D.A. Buttry, M.D. Ward, 'Measurement of Interfacial Processes at Electrode Surfaces with the Electrochemical Quartz Crystal Microbalance, *Chem. Rev.*, **92**, 1355 (1992).
3. D.A. Buttry, 'Applications of the Quartz Crystal Microbalance to Electrochemistry', *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 2, Vol. 17, 1991.
4. P. Curie, J. Curie, 'Cristallophysique-Development de L'electricite Polaire dans les Cristaux Hemiedes a Faces Inclines', *C.R. Acad. Sci.*, **91**, 294 (1880).
5. C.S. Lu, O. Lewis, 'Investigation of Film-thickness Determination by Oscillating Quartz Resonators with Large Mass Load', *J. Appl. Phys.*, **43**, 4385 (1972).
6. Z. Lin, C. Yip, I.S. Joseph, M.D. Ward, 'Operation of an Ultrasensitive 30 MHz Quartz Crystal Microbalance in Liquids', *Anal. Chem.*, **65**, 1646 (1993).
7. K.K. Kanazawa, J.G. Gordon III, 'Frequency of a Quartz Microbalance in Contact with Liquid Sensors to the Liquid Phase by Interfacial Viscosity', *Anal. Chem.*, **57**, 1770 (1985).
8. L.V. Rajakovic, B.A. Cavic-Vlasak, V. Ghaemmaghami, K.M.R. Kallury, A.L. Kipling, M. Thompson, 'Mediation of Acoustic Energy Transmission from Acoustic Wave', *Anal. Chem.*, **63**, 615 (1991).
9. A.L. Kipling, M. Thompson, 'Network Analysis Method Applied to Liquid Phase Acoustic Sensors', *Anal. Chem.*, **62**, 1514 (1990).
10. J. Wang, L.M. Frostman, M.D. Ward, 'Self-assembled Thiol Monolayers with Carboxylic Acid Functionality, Measuring pH-dependent Phase Transitions with the Quartz Crystal Microbalance', *J. Phys. Chem.*, 96 (1992).
11. R. Schumacher, G. Borges, K.K. Kanazawa, *Surf. Sci.*, **163**, L621 (1985).
12. R. Schumacher, J. Gordon, O. Melroy, 'Observation of Morphological Relaxation of Copper and Silver Electrodes in Solution Using a Quartz Crystal Microbalance', *J. Electroanal. Chem.*, **216**, 127 (1987).
13. A. Muller, M. Wicker, R. Schumacher, R.N. Schindler, *Ber. Bunsenges. Phys. Chem.*, **92**, 1395 (1988).
14. T. Nomura, M. Iijima, 'Electrolytic Determination of Nanomolar Concentrations of Silver in Solution with a Piezoelectric Quartz Crystal', *Anal. Chim. Acta*, **131**, 97 (1981).
15. S. Bruckenstein, M. Shay, 'Experimental Aspects of Use of the Quartz Crystal Microbalance in Solution', *Electrochim. Acta*, **30**, 1295 (1985).
16. R. Schumacher, 'The Quartz Microbalance, a Novel Approach to the In-situ Investigation of Interfacial Phenomena at the Solid/Liquid Junction', *Angew. Chem. Int. Ed. Engl.*, **29**, 329 (1990).
17. M.D. Ward, 'Investigation of Open Circuit Reactions of Polymer Films Using the Quartz Crystal Microbalance. Reactions of Poly(vinylferrocene) Films', *J. Phys. Chem.*, **92**, 2049 (1988).
18. R. Bechmann, 'Single Response Thickness-shear Mode Resonators Using Circular Bevelled Plates', *J. Sci. Instrum.*, **29**, 73 (1952).
19. R. Bechmann, 'Quartz AT-type Filter Crystals for Frequency Range 0.7 to 60 MC', *Proc. Inst. Radio Engineers*, **49**, 523 (1961).
20. H.F. Tiersten, 'Analysis of Trapped Energy Resonators Operating in Overtones of Thickness-shear', *Proc. Annu. Freq. Control Symp.*, **28**, 44 (1974).
21. K.E. Heusler, A. Grzegorzewski, L. Jackel, J. Pietrucha, 'Measurement of a Mass and Surface Stress at One Electrode of a Quartz Oscillator', *Ber. Bunsenges. Phys. Chem.*, **92**, 1395 (1988).
22. E.P. EerNisse, *Methods and Phenomena*, eds. C. Lu, A. Czanderna, Elsevier Sequoia, Lausanne, Vol. 4, 1984.
23. G. Sauerbrey, 'The Use of a Quartz Crystal Oscillator for Weighing Thin Layers and for Microweighing Applications', *Z. Physik*, **155**, 206 (1959).
24. D.M. Ullevig, J.F. Evans, M.G. Albrecht, 'Effects of Stressed Materials on the Radial Sensitivity Function of a Quartz Crystal Microbalance', *Anal. Chem.*, **54**, 2341 (1982).
25. L. Koga, Y. Tsuzuki, S.N. Witt, A.L. Bennet, 'Measurements of the Vibrations of Quartz Plates', *Proc. Annu. Freq. Control Symp.*, **14**, 53 (1960).
26. H. Fukuyo, A. Yokoyama, N. Ooura, S. Nonaka, 'Vibration of Biconvex Circular AT-Cut Plate', *Bull. Tokyo Inst. Technol.*, **72**, 1 (1965).
27. K.S. van Dyke, 'Strain Patterns in Thickness-shear Quartz Resonators', *Proc. Annu. Freq. Control Symp.*, **111**, 1 (1957).
28. A.C. Hillier, M.D. Ward, 'Scanning Electrochemical Mass Sensitivity Mapping of the Quartz Crystal Microbalance in Liquid Media', *Anal. Chem.*, **64**, 2539 (1992).

29. M.D. Ward, E.J. Delawski, 'Radial Sensitivity of the Quartz Crystal Microbalance in Liquid Media', *Anal. Chem.*, **63**, 886 (1991).
30. C. Gabrielli, M. Keddam, R. Torresi, 'Calibration of the Electrochemical Quartz Crystal Microbalance', *J. Electrochem. Soc.*, **138**, 2657 (1991).
31. B.A. Martin, H.E. Hager, 'Velocity Profile On Quartz Crystals Oscillating in Liquids', *J. Appl. Phys.*, **65**, 2630 (1989).
32. O. Melroy, K.K. Kanazawa, J.G. Gordon, D.A. Buttry, 'Direct Determination of the Mass of an Underpotentially Deposited Monolayer of Lead On Gold', *Langmuir*, **2**, 697 (1987).
33. G.K. Guttwein, A.D. Ballato, T.J. Lukaszek, US Patent, 3, 694, 677; Sept. 26, 1972.
34. W.W. Lee, H.S. White, M.D. Ward, 'Depletion Layer Effects on the Response of the Electrochemical Quartz Crystal Microbalance', *Anal. Chem.*, **65**, 3232 (1993).
35. S. Bruckenstein, M. Shay, 'Experimental Aspects of Use of the Quartz Crystal Microbalance in Solution', *Electrochim. Acta*, **30**, 1295 (1985).
36. H.E. Hager, R.D. Ruedisueli, M.E. Buehler, 'The Use of Piezoelectric Crystals as Electrode Substrates in Iron Corrosion Studies, the Real-time, In Situ Determination of Dissolution and Film Formation Reaction Rates', *Corrosion*, **42**, 345 (1986).
37. M.R. Deakin, O. Melroy, 'Underpotential Metal Deposition on Gold, Monitored In Situ with a Quartz Crystal Microbalance', *J. Electroanal. Chem.*, **239**, 321 (1988).
38. M. Hepel, S. Bruckenstein, 'Tracking Anion Expulsion During Underpotential Deposition of Lead at Silver Using the Quartz Microbalance', *Electrochim. Acta*, **34**, 1499 (1989).
39. M. Hepel, K. Kanige, S. Bruckenstein, 'In Situ Underpotential Deposition Study of Lead on Silver Using the Electrochemical Quartz Crystal Microbalance. Direct Evidence for Lead(II) Adsorption Before Spontaneous Charge Transfer', *J. Electroanal. Chem.*, **266**, 409 (1989).
40. M. Hepel, K. Kanige, S. Bruckenstein, 'Expulsion of Borate Ions from the Silver/Solution Interfacial Region During Underpotential Deposition Discharge of Pb(II) in Borate Buffers', *Langmuir*, **6**, 1063 (1990).
41. M. Deakin, T. Li, O. Melroy, 'A Study of the Electrosorption of Bromide and Iodide Ions on Gold Using the Quartz Crystal Microbalance', *J. Electroanal. Chem.*, **243**, 343 (1988).
42. M. Benje, M. Eiermann, U. Pitterman, K.G. Weil, 'An Improved Quartz Crystal Microbalance. Applications to the Electrocrystallization and -dissolution of Nickel', *Ber. Bunsenges. Phys. Chem.*, **90**, 435 (1986).
43. M. Benje, U. Hofmann, U. Pitterman, K.G. Weil, 'Anodic Dissolution of Thin Nickel–Phosphorus Films', *Ber. Bunsenges. Phys. Chem.*, **92**, 1257 (1988).
44. R. Schumacher, A. Muller, W. Stockel, 'An In Situ Study of the Mechanism of the Electrochemical Dissolution of Copper in Oxygenated Sulfuric Acid. An Application of the Quartz Microbalance', *J. Electroanal. Chem.*, **219**, 311 (1987).
45. G.T. Cheek, W.E. O'Grady, 'Measurement of Hydrogen Uptake by Palladium Using a Quartz Crystal Microbalance', *J. Electroanal. Chem.*, **277**, 341 (1990).
46. L. Grasjo, M. Seo, 'Measurement of Absorption of Hydrogen and Deuterium into Palladium During Electrolysis by a Quartz Crystal Microbalance', *J. Electroanal. Chem.*, **296**, 233 (1990).
47. N. Yamamoto, T. Ohsaka, T. Terashima, N. Oyama, 'In Situ Electrochemical Quartz Crystal Microbalance Studies of Water Electrolysis at a Palladium Cathode in Acidic Aqueous Media', *J. Electroanal. Chem.*, **274**, 313 (1989).
48. E.P. EerNisse, 'Simultaneous Thin Film Stress and Mass-change Measurements Using Quartz Resonators', *J. Appl. Phys.*, **43**, 1330 (1972).
49. E.P. EerNisse, 'Extension of the Double Resonator Technique', *J. Appl. Phys.*, **44**, 4482 (1973).
50. J.J. Donahue, D.A. Buttry, 'Adsorption and Micellization Influence the Electrochemistry of Redox Surfactants Derived from Ferrocene', *Langmuir*, **5**, 671 (1989).
51. L. Nordyke, D.A. Buttry, 'Redox Surfactants are Chemical Probes of Electrode Surface Functionalization Derived from Disulfide Immobilization on Gold', *Langmuir*, **7**, 380 (1991).
52. H.C. De Long, J.J. Donahue, D.A. Buttry, 'Ionic Interactions in Electroactive Self-assembled Monolayers of Ferrocene Species', *Langmuir*, **7**, 2196 (1991).
53. G.S. Ostrom, D.A. Buttry, 'Quartz Crystal Microbalance Studies of Deposition and Dissolution Mechanisms of Electrochromic Films of Diheptylviologen Bromide', *J. Electroanal. Chem.*, **256**, 411 (1988).
54. E. Mori, C.K. Baker, J.R. Reynolds, K. Rajeshwar, 'Aqueous Electrochemistry of Tellurium at Glassy Carbon and Gold. A Combined Voltammetry–Oscillating Quartz Crystal Microgravimetry Study', *J. Electroanal. Chem.*, **252**, 441 (1988).
55. L.A. Larew, J.S. Gordon, Y.-L. Hsiao, D.C. Johnson, D.A. Buttry, 'Application of an Electrochemical Quartz Crystal Microbalance to a Study of Pure and Bismuth-doped B-Lead Dioxide Film Electrodes', *J. Electrochem. Soc.*, **137**, 3071 (1990).
56. B.J. Feldman, O.R. Melroy, 'Ion Flux During Electrochemical Charging of Prussian Blue Films', *J. Electroanal. Chem.*, **234**, 213 (1987).
57. S.J. Lasky, D.A. Buttry, 'Mass Measurements Using Isotopically Labeled Solvents Reveal the Extent of Solvent Transport During Redox in Thin Films on Electrodes', *J. Am. Chem. Soc.*, **110**, 6258 (1988).
58. K. Aoki, T. Miyamoto, Y. Ohsawa, 'The Determination of the Selectivity Coefficient of Na^+ Versus Li^+ on Prussian Blue Thin Film in Propylene Carbonate by Means of a Quartz Crystal Microbalance', *Bull. Chem. Soc. Jpn.*, **62**, 1658 (1989).

59. M.R. Deakin, H. Byrd, 'Prussian Blue Coated Quartz Crystal Microbalance as a Detector for Electroinactive Cations in Aqueous Solution', *Anal. Chem.*, **61**, 290 (1989).
60. A.R. Hillman, D.C. Loveday, S. Bruckenstein, 'Electrochemical Quartz Crystal Microbalance Monitoring of Poly(vinylferrocene) Electrodeposition', *Langmuir*, **7**, 191 (1991).
61. J.N. Kaufman, K.K. Kanazawa, G.B. Street, 'Gravimetric Electrochemical Voltage Spectroscopy, In Situ Mass Measurements During Electrochemical Doping of the Conducting Polymer Polypyrrole', *Phys. Rev. Lett.*, **53**, 2461 (1984).
62. P.T. Varineau, D.A. Buttry, 'Applications of the Quartz Crystal Microbalance to Electrochemistry. Measurement of Ion and Solvent Populations in Thin Films of Poly(vinylferrocene) as Functions of Redox State', *J. Phys. Chem.*, **91**, 1292 (1987).
63. P.T. Varineau, 'Determination of the Ion and Solvent Content of Thin Films of Poly(vinylferrocene) on Gold [Electrode] Using the Quartz Crystal Microbalance', PhD Thesis, University of Wyoming, 1989.
64. S. Bruckenstein, A.R. Hillman, 'Consequences of Thermodynamic Restraints on Solvent and Ion Transfer During Redox Switching of Electroactive Polymers', *J. Phys. Chem.*, **92**, 4837 (1988).
65. A.R. Hillman, D.C. Loveday, M.J. Swann, S. Bruckenstein, C.P. Wilde, 'Transport of Neutral Species in Electroactive Polymer Films', *J. Chem. Soc., Faraday Trans.*, **87**, 2047 (1991).
66. A.R. Hillman, D.C. Loveday, S. Bruckenstein, 'A General Approach to the Interpretation of Electrochemical Quartz Crystal Microbalance Data, Part II. Chronoamperometry, Temporal Resolution of Mobile Species Transport in Poly(vinylferrocene) Films', *J. Electroanal. Chem.*, **300**, 67 (1991).
67. S. Bruckenstein, C.P. Wilde, M. Shay, A.R. Hillman, D.C. Loveday, 'Observation of Kinetic Effects During Interfacial Transfer at Redox Polymer Films Using the Quartz Microbalance', *J. Electroanal. Chem.*, **258**, 457 (1989).
68. A.R. Hillman, D.C. Loveday, S. Bruckenstein, C.P. Wilde, 'Criteria Governing Ion and Solvent Transport Rates in Electroactive Polymers, the Existence of Kinetic Permselectivity', *J. Chem. Soc., Faraday Trans.*, **86**, 437 (1990).
69. R. Borjas, D.A. Buttry, 'Solvent Swelling Influences the Electrochemical Behavior and Stability of Thin Films of Nitrated Poly(styrene)', *J. Electroanal. Chem.*, **280**, 73 (1990).
70. S. Bruckenstein, C.P. Wilde, M. Shay, A.R. Hillman, 'Experimental Observations on Transport Phenomena Accompanying Redox Switching in Polythionine Films Immersed in Strong Acid Solutions', *J. Phys. Chem.*, **94**, 787 (1990).
71. A.R. Hillman, M.J. Swann, S. Bruckenstein, 'General Approach to the Interpretation of Electrochemical Quartz Crystal Microbalance Data. 1. Cyclic Voltammetry, Kinetic Subtleties in the Electrochemical Doping of Polybithiophene Films', *J. Phys. Chem.*, **95**, 3271 (1991).
72. S. Bruckenstein, C.P. Wilde, A.R. Hillman, 'Transport Phenomena Accompanying Redox Switching in Polythionine Films Immersed in Aqueous Acetic Acid Solutions', *J. Phys. Chem.*, **94**, 6458 (1990).
73. S. Bruckenstein, A.R. Hillman, M.J. Swann, 'Transient Mass Excursions During Switching of Polythionine Films Between Equi-mass Equilibrium Redox States', *J. Electrochem. Soc.*, **137**, 1323 (1990).
74. A.R. Hillman, D.C. Loveday, M.J. Swann, R.M. Eales, A. Hamnett, S.J. Higgins, S. Bruckenstein, C.P. Wilde, 'Charge Transport in Electroactive Polymer Films', *Faraday Discuss. Chem. Soc.*, **88**, 151 (1989).
75. A.R. Hillman, N.A. Hughes, S. Bruckenstein, 'Solvation Phenomena in Poly(vinylferrocene) Films, Effect of History and Redox State', *J. Electrochem. Soc.*, **139**, 74 (1992).
76. M.D. Ward, 'Ion Exchange of Ferro(Ferricyanide) into Poly(vinylferrocene) Films', *J. Electrochem. Soc.*, **135**, 2747 (1988).
77. C.K. Baker, J.R. Reynolds, 'Use of the Quartz Microbalance in the Study of Polyheterocycle Electrosynthesis', *Synth. Met.*, **28**, C21 (1989).
78. C.K. Baker, J.R. Reynolds, 'A Quartz Microbalance Study of the Electrosynthesis of Polypyrrole', *J. Electroanal. Chem.*, **251**, 307 (1988).
79. C.K. Baker, Y.-J. Qiu, J.R. Reynolds, 'Electrochemically Induced Charge and Mass Transport in Polypyrrole/Poly(styrenesulfonate) Molecular Composites', *J. Phys. Chem.*, **95**, 4446 (1991).
80. K. Naoi, M.M. Lien, W.H. Smyrl, 'Quartz Crystal Microbalance Study, Ionic Motion Across Conducting Polymers', *J. Electrochem. Soc.*, **138**, 440 (1991).
81. K. Naoi, M.M. Lien, W.H. Smyrl, 'Quartz Crystal Microbalance Analysis. Part 1. Evidence of Anion or Cation Inversion into Electropolymerized Conducting Polymers', *J. Electroanal. Chem.*, **272**, 273 (1989).
82. J.R. Reynolds, N.S. Sundaresan, M. Pomerantz, S. Basak, C.K. Baker, 'Self-doped Conducting Copolymers, a Charge and Mass Transport Study of Poly{pyrrole-co[3-(pyrrol-1-yl)propanesulfonate]}', *J. Electroanal. Chem.*, **250**, 355 (1988).
83. R. Borjas, D.A. Buttry, 'EQCM Studies of Film Growth, Redox Cycling and Charge Trapping of a N-Doped and P-Doped Poly(thiophene)', *Chem. Mater.*, **3**, 872 (1991).
84. S. Servagent, E. Vieil, 'In Situ Quartz Microbalance Study of the Electrosynthesis of Poly(3-Methylthiophene)', *J. Electroanal. Chem.*, **280**, 227 (1990).
85. A.R. Hillman, M.J. Swann, S. Bruckenstein, 'Ion and Solvent Transfer Accompanying Polybithiophene

Doping and Undoping', *J. Electroanal. Chem.*, **291**, 147 (1990).
86. D.O. Orata, D.A. Buttry, 'Determination of Ion Populations and Solvent Content as Functions of Redox State and pH in Polyaniline', *J. Am. Chem. Soc.*, **109**, 3574 (1987).
87. S.H. Glarum, J.H. Marshall, 'In Situ Potential Dependence of Poly(aniline) Paramagnetism', *J. Phys. Chem.*, **90**, 6076 (1986).
88. H. Daifuku, T. Kawagoe, N. Yamamoto, T. Ohsaka, N. Oyama, 'A Study of the Redox Reaction Mechanisms of Polyaniline Using a Quartz Crystal Microbalance', *J. Electroanal. Chem.*, **274**, 313 (1989).
89. E.S. Grabbe, R.P. Buck, O.R. Melroy, 'Cyclic Voltammetry and Quartz Microbalance Electrogravimetry of Igg and Anti-igg Reactions on Silver', *J. Electroanal. Chem.*, **223**, 67 (1987).
90. J. Hodak, R. Etchenique, E.J. Calvo, K. Singhal, P.N. Bartlett, 'Layer-by-layer Self-assembly of Glucose Oxidase with a Poly(allylamine)ferrocene Redox Mediator', *Langmuir*, **13**, 2708 (1997).
91. Y.M. Lvov, Z. Lu, J.B. Schenkman, X. Zu, J.F. Rusling, 'Direct Electrochemistry of Myoglobin and Cytochrome $P450_{cam}$ in Alternate Layer-by-layer Films with DNA and Other Polyions', *J. Am. Chem. Soc.*, **120**, 4073 (1998).
92. Y. Wang, G. Zhu, E. Wang, 'Electrochemical Behavior of FAD at a Gold Electrode Studied by Electrochemical Quartz Crystal Microbalance', *Anal. Chim. Acta*, **338**, 97 (1997).
93. A.J. Tudos, P.J. Vandeberg, D.C. Johnson, 'Evaluation of EQCM Data from a Study of Cysteine Adsorption on Gold Electrodes in Acidic Media', *Anal. Chem.*, **67**, 552 (1995).
94. G. Gao, D.B. Wurm, Y.-T. Kim, L.D. Kispert, 'Electrochemical Quartz Crystal Microbalance, Voltammetry, Spectroelectrochemical, and Microscopic Studies of Adsorption Behavior for (7*E*,7′*Z*)-Diphenyl-7,7′-diapocarotene Electrochemical Oxidation Product', *J. Phys. Chem. B*, **101**, 2038 (1997).
95. V.G. Bottom, *Introduction to Quartz Crystal Unit Design*, Van Nostrand Reinhold, New York, 1982.
96. W.P. Mason, *Electromechanical Transducers and Wave Filters*, Van Nostrand, New York, 1948.
97. W.G. Cady, *Piezoelectricity*, Dover, New York, 1964.
98. S.J. Martin, V.E. Granstaff, G.C. Frye, 'Characterization of a Quartz Microbalance with Simultaneous Mass and Liquid Loading', *Anal. Chem.*, **63**, 2272 (1991).
99. C.E. Reed, K.K. Kanazawa, J.H. Kaufman, 'Physical Description of a Viscoelastically Loaded AT-cut Quartz Resonator', *J. Appl. Phys.*, 68, 1993 (1990).
100. V.E. Granstaff, S.J. Martin, 'Characterization of a Thickness-shear Mode Quartz Resonator with Multiple Nonpiezoelectric Layers', *J. Appl. Phys.*, **75**, 1319 (1994).
101. K. Tozaki, M. Kimura, N. Komatsu, S. Itou, 'An Improved Method to Determine Complex Acoustic-impedance Using a Piezoelectric Resonator', *Jpn. J. Appl. Phys.*, **31**, 2592 (1992).
102. K.K. Kanazawa, ANTEC'90, 1049 (1990).
103. W.G. Cady, 'Piezoelectric Resonator and Effect of Electrode Spacing upon Frequency', *Physics*, **7**, 237 (1936).
104. T. Nomura, F. Tanaka, 'Frequency Characteristics of Piezoelectric Quartz Crystals Having a Separated Electrode in Liquids', *Bunseki Kagaku*, **39**, 773 (1990).
105. T. Nomura, T. Yanagihara, T. Mitsui, 'Electrode-separated Piezoelectric Quartz Crystal and its Application as a Detector for Liquid Chromatography', *Anal. Chim. Acta*, **248**, 329 (1991).
106. Z. Mo, L. Nie, S.J. Yao, 'A New Type of Piezoelectric Detector in Liquid. 1. Theoretical Considerations and Measurements of Resonance Behavior Dependent on Liquid Properties', *Electroanal. Chem.*, **316**, 79 (1991).
107. K. Takada, T. Tatsuma, N. Oyama, T. Nomura, 'Impedance and Frequency Characteristics of Electrode-separated Piezoelectric Quartz Crystals in Liquids', *Electroanal. Chem.*, **370**, 103 (1994).
108. N. Oyama, T. Tatsuma, S. Yamaguchi, M. Tsukahara, 'Scanning Electrode Quartz Crystal Analysis', *Anal. Chem.*, **69**, 1023 (1997).
109. J. Rishpon, A. Redondo, C. Derouin, S. Gottesfeld, 'Simultaneous Ellipsometric and Microgravimetric Measurements During the Electrochemical Growth of Polyaniline', *J. Electroanal. Chem.*, **294**, 73 (1990).
110. I. Rubinstein, J. Rishpon, E. Sabatini, A. Redondo, S. Gottesfeld, 'Morphology Control in Electrochemically Grown Conducting Polymer Films. 1. Precoating the Metal Substrate with an Organic Monolayer', *J. Am. Chem. Soc.*, **112**, 6135 (1990).
111. W. Stockel, R. Schumacher, 'In Situ Microweighing at the Metal/Electrolyte Junction', *Ber. Bunsenges. Phys. Chem.*, **345**, 91 (1987).

Neurotransmitters, Electrochemical Detection of

Thomas L. Colliver and Andrew G. Ewing
Pennsylvania State University, PA, USA

1 Introduction **1**
2 Overview of Electrochemical Methods **2**
2.1 Potentiometric Compared with Voltammetric Measurements 2
2.2 Electrode Types 3
3 In Vivo Voltammetry **4**
3.1 Introduction 4
3.2 Microdialysis Compared with In Vivo Voltammetry 4
3.3 Development of In Vivo Voltammetry to Detect Stimulated Neurotransmitter Release 4
3.4 Fast Cyclic Voltammetry and Electrical Stimulation: Applications In Vivo 7
4 Voltammetry in Brain Slices **8**
4.1 Introduction: Use of Brain Slices Compared with In Vivo Techniques 8
4.2 Technical Issues 8
4.3 Applications 8
5 Intracellular Voltammetry at Single Cells **10**
5.1 Overview 10
5.2 Monitoring Intracellular Dopamine 11
6 Extracellular Voltammetry at Single Cells **13**
6.1 Introduction 13
6.2 Overview of Systems Used for Single-cell Analysis 13
7 Concluding Remarks and Future Directions **20**
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

In this article, various electrochemical methods used in neuronal cell studies are examined. Electrochemical methods useful for investigating neuronal communication directly in the brain and in ex vivo brain slices are discussed. Additionally, we describe how voltammetric methods can be used to make intracellular and extracellular measurements at single cells. Technical aspects of each method are emphasized and selected applications are highlighted.

Although techniques such as electron microscopy and patch clamp can be used to visualize and monitor cellular events, electrochemical methods can provide chemical information about neurotransmitters being released and transported in brain tissue or even at single cells. The main competing technique for in vivo measurements of neurotransmitters is microdialysis. Although microdialysis coupled to separation techniques can be more selective and more sensitive than electrochemical methods, the latter methods are considerably faster and more accurate in determining the in vivo concentration. At single cells, electrochemistry provides extremely small probes compared to microdialysis thus facilitating measurements of single-cell events. Here, again, the response time associated with electrochemical methods is important for measurements of millisecond exocytosis events.

Electrochemical methods are clearly a powerful means to measure rapid changes in neurotransmitters that are electroactive and found in tissue or cellular microenvironments.

1 INTRODUCTION

The study of the brain and its function has been an area of research for many years. Analytical techniques have provided unique information regarding the neurophysiology and neuropharmacology of the brain. One technique, electroanalysis, has found extensive applications since many neurochemicals are easily oxidized or reduced. Electroanalytical techniques possess unique characteristics that make them useful for studying chemical processes occurring in complex biological matrices. Attributes that are particularly advantageous include the ability to provide qualitative and quantitative information for electroactive substances in the brain.

The fundamental building block of the brain is the single nerve cell or neuron.[1,2] It has been estimated that the human brain contains 10^{10} to 10^{12} neurons.[1] Nerve cells are unique in that they are specialized to communicate with each other. A typical neuron is comprised of a cell body, dendrites and an axon (Figure 1).[3] According to the classical model of neurotransmission, neuronal signals are received at the dendrites, integrated at the cell body, and then sent down the axon to the terminal where they are relayed to the next cell or cells. The site of information transfer between two cells is called the synapse. This structure is essentially a very small gap between cells which ranges in size from 1 to 100 nanometers.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

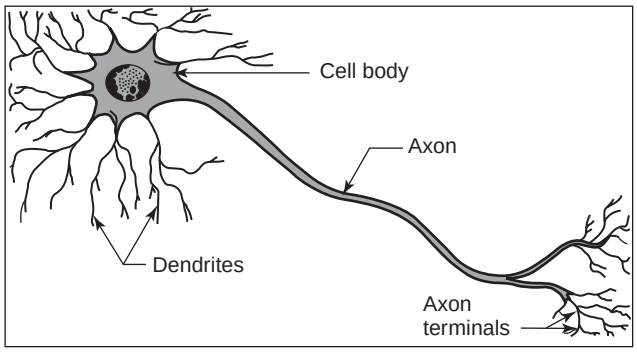

Figure 1 "Typical" neuron in mammalian brain. (Reproduced with permission from ref. 3. © 1976 American Chemical Society.)

Electron microscopy has revealed that storage granules or vesicles aggregate at the synapse near the cell membrane inside the presynaptic neuron (i.e. the neuron sending the signal). These storage compartments are where neurotransmitters are packaged, stored, and sometimes synthesized. When an electrical signal, called an action potential, travels down the axon and reaches the axon terminal, a series of events are initiated which cause the vesicles to transiently fuse with the cell membrane and expel their contents into the synapse. This process is termed exocytosis (Figure 2).[4,5]

After exocytosis, neurotransmitters diffuse across the synapse and bind to receptors on the postsynaptic cell. In some instances, the binding of neurotransmitters to these receptors causes a change in the permeability of the postsynaptic membrane to certain ions leading to the generation of a small postsynaptic potential. The magnitude of this potential is directly dependent on the amount of neurotransmitter that has been released, and can be either excitatory or inhibitory. At the cell body, postsynaptic potentials are integrated and the net sum of these potential changes determines whether the receiving neuron initiates a new action potential.

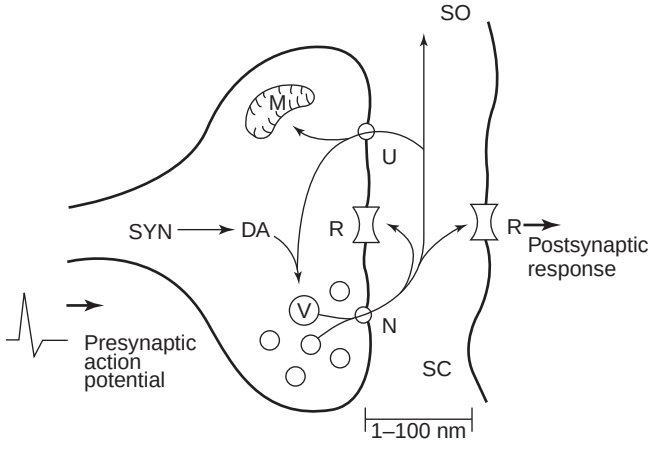

Figure 2 The terminal (or synaptic) region coupling two cells. (SYN = synthesis; DA = dopamine; V = dopamine storage vesicles; N = neuronal release; SC = synaptic cleft; R = pre- and postsynaptic dopamine receptors; U = dopamine uptake; SO = synaptic overflow; M = metabolic enzymes.) (Reproduced with permission from ref. 5. © 1988 American Chemical Society.)

Once released, the excess neurotransmitter needs to be deactivated to prevent continual stimulation of the postsynaptic cell. Deactivation usually occurs via three mechanisms: diffusion of the neurotransmitter out of the synapse, uptake of the neurotransmitter into the presynaptic cell (or surrounding glial cells), and/or metabolic breakdown of the neurotransmitter.

Neurons utilize a variety of different transmitters and in some cases an individual neuron can release more than one type of transmitter. The general classes of neurotransmitters include acetylcholine, catecholamines, indoleamines, amino acids, and peptides.[1] Since biogenic amines are electroactive, electroanalytical techniques have mainly been used in the study of these transmitters. In this category are the catecholamines; dopamine, norepinephrine, epinephrine, and the indoleamine serotonin.

In the 1970s, Adams et al. began to develop voltammetric methods to monitor dopamine in a dopamine-rich portion of the mammalian brain.[6] This work sparked interest in the use of electroanalysis in neuroscience-related research and has led to the continuing development of methods to monitor neurotransmitter action more effectively in neuronal microenvironments. Throughout the course of this article, various electrochemical methods used in neuronal cell studies will be examined. Specifically, the development of electrochemical techniques useful for investigating neuronal communication directly in the brain and in ex vivo brain slices are discussed. Additionally, the final two sections will examine voltammetric methods used for intracellular and extracellular measurements at single cells, both in vivo and in vitro. At the end of relevant sections, selected applications of these methods will be highlighted.

2 OVERVIEW OF ELECTROCHEMICAL METHODS

2.1 Potentiometric Compared with Voltammetric Measurements

Potentiometry and voltammetry are two of the basic electroanalysis methods available for chemical analysis. These are broad categories that include many different techniques developed for use in different experimental situations. Potentiometric ion-selective electrodes (ISEs)[7] are commonly used to measure concentrations of ions both outside and inside cells. The most common ISEs

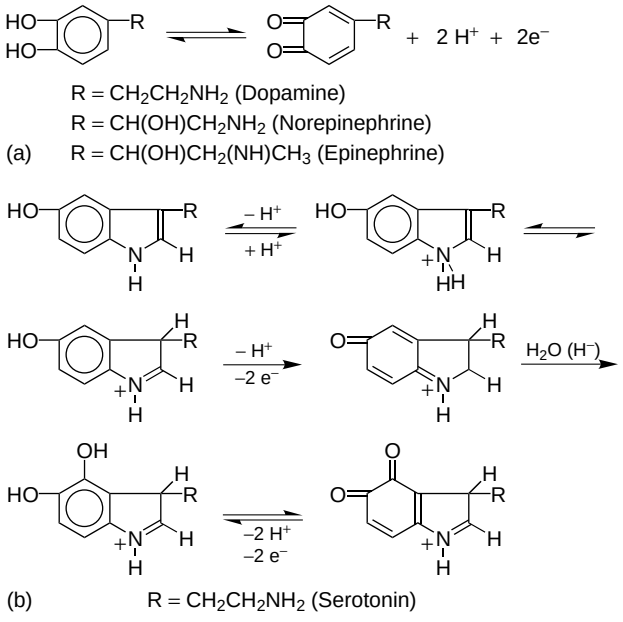

Figure 3 Oxidation reactions for (a) catecholamines and (b) indoleamines. (Reproduced from ref. 104, 225. © 1998 with permission from Marcel Dekker Inc.)

are used to measure pH, K^+, Na^+, or Ca^{2+}[8] levels in resting and stimulated cells.[9,10] These types of electrodes have been used to study diffusion characteristics as well as to determine the volume of the extracellular space in the brain.[11,12] Although potentiometric methods can be used to study ion fluxes and concentrations, these techniques are not capable of detecting neurotransmitters at neuronal cells. For these types of measurements voltammetry must be employed.

Several neuroactive species are easily oxidized (Figure 3) and are therefore amenable to detection by voltammetry. The current created by the oxidation of these species in response to an applied voltage is the common operating principle behind the various voltammetric techniques. Voltammetric techniques can be put into one of two broad categories: potential step (or pulse) and continuously applied potentials. The pulse techniques include chronoamperometry, normal pulse voltammetry, and differential pulse voltammetry, while linear sweep voltammetry, cyclic voltammetry, and amperometry can be categorized as continuous potential methods. A comparison of the various electrochemical methods reveals that each method has specific advantages and disadvantages in particular protocols. Generally, the voltammetric method used depends on the needs of the user.

2.2 Electrode Types

Platinum wire and carbon fibers are just two of the many materials that have been used in the construction of microelectrodes. Both platinum and carbon can be fabricated into durable electrodes with micron or smaller size tips and used in minute environments.[13,14] Platinum and carbon have also been combined to construct platinized carbon microelectrodes.[15–17] Selectivity, sensitivity, stability, and reproducibility of the signals are the major differences between these two types of electrodes, especially when working in biological environments such as the brain or with cells in culture. The electrode material of choice for a given experiment depends upon the nature of the substance(s) being investigated.

2.2.1 Micro Platinum Electrodes

Platinum possesses partially unsaturated surface d-orbitals which facilitate the adsorption stabilization of free radical products in slow electrooxidation reactions and are beneficial in the detection of aliphatic alcohols and amines. Platinum electrodes are also very sensitive to the presence of hydrogen and oxygen.[15] This is advantageous if one wishes to detect H_2 or O_2, but is detrimental if these gases are background signals that obscure detection of the analyte. The strong adsorption characteristics of Pt makes it useful in a "clean" environment, but much less so in biological matrices where proteins and other biological molecules can adsorb to the surface. Ultimately, electrode fouling decreases the sensitivity, stability, and reproducibility of the measurements. These issues have been addressed by depositing thin layers of platinum on carbon surfaces as well as by coating the platinum with various permselective membranes. Although these modifications have improved the performance of platinum electrodes, there is still room for further improvements.

2.2.2 Carbon-fiber Electrodes

Owing to their resistance to drift[18] and inert nature in biological environments, carbon fibers have been used to develop carbon-fiber microelectrodes with dimensions in the micrometer range. The small size of these electrodes provides many advantages in terms of electrochemistry and the practical aspects of their use in biological environments. The improved faradaic to nonfaradaic current ratio at microelectrodes compared to macroelectrodes, for example, enhances the signal-to-noise ratio.[19] Additionally, currents generated at microelectrodes typically lie in the picoampere to nanoampere (pA to nA) range. Such small currents facilitate the use of simpler two-electrode configurations in voltammetric studies.[20] In terms of the biological environment, the smaller currents also mean fewer products are generated at the electrode. This is beneficial since many of these products are detrimental to cells at high concentrations.[21] Small currents

created at microelectrodes are not as disruptive to normal cellular function and are less likely to interfere with the normal activity of the surrounding tissue or cells.[22]

Like platinum, some adsorption of large biomolecules does occur on carbon-fiber electrodes[23] causing the electrode's response to deteriorate over time. Usually there is a 30–50% decrease in electrode sensitivity following implantation into brain tissue. This decrease levels off after approximately two hours and reproducible, stable measurements can be collected for six to twelve hours.[21,24,25,26] In culture, sensitivity loss is much less.[18] In both instances, the decrease in electrode sensitivity requires that electrodes be calibrated after experiments.

3 IN VIVO VOLTAMMETRY

3.1 Introduction

The brain is a chemically heterogeneous organ with a fluid environment containing various chemical substances at different concentrations that can change on the second to millisecond timescale. An ideal technology to investigate the neurochemistry of the brain should be able to sample from spatially discrete locations without perturbing the local neuronal environment and have the sensitivity and selectivity to detect and discriminate between chemicals that are present at a wide range of concentrations. This technology should also be able to make measurements on a timescale similar to that on which neuronal signaling occurs. Over the last three decades, two methods have been developed which are complementary to each other and collectively satisfy these requirements. These methods are microdialysis and in vivo voltammetry.

3.2 Microdialysis Compared with In Vivo Voltammetry

In vivo voltammetry and microdialysis sampling have played a large role in detecting neurotransmitter release and turnover. For both these techniques, stereotaxic coordinates are used to implant the sensors into a specific region of the brain. Probes for microdialysis consist of two concentric tubes. The tips of these probes are covered with a low-molecular-weight cut-off dialysis membrane. Small molecules are removed from the extracellular fluid (ECF) by diffusing across this membrane. Fluid containing the analytes is pumped through the probe and out of the brain for subsequent analysis. High-performance liquid chromatography (HPLC) can be used to simultaneously detect many analytes in the dialysis samples. The combination of a sample volume requirement of microliters for HPLC and slow diffusion of molecules across the dialysis membrane results in sampling rates that are generally five to ten minutes, although this has been pushed into the range of a few seconds in recent years.[27,28]

Voltammetry permits measurements to be made more rapidly by applying a small potential to an electrode and measuring the current resulting from the oxidation of electroactive species. In contrast to microdialysis, voltammetric techniques analyze the sample at the surface of the probe. The rate at which molecules can be detected is limited only by diffusion and the response time of the electrode. Another distinction between in vivo voltammetry and microdialysis is the small size of the recording electrodes used in voltammetry. Most electrodes used in vivo have a 10–20 μm total tip diameter. This is approximately ten times smaller than the outer diameter of a microdialysis probe. The size of both types of probes is large compared to cells and synapses. As a result, measurements obtained from both of these techniques reflect the dynamics of chemical events occurring in the extracellular space of the brain. However, the smaller electrochemical probes make it possible to get closer to the synaptic region while minimizing damage to the surrounding brain tissue.

The fast acquisition rates and small size of voltammetric probes allow more rapid and spatially discrete measurements of chemical dynamics. Although microdialysis, in contrast, is better suited to monitor concentration changes for a wider dynamic range of compounds at low levels over prolonged periods of time, the temporal and spatial resolution of this technique are not as good as for voltammetry at microelectrodes.

Since its inception, in vivo voltammetry has been used to measure the dynamic changes of many different substances in the brain that could be considered ''informational''. One goal which has driven the development of this technology has been to measure directly, in real time, changes in the concentration of neurotransmitter(s) that result from neural activity. Throughout this section of the article, key developments that have helped to push this technology toward this goal will be highlighted and findings from select studies using in vivo voltammetry will be discussed.

3.3 Development of In Vivo Voltammetry to Detect Stimulated Neurotransmitter Release

3.3.1 Selectivity of In Vivo Voltammetry

In vivo voltammetry is a general term used to describe a group of loosely related electrochemical techniques developed to measure electroactive compounds in the brain. These techniques can vary with the type of scanning method (i.e. pulse or continuous), the rate at which they scan, and the type of electrode used. As a result of these different experimental approaches, each technique varies in selectivity, rate of data acquisition, and sensitivity.

Although it is not the intent here to compare the merits of every technique, (this has been done elsewhere)[21,29] in general, the voltammetric method used depends on the needs of the user. Despite the varied approaches, one of the main disadvantages of using voltammetry is its limited selectivity. The lack of chemical selectivity in voltammetric techniques results mainly from the fact that electrodes respond to any substance that can be electrolyzed at the applied potential.

Initial studies reporting the voltammetric detection of neurotransmitters in the rat brain attributed the increase in oxidation current following intraperitoneal (i.p.) amphetamine injection to an increase in the catecholamine dopamine.[30,31] Gonon et al. later exploited the ability of electrochemically modified electrodes to clearly resolve ascorbic acid and dihydroxyphenylacetic acid (DOPAC) to provide evidence that ascorbic acid, not dopamine, was responsible for the increased electrochemical signal observed following administration of amphetamine.[32] Using an electrode that could discriminate between ascorbic acid, dopamine, and DOPAC, investigators from Wightman's laboratory also concluded that ascorbic acid increased in the ECF of the brain following amphetamine injection.[33,34] The conclusions from these in vivo voltammetry studies were later independently confirmed using push–pull perfusion and microdialysis methods. Using the superior selectivity of HPLC, Salamone et al. showed that ascorbic acid increases in concentration in response to amphetamine.[35] It should be noted that further technical developments did make it possible to detect the amphetamine-induced release of dopamine using voltammetry (see below). Additionally, voltammetry continues to be a very useful method for following changes in ECF levels of ascorbic acid in vivo.[36,37]

Ascorbic acid and DOPAC are both electroactive and are always present in ECF at concentrations 10^4 to 10^3 greater than biogenic amines. Detecting these neurotransmitters in the presence of ascorbic acid and metabolites such as DOPAC therefore remains an analytical challenge. Another way to enhance the selectivity of voltammetric electrodes is to coat the electrode with a permeation selective film. Nafion™ is a sulphonated polymer that is selectively permeable to cations, but repels anions. When coated onto the tip of an electrode, Nafion™ permits cations such as dopamine and serotonin to reach the electrode surface, but rejects anions such as ascorbic acid and DOPAC. Thus, the development of electrodes coated with Nafion™[38] has been a critical component of selective measurements with in vivo voltammetry.

3.3.2 *Fast Compared with Slow Voltammetric Methods*

By utilizing the above techniques to improve selectivity, several investigators have been able to successfully demonstrate that neurotransmitters can be directly detected in the brain using voltammetry. However, exocytosis occurs on a millisecond timescale. In these early examples, data was acquired on the second to minute timescale.[32,33,39] Therefore, these relatively slower voltammetric methods were not capable of detecting faster signals that could have been present. The merits of fast and slow in vivo voltammetric methods have been reviewed elsewhere.[40] In general, slower voltammetric methods, and microdialysis, are best suited to monitor slow changes in concentration of metabolites and related species that result from pharmacological or behavioral stimuli.[41] However, these types of changes are slow and long lasting (minutes to hours) compared to release processes and do not require that high-resolution sampling methods be used. Furthermore, changes in metabolite levels do not reliably reflect release of the parent neurotransmitter.[42,43]

To accurately measure neurotransmitter dynamics following direct rapid stimulation of nerve terminals, voltammetric methods need to acquire data on a sub-second timescale. Owing to their high rate of data acquisition, chronoamperometry and fast cyclic voltammetry (FCV) are both well suited to measure the acute kinetic characteristics of release processes in ''real time''. Before the advent of very small electrodes, which facilitated faster scanning methods, the most popular voltammetric methods for rapid electrochemical measurements were potential step techniques such as chronoamperometry.[5] For typical chronoamperometry, data acquisition usually requires 50 to 100 ms. Compared to pulse techniques, however, scanning methods provide a higher degree of selectivity and are more effective at discriminating possible interferents (see below). Data for an entire FCV can be acquired in about 10 to 15 ms using a scan rate of $300\,V\,s^{-1}$. Since small carbon-fiber microelectrodes can be used in combination with FCV, this method provides excellent temporal and spatial resolution. Consequently, FCV has been the most widely utilized electrochemical technique for in vivo voltammetry. In the following sections, issues related to the selectivity and sensitivity of in vivo voltammetry will be discussed with an emphasis on FCV.

3.3.3 *Selectivity of Fast Cyclic Voltammetry*

There are two ways FCV provides selectivity. FCV at microelectrodes is dominated by double-layer charging current which masks the faradaic current arising from the oxidation–reduction of electroactive neurotransmitters. This background current can be digitally subtracted from the faradaic current to give a ''background subtracted'' voltammogram. During periods between stimulation, the majority of the FCV signal is presumably due to the

detection of ascorbic acid. Since individual fast cyclic voltammograms are obtained on a timescale similar to release processes, following neuronal stimulation, neurotransmitters are the most likely compounds to change during the time of each scan. Metabolite and ascorbic acid concentrations should change on a much slower timescale.

Generally, the shape of background corrected voltammograms vary with different chemical species. At the fast scan rates used in FCV, voltammogram shapes are kinetically controlled. As a result, most electroactive analytes in the ECF have a unique ''fingerprint'' which can be used to tentatively identify what has changed in concentration during the time course of the scan. Unfortunately, qualitative chemical analysis cannot be solely based on the shape of the voltammogram. The neurotransmitters serotonin, epinephrine, and dopamine for example, all oxidize at potentials between +0.2 and 0.5 V against Ag–AgCl with untreated electrodes, and the shape of the signal from these compounds is not always readily distinguishable. Additionally, interpreting the shape of the voltammogram can become complicated if more than one neurotransmitter changes in concentration. When using FCV to detect electroactive neurotransmitters in the brain, it is essential that independent neuropharmacological interventions be used to verify the identity of the detected neurotransmitter.[41,44]

3.3.4 Sensitivity of In Vivo Voltammetry

Realizing the limited sensitivity of voltammetric methods in general, Ewing et al.[33] used elevated levels of K^+ to stimulate massive amounts of dopamine release (Figure 4). This method of stimulation increased ECF levels of dopamine from basal levels of approximately 5–10 nanomolar to the low micromolar range. Although a relatively slower method of voltammetry (normal pulse voltammetry) was used, these results demonstrated that, with an adequate stimulus, voltammetry could be used to monitor the stimulated release of dopamine.

Owing to the limited spatial and temporal control provided by elevated levels of K^+, this method of stimulation was later replaced by the use of a bipolar stimulating electrode.[45] Stimulating electrodes can be used to artificially trigger action potentials within a specific group of neurons. Stimulating and recording electrodes can be placed in different locations in the brain of an animal during an experiment. For example, with the working electrode in the terminal field where release occurs, the stimulating electrode could be placed in the cell body region, along ascending fiber tracts, or directly in the terminal fields to stimulate release. Despite these options, most workers stimulate ascending pathways. One approach that has been used extensively is to electrically stimulate the medial forebrain bundle (MFB) while detecting dopamine concentration changes in the rat striatum with a carbon-fiber electrode.

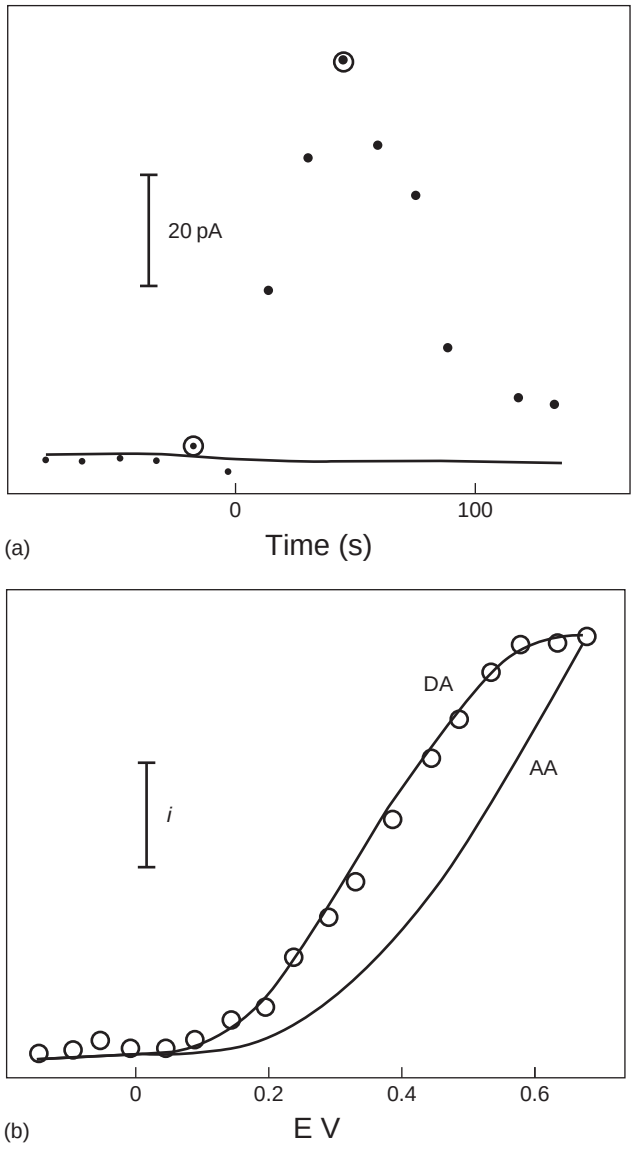

Figure 4 Electrochemical response to injection at 0 time of 0.6 μmol of potassium chloride in 1.0 μL of physiological buffer into the caudate nucleus. Distance from electrode to syringe tip is 0.8 mm. (a) Current vs time response at 0.5 V. Solid line, injection of physiological buffer only; points, injection of potassium chloride. (b) Difference voltammograms. Solid lines, dopamine (25 μM) and ascorbic acid (200 μM) in pH 7.4 buffer after in vivo use; circles, in vivo result from subtracting voltammograms obtained at the circles shown in (a). Current scales: $i = 25$ pA, dopamine; $i = 28$ pA, ascorbic acid; and $i = 22$ pA, in vivo. (Reproduced from Brain Research, A.G. Ewing, R.M. Wightman, M.A. Dayton 'In vivo Voltammetry with Electrodes that Discriminate Between Dopamine and Ascorbate', **249**, 361–370. © 1982 with permission from Elsevier Science.)

3.4 Fast Cyclic Voltammetry and Electrical Stimulation: Applications In Vivo

Combined, the use of local electrical stimulation and FCV make it possible to directly measure, in real time, changes in the concentration of neurotransmitter(s) that result from neural activity. Stamford et al.[46] were the first to realize the advantages of combining these two techniques and, since then, FCV and electrical stimulation have been the most widely used methods for examining the dynamics of electroactive neurotransmitters in vivo and in vitro (see below). Detailed and exhaustive discussion of the neurochemistry examined in all of these studies is beyond the scope of this article. However, selected findings are discussed here to illustrate the utility of this methodology.

Kuhr and Wightman[47] have used fast-scan voltammetry at Nafion™-coated, beveled, carbon-disk electrodes with electrical stimulation of the rat MFB to monitor the time course of stimulated dopamine release in the caudate nucleus. The maximum concentration of dopamine detected at a single electrode location during stimulation has been found to be directly proportional to the frequency of pulses in the stimulation train. Dopamine is observed in the ECF for approximately 1.5 s when short stimulation times are employed demonstrating the brief time that detectable dopamine levels are present following stimulation. The disappearance of dopamine appears to occur exponentially, which is characteristic of first-order uptake kinetics. Using the uptake inhibitor, nomifensine, a large increase in current is observed when cells are stimulated with a small electrical stimulus (15 pulses in 250 ms), which normally gives only a barely detectable signal (Figure 5). These results confirm the large effect that uptake has on the detected dopamine signal in vivo.

In another example, in vivo voltammetry has been used to suggest that the brain is not "hard-wired" meaning that some neurons communicate by pathways outside of the synapse. Garris et al.[48] have used FCV with Nafion™-coated, beveled, disk electrodes and electrical stimulation to examine dopamine efflux from the synaptic junction. It has been determined that binding of released dopamine to receptors and uptake sites does not appreciably alter the diffusion of dopamine from the synaptic cleft in the rat nucleus accumbens. This implies that the dopamine synapse is designed for efficient transmitter efflux from the synaptic cleft so that dopamine can escape the synapse and affect a greater number of cells. This proposed scheme contrasts with the established model of neurotransmission derived from experiments at the neuromuscular junction where communication occurs only inside the synapse (Figure 6).

The simultaneous detection of two neurochemically relevant substances in vivo has also been accomplished using FCV. Zimmerman and Wightman[49] have demonstrated that both dopamine and O_2 can be independently and simultaneously monitored at a single

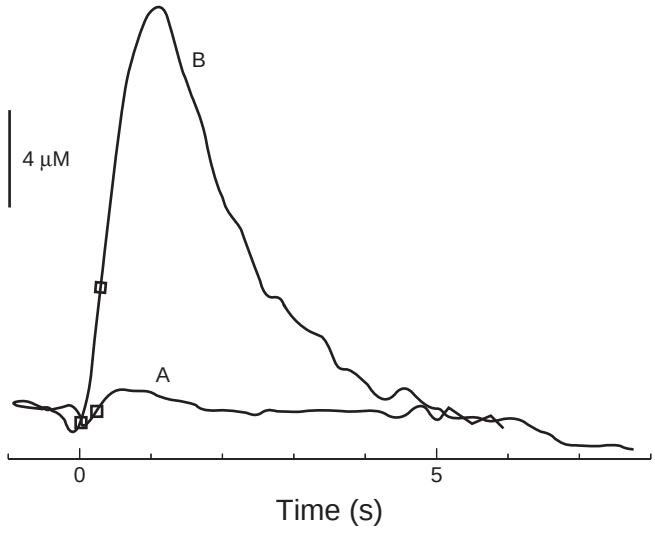

Figure 5 Time course of dopamine release before (a), and 8 min after (b), administration of 20 mg kg^{-1} nomifensine. The release was induced by stimulation consisting of fifteen 300-μA bi-phasic pulses delivered at a frequency of 60 Hz. Although only one time course is shown, identical temporal profiles were obtained in 3 additional rats. (Reproduced from Brain Research, W.G. Kuhr, R.M. Wightman, 'Real-time Measurement of Dopamine Release in Rat Brain', **381**, 168–171. © 1986 with permission from Elsevier Science.)

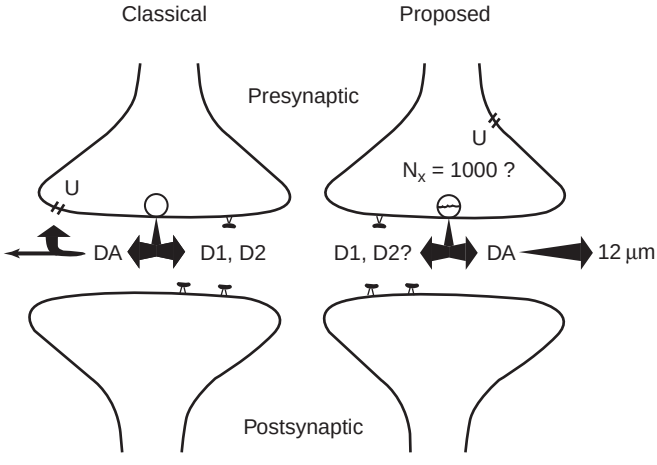

Figure 6 Schematic drawings of dopamine neurotransmission in the core of the nucleus accumbens. Schematics contrast the classical and proposed model of dopamine efflux from dopamine synapses in the nucleus accumbens of the rat brain. For the proposed model, it is suggested that the dopamine transporter (DAT) (U) mainly functions to regulate extrasynaptic dopamine levels as opposed to preventing the efflux of transmitter from the synaptic cleft. Note the difference in the size of the arrows representing the efflux of dopamine from the synaptic cleft. (Reproduced from *J. Neuroscience*, **14**(10). © 1994 with permission.)

Nafion™-coated disk electrode by scanning from 0.0 V to +0.8 V followed by a cathodic sweep to −1.4 V before returning to a resting potential of 0.0 V. Several laboratories have also used voltammetry at carbon-fiber electrodes, in combination with other techniques, to simultaneously acquire electrochemical and electrophysiological information in vivo and in vitro.[22,50,51] In one particular study, Ewing et al. used carbon-fiber microelectrodes combined with tungsten microelectrodes to simultaneously measure the effects of *d*-amphetamine and ascorbic acid on the electrochemical and unit activity of neostriatal neurons.[22] Neurotransmitter efflux and unit activity may be considered respectively as indices of presynaptic and postsynaptic neuronal function. The ability to record both of these phenomena makes it possible to correlate the release of neurotransmitter from a discrete population of release sites with changes in neuronal activity at a specific region in the brain.[51]

4 VOLTAMMETRY IN BRAIN SLICES

4.1 Introduction: Use of Brain Slices Compared with In Vivo Techniques

The first report of voltammetric techniques being used in combination with brain slices described experiments that were designed to shed light on data obtained in vivo.[52] Owing to the many technical advantages of using this approach, slice-voltammetry methods have become very popular and have nicely complemented work done in vivo. For these types of experiments, voltammetry is performed in slices of brain tissue that are created by taking ''sections'' from a whole brain. This approach has several advantages that make it attractive. It avoids any potential confounding effects that may arise from the use of anesthetics and inherently reduces/simplifies the complexities of the neuronal circuits present during an experiment. Additionally, when using brain slices one does not need to worry about the ability of a given agent to cross the blood brain barrier or the possible toxic side effects that an agent may have on the whole animal. Combined, these two factors widen the range of agents which can be investigated. Since the investigator can control the composition of the fluid perfusing the tissue, it is also possible to minimize potential interferents such as ascorbic acid.

4.2 Technical Issues

4.2.1 Slice Preparation

In general, when slices of brain tissue are prepared for an experiment, the brain is quickly removed from an animal and immediately placed in chilled physiological saline. A block of tissue containing the desired region of the brain is then isolated from the whole brain and slices of tissue are created by sectioning this block using a vibratome. A slice is then placed into a tissue bath where the tissue has complete access to the physiological buffer. In the bath, the tissue is continuously superfused with warm physiological buffer saturated with oxygen and carbon dioxide (32–37 °C; some investigators have reported that the temperature of the slice affects the long-term stability of the electrochemical signal[51]). Slices are usually cut to a thickness of 150–400 μm and the physiological buffer is flowed through the bath at a rate of approximately 0.25–1 mL min^{-1}. The dimensions of the slice allow oxygen to easily diffuse into the tissue and the constant exposure to fresh buffer helps to establish a metabolically stable system.[53] Once situated within the tissue bath, the slice is allowed to incubate for 30 to 60 minutes prior to an experiment. This allows the preparation to stabilize and equilibrate with its surrounding environment.

4.2.2 Experimental Set-up

With the aid of a dissecting microscope, the working electrode is lowered into the slice, using a micromanipulator, to a position of 75 to 100 μm below the surface of the tissue. This ensures that the whole sensing area of the electrode is completely enclosed within the slice. Auxiliary and reference electrodes are placed in convenient positions within the tissue bath. While solutions of elevated K^+ have been used to stimulate neurotransmitter release in slice-voltammetry experiments,[54] as with experiments in vivo, the most popular method used in slices has been electrical stimulation. This type of stimulation is accomplished by placing a bipolar stimulating electrode in or on the slice approximately 100 to 400 μm away from the working electrode. The stimulating electrodes can be accurately placed, with the aid of the microscope, at a specific region within the slice. Interestingly, the relative efficacy of this method of stimulation varies with the area of the brain being studied. Single pulse stimulations (0.1 ms width, 20 V) for example, consistently elicit detectable levels of dopamine in slices of rat striatum[55] while trains of stimuli (i.e. 25 pulses, 50 Hz, 0.1 ms, 20 V) are required to give similar voltammetric signals for serotonin in slices of rat dorsal raphe or suprachiasmatic nuclei.[56] For a more detailed discussion of the technical issues concerning the use of voltammetry in brain slices, the reader is referred to an excellent review by Stamford et al.[57]

4.3 Applications

Although FCV is not the only voltammetric technique that has been used with slices,[58,59] it is the method

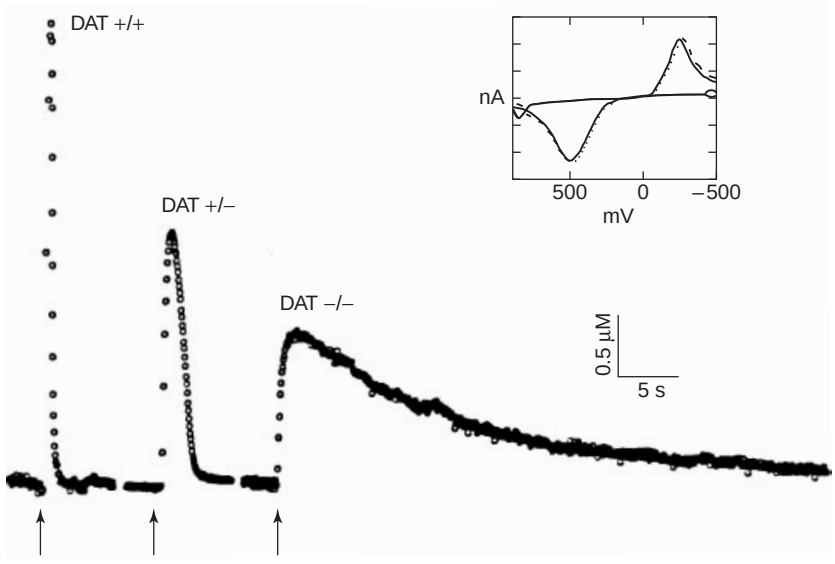

Figure 7 Typical recordings of electrically stimulated dopamine efflux in striatal slices from DAT +/+, DAT +/– and DAT –/– mice. DAT +/+ represents mice with both alleles of the DAT; DAT +/– represents mice missing one allele; DAT –/– represents mice missing both alleles. A single biphasic electrical pulse (4 ms, 350 μA) was applied at the time indicated by the arrows with a locally placed bipolar electrode. Data points (open circles) were collected from a recording electrode every 100 ms. Inset shows cyclic voltammograms recorded at the peak of dopamine efflux in each slice. Voltammograms recorded from DAT +/+ (solid line), DAT +/– (broken line) and DAT –/– (dotted line) mice were scaled to approximately the same peak amplitude for comparison. These cyclic voltammograms are not significantly different from those obtained with authentic dopamine. Measurements were made in at least 22 slices from at least 5 animals for each group. The average time to clear released dopamine was 1, 3, and 100 s in DAT +/+, DAT +/– and DAT –/–, respectively. (Reproduced from ref. 67. © 1996 with permission from *Nature*.)

that has been most widely applied to these preparations. Since voltammetry was first used in slices, there have been a variety of experiments performed using different voltammetric techniques. The focus of the majority of these studies can be loosely categorized as: examining neurotransmitter release and uptake in different regions of the brain,[55,56,60] comparing neurotransmitter release and uptake between different regions of the brain,[61,62] and investigating the presynaptic regulation of neurotransmitter release by pharmacologically manipulating autoreceptors.[63,64] Recent studies have even used voltammetry in slices to better understand the physiological mechanisms regulating ischaemia-induced dopamine release.[65,66] Since it is beyond the scope of this article to cover in detail all the findings from these studies, we have chosen to illustrate the utility of using voltammetry in slices by focusing on work recently done in brain slices from mice lacking the DAT. The DAT is associated with presynaptic dopamine neurons and mediates uptake processes to clear dopamine from the synaptic cleft (see Figure 2).

The central importance of cellular uptake mechanisms for regulating dopamine levels in the ECF has been investigated by disrupting the mouse DAT gene. Voltammetry experiments in brain slices from these ''knockout'' mice have shown that disruption of the DAT significantly increases the time required for dopamine to be cleared from the ECF. Indeed, when both alleles of the DAT are disrupted, dopamine remains in the ECF up to 100 times longer than normal (Figure 7). These results directly demonstrate that the DAT is critically involved in regulating dopamine ECF levels, and hence dopamine signaling, and provide a biochemical basis for the spontaneous hyperactivity observed in DAT knockout mice.[67]

Subsequent experiments in slices from DAT knockout mice have taken advantage of the unique genetic background of these animals to help unravel the complex actions of the psychostimulant amphetamine.[68] Amphetamine can cause the release of dopamine from neurons in a manner which is impulse independent and has little calcium dependence. The neurochemical mechanism responsible for the releasing action of amphetamine is attributed to both the drug's ability to redistribute dopamine from secretory vesicles and its ability to cause the release of cytoplasmic dopamine into the ECF via the DAT (a process that has been called reverse transport[69]). The relative importance of these two mechanisms in the releasing action of amphetamine, however, has been controversial.

Figure 8 shows how amphetamine affects the electrically stimulated and baseline levels of dopamine when applied to striatal slices from wild-type (DAT +/+,

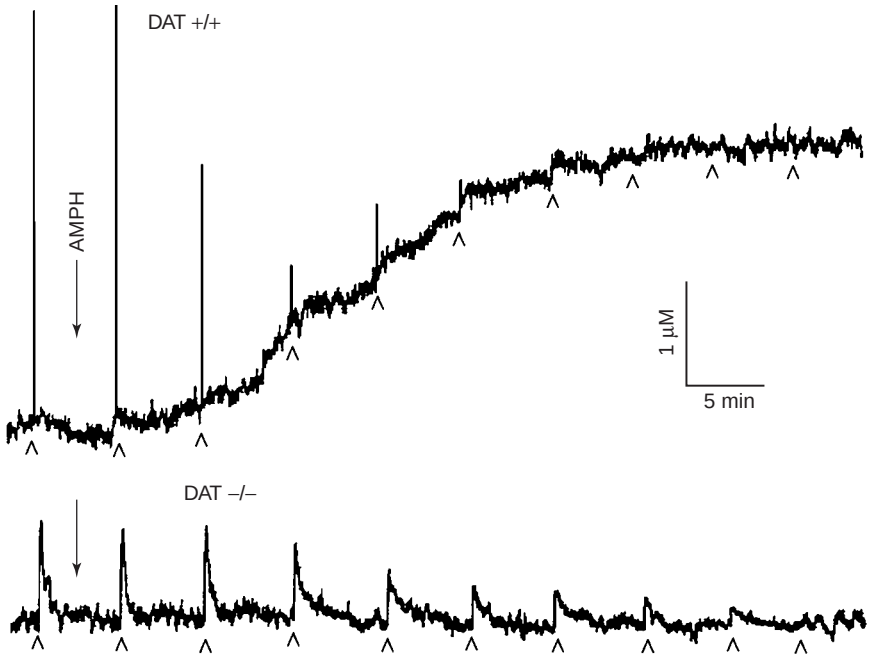

Figure 8 Effect of amphetamine (10 μM) on dopamine efflux in striatal slices from wild-type (DAT +/+) and homozygote DAT knockout (DAT −/−) mice. Current was measured by cyclic voltammetry at microelectrodes implanted in the slices. Stimulated dopamine release was elicited by single electrical pulses applied to the slice at the times indicated by small arrowheads. On the timescale shown, electrically stimulated dopamine release events appear as sharp spikes. Baseline release was monitored as any increase in the current measured between stimulations that was identified as dopamine. In the absence of pharmacological or electrical intervention, baseline and electrically stimulated dopamine recordings were stable for >3 h. (Reproduced from *J. Neuroscience*, **18**(16). © 1998 with permission.)

control) and DAT knockout (DAT −/−) mice. As shown in Figure 8, approximately 10 min after amphetamine is applied to slices from wild-type mice there is a distinct rise in baseline dopamine levels. At the same time, as can be seen by the decrease in height of the sharp 'spikes' of dopamine efflux, amphetamine also decreases the amount of electrically stimulated dopamine release. Approximately 25 min after the drug has been administered, the steady increase in the baseline reaches a plateau. In contrast to these results, when amphetamine is applied to slices from DAT −/− mice, although there is also a decrease in the stimulated efflux of dopamine, no change in baseline dopamine levels are observed. These results directly demonstrate that the DAT is required for the releasing action of amphetamine.

To examine vesicle depletion and its effect on reverse transport independently, slices from control and DAT deficient mice were exposed to the vesicle depleting agent Ro4-1284. As can be seen in Figure 9, this drug, like amphetamine, caused a gradual reduction in the stimulated release of dopamine. Unlike amphetamine however, Ro4-1284 did not cause an increase in baseline dopamine levels. When amphetamine was applied to slices from wild-type mice, after Ro4-1284 had completely blocked stimulated dopamine release, there was a rapid increase in baseline dopamine levels that reached values similar to those observed when amphetamine was used alone. The rate of this increase, however, was much faster with baseline concentrations of dopamine reaching their maximum values 5 to 10 min after the application of amphetamine.

From these two sets of experiments it has been estimated that it takes approximately 25 min for amphetamine to exert its depleting action on synaptic vesicles and only 5 min for amphetamine to release cytoplasmic dopamine via reverse transport. These results demonstrate that both vesicular depletion and reverse transport play central roles in the releasing action of amphetamine. However, vesicle depletion appears to be the rate-limiting step.

5 INTRACELLULAR VOLTAMMETRY AT SINGLE CELLS

5.1 Overview

The ability to detect species inside single nerve cells is very important in order to gain further insight into the metabolism, function and regulation of individual

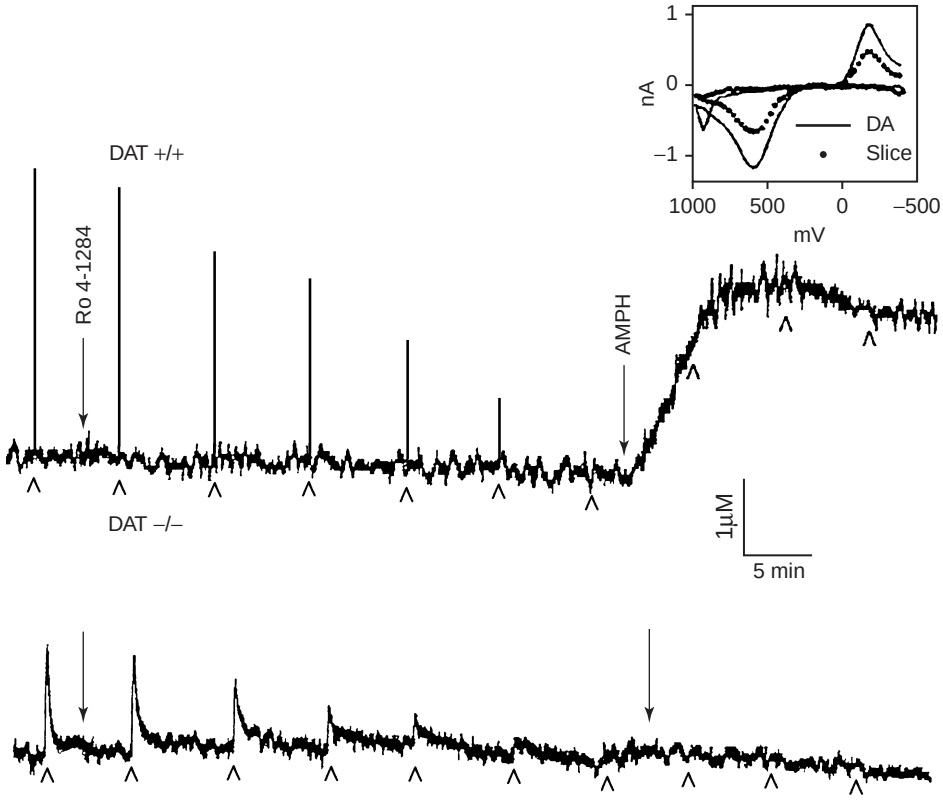

Figure 9 Effect of a fast-acting reserpine-like drug, Ro4-1284 (10 μM), and amphetamine (104 μM) on stimulated and baseline release of dopamine. Individual recordings of dopamine efflux were measured by cyclic voltammetry in striatal slices from wild-type (DAT +/+) and homozygote DAT knockout mice (DAT −/−). Drugs were applied to slices at the times indicated by the large arrows. Top, Single-pulse stimulations were applied to a slice from a wild-type mouse at the times indicated by the small arrowheads. Inset, Cyclic voltammograms, current (nA) versus potential (mV) plots, recorded during postcalibration with authentic dopamine (solid line) and during the plateau phase of the release induced by amphetamine (filled circles), are plotted for comparison. Bottom, Single-pulse stimulations were applied to a slice from a homozygote DAT knockout mouse at the times indicated by the small arrowheads. (Reproduced from *J. Neuroscience*, **18**(16). © 1998 with permission.)

neurons. Electrode tips 10 μm or less in diameter are small enough to impale fairly large cells (100–200 μm diam.) without causing excessive damage or interfering with cellular function. The development of carbon and platinum microelectrodes with total overall dimensions in the micron range has made it possible to directly detect several different electroactive substances in the cytoplasm of single nerve cells.[14,70,71]

Initial intracellular voltammetric studies, conducted by Meulmans et al.,[14] to measure the endogenous ascorbic acid levels in single *Aplysia* neurons, demonstrated the potential of electrochemical methods to intracellularly monitor electroactive neurochemicals. Using 0.5- to 2-μm tip diameter carbon-fiber or platinum electrodes in combination with differential pulse voltammetry, it has been determined that the endogenous ascorbic acid concentration in these single neurons is approximately 100 μM. Similar approaches have also been used to monitor the kinetics of uptake, clearance of two electroactive drugs,[72] and levels of serotonin in single *Aplysia* neurons.[14]

Development of glucose microsensors has provided the ability to qualitatively monitor glucose transients in the cytoplasm of a single cell. Glucose metabolism is of particular interest because glucose is a major energy source for the brain.[16] Abe et al.[73,74] have used 2-μm diameter platinized-carbon ring electrodes with amperometric detection (0.6 V versus SSCE (sodium saturated calomel electrode)) to monitor intentional manipulations of cytoplasmic glucose in the giant dopamine cell of the pond snail *Planorbis corneus*. Intracellular oxygen levels in the giant dopamine cell of *Planorbis* have also been measured using Nafion™-coated platinized-carbon ring electrodes with differential pulse voltammetry.[16] In the following section, experiments investigating intracellular dopamine in the giant dopamine cell of *Planorbis* will be described.

5.2 Monitoring Intracellular Dopamine

Carbon and platinum microelectrodes and various electrochemical techniques have been used to monitor

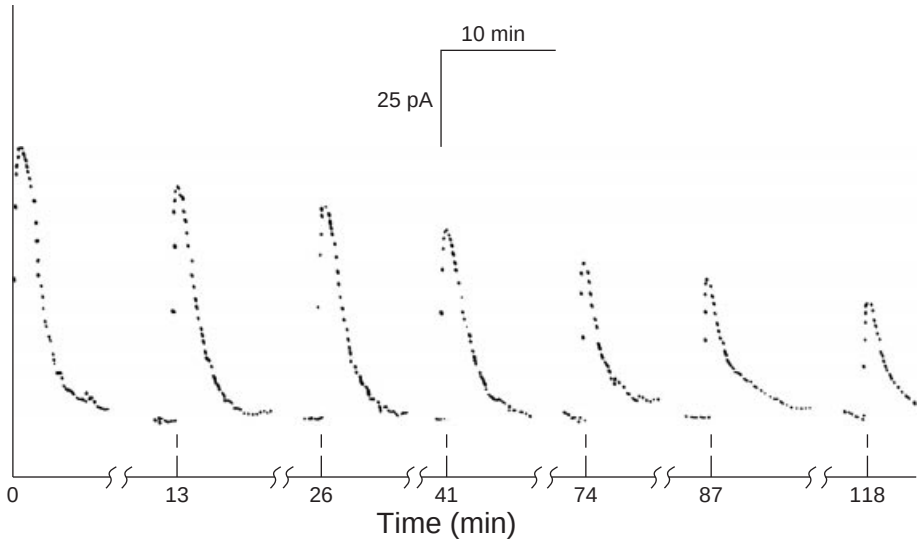

Figure 10 Observed time course plot of repeated dopamine bathings (0.5 mM, 30 μL each arrow) with the electrode placed in the cell body of the identified dopamine neuron of *Planorbis corneus*. The dopamine solution used for extracellular bathings was a modified snail saline solution. The current was monitored at +0.78 V vs SSCE with 7 s between measurements. (Reproduced from ref. 70. © 1991 with permission from Wiley-VCH.)

dopamine inside the giant dopamine cell from the pond snail *Planorbis corneus*. Active transport of dopamine into the cell has been monitored using linear scan voltammetry and poly(ester sulfonic acid)-coated carbon ring electrodes. In these experiments, the electrode is implanted in the neuron and the cell is bathed in 0.5 mM dopamine.[70] When the oxidation current at the electrode is monitored at +0.78 V and plotted against time, distinct peaks can be seen following extracellular application of the dopamine solution (Figure 10). The rising portion of each peak represents the active transport of dopamine across the cell membrane, and the decreasing portion of each peak represents the clearance of cytoplasmic dopamine through metabolism and/or vesicularization.

In these experiments, the peak current for repeated bathings with dopamine gradually decreases over time. The average loss of response in the cell is 36% after 20 min of continuous voltammetry. This loss is most likely due to a decrease in electrode sensitivity brought on by specific adsorption of large biomolecules. Loss of sensitivity from fouling of the electrode surface has been largely overcome by use of integrated pulse linear scan voltammetry, at platinized-carbon ring electrodes.[17] This method, also referred to as pulsed amperometric detection (PAD), minimizes electrode fouling while maintaining a reactive electrode surface. The applied waveform, based on the pulse techniques developed by Johnson's group,[75,76] consists of a rapid cyclic potential sweep followed by a large positive potential pulse to oxidatively clean the electrode surface and then a large negative potential pulse to restore the reactivity of the electrode.[17] This technique allows continuous amperometric detection that is both sensitive and reproducible. The average peak concentration of dopamine uptake in the *Planorbis* studies has been found to be 44 ± 2 μM, and the average rate of dopamine clearance, estimated from the linear portion of the declining limiting current, is 0.29 μM s^{-1}.[17]

To gain insight into the distribution of dopamine inside the cell, intracellular studies measuring basal dopamine levels have also been carried out. Carbon ring electrodes with electrode tips of approximately 6 μm have been placed inside the giant dopamine cell of *Planorbis* to monitor dopamine using staircase voltammetry.[71] Basal levels of free dopamine are below the detection limit of this method, so 0.5 mM dopamine is added extracellularly. The uptake of dopamine by the cell causes a sharp increase in the oxidation current detected in the cytoplasm, results in agreement with those presented by Lau et al.[17,70] The extracellular application of nominfensine (700 μL, 0.6 mM), an uptake inhibitor, diminishes the oxidation current, verifying that the rise in current observed is due to the uptake of extracellular dopamine and not free intracellular dopamine.

Bathing the cell in 300 μL of 50% ethanol disrupts the cellular membranes and results in a large increase in intracellular dopamine. Capillary electrophoresis has been used to verify the identification of the electroactive compound released upon ethanol exposure. In the resting state, cytoplasmic dopamine levels are below the detection limits of this method. However, upon exposure to ethanol, 14 fmol of dopamine is detected. Assuming an injection volume of 100 to 300 pL, the

cytoplasmic dopamine concentration following ethanol treatment is approximately 1.4×10^{-4} to 4.7×10^{-5} M.[71] This technique has also been coupled to off-column amperometric detection[77] with a detection limit of 10^{-8} M in 50 pL samples.[78] These lower detection limits are necessary for detection of basal levels of cytoplasmic free dopamine. Olefirowicz and Ewing[78] have determined basal dopamine concentration in *Planorbis* to be 2.2 ± 0.5 μM.

6 EXTRACELLULAR VOLTAMMETRY AT SINGLE CELLS

6.1 Introduction

As described above, carbon-fiber microelectrodes have been used in a number of studies to monitor extracellular changes in neurotransmitter levels both in the brain and in brain slices. One limitation to these types of studies is that the observed signal arises from multiple cells. Furthermore, since the electrodes are much larger than the synapse, one can only infer what is going on at the site of neurotransmitter release. The next frontier for voltammetry has been to monitor the exocytotic release of neurotransmitters at the level of a single cell.

Several different examples are given below illustrating how carbon-fiber microelectrodes can be used to monitor exocytosis at a single cell. In all examples the experimental protocols are similar. In most cases, with the exception of two examples, tissue containing the cell(s) of interest is dissociated into a single-cell suspension and cultured. The distance between single cells in culture facilitates electrode placement onto a cell and minimizes possible interferences from neighboring cells. Carbon-fiber microelectrodes, which are usually beveled for these experiments, are positioned via a micromanipulator next to or directly against a single cell (see Figure 11). To stimulate exocytosis, nanoliter volumes of a chemical stimulant (i.e. elevated K^+ and/or nicotine) are administered to a cell via pressure injection through a micropipette placed approximately 20–30 μm from the cell. The electrode is held at a constant potential and the current is sampled at high rates. This electrochemical technique is referred to as amperometry. When electroactive neurotransmitters are released from a cell and approach the electrode surface they are oxidized. The resulting current gives rise to a positive deviation in the baseline.

6.2 Overview of Systems Used for Single-cell Analysis

Single-cell amperometry has been used to investigate many types of cells that release easily oxidized substances by exocytosis. Some of the nonsynaptic cell systems that have been investigated include bovine adrenal chromaffin cells, rat pheochromocytoma (PC12) cells, beige mouse mast cells, and human pancreatic β-cells. Synaptic systems including invertebrate and mammalian neurons have also been investigated. A brief overview of some of the key studies using these different cell types will be presented in the following sections.

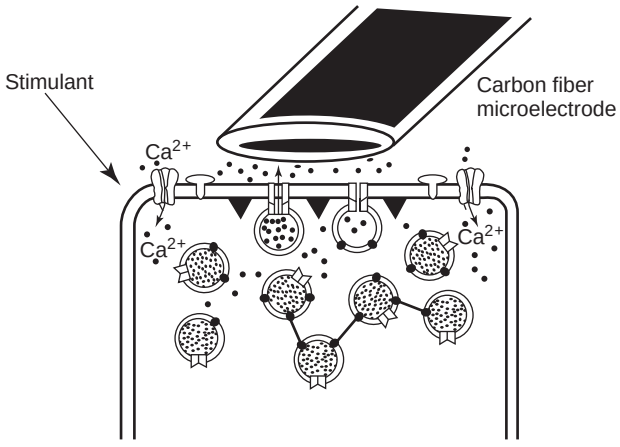

Figure 11 Schematic showing the experimental arrangement used for measuring neurotransmitter secretion from a single cell. (Reproduced from ref. 104, 225. © 1998 with permission from Marcel Dekker Inc.)

6.2.1 Bovine Adrenal Chromaffin Cells

The direct measurement of catecholamine secretion from a single cell was first reported by workers in Wightman's laboratory.[79] Bovine adrenal chromaffin cells were used in this study. Adrenal chromaffin cells are useful models of neurotransmitter biosynthesis, metabolism, and secretion owing to their neuroectodermal origin and biochemical and functional similarities with postganglionic sympathetic neurons.[80] Changes in the chemistry surrounding a single chromaffin cell are monitored using FCV at an electrode placed next to a cell. The application of 100 μM nicotine results in a series of sharp irregular concentration spikes superimposed on a secretion envelope (Figure 12d). Individual cyclic voltammograms confirm that the substances secreted and detected are catecholamines (Figure 12b). The secretion of catecholamines at a single cell has been shown to be induced by stimulants such as nicotine, the nicotinic agonist carbamylcholine, and K^+.[81] Furthermore, K^+- and nicotine-stimulated secretion could be inhibited by Cd^{2+}, a calcium channel blocker, while nicotine-stimulated secretion could be selectively inhibited by hexamethonium, a sympathetic ganglionic blocker.[81]

Many of these observations of catecholamine secretion are consistent with similar measurements made at multiple-cell preparations such as the intact adrenal

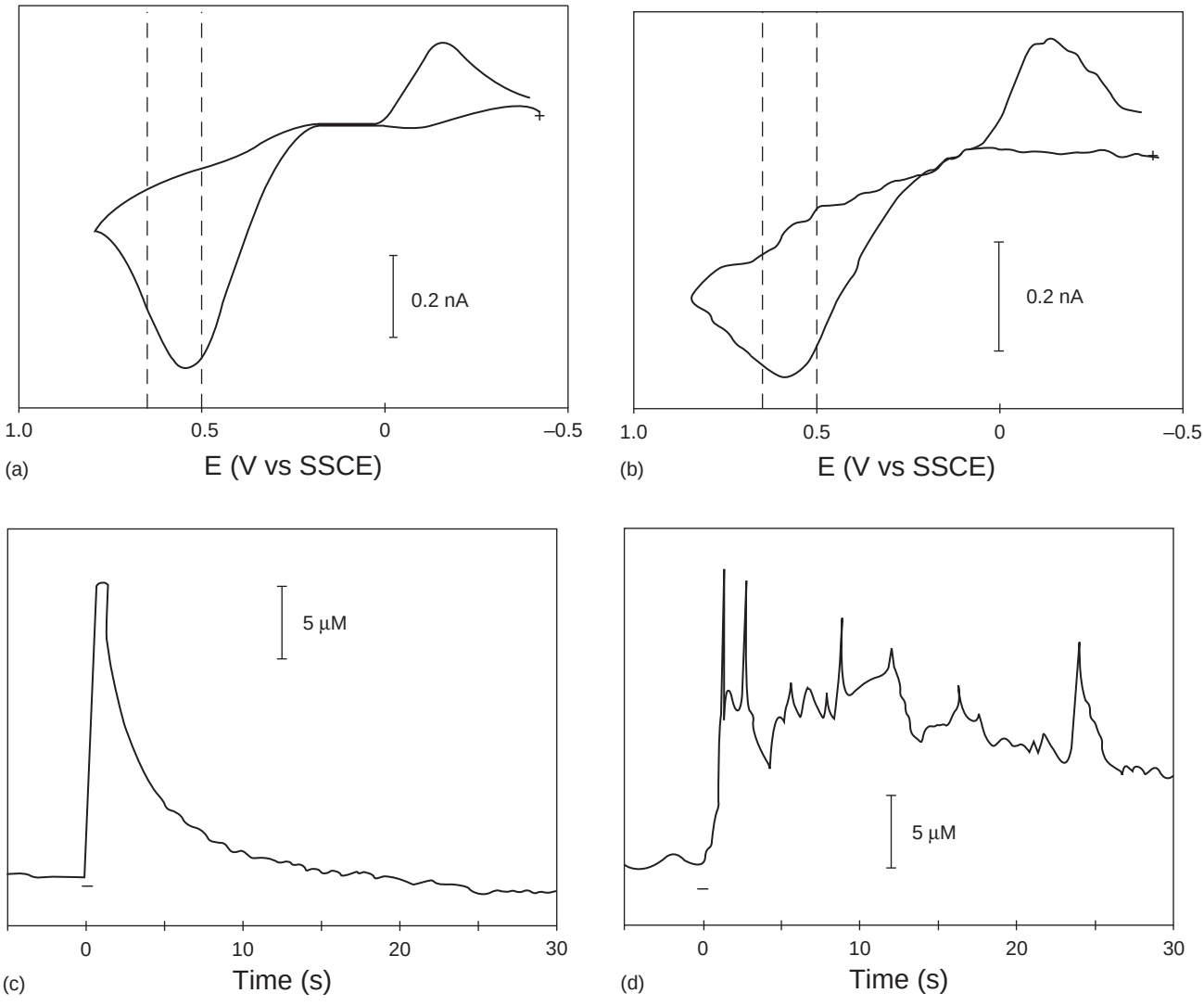

Figure 12 Cyclic voltammetric response ($200\,V\,s^{-1}$, repeated at 100-ms intervals) of a carbon-fiber electrode to norepinephrine ejection from an adjacent micropipette and to catecholamine secretion from a single chromaffin cell. Panels (a) and (b) are averaged background-subtracted voltammograms of the substances whose concentration changed during the measurement interval of panels (c) and (d). Each time point in panels c and d is the integrated current recorded from 0.5–0.6 V from individual voltammograms (hatched lines in panels (a) and (b)); bars to the right in panels (c) and (d) are the conversion of current to catecholamine concentration based on calibration curves constructed with standards. Panels (a) and (c) are the electrode response to a 1-s ejection (3 nL) of 20 μM norepinephrine applied at $t = 0$ with the ejection pipette 20 μm from the electrode. Panels (b) and (d) are the electrochemical response obtained with the electrode tip adjacent to a single cell; at $t = 0$, a 1-s ejection of nicotine (100 μM) was made 20 μm away from the cell. (Reproduced with permission from ref. 79. © 1990 The American Society for Biochemistry.)

gland, adrenal slices, and whole cultures of adrenal cells. Additionally, since vesicular fusion events should result in discrete packets of catecholamines being released from the cell surface, the observed chemical concentration spikes are the expected consequence of exocytosis.

Combined, the above evidence suggested that the single-cell electrochemical signals were the result of secretion of catecholamines via exocytosis. To improve the temporal resolution of the measurements made with FCV and to more accurately measure exocytosis, which can occur on the millisecond to microsecond timescale, single-cell measurements have been made in the amperometric mode. The faster sampling rates possible with this technique make it possible to more clearly resolve the chemical spikes. For each spike, the integrated current, or charge, is proportional to the mass of the analyte detected and can be used in Faraday's law, Equation (1), to calculate the number of molecules detected per each spike event. Faraday's law can be expressed as

$$\frac{Q}{nF} = N \tag{1}$$

where Q is the charge in coulombs, n is the number of moles of electrons transferred per mole of analyte oxidized, F is the Faraday constant (96 485 C equiv^{-1}) and N is the total number of moles of substance oxidized.

Several results from the above studies have demonstrated that the properties of the electrochemically detected spikes at single bovine adrenal cells correlated well with expectations based on the exocytotic secretion of catecholamines from individual vesicles. For example, cyclic voltammograms at individual concentration spikes confirm that the spikes are due to the detection of concentration packets of catecholamines rather than an electrical artifact.[79,81] Additionally, in agreement with the requirements of Ca^{2+} for exocytosis, spikes do not occur in the absence of this ion.[80] Within each vesicle there is a discrete or quantal amount of neurotransmitter. If each of the observed chemical spikes was due to the exocytotic release of neurotransmitter(s) from a single vesicle, the amount detected per each spike event should be independent of the method of stimulation. Different methods of stimulation have been used to test this hypothesis.[80] Stimulation with 100 μM nicotine, results in the highest frequency response (1.2 ± 0.2 Hz), whereas at a lower concentration (10 μM) the response is much slower (0.5 ± 0.2 Hz). Both the nicotinic agonist carbamylcholine (1 mM), and KCl (60 mM) give a similar response (0.7 ± 0.2 and 0.6 ± 0.2 Hz, respectively). The mean ± SEM (standard error of the mean) spike areas for the different stimulants is 1.19 ± 0.9 pC for K^+, 0.98 ± 0.13 pC for carbamylcholine, 1.04 ± 0.16 pC for 10 μM nicotine, and 1.19 ± 0.10 pC for 100 μM nicotine. Clearly, the frequency of the spikes increases with the intensity of the stimulus. However, the average quantity of catecholamine in each spike does not significantly differ with the intensity or type of stimulant used. While previous experiments in other laboratories have provided evidence for the quantized nature of neurotransmitter release, results from these studies were the first to provide direct chemical evidence for exocytosis. Further support for these conclusions was provided by Chow et al.[82] In this study, it was found that the time-averaged signal of the carbon-fiber microelectrode closely resembled the derivative of the membrane capacitance trace (Figure 13). Membrane capacitance is proportional to the cell surface area and increases when individual vesicles fuse with the plasma membrane during exocytosis.[83]

Since the initial studies, bovine adrenal cells have been one of the most widely used cell types in single-cell amperometry experiments. From these experiments, an incredibly large amount of information has been generated which has helped further our basic understanding of exocytosis, the fundamental means by which chemical communication takes place between cells in the nervous system, at the single-cell level. Several studies have investigated

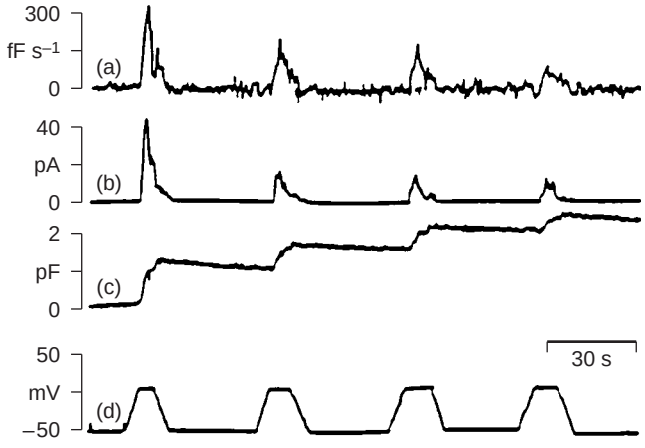

Figure 13 Correlation between the amperometric signal and the rate of capacitance change at a single adrenal cell. A cell was held at −60 mV and repetitively depolarized to 0 mV for about 8 s at a time (d). Capacitance (c) increased during depolarizations, and the amperometric signal (b) displayed peaks. The time derivative of the capacitance signal, which, like the amperometric signal, should be proportional to secretion rate, is displayed in the top trace (in femtofarads per s) for comparison with the latter. (Reproduced from ref. 82. © 1992 with permission of Nature.)

how conditions of the extracellular environment, such as temperature,[84] osmolarity,[85] and pH,[86] affect the release of catecholamines from single chromaffin vesicles. Amperometry has also been used to dissect different temporal stages of vesicular catecholamine release at bovine adrenal cells.[87–90] Chow et al.[82] for example, reported a delay in vesicle fusion compared to neuronal systems and a small peak or ''foot'' at the onset of vesicular release events. Schroeder et al.[91] also proposed that there are three distinct stages of exocytosis at bovine adrenal cells that include formation of a fusion pore, pore expansion to release catecholamine, and a final dissociation of the intravesicular matrix. Bovine adrenal cells have been used with amperometry and in conjunction with other single-cell techniques to learn more about the role of Ca^{2+} and Ca^{2+} channels in the exocytotic process.[92–97] By using small (1 μm) flame-etched carbon-fiber electrodes, it has been possible to map the surface of adrenal cells and identify the existence of spatially localized release sites.[98] Amperometry at single bovine adrenal cells has also been used to study the regulatory role of autoreceptors in exocytosis.[99] More detailed discussions of findings from these studies and those discussed in the following sections may be found in the literature.[100–104]

6.2.2 Rat Pheochromocytoma (PC12) Cells

PC12 cells are an immortalized cell line derived from rat adrenal chromaffin cells.[105] These cells have been widely used as a model for adrenal chromaffin cells and

sympathetic neurons. PC12 cells are an attractive model neuronal system owing to their ability to differentiate into neuronal-type cells upon treatment with nerve growth factor (NGF). Additionally, like sympathetic neurons, PC12 cells can synthesize acetylcholine, as well as the catecholamines dopamine and norepinephrine.[106,107]

Using amperometric detection at carbon-fiber electrodes, catecholamine release has been observed at the zmol level from single PC12 cells in culture.[108] The amperometric response at a single PC12 cell following three successive stimulations with both 1 mM nicotine and 105 mM K^+ is compared to that of a control in Figure 14. The shape of each transient is similar to those observed from bovine adrenal cells[80] except that the average charge for each transient is significantly smaller. Individual current spikes have an average half-width of 9.3 ± 0.1 ms and the average catecholamine content calculated using the area under each current transient and Faraday's law is 190 ± 3.5 zmol.[108]

Following treatment with NGF, PC12 cells extend processes.[109] Along these processes, are bulbous regions (1–2 μm diam.), referred to as varicosities. Zerby and Ewing[110] have used PC12 cells treated with NGF to investigate how the site of exocytosis changes during the differentiation process. To investigate these changes, beveled carbon-fiber microelectrodes have been placed at several different locations on differentiated PC12 cells. Experiments carried out between days 10 and 14 of culture show no release from the cell body ($n = 3$), only occasional responses from the smooth regions of the neurites ($n = 5$), and frequent release when the electrode is placed at a varicosity ($n = 16$). Additionally, the average vesicular catecholamine content observed at varicosities is not significantly different from that observed at the cell body of undifferentiated cells. These results lend some insight into functional changes that occur during differentiation and indicate that during differentiation the relocation of release sites to PC12 varicosities does not significantly alter the mean catecholamine content of individual vesicles.

The time resolution of carbon-fiber microelectrodes not only allows the detection of single exocytotic events, but also provides an excellent means by which to monitor the time course of the stimulus–secretion process. PC12 cells possess both nicotinic and muscarinic receptors that trigger exocytosis through two different mechanisms. Nicotine causes a change in the conformation of the nicotine receptor opening channels permeable to Na^+. The influx of Na^+ causes depolarization of the cell membrane which opens voltage-sensitive Ca^{2+} channels that allow Ca^{2+} into the cell to cause exocytosis.[111] Muscarine activates exocytosis through muscarinic receptors that act through intracellular second messengers to release Ca^{2+} from intracellular stores.[112] In contrast to nicotine and muscarine, K^+ causes exocytosis through direct membrane depolarization. Zerby and Ewing have used amperometry at single PC12 cells to investigate how these different mechanisms affect the time course or latency of the

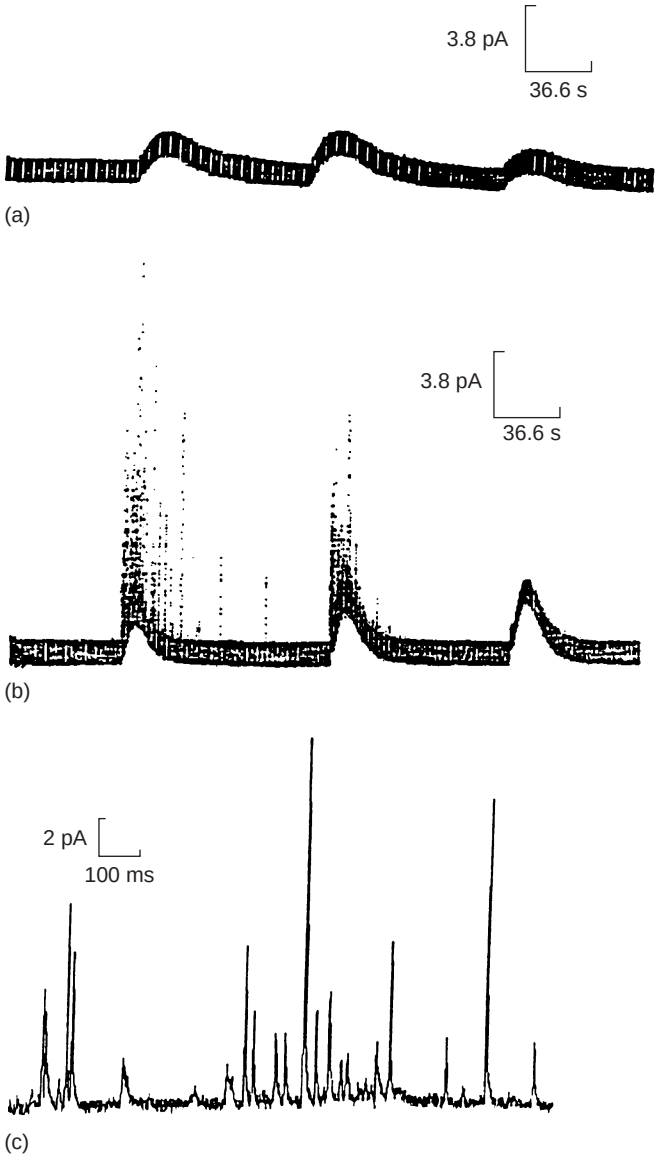

Figure 14 Amperograms of individual vesicular exocytosis events from PC12 cells. (a) Amperogram of a control experiment with the tip of the working electrode 200 μm from the cell. (b) Amperogram of vesicular exocytosis induced by bathing a cell with 1 mM nicotine in 105 mM K^+ balanced salt solution while the electrode was placed on top of the cell. (c) Enlargement of a 1-s period of the amperogram from the first stimulation in (b). Data displayed correspond to the time period near the middle of the first baseline disturbance of (b). Traces (a) and (b) were obtained from computer screen dumps; each printed line represents 600 data points. (Reproduced with permission from ref. 108. © 1994 American Chemical Society.)

release process.[113] Their results have shown that while the average vesicle catecholamine content is unaltered by the different mechanisms of release, their latencies (time between application of the stimulant and onset of exocytotic events) vary significantly. The mean latencies for each type of stimulation are reported as: 6 ± 1 s for K^+ (105 mM), 37 ± 5 s for nicotine (1 mM), and 103 ± 11 s for muscarine (1 mM). Figure 15 shows representative amperograms obtained using the three different stimulants. Interestingly, these results are in agreement with similar measurements made at single rat adrenal cells.[114]

Several studies have used PC12 cells to investigate how different pharmacological agents effect the average catecholamine content of PC12 vesicles. For example, PC12 vesicular dopamine levels have been shown to decrease when cells are exposed to amphetamine and reserpine.[69,115] Recently, Pothos et al. have used amperometry at single PC12 cells to investigate the effects of D_2 autoreceptors on the exocytotic release of neurotransmitters. By treating PC12 cells with the D_2 agonist quinpirole, it was possible to decrease the quantal size of release events by approximately 50%.[116] In contrast to results obtained with reserpine and amphetamine, these results provide evidence for receptor-mediated mechanisms that can decrease quantal release. The size of release events detected at PC12 cells has also been modulated by exposing cells to the catecholamine precursor L-3,4-dihydroxyphenylalanine (L-Dopa). Following exposure to 50 μM L-Dopa for 40 to 90 min, the average size of quantal events is increased to 251% of control values.[117]

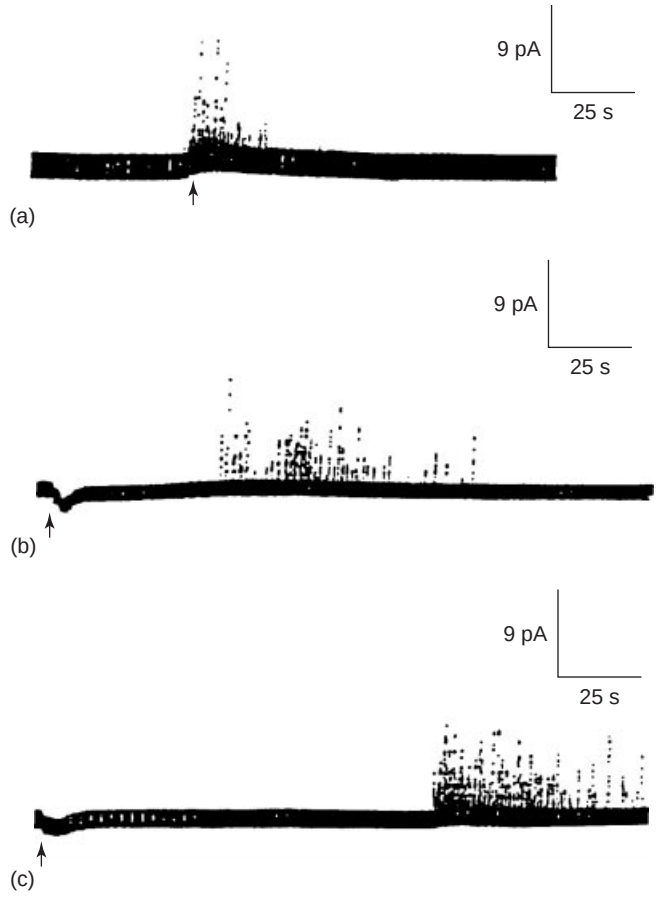

Figure 15 Current–time traces for exocytosis at single PC12 cells. A 6-s ejection of stimulant [105 mM K^+ (a), 1 mM nicotine (b), or 1 mM muscarine (c)] from a microinjector was administered at each arrow. The resulting current transients correspond to the oxidation of dopamine at the electrode tip as it is released from the cell. Detection was performed in the amperometric mode at 650 mV vs SSCE. (Reproduced from *J. Neurochem.*[113] © 1996 with permission from Lippincott, Williams and Wilkins.)

6.2.3 Mast Cells, Pancreatic β-cells, and Rat Melanotrophs

Beige mouse mast cells have been utilized for amperometric investigations of the secretory agents histamine and serotonin. Histamine and serotonin are easily oxidized molecules and are important in the nervous and immune systems of mammals. One advantage of using this cell system is the large size of the vesicles present. The larger vesicles allow release to be more easily correlated with simultaneous changes in membrane capacitance.[118] Using amperometric methods, histamine and serotonin have been shown to be co-released from single mast cell vesicles.[119]

Amperometric techniques have also been extended to pancreatic β-cells and rat melanotrophs, secretory cells not normally considered as model systems for studying neurotransmission. Pancreatic β-cells secrete insulin in response to glucose levels in the body. The common link between insulin secretion from β-cells and the other secretory cells described is that insulin is stored in vesicles and secreted via exocytosis. Membrane-potential experiments with intracellular microelectrodes[120] and amperometric monitoring[121] of insulin secretion, using chemically modified electrodes, have been carried out at individual pancreatic β-cells in an effort to better understand how these cells regulate blood sugar levels in the body. Kennedy et al. have further investigated secretion from β-cells using chromatographic methods and have confirmed that the secreted products are indeed insulin.[122]

Peptide hormones are necessary for many biological functions and understanding how they are stored and secreted at the single-cell level is extremely important. Kennedy et al. have demonstrated that it is possible to detect the exocytotic release of peptides at single rat melanotrophs. The secretion of peptide hormones can be detected electrochemically by the oxidation of tryptophan and tyrosine residues found in the proopiocortin cleavage

products. The calcium-mediated release of ∂-melanocyte stimulating hormone has also been measured at a carbon-fiber electrode placed on a single rat melanotroph.[123]

6.2.4 Invertebrate and Mammalian Neurons

Exocytotic events at single neurons from the pond snail *Planorbis corneus*[124,125] and the leech *Hirudo medicinalis*[126] have been studied using extracellular voltammetry. Bruns and Jahn[126] have described electrochemical detection of serotonin release from isolated Retzius cells of the leech *Hirudo medicinalis*. Single leech Retzius neurons have been cultured and a carbon-fiber electrode placed at the axonal stump and the cell body. Two types of exocytotic responses have been observed, one interpreted as resulting from serotonin release from small clear synaptic vesicles and the other from large dense-core vesicles. Release from the small clear vesicles appears to occur more rapidly (faster time constant) than from the large dense-core vesicles, which are more randomly distributed throughout the cell. It has been suggested that the faster rate observed for the small clear vesicles is due to the discharge of their contents on a submillisecond timescale through an undilated fusion pore.

Using single-cell amperometry, Chen et al.[124] have found that stimulation of the giant dopamine neuron of the pond snail *Planorbis corneus*, with elevated KCl, results in massive exocytotic release from the cell body (Figure 16). Since norepinephrine and epinephrine have not been detected in the giant dopamine cell,[127] it appears that the observed current transients are due to the release of dopamine from the cell body. Capillary electrophoresis with electrochemical detection has been used to verify that dopamine, indeed, is being released. The current transients observed from the cell body have rise times between 2 and 5 ms, an average base width of 14 ± 0.8 ms ($n = 13$ cells, 12 324 transients), and are not observed in the absence of Ca^{2+}.

In addition to the large number of vesicular release events at the cell body, the amount of dopamine released from the vesicles has an interesting distribution.[124,125] When the area of each spike is converted into vesicular content and plotted as a histogram, a bimodal distribution is observed in which each phase drops off exponentially (Figure 17a). To correlate the content of release events with vesicle radii, it has been assumed that the catecholamine concentration within each vesicle is constant.[128] In contrast to results obtained from other cell types,[128] when release events from the giant dopamine cell are plotted against the cube root of vesicle content two distinct Gaussians can clearly be seen in the histogram. These results suggest that there are at least two distinct classes of vesicles or release events occurring at this cell (Figure 17b). Results from the *Planorbis* system have also shown that amphetamine differentially alters the catecholamine content of the two classes of vesicles[125] and that release events at the cell body occur in a bursting pattern with regular time intervals between individual events (Figure 16).[129]

While much has been learned by studying exocytosis at model neuronal systems and invertebrate neurons, efforts have also been focused on directly analyzing exocytosis at single mammalian neurons from the central nervous system (CNS) and peripheral nervous system (PNS). Developing superior cervical ganglion neurons from

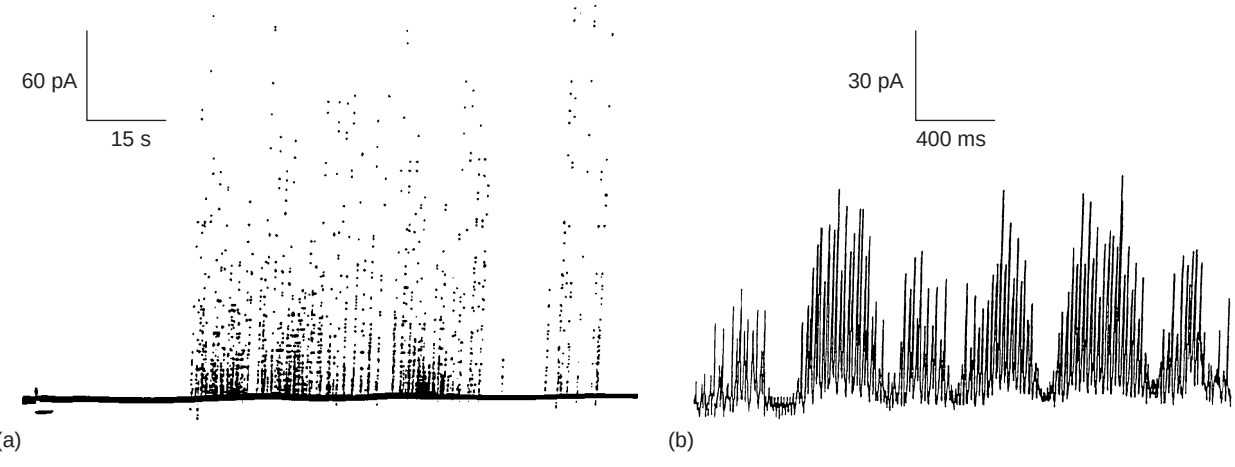

Figure 16 Dopamine current transients detected by a carbon-fiber disk electrode placed on the cell body of the dopamine neuron in *Planorbis corneus*. (a) An example of current transients recorded by amperometry. A large *Planorbis* dopamine cell (diameter about 100 μm) was stimulated with a 4-s potassium chloride (1 M) pulse (87 nL) delivered from a glass pipette that was placed about 15 μm from the cell body. The stimulation is shown by the horizontal bar below the trace. (b) Bursting release events were observed in 24 out of 29 cells that showed release transients. The overall success rate for observation of current transients from cells sampled was approximately 50%. (Reproduced from *J. Neuroscience*, **15**(11). © 1995 with permission.)

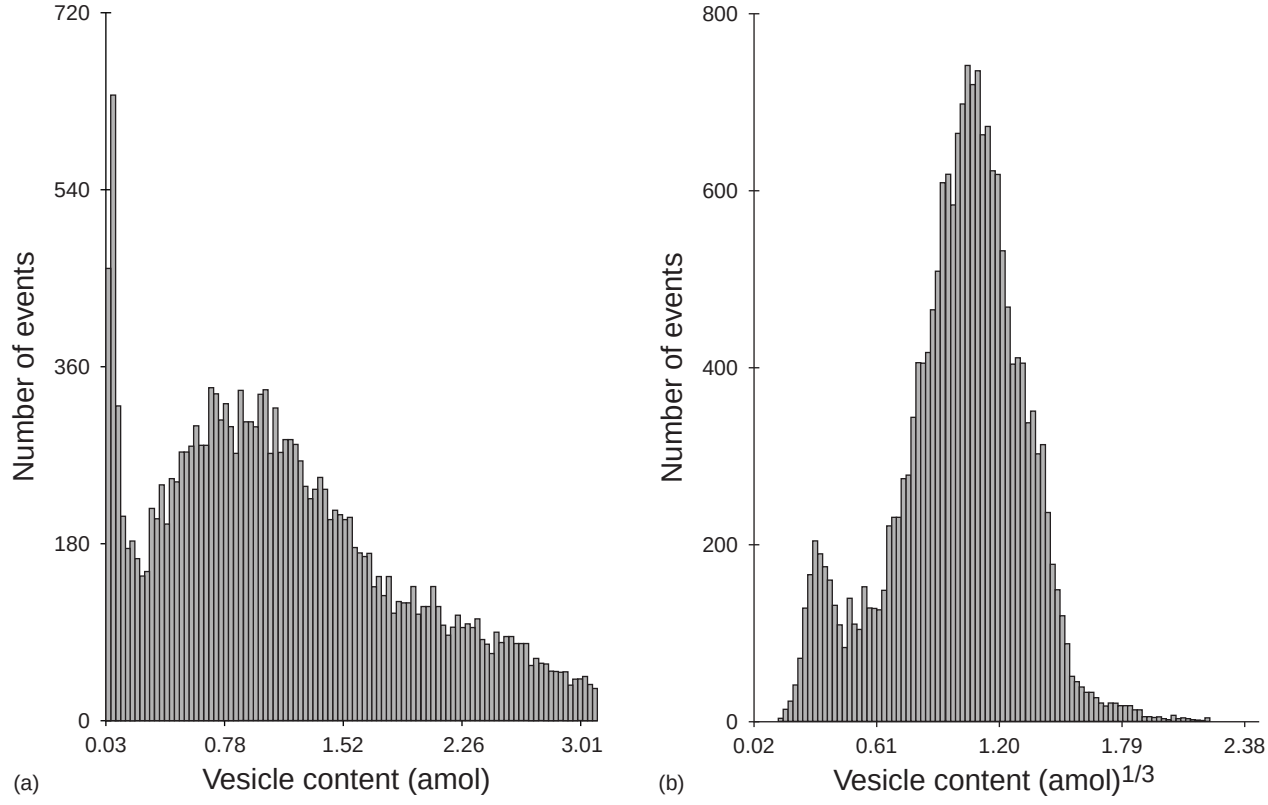

Figure 17 Histograms of the frequency of release events vs the amount of neurotransmitter released from the dopamine cell of *Planorbis*. (a) Frequency vs attomoles of dopamine released (16 cells, 18 456 events). Only release events with base widths less than 40 ms were considered. Wider transients, which made up less than 2% of the total number and mostly appeared to contain contributions from more than one vesicle (appearing as doublets and shoulders), were excluded in the calculations. Based on the results obtained from 16 cells, the average vesicle content was 1.36 ± 0.53 amol (means $\pm$ s.d.), equivalent to $818\,000 \pm 319\,000$ molecules of dopamine. (b) Frequency vs the cubed root of attomoles of dopamine released (16 cells and 18 456 events). (Reproduced from *J. Neuroscience*, **15**(11). © 1995 with permission.)

neonatal rats, for example, have been investigated with amperometric methods and vesicular release has been detected at varicosities along axons from these neurons following the application of potassium or black widow spider venom.[130] Zhou and Misler[130] report a median spike charge of 11.3 femtocoulombs, which corresponds to 35 000 catecholamine molecules per release event (58 zmol). These values are similar to those obtained for release events associated with small clear vesicles from leech Retzius neurons.[126]

Recently, two studies have used amperometry to investigate the quantal release of neurotransmitters from CNS neurons. Pothos et al. adapted amperometric methods to directly observe quantal dopamine release from axonal varicosites of midbrain dopamine neurons.[131] Events were elicited by high K^+ or α-latrotoxin, required extracellular Ca^{2+}, and were abolished by reserpine. The characteristics of these events indicated that, on average, 3000 molecules were detected over 200 µs, parameters much smaller and faster than those associated with release events detected at other vertebrate preparations. The number of dopamine molecules detected per event increased to 380% of control values after cells were exposed to glial-derived neurotrophic factor and to 350% of control values after exposure to L-Dopa. These results indicate that quantal size is not invariant in CNS neurons.

Substantia nigra neurons release dopamine from their somatodendritic regions. However, previous studies had not been able to determine whether dopamine release occurred via exocytosis or some nonvesicular mechanism(s). To shed light on this process, Jaffe et al.[132] have used amperometry at carbon-fiber microelectrodes to assay catecholamine secretion from the cell bodies of substantia nigra neurons in rat midbrain slices. Detected secretion events had charge integrals ranging from a few to several hundred femtocoulombs. While local application of glutamate enhanced the frequency of events, the mean area of the spikes was not changed. The addition of Cd^{2+}, a blocker of voltage-dependent Ca^{2+} channels, to the cells blocked the stimulatory effects of glutamate. Together, these results suggest that dopamine

is released from the cell body region of substantia nigra neurons via exocytosis.

7 CONCLUDING REMARKS AND FUTURE DIRECTIONS

Understanding communication between mammalian neurons is the ultimate goal of most neuronal studies. Since the initial experiments in Adams' laboratory some 30 years ago, several developments have helped to push voltammetry to the forefront of available technologies. Today, voltammetry is being used in many different ways by chemists and neuroscientists to study fundamental processes related to neuronal communication. Fast scan rate cyclic voltammetry at microelectrodes combined with electrical stimulation has proven to be a very powerful means for detecting neurotransmitter dynamics in vivo and in vitro. These methods have been used to identify released neurochemicals, monitor how drugs affect neuronal communication, investigate the kinetics of neurotransmitter release and uptake, and further our understanding of the basic process of neurotransmission in the brain.

With the development of microelectrodes, many of the advantages of voltammetry have been exploited and used to study neurochemical processes both inside and at the surface of single cells and neurons. The small size of microelectrodes makes it easy to investigate spatially discrete regions of release on a single cell with the sensitivity and response time necessary to detect the rapid exocytotic release of minute amounts of neurotransmitters. The large amount of information that has been generated using single-cell electrochemical methods attests to the utility of this technology. In its current state, voltammetric methods are best suited to study the neurochemistry of biological systems that utilize easily oxidized neurotransmitters. Combined with enzyme-based electrodes and other chemically modified electrodes, a new group of electrochemical sensors promises to move this technology into the area of monitoring molecules that are normally not easily oxidized in biological microenvironments (key developments in this area have been published[104,133]). It seems likely that some of the important molecules to be examined in the future will be glutamate, acetylcholine, peptides, and NO.[134,135]

Future developments will probably also include the development of nanometer-size electrochemical probes to access the small synaptic gap. Looking back on the developments that have already taken place, it seems logical that intra-synaptic measurements will be the next frontier in which microelectrodes will be used for the electrochemical detection of neurotransmitters.

ACKNOWLEDGMENTS

The contributions by our coworkers that are referenced herein are gratefully acknowledged. This work was supported, in part, by grants from the National Science Foundation, and the National Institutes of Health. T.L.C. is a National Institute of Mental Health Pre-doctoral Fellow.

ABBREVIATIONS AND ACRONYMS

CNS	Central Nervous System
DAT	Dopamine Transporter
DOPAC	Dihydroxyphenylacetic Acid
ECF	Extracellular Fluid
FCV	Fast Cyclic Voltammetry
HPLC	High-performance Liquid Chromatography
i.p.	Intraperitoneal
ISEs	Ion-selective Electrodes
L-Dopa	L-3,4-dihydroxyphenylalanine
MFB	Medial Forebrain Bundle
NGF	Nerve Growth Factor
PAD	Pulsed Amperometric Detection
PNS	Peripheral Nervous System
SEM	Standard Error of the Mean
SSCE	Sodium Saturated Calomel Electrode

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Pulse Voltammetry • Selective Electrode Coatings for Electroanalysis • Ultrafast Electrochemical Techniques

REFERENCES

1. G.M. Shepherd, *Neurobiology*, 2nd edition, Oxford University Press, Oxford, 1988.
2. J.R. Cooper, F.E. Bloom, R.H. Roth, *The Biochemical Basis of Neuropharmacology*, Oxford University Press, Oxford, 1986.
3. R.N. Adams, 'Probing Brain Chemistry with Electroanalytical Techniques', *Anal. Chem.*, **48**, 1128–1138A (1976).
4. F. Valtorta, R. Fesce, F. Grohovaz, C. Haimann, W.P. Hurlbut, N. Iezzi, F. Torri Tarelli, A. Villa, B. Ceccarelli,

'Neurotransmitter Release and Synaptic Vesicle Recycling', *Neuroscience*, **35**, 477–489 (1990).
5. R.M. Wightman, L.J. May, A.C. Michael, 'Detection of Dopamine Dynamics in the Brain', *Anal. Chem.*, **60**, 769–779A (1988).
6. P.T. Kissinger, J.B. Hart, R.N. Adams, 'Voltammetry in Brain Tissue: A New Neurophysiological Measurement', *Brain Res.*, **55**, 209–213 (1973).
7. C. Nicholson, 'Ion-selective Microelectrodes and Diffusion Measurements as Tools to Explore the Brain Cell Microenvironment', *J. Neurosci. Methods*, **48**, 199–213 (1993).
8. S. Baudet, L. Hove-Madsen, D.M. Bers, 'How to Make and Use Calcium-specific Mini- and Microelectrodes', *Methods Cell Biol.*, **40**, 93–113 (1994).
9. J.O. Schenk, E. Miller, R.N. Adams, 'Electrochemical Techniques for the Study of Brain Chemistry', *J. Chem. Ed.*, **60**, 311–314 (1983).
10. R.M. Wightman, 'In Vivo Electrochemistry', in *Electrochemistry in Research and Development*, eds. R. Kalvoda, R. Parsons, Plenum Press, New York, 189–202, 1985.
11. C. Nicholson, J.M. Phillips, 'Ion Diffusion Modified by Tortuosity and Volume Fraction in the Extracellular Microenvironment of the Rat Cerebellum', *J. Physiol.*, **321**, 225–257 (1981).
12. M.E. Rice, C. Nicholson, 'Diffusion Characteristics and Extracellular Volume Fraction During Normoxia and Hypoxia in Slices of Rat Neostriatum', *J. Neurophysiol.*, **65**, 264–272 (1991).
13. R.M. Penner, M.J. Heben, T.L. Longin, N.S. Lewis, 'Fabrication and Use of Nanometer-sized Electrodes in Electrochemistry', *Science*, **250**, 1118–1121 (1990).
14. A. Meulemans, B. Poulain, G. Baux, L. Tauc, 'Changes in Serotonin Concentration in a Living Neurone: A Study by On-line Intracellular Voltammetry', *Brain Res.*, **414**, 158–162 (1987).
15. N. Georgolios, D. Jannakoudakis, P. Karabinas, 'Pt Electrodeposition on PAN-based Carbon Fibers', *J. Electroanal. Chem.*, **264**, 235–245 (1989).
16. Y.Y. Lau, T. Abe, A.G. Ewing, 'Voltammetric Measurement of Oxygen in Single Neurons Using Platinized Carbon Ring Electrodes', *Anal. Chem.*, **64**, 1702–1705 (1992).
17. Y.Y. Lau, D.K.Y. Wong, A.G. Ewing, 'Intracellular Voltammetry at Ultrasmall Platinum Electrodes', *Microchem. J.*, **47**, 308–316 (1993).
18. K.T. Kawagoe, J.B. Zimmerman, R.M. Wightman, 'Principles of Voltammetry and Microelectrode Surface States', *J. Neurosci. Methods*, **48**, 225–240 (1993).
19. R.M. Wightman, D.O. Wipf, 'Voltammetry at Ultramicroelectrodes', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 267–353, Vol. 15, 1989.
20. A. Fitch, D.H. Evans, 'Use of Microelectrodes for the Study of a Fast Chemical Step in an Electrode Reaction', *J. Electroanal. Chem.*, **202**, 83–92 (1986).
21. J.A. Stamford, 'In Vivo Voltammetry: Some Methodological Considerations', *J. Neurosci. Methods*, **17**, 1–29 (1986).
22. A.G. Ewing, K.D. Alloway, S.D. Curtis, M.A. Dayton, R.M. Wightman, G.V. Rebec, 'Simultaneous Electrochemical and Unit Recording Measurements: Characterization of the Effects of D-Amphetamine and Ascorbic Acid on Neostriatal Neurons', *Brain Res.*, **261**, 101–108 (1983).
23. P.M. Plotsky, 'Differential Voltammetric Measurement of Catecholamines and Ascorbic Acid at Surface-modified Carbon Filament Microelectrodes', *Brain Res.*, **235**, 179–184 (1982).
24. A.G. Ewing, M.A. Dayton, R.M. Wightman, 'Pulse Voltammetry with Microvoltammetric Electrodes', *Anal. Chem.*, **53**, 1842–1847 (1981).
25. F.G. Gonon, F. Navarre, M.J. Buda, 'In Vivo Monitoring of Dopamine Release in the Rat Brain with Differential Normal Pulse Voltammetry', *Anal. Chem.*, **56**, 573–575 (1984).
26. P. Capella, B. Ghasemzadeh, K. Mitchell, R.N. Adams, 'Nafion-coated Carbon Fiber Electrodes for Neurochemical Studies in Brain Tissue', *Electroanalysis*, **2**, 175–182 (1990).
27. M.W. Lada, R.T. Kennedy, 'Quantitative In Vivo Monitoring of Primary Amines in Rat Caudate Nucleus Using Microdialysis Coupled by a Flow-gated Interface to Capillary Electrophoresis with Laser-induced Fluorescence Detection', *Anal. Chem.*, **68**, 2790–2797 (1996).
28. B.L. Hogan, S.M. Lunte, J.F. Stobaugh, C.E. Lunte, 'On-line Coupling of In Vivo Microdialysis Sampling with Capillary Electrophoresis', *Anal. Chem.*, **66**, 596–602 (1994).
29. C.A. Marsden, M.P. Brazell, N.T. Maidment, 'An Introduction to In Vivo Electrochemistry', in *Measurement of Neurotransmitter Release In Vivo*, ed. C.A. Marsden, J. Wiley and Sons, New York, 127–151, 1984.
30. J.C. Conti, E. Strope, R.N. Adams, C.A. Marsden, 'Voltammetry in Brain Tissue: Chronic Recording of Stimulated Dopamine and 5-Hydroxytryptamine Release', *Life Sciences*, **23**, 2705–2715 (1978).
31. R.F. Lane, A.T. Hubbard, K. Fukunaga, R.J. Blanchard, 'Brain Catecholamines: Detection In Vivo by Means of Differential Pulse Voltammetry at Surface-modified Platinum Electrodes', *Brain Res.*, **114**, 346–352 (1976).
32. F. Gonon, M. Buda, R. Cespuglio, M. Jouvet, J.F. Pujol, 'In Vivo Electrochemical Detection of Catechols in the Neostriatum of Anaesthetized Rats: Dopamine or DOPAC?', *Nature*, **286**, 902–904 (1980).
33. A.G. Ewing, R.M. Wightman, M.A. Dayton, 'In Vivo Voltammetry with Electrodes that Discriminate between Dopamine and Ascorbate', *Brain Res.*, **249**, 361–370 (1982).
34. M.A. Dayton, A.G. Ewing, R.M. Wightman, 'Evaluation of Amphetamine-induced In Vivo Electrochemical Response', *Eur. J. Pharmacol.*, **75**, 141–144 (1981).

35. J.D. Salamone, L.S. Hamby, D.B. Neill, J.B. Justice, 'Extracellular Ascorbic Acid Increases in Striatum Following Systemic Amphetamine', *Pharmacol. Biochem. Behav.*, **20**, 609–612 (1984).
36. R.L. Wilson, R.M. Wightman, 'Systemic and Nigral Application of Amphetamine Both Cause an Increase in Extracellular Concentration of Ascorbate in the Caudate Nucleus of the Rat', *Brain Res.*, **339**, 219–226 (1985).
37. R.D. O'Neill, 'The Measurement of Brain Ascorbate In Vivo and its Link with Excitatory Amino Acid Neurotransmission', in *Neuromethods, Voltammetric Methods in Brain Systems*, eds. A.B. Boulton, G.B. Baker, R.N. Adams, Humana Press, Totowa, NJ, 221–268, Vol. 27, 1995.
38. G.A. Gerhardt, A.F. Oke, G. Nagy, B. Moghaddam, R.N. Adams, 'Nafion-coated Electrodes with High Selectivity for CNS Electrochemistry', *Brain Res.*, **290**, 390–395 (1984).
39. G. Rose, G. Gerhardt, I. Stromberg, L. Olson, B. Hoffer, 'Monoamine Release from Dopamine-depleted Rat Caudate Nucleus Reinnervated by Substantia Nigra Transplants: An In Vivo Electrochemical Study', *Brain Res.*, **341**, 92–100 (1985).
40. J.A. Stamford, 'In Vivo Voltammetry: Prospects for the Next Decade', *Trends Neurosci.*, **12**, 407–412 (1989).
41. R.N. Adams, 'In Vivo Electrochemical Measurements in the CNS', *Prog. Neurobiol.*, **35**, 297–311 (1990).
42. F. Crespi, 'In Vivo Voltammetry with Micro-biosensors for Analysis of Neurotransmitter Release and Metabolism', *J. Neurosci. Methods*, **34**, 53–65 (1990).
43. J.W. Commissiong, 'Monoamine Metabolites: Their Relationship and Lack of Relationship to Monoaminergic Neuronal Activity', *Biochem. Pharmacol.*, **34**, 1127–1131 (1985).
44. R.M. Wightman, D.S. Brown, W.G. Kuhr, R.L. Wilson, 'Molecular Specificity of In Vivo Electrochemical Measurements', in *Voltammetry in the Neurosciences*, ed. J.B. Justice, Humana Press, Totowa, NJ, 103–138, 1987.
45. A.G. Ewing, J.C. Bigelow, R.M. Wightman, 'Direct In Vivo Monitoring of Dopamine Released from Two Striatal Compartments in the Rat', *Science*, **221**, 169–171 (1983).
46. J.A. Stamford, Z.L. Kruk, J. Millar, R.M. Wightman, 'Striatal Dopamine Uptake in the Rat: In Vivo Analysis by Fast Cyclic Voltammetry', *Neurosci. Lett.*, **51**, 133–138 (1984).
47. W.G. Kuhr, R.M. Wightman, 'Real-time Measurement of Dopamine Release in Rat Brain', *Brain Res.*, **381**, 168–171 (1986).
48. P.A. Garris, E.D. Ciolkowski, P. Pastore, R.M. Wightman, 'Efflux of Dopamine from the Synaptic Cleft in the Nucleus Accumbens of the Rat Brain', *J. Neurosci.*, **14**, 6084–6093 (1994).
49. J.B. Zimmerman, R.M. Wightman, 'Simultaneous Electrochemical Measurements of Oxygen and Dopamine In Vivo', *Anal. Chem.*, **63**, 24–28 (1991).
50. J. Millar, M. Armstrong-James, Z.L. Kruk, 'Polarographic Assay of Iontophoretically Applied Dopamine and Low-noise Unit Recording Using a Multibarrel Carbon Fibre Microelectrode', *Brain Res.*, **205**, 419–424 (1981).
51. J.A. Stamford, P. Palij, C. Davidson, C.M. Jorm, J. Millar, 'Simultaneous ''Real-time'' Electrochemical and Electrophysiological Recording in Brain Slices with a Single Carbon-fibre Microelectrode', *J. Neurosci. Methods.*, **50**, 279–290 (1993).
52. J.O. Schenk, E. Miller, M. Rice, R.N. Adams, 'Chronoamperometry in Brain Slices: Quantitative Evaluations of In Vivo Electrochemistry', *Brain Res.*, **277**, 1–8 (1983).
53. D.N. Middlemiss, P.H. Hutson, 'Measurement of the In Vitro Release of Endogenous Monoamine Neurotransmitters as a Means of Identification of Prejunctional Receptors', *J. Neurosci. Methods*, **34**, 23–28 (1990).
54. M. Rice, A.F. Oke, C.W. Bradberry, R.N. Adams, 'Simultaneous Voltammetric and Chemical Monitoring of Dopamine Release In Situ', *Brain Res.*, **340**, 151–155 (1985).
55. D.R. Bull, P. Palij, M.J. Sheehan, J. Millar, J.A. Stamford, Z.L. Kruk, P.P.A. Humphrey, 'Application of Fast Cyclic Voltammetry to Measurement of Electrically Evoked Dopamine Overflow from Brain Slices In Vitro', *J. Neurosci. Methods*, **32**, 37–44 (1990).
56. J.J. O'Connor, Z.L. Kruk, 'Fast Cyclic Voltammetry can be used to Measure Stimulated Endogenous 5-Hydroxytryptamine Release in Untreated Rat Brain Slices', *J. Neurosci. Methods*, **38**, 25–33 (1991).
57. J.A. Stamford, P. Palij, C. Davidson, C.M. Jorm, P.E.M. Phillips, 'Fast Cyclic Voltammetry in Brain Slices', in *Neuromethods, Voltammetric Methods in Brain Systems*, eds. A.B. Boulton, G.B. Baker, R.N. Adams, Humana Press, Totowa, NJ, 81–116, Vol. 27, 1995.
58. F. Crespi, C. Pietra, 'Middle Cerebral Artery Occlusion Alters Neurotransmitter Activities in Ipsilateral and Contralateral Rat Brain Regions: An Ex Vivo Voltammetric Study', *Neurosci. Lett.*, **230**, 77–80 (1997).
59. R.S. Kelly, R.M. Wightman, 'Detection of Dopamine Overflow and Diffusion with Voltammetry in Slices of Rat Brain', *Brain Res.*, **423**, 79–87 (1987).
60. P. Palij, J.A. Stamford, 'Real-time Monitoring of Endogenous Noradrenaline Release in Rat Brain Slices Using Fast Cyclic Voltammetry: 3. Selective Detection of Noradrenaline Efflux in the Locus Coeruleus', *Brain Res.*, **634**, 275–282 (1994).
61. M.A. Bunin, C. Prioleau, R.B. Mailman, R.M. Wightman, 'Release and Uptake Rates of 5-Hydroxytryptamine in the Dorsal Raphe and Substantia Nigra Reticulata of the Rat Brain', *J. Neurochem.*, **70**, 1077–1087 (1998).
62. S. Cragg, M.E. Rice, 'Heterogeneity of Electrically Evoked Dopamine Release and Reuptake in Substantia Nigra, Ventral Tegmental Area, and Striatum', *J. Neurophysiol.*, **77**, 863–873 (1997).

63. J.J. O' Connor, Z.L. Kruk, 'Frequency Dependence of 5-HT Autoreceptor Function in Rat Dorsal Raphe and Suprachiasmatic Nuclei Studied Using Fast Cyclic Voltammetry', *Brain Res.*, **568**, 123–130 (1991).
64. P. Palij, J.A. Stamford, 'Real-time Monitoring of Endogenous Noradrenaline Release in Rat Brain Slices Using Fast Cyclic Voltammetry. 2. Operational Characteristics of the α_2 Autoreceptor in the Bed Nucleus of Stria Terminalis, Pars Ventralis', *Brain Res.*, **607**, 134–140 (1993).
65. C.C. Toner, J.A. Stamford, 'Sodium Channel Blockade Unmasks Two Temporally Distinct Mechanisms of Striatal Dopamine Release during Hypoxia/Hypoglycaemia In Vitro', *Neuroscience*, **81**, 999–1007 (1997).
66. C.C. Toner, J.A. Stamford, 'Involvement of N- and P/Q- but not L- or T-type Voltage-gated Calcium Channels in Ischaemia-induced Striatal Dopamine Release In Vitro', *Brain Res.*, **748**, 85–92 (1997).
67. B. Giros, M. Jaber, S.R. Jones, R.M. Wightman, M.G. Caron, 'Hyperlocomotion and Indifference to Cocaine and Amphetamine in Mice Lacking the Dopamine Transporter', *Nature*, **379**, 606–612 (1996).
68. S.R. Jones, R.R. Gainetdinov, R.M. Wightman, M.G. Caron, 'Mechanisms of Amphetamine Action Revealed in Mice Lacking the Dopamine Transporter', *J. Neurosci.*, **18**, 1979–1986 (1998).
69. D. Sulzer, T.K. Chen, Y.Y. Lau, H. Kristensen, S. Rayport, A.G. Ewing, 'Amphetamine Redistributes Dopamine from Synaptic Vesicles to the Cytosol and Promotes Reverse Transport', *J. Neurosci.*, **15**, 4102–4108 (1995).
70. Y.Y. Lau, J.B. Chien, D.K.Y. Wong, A.G. Ewing, 'Characterization of the Voltammetric Response at Intracellular Carbon Ring Electrodes', *Electroanalysis*, **3**, 87–95 (1991).
71. J.B. Chien, R.A. Wallingford, A.G. Ewing, 'Estimation of Free Dopamine in the Cytoplasm of the Giant Dopamine Cell of *Planorbis corneus* by Voltammetry and Capillary Electrophoresis', *J. Neurochem.*, **54**, 633–638 (1990).
72. A. Meulemans, B. Poulain, G. Baux, L. Tauc, D. Henzel, 'Micro Carbon Electrode for Intracellular Voltammetry', *Anal. Chem.*, **58**, 2088–2091 (1986).
73. T. Abe, Y.Y. Lau, A.G. Ewing, 'Characterization of Glucose Microsensors for Intracellular Measurements', *Anal. Chem.*, **64**, 2160–2163 (1992).
74. T. Abe, Y.Y. Lau, A.G. Ewing, 'Intracellular Analysis with an Immobilized-enzyme Glucose Electrode Having a 2-μm Diameter and Subsecond Response Times', *J. Am. Chem. Soc.*, **113**, 7421–7423 (1991).
75. W.R. LaCourse, D.A. Mead, D.C. Johnson, 'Anion-exchange Separation of Carbohydrates with Pulsed Amperometric Detection Using a pH-selective Reference Electrode', *Anal. Chem.*, **62**, 220–224 (1990).
76. G.G. Neuburger, D.C. Johnson, 'Pulsed Coulometric Detection with Automatic Rejection of Background Signal in Surface-oxide-catalyzed Anodic Detections at Gold Electrodes in Flow-through Cells', *Anal. Chem.*, **60**, 2288–2293 (1988).
77. R.A. Wallingford, A.G. Ewing, 'Capillary Zone Electrophoresis with Electrochemical Detection', *Anal. Chem.*, **59**, 1762–1766 (1987).
78. T.M. Olefirowicz, A.G. Ewing, 'Dopamine Concentration in the Cytoplasmic Compartment of Single Neurons Determined by Capillary Electrophoresis', *J. Neurosci. Methods*, **34**, 11–15 (1990).
79. D.J. Leszczyszyn, J.A. Jankowski, O.H. Viveros, E.J. Diliberto, Jr, J.A. Near, R.M. Wightman, 'Nicotinic Receptor-mediated Catecholamine Secretion from Individual Chromaffin Cells: Chemical Evidence for Exocytosis', *J. Biol. Chem.*, **265**(25), 14736–14737 (1990).
80. R.M. Wightman, J.A. Jankowski, R.T. Kennedy, K.T. Kawagoe, T.J. Schroeder, D.J. Leszczyszyn, J.A. Near, E.J. Diliberto, Jr, O.H. Viveros, 'Temporally Resolved Catecholamine Spikes Correspond to Single Vesicle Release from Individual Chromaffin Cells', *Proc. Natl. Acad. Sci. USA*, **88**, 10754–10758 (1991).
81. D.J. Leszczyszyn, J.A. Jankowski, O.H. Viveros, E.J. Diliberto, Jr, J.A. Near, R.M. Wightman, 'Secretion of Catecholamines from Individual Adrenal Medullary Chromaffin Cells', *J. Neurochem.*, **56**, 1855–1863 (1991).
82. R.H. Chow, L. von Rüden, E. Neher, 'Delay in Vesicle Fusion Revealed by Electrochemical Monitoring of Single Secretory Events in Adrenal Chromaffin Cells', *Nature*, **356**, 60–63 (1992).
83. E. Neher, A. Marty, 'Discrete Changes of Cell Membrane Capacitance Observed Under Conditions of Enhanced Secretion in Bovine Adrenal Chromaffin Cells', *Proc. Natl Acad. Sci. USA*, **79**, 6712–6716 (1982).
84. K. Pihel, E.R. Travis, R. Borges, R.M. Wightman, 'Exocytotic Release from Individual Granules Exhibits Similar Properties at Mast and Chromaffin Cells', *Biophys. J.*, **71**, 1633–1640 (1996).
85. R. Borges, E.R. Travis, S.E. Hochstetler, R.M. Wightman, 'Effects of External Osmotic Pressure on Vesicular Secretion from Bovine Adrenal Medullary Cells', *J. Biol. Chem.*, **272**, 8325–8331 (1997).
86. J.A. Jankowski, J.M. Finnegan, R.M. Wightman, 'Extracellular Ionic Composition Alters Kinetics of Vesicular Release of Catecholamines and Quantal Size During Exocytosis at Adrenal Medullary Cells', *J. Neurochem.*, **63**, 1739–1747 (1994).
87. Z. Zhou, S. Misler, R.H. Chow, 'Rapid Fluctuations in Transmitter Release from Single Vesicles in Bovine Adrenal Chromaffin Cells', *Biophys. J.*, **70**, 1543–1552 (1996).
88. R.M. Wightman, T.J. Schroeder, J.M. Finnegan, E.L. Ciolkowski, K. Pihel, 'Time Course of Release of Catecholamines from Individual Vesicles during Exocytosis at Adrenal Medullary Cells', *Biophys. J.*, **68**, 383–390 (1995).

89. A.F. Oberhauser, I.M. Robinson, J.M. Fernandez, 'Simultaneous Capacitance and Amperometric Measurements of Exocytosis: A Comparison', *Biophys. J.*, **71**, 1131–1139 (1996).
90. A. Albillos, G. Dernick, H. Horstmann, W. Almers, G. Alvarez de Toledo, M. Lindau, 'The Exocytotic Event in Chromaffin Cells Revealed by Patch Amperometry', *Nature*, **389**, 509–512 (1997).
91. T.J. Schroeder, R. Borges, J.M. Finnegan, K. Pihel, C. Amatore, R.M. Wightman, 'Temporally Resolved, Independent Stages of Individual Exocytotic Secretion Events', *Biophys. J.*, **70**, 1061–1068 (1996).
92. J.A. Jankowski, T.J. Schroeder, R.W. Holz, R.M. Wightman, 'Quantal Secretion of Catecholamines Measured from Individual Bovine Adrenal Medullary Cells Permeabilized with Digitonin', *J. Biol. Chem.*, **267**, 18329–18335 (1992).
93. J.M. Finnegan, R.M. Wightman, 'Correlation of Real-time Catecholamine Release and Cytosolic Ca^{2+} at Single Bovine Chromaffin Cells', *J. Biol. Chem.*, **270**, 5353–5359 (1995).
94. I.M. Robinson, J.M. Finnegan, J.R. Monck, R.M. Wightman, J.M. Fernandez, 'Colocalization of Calcium Entry and Exocytotic Release Sites in Adrenal Chromaffin Cells', *Proc. Natl Acad. Sci. USA*, **92**, 2474–2478 (1995).
95. R.H. Chow, J. Klingauf, E. Neher, 'Time Course of Ca^{2+} Concentration Triggering Exocytosis in Neuroendocrine Cells', *Proc. Natl Acad. Sci. USA*, **91**, 12765–12769 (1994).
96. A. Elhamdani, Z. Zhou, C.R. Artalejo, 'Timing of Dense-core Vesicle Exocytosis Depends on the Facilitation L-type Ca Channel in Adrenal Chromaffin Cells', *J. Neurosci.*, **18**, 6230–6240 (1998).
97. K.L. Engisch, N.I. Chernevskaya, M.C. Nowycky, 'Short-term Changes in the Ca^{2+}–Exocytosis Relationship during Repetitive Pulse Protocols in Bovine Adrenal Chromaffin Cells', *J. Neurosci.*, **17**, 9010–9025 (1997).
98. T.J. Schroeder, J.A. Jankowski, J. Senyshyn, R.W. Holz, R.M. Wightman, 'Zones of Exocytotic Release on Bovine Adrenal Medullary Cells in Culture', *J. Biol. Chem.*, **269**(25), 17215–17220 (1994).
99. R. Zhou, G. Luo, A.G. Ewing, 'Direct Observation of the Effect of Autoreceptors on Stimulated Release of Catecholamines from Adrenal Cells', *J. Neurosci.*, **14**, 2402–2407 (1994).
100. A.G. Ewing, T.K. Chen, G. Chen, 'Voltammetric and Amperometric Probes for Single-cell Analysis', in *Neuromethods, Voltammetric Methods in Brain Systems*, eds. A.B. Boulton, G.B. Baker, R.N. Adams, Humana Press, Totowa, NJ, 269–304, Vol. 27, 1995.
101. R.M. Wightman, J.M. Finnegan, K. Pihel, 'Monitoring Catecholamines at Single Cells', *Trends Anal. Chem.*, **14**, 154–158 (1995).
102. G. Chen, A.G. Ewing, 'Chemical Analysis of Single Cells and Exocytosis', *Crit. Rev. Neurobiol.*, **11**, 59–90 (1997).
103. R.A. Clark, A.G. Ewing, 'Quantitative Measurements of Released Amines from Individual Exocytosis Events', *Mol. Neurobiol.*, **15**, 1–16 (1997).
104. R.A. Clark, S.E. Zerby, A.G. Ewing, 'Electrochemistry in Neuronal Microenvironments', in *Electroanalytical Chemistry a Series of Advances*, eds. A.J. Bard, I. Rubinstein, Marcel Dekker, New York, 227–294, 1998.
105. L.A. Greene, A.S. Tischler, 'Establishment of a Noradrenergic Clonal Line of Rat Adrenal Pheochromocytoma Cells which Respond to Nerve Growth Factor', *Proc. Natl Acad. Sci. USA*, **73**, 2424–2428 (1976).
106. D. Schubert, F.G. Klier, 'Storage and Release of Acetylcholine by a Clonal Cell Line', *Proc. Natl Acad. Sci. USA*, **74**, 5184–5188 (1977).
107. J.A. Wagner, 'Structure of Catecholamine Secretory Vesicles from PC12 Cells', *J. Neurochem.*, **45**, 1244–1253 (1985).
108. T.K. Chen, G. Luo, A.G. Ewing, 'Amperometric Monitoring of Stimulated Catecholamine Release from Rat Pheochromocytoma (PC12) Cells at the Zeptomole Level', *Anal. Chem.*, **66**(19), 3031–3035 (1994).
109. R. Levi-Montalcini, P.U. Angeletti, 'Nerve Growth Factor', *Physiol. Rev.*, **48**, 534–569 (1968).
110. S.E. Zerby, A.G. Ewing, 'Electrochemical Monitoring of Individual Exocytotic Events from the Varicosities of Differentiated PC12 Cells', *Brain Res.*, **712**, 1–10 (1996).
111. W.B. Stallcup, 'Sodium and Calcium Fluxes in a Clonal Nerve Cell Line', *J. Physiol.*, **286**, 525–540 (1979).
112. M.J. Berridge, R.F. Irvine, 'Inositol Trisphosphate, a Novel Second Messenger in Cellular Signal Transduction', *Nature*, **312**, 315–321 (1984).
113. S.E. Zerby, A.G. Ewing, 'The Latency of Exocytosis Varies with the Mechanism of Stimulated Release in PC12 Cells', *J. Neurochem.*, **66**, 651–657 (1996).
114. P.S. Chowdhury, X. Guo, T.D. Wakade, D.A. Przywara, A.R. Wakade, 'Exocytosis from a Single Rat Chromaffin Cell by Cholinergic and Peptidergic Neurotransmitters', *Neuroscience*, **59**, 1–5 (1994).
115. K.D. Kozminski, D.A. Gutman, V. Davila, D. Sulzer, A.G. Ewing, 'Voltammetric and Pharmacological Characterization of Dopamine Release from Single Exocytotic Events at Rat Pheochromocytoma (PC12) Cells', *Anal. Chem.*, **70**, 3123–3130 (1998).
116. E.N. Pothos, S. Przedborski, V. Davila, Y. Schmitz, D. Sulzer, 'D_2-like Dopamine Autoreceptor Activation Reduces Quantal Size in PC12 Cells', *J. Neurosci.*, **18**, 5575–5585 (1998).
117. E. Pothos, M. Desmond, D. Sulzer, 'L-3, 4-Dihydroxyphenylalanine Increases the Quantal Size of Exocytotic Dopamine Release In Vitro', *J. Neurochem.*, **66**, 629–636 (1996).
118. G.A. deToledo, R. Fernández-Chacón, J.M. Fernández, 'Release of Secretory Products During Transient Vesicle Fusion', *Nature*, **363**, 554–558 (1993).

119. K. Pihel, S. Hsieh, J.W. Jorgenson, R.M. Wightman, 'Electrochemical Detection of Histamine and 5-Hydroxytryptamine at Isolated Mast Cells', *Anal. Chem.*, **67**, 4514–4521 (1995).
120. H.P. Meissner, 'Membrane Potential Measurements in Pancreatic Beta Cells with Intracellular Microelectrodes', *Methods Enzymol.*, **192**, 235–246 (1990).
121. R.T. Kennedy, L. Huang, M.A. Atkinson, P. Dush, 'Amperometric Monitoring of Chemical Secretions from Individual Pancreatic β-Cells', *Anal. Chem.*, **65**, 1882–1887 (1993).
122. L. Huang, H. Shen, M.A. Atkinson, R.T. Kennedy, 'Detection of Exocytosis at Individual Pancreatic β-Cells by Amperometry at a Chemically Modified Microelectrode', *Proc. Natl Acad. Sci. USA*, **92**, 9608–9612 (1995).
123. C.D. Paras, R.T. Kennedy, 'Electrochemical Detection of Exocytosis at Single Rat Melanotrophs', *Anal. Chem.*, **67**, 3633–3637 (1995).
124. G. Chen, P.F. Gavin, G. Luo, A.G. Ewing, 'Observation and Quantitation of Exocytosis from the Cell Body of a Fully Developed Neuron in *Planorbis corneus*', *J. Neurosci.*, **15**, 7747–7755 (1995).
125. G. Chen, A.G. Ewing, 'Multiple Classes of Catecholamine Vesicles Observed During Exocytosis from the Planorbis Cell Body', *Brain Res.*, **701**, 167–174 (1995).
126. D. Bruns, R. Jahn, 'Real-time Measurement of Transmitter Release from Single Synaptic Vesicles', *Nature*, **377**, 62–65 (1995).
127. B. Powell, G.A. Cottrell, 'Dopamine in an Identified Neuron of *Planorbis corneus*', *J. Neurochem.*, **22**, 605–606 (1974).
128. J.M. Finnegan, K. Pihel, P.S. Cahill, L. Huang, S.E. Zerby, A.G. Ewing, R.T. Kennedy, R.M. Wightman, 'Vesicular Quantal Size Measured by Amperometry at Chromaffin, Mast, Pheochromocytoma, and Pancreatic β-Cells', *J. Neurochem*, **66**, 1914–1923 (1996).
129. G. Chen, D.A. Gutman, S.E. Zerby, A.G. Ewing, 'Electrochemical Monitoring of Bursting Exocytotic Events from the Giant Dopamine Neuron of *Planorbis corneus*', *Brain Res.*, **733**, 119–124 (1996).
130. Z. Zhou, S. Misler, 'Amperometric Detection of Stimulus-induced Quantal Release of Catecholamines from Cultured Superior Cervical Ganglion Neurons', *Proc. Natl Acad. Sci. USA*, **92**, 6938–6942 (1995).
131. E.N. Pothos, V. Davila, D. Sulzer, 'Presynaptic Recording of Quanta from Midbrain Dopamine Neurons and Modulation of the Quantal Size', *J. Neurosci.*, **18**, 4106–4118 (1998).
132. E.H. Jaffe, A. Marty, A. Schulte, R.H. Chow, 'Extrasynaptic Vesicular Transmitter Release from the Somata of Substantia Nigra Neurons in Rat Midbrain Slices', *J. Neurosci.*, **18**, 3548–3553 (1998).
133. L. Huang, R.T. Kennedy, 'Exploring Single-cell Dynamics Using Chemically-modified Microelectrodes', *Trends Anal. Chem.*, **14**, 158–164 (1995).
134. T. Malinski, Z. Taha, S. Grunfeld, A. Burewicz, P. Tomboulian, F. Kiechle, 'Measurements of Nitric Oxide in Biological Materials Using a Porphyrinic Microsensor', *Anal. Chim. Acta*, **279**, 135–140 (1993).
135. F. Lantoine, S. Trevin, F. Bedioui, J. Devynck, 'Selective and Sensitive Electrochemical Measurement of Nitric Oxide in Aqueous Solution: Discussion and New Results', *J. Electroanal. Chem.*, **392**, 85–89 (1995).

Organic Electrochemical Mechanisms

Claude P. Andrieux
Université Paris 7 – Denis Diderot, Paris, France

1 Introduction **1**
2 Electrochemical Techniques Used for the Determination of Organic Electrochemical Mechanisms **2**
2.1 Cells and Electrodes 2
2.2 Stationary Techniques 3
2.3 Transient Techniques 4
2.4 Comparison of the Time Windows 6
3 Characteristics of Electron Transfers for Organic Compounds **9**
3.1 Theory 9
3.2 Characterization Procedures 10
3.3 Some Results 12
4 Simple Organic Electrochemical Mechanisms **13**
4.1 Nature and Classification of Coupled Chemical Reactions 13
4.2 Kinetic Analysis for a Single Chemical Reaction Acting as the Determining Step 14
5 Complex Mechanisms **16**
5.1 Sequence of Chemical Reactions After One Electron Transfer 16
5.2 Two-electron Process: Electron Transfer at the Electrode or in Solution 17
5.3 Redox Catalysis 18
6 Examples of Mechanism Analysis and Rate Constant Determination **19**
6.1 Kolbe Reaction 20
6.2 Reduction of Halo Compounds 21
6.3 Organic Electrodimerization 23
6.4 Chemical Reactions Induced by Electrochemistry 23
7 Conclusion **25**
Acknowledgments **25**
Abbreviations and Acronyms **25**
Related Articles **25**
References **26**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The determination of electrochemical mechanisms by electrochemical techniques is very efficient in organic electrochemistry. In contrast to the electrochemistry of small molecules and ions, the electron transfers on organic compounds are mainly outer-sphere electron transfers following a Marcus–Hush law. Hence these fast electron transfers open up the possibility of obtaining kinetic limitation by the associated homogeneous chemical reactions.

The classical methods for these determinations are stationary techniques such as rotating disk electrode voltammetry (RDEV) and polarography but also more powerful methods, involving large-amplitude transient methods: linear sweep voltammetry (LSV), cyclic voltammetry (CV) and double-step chronoamperometry (DSC) allow the time window to be extended to thousands of microseconds; 1 μs can be reached with the use of ultramicroelectrodes. When the chemical reaction is the rate-determining step, the response, i.e. current as a function of time (or potential), provides information about the nature of this reaction and in many case allows the measurement of its rate constant. Characteristic behaviors are detailed for the principal chemical reactions preceding (C + E) or following (E + C) the electron transfer.

Organic electrochemical mechanisms often involve several electron transfers. The influence of homogeneous electron transfers is underlined in particular for two-electron reductions or oxidations and also in redox catalysis. The possibility of an electron transfer concerted with a chemical reaction is characterized. Some examples of mechanisms and kinetic determinations are given in the field of widely used organic electrosynthetic processes.

1 INTRODUCTION

An organic electrochemical process involves electron transfers on an organic molecule. Such processes are possible with a great number of organic compounds and are usually conducted in an electrolytic cell where electron transfers are located at the interface between an electrode and a conducting solution containing the compound to be reduced (electron addition) or to be oxidized (electron removal).

In the field of organic chemistry, a simple electron transfer never occurs alone, with the exception of the electrochemical preparation of stable free radicals. An organic electrochemical mechanism is an association of electron transfers and chemical reactions. Hence an electrochemical reaction appears as a sequence of three kinds of simple phenomena:

- electron transfer on organic species at an electrode surface; these species are usually in the solution but

could be located in somes cases in a particular layer (adsorption layer, for example) near the electrode;

- chemical reactions from organic species produced after one or several electron transfers; these reactions can involve other reactants present in the electrolytic solution;
- transport of reactants, intermediates and products from and to the electrode surface.

This article deals with the description of the sequence of electron transfers and chemical reactions for organic species when particular conditions of potential or current are fixed on an electrode. The nature of this electrode, the solvent used and the other reactants in the medium are important parameters in allowing a particular global electrochemical reaction to occur from the reactant to one or several products. Some books deal with the more interesting processes in this field.[1–3]

The purpose here is mainly to describe how it is possible to determine this sequence (or some part of this sequence) from electrochemical measurements by analyzing the electrical response when using electrochemical techniques at the level of small electrodes (microelectrodes) set in the initial solution. As described in several textbooks,[4,5] the kinetics of the overall process are the result of the kinetics of the above three kinds of phenomena, but in many cases the global kinetics are a function of only a few key steps. The analysis of the electrical response on the microelectrode, associated with the description of the final products, is an efficient tool in the determination of this mechanism.

2 ELECTROCHEMICAL TECHNIQUES USED FOR THE DETERMINATION OF ORGANIC ELECTROCHEMICAL MECHANISMS

2.1 Cells and Electrodes

As organic compounds are usually insoluble in water, most electrochemical studies of organic compounds are carried out in solvents other than water. These solvents must also permit the dissolution of organic species and of some strong electrolytes (supporting electrolytes) at about 0.1 or 1 M concentrations, leading to an electrolytic solution with sufficiently good conductivity. Thus, solvents with a dielectric constant ε_r in the range 20–50 are the most commonly used, e.g. acetonitrile $\varepsilon_r = 37.4$, dimethylformamide (DMF) $\varepsilon_r = 36.7$ and dimethyl sulfoxide (DMSO) $\varepsilon_r = 46.7$ as aprotic solvents and methanol $\varepsilon_r = 32.6$ and ethanol $\varepsilon_r = 24.3$ as protic organic solvents, the latter able to be used pure or mixed with water.[6] The possibilities opened up by the addition of a surfactant to an aqueous solution that allows the solubilization of organic compounds have often been described and pose a particular challenge at the level of industrial preparation.[7]

Because these electrolytes must remain inert within the potential range of interest, tetraalkylammonium or lithium cations are used when a large potential range is necessary in reduction; perchlorate, tetrafluoroborate and hexafluorophosphate are used when a large potential range is necessary in oxidation.

For the electrode material, the same metallic material must be used for the analytical probe and for the preparative electrolysis, so a wide number of conducting materials are used. For reduction processes, mercury or lead are often used because they allow a high hydrogen overvoltage. In oxidation processes the inertness to most chemical environments calls for the use of platinum or glassy carbon. The latter allow both oxidation and reduction.[1–3,8]

From an analytical point of view, a three-electrode device cell is generally used. The electrical potential of the working electrode is controlled regardless of what current is flowing through it, independent of the reactions occurring at the other electrode (counter electrode or auxiliary electrode) by a potentiostat.[5,9] This electronic device, built from operational amplifiers, assumes a potential difference automatically adjusted between the working and the auxiliary electrodes in such a way that the potential difference between the working electrode and the reference electrode is fixed at the desired value. The auxiliary electrode is commonly an inert, larger sized electrode compared with the working electrode.

The reference electrode should satisfy the following requirements: its potential should remain constant with time, and if small currents are passed through it, they must have little effect on its potential, with an immediate return to the initial value upon elimination of polarization. Therefore, its impedance must be high and fast electron transfer between two stable species is necessary.

The most common reference electrode is a saturated calomel electrode (SCE) in water. In fact, it is very often used in all media provided that a salt-conducting separator is introduced between the two solutions. In an aprotic medium, other reference electrodes are used, e.g. silver–silver ions in acetonitrile or silver–silver chloride.[10] The normal hydrogen electrode is never used in practical experiments.

According to readings from the potentiostat and from the electrochemical method used, it is possible to indicate suitable dimensions of the working electrode. For classical methods, millimeter-sized electrodes are used, but for very fast transient methods the dimensions can reach 1 μm. In order to interpret the electrochemical response,

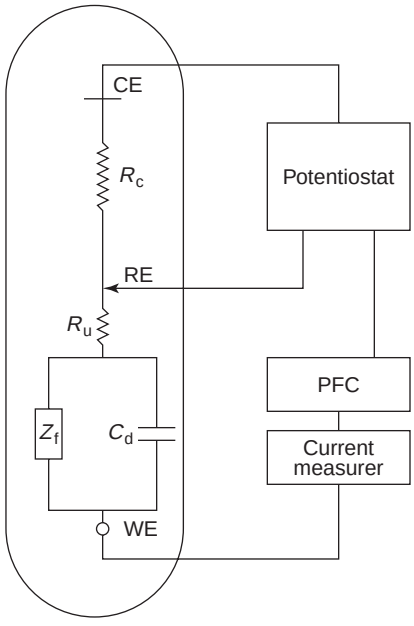

Figure 1 Scheme of the equivalent electrical circuit of an electrochemical cell and its control of potential by a three-electrode potentiostat with a positive feedback compensation (PFC). WE, CE and RE are the working, counter and reference electrodes, R_u and R_c the cell resistances, Z_f the faradaic impedance and C_d the double-layer capacitance.

an analysis of the electrical circuit of an electrochemical cell with its connection to the potentiostat must be given (Figure 1).[4,5,11] The effect of the double-layer capacitance will be discussed later. Recent potentiostats permit the minimization (without being able to suppress completely the ohmic drop for stability reasons) of the value of the uncompensated resistance R_u by a positive feedback compensation.[12] A current measurer is also associated with the electronic device. The readings of these electronic devices are a function of the technique used. If rapidly varying working potentials are involved, the potentiostat response has to be short and requires the use of large bandpass amplifiers.

It must be noted that although in the preparative scale-up some separation between the working compartment and the auxiliary compartment often occurs, this separation is not necessarily at the level of the mechanism determination since the concentration perturbation stays in the vicinity of the working electrode and therefore the reactions at the auxiliary electrode do not affect the process. Consequently, the analysis that is developed here is not directly applicable to coupled reduction and oxidation reactions where products produced on one electrode interfere with the electrochemical reactions on the other electrode.

In order to describe the different techniques that are used, we shall first describe, briefly, the electrical response for a single, fast, one-electron transfer neglecting secondary phenomena and more specifically ohmic drop and double-layer capacitance. Fast electron transfer[5,13] means that the kinetics of the electron transfer do not interfere with other limiting kinetics and that we can apply the Nernst law to this simple electron transfer (Equations 1 and 2):

$$A + e \rightleftharpoons B \tag{1}$$

$$E = E^0 + \frac{RT}{F} \ln\left(\frac{C_A^0}{C_B^0}\right) \tag{2}$$

where C_A^0 and C_B^0 are the concentrations of A and B at the electrode, E is the electrode potential and E^0 is the standard potential of the couple A/B. The initial concentration of only electroactive species A present in solution is C^0, with an order of magnitude for this concentration of 10^{-3} M. The example here is of a reductive process but transposition to oxidation is easy.

Because we are assuming that no chemical complications are involved, only transport phenomena are kinetically determinant. In fact, the transport phenomenon is mainly diffusion and under certain conditions convection for the reactants and the products. This is because in the solution as a large excess of supporting electrolyte is present (the concentration C^0 remains very low compared with the concentration of electrolyte), so the migration of the charged species involved in the electrode reaction is restricted to a very narrow double-layer region (a few ångstroms) and, as seen below, very narrow compared with the diffusion layer.[5,14] Hence we can neglect this migration effect for all the described techniques and the diffusion phenomenon can be explained by a simple relationship according to Fick's law.

2.2 Stationary Techniques

RDEV[5,14,15] is a method where the working electrode is a small conducting disk with a surface ranging from 1 to 10 mm^2. Thus a large amount of material can be used. Laminar flow is involved for moderate rotation speeds and the concentration profiles for the reactant and the product are calculated, taking into account the forced convection produced by the rotation of the electrode. The current at the electrode is dependent on potential but reaches a stationary value for any imposed potential (Figure 2a and b).

The Nernst approximation applies and we can replace the actual profile by a linear gradient in a diffusion layer (Equation 3):

$$\delta = 4.98\, D^{1/3} \nu^{1/6} w^{-1/2} \tag{3}$$

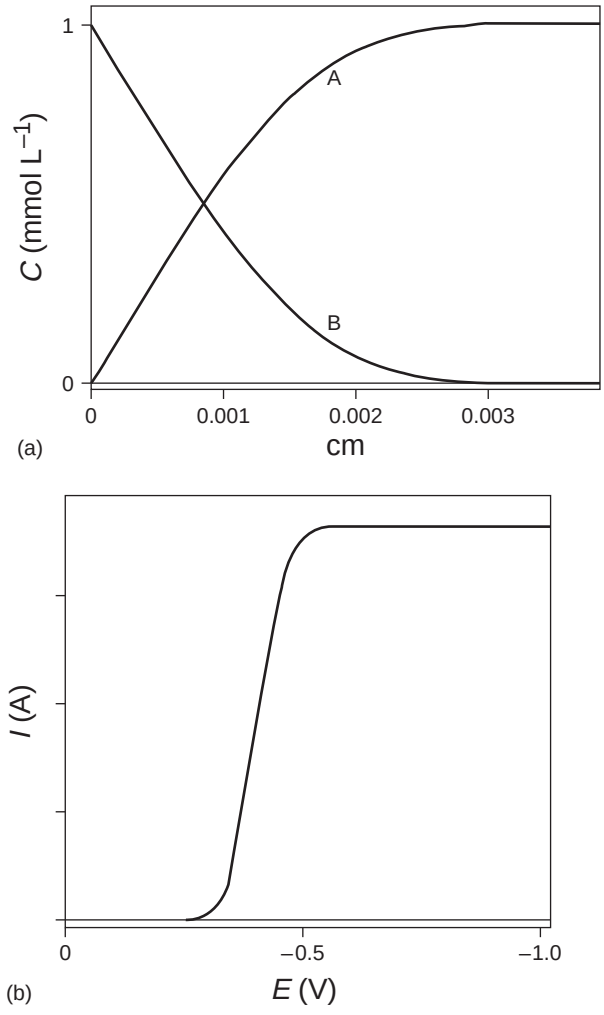

Figure 2 (a) Concentration profile for reactant A and product B in RDEV on the plateau current. (b) Current potential curve in RDEV for the reaction $A + e^- \rightleftharpoons B$ with $E^0 = -0.4\,V$.

where D is the diffusion coefficient (cm^2), v the kinematic viscosity ($cm^2\,s^{-1}$), w the angular rotation rate (rpm) and δ is expressed in centimeters. In this diffusion layer, a stationary profile is obtained and a linear variation of the two concentrations is observed. The possible rotation rate is between 50 and 20 000 rpm and the corresponding thickness of the diffusion layer is between 7×10^{-3} and 4×10^{-4} cm.

Polarography[5] is the oldest of all electrochemical techniques. A dropping mercury electrode serves as the working electrode, imposing severe limitations on reactions on this metal. The current law seems rather complex because this method involves a moving sphere, the surface area increasing until the drop falls. The current varies with time yet at the level of the maximum (or the average) current, a satisfactory approximation is a stationary regime in which the Nernst approximation (linear profiles) is applied, with a diffusion layer thickness in relation to the drop time θ (Equation 4):

$$\delta = \left(\frac{3\pi D\theta}{7}\right)^{1/2} \tag{4}$$

As the drop time is variable only within a narrow window (1–10 s), the thickness of the diffusion layer is only between 4×10^{-3} and 10^{-2} cm, leading to a very small variation in the experimental parameters.

With both stationary techniques, it is possible to scan the potential very slowly and thus obtain a current–potential curve (voltammetric curve). This gives a clear picture of the range of potential that must be applied in order to conduct the electrochemical reaction. The characteristic values of this plateau-shaped curve are the half-wave potential $E_{1/2} = E^0$ and plateau current $i_l = FSDC^0/\delta$.

As a complement to the RDEV and with the purpose of detecting the product obtained on the disk in the electrochemical reaction and to see if this intermediate is decomposed by a chemical reaction, rotating ring-disk electrode voltammetry (RRDEV) is used.[16] The working electrode is now a disk as in RDEV, surrounded by a concentric ring separated from the disk by an insulator. The gap between disk and ring must be small to ensure that species produced on the disk can easily reach the ring. The potentials of the disk and the ring are controlled independently. In the case of a simple electron exchange, if the potential of the ring is set at a value more positive than the reduction potential of A and if the potential of the disk is slowly changed from positive to negative as in an RDEV experiment, a wave of similar shape (with a plateau for a negative potential of the disk) is also obtained on the disk during this experiment. The current obtained on the ring is the current of reoxidation of B and the height of this wave depends on the geometric dimensions of the electrode but will always be smaller than the reduction current.

2.3 Transient Techniques

Two kinds of transient techniques are described. At the level of mechanism investigations, large-amplitude techniques are usually more adapted to the problem, since they have the same advantage as stationary techniques of easily visualizing the electrochemical process. These techniques also allow a large range for the time window.[11,17] LSV and CV are powerful tools in this area.

The working electrode is a stationary disk or sphere with an area in the range of 1 mm^2 and the working electrode potential is imposed as a linear function of time (LSV). Very often, the potential returns linearly to the initial value (CV). The sweep rate v is the most important

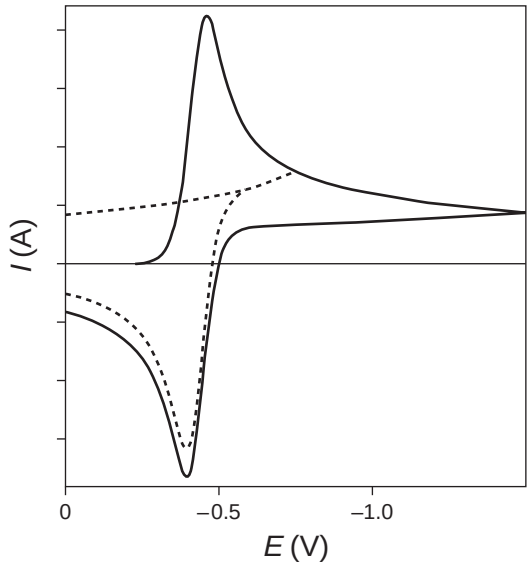

Figure 3 LSV and CV. Diffusion-controlled voltammogram (solid curve) and the measurement of the anodic part from the prolongation of the cathodic part (dashed curve).

parameter because it controls the rate of diffusion of the species at the electrode and thus plays a role analogous to the rotation speed in RDEV. This sweep rate can be varied between 0.1 and a few thousand $V\,s^{-1}$.

In the case of a simple electron transfer, when the potential is swept within the potential range where the electrochemical reaction is, at the beginning (initial potential), nonoperative and completely occurring (inversion potential), the curve shown in Figure 3 is obtained.

A convenient experimental characterization of the current–potential curve is measured according to the height, the location and the width of the peak for the first and second parts of the sweep.[11,17,18]

For the first scan, these cathodic peak characteristics for a single fast transfer are expressed by Equations (5–7):

Peak current: $$i_p = 0.446FSC^0D^{1/2}\left(\frac{Fv}{RT}\right)^{1/2} \quad (5)$$

Peak potential: $$E_p = E^0 - 1.11\left(\frac{RT}{F}\right) \quad (6)$$

Peak width: $$E_{p/2} - E_p = 2.20\left(\frac{RT}{F}\right) \quad (7)$$

where $E_{p/2}$ is the potential corresponding to a current intensity equal to half that peak current.

For the reverse part of the potential sweep, if the curve is calculated from the prolongation of the diffusion part of the first part, the same peak current (anodic current) and peak width are found with an anodic peak potential (Equation 8) of

$$(E_p)_2 = E^0 + 1.11\left(\frac{RT}{F}\right) \quad (8)$$

leading to a cathodic and anodic peak separation (Equation 9) of

$$\Delta E_p = 2.22\left(\frac{RT}{F}\right) \quad (9)$$

Other useful large-amplitude techniques in the field of organic electrochemical mechanism determination are potential step chronoamperometry (PSC) and DSC.[5,11,17] In PSC, the stationary electrode is stepped from a rest potential (no reduction of A) to a more negative value, markedly more negative than the standard potential E^0 so that the concentration of A at the electrode is zero. The cathodic current is only controlled by diffusion and can be expressed by Equation (10):

$$i = \frac{FSC^0D^{1/2}}{(\pi t)^{1/2}} \quad (10)$$

Moreover, in DSC, after a time θ, the potential is stepped back to the initial potential and, as for a CV experiment, the reoxidation curve is exactly the same as the cathodic curve if the current is measured starting from the prolongation of the cathodic curve (Figure 4).

An easy way to analyze the reoxidation curve is to measure the anodic current from the current axis at a time 2θ (double the inversion time). The ratio between the anodic and the cathodic currents for θ is 0.293.

Small-amplitude transient techniques are often used for the analytical aspects of these techniques.[19] For the purpose of determining the characteristics of the electron transfers, some aspects of square- or sinusoidal-superimposed potential [alternating current (ac) technique, also called the impedance technique][20]

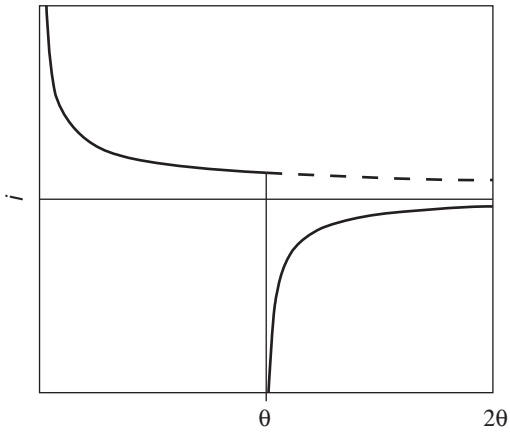

Figure 4 DSC. Current as a function of time with inversion time θ. Reconstruction of the anodic current from the prolongation of the cathodic part (dashed line).

have been developed with a range of frequency of 10–10 000 Hz. It has been claimed that the second-harmonic ac technique is very efficient.[21] These techniques consist of applying on the electrode a small-amplitude sinusoidal potential (frequency v) and a slow variation of a dc potential and detecting the amplitude and the phase of the first- (frequency v) or second-harmonic (frequency $2v$) current amplitude as a function of the dc potential. The two components in-phase and out-of-phase with the input potential are then separated. It must be noted that the second-harmonic ac technique contains the same limitation as the simple ac technique.[22] At the level of mechanism determinations for complex sequences, these techniques are very rarely used.

2.4 Comparison of the Time Windows

For stationary techniques, the time window is determined from the mean transport (diffusion) time from the solution to the electrode and thus is D/δ^2. The limitations are dependent on the hydrodynamic limitation with respect to the rotation rate in RDEV, as mentioned earlier, and dependent on the drop time in polarography.

For transient techniques, the limitations that occur are essentially due to secondary phenomena.[11] In the preceding sections, we have assumed that the current at the electrode surface results only from diffusion in solution when an electron transfer occurs between an electrode and the organic substrate (faradaic current). However, several side effects can disturb this assumption.

To start with, a double layer is always present at the electrode interface, which leads to an additional capacitor SC_d in parallel with the faradaic impedance, as indicated in Figure 1 (C_d is the capacitance per unit of area). At the level of the stationary techniques, this capacitor does not give any additional current, but for all the transient techniques its presence imposes severe limitations.

For example, in LSV, a constant capacitive current $i_c = SC_d v$ is obtained and it should be noted that this current is a function of the sweep rate v as compared with the faradaic current (that is a function the square root of v). Hence, for large values for v, the capacitive component increases as compared with the faradaic current. For a concentration of substrate of about 10^{-3} M, these two components have the same order of magnitude for a sweep rate of a few thousand volts per second.

As previously remarked, an uncompensated resistance between the working and reference electrodes remains in a three-electrode potentiostat; even when this potentiostat is equipped with feedback compensation, part of this resistance remains.[12] The total capacitive current is expressed by Equation (11):

$$i_c = SC_d v \quad 1 - \exp\left(\frac{-t}{R_u SC_d}\right) \qquad (11)$$

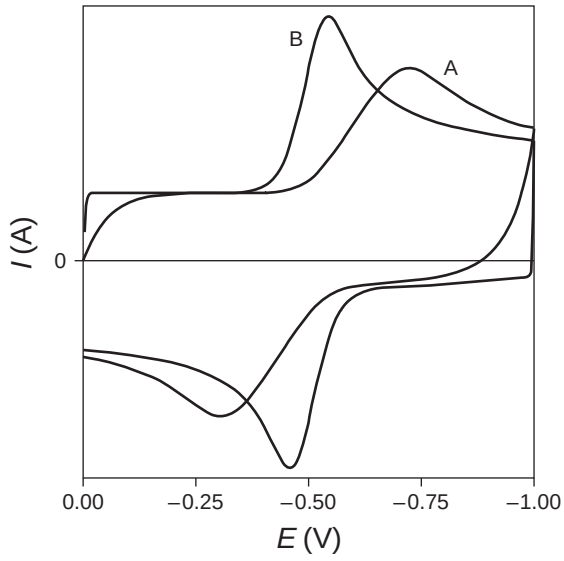

Figure 5 CV at $1000\,V\,s^{-1}$ (A) without positive feedback compensation, $R_u = 500\,\Omega$, and (B) with positive feedback compensation, $R_u = 20\,\Omega$ ($S = 0.01\,cm^2$, $C_d = 10\,mF\,cm^{-2}$, $D = 10^{-5}\,cm^2\,s^{-1}$).

The resulting voltammetry is represented in Figure 5. The duration of the increasing part of the capacitive current is approximated by $R_u SC_d$ and a faradaic signal can be accurately measured if the time of the experiment is long compared with $R_u SC_d$. The uncompensated resistance also leads to important modifications on the faradaic curve itself owing to an effective potential applied where the ohmic drop $R_u i$ must be taken into account. Let us estimate the order of magnitude of these different effects. If a potentiostat with no feedback compensation is used in an aprotic medium, R_u is about $500\,\Omega$, where $S = 0.01\,cm^2$, $C_d = 10\,mF\,cm^{-2}$ and $D = 10^{-5}\,cm^2\,s^{-1}$, using a sweep rate of $1000\,V\,s^{-1}$, then $i_F = 283\,mA$, $i_c = 100\,mA$, $R_u i = 161\,mV$ and $R_u SC_d = 50\,\mu s$. With a reduced resistance at $20\,\Omega$ by feedback compensation, the rising time becomes $2\,\mu s$ and the residual ohmic drop falls at 8 mV. This value is not sufficient for an accurate determination of peak potential, which would require 1–2 mV. However, by taking into account the possibility of approximating the ohmic drop via a simple calculation, since the residual resistance can be measured, we can determine this peak potential with the desired accuracy. It must be noted that a complete resistance compensation can never be obtained because when the resistance is reduced to a very small value, oscillations at the beginning of the capacitive current curve appear and it is therefore possible theoretically to show that for zero residual resistance a complete unstable system is involved.

For other transient techniques the same kind of limitation occurs, which prohibits a very short experiment time. For example, in PSC, even if the effect of ohmic

drop can theoretically be neglected because the step value in potential is chosen far enough away from the standard potential value, the effect is still present on the shape of the curve, mainly for short observation times. Moreover, the capacitive and the faradaic currents in this short-time experiment are high enough that the limitation for short times cannot be estimated merely in terms of R_uSC_d but by about 10 or 20 times these values before which the diffusion current is not affected by the capacitive current. These high currents also prevent the use of very efficient ohmic drop compensation. Thus, the short-time limitation for this technique is 1 ms without ohmic drop compensation and 0.1 ms otherwise.

Another limitation is the effect of nonlinear diffusion on the small disk or on the sphere acting as the working electrode. This effect appears for long-time values or for low sweep rates and it can be neglected on millimeter-sized electrodes down to 0.1 V s^{-1} (or up to 1 s in PSC). Thus the sweep rate window in LSV ranges from 0.1 to 1000 V s^{-1}. The time window corresponding to this sweep rate is RT/Fv, a value that is of the same order of magnitude as the time occurring between the standard potential and peak potential. Hence the time window for an LSV experiment on a millimeter-sized electrode ranges from 25 µs to 0.25 s.

Recently, a means of obtaining higher sweep rates was found using ultramicroelectrodes[23–25] provided that large-bandpass amplifiers were used. Disk electrodes between 1 and 10 µm are now commercially available for this purpose. As shown previously, three parameters must be minimized in order to shorten response times (or increase sweep rates): R_uSC_d, R_ui and i_c as compared with i_F. A possibility is to decrease the size of the electrode. In this way not only R_uSC_d is decreased but also the resistance near the working electrode (the important part of the resistance between the working and reference electrodes) is more or less a function of the inverse of the radius of this electrode. Thus R_u is decreased by a factor of 100 between a millimeter-sized and a 10-µm electrode. Figure 6 illustrates the possibility of the use of these ultramicroelectrodes, shown with a sweep rate up to 10^6 V s^{-1}.

The use of an ultramicroelectrode is not restricted to LSV; we can also use it in DSC. The gain obtained

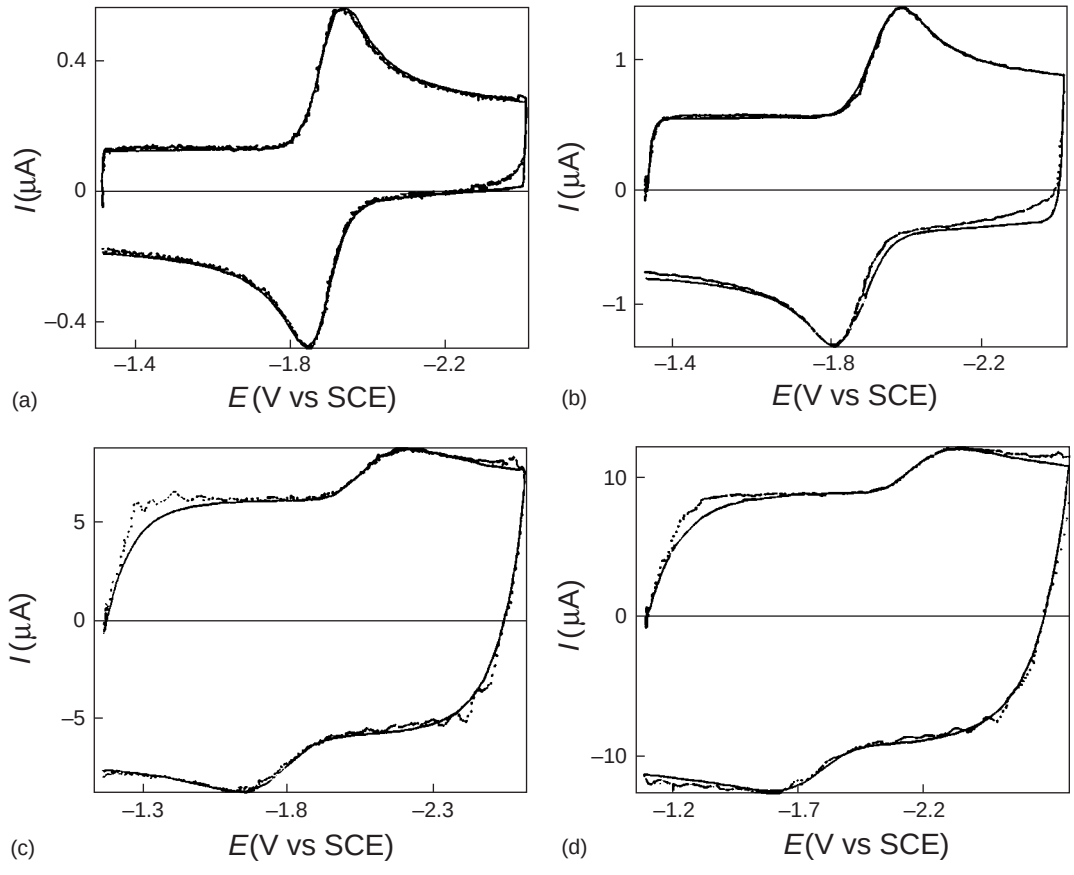

Figure 6 CV of anthracene (10 mM) in acetonitrile at a 5-µm diameter gold electrode. Scan rate: (a) 22 100, (b) 113 400, (c) 1 191 000 and (d) 1 724 000 V s^{-1}. [Reproduced by permission of the American Chemical Society from C.P. Andrieux, P. Hapiot, J.-M. Savéant, *Chem. Rev.*, **90**, 723–736 (1990).]

Table 1 Time window for stationary and transient techniques

Technique	Experimental parameter	Range of accessible values	Time window (s)
Polarography	Drop time	1–10 s	1–10
RDEV	Rotation speed w	50–20 000 rpm	2×10^{-2}–5
LSV (millimeter electrode)	Sweep rate	0.1–1000 V s^{-1}	3×10^{-5}–0.3
DSC (millimeter electrode)	Step time	1 s–1 ms	10^{-3}–1
Sinusoidal technique	Frequency	1–10 000 Hz	10^{-4}–1
LSV (ultramicroelectrode)	Sweep rate	10^{2}–10^{6} V s^{-1}	3×10^{-8}–3×10^{-4}
DSC (ultramicroelectrode)	Step time	10 ms–1 µs	10^{-6}–10^{-2}

compared with a millimeter-sized electrode is about two orders of magnitude, which leads to short time values of about 1 µs.

With an ultramicroelectrode, a limitation due to spherical diffusion occurs at a higher sweep rate (about 100 V s^{-1}) but does not present a real problem because at lower sweep rates a stationary response is obtained even if a stationary electrode is used. From this change in the diffusion regime, an elegant method for determining the exchange number of electrons in a complex mechanism is developed.[25]

Table 1 summarizes the various possibilities of the more commonly used methods in the analysis of organic electrochemical mechanisms.

In addition to the classical electrochemical methods, an effective method using photoinjection of electrons by means of a laser flash has been described,[26] as shown in Figure 7.

The advantage of this method is that it produces solvated electrons that react with a reducible neutral compound very close to the surface. Therefore, analysis of the chemical reaction in this region near the electrode is possible. The potential of the electrode can be fixed in a potential region beneath the reduction potential of the initial compound. The measurement is the residual photoinduced charge transferred from the electrode to the solution. If the intermediate product (anion radical) is stable, the response is similar to the response of the recombination of the free electrons produced in the solution without any reducible compound (the solvated electron and the radical anion are reoxidized at the electrode). However, if the intermediate compound cleaves very quickly, yielding an unoxidizable compound, the residual photoinduced charge increases. According to the analysis time for the residual photoinduced charge at the end of the laser flash, a first-order kinetic constant can be obtained between 2×10^{-3} and 10^{-8}–10^{-9} s^{-1}, leading to the shortest time window of only a few nanoseconds.

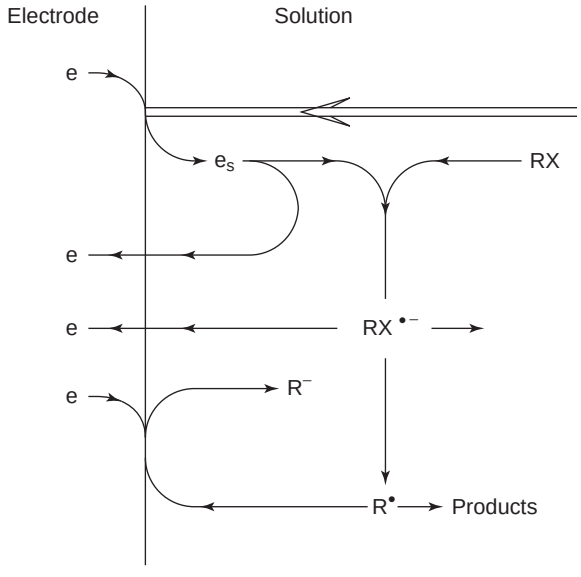

Figure 7 Principle of the photoinjection of electrons method in the case of the reduction of a substrate RX.

However, with the preceding performance, this method is restricted to an aqueous solution and even in DMF this application is difficult. On the other hand, this method allows the study of the second electron transfer when a very fast cleavage occurs at the level of the first step.[27] The reduction of radicals giving anions can be characterized via its standard potential and its kinetics of electron transfer even in an aprotic medium such as DMF. Some results are given in the following section. At the level of the determination of electrochemical properties of transient radicals, another effective method is used, photomodulation voltammetry (PMV),[28] which, yields either cations in oxidation or anions in reduction. If these radicals are destroyed by a dimerization process, the measurement is always possible even when this reaction reaches the diffusion limit.

In addition to the electrochemical methods described here, the determination of electrochemical reaction mechanisms is often assisted by spectrometric methods ex situ or in situ. In the latter case, spectroelectrochemical techniques with ultraviolet/visible (UV/VIS), infrared (IR) or electron spin resonance (ESR) spectroscopy are used. These techniques are described in other articles.

3 CHARACTERISTICS OF ELECTRON TRANSFERS FOR ORGANIC COMPOUNDS

Two characteristics of the electron transfer are of particular importance because they open up possibilities for the use of electrochemical techniques in mechanism determinations. The first is the weak adsorption of organic reactants, products and intermediates when aprotic solvents and classical inert electrodes are used. Hence most organic mechanisms deal only with homogeneous chemical reactions and electron transfer reactions at the electrode. The second is related to the kinetics of electron transfer in organic electrochemistry.

3.1 Theory

Electron transfers constitute one or several elementary steps of the mechanism of electrochemical reactions. As organic species are molecules of significant size, the Marcus–Hush theory[29,30] is especially adapted to the description of the kinetic laws of this transfer. This theory was developed in the case of an electron transfer occurring without any breaking or any formation of chemical bonds, still taking into account changes in the solvation (changes in orientation of solvent molecules near the molecule) and changes in bond lengths or angles. The theory of electron transfer at the electrode is completely associated with the theory of homogeneous electron transfer when the electron exchange occurs between two molecules differing only by one oxidation degree.[30–32]

From a theory similar to that concerning activated complexes, the electron transfer rate can be written for an adiabatic process, and according to the Franck–Condon principle, which states that the slow step involves the heavier particle, leading to kinetics determined by the changes at the level of the reactants and the product. The kinetics can be expressed using Equations (12) and (13):

$$k_S = Z_{el} \exp\left(\frac{-\Delta G^*}{kT}\right. \tag{12}$$

with

$$\Delta G^* = \left(\frac{\lambda}{4}\right.\left(1 + \frac{\Delta G^0}{\lambda}\right.^2 \tag{13}$$

where Z_{el} is the collision number for the heterogeneous reaction, ΔG^* the free energy of activation, ΔG^0 the standard free energy of the reaction ($\Delta G^0 = 0$ at the standard potential) and λ a reorganization term composed of solvational (λ_0) and internal (λ_i) components. The standard activation free energy (intrinsic barrier) for $\Delta G^0 = 0$ is then given by Equation (14):

$$(\Delta G^*)^0 = \frac{\lambda_0 + \lambda_i}{4} \tag{14}$$

In an outer-sphere electron transfer with small internal reorganization, so that the solvent reorganization term λ_0 is the largest, this term can be estimated from the ionic radii of reactant and product and a distance d that expresses the location of the electron exchange (Equation 15):

$$\lambda_0 = (\Delta e)^2 \left(\frac{1}{a} - \frac{1}{d}\right.\left(\frac{1}{\varepsilon_{op}} - \frac{1}{\varepsilon_S}\right. \tag{15}$$

where Δe is the change in charge between reactants and product, ε_{op} and ε_S are the optical and static dielectric constant of the solvent and the estimation of d is $2a$ according to Marcus[30] and $1/d = 0$ according to Hush.[29]

From this theory and for large-sized molecules such as aromatic compounds in aprotic solvents where only solvent reorganization is involved, the rate constant of the electron transfer can be estimated. For radii in the range 2–5 Å, values of k_S between 1 and 10 cm s^{-1} were estimated. It must be noted that the rate of the electron transfer is quadratic with the potential having a value of $\alpha(d\Delta G^*/d\Delta G^0)$ equal to 0.5 at the standard potential of the reaction.

The internal or vibrational term provides an important contribution when some bond lengths or bond angles are changed during the electron transfer. This is important, for example, in the reduction of an aliphatic radical in an anion.

It is also possible to estimate the kinetics of the electron transfer when this transfer is concerted with a bond cleavage (Equation 16):

$$A{-}X + e^- \rightleftharpoons A^\bullet + X^- \tag{16}$$

As with the Marcus–Hush theory, the Born–Oppenheimer approximation is assumed to be applicable and the reaction to be adiabatic. The potential energy surfaces for the reactants and products depend now on three kinds of reaction coordinates: (1) the solvent fluctuational configuration as in the simple outer-sphere electron transfer, (2) the vibration coordinates for bonds that are not cleaved during the reaction and (3) the stretching of one bond cleaved during the reaction. These three parts are additive and for the last one the potential energy of the reactant is assumed to depend on the bond length according to a Morse curve whereas the product potential energy curve is assumed to be the same as the repulsive part of the reactant Morse curve.[33,34] A comparison of the potential energies for the outer-sphere mechanism of electron transfer and dissociative electron transfer is illustrated in Figure 8(a) and (b).

As a result, the standard activation free energy is now given by Equation (17):

$$(\Delta G^*)^0 = \frac{D_e + \lambda_0 + \lambda_i}{4} \tag{17}$$

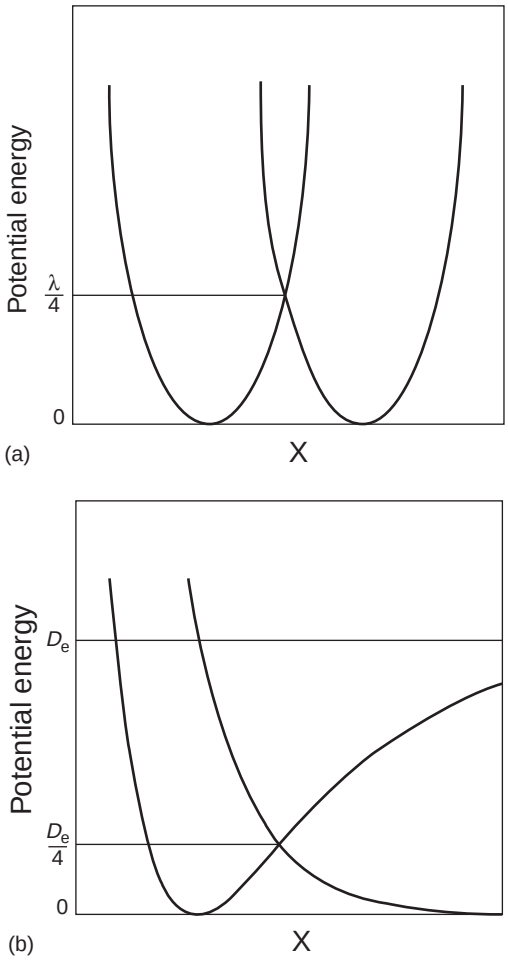

Figure 8 Potential energy (a) for an outer-sphere mechanism of electron transfer (Marcus–Hush theory) and (b) for a dissociative electron transfer.

where D_e represents the bond dissociation energy. It can easily be shown that $(\Delta G^*)^0$ can have large values compared with the outer-sphere case and that the standard potential of the redox couple involved (Equation 16) can be very different to the standard potential from the possible simple electron transfer (Equation 18):

$$A{-}X + e^- \rightleftharpoons AX^{\bullet -} \qquad (18)$$

3.2 Characterization Procedures

3.2.1 Linear Sweep Voltammetry and Convolution

The analysis of the kinetics of the electron transfer must be taken into account in comparison with the always present rate of diffusion. This depends on the method used and on the time window of this method. It should be emphasized that there is a possible difficulty in ensuring that the kinetics are dependent on the electron transfer and not on other phenomena such as chemical evolution of an intermediate product.

The effect of the kinetics of the electron transfer on the curves obtained by the different electrochemical methods is illustrated via LSV and CV. As a first approach, the electronic rate law is often described with the empirical Butler–Volmer law (Equation 19):[5,14]

$$i = FSk_S \exp\left[\frac{-\alpha F(E - E^0)}{RT}\right] \times \left\{(C_A)_0 - (C_B)_0 \exp\left[\frac{F(E - E^0)}{RT}\right]\right\} \qquad (19)$$

(for a reduction) where k_S is the standard rate constant and α the transfer coefficient. It must be noted that this approach can be related to the Marcus law and, for example, when the potential is approximately equal to the standard potential, $k_S = Z_{el} \exp(-(\Delta G^*)^0/kT)$ and $\alpha = 0.5$.

The change in the voltammetric curve occurs according to the change in k_S and to the time window of the method. The operational parameter is, in fact, $k_S/(DFv/RT)^{1/2}$. As the values of k_S become smaller and smaller, the more the cathodic peak potential becomes negative and the more the cathodic and anodic peaks are separated. A cathodic curve depending only on the shape of α is obtained for lower values of k_S.[5,11]

However, as mentioned earlier, the quadratic law is predicted by the Marcus–Hush theory and its extension. Under these conditions, some different evolutions are obtained that are characterized by the standard activation free energy. The value of the apparent transfer coefficient is then a function of the gap between peak potential and standard potential. Figure 9(a) and (b) shows a comparison between these two possible approaches. The apparent transfer coefficient decreases from 0.5 in the vicinity of the standard potential when the potential is set to a more negative value.

The shape of the curve is determined by the value of the peak width (gap between peak potential and half-peak potential) (Equation 20):

$$E_{p/2} - E_p = 1.85\left(\frac{RT}{\alpha F}\right) \qquad (20)$$

and for all the values of $\alpha \leq 0.5$ this peak width is larger than for a fast electron transfer where only the diffusion is rate-limiting step. This measurement of the peak width is a good way to determine the experimental value of α_{ap} but not very accurate, particularly when this value is near 0.5.

A way in which to increase this accuracy is to take the convolution of the LSV curve (Equation 21):[35]

$$I = \frac{1}{\pi^{1/2}} \int_0^{\tau} \frac{i(v)}{(\tau - v)^{1/2}}\, dv \qquad (21)$$

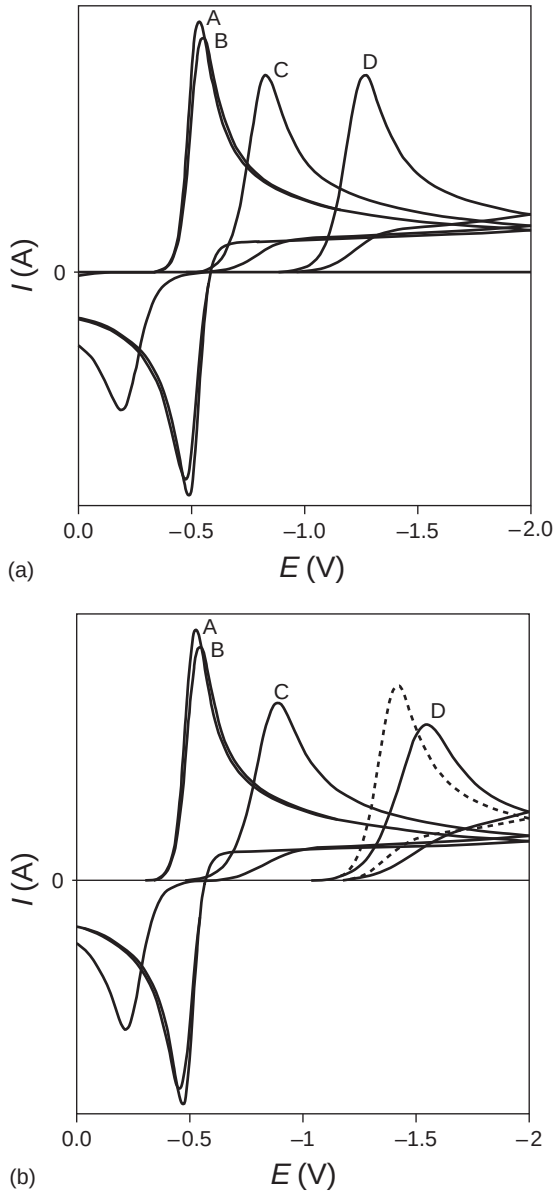

Figure 9 CV with a slow-rate electron transfer, (a) according a Butler–Volmer law with $k_S(\text{cm s}^{-1}) = (\text{A})\ \infty$, (B) 0.03 (C) 5×10^{-5} and (D) 1×10^{-8} and (b) according a Marcus–Hush law. For curve D, comparison of the shape of the curve corresponding to the Butler–Volmer law (dashed line).

This quantity is easily calculated and is directly proportional to the diffusing reactant concentration. Hence a plateau curve identical with what is obtained using a stationary technique is obtained regardless of the kinetic law of the electron transfer that is followed (Figure 10a).

From the shape of the curve and from the different convolution curves obtained using LSV curves at several sweep rates, the kinetics of the electron transfer are measured by determining the value of $\ln k(E)$

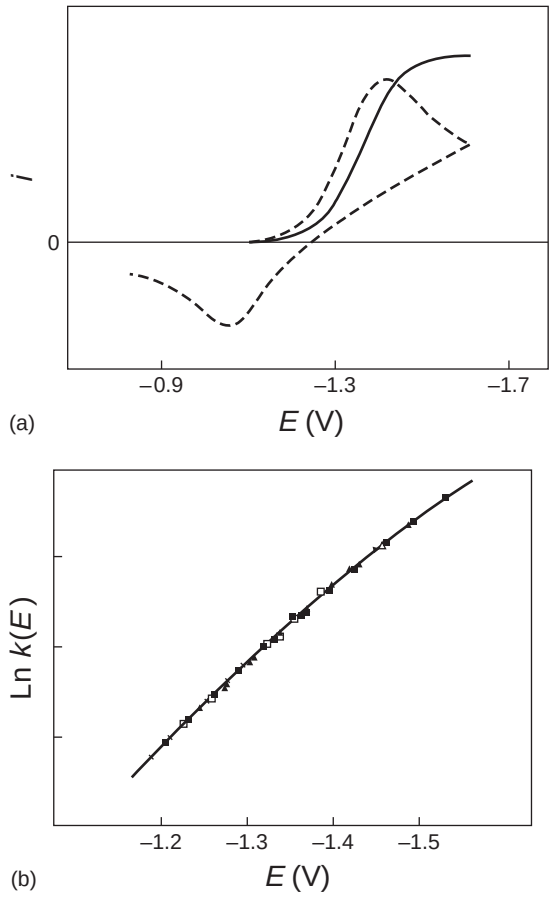

Figure 10 (a) CV (dashed line) and the corresponding convolution curve (solid line) for *tert*-nitrobutane in DMF. (b) Charge transfer rate $k(E)$ for *tert*-nitrobutane in DMF, using sweep rates between 0.59 and 189 V s^{-1}. [Reproduced by permission of Elsevier Sequoia from J.-M. Savéant, D. Tessier, *J. Electroanal. Chem.*, **65**, 57–66 (1975).]

(Equation 22):

$$i = FSk(E)\left\{(C_A)_0 - (C_B)_0 \exp\ \frac{F(E - E^0)}{RT}\ \right\} \qquad (22)$$

An example of the determination of this quantity is given in Figure 10(b), where a curvature of this relationship with potential, resulting from a Marcus-like law of the electron transfer, can be noted. The apparent transfer coefficient α_{ap} is obtained from Equation (23):

$$\alpha_{ap} = -\left(\frac{RT}{F}\ \ \frac{\mathrm{d}[\ln k(E)]}{\mathrm{d}E}\right. \qquad (23)$$

The standard potential is the value of the potential when α_{ap} is exactly 0.5.

3.2.2 Impedance Technique

This kind of determination may also be conducted using an impedance technique,[20] mainly in the vicinity of the

standard potential for fairly fast electron transfers (quasi-reversible system). By superimposing an alternating current on a slowly variable dc potential, the faradaic resistance R_f and the faradaic capacitance C_f are analyzed taking into account the double-layer capacitance and the cell resistance that are obtained from a test experiment. The apparent rate constant $k(E)$ is obtained from Equation (24):

$$\log \frac{k(E)}{D^{1/2}} = -\log \frac{(R_f C_f w - 1)[1 + \exp\{(F/RT)(E - E_{1/2})\}]}{(2w)^{1/2}} \quad (24)$$

The results obtained by this technique for a quasi-reversible system have comparable accuracy to those is obtained via convolution.[36]

3.3 Some Results

Organic species producing stable radicals (mainly anion radicals in reduction and cation radicals in oxidation) are for the most part aromatic and bulky molecules. Thus, the electron transfer for these molecules can be explained using the Marcus–Hush theory where the predominant effect is solvation reorganization. The transfers that occur are fairly fast and their measurement requires very short time window techniques.

Some examples of the determination of the kinetics of simple reduction processes as measured using the standard rate constant and the standard free energies of activation where DMF is the solvent are given in Table 2.

It is possible to show that a quadratic Marcus-type law is followed for the reduction of nitrodurene and *tert*-nitrobutane where the electron transfer is simple but slower than for the first ones. There the change in the solvation is localized at the level of the nitro group, which leads to a significant change in the solvation reorganization term.[38] For the reduction of aliphatic radicals a significant change in the internal reorganization term explains the large standard activation energy that results.[39]

When chemical reactions occur (the product of electron transfer is not stable), the kinetics of the electron transfer are difficult to determine, yet, as explained below, the measurement of the kinetics for very slow electron transfer is possible whatever the following chemical or electrochemical processes. For this reason, it is easy to characterize dissociative electron transfer where the $(\Delta G^*)^0$ values are high owing to the influence of the dissociation energy. Some examples of the determination of $(\Delta G^*)^0$, of values of α_{ap}, of transfer coefficients and of the order of magnitude of the gap between the standard potential and the effective reduction potential in a voltammetric experiment are presented for the case of reductive cleavage of organic halides[40,41] in Table 3.

All these electron transfers involve the following electrochemical elementary reaction (Equation 25):

$$R{-}X + e^- \rightleftharpoons R^\bullet + X^- \quad (25)$$

with a standard potential $E^0(RX/R^\bullet + X^-)$. It is important to note that this standard potential is absolutely different from the standard potential $E^0(RX/RX^{\bullet -})$ in the couple (Equation 26)

$$R{-}X + e^- \rightleftharpoons RX^{\bullet -} \quad (26)$$

which corresponds only to a simple electron exchange on the same molecule. For example, in the case of the reduction in DMF of 9-chloromethylanthracene,[41] $E^0(RX/R^\bullet + X^-) = -0.15\,V$ versus SCE and $E^0(RX/RX^{\bullet -})$ is estimated by comparison with other anthracenyl compounds to be about -1.8 to $-1.9\,V$ versus SCE.[42] The fact that the reduction potential is $-1.36\,V$ in LSV at $1\,V\,s^{-1}$ and that this value is less negative than for the possible simple reduction reaction is proof of the efficiency of the reduction via dissociative electron transfer.

Table 2 Kinetics of electron transfer for a simple reduction process

Compound	k_S ($cm\,s^{-1}$)	$(\Delta G^*)^0$ (eV)	Technique	Refs.
Anthracene	5	0.13	Impedance	32
Phthalonitrile	1.8	0.17	Impedance	32, 37
Nitrobenzene	2.2	0.16	Impedance	32
Nitrodurene	0.12	0.26	Impedance	37
tert-Nitrobutane	4×10^{-4}	0.36	Convolution	38
tert-Butyl radical	9×10^{-7}	0.58	R[a]	39
Benzyl radical	0.7	0.23	P[b]	27

[a] By analysis of the second step of reduction of *tert*-butyl iodide by LSV.
[b] By photoinjection of electrons.

Table 3 Values of standard free energy of dissociative electron transfer $(\Delta G^*)^0$, of values of α_{ap} and of the gap between the standard potential and the effective reduction potential

Compound	$(\Delta G^*)^0$ (eV)	α_{ap}	$E^0(RX/RX^{\bullet -}) - E_m$ (V)
9-Chloromethylanthracene	0.77	0.38	1.14
4-Cyanobenzyl bromide	0.72	0.33	0.97
Benzyl bromide	0.74	0.34	1.03
n-Butyl bromide	0.93	0.25	1.60
n-Butyl iodide	0.80	0.30	1.10

E_m is the mean value of the reduction peak in an LSV experiment at $0.1\,V\,s^{-1}$.

4 SIMPLE ORGANIC ELECTROCHEMICAL MECHANISMS

4.1 Nature and Classification of Coupled Chemical Reactions

In this section, the different kinds of chemical reactions that occur in organic electrochemistry and their influence on the kinetics of the overall process will be presented. For this purpose, we have chosen some simple examples where the nature of these reactions is clearly demonstrated and not merely postulated. This is an important point because in complex mechanisms not all the steps are fully elucidated as they are out of reach of characterization techniques.[5,11]

The way in which these chemical reactions may be qualitatively characterized will be explained for each particular case.

4.1.1 Preceding Chemical Reaction

The reaction scheme in Scheme 1 is followed:

$$C \rightleftharpoons A$$
$$A \pm e^- \rightleftharpoons B$$

Scheme 1

Compound C present in solution is not electroactive in the potential range of interest and compound A must be formed through prior conversion of C. It must be noted that compound C is sometimes reducible or oxidizable directly in another range of potential. The reduction of carbonyl compound in acidic medium illustrates this behavior (Scheme 2):[1,2]

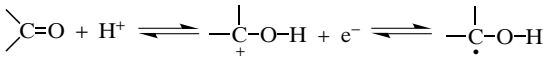

Scheme 2

Reactions that thermodynamically favor C are the only ones that are easily observable and the main manifestation of this process is the height of the wave in RDEV, the peak current in LSV or the current values in a single potential step technique. If the chemical reaction is slow compared with the diffusion, the result will be a measurement of the equilibrium concentration of A, whereas if the reaction is fast, the result will be proportional to the total concentration of A and C. So, for example, in RDEV the height of a wave does not follow $w^{1/2}$ when the time window is of the same order of magnitude as that of the rate of reaction.

4.1.2 Consecutive Reaction

A chemical reaction occurs after the electron transfer (Scheme 3):

$$A \pm e^- \rightleftharpoons B$$
$$B \rightleftharpoons C$$

Scheme 3

In the case where C is thermodynamically stable compared with B, this intermediate B is not stable and cannot accumulate near the electrode. As written above, this reaction is first order (or pseudo-first order if B reacts with solvent or a species in high concentration in the solution), but a second-order reaction (dimerization) is also possible. The characterization of such a process is not possible at the level of the wave or peak current but other techniques that allow the detection of the reverse current (reoxidation of B for a reduction of A) can help to visualize this type of process. Thus, with CV or RRDEV, if the lifetime of the intermediate is small compared with the time window of the method, no reoxidation current is observed. However, if the time window of the technique can be shortened compared with the lifetime of B, then nothing changes in the pattern obtained from a simple electron transfer.

Another way to detect a fast chemical reaction involves the displacement of the potential in response to the chemical reaction. The electrochemical reaction is facilitated in the presence of a chemical reaction (more positive reduction).

At the level where it is possible to determine the order of the reaction, an accurate determination of the peak potential in LSV or of the half-wave in RDEV provides an answer to the question. The displacement as a function of the initial concentration also provides information.

Some examples of consecutive reactions are given in Schemes 4–6.

First-order, bond cleavage reaction:[43]

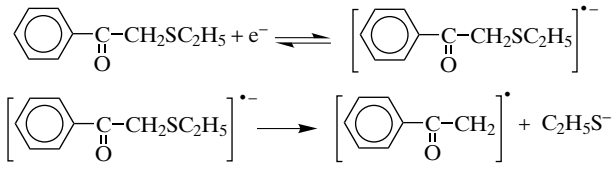

Scheme 4

Pseudo-first-order reaction, protonation of a ketone by phenol:[44]

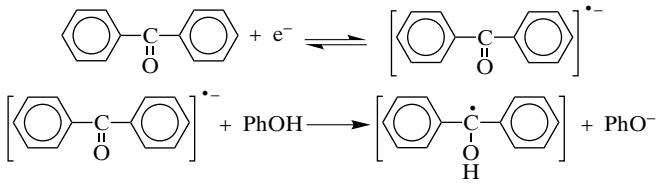

Scheme 5

Dimerization:[45]

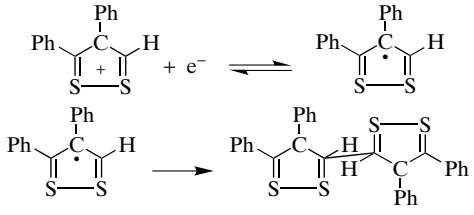

Scheme 6

4.2 Kinetic Analysis for a Single Chemical Reaction Acting as the Determining Step

As explained earlier, in organic electrochemistry the electron transfers are fairly fast and homogeneous chemical reactions occur more frequently than surface chemical reactions. When electron transfer is fast, it is possible to express, using efficient methods, how the homogeneous chemical reactions (acting only with the diffusion), influence the electrochemical response. For each method used the same kind of kinetic limits are obtained but, as LSV and CV are often used, the results are presented for these methods.[46] The sequence of electrochemical and chemical reactions are expressed by the letters E for electrochemical and C for chemical; if necessary, the order of the reaction is indicated after the letter C.[13]

4.2.1 First-order Preceding Chemical Reaction (C + E) (Scheme 7)

As explained above, only cases where the equilibrium is displaced towards C are of interest. For each value of

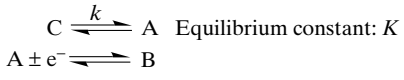

Scheme 7

this equilibrium constant, three main kinetic behaviors are singled out according to the value of the forward rate constant k:

- $k \approx 0$, curve corresponding to a simple diffusion of the A species in the thermodynamically low concentration: ''pure diffusion'';
- $k \approx \infty$, curve corresponding to the reduction of the whole concentration of species A + C: ''pure extraordinary diffusion'';
- intermediate value of k where a plateau-shaped curve is obtained at negative potential; this ''pure kinetics'' case depends on the parameter $K\sqrt{k}$ and under these conditions the plateau current obtained is independent of the sweep rate (Equation 27):

$$i_{pl} = FSC^0 D^{1/2} K\sqrt{k} \tag{27}$$

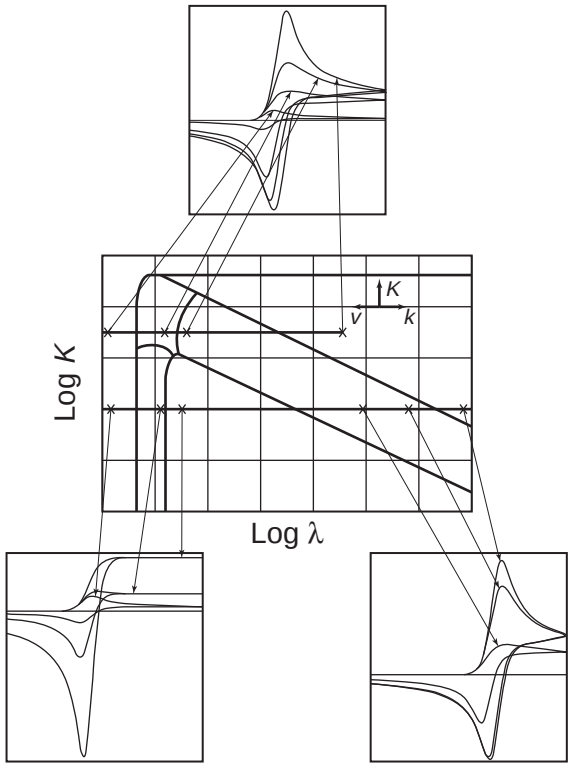

Figure 11 First-order C + E mechanism. Kinetic zone diagram for LSV with $\lambda = (RT/F)(k/v)$ and the characteristic cyclic voltammograms.

These limiting cases are easily visualized in a kinetic diagram (Figure 11) where kinetic control intermediate zones depending on the accuracy of the determination are present between the three pure limiting zones. The two parameters are the thermodynamic constant K and the kinetic constant k.

4.2.2 Consecutive Chemical Reaction

4.2.2.1 First-order Reaction (E + C1) (Scheme 8)

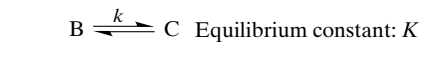

Scheme 8

The kinetic diagram with the two parameters k and K is shown in Figure 12. Here also a ''pure kinetic'' zone is obtained for intermediate values of k in the region of an equilibrium displaced towards C. In this pure kinetic zone, chemical irreversibility occurs (no reverse current

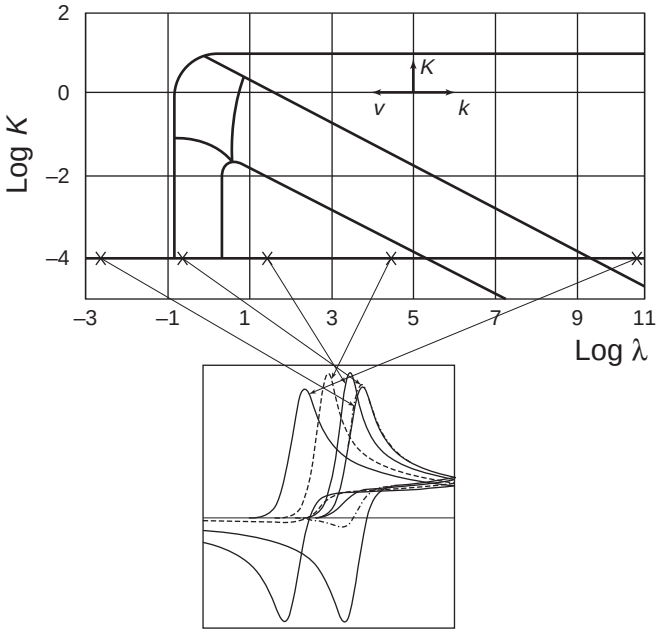

Figure 12 First-order E + C mechanism. Kinetic zone diagram for LSV with $\lambda = (RT/F)(k/v)$ and the characteristic cyclic voltammograms.

in CV) and the characteristics of the peak current are expressed by Equations (28–30):

Peak current:

$$i_p = 0.496FSC^0D^{1/2}\left(\frac{Fv}{RT}\right)^{1/2} \tag{28}$$

Peak potential:

$$E_p = E^0 - 0.78\left(\frac{RT}{F}\right) + \left(\frac{RT}{2F}\right)\ln\left(\frac{RT}{F}\right)\left(\frac{k}{v}\right) \tag{29}$$

Peak width:

$$E_{p/2} - E_p = 1.85\left(\frac{RT}{F}\right) \tag{30}$$

Thus, the peak current in a ''pure kinetic'' condition is proportional to the square root of the sweep rate and the peak potential is displaced to $-(RT/2F)\ln 10$ (−30 mV at 302 K) when the sweep rate is increased by one order of magnitude. This value is for a reduction reaction as in the other following schemes. For an oxidation a change in the direction of the variations is obvious.

The displacement to −30 mV at 302 K when the sweep rate is increased by one order of magnitude without any variation with the concentration of the initial species is a characteristic of the order of the chemical reaction.

4.2.2.2 Second-order Reactions (E + C2)[47] If the chemical reaction is a second-order reaction, the same kind of kinetic zone diagram is obtained and a pure kinetic zone is also found. In fact, several coupling processes are possible and each of them can be characterized, in the ''pure kinetic'' zone, by the shape of the voltammetric curve in CV, by the peak potential variations with sweep rate and concentration in LSV.

$$A \pm e^- \rightleftharpoons B$$
$$2B \overset{k}{\rightleftharpoons} D$$

Scheme 9

Dimerization Between Two Reduced (or Oxidized) Species [E + C2(Arr)] (Scheme 9). The characteristics of the peak current are expressed by Equations (31–33):

Peak current: $$i_p = 0.527FSC^0D^{1/2}\left(\frac{Fv}{RT}\right)^{1/2} \tag{31}$$

Peak potential: $$E_p = E^0 - 0.902\left(\frac{RT}{F}\right) + \left(\frac{RT}{3F}\right)\times \ln\left(\frac{2RT}{3F}\right)\left(\frac{kC^0}{v}\right) \tag{32}$$

Peak width: $$E_{p/2} - E_p = 1.51\left(\frac{RT}{F}\right) \tag{33}$$

The displacements to −20 mV at 302 K when the sweep rate is increased by one order of magnitude and to 20 mV when the concentration of the initial species is increased by one order of magnitude characterize this dimerization process.

Dimerization Between One Reduced (or Oxidized) Species and the Substrate [E + C2(Ars)] (Scheme 10). The characteristics of the peak current are expressed by

$$A \pm e^- \rightleftharpoons B$$
$$A + B \overset{k}{\rightleftharpoons} D$$

Scheme 10

Equations (34–36):

Peak current: $$i_p = 0.430FSC^0D^{1/2}\left(\frac{Fv}{RT}\right)^{1/2} \tag{34}$$

Peak potential: $$E_p = E^0 - 0.457\left(\frac{RT}{F}\right) + \left(\frac{RT}{2F}\right)\times \ln\left(\frac{RT}{F}\right)\left(\frac{kC^0}{v}\right) \tag{35}$$

Peak width: $$E_{p/2} - E_p = 2.27\left(\frac{RT}{F}\right) \tag{36}$$

The displacements to −30 mV at 302 K when the sweep rate is increased by one order of magnitude and to 30 mV when the concentration of the initial species is increased by one order of magnitude characterize this dimerization process.

When only one compound is electroactive, from these four simple kinetic situations (C + E, E + C1 and the two E + C2), it is possible to analyze the complex mechanisms obtained in organic electrochemistry.

One question remains concerning the possibility of considering the kinetics of the electron transfer as a non-limiting kinetic phenomenon. For example, in chemical reactions following the electron transfer in reduction, the CV curve is displaced towards increasingly positive potentials as the chemical rate constant increases. Taking into account the law of electron transfer, it appears that the apparent electron rate constant is lower at more positive potential values than at standard potentials. Hence the faster the chemical reactions are, the more the limitation by electronic transfer becomes significant.[11] However, in many experiments, in a range up to 10^4–10^6 s for the first-order chemical reaction constant, the hypothesis of fast electron transfer still remains valid.

5 COMPLEX MECHANISMS

The mechanism for electrochemical reactions is the pathway leading from the initial reactants to the final products where at least one electron transfer occurs. In many cases the electrochemical reactions on a substrate can yield several products. On the one hand, the process is potential dependent and, according to the applied potential, the nature of the products is different. On the other hand, for a certain potential, competitive sequences of reactions are possible. In this section we limit ourselves to mechanisms where only one sequence is involved (no branching reactions) and also to cases where only one substrate is electroactive.

An electrochemical mechanism is written as a succession of electrochemical steps and chemical steps and we shall focus on two main problems: (1) how to understand the reaction sequence for complex reactions and (2) to elucidate what happens when the mechanism involves several electron transfers.

5.1 Sequence of Chemical Reactions After One Electron Transfer

Consider the formation of pinacol from the reduction of a ketone in a slightly acidic medium (the ketone cannot be protonated prior to its reduction). The first electrochemical reaction is (Equation 37):

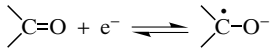

(37)

and successive reactions lead to the pinacol (Equation 38):

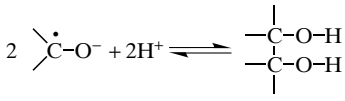

(38)

Because the dimerization of the two anions is slow, the possible sequence is as shown in Scheme 11:

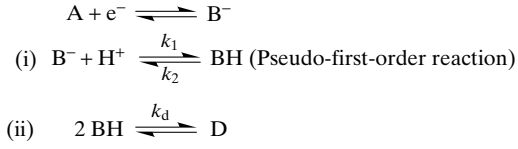

Scheme 11

and several kinetic limiting behaviors are possible according to what the slow chemical step (rate-determining step) is, as in chemical kinetics.[44]

If reaction (i) is slow compared with reaction (ii), the kinetics and the characteristics of the electrochemical curves are those of the mechanism E + C1. For example, in LSV, the characteristics are expressed by Equations (39–41):

$$\text{Peak current:} \quad i_p = 0.496FSC^0D^{1/2}\left(\frac{Fv}{RT}\right)^{1/2} \tag{39}$$

$$\text{Peak potential:} \quad E_p = E^0 - 0.78\left(\frac{RT}{F}\right) + \left(\frac{RT}{2F}\right) \times \ln\left(\frac{RT}{F}\right)\left(\frac{k}{v}\right) \tag{40}$$

$$\text{Peak width:} \quad E_{p/2} - Ep = 1.85\left(\frac{RT}{F}\right) \tag{41}$$

Hence the peak current is, as in pure diffusion conditions, proportional to the square root of the sweep rate and the peak potential is displaced to $-(RT/2F)\ln 10$ (−30 mV at 302 K) when the sweep rate is increased by one order of magnitude.

If reaction (i) is fast compared with reaction (ii) ($k_2 > k_d$), the reaction acting as a pre-equilibrium for reaction (ii), the kinetics and the characteristics of the electrochemical curves are those of the mechanism $E + C2_{(Arr)}$ resulting in the pre-equilibrium that replaces k_d by $(k_d k_1/k_2)[H^+]$, which introduces a variation with pH and more generally with concentration of other species

occuring before the slow step. In LSV, the characteristics are expressed by Equations (42–44):

$$\text{Peak current: } i_p = 0.430FSC^0D^{1/2}\left(\frac{Fv}{RT}\right)^{1/2} \quad (42)$$

$$\text{Peak potential: } E_p = E^0 - 0.457\left(\frac{RT}{F}\right) + \left(\frac{RT}{2F}\right) \times \ln\left(\frac{RT}{F}\right)\left(\frac{k_d k_1}{k_2}\right)[H^+]\left(\frac{C^0}{v}\right) \quad (43)$$

$$\text{Peak width: } E_{p/2} - E_p = 2.27\left(\frac{RT}{F}\right) \quad (44)$$

The displacement is to −30 mV at 302 K when the sweep rate is increased by one order of magnitude and to 30 mV when the concentration of the initial species or the concentration in protons is increased by one order of magnitude. The preceding results are applicable when the proton concentration is sufficiently high to keep its concentration constant in the process. This is also the case when a buffer medium is used.

More generally, all the reactions occurring after the rate-determining step do not play any role in the kinetics and all the reactions before the rate-determining step act as a pre-equilibrium for the rate-determining step. In an unbuffered medium, a more complicated law results. This is not considered here.

5.2 Two-electron Process: Electron Transfer at the Electrode or in Solution

In organic electrochemistry, the ability to transfer more than one electron during a single step is impossible without a chemical reaction between the two electron transfers. This result is explained as follows. When two electron transfers occur in the same part of the molecule, adding (or removing) an electron leads to a group that is more difficult to reduce (or to oxidize) (repulsion of charge). Hence the second step will occur at a potential that is distinct from the first one and the two steps can be separated by a technique such as LSV.

This result is illustrated by the reduction potentials for a series of dinitro compounds.[48] When the two groups are on the same phenyl ring, a separation of 300–500 mV is observed according to the position of these groups on the ring. When a saturated carbon chain is introduced this separation decreases and it must be noted that for a long chain (three or four CH_2 groups are sufficient) the curve obtained is exactly the same in shape and potential as for a one-electron transfer with a double current. In fact, the two standard potentials are not equal but separated by $(RT/F)\ln 4$ as a function of the statistical factors for the two steps.

However, in many cases a first single two-electron process is observed. This is due to the influence of a chemical reaction that occurs between the two-electron steps that provides for a second step easier than the first one.

This is often the case because injection of an electron into an organic molecule increases its Lewis basicity and thus increases its ability either to capture a proton (or an electrophile) or to lose a base. On the oxidation side, removal of an electron from the molecule increases its Lewis basicity and thus increases its ability either to capture a nucleophile or to lose an acid. This explains the choice of a weakly acidic solvent for reduction (and the use of weakly basic medium in oxidation) in order to stabilize the initially formed ion radical and to yield useful chemical reactions, provided that some nucleophiles are specially added.

An example of a two-electron process is illustrated by the reduction process of anthracene in the presence of an acid (phenol) leading to dihydroanthracene in only one step[49] (Equation 45):

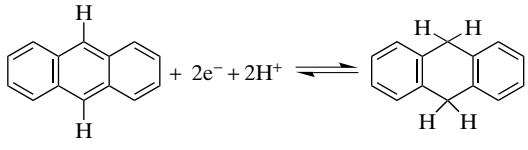

(45)

What are the possible mechanisms for this reaction, involving two-electron transfers? Often, a mechanism involving the sequence shown in Scheme 12 is postulated:

$$\begin{aligned} An + e^- &\rightleftharpoons An^{\bullet -} \quad (E^0{}_1)\\ An^{\bullet -} + H^+ &\rightleftharpoons AnH^\bullet\\ AnH^\bullet + e^- &\rightleftharpoons AnH^- \quad (E^0{}_2)\\ AnH^- + H^+ &\rightleftharpoons AnH_2 \end{aligned}$$

Scheme 12

This mechanism is a succession of electron transfers at the electrode and of chemical reactions in solution. It is an E + C + E + C (often called an ECE) mechanism.

By the fact that a two-electron transfer is obtained, the second electron transfer is easier than the first. This means that E_2^0 is more positive than E_1^0. The reaction, an electron transfer in solution (Equation 46):

$$AnH^\bullet + An^{\bullet -} \rightleftharpoons AnH^- + An \quad (46)$$

is a disproportionation of two once-reduced species. This is displaced to the right, thus making the sequence in Scheme 13 also possible:

It appears that three limiting kinetic behaviors are possible, leading to characteristic variations within the experimental parameters (Table 4).[50,51]

$$\begin{aligned} An + e^- &\rightleftharpoons An^{\bullet -} \quad (E^0_1) \\ An^{\bullet -} + H^+ &\rightleftharpoons AnH^{\bullet} \\ AnH^{\bullet} + An^{\bullet -} &\rightleftharpoons AnH^- + An \\ AnH^- + H^+ &\rightleftharpoons AnH_2 \end{aligned}$$

Scheme 13

As is easily explained by the fact that the same reaction is the rate-determining step, the first two mechanisms are not easily separable by electrochemical techniques and under pure kinetic conditions (as in Table 4) it is impossible to deduce this determination. Yet at the level of the transition between the two-electron process and the one-electron process when the effect of the chemical reaction is increasingly weakened, some discrepancies appear between these two mechanisms. For example, in DSC[52] the law of variation with respect to inversion time is given in Figure 13.

It is also possible to predict what the mechanism followed is according to the values of the kinetic constants of the two reactions and according to the method used. In LSV (Scheme 14) the two parameters governing the competition between these three mechanisms are k_2/k_1, the equilibrium constant $K = [B]/[C]$ and $p = k_d C^0 (RT/Fv)^{-1/2} (k)^{-3/2}$, where $k = k_1 + k_2$.

$$\begin{aligned} B &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \\ B + C &\xrightarrow{k_d} D + A \end{aligned}$$

Scheme 14

Figure 14 shows the predicted kinetic control diagram.[51] From this diagram and taking into account that the value of k_d is very high since the equilibrium has been displaced toward A + D, and therefore can reach the diffusion limit in the solvent, it is possible to show that a second-order disproportionation mechanism is only possible in the case where an equilibrium B/C is displaced towards B and where k_2 has very high values. An ECE mechanism is often written in the case of a two-electron

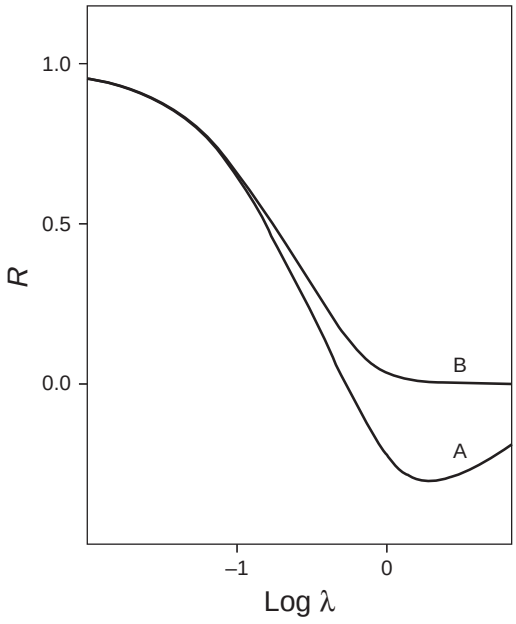

Figure 13 Two-electron process. Comparison of (A) ECE and (B) first-order disproportionation mechanisms in DSC ($\lambda = k_1\theta$, $R = [i(2\theta)/i(\theta)]/0.293$).

mechanism without a real proof of the mechanism. It appears that this case is mainly possible for an equilibrium displaced towards C and where the values of the kinetic constant k_1 are high. At a moderate sweep rate of $1\ V\ s^{-1}$ at a millimolar concentration, if the disproportionation is at the diffusion limit ($\sim 10^{10}\ s^{-1}$), the value of k_1 must be higher than $2 \times 10^5\ s^{-1}$. This is explained qualitatively by the fact that if the reaction B → C is fast the production of C occurs near the electrode and the second transfer can take place another time at the electrode. However if the reaction is slow, the production of C takes place far from the electrode and the solution reaction of electron exchange becomes predominant.

5.3 Redox Catalysis

From a practical point of view, e.g. in order to avoid passivation of the electrodes, or for kinetic purposes, it is

Table 4 Two-electron process: mechanisms and diagnostic criteria in LSV

ECE	DISP 1	DISP 2
$A + e^- \rightleftharpoons B$	$A + e^- \rightleftharpoons B$	$A + e^- \rightleftharpoons B$
$B \longrightarrow C$	$B \longrightarrow C$	$B \rightleftharpoons C$
$C + e^- \rightleftharpoons D$	$B + C \rightleftharpoons D + A$	$B + C \longrightarrow D + A$
$D \rightleftharpoons E$	$D \rightleftharpoons E$	$D \rightleftharpoons E$
$\partial E_p/\partial \log v$ in LSV = −30 mV at 302 K	$\partial E_p/\partial \log v$ in LSV = −30 mV at 302 K	$\partial E_p/\partial \log v$ in LSV = −20 mV at 302 K
$\partial E_p/\partial C^0$ in LSV = 0	$\partial E_p/\partial C^0$ in LSV = 0	$\partial E_p/\partial C^0$ in LSV = 20 mV at 302 K

DISP 1 and 2 represent first- and second-order disproportionation, respectively.

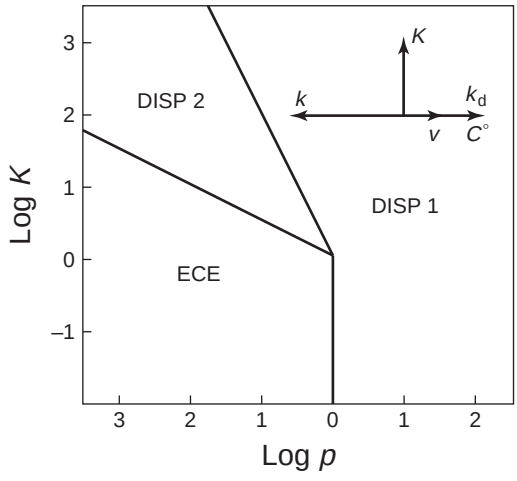

Figure 14 Two-electron process. Kinetic zone diagram showing the possibilities for the three mechanisms; $p = k_d C^0 (RT/Fv)^{-1/2} (k)^{-3/2}$.

sometimes possible to realize an electrochemical reaction via the intermediary of a homogeneous electron transfer from a mediator acting as an electron carrier from the electrode to the solution.

The principle of this process is described in the example of a reductive first-order E + C reaction scheme[53] (Scheme 15):

Direct electrochemical process:

$$A \pm e^- \rightleftharpoons B$$

$$B \xrightarrow{k} C$$

Mediated process:

$$P + e^- \rightleftharpoons Q \quad (E^0_{PQ})$$

$$Q + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B + P$$

$$B \xrightarrow{k} C$$

Scheme 15

The mediator couple (or catalyst couple) is chosen not to undergo any chemical reaction alone. Species P and A are only initially present at concentrations C^0_P and C^0_A. It is possible to obtain a mediated process at a potential where the direct process is not operative. This would be true when the direct process is kinetically slow, that is, either when the intrinsic rate constant is low and/or if the kinetic constant k is large.

Two limiting behaviors are reached according to the values of the kinetic constants and the parameters of concentration and diffusion rate. Thus, in LSV the system depends upon three parameters (Equations 47–49):

$$\lambda_1 = k_1 C^0_P \left(\frac{RT}{Fv} \right. \tag{47}$$

$$\sigma = \frac{k}{k_2 C^0_P} \tag{48}$$

$$\gamma = \frac{C^0_A}{C^0_P} \tag{49}$$

Two limiting behaviors are distinguished accordingly the value of σ. When $\sigma \gg 1$, the reaction of homogeneous electron exchange in the direct direction is the rate-determining step and the observed curve is a function of λ_1 and γ. An example of a curve and peak current evolution is presented in Figure 15(a–c). It must be noted that for large values of γ, a plateau-shaped curve is obtained for a wide range of values of λ_1. When $\sigma \ll 1$, the important parameters are $\lambda_1 \sigma$ [$= (kk_1/k_2)(RT/Fv)$] and γ. It is interesting that as the parameter is experimentally varied by changing the mediator concentration, it is therefore possible, under favorable circumstances, to shift the kinetic control from the homogeneous forward reaction to an intervention of the deactivation of B by decreasing the concentration of the mediator. This fact is used to determine of the rate constant k in the range 10^6–$10^9\ s^{-1}$, out of reach of all direct electrochemical techniques, even the most modern ones.

Another possible catalytic process, called chemical catalysis, is obtained when an adduct is formed between the catalyst (or a species obtained by electron transfer from the catalyst) and the substrate.[54] The reaction induced by this adduct leads to the products and to a regeneration of the catalytic species after several steps and perhaps after several electron transfers. The shape of the voltammetric curves obtained in these cases can be explained by a sequence of chemical and electrochemical reactions.

6 EXAMPLES OF MECHANISM ANALYSIS AND RATE CONSTANT DETERMINATION

The ability to reduce or to oxidize electrochemically organic molecules has been the topic of numerous books and even series of books. It is not the purpose of this article to study in detail the possible reactions for each class of compounds. In many cases, the mechanism is only described by the overall reactions leading from the substrates to the products. These products are analyzed using traditional methods of organic chemistry.

We present here only a selection of some important electrochemical reactions where the mechanism is analyzed by means of electrochemical methods, leading in some cases to the determination of kinetic constants.

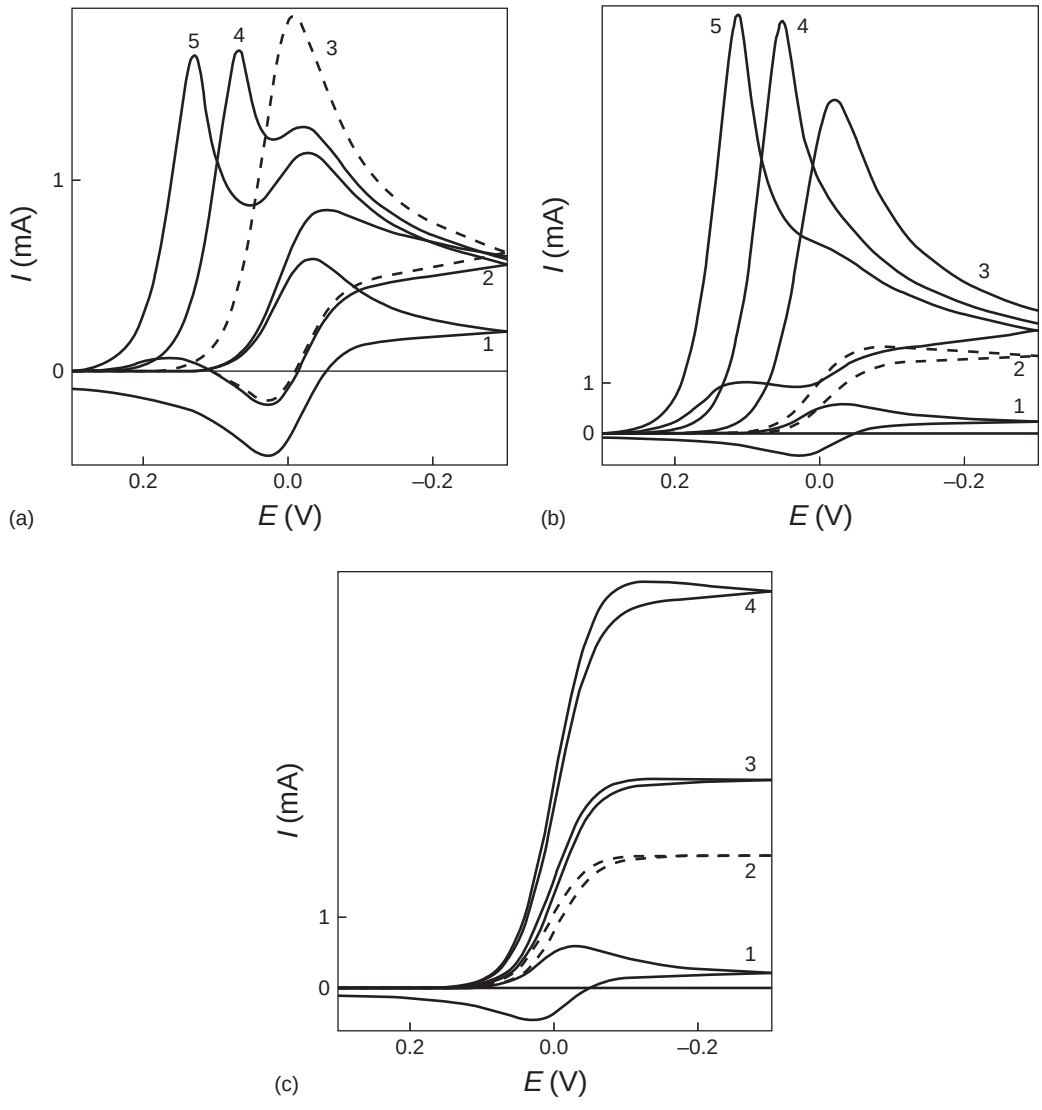

Figure 15 Redox catalysis. CV when the reaction of homogeneous electron exchange in the direct direction is the rate-determining step with $\lambda_1 = k_1 C_P^0 (RT/Fv)$, $\gamma = C_A^0/C_P^0$. (a) $\gamma = 2$, $\gamma_1 =$ (1) 0, (2) 0.2, (3) 20, (4) 2000, (5) 200 000. (b) $\gamma = 10$, $\gamma_1 =$ (1) 0, (2) 0.2, (3) 20, (4) 2000, (5) 200 000. (c) $\gamma = 1000$, $\gamma_1 =$ (1) 0, (2) 0.02, (3) 0.05, (4) 0.2.

6.1 Kolbe Reaction

This first example deals with the oldest electrosynthetic process in organic chemistry. In an aqueous alcoholic medium without any particular precautions, salts of organic acids by oxidation undergo a decarboxylation and then the formation of a carbon–carbon bond between the two organic groups[1,2,55] (Equation 50):

$$2R{-}COO^- \longrightarrow R{-}R + 2CO_2 + 2e^- \qquad (50)$$

This reaction is often used with long-chain carboxylates. At the level of the mechanism determination this reaction is very difficult to analyze because these salts are difficult to oxidize and furthermore the solvents are oxidized in the same range of potential. Two kinds of mechanisms have been postulated for the oxidation of these compounds:

- radical–radical coupling leading to a dimer compound after a one-electron transfer (Scheme 16):
- the formation of a carbocation after a two-electron transfer leading to several kinds of compounds such as esters, alcohols and ethers.

$$R{-}COO^- \longrightarrow R^\bullet + CO_2 + e^-$$

$$2R^\bullet \longrightarrow RR$$

Scheme 16

These results are supported by the analysis of the products obtained after electrolysis taking into account the structure of the carboxylate and the nature of the solvent.

However, information about the first step of the reaction or about the rate of dimerization of the two radicals is very difficult to obtain in most cases. In terms of the possibility of detecting the intermediate $RCOO^{\bullet}$, nothing was possible through electrochemistry for aliphatic or for unsubstituted aromatic carboxylates and the observation of a slow electron transfer prevents further kinetic determinations. For example, it is not possible to say either that the departure of CO_2 is concerted with the electron transfer or that it occurs in a second step. Even in a nonaprotic medium for easily oxidizable compounds such as aryl acetates an answer is not possible owing to the grafting process on the electrode, except in the case of the oxidation of *N*-dimethylaminophenyl acetate where a sweep rate of $20\,V\,s^{-1}$ led to a fully chemically reversible curve in CV.[56] In this last case the first two steps of the reaction are as shown in Scheme 17.

6.2 Reduction of Halo Compounds

These studies are conducted in an aprotic medium such as acetonitrile or DMF, and from the comparison of the heterogeneous and homogeneous electron transfer an accurate answer to the question could be given concerning the first steps of the reduction process according to the structure of the compounds.

For polyaromatic or easily reducible benzenic compounds bearing an electron-withdrawing group, a two-step mechanism is demonstrated by electrochemical methods or by a redox catalysis using as mediator a stable anion radical of aromatic compounds that permit the determination of the cleavage rate constant.[42] In Table 5 some determinations of standard potential and cleavage rate constants are shown.

For simple phenyl halides or halopyridines, the two-step mechanism is not directly determined without a redox catalysis study showing kinetic control by the forward homogeneous exchange for all the catalysts used. From the determination of these rate constants and by observing the law showing the predicted activation and diffusion lines, homogeneous electron exchange can be characterized as an equilibrium and the standard potential of the aromatic halide/aromatic halide anion couple can be determined (Equation 51):

$$\frac{1}{k_1} = \frac{1}{k_1^{act}} + \frac{1}{k^{dif}} \left\{ 1 + \exp \left(\frac{F}{RT} \left(E^0_{PQ} - E^0_{AB} \right) \right) \right\} \tag{51}$$

where k_1^{act} is the kinetic constant under activation control, which in the vicinity of the standard potential E^0_{PQ} is approximately a straight line with a slope of half the diffusion line when the backward reaction k_2 reaches k^{dif}, the diffusion limit.[57]

Although these results clearly indicate that the rate constant of cleavage is not measurable, the two-step mechanism is still followed (Figure 16a and b).

In contrast to aliphatic halides such as butyl halides, only the activation part of the reaction is observed in the redox catalysis of the compounds.[40] The shape differs significantly from a straight line and the slope as compared with a reverse diffusion is less than 0.5 (Figure 16c and d). The shape of the curves obtained by direct electrochemical methods (for bromides and iodides; the chloro compounds are too difficult to reduce at the electrode) demonstrates a slow electron transfer, which is also attested by a transfer coefficient significantly less

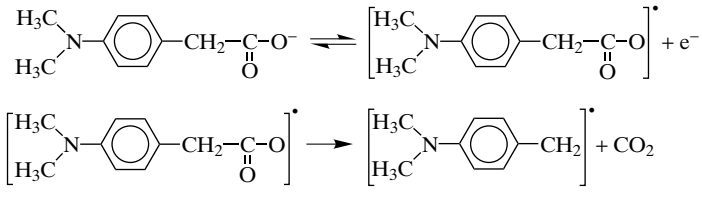

Scheme 17

Table 5 Standard potentials and cleavage rate constants for halo compounds in DMF

Chloro compound	E^0 (V vs SCE)	Log k (s^{-1})	Bromo compound	E^0 (V vs SCE)	Log k (s^{-1})
4-Chloronitrobenzene	−1.05	−2	4-Bromonitrobenzene	−0.98	<−2
4-Chlorobenzophenone	−1.64	1.6	4-Bromobenzophenone	−1.63	5
9-Chloroanthracene	−1.73	2.2	9-Bromoanthracene	−1.70	6.4
1-Chloronaphthalene	−2.26	7.2	1-Bromonaphthalene	−2.19	7.8

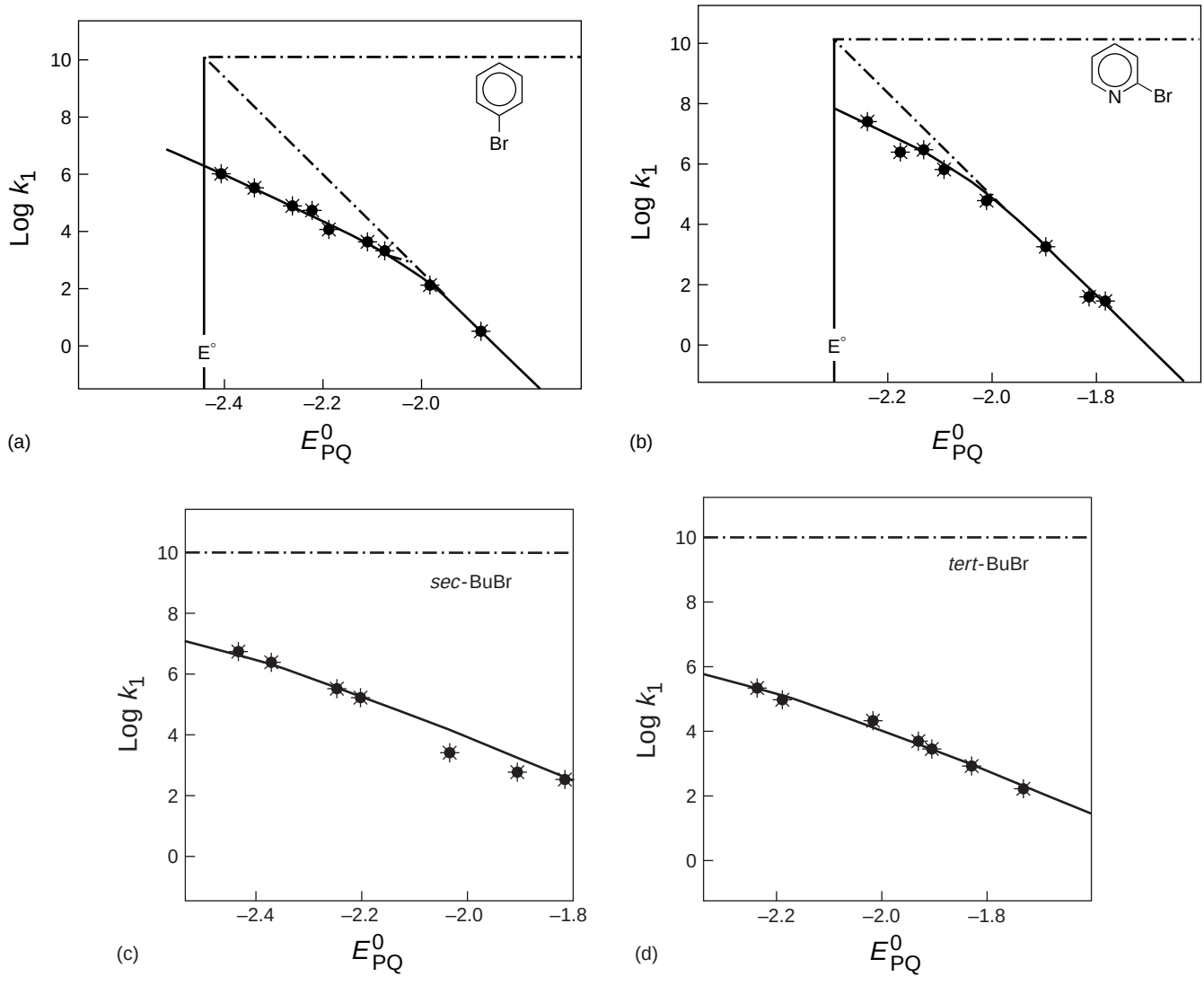

Figure 16 Forward homogeneous electron exchange in the redox catalysis of halo compounds. (a) Bromobenzene; (b) 2-bromopyridine; (c) *sec*-butyl bromide; (d) *tert*-butyl bromide.

than 0.5. In the context of the Marcus–Hush theory, this leads to a large gap between the reduction potential and the standard potential. All these results are characteristic of a concerted reduction and cleavage for the first step of the mechanism.

For bromides, the electron stoichiometry is 2, since the radical produced at the electrode is more easily reducible than the initial substrate. For example, with *n*-butyl bromide (Scheme 18):

$$n\text{-}C_4H_9\text{–}Br + e^- \longrightarrow n\text{-}C_4H_9^\bullet + Br^-$$
$$n\text{-}C_4H_9^\bullet + e^- \rightleftharpoons n\text{-}C_4H_9^-$$
$$n\text{-}C_4H_9^- + AH \longrightarrow n\text{-}C_4H_{10} + A^-$$

Scheme 18

where AH could be an acidic compound or the solvent itself.

For the tertiary and secondary iodides the first step yields a dimer compound and only in a second step (with a more negative potential) is the radical reduced[39] (Scheme 19):

First step:

$$t\text{-}C_4H_9\text{–}I + e^- \longrightarrow t\text{-}C_4H_9^\bullet + I^-$$
$$2\,t\text{-}C_4H_9^\bullet \longrightarrow \begin{matrix} t\text{-}C_4H_9 \\ | \\ t\text{-}C_4H_9 \end{matrix}$$

Second step:

$$t\text{-}C_4H_9^\bullet + e^- \rightleftharpoons t\text{-}C_4H_9^-$$
$$t\text{-}C_4H_9^- + AH \longrightarrow t\text{-}C_4H_{10} + A^-$$

Scheme 19

From these results, it is possible to estimate the standard potential of the anion radical–anion couple for *sec*-butyl (-1.4_4 V vs SCE) and *t*-butyl(-1.5_4 V vs SCE).

6.3 Organic Electrodimerization

This reaction is a powerful method for the industrial preparation of adiponitrile resulting from the reduction of acrylonitrile[1–3,55] (Equation 52):

$$2\ CH_2{=}CHCN \xrightarrow{+2e^- + 2H^+} \begin{matrix} CH_2CH_2CN \\ | \\ CH_2CH_2CN \end{matrix} \qquad (52)$$

In fact, this is a hydrodimerization and with this sample compound, as with many other compounds, the mechanism is very difficult to analyze owing to a reduction that is very close to the solvent reduction.

For more conjugated compounds such as *p*-methylbenzylidenemalononitrile in an aprotic solvent, it is possible to elucidate the sequence of reactions. One important question that arises from this reaction is how to determine whether the dimerization occurs between two reduced species [two anion radicals, $E + C2_{(Arr)}$] or between one reduced species reacting on the substrate [$E + C2_{(Ars)}$]. It is also necessary to indicate at what level the protonations by the residual water are involved.

In the present case, via LSV and via convolution,[58] it is possible to demonstrate that the first reaction is the coupling of two radical anions followed by two protonations (Scheme 20):

$$Ph{-}CH{=}CHCN + e^- \rightleftharpoons Ph{-}\dot{C}H{-}\bar{C}HCN$$

$$2\ Ph{-}\dot{C}H{-}\bar{C}HCN \xrightarrow{k_d} \begin{matrix} Ph{-}CH{-}\bar{C}HCN \\ | \\ Ph{-}CH{-}\bar{C}HCN \end{matrix}$$

$$k_d = 10^7\ M^{-1}\ s^{-1}$$

Scheme 20

The same kind of mechanism occurs during oxidation for the coupling of two cation radicals and during the oxidation of aromatic amines,[59] the loss of proton occurring in a second chemical step.

It is possible to obtain mixed coupling when two compounds are electroactive in the same range of potential and also to obtain intramolecular coupling, yielding cyclic compounds.

The formation of electroactive polymers with compounds such as pyrrole and thiophene is an important electrosynthetic process.[60] As the formation of long-chain oligomers is described, it has often been postulated that the propagation mechanism is the result of the addition of a cation radical or of a deprotonated radical to the monomer or to the polymer moieties. However, for example, polypyrrole is obtained as a long-chain insoluble product and even the use of the fastest method in the electrochemistry of pyrrole does not permit any way of visualizing the intermediates.

In contrast, when substituted pyrroles[61] or bipyrroles are used, it is possible to demonstrate by CV or DSC that the initial step of this polymerization is a dimerization of two cation radicals; the deprotonation steps occur consecutively[62] (Scheme 21):

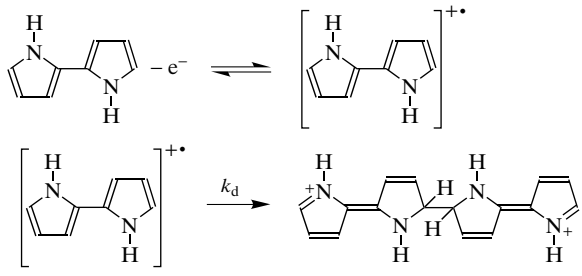

Scheme 21

The same mechanism was demonstrated in the polythiophene series for long-chain polythiophene (three or more units) owing to the strong reactivity of thiophenes as compared with that of pyrroles.

6.4 Chemical Reactions Induced by Electrochemistry

Electrochemistry is also used for inducing chemical reactions when these reactions are impossible or slow when unaccompanied by an activating process. Under these conditions, the electrochemical methods described earlier yield interesting results about the mechanism.

These processes can be induced electrochemically without any addition of other compounds (only by some change in the oxidation degree of one reactant) or with the addition of some chemical catalysts. In the first case the process can be described by Scheme 22.

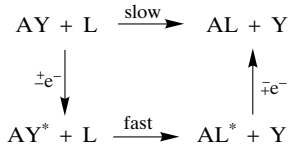

Scheme 22

The "zero-electron" reaction is a nucleophilic substitution induced by electrochemistry when the reduction of AY leads to the replacement of the Y substituent by L, which leads to the reoxidation of $AL^{\bullet -}$ in AL.[63,64] The reaction is called an $S_{RN}1$ reaction (first-order radical nucleophilic substitution) when a one-electron reduction is involved, e.g. in the substitution by phenoxy or

phenylthio groups in halo compounds (Equation 53):

$$RX + Nu^- \longrightarrow RNu + X^- \qquad (53)$$

This reaction is easily studied via CV in the absence or presence of nucleophiles as illustrated in Figure 17(a) and (b). The mechanism of this reaction and good working conditions are determined by the following observations:

- the potential of initiation is the potential of reduction of RX;
- the potential of reduction of RNu must be more negative than the potential of reduction of RX;
- the ''zero-electron'' is more efficient for RX compounds that give a slow cleavage reaction because the faster the cleavage, the closer to the electrode the reaction occurs.

Taking into account the preceding results about the rate of cleavage of halo compounds, polyaromatic compounds must be suitable for this process. The reduction of the radical into an anion is a competing process that must be minimized, either by using chloride rather than bromide or iodide (the cleavage is faster for bromide or iodide than for chloride) or by redox catalysis of the process that also leads to radical formation far from the electrode. On the other hand, we must avoid the possibility of atom transfer from the solvent on the radical; for this purpose,

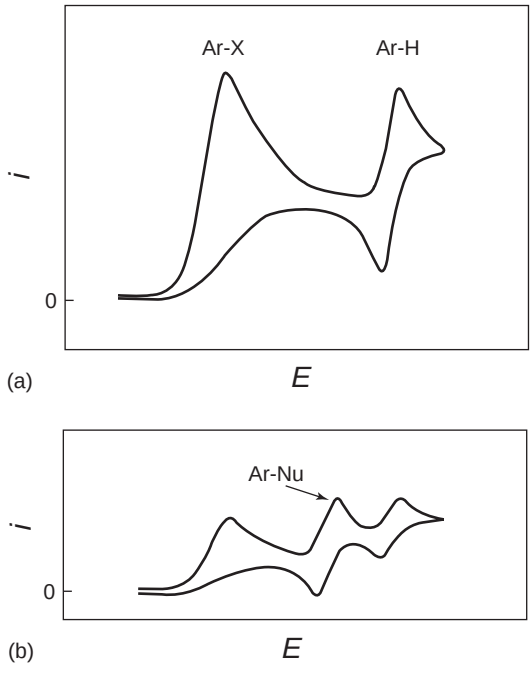

Figure 17 Schematic CV for an $S_{RN}1$ reaction on an aromatic halide. (a) CV without nucleophile and (b) CV in the presence of nucleophile.

liquid ammonia would be the best solvent but we can also use DMSO or acetonitrile[65] (Scheme 23):

$$RX + e^- \rightleftharpoons RX^{\bullet -}$$
$$RX^{\bullet -} \longrightarrow R^{\bullet} + X^-$$
$$R^{\bullet} + Nu^- \longrightarrow RNu^{\bullet -}$$
$$RNu^{\bullet -} \rightleftharpoons RNu + e^-$$

Scheme 23

A good example is the following one shown in Equation (54):

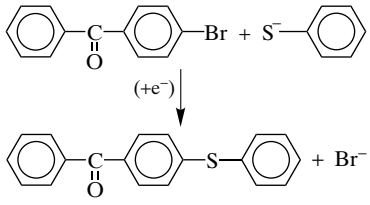

(54)

In many cases the simple activation by reduction (or oxidation) of one reactant does not lead to a significant reaction. A catalyst acting by a chemical reactant on the substrate can be used. For example, in the same kind of reactions, by using palladium complexes,[66,67] some new reactions are possible. The electrochemistry can play on two registers: it is often necessary to prepare the active species at the right degree of oxidation and on the other hand the mechanism of the reaction is followed by electrochemical methods (Scheme 24).

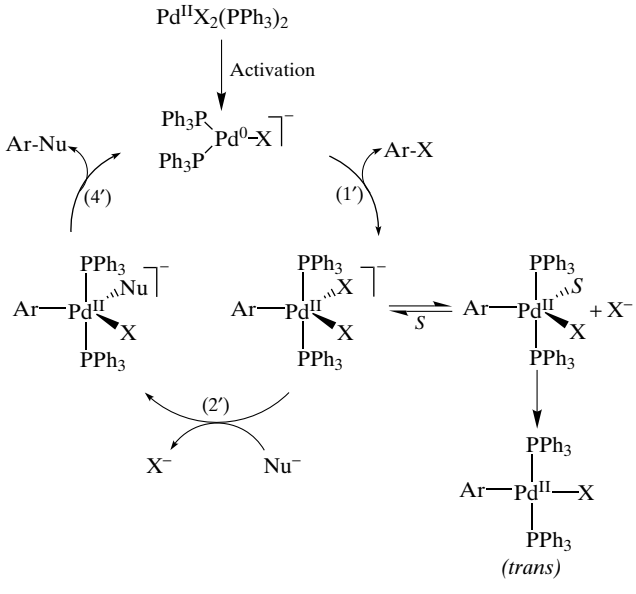

Scheme 24 Catalytic cycle for an $S_{RN}1$ reaction on an aromatic halide using a palladium catalyst. [Reproduced by permission of La Société Française de Chimie from C. Amatore, A. Jutand, L. Thouin, J.-N. Verpeaux, *Actual. Chim.*, (8–9), 43–62 (1998).]

7 CONCLUSION

In contrast to small-molecule electrochemistry, the mechanisms of organic electrochemical reactions are slightly influenced by the adsorption of reactants or of intermediates on the electrodes. Hence it is possible to determine a great number of mechanisms that involve a sequence of electrochemical reactions and homogeneous chemical reactions and to characterize the key steps of these processes.

Our aim was not to list the mechanisms of the electrochemistry of each class of organic compounds but to show what can be determined using different electrochemical methods.

Because simple electron transfer steps are rather fast for organic molecules, the chemical reactions acting before or after the electron transfers are easily characterized. Using double-step techniques, mainly CV and also RRDEV or DSC, when the follow-up chemical reactions are not too fast, it is easy first to visualize the stability of the intermediates with half-lives of about 10^{-4} s. Under these conditions the rate constants of these chemical reactions are accurately measured using either double-step or single-step techniques (LSV, RDEV). Recent advances in electronic devices and in the miniaturization of electrodes allow the detection of intermediates with a lower limit of the detectable half-life of a few microseconds.

Preceding reaction, catalytic reaction or multielectron reactions with the same range of half-life of intermediates are easier to study because the loss of stability of the intermediates can also be followed by the significant changes in the currents giving a clearer determination of the rate constants.

Even when intermediates cannot be detected, the nature of the chemical reactions is accurately demonstrated by the use of single-step techniques as measured by the order of the reaction owing to the possibility of changing the operational window time that these techniques afford.

Even if the limitations due to the electron transfer occur less for organic than for inorganic species, very fast chemical reactions nevertheless lead to kinetic control by electron transfer and prevent kinetic information about the chemical step from being obtained. An important question still remains regarding the fact that electron transfer and chemical reaction are either two separate reactions or only one concerted process. The kinetic behavior of each electron transfer is characteristic and leads to the conclusion that contrary to common knowledge, concerted electron transfer and chemical reaction exist in many cases.

Beyond the description of the reaction sequence, the electrochemical methods described here open up the possibility of obtaining accurate information about the elementary steps in organic chemistry and can contribute to the development of structure–reactivity relationships concerning the kinetics of both electron transfers and coupled chemical reactions.

ACKNOWLEDGMENTS

I am indebted to all the members of the Laboratoire d'Electrochimie Moléculaire who participated in the work on organic electrochemical mechanisms and whose names may be found in the reference list. The invaluable contribution of Dr Jean-Michel Savéant to this field should be particularly emphasized.

For the simulation of cyclic voltammograms, Digisim® from BAS was used.

ABBREVIATIONS AND ACRONYMS

CV	Cyclic Voltammetry
DMF	Dimethylformamide
DMSO	Dimethyl Sulfoxide
DSC	Double-step Chronoamperometry
ESR	Electron Spin Resonance
IR	Infrared
LSV	Linear Sweep Voltammetry
PMV	Photomodulation Voltammetry
PSC	Potential Step Chronoamperometry
RDEV	Rotating Disk Electrode Voltammetry
RRDEV	Rotating Ring-disk Electrode Voltammetry
SCE	Saturated Calomel Electrode
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Infrared Spectroelectrochemistry • Pulse Voltammetry • Scanning Tunneling Microscopy, In Situ, Electrochemical • Ultrafast Electrochemical Techniques • Ultraviolet/Visible Spectroelectrochemistry

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Catalytic Kinetic Determinations: Nonenzymatic

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics

REFERENCES

1. H. Lund, M.M. Baizer, *Organic Electrochemistry*, 3rd edition, Marcel Dekker, New York, 1991.
2. A.J. Bard, H. Lund, *Encyclopedia of the Electrochemistry of the Elements*, Marcel Dekker, New York, Vols. 11–15, 1978–1984.
3. T. Shono, *Electroorganic Synthesis*, Academic Press, London, 1991.
4. P. Delahay, *New Instrumental Methods in Electrochemistry*, Interscience, New York, 1952.
5. A.J. Bard, L.R. Faulkner, *Electrochemical Methods*, John Wiley, New York, 1980.
6. C.K. Mann, 'Nonaqueous Solvents for Electrochemical Use', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 57–134, Vol. 3, 1969.
7. H. Feess, H. Wendt, 'Performance of Two-phase-electrolyte Electrolysis', in *Techniques of Electroorganic Synthesis, Part III*, eds. N.L. Weinberg, B.V. Tilak, John Wiley & Sons, New York, 81–177, 1982.
8. R.N. Adams, *Electrochemistry at Solid Electrodes*, Marcel Dekker, New York, 1969.
9. D.T. Sawyer, J.L. Roberts, *Experimental Electrochemistry for Chemists*, John Wiley, New York, 1974.
10. D.J.G. Ives, G.J. Janz, *Reference Electrodes*, Academic Press, New York, 1961.
11. C.P. Andrieux, J.-M. Savéant, 'Electrochemical Reactions', in *Investigations of Rates and Mechanisms of Reactions*, ed. C. Bernasconi, John Wiley, New York, 305–390, Vol. 6, 4/E, part 2, 1986.
12. D. Garreau, J.-M. Savéant, 'Linear Sweep Voltammetry. Compensation of Cell Resistance and Stability. Determination of the Residual Uncompensated Resistance', *J. Electroanal. Chem.*, **35**, 309–331 (1972).
13. C.P. Andrieux, 'Terminology and Notations for Multistep Electrochemical Reaction Mechanisms', *Pure Appl. Chem.*, **66**, 2245–2250 (1994).
14. W.J. Albery, *Electrode Kinetics*, Clarendon Press, Oxford, 1975.
15. F. Opekar, P. Beran, 'Rotating Disk Electrodes', *J. Electroanal. Chem.*, **69**, 1–115 (1976).
16. W.J. Albery, M.L. Hitchman, *Ring-disk Electrodes*, Clarendon Press, Oxford, 1971.
17. D.D. MacDonald, *Transient Techniques in Electrochemistry*, Plenum Press, New York, 1977.
18. R.S. Nicholson, I. Shain, 'Theory of Stationary Electrode Polarography. Single Scan and Cyclis Methods Applied to Reversible, Irrevesible and Kinetic Systems', *Anal. Chem.*, **36**, 706–723 (1964).
19. J. Osteryoung, J.J. O'Dea, 'Square-wave Voltammetry', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 209–308, Vol. 14, 1986.
20. D.E. Smith, 'AC Polarography and Related Techniques: Theory and Practice', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–148, Vol. 1, 1966.
21. E.M. Arnett, R.A. Flowers, 'Bond Cleavage Energies for Molecules and Their Associated Radical Ions', *Chem. Soc. Rev.*, **22**, 9–15 (1993).
22. C.P. Andrieux, P. Hapiot, J. Pinson, J.-M. Savéant, 'Determination of Formal Potentials of Chemically Unstable Redox Couples by Second Harmonic Alternating Current Voltammetry and Cyclic Voltammetry. Application to the oxidation of thiophenoxide ions', *J. Am. Chem. Soc.*, **115**, 7783–7788 (1993).
23. C.P. Andrieux, D. Garreau, P. Hapiot, J. Pinson, J.-M. Savéant, 'Fast Sweep Cyclic Voltammetry at Ultramicroelectrodes. Evaluation of the Method for Fast Electron-transfer Kinetic Measurements', *J. Electroanal. Chem.*, **243**, 321–335 (1988).
24. C.P. Andrieux, D. Garreau, P. Hapiot, J.-M. Savéant, 'Ultramicroelectrodes. Cyclic Voltammetry Above One Million Volts per Second', *J. Electroanal. Chem.*, **248**, 447–450 (1988).
25. I. Montenegro, M.A. Queiros, J.L. Daschbach (eds.), *Microelectrodes: Theory and Applications*, NATO ASI Series, Series E, Kluwer Academic, Dordrecht, Vol. 197, 1993.
26. V.V. Konavalov, A.M. Raitsimring, Y.D. Tsvetkov, I.I. Bilkis, 'Photoelectrochemical Study of the Radical Anion Cleavage Rates in Aromatic Molecules: Halobenzoic Acids', *Chem. Phys. Lett.*, **157**, 257–264 (1989).
27. D.D.M. Wayner, D.J. McPhee, D. Griller, 'Oxidation and Reduction Potentials of Transient Free Radicals', *J. Am. Chem. Soc.*, **110**, 132–137 (1988).
28. P. Hapiot, V. Konovalov, J.-M. Savéant, 'Application of Laser Pulse Photoinjection of Electrons from Metal Electrodes to the Determination of Reduction Potentials of Organic Radicals in Aprotic Solvents', *J. Am. Chem. Soc.*, **117**, 1428–1434 (1995).
29. N.S. Hush, 'Adiabatic Theory of Outer-sphere Electron-transfer Reactions in solution', *Trans. Faraday Soc.*, **57**, 557–580 (1961).
30. R.A. Marcus, 'On the Theory of Electron-transfer Reactions VI.Unified Treatment for Homogeneous and Electrode Reactions', *J. Chem. Phys.*, **43**, 679–701 (1965).
31. L. Eberson, *Electron Transfer Reactions in Organic Chemistry*, Springer, Berlin, 1987.
32. H. Kojima, A.J. Bard, 'Determination of Rate Constants for the Electroreduction of Aromatic Compounds and Their Correlation with Homogeneous Electron Transfer Rates', *J. Am. Chem. Soc.*, **97**, 6317–6324 (1975).
33. J.-M. Savéant, 'A Simple Model for the Kinetics of Dissociative Electron Transfer in Polar Solvents. Application to the Homogeneous and Heterogeneous Reduction of Alkyl Halides', *J. Am. Chem. Soc.*, **109**, 6788–6795 (1987).
34. J.-M. Savéant, 'Dissociative Electron Transfer', in *Adv. Electron Transfer Chem.*, **4**, 53–116 (1994).
35. J.C. Imbeaux, J.-M. Savéant, 'Convolution Potential Sweep Voltammetry. I. Introduction', *J. Electroanal. Chem.*, **44**, 169–187 (1973).

36. D. Garreau, J.-M. Savéant, D. Tessier, 'A Computer Coupled High Frequency AC Technique for Investigating Fast Non-volmerian Electron Transfers', *J. Electroanal. Chem.*, **103**, 321 (1979).
37. D. Garreau, J.-M. Savéant, D. Tessier, 'Potential Dependence of the Electrochemical Transfer Coefficient. An Impedance Study of the Reduction of Aromatic Compounds', *J. Phys. Chem.*, **83**, 3003–3007 (1979).
38. J.-M. Savéant, D. Tessier, 'Potential Dependence of the Electrochemical Transfer Coefficient. Reduction of Some Nitro-compounds in Aprotic Media', *J. Phys. Chem.*, **81**, 2192–2197 (1977).
39. C.P. Andrieux, I. Gallardo, J.-M. Savéant, 'Outer-sphere Electron-transfer Reduction of Alkyl Halides: A Source of Alkyl Radicals or of Carbanions? Reduction of Alkyl Radicals', *J. Am. Chem. Soc.*, **111**, 1620–1626 (1989).
40. C.P. Andrieux, I. Gallardo, J.-M. Savéant, K.B. Su, 'Dissociative Electron Transfer. Homogeneous and Heterogeneous Reductive Cleavage of the Carbon–Halogen Bond in Simple Aliphatic Halides', *J. Am. Chem. Soc.*, **108**, 638–647 (1986).
41. C.P. Andrieux, A. Le Gorande, J.-M. Savéant, 'Electron Transfer and Bond Breaking. Examples of Passage from a Sequential to a Concerted Mechanism in the Electrochemical Reductive Cleavage of Arylmethyl Halides', *J. Am. Chem. Soc.*, **114**, 6892–6904 (1992).
42. C.P. Andrieux, J.-M. Savéant, D. Zann, 'Relationship Between Reduction Potentials and Cleavage Rates in Aromatic Molecules', *Nouv. J. Chim.*, **8**, 107–115 (1984).
43. C.P. Andrieux, J.-M. Savéant, A. Tallec, R. Tardivel, C. Tardy, 'Concerted and Stepwise Dissociative Electron Transfers. Oxidizability of the Leaving Group and Strength of the Breaking Bond as Mechanism and Reactivity Governing Factors Illustrated by the Electrochemical Reduction of α-Substituted Acetophenones', *J. Am. Chem. Soc.*, **119**, 2420–2429 (1997).
44. L. Nadjo, J.-M. Savéant, 'Dimerization, Disproportionation and ECE Mechanism in the Reduction of Aromatic Carbonyl Compounds in Alkaline Media', *J. Electroanal. Chem.*, **33**, 419–451 (1971).
45. C.T. Pedersen, V.D. Parker, 'The Electrochemistry of Organic Sulfur Compounds. Part II. Cathodic Reduction of 1,2-Dithiolylium Ions. Redox Analogy to the Tropylium Bitrotenyl System', *Tetrahedron Lett.*, **9**, 767–770 (1972).
46. J.-M. Savéant, E. Vianello, 'Potential Sweep Voltammetry. General Theory of Chemical Polarization', *Electrochim. Acta*, **12**, 629–646 (1967).
47. C.P. Andrieux, L. Nadjo, J.-M. Savéant, 'Electrodimerization. I. One Electron Irreversible Dimerization. Diagnostic Criteria and Rate Determination Procedures for Voltammetric Studies', *J. Electroanal. Chem.*, **26**, 147–186 (1970).
48. F. Ammar, J.-M. Savéant, 'Thermodynamics of Successive Electron Transfers. Entropy Effects in the Series of Polynitro Compounds', *J. Electroanal. Chem.*, **47**, 115–125 (1973).
49. M.E. Peover, 'Electrochemistry of Aromatic Hydrocarbons and Related Substances', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–29, Vol. 2, 1967.
50. M. Mastragostino, L. Nadjo, J.-M. Savéant, 'Disproportionation and ECE Mechanisms. I. Theoretical Analysis for Linear Sweep Voltammetry', *Electrochim. Acta*, **13**, 721–749 (1968).
51. C. Amatore, J.-M. Savéant, 'ECE and Disproportionation. V. Stationary State General Solution. Application to Linear Sweep Voltammetry', *J. Electroanal. Chem.*, **85**, 27–46 (1977).
52. C. Amatore, J.-M. Savéant, 'Electrochemical Hydrogenation of Aromatic Hydrocarbons. Discrimination Between ECE and Disproportionation Mechanisms by Double Potential Step Chronoamperometry', *J. Electroanal. Chem.*, **107**, 353–364 (1980).
53. C.P. Andrieux, C. Blocman, J.-M. Dumas-Bouchiat, F. M'Halla, J.-M. Savéant, 'Homogeneous Redox Catalysis of Electrochemical Reactions. V. Cyclic Voltammetry', *J. Electroanal. Chem.*, **113**, 19–40 (1980).
54. D. Lexa, J.-M. Savéant, K.B. Su, D.L. Wang, 'Chemical Versus Redox Catalysis of Electrochemical Reactions. Reduction of *trans*-1,2-Dibromocyclohexane by Electrogenerated Aromatic Anion Radicals and Low Oxidation State Metallo-porphyrins', *J. Am. Chem. Soc.*, **109**, 6464–6470 (1987).
55. R.D. Little, M.K. Schwaebe, 'Reductive Cyclizations at the Electrode', *Top. Curr. Chem.*, **185**, 1–48 (1997).
56. C.P. Andrieux, F. Gonzales, J.-M. Savéant, 'Derivatization of Carbon Surfaces by Anodic Oxidation of Arylacetates. Electrochemical Manipulation of the Grafted Films', *J. Am. Chem. Soc.*, **119**, 4292–4300 (1997).
57. C.P. Andrieux, C. Blocman, J.-M. Dumas-Bouchiat, J.-M. Savéant, 'Heterogeneous and Homogeneous Electron Transfers to Aromatic Halides. A Redox Catalysis Study in the Halobenzene and Halopyridine Series', *J. Am. Chem. Soc.*, **101**, 3431–3441 (1979).
58. L. Nadjo, J.-M. Savéant, D. Tessier, 'Electrodimerization. XI. Coupling Mechanism of an Activated Olefin: *p*-Methylbenzylidenemalononitrile as Studied by Convolution Potential Sweep Voltammetry', *J. Electroanal. Chem.*, **64**, 143–154 (1975) .
59. D. Larumbe, I. Gallardo, C.P. Andrieux, 'Anodic Oxidation of Some Tertiary Amines', *J. Electroanal. Chem.*, **304**, 241–247 (1991).
60. H.S. Nalwa, *Handbook of Organic Conductive Molecules and Polymers*, John Wiley, New York, 1997.
61. C.P. Andrieux, P. Audebert, P. Hapiot, J.-M. Savéant, 'Identification of the First Steps of the Electrochemical Polymerization of Pyrroles by Means of Fast Potential-step Techniques', *J. Phys. Chem.*, **95**, 10158–10164 (1991).

62. L. Guyard, P. Hapiot, P. Neta, 'Redox Chemistry of Bipyrroles: Further Insights into the Oxidative Polymerization Mechanism of Pyrrole and Oligopyrroles', *J. Phys. Chem. B*, **101**, 5698–5706 (1997).
63. J. Pinson, J.-M. Savéant, 'Electrochemically Induced Aromatic Nucleophilic Substitution', *J. Am. Chem. Soc.*, **100**, 1506–1510 (1978).
64. R.A. Rossi, R.H. Rossi, *Aromatic Nucleophilic Substitution by the SRN1 Mechanism*, ACS Monograph 178, American Chemical Society, Washington, DC, 1983.
65. C. Amatore, J. Chaussard, J. Pinson, J.-M. Savéant, A. Thiébault, 'Electrochemically Induced Aromatic Nucleophilic Substitution in Liquid Ammonia. Competition with Electron Transfer', *J. Am. Chem. Soc.*, **101**, 6012–6020 (1979).
66. C. Amatore, M. Azzabi, A. Jutand, 'Rates and Mechanism of the Reversible Oxidative Addition of (*Z*)- and (*E*)-1,2 Dichloroethylene to Low Liganded Zerovalent Palladium', *J. Am. Chem. Soc.*, **113**, 1640–1677 (1991).
67. C. Amatore, A. Jutand, A. Suarez, 'Intimate Mechanism of Oxidative Addition of Zerovalent Palladium Complexes in Presence of Halide Ions and Its Relevance to the Mechanism of Palladium-catalyzed Nucleophilic Substitutions', *J. Am. Chem. Soc.*, **115**, 9531–9541 (1993).

Pulse Voltammetry

Michael T. Carter
Eltron Research Inc., Boulder, USA

Robert A. Osteryoung
North Carolina State University, Raleigh, USA

1 Introduction **1**
2 Theory and Operating Principles **2**
2.1 Chronoamperometric Response to a Potential Step 2
2.2 Pulse Voltammetry at Electrodes of Classical Dimensions 5
2.3 Pulse Polarography and Voltammetry at Mercury Electrodes 10
2.4 Pulse Voltammetry at Ultramicroelectrodes 12
3 Instrumentation and Operation **12**
3.1 Commercial Instrumentation 13
3.2 General Considerations 14
4 Applications and Examples **15**
4.1 Electrochemical Studies 15
4.2 Analysis and Detection 18
5 Future Prospects **21**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **23**

This article covers fundamentals and applications of pulse voltammetry. These techniques use a train of timed potential pulses to perturb the electrode–solution interface. Electron transfer reactions are driven at the electrode surface during each pulse. Current flowing as a result of electron transfer is measured at specified times during the pulse. Popular pulse voltammetric methods include normal pulse voltammetry (NPV), reverse pulse voltammetry (RPV), square wave voltammetry (SWV) and differential pulse voltammetry (DPV). Pulse voltammetry generally provides improved signal-to-noise ratio (S/N) and a decreased lower limit of detection (LLD) compared with continuous electrolysis methods such as cyclic voltammetry (CV), by controlling the extent of electrolysis prior to current measurement and discrimination against background processes. A concise discussion of the basic concepts behind the various pulse voltammetric methods is presented, followed by method-specific discussions of the most popular methods. Instrumentation and practical experimental considerations are covered. Finally, a discussion of selected illustrative applications of pulse voltammetry to analysis and other chemical problems is presented, demonstrating the use of these techniques as a laboratory tool.

1 INTRODUCTION

Pulse voltammetry comprises a family of electrochemical methods which use a train of potential pulses and corresponding current responses to generate a current–potential curve or voltammogram. Pulse voltammetry takes advantage of the difference between Faradaic (electron transfer) and nonFaradaic (interfacial charging) processes to discriminate against background, resulting in improved S/N compared with continuous electrolysis [direct current (DC)] methods, such as CV. A variety of specialized pulse sequences, conditioning potentials and other steps can be included in the complete experimental waveform, making these methods very versatile for electrochemical studies. The applied potential required to oxidize or reduce a particular species can be used to identify it in many cases. This is particularly useful in differential methods, including SWV and DPV, which provide a peaked voltammogram. The analytical utility of pulse voltammetry, as with other voltammetric methods, comes from proportionality between concentration of redox-active species in solution and measured current.

The predominant modern techniques of pulse voltammetry are SWV, NPV, RPV and DPV. SWV is a method which combines features of the pulse method with continuous electrolysis. Strictly, ''voltammetry'' refers to techniques applied at solid electrodes, e.g. platinum, carbon or gold and to mercury electrodes exclusive of the dropping mercury electrode (DME), such as the static mercury drop electrode (SMDE), the hanging mercury drop electrode (HMDE) and the thin-film mercury electrode (TFME). The term ''polarography'' strictly applies to the DME only. Normal and reverse pulse methods at the DME, for example, will be referred to as normal pulse polarography (NPP) and reverse pulse polarography (RPP), respectively, while the corresponding experiments performed at, say, a glassy carbon (GC) disk electrode are NPV and RPV. The reader will no doubt encounter many violations of this rule and a confusing array of specialized nonstandard jargon in the electrochemical literature, but one should keep in mind that all these techniques have a common basis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 THEORY AND OPERATING PRINCIPLES

Pulse voltammetric methods are potentially very powerful analysis tools, sometimes capable of detection limits rivaling more popular analytical methods such as atomic absorption spectroscopy. Voltammetric measurements may also provide information on the thermodynamics and kinetics of an interfacial electrochemical reaction. Information is contained in the redox potential at which an electron transfer reaction proceeds, which is often characteristic of the particular species reacting and in the current passed, which is a direct measure of the overall rate of the reaction and is commonly proportional to analyte concentration.

We shall concentrate on basic operating principles, which will therefore be limited to simple reversible cases to illustrate the main concepts, although common complications of which the reader should be aware are mentioned where appropriate. The reader is directed to the literature citations for additional detail.

2.1 Chronoamperometric Response to a Potential Step

Chronoamperometry, measurement of current flowing in response to a potential step as a function of time, is the fundamental experiment behind all pulse voltammetry. The most basic results for a simplified redox system are described here. For convenience, electrochemists are fond of discussing generic electron transfer reactions in terms of the O/R redox couple (Equation 1):

$$O + ne^- \rightleftharpoons R \qquad (1)$$

where O is the oxidized form of the couple, R is the reduced form and n is the number of electrons transferred in the reaction. In the simplest case, O and R differ by one electron, but are otherwise essentially identical, e.g. Fe^{3+}/Fe^{2+}. We also usually make assumptions that both O and R are soluble, i.e. neither adsorbs on the electrode, and that both are chemically stable on the *timescale of the measurement*. The experimental timescale is an important aspect of pulse voltammetry. Furthermore, the electron transfer reaction is assumed to be reversible, i.e. the concentrations of O and R at the electrode–solution interface obey the Nernst equation. Clearly, this is a lot to expect from a real system. We shall discuss caveats and exceptions to the simple case as appropriate.

2.1.1 Reversible Systems Under Diffusion Control

We shall consider a case in which the working electrode is immersed in a solution of O.[1–4] We generally ignore the rest of the electrochemical cell, concentrating on what happens at the working electrode, which is the detector in an analytical application. It is common to speak of reduction of O to R, but the reasoning applies equally to oxidations. During the electrochemical reaction, O is converted to R at the electrode surface. This conversion depends on time-dependent phenomena such as diffusion and heterogeneous reaction rate, and so pulsed electrochemical methods are characterized by currents which vary with time. As conversion proceeds, the total analytical concentration of O dissolved in solution, denoted C_O^*, is partitioned between O and R. These concentrations, denoted C_O and C_R, respectively, change with time and distance from the electrode surface, with the sum of C_O and C_R equaling C_O^*. Designation of concentration of O or R as a function of distance (x) and time (t) is often given in shorthand as $C_O(x, t)$ and $C_R(x, t)$, or some variation thereof.[1]

When the electrode–solution interface is perturbed by application of a potential, denoted E (volts), and E is sufficiently extreme compared with $E^{\circ'}_{O/R}$, the standard formal potential of the O/R system, to drive redox Equation (1) at its maximum rate from left to right, it is common (but not necessarily always the case) that the current, usually denoted i (amperes), is controlled by diffusion of O to the electrode surface. A typical potential step traversing the $E^{\circ'}_{O/R}$ region and resulting current response are shown in Figure 1(a) and (b). C_O at the surface of the electrode rapidly decreases to zero after pulse application and C_R increases correspondingly, as shown in Figure 2(a). The exact concentrations of O and R at the electrode surface are fixed by the Nernst equation for a reversible reaction (Equation 2):

$$E = E^{\circ'} - \frac{0.059}{n} \log \frac{C_R(x = 0)}{C_O(x = 0)} \qquad (2)$$

where E is the applied potential, $E^{\circ'}$ is the standard formal potential of the O/R couple and $C_R(x = 0)$ and $C_O(x = 0)$ are the concentrations of O and R at the electrode surface. The term $0.059/n$ is RT/nF at 25 °C, where R is the gas

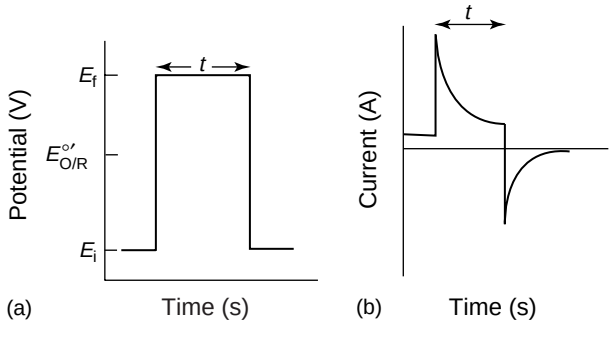

Figure 1 (a) A potential step experiment and (b) the resulting chronoamperometric response. $E^{\circ'}_{O/R}$ is the formal potential of the O/R redox couple; E_i and E_f denote initial and final values of the potential, respectively.

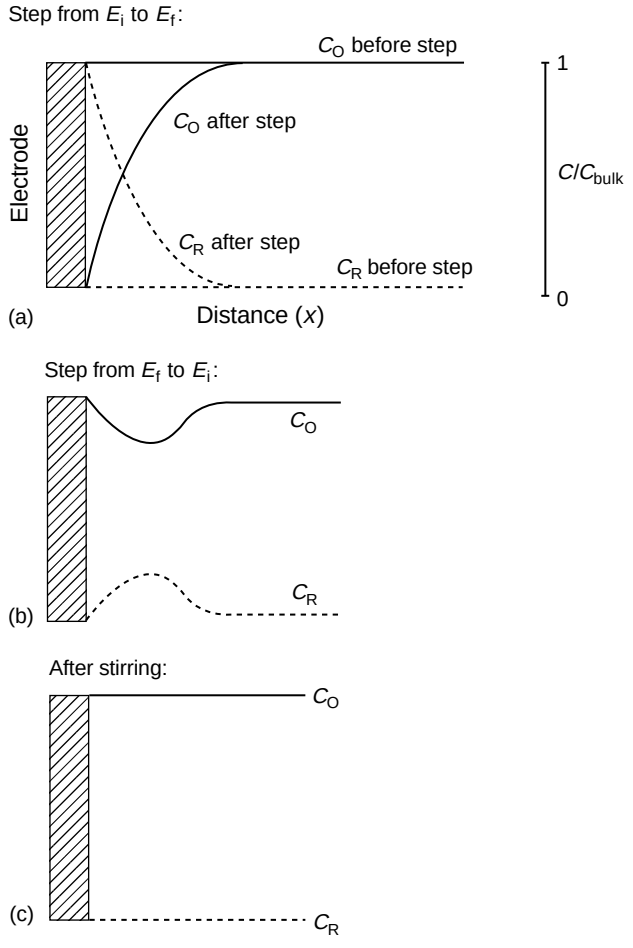

Figure 2 Concentration–distance profiles at various stages of a potential step: (a) changes in concentrations of oxidized (C_O) and reduced (C_R) forms following a potential step to a diffusion-limited value; (b) effect of returning the potential to the initial value; (c) re-establishment of initial conditions by convection.

constant, T is absolute temperature (kelvin) and F is the Faraday constant. Once C_O has reached zero at the electrode surface, a concentration gradient of O exists and material must diffuse to the electrode to sustain the reduction, hence ''diffusion-limited current''. When the reaction is fast, current is controlled by the diffusion gradient of O established at the electrode surface, which is characterized by the diffusion coefficient, D, of the diffusing species. The current, $i(t)$, at any time during the potential step is given by Equation (3):

$$i(t) = \frac{nFAD^{1/2}C_O^*}{\pi^{1/2}t^{1/2}} \tag{3}$$

where A is electrode area and the other terms were defined previously. Equation (3) is commonly known as the Cottrell equation.[1–4] The main points to take from this are that diffusion-controlled currents decay with a $t^{-1/2}$ dependence and they are proportional to analyte concentration.

Most pulse voltammetric experiments rely on the re-establishment of ''initial boundary conditions'' at the electrode surface between potential steps. Following the step from E_i to E_f, above, returning the potential to E_i (Figure 2b) drives R at the electrode surface back to O. A slight concentration gradient still exists, however, since some R has diffused away from the electrode surface during the step. Stirring or other modes of forced convection, as shown in Figure 2(c), or waiting for a sufficiently long time at E_i are commonly used to remove residual gradients and return the interface to its initial state.

2.1.2 Charging Current

An electrochemical interface has the properties of a capacitor.[5] When an electrode in contact with an aqueous medium containing dissolved ionic species is subjected to a potential step, the ions redistribute in response to change in interfacial potential. This redistribution results in a structure called the electrical double layer. Charging of the interface can be thought of conveniently as charge separation in a capacitor (Equation 4):

$$Q = CE \tag{4}$$

where charge built up on the plates of the capacitor (Q, coulombs) is proportional to the potential applied (E) and the capacitance (C, farads), which depends on the dielectric properties of the medium. Experimentally, a potential step applied to the interface results in current–time behavior described by a simple RC circuit (Equation 5):

$$i(t) = \frac{\Delta E}{R_s} e^{-t/R_s C_{dl}} \tag{5}$$

where $i(t)$ is the time-dependent current, t is time, ΔE is applied potential, R_s is solution resistance and C_{dl} is the double-layer capacitance. Typical capacitances are of the order of several $\mu F\,cm^{-2}$ for metal electrodes in contact with aqueous electrolytes. Charging current decays as e^{-t} whereas current from a Faradaic process decays with a slower $t^{-1/2}$ dependence, as shown in Figure 3. This is an important result for pulse voltammetry because it suggests that proper timing will allow significant Faradaic current to be measured at times when charging contributions are diminished. The ability to discriminate against background processes by such an approach is one feature that gives pulse methods significantly improved S/N over DC waveforms.

2.1.3 Kinetic Complications

It is convenient to think of most electrochemical processes as reversible, since this is the simplest case conceptually

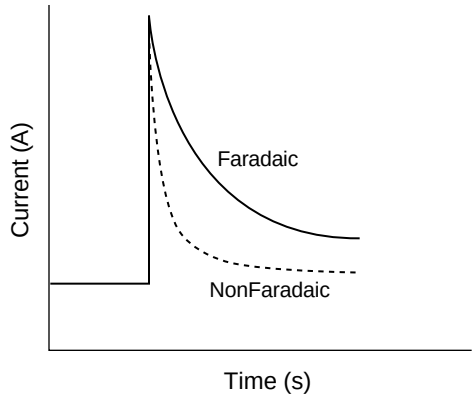

Figure 3 Chronoamperometry of Faradaic (electron transfer) and nonFaradaic (charging) processes.

and mathematically. Reversibility depends on the rate of the redox reaction vs timescale of the experiment. In the real world there are not many perfectly reversible systems. Many nonreversible systems, however, can be treated as reversible without introducing significant error. It is also possible for a redox reaction to appear reversible by slowing the measurement speed, but this often comes at the cost of reduced signal.

The treatment of electrode kinetics is too involved for a detailed discussion here, but it is useful to understand the meaning of kinetic limitations in terms of analytical applications. A main concept of all voltammetric methods is that the current is a direct reflection of the rate of electrolysis of O to R[6] (Equation 6):

$$i = nFAv \tag{6}$$

where v is rate of conversion of O to R in $\mathrm{mol\,s^{-1}\,cm^{-2}}$. The area term comes from the heterogeneous nature of the interfacial process, conversion per unit electrode area per unit time. Kinetic complications enter into v. If one thinks of the O/R conversion of Equation (1) as a general kinetic process (Equation 7):

$$\mathrm{O} + n\mathrm{e}^- \underset{k_\mathrm{b}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} \mathrm{R} \tag{7}$$

where k_f and k_b are the forward and backward rate constants, respectively, then the current, i_net, which reflects the net rate of conversion of O to R is given by Equation (8):

$$i_\mathrm{net} = i_\mathrm{c} - i_\mathrm{a} = nFAk_0 \times C_\mathrm{O}(0,t)\mathrm{e}^{\frac{-\alpha nF(E-E^{\circ'})}{RT}} - C_\mathrm{R}(0,t)\mathrm{e}^{\frac{(1-\alpha)nF(E-E^{\circ'})}{RT}} \tag{8}$$

where i_c and i_a are currents for the forward and reverse parts of Equation (7) at the given potential, E, and k_0 is the standard heterogeneous rate constant ($\mathrm{cm\,s^{-1}}$), the reaction rate constant when $E = E^{\circ'}$. $C_\mathrm{O}(0,t)$ and $C_\mathrm{R}(0,t)$ are the concentrations of O and R at the electrode surface at time t, α is the transfer coefficient (which usually has a value of ca. 0.5) and $E - E^{\circ'}$ is the overpotential, denoted η. Overpotential corresponds to reaction driving force (applied potential) in excess of the standard potential that must be supplied to drive the O/R conversion. The current depends on the net rate of conversion of O to R, which in turn depends on applied potential. The determination of kinetic parameters from voltammetric data has been greatly simplified by computer-assisted fitting of theory to experiment. Many modern electrochemical instruments incorporate such procedures as software options.

Many physical and chemical factors manifest themselves as slow electron transfer. Specific adsorption of ions from solution on to an electrode surface, distance requirements for electron transfer at the electrode–solution interface, chemical reactions coupled to electron transfer and adsorption of macromolecular contaminants on the electrode can all result in sluggish kinetics.

Real samples often contain species which can strongly adsorb on an electrode surface, e.g. proteins in biological matrices or humic substances in fresh water. Adsorption is often strongly dependent on the nature of the electrode material. The effect of these adsorbates is to slow the rate of electron transfer at any applied potential by increasing the distance over which this event must take place. This means that more extreme potentials must be applied to drive the reaction at a rate comparable with that observed in the absence of the blockage. Such a kinetic effect typically broadens the volammetric wave, decreases its magnitude and shifts the wave to more extreme potentials. This has important implications for detection since it can degrade both S/N and resolution.

2.1.4 Surface-confined Redox Systems

While adsorption of macromolecular interferents is usually a hindrance to analytical detection, confinement of an analyte at the electrode surface by adsorption[7,8] or other types of intentional confinement[9] can actually be beneficial. Adsorption offers the opportunity to concentrate reactant at the interface and eliminate diffusion from the analytical signal. Anodic stripping voltammetry (ASV)[10] and cathodic stripping voltammetry (CSV)[11] both take advantage of surface effects and preconcentration of analyte to allow extremely low detection limits and high analytical sensitivity.

2.2 Pulse Voltammetry at Electrodes of Classical Dimensions

We make the distinction between "classically" sized electrodes, i.e. electrodes with dimensions of the order of 1–0.1 mm, and microelectrodes, having at least one microscopic dimension (i.e. <0.1 mm), because the nature of analyte mass transfer and hence current response change as the dimensions of the electrode decrease. Most common electrodes that one will encounter are the "macro" variety, but the use of microelectrodes, discussed in section 2.4, is increasing owing to their advantageous properties.

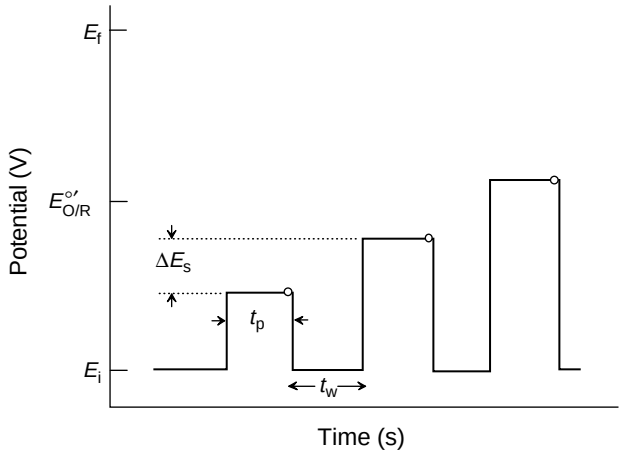

Figure 4 The NPV waveform with experimental parameters. Current measurement at the end of each pulse is denoted by an open circle.

2.2.1 Normal and Reverse Pulse Voltammetry

The NPV[1,2,12–15] waveform is illustrated in Figure 4. The electrode is subjected to series of potential steps, each originating from the same initial potential (E_i), where O is stable. The potential step duration is t_p. At a point late in each step, the current is measured which is denoted by an open circle superimposed on the E–t waveform in Figure 4. The potential is then returned to E_i for a waiting period, t_w. Returning the potential to E_i drives any R at the electrode surface back to O, as shown in Figure 2(b). During t_w, any residual R further removed from the interface diffuses away from the electrode. It is often useful during this period to stir the solution or otherwise renew the initial boundary conditions, so that the concentration of O is everywhere equal to the original bulk concentration. If $t_w \gg t_p$, stirring may not be necessary. Following the first step, the potential is stepped incrementally to more extreme values, $E_i + 2\Delta E_s$, $E_i + 3\Delta E_s$ and so forth, until E_f is reached. This sequence of evenly spaced, progressively more extreme potential steps and corresponding current measurements span the region of $E^{\circ\prime}_{O/R}$. The early steps produce no electrolysis of O to R, as shown in Figure 5(a), and so generate only background current by double-layer charging. Successive potential steps eventually perturb the concentration profile of O at the electrode surface as the step approaches $E^{\circ\prime}$ and currents increase accordingly. At sufficiently extreme potentials, the diffusion layer is established quickly on the timescale of the potential step and the flux of O at the electrode surface is not changed

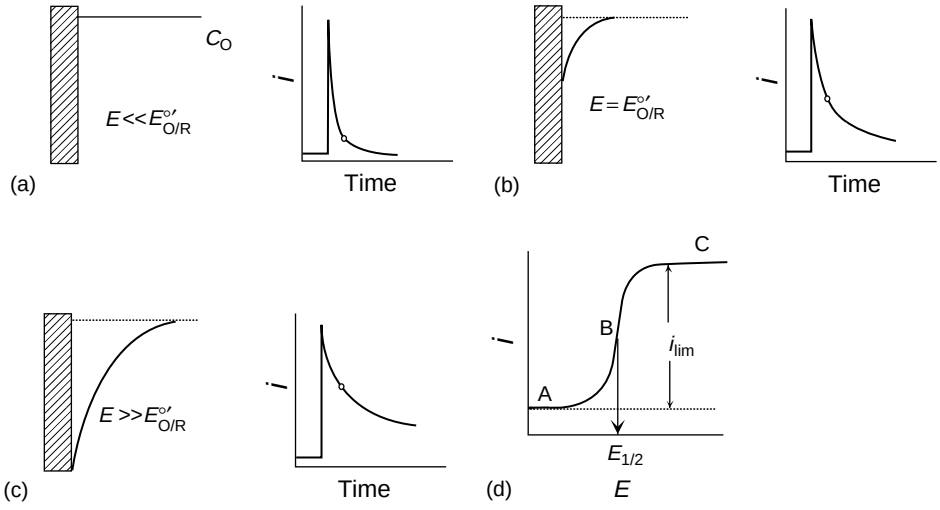

Figure 5 Origin of the NPV in a series of potential steps to various values vs $E^{\circ\prime}_{O/R}$. (a), (b) and (c) show potential steps to values positive of, equal to and negative of $E^{\circ\prime}_{O/R}$ and the corresponding chronoamperometric trace. Plotting currents sampled at t_p vs the step potential yields the sigmoid NPV curve (d). (Reproduced in part by permission of Marcel Dekker from W.R. Heineman, P.T. Kissinger, 'Large Amplitude Controlled-potential Techniques', in *Laboratory Techniques in Electroanalytical Chemistry*, eds. P.T. Kissinger, W.R. Heineman, Marcel Dekker, New York, Chapter 3, 51–127, 1984.)

appreciably at potentials well beyond $E^{\circ'}_{O/R}$. The current reaches a plateau in this regime.

The NPV curve (Figure 5d) has a sigmoid shape centered around the half-wave potential, $E_{1/2}$, which corresponds to $E^{\circ'}$ when the diffusion coefficients of the oxidized (D_O) and reduced (D_R) forms are equal (Equation 9):

$$E_{1/2} = E^{\circ'} + \frac{RT}{nF} \ln \left(\frac{D_O}{D_R} \right)^{1/2} \quad (9)$$

For most practical circumstances, the difference between $E_{1/2}$ and $E^{\circ'}$ is negligible. The overall shape of the NPV trace can be cast in terms of $E_{1/2}$ and the limiting current, i_{lim}, at the plateau of the curve (Equation 10):

$$E = E_{1/2} + \frac{RT}{nF} \ln \left(\frac{i_{lim} - i}{i} \right) \quad (10)$$

A semilogarithmic plot of the rising portion of the NPV curve in the vicinity of $E_{1/2}$ therefore should have a slope of $59.1/n$ mV at 25 °C for a reversible system. This serves as a useful diagnostic criterion for reversibility, as does an absence of dependence of $E_{1/2}$ on t_p. Reversibility is not necessary for analytical detection methods based on pulse voltammetry, but resolution is maximized under these conditions.

The limiting current of an NPV curve is given by Equation (3), where t is replaced by the pulse time, t_p (Equation 11):

$$i_{lim} = \frac{nFAD^{1/2}C}{\pi^{1/2} t_p^{1/2}} \quad (11)$$

i_{lim} is linearly proportional to concentration, hence its analytical utility. For a fixed pulse time the simple Equation (12):

$$i_{lim} = (\text{slope})C \quad (12)$$

may be used to relate signal to concentration, where the slope is $nFAD^{1/2}/\pi^{1/2}t_p^{1/2}$. Often in analysis the nature of the slope will not be important, but it will vary with electrode area and diffusion coefficient. Since limiting current is proportional to $D^{1/2}$, the dependence of i_{lim} on t_p can also be used to determine the diffusion coefficient of an electroactive species.

Figure 6 shows consecutive one-electron transfers[16] (Equations 13 and 14):

$$\text{TTF} \rightleftharpoons \text{TTF}^+ + e^- \quad (13)$$

$$\text{TTF}^+ \rightleftharpoons \text{TTF}^{2+} + e^- \quad (14)$$

where TTF is tetrathiafulvalene (CAS 31366-25-3). This figure illustrates the NPV behavior of consecutive, simple,

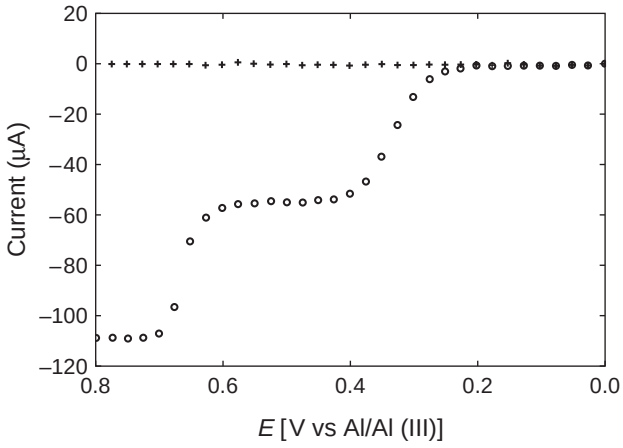

Figure 6 Two-step NPV oxidation of 4.8 mM tetrathiafulvalene at a GC disk electrode in an ambient temperature ionic liquid (25 °C). Circles are currents at the end of each pulse and crosses are currents at E_i prior to the pulse. Waiting time, t_w, at E_i was 12 s and pulse time, t_p, was 50 ms. (Reproduced by permission of The Electrochemical Society from M.T. Carter, R.A. Osteryoung, 'Heterogeneous and Homogeneous Electron Transfer Reactions of Tetrathiafulvalene in Ambient Temperature Chloroaluminate Molten Salts', *J. Electrochem. Soc.*, **141**, 1713–1720 (1994).)

reversible one-electron transfer reactions. Circles denote currents measured at the end of t_p and crosses denote currents prior to the pulse. These latter values can be used to assess boundary condition renewal. Deviations from zero indicate inconsistent renewal of boundary conditions. The value of i_{lim} for each reaction varied linearly with $t_p^{-1/2}$. $E_{1/2}$ was essentially independent of t_p for each wave, suggesting reversibility of the electron transfers.

One drawback of NPV at solid electrodes, as in Figure 6, is the time needed for proper experimental control. Following each potential pulse, the working electrode was returned to 0 V and the solution stirred to renew the boundary conditions. Allowing 12 s for this process at each pulse took 6.6 min to acquire 33 data points. The use of renewable electrodes, such as the SMDE, can significantly speed data acquisition. Although NPV is useful for certain measurements, it is extremely slow when performed at solid, stationary electrodes. Of course, if one decreases t_p then t_w can be decreased commensurately.

RPV[12,14,17,18] is complementary to NPV. Typical RPV waveforms are shown in Figure 7(a) and (b). The two common variants are RPV and reverse pulse voltammetry with waiting (RPW). RPV is essentially an inverted NPV experiment, whereas RPW entails additional potential steps in the nonFaradaic region positive of $E^{\circ'}_{O/R}$. Like NPV, the potential waveform is chosen to span the region of $E^{\circ'}_{O/R}$. However, unlike NPV, the experiment

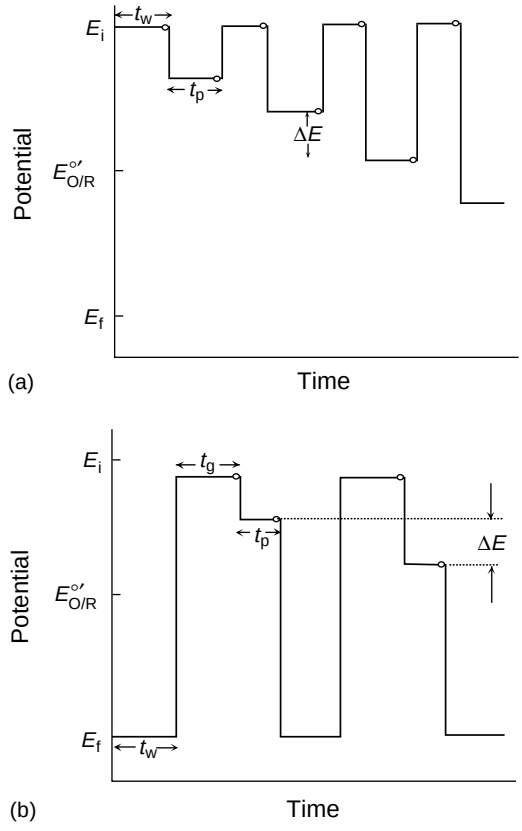

Figure 7 (a) Reverse pulse and (b) reverse pulse with waiting voltammetric waveforms and relevant parameters. Points of current measurement are denoted by open circles.

is initiated at a potential where R is generated, typically at a diffusion-controlled rate. After holding the potential for t_w at a value E_f where O is stable, the potential is stepped to E_i, where R is generated for time t_g. This generates a diffusion layer of R at the electrode surface and current flows according to Equation (3). This current is usually not of analytical interest. After generating R, the potential is pulsed for time t_p to successively more positive values with current measurement at the end of each pulse. The potential is finally returned to E_f and stirring initiated to renew the initial boundary conditions between pulses.

The RPV curve is shown in Figure 8. A corresponding NPV trace is shown for comparison. If both forms of the redox couple are stable, the RPV curve will have a limiting current, $i_{lim,RPV}$, equal to that of the corresponding NPV experiment. The full $i-E$ curve is displaced on the current axis from the NPV experiment, however, because of the current that flows at E_i owing to generation of R. This current, denoted i_{DC}, is given by Equation (11) for a reversible redox couple where the pulse time is t_g. Under ideal conditions, i_{DC} should be constant at all points in the experiment. The sum of i_{DC} and the reverse pulse

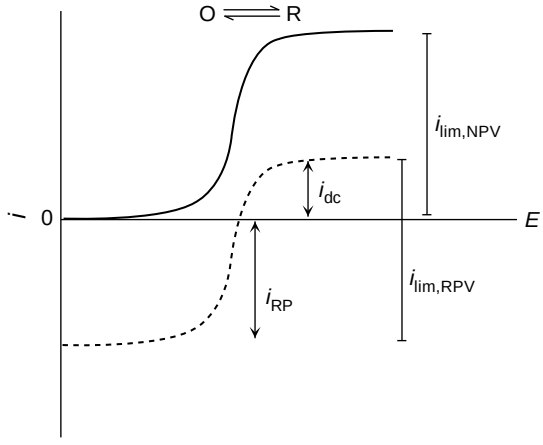

Figure 8 NPV and RPV curves for a reversible system.

current i_{RP} for electrolysis of R to O in the diffusion layer is equal to the corresponding NPV i_{lim} if all species are stable (Equation 15):

$$i_{RP} + i_{DC} = i_{lim,RPV} = i_{lim,NPV} \quad (15)$$

The value of $E_{1/2}$ from RPV should equal the NPV value in the reversible case and the shape of the curve is described completely analogously to NPV.

RPV can be used to electrogenerate a species under controlled conditions and examine its stability. For example, if the value of t_p is tens of milliseconds, but the product R decays on a timescale of seconds, then NPV will behave reversibly, since R is essentially stable on the timescale of that experiment. However, in RPV, if R is generated for several seconds, i.e. long enough to allow measurable decay, a fraction of R will not be available for reoxidation during t_p, resulting in a decrease in i_{RP} compared with the case where R is completely stable. Instability of R on the timescale of the RPV measurement could be probed by systematic variation of t_g from short times, where R is completely stable to long times, where significant loss of R occurs and $i_{RP} + i_{DC} < i_{NPV}$. RPV applications have included detection of halogenated organics,[19] determination of OH^-,[20] evaluation of stability of organometallics[21] and characterization of chemical reactions coupled to electron transfer.[22]

The use of RPV to evaluate the stability of an organometallic complex is shown in Figure 9(a) and (b).[21] Here, decomposition of $[Ru^{III}(EDTA)]_2$ dimer, the product of electroreduction of a related starting material, was followed as a function of t_p (Equation 16):

$$[Ru^{III}(EDTA)]_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2Ru^{III}(EDTA)(OH)_2 \quad (16)$$

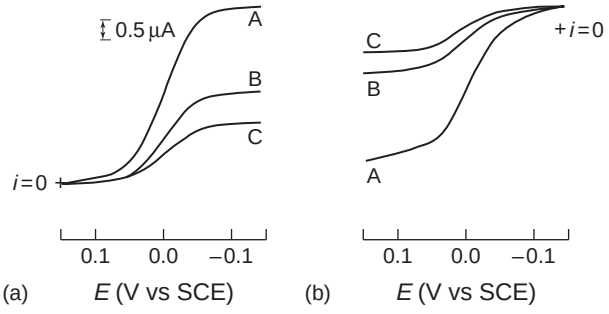

Figure 9 (a) NPV and (b) RPV of 0.25 mM $[Ru^{III\frac{1}{2}}(EDTA)_2]_2$ dimer for t_p = (a) 10.4, (b) 39.7 and (c) 90 ms. (Reprinted from C.-L. Ni, F.C. Anson, J.G. Osteryoung, 'A Pulse Polarographic Measurement of the Kinetics of Dissociation of the $[Ru^{III}(EDTA)]_2$ Dimer', *J. Electroanal. Chem.*, **202**, 101–109 (1986), Copyright 1986, with permission from Elsevier Science.)

Removal of electroactive $[Ru^{III}(EDTA)]_2$ product by Equation (16) on a timescale comparable to t_p was reflected in decrease in the ratio i_{RPV}/i_{NPV} as t_p increased. A rate constant of 0.75 s^{-1} at pH 4.8 was calculated based on this trend.

2.2.2 *Differential Pulse Voltammetry*

DPV is one of the most popular pulse voltammetric methods.[1,23–28] The DPV waveform is shown in Figure 10. Each pulse of time t_p and magnitude ΔE_p is preceded by a uniform waiting time, t_w, as in NPV, but instead of making one current measurement at the end of t_p, two measurements are made. The first (i_1) is made at τ' just prior to initiation of the pulse and the second (i_2) at τ toward the end of the pulse. After the pulse is completed, the potential does not return to E_i, as in NPV, but is incremented by a step ΔE_s, typically

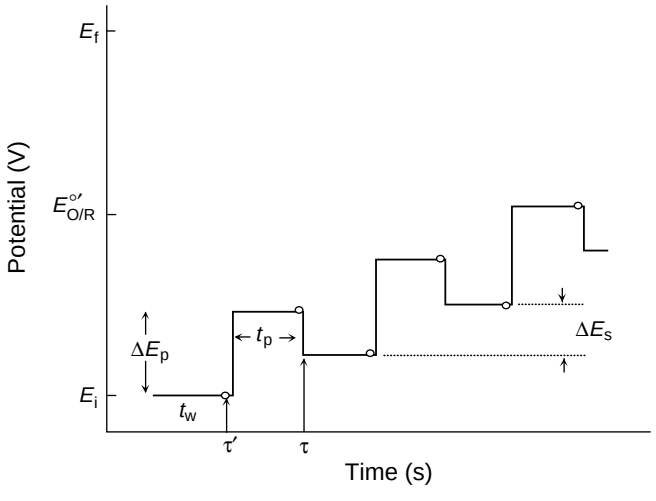

Figure 10 The DPV waveform and associated experimental parameters.

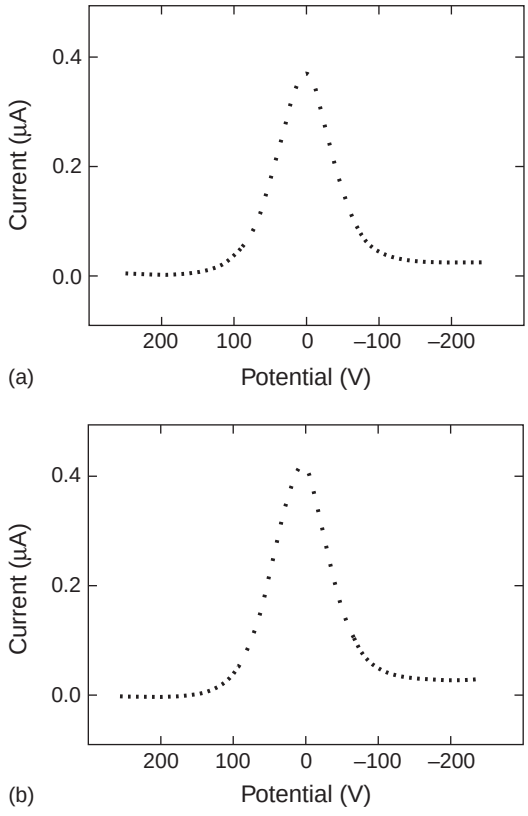

Figure 11 Simulated differential pulse voltammograms at (a) a planar electrode and (b) a DME with sphericity. Drop time, 0.5 s; pulse time, 40 ms; ΔE_s, 10 mV; ΔE_p, 10 mV. (Reproduced by permission of The American Chemical Society from J.W. Dillard, J.A. Turner, R.A. Osteryoung, 'Digital Simulation of Differential Pulse Polarography with Incremental Time Change', *Anal. Chem.*, **49**, 1246–1250 (1977).)

1–10 mV. Pulse widths and waiting times are similar to those in NPV. In differential pulse polarography (DPP) a new drop is generated at the end of t_p, as discussed in section 2.3.

The output of the experiment is the difference $i_2 - i_1$ vs the base potential E_s, hence the name of the method. DPV yields a peaked output as shown in Figure 11(a) and (b). At the beginning of the experiment, no Faradaic current flows, so $i_2 - i_1$ is very close to zero. At the other extreme, i.e. well beyond the peak, the O/R reaction is driven at its diffusion-limited rate. Further pulses do not appreciably raise the rate beyond that driven by the base potential, so again $i_2 - i_1$ is near zero. It is only in the region near $E^{o'}$ that small changes in potential between the base value and the pulse value result in large changes in current and nonzero $i_2 - i_1$.

The peaked DPV output has distinct advantages for quantitation and speciation in analytical applications. The differential nature of DPV tends to flatten sloping baselines, in addition to increasing S/N. The height of the

peak varies with pulse amplitude, ΔE_p[23] (Equation 17):

$$i_p = \frac{nFAD_O^{1/2}C_O}{\pi^{1/2}(\tau - \tau')^{1/2}} \frac{1-\sigma}{1+\sigma} \quad (17)$$

where $(1-\sigma)/(1+\sigma)$ is a function of ΔE_p and n.[23] At large pulse amplitudes $(1-\sigma)/(1+\sigma) = 1$ and Equation (17) reduces to Equation (11). The DPV curve has a maximum current equal to that in the corresponding NPV experiment under these conditions. Usually, the DPV peak current is smaller than the corresponding NPV limiting current, at typical values of ΔE_p. The enhanced sensitivity and detection limit of DPV over NPV, as much as an order of magnitude, is not due to enhancement of the Faradaic signal but to a reduction in charging current.[1] Maximization of sensitivity is essentially a trade-off between larger currents obtained at large pulse amplitudes and the increased background component that accompanies it. Larger pulse widths also broaden the wave, which decreases the ability to resolve multiple analytes. DPV has been used most extensively in conjunction with the DME and SMDE (section 2.3).

The peak potential, E_p, corresponds approximately to the NPV $E_{1/2}$. Resolution of multicomponent mixtures is defined by peak width, $W_{1/2}$, which reaches a limiting value of $90.4/n$ mV at 25 °C, as ΔE_p approaches zero.

2.2.3 Square Wave Voltammetry

SWV[29–39] employs the waveform shown in Figure 12. Following an initial period of quiet time, t_w, a symmetrical train of pulses of amplitude $2\Delta E_{SW}$ is applied to the electrode, superimposed on a small DC potential increment, ΔE_s, changed at a frequency (f) of $1/\tau$, where

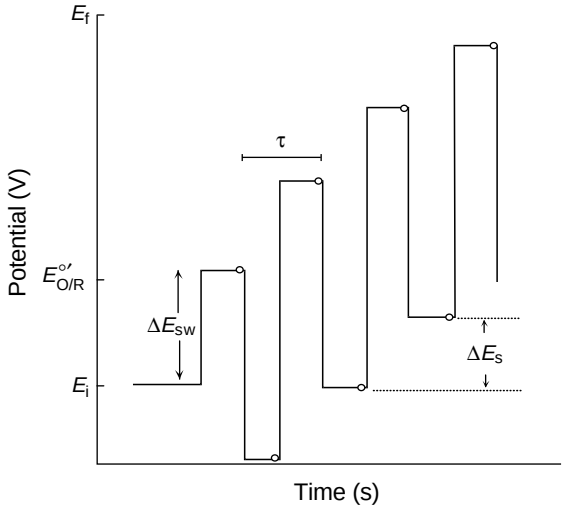

Figure 12 The SWV waveform and associated experimental parameters.

τ is the period of the square wave. Current is sampled at the end of each forward (i_f) and reverse (i_r) half-cycle. The difference $i_f - i_r$, called the net current, is plotted against the DC potential to yield the net voltammogram, where $i_{net} = i_f - i_r$.

SWV can be performed very quickly compared with NPV and RPV since boundary conditions do not need to be renewed before each pulse. For example, a typical square wave frequency of 50 Hz and step height of 10 mV complete a 1-V scan in 2 s, whereas the corresponding NPV experiment would take 1000 s if a 10-s waiting period was used between pulses. Measurement speed makes SWV a very promising electrochemical technique for applications including chromatography detectors and on-line monitors where real-time data acquisition may be necessary.[40,41]

The differential nature of SWV imparts enhanced S/N and sensitivity. SWV is more sensitive than DPV by as much as 24% under optimum conditions since it uses signals from both the forward and reverse reactions in the output, whereas DPV only uses the forward component. Noise is reduced because charging components in the forward and reverse half-cycles are nearly identical, and are zeroed in the calculation of i_{net}.

Figure 13 shows a calculated reversible SWV curve with forward, reverse and net currents. Here the current is given as ψ vs $n(E - E_{1/2})$, which are simply dimensionless variables which correspond to current and potential,

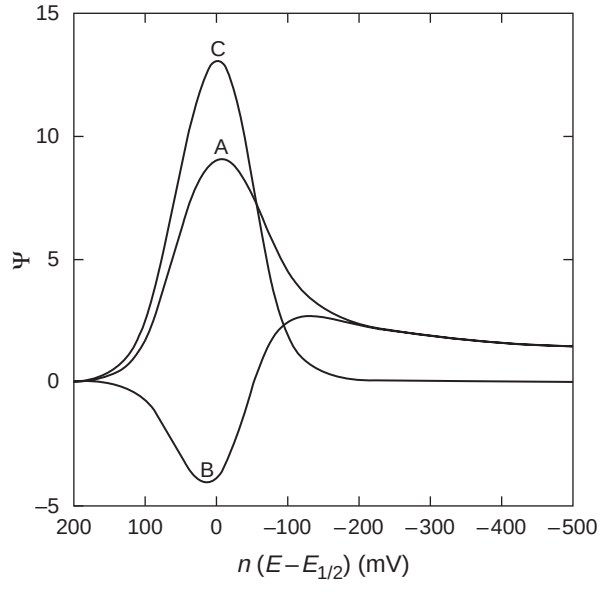

Figure 13 Calculated square wave voltammograms for (A) forward, (B) reverse and (C) net currents in dimensionless units. (Reproduced by permission of The American Chemical Society from J.J. O'Dea, J. Osteryoung, R.A. Osteryoung, 'Theory of Square-wave Voltammetry for Kinetic Systems', *Anal. Chem.*, **53**, 695–701 (1981).)

Table 1 Pulse voltammetry methods

Method	Acronym	Parameter	Designation	Typical ranges[a]
Normal pulse voltammetry	NPV	Pulse time	t_p	1–100 ms
		Waiting time	t_w	1–10 s
		Step height	ΔE_s	1–25 mV
Reverse pulse voltammetry	RPV[b]	Waiting time	t_w	As for NPV
		Pulse time	t_p	As for NPV
		Step height	ΔE_s	As for NPV
		Generation time	t_g	1–10 s
Differential pulse voltammetry	DPV	Pulse height	ΔE_p	1–10 mV
		Pulse width	t_p	As in NPV
		Step increment	ΔE_s	1–10 mV
Square wave voltammetry	SWV	Step height	ΔE_s	1–10 mV
		Square wave amplitude	ΔE_{sw}	1–100 mV
		Frequency	f	1–1000 Hz

[a] Ranges indicate order of magnitude for various parameters. Drop times for Hg electrodes have not been included in the table but are typically of the order of 1–5 s.
[b] RPV consists of two different methods, reverse pulse (RP) and reverse pulse with waiting (RPW).

respectively. The peak current of a reversible SWV curve is given by Equation (18):[29]

$$i_p = \frac{nFAD_O^{1/2}C_O}{\pi^{1/2}(t_p)^{1/2}}\psi_p \qquad (18)$$

where ψ_p is the dimensionless current function for SWV, which depends on nE_{SW} and $n\Delta E_s$. Values of ψ_p have been tabulated.[29] The peak height of the net voltammogram depends linearly on $f^{1/2}$ for reversible systems. The peak height of a DPV curve is approximately the same as that of the forward SWV component.

The width at half-height for a reversible system is $90.5/n$ mV, at 25 °C, for small values of E_{SW}. The height and width of the wave both depend on E_{SW}. For analytical applications, the peak width-to-height ratio is maximized using $nE_{SW} = 50$ mV.[27] Table 1 summarizes typical ranges of experimental parameters for commonly used pulse voltammetric methods.

2.3 Pulse Polarography and Voltammetry at Mercury Electrodes

Mercury electrodes have played an important role in the development of electroanalytical methods and are still used today.[42,43] The ease of oxidation of Hg in aqueous media limits its use at positive potentials. Nonetheless, Hg is an excellent material for the electrochemical detection of species with fairly negative redox potentials and for the detection of metal ions in particular. Hg also has an advantage over solid electrodes that the electrode surface can be renewed during an experiment by replacing an old drop with a new one. This is particularly useful in situations where the electrode surface may be degraded or fouled during the measurement. Recent advances in renewable mercury electrodes have been reviewed.[44]

2.3.1 Dropping Mercury Electrode

The DME was the original working electrode of polarography. A DME is a gravity-fed device in which Hg flows from a reservoir through a narrow-bore glass capillary having an inner diameter of the order of 0.1 mm. Hg drops form at the open end of the capillary, grow and finally fall from the capillary tip. Hg drops extruded from a capillary tip are spherical and their size generally falls in the macroelectrode range. The rate of drop formation and growth is dependent upon the height of the Hg reservoir above the tip and the diameter of the capillary. The ultimate size of the Hg drop formed depends on the surface tension between the drop and the surrounding aqueous medium into which it grows. A ''drop knocker'' is often used to dislodge mechanically the drop at a fixed point during growth, i.e. at constant surface area. Prior to the advent of pulse methods, continuous measurement of current was performed during the entire cycle of drop birth, growth and removal. Since the surface area of the electrode increased during the drop's lifetime, periodic current oscillations and low S/N were facts of life for polarographic methods. Developments in pulse voltammetry addressed this problem by timing potential application and current measurement to the cycle of the DME.[1,23] The basic strategy for pulse timing, e.g. in DPP, was to apply pulses and measure currents at the very end of a drop's life, so that surface area changes between measurements were minimized. This was possible using short pulse widths, 10–100 ms, which allowed the experiment to be performed under nearly constant electrode area conditions. Significant reduction in charging current was also realized.

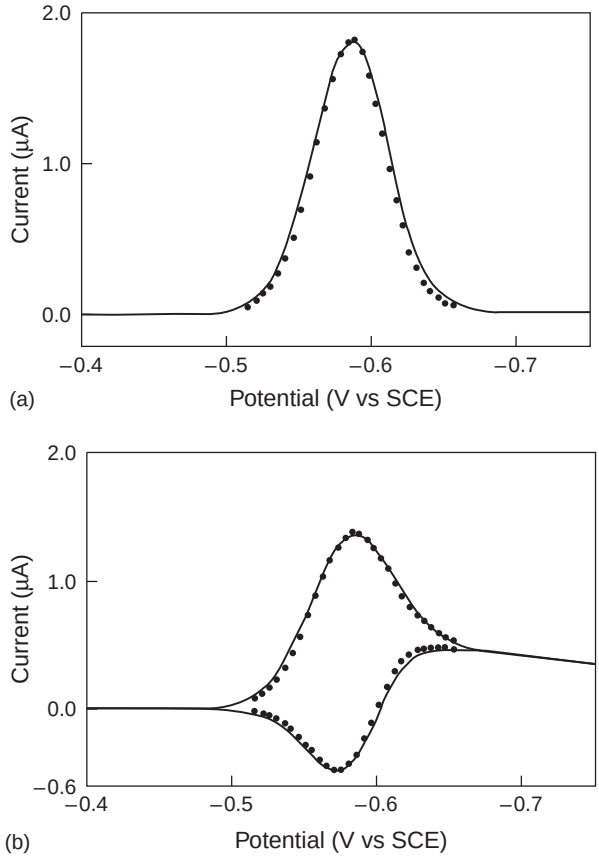

Figure 14 Experimental SWV of Cd^{2+}/Cd(Hg) in 0.1 M HCl for 8.43×10^{-5} M Cd^{2+} at a single drop (t_d, 2 s) of a DME. (a) Net current. (b) Forward and reverse currents. $\Delta E_s = 5$ mV; $E_{SW} = 30$ mV; $f = 30$ Hz. (Reproduced by permission of The American Chemical Society from J.A. Turner, J.H. Christie, M. Vukovic, R.A. Osteryoung, 'Square-wave Voltammetry at the Dropping Mercury Electrode: Experimental', *Anal. Chem.*, **49**, 1904–1908 (1977).)

An example of SWV at a DME[45] is given in Figure 14(a) and (b). Here, the redox reaction is (Equation 19):

$$Cd^{2+} + 2e^- \rightleftharpoons Cd(Hg) \qquad (19)$$

in which aqueous Cd^{2+} (CAS 10022-68-1) is reduced by $2e^-$ to Cd metal, which dissolves into the Hg electrode. The entire SWV scan was performed on a single Hg drop after it had grown for 2 s. Drop surface area growth was compensated by normalizing currents to $t_d^{2/3}$, where t_d is the drop time, since the drop surface area changes with $t^{2/3}$. Today, the DME is mostly of historical interest.

2.3.2 *Static Mercury Drop Electrode*

The SMDE[46] was developed by EG&G Princeton Applied Research (PAR) to address the problem of time-dependent electrode surface area and is still in widespread use today. This device can be used in either a repetitive mode analogous to the DME or as an HMDE where a single drop is used for an entire experiment. The HMDE is a spherical, stationary mercury electrode. The SMDE shares some characteristics of the DME in that it retains the Hg reservoir and uses a capillary tip to form the Hg drop.

When the SMDE is used in repetitive mode, Hg drops are extruded from the end of the large bore capillary by activation of a solenoid and plunger which prevents Hg flow when closed. When opened Hg flows, rapidly forming a drop. Deactivation of the solenoid closes the plunger and Hg flow ceases. The drop is grown to its final size and surface area very quickly compared with the DME. After a predetermined time, the drop is dislodged and the cycle repeated. Advantages realized by a renewable electrode are countered somewhat by accumulation of large amounts of waste mercury generated by the SMDE.

NPV and RPV of $Fe^{2+/3+}$ (CAS 10025-77-1) and Pb^{2+}/Pb(Hg) (CAS 10099-74-8)/(CAS 7439-92-1) at an SMDE are shown in Figure 15.[18] For $Fe^{2+/3+}$, RPV was performed starting at −0.5 V vs a saturated calomel electrode (SCE) to generate Fe^{2+}. Pb was generated at −0.8 V vs SCE. The SMDE drop grew for 1.5 s at the initial potential, 0 V ($Fe^{2+/3+}$) or −0.3 V (Pb^{2+}), prior to the RPV run. Figure 15 shows that spherical diffusion complications were apparent for RPV of Pb^{2+}/Pb(Hg), i.e. $i_{RP} + i_{DC} > i_{NP}$, when Pb dissolved in the Hg drop,

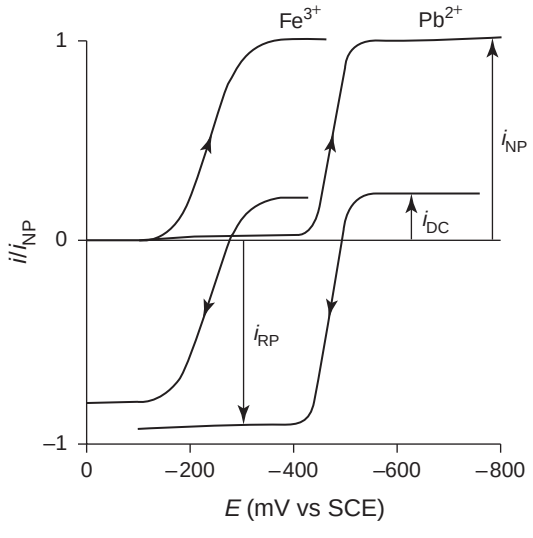

Figure 15 Normalized NPV and RPV of 0.639 mM $Fe^{3+/2+}$ in 0.1 M sodium oxalate–0.1 M $NaClO_4$ and 0.488 mM $Pb^{2+/0}$ in 0.5 M acetate buffer (pH 4.8) at an SMDE. (Reproduced by permission of The American Chemical Society from T.R. Brumleve, J. Osteryoung, 'Spherical Diffusion and Shielding Effects in Reverse Pulse Voltammetry', *J. Phys. Chem.*, **86**, 1794–1801 (1982).)

but were absent for $Fe^{2+/3+}$ where neither oxidation state was soluble in Hg.

2.3.3 Thin-film Mercury Electrode

TFMEs are fabricated by plating Hg on solid substrate electrodes such as carbon, platinum or gold. TFMEs allow the use of Hg in a convenient, planar format when the use of Hg drop electrodes is not practical. Such applications could include chromatographic flow cell detectors and monitors for flowing streams. The TFME is not renewable, but imparts the cathodic advantages of Hg to a detection scheme.

Pulse voltammetry has been carried out at TFMEs in DPV[47,48] and SWV modes.[49–52] TFMEs have been shown to improve the analytical detection limit, sensitivity and resolution compared with Hg drops, particularly in ASV discussed in section 4.2.1, by virtue of exhaustive removal of electrodeposited material, which often is not achieved with a mercury drop.

2.4 Pulse Voltammetry at Ultramicroelectrodes

Ultramicroelectrodes (UMEs) have proven useful in electroanalytical pulse voltammetry. Once at the cutting edge of electrochemical research, UMEs are now fairly standard devices. A variety of geometries have been characterized, including disks, rings, bands, cylinders and arrays of these.[53] UMEs have been prepared from solid electrode materials such as Pt, Au and C. Hemispherical Hg drop UMEs have also been fabricated. We shall discuss the characteristics of UMEs that make them useful in electroanalytical work and illustrate these points with examples of pulse voltammetric applications.

UMEs are small, with at least one microscopic dimension. An ultramicrodisk electrode, for example, may have a diameter of 20 μm or less. An ultramicroband electrode, on the other hand, could have a width of 20 μm and a length of 1–2 mm. Small physical size makes the UME obviously useful in sampling small volumes and restricted spaces. Since UMEs have a small surface area, the current flowing at them is also small. This characteristic tends to reduce the effects of uncompensated resistance (iR_u) in the electrochemical cell[5,54] and allows undistorted voltammetry in media or physical vicinities of high resistance, e.g. a nonaqueous solvent or a very thin flow channel. UMEs feature reduced double-layer capacitance compared with larger electrodes and have smaller RC time constants which allow faster electrochemical measurements (microseconds or less) to be performed compared with their macroscale brethren.

The nature of diffusion of electroactive species to a very small electrode is different from that of its conventional counterparts. Generally UMEs show enhanced mass transfer in the form of significant radial diffusion contributions in addition to planar diffusion. This feature actually simplifies and speeds NPV and RPV experiments by eliminating the necessity for convection to achieve boundary condition renewal, and enables true steady-state limiting currents to be obtained.

SWV,[55,56] NPV and RPV[57–68] have been studied extensively at microelectrodes in both their theoretical and experimental aspects. Hg-based microelectrodes have also been characterized.[69–71] Reduced iR_u at microelectrodes has allowed pulse voltammetry to be applied to resistive media[67,68] and media with high concentrations of electroactive species.[70] The small RC time constant of microelectrodes has enabled fast pulse voltammetry on the microsecond timescale.[62,64] Radial diffusion contributions allow rapid renewal of initial boundary conditions in the NPV/RPV experiment.[57,58,63]

Radial diffusion contributions to mass transfer at a microdisk electrode have important implications for the timescale on which NPV/RPV experiments can be performed. A stationary 5 μm radius microdisk electrode, for example, has been shown to have a mass transfer coefficient comparable to that which would be obtained by rotating a conventional size electrode at 10 000 rpm.[57] Simply stated, this means that it is only necessary to wait sufficiently long during t_w for re-establishment of initial conditions. If $t_p/t_w \leq 0.1$, initial boundary conditions can be restored without stirring for electrode diameters up to 25 μm.[57]

Figure 16(a–f) shows RPV curves taken in the RPW mode for oxidation of 5 mM $Fe(CN)_6^{4-}$ (CAS 14459-95-1) in 0.4 M $Sr(NO_3)_2$.[57] The curves show variation of t_g between 0.01 and 5.00 s for a constant t_w of 5 s. The magnitudes of i_{dc} and i_{RP} were within 15% of theory even though no stirring was used to renew the boundary conditions.

3 INSTRUMENTATION AND OPERATION

This section covers some basic considerations in carrying out pulse voltammetry experiments in the laboratory. The advent of widely available commercial benchtop instruments controlled by PCs has greatly simplified the implementation of pulse voltammetry in the laboratory. It has also virtually eliminated the need to construct one's own potentiostatic instrumentation. User friendliness has reached the stage where practically anyone can setup and use electrochemistry as a research tool. The reader will find much helpful information about the particulars of doing specific electrochemical experiments in the volumes by Heineman and Kissinger[2] and Sawyer and Roberts,[3] both of which have been updated in new editions within the past few years. With the

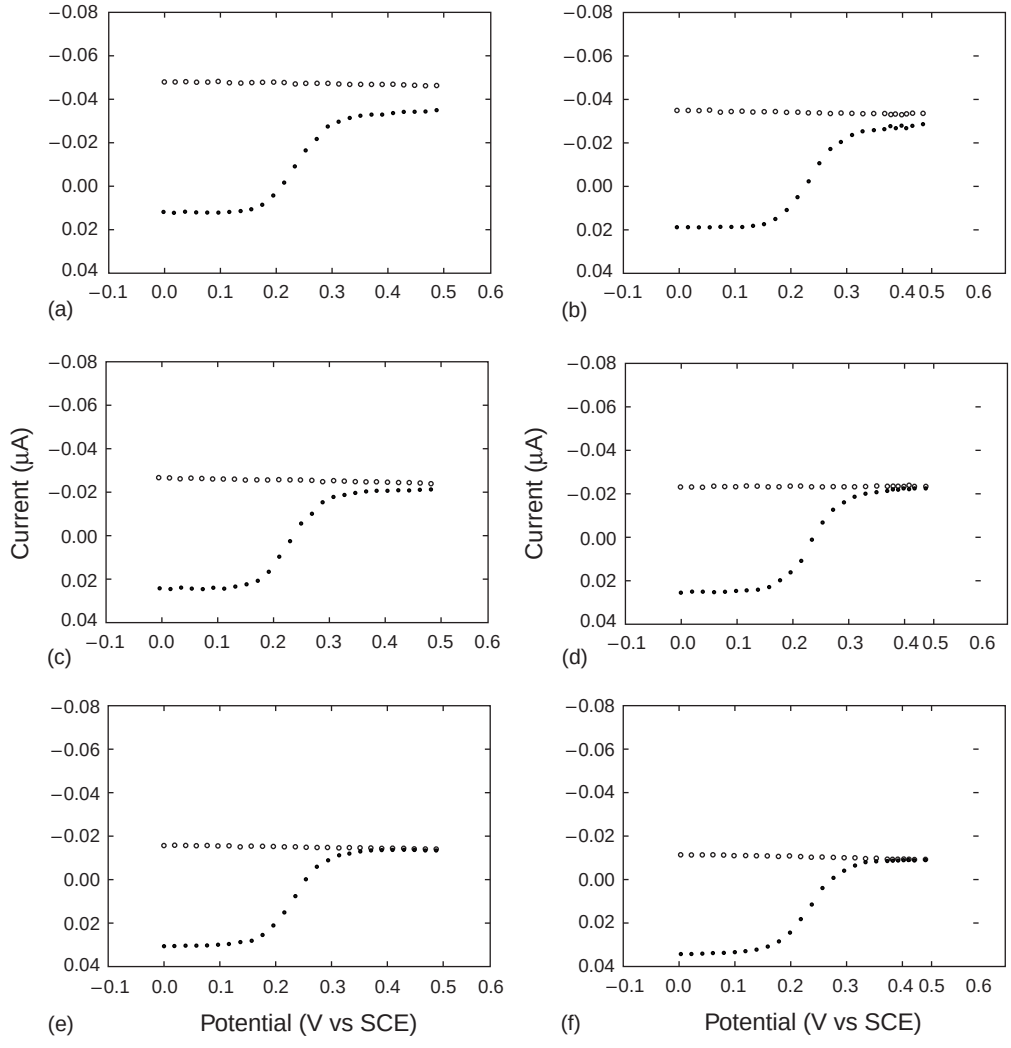

Figure 16 Reverse pulse voltammograms with waiting for 5 mM $Fe(CN)_6^{4-}$ in 0.4 M $Sr(NO_3)_2$ at a 12.98-μm radius Pt electrode with $t_w = 5$ s and $t_p = 0.1$ s; $t_g =$ (a) 0.01, (b) 0.02, (c) 0.03, (d) 0.10, (e) 1.00 and (f) 5 s. (Reproduced by permission of The American Chemical Society from L. Sinru, J. Osteryoung, J.J. O'Dea, R.A. Osteryoung, 'Normal and Reverse Pulse Voltammetry at Microdisk Electrodes', *Anal. Chem.*, **60**, 1135–1141 (1988).)

appearance of microprocessor-controlled potentiostats in the early 1980s, the potential of analytical voltammetry as a working tool has increased significantly. Modern Windows®-based graphic user interfaces have made many sophisticated electrochemical techniques available to a wider spectrum of users from diverse disciplines.

3.1 Commercial Instrumentation

We mainly discuss commercially available instrumentation from US manufacturers and among these only several representative examples. A recent review of electrochemical literature by Anderson et al.[72] lists many useful Web sites for electrochemical instrument vendors. The reader is directed to these sources for additional information.

PAR (*http://www.egginc.com*) has long been a leading manufacturer of electrochemical instruments. PAR offers a line of no less than nine different potentiostat models, the most prevalent of which is probably Model 273A, which is scheduled to be discontinued. Its replacement is Model 283. Most PAR potentiostats are computer controlled from a Windows® environment. PAR offers two highly sophisticated software packages, M250 and M270, which allow the user to employ a wide array of electrochemical methods and exercise advanced data acquisition and analysis including curve fitting for mechanism and kinetics. All of the popular pulse voltammetric methods discussed in this article are represented. Two units targeted for analytical applications are the Model 394 Electrochemical Trace Analysis System and the Model 264B Polarographic Analyzer, both of which

handle NPV and NPP. Model 394 also performs SWV, DPP and DPV.

Bioanalytical Systems (BAS) (*http://bioanalytical.com*) markets two multi-technique, research-grade potentiostats, the 100 B/W and the CV-50 W. BAS was the first company to attempt to make electrochemical methods more transparent and user friendly with introduction of the BAS 100 in the early 1980s. These units enable a wide range of methods to be performed, including NPV, NPP, DPV, DPP and SWV. Many automatic functions are included in these instruments, such as automatic curve analysis for peak potential and current calculation and quantitation from calibration curves. BAS also sells software for fitting of electrochemical data to mechanisms.

Cypress Systems (*http://www.cypresshome.com*) sells a variety of electrochemical instruments ranging from models suitable for an undergraduate laboratory to full research-grade instruments. The CS-1190 and CS-1090, for example, support SWV, DPV and DPP, but not NPV and NPP.

CH Instruments (*http://www.chinstruments.com*) is a relatively new company that offers a full line of electrochemical instruments in three different series, Models 600, 700 and 800. Within each series a huge array of choices may be made in terms of available techniques and other system capabilities, from units that are fairly limited for specific applications, to fully capable research-grade workstations. NPV, NPP, DPV, DPP and SWV are all supported. Many models incorporate extensive data analysis and curve fitting capabilities.

ECO CHEMIE (*http://www.ecochemie.nl*) is a very innovative and successful vendor of electrochemical instrumentation in Europe. Their Autolab instrument is particularly useful for implementation of modern pulse voltammetric methods.

3.2 General Considerations

Basic electrochemical experiments are actually fairly easy to setup once the electrochemical instrument is available. A real experiment could involve no more than a beaker or other suitable container, a few electrodes including a reference electrode for the three-electrode potentiostat and some solutions or samples. Pulse voltammetry will also require efficient stirring which can be actuated, e.g. by a transistor-transistor logic (TTL) pulse. SMDE units are also run via the potentiostat. The books by Heineman and Kissinger[2] and Sawyer and Roberts[3] cited above contain ample detail on specifics of electrochemical experiments for the interested reader.

Renewal of the initial boundary conditions is probably the most important factor in the successful implementation of analytical pulse voltammetry. If one uses an SMDE, this requirement is automatically satisfied and, as we saw above, microelectrodes also satisfy the condition easily, provided that appropriate time parameters are chosen. If solid electrodes of conventional size are employed, one must ensure that adequate stirring and time to achieve quiescence are incorporated into the pulse voltammetry protocol.

Selection of the appropriate pulse voltammetric method and experimental parameters depends on the task at hand. For a laboratory researcher wanting to evaluate $E^{\circ'}$ and diffusion coefficients for newly synthesized organometallic complexes, NPV will probably be the method of choice. The sigmoidal shape of the NPV waveform is naturally adapted to facile comparison of wave heights and half-wave potentials. For an environmental researcher wanting to determine what electroactive trace metals are present in river water samples, then a differential method such as DPV or SWV may be more appropriate because the peaked response of these techniques is more amenable to identification and quantification of components in multianalyte mixtures. Differential methods also have a certain tolerance for irreversible background electrochemical processes such as O_2 reduction. O_2 is present at ca. 25 ppm in air-equilibrated aqueous samples and is reducible at potentials where many common analytes are electroactive. In the laboratory, one may simply deoxygenate a solution with N_2 or Ar, but in field applications this practice should be avoided since it adds extra steps and cost to the analysis process. Differential methods are therefore the choice if one wishes to develop a fieldable unit.

Selection of the appropriate electrode for a particular application is often extremely important since electrode materials have widely disparate properties. Which electrode material will work acceptably will depend on the particular problem to be solved. Among the solid metals, Pt and Au have been popular. Pt is harder than Au and is therefore easier to polish to a mirror finish. Both have relatively low overpotentials for H^+ reduction to H_2, which means that their use in acidic media can be problematic. Some relief from proton reduction background can be gained by using the differential methods for the same reasons as for O_2 reduction. Pt and Au both form surface oxide films at positive potentials which may interfere with some analytical applications.

Carbon electrodes, including GC and highly oriented pyrolytic graphite (HOPG), display high overpotential for H^+ reduction and are therefore useful in acidic media. Carbon is more difficult to fabricate as a microelectrode owing to its fragility compared with common metals, but the extra effort is often worthwhile. Carbon electrodes are used extensively in aqueous and nonaqueous electrochemistry. One problem that can be encountered with carbon is the chemical nature of the electrode surface. Carbon surfaces generally are

covered with functional groups such as quinones, carboxylic acids and so forth, the presence of which can affect electron transfer reversibility for certain analytes. Electrochemical pretreatment of the electrode is often necessary to achieve repeatable results. These pretreatments change the chemical nature and distribution of functional groups on the carbon surface. The effect of pretreatment protocols is usually determined empirically for a particular problem.

Hg electrodes have a very large overpotential for H^+ reduction, which means that they are ideal for cathodic reduction processes in acidic media. However, Hg is easily oxidized, especially in the presence of complexing anions such as Cl^-. Therefore, one cannot very often study oxidative chemistry at an Hg electrode. In certain nonaqueous media, the anodic range of an Hg electrode can be extended to more positive values. A major drawback of Hg, aside from the negative environmental connotations involved in using this metal, is the large amount of waste Hg generated by devices such as the SMDE. One must store and either clean up or dispose of this Hg, which can be a nuisance if a large amount of activity is anticipated. Electrodes in conventional formats, such as the SMDE, are not useful for field work, but the TFME can be used in the field.

Selection of pulse voltammetric parameters, within reasonable limits, is not as critical to the success of a method as selection of appropriate technique and electrode provided that common sense is exercised. For example, if NPV of a dissolved species is being performed to determine its diffusion coefficient and the species is stable, one would not want to select $t_w = 500\,s$ to renew the initial conditions between pulses when $t_w = 5\,s$ will do. In SWV, $\Delta E_{sw} = 1\,V$ probably will not be useful since one will quickly exceed the available potential window of the aqueous medium. Values in range 25–100 mV are more useful. Common ranges of timing parameters for various pulse voltammetric methods are given in Table 1. Within these ranges, empirical adjustment of pulse amplitudes and timing can be performed to optimize the method for a particular problem, although the effects of the parameters on the shape of the resulting voltammogram are known quantitatively and exactly for the various pulse voltammetric methods, as noted in previous sections.

In general, one cannot expect more than a ca. $\pm 1\,V$ window for performing aqueous electrochemistry. The available potential window depends ultimately on the potential needed to oxidize or reduce the solvent, but other limiting processes, such as redox chemistry of dissolved electrolytes, usually occur prior to solvent discharge. Redox reactions occurring at potentials more extreme than the solvent limits are not generally accessible by electrochemical methods. Electrochemical processes which occur on the edge of the solvent limit, e.g. reduction of Zn(II) in aqueous media, do not typically provide reliable quantitative information unless careful background subtraction can be performed. Nonaqueous media have significantly extended potential ranges compared with water: as much as $\pm 2\,V$ or more.

4 APPLICATIONS AND EXAMPLES

In this section we present a variety of illustrative examples of the use of various pulse voltammetric methods for examination of problems in chemistry, analysis and monitoring. These problems range from the most basic to the most applied and hopefully will give the reader a flavor of the possibilities for using pulse voltammetry in their own work. In the course of reviewing the literature one of us (M.T.C.) found 1703 papers containing some use of pulse voltammetric methods published in the chemical literature between January 1991 and June 1998 using only a single bibliographic search source (Chemistry Citation Index, Institute for Scientific Information, Philadelphia, PA). This is an average of 227 papers per year! Examples presented here are therefore selective, but hopefully show a spectrum of uses for pulse voltammetry.

4.1 Electrochemical Studies

4.1.1 Mass Transport

A common theme in fundamental experimental studies in electrochemistry is comparison of the shape of a voltammogram with that predicted by theory. Knowledge of the shape of a voltammogram can provide significant information about the thermodynamics and rates of the electron transfer reactions, magnitudes of diffusion coefficients and coupled chemical reaction rates and equilibrium constants. The following examples show how pulse voltammetric methods have been applied to problems involving mass transport.

Steady-state pulse voltammetry at microelectrodes has been used to probe the properties of strong and weak acids, both in simple molecular form, such as phosphoric acid (H_3PO_4, CAS 7664-38-2), and in polyelectrolytes, e.g. poly(styrene sulfonate).[67,73–75] Steady-state current for proton reduction at an 11.25-μm radius platinum microelectrode was simply described by Equation (20):

$$i_{ss} = 4nFDCr \tag{20}$$

where i_{ss} is the steady-state limiting current, D is the diffusion coefficient of proton, C is the acid concentration and r is the microelectrode radius. In the case of a simple strong acid, e.g. $HClO_4$ (CAS 7601-90-3), the limiting current for proton reduction was controlled by the total

acid concentration and displayed variation with ionic strength consistent with the influence of ionic strength on the diffusion coefficient of proton.[72] In the case of strong acids, which are completely dissociated in water, the only reaction of interest was proton reduction (Equation 21):

$$H^+ + e^- \longrightarrow \tfrac{1}{2}H_2 \qquad (21)$$

However, in the case of a weak acid, such as acetic acid (CAS 64-19-7), the additional step shown in Equation (22):

$$CH_3CO_2H \rightleftharpoons CH_3CO_2^- + H^+ \qquad (22)$$

had to be taken into account because i_{ss} was controlled by the slower diffusion coefficient of undissociated acid rather than that of H^+. This is illustrated in Figure 17, where steady-state limiting currents are compared for the strong acid $HClO_4$ and weak acids acetic and ascorbic acid (CAS 50-81-7).[75]

4.1.2 Chemistry Coupled to Electron Transfer

Pulse voltammetry has proven useful for study of chemical reactions coupled to electron transfer.[34,38,39] Some of the common reaction schemes encountered in practice include slow heterogeneous kinetics, preceding chemical reaction (prior to electron transfer) and following chemical reaction. Chemical conversions are commonly denoted C and electron transfers E in the electrochemical literature. Therefore, an ECE mechanism, for example, involves an electron transfer followed by a chemical reaction whose product undergoes a second electron transfer.

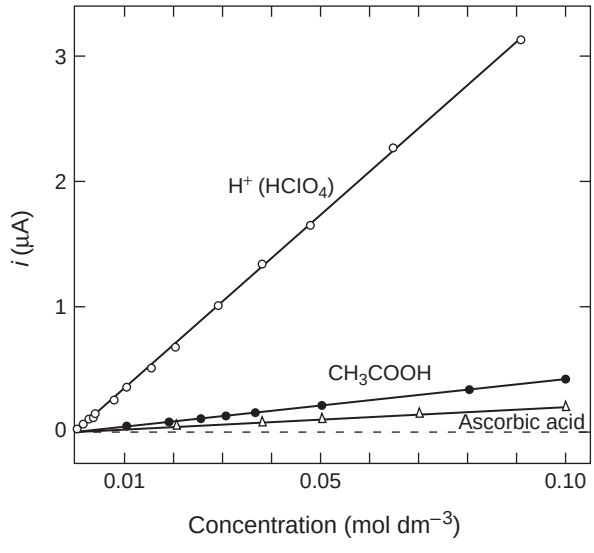

Figure 17 Mass transport-limited NPV currents for proton reduction in strong ($HClO_4$) and weak (acetic, ascorbic) acids as a function of concentration. (Reproduced by permission of The American Chemical Society from M. Ciszkowska, Z. Stojek, S.E. Morris, J.G. Osteryoung, 'Steady-state Voltammetry of Strong and Weak Acids With and Without Supporting Electrolyte', *Anal. Chem.*, **64**, 2372–2377 (1992).)

In slow electron transfer, the only new parameter to consider is the rate of the heterogeneous process, which is characterized by k^0, the rate of the reaction at zero overpotential and the transfer coefficient, α. While slow electron transfer is not coupled to chemistry per se, it is conveniently treated along with the other deviations from ideal reversible behavior.

Chemical reactions can either precede or follow an electron transfer. The voltammetric response will depend on whether the initial and final chemical forms are redox active and the rates and reversibility of the chemical steps will depend on the timescale of the pulse method. For example, in the scheme shown in Equations (23) and (24):

$$Y \longrightarrow O \qquad (23)$$

$$O + ne^- \longrightarrow R \qquad (24)$$

where Y is electrochemically silent (CE mechanism), one may expect that the current obtained for the reduction of O to R (Equation 23) will depend on the rate at which Y converts to O compared with the experimental timescale. Similarly, for an electron transfer (Equation 24), followed by (EC mechanism) Equation (25):

$$R \longrightarrow Y \qquad (25)$$

the current for oxidation of R back to O when viewed by RPV will depend on the rate of Equation (25) compared with the experimental timescale. More complicated situations, such as ECE (two electron transfers separated by a chemical conversion), disproportionation, etc., with varying degrees of reversibility are also possible and are accordingly complicated to analyze.

SWV responses have been defined for many of the basic mechanistic cases.[38] A well-known example of homogeneous chemistry coupled to redox reactions is electroreduction of anthracene (CAS 120-12-7) (An) in acetonitrile[65] (Equations 26–28):

$$An \rightleftharpoons An^{+\cdot} + e^- \qquad (26)$$

$$An^{+\cdot} + CH_3CN \longrightarrow An{-}CH_3CN^{+\cdot} \qquad (27)$$

$$An{-}CH_3CN^{+\cdot} \rightleftharpoons AnCH_3CN^+ + H^+ + e^- \qquad (28)$$

The radical cation generated in Equation (26) can react with nucleophilic CH_3CN to form an oxidizable intermediate. Disproportionation of the intermediates

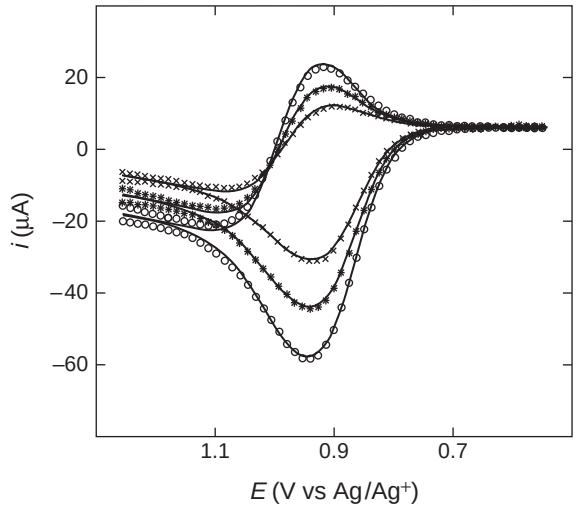

Figure 18 Forward and reverse square wave voltammograms with best-fit theoretical curves for an ECE mechanism. Square wave frequencies were (○) 3401, (∗) 2000 and (×) 1000 Hz, with $\Delta E_s = 10\,mV$ and $E_{SW} = 50\,mV$. (Reprinted from M.M. Murphy, Z. Stojek, J.J. O'Dea, J.G. Osteryoung, 'Pulse Voltammetry at Cylindrical Electrodes: Oxidation of Anthracene', *Electrochim. Acta*, **36**, 1475–1484 (1991), Copyright 1991, with permission from Elsevier Science.)

(Equation 29):

$$An^{+\cdot} + An{-}CH_3CN^{+\cdot} \rightleftharpoons An + AnCH_3CN^{+} \quad (29)$$

which reforms 1 mol of starting material and produces 1 mol of final product could affect the pulse voltammetric results if these conversions proceed significantly on the voltammetric timescale. Figure 18 shows, however, that the above reaction sequence was adequately explained by an ECE mechanism, i.e. the sequence of Equations (26–28), and that Equation (29) only had a small effect on the outcome.

An example of the application of RPV to the evaluation of unstable intermediates is provided by the electrochemical reduction of 1-iodoalkanes at Hg in nonaqueous solvents[68] (Equations 30 and 31):

$$RI + Hg + e^- \longrightarrow RHg^{\bullet}_{(ads)} + I^- \quad (30)$$

$$RI + 2e^- \longrightarrow R^- + I^- \quad (31)$$

Reduction of the iodoalkane (RI) at Hg by one electron generates an adsorbed radical (Equation 30), while two-electron reduction at more extreme potentials produces the anion, R^- and I^-. The radical can dimerize to form R_2 (Equation 32):

$$2RHg^{\bullet}_{(ads)} \longrightarrow HgR_2 \quad (32)$$

and the anion R^- can react with a proton donor impurity (e.g. water) to form the alkane or with the original iodoalkane to form dimer and iodide (Equations 33 and 34):

$$R^- + HD \longrightarrow RH + D^- \quad (33)$$

$$R^- + RI \longrightarrow R_2 + I^- \quad (34)$$

where HD is some proton donor. Figure 19 shows RPV of 2 mM 1-iododecane (CAS 2050-77-3) in pure propylene carbonate without supporting electrolyte at an Hg hemisphere microelectrode which was formed on a 12.5-μm radius Pt disk. Note how the use of a microelectrode in this highly resistive medium gives well-defined pulse voltammetric waves. This RPV experiment probed the reoxidation of products such as $RHg^{\bullet}$ and R^- generated at negative potentials. Two oxidation waves were found between ca. −0.5 and +1.0 V. The magnitude of the more negative wave decreased with increasing t_p, suggesting that the product formed during t_g at the negative limit was unstable on a timescale comparable to t_p. Furthermore, the ratio $i_{DC}/(i_{DC} - i_{RP})$ for the more negative wave increased with increasing RI concentration. It was suggested that this was caused by the effect of Equation (33), the rate of which would increase with increasing concentration of RI and subsequently attenuate i_{RP}. Additional information was obtained by examination of the effect of the negative potential limit at which the initial products were generated, shown in Figure 19. As this value was made less negative, the magnitude of both RPV waves decreased and the height of the second wave decreased at the expense of the first. This was consistent with increased rate of radical production by Equation (30) at less negative generation

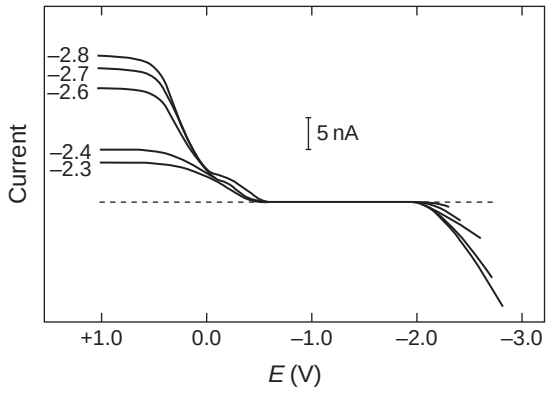

Figure 19 RPV of 2 mM 1-iododecane in propylene carbonate at various generation potentials, indicated at the left of each curve, at an Hg hemisphere electrode plated on to a 12.5-μm radius Pt disk. $t_p = 50\,ms$, $t_g = 1\,s$. (Reproduced by permission of The American Chemical Society from M. Ciszkowska, Z. Stojek, J. Osteryoung, 'Pulse Voltammetric Techniques at Microelectrodes in Pure Solvents', *Anal. Chem.*, **62**, 349–353 (1990).)

potentials and shows that the radical is actually fairly long lived in the pure solvent.

4.1.3 Corrosion

The study of corrosion processes is a broad field of electrochemical research and an extremely important one. Corrosion causes billions of dollars in losses per year in failure of equipment and civil infrastructure and is becoming an increasingly pressing problem for the military in the face of tight budgets for replacement of aging equipment such as military aircraft. Unfortunately, real corrosion phenomena are extremely complex and very difficult to simulate or reproduce in the laboratory.

Pulse voltammetry has been applied to the study of some extremely simple corrosion systems including anodic dissolution of copper (CAS 7440-50-8)[59,60] and silver (CAS 7440-22-4).[61] The virtue of the pulse voltammetric approach, in this case, is the ability to control precisely the potential applied to the corroding electrode and the timescale of potential application. Additionally, if metallic microelectrodes are used, as in the example below, a low current density allows high overpotentials to be investigated without significant iR_u distortion.

Anodizaton of a copper microelectrode (25-μm diameter, sealed in glass) was performed in 50% (v/v) ethylene glycol–water containing 1 M NaCl[60] (Equation 35):

$$Cu + 2Cl^- \rightleftharpoons CuCl_2^- + e^- \qquad (35)$$

A typical NPV trace for this system is shown in Figure 20. The potential was held between pulses at −0.4 V and was progressively stepped (t_p, 20 ms) to more positive values to drive Equation (35). The flat line in Figure 20 is the current just prior to pulse application. Constancy of this value is a good measure of the degree to which initial boundary conditions of the experiment are being renewed from pulse to pulse. The use of a microelectrode allowed the analysis of the entire anodic dissolution process. The slopes of semilogarithmic plots may be related to the stoichiometry and formation constants of the products formed.[60] At low current densities (i.e at the foot of the wave), Equation (35) was operative and generated a soluble product. At more extreme values, neutral CuCl was formed, which precipitated on the electrode.

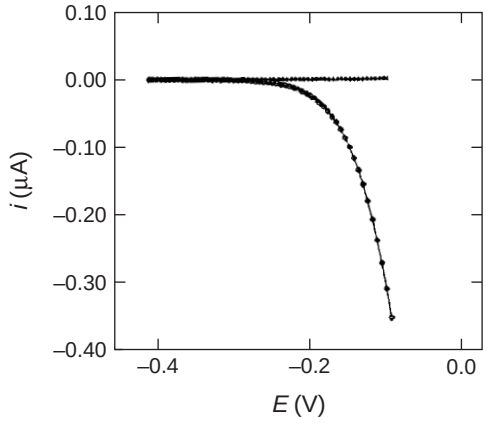

Figure 20 NPV for anodization of a 12.5-μm radius Cu electrode in 50% (v/v) ethylene glycol–water containing 1 M NaCl. t_p, 20 ms; t_w, 400 ms at −0.4 V. (Reprinted from K. Wikiel, J. Osteryoung, 'Pulse Voltammetric Techniques Applied to the Anodization of Copper Microelectrodes', *Electrochim. Acta*, **38**, 2291–2296 (1993), Copyright 1993, with permission from Elsevier Science.)

4.2 Analysis and Detection

Pulse voltammetric methods hold great promise for analytical applications in laboratory and field monitoring instruments. Of particular importance is the ability of the pulse techniques to discriminate against background processes to produce high S/N detection methods. Coupled to microelectrodes and other technologies, these methods are sometimes capable of the identification and quantification of multianalyte mixtures with minimum or no sample preparation. Under favorable circumstances, pulse voltammetry is capable of sensitivity and detection limits rivaling or surpassing those of other common analytical methodologies. A comparison of detection limits for some of these methods is presented in Table 2.

4.2.1 Pulse Voltammetry Coupled to Stripping Analysis for Metal Detection

A large fraction of papers published using pulse voltammetric detection methods deal with metal detection in water. This indicates not only the importance of environmental metal pollution, especially heavy metals such as Pb (CAS 7439-92-1) and Hg (CAS 7439-97-6), but also the ideal fit between these problems and electrochemical detection. A wide variety of soluble toxic metal ions are addressable by voltammetric methods,

Table 2 Comparison of detection limits for selected analytical methods

Technique	LLD (mol L^{-1})
DC voltammetry	10^{-6}
Visible spectrophotometry	10^{-6}
Atomic absorption spectroscopy	10^{-7}
Mass spectrometry	10^{-9}
Neutron activation analysis	10^{-10}
SWV	10^{-8}–10^{-9}
DPV	10^{-8}–10^{-9}
Stripping analysis with pulse voltammetry	10^{-9}–10^{-10}

including Pb(II) (CAS 10099-74-8), Hg(II) (CAS 7783-34-8), Cu(II) (CAS 10125-13-0), Zn(II) (CAS 10196-18-6), Cr(VI) (CAS 7440-47-3), Cd(II) (CAS 10022-68-1) and As(III) (CAS 7784-34-1).

Identification and quantification of these metals are possible with a sub-part per billion detection limit in some cases. A particularly noteworthy demonstration of the capability of pulse voltammetric detection was provided by Meyer et al.,[76] for the detection of Hg at the 5×10^{-14} M level, or ca. 0.01 ppt. This extremely low level of detection was achieved by ASV in conjunction with differential pulse anodic stripping voltammetry (DPASV) detection. The authors deposited Hg(II) from an acidic thiocyanate solution on to a GC working electrode. Unusually long deposition times, up to 40 min, were necessary to achieve the low detection limits reported. Nonetheless, this impressive demonstration of ultratrace metal detection shows how the very simple and relatively inexpensive electrochemical detection can rival or even surpass the performance of more complex and much more expensive instrumentation.

It is worthwhile digressing briefly to explain the ASV process[10] since it is commonly employed for electrochemical metal detection. ASV involves two steps. First, metal is plated by cathodic deposition on to the working electrode at a prescribed potential and for a fixed time. This step is a preconcentration in which metal ions are removed from a dilute solution and plated as elemental metal on to the working electrode. Since the working electrode is small compared with the volume of sample, the metal is concentrated by several orders of magnitude during this step. If a mercury electrode is used, the metal may dissolve in the electrode if it forms an amalgam. To a first approximation, the longer the deposition phase, the more metal is deposited at the electrode. Stirring often accompanies the preconcentration step to increase mass transfer of metal ion to the electrode surface. Following the preconcentration phase, metal is oxidatively removed from the electrode or ''stripped'' by application of a positive-going waveform. While the waveforms used to strip metal are identical with those described earlier, the difference between ASV and common pulse experiments is that the metal to be oxidized is located on or in the electrode, rather than in the solution at the electrode–solution interface. This has the important consequence that (ideally) the metal is quantitatively removed from the electrode during oxidative stripping, resulting in a symmetrical, peaked current–potential curve. This is true even if a nondifferential technique such as NPV is employed. True quantitative stripping of metal is usually achieved only with TFMEs (section 2.3.3). In the majority of recent applications either DPASV or square wave anodic stripping voltammetry (SWASV) waveforms are used, because of their favorable background discriminating abilities. The peak current or charge under the stripping peak can then be calibrated back to the original metal ion concentration in the sample. If the peaks are not overlapped on the potential axis, stripping voltammetry can be used to identify individual components of mixtures.[77,78] Detection limits in the single part per billion range are achievable, with care, in most cases.

CSV[11] is a related method which also has been used extensively in metal detection. This method is similar to ASV in its detection step, but metals (or organics) are adsorbed in their oxidized forms on an electrode surface, often with the help of a complexing agent. The adsorbed metal ion is then reductively stripped from the electrode in the detection step. CSV is typically employed for metal ions which are difficult to reduce directly to metal in aqueous media, including Cr(VI), Zn(II) and UO_2^{2+} (CAS 36478-76-9). CSV has also been employed widely in the analysis of organics, to be described in section 4.2.2.

Environmental trace metal analysis is becoming an increasingly important endeavor. Electrochemical methods provide a natural avenue to devices and instruments for field monitoring of metals. Stripping analysis by pulse voltammetry has played a large role in progress to move monitoring operations from the laboratory to the field.[79–85] Microelectrodes and microelectrode arrays prepared by photolithographic methods, coupled to DPASV and SWASV, offer the promise of fast, sensitive, selective and perhaps disposable devices compatible with field systems.[80,86–93]

Figure 21 illustrates the simultaneous detection of several metals in water using SWASV.[88] The working

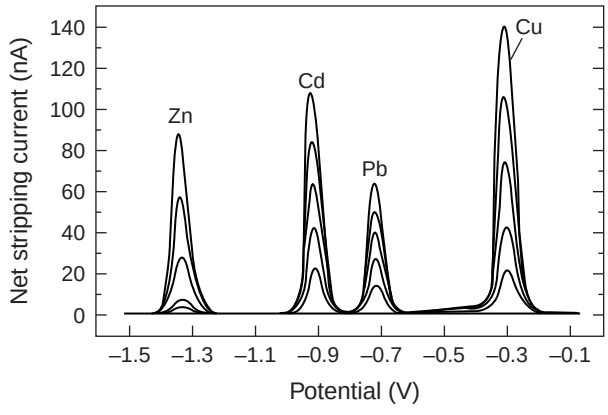

Figure 21 Simultaneous square wave anodic stripping voltammetric detection of Zn(II), Cd(II), Pb(II) and Cu(II) in 2 mM acetate buffer at 1, 2, 3, 4 and 5 ppb for each metal ion. $E_{SW} = 25$ mV; $\Delta E_s = 5$ mV; $f = 120$ Hz; deposition potential = −1.4 V for 260 s. (Reprinted from G.T.A. Kovacs, C.W. Storment, S.P. Kounaves, 'Microfabricated Heavy Metal Ion Sensor', *Sens. Actuators B*, **23**, 41–47 (1995), Copyright 1995, with permission from Elsevier Science.)

electrode in this case was an array of 19 Ir microdisks, each 10 μm in diameter, overcoated with electrodeposited Hg to form microhemispherical Hg electrodes. The metal ions Zn(II), Cd(II), Pb(II) and Cu(II) were codeposited on to the Hg–Ir array at −1.4 V vs Ag/AgCl for 4.33 min prior to the SWV scan. Figure 21 illustrates how multiple metals can be identified and quantified by SWASV. The positions of the SWASV peaks on the potential axis identify the metal and the charge under the stripping peak, or equivalently the magnitude of the peak, quantifies the amount of each metal stripped, which is calibrated to the original sample concentration by use of appropriate calibration curves or tables. The extremely high S/N available by SWASV is clearly shown in Figure 21.

Figure 22 shows the use of SWASV for the detection of Pb(II) and Cd(II) in raw river water.[92] The electrode was a 5 × 20 array of 12.2-μm diameter Hg hemispheres on 5-μm diameter Ir disks, and the detection method was SWASV. The concentrations shown in the figure amount to 60 ppt Pb(II) and 22.4 ppt Cd(II). This demonstrates again that it is possible to use SWASV to detect trace components with high S/N at levels competitive with conventional instruments which are not commonly amenable to portability and field analysis. Table 3 summarizes some representative detection strategies for toxic metal ions in water using pulse voltammetry.

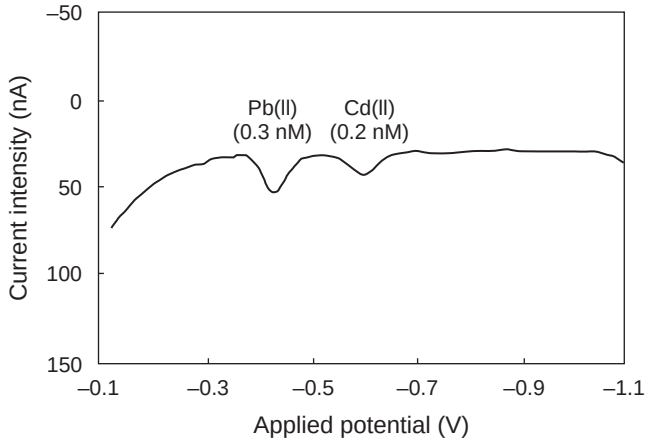

Figure 22 Square wave anodic stripping detection of Pb(II) and Cd(II) in raw river water at a gel-coated Hg-plated Ir microelectrode array composed of 12.2-μm diameter Hg hemispheres. Deposition potential, −1.1 V for 15 min; $E_{SW} = 25$ mV; $\Delta E_s = 8$ mV; $f = 50$ Hz. (Reproduced by permission of The American Chemical Society from C. Belmont-Hébert, M.L. Tercier, J. Buffle, G.C. Fiaccabrino, N.F. de Rooij, M. Koudelka-Hep, 'Gel-integrated Microelectrode Arrays for Direct Voltammetric Measurements of Heavy Metals in Natural Waters and Other Complex Media', *Anal. Chem.*, **70**, 2949–2956 (1998).)

4.2.2 Analysis of Organic Compounds

Pulse voltammetric detection has been applied successfully to a wide range of organic compounds of biomedical and environmental interest. Various forms of nonelectrolytic preconcentration[11] coupled to CSV or ASV are applied to adsorb the analyte on an electrode, followed by SWV, DPV or similar detection. In some cases, selection of an appropriate potential, analogous to the deposition phase of ASV, is all that is necessary to facilitate adsorption of analyte. Hg electrodes have proven useful in potential-dependent adsorption of organics for preconcentration prior to CSV. In some cases a reagent may be

Table 3 Stripping analysis of selected metals with pulse voltammetric detection

Metal ion	Method	Application/conditions	LLD (mol L^{-1})[a]	Working electrode	Ref.
Hg(II)	DPASV	Tap water acidified to pH 3.6	5×10^{-14}	Rotating GCE[b]	76
Hg(II)	SWASV	$HNO_3/NaNO_3$ media, pH 0–7	5×10^{-7}	Au microband/flow cell	93
Pb(II)	SWASV	Tap water Unmodified sample	5×10^{-9}	Hg UME array	90
Pb(II)	SWCSV	Seawater Acetate buffer, pH 5, xylenol orange	3×10^{-11}	HMDE	94
Pb(II)	DPASV	Acetate buffer, pH 4	2.4×10^{-8}	Hg UME array	89
Pb(II)	SWASV	0.1 M acetate buffer, pH 4	5×10^{-10}	Hg UME array	80
Cd(II)	DPASV	Acetate buffer, pH 4	$<1.8 \times 10^{-7}$	Hg UME array	89
Cd(II)	SWCSV	Seawater HEPES, pH 7.7, 8-hydroxyquinoline	4×10^{-10}	HMDE	95
As(III)	DPASV	1 M HCl	2.7×10^{-10}	Au disk	81
As(III)	SWCSV	Natural water Hydrazine sulfate, 2 M HCl	0.3×10^{-9}	HMDE	96
Cr(VI)	DPCSV	Groundwater Cupferron, PIPES buffer, pH 7	2×10^{-11}	HMDE	97

[a] ppb $= M \times 10^6 AW_{metal}$, where M is molarity (mol L^{-1}) and AW is atomic weight of metal (g mol^{-1}).
[b] GCE, glassy carbon electrode.

Table 4 AdSV of selected organics with pulse voltammetric detection

Compound	Method	Application/conditions	LLD (mol L^{-1})[a]	Working electrode	Ref.
Uric acid (CAS 69-93-2), dopamine (CAS 62-31-7)	SWASV	Biomolecules, citrate–phosphate buffers	2×10^{-7} uric acid 2.7×10^{-9} dopamine	Clay–Nafion modified carbon UME	96
Terbinafine (CAS 91161-71-6)	SWCSV	Antimycotic, pH 6	1.7×10^{-10}	HMDE	97
Sunset Yellow	DPCSV	Synthetic colorants, phosphate–citrate buffer, KCl, pH 5.7	0.03–0.2 μg mL^{-1}	HMDE	98
Ceftazidime	DPCSV	Pharmaceuticals, Britton–Robinson (BR) buffer, pH 4	5×10^{-9}	HMDE	99
Nitralin	SWCSV	Herbicides, pH >10	4.4×10^{-10}	HMDE	100
Doxazosin (CAS 77883-43-3)	DPCSV	Pharmaceuticals, BR buffer, pH 6.6	4×10^{-11}	Carbon paste	101
Atrazine	DPCSV	Herbicides, untreated groundwater	5 μg L^{-1}	SMDE	102
Metamitron	DPCSV	Herbicides, BR buffer, pH 1.9	$<4 \times 10^{-9}$	Carbon paste	103
Naringin (CAS 10236-47-2)	DPCSV	Flavonoid, grapefruit juice	5.5×10^{-8}	HMDE	104
Methylamphetamine (CAS 51-57-0)	DPCSV	Stimulant, ethanol–water–0.25 M ammonium acetate	0.125 μg L^{-1}	HMDE	105
Myoglobin (CAS 9008-45-1)	DPCSV	Biomolecules	0.1 μg mL^{-1}	Chemically modified Ag electrode	106
Ofloxacin (CAS 82419-36-1)	DPCSV	Pharmaceuticals, BR buffer, pH 4–10	3×10^{-7}	DME	107

[a] detection limits in mol L^{-1} unless indicated otherwise.

added to effect adsorption of analyte, or the electrode itself can be modified to favor analyte adsorption. Detection limits in the 10^{-7}–10^{-9} M range are achievable.

A recent review of adsorptive stripping voltammetry (AdSV) in the trace analysis of organic molecules and biomolecules tabulates detection limits and conditions for approximately 80 biologically active compounds.[98] Many of these methods use pulse voltammetric detection because of their superior S/N and background discrimination. Additional details on specific analytical methods can be found in the reviews and monographs by Wang.[11,99] Table 4 shows a range of representative examples of CSV organic detection by pulse voltammetry.[98–111]

4.2.3 Chromatographic and Flow Injection Analysis Detectors

Chromatography and flow injection analysis (FIA) represent mature practical application areas in which pulse voltammetry has made contributions. Electrochemical detection in high-performance liquid chromatography (HPLC), for example, has been carried out successfully for many years and commercial units are available. The typical arrangement uses constant-potential amperometry as the detection scheme, in which the detector electrode is held at a potential where oxidizable or reducible components of a mixture are detected following separation. In principle, very rapid scan pulse methods, such as SWV, should provide additional potential resolution within the timescale of a chromatographic peak, although these methods have not been implemented in commercial devices.[112]

FIA is another relatively mature process analytical methodology in which samples are taken from a pot or pipe, injected into a flowing electrolyte and sent to a detector.[113] Pulse voltammetric applications have been reported for ASV of metals[114–117] and organics.[118,119] A typical experimental set-up for FIA with pulse electrochemical detection is shown in Figure 23.[116]

5 FUTURE PROSPECTS

Pulse voltammetry has been demonstrated to be useful and versatile in a variety of basic problems and applications. Further advances in instrumentation and

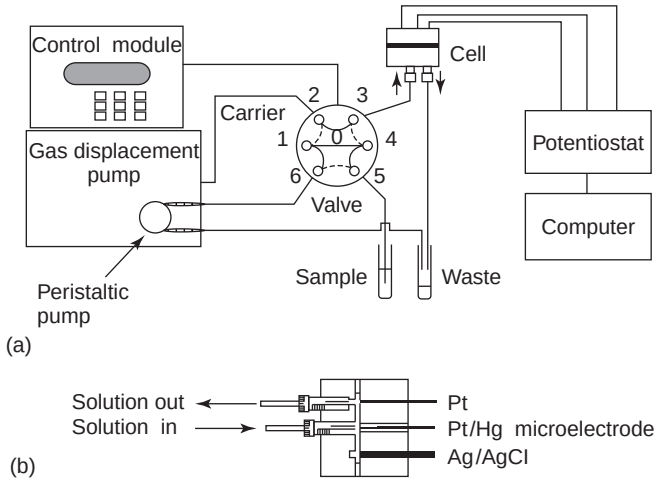

Figure 23 Schematic diagram of (a) a flow injection system and (b) a microelectrode flow cell for pulse voltammetric electrochemical detection in FIA. (Reproduced by permission of The American Chemical Society from F. Zhu, J.W. Ironstone, M.R. Ruegnitz, 'High Throughput Fast-scan Anodic Stripping Voltammetry in a Microflow System', *Anal. Chem.*, **69**, 728–733 (1997).)

techniques will no doubt occur in the future, in addition to many more convincing demonstrations of the utility of these methods for solving chemical problems. Mating sensitive pulse detection methods with other advanced technologies will, hopefully, speed the use of these methods in solving practical problems. Exciting recent uses and possibilities of pulse voltammetric detection in environmental analysis are capitalizing on the microfabrication methods of the semiconductor industry, for example, to produce arrays of electrodes which can do what single electrodes cannot. There is clearly sufficient interest and activity to generate many new innovations and applications. The key for future progress and wide use of pulse voltammetry will be to apply the techniques, which are now relatively mature, to new problems and new disciplines, particularly the difficult ones, i.e. real samples and complex matrices in chemistry, biology and the environment.

ACKNOWLEDGMENTS

Work performed in the authors' laboratories was supported by the US Department of Energy (M.T.C.) and by the Air Force Office of Scientific Research (R.A.O.).

ABBREVIATIONS AND ACRONYMS

AdSV	Adsorptive Stripping Voltammetry
ASV	Anodic Stripping Voltammetry
BAS	Bioanalytical Systems
CSV	Cathodic Stripping Voltammetry
CV	Cyclic Voltammetry
DC	Direct Current
DME	Dropping Mercury Electrode
DPASV	Differential Pulse Anodic Stripping Voltammetry
DPP	Differential Pulse Polarography
DPV	Differential Pulse Voltammetry
FIA	Flow Injection Analysis
GC	Glassy Carbon
GCE	Glassy Carbon Electrode
HMDE	Hanging Mercury Drop Electrode
HOPG	Highly Oriented Pyrolytic Graphite
HPLC	High-performance Liquid Chromatography
LLD	Lower Limit of Detection
NPP	Normal Pulse Polarography
NPV	Normal Pulse Voltammetry
PAR	EG&G Princeton Applied Research
RPP	Reverse Pulse Polarography
RPV	Reverse Pulse Voltammetry
RPW	Reverse Pulse Voltammetry with Waiting
SCE	Saturated Calomel Electrode
SMDE	Static Mercury Drop Electrode
S/N	Signal-to-noise Ratio
SWASV	Square Wave Anodic Stripping Voltammetry
SWV	Square Wave Voltammetry
TFME	Thin-film Mercury Electrode
TTL	Transistor-transistor Logic
UME	Ultramicroelectrode

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Voltammetry In Vivo for Chemical Analysis of the Living Brain • Voltammetry In Vivo for Chemical Analysis of the Nervous System

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis

Environment: Water and Waste **(Volume 3)**
Heavy Metals Analysis in Seawater and Brines • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Electrochemical Methods

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Organic Electrochemical Mechanisms

Electronic Absorption and Luminescence **(Volume 12)**
Indirect Detection Methods in Capillary Electrophoresis

REFERENCES

1. A.J. Bard, L.R. Faulkner, 'Controlled Potential Microelectrode Techniques – Potential Step Methods', *Electrochemical Methods*, John Wiley and Sons, New York, Chapter 5, 136–212, 1980.
2. W.R. Heineman, P.T. Kissinger, 'Large Amplitude Controlled-potential Techniques', in *Laboratory Techniques in Electroanalytical Chemistry* 2nd edition, eds. P.T. Kissinger, W.R. Heineman, Marcel Dekker, New York, Chapter 3, 51–125, 1996.
3. D.T. Sawyer, J.L. Roberts, 'Controlled Potential Methods', *Experimental Electrochemistry for Chemists*, Wiley-Interscience, New York, Chapter 7, 329–395, 1974.
4. R.N. Adams, 'Mass Transfer to Stationary Electrodes in Quiet Solution', *Electrochemistry at Solid Electrodes*, Marcel Dekker, New York, Chapter 3, 43–66, 1969.
5. A.J. Bard, L.R. Faulkner, 'Introduction and Overview of Electrode Processes', *Electrochemical Methods*, John Wiley and Sons, New York, Chapter 1, 1–43, 1980.
6. A.J. Bard, L.R. Faulkner, 'Kinetics of Electrode Reactions', *Electrochemical Methods*, John Wiley and Sons, New York, Chapter 3, 86–118, 1980.
7. A.J. Bard, L.R. Faulkner, 'Double Layer Structure and Adsorbed Intermediates', *Electrochemical Methods*, John Wiley and Sons, New York, Chapter 12, 488–552, 1980.
8. E. Laviron, 'Voltammetric Methods for the Study of Absorbed Reactants', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 53–157, Vol. 12, 1982.
9. R.W. Murray (ed.), 'Molecular Design of Electrode Surfaces', *Techniques of Chemistry*, John Wiley and Sons, New York, Chapter 1, 1–48, Vol. 22, 1992.
10. J. Wang, *Stripping Analysis*, VCH, Deerfield Beach, FL, 1985.
11. J. Wang, 'Voltammetry Following Nonelectrolytic Preconcentration', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–88, Vol. 16, 1989.
12. J. Osteryoung, 'Pulse Voltammetry', *J. Chem. Educ.*, **60**, 296–298 (1983).
13. J. Osteryoung, R.A. Osteryoung, 'Analytical Pulse Voltammetry', in *Electrochemistry, Sensors and Analysis*, eds. M.R. Smyth, J.G. Vos, Elsevier, Amsterdam, 3–12, 1986.
14. J. Osteryoung, M. Schreiner, 'Recent Advances in Pulse Voltammetry', *CRC Crit. Rev. Anal. Chem.*, **19**, S1–S27 (1988).
15. R.A. Osteryoung, J.G. Osteryoung, 'Pulse Voltammetric Methods of Analysis', *Philos. Trans. R. Soc. London, Ser. A*, **302**, 315–326 (1981).
16. M.T. Carter, R.A. Osteryoung, 'Heterogeneous and Homogeneous Electron Transfer Reactions of Tetrathiafulvalene in Ambient Temperature Chloroaluminate Molten Salts', *J. Electrochem. Soc.*, **141**, 1713–1720 (1994).
17. J. Osteryoung, E. Kirowa-Eisner, 'Reverse Pulse Polarography', *Anal. Chem.*, **52**, 62–66 (1980).
18. T.R. Brumleve, J. Osteryoung, 'Spherical Diffusion and Shielding Effects in Reverse Pulse Voltammetry', *J. Phys. Chem.*, **86**, 1794–1801 (1982).
19. M. Wojciechowski, J. Osteryoung, 'Electrochemistry and Reverse Pulse Polarographic Determination of 1,2-Dibromo-2,4-dicyanobutane', *Anal. Chem.*, **57**, 927–933 (1985).
20. A. Brestovisky, E. Kirowa-Eisner, J. Osteryoung, 'Direct and Titrimetric Determination of Hydrogen Peroxide by Reverse Pulse Polarography', *Anal. Chem.*, **55**, 2063–2066 (1983).
21. C.-L. Ni, F.C. Anson, J.G. Osteryoung, 'A Pulse Polarographic Measurement of the Kinetics of Dissociation of the $[Ru^{III}(EDTA)]_2$ Dimer', *J. Electroanal. Chem.*, **202**, 101–109 (1986).
22. J. Osteryoung, D. Talmore, J. Hermolin, E. Kirowa-Eisner, 'Reverse Pulse Voltammetry. Application to Second Order Following Reactions', *J. Phys. Chem.*, **85**, 285–289 (1981).
23. E.P. Parry, R.A. Osteryoung, 'Evaluation of Analytical Pulse Polarography', *Anal. Chem.*, **37**, 1634–1637 (1965).
24. J.H. Christie, J. Osteryoung, R.A. Osteryoung, 'Instrumental Artifacts in Differential Pulse Polarography', *Anal. Chem.*, **45**, 210–215 (1973).
25. J.W. Dillard, J.A. Turner, R.A. Osteryoung, 'Digital Simulation of Differential Pulse Polarography with Incremental Time Change', *Anal. Chem.*, **49**, 1246–1250 (1977).
26. K. Aoki, J. Osteryoung, 'Theory of Differential Pulse Polarography at Expanding or Stationary Planar Electrodes for Quasi-reversible or Totally Irreversible Reactions', *J. Electroanal. Chem.*, **110**, 19–36 (1980).
27. D.J. Myers, J. Osteryoung, 'Amperometric Titrations Employing Differential Pulse Polarography', *Anal. Chem.*, **46**, 356–359 (1974).
28. D.J. Myers, R.A. Osteryoung, J. Osteryoung, 'Pulse Voltammetry at Rotated Electrodes', *Anal. Chem.*, **46**, 2089–2092 (1974).
29. J. Osteryoung, J.J. O'Dea, 'Square Wave Voltammetry', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 209–308, Vol. 14, 1986.
30. J. Osteryoung, R.A. Osteryoung, 'Square Wave Voltammetry', *Anal. Chem.*, **57**, 101A – 110A (1985).

31. E. Zachowski, M. Wojciechowski, J. Osteryoung, 'The Analytical Application of Square Wave Voltammetry', *Anal. Chim. Acta*, **183**, 47–57 (1986).
32. K. Aoki, K. Tokuda, H. Matsuda, J. Osteryoung, 'Reversible Square Wave Voltammograms: Independence of Electrode Geometry', *J. Electroanal. Chem.*, **207**, 25–39 (1986).
33. W. Go, J.J. O'Dea, J. Osteryoung, 'Square Wave Voltammetry for the Determination of Kinetic Parameters. The Reduction of Zinc(II) at Mercury Electrodes', *J. Electroanal. Chem.*, **255**, 21–44 (1988).
34. J.J. O'Dea, K. Wikiel, J. Osteryoung, 'Square Wave Voltammetry for ECE Mechanisms', *J. Phys. Chem.*, **94**, 3628–3636 (1990).
35. M. Wojciechowski, W. Go, J. Osteryoung, 'Square Wave Anodic Stripping Analysis in the Presence of Dissolved Oxygen', *Anal. Chem.*, **57**, 155–158 (1985).
36. K. Aoki, K. Maeda, J. Osteryoung, 'Characterization of Nernstian Square-wave Voltammograms', *J. Electroanal. Chem.*, **272**, 17–28 (1989).
37. M.J. Nuwer, J.J. O'Dea, J. Osteryoung, 'Analytical and Kinetic Investigations of Totally Irreversible Electron Transfer Reactions by Square-wave Voltammetry', *Anal. Chim. Acta*, **261**, 13–25 (1991).
38. J.J. O'Dea, J. Osteryoung, R.A. Osteryoung, 'Theory of Square-wave Voltammetry for Kinetic Systems', *Anal. Chem.*, **53**, 695–701 (1981).
39. J. Zeng, R.A. Osteryoung, 'Square-wave Voltammetry for a Pseudo-first-order Catalytic Process', *Anal. Chem.*, **58**, 2766–2771 (1986).
40. R. Samuelsson, J.J. O'Dea, J. Osteryoung, 'Rapid Scan Square Wave Voltammetric Detector for HPLC', *Anal. Chem.*, **52**, 2215–2216 (1980).
41. L. Mahoney, J. O'Dea, J. Osteryoung, 'Development and Characterization of an Automated Flow System for Voltammetric Analysis', *Anal. Chim. Acta*, **281**, 25–33 (1993).
42. Z. Galus, 'Mercury Electrodes', in *Laboratory Techniques in Analytical Chemistry*, 2nd edition, eds. W.R. Heineman, P.T. Kissinger, Marcel Dekker, New York, Chapter 9, 267–288, 1996.
43. A. Bond, *Modern Polarographic Methods of Analysis*, Marcel Dekker, New York, 1980.
44. L. Novotny, M. Heyrovsky, 'Renewable Mercury Electrodes – Versatile Research Tools in General Chemistry', *Croat. Chem. Acta*, **70**, 151–165 (1997).
45. J.A. Turner, J.H. Christie, M. Vukovic, R.A. Osteryoung, 'Square-wave Voltammetry at the Dropping Mercury Electrode: Experimental', *Anal. Chem.*, **49**, 1904–1908 (1977).
46. W.M. Peterson, 'Static Mercury Drop Electrode', *Am. Lab.*, **11**(12), 69–70 (1979).
47. J.A. Turner, U. Eisner, R.A. Osteryoung, 'Pulsed Voltammetric Stripping at the Thin-film Mercury Electrode', *Anal. Chim. Acta*, **90**, 25–34 (1977).
48. T.R. Copeland, J.H. Christie, R.A. Osteryoung, R.K. Skoerboe, 'Analytical Applications of Pulsed Voltammetric Stripping at Thin Mercury Electrodes', *Anal. Chem.*, **45**, 2171–2174 (1973).
49. K. Wikiel, J. Osteryoung, 'Square Wave Voltammetry at a Mercury Film Electrode: Experimental Results', *Anal. Chem.*, **61**, 2086–2092 (1989).
50. C. Wechter, J. Osteryoung, 'Square Wave and Linear Scan Anodic Stripping Voltammetry at Iridium Based Mercury Film Electrodes', *Anal. Chem.*, **61**, 2092–2097 (1989).
51. S. Kounaves, J.J. O'Dea, P. Chandrasekhar, J. Osteryoung, 'Square-wave Voltammetry at the Mercury Film Electrode: Theoretical Treatment', *Anal. Chem.*, **58**, 3199–3202 (1986).
52. S. Kounaves, J.J. O'Dea, P. Chandrasekhar, J. Osteryoung, 'Square Wave Anodic Stripping Voltammetry at the Mercury Film Electrode: Theoretical Treatment', *Anal. Chem.*, **59**, 386–389 (1987).
53. R.M. Wightman, D.O. Wipf, 'Voltammetry at Ultramicroelectrodes', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 267–353, Vol. 15, 1989.
54. D.K. Roe, 'Overcoming Solution Resistance with Stability and Grace in Potentiostatic Circuits', in *Laboratory Techniques in Analytical Chemistry*, 2nd edition, eds. W.R. Heineman, P.T. Kissinger, Marcel Dekker, New York, Chapter 7, 193–234, 1996.
55. D. Whelan, J.J. O'Dea, J. Osteryoung, K. Aoki, 'Square Wave Voltammetry at Small Disk Electrodes: Theory and Experiment', *J. Electroanal. Chem.*, **202**, 23–36 (1986).
56. J. O'Dea, M. Wojciechowski, J. Osteryoung, K. Aoki, 'Square Wave Voltammetry at Electrodes Having Small Dimensions', *Anal. Chem.*, **57**, 954–955 (1985).
57. L. Sinru, J. Osteryoung, J.J. O'Dea, R.A. Osteryoung, 'Normal and Reverse Pulse Voltammetry at Microdisk Electrodes', *Anal. Chem.*, **60**, 1135–1141 (1988).
58. L. Sinru, R.A. Osteryoung, 'Normal and Reverse Pulse Voltammetry for Poised Systems at Microdisk Electrodes', *Anal. Chem.*, **60**, 1845–1850 (1988).
59. K. Wikiel, M.M. dos Santos, J. Osteryoung, 'Determination of Stability Constants by Using Normal Pulse Voltammetry at Microelectrodes', *Electrochim. Acta*, **38**, 1555–1558 (1993).
60. K. Wikiel, J. Osteryoung, 'Pulse Voltammetric Techniques Applied to the Anodization of Copper Microelectrodes', *Electrochim. Acta*, **38**, 2291–2296 (1993).
61. M. Donten, J. Osteryoung, 'Pulse Techniques in Studies of Metal Dissolution: Anodization of Silver', *J. Electrochem. Soc.*, **13**, 82–88 (1991).
62. Z.J. Karpinski, R.A. Osteryoung, 'Short Time Pulse Voltammetric Studies of Fast Heterogeneous Electron Transfer Reactions', *J. Electroanal. Chem.*, **349**, 285–297 (1993).

63. Z.J. Karpinski, R.A. Osteryoung, 'Renewal of Boundary Conditions in Pulse Voltammetry at Microdisk Electrodes for Non-reversible Systems', *J. Electroanal. Chem.*, **307**, 47–62 (1991).
64. M.A.M. Nöel, J.J. O'Dea, R.A. Osteryoung, 'Short Time Pulse Voltammetry in Ambient Temperature Chloroaluminate Ionic Liquids', *J. Electrochem. Soc.*, **139**, 1231–1236 (1992).
65. M.M. Murphy, Z. Stojek, J.J. O'Dea, J.G. Osteryoung, 'Pulse Voltammetry at Cylindrical Electrodes: Oxidation of Anthracene', *Electrochim. Acta*, **36**, 1475–1484 (1991).
66. M.M. Murphy, J.J. O'Dea, J. Osteryoung, 'Pulse Voltammetry at Microcylinder Electrodes', *Anal. Chem.*, **63**, 2743–2750 (1991).
67. S.E. Morris, M. Ciszkowska, J.G. Osteryoung, 'Steady-state Voltammetry of Simple and Polyelectrolyte Strong Acids With and Without Supporting Electrolyte', *J. Phys. Chem.*, **97**, 19 453–19 457 (1993).
68. M. Ciszkowska, Z. Stojek, J. Osteryoung, 'Pulse Voltammetric Techniques at Microelectrodes in Pure Solvents', *Anal. Chem.*, **62**, 349–353 (1990).
69. J. Golas, Z. Galus, J. Osteryoung, 'Iridium-based Small Mercury Electrodes', *Anal. Chem.*, **59**, 389–392 (1987).
70. J. Golas, Z. Kowalski, 'Applications of Small Iridium-based Mercury Electrodes to High Concentrations of Depolarizers', *Anal. Chim. Acta*, **221**, 305–318 (1989).
71. C. Wechter, J. Osteryoung, 'Voltammetric Characterization of Small Platinum–Iridium-based Mercury Film Electrodes', *Anal. Chim. Acta*, **234**, 275–284 (1990).
72. J.L. Anderson, L.A. Coury, J. Leddy, 'Dynamic Electrochemistry: Methodology and Application', *Anal. Chem.*, **70**, 519R–589R (1998).
73. M. Ciszkowska, A. Jaworski, J.G. Osteryoung, 'Voltammetric Reduction of Hydrogen Ion in Solutions of Polyprotic Strong Acids With and Without Supporting Electrolyte', *J. Electroanal. Chem.*, **423**, 95–101 (1997).
74. M. Ciszkowska, Z. Stojek, J.G. Osteryoung, 'Voltammetric Reduction of Polyprotic Acids at the Platinum Microelectrode: Dependence on Supporting Electrolyte', *J. Electroanal. Chem.*, **398**, 49–56 (1995).
75. M. Ciszkowska, Z. Stojek, S.E. Morris, J.G. Osteryoung, 'Steady-state Voltammetry of Strong and Weak Acids With and Without Supporting Electrolyte', *Anal. Chem.*, **64**, 2372–2377 (1992).
76. S. Meyer, F. Scholz, R. Trittler, 'Determination of Inorganic Ionic Mercury Down to 5×10^{-14} mol L^{-1} by Differential Pulse Anodic Stripping Voltammetry', *Fresenius' J. Anal. Chem.*, **356**, 247–252 (1996).
77. W. Jin, V.D. Nguyen, P. Valenta, H.W. Nürnberg, 'Simultaneous Determination of Seven Toxic Trace and/or Ultratrace Metals in Environmental Plants by Differential Pulse Voltammetry Without Change of Solution and Electrode', *Anal. Lett.*, **30**, 1235–1254 (1997).
78. C. Locatelli, 'Anodic and Cathodic Stripping Voltammetry in the Simultaneous Determination of Toxic Metals in Environmental Samples', *Electroanalysis*, **9**, 1014–1017 (1997).
79. J. Wang, B. Greene, 'Characteristics of a Flow Cell for the Determination of Arsenic(III) by Stripping Voltammetry', *J. Electroanal. Chem.*, **154**, 261–268 (1983).
80. J. Wang, J. Lu, B. Tian, C. Yarnitzky, 'Screen-printed Ultramicroelectrode Arrays for On-site Stripping Measurements of Trace Metals', *J. Electroanal. Chem.*, **361**, 77–83 (1993).
81. G. Forsberg, J.W. O'Laughlin, R.G. Megargle, S.R. Koirtyohann, 'Determination of Arsenic by Anodic Stripping Voltammetry and Differential Pulse Anodic Stripping Voltammetry', *Anal. Chem.*, **47**, 1586–1591 (1979).
82. A.M. Bond, W.A. Czerwinski, M. Llorente, 'Comparison of Direct Current, Derivative Direct Current, Pulse and Square Wave Voltammetry at Single Disc, Assembly and Composite Carbon Electrodes: Stripping Voltammetry at Thin Film Mercury Microelectrodes with Field-based Instrumentation', *Analyst*, **123**, 1333–1337 (1998).
83. R. Pongartz, K.G. Herrmann, 'Determination of Monomethylcadmium in the Environment by Differential Pulse Anodic Stripping Voltammetry', *Anal. Chem.*, **68**, 1262–1266 (1996).
84. D. Sancho, M. Vega, L. Debán, R. Pardo, G. González, 'Determination of Copper and Arsenic in Refined Beet Sugar by Stripping Voltammetry Without Sample Pretreatment', *Analyst*, **123**, 743–747 (1998).
85. M.-L. Tercier, N. Parthamaraty, J. Buffle, 'Reproducible, Reliable and Rugged Hg-plated Ir-based Microelectrode for In Situ Measurements in Natural Waters', *Electroanalysis*, **7**, 55–63 (1995).
86. S.P. Kounaves, W. Deng, P.R. Hallock, G.T.A. Kovacs, C.W. Storment, 'Iridium-based Ultramicroelectrode Array Fabricated by Microlithography', *Anal. Chem.*, **6**, 418–423 (1994).
87. B.J. Seddon, Y. Shao, N.H. Girault, 'Printed Microelectrode Array and Amperometric Sensor for Environmental Monitoring', *Electrochim. Acta*, **39**, 2377–2386 (1994).
88. G.T.A. Kovacs, C.W. Storment, S.P. Kounaves, 'Microfabricated Heavy Metal Ion Sensor', *Sens. Actuators B*, **23**, 41–47 (1995).
89. A. Uhlig, M. Paeschke, V. Schnakenberg, R. Hintsche, H.J. Diederich, F. Scholz, 'Chip-array Electrodes for Simultaneous Stripping of Trace Heavy Metals', *Sens. Actuators B*, **24–25**, 899–903 (1995).
90. R.J. Reay, A.F. Flannery, C.W. Storment, S.P. Kounaves, G.T.A. Kovacs, 'Microfabricated Electrochemical Analysis System for Heavy Metal Detection', *Sens. Actuators B*, **34**, 450–455 (1996).
91. M.A. Augelli, V.B. Nascimento, J.J. Pedrotti, I.G.R. Gutz, L. Angnes, 'Flow-through Cell Based on an Array of Gold Microelectrodes Obtained from Modified Integrated Circuit Chips', *Analyst*, **122**, 843–847 (1997).

92. C. Belmont-Hébert, M.L. Tercier, J. Buffle, G.C. Fiaccabrino, N.F. de Rooij, M. Koudelka-Hep, 'Gel-integrated Microelectrode Arrays for Direct Voltammetric Measurements of Heavy Metals in Natural Waters and Other Complex Media', *Anal. Chem.*, **70**, 1949–2956 (1998).
93. M.T. Carter, E.D. Cravens, 'Hybrid Electrochemical/ Microfluidic Monitors for Trace Heavy Metals', *Proc. SPIE*, **3534**, 251–260 (1999).
94. Q.G. Wu, G.E. Batley, 'Determination of Sub-nanomolar Concentrations of Lead in Seawater by Adsorptive Stripping Voltammetry with Xylenol Orange', *Anal. Chim. Acta*, **309**, 95–101 (1995).
95. O. Abollino, M. Aceto, G. Sacchero, C. Sarzanini, E. Mentasti, 'Determination of Copper, Cadmium, Iron, Manganese, Nickel and Zinc in Antarctic Seawater. Comparison of Electrochemical and Spectroscopic Procedures', *Anal. Chim. Acta*, **305**, 200–206 (1995).
96. H. Li, R.B. Smart, 'Determination of Sub-nanomolar Concentration of Arsenic(III) in Natural Waters by Square Wave Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **325**, 25–32 (1996).
97. J. Wang, J. Lu, K. Olsen, 'Measurement of Ultratrace Levels of Chromium by Adsorptive Stripping Voltammetry in the Presence of Cupferron', *Analyst*, **117**, 1913–1917 (1992).
98. A.Z. Abuzuhri, W. Voelter, 'Applications of Adsorptive Stripping Voltammetry for the Trace Analysis of Metals, Pharmaceuticals and Biomolecules', *Fresenius' J. Anal. Chem.*, **360**, 1–9 (1998).
99. J. Wang, *Electroanalytical Techniques in Clinical Chemistry and Laboratory Medicine*, VCH, New York, 1988.
100. J.M. Zen, P.J. Chen, 'A Selective Voltammetric Method of Uric Acid and Dopamine Detection Using Clay-modified Electrodes', *Anal. Chem.*, **69**, 5087–5093 (1997).
101. A. Arranz, S.F. Debetono, J.M. Moreda, A. Cid, J.F. Arranz, 'Voltammetric Behaviour of the Antimycotic Terbinafine at the Hanging Mercury Drop Electrode', *Anal. Chim. Acta*, **351**, 97–103 (1997).
102. Y.N. Ni, J.L. Bai, L. Jin, 'Simultaneous Adsorptive Voltammetric Analysis of Mixed Colorants by Multivariate Calibration Approach', *Anal. Chim. Acta*, **329**, 65–72 (1996).
103. V.S. Ferreira, M.V.B. Zanoni, A.G. Fogg, 'Cathodic Stripping Voltammetric Determination of Ceftazidime in Urine at a Hanging Mercury Drop Electrode', *Microchem. J.*, **57**, 115–122 (1997).
104. A. Arranz, S.F. Debetono, J.M. Moreda, J.F. Arranz, 'Study of the Electro-adsorptive Behavior of the Herbicide Nitralin by Means of Voltammetric Techniques', *Talanta*, **45**, 417–424 (1997).
105. A. Arranz, S.F. Debetono, J.M. Moreda, A. Cid, J.F. Arranz, 'Cathodic Stripping Voltammetric Determination of Doxazosin in Urine and Pharmaceutical Tablets Using Carbon Paste Electrodes', *Analyst*, **122**, 849–854 (1997).
106. C.M.P. Vaz, S. Crestana, S.A.S. Machado, L.H. Mazo, L.A. Avaca, 'Electroanalytical Determination of the Herbicide Atrazine in Natural Waters', *Int. J. Environ. Anal. Chem.*, **62**, 65–76 (1996).
107. A. Arranz, S.F. Debetono, J.M Moreda, A. Cid, J.F. Arranz, 'Preconcentration and Voltammetric Determination of the Herbicide Metamitron with a Silica-modified Carbon Paste Electrode', Mikrochim. Acta, **127**, 273–279 (1997).
108. E. Reichart, D. Obendorf, 'Determination of Naringin in Grapefruit Juice by Cathodic Stripping Differential Pulse Voltammetry at the Hanging Mercury Drop Electrode', *Anal. Chim. Acta*, **360**, 179–187 (1997).
109. Y. Fushinuki, I. Taniguchi, 'Determination of Methylamphetamine in Urine by Differential Pulse Polarography', *Anal. Sci.*, **14**, 265–268 (1998).
110. Y.T. Long, J.J. Zhu, H.Y. Chen, 'Preconcentration and Voltammetric Determination of Trace Myoglobin at a 6-Mercaptopurine Modified Silver Electrode', *Fresenius, J. Anal. Chem.*, **360**, 614–617 (1998).
111. M. Rizk, F. Belal, F.A. Aly, N.M. Elenany, 'Differential Pulse Polarographic Determination of Ofloxacin in Pharmaceuticals and Biological Fluids', *Talanta*, **46**, 83–89 (1998).
112. J. O'Dea, J. Osteryoung, 'Rapid Scan Square Wave Voltammetric Detector for High Performance Liquid Chromatography', *Anal. Chem.*, **52**, 2215–2216 (1980).
113. E.D. Yalvac, R.A. Bredeweg, 'Status of Process Flow-injection Analysis and Current Trends', *Proc. Control Anal.*, **6**, 41–95 (1994).
114. C. Wechter, M. Sleszynski, J.J. O'Dea, J. Osteryoung, 'Anodic Stripping Voltammetry with Flow Injection Analysis', *Anal. Chim. Acta*, **175**, 45–53 (1985).
115. J. Alpízar, A. Cladera, V. Cerdá, E. Lastres, L. Garcia, M. Caduceus, 'Simultaneous Flow Injection Analysis of Cadmium and Lead with Differential Pulse Voltammetric Detection', *Anal. Chim. Acta*, **340**, 149–158 (1997).
116. F. Zhu, J.W. Ironstone, M.R. Ruegnitz, 'High Throughput Fast-scan Anodic Stripping Voltammetry in a Microflow System', *Anal. Chem.*, **69**, 728–733 (1997).
117. L.A. Mahoney, J. O'Dea, J.G. Osteryoung, 'Development and Characterization of an Automated Flow System for Voltammetric Analysis', *Anal. Chim. Acta*, **281**, 25–33 (1993).
118. R. Samuelsson, J. Osteryoung, 'Determination of *N*-Nitrosamines by High Performance Liquid Chromatographic Separation with Voltammetric Detection', *Anal. Chim. Acta*, **123**, 97–105 (1981).
119. L.A. Mahoney, J. O'Dea, J.G. Osteryoung, 'Determination of Charge Transfer Parameters Using an Automated Flow System', *J. Electroanal. Chem.*, **366**, 81–92 (1994).

Scanning Tunneling Microscopy, In Situ, Electrochemical

Kingo Itaya
Tohoku University, Sendai, Japan

1 Introduction 1
2 Experimental Aspects 2
2.1 Principle of Scanning Tunneling Microscopy 2
2.2 Principle of In Situ Scanning Tunneling Microscopy 2
2.3 Preparation of Well-defined Electrode Surfaces 2
3 Structure of Specifically Adsorbed Anions 4
3.1 Iodine Adlayers on Platinum(111) 5
3.2 Iodine on Gold(111) and Silver(111) 5
3.3 Sulfate/Hydrogensulfate on Gold(111), Platinum(111), and Rhodium(111) 7
3.4 Cyanide 9
4 Underpotential Deposition 11
4.1 Copper Underpotential Deposition on Gold(111) 11
5 Adsorption of Organic Molecules on Modified Electrodes 12
5.1 5,10,15,20-Tetrakis(*N*-methyl-pyridinium-4-yl)-21*H*,23*H*-porphyrine on Iodine–Gold(111) 14
5.2 Other Molecules on Iodine–Gold(111) 15
6 Adsorption of Aromatic Molecules on Clean Bare Electrodes 17
6.1 Benzene on Rhodium(111) in Hydrofluoric Acid 18
6.2 Other Molecules 22
7 Electrochemical Dissolution Processes of Semiconductors 24
7.1 Silicon(111) 24
7.2 Other Semiconductors 27
8 Conclusion 30
Acknowledgments 30
Abbreviations and Acronyms 30
Related Articles 31
References 31

This article describes the in situ scanning tunneling microscopy (STM) operated at electrode–electrolyte interfaces. It is demonstrated that in situ STM makes it possible to monitor with atomic resolution a wide variety of electrode processes such as the adsorption of inorganic and organic species and the dissolution and deposition of metals and semiconductors. Owing to limitations on space, the focus is on selected topics and mainly on our own experimental results.

1 INTRODUCTION

Of the phenomena which occur at the interface between a solid and a liquid, common examples include the deposition and corrosion of metals, the charging and discharging of storage batteries, and the wet processing of semiconductor devices. Such processes, and many others of a similar nature, involve electrochemical (EC) oxidation–reduction reactions that take place at solid–electrolyte interfaces. Until recently, there had been few in situ methods available for the structural determination of an electrode surface, in solution, at the atomic level. Atomic level information had previously been acquired only via surface spectroscopic techniques under ultrahigh vacuum (UHV).[1,2]

However, since its invention by Binnig and Rohrer,[3] STM was immediately established as an invaluable and powerful surface analysis technique with atomic resolution in UHV. Belatedly, but assuredly, developments in STM operated at solid–liquid interfaces led to its valuation as arguably the premier technique for atomic-level surface structural investigations of chemical processes taking place at solid–liquid interfaces. It has been demonstrated that in situ STM makes it possible to monitor, under reaction conditions, a wide variety of electrode processes such as the adsorption of inorganic and organic species, the reconstruction of electrode surfaces, and the dissolution and deposition of metals and semiconductors. Several review articles on in situ STM and related techniques such as in situ atomic force microscopy (AFM) have been published.[4–7] The reviews by Gewirth and Niece[6] and the present author[7] are the most comprehensive in terms of results obtained on various substrates of metals and semiconductors.

This article describes the current status of in situ STM. Owing to limitations on space, the focus is on selected

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

topics and mainly on our own experimental results. Experimental procedures are only briefly described, because detailed reviews on this aspect have already been published.[5]

2 EXPERIMENTAL ASPECTS

2.1 Principle of Scanning Tunneling Microscopy

In conventional STM operated in vacuum or air, a voltage (V) is applied between the substrate and tip electrodes as shown schematically in Figure 1. The tunneling current density (i_T) can be expressed by Equation (1) when a low-bias voltage is applied:[8]

$$i_T = \frac{e^2}{h^2}\frac{(2m\bar{\phi})^{1/2}}{\Delta s}V\exp(-A\bar{\phi}^{1/2}\Delta s) \quad (1)$$

where e is the charge of an electron, h is Planck's constant, $\bar{\phi}$ is the mean barrier height and Δs is the distance between the two electrodes. A is defined by Equation (2):

$$A = 4\pi\frac{(2m)^{1/2}}{h} \quad (2)$$

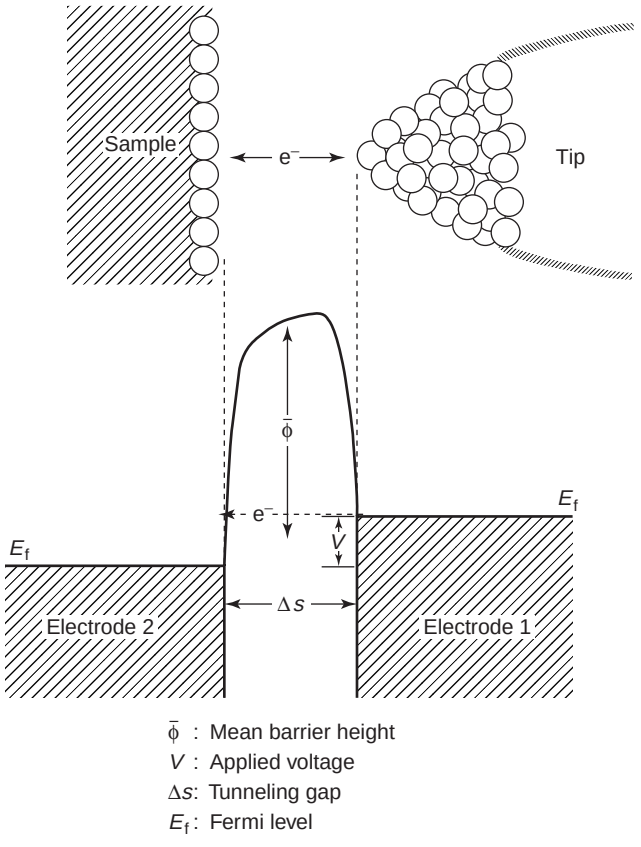

Figure 1 Tunneling barrier between two metal electrodes (substrate and tip).

where m is the mass of an electron. As can be seen in Equation (1), the tunneling current is exponentially dependent on the width of the potential barrier (Δs) and the square root of the mean barrier height. This characteristic exponential dependence allows STM to achieve high resolution on the z-axis. For typical metals ($\bar{\phi} = 4-5$ eV) a predicted change in the tunneling current (i_T) of one order of magnitude with a change $\Delta s = 0.1$ nm has been verified. If the tunneling current is kept constant to within 2%, then the tunneling gap (Δs) remains constant to within 0.001 nm.

2.2 Principle of In Situ Scanning Tunneling Microscopy

The review by Siegenthaler describes a detailed comparison between various types of electric circuits to control the electrode potentials of the tunneling tip and the substrate independently using the so-called bipotentiostat.[5] Figure 2 illustrates the apparatus for in situ STM with the four-electrode configuration and shows the EC cell. Using a bipotentiostat, the electrode potentials of the substrate [working electrode (WE) one (WE_1)] and the tunneling tip (WE_2) can be controlled independently with respect to a reference electrode (RE). The EC current (i_F) flowing through the substrate and the counter electrode (CE) can be monitored from the output of a current follower. The tunneling current (i_T) can be measured by the other amplifier. The potential difference between WE_1 and WE_2 is equivalent to the bias voltage (V) in Figure 1.

The side wall of the tip must be isolated in order to reduce the background EC current flowing through the tip. Soft glass, organic polymers, and Apiezon wax have been used. Details of tip-coating methods have been described by Siegenthaler.[5]

2.3 Preparation of Well-defined Electrode Surfaces

As a fundamental basis for all STM studies, electrode–electrolyte interfaces must be prepared reproducibly, and methods must be established to observe these interfaces accurately. Well-defined single crystalline surfaces must be exposed to solution in order to understand

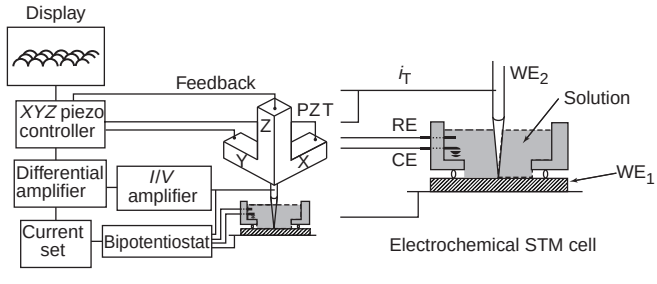

Figure 2 Apparatus for in situ STM with the four-electrode configuration.

surface structure–reactivity relationships on the atomic scale. It is still difficult to elucidate EC reactions on the atomic scale using polycrystalline electrodes. Efforts have succeeded in producing extremely well-defined, atomically flat surfaces of various electrodes made of noble metals, base metals, and semiconductors without either oxidation or contamination in solution.

2.3.1 Flame-annealing and Quenching Method

A unique and very convenient way to expose well-defined clean Pt to aqueous solution was proposed by Clavilier et al. in 1980, in which mechanically exposed single-crystal Pt was annealed in an oxygen flame and quenched in pure water.[9] They also established a method for preparing a single-crystal Pt electrode by melting a Pt wire in a flame. This technique was extended by Hamelin for Au,[10] by Motoo and Furuya for Ir,[11] and by us for Rh and Pd.[12]

Figure 3 shows typical examples of a cyclic voltammogram (CV) of the three low-indexed Pt surfaces in sulfuric acid solution obtained in our laboratory. The results show that the hydrogen adsorption–desorption reaction is a very structure-sensitive reaction on the Pt surfaces, as indicated by the different shapes and peak positions for the CVs of the different crystalline faces. Although Clavilier et al. quantitatively analyzed the binding sites of hydrogen using systematically prepared stepped surfaces,[13] more recent investigations using a CO replacement technique clearly indicated that the charges shown in Figure 3 include a significant contribution of the adsorption and desorption of sulfate/bisulfate.[14]

Nevertheless, direct evidence to support the existence of a well-defined surface in solution was presented by us in 1990 using in situ STM.[15] Figure 4(a) shows our first STM image of a flame-annealed Pt(111) in sulfuric acid solution. The height of each step is ca. 0.23 nm, in accord

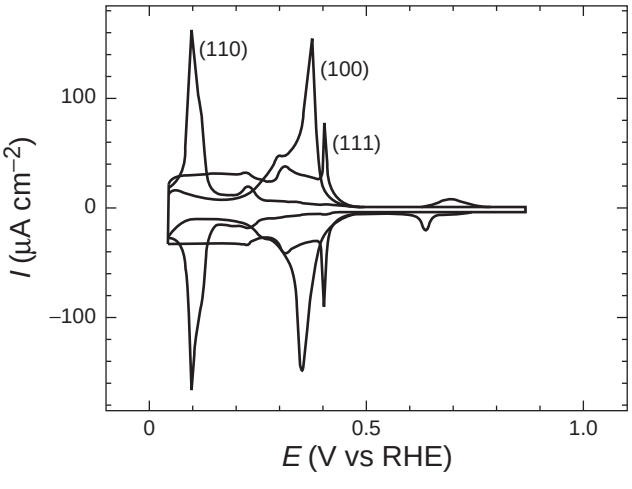

Figure 3 CVs of Pt(111), Pt(110), and Pt(100) in 0.5 M H_2SO_4. Scan rate: 50 mV s^{-1}. RHE, reversible hydrogen electrode.

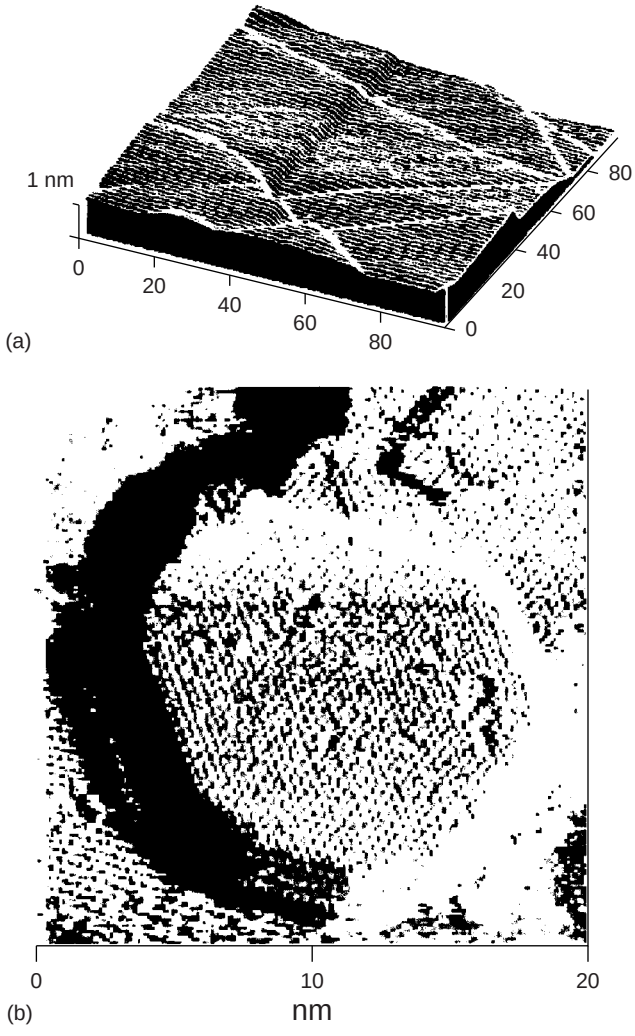

Figure 4 In situ STM images of flame-annealed Pt(111) in solution. (Figure 4(a) reprinted from K. Itaya, S. Sugawara, K. Sashikata, N. Furuya, 'In Situ Scanning Tunneling Microscopy of Platinum(III) Surface with the Observation of Monoatomic Steps', *J. Vac. Sci. Technol.*, **A8**, 515–519,[15] Copyright (1990) American Vacuum Society. Figure 4(b) reprinted from S. Tanaka, S.-L. Yau, K. Itaya, 'In-situ Scanning Tunneling Microscopy of Bromine Adlayers on Pt(111)', *J. Electroanal. Chem.*, **396**, 125–130,[16] Copyright (1995), with permission from Elsevier Science.)

with the monatomic step height of 0.238 nm on a Pt(111) surface. The monatomic steps observed on the surface are usually located on nearly parallel straight lines or form an angle of 60°, as expected for a surface with threefold symmetry. The terraces seem to be atomically flat. Later it was shown that the terrace was composed of Pt atoms forming a (1×1) structure as shown in Figure 4(b).[16] On the upper and lower terraces, the Pt(111–1×1) structure was clearly discerned at potentials near the hydrogen evolution reaction. The nearest-neighbor spacing and corrugation height were 0.28 and 0.03 nm, respectively.

Flame-annealed Au single crystals have been more frequently used for various studies including the potential-induced reconstruction investigated by several groups.[17] However, it must be emphasized that the flame-annealing method can be applied only to Au, Pt, Rh, Pd, Ir, and possibly Ag.

2.3.2 Ultrahigh Vacuum Electrochemical Methods

It is well established that clean surfaces are exposed in UHV by cycles of Ar-ion bombardment and high-temperature annealing. The surface structure and composition are usually determined by low-energy electron diffraction (LEED) and Auger electron spectroscopy (AES). By using an ultrahigh vacuum electrochemical (UHVEC) system, in which a chamber for EC measurements was interfaced to a UHV apparatus, well-defined substrates can be transferred into an EC cell in a purified Ar atmosphere. This experimental procedure was successfully applied to various metals such as Pt, Au, Pd, and Rh.[1,2] However, for some metals, such as Ni, the oxidation of the surface took place in the EC chamber before immersion of the electrode in electrolyte solutions owing to the presence of trace amounts of oxygen and water vapor.[18] The same difficulty was encountered with Cu electrodes.[19] It is clear that the problem of substrate oxidation of reactive metals occurring during the immersion and emersion processes is still unsolved in the UHVEC method.

2.3.3 Iodine–Carbon Monoxide Replacement Technique

It is important to note that iodine adlayers are known to protect highly sensitive surfaces of metal single crystals from oxidation and contamination in the ambient atmosphere, providing easy preparation and handling of well-defined surfaces during EC measurements.[1,2] The iodine–CO replacement is known to be a method for exposing well-defined and clean surfaces of such electrodes as Pt and Rh in solution.[20,21] The adsorbed iodine on these surfaces can be replaced by a CO adlayer. Clean surfaces are then exposed in solution by the EC oxidation of CO from the surface.

2.3.4 Electrochemical Etching Method

As described above, the flame-annealing and quenching method can only be applied to limited metals such as Pt, Au, Rh, Pd, and Ir and cannot be used for more industrially important less noble metals such as Ni, Co, Fe, and Cu, because they are heavily oxidized in the flame as well as in air. These metals are also difficult to transfer into electrolyte solutions without oxidation even by using UHVEC.

However, it was recently found that the anodic dissolution of various metals and semiconductors occurs only at the step edge under carefully adjusted EC conditions, resulting in atomically flat terrace-step structures. Although the etching method has not yet been well recognized as a promising method for exposing well-defined surfaces of various metals, we demonstrate in this article that layer-by-layer dissolution occurs on various semiconductors, such as Si, GaAs, and InP, resulting in the formation of atomically flat terrace-step structures. It was also found that the layer-by-layer dissolution occurs on various metals such as Ni, Ag, Co, Pd, and Cu.[7]

3 STRUCTURE OF SPECIFICALLY ADSORBED ANIONS

The adsorption of anions such as iodide, bromide, cyanide, and sulfate/hydrogensulfate on electrode surfaces is currently one of the most important subjects in electrochemistry. It is well known that various EC surface processes, such as the underpotential deposition (UPD) of hydrogen and metal ions are strongly affected by coadsorbed anions.[1,2] In particular, structures of the iodine adlayers on Pt, Rh, Pd, Au, and Ag surfaces have been extensively investigated using UHVEC techniques such as LEED.[1] For example, the commensurate $(\sqrt{3} \times \sqrt{3})R30^\circ$, (3×3), and $(\sqrt{7} \times \sqrt{7})R19.1^\circ$ adlattices were found to form on the well-defined Pt(111–1 × 1) surface, depending on the electrode potential and pH of the solution.[22] More recently, these structures were confirmed by STM in both air[23] and solution.[24] In contrast to Pt(111), only one phase of the commensurate $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure was observed on Pd(111) and Rh(111) surfaces with in situ STM.[25,26]

On the other hand, it has recently been recognized that iodine adlayer structures are more complicated on Au and Ag surfaces. Although several discrepancies about the iodine adlayer structure on Au(111) [I–Au(111)] are found in the literature,[27] surface X-ray scattering (SXS) studies carried out by Ocko et al. revealed structural changes of I–Au(111) in KI solution.[28] They found an increasing degree of compression, the so-called electrocompression, of the iodine adlattice with increasing iodine coverage and electrode potential. Instead of commensurate structures found on Pt(111), Rh(111), and Pd(111) as described above, they proposed that the iodine adlayer should be characterized as two distinct series of incommensurate adlattices, a centered rectangular phase and a rotated hexagonal phase.[28] We have recently reported the structures of I–Au(111) in KI solution determined by both ex situ LEED and in situ STM,[27] which agree with Ocko et al.'s SXS results.

Similar electrocompression was also found recently on Ag(111) using the LEED and in situ STM techniques.[29] Our results clearly demonstrate that the complementary use of LEED and in situ STM is a powerful technique for determining atomic structures of iodine adlayers on single-crystal electrodes.

3.1 Iodine Adlayers on Platinum(111)

The objective of this section is to describe the potential dependence of the structure of iodine on Pt(111). Hubbard et al. have extensively investigated the structure of the iodine adlayer formed on Pt(111) in aqueous iodide solutions using the UHVEC technique.[1,22] They demonstrated that the adlayer structures, mainly the commensurate (3×3) and $(\sqrt{7} \times \sqrt{7})R19.1°$, were formed on the well-defined Pt(111–1×1) surface in the double-layer potential range, depending on the electrode potential and pH of the solution. For example, in a solution containing 10 mM $KClO_4$ and 0.1 mM KI, adjusted to pH 4 with HI, they found the (3×3) and $(\sqrt{7} \times \sqrt{7})R19.1°$ structures in anodic and cathodic potential ranges, respectively [see Figure 4 in Hubbard et al.[22]].

Although previous STM studies revealed various atomic structures on Pt in air and in solution, no direct in situ STM investigation has been carried out for the structural change expected from the results reported by Hubbard et al. However, as described in our recent paper,[30] the structural transformation induced by changing the electrode potential did occur, but was surprisingly slow. Nearly perfect (3×3) and $(\sqrt{7} \times \sqrt{7})R19.1°$ adlayers could be prepared by immersion of the electrode in a solution containing iodide ions at anodic and cathodic potentials, which is consistent with the previous result.[22] However, in situ STM showed that both of the structures co-exist on the Pt(111) surface, when the clean Pt(111) electrode was immersed at potentials in the middle of the double-layer potential range. A typical example is shown in Figure 5(a), in which a $(\sqrt{7} \times \sqrt{7})$ domain can be clearly seen at the center of the image, surrounded by (3×3) domains. The potential-dependent structural change was directly probed by a time-dependent in situ STM experiment. When the electrode potential was stepped to the cathodic potential limit, the $(\sqrt{7} \times \sqrt{7})$ structure was expected to appear upon consuming the (3×3) domains according to previous work.[22] When the electrode potential was stepped after the acquisition of the image shown in Figure 5(a), the image shown in Figure 5(b) was obtained after 10 min, indicating that the interconversion between the two structures was very slow. Only a few iodine atoms marked by arrows were incorporated into the $(\sqrt{7} \times \sqrt{7})$ domain.

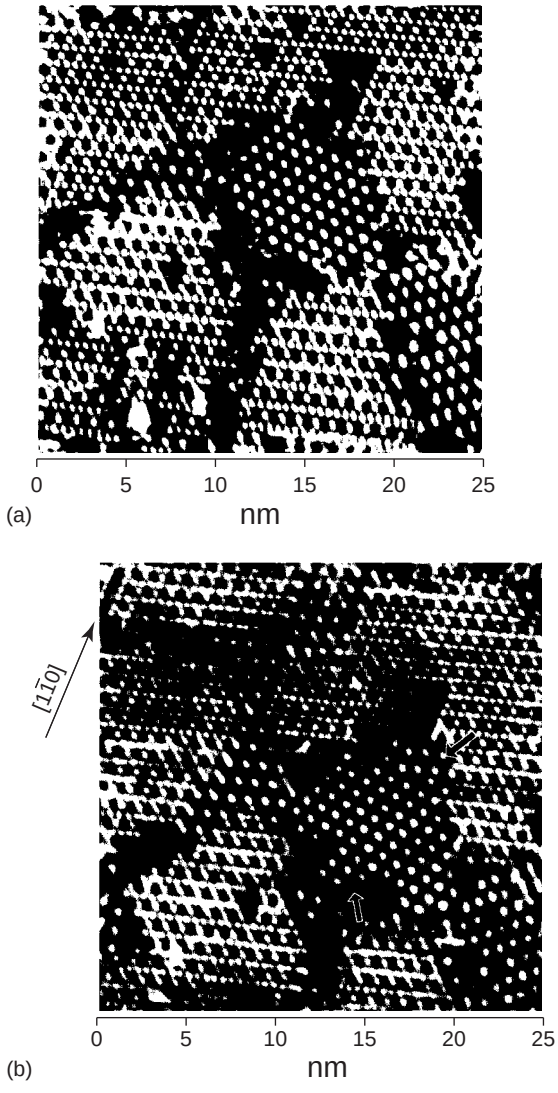

Figure 5 Adlayer structures of iodine on Pt(111): (a) at 0.45 V; (b) 10 min after the potential step from 0.45 to 0.15 V. (Reprinted with permission from J. Inukai, Y. Osawa, M. Wakisaka, K. Sashikata, Y.-G. Kim, K. Itaya, 'Underpotential Deposition of Copper on Iodine-Modified Pt(111): In Situ STM and Ex Situ LEED Studies', *J. Phys. Chem.*, **102**, 3498–3505,[30] Copyright (1998) American Chemical Society.)

The above result provides direct evidence that the surface diffusion of iodine atoms is very slow on Pt(111), suggesting that the iodine atoms are attached on Pt(111) through a strong chemical bond. The initially formed structure was almost insensitive to changes in the electrode potential under our experimental conditions.[30]

3.2 Iodine on Gold(111) and Silver(111)

Although the structures of the iodine adlayers are more complicated on Au(111) and Ag(111), in situ SXS studies by Ocko et al. revealed a series of I–Au(111)

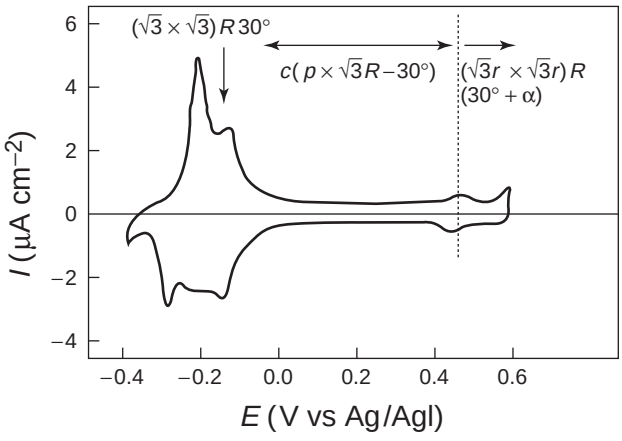

Figure 6 CV of Au(111) in 1 mM KI at a scan rate of 50 mV s^{-1}. (Reproduced from T. Yamada, N. Batina, K. Ogaki, S. Okubo, K. Itaya, 'Lateral Structure of Iodine Adlattices on Au(III) and Ag(III) Electrodes', in *Electrode Processes VI, Proceedings of the Sixth International Symposium*, eds. A. Wieckowski, K. Itaya, 1996, by permission of the Electrochemical Society, Inc.)

adlattices.[28] The adlattice constants varied continuously with the electrode potential in each of the two-dimensional phases designated by Ocko et al., the rectangular $c(p \times \sqrt{3})$ phase and the rotated hexagonal phase.[28]

We have extensively investigated the same system by the complementary use of in situ STM and ex situ LEED. Figure 6 shows a CV of a well-defined Au(111) in 1 mM KI. A pair of anodic–cathodic peaks associated with adsorption–desorption of iodine appear below 0 V vs Ag/AgI. The small pair of peaks at 0.5 V vs Ag/AgI is due to a phase transition between the two phases described in detail below. The transition between these two phases occurred at ca. 0.5 V, where a small reversible peak appears in the voltammogram in Figure 6. Adatoms were at the same height with atomic corrugation about 0.03–0.04 nm at potentials more negative than 0.4 V. The I adlattice obtained at −0.2 V possessed $(\sqrt{3} \times \sqrt{3})R30^\circ$ symmetry with a characteristic interatomic distance of 0.50 nm. No distortion from the threefold symmetry was observed. A positive shift of the electrode potential resulted in the formation of a more densely packed adlattice. The spacing of atoms in the $\sqrt{3}$ direction remained unchanged within the experimental error, regardless of the electrode potential between 0 and 0.4 V. Images obtained at more positive potentials contain more atoms in the direction perpendicular to the $\sqrt{3}$ direction, resulting in a uniaxial compression as described by Ocko et al. using SXS.

In the phase observed at potentials more positive than 0.5 V, the atomic adlattices of I possess a true sixfold symmetry. The nearest I–I distance was smaller than that of $(\sqrt{3} \times \sqrt{3})R30^\circ$ (0.50 nm), and the entire lattice seemed to be rotated by several degrees with respect to the $(\sqrt{3} \times \sqrt{3})R30^\circ$. This type of adlattice has been denoted as rot–hex, $(\sqrt{3}r \times \sqrt{3}r)R(30^\circ + \alpha)$ by Ocko et al.[28] It is interesting that the adlattices are furthermore modulated with periodically arranged surface features. These features, namely groups of slightly elevated I atoms, are interpreted as Moiré patterns resulting from the mismatch between the adlattice and the lattice of Au(111).

The Moiré pattern can be analyzed by simulation to determine the adlattice constant.[31,32] The STM images were simulated by computer calculation based on a simple "hard-ball contact model". Various compression ratios (0–15% subtracted from the values for $(\sqrt{3} \times \sqrt{3})R30^\circ$) and rotation angles [0–5° from $(\sqrt{3} \times \sqrt{3})R30^\circ$] were tested. The rot–hex I–I distance obtained by this simulation varied from 0.45 to 0.43 nm and the rotation angle from 2 to 5°, the variations corresponding to the change of potential from 0.45 to 0.55 V.

We also carried out quantitative analysis of LEED patterns using a UHVEC system as described in our previous papers.[27,31,32] A $c(p \times \sqrt{3}R - 30^\circ)$ structure was found in the potential range between −0.1 and 0.4 V, as shown in Figure 6. The variation of the adlattice constants is only along a particular direction of the Au atomic rows. As the largest value of p is equal to 3, at which the adlattice is identical with $(\sqrt{3} \times \sqrt{3})R30^\circ$, a value of p smaller than 3 signifies compression of the $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure only in the direction of the Au(111) atom row. Hence the variation of p is referred to as uniaxial compression.

The lattice parameters determined by in situ STM and ex situ LEED are fairly consistent with those obtained by SXS, which is usually believed to give the most accurate values. Nevertheless, our results are consistent with those obtained by SXS, demonstrating that the complementary use of in situ STM and ex situ LEED is a powerful technique for characterizing the atomic structure of iodine on Au(111).

The same phase transition from the $c(p \times \sqrt{3} - R30^\circ)$ to the rotated hexagonal structures was found and characterized by in situ STM and ex situ LEED for an iodine adlayer on Ag(111) in an HI solution.[29] Our result is consistent with that obtained using SXS carried out by Ocko et al.[33] LEED patterns with six splitting spots were observed at potentials before the bulk formation of AgI, indicating that the rotated hexagonal structure was stable even in UHV.[29] The iodine adlayer on Ag(111) was complicated in an alkaline solution, showing several structures including square $(\sqrt{3} \times \sqrt{3}R - 30^\circ)$ and $(\sqrt{3} \times \sqrt{3})R30^\circ$,[29] suggesting that there is a remarkable pH dependence of the structure of iodine on Ag(111).

3.3 Sulfate/Hydrogensulfate on Gold(111), Platinum(111), and Rhodium(111)

It has been demonstrated by several groups that in situ STM can be used to visualize adsorbed sulfate/hydrogensulfate species on Au(111), Pt(111), and Rh(111).

3.3.1 Gold(111)

An ordered structure with a $(\sqrt{3} \times \sqrt{7})$ symmetry was first observed for the adsorbed sulfate/hydrogensulfate on Au(111) in sulfuric acid by Magnussen et al.,[34] who proposed a model structure based on the assumption that the adsorbed species is hydrogensulfate, not sulfate, with a surface coverage of 0.4. More recently, Weaver et al. reported STM images with the same symmetry of $(\sqrt{3} \times \sqrt{7})$ on Au(111)[35] as that observed by Magnussen et al. However, they proposed a possibility of incorporation of hydronium cations in the ordered sulfate adlayer by taking into account the result that the surface coverage of sulfate on Au(111) determined by chronocoulometry and radiochemical assay is 0.2.[36] Note that this surface coverage is half of that for the structure proposed by Magnussen et al. as described above.

3.3.2 Platinum(111)

Stimming et al. found by in situ STM that adsorbed sulfate ions form the same adlayer structure as that found on Au(111).[37] Ordered domains with $(\sqrt{3} \times \sqrt{7})$ symmetry appeared in the potential range 0.5–0.7 V vs a RHE in 0.05 M H_2SO_4. As shown in Figure 3, only the (111) surface shows the characteristic butterfly peaks at potentials slightly more negative than 0.5 V. Their STM observations confirmed that the butterfly peaks are due to the adsorption and desorption of sulfate ions as indicated with the CO replacement technique by Clavilier et al. as described above.[14] STM images obtained on Pt(111) were interpreted in terms of the coadsorption of sulfate anions and water.

3.3.3 Rhodium(111)

High-resolution STM imaging conducted on atomically flat terraces at 0.5 V in H_2SO_4 readily discerned atomic features as shown in Figure 7(a) obtained near the step edges.[38] The image areas include three terraces with monatomic steps. It is clearly seen that parallel atomic rows in each domain are located in the directions forming angles of nearly 60° or 120°. It is also recognized that individual bright spots exist very near the monatomic step. This observation strongly indicates that the entire surface of Rh(111) is almost completely covered by adsorbed sulfate ions even very near the end of the terraces.

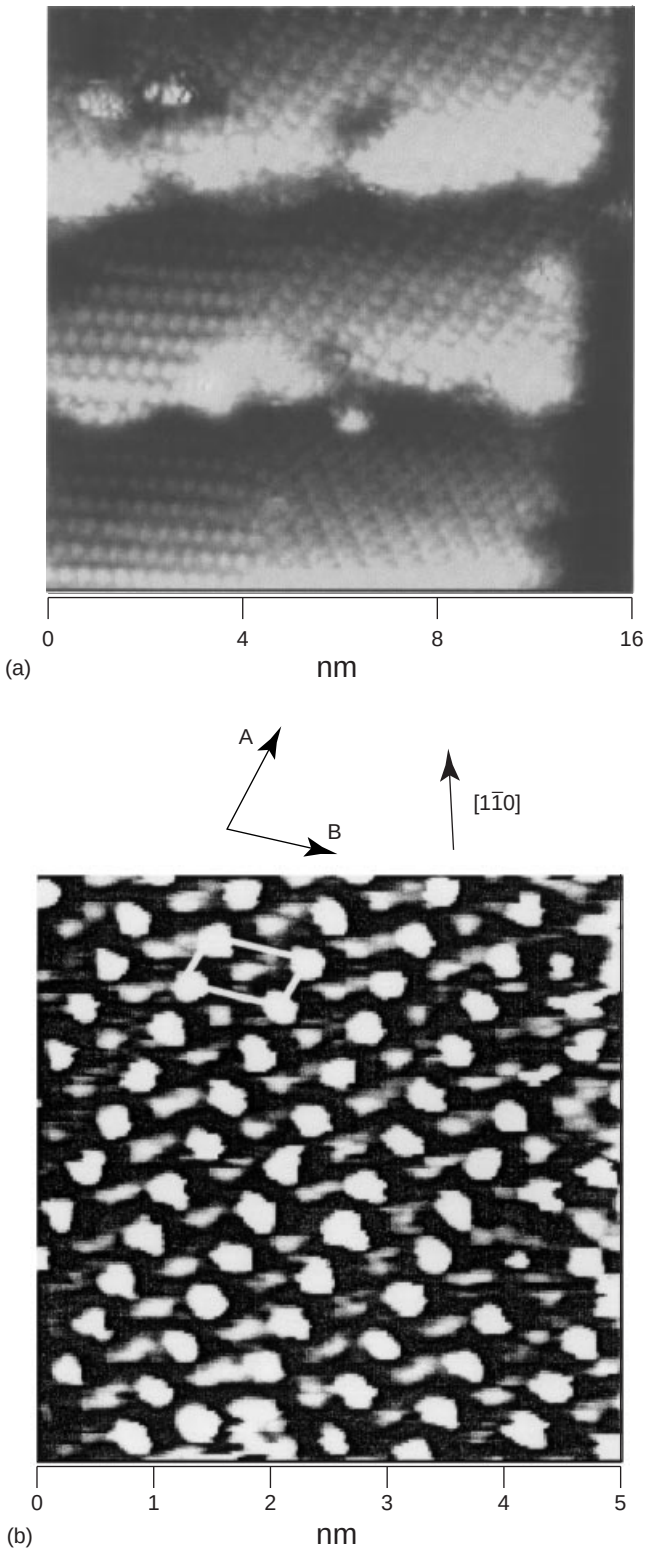

Figure 7 High-resolution STM images of sulfate adlayer on Rh(111) obtained in H_2SO_4. (Reprinted with permission from L.-J. Wan, S.-L. Yau, K. Itaya, 'Atomic Structure of Adsorbed Sulfate on Rh(111) in Sulfuric Acid Solution', *J. Phys. Chem.*, **99**, 9507–9513,[38] Copyright (1995) American Chemical Society.)

Figure 7(b) presents an STM image showing a more detailed internal structure acquired in an area where a single domain appeared on a wide terrace. It can be seen that there are two different parallel rows with a 30° rotation relative to the underlying Rh lattice. One appears as bright spots. The observed atomic distance in these bright rows along the A direction is equal to 0.46 nm. The average distance between neighboring bright rows is ca. 0.7–0.73 nm. The interatomic distance of 0.74–0.75 nm observed along the B direction in this particular STM image in Figure 7(b) is slightly larger than that of the $\sqrt{7}a_{Rh}$ (0.707 nm), probably owing to a small thermal drift during the acquisition of the image. However, it was ascertained that the distance along the B direction is very close to the $\sqrt{7}a_{Rh}$ based on the averaging of all atomic images obtained in this study. The angle between the directions marked by arrows A and B in Figure 7(b) is ca. 72°. The above results strongly indicate that the unit cell can be defined by the so-called $(\sqrt{3} \times \sqrt{7})$ structure. The rows along the direction marked by the arrow B are constituted of alternative bright and dark spots in the image shown in Figure 7(b). The dark spots appeared almost at the center between neighboring bright spots in the direction of B. Magnussen et al. have also found the position for the darker spots equidistant between two brighter spots along the B direction.[34]

The results described above are almost the same as those reported for Au(111) by Magnussen et al.[34] and other investigators[35] and those observed on Pt(111).[37] Magnussen et al. proposed a model structure with a unit cell, the so-called $(\sqrt{3} \times \sqrt{7})$, for the adlayer of hydrogensulfate (HSO_4^-) on Au(111). Both bright and dark spots were interpreted as hydrogensulfate ions adsorbed on Au(111). Therefore, the surface coverage of this proposed structure is 0.40.

According to the coverage value of ca. 0.2 obtained on Rh(111) by Zelenay and Wiekowski,[39] it is reasonable to expect that only the bright spots in the STM images observed on Rh(111) correspond to the adsorbed sulfate or hydrogensulfate. If sulfate or hydrogensulfate is assumed to be also trigonally coordinated on Rh(111), a ball-model can be presented as shown in Figure 8(a), where the SO_4^{2-} (or HSO_4^-) is positioned at the threefold hollow sites. It can be seen in Figure 8(a) that the sulfate ions along the $\sqrt{3}$ direction form an almost close-packed row. On the other hand, an open space can be found between neighboring rows of the sulfates. As described above, Weaver et al. proposed a model where coadsorbed hydronium cations exist along the $\sqrt{3}$ direction between neighboring rows of sulfates.[35] The dark spots which appeared in the STM images were assigned to be the coadsorbed hydronium cations. Such cation coadsorption was expected to minimize the coulombic repulsion between adjacent SO_4^{2-} on Au(111). Although the coadsorption of hydronium cations is thought to be a factor which explains the nonuniform interatomic distances of the $(\sqrt{3} \times \sqrt{7})$ structure, it is not clear why the adsorbed sulfates have the different spacings along the $\sqrt{3}$ and $\sqrt{7}$ directions.

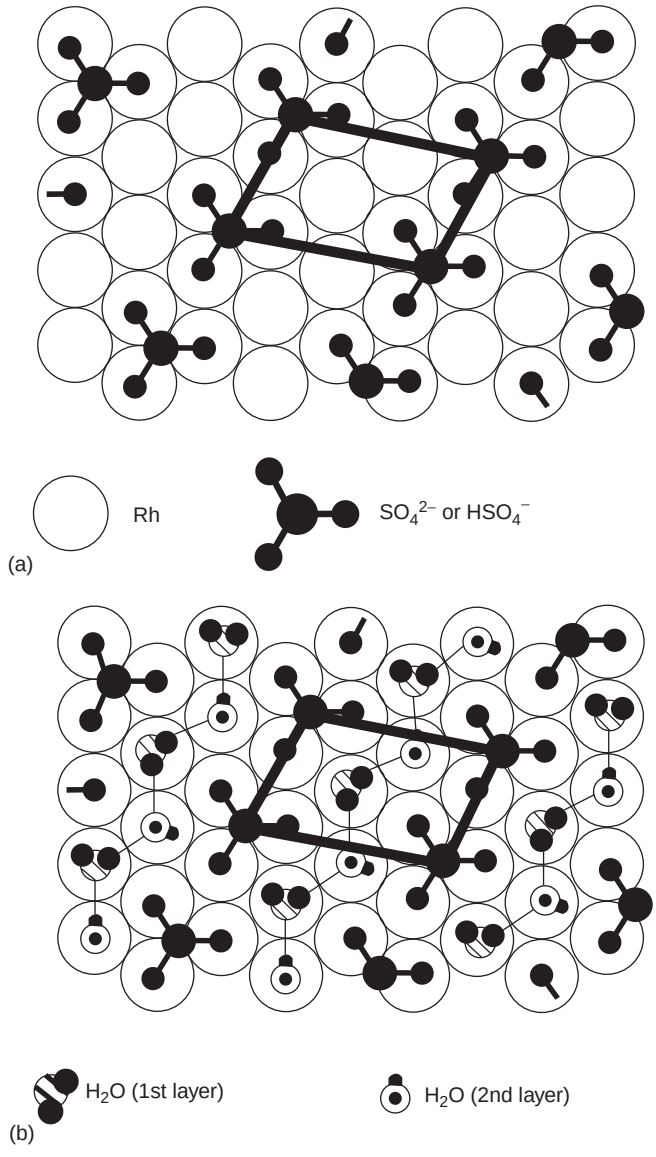

Figure 8 Model structures of the sulfate adlayer on Rh(111) surface: (a) sulfate on the threefold site; (b) coadsorption of sulfate and hydrogen-bonded water chains. (Reprinted with permission from L.-J. Wan, S.-L. Yau, K. Itaya, 'Atomic Structure of Adsorbed Sulfate on Rh(111) in Sulfuric Acid Solution', *J. Phys. Chem.*, **99**, 9507–9513,[38] Copyright (1995) American Chemical Society.)

In a previous paper, we proposed a new model to explain the nonuniform spacing in the unit cell of $(\sqrt{3} \times \sqrt{7})$.[38] It can be seen in Figure 8(a) that uncoordinated Rh atoms are arranged in a zig-zag form

in the $\sqrt{3}$ direction between neighboring rows of the adsorbed sulfates. In the new model shown in Figure 8(b), hydrogen-bonded water chains are simply inserted along the $\sqrt{3}$ direction between neighboring rows of the sulfates. The model shown in Figure 8(b) includes the adsorbed sulfate/hydrogensulfate and hydrogen-bonded water chains formed along the $\sqrt{3}$ direction. The model shows only the first water bilayer. Water molecules in the first layer are bonded directly to Rh atoms at the top site via the oxygen lone pair. For the ice-like lattices on the fcc(111) surfaces, two hydrogen bonds form to oxygen lone pairs of two water molecules in the second layer. It is assumed, however, that only a hydrogen bond forms to an oxygen lone pair of a water molecule in the second layer as shown in Figure 8(b), although it is expected that the water in the second layer forms a hydrogen bond with an oxygen lone pair of sulfate, which is not drawn in Figure 8(b) for the sake of clarity. According to the model, it is possible that the dark spots which appeared in the STM image shown in Figure 7(b) arise from the water molecules in the second layer. Although the model presented here would seem to be equivalent or similar to the model proposed by Weaver et al.,[35] our model more confidently explains the feature of the nonuniform interatomic distances in the $(\sqrt{3} \times \sqrt{7})$ structure. It is also noteworthy that the hydrogen-bonded water chains are expected to form hydronium cations in acidic solutions. If the water molecule in the second layer is protonated to form hydronium cations, the model presented here is equivalent to the model proposed by Weaver et al.[35] We have calculated the heights of the adsorbed sulfate and the water molecule in the second layer for the model shown in Figure 8(b) using a hard-ball contact model, yielding values of 0.35 and 0.30 nm respectively. The outer sphere of the water molecule in the second layer is slightly lower than that of the sulfate. It is interesting to compare these values with the corrugation heights described above, although it is true from a theoretical point of view that corrugation heights observed in STM images should arise from electronic factors, such as wave functions of adsorbates rather than spatial structural factors. We believe that a similar coadsorption of water chains and sulfate might occur on Au(111) and also possibly on Pt(111).

The adlayer structure almost abruptly disappeared at potentials near the hydrogen evolution reaction, showing the Rh(111–1 × 1) structure, indicating that the adsorbed sulfate is completely desorbed at the potential of the hydrogen evolution.

Nevertheless, it is now clear that sulfate/hydrogensulfate form adlayers with the same structure and symmetry on at least three different substrates, Au(111), Rh(111), and Pt(111). It is noteworthy that the atomic diameters of Au, Rh, and Pt are 0.289, 0.268, and 0.278 nm, respectively, i.e. the diameter of Rh is the smallest. In general, structures of many adlayers depend strongly on the diameter of substrates. The appearance of the same $(\sqrt{3} \times \sqrt{7})$ structure on the substrates with different lattice parameters might suggest that the coadsorption of sulfate/hydrogensulfate and water illustrated in Figure 8(b) is flexible with respect to the change in lattice parameter of the substrate. Such flexibility may be due to the existence of water molecules between the sulfate/hydrogensulfate chains with a relatively weak hydrogen bonding.

3.4 Cyanide

In situ STM was recently employed by Weaver et al., for the first time, to examine the CN adlayer on Pt(111).[40] It was reported that six CN functional groups form a hexagonal ring with an additional CN in the center of the ring (see Figure 9b).

Our in situ STM observations revealed a new structure for the CN adlayer and the complexation of alkali metal cations such as Na^+ and K^+ with the CN adlayer on Pt(111).[41] An atomically resolved STM image acquired in an alkaline CN solution containing Na^+ cations is shown in Figure 9(a). Figure 9(a) shows an STM image of the hollow hexagonal pattern. This image (5 × 5 nm) was acquired in 0.1 mM NaCN + 0.1 M $NaClO_4$ (pH 9.5) at 0.6 V with a bias voltage of −50 mV and a tunneling current of 20 nA.

Under these experimental conditions, well-arranged hexagonal rings, aligned in a direction 30° rotated from the close-packed directions of Pt(111) lattice, are observed. The 0.95 ± 0.02 nm distance between the nearest-neighbor hexagonal rings, as measured from their centers, is roughly twice as large as the $\sqrt{3}$ lattice spacing of the Pt (0.2778 nm). This ordered atomic feature can be characterized as the $(2\sqrt{3} \times 2\sqrt{3})R30°$ structure. However, Weaver et al. reported an STM image [see Figures 3 and 4 in Stuhlmann et al.[40]] which was interpreted as the $(2\sqrt{3} \times 2\sqrt{3})R30°$–7CN structure with cyanides bound in symmetric top sites surrounded by hexagonal rings of near-top CN. It is important to see in Figure 9(a) that the center spots are now essentially invisible in the image, strongly suggesting that there is no cyanide in the center of the hexagonal ring.

Based on the results described above, two model structures are presented in Figures 9(b) and (c) to outline atomic arrangements of the adsorbed CN on Pt(111). The model structure shown in Figure 9(b) is essentially a replica of the model proposed by Weaver et al.[40] in which the center spot was attributed to the adsorbed CN. However, the image shown in Figure 9(a) suggests that the adsorbed CN is not located in the center of the six-membered ring. Figure 9(c) is a new model where the adsorbed CN in the center is removed.

It is interesting that the six-membered ring is similar in structure to crown ethers. Crown ethers are known

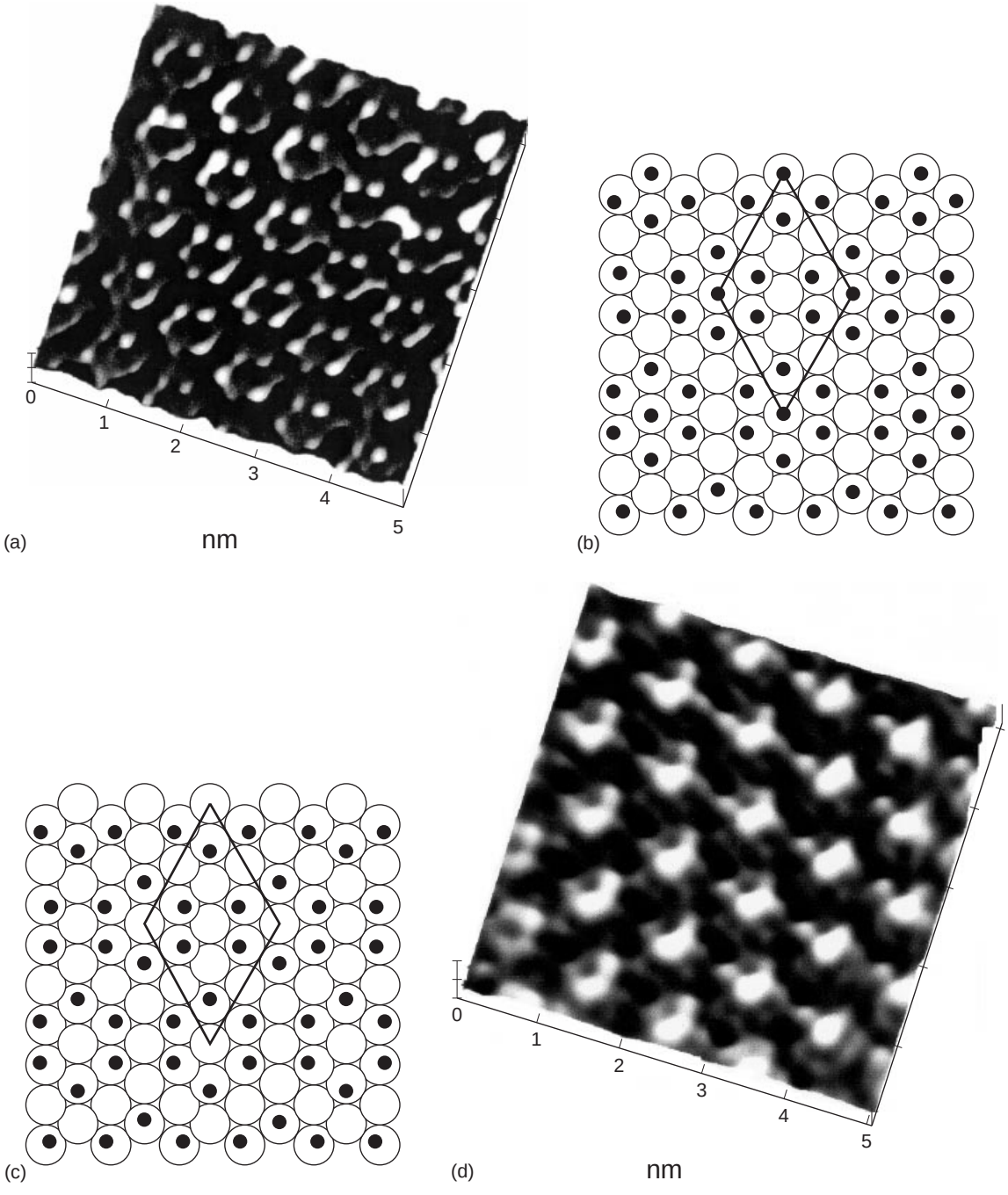

Figure 9 (a) High-resolution STM image of the hollow CN hexagonal arrangement. Two ball models of the $(2\sqrt{3} \times 2\sqrt{3})R30°$ structure are shown in (b) and (c). (d) STM image in the solution containing K^+ ions. (Reprinted with permission from Y.-G. Kim, S.-L. Yau, K. Itaya, 'Direct Observation of Alkali Cations on Cyanide-modified Pt(111) by Scanning Tunneling Microscopy', *J. Am. Chem. Soc.*, **118**, 393–400,[41] Copyright (1996) American Chemical Society.)

to complex effectively with alkali metal cations. The configuration of the CN adlayer is such that the C is bound to the Pt electrode with the N facing the solution side. Each nitrogen atom contains a lone pair of electrons which is expected to act as a binding site similar to the oxygen atoms in crown ethers.

It was surprising to find that the bright spots appear in the center of the six-membered ring in a solution containing K^+ as shown in Figure 9(d)[41] It was found that K^+ cations are more strongly bound than Na^+ in the center of the CN ring, because bright spots due to the coordinated Na^+ cations were only sparsely observed in

the center of the hexagonal ring of CN. Our STM result for the coordinated K^+ described above is probably the first case to describe an outer Helmholtz layer, because almost all previous STM studies have elucidated the adlayer structure directly attached on the electrode surface.

4 UNDERPOTENTIAL DEPOSITION

The EC adsorption of hydrogen and metals on a foreign metal substrate taking place in a potential region positive to the thermodynamically reversible potential is called the UPD. Particularly the UPD of hydrogen on single-crystal Pt electrodes was intensively and systematically investigated by Clavilier et al.[13] The UPD of a metal, M, on a different metal substrate, Ms, is expected to occur at potentials more positive than the reversible potential for the bulk deposition of M when an interaction between M and Ms is greater than that of Ms–Ms. The UPD process is important in EC reactions, such as metal deposition, as the initial step of a series of reactions and also because of electrocatalytic effects induced by adatoms formed by the process. Although a large number of UPD systems have been investigated by using conventional EC techniques, such as CV, to evaluate the thermodynamics and kinetics, the structural information of UPD layers was first obtained mainly by Hubbard et al. using the UHVEC technique.[1,2] With a considerable amount of previous knowledge available on the UPD phenomenon itself, in situ STM was applied, for the first time, to determine the structure of the adlayer of Cu on Au(111) in sulfuric acid solutions with atomic resolution.

4.1 Copper Underpotential Deposition on Gold(111)

Magnussen et al. reported the first atomic image of a Cu adlayer on Au(111) in sulfuric acid solution.[42] They found a $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure after the first UPD peak which transformed into a second phase of (5×5) structures. However, the appearance of the (5×5) structure was confirmed to be due to chloride contamination,[43] suggesting that the UPD process is extremely sensitive to co-adsorbates.

Figures 10(a) and (b) show CVs of an Au(111) electrode obtained in pure 0.05 M H_2SO_4 solution in the presence of 1 mM $CuSO_4$.[44] The main oxidation peak at 1.25 V and the cathodic peak at 0.82 V are due to the oxidation of the surface of Au(111) and the reduction of the oxide layers, respectively. On the other hand, two different waves for the UPD of Cu are clearly observed in the potential region between 0.35 and 0 V vs a saturated calomel electrode (SCE) before the beginning of the bulk deposition. Figure 10(b) shows a detailed CV for the UPD observed at a scan rate (v) of 1 mV s^{-1}. Two distinctly

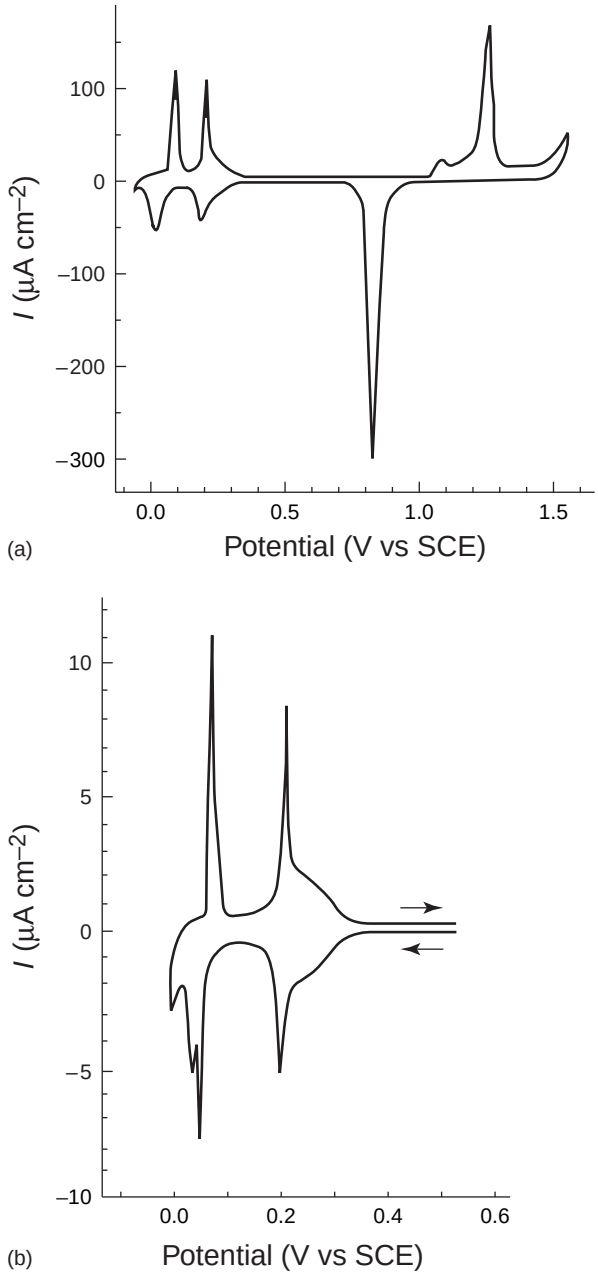

Figure 10 CVs for a Au(111) electrode in 0.05 M H_2SO_4 + 1 mM $CuSO_4$. Scan rate: (a) 20 and (b) 1 mV s^{-1}. (Reprinted from T. Hachiya, H. Honbo, K. Itaya, 'Detailed Underpotential Deposition of Copper on Gold(111) in Aqueous Solutions', *J. Electroanal. Chem.*, **315**, 275–291,[44] Copyright (1991), with permission from Elsevier Science.)

different processes can be seen clearly in the CV shown in Figure 10(b). The peak current is proportional to the scan rate only up to 5 mV s^{-1}, and then is approximately proportional to $v^{1/2}$, suggesting that the UPD of Cu on Au(111) is a surprisingly slow process.

Figure 11(a) shows a high-resolution STM image obtained in 0.05 M H_2SO_4 + 1 mM $CuSO_4$ solution.[44]

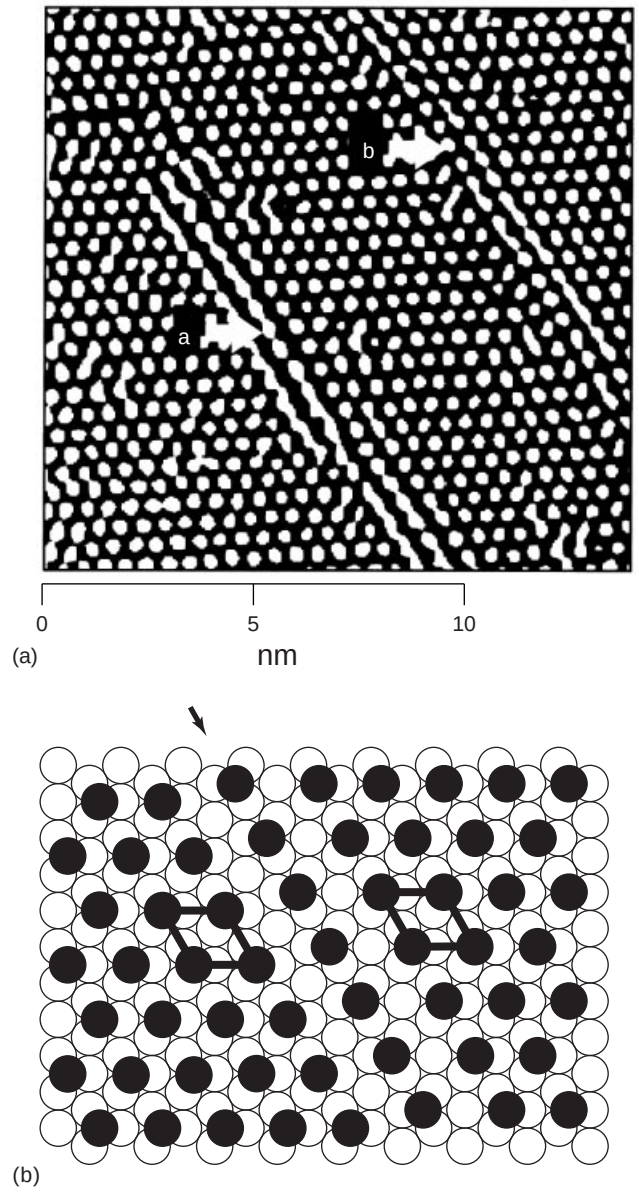

Figure 11 (a) High-resolution STM image and (b) a model structure of a copper adlattice on Au(111). (Reprinted from T. Hachiya, H. Honbo, K. Itaya, 'Detailed Underpotential Deposition of Copper on Gold(111) in Aqueous Solutions', *J. Electroanal. Chem.*, **315**, 275–291,[44] Copyright (1991), with permission from Elsevier Science.)

Although the wide terrace of the Au(111) surface was almost completely covered by the Cu adlayer with the $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure, several types of phase boundary can be seen in Figure 11(a). Figure 11(b) shows a model structure of the phase boundary marked by arrow (a), in which two $(\sqrt{3} \times \sqrt{3})R30^\circ$ domains are shifted by a half position along the direction indicated by the arrow. In the model structure shown in Figure 11(b), it was assumed that the solid circles represent Cu atoms. The same $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure was also found by in situ AFM.[45]

However, the coulometric curve obtained simultaneously with CV shown in Figure 10(b) showed that the ratio of the charges consumed during the first and second UPD processes was roughly 2 : 1, suggesting that the surface coverage of Cu was about 2/3 after the first UPD peak.[44] According to the model structure shown in Figure 11(b), the surface coverage must be 1/3 because of the $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure.

This discrepancy was carefully investigated by Shi and Lipkowski using a chronocoulometric technique. They measured the Gibbs excess of Cu adatoms and that of coadsorbed sulfate (SO_4^{2-}) as a function of the electrode potential, and concluded that Cu adatoms are packed to form a honeycomb $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure with the center of each honeycomb cell occupied by a sulfate ion.[46]

Finally, Toney et al. examined the above system using an SXS technique,[47] and concluded that the Cu atoms form a honeycomb lattice and are adsorbed on threefold hollow sites with sulfate ions located at the honeycomb centers. They also concluded that three oxygens of each sulfate ion bond to Cu atoms. According to all of the results described above, the most reliable model structure can be presented as shown in Figures 12(a) and (b), in which top and perspective views, respectively, are given. This model structure is essentially the same as that shown by Toney et al.[48] According to the model structure, the corrugation observed by in situ STM and AFM must be considered to be due to the coadsorbed sulfate ions, and not the Cu atoms. Note that Blum et al. studied a statistical mechanical model for the UPD of Cu.[48] Their proposed structure is similar to that shown in Figures 12(a) and (b).[48] The study of the UPD of Cu on Au(111) was a very important lesson for understanding limitations and strengths of various in situ techniques. It is clear that STM and AFM cannot distinguish chemical species.

Review articles are useful for finding the literature on UPD of various metals on different substrates.[6,7] Here we only describe the UPD of Cu on Au(111), which is the most extensively studied system.

5 ADSORPTION OF ORGANIC MOLECULES ON MODIFIED ELECTRODES

STM has also made it possible to determine directly orientations, packing arrangements, and even internal structures of organic molecules adsorbed both on surfaces in UHV and at solid–liquid interfaces. For example, individual molecules and distinguishable molecular shapes of benzene on Pt(111), coadsorbed benzene and CO on

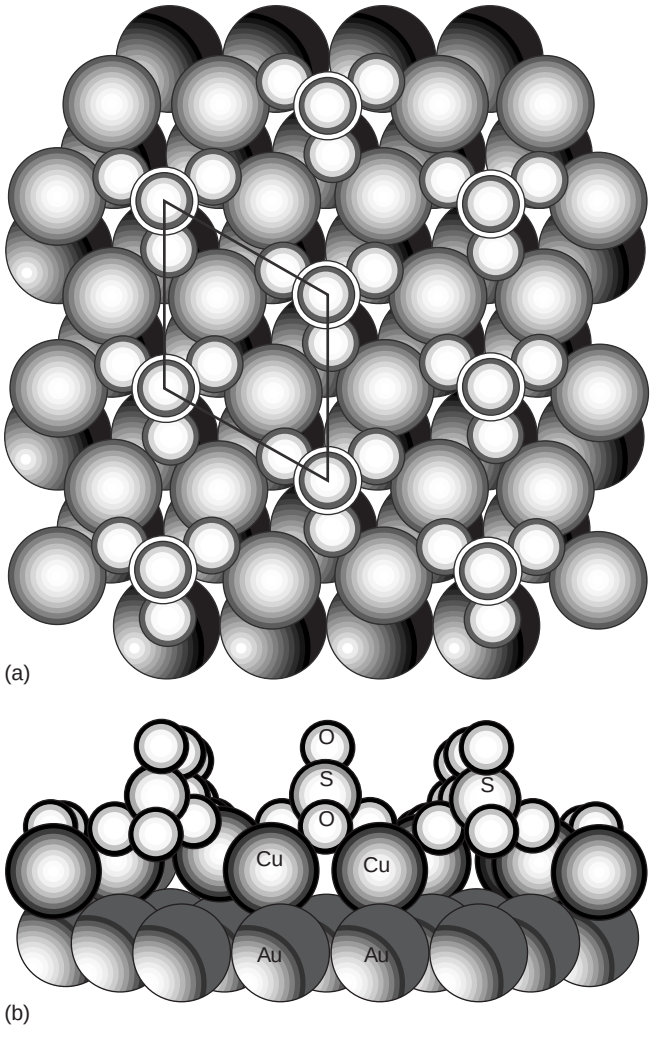

Figure 12 (a) Top and (b) perspective view of the coadsorbed Cu^{2+} and SO_4^{2-} adlayer on Au(111).

Rh(111), naphthalene on Pt(111), and copper phthalocyanine on Cu(100) have successfully been resolved with STM under UHV conditions.[49] These results have stimulated a large number of STM studies of ordered molecular adlayers in UHV and air and at solid–liquid interfaces.

A variety of experimental procedures have been reported for the preparation of ordered molecular adlayers on well-defined substrates including single crystals of metals and layered materials, such as highly ordered pyrolytic graphite (HOPG) and MoS_2.[50] Alkanethiols have been intensively investigated on metals such as Au, because the –SH group is known to be chemically attached to the Au surface through the formation of a covalent bond between S and Au atoms, producing densely packed adlayers. On the other hand, it is well known that simple physical adsorption can also provide ordered adlayers of molecules, such as liquid crystals and *n*-alkanes on inert substrates such as HOPG and MoS_2.[50]

From the EC point of view, the adsorption of organic molecules at electrode–electrolyte interfaces can be considered as one of the most promising approaches not only for the preparation of ordered adlayers, but also for elucidating the role of properties of adsorbed molecules and the nature of electrode–electrolyte interfaces.[1,51] In spite of a large number of reports describing observations by STM and related techniques, such as AFM of adsorbed organic molecules in UHV, air, and organic liquids, only a few in situ STM studies have been carried out for organic molecules adsorbed at electrode–electrolyte interfaces under EC conditions. Recently, pioneering studies have shown high-resolution images of molecules such as DNA bases (adenine, guanine, and cytosine) adsorbed on HOPG and Au(111) in electrolyte solutions.[6,7] In another study, xanthine and its oxidized form[52] and porphyrins[53] were found to form ordered adlayers on HOPG. Further, the order–disorder transition in a monolayer of 2,2′-bipyridine on Au(111) was reported as a function of electrode potential.[54]

Although a number of successful in situ experiments using STM and AFM have been performed to determine the structure of organic adlayers, HOPG, similar layered crystals, and Au electrodes have almost exclusively been used as the substrate. Therefore, the role of the interaction between organic molecules on one hand and substrates on the other in ordering processes is not yet fully understood. We have long been interested in finding a more appropriate substrate to investigate the adsorption of organic molecules. Recently, we disclosed a novel property of iodine-modified electrodes for the adsorption of organic molecules.[55–59] It was found, for the first time, that a water-soluble porphyrine, 5,10,15,20-tetrakis(*N*-methylpyridinium-4-yl)-21*H*,23*H*-porphyrine (TMPyP), formed highly ordered molecular arrays via self-ordering on the iodine-modified Au(111) electrode in $HClO_4$ solution.[55] As described in the preceding section, the iodine adlayers are thought to protect metal surfaces from oxidation and contamination in the ambient atmosphere, providing an easy method of preparation of well-defined surfaces.[1,2] However, it should now be recognized that the I–Au(111) electrode is one of the most promising substrates for the investigation of organic molecules in solution. Indeed, we discovered this electrode, with great generality, to be a suitable one on which to form highly ordered adlayers not only of TMPyP but also various other molecules such as crystal violet.[56] It was also demonstrated in our papers that various iodine-modified metal electrodes such as Ag(111)[58] and Pt(100)[59] can also be employed as a substrate on which to investigate the adsorption of organic molecules. In this section, we briefly

summarize in situ STM of organic molecules adsorbed on iodine-modified electrodes.

5.1 5,10,15,20-Tetrakis(*N*-methylpyridinium-4-yl)-21*H*,23*H*-porphyrine on Iodine–Gold(111)

The experimental procedure was fairly simple. A well-defined Au(111) surface prepared by the flame-annealing and quenching method was immersed in 1 mM KI solution for several minutes and then thoroughly rinsed with 0.1 M $HClO_4$ solution. The iodine-modified electrode, thus prepared, was installed in an EC cell containing pure $HClO_4$ solution for in situ STM measurements. Under potential control, the iodine adlayer structures can be determined by in situ STM as described above. After achieving an atomic resolution, a dilute solution of TMPyP was injected into the $HClO_4$ solution. After the addition of TMPyP, STM images for the iodine adlayer usually became unclear within the first 5–10 min, because of the adsorption of TMPyP, and then ordered adlayers became visible, extending over atomically flat terraces.[55]

Molecular orientations, packing arrangements and even internal molecular structures of the adsorbed TMPyP molecules can be seen in high-resolution STM images. Two typical STM images are presented in Figures 13(a) and (b). These high-resolution images directly demonstrate that the flat-lying TMPyP molecule can be recognized in the images as a square with four additional bright spots. The shape of the observed features in the image clearly corresponds to the chemical structure of the TMPyP molecule. The characteristic four bright spots located at the four corners of a square correspond to the pyridinium units of TMPyP. The center to center distance between the bright spots was found to be 1.3 nm measured diagonally, which is nearly equal to the distance between two diagonally located pyridinium units. In addition to the internal structure, the STM image shown in Figure 13(a) reveals details of the symmetry and the packing arrangement. It can be seen that there are three different molecular rows marked by arrows I, II, and III. In the row marked I, all TMPyP molecules show an identical orientation with an intermolecular distance of ca. 1.8 nm. An alternative orientation can be seen along the row II in which every second molecule shows the same orientation. On the other hand, the rotation angle of ca. 45° can be recognized between two neighboring molecules in row III.

The well-ordered TMPyP arrays were found to extend over atomically flat terraces as shown in Figure 13(b), which was obtained in areas involving monatomic step edges. It is clearly seen that the same molecular feature as that seen in Figure 13(a) extends over the lower and upper terraces. The ordered structure could be seen even on the relatively narrow terrace in the upper part of the image. It is also surprising to find that individual TMPyP molecules exist very near the monatomic step. The result shown in Figure 13(b) indicates that the entire

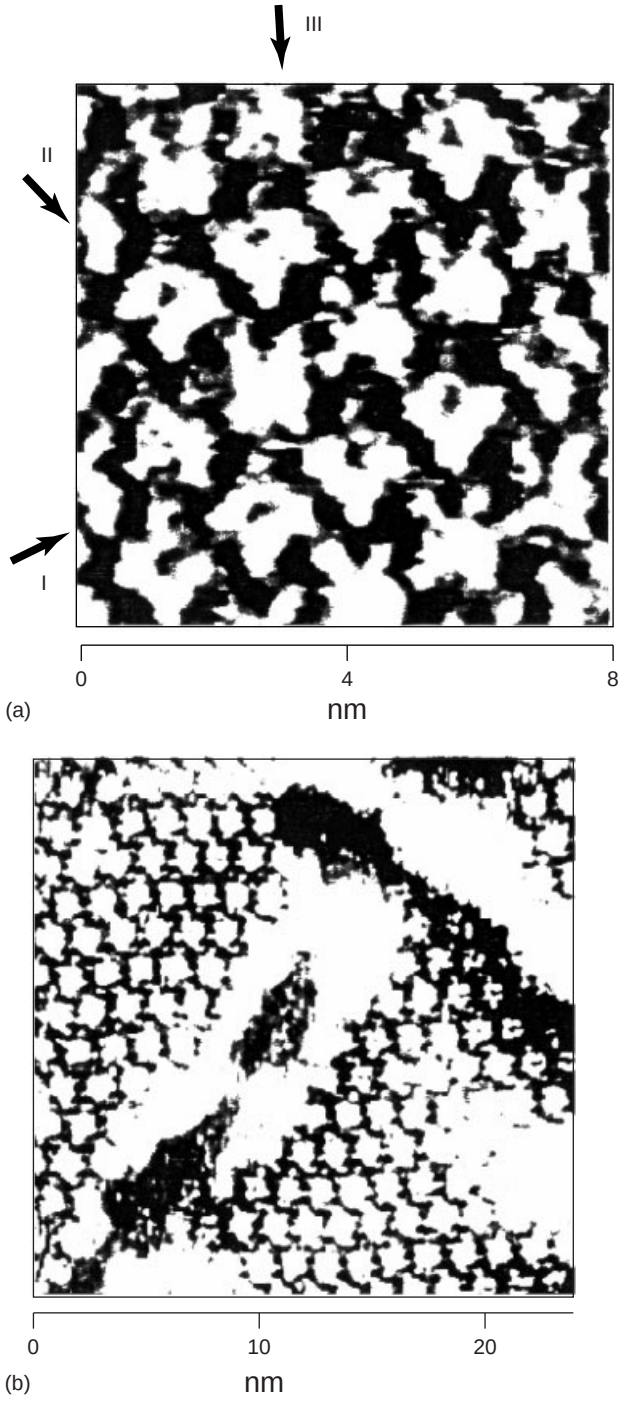

Figure 13 High-resolution STM images of TMPyP on I-modified Au(111) in $HClO_4$. (Reprinted from M. Batina, M. Kunitake, K. Itaya, 'Highly Ordered Molecular Arrays Formed on Iodine-modified Au(111) in Solution: In Situ STM Imaging', *J. Electroanal. Chem.*, **405**, 245–250,[56] Copyright (1996), with permission from Elsevier Science.)

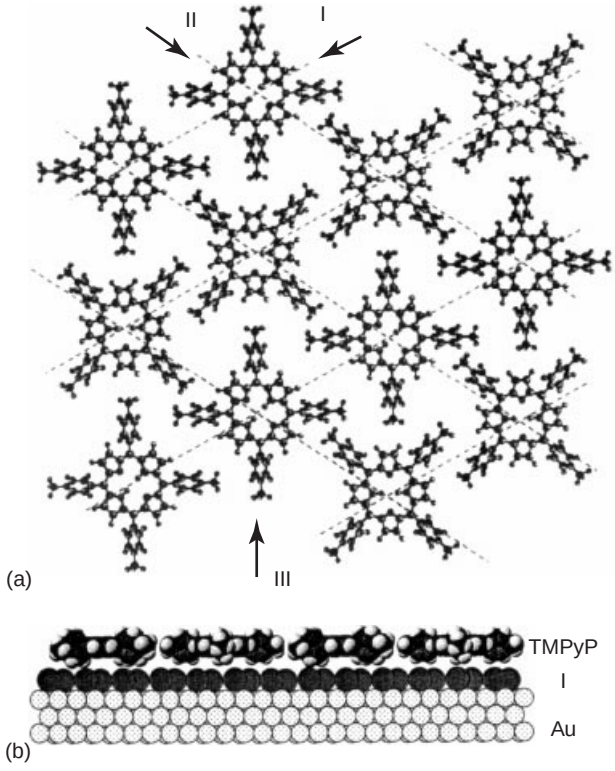

Figure 14 Illustrative depictions of the structure of TMPyP adlayer on I–Au(111): (a) top view; (b) side view. (Reprinted with permission from M. Kunitake, U. Akiba, N. Batina, K. Itaya, 'Structures and Dynamic Formation Processes of Porphyrine Adlayers on Iodine-modified Au(111) in Solution: In Situ STM Study', *Langmuir*, **13**, 1607–1615,[57] Copyright (1997), American Chemical Society.)

surface of the I–Au(111) is almost completely covered by ordered TMPyP molecules even near the end of the terraces. Figures 14(a) and (b) illustrate a structural model showing a top view and a side view of the ordered TMPyP adlayer on the iodine on Au(111), respectively. We have also investigated the adsorption of TMPyP on a well-defined Au(111) in the absence of iodine adlayer.[57] After achieving an atomic resolution for the Au(111–1 × 1) structure in 0.1 M $HClO_4$, a dilute solution of TMPyP was injected into the solution in a manner similar to that described above. Although the TMPyP molecules adsorbed directly on Au(111) could be seen by STM, the adsorbed molecules did not form ordered adlayers. Disordered adlayers formed on bare Au(111) suggest that strong interactions including chemical bonds between the Au substrate and the organic molecules prevented self-ordering processes from occurring, which must involve surface diffusion of the adsorbed molecules. The surface diffusion of the molecules adsorbed on bare Au was found to be very slow. Relatively weak van der Waals-type interactions between the hydrophobic iodine adlayer and the organic molecules could be the key factor promoting self-ordering processes on the I–Au(111) substrate.

It should particularly be emphasized that the iodine layer plays a crucial role in the formation of highly ordered TMPyP arrays. The relatively weak van der Waals-type interaction on the iodine adlayer seems to be a key factor in the formation of ordered molecular arrays of such large molecules. However, the relationship between the TMPyP and iodine adlayer structures is not fully understood, as described above,[57,58] because the iodine adlayer structures on Au(111) and Ag(111) are complicated by a potential-dependent compression in the adlayers.[27,29]

5.2 Other Molecules on Iodine–Gold(111)

Here, we briefly describe further evidence that the I–Au(111) electrode can be employed as an ideal substrate for in situ STM imaging of various adsorbed organic molecules in solution. Organic substances investigated were water-soluble cationic molecules purposely selected based on their characteristic shapes: triangular and linear. Hexamethylpararosaniline (crystal violet) and 4,4′-bis(*N*-methylpyridinium)-*p*-phenylenedivinylene (PPV) were also found to form highly ordered molecular arrays on top of the iodine monolayer adsorbed on Au(111).[56]

In situ STM with near-atomic resolution revealed their orientation, packing arrangement, and internal structure of each molecule. A typical high-resolution STM image of the molecular arrays of crystal violet (Figure 15b) is shown in Figure 15(a). It is also surprising to see that the STM image shows a distinctly characteristic, propeller-shaped feature for each molecule with highly ordered arrays. Each molecule has three benzene rings located at the apexes of a triangle with an equal distance from the central carbon atom. Three bright spots seem to correspond to the benzene rings. The center of each spot is located at a distance of ca. 0.35 nm from the center of the triangle. An additional spot can also be seen at the position of the central carbon atom of crystal violet. According to the STM image, it is clear that all molecules are oriented in the same direction. The unit cell shown in Figure 15(a) can be characterized by the lattice parameters, $a = 0.9$ nm, $b = 1.1$ nm, and the angle of ca. 75°, indicating that the crystal violet adlayer was slightly deformed from threefold symmetry.

The third compound investigated is the highly symmetric cationic PPV molecule with two terminal pyridinium rings connected with the straight phenylenedivinylene core. The image shown in Figure 16 is a typical STM image of the ordered PPV adlayers formed on the I–Au(111) surface. The individual flat-lying PPV molecule can be

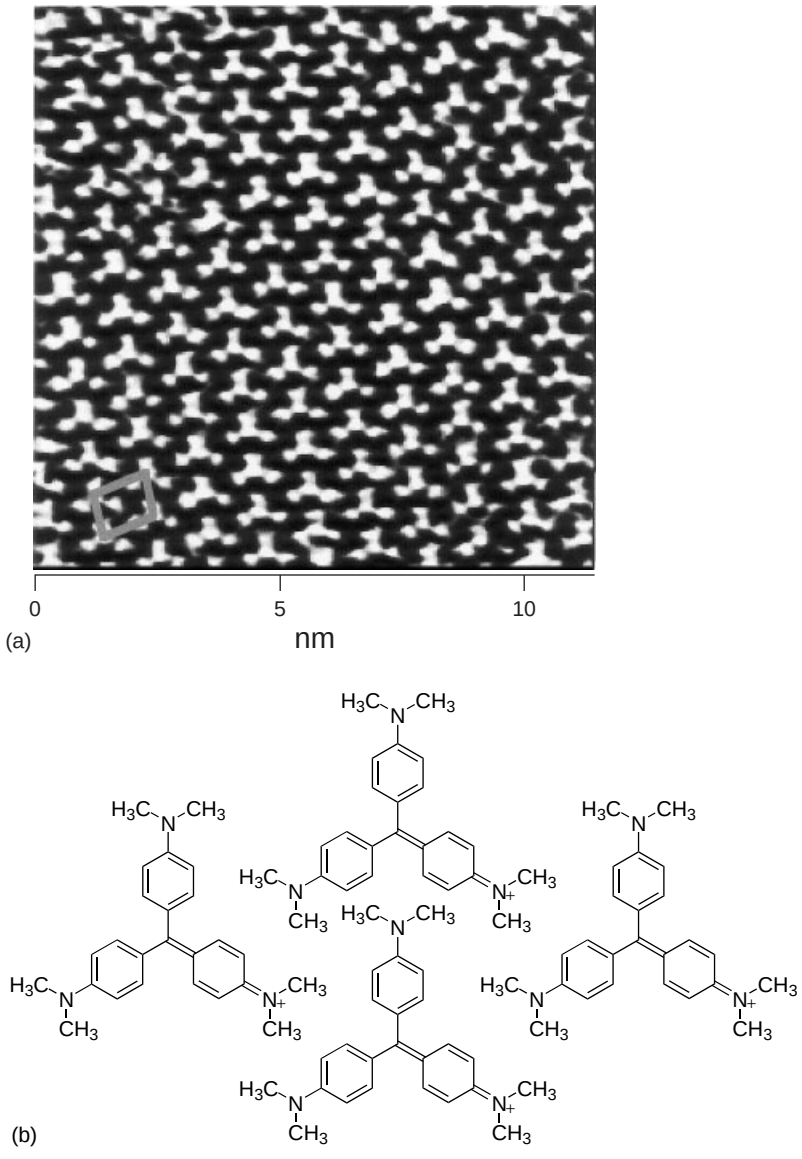

Figure 15 High-resolution STM image of crystal violet on I–Au(111). (Reprinted from M. Batina, M. Kunitake, K. Itaya, 'Highly Ordered Molecular Arrays Formed on Iodine-modified Au(111) in Solution: In Situ Scanning Tunneling Microscopy Imaging', *J. Electroanal. Chem.*, **405**, 245–250,[56] Copyright (1996), with permission from Elsevier Science.)

seen as a linearly aligned feature consisting of three bright spots that can be attributed to the three aromatic rings in PPV. Three bright spots aligned in a straight line suggest that the PPV molecules adsorb on the I–Au(111) surface with a straight configuration. PPV molecules are expected to form straight or bent configurations, depending on the relative orientation of the two trans CH=CH double bonds. The STM image shown in Figure 16(a) indicates that the two trans CH=CH double bonds are located on opposite sides, forming the straight configuration in the adlayer shown in Figure 16. It is also clearly seen that the tightly packed arrangement forms long striped domains. In each domain, all molecules show the same orientation as indicated by the model in Figure 16(b). The width of each domain along the molecular axis was found to be ca. 2.1 nm from the STM image, which corresponds to the total molecular length of PPV. It is also interesting that a zig-zag arrangement appears alternately in these striped domains.

In this section, we have presented further evidence that the I–Au(111) electrode can be employed as an ideal substrate for in situ STM imaging of adsorbed organic molecules in solution. The organic substances investigated were all water-soluble cationic molecules with characteristic shapes: triangular, linear, and square. Molecules of crystal violet, PPV, and TMPyP were all

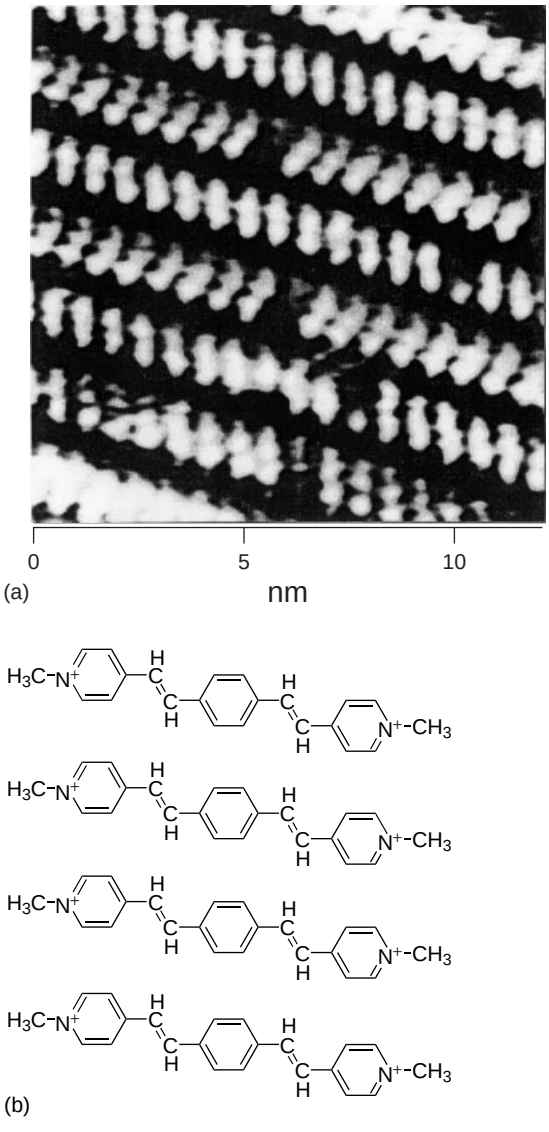

Figure 16 High-resolution STM image of PPV on I–Au(111). (Reprinted from M. Batina, M. Kunitake, K. Itaya, 'Highly Ordered Molecular Arrays Formed on Iodine-modified Au(111) in Solution: In Situ STM Imaging', *J. Electroanal. Chem.*, **405**, 245–250,[56] Copyright (1996), with permission from Elsevier Science.)

found to form highly ordered arrays on top of the iodine monolayer adsorbed on Au(111). The novel approach, using the iodine monolayer as an intermediate layer for the adsorption and formation of molecular arrays, has great potential for investigations of many organic molecules, including more complexed molecules and native biological materials.

In evaluating the structural relationship between iodine and TMPyP adlayers, we also investigated the structure of TMPyP on I–Pt(100).[59] It was found that adsorbed TMPyP molecules formed a highly ordered adlayer with a side-by-side configuration. The adlayer structure of TMPyP on I–Pt(100) is controlled by the interaction between the iodine adlayer and TMPyP.[59]

6 ADSORPTION OF AROMATIC MOLECULES ON CLEAN BARE ELECTRODES

The adsorption of organic molecules on bare electrode surfaces in electrolyte solutions under potential control has long been investigated for elucidating the role of the structures and properties of adsorbed molecules in EC reactions.[1,2,51] Although conventional EC and optical techniques, such as infrared (IR), Raman, and second harmonic generation (SHG) spectroscopy, have been extensively applied to the investigation of the molecular adsorption at electrode surfaces in solution,[51] they usually can provide only averaged information on the molecular orientation and packing within an adlayer. Ex situ techniques such as LEED and AES, using the UHVEC technique, have also been extensively employed for generating understanding of the relationship between the adsorbed molecules and the atomic structure of the electrode surfaces.[1,2] More recently, in situ STM has been well recognized as an important in situ method for structural investigation of adsorbed chemical species on well-defined electrode surfaces in electrolyte solution with atomic resolution.

Although small inorganic species, such as halide, sulfate, cyanide, and thiocyanate, adsorbed on the metal electrode surface can be visualized relatively easily by in situ STM as described already, high-resolution STM images have rarely been reported for organic molecules adsorbed on bare metal surfaces.

In the previous section, it was demonstrated that highly ordered molecular arrays of porphyrine, crystal violet, and a linear aromatic molecule, PPV, were easily formed on iodine-modified Au(111) rather than on bare Au, and they were visualized in solution by in situ STM with near-atomic resolution, revealing packing arrangements and even internal molecular structures. Such an extraordinarily high resolution achieved in solution strongly encouraged us to investigate the adsorption of relatively small organic molecules such as benzene directly attached to the electrode surface in order to understand electrocatalytic activities of noble metals such as Pt and Rh.

On the other hand, many reports have describe the investigation of the adsorption in UHV of aromatics, such as benzene and its derivatives on Pt, Rh, Ni, Ir, Ru, and Pd, which were performed by using various surface-sensitive techniques such as LEED, AES, and electron energy-loss spectroscopy (EELS). The purpose of those investigations was to evaluate gas-phase catalytic

reactions, such as hydrogenation, dehydrogenation, and dehydrocyclization.[60] In UHV the (3×3) superlattice of benzene and CO coadsorbed on Rh(111) revealed a well-ordered array of ring-like features associated with adsorbed benzene molecules, whereas CO did not appear in STM images.[61] A more detailed STM study has been reported by Somorjai et al.[62]

Nevertheless, we described, for the first time, the adlayer structures of benzene adsorbed on Rh(111) and Pt(111) in HF solutions.[63] High-resolution STM images allowed us to determine the packing arrangement and even the internal structure of each benzene molecule in solution.

6.1 Benzene on Rhodium(111) in Hydrofluoric Acid

Figures 17(a) and (b) show CVs of Rh(111) and Pt(111) electrodes in the absence and presence of benzene in 0.01 M HF, respectively. In the absence of benzene, the CV obtained on the well-defined Rh(111) and Pt(111) exhibited several highly reversible characteristic peaks. It was noted that the heights and widths of these characteristic peaks depended on the quality of the surface of Rh(111) prepared by the flame-annealing and quenching method.

After the Rh(111) electrode had been subjected to the CV measurement in the pure HF solution, the electrode was transferred into 0.01 M HF solution containing ca. 1 mM benzene. The CV indicated a featureless double-layer region between 0.3 and 0.7 V, as shown in Figure 17(a). The cathodic current commencing at about 0.3 V was considered to be due to simultaneously occurring processes, such as the desorption of adsorbed benzene, the adsorption of hydrogen, and the irreversible hydrogenation of benzene to cyclohexane, according to previous studies using differential electrochemical mass

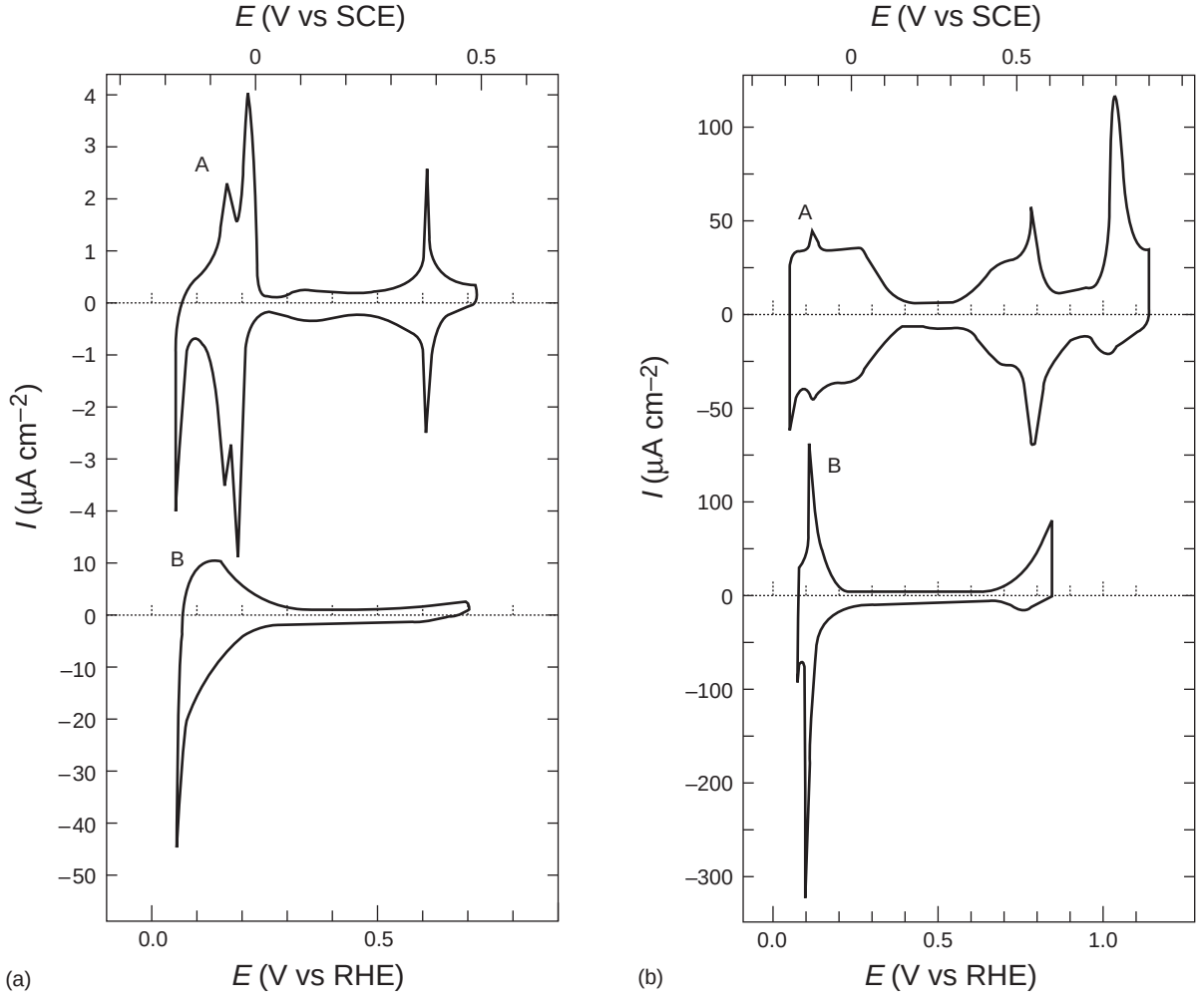

Figure 17 CV of (a) Rh(111) and (b) Pt(111), (A) without and (B) with 1 mM benzene. (Reprinted with permission from S.-L. Yau, Y.-G. Kim, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Benzene Adsorbed on Rh(111) and Pt(111) in HF Solution', *J. Am. Chem. Soc.*, **118**, 7795–7803,[63] Copyright (1996), American Chemical Society.)

spectrometry (DEMS).[64] A similar featureless CV was also obtained with a benzene-dosed Rh(111) electrode in pure 0.01 M HF. The Rh(111) electrode was immersed in 0.01 M HF containing 1 mM benzene for 1 min at the open-circuit potential (OCP) and then transferred to the pure HF solution. These results strongly suggest that benzene is chemisorbed and remains on the surface of Rh(111), at least in the potential range between 0.3 and 0.7 V.

A well-defined terrace-step structure was easily observed on the well-prepared Rh(111) face. The atomic image of Rh(111–1 × 1) was routinely discerned on the terrace in the pure HF solution. The almost perfectly aligned hexagonal structure can be seen with an interatomic distance of 0.27 nm, indicating that the structure of the Rh(111) surface is (1 × 1). Identical atomic images were consistently observed in the potential range between 0.1 and 0.75 V. No additional species were found in STM images at potentials corresponding to the butterfly peaks or the hydrogen adsorption and desorption peaks.[63]

After achieving the atomic resolution, a small amount of 1 mM benzene solution was directly added to the STM cell at 0.45 V. The average concentration of benzene in 0.01 M HF was 10 μM. Immediately after the injection of benzene, completely different patterns appeared in STM images. Figure 18(a) shows an example of the STM images acquired at 0.45 V. It is evident that the atomically flat terraces are now covered by ordered benzene adlayers. An averaged domain size was about 10 × 10 nm. The adsorbed benzene molecules appear to form a square adlattice in each domain. Furthermore, the molecular rows in a given domain cross each other, forming boundaries at an angle of either 60° or 120°.

More details of the orientation of benzene in the adlayer are revealed by the higher-resolution STM image shown in Figure 18(b). The acquisition of the STM image was performed specifically under conditions with minimal thermal drift in the *x* and *y* directions in order to determine the unit cell of the adlayer as accurately as possible. It is seen in Figure 18(b) that the molecular rows along the direction of arrows A and B cross each other at 90°, and they are always parallel with the close-packed and $\sqrt{3}$ directions of the Rh(111) substrate, respectively. The intermolecular distances along these directions are not equal to each other and were found to be, on average, 0.8 and 0.9 nm, respectively. Based on the orientation of molecular rows and the intermolecular distances, we concluded that the benzene adlayer was composed of rectangular unit cells, namely $c(2\sqrt{3} \times 3)$*rect* ($\theta = 0.17$), as shown in Figure 18(b). The known lattice spaces of $2\sqrt{3}$ and 3 on Rh(111) (0.268 nm) correspond to 0.93 and 0.80 nm, respectively, which are consistent with our experimental values.

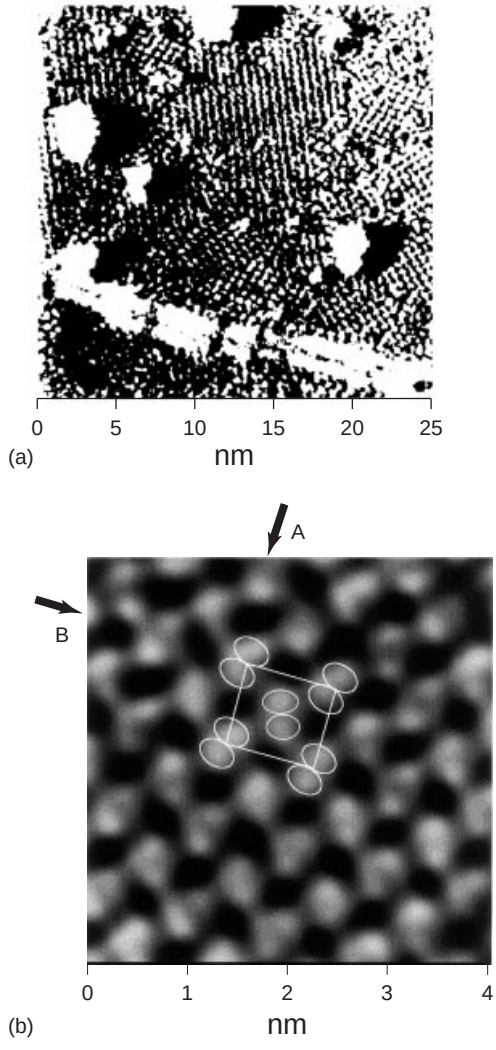

Figure 18 High-resolution STM images of the $c(2\sqrt{3} \times 3)$*rect* benzene adlayer in HF. (Reprinted with permission from S.-L. Yau, Y.-G. Kim, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Benzene Adsorbed on Rh(111) and Pt(111) in HF Solution', *J. Am. Chem. Soc.*, **118**, 7795–7803,[63] Copyright (1996), American Chemical Society.)

Surprisingly, the STM image allowed us to determine the internal structure and micro-orientation of each benzene molecule adsorbed on Rh(111). It is clear that each spot is split into two bright spots, forming a characteristic dumbbell shape for each benzene molecule. The STM discerned a 0.01 nm corrugation between the valley and ridge of each benzene molecule. It can also be seen in Figure 18(b) that the orientation of dumbbell shaped benzene is not the same for all molecules, but depends on their positions. The dumbbell-shape of the central benzene molecule in the unit cell shown in Figure 18(b) is clearly rotated by 60° with respect to the molecules located on the four corners of the unit cell. The molecules on the corners of the unit cell

appeared with an identical feature, suggesting that they are situated on an identical binding site. It is also seen that the orientation of these dumbbells is always rotated by 30° with respect to the direction of close-packed rows [arrow A in Figure 18(b)] of the Rh(111) substrate. The STM image shown in Figure 18(b) provides more detailed information on the orientation of molecule in the unit cell as discussed below.

The $c(2\sqrt{3} \times 3)$*rect* structure described above was consistently observed in the potential range between 0.4 and 0.7 V without additional structural transitions. On the other hand, it was found that the adlayer structure changed at negative potentials. A negative potential step from 0.45 to 0.35 V induced a reconstruction in the benzene adlayer from $c(2\sqrt{3} \times 3)$*rect* symmetry to an ordered hexagonal pattern. The electrode potential of 0.35 V is near the onset potential of the cathodic current as shown in Figure 17(a).

Figures 19(a) and (b) show a set of STM images acquired in almost the same area in order to reveal the dynamic process of phase transition. It is clearly seen in Figure 19(a) that a new domain appeared with the hexagonal array of benzene on the upper-right corner marked by solid lines, while the $c(2\sqrt{3} \times 3)$*rect* structure remained as the main phase. A further cathodic step to 0.25 V resulted in a predominantly hexagonal phase, while eliminating the $c(2\sqrt{3} \times 3)$*rect* domains as shown in Figure 19(b). Such a long-range ordered hexagonal pattern could be seen over almost the entire area of the terrace at 0.25 V. All benzene molecules exhibited the same corrugation height of 0.07 nm, similar to that in the $c(2\sqrt{3} \times 3)$*rect* structure.

To reveal the internal molecular structure in the hexagonal phase, STM images were acquired under particularly carefully adjusted experimental conditions with minimal thermal drift. Figure 20(a) shows one of the highest resolution images acquired on the terrace shown in Figure 19(b). Compared with the crystal orientation, $[1\bar{1}0]$, determined by the Rh(111–1 × 1) atomic image, it can be seen that all benzene molecules are almost perfectly aligned along three close-packed directions of Rh(111). The molecular rows cross each other at an angle of either 60° or 120° within experimental error (±2°). The intermolecular distance along these rows was found to be 0.8 nm, which corresponds to three times the lattice parameter of Rh(111). Therefore, we conclude that the hexagonal structure is (3×3)-C_6H_6 ($\theta = 0.11$), as shown by the unit cell superimposed in Figure 20(a). Moreover, a careful examination of the image reveals that each benzene molecule appears as a set of three spots with similar intensities. It can also be seen that a clear dip exists in the center of each triangle with three lobes. These features can be more clearly seen in the height-shaded surface plot obtained by applying a mild 2D Fourier transform filter method as shown in Figure 20(b). The spacing between the two lobes in each molecule was found to be about 0.3 nm. In addition, a weaker additional spot with a small corrugation of about 0.02 nm can be seen in the unit cell. It is important to note that all these features of the STM image for the (3×3) adlayer observed at 0.25 V are essentially identical with those found for the coadsorbed benzene and CO adlayer on Rh(111) in UHV reported by Ohtani et al.[61] They also found the weaker spot, which was attributed to the coadsorbed CO or artifacts caused by asymmetric tips.[61]

When the electrode potential was stepped further in the negative direction, the ordered (3×3) domain became

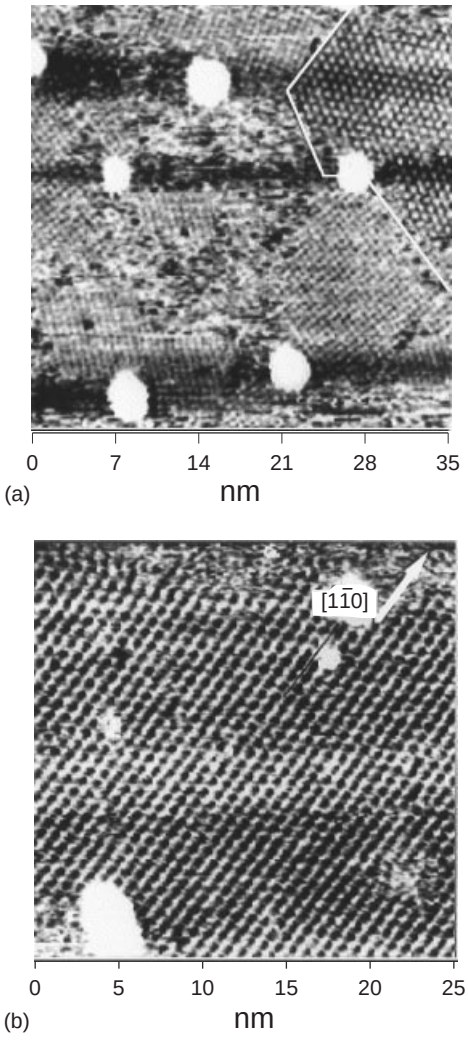

Figure 19 STM images of (a) domain boundaries of $c(2\sqrt{3} \times 3)$ *rect* and (3×3) benzene adlayers and (b) the pure (3×3) structure on Rh(111). (Reprinted with permission from S.-L. Yau, Y.-G. Kim, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Benzene Adsorbed on Rh(111) and Pt(111) in HF Solution', *J. Am. Chem. Soc.*, **118**, 7795–7803,[63] Copyright (1996), American Chemical Society.)

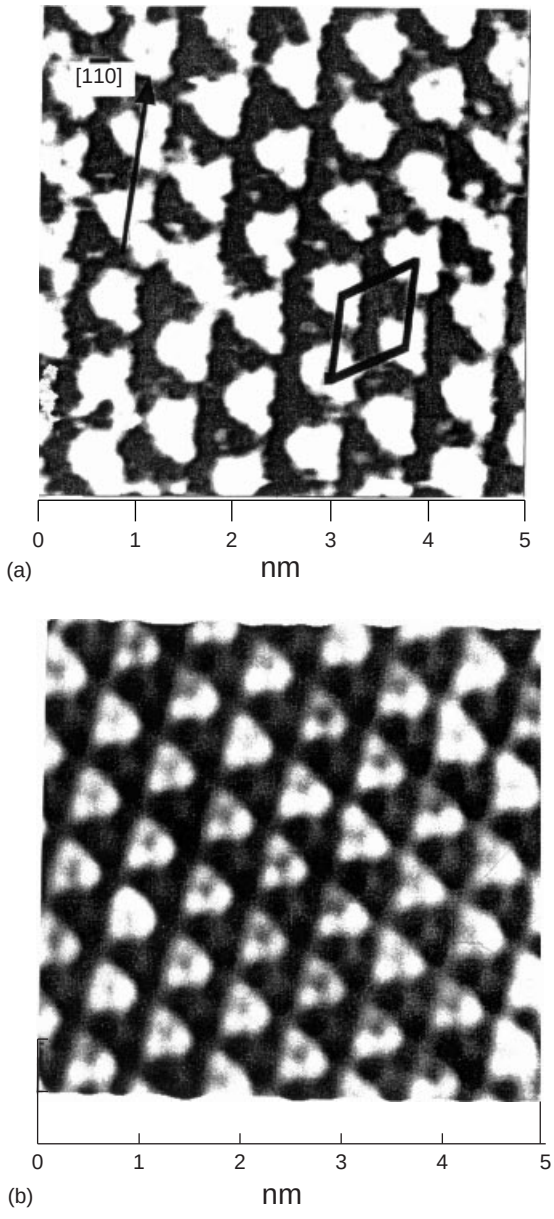

Figure 20 High-resolution STM images of the (3×3) structure on Rh(111). (a) Top view; (b) height-shaded plot. (Reprinted with permission from S.-L. Yau, Y.-G. Kim, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Benzene Adsorbed on Rh(111) and Pt(111) in HF Solution', *J. Am. Chem. Soc.*, **118**, 7795–7803,[63] Copyright (1996), American Chemical Society.)

islands with the same internal structure, suggesting that the desorption of benzene occurred preferentially at edges of the islands of ordered (3×3) domains.[63] Eventually, all adsorbed benzene molecules were desorbed from the surface at 0.1 V owing partially to the hydrogen adsorption and partially to the hydrogenation as expected from the result obtained by DEMS,[64] and the Rh(111–1×1) structure was consistently discerned at 0.1 V. The structural changes described above were reversible. When the electrode potential was stepped back to the positive region, the (3×3) and $c(2\sqrt{3} \times 3)$*rect* phases returned at the potentials described above.

The structures and registries of chemisorbed benzene on Rh(111) have been thoroughly scrutinized by the surface-sensitive techniques, such as LEED, EELS, and angle-resolved ultraviolet photoemission spectroscopy (ARUPS), in UHV.[60] These previous studies revealed various structures for benzene, including well-known structures such as $c(2\sqrt{3} \times 4)$*rect* and (3×3), depending on whether CO was present unintentionally or intentionally in the UHV chambers. Although it has been repeatedly demonstrated that the adlayer structures of benzene on Rh and Pt were greatly affected by the presence of CO in the adlayer, the structure of the pure benzene adlayer is not yet fully understood. Neuber et al. have reported that a completely new structure with a $(\sqrt{19} \times \sqrt{19})R23.4^\circ$ symmetry appeared for the pure benzene adsorption on Rh(111) under cleaner UHV conditions in the absence of CO, and the previously known structures of $c(2\sqrt{3} \times 4)$*rect* and (3×3) were found to appear upon admission of CO.[65]

However, it was found in our study[63] that the anodic peak due to the oxidation of CO was hardly detectable in CV even after a prolonged STM experiment for several hours in an air-saturated HF solution. We strongly believe that the adlayer structures found in HF solution described above did not result from contamination with CO.

It is extremely important to recognize that in the previous study of the adsorbed benzene on Rh(111) in UHV,[65] one of the structures of the pure benzene adlayer was attributed to the $c(2\sqrt{3} \times 3)$*rect* structure, which was found in solution. This result strongly suggests that the existence of water molecule on top of the benzene adlayer plays a minor role in determining the structure of benzene.

Figure 21(a) shows a proposed model for the $c(2\sqrt{3} \times 3)$*rect* structure. All of the adsorbed benzene molecules are assumed to be located on the twofold bridging sites. The benzene molecule at the center of the unit cell also occupies a twofold site, but it is rotated by 60° from the orientation of the molecules at the corners. Weiss and Eigler reported three distinct types of STM images for isolated benzene molecules located at threefold hollow, atop, and bridge sites on Pt(111) at 4 K.[66] They assigned the single bump elongated perpendicularly to the bridge to the bridge-bonded benzene. In Figure 18(b) each benzene molecule is seen with the dumbbell shape on Rh(111) and elongated perpendicularly to the bridge. It is clear that the direction of each elongated dumbbell is always rotated by approximately 30° with respect to that of the corresponding atomic row of Rh(111).

These detailed features can be explained by the model structure shown in Figure 21(a), where two lobes marked

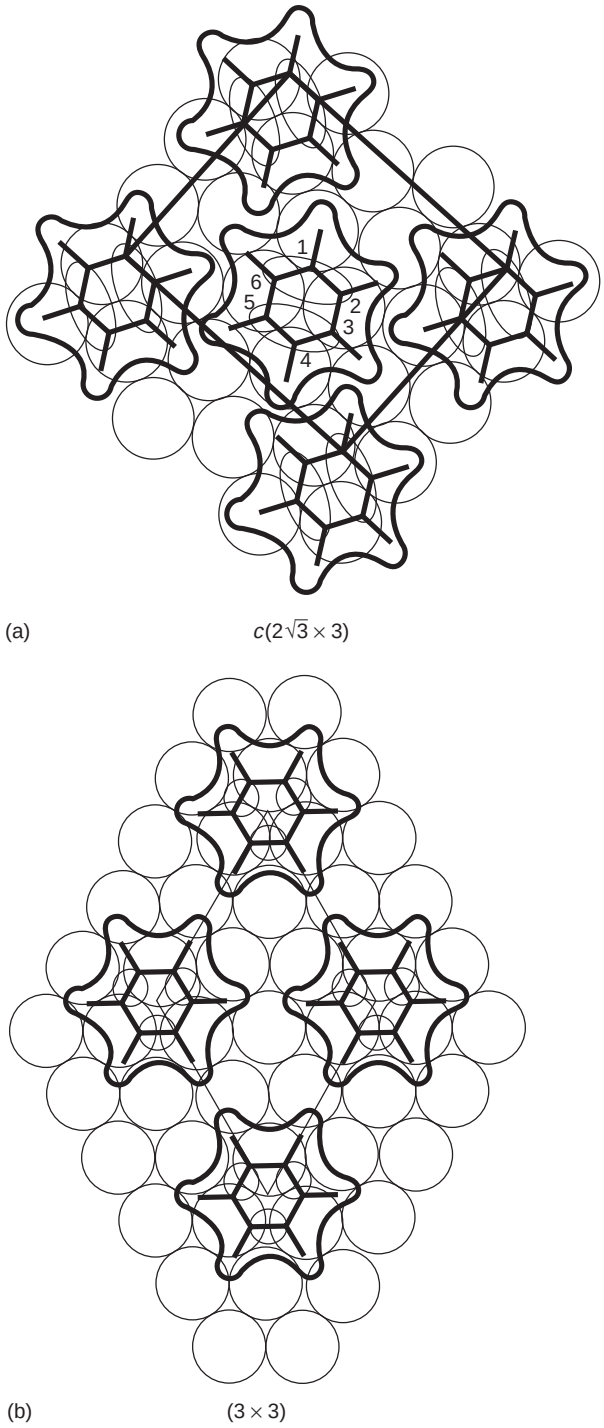

Figure 21 Space models of (a) the $c(2\sqrt{3} \times 3)$*rect* and (b) the (3×3) structures. (Reprinted with permission from S.-L. Yau, Y.-G. Kim, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Benzene Adsorbed on Rh(111) and Pt(111) in HF Solution', *J. Am. Chem. Soc.*, **118**, 7795–7803,[63] Copyright (1996), American Chemical Society.)

by the circles are assumed to be localized to near carbon atoms (1, 2, 6 and 3, 4, 5) bonded across the Rh atoms.

The STM image obtained at 0.25 V shown in Figure 20 can be explained by the structural model with the (3×3) symmetry illustrated in Figure 21(b). Although the structure proposed here is basically the same as that proposed previously, based on LEED, EELS,[60] and STM[61] studies in UHV, for the adlayer of coadsorbed benzene and CO on Rh(111), two CO molecules thought to be located at the threefold hollow sites in the unit cell are omitted in Figure 21(b). Each benzene molecule is assumed to bond at the threefold hollow site. The coadsorption of CO was unlikely to take place in the solution under the present conditions, because no oxidation peak was observed, as described above. Instead of CO, water molecules or hydronium cations might be coadsorbed near the uncoordinated threefold hollow sites to stabilize the (3×3) structure, their function being similar to that of the coadsorbed CO. The weak small spots seen in Figure 20 might be due to such coadsorbed water molecules or hydronium cations.

6.2 Other Molecules

The in situ STM imaging of benzene adlayers on Pt(111) was carried out in the same manner as that on Rh(111).[63] It was found in our study that benzene adlayers on Pt(111) mostly appeared as less ordered phases than those on Rh(111). The intermolecular distances and directions of molecular rows indicate that the structure of the benzene adlayer at 0.35 V is $c(2\sqrt{3} \times 3)$*rect*, the same as that found on Rh(111).

More interestingly, in situ STM revealed the reconstruction of the benzene adlayer that occurs upon the cathodic potential step from 0.35 to 0.25 V. On the basis of detailed results, we proposed that the benzene adlayer has a $(\sqrt{21} \times \sqrt{21})R10.9^\circ$ structure ($\theta = 0.14$).[63]

After the atomic resolution was achieved with Rh(111), a saturated naphthalene solution was added to the STM cell at 0.3 V.[67] A high-resolution image acquired in an ordered domain is shown in Figures 22(a) and (b). It is clearly seen that the molecular rows parallel the ⟨110⟩ direction of the substrate indicated by the arrows in Figure 22(a). More importantly, the STM image allowed us to determine the internal structure and orientation of each naphthalene molecule. The elongated features along the longer molecular axis (C_2) were discerned for each molecule. In addition, the images of some molecules clearly show a two-ring structure expected from the molecular model. It can be seen that naphthalene molecules are perfectly aligned with a regular micro-orientation along the molecular rows.

A typical arrangement of naphthalene molecules on Pt(111) was found to be a full monolayer of flat-lying naphthalene molecules. Although the overall appearance of the naphthalene adlayer on Pt(111) on the large

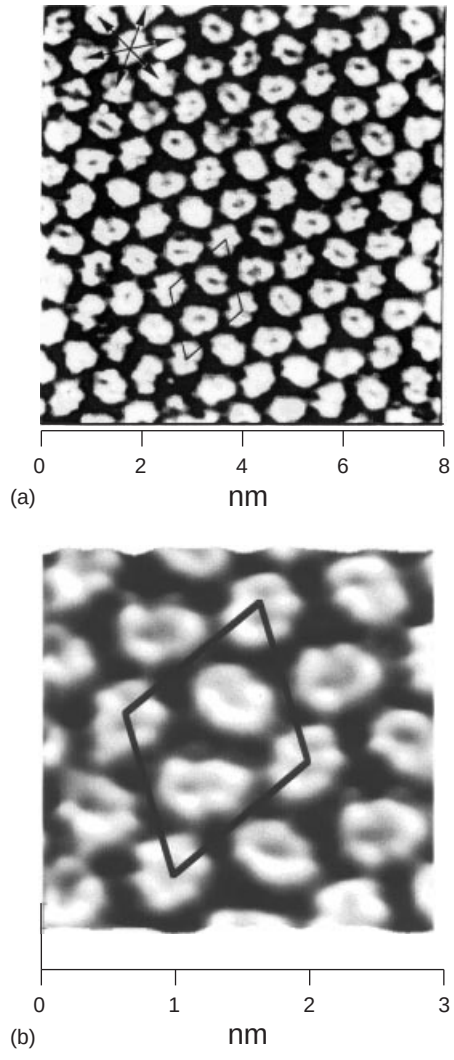

Figure 22 High-resolution STM images of naphthalene on Rh(111). (a) Top view; (b) perspective view. (Reprinted with permission from S.-L. Yau, Y.-G. Kim, K. Itaya, 'High-resolution Imaging of Aromatic Molecules Adsorbed on Rh(111) and Pt(111) in Hydrofluoric Acid Solution: In Situ STM Study', *J.Phys. Chem. B*, **101**, 3547–3553,[67] Copyright (1997) American Chemical Society.)

scale was similar to that on Rh(111), close inspection revealed that the adlayer included many randomly oriented molecules. Because the molecular rows were nearly parallel to the close-packed Pt atomic rows, and because the intermolecular distance was three times that of the Pt substrate, the structure roughly fitted a (3×3) symmetry.

Periodic rotation of the molecules of naphthalene by 60° is seen within each molecular row with every third molecule being in the same orientation. A further magnified view in height-shaded mode is shown in Figure 22(b), in which the two-ring structure can be more clearly seen. The nearest-neighbor distance of 0.82 nm on average is equivalent to three times the Rh lattice parameter of 0.268 nm.

According to the results described above, the unit cell can be defined as a $(3\sqrt{3} \times 3\sqrt{3})R30^\circ$ symmetry as shown in Figure 22(b).[67] It is now clear that all naphthalene molecules align their C_2 axes along the close-packed directions of Rh substrate. The molecules aligned along the $\langle 112 \rangle$ direction, which is the so-called $\sqrt{3}$ direction, have the same orientation. The spacing between two adjacent molecules along the $\sqrt{3}$ direction is measured to be 1.4 nm, which is three times the $\sqrt{3}$ spacing. In our model structure,[67] two carbon atoms at the 9- and 10-positions are assumed to be attached directly to a Rh atom. It is noteworthy that this structure is identical with that previously proposed from LEED results by UHV workers.[68] If one recalls the identical results for benzene adsorbed on Rh(111) in UHV and in HF solution as described above, the results obtained with naphthalene further support the predominant adsorbate–substrate interaction for hydrophobic molecules and the minor role of water molecules.

1,2- and 1,4-naphthoquinone can be considered as representative derivatives of naphthalene in order to understand the effect of functional groups on the molecular organization.[67] Generally, these quinones, similarly to naphthalene, formed ordered adlayers on Rh(111) and mostly disordered adlayers with the flat-lying orientation on Pt(111). The high-resolution STM images revealed the details of internal molecular structures. It was demonstrated that 1,2-naphthoquinone formed a well-ordered adlayer with the structure $(3\sqrt{3} \times 3\sqrt{3})R30^\circ$, which was identical with that found for naphthalene. More interestingly, it was found that an additional bright spot, unseen in the image for naphthalene, is seen at the 2-position of each 1,2-naphthoquinone molecule.[67] This spot exhibits a ca. 0.03 nm higher corrugation with respect to the naphthalene ring, which is likely to be due to the oxygen at the 2-position. The oxygen at the 1-position was not clearly seen in the STM images.

An STM image of anthracene on Rh(111) unambiguously disclosed the internal molecular structure of anthracene.[67] It was also found that molecules of anthraquinone and 1,4,9,10-anthracenetetrol adsorbed in a manner similar to those of anthracene.

To elucidate the effect of molecular structure on the packing arrangement, biphenyl was further investigated on Rh(111).[69] In contrast to the planar structure of naphthalene and anthracene, the two aromatic rings of biphenyl are slightly off the coplanar configuration because of the restriction of hydrogen atoms at the 2,2′- and 6,6′-positions. However, it was found that biphenyl formed disordered adlayers on Rh(111) in HF.[69] Although STM images revealed the internal molecular structure of biphenyl with clear identification of

two rings in each molecule, the nonplanar configuration was not clearly seen. We expected that two rings of biphenyl behave like two benzene molecules, and they prefer to be attached on the bridge sites as benzene molecules do.[69] The appearance of the disordered adlayer of biphenyl suggests that biphenyl is more strongly attached on Rh(111) than naphthalene.

It is of special interest to distinguish structures among a series of benzene derivatives such as phthalic acid, terephthalic acid, and hydroquinone.

7 ELECTROCHEMICAL DISSOLUTION PROCESSES OF SEMICONDUCTORS

The preparation of clean and stable semiconductor surfaces is the first step in the manufacture of semiconductor devices. The drive towards submicron technology for ultralarge-scale integrated circuits has focused special attention on wet chemical processes, since high-temperature-based procedures often lead to adverse effects that arise from new and difficult-to-control reaction channels; the pursuit of nanometer-scale technology has likewise necessitated the development of surface characterization methods that permit atomic-scale resolution. Since commercial integrated circuits are still based exclusively on silicon, the wet-chemical processing of Si single-crystal surfaces has been and continues to be widely investigated; much of the interest centers around the nature of the hydrogen termination of the Si surface atoms.

The first work to demonstrate that an Si(111) surface etched in aqueous NH_4F is ideally terminated with Si monohydride was based upon IR spectroscopic studies of the Si–H vibrational modes.[70] This original work stimulated further investigations, likewise based upon IR and other spectroscopic techniques, on Si(100) and Si(110) substrates. STM in vacuum directly established the fact that the NH_4F-etched Si(111) surface was atomically well defined with an ideal H-terminated $Si(111):H-(1\times1)$ structure.[71,72] While these studies provided critical information on the post-etched Si surfaces, it was clear that in situ investigations had to be undertaken if the chemical etching process is to be understood at the atomic level.

In response to this need, in situ STM and AFM were adopted for semiconductor-etching studies. Our first in situ STM observation[73] of the $Si(111):H-(1\times1)$ atomic structure in a noncorrosive solution (aqueous H_2SO_4) spurred investigation of the etching of Si(111) in corrosive solutions, such as aqueous NH_4F,[74–76] and NaOH.[77–79]

It was soon determined that the etching of Si(111) was potential dependent. At potentials markedly negative of the open-circuited potential (OCP), the etching rate decreased and the dissolution proceeded via a step-selective layer-by-layer mechanism.[74–78] At potentials near or more positive than the OCP, the etching rate increased and pit corrosion occurred on the terraces, which resulted in atomically roughened surfaces.[77,78] The potential dependence of the Si-etching process is a technologically relevant issue. Industrial wet-etching processes are usually performed without potential control, hence the surface chemical reactions that characterize the simple dipping of Si wafers in an etching bath are expected to bear a strong resemblance to what is observed from EC etching at the OCP. The ability to control the applied potential during the etching process may serve as an additional critical factor in the preparation of atomically well-defined semiconductor surfaces.

We have also demonstrated that in situ STM can be employed to monitor atomic-scale features of the etching process.[74–76] For example, we discovered that, in general, multiple H-terminated Si atoms at the kink and step sites were eroded more rapidly than the monohydride-capped atoms.

In this section, we first describe detailed etching processes of Si(111) in NH_4F solutions. Atomic images of GaAs and InP surfaces are also briefly discussed.

7.1 Silicon(111)

The STM imaging of the n-Si(111) electrode was performed immediately after the electrode was etched. After the Si electrode had been immersed in the NH_4F solution, the electrode potential of Si was immediately set to −1.1 V. Note that the potential effect on the etching rate of Si was reported in NaOH solutions.[77,78] It was found that both chemical and EC etching mechanisms operate at the OCP, leading to a higher etching rate.[78,79] We also found that applying a cathodic potential of −1.1 V drastically reduced the corrosion rate of Si in NH_4F solution.[74–76]

Figures 23(a) and (b) show a crystallographic orientation and a ball-and-stick model, respectively, of an H-terminated Si(111) with $[11\bar{2}]$ and $[\bar{1}\bar{1}2]$ oriented steps. One of the most interesting features of an Si(111) surface is the existence of two structurally different steps where the Si atoms have monohydride and dihydride configurations. They can be exemplified by the steps in the $[11\bar{2}]$ and $[\bar{1}\bar{1}2]$ directions, respectively (Figure 23b). Ex situ STM results have shown that the dihydride-bound Si atoms are more reactive than those of the monohydride ones in weakly alkaline HF solutions, resulting in the appearance of the most stable $[11\bar{2}]$ steps.[72] Those experiments were conducted with Si(111) samples tilted towards $[11\bar{2}]$ and $[\bar{1}\bar{1}2]$.[72] Furthermore, there are two possible dihydride structures for the Si atoms at the $[\bar{1}\bar{1}2]$ step. The dihydride axis is either perpendicular or parallel to the (111) plane.

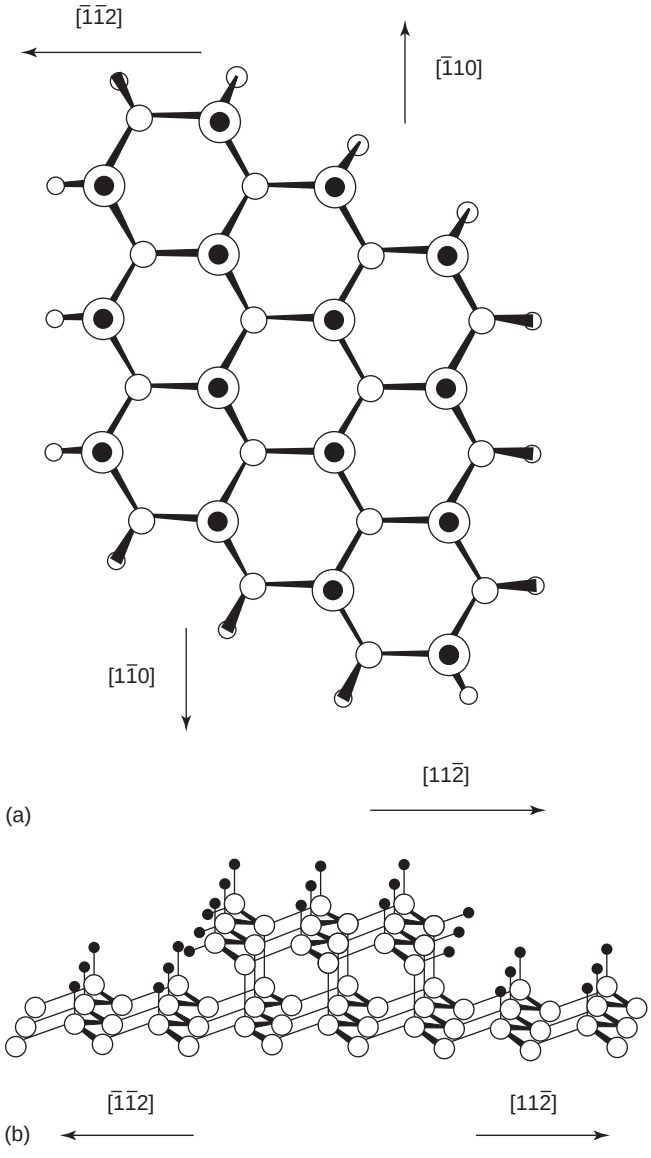

Figure 23 (a) Top view and (b) side view of the Si(111) surface. (Reprinted with permission from K. Kaji, S.-L. Yau, K. Itaya, 'Atomic Scale Etching Processes of n-Si(111) in NH_4F Solutions: In Situ Scanning Tunneling Microscopy', *J. Appl. Phys.*, **78**, 5727–5733,[75] (1995).)

It is believed that there is a strong repulsive interaction among the horizontal dihydride structures, so that the perpendicular dihydride, as depicted in Figure 23(b), is in fact more stable than the horizontal one, as confirmed by an IR spectroscopic study.[80] The horizontal dihydride[72] might be too reactive to exist on an Si(111) surface in the presence of etching species of H_2O and F^-.

We focused our attention on the evaluation of the reactivity difference between the microscopically different steps on Si(111).[75] We were able to use in situ STM to locate some areas which contain both types of steps so that the reactivity of these steps can be simultaneously examined under identical conditions. This approach is evidently more advantageous than that of the previous study which used differently tilted Si(111) substrates in separated experiments.[72] As the initial surface feature of Si controlled the subsequent etching process, we first recorded an STM image at −1.1 V to show the initial surface morphology as shown in Figure 24(a), followed by stepping the electrode potential to a less negative value of −1.04 V to accelerate the erosion of Si. A series of STM images shown in Figures 24(b–f) was acquired successively with a time interval of 13 s.[75]

Figure 24(a) shows well-defined double-layer steps of 0.32 nm in height and terraces extending more than 25 nm on the (111) surface. The Si(111) was etched in 11 M NH_4F for 3 min at room temperature. The relative heights of terraces are reflected by their brightness in the STM image. The internal atomic structure of terrace (marked T) was readily discerned by a high-resolution STM scan. A well-ordered hexagonal pattern with an interatomic spacing of 0.38 nm was in good agreement with the ideal Si(111) : H–1 × 1 structure.[73] Consequently, the treatment in 40% NH_4F yielded a long-range ordered monohydride-terminated Si(111) surface with no discernable vacancy defect in the hexagonal network. The step orientations, as defined by their outward normals, are shown in Figure 24(a). It is important to note that both the mono- and dihydride steps were probed by the in situ STM imaging at the same time. The shape of the terraces shown in Figure 24(a) was determined by the morphology of the step ledges, i.e. the monohydride steps are mostly straight, in strong contrast to the typical zig-zag pattern for the dihydride ones.

The small islands (3 nm in diameter, probably impurities) at the upper edge of the STM image were used as a guide against themal drift during the STM measurement. Their unchanged locations indicate low thermal drift. The time-dependent STM results presented below demonstrate the important role played by the atomic structure at the steps in controlling the etching rate.

Figures 24(b–f) present the time-dependent etching process of Si(111) after acquiring the image of Figure 24(a). During the first 13 s etching of the Si from Figures 24(a) to (b), the width of the upper portion of the terrace T marked by D decreased from 16 to 8 nm, while the lower portion marked by D′ retracted from 18 to 12.5 nm. The relatively faster erosion of the upper half of the [$\bar{1}\bar{1}$2] dihydride step is thought to be due to the higher kink density within the zig-zag pattern step ledge. On the other hand, the monohydride step in the direction of [11$\bar{2}$] seems to be unchanged from Figures 24(a) to (b). In particular, the steps within the circle marked C in Figure 24(b) remained still in both images, indicating that the ideal monohydride [11$\bar{2}$] and [$\bar{2}$11] step in the absence of kink sites is inactive under

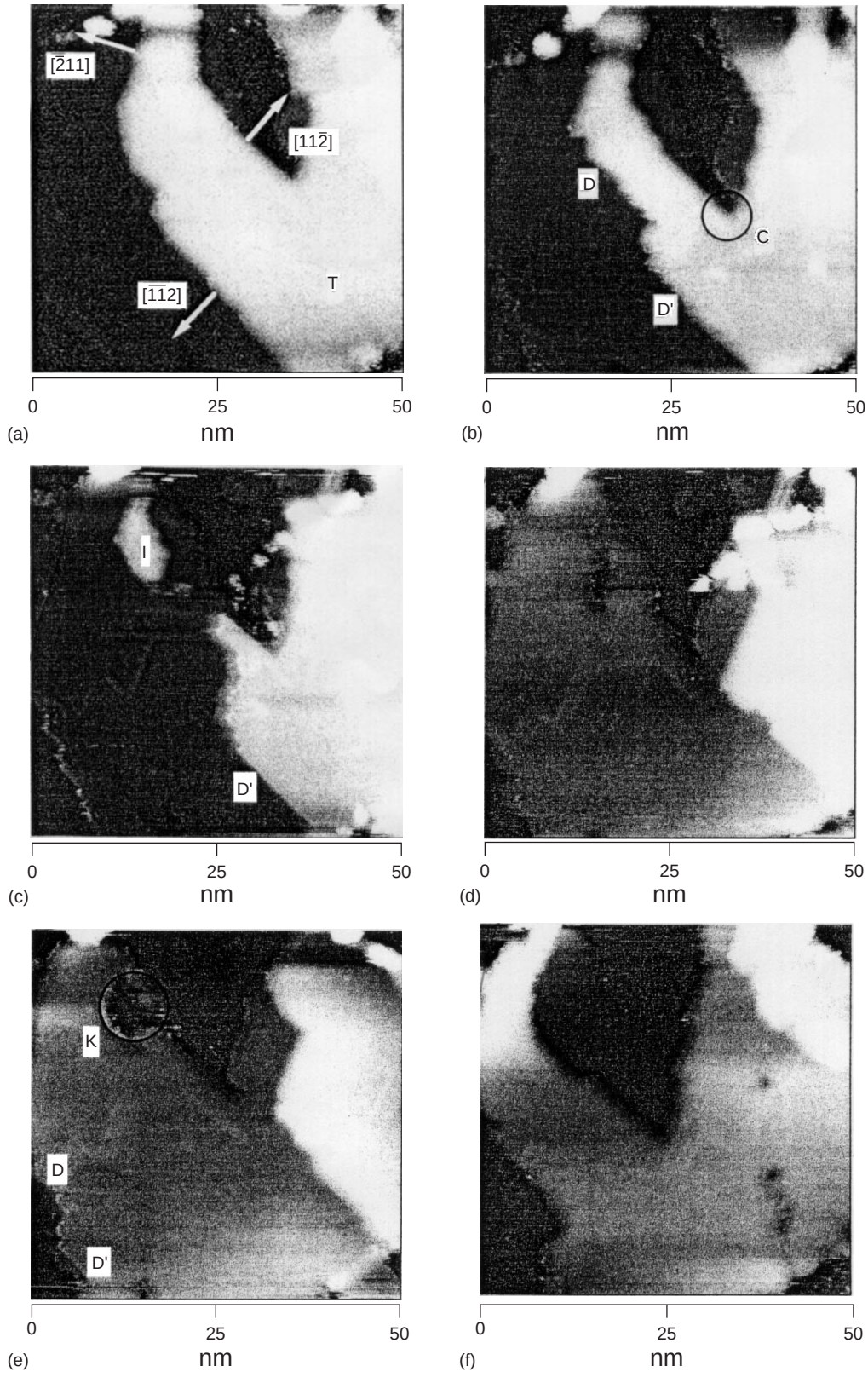

Figure 24 Successively recorded STM images for the etching process of Si(111) in NH_4F. The images were acquired at time intervals of 12.8 s. (Reprinted with permission from K. Kaji, S.-L. Yau, K. Itaya, 'Atomic Scale Etching Processes of n-Si(111) in NH_4F Solutions: In Situ Scanning Tunneling Microscopy', *J. Appl. Phys.*, **78**, 5727–5733,[75] (1995).)

the present etching conditions. The dihydride-terminated steps continued to retract rapidly to dissolve the terrace T, leaving a small isolated island I in Figure 24(c). However, it was surprising to find that the $[\bar{1}\bar{1}2]$ step with the dihydride configuration marked by D′ in Figure 24(c) was essentially unchanged from Figures 24(b) to (c). The well-defined step ledge, marked D′ in Figure 24(c), suggests that an ideal dihydride step without kinks is also stable.

After the complete removal of the island I in Figure 24(c), a small, bilayer deep (0.32 nm) pit evolved in Figure 24(d). This newly formed depression is apparently a real pit, not an imaging artifact, because it expanded and coalesced with an adjacent $[11\bar{2}]$ step, as shown in Figure 24(e). This coalescence introduced many kink sites (marked K) in Figure 24(e) at the almost inactive $[11\bar{2}]$ step. A close examination of Figures 24(e) and (f) reveals the very important fact that the etching rate of the $[11\bar{2}]$ monohydride step is now increased by the introduction of the kinks into the nearly ideal monohydride step. The $[11\bar{2}]$ step line retracted by ca. 3 nm within 13 s, which corresponds to an etching rate of 14 nm min^{-1}. This is a significant increase from a negligible level for the ideal monohydride step as observed in Figures 24(a) and (b).

The aforementioned results provide compelling evidence that the difference in the chemical reactivity of the monohydrogen-terminated and dihydrogen-capped Si surface atoms profoundly influences the Si-etching process. We extended our investigations of the EC etching of Si(110) and Si(100).[74,76]

7.2 Other Semiconductors

It is becoming more urgent for workers in semiconductor technology to understand wet-chemical etching processes, particularly those of Si, GaAs, and InP with atomic resolution.

We acquired the first atomic STM images of GaAs surfaces in 0.05 M H_2SO_4 solution.[81–83] The results clearly demonstrate that the ideal GaAs(001–1 × 1) and (111–1 × 1) structures exist in a pure H_2SO_4 solution in a cathodic potential region. The samples were n-type Si-doped GaAs(001), (111)A and (111)B wafers, grown by the horizontal Bridgman method. The GaAs(001) and (111)B samples were treated in 1 M HCl for 10 min at room temperature. The GaAs(111)A sample was etched with an etching solution [H_2SO_4–H_2O_2–H_2O (1 : 8 : 1 by volume)]. The etching rate of the (111)A surface has been reported to be the lowest among all low-index planes in this mixed solution. After the etching, the solution was completely replaced with 0.05 M H_2SO_4. The replacement of the etching solution was carried out repeatedly to exclude HCl in the solution. It is important that the GaAs surface should always be kept submerged in the solution in order to protect it from oxidation and contamination in the ambient atmosphere.

Figure 25(a) shows a typical surface topography of a chemically etched GaAs(001) surface acquired in an area of 50 × 50 nm. It can be clearly seen that the surface of the (001) exhibits a well-defined step-terrace structure extending over a large area. Wider terraces are seen to extend over 30 nm. The relative brightness of the terraces in the STM image reflects their heights on the surface, i.e. the surface ascends from left to right. The rather uniform appearance of the terraces strongly suggests that the (001) surface has a structure that is well defined on an atomic scale. It is also clear that the steps intersect each other to form an angle of 90°, as expected for a surface with fourfold symmetry. These steps appearing as straight lines were confirmed as double-layer steps on the (001) surface based on the observed height of 0.28 nm obtained by a cross-section analysis. This unique height of 0.28 nm for the steps indicates straightforwardly that the (001) surface prepared by etching in HCl must be either Ga- or As-terminated. According to the crystallographic orientation of the GaAs(001) electrode, these steps were found to be parallel to either [110] or $[\bar{1}10]$ directions.

Figure 25(b) shows our first atomic STM image of an atomically flat terrace on a GaAs(001) surface. It is clear that the ideal square arrangement expected for the (001) surface with fourfold symmetry is discerned by in situ STM. The observed nearest interatomic distances in the [110] and $[\bar{1}10]$ directions were found to be 0.4 ± 0.02 nm. The atomic image shown in Figure 25(b) clearly demonstrates that the ideal, nonreconstructed GaAs(001–1 × 1) structure is exposed in H_2SO_4 solution under the cathodic polarization. Note that the ideal (1 × 1) structure seemed to be extended over the entire region of the terrace, because pits or even single atomic defects were rarely observed.

The (111)A surface etched in the mixed solution containing H_2O_2 was also found to have an atomically flat terrace-step structure in H_2SO_4 solution as shown in Figure 26(a). All steps observed were double-layer steps with a height of 0.33 nm. The steps in local areas were straight and parallel to the close-packed atomic row direction of (111) surface. Figure 26(b) shows a typical atomic STM image, revealing an interatomic distance of 0.4 nm with an almost perfect hexagonal arrangement. This result clearly demonstrates that the ideal GaAs(111)A–(1 × 1) structure, as shown in Figure 27(c), is exposed in H_2SO_4 solution. It is reasonably expected that the uppermost layer on the (111)A surface consists of Ga atoms.

Finally, it is noteworthy that the GaAs(111)B surface prepared by etching in 1 M HCl also showed an atomically flat terrace-step structure in H_2SO_4 with a step height identical with that observed on the (111)A

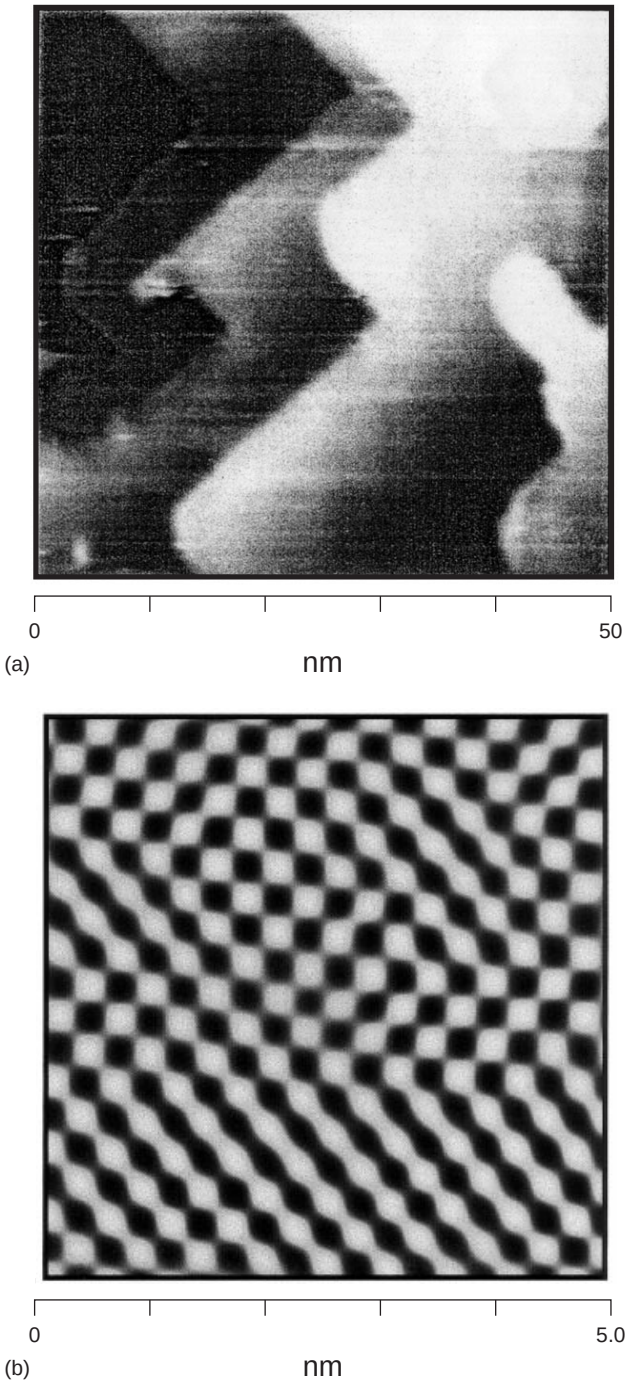

Figure 25 (a) Topographic and (b) high-resolution STM images of n-GaAs(001) in H_2SO_4. (Reprinted with permission from H. Yao, S.-L. Yau, K. Itaya, 'In Situ Scanning Tunneling Microscopy of GaAs(001), (111)A, and (111)B Surfaces in Sulfuric Acid Solution', *Appl. Phys. Lett.*, **68**, 1473–1475[82] (1996).)

surface. Although the average terrace width was typically 5–10 nm, obviously smaller than that found on the A surface, an atomic STM image revealed a hexagonal

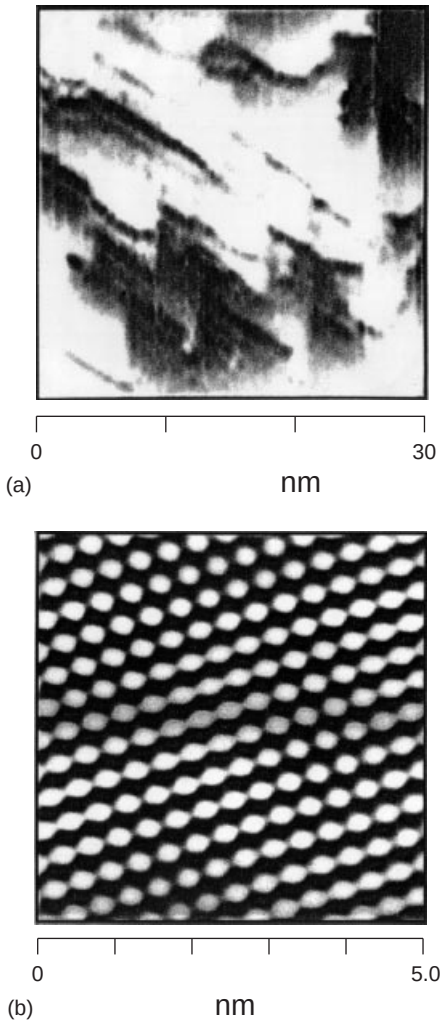

Figure 26 STM images of n-GaAs(111)A in 0.05 M H_2SO_4. (Reprinted with permission from H. Yao, S.-L. Yau, K. Itaya, 'In Situ Scanning Tunneling Microscopy of GaAs(001), (111)A, and (111)B Surfaces in Sulfuric Acid Solution', *Appl. Phys. Lett.*, **68**, 1473–1475[82] (1996).)

arrangement of As atoms with an interatomic distance of 0.4 nm. The above results indicate that the GaAs(111)B surface has also the ideal (1×1) structure shown in Figure 27(d).[82]

In summary, it was demonstrated that the well-defined GaAs(001), (111)A, and (111)B surfaces can be prepared by chemical etching in solutions. Atomically flat terrace-step structures were consistently observed by in situ STM on all three surfaces in H_2SO_4 solution under potential control. Furthermore, we successfully obtained the first atomic STM images which showed that the ideal GaAs(001–1×1), GaAs(111)A–(1×1), and GaAs(111)B–(1×1) structures are exposed and persist in H_2SO_4 solution.

In spite of the fact that InP is a very important material for both optoelectronic and electronic device applications,

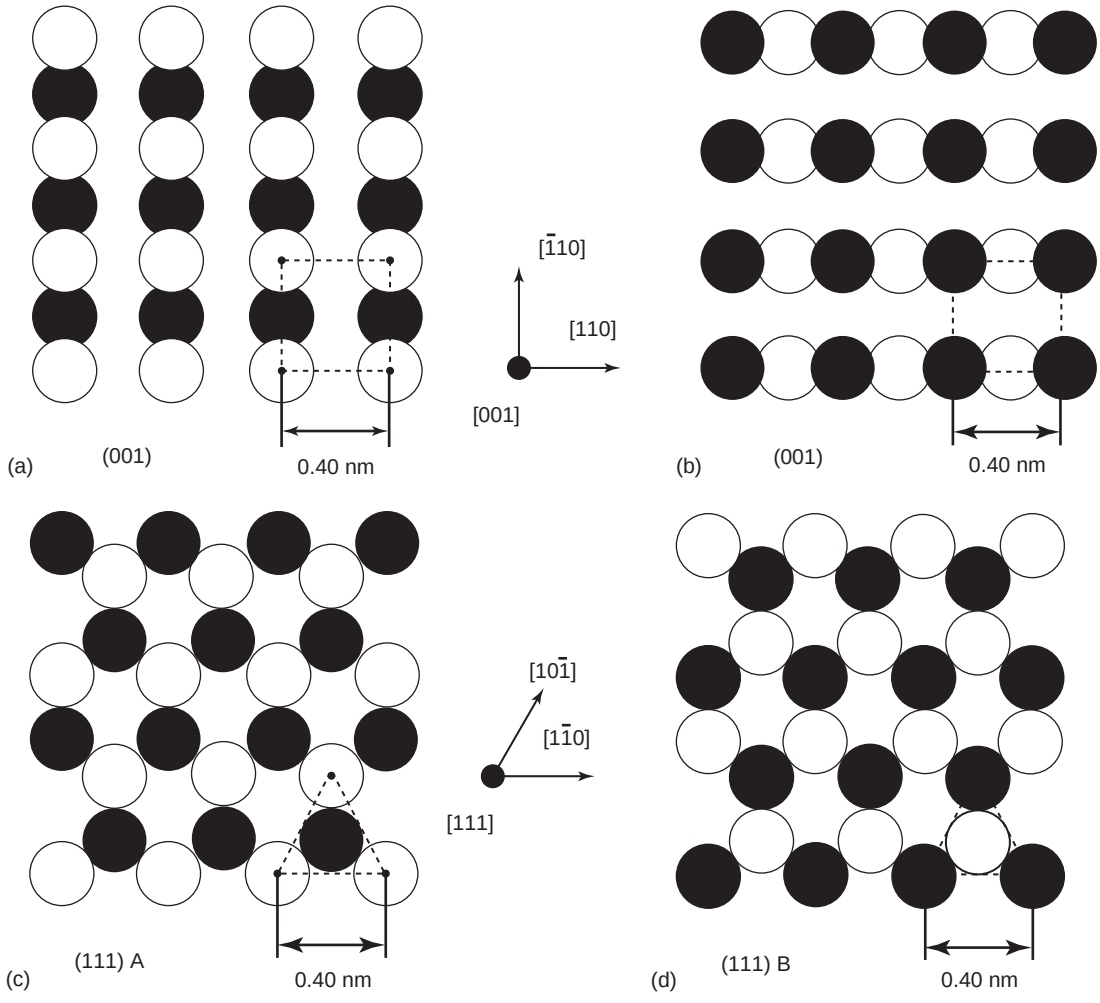

Figure 27 Top views of ball models for ideal (a) Ga-terminated GaAs(001), (b) As-terminated GaAs(001), (c) GaAs(111)A, and (d) GaAs(111)B. (Reprinted with permission from H. Yao, S.-L. Yau, K. Itaya, 'In Situ Scanning Tunneling Microscopy of GaAs(001), (111)A, and (111)B Surfaces in Sulfuric Acid Solution', *Appl. Phys. Lett.*, **68**, 1473–1475[82] (1996).)

there have been a fewer STM studies on InP than on GaAs. Previous STM studies of InP were mostly carried out in UHV.

We have recently shown the first atomically resolved STM images of InP(001), (111)A, and (111)B surfaces in an H_2SO_4 solution under cathodic potential control, which effectively protected the surfaces from oxidation.[84] These images clearly demonstrate that well-defined InP surfaces can be prepared by chemical etching in HCl solution. In STM images, individual atoms were relatively clearly observed on the atomically flat terraces of the (111) surface with a corrugation height of ca. 0.02 nm. Monolayer steps were found to be exactly parallel to the close-packed atomic row direction of the (111) surface. An atomically resolved STM image of the (111)A surface revealed a perfect hexagonal arrangement of In atoms with an interatomic distance of 0.42 nm, as expected for the ideal InP (111)A surface. These results clearly demonstrate that the ideal InP(111)A–(1×1) structure is exposed in the H_2SO_4 solution after the chemical etching. This (1×1) structure seemed to extend over the entire region of the terrace; even atomic defects were rarely observed.

On the other hand, the chemically etched (111)B surface also possessed a well-defined structure in H_2SO_4 solution. An atomically flat terrace-step structure was observed with a step height identical with that on the A surface, although the average terrace width was typically 5–10 nm, which was smaller than that on the A surface. The appearance of wider terraces on the (111)A surface is reasonable, because the etching rate of (111)A surface in HCl solution is the slowest of all low-index planes.

An almost ideal square arrangement expected for the (001) surface was clearly discerned by in situ

STM. The observed interatomic distances in the [110] and [1$\bar{1}$0] directions were found to be equal to 0.42 nm. Atomic images clearly demonstrate that the ideal InP(001–1 × 1) structure exists in H_2SO_4 solution under cathodic potential control. We also presume the InP(001–1 × 1), (111)A–(1 × 1) and (111)B–(1 × 1) surfaces to be terminated by hydrogen, at least under cathodic polarization.

In summary, it has been demonstrated that well-defined InP(001), (111)A, and (111)B surfaces can be prepared by chemical etching in HCl solution.[84] In situ STM revealed atomically flat terrace-step structures on each surface in H_2SO_4 solution under proper potential control. Furthermore, we successfully obtained the first atomic resolution STM images of InP surfaces, which showed that the ideal InP(001–1 × 1), (111)A–(1 × 1) and (111)B–(1 × 1) structures are exposed and persist in H_2SO_4 solution under EC conditions.

8 CONCLUSION

The methods for exposing well-defined electrode surfaces in solution were reviewed. The flame-annealing and quenching method can be applied to Au, Pt, Rh, Pd, and Ir single-crystal electrodes. The UHVEC method can be used for Pt, Au, Pd and Rh. For some metals, such as Ni and Cu, surface oxidation takes place in the EC chamber before immersion of the electrode in electrolyte solutions. It was demonstrated that the EC etching method produced atomically flat terrace-step structures of semiconductors under carefully adjusted EC conditions. The method of anodic dissolution is expected to become an important in situ technique for exposing well-defined surfaces of various semiconductors and metals.[7]

The structures of specifically adsorbed iodine on Pt(111) and Au(111) and briefly on Ag(111) were discussed, demonstrating that complementary use of in situ STM and ex situ LEED is a powerful combination to characterize the atomic structure of adsorbed iodine. The adsorption of sulfate/hydrogensulfate adsorbed on Au(111), Pt(111), and Rh(111) was described, with emphasis on the fact that the same ($\sqrt{3} \times \sqrt{7}$) structure is formed on these three substrates. Our model indicates that hydrogen-bonded water chains are inserted along the $\sqrt{3}$ direction between neighboring rows of the adsorbed sulfates. The detailed structural analysis of the CN adlayer on Pt(111) was carried out by in situ STM, revealing that six CN groups form a hexagonal ring without an additional CN at the center of the ring. The complexation of K^+ with the CN adlayer was also discussed.

The UPD of Cu on Au(111) in H_2SO_4 was discussed in depth. In situ SXS indicated that the Cu atoms form a honeycomb lattice and are adsorbed on threefold hollow sites with sulfate ions located at the honeycomb centers. According to the model structure, the corrugation observed by in situ STM and AFM should be ascribed to the coadsorbed sulfate ions, and not the Cu atoms.

It was shown that, in general, the iodine-modified electrodes are suitable for producing highly ordered adlayers of various organic molecules. TMPyP forms highly ordered adlayers on I–Au(111), I–Ag(111), and I–Pt(100).

The adlayer structure of benzene on Rh(111) and Pt(111) was also described in detail; it was found to be dependent on the electrode potential. The (3 × 3) structure found on Rh(111) in the cathodic potential range is almost identical with that found in UHV for the coadsorbed benzene and CO. The molecular shapes of naphthalene and anthracene could be clearly discerned by in situ STM.

The EC etching processes of Si(111), Si(110), and Si(100) were discussed in relation to the atomic structures of the step-edges. It was shown that the chemical etching produces well-defined GaAs and InP single-crystal electrodes. The atomic structures of these electrodes in solution could be clearly seen.

This article clearly demonstrates that STM allows us not only to determine interfacial structures but also to follow EC reactions. It is certain that in situ STM will continue to be the premier technique in the study of the relationship between the reactivity and the structure of electrode surfaces.

ACKNOWLEDGMENTS

This work was supported by the Exploratory Research for Advanced Technology (ERATO)–Itaya Electrochemiscopy Project organized by Japan Science and Technology Corporation (JST) and partially by the Ministry of Education, Science, Sports and Culture, Japan, with a Grant-in-Aid for Science Research on the Priority Area of ''Electrochemistry of Ordered Interfaces''.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
ARUPS	Angle-resolved Ultraviolet Photoemission Spectroscopy
CE	Counter Electrode
CV	Cyclic Voltammogram
DEMS	Differential Electrochemical Mass Spectrometry

EC	Electrochemical
EELS	Electron Energy-loss Spectroscopy
HOPG	Highly Ordered Pyrolytic Graphite
IR	Infrared
LEED	Low-energy Electron Diffraction
OCP	Open-circuit Potential
PPV	4,4′-bis(*N*-methylpyridinium)-*p*-phenylenedivinylene
RE	Reference Electrode
RHE	Reversible Hydrogen Electrode
SCE	Saturated Calomel Electrode
SHG	Second Harmonic Generation
STM	Scanning Tunneling Microscopy
SXS	Surface X-ray Scattering
TMPyP	5,10,15,20-tetrakis(*N*-methylpyridinium-4-yl)-21*H*,23*H*-porphyrine
UHV	Ultrahigh Vacuum
UHVEC	Ultrahigh Vacuum Electrochemical
UPD	Underpotential Deposition
WE	Working Electrode

RELATED ARTICLES

Surfaces **(Volume 10)**
Surfaces: Introduction • Proximal Probe Techniques • Scanning Probe Microscopy, Industrial Applications of • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Infrared Spectroelectrochemistry • Self-assembled Monolayers on Electrodes • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques • Ultraviolet/Visible Spectroelectrochemistry • X-ray Methods for the Study of Electrode Interaction

REFERENCES

1. A.T. Hubbard, 'Electrochemistry at Well-characterized Surfaces', *Chem. Rev.*, **88**, 633–656 (1988).
2. M.P. Soriaga, 'Ultra-high Vacuum Techniques in the Study of Single-crystal Electrode Surfaces', *Prog. Surf. Sci.*, **39**, 325–443 (1992).
3. H.-J. Guntherodt, R. Wiesendanger (eds.), *Scanning Tunneling Microscopy I*, Springer-Verlag, Berlin, 1991.
4. P.A. Christensen, 'Electrochemical Aspects of STM and Related Techniques', *Chem. Soc. Rev.*, **21**, 197–208 (1992).
5. H. Siegenthaler, 'STM in Electrochemistry', in *Scanning Tunneling Microscopy II*, eds. R. Wiesendanger, H.-J. Guntherodt, Springer-Verlag, Berlin, 7–49, 1992.
6. A.A. Gewirth, B.K. Niece, 'Electrochemical Applications of In Situ Scanning Probe Microscopy', *Chem. Rev.*, **97**, 1129–1162 (1997).
7. K. Itaya, 'In Situ Scanning Tunneling Microscopy in Electrolyte Solutions', *Prog. Surf. Sci.*, **58**, 121–247 (1998).
8. J.G. Simmons, 'Electric Tunnel Effect between Dissimilar Electrodes Separated by a Thin Insulating Film', *J. Appl. Phys.*, **34**, 2581–2590 (1963).
9. J. Clavilier, R. Faure, G. Guinet, R. Durand, 'Preparation of Monocrystalline Pt Microelectrodes and Electrochemical Study of the Plane Surfaces Cut in the Direction of the {111} and {110}', *J. Electroanal. Chem.*, **107**, 205–209 (1980).
10. A. Hamelin, 'Double-layer Properties at sp and sd Metal Single-crystal Electrodes', in *Modern Aspects of Electrochemistry*, eds. B.E. Conway, R.E. White, J.O'M. Bockris, Plenum Press, New York, 1–101, Vol. 16, 1985.
11. S. Motoo, N. Furuya, 'Hydrogen and Oxygen Adsorption on Ir(111), (100) and (110) Planes', *J. Electroanal. Chem.*, **167**, 309–315 (1984).
12. K. Sashikata, N. Furuya, K. Itaya, 'In Situ Electrochemical Scanning Tunneling Microscopy of Single-crystal Surfaces of Pt(111), Rh(111), and Pd(111) in Aqueous Sulfuric Acid Solution', *J. Vac. Sci. Technol.*, **B9**, 457–464 (1991).
13. J. Clavilier, A. Rodes, K. El. Achi, M.A. Zamakhchari, 'Electrochemistry at Platinum Single Crystal Surfaces in Acidic Media: Hydrogen and Oxygen Adsorption', *J. Chim. Phys.*, **88**, 1291–1337 (1991).
14. J.M. Feliu, J.M. Orts, R. Gomez, A. Aldaz, J. Clavilier, 'New Information on the Unusual Adsorption States of Pt(111) in Sulfuric Acid Solutions from Potentiostatic Adsorbate Replacement by CO', *J. Electroanal. Chem.*, **372**, 265–268 (1994).
15. K. Itaya, S. Sugawara, K. Sashikata, N. Furuya, 'In Situ Scanning Tunneling Microscopy of Platinum(111) surface with the Observation of Monatomic Steps', *J. Vac. Sci. Technol.*, **A8**, 515–519 (1990).
16. S. Tanaka, S.-L. Yau, K. Itaya, 'In-situ Scanning Tunneling Microscopy of Bromine Adlayers on Pt(111)', *J. Electroanal. Chem.*, **396**, 125–130 (1995).
17. D.M. Kolb, 'Reconstruction Phenomena at Metal–Electrolyte Interfaces', *Prog. Surf. Sci.*, **51**, 109–173 (1996).
18. K. Wang, G.S. Chottiner, D.A. Scherson, 'Electrochemistry of Nickel(111) in Alkaline Electrolytes', *J. Phys. Chem.*, **97**, 10 108–10 111 (1993).
19. J.L. Stickney, C.B. Ehlers, B.W. Gregory, 'Adsorption of Gaseous and Aqueous HCl on the Low-index Planes of Copper', *Langmuir*, **4**, 1368–1373 (1988).
20. M. Hourani, A. Wieckowski, 'Electrochemistry of the Ordered Rh(111) Electrode: Surface Preparation and Voltammetry in $HClO_4$ Electrolyte', *J. Electroanal. Chem.*, **227**, 259–264 (1987).
21. C.M. Vitus, S.-C. Chang, B.C. Schardt, M.J. Weaver, 'In Situ Scanning Tunneling Microscopy as a Probe of

Adsorbate-induced Reconstruction at Ordered Monocrystalline Electrodes: CO on Pt(100)', *J. Phys. Chem.*, **95**, 7559–7563 (1991).
22. F. Lu, G.N. Salaita, H. Baltruschat, A.T. Hubbard, 'Adlattice Structure and Hydrophobicity of Pt(111) in Aqueous Potassium Iodide Solutions: Influence of pH and Electrode Potential', *J. Electroanal. Chem.*, **222**, 305–320 (1987).
23. B.C. Schardt, S.-L. Yau, F. Rinaldi, 'Atomic Resolution Imaging of Adsorbates on Metal Surfaces in Air: Iodine Adsorption on Pt(111)', *Science*, **243**, 1050–1053 (1989).
24. N. Shinotsuka, K. Sashikata, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Underpotential Deposition of Ag on Pt(111) ($\sqrt{7} \times \sqrt{7}$)*R*19.1°–I', *Surf. Sci.*, **335**, 75–82 (1995).
25. K. Sashikata, Y. Matsui, K. Itaya, M.P. Soriaga, 'Adsorbed-iodine-catalyzed Dissolution of Pd Single-crystal Electrodes: Studies by Electrochemical Scanning Tunneling Microscopy', *J. Phys. Chem.*, **100**, 20 027–20 034 (1996).
26. L.-J. Wan, S.-L. Yau, G.M. Swain, K. Itaya, 'In-situ Scanning Tunneling Microscopy of Well-ordered Rh(111) Electrodes', *J. Electroanal. Chem.*, **381**, 105–111 (1995).
27. T. Yamada, N. Batina, K. Itaya, 'Structure of Electrochemically Deposited Iodine Adlayer on Au(111) Studied by Ultrahigh-vacuum Instrumentation and In Situ STM', *J. Phys. Chem.*, **99**, 8817–8823 (1995).
28. B.M. Ocko, G.M. Watson, J. Wang, 'Structure and Electrocompression of Electrodeposited Iodine Monolayers on Au(111)', *J. Phys. Chem.*, **98**, 897–906 (1994).
29. T. Yamada, K. Ogaki, S. Okubo, K. Itaya, 'Continuous Variation of Iodine Adlattices on Ag(111) Electrodes: In Situ STM and Ex Situ LEED Studies', *Surf. Sci.*, **369**, 321–335 (1996).
30. J. Inukai, Y. Osawa, M. Wakisaka, K. Sashikata, Y.-G. Kim, K. Itaya, 'Underpotential Deposition of Copper on Iodine-Modified Pt(111): In Situ STM and Ex Situ LEED Studies', *J. Phys. Chem.*, **102**, 3498–3505 (1998).
31. N. Batina, T. Yamada, K. Itaya, 'Atomic Level Characterization of the Iodine-modified Au(111) Electrode Surface in Perchloric Acid Solution by In-situ STM and Ex-situ LEED', *Langmuir*, **11**, 4568–4576 (1995).
32. T. Yamada, N. Batina, K. Ogaki, S. Okubo, K. Itaya, 'Lateral Structure of Iodine Adlattices on Au(111) and Ag(111) Electrodes', in *Electrode Processes VI, Proceedings of the Sixth International Symposium*, eds. A. Wieckowski, K. Itaya, Electrochemical Society, Pennington, NJ, 43–57, 1996.
33. B.M. Ocko, O.M. Magnussen, J.X. Wang, R.R. Adzic, Th. Wandlowski, 'The Structure and Phase Behavior of Electrodeposited Halides on Single-crystal Metal Surfaces', *Physica B*, **221**, 238–244 (1996).
34. O.M. Magnussen, J. Hagebock, J. Hotlos, R.J. Behm, 'In Situ Scanning Tunneling Microscopy Observations of a Disorder–Order Phase Transition in Hydrogensulfate Adlayers on Au(111)', *Faraday Discuss. Chem. Soc.*, **94**, 329–338 (1992).
35. G.J. Edens, X. Gao, M.J. Weaver, 'The Adsorption of Sulfate on Gold(111) in Acidic Aqueous Media: Adlayer Structural Inferences from Infrared Spectroscopy and Scanning Tunneling Microscopy', *J. Electroanal. Chem.*, **375**, 357–366 (1994).
36. Z. Shi, J. Lipkowski, M. Gamboa, P. Zelenay, A. Wieckowski, 'Investigations of SO_4^{2-} Adsorption at the Au(111) Electrode by Chronocoulometry and Radiochemistry', *J. Electroanal. Chem.*, **366**, 317–326 (1994).
37. A.M. Funtikov, U. Stimming, R. Vogel, 'Anion Adsorption from Sulfuric Acid Solutions on Pt(111) Single Crystal Electrodes', *J. Electroanal. Chem.*, **428**, 147–153 (1997).
38. L.-J. Wan, S.-L. Yau, K. Itaya, 'Atomic Structure of Adsorbed Sulfate on Rh(111) in Sulfuric Acid Solution', *J. Phys. Chem.*, **99**, 9507–9513 (1995).
39. P. Zelenay, A. Wieckowski, 'Radiochemical Assay of Adsorption at Single Crystal/Solution Interfaces' *J. Electrochem. Soc.*, **139**, 2552–2558 (1992).
40. C. Stuhlmann, I. Villegas, M.J. Weaver, 'Scanning Tunneling Microscopy and Infrared Spectroscopy as Combined In Situ Probes of Electrochemical Adlayer Structure. Cyanide on Pt(111)', *Chem. Phys. Lett.*, **219**, 319–324 (1994).
41. Y.-G. Kim, S.-L. Yau, K. Itaya, 'Direct Observation of Alkali Cations on Cyanide-modified Pt(111) by Scanning Tunneling Microscopy', *J. Am. Chem. Soc.*, **118**, 393–400 (1996).
42. O.M. Magnussen, J. Hotlos, R.J. Nichols, D.M. Kolb, R.J. Behm, 'Atomic Structure of Cu Adlayers on Au(100) and Au(111) Electrodes Observed by In Situ Scanning Tunneling Microscopy', *Phys. Rev. Lett.*, **64**, 2929–2932 (1990).
43. O.M. Magnussen, J. Hotlos, G. Beitel, D.M. Kolb, R.J. Behm, 'Atomic Structure of Ordered Copper Adlayers on Single-crystalline Gold Electrodes', *J. Vac. Sci. Technol.*, **B9**, 969–975 (1991).
44. T. Hachiya, H. Honbo, K. Itaya, 'Detailed Underpotential Deposition of Copper on Gold(111) in Aqueous Solutions', *J. Electroanal. Chem.*, **315**, 275–291 (1991).
45. S. Manne, P.K. Hansma, J. Massie, V.B. Elings, A.A. Gewirth, 'Atomic-resolution Electrochemistry with the Atomic Force Microscopy: Copper Deposition on Gold', *Science*, **251**, 183–186 (1991).
46. Z. Shi, J. Lipkowski, 'Coadsorption of Cu^{2+} and SO_4^{2-} at the Au(111) Electrode', *J. Electroanal. Chem.*, **365**, 303–309 (1994).
47. M.F. Toney, J.N. Howard, J. Richer, G.L. Borges, J.G. Gordon, O.R. Melroy, D. Yee, L. B. Sorensen, 'Electrochemical Deposition of Copper on a Gold Electrode in Sulfuric Acid: Resolution of the Interfacial Structure', *Phys. Rev. Lett.*, **75**, 4472–4475 (1995).
48. D.A. Huckaby, L. Blum, 'A Model for Sequential First-order Phase Transitions Occurring in the Underpotential Deposition of Metals', *J. Electroanal. Chem.*, **315**, 255–261 (1991).

49. S. Chiang, 'Scanning Tunneling Microscopy Imaging of Small Adsorbed Molecules on Metal Surfaces in an Ultrahigh Vacuum Environment', *Chem. Rev.*, **97**, 1083–1096 (1997).
50. A. Ikai, 'STM and AFM of Bio/Organic Molecules and Structures', *Surf. Sci. Rep.*, **26**, 261–332 (1996).
51. J. Lipkowski, P.N. Ross (eds.), *Adsorption of Molecules at Metal Electrodes*, VCH, New York, 1992.
52. N.J. Tao, Z. Shi, 'Real-time STM/AFM Study of Electron Transfer Reactions of an Organic Molecule: Xanthine at the Graphite–Water Interfaces', *Surf. Sci. Lett.*, **321**, L149–L156 (1994).
53. N.J. Tao, G. Cardenas, F. Cunha, Z. Shi, 'In Situ STM and AFM Study of Protoporphyrin and Iron(III) and Zinc(II) Protoporphyrins Adsorbed on Graphite in Aqueous Solutions', *Langmuir*, **11**, 4445–4448 (1995).
54. F. Cunha, N.J. Tao, 'Surface Charge Induced Order–Disorder Transition in an Organic Monolayer', *Phys. Rev. Lett.*, **75**, 2376–2379 (1995).
55. M. Kunitake, N. Batina, K. Itaya, 'Self-organized Porphyrine Array on Iodine-modified Au(111) in Electrolyte Solutions: In Situ Scanning Tunneling Microscopy Study', *Langmuir*, **11**, 2337–2340 (1995).
56. M. Batina, M. Kunitake, K. Itaya, 'Highly Ordered Molecular Arrays Formed on Iodine-modified Au(111) in Solution: In Situ STM Imaging', *J. Electroanal. Chem.*, **405**, 245–250 (1996).
57. M. Kunitake, U. Akiba, N. Batina, K. Itaya, 'Structures and Dynamic Formation Processes of Porphyrine Adlayers on Iodine-modified Au(111) in Solution: In Situ STM Study', *Langmuir*, **13**, 1607–1615 (1997).
58. K. Ogaki, N. Batina, M. Kunitake, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Ordering Processes of Adsorbed Porphyrine on Iodine-modified Ag(111)', *J. Phys. Chem.*, **100**, 7185–7190 (1996).
59. K. Sashikata, T. Sugata, M. Sugimasa, K. Itaya, 'In Situ Scanning Tunneling Microscopy Observation of a Porphyrine Adlayer on an Iodine-modified Pt(100) Electrode', *Langmuir*, **14**, 2896–2902 (1998).
60. G.A. Somorjai, *Introduction to Surface Chemistry and Catalysis I*, J. Wiley & Sons, New York, 1994.
61. H. Ohtani, R.J. Wilson, S. Chiang, C.M. Mate, 'Scanning Tunneling Microscopy Observations of Benzene Molecules on the Rh(111–3 × 3)(C_6H_6 + 2CO) Surface', *Phys. Rev. Lett.*, **60**, 2398–2401 (1988).
62. H.A. Yoon, M. Salmeron, G.A. Somorjai, 'Scanning Tunneling Microscopy (STM) Study of Benzene and its Coadsorption with Carbon Monoxide on Rh(111)', *Surf. Sci.*, **373**, 300–306 (1997).
63. S.-L. Yau, Y.-G. Kim, K. Itaya, 'In Situ Scanning Tunneling Microscopy of Benzene Adsorbed on Rh(111) and Pt(111) in HF Solution', *J. Am. Chem. Soc.*, **118**, 7795–7803 (1996).
64. H. Baltruschat, U. Schmiemann, 'The Adsorption of Unsaturated Organic Species at Single Crystal Electrodes Studied by Differential Electrochemical Mass Spectrometry', *Ber. Bunsenges. Phys. Chem.*, **97**, 452–460 (1993).
65. M. Neuber, F. Schneider, C. Zubragel, M. Neumann, 'A Dense and CO Free Benzene Structure on Rh(111)', *J. Phys. Chem.*, **99**, 9160–9168 (1995).
66. P.S. Weiss, D.M. Eigler, 'Site Dependence of the Apparent Shape of a Molecule in Scanning Tunneling Microscope Images: Benzene on Pt{111}', *Phys. Rev. Lett.*, **71**, 3139–3142 (1993).
67. S.-L. Yau, Y.-G. Kim, K. Itaya, 'High-resolution Imaging of Aromatic Molecules Adsorbed on Rh(111) and Pt(111) in Hydrofluoric Acid Solution: In Situ STM Study', *J. Phys. Chem. B*, **101**, 3547–3553 (1997).
68. R.F. Lin, R.J. Koestner, M.A. Van Hove, G.A. Somorjai, 'The Adsorption of Benzene and Naphthalene on the Rh(111) Surface: A LEED, AES and TDS Study', *Surf. Sci.*, **134**, 161–183 (1983).
69. S.-L. Yau, K. Itaya, 'Structures and Dynamic Processes of Molecular Adlayers on Rh(111) and Pt(111) in HF Solution: Naphthalene and Biphenyl', *Colloids Surf. A: Physicochem. Eng. Aspects*, **134**, 21–30 (1998).
70. G.S. Higashi, Y.J. Chabal, G.W. Trucks, K. Raghavachari, 'Ideal Hydrogen Termination of the Si(111) Surface', *Appl. Phys. Lett.*, **56**, 656–658 (1990).
71. R.S. Becker, G.S. Higashi, Y.J. Chabal, A.J. Becker, 'Atomic Scale Conversion of Clean Si(111):H–1 × 1 to Si(111)–2 × 1 by Electron-stimulated Desorption', *Phys. Rev. Lett.*, **65**, 1917–1920 (1990).
72. H.E. Hessel, A. Feltz, M. Reiter, U. Memmert, R.J. Behm, 'Step-flow Mechanism versus Pit Corrosion: Scanning-tunneling Microscopy Observations on Wet Etching of Si(111) by HF Solutions', *Chem. Phys. Lett.*, **186**, 275–280 (1991).
73. K. Itaya, R. Sugawara, Y. Morita, H. Tokumoto, 'Atomic Resolution Images of H-terminated Si(111) Surfaces in Aqueous Solutions', *Appl. Phys. Lett.*, **60**, 2534–2536 (1992).
74. S.-L. Yau, K. Kaji, K. Itaya, 'Electrochemical Etching of Si(001) in NH_4F Solutions: Initial Stage and {111} Microfacet Formation', *Appl. Phys. Lett.*, **66**, 766–768 (1995).
75. K. Kaji, S.-L. Yau, K. Itaya, 'Atomic Scale Etching Processes of n-Si(111) in NH_4F Solutions: In Situ Scanning Tunneling Microscopy', *J. Appl. Phys.*, **78**, 5727–5733 (1995).
76. J.H. Ye, K. Kaji, K. Itaya, 'Atomic-scale Elucidation of the Anisotropic Etching of (110) n-Si in Aqueous NH_4F: Studies by In Situ Scanning Tunneling Microscopy', *J. Electrochem. Soc.*, **143**, 4012–4019 (1996).
77. P. Allongue, V. Kieling, H. Gerischer, 'Etching Mechanism and Atomic Structure of H-Si(111) Surfaces Prepared in NH_4F', *Electrochimica Acta*, **40**, 1353–1360 (1995).
78. P. Allongue, V. Costa-Kieling, H. Gerischer, 'Etching of Silicon in NaOH Solutions: I. In Situ Scanning Tunneling

Microscopic Investigation of n-Si (111)', *J. Electrochem. Soc.*, **140**, 1009–1018 (1993).
79. P. Allongue, V. Costa-Kieling, H. Gerischer, 'Etching of Silicon in NaOH Solutions: II. Electrochemical Studies of n-Si(111) and (100) and Mechanism of the Dissolution', *J. Electrochem. Soc.*, **140**, 1018–1026 (1993).
80. P. Jakob, Y.J. Chabal, K. Kuhnke, S.B. Christman, 'Monohydride Structures on Chemically Prepared Silicon Surfaces', *Surf. Sci.*, **302**, 49–56 (1994).
81. H. Yao, S.-L. Yau, K. Itaya, 'Atomic Resolution Images of GaAs(111)A Surfaces in Sulfuric Acid Solution', *Surf. Sci.*, **335**, 166–170 (1995).
82. H. Yao, S.-L. Yau, K. Itaya, 'In Situ Scanning Tunneling Microscopy of GaAs(001), (111)A, and (111)B Surfaces in Sulfuric Acid Solution', *Appl. Phys. Lett.*, **68**, 1473–1475 (1996).
83. K. Uosaki, S. Ye, N. Sekine, 'Study of the Electronic Structure of GaAs(100) Single Crystal Electrode/Electrolyte Interfaces by Electrochemical Tunneling Spectroscopy', *Bull. Chem. Soc. Jpn.*, **69**, 275–288 (1996).
84. H. Yau, K. Itaya, 'Atomically Resolved Scanning Tunneling Microscopy Images of InP(001), (111)A, and (111)B Surfaces in Sulfuric Acid Solution', *J. Electrochem. Soc.*, **145**, 3090–3094 (1998).

Selective Electrode Coatings for Electroanalysis

Lee J. Klein and Dennis G. Peters
Indiana University, Bloomington, USA

1 Introduction **1**
2 Self-assembled Monolayers on Electrodes **1**
2.1 Types of Materials Used 2
2.2 Methods of Preparation 2
2.3 Methods of Characterization 3
2.4 Analytical Applications 3
3 Polymer-coated Electrodes **4**
3.1 Organic Conducting Polymers 4
3.2 Metal Ion-containing Polymers 5
3.3 Ion-exchange Polymers 8
4 Langmuir–Blodgett Films on Electrodes **10**
4.1 Types of Materials Used 11
4.2 Methods of Preparation 11
4.3 Methods of Characterization 12
4.4 Analytical Applications 13
5 Immobilized Enzymes on Electrodes **13**
5.1 Types of Materials and Methods of Preparation 13
5.2 Methods of Characterization 15
5.3 Analytical Applications 15
Abbreviations and Acronyms **15**
Related Articles **16**
References **16**

Tremendous strides have taken place in recent years in our ability to modify electrodes chemically to enhance their selectivity and sensitivity for electroanalyses. Four different categories of selective electrode coatings have been chosen for discussion in this article: self-assembled monolayers (SAMs); polymer coatings; Langmuir–Blodgett films; and immobilized enzymes. Each of these chemical modifications is discussed in terms of the necessary materials, the methods of preparation and characterization, and some examples of analytical applications.

1 INTRODUCTION

In the realm of electroanalysis, one of the greatest challenges is to design and fabricate an electrode that ideally will respond both selectively and sensitively to a particular chemical species (e.g. a simple inorganic cation or anion, a metal–ion complex, or an organic or biological material). Over the past 30 years or so, tremendous advances have been made in our ability to tailor or modify chemically the surfaces of electrodes to enhance their selectivity and sensitivity for a variety of analyses.[1] In a sense, these advances constitute a form of molecular engineering, in which one knowledgeably and purposefully redesigns or alters the surface of an electrode to try to define and control how that electrode responds to one targeted species or to a family of species. In this short article, we attempt to introduce and describe some of these advances.

Four different categories of selective electrode coatings have been chosen for discussion. First, we begin with SAMs on electrodes; a SAM is an organized array of molecules strongly adsorbed or chemically bonded to an electrode surface, and such a monolayer can form spontaneously when a carefully prepared electrode surface is brought into contact with an appropriate species that can interact with the surface. Second, we consider polymer-coated electrodes; most simply, one can create a polymer-coated electrode (a) by electropolymerizing a selected (and perhaps functionalized) monomer directly onto the surface of an electrode or (b) by coating or depositing a preformed (and functionalized) polymer onto an electrode surface. Third, we describe the preparation of Langmuir–Blodgett films on electrodes; a Langmuir–Blodgett film is an ordered monolayer of amphiphilic molecules formed at a liquid/gas interface, and this film is subsequently transferred mechanically to the surface of an electrode. Fourth, we discuss the immobilization of enzymes onto electrodes, which can be accomplished in a variety of ways, including electrostatic self-adsorption and incorporation into SAMs, polymer films, or Langmuir–Blodgett films.

Recent progress in this area of research and development has been so great that we have been forced to be selective in choosing the topics, examples, and references included in this article. Thus, we can offer only a brief glimpse into this highly interesting subject. Obviously, by referring to the cited reviews and original papers, and by seeking out other publications by the major contributors to this field, the interested reader can find more information.

2 SELF-ASSEMBLED MONOLAYERS ON ELECTRODES

When an electrode (substrate) is immersed into a solvent in which an appropriate surfactant is dissolved, the molecules of surfactant can interact (react) spontaneously

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

at the surface of the electrode to give a SAM. From the viewpoint of electroanalysis, an excellent example of such a self-assembly process is seen in the spontaneous formation of a SAM consisting of a functionalized alkanethiol on the surface of a gold electrode. For purposes of illustration, let us assume that we have a solution of $(C_5H_5)Fe(C_5H_4)CONH(CH_2)_6SH$ in ethanol. Typical of many surfactants, $(C_5H_5)Fe(C_5H_4)CONH(CH_2)_6SH$ consists of three moieties: (a) the surface-active head group, $-SH$, which tethers the molecule to the gold surface via the formation of an $S-Au$ bond; (b) the alkyl group, $-(CH_2)_6-$, which is essentially a spacer that separates the head group from the end group; and (c) the end group itself, $(C_5H_5)Fe(C_5H_4)CONH-$, which is farthest from the surface of the gold electrode and which imparts desirable characteristics to the SAM on gold (such as selectivity toward a particular species in a solution to be analyzed). Figure 1 depicts the spontaneous process by which this functionalized alkanethiol, originally dissolved in ethanol, forms a SAM on gold.

Although details of the mechanism of adsorption of an alkanethiol are still unclear, the most important step in the formation of a SAM on gold is chemisorption of the deprotonated (thiolate) form, $(C_5H_5)Fe(C_5H_4)CONH(CH_2)_6S^-$, onto the gold surface.[2,3] Then, as more and more of the thiolate form becomes attached (adsorbed) to the gold surface, the assembly of a closely packed (ordered) SAM takes place.

2.1 Types of Materials Used

Since the pioneering work of Zisman,[4] who investigated the wetting behavior of solutions of surfactants, several classes of compounds have been used to prepare SAMs. First, long-chain carboxylic (*n*-alkanoic) acids can be adsorbed onto a variety of substrates, including aluminum oxide,[5–7] silver,[8,9] and iron oxide.[10] In addition, diacetylenic carboxylic acids, such as $CH_3(CH_2)_8C{\equiv}C-C{\equiv}C(CH_2)_8COOH$, have been used to prepare SAMs on glass, quartz, and silicon;[11] although such substrates cannot serve as electrodes, they could be rendered conductive by being coated with thin layers of indium-doped tin oxide (ITO) prior to the formation of the SAM. Second, alkylsilane derivatives can be bonded onto hydroxylated surfaces; for example, compounds such as $RSiX_3$, R_2SiX_2, and R_3SiX (where R is an alkyl chain, which may or may not be functionalized with a suitable electroactive end group, and X is a chloro or alkoxy moiety) can be attached to glass,[12] silicon dioxide,[13] or aluminum oxide,[14] via the formation of stable $Si-O$ bonds. Once again, the strategy of coating the substrate with a thin layer of ITO (which is also a hydroxylated surface) can provide a conductive electrode onto which the SAM can form. When a trichlorosilane ($RSiCl_3$) is used to prepare a SAM on one of these hydroxylated substrates, it is possible to obtain a cross-linked network, as illustrated in Figure 2.

Third, as described earlier, functionalized organosulfur compounds such as alkanethiols, dialkyl sulfides, and dialkyl disulfides can form SAMs on gold,[15–18] silver,[19] copper,[20] nickel,[21] mercury,[22] and platinum[23] surfaces. Gold is much preferred, however, because it has the least tendency to form a surface oxide and is, therefore, the easiest substrate to prepare and handle.

2.2 Methods of Preparation

Many of the articles cited in the preceding section contain specific procedures employed to prepare SAMs; the reader is encouraged to consult these sources. However, for illustrative purposes, it is useful to examine some key steps of a general protocol for the formation of a SAM consisting of a functionalized alkanethiol on gold. First, the particular gold substrate must be selected; one may use either a polycrystalline (rough) gold surface or an evaporated (smooth) gold surface. Second, the gold surface is usually cleaned (freed of adventitious impurities) by being heated, immersed in solutions of strong oxidants, exposed to an argon or

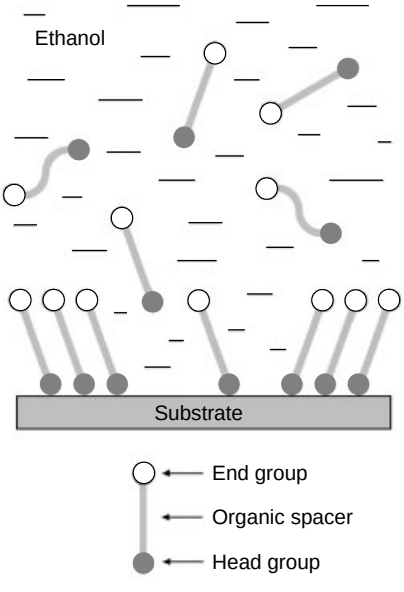

Figure 1 Spontaneous formation of a SAM of functionalized alkanethiol onto a gold surface.

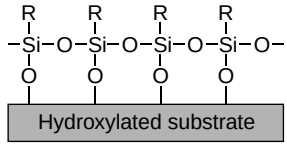

Figure 2 Cross-linked network formed by reaction of a trichlorosilane ($RSiCl_3$) with a hydroxylated substrate.

oxygen plasma, or placed in a suitable electrolytic cell whereupon the potential of the gold electrode is cycled; often a combination of these techniques is used. Third, the desired functionalized alkanethiol (which must often be synthesized) is dissolved in an appropriate organic solvent; typically, the concentration of the alkanethiol is 10^{-3} M. Fourth, the gold substrate is immersed into the alkanethiol solution for times ranging from minutes to 12 h or longer at room temperature to promote the self-assembly process. Finally, after being rinsed, the monolayer-covered gold electrode may be annealed thermally to cause the monolayer to rearrange into a more ordered structure.

2.3 Methods of Characterization

Several electroanalytical techniques can be used to study and characterize SAMs on electrodes. Cyclic voltammetry is the most popular method, but differential pulse voltammetry, chronoamperometry, and hydrodynamic voltammetry deserve mention.

Ulman[24] and Finklea[25] have reviewed the literature on the various nonelectrochemical techniques that can be employed to characterize SAMs. Although space limitations prevent any detailed discussion of these methods, it is useful at least to list some of these techniques, most of which are completely described elsewhere in this encyclopedia. Many workers have utilized wetting contact angle to elucidate the structure of SAMs. Both infrared and Raman spectroscopy as well as ellipsometry and surface plasmon resonance techniques have seen extensive use. High-vacuum surface methods such as X-ray photoelectron spectroscopy (XPS), Auger electron spectroscopy (AES), high-resolution electron energy loss spectroscopy (HREELS), and low-energy electron diffraction spectroscopy (LEEDS) have been employed. Scanning tunneling microscopy (STM) and atomic force microscopy (AFM) can provide images of alkanethiol-based SAMs on gold substrates. Some investigators have used the electrochemical quartz crystal microbalance (EQCM) to probe the kinetics of self-assembly of alkanethiols onto gold surfaces.

2.4 Analytical Applications

Despite enormous interest in the preparation and properties of SAMs, especially those involving functionalized alkanethiols on gold electrodes, the list of practical analytical applications is still somewhat limited. A review by Mandler and Turyan[26] summarizes the literature through mid-1995 on the use of SAMs in electroanalytical chemistry. Citing a few examples of analytical applications will help to convey the excitement and promise of this area of research.

Steinberg and Rubinstein[27] found that a mixed SAM, composed of 2,2′-thiobisethyl acetoacetate and *n*-octadecyl mercaptan, on a gold substrate can be used to determine Cu^{2+} in the presence of Fe^{3+}, and that replacing *n*-octadecyl mercaptan with *n*-octadecyltrichlorosilane afforded an electrode system for the detection of trace levels of Cu^{2+} (10^{-7} M), Pb^{2+} (10^{-5} M), and Zn^{2+} (10^{-9} M). A novel microsensor for pH measurements has been developed that is based on the coadsorption of ferrocenyl and quinone thiols onto a gold microelectrode; the redox potential for the ferrocenyl moiety is independent of pH, but the redox potential for the quinone moiety is pH dependent, so the difference in these redox potentials is related to the pH of the sample.[28]

Turyan and Mandler[29] have devised a highly sensitive and selective method for the determination of Cd^{2+}; a detection limit as low as 4×10^{-12} M was obtained. These workers prepared SAMs, consisting of *w*-mercaptocarboxylic acids ($HS(CH_2)_nCOOH$, where $n = 2$, 3, 6, and 11) adsorbed onto both thin mercury film and gold electrodes. When placed into a buffered solution containing Cd^{2+}, the modified electrode acts to preconcentrate Cd^{2+}, possibly through formation of a complex involving the interaction of Cd^{2+} with both the sulfur and carboxylate moieties of the SAM. After the preconcentration step, the complexed cadmium(II) can be determined by means of voltammetry.

Another approach for gaining selectivity for electroanalysis involves the incorporation of so-called ''molecular gates'' in an otherwise blocking monolayer.[30] Two mercaptans, 1-hexadecanethiol and 4-hydroxythiophenol, are coadsorbed onto a gold surface to form the SAM; the first compound is a passivating (blocking) substance, whereas the second species acts as an electron-transfer site (the ''molecular gate''). These ''molecular gates'' allow $[Ru(NH_3)_6]^{3+}$, but not $[Fe(CN)_6]^{3-}$, to undergo a reversible one-electron reduction.

Chlorpromazine in human urine has been measured by means of an amperometric flow detector coated with unsubstituted *n*-alkanethiols on gold.[31] By varying the chain length of the alkanethiol, one can tailor the permselective properties of the SAM so that other constituents of the urine are prevented from reaching the gold surface. In the same study, it was shown that such a SAM-coated detector can distinguish dopamine (a neurotransmitter) from ascorbate; this approach overcomes the difficult problem of detecting dopamine in cerebral fluid, which contains relatively high concentrations of ascorbate.

A method to determine glucose relies on the mediated oxidation of the reduced form of glucose oxidase (GOx(red)), by dicationic (ferricenylmethyl)dimethyloctadecyl ammonium ion ($C_{18}Fc^{2+}$), which is incorporated into a self-assembled bilayer.[32]

Preparation of the bilayer begins with the spontaneous reaction of octadecyltrichlorosilane with a film of microporous aluminum oxide to form a cross-linked SAM. After gold is evaporated onto the back of the SAM–Al_2O_3 assembly, (ferrocenylmethyl)dimethyloctadecyl ammonium ion ($C_{18}Fc^+$) is allowed to interleave with the silanized surface to form a bilayer, and glucose oxidase (GOx) is immobilized on the bilayer. In the presence of glucose, the following reactions (Equations 1–3) occur

$$\text{GOx} + \text{glucose} \longrightarrow \text{GOx(red)} + \text{gluconolactone} \quad (1)$$

$$\text{GOx(red)} + 2C_{18}Fc^{2+} \longrightarrow \text{GOx} + 2C_{18}Fc^{+} + 2H^{+} \quad (2)$$

$$C_{18}Fc^{+} \rightleftharpoons C_{18}Fc^{2+} + e^{-} \quad (3)$$

where the last process gives rise to a steady-state voltammetric current that is related to the original concentration of glucose.

3 POLYMER-COATED ELECTRODES

Although electrodes coated with a variety of polymer films have been prepared, there are essentially three major types of polymers to consider here: electronically conducting organic polymers, polymers containing electroactive metal-ion sites, and ion-exchange polymers.

3.1 Organic Conducting Polymers

3.1.1 Types of Materials Used

Two of the best examples of organic conducting polymers are polypyrrole **(1)** and polythiophene **(2)**. Other conducting polymers include polyacetylene, polyaniline, poly-*p*-phenylene, polyazulene, polyindole, and polyfuran. Moreover, it is possible to prepare copolymers consisting of two different monomers (such as a mixture of thiophene and pyrrole[33,34]), and even more exotic starting materials (such as 1-thienyl-2-cyano-2-phenylethylene[35,36]) can be used. Information about the properties of conducting polymers on electrodes can be found in two reviews.[37,38]

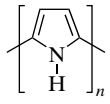

(1)

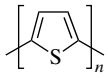

(2)

3.1.2 Methods of Preparation

There are several important requirements for the successful electrochemical preparation of a polymer film on an electrode: the parent monomer must be readily oxidizable (reducible) at an electrode in the chosen solvent–electrolyte; the radical cation (anion) arising from oxidation (reduction) of the monomer must react rapidly with other monomeric species to grow the polymer; the polymer must have a relatively low solubility so that it will deposit on the electrode; and the polymer must have an oxidation (reduction) potential less (more) positive than that of the monomer if the polymer is to have a high conductivity. A common procedure[39] for the formation of a film of polypyrrole on a platinum substrate involves the controlled-potential electrolysis of a solution of 50 mM pyrrole in acetonitrile containing 0.10 M tetraethylammonium tetrafluoroborate. Films are grown slowly by proper choice of the potential, the thickness of a film (typically 20–80 nm) is controlled by the amount of charge passed, and oxygen is excluded from the system. Polypyrrole films can also be prepared by means of constant-current electrolysis, and polypyrrole can be deposited onto other substrates including carbon, gold, palladium, and ITO.

Electrochemical techniques for the preparation of some of the other polymeric films mentioned above are similar to those used to obtain polypyrrole coatings. In addition, one can prepare some conducting polymers by casting a soluble precursor polymer onto a chosen substrate and by heating the system to form the desired material.[37] Alternatively, if the conducting polymer can be synthesized by a convenient chemical procedure and if the polymer is soluble in a volatile solvent, a solution of the polymer can be applied to an electrode surface by means of dip- or spin-coating, and the solvent can be evaporated to produce the polymer-coated electrode.

3.1.3 Methods of Characterization

Most of the methods mentioned in section 2.3 can be utilized to characterize polymer films on electrodes. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) are other useful approaches. A rich source of both electrochemical and nonelectrochemical techniques for the study of conducting polymers is a review by Doblhofer and Rajeshwar.[40]

One of the most important properties of a conducting polymer film is its conductivity, which can be measured by means of the classic four-probe method.[41] Conductivity of a conjugated polymer film arises from partial oxidation (p-doping) or partial reduction (n-doping). In electrochemical p-doping – which all of the polymers mentioned so far undergo – electrons are withdrawn from the polymer, and the polymer becomes positively charged. Then anions (usually from the supporting electrolyte) diffuse into the polymer to provide charge compensation; thus, the choice of supporting electrolyte can affect the nature of the conducting polymer. In combination, all of these processes profoundly influence the formation, structure,

and conductivity of organic polymers; these issues are discussed in two previously cited reviews.[37,38]

3.1.4 Analytical Applications

Several kinds of analytical applications have been developed that involve the use of an electrode coated with a film of conducting organic polymer. Examples in the present section have been chosen to exclude situations where the polymer film also contains either trapped transition-metal sites (see section 3.2) or an immobilized enzyme (see section 5).

Shiu et al.[42] have reported that, when pyrrole is oxidatively polymerized onto a glassy carbon electrode in the presence of either bathophenanthroline disulfonate or bathocuproine disulfonate, these electroinactive anionic complexing agents are incorporated into the polymer. When Cu^+ or Cu^{2+} is extracted (via formation of a complex) from a sample solution into the polymer, a cyclic voltammetric peak current is seen that can be used to determine these cations; the electrode responds better to Cu^+ than to Cu^{2+}, and the detection limit for Cu^+ is approximately 2×10^{-6} M. In their paper, the authors review earlier efforts to develop polymer-coated electrodes for the determination of other metal cations.

Because p-doping of a polymer is associated with incorporation of anions, this phenomenon can be utilized to fabricate polymer-coated electrodes that respond to anions in a sample solution. This principle has been employed for the development of a polypyrrole-coated platinum detector for the flow injection analysis (FIA) of phosphate, carbonate, and acetate;[43] the electrode response to phosphate and carbonate is linear over the concentration range from 10^{-5} to 10^{-3} M. Similarly, a polyaniline-coated glassy carbon detector for the FIA of iodide, bromide, thiocyanate, and thiosulfate has been devised, with detection limits of 1, 5, 10, and 10 mg L^{-1}, respectively.[44] A poly(3-methylthiophene)-coated graphite electrode has been constructed for the potentiometric determination of iodide, chloride, bromide, and sulfite;[45] for the measurement of iodide, the linear dynamic range of the electrode is 1×10^{-7} to 5×10^{-1} M, and the detection limit for iodide is 1×10^{-8} M.

A platinum electrode coated with polypyrrole with incorporated anti-human serum albumin (serving as a charge-compensating anion) has been found to respond linearly in alternating current (AC) voltammetric experiments to human serum albumin at concentrations ranging from 10 to 50 μg mL^{-1}.[46] Polypyrrole on glassy carbon, as well as polypyrrole on mercury-filmed glassy carbon, provides a sensor for the amperometric determination of dichloramine separated from chloramine in potable water by means of high-performance liquid chromatography (HPLC); a detection limit of 4×10^{-6} M dichloramine is possible by use of the polypyrrole–mercury film–glassy carbon electrode.[47]

Slater et al.[48] have devised sensors consisting of platinum microband-electrode arrays coated with polypyrrole (or an alkylated polypyrrole). When such a detector is exposed to the vapors of an alcohol, molecules of the alcohol permeate the polymer, causing the polypyrrole film to swell and to undergo a change in conductivity. Time-resolved measurement of the change in conductivity of the polypyrrole sensors was used to separate and classify methanol, ethanol, and propanol vapors.

3.2 Metal Ion-containing Polymers

Several strategies can be employed to incorporate electroactive transition-metal sites into polymer films on electrodes; these kinds of polymer films are often referred to as *redox polymers*, although this terminology can also embrace polymer films bearing electroactive organic groups (e.g. naphthalene, anthracene, or viologen). First, it is possible to polymerize onto the surface of an electrode a monomer in which the transition-metal site is *covalently bonded* to the polymerizable moiety; a classic example is vinylferrocene, which gives rise to poly(vinylferrocene) **(3)**.[49] Second, the monomer to be polymerized can consist of a transition-metal complex *coordinately linked* to the polymerizable moiety; an example of such a monomer is $[Ru(bpy)_2(vbpy)]^{2+}$ (where vbpy denotes 4-methyl-4′-vinyl-2,2′-bipyridine and bpy denotes 2,2′-bipyridine) which forms poly$[Ru(bpy)_2(vbpy)]^{2+}$ **(4)**.[50] Third, a transition-metal complex can be *coordinately linked* to a polymer after the polymer has been deposited onto an electrode; an example is the interaction of $[Ru(NH_3)_5OH_2]^{2+}$ with a pyridine moiety of a film of poly(4-vinylpyridine) (PVP) **(5)**.[51]

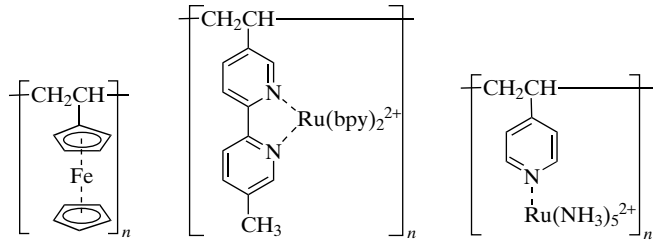

(3) **(4)** **(5)**

Fourth, an ion-exchange polymer can be coated onto an electrode, and the fixed charged sites (cationic or anionic) *electrostatically interact* with a metal complex; an example is the attachment of $[Ru(bpy)_3]^{2+}$ to sulfonate groups of Nafion®, a perfluorinated polysulfonated ion-exchange polymer. Only the first three kinds of redox polymers will be discussed further in this section; films of ion-exchange polymers on electrodes are described in section 3.3.

Before leaving this discussion, we should mention a few other ways to immobilize electroactive transition-metal sites onto electrode surfaces, although these approaches do not strictly involve polymer-coated electrodes. One method[52] entails adsorption of a molecule with an extended π-electron system (e.g. 1-(9-phenanthrene)-2-(4-pyridine)ethene) onto a suitable electrode (e.g. pyrolytic graphite or glassy carbon); then an appropriately substituted transition-metal species (e.g. $[Ru(NH_3)_5OH_2]^{2+}$) is coordinately linked to the adsorbate. In another approach, a functionalized silane (e.g. 4-[β-(trichlorosilyl)ethyl]pyridine) is attached to an electrode possessing surface hydroxy groups (e.g. Pt−OH), after which a transition-metal complex (e.g. $[Ru(bpy)_2(DME)Cl]^+$, where DME denotes dimethoxyethyl) is bonded to the functionalized (pyridyl) end of the silane.[53] In addition, one can functionalize the surface groups on edge planes of pyrolytic graphite by treating the graphite with a suitable reagent (e.g. thionyl chloride or cyanuric chloride); subsequent reaction with tetra(aminophenyl)porphyrin, followed by metallation of the porphyrin, completes the process.[54]

Many aspects of redox polymers bearing electroactive transition-metal sites, as well as other features of chemically modified electrodes, have been reviewed by Abruña.[55]

3.2.1 Types of Materials Used

Three examples of polymerizable species with covalently bonded transition-metal sites are (a) the aforementioned vinylferrocene (which can be polymerized onto platinum), (b) *N*-(2-ferrocenylethyl)pyrrole **(6)** (which can be copolymerized with pyrrole to produce a ferrocene-containing polypyrrole film on platinum),[56] and (c) *N*,*N*′-bis-[3-(pyrrol-1-yl)propyl]ferrocene-1,1′-dicarboxamide **(7)** (which can be polymerized onto glassy carbon).[57]

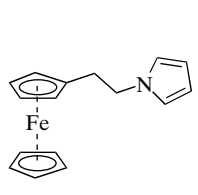

(6)

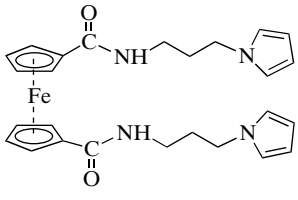

(7)

We now turn our attention to electrodes coated with polymer films bearing coordinately linked redox-active transition-metal sites; the transition metal can be (a) part of the original molecule (monomer) from which the polymer film is created or (b) introduced after the polymer film has been deposited onto the surface of the electrode.

Included in the first subclass are monomers having metal-containing ligands bound to pyrrole.[58] One example is $[Ru(bpy)L_2]^{2+}$, where L is 4-(2-pyrrol-1-ylethyl)-4′-methyl-2,2′-bipyridine, which can be anodically polymerized onto a platinum electrode; when either ferrocene or decamethyl ferrocene permeates the polymer film, there can be direct oxidization at the underlying platinum or catalytic oxidation by electrogenerated ruthenium(III) sites.[59] A second example is a nickel(II) pyridyltritolylporphyrin linked to *N*-(3-propyl)pyrrole; the resulting compound **(8)** can be oxidatively polymerized onto platinum, and the resulting polymer film exhibits both reversible oxidation and reduction processes.[60]

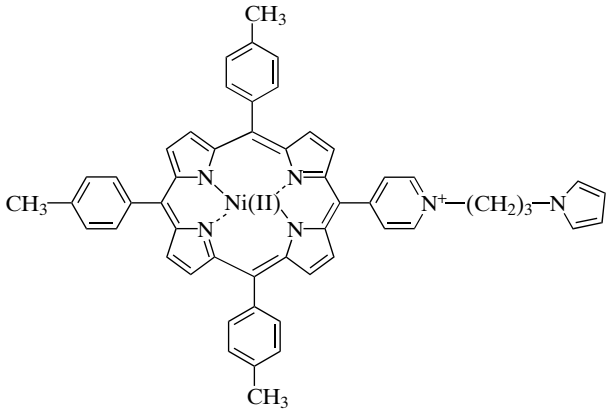

(8)

There has been much additional research dealing with polymer films that are formed from monomers with already incorporated transition-metal sites. Earlier we mentioned work[50] in which the monomer $[Ru(bpy)_2(vbpy)]^{2+}$ can be reductively polymerized to form poly$[Ru(bpy)_2(vbpy)]^{2+}$ on platinum, glassy carbon, tin oxide, and titanium oxide electrodes; in the same study, related monomers containing ruthenium or iron were polymerized, and it was shown that pairs of such monomers can yield site-mixed copolymers as well as spatially segregated two-layer films. Glassy carbon electrodes dip-coated with aquobis(bpy)(poly-4-vinylpyridine)ruthenium(II) have been used by Samuels and Meyer[61] as a way to generate immobilized ruthenium(IV) sites for the catalytic oxidation of 2-propanol. Anodic polymerization of cobalt- and nickel-containing tetrakis(*o*-, *m*-, and *p*-aminophenyl)porphyrin, tetrakis(*p*-(dimethylamino)phenyl)porphyrin, tetrakis(*p*-hydroxyphenyl)porphyrin, and tetrakis(*p*-*N*-pyrrolylphenyl)porphyrin onto platinum has been described;[62] a film of electropolymerized cobalt tetrakis(*o*-aminophenyl)porphyrin catalyzes the reduction of oxygen. Electroreductive polymerization of $[Fe(vbpy)_3]^{2+}$ and $[Ru(4,4'\text{-dimethyl-bpy})_2(vbpy)]^{2+}$ has been accomplished

at a platinum cathode; emphasis was placed on the structure of the polymers and their mechanism of formation.[63] When a copolymer consisting of 1-vinyl-2-pyrrolidone and *meso*-[tri(phenyl)mono(*p*-methacrylamidophenyl)]porphine is treated with iron(II), the resulting solid can be dissolved in tetrahydrofuran (THF), and the solution can be applied to the surface of a graphite electrode; evaporation of the solvent gives a coated electrode with redox-active iron sites that catalyze the reduction of oxygen.[64] Treatment of poly-*N*-vinylimidazole (pnvi) with $[Ru(bpy)_2Cl_2]$ gives several metallopolymers, such as $[Ru(bpy)_2(pnvi)_2]^{2+}$ and $[Ru(bpy)_2(pnvi)Cl]^{+}$, which can be coated onto carbon electrodes.[65]

For the second subclass, for which an electroactive transition-metal species is introduced after the polymer is deposited onto an electrode, we will cite just two representative examples. Oyama and Anson[51] dip-coated pyrolytic graphite with films of PVP or polyacrylonitrile (PAN), and demonstrated that ruthenium(III)(edta)OH_2 (where edta is ethylenediaminetetraacetate) loses H_2O in coordinating to a pendant pyridine group of PVP and that $[Ru(NH_3)_5OH_2]^{2+}$ loses H_2O in binding to a $-CN$ moiety of PAN; in addition, the interaction of $[Ru(NH_3)_5OH_2]^{2+}$ with a film consisting of a mixture of PVP and PAN was examined. Platinum electrodes coated with *p*-chlorosulfonated polystyrene have been treated with polypyridyl complexes of iron, ruthenium, and osmium as well as a ferrocene derivative, a nickel macrocycle, and a nickel porphyrin;[66] these inorganic species possess unbound amine, hydroxy, and carboxylate moieties that can react with the *p*-chlorosulfonated polystyrene to anchor the redox-active transition-metal complex to the polymer.

As a final group of metal-containing species that can be used to modify an electrode surface, metal phthalocyanines **(9)** and metal salens **(10)** deserve some mention.

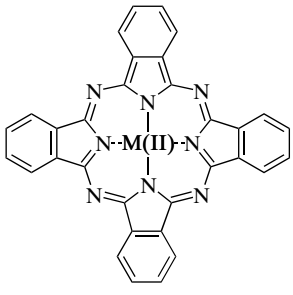

(9)

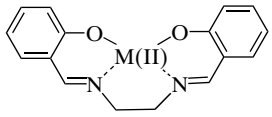

(10)

Over the years, metal phthalocyanines have received much attention as catalysts for the reduction of oxygen. Films of polypyrrole and polyaniline, each intercalated with tetrasulfonated iron phthalocyanine, have been electrodeposited onto gold; these chemically modified electrodes appear to catalyze the reduction of oxygen to water.[67] In addition, the reduction of oxygen has been examined at pyrolytic graphite electrodes spin-coated with a number of different metal phthalocyanines,[68] and at highly oriented pyrolytic graphite electrodes with adsorbed layers of iron and cobalt crown phthalocyanines.[69] Furthermore, oxidative electropolymerization of cobalt and nickel tetraaminophthalocyanines onto glassy carbon electrodes has been accomplished.[70] Metal salen complexes are distinguished by the fact that the ligand (a Schiff base) which coordinates the metal cation can be oxidatively polymerized onto a variety of electrodes; thus, the resulting polymers are conductive (due to their extended π-conjugated systems) as well as redox-active (due to the presence of a transition metal). A number of investigators have studied the formation and properties of anodically polymerized films of metal salens.[71–74]

3.2.2 Methods of Preparation

As the preceding section has revealed, metal ion-containing polymers on electrodes have been prepared in many ways.[55] Anodic or cathodic electropolymerization of a metal-containing monomer onto an appropriate electrode is the most common and straightforward procedure, and many of the references cited above provide details.[49,50,56–60,62,63,70–74] Usually, it is most convenient to carry out the direct electropolymerization by means of cyclic voltammetry, because the thickness of the polymer film can be controlled by the number of consecutive scans and by the scan rate; an excellent example of this technique is provided by Figure 3, which depicts the oxidative polymerization of copper(II) salen onto a platinum electrode.[72]

Other methods of preparing metal-containing polymer films on electrodes include metallation of an already electrodeposited polymer film;[51] adsorption of an organic species, followed by metallation;[52] attachment of a complexing ligand via silanization, followed by metallation;[53] functionalization of edge-plane sites of graphite, followed by reaction with a complexing ligand and introduction of a metal ion;[54] dip-coating of an electrode with a metallated polymer;[61,64,65] covalent linkage of a functionalized metal-containing species to a functionalized polymer film;[66] intercalation of a polymer with a metal-containing compound;[67] spin-coating of an electrode with a metal-containing species;[68] and adsorption of a metallated compound.[69]

3.2.3 Methods of Characterization

Virtually all of the electrochemical and spectroscopic techniques referred to in section 3.1.3 are suitable for

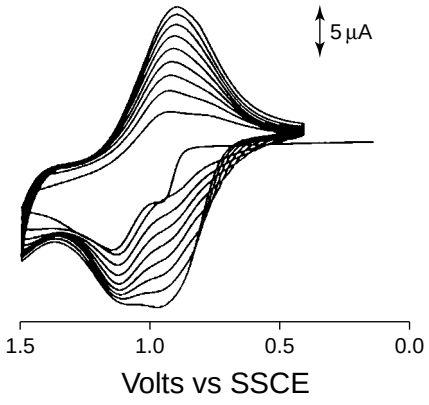

Figure 3 Consecutive cyclic voltammograms showing the oxidative polymerization of copper(II) salen in acetonitrile containing 0.10 M tetra-*n*-butylammonium hexafluorophosphate at a platinum disk electrode and at a scan rate of $100\,mV\,s^{-1}$. (Reproduced by permission from Hofercamp and Goldsby.[72] Copyright 1989, American Chemical Society.)

the characterization of metal-containing polymer films on electrodes. Many of the references listed in the preceding section contain specific information about the methods employed to characterize the chemically modified electrodes that are the subjects of those studies. Most importantly, since the metal-ion sites within a polymer coating are ultimately what is essential to its redox activity, electroanalytical techniques are especially informative.

3.2.4 Analytical Applications

Most studies of metal-containing polymers on electrodes have focused on (a) understanding charge-transport (electron-transfer) mechanisms within such films, (b) investigating catalytic processes that take place within and at the surface of these films, or (c) characterizing the electrochemistry of the immobilized metal center itself. Practical analytical applications remain an area ripe for further development. We can, however, summarize several analytical uses for electrodes coated with metal-containing polymers.

Earlier, we indicated that poly-*N*-vinylimidazole reacts with $[Ru(bpy)_2Cl_2]$ to give ruthenium-containing polymers.[65] When coated onto a rotating disk electrode, these polymers provide a sensor that responds linearly to either $[Fe(CN)_6]^{4-}$ or $[Fe(H_2O)_6]^{2+}$ at concentrations as low as 5×10^{-6} M.[75] A glassy carbon electrode coated with polymerized nickel(II) phthalocyanine has been used for amperometric determination of phenolic antioxidants – *tert*-butylhydroxytoluene (BHT), *tert*-butylhydroxyanisole (BHA), propyl gallate (PG), and *tert*-butylhydroquinone (TBHQ) – after separation by HPLC;[76] detection limits for BHA, PG, and BHT are 0.11, 0.15, and 0.60 ppm, respectively. A sensor has been developed that utilizes a glassy carbon electrode dip-coated with $[Ru(bpy)_2(PVP)_{10}Cl]Cl$ for the flow injection measurement of nitrite;[77] this procedure has been applied to the determination of nitrate in fertilizer, after the nitrate in the sample is first converted to nitrite with the aid of a copper–cadmium reductor.

A carbon electrode bearing a film of anodically polymerized tetraaminophthalocyanatocobalt(II) responds potentiometrically in a Nernstian fashion to sulfide ion (3×10^{-6} to 10^{-3} M) and to mercaptoethanol (6×10^{-6} to 10^{-3} M).[78] Amperometric determination of BHA (an antioxidant) has been accomplished with a poly[tetraaminophthalocyanatonickel(II)]-modified glassy carbon electrode;[79] a linear response was observed for the concentration range from 5.6×10^{-8} to 5.6×10^{-5} M, and the detection limit was 1.5×10^{-8} M. Glassy carbon, coated with electropolymerized tetraaminophthalocyanatonickel(II), is a sensor for the determination of dopamine at concentrations from 2×10^{-7} to 2×10^{-5} M, with a detection limit of 9×10^{-8} M;[80] one can eliminate the interference due to ascorbic acid by coating the modified electrode with a layer of Nafion®.

3.3 Ion-exchange Polymers

As mentioned previously, Nafion® **(11)** is a perfluorinated, polysulfonated cation-exchange polymer.

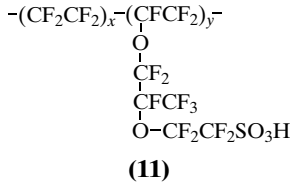

(11)

One can purchase Nafion®, in its protonated form, as a dilute solution of the polymer in alcohol–water. A measured small volume of this solution can be applied to the surface of a glassy carbon electrode; then, after the solvent has evaporated, a film of the polymer adheres to the electrode. This coated electrode can be soaked in an appropriate electrolyte solution (e.g. dilute sulfuric acid), and can then be inserted into an electrochemical cell containing a solution of some cationic metal complex (e.g. $[Ru(bpy)_3]^{2+}$) along with the same supporting electrolyte. Quickly, the $[Ru(bpy)_3]^{2+}$ becomes incorporated into the polymer film via a cation-exchange process (two protons at two adjacent $-SO_3^-$ sites are replaced by one $[Ru(bpy)_3]^{2+}$ moiety) until a steady state is attained, whereupon a variety of electroanalytical measurements can be performed.[81] One can employ an alternative approach, which involves preloading the Nafion® with

$[Ru(bpy)_3]^{2+}$ (or some other electroactive species), by adding the latter to the alcohol–water solution of the polymer before it is applied to the electrode surface. This is the essence of many procedures for the preparation and use of ion-exchange polymer films on electrodes.

3.3.1 Types of Materials Used

Both cationic and anionic ion-exchange polymer films have been coated onto electrodes. Despite the overwhelming popularity of Nafion®, several other types of cation-exchange polymers have been employed. Majda and Faulkner[82] used platinum electrodes spin-coated with the sodium salt of poly(styrene sulfonate) to investigate the electrochemistry of incorporated $[Ru(bpy)_3]^{2+}$. When a pyrolytic carbon electrode filmed with deprotonated polyacrylic acid is placed into a solution of $[Ru(NH_3)_6]^{3+}$, the $-COO^-$ sites can electrostatically bind the positively charged metal complex.[83] Another approach taken by Vining and Meyer[84] involved coating glassy carbon electrodes with *p*-chlorosulfonated polystyrene; then the polymer-filmed electrode was exposed to an aqueous medium of pH 9 in order to convert the $-SO_2Cl$ sites to $-SO_3^-$ sites which can bind catalytically active, cationic ruthenium and osmium oxo-bridged dimers. In addition, poly(vinyl sulfate) has been proposed as a coating that can bind cations, including $[Ru(NH_3)_6]^{3+}$, $[Ru(bpy)_3]^{2+}$, $[Co(bpy)_3]^{3+}$, and $[Co(phen)_3]^{3+}$.[85] A mixture of a cystine-derivatized pyrrole and a tetraethylammonium-substituted pyrrole can be oxidatively copolymerized onto platinum or glassy carbon electrodes; chemical treatment of the polymer-coated electrode with dithiothreitol converts the polymer to its cysteine-SH form which tightly binds $[Fe_4S_4]^{2+}$ centers to give an electrode–polymer assembly with electroactive, cysteinyl-ligated, ferredoxin-like units.[86] Carbon electrodes coated with Eastman Kodak AQ 55 (a polysulfonated block copolymer) incorporate and strongly retain electroactive counter-cations (e.g. $[Ru(1,10\text{-phenanthroline})_3]^{2+}$ and $[Ru(9,10\text{-phenanthroline-5,6-dione})_3]^{2+}$) in aqueous and acetonitrile solutions.[87]

Electrodes coated with anion-exchange polymers that can electrostatically bind anionic metal complexes have received considerable attention. Oyama and Anson[88] demonstrated that graphite electrodes filmed with protonated PVP can strongly bind multiply charged anions such as $[Fe(CN)_6]^{3-}$ and $[IrCl_6]^{2-}$. Methyl- or benzyl-quaternized PVP provides a polyelectrolyte coating for graphite that is superior to PVP, because the density of cationic sites available for anion binding is independent of pH.[85] Protonated poly(L-lysine), coated onto a freshly cleaved graphite electrode, has been shown to interact electrostatically with both $[Fe(III)(edta)]^-$ and $[Fe(II)(edta)]^{2-}$, and a model to explain charge propagation through the film has been proposed.[89] A platinum electrode coated with poly[5,15-bis(2-aminophenyl)porphyrin], with and without metallation by cobalt(II), responds potentiometrically at pH 5.5 to a variety of anions, including iodide, bromide, fluoride, thiocyanate, salicylate, nitrate, and acetate.[90] An interesting new derivative of polyacetylene with pendant cationic moieties – namely, poly(dihexyldipropargyl ammonium bromide) – has been synthesized and has been coated (as a mixed film with Nafion®) onto an ITO electrode; however, its binding of anionic transition-metal complexes has not yet been explored.[91]

3.3.2 Methods of Preparation

By far the most common methods to prepare electrodes with films of ion-exchange polymers are the techniques of dip-coating, drop-coating, and spin-coating. These procedures are literally self-explanatory, but details can be found in the various references cited in the preceding section. Brumlik et al.[92] have described a new approach for the fabrication of sulfonated fluorochlorocarbon ionomer films on surfaces of stainless steel and microporous alumina, which entails the radiofrequency plasma polymerization of trifluorochloroethylene (TFCE) and trifluoromethane sulfonic acid (TFMSA); this technique offers the advantage of providing anion-exchange polymers that are not commercially available, and so it deserves attention by electroanalytical chemists.

3.3.3 Methods of Characterization

As for other previously described kinds of selective coatings on electrodes, the same methods of characterization (including all of the electroanalytical techniques) outlined in sections 2.3 and 3.1.3 can be employed for surfaces coated with films of ion-exchange polymers. In addition, some of the references cited in section 3.1.1 provide valuable insights. In particular: film thicknesses have been measured;[82,87] diffuse-reflectance Fourier transform infrared (FTIR) spectroscopy has been used to characterize a polymer;[86,92] the quantities of electrostatically bound transition-metal species have been determined spectrophotometrically through the use of optically transparent, polymer-coated electrodes;[89] the electrical conductivities of polymer films, along with their absorption, fluorescence, electron paramagnetic resonance (EPR), and X-ray diffraction spectra, have been determined;[91,92] and SEM and XPS of polymer films have proven to be useful.[92]

3.3.4 Analytical Applications

Electrodes coated with ion-exchange polymers, especially Nafion®, have been employed for many different kinds of electrochemical studies, including electroanalysis. In some applications the purpose of a Nafion® film has been to provide or to improve selectivity toward one or more specific analytes, whereas in other instances a Nafion® film has served to protect some other coating on the surface of an electrode.

To alleviate problems associated with an adventitious surfactant (alkaline phosphatase) during the adsorptive stripping voltammetric determination of copper(II) at the 10 nM concentration level, Economou and Fielden[93] used Nafion® to coat a mercury-filmed glassy carbon electrode. In a method for the measurement of copper(II) in natural waters,[94] a carbon electrode, first coated with a complexing agent (1-phenyl-3-methyl-4-octanoylpyrazol-5-one), was subsequently modified with a Nafion® film; the procedure is successful for the determination of nanomolar to micromolar concentrations of copper(II), with a relative standard deviation of 1–6%. Lead(II) has been measured by means of square-wave voltammetry with a glassy carbon electrode modified with a mercury film coated with Nafion® that contains 2,2′-bipyridine;[95] for a 5-min preconcentration step, the anodic stripping method gives a calibration curve for lead(II) that is linear from approximately 2×10^{-8} to 2×10^{-6} M, with a detection limit of 2×10^{-9} M.

A Nafion®-coated glassy carbon electrode, treated with ruthenium(III), can be employed as an amperometric sensor for the determination of hydrazine, dimethylhydrazine, and phenylhydrazine in drinking and river waters; the method provides a linear response for hydrazine in the range from 1.8×10^{-7} to 4.6×10^{-4} M, with a detection limit of 1.0×10^{-7} M, and similar analytical results can be achieved with the other hydrazines.[96] A method for the amperometric determination of nitric oxide has been devised that utilizes an ultramicroelectrode chemically modified with a film of poly[nickel(II) salen] and Nafion®; the electrode responds linearly to in vivo nitric oxide at concentrations from 1.0×10^{-8} to 4.0×10^{-6} M in biological systems.[97] A Nafion®-modified carbon paste electrode has been developed for the square-wave voltammetric determination of Doxazosin (an antihypertensive substance), after the compound is adsorptively preconcentrated onto the electrode;[98] the detection limit is 2.3×10^{-11} M, which is better (by five orders of magnitude) than a spectrophotometric method.

Zen and Wang[99] have employed a graphite electrode spin-coated with Nafion® which is (a) impregnated with different ratios of Pb^{2+} and Ru^{3+}, (b) treated with potassium hydroxide to form $Pb(OH)_2$ and $Ru(OH)_3$, and (c) conditioned in the presence of oxygen at 53 °C; this electrode responds linearly to dissolved oxygen at concentrations ranging from 1 to 10 ppm in both acidic and alkaline environments. Finally, nanoelectrode ensembles filmed with Eastman Kodak AQ 55 (a polysulfonated block copolymer) have been studied by Ugo et al.[100] In separate experiments, (ferrocenylmethyl)trimethylammonium cation and $[Ru(NH_3)_6]^{3+}$ were preconcentrated (partitioned) into these modified electrodes, and the cyclic voltammetric response for each system was examined. For the first cation, linear calibration curves for concentrations ranging from 20 nM to 50 μM were obtained (with a detection limit of 5 nM), and similar results were achieved for $[Ru(NH_3)_6]^{3+}$ (with a detection limit of 1 nM). This approach to highly sensitive analysis merits extension to systems containing other cations.

4 LANGMUIR–BLODGETT FILMS ON ELECTRODES

For many types of electrode coatings, the dual nature of surfactants plays a key role. In the language of Langmuir–Blodgett films, the surfactant molecule is termed an ''amphiphile'' and is mechanically forced into an organized monolayer within a liquid/gas interface. Typically, one accomplishes this by first preparing a dilute solution (~1%) of pure amphiphile(s) in a solvent of high vapor pressure (e.g. chloroform). Aliquots of this solution are then allowed to evaporate over a liquid (the subphase) in which the amphiphiles as a whole are insoluble. Usually, the subphase consists of purified water of controlled pH and ionic strength, but mercury, glycerol, hydrocarbons, and other liquids have been employed as well.[101] After the volatile solvent evaporates, the amphiphiles are in a state known as a two-dimensional gas in which they are widely and randomly distributed along the surface of the subphase. With aqueous subphases, the amphiphiles are oriented with their hydrophilic regions immersed, forcing hydrophobic side chains up and out of the water. As shown in Figure 4, the two-dimensional gas and subphase are contained in a specially designed (commercially available) trough with a movable barrier traversing the interface.

At this stage, the barrier is then moved along the trough to force the amphiphiles into successively more ordered states, while the surface pressure and area per molecule are recorded. A pressure–area isotherm results which frequently displays marked phase transitions. A theoretical isotherm for stearic acid on 0.01 M hydrochloric acid, depicted in Figure 5, illustrates the various phases of film formation.

If the area of a solid-phase film is decreased further by the movable barrier, a sharp decline in surface pressure results, indicating buckling of the monolayer and the

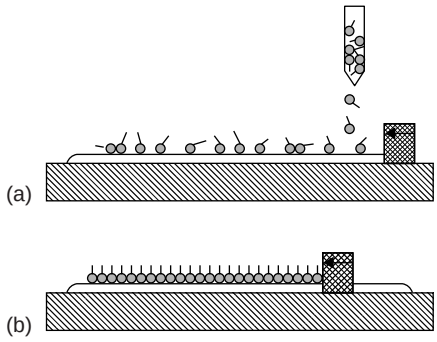

Figure 4 (a) Formation of the two-dimensional gas and (b) compression into an ordered monolayer. (Reproduced by permission from Ulman.[107] Copyright 1991, Academic Press.)

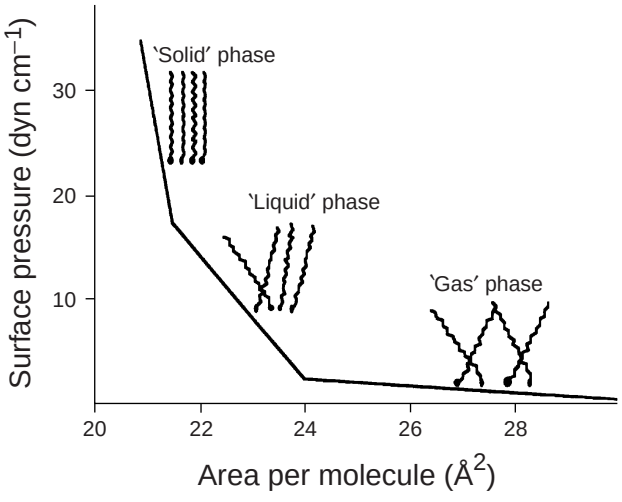

Figure 5 Theoretical isotherm for stearic acid on 0.01 M hydrochloric acid. (Reproduced by permission from Ulman.[107] Copyright 1991, Academic Press.)

formation of multilayer structures. For the purposes of electrode coatings, such films are to be avoided since they are not reproducible, and multilayer coatings on electrodes can easily be achieved through the repeated application of monolayers.

4.1 Types of Materials Used

A wide variety of molecular structures can be incorporated into Langmuir–Blodgett films. DeArmond and Fried[102] and Goldenberg[103] have summarized much of the work that has been done with electroactive amphiphiles (which usually must be synthesized). Space limitations preclude a detailed discussion of these species, but some noteworthy examples include derivatized fullerenes, tetrathiafulvalenes (TTFs), quinones, and various metal-containing complexes such as porphyrins, phthalocyanines, cyclams, bipyridine complexes, and metallocenes. In addition, several polymerizable amphiphiles have been employed in electrode films. With these species, polymerization may be initiated either before or after transfer of the monolayer to the electrode, a feature which is unique to the Langmuir–Blodgett method.

In some applications, the amphiphile need not be electroactive, in which case a very large array of molecules may be employed for a Langmuir–Blodgett film. Hann[101] has provided detailed coverage of many of these species: long-chain (C_{16} or higher) fatty acids and their various derivatives comprise the largest and most well studied class. For biological applications, phospholipids and sterols are most commonly employed.[104]

4.2 Methods of Preparation

Once an ordered monolayer has been formed, it must be transferred onto an electrode surface; the method employed for this transfer depends on the shape and wetting characteristics of the electrode surface. For planar electrodes, transfer may be accomplished in two ways. First, but less well understood, is the method involving vertical immersion and emersion of the electrode through the interface. Figure 6 illustrates how this method was recently applied to the coating of optically transparent ITO electrodes.[105,106]

In the first stage of the process, the ITO electrode (a hydrophilic substrate) penetrates an ordered monolayer on water from above. Formation of an upward-curving meniscus prevents the amphiphiles from adsorbing as the electrode is immersed. Slow emersion of the electrode from the subphase then takes place (5 mm min^{-1}), as the surface pressure is maintained constant by the movable barrier. If the substrate is sufficiently smooth, water will drain away and evaporate from the surface of the electrode as an even monolayer is transferred. Reimmersion of such an electrode into the interface can then be used to produce bilayers and multilayers whose structures closely resemble those of biological cell membranes.[104] One can make a preliminary evaluation of the quality and quantity of film transferred by simply comparing the area of the

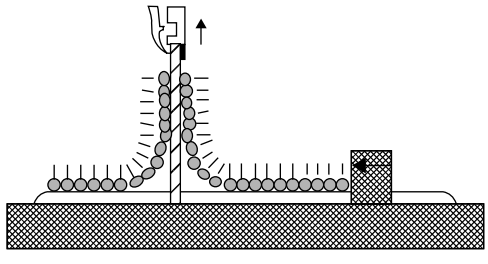

Figure 6 Transfer of a monolayer to a planar electrode by slow emersion. (Reproduced by permission from Ulman.[107] Copyright 1991, Academic Press.)

electrode to the change in area of the trough as dictated by the barrier (the transfer ratio).[107]

Alternatively, or if the electrode is not conveniently submerged, a horizontal-lifting method of transfer may be employed. Figure 7 illustrates how this procedure (Schaefer's method) has been applied to the coating of freshly cleaned (hydrophobic) glassy carbon electrodes.[108,109]

One coats the electrode by simply lowering it onto a monolayer parallel with the surface. Following reversed-phase adsorption of the film, the electrode is either withdrawn with a monolayer (or bilayer)[102] intact on its surface or held at the interface as excess monolayer is swept away. In the former case, recompression of the interface by means of the movable barrier generates a new ordered monolayer.

To perform electrochemical examination of the film, it is necessary to transfer the electrode to a different solution. In some cases, however, withdrawal of the electrode from the subphase is unnecessary, and the electrochemistry can be done in situ. In these experiments, counter- and reference electrodes are pre-positioned in a subphase containing a suitable electrolyte. So-called ''horizontal-touch'' techniques have been successfully employed with both glassy carbon[110] and ITO[111] electrodes.

Horizontal transfers avoid the surface flow of the monolayer inherent with vertical methods, and thus can preserve the original structure of the monolayer to a larger degree. Furthermore, surface-pressure requirements are much less stringent, making the technique applicable to a wider variety of amphiphiles.[111]

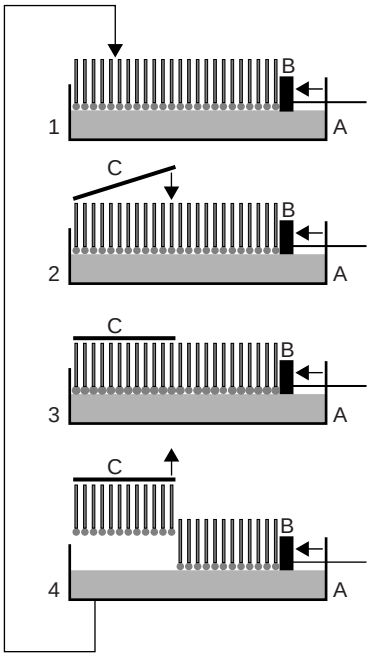

Figure 7 Transfer of a monolayer to an electrode by the horizontal-lifting technique (Schaefer's method): (A) trough; (B) movable barrier; (C) substrate. (Reproduced by permission from Ulman.[107] Copyright 1991, Academic Press.)

Nelson and Benton[112] described a vertical-type procedure for the preparation of a spherical hanging mercury drop electrode (HMDE) modified with a monolayer of naturally occurring phospholipids. In this case, film deposition occurs during the immersion step, with the long alkyl chains of the lipids next to the mercury.

As a consequence of their formation and structure, Langmuir–Blodgett films exhibit several clear advantages over other electrode coatings in terms of flexibility, ease of application, and sensitivity. First, one can easily make mixed films in any proportion by combining the appropriate ratio of one amphiphile with another in the evaporated solvent. As long as the amphiphiles are miscible with one another, the resulting films will have a large degree of homogeneity with no dependence on competing chemical reactions (as with mixed SAMs and copolymers). Second, the Langmuir–Blodgett technique allows one to prepare both highly ordered and multiply layered coatings on electrodes; the number of discrete layers can range from one to as many as 50. Third, the monolayer approach is very efficient, consuming only tiny quantities of synthetic amphiphiles,[111] while exposing the vast majority of these molecules to electroanalysis. Expensive amphiphiles can be further conserved by being mixed with well behaved long-chain carboxylic acids.[106] Casting and bulk polymerization methods tend to be far more wasteful. Fourth, several manufacturers offer fully automated, programmable-trough instruments, enabling one to prepare films with little operator involvement. Some of these are even equipped with multiple troughs, allowing one to prepare alternate-layered electrodes.

For some Langmuir–Blodgett films, a possible disadvantage lies in the fact that they are inherently less stable than other electrode coatings. Inclusion of solvent during deposition of a Langmuir–Blodgett film can lead to its post-transfer instability; evaporation of solvent trapped in irregularities or imperfections on the surface can cause collapse of the monolayer.[101] Another problem is that forming the film and transferring it to the surface of an electrode are both very time-consuming processes.

4.3 Methods of Characterization

An excellent summary of the nonelectrochemical techniques employed to characterize Langmuir–Blodgett and other organic thin films has been written by Ulman.[113] One of the most powerful techniques that can be applied in situ involves optical second harmonic generation (SHG). This method has been reviewed by Corn[114] and by Richmond,[115] and features submonolayer sensitivities, instantaneous response time, the ability to discriminate between surface and bulk species, and access

to both film thickness and orientation. Although Langmuir–Blodgett films adhere relatively weakly to electrode surfaces, many are still amenable to all standard surface analyses requiring high vacuum. One recent example included the secondary ion mass spectrometric analysis of cadmium arachidate films on various metal substrates, with little or no sample preparation being required.[116]

4.4 Analytical Applications

Although electroanalytical applications for Langmuir–Blodgett films are largely unrealized, it is clear that electrochemical research in this area has only just begun. Many reports have focused either on the dynamic transport of electroactive species across a bilayer membrane or on the enzymatic generation of electroactive species at the surface of a modified electrode. Results from these systems are included in section 5 below.

Nelson[117] prepared phosphatidylserine-coated mercury electrodes (HMDEs) to study the reduction of Tl^+, Cd^{2+}, Cu^{2+}, Pb^{2+}, and Eu^{3+} in aqueous systems. It was found that the organization of the film is potential dependent and that the reduction potentials of the various metal ions depend strongly on the extent of film protonation (pH) and the presence of other metal species. Nelson et al.[118] observed that a variety of common aqueous pollutants, including polynuclear aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs) and various pesticides, can be concentrated from contaminated water samples into dioleoylphospholipid monolayers on mercury; for PAHs, these investigators established that the detection limit is 0.4 $\mu g\ L^{-1}$, with a relative standard deviation of 7%.

In a unique application of the Langmuir–Blodgett technique, Fujihira and Poosittisak[109] employed a glassy carbon electrode filmed with an electroactive bipyridinium amphiphile and coated with platinum. This two-layer electrode catalyzed the reduction of oxygen when used for rotating-disk voltammetry, a fact which simultaneously attests to the versatility and stability of Langmuir–Blodgett films.

Finally, Miller and Bard[111] designed a special trough to accommodate the filming of a small horizontally oriented ITO electrode. They demonstrated picomole sensitivities of this electrode toward alkylated $[Ru(bpy)_3]^{2+}$ species, noting that the current efficiency for the Langmuir–Blodgett method far exceeds that attainable by traditional casting methods with the same complexes.

5 IMMOBILIZED ENZYMES ON ELECTRODES

Chemical selectivity remains one of the great challenges facing analytical electrochemists today. In this regard, electrodes modified with enzymes (bioelectrochemical sensors) continue to hold the most promise for progress and are presently one of the most active areas of research in electrochemistry.[119–122] Over 3300 references (in English) from 1987 to mid-1999 were retrieved on this subject from the Institute of Scientific Information database alone.[123] These reports cover a host of enzymes (from multiple sources) and immobilization techniques for virtually every type of electrode. Common themes include (a) the effect of various deposition processes on the activity and stability of the enzyme(s), (b) the influence of any mediators and/or required cofactors on electronic communication between the enzyme and the electrode, (c) the sensitivity and selectivity of the modified electrode for a given substrate (especially within a complex matrix), (d) the effect of polymers, membranes, and various other supports on the performance of the electrode system, and (e) the overall stability and applicability of the modified electrode system. With this in mind, our goal is merely to provide the reader with an entry point into the literature of enzyme-modified electrodes along with only the most basic of overviews.

5.1 Types of Materials and Methods of Preparation

Although enzymes typically do not exhibit reversible redox behavior at bare electrodes, they frequently do when immobilized in polymers, membranes, cross-linked networks, and a host of other media. Reasons for this behavior have been discussed by several authors.[124,125] In addition, it is usually necessary to employ "mediators" (small, organic, electron-transfer partners) to establish electronic communication between the electrode and the immobilized enzyme. Figure 8 illustrates a general scheme that applies to an electrode modified with an oxidase.

Usually, the link between electrode and enzyme is achieved with a combination of a synthetic mediator and an enzyme cofactor (e.g. reduced forms of nicotinamide adenine dinucleotide or flavin adenine dinucleotide). Enzyme-modified electrodes of all kinds are prepared from aqueous starting materials, but are not restricted to use in aqueous media.[126] A brief overview of some common materials and methods for the construction of enzyme-modified electrodes follows. Although this list is

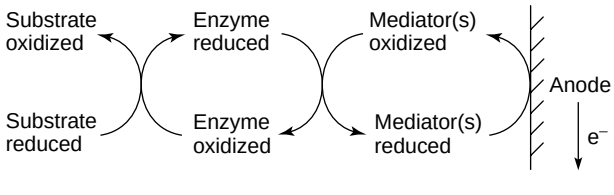

Figure 8 General format of an electrode modified with an oxidase and mediator(s).

far from complete, recent examples and reviews therein should serve as an adequate starting point.

5.1.1 Enzymes and Mediators

Glucose sensors based on GOx derived from *Aspergillus niger* form the cornerstone of enzyme research in electrochemistry. Reports featuring a new bioelectrochemical sensor design typically employ GOx as a benchmark for the performance of the electrode – usually one notes the effect of ascorbate and uric acid since these are problematic interferents found in serum samples. Additional commonly employed enzymes include other oxidases (especially horseradish peroxidase),[127] dehydrogenases,[128] cytochromes,[125] and myoglobins.[125] For systems with synthetic mediators, derivatives of ferrocene appear to be the most common, but tetracyanoquinodimethane (TCNQ), TTF, ferricyanide, and ruthenium and osmium complexes are also frequently employed. Enzymes and the mediators used with them have been extensively reviewed.[119–122,124,126–130]

5.1.2 Polymers for Electrodeposition

Electroactive polymers provide the experimenter with a rapid and simple method for the immobilization of enzymes onto an electrode surface. Pyrroles and aromatic amines (especially diamines) are the most commonly employed monomers and may be polymerized onto an electrode from an aqueous solution containing the enzyme of interest. This results in the entrapment of enzymes (and any other added reagents) within the bulk of the polymer film, which can also serve as part of the transduction mechanism. For example, the conductivity of polypyrrole is pH sensitive,[131] so it is the medium of choice for enzyme–substrate interactions which proceed with a net change in pH. Applications of electroactive polymers (and incorporated mediators) to enzyme-modified electrodes have been reviewed.[129,130]

5.1.3 Chemical Cross-linking

Glutaraldehyde can be employed for the chemical cross-linking of enzymes in solution. Typically, a mixture of enzyme(s), mediator(s), and frequently a stabilizer (such as bovine serum albumin (BSA)) is treated with glutaraldehyde and then cast onto an electrode surface. A recent paper by Sarkar et al.[132] illustrates the use of glutaraldehyde in an amino acid sensor based on screen-printed carbon electrodes. Cross-linking and polymeric supports are often used either in tandem or sequentially to enhance the stability and selectivity of an electrode. Thus, Moore et al.[133] utilized a mixture of glutaraldehyde, poly(vinyl alcohol), and GOx in a voltammetric sensor for acetaminophen and norepinephrine. Nishizawa et al.[131] prepared layered amperometric sensors for penicillin by coating interdigitated platinum arrays with polypyrrole followed by cross-linking of penicillinase with BSA.

5.1.4 Covalent Attachment

When a glassy carbon electrode is treated with dilute nitric acid and dichromate at positive potentials (+2.2 V versus saturated calomel electrode (SCE)), carboxylic acid functionalities are rapidly formed along the surface. Carbodiimide reagents may then be employed in a two-step synthesis which results in the covalent attachment of an enzyme to the surface (via an amide linkage). Bourdillon et al.[134] have provided an example pertaining to GOx.

5.1.5 Immobilization in Self-assembled Monolayers

Willner and Ricklin[135] recently prepared an amperometric biosensor based on an enzyme covalently attached to both an electron-transfer mediator (pyrroloquinolinequinone (PQQ)) and a monolayer of cysteamine on gold. They demonstrated that PQQ in this configuration is capable of oxidizing the reduced form of nicotinamide adenine dinucleotide in solution, while being oxidized at the electrode, thus linking the electrode to the activity of the enzyme. Willner et al.[136] have summarized other methods pertaining to enzymes on SAMs.

5.1.6 Langmuir–Blodgett Films

Biotinylated phospholipids are readily assembled into monolayer Langmuir–Blodgett films. These films can then be treated with streptavidin-modified enzymes prior to transfer onto an electrode surface. Rehak et al.[137] have employed this technology to prepare a xanthine-selective electrode. Applications of Langmuir–Blodgett films for other biosensors have been reviewed by Osa and Anzai[138] and by Aizawa.[139]

5.1.7 Direct Adsorption

Certain carbon electrode surfaces (such as pyrolytic graphite edge planes) have been shown to adsorb enzymes from solution electrostatically. In some cases, this results in the direct communication of an enzyme with the electrode (allowing for potential-controlled enzyme activity). Using this approach, Hirst et al.[140] demonstrated a potential-dependent fumarate–succinate equilibrium with succinate dehydrogenase.

5.1.8 Lipid and Surfactant Membranes (Bilayers and Multilayers)

Rusling[125] and Nikolelis et al.[141] have summarized the application of casting techniques to the study

of membrane-bound proteins (especially cytochrome $P450_{cam}$ and myoglobin) on various electrode surfaces. Typically, phospholipids (or other synthetic surfactants) are initially cast onto the electrode surface from a volatile organic solvent (e.g. chloroform). As the solvent evaporates, the surfactant molecules spontaneously form multiple bilayer membranes on the surface. This modified electrode surface is then allowed to equilibrate (~20 min) with an aqueous solution of an enzyme, resulting in the incorporation of the protein into the bilayers. Alternatively, an aqueous vesicular dispersion of pure surfactant(s) is combined with a solution of the enzyme, and the mixture is cast onto the surface. As the water evaporates, vesicles in the mixture collapse, again resulting in the formation of multiple bilayers containing an enzyme. Lipid bilayers provide the enzyme with a more native environment, and so preserve the original structure and activity to a larger degree.

5.1.9 Zeolites and Clays

Recently, it has been demonstrated that certain zeolites and clays (immobilized onto the surfaces of electrodes) can act as effective hosts for both enzymes and electron-transfer mediators.[142] These ''reagentless'' sensors feature simple construction and enhanced strength, but are essentially restricted to neutral aqueous solutions. Responses to interferents may be diminished with further modifications of the zeolite surface.

5.2 Methods of Characterization

Primary concerns for any enzyme-modified electrode are for the activity and stability of the enzyme. To this end FIA (described elsewhere in this encyclopedia) is ordinarily employed. Response of the electrode to multiple injections over a period of minutes to weeks can indicate whether denaturing or leaching of enzyme or mediator occurs. FIA also affords rapid determination of the limit of detection and linear range, while revealing the response time of the electrode. In addition, known inhibitors and interferents can be introduced into a solution of substrate to show that the enzyme behaves normally and that the initial response can be recovered.

Several standard surface analytical techniques have also been applied to enzyme-modified electrodes. Recent investigations have included XPS,[143] SEM,[144] FTIR spectroscopy,[145] and STM.[146]

5.3 Analytical Applications

Immobilized enzymes impart unprecedented chemical selectivity to virtually any electrode and, in principle, allow for repeated use of the catalysts in continuous processes without contaminating the products. As a result, enzyme-modified electrodes are finding application in the food industry,[147,148] in clinical pharmacology (especially for glucose, lactate, urea, xanthine, glutamine, and various neurotransmitters),[149–151] and in environmental chemistry for the detection of pesticides[152,153] and pollutants.[154]

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
BHA	*tert*-Butylhydroxyanisole
BHT	*tert*-Butylhydroxytoluene
BSA	Bovine Serum Albumin
EPR	Electron Paramagnetic Resonance
EQCM	Electrochemical Quartz Crystal Microbalance
FIA	Flow Injection Analysis
FTIR	Fourier Transform Infrared
GOx	Glucose Oxidase
HMDE	Hanging Mercury Drop Electrode
HPLC	High-performance Liquid Chromatography
HREELS	High-resolution Electron Energy Loss Spectroscopy
ITO	Indium-doped Tin Oxide
LEEDS	Low-energy Electron Diffraction Spectroscopy
PAHs	Polynuclear Aromatic Hydrocarbons
PAN	Polyacrylonitrile
PCBs	Polychlorinated Biphenyls
PG	Propyl Gallate
PQQ	Pyrroloquinolinequinone
PVP	Poly(4-vinylpyridine)
SAM	Self-assembled Monolayer
SCE	Saturated Calomel Electrode
SEM	Scanning Electron Microscopy
SHG	Second Harmonic Generation
STM	Scanning Tunneling Microscopy
TBHQ	*tert*-Butylhydroquinone
TCNQ	Tetracyanoquinodimethane
TEM	Transmission Electron Microscopy
TFCE	Trifluorochloroethylene
TFMSA	Trifluoromethane Sulfonic Acid
THF	Tetrahydrofuran
TTF	Tetrathiafulvalene
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Voltammetry In Vivo for Chemical Analysis of the Living Brain • Voltammetry In Vivo for Chemical Analysis of the Nervous System

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Glucose Measurement

Environment: Water and Waste **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis

Field-portable Instrumentation **(Volume 4)**
Electrochemical Sensors for Field Measurements of Gases and Vapors

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Ion-selective Electrodes: Fundamentals • Self-assembled Monolayers on Electrodes • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

REFERENCES

1. R.W. Murray (ed.), *Techniques of Chemistry*, Wiley-Interscience, New York, Vol. 22, 1992.
2. C.J. Zhong, M.D. Porter, 'Evidence for Carbon–Sulfur Bond Cleavage in Spontaneously Adsorbed Organosulfide-based Monolayers at Gold', *J. Am. Chem. Soc.*, **116**(25), 11 616–11 617 (1994).
3. C.J. Zhong, M.D. Porter, 'Designing Interfaces at the Molecular Level', *Anal. Chem.*, **67**(23), 709A–715A (1995).
4. W.A. Zisman, 'Relation of the Equilibrium Contact Angle to Liquid and Solid Constitution', *Adv. Chem. Ser.*, **43**, 1–51 (1964).
5. D.L. Allara, R.G. Nuzzo, 'Spontaneously Organized Molecular Assemblies. 1. Formation, Dynamics, and Physical Properties of *n*-Alkanoic Acids Adsorbed from Solution on an Oxidized Aluminum Surface', *Langmuir*, **1**(1), 45–52 (1985).
6. D.L. Allara, R.G. Nuzzo, 'Spontaneously Organized Molecular Assemblies. 2. Quantitative Infrared Spectroscopic Determination of Equilibrium Structures of Solution-adsorbed *n*-Alkanoic Acids on an Oxidized Aluminum Surface', *Langmuir*, **1**(1), 52–66 (1985).
7. H. Ogawa, T. Chihara, K. Taya, 'Selective Monomethyl Esterification of Dicarboxylic Acids by Use of Monocarboxylate Chemisorption on Alumina', *J. Am. Chem. Soc.*, **107**(5), 1365–1369 (1985).
8. N.E. Schlotter, M.D. Porter, T.B. Bright, D.L. Allara, 'Formation and Structure of a Spontaneously Adsorbed Monolayer of Arachidic Acid on Silver', *Chem. Phys. Lett.*, **132**(1), 93–98 (1986).
9. D.L. Allara, S.V. Atre, C.A. Elliger, R.G. Snyder, 'Formation of a Crystalline Monolayer of Folded Molecules by Solution Self-assembly of α,w-Alkanedioic Acids on Silver', *J. Am. Chem. Soc.*, **113**(5), 1852–1854 (1991).
10. Q. Liu, Z. Xu, 'Self-assembled Monolayer Coatings on Nanosized Magnetic Particles Using 16-Mercaptohexadecanoic Acid', *Langmuir*, **11**(12), 4617–4622 (1995).
11. D.Y. Huang, Y.T. Tao, 'Self-assembled Monolayer: Behavior of Diacetylenic Amphiphiles', *Bull. Inst. Chem., Acad. Sin.*, **33**, 73–80 (1986).
12. C.A. Goss, D.H. Charych, M. Majda, 'Application of (3-Mercaptopropyl)trimethoxysilane as a Molecular Adhesive in the Fabrication of Vapor-deposited Gold Electrodes on Glass Substrates', *Anal. Chem.*, **63**(1), 85–88 (1991).
13. R. Maoz, J. Sagiv, 'Penetration-controlled Reactions in Organized Monolayer Assemblies. 1. Aqueous Permanganate Interaction with Monolayer and Multilayer Films of Long-chain Surfactants', *Langmuir*, **3**(6), 1034–1044 (1987).
14. C.J. Miller, C.A. Widrig, D.H. Charych, M. Majda, 'Microporous Aluminum Oxide Films at Electrodes. 4. Lateral Charge Transport in Self-organized Bilayer Assemblies', *J. Phys. Chem.*, **92**(7), 1928–1936 (1988).
15. R.G. Nuzzo, D.L. Allara, 'Adsorption of Bifunctional Organic Disulfides on Gold Surfaces', *J. Am. Chem. Soc.*, **105**(13), 4481–4483 (1983).
16. R.G. Nuzzo, F.A. Fusco, D.L. Allara, 'Spontaneously Organized Molecular Assemblies. 3. Preparation and Properties of Solution Adsorbed Monolayers of Organic Disulfides on Gold Surfaces', *J. Am. Chem. Soc.*, **109**(8), 2358–2368 (1987).
17. E.B. Troughton, C.D. Bain, G.M. Whitesides, R.G. Nuzzo, D.L. Allara, M.D. Porter, 'Monolayer Films Prepared by the Spontaneous Self-assembly of Symmetrical and Unsymmetrical Dialkyl Sulfides from Solution onto Gold Substrates: Structure, Properties, and Reactivity of Constituent Functional Groups', *Langmuir*, **4**(2), 365–385 (1988).
18. G.M. Whitesides, P.E. Laibinis, 'Wet Chemical Approaches to the Characterization of Organic Surfaces: Self-assembled Monolayers, Wetting, and the Physical–Organic Chemistry of the Solid–Liquid Interface', *Langmuir*, **6**(1), 87–96 (1990).
19. J.N. Richardson, S.R. Peck, L.S. Curtin, L.M. Tender, R.H. Terrill, M.T. Carter, R.W. Murray, G.K. Rowe, S.E. Creager, 'Electron-transfer Kinetics of Self-assembled Ferrocene Octanethiol Monolayers on Gold and Silver Electrodes from 115 to 170 K', *J. Phys. Chem.*, **99**(2), 766–772 (1995).
20. Y. Feng, W.K. Teo, K.S. Siow, Z. Gao, K.L. Tan, A.K. Hsieh, 'Corrosion Protection of Copper by

a Self-assembled Monolayer of Alkanethiol', *J. Electrochem. Soc.*, **144**(1), 55–64 (1997).
21. Z. Mekhalif, J. Riga, J.J. Pireaux, J. Delhalle, 'Self-assembled Monolayers of *n*-Dodecanethiol on Electrochemically Modified Polycrystalline Nickel Surfaces', *Langmuir*, **13**(8), 2285–2290 (1997).
22. K. Slowinski, R.V. Chamberlain, R. Bilewicz, M. Majda, 'Evidence for Inefficient Chain-to-chain Coupling in Electron Tunneling Through Liquid Alkanethiol Monolayer Films on Mercury', *J. Am. Chem. Soc.*, **118**(19), 4709–4710 (1996).
23. M.A. Hines, J.A. Todd, P. Guyot-Sionnest, 'Conformation of Alkanethiols on Au, Ag(111), and Pt(111) Electrodes: a Vibrational Spectroscopy Study', *Langmuir*, **11**(2), 493–497 (1995).
24. A. Ulman, *Ultrathin Organic Films*, Academic Press, San Diego, CA, 1–99, 1991.
25. H.O. Finklea, 'Electrochemistry of Organized Monolayers of Thiols and Related Molecules on Electrodes', in *Electroanalytical Chemistry*, eds. A.J. Bard, I. Rubinstein, Dekker, New York, 109–335, Vol. 19, 1996.
26. D. Mandler, I. Turyan, 'Applications of Self-assembled Monolayers in Electroanalytical Chemistry', *Electroanalysis*, **8**(3), 207–213 (1996).
27. S. Steinberg, I. Rubinstein, 'Ion-selective Monolayer Membranes Based upon Self-assembling Tetradentate Ligand Monolayers on Gold Electrodes. 3. Application as Selective Ion Sensors', *Langmuir*, **8**(4), 1183–1187 (1992).
28. J.J. Hickman, D. Ofer, P.E. Laibinis, G.M. Whitesides, M.S. Wrighton, 'Molecular Self-assembly of Two-terminal, Voltammetric Microsensors with Internal References', *Science*, **252**(5006), 688–691 (1991).
29. I. Turyan, D. Mandler, 'Self-assembled Monolayers in Electroanalytical Chemistry: Application of *w*-Mercaptocarboxylic Acid Monolayers for Electrochemical Determination of Ultralow Levels of Cadmium(II)', *Anal. Chem.*, **66**(1), 58–63 (1994).
30. O. Chailapakul, R.M. Crooks, 'Synthesis and Characterization of Simple Self-assembling Nanoporous Monolayer Assemblies: a New Strategy for Molecular Recognition', *Langmuir*, **9**(4), 884–888 (1993).
31. J. Wang, H. Wu, L. Angnes, 'On-line Monitoring of Hydrophobic Compounds at Self-assembled Monolayer Modified Amperometric Flow Detectors', *Anal. Chem.*, **65**(14), 1893–1896 (1993).
32. C. Bourdillon, M. Majda, 'Microporous Aluminum Oxide Films at Electrodes. 7. Mediation of the Catalytic Activity of Glucose Oxidase via Lateral Diffusion of a Ferrocene Amphiphile in a Bilayer Assembly', *J. Am. Chem. Soc.*, **112**(5), 1795–1799 (1990).
33. O. Inganäs, B. Liedberg, W. Chang-Ru, 'A New Route to Polythiophene and Copolymers of Thiophene and Pyrrole', *Synth. Met.*, **11**(4), 239–249 (1985).
34. J.P. Ferraris, D.J. Guerrero, 'Recent Advances in Heteroaromatic Copolymers', in *Handbook of Conducting Polymers*, 2nd edition, eds. T.A. Skotheim, R.L. Elsenbaumer, J.R. Reynolds, Dekker, New York, 259–276, 1998.
35. J. Roncali, 'Synthetic Principles for Bandgap Control in Linear π-Conjugated Systems', *Chem. Rev.*, **97**(1), 173–205 (1997).
36. J. Rault-Berthelot, C. Rozé, M.M. Granger, E. Raoult, 'Anodic Oxidation of Various Arylene–Cyanovinylenes Made of Alternating Fluorenyl, Thienyl and/or Phenyl Units', *J. Electroanal. Chem.*, **466**(2), 144–154 (1999).
37. G.P. Evans, 'The Electrochemistry of Conducting Polymers', in *Advances in Electrochemical Science and Engineering*, eds. H. Gerischer, C.W. Tobias, VCH, Weinheim, 1–74, Vol. 1, 1990.
38. G. Inzelt, 'Mechanism of Charge Transport in Polymer-modified Electrodes', in *Electroanalytical Chemistry*, ed. A.J. Bard, Dekker, New York, 89–241, Vol. 18, 1994.
39. A.F. Diaz, J.I. Castillo, J.A. Logan, W.Y. Lee, 'Electrochemistry of Conducting Polypyrrole Films', *J. Electroanal. Chem.*, **129**(1), 115–132 (1981).
40. K. Doblhofer, K. Rajeshwar, 'Electrochemistry of Conducting Polymers', in *Handbook of Conducting Polymers*, 2nd edition, eds. T.A. Skotheim, R.L. Elsenbaumer, J.R. Reynolds, Dekker, New York, 531–588, 1998.
41. L.B. Coleman, 'Technique for Conductivity Measurements on Single Crystals of Organic Materials', *Rev. Sci. Instrum.*, **46**(8), 1125–1126 (1975).
42. K.K. Shiu, S.K. Pang, H.K. Cheung, 'Electroanalysis of Metal Species at Polypyrrole-modified Electrodes', *J. Electroanal. Chem.*, **367**(1), 115–122 (1994).
43. Y. Ikariyama, W.R. Heineman, 'Polypyrrole Electrode as a Detector for Electroinactive Anions by Flow Injection Analysis', *Anal. Chem.*, **58**(8), 1803–1806 (1986).
44. E. Wang, A. Liu, 'Polyaniline Chemically Modified Electrode for Detection of Anions in Flow-injection Analysis and Ion Chromatography', *Anal. Chim. Acta*, **252**(1), 53–57 (1991).
45. A. Galal, Z. Wang, A.E. Karagözler, H. Zimmer, H.B. Mark, Jr, P.L. Bishop, 'A Potentiometric Iodide (and Other) Ion Sensor Based on a Conducting Polymer Film Electrode. Part II. Effect of Electrode Conditioning and Regeneration Techniques', *Anal. Chim. Acta*, **299**(2), 145–163 (1994).
46. R. John, M. Spencer, G.G. Wallace, M.R. Smyth, 'Development of a Polypyrrole-based Human Serum Albumin Sensor', *Anal. Chim. Acta*, **249**(2), 381–385 (1991).
47. Y. Lin, G.G. Wallace, 'Development of a Polymer-based Electrode for Sensitive Detection of Dichloramine', *Anal. Chim. Acta*, **263**(1), 71–75 (1992).
48. J.M. Slater, J. Paynter, E.J. Watt, 'Multi-layer Conducting Polymer Gas Sensor Arrays for Olfactory Sensing', *Analyst*, **118**(4), 379–384 (1993).

49. P.J. Peerce, A.J. Bard, 'Polymer Films on Electrodes. Part II. Film Structure and Mechanism of Electron Transfer with Electrodeposited Poly(vinylferrocene)', *J. Electroanal. Chem.*, **112**(1), 97–115 (1980).
50. H.D. Abruña, P. Denisevich, M. Umaña, T.J. Meyer, R.W. Murray, 'Rectifying Interfaces Using Two-layer Films of Electrochemically Polymerized Vinylpyridine and Vinylbipyridine Complexes of Ruthenium and Iron on Electrodes', *J. Am. Chem. Soc.*, **103**(1), 1–5 (1981).
51. N. Oyama, F.C. Anson, 'Polymeric Ligands as Anchoring Groups for the Attachment of Metal Complexes to Graphite Electrode Surfaces', *J. Am. Chem. Soc.*, **101**(13), 3450–3456 (1979).
52. A.P. Brown, F.C. Anson, 'Molecular Anchors for the Attachment of Metal Complexes to Graphite Electrode Surfaces', *J. Electroanal. Chem.*, **83**(1), 203–206 (1977).
53. H.D. Abruña, T.J. Meyer, R.W. Murray, 'Chemical and Electrochemical Properties of 2,2′-Bipyridyl Complexes of Ruthenium Covalently Bound to Platinum Oxide Electrodes', *Inorg. Chem.*, **18**(11), 3233–3240 (1979).
54. R.D. Rocklin, R.W. Murray, 'Chemically Modified Carbon Electrodes. Part XVII. Metallation of Immobilized Tetra(aminophenyl)porphyrin with Manganese, Iron, Cobalt, Nickel, Copper, and Zinc, and Electrochemistry of Diprotonated Tetraphenylporphyrin', *J. Electroanal. Chem.*, **100**(1), 271–282 (1979).
55. H.D. Abruña, 'Coordination Chemistry in Two Dimensions: Chemically Modified Electrodes', *Coord. Chem. Rev.*, **86**(1), 135–189 (1988).
56. A. Haimerl, A. Merz, 'Ferrocene-modified Polypyrrole Films by Electrochemical Copolymerization', *Angew. Chem. Int. Ed. Engl.*, **25**(2), 180–181 (1986).
57. J.G. Eaves, R. Mirrazaei, D. Parker, H.S. Munro, 'Redox-active Films of Ferrocene Covalently Attached to Polypyrrole', *J. Chem. Soc., Perkin Trans. 2*, (4), 373–376 (1989).
58. A. Deronzier, J.C. Moutet, 'Functionalized Polypyrroles. New Molecular Materials for Electrocatalysis and Related Applications', *Acc. Chem. Res.*, **22**(7), 249–255 (1989).
59. S. Cosnier, A. Deronzier, J.C. Moutet, 'Oxidative Electropolymerization of Polypyridinyl Complexes of Ruthenium(II)-containing Pyrrole Groups', *J. Electroanal. Chem.*, **193**(1), 193–204 (1985).
60. A. Deronzier, J.M. Latour, 'A Poly(pyrrole – Nickel(II)-pyridiniumporphyrin)-modified Electrode', *J. Electroanal. Chem.*, **224**(1), 295–301 (1987).
61. G.J. Samuels, T.J. Meyer, 'An Electrode-supported Oxidation Catalyst Based on Ruthenium(IV). pH ''Encapsulation'' in a Polymer Film', *J. Am. Chem. Soc.*, **103**(2), 307–312 (1981).
62. A. Bettelheim, B.A. White, S.A. Raybuck, R.W. Murray, 'Electrochemical Polymerization of Amino-, Pyrrole-, and Hydroxy-substituted Tetraphenylporphyrins', *Inorg. Chem.*, **26**(7), 1009–1017 (1987).
63. C.M. Elliott, C.J. Baldy, L.M. Nuwaysir, C.L. Wilkins, 'Electrochemical Polymerization of 4-Methyl-4′-vinyl-2,2′-bipyridine-containing Metal Complexes: Polymer Structure and Mechanism of Formation', *Inorg. Chem.*, **29**(3), 389–392 (1990).
64. G.X. Wan, K. Shigehara, E. Tsuchida, F.C. Anson, 'Virtues of a Copolymer Containing Pyrrolidone and Iron Porphyrin Groups in the Catalysis of the Reduction of Dioxygen at Graphite Electrodes', *J. Electroanal. Chem.*, **179**(1), 239–250 (1984).
65. S.M. Geraty, J.G. Vos, 'Synthesis, Characterization, and Photochemical Properties of a Series of Ruthenium Containing Metallopolymers Based on Poly-*N*-vinylimidazole', *J. Chem. Soc., Dalton Trans.*, (12), 3073–3078 (1987).
66. C.D. Ellis, T.J. Meyer, 'Incorporation of Redox Couples into *p*-Chlorosulfonated Polystyrene Coated Electrodes by Chemical Binding', *Inorg. Chem.*, **23**(12), 1748–1756 (1984).
67. A. El Hourch, S. Belcadi, P. Moisy, P. Crouigneau, J.M. Léger, C. Lamy, 'Electrocatalytic Reduction of Oxygen at Iron Phthalocyanine Modified Polymer Electrodes', *J. Electroanal. Chem.*, **339**(1), 1–12 (1992).
68. J. Zagal, M. Páez, A.A. Tanaka, J.R. dos Santos, Jr, C.A. Linkous, 'Electrocatalytic Activity of Metal Phthalocyanines for Oxygen Reduction', *J. Electroanal. Chem.*, **339**(1), 13–30 (1992).
69. N. Kobayashi, P. Janda, A.B.P. Lever, 'Cathodic Reduction of Oxygen and Hydrogen Peroxide at Cobalt and Iron Crowned Phthalocyanines Adsorbed on Highly Oriented Pyrolytic Graphite Electrodes', *Inorg. Chem.*, **31**(25), 5172–5177 (1992).
70. H. Li, T.F. Guarr, 'Formation of Electronically Conductive Thin Films of Metal Phthalocyanines via Electropolymerization', *J. Chem. Soc., Chem. Commun.*, (13), 832–834 (1989).
71. C.P. Horwitz, R.W. Murray, 'Oxidative Electropolymerization of Metal Schiff-base Complexes', *Mol. Cryst. Liq. Cryst.*, **160**, 389–404 (1988).
72. L.A. Hofercamp, K.A. Goldsby, 'Surface-modified Electrodes Based on Nickel(II) and Copper(II) Bis(salicylaldimine) Complexes', *Chem. Mater.*, **1**(3), 348–352 (1989).
73. P. Audebert, P. Capdevielle, M. Maumy, 'Redox and Conducting Polymers Based on Salen-type Metal Units: Electrochemical Study and Some Characteristics', *New J. Chem.*, **16**(6), 697–703 (1992).
74. C.E. Dahm, D.G. Peters, J. Simonet, 'Electrochemical and Spectroscopic Characterization of Anodically Formed Nickel Salen Polymer Films on Glassy Carbon, Platinum, and Optically Transparent Tin Oxide Electrodes in Acetonitrile Containing Tetramethylammonium Tetrafluoroborate', *J. Electroanal. Chem.*, **410**(2), 163–171 (1996).
75. S.M. Geraty, D.W.M. Arrigan, J.G. Vos, 'The Use of Electrodes Coated with Ruthenium-containing Polymers

as Sensors for Iron(II)', *Anal. Chem. Symp. Ser.*, **25**, 303–308 (1986).
76. M.A. Ruiz, E. Garcia-Moreno, C. Barbas, J.M. Pingarrón, 'Determination of Phenolic Antioxidants by HPLC with Amperometric Detection at a Nickel Phthalocyanine Polymer Modified Electrode', *Electroanalysis*, **11**(7), 470–474 (1999).
77. A.P. Doherty, M.A. Stanley, D. Leech, J.G. Vos, 'Oxidative Detection of Nitrite at an Electrocatalytic [$Ru(bpy)_2$poly-(4-vinylpyridine)$_{10}$Cl]Cl Electrochemical Sensor Applied for the Flow Injection Determination of Nitrate Using a Cu/Cd Reductor Column', *Anal. Chim. Acta*, **319**(1), 111–120 (1996).
78. Y.H. Tse, P. Janda, H. Lam, A.B.P. Lever, 'Electrode with Electropolymerized Tetraaminophthalocyanatocobalt(II) for Detection of Sulfide Ion', *Anal. Chem.*, **67**(5), 981–985 (1995).
79. M.A. Ruiz, M.G. Blázquez, J.M. Pingarrón, 'Electrocatalytic and Flow-injection Determination of the Antioxidant *tert*-Butylhydroxyanisole at a Nickel Phthalocyanine Polymer Modified Electrode', *Anal. Chim. Acta*, **305**(1), 49–56 (1995).
80. T.F. Kang, G.L. Shen, R.Q. Yu, 'Voltammetric Behavior of Dopamine at Nickel Phthalocyanine Polymer Modified Electrodes and Analytical Applications', *Anal. Chim. Acta*, **354**(1), 343–349 (1997).
81. C.R. Martin, I. Rubinstein, A.J. Bard, 'Polymer Films on Electrodes. 9. Electron and Mass Transfer in Nafion® Films Containing $Ru(bpy)_3^{2+}$', *J. Am. Chem. Soc.*, **104**(18), 4817–4824 (1982).
82. M. Majda, L.R. Faulkner, 'Electrochemical Behavior of Tris(2,2′-Bipyridine)ruthenium Complexes in Films of Poly(styrenesulfonate) on Electrodes', *J. Electroanal. Chem.*, **169**(1), 77–95 (1984).
83. N. Oyama, F.C. Anson, 'Electrostatic Binding of Metal Complexes to Electrode Surfaces Coated with Highly Charged Polymeric Films', *J. Electrochem. Soc.*, **127**(1), 249–250 (1980).
84. W.J. Vining, T.J. Meyer, 'Redox Properties of the Water Oxidation Catalyst $(bpy)_2(H_2O)RuORu(H_2O)(bpy)_2^{4+}$ in Thin Polymeric Films. Electrocatalytic Oxidation of Cl^- to Cl_2', *Inorg. Chem.*, **25**(12), 2023–2033 (1986).
85. N. Oyama, T. Shimomura, K. Shigehara, F.C. Anson, 'Electrochemical Responses of Multiply Charged Transition Metal Complexes Bound Electrostatically to Graphite Electrode Surfaces Coated with Polyelectrolytes', *J. Electroanal. Chem.*, **112**(2), 271–280 (1980).
86. C.J. Pickett, K.S. Ryder, J.C. Moutet, 'Synthesis and Anodic Polymerization of an L-Cystine Derivatized Pyrrole; Copolymerization with a Tetraalkylammonium Pyrrole Allows Reduction of the Cystinyl Film to a Cysteinyl State that Binds Electroactive $\{Fe_4S_4\}^{2+}$ Centres', *J. Chem. Soc., Chem. Commun.*, (9), 694–697 (1992).
87. F. Nguyen, F.C. Anson, 'Electrochemical Examination of Electrostatic Attraction of Cations and the Ejection of Anions by Polyanionic Electrode Coatings Stable in Non-aqueous Solvents', *Electrochim. Acta*, **44**(2), 239–245 (1998).
88. N. Oyama, F.C. Anson, 'Electrostatic Binding of Metal Complexes to Electrode Surfaces with Highly Charged Polymeric Films', *J. Electrochem. Soc.*, **127**(1), 247–248 (1980).
89. F.C. Anson, J.M. Savéant, K. Shigehara, 'New Model for the Interior of Polyelectrolyte Coatings on Electrode Surfaces. Mechanisms of Charge Transport Through Protonated Poly(L-lysine) Films Containing $Fe^{III}(edta)^-$ and $Fe^{II}(edta)^{2-}$ as Counterions', *J. Am. Chem. Soc.*, **105**(5), 1096–1106 (1983).
90. R. Volf, T.V. Shishkanova, P. Matejka, M. Hamplová, V. Král, 'Potentiometric Anion Response of Poly[5, 15-bis(2-aminophenyl)porphyrin] Electropolymerized Electrodes', *Anal. Chim. Acta*, **381**(2), 197–205 (1999).
91. K. Yoshino, K. Yoshimoto, S. Morita, T. Kawai, S.H. Kim, K.L. Kang, S.K. Choi, 'Electrical and Electrochemical Properties of Polyacetylene Derivatives with Pendant Cationic Groups', *Synth. Met.*, **69**(1), 81–82 (1995).
92. C.J. Brumlik, A. Parthasarathy, W.J. Chen, C.R. Martin, 'Plasma Polymerization of Sulfonated Fluorochlorocarbon Ionomer Films', *J. Electrochem. Soc.*, **141**(9), 2273–2279 (1994).
93. A. Economou, P.R. Fielden, 'Adsorptive Stripping Voltammetry on Mercury Film Electrodes in the Presence of Surfactants', *Analyst*, **118**(11), 1399–1404 (1993).
94. J. Labuda, M. Vanìcková, E. Uhlemann, W. Mickler, 'Applicability of Chemically Modified Electrodes for Determination of Copper Species in Natural Waters', *Anal. Chim. Acta*, **284**(3), 517–523 (1994).
95. J.M. Zen, S.Y. Huang, 'Square-wave Voltammetric Determination of Lead(II) with a Nafion® 2,2′-Bipyridyl Mercury Film Electrode', *Anal. Chim. Acta*, **296**(1), 77–86 (1994).
96. I.G. Casella, M.R. Guascito, A.M. Salvi, E. Desimoni, 'Catalytic Oxidation and Flow Detection of Hydrazine Compounds at a Nafion®/Ruthenium(III) Chemically Modified Electrode', *Anal. Chim. Acta*, **354**(1), 333–341 (1997).
97. L.Q. Mao, Y. Tian, G.Y. Shi, H.Y. Liu, L.T. Jin, K. Yamamoto, S. Tao, J.Y. Jin, 'A New Ultramicrosensor for Nitric Oxide Based on Electropolymerized Films of Nickel Salen', *Anal. Lett.*, **31**(12), 1991–2007 (1998).
98. S.F. de Betono, A.A. Garcìa, J.F.A. Valentin, 'UV Spectrophotometry and Square-wave Voltammetry at Nafion®-modified Carbon Paste Electrode for the Determination of Doxazosin in Urine and Formulations', *J. Pharm. Biomed. Anal.*, **20**(4), 621–630 (1999).
99. J.M. Zen, C.B. Wang, 'Determination of Dissolved Oxygen by Catalytic Reduction on a Nafion®/Ruthenium Oxide Pyrochlore Chemically Modified Electrode', *J. Electroanal. Chem.*, **368**(1), 251–256 (1994).

100. P. Ugo, L.M. Moretto, S. Bellomi, V.P. Menon, C.R. Martin, 'Ion-exchange Voltammetry at Polymer Film-coated Nanoelectrode Ensembles', *Anal. Chem.*, **68**(23), 4160–4165 (1996).
101. R.A. Hann, 'Molecular Structure and Monolayer Properties', in *Langmuir–Blodgett Films*, ed. G. Roberts, Plenum Press, New York, 17–92, 1990.
102. M.K. DeArmond, G.A. Fried, 'Langmuir–Blodgett Films of Transition Metal Complexes', in *Progress in Inorganic Chemistry*, ed. K.D. Karlin, Wiley, New York, 97–142, 1997.
103. L.M. Goldenberg, 'Electrochemical Properties of Langmuir-Blodgett Films', *J. Electroanal. Chem.*, **379**(1), 3–19 (1994).
104. R.M. Swart, 'Monolayers and Multilayers of Biomolecules', in *Langmuir–Blodgett Films*, ed. G. Roberts, Plenum Press, New York, 273–316, 1990.
105. A.J. Fernández, M.T. Martìn, J.J. Ruiz, E. Muñoz, L. Camacho, 'Electrochemical Behavior of LB Films Containing a Mixture of Viologen and a Phospholipid', *J. Phys. Chem. B*, **102**(35), 6799–6803 (1998).
106. L.M. Goldenberg, C. Pearson, M.R. Bryce, M.C. Petty, 'Preparation and Characterization of Conductive Langmuir–Blodgett Films of a Tetrabutylammonium-Ni(dmit)$_2$ Complex', *J. Mater. Chem.*, **6**(5), 699–704 (1996).
107. A. Ulman, *Ultrathin Organic Films*, Academic Press, San Diego, CA, 124–125, 1991.
108. S. Nagase, M. Kataoka, R. Naganawa, R. Komatsu, K. Odashima, Y. Umezawa, 'Voltammetric Anion Responsive Sensors Based on Modulation of Ion Permeability Through Langmuir–Blodgett Films Containing Synthetic Anion Receptors', *Anal. Chem.*, **62**(13), 1252–1259 (1990).
109. M. Fujihira, S. Poosittisak, 'Precise Control of Amount of Electrodeposited Platinum by Using the Langmuir–Blodgett Film and its Application to Electrocatalysis of Molecular Oxygen Reduction', *Chem. Lett.*, (2), 251–252 (1986).
110. M. Fujihira, T. Araki, 'In Situ Electrochemical Measurements of the Redox-active Monolayer by the Horizontal Touching Method Under a Controlled Surface Pressure', *Chem. Lett.*, (6), 921–926 (1986).
111. C.J. Miller, A.J. Bard, 'Horizontal Touch Voltammetric Analysis: Determination of Insoluble Electroactive Species in Films at the Air/Water Interface', *Anal. Chem.*, **63**(17), 1707–1714 (1991).
112. A. Nelson, A. Benton, 'Phospholipid Monolayers at the Mercury/Water Interface', *J. Electroanal. Chem.*, **202**(1), 253–270 (1986).
113. A. Ulman, *Characterization of Organic Thin Films*, Butterworth-Heinemann, Stoneham, MA, 1995.
114. R.M. Corn, 'Optical Second Harmonic Generation Studies of Adsorption, Orientation, and Order at the Electrochemical Interface', *Anal. Chem.*, **63**(2), 285A–295A (1991).
115. G.L. Richmond, 'Optical Second Harmonic Generation as an In Situ Probe of Electrochemical Interfaces', in *Electroanalytical Chemistry*, ed. A.J. Bard, Dekker, New York, 87–180, Vol. 17, 1991.
116. N. Ogura, Y. Ichinohe, S. Yoshida, T. Watanabe, T. Hoshi, K. Endo, M. Kudo, 'Substrate Dependence of Secondary Ion Intensities from Langmuir–Blodgett Films Investigated by TOF-SIMS', *J. Surf. Anal.*, **5**(2), 318–321 (1999).
117. A. Nelson, 'Voltammetry of Tl^{I}, Cd^{II}, Cu^{II}, Pb^{II}, and Eu^{III} at Phosphatidylserine-coated Mercury Electrodes', *J. Chem. Soc., Faraday Trans.*, **89**(16), 3081–3090 (1993).
118. A. Nelson, N. Auffret, J. Readman, 'Initial Applications of Phospholipid-coated Mercury Electrodes to the Determination of Polynuclear Aromatic Hydrocarbons and Other Organic Micropollutants in Aqueous Systems', *Anal. Chim. Acta*, **207**(1), 47–57 (1988).
119. J. Wang, 'Electroanalysis and Biosensors', *Anal. Chem.*, **71**(12), 328R–332R (1999).
120. J. Janata, M. Josowicz, P. Vanýsek, D.M. DeVaney, 'Chemical Sensors', *Anal. Chem.*, **70**(12), 179R–208R (1998).
121. J.L. Anderson, L.A. Coury, Jr, J. Leddy, 'Dynamic Electrochemistry: Methodology and Application', *Anal. Chem.*, **70**(12), 519R–589R (1998).
122. M. Aizawa, 'New Trends in Bioelectrochemistry', in *New Challenges in Organic Electrochemistry*, ed. T. Osa, Gordon and Breach Science Publishers, Amsterdam, 339–356, 1998.
123. The Institute for Scientific Information databases are site-configured and can be accessed at http://www.webofscience.com/.
124. P.N. Bartlett, P. Tebbutt, R.G. Whitaker, 'Kinetic Aspects of the Use of Modified Electrodes and Mediators in Bioelectrochemistry', *Prog. React. Kinet.*, **16**(2), 55–155 (1991).
125. J.F. Rusling, 'Enzyme Bioelectrochemistry in Cast Biomembrane-like Films', *Acc. Chem. Res.*, **31**(6), 363–369 (1998).
126. L. Stancik, 'Amperometric Biosensors for Non-aqueous Media', *Chem. Listy*, **91**(1), 30–37 (1997).
127. T. Ruzgas, E. Csöregi, J. Emnéus, L. Gorton, G. Marko-Varga, 'Peroxidase-modified Electrodes: Fundamentals and Application', *Anal. Chim. Acta*, **330**(2), 123–138 (1996).
128. I. Katakis, E. Dominguez, 'Catalytic Electrooxidation of NADH for Dehydrogenase Amperometric Biosensors', *Mikrochim. Acta*, **126**(1), 11–32 (1997).
129. P.N. Bartlett, J. Cooper, 'Applications of Electroactive Polymers in Biochemistry and Bioelectronics', in *Electroactive Polymer Chemistry*, ed. M.E.G. Lyons, Plenum Press, New York, 233–267, Vol. 2, 1996.
130. T.W. Lewis, G.G. Wallace, M.R. Smyth, 'Electrofunctional Polymers: their Role in the Development of New Analytical Systems', *Analyst*, **124**(3), 213–219 (1999).

131. M. Nishizawa, T. Matsue, I. Uchida, 'Penicillin Sensor Based on a Microarray Electrode Coated with pH-responsive Polypyrrole', *Anal. Chem.*, **64**(12), 2642–2644 (1992).
132. P. Sarkar, I.E. Tothill, S.J. Setford, A.P.F. Turner, 'Screen-printed Amperometric Biosensors for the Rapid Determination of L- and D-Amino Acids', *Analyst*, **124**(6), 865–870 (1999).
133. T.J. Moore, G.G. Nam, L.C. Pipes, L.A. Coury, Jr, 'Chemically Amplified Voltammetric Enzyme Electrodes for Oxidizable Pharmaceuticals', *Anal. Chem.*, **66**(19), 3158–3163 (1994).
134. C. Bourdillon, J.P. Bourgeois, D. Thomas, 'Covalent Linkage of Glucose Oxidase on Modified Glassy Carbon Electrodes. Kinetic Phenomena', *J. Am. Chem. Soc.*, **102**(12), 4231–4235 (1980).
135. I. Willner, A. Ricklin, 'Electrical Communication between Electrodes and $NAD(P)^+$-dependent Enzymes Using Pyrroloquinolinequinone–Enzyme Electrodes in a Self-assembled Monolayer Configuration: Design of a New Class of Amperometric Biosensors', *Anal. Chem.*, **66**(9), 1535–1539 (1994).
136. I. Willner, E. Katz, B. Willner, 'Electrical Contact of Redox Enzyme Layers with Electrodes: Routes to Amperometric Biosensors', *Electroanalysis*, **9**(13), 965–977 (1999).
137. M. Rehak, M. Snejdarkova, M. Otto, 'Application of Biotin–Streptavidin Technology in Developing a Xanthine Biosensor Based on a Self-assembled Phospholipid Membrane', *Biosens. Bioelectron.*, **9**(4), 337–341 (1994).
138. T. Osa, J. Anzai, 'Preparation of Biosensors by Immobilizing Enzymes on Surface of Langmuir–Blodgett Membrane', *Hyomen*, **30**(12), 985–990 (1992).
139. M. Aizawa, 'LB (Langmuir–Blodgett) Film Technology for Biosensors', *Yukagaku*, **39**(3), 166–170 (1990).
140. J. Hirst, A. Sucheta, B.A.C. Ackrell, F.A. Armstrong, 'Electrocatalytic Voltammetry of Succinate Dehydrogenase: Direct Quantification of the Catalytic Properties of a Complex Electron-transport Enzyme', *J. Am. Chem. Soc.*, **118**(21), 5031–5038 (1996).
141. D.P. Nikolelis, T. Hianik, U.J. Krull, 'Biosensors Based on Thin Lipid Films and Liposomes', *Electroanalysis*, **11**(1), 7–15 (1999).
142. B. Liu, F. Yan, J. Kong, J. Deng, 'A Reagentless Amperometric Biosensor Based on the Co-immobilization of Horseradish Peroxidase and Methylene Green in a Modified Zeolite Matrix', *Anal. Chim. Acta*, **386**(1), 31–39 (1999).
143. A. Griffith, A. Glidle, G. Beamson, J.M. Cooper, 'Determination of the Biomolecular Composition of an Enzyme-polymer Biosensor', *J. Phys. Chem. B*, **101**(11), 2092–2100 (1997).
144. P.C. Pandey, S. Upadhyay, H.C. Pathak, 'A New Glucose Biosensor Based on Sandwich Configuration of Organically Modified Sol-gel Glass', *Electroanalysis*, **11**(1), 59–64 (1999).
145. B. Wang, B. Li, Z. Wang, G. Xu, Q. Wang, S. Dong, 'Sol-gel Thin-film Immobilized Soybean Peroxidase Biosensor for the Amperometric Determination of Hydrogen Peroxide in Acid Medium', *Anal. Chem.*, **71**(10), 1935–1939 (1999).
146. Q. Chi, J. Zhang, S. Dong, E. Wang, 'Direct Electrochemistry and Surface Characterization of Glucose Oxidase Adsorbed on Anodized Carbon Electrodes', *Electrochim. Acta*, **39**(16), 2431–2438 (1994).
147. W.A.C. Somers, W. Van Hartingsveldt, E.C.A. Stigter, J.P. Van Der Lugt, 'Bio-electrochemistry: a Powerful Tool for the Production of Ingredients in the Food Industry', *Agro Food Ind. Hi-Tech*, **8**(2), 32–35 (1997).
148. A. Maines, D. Ashworth, P. Vadgama, 'Enzyme Electrodes for Food Analysis', *Food Technol. Biotechnol.*, **34**(1), 31–42 (1996).
149. J. Wang, 'Amperometric Biosensors for Clinical and Therapeutic Drug Monitoring: a Review', *J. Pharm. Biomed. Anal.*, **19**(1), 47–53 (1999).
150. L.V. Sigolaeva, A.V. Eremenko, A. Makower, G.F. Makhaeva, V.V. Malygin, I.N. Kurochkin, 'A New Approach for Determination of Neuropathy Target Esterase Activity', *Chem.-Biol. Interact.*, **120**, 559–565 (1999).
151. Y. Hu, K.M. Mitchell, F.N. Albahadily, E.K. Michaelis, G.S. Wilson, 'Direct Measurement of Glutamate Release in the Brain Using a Dual Enzyme-based Electrochemical Sensor', *Brain Res.*, **659**(1), 117–125 (1994).
152. P. Skladal, 'Biosensors Based on Cholinesterase for Detection of Pesticides', *Food Technol. Biotechnol.*, **34**(1), 43–49 (1996).
153. T. Noguer, B. Leca, G. Jeanty, J.L. Marty, 'Biosensors Based on Enzyme Inhibition: Detection of Organophosphorus and Carbamate Insecticides and Dithiocarbamate Fungicides', *Field Anal. Chem. Tech.*, **3**(3), 171–178 (1999).
154. W. Del Carlo, M. Mascini, 'Immunoassay for Polychlorinated Biphenyls (PCB) Using Screen Printed Electrodes', *Field Anal. Chem. Tech.*, **3**(3), 179–184 (1999).

Self-assembled Monolayers on Electrodes

Harry O. Finklea
West Virginia University, Morgantown, USA

1 Introduction **1**
1.1 Definition of Self-assembled Monolayer 1
1.2 Types of Self-assembled Monolayers on Electrodes 2
1.3 Advantages of Self-assembled Monolayers Based on Thiols and Related Molecules 2
1.4 Scope and Previous Reviews 3
2 Preparation of Self-assembled Monolayers on Electrodes **3**
2.1 Substrate Choice and Preparation 3
2.2 Solution Adsorption 4
2.3 Other Deposition Methods 4
3 Nonelectrochemical Characterization of Self-assembled Monolayers **5**
4 Postdeposition Modification of Self-assembled Monolayers **5**
5 Electrochemical Characterization of Self-assembled Monolayers on Electrodes **6**
5.1 Double-layer Structure and Capacitance 6
5.2 Electrochemical Stripping and Deposition of Self-assembled Monolayers 7
5.3 Blocking Behavior 8
5.4 Attached Redox Centers 9
6 Electrochemical Applications of Self-assembled Monolayers on Electrodes **11**
6.1 Microarray Electrodes 11
6.2 Selective Permeation 12
6.3 Preconcentration and Selective Binding 12
6.4 Electrocatalysis 13
6.5 Long-range Electron Transfer 14
6.6 Corrosion Control 15
6.7 Adhesion Control 15
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Alkanethiols and related molecules spontaneously adsorb from solution or vapor phase onto oxide-free metals, especially gold, to form close-packed oriented monolayers. The ease and flexibility of the self-assembly process provides a convenient method for altering the properties of the metal as an electrode. The close-packed hydrocarbon layer blocks access of most solution species to the electrode surface. Consequently, interfacial capacitances are markedly reduced and most electron-transfer reactions are strongly inhibited. The monolayers also provide a foundation for attachment of additional molecules or molecular layers to electrodes. Self-assembled monolayers (SAMs) on electrodes have been used in voltammetric methods to improve the analytical sensitivity and to impart greater selectivity towards specific analytes. Redox molecules have been anchored at fixed and controllable distances from the electrode surface, thereby permitting the study of electron-transfer kinetics over long distances and at large driving forces. The ability of SAMs to inhibit metal corrosion and to promote better adhesion of electroactive polymers has been demonstrated.

1 INTRODUCTION

1.1 Definition of Self-assembled Monolayer

An organized SAM is a single layer of molecules on a substrate in which the molecules exhibit a high degree of orientation, molecular order and packing (Figure 1). There are two common methods for depositing a monolayer. In the Langmuir–Blodgett (LB) method, amphiphilic molecules are spread on the air–water interface, compressed laterally, and transferred to the substrate either by dipping the substrate through the interface or by touching it to the interface.[2] In the self-assembly method, the monolayer spontaneously forms upon exposure of the substrate to a solution or vapor containing the molecules. Successful self-assembly requires a relatively strong bond between the substrate and an atom or moiety in the molecule, and an additional lateral interaction between molecules in the monolayer. The strength of the head group–substrate bonds, the lateral interactions and the density of packing result in sufficient stability that the monolayer resists removal by a solvent rinse. Unlike the popular LB monolayers, SAMs remained relatively obscure until the 1980s when several researchers discovered that long-chain thiols and disulfides spontaneously formed remarkably well-packed and stable monolayers on gold substrates.[3–9] Because SAMs provide a facile means of defining the chemical composition and structure of a surface (see below), they have become the focus of intensive investigation. Potential technological applications can be

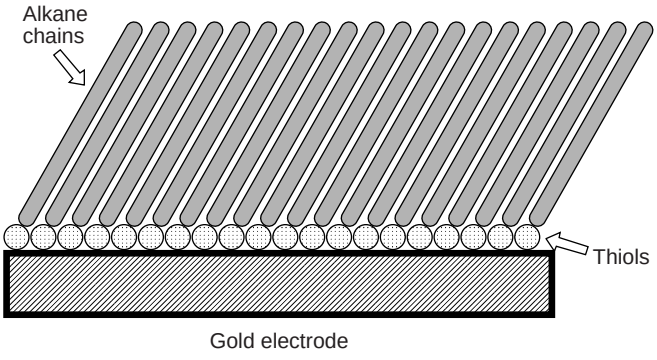

Figure 1 An organized monolayer on a substrate (electrode). The monolayer can be deposited by the LB method or by self-assembly. (Reprinted from Finklea,[1] p. 110, by courtesy of Marcel Dekker, Inc.)

found in areas such as wetting, lubrication, adhesion, corrosion, biocompatibility, catalysis, chemical sensing and nanoscale lithography.

1.2 Types of Self-assembled Monolayers on Electrodes

SAMs have been formed on every common electrode material with the possible exception of carbon. The common electrode metals (gold, silver, platinum and mercury) have all been coated with SAMs containing sulfur compounds, especially thiols, disulfides, and sulfides. The sulfur compounds most commonly contain pendant alkane chains of varying lengths. Less common are SAMs based on adsorption of isonitriles.[10,11] Chlorosilanes with long alkyl chains self-assemble on doped metal oxides such as SnO_2,[12] silicon with a thin oxide coating[13] and even gold without any surface oxide.[14,15] Generation of alkane radicals near an oxide-free silicon surface results in the formation of a densely packed SAM in which a methylene carbon is directly bonded to a silicon surface atom.[16,17]

1.3 Advantages of Self-assembled Monolayers Based on Thiols and Related Molecules

The first advantage is the ease with which SAMs are formed when gold and other metals are exposed to thiols and related molecules (in subsequent discussion, the word thiols will also imply disulfides and sulfides). A monolayer is deposited on the metal in a matter of seconds to minutes. The self-assembly method does not require anaerobic or anhydrous conditions; nor does it require a vacuum. Self-assembly is relatively insensitive to the choice of solvent. While organic-free metal surfaces are desirable, the high affinity of the sulfur for the metal enables the assembling layer to displace more weakly adsorbed impurities. Curvature or accessibility of the metal surface is not a factor; substrates can range from macroscopic to submicroscopic, and from smooth to highly porous.

A second advantage arises from the affinity of the sulfur for the metal and the strength of the bond formed. SAMs survive prolonged exposure to vacuum. It is possible to have a wide range of functional groups in the adsorbing molecule without disrupting the self-assembly process or destabilizing the SAM. Considering just the family of w-substituted alkanethiols, the terminal substituent can be an alkane (linear, branched, perfluorinated, perdeuterated), alkene, alkyne, aromatic, halide, ether, alcohol, aldehyde, carboxylic acid, amide, ester, amine or nitrile.[18–20] The ''body'' of the molecule can contain, for example, heteroatoms,[21–24] aromatic groups,[25–27] conjugated unsaturated links[28–30] and other rigid rod structures,[31–33] sulfones[24,34] and amides.[35,36] If the SAM is uniform in composition and densely packed, then a single functional group is exposed on the external surface. This property permits the exploration of the effect of surface composition on such surface-sensitive properties as wetting, friction and adhesion.[20,37–40] The diversity of sulfur-based SAMs, their autophobic behavior and the slow exchange with solution molecules allows the preparation of micropatterned surfaces in which the SAM is laterally heterogeneous.[41]

Mixed SAMs (Figure 2) can be prepared by depositing different molecules simultaneously or sequentially. When

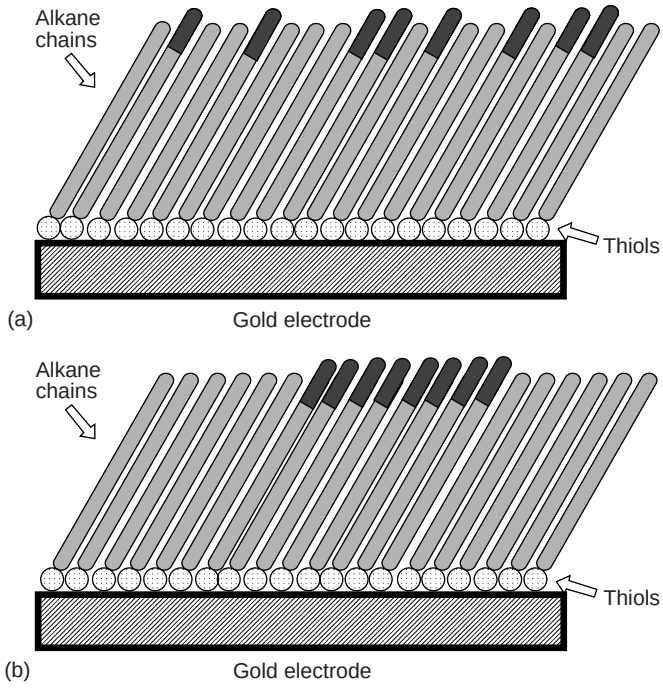

Figure 2 Two mixed SAMs exhibiting homogeneous mixing (a) and phase separation (b). Differences in chain length and terminal groups are shown.

both molecules are present in the deposition solution, the mole fraction of each molecule in the SAM can be controlled via the mole ratio of the two molecules in the solution, the identity and temperature of the solvent and the time of the deposition.[42] The distribution of the two molecules may vary from intimately mixed to completely phase-separated. It is possible to deposit SAMs with a 1:1 mole ratio of two chains and terminal groups with presumably perfect mixing via asymmetric sulfides or disulfides or thiols with two pendant chains.[43] In sequential deposition, a second thiol is incorporated into an existing SAM by prolonged immersion of the substrate in the second deposition solution.

In electrochemistry, the utility of thiol-based SAMs arises from their ability to survive the electrochemical experiment. The sulfur atoms resist oxidation, reduction and desorption. SAMs on electrodes are stable over a wide range of potentials and electrolyte compositions (especially aqueous electrolytes). They afford a means of controlling the electrode/electrolyte interfacial properties and the accessibility of the electrode surface to solution molecules. SAMs also provide a means of attaching to the electrode a diverse set of structures ranging from modified monolayers to multilayers. Specific applications include the development of more selective and sensitive electrochemical sensors (especially biosensors), control of faradaic reaction mechanisms, and a better understanding of the factors controlling electron transfer over long distances and under large driving forces.

1.4 Scope and Previous Reviews

The scope of this article will be primarily limited to SAMs composed of thiols and related molecules on metal electrodes. Since there is already a vast literature on the characterization of SAMs on metals, and since the findings therein are relevant to the behavior of SAMs on electrodes, a brief overview of nonelectrochemical characterization of SAMs follows the section on SAM preparation. Postdeposition modification reactions are summarized before discussion of the electrochemical characterization of the SAMs on electrodes. The final section covers a range of electrochemical applications of SAMs.

A previous review of SAMs on electrodes provides a more in-depth treatment of many topics in this article.[1] Two books by Ulman survey the field of monolayers prepared by both LB and self-assembly methods.[2,44] Reviews by Nuzzo cover both the use of SAMs as model surfaces and their technological applications.[45,46] Other discussions of applications of SAMs have been provided by Zhong and Porter,[47] Whitesides and Laibinis,[37] and Wink et al.[48]

2 PREPARATION OF SELF-ASSEMBLED MONOLAYERS ON ELECTRODES

2.1 Substrate Choice and Preparation

Gold is the most popular substrate for thiol SAMs. Owing to its noble character, gold substrates can be handled in air without the formation of an oxide surface layer, and can survive harsh chemical treatments which remove organic contaminants. Electrodes may be composed of bulk gold, either polycrystalline or single crystal, or thin films deposited on various substrates. To some extent, the choice of substrate is dictated by the application of the SAM. For the most highly oriented crystalline SAM, a single crystalline and atomically flat substrate is desired. When the SAM's function is to block access of molecules to the gold, a minimum density of pinholes is sought. While the origin of pinholes in SAMs is still unclear, impurities such as inorganic oxides are a likely source. Hence, the purity of the gold surface appears to be more important than its smoothness.[49] Both requirements are met by evaporated or sputtered gold films (typically 50–200 nm thick) on glass, silicon or cleaved mica substrates. To improve the adhesion of the gold film to the oxide surface, a thin (1–5 nm) layer of chromium or titanium is often deposited first. There is evidence that defects in the underlayer of Cr or Ti can be a source of pinholes in the final SAM.[50] If the deposited film is annealed at high temperatures, then the gold surface is largely composed of Au(111) domains and is frequently labeled as a Au(111) substrate. A comparison of grain size and crystallinity of thermally evaporated and sputtered gold on glass, silicon and cleaved mica substrates has been provided by Golan et al.[51] Some subtle effects of the substrate identity and the gold deposition protocol on the behavior of the subsequently deposited SAMs have been noted.[52–54] Procedures are available for the preparation of ultraflat[55,56] or optically transparent gold substrates.[57,58]

Cleaning and etching steps are often part of the deposition protocol. For bulk gold, heating the gold in a gas/air flame results in a hydrophilic surface, indicating that all organics have been removed. Electrochemical cycling into the oxide formation region in dilute acid functions as both a cleaning and an annealing process; the resulting voltammogram provides an assay of the cleanliness and crystallinity of the gold and its true surface area.[1] Etching the bulk gold in dilute aqua regia removes polishing damage and improves the blocking properties of the SAM.[59] Evaporated or sputtered gold films are frequently immersed into the thiol deposition solution as soon as they are removed from the gold deposition unit. Despite the fact that the gold surfaces are hydrophobic and therefore contaminated, SAM

deposition proceeds unimpeded. Alternatively, organic contaminants can be removed via exposure of the gold to a powerful oxidant. Popular oxidants include ''piranha'' solution (a 1 : 3 mixture of 30% hydrogen peroxide and concentrated sulfuric acid at ca. 100 °C) (*Caution: this mixture reacts violently with organic material and has been known to explode when stored in closed containers!*), an oxygen plasma and ozone generated by UV (ultraviolet) light. These treatments leave a surface oxide which can become trapped under a self-assembling thiol monolayer, and which can affect the properties of the SAM deleteriously.[60,61] An ethanol rinse rapidly removes the gold oxide.

The other coinage metals, silver and copper, and platinum have been used as substrates for thiol SAMs. For freshly evaporated silver or copper thin films, surface oxides exist prior to SAM deposition. The self-assembly process appears to remove the oxide layer from silver but not from copper, yielding lower quality SAMs on the latter substrate.[62–64] The oxide layer can be removed by an acid soak[65] or electrochemical reduction.[66] A recent innovation has been to form a single atomic layer of copper or silver on a gold surface by underpotential deposition (UPD) and to assemble the monolayer on the UPD layer.[67–70] There is growing evidence that the bond between the sulfur and the UPD atom is stronger than the bond between the sulfur and gold atoms, resulting in a more stable SAM. It is even possible to form a UPD Cu layer after deposition of a short-chain alkanethiol SAM.[71]

Liquid mercury avoids the issues of surface crystallinity, roughness and morphology, but is easily oxidized in the presence of thiols. It is possible to prepare homogenous and densely packed SAMs of alkanethiols on mercury via solution deposition,[72–75] but not by vapor deposition.[76]

Other substrates for thiol SAMs include nickel,[77,78] indium tin-oxide,[79] indium phosphide,[80] and a Tl-Ba-Cu-O high-temperature superconductor.[81]

2.2 Solution Adsorption

Immersion of the substrate into a homogenous solution of the self-assembling molecule at room temperature followed by rinsing is the most common approach for depositing the SAM. Usually any solvent capable of dissolving the molecule is suitable. Ethanol is the most popular solvent.[19] The thiol concentration can be varied from micromolar levels to that of the neat thiol liquid. Very low concentrations of thiols are favored by those seeking large crystalline domains of alkanethiols via slow self-assembly,[82] but are not a guarantee of SAM quality.[83] There is evidence that the solvent can be trapped in the SAM and not removed by the subsequent rinsing step.[19,84] Insoluble thiols can be dispersed into water with the aid of surfactants[85,86] or cyclodextrins[87] and subsequently assembled on gold. Considerable effort has been made to ascertain the time needed to form a well-organized layer. For millimolar or higher concentrations of thiols, a disordered monolayer is deposited in a few seconds. There is a much slower transformation over a period of hours to days into a highly oriented and densely packed monolayer.[19,50] Scanning tunneling microscopy (STM) studies (see section 3) suggest that the gold is slightly etched during self-assembly; gold has been detected in the deposition solution.[88,89] To remove kinetically trapped disordered states, the SAM is sometimes annealed by soaking it in hot deposition solution,[90] exposing it to warm temperatures in a gaseous ambient,[91] or subjecting the electrode to a cyclic voltammogram (CV) and repeated immersions in the deposition solution.[92]

When two thiols are being co-deposited, the solvent, time and temperature all affect the mole fraction of each component in the SAM.[42,64,93,94] The less soluble thiol is preferentially deposited. High concentrations of the thiol, long adsorption times and high temperatures encourage the SAM composition to approach equilibrium with the solution; the result is an abrupt transition of the SAM composition from one component to the other as a function of the mole ratio of the two components in solution.[42]

2.3 Other Deposition Methods

Deposition of the SAM from an electrolyte with the substrate under potential control ensures that the metal is in the reduced state. A thiol SAM desorbs at very negative potentials in alkaline electrolytes (section 5.2). Consequently, a SAM can be formed in an alkaline electrolyte with dissolved alkanethiols by slowly shifting the electrode potential in the positive direction from the desorption potential.[95–97] When the thiol is sufficiently volatile, SAM deposition can be performed from the vapor phase, both in a vacuum[98] and at ambient pressures.[99] The resulting SAMs appear to be virtually identical to those obtained by solution deposition. Majda et al. have demonstrated that monolayers very similar to those obtained by self-assembly can be obtained by spreading a mixture of alkanethiol and long-chain alkanol on a LB trough, applying compression and transferring the monolayer to a gold substrate.[100–102] There are numerous approaches to producing micropatterned SAMs, but the most common is to press a patterned polysiloxane stamp inked with a thiol onto a gold substrate and then to fill in the gaps with solution deposition of a second thiol.[41]

3 NONELECTROCHEMICAL CHARACTERIZATION OF SELF-ASSEMBLED MONOLAYERS

SAMs have been subjected to virtually every surface analytical method known. Among the more frequently applied tools are wetting contact angle, ellipsometry, surface plasmon resonance (SPR) spectroscopy, surface IR (infrared) spectroscopy, Raman spectroscopy, X-ray absorption spectroscopy, X-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES), temperature programmed desorption (TPD), scanning electron microscopy (SEM), surface ionization mass spectroscopy (SIMS) and laser desorption mass spectroscopy (LDMS), STM and atomic force microscopy (AFM), assorted diffraction (X-ray, electron, atom) and piezoelectric methods. The following discussion is a summary of observations from all of these methods.

The mechanism of self-assembly appears to follow two stages. Initially, alkanethiols adsorb horizontally onto the metal substrate. Subsequently, the thiols lift up to form the vertically oriented layer. The thickness of the SAM is linearly related to the chain length of the thiol. On single-crystal surfaces, SAMs generally form highly crystalline lattices which are commensurate with the metal atom lattice. Chain ''melting'' to a more liquid-like structure has been reported in several instances, but other observations indicate only a slow increase in the degree of chain disorder as the temperature is raised. The exact binding sites and nature of bonding between the sulfur and the surface metal atoms is still not clear. The Au–S bond strength is about $40\,\mathrm{kcal\,mol^{-1}}$.[18] On the most common metal substrate, Au(111), alkanethiols form a $(\sqrt{3}\times\sqrt{3})R30^\circ$ hexagonal lattice with an average spacing of 5.0 Å (1 Å $= 10^{-10}$ m) between alkane chains. The alkane chains are in the *trans*-conformation with very few gauche defects. The average tilt of the alkane chain with respect to the surface normal is typically less than or equal to 30°. The tilt angle is controlled by the headgroup spacing combined with a minimization of free volume in the alkane chain domain. A detectable $C(4\times 2)$ unit cell of four thiols within the lattice arises from variations in the twist of the alkane chains. In highly ordered SAMs, the orientation of the terminal group can depend on the number of methylenes separating the terminal group from the thiol (the ''odd–even'' effect). Domains of SAMs with nanometer dimensions are separated by grain boundaries corresponding to changes in chain tilt and/or registry with the substrate surface atoms. Even when the surface of the substrate is not ordered or when a large group in the chain perturbs the packing, the exposed alkane chains often adopt a hexagonal close-packed structure. There is considerable evidence that thiols lose the hydrogen to form a surface thiolate. Disulfides cleave to form the identical thiolate. Surface metal atoms should carry a corresponding positive charge, but spectroscopic evidence for the oxidized metal layer has not been found. An X-ray diffraction study[103] suggests that the thiolates exist as a dimer on a Au(111) surface, but there has been no other supporting evidence for this hypothesis. The degree of molecular order is very similar in SAMs prepared from thiols and from disulfides, but noticeably lower in SAMs prepared from sulfides. A controversial proposal by Zhong and Porter suggests that the C–S bond can also cleave during the self-assembly process of sulfides.[104] STM images reveal the presence of ''pits'' after SAM deposition from either solution or vapor. These pits have been identified as missing single atom layers in flat terraces. The pits contain attached thiolates and are not the origin of pinholes discussed in section 5.3. Postdeposition thermal annealing causes migration of the surface gold atoms, and promotes Ostwald ripening of the ''pits''. The SAMs are air stable unless they are exposed to ozone, in which case the thiolates are rapidly oxidized to sulfonates.[66]

Most of the characterization methods indicate that mixed SAMs composed of molecules with similar chain lengths are homogeneously mixed rather than phase separated. Only in high-resolution STM images is there evidence of imperfect mixing.[105,106] When the mixed SAM contains molecules of greatly differing chain length, phase separation into microscopic domains is detectable. The longer-chain component exhibits a greater degree of disorder since it is not stabilized by packing over the entire surface.

Slight increases in disorder of the outer part of the SAM can be detected when the SAM is brought into contact with a liquid, depending on the liquid and the terminal group of the SAM. SAMs with hydrophilic surface groups, particularly an ethylene oxide oligomer, are very resistant to adsorption of proteins from solution.[22,57] Thiols with terminal acid or base groups exhibit marked changes in the ionization behavior when they are incorporated into SAMs. The ionization reaction which forms a charged species is less favored thermodynamically. Thus, alkyl carboxy groups (–COOH), which in homogeneous solution have a pK_a of about five, do not start to ionize significantly in SAMs until the pH of the contacting solution exceeds six. Complete ionization of the carboxy groups requires a much higher pH.

4 POSTDEPOSITION MODIFICATION OF SELF-ASSEMBLED MONOLAYERS

Because a diverse variety of moieties can be attached to the thiol molecule without inhibiting SAM formation,

there is an abundance of chemical transformations available to tailor the surface composition and properties. For both bond cleavage and bond formation on a SAM, both steric factors and electrostatic attraction/repulsion strongly affect the reactivity of the terminal groups. A topic not covered here is the physical transformation of SAMs via lithographic methods into patterned arrays.

Thiol exchange is accomplished by immersing the SAM in a new thiol deposition solution. The kinetics of exchange indicate the presence of two populations of molecules in the SAM, one of which is rapidly exchanged. In general, the second population is not fully replaced by the new thiol even after prolonged exposure (days) to the second deposition solution.

Ester groups near the exposed surface of a SAM can be hydrolyzed by small nucleophiles,[107,108] but not by an enzyme.[109]

Internal polymerizations of SAMs include conversion of diacetylenes to a conjugated ene–yne structure,[110,111] cross-linking of a terminal vinyl group,[112] hydrolysis and cross-linking of a trimethoxysilane,[113] dehydration of a boronic acid to a borate glass,[114] and electrochemical polymerization of a terminal pyrrole.[115,116] The polymerized SAMs exhibit increased stability towards thermal desorption and greater resistance to exchange with solution thiols.

Numerous examples exist of reactions designed to couple a new moiety covalently to the surface of a SAM. Among the more common covalent transformations are the formation of amides and esters via acid chlorides or anhydrides or with the aid of carbodiimide coupling agents. Proteins, enzymes and other biomolecules are attached via reaction of exposed amines to SAMs with terminal COOH. Strategies for more site-specific binding of biomolecules take advantage the biotin–streptavidin reaction,[117] the high affinity of a dihistidine tag for Ni^{2+} complexed with nitriloacetic acid[118,119] and the reconstitution of flavin or nicotinamide adenine dinucleotides (NAD) with their enzymes.[120,121] Pendant DNA oligomers on a SAM can be hybridized with their complementary oligomers with remarkable specificity.[122,123]

SAMs with ionized terminal groups electrostatically attract ions of the opposite charge. For polyions (polymers with charge groups, charged planar inorganic oxide fragments and proteins), the binding is sufficiently strong that the attached layer resists removal by a water rinse. Because a bound polyion reverses the surface charge on an ionized SAM, it is possible to build up multiple layers of polyions with sequential immersion of the substrate in aqueous solutions of each component.[124,125] Multilayers can be also constructed on SAMs via sequential deposition of a metal cation and then a difunctional molecule.[126,127]

A second highly oriented layer of surfactant molecules can be deposited on the top of a SAM via LB transfer. The orientation of the outer layer depends on whether the SAM surface is hydrophobic[128] or hydrophilic.[129]

The importance of cell membrane phospholipid bilayers in controlling ion transport and signalling has led to a considerable effort to create analogous structures on or near solid substrates. Phospholipids spontaneously form a second layer on top of a methyl-terminated SAM from either a suspension of phospholipid vesicles,[130] a suspension of phospholipid and detergent micelles[131] or from the evaporation of a phospholipid solution.[132] The resulting structure is called a hybrid bilayer. To avoid the obvious disadvantage of a bilayer with one side pinned to the substrate, suspended bilayers have been formed by creating a SAM with a remote pendant phospholipid[133–135] or cholesterol[136] and assembling the bilayer around the remote seed molecules.

5 ELECTROCHEMICAL CHARACTERIZATION OF SELF-ASSEMBLED MONOLAYERS ON ELECTRODES

5.1 Double-layer Structure and Capacitance

Electrode/electrolyte interfaces exhibit a capacitance whose magnitude reflects the distribution of ions on the solution side of the interface. The electrolyte double layer is composed of the Helmholtz layer, a layer of ions and solvent in physical contact with the electrode, and the diffuse layer, a layer of ions near the electrode whose concentration deviates from bulk concentrations. In relatively concentrated electrolytes, the capacitance of the Helmholtz layer dominates the interfacial capacitance. For most metals, typical Helmholtz capacitances range from $10–100\,\mu F\,cm^{-2}$, and are dependent on potential. When a long-chain alkanethiol self-assembles on a metal electrode, the Helmholtz layer changes from a mixture of ions and solvent with a high dielectric constant to an ion-free hydrocarbon layer with a low dielectric constant. Consequently, the interfacial capacitance is dramatically reduced and becomes virtually independent of potential.

Interfacial capacitances are often measured via the charging current in a CV. More detailed studies of capacitance behavior are obtained by AC (alternating current) impedance spectroscopy or AC voltammetry. A simple parallel-plate capacitor model predicts that the reciprocal capacitance increases linearly with the thickness of the dielectric layer. Plots of C^{-1} versus n (number of CH_2 groups in the alkane chain) are indeed linear for the longer chain lengths. Shorter chain-length SAMs appear to be permeable to some electrolyte

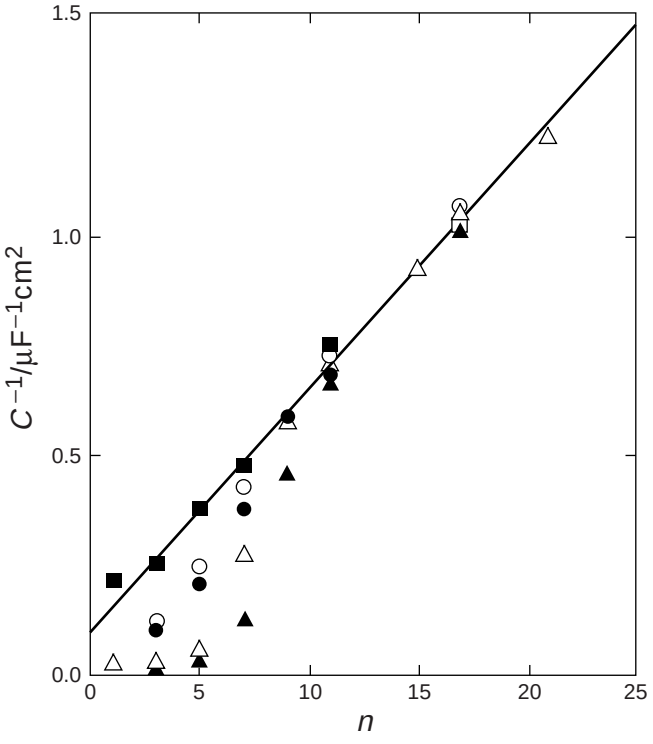

Figure 3 Reciprocal capacitance of alkanethiol SAMs versus chain length *n*. The symbols represent capacitances obtained from CVs. Filled symbols indicate 10 mV s^{-1} and empty symbols indicate 100 mV s^{-1} scan rate. △, ▲ 1 M KCl; ○, ● 1 M $HClO_4$; ◯, ■ 1 M NaF. (Reprinted with permission from Porter et al.,[6] Copyright 1987, American Chemical Society.)

ions and exhibit greater capacitances than expected (Figure 3).[6,85,137,138] Extraction of the SAM dielectric constant requires an assumption about the tilt of the alkane chains. Observed values fall in the range of 2.3–2.6. The dielectric constants are consistent with a close-packed layer of alkane chains with essentially no penetration by electrolyte solvent or ions. Mathematical models have been developed to predict the capacitances of a SAM-coated electrode when the surface charge of the SAM changes because of either an acid–base reaction or a redox reaction.[139–141] Deviations from the capacitances of simple alkanethiol SAMs are useful in monitoring phenomena such as poorly packed SAMs, increased solvent and/or ion penetration and variations in the tilt of the alkane chains. Capacitance measurements are also helpful in monitoring the self-assembly process in situ and in confirming the formation of a hybrid bilayer on a SAM.

5.2 Electrochemical Stripping and Deposition of Self-assembled Monolayers

SAMs from thiols, disulfides and sulfides resist desorption over a wide potential range, but at very negative potentials and in strongly alkaline electrolytes, they are desorbed quantitatively.[26,52,72,96,97,137,142] On single crystal surfaces or on mercury, a typical CV exhibits a rather sharp cathodic peak (or sometimes two peaks), corresponding to stripping of the thiols, and a somewhat broader anodic peak corresponding to readsorption of the thiol (Figure 4). The half reaction is thus written as Equation (1):

$$\mathbf{M}|S(CH_2)_nX + \mathbf{n}e^- \rightleftharpoons \mathbf{M}| + {}^-S(CH_2)_nX \qquad (1)$$

where **M** is the electrode metal. The desorption proceeds either from a few nucleation centers or nearly homogeneously across the electrode, depending on the chain length of the thiol and the applied potential.[143]

A key observation is that the area under the desorption peak, corrected for charging current and true surface area, is independent of the chain length of the alkanethiol. The area, converted to moles of electrons per cm^2, is 7.8×10^{-10} for Au(111), 7.7×10^{-10} for Ag(111), and 9.8×10^{-10} for Hg. The theoretical coverages of thiols (in moles cm^{-2}) for a $(\sqrt{3} \times \sqrt{3})R30°$ lattice are 7.7×10^{-10} for Au(111) and 7.6×10^{-10} for Ag(111). The data support the conclusion that $\mathbf{n} = 1$ in Equation (1).

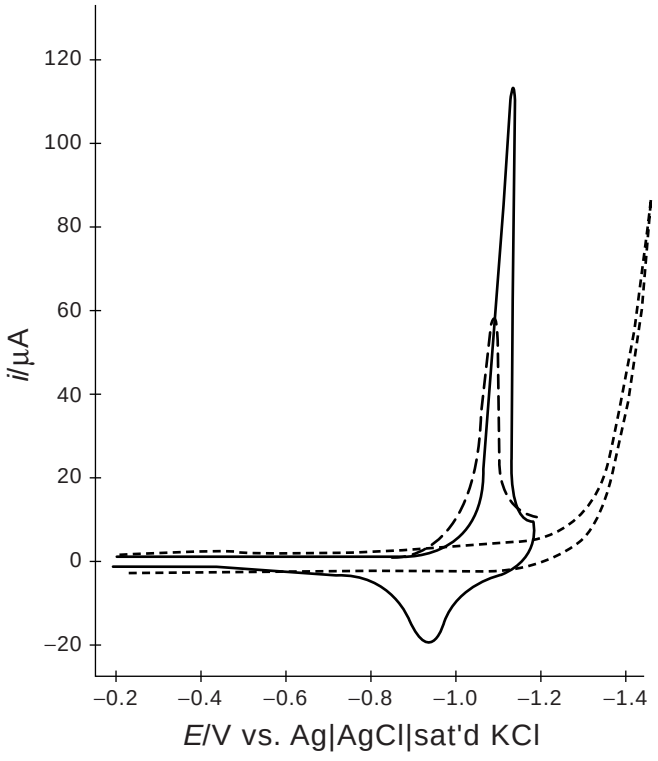

Figure 4 Electrochemical stripping and redeposition of a dodecanethiol SAM on Au(111) in 0.5 M KOH. The solid and dashed lines are the first and second scans, respectively. The dotted line is a CV of bare Au(111). (Reprinted with permission from Walczak et al.,[142] Copyright 1991, American Chemical Society.)

Schneider and Buttry argue that the correction for charging current is underestimated, so that **n** is smaller than 1.[144] Even with this caveat, the area of the cathodic stripping peak is a useful measure of thiol coverage as a function of SAM composition[26,104] and electrochemical history.[145] For example, a racemic deposition mixture of an optically active thiol yields a higher coverage on mercury than do deposition solutions containing either pure enantiomer.[146]

The potentials of the cathodic stripping peaks supply information about the strength of the metal–sulfur bond and the interchain interaction, the accessibility of the metal–sulfur bond to cations from the solution, and the presence of any intermediate or weakly adsorbed states.[96] Gold substrates with a high density of steps connecting Au(111) terraces exhibit two stripping waves, suggesting that thiols are more strongly bonded to step sites.[52] The presence of two stripping peaks for a SAM formed from an asymmetrical sulfide is part of the evidence for cleavage of the C–S bond in the sulfide.[104] In some mixed SAMs, distinct stripping peaks can be assigned to each component and that component thiol selectively desorbed; both observations support the existence of phase-separated domains.[147,148]

The desorbed thiols tend to remain near the electrode and are readily readsorbed when the electrode potential is shifted to a more positive value. It is possible to create partial monolayers by controlling the electrode potential prior to and during emersion.[95]

5.3 Blocking Behavior

Alkanethiol SAMs suppress faradaic processes such as electrode oxidation and the exchange of electrons between the electrode and solution redox couples. This blocking property is attributed to the densely packed structure of the hydrocarbon chains which impede the approach of solution ions and molecules to the electrode surface. Possible applications of blocking SAMs can be found in the areas of corrosion prevention, nanoscale lithography and selective electrodes. However, closer examination of SAM-coated electrodes reveals the presence of pinholes (bare metal sites) and other defects, which permit a close approach of solution species. It is important to understand the nature, size and distribution of the pinholes and other defects before the applications mentioned above are achievable.

Of all the methods for probing SAM structures, voltammetry is the most sensitive tool for detection of pinholes and defects in a SAM. The extraordinary sensitivity arises from the ability to detect currents corresponding to oxidation or reduction of a fraction of a monolayer, and from the high rates of mass transfer of redox couples to small "hot spots" on an otherwise blocked electrode. For the subsequent discussion, we define the area fraction of pinholes as $1 - \Theta$, where Θ is the fractional coverage relative to a complete monolayer. Gold oxidation is suppressed by the thiol SAM, except at pinholes. Consequently, a measurement of $1 - \Theta$ can be obtained from the charge needed to reduce the oxide.[7] Cyclic voltammetry of solution redox couples at electrodes with imperfect SAMs exhibits current peaks or small current plateaus at low overpotentials (near the formal potential $E^{0'}$) (Figure 5). These current peaks or plateaus greatly exceed the extremely small tunneling currents found on electrodes with perfect SAMs (section 6.5). The degree of attenuation of the current by the SAM at low overpotentials relative to the currents at a bare electrode is often used as a qualitative assessment of pinholes and defects in the SAM.

Quantitative measurements of $1 - \Theta$ require a model. One model treats the pinholes as an array of microelectrodes in an insulating plane. Given the somewhat

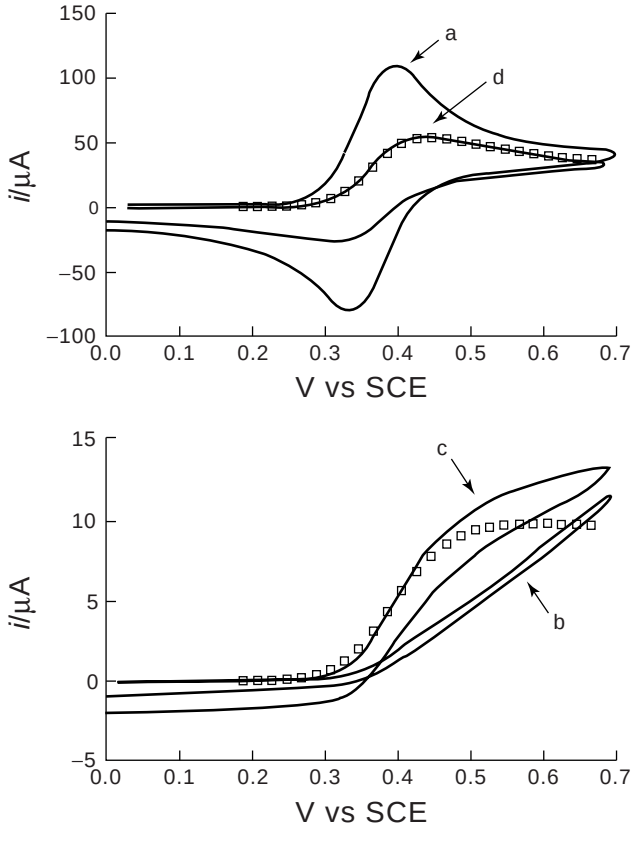

Figure 5 Electrochemical blocking of a solution redox couple by an octadecanethiol SAM. Ferrocenylmethyl-trimethylammonium in 0.5 M H_2SO_4, polycrystalline gold, 0.1 V s^{-1} scan rate. Curve (a) is the reversible CV obtained on bare gold, curves (b), (c) and (d) are obtained on the same electrode with different monolayers. The symbols are theoretical fits to a microarray electrode model. (Reprinted with permission from Finklea et al.,[149] Copyright 1993, American Chemical Society.)

unrealistic assumptions of uniform size and uniform spacing for the microelectrodes, CVs can be fitted with calculated voltammograms (the symbols in Figure 5) to obtain the pinhole parameters (the pinhole radius R_a, the radius of the blocked area around the pinhole R_0, and $1 - \Theta$).[149] In alternating current impedance spectroscopy (ACIS), the electrode is biased at the Nernst potential (at or close to $E^{0'}$) in an electrolyte containing both the oxidized and the reduced forms of a redox couple and the impedance of the system is measured as a function of frequency. The same model is used to interpret the data.[149] Typical values obtained are 10^{-2} to 10^{-4} for $1 - \Theta$ and micrometer to submicrometer dimensions for R_a. Deviations between the impedance behavior predicted by the model and the observed impedance behavior suggest that the pinholes are not uniformly distributed in size and in spacing.

Alternatives to the electrochemical approach for pinhole detection and mapping include deposition of metals[150] or etching the substrate metal in the presence of the SAM and then scanning the surface with STM, AFM or SEM.[50,151] These experiments suggest that the pinholes are of atomic dimensions, subnanometer rather than micrometers. Scanning electrochemical microscopy (SECM), a method which maps electroactive ''hot spots'' on an electrode, indicates that pinholes are less than 0.5 μm in diameter.[152] It is likely that the apparent submicrometers to micrometers dimensions of pinholes obtained by voltammetric methods arise from areas with a high density of subnanometer pinholes.

Unfortunately, no procedure has been developed which reliably yields pinhole-free SAMs. A survey of the literature indicates that the least defective SAMs are obtained most consistently on freshly deposited thin films of metal or on mercury. For bulk solid metal electrodes, surfaces should be annealed, cleaned of organic contaminants, and etched to expose fresh metal and to remove inorganic oxides, a potential source of pinholes.[153] In general, flexible chains yield better blocking properties than rigid ones. Postdeposition procedures that reduce the defectiveness of a SAM include electrochemical polymerization of phenol[154] and polymerization of a diacetylene SAM.[110] However, polymerization of a terminal vinyl group has the opposite effect.[112]

Nonaqueous electrolytes tend to increase the permeability of SAMs, presumably because of increased solvent interactions with the pendant chains. Some redox couples (e.g. ferrocenemethanol, benzoquinone) are not blocked by even pinhole-free SAMs in aqueous electrolytes by virtue of their hydrophobicity. Electrostatic attraction or repulsion between surface moieties on the SAM and solution redox couples have a powerful effect on the blocking behavior of SAMs. These last two observations form the basis of the creation of selective electrodes (section 6.2).

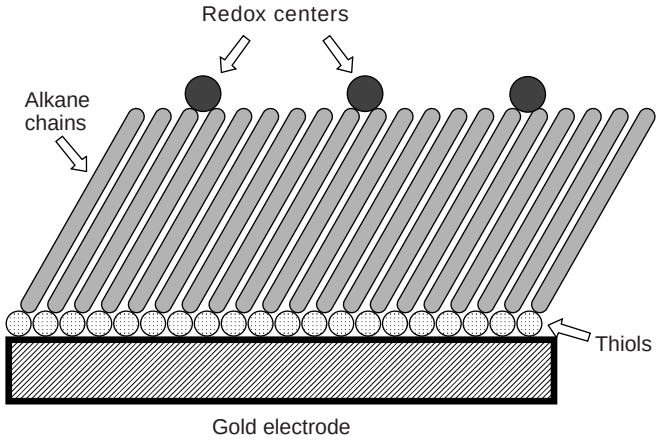

Figure 6 An electroactive SAM containing both redox and diluent thiols. (Reprinted from Finklea,[1] p. 240, by courtesy of Marcel Dekker, Inc.)

5.4 Attached Redox Centers

SAMs afford a highly flexible and convenient method for attaching redox centers to electrodes. The coverage of the redox center can be varied without lowering the packing density of the SAM, thereby controlling the lateral spacing between redox centers (Figure 6). The spacing between the electrode and the redox center is determined by the chain length; close packing prevents motion of the redox center towards the electrode. It is possible to control the local environment near the redox center by either partially burying the redox center in the hydrocarbon domain, or by varying the terminal group of diluent thiols. In particular, proteins are less likely to suffer denaturing structural changes when they are attached to a hydrophilic SAM surface compared to their interaction with a bare metal. The redox centers serve as a sensitive and nondestructive probe of the structure of SAMs (especially in mixed SAMs), their dynamics (thiol exchange) and their reactivity (surface attachment and hydrolysis reactions). For electron-transfer kinetic studies, electroactive SAMs are much less sensitive to pinholes and other defects than the blocking SAMs. The redox centers provide a means of catalyzing electron transfer to kinetically slow species (see section 6.4). A wide range of redox centers have been incorporated into SAMs, including ferrocenes, ruthenium and osmium complexes, viologens, porphyrins, heme proteins, (hydro)quinones, azobenzenes and fullerenes. Several strategies exist for the formation of multilayers of redox centers.

There are two main approaches to creating an electroactive SAM. The redox center may be prepared with a pendant thiol prior to deposition, or it may be coupled to the SAM via electrostatic binding, or amide or ester bond formation after deposition. In either approach, there are a

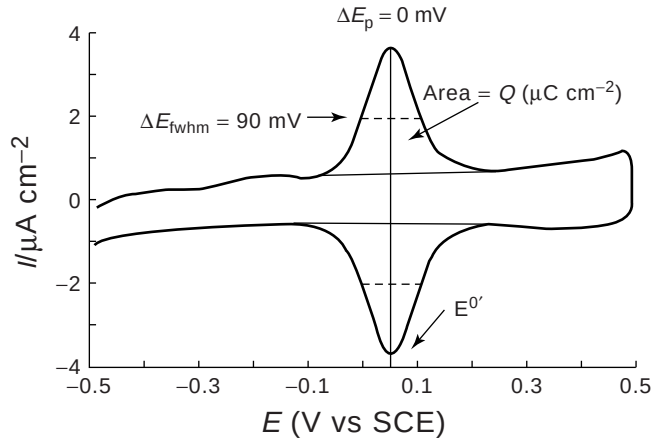

Figure 7 Reversible CV of an electroactive SAM. The redox thiol is $HS(CH_2)_{11}CONHCH_2pyRu(NH_3)_5^{2+/3+}$ and the diluent thiol is $HS(CH_2)_{11}COOH$. The electrolyte is 1 M Na_2SO_4 adjusted to pH 4. A linear charging current baseline is extrapolated under the faradaic waves to obtain the peak areas. (Reprinted from Finklea,[1] p. 241, by courtesy of Marcel Dekker, Inc.)

number of issues that need to be addressed: the coverage of the redox center, the formal potential, the shape of the CV compared to that of an ideal voltammogram and causes for deviations from the ideal shape.

A CV of an electroactive SAM usually exhibits matching anodic and cathodic current peaks (Figure 7). The coverage of the redox center Γ on the electrode surface is obtained by current integration after extrapolating and subtracting the charging current baseline underneath either the anodic or the cathodic current peak, Equation (2).

$$\Gamma = \frac{Q}{\mathbf{n}FA} \tag{2}$$

Q is the peak area in coulombs, $\mathbf{n}$ is the number of electrons transferred in the redox center half reaction, F is Faraday's constant, and A is the electrode area. The peak area and coverage are independent of the scan rate. Coverages down to 10^{-12} mol cm^{-2} can be measured.

The CV is reversible if its shape and the peak positions do not change with scan rate. The peak current i_p is proportional to the scan rate v, Equation (3)

$$i_p = \frac{\mathbf{n}^2F^2A\Gamma v}{4RT} \tag{3}$$

For an ideally shaped CV, the current obeys Equation (4):

$$i = \frac{4i_p e^{\theta}}{(1+e^{\theta})^2} \tag{4}$$

where θ is given by Equation (5)

$$\theta = \frac{\mathbf{n}F(E-E^{0'})}{RT} \tag{5}$$

The anodic and cathodic peak potentials are identical ($\Delta E_p = 0$) and are equal to $E^{0'}$. The peak half-width ΔE_{fwhm} is $3.53RT/\mathbf{n}F$, or $90.6/\mathbf{n}$ mV at 25 °C.

Departures from the ideal CV diagnostics are common and reveal molecular detail about the redox center and/or the SAM. A nonzero peak splitting indicates possible intermolecular interactions between the redox centers or a change in SAM structure with respect to the oxidation state of the redox centers (e.g. precipitation of the redox center with a counter ion). ΔE_{fwhm} can be either greater or smaller than the theoretical value. Large values of ΔE_{fwhm} may be caused by a spread of formal potentials (possibly indicating a disorganized SAM structure with a range of local environments about the redox centers), intermolecular interactions between redox centers or double-layer effects. In the latter case, the change in charge density of the SAM surface causes a variation in the local electrostatic potential.[139,155] The distortion of the CV is reduced when the coverage of redox centers is low and/or ion-pairing reduces the change in charge density with oxidation state. Sharp CVs with small ΔE_{fwhm} values are obtained if the redox centers tend to form a solid-like phase; this occurs only in the most densely packed electroactive SAMs.[156]

The formal potential is also informative. Generally, attached redox couples on the external surface of the SAM exhibit $E^{0'}$ values within 100 mV of the $E^{0'}$ of a solution analog, suggesting that the local solvation and dielectric constant are similar to that found at a bare electrode. For high coverages of redox centers with at least one charged oxidation state, $E^{0'}$ shifts at a Nernstian rate ($60/\mathbf{n}$ mV) with each decade change in concentration of the oppositely charged counter ion in the electrolyte.[157] This shift is due to the migration of the counter ion in or out of the SAM to provide charge compensation during the oxidation or reduction. The ion migration is also detected as mass changes for SAMs on quartz crystal microbalances (QCMs).[158,159] Formal potentials shift markedly as the redox center becomes progressively more buried in the hydrocarbon domain of the SAM.[160] The direction of the shift is consistent with the destabilization of the more highly charged oxidation state. Ion pairing is invoked to explain shifts in $E^{0'}$ with respect to the identity of the electrolyte counter ions.[160]

A brief overview of the behavior of some of the more common redox centers will be given; see Table 1 for characteristic parameters.

Ferrocene derivatives constitute the largest class of attached redox centers. Maximum coverages in SAMs are consistent with a close-packed layer of ferrocenes. The formal potential can be shifted over a 0.5 V range, depending on the presence of electron-donating or

Table 1 Properties of common redox centers attached to SAMs

Redox center	Max. Γ[a]	$E^{0'}$ vs SCE	Comments on peak shapes	Refs.
Ferrocenes	4–5	+0.2 to +0.7 V	Ranges from narrow to ideal to broad	79, 158, 161
$pyRu(NH_3)_5$	1–2	0 V	Nearly ideal at all coverages	92, 162
Cytochrome *c*	0.2	0 V	Slightly broader than ideal	163–165
Microperoxidase-11	2	−0.4 V	Broad CVs	166, 167
Viologens	4	−0.3 to −0.5 V −0.8 to −1.0 V	Two waves, broad, distorted by dimerization reaction	159
Quinones	3–5	pH-sensitive	Very broad	168
Azobenzenes	3–5	pH-sensitive	Very broad, sensitive to packing	53, 169, 170

[a] Coverage $\times 10^{-10}$ mol cm^{-2}.

electron-withdrawing substitutents. Thiol exchange reactions in SAMs have been monitored by ferrocene thiols with distinguishable formal potentials. The reduced ferrocene is chemically stable, but the oxidized ferricenium is subject to nucleophilic attack by hydroxide with concomitant loss of coverage, so acidic electrolytes are preferred. The peak shape of 100% ferrocene thiol SAMs is usually distorted by double-layer effects or the formation of a solid-like phase. However, carefully prepared mixed SAMs containing a low density of ferrocene thiols exhibit ideal CV shapes, suggesting that the redox centers reside on the external surface of the SAM.[161]

The $pyRu(NH_3)_5^{2+/3+}$ (py = pyridine) redox center has been thoroughly studied.[92] The reduced coverage relative to that of ferrocene SAMs reflects the electrostatic repulsion between cationic redox centers and the need to incorporate counter ions in the SAM. Like ferrocene, the Ru redox centers are unstable in the oxidized state in alkaline electrolytes. Unlike the ferrocenes, nearly ideal CV peak shapes are obtained at all coverages in aqueous electrolytes. In nonaqueous electrolytes, the CV peak shapes become broader and more distorted; the change is reversible upon reimmersion in aqueous electrolytes.[162]

CVs of attached metalloporphyrins are generally poorly defined unless the porphyrin ring is incorporated into a protein or protein fragment. Cytochrome *c* has been studied extensively. This protein can be attached via amide formation or by electrostatic binding.[163] It is believed that the electrostatic binding orients the heme-containing cleft towards the SAM surface. The observation that ΔE_{fwhm} is consistently larger than 100 mV has been interpreted in terms of a Gaussian spread of formal potentials, suggesting a heterogeneity in the orientation of the proteins.[164,165] Microperoxidase-11 is an oligopeptide containing a heme group and several carboxy groups which are usually attached to a cystamine SAM.[166] This protein fragment is a useful electrocatalyst.[167]

Viologens (4,4′-bipyridiniums) exhibit two distinct one-electron waves at negative potentials. In SAMs, the viologens yield high coverages and nonideal CV shapes due in part to a tendency of the radical cation to dimerize and the hydrophobicity of the two reduction products. The CVs are sensitive to the identity and concentration of the counter ions.[159] Quinones and azobenzenes generally exhibit high and stable coverages, very slow electron-transfer kinetics and very broad peaks, due in part to the dense packing achieved by these neutral organic redox centers. The formal potentials show the expected Nernstian shift with pH for a $2H^+$, $2e^-$ half reaction.[168,169] In very densely packed SAMs, azobenzenes are electrochemically silent, presumably because of strong inhibition of ion ingress and the significant structural change during reduction.[170]

6 ELECTROCHEMICAL APPLICATIONS OF SELF-ASSEMBLED MONOLAYERS ON ELECTRODES

The following discussion focuses on applications of SAMs for the enhancement of signal-to-noise in voltammetry, the development of selective electrodes, electrocatalysis, the study of electron-transfer kinetics over long distances and at large driving forces and the control of corrosion and adhesion at electrodes.

6.1 Microarray Electrodes

Microarray electrodes offer advantages for both analytical and kinetic studies because the electrode capacitance is greatly reduced and mass transfer is greatly enhanced relative to surface faradaic processes. The resulting improved signal-to-noise on a SAM-coated electrode has been demonstrated for the voltammetry of $Ru(bpy)_3^{2+/3+}$ (bpy = 2, 2′-bipyridine) in the presence of gold oxidation and oxide reduction.[171] However, for practical applications, it is necessary to control the density and size of the pinholes in the SAM. Mixed SAMs composed of a long-chain alkane thiol and a short thiol (3-mercaptoproprionic acid,[148] 4-aminothiophenol[172]) form phase-separated

domains of the short thiol with dimensions on the order of $10^4\,nm^2$ under the right conditions. Reproducible microarray behavior has been obtained for mixed SAMs of hexadecanethiol and 4-hydroxythiophenol.[153,173] It is probable that the ''microelectrodes'' are phase-separated domains of 4-hydroxythiophenol. At low microelectrode density, the mixed SAM shows a preferential response to neutral redox species over charged ones, suggesting the presence of adsorption or permselectivity through the hydrophobic SAM. A similar microarray electrode has been fabricated by LB deposition of a mixture of octadecanethiol plus octadecanol containing a small percentage of ubiquinone.[102,174] In this system, the density of ''molecular gates'' is readily controlled. At ubiquinone coverages less than $10^{-15}\,mol\,cm^{-2}$, the current per ubiquinone becomes independent of the ubiquinone coverage. This is evidence for electron transfer at isolated ''molecular gates''. Like the preceding system, the microarray electrode exhibits greater response for hydrophobic redox couples than for hydrophilic ones.

6.2 Selective Permeation

For SAMs that are less than perfectly blocking, selectivity arises from the inherent affinity of hydrophobic molecules for the alkane domain. This property has been applied in the development of a liquid chromatography (LC) electrochemical detector which is selective for neutral organic analytes over ionic compounds.[175] Even more dramatic are electrostatic effects. The current is greatly enhanced at an ionic SAM for redox species of the opposite charge and strongly suppressed for a redox species of the same charge.[176–178] For example, cationic dopamine can be detected in the presence of 100-fold higher concentration of anionic ascorbic acid at a gold electrode coated with w-mercaptoalkanoic acid.[179] Interestingly, a highly charged and hydrophilic SAM containing w-mercaptodecanesulfonic acid strongly blocks organic redox couples, regardless of their charge.[180]

By extension, a bilayer of molecules on or near the electrode surface should provide even greater selectivity for analytes if ion transport agents are incorporated into the bilayer.[133,181] A phospholipid bilayer can be formed near but not on an electrode coated with a SAM whose constituent molecules possess a long hydrophilic body and a hydrophobic phytanyl tail.[135] The hydrophobic tails serve as nucleation sites for the spontaneous assembly of a bilayer. When valinomycin is incorporated into the bilayer, the conductivity of the assembly exhibits a marked increase in the presence of K^+, and a smaller response in the presence of other cations.

A hybrid bilayer containing an alkanethiol and a phospholipid is strongly blocking towards ferricyanide until melittin is added to the solution.[130] Melittin, a cationic oligopeptide isolated from bee venom, is believed to aggregate in bilayers and to create pores that are much larger than most ions. Thus, selectivity should be minimal, except possibly for electrostatic effects due to the positive charge of melittin.

6.3 Preconcentration and Selective Binding

Preconcentration implies that the analyte partitions preferentially into the SAM or to its external surface, so that the analytical signal is enhanced. Preconcentration without designed binding sites depends on hydrophobic/hydrophilic or electrostatic forces. Thus, SAMs with a terminal COOH group preconcentrate Cd^{2+}, while SAMs with a terminal pyridinium bind chromate, allowing detection of these ions down to $10^{-10}\,M$ with cathodic stripping voltammetry.[182,183] Cationic surfactants are quantified by the current obtained for the reduction of ferricyanide at an octadecanethiol-coated electrode; adsorption of the cationic surfactant on the SAM enhances the concentration of ferricyanide and hence the current.[184]

Designed binding sites in the SAM can greatly enhance the selectivity of the electrode (Table 2). Cyclodextrins, either with a pendant thiol or attached to an existing SAM, bind electroactive analytes like ferrocene[185] or quinone.[186] Nonelectroactive *trans*-azo dyes, which also bind to cyclodextrins, are detected either by competitive inhibition[187] or by attachment of an electroactive viologen to the dye.[188] The formal potential of a cyclic bis(bipyridinium) molecule in a SAM shifts in the negative direction in the presence of indole or catechol.[189] A bis(acetylacetone)sulfide forms SAMs that bind Cu^{2+}, Pb^{2+} and Zn^{2+} selectively in the presence of Fe^{2+} or Fe^{3+}.[190–192] The first three ions readily form four-coordinate complexes, while iron cations prefers a six-coordinate geometry. The bound Cu^{2+} or Pb^{2+} act as nucleation sites for the electrodeposition of bulk copper or lead. Consequently, direct measurements are possible for Cu down to $10^{-7}\,M$ and Pb down to $10^{-5}\,M$, and an indirect measurement of Zn can be made down to $10^{-9}\,M$. A mixed SAM containing octadecanethiol and a flat oriented steroid partially blocks ferricyanide

Table 2 SAMs with designed binding sites and the targeted analytes

Binding site	Analyte(s)	Refs.
Cyclodextrin	Ferrocenes	185, 187
Cyclodextrin	Quinone, azo dyes	186
Cyclodextrin	Azo dyes	188
Cyclic bis(bipyridinium)	Indole, catechol	189
bis(Acetylacetone)	4-Coordinate metal ions	190, 191, 192
Steroids	Cyclic polyols	193

reduction.[193] The reduction current is greatly attenuated when cyclic polyols with equatorial OH groups are present in the solution. The polyols are thought to form ice-like structures in the cavities defined by the steroid molecules.

Incorporation of enzymes into SAM structures affords highly selective electrodes. Direct electron transfer between the enzyme and the electrode is relatively rare; most biosensors use a freely diffusing or attached redox couple to catalyze electron transfer to the enzyme (see section 6.4). Examples of the former approach include electrodes selective for the reduction of fumarate,[194] the oxidation of hydrogen peroxide,[167] and the oxidation of lactate.[195]

A particularly elegant approach to a biosensor is based on a suspended bilayer near a gold surface (Figure 8).[134] The bilayer contains gramicidin fragments in the inner layer which are tethered to the gold surface, and hence are fixed. Gramicidin fragments in the outer layer contain a biotin tag which is connected via streptavidin to a biotin-tagged antibody fragment. In the absence of the analyte, the outer-layer gramicidin molecules diffuse laterally and intermittently form ionically conducting dimers with the inner layer gramicidin; the conductivity of the bilayer, measured by AC impedance, is relatively high. In the presence of the analyte, the outer-layer gramicidins become locked to fixed antibody fragments and thus can no longer form conducting dimers. The signal is the rate of decrease in the conductance of the bilayer. This sensor is potentially competitive with enzyme-linked immunosorbent assay in terms of flexibility, selectivity and sensitivity, and does not require the washings or reagent additions of the latter method.

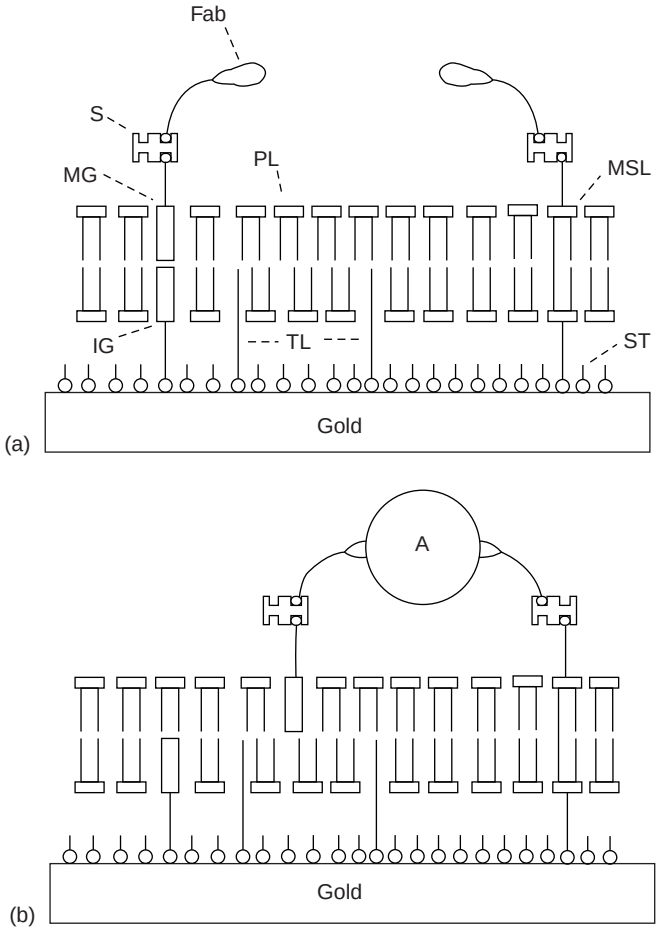

Figure 8 Biosensor based on a suspended lipid bilayer.[134] A = analyte; Fab = biotinylated antibody fragment; IG = immobilized gramacidin; MG = mobile gramacidin with pendant biotin; MSL = membrane-spanning lipid with pendant biotin; PL = phospholipid; S = streptavidin; ST = spacer thiol; TL = tethered lipid. The tethered lipid anchors the phospholipid bilayer to the gold electrode; a short spacer thiol (e.g. mercaptoacetic acid) helps to define the ionic reservoir under the lipid bilayer. The immobilized gramacidins are attached to the electrode via a thiol. Biotinylated antibody fragments are attached via streptavidin to both membrane-spanning lipids with fixed positions and gramacidins which diffuse freely across the external surface. When mobile and immobilized gramacidins coincide, ions can pass across the bilayer and conductivity is high (a). In the presence of the analyte, the Fab's are cross-linked, and the attached gramacidins are no longer able to diffuse freely. Ion channels cannot form, and the conductivity is low (b).

6.4 Electrocatalysis

In electrocatalysis (also known as mediated electron transfer), a redox couple, either attached or freely diffusing, transfers electrons between the electrode and target redox couple. The target redox couple is thermodynamically capable of being oxidized or reduced at the electrode, but the electron-transfer kinetics are extremely slow. Kinetically slow systems include multielectron redox centers and enzymes whose redox centers are inaccessible to the electrode. Because electron transfer between the electrocatalyst and the target molecule is usually thermodynamically downhill, a characteristic feature of voltammograms is catalytic current in one direction only (anodic or cathodic).

SAMs serve two roles in electrocatalysis. First, they provide a flexible means of anchoring the electrocatalyst and/or the enzyme. Second, SAMs block access of the target redox couple to the electrode, a useful feature for fundamental studies. For example, SAMs with pendant ferrocenes block reduction of Fe^{3+} at the $E^{0'}$ of the $Fe^{2+/3+}$ redox couple, but catalyze the reduction of Fe^{3+} at potentials where the attached ferriceniums are reduced.[196] It is possible to measure the rate constant for electron exchange between the catalyst and the target.[197] A SAM containing a remarkable quinone that exhibits photoisomerization between an electroactive and a nonelectroactive form is the basis of a light-switchable electrocatalytic electrode.[198]

Numerous examples exist of electrocatalysis of multielectron redox systems. The two-electron reduction of O_2 to H_2O_2 is catalyzed by attached cobalt porphyrins,[199,200] while the four-electron reduction to water is catalyzed by an attached tetrone (diquinone), albeit at potentials well negative of the thermodynamic potential.[201] NADH oxidation is catalyzed by attached catechols,[202,203] NO reduction is catalyzed by $Cr(terpyridine)_2^{3+}$,[204] H_2O_2 reduction is catalyzed by iron porphyrin or heme protein,[166,167] and both ascorbate and hydrazine oxidation are catalyzed by immobilized ferricyanide (possibly as a Prussian blue-like complex).[205]

Likewise, enzymes have been immobilized on SAMs and their catalytic reaction driven by electrocatalysis. If the electrocatalyst is freely diffusing in the electrolyte, then the SAM must be made permeable to the electrocatalyst. A mixed SAM containing octadecanethiol and dibenzyldisulfide is the base for a hybrid bilayer containing pyruvate oxidase; the benzylthiolate components enhance access of the electrocatalyst ferrocenemethanol to the electrode.[206] Alternately, the electrocatalyst and the enzyme are both attached to the SAM. Glucose oxidase can be attached to a mixed SAM of 2-aminoethanethiol and a long chain ferrocenethiol.[207] More complex structures containing multiple layers of enzymes and electron catalysts have been constructed with the purpose of improving the sensitivity of the sensor or making the sensor light-switchable; in each case, the base coat is a SAM of cystamine.[120,121,208–210]

6.5 Long-range Electron Transfer

The close-packed structure of the alkanethiol SAM provides a useful spacer for anchoring redox centers at fixed and controllable distances from the electrode surface (Figure 6). Mixed SAMs allow adequate spacing between redox centers without loss of packing density. The redox centers can also be freely diffusing if the SAM is completely free of pinholes and other defects (section 5.3), but that condition is difficult to prove.[85,211,212] The alkane chain is an insulator, and as a consequence, electron transfer must proceed by a tunneling mechanism. The rate of electron transfer is diminished to the point that the rate constants can easily be measured over a wide range of overpotentials ($\eta = E - E^{0'}$). Both of these facts have been widely exploited by researchers interested in long-range electron transfer and in fundamental theories of electron transfer. The prevailing theory is the Marcus DOS (density of states) model.[213,214] This model focuses on the overlap of donor and acceptor energy levels in the electrode and in the redox center. It includes an important parameter in Marcus theory called the reorganization energy (λ), which is the energy needed to change the redox-center structure and its solvation sphere between the equilibrium states for the two oxidation states.

A number of electrochemical methods provide access to the rate of electron transfer: cyclic voltammetry,[215] ACIS,[216] square-wave voltammetry,[217] and chronoamperometry.[213,218] These electrochemical methods are useful for rate constants in the range 10^{-2} to $10^5\ s^{-1}$. Measurements of rate constants outside this range require spectroscopic[219] or temperature-jump methods.[220]

Three types of experimental results are of interest. A plot of $\ln k^0$ (k^0 is the standard rate constant at $\eta = 0$) versus the number of repeating units in the tether (i.e. the number of CH_2 groups in the alkane chain) is usually linear, indicating that tunneling rates decay exponentially with distance (Figure 9).[213,220–224] The slope of the plot yields the tunneling parameter β. For a single monolayer, a plot of $\ln k$ versus η (Tafel plot) is nearly linear, but with some curvature (Figure 10). Fitting the curvature to the Marcus DOS model yields λ, the reorganization energy. The reorganization energy can also be obtained from a plot of $\ln k^0$ versus T^{-1} (Arrhenius plot).

Detailed experimental studies have been performed on SAMs with attached ferrocenes, $pyRu(NH_3)_5$, and cytochrome *c*.[1] For alkanethiols on gold, plots of $\ln k^0$ versus the number of CH_2 groups are linear. Values for the tunneling parameter β are 1.0–1.1 per CH_2. The same β is obtained for blocking SAMs and various redox couples in solution.[212] Assuming a 30° tilt for an all-*trans* alkyl chain, β is 1.0–1.1 $Å^{-1}$. Theoretical calculations are in agreement with this value.[225,226] The tunneling mechanism invokes electronic coupling of the metal and redox-center orbitals via the highest filled and lowest unoccupied molecular orbitals of the connecting alkane chain. From this model, a substantial

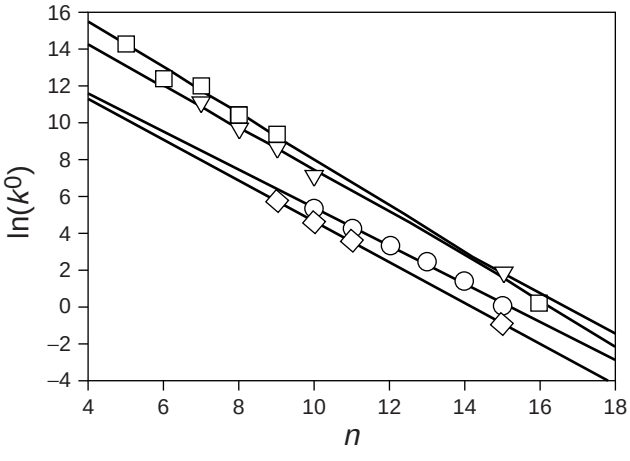

Figure 9 Semilog plots of standard rate constant k^0 versus n in alkanethiol SAM with an attached redox center. ○: $HS(CH_2)_nCONHCH_2pyRu(NH_3)_5^{2+}$ ($\beta = 1.0/CH_2$);[221] ▽: $HS(CH_2)_nNHCOFc$ (Fc = ferrocene) ($\beta = 1.1/CH_2$);[222] : $HS(CH_2)_nOOCFc$ ($\beta = 1.3/CH_2$);[213,220] ◇: cytochrome *c* electrostatically adsorbed on $HS(CH_2)_nCOOH$ ($\beta = 1.1/CH_2$).[223,224]

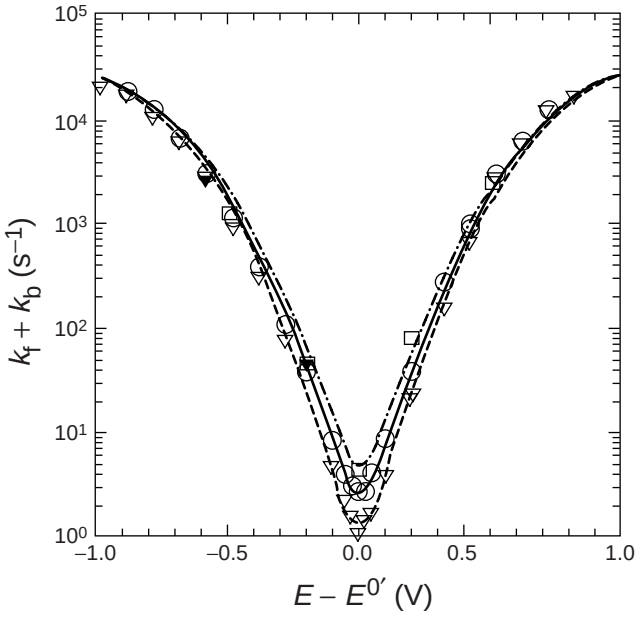

Figure 10 Tafel plots for a mixed SAM containing $HS(CH_2)_{16}OOCFc$ (Fc = ferrocene) and $HS(CH_2)_{15}CH_3$ in 1 M $HClO_4$ at three temperatures: 1 °C (▽), 25 °C (○), and 47 °C ( ). The solid lines are predicted from Marcus DOS theory for a standard rate constant of 1.25 s^{-1} at 25 °C and a reorganization energy of 0.85 eV. (Reprinted with permission from Chidsey, Science, **251**, 919–922,[213] Copyright 1991 by the American Association for the Advancement of Science.)

lowering of β is expected if a fully conjugated spacer is used in the SAM. As predicted, a SAM composed of repeating phenyl–alkynyl spacer units yields a β of 0.6 $Å^{-1}$.[29] Insertion of a heteroatom, a double bond or a triple bond into the repeating CH_2 spacer unit in an alkane chain reduces the electronic coupling.[227] Several experiments suggest that electronic coupling through alkane chains not attached to or in contact with the redox center can contribute significantly to the total electronic coupling.[75,221]

In general, reorganization energies obtained from Tafel plots and Arrhenius plots are in agreement. The λ values are also in surprisingly good agreement with a simple theoretical equation based on the energy of a charged sphere in a dielectric continuum.[162,228] Table 3 lists selected λ values obtained in aqueous electrolytes.

Table 3 Reorganization energies of selected redox centers in water

Redox center	λ (eV)	Refs.
Ferrocene	0.8–1.0	29, 213, 220, 229
Ferri/ferrocyanide	1.1	228
$M(bpy)_3$ (M = Fe, Ru, Os)	0.6	228
$pyRu(NH_3)_5$	0.8–1.2	221, 230, 231
Cytochrome *c*	0.6	232, 233

6.6 Corrosion Control

The most popular metal substrate, gold, is hardly in need of corrosion protection in everyday use. However, protection from chemical etchants is a useful property in the preparation of patterned gold arrays. Two etchants for gold that are effectively blocked by long-chain alkanethiol SAMs are an oxygen-saturated alkaline cyanide bath and a solution of ferricyanide combined with thiocyanate or thiosulfate.[234,235] Various gold microelectrode geometries have been fabricated by this means.[236]

Corrosion protection of copper by SAMs has received more attention, principally by Aramaki et al.[237–241] An alkanethiol SAM affords modest protection against corrosion in aqueous electrolytes. Better corrosion protection is obtained by first depositing a SAM of 11-hydroxyundecanethiol, and then reacting the exposed OH groups with octadecyltrichlorosilane. The 5-nm bilayer inhibits both O_2 reduction and copper oxidation in solution and provides excellent protection against indoor atmospheric corrosion for nearly a year.[231] An XPS study of copper corrosion shows that both the copper and the sulfur are oxidized and that the rate of corrosion in air decreases dramatically with a small increase in SAM thickness.[242] Pit corrosion is clearly evident when a SAM-coated Cu(100) surface is exposed to dilute HCl solution, suggesting that chemical attack occurs at defects in the SAM.[65]

6.7 Adhesion Control

Several organic compounds (aniline, pyrrole, thiophene) undergo polymerization reactions when oxidized to form conducting polymers. When the monomers are oxidized at an electrode, the polymers precipitate onto the electrode to form a coating. The conducting polymer coatings have potential applications in the areas of electrochromism, charge storage and organic semiconductors. It is desirable to create smooth and adherent coatings on the electrodes.

SAMs containing the monomer greatly improve the adhesion, smoothness and density of the conducting polymer coatings. The SAM appears to encourage the formation of many nucleation sites for growth of the polymer. Thus, a poly(aniline) film deposited on a gold electrode coated with a *p*-aminothiophenol SAM is denser and more uniform and exhibits a more uniform discharge than does a poly(aniline film) deposited on a bare gold electrode.[243,244] In a phase-separated mixed SAM of *p*-aminothiophenol and octadecanethiol, poly(aniline) is selectively deposited on the aminothiophenol domains.[172,245] A mechanism for initiation and growth of poly(aniline) on the SAM has been

proposed.[246] SAMs of w-(N-pyrrole)alkanethiol greatly improve the adhesion[247,248] and smoothness of deposited poly(pyrrole).[249] A linear mass change versus charge obtained on a QCM indicates that the poly(pyrrole) is depositing layer by layer, rather than by growth of clumps.[250]

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
ACIS	Alternating Current Impedance Spectroscopy
AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
CV	Cyclic Voltammogram
DOS	Density of States
IR	Infrared
LB	Langmuir–Blodgett
LC	Liquid Chromatography
LDMS	Laser Desorption Mass Spectroscopy
NAD	Nicotinamide Adenine Dinucleotides
QCM	Quartz Crystal Microbalances
SAM	Self-assembled Monolayer
SECM	Scanning Electrochemical Microscopy
SEM	Scanning Electron Microscopy
SIMS	Surface Ionization Mass Spectroscopy
SPR	Surface Plasmon Resonance
STM	Scanning Tunneling Microscopy
TPD	Temperature Programmed Desorption
UPD	Underpotential Deposition
UV	Ultraviolet
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Selective Electrode Coatings for Electroanalysis • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques • Ultrafast Electrochemical Techniques

REFERENCES

1. H.O. Finklea, 'Electrochemistry of Organized Monolayers of Thiols and Related Molecules on Electrodes', in *Electroanalytical Chemistry*, eds. A.J. Bard, I. Rubinstein, Marcel Dekker, New York, 109–335, Vol. 19, 1996.
2. A. Ulman, *An Introduction to Ultrathin Organic Films: from Langmuir–Blodgett to Self-assembly*, Academic Press, San Diego, 1991.
3. D.L. Allara, R.G. Nuzzo, 'Modifications of Properties of Metals', U.S. Patent Application 389/775 (June 18), 1982.
4. R.G. Nuzzo, D.L. Allara, 'Adsorption of Bifunctional Organic Disulfides on Gold Surfaces', *J. Am. Chem. Soc.*, **105**, 4481–4483 (1983).
5. R.G. Nuzzo, F.A. Fusco, D.L. Allara, 'Spontaneously Organized Molecular Assemblies. 3. Preparation and Properties of Solution Adsorbed Monolayers of Organic Disulfides on Gold Surfaces', *J. Am. Chem. Soc.*, **109**, 2358–2368 (1987).
6. M.D. Porter, T.B. Bright, D. Allara, C.E.D. Chidsey, 'Spontaneously Organized Molecular Assemblies. 4. Structural Characterization of *n*-Alkyl Thiol Monolayers on Gold by Optical Ellipsometry, Infrared Spectroscopy, and Electrochemistry', *J. Am. Chem. Soc.*, **109**, 3559–3568 (1987).
7. H.O. Finklea, S. Avery, M. Lynch, T. Furtsch, 'Blocking Oriented Monolayers of Alkyl Mercaptans on Gold Electrodes', *Langmuir*, **3**, 409–413 (1987).
8. E. Sabatani, I. Rubinstein, R. Maoz, J. Sagiv, 'Organized Self-assembling Monolayers on Electrodes. Part I. Octadecyl Derivatives on Gold', *J. Electroanal. Chem.*, **219**, 365–371 (1987).
9. E.B. Troughton, C.D. Bain, G.M. Whitesides, R.G. Nuzzo, D.L. Allara, M.D. Porter, 'Monolayer Films Prepared by the Spontaneous Self-assembly of Symmetrical and Unsymmetrical Dialkyl Sulfides from Solution onto Gold Substrates: Structure, Properties, and Reactivity of Constituent Functional Groups', *Langmuir*, **4**, 365–385 (1988).
10. U.B. Steiner, W.R. Caseri, U.W. Suter, 'Adsorption of Alkanenitriles and Alkanedinitriles on Gold and Copper', *Langmuir*, **8**, 2771–2777 (1992).
11. J.J. Hickman, C. Zou, D. Ofer, P.D. Harvey, M.S. Wrighton, P.E. Laibinis, C.D. Bain, G.M. Whitesides, 'Combining Spontaneous Molecular Assembly With Microfabrication to Pattern Surfaces: Selective Binding of Isonitriles to Platinum Microwires and Characterization by Electrochemistry and Surface Spectroscopy', *J. Am. Chem. Soc.*, **111**, 7271–7272 (1989).
12. I. Tabushi, K. Kurihara, K. Naka, K. Yamamura, H. Hatakeyama, 'Supramolecular Sensor Based on SnO_2 Electrode Modified With Octadecasilyl Monolayer Having Molecular Binding Sites', *Tetrahedron Lett.*, **28**, 4299–4302 (1987).
13. Z.H. Jin, D.V. Vezenov, Y.W. Lee, J.E. Zull, C.N. Sukenik, R.F. Savinell, 'Alternating Current Impedance Characterization of the Structure of Alkylsiloxane Self-assembled Monolayers on Silicon', *Langmuir*, **10**, 2662–2671 (1994).
14. H.O. Finklea, L.R. Robinson, A. Blackburn, B. Richter, D. Allara, T. Bright, 'Formation of an Organized Monolayer by Solution Adsorption of Octadecyltrichlorosilane on Gold: Electrochemical Properties and Structural Characterization', *Langmuir*, **2**, 239–244 (1986).

15. D.L. Allara, A.N. Parikh, F. Rondelez, 'Evidence for a Unique Chain Organization in Long Chain Silane Monolayers Deposited on Two Widely Different Substrates', *Langmuir*, **11**, 2357–2360 (1995).
16. M.R. Linford, C.E.D. Chidsey, 'Alkyl Monolayers Covalently Bonded to Silicon Surfaces', *J. Am. Chem. Soc.*, **115**, 12631–12632 (1993).
17. M.R. Linford, P. Fenter, P.M. Eisenberger, C.E.D. Chidsey, 'Alkyl Monolayers on Silicon Prepared from 1-Alkenes and Hydrogen-terminated Silicon', *J. Am. Chem. Soc.*, **117**, 3145–3155 (1995).
18. R.G. Nuzzo, L.H. Dubois, D.L. Allara, 'Fundamental Studies of Microscopic Wetting on Organic Surfaces. 1. Formation and Structural Characterization of a Self-consistent Series of Polyfunctional Organic Monolayers', *J. Am. Chem. Soc.*, **112**, 558–569 (1990).
19. C.D. Bain, E.B. Troughton, Y.-T. Tao, J. Evall, G.M. Whitesides, R.G. Nuzzo, 'Formation of Monolayer Films by the Spontaneous Assembly of Organic Thiols from Solution onto Gold', *J. Am. Chem. Soc.*, **111**, 321–335 (1989).
20. C.D. Bain, G.M. Whitesides, 'Modeling Organic Surfaces With Self-assembled Monolayers', *Angew. Chem. Int. Ed. Engl.*, **28**, 506–512 (1989).
21. C.D. Bain, G.M. Whitesides, 'Depth Sensitivity of Wetting: Monolayers of w-Mercapto Ethers on Gold', *J. Am. Chem. Soc.*, **110**, 5897–5898 (1988).
22. C. Pale-Grosdemange, E.S. Simon, K.L. Prime, G.M. Whitesides, 'Formation of Self-assembled Monolayers by Chemisorption of Derivatives of Oligo(ethylene glycol) of Structure $HS(CH_2)_{11}(OCH_2CH_2)_mOH$ on Gold', *J. Am. Chem. Soc.*, **113**, 12–20 (1991).
23. P.E. Laibinis, C.D. Bain, R.G. Nuzzo, G.M. Whitesides, 'Structure and Wetting Properties of w-Alkoxy-n-alkanethiolate Monolayers on Gold and Silver', *J. Phys. Chem.*, **99**, 7663–7676 (1995).
24. S.D. Evans, E. Urankar, A. Ulman, N. Ferris, 'Self-assembled Monolayers of Alkanethiols Containing a Polar Aromatic Group: Effects of the Dipole Position on Molecular Packing, Orientation, and Surface Wetting Properties', *J. Am. Chem. Soc.*, **113**, 4121–4131 (1991).
25. S.-C. Chang, I. Chao, Y.-T. Tao, 'Structures of Self-assembled Monolayers of Aromatic-derivatized Thiols on Evaporated Gold and Silver Surfaces: Implication on Packing Mechanism', *J. Am. Chem. Soc.*, **116**, 6792–6805 (1994).
26. Y.-T. Tao, C.-C. Wu, J.-Y. Eu, W.-L. Lin, K.-C. Wu, C. Chen, 'Structure Evolution of Aromatic-derivatized Thiol Monolayers on Evaporated Gold', *Langmuir*, **13**, 4018–4023 (1997).
27. E. Sabatani, J. Cohen-Boulakia, M. Bruening, I. Rubinstein, 'Thioaromatic Monolayers on Gold: A New Family of Self-assembling Monolayers', *Langmuir*, **9**, 2974–2981 (1993).
28. T. Kim, R.M. Crooks, M. Tsen, L. Sun, 'Polymeric Self-assembled Monolayers. 2. Synthesis and Characterization of Self-assembled Polydiacetylene Mono- and Multilayers', *J. Am. Chem. Soc.*, **117**, 3963–3967 (1995).
29. S.B. Sachs, S.P. Dudek, R.P. Hsung, L.A. Sita, J.F. Smalley, M.D. Newton, S.W. Feldberg, C.E.D. Chidsey, 'Rates of Interfacial Electron Transfer Through π-Conjugated Spacers', *J. Am. Chem. Soc.*, **119**, 10563–10564 (1997).
30. R.W. Zehner, L.R. Sita, 'Electrochemical Evaluation and Enhancement via Heterogeneous Exchange of the Passivating Properties and Stability of Self-assembled Monolayers Derived from the Rigid Rod Arenethiols', *Langmuir*, **13**, 2973–2979 (1997).
31. J.M. Tour, L.J. Jones II, D.L. Pearson, J.J.S. Lamba, T.P. Burgin, G.M. Whitesides, D.L. Allara, A.N. Parikh, S.V. Atre, 'Self-assembled Monolayers and Multilayers of Conjugated Thiols, α, w-Dithiols, and Thioacetyl-containing Adsorbates. Understanding Attachments Between Potential Molecular Wires and Gold Surfaces', *J. Am. Chem. Soc.*, **117**, 9529–9534 (1995).
32. B. Liedberg, Z. Yang, I. Engquist, M. Wirde, U. Gelius, G. Gotz, P. Bauerle, R.-M. Rummel, Ch. Ziegler, W. Gopel, 'Self-assembly of Alpha-functionalized Terthiophenes on Gold', *J. Phys. Chem. B*, **101**, 5951–5962 (1997).
33. Y.S. Obeng, M.E. Laing, A.C. Freidli, H.C. Yang, D.N. Wang, E.W. Thulstrup, A.J. Bard, J. Michl, 'Self-assembled Monolayers of Parent and Derivatized [n]Staffane-3,3(n − 1)-dithiols on Polycrystalline Gold Electrodes', *J. Am. Chem. Soc.*, **114**, 9943–9952 (1992).
34. S.D. Evans, K.E. Goppert-Bearducci, E. Urankar, L.J. Gerenser, A. Ulman, 'Monolayers Having Large In-plane Dipole Moments: Characterization of Sulfone-containing Self-assembled Monolayers of Alkanethiols on Gold by Fourier Transform Infrared Spectroscopy, X-ray Photoelectron Spectroscopy, and Wetting', *Langmuir*, **7**, 2700–2709 (1991).
35. S.-W. Tam-Chang, H.A. Biebuyck, G.M. Whitesides, 'Self-assembled Monolayers on Gold Generated from Alkanethiols With the Structure $RNHCOCH_2SH$', *Langmuir*, **11**, 4371–4382 (1995).
36. R.S. Clegg, J.E. Hutchison, 'Hydrogen-bonding, Self-assembled Monolayers: Ordered Molecular Films for Study of Through-peptide Electron Transfer', *Langmuir*, **12**, 5239–5243 (1996).
37. G.M. Whitesides, P.E. Laibinis, 'Wet Chemical Approaches to the Characterization of Organic Surfaces: Self-assembled Monolayers, Wetting, and the Physical–Organic Chemistry of the Solid–Liquid Interface', *Langmuir*, **6**, 87–96 (1990).
38. M.T. McDermott, J.-B.D. Green, M.D. Porter, 'Scanning Force Microscopic Exploration of the Lubrication Capabilities of n-Alkanethiolate Monolayers Chemisorbed at Gold: Structural Basis of Microscopic Friction and Wear', *Langmuir*, **13**, 2504–2510 (1997).

39. R.C. Thomas, J.E. Houston, R.M. Crooks, T. Kim, T.A. Michalske, 'Probing Adhesion Forces at the Molecular Scale', *J. Am. Chem. Soc.*, **117**, 3830–3834 (1995).
40. H.I. Kim, T. Koini, T.R. Lee, S.S. Perry, 'Systematic Studies of the Frictional Properties of Fluorinated Monolayers With Atomic Force Microscopy: Comparison of CF_3- and CH_3-terminated Films', *Langmuir*, **13**, 7192–7196 (1997).
41. G.P. Lopez, H.A. Biebuyck, G.M. Whitesides, 'Scanning Electron Microscopy Can Form Images of Patterns in Self-assembled Monolayers', *Langmuir*, **9**, 1513–1516 (1993).
42. J.P. Folkers, P.E. Laibinis, G.M. Whitesides, J. Deutch, 'Phase Behavior of Two-component Self-assembled Monolayers of Alkanethiolates on Gold', *J. Phys. Chem.*, **98**, 563–571 (1994).
43. V. Chechik, H. Schonherr, G.J. Vancso, C.J.M. Stirling, 'Self-assembled Monolayers of Branched Thiols and Disulfides on Gold: Surface Coverage, Order and Chain Orientation', *Langmuir*, **14**, 3003–3010 (1998).
44. A. Ulman, *Characterization of Organic Thin Films: Materials Characterization Series–Surfaces, Interfaces, Thin Films*, Butterworth-Heinemann, Stoneham, MA, 1995.
45. L.H. Dubois, R.G. Nuzzo, 'Synthesis, Structure, and Properties of Model Organic Surfaces', *Annu. Rev. Phys. Chem.*, **43**, 437–463 (1992).
46. A.R. Bishop, R.G. Nuzzo, 'Self-assembled Monolayers: Recent Developments and Applications', *Curr. Opin. Colloid Interface Sci.*, **1**, 127–136 (1996).
47. C.-J. Zhong, M.D. Porter, 'Designing Interfaces at the Molecular Level', *Anal. Chem.*, **67**, 709A–715A (1995).
48. Th. Wink, S.J. van Suilen, A. Bult, W.P. van Bennekom, 'Self-assembled Monolayers for Biosensors', *Analyst*, **122**, 43R–50R (1997).
49. L.-H. Guo, J.S. Facci, G. McLendon, R. Mosher, 'Effect of Gold Topography and Surface Pretreatment on the Self-assembly of Alkanethiol Monolayers', *Langmuir*, **10**, 4588–4593 (1994).
50. X.-M. Zhao, J.L. Wilbur, G.M. Whitesides, 'Using Two-stage Chemical Amplification To Determine the Density of Defects in Self-assembled Monolayers of Alkanethiolates on Gold', *Langmuir*, **12**, 3257–3264 (1996).
51. Y. Golan, L. Margulis, I. Rubinstein, 'Vacuum-deposited Gold Films. I. Factors Affecting Film Morphology', *Surf. Sci.*, **264**, 312–326 (1992).
52. M.M. Walczak, C.A. Alves, B.D. Lamp, M.D. Porter, 'Electrochemical and X-ray Photoelectron Spectroscopic Evidence for Differences in the Binding Sites of Alkanethiolate Monolayers Chemisorbed on Gold', *J. Electroanal. Chem.*, **396**, 103–114 (1995).
53. W.B. Caldwell, K. Chen, B.R. Herr, C.A. Mirkin, J.C. Hulteen, R.P. Van Duyne, 'Self-assembled Monolayers of Ferrocenylazobenzenes on Au(111)/Mica Films: Surface-enhanced Raman Scattering Response vs Surface Morphology', *Langmuir*, **10**, 4109–4115 (1994).
54. V.K. Gupta, N.L. Abbott, 'Uniform Anchoring of Nematic Liquid Crystals on Self-assembled Monolayers Formed from Alkanethiols on Obliquely Deposited Films of Gold', *Langmuir*, **12**, 2587–2593 (1996).
55. D. Stamou, D. Gourdon, M. Liley, N.A. Burnham, A. Kulik, H. Vogel, C. Duschl, 'Uniformly Flat Gold Surfaces: Imaging the Domain Structure of Organic Monolayers Using Scanning Force Microscopy', *Langmuir*, **13**, 2425–2428 (1997).
56. P. Wagner, M. Hegner, H.-J. Gůntherodt, G. Semenza, 'Formation and in situ Modification of Monolayers Chemisorbed on Ultraflat Template-stripped Gold Surfaces', *Langmuir*, **11**, 3867–3875 (1995).
57. P.A. DiMilla, J.P. Folkers, H.A. Biebuyck, R. Harter, G.P. Lopez, G.M. Whitesides, 'Wetting and Protein Adsorption of Self-assembled Monolayers of Alkanethiolates Supported on Transparent Films of Gold', *J. Am. Chem. Soc.*, **116**, 2225–2226 (1994).
58. K.A. Peterlinz, R. Georgiadis, 'In situ Kinetics of Self-assembly by Surface Plasmon Resonance Spectroscopy', *Langmuir*, **12**, 4731–4740 (1996).
59. S.E. Creager, L.A. Hockett, G.K. Rowe, 'Consequences of Microscopic Surface Roughness for Molecular Self-assembly', *Langmuir*, **8**, 854–861 (1992).
60. H. Ron, I. Rubinstein, 'Alkanethiol Monolayers on Preoxidized Gold. Encapsulation of Gold Oxide Under an Organic Monolayer', *Langmuir*, **10**, 4566–4573 (1994).
61. H. Ron, S. Matlis, I. Rubinstein, 'Self-assembled Monolayers on Oxidized Metals. 2. Gold Surface Oxidative Pretreatment, Monolayer Properties, and Depression Formation', *Langmuir*, **14**, 1116–1121 (1998).
62. P.E. Laibinis, C.D. Bain, G.M. Whitesides, 'Attenuation of Photoelectrons in Monolayers of *n*-Alkanethiols Adsorbed on Copper, Silver, and Gold', *J. Phys. Chem.*, **95**, 7017–7021 (1991).
63. P.E. Laibinis, G.M. Whitesides, D.L. Allara, Y.-T. Tao, A.N. Parikh, R.G. Nuzzo, 'Comparison of the Structure and Wetting Properties of Self-assembled Monolayers of *n*-Alkanethiols on the Coinage Metal Surfaces, Cu, Ag, Au', *J. Am. Chem. Soc.*, **113**, 7152–7167 (1991).
64. P.E. Laibinis, G.M. Whitesides, 'ω-Terminated Alkanethiolate Monolayers on Surfaces of Copper, Silver, and Gold Have Similar Wettabilities', *J. Am. Chem. Soc.*, **114**, 1990–1995 (1992).
65. J. Scherer, M.R. Vogt, O.M. Magnussen, R.J. Behm, 'Corrosion of Alkanethiol-covered Cu(100) Surfaces in Hydrochloric Acid Solution by in situ Scanning Tunneling Microscopy', *Langmuir*, **13**, 7045–7051 (1997).
66. M.H. Schoenfisch, J.E. Pemberton, 'Air Stability of Alkanethiol Self-assembled Monolayers on Silver and Gold Surfaces', *J. Am. Chem. Soc.*, **120**, 4502–4513 (1998).
67. G.K. Jennings, P.E. Laibinis, 'Underpotential Deposited Metal Layers of Silver Provide Enhanced Stability

to Self-assembled Alkanethiol Monolayers on Gold', *Langmuir*, **12**, 6173–6175 (1996).
68. G.K. Jennings, P.E. Laibinis, 'Self-assembled *n*-Alkanethiolate Monolayers on Underpotentially Deposited Adlayers of Silver and Copper on Gold', *J. Am. Chem. Soc.*, **119**, 5208–5214 (1997).
69. J.D. Burgess, F.M. Hawkridge, 'Octadecyl Mercaptan Sub-monolayers on Silver Electrodeposited on Gold Quartz Crystal Microbalance Electrodes', *Langmuir*, **13**, 3781–3786 (1997).
70. F.P. Zamborini, J.K. Campbell, R.M. Crooks, 'Spectroscopic, Voltammetric, and Electrochemical Scanning Tunneling Microscopic Study of Underpotentially Deposited Cu Corrosion and Passivation With Self-assembled Organomercaptan Monolayers', *Langmuir*, **14**, 640–647 (1998).
71. M. Nishizawa, T. Sunagawa, H. Yoneyama, 'Underpotential Deposition of Copper on Gold Electrodes Through Self-assembled Monolayers of Propanethiol', *Langmuir*, **13**, 5215–5217 (1997).
72. A. Demoz, D.J. Harrison, 'Characterization of Extremely Low Defect Density Hexadecanethiol Monolayers on Hg Surfaces', *Langmuir*, **9**, 1046–1050 (1993).
73. N. Muskal, I. Turyan, D. Mandler, 'Thiol Self-assembled Monolayers on Mercury Surfaces', *J. Electroanal. Chem.*, **409**, 131–136 (1996).
74. O.F. Magnussen, B.M. Ocko, M. Deutsch, M.J. Regan, P.S. Pershan, D. Abernathy, G. Grubel, J.-F. Legrand, 'Self-assembly of Organic Films on a Liquid Metal', *Nature*, **384**, 250–252 (1996).
75. K. Slowinski, R.V. Chamberlain, C.J. Miller, M. Majda, 'Through-bond and Chain-to-chain Coupling. Two Pathways in Electron Tunneling Through Liquid Alkanethiol Monolayers on Mercury Electrodes', *J. Am. Chem. Soc.*, **119**, 11910–11919 (1997).
76. C. Bruckner-Lea, J. Janata, J. Conroy, A. Pungor, K. Caldwell, 'Scanning Tunneling Microscopy on a Mercury Sessile Drop', *Langmuir*, **9**, 3612–3617 (1993).
77. Z. Mekhalif, J.-J. Pireaux, J. Delhalle, 'Self-assembled Monolayers of *n*-Dodecanethiol on Electrochemically Modified Polycrystalline Nickel Surfaces', *Langmuir*, **13**, 2285–2290 (1997).
78. A.D. Vogt, T. Han, T.P. Beebe, 'Adsorption of 11-Mercaptoundecanoic Acid in Ni(111) and Its Interaction With Probe Molecules', *Langmuir*, **13**, 3397–3403 (1997).
79. T. Kondo, M. Takechi, Y. Sato, K. Uosaki, 'Adsorption Behavior of Functionalized Ferrocenylalkane Thiols and Disulfide on Au and ITO and Electrochemical Properties of Modified Electrodes: Effects of Acyl and Alkyl Groups Attached to the Ferrocene Ring', *J. Electroanal. Chem.*, **381**, 203–209 (1995).
80. Y. Gu, Z. Lin, R.A. Butera, V.S. Smentkowski, D.H. Waldeck, 'Preparation of Self-assembled Monolayers on InP', *Langmuir*, **11**, 1849–1851 (1995).
81. K. Chen, F. Xu, C.A. Mirkin, R.-K. Lo, K.S. Nanjundaswamy, J.-P. Zhou, J.T. McDevitt, 'Do Alkanethiols Adsorb onto the Surfaces of Tl-Ba-Cu-O-based High-temperature Superconductors? The Critical Role of H_2O Content in the Adsorption Process', *Langmuir*, **12**, 2622–2624 (1996).
82. P. Fenter, A. Eberhardt, K.S. Liang, P. Eisenberger, 'Epitaxy and Chainlength Dependent Strain in Self-assembled Monolayers', *J. Chem. Phys.*, **106**, 1600–1608 (1997).
83. N. Camillone III, T.Y.B. Leung, G. Scoles, 'A Low Energy Helium Atom Diffraction Study of Decanethiol Self-assembled on Au(111)', *Surf. Sci.*, **373**, 333–349 (1997).
84. A. Kudelski, P. Krysinski, 'Solvent Trapping During the Self-assembly of Octadecanethiol Monolayer on Roughened Gold Electrodes from Surface-enhanced Raman Scattering Studies', *J. Electroanal. Chem.*, **443**, 5–7 (1998).
85. C. Miller, P. Cuendet, M. Gratzel, 'Adsorbed *w*-Hydroxy Thiol Monolayers on Gold Electrodes: Evidence for Electron Tunneling to Redox Species in Solution', *J. Phys. Chem.*, **95**, 877–886 (1991).
86. J. Liu, A.E. Kaifer, 'Preparation of Self-assembled Monolayers from Micellar Solutions', *Israel J. Chem.*, **37**, 235–239 (1997).
87. J. Yan, S. Dong, 'Formation of Surface Inclusion Complexes Between Cyclodextrins and *n*-Alkanethiols and Their Self-assembled Behaviors on Gold', *Langmuir*, **13**, 3251–3255 (1997).
88. J.A.M. Sondag-Huethorst, C. Schonenberger, L.G.J. Fokkink, 'Formation of Holes in Alkanethiol Monolayers on Gold', *J. Phys. Chem.*, **98**, 6826–6834 (1994).
89. K. Edinger, A. Golzhauser, K. Demota, Ch. Woll, M. Grunze, 'Formation of Self-assembled Monolayers of *n*-Alkanethiols on Gold: A Scanning Tunneling Microscopy Study on the Modification of Substrate Morphology', *Langmuir*, **9**, 4–8 (1993).
90. E. Delamarche, B. Michel, Ch. Gerber, D. Anselmetti, H.-J. Guntherodt, H. Wolf, H. Ringsdorf, 'Real-space Observation of Nanoscale Molecular Domains in Self-assembled Monolayers', *Langmuir*, **10**, 2869–2871 (1994).
91. N. Camillone, C.E.D. Chidsey, P. Eisenberger, P. Fenter, J. Li, K.S. Liang, G.Y. Liu, G. Scoles, 'Structural Defects in Self-assembled Organic Monolayers via Combined Atomic Beam and X-ray Diffraction', *J. Chem. Phys.*, **99**, 744–747 (1993).
92. H.O. Finklea, D.D. Hanshew, 'Preparation and Reversible Behavior of Organized Thiol Monolayers With Attached Pentaminepyridineruthenium Redox Centers', *J. Electroanal. Chem.*, **347**, 327–340 (1993).
93. C.D. Bain, J. Evall, G.M. Whitesides, 'Formation of Monolayers by the Coadsorption of Thiols on Gold: Variation in the Head Group, Tail Group, and Solvent', *J. Am. Chem. Soc.*, **111**, 7155–7164 (1989).

94. C.D. Bain, G.M. Whitesides, 'Formation of Monolayers by the Coadsorption of Thiols on Gold: Variation in the Length of the Alkyl Chain', *J. Am. Chem. Soc.*, **111**, 7164–7175 (1989).
95. D.E. Weisshaar, B.D. Lamp, M.D. Porter, 'Thermodynamically Controlled Electrochemical Formation of Thiolate Monolayers at Gold: Characterization and Comparison to Self-assembled Analogs', *J. Am. Chem. Soc.*, **114**, 5860–5862 (1992).
96. D.W. Hatchett, R.H. Uibel, K.J. Stevenson, J.M. Harris, H.S. White, 'Electrochemical Measurement of the Free Energy of Adsorption of *n*-Alkanethiolates on Ag(111)', *J. Am. Chem. Soc.*, **120**, 1062–1069 (1998).
97. K.J. Stevenson, M. Mitchell, H.S. White, 'Oxidative Adsorption of *n*-Alkanethiolates at Mercury. Dependence of Adsorption Free Energy on Chain length', *J. Phys. Chem. B*, **102**, 1235–1240 (1998).
98. G.E. Poirier, E.D. Pylant, 'The Self-assembly Mechanism of Alkane Thiols on Au(111)', *Science*, **272**, 1145–1148 (1996).
99. O. Chailapakul, L. Sun, C. Xu, R.M. Crooks, 'Interactions Between Organized, Surface-confined Monolayers and Vapor-phase Probe Molecules. 7. Comparison of Self-assembling *n*-Alkanethiol Monolayers Deposited on Gold from Liquid and Vapor Phases', *J. Am. Chem. Soc.*, **115**, 12459–12467 (1993).
100. R. Bilewicz, M. Majda, 'Monomolecular Langmuir–Blodgett Films at Electrodes. Formation of Passivating Monolayers Incorporating Electroactive Reagents', *Langmuir*, **7**, 2794–2802 (1991).
101. P. Krysinski, R.V. Chamberlain, M. Majda, 'Partial Electron Transfer in Octadecanethiol Binding to Gold', *Langmuir*, **10**, 4286–4294 (1994).
102. R. Bilewicz, T. Sawaguchi, R.V. Chamberlain II, M. Majda, 'Monomolecular Langmuir–Blodgett Films at Electrodes. Electrochemistry at Single Molecule ''Gate Sites'' ' *Langmuir*, **11**, 2256–2266 (1995).
103. P. Fenter, A. Eberhardt, P. Eisenberger, 'Self-assembly of *n*-Alkyl Thiols as Disulfides on Au(111)', *Science*, **266**, 1216–1218 (1994).
104. C.-J. Zhong, M.D. Porter, 'Evidence for Carbon–Sulfur Cleavage in Spontaneously Adsorbed Organosulfide-based Monolayers at Gold', *J. Am. Chem. Soc.*, **116**, 11616–11617 (1994).
105. S.J. Stranick, A.N. Parikh, Y.-T. Tao, D.L. Allara, P.S. Weiss, 'Phase Separation of Mixed-composition Self-assembled Monolayers into Nanometer Scale Molecular Domains', *J. Phys. Chem.*, **98**, 7636–7646 (1994).
106. S.J. Stranick, S.V. Atre, A.N. Parikh, M.C. Wood, D.L. Allara, N. Winograd, P.S. Weiss, 'Nanometer-scale Phase Separation in Mixed Composition Self-assembled Monolayers', *Nanotechnology*, **7**, 438–442 (1996).
107. H. Van Ryswyk, E.D. Turtle, R. Watson-Clark, T.A. Tanzer, T.K. Herman, P.Y. Chong, P.J. Waller, A.L. Taurog, C.E. Wagner, 'Reactivity of Ester Linkages and Pentaammineruthenium(III) at the Monolayer Assembly/Solution Interface', *Langmuir*, **12**, 6143–6150 (1996).
108. V. Chechik, C.J.M. Stirling, 'Reactivity in Monolayers versus Bulk Media: Intra- and Intermolecular Aminolysis of Esters', *Langmuir*, **13**, 6354–6356 (1997).
109. P. Neogi, S. Neogi, C.J.M. Stirling, 'Reactivity of Carboxy Esters in Gold–Thiol Monolayers', *J. Chem. Soc., Chem. Commun.*, 1134–1136 (1993).
110. T. Kim, Q. Ye, L. Sun, K.C. Chan, R.M. Crooks, 'Polymeric Self-assembled Monolayers: 5. Synthesis and Characterization of w-Functionalized, Self-assembled Diacetylenic and Polydiacetylenic Monolayers', *Langmuir*, **12**, 6065–6073 (1996).
111. T. Kim, K.C. Chan, R.M. Crooks, 'Polymeric Self-assembled Monolayers. 4. Chemical, Electrochemical, and Thermal Stability of w-Functionalized, Self-assembled Diacetylenic and Polydiacetylenic Monolayers', *J. Am. Chem. Soc.*, **119**, 189–193 (1997).
112. J.S. Peanasky, R.L. McCarley, 'Surface-confined Monomers on Electrode Surfaces. 4. Electrochemical and Spectroscopic Characterization of Undec-10-ene-1-thiol Self-assembled Monolayers on Au', *Langmuir*, **14**, 113–123 (1998).
113. W.R. Thompson, M. Cai, M. Ho, J.E. Pemberton, 'Hydrolysis and Condensation of Self-assembled Monolayers of (3-Mercaptopropyl)trimethoxysilane on Ag and Au Surfaces', *Langmuir*, **13**, 2291–2302 (1997).
114. R.I. Carey, J.P. Folkers, G.M. Whitesides, 'Self-assembled Monolayers Containing w-Mercaptoalkylboronic Acids Adsorbed onto Gold Form a Highly Cross-linked, Thermally Stable Borate Glass Surface', *Langmuir*, **10**, 2228–2234 (1994).
115. R.J. Willicut, R.L. McCarley, 'Electrochemical Polymerization of Pyrrole-containing Self-assembled Alkanethiol Monolayers on Au', *J. Am. Chem. Soc.*, **116**, 10823–10824 (1994).
116. R.J. Willicut, R.L. McCarley, 'Surface-confined Monomers on Electrode Surfaces. Part 3. Electrochemical Reactions and Scanning Probe Microscopy Investigations of w-(*N*-Pyrrolyl)alkanethiol Self-assembled Monolayers on Gold', *Anal. Chim. Acta*, **307**, 269–276 (1995).
117. J. Spinke, M. Liley, F.-J. Schmitt, H.-J. Guider, L. Angermaier, W. Knoll, 'Molecular Recognition at Self-assembled Monolayers: Optimization of Surface Functionalization', *J. Chem. Phys.*, **99**, 7012–7019 (1993).
118. G.B. Sigal, C. Bamdad, A. Barberis, J. Strominger, G.M. Whitesides, 'A Self-assembled Monolayer for the Binding and Study of Histidine-tagged Proteins by Surface Plasmon Resonance', *Anal. Chem.*, **68**, 490–497 (1996).
119. M. Liley, T.A. Keller, C. Duschl, H. Vogel, 'Direct Observation of Self-assembled Monolayers, Ion Complexation, and Protein Conformation at the Gold/Water Interface: An FTIR Spectroscopic Approach', *Langmuir*, **13**, 4190–4192 (1997).

120. E. Katz, V. Heleg-Shabtai, B. Willner, I. Willner, A.F. Buckmann, 'Electrical Contact of Redox Enzymes With Electrodes: Novel Approaches for Amperometric Biosensors', *Bioelectrochem. Bioenerg.*, **42**, 95–104 (1997).
121. A. Bardea, E. Katz, A.F. Buckmann, I. Willner, 'NAD^+-dependent Enzyme Electrodes: Electrical Contact of Cofactor-dependent Enzymes and Electrodes', *J. Am. Chem. Soc.*, **119**, 9114–9119 (1997).
122. K.A. Peterlinz, R.M. Georgiadis, T.M. Herne, M.H. Tarlov, 'Observation of Hybridization and Dehybridization of Thiol-tethered DNA Using Two-color Surface Plasmon Resonance Spectroscopy', *J. Am. Chem. Soc.*, **119**, 3401–3402 (1997).
123. J.J. Storhoff, R. Elghanian, R.C. Mucic, C.A. Mirkin, R.L. Letsinger, 'One-pot Colorimetric Differentiation of Polynucleotides With Single Base Imperfections Using Gold Nanoparticle Probes', *J. Am. Chem. Soc.*, **120**, 1959–1964 (1998).
124. F. Caruso, K. Niikura, D.N. Furlong, Y. Okahata, 'Ultrathin Multilayer Polyelectrolyte Films on Gold: Construction and Thickness Determination', *Langmuir*, **13**, 3422–3426 (1997).
125. F. Caruso, K. Niikura, D.N. Furlong, Y. Okahata, 'Assembly of Alternating Polyelectrolyte and Protein Multilayer Films for Immunosensing', *Langmuir*, **13**, 3427–3433 (1997).
126. T.L. Freeman, S.D. Evans, A. Ulman, 'XPS Studies of Self-assembled Multilayer Films', *Langmuir*, **11**, 4411–4417 (1995).
127. H. Lee, L.J. Kepley, H.-G. Hong, S. Akhter, T.E. Mallouk, 'Adsorption of Ordered Zirconium Phosphonate Multilayer Films on Silicon and Gold Surfaces', *J. Phys. Chem.*, **92**, 2597–2601 (1988).
128. L.-H. Guo, J.S. Facci, G. McLendon, 'Distance Dependence of Electron Transfer Rates in Bilayers of a Ferrocene L–B Monolayer and a Self-assembled Monolayer on Gold', *J. Phys. Chem.*, **99**, 8458–8461 (1995).
129. P. Sanassy, S.D. Evans, 'Mixed Alkanethiol Monolayers on Gold Surfaces: Substrates for Langmuir–Blodgett Film Deposition', *Langmuir*, **9**, 1024–1027 (1993).
130. A.L. Plant, 'Self-assembled Phospholipid/Alkanethiol Biomimetic Bilayers on Gold', *Langmuir*, **9**, 2764–2767 (1993).
131. S. Terrettaz, T. Stora, C. Duschl, H. Vogel, 'Protein Binding to Supported Lipid Membranes: Investigation of the Cholera Toxin–Ganglioside Interaction by Simultaneous Impedance Spectroscopy and Surface Plasmon Resonance', *Langmuir*, **9**, 1361–1369 (1993).
132. L. Ding, J. Li, S. Dong, E. Wang, 'Supported Phospholipid Membranes: Comparison Among Different Deposition Methods for a Phospholipid Monolayer', *J. Electroanal. Chem.*, **416**, 105–112 (1996).
133. N. Bunjes, E.K. Schmidt, A. Jonczyk, F. Rippmann, D. Beyer, H. Ringsdorf, P. Graber, W. Knoll, R. Naumann, 'Thiopeptide-supported Lipid Layers on Solid Substrates', *Langmuir*, **13**, 6188–6194 (1997).
134. B.A. Cornell, V.L.B. Braach-Maksvytis, L.G. King, P.D. J. Osman, B. Ragues, L. Wieczorek, R.J. Pace, 'A Biosensor That Uses Ion-channel Switches', *Nature*, **387**, 580–583 (1997).
135. B. Raguse, V. Braach-Maksvystis, B.A. Cornell, L.G. King, P.D.J. Osman, R.J. Pace, L. Weiczorek, 'Tethered Lipid Bilayer Membranes: Formation and Ionic Reservoir Characterization', *Langmuir*, **14**, 648–659 (1998).
136. L.M. Williams, S.D. Evans, T.M. Flynn, A. Marsh, P.F. Knowles, R.J. Bushby, N. Boden, 'Kinetics of the Unrolling of Small Unilamellar Phospholipid Vesicles onto Self-assembled Monolayers', *Langmuir*, **13**, 751–757 (1997).
137. C.A. Widrig, C. Chung, M.D. Porter, 'The Electrochemical Desorption of *n*-Alkanethiol Monolayers from Polycrystalline Au and Ag Electrodes', *J. Electroanal. Chem.*, **310**, 335–359 (1991).
138. K. Slowinski, R.V. Chamberlain, R. Bilewicz, M. Majda, 'Evidence for Inefficient Chain-to-chain Coupling in Electron Tunneling Through Liquid Alkane Thiol Monolayer Films on Mercury', *J. Am. Chem. Soc.*, **118**, 4709–4710 (1996).
139. C.P. Smith, H.S. White, 'Theory of the Interfacial Potential Distribution and Reversible Voltammetric Response of Electrodes Coated With Electroactive Molecular Films', *Anal. Chem.*, **64**, 2398–2405 (1992).
140. C.P. Smith, H.S. White, 'Voltammetry of Molecular Films Containing Acid/Base Groups', *Langmuir*, **9**, 1–3 (1993).
141. W.R. Fawcett, M. Fedurco, Z. Kovacova, 'Double Layer Effects at Molecular Films Containing Acid/Base Groups', *Langmuir*, **10**, 2403–2408 (1994).
142. M.M. Walczak, D.D. Popenoe, R.S. Deinhammer, B.D. Lamp, C. Chung, M.D. Porter, 'Reductive Desorption of Alkanethiolate Monolayers at Gold: A Measure of Surface Coverage', *Langmuir*, **7**, 2687–2693 (1991).
143. D.-F. Yang, M. Morin, 'Chronoamperometric Study of the Reductive Desorption of Alkanethiol Self-assembled Monolayers', *J. Electroanal. Chem.*, **441**, 173–181 (1998).
144. T.W. Schneider, D.A. Buttry, 'Electrochemical Quartz Microbalance Studies of Adsorption and Desorption of Self-assembled Monolayers of Alkyl Thiols on Gold', *J. Am. Chem. Soc.*, **115**, 12391–12397 (1993).
145. W.R. Everett, I. Fritsch-Faules, 'Factors That Influence the Stability of Self-assembled Organothiols on Gold under Electrochemical Conditions', *Anal. Chim. Acta*, **307**, 253–268 (1995).
146. N. Muskal, I. Turyan, A. Shurky, D. Mandler, 'Chiral Self-assembled Monolayers', *J. Am. Chem. Soc.*, **117**, 1147–1148 (1995).

147. S. Imabayashi, D. Hobara, T. Kakiuchi, W. Knoll, 'Selective Replacement of Adsorbed Alkanethiols in Phase-separated Binary Self-assembled Monolayers by Electrochemical Partial Desorption', *Langmuir*, **13**, 4502–4504 (1997).
148. D. Hobara, M. Ota, S. Imabayashi, K. Niki, T. Kakiuchi, 'Phase Separation of Binary Self-assembled Thiol Monolayers Composed of 1-Hexadecanethiol and 3-Mercaptopropionic Acid on Au(111) Studied by Scanning Tunneling Microscopy and Cyclic Voltammetry', *J. Electroanal. Chem.*, **444**, 113–119 (1998).
149. H.O. Finklea, D.A. Snider, J. Fedyk, E. Sabatani, Y. Gafni, I. Rubinstein, 'Characterization of Octadecanethiol-coated Gold Electrodes as Microarray Electrodes by Cyclic Voltammetry and ac Impedance Spectroscopy', *Langmuir*, **9**, 3660–3667 (1993).
150. L. Sun, R.M. Crooks, 'Imaging of Defects Contained Within *n*-Alkylthiol Monolayers by Combination of Underpotential Deposition and Scanning Tunneling Microscopy: Kinetics of Self-assembly', *J. Electrochem. Soc.*, **138**, L23–L25 (1991).
151. L. Sun, R.M. Crooks, 'Indirect Visualization of Defect Structures Contained Within Self-assembled Organomercaptan Monolayers: Combined Use of Electrochemistry and Scanning Tunneling Microscopy', *Langmuir*, **9**, 1951–1954 (1993).
152. F. Forouzan, A.J. Bard, M.V. Mirkin, 'Voltammetric and Scanning Electrochemical Microscopic Studies of the Adsorption Kinetics and Self-assembly of *n*-Alkanethiol Monolayers on Gold', *Israel J. Chem.*, **37**, 155–163 (1997).
153. O. Chailapakul, R.M. Crooks, 'Interactions Between Organized, Surface-confined Monolayers and Liquid-phase Probe Molecules. 4. Synthesis and Characterization of Nanoporous Molecular Assemblies: Mechanism of Probe Penetration', *Langmuir*, **11**, 1329–1340 (1995).
154. H.O. Finklea, D.A. Snider, J. Fedyk, 'Passivation of Pinholes in Octadecanethiol Monolayers on Gold Electrodes by Electrochemical Polymerization of Phenol', *Langmuir*, **6**, 371–376 (1990).
155. R. Andreu, J.J. Calvente, W.R. Fawcett, M. Molero, 'Role of Ion Pairing in Double-layer Effects at Self-assembled Monolayers Containing a Simple Redox Couple', *J. Phys. Chem. B*, **101**, 2884–2894 (1997).
156. Y. Sato, K. Uosaki, 'Self-assembled Monolayer as a Mediator and Barrier for Electron Transfer Between Electrodes and Redox Species in Solution', in *Redox Mechanisms and Interfacial Properties of Molecules of Biological Importance*, eds. F.A. Schultz, I. Taniguchi, The Electrochemical Society, Pennington, NJ, 299–311, 1993.
157. J. Redepenning, H.M. Tunison, H.O. Finklea, 'Influence of Donnan Potentials on Apparent Formal Potentials Measured for Organized Thiol Monolayers With Attached Pentaminepyridineruthenium Redox Centers', *Langmuir*, **9**, 1404–1407 (1993).
158. K. Shimazu, I. Yagi, Y. Sato, K. Uosaki, 'In Situ and Dynamic Monitoring of the Self-assembling and Redox Processes of a Ferrocenylundecanethiol Monolayer by Electrochemical Quartz Crystal Microbalance', *Langmuir*, **8**, 1385–1387 (1992).
159. H.C. De Long, D.A. Buttry, 'Environmental Effects on Redox Potentials of Viologen Groups Embedded in Electroactive Self-assembled Monolayers', *Langmuir*, **8**, 2491–2496 (1992).
160. G.K. Rowe, S.E. Creager, 'Redox and Ion-pairing Thermodynamics in Self-assembled Monolayers' *Langmuir*, **7**, 2307–2312 (1991).
161. C.E.D. Chidsey, C.R. Bertozzi, T.M. Putvinski, A.M. Mujsce, 'Coadsorption of Ferrocene-terminated and Unsubstituted Alkane Thiols on Gold: Electroactive Self-assembled Monolayers', *J. Am. Chem. Soc.*, **112**, 4301–4306 (1990).
162. M.S. Ravenscroft, H.O. Finklea, 'Electron-transfer Kinetics to Attached Redox Centers on Gold Electrodes in Nonaqueous Electrolytes', *J. Phys. Chem.*, **98**, 3843–3850 (1994).
163. M.J. Tarlov, E.F. Bowden, 'Electron-transfer Reaction of Cytochrome *c* Adsorbed on Carboxylic Acid Terminated Alkanethiol Monolayer Electrodes' *J. Am. Chem. Soc.*, **113**, 1847–1849 (1991).
164. R.A. Clark, E.F. Bowden, 'Voltammetric Peak Broadening for Cytochrome *c*/Alkanethiolate Monolayer Structures: Dispersion of Formal Potentials', *Langmuir*, **13**, 559–565 (1997).
165. T.M. Nahir, E.F. Bowden, 'The Distribution of Standard Rate Constants for Electron Transfer Between Thiol-modified Gold Electrodes and Adsorbed Cytochrome *c*', *J. Electroanal. Chem.*, **410**, 9–13 (1996).
166. T. Lötzbeyer, W. Schuhmann, E. Katz, J. Falter, H.-L. Schmidt, 'Direct Electron Transfer Between the Covalently Immobilized Enzyme Microperoxidase MP-11 and a Cystamine-modified Gold Electrode', *J. Electroanal. Chem.*, **377**, 291–294 (1994).
167. T. Lötzbeyer, W. Schuhmann, H.-L. Schmidt, 'Minizymes. A New Strategy for the Development of Reagentless Amperometric Biosensors Based on Direct Electron-transfer Processes', *Bioelectrochem. Bioenerg.*, **42**, 1–6 (1997).
168. J.J. Hickman, D. Ofer, P.E. Laibinis, G.M. Whitesides, M.S. Wrighton, 'Molecular Self-assembly of Two-terminal Voltammetric Microsensors With Internal References', *Science*, **252**, 688–691 (1991).
169. H.-Z. Yu, Y.-Q. Wang, J.-Z. Cheng, J.-W. Zhao, S.-M. Cai, H. Inokuchi, A. Fujishima, Z.-F. Liu, 'Electrochemical Behavior of Azobenzene Self-assembled Monolayers on Gold', *Langmuir*, **12**, 2843–2848 (1996).
170. W.B. Caldwell, D.J. Campbell, K. Chen, B.R. Herr, C.A. Mirkin, A. Malik, M.K. Durbin, P. Dutta, K.G. Huang, 'A Highly Ordered Self-assembled Monolayer Film of an Azobenzenealkanethiol on Au(111): Electrochemical Properties and Structural Characterization

by Synchrotron In-plane X-ray Diffraction, Atomic Force Microscopy, and Surface-enhanced Raman Spectroscopy', *J. Am. Chem. Soc.*, **117**, 6071–6082 (1995).
171. E. Sabatani, I. Rubinstein, 'Organized Self-assembling Monolayers on Electrodes. 2. Monolayer-based Ultramicroelectrodes for the Study of Very Rapid Electrode Kinetics', *J. Phys. Chem.*, **91**, 6663–6669 (1987).
172. W.A. Hayes, H. Kim, X. Yue, S.S. Perry, C. Shannon, 'Nanometer-scale Patterning of Surfaces Using Self-assembly Chemistry. 2. Preparation, Characterization, and Electrochemical Behavior of Two-component Organothiol Monolayers on Gold Surfaces', *Langmuir*, **13**, 2511–2518 (1997).
173. O. Chailapakul, R.M. Crooks, 'Synthesis and Characterization of Simple Self-assembling, Nanoporous Monolayer Assemblies: A New Strategy for Molecular Recognition', *Langmuir*, **9**, 884–888 (1993).
174. R. Bilewicz, M. Majda, 'Bifunctional Monomolecular Langmuir–Blodgett Films at Electrodes. Electrochemistry at Single Molecule Gate Sites', *J. Am. Chem. Soc.*, **113**, 5464–5466 (1991).
175. J. Wang, H. Wu, L. Angnes, 'On Line Monitoring of Hydrophobic Compounds at Self-assembled Monolayer Modified Amperometric Flow Detectors', *Anal. Chem.*, **65**, 1893–1896 (1993).
176. K. Takehara, H. Takemura, Y. Ide, 'Electrochemical Studies of the Terminally Substituted Alkanethiol Monolayers Formed on a Gold Electrode: Effects of the Terminal Group on the Redox Responses of $Fe(CN)_6^{3-}$, $Ru(NH_3)_6^{3+}$, and Ferrocenedimethanol', *Electrochim. Acta*, **39**, 817–822 (1994).
177. K. Doblhofer, J. Figura, J.-H. Fuhrhop, 'Stability and Electrochemical Behavior of ''Self-assembled'' Adsorbates With Terminal Ionic Groups', *Langmuir*, **8**, 1811–1816 (1992).
178. A. Doron, E. Katz, G. Tao, I. Willner, 'Photochemically, Chemically, and pH-controlled Electrochemistry at Functionalized Spiropyran Monolayer Electrodes', *Langmuir*, **13**, 1783–1790 (1997).
179. F. Malem, D. Mandler, 'Self-assembled Monolayers in Electroanalytical Chemistry: Application of *w*-Mercapto Carboxylic Acid Monolayers for the Electrochemical Detection of Dopamine in the Presence of a High Concentration of Ascorbic Acid', *Anal. Chem.*, **65**, 37–41 (1993).
180. I. Turyan, D. Mandler, 'Characterization and Electroanalytical Application of Omega-mercaptoalkanesulfonic Acid Monolayers on Gold', *Israel J. Chem.*, **37**, 225–233 (1997).
181. J. Li, L. Ding, E. Wang, S. Dong, 'The Ion Selectivity of Monensin Incorporated Phospholipid/Alkane Thiol Bilayers', *J. Electroanal. Chem.*, **414**, 17–21 (1996).
182. I. Turyan, D. Mandler, 'Self-assembled Monolayers in Electroanalytical Chemistry: Application of *w*-Mercaptocarboxylic Acid Monolayers for Electrochemical Determination of Ultralow Levels of Cadmium(II)', *Anal. Chem.*, **66**, 58–63 (1994).
183. I. Turyan, D. Mandler, 'Selective Determination of Cr(VI) by a Self-assembled Monolayer-based Electrode', *Anal. Chem.*, **69**, 894–897 (1997).
184. M. Gerlache, Z. Senturk, G. Quarin, J.-M. Kauffmann, 'Self-assembled Monolayer Gold Electrode for Surfactant Analysis', *J. Solid State Electrochem.*, **1**, 155–160 (1997).
185. M.T. Rojas, R. Königer, J.F. Stoddart, A.E. Kaifer, 'Supported Monolayers Containing Preformed Binding Sites. Synthesis and Interfacial Binding Properties of a Thiolated β-Cyclodextrin Derivative', *J. Am. Chem. Soc.*, **117**, 336–343 (1995).
186. Y. Maeda, T. Fukuda, H. Yamamoto, H. Kitano, 'Regio- and Stereoselective Complexation by a Self-assembled Monolayer of Thiolated Cyclodextrin on a Gold Electrode', *Langmuir*, **13**, 4187–4189 (1997).
187. A.E. Kaifer, 'Functionalized Self-assembled Monolayers Containing Preformed Binding Sites', *Israel J. Chem.*, **36**, 389–397 (1996).
188. M. Lahav, K.T. Ranjit, E. Katz, I. Willner, 'Photostimulated Interactions of Bipyridinium-Azobenzene With a beta-Aminocyclodextrin Monolayer-functionalized Electrode: An Optoelectronic Assembly for the Amperometric Transduction of Recorded Optical Signals', *Israel J. Chem.*, **37**, 185–195 (1997).
189. M.T. Rojas, A.E. Kaifer, 'Molecular Recognition at the Electrode–Solution Interface. Design, Self-assembly, and Interfacial Binding Properties of a Molecular Sensor', *J. Am. Chem. Soc.*, **117**, 5883–5884 (1995).
190. I. Rubinstein, S. Steinberg, Y. Tor, A. Shanzer, J. Sagiv, 'Ionic Recognition and Selective Response in Self-assembling Monolayer Membranes on Electrodes', *Nature*, **332**, 426–429 (1988).
191. S. Steinberg, Y. Tor, E. Sabatani, I. Rubinstein, 'Ion-selective Monolayer Membranes Based on Self-assembling Tetradentate Ligand Monolayers on Gold Electrodes. 2. Effect of Applied Potential on Ion Binding', *J. Am. Chem. Soc.*, **113**, 5176–5182 (1991).
192. S. Steinberg, I. Rubinstein, 'Ion-selective Monolayer Membranes Based upon Self-assembling Tetradentate Ligand Monolayers on Gold Electrodes. 3. Application as Selective Ion Sensors', *Langmuir*, **8**, 1183–1187 (1992).
193. J.-H. Fuhrhop, T. Bedurke, M. Gnade, J. Schneider, K. Doblhofer, 'Hydrophobic Gaps of Steroid Size in a Surface Monolayer Collect 1,2-*trans*-Cyclohexanediol and Glucose from Bulk Water', *Langmuir*, **13**, 455–459 (1997).
194. K.T. Kinnear, H.G. Monbouquette, 'Direct Electron Transfer to *Escherichia coli* Fumarate Reductase in Self-assembled Alkanethiol Monolayers on Gold Electrodes', *Langmuir*, **9**, 2255–2257 (1993).

195. D.D. Schlereth, R.P.H. Kooyman, 'Self-assembled Monolayers With Biospecific Affinity for NAD(H)-dependent Dehydrogenases: Characterization by Surface Plasmon Resonance Combined With Electrochemistry ''in situ'' ', *J. Electroanal. Chem.*, **444**, 231–240 (1998).
196. Y. Sato, H. Itoigawa, K. Uosaki, 'Unidirectional Electron Transfer at Self-assembled Monolayers of 11-Ferrocenyl-1-undecanethiol on Gold', *Bull. Chem. Soc. Jpn*, **66**, 1032–1037 (1993).
197. H.-G. Hong, 'Kinetic Behavior of a Rotating Gold Disk Electrode Modified With a Self-assembled Ferrocenecarboxamidoyl Monolayer Based on Zirconium Phosphonate', *Electrochim. Acta*, **42**, 2319–2326 (1997).
198. A. Doron, M. Portnoy, M. Lion-Dagan, E. Katz, I. Willner, 'Amperometric Transduction and Amplification of Optical Signals Recorded by a Phenoxynaphthacenequinone Monolayer Electrode: Photochemical and pH-gated Electron Transfer', *J. Am. Chem. Soc.*, **118**, 8937–8944 (1996).
199. J.E. Hutchison, T.A. Postlethwaite, C. Chen, K.W. Hathcock, R.S. Ingram, W. Ou, R.W. Linton, R.W. Murray, D.A. Tyvoll, L.L. Chang, J.P. Collman, 'Electrocatalytic Activity of an Immobilized Cofacial Diporphyrin Depends on the Electrode Material', *Langmuir*, **13**, 2143–2148 (1997).
200. J. Zak, H. Yuan, M. Ho, L.K. Woo, M.D. Porter, 'Thiol-derivatized Metalloporphyrins: Monomolecular Films for the Electrocatalytic Reduction of Dioxygen at Gold Electrodes', *Langmuir*, **9**, 2772–2774 (1993).
201. E. Katz, H.-L. Schmidt, 'Gold Electrode Modification With a Monolayer of 1,4,5,8-Naphthalenetetrone as a Four-electron Transfer Mediator: Electrochemical Reduction of Dioxygen to Water', *J. Electroanal. Chem.*, **368**, 87–94 (1994).
202. E. Lorenzo, L. Sanchez, F. Pariente, J. Tirado, H.D. Abruna, 'Thermodynamics and Kinetics of Adsorption and Electrocatalysis of NADH Oxidation With a Self-assembling Quinone Derivative', *Anal. Chim. Acta*, **309**, 79–88 (1995).
203. M. Kunitake, K. Akiyoshi, K. Kawatana, N. Nakashima, O. Manabe, 'Transmembrane Electron Transfer Through a New Hydrophobic Mediator Incorporated in Monolayer Assemblies on Metal Electrodes', *J. Electroanal. Chem.*, **292**, 277–280 (1990).
204. M. Maskus, H.D. Abruna, 'Syntheses and Characterization of Redox-active Metal Complexes Sequentially Self-assembled onto Gold Electrodes via a New Thiol-terpyridine Ligand', *Langmuir*, **12**, 4455–4462 (1996).
205. D.N. Upadhyay, V. Yegnaraman, G.P. Rao, 'Hexacyanoferrate Modification of Gold Electrode Through Monolayer Approach', *Langmuir*, **12**, 4249–4252 (1996).
206. O. Pierrat, C. Bourdillon, J. Moiroux, J.-M. Laval, 'Enzymatic Electrocatalysis Studies of *Escherichia coli* Pyruvate Oxidase, Incorporated into a Biomimetic Supported Bilayer', *Langmuir*, **14**, 1692–1696 (1998).
207. S. Rubin, J.T. Chow, J.P. Ferraris, T.A. Zawodzinski, 'Electrical Communication Between Components of Self-assembled Mixed Monolayers', *Langmuir*, **12**, 363–370 (1996).
208. V. Heleg-Shabtai, E. Katz, I. Willner, 'Assembly of Microperoxidase-11 and Co(II)-Protoporphyrin IX Reconstituted Myoglobin Monolayers on Au-electrodes: Integrated Bioelectrocatalytic Interfaces', *J. Am. Chem. Soc.*, **119**, 8121–8122 (1997).
209. V. Heleg-Shabtai, E. Katz, S. Levi, I. Willner, 'Microperoxidase-11 Functionalized Electrodes: An Active Monolayer Interface for the Electrocatalyzed Reduction of Co(II)-Protoporphyrin IX Reconstituted Myoglobin and for the Generation of Integrated Protein Electrodes for Bioelectrocatalyzed Hydrogenation of Acetylenes', *J. Chem. Soc., Perkin Trans.*, **2**, 2645 (1997).
210. I. Willner, B. Willner, 'Photoswitchable Biomaterials as Grounds for Optobioelectronic Devices', *Bioelectrochem. Bioenerg.*, **42**, 43–57 (1997).
211. C. Miller, M. Gratzel, 'Electrochemistry at w-Hydroxy Thiol Coated Electrodes. 2. Measurement of the Density of Electronic States Distributions for Several Outer-sphere Redox Couples', *J. Phys. Chem.*, **95**, 5225–5233 (1991).
212. A.M. Becka, C.J. Miller, 'Electrochemistry at w-Hydroxy Thiol Coated Electrodes. 3. Voltage Independence of the Electron Tunneling Barrier and Measurements of Redox Kinetics at Large Overpotentials', *J. Phys. Chem.*, **96**, 2657–2668 (1992).
213. C.E.D. Chidsey, 'Free Energy and Temperature Dependence of Electron Transfer at the Metal–Electrolyte Interface', *Science*, **251**, 919–922 (1991).
214. C.J. Miller, 'Heterogeneous Electron-transfer Kinetics at Metallic Electrodes', in *Physical Electrochemistry: Principles, Methods, and Applications*, ed. I. Rubinstein, Marcel Dekker, New York, 27–79, 1995.
215. E. Laviron, 'General Expression of the Linear Potential Sweep Voltammogram in the Case of Diffusionless Electrochemical Systems', *J. Electroanal. Chem.*, **101**, 19–28 (1979).
216. E. Laviron, 'A.C. Polarography and Faradaic Impedance of Strongly Adsorbed Electroactive Species. Part I. Theoretical and Experimental Study of a Quasi-reversible Reaction in the Case of a Langmuir Isotherm', *J. Electroanal. Chem.*, **97**, 135–149 (1979).
217. J.H. Reeves, S. Song, E.F. Bowden, 'The Application of Square wave Voltammetry to Strongly Adsorbed Quasireversible Redox Molecules', *Anal. Chem.*, **65**, 683–688 (1992).
218. T.M. Nahir, R.A. Clark, E.F. Bowden, 'Linear Sweep Voltammetry of Irreversible Electron Transfer in Surface-confined Species Using Marcus Theory', *Anal. Chem.*, **66**, 2595–2598 (1994).
219. S. Ye, A. Yashiro, Y. Sato, K. Uosaki, 'Electrochemical in situ FT-IRRAS Studies of a Self-assembled

Monolayer of 2-(11-Mercaptoundecyl)hydroquinone', *J. Chem. Soc., Faraday Trans.*, **92**, 3813–3821 (1996).
220. J.F. Smalley, S.W. Feldberg, C.E.D. Chidsey, M.R. Linford, M.R. Newton, Y.-P. Liu, 'The Kinetics of Electron Transfer Through Ferrocene-terminated Alkanethiol Monolayers on Gold', *J. Phys. Chem.*, **99**, 13141–13149 (1995).
221. H.O. Finklea, L. Liu, M.S. Ravenscroft, S. Punturi, 'Multiple Electron Tunneling Paths across Self-assembled Monolayers of Alkane Thiols With Attached Ruthenium(II/III) Redox Centers', *J. Phys. Chem.*, **100**, 18852–18858 (1996).
222. K. Weber, L. Hockett, S. Creager, 'Long-range Electronic Coupling Between Ferrocene and Gold in Alkanethiolate-based Monolayers on Electrodes', *J. Phys. Chem. B*, **101**, 8286–8291 (1997).
223. Z.Q. Feng, S. Imabayashi, T. Kakiuchi, K. Niki, 'Long Range Electron-transfer Reaction Rates to Cytochrome *c* Across Long- and Short-chain Alkanethiol Self-assembled Monolayers: Electroreflectance Studies', *J. Chem. Soc., Faraday Trans.*, **93**, 1367–1370 (1997).
224. S. Song, R.A. Clark, E.F. Bowden, M.J. Tarlov, 'Characterization of Cytochrome *c*/Alkanethiolate Structures Prepared by Self-assembly on Gold', *J. Phys. Chem.*, **97**, 6564–6572 (1993).
225. C. Liang, M.D. Newton, 'Ab Initio Studies of Electron Transfer. 2. Pathway Analysis for Homologous Organic Spacers', *J. Phys. Chem.*, **97**, 3199–3211 (1993).
226. C.-P. Hsu, R.A. Marcus, 'A Sequential Formula for Electronic Coupling in Long Range Bridge-assisted Electron Transfer. Formulation of Theory and Application to Alkane Thiol Monolayers', *J. Chem. Phys.*, **106**, 584–598 (1997).
227. J. Cheng, G. Saghi-Szabo, J.A. Tossell, C.J. Miller, 'Modulation of Electronic Coupling Through Self-assembled Monolayers via Internal Chemical Modification', *J. Am. Chem. Soc.*, **118**, 680–684 (1996).
228. S. Terrettaz, A.M. Becka, M.J. Traub, J.C. Fettinger, C.J. Miller, '*w*-Hydroxythiol Monolayers at Au Electrodes. 5. Insulated Voltammetric Studies of Cyano/Bipyridyl Iron Complexes', *J. Phys. Chem.*, **99**, 11216–11224 (1995).
229. K. Weber, S.E. Creager, 'Voltammetry of Redox-active Groups Irreversibly Adsorbed onto Electrodes. Treatment Using the Marcus Relation Between Rate and Overpotential', *Anal. Chem.*, **66**, 3164–3172 (1994).
230. H.O. Finklea, D.D. Hanshew, 'Electron-transfer Kinetics in Organized Thiol Monolayers With Attached Penta-ammine(pyridine)ruthenium Redox Centers', *J. Am. Chem. Soc.*, **114**, 3173–3181 (1992).
231. H.O. Finklea, M.S. Ravenscroft, D.A. Snider, 'Electrolyte and Temperature Effects on Long Range Electron Transfer Across Self-assembled Monolayers', *Langmuir*, **9**, 223–227 (1993).
232. S. Terrettaz, J. Cheng, C.J. Miller, 'Kinetic Parameters for Cytochrome *c* via Insulated Electrode Voltammetry', *J. Am. Chem. Soc.*, **118**, 7857–7858 (1996).
233. J. Cheng, S. Terrettaz, J.I. Blankman, C.J. Miller, B. Dangi, R.D. Guiles, 'Electrochemical Comparison of Heme Proteins by Insulated Electrode Voltammetry', *Israel J. Chem.*, **37**, 259–266 (1997).
234. A. Kumar, H.A. Biebuyck, N.L. Abbott, G.M. Whitesides, 'The Use of Self-assembled Monolayers and a Selective Etch to Generate Patterned Gold Features', *J. Am. Chem. Soc.*, **114**, 9188–9189 (1992).
235. A. Kumar, H.A. Biebuyck, G.M. Whitesides, 'Patterning Self-assembled Monolayers: Applications in Materials Science', *Langmuir*, **10**, 1498–1511 (1994).
236. N.L. Abbott, D.R. Rolison, G.M. Whitesides, 'Combining Micromachining and Molecular Self-assembly to Fabricate Microelectrodes', *Langmuir*, **10**, 2672–2682 (1994).
237. Y. Yamamoto, H. Nishihara, K. Aramaki, 'Self-assembled Layers of Alkanethiols on Copper for Protection Against Corrosion', *J. Electrochem. Soc.*, **140**, 436–443 (1993).
238. M. Itoh, H. Nishihara, K. Aramaki, 'A Chemical Modification of Alkanethiol Self-assembled Monolayers With Alkyltrichlorosilanes for the Protection of Copper Against Corrosion', *J. Electrochem. Soc.*, **141**, 2018–2023 (1994).
239. M. Ishibashi, M. Itoh, H. Nishihara, K. Aramaki, 'Permeability of Alkane Thiol Self-assembled Monolayers Adsorbed on Copper Electrodes to Molecular Oxygen Dissolved in 0.5 M Na_2SO_4 Solution', *Electrochim. Acta*, **41**, 241–248 (1996).
240. R. Haneda, H. Nishihara, K. Aramaki, 'Chemical Modification of an Alkanethiol Self-assembled Layer to Prevent Corrosion of Copper', *J. Electrochem. Soc.*, **144**, 1215–1221 (1997).
241. R. Haneda, K. Aramaki, 'Protection of Copper Corrosion by an Ultrathin Two-dimensional Polymer Film of Alkanethiol Monolayer', *J. Electrochem. Soc.*, **145**, 1856–1861 (1998).
242. P.E. Laibinis, G.M. Whitesides, 'Self-assembled Monolayers of *n*-Alkanethiolates on Copper are Barrier Films That Protect the Metal against Oxidation by Air', *J. Am. Chem. Soc.*, **114**, 9022–9028 (1992).
243. I. Rubinstein, J. Rishpon, E. Sabatani, A. Redondo, S. Gottesfeld, 'Morphology Control in Electrochemically Grown Conducting Polymer Films. 1. Precoating the Metal Substrate With an Organic Monolayer', *J. Am. Chem. Soc.*, **112**, 6135–6136 (1990).
244. E. Sabatani, Y. Gafni, I. Rubinstein, 'Morphology Control in Electrochemically Grown Conducting Polymer Films. 3. A Comparative Study of Polyaniline Films on Bare Gold and on Gold Pretreated With *p*-Aminothiophenol', *J. Phys. Chem.*, **99**, 12305–12311 (1995).

245. W.A. Hayes, C. Shannon, 'Nanometer-scale Patterning of Surfaces Using Self-assembly Chemistry. 3. Template-directed Growth of Polymer Nanostructures on Organothiol Self-assembly Mixed Monolayers', *Langmuir*, **14**, 1099–1102 (1998).
246. J. Lukkari, K. Kleemola, M. Meretoja, T. Ollonqvist, J. Kankare, 'Electrochemical Post-self-assembly Transformation of 4-Aminothiophenol Monolayers on Gold Electrodes', *Langmuir*, **14**, 1705–1715 (1998).
247. R.J. Willicut, R.L. McCarley, 'Surface-confined Monomers on Electrode Surfaces. 1. Electrochemical and Microscopic Characterization of w-(*N*-Pyrrolyl)alkanethiol Self-assembled Monolayers on Au', *Langmuir*, **11**, 296–301 (1995).
248. C.N. Sayre, D.M. Collard, 'Self-assembled Monolayers of Pyrrole-containing Alkanethiols on Gold', *Langmuir*, **11**, 302–306 (1995).
249. D.B. Wurm, S.T. Brittain, Y.-T. Kim, 'Organic Monolayers as Nucleation Sites for Epitaxial Growth. 1. Electrochemical Polymerization of *N*-Alkylpyrrole', *Langmuir*, **12**, 3756–3758 (1996).
250. D.B. Wurm, K. Zong, Y.-H. Kim, Y.-T. Kim, M. Shin, I.C. Jeon, 'Electrochemical Quartz Crystal Microbalance Studies of Polypyrrole Growth on Self-assembled Nucleation Sites', *J. Electrochem. Soc.*, **145**, 1483–1488 (1998).

Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

Manuel P. Soriaga
Texas A&M University, Texas, USA

1 Surface Characterization Techniques 2
1.1 Surface Spectroscopy with Low-energy Electrons 2
1.2 Low-energy Electron Diffraction 3
1.3 Reflection High-energy Electron Diffraction 3
1.4 Auger Electron Spectroscopy 4
1.5 X-ray Photoelectron Spectroscopy 6
1.6 High-resolution Electron Energy Loss Spectroscopy 8
1.7 Work Function Measurements 9
1.8 Temperature Programmed Desorption 10
1.9 Instrument Designs 11
2 Fundamental Aspects 14
2.1 The Emersion Process 14
2.2 Perturbations Caused by Evacuation and Surface Analysis 14
3 Representative Studies 16
3.1 Electrochemical Double Layer 16
3.2 Underpotential Electrodeposition and Anodic Dissolution 16
3.3 Molecular Adsorption 18
3.4 Solvent–Electrode Interactions 18
3.5 Group IB Electrodes 19
3.6 Group VIII Electrodes 19
3.7 Carbon Monoxide 19
3.8 Other Organic Compounds 19
4 Epilogue 21
Acknowledgments 21
Abbreviations and Acronyms 21
Related Articles 21
Appendix 1 21
Ultrahigh Vacuum Electrochemistry Bibliography 21
References 24

The study of processes that transpire at heterogeneous interfaces is an exceedingly difficult proposition. No single experimental technique can ever hope to unravel all the nuances of heterogeneous reactions; hence, in surface science, the use of multiple complementary methods is not uncommon. Ultrahigh vacuum electrochemistry (UHV/EC) is a term ascribed to the approach that rests upon the integration of classical electrochemical methods with surface-sensitive analytical techniques; this strategy parallels that successfully implemented in the study of gas–solid heterogeneous catalysis. The unique surface sensitivity of the techniques adopted emanates from the use of particles (e.g. ions or electrons) that serve to interrogate the outermost layer(s) of the electrode. This surface sensitivity is tempered by the requirement that the analysis be performed in an environment (outside the electrochemical cell) that does not impede the mean-free paths of the probe particles. Since its inception in the early 1970s, more than a thousand UHV/EC-based studies have been published; most of the work involved polycrystalline materials and focused on the elemental composition at the electrode surface. While its importance in the study of polycrystalline surfaces cannot be trivialized, the greater value of UHV/EC appears to be in its ability to help resolve fundamental issues that intertwine interfacial structure and composition with electrochemical reactivity. It is in this context that the present review is written.

A complete mechanism of an electrochemical reaction must incorporate all the physical and chemical interactions that arise between an electrified surface and its environment. The extent of such interactions depends upon several factors such as solvent, supporting electrolyte, electrode potential, reactant concentration, electrode material and surface crystallographic orientation. The traditional approach is based upon a thermodynamic treatment of the interface and its response to external perturbations. Interpretation of the results relies on phenomenological models of the interface. Although a thermodynamic treatment cannot be ignored, the need for an atomic-level view has long been realized. One approach[1–12] towards the establishment of an atomic-level description parallels that successfully implemented in the study of gas–solid heterogeneous catalysis; it rests upon the integration of classical electrochemical methods with surface-sensitive analytical techniques. The analytical methods that exhibit surface sensitivity are based upon the mass-selection and/or energy-discrimination of electrons, ions, atoms or molecules scattered from solid surfaces. These particles have shallow escape depths; hence the information they bear is characteristic of the near-surface layers. Their short mean-free paths, however, necessitate a high vacuum ($<10^{-6}$ torr) environment. The application of such surface techniques to electrochemistry requires that the analysis be performed outside the electrochemical cell. The possibility of structural and compositional changes that accompany the removal of the electrode from solution is

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the major concern in UHV/EC studies. Although a myriad of surface analytical techniques is currently available, those actually employed in UHV/EC have been limited to low-energy electron diffraction (LEED), Auger electron spectroscopy (AES), X-ray photoelectron spectroscopy (XPS), high-resolution electron energy loss spectroscopy (HREELS), reflection high-energy electron diffraction (RHEED), work-function changes, and thermal desorption mass spectrometry (TDMS).

While most vacuum-based analytical methods do not require single-crystal surfaces, the use of uniform (monocrystalline) surfaces is a necessary aspect for fundamental studies. The low-index crystallographic faces [(100), (110) and (111)] have been widely used because of their low free energies, high symmetries, and relative stabilities. In addition, it may be possible to reconstruct the overall behavior of polycrystalline electrodes from the individual properties of the low-index planes.[1–4] *A handful of procedures for the preparation and preservation of well-defined single-crystal surfaces have been described.*[5–8] *The verification or identification of initial, intermediate, and final interfacial structures and compositions is an essential ingredient in electrochemical surface science.*

1 SURFACE CHARACTERIZATION TECHNIQUES

1.1 Surface Spectroscopy with Low-energy Electrons

The main difficulty in surface analysis lies in the exceedingly low population of atoms at the interface (10^{15} atoms cm^{-2}) relative to that in the bulk (10^{23} atoms cm^{-3}). Experiments intended to examine the physical and chemical properties of surfaces must employ methods that interact only with the interfacial layers; the majority of such techniques take advantage of the surface sensitivity of *low-energy* electrons. The surface sensitivity arises because the mean-free-path of an electron through a solid is dependent upon its kinetic energy. As shown in the "universal curve" reproduced in Figure 1, the electron mean-free-path falls to a minimum (0.4 to 2 nm) when the kinetic energy is between 10 and 500 eV. That is, information derived from experimental techniques based upon low-energy-electron incidence onto and/or emergence from surfaces will exclusively be from the topmost surface layers.

A solid surface subjected to a beam of electrons of incident (primary) energy E_p gives rise to the appearance of backscattered and emitted electrons; a plot of the number of electrons $N(E)$ as a function of energy E of these electrons is shown in Figure 2.[9–15] This spectrum can be divided into four regions according to the origin of the scattered electrons: (i) true secondary electrons, created as a result of inelastic interactions between the incident and bound electrons, give rise to the prominent broad band at the lower end of the spectrum; (ii) Auger electrons emitted and primary electrons scattered due to interactions with electronic states in the solid account for the small peaks in the medium-energy range of the spectrum; (iii) primary electrons scattered upon interactions with the vibrational states of the surface yield peaks close to the elastic peak; (iv) primary electrons scattered

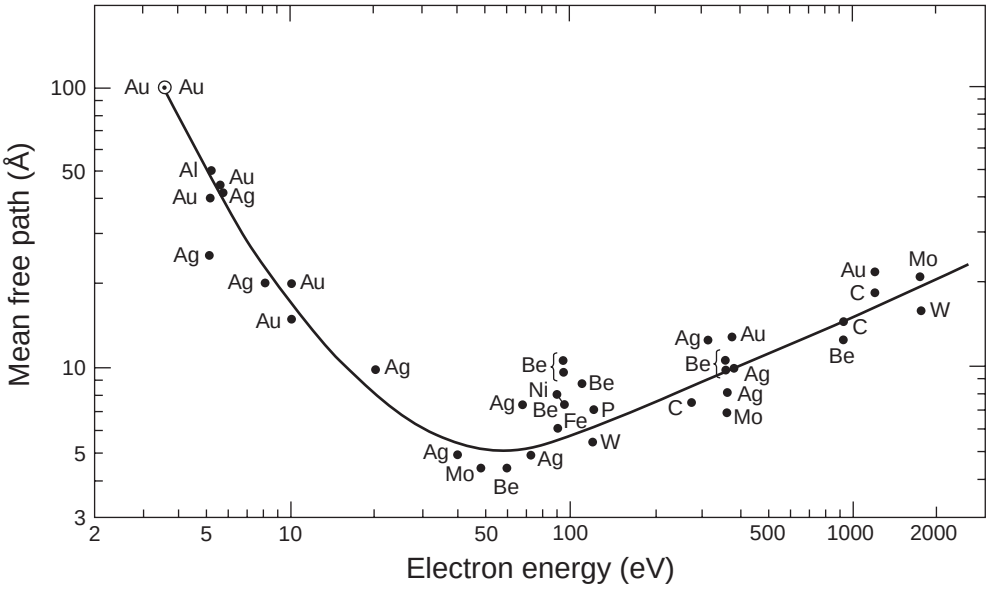

Figure 1 "Universal curve": electron mean free path as a function of electron kinetic energy. (Reproduced by permission from Somorjai.[9])

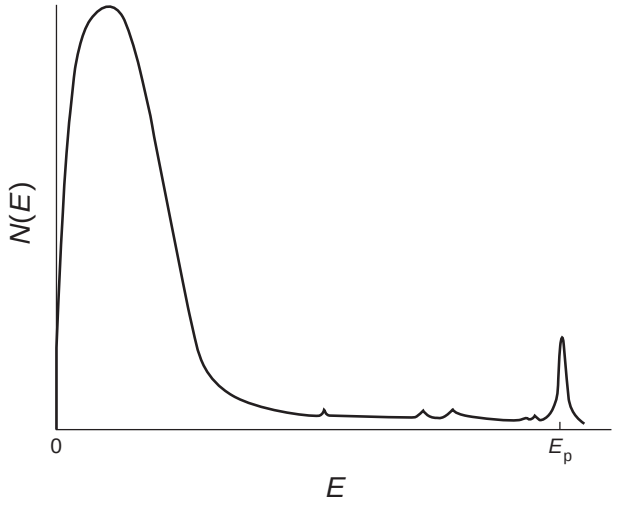

Figure 2 Experimental number $[N(E)]$ of scattered electrons of energy E as a function of electron energy. (Reproduced by permission from Somorjai.[9])

elastically, which comprise only a few percent of the total incident electrons, appear at E_p. The spectral regions (ii) to (iv) have been exploited in modern surface analysis.

1.2 Low-energy Electron Diffraction

In this method,[9–15] the surface is irradiated with a monoenergetic beam of electrons and the elastically backscattered electrons are collected onto a phosphor screen. The virtue of LEED as a surface structural technique is a result of the low kinetic energies used (50 to 500 eV): (i) the electron mean free path is at a minimum; (ii) the de Broglie wavelengths, $\lambda_e \approx (150/E_e)^{1/2}$ (where E_e is in electronvolts and λ_e is in angstroms) correspond to crystal lattice dimensions; and (iii) electron backscattering is strong which minimizes incident electron fluxes at, and subsequent scattering from, nonsurface layers. In LEED, the presence (or absence) of diffraction patterns on the fluorescent screen is a consequence of the order (or disorder) of the atomic arrangements at the surface.

The locations of the diffraction spots define the reciprocal lattice of the real surface. The real-space surface structure can be reconstructed from the reciprocal lattice vectors.[9–15] The *coherence width* of electron beam sources in LEED is typically 10 nm. That is, sharp diffraction features appear only if well-ordered domains are *at least* 10 nm × 10 nm in size; diffraction from smaller domains leads to beam broadening. The analysis of LEED data based solely upon the geometry of the diffraction spots provides information on the periodicity of the electron scatterers on the surface. In some favorable instances, other information such as adsorbate coverages or point group symmetries can also be inferred. However, the actual location of the atoms within the surface lattice cannot be determined without an analysis of the intensities of the diffracted beams. Surface crystallography by LEED relies upon a comparison of the measured diffraction intensities with those calculated for model structures. LEED simulations are difficult because of multiple electron scattering.[9–15]

Two schemes are employed for the notation of interfacial adlattice structures. The matrix notation, which is applicable to any system, is based upon the relationship between the real-space lattice vectors of the *adsorbate* mesh (two-dimensional lattice) and the *substrate* mesh. For example, if the adsorbate unit cell vectors $\mathbf{a}'$ and $\mathbf{b}'$ are related to those of the substrate mesh, $\mathbf{a}$ and $\mathbf{b}$, according to Equations (1) and (2):

$$\mathbf{a}' = m_{11}\mathbf{a} + m_{12}\mathbf{b} \tag{1}$$

$$\mathbf{b}' = m_{21}\mathbf{a} + m_{22}\mathbf{b} \tag{2}$$

Then the matrix $\mathbf{M}$ defined by the coefficients m_{ij}, $\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$, denotes the real-space surface structure. The other method, known as the Wood notation,[16] is more widely used but is applicable only if the angle between $\mathbf{a}'$ and $\mathbf{b}'$ is the same as that between $\mathbf{a}$ and $\mathbf{b}$. The adlattice structure is labeled using the general form $(n \times m)R\phi^\circ$ or $c(n \times m)R\phi^\circ$, where c designates a centered unit cell, $R\phi^\circ$ the angle of rotation of the adsorbate unit cell relative to the substrate unit mesh, and n and m are scale factors relating the adsorbate and substrate unit cell vectors $|\mathbf{a}'| = n|\mathbf{a}|$ and $|\mathbf{b}'| = m|\mathbf{b}|$.

A schematic diagram of a typical LEED instrument is shown in Figure 3.[17] The LEED "optics" consists of a phosphor-coated hemispherical screen at the center of which is a normal-incidence, electrostatically focused electron gun. In front are three concentric grids; the outer grid is held at ground potential, while the inner two are maintained at a voltage just below that of the electron gun in order to reject *inelastically* backscattered electrons. The elastically diffracted electrons which pass through the suppressor grids are accelerated onto the fluorescent screen by a 5-kV potential applied to the screen. For quantitative intensity measurements, additional provisions (e.g. Faraday cup, spot photometer, or digital camera) will be necessary.

A typical LEED pattern, that for an iodine-coated Pd(111) surface, is shown in Figure 4 along with the suggested real-space adlattice structure, Pd(111) – $(\sqrt{3} \times \sqrt{3})R30^\circ$ – I.

1.3 Reflection High-energy Electron Diffraction

RHEED[2,12–15] represents an alternative to LEED. The principal difference between the two techniques is that,

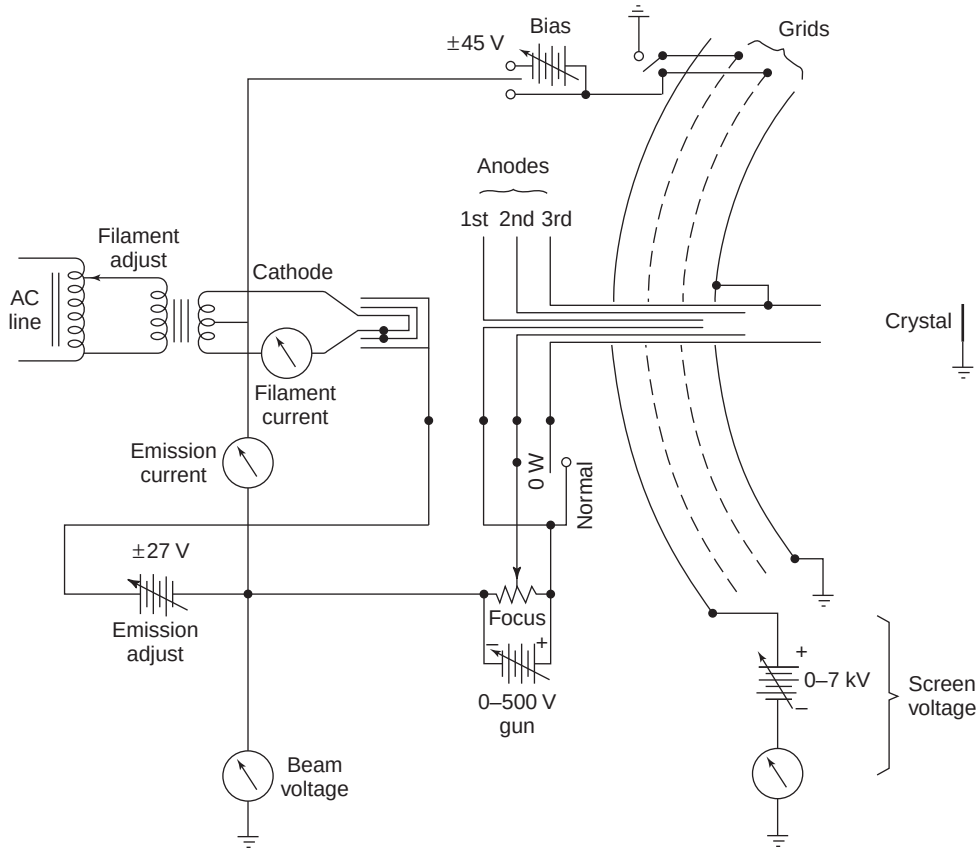

Figure 3 Schematic diagram of a LEED apparatus. (Reproduced by permission from Adamson.[17])

whereas low-energy electrons are utilized in LEED, high-energy (30 to 100 keV) electrons are employed in RHEED. At such high energies, the mean-free-paths of the incident electrons are long (10 to 100 nm) and elastic scattering is predominantly in the forward direction. Hence, to afford the required surface sensitivity, RHEED experiments are performed at small angles ($<5^\circ$) of incidence and diffraction. Energy filtering is not a requirement in RHEED because of the large energy difference between the elastically and inelastically scattered electrons; post-acceleration is likewise unnecessary as the primary electrons are sufficiently energetic to produce fluorescence on the phosphor screen.

Figure 5 shows LEED and RHEED patterns of gold films evaporated on glass and on mica;[18] the film deposited on glass (at room temperature) is rough but that on mica (at elevated temperatures) is smooth and well-ordered. The ordered-surface diffraction is manifested in LEED by distinct spots and by sharp streaks in RHEED. This difference is due to the fact that, in reciprocal space, the surface layer is represented by perpendicularly-oriented *rods* that pass through the points in the reciprocal net. In normal-incidence LEED, only a section *perpendicular* through these reciprocal lattice rods is displayed, leading to the spot pattern. In grazing-incidence RHEED, a section *parallel* through the rods is displayed, resulting in a streak pattern. For the same reason, RHEED is unable to detect changes in periodicity along the plane of incidence. Hence, if the full two-dimensional periodicity is to be established, the sample needs to be rotated.

RHEED is most useful in studies related to the structure and morphology of thin films and surface coatings. It is possible to continuously monitor film formation under deposition conditions since the front of the sample is unimpeded by either the electron source or analyzer. In view of its low-scattering-angle geometry, RHEED is sensitive to surface asperities.

1.4 Auger Electron Spectroscopy

AES is one of the more widely used techniques for surface elemental analysis.[9–15] In the Auger process, illustrated schematically in Figure 6, a core (K) level electron is emitted when a beam of electrons, typically with energies between 2 and 10 keV, is impinged onto the sample surface. In the decay process, an electron in an upper (L_I) level falls into the vacant core level and another electron

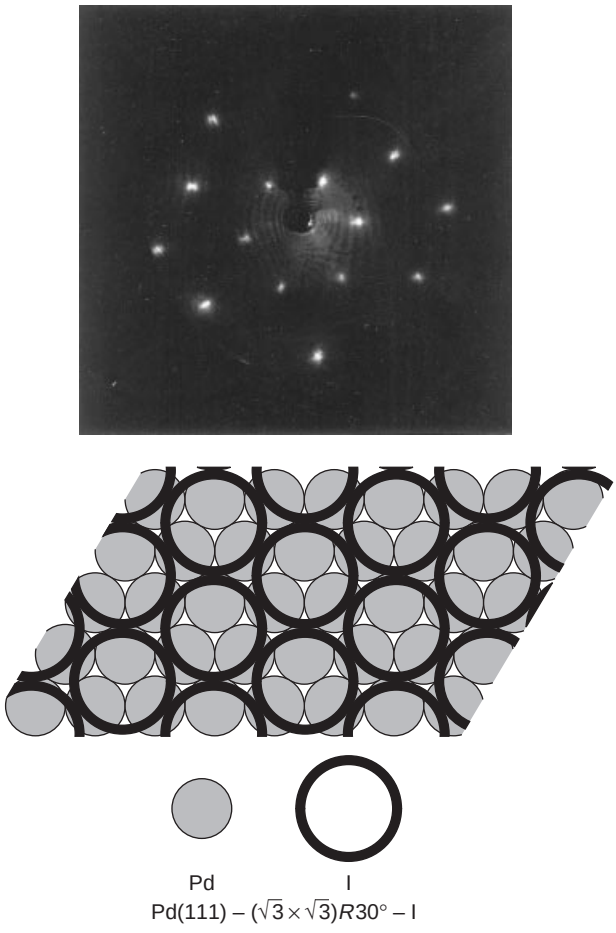

Figure 4 LEED pattern for a Pd(111) electrode coated with iodine from dilute aqueous NaI; also shown is the postulated adlattice structure.

in a different upper (L_{III}) level is ejected; the latter is the Auger electron. This particular Auger process is labeled KL_IL_{III} in order to specify the energy levels involved. The kinetic energy of the Auger electron is dependent upon the binding energies of the K, L_I, and L_{III} electrons but not on E_p, as shown in Equation (3):

$$E_{KL_IL_{III}} = E_K - E_{L_I} - E_{L_{III}} - e\varphi_{sp} \qquad (3)$$

where e is the electronic charge and φ_{sp} the spectrometer work function. The exact application of Equation (3) must realize that the energy difference is actually between singly ionized and doubly ionized states. Nevertheless, $E_{KL_IL_{III}}$, as obtained from empirical spectra, is characteristic of a given atom, which affords AES its element-specificity. It should be noted that, although the incident electrons are of high energies, AES is still a surface-sensitive method because the *emitted* Auger electrons are of much lower energies that fall within the minimum of the ''universal curve'' (Figure 1).

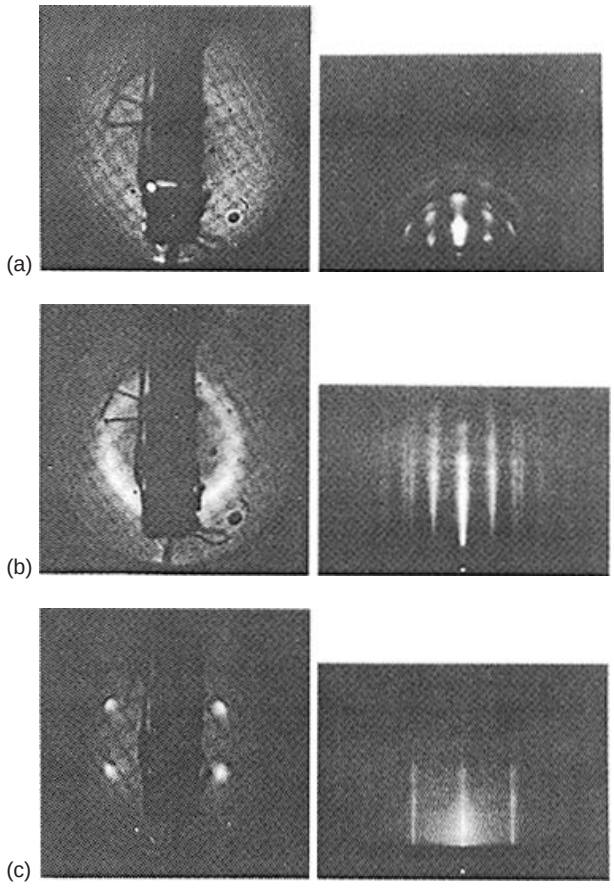

Figure 5 LEED and RHEED patterns for Au vapor deposited onto (a) glass at room temperature, (b) glass at 400 °C, and (c) mica at 360 °C. (Reproduced by permission of Zei et al.[18])

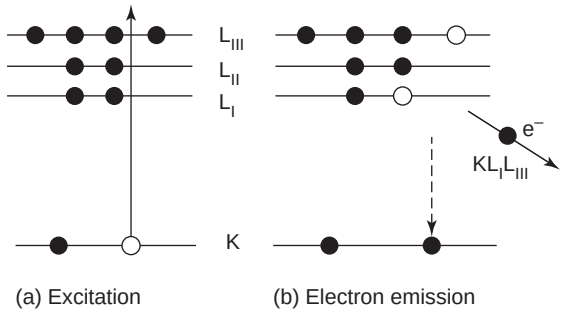

Figure 6 Schematic diagram of the Auger emission process: (a) excitation; (b) electron emission. The core (K) level electron is that which is ejected in XPS, while the Auger electron is the ejected electron designated as KL_IL_{III}.

An inherent difficulty in AES is that the Auger emission peaks are of very low intensities superimposed on a large (secondary emission) background. The usual approach to negate the background is a combination of electron-energy analysis with suitable modulation techniques. In multi-technique instrumentation that includes LEED,

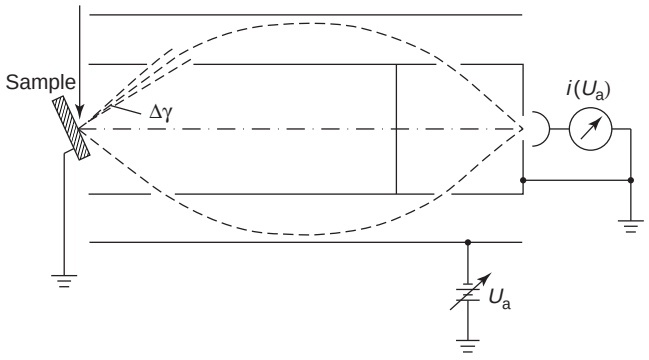

Figure 7 Schematic cross-section of a CMA. U_a is the potential applied between the two coaxial cylindrical electrodes, $\Delta\gamma$ is the angular spread of the electrons at the entrance slit, and $i(U_a)$ is the current at the exit aperture. (Reproduced by permission from Ertl and Kuppers.[11])

a retarding field analyzer (RFA) is most affordable since it makes use of the LEED optics. The sensitivity and resolution of an RFA, however, leave much to be desired. A cylindrical mirror analyzer (CMA), Figure 7, is thus more widely used. Energy analysis with a CMA is achieved by a negative ramp voltage applied to the outer cylinder while the inner cylinder is held at ground potential. Only electrons of the appropriate energy can pass unhindered through the CMA and into the detector, which is usually a channel electron multiplier. The pass energy of the CMA is modulated and then synchronously demodulated with a lock-in amplifier; the resultant spectrum is a derivative spectrum, $dN(E)/dE$, devoid of the large background. Newer instruments employ software-based modulation and filtering. Higher resolution can be achieved via a double-pass CMA or by the use of a cylindrical hemisphere analyzer (CHA), which is a double-focusing analyzer (see below).

A typical Auger spectrum is shown in Figure 8 for a Pd(100) surface exposed *serially* to dilute aqueous solutions of NaCl, NaBr and NaI. For quantitative and/or molecular compositional analysis, the derivative spectrum is difficult to process. An alternative approach, the collection of nonderivative Auger spectra, involves pulse counting electronics or direct current measurements; spectra thus generated can be deconvoluted by a fast Fourier transform algorithm[19] to obtain information on chemical shifts and lineshapes. Changes in Auger lineshapes reflect modifications in the valence band density of states.[20,21]

The use of derivative Auger spectra for the determination of adsorbate surface coverages has been the subject of numerous studies.[22] One method, the first to be used in surface electrochemical studies, makes use of Equation (4):[23–25]

$$\Gamma_a = \frac{I_a}{I_p \phi_c G_a} \tag{4}$$

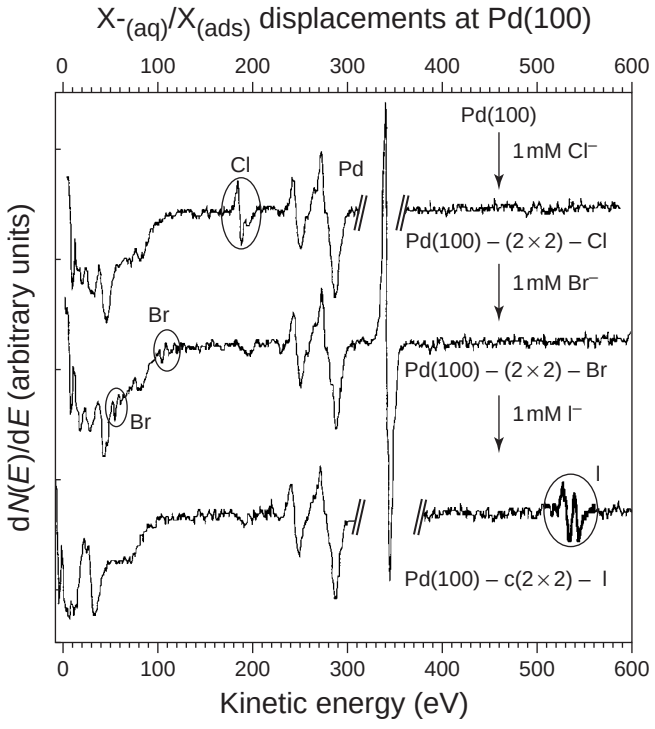

Figure 8 Auger electron spectra for a Pd(100) electrode after sequential exposure to 1 mM NaCl, NaBr, and NaI.

where Γ_a is the absolute packing density (mol cm^{-2}) of the adsorbate, I_a the Auger current for the adsorbate, I_p the primary beam current, ϕ_c the measured collection efficiency of the Auger spectrometer, and G_a is the calculated Auger electron yield factor.[26] I_a is obtained using Equation (5) by double integration of the adsorbate second-harmonic amplitude A_2 corrected for the clean-surface signal A_{2c}:[23]

$$I_a = \frac{4}{k^2} \int_0^{E_p} \int_0^{E} (A_2 - \phi_b A_{2c})\, dE'\, dE \tag{5}$$

where ϕ_b is the observed attenuation of the substrate signal by the adsorbed species, and k is the modulation amplitude. For simple adsorbates for which well-defined adlayers are available for calibration purposes, Equation (4) can be expressed in purely empirical terms by Equation (6):[24,25]

$$\Gamma_a = \left(\frac{I_a}{I_M^0} \quad \frac{1}{B_a} \right. \tag{6}$$

where I_M^0 is the Auger signal for the clean substrate, and B_a is a calibration factor.

1.5 X-ray Photoelectron Spectroscopy

This technique, originally referred to as electron spectroscopy for chemical analysis (ESCA),[27] is the other

widely used method for surface compositional analysis. In XPS, the solid surface is irradiated with X-rays which results in the ejection of a core-level electron. The kinetic energy E_{Kin} of the emitted photoelectron is given by Equation (7):

$$E_{Kin} = h\nu - E_B - e\varphi_{sp} \tag{7}$$

where $h\nu$ is the energy of the incident X-ray photon and E_B, within the framework of Koopman's theorem,[11,12] is the binding energy of the core-level electron. For studies with metals, it is convenient to reference E_B with respect to the Fermi level; the latter is determined from the onset of electron emission at the highest kinetic energy.

The XPS source consists of an anode material which, upon bombardment by high-energy electrons, emits X-rays. The emitted radiation can be rendered monochromatic either by Bragg diffraction or by the use of the characteristic emission lines of the anode; for Mg and Al anodes, these lines are 1253.6 eV (Mg Kα) and 1486.6 eV (Al Kα), respectively. It is important to note that, for $E_B < 700$ eV, the E_{Kin} of the ejected photoelectron will *not* fall within the minimum of the universal curve. In such a case, the surface sensitivity of XPS becomes minimal. This can be remedied either by the use of near-grazing incidence or by the detection of electrons emitted at small angles with respect to the surface plane.

To afford the high resolution required for meaningful XPS studies, energy analysis is usually based upon a CHA (Figure 9). A potential difference U_k is applied across the inner and outer hemispheres of radii R_1 and R_2, respectively. Electrons of energy eV_e are focused at the exit slit only if Equation (8) is satisfied:

$$U_k = V_e \left(\frac{R_2}{R_1} - \frac{R_1}{R_2} \right) \tag{8}$$

The CHA is double focusing since it focuses in two planes. The resolution of a CHA can be improved significantly by electron pre-retardation via either an RFA or a retarding lens system. XPS has also been performed with a double-pass CMA. Detection is typically done with a channel electron multiplier. Due to inherently weak intensities, signal averaging and other data processing routines are always employed.

Qualitative elemental analysis of sample surfaces relies upon the comparison of measured E_B values with those for reference materials. In this regard, it is critical to note that XPS of an insulating layer (e.g. a thick oxide film) will result in a net loss of electrons and, hence, an excess of positive charge at the interface; such positive-charge excess leads to shifts in the experimental E_B values. This problem can be effectively countered by irradiation of the surface, at low currents, with low-energy electrons.

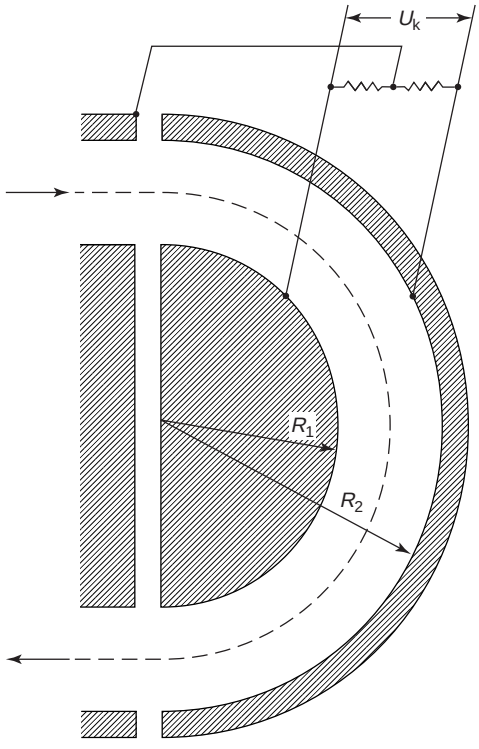

Figure 9 Schematic cross-section of the concentric hemisphere analyzer. (Reproduced by permission from Ertl and Kuppers.[11])

Quantitative analysis is based on the fact that the ionization cross-section of a core electron is essentially independent of the valence state of the element. Hence, the intensity will always be proportional to the number of atoms within the detected volume. For quantitative purposes, the area under the background-corrected peak is taken as the intensity. The intensity is a complicated function of several parameters some of which can be eliminated by the use of a reference state analyzed under identical conditions as the sample. If the spectrometer has a small aperture and the surface is uniformly irradiated, the equation for the intensity can be simplified to Equation (9):[11,28]

$$I_A = \sigma_A D L_A J_0 N_A \lambda_M G_1 \cos\theta_1 \tag{9}$$

where I_A is the integrated peak intensity for an element A, σ_A the photoionization cross-section, D the spectrometer detection efficiency, L_A the angular asymmetry of the emitted intensity with respect to the angle between the incidence and detection directions, J_0 the flux of primary photons, N_A the density of atoms A, λ_M the escape depth of the photoelectron, G_1 the spectrometer transmission, and θ_1 is the angle between the surface normal and the detection direction.

XPS is complementary to AES. The ionization cross-section for an Auger process decreases with E_B; the

latter, in turn, increases with atomic number Z. Hence, AES is most sensitive for $Z < 45$ elements. The one distinct advantage that XPS offers is in the determination of oxidation states. This is possible because E_B of the core-level electrons is influenced by changes in the chemical environment. In principle, identical information can be obtained from Auger peak energy shifts and lineshapes; in practice, however, the task is nontrivial.

XPS spectra of a smooth polycrystalline Ir foil electrode surface before and after pretreatment with iodine, are shown in Figure 10; the peaks at 62.1 and 65.2 eV represent surface iridium oxide.

1.6 High-resolution Electron Energy Loss Spectroscopy

Almost all of the incident electrons impinged at a solid surface undergo inelastic events that cause them to be backscattered at energies lower than the primary energy E_p. If E_l is the energy lost to the surface, peaks would appear in the energy distribution spectrum (Figure 2) at energies $\Delta E = E_p - E_l$. Such peaks, commonly referred to as electron energy loss peaks, are of several types according to the origin of the energy loss. For vibrational excitations, the energy losses are small since $E_{vib} < 4000\,cm^{-1} < 0.5\,eV$. Hence, the loss peaks due to vibrational interactions lie close to the elastic peak and can be observed only if the energy loss measurements are done at high resolution.

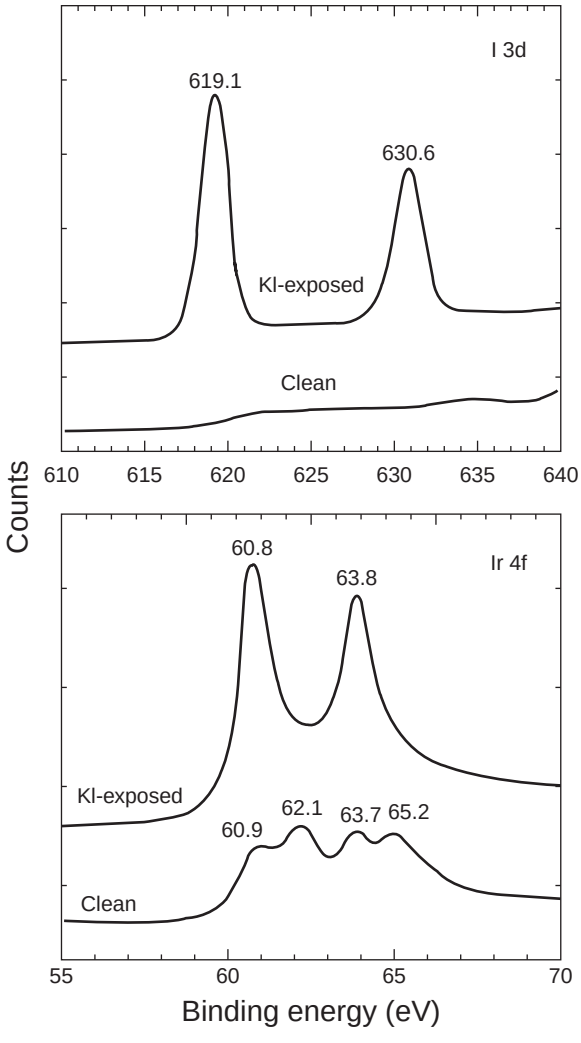

Figure 10 X-ray photoelectron spectra for a thermally annealed polycrystalline Ir foil electrode before and after pretreatment with iodine.

There are two mechanisms that give rise to vibrational HREELS:[9,11,12,29,30] dipole scattering and impact scattering. In dipole scattering, the incident electron interacts with the oscillating electric dipole moment induced by the vibration of surface species. Such interactions occur at long range and can be described either classically or quantum mechanically. Two important selection rules apply for surface dipole scattering: (i) only vibrations whose dynamic dipole moments perpendicular to the surface are nonzero contribute to HREELS spectra. This selection rule is the same as that for infrared reflection–absorption spectroscopy (IRAS);[31–34] (ii) the intensity distribution with respect to scattering angle is sharply peaked in the specular direction; that is, dipole-scattering loss peaks disappear when the backscattered electrons are collected at nonspecular angles.

Impact scattering, which can only be treated quantum mechanically, involves exceedingly short-range interactions between the incident electron and the oscillator at the surface. The surface dipole selection rules do not apply to impact scattering. Theoretical considerations have predicted, and experimental studies have confirmed, the following properties of this type of scattering mechanism:[29] (i) impact scattering vanishes in the specular direction; (ii) impact scattering is more likely to prevail at higher energies; (iii) strong dipole scatterers are weak impact scatterers; conversely, weak dipole scatterers are strong impact scatterers.

HREELS is an extremely sensitive technique. The limit of detection for strong dipole scatterers such as CO can be as low as 0.0001 monolayer; for weak scatterers such as hydrogen, the limit is 0.01 monolayer. In comparison, IRAS for chemisorbed CO, a strong absorber, is restricted to coverages above 0.1 monolayer. HREELS studies of non-CO organic molecules adsorbed at atomically smooth electrode surfaces are abundant; similar experiments using IRAS are meager. The energy accessible to HREELS ranges from $100\,cm^{-1}$ to $4000\,cm^{-1}$; IRAS detectors are not useful below $600\,cm^{-1}$. On the other hand, IRAS has higher resolution and can be utilized for experiments under electrochemical conditions.[33]

Figure 11 shows a schematic diagram of an HREELS spectrometer.[35] The energy of incident electrons can

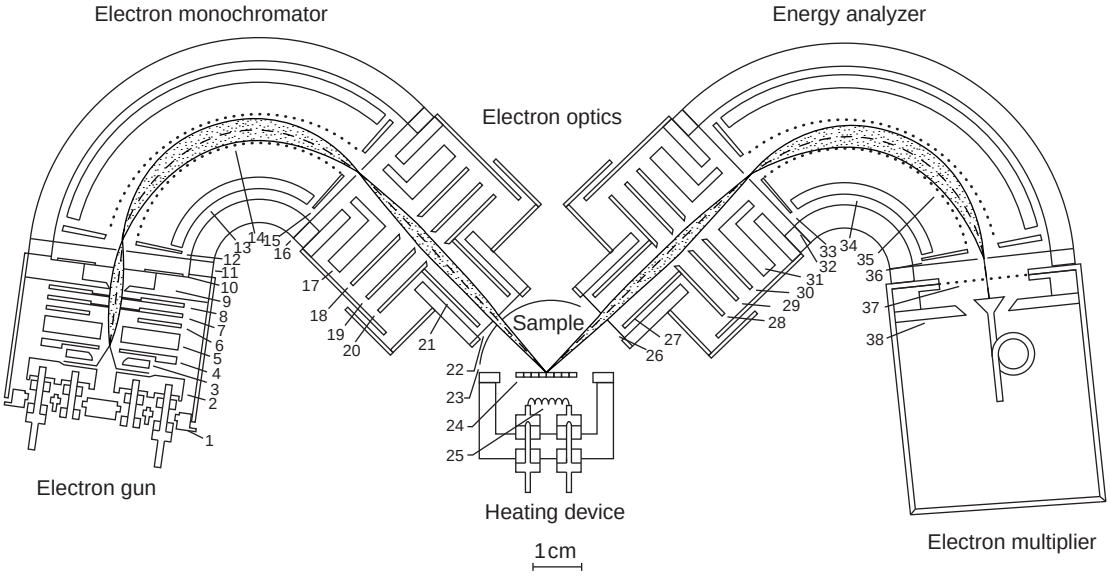

Figure 11 Schematic drawing of a tandem cylindrical deflector spectrometer used for HREELS. (Reproduced by permission from Froitzheim.[35])

be varied from 1 to 10 eV. To afford high resolution, energy monochromation and analysis are done either with a CMA, cylindrical deflector, or spherical deflector analyzers in combination with retarding field optics. Off-specular collection of the backscattered electrons is afforded by rotation of either the sample or the analyzer. Due to extremely low signals (10^{-10} A), continuous dynode electron multiplier detectors are widely used.

An example HREEL spectrum, that of 2,5-dihydroxybenzenesulfonate chemisorbed on a Pd(100) surface, is shown in Figure 12.

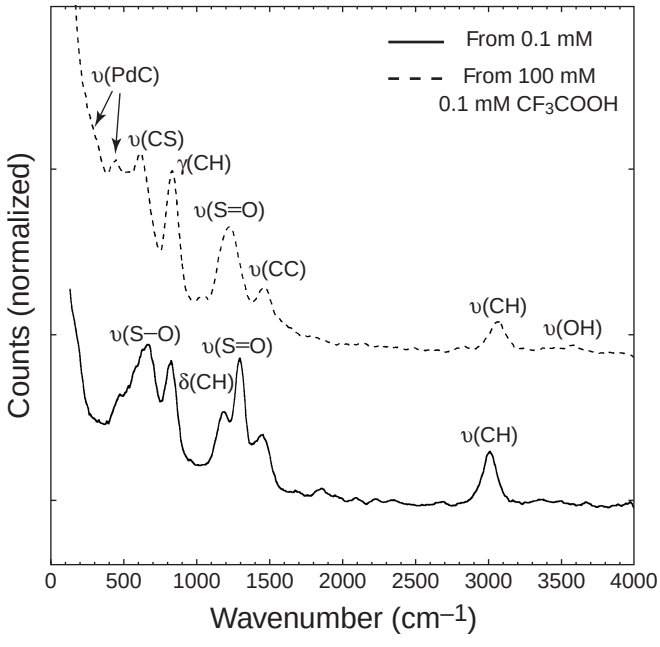

Figure 12 High-resolution electron energy loss spectra of the potassium salt of hydroquinone sulfonate chemisorbed onto Pd(100) from two different concentrations.

1.7 Work Function Measurements

The work function is the energy required to remove a Fermi-level electron from the bulk to the vacuum just outside the surface.[36,37] In the bulk, the electron has an electrochemical potential $\bar{\mu}_e$ which is equal to the Fermi energy. Once the electron is removed from the crystal and is at rest, its electrochemical potential is simply the electrostatic potential energy $-e\Phi_0$, where Φ_0 is the electrostatic potential just outside the surface (where the electron does not feel its image charge). The work function is thus defined by Equation (10):

$$\phi = -e\Phi_0 - \bar{\mu}_e = -e(\Phi_0 - \Phi_i) - \mu_e \qquad (10)$$

It should be noted that the Fermi energy is not a property only of the bulk, because it contains the electrostatic potential Φ_i inside the metal, which is determined by the surface dipole layers.

In surface adsorption studies, only the surface dipole part $e(\Phi_0 - \Phi_i)$ changes, and consequently, the change in work function is equal to the change in this quantity; absolute work functions are less important. For a clean metal surface, the exponential decay of the wave function into the vacuum (electron ''overspill'') creates the surface double layer. The dipole has the negative end outward, and is dependent on the surface crystallography.

Work function measurements lead to useful correlations between $\Delta\phi$ and the nature of adsorption. For example, the adsorption of an electronegative atom such as oxygen changes the surface dipole and renders the outside of the crystal more negative, thereby increasing the work function. Conversely, an electropositive adsorbate such as Cs increases the work function. For atomic adsorbates, the sign of $\Delta\phi$ is therefore correlated with the direction of charge transfer: positive $\Delta\phi$ is associated with adsorbate-to-substrate charge transfer and negative $\Delta\phi$ with substrate-to-adsorbate charge transfer. In the case of molecular adsorbates, it must be realized that the dipole moment of the molecule is a more important contributor to $\Delta\phi$ than the surface chemical bond.

Experimental measurements of ϕ and $\Delta\phi$ have been based upon the diode method, field emission, contact potential difference, and the photoelectric effect; the latter two are more common. The photoelectric method, which measures absolute values of the work function, is based upon the determination of the threshold energy $h\nu_0$ for photoelectron ejection; the work function is then calculated from the equation $h\nu_0 = e\phi$. The contact potential difference method, which monitors changes in work function, depends on the measurement of the potential difference between two plates in electrical contact. If one of the plates is used as a reference of constant work function, $\Delta\phi$ at the other plate is manifested as a change in the contact potential. The most common way of measuring the change in contact potential uses a vibrating tip close to the surface as the reference plate (the Kelvin probe method). In its balanced condition there is an electric-field-free region between the sample and the tip, and no induced alternating current flows in the circuit. Adsorption leads to a momentary field and, hence, an alternating current. Electronic feedback is used to adjust the potential on the sample until the field-free condition again applies.

1.8 Temperature Programmed Desorption

Thermal desorption techniques[9–15] exploit the fact that species adsorbed on a surface will desorb at a rate which increases with temperature; the temperature dependence of the desorption rate yields data on desorption energies which, for most cases, yield information on adsorption binding energy states. Thermal desorption can also be used to obtain surface coverages and, in combination with mass spectrometry, determine the identity of species desorbed from the surface.

Temperature-induced desorption methods can be classified according to whether the rise in temperature is fast (flash desorption) or gradual (temperature programmed desorption (TPD)). In flash desorption, the desorption rate is much greater than the rate at which desorbed gas is pumped out of the system; in this case, the desorption of a given binding state is marked by a plateau in the pressure–temperature curve. In TPD, the slow heating (desorption) rate allows the evolved gases to be pumped out; as a result, the desorption of a particular binding state appears as a peak instead of a plateau in the desorption curve. The most common method of extracting the activation energy for desorption (E_d) from the peak temperature (T_p) involves the use of the so-called Redhead equations.[38] Other approaches exist that, while more rigorous, are also more difficult.[39] The desorption activation energy is related to the heat of adsorption ΔH_{ads} as shown in Equation (11):

$$E_d = E_a + \Delta H_{ads} \tag{11}$$

where E_a is the activation energy for adsorption. Several instances can be found in which the adsorption is nonactivated; for such cases, E_d is equal to ΔH_{ads}.

The use of a mass spectrometer to monitor species emitted from the surface simplifies TPD measurements. A comparatively inexpensive means for mass detection is afforded by a quadrupole mass analyzer, a schematic diagram of which is shown in Figure 13.[40] This mass filter consists of four parallel rod-shaped electrodes arranged at the apices of a diamond. The existence of an appropriately varying electrical field between the pairs of opposite electrodes will cause all ions to impact on the rods during transit except those of a particular mass-to-charge ratio m/z.

In TDMS, the substrate is positioned as close as possible to the mass spectrometer; provision must be made to minimize temperature gradients at the sample surface. The temperature is monitored by a thermocouple wire placed in direct contact with the crystal. Problems associated with degassing from parts of the sample

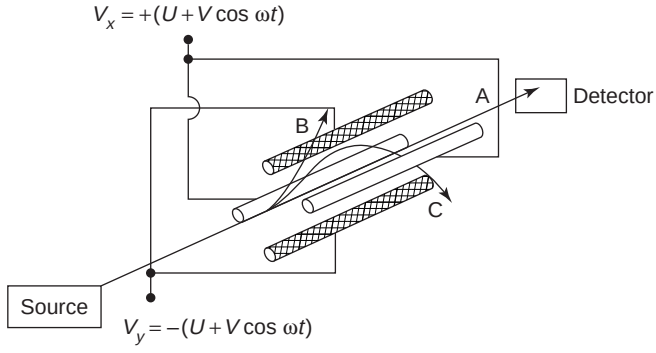

Figure 13 Schematic diagram of a quadrupole mass filter. Ions B and C have incorrect m/z ratio and are thrown against the rods; A has the proper m/z ratio and is transmitted through the rods onto the detector. (Reproduced by permission from Strobel and Heineman.[40])

manipulator close to the crystal can be solved by masking the mass spectrometer with a small aperture such that only line-of-sight detection is possible. Microprocessor control allows multiplexed data acquisition.

A thermal desorption mass spectrum is shown in Figure 14 for CO desorbed from a Pd(111)–c(4 × 2)–CO adlattice.

1.9 Instrument Designs

The single most critical step in UHV/EC experiments is the transfer of the electrode between the electrochemical cell (at ambient pressures) and the surface analysis chamber (under ultrahigh vacuum (UHV)). Ideally, the transfer is not accompanied by changes in surface structure and/or composition. The simplest approach, transfer through air is applicable only if the surface is inert (such as an oxide film) or is covered with a protective film (of solvent or electrolyte)[3] that can be removed by evacuation inside the UHV chamber. An alternative procedure that would allow electrode transfer in a controlled environment makes use of an inert-atmosphere glovebox attached to the surface-analysis instrument; electrochemical experiments are performed inside the box. However, regardless of the purity of the inert gas, the environment is not free of contaminants since the walls of the glovebox are replete with impurities that are slowly desorbed. The transfer-through-air and drybox methods have been used extensively in XPS and AES studies with oxided or chemically modified polycrystalline surfaces;[41,42] neither has been adopted for single-crystal work.

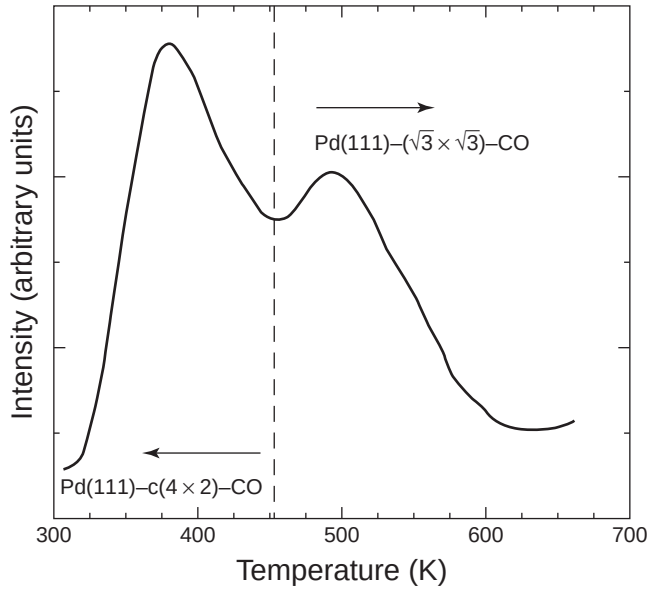

Figure 14 Thermal desorption mass spectrum of Pd(111)–c(4 × 2)–CO.

The approach employed most successfully in UHV/EC studies of single-crystal electrode surfaces involves the fabrication of a multi-technique surface-analysis apparatus to which an electrochemistry chamber is physically appended. The entire assembly is constructed of stainless steel and can be baked to about 200 °C in UHV to attain ultra-clean conditions. UHV is maintained by a combination of a titanium sublimation pump and either an ion or a turbomolecular pump. Transfer of the electrode between the analysis and electrochemistry chambers is accomplished by a sample manipulator-translator. In some instruments, the crystal remains attached to the same sample holder as it is moved between the two compartments; in other systems, the crystal is actually transferred between two different manipulators. A gate valve isolates the electrochemistry compartment, whenever necessary, from the rest of the system. It is preferable to keep the electrochemistry chamber under UHV when not in use in order to preserve its cleanliness. The electrochemical cell itself is located inside a bellows-enclosed compartment separated from the electrochemistry chamber by another gate valve; the cell is inserted only after the electrochemistry chamber is brought to ambient pressures with ultra-high purity inert gas. Based upon these considerations, various types of UHV/EC instruments have been constructed;[1,5,43–55] three of these are shown in Figures 15–17[1,52,55] for illustrative purposes.

In the UHV/EC system depicted in Figure 17,[55] an isolable, differentially pumped antechamber is situated between the UHV and electrochemistry compartments. The main function of this antechamber is to minimize the influx of solvent and/or electrolyte vapor into the surface analysis compartment. In this context, it is important to mention that the pressure in the electrochemistry chamber is usually an order of magnitude higher than in the UHV chamber; mass spectrometric analysis of the residual gas has revealed that the pressure difference arises primarily from higher amounts of water in the electrochemistry compartment. Because water is only weakly surface-active, it is generally not of major concern in UHV/EC studies. However, in the presence of comparatively high quantities of water, impurity species may be dislodged from the walls of the chamber and onto the sample surface. Similar ''knock-off'' effects can arise when the chamber is backfilled with high purity inert gas. Hence, it is critical to maximize the cleanliness of the electrochemistry chamber and its associated manifold; this can be accomplished by frequent bakeout and continuous evacuation of the electrochemistry chamber when it is not in use.

It is also important to ensure that the backfill gas is of the highest purity to minimize surface contamination by trace-level impurities; argon of at least 99.99% purity is usually employed. Background contamination is metal specific; for example, Cu is more sensitive to residual O_2

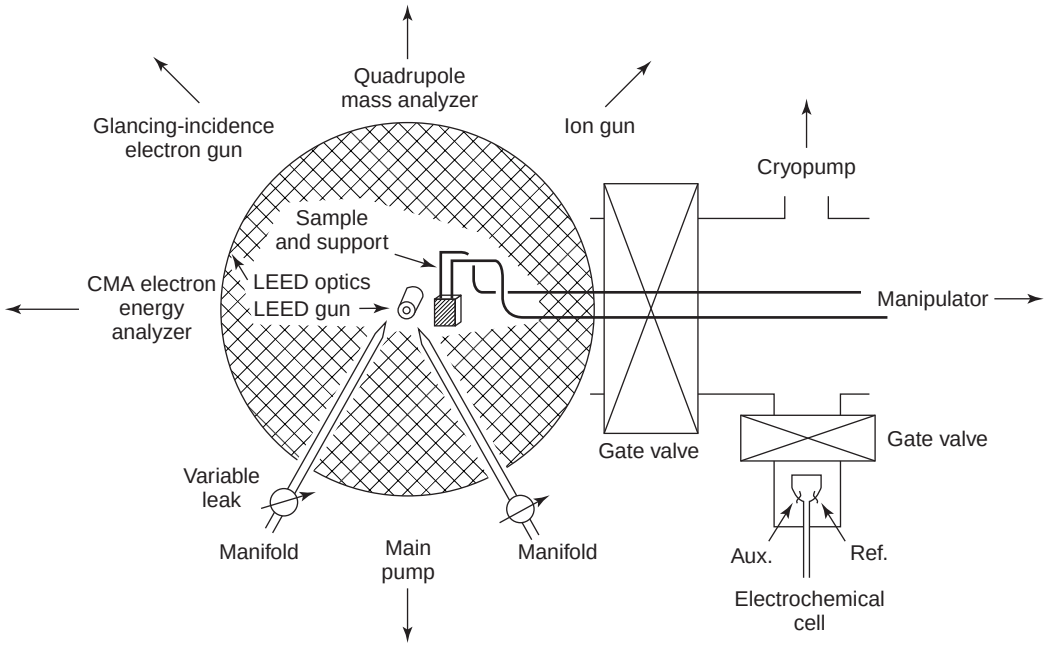

Figure 15 Schematic drawing of an experimental arrangement for UHV/EC studies. In this design, the crystal remains attached to only one sample manipulator. (Reproduced by permission from Hubbard et al.[1])

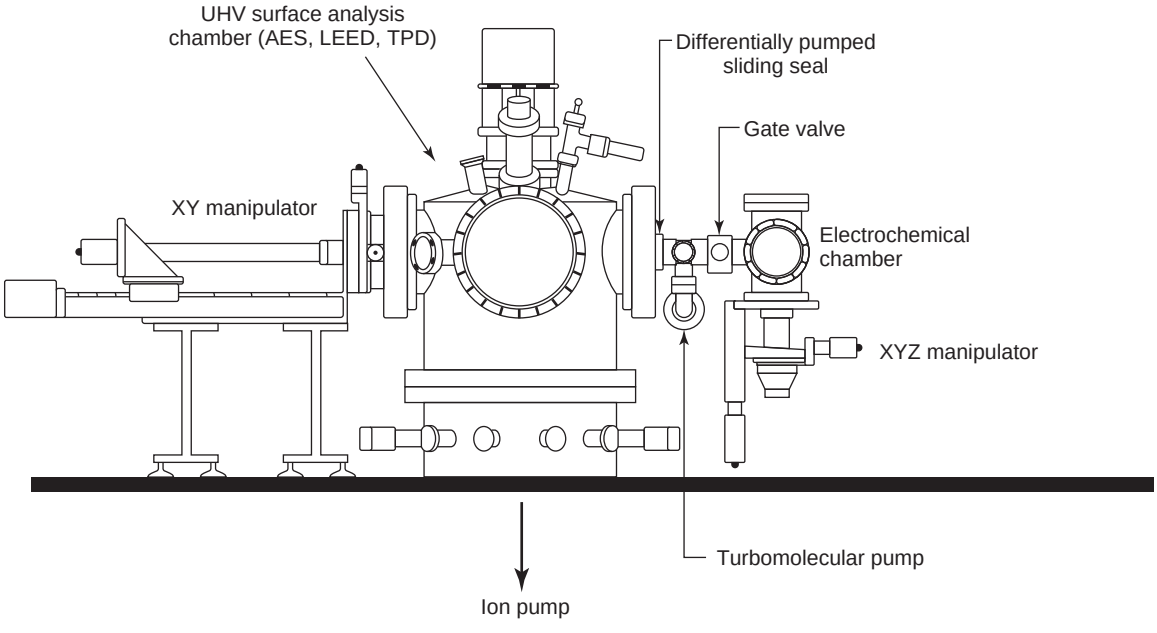

Figure 16 Schematic drawing of an experimental arrangement for UHV/EC studies. In this design, the electrode is transferred between different manipulators. (Reproduced by permission from Leung et al.[55])

while Pt is more susceptible to carbonaceous impurities. Hence, depending upon the nature of the investigation, it may be necessary to pass the high-purity inert gas through molecular scavengers (e.g. a Ti sponge heated to 900 °C) for still further purification. It must also be realized that electrode-surface contamination can also result from trace-level impurities in the electrolyte solution. Such impurities can originate from the solvent, electrolyte, glassware, and/or the inert gas employed for solution deaeration. The level of solution-based impurities can be minimized by the use of highly purified chemical reagents and gases; in the case of aqueous solutions, the utilization

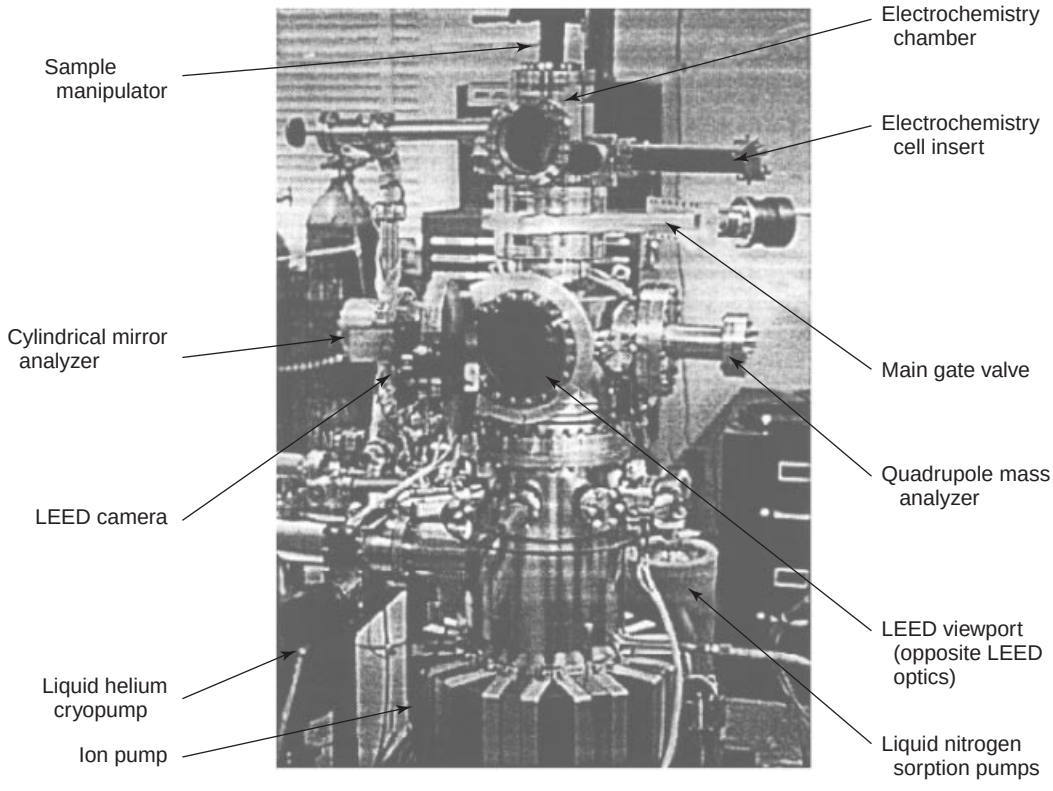

Figure 17 Photograph of a UHV/EC apparatus. In this design, the sample electrode remains attached to only one sample manipulator.

of pyrolytically triply distilled water is recommended,[56] although the use of water eluted through (commercial) multiple-filtration stages is now an acceptable alternative.

A typical UHV/EC experiment following instrument ''bakeout'' (at which point the base pressure should be less than 5×10^{-10} torr) would include the following steps. After surface preparation and characterization, the electrode is transferred into the electrochemistry chamber (cf. Figure 17) which, by closure of the main gate valve, is isolated from the UHV system and then backfilled with high-purity inert gas. The external gate valve is opened and the cell is inserted into the electrochemistry chamber. After completion of the electrochemical experiments, the cell is retracted, the external gate valve closed, and the chamber evacuated by a liquid helium cryogenic or turbomolecular pump to less than 10^{-6} torr. At this point, the main gate valve can be opened to complete the evacuation of the electrochemistry chamber and permit the transfer of the electrode into the analysis compartment. Pumpdown from ambient pressure to 10^{-8} torr vacuum is usually achieved in less than 15 min.

Electrochemistry experiments have been performed with cells in either the standard or thin-layer arrangement. The latter significantly reduces the level of surface contamination from solution-borne impurities. If the

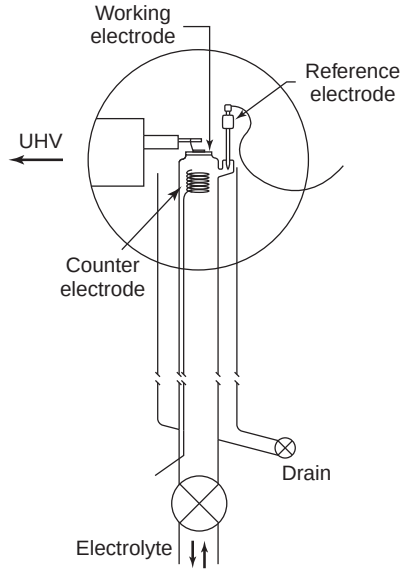

Figure 18 Schematic view of an electrochemical cell for use with a single-crystal disk electrode. (Reproduced by permission from Leung et al.[55])

entire electrode is to be immersed in solution, all faces of the single-crystal should be oriented identically in order to obtain voltammetry characteristic of only one

crystallographic face. As an alternative, the electrode can be positioned on top of the electrochemistry cell in such a way that only one crystal face is exposed to solution (Figure 18).[47] Such a configuration, however, often results in the adherence of a droplet of electrolyte on the crystal surface when the electrode is withdrawn from the solution; this problem does not arise if the electrode is withdrawn slowly ($1\,mm\,s^{-1}$) in the vertical position, as the droplet would then form at the bottom of the crystal.[47]

2 FUNDAMENTAL ASPECTS

2.1 The Emersion Process

It is important to determine the changes at the electrode–electrolyte interface when the electrode is *emersed* (removed at a given potential) from the electrolyte solution. In the ideal process, the emersed electrode retains an interfacial layer identical in composition and structure to that when the electrode was still in solution. The electrode–solution interface, or the electrochemical double layer, is a structured assembly of solvent, electrolyte, and reactant. In the traditional view,[57–60] this ensemble, nominally 1 nm in thickness, is subdivided into an inner (compact) layer that consists of field-oriented adsorbed solvent molecules and *specifically* adsorbed anions, and an outer layer composed of solvated cations. The locus of the centers of the adsorbed anions delineates the so-called inner Helmholtz plane (IHP), whereas the line of centers of the nearest solvated cations defines the outer Helmholtz plane (OHP). Charge transfer reactions are thought to occur at this outer (reaction) plane. The solvated ions interact with the charge metal only through long-range electrostatic forces and, because of thermal agitation in the solution, are distributed in a three-dimensional region that extends from the OHP into the bulk of the solution. This region is identified as the *diffuse layer*, and its thickness is a function of electrolyte concentration; it is less than 30 nm for concentrations greater than 10^{-2} M.

Clearly, the electrode-withdrawal process involves a delicate balance with respect to the thickness of the emersion layer: it must be sufficiently thick to incorporate the intact electrochemical double layer but it should also be thin enough to exclude *residual* (bulk) electrolyte. Numerous studies have helped establish the fact that the electrochemical double layer can, under appropriate electrolyte concentrations, be retained intact when the electrode is withdrawn from solution under potential control. The optimum concentration depends upon whether emersion is hydrophobic or hydrophilic.[61–64] For the latter type, the concentration must not be much *higher* than 10^{-3} M if inclusion of bulk electrolyte is to be circumvented. For hydrophobic emersion, the concentration must not be much *lower* than 10^{-3} M if double-layer discharge is to be avoided. In cases where the mode of emersion is not known, an electrolyte concentration of 10^{-3} M appears to be a logical choice.

Investigations of hydrophobic emersion based upon electrode resistance measurements,[64] electroreflectance spectroscopy,[61,62,65] XPS,[66] and work function changes[36,37,61,63] have been able to: (i) demonstrate the existence of an emersed double layer; (ii) determine its stability; and (iii) monitor changes in its structure and composition brought about by the emersion process. The evidence has been compelling that the structure and composition of the double layer in the emersed phase are very similar, if not identical, to those in the solution state; that is, only a little or no double-layer discharge occurs upon emersion. Later studies focused on the effect of the emersion process on the structure of adsorbed molecular species;[67–71] the data obtained demonstrate that the structure and orientation of molecular adsorbates are essentially unperturbed by the emersion process.

2.2 Perturbations Caused by Evacuation and Surface Analysis

Another important issue in UHV/EC centers around the perturbations caused by the evacuation and subsequent surface analytical processes. Alterations in the surface electronic structure can be studied by work function change measurements; representative results are shown in Figure 19, in which a plot of the work function of a polycrystalline Au emersed from 0.1 M $HClO_4$ into UHV as a function of the emersion potential is presented.[37,71] It can be seen here that the work function tracks the applied potential over a wide range, even into the oxide formation region. This, and other sets of pertinent data, demonstrates that the electronic properties of the electrochemical double layer are unaffected by emersion into either the ambient or UHV.

An expected effect of evacuation are changes in composition due to UHV-induced desorption; the extent will depend upon the heats of vaporization ΔH_{vap} or sublimation ΔH_{sub} of the *unbound* materials (excess water, unadsorbed gases, liquids, and sublimable solids) entrapped within the emersed layer. Water retained as part of the hydration sphere can survive the evacuation process if the hydration enthalpies ΔH_{hyd} are substantial.[72,73] Strongly chemisorbed species, such as iodine at the noble-metal electrodes, are expected to form stable well-ordered adlattices in solution[74–76] that would not reconstruct in vacuum. Similarly, the coverages and structures of irreversibly adsorbed molecules are expected to withstand the evacuation process. One

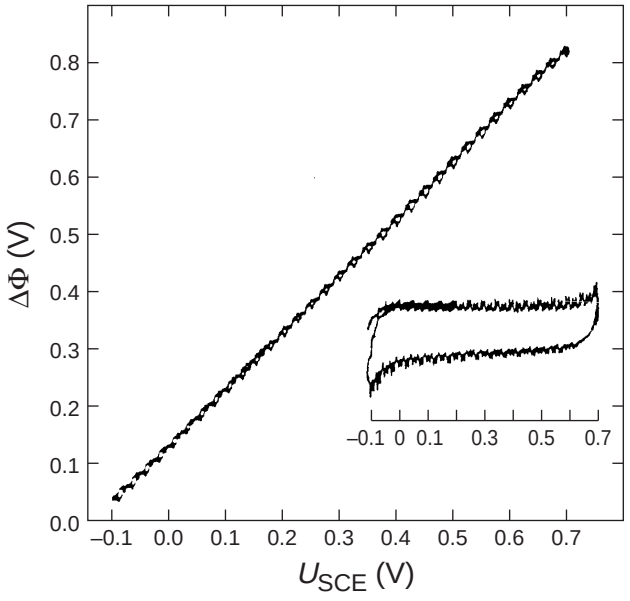

Figure 19 A plot of the work function of a polycrystalline Au emersed from 0.1 M $HClO_4$ into UHV as a function of the emersion potential. The work function of the clean metal was 5.2 eV. The lower and upper lines, respectively, represent the solution inner potential if the absolute normal hydrogen electrode (NHE) potential is 4.45 or 4.85 V. (Reproduced by permission from Hansen and Hansen.[37])

example is provided by 3-pyridylhydroquinone (Py-H_2Q) which, on Pt(111), is chemisorbed through the N-heteroatom.[77] In such a mode of surface attachment, the diphenol group is pendant and able to undergo reversible quinone/hydroquinone redox reaction. It has been shown that the redox current–potential curves for chemisorbed Py-H_2Q, before and after a 1-h exposure to UHV, are identical (Figure 20).[77]

Perturbations can also arise from the surface analysis itself. For example, implicit in TPD is the requirement for complete desorption; hence, TPD and TDMS are inherently destructive techniques. On the other hand, analytical methods based upon electron and optical spectroscopies are not intended to damage the surface layer. Unless exceedingly high photon fluxes are used, optical methods are nondeleterious when compared with particle-based techniques. Several surface processes are known to be stimulated by electron impact. Examples are binding-site conversions, dissociative chemisorption, and particle desorption.[12] Such processes take place even at minimal sample heating (low electron power densities), which indicate that surface thermal effects are insignificant. It is now accepted that electron-stimulated reactions occur mainly via electronic excitations. These excitations can lead to bond dissociation and form the basis of the surface spectroscopic technique known as electron stimulated desorption ion angular distribution (ESDIAD).[78]

Pendant functional groups not directly bonded to the substrate surface, such as the diphenol moiety in Py-H_2Q, are most prone to electron-stimulated desorption. In other instances, electron irradiation can induce surface displacement reactions that involve species present as residual gas in the analysis chamber. It has been reported, for example, that prolonged electron-beam irradiation during LEED and/or AES in the presence

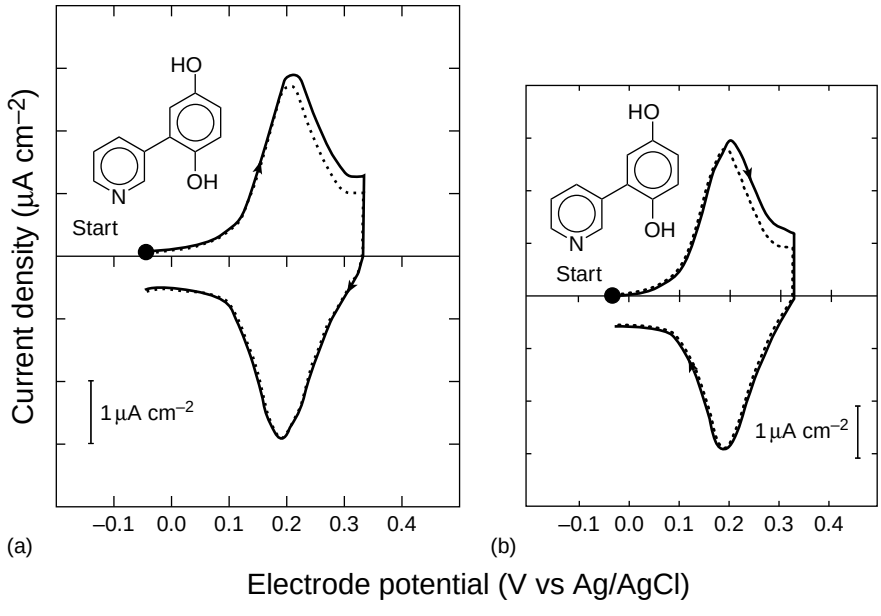

Figure 20 Cyclic voltammetry of Py-H_2Q at Pt(111). (a) Solid curve: first scan; dotted curve: after 1 h in UHV. (b) Solid curve: first scan; dotted curve: second scan. (Reproduced by permission from Stern et al.[77])

of residual water vapor caused a $Ba(CN)_2$ adlayer to undergo a reaction that formed gaseous HCN and solid $Ba(OH)_2$.[72,73]

Even if changes in the emersed double layer are induced by the surface analysis, it must be realized that such alterations would be deleterious only if the post-analysis layers are to be used for further experiments. In those instances when additional experiments have to be performed, it is a simple matter to regenerate the surface to exactly the point just prior the surface analysis. If desired, beam damage can be assessed by repeated analysis over a period of time followed by extrapolation of the data to *zero* time.

3 REPRESENTATIVE STUDIES

UHV/EC investigations with single-crystal electrode surfaces can be broadly classified into three groups. The first is focused on the structure and constitution of the electrochemical double layer as functions of electrode potential and solution composition. The second centers on electrodeposition and dissolution processes; included in this category are extensive studies on hydrogen and oxygen adsorption/desorption. The third deals with the interfacial structure and reactivity of chemisorbed molecules.

3.1 Electrochemical Double Layer

Two general strategies have been adopted in UHV/EC studies of the electrical double layer. One, strictly a model approach, involves the synthesis of the double layer in UHV by sequential cryogenic adsorption of its constituents;[79–85] the temperature must be maintained below 160 K at all times in order to prevent the evaporation of unbound solvent. The other approach is based upon the structural and compositional analysis of the emersed layer; since surface characterization is done at ambient temperatures, excess water in the diffuse layer is pumped away.

The viability of the cryogenic coadsorption approach[79–82] was first demonstrated by comparison of work-function changes $\Delta\varphi$ for UHV-synthesized and actual $Ag(110)-X-H_2O$ layers, where X denotes Cl^- or Br^-. The results showed that, as long as *coadsorbed* water is present in the cryogenic layer, good agreement exists between the ex situ and in situ results. The satisfactory agreement indicates that, at least under zero diffuse-layer-charge conditions, information from the UHV simulation work may have relevance to actual electrochemical systems. The requirement of solvation implies that the electronic properties of the unsolvated $Ag(110)-Br$ interface are not identical to those of the fully solvated $Ag(110)-Br-H_2O$ layer. Other more complex interfacial systems have also been modeled via cryogenic coadsorption. For example, UHV-synthesized $H_2O-HF-CO$ coadsorbed layers were studied at Pt(111) and Rh(111) surfaces by HREELS, LEED, TPD and XPS.[83,84] In such work, the ''control'' of electrode potential was based upon the amount of coadsorbed H_2. The cryogenic coadsorption approach offers two main advantages: (i) the control of interfacial parameters far more precisely than can be achieved in solution; and (ii) the detailed characterization of fully solvated species by a host of surface-sensitive spectroscopic methods. The approach, however, provides little beyond structural models for the interfacial layers; its usefulness as a surface analytical technique is unclear.

The direct approach to the study of the electrochemical double layer involves the characterization of the electrolyte layer retained at the emersed surface. This approach is applicable only to cases in which the compact layer consists of materials that remain on the surface even after evacuation to UHV. For example, only limited information can be expected when emersion is from aqueous HF since both H_2O and HF will be completely desorbed, at ambient temperatures, in vacuum. It is also implied that the molecules pumped away are inconsequential in the formation and preservation of the electrochemical double layer. This is not an unreasonable premise since the double layer composed of chemisorbed species is governed by strong chemical bonds, and should only be minimally perturbed by physisorbed species.

UHV/EC studies of anions specifically adsorbed from aqueous solutions have been carried out at well-defined Pt(111),[86–92] Pt(100),[93,94] stepped Pt(s)[6(111) × (111)],[95] Cu(111),[96] Ag(111),[97] Au(111),[98] and Pd(111)[99] electrodes. The anions studied include the halides,[86–90,93–99] SH^-,[89,99] CN^-,[91,92] SCN^-,[100] and SO_4^{2-}.[96] All of these anions yield surface coverages and well-ordered structures that depend upon the solution pH and the applied potential.

3.2 Underpotential Electrodeposition and Anodic Dissolution

The deposition of *monolayer* quantities of one metal onto another generally occurs at potentials positive of that for *bulk* deposition because of preferential interactions between the substrate and the foreign-metal electrodeposit. The underpotential deposition (UPD) process is strongly influenced by the structure and composition of the substrate; hence, UPD research is an area in which UHV/EC methods have been widely adopted. The published literature on UHV/EC studies of UPD can be categorized according to whether the experiments were used to correlate substrate structure with

the electrodeposition voltammograms, or to determine the interfacial properties of the adatom-modified substrate. Investigations devoted to structure–voltammetry correlations help establish reference states against which new experiments can be calibrated; those focused on post-deposition characterization yield important information concerning the electrocatalytic selectivity of the mixed-metal interfaces.

The first applications of LEED and AES in electrochemistry involved the correlation of the surface crystallographic orientation with the underpotential hydrogen deposition at Pt electrodes.[1,101–105] Those studies were motivated by earlier work with polycrystalline Pt electrodes whose cyclic voltammograms showed two hydrogen deposition peaks; the peaks were identified simply as weakly and strongly bound states of adsorbed hydrogen. Subsequent studies with single-crystalline but *uncharacterized* Pt(111), Pt(110), and Pt(100) electrodes associated the weakly bound hydrogen peak with the Pt(111) electrode, and the strongly bound hydrogen with the Pt(100) electrode. These studies, though not definitive enough due to the existence of atomically disordered surfaces because of multiple oxidation–reduction cycles, provided the impetus for further UHV/EC studies.

Later work based upon flame-annealed single-crystal surfaces[3] led to the discovery of new voltammetric features for Pt(111) in the form of highly reversible pseudocapacitance peaks at potentials *well positive* of the usual hydrogen deposition peaks.[106–108] UHV/EC studies that employed apparatus equipped with improved UHV-to-EC transfer technology were able to reproduce the new voltammetric results. Extensive follow-up work then ensued that clarified several aspects of this surface-sensitive reaction.[101,107,108]

While studies of hydrogen chemisorption at Pt single-crystal electrode surfaces have been extensive, investigations on the formation of underpotential states of oxygen have not been as numerous. Electrochemical experiments have been performed only with Pt(100) and Pt(111) electrodes.[109–111] Gas-phase and solution-state reactions with oxygenous species have been carried out at stainless-steel single crystals.[112–114] The occurrence of place-exchange during anodic film formation has been studied via LEED spot-profile analysis.[111] This irreversible place-exchange reaction accounts for the observation that the electrode surface loses its single crystallinity even after only minimal anodic oxidation.

The literature on monolayer metal deposits is extensive. Most of the work pertains to the geometric, electronic, and catalytic properties of foreign metals *vapor deposited in UHV* onto single-crystal substrates; a compilation of the adlattice structures of such metal adlattices has been published.[115] Studies of foreign metal monolayers *deposited electrochemically* have been primarily with polycrystalline substrates. The first UHV-based investigation of electrodeposited admetals employed XPS to determine the core-level shifts of submonolayer Cu and Ag on *polycrystalline* Pt.[43,116] Quite a few investigations have been reported of admetal deposits on single-crystal surfaces but only a minority of these involve UHV/EC technology.

The first UHV/EC work on electrodeposition at well-defined electrode surfaces involved Ag at an iodine-coated Pt(111) electrode.[117,118] The iodine pretreatment was done in UHV to form a protective Pt(111)($\sqrt{7} \times \sqrt{7}$)R19.1°–I adlattice before immersion into a solution containing dilute Ag^+ in 1 M $HClO_4$. Subsequent studies included Ag electrodeposition on I-coated Pt(100)[119] and stepped Pt(s)[6(111) × (111)],[95] Cu on I-pretreated Pt(111),[120] and Pb on I-covered Pt(111).[121] Sn[122] and Pb[123] deposition onto *iodine-free* Pt(111) in Br^- or Cl^- solutions[124] has also been studied. Although the Pt substrate was not pretreated with I, the presence of halide ions in the plating solution led to specific adsorption of anions prior to the deposition process.

Electrodeposition from solutions free of surface-active anions has been studied. These investigations, carried out in ClO_4^- or F^- electrolyte, included UPD of Cu on Pt(111),[125–127] Tl, Pb, Bi, and Cu on Ag(111).[128] Invariably, the underpotentially deposited films showed unique adlattice geometries that were dependent upon the substrate orientation and the admetal coverage.

The atomic layer epitaxy (ALE) approach to deposition of a compound film, based upon the alternate layer-by-layer deposition of the elements of the compound, has been adopted in the electrochemical synthesis of compound semiconductors. This electrochemical analog, referred to as electrochemical atomic layer epitaxy (ECALE).[129,130] takes advantage of the fact that only monolayer quantities are produced by UPD. For Example, the UPD-based epitaxial growth of CdTe on Au(111) has been monitored by LEED and AES.[129]

While UHV/EC-based investigations on electrodeposition abound, little has been published on the reverse reaction, anodic dissolution; most of the studies reported have made use of electrochemical scanning tunneling microscopy (ECSTM).[131] One system, that of the adsorbate-catalyzed corrosion of Pd, has been extensively studied using both UHV/EC and ECSTM.[132,133] In the absence of a layer of chemisorbed iodine, anodic oxidation of Pd in sulfuric acid yields only a passivating metal oxide film; when iodine is chemisorbed on the surface, dissolution of Pd occurs without alteration of the structure or composition of the iodine adlattice (Figure 20).[132] It

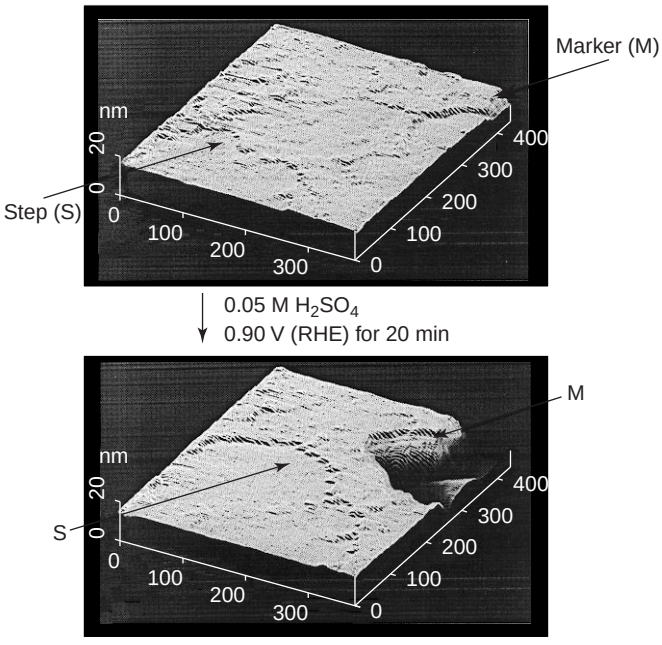

Figure 21 Wide-area ECSTM images of a Pd(111)–$(\sqrt{3} \times \sqrt{3})R30°$–I facet on a single-crystal bead at low-current adsorbate-catalyzed dissolution. The potential of the tip was held at 0.9 V; the tunneling current was 1 nA.

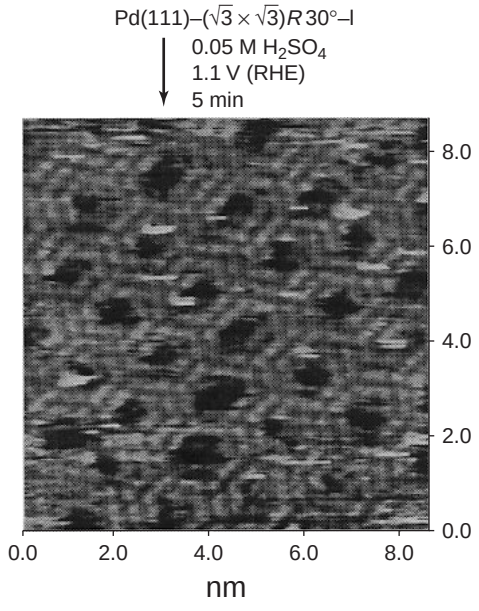

Figure 22 Atomic-resolution ECSTM image of a Pd(111)–$(\sqrt{3} \times \sqrt{3})R30°$–I facet on a single-crystal bead after medium-current anodic dissolution.

has been found that, at low dissolution rates, the reaction proceeds at the step-edges (Figure 21) at much higher currents, I–Pd place-exchange occurs that leads to the formation of pits (Figure 22).[133]

3.3 Molecular Adsorption[134]

The capability to prepare single-crystal surfaces by thermal treatment at ambient pressures[3] has fostered the proliferation of non-UHV studies of the adsorption of molecules at monocrystalline electrodes. The detail of information obtained from such in situ work, however, falls short of that provided by UHV/EC experiments. As one example, although in situ IRAS has provided much information about the structure-sensitivity of the chemisorption and anodic oxidation of CO, its sensitivity is too low to permit meaningful investigations with other molecules as simple as ethylene.

3.4 Solvent–Electrode Interactions

The nature of the interactions between the solvent and the electrode surface has significant ramifications in electrochemical surface science. For instance, the use of strongly surface-active solvents would severely repress electrocatalytic processes that rely on a direct interaction between the reactant and the metal surface. The bonding of water to metal surfaces is an important issue in aqueous electrochemistry. In models suggested to explain the potential dependence of double-layer capacity, the existence has been postulated of monomeric and clustered water molecules, both of which are able to adopt two opposite dipolar orientations.[135,136] The studies of water adsorption on single-crystal electrodes are all based upon vapor deposition in UHV, usually at cryogenic temperatures since water is not adsorbed on clean metal surfaces at ambient temperatures. Of significant interest to electrochemistry is the observation that, on Ni, Pt, Ag, Cu, and Pd, water is dissociatively chemisorbed if the surface contains *submonolayer* coverages of oxygen.[137] The reaction is thought to occur by hydrogen abstraction. This reaction is metal-specific since at other noble metals such as Ru(001), adsorbed oxygen is inactive towards water dissociation.

Nonaqueous solvents commonly used in electrochemistry include acetonitrile, dimethylformamide, *p*-dioxane, sulfolane, dimethylsulfoxide, pyridine, acetic acid, propylene carbonate, liquid ammonia, and dichloromethane.[138] Work on such materials can be categorized according to how the electrode is allowed to interact with the nonaqueous solvent: by vapor dosing in vacuum, or by exposure to aqueous solutions containing small quantities of nonaqueous-solvent material. Studies under the latter category are more abundant, but those are more accurately classified under electrode–solute, not electrode–solvent, interactions. Except for one case, all UHV-based studies with nonaqueous-solvent compounds were carried out purely in the context of gas–solid interactions.[139] The intent of the one exception[140] was

to use the reactions between the solvent vapor and the metal surface as models for the electrochemical analogs; for better simulation of solution conditions, vapor dosing was up to 0.3 torr, approaching the vapor pressures of the liquid solvents. UHV/EC work in which the electrode is immersed in pure nonaqueous solvent has not been pursued.

3.5 Group IB Electrodes

Most organic compounds are only weakly adsorbed on Cu, Ag, and Au electrode surfaces; hence, unless the adsorbate itself is a solid or when adsorption is carried out at cryogenic temperatures, meaningful UHV/EC experiments with the coinage metals are limited. One study, which took advantage of the strong interaction of the -SH functional group with the coinage metals, used HREELS, LEED, AES, and voltammetry to determine the influence of the location of the N heteroatom on the adsorption properties of the isomers 2-mercaptopyridine and 4-mercaptopyridine at Ag(111) in aqueous HF.[141] The subject compounds were thought to undergo isomerization upon oxidative adsorption through the -SH moiety.

3.6 Group VIII Electrodes

The abundance of studies of organic molecular adsorption at electrode surfaces involves the platinum metals. This is not surprising since these metals are well known for their electrocatalytic activities and an immense body of work has already been amassed for these materials in their polycrystalline states.[142–145] Surface electrochemical studies of metal-organic compounds at single-crystal electrodes can be broadly classified according to whether the work was done with CO (and related small molecules) or with more complex molecules. The former are more numerous, although a vast majority of the studies have been carried out without UHV-based surface characterization. Work with well-defined surfaces have been limited to LEED of CO adlattices on Pt(111)[146] and Pd(111),[147,148] and HREELS, LEED, TPD and XPS of mixed H_2O–HF–CO layers generated in UHV by cryogenic adsorption at Pt(111) and Rh(111) surfaces.[83,84]

An impressive amount of detailed information on a wide variety of complex organic compounds chemisorbed at well-defined Pt(111) and Pt(100) electrode surfaces has been furnished by LEED, AES, TPD, and HREELS.[134] Electrocatalytic reactivity studies which accompanied these investigations were limited to anodic oxidation reactions; only correlations between the mode of adsorbate bonding and the extent of anodic oxidation were attempted.

3.7 Carbon Monoxide

Much of what is known about the structure and reactivity of CO chemisorbed at single-crystal electrodes, and their dependencies on surface crystallographic orientation, electrode potential, and adsorbate coverage are based almost entirely upon in situ IRAS measurements.[134] Two UHV/EC studies on CO are noteworthy. One made use of a well-defined Pt(111) surface and sought to correlate anodic peak potentials with observed LEED structures.[146] The other, based upon LEED, AES, TPD, voltammetry, and coulometry, examined the chemisorption of CO at well-defined and anodically disordered Pd(111).[147] In this study, it was shown that CO adsorption from solution onto a UHV-prepared Pd(111) surface yielded an ordered adlattice, Pd(111)–c(4 × 2)–CO, in which the CO molecules occupy two-fold hollow sites. Two CO-to-CO_2 anodic oxidation peaks were observed: the first peak is partial oxidation of the c(4 × 2) layer followed by an *adlattice reconstruction* to form a Pd(111)($\sqrt{3} \times \sqrt{3}$)$R30^\circ$–CO structure; the second peak is due to complete oxidative desorption of the Pd(111)($\sqrt{3} \times \sqrt{3}$)$R30^\circ$–CO adlayer to yield a clean and well-ordered Pd(111) single-crystal surface.[148] At the surface disordered by extensive anodic oxidation, chemisorption of CO occurred spontaneously but no ordered CO adlayers were produced; on the disordered surface, the CO molecules were thought to reside on atop sites.

3.8 Other Organic Compounds

UHV/EC investigations with organic compounds more complex than CO have been focused on the nature of the adsorbate–substrate chemical interactions as a function of interfacial parameters, and on the influence of the mode of attachment on the reversible and/or catalytic electrochemistry of the adsorbed species. For example, the differences between gas-phase and solution-state chemisorption and catalytic hydrogenation of ethylene have been documented:[149–152] variations in the structures of ethylene chemisorbed at the solid–solution and gas–solid interfaces lead to different reaction pathways. In solution, ethylene chemisorption occurs molecularly through its π-electron system, whereas chemisorption in UHV is accompanied by molecular rearrangements to form a surface ethylidyne species. In electrocatalytic hydrogenation, ethylene is reduced on the Pt surface by adsorbed H atoms; in gas-phase hydrogenation, H atoms must be transferred from the Pt surface through a layer of irreversibly adsorbed ethylidyne to ethylene adsorbed on top of the ethylidyne layer. Further work with alkenes[153–157] has been centered on the effects of hydrocarbon chain length and the presence of weakly surface-active substituents such as

carboxylates[158] and alcohols.[159] These studies showed that: (i) the primary mode of surface coordination of terminal alkenes, alkenols, and alkenoic acids is through the olefinic double bond; and (ii) the pendant alkyl chain is always extended outward on top of the propylene moiety. From coulometric measurements, it was concluded that anodic oxidation of the chemisorbed alkenes is limited to the anchor group, unless the pendant moieties are too close in close proximity to the electrode surface.

Early studies with smooth polycrystalline Pt indicated that aromatic compounds such as 1,4-dihydroxybenzene are chemisorbed in discrete, nonrandom orientations that depend upon various interfacial parameters.[160–162] Subsequent UHV/EC experiments with well-defined Pt(111) electrodes[163,164] supported the earlier findings, although the conditions at which the multiple orientational transitions occur are different for the polycrystalline and single-crystal electrodes. The electrocatalytic oxidation of multiply oriented aromatic molecules has been shown to be strongly dependent on their *initial* adsorbed orientations.

Sulfur-containing compounds investigated[165,166] include thiophenol, pentafluorothiophenol, 2,3,5,6-tetrafluorothiophenol, 2,3,4,5-tetrafluorothiophenol, 2,5-dihydroxythiophenol, 2,5-dihydroxy-4-methylbenzyl mercaptan, and benzyl mercaptan; chemisorption of these compounds occurs oxidatively through the sulfur group with loss of the sulfhydryl hydrogen. The tethered diphenolic moieties in the adsorbed dihydroxythiophenols show reversible quinone/diphenol redox chemistry. The S-heterocyclic compounds studied were thiophene, bithiophene, and their carboxylate and methyl derivatives. Experimental evidence indicates that these compounds are bound exclusively through the S heteroatom, although the chemisorption process may be accompanied by self-desulfurization. The electropolymerization of 3-methylthiophene at clean Pt(111) and monomer-treated Pt(111) pretreated has been studied, and the properties of the two types of polymer film were compared. In terms of the HREELS spectra, two major differences were noted which were attributed to changes in the physical nature of the polymer film.

The chemisorption of pyridine, bipyridine, multinitrogen heteroaromatic compounds, and their derivatives[165–168] has been examined as a function of isomerism and substituents. Pyridine forms a well-ordered layer of admolecules chemisorbed through the N heteroatom in a tilted vertical orientation. The derivatives are coordinated similarly unless the ring nitrogen is sterically hindered such as in 2,6-dimethylpyridine where chemisorption is in the flat orientation. Pyrazine, pyrimidine, and pyridazine are chemisorbed through only *one* nitrogen heteroatom in a tilted-vertical orientation. For the derivatives, adsorption occurs through the least hindered ring nitrogen. Carboxylate substituents located in positions *ortho* or *meta* to the nitrogen heteroatom interact with the Pt(111) surface at positive potentials. The chemisorbed layers were disordered as indicated by the absence of LEED patterns and were observed to be electrochemically unreactive. The adsorption behavior of the bipyridyls was found to be sensitive to steric hindrance at the

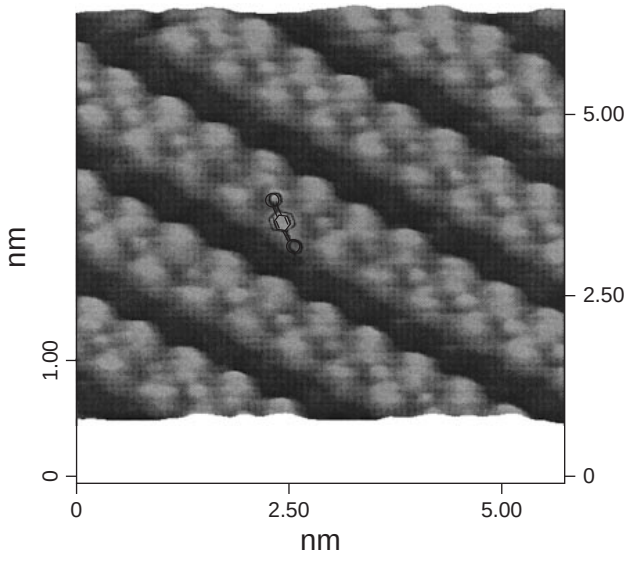

Figure 23 High-resolution ECSTM image of Pd(111) immersed in a 0.01 mM aqueous solution of hydroquinone in 0.05 M H_2SO_4.

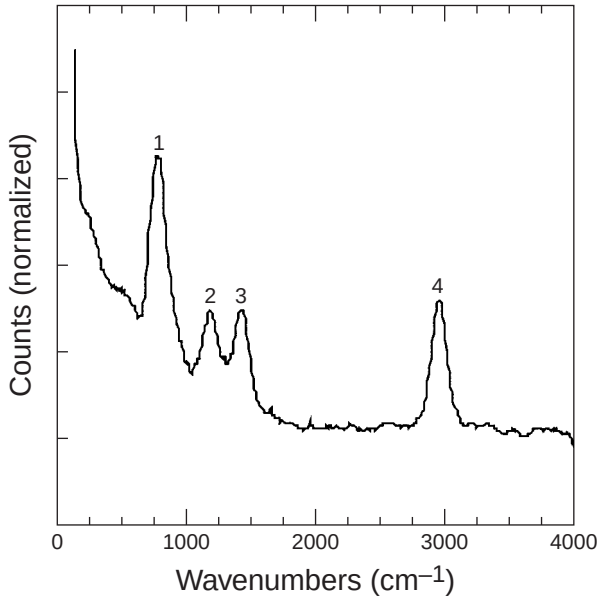

Figure 24 HREEL spectrum of a Pd(111) surface after emersion from a 0.1 mM aqueous solution of benzoquinone in 1 mM tetrafluoroacetic acid. Peaks 2, 3 and 4 correspond to *in-plane* C–H bending, C–C and C–H stretch modes, respectively; peak 1 is due to an *out-of-plane* C–H bending mode.[170]

positions *ortho* to the nitrogen heteroatom. Studies on the mode of chemisorption at well-defined Pt(111) of L-dopa, L-tyrosine, L-cysteine, L-phenylalanine, alanine, and dopamine have been reported.[167,168]

The chemisorption of aromatic molecules on well-defined Pd(*hkl*) surfaces has recently been studied by a combination of UHV/EC and ECSTM.[169,170] Figure 23 shows a high-resolution STM image of a Pd(111) electrode immersed in a dilute solution of 1,4-dihydroxybenzene; Figure 24 shows the corresponding HREEL spectrum. Analysis of both sets of data indicate that hydroquinone is oxidatively chemisorbed as benzoquinone and, although bound in the flat orientation, is slightly tilted.[170]

4 EPILOGUE

The aim of modern electrochemical surface science is the design of superior electrode materials not only for the characterization but also for the control of important electron-transfer reactions. This goal can best be accomplished by the establishment of fundamental correlations between the structure, composition and chemical reactivity at the electrode–solution interface. As should be evident from the above review and the cited literature, the UHV/EC approach provides an arsenal of surface-sensitive techniques that help in the establishment of such fundamental correlations.

The UHV/EC approach offers three essential functions: (i) it can be used to tailor surfaces of well-defined structure and composition for the performance of specific electrochemical tasks; (ii) it can be employed to determine (or model via cryogenic deposition of the constituent species) the structure and composition of the compact layer; (iii) it can be exploited to interrogate the composition and lateral structure of nonspecifically absorbed materials in the diffuse layer. These tasks will always be paramount in fundamental electron-transfer investigations, whether those are directed towards biotechnology, electroanalysis, fuel-cell technology, or materials research. That UHV/EC-based results have been supported and complemented by in situ experiments essentially assures the continued utilization of this approach in future studies centered around the electrode–electrolyte interface.

ACKNOWLEDGMENTS

The author wishes to thank the National Science Foundation and the Robert A. Welch Foundation for their support of work cited in this review.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
ALE	Atomic Layer Epitaxy
CHA	Cylindrical Hemisphere Analyzer
CMA	Cylindrical Mirror Analyzer
ECALE	Electrochemical Atomic Layer Epitaxy
ECSTM	Electrochemical Scanning Tunneling Microscopy
ESCA	Electron Spectroscopy for Chemical Analysis
ESDIAD	Electron Stimulated Desorption Ion Angular Distribution
HREELS	High-resolution Electron Energy Loss Spectroscopy
IHP	Inner Helmholtz Plane
IRAS	Infrared Reflection–Absorption Spectroscopy
LEED	Low-energy Electron Diffraction
OHP	Outer Helmholtz Plane
Py-H_2Q	3-Pyridylhydroquinone
RFA	Retarding Field Analyzer
RHEED	Reflection High-energy Electron Diffraction
TDMS	Thermal Desorption Mass Spectrometry
TPD	Temperature Programmed Desorption
UHV	Ultrahigh Vacuum
UHV/EC	Ultrahigh Vacuum Electrochemistry
UPD	Underpotential Deposition
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

APPENDIX 1

Ultrahigh Vacuum Electrochemistry Bibliography

Presented in this Appendix is a near-exhaustive bibliography of UHV/EC-based investigations of single-crystal electrodes published since 1991. A compilation of earlier work has been published previously (G.J. Cali, J.R. McBride, M.P. Soriaga, *Prog. Surf. Sci.*, **39**, 422 (1992)). Owing to space limitations, no listing is provided for studies with polycrystalline electrodes which are more numerous. Dr Y.-G. Kim and Dr J.E. Soto assisted in the compilation of this bibliography.

A.T. Hubbard, J.Y. Gui, 'Molecular Chemistry at Well-defined Surfaces', *J. Chim. Phys.*, **88**, 1547 (1991).

S.C. Chang, M.J. Weaver, 'In-situ IR Spectroscopy at Single-crystal Electrodes: An Emerging Link Between EC and UHV Surface Science', *J. Phys. Chem.*, **95**, 5391 (1991).

J.Y. Gui, D.A. Stern, C.H. Lin, A.T. Hubbard, 'Potential-dependent Surface Chemistry of Hydroxypyridines Adsorbed at Pt(111) Studied By EELS, LEED, AES, and Electrochemistry', *Langmuir*, **7**, 3183 (1991).

C. Shannon, D.G. Frank, A.T. Hubbard, 'Electrode-reactions of Well-characterized Adsorbed Molecules', *Annu. Rev. Phys. Chem.*, **42**, 393 (1991).

J.Y. Gui, D.A. Stern, F. Lu, A.T. Hubbard, 'Surface-chemistry of 5-Membered Heteroaromatics at Pt(111) Studied by EELS, LEED, AES and Electrochemistry: Furan, Pyrrole and Thiophene', *J. Electroanal. Chem.*, **305**, 37 (1991).

D. Aberdam, R. Durand, R. Faure, 'Study of UPD on Metal Single-crystals by Surface Techniques', *J. Chim. Phys.*, **88**, 1519 (1991).

J.Y. Gui, D.A. Stern, D.G. Frank, A.T. Hubbard, 'Adsorption and Surface Structural Chemistry of Thiophenol, Benzyl Mercaptan, and Alkyl Mercaptans: Comparative-studies at Ag(111) and Pt(111): AES, EELS, LEED, and EC', *Langmuir*, **7**, 955 (1991).

N. Kizhakevariam, E.M. Stuve, R. Dohloelze, 'Coadsorption of Water and Chlorine on Ag(110): Evidence for Adsorbate-induced Hydrophilicity', *J. Chem. Phys.*, **94**, 670 (1991).

J.K. Sass, D. Lackey, J. Schott, 'Electrochemical Double-layer Simulations by Halogen, Alkali and Hydrogen Coadsorption with Water on Metal Surfaces', *Surf. Sci.*, **247**, 239 (1991).

T. Solomun, 'The Role of the Electrolyte Anion in Anodic Dissolution of Pd(100)', *J. Electroanal. Chem.*, **302**, 31–46 (1991).

A.T. Hubbard, J.Y. Gui, 'Molecular Chemistry at Well-defined Surfaces', *J. Chim. Phys.*, **88**, 1547 (1991).

X.P. Gao, A. Hamelin, M.J. Weaver, 'Atomic Relaxation at Ordered Electrode Surfaces Probed by STM: Au(111) in Aqueous-solution Compared with UHV Environments', *J. Chem. Phys.*, **95**, 6993 (1991).

K.R. Zavadil, D. Ingersoll, J.W. Rogers, 'Electrochemical and Surface Spectroscopic Investigation of Cu(111) Films on Ru(0001)', *J. Electroanal. Chem.*, **318**, 223 (1991).

I. Villegas, J.L. Stickney, 'GaAs Deposition on Au(100) and Au(110): A LEED, AES, and STM Study', *J. Vac. Sci. Technol. A*, **10**, 3032 (1992).

P.N. Ross, A.T. D'Agostino, 'The Effect of Surface Reconstruction on the Capacitance of Au(100) Surfaces', *Electrochim. Acta*, **37**, 615 (1992).

J.R. McBride, J.A. Schimpf, M.P. Soriaga, 'In-situ Chemisorption-induced Reordering of Disordered Pd(100) Surfaces', *J. Am. Chem. Soc.*, **114**, 10 950 (1992).

M.P. Soriaga, 'UHV Techniques in the Study of Single-crystal Electrode Surfaces', *Prog. Surf. Sci.*, **39**, 325 (1992).

N. Kizhakevariam, E.M. Stuve, 'Coadsorption of Water and Hydrogen on Pt(100): Formation of Adsorbed Hydronium Ions', *Surf. Sci.*, **275**, 223 (1992).

E.Y. Cao, P. Gao, J.Y. Gui, A.T. Hubbard, 'Adsorption and Electrochemistry of SCN^- at Ag(111) and Pt(111): AES, CV, HREELS and LEED', *J. Electroanal. Chem.*, **339**, 311 (1992).

G.N. Salaita, A.T. Hubbard, 'Surface Characterization of Molecules at Pt(111) by LEED, Auger, HREELS and EC', *Catal. Today*, **12**, 465 (1992).

D.C. Zapien, J.Y. Gui, D.A. Stern, A.T. Hubbard, 'Surface Electrochemistry and Molecular Orientation: Studies of Pyridyl Hydroquinones at Pt(111)', *J. Electroanal. Chem.*, **330**, 469 (1992).

E.M. Stuve, N. Kizhakevariam, 'Chemistry and Physics of the Liquid–Solid Interface: A Surface Science Perspective', *Vac. Sci. Technol. A*, **11**, 2217 (1993).

R.A. Bradley, R. Georgiadis, S.D. Kevan, 'Nonlinear-optical Spectroscopy of the Ag(111) Surface in an Electrolyte and in Vacuum', *J. Chem. Phys.*, **99**, 5535 (1993).

K. Tanaka, 'UHV-EC: Electrochemical Preparation of Specific Metal Surfaces', *Denki Kagaku*, **61**, 392 (1993).

S.G. Sun, D.F. Yang, S.J. Wu, 'Electrochemical, AES and LEED of Au(210) in the Presence of Adsorbed Pyridine', *J. Electroanal. Chem.*, **349**, 211 (1993).

W.H. Hung, J. Schwartz, S.L. Bernasek, 'Adsorption of H_2O on Oxidized Fe(100) Surfaces', *Surf. Sci.*, **294**, 21 (1993).

J.A. Schimpf, J.R. McBride, M.P. Soriaga, 'Adsorbate-catalyzed Layer-by-layer Metal Dissolution in Halide-free Solutions', *J. Phys. Chem.*, **97**, 10 518 (1993).

P. Len, T. Solomun, 'Underpotential Deposition of Cu on Pd(100): An Electron Spectroscopy Study', *J. Electroanal. Chem.*, **353**, 131 (1993).

D.G. Frank, O.M.R. Chyan, T. Golden, 'I Monolayer Structures at Pt(111) by Means of Angular-distribution Auger Microscopy', *J. Phys. Chem.*, **97**, 5444 (1993).

J.F. Rodriguez, D.L. Taylor, H.D. Abruna, 'Concentration-dependence of the UPD of Ag on Pt(111): Electrochemical and UHV Studies', *Electrochim. Acta*, **38**, 235 (1993).

R.L. Borup, D.E. Sauer, E.M. Stuve, 'Electrodeposited Pb on Pt(111). Surface Redox Behavior of Lead and Dynamic Emersion', *Surf. Sci.*, **293**, 10 (1993).

R.L. Borup, D.E. Sauer, E.M. Stuve, 'Electrodeposited Pb on Pt(111). Vacuum Characterization and Thermal Desorption of the Emersed Adlayer', *Surf. Sci.*, **293**, 27 (1993).

E.Y. Cao, D.A. Stern, J.Y. Gui, A.T. Hubbard, 'Studies of Ru(001) Electrodes in Aqueous Electrolytes', *J. Electroanal. Chem.*, **354**, 71 (1993).

A. Krasnopoler, E.M. Stuve, 'Stabilization of Water by Fluorine on Ag(110)', *Surf. Sci.*, **303**, 355 (1994).

J.A. Schimpf, J.B. Abreu, M.P. Soriaga, 'In-situ Reordering of Extensively Disordered Pd(100) Surfaces', *Electrochim. Acta*, **39**, 2445 (1994).

J.A. Schimpf, J.B. Abreu, A. Carrasquillo, M.P. Soriaga, 'Layer-by-layer Metal Dissolution in Inert Electrolyte: Pd(100)–c(2 × 2)–I', *Surf. Sci.*, **314**, L909 (1994).

J.A. Schimpf, J.B. Abreu, M.P. Soriaga, 'Electrochemical Regeneration of Clean and Ordered Pd(100) Surfaces By

Iodine Adsorption–Desorption', *J. Electroanal. Chem.*, **364**, 247 (1994).

N. Kizhakevariam, X.D. Jiang, M.J. Weaver, 'IR Spectroscopy of Model Electrochemical Interfaces in UHV', *J. Chem. Phys.*, **100**, 6750 (1994).

R.L. Borup, D.E. Sauer, E.M. Stuve, 'Electrochemical and Vacuum Behavior of CO and Pb Coadsorbed on Pt(111)', *J. Electroanal. Chem.*, **374**, 235 (1994).

R.L. Borup, D.E. Sauer, E.M. Stuve, 'Electrochemical and Vacuum Coadsorption of CO and Pb on Pt(111)', *J. Vac. Sci. Technol. A*, **12**, 1886 (1994).

A.T. Hubbard, 'Surface Electrochemistry of Metals', *Heterogen. Chem. Rev.*, **1**, 3 (1994).

M.P. Soriaga, 'Electrochemical Surface Science: Applications of UHV Techniques', *Kikan Kagaku Sosetsu*, **26**, 76 (1995).

E.M. Stuve, A. Krasnopoler, D.E. Sauer, 'Relating the In-situ, Ex-situ, and Non-situ Environments in Surface Electrochemistry', *Surf. Sci.*, **335**, 177 (1995).

A. Krasnopoler, E.M. Stuve, 'Evidence of Specific and Nonspecific Adsorption of Perchlorate (ClO_4) on Silver(110)', *J. Vac. Sci. Technol. A*, **13**, 1681 (1995).

A. Krasnopoler, A.L. Johnson, E.M. Stuve, 'Hydrogen Bonding and Molecular Orientation in Water–Fluorine Adlayers on Ag(110)', *Surf. Sci.*, **328**, 186 (1995).

P. Mrozek, Y.E. Sung, A. Wieckowski, 'Ag Deposition on Au(111)', *Surf. Sci.*, **335**, 44 (1995).

F. Bensebaa, T.H. Ellis, 'Water at Surfaces: What Can We Learn from Vibrational Spectroscopy?', *Prog. Surf. Sci.*, **50**, 173 (1995).

Y. Hori, H. Wakebe, T. Tsukamoto, 'Adsorption of CO Accompanied with Simultaneous Charge-transfer on Cu Single-crystals', *Surf. Sci.*, **335**, 258 (1995).

G.M. Brisard, E. Zenati, H.A. Gasteiger, 'UPD of Pb on Cu(111)', *Langmuir*, **11**, 2221 (1995).

M.P. Soriaga, J.A. Schimpf, A. Carrasquillo, 'Electrochemistry of the I-on-Pd Single-crystal Interface: Studies by UHV-EC and In-situ STM', *Surf. Sci.*, **335**, 273 (1995).

J.A. Schimpf, A. Carrasquillo, M.P. Soriaga, 'Electrochemical Digital Etching in Inert Electrolyte', *Electrochim. Acta*, **40**, 1203 (1995).

J.B. Abreu, R.J. Barriga, W. Temesghen, 'I-Catalyzed Dissolution and Reordering of Ion-bombarded Pd(111)', *J. Electroanal. Chem.*, **381**, 239 (1995).

R. Dalbeck, W. Vielstich, 'Ex-situ Analysis of Sulfate Adsorption on Pt via Thermal Desorption Spectroscopy in UHV', *Electrochim. Acta*, **40**, 2687 (1995).

I. Villegas, R. Gomez, M.J. Weaver, 'NO as a Probe Adsorbate for Linking Pt(111) Electrochemical and Model UHV Interfaces', *J. Phys. Chem.*, **99**, 14 832 (1995).

N. Kizhakevariam, I. Villegas, M.J. Weaver, 'Model Electrochemical Interfaces in UHV: Solvent-induced Surface-potential Profiles on Pt(111) from Work-function Measurements and Infrared Stark Effects', *Surf. Sci.*, **336**, 37 (1995).

N.M. Markovic, H.A. Gasteiger, C.A. Lucas, 'The Effect of Cl^- on the UPD of Cu on Pt(111)', *Surf. Sci.*, **335**, 91 (1995).

U.W. Hamm, V. Lazarescu, D.M. Kolb, 'Adsorption of Pyrazine on Au(111) and Ag(111): An XPS Study', *J. Chem. Soc. Faraday Trans.*, **T92**, 3785 (1996).

A. Krasnopoler, N. Kizhakevariam, E.M. Stuve, 'Hydrogen Bonding and Surface Interactions in Protic Solvents: Coadsorption of NH_3 and HF with Water on Ag(110)', *J. Chem. Soc. Faraday Trans.*, **T92**, 2445 (1996).

M.P. Soriaga, R.J. Barriga, M.R. Marrero, J.E. Sotó, B. Bravo-Cisneros, M.A. del Valle, 'Tecnicas de Ultravacio En La Electroquimica De Superficie', *Quim. Actualid. Futuro.*, **5**, 69 (1996).

J.B. Abreu, J.-J. Jeng, M.P. Soriaga, J.F. Garst, T.L. Wade, J.L. Stickney, 'The Solid–Liquid Interface in Grignard Reagent Formation', in *Sixth International Symposium on Electrode Processes*, eds. K. Itaya, A. Wieckowski, The Electrochemical Society, Pennington, NJ, 233–242, 1996.

I. Villegas, M.J. Weaver, 'IR Spectroscopy of Model Electrochemical Interfaces in UHV: Evidence for Coupled Cation–Anion Hydration in the Pt(111)/K^+,Cl^- System', *J. Phys. Chem.*, **100**, 19 502 (1996).

I. Villegas, M.J. Weaver, 'IR Spectroscopy of Model Electrochemical Interfaces in UHV: Interfacial Cation Solvation by NH_3 on Pt(111)', *Surf. Sci.*, **367**, 162 (1996).

W.F. Temesghen, J.E. Soto, M.P. Soriaga, 'Coordination Chemistry Pd(110) Single-crystal Surfaces', in *Sixth International Symposium on Electrode Processes*, eds. K. Itaya, A. Wieckowski, The Electrochemical Society, Pennington, NJ, 58–67, 1996.

T.E. Lister, L.P. Colletti, J.L. Stickney, 'Electrochemical Formation of CdSe Monolayers on the Low-index Planes of Au', *Israel J. Chem.*, **37**, 287 (1997).

I.S.M. Thurgate, T. Naumovski, P. Hale, 'Surface Reconstructions on Au in Contact with Solution', *Surf. Rev. Lett.*, **4**, 1375 (1997).

G.M. Brisard, E. Zenati, H.A. Gasteiger, 'UPD of Pb on Cu(100) in the Presence of Cl^-: LEED, AES, and EC Studies', *Langmuir*, **13**, 2390 (1997).

G.M. Brisard, E. Zenati, H.A. Gasteiger, 'LEED, AES and EC Studies of the UPD of Pb on Cu(100) and Cu(111)', *ACS Symp. Ser.*, **656**, 142 (1997).

A. Carrasquillo, J.-J. Jeng, R.J. Barriga, 'Ligand Substitution and Competitive Coordination of Halides at Well-defined Pd(100) and Pd(111)', *Inorg. Chim. Acta*, **255**, 249 (1997).

R.G. Jones, M. Kadodwala, 'Br Adsorption on Cu(111)', *Surf. Sci.*, **370**, L219 (1997).

R.L. Borup, D.E. Sauer, E.M. Stuve, 'Electrolyte Interactions with Vapor- and Solution-dosed CO on Pt(111)', *Surf. Sci.*, **374**, 142 (1997).

D.E. Sauer, R.L. Borup, E.M. Stuve, 'Electrocatalysis of HCOOH and CO with Probe Adlayers of C and Ethylidyne on Pt(111)', *ACS Symp. Ser.*, **656**, 283 (1997).

I. Villegas, M.J. Weaver, 'Modeling Electrochemical Interfaces in UHV: Molecular Roles of Solvation in Double-layer Phenomena', *J. Phys. Chem. B*, **101**, 10 166 (1997).

I. Villegas, M.J. Weaver, 'IR Spectroscopy of Model Electrochemical Interfaces in UHV: Acetone Chemisorption on Pt(111)', *J. Electoanal. Chem.*, **426**, 55 (1997).

A. Carrasquillo, J.-J. Jeng, M.P. Soriaga, 'The Interfacial Chemistry of Pd(100) Single-crystal Electrodes in Aqueous Cl^- Solutions', *Proc. Electrochem. Soc.*, **17**, 263 (1997).

M.P. Soriaga, K. Itaya, J.L. Stickney, 'STM and UHV-EC: Complementary, Noncompeting Techniques', in *Electrochemical Nanotechnology*, eds. W. Lorenz, W. Plieth, Springer-Verlag, Berlin, 267–276, 1998.

E.A. Lafferty, Y.-G. Kim, M.P. Soriaga, 'Selective and Quantitative Removal of Pd Films from Pt Substrates by Adsorbed-I-catalyzed Anodic Stripping', *Electrochim. Acta*, **44**, 1031 (1998).

J.B. Abreu, A.A. Ashley-Facey, J.E. Soto, M.P. Soriaga, 'The Interfacial Chemistry of the Grignard Reaction', *J. Coll. Interf. Sci.*, **206**, 247 (1998).

Y.-G. Kim, M.P. Soriaga, 'Evidence of I–Pd Place Exchange in the I-catalyzed Dissolution of Pd(111)', *J. Phys. Chem. B*, **102**, 6188 (1998).

G. Pirug, H.P. Bonzel, 'UHV Simulation of the Electrochemical Double Layer: Adsorption of $HClO_4/H_2O$ on Au(111)', *Surf. Sci.*, **405**, 87 (1998).

J. Inukai, Y. Osawa, K. Itaya, 'Adlayer Structures of Chlorine, Bromine, and Iodine on Cu(111) Electrode in Solution: STM and LEED Studies', *J. Phys. Chem. B*, **102**, 10 034 (1998).

C. Stuhlmann, Z. Park, C. Bach, 'An Ex-situ Study of Cd UPD on Cu(111)', *Electrochim. Acta*, **44**, 993 (1998).

A. Wieckowski, A. Kolics, J.C. Polkinghorne, 'Radioactive Labeling, EC, and UHV Study of Sulfate Adsorption on Type 316 Stainless Steel', *Corrosion*, **54**, 800 (1998).

C. Tang, S. Zou, M.W. Severson, 'Coverage-dependent IR Spectroscopy of CO on Ir(111) in Aqueous Solution: Comparison Between EC and UHV Environments', *J. Phys. Chem. B*, **102**, 8796 (1998).

R. Gomez, J.M. Feliu, A. Aldaz, 'Double-layer Charge-corrected Voltammetry for Assaying CO Coverages: Comparisons Between EC and UHV Environments', *Surf. Sci.*, **410**, 48 (1998).

L.F. Li, D. Totir, Y. Gofer, 'The Electrochemistry of Ni in a Li-based Solid Polymer Electrolyte in UHV', *Electrochim. Acta*, **44**, 949 (1998).

L.F. Li, D. Totir, G.S. Chottiner, 'Electrochemical Reactivity of CO and S Adsorbed on Ni(111) and Ni(110) in a Li-based Solid Polymer Electrolyte in UHV', *J. Phys. Chem. B*, **102**, 8013 (1998).

M.J. Weaver, 'Potentials of Zero Charge for Pt(111)-aqueous Interfaces: A Combined Assessment from EC and UHV Measurements', *Langmuir*, **14**, 3932 (1998).

D.Y. Zemlyanov, E. Savinova, A. Scheybal, 'XPS Observation of OH Groups Incorporated in an Ag(111) Electrode', *Surf. Sci.*, **418**, 441 (1998).

T.D. Jarvi, T.H. Madden, E.M. Stuve, 'UHV and EC Behavior of Vapor-deposited Ru on Pt(111)', *Electrochem. Solid St.*, **2**, 224 (1999).

Y.-G. Kim, M.P. Soriaga, 'The Interfacial Chemistry of Pd Electrodes', in *Interfacial Electrochemistry*, ed. A. Wieckowski, Marcel Dekker, New York, 249–268, 1999.

J.E. Soto, Y.-G. Kim, M.P. Soriaga, 'UHV-EC and EC-STM Studies of Molecular Chemisorption at Well-defined Pd(hkl) Surfaces', *Electrochem. Commun.*, **1**, 135 (1999).

REFERENCES

1. A.T. Hubbard, R.P. Ishikawa, J. Katekaru, *J. Electroanal. Chem.*, **86**, 271 (1978).
2. G.G. Will, *J. Electrochem. Soc.*, **112**, 451 (1965).
3. J. Clavilier, *J. Electroanal. Chem.*, **107**, 205 (1980).
4. R. Adzic, in *Modern Aspects of Electrochemistry XXI*, eds. R.E. White, J.O'M. Bockris, B.E. Conway, Plenum, New York, 1990.
5. M.P. Soriaga, *Prog. Surf. Sci.*, **39**, 325 (1992).
6. M.P. Soriaga, D.A. Harrington, J.L. Stickney, A. Wieckowski, in *Modern Aspects of Electrochemistry XXVII*, eds. B.E. Conway, R.E. White, J.O'M. Bockris, Plenum, New York, 1996.
7. I.M. Villegas, M.J. Weaver, *J. Phys. Chem. B*, **101**, 10 166 (1997).
8. M.P. Soriaga, R.J. Barriga, M.R. Marrero, J.E. Sotó, B. Bravo-Cisneros, M.A. del Valle, *Quim. Actualid. Futuro.*, **5**, 69 (1996).
9. G.A. Somorjai, *Chemistry in Two Dimensions: Surfaces*, Cornell University, Ithaca, 1981.
10. T.H. Rhodin, G. Ertl (eds.), *The Nature of the Surface Chemical Bond*, North-Holland Publishing, New York, 1979.
11. G. Ertl, J. Kuppers, *Low Energy Electrons and Surface Chemistry*, VCH Publishers, New York, 1985.
12. D.P. Woodruff, T.A. Delchar, *Modern Techniques of Surface Science*, Cambridge University Press, New York, 1986.
13. A. Zangwill, *Physics at Surfaces*, Cambridge University Press, New York, 1988.
14. M.A. Van Hove, S.W. Wang, D.F. Ogletree, G.A. Somorjai, *Adv. Quantum Chem.*, **20**, 1 (1989).
15. G.A. Somorjai, *Introduction to Surface Chemistry and Catalysis*, Wiley, New York, 1994.
16. E.A. Wood, *Crystal Orientation Manual*, Columbia University Press, New York, 1963.
17. A.W. Adamson, *Physical Chemistry of Surfaces*, Wiley, New York, 1990.
18. M.S. Zei, Y. Nakai, G. Lehmpfuhl, D.M. Kolb, *J. Electroanal. Chem.*, **150**, 201 (1983).
19. K.W. Nebesny, N.R. Armstrong, *J. Electron Spectrosc.*, **37**, 355 (1986).
20. J.M. Morabito, *Surf. Sci.*, **49**, 318 (1975).
21. W.A. Coghlan, R.E. Clausing, *At. Data*, **5**, 317 (1973).
22. J. Houston, *Surf. Sci.*, **38**, 283 (1973).

23. J.A. Schoeffel, A.T. Hubbard, *Anal. Chem.*, **49**, 2330 (1977).
24. B.C. Schardt, J.L. Stickney, D.A. Stern, D.G. Frank, J.Y. Katekaru, S.D. Rosasco, G.N. Salaita, M.P. Soriaga, A.T. Hubbard, *Inorg. Chem.*, **24**, 1419 (1985).
25. G.N. Salaita, A.T. Hubbard, in *Molecular Design of Electrode Surfaces*, ed. R.W. Murray, Wiley, New York, 1989.
26. M. Gryzinski, *Phys. Rev.*, **A138**, 305 (1965).
27. K. Siegbahn, *ESCA, Atomic, Molecular and Solid State Structure Studies by Means of Electron Spectroscopy*, Almqvist-Wiksell Boktryckeri, Sweden, 1967.
28. M.P. Seah, in *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, eds. D. Briggs, M.P. Seah, Wiley, New York, 1983.
29. H. Ibach, D.A. Mills, *Electron Energy Loss Spectroscopy*, Academic Press, New York, 1982.
30. N.R. Avery, in *Vibrational Spectroscopy of Molecules on Surfaces*, eds. J.T. Yates, Jr, T.E. Madey, Plenum, New York, 1987.
31. F.M. Hoffman, *Surf. Sci. Rep.*, **3**, 2 (1983).
32. H.A. Pearce, N. Sheppard, *Surf. Sci.*, **59**, 205 (1976).
33. A. Bewick, B.S. Pons, in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Hester, R.E. Clark, Wiley-Hayden, London, 1985.
34. H.G. Tompkins, in *Methods of Surface Analysis*, ed. A. Czanderna, Elsevier, New York, 1975.
35. H. Froitzheim, in *Topics in Current Physics. IV*, ed. H. Ibach, Springer-Verlag, New York, 1977.
36. W.N. Hansen, *J. Electroanal. Chem.*, **150**, 133 (1983).
37. W.N. Hansen, G.J. Hansen, in *Electrochemical Surface Science*, ed. M.P. Soriaga, American Chemical Society, Washington, DC, 1988.
38. P.A. Redhead, *Vacuum*, **12**, 203 (1962).
39. C.N. Chittenden, E.D. Pylant, A.L. Schwaner, J.M. White, in *Surface Imaging and Visualization*, ed. A.T. Hubbard, CRC Press, Boca Raton, 1995.
40. H.A. Strobel, W.R. Heineman, *Chemical Instrumentation*, Wiley, New York, 1989.
41. J. Augustynski, L. Balsenc, in *Modern Aspects of Electrochemistry XIV*, eds. J.O'M. Bockris, B.E. Conway, Plenum, New York, 1979.
42. P.M.A. Sherwood, in *Contemporary Topics in Analytical and Clinical Chemistry*, eds. D.M. Hercules, G.M. Hieftje, L.R. Snyder, M.E. Evenson, Plenum, New York, 1982.
43. J.S. Hammond, N. Winograd, *J. Electroanal. Chem.*, **78**, 55 (1977).
44. A.T. Hubbard, *Acc. Chem. Res.*, **13**, 177 (1980).
45. P.N. Ross, in *Chemistry and Physics of Solid Surfaces*, eds. R. Vanselow, R. Howe, Springer-Verlag, New York, 1982.
46. E. Yeager, A. Homa, B.D. Cahan, D. Scherson, *J. Vac. Sci. Technol.*, **20**, 628 (1982).
47. D.M. Kolb, *Z. Phys. Chem. Neue Folge*, **154**, 179 (1987).
48. A.T. Hubbard, M.P. Soriaga, J.L. Stickney, in *New Directions in Chemical Analysis*, ed. B. Shapiro, Texas A&M University Press, College Station, 1985.
49. A.T. Hubbard, *Chem. Rev.*, **88**, 633 (1988).
50. M.E. Hanson, E. Yeager, in *Electrochemical Surface Science*, ed. M.P. Soriaga, American Chemical Society, Washington, DC, 1988.
51. J.L. Stickney, C.B. Ehlers, B.W. Gregory, in *Electrochemical Surface Science*, ed. M.P. Soriaga, American Chemical Society, Washington, DC, 1988.
52. T. Mebrahtu, J.F. Rodriguez, M.E. Bothwell, I.F. Cheng, D.R. Lawson, J.R. McBride, C.R. Martin, M.P. Soriaga, *J. Electroanal. Chem.*, **267**, 351 (1989).
53. T. Solomun, W. Richtering, H. Gerischer, *Ber. Bunsen-Ges. Phys. Chem.*, **91**, 412 (1987).
54. K.R. Zavadil, N.R. Armstrong, *J. Electrochem. Soc.*, **137**, 2371 (1990).
55. L.-W.H. Leung, T.W. Gregg, D.W. Goodman, *Rev. Sci. Instrum.*, **62**, 1857 (1991).
56. B.E. Conway, H. Angerstein-Kozlowska, W.B.A. Sharp, E.E. Criddle, *Anal. Chem.*, **45**, 1331 (1973).
57. A.J. Bard, H.D. Abruña, C.E. Chidsey, L.R. Faulkner, S. Feldberg, K. Itaya, O. Melroy, R.W. Murray, M.D. Porter, M.P. Soriaga, H.S. White, *J. Phys. Chem.*, **97**, 7147 (1993).
58. A.J. Bard, L.R. Faulkner, *Electrochemical Methods*, John Wiley, New York, 1980.
59. M.P. Soriaga (ed.), *Electrochemical Surface Science*, American Chemical Society, Washington, DC, 1988.
60. E. Yeager, J. Kuta, in *Physical Chemistry, An Advanced Treatise. IXA*, eds. H. Eyring, D. Henderson, W. Jost, Academic Press, New York, 1970.
61. W. Hansen, D.M. Kolb, *J. Electroanal. Chem.*, **100**, 493 (1979).
62. D. Rath, D.M. Kolb, *Surf. Sci.*, **109**, 641 (1981).
63. G.J. Hansen, W.N. Hansen, *J. Electroanal. Chem.*, **150**, 193 (1983).
64. W.N. Hansen, *Surf. Sci.*, **101**, 109 (1980).
65. J.D.E. McIntyre, *Adv. Electrochem. Electrochem. Engr.*, **9**, 61 (1973).
66. W. Hansen, D.M. Kolb, D. Rath, R. Wille, *J. Electroanal. Chem.*, **110**, 369 (1980).
67. O. Hofman, K. Doblhofer, H. Gerischer, *J. Electroanal. Chem.*, **161**, 337 (1984).
68. G.J. Hansen, W.N. Hansen, *Ber. Bunsen-Ges. Phys. Chem.*, **91**, 317 (1987).
69. J.E. Pemberton, R.L. Sobocinksi, M.A. Bryant, *J. Am. Chem. Soc.*, **112**, 6177 (1990).
70. J.E. Pemberton, R.L. Sobocinksi, *J. Electroanal. Chem.*, **318**, 157 (1991).
71. E.R. Koetz, H. Neff, R.H. Muller, *J. Electroanal. Chem.*, **215**, 331 (1986).
72. S.D. Rosasco, J.L. Stickney, G.N. Salaita, D.G. Frank, J.Y. Katekaru, B.C. Schardt, M.P. Soriaga, D.A. Stern, A.T. Hubbard, *J. Electroanal. Chem.*, **188**, 95 (1985).

73. D.G. Frank, J.Y. Katekaru, S.D. Rosasco, G.N. Salaita, B.C. Schardt, M.P. Soriaga, J.L. Stickney, A.T. Hubbard, *Langmuir*, **1**, 587 (1985).
74. J.F. Rodriguez, T. Mebrahtu, M.P. Soriaga, *J. Electroanal. Chem.*, **264**, 291 (1979).
75. T.E. Felter, A.T. Hubbard, *J. Electroanal. Chem.*, **100**, 473 (1979).
76. B.C. Schardt, S.L. Yau, F. Rinaldi, *Science*, **243**, 1050 (1989).
77. D.A. Stern, L. Laguren-Davidson, F. Lu, C.H. Lin, D.G. Frank, G.N. Salaita, N. Walton, J.Y. Gui, D.C. Zapien, A.T. Hubbard, *J. Am. Chem. Soc.*, **111**, 877 (1989).
78. T.E. Madey, F.P. Netzer, *Surf. Sci.*, **117**, 549 (1980).
79. J.K. Sass, K. Kretzschmar, S. Holloway, *Vacuum*, **31**, 483 (1981).
80. J.K. Sass, *Vacuum*, **33**, 741 (1983).
81. J.K. Sass, K. Bange, in *Electrochemical Surface Science*, ed. M.P. Soriaga, American Chemical Society, Washington, DC, 1988.
82. J.K. Sass, K. Bange, R. Dohl, E. Plitz, R. Unwin, *Ber. Bunsenges, Phys. Chem.*, **88**, 354 (1984).
83. F.T. Wagner, S.J. Schmieg, T.E. Moylan, *Surf. Sci.*, **195**, 403 (1988).
84. F.T. Wagner, T.E. Moylan, in *Electrochemical Surface Science*, ed. M.P. Soriaga, American Chemical Society, Washington, DC, 1988.
85. E.M. Stuve, A. Krasnopoler, D.E. Sauer, *Surf. Sci.*, **335**, 177 (1995).
86. D.A. Stern, H. Baltruschat, M. Martinez, J.L. Stickney, D. Song, S.K. Lewis, D.G. Frank, A.T. Hubbard, *J. Electroanal. Chem.*, **217**, 101 (1987).
87. G.N. Salaita, D.A. Stern, F. Lu, H. Baltruschat, B.C. Schardt, J.L. Stickney, M.P. Soriaga, D.G. Frank, A.T. Hubbard, *Langmuir*, **2**, 828 (1986).
88. F. Lu, G.N. Salaita, H. Baltruschat, A.T. Hubbard, *J. Electroanal. Chem.*, **222**, 305 (1987).
89. N. Batina, J.W. McCargar, L. Laguren-Davidson, C.-H. Lin, A.T. Hubbard, *Langmuir*, **5**, 123 (1989).
90. J.L. Stickney, S.D. Rosasco, G.N. Salaita, A.T. Hubbard, *Langmuir*, **1**, 66 (1985).
91. B.C. Schardt, J.L. Stickney, D.A. Stern, D.G. Frank, J.Y. Katekaru, S.D. Rosasco, G.N. Salaita, M.P. Soriaga, A.T. Hubbard, *Inorg. Chem.*, **24**, 1419 (1985).
92. S.D. Rosasco, J.L. Stickney, G.N. Salaita, D.G. Frank, J.Y. Katekaru, B.C. Schardt, M.P. Soriaga, D.A. Stern, A.T. Hubbard, *J. Electroanal. Chem.*, **188**, 95 (1985).
93. G.A. Garwood, A.T. Hubbard, *Surf. Sci.*, **112**, 281 (1982).
94. G.A. Garwood, A.T. Hubbard, *Surf. Sci.*, **121**, 396 (1982).
95. T. Solomun, A. Wieckowski, S.D. Rosasco, A.T. Hubbard, *Surf. Sci.*, **147**, 241 (1984).
96. C.B. Ehlers, J.L. Stickney, *Surf. Sci.*, **239**, 85 (1990).
97. G.N. Salaita, F. Lu, L. Laguren-Davidson, A.T. Hubbard, *J. Electroanal. Chem.*, **229**, 1 (1987).
98. B.G. Bravo, S.L. Michelhaugh, M.P. Soriaga, I. Villegas, D.W. Suggs, J.L. Stickney, *J. Phys. Chem.*, **95**, 5245 (1991).
99. T. Mebrahtu, M.E. Bothwell, J.E. Harris, G.J. Cali, M.P. Soriaga, *J. Electroanal. Chem.*, **300**, 487 (1991).
100. D.G. Frank, J.Y. Katekaru, S.D. Rosasco, G.N. Salaita, B.C. Schardt, M.P. Soriaga, J.L. Stickney, A.T. Hubbard, *Langmuir*, **1**, 587 (1985).
101. P.N. Ross, in *Electrochemical Surface Science*, ed. M.P. Soriaga, American Chemical Society, Washington, DC, 1988.
102. W.E. O'Grady, M.Y.C. Woo, P.L. Hagans, E. Yeager, *J. Vac. Sci. Technol.*, **14**, 365 (1977).
103. K. Yamamoto, D.M. Kolb, R. Kotz, G. Lempfuhl, *J. Electroanal. Chem.*, **96**, 233 (1979).
104. P.N. Ross, *J. Electroanal. Chem.*, **126**, 67 (1979).
105. A.S. Homa, E. Yeager, B.D. Cahan, *Electroanal. Chem.*, **150**, 181 (1983).
106. J. Clavilier, A. Rodes, K. El Achi, M.A. Zamakhchari, *J. Chim. Phys.*, **88**, 1291 (1991).
107. P.N. Ross, *J. Chim. Phys.*, **88**, 1353 (1991).
108. G. Jerkiewicz, B.E. Conway, *J. Chim. Phys.*, **88**, 1291 (1991).
109. F.T. Wagner, P.N. Ross, *Surf. Sci.*, **160**, 305 (1985).
110. J. Horkans, B.D. Cahan, E. Yeager, *Surf. Sci.*, **37**, 559 (1973).
111. F.T. Wagner, P.N. Ross, *J. Electroanal. Chem.*, **150**, 141 (1985).
112. J.B. Lumsden, G.A. Garwood, A.T. Hubbard, *Surf. Sci.*, **121**, 1524 (1982).
113. D.A. Harrington, A. Wieckowski, S.D. Rosasco, B.C. Schardt, G.N. Salaita, J.B. Lumsden, A.T. Hubbard, *Corrosion Sci.*, **25**, 849 (1985).
114. D.A. Harrington, A. Wieckowski, S.D. Rosasco, B.C. Schardt, G.N. Salaita, A.T. Hubbard, *Langmuir*, **1**, 232 (1985).
115. J.P. Biberian, G.A. Somorjai, *J. Vac. Sci. Technol.*, **16**, 2073 (1979).
116. J.S. Hammond, N. Winograd, *J. Electrochem. Soc.*, **124**, 826 (1977).
117. A.T. Hubbard, J.L. Stickney, S.D. Rosasco, M.P. Soriaga, D. Song, *J. Electroanal. Chem.*, **150**, 165 (1983).
118. J.L. Stickney, S.D. Rosasco, M.P. Soriaga, D. Song, A.T. Hubbard, *Surf. Sci.*, **130**, 326 (1983).
119. J.L. Stickney, S.D. Rosasco, B.C. Schardt, A.T. Hubbard, *J. Phys. Chem.*, **88**, 251 (1984).
120. J.L. Stickney, S.D. Rosasco, A.T. Hubbard, *J. Electrochem. Soc.*, **131**, 260 (1984).
121. J.L. Stickney, D.A. Stern, B.C. Schardt, D.C. Zapien, A. Wieckowski, A.T. Hubbard, *J. Electroanal. Chem.*, **213**, 293 (1986).
122. J.L. Stickney, B.C. Schardt, D.A. Stern, A. Wieckowski, A.T. Hubbard, *J. Electrochem. Soc.*, **133**, 648 (1986).
123. B.C. Schardt, J.L. Stickney, D.A. Stern, A. Wieckowski, D.C. Zapien, A.T. Hubbard, *Langmuir*, **3**, 239 (1987).

124. B.C. Schardt, J.L. Stickney, D.A. Stern, A. Wieckowski, D.C. Zapien, A.T. Hubbard, *Surf. Sci.*, **175**, 520 (1986).
125. P.C. Andricacos, P.N. Ross, *J. Electroanal. Chem.*, **167**, 301 (1984).
126. D. Aberdam, R. Durand, R. Faure, F. El-Omar, *Surf. Sci.*, **162**, 782 (1985).
127. L.-W.H. Leung, T.W. Gregg, D.W. Goodman, *Langmuir*, **7**, 3205 (1991).
128. L. Laguren-Davidson, F. Lu, G.N. Salaita, A.T. Hubbard, *Langmuir*, **4**, 224 (1988).
129. B.W. Gregory, J.L. Stickney, *J. Electroanal. Chem.*, **300**, 543 (1991).
130. D.W. Suggs, J.L. Stickney, *J. Phys. Chem.*, **95**, 10 056 (1991).
131. K. Itaya, *Prog. Surf. Sci.*, **58**, 121 (1998).
132. J.A. Schimpf, J.R. McBride, M.P. Soriaga, *J. Phys. Chem.*, **97**, 10 518 (1993).
133. Y.-G. Kim, M.P. Soriaga, *J. Phys. Chem. B*, **102**, 6188 (1998).
134. M.P. Soriaga, in *Frontiers of Electrochemistry II*, eds. P.N. Ross, J. Lipkowski, VCH, New York, 1992.
135. B.B. Damaskin, A.N. Frumkin, *Electrochim. Acta*, **19**, 173 (1974).
136. R. Parsons, *J. Electroanal. Chem.*, **59**, 229 (1975).
137. P.A. Thiel, T.E. Madey, *Surf. Sci. Rep.*, **7**, 211 (1987).
138. C.K. Mann, K.K. Barnes, *Electrochemical Reactions in Nonaqueous Solvents*, Marcel Dekker, New York, 1970.
139. H. Ohtani, C.T. Kao, M.A. Van Hove, G.A. Somorjai, *Prog. Surf. Sci.*, **23**, 155 (1987).
140. G.A. Garwood, Jr, A.T. Hubbard, *Surf. Sci.*, **118**, 233 (1982).
141. J.Y. Gui, F. Lu, D.A. Stern, A.T. Hubbard, *J. Electroanal. Chem.*, **292**, 245 (1990).
142. B.B. Damaskin, O.A. Petrii, V.V. Batrakov, *Adsorption of Organic Compounds on Electrodes*, Plenum, New York, 1971.
143. R.D. Snell, A.G. Keenan, *Chem. Soc. Rev.*, **8**, 259 (1979).
144. J.L. Stickney, M.P. Soriaga, A.T. Hubbard, S.E. Anderson, *J. Electroanal. Chem.*, **125**, 73 (1981).
145. S.L. Michelhaugh, C. Bhardwaj, G.J. Cali, B.G. Bravo, M.E. Bothwell, G.M. Berry, M.P. Soriaga, *Corrosion*, **47**, 322 (1991).
146. D. Zurawski, M. Wasberg, A. Wieckowski, *J. Phys. Chem.*, **94**, 2076 (1990).
147. G.M. Berry, M.E. Bothwell, S.L. Michelhaugh, J.R. McBride, M.P. Soriaga, *J. Chim. Phys.*, **88**, 1591 (1991).
148. G.M. Berry, J.R. McBride, J.A. Schimpf, M.P. Soriaga, *J. Electroanal. Chem.*, **353**, 281 (1993).
149. A. Wieckowski, S.D. Rosasco, G.N. Salaita, A.T. Hubbard, B.E. Bent, F. Zaera, D. Godbey, G.A. Somorjai, *J. Am. Chem. Soc.*, **107**, 5910 (1985).
150. A.T. Hubbard, M.A. Young, J.A. Schoeffel, *J. Electroanal. Chem.*, **114**, 273 (1980).
151. F. Zaera, G.A. Somorjai, *J. Am. Chem. Soc.*, **106**, 2288 (1984).
152. M. Hourani, A. Wieckowski, *Langmuir*, **6**, 379 (1990).
153. J.Y. Gui, D.A. Stern, D.C. Zapien, G.N. Salaita, F. Lu, C.H. Lin, B.E. Kahn, A.T. Hubbard, *J. Electroanal. Chem.*, **252**, 169 (1988).
154. B.E. Kahn, S.A. Chaffins, J.Y. Gui, F. Lu, D.A. Stern, A.T. Hubbard, *Chem. Phys.*, **141**, 21 (1990).
155. G.N. Salaita, C. Lin, P. Gao, A.T. Hubbard, *Arabian J. Sci. Engr.*, **15**, 319 (1990).
156. J.Y. Gui, L. Laguren-Davidson, C.H. Lin, F. Lu, G.N. Salaita, D.A. Stern, B.E. Kahn, A.T. Hubbard, *Langmuir*, **5**, 819 (1989).
157. N. Batina, D.C. Zapien, F. Lu, C.H. Lin, J.W. McCargar, B.E. Kahn, J.Y. Gui, D.G. Frank, G.N. Salaita, D.A. Stern, A.T. Hubbard, *Electrochim. Acta*, **34**, 1031 (1989).
158. N. Batina, S.A. Chaffins, J.Y. Gui, F. Lu, J.W. McCargar, J.W. Rovang, D.A. Stern, A.T. Hubbard, *J. Electroanal. Chem.*, **284**, 81 (1990).
159. N. Batina, J.W. McCargar, C.H. Lin, G.N. Salaita, B.E. Kahn, A.T. Hubbard, *Electroanal.*, **1**, 213 (1989).
160. M.P. Soriaga, E. Binamira-Soriaga, A.T. Hubbard, J.B. Benziger, K.W.P. Pang, *Inorg. Chem.*, **24**, 65 (1985).
161. M.P. Soriaga, J.H. White, V.K.F. Chia, D. Song, P.O. Arrhenius, A.T. Hubbard, *Inorg. Chem.*, **24**, 73 (1985).
162. M.P. Soriaga, J.L. Stickney, A.T. Hubbard, *J. Mol. Catal.*, **21**, 211 (1983).
163. N. Batina, B.E. Kahn, J.Y. Gui, F. Lu, J.W. McCargar, H.B. Mark, C.H. Lin, B.N. Salaita, H. Zimmer, D.A. Stern, A.T. Hubbard, *Langmuir*, **5**, 588 (1989).
164. D.A. Stern, L. Laguren-Davidson, F. Lu, C.H. Lin, D.G. Frank, G.N. Salaita, N. Walton, J.Y. Gui, D.C. Zapien, A.T. Hubbard, *J. Am. Chem. Soc.*, **111**, 877 (1989).
165. S.A. Chaffins, J.Y. Gui, B.E. Kahn, C.H. Lin, F. Lu, G.N. Salaita, D.A. Stern, D.C. Zapien, A.T. Hubbard, C.M. Elliott, *Langmuir*, **6**, 951 (1990).
166. S.A. Chaffins, J.Y. Gui, C.H. Lin, F. Lu, G.N. Salaita, D.A. Stern, A.T. Hubbard, *Langmuir*, **6**, 1273 (1990).
167. A.T. Hubbard, D.G. Frank, D.A. Stern, M.J. Tarlov, N. Batina, N. Walton, E. Wellner, J.W. McCargar, in *Redox Chemistry and Interfacial Behavior of Biological Molecules*, eds. G. Dryhurst, R. Niki, Plenum, New York, 1988.
168. D.A. Stern, N. Walton, J.W. McCargar, G.N. Salaita, L. Laguren-Davidson, F. Lu, C.H. Lin, J.Y. Gui, N. Batina, D.G. Frank, A.T. Hubbard, *Langmuir*, **4**, 711 (1988).
169. M.P. Soriaga, Y.-G. Kim, in *Interfacial Electrochemistry*, ed. A. Wieckowski, Marcel Dekker, New York, 1999.
170. J.E. Soto, Y.-G. Kim, M.P. Soriaga, *Electrochem. Commun.*, **1**, 135 (1999).

Ultrafast Electrochemical Techniques

Robert J. Forster
Dublin City University, Dublin, Ireland

1 Principles and Objectives of Ultrafast Electrochemical Techniques **1**
1.1 Introduction 1
1.2 The Need for Speed 1
2 Direct Electrochemical Methods **2**
2.1 Microelectrode Fabrication and Characterization 2
2.2 Cell Time Constants 2
2.3 Ohmic Effects 5
2.4 Mass Transport 5
2.5 Instrumental Challenges 7
2.6 Transient Techniques 8
2.7 Steady-state Electrochemistry 14
3 Indirect Methods **17**
3.1 Redox Catalysis 18
3.2 Indirect Laser-induced Temperature Jump Method 19
3.3 Scanning Electrochemical Microscopy 22
4 Future Directions **25**
4.1 Analytical Applications of Fast Electrochemistry 25
4.2 Ultimate Limits on Ultrafast Electrochemical Techniques 25
5 Conclusions **26**
Acknowledgments **26**
Abbreviations and Acronyms **26**
Related Articles **27**
References **27**

Ultrafast electrochemical techniques provide information about the kinetics and thermodynamics of redox processes that occur at submillisecond or even nanosecond timescales. This short timescale is achieved either by making very rapid transient measurements or by using ultrasmall probes to achieve very high rates of diffusion under steady-state conditions. Microelectrodes (i.e. electrodes with critical dimensions in the micrometer range) play pivotal roles in both approaches. Electrochemistry has several advantages over spectroscopy in that it provides direct information about electron transfer (E) and coupled chemical (C) reactions. Ultrafast electrochemical techniques now allow it to do so at times as short as 10 ns. In transient measurements, decreasing the lower accessible timescale depends critically on fabricating ultramicroelectrodes that continue to respond ideally as their critical dimension (e.g. the radius of a microdisc) decreases. It is now possible to assemble microelectrodes that respond to changes in applied potential within less than a few nanoseconds. In steady-state approaches, ultrasmall probes are required to make short timescale measurements and various approaches that yield nanodes (i.e. electrodes of nanometer dimension) have been proposed. However, beyond the need for smaller probes and faster instrumentation, the continued development of new theory describing electron transfer is essential, where the dimensions of the zone that is depleted of reactant because of a Faradaic reaction and the electrochemical double layer become comparable.

1 PRINCIPLES AND OBJECTIVES OF ULTRAFAST ELECTROCHEMICAL TECHNIQUES

1.1 Introduction

Irrespective of the approach taken, from the detection of highly reactive intermediates to the tantalizing possibility of directly probing the energetics and dynamics of facile electron transfers, ultrafast electrochemical techniques are revolutionizing investigations into redox processes. In doing so, they promise to revolutionize the type, quality and range of information available to test and develop new theories of electron transfer. Moreover, these high-speed approaches will allow new devices to be developed, such as sensors that exploit differences in reaction dynamics rather than energetics to achieve a selective response.

1.2 The Need for Speed

Many significant electrochemical events, such as electron and proton transfers, ligand exchanges, isomerizations, and ejection of leaving groups, occur on the low microsecond and nanosecond time domains. To achieve a meaningful insight into these redox processes, it must be possible to measure rate constants under a wide range of experimental conditions, such as driving force, temperature and so on. However, conventional electrochemical methods cannot fulfil this role since they are restricted to millisecond, or longer, timescales. Thus, while modern laser-based spectroscopy has provided a powerful new insight into chemical processes that occur at picosecond and even femtosecond timescales, it is only recently that

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

electrochemists have meaningfully probed redox processes occurring on the submicrosecond timescale.

The objective of this article is not to provide an extensive catalogue of recent studies conducted in this area. Rather, the intention is to describe the theory and practice of ultrafast electrochemical techniques, to convey a sense of the current state of the art in the area, and to look at likely future developments.

Much of our discussion will focus on microelectrodes; the ability routinely to fabricate electrodes with radii smaller than one hundredth of the thickness of a human hair has profoundly changed the way electrochemistry is undertaken. The impact of microelectrodes in redefining the spatial and time limits of electrochemistry began in the late 1960s with Fleischmann's pioneering work.[1] Since the 1970s, researchers have used these ultrasmall probes to study the kinetics and mechanism of fast heterogeneous electron transfers and homogeneous chemical reactions, to extend the range of electroanalytical measurements, to map redox activity topographically using scanning electrochemical microscopy (SECM), and to provide platforms for the fabrication of sensors for both in vitro and in vivo applications.[2] The areas of microelectrodes are more than one million times smaller than conventional electrodes, allowing electrochemical measurements to be performed under unusual conditions and in nonconventional media. For example, it is now possible to monitor the concentrations of radicals in organic solvents that live for only a few nanoseconds. Moreover, measurements can be performed with a high degree of spatial resolution, e.g. the detection of neurotransmitter release from single brain cells.

2 DIRECT ELECTROCHEMICAL METHODS

2.1 Microelectrode Fabrication and Characterization

Microelectrodes are miniature electrodes where the critical electrode dimension is less than about 10 μm yet remains much greater than the thickness of the electrical double layer, which is typically 10–100 Å (1 Å = 10^{-10} m). The physicochemical properties of microelectrodes are distinctly different from the properties observed for macroelectrodes and from those predicted for electrodes whose dimensions approach the molecular level. Although microelectrodes have been used since the 1940s, e.g. to measure oxygen concentrations within tissue,[3] it is only since the 1980s that the availability of microscopic gold and platinum wires has allowed them to be widely used.

Figure 1 illustrates the five common microelectrode geometries. The microdisc is the most popular geometry and is employed in approximately 50% of all

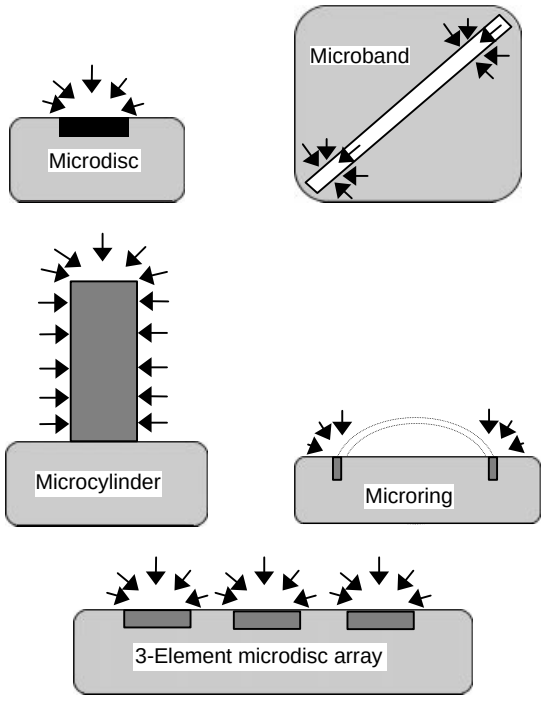

Figure 1 The most common microelectrode geometries and their associated diffusion fields.

investigations. Other common geometries include cylinders (20%) and arrays (20%), with the remaining 10% comprising bands and rings and less frequently spheres, hemispheres, and more unusual assemblies. The most popular materials include platinum, carbon fibers, and gold, although mercury, iridium, nickel, silver, and superconducting ceramics have also been used.[2] Microdisc electrodes predominate because of their ease of construction and because the sensing surface of the electrode can be mechanically polished. Microelectrodes in the form of discs, cylinders, and bands are commonly fabricated by sealing a fine wire or foil into a nonconducting electrode body such as glass. Microlithographic techniques are perhaps the best method of producing well-defined microelectrode arrays. Other array fabrication methods include immobilizing large numbers of metal wires within a nonconducting support and electrodeposition of mercury or platinum within the pores of a polymer membrane. Spherical and hemispherical microelectrodes are typically formed by electrodepositing mercury onto platinum or iridium microdiscs.

2.2 Cell Time Constants

Every electrochemical measurement has a lower timescale limit that is imposed by the *RC* cell time constant, i.e. the product of the solution resistance, *R*, and the double layer capacitance, *C*, of the working electrode. Meaningful electrochemical data can only be extracted at timescales that

are longer than the cell time constant, typically five to ten times the *RC* time constant.[4] Therefore, an important objective when seeking to make high-speed transient measurements is to minimize the cell time constant.

As discussed in detail by Newman[5] and described in Equation (1), the solution resistance for a disc-shaped ultramicroelectrode is inversely proportional to the electrode radius,

$$R = \frac{1}{4\kappa r} \tag{1}$$

where κ is the conductivity of the solution and r is the radius of the microdisc. Equation (1) shows that R increases as the electrode radius decreases. Thus, changes in the cell resistance with decreasing electrode radius do not have the desired effect of reducing the *RC* cell time constant.

Altering the potential that is applied to an electrode causes the charge on the metal side of the interface to change and some reorganization of the ions and solvent dipoles in the double layer on the solution side of the interface will occur. This process causes electrons to flow into or out of the surface giving rise to the charging or capacitive response. The double layer capacitance for a disc-shaped ultramicroelectrode is proportional to the area of the electrode surface and is given by Equation (2),

$$C = \pi r^2 C_o \tag{2}$$

where C_o is the specific double layer capacitance of the electrode. Thus, shrinking the size of the electrode causes the interfacial capacitance to decrease with decreasing r^2.

The existence of the double layer capacitance at the working electrode complicates electrochemical measurements at short timescales. Figure 2 is an equivalent circuit of an electrochemical cell, where Z_F is the Faradaic impedance corresponding to the electrochemical reaction. In seeking to make transient measurements, the electrochemical cell must respond to the applied potential waveform much more rapidly than the process one is seeking to measure. However, the potential across a capacitor cannot be changed instantaneously and the double layer capacitance must be charged through the solution resistance in order to change the potential across the Faradaic impedance.

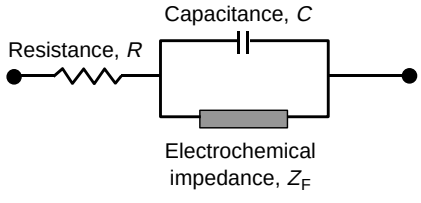

Figure 2 Equivalent circuit for an electrochemical cell and associated electrochemical process.

The time constant for this charging process is given by Equation (3):

$$RC = \frac{\pi r C_o}{4\kappa} \tag{3}$$

and is typically hundreds of microseconds for a conventional millimeter-size electrode, placing a lower limit on the useful timescale of the order of several milliseconds. The use of ultramicroelectrodes with critical dimensions in the micrometer and even nanometer range has opened new possibilities for fast kinetic studies because of the greatly diminished capacitance of these ultrasmall probes.

For example, Figure 3 shows how the *RC* cell time constant as measured in 0.1 M HCl depends on the radius of platinum microdisc electrodes. In these experiments, the double layer charging process was monitored using chronoamperometry conducted at microsecond timescales following a potential step within the double layer region of the voltammogram. As the electrode radius decreases from 25 to 1 μm, the cell time constant decreases linearly from approximately 2 μs to 80 ns. The slope of the best fit line is consistent with Equation (3) where the double layer capacitance is about 40 μF cm^{-2}. Moreover, the intercept is approximately 4.3 ns indicating that the stray capacitance (see below) of these microelectrodes is very small. In conclusion, it is apparent that by using ultramicroelectrodes with radii of microns, cell time constants of tens of nanoseconds can be achieved.

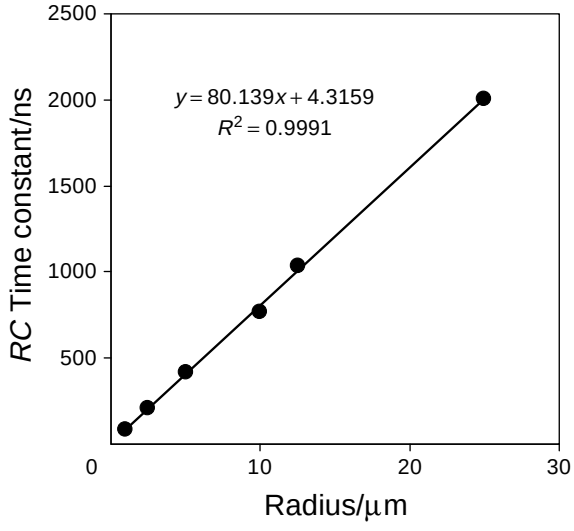

Figure 3 Relationship between the *RC* cell time constant and the radius of platinum microdiscs where the supporting electrolyte is 0.1-M HCl. Cell time constants were measured using chronoamperometry conducted on a microsecond to submicrosecond timescale by stepping the potential from 0.200 to 0.250 V vs Ag/AgCl.

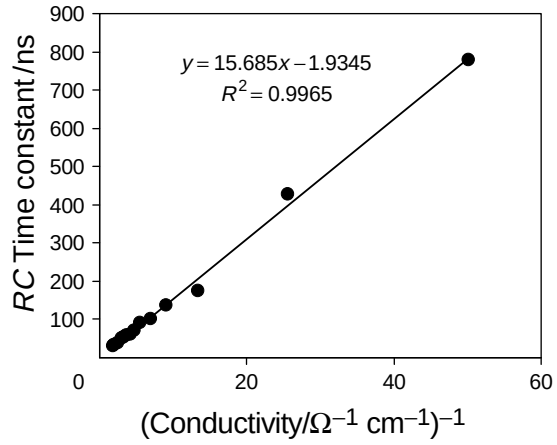

Figure 4 Relationship between the *RC* cell time constant of a 5 μm radius platinum microdisc and the reciprocal of the solution conductivity. Cell time constants were measured using chronoamperometry conducted on a microsecond to submicrosecond timescale by stepping the potential from 0.200 to 0.250 V vs Ag/AgCl.

The preceding analysis indicates that the overall cell resistance influences the cell time constant. Therefore, decreasing the resistance of the solution, through which the Faradaic and charging currents must flow, will decrease the cell time constant. Figure 4 illustrates the decrease in cell time constant that is observed for a 5 μm radius platinum microdisc as the solution conductivity is systematically varied by changing the supporting electrolyte concentration from 0.05 to 2.0 M. As predicted by Equation (3), this figure shows that a linear response is obtained, and the specific double layer capacitance is estimated as $41.7 \pm 2.1\,\mu F\,cm^{-2}$ over this range of supporting electrolyte concentrations. Significantly, Figures 3 and 4 indicate that microdisc electrodes can be manufactured that respond ideally to changes in the applied potential at timescales as short as 30 ns. Experiments performed in more conducting solutions, e.g. highly concentrated acids, indicate that *RC* cell time constants as short as 5 ns can be achieved.

However, there are a number of practical problems associated with the design and fabrication of microelectrodes that cause micrometer-sized electrodes to have *RC* time constants that greatly exceed those predicted by Equation (3).

An important cause of nonideal responses is stray capacitance within the electrochemical system that may arise from the electrode itself, the leads, or electrical connections. Stray capacitance will increase the cell time constant as described by Equation (4).

$$RC = \frac{1}{4\kappa r}(\pi r^2 C_0 + C_{\text{Stray}}) \qquad (4)$$

where C_{Stray} is the stray capacitance. Although it depends on the microelectrode design and the experimental setup, this stray capacitance is typically between a few picofarads and several tens of picofarads. The cell time constant observed in these circumstances depends strongly on the relative magnitudes of the double layer and stray capacitances. At a normal size electrode, the stray capacitance is negligible compared to the double layer capacitance and therefore does not significantly affect the observed cell time constant. However, when the magnitude of the double layer capacitance is reduced by shrinking the size of the electrode to micrometer and submicrometer dimensions, the stray and double layer capacitances can become comparable. For example, taking a typical value of $40\,\mu F\,cm^{-2}$ as the specific double layer capacitance for a platinum electrode in contact with 1 M aqueous electrolyte solution, the double layer capacitance of a 1 mm radius disc will be approximately 1 μF. This value is significantly larger than the picofarads stray capacitance found in a typical electrochemical experiment. However, for a 1 μm radius microdisc, the interfacial capacitance will decrease by six orders of magnitude to approximately 1 picofarad. Therefore, stray capacitance of even a few picofarads will cause the observed *RC* time constant to increase significantly beyond the minimum value dictated by double layer charging alone. This increased cell time constant will cause the transient response characteristics of the electrodes to become nonideal. Thus, an important objective in seeking to implement ultrafast transient techniques is to minimize the stray capacitance.

There are two major sources of stray capacitance. First, the capacitance of the cell leads and capacitive coupling between leads. Second, the microelectrode itself. By using high-quality cable of minimum length, e.g. by mounting the current-to-voltage converter directly over the electrochemical cell, and by avoiding the use of switches as far as possible, stray capacitance from the electrochemical system can be minimized. However, the importance of good microelectrode fabrication and design should not be overlooked. For example, if there is a small imperfection in the seal between the insulator and the electrode material then solution leakage will cause the *RC* cell time constant to increase massively and the Faradaic response may become obscured by charging/discharging processes. Moreover, as shown by Wightman's group[6] and Faulkner et al.,[7] using silver epoxy or mercury to make the electrical connection between the microwire and a larger hook-up wire can cause the *RC* cell time constant to increase dramatically. This increase arises because the electronically conducting mercury/glass insulator/ionically conducting solution junctions cause significant stray capacitance. It is important to note that

these effects may only become apparent in high-frequency measurements.

2.3 Ohmic Effects

When Faradaic and charging currents flow through a solution, they generate a potential that acts to weaken the applied potential by an amount iR, where i is the total current. This is an undesirable process that leads to distorted voltammetric responses. It is important to note that, as described by Equation (1), microelectrodes exhibit higher resistances than macroelectrodes because of their smaller size. However, the currents observed at microelectrodes are typically six orders of magnitude smaller than those observed at macroelectrodes. These small currents often completely eliminate ohmic drop effects even when working in organic solvents. For example, the steady-state current observed at a 5 μm radius microdisc is approximately 2 nA for a 1.0 mM solution of ferrocene. Taking a reasonable value of 0.01 $\Omega^{-1}\,cm^{-1}$ as the specific conductivity, then Equation (1) indicates that the resistance will be of the order of 50 000 Ω. This analysis suggests that the iR drop in this organic solvent is a negligible 0.09 mV. In contrast, for a conventional macroelectrode the iR drop would be of the order of 5–10 mV. Under these circumstances, distorted current responses and shifted peak potentials would be observed in cyclic voltammetry.

It is useful at this point to investigate the effect of experimental timescale on the iR drop observed at microelectrodes. The following section discusses the way that the diffusion field at microelectrodes depends on the characteristic time of the experiment. However, in general, at short times, the dominant mass transport mechanism is planar diffusion and the microelectrode behaves like a macroelectrode. Therefore, at short times the current i decreases with decreasing electrode area (r^2). Since the resistance increases with decreasing electrode radius rather than electrode area, the product iR decreases with decreasing electrode radius in short timescale experiments. Thus, beyond the reduced iR drop because of low currents, decreasing the electrode radius from say 1 mm to 10 μm decreases the ohmic iR drop observed at short times by a factor of one hundred. In contrast, at long experimental timescales the Faradaic current depends directly on the radius making the product iR independent of the electrode radius. Thus, while the low currents observed at microelectrodes reduce ohmic effects for all experimental timescales, using the transient rather than the long timescale response offers even better performance.

2.4 Mass Transport

Oxidation or reduction of a redox-active species at an electrode surface generates a concentration gradient between the interface and the bulk solution. This redox process requires electron transfer across the electrode/solution interface. The rate at which electron transfer takes place across the interface is described by the heterogeneous electron transfer rate constant, k. If this rate constant is large, then diffusional mass transport will control the current observed. Our objective is to describe how these diffusion fields evolve in time. The experiment of interest involves stepping the potential from an initial value where no electrode reaction occurs, to one where electrolysis proceeds at a diffusion controlled rate. Consider the case of a spherical electrode of radius r_s placed in a solution that contains only supporting electrolyte and a redox-active species of concentration C. The concentration gradient at the electrode surface is obtained by solving Fick's second law in spherical coordinates, Equation (5)

$$\frac{\partial C(r,t)}{\partial t} = D\ \frac{\partial^2 C(r,t)}{\partial r^2} + \frac{2}{r}\frac{\partial C(r,t)}{\partial r} \tag{5}$$

The boundary conditions for the potential step experiments described above are shown in Equation (6)

$$\lim_{r\to\infty} C(r,t) = C^{\infty}$$
$$C(r,0) = C^{\infty}$$
and $$C(r,t) = 0 \quad \text{for } t > 0 \tag{6}$$

where r is the distance from the center of the sphere, D is the diffusion coefficient for the redox active species, and C is the concentration as a function of distance r and time t.

Equation (5) can be solved using Laplace transform techniques to give the time evolution of the current, $i(t)$, subject to the boundary conditions described. Equation (7)

$$i(t) = \frac{nFADC^{\infty}}{r_s} + \frac{nFAD^{1/2}C^{\infty}}{\pi^{1/2}t^{1/2}} \tag{7}$$

is obtained where n is the number of electrons transferred in the redox reaction, F is Faraday's constant, and A is the geometric electrode area.

Equation (7) shows that the current response following a potential step contains both time-independent and time-dependent terms. The differences in the electrochemical responses observed at macroscopic and microscopic electrodes arise because of the relative importance of these terms at conventional electrochemical timescales. It is possible to distinguish two limiting regimes depending on whether the experimental timescale is short or long.

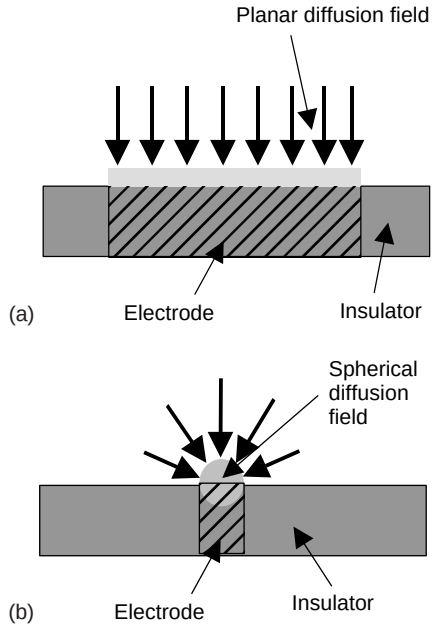

Figure 5 Diffusion fields observed at microelectrodes. (a) Linear diffusion observed at short times. (b) Radial (convergent) diffusion observed at long times.

2.4.1 Short Times

At sufficiently short times, the thickness of the diffusion layer that is depleted of reactant is much smaller than the electrode radius and the spherical electrode appears to be planar to a molecule at the edge of this diffusion layer. Under these conditions, the electrode behaves like a macroelectrode and mass transport is dominated by linear diffusion to the electrode surface as illustrated in Figure 5(a). At these short times, the $t^{-1/2}$ dependence of the second term in Equation (7) makes it significantly larger than the first term and the current response induced by the potential step initially decays in time according the Cottrell equation, Equation (8)

$$i(t) = \frac{nFAD^{1/2}C^{\infty}}{\pi^{1/2}t^{1/2}} \tag{8}$$

Figure 6 shows the relationship that exists between the range of useable scan rates and electrode radius subject to the condition that ohmic drop is negligible and that the dominant mass transport regime is linear diffusion.[8]

2.4.2 Long Times

At long times, the transient contribution given by the second term of Equation (7) has decayed to the point where its contribution to the overall current is negligible. At these long times, the spherical character of the electrode becomes important and the mass transport process is dominated by radial (spherical) diffusion as illustrated in Figure 5(b).

The current attains a time-independent steady-state value given by Equation (9)

$$i_{ss} = \frac{nFADC^{\infty}}{r_s} \tag{9}$$

The steady-state response arises because the electrolysis rate is equal to the rate at which molecules diffuse to the electrode surface.

Since *short* and *long* times are relative terms, it is useful to determine the times over which transient and steady-state behaviors will predominate and

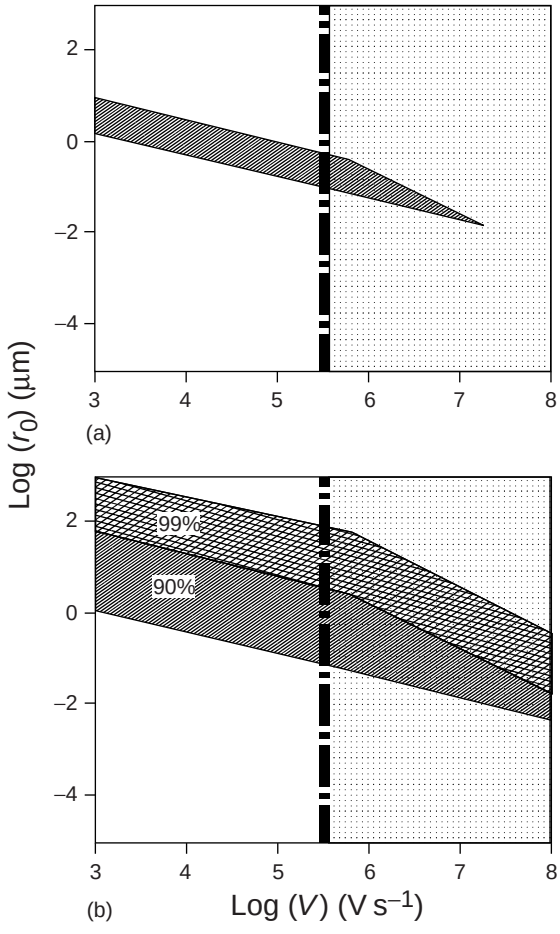

Figure 6 Theoretical limitations to ultrafast cyclic voltammetry. The shaded (or cross-hatched) area between the slanted lines represents the radius that a microdisc must have if the ohmic drop is to be less than 15 mV and distortions due to nonplanar diffusion account for less than 10% of the peak current. (a) Without *iR* drop compensation by positive feedback; (b) with 90% and 99% ohmic drop compensation. The dotted areas in (a) and (b) represent the regions where transport within the double layer affects the voltammetric response. Limits are indicative and correspond approximately to a 5-mM anthracene solution in acetonitrile, with 0.3-M tetrafluoroborate as supporting electrolyte. (Reprinted from C. Amatore, *Electrochemistry at Microelectrodes*, ed. I. Rubenstein, Chapter 4, 1995 by courtesy of Marcel Dekker, Inc.)

how this time regime is affected by the electrode radius.

This objective can be achieved by considering the ratio of the transient to steady-state current contributions (Equations (8) and (9), respectively). This analysis gives a dimensionless parameter $(\pi Dt)^{1/2}/r_s$ that can be used to calculate a lower time limit at which the steady-state contribution will dominate the total current to a specified extent. For example, the time required for the steady-state current contribution, i_{ss}, to be ten times larger than the transient component, i_t can be calculated. Taking a typical value of D as $1 \times 10^{-5}\,cm^2\,s^{-1}$ for an aqueous solution, then for an electrode of radius 5 mm, the experimental timescale must be longer than 80 s. Therefore, steady state is not observed for macroelectrodes at the tens of millivolts per second timescale typical of conventional cyclic voltammetry experiments. However, reducing the electrode radius by a factor of a thousand to 5 μm, means that a steady-state response can be observed for times longer than 80 μs. Since the steady-state current becomes more dominant with increasing time, steady-state responses are easily observed for microelectrodes in electrochemical experiments run at conventional timescales. Figure 7(a) shows the sigmoidal-shaped responses that characterize steady-state mass transfer in slow scan rate cyclic voltammetry. In contrast, at short experimental timescales (high scan rates) peaked responses (Figure 7b) similar to those observed at conventional macroelectrodes are seen.

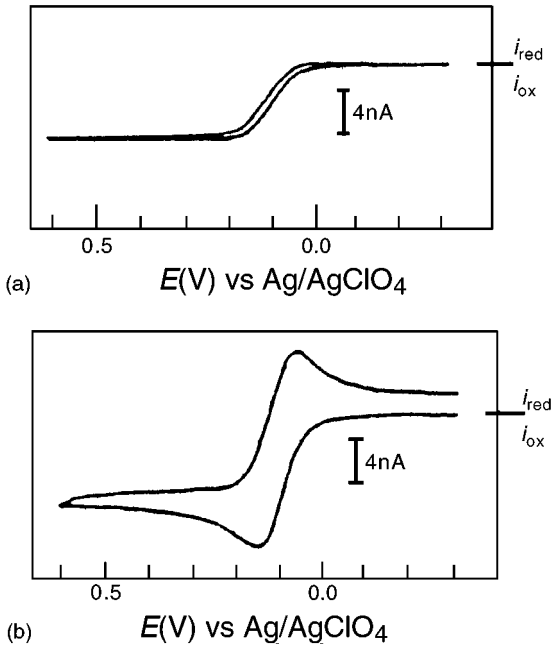

Figure 7 Effect of scan rate on the cyclic voltammetry of 1.0 mM ferrocene at a 6.5-μm gold microdisc where the supporting electrolyte is 0.1-M tetrabutyl ammonium perchlorate in acetonitrile. (a) Scan rate is $0.1\,V\,s^{-1}$; (b) scan rate is $10\,V\,s^{-1}$. (Reprinted with permission from J.O. Howell and R.M. Wightman, *Anal. Chem.*, **56**, 524. Copyright (1984) American Chemical Society.)

The preceding analysis considered a spherical electrode because its surface is uniformly accessible and a simple closed-form solution to the diffusion equation exists.[9] The microdisc is the most widely used geometry, but derivation of rigorous expressions describing their experimental response is complicated because the surface is not uniformly accessible. For discs, electrolysis at the outer circumference of the disc diminishes the flux of the electroactive material to the center of the electrode. However, microdisc and microring geometries share the advantage of spherical microelectrodes in that quasi-spherical diffusion fields are established in relatively short periods of time. The steady-state current for a disc is given by Equation (10)

$$i_{ss} = 4nFDCr, \tag{10}$$

where r is the radius of the disc.

Observing a steady-state response depends on all the electrode dimensions being small, not just the radius, and is therefore not achieved for every geometry at the timescales considered above. For example, band electrodes whose thickness is in the micrometer range, but whose length is several millimeters, do not exhibit true steady-state responses. However, a high analyte flux to the ends of the band often makes it possible to observe a pseudo-steady-state condition in a practical sense.

Radial diffusion gives very high rates of mass transport to the electrode surface with a mass transport coefficient of the order of D/r. Therefore, even at rotation rates of 10^4 rpm, convective transport to a rotating macroelectrode is smaller than diffusion to a 1 μm microdisc. The high flux at a microelectrode means that a reverse wave under steady-state conditions is not observed (Figure 5a), because the electrolysis product leaves the diffusion layer at an enhanced rate.

2.5 Instrumental Challenges

Taking typical values of 1 mM and $1 \times 10^{-5}\,cm^2\,s^{-1}$ for the concentration and diffusion coefficient, respectively, the steady-state current is of the order of 2 nA for a 5 μm radius microdisc. As described by Equation (10), the magnitude of this current will decrease with decreasing electrode radius and picoamp currents will be observed for a 10 nm radius electrode. These are certainly small currents but can be measured with relatively simple circuits based on operational amplifiers. This success can only be achieved because the measurement is performed at long timescales.

In contrast, trying to measure the currents generated at microelectrodes at short timescales poses real technological challenges. The current observed at short timescales is typically in the microamp range. For example, when the potential is stepped to a value where electrolysis proceeds at a diffusion-controlled rate, the Faradaic current observed 1 μs after the application of the potential step will be approximately 135 nA for a 5 μm radius microdisc, i.e. almost two orders of magnitude larger than the steady-state value. However, there is a difficulty in using operational amplifiers to measure these short timescale currents. For a given operational amplifier, the product of the gain and frequency bandwidth is a constant. Therefore, if a large amplification of a small current is required then the time resolution will be adversely affected and it will become increasingly more difficult to probe fast electron transfer dynamics accurately. The development of faster electronic components, notably for the capture, processing, and streaming of video images, will certainly continue to decrease the lower accessible limit. However, even today, measuring submicroamp currents with nanosecond time resolution remains a significant technological challenge.

Another approach to solving this amplification problem is to use microelectrode arrays. Arrays are ensembles of microelectrodes that may consist of regularly or irregularly spaced assemblies of identical electrodes, ensembles of electrodes with identical shape but uneven dimensions, or disordered arrangements of irregularly shaped electrodes. In seeking to make ultrafast measurements, the advantage of using a microelectrode array is that it provides a nonelectronic approach to current amplification. Often the individual elements are not independently addressable and all of the electrodes operate at the same applied potential. Under these conditions, the total response of the array depends on the relative size of the individual electrodes and the thickness of the diffusion layer that develops around each element during electrolysis. When seeking to make ultrafast measurements, the individual diffusion fields need to act independently of one another, i.e. the separation of the electrodes must be much larger than the diffusion layer thickness throughout the experiment. If the diffusion layers that develop at the individual electrodes merge, then a planar diffusion layer is created that extends over the entire surface of the array. Under these conditions, the array behaves like a large electrode whose surface area is given by the sum of the electrochemically active and inactive areas. This condition does not yield any benefit in terms of probing fast electron transfer dynamics.

Thus, arrays used for dynamic measurements typically consist of N widely separated and noninteracting discs that will provide N times the current observed at a single electrode. In terms of the RC cell time constant, such a microelectrode array can be considered as a circuit of N parallel combinations of each single microelectrode. Thus, the equivalent resistance and capacitance are R/N and $N \times C$, respectively. The product, RC, of the array remains unchanged from that found at a single element of the assembly. Microelectrode arrays appear to offer the same advantages as a single microelectrode, but it is not necessary to construct such a high-gain current-to-voltage converter, making measurements at shorter timescales possible. However, in the short timescale limit (nonoverlapping diffusion fields), the gain achieved is necessarily finite and fixed by the total number of microelectrodes in the array. In this sense, the amplification achieved is less flexible and of lower quality than is possible using electronic means. Moreover, fabrication of microelectrode arrays with electrode radii smaller than ten micrometers, in which each electrode exhibits its theoretical RC time constant, remains technologically challenging.

2.6 Transient Techniques

In transient electrochemical measurements involving a solution-phase redox couple, one seeks to create a competition between the reaction of interest, electron transfer at the electrode surface or coupled homogeneous steps, and diffusion of the species to and from the electrode surface.[10] In the following sections, consideration is given to how high diffusion rates that allow fast reactions to be investigated are obtained by controlling the experimental timescale directly, i.e. by applying a shorter and shorter time perturbation to the system.

In order to extract useful kinetic information from transient electrochemical responses, the Faradaic and charging currents must be separated. This requirement presents a number of challenges when attempting to perform experiments at very short timescales. For example, in fast scan rate cyclic voltammetry, the charging current increases proportionally to the scan rate while the Faradaic current for diffusive species is proportional to the square root of the scan rate. These dependencies cause the ratio of the Faradaic to charging current ratio to *decrease* with *increasing* scan rate. In some sense, this process corresponds to a decreasing signal-to-noise ratio since the magnitude of the charging current places a lower limit on the detectable Faradaic current. One strategy for dealing with capacitive charging in high frequency experiments is to measure the double layer charging current in blank electrolyte and then to subtract this signal from the total current observed in the presence of the analyte of interest. However, the usefulness of this approach is limited since at high scan rates the magnitudes of the charging and Faradaic currents are often similar.

While microelectrodes can significantly decrease the deleterious effects of ohmic drop, the large currents

observed in short-timescale transient experiments can result in significant *iR* drop and distorted voltammograms. Determination of heterogeneous electron transfer rate constants often relies on measuring the scan rate dependence of the peak-to-peak separation, ΔE_P. Ohmic effects represent a serious problem, not only because they will cause a significant ΔE_P to be observed even when heterogeneous electron transfer is fast, but also because the magnitude of the ohmic effect depends on the experimental timescale. Several strategies have been used to decrease ohmic effects. First, the Faradaic information is extracted by means of convolution of the voltammograms with the diffusion operator $(\pi t)^{-1/2}$. Second, the experimental voltammogram is simulated using a model that incorporates ohmic and capacitive factors. Third, corrections for ohmic drop are made on-line, using positive feedback circuitry. Savéant et al.[11] have developed high-speed potentiostats capable of providing positive feedback to compensate on-line for ohmic drop at scan rates up to $5 \times 10^5\,\mathrm{V\,s^{-1}}$.

In contrast to cyclic voltammetry, for chronoamperometric experiments the decay rates for double layer charging and Faradaic processes are different, allowing the two processes to be separated on a kinetic basis. Ideally, the double charging current undergoes a single exponential in time while the Faradaic current for a diffusive species decreases with the square root of time. Therefore, at short times the current will be dominated by the charging current while at long times the diffusion controlled Faradaic current will be relatively more significant.

2.6.1 *Heterogeneous Electron Transfer Dynamics*

As discussed above, at short times the diffusion layer thickness is much smaller than the microelectrode radius and the dominant mass transport mechanism is planar diffusion. As observed for macroelectrodes at conventional millivolt per second scan rates, the Faradaic current increases in proportion to the electrode area. Under these conditions, the classical theories, e.g. that of Nicholson and Shain,[12] can be used to extract kinetic parameters from the scan rate dependence of the separation between the anodic and cathodic peak potentials. The peak-to-peak separation increases with increasing scan rate thus improving the accuracy of the rate constant determination. However, as discussed above, stray capacitance, ohmic drop, double layer charging currents, and difficulties posed by bandwidth limitations of the instrument must be recognized. These effects can be recognized by performing measurements across the whole dynamic range available, i.e. from kilovolts per second to megavolts per second.

As shown in Table 1, the heterogeneous electron transfer dynamics of a diverse range of organic and inorganic species have been investigated using transient techniques. However, cyclic voltammetry of the anthracene/anthracene anion radical has been used extensively as a reference system for characterizing new electrodes or instruments. This focus arises because the aromatic nature of the molecule leads to very small changes in the bond lengths and angles, i.e. the Marcus inner sphere reorganization energy is small,[13] causing the heterogeneous electron transfer rate constant to be large. Howell and Wightman[14] have studied this reaction using cyclic voltammetry at scan rates up to $10^5\,\mathrm{V\,s^{-1}}$ and good agreement between the predictions of the Nicholson and Shain theory and experiment was found after correcting for the nonspherical nature of the microdisc used.

A useful strategy in trying to extend the upper limit of measurable electron transfer rate constants is to perform measurements at lower temperatures. This strategy is successful because even for heterogeneous electron transfers with negligible inner sphere reorganization energies, activation barriers of the order of $20–25\,\mathrm{kJ\,mol^{-1}}$ are expected. Therefore, considerably slower rates of heterogeneous electron transfer ought to be observed even by decreasing the temperature of the electrochemical cell by a few tens of degrees. Measurements of this type are facilitated greatly by microelectrodes, since solvents, such as alcohols or nitriles that remain liquid over a wide temperature range, can be used without catastrophic ohmic effects. For example, Weaver and coworkers investigated the ferrocene, *o*-nitrotoluene, and nitromesitylene systems in acetonitrile, propionitrile, and butyronitrile at a gold microdisc using scan rates up to $10^4\,\mathrm{V\,s^{-1}}$ between 200 and 300 K.[15] The experimental voltammograms were interpreted with the aid of simulated responses that accounted for the activation enthalpy, temperature-dependent diffusion coefficients, and double layer capacitance. For ferrocene, the standard heterogeneous electron transfer rate constants ranged from $0.083\,\mathrm{cm\,s^{-1}}$ at 198 K to approximately $5.5\,\mathrm{cm\,s^{-1}}$ at 298 K yielding an activation enthalpy of $20\,\mathrm{kJ\,mol^{-1}}$.

There is little doubt that temperature-dependent studies offer a useful new insight into the electrochemical behavior of fast redox couples especially by providing activation enthalpies, entropies, and reaction free energies that allow electronic coupling terms to be evaluated. However, as exemplified by the careful work of Evans et al.[32] and Safford and Weaver,[15] attention must be paid to two important processes. First, the rate of diffusional mass transport will decrease with decreasing temperature. This process will make the voltammetric response less sensitive to the electrode kinetics. Second, lower temperatures will cause the solution resistance to increase, which can cause an enhanced peak-to-peak separation to be observed that could be incorrectly interpreted as slow heterogeneous electron transfer.

Table 1 Rate constants for heterogeneous electron transfer as determined using transient methods

Analyte	Electrode[a]; Element, Dimension (μm)	k^o(s^{-1})	Refs.
Anthracene	Au, 6.5	3.46 ± 0.55	14
	Au, 5	3.3	16
	Au, 5	2.6	17
	Au, 3.0 and 8.5	3–4.8	18, 19
9-Bromoanthracene	Au, 5	2.9 ± 0.3	31
	Au, 6	10	20
Anthraquinone	Pt cylinder, $r = 25.4\,\mu m$, $\ell < 0.25\,cm$	1.78 ± 0.35	14
	Pt, 5	1.5	17
	Au, 6.5	2.4	21
Benzoquinone	Au, 6.5	0.39 ± 0.1	14
	Pt, 5	0.14	17
Ferrocene	Pt, 5	1.1	17
	Au, 5	3.1 ± 1.1	22
	Pt, 5–25	1.4–3.6	21
	Au, 5	5.5	15
	Pt, 10		23
Ferrocyanide	C cylinder, $r = 15\,\mu m$, $\ell = 500\,\mu m$	0.0114 ± 0.0022	24
	Pt, 10, 50	0.42 ± 0.03	25
	C, 5	0.06 ± 0.05	26
	Pt, 1–30	0.64–0.79	27
	Pt, 20	0.2	28
9-Fluorenone	Pt, 6	3	29
10-Methylacridan cation radical	Au, Pt, 8.5, 10	3.2 ± 0.5	30
$[Ru(bpy)_3]^{2+}$	Hg, 5.5	0.45	22
	Au, 5	2.5	22

[a] Dimension given is the radius of a microdisc electrode unless otherwise stated.

Irrespective of whether steady-state or transient approaches are used, the electrode surface activation, pretreatment, roughness, degree of sample adsorption, and electrode material as well as the nature and concentration of the supporting electrolyte, sample concentration, and solvent type and purity, can have a profound effect on the magnitude of the heterogeneous electron transfer rate observed.

Despite the many elegant investigations that have been conducted into heterogeneous electron transfer dynamics using ultrafast electrochemical techniques, the magnitude of the diffusion-controlled current at short times ultimately places a lower limit on the accessible timescale. For diffusive species, the thickness of the diffusion layer, δ, is defined as $\delta = (\pi Dt)^{1/2}$ and is therefore proportional to the square root of the polarization time, t. The diffusion layer thickness can be estimated to be approximately 50 Å if the diffusion coefficient is $1 \times 10^{-5}\,cm^2\,s^{-1}$ and the polarization time is 10 ns. Given a typical bulk concentration of the electroactive species of 1 mM, this analysis reveals that only 10 000 molecules would be oxidized or reduced at a 1 μm radius microdisc under these conditions. The average current for this experiment is only 170 nA, which is too small to be detected with low-nanosecond time resolution.

Therefore, in order to probe the dynamics and energetics of ultrafast heterogeneous electron transfer dynamics this diffusion limitation must be eliminated. One successful approach to achieving this objective is to use self-assembled or spontaneously adsorbed monolayers. When immobilized on an electrode surface the electroactive species no longer needs to diffuse to the electrode to undergo electron transfer. Moreover, the electroactive species is preconcentrated on the electrode surface. For example, in the situation considered above, there will be approximately 1.7×10^{-20} mol of electroactive material within the diffusion layer. Given that the area of a 1 μm disc is approximately $3.1 \times 10^{-8}\,cm^2$, this translates into an "equivalent surface coverage" of about $5.4 \times 10^{-13}\,mol\,cm^{-2}$. In contrast, the surface coverage, Γ, observed for dense monolayers of adsorbates is typically more than two orders of magnitude larger with coverages of the order of $10^{-10}\,mol\,cm^{-2}$ being observed. This higher concentration gives rise to much larger currents that are easier to detect at short timescales.

For example, Faulkner et al. have investigated the dynamics and energetics of electron transfer across mercury/anthraquinone monolayer interfaces.[33] The surface coverage of an anthraquinone-2,6-disulfonic acid (2,6-AQDS) monolayer formed from a solution containing micromolar concentrations of the quinone was approximately 1.6×10^{-10} mol cm^{-2}. Therefore, in transient electrochemical experiments performed on nanosecond timescales, the average Faradaic current would be expected to be of the order of 50 μA. It is technologically feasible to detect currents of this magnitude with circuitry based on existing wide bandwidth operational amplifiers.

As exemplified by the work of Chidsey,[34] Acevedo and Abruña,[35] Forster and Faulkner,[36] and Finklea and Hanshew,[37] electroactive adsorbed monolayers have been developed that exhibit close to ideal reversible electrochemical behavior under a wide variety of experimental conditions of timescale, temperature, solvent and electrolyte. In order for a monolayer to be an attractive model system for understanding those factors that control the rate of heterogeneous electron transfer, it should have a number of properties. The nature and physical location of the electroactive center within the molecule should be well defined. The electron transfer should be mechanistically simple with the electrochemical responses being close to ideally reversible and stable over prolonged periods. The molecules should adsorb strongly onto electrode surfaces in both oxidized and reduced forms. It should be possible to control the surface coverage by controlling the concentration of the adsorbate in solution (reversible physisorption), by controlling the time that the pristine electrode is exposed to the deposition solution, by using electroinactive diluents, or by controlled desorption from a high-coverage monolayer. The layers should be stable in a wide variety of solvents, especially water, to allow the effect of solvent parameters, e.g. relaxation rate and dielectric constant, on simple electron transfer processes to be investigated.

In these respects, monolayers containing ferrocene or ruthenium hexamine are attractive model systems for studying heterogeneous electron transfer across electrode/monolayer interfaces. However, in many of these studies the bridging ligand linking the electroactive group to the electrode surface has been carefully chosen with regard to its length and electronic structure so that electron transfer occurs on a millisecond or longer timescale. Therefore, ultrafast electrochemical techniques are not required to probe the details of electron transfer. In contrast, species that are directly adsorbed onto the electrode surface (e.g. anthraquinones) or where the bridging ligand is shorter (e.g. 4,4′-dipyridyl-type linkers) can undergo electron transfer on the submicrosecond timescale. For example, Faulkner et al.[33] found that 2,6-AQDS monolayers can be reversibly converted to the hydroquinone form in a two-electron, two-proton mechanism and virtually ideal cyclic voltammetry and chronoamperometry are observed. The standard heterogeneous electron transfer rate constant is in the range of 10^4 to 10^5 s^{-1} depending on the driving force (overpotential, $\eta \equiv E_{app} - E^{o'}$) for the reaction and the proton concentration.

Acevedo and Abruña[35] formed spontaneously adsorbed monolayers of the complex $[Os(bpy)_2\ Cl\ (dipy)]^+$, where dipy is 4,4′-trimethylenedipyridine, and investigated their dynamic and energetic properties. The voltammetric responses observed for the $Os^{2+/3+}$ redox reaction at scan rates over 8000 V s^{-1} were close to the behavior expected for an ideal reversible one-electron transfer reaction involving a surface-confined species. They used Laviron's formalism[38] describing the variation of the peak potential with scan rate to estimate the rate of heterogeneous electron transfer as 2×10^5 s^{-1}. The formation and structure of these monolayers on Pt(111) surfaces was investigated using scanning tunneling microscopy (STM) and electrochemical scanning tunneling microscopy (ECSTM). Molecularly resolved images were obtained that demonstrated that the monolayer exists as a tightly packed two-dimensional crystal on the electrode surface. Abruña also demonstrated that the exchange dynamics associated with the desorption and displacement of these adsorbates are controlled by the rate of desorption via a dissociative mechanism and that the applied potential does not significantly affect the free energy of adsorption.

The ideality of the electrochemical response exhibited by these osmium-containing monolayers makes them attractive model systems for probing the effect of solvent, electrolyte, and temperature on the electron transfer dynamics. Faulkner and Forster[36] investigated these effects for $[Os(bpy)_2\ Cl\ (pNp)]^+$, where pNp is 4,4′-dipyridyl, 1,2-bis(4-pyridyl)ethane, or dipy, monolayers on platinum microelectrodes. The rate of heterogeneous electron transfer was measured using chronoamperometry conducted on a low microsecond timescale.

For the $Os^{2+/3+}$ redox reaction, heterogeneous electron transfer is a rapid first-order process characterized by a single unimolecular rate constant (k/s^{-1}). Tafel plots of the dependence of ln k on the electrochemical driving force (overpotential) show curvature, indicating that the transfer coefficient is potential-dependent. For sufficiently large overpotentials, k tends to become independent of the free energy driving force, which is consistent with Marcus theory. The response is asymmetric with respect to overpotential, with the slope for the oxidation process tending towards zero more rapidly than that for the reduction process. This response was modeled as a tunneling process between electronic manifolds on the two sides of the interface. Temperature-resolved measurements of k and the formal potential were made

from −5 to +40 °C to provide enthalpies and entropies of activation, respectively. The corresponding free energies of activation ranged from 12.3 kJ mol^{-1} in acetonitrile to 6.4 kJ mol^{-1} in chloroform. There is weak coupling between the electronic manifolds on the two sides of the electrochemical interface. Surprisingly however, k depends linearly on the longitudinal relaxation rate of the solvent. This behavior is not predicted on the basis of contemporary theories of electron transfer. These osmium- and related ruthenium-containing monolayers have been used to investigate interfacial field effects on reductive chloride elimination and as model systems for investigating how Faradaic responses that occur at identical formal potentials can be separated on a kinetic basis.[39]

More recently, Forster has probed the dynamics of both metal-centered oxidation and ligand-based reduction processes to address how electronic states of the bridging ligand contribute to the electron tunneling pathway.[40] This work indicates that the close proximity of the redox potentials of the bridge and the remote bipyridyl ligands leads to stronger electronic coupling between the electrode and bipyridyl ligands than between the electrode and the osmium center. Forster and O'Kelly also investigated monolayers that incorporated a ligand capable of undergoing protonation/deprotonation reactions within the coordination shell of the metal center.[41] This study showed that the rate of heterogeneous electron transfer across the metal/monolayer interface depends on the pH of the contacting electrolyte. From a molecular electronics perspective, pH-induced *conformational gating* of this type offers the possibility of developing pH-triggered electrical switches.

There are few literature reports addressing the issue of separating charging and Faradaic currents in chronoamperometric experiments involving surface-confined reactants. Separating these contributions is complicated by the fact that both are expected to exhibit single exponential decays in time. Measuring the capacitive response in blank supporting electrolyte is not a useful strategy since the double-layer capacitance is altered by adsorption. Moreover, the interfacial capacitance depends on the redox composition of the monolayer. Much of the work discussed above relies on a kinetic separation of these currents, i.e. a microelectrode with an RC cell time constant at least five times smaller than the redox process under investigation is used causing the short and long timescale currents to be dominated by double-layer charging and the Faradaic reactions, respectively. However, approaches based on digital simulation and Laplace transform methods have been developed.[42]

2.6.2 Homogeneous Chemical Kinetics

There are two important advantages in using micro- rather than macroelectrodes for probing the dynamics of homogeneous chemical reactions. First, because of their relative immunity to ohmic effects, microelectrodes have greatly extended the range of useful media, e.g. low-dielectric-constant solvents, solids, and solutions with no deliberately added electrolyte. This capability often makes it feasible to compare electrochemical rate constants directly with those obtained using other techniques, notably transient absorption or luminescence spectroscopy. Second, very high diffusion rates and high-quality data can be obtained at short experimental timescales. These decreases in the lower accessible time limit have important implications for probing the dynamics of rapid heterogeneous electron transfer and homogeneous chemical reactions alike. For example, bimolecular reactions in solution cannot proceed faster than the rate at which molecules come into close contact. Thus, bimolecular rate constants cannot exceed the diffusion-limited rate constant that is of the order of 10^9–10^{10} M^{-1} s^{-1} in most organic solvents. Since the characteristic time of cyclic voltammetry is RT/Fv, where v is the scan rate, experiments performed at mV s^{-1} scan rates allow kinetic information, such as lifetimes that are close to the diffusion limit, to be obtained.[43]

In the past, when the rate constant for the following reaction was large, the voltammetric response was irreversible because the intermediate underwent a following chemical reaction before the direction of the potential scan was reversed. With the development of microelectrodes and high-scan-rate voltammetry, it is now possible to detect the fast-decaying intermediate. In many circumstances, it is even possible to observe reversible electrochemical responses since the experimental timescale can be made shorter than the lifetime of the electrogenerated reactant. For example, as illustrated in Figure 8, Howell and Wightman[44] have shown that the irreversible response observed for the oxidation of anthracene at slow scan rates becomes fully reversible at a scan rate of 10^4 V s^{-1}. This behavior is opposite to that expected when heterogeneous electron transfer is slow and suggests that the cation radical undergoes a following chemical reaction. The ability to make the voltammetric response reversible means that the formal potentials of highly reactive species can be measured accurately. However, it is important that ohmic drop is negligible, or is made so using on-line current interrupt or positive feedback approaches, otherwise the experimental responses have to be modeled if true thermodynamic information is to be obtained.

Beyond this thermodynamic information, transient measurements can be used to probe the mechanism of the following chemical reaction. For example, the oxidation of anthracene has been reported widely as proceeding through an ECE mechanism, where E denotes an electron transfer step and C a chemical reaction. However,

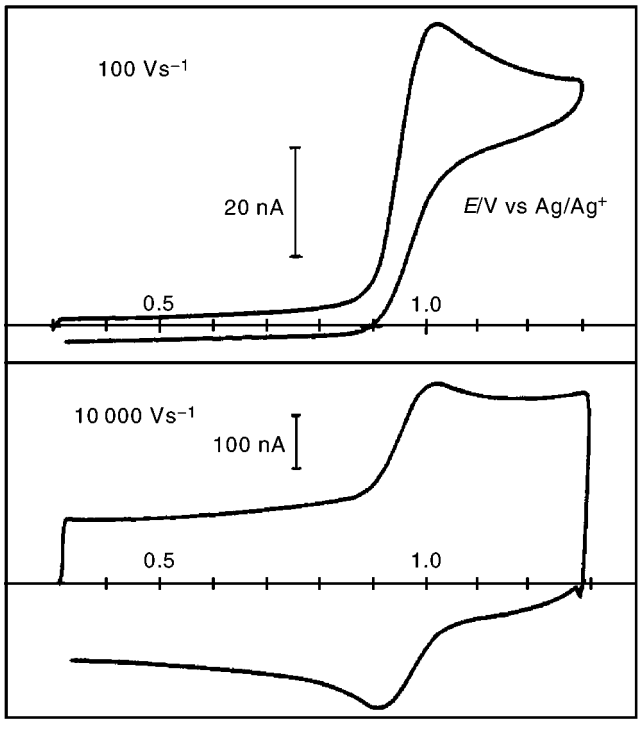

Figure 8 Effect of scan rate on the voltammetric response observed at a 5-μm platinum microdisc for the oxidation of 2.36-mM anthracene in dimethylformamide (DMF) containing 0.6-M tetraethyl ammonium perchlorate. (Reprinted with permission from J.O. Howell and R.M. Wightman, *J. Phys. Chem.*, **88**, 3915. Copyright (1984) American Chemical Society.)

Osteryoung et al.[45] report that closer agreement is found between staircase, square wave, normal pulse, and reverse pulse responses and a DISP1 mechanism, where DISP denotes a disproportionation (DISP) reaction. First-order rate constants range from 310–750 s^{-1}. However, for true first-order kinetics, the same rate constant ought to be obtained irrespective of the time at which the transient response is analyzed. In the case of anthracene oxidation, smaller rate constants are observed at long times. This behavior may arise because of low concentrations of reactive impurities in the sample that compete with the main reaction at long times.

As indicated in Table 2, the kinetics of many following chemical reactions have been investigated using transient methods at microelectrodes including the oxidation of polyalkylbenzenes, anthracene, aromatic hydrocarbons, and ascorbic acid; the reduction of quinones, aryl halides, chloroquinoline, acetophenones, and butylpyridinium derivatives; and the initial stages of electropolymerization of the conducting polymer polypyrrole. The shortest lifetime attainable is currently in the tens of nanoseconds range, e.g. the lifetime for reductive dimerization of the 2,6-diphenylpyrillium cation in acetonitrile has been determined to be 50 ns.[43]

The ability of individual solvents to stabilize transition states can have a profound effect on the observed rate constants. For example, Wipf and Wightman[52] studied the cleavage reaction involved in the reduction of 9-chloroanthracene in low dielectric solvents including tetrahydrofuran, dichloromethane, and chlorobenzene. These measurements would simply be impossible using conventional macroelectrodes. The transition state in this reduction involves transfer of the negative charge of the anion radical to the chlorine atom. The localized negative charge can be stabilized by the solvent thus lowering the energy of the transition state and yielding a more rapid cleavage. Measurements of this type not only provide fundamental mechanistic information but also play an important role in seeking to optimize industrial processes.

Table 2 Rate constants for homogeneous chemical reactions as determined using transient methods

Analyte	Electrode[a]; element, dimension (μm)	k	Refs.
Anthracene oxidation	Pt, 5	$7.6 \pm 1.0 \times 10^3$ s^{-1}	44, 46
Ascorbic acid oxidation	Hg, 5	1.4×10^3 s^{-1}	47
3-Bromoacetophenone reduction	Au, 5	6.3×10^4 s^{-1}	31
	Au, 5	8×10^3 s^{-1}	20
Acetophenone oxidation	Au, 5	6.3×10^4 s^{-1}	31
Chlorpromazine + dopamine	Pt, 25	10^8 M^{-1} s^{-1}	48
N,*N*-dimethylaniline oxidation	Pt, 12.5	6.3×10^5 M^{-1} s^{-1}	49
$Fe(CO)_5 + CO$	Au, 100	$6 \pm 2 \times 10^6$ M^{-1} s^{-1}	50
Methylbenzenes	Pt, 5	10^9 M^{-1} s^{-1}	51

[a] Dimension given is the radius of a microdisc electrode unless otherwise stated.

Bond formation is also amenable to investigation using ultrafast transient electrochemical techniques. For example, dimerization reactions triggered by electron transfer have been investigated using both cyclic voltammetry and potential-step methods. Andrieux, Hapiot, and Savéant[31] have made seminal contributions in this area. For example, using chronoamperometry conducted on a microsecond timescale, these authors showed that reduction of 1-methyl-4-*tert*-butylpyridinium cations in DMF leads to radical–radical coupling at a rate of $6.4 \pm 0.4 \times 10^6\ M^{-1}\ s^{-1}$. Their method of analysis involves stepping the potential to a value that is sufficiently negative of the formal potential for reduction to proceed at a diffusion-controlled rate. At a transition time, τ, the current is measured and the potential stepped back to its original value. The current is measured again at 2τ. The is essentially a generator–collector experiment and, as shown in Equation (11), the ratio of the currents obtained at τ and 2τ are normalized relative to the values that would be obtained in the absence of a following chemical reaction

$$R = \frac{I(2\tau)/I(\tau)}{[I(2\tau)/I(\tau)]_{dif}} \tag{11}$$

Figure 9 shows how R depends on the characteristic time of the experiment, τ, and the concentration of 1-methyl-4-*tert*-butylpyridinium cation. By combining cyclic voltammetric and potential-step techniques, Savéant has studied more complicated chemical systems such as the methylacridan (AH)/10-methylacridinium (A^+) couple in acetonitrile, which is an analogue of the coenzyme NADH/NAD^+ couple.[30] The formal potentials and the heterogeneous electron transfer rates, as well as the reaction rates of the following chemical reactions, were determined. These investigations revealed that oxidation of AH, as well as reduction of A^+, were kinetically controlled by following homogeneous chemical steps, rather than by the initial electron transfer, which appears to be quite fast in both cases.

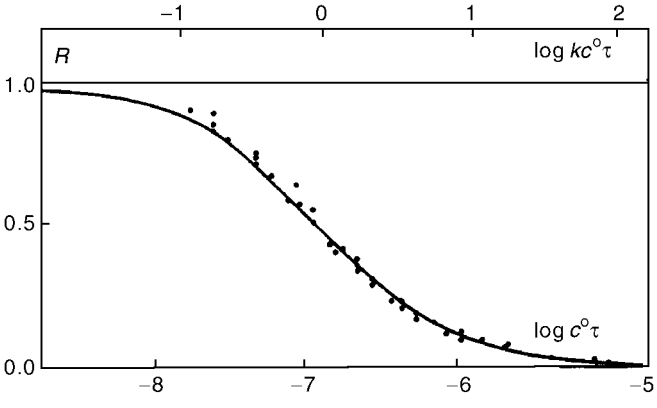

Figure 9 Variation of R with τ in double potential step chronoamperometry of 1-methyl-4-*tert*-butylpyridinium cation at concentrations of 2.5, 5, and 10 mM at 2.5- and 8.5-μm radius gold disc electrodes. (Reprinted with permission from C.P. Andrieux, P. Hapiot, J.-M. Savéant, *J. Phys. Chem.*, **92**, 5992. Copyright (1988) American Chemical Society.)

2.7 Steady-state Electrochemistry

In the preceding sections, we have focused on the use of transient methods to address short timescales in electrochemistry. However, in an electrochemical measurement involving a solution-phase reactant, the appropriate timescale is not necessarily the actual duration of the experiment. Typically, the timescale is imposed by the time taken for diffusion, t_{dif}, of the analyte of interest from the bulk solution to the electrode surface or vice versa. In seeking to measure electrochemical kinetics, one seeks to match this diffusion time to the half-life of the chemical process (homogeneous reaction dynamics) or of the interfacial electron transfer process (heterogeneous electron transfer dynamics). In transient methods, t_{dif} is directly related to the duration of the experiment. However, under steady-state conditions, t_{dif} is not related to the actual duration of the experiment since the diffusion-layer thickness depends only on the radius. Under these conditions, t_{dif} is approximately equal to r^2/D. This property has a profound impact on the size of microelectrodes required to make measurements at steady state, e.g. given a typical diffusion coefficient of $10^{-5}\ cm^2\ s^{-1}$, microelectrodes with radii of less than 30 nm are required to address submicrosecond timescales. Thus, it is clear that probing ultrafast electrochemical processes demands the fabrication of vanishingly small electrodes. However, although specialized, this is not an impossible task and well-characterized band electrodes with widths of 2 nm have been fabricated by sealing a deposited metal film between insulators.

In contrast to transient techniques, a steady-state response is unaffected by charging currents, is insensitive to low levels of reactant adsorption, and requires less complex instrumentation. The choice between transient or steady-state approaches centers on the need to produce a relatively larger, but ideally responding, electrode for transient measurements, e.g. a 25-μm radius microelectrode will exhibit a submicrosecond RC time constant in high concentrations of supporting electrolyte, compared to the need to fabricate nanometer or even angstrom dimension electrodes for use in a steady-state approach. However, the ability of transient reversal techniques, such as double step chronoamperometry or cyclic voltammetry, to create and subsequently to detect electrochemical intermediates, represents a significant advantage when attempting to characterize reactive intermediates or elucidate reaction mechanisms.

Perhaps the most convenient way to measure the steady-state current is to apply a potential sweep to the microelectrode. The appropriate timescale can be determined as outlined in section 2.4 and this time can then be converted into a scan rate by noting that the characteristic time of cyclic voltammetry is RT/Fv. It is important to stress that the magnitude of the steady-state current is intrinsic to the size of the microelectrode. It is independent of the technique (e.g. slow-scan voltammetry or long-timescale potential step) and of the direction of the potential scan or step. Moreover, at steady state all concentrations in the vicinity of the electrode surface are unchanging over time.

Figure 10 shows that while reversible, quasireversible, and irreversible systems all exhibit sigmoidal-shaped current responses and the same limiting currents in slow-scan voltammetric experiments, the experimental formal potential shifts from its thermodynamic value. This shift arises from competition between kinetic and diffusion control, with less steep curves being obtained for quasi-reversible and irreversible systems. Quasi-reversibility is observed if the kinetic distance of the reaction, D/k^o, where k^o is the standard heterogeneous electron transfer rate constant, sufficiently exceeds the dimension of the microdisc (e.g. $D/k^o = 10^4\, r$). Under these conditions, the difference between the operational and thermodynamic formal potentials can be used to determine k^o without any interference from the charging current. Reversible responses are observed when D/k^o is much smaller than the radius of the microdisc.

This analysis places upper limits on the magnitude of k^o that can be determined using a particular radius microdisc under steady-state conditions. For example, for a 1-μm radius microdisc, the upper limit on the measurable k^o is approximately $10^{-5}\,\mathrm{cm\,s^{-1}}$. Using microdiscs of 50 nm, which are about the smallest microdiscs currently routinely available, the upper limit can be extended to about $2 \times 10^{-4}\,\mathrm{cm\,s^{-1}}$. However, in practice this upper limit cannot be extended indefinitely by shrinking the electrode. Imperfect seals, often caused by mismatching of the thermal expansion coefficients of the glass shroud and the metallic microwire, often enhance non-Faradaic currents.

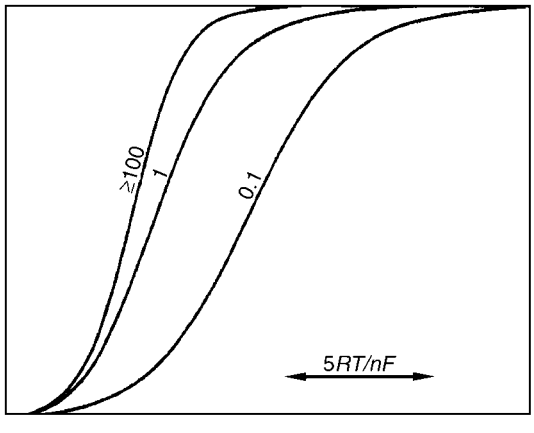

Figure 10 Shapes of the reversible, quasi-reversible, and irreversible steady-state voltammograms. The value of $k^o r/D$ is given on each curve. (Reproduced with kind permission of Kluwer Academic Publishers, Dordrecht, The Netherlands from K.B. Oldham *Microelectrodes: Theory and Applications, Steady-state Voltammetry*, eds. M.I. Montenegro, M.A. Queirós, J.L. Daschbach, NATO ASI Series E, Vol. 197, Figure 6, 1991.)

Another important consideration when trying to fabricate increasingly smaller microelectrodes is the formation of a microcavity around the electrode tip. Creating a slightly recessed electrode (e.g. because a gold microwire is used that is much softer than the insulating glass shroud), may not significantly compromise the response observed for a 10 μm radius microdisc. However, such a cavity can become very important if the electrode size is reduced to nanometer dimensions, causing distortions that are attributed to kinetic effects to become significant. Moreover, the use of an ion-exchange material as an insulator may lead to a double layer being set up on the shroud, generating an electrical potential that would compromise the measurement of the interfacial kinetics. However, these effects are likely to be important only when the Debye length of the electrolyte (typically 10–100 Å) is larger than the electrode radius. A more important effect arises from surface contamination by impurities found within the analyte, solvent, or supporting electrolyte. It is important therefore to maintain high levels of cell cleanliness and to confirm the pristine state of the electrode surface by measuring the kinetics of a reference system before and after a new analyte is investigated.

2.7.1 Heterogeneous Electron Transfer Dynamics

Understanding the elementary steps involved in heterogeneous electron transfer across a metal/solution interface is of considerable technological importance in areas ranging from battery performance to corrosion inhibition.[53] In trying to use steady-state methods to determine kinetic parameters for fast reactions, experimental conditions are usually chosen so that there is mixed control by kinetics and diffusion. Aoki et al.,[54] Fleischmann et al.,[55] and Oldham and Zoski[56] have addressed this difficult issue and provided several equivalent approaches for analyzing the experimental current–voltage curves. More recently, Mirkin and Bard[57] developed a new approach to determining kinetic parameters for a simple quasi-reversible electron transfer reaction. In principle, only the one-quarter ($E_{1/4}$), one-half ($E_{1/2}$), and three-quarters ($E_{3/4}$) potentials from a single steady-state voltammogram are required, although more precise values of k^o and α, the

Table 3 Rate constants for heterogeneous electron transfer determined using steady-state methods

Analyte	Electrode[a]; element, dimension (μm)	k^o (cm s^{-1})	Ref.
Anthracene	Au ring $\Delta r = 0.09\,\mu m$ $r = 5\,\mu m$	3.33 ± 0.05	58
(C_6H_6) Cr $(CO)_3^+$	Pt, 25	≥ 0.3	59
$(CpCOOCH_3)_2C_o^+$	Pt, 23 Å to 4.7 μm	130 ± 70	60
Cytochrome *c*	C, 6.3	>0.4	61
9,10-Diphenylanthracene	Au ring $\Delta r = 90\,\mu m$ $r = 5\,mm$	5.7 ± 0.1	58
Ferrocene	Au ring $\Delta r = 0.09\,\mu m$ $r = 5\,\mu m$	0.09 ± 0.005	58
	C, 6	2.3 ± 0.8	62
	Pt, 0.3–25	≥ 6	63
	Pt, 1	>2	64
	Pt, 16 Å to 2.6 μm	220 ± 120	60
Fe(OEP)(N-MeIM)$_2^+$	Pt, 1–25	0.4	65
	Pt, 0.5–12.5	0.38	66
	Pt, 1–25	0.35	65
Fe(TPP) $(HIM)_2^+$	Pt, 1–25	0.5	65
Fe (TPP) py_2^+	Pt, 1–25	0.6	65
$MV^{2+(F)}$	Pt, 22 Å to 0.21 μm	170 ± 90	60
Naphthalene	Au ring $\Delta r = 0.2\,\mu m$ $r = 20.5\,\mu m$	0.88 ± 0.02	58
Oxygen	Pt, 12.7–250	0.63 ± 0.05	67
Ru $(NH_3)_6^{3+}$	Au, 5	0.076	68
	Pt, 1.3–4.6	0.26 ± 0.13	60
	Pt, 11 Å to 11.1 μm	79 ± 44	60
$[Ru(bpy)_3]^{2+}$	Pt, 11.2		69
Tetracyanoethylene	Au ring $\Delta r = 0.2\,\mu m$ $r = 20.5\,\mu m$	0.15 ± 0.01	58
Tetracyanoquinodiethane	Au ring $\Delta r = 0.2\,\mu m$ $r = 20.5\,\mu m$	0.23 ± 0.01	58
Zn(TPP)	Pt, 1 to 25 μm	>1	65

[a] Dimension given is the radius of a microdisc electrode unless otherwise stated. OEP, octa-ethyl-porphyrin; MeIM, methyl imidazole; TPP, tetra-phenyl porphyrin; HIM, imidazole; MV, methyl viologen.

transfer coefficient, can be obtained by fitting the full voltammogram. Moreover, the analysis is independent of the electrode area and the concentration of the electroactive species that improves the reliability of the analysis.

The parameters describing the kinetics of heterogeneous electron transfer for several solution-phase redox-active molecules have been reported and a representative sample is given in Table 3. As expected, these data show that the rate of electron transfer across the electrode/solution interface is significantly influenced by the identity of the redox couple. These variations reflect differences in the reaction adiabaticity and activation barriers that exist for the individual systems. However, even for a single species, a considerable range of k^o values are observed, e.g. for the archetypal reversible couple ferrocene, reported k^o values range from 0.09 to 220 cm s^{-1}. There are numerous reasons for these discrepancies, including the electrode preparation regime, the purity of the ferrocene, solvent and electrolyte, and the method used to analyze the data. Moreover, for high ferrocene concentrations the electrode becomes pacified by an insoluble layer. This blocking of the electrode surface impedes diffusion and prevents the current observed from increasing proportionally with increasing ferrocene concentration. Thus, even for the standard electrochemical test system considerable care must be taken over all aspects of the experiment, chemicals, electrode preparation, experimental setup, as well as data collection and analysis if accurate kinetic parameters are to be obtained.

2.7.2 Homogeneous Chemical Kinetics

A second important application of steady-state measurements is in studies of chemical reactivity. Steady-state measurements using electrodes of different radii can provide a powerful insight into the kinetics of homogeneous reactions where the limiting current density depends on the magnitude of the homogeneous rate constant. Hence, as described in Table 4, coupled chemical (C) and electron transfer (E) reactions (e.g. EC mechanisms), catalytic follow-up processes, as well as reactions involving DISP have been characterized. It is important to note that reactions, such as chemical reactions that follow electron transfer mechanisms, cannot be investigated in the same way since the current density is not influenced by the following chemical reaction. In these circumstances, the

Table 4 Rate constants for homogeneous chemical reactions determined using steady-state methods

Analyte	Electrode[a]; element, dimension (μm)	k	Refs.
Anthracene oxidation	Pt, 0.3–62.5	$190 \pm 50\,s^{-1}$	70
Ascorbic acid oxidation at a Prussian blue film	Pt, 2.5 to 25	$1.3 \times 10^5\,M^{-1}\,s^{-1}$	71
Ferrocyanide oxidation in the presence of ascorbic acid	Pt band pair, gap = 2–12	$27 \pm 4\,M^{-1}\,s^{-1}$	72
9,10-Diphenylanthracene + 4,4-dibromodiphenyl	C, 6–9	$3.9 \pm 0.6\,M^{-1}\,s^{-1}$	73
Anion radicals + alkyl halides	Pt, 0.25	9×10^{-4} to 1.7×10^4	74, 75
$[Fe(CN)_6]^{3-}$ + aminopyridine	Pt, 0.3–25	$3.0 \pm 0.6 \times 10^3\,M^{-1}\,s^{-1}$	76
	Pt, 2.5–432	$1.8 \times 10^3\,M^{-1}\,s^{-1}$	77
	Pt band pair, gap = 2–12	$8 \pm 1 \times 10^2\,M^{-1}\,s^{-1}$	72
Hexamethylbenzene oxidation	Pt, 0.3–25	$720 \pm 100\,s^{-1}$	70
H_2 evolution on Pt from acetic acid solution	Pt, 0.3–25	$4.1 \times 10^{10}\,M^{-1}\,s^{-1}$	76
1-Napthylamine oxidation	Pt, 0.5–12.5	$4.1 \times 10^3\,s^{-1}$	78
Triphenylamine oxidation	Pt, 0.3–20	$> 3 \times 10^4\,M^{-1}\,s^{-1}$	70
Thioselenanthrene		$8.87 \pm 1.1\,s^{-1}$	79
Dibenzo-1,2-diselenine		$20.7 \pm 2.8\,s^{-1}$	80

[a] Dimension given is the radius of a microdisc electrode unless otherwise stated.

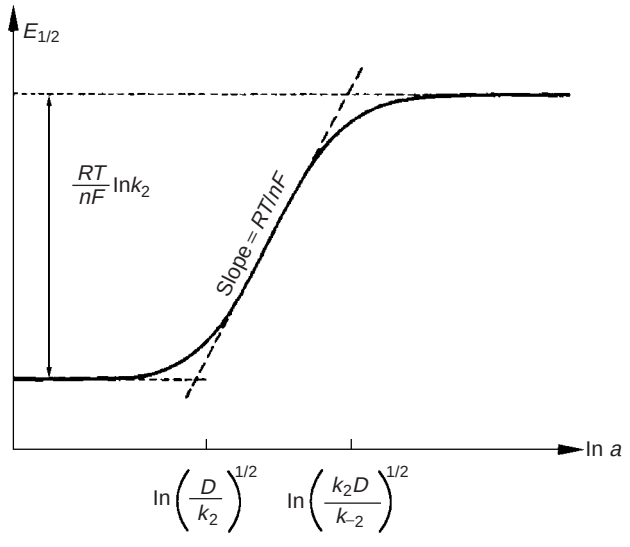

Figure 11 Dependence of $E_{1/2}$ for an electron transfer reaction on the logarithm of the electrode radius. (Reprinted from K.B. Oldham, *J. Electroanal. Chem.*, **313**, 3. Copyright (1991) with the permission of Elsevier Science.)

homogeneous reaction does not affect the height or shape of the reversible steady-state voltammogram. However, as indicated by Figure 11, the position of the wave on the potential axis depends on the homogeneous reaction rate, and kinetic information can be obtained by probing how $E_{1/2}$ depends on the electrode radius.

The technique is best applied to reactions whose rate constants are of the same magnitude as D/r^2. Therefore, given that it is now feasible to fabricate microelectrodes with submicrometer critical dimensions, the dynamics of first-order chemical reactions with rate constants between 10^2 and $10^4\,s^{-1}$ can be investigated.

3 INDIRECT METHODS

As discussed above, with direct electrochemical measurements a competition is created between the reaction of interest and diffusional mass transport either by controlling the time over which the system is electrochemically perturbed (transient methods) or by controlling the electrode size (steady-state methods). In both approaches, diffusion provides a time reference for the measurement. In contrast, for indirect methods, the reaction of interest competes with another chemical reaction whose rate constant is known from independent measurements. Thus, this second reaction acts as a kinetic reference instead of the diffusion to or from the electrode that applies in the direct methods.

Competitive preparative-scale electrolysis is an approach that can provide a limited insight into the relative rates of reactions. Here, the distribution of products between two competitive reactions is measured, one of which is the reaction of interest and the other is known from the independent sources and acts as reference. For example, Savéant et al.[81] have investigated the competition that takes place between electron and

hydrogen atom transfer pathways for the reduction of aryl halides in H-atom donor solvents such as acetonitrile and dimethylsulfoxide (DMSO). Mathematical models describing how the product distributions depend on the reaction mechanism and rate constant have been determined. Not only do these analyses provide important fundamental information about those factors that control chemical reactivity, they also allow important industrial processes to be optimized, e.g. reduction of CO_2 in media of low proton mobility.

3.1 Redox Catalysis

Redox catalysis can be viewed as the ultimate limit of small probe electrochemistry in which the electrode is decreased to the size of a single molecule. Figure 12 illustrates how the dynamics of an electron transfer reaction can be probed using a mediator in a redox catalysis scheme. Only the case in which the dynamics of a reduction reaction are being probed is considered.

In this approach, a substrate A that would normally be reduced at an electrode surface is now reduced by a mediator Q, i.e. the mediator Q replaces the electrode surface and reduces the substrate A. If the reduced form of the redox couple of interest undergoes a following chemical reaction, then its decay can be monitored using this approach. Figure 13 illustrates the differences between redox catalysis and direct electrochemistry.

While it provides great insight into a wide variety of reaction mechanisms, the useful time domain of this technique is somewhat unusual in light of the discussion on direct methods. Redox catalysis is only useful for monitoring species with lifetimes less than approximately 10 μs. As discussed previously, the rate of diffusion increases with decreasing electrode size. In redox catalysis, the electrode has been reduced to the dimensions of a single molecule allowing extremely low timescales to be accessed. In the case where the driving force is large (i.e. large values of $E^{o}_{PQ} - E^{o}_{AB}$), k_+ will have a value close to the diffusion limit. Taking a maximum value of 50 mM for $[P]_0$, it is apparent that the shortest measurable lifetime, $k\,[P]_0$, is of the order of 50 ns.

In redox catalysis, the current flowing through the electrode is monitored. The mediating P/Q couple is selected to be electrochemically fast and reversible as well as having a standard potential located positive of the substrate (A) reduction wave. Thus, in cyclic voltammetry experiments performed on the mediator alone, reversible on-electron waves are observed even at high scan rates because the rate of heterogeneous electron transfer is fast. However, when the analyte of interest is added, the voltammetry for the mediator becomes irreversible and the reduction current increases. These effects provide kinetic information about the electron transfer and coupled chemical reactions. As illustrated in Figure 13, the rate-determining step is either the forward electron transfer or the homogeneous chemical reaction with the electron transfer step acting as a pre-equilibrium. It is the competition that exists between the backward electron transfer step and the chemical reaction that dictates which process will be rate determining.

The work of Utley et al. on *ortho*-quinodimethanes (*o*-QDM) provides a good illustration of how redox catalysis affects the voltammetric response.[82] As illustrated in Figure 14, co-electrolysis of 1,2-bis(bromomethylarenes) and dienophiles (e.g. maleic anhydrides or quinones), gives the Diels-Alder adduct expected for reaction of the *o*-QDM derived from the dihalide. It is known that direct reduction of the dihalides gives *o*-QDMs but in the absence of a dienophile they polymerize. However, Utley performed co-electrolysis experiments at a potential several hundred millivolts more positive than the reduction potential of the dihalide. Cyclic

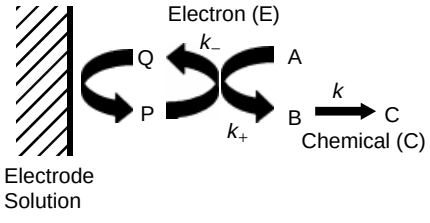

Figure 12 Dynamics of an EC reaction.

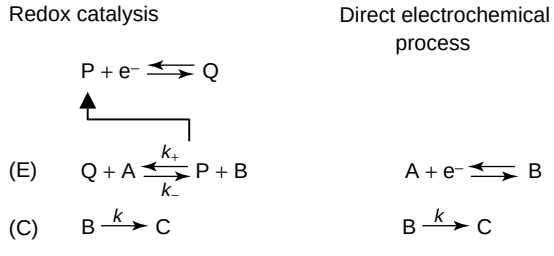

Figure 13 Comparison of redox catalysis and direct electrochemistry.

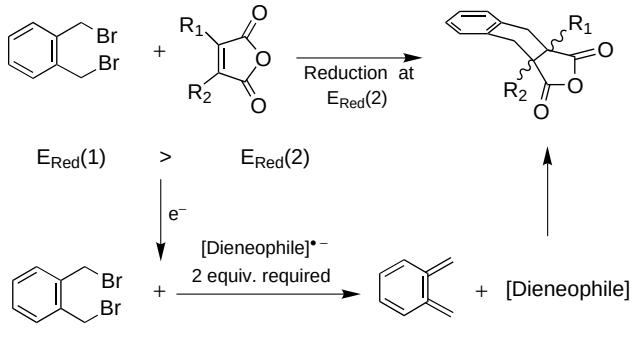

Figure 14 Redox catalyzed reaction scheme for 1,2-bis(bromomethylarenes) and functionalized maleic anhydride.

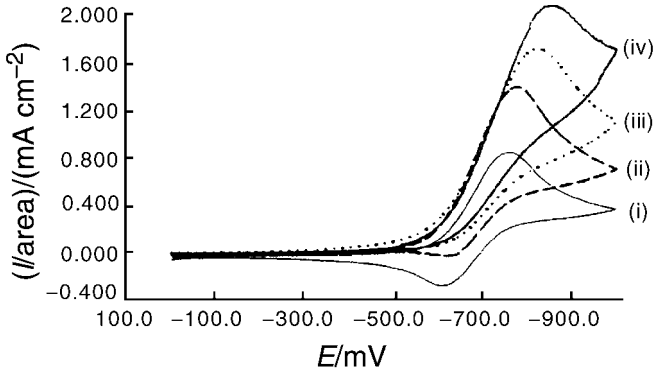

Figure 15 Characteristic cyclic voltammetry of a redox catalyzed system at a mercury electrode where the supporting electrolyte is 0.1-M tetraethyl ammonium perchlorate in DMF. The scan rate in all cases is $0.3\,V\,s^{-1}$. (i) Anhydride only, (ii), (iii), and (iv) contain 0.5, 1.0 and 2.0 equiv. of dibromide, respectively. (Reprinted from E. Oguntoye, S. Szunerits, J.H.P. Utley and P.B. Wyatt, *Tetrahedron*, **52**, 7771. Copyright (1996) with permission of Elsevier Science.)

voltammetry of this system is illustrated in Figure 15 which shows that a chemically reversible reduction is observed for the mediator (dienophile) in the absence of the dihalide. Moreover, this reversible response is observed at slow scan rates, indicating that the radical-anion is stable over a significant lifetime. However, in the presence of the analyte of interest, the radical-anion can transfer an electron to the added dihalide, which rapidly undergoes a following chemical reaction, i.e. it loses bromide anion. As shown in Figure 14, at long times an *o*-QDM is formed, which can react relatively slowly with the regenerated mediator. On the cyclic voltammetry timescale the regenerated mediator is rapidly reduced, and it is this process that is responsible for the loss of reversibility and the more than doubling of the peak current seen in Figure 15. This increased current is known as the catalytic current.

The effect of kinetics on the diffusion processes of P, Q and A is given by Equation (12)

$$\frac{d[P]}{dt} = -\frac{d[Q]}{dt} = \frac{-d[A]}{dt} = k_+ k \frac{[A][Q]}{k_-[P] + k} \qquad (12)$$

Kinetic control by the forward electron transfer step, Equation (13)

$$\frac{d[P]}{dt} = -\frac{d[Q]}{dt} = \frac{-d[A]}{dt} = k_+[A][Q] \qquad (13)$$

is achieved when $k_-[P]_0 \ll k$. In contrast, when $k_-[P]_0 \gg k$, the kinetics of the homogeneous chemical reaction are rate determining, Equation (14)

$$\frac{d[P]}{dt} = -\frac{d[Q]}{dt} = \frac{-d[A]}{dt} = \frac{k_+ k[A][Q]}{k_-[P]} \qquad (14)$$

This analysis reveals that the rate-determining step depends on the concentration of the mediator, P, and can be switched from electron transfer to homogeneous chemical reaction by changing $[P]_0$. Solutions to these systems of differential equations have been obtained for a wide variety of electrochemical techniques, including cyclic voltammetry and chronoamperometry, allowing rate data to be obtained for unstable intermediates having lifetimes down to the low nanosecond time range.

The technique has certain limitations, e.g. if the chemical reaction is so fast that the kinetic control is always by the forward electron transfer step then it is not possible to characterize the homogeneous chemical reaction kinetically. However, under these conditions it is often still possible to measure the formal potential and heterogeneous electron transfer rate constant for the A/B couple if experiments are conducted using a range of mediators with different formal potentials. Savéant et al.[81] have elucidated the relationship between the electron transfer rate constant and the formal potential of the PQ couple where the homogeneous chemical reaction is so fast that it occurs within the molecular diffusion layer of the mediator. Moreover, this group has extended the theory to allow more complex reaction schemes, including bimolecular following reactions, than the electron transfer reaction described above to be characterized. These advances are significant since simple electron transfer reaction schemes are rarely found in real experimental systems.

3.2 Indirect Laser-induced Temperature Jump Method

There is a long and distinguished record of using temperature jump to elucidate the kinetics of homogeneous chemical phenomena, e.g. Sutin et al.[83] used a laser-induced temperature jump method to achieve rise times of tens of nanoseconds. Moreover, approaches have been developed to allow the temperature of the electrode/solution interface to be rapidly changed. However, electrical heating methods typically allow the interfacial temperature to be perturbed only at the millisecond or longer timescale. The use of lasers has radically changed this field allowing much more rapid temperature jumps to be achieved allowing fast electron transfer processes to be investigated. These laser-based approaches stem from early work in which laser flashes impinged on the electrode/solution interface causing electrons to be photoemitted. For example, Barker et al.[84] studied the effects of irradiation by a 20 ns laser pulse ($\lambda = 694.3$ nm) on mercury and suggested that laser-induced temperature changes might be the cause of some of the observed potential changes since the laser energy was insufficient to cause photoemission at the applied potentials investigated. This "front side" irradiation approach without stimulated photoemission was also used by Benderskii

et al.[85] to examine double-layer phenomena at mercury electrodes with nanosecond time resolution. Coufal and Hefferle[86] have also demonstrated the ability of laser-induced temperature steps to resolve nonelectrochemical events at subnanosecond timescales.

An extremely important variation on these approaches has been developed by Feldberg and co-workers.[87] Their technique, the indirect laser-induced temperature jump method (ILIT) involves indirect irradiation of the electrode, i.e. the laser pulse impinges on the "back face" or nonelectrolyte electrode surface. In this method, some of the incident photons are absorbed by the thin foil or film electrode and thermalise virtually instantaneously causing a rapid rise in the interfacial temperature. The time dependence of the thermal perturbation is determined by the shape of the laser pulse and by the thickness, thermal diffusion constant, and heat capacity of the electrode material, also the heat capacities and thermal diffusion constants of the two adjacent media (electrolyte at the front face, and dielectric surrounding the rear of the electrode). Typically, this interfacial temperature jump destroys the equilibrium between the double-layer potential and the redox potential. This process changes the open-circuit potential as the system attempts to re-establish equilibrium under the new conditions. In contrast to classical charge-injection methods, thermal perturbation does not cause a large *iR* spike, making data interpretation more straightforward. Moreover, the ILIT offers the possibility of extracting information about thermal as well as electrochemical properties of electrodes and adsorbed layers.

When compared with direct irradiation of the electrode/solution interface, Feldberg's indirect approach has several advantages. Photoemission of electrons into the solution is not possible; the optical density of the solution is irrelevant and photolysis and/or direct thermal heating of the solution is avoided; the form of the interfacial temperature perturbation approximates a step function more closely than that associated with direct irradiation; deconvolution of the data is facilitated and the rates of the rise and fall of the interfacial temperature can be adjusted (but not independently) by changing the thickness of the electrode.

There are three major processes that can cause the open-circuit potential to change in response to the temperature jump. First, there is a junction potential between the (hot) electrode and the (cold) contact wire. Second, a change occurs in the potential across the electrode double layer which can be caused by a change in the capacitance or dipole reorientation (equivalent to a change in the potential of zero charge); by charge transfer (electron transfer between the electrode and a redox system) located at the outer Helmholtz plane (solution-phase reactant) or inner Helmholtz plane (adsorbed reactant), or ion transfer between the inner and outer Helmholtz planes. Third, a Soret potential arises from the temperature gradient between the interfacial region and the bulk electrolyte. Therefore, this approach can be used to determine kinetic and thermodynamic parameters of potential-dependent double-layer restructuring, heterogeneous electron transfer to the solution phase and adsorbed electroactive species and ion transfer from the inner to outer Helmholtz planes, all at nanosecond or shorter timescales. The theoretical analysis for each of these situations yields an analytic equation provided that the temperature changes are small (controllable through the laser power) and the charge-transfer resistance is constant.

As an example, determination of the heterogeneous electron transfer rate constant for ferro/ferri cyanide and the ways in which ILIT can provide a powerful insight into adsorption phenomena are reviewed here. As illustrated in Figure 16, the typical ILIT experiment uses a Nd:YAG laser ($\lambda = 1.06\,\mu m$, full width at half height maximum of 8 ns) that is capable of resolving relaxations as short as 10 ns. The pulse from this laser passes through a neutral density filter that allows the incident power to be controlled before being directed by a mirror onto the back surface of the working electrode. As shown in Figure 17, the cell contains two cylinders of 25 μm Pt foil approximately 1 cm wide. One cylinder acts as a pseudoreference electrode against which the change in the open-circuit potential is measured while the other is used as an auxiliary electrode for a three-electrode potentiostat.

Figure 18, illustrates examples of the ILIT responses obtained for a 1.07 μm gold film electrode in contact with 1.0 M KF before and after the addition of 4×10^{-5} M of $K_3[Fe(CN)_6]$ and $K_4[Fe(CN)_6]$. In both cases the electrode potential is poised at the formal potential of the redox couple. The magnitude of the ILIT response is of the order of a few millivolts and is typical of these experiments. In the absence of

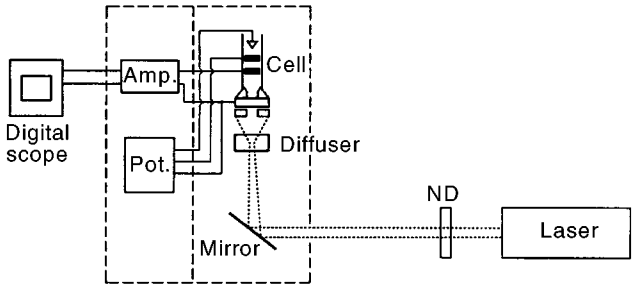

Figure 16 Block diagram of the ILIT apparatus. ND, neutral density filter; Amp., amplifier; Pot., potentiostat. The broken line denotes the Faraday cage. (Reprinted from J.F. Smalley, L. Geng, S.W. Feldberg, L.C. Rogers, J. Leddy, *J. Electroanal. Chem.*, **356**, 181. Copyright (1993) with the permission of Elsevier Science.)

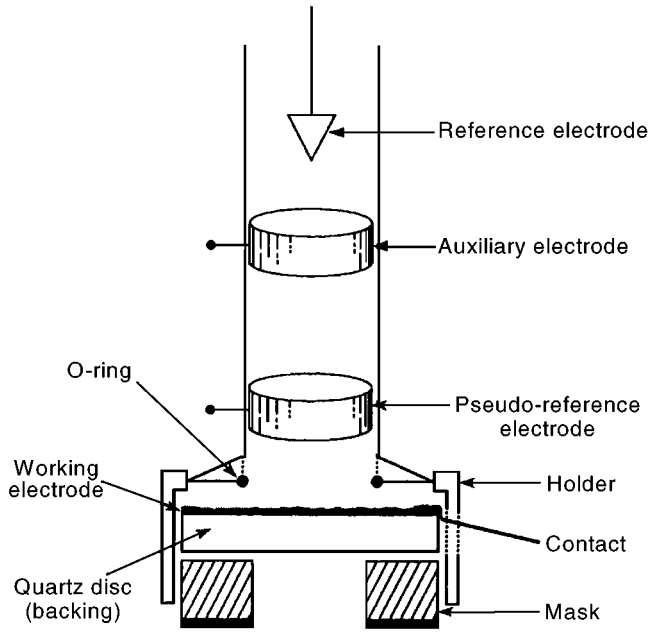

Figure 17 Diagram of the ILIT cell. (Reprinted from J.F. Smalley, L. Geng, S.W. Feldberg, L.C. Rogers, J. Leddy, *J. Electroanal. Chem.*, **356**, 181. Copyright (1993) with the permission of Elsevier Science.)

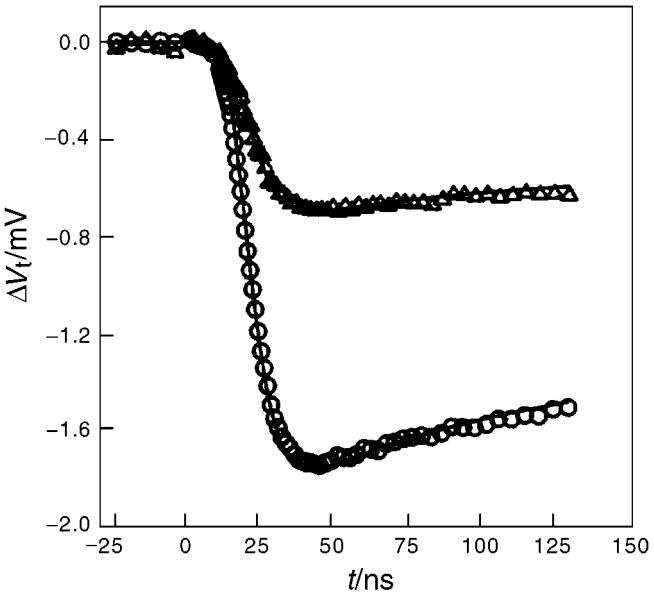

Figure 18 The ILIT responses of a gold electrode in 1.0-M KF at 0.2 V vs saturated calomel electrode (SCE) before (Δ) and after (◦) addition of 4×10^{-5} M each of $K_3[Fe(CN)_6]$ and $K_4[Fe(CN)_6]$ at $E = E^o = 0.214$ V vs SCE. The solid curves represent theoretical fits. (Reprinted from J.F. Smalley, L. Geng, S.W. Feldberg, L.C. Rogers, J. Leddy, *J. Electroanal. Chem.*, **356**, 181. Copyright (1993) with the permission of Elsevier Science.)

any dissolved electroactive species, the open-circuit potential changes abruptly within a few nanoseconds and then relaxes over several hundred nanoseconds. Both before and after addition of the electroactive species, simple thermal responses are obtained, i.e. the change in the open-circuit potential precisely follows the change in the interfacial temperature. Moreover, the time dependence of these responses follow, almost exactly, that predicted by theory. Analyzing the ILIT response in the presence of the electroactive species provides a value of 1.64×10^{-2} cm s^{-1} for the standard heterogeneous electron transfer rate constant which is in reasonable agreement with previous results obtained using traditional electrochemical techniques.

Implicit in the analysis is the assumption that the change in the interfacial temperature does not cause the rate constant to change significantly. For the case of the $Fe(CN)_6^{3-/4-}$ couple, the activation energy is of the order of 13 kJ mol^{-1} so that changes in the electrode temperature of up to 4 K will result in k^o being constant to within 10%. However, for reactions featuring large activation energies, e.g. those with large inner or outer sphere reorganization energies, temperature-induced changes in k^o may be significant.

Although the open-circuit potential becomes constant with time in all cases, suggesting that equilibrium is achieved, the magnitude of the ILIT response depends on the concentration of the ferrocyanide and ferricyanide. While effects such as adsorption of an impurity, electron transfer between the electrode, and solution-phase reactant cannot be excluded, this observation is considered to arise because of $Fe(CN)_6^{3-/4-}$ adsorption onto the electrode surface. Adsorption will modify the structure of the electrochemical double layer thus changing the thermal response. The concentration-dependent ΔV_t can be used as a measure of the surface coverage of the adsorbate and the concentration-dependent surface coverage modeled using the Frumkin adsorption isotherm. This analysis reveals that the $Fe(CN)_6^{3-/4-}$ layer is strongly bound, $K_{ads} = 4.7 \times 10^9$ M^{-1}. That positive values of the interaction parameter, g, are observed suggests either that there are strong repulsive interactions between the adsorbates, or that the potential difference between the inner and outer Helmholtz planes changes with changing surface coverage.

A very important example of the ultrafast capabilities of this method is in the determination of heterogeneous electron transfer rate constants for electroactive self-assembled monolayers. As discussed in the section dealing with fast transient methods, when a short bridge is used to tether the redox active species to the electrode surface, the rate of heterogeneous electron transfer becomes immeasurably fast even for chronoamperometry at microelectrodes. The ILIT was used to probe the kinetics of electron transfer between a gold electrode and a self-assembled monolayer formed from $CH_3(CH_2)_{n-1}SH$ and

$(\eta^5-C_5H_5)Fe(\eta^5-C_5H_4)CO_2(CH_2)_nSH$ (i.e. a diluted ferrocene alkane thiol monolayer), where n is the number of methylene groups in the bridge and $5 \leq n \leq 9$.[88] In these experiments, as well as recording the time-dependent change in the open-circuit potential, the size of the temperature jump at the electrode/electrolyte interface was monitored using a pressure transducer.

Figure 19 shows the ILIT response obtained close to the formal potential of the ferrocene/ferricenium redox reaction for a gold film electrode modified with a ferrocene alkane thiol monolayer in which the bridge contains five methylene groups. Fitting these data using theory developed for heterogeneous electron transfer involving a surface confined species provides an estimate of k as $7.8 \times 10^7\,s^{-1}$. This value is comparable with the upper limit of that possible using state-of-the-art transient chronoamperometry. By controlling the applied potential, the potential dependence of the heterogeneous electron transfer rate constant can be determined. For this system, this dependence appears to be adequately described by the Butler–Volmer formalism of electrode kinetics.

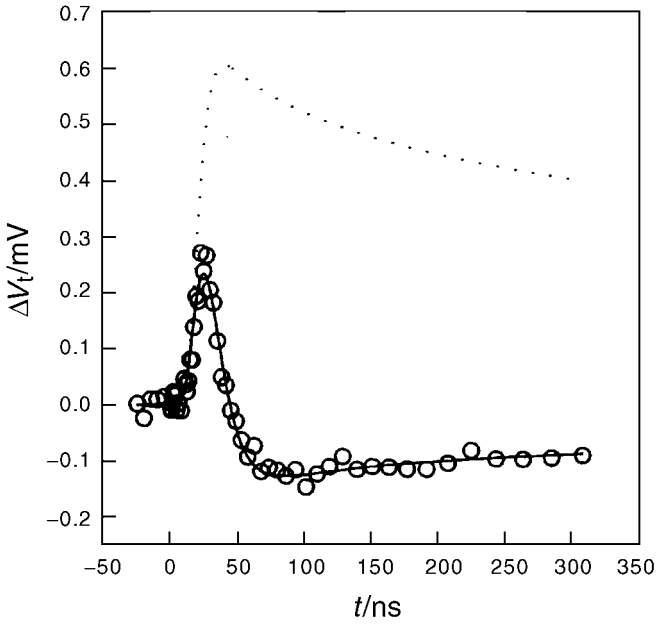

Figure 19 ILIT response obtained for a gold electrode coated with a mixed monolayer formed from $CH_3(CH_2)_4SH$ and $(\eta^5-C_5H_5)Fe(\eta^5-C_5H_4)CO_2(CH_2)_5SH$ with $N_T = 1.3 \times 10^{-10}$ mol, $T = 31.0\,^{\circ}C$, thickness of the electrode $= 0.96 \times 10^{-4}$ cm. Open circles are the experimental data. $E_{app} = 0.475$ V vs standard SCE, solid line is the theoretical fit. The dotted line represents the response which would be observed if there were no relaxation of the ILIT signal caused by electron transfer between the electrode and the redox couple. (Reprinted with permission from J.F. Smalley, S.W. Feldberg, C.E.D. Chidsey, M.R. Linford, M.D. Newton, Y.-P. Liu, *J. Phys. Chem.*, **99**, 13 141. Copyright (1995) American Chemical Society.)

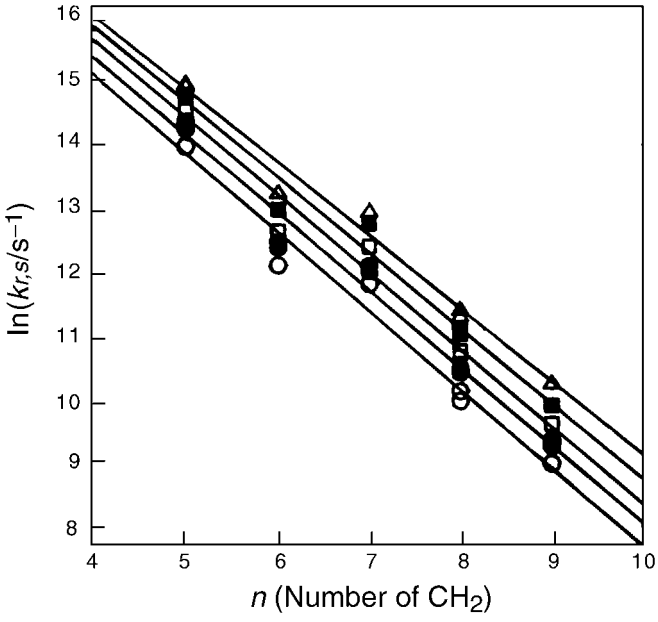

Figure 20 Plots of $k_{\Gamma,s}$ versus n at five different temperatures. Open circles, $T = 14.9\,^{\circ}C$; filled circles, $T = 25\,^{\circ}C$; open squares, $T = 35.0\,^{\circ}C$; filled squares, $T = 44.8\,^{\circ}C$; open triangles, $T = 55.2\,^{\circ}C$. (Reprinted with permission from J.F. Smalley, S.W. Feldberg, C.E.D. Chidsey, M.R. Linford, M.D. Newton, Y.-P. Liu, *J. Phys. Chem.*, **99**, 13 141. Copyright (1995) American Chemical Society.)

Beyond potential-dependent rate constants, the approach was extended to measure activation (reorganization) energies by using a thermostated cell. Figure 20 shows that the electron transfer rate constant depends on the number of methylene groups in the bridge. At $25\,^{\circ}C$, the standard electron transfer rate constants vary according to $k_{\Gamma,s,n=0}\exp[-\beta_n n]$ where $k_{\Gamma,s,n=0}$ is the extrapolated rate constant for the electron transfer at $n = 0$ and is equal to $6 \times 10^8\,s^{-1}$ and β_n is 1.21 ± 0.05. Moreover, for each of the monolayers investigated, linear Arrhenius plots are observed. The slopes of these plots have been used to determine the reorganization energy, λ, assuming that λ is the sole contributor to the activation energy for electron transfer. The reorganization energies range from 0.70 ± 0.04 eV for $n = 5$ to 0.91 ± 0.05 for $n = 9$.

3.3 Scanning Electrochemical Microscopy

Bard et al. have developed an important technique, SECM,[89] in which a small electrode is precisely positioned close to a surface in an electrochemical cell arrangement to obtain information about the surface topography and reactions that occur in the solution space between tip and sample. As described in Equation (10), at long timescales a steady-state potential-independent current is observed for the oxidation or reduction of an

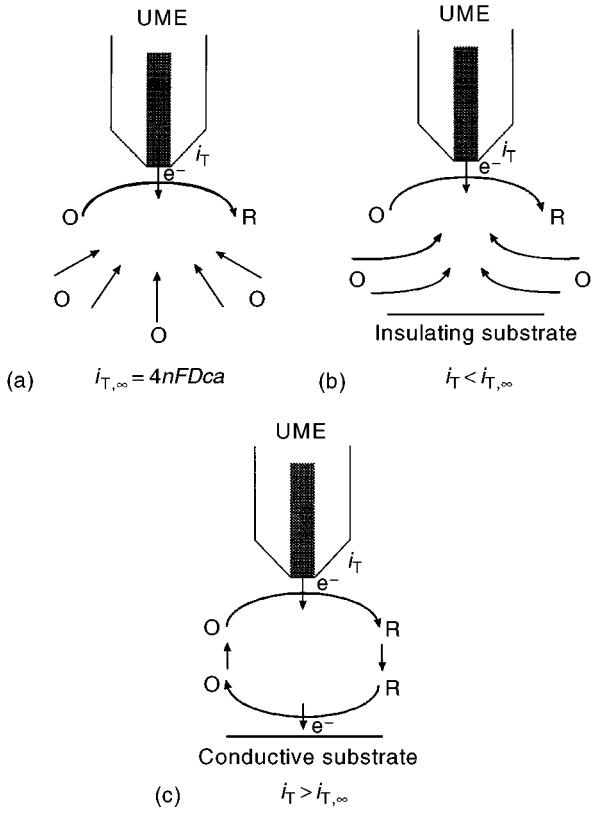

Figure 21 Basic principles of SECM. (a) With the microelectrode far from the substrate, diffusion of O leads to a steady-state current, $i_{T,\infty}$. (b) With the microelectrode near an insulating substrate, hindered diffusion of O leads to $i_T < i_{T,\infty}$. (c) With the microelectrode near a conducting surface, positive feedback of O to the tip leads to $i_T > i_{T,\infty}$. (Reproduced with the permission of the Royal Society of Chemistry from M. Arca, A.J. Bard, B.R. Horrocks, T.C. Richards, D.A. Treichel, *Analyst*, **119**, 719 (1994).)

electroactive species at a microelectrode. Figure 21 shows how the current to this microelectrode tip is perturbed by the presence of a substrate near the tip. These effects form the basis of the SECM technique. When the tip is brought near a nonconducting surface, the current will be smaller than the steady-state value of Equation (10) because the surface blocks diffusion of the redox active species to the tip. The closer the tip is to the surface, the smaller the current, until when the tip-to-surface spacing, d, becomes zero, the tip current will approach zero. The tip current, i_T, as a function of separation from the surface can be approximately described by Equation (15)

$$\frac{i_T}{i_{T,\infty}} = 0.292 + \frac{1.515}{L} + 0.655 \exp\left(\frac{-2.4035}{L}\right)^{-1} \quad (15)$$

where the normalized tip–substrate separation, L, is equal to d/r. Thus, the presence of a nonreactive substrate near the tip can always be recognized by the condition $i_T < i_{T,\infty}$.

A different behavior is observed when the tip is near a surface where tip-generated product can be re-oxidized or re-reduced, e.g. when the substrate is a conductive material at a sufficiently positive or negative potential. Under these circumstances, there is a flux of the analyte to the tip from both the substrate and the bulk solution. Owing to this positive feedback from the substrate, $i_T > i_{T,\infty}$. When the redox reaction occurring at the surface proceeds at a diffusion-controlled rate, Equation (16) applies

$$\frac{i_T}{i_{T,\infty}} = 0.68 + \frac{0.738}{L} + 0.3315 \exp\left(\frac{-1.0672}{L}\right) \quad (16)$$

Figure 22 shows approach curves, i.e. plots of $i_T/i_{T,\infty}$ versus L for the cases where the microelectrode tip approaches conducting and nonconducting surfaces.

An important attribute of the technique is that if the electrode radius is known, the Faradaic current can be used to determine the tip–surface separation without having prior knowledge of the bulk concentration or diffusion coefficients.

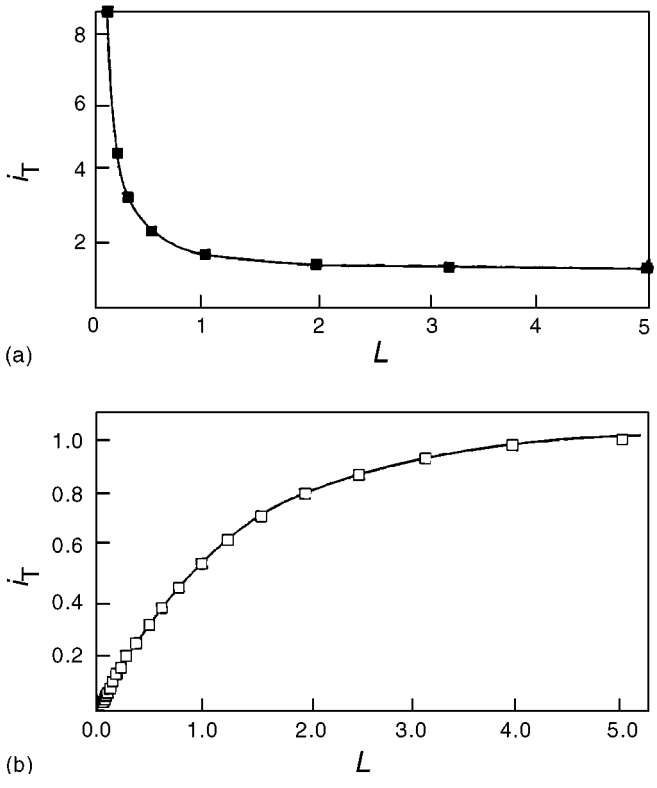

Figure 22 Diffusion controlled steady-state tip current as a function of tip–substrate separation. $i_T = i_T/i_{T,\infty}$, $L = d/r$. (a) Substrate is a conductor, (b) conductor is an insulator. The solid lines are theoretical fits. (Reproduced with the permission of the Royal Society of Chemistry from M. Arca, A.J. Bard, B.R. Horrocks, T.C. Richards, D.A. Treichel, *Analyst*, **119**, 719 (1994).)

3.3.1 Heterogeneous Electron Transfer Dynamics

SECM has recently emerged as a useful technique for measuring the rate constants of fast heterogeneous electron transfer reactions. To achieve this objective, the SECM apparatus is used to form a twin-electrode thin-layer cell between the microelectrode tip and a conducting substrate. During voltammetric scans, this configuration induces high rates of mass transfer between the substrate and tip electrodes because large concentration gradients exist between the two closely spaced electrodes. The overall current observed can be limited by either the rate of mass transfer to the electrode or the intrinsic heterogeneous electron transfer rate. Therefore, the increase in mass transfer of the thin layer cell allows faster rates of electron transfer to be measured in the SECM arrangement than can be accomplished with the same electrode in bulk solution.

Most electrochemical responses are sensitive to the heterogeneous electron transfer rate whenever the rate of mass transfer is of the same order of magnitude or greater than the standard heterogeneous electron transfer rate constant. The advantage of SECM can be seen by comparing mass-transfer coefficients for two different electrode configurations. For a microdisc in bulk solution, the steady-state mass-transfer coefficient is approximately D/r. For thin-layer cells, the mass transfer coefficient is approximately D/d. However, while the smallest well-characterized microdisc electrodes are of the order of 500 nm, stable layers as thin as 40 nm have been obtained for a 1-μm diameter Pt disc electrode opposed by a mercury substrate. Clearly, the rate of mass transfer can be enhanced by more than an order of magnitude in going from a microelectrode in bulk solution to the thin-layer configuration in SECM. Achieving a 40 nm substrate to tip separation allows rate constants of the order of several centimeters per second to be measured under steady-state conditions.

Figure 23 shows experimental data and theoretical calculations for the oxidation of ferrocene system taken at different tip–substrate separations.[90] This figure shows that the wave broadens as the tip–substrate separation decreases and the limiting current increases owing to enhanced feedback of the analyte by reaction at the conducting substrate. More significantly however, the waves broaden because the mass-transfer rate increases to the point where is becomes comparable to, or greater than, the electron transfer rate. It is this increased sensitivity of the voltammetric response to k that allows large rate constants to be measured under steady-state conditions. The measured k^{o} for ferrocene was $3.7 \pm 0.6\,\mathrm{cm\,s^{-1}}$, confirming the rapid nature of heterogeneous electron transfer in this system.

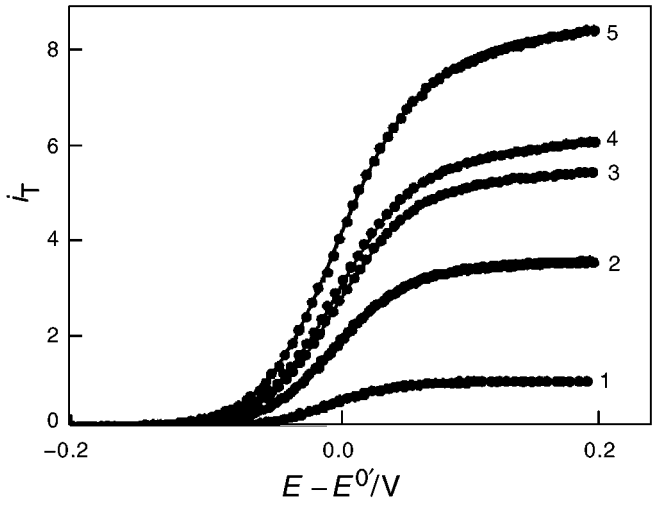

Figure 23 Experimental data and theoretical fits for the oxidation of ferrocene at a 2-μm-diameter platinum tip opposed by a 1.5-mm-diameter platinum substrate electrode. Voltammograms 1–5 show the transition from data recorded in bulk solution (curve 1) to those recorded with the tip electrode 100 nm from the substrate. Currents are normalized such that $i_T = i_T/i_{T,\infty}$. (Reproduced with the permission of the Royal Society of Chemistry from M. Arca, A.J. Bard, B.R. Horrocks, T.C. Richards, D.A. Treichel, *Analyst*, **119**, 719 (1994).)

3.3.2 Homogeneous Chemical Kinetics

SECM can also be used to measure the rate of homogeneous chemical reactions that follow oxidation or reduction by monitoring the effect of the reaction of interest on the tip current. For example, in the case of a reversible redox reaction, when d is small and the substrate area is large in comparison to the microelectrode tip, essentially all of the material that is electrogenerated at the tip is collected by the substrate. However, a following chemical reaction will consume some of the tip product, decreasing the collection efficiency. In this sense, the tip and substrate electrodes act as a generator–collector assembly in an analogous way to the ring and disc electrodes in a rotating ring-disc electrode. Theory that describes both first- and second-order homogeneous chemical reactions following electron transfer has been developed.[89] This theory is most accessible in the form of working curves describing how the currents observed at both tip and substrate depend on their mutual separation.

A typical reaction scheme which can be studied by SECM is shown by Equation (17)

$$\begin{aligned} &\text{tip:} && \mathrm{R} \longrightarrow \mathrm{O} + n\mathrm{e}^- \\ &\text{gap:} && 2\mathrm{O} \longrightarrow \text{products} \\ &\text{substrate:} && \mathrm{O} + n\mathrm{e}^- \longrightarrow \mathrm{R} \end{aligned} \qquad (17)$$

Rate constants can be determined if the tip and substrate are positioned sufficiently close so that a detectable amount of R can cross the gap before reacting. The time required for a species to cross the gap is approximately d^2/D. Given that the half-life of the electrogenerated intermediate undergoing a second-order following reaction is $1/kC$, where C is the bulk concentration, rate constants as large as $10^8\,M^{-1}\,s^{-1}$ can be measured using separations of the order of 100 nm. For unimolecular reactions, first-order rate constants as large as $10^5\,s^{-1}$ are measurable.

Bard et al.[91] used the SECM technique to investigate the reductive coupling of dimethyl fumarate as well as fumaronitrile in DMF. For fumaronitrile, the rate constant was shown to be approximately constant as the analyte concentration was changed over two orders of magnitude. This consistency indicates that second-order kinetics are being followed for the following chemical reaction of the fumaronitrile anion radical. The measured rate constant is large, $2 \times 10^5\,M^{-1}\,s^{-1}$, demonstrating the ability of SECM to provide high-quality kinetic information about fast homogeneous chemical reactions. In contrast, the other alkene, dimethyl fumarate, appears to be less activated and undergoes a much slower chemical reaction when reduced and a rate constant of $170\,M^{-1}\,s^{-1}$ was determined using both tip feedback and substrate current measurements. Good agreement of experiment with theory was seen for both modes of SECM operation.

Despite the rapid nature of these processes, Bard and his group have extended the use of SECM to the study of even faster reactions, e.g. the dimerization of 4-nitrophenolate,[92] which has been shown to undergo a fast irreversible dimerization following oxidation. The calculated rate constant, approximately $8 \times 10^7\,M^{-1}\,s^{-1}$, agrees well with that measured by fast-scan cyclic voltammetry.

4 FUTURE DIRECTIONS

4.1 Analytical Applications of Fast Electrochemistry

When attempting to determine the concentration of a particular redox active species in a complex mixture, the response of the target analyte is typically separated from redox active interferences on the basis of different formal potentials. In fact, there have been relatively few reports using differences in electrochemical reactivity (i.e. electrode kinetics) to determine the concentration of a target analyte by separating its voltammetric response from that of an interferent on the basis of different time constants for the two reactions. Since the width of the electrochemical response for any species is a sizeable fraction of the potential scale, relying on the potential axis alone to generate a selective response provides only a very limited ability to resolve an analyte's response from that of an interfering species. If the time axis as well as the potential axis can be used, separating the analyte's response from that of an interfering species becomes considerably more likely. This time-resolved approach ought to benefit significantly from the dramatic expansion in the range of timescales that can be resolved and exploited in electrochemistry with the advent of high-speed instrumentation and microelectrodes.

It is widely recognized that double-layer charging can represent a significant obstacle to achieving very low limits of detection in electroanalysis. Often, background subtraction is achieved by measuring the voltammetric response in the absence of the electroactive species of interest. However, differences in rates at which charging and Faradaic processes occur can be exploited to time gate the response so that an enhanced Faradaic to capacitive current ratio is obtained.

4.2 Ultimate Limits on Ultrafast Electrochemical Techniques

Decreasing the timescale of t_{dif} of an electrochemical experiment involving a solution-phase reactant can be achieved either by shortening the experimental timescale (faster voltammetric scan rates or shorter potential steps), or by shrinking the critical dimension of the microelectrode in steady-state measurements. Both of these effects cause the thickness of the diffusion layer to shrink and it will eventually become less than that of the electrochemical double layer. This effect has profound consequences for migration of the analyte within the electric field of the double layer.

In conventional electrochemical cells containing a dilute ($\leq$0.1 M) supporting electrolyte, the electrical field that exists because of the electrode–solution potential difference decays exponentially from the electrode surface out into solution over a distance of a few tens of angstroms. Under these conditions, the effects of diffusion or migration within the double layer are negligible. In classical electrochemical theories, it is assumed that the concentrations of charged species within the interfacial region are described by a Boltzmann distribution imposed by the electrical potential of the double layer. Further out in solution, diffusion is the mechanism of mass transport of molecules and ions. The Frumkin correction allows the true electrode concentrations to be related to those existing at the end of the electrochemical double layer. While these concentrations do not represent those actually found at the electrode surface, they are usually used as boundary conditions in solving diffusion problems.

The difficulty with the Frumkin approximation arises when dealing with ultrafast electrochemical techniques since the sizes of the double and diffusion layers become comparable. Therefore, using the concentrations found at the end of the interfacial region as a boundary condition ceases to be appropriate. For example, given a typical diffusion coefficient of $10^{-5}\,cm^2\,s^{-1}$, the diffusion layer thickness will be of the order of 100 Å for a cyclic voltammetry experiment in which the scan rate is $300\,000\,V\,s^{-1}$ or for a 10 nm electrode in a steady-state measurement. This diffusion-layer thickness is comparable to the typical double-layer thickness. The difficulty arises because the voltammetric response will no longer be controlled by diffusion alone but migration within the double layer will influence the observed response.

Recently, two models describing this mixed control situation have been developed for transient and steady-state methods. Both of these models predict that the voltammetric waves appear to be kinetically controlled by heterogeneous electron transfer when analyte migration within the double layer becomes important. However, the predicted behavior contrasts with that predicted by conventional theories of electrode kinetics. In the Butler–Volmer formalism of electrode kinetics, the heterogeneous electron transfer rate constant is expected to increase without limit as the overpotential (electrochemical driving force) is increased. However, coupling between diffusion and migration in the interfacial region causes the apparent heterogeneous electron transfer rate constant eventually to become potential independent, corresponding to the maximum rate, not of electron transfer, but of the reactant crossing the double layer. Further decreases in the critical dimension of ultramicroelectrodes used in steady-state measurements, and the timescale of transient electrochemical techniques, will allow further testing of these theories to be performed.

In order to achieve faster perturbation of the chemical system in transient measurements, it will be necessary to generate pulse signals with rise and fall times on the picosecond timescale if deconvolution of the instrumental and electrode responses is to be avoided. With the development of ever faster chips (e.g. even personal computers have cycle times on the order of nanoseconds), programmable pulse generators with this performance level are becoming more widely available commercially. Another key issue is the transduction of the fast current response. The transient current response for electrodes with subnanosecond *RC* time constants will be on the nano- and even picoamp scale. One strategy for detecting these very small currents with high speed is to insert a second stage of amplification following fast current-to-voltage conversion. Amplifiers of this type that operate at gigahertz frequencies are also available, although their cost is high. Ultrahigh-speed transient digitizers with gigahertz bandwidth and gigahertz sampling rate are also available. However, the total cost of assembling such a system with low to subnanosecond time resolution would be of the order of $175 000.

5 CONCLUSIONS

This article has attempted to convey the significant new insights and opportunities that ultrafast electrochemical techniques have provided. This revolution in electrochemistry has greatly extended the range of conditions of solvent, temperature, and timescale under which it is now possible to obtain information about redox processes directly. In particular, today microelectrodes allow experiments (e.g. voltammetry in oil or concrete) to be performed that would simply have been impossible a few years ago. This advance has not only revolutionized the field internally, it has broadened the impact of electrochemistry into new dimensions of space and time, e.g. microsecond monitoring of neurotransmitter release with single-cell spatial resolution.

ACKNOWLEDGMENTS

The Irish Science and Technology Agency, Enterprise Ireland, that has funded our research in this area over several years is gratefully acknowledged. Sincere thanks go to Professor Larry R. Faulkner of the University of Austin at Texas for a wonderful introduction to this area of study.

ABBREVIATIONS AND ACRONYMS

dipy	4,4′-Trimethylenedipyridine
DISP	Disproportionation
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
ECSTM	Electrochemical Scanning Tunneling Microscopy
HIM	Imidazole
ILIT	Indirect Laser-induced Temperature Jump Method
MeIM	Methyl Imidazole
MV	Methyl Viologen
OEP	Octa-ethyl-porphyrin
o-QDM	*ortho*-Quinodimethanes
SCE	Saturated Calomel Electrode
SECM	Scanning Electrochemical Microscopy

STM	Scanning Tunneling Microscopy
TPP	Tetra-phenyl Porphyrin
2,6-AQDS	Anthraquinone-2,6-disulfonic acid

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Voltammetry In Vivo for Chemical Analysis of the Living Brain

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Organic Electrochemical Mechanisms • Pulse Voltammetry

REFERENCES

1. M. Fleischmann, S. Pons, D.R. Rolison, P. Schmidt (eds.), *Ultramicroelectrodes*, Datatech Science, Morgantown, NC, 1987.
2. R.J. Forster, 'Microelectrodes: New Dimensions in Electrochemistry', *Chem. Soc. Rev.*, 289–297 (1994).
3. P.W. Davies, F. Brink, 'Microelectrodes for Measuring Local Oxygen Tension in Animal Tissues', *Rev. Sci. Instrum.*, **13**, 524–533 (1942).
4. A.J. Bard, L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, J. Wiley & Sons, New York, 1980.
5. J. Newman, 'Resistance for Flow of Current to a Disk', *J. Electrochem. Soc.*, **113**, 501–502 (1968).
6. D.O. Wipf, A.C. Michael, R.M. Wightman, 'Microdisk Electrodes. Part II Fast Scan Cyclic Voltammetry with Very Small Electrodes', *J. Electroanal. Chem.*, **269**(1), 15–27 (1989).
7. L.R. Faulkner, M.R. Walsh, C. Xu, 'New Instrumental Approaches to Fast Electro-chemistry at Ultramicroelectrodes', in *Contemporary Electroanalytical Chemistry*, ed. A. Ivaska, Plenum Press, New York, 5–14, 1990.
8. W.J. Bowyer, E.E. Engelman, D.H. Evans, 'Kinetics by Cyclic Voltammetry at Low Temperatures Using Microelectrodes', *J. Electroanal. Chem.*, **262**(1–2), 67–83 (1989).
9. C. Amatore, B. Fosset, 'Equivalence Between Microelectrodes of Different Shapes: Between Myth and Reality', *Anal. Chem.*, **68**(24), 4377–4388 (1996).
10. M.I. Montenegro, 'Applications of Microelectrodes in Kinetics', *Research in Chemical Kinetics*, eds. R.G. Compton, G. Hancock, Elsevier, The Netherlands, 1, Vol. 2, 1994.
11. D. Garreau, P. Hapiot, J.-M. Savéant, 'Instrumentation for Fast Voltammetry at Ultramicroelectrodes. Stability and Bandpass Limitations', *J. Electroanal. Chem.*, **272**(1–2), 1–16 (1989).
12. R.S. Nicholson, I. Shain, 'Theory of Stationary Electrode Polarography. Single Scan and Cyclic Methods Applied to Reversible, Irreversible and Kinetic Systems', *Anal. Chem.*, **36**(4), 706–723 (1964).
13. R.A. Marcus, 'On the Theory of Electron-transfer Reactions. VI. Unified Treatment for Homogeneous and Electrode Kinetics', *J. Chem. Phys.*, **43**(2), 679–701 (1965).
14. J.O. Howell, R.M. Wightman, 'Ultra-fast Voltammetry and Voltammetry in Highly Resistive Solutions with Microvoltammetric Electrodes', *Anal. Chem.*, **56**(3), 524–529 (1984).
15. L.K. Safford, M.J. Weaver, 'The Evaluation of Rate Constants for Rapid Electrode-reactions Using Microelectrode Voltammetry–Virtues of Measurements at Lower Temperatures', *J. Electroanal. Chem.*, **331**(1–2), 857–876 (1992).
16. C.P. Andrieux, D. Garreau, P. Hapiot, J.Pinson, J.M. Savéant, 'Fast Sweep Cyclic Voltammetry at Ultramicroelectrodes. Evaluation of the Method for Fast Electron Transfer Kinetic Measurements', *J. Electroanal. Chem.*, **243**, 321–335 (1988).
17. M.I. Montenegro, D. Pletcher, 'The Determination of the Kinetics of Electron Transfer Using Fast Sweep Cyclic Voltammetry at Microdisk Electrodes', *J. Electroanal. Chem.*, **200**, 371–374 (1986).
18. C. Amatore, C. Lefrou, F. Pfluger, 'On-line Compensation of Ohmic Drop in Submicrosecond Time Resolved Cyclic Voltammetry at Ultramicroelectrodes', *J. Electroanal. Chem.*, **270**, 43–59 (1989).
19. C. Amatore, C. Lefrou, 'New Concept for a Potentiostat for Online Ohmic Drop Compensation in Cyclic Voltammetry Above 300 kV s^{-1}', *J. Electroanal. Chem.*, **324**(1–2), 33–58 (1992).
20. D.O. Wipf, R.M. Wightman, 'Rapid Cleavage Reactions of Haloaromatic Radical Anions Measured with Fast Scan Cyclic Voltammetry', *J. Phys. Chem.*, **93**, 4286–4291 (1989).
21. I. Lavagnini, P. Pastore, F. Magno, 'Application of Cyclic Voltammograms Under Mixed Spherical Semiinfinite Linear Diffusion at Microdisk Electrodes for Measurement of Fast Electrode-kinetics', *J. Electroanal. Chem.*, **333**(1–2), 1–10 (1992).
22. D.O. Wipf, E.W. Kristensen, M.R. Deakin, R.M. Wightman, 'Fast Scan Cyclic Voltammetry as a Method to Measure Rapid Heterogeneous Electron Transfer Kinetics', *Anal. Chem.*, **60**, 306–310 (1988).
23. H.F. Zhou, N.Y. Gu, S.J. Dong, 'Studies of Ferrocene Derivative Diffusion and Heterogeneous Kinetics in Polymer Electrolyte by Using Microelectrode Voltammetry', *J. Electroanal. Chem.*, **441**(1–2), 153–160 (1998).
24. A. Neudeck, J. Dittrich, 'The Determination of Diffusion Coefficients and Rate Constants from the Dependence of the Peak Separation and Peak Current on the Scan Rate of Cyclic Voltammograms at Micro-cylindrical Electrodes', *J. Electroanal. Chem.*, **313**(1–2), 37–59 (1991).

25. W. Huang, R. McCreery, 'Electron-transfer Kinetics of $[Fe(CN)_6]^{3-/4-}$ on Laser-activated and CN^- Modified Pt Electrodes', *J. Electroanal. Chem.*, **326**(1–2), 1–12 (1992).
26. N. Oyama, T. Ohsaka, N. Yamamoto, J. Matsui, O. Hatozaki, 'Determination of the Heterogeneous Electron Transfer Rate Constants for the Redox Couples Octacyanomolybdate $[Mo(CN)_8]^{4-/3-}$, Octacyanotungstate $[W(CN)_8]^{4-/3-}$, Hexacyanoferrate $[Fe(CN)_6]^{4-/3-}$, Hexacyanoosmate $[Os(CN)_6]^{4-/3-}$, Hexachloroiridate $[Ir(Cl)_6]^{3-/2-}$, Using Fast Sweep Cyclic Voltammetry at Carbon Fiber Electrodes', *J. Electroanal. Chem.*, **265**(1), 297–304 (1989).
27. C. Beriet, D. Pletcher, 'A Further Microelectrode Study of the Influence of Electrolyte Concentration on the Kinetics of Redox Couples', *J. Electroanal. Chem.*, **375**(1–2), 213–218 (1994).
28. K. Winkler, 'The Kinetics of Electron-transfer in $[Fe(CN)_6]^{4-/3-}$ Redox System on Platinum Standard-size and Ultramicroelectrodes', *J. Electroanal. Chem.*, **388**(1–2), 151–159 (1995).
29. D.O. Wipf, M.R. Wightman, 'Submicrosecond Measurements with Cyclic Voltammetry', *Anal. Chem.*, **60**, 2460–2464 (1988).
30. P. Hapiot, J. Moiroux, J.-M. Savéant, 'Electrochemistry of NADH–NAD^+ Analogs – A Detailed Mechanistic Kinetic and Thermodynamic Analysis of the 10-Methylacridan–10-Methylacridinium Couple in Acetonitrile', *J. Am. Chem. Soc.*, **112**(4), 1337–1343 (1990).
31. C.P. Andrieux, P. Hapiot, J.-M. Savéant, 'Fast Chemical Steps Coupled with Outer-sphere Electron Transfers: Application of Fast Scan Voltammetry at Ultramicroelectrodes to the Cleavage of Aromatic Halide Anion Radicals in the Microsecond Lifetime Range', *J. Phys. Chem.*, **92**(21), 5987–5992 (1988).
32. W.J. Bowyer, E.E. Engelman, D.H. Evans, 'Kinetic Studies by Cyclic Voltammetry at Low Temperatures Using Ultramicroelectrodes', *J. Electroanal. Chem.*, **262**(1–2), 67–82 (1989).
33. P. He, R.M. Crooks, L.R. Faulkner, 'Adsorption and Electrode-reactions of Disulfonated Anthrquinones at Mercury-electrodes', *J. Phys. Chem.*, **94**(3), 1135–1141 (1990).
34. C.E.D. Chidsey, 'Free-energy and Temperature-dependence of Electron-transfer at the Metal–Electrolyte Interface', *Science*, **251**(4996), 919–922 (1991).
35. D. Acevedo, H.D. Abruña, 'Electron-transfer Study and Solvent Effects on the Formal Potential of a Redox-active Self-assembling Monolayer', *J. Phys. Chem.*, **95**(23), 9590–9594 (1991).
36. R.J. Forster, L.R. Faulkner, 'Electrochemistry of Spontaneously Adsorbed Monolayers–Equilibrium Properties and Fundamental Electron-transfer Characteristics', *J. Am. Chem. Soc.*, **116**(20), 5444–5452 (1994).
37. H.O. Finklea, D.D. Hanshew, 'Electrochemistry of Spontaneously Adsorbed Monolayers – Effects of Solvent, Potential, and Temperature on Electron-transfer', *J. Am. Chem. Soc.*, **114**(20), 5453–5461 (1992).
38. E. Laviron, 'General Expression of the Linear Potential Sweep Voltammogram in the Case of Diffusionless Electrochemical Systems', *J. Electroanal. Chem.*, **101**(1), 19–28 (1979).
39. R.J. Forster, L.R. Faulkner, 'Interfacial Field Effects on Reductive Chloride Elimination from Spontaneously Adsorbed Monolayers', *Langmuir*, **11**(3), 1014–1023 (1995).
40. R.J. Forster, 'Heterogeneous Kinetics of Metal and Ligand Based Redox Reactions within Adsorbed Monolayers', *Inorg. Chem.*, **35**, 3394–3401 (1996).
41. J.P. O'Kelly, R.J. Forster, 'Potential Dependent Adsorption of Anthraquinone-2,7-disulfonate on Mercury', *Analyst*, **123**(10), 1987–1993 (1998).
42. A.A. Pilla, *Computers in Chemistry and Instrumentation*, Vol. 2, eds. J.S. Mattson, H.B. Mark, H.C. MacDonald, Marcel Dekker, New York, 1972.
43. C.A. Amatore, A. Jutand, F. Pflüger, 'Nanosecond Time-resolved Cyclic Voltammetry Direct Observation of Electrogenerated Intermediates with Biomolecular Diffusion Controlled Decays Using Scan Rates in the Megavolt per Second Range', *J. Electroanal. Chem.*, **218**(1–2), 361–365 (1987).
44. J.O. Howell, R.M. Wightman, 'Ultrafast Voltammetry of Anthracene and 9,10-Diphenylanthracene', *J. Phys. Chem.*, **88**(18), 3915–3922 (1989).
45. M. Murphy, Z. Stojek, J. O'Dea, J. Osteryoung, 'Pulse Voltammetry at Cylindrical Electrodes: Oxidation of Anthracene', *Electrochim. Acta.*, **36**(9), 1475–1484 (1991).
46. S. Okazaki, N. Oyama, S. Nomura, 'Correlation Between the Redox Potentials of 9-Substituted Anthracenes and the Results of PM3 Calculation', *Electroanalysis*, **9**(16), 1242–1246 (1997).
47. K.R. Wehmeyer, R.M. Wightman, 'Cyclic Voltammetry and Anodic Stripping Voltammetry with Mercury Microelectrodes', *Anal. Chem.*, **57**, 1989–1993 (1985).
48. R.S. Robinson, R.L. McCreery, 'Submicrosecond Spectroelectrochemistry Applied to Chlorpromazine Cation Radical Charge Transfer Reactions', *J. Electroanal. Chem.*, **182**(1), 61–72 (1985).
49. H. Yang, D.O. Wipf, A.J. Bard, 'Application of Rapid Scan Cyclic Voltammetry to a Study of the Oxidation and Dimerization of *N*,*N*-Dimethylaniline in Acetonitrile', *J. Electroanal. Chem.*, **331**(1–2), 913–924 (1992).
50. C. Amatore, J.N. Verpeaux, P.J. Krusic, 'Electrochemical Reduction of Iron Pentacarbonyl Revisited', *Organometallics*, **7**(11), 2426–2428 (1988).
51. C.J. Schlesener, C. Amatore, J.K. Kochi, 'Marcus Theory in Organic Chemistry. Mechanisms of Electron and Proton Transfer for Aromatics and Their Cation Radicals', *J. Phys. Chem.*, **90**, 3747–3756 (1986).
52. D.O. Wipf, R.M. Wightman, 'Voltammetry with Microvoltammetric Electrodes in Resistive Solvents Under

Linear Diffusion Conditions', *Anal. Chem.*, **62**(2), 98–102 (1990).
53. A.J. Bard, H.D. Abruña, C.E.D. Chidsey, L.R. Faulkner, S.W. Feldberg, K. Itaya, M. Majda, O. Melroy, R.W. Murray, M.D. Porter, M.P. Soriaga, H.S. White, 'The Electrode–Electrolyte Interface – A Status-report', *J. Phys. Chem.*, **97**(28), 7147–7173 (1993).
54. K. Aoki, K. Tokuda, H. Matsuda, 'Simulation in the Theory of Step Experiments at Microelectrodes', *J. Electroanal. Chem.*, **235**(1–2), 87–89 (1987).
55. M. Fleischmann, S. Bandyopadhyay, S. Pons, 'The Behavior of Microring Electrodes', *J. Phys. Chem.*, **89**(25), 5537–5541 (1985).
56. K.B. Oldham, C.G. Zoski, 'Simulation of Voltammetric Steady States at Hemispherical and Disk Microelectrodes', *J. Electroanal. Chem.*, **256**(1), 11–21 (1988).
57. M.V. Mirkin, A.J. Bard, 'Simple Analysis of Quasi-reversible Steady-state Voltammograms', *Anal. Chem.*, **64**(19), 2293–2302 (1992).
58. A. Russell, K. Repka, T. Dibble, J. Ghoroghchian, J. Smith, M. Fleischmann, C. Pitt, S. Pons, 'Determination of Electrochemical Heterogeneous Electron Transfer Reaction Rates from Steady-state Measurements at Ultramicroelectrodes', *Anal. Chem.*, **58**(14), 2961–2964 (1986).
59. C.G. Zoski, D.A. Sweigart, N.J. Stone, P.H. Rieger, E. Mocellin, T.F. Mann, D.R. Mann, D.K. Gosser, M.M. Doeff, A.M. Bond, 'An Electrochemical Study of the Substitution and Decomposition Reactions of (Arene) Tricarbonyl Chromium Radical Cations', *J. Am. Chem. Soc.*, **110**(7), 2109–2116 (1988).
60. R.M. Penner, M.J. Heben, T.L. Longin, N.S. Lewis, 'Fabrication and Use of Nanometer-sized Electrodes in Electrochemistry', *Science*, **250**(4984), 1118–1121 (1990).
61. F.N. Büchi, A.M. Bond, 'Interpretation of the Electrochemistry of Cytochrome-*c* at Macro-sized and Micro-sized Carbon Electrodes Using a Microscopic Model Based on a Partially Blocked Surface', *J. Electroanal. Chem.*, **314**(1–2), 191–206 (1991).
62. A. Owlia, J.F. Rusling, 'Non-linear Regression Analysis of Steady-state Voltammograms Obtained at Microelectrodes', *Electroanalysis*, **1**(2), 141–149 (1989).
63. A.M. Bond, T.L.E. Henderson, D.R. Mann, T.F. Mann, W. Thormann, C.G. Zoski, 'Fast Electron Transfer Rate for the Oxidation of Ferrocene in Acetonitrile or Dichloromethane at Platinum Disk Ultra-microelectrodes', *Anal. Chem.*, **60**(18), 1878–1882 (1988).
64. J. Daschbach, D. Blackwood, J.W. Pons, S. Pons, 'Electrochemistry of Ferrocene in Acetonitrile. Evidence for Irreversible Kinetic due to Passive Film Formation', *J. Electroanal. Chem.*, **237**(2), 269–275 (1987).
65. Y. Zhang, C.D. Baer, C. Camaioni-Neto, P. O'Brien, D.A. Sweigart, 'Steady-state Voltammetry with Microelectrodes – Determination of Heterogeneous Charge-transfer Rate Constants for Metalloporphyrin Complexes', *Inorg. Chem.*, **30**(8), 1682–1685 (1991).
66. K.B. Oldham, C.G. Zoski, A.M. Bond, D.A. Sweigart, 'Measurement of Ultrafast Electrode Kinetics via Steady-state Voltammograms at Microdisk Electrodes', *J. Electroanal. Chem.*, **248**(2), 467–473 (1988).
67. B.R. Scharifker, P. Zelenay, J. O'M Bockris, 'The Kinetics of Oxygen Reduction in Molten Phosphoric Acid at High Temperatures', *J. Electrochem. Soc.*, **134**(11), 2714–2725 (1987).
68. A.J. Bard, M.V. Mirkin, P.R. Unwin, D.O. Wipf, 'Scanning Electrochemical Microscopy.12. Theory and Experiment of the Feedback Mode with Finite Heterogeneous Electron-transfer Kinetics and Arbitrary Substrate Size', *J. Phys. Chem.*, **96**(4), 1861–1868 (1992).
69. C. Amatore, S.C. Paulson, H.B. White, 'Successive Electron-transfers in Low Ionic Strength Solutions. Migration Flux Coupling by Homogeneous Electron Transfer Reactions', *J. Electroanal. Chem.*, **439**(1), 173–182 (1997).
70. M. Fleischmann, F. Lassere, J. Robinson, 'The Application of Microelectrodes to the Study of Homogeneous Processes Coupled to Electrode Reactions. Part II ECE and DISP1 Reactions', *J. Electroanal. Chem.*, **177**(1–2), 115–127 (1984).
71. S. Dong, G. Che, 'Electrocatalytic Oxidation of Ascorbic-acid at a Prussian Blue Film Modified Microdisk Electrode', *J. Electroanal. Chem.*, **315**(1–2), 191–199 (1991).
72. T.V. Shea, A.J. Bard, 'Digital Simulation of Homogeneous Chemical Reactions Coupled to Heterogeneous Electron Transfer and Applications at Platinum/Mica/Platinum Ultramicroband Electrodes', *Anal. Chem.*, **59**, 2101–2111 (1987).
73. C.L. Miaw, J.F. Rusling, A. Owlia, 'Simulation of 2-electron Homogeneous Electrocatalysis for Steady-state Voltammetry at Hemispherical Microelectrodes', *Anal. Chem.*, **62**(3), 268–273 (1990).
74. S.U. Pedersen, K. Daasbjerg, 'Ultramicroelectrodes for Electrochemical Monitoring of Homogeneous Reactions', *Acta Chem. Scand.*, **43**, 301–303 (1989).
75. K. Daasbjerg, S.U. Pedersen, H. Lund, 'Electrochemical Measurements of Rate Constants for the Electron Transfer Reaction to Sterically Hindered Alkyl Halides', *Acta Chem. Scand.*, **43**, 876–881 (1989).
76. M. Fleischmann, F. Lassere, J. Robinson, D. Swan, 'The Application of Microelectrodes to the Study of Homogeneous Processes Coupled to Electrode Reactions. Part II ECE and DISP1 Reactions', *J. Electroanal. Chem.*, **177**(1–2), 97–115 (1984).
77. Dong, G. Che, 'Application of Ultramicroelectrodes in Studies of Homogeneous Catalytic Reactions. 2. A Theory of Quasi-1st and 2nd-order Homogeneous Catalytic Reactions', *Electrochim. Acta*, **37**(15), 2695–2699 (1992).
78. S. Daniele, P. Ugo, G.A. Mazzocchin, 'The Use of Microelectrodes for Studying the Processes Involved in 1-naphthylamine Oxidation in Dimethyl Sulfoxide', *J. Electroanal. Chem.*, **267**(1–2), 129–140 (1989).

79. R. Muller, L. Lamberts, M. Evers, 'The Electrochemical Oxidation of Thioselenanthrene in Acetonitrile at Conventional Electrodes and Microelectrodes', *J. Electroanal. Chem.*, **417**(1–2), 35–43 (1996).
80. R. Muller, L. Lamberts, M. Evers, 'The Electrochemical Oxidation of Dibenzo(C,E)-1,2-Diselenine to its Cation Radical. A Voltammetric Study in Acetonitrile at Conventional Electrodes and Microelectrodes', *J. Electroanal. Chem.*, **401**(1–2), 183–189 (1996).
81. C.P. Andrieux, P. Hapiot, J.-M. Savéant, 'Fast Kinetics by means of Direct and Indirect Electrochemical Techniques', *Chem. Rev.*, **90**(5), 723–738 (1990).
82. E. Oguntoye, S. Szunertis, J.H.P. Utley, P.B. Wyatt, 'Electro-organic Reactions. 46. Diels–Alder Trapping of *O*-Quinodimethane Generated by Redox-mediated Cathodic Reduction of α, α'-Dibromo-*O*-xylene in the Presence of Hindered Dienophiles', *Tetrahedron*, **52**(22), 7771–7778 (1996).
83. D.H. Turner, G.W. Flynn, N. Sutin, J.V. Beitz, 'Laser Raman Temperature-jump Study of the Kinetics of the Triiodide Equilibrium. Relaxation Times in the 10^{-8}–10^{-7} Second Range', *J. Am. Chem. Soc.*, **94**(5), 1554–1562 (1972).
84. G.C. Barker, A.W. Gardner, G. Bottura, 'Laser Induced Potential Changes at a Mercury Electrode', *J. Electroanal. Chem.*, **45**(1), 21–30 (1973).
85. V.A. Benderskii, G.I. Velichko, 'Temperature Jump in Electric Double-layer Study. Part I. Method of Measurements', *J. Electroanal. Chem.*, **140**, 1–22 (1982).
86. H. Coufal, P. Hefferle, 'Photothermal Analysis of Thin Films', *Springer Ser. Chem. Phys*, **39**, 497–501 (1984).
87. J.F. Smalley, C.V. Krishnan, M. Goldman, S.W. Feldberg, I. Ruzic, 'Laser Induced Temperature Jump Coulostatics for the Investigation of Heterogeneous Rate Processes. Theory and Application', *J. Electroanal. Chem.*, **248**(2), 255–283 (1988).
88. J.F. Smalley, S.W. Feldberg, C.E.D. Chidsey, M.R. Linford, M.D. Newton, Y.-P. Liu, 'The Kinetics of Electron Transfer Through Ferrocene-terminated Alkane Thiol Monolayers on Gold', *J. Phys. Chem.*, **99**(35), 13 141–13 149 (1995).
89. A.J. Bard, F.-R.F. Fan, M.V. Mirkin, *Electroanalytical Chemistry*, Vol. 18, ed. A.J. Bard, Marcel Dekker, New York, 243, 1993.
90. M.V. Mirkin, L.O.S. Bulhões, A.J. Bard, 'Determination of the Kinetic-parameters for the Electroreduction of C_{60} by Scanning Electrochemical Microscopy and Fast Scan Cyclic Voltammetry', *J. Am. Chem. Soc.*, **115**(1), 201–204 (1993).
91. F. Zhou, P.R. Unwin, A.J. Bard, *J. Phys. Chem.*, **96**, 4917 (1992).
92. D.A. Treichel, M.V. Mirkin, A.J. Bard, 'Scanning Electrochemical Microscopy. 27. Application of a Simplified Treatment of an Irreversible Homogeneous Reaction Following Electron-transfer to the Oxidative Dimerization of 4-Nitrophenolate in Acetonitrile', *J. Phys. Chem.*, **98**(22), 5751–5757 (1994).

Ultraviolet/Visible Spectroelectrochemistry

Daniel A. Scherson, Yuriy V. Tolmachev, and Ionel C. Stefan
Case Western Reserve University, Cleveland, USA

1 Introduction 1
2 Theoretical Considerations 2
2.1 Optics 2
2.2 Mass Transport 3
3 Experimental Considerations 7
4 Applications 8
4.1 Transmission Thin-layer Spectroelectrochemistry 9
4.2 Transmission Spectroelectrochemistry in Semi-infinite Media 17
4.3 Internal Reflection 18
4.4 External Reflection 20
4.5 Spatially Resolved Spectroelectrochemistry 30
4.6 Ultraviolet/Visible Spectroelectrochemistry in the Presence of Convective Flow 34
4.7 Modulation Techniques 43
Concluding Remarks 50
Acknowledgments 51
Abbreviations and Acronyms 51
Related Article 51
References 51

This article reviews the field of solution-phase ultraviolet/visible (UV/VIS) absorption spectroelectrochemistry and its application to problems in thermodynamics, kinetics, and mass transport, with emphasis on both experimental and theoretical aspects. Examples are provided for transmission in thin layer and semi-infinite media, using optically transparent electrodes, and external and internal reflection in quiescent media. Also illustrated is the coupling of UV/VIS spectroscopy to systems under forced convection, such as rotating disk and channel electrodes, under steady state and transient conditions. Attention is also focused on spatially resolved spectroelectrochemistry for the imaging of diffusion and reaction layers. The last section is devoted to the more complex and particularly powerful modulation techniques involving diffraction, refraction and absorption in stagnant media and in the presence of convective flow. Factors that limit the sensitivity and spatial and temporal resolution of UV/VIS absorption spectroelectrochemistry, as well as its future prospects in analytical chemistry are briefly discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Light in the UV/VIS range, i.e. 190–700 nm, can promote excitation of electronic states in atomic and molecular species in gas and solution phases, and also in condensed matter, including crystalline solids and pure liquids. As such, it represents a valuable probe of the structure and properties of materials.[1–3] UV/VIS spectroelectrochemistry refers collectively to a wide array of techniques that employ radiation in this frequency domain for the study of the optical properties of electrodes and electrolyte solutions, particularly those induced by changes in the applied potential across electrode–solution interfaces.[4–8] Implicit in this definition is the fact that measurements are performed in situ, i.e. with the electrode immersed in the solution under potential control. Particularly amenable to UV/VIS spectroelectrochemical investigation are electrode processes that generate soluble species that absorb radiation in this spectral region, elicit spatial variations in the index of refraction, or modify the optical properties of interfaces. The dependence of the molar absorptivity on the wavelength of the probing beam, which is often specific to each and every chromophore, provides an added dimension for the identification and monitoring of species involved in electrochemical reactions as compared to simple current–potential relationships.

This article reviews the field of solution phase UV/VIS absorption spectroelectrochemistry and its application to problems in kinetics, mechanisms and mass transport. Within this rather narrow scope, it excludes equally important areas that rely on the analysis of light reflected from the electrode surface, such as ellipsometry[9,10] and electroreflectance,[11,12] as well as simple refraction, notably probe beam deflection[13] and interferometry,[14] for the study of interfacial and bulk phenomena. As the vast majority of heterogeneous electron transfer reactions involve generation or depletion of species in solution, diffusion plays an important role in controlling the current that flows across the interface and, thus, the time evolution of concentration profiles.[15] Hence, it is of key importance to highlight at the outset fundamental aspects of the two most relevant physical phenomena, namely, the interaction of light and matter, and mass transport of species in condensed media, to gain a full appreciation of the various factors associated with the design and interpretation of spectroelectrochemical experiments.

2 THEORETICAL CONSIDERATIONS

2.1 Optics

2.1.1 Light Propagation through Single-phase Materials

The optical properties of an isotropic nonmagnetic media are defined in terms of its complex refractive index $\hat{n} = n - ik$, or its complex dielectric function, $\hat{\varepsilon} = \varepsilon' - i\varepsilon''$, where the index of refraction n and the extinction coefficient k are both functions of the wavelength λ, and $\varepsilon' = (n^2 - k^2)$ and $\varepsilon'' = 2nk$.[16,17] Modifications in the composition of an otherwise homogeneous solution, such as those derived from electrode processes, will often bring about variations in the local magnitudes of n and k, so that the intensity, as well as the direction of light propagating through the media may be altered. These effects can be exploited to obtain time- and space-resolved concentration profiles, allowing important aspects of electrochemical systems to be investigated, including identification of products of redox reactions, measurements of diffusion coefficients, elucidation of reaction mechanisms, and determination of the rates of both heterogeneous and homogeneous electron transfer reactions involving electrogenerated species.

Correlations between n and k, and the concentration of a single species in solution are usually linear, a factor that simplifies considerably the analysis of experimental data. Furthermore, if ionic migration in an electric field is sufficiently minimized, for example by using an excess of supporting electrolyte, mass transport under either quiescent or laminar flow conditions can be approximated by relatively simple differential equations, allowing solutions to be expressed in terms of common analytic functions. It is precisely the interaction of light with such spatially and temporally varying concentration fields that constitutes the basis for a rigorous mathematical treatment of spectroelectrochemical experiments that rely on absorption, refraction, and diffraction of light. In view of the nature of the material being reviewed, it seems appropriate to focus general attention on light absorption, and defer discussion of refraction and diffraction to those few specialized sections where methods based on these phenomena are introduced.

2.1.2 Absorption

The instantaneous attenuation in the intensity I of a collimated light beam propagating along an axis y, $-dI$, through a media of thickness dy, containing, without loss of generality, a single absorbing species, is most often proportional to the intensity of the light incident on the infinitesimal volume, $dx\,dy\,dz$, $I(x, y, z, t)$, to the local concentration $c(x, y, z, t)$ and to the molar absorptivity of that species $\varepsilon(\lambda)$:

$$-dI(x, y, z, t) = I(x, y, z, t)\kappa(\lambda)c(x, y, z, t)\,dy \tag{1}$$

In Equation (1) $\kappa(\lambda) = k\varepsilon(\lambda)$, $k = 2.303$, and λ is the wavelength of light at which the measurements are being carried out. The intensity of the light emerging from a cell of length d along y, $I(x, d, z, t)$, is given by Equation (2),

$$I(x, d, z, t) = I(x, 0, z)\exp\left\{-\int_0^d \kappa(\lambda)c(x, y, z, t)\,dy\right\} \tag{2}$$

where, for simplicity, the intensity of the beam incident on the cell, $I(x, 0, z)$ has been assumed independent of time. Equation (2) can be integrated over the entire cross-sectional area of the beam xz, to yield Equation (3)

$$\int_x\int_z I(x, d, z, t)\,dx\,dz = \int_x\int_z I(x, 0, z)\exp\left\{-\int_0^d \kappa(\lambda) \times c(x, y, z, t)\,dy\right\} dx\,dz \tag{3}$$

If absorption is weak,[18] as is assumed throughout this article, the exponential function can be approximated by the first two terms in its Taylor's series expansion. On this basis, the instantaneous absorbance $A(t)$, i.e. the log (base 10) of the ratio of the light intensity incident on, and that emerging from the cell, is expressed by Equation (4):

$$A(t) = \frac{\int_x\int_z I(x, 0, z)\left\{\varepsilon(\lambda)\int_0^d c(x, y, z, t)\,dy\right\} dx\,dz}{\int_x\int_z I(x, 0, z)\,dx\,dz} \tag{4}$$

In cases in which the concentration is only a function of y and t, $A(t)$ reduces to the much simpler Equation (5):

$$A(t) = \varepsilon\int_0^d c(y, t)\,dy \tag{5}$$

where the explicit dependence of ε on λ will, hereafter, be omitted. This expression specifies that the intensity of the light emerging from the cell is proportional to the integral of the concentration profile along y, and would be applicable for measurements in which a beam of light is incident normal to a semitransparent, planar electrode placed in the xz plane (Figure 1a). Alternatively, the beam could propagate through the solution and reflect off the surface of a highly polished electrode at an angle strictly smaller than 90° (Figure 1b). This geometry will increase the effective optical path compared to the transparent

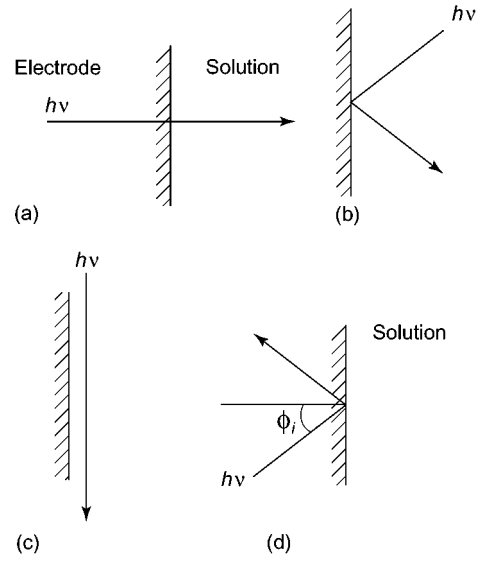

Figure 1 Experimental arrangements for UV/VIS spectroelectrochemical experiments: (a) optically transparent electrode transmission; (b) external reflection; (c) parallel geometry; (d) internal reflection.

electrode, which will appear as a correction factor to Equation (5) (see section 4.4.1).

A far more powerful strategy involves the use of a well-collimated beam of monochromatic light of finite width propagating parallel to and very close to the surface of a planar electrode (Figure 1c). This can then be imaged onto a photodiode array detector, for example, to yield spatially resolved profiles of absorbing species.

It becomes evident from these arguments, that a quantitative analysis of dynamic spectroelectrochemical experiments involving solution-phase chromophores requires time dependent concentration profiles to be calculated. Such problems are inextricably linked to finding solutions of the equations that govern the mass transport of species in solution, subject to the appropriate boundary conditions, especially those prescribed at the electrode surface.[15]

2.1.3 Internal Reflection

Consider a beam of light propagating in the xy plane in phase 1 with index of refraction n_1 impinging onto the plane xz that separates phase 1 from another phase 2 with different index of refraction n_2, at an angle of incidence ϕ_i with respect to the normal to the interface (Figure 1d). Classical optics predicts that ϕ_i is related to the angle the transmitted light will make with respect to the same plane, ϕ_t, through the familiar Snell's law (Equation 6),

$$\sin\phi_t = \frac{\sin\phi_i}{n_{12}} \tag{6}$$

where $n_{12} = n_2/n_1$. If $n_2 < n_1$ and $\sin\phi_i > n_{12}$, ϕ_t becomes imaginary leading physically to a condition known as total internal reflection.[16] Based on purely trigonometric arguments, Equation (6) can be rewritten as

$$\cos\phi_t = \pm i \quad \overline{\frac{\sin^2\phi_i}{n_{12}^2} - 1} \tag{7}$$

Based on Equations (6) and (7), the amplitude of the electric field, E, may be shown to be proportional to Equation (8):

$$e^{-i\omega\left(t-((x\sin\phi_i)/(n_{12}v_2))\right)}e^{-((\omega y)/(v_2))\ \sqrt{(\sin^2\phi_i)/(n_{12}^2)-1}} \tag{8}$$

where $\omega = 2\pi v_j/\lambda_j$ and v_j and λ_j are the velocity and wavelength of light in medium j, respectively. This expression represents a wave travelling along the x-axis within the interfacial plane, for which its intensity decreases exponentially along the y-axis, i.e. normal to the plane. The distance at which the magnitude of the electric field E decreases to 1/e of its value at the precise interface is known as the depth of penetration δ (Equation 9):

$$\frac{E_\delta}{E_o} = \frac{1}{e} = e^{-((\omega\delta)/(v_2))\ \sqrt{(\sin^2\phi_i)/(n_{12}^2)-1}} \tag{9}$$

Hence, in terms of $\omega = 2\pi v_j/\lambda_j$,

$$\delta = \frac{\lambda_2}{2\pi\ \overline{\frac{\sin^2\phi_i}{n_{12}^2} - 1}} = \frac{\lambda_1}{2\pi\sqrt{\sin^2\phi_i - n_{12}^2}} \tag{10}$$

where $\lambda_1 = n_{12}\lambda_2$ in Equation (10).

This evanescent wave can be used to probe species present at, and in the near vicinity of the interface, as has been widely popularized in the infrared spectral region under the acronym of attenuated total reflection (ATR).[19] It thus follows from Equation (10) that λ and δ are of the same order of magnitude, i.e. a few hundred nanometers in the UV/VIS region.

2.2 Mass Transport

2.2.1 Quiescent Solutions

In the absence of complications derived from ionic migration in an electric field, natural convection, and homogeneous chemical reactions, mass transport in stagnant solutions is governed to a good degree of approximation by Fick's second law,[20]

$$\frac{\partial c}{\partial t} = D\nabla^2 c \tag{11}$$

where D is the diffusion coefficient of the species in question. The symbol ∇^2 in Equation (11) represents the

Laplacian, which in cartesian coordinates is given by Equation (12):

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{12}$$

For measurements performed over relatively short periods of time, as is often the case, the changes in concentration of reactants and products are confined to a small volume of solution close to the electrode surface. Under these conditions, the composition of the media far away from the electrode may be assumed to remain unaltered during the entire data acquisition period, and thus the actual size of the container may be regarded as being arbitrarily large.

The theoretical treatment of a vast number of UV/VIS spectroelectrochemical experiments involving planar, cylindrical and spherical electrodes reduces to finding solutions of Equation (11) in one dimension under the initial and boundary conditions specified in the left column of Table 1, where C_R is the concentration of an electroactive reactant R, and the superscript 'o' represents its bulk value. As will be illustrated for a few simple cases in this article, the Laplace transform method[21] affords a particularly powerful tool for solving problems of this type.[15,22]

It is convenient from a computational viewpoint to define a dimensionless concentration, i.e.

$$\theta_R(y, t) = \frac{C_R^o - C_R}{C_R^o} \tag{13}$$

Substitution of Equation (13) into Equation (11) renders a similar differential equation subject to initial and boundary conditions shown in the center column in Table 1, which do not depend explicitly on the value of C_R^o.

Subsequent application of the Laplace transform with respect to t reduces the problem to solving a simple linear differential equation, as prescribed in the third column in that same table, for which a solution can be easily obtained to yield

$$\bar{\theta}_R(y, s) = \bar{\theta}_R(0, s)e^{-(s/D_R)^{1/2}y} \tag{14}$$

In Equation (14) the bar above indicates that the function has been Laplace transformed, and $\bar{\theta}_R(0, s)$ is the concentration of the reactant at the boundary for $t > 0$ in Laplace space. As will be shown, the quantitative analysis of a variety of electrochemical techniques requires for $\bar{\theta}_R(0, s)$ to be explicitly specified.

Many of the systems to be described in this article involve simple reactions of the Equation (15) type:

$$R \longrightarrow P \tag{15}$$

where, unless otherwise stated, both R and P are solution phase species. A relationship between the concentration profiles of R and P can be obtained from conservation of mass at the surface (Equation 16)

$$D_R \left.\frac{\partial C_R(y, t)}{\partial y}\right|_{y=0} = -D_P \left.\frac{\partial C_P(y, t)}{\partial y}\right|_{y=0} \tag{16}$$

The terms on the right and left of this equation represent the fluxes of P and R through the electrode, respectively, bearing opposite signs, as required.

2.2.1.1 Chronocoulometry Consider a planar electrode embedded in the xz plane immersed in a quiescent, homogeneous solution containing an electrochemically active species R, and assume a potential step is applied to the electrode of a magnitude large enough to reduce C_R to zero at the surface. This experimental protocol constitutes the basis of chronocoulometry, an electrochemical technique in which the current generated following the potential step is monitored as a function of time. The time evolution of the concentration profile for R in this case may be obtained by setting $\theta_R(0, t) = 1$. Inserting its Laplace transform, $\bar{\theta}_R(0, s) = 1/s$, in Equation (14) gives Equation (17),

$$\bar{\theta}_R(y, s) = \frac{1}{s}e^{-(s/D_R)^{1/2}y} \tag{17}$$

which yields Equation (18) upon Laplace inversion:

$$C_R = C_R^o \mathrm{erf}\left(\frac{y}{2\sqrt{D_R t}}\right. \tag{18}$$

where $\mathrm{erf}(z)$ is the error function of argument z.[23]

Table 1 Differential equations and boundary conditions for mass transport of species in quiescent media in real and Laplace spaces

Real space		Laplace space
$\dfrac{\partial C_R(y, t)}{\partial t} = D_R \dfrac{\partial^2 C_R(y, t)}{\partial y^2}$	$\dfrac{\partial \theta_R(y, t)}{\partial t} = D_R \dfrac{\partial^2 \theta_R(y, t)}{\partial y^2}$	$D_R \dfrac{d^2 \bar{\theta}_R(y, s)}{dy^2} = s\bar{\theta}_R(y, s)$
$C_R(y, 0) = C_R^o$	$\theta_R(y, 0) = 0$	$\bar{\theta}_R(y, s) = 0$
$C_R(\infty, t) = C_R^o$	$\theta_R(\infty, t) = 0$	$\bar{\theta}_R(\infty, s) = 0$

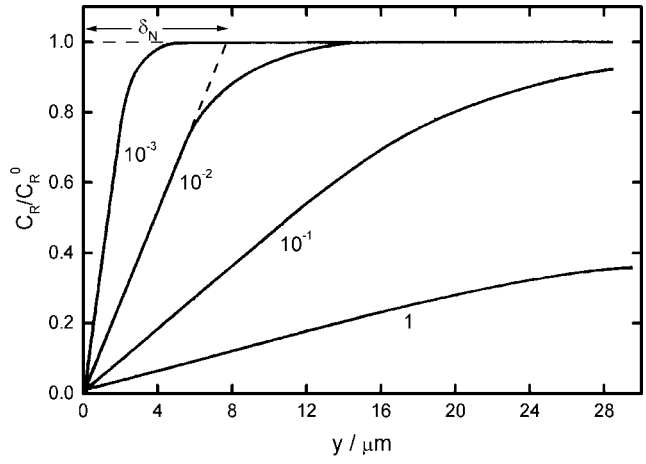

Figure 2 Time evolution of the concentration profile for a reactant R following application of a potential step to a planar electrode based on Equation (18) for $D_R = 10^{-5}$ cm^2 s^{-1}. Each of the curves was calculated for the specified times in units of seconds.[73]

A series of C_R/C_R^o versus y plots show that for fixed values of t, the concentration near the surface is proportional to the distance from the electrode (Figure 2). Linear extrapolation to the bulk concentration, i.e. $C_R/C_R^o = 1$, as indicated in that figure defines an imaginary region known as the Nernst diffusion layer. This simple diffusional theory predicts that, as time elapses, the thickness of this layer δ_N extends progressively into the bulk solution. In practice, however, variations in concentration give rise to differences in the density of the media, and thus to the onset natural convection, a phenomenon not considered by Fick's second law, leading ultimately to the formation of a rather static boundary layer.[24]

If the product of the reaction P is not present at the beginning of the experiment, its time-varying concentration profile can be calculated in a very similar way by defining an appropriate dimensionless concentration (Equation 19),

$$\theta_P(y, t) = -\frac{C_P}{C_R^o} \tag{19}$$

rendering a problem identical to that specified for R in Table 1. In Laplace space this gives Equation (20),

$$\bar{\theta}_P(y, s) = \bar{\theta}_P(0, s)e^{-(s/D_P)^{1/2}y} \tag{20}$$

where $\bar{\theta}_P(0, s)$ is the dimensionless concentration of P at the electrode surface. The latter can be obtained from Equation (16) in Laplace space, as the additional boundary condition, which in terms of dimensionless concentrations reads

$$\bar{\theta}_P(0, s) = -\bar{\theta}_R(0, s)\left(\frac{D_R}{D_P}\right)^{1/2} \tag{21}$$

Insertion of Equation (21) into Equation (20) and replacing $\bar{\theta}_R(0, s)$ by $1/s$ yields Equation (22),

$$\bar{\theta}_P(y, s) = -\frac{1}{s}\left(\frac{D_R}{D_P}\right)^{1/2} e^{-(s/D_P)^{1/2}y} \tag{22}$$

which upon subsequent Laplace inversion gives the time-dependent concentration profile of P for a chronocoulometric-type experiment (Equation 23):

$$C_P(y, t) = \left(\frac{D_R}{D_P}\right)^{1/2} C_R^o \operatorname{erfc}\left(\frac{y}{2\sqrt{D_P t}}\right) \tag{23}$$

where $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$ is the complementary error function of argument z.[23]

An explicit expression for the flux of P at the surface can be obtained from Equation (20) as Equation (24),

$$-D_P\left(\frac{\partial\bar{\theta}_P(y, s)}{\partial y}\right)_{x=0} = \sqrt{D_P s}\,\bar{\theta}_P(0, s) \tag{24}$$

a quantity proportional to the charge that flows across the electrode per unit time due to a heterogeneous electron transfer reaction, or faradaic current.

The total amount of material produced by the electrode reaction up to time t is then proportional to the integral of the flux with respect to t, which in Laplace space is equivalent to simply dividing Equation (24) by s, to yield a quantity denoted as $Q(s)$.

$$Q(s) = \sqrt{\frac{D_P}{s}}\,\bar{\theta}_P(0, s) \tag{25}$$

Equation (25) is also obtained by integrating the dimensionless concentration profile along the entire y-axis, a quantity proportional to the absorbance (see Equation 5), provided P is the only absorbing species in the media (Equation 26):

$$\int_0^\infty \bar{\theta}_P(y, s)\,dy = \int_0^\infty \bar{\theta}_P(0, s)e^{-(s/D_R)^{1/2}y}\,dy = \sqrt{\frac{D_P}{s}}\,\bar{\theta}_P(0, s) \tag{26}$$

Such an identity should not be surprising, as the total amount of solution-phase material produced by the electrode, is indeed directly proportional to that calculated based on the flux at the interface. However, no such rigorous relationship exists involving the *measured* charge, as in many cases the current may contain nonfaradaic contributions derived primarily from interfacial double layer capacitive effects.

2.2.1.2 Coupled Chemical Reactions Many electrochemical processes of fundamental and technological interest involve generation of species that are

either intrinsically unstable, i.e. undergo spontaneous decomposition, or are capable of reacting with other solution-phase species. The latter include redox-mediated catalysis in which reduction or oxidation of the reactant R at an electrode produces P (Equation 27), which in turn reacts with Q via homogeneous electron transfer to yield a desired product Z, regenerating R (Equation 28):

$$\mathrm{R} \longrightarrow \mathrm{P} \tag{27}$$

$$\mathrm{P} + \mathrm{Q} \xrightarrow{k} \mathrm{Z} + \mathrm{R} \tag{28}$$

A mathematical analysis of this reaction sequence requires that the appropriate diffusion equations for R and P are solved simultaneously. Under conditions in which there is great excess of Q in solution, the homogeneous electron transfer reaction may be assumed to be pseudo first order in P. Furthermore, if both Q and Z are electrochemically inactive in the potential range in which R is either oxidized or reduced, the system of differential equations governing the overall process may be shown to be given by Equations (29) and (30),

$$\frac{\partial C_\mathrm{R}}{\partial t} = D_\mathrm{R}\frac{\partial^2 C_\mathrm{R}}{\partial y^2} + \beta^2 C_\mathrm{P} \tag{29}$$

$$\frac{\partial C_\mathrm{P}}{\partial t} = D_\mathrm{P}\frac{\partial^2 C_\mathrm{P}}{\partial y^2} - \beta^2 C_\mathrm{P} \tag{30}$$

where $\beta^2 = kC_\mathrm{Q}$ is a pseudo-first-order rate constant. A solution to the concentration step problem for P in Laplace space, assuming $D_\mathrm{R} = D_\mathrm{P} = D$, may be obtained by the same techniques introduced above,[25] yielding Equation (31):

$$\bar{\theta}_\mathrm{P}(y, s) = \mathrm{e}^{-([s+\beta^2]/D)^{1/2}y} \tag{31}$$

where $\bar{\theta}_\mathrm{P}(y, s) = \overline{C}_P/C^\mathrm{o}_\mathrm{R}$ and $\overline{C}_\mathrm{P}$ is the Laplace transform of C_P.

2.2.2 *Diffusion in the Presence of Convective Flow*

The relationship that governs convective diffusion of solution-phase species that do not participate in any homogeneous phase reactions may be written in general as Equation (32):

$$\frac{\partial c}{\partial t} = D\nabla^2 c - \mathbf{v}\nabla c \tag{32}$$

where $\mathbf{v}$ is a vector that represents the fluid velocity. Considerable mathematical simplifications to this rather formidable problem can be obtained provided that two conditions are fulfilled:[26]

1. The thickness of the diffusion boundary layer must be small compared to the thickness of the hydrodynamic boundary layer, i.e. that region of the solution in which velocity gradients occur, so that the velocity components along the relevant axes can be approximated by the first terms of their power series expansion along an axis normal to the electrode surface.
2. Convection, rather than diffusion, is the predominant mode of mass transport along the axis parallel to the electrode surface in the direction of fluid flow.

Under the same conditions specified for diffusion in quiescent media (section 2.2.1), the convective diffusion equations for a rotating disk (and ring-disk), channel-type, and tube-type geometries under steady state are given by Equations (33) in Table 2.[27] Solutions for transient convective diffusion for rotating disk and channel-type electrodes are discussed in section 4.

The dimensionless variables defined in Equations (34–36) enable a single equation to be obtained that is applicable to a reactant (and under certain conditions to a product[27]) for all three electrode arrangements (see Equation 37), subject to the boundary conditions specified in Equations (38) and (39) in that same table. By taking the Laplace transform of Equations (37) and (39) with respect to X, denoted as L_X, and using Equation (38), yields Equations (40) and (41):

$$s\bar{\theta}(s, Y) = \frac{1}{Y}\frac{\partial^2 \bar{\theta}(s, Y)}{\partial Y^2} \tag{40}$$

$$\lim_{Y\to\infty} \bar{\theta}(s, Y) = 0 \tag{41}$$

where $L_X\{\theta(X, Y)\} = \bar{\theta}(s, Y)$

The general solution to this problem may be written as Equation (42),

$$\bar{\theta}(s, Y) = A_1(s)\mathrm{Ai}(s^{1/3}Y) \tag{42}$$

where $\mathrm{Ai}(z)$ is the Airy function of argument z.[23]

At $Y = 0$ Equation (43) holds

$$\bar{\theta}(s, 0) = \frac{A_1(s)}{3^{1/3}\Gamma(2/3)} \tag{43}$$

Hence, $A_1(s) = 3^{2/3}\Gamma(2/3)\bar{\theta}(s, 0)$, and therefore,

$$\bar{\theta}(s, Y) = 3^{2/3}\Gamma\left(\frac{2}{3}\right.\ \bar{\theta}(s, 0)\mathrm{Ai}(s^{1/3}Y) \tag{44}$$

From Equation (44), and recalling Equation (45),

$$\left.\frac{\partial \mathrm{Ai}(s^{1/3}Y)}{\partial Y}\right|_{Y=0} = -\frac{3^{-1/3}}{\Gamma(1/3)}s^{1/3} \tag{45}$$

the flux at the surface is given by

$$\left.\frac{\partial \bar{\theta}(s, Y)}{\partial Y}\right|_{Y=0} = -3^{1/3}\frac{\Gamma(2/3)}{\Gamma(1/3)}\bar{\theta}(s, 0)s^{1/3} \tag{46}$$

Table 2 Steady-state convective diffusion equations, dimensionless variables, boundary conditions and universal dimensionless differential equation for rotating disk, channel and tube electrodes

Rotating disk	Channel	Tube	
Steady-state convective diffusion equation			
$ay\Omega\left(\frac{\Omega}{\nu}\right)^{1/2}\left\{r\frac{\partial C_i}{\partial r}-y\frac{\partial C_i}{\partial y}\right\}=D_i\frac{\partial^2 C_i}{\partial y^2}$	$2v_o\frac{y}{h}\frac{\partial C_i}{\partial x}=D_i\frac{\partial^2 C_i}{\partial y^2}$	$2v_o\left(1-\frac{r}{R}\right)\frac{\partial C_i}{\partial x}=D_i\frac{\partial^2 C_i}{\partial r^2}$	(33)
Dimensionless variables			
$X=\left(\frac{r}{r_0}\right)^3=\rho^3$	$X=\frac{x}{l}$	$X=\frac{x}{l}$	(34)
$Y_i=\left(\frac{r}{r_0}\right)\left(\frac{3av}{D_i}\right)\left(\frac{\Omega}{\nu}\right)^{1/2}y$	$Y_i=\left(\frac{2v_oh^2}{D_il}\right)\frac{y}{h}$	$Y_i=\left(\frac{2v_oR^2}{D_il}\right)\left(1-\frac{r}{R}\right)$	(35)
i = R or P	$\theta_R=\frac{C_R^o-C_R}{C_R^o}$	$\theta_P=K\frac{C_P}{C_R^o}$	(36)
Universal equation			
	$Y\frac{\partial\theta}{\partial X}=\frac{\partial^2\theta}{\partial Y^2}$		(37)
Boundary conditions			
	$\lim_{X\to\infty}\theta(X,Y)=0$		(38)
	$\lim_{Y\to\infty}\theta(X,Y)=0$		(39)

Table 3 Fundamental relationships for concentrations and fluxes at the surface in Laplace and real spaces

Laplace space	Real space
$\bar\theta(s,Y)=3^{2/3}\Gamma(2/3)\mathrm{Ai}(s^{1/3}Y)\bar\theta(s,0)$	$\theta(X,Y)=\frac{Y}{3^{2/3}\Gamma(1/3)}\int_0^X\theta(x,0)(X-x)^{-4/3}\exp\left\{-\frac{Y^3}{9(X-x)}\right\}dx$
$\bar\theta(s,Y)=-3^{1/3}\Gamma(1/3)\frac{\mathrm{Ai}(s^{1/3}Y)}{s^{1/3}}\left(\frac{\partial\bar\theta}{\partial Y}\right)_{Y=0}$	$\theta(X,Y)=\frac{-1}{3^{1/3}\Gamma(2/3)}\int_0^X\left(\frac{\partial\theta}{\partial Y}\right)_{Y=0}(X-x)^{-2/3}\exp\left\{-\frac{Y^3}{9(X-x)}\right\}dx$
$\bar\theta(s,0)=-3^{-1/3}\frac{\Gamma(1/3)}{\Gamma(2/3)}s^{-1/3}\left(\frac{\partial\bar\theta}{\partial Y}\right)_{Y=0}$	$\theta(X,0)=\frac{-1}{3^{1/3}\Gamma(2/3)}\int_0^X\left(\frac{\partial\theta}{\partial Y}\right)_{Y=0}(X-x)^{-2/3}dx$
$\left(\frac{\partial\bar\theta}{\partial Y}\right)_{Y=0}=-3^{1/3}\frac{\Gamma(2/3)}{\Gamma(1/3)}s^{1/3}\bar\theta(s,0)$	$\left(\frac{\partial\theta}{\partial Y}\right)_{Y=0}=-\frac{3^{1/3}}{\Gamma(1/3)}\int_0^X(X-x)^{-1/3}d\theta(x,0)$

Equation (46) can be rearranged as Equation (47)

$$\bar\theta(s,0)=-3^{-1/3}\frac{\Gamma(1/3)}{\Gamma(2/3)}\left(\frac{\partial\bar\theta(s,Y)}{\partial Y}\right)_{Y=0}s^{-1/3} \tag{47}$$

and then replaced into Equation (44) to yield Equation (48):

$$\bar\theta(s,Y)=-3^{1/3}\Gamma(1/3)\left(\frac{\partial\bar\theta(s,Y)}{\partial Y}\right)_{Y=0}s^{-1/3}\mathrm{Ai}(s^{1/3}Y) \tag{48}$$

Equations (44) and (46–48) and their corresponding inverse Laplace transforms (see Table 3),[28] provide useful relationships between fluxes and concentrations, which can be used to solve a variety of complex problems of electrochemical and spectroelectrochemical interest.

3 EXPERIMENTAL CONSIDERATIONS

Spectroelectrochemical experiments in the UV/VIS range may be broadly classified as transmission, external reflection, and internal reflection (see Figure 1). This brief section summarizes general aspects of these measurements with particular emphasis on factors that limit their capabilities in terms of spatial and temporal resolution, as well as detection sensitivity. More specialized issues are discussed in the applications section.

With the exception of the thin layer technique in the static mode, for which cells can be readily adapted to fit in the sample compartment of commercial spectrophotometers, implementation of all other UV/VIS spectroelectrochemical methodologies requires the custom assembly of optical components to bring radiation

into the cell and, after interaction with the electrolyte solution, redirect the beam to the detection stage.

The spatial resolution of profile images is governed primarily by diffraction limits imposed by the optics, which set bounds to how small a volume can be reliably imaged, about 2–4 μm for this type of application, although in some cases the size of individual pixels or diodes in arrays may become the limiting factor. Beam collimation, an important aspect of this type of experiment, can be most readily achieved with lasers. However, wavelength tunability may at times be required and, in the absence of expensive instrumentation such as tunable solid-state and dye lasers, conventional optical components can be utilized to modify beams from arc or filament lamps into the desired shape. Even when collimation is optimum, a beam traveling parallel and too close to the electrode surface will undergo diffraction. Although analysis of the pattern generated is closely related to the profile itself, it has been difficult to obtain reliable quantitative spatial information using this approach. Advantage has been taken, however, of the spatial specificity of the diffracted beam to probe a small volume of solution, within less than 10 μm from an electrode surface especially for transient type measurements.[29]

Technical advances in electronics have rendered electrochemistry (more precisely, the ability to establish a specific working electrode potential over a short period of time) as the factor that limits time-resolution, rather than data acquisition. Conventional potentiostats suffer from serious limitations, in that the powers required to achieve potential control are higher than the compliance limits of regular operational amplifiers. For example, if a 0.5 V step is applied to an electrode with a double layer capacity of 1 μF, over a period of 1 μs, the potentiostat must supply a current of 0.5 A. If, in addition, the resistance due to the electrolyte is of about 0.5 kΩ, the voltage required to drive the system would be 250 V, or a peak power of 125 W.

A means of overcoming some of these restrictions, implemented by Winograd,[30] involves the use of a capacitor to inject charge into the double layer at very fast rates, coupled to a potentiostat, to establish potential control at a later time. Further reductions in the time constant of the cell τ_{cell}, defined as the product of the cell resistance and the cell capacitance, can be achieved using microelectrodes. Their size leads to smaller currents and, thus, reduced power requirements. This strategy allows use of a two-electrode cell arrangement with a comparatively much larger counter electrode, so that its potential may not be greatly affected following the potential step. Using this approach, charge can be delivered to the working electrode over tens of nanoseconds. It is indeed serendipitous that UV/VIS spectroelectrochemical methods have played a key role in verifying that potentials can indeed be controlled over such short periods of time.[31]

Detection sensitivity is determined by the response of the photomultiplier, photodiode or charge-coupled device (CCD), including their noise characteristics. Most often, signal averaging or modulation schemes are required to extract signals for systems involving species displaying small molar absorptivities at low concentrations, or cells with thin pathlengths. Although double-beam instruments have been used for UV/VIS spectroelectrochemical studies,[12] it is usually not possible to employ a reference beam to normalize absorption from the cell. Under such conditions, special efforts must be made to stabilize the single beam by using, for instance, a split beam in conjunction with its own photomultiplier to control the voltage at the primary photomultiplier.

A powerful experimental approach to enhancing sensitivity via signal averaging involves the use of spectrophotometers capable of acquiring spectra over a wide wavelength region in times of the order of milliseconds. Some of these instruments are based on mechanical motion, such as the vibrating grating/mirror developed by Harrick, the ingenious ScanDisk of OLIS's subtractive monochromator, or on the use of CCD such as that described by Park.[32,33] Wavelength modulation provides yet another viable means of extracting very weak spectral signals.[34] This technique relies on the application of a periodic perturbation to the wavelength by, for example, mechanically vibrating a grating or a slit.[35] Hence, using a conventional lock-in amplifier it becomes possible to measure a signal proportional to the derivative of the response, which can then be integrated to generate, within an additive constant, a more conventional spectrum. Another type of signal averaging involves single wavelength measurements in which the same experiment is performed repetitively. Implementation of this methodology relies on returning the system to its original state following each acquisition over as short a time as possible. Such conditions are most readily achieved by probing volumes of solution very close to the electrode surface using, for example, internal reflection or diffraction, which can be ''filled'' and ''dumped'' by potential control.

Because of differences in the acquisition times and other considerations, it is difficult to assess absolute sensitivities of the various techniques developed to date. Nevertheless, a very useful semiquantitative comparison of these methods has been provided by McCreery et al.[29]

4 APPLICATIONS

This section presents experimental and theoretical aspects of the techniques that constitute the foundations of

modern UV/VIS solution-phase spectroelectrochemistry. The intent has not been to review exhaustively the literature in this area, but rather to provide illustrative examples that may be used as a basis for the design and execution of experiments and interpretation of data.

The material selected has been organized in increasing order of complexity, which coincides, by and large, with their chronological development. In this spirit, the first subsection introduces the now classical thin-layer cell technique in the static mode, a method that provides primarily thermodynamic and kinetic information of strictly solution-phase reactions, followed by dynamic measurements in quiescent media involving optically transparent electrodes. Attention is then focused on the monitoring of solution composition changes in the immediate vicinity of the electrode by internal reflection, and of the entire diffusion layer by the more powerful and versatile specular reflection mode. This latter approach enhances the detection sensitivity by increasing the effective optical path allowing, in addition, use of a much wider variety of electrode materials. Illustrations of reflection techniques are provided for a number of electrode geometries, including microelectrodes, under stagnant and forced convection conditions. Described in the subsequent section are imaging techniques to obtain detailed dynamic information regarding evolution of concentration profiles using mostly multidiode array detectors. Owing to its own mathematical nuances, steady and transient UV/VIS spectroelectrochemical methods under forced convection were grouped in a separate subsection. Finally, the last subsection addresses methods that rely on potential modulation for the study of kinetics and mass transport phenomena associated with electron-transfer reactions.

Because of their unique spectroscopic and electrochemical properties, a few systems have been used as models for the development of most of these techniques. The most useful absorption features and the corresponding diffusion coefficients of some of the most representative species involved are compiled in Table 4.

Table 4 Molar absorptivities at the specified wavelengths and diffusion coefficients for selected chromophores commonly used in spectroelectrochemical experiments

Chromophore	$\varepsilon(M^{-1}\,cm^{-1}) \times 10^{-3}(\lambda(nm))$	$D(cm^2\,s^{-1}) \times 10^6$	Ref.
CPZ cation	12 (520)	4.1[a]	36
Orthodianisidine cation	22.7 (515)	4.4[a]	37
Ferricyanide	1.0 (425)	6.32	38
MB	56 (633)	4.1	39
TAA cation[b]	11 ± 0.1(633)	1.25	40
Methyl promazine	5.46 (600)	2.83	41
Methyl viologen cation	9.7 (632.8)	8.6[a]	42

[a] Values correspond to *D* of the parent neutral species.
[b] TAA, tris(4-methoxyphenyl)amine; CPZ, chlorpromazine; MB, methylene blue.

4.1 Transmission Thin-layer Spectroelectrochemistry

The simplest type of UV/VIS spectroelectrochemical experiment, both from experimental and theoretical viewpoints, involves measurements of fully homogeneous solutions that contain a single species that absorbs light in this spectral region, either a redox-active reactant or a product generated directly or indirectly by an electrochemical process. Such conditions can be achieved by using the electrode as a source or sink of electrons to convert reactants into products to the desired extent, rendering a solution for which the concentration of all species is uniform within the thin layer cell and in equilibrium with the applied potential. On this basis, the absorbance may be interpreted in the same fashion as in conventional nonelectrochemical spectroscopic measurements.

Owing primarily to their ease of construction, high versatility and compatibility with commercial spectroscopic instrumentation, thin-layer cells have become a common tool for thermodynamic and also for certain kinetic studies of redox-active materials containing UV/VIS chromophores.[5,7] In particular, the lack of spatial dependence allows conventional kinetic analyses for determination of rate constants and reaction orders and elucidation of mechanisms.

The archetypal UV/VIS thin-layer spectroelectrochemical cell is comprised of two flat plates made out of a material transparent in this wavelength range, such as fused silica or quartz, placed parallel to and at a small distance from each other (10–200 μm) forming a small volume cavity.[8] Either a thin metal, or a conductive oxide layer, such as In-doped Sn oxide, deposited on one of the plates, or a thin self-standing open metal grid interposed between the plates is used as the optically transparent working electrode. Structural integrity and isolation from the ambient atmosphere are achieved by means of adhesives, gaskets or mechanical compression. Depending on the design, the cell can be filled by suction or by capillary action through an opening at the bottom in contact with a bulk solution reservoir, where the reference and auxiliary electrodes are placed (Figure 3). Various types of optically transparent thin-layer electrode (OTTLE) cells have been designed to fit in the sample compartment of conventional spectrophotometers enabling transmission spectroelectrochemical data to be acquired normal to the electrode in a rather straightforward fashion.[7]

Vast improvements in the spectral sensitivity of thin layer cells can be obtained by placing the beam parallel as opposed to normal to the electrode surface, thereby

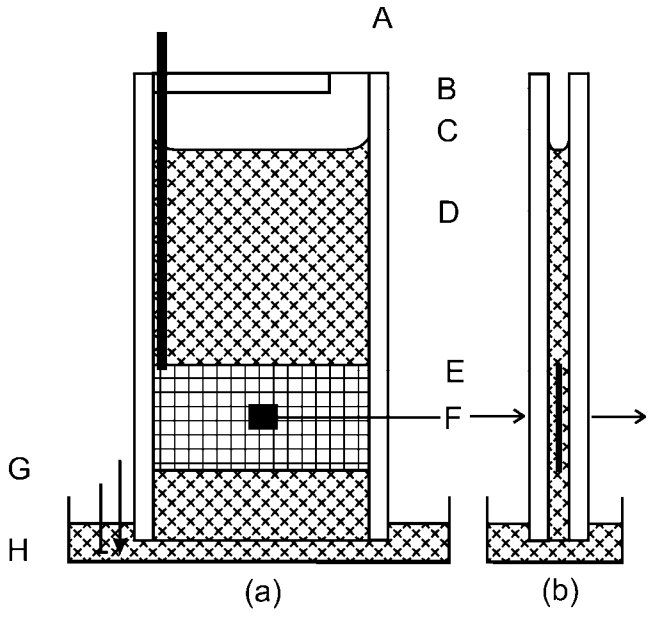

Figure 3 Optically transparent thin-layer electrochemical cell: (a) front view; (b) side view. A = point of solution suction; B = Teflon tape spacers; C = microscope slides; D = solution; E = gold minigrid electrode; F = optical path of spectrometer; G = reference and auxiliary electrodes; H = solution exit.[5]

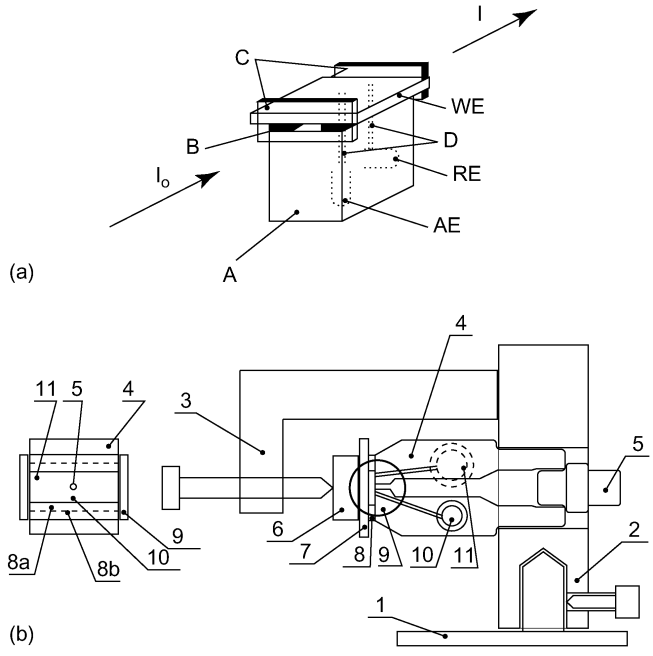

Figure 4 (a) View of the LOPTLC: A = cell body; B = thickness spacers; C = quartz windows; D = solution contact channels; WE = working electrode; RE = reference electrode; AE = auxiliary electrode. (b) Diagrams of the LOPTLC and the interior of the cell cavity: 1 = base plate; 2 = cell mount; 3 = thumbscrew and mount; 4 = cell; 5 = auxiliary electrode chamber; 6 = pressure plate; 7 = working electrode; 8a = Teflon thickness spacer; 8b = silicone rubber seal; 9 = quartz window; 10 = reference electrode chamber; 11 = solution inlet chamber.

increasing the optical pathlength, a configuration that also allows the use of nontransparent electrodes. At least two versatile long optical pathlength thin-layer cells (LOPTLCs) have been described in the literature using either conventional windows[43] (Figure 4) or arrays of optical fibers[44] to bring the beam in and out of the cell.

Among its many advantages, transmission thin layer spectroelectrochemistry enables expedient and quantitative conversion of reactant into products, which in many cases may be difficult to synthesize, handle or stabilize over long periods of time. In addition, the small area of contact between the solution in the thin layer and that in the bulk reservoir increases the time required to achieve chemical equilibration. In other words, the composition of the solution within the layer following polarization remains virtually unaltered after the electrode is open circuited for periods of time on the order of minutes. As the electrolyte solutions are at all times homogeneous, the absorbance in Equation (5) reduces to the most familiar form of Beer's law, allowing the spectral properties of electrogenerated species to be determined.

However, the thin character of the cell brings about large ohmic resistances. Consequently, the time required for a uniform potential to be established within the cell can be relatively long, thereby restricting the range of accessible homogeneous rate constants that can be measured with this technique. In fact, about 30 s are often required for complete electrolysis of material confined within the OTTLE cell.

Despite these shortcomings, transmission UV/VIS spectroelectrochemistry has become a well-established tool in inorganic chemistry and biologically oriented areas for the characterization of redox-active species under equilibrium conditions, or for reactions that proceed via fast electrochemical steps followed by a relatively slower chemical step. Examples of some of these applications are given below.

4.1.1 Thermodynamics

4.1.1.1 Determination of the Formal Reduction Potential $E^{o\prime}$ *and the Number of Electrons Transferred* n *for Redox Reactions Involving Stable Solution-phase Species*

Consider a generalized reversible redox couple (Equation 49)

$$R + ne^- \rightleftharpoons P \tag{49}$$

where both R and P represent stable solution phase species, for which the kinetics of electron transfer for

both the forward and reverse process are relatively fast. Under these conditions, the ratio of the concentrations of R and P at the electrode–solution interface will be determined by the applied potential E_{appl}, via the Nernst equation. If the electrode is polarized at a fixed value of E_{appl} for a sufficiently long time, so that C_R and C_P become uniform throughout the region probed by the beam, an explicit relationship can be obtained between E_{appl} and the relative concentrations of R and P (Equation 50):

$$E_{appl} = E^{o\prime} + \left(\frac{RT}{nF}\right) \ln\left(\frac{C_R}{C_P}\right) \qquad (50)$$

where $E^{o\prime}$ is the formal reduction potential of the redox couple.

If it is further assumed that at least one of these species displays one or more clearly defined features in the UV/VIS region, it becomes possible to determine C_R/C_P as a function of E_{appl} directly from the spectroscopic data, i.e.

$$\frac{C_R}{C_P} = \frac{A_i - A_P}{A_R - A_i} \qquad (51)$$

In Equation (51) A_P and A_R are the absorbances due to P and R, respectively, at judiciously selected wavelengths, measured at potentials sufficiently positive or negative so that only P or R are present in the solution, and A_i is the corresponding absorbance at some intermediate E_{appl} value. Combination of Equations (50) and (51) yields Equation (52)

$$E_{appl} = E^{o\prime} + \left(\frac{RT}{nF}\right) \ln\left(\frac{A_i - A_P}{A_R - A_i}\right) \qquad (52)$$

from which $E^{o\prime}$ and n can be determined, respectively, from the intercept and slope of E_{appl} versus $(A_i - A_P)/(A_R - A_i)$, or Nernstian plots. The subsections to follow will show examples of application of these principles.

Reversible Systems. Figure 5 shows in situ UV/VIS spectra of a 2.0 mM $K_3[Fe(CN)_6]$ aqueous solution in 1 M KCl at different potentials recorded in a thin layer cell.[45] As indicated, $[Fe(CN)_6]^{3-}$ displays two rather prominent bands centered at 213 and 420 nm, which can be used to construct Nernstian plots, such as those shown in the insert to Figure 5. Analysis of these data yielded values of $E^{o\prime} = 0.256 \pm 0.001$ V versus Ag/AgCl and $n = 0.965 \pm 0.0015$, in fairly good agreement with results obtained by other means.

Quasireversible Systems. The virtual diffusional decoupling between the solution contained within the thin-layer cell and that in the bulk reservoir allows spectroelectrochemical experiments to be conducted over periods of time long enough to investigate slow

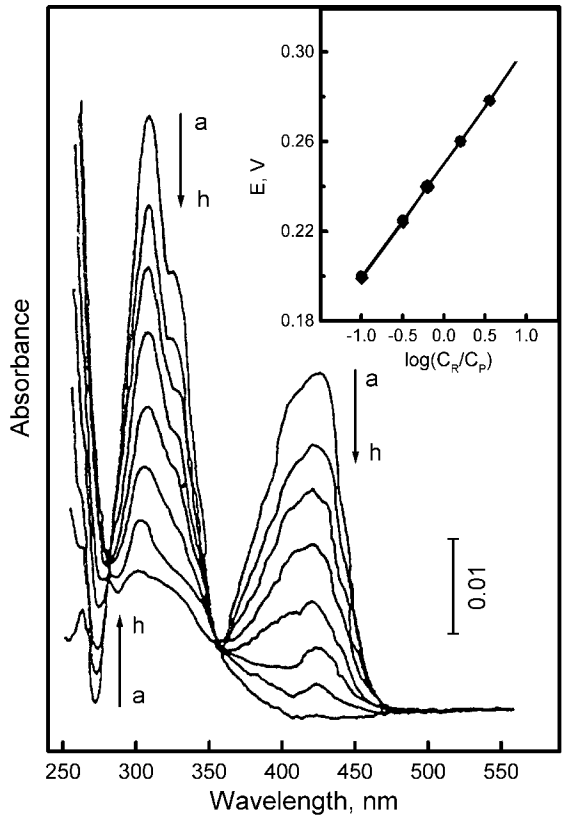

Figure 5 UV/VIS absorption spectra of 2.0 mM $K_3Fe(CN)_6$ in aqueous 1 M KCl at different potentials in the range +0.5 (a) and 0.0 (h). The inset shows a plot of E_{appl} versus $\log(C_R/C_P)$ at (•) $\lambda = 312$ nm and ( ) $\lambda = 420$ nm obtained in a thin-layer cell.[45]

redox reactions. For example, vitamin B_{12}, cyano-[Co^{III}]cobalamin, exhibits no cyclic voltammetric reduction features at potentials positive to the observed two-electron reduction wave, (Figure 6), regardless of scan rate.[46] Clear evidence that this reaction proceeds via two consecutive one-electron transfer steps was obtained by Mark et al.[46,47] in buffered solutions (pH = 6.86), using an Hg–Au minigrid working electrode in an OTTLE cell. For these measurements, the potential was stepped to different values in the range $-1.15 < E < -0.1$ V versus SCE (saturated calomel electrode) and the UV/VIS spectrum recorded after equilibrium had been established. Two distinct absorption peaks were observed for potentials in the range $-0.58 < E < -0.75$ V at 311 and 361 nm (Figure 7a), and $-0.77 < E < -0.95$ V versus SCE (Figure 7b) at 368 nm. Nernstian plots of the absorbance measured at these wavelengths yielded for the two one-electron transfer processes values of $E_{1/2} = -0.655$ V and $E_{1/2} = -0.88$ V versus SCE.

Mediated Electron Transfer. The rates of heterogeneous electron transfer of large redox-active molecules,

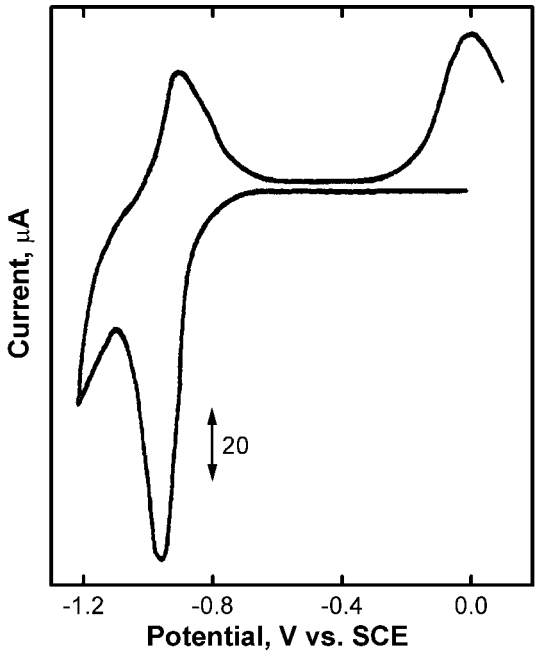

Figure 6 Cyclic voltammogram of 1 mM vitamin B_{12}, Britton Robinson buffer pH = 6.86, 0.5 M Na_2SO_4 obtained with an Hg–Au minigrid in a thin-layer cell. Scan rate 2 mV s^{-1}.[46]

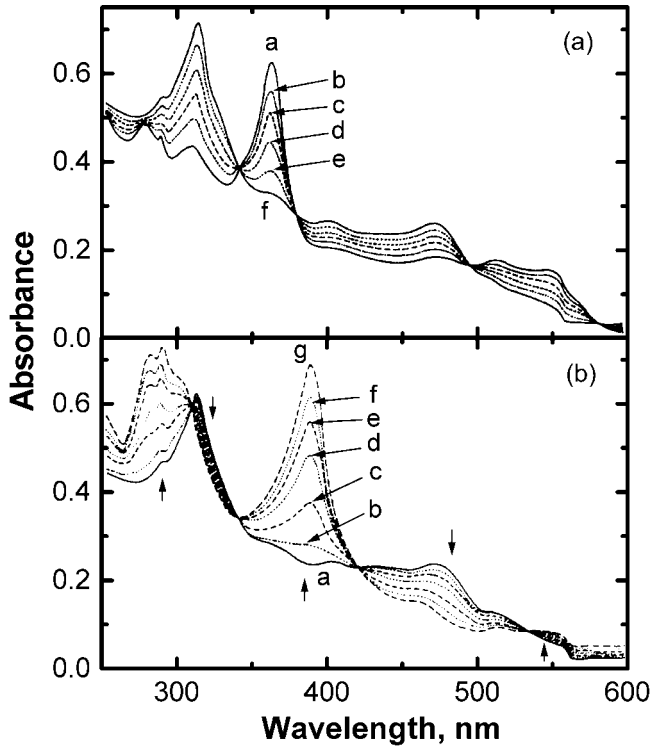

Figure 7 UV/VIS spectra obtained in a thin-layer cell for (a) the one-electrode reduction of vitamin B_{12} under the same conditions specified in the caption to Figure 6. Applied potentials: a = −0.550; b = −0.630; c = −0.660; d = −0.690; e = −0.720; f = −0.770 V versus SCE and (b) the second one-electron reduction. Applied potentials: a = −0.770; b = −0.820; c = −0.860; d = −0.880; e = −0.900; f = −0.920; g = −1.000 V versus SCE.[46]

such as proteins, are very small, due at least in part to the insulating character of the chemical environment surrounding the redox site. A common strategy to overcome this difficulty, shown schematically in Scheme 1, is to employ an electron-transfer mediator M, either in solution or bound to the electrode surface, capable of exchanging electrons with such larger molecules in solution at much higher rates.

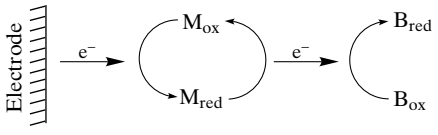

Scheme 1

This approach enables spectra of redox-active proteins or other large molecules to be recorded as a function of E_{appl}, allowing values of $E^{o\prime}$ and n to be determined by the same method described above. A classic example of this principle is provided by cytochrome c, a species that exhibits no clear voltammetric features on Au electrodes. However, large enhancements in the electron transfer rates can be obtained by using 2,6-dichlorophenolindophenol (DCIP) as a mediator, as shown in Scheme 2.

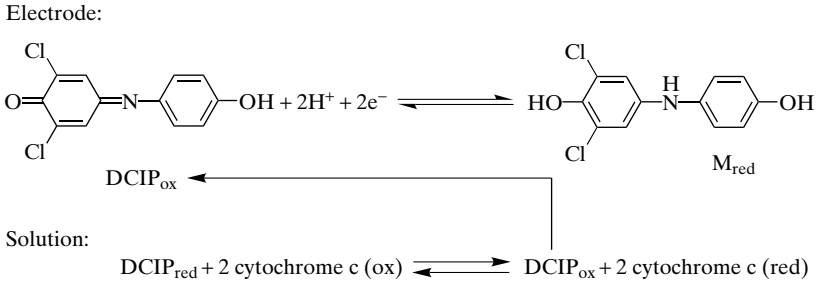

Scheme 2

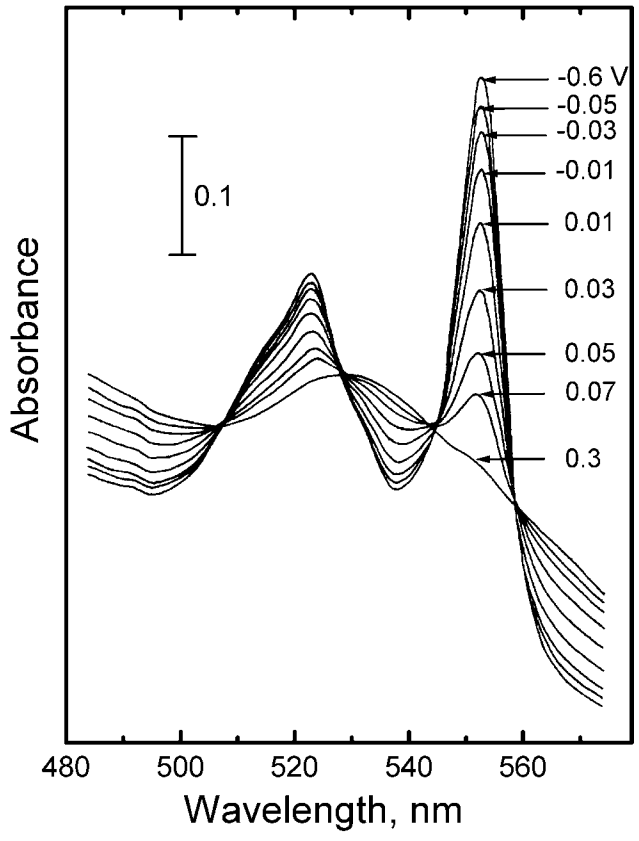

Figure 8 UV/VIS spectra of cytochrome c in 0.5 phosphate buffer (pH = 7.0), obtained in a thin-layer cell in the presence of 2,6-dichlorophenol–indophenol mediator–titrant, for a series of applied potentials versus SCE.[48]

This mediated process was monitored spectroelectrochemically in phosphate buffer (pH = 7), using a Au minigrid working electrode in an OTTLE cell, at potentials in the range −0.600 (fully reduced cytochrome c) to 0.300 V (fully oxidized cytochrome c) (Figure 8).[48] Nernstian plots for absorbance measured at 520 and 550 nm, the characteristic absorption maxima for cytochrome c(oxidized) and cytochrome c(reduced), yielded values of $E^{o\prime} = +0.264$ V versus SHE and $n = 1$.

4.1.2 Coupled Chemical Reactions

Electrochemistry provides an expedient means of preparing solutions of a given composition by controlling the charge that flows through the interface over a certain period of time. It thus becomes possible to use this method to monitor rates of some chemical reactions involving electrogenerated species under various initial conditions as a function of time. Application of this simple methodology assumes that the total time of charge injection is long enough for the concentration of all species to be uniform throughout the thin-layer cell before any significant bulk reaction occurs.

This approach has been exploited for studies of certain inorganic and biological processes that proceed via complex mechanisms, including chemical reactions that precede or follow an electron-transfer step (Scheme 3), for which one of the species involved absorbs light in the UV/VIS range.

$$\mathrm{R} + n\mathrm{e}^- \rightleftharpoons \mathrm{P}$$
$$\mathrm{P} \xrightarrow{k} \mathrm{Q}$$

Scheme 3

Three different techniques will be featured in this section to illustrate the principles involved.

4.1.2.1 Single Potential Step Perhaps the simplest transient spectroelectrochemical experiment of the type discussed above involves monitoring the absorbance of one or both of the redox species as a function of time following a potential step. An excellent illustration of this technique is provided by the work of Owens and Dryhurst,[49] who studied the electrochemical oxidation of 5,6-diaminouracil in aqueous buffered solutions pH = 4–6, for which the product, the corresponding diimine, undergoes hydrolysis, as depicted in Scheme 4. This system may be regarded as a model for a variety of biologically fundamental redox processes, including the enzymatic oxidation of purines.

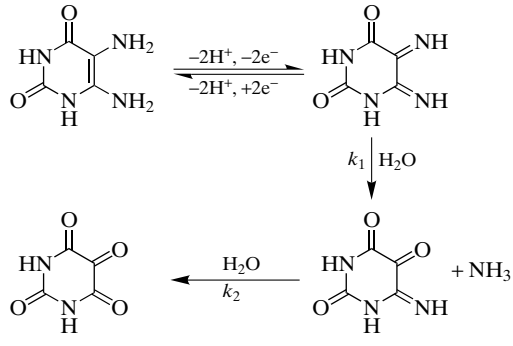

Scheme 4

UV/VIS spectra obtained using an OTTLE cell incorporating a Au minigrid electrode, recorded at different time intervals after the potential step was applied, revealed a broad band centered at 320 nm (not shown here) attributed to the diimine intermediate. If it is assumed that the decomposition of diimine follows first-order kinetics, the time dependence of its decay can be related to the change in absorbance via Equation (53):

$$\ln(A_t - A_\infty) = -kt + \ln(A_0 - A_\infty) \qquad (53)$$

where A_0, A_t and A_∞ are the absorbances at times 0, t and at the end of the experiment. A linear relationship

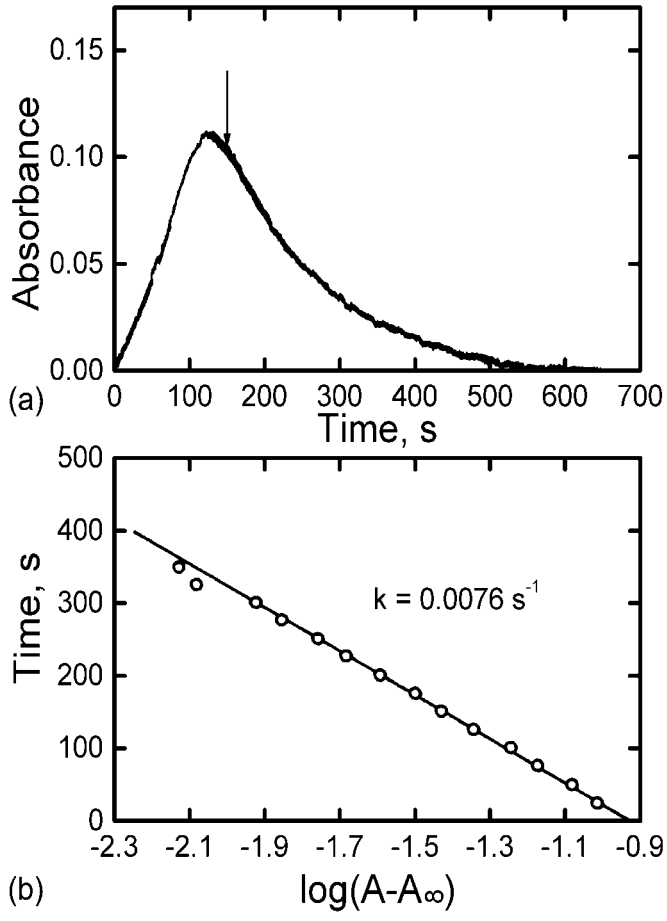

Figure 9 (a) Absorbance versus time recorded at $\lambda = 320\,\text{nm}$ during the electrooxidation of 10 mM 5,6-diaminouracil in McIlvaine buffer (pH = 5) at 0.35 V in a thin-layer cell with an Au minigrid electrode. The arrow indicates the time at which 5,6-diaminouracil was fully electrolyzed. (b) Plot of t versus $\log(A - A_\infty)$ obtained from the data in (a) by setting $t = 0$, at the point specified by the arrow.[49]

was observed between t and $\ln(A_t - A_\infty)$ recorded at $\lambda = 320\,\text{nm}$ (Figure 9), from which the rate constant k_1 was calculated, yielding a value of $0.04\,\text{s}^{-1}$ at pH = 4.

4.1.2.2 Double Potential Step A slight variation of the method just described involves application of a second potential step, after allowing a period of time t_R to elapse following the initial potential step. Assume, for example, that a certain absorbing species produced during the initial step, is then partially consumed via a first-order chemical reaction as specified in Scheme 3 above. If a potential step is applied in precisely the opposite direction at t_R, so as to convert all remaining P back to R, the difference in the change in absorbance between the first, ΔA_f (proportional to the amount of P generated), and the second step, ΔA_b (proportional to the amount of P remaining in solution), provides a measure of the amount of product decomposed during t_R. Based on information collected for various values of t_R, the rate of chemical decomposition of the product can be determined from Equation (54):

$$\ln \frac{\Delta A_f}{\Delta A_b} = k t_R \tag{54}$$

This method was employed to determine k_f for benzidine rearrangement[50] in an aqueous acidic electrolyte, a process that proceeds via the electrochemical–chemical (EC) mechanism shown in Scheme 5, using an Au minigrid electrode in an OTTLE cell.

For these experiments, azobenzene was first completely reduced to hydrazobenzene during the first potential step from 0.0 to −0.6 V versus SCE, the potential was then maintained at −0.6 for a reaction time t_R, and the remaining hydrazobenzene reoxidized to azobenzene during the second potential step to 0.3 V. All spectroscopic data were acquired at $\lambda = 325\,\text{nm}$, a value at which only azobenzene displays significant absorption. Typical optical response curves are shown in Figure 10, and the plot based on Equation (54) is given in Figure 11.

The advantage of the double-potential step versus single-potential step method is that either P or R can be monitored spectroscopically. However, the results obtained with this approach are less precise than for

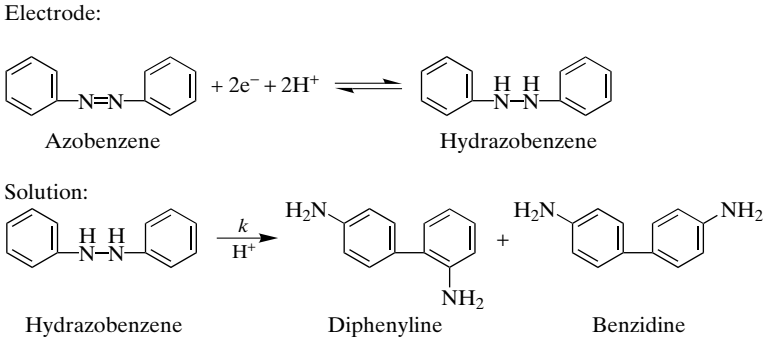

Scheme 5

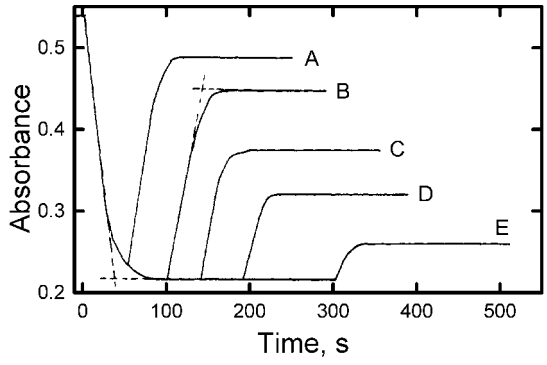

Figure 10 Absorbance versus time curves obtained at $\lambda = 325$ nm for a double-potential-step experiment performed in a thin-layer cell in 1 mM azobenzene, 0.100 M HCl, 0.150 M KCl, 44% ethanol aqueous solution. First step from +0.0 to −0.6 V and second step to +0.3 V versus SCE; $t_R = 50$ (A), 100 (B), 150 (C), 200 (D), and 300 s (E).[50]

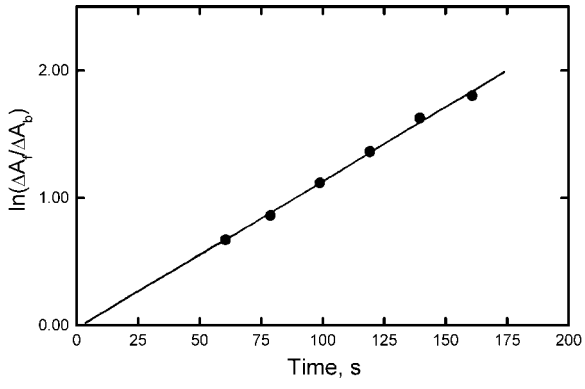

Figure 11 Kinetic plot, $\ln(\Delta A_f/\Delta A_b)$ versus time based on the absorbance versus time data for the double-potential-step experiment described in the caption to Figure 10.[50]

single potential step, as t_R has to be calculated graphically (dotted lines in Figure 10).

4.1.2.3 Single Potential Step Open Circuit Relaxation In some cases, species P in Scheme 3 can yield upon subsequent chemical reaction an electroactive species Q, which might be the original electroactive reactant R itself, a factor that can complicate analysis of spectroscopic data of the type introduced in previous sections. If k in Scheme 3 is sufficiently small, it becomes possible to monitor the reaction progress by bringing the electrochemical reaction to completion and then opening the electrical circuit enabling P to undergo further chemical transformations. This technique, known as single potential step open circuit relaxation, has been used to elucidate the mechanism of oxidation of the psychoactive drug CPZ.[51] This material can be electrochemically oxidized in strongly acidic media via two sequential one-electron steps to produce first, a deep red cation radical chlorpromazine cation radical ($CPZ^{+\bullet}$) (see Table 4), and, at more negative potentials, a colorless sulfoxide-type derivative (CPZO). However, the cation radical undergoes disproportionation according to the reaction sequence shown in Scheme 6.

$$\text{Electrode: CPZ} \longrightarrow \text{CPZ}^{+\bullet} + e^-$$

$$\text{Solution: } 2\,\text{CPZ}^{+\bullet} \xrightarrow[k]{H_2O} \text{CPZ} + \text{CPZO}$$

Scheme 6

For this process, Q in Scheme 3, happens to be identical to R. Evidence in support of this mechanism was obtained by spectroelectrochemical experiments using an Au minigrid working electrode in an OTTLE cell in a strongly acidic solution. Figure 12 shows a series of spectra recorded for a potential step from 0.250 V to 0.620 mV versus SCE. Curves a and b in this figure represent spectra collected before (CPZ) and after the step allowing 2.8 min for equilibrium to be reached, whereas all other spectra were recorded after the circuit was opened. The peaks at 262 and 525 nm, characteristic of $CPZ^{+\bullet}$, decreased during open circuit, and new features appeared at 292 and 337 nm, attributed to the final reaction product CPZO.

Plots of $1/C$ versus t for absorbance measured at 525 and 337 nm were linear, and therefore consistent with a

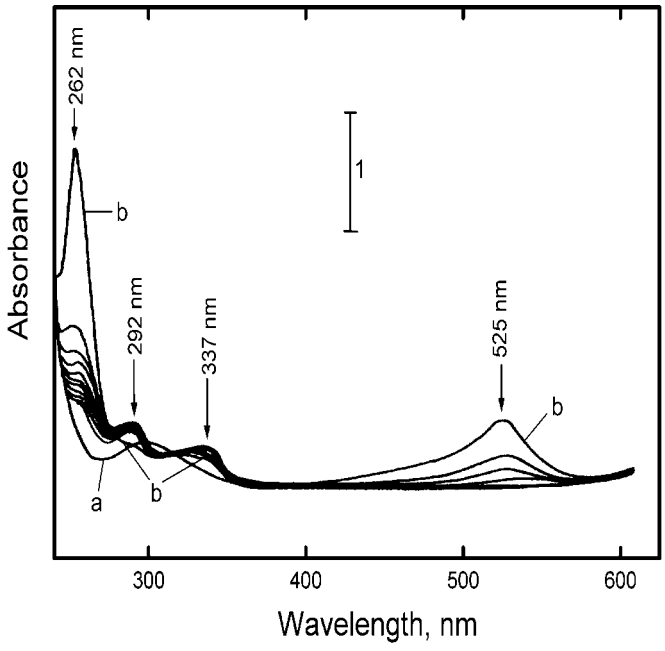

Figure 12 UV/VIS spectra recorded during a thin-layer chronoabsorptometry experiment with open-circuit relaxation in a 4.8 mM CPZ in 3 M H_2SO_4: (a) spectrum of CPZ at 250 mV; (b) spectrum of CPZ at open circuit. Spectra were acquired every 2.8 min.[51]

second order mechanism for $CPZ^{+\bullet}$ decomposition, with slopes of $8.6\,M^{-1}s^{-1}$ and $4.4\,M^{-1}s^{-1}$, i.e. a ratio of 2 : 1, in accordance with the reaction stoichiometry.

4.1.3 Slow Scan Rate Cyclic Voltammetry

The small amount of electroactive material contained within a thin-layer cell allows changes in composition induced by the applied potential to occur within rather short times, of the order of tens of seconds. Under these conditions it becomes possible to follow optically electrochemical processes involving absorbing species, by scanning the electrode potential linearly between two prescribed limits, while monitoring the current, a technique known as cyclic voltammetry. Analysis of these measurements can be greatly simplified by scanning the potential at slow enough rates to achieve at every instant quasi-equilibrium conditions throughout the cell.

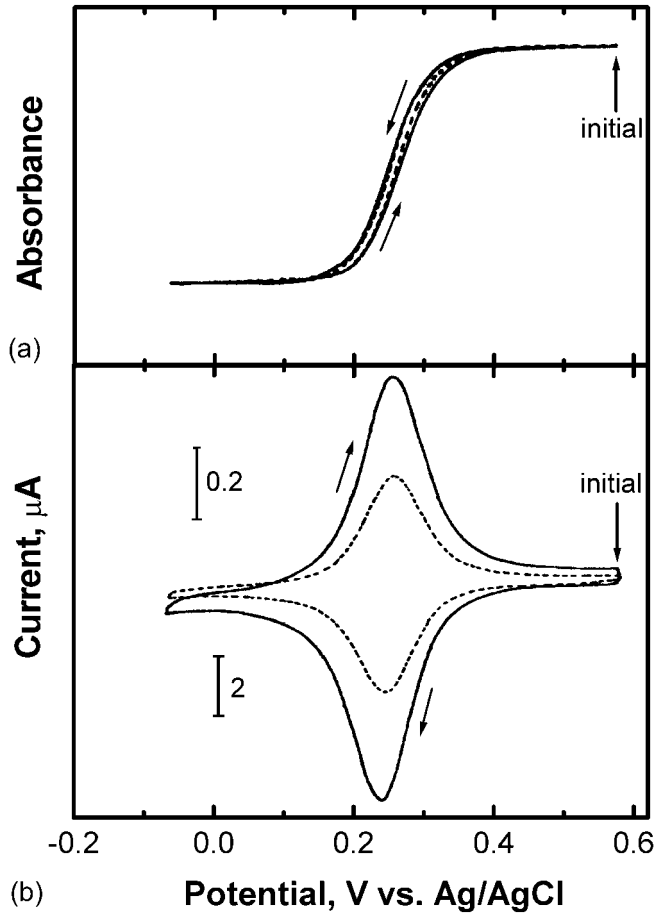

Figure 13 Curve (a) is the absorption versus time plot recorded during the cyclic voltammogram shown by curve (b) at $\lambda = 420\,nm$ using a glassy carbon electrode in a 0.4 mM ferricyanide in 0.5 M KCl solution at linear scan rates of (——) $2\,mV\,s^{-1}$ and (- - -) $1\,mV\,s^{-1}$.[43]

Figure 13 shows simultaneous absorbance and current data obtained on a glassy carbon electrode in the long pathlength thin layer cell (LPTLC) shown in Figure 4 in a solution of 0.4 mM ferricyanide in 0.5 M KCl at $\lambda = 420\,nm$.[43] For this LPTLC, the absorbance change, following full reduction or subsequent oxidation, is about two orders of magnitude larger than those observed for OTTLE cells described earlier.

4.1.4 Molecular Adsorption on Solid Electrodes

Advantage can be taken of the much enhanced sensitivity of LPTLC spectroelectrochemical cells and the possibility of using flat opaque electrodes to measure minute changes in concentration ΔC, such as those derived from adsorption of UV/VIS chromophores on electrode surfaces, by monitoring the corresponding decrease in the absorbance ΔA. As may be inferred from simple arguments, these two quantities are related to the width of the optical cell w and $\varepsilon(\lambda)$ (Equation 55):

$$\Delta C = \frac{\Delta A}{w\varepsilon(\lambda)} \tag{55}$$

On this basis, material lost from the solution phase, as measured by spectroscopic means, may be assumed to be adsorbed on the electrode surface, and the coverage Γ may be calculated from Equation (56),

$$\Gamma = \frac{V\Delta C}{a} \tag{56}$$

where V is the volume of the cell, and a is the area of the electrode. The ability to measure Γ accurately is thus determined by how small an absorbance change the instrument is capable of detecting. Saturation coverages of small size organic molecules on common metals are about $10^{-10}\,mol\,cm^{-2}$; hence, assuming a cell volume of $10\,\mu L$, the change in concentration associated with formation of a full monolayer would be $10^{-5}\,M$. Minimum detectable values of ΔA are of the order of 0.002 at a signal-to-noise ratio of 2; therefore, for $\varepsilon(\lambda)$ of $1000\,M^{-1}cm^{-1}$, $w = 1$ and $a = 1\,cm^2$, ΔC as small as $2 \times 10^{-6}\,M$, and, thus, smaller that those corresponding to $\Gamma = 1$ could be observed.

Measurements involving a series of heterocyclic aromatics were performed by Kuwana et al. with the same LPTLC spectroelectrochemical cell shown in Figure 4 incorporating a Pt electrode.[52] Figure 14(a) shows absorbance versus time plots obtained after injecting a solution $32.9\,\mu M$ *trans*-1,2-bis-(4-pyridine)ethylene (Pyc=cPy) in buffered (pH = 7) 0.1 M $NaClO_4$ (first curve) monitored at $\lambda = 299\,nm$ with the Pt electrode polarized at −0.1 V versus Ag/AgCl. The two other curves were obtained after a second and third injection without removing Pyc=cPy adsorbed on the Pt surface during the

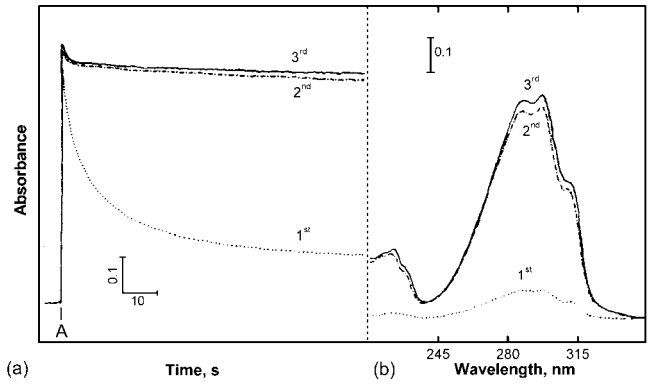

Figure 14 (a) Plots of absorbance versus time recorded at $\lambda = 299$ nm after the first, second, and third fillings of a thin-layer cell with a solution 32.9 μM PyC-CPy (*trans*-1,2-bis(4-pyridine)-ethylene) in 0.1 M $NaClO_4$ in 0.02 M phosphate buffer (pH = 7) using a Pt electrode polarized at −0.1 V versus Ag/AgCl immediately after each filling. Experiments began at point A. (b) Absorbance spectra acquired at $E = -0.1$ V after the corresponding absorbance versus time response has reached a steady state value for the 1st (· · ·), 2nd (-·-·-·-) and 3rd (——) fillings.[52]

first injection. Figure 14(b) shows the corresponding spectra of the solution phase recorded after the absorbance had reached steady state following each injection. The very small differences observed between the second and third injections indicate that the surface had attained saturation, i.e. $\Gamma = 1$.

These data may be used to determine molecular packing densities and from these to infer the mode of bonding of species to the surface as thoroughly discussed by Soriaga and Hubbard.[53] Proposed orientations of four of the species investigated by Kuwana et al. are shown in Scheme 7.[52]

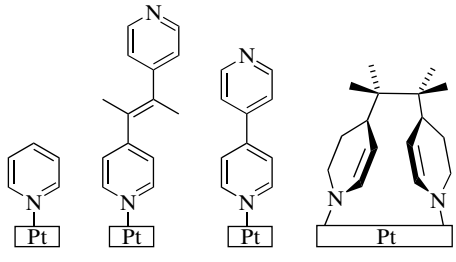

Scheme 7

4.2 Transmission Spectroelectrochemistry in Semi-infinite Media

Methods based on the dynamic monitoring of absorbance changes induced by electrochemical perturbations can provide kinetic information regarding rates of homogeneous electron transfer rates involving electrogenerated species, as well as mass transport phenomena, not accessible by the static techniques described in the previous section. For simplicity, consider an OTE immersed in a solution containing a single, nonabsorbing solution-phase species R capable of undergoing heterogeneous electron transfer to yield an absorbing solution phase product P (Equation 57):

$$R + ne^- \longrightarrow P \tag{57}$$

If C_P is independent of x and z, the instantaneous absorbance $A(t)$ in Equation (4) reduces to Equation (58),

$$A(t) = \varepsilon \int_0^d C_P(y, t)\, dy \tag{58}$$

where d is the length of the optical cell along the axis of propagation of the light. Provided the measurements are made over relatively shorts periods of time, so that the diffusion layer thickness is always smaller than d, the media may be regarded as semi-infinite and $A(t)$ may be rewritten as Equation (59),

$$A(t) = \varepsilon \int_0^\infty C_P(y, t)\, dy \tag{59}$$

or, in Laplace space (Equation 60), as

$$\overline{A}(s) = \varepsilon \int_0^\infty \overline{C}_P(y, s)\, dy \tag{60}$$

If it is further assumed the overall measurement time is short so that no complications arise from natural convection, an explicit form for $\overline{C}_P(y, s)$ for a potential step experiment of the type described in section 2.2.2 may be obtained from Equation (22):

$$\overline{C}_P(y, s) = -\frac{C_R^o}{s}\left(\frac{D_R}{D_P}\right)^{1/2} e^{-(s/D_P)^{1/2}y} \tag{61}$$

Equation (61) can be substituted into Equation (60) to yield Equation (62),

$$\overline{A}(s) = \frac{\varepsilon C_R^o D_R^{1/2}}{s^{3/2}} \tag{62}$$

and inverse Laplace transformed to afford an expression for the time dependence of the absorbance due to P for such diffusion-controlled process (Equation 63):

$$A(t) = \varepsilon C_R^o \left(\frac{4D_R t}{\pi}\right)^{1/2} \tag{63}$$

Based on the analysis of such chronoabsorptometric-type experiment, it becomes possible to determine D_R if ε is known or, conversely, ε if D_R is known.

A rather versatile cell design reported by Kuwana et al.[4] suitable for this type of measurements incorporates an OTE in a glass sandwich-type cell (Figure 15), in

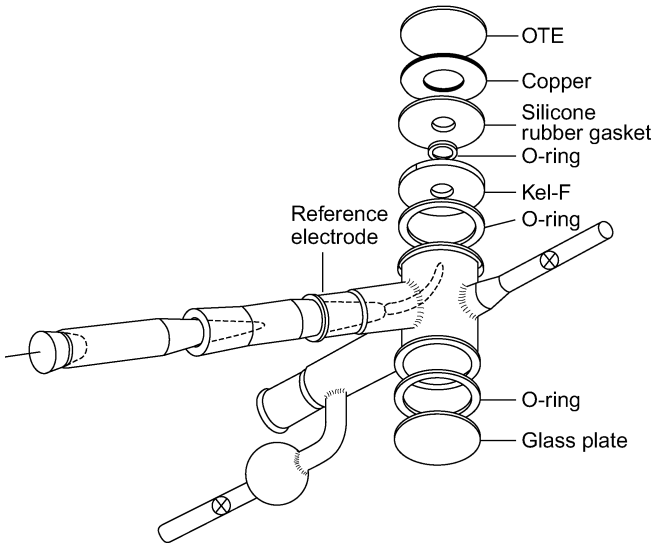

Figure 15 Sandwich cell for transmission UV/VIS spectroelectrochemical experiments.[4]

which the light travels along the axis of the cylinder. The small bent tube, known as a Luggin capillary, allows electrolytic contact between the main and reference electrode compartments.

Potential step chronoabsorptometry at an optically transparent tin oxide electrode has been used[54] to study the oxidation of $[Fe(CN)_6]^{4-}$ to $[Fe(CN)_6]^{3-}$ in aqueous electrolytes by monitoring $A(t)$ at $\lambda = 420\,\text{nm}$ (Table 4). In accordance with theoretical predictions, plots of $A(t)$ versus $t^{1/2}$ (Figure 16) were found to be linear, yielding values of D for $[Fe(CN)_6]^{4-}$ determined from the slope, of $6.5 \times 10^{-6}\,\text{cm}^2\,\text{s}^{-1}$, in reasonable agreement with those obtained by other methods.

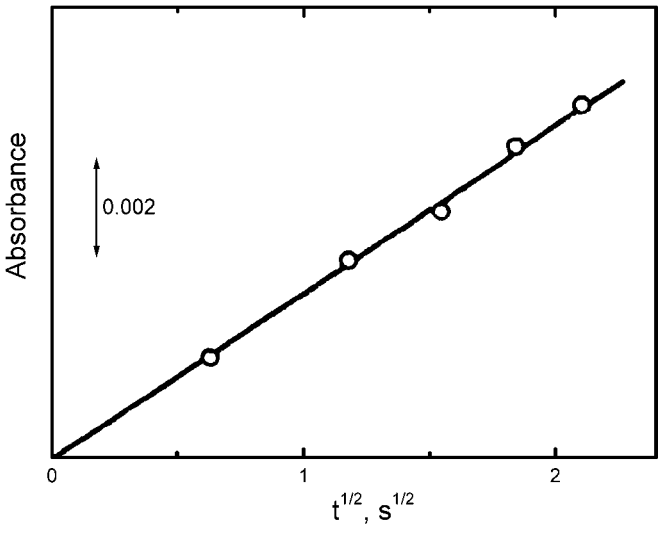

Figure 16 Plot of A versus $t^{1/2}$ for the chronoamperometric oxidation of $[Fe(CN)6]^{4-}$ at a tin oxide optically transparent electrode.[54]

One of the advantages of using large-volume cells, as opposed to thin-layer cells, is the decrease in τ_{cell}, which leads to improvements in transient response. However, the relatively low spectral sensitivity due to the very short optical paths associated with OTE, the restricted number of electrode materials with suitable electronic and optical characteristics, and the high resistance associated with thin films, severely limit the applicability of the technique described in this section. As explored in detail later below, some of these difficulties can be overcome by implementation of dynamic methods in both internal and external reflection modes.

4.3 Internal Reflection

As briefly discussed in section 2.1.3, light traveling through a media of high index of refraction, impinging on a planar interface with a media of smaller index of refraction at a sufficiently large angle, will penetrate slightly into the adjacent phase and interact with species present therein. For spectroelectrochemical applications, the surface of the internal reflection element (IRE) is coated with a thin film of an electronically conducting material, such as a metal or a highly doped semiconductor, which serves as the optically transparent working electrode. This technique was pioneered by Kuwana, Winograd et al. in the 1970s.[4]

The volume of solution probed by the evanescent wave is of the same order of magnitude as the wavelength of light, about 100–300 nm in the UV/VIS range, and therefore the size of the cell becomes immaterial. Under such conditions, the time response of the cell is limited not by the solution resistance, but by the thin character of the electrode, as was the case with the OTE-based techniques described in the previous sections.

Owing to the very short optical path and the relatively low concentrations of chromophores often used, the extent to which the exponentially decaying evanescent wave in the solution is attenuated is very small, and therefore its spatial dependence may be regarded as unperturbed by the presence of absorbing material therein. Under these conditions, the instantaneous absorbance is given by Equation (64),

$$A(t) = \varepsilon N_{\text{eff}} \int_0^\infty C_P(y, t) \exp\left(-\frac{y}{\delta}\right) dy \qquad (64)$$

where δ is the penetration depth (see Equation 10), and N_{eff} is a sensitivity factor that depends on the geometry and electrode material.

Insertion of Equation (22) into Equation (64) in Laplace space and subsequent inversion, yields Equation (65) for a chronoabsorptometric experiment:

$$A(t) = \varepsilon\delta N_{\text{eff}} C_{\text{R}}^{\text{o}} \left(\frac{D_{\text{R}}}{D_{\text{P}}}\right)^{1/2} 1 - \exp(\beta^2 t)\text{erfc}(\beta t^{1/2}) \quad (65)$$

where $\beta = D_{\text{R}}^{1/2}/\delta$.

For $\delta \sim 100\,\text{nm}$, $A(t)$ reaches steady state in about 1 ms, which may be regarded as the time needed to fill the cell. If the potential is then returned to its original value by applying a reverse step, the cell can be dumped over as short a time, making it possible to perform numerous *identical* experiments that can then be co-added and averaged to improve signal detection.

A multiple reflection internal reflection spectroscopy (IRS) cell designed by Kuwana and Winograd[55,56] is shown in Figure 17. In this optical arrangement two prisms are used to guide the beam into and out of the IRE, allowing five internal reflections to probe the solution phase of the interface, thereby increasing the effective pathlength.

Sensitivities better than one part in 10^5 with time resolutions in the microsecond range have been obtained for systems in which $N_{\text{eff}} \gg 1$, and long averaging times. This is illustrated in Figure 18 for MV^{++} reduction, for which the time resolution, using charge injection to achieve fast potential control, was of the order of 4 μs.[57] Advantage was taken of the extraordinary capabilities of this measurement scheme for kinetics studies involving fast processes,[58] such as the second-order catalytic mechanism (Scheme 8) for the reaction of cyanide with electrogenerated tri-*p*-anisylamine cation radical ($TAA^{+\bullet}$) in acetonitrile.

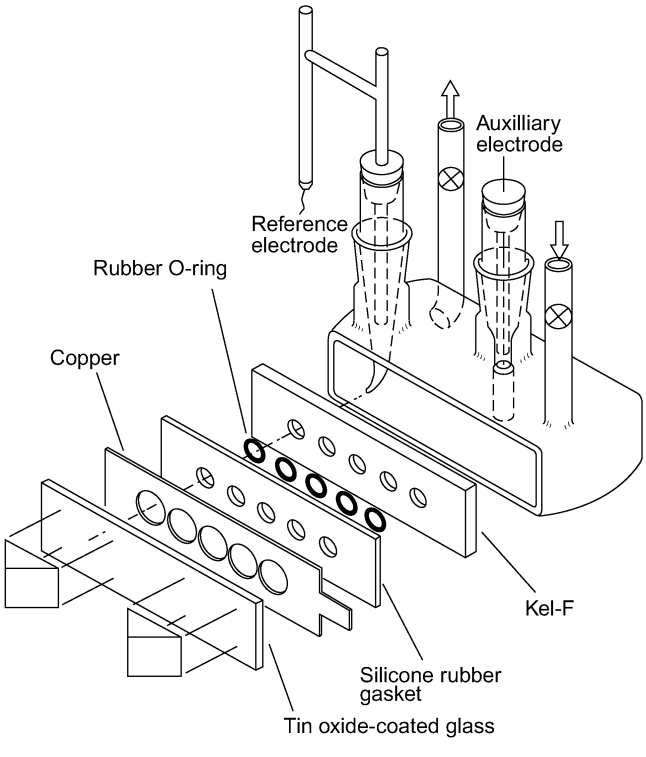

Figure 17 Cell for UV/VIS IRS measurements.[59]

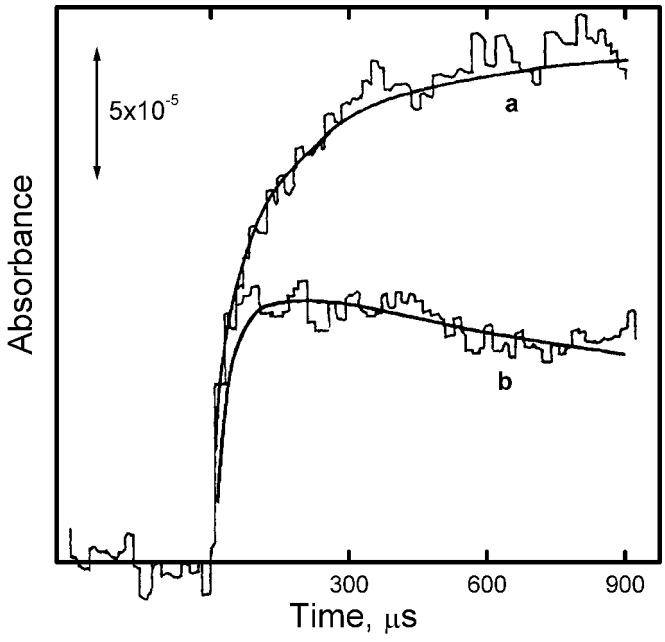

Figure 18 Absorbance versus time curves for the reduction of methyl viologen dication (MV^{++}) reduction, monitored at $\lambda = 605\,\text{nm}$ in the internal reflection mode. Curve (a) corresponds to the diffusion-limited reduction of MV^{2+} to its cation radical. Curve (b) is for the reduction to MV^0. Solid lines are theoretical transients; noisy curves are experimental.[57]

$$TAA \rightleftharpoons TAA^{+\bullet} + e^-$$
$$TAA^{+\bullet} + CN^- \xrightarrow{k_1} TAA + CN^{\bullet}$$
$$2\,CN^{\bullet} \xrightarrow{k_2} (CN)_2$$

Scheme 8

Plots of absorbance versus time following a potential step recorded at $\lambda = 715\,\text{nm}$, a wavelength at which $TAA^{+\bullet}$ exhibits a characteristic absorption band, were recorded both in the presence and in the absence of tetraethylammonium cyanide (TEACN; Figure 19). Analysis of these data yielded values for the k_1 in Scheme 8 of $2.7 \times 10^5\,M^{-1}\,s^{-1}$, in agreement with those obtained by other means.

IRS also lends itself to the study of homogeneous electron-exchange reactions involving electrogenerated species for values of k in Scheme 8 larger than $10^6\,M^{-1}\,s^{-1}$. Under such conditions, the intermediate is formed in close

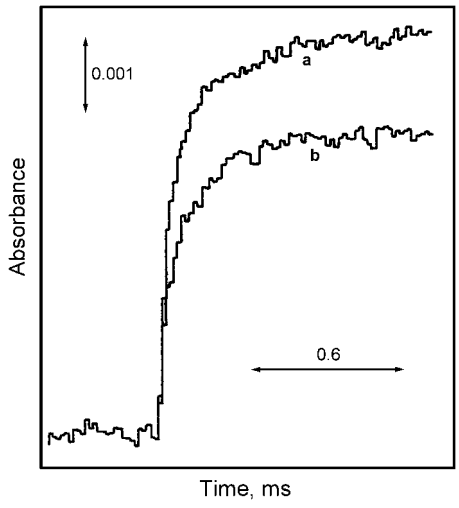

Figure 19 Absorbance versus time curves obtained at $\lambda = 715$ nm in the internal reflection mode following a potential step for (a) 0.5 mM TAA and (b) 0.5 mM TAA with 15 mM TEACN added, in 0.3 M tetraethylammonium perchlorate (TEAP) in acrylonitrile (AN) using an optically transparent Pt electrode for 800 repetitions. $N_{\text{eff}} = 15$, $\sqrt{D}/\delta = 115$.[58]

proximity to the electrode surface and can, therefore, be detected by the evanescent wave.

From a general perspective, quantitative analysis of experimental data can be pursued using digital simulation techniques to predict, for a proposed mechanism, time-dependent concentration profiles for all species involved,[59] as illustrated for an electrochemical–electrochemical–chemical (EEC)-type process in Figure 20. Numerical values for the various parameters involved can then be calculated based on best fits to the experimental curves. In fact, for the reduction of MV^{++} mentioned above, which has been found to follow the mechanism in Scheme 6, a fit to the data obtained by monitoring $A(t)$ at $\lambda = 605$ nm, associated with the radical cation intermediate $MV^{+\bullet}$, yielded values of $k_f = 3 \times 10^9\ M^{-1}\ s^{-1}$.[57]

More recently, Heineman et al.[60–63] have implemented the use of IRS to develop spectroelectrochemical sensors with high selectivity and sensitivity. For this application, the IRE is covered first with a thin conducting layer as before, and then with a film displaying molecular specificity or selective film, as shown in Figure 21. This device allows three independent sensing techniques to be employed simultaneously, i.e. electrochemistry, spectroscopy, and selective partitioning, thereby enhancing the degree of specificity for chemical detection. As a means of illustration, Figure 22 shows results obtained with simultaneous optical (panels a and b) and electrochemical techniques (panel c) for the detection of ferrocyanide partitioned into a poly(dimethyldiallyl ammonium chloride)-SiO_2 composite sol-gel film from an aqueous solutions at two different concentrations.

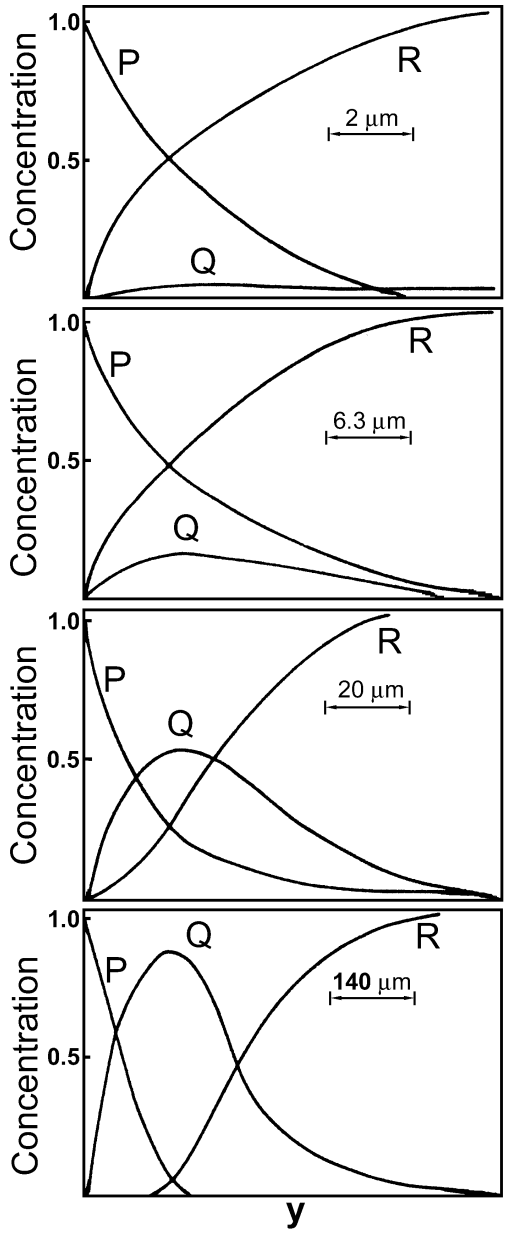

Figure 20 Concentration profiles obtained by digital simulation for an EEC mechanism with $k_f C = 10^2\ s^{-1}$. The different curves are drawn for $t = 1$, 10, 100, and 500 ms, from (a) to (d).[59]

4.4 External Reflection

The rather short effective pathlengths associated with techniques based on OTEs limits their applicability to systems involving strong chromophores, or long times for sufficient absorbing material to accumulate and be detected. Although gains in sensitivity can indeed be

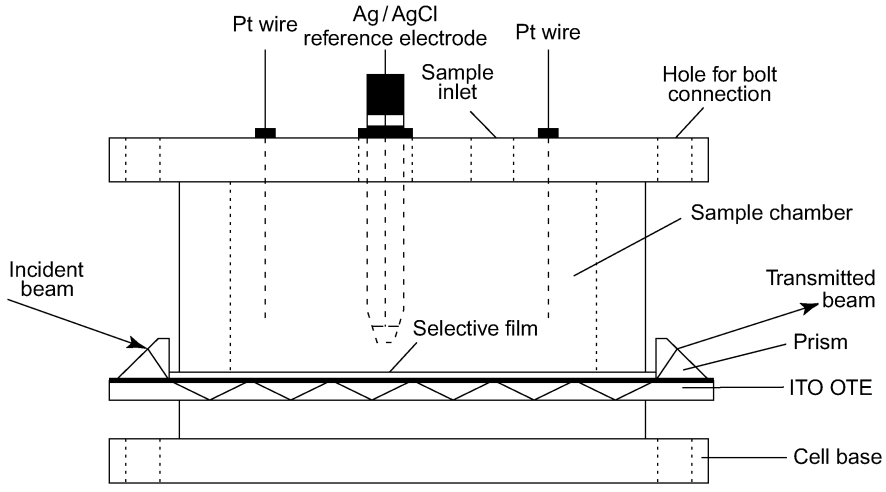

Figure 21 Simplified diagram of an internal reflection-based spectroelectrochemical sensor cell. ITO, indium-tin oxide.[60]

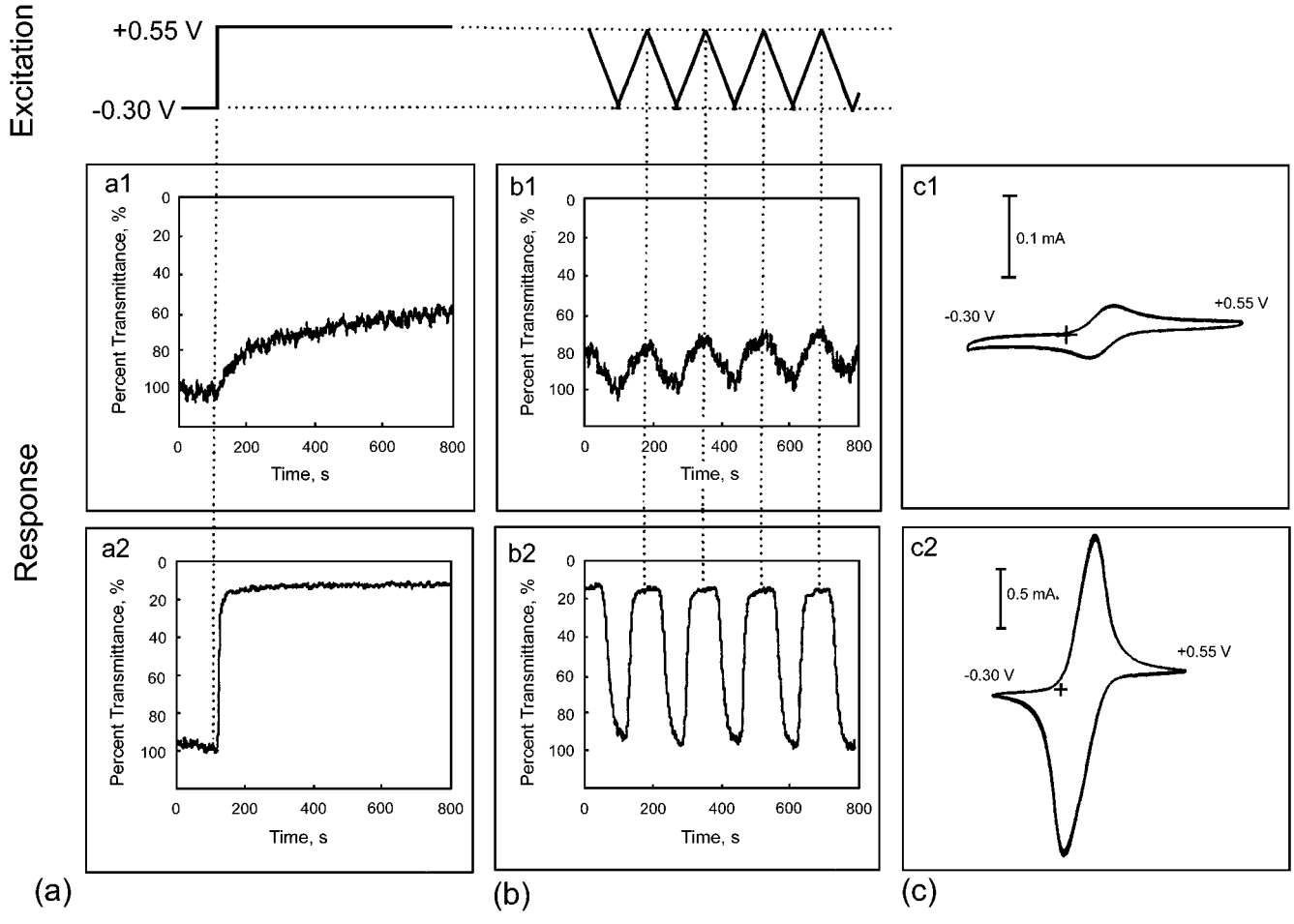

Figure 22 Optical response at $\lambda = 420\,\text{nm}$ of the sensor in Figure 20 to (a) potential step and (b) cyclic voltammetry for two concentrations of $Fe(CN)_6^{4-}$. Panel 1, 0.025 mM; panel 2, 2.5 mM. Panels (c) show cyclic voltammograms recorded in those solutions recorded at a scan rate of $10\,\text{mV}\,\text{s}^{-1}$. Potentials are referred to an Ag/AgCl reference electrode.[60]

achieved by IRS using extensive signal averaging, the region probed by the evanescent wave is very close to the electrode surface. Hence, it is not possible to monitor changes in composition at times *longer* than a few milliseconds. In addition, IRS is very sensitive to wavelength and refractive index effects on the pathlength, compromising the analysis of spectral data, particularly when N_{eff} is large.

Many of these problems can be minimized by experiments in which light propagating in the solution is reflected from, as opposed to transmitted through, the electrode. This approach allows transient studies to be performed under either quiescent or convective flow conditions, involving weakly absorbing and short-lived species. Such improvements are derived primarily, but not exclusively, from increases in the optical pathlength, achieved for example by employing grazing incidence or multiple reflection, better current distribution, and from the possibility of using microelectrodes.

Implementation of external reflection geometries often requires custom assembly of components, including a light source, either a laser or a conventional arc lamp, a monochromator, a detector, and other optical elements, as prescribed by the specific application. A few optical arrangements designed for this type of spectroelectrochemical experiments are depicted in Figure 23.[31] Considerable care has been taken to develop electrochemical cells displaying optimum optical and electrochemical response. For example, Figure 24[39,64] shows two cells reported in the literature for (a) external reflection at constant angle of incidence and (b) for multiple reflection experiments. One of the most interesting applications of external reflection, reported by McCreery et al.,[31]

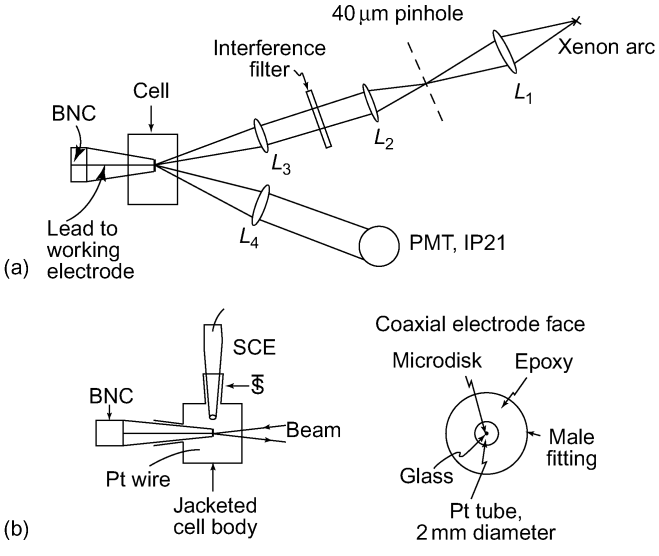

Figure 23 (a) Optical arrangement and (b) schematic diagram of a cell for spectroelectrochemistry at a microdisk electrode.[31]

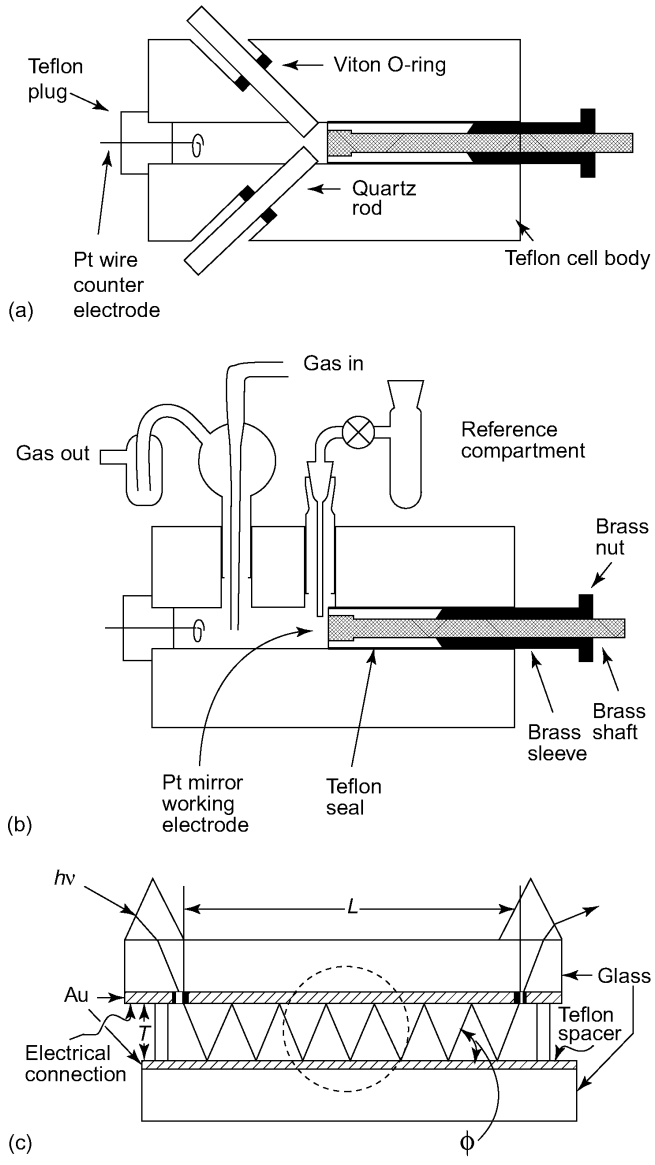

Figure 24 (a) Spectroelectrochemical cell for chronoabsorptometric studies: (a) top view, (b) side view,[64] (c) cross-section of a multiple specular reflection spectroelectrochemical cell.[39]

exploits the advantages of the small currents associated with micro-electrodes to eliminate the need for a potentiostat and use instead a capacitor-based injection circuit to charge the double layer, coupled to a pulse generator. As shown by these workers, and discussed later in this chapter, such a strategy enables a fixed electrode potential to be established following a step, within times shorter than 35 ns, making it possible to follow the time evolution of the absorbance from times as short as 0.15 μs.

Metals, even when optically polished are not perfect reflectors in the UV/VIS region; in addition, their reflectance is often a function of the state of charge and coverage of adsorbed species on the surface.[12] However,

it is possible to either minimize, or account for these effects, allowing quantitative information to be acquired regarding the identity and reactivity of electrogenerated solution phase species produced at conventional solid electrodes.

4.4.1 Planar Electrodes

The solution-phase absorbance measured with light reflected from a planar electrode at an angle of incidence ϕ (Figure 24a[64]) is $2/\cos\phi$ larger than that observed, under otherwise identical conditions, using a beam normal to an OTE in the standard transmission configuration. The factor of 2 accounts for the beam crossing the diffusion layer twice, whereas the divisor of $\cos\phi$ reflects the increase in pathlength for a single crossing of the layer. This enhancement factor, denoted as η, arises purely from geometric considerations, assuming the electrode surface to be perfectly reflecting, and that the incident and reflected beams traverse a diffusion layer uniform along an axis normal to the surface, i.e. possible edge effects are ignored. Based on these considerations, η will range from 2, for normal incidence, up to about 400 for ϕ slightly less than 89.7°. Comparisons between theoretical and experimental data, as well as quantitative calculations of edge effects, have been provided in the literature.[38,42,65]

The higher absorbances achieved at glancing incidence enable detection of short-lived electrogenerated chromophores with increased sensitivity, making it possible to monitor, for example, species with small ε, or strongly absorbing materials at lower concentrations. In addition, the higher electrical conductivity of massive, compared to thin film electrodes, allows measurements to be performed with better time resolution. Some of these aspects are discussed in more detail later in this section.

4.4.1.1 Identification of a Solution-phase Corrosion Product Polarization of metals at sufficiently high potentials may result in the formation of species derived from the electrode displaying finite solubility in the electrolyte solution. An illustrative example of the use of UV/VIS spectroscopy in the external reflectance mode for the identification of a metal corrosion product was reported by Kolb et al.[66] for the case of Ru in acid solutions. For these experiments, a Harrick rapid scan spectrophotomer was employed to collect spectra by reflecting light from an optically smooth Ru film electrode sputtered on glass as a function of the applied potential. Figure 25 shows the normalized reflectivity of Ru polarized at 1.17 V versus SCE in 0.5 M H_2SO_4 using the reflectance spectrum of pristine Ru at −0.1 V as a reference. Based on a comparison of the features observed with those of the genuine material (shown by the inset in the figure), the corrosion product could be identified as RuO_4.

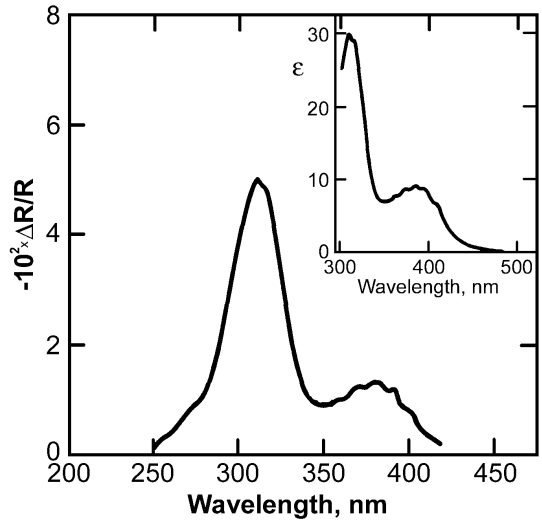

Figure 25 Background-corrected differential reflectance spectra for an Ru electrode in 0.5 M H_2SO_4. (· · ·) Result obtained following a potential step from −0.1 to +1.17 V versus SCE. The inset shows the molar extinction coefficient ε versus λ for RuO_4.[66]

4.4.1.2 Chronoabsorptometry

Glancing Incidence – Systems Involving a Single Absorbing Product. Some of the advantages associated with external reflection compared to transmission through an OTE may be illustrated using potential step chronoabsorptometry (see section 4.1) as an example. Assuming that the only optically absorbing species at the wavelength selected for the measurements (λ) is the product P of the electrochemical reaction, and that the heterogeneous electron transfer process is under strict diffusion control, $A(t)$ is given by Equation (63) corrected for the difference in optical path, i.e. $\eta = 2/\cos\phi$, namely

$$A(t) = \varepsilon_P C_R^o \frac{2}{\cos\phi}\left(\frac{4D_R t}{\pi}\right)^{1/2} \qquad (66)$$

Equation (66) has been verified in a number of laboratories for simple electrochemical processes. In particular, Figure 26(a)[42] shows absorbance versus time and Figure 26(b) the corresponding absorbance versus $t^{1/2}$ plots calculated from these data for a buffered solution (pH = 7) of 0.59 mM MV^{++}. The potential in this case was stepped from −0.3 to −0.8 V versus SCE using an Au electrode with an He−Ne laser beam ($\lambda = 632.8$ nm) incident on the sample at $\phi = 88.52°$. At this energy $MV^{+\bullet}$ displays significant absorption. The straight line in the figure was calculated based on values of ε_P and D_R reported in the literature (see Table 4). It is evident from these results that the agreement between theory and experiment is excellent for times longer than 10 ms. The discrepancies found at

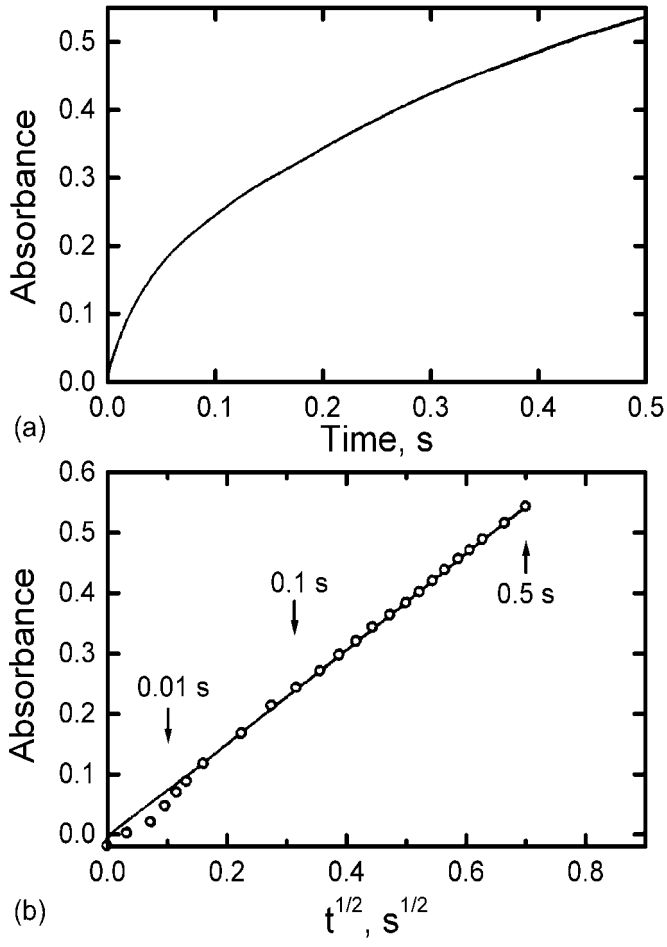

Figure 26 (a) Absorption versus time and (b) absorption versus $t^{1/2}$ plots obtained in a solution 0.59 mM methyl viologen in 0.3 M phosphate buffer (pH = 7.0) for a potential step from −0.3 to −0.8 V versus SCE. The data represent the average of three runs.[42]

shorter times were attributed to a variety of factors, a thorough discussion of which may be found in the original literature.[42] Good quantitative agreement (<2%) was also observed between the theoretical and experimental values of η for ϕ in the range 88–89°, which corresponds to η values of about 50–100, for electrolysis times of the order of 50 ms. However, deviations were detected at shorter times and smaller ϕ. This technique has been successfully extended to the entire UV/VIS range by using a continuum light source and a well-collimated beam, augmenting significantly its overall versatility.

Glancing Incidence – Systems Involving an Absorbing Reactant and an Absorbing Product. In the case of systems in which both R and P in Equation (57) absorb at the same wavelength λ, Equation (66) must be modified to Equation (67):

$$A(t) = \Delta\varepsilon(\lambda) C_R^o \frac{2}{\cos\phi} \left(\frac{4D_R t}{\pi}\right)^{1/2} \tag{67}$$

where $\Delta\varepsilon(\lambda) = \varepsilon_P(\lambda) - \varepsilon_R(\lambda)$. Assume that after a certain period of time τ, following application of the step, the electrode potential is stepped to a new value at which the reverse reaction occurs under diffusion control. Equation (68) presents this in mathematical terms (see section 2.2.1),

$$C_R(0,t) = S_\tau \text{ where } S_\tau = \begin{cases} 0, & t \le \tau \\ 1, & t > \tau \end{cases}, \theta_R(0,t) = 1 - S_\tau,$$
$$\text{and} \quad \bar{\theta}_R(0,s) = \frac{1}{s} - \frac{e^{-\tau s}}{s} \tag{68}$$

and, hence,

$$Q(s) = \left(\frac{D_R}{s}\right)^{1/2} \bar{\theta}_R(0,s) = \frac{D_R^{1/2}}{s^{3/2}} - D_R^{1/2} \frac{e^{\tau s}}{s^{3/2}} \tag{69}$$

After inversion of Equation (69) and substitution in Equation (25) the transient absorbance is given by Equation (70):

$$A(t) = \Delta\varepsilon(\lambda_o) C_R^o \frac{2}{\cos\phi} \left(\frac{4D_R}{\pi}\right)^{1/2} \times t^{1/2} - S_\tau(\tau - t)^{1/2} \tag{70}$$

which reduces to Equation (54) for $t \le \tau$, and, for the second step only Equation (71) holds:

$$A(t > \tau) = A(\tau) - \Delta\varepsilon(\lambda_o) C_R^o \frac{2}{\cos\phi} \left(\frac{4D_R}{\pi}\right)^{1/2} \times \tau^{1/2} - t^{1/2} + (\tau - t)^{1/2} \tag{71}$$

An interesting application of this formalism was described by Jones and Hinman,[64] who examined the redox properties of iron tetraphenylporphyrin chloride (FeTPPCl) in a nonaqueous solution, using the cell shown in Figure 24(a). Figure 27 shows a plot of ΔA versus t obtained for a 10.6 μM FeTPPCl solution in 0.1 M tetra-*n*-butyl ammonium perchlorate (TBAP) in CH_2Cl_2 for an experiment in which the potential was stepped from 0.9 V to 1.3 V versus SCE for 1 min, and then stepped back to 0.9 V. A clear indication that the system behaves in an ideal way was obtained from the linear character of the plots of ΔA versus $t^{1/2}$, and ΔA versus $\tau^{1/2} - t^{1/2} + (\tau - t)^{1/2}$ (Figure 28), which may be used as a criterion to ascertain that, within the timescale of the measurements, the product of the reaction P is stable. As the spectrum of R is often known, measurements of this type at different values of λ, allow the spectrum of the other species to be determined, yielding in this case

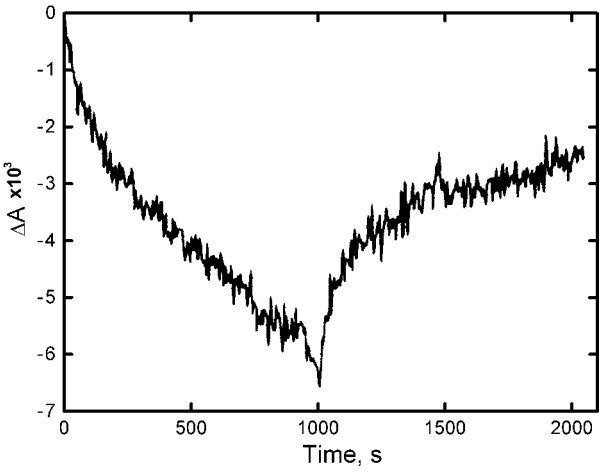

Figure 27 Plot of ΔA versus t obtained at $\lambda = 410$ nm in a 10 μM FeTPPCl solution in CH_2Cl_2 containing 0.1 M TBAP during a double potential step chronoabsorptometric experiment. The data represent an average of 64 measurements.[64]

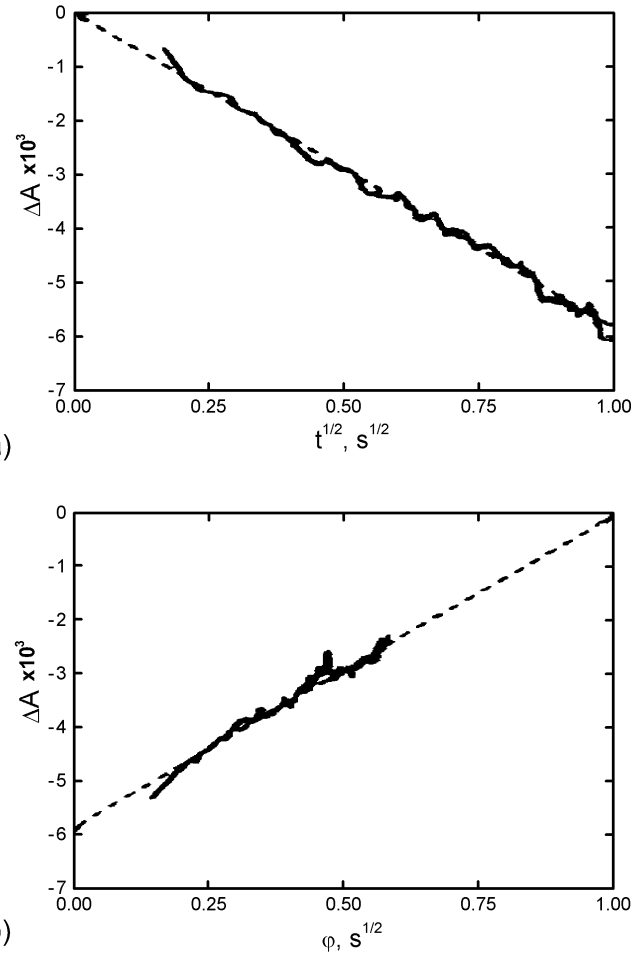

Figure 28 Plots of (a) ΔA versus $t^{1/2}$ and (b) ΔA versus $\varphi = \tau^{1/2} - t^{1/2} + (\tau - t)^{1/2}$ obtained from the data in Figure 27. The broken lines represent a linear least-squares fit.[64]

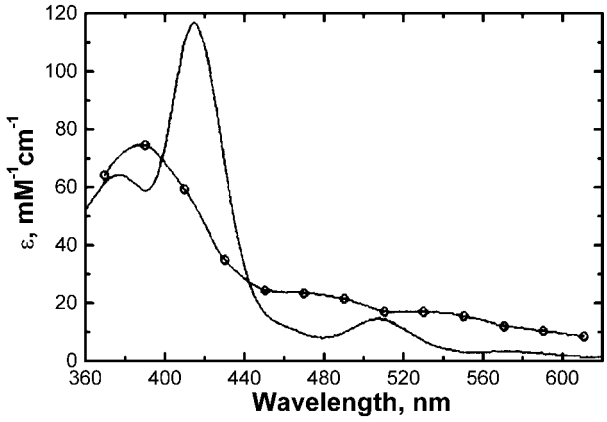

Figure 29 UV/VIS spectrum of FeTPPCl (——) and its one-electron oxidation product (○) calculated from chronoabsorptometric data.[64]

excellent agreement with that of the oxidized form of FeTPPCl obtained by other means (Figure 29).

Multiple Reflection. A different approach[39] to increasing the pathlength is to introduce a well-collimated beam into a cell consisting of two working electrodes placed parallel, and at close distance from each other, as shown schematically in Figure 24. Under these conditions and assuming P is the only absorbing species, $A(t)$ for this multiple reflection configuration is given to a good degree of approximation by

$$A(t) = \varepsilon_P C_R^o \frac{2w}{T \sin\phi} \left(\frac{4D_R t}{\pi} \right)^{1/2} \qquad (72)$$

where w and T are the cell length and cell thickness, respectively. According to Equation (72) the further enhancement in absorbance derived from multiple, compared to single, reflection amounts to a factor $L/(T \tan\phi)$. For systems in which the only absorbing species at λ is the reactant R, the decrease in absorbance after a potential step will also be linear in $t^{1/2}$, namely,

$$A(t) = \frac{\varepsilon_R C_R^o L}{\sin\phi} - \frac{2L}{T \sin\phi} \varepsilon_R C_R^o \left(\frac{4D_R t}{\pi} \right)^{1/2} \qquad (73)$$

The leading term in Equation (73) represents the initial absorbance of the solution, a factor which is independent of T, and $L/\sin\phi$ represents the *total* pathlength.

Shown in Figure 30 are A versus $t^{1/2}$ plots for the reduction of MB on Au in 0.1 M KCl solutions buffered (pH = 7) obtained with the cell shown in Figure 24(c), using an He–Ne laser, following a potential step as a function of MB concentration. Excellent linearity was observed within the timescale of the measurements, except for the most concentrated solution, for which the electrogenerated radical undergoes precipitation.

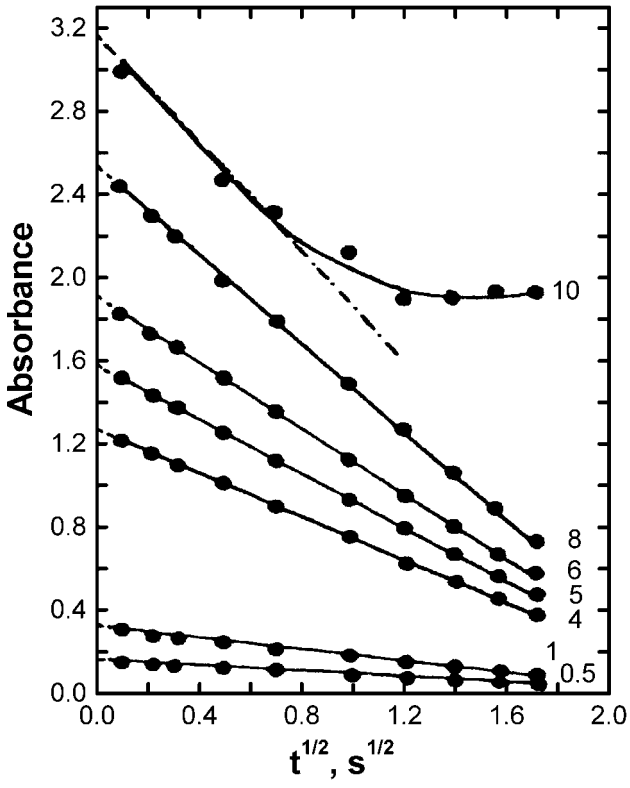

Figure 30 Plots of A versus $t^{1/2}$ for the reduction of MB at concentrations in the range 0.5 to 10 μM.[39]

Furthermore, the magnitude of A, based on ε for MB at 632.8 nm (see Table 4), the geometry of the cell and the angle of incidence, was in agreement with that found experimentally with $\phi < 45°$ for about 330 reflections.

4.4.1.3 Determination of Rate Constants of Homogeneous Electron Transfer Reactions Involving Electrogenerated Species Consider a reaction involving exclusively solution-phase species, A, $A^{+\bullet}$, B and B′, in which A is reduced at the electrode surface to yield $A^{+\bullet}$, which then reacts with B to yield B′ regenerating A (Equation 74):

$$2A^{+\bullet} + B \longrightarrow 2A + B' \tag{74}$$

Two simple reaction mechanisms consistent with this stoichiometry may be envisaged as Equation (75) with (76), and Equation (75) with (77):

$$A^{+\bullet} + B \xrightarrow{k_2} B^{+\bullet} + A \tag{75}$$

$$A^{+\bullet} + B^{+\bullet} \xrightarrow{\text{fast}} B' + A \tag{76}$$

$$2B^{+\bullet} \xrightarrow{\text{fast}} B + B' \tag{77}$$

Assume further, that the rate-determining step is first order both in $A^{+\bullet}$ and B, such as a collision between these species, and that the equilibrium constant for the overall reaction is large so that all reactions following the slow step will proceed to completion. Hence, if Equation (75) is slower than (76) and (77), then Equation (78) holds:

$$\frac{d[A^{+\bullet}]}{dt} = -2k_2[A^{+\bullet}][B] \tag{78}$$

If an electrode immersed in a homogeneous solution containing A and B at known concentrations is stepped to a potential sufficiently positive to form the corresponding radical cation $A^{+\bullet}$ under strict diffusion control diffusing into the solution, then $A^{+\bullet}$ will encounter and react with B diffusing toward the electrode in an electron transfer-type process to yield $B^{+\bullet}$. As the amount of B at, and near the surface is expected to be small, Equation (75) will proceed a distance away from the electrode. Analytic solutions of this rather complex problem do not exist in general; however, it is possible to generate concentration profiles of each of the species involved by digital simulation techniques, provided that all the relevant parameters are specified and, from these, to predict the absorbance associated with the chromophores.

The power of UV/VIS spectroelectrochemistry as a tool to determine rate constants of reactions between an electrogenerated radical with a species in solution, i.e. k_2 in Equations (76) and (77), was demonstrated by McCreery et al.,[67] who measured the rates of $CPZ^{+\bullet}$ with a catecholamine, using, among other techniques, grazing incidence external reflection at a Pt electrode.

Figure 31 shows the results of numerical simulations for the reaction sequence shown above, as $C_{CPZ^{+\bullet}}$ observed in the presence and in the absence of catecholamine, denoted as DA, as a function of distance, expressed in terms of units of $(Dt)^{1/2}$, for four different values of kt assuming equal bulk concentrations for both CPZ and DA. For simplicity, the diffusion coefficients of all species involved were assumed to be the same. Figure 31(a) corresponds to a situation in which $k_2 = 0$, i.e. no subsequent reaction involving $CPZ^{+\bullet}$. For k_2 finite and $t \neq 0$, the concentrations of both $C_{CPZ^{+\bullet}}$ and C_{DA} at a fixed distance from the electrode decrease, and therefore the extent of overlap between the two profiles of the two species, also diminishes.

Based on arguments set forward in previous sections, the absorbance of a chromophore is proportional to the integral of the concentration profile, enabling comparisons to be made with experimental measurements. Figure 32 shows normalized absorbances, i.e. those obtained in the presence and in the absence of equimolar amounts of DA, monitored at wavelengths at which only DOQ, the oxidation product of DA ($\lambda = 400$ nm, upper curve), and $CPZ^{+\bullet}$ ($\lambda = 525$ nm, all

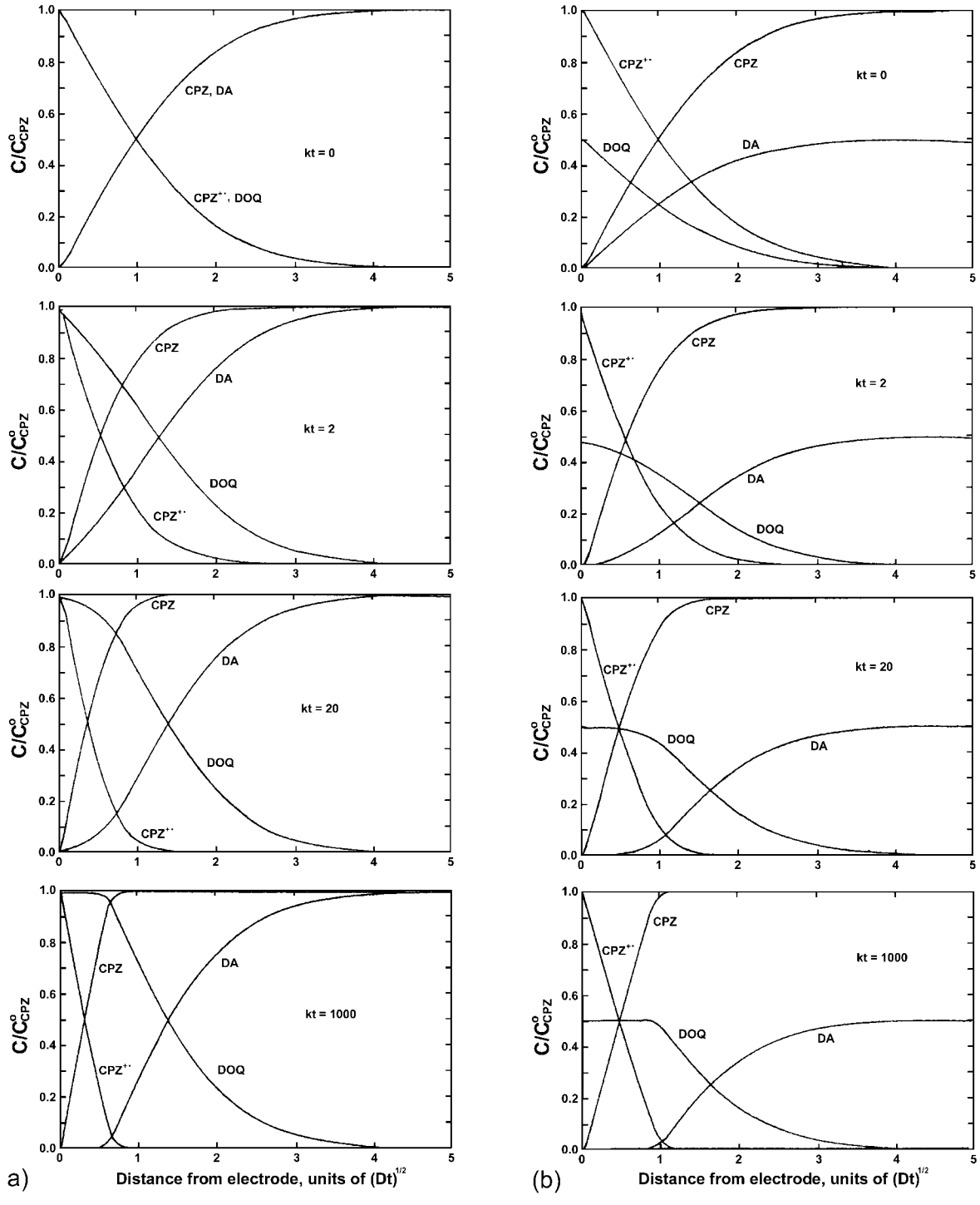

Figure 31 Simulated concentration versus distance profiles for the EE′C reaction sequence at various values of β and *kt*: (a) $C^{o}_{CPZ} = C^{o}_{DA}$ ($\beta = 1$); (b) $C^{o}_{CPZ} = 2C^{o}_{DA}$ ($\beta = 2.0$). In all cases the ordinate represents fractional concentration relative to the C^{o}_{CPZ} value.[67] DOQ, dopamine quinone.

other curves) absorb, respectively. Also included in these plots as solid lines are best-fit numerical simulations for $\beta = C_{CPZ}/C_{DA}$, of 0.5, 1, 2, and 5. 5, yielding k_2 values of $2.12 \pm 0.25\ M^{-1}\ s^{-1}$.

4.4.2 Microelectrodes

Planar electrodes of dimensions larger than about 0.1 mm suffer from relatively slow diffusional relaxation; hence, long times are required to recover the initial conditions

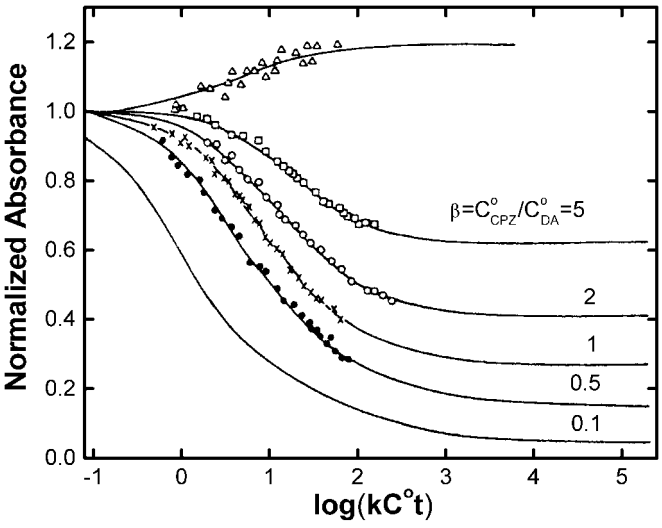

Figure 32 Normalized absorbance versus $\log(kC^{b}t)$ for various values of β. The solid curves are theoretical plot for DOQ monitored at $\lambda = 400\,\text{nm}$ (top curve) or $CPZ^{+\bullet}$ monitored at $\lambda = 525\,\text{nm}$ (remaining curves). Points are best fits to experimental data.[67]

and, thus, acquire sufficient measurements over reasonable times for efficient signal averaging. Considerable enhancements in mass transport in quiescent solutions can be realized by using electrodes of micrometer dimensions, including disks and cylinders. Despite their size, the absorbance observed for such microelectrodes in the external reflection mode is larger than that using an OTE for similar electrolysis times. Because of their small area, both capacitive and faradaic currents are small, placing lower demands on the potentiostat power, increased time averaging, higher duty cycle, low cell-time constants, more precise solution resistance and, most importantly, better conditions for studies of fast reactions.

4.4.2.1 Microcylinders An analytic expression for the time-dependent concentration profile of a species P in a stagnant solution, produced at the circumference of a circular electrode of radius r_0, via oxidation or reduction of R under chronocoulometric conditions, can be obtained by solving the appropriate Fick's law using the Laplace transform method. The solution to this problem may be shown to be given by Equation (79):[37]

where $J_0(\zeta)$, and $Y_0(\zeta)$ are Bessel functions of order zero of the first and second kind of argument ζ, respectively.[23] Figure 33 shows a plot of A_{cyl}/A_{pl} where A_{cyl} and A_{pl} represent the absorbances observed with a cylinder by integration of Equation (79) along r, and with a planar electrode (see Equation 64), respectively, as a function of the dimensionless variable $(D_R t)^{1/2}/r_0$. As would be expected, A_{cyl}/A_{pl} is close to unity for $\delta_N < r_0$ (see Figure 33), but deviates significantly at longer times for fixed r_0.

Figure 34 shows average absorbance versus $t^{1/2}$ plots in the range 0.1–100 ms following a potential step from 0.3 to 1.1 V versus SCE in solutions of 3.5 mM OD in 1 M H_2SO_4 using a Pt microcylinder with $r_0 = 12.5\,\mu\text{m}$. The open circles in this figure were obtained based on the theory introduced above without adjustable parameters. Similar data over a much shorter timescale, down to 4 μs, could be recorded using a Pt microcylinder with $r_0 = 5\,\mu\text{m}$ in 2.2 mM OD in 1 M H_2SO_4 after 2700 averages, under otherwise identical conditions. In this latter case, deviations of only up to 5% could be observed at 1 ms. Both of these experiments involved the use of a laser focused onto the cylinder wall as the radiation source, and optical fibers to collect the reflected light in a two-electrode configuration (Figure 35). The primary factor that limits a quantitative analysis of data acquired at longer times relates to thermal effects induced by heat dissipation within the electrode. Such phenomena restrict the range of validity of the model to times at which δ_N is less than twice the electrode radius, i.e.

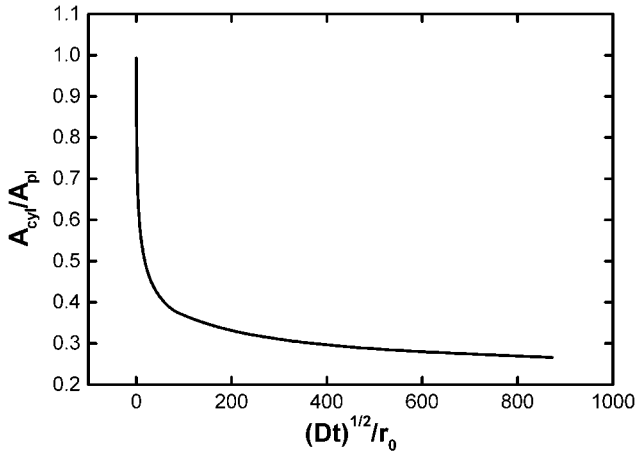

Figure 33 Ratio of absorbance for a cylindrical (cyl) and a planar (pl) electrode as a function of the dimensionless variable $(D_R t)^{1/2}/r_0$.[37]

$$C_P(r,t) = C_R^o \left[1 + \frac{2}{\pi} \int_0^\infty \frac{J_0\left(\frac{r}{r_0}\xi\right) Y_0(\xi) - Y_0\left(\frac{r}{r_0}\xi\right) J_0(\xi)}{J_0^2(\xi) + Y_0^2(\xi)} \left(\frac{\exp[(-D_R t/r_0^2)\xi^2]}{\xi}\right) d\xi \right] \quad (79)$$

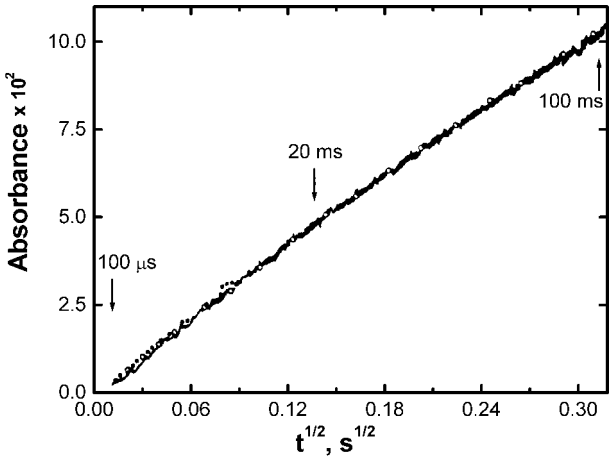

Figure 34 Absorbance versus $t^{1/2}$ for a 12.5 μm radius platinum electrode in 1 M H_2SO_4 containing 3.5 mM OD (*o*-dianisidine): open circles = theoretically predicted response; solid line = experimental response for a conventional three-electrode arrangement, average of 100 runs; solid circles = response for a two-electrode configuration, average of 60 runs. E_{app} was stepped from 0.3 to 1.1 V versus SCE.[37]

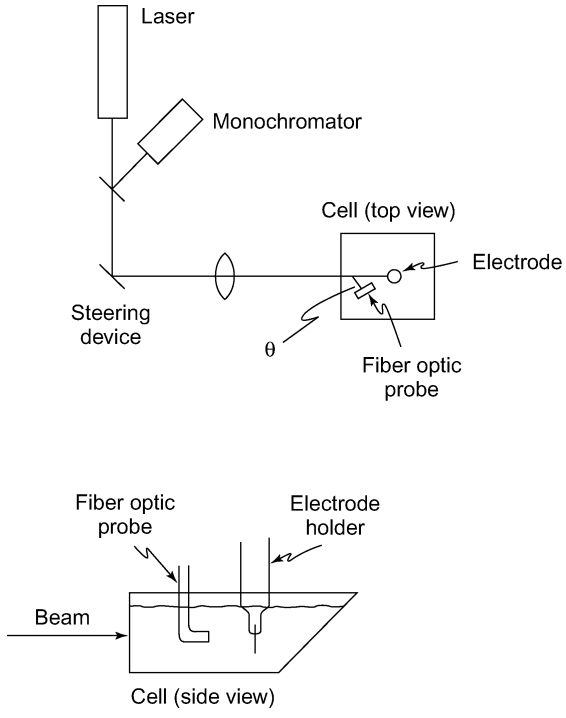

Figure 35 Experimental arrangement for UV/VIS spectroelectrochemical measurements involving a microcylinder electrode.[37]

$(Dt)^{1/2}/r_0 < 2$. As discussed by these authors, both the IR drop, and the time constant of the cell increase as r_0 decreases; hence, smaller-diameter electrodes will not allow access to measurements for times shorter than a few microseconds. Nevertheless, it is possible to achieve submicrosecond resolution by employing a microdisk electrode, as described in the next subsection.

4.4.2.2 Microdisk Electrodes Unlike the case of cylindrical electrodes, the solution resistance R_{sol} for a microdisk is directly, as opposed to inversely, proportional to r_0. As for a typical microdisk electrode, the contribution of the capacity to τ_{cell} is very small compared to that ascribed to R_{sol}, decreasing r_0 should lead to a better transient response. Indeed, chronoabsorptometric experiments involving the use of a coaxial Au microdisk electrode with $r_0 = 30\,\mu m$ in a two-electrode cell configuration, allowed for absorbances associated with electrogenerated $MV^{+\bullet}$ to be monitored in the millimeter range with an Xe arc lamp down to 150 ns following extensive signal averaging (Figure 36).[31] This improved methodology enables measurements of second-order homogeneous rate constants of reactions involving electrochemically generated strong chromophores (k_2 in Equation 75), of the order of $10^8\,M^{-1}\,s^{-1}$, as illustrated by the reaction of CPZ with DA at physiological pH. It should be emphasized that these values exceed those achievable with more conventional stopped flow experiments.

Shown in Figure 37 are absorbance versus time plots for generation of $CPZ^{+\bullet}$ both in the absence and in the presence of DA. The solid lines in the lower curve represent best digital simulation fits to the average of 1120 runs, for which k_2 for the reduction of the radical

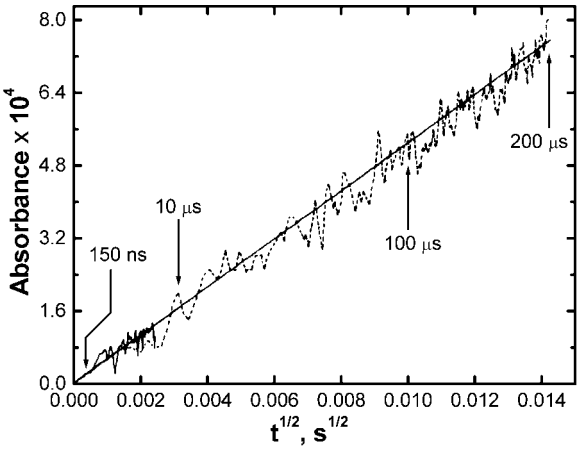

Figure 36 Absorbance versus time curves for the generation of $MV^{+\bullet}$ from 5.6 mM $MVCl_2$ in 2 M phosphate buffer, (pH = 7.0) recorded using a coaxial Au microdisk working electrode (30 μm radius) with an He–Ne laser at λ = 632.8 nm. The dashed line is an average of 60 000 runs covering 1–200 μs. The solid noisy line is a shorter transient at higher time resolution. The straight line is a least squares fit to all of the data, and has the same slope as 5 ms runs.[31]

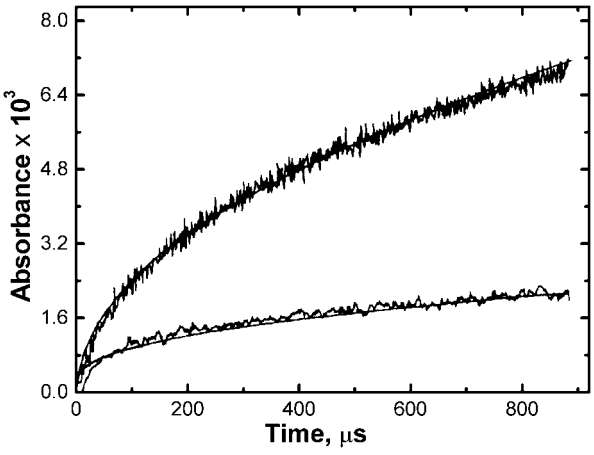

Figure 37 Absorbance versus time curves for the generation of $CPZ^{+\bullet}$ in the presence (lower curve) and absence (upper curve) of equimolar DA in 40% $MeOH/H_2O$, containing 0.2 M dimethylarsenate buffer (pH = 6.8) plus KCl for a total ionic strength of 1.8 M. The noisy lines are averages of 1120 runs. The upper smooth curve was calculated for $k = 0$, the lower curve for $k = 6.2 \times 10^7\ M^{-1}\ s^{-1}$ for the reduction of the radical by dopamine.[31]

by DA is $6.2 \times 10^7\ M^{-1}\ s^{-1}$ (lower curve), whereas that in the upper curves assumes $k_2 = 0$.

4.5 Spatially Resolved Spectroelectrochemistry

Transmission and external reflection spectroelectrochemical methods afford information about integrated profiles of the species being monitored. Additional insight into mechanistic aspects of homogeneous-phase redox reactions and transport phenomena can be obtained from the concentration profiles themselves. Not surprisingly, considerable effort has been devoted to monitor their temporal and spatial evolution using both interferometry[14] and transmission spectroscopy. This section summarizes the most salient features of the latter of these two methodologies and discusses with various illustrations its scope of applicability.

The problems associated with the lack of selectivity and rather weak sensitivity of solution-phase imaging techniques based on interferometry so far reported can be circumvented to a significant extent by using absorption, as opposed to refraction, as the physical phenomenon being monitored. Although restricted to systems involving one or more chromophores, absorption-imaging methods have provided a wealth of high spatial resolution information regarding the time evolution of diffusion profiles induced by heterogeneous electron transfer reactions at electrode–solution interfaces approaching the limits imposed by optical diffraction. Because of the rather low concentrations of reactants often involved in spectroelectrochemical experiments, distortions in

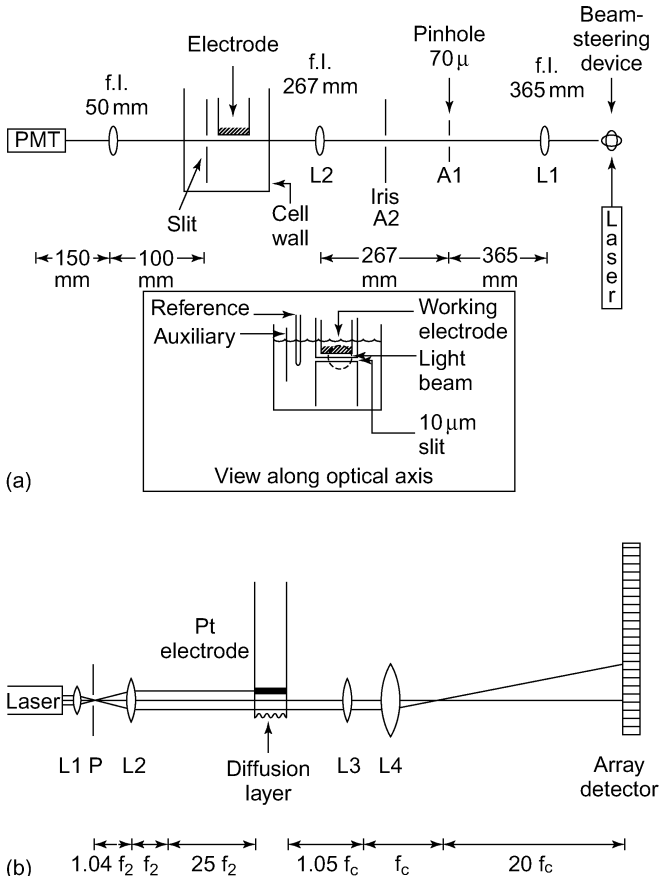

Figure 38 Diagrams of experimental arrangements for profile imaging using (a) a scanning slit[68] and (b) a multidiode array.[40]

the spatially resolved profiles measured by absorption, induced by changes in the index of refraction, can be neglected.

Two types of absorption-based profile imaging methods[40,68] have been described in the literature, involving in both cases weakly focused light propagating parallel to the electrode surface. The first, depicted in Figure 38(a),[68] employs a moving mechanical slit aligned along an axis normal to the electrode, and a photomultiplier to measure the extent of light attenuation through small volumes of solution. The second approach[40,69] (Figure 38b), relies on a photodiode array detector to capture the magnified image of the beam produced either by a laser or a lamp, after crossing the diffusion boundary layer, as a function of time. Analysis of the projection of an expanded image onto such an array enables, within diffraction-limited resolution, construction of time-resolved and spatially resolved maps of light intensities, which can then be related to a physical variable of the system being examined, such as concentration, as a function of distance normal to the probing beam.

4.5.1 Imaging of Diffusion Layers

4.5.1.1 Potential Step

Planar Electrodes. Figure 39 shows a series of $C_{TAA^{+\bullet}}/C^{o}_{TAA}$ versus y (the distance from the electrode surface) plots obtained for single and double potential step experiments in 2.56 and 4.06 mM TAA in acetonitrile solutions, respectively.[40] As shown in Table 4, the species $TAA^{+\bullet}$ exhibits a large ε at $\lambda = 632.8$ nm. Each of the solid curves in (a) and (b) were acquired with an He–Ne laser at different times after application of a potential step from 0.0 to +0.8 V versus SCE, whereas those in (c) and (d) were recorded at various times after returning the potential to 0 V, following polarization at +0.8 for 2 s (see caption). Excellent agreement was found between the results of these experiments using values of ε and D obtained independently, based on solutions of Fick's law (see Equation 23). Deviations observed at times longer than 8 s in (a) and (b), are attributed to the onset of natural convection.

Modifications to the optical set-up[70] in Figure 40, to include an Xe arc lamp instead of a laser, makes it possible to gain access to a much wider spectral region and thereby monitor profiles of a more general class of chromophores. Using such an approach, direct verification of theoretical predictions for times up to 10 s could be obtained for the potential step oxidation of ferrocyanide ion $[Fe(CN)_6]^{4-}$ to the corresponding ferric species.

Overall, analysis of these type of data allows for D for both R and P to be determined without specific knowledge of electrode areas, concentration, molar absorptivity, absorbance, number of electrons transferred or current, as would be required by chronoamperometry and chronoabsorptometry.

Microelectrodes. Imaging techniques have also been implemented for the analysis of cylindrical diffusion fields around thin wires.[71] In this case, a parallel beam of light of dimensions larger than the diameter of the wire is directed normal to the cylinder axis, as

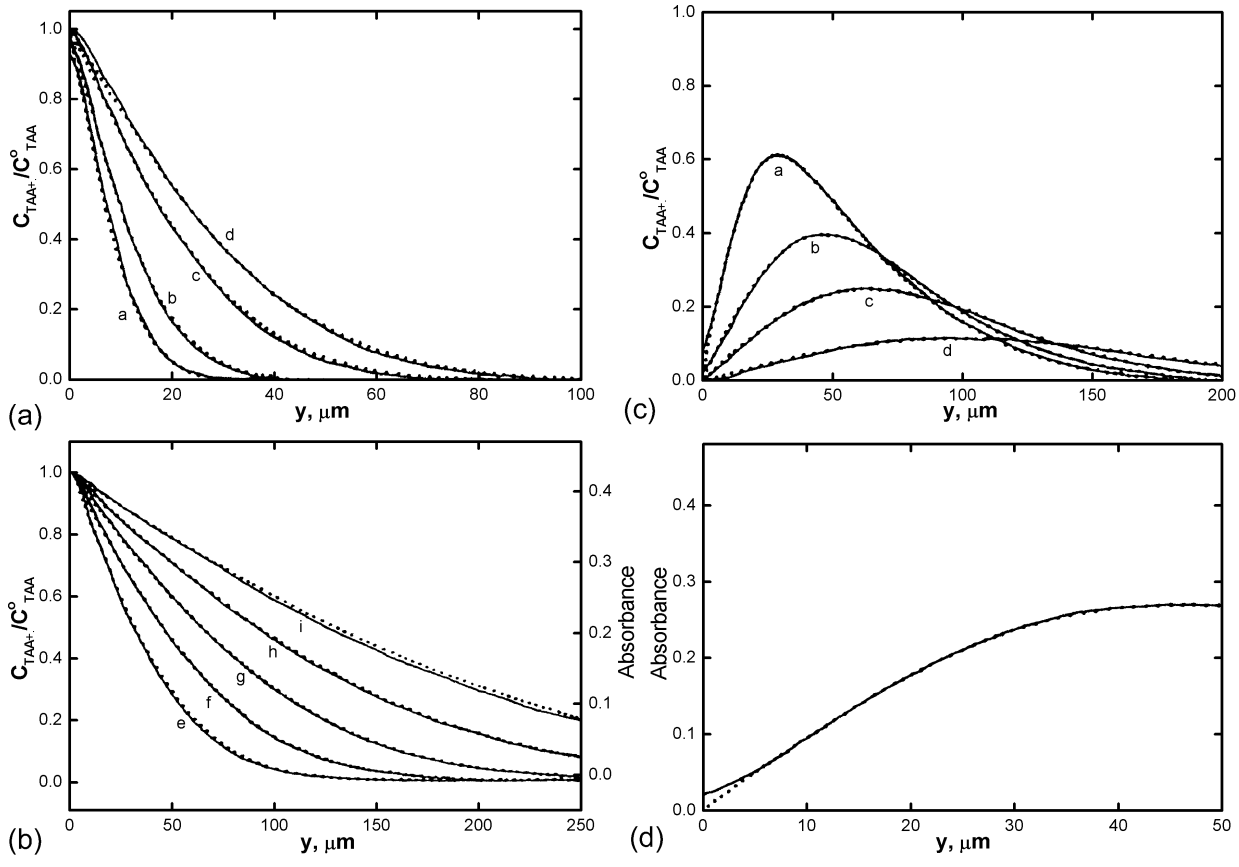

Figure 39 (a and b) Plots of $C_{TAA^{+\bullet}}$ relative to C^{o}_{TAA} as a function of y after various times following application of a single potential step from 0.0 to 0.8 V versus SCE: (a) 0.05 s, (b) 0.1 s, (c) 0.3 s, (d) 0.5 s, (e) 1.0 s, (f) 2.0 s, (g) 4.0 s, (h) 8.0 s, (i) 16 s. Solid lines are experimental profiles: points were calculated from Fick's law and Beer's law with pathlength = 0.015 cm and $\varepsilon = 11\,000\ M^{-1}\ cm^{-1}$; $C^{o}_{TAA} = 2.56$ mM. (c) As for (a) following application of a subsequent step after 2 s at 0.8 V, back to 0.0 V versus SCE: (a) 0.1 s, (b) 0.4 s, (c) 1.0 s, (d) 3 s. (d) Expanded profile of curve b in (c) above; $C^{o}_{TAA} = 4.06$ mM.[40]

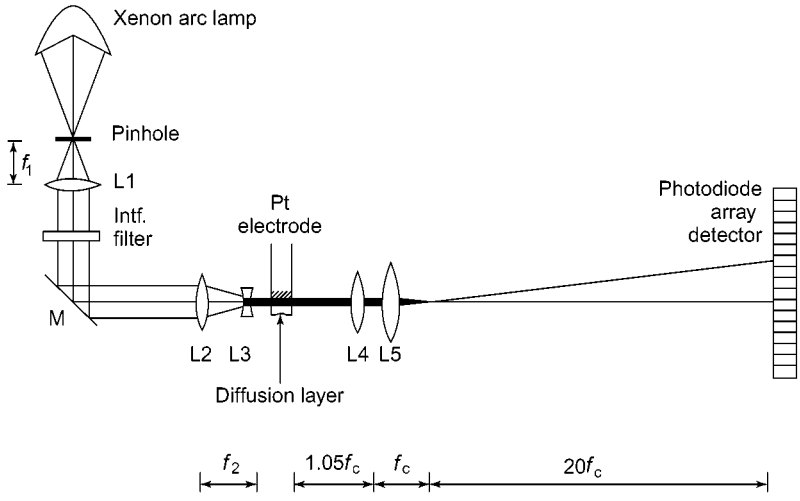

Figure 40 Optical arrangement for profile imaging using an arc lamp and a photodiode array.[70]

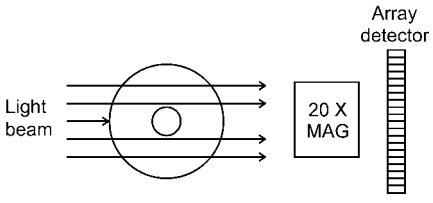

Figure 41 Orientation of microwire (cross-hatched) diffusion field, and detector. Magnification was accomplished with a two-element simple magnifier.[71]

shown in Figure 41, and then magnified before striking the array detector. Concentration profiles based on information obtained from lateral absorbances can be obtained by a mathematical technique known as Abel inversion, a procedure readily amenable to computer implementation. Experimental results in agreement with theoretical predictions have been obtained for microwires of radii 6–25 μm and electrolysis times ranging from 50 ms to several seconds. Good agreement between theory and experiments were obtained for $C(r)$ for various times following the potential step, as shown in Figure 42.

4.5.1.2 Cyclic Voltammetry The same experimental approach enables images of concentration profiles to be obtained for planar electrodes during cyclic voltammetry.[40] For example, Figure 43(a) shows profiles recorded during the forward and Figure 43(b) reverse sweeps for the TAA/$TAA^{+\bullet}$ redox system in acetonitrile. Also depicted in the insert to Figure 43(a) is a plot of absorbance versus potential obtained with a photomultiplier tube (PMT), instead of a diode array, through a slit (see Figure 38) illuminating the region 0.0 ± 0.2 μm, i.e. very close to the electrode surface, under the same conditions. For scan rates up to ca. $80\,mV\,s^{-1}$, these

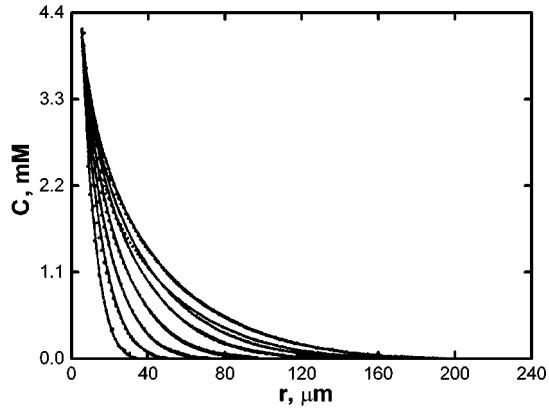

Figure 42 Radial concentration profiles at a 6 μm radius microwire, determined by Abel inversion. Dotted curves are experimental; solid lines are simulated $C(r)$. From the bottom, electrolysis times are 0.2, 0.5, 1, 2, 4, 6, and 10 s; $C^{o}_{CPZ} = 4.2$ mM in 1 M $HClO_4$, 40% MeOH.[71]

curves displayed virtually no hysteresis. At the half-wave potential, i.e. $E_{1/2} = 0.54$ V versus SCE, the surface concentrations of TAA and $TAA^{+\bullet}$ are equal. Under these conditions, the surface absorbance should then be half that of the maximum absorbance for full TAA oxidation, which can be measured at the most positive potentials (see Figure 43). Furthermore, the overall magnitude and shapes of these curves are consistent with the predicted Nernstian behavior for fast heterogeneous electron transfer.

Although the concentration profiles of $TAA^{+\bullet}$ recorded during the forward and reverse scans show common values for the surface concentration, large differences are observed away from the electrode (see Figure 43). This is not surprising as $TAA^{+\bullet}$ accumulates

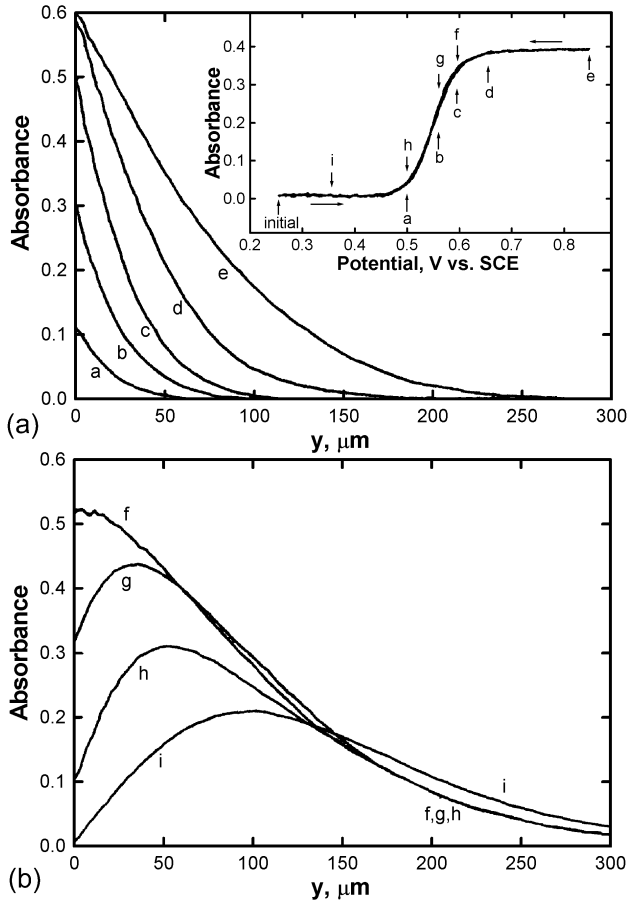

Figure 43 (a) Absorbance versus y, for the forward scan. (b) Same as (a) for the reverse scan during a cyclic voltammogram recorded at a scan rate of $80\,mV\,s^{-1}$ in solution, $C^o_{TAA} = 3.6\,mM$ (see text for details). The inset shows the potentials at which the diode array was triggered to obtain the absorbance profiles.[40]

during the forward scan, diffusing away from the electrode.

4.5.2 *Imaging of Reaction Layers*

The ability of measuring time-, and space-resolved profiles can be advantageous for the evaluation of rate constants of homogeneous electron transfer reactions involving electrogenerated species and the diagnosis of reaction mechanisms.

4.5.2.1 Determination of Rate Constants Scheme 9 represents a process in which a species S, produced by the oxidation of R, reacts with another electroactive species P to regenerate R and the oxidized form of P, denoted as Q. The equilibrium constant of the overall reaction is given by Equation (80):

$$\frac{[R]^{n_2}[Q]^{n_1}}{[S]^{n_2}[P]^{n_1}} = K_{eq} \qquad (80)$$

$$R \longrightarrow S + n_1 e^-$$
$$P \longrightarrow Q + n_2 e^-$$
$$n_2 S + n_1 P \Longleftrightarrow n_2 R + n_1 Q$$

Scheme 9

The magnitude of K_{eq} can be calculated directly from the values of the formal reduction potentials of the two redox couples involved. As demonstrated by theoretical calculations, the concentration profiles of these species are very sensitive to the reaction parameters, including stoichiometry (i.e. n_2/n_1), C^o_P/C^o_R, K_{eq} and the rate constant, often expressed in terms of a dimensionless quantity $\gamma = k_1 C^o_R t$.

An example of this type of mechanism[41,72] is provided by the oxidation of methoxypromazine (MPZ) by the electrochemically generated $CPZ^{+\bullet}$ (=S) at potentials at which $MPZ^{+\bullet}$ (=Q) is also produced as shown in Scheme 10. Equation (81) gives the equilibrium constant for the overall reaction:

$$K_{eq} = \frac{k_1}{k_{-1}} = \exp\left(\frac{nF}{RT}(E_1 - E_2)\right) = 159 \qquad (81)$$

$$CPZ \rightleftharpoons CPZ^{+\bullet} + e^- \qquad E_1 = +0.68\ V \text{ versus SCE}$$
$$MPZ \rightleftharpoons MPZ^{+\bullet} + e^- \qquad E_2 = +0.55\ V \text{ versus SCE}$$
$$CPZ^{+\bullet} + MPZ \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} CPZ + MPZ^{+\bullet}$$

Scheme 10

For these studies, the He–Ne laser was replaced by an Xe arc lamp, thereby allowing the beam to be tuned to the specific absorption bands of the chromophores, i.e. $CPZ^{+\bullet}$ and $MPZ^{+\bullet}$. The contributions due to $MPZ^{+\bullet}$ to the absorbance at $\lambda = 520\,nm$, the absorption maxima of $CPZ^{+\bullet}$, can be determined from the results obtained at 600 nm, a wavelength at which $CPZ^{+\bullet}$, CPZ, and MPZ do not absorb. A comparison between experimental results (solid lines) and best fit simulations (dotted lines) in the form of $C(y,t)/C^o_{CZP}$ versus y, where y is the distance normal from the electrode surface, for $\lambda = 600\,nm$ and 520 nm, are given in Figure 44(a) and (b), respectively, for $t = 1$, 3 and 6 s, assuming $K_{eq} = 159$, $n_1 = n_2$, $\gamma = k_1 C^o_R t = \infty$. The latter condition simply implies that the reaction rate is very large, i.e. no kinetic information can be inferred from this analysis.

Equally large rate constants were found for the cross-reaction between triflupromazine hydrochloride (TPZ) and hydroquinone (H_2Q) for which $n_2/n_1 = 2$. As for K_{eq}, absolute values of n_1 and n_2, can be obtained using conventional electrochemical techniques, or spatially resolved spectroelectrochemistry for the individual redox couples.

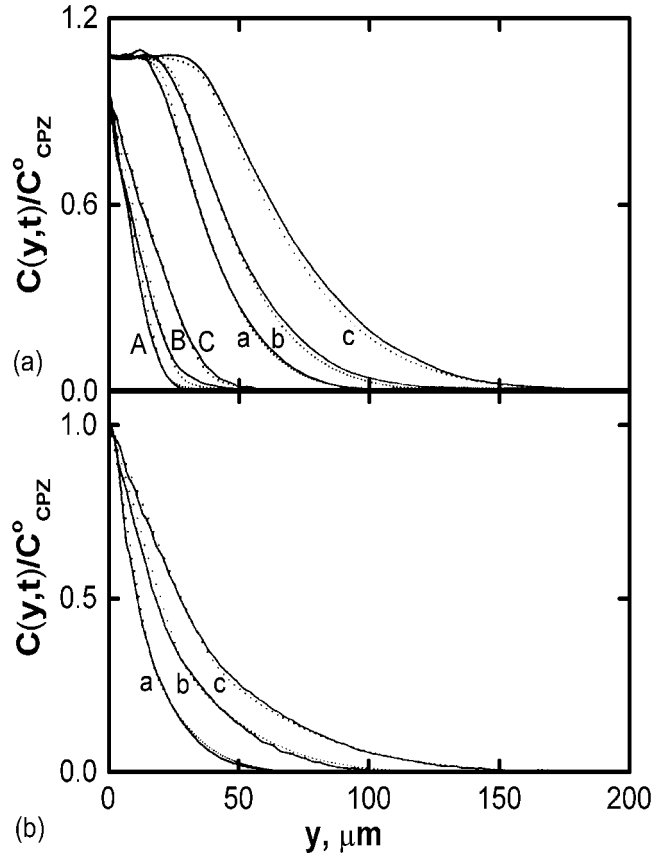

Figure 44 (a) Experimental (——) and simulated (· · ·) concentration profiles for the $CPZ^{+\bullet}$/MPZ cross-reaction: curves A–C are $CPZ^{+\bullet}$, curves a–c are $MPZ^{\bullet+}$. Simulations are for k_1, $\gamma = \infty$, $K_{eq} = 159$. Curves: (A, a) $t = 2$ s; (B, b) $t = 3$ s; (C, c) $t = 6$ s. Experimental profiles determined at $\lambda = 600$ nm for $MPZ^{+\bullet}$, and at $\lambda = 520$ nm for $CPZ^{+\bullet}$, corrected for absorption due to $MPZ^{\bullet+}$. (b) Experimental (——) and simulated (· · ·) absorbance profiles determined at $\lambda = 520$ nm for the CPZ/$CPZ^{+\bullet}$ cross-reaction. γ, $k_1 = \infty$, $K_{eq} = 159$. Curves: (a) 1 s; (b) 3 s; (c) 6 s.[41]

Careful inspection of simulated profiles for the reaction mechanism in Scheme 10 reveals for large values of K_{eq} and infinitely large values of γ, a rather linear region extending from the electrode, at which the dimensionless concentration of S under the operating diffusion control conditions is unity, into the solution. Extrapolation of this profile onto the abscissa may be used to define a dimensionless reaction layer thickness, denoted δ_r, a parameter that depends on n_2/n_1, and the relative bulk concentrations of P and R.

4.5.2.2 Mechanistic Diagnosis Methods based on the measurement of integrated profiles by absorption techniques yielded data consistent with at least two proposed mechanisms for the oxidation of DA by $CPZ^{+\bullet}$, already given in Equations (75–77). In other words, it was not possible to discern whether $DA^{+\bullet}$ reacts with $CPZ^{+\bullet}$ or undergoes dismutation (see Equations 76 and 77) to yield the specified products. As evidenced by numerical simulations,[36] however, the spatial and temporal profiles of $CPZ^{+\bullet}$ do indeed show differences for these two mechanisms and, therefore, direct observation could help identify the preferred pathway.

Comparisons between experimental and theoretical data[36] provided unambiguous proof that Equation (75) with (77) is not operative, thus favoring stepwise electron transfer (Equations 75 and 76) as being the most likely mechanism. Moreover, analysis of the data obtained with a wire electrode revealed that under very acidic conditions the reaction between the neutral species CPZ and DOQ cannot be neglected, i.e. the rate constant is indeed significant. Overall, there appears to be no other method available to date capable of providing this level of detail regarding such a complex reaction.

4.6 Ultraviolet/Visible Spectroelectrochemistry in the Presence of Convective Flow

The specificity of spectroscopic techniques, together with the well-defined mass transport characteristics of the rotating disk electrode (RDE), the rotating ring-disk electrode (RRDE), the channel electrode and, to a lesser extent tube electrodes,[73] provide excellent experimental tools for the quantitative analysis of complex electrode processes. This section illustrates with representative examples, theoretical and experimental aspects of the coupling of UV/VIS spectroscopy and forced convection systems, both under steady-state and transient conditions.

4.6.1 Steady-state Measurements

4.6.1.1 Near Normal Incidence Ultraviolet/Visible Reflection–Absorption Spectroscopy at Rotating Disk Electrodes Optical access to the diffusion boundary layer of an RDE under near-normal incidence conditions requires an experimental arrangement of the type depicted in Figure 45.[74] The beam enters through a window parallel to the electrode surface, travels along the axis of rotation of the disk, reflects off the electrode surface and, after emerging from the cell through the same window, is directed toward the detector. Care must be exercised to set the distance between the electrode surface and the window (Figure 45) sufficiently long to preserve the laminar fluid flow characteristics undisturbed. The amount of product generated during the experiments will be assumed to be sufficiently small so that the bulk solution composition remains unaltered during data acquisition. Such requirements will be closely fulfilled provided the currents are small, the experiments short and the volume of the solutions large. For small incidence angles, such as those used in practice, the

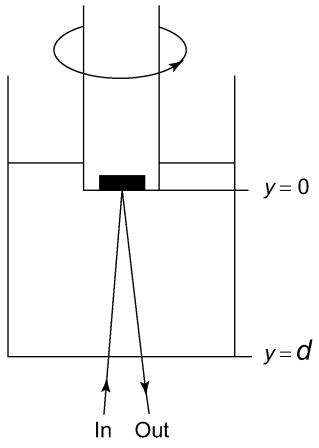

Figure 45 Diagram of the experimental arrangement for reflection–absorption UV/VIS spectroscopy at a RDE at near normal incidence angle. The labels 'In' and 'Out' refer to the incoming and outgoing beams, respectively. Other components have been omitted for clarity.[74]

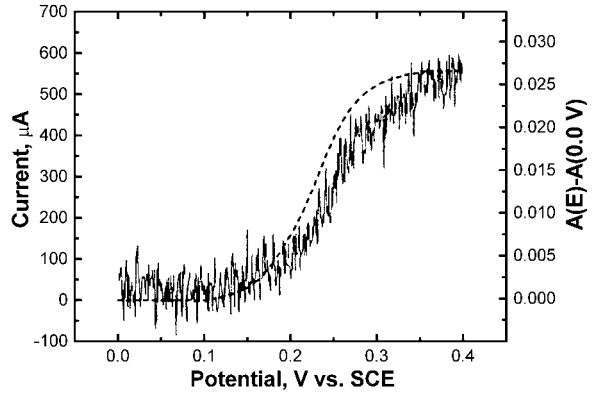

Figure 46 Combined plot of the current (dashed line) and $A(E) - A(0.0\,\mathrm{V})$, as defined in Equation (81), as a function of the applied potential for the oxidation of 0.01 M $[Fe(CN)_6]^{4-}$ in 0.5 M K_2SO_4 aqueous solutions. These curves were recorded at $\omega = 530$ rpm and $\lambda = 420$ nm. The optical and electrochemical data were recorded at scan rates of 100 and 2 mV s^{-1}, respectively.[74]

errors introduced by assuming strict normal incidence are relatively small and therefore will be neglected.

Simple Redox Reactions. Steady-state concentration profiles for a stable product P generated at a RDE along the rotation axis of the disk can be obtained by solving Equation (33) in Table 2, assuming variations in the radial coordinate can be ignored, to yield Equation (82):

$$C_P(\zeta) = \frac{C_P^s}{\Gamma(4/3)} \int_{\zeta_P}^{\infty} \exp(-u^3)\,du \qquad (82)$$

On this basis, the absorbance due to P is given by Equation (83):[74]

$$A(E) = 1.62936\varepsilon\nu^{1/6}D_R^{2/3}D_P^{-1/3}\omega^{-1/2}\frac{1 - C_R^s(E)}{C_R^o} \qquad (83)$$

where the superscript 's' refers to the concentration at the surface, and the potential E in this case is positive (or negative) enough for oxidation (or reduction) to ensue. Shown in Figure 46 is a comparison of plots of the current I versus E (dashed line, left ordinate) and A versus E (solid line, right ordinate) for an Au RDE in a 0.01 M $Fe(CN)_6^{4-}$ solution in aqueous 0.5 M K_2SO_4 recorded at $\omega = 530$ rpm[74] using the arrangement shown in Figure 45. For these experiments, the optical response was monitored while scanning E at relatively high rates of 100 mV s^{-1} whereas the current was recorded at 2 mV s^{-1}. At high scan rates, the concentration profiles cannot achieve steady state; hence, the optical signal is smaller than that expected from the current. This effect is particularly pronounced for potentials at which the current is below its diffusion-limited value i_{lim}, i.e. $C_R^s = 0$. Nevertheless, as predicted by theory, a plot of absorbance at the diffusion-limited current as a function of $\omega^{-1/2}$ was found to be linear, yielding values for ε of $[Fe(CN)_6]^{3-}$ within about 1% of those reported in the literature.

Determination of Faradaic Efficiencies of Complex Processes. Electrosynthesis affords an expedient route for the industrial production of a variety of high-valued chemicals. In addition to cell design considerations and electrode stability, the *overall* efficiency of these processes is often controlled by the kinetics of heterogeneous electron transfer, as well as by the rates of preceding and subsequent chemical reactions. Of common occurrence are situations in which materials, other than the desired product, are generated during operation, thereby decreasing the *specific* faradaic efficiencies, i.e. the fraction of the total current (or charge) involved in the generation of a given product. Quantitative aspects of this phenomenon may be assessed by determining product distributions using conventional analytical techniques external to the reactor or cell. A far more desirable strategy is to acquire such information in situ or on line, i.e. by probing directly the solution adjacent to the electrode surface under well-defined conditions of mass transfer, especially when dealing with rather unstable products.

Within the Levich formalism, the flux j_P of a non-adsorbing product P at the surface of a RDE, may be expressed in terms of the partial current density due to the formation of P (Equation 84):

$$D_P\left(\frac{dC_P}{dy}\right)_0 = \frac{-j_P}{n_P F} \qquad (84)$$

where n_P is the number of electrons required to form one molecule of P. This parameter is defined as negative or

positive depending on whether the process is a reduction or an oxidation, respectively. As is customary, the sign of the current is positive for an oxidation and negative for a reduction.

In terms of the dimensionless variable $\zeta_P = y/\delta_P$, where $\delta_P = 1.805\, D_P^{1/3} \nu^{1/6} \omega^{-1/2}$ is the thickness of the diffusion boundary layer for P, Equation (84) can be rewritten as Equation (85):

$$\left(\frac{dC_P}{d\zeta_P}\right)_0 = \frac{-j_P \delta_P}{n_P F D_P} \tag{85}$$

The solution for the profile of a stable product P, i.e. displaying no decomposition, along an axis normal to the electrode surface at steady state may be expressed in terms of the flux at the surface (Equation 86):[26]

$$C_P(\zeta_P) = -\left(\frac{dC_P}{d\zeta_P}\right)_0 \int_{\varsigma_P}^{\infty} \exp(-u^3)\, du = \frac{-j_P \delta_P}{n_P F D_P} \int_{\varsigma_P}^{\infty} \exp(-u^3)\, du \tag{86}$$

For normal incidence reflection absorption experiments at an RDE, and assuming P is the only optically absorbing species in the media, the absorbance A is given by

$$A = 2\varepsilon_P \int_0^{\infty} C_P(y)\, dy = 2\varepsilon_P \delta_P \int_0^{\infty} C_P(\varsigma_P)\, d\varsigma_P = 2\varepsilon_P \frac{\delta_P^2}{D_P} \frac{j_P}{n_P F} \int_0^{\infty} \int_{\varsigma_P}^{\infty} \exp(-u^3)\, du\, d\varsigma_P \tag{87}$$

The value of the double integral in the right-hand side of Equation (87) is $\Gamma(2/3)/3$, where $\Gamma(2/3)$ is the gamma function of argument 2/3.[23] Hence, upon rearrangement, Equation (87) may be written as

$$\frac{A\omega}{j_P} = 2.9408 \frac{\varepsilon_P}{n_P F} \left(\frac{\nu}{D_P}\right)^{1/3} \tag{88}$$

As the right-hand side of Equation (88) involves parameters intrinsic either to the product (ε_P, n_P, D_P) or to the solution (ν), the quantity $A\omega/j_P$ is constant. In fact, the constancy of this ratio for different E, ω, and C_R may be regarded as proof that the process indeed satisfies the requirements of the model. Note that in the derivation of Equation (88), no restrictions were imposed on the nature of the step-limiting the values of j_P, such as diffusion, heterogeneous or homogeneous reactions, or a combination thereof.

The reduction of bisulfite in mildly acidic solutions generates dithionite $S_2O_4^{2-}$, a material that displays a high ε band centered at $\lambda = 316$ nm, i.e. within a spectral region that does not interfere with other species in the media.

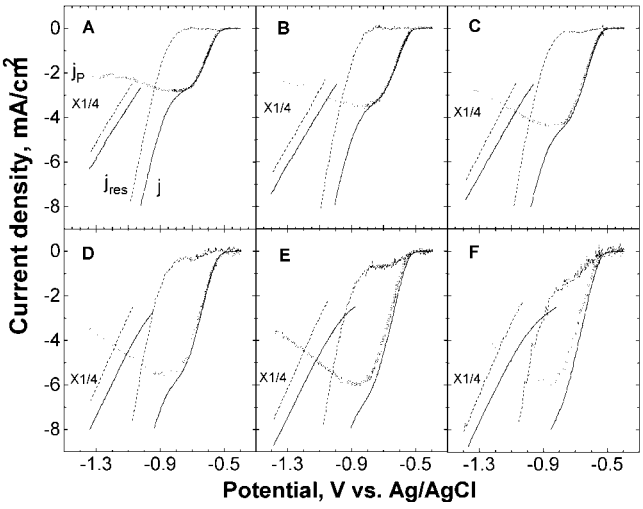

Figure 47 Current density j (solid line), partial current density for dithionite generation j_P (points), and difference between these two currents j_{res} (dotted line), obtained with an Au disk (area 0.452 cm^2) of an Au–Au RRDE assembly in 10 mM Na_2SO_3 in 0.50 M phosphate buffer solution (pH = 5.25), at different rotation rates: A = 200, B = 400, C = 900, D = 1600, E = 2500, and F = 3600 rpm. Values of j_P were calculated from Equation (86) based on the absorbance measured in the reflection absorption mode at near-normal incidence at $\lambda = 316$ nm.[92]

Plots of j versus E, known as polarization curves, where j is the total current density, were recorded with an Au RDE in solutions of 10 mM Na_2SO_3 in 0.50 M NaH_2PO_4 adjusted to pH 5.25 with NaOH using the arrangement shown in Figure 45. At all rotation rates examined, the currents at the inflection (solid lines in Figure 47), were significantly smaller than the diffusion-limited current densities (j_{lim}) for the two-electron oxidation of sulfite in the same electrolyte at 1.00 V, e.g. $j_{lim} = 19.1$ mA cm^{-2} at 900 rpm.

Plots of A at 316 nm versus E recorded during acquisition of I versus E curves in Figure 47, were used to construct j_P versus E plots using Equation (86), and are given in scattered form in the same figure. For these calculations, the constant $A\omega/j_P$ was determined based on ε of $S_2O_4^{2-}$, i.e. $7.3 \pm 0.3 \times 10^6$ cm^2 mol^{-1}, and the diffusion coefficient of $S_2O_4^{2-}$, i.e. $9.1 \pm 0.03 \times 10^{-6}$ cm^2 s^{-1}, respectively, $\nu = 0.010$ cm^2 s^{-1}, the kinematic viscosity of water, and $n = 2$, yielding a value of 9.9 ± 0.6 rpm cm^2 mA^{-1}. Also given in this figure are plots of j_{res} versus E, where j_{res} is the contribution to the total current j due to processes other than bisulfite reduction to dithionite, $j_{res} = j - j_P$.

In the range $-0.60 > E > -0.75$ V, j and j_P were found to virtually coincide at j_{res} about 0. This observation provides convincing evidence that j and j_P are not only proportional, but also that the *specific* faradaic efficiency (j_P/j) for dithionite generation under these conditions is

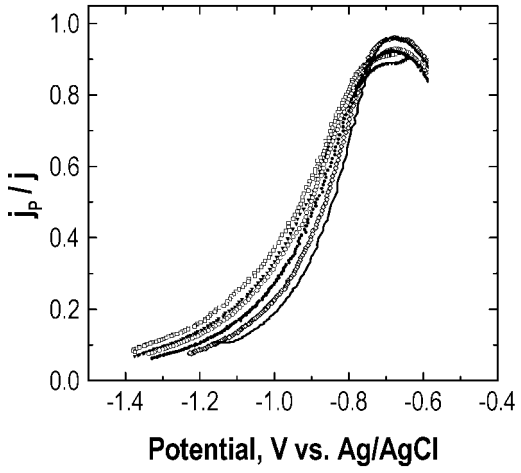

Figure 48 Faradaic efficiency for dithionite generation, i.e. j_P/j, as a function of the applied potential based on the data shown in Figure 47 for six different rotation rates: solid line = 200, diamonds = 400, solid circles = 900, open circles = 1600, solid triangles = 2500, and open squares = 3600 rpm.[92]

within experimental error about 100% (Figure 48). The decrease in the *specific* faradaic efficiency at more positive potentials is probably caused by contributions to the total current due to the reduction of adventitious oxygen in the solution. Furthermore, j_P increased with ω over the entire potential range, indicating that the reaction proceeds under partial mass transport control. However, despite the increase in j as the potential was made more negative, j_P reached a maximum at about -0.80 V (j_P^{max}) for all ω, and decreased steadily thereafter. In addition, plots of j_P^{max} versus $\omega^{1/2}$ and $1/j_P^{max}$ versus $1/\omega^{1/2}$ yielded straight lines with nonzero intercepts, and j_{res} was found to be independent of ω.

4.6.1.2 Rotating Disk Electrode with a Concentric Transparent Ring An interesting forced convection system introduced by Debrodt and Heusler,[75] involves a conventional RRDE assembly in which the metal ring is replaced by a material transparent in the UV/VIS region. This device makes it possible to detect optically absorbing species generated at the disk electrode (Figure 49). Although certain theoretical aspects of this RRDE electrode have been analyzed by digital simulation techniques,[76] analytical solutions can be obtained for simple reactions via the unified formalism given in Tables 2 and 3.[77] Specifically, the dimensionless concentration and flux of a reactant, or a product along the surface of an RDE are constant, and given by $\theta^D = \{1 + \beta^{-1}\}^{-1}$, and $(\partial\theta/\partial\zeta)|_{\zeta=0} = -3\theta^D/\Gamma(1/3)$, respectively, where $\beta = [\Gamma(1/3)/3]k[3/(a\nu D_R^2)]^{1/3}(\nu/\Omega)^{1/2}$.

Based on the second equation for real space in Table 3, the concentration along an axis normal to the surface of an RDE beyond the electrode edge ($1 < \rho, \zeta$) can be

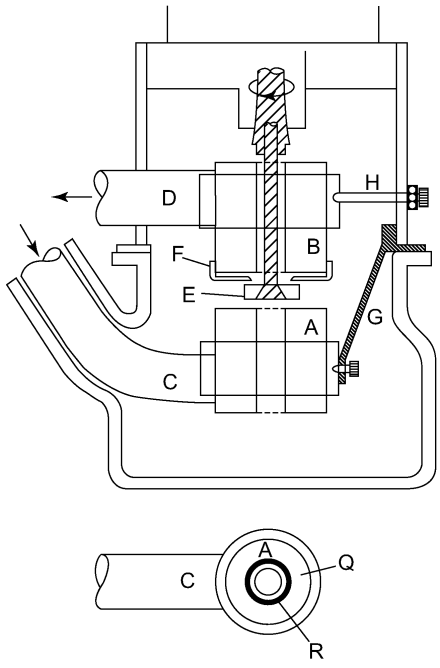

Figure 49 Diagram of a transparent ring RRDE.[75] A, circular light inlet; Q, quartz glass cover; C, optical fiber cable; R, ring; B, D, receivers built similarly; E, rotating disk with quartz ring and electrode; F, aperture; G, H, adjusting devices.

expressed as Equation (89),

$$\Theta = \frac{\theta}{\theta^D} = 3\left[\frac{\Gamma(1/3)}{\Gamma(2/3)}\right]^{-1}\int_0^1 \exp\left[\frac{-\varsigma^3}{(1-\rho/\rho')}\right] \times \left[1-\left(\frac{\rho}{\rho'}\right)^3\right]^{-2/3}\rho'\,d\rho' \quad (89)$$

which at $\zeta = 0$ reduces to Equation (90):

$$\Theta(\rho > 1, \varsigma = 0) = \frac{3}{4} + \sqrt{\frac{3}{4}}\pi \ln\left[\frac{1+g^3}{(1+g)^3}\right] - \frac{3}{2}\pi \arctan\left[(2g-1)\sqrt{3}\right] \quad (90)$$

where $g = (\rho^3 - 1)^{1/3}$. Concentration profiles along ζ for three values of ρ are given in Figure 50.

The integrated profile along ζ, defined as $I(\rho)$ may be written as Equation (91):

$$I(\rho) = \int_0^\infty \Theta(\rho, \varsigma)\,d\varsigma = \left[\frac{\Gamma(1/3)}{\Gamma(2/3)}\right]^{-1}\int_0^1 \rho'(\rho^3 - \rho'^3)^{-2/3} \times \int_0^\infty \exp\left\{\frac{-\varsigma^3}{1-(\rho'/\rho)^3}\right\} d\varsigma\,d\rho' \quad (91)$$

A plot of $I(\rho)$ versus ρ is shown in Figure 51(a).

An explicit expression for the total amount of material probed by the light beam, defined by a cylindrical shell

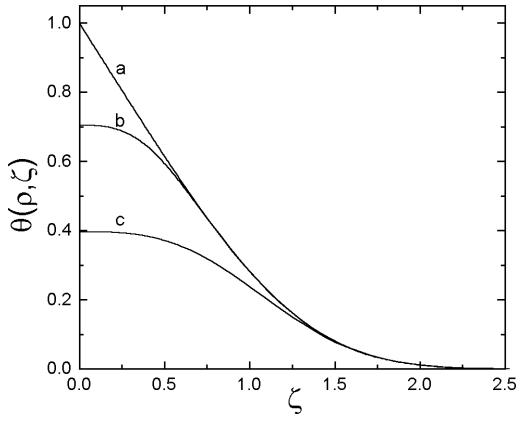

Figure 50 Plot of $\theta(\rho, \zeta)$ versus ζ for three values of ρ, i.e. curve a = 1, b = 65.5/64.5, and c = 74.5/64.5.[77]

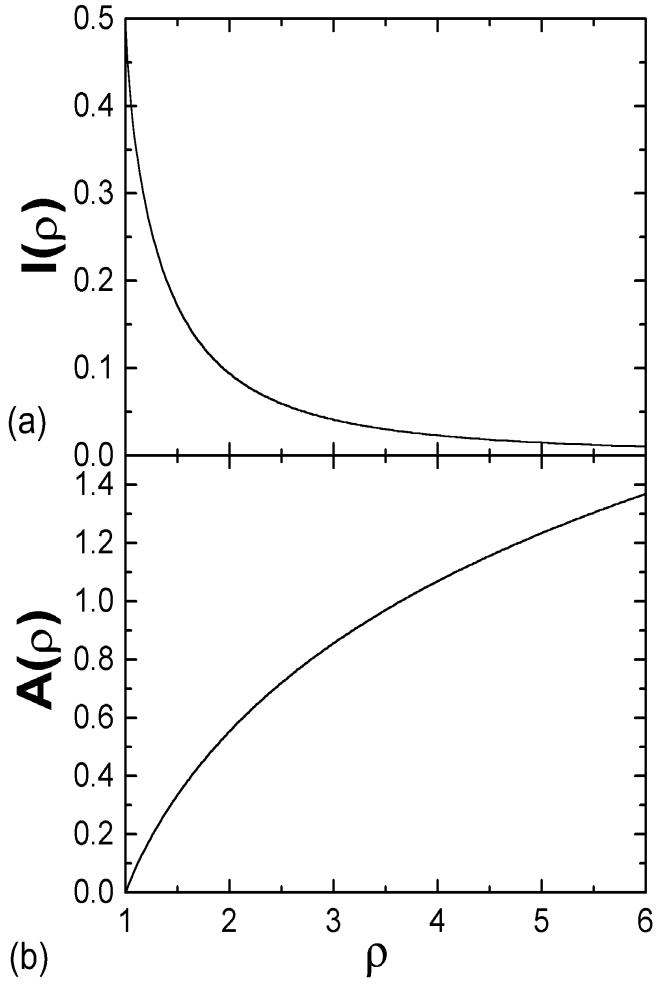

Figure 51 Plots of (a) $I(\rho)$ versus ρ and (b) $A(\rho)$ versus ρ for a transparent ring RRDE.[77]

with inner, and outer dimensionless radii, $\rho_1 = r_1/r_0$ and $\rho_2 = r_2/r_0$, respectively, can be shown to be given by Equation (92):

$$N_{\text{vol}} = N_o\theta^D \left\{ 2 \int_{\rho_1}^{\rho_2} I(\rho;\beta)\rho \, d\rho \right\} = N_o\theta^D[A(\rho_2) - A(\rho_1)] \quad (92)$$

where $A(\rho)$ is given by Equation (93),

$$A(\rho) = 2 \int_0^{\rho} I(\rho)\rho \, d\rho \quad (93)$$

and N_o is the same as for the RDE.[77] A plot of $A(\rho)$ versus ρ is shown in Figure 51(b).

4.6.1.3 Channel Electrodes

Profile Imaging Downstream from the Electrode Surface. The composition of a solution downstream from a channel-type electrode depends, among other factors, on the rates of heterogeneous electron transfer and homogeneous reactions involving electrogenerated species, and their mass transport characteristics. Information regarding various aspects of these processes can be obtained from the analysis of spectroscopic measurements in which a beam oriented normal to the electrode surface is used to monitor the absorbance of the solution along an axis parallel to the fluid flow. A cell arrangement suitable for this type of measurements is shown in Figure 52.[78]

Simple Electrode Reaction. In the case of uniform surface concentration, the integrated concentrations profile of the absorbing product P, denoted as I, at a distance X from the downstream edge of the electrode ($X > 1$) along an axis normal to its surface may be shown to be given by Equation (94):

$$I(X > 1; \infty) = \int_0^{\infty} \theta_P(X, Y)\, dY = \frac{1}{3^{1/3}}\Gamma\left(\frac{2}{3}\right) \times \int_0^{1} x^{-1/3}(X - x)^{-1/3}\, dx \quad (94)$$

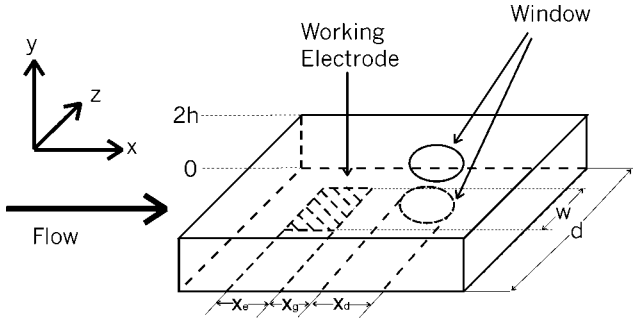

Figure 52 Diagram of a channel cell for spectroelectrochemical studies.[78]

where $\theta_P = (D_P/D_R)^{2/3}(C_P/C_R^o)$. This requirement is fulfilled when the reaction is reversible and, thus, C_R and C_P are prescribed by the Nernst equation, or under diffusion limited conditions.

Figure 53 shows plots of current and absorbance versus E obtained in a solution 0.01 M $K_4Fe(CN)_6$ in 0.25 M K_2SO_4 recorded simultaneously using a potential scan rate of 1 mV s^{-1} and a flow rate of 1.74 mL min^{-1},[79] where the optical monitoring was performed at 1.1 cm from the downstream edge of the electrode. The ordinates of the two curves in this figure are scaled based on the values of A at i_{lim}.

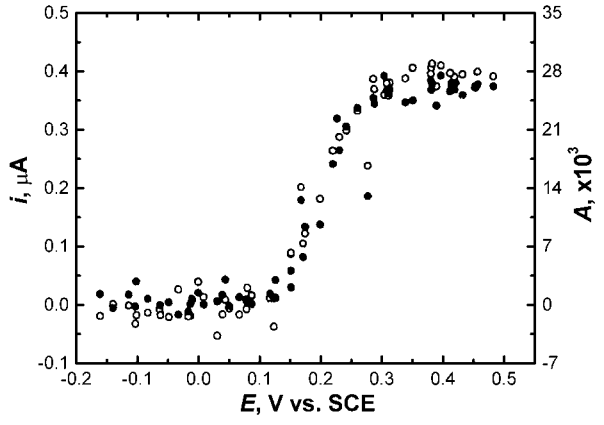

Figure 53 Combined plots of current (left ordinate, solid circles) and absorbance (right ordinate, open circles) versus E for a solution 0.01 M $K_4Fe(CN)_6$ in 0.25 M K_2SO_4 solution. Data were collected simultaneously as E was scanned at 1 mV s^{-1} at a flow rate of 1.74 mL min^{-1}. The ordinates of the two curves have been scaled using an appropriate conversion factor.[79]

Experiments performed under diffusion-limited conditions[80] showed that the absorbance as a function of X for $X > 1$ (solid squares in Figure 54), which may be regarded as an image of the integrated concentration profile, was in quantitative agreement with that predicted

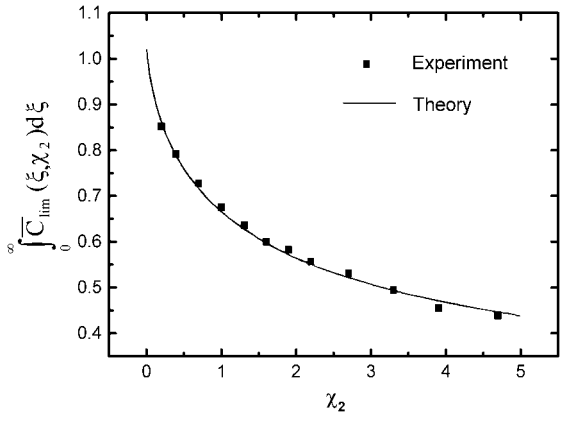

Figure 54 Plot of the steady-state dimensionless integrated concentration profile of $[Fe(CN)_6]^{3-}$, generated at the surface of an Au electrode by the oxidation of 0.01 M $[Fe(CN)_6]^{4-}$ in 0.25 M K_2SO_4, as a function of the dimensionless distance from the downstream edge of the electrode $\chi_2 = x_2/l$. All measurements were performed at $\lambda = 420$ nm along the center axis of the channel in the direction of the flow (solid squares) under diffusion-limited conditions for the oxidation of $[Fe(CN)_6]^{4-}$. The solid curve represents theoretical results (the original paper gives full details).[80]

Table 5 Differential equations and boundary conditions for an EC-type mechanism at a channel-type electrode cell (note that $K = kx_e/6U$)

$D\frac{\partial^2 C_R}{\partial y^2} = U_x\frac{\partial C_R}{\partial x}$			$w\frac{\partial^2 \overline{C}_R}{\partial Y^2} = (Y - Y^2)\frac{\partial \overline{C}_R}{\partial X}$		
$D\frac{\partial^2 C_P}{\partial y^2} - kC_P = U_x\frac{\partial C_P}{\partial x}$			$w\frac{\partial^2 \overline{C}_P}{\partial Y^2} = (Y - Y^2)\frac{\partial \overline{C}_P}{\partial X} + K\overline{C}_P$		
$x = 0$	$y \geq 0$	$C_R = C_R^o$ $C_P = 0$	$X = 0$	$Y \geq 0$	$\overline{C}_R = 1$ $\overline{C}_P = 0$
$0 < x \leq x_e$	$y = 0$	$C_R = C_R^s$	$0 < X \leq 1$	$Y = 0$	$\overline{C}_R = \frac{C_R^s}{C_R^o}$
		$C_P = C_P^s$			$\overline{C}_P = \frac{C_P^s}{C_R^o}$
$x > x_e$	$y = 0$	$\frac{\partial C_R}{\partial y} = 0$	$X > 1$	$Y = 0$	$\frac{\partial \overline{C}_R}{\partial Y} = 0$
		$\frac{\partial C_P}{\partial y} = 0$			$\frac{\partial \overline{C}_P}{\partial Y} = 0$
$x > 0$	$y = 2h$	$\frac{\partial C_R}{\partial y} = 0$	$X > 0$	$Y = 1$	$\frac{\partial \overline{C}_R}{\partial Y} = 0$
		$\frac{\partial C_P}{\partial y} = 0$			$\frac{\partial \overline{C}_P}{\partial Y} = 0$

by theory (solid line in the same figure), lending strong support to the validity of the underlying assumptions of the model.

Determination of the Rate Constant of an Electrochemical–Chemical Mechanism. Consider a simple EC mechanism of the type shown in Scheme 2, where k is the rate constant of the homogeneous chemical reaction. Within the framework of approximations specified in section 2.2.2, and assuming $D_P = D_R = D$, the steady state convective diffusion equations for R and P are given in the right-hand panel of Table 5, where the fluid velocity U_x is given by

$$U_x = \frac{3}{2}U \quad 1 - \frac{(h-y)^2}{h^2} \tag{95}$$

In Equation (95) U is the mean flow velocity, and h is the half-height of the channel.

The appropriate boundary conditions, summarized in the lower section of the same panel, specify that the concentration of the two species is constant over the entire surface of the electrode, which implies that the electrochemical process is infinitely fast. The last entry in that table means that no reactions occur at the optically transparent surface directly opposite to the electrode along the entire length of the channel. This problem, cast in terms of the dimensionless variables $Y = y/2h$, $X = x/x_e$, $\overline{C}_R = C_R^s/C_R^o$, and $\overline{C}_P = C_P^s/C_R^o$ is given on the right-hand panel in the table.

If P is the only absorbing species at the selected λ, plots of absorbance versus $-\ln w$, where $w = (Dx_e)/(24Uh^2)$, $K = (kx_e)/(6U)$, determined by digital simulations yielded for different X and $w/K = 0.01$, showed clearly defined maxima for each of the curves (Figure 55a).[81] Maxima were also observed for a fixed value of $X = 1.75$ and different values of w/K, as shown in Figure 55(b).

Based on these results, an empirical relationship was obtained between the flow rate and the position along the x-axis at the absorption maxima, V_m and x, respectively (Equation 96),

$$\log V_m = \log\left(\frac{2}{3}h^2 x_e b \quad + 1.50\log k_1 - 0.50\log D + \log\left(-4.45 + 4.72\frac{x}{x_e}\right)\right. \tag{96}$$

where, as before, x_e is the length of the electrode, and b the channel width. Hence, it becomes possible to extract values for k_1 by determining experimentally such maxima using the required physical dimensions of the cell and the electrode, as well as the diffusion coefficient. From a practical viewpoint it is more convenient to fix x and measure the relative absorbance for various values of V_m.

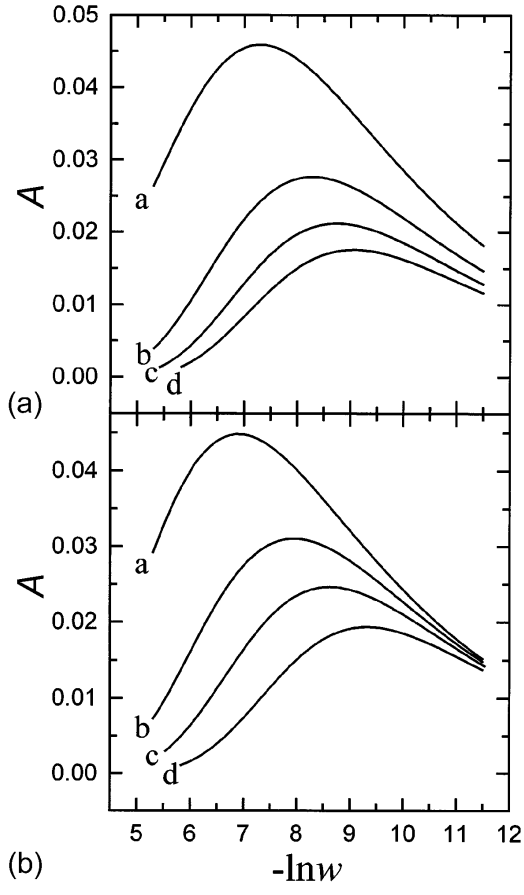

Figure 55 (a) Absorbance versus $-\ln w$ at $w/K = 0.01$ for different X: curve a = 1.25, b = 1.75, c = 2.25, d = 2.75. (b) Absorbance versus $-\ln w$ at $X = 1.75$ for different w/K: curve a = 0.025, b = 0.0125, c = 0.008, d = 0.005.[81]

Figure 56 shows a plot of f/q, a quantity proportional to the normalized absorbance ratio at two values of V_m, corrected for contributions due to other absorbing species in the media, as a function of V_m, for the oxidation of *p*-aminophenol (PAP) to yield *p*-benzoquinoneimine (BQI). This latter species undergoes hydrolysis to form benzoquinone (BQ).[82] The solid line in this figure represents the best fit to the experimental points. Based on the maximum in this curve, the pseudo-first-order rate constant k for the irreversible hydrolysis of BQI can be evaluated using Equation (96), yielding values of about $0.33\,s^{-1}$.

4.6.2 Chronoamperometry under Forced Convection

4.6.2.1 Rotating Disk Electrodes Consider a chronoamperometric experiment in which the potential of an RDE immersed in a solution containing a nonabsorbing

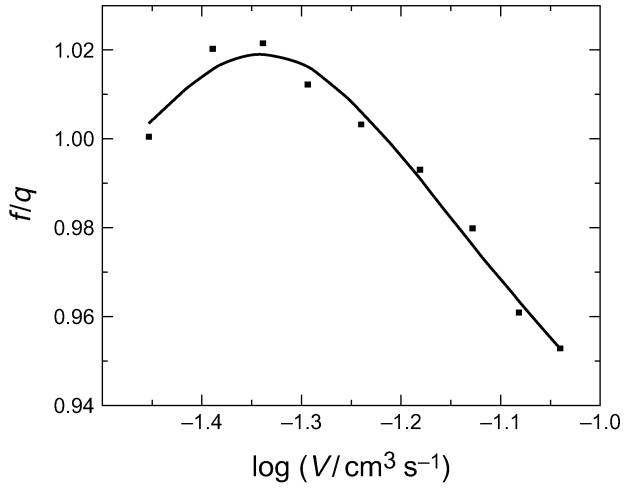

Figure 56 Plot of f/q versus $\log V$ for $X = 2.5$. Solid squares are experimental data and the solid line is a fit to obtain the maximum of f/q.[82]

electroactive species R is stepped from a value E_i, at which no reaction occurs, to a potential sufficiently positive (or negative) E_f for R to undergo oxidation (or reduction) generating an optically absorbing product, P. According to Beer's law, the instantaneous change in absorbance A along y can be written as

$$A(t) - A(t=0) = 2\varepsilon_P \int_0^\infty C_P(y,t)\,dy \qquad (97)$$

In Equation (97), $C_P(y,t)$ is the transient concentration profile of P, which can be determined based on the solution of the transient convective diffusion equation subject to the appropriate initial and boundary conditions. Solutions to this problem, for a reactant, have been reported for short times by Krylov and Babak,[83] in terms of parabolic cylinder functions, $D_k(z)$, and for long times by Nicancioglu and Newman in terms of (numerically evaluated) eigenvectors and eigenfunctions

Table 6 Transient convective diffusion to a RDE

Governing equation

$$D\frac{\partial^2 C_R}{\partial y^2} - v_y\frac{\partial C_R}{\partial y} - \frac{\partial C_R}{\partial t} = 0$$

Boundary conditions for concentration step problem

$$C_R(0,t) = 0$$
$$C_R(\infty,t) = C_o$$
$$C_R(y,0) = C_o$$

Short-time solution	Long-time solution
Dimensionless variables	
$c = \frac{C_R}{C_o}, \tau = (DA^2)^{1/3}t, x = \left(\frac{A}{D}\right)^{1/3} y; A = a\left(\frac{\omega^3}{\nu}\right)^{1/2};$ $z = x(2\tau)^{-1/2}$	$\Theta = \frac{C_o - C_R}{C_o}, \theta = \omega\left(\frac{D}{\nu}\right)^{1/3}\left(\frac{a}{3}\right)^{2/3} t; \varsigma = \left(\frac{a\nu}{3D}\right)^{1/3}\left(\frac{\omega}{\nu}\right)^{1/2} y$
Convective diffusion equation	
$\frac{\partial^2 c}{\partial x^2} + x^2\frac{\partial c}{\partial x} - \frac{\partial c}{\partial \tau} = 0$	$\frac{\partial^2 \Theta}{\partial \zeta^2} + 3\zeta^2\frac{\partial \Theta}{\partial \zeta} - \frac{\partial \Theta}{\partial \theta} = 0$
Boundary conditions	
$c(0,\tau) = 0$ $c(\infty,\tau) = 1$ $c(z,0) = 1$	$\Theta(0,\theta) = 1$ $\Theta(\infty,\theta) = 0$ $\Theta(\zeta,0) = 0$
Dimensionless profile	
$c(z,\tau) = \sum_{n=0}^{\infty} \tau^{3n/2} G_n(z)$	$\Theta(\zeta,\theta) = \Theta^{ss} - \Theta^{t}$
$G_n(z) = e^{-z^2/4} \sum_{k=-3n-1}^{3n-1}{}^{**} b_k^{(n)} D_k(z)$	$\Theta^{ss} = \frac{1}{\Gamma(4/3)}\int_\zeta^\infty e^{-x^3}\,dx, \quad \Theta^{t} = \sum_{n=0}^{\infty} B_n Z_n(\zeta) e^{\lambda_n \theta}$ $B_n = \frac{\int_0^\infty \Theta^{ss} e^{\zeta^3} Z_n(\zeta)\,d\zeta}{\int_0^\infty e^{\zeta^3} Z_n^2(\zeta)\,d\zeta}$
Dimensionless flux	
$\frac{j(\tau)}{j(\infty)} = \frac{a^{1/3}}{0.62(\pi\tau)^{1/2}}\left[1 + \sum_{n=1}^{\infty}\left(-\pi \sum_{k=-3n-1}^{3n-1}{}^{**} \frac{2^{k/2} b_k^{(n)}}{\Gamma(-k/2)}\right)\tau^{3n/2}\right]$	$\frac{j(\theta)}{j(\infty)} = -\frac{3a^{1/3}}{0.62}\left(\left.\frac{\partial \Theta^{ss}}{\partial \zeta}\right\|_{\zeta=0} - \left.\frac{\partial \Theta^{\tau}}{\partial \zeta}\right\|_{\zeta=0}\right.$

of the associated Sturm-Liouville system.[84] A summary of the two mathematical formalisms, which include the appropriate dimensionless variables and the dimensionless profiles, are given in Table 6. Explicit expressions for the coefficients $b_k(n)$ may be found in the original literature.[83] Exact concentration profiles based on these two solutions spanning a wide range of dimensionless times are shown in Figure 57(a). These were used to calculate the corresponding transient integral profiles involved in the analysis of the spectroscopic data, assuming $D_R = D_P$ (Figure 57b).

Figure 58 shows curves of $[A(t) - A(t = 0)]$ versus t for measurements performed in 0.01 M $K_4Fe(CN)_6$ in 0.5 M K_2SO_4 aqueous solutions using an Au RDE in the arrangement shown in Figure 45,[85] with the spectrophotometer set at $\lambda = 420$ nm (see Table 4), where the ε value of $Fe(CN)_6^{4-}$ is very small. Reflection absorption data were acquired at various rotation rates following a potential step from 0.0 to 0.4 V versus Ag/AgCl, which is positive enough for the oxidation of the ferrous species to occur under complete mass transport control. The results obtained for each cycle, consisting of a forward and backward step, were co-added using a signal averager. Also shown in this figure as dotted lines are the theoretical results based on values of ε for $[Fe(CN)_6]^{3-}$ at 420 nm $= 1.14 \times 10^6$ mol^{-1} cm^2, $D = 7.3 \times 10^{-6}$ cm^2 s^{-1}, $\nu = 0.01$ cm^2 s^{-1}, and the specified rotation rate ω (radians per second), using the appropriate conversion between dimensionless and actual integral profiles (Equation 98):

$$(1 - C)\,\mathrm{d}(z\tau^{1/2}) = a^{1/3} \left(\frac{\omega}{2}\right)^{1/2} D^{-1/3}\nu^{-1/6}\, c\,\mathrm{d}y \qquad (98)$$

where $a = 0.51024$. As can be seen, the agreement between theory and experiment for the three rotation rates examined is very good.

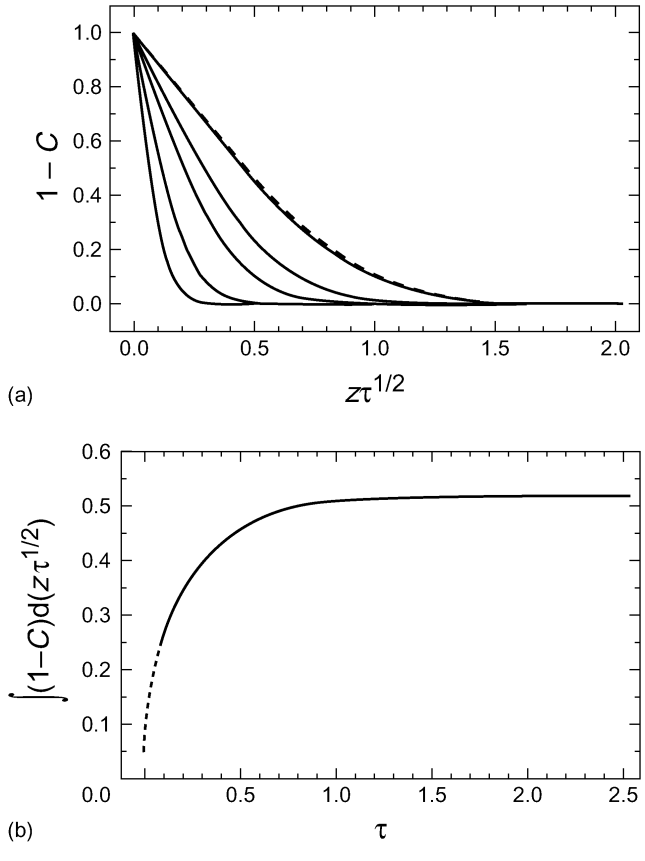

Figure 57 (a) Exact concentration profiles obtained using the short- and long-time solutions, for (curves from left to right) $\tau = 0.01$, 0.03, 0.1, 0.2, 1.0, and the steady state (dashed line). (b) Theoretical integral concentration profiles versus time using the data shown in (a); the dashed and the solid lines were obtained from the short-time and long-time solutions.[85]

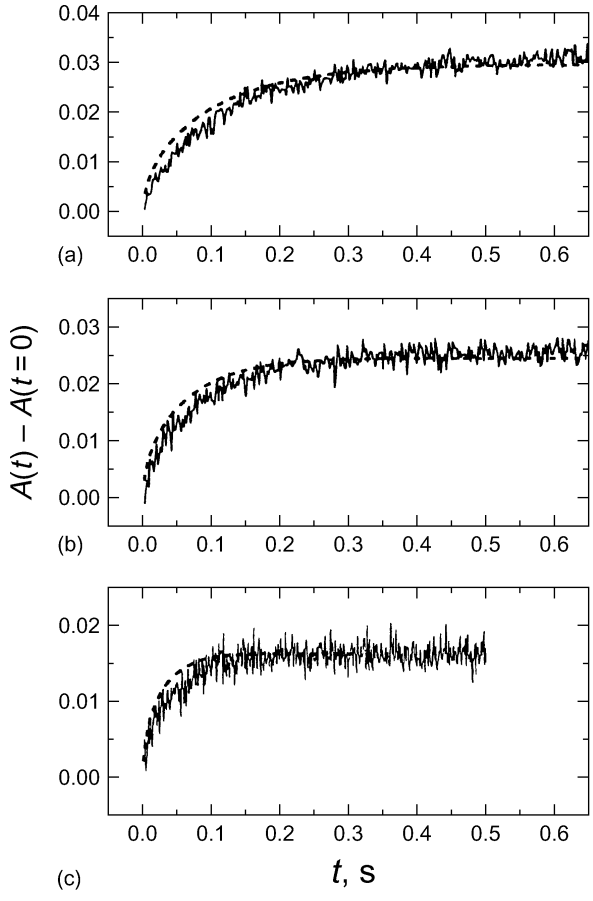

Figure 58 Experimental (difference) absorbance $[A(t) - A(t = 0)]$ versus t: (a) $\omega = 400$ rpm, (b) $\omega = 650$ rpm, (c) $\omega = 1400$ rpm; the dashed lines are the corresponding theoretical curves.[85]

4.6.2.2 Channel Electrodes The transient convective diffusion equation governing mass transport of a solution

phase species C in a channel is given by Equation (99):

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial y^2} - v_x\frac{\partial C}{\partial x} \tag{99}$$

where $v_x = (3V_f/4hd)[1 - (h-y)^2/h^2]$ and other terms have their usual meanings.

Consider a chronopotentiometric experiment in which a potential of large enough magnitude is applied to an electrode in the channel to reduce instantaneously C_R to zero at the surface, assuming, that $C_P = 0$ before the potential step is applied. The initial and boundary conditions for R and P may thus be summarized as Equations (100–103):

$$\text{for } t < 0: \quad y = 0,\ 0 < x < (x_e + x_g + x_d),$$
$$C_R = C_R^o,\ C_P = 0 \tag{100}$$

$$\text{for } t > 0: \quad 0 < x < x_e;\ C_R = 0,$$
$$D_R \left.\frac{\partial C_R(y,t)}{\partial y}\right|_{y=0} = -D_P \left.\frac{\partial C_P(y,t)}{\partial y}\right|_{y=0} \tag{101}$$

$$\text{for all } t: \quad y = 0,\ x_e < x < (x_e + x_g + x_d);$$
$$D_R \left.\frac{\partial C_R(y,t)}{\partial y}\right|_{y=0} = D_P \left.\frac{\partial C_P(y,t)}{\partial y}\right|_{y=0} = 0 \tag{102}$$

$$y = 2h,\ 0 < x < (x_e + x_g + x_d):$$
$$D_R \left.\frac{\partial C_R(y,t)}{\partial y}\right|_{y=0} = D_P \left.\frac{\partial C_P(y,t)}{\partial y}\right|_{y=0} = 0 \tag{103}$$

where x_e is the width of the electrode and x_g and x_d are the distances from the downstream edge of the electrode to the area of the solution being monitored optically along an axis normal to the plane of the electrode, as shown in Figure 52.

The time-dependent concentration profiles for all species involved can be obtained by digital simulation techniques, and the values of the parameters then adjusted until a best fit to the experimental data is found. Figure 59(a) shows values of the normalized absorbance for the one-electron oxidation of ferrocyanide, and Figure 59(b) is the reduction of ferricyanide via chronopotentiometric techniques acquired at wavelengths at which the absorbance of each of the species is maximum.[78] The solid lines in these figures are calculated best-fit curves to the experimental data. Using the value of D_R determined from i_{lim} (Equation 104),

$$i_{lim} = 0.925nFwC_R^o\left(\frac{V_f D_R^2 x_e^2}{h^2 d}\right)^{1/3} \tag{104}$$

for different flow rates and the physical dimensions of the channel, the ratio of the diffusion coefficients of the two species was found to be in excellent agreement with data reported in the literature, i.e. $D_{ferro}/D_{ferri} = 1.17 \pm 0.05$.

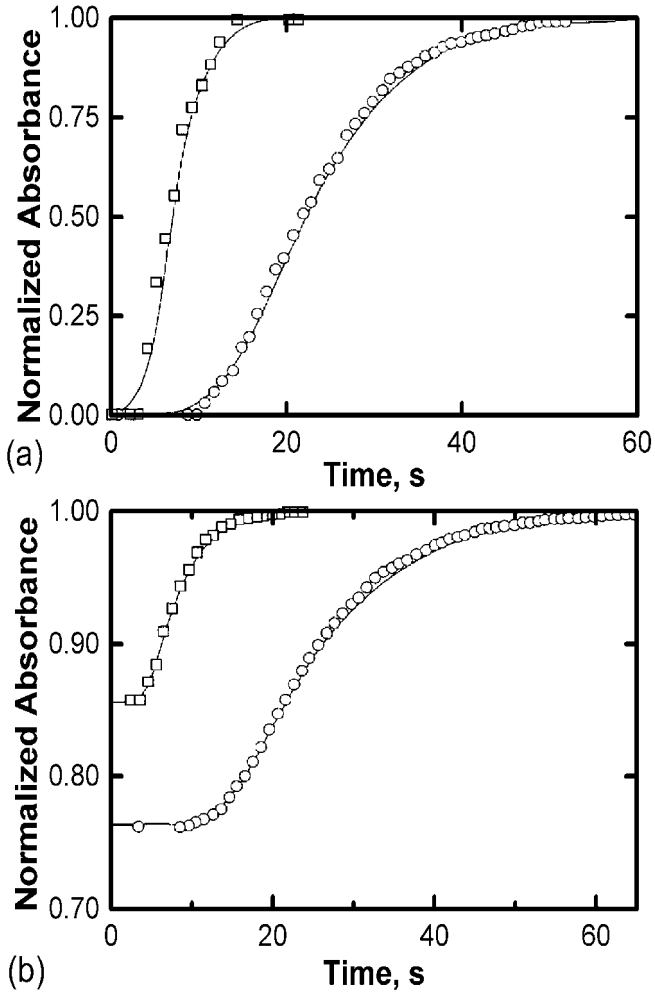

Figure 59 (a) Transient absorbance at $\lambda = 420$ nm for the one-electron oxidation of $Fe(CN)_6^{4-}$ and (b) $\lambda = 218$ nm for the one-electron reduction of $Fe(CN)_6^{3-}$ measured using the channel cell with cell depths of 0.0805 cm and 0.0792 cm, respectively, following a potential step at the working electrode. Solution flow rates: $= 1.44 \times 10^{-3}$ cm^3 s^{-1}; (○) 7.51×10^{-3} cm^3 s^{-1}. Solid lines represent theoretical curves.[78]

Application of this method revealed very small differences between the diffusing coefficients of electrogenerated radical species and the neutral precursors in nonaqueous solutions, which suggests that the interactions of the two types of species with the solvent are indeed quite similar.

4.7 Modulation Techniques

Application of an external periodic potential to an electrical system, such as an electrochemical cell, will elicit corresponding periodic variations in the current. Analysis of the amplitudes and phases of the response as

a function of the perturbation frequency constitutes the basis of a general method for the study of the dynamic behavior of electronic circuitry known as impedance spectroscopy. The mathematical treatment of modulation techniques can be simplified considerably by employing periodic signals of a small enough magnitude so that the behavior of the system may be represented in terms of linear equations. This approach allows contributions to the observed signals derived from other sources, such as noise, to be minimized or eliminated, by extracting that component of the response that oscillates at the same frequency as the exciting perturbation. Also amenable to experimental determination are periodic changes in the concentration profiles of species consumed or generated by faradaic processes brought about by the modulating potential, which can be monitored optically by relying either on absorption of UV/VIS chromophores, or on beam-bending effects derived from local changes in the index of refraction along an axis normal to the beam direction.

4.7.1 General Considerations

Consider an experimental arrangement in which light propagating through the solution, reflects at normal incidence from a flat electrode, placed parallel to and at a distance d from a planar transparent cell window. It will be assumed, for simplicity, that the redox process involves consumption of a nonabsorbing species R, and generation of a single absorbing product P, where both R and P are solution phase species, and that only R is present at the beginning of the experiments. If the attenuation in light intensity induced by other components in the path of the beam, including reflection from the electrode, can be ignored, the intensity I of the light exiting the cell is given by (Equation 105),

$$I = I_o \exp\left(-2\varepsilon k \int_0^d c(y,t)\,dy\right) \tag{105}$$

where I_o is the light intensity entering the cell. Attention is focused hereafter on solutions of one-dimensional mass transport equations of the form

$$c(y,t) = \bar{c}(y,t) + \tilde{c}(y,t) \tag{106}$$

where in Equation (106) $\bar{c}(y,t)$ is a slowly varying function of time, or steady state, and $\tilde{c}(y,t)$ is the oscillatory contribution. Introducing Equation (106) into Equation (105), a normalized absorbance may be defined as follows (Equation 107):

$$\frac{I}{\bar{I}} = \exp\left(-2\varepsilon k \int_0^d \tilde{c}(y,t)\,dy\right) \tag{107}$$

Provided that the changes in concentration induced by the periodic perturbation are small, the argument in the exponential can be approximated by the first two terms in the Taylor series expansion, yielding Equation (108) after rearrangement:

$$\frac{\bar{I} - I}{\bar{I}} = \frac{I_{ac}}{I_{dc}} = 2\varepsilon k \int_0^d \tilde{c}(y,t)\,dy \tag{108}$$

If it is further assumed that the changes in concentration are confined to a volume of solution very close to the electrode surface, d may be regarded as arbitrarily large, and therefore the upper limit of integration may be set as infinity, leading to much simpler mathematical expressions.

From an experimental viewpoint, the ratio I_{ac}/I_{dc}, can be measured very accurately, such as by using a lock-in amplifier. A quantitative analysis of the results obtained using this modulation technique requires the integral on the right-hand side of Equation (108) to be evaluated, a problem that is equivalent to finding solutions for the mass transport equations subject to the appropriate boundary conditions.

Of particular interest here is derivation of expressions for the oscillatory component of the response of the form

$$\tilde{c}(y,t) = A^* \Psi(y) \exp(j\omega t) \tag{109}$$

where A^* in Equation (109) is the maximum amplitude of the sinusoidally varying concentration, and $\Psi(y)$ is a function of distance.

4.7.2 Quiescent Solutions

Substitution of Equation (109) into Fick's second law leads to a differential equation in y in total derivatives, which can be easily solved analytically to yield Equation (110):

$$\Psi(y) = \exp\left(-\left(\frac{j\omega}{D}\right)^{1/2} y\right) \tag{110}$$

Hence, I_{ac}/I_{dc}, also denoted as $I(\omega t)$, may be written as Equation (111):

$$\begin{aligned}\frac{I_{ac}}{I_{dc}} = I(\omega t) &= 2\varepsilon k \int_0^\infty A^* \exp\left(-\left(\frac{j\omega}{D}\right)^{1/2} y\right) \exp(j\omega t)\,dy \\ &= 2\varepsilon k A^* \left(\frac{D}{j\omega}\right)^{1/2} \exp(j\omega t) \\ &= 2\varepsilon k A^* \left(\frac{D}{\omega}\right)^{1/2} \exp\left(j\left(\omega t - \frac{\pi}{4}\right)\right)\end{aligned} \tag{111}$$

that is, the optical response lags the modulated potential by $\pi/4$.

In the case of very fast heterogeneous electron transfer reactions, the applied potential prescribes the ratio of concentrations of reactants and products through the Nernst equation. For perturbations of very small amplitude, this equation can be linearized, to yield Equation (112) that relates the amplitudes of the applied potential E^* and those of the concentrations of R and P, A_R^* and A_P^*, respectively:

$$E = \overline{E} + E = E_{dc} + E^* \exp(j\omega t) \tag{112}$$

where E_{dc} and E^* are given by Equation (113):

$$E_{dc} = E^{o\prime} + \frac{RT}{nF} \ln \frac{\bar{c}_R(0,t)}{\bar{c}_P(0,t)};$$

$$\text{and} \quad E^* = \frac{RT}{nF} \left(\frac{A_R^*}{\bar{c}_R(0,t)} - \frac{A_P^*}{\bar{c}_P(0,t)} \right. \tag{113}$$

Because of conservation of mass at the electrode surface (see Equation 16) $A_P^* = -A_R^* \xi$, where $\xi = (D_R/D_P)^{1/2}$; hence, it becomes possible to express either one of the A_i^* coefficients in terms of E^*, to yield Equation (114)

$$E^* = \frac{4RTA_R^*}{nFc_R^o} \cosh^2 \quad \frac{a}{2} \tag{114}$$

where a is given by Equation (115):

$$a = \frac{nF}{RT}(E_{dc} - E_{1/2}); \text{ and } E_{1/2} = E^{o\prime} + \frac{RT}{nF} \ln \left(\frac{D_P}{D_R} \right)^{1/2} \tag{115}$$

The current associated with the faradaic process is related to the flux of material at the electrode surface via Equation (116),

$$i = -nFAD_R \left. \frac{\partial c_R}{\partial y} \right|_{y=0} = -nFAA_P^*(j\omega D_P)^{1/2} \exp(j\omega t)$$

$$= -nFAA_P^*(\omega D_P)^{1/2} \exp \quad j \quad \omega t + \frac{\pi}{4} \tag{116}$$

where A is the area of the electrode. Hence, from Equations (111) and (116), the current and the optical response are 90° out of phase with respect to one another.

4.7.2.1 Alternating Current Solution-phase Reflectance Spectroscopy Based on Equations (105) and (109), the ratio of the magnitudes of the current and reflectance signal induced by the modulated applied potential is given by Equation (117),

$$\frac{|i(\omega t)|}{|I(\omega t)|} = \frac{nFA\omega}{2\varepsilon k} \tag{117}$$

and, therefore, for a fixed perturbation frequency, the electrical and optical responses are proportional to each other. In the case of solution-phase redox couples involving at least one UV/VIS chromophore, these considerations afford a basis for implementing a spectroscopic analog of a highly sensitive electrochemical technique known as alternating current (AC) voltammetry. This method relies on the superposition of an AC signal of small amplitude onto a slow linear potential scan, while monitoring the amplitude of the current response. Although restricted to optically absorbing species, this spectroscopic approach offers a definite advantage over current measurements in that the response is not affected by contributions to the electrical signal derived from other sources, particularly double layer capacity effects.

By way of illustration,[86] Figure 60 displays a plot of $I(\omega t)$ versus E_{dc} for the oxidation of an unbuffered solution 10 mM $K_4Fe(CN)_6$ in aqueous 1 M KCl obtained at a frequency of 43.9 Hz at a scan rate of 2 mV s^{-1}, for which both for the forward and reverse scans, as indicated by the arrows, were found to superimpose. The position of the peak at 0.219 V versus SCE agrees very well with the reversible half-wave potential for the redox couple. Moreover, the width of the peak at half-height was 94 mV, in harmony with that predicted by theory, i.e. 90 mV. This overall behavior is characteristic of the AC response for a reversible one-electron system.

The direct relationship between the optical and electrical response can also be valuable for the determination of rate constants by examining the dependence of the phase angle ϕ on the perturbation frequency. In particular, in the absence of effects due to IR drop in the solution, ϕ may be shown to be related to the rate of a simple heterogeneous

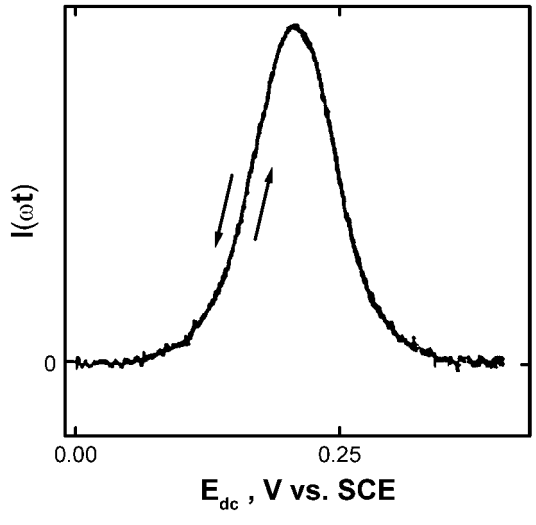

Figure 60 AC reflectance spectra for the oxidation of 10 mM $K_4Fe(CN)_6$ in 1 M KCl, at $\lambda = 420$ nm, $\omega/2\pi = 43.9$ Hz, $\nu = 2$ mV s^{-1}.[86]

electron transfer reaction k_s (Equation 118):

$$\cot\phi = 1 + \frac{(\omega D/2)^{1/2}}{k_s} \tag{118}$$

In fact, a plot of the *iR*-corrected $\cot\phi$ versus $\omega^{1/2}$ (Figure 61) for the ferro–ferricyanide couple obtained from potential modulated spectroelectrochemical measurements was found to be linear, yielding values for k_s around $0.096 \pm 0.008\,\mathrm{cm\,s^{-1}}$ in agreement with data obtained by more conventional means.[86]

4.7.2.2 Diffractive Alternating Current Modulation Light striking the edge of a solid object will be scattered producing a diffraction pattern on a plane normal to the beam direction, including regions in the shadow of the object, with much of the intensity of the diffracted beam originating from areas very close to the edge. Furthermore, the diffracted light will be attenuated by a UV/VIS absorber present very near the edge, and the extent of attenuation will be proportional to the concentration of the absorber therein, as predicted by Beer's law. Assuming the concentration of product in the area sampled by the diffracted beam, i.e. at and near the electrode edge C_P^s is uniform, the attenuation of the beam may be written as Equation (119):

$$A = \log\left(\frac{I_o}{I_t}\right. \quad = \varepsilon w C_P^s \tag{119}$$

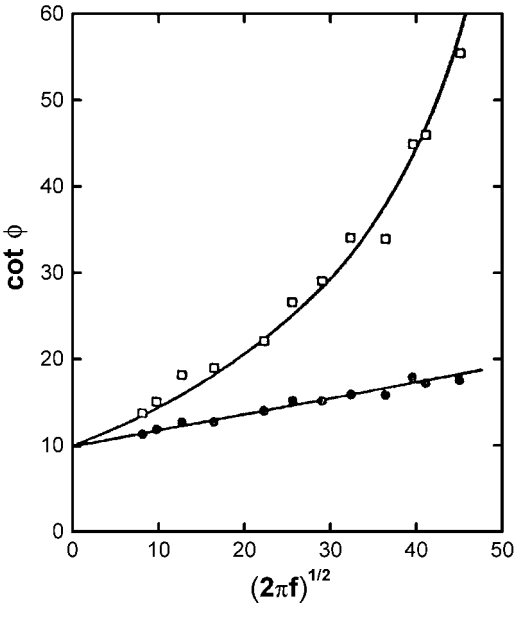

Figure 61 Plots of $\cot\phi$ versus $\omega^{1/2}$ for the oxidation of 10 mM $K_4Fe(CN)_6$ in 1 M KCl. Squares represent $\cot\varphi_S$, circles represent calculated $\cot\phi$; $E_{dc} = 0.219$ V versus SCE, $\Delta E = 18$ mV.[86]

where w is the electrode length along the optical axis. Furthermore, I_o and I_t are the intensities of the diffracted beam at a specific angle of observation measured before P is generated, and after C_P^s reaches a limiting value, respectively. This situation is somewhat equivalent to that of a conventional spectroelectrochemical cell of a pathlength equal to the length of the electrode along the direction of beam propagation, which can be electrochemically filled or emptied with absorber over a very short period of time.

If A is small, the normalized difference in light intensities $I_o - I_t$, denoted as $\Delta I_{diff}/I_o$, may be shown to be given by Equation (120):

$$\frac{\Delta I_{diff}}{I_o} = k\varepsilon b\xi C_P^o \tag{120}$$

where $C_R^o\xi = C_P$, C_R^o is the bulk concentration of R, and $\xi = (D_R/D_P)^{1/2}$. On this basis, it becomes possible to monitor the modulated diffracted light intensity, and, thus, the concentration of electrogenerated product P in a region of the solution very close to electrode surface. More specifically, the amplitude of the modulated diffracted light intensity I_{ac}^{diff} for AC voltammetry is related to the amplitude of the applied voltage by Equation (121):

$$\frac{I_{ac}^{diff}}{I_o} = \frac{k\varepsilon bn\xi F C_R^o E^*}{4RT\cosh^2(a/2)} \tag{121}$$

As before, this latter expression is valid provided the heterogeneous electron transfer reaction is infinitely fast.

Experiments involving TAA in TEAP/acetonitrile and square wave modulation of relatively large amplitude, i.e. 0.5 V about $E^{o\prime}$, yielded values of I_{ac}^{diff} lower than those predicted theoretically.[29,87] However, the output of the lock-in amplifier using a modulation frequency of 10 Hz was found to be proportional to the concentration of TAA in the range from 3×10^{-7} to 1×10^{-4} M. Furthermore, judicious choice of experimental conditions, including higher modulation frequencies, lower angles and phase adjustments, made it possible to increase the detection sensitivity to less than 2×10^{-8} M. Also in qualitative agreement with the expected response, were the results of spectroelectrochemical AC voltammetry experiments, which yielded superimposable curves for forward and backward scans for slow scan rates, about $2\,\mathrm{mV\,s^{-1}}$, and E^* value in the range 10–50 mV (Figure 62).

4.7.2.3 Transmission Alternating Current Voltammetry Application of a sinusoidally varying potential to a planar electrode induces a modulation in the concentration of an optically absorbing redox product. Changes along an axis parallel to the electrode surface can be monitored spectroscopically using a CCD or a photomultiplier in

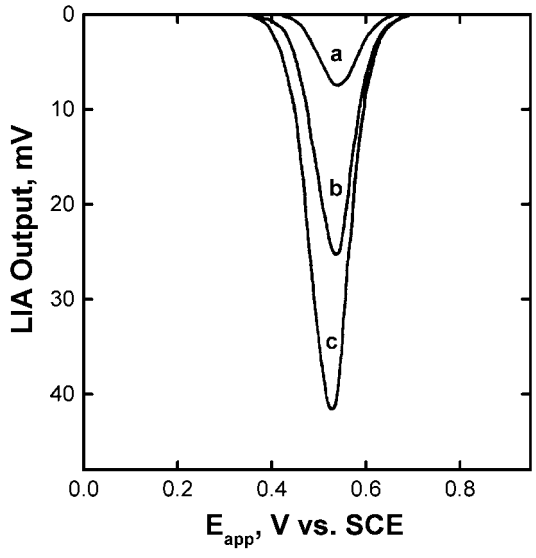

Figure 62 Lock-in amplifier output versus direct current (DC) potential for small-amplitude square-wave AC potential superimposed on a slow DC scan ($2\,\mathrm{mV\,s^{-1}}$): modulation frequency 10 Hz; phase shift $-20°$; time constant 0.3 s; $C^{o}_{TAA} = 0.257$ mM; curve a, $\Delta E = 10$ mV; b, $\Delta E = 30$ mV; c, $\Delta E = 50$ mV. Reverse scans were superimposed on forward scans within the pen width. Peaks occur at +0.52 V versus SCE.[87]

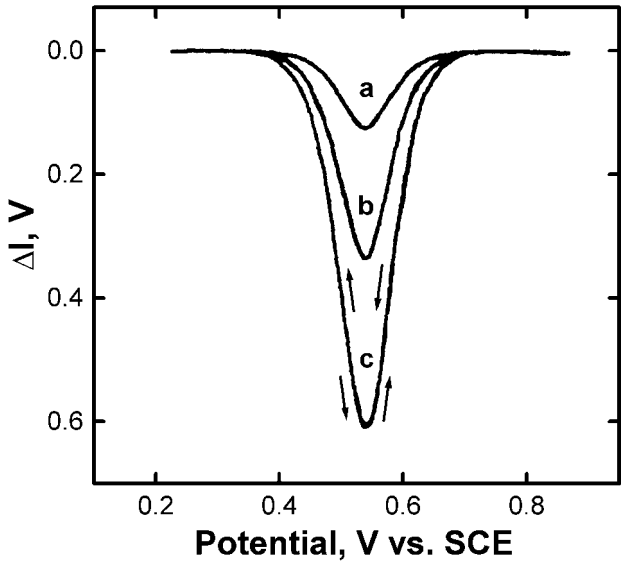

Figure 63 Optically monitored AC voltammograms for TAA solutions. Square-wave perturbations of varying height were superimposed on a slow DC ramp ($2\,\mathrm{mV\,s^{-1}}$). Frequency = 5.0 Hz, phase = $-20°$. Curve a, $\Delta E = 10$ mV; b, $\Delta E = 30$ mV; c, $\Delta E = 60$ mV. $y = 3\,\mu$m, $C^{o}_{TAA} = 0.38$ mM.[87]

conjunction with a narrow slit to illuminate a small volume of solution. The normalized optically modulated signal may thus be expressed as Equation (122):

$$\frac{I_{ac}}{I_{dc}} = \varepsilon k w \tilde{c}(y,t) = \varepsilon k w A^* \exp\left[-\left(\frac{j\omega}{D}\right)^{1/2} y\right] \exp(j\omega t)$$

$$= \varepsilon k w A^* \exp\left[-\left(\frac{\omega}{2D}\right)^{1/2} y\right] \times \exp\left\{j\left[\omega t + \left(\frac{\omega}{2D}\right)^{1/2} y\right]\right\} \qquad (122)$$

where w is the length of the electrode along the direction of light propagation, and y is the distance normal to the electrode surface. Although results have been reported for a CCD, the most interesting data have been collected with the slit arrangement.[87] In particular, Figure 63 shows results for the optically monitored AC spectroelectrochemical voltammogram of TAA obtained for three different scan rates at a distance of 3 μm from the surface, which are similar to those obtained using diffraction as described in the previous section. The expression above indicates that the phase is a function not only of the frequency of the perturbation, but also of the specific distance from the electrode surface.

4.7.2.4 Alternating Current Modulated Solution-phase Refraction According to Fermat's principle, a beam travelling along an axis z, in a medium in which the index of refraction n varies normal to the direction of propagation, say y, will bend toward regions having larger n, and that the extent of bending will be proportional to $\partial c/\partial y$. Assume, more specifically, that z and y are parallel and perpendicular to the electrode surface. A modulation in the concentration profile of solution phase species induced by the applied potential will modify the index of refraction along y, and consequently the extent of beam bending. Although refraction is insensitive to the nature of the species that causes the change in the index of refraction, it is useful from a didactic viewpoint to regard these effects as caused by a single species, and that the changes in n are linear in the concentration of that species, i.e. $\partial n/\partial c = N$. This situation can be realized experimentally for systems in which a single species is consumed or generated at the electrode surface. On this basis, and neglecting the effects due to the much slower nonoscillatory transient contribution, the oscillatory degree of bending Ψ will be given by Equation (123),

$$\Psi = \frac{w}{n} N \frac{\partial \tilde{c}}{\partial y} \qquad (123)$$

where w is the length of the electrode along z. An explicit expression for the derivative can be obtained from Equations (109) and (110), yielding

$$\frac{\partial \tilde{c}}{\partial y} = A^* \left[-\left(\frac{j\omega}{D}\right)^{1/2}\right] \exp\left[-\left(\frac{j\omega}{D}\right)^{1/2} y\right] \exp(j\omega t)$$

$$= -A^* \left(\frac{\omega}{D}\right)^{1/2} \exp\left(-\left(\frac{\omega}{2D}\right)^{1/2} y\right)$$

$$\times \exp\left(j\left\{\omega t + y\left(\frac{\omega}{2D}\right)^{1/2} + \frac{\pi}{4}\right\}\right) \quad (124)$$

Based on Equations (109) and (124) above, this analysis predicts that $\tilde{c}$ and $\partial\tilde{c}/\partial y$ should be 45° out of phase with respect to one another. In fact, large differences in the magnitude of the lock-in amplifier output as a function of phase angle have been reported[87] for potential modulation experiments in solutions containing 0.66 mM TAA and 10 mM BQ, which generate, respectively, an absorbing, and a nonabsorbing product upon oxidation and reduction (Figure 64). This phenomenon makes it possible to enhance the response of the absorbing versus the nonabsorbing product, by simply adjusting the phase angle. This is illustrated in Figure 64(b–d) in the form of amplitude of the measured light intensity at the image plane ΔI_{image}, as a function of applied DC potential (E_{dc}) for three different phase angles. Figure 64(a) is the magnitude of the AC current as a function of the applied DC potential, which reflect the much higher concentration of BQ compared to TAA.

4.7.3 Warburg Impedance

The dynamic properties of electronic circuits are customarily defined in terms of a frequency-dependent function known as the impedance, Z (Equation 125):

$$Z = \frac{\mathrm{d}E}{\mathrm{d}i} \quad (125)$$

where E is the potential and i the current. It is then possible to define a diffusional impedance, also known as the Warburg impedance, Z_{w} as the ratio of the time derivatives of the voltage and the current.

4.7.3.1 Quiescent Solutions The Warburg impedance for quiescent solutions can be obtained from Equations (112) and (116), namely Equation (126),

$$Z_{\text{w}} = \frac{\sigma\sqrt{2}}{(j\omega)^{1/2}} \quad (126)$$

where σ is defined by Equation (127):

$$\sigma = \frac{RT}{n^2F^2A\sqrt{2}}\left(\frac{1}{D_{\text{R}}^{1/2}\bar{c}_{\text{R}}(0,t)} + \frac{1}{D_{\text{P}}^{1/2}\bar{c}_{\text{P}}(0,t)}\right) \quad (127)$$

For a reversible reaction, Z_{w} reduces to Equation (128):

$$Z_{\text{w}} = \frac{4RT}{n^2F^2A(j\omega D_{\text{o}})^{1/2}c_{\text{o}}^*}\cosh^2\left(\frac{a}{2}\right) \quad (128)$$

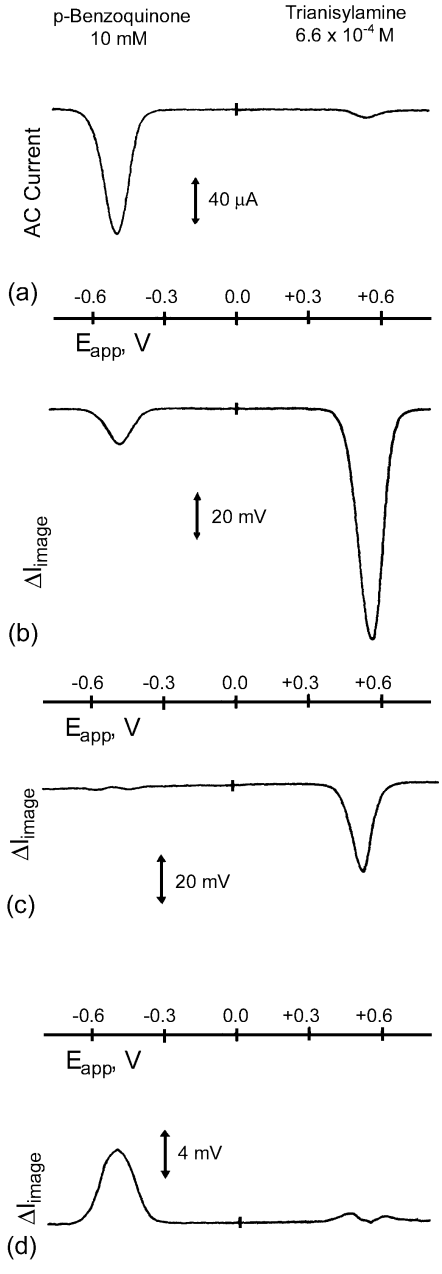

Figure 64 The AC current and AC voltammetric response for a mixture of TAA (0.66 mM) and BQ (0.01 M) for several phase angles. $\Delta E = 70$ mV, scan rate $= 2$ mV s^{-1}, $y = 3$ μm: (a) AC current on a 3.2 mm diameter Pt electrode (phase angle = 0°); (b) optical monitoring at a phase of −20°; (c) at −100°; (d) at −130°.[87]

Based on the results presented in the previous section, the Warburg impedance is in phase with the optical response (see Equation 111).

4.7.3.2 Forced Convection Expressions for the Warburg impedance of an RDE can be obtained[88] by taking the Laplace transform of the concentration step problem

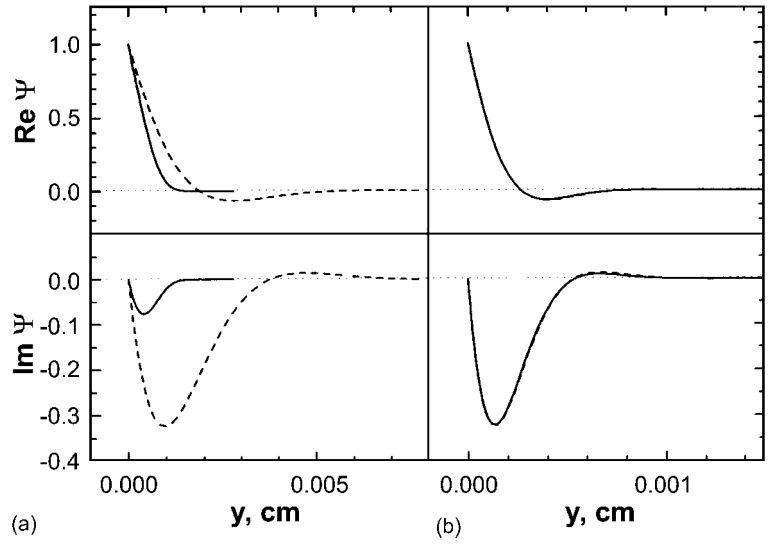

Figure 65 Plots of Re ψ and Im ψ for (a) $K = 1.2$ and (b) $K = 60$ as functions of the real axial distance for $\Omega = 0$ (dashed line) and $\Omega = 377\,\text{rad}\,\text{s}^{-1}$ (3600 rpm; solid line).[88]

as specified in section 2.2.2, and regarding the complex parameter s as $j\omega$.[89] Plots of the real and imaginary parts of the integral of the concentration profiles as a function of the dimensionless frequency K are shown in Figure 65. These data can also be represented by plotting the real versus the imaginary part of the function in question and specifying the values of the frequency directly on the curve (Figure 66), also known as Cole–Cole or Nyquist diagrams. It is significant to note that for $K > 40$, the angle reaches $-45°$, which implies that at high enough frequencies, the response of the electrochemical system in the presence of forced convection approaches that in a quiescent media.

4.7.4 Determination of Molar Absorptivities of Electrogenerated Species

Provided that $D_R = D_P$, rearrangement of Equation (102) yields Equation (129) for $\varepsilon(k)$ in terms of the integral of the sinusoidal profile:

$$\varepsilon(\lambda) = \frac{I_{ac}}{I_{dc}} \; 2k \int_0^{\infty} \tilde{c}(y, t)\,\mathrm{d}y \Bigg)^{-1} \qquad (129)$$

Setting $I_{ac} = I^* \exp[j(K\theta + \phi)]$, where ϕ is the phase, this equation can be rearranged to Equation (130),

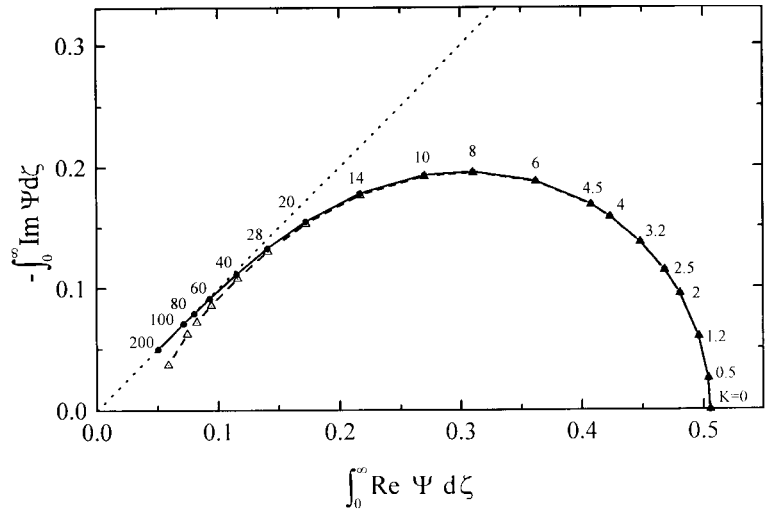

Figure 66 Plots of the real and imaginary components of the integral of the dimensionless concentration profile in the Cole–Cole representation: • = exact and Δ = long-time solutions.[88]

$$\varepsilon(\lambda) = \frac{I^*}{I_{dc}} \left(2A^* k\gamma \int_0^\infty \Psi \, d\varsigma \right)^{-1} \quad (130)$$

where A^* is the amplitude of the time-varying concentration, and γ the factor that converts y into the dimensionless variable ζ. As $\varepsilon(\lambda)$ is real, the phase may be expressed as Equation (131):[90]

$$\phi = \arctan \frac{\int_0^\infty \mathrm{Im}\Psi \, d\varsigma}{\int_0^\infty \mathrm{Re}\Psi \, d\varsigma} \quad (131)$$

Plots of $\int_0^\infty \Psi \, d\varsigma$ versus K, and ϕ versus K for an RDE are shown Figure 67.

The most striking verification of the validity of this theory, is provided by the excellent agreement between $\varepsilon(\lambda)$ obtained for ferricyanide from potential modulation spectroelectrochemical measurements (Figure 68) at an RDE in $K_4Fe(CN)_6$/0.5 M K_2SO_4 aqueous solutions (see scattered plots), and from conventional transmission experiments given by the continuous curve in the same figure.[91]

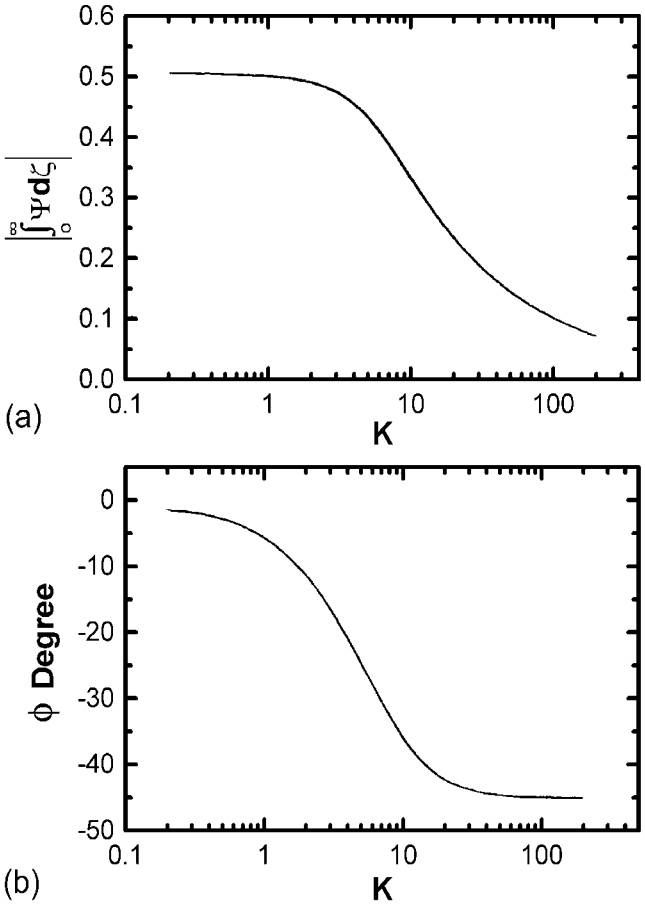

Figure 67 (a) Plot of $\int_0^\infty \Psi \, d\zeta$ as a function of K, based on the exact solutions. (b) Plot of the phase ϕ, as a function of the dimensionless frequency K.[91]

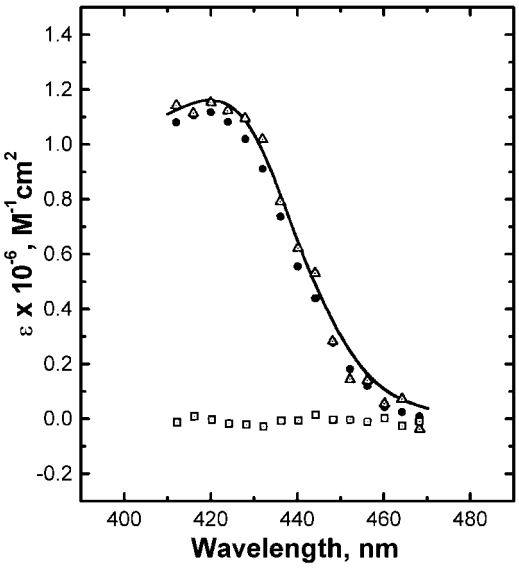

Figure 68 Comparison between $\varepsilon(\lambda)$ versus λ data obtained by conventional transmission spectroscopy (solid line) and that derived from the magnitude of I^*/I_{dc} determined from potential modulation experiments and the value of $\int_0^\infty \Psi \, d\zeta$ determined by numerical means for $c_o = 10$ mM (triangles) and $c_o = 3$ mM (solid circles). The open squares were obtained in measurements in which the potential was modulated about 0.0 V.[88]

CONCLUDING REMARKS

The quantitative analysis of UV/VIS spectroelectrochemical techniques relies on well-known optical and mass transport principles. Although analytic solutions for transient problems appear to be available only for the simplest of systems, digital simulations involving an inexpensive desk computer provide a means of diagnosing mechanisms and extracting kinetic parameters of interest from experimental data. Furthermore, the advent of computer-controlled instrumentation including data acquisition makes it possible to collect large amounts of information so as to achieve reliable statistics. Intrinsic theoretical and practical limitations imposed by the very nature of these measurements, sets limits on the space-, and time-resolution of absorption-based UV/VIS spectroelectrochemical techniques. Without diminishing its importance and future impact in scientific and technological fields, it seems reasonable to conclude that this specialized subfield of analytical chemistry has reached a solid level of maturity. On this basis, new developments may not be expected in instrumental principles, but in the applications area, especially sensor miniaturization.

It may be envisaged that new generations of inexpensive, fully tunable lasers and widespread implementation of chemometrics, will lead to versatile and expedient means for quantitative detection of weak chromophores and thus join the array of existing analytical techniques. Also to be expected over the next few years, are applications of UV/VIS imaging analysis to the study of current distribution in electrochemical cells, which are expected to have an impact on the design of electrochemical reactors for electrosynthesis. It may be interesting to note that in contrast to the absorption-based techniques featured in this article, the analysis of in situ UV/VIS ellipsometric and electroreflectance measurements, continues to pose formidable problems in terms of a quantitative theoretical interpretation. It is hoped that the newly discovered links between electrochemistry and surface science will prompt scientists in both fields to find novel approaches for gaining new insights into the fundamental basis of this phenomenon and thus fully exploit its exquisite surface specificity and sensitivity, which greatly surpasses our theoretical understanding.

ACKNOWLEDGMENTS

Support for this work was provided by the National Science Foundation, and the National Institutes of Health. The authors are deeply indebted to Prof. Richard L. McCreery from Ohio State University for his insightful comments during the preparation of this manuscript.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
AN	Acrylonitrile
ATR	Attenuated Total Reflection
BQ	Benzoquinone
BQI	*p*-Benzoquinoneimine
CCD	Charge-coupled Device
CPZ	Chlorpromazine
$CPZ^{+\bullet}$	Chlorpromazine Cation Radical
DC	Direct Current
DCIP	2,6-Dichlorophenolindophenol
DOQ	Dopamine Quinone
EC	Electrochemical–Chemical
EEC	Electrochemical–Electrochemical–Chemical
FeTPPCl	Iron Tetraphenylporphyrin Chloride
H_2Q	Hydroquinone
IRE	Internal Reflection Element
IRS	Internal Reflection Spectroscopy
LOPTLC	Long Optical Pathlength Thin-layer Cell
LPTLC	Long Pathlength Thin Layer Cell
MB	Methylene Blue
MPZ	Methoxypromazine
MV^{++}	Methyl Viologen Dication
OD	*o*-Dianisidine
OTTLE	Optically Transparent Thin-layer Electrode
PAP	*p*-Aminophenol
PMT	Photomultiplier Tube
Pyc=cPy	1,2-Bis-(4-pyridine)ethylene
RDE	Rotating Disk Electrode
RRDE	Rotating Ring-disk Electrode
SCE	Saturated Calomel Electrode
$TAA^{+\bullet}$	Tri-*p*-anisylamine Cation Radical
TBAP	Tetra-*n*-butyl Ammonium Perchlorate
TEACN	Tetraethylammonium Cyanide
TEAP	Tetraethylammonium Perchlorate
TPZ	Triflupromazine Hydrochloride
UV/VIS	Ultraviolet/Visible

RELATED ARTICLE

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

REFERENCES

1. F. Wooten, *Optical Properties of Solids*, Academic Press, New York, 1972.
2. A.V. Sokolov, *Optical Properties of Metals*, American Elsevier Publishing Company, Inc., New York, 1967.
3. J.I. Steinfeld, *Molecules and Radiation*, The MIT Press, Cambridge, MA, 1985.
4. T. Kuwana, N. Winograd, 'Spectroelectrochemistry at Optically Transparent Electrodes. 1. Electrodes Under Semi-infinite Diffusion Conditions', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–78, Vol. 7, 1974.
5. W.R. Heineman, F.M. Hawkridge, H.N. Blount, 'Spectroelectrochemistry at Optically Transparent Electrodes. II. Electrodes Under Thin-layer and Semi-infinite Diffusion Conditions and Indirect Coulometric Titrations', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 1–113, Vol. 13, 1984.
6. R.L. McCreery, 'Spectroelectrochemistry', in *Physical Methods of Chemistry*, eds. B.W. Rossiter, S.F. Hamilton, John Wiley & Sons, New York, 591–661, Vol. II, 1986.
7. J. Niu, S. Dong, 'Transmission Spectroelectrochemistry', *Reviews in Analytical Chemistry*, **XV1**, 1 (1996).
8. W.R. Heineman, 'Spectroelectrochemistry', *Anal. Chem.*, **50**, 390A (1978).

9. R.H. Muller, 'Principles of Ellipsometry', in *Advances in Electrochemistry and Electrochemical Engineering*, ed. R.H. Muller, J. Wiley & Sons, New York, 167, Vol. 9, 1973.
10. S. Gottesfeld, 'Ellipsometry', in *Physical Electrochemistry: Principles, Methods and Applications*, ed. I. Rubinstein, Marcel Dekker, New York, 1995.
11. J.D.E. McIntyre, 'Specular Reflection Spectroscopy of the Electrode Solution Interphase', in *Advances in Electrochemistry and Electrochemical Engineering*, ed. R.H. Muller, J. Wiley & Sons, New York, 61, Vol. 9, 1973.
12. D.M. Kolb, 'UV Visible Reflectance Spectroscopy', in *Spectroelectrochemistry Theory and Practice*, ed. R.J. Gale, Plenum Press, New York, 87, 1988.
13. C. Barbero, M.C. Miras, R. Kotz, 'Electrochemical Mass Transport Studied by Probe Beam Deflection: Potential Step Experiments', *Electrochim. Acta*, **37**, 429 (1992).
14. R.H. Muller, 'Double Beam Inteferometry for Electrochemical Studies', in *Advances in Electrochemistry and Electrochemical Engineering*, ed. R.H. Muller, J. Wiley & Sons, New York, 281, Vol. 9, 1973.
15. A.J. Bard, L.R. Faulkner, *Electrochemical Methods, Fundamentals and Applications*, Wiley, New York, 1980.
16. M. Born, E. Wolf, *Principles of Optics*, 6th edition, Pergamon Press, Oxford, 1987.
17. M.V. Klein, T.E. Furtak, *Optics*, 2nd edition, Wiley, New York, 1986.
18. J.D. Brewster, J.L. Anderson, 'On Absorbance Measurements in Spatially Inhomogeneous Fields', *Appl. Spectrosc.*, **43**, 710–714 (1989).
19. N.J. Harrick, *Internal Reflection Spectroscopy*, Interscience Publishers, New York, 1967.
20. H.S. Carslaw, J.C. Jaeger, *Conduction of Heat in Solids*, 2nd edition, Clarendon Press, Oxford, 1967.
21. R.V. Churchill, R. Vance, *Operational Mathematics*, 2nd edition, McGraw-Hill, New York, 1958.
22. D.D. Macdonald, *Transient Techniques in Electrochemistry*, Plenum Press, New York, 1977.
23. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, U.S. Govt. Print. Off., Washington, 1964.
24. R.B. Bird, W.E. Stewart, E.N. Lightfoot, *Transport Phenomena*, Wiley, New York, 1960.
25. N. Winograd, H. Blount, T. Kuwana, 'Spectroelectrochemical Measurement of Chemical Reaction Rates First-order Catalytic Processes', *J. Phys. Chem.*, **73**, 3456 (1969).
26. V.G. Levich, *Physicochemical Hydrodynamics*, Prentice-Hall Inc., Englewood Cliffs, 1962.
27. Y.V. Tolmachev, Z. Wang, D.A. Scherson, 'In Situ Spectroscopy in the Presence of Convective Flow Under Steady-state Conditions: A Unified Mathematical Formalism', *J. Electrochem. Soc.*, **143**, 3539 (1996).
28. A. Erdelyi, *Table of Integral Transforms*, McGraw-Hill, New York, Vol. 1, 1954.
29. C.-C. Jan, B.K. Lavine, R.L. McCreery, 'High-sensitivity Spectroelectrochemistry Based on Electrochemical Modulation of an Absorbing Analyte', *Anal. Chem.*, **57**, 752–758 (1985).
30. J.E. Davis, N. Winograd, 'Application of Coulostatic Charge Injection. Techniques to Improve Potentiostat Risetimes', *Anal. Chem.*, **44**, 2152 (1972).
31. R.S. Robinson, R.L. McCreery, 'Submicrosecond Spectroelectrochemistry Applied to Chlorpromazine Cation Radical Charge Transfer Reactions', *J. Electroanal. Chem.*, **182**, 61–72 (1985).
32. C.H. Pyun, S.M. Park, 'Construction of a Microcomputer Controlled Near Normal Incidence Reflectance Spectroelectrochemical System and Its Performance Evaluation', *Anal. Chem.*, **58**, 251 (1986).
33. C. Zhang, S.M. Park, 'Simple Technique for Constructing Thin-layer Electrochemical-cells', *Anal. Chem.*, **60**, 1639 (1988).
34. Y.R. Shen, 'Optical Wavelength-modulation Spectroscopy', *Surface Sci.*, **37**, 522–539 (1973).
35. S. Kim, D.A. Scherson, 'In-situ UV-visible Reflection Absorption Wavelength Modulation Spectroscopy of Species Irreversible Adsorbed on Electrode Surfaces', *Anal. Chem.*, **64**, 3091 (1992).
36. A. Deputy, H.-P. Wu, R.L. McCreery, 'Spatially Resolved Spectroelectrochemical Examination of the Oxidation of Dopamine by Chlorpromazine Cation Radical', *J. Phys. Chem.*, **94**, 3620–3624 (1990).
37. R.S. Robinson, C.W. McCurdy, R.L. McCreery, 'Microsecond Spectroelectrochemistry by External Reflection from Cylindrical Microelectrodes', *Anal. Chem.*, **54**, 2356–2361 (1982).
38. J.P. Skully, R.L. McCreery, 'Glancing Incidence External Reflection Spectroelectrochemistry with a Continuum Source', *Anal. Chem.*, **52**, 1885–1889 (1980).
39. C.E. Baumgartner, G.T. Marks, D.A. Alkens, H.H. Richtol, 'Spectroelectrochemical Monitoring of Electrode Reactions by Multiple Specular Reflection Spectroscopy', *Anal. Chem.*, **52**, 167–270 (1980).
40. C.-C. Jan, R.L. McCreery, 'High-resolution Spatially Resolved Visible Absorption Spectrometry of the Electrochemical Diffusion Layer', *Anal. Chem.*, **58**, 2771–2777 (1986).
41. A.L. Deputy, R.L. McCreery, 'Spatially Resolved Spectroelectrochemistry for Examining an Electrochemically Initiated Homogeneous Electron Transfer Reaction', *J. Electroanal. Chem.*, **257**, 57–70 (1988).
42. R.L. McCreery, R. Pruiksma, R. Fagan, 'Optical Monitoring of Electrogenerated Species Via Specular Reflection at Glancing Incidence', *Anal. Chem.*, **51**, 749 (1979).
43. J. Zak, M.D. Porter, T. Kuwana, 'Thin-layer Electrochemical Cell for Long Optical Path Length Observation of Solution Species', *Anal. Chem.*, **55**, 2219–2222 (1983).
44. J.D. Brewster, J.L. Anderson, 'Fiber Optic Thin-layer Spectroelectrochemistry with Long Optical Path', *Anal. Chem.*, **54**, 2560–2566 (1982).

45. J. Niu, S. Dong, 'An Integrated Calcium Fluoride Crystal Thin-layer Cell and its Application to Identification of Electrochemical Reduction Product of Bilirubin', *Electrochim. Acta*, **40**, 823–828 (1995).
46. K.A. Rubinson, E. Itabashi, H.B.J. Mark, 'Electrochemical Oxidation and Reduction of Methylcobalamin and Coenzyme B12', *J. Am. Chem. Soc.*, (1982).
47. T.M. Kenyhercz, T.P. DeAngelis, B.J. Norris, W.R. Heineman, J. Mark, B. Harry, 'Thin Layer Spectroelectrochemical Study of Vitamin B12 and Related Cobalamin Compounds in Aqueous Media', *J. Am. Chem. Soc.*, **98**, 2469–2477 (1975).
48. W.C. Anderson, B.H. Halsall, W.R. Heineman, 'A Small-volume Thin-layer Spectroelectrochemical Cell for the Study of Biological Components', *Anal. Biochem.*, **93**, 366–372 (1979).
49. J.L. Owens, G. Dryhurst, 'Electrochemical Oxidation of 5,6-Diaminouracil an Investigation by Thin-layer Spectroelectrochemistry', *J. Electroanal. Chem.*, **80**, 171 (1977).
50. E. Blubaugh, M.A. Yacynych, W.R. Heineman, 'Thin-layer Spectroelectrochemistry for Monitoring Kinetics of Electrogenerated Species', *Anal. Chem.*, **51**, 561 (1979).
51. T.B. Jarbawi, W.R. Heineman, 'Thin-layer Spectroelectrochemistry of a Regenerative Mechanism', *J. Electroanal. Chem.*, **132**, 323–328 (1982).
52. Y.P. Gui, T. Kuwana, 'Long Optical Path Length Thin-layer Spectroelectrochemistry. Quantitation and Potential Dependence of Electroinactive Species Adsorbed on Platinum', *J. Electroanal. Chem.*, **222**, 321–330 (1987).
53. M.P. Soriaga, A.T. Hubbard, 'Determination of the Orientation of Aromatic-molecules Adsorbed on Platinum-electrode – The Effect of Solute Concentration', *J. Am. Chem. Soc.*, **104**, 3937 (1982).
54. J.W. Strojek, T. Kuwana, 'Electrochemical-spectroscopy Using Tin Oxide-coated Optically Transparent Electrodes', *J. Electroanal. Chem.*, 741 (1968).
55. N. Winograd, T. Kuwana, 'Characteristics of the Electrode-solution Interface Under Faradaic and Non-faradaic Conditions as Observed by Internal Reflection Spectroscopy', *J. Electroanal. Chem.*, **23**, 333 (1969).
56. N. Winograd, T. Kuwana, 'High Sensitivity Internal Reflection Spectroelectrochemistry for Direct Monitoring of Diffusing Species Using Signal Averaging', *Anal. Chem.*, **43**, 252 (1971).
57. N. Winograd, T. Kuwana, 'Evaluation of Fast Homogeneous Electron-exchange Reaction Rates Using Electrochemistry and Reflection Spectroscopy', *J. Am. Chem. Soc.*, **92**, 224 (1970).
58. H.N. Blount, N. Winograd, T. Kuwana, 'Spectroelectrochemical Measurements of Second-order Catalytic Reaction Rates Using Signal Averaging', *J. Phys. Chem.*, **74**, 3231 (1970).
59. N. Winograd, T. Kuwana, 'Homogeneous Electron-transfer Reactions Studied by Internal Reflection Spectroelectrochemistry', *J. Am. Chem. Soc.*, **93**, 4343 (1971).
60. Y. Shi, A.F. Slaterbeck, C.J. Seliskar, W.R. Heineman, 'Spectroelectrochemical Sensing Based on Multimode Selectivity Simultaneously Achievable in a Single Device. 1. Demonstration of Concept with Ferricyanide', *Anal. Chem.*, **69**, 3679–3686 (1997).
61. Y. Shi, C.J. Seliskar, W.R. Heineman, 'Spectroelectrochemical Sensing Based on Multimode Selectivity Simultaneously Achievable in a Single Device. 2. Demonstration of Selectivity in the Presence of Direct Interferences', *Anal. Chem.*, **69**, 4819–4827 (1997).
62. A.F. Slaterbeck, T.H. Ridgway, C.J. Seliskar, W.R. Heineman, 'Spectroelectrochemical Sensing Based on Multimode Selectivity Simultaneously Achievable in a Single Device. 3. Effect of Signal Averaging on Limit of Detection', *Anal. Chem.*, **71**, 1196–1203 (1999).
63. L. Gao, C.J. Seliskar, W.R. Heineman, 'Spectroelectrochemical Sensing Based on Multimode Selectivity Simultaneously Achievable in a Single Device 4. Sensing with Poly(vinyl alcohol) – Polyelectrolyte Blend Modified Optically Transparent Electrodes', *Anal. Chem.*, **71**, 4061–4068 (1999).
64. D.H. Jones, S. Hinman, 'Determination of the Visible Spectra of Electrode Reaction Products in Strongly Absorbing Media by Diffusion-controlled Chronoabsorptometry', *Can. J. Chem.*, **74**, 1403 (1996).
65. E. Ahlberg, D.P. Parker, V.D. Parker, 'Sensitivity Enhancement During Electrochemical Specular Reflectance Measurements', *Acta Chem. Scand.*, **B33**, 760–762 (1979).
66. R. Kotz, S. Stucki, D. Scherson, D.M. Kolb, 'In-situ Identification of RuO_4 as the Corrosion Product During Oxygen Evolution on Ruthenium in Acid Media', *J. Electroanal. Chem.*, **172**, 211–219 (1984).
67. J.S. Mayausky, R.L. McCreery, 'Spectroelectrochemical Examination of Chloripromazine Cation Radical Reactions with Mono- and Bifunctional Nucleophiles', *J. Electroanal. Chem.*, **145**, 117–126 (1983).
68. R. Pruiksma, R.L. McCreery, 'Observation of Electrochemical Concentration Profiles by Absorption Spectroelectrochemistry', *Anal. Chem.*, **51**, 2253 (1979).
69. C.-C. Jan, R.L. McCreery, F.T. Gamble, 'Diffusion Layer Imaging: Spatial Resolution of Electrochemical Concentration Profiles', *Anal. Chem.*, **57**, 1763–1765 (1985).
70. H.-P. Wu, R.L. McCreery, 'Spatially Resolved Absorption Spectroelectrochemistry', *J. Electrochem. Soc.*, **136**, 1375 (1989).
71. H.-P. Wu, R.L. McCreery, 'Observation of Concentration Profiles at Cylindrical Microelectrodes by a Combination of Spatially Resolved Absorption Spectroscopy and the Abel Inversion', *Anal. Chem.*, **61**, 2347–2352 (1989).
72. A.L. Deputy, R.L. McCreery, 'Spatially Resolved Absorption Examination of the Redox Catalysis Mechanism: Equilibrium and Near Equilibrium Cases', *J. Electroanal Chem.*, **285**, 1–9 (1989).
73. C.M.A. Brett, M. Oliveira Brett, *Electrochemistry Principles Methods and Applications*, Oxford University Press, Oxford, 1993.

74. M. Zhao, D.A. Scherson, 'UV-Visible Reflection Absorption Spectroscopy in the Presence of Convective Flow', *Anal. Chem.*, **64**, 3964 (1992).
75. H. Debrodt, K.E. Heusler, 'Eine rotierende Scheibenelektrode mit optisch durchlassigem Ring zur spektrophotometrischen Untersuchung von Produkten elektrochemischer Reaktionen', *Ber. Bunsen-Ges. Phys. Chem.*, **81**, 1172 (1977).
76. R. Dorr, E.W. Grabner, 'Digital Simulation of a Rotating Disk Electrode with Optically Transparent Ring', *Ber. Bunsen-Ges. Phys. Chem.*, **82**, 164 (1978).
77. Y.V. Tolmachev, Z. Wang, D.A. Scherson, 'In-situ Spectroscopy in the Presence of Convective Flow Under Steady State Conditions: A Unified Mathematical Approach', *J. Electrochem. Soc.*, **143**, 3539 (1996).
78. R.L. Wang, K.Y. Tam, R.G. Compton, 'Applications of the Channel Flow Cell for UV Visible Spectroelectrochemical Studies. Part 3. Do Radical Cations and Anions have Similar Diffusion Coefficients to their Neutral Parent Molecules?', *J. Electroanal. Chem.*, **434**, 105 (1997).
79. Z. Wang, M. Zhao, D. Scherson, 'Channel Flow Cell for UV/Visible Spectroelectrochemistry', *Anal. Chem.*, **66**, 4560 (1994).
80. Z. Wang, D. Scherson, 'In Situ UV-visible Spectroscopic Imaging of Laminar Flow in a Channel-type Electrochemical Cell', *J. Electrochem. Soc.*, **142**, 4225 (1995).
81. S. Ma, Y. Wu, Z. Wang, 'Spectroelectrochemistry for a Coupled Chemical Reaction in the Channel Cell. Part I. Theoretical Simulation of an EC Reaction', *J. Electroanal. Chem.*, **464**, 176 (1999).
82. Z. Wang, X. Li, Y. Wu, Y. Tang, S. Ma, 'Spectroelectrochemistry for a Coupled Chemical Reaction in the Channel Cell. Part II. Kinetics of Hydrolysis and the Absorption Spectrum of *p*-Benzoquinoneimine', *J. Electroanal. Chem.*, **464**, 181 (1999).
83. V.S. Krylov, V.N. Babak, 'Nonsteady-state Diffusion to the Surface of a Rotating Disk', *Soviet Electrochem.*, **7**, 626 (1971).
84. K. Nicancioglu, J.S. Newman, 'Transient Convective Diffusion to a Disk Electrode', *J. Electroanal. Chem.*, **50**, 23 (1974).
85. M. Zhao, D.A. Scherson, 'Transient Convective Diffusion to a Rotating Disk Electrode as Monitored by Near-normal Incidence Reflection Absorption Ultraviolet-visible Spectroscopy', *J. Electrochem. Soc.*, **140**, 729 (1993).
86. A.S. Hinman, J.F. McAleer, S. Pons, 'Spectroscopic Determination of the AC Voltammetric Response', *J. Electroanal Chem.*, **154**, 45–56 (1983).
87. C.-C. Jan, R.L. McCreery, 'Spectroelectrochemical Determination of Trace Concentrations by Diffusion Layer Imaging', *J. Electroanal. Chem.*, **220**, 41 (1987).
88. M. Zhao, D.A. Scherson, 'Theoretical Aspects of Potential Modulation Normal Incidence Reflection Absorption UV-visible Spectroscopy Under Forced Convection', *J. Electrochem. Soc.*, **140**, 1671–1676 (1993).
89. D. Scherson, J. Newman, 'The Warburg Impedance in the Presence of Convective Flow', *J. Electrochem. Soc.*, **127**, 110 (1980).
90. M. Zhao, D.A. Scherson, 'Theoretical Aspects of Potential Modulation Reflection Absorption UV-visible Spectroscopy Under Forced Convection', *J. Electrochem. Soc.*, **140**, 1671 (1993).
91. M. Zhao, D.A. Scherson, 'Potential Modulation Normal Incidence Reflection Absorption UV-visible Spectroscopy Under Forced Convection', *J. Electrochem. Soc.*, **140**, 2877–2879 (1993).
92. Y.V. Tolmachev, Z. Wang, Y. Hu, I.T. Bae, D.A. Scherson, 'In-situ Spectroscopic Determination of Faradaic Efficiencies in Systems with Forced Convection Under Steady State: Electroreduction of Bisulfite to Dithionite on Gold in an Aqueous Electrolyte', *Anal. Chem.*, **70**, 1149–1155 (1998).

X-ray Methods for the Study of Electrode Interaction

Enrique Herrero
Universidad de Alicante, Alicante, Spain

1 Introduction **1**
1.1 X-ray Sources 2
1.2 Synchrotron Radiation 3
2 X-ray Absorption Techniques **4**
2.1 Interaction of X-rays with Matter 4
2.2 Extended X-ray Absorption Fine Structure and X-ray Absorption Near Edge Structure Theory 6
2.3 Data Analysis 7
2.4 Experimental Aspects 10
3 X-ray Scattering Techniques **11**
3.1 X-ray Diffraction 11
3.2 Crystal Truncation Rods and Bragg Rods 13
3.3 X-ray Standing Waves 14
3.4 Experimental Aspects 15
4 Application of X-ray Methods to Electrochemistry **17**
4.1 Surface Structure of Metal Electrodes 17
4.2 Adsorbed Layers on Metal Surfaces 18
4.3 Fuel Cell Electrodes 23
4.4 Battery and Oxide Electrodes 23
4.5 Corrosion 23
4.6 Other Examples 24
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **24**
References **24**

X-ray methods provide an excellent tool for the determination of the structure and composition of the electrode/solution interphase. The possibility of carrying out in situ experiments, because of the low absorbability of the X-rays by the solution, allows direct correlation to be made between the structural changes observed in the sample and the potential applied. Depending on the interaction of the X-rays with matter, these techniques can be classified in two main groups: X-ray absorption techniques and X-ray scattering techniques. Within the absorption techniques, extended X-ray absorption fine structure (EXAFS) and X-ray absorption near edge structure (XANES) will be considered. EXAFS allows determination of the short-range order of the sample, i.e. it provides information about the nature and distances of the closest neighbors to the absorbing atom. On the other hand, identification of the oxidation state of the absorbing atom can be done with XANES. Grazing incident X-ray diffraction (GIXD) and crystal truncation rod (CTR) measurements belong to the scattering techniques. The information obtained from these measurements can be considered complementary since GIXD allows precise determination of the long-range order structure of the interphase whereas CTR analysis gives information about the out-of-plane structure of the interphase. The last technique covered, X-ray standing waves (XSW), is not a pure scattering technique, but rather one that combines X-ray scattering and X-ray interference. With this last technique, the exact position of a foreign atom in the interphase can be determined.

1 INTRODUCTION

Understanding the electrode/electrolyte interphase is one of the major areas of study in electrochemistry.[1] The structure and properties of the interphase and the interactions both between the electrode and electrolyte, and within the different species in the electrolyte, affect and control the behavior of electrochemical systems. The complex nature of the electrochemical interphase requires the use of different techniques, since a single technique is able to render only a partial picture of the system. The first available techniques, pure electrochemical ones, provided macroscopic information about the reactions occurring at the interphase, and the thermodynamics and kinetics of these processes. It is clear that a complete picture of the interphase cannot be achieved with the information collected by these techniques, since no information about the interactions between the electrode and electrolyte and the interphase structure is obtained.

The existence of a condensed phase in contact with the electrode caused a major problem in using the techniques employed in other fields of surface chemistry, since it constituted a barrier that limited the penetration of the probes. The first approach to studying the surface structure of an electrode was the ex situ approach, using the techniques for surface characterization that the ultrahigh vacuum (UHV) environment provided. Although UHV techniques, such as low-energy electron diffraction (LEED) and Auger electron spectroscopy (AES), are powerful surface characterization techniques, the removal of the electrolyte and the loss of potential control can lead to changes in the structure and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

composition in the interphase. This drawback has limited the application of UHV techniques to the characterization of very stable layers, normally metal layers, on the electrode surface.

In recent years, the advent of in situ spectroscopic scanning probe microscopy (SPM) and X-ray techniques for electrochemical systems have started to fill in the gap in information about the interphase. The spectroscopic techniques, such as in situ Fourier transform infrared spectroscopy (FTIRS) and related techniques,[2,3] enhanced Raman spectroscopy,[4–7] and nonlinear optical spectroscopy,[8,9] give information about the chemical composition and the orientation of the different species on the interphase. The SPM techniques[10] (scanning tunneling microscopy (STM) and atomic force microscopy (AFM)) provide information about the surface structure on the atomic level, when ordered structures are found. However, interpretation of the images obtained with these techniques is not always straightforward, since the maxima appearing in the image cannot be unequivocally identified with a single species in the system.

Details about the surface structure can also be obtained by X-ray techniques. When compared with SPM techniques, X-ray techniques present several advantages:

- Distances and angles calculated from X-ray measurements have a higher accuracy than those calculated from SPM techniques. The typical error in the distance measurements with STM or AFM is ca. 10%, whereas errors below 2% can be routinely obtained with X-ray techniques. This leads to a better distinction between different possible structures and the detection of small shifts in the positions of the atoms with different experimental conditions.
- Penetration of the X-rays in matter allows collection of data about the surface structure of the electrode, which can also be affected by the experimental conditions. SPM techniques provide information about the outermost layer (only in extreme conditions can the underlying substrate be imaged).
- They can also be used to determine the out-of-plane structure (the structure of the interphase perpendicular to the electrode surface), the short-range order in nonordered interphases, and the chemical nature of the species in the interphase.

The main disadvantage of the X-ray techniques is the requirement of a source bright enough to be surface sensitive. Although some studies have been done using laboratory X-ray sources, their brightness does not allow data about the structure of the interphase to be obtained without very long acquisition times, and a more powerful source has to be used: a synchrotron.

X-ray techniques can be divided in two main families, depending on the interaction of the X-rays with matter: absorption techniques and scattering techniques. Within the absorption techniques,[11–14] we can consider the surface extended X-ray absorption fine structure (SEXAFS), which provides information about the short-range order structure of the substrate, and XANES, which allows identification of the oxidation state of the atoms. The main scattering techniques[11–13,15–19] are GIXD that allows a precise determination of the long-range order structure of the interphase and CTR analysis, a reflectivity technique that gives information about the out-of-plane structure of the interphase. The XSW technique is not a pure scattering technique, but rather one that combines X-ray scattering and X-ray interference. However, in the present article, it has been included as a scattering technique. With this technique, the atomic positioning of foreign atoms in the interphase can be determined.

1.1 X-ray Sources

X-rays were discovered by Röntgen in 1895 using an evacuated tube with two metal plates (anode and cathode). When a voltage is applied between anode and cathode, the residual gas in the tube is ionized and the cations generated collide with the cathode, which releases electrons under this bombardment. These electrons are accelerated by the electric field to the anode and collide with it, generating the X-rays in the sudden deceleration of the electrons. If all the energy of the impinging electrons is converted to X-ray photon energy, the wavelength of the resulting X-rays will be given by the Duane–Hunt law (Equation 1):

$$\lambda_{\min} = \frac{hc}{eV} = \frac{12\,398\ \text{Å V}^{-1}}{V} \qquad (1)$$

where h is Planck's constant, c is the velocity of light, e is the electron charge and V the applied voltage between anode and cathode. ($1\text{Å} = 10^{-10}$ m.) However, not all the kinetic energy of the electrons is transformed into X-ray photon energy. Part of the incident energy can be lost in multiple collisions and inelastic scattering, resulting in a continuous ''bremsstrahlung'' spectrum (white radiation). Therefore, the wavelength obtained according to Equation (1) ($\lambda_{\min}$) represents the minimum wavelength emitted by the tube.

When the accelerating voltage reaches a specific value (which depends on the target material), the electrons are able to knock out a core electron from the atoms of the target material, producing a core hole. If an electron from the outer shells descends to fill this hole, the emitted energy depends on the energy of the two levels involved. This results in a sharp spike in the emission spectrum (Figure 1). These characteristic lines are well defined

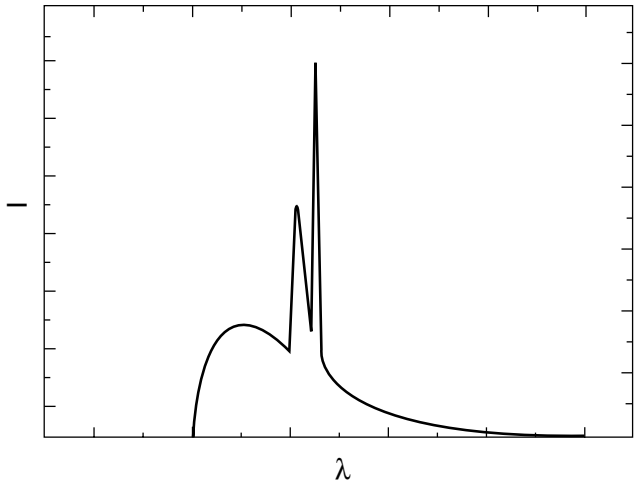

Figure 1 X-ray spectrum with the characteristic lines emitted by an X-ray tube.

and have an intensity that can be up to three orders of magnitude higher than that of the bremsstrahlung radiation. Thus, conventional X-ray tubes can be used as sources of both a continuous X-ray spectrum and discrete characteristic lines.

The main disadvantage of these X-ray tubes is the low brightness of the generated X-rays. Rotating anode sources, which take advantage of bent crystal optics, are able to generate 10^2–10^3 times higher intensity than a conventional tube. However, the application of these sources to the study of the solid/liquid interphase requires long acquisition times. For that reason, most of the X-ray studies of the electrochemical interphase have been carried out using synchrotron radiation, which provides brightness at least 10^3 times higher than that obtained with a rotating anode.

1.2 Synchrotron Radiation

The basis of synchrotron radiation is the electromagnetic radiation emitted by a charged particle under acceleration. When a charged particle (in this case, an electron or positron) describes a circular trajectory as a result of a centripetal acceleration, it emits radiation perpendicular to the trajectory. For electrons traveling at low velocity, the radiation is emitted isotropically in all the directions perpendicular to the acceleration, as shown in Figure 2(a). However, at relativistic velocities the radiation is focused into a cone in the plane of orbit of the electrons and in the forward direction (Figure 2b). The divergence angle (θ_v) is approximately inversely proportional to the wavelength and given by Equation (2):

$$\theta_v \propto \frac{1}{\lambda}, \quad \lambda = \frac{E}{m_0 c^2} \tag{2}$$

where m_0 is the rest mass of the electron.

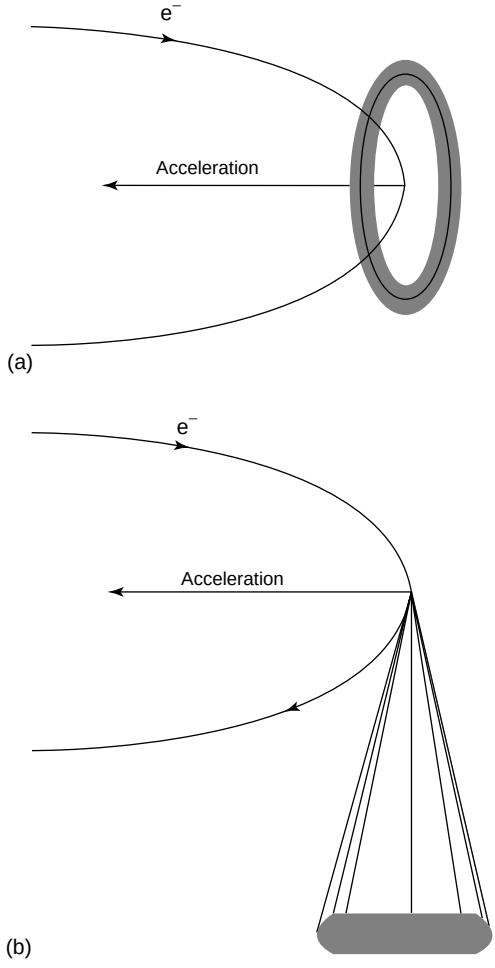

Figure 2 Radiation emitted by an orbiting electron at (a) low and (b) relativistic velocities.

The radiation emitted by relativistic electrons has a very small divergence (around 0.2 milliradian) which results in a high brightness (defined as the number of photons per unit band width, per unit solid angle, per unit area source) of the radiation. Another important quality of the radiation is that it is almost 100% polarized in the plane of the orbit.

The spectrum of the radiation emitted by a relativistic electron in a circular orbit is continuous down to a critical wavelength, which in turn depends on the energy of the electrons and the radius of the orbit that it follows. The high brightness of continuous radiation allows a specific wavelength to be selected for the experiment without being limited to the characteristic lines of a tube generator. Moreover, the intensity of synchrotron radiation is at least 10^3 times higher than that obtained by a rotating anode tube.

A synchrotron (Figure 3) is an approximately circular vacuum chamber with bending magnets for directing the beam, quadrupole magnets for focusing the beam

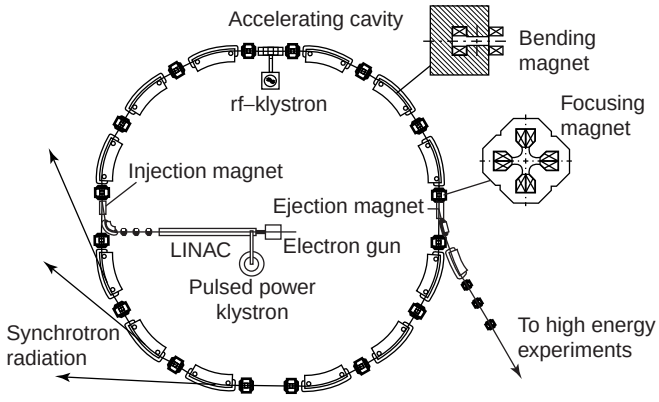

Figure 3 Sketch of a synchrotron. (Reproduced from K. Wilke, 'Synchrotron Radiation Sources', *Rep. Prog. Phys.*, **54**, 1005–1068. © 1991 by permission of the Institute of Physics and the author.)

and radiofrequency cavities for accelerating the electron beam. A synchrotron working properly requires that the energy of the beam be at least 10 MeV. Therefore, a linear accelerator (LINAC) with radiofrequency cavities injects the electron beam continuously in the ring. This classical synchrotron has some disadvantages:[20]

- low stability of the beam;
- the intensity and the critical energy of the radiation changes very quickly depending on the accelerating cycle;
- low beam currents and therefore low intensity;
- the beam is not focused to the extremely small dimensions required by some experiments.

For those reasons, synchrotrons have been replaced by storage rings, which in design are similar to synchrotrons. In these storage rings, the accelerated beam is injected at a given energy, and the machine keeps the beam circulating in the ring for several hours without change in energy. Therefore, the spectrum and the intensity are extremely stable. A radiofrequency cavity is also needed to compensate losses of energy due to the emission of radiation. The electrons circulating in the ring are grouped in bunches, conferring on the beam a time structure which can be used for several experiments, especially in kinetic studies.

In the storage ring, the X-rays are generated in the insertion devices: wigglers, undulators and wavelength sifters. In general, they consist of magnetic structures with alternating poles. As the critical wavelength is proportional to the bending radius, these devices force the electron beam to move with very short radius of curvature. A detailed discussion of the advantage of each insertion device can be found in Wilke.[20]

2 X-RAY ABSORPTION TECHNIQUES

2.1 Interaction of X-rays with Matter

X-rays can interact in two different ways with matter: the X-rays are either scattered or absorbed by the matter. In the first case, the electromagnetic wave of the X-ray beam makes the electrons of matter oscillate, so that they emit radiation in all directions. When no momentum is transferred to the electrons (elastic scattering), the scattered beam has the same wavelength as the incident beam. On the other hand, when momentum is transferred to loosely bound electrons (Compton effect), the scattered beam has a higher wavelength than the incident beam. The scattering phenomenon will be discussed extensively in section 3, therefore in this section we will concentrate on the absorption phenomena.

When monochromatic X-rays travel through matter, part of the incident beam is absorbed by it. The intensity of transmitted beam (I) follows the Beer–Lambert Law (Equation 3):

$$I = I_0 e^{-\mu x} \tag{3}$$

where μ is the linear absorption coefficient (which depends on the wavelength), I_0 the intensity of the incident beam and x the thickness of the sample. Sometimes, Equation (3) is expressed in terms of the mass absorption coefficient μ/ρ, which is obtained by dividing the linear absorption coefficient by the density (ρ). Therefore (Equation 4):

$$I = I_0 e^{-\left(\frac{\mu}{\rho}\right)\rho x} \tag{4}$$

The mass absorption coefficient for an element is almost independent of the physical state and approximately additive. This way, the mass absorption coefficient for a compound, mixture or alloy can be expressed as Equation (5):

$$\frac{\mu}{\rho} = \sum_i w_i \left(\frac{\mu}{\rho}\right)_i \tag{5}$$

where w_i is the mass fraction of the element i and $(\mu/\rho)_i$ its mass absorption coefficient. Except for the discontinuities for the absorption edges (see below), the mass absorption coefficient approximately follows the relationship (Equation 6):

$$\frac{\mu}{\rho} = CZ^4\lambda^3 \tag{6}$$

where C is a constant and Z the atomic number of the element. For practical uses, values of the mass absorption coefficients are tabulated in Ibers and Hamilton.[21]

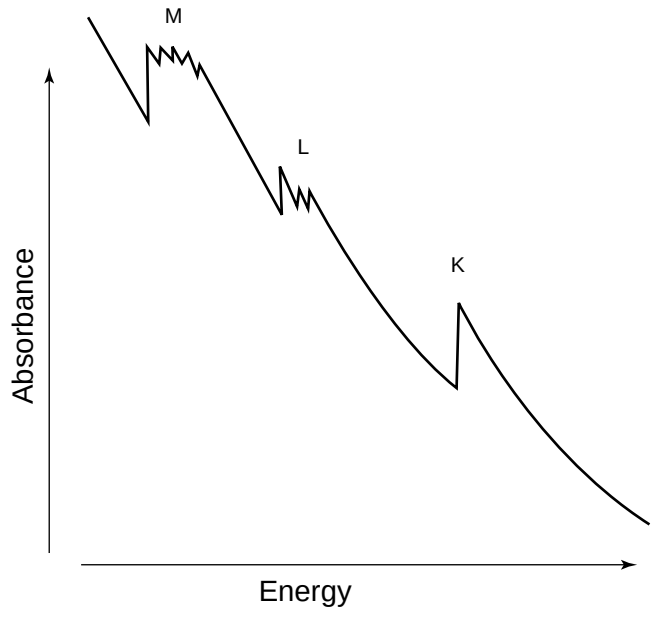

Figure 4 Absorbance spectrum of an atom as a function of the energy of the incident beam.

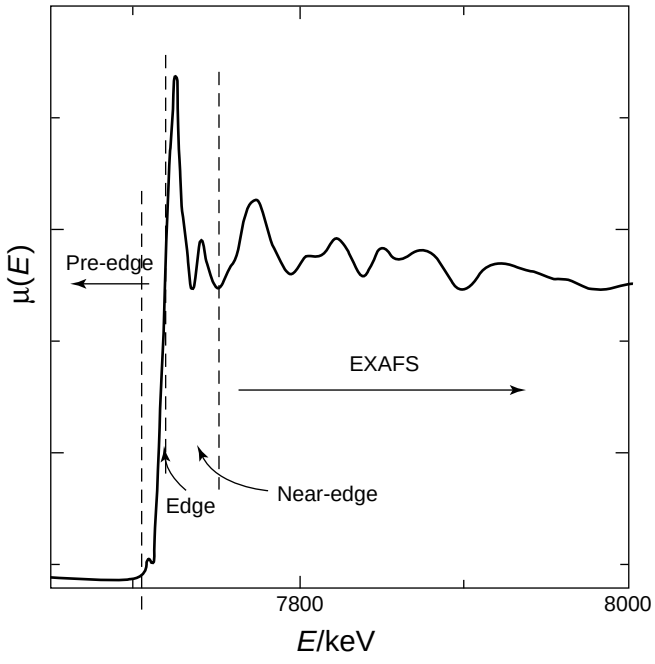

Figure 5 Absorbance spectrum of a CoO foil divided into different regions.

If the energy of the incident beam is increased, it can reach a value where a core electron from an atom in the sample can be ejected, originating an abrupt increase in the absorption coefficient, which is called an absorption edge. There is an absorption edge for each shell and subshell in the excited atom, i.e. there is a K absorption edge, three L edges, five M edges etc. (Figure 4). As expected, the energy of the absorption edge increases in the order M < L < K. After the edge, the diminution in the absorption mass coefficient can also be described by Equation (6), but the constant is different to that obtained at energies below the edge.

Once a hole in the core levels has been created, the excited atom can relax in several ways. First, an electron of the outer shells can descend to fill in the hole, emitting radiation. The energy of this radiation depends on the energy difference between the two levels involved in the electronic transition. This phenomenon is known as fluorescence. The allowed transitions between two electronic levels are restricted to changes in the quantum number of $\Delta l = \pm 1$, and $\Delta m = \pm 1, 0$. The intensity of the fluorescence signal generated depends on the probability of the transition between the two electronic levels.

The relaxation of the atom can also give rise to the emission of Auger electrons or secondary electrons. In the Auger process, one electron descends to occupy the hole in the core levels, and the energy emitted is used to eject an electron from another level. The secondary electrons are emitted when the Auger electrons or fluorescence radiation interacts with the absorbing atom, ejecting one electron from the outer shells.

When the absorption (or fluorescence) spectrum of a substance is recorded, some oscillations are observed after the edge, superimposed on the decay in absorbance that should approximately follow Equation (6). Figure 5 shows the typical absorption spectrum of the K edge of a CoO foil. Although the presence of oscillations (EXAFS) was first discovered by Kronig in 1931,[22,23] the phenomenon was not fully understood until the 1970s, when Stern, Lytle and Sayers developed their theory.[24] From the beginning, it was known that the oscillations contained structural information, since the absorption coefficient for monoatomic diluted gases, for which no interactions between the atoms are expected, showed no oscillations. However, it was not clear whether the oscillations were due to long- or short-range order interactions. The theories developed in the 1970s confirmed that the nature of the oscillations were short-range order interactions.[24–29]

The oscillation in the EXAFS region is the result of an interference phenomenon. The electron ejected from the core level can be treated as a wave because of wave–particle duality. In the presence of another atom in the near neighborhood, this wave can be backscattered and interfere with the outgoing wave (Figure 6). From the interference nature of the EXAFS phenomenon, it is clear that the frequency of the oscillations depends on the distance between the absorbing atom and the backscatterer, and the amplitude is function of the nature and number of backscatterers. Therefore, a detailed

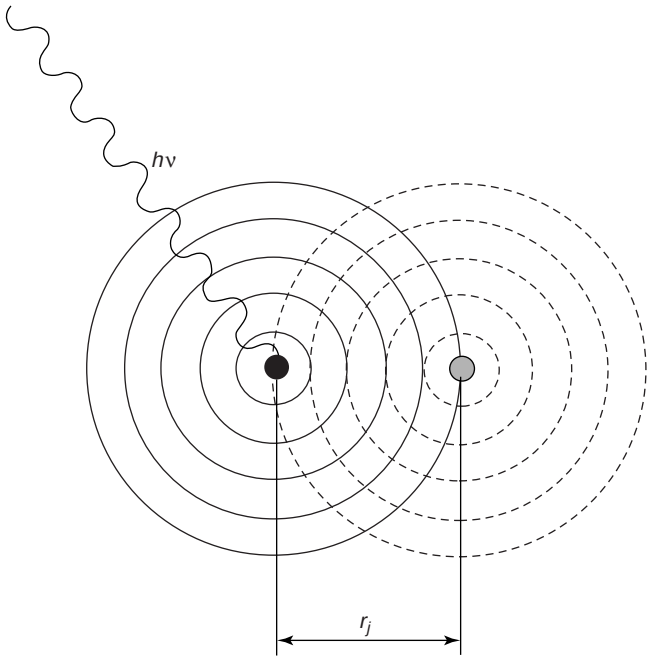

Figure 6 Scheme of the backscattering phenomenon that originates the EXAFS spectra. r_j is the distance between the absorbing and backscattering atoms.

analysis of the EXAFS zone can provide information about the short-range order of the sample.

The theories for describing EXAFS are based on the single-electron single-scattering formalism.[24–29] This assumes that the photoelectron has an energy high enough to suffer only one backscattering event. This condition is normally fulfilled for energies 40 eV above the edge. However, the absorption spectrum of a substance has other regions with important features (Figure 5). Near or below the edge, some absorption peaks may appear, caused by the excitation of core electrons to bond states. The edge position gives information about the effective charge of the absorbing atom, and can be used in the assignment of oxidation states of the atom. In the near-edge region, the photoelectron has a very small momentum and multiple backscattering events take place. All these regions are normally termed XANES. Although the structural information that can be obtained from this region can be very rich, the complex nature of the interaction makes the development of a complete mathematical model difficult. The present status of XANES theory has been reviewed in Bianconi[30] and Stöhr.[31]

2.2 Extended X-ray Absorption Fine Structure and X-ray Absorption Near Edge Structure Theory

As already mentioned, EXAFS refers to the region in the absorption spectrum between 40 and 1000 eV above the edge. In order to obtain the structural information contained in the spectrum, the absorption coefficient as a function of the energy has to be known ($\mu(E)$). The first step towards obtaining the structural information is to isolate the oscillatory part of the spectrum from the rest. This process is carried out by subtracting the absorption coefficient obtained for isolated atoms, $\mu_0(E)$, from the experimental absorption coefficient $\mu(E)$. $\mu_0(E)$ could be obtained for a diluted monoatomic gas, since the interactions between the different atoms are very small and the atoms can be considered as isolated. The result is normalized to $\mu_0(E)$ to obtain the EXAFS function in the energy space ($\chi(E)$) on a per atom basis (Equation 7):

$$\chi(E) = \frac{\mu(E) - \mu_0(E)}{\mu_0(E)} \tag{7}$$

The dual behavior of the emitted electron allows its wavelength to be calculated from Equation (8):

$$\lambda = \frac{2\pi}{k} \tag{8}$$

where k is the wave vector, calculated according to Equation (9):

$$k = \sqrt{\frac{2m}{\hbar^2}(E - E_0)} \tag{9}$$

Here $\hbar$ is $h/2\pi$, m is the mass of the electron and E_0 the energy of the absorption edge. When E is in eV and k in Å^{-1}, Equation (9) is transformed to Equation (10):

$$k = \sqrt{0.2625(E - E_0)} \tag{10}$$

The structural information of the EXAFS can only be extracted after transformation of the EXAFS function to the wave vector space ($\chi(k)$) using Equation (9).

As each shell of atoms surrounding the absorber will give an interference event, the EXAFS spectrum will involve the summation over all these interference events (Equation 11):

$$\chi(k) = \sum_j A_j(k) \sin\left(2kr_j + \phi_{ij}(k)\right) \tag{11}$$

where the subscripts j and i refer to the jth shell of neighboring atoms and the absorbing atom, respectively, $A_j(k)$ is the amplitude term, $\sin(2kr_j + \phi_{ij}(k))$ the oscillatory term, r_j the distance between the atoms in the jth shell to the absorbing atom and $\phi_{ij}(k)$ the total phase shift experienced by the photoelectron. We will discuss each term separately.

2.2.1 *Oscillatory Term:* $\sin\left(2kr_j + \phi_{ij}(k)\right)$

This term contains the information about the distance between the backscattering and absorbing atoms. As seen

in Figure 6, the wave travels twice the distance between the absorbing and backscattering atom. Therefore, in the final state, when it returns to the absorbing atom, it has experienced a phase shift of $2kr_j$. However, the backscattering phenomenon is mainly due to the repulsion between the photoelectron and the electronic cloud of the backscattering atom and the effective distance traveled by the photoelectron wave is shorter than $2kr_j$. An additional term, $\phi(k)$, is required to account for this effect.

2.2.2 The Amplitude Term: $A_j(k)$

This term, which contains information about the nature and number of the atoms in the jth shell, can be written as Equation (12):

$$A_j(k) = \frac{N_j}{kr_j^2} S_i(k) F_j(k) e^{-2\sigma_j^2 k^2} e^{-2r_j/\lambda_j} \tag{12}$$

where N_j is the number of atoms in the jth shell, $F_j(k)$ the backscattering amplitude for a single atom, $S_i(k)$ an amplitude reduction associated to many-body effects, σ_j^2 the Debye–Waller factor and $\lambda_j(k)$ the mean free path of the electron.

As can be seen, the maximum amplitude, given by $N_j F_j(k)$ is reduced by a series of factors. The first one, $1/(kr_j^2)$, has a double diminution effect. First, the EXAFS function will decrease rapidly as the distance between the absorbing and the backscattering atom increases. This means that the contributions of the farther backscattering atoms to the total EXAFS function will be smaller than those arising from the closer atoms, restricting the EXAFS phenomenon to the closest shells and justifying the short-range nature of the EXAFS. The second damping factor is related to k, implying a diminution of the oscillation amplitude as the energy increases. This is one of the factors that restricts the presence of oscillations to approximately 1000 eV above the absorption edge. Another important factor in the decrease of the oscillations with increasing energy is the Debye–Waller factor (see below).

An additional damping factor is related to many-body effects ($S_i(k)$). As the subscript indicates, this term is associated with the absorbing atom and includes losses in the photoelectron energy due to electron shake-up (excitation of other electrons in the absorber) and shake-off (ionization of the electrons of the outer shells).

The reduction in the amplitude due to the thermal vibration and static disorder is contained in $e^{-2\sigma_j^2 k^2}$. When the backscattering event takes place, not all the backscattering atoms in the shell are exactly at the same distance. On the contrary, they may appear displaced from their ''equilibrium'' positions with a mean standard deviation of σ_j^2. The presence of a negative exponential means that the effect of the Debye–Waller factor is more intense at high-energy values (high k values). The EXAFS spectra of samples with a higher Debye–Waller factor (because of vibrations or static disorder) will exhibit oscillations that disappear rapidly above the edge.

The last damping term corresponds to the inelastic losses in the photoelectron energy (e^{-2r_j/λ_j}). The photoelectron that suffers an inelastic loss is not able to contribute to the interference phenomenon. This term is also responsible for the short-range limitation of the EXAFS phenomenon, since the probability of an inelastic loss increases with the distance between the adsorbing and backscattering atoms.

As already mentioned, there is not a complete theory that describes the XANES region of the spectra.[30,31] The information contained in this region is mainly related to the absorbing atom. The position of the edge is related to the electronic density of the absorbing atom, in a similar way to X-ray photoelectron spectroscopy (XPS). Other features of the region, such as the so-called ''white line'' (a sharp peak appearing after the edge) is associated with transitions to unoccupied states. These transitions are also affected by the symmetry of the adsorbing atom.

2.3 Data Analysis

In order to obtain the structural information contained in the EXAFS spectrum, the data have to be treated and manipulated. The reliability of the data obtained from the EXAFS spectrum depends on the correctness of the data manipulation. A very detailed data analysis can be found in Teo.[32] It normally comprises the following steps:

1. background removal and normalization;
2. conversion to the k space;
3. k weighting;
4. Fourier transforming and filtering;
5. fitting for the phase and amplitude of the filtered spectrum.

We will discuss now in detail every point in the data analysis.

2.3.1 Background Removal and Normalization

Prior to background subtraction, the measured data has to be converted to absorption coefficient. When the measured property is the transmitted intensity, the total linear absorption coefficient can be calculated from Equation (13):

$$\mu(E)x = \ln \frac{I_0}{I} \tag{13}$$

Another possibility is to measure the fluorescence intensity (F). As the probability of emitting a photoelectron is

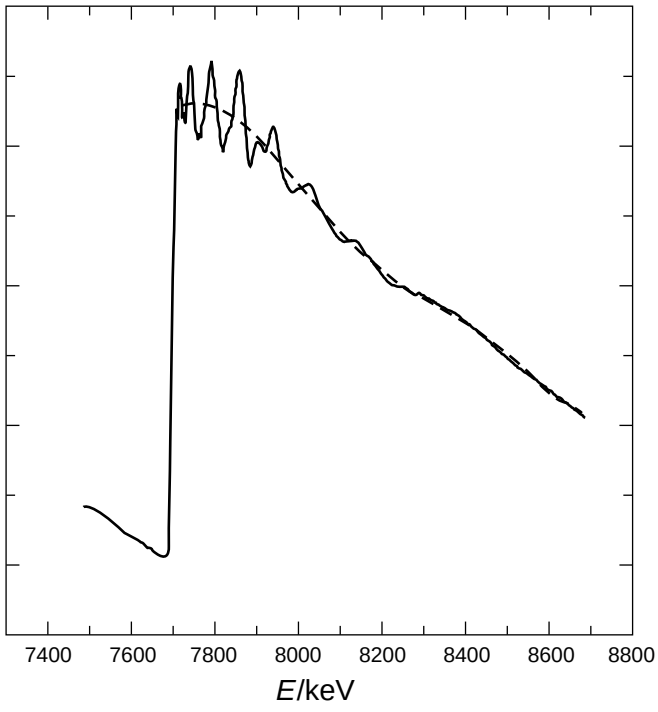

Figure 7 EXAFS spectrum of Co. The dashed line represents the fitted spline to calculate $[\mu(E) - \mu_0(E)]$.

directly proportional to the probability of absorption, the absorption coefficient is given by Equation (14):

$$\mu(E)x = \frac{F}{I_0} \tag{14}$$

In order to calculate the EXAFS function ($\chi(E)$) according to Equation (7), the value of $\mu_0(E)$ has to be known. Normally $\mu_0(E)$ values are not available and thus other strategies have to be used to calculate $\chi(E)$. Moreover, the spectrum also contains additional background elements, such as the spectrometer baseline, beam harmonics, elastic scattering, etc. The strategy normally followed to calculate $\mu_0(E)$ and to remove all the additional background is to fit a polynomial spline to the EXAFS part of spectrum. Figure 7 shows the spline obtained for the Co spectrum.

As mentioned above, the spline obtained is not only $\mu_0(E)$ but also contains the additional background, thus it is not possible to use it as normalization factor. One of the most used strategies is to divide $[\mu(E) - \mu_0(E)]$ by the edge jump ($\mu_{E'}$). Figure 8(a) shows the EXAFS function obtained for a Co film after background subtraction and normalization.

2.3.2 Conversion to k Space

The conversion to k space according to Equation (9) requires the determination of the edge threshold (E_0). However, the edge position cannot be identified with any feature of the spectrum. Several points can be chosen as the edge threshold, i.e. the onset of the jump, the inflexion point, the mid-point in the edge jump, the edge peak, etc. Despite such an unclear definition of edge threshold, the exact position of the edge is not very important, provided that it is chosen similarly for related compounds. The final analysis of the EXAFS spectrum normally uses comparison with model compounds or theoretical models, in which the value of E_0 is treated as an adjustable parameter.

2.3.3 k Weighting

The oscillations in the EXAFS function are attenuated at high k values by a series of factors. The amplitude of the oscillation has $1/k$ factor and normally the backscattering amplitude $[F_j(k)]$ also depends on $1/k^2$. Thus, the oscillations appearing at low energies (low wave vector) dominate over those at high energies. On the other hand, the interatomic distances depend only on the frequency, not on the amplitude of the oscillations. To prevent any effect of the amplitude on the distance determination, the EXAFS function is normally weighted by k^1, k^2 or k^3. k^1 factor cancels the $1/k$ factor in the amplitude functions and k^3 cancels both the $1/k$ and the $1/k^2$ factors. k^3 is normally preferred for light absorbing atoms and k^1 for heavy absorbing atoms. When comparing the k weighted function (Figure 8b) to the original EXAFS function (Figure 8a), it can be seen that the oscillations on the k weighted function shows a higher relative amplitude for high energies.

2.3.4 Fourier Transform and Filtering

Fourier transform allows the EXAFS function to be translated from the k space to the frequency (or distance) space (Equation 15):

$$\phi(r) = \frac{1}{\sqrt{2\pi}} \int_{k_{\min}}^{k_{\max}} w(k)k^n\chi(k)e^{2ikt}\,dk \tag{15}$$

where $w(k)$ is a window function which selects the k range to be transformed. Fourier transform yields a function that resembles a radial distribution function, in which a number of peaks appear at given distances (Figure 8c). The advantage of Fourier transform is that the transform of a summation of functions is the summation of the Fourier transform of the individual functions. Thus every peak appearing in the radial distribution function can be assigned to a given shell (unless the distances between two shells are small enough to prevent peak resolution in the frequency space). The distance at which each peak appears does not correspond exactly to the distance between the absorbing atom and the shell since

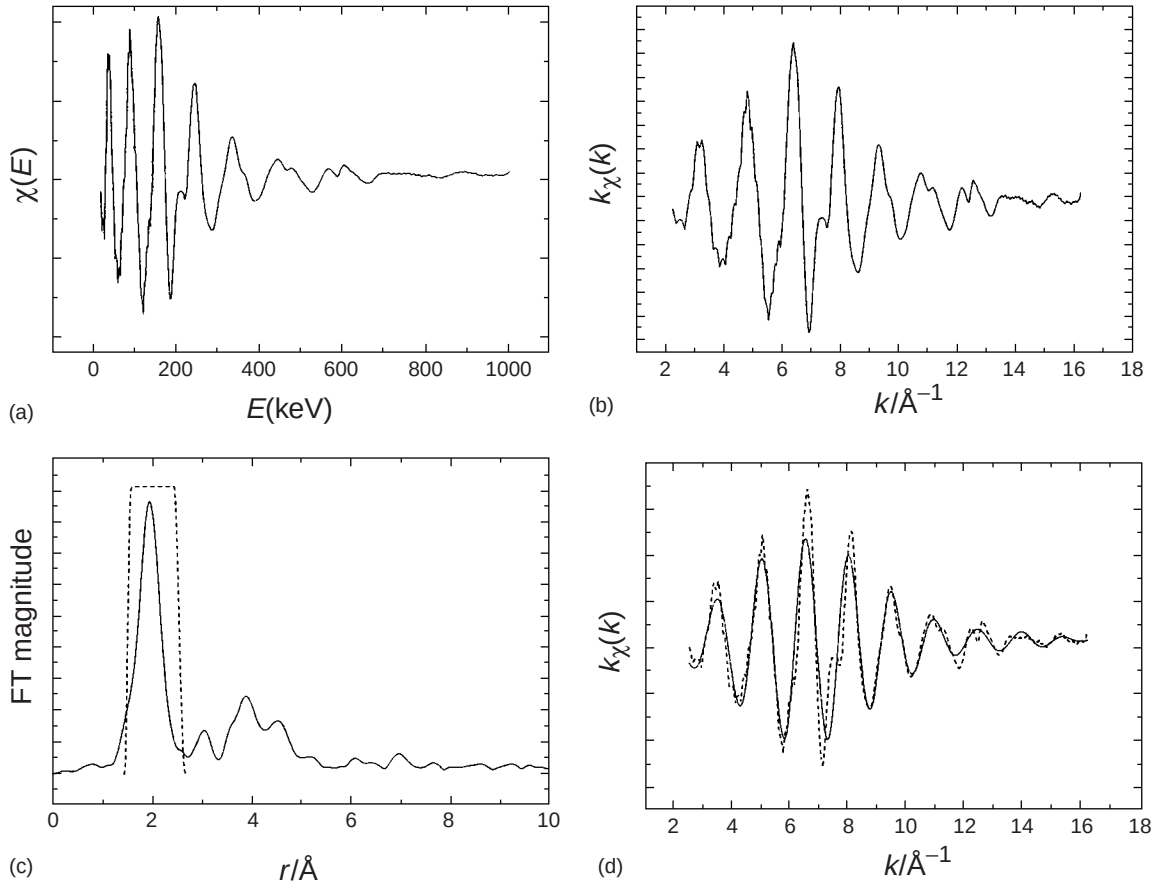

Figure 8 Four steps in the data analysis of an EXAFS spectrum: (a) EXAFS spectrum after background removal and normalization, (b) k^1 weighting, (c) Fourier transform and (d) EXAFS function after Fourier filtering. The dashed line in (c) represents the window used for the Fourier filtering and in (d) represents the original k weighted function (panel (b)).

these distances have to be corrected for the total phase shift ($\phi_{ij}(k)$) (see section 2.2). Fourier back transforming the desired shell or shells using a window as shown in Figure 8(c) (dashed line) allows filtering the $k^n\chi(k)$ function and eliminating the undesired contributions (noise, high frequencies...) (Figure 8d).

2.3.5 Fitting for Phase and Amplitude

The last part of the data analysis is obtaining the structural parameters. As shown in section 2.2, the EXAFS function depends on several parameters: r_j, N_j, σ_j and λ_j. In order to obtain these parameters from the spectrum, the total phase shift ($\phi_{ij}(k)$) and backscattering amplitude ($F_j(k)$) have to be known. One of the usual approaches to calculating both functions is to use the spectrum of model compounds, from which the different parameters are known. Another possibility is to generate these functions from theoretical models, such as the one proposed by Teo and Lee[33] or from the program code FEFF.[34–37]

The use of $\phi_{ij}(k)$ and $F_j(k)$ obtained theoretically or from model compounds implies that they are similar to those found in our spectrum, and this is normally termed as phase and amplitude transferability. The phase transferability is normally regarded as very good, since the phase shift depends on the core electrons of the absorbing and backscattering atoms. Therefore, it is almost independent of the chemical state of both atoms and only depends on their nature. This allows very accurate determination of the interatomic distances, with an estimated error of ±0.01 Å. On the other hand, amplitude transferability is considered worse than the phase transferability, since it depends on several factors. Except when very similar model compounds are used, the determination of N_j is not better than ±20%.

The strategy for obtaining the different structural parameters is a nonlineal least-squares fitting with the following function (Equation 16) to the k^n-weighted filtered function:

$$Y(k) = \sum_j \frac{N_j}{r_j^2} F_j(k_j) k_j^{n-1} e^{-2\sigma_j^2 k_j^2} e^{-2\frac{r_j}{\lambda_j(k_j)}} \times \sin\left(2k_j r_j + \phi_{ij}(k_j)\right) \quad (16)$$

where $F_j(k_j)$ and $\phi_{ij}(k_j)$ functions are obtained from model compounds or theoretical models. Here the value of E_0 is allowed to vary for every shell, giving a different k_j value for the shell according to Equation (17):

$$k_j = \sqrt{k^2 - 0.2625(E_{0_j} - E_0)} \tag{17}$$

2.4 Experimental Aspects

2.4.1 Detection

When the X-ray beam interacts with matter, four different processes occur: absorption, fluorescence, Auger electron emission and secondary electron emission.[38] As they are proportional to the absorption coefficient, any of these processes can be use to record an EXAFS spectrum. However, the solvent in the electrochemical cell readily absorbs the Auger or secondary electrons. Therefore, the use of Auger or secondary electrons in in situ electrochemical studies is completely precluded.

The mode of detection is generally dictated by the concentration of the species of interest. In electrochemical environments, we are normally interested in the species in the interphase area, which represent a very small amount of the total number of species in the whole electrochemical system. For that reason, the preferred method of detection is fluorescence using a solid-state detector, which has a high energy resolution and sensitivity. The main drawback of the solid-state detector is the long counting times required to record a spectrum. Total reflection geometry can be used to increase sensitivity.[11,12] The absorption detection strategy is normally restricted to very porous materials, which have a very high surface-to-volume ratio.

In recording an EXAFS spectrum of a surface species (SEXAFS), it is also important to consider the polarization geometry. Synchrotron radiation is highly polarized in the orbit plane. Because of selection rules, a backscattering event will take place only when the position vector of the backscattering atom has a component in the polarizing plane. The number of neighbors obtained from the spectrum (N_{eff}) is related to the real number of neighbors (N) for a K edge according to Equation (18):

$$N_{\text{eff}} = 3\sum_{i=1}^{N} \cos^2 \alpha_i \tag{18}$$

where α_i is the angle between the position vector of the i-esim atom and the polarization vector. In-plane (polarization vector parallel to the surface) and out-of-plane (polarization vector perpendicular to the surface) polarization geometries have been used to characterize the exact position of the atoms of an adsorbed layer with respect to the surface atoms.[39]

2.4.2 Time Resolved Extended X-ray Absorption Fine Structure

In a normal EXAFS experiment, the energy of the incident beam is scanned with the use of a monochromator. The time required to cover the full EXAFS region by the monochromator prevents the use of this technique in kinetic studies. The alternative for time resolved studies is the use of dispersive arrangements.[40–44] In this case, the use of bend optics permits a range of energies to be focused on the sample (normally 600 eV). A photodiode array is employed for the simultaneous detection of the full range and the whole spectrum can be recorded in a millisecond range.

2.4.3 Electrochemical Cell

The design of an electrochemical cell depends on the electrode material and the type of research. There are two main categories, classified by the type of electrode surface: smooth or rough. Figure 9 depicts the basic design used in all the cells employed for smooth electrodes (generally single crystal electrodes).[45–47] The cell body has two ports for electrolyte exchange, two ports for the counter and reference electrodes and an X-ray transparent polymer film which serves as a window. The polymer film has two positions, inflated and deflated. The deflated position, in which the cell attains a thin layer configuration, is used during the data acquisition. This configuration is required to minimize the scattering due to the solvent. When the electrolyte or potential is changed, the polymer is set in the inflated position, since the equilibrium conditions are reached faster.

A typical cell used in the studies of rough electrodes is depicted in Figure 10.[48] In this cell, the working electrode is set perpendicular to the beam. Unlike the previous cell, this design allows working in absorption or fluorescence mode. The cell is also in thin layer configuration.

A different approach has been used in the study of passive films.[49,50] In this case, the electrodes were

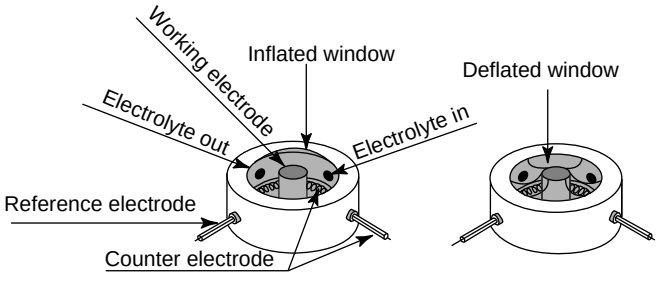

Figure 9 Typical cell design for an X-ray cell used with smooth electrodes, with the polymer window in the inflated and deflated positions.

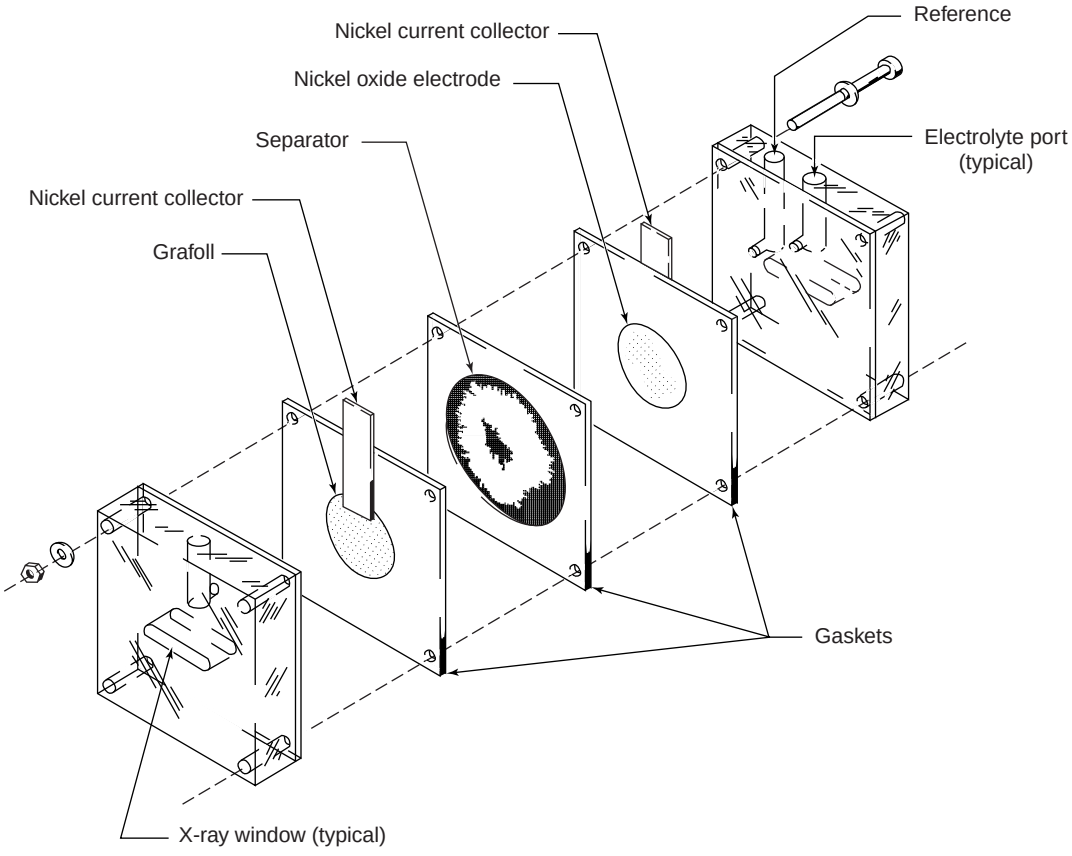

Figure 10 EXAFS cell used in the study of nickel oxide electrodes. (Reprinted with permission from J. McBreen, W.E. O'Grady, K.I. Pandya, R.W. Hoffman, D.E. Sayers, *Langmuir*, **3**, 428–433 (1987). © American Chemical Society.)

prepared by sputtering thin films of the electrode material (iron) onto Mylar™ film, in which a thin film of gold or tantalum had previously been sputtered to improve the electrical conductivity. This film is then mounted in the electrochemical cell. The EXAFS and XANES spectra are taken in fluorescence mode, through the back of the electrode (the polymer film). The advantage of this configuration is that the X-ray does not travel through the electrolyte solution, and the electrochemical cell is not restricted to a thin layer configuration. The main disadvantage is that the EXAFS and XANES spectra contain information not only of the electrode surface but also from the bulk electrode material. For this reason, the studies are restricted to thin films.

A complete review of the different cell designs can be found in Sharpe et al.[14]

3 X-RAY SCATTERING TECHNIQUES

As mentioned in section 2.1, the X-rays can be absorbed or scattered by matter. In the latter case, the different waves scattered from different atoms can interfere and maxima in the scattered intensity can be obtained at given incident angles. This is normally known as diffraction. However, this is not the only use of the scattering measurements. Additional information can be obtained by studying the scattered intensity in the regions between the diffraction peaks, normally known as CTR and Bragg rods. In the following text, the diffraction and the CTR and Bragg rod measurements will be treated in two different sections. However, the separation is just made for clarity, since both diffraction and rod studies are different ways of measuring the same property: the X-ray scattering.

3.1 X-ray Diffraction

When elastic scattering occurs, the different waves scattered from different atoms can interfere. As the path length is different for each wave, they can interfere constructively or destructively, depending on the incident and scattering angles. In crystalline materials, which have the atoms in fixed positions, the interference phenomenon will give rise to very characteristic diffraction patterns.

A crystalline material is composed of a lattice of identical structural units. Considering a material that has only one atom per unit cell, the position of each atom in the crystal can be found according to Equation (19):

$$\mathbf{R}_{xyx} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \tag{19}$$

where **a**, **b** and **c** are the lattice vectors and x, y and z integers. In the Bragg approach, the atoms are considered occupying a set of parallel planes with a d spacing. A maximum in the scattered radiation will take place when the reflections from successive planes interfere constructively (Figure 11). The difference in the path length for the two waves is $2d\sin\theta$. Thus, a constructive interference will occur when this difference be equal to an integral number of wavelengths (Bragg's law) (Equation 20):

$$n\lambda = 2d\sin\theta \tag{20}$$

In X-ray diffraction, it is convenient to define the scattering vector $\mathbf{Q} = \mathbf{K_0} - \mathbf{K_H}$, as the difference between the incident ($\mathbf{K_0}$) and scattered ($\mathbf{K_H}$) wave vector. In elastic scattering, $\mathbf{K_0}$ and $\mathbf{K_H}$ have the same magnitude ($2\pi/\lambda$). Therefore, **Q** is given by Equation (21):

$$\mathbf{Q} = \frac{4\pi}{\lambda}\sin\theta \tag{21}$$

which leads to an alternate expression (Equation 22) for Bragg's law using Equation (20):

$$\mathbf{Q} = n\frac{2\pi}{d} \tag{22}$$

In order to study the X-ray diffraction is useful to employ the Von Laue approach. In the first step it is convenient to define **Q** as a linear combination of the primitive reciprocal lattice vectors, $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ (Equation 23):

$$\mathbf{Q} = Q_a\mathbf{a}^* + Q_b\mathbf{b}^* + Q_c\mathbf{c}^* \tag{23}$$

$\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ are defined by Equation (24):

$$\mathbf{a}^* = 2\pi\frac{\mathbf{b}\times\mathbf{c}}{V} \quad \mathbf{b}^* = 2\pi\frac{\mathbf{c}\times\mathbf{a}}{V} \quad \mathbf{c}^* = 2\pi\frac{\mathbf{a}\times\mathbf{b}}{V} \tag{24}$$

where V is the unit cell volume given by $(\mathbf{a}\cdot\mathbf{b}\times\mathbf{c})$.

Within the kinematic approximation (also known as the Born approximation), which assumes a single scattering event for the photon, the scattering amplitude (A) can be calculated from Equation (25):

$$A = \frac{e^2}{mc^2R} \int e^{iQr}\rho(\mathbf{r})\,d^3\mathbf{r} \tag{25}$$

where e^2/mc^2R is the scattering from a single electron neglecting the polarization factor $(1 + 1/2\cos^2\theta)$, R is the distance between the sample and detector and $\rho(\mathbf{r})$ is the electron density. The total electron density is obtained by summing the electron density for each atom in the crystal (Equation 26), giving a scattered amplitude of:

$$\begin{aligned} A &= \frac{e^2}{mc^2R} \int e^{iQr}\rho(\mathbf{r})\,d^3r \sum_{xyz} e^{iQR_{xyz}} \\ &= A_e f(\mathbf{Q})\mathrm{F}(\mathbf{Q}) \end{aligned} \tag{26}$$

where the atomic form factor ($f(\mathbf{Q})$) is the Fourier transform of the atomic electron density ($\rho(\mathbf{r})$) and the interference function ($\mathrm{F}(\mathbf{Q})$) is the summation term. Thus (Equation 27), the scattering intensity(I) is equal to:

$$I = |A|^2 = A_e^2|f(\mathbf{Q})|^2|\mathrm{F}(\mathbf{Q})|^2 e^{-2\sigma^2|\mathbf{Q}|^2} \tag{27}$$

where the Debye–Waller factor has been included to account for thermal vibrations. Since in Equation (27), the $|\mathrm{F}(\mathbf{Q})|^2$ changes rapidly with **Q**, the position of the diffraction spots will be given by the maxima in $|\mathrm{F}(\mathbf{Q})|^2$. Using Equation (23) the interference function can be written as Equation (28):

$$\begin{aligned} |\mathrm{F}(\mathbf{Q})|^2 &= |\mathrm{F}(Q_a, Q_b, Q_c)|^2 \\ &= \sum_{x=1}^{N_a}\sum_{y=1}^{N_b}\sum_{z=1}^{N_c} e^{2\pi i(xQ_a + yQ_b + zQ_c)}{}^2 \\ &= \frac{\sin^2\pi N_aQ_a}{\sin^2\pi Q_a}\frac{\sin^2\pi N_bQ_b}{\sin^2\pi Q_b}\frac{\sin^2\pi N_cQ_c}{\sin^2\pi Q_c} \end{aligned} \tag{28}$$

where N_a, N_b and N_c are the total number of unit cells in the **a**, **b** and **c** directions, respectively. This function maximizes (Equation 29) when:

$$Q_a = h \quad Q_b = k \quad Q_c = l \tag{29}$$

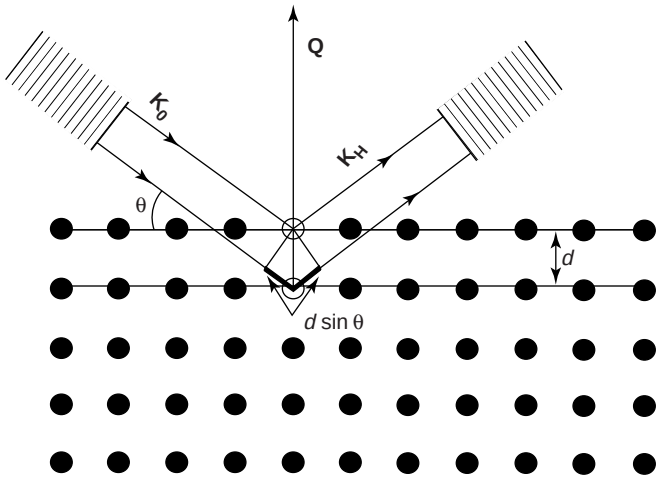

Figure 11 Bragg diffraction from lattice planes with a d spacing.

with h, k and l integers. The position of the diffraction (Bragg) peak is normally termed as (h, k, l). Since N_a, N_b and N_c tend to infinity, Equation (28) can be transformed to Equation (30):

$$|\mathrm{F}(\mathbf{Q})|^2 \longrightarrow (N_a N_b N_c)\delta(Q_a - h)\delta(Q_b - k)\delta(Q_c - l) = N\delta(\mathbf{Q} - \mathbf{G}_{hkl}) \tag{30}$$

where δ is the Dirac function, N the total number of atoms in the crystal and $\mathbf{G}_{hkl}$ is defined as $h\mathbf{a}^* + k\mathbf{b}^* + k\mathbf{c}^*$. The intensity of the Bragg peak is therefore (Equation 31):

$$I_{hkl} = |A|^2 = A_e^2 |f(\mathbf{Q})|^2 (N_a N_b N_c)^2 \mathrm{e}^{-2\sigma^2|\mathbf{Q}|^2} \tag{31}$$

For ideal crystals, the diffraction pattern consists of perfect δ-functions, with infinite value for a Bragg diffraction peak and zero value for the rest of the points of the reciprocal space. However, real crystals present imperfections, defects, grain boundaries, etc. A real crystal can be considered as a set of domains (a lattice of atoms perfectly oriented), in which each domain can have an orientation that differs from the next, giving different diffraction patterns. This results in a broadening of the diffraction peaks. Defining $\Delta\mathbf{Q}$ as the full width at half-maximum (fwhm) of the diffraction peak, the domain size (L) can be estimated from Equation (32):

$$L \approx \frac{2\pi}{\Delta\mathbf{Q}} \tag{32}$$

When the unit cell has several atoms, the atomic form factor ($f(\mathbf{Q})$) is replaced by the structure factor defined by Equation (33):

$$S(\mathbf{Q}) = \sum_{j=1}^{n} f_j(\mathbf{Q}) \mathrm{e}^{i\mathbf{Q}r_j} \tag{33}$$

where n represents the total number of atoms in the unit cell and the subscript j the jth atom in the cell.

When considering diffraction from monolayers and surfaces, they can be treated as a two-dimensional lattice structure, that is, a crystal that has $N_c = 1$. Substituting that value in Equation (28) provides the new diffraction conditions (Equation 34):

$$|\mathrm{F}(\mathbf{Q})|^2 = \frac{\sin^2 \pi N_a Q_a}{\sin^2 \pi Q_a} \frac{\sin^2 \pi N_b Q_b}{\sin^2 \pi Q_b} \tag{34}$$

with maxima at $Q_a = h$ and $Q_b = k$. No condition is limiting the value of Q_c, so the maxima are obtained regardless of its value. This results in the transformation of the diffraction spots into rods with $Q_a = h$ and $Q_b = k$. Such rods are normally termed Bragg rods. This way, the diffracted intensity of the rod will be (Equation 35):

$$I_{hk} = |A|^2 = A_e^2 |f(\mathbf{Q})|^2 (N_a N_b)^2 \mathrm{e}^{-2\sigma^2|\mathbf{Q}|^2} \tag{35}$$

This equation can be used to calculate the different intensity of the diffraction spots from an adsorbed single layer on a surface. As the diffraction from monolayers and surfaces is independent of the value of l, surface diffraction spots are normally termed as (h, k).

3.2 Crystal Truncation Rods and Bragg Rods

The termination of a crystal in a surface or interface will give to the intensity profile some characteristics of the Bragg rods. Thus, the scattered intensity is not zero in the regions connecting two Bragg diffraction spots along directions normal to the surface, giving rise to the CTRs. For a given diffraction peak at (h, k, l), a CTR will occur at (h, k, Q_c). The intensity of such rods can be used to probe the structure of the interphase normal to the electrode surface. Normally, the CTR is measured as reflectivity (defined as the ratio of the integrated reflected intensity to the integrated incident intensity).[51–55] The integrated intensity is preferred in CTR studies since it is independent of the crystal imperfections and instrumental resolution. Thus, integrating Equation (27) for a given value of h, k over the angles accepted by the detector, Equation (36) is obtained:

$$R = T^4(Q_z) \frac{4\pi^2 r_0^2}{\Gamma^2 |\mathbf{k_0}|^2 \sin^2 \alpha} |f(Q_z)|^2 \mathrm{e}^{-2\sigma^2 Q_z^2} \sum_{n=0}^{\infty} \mathrm{e}^{iQ_z dn} \, \mathrm{e}^{-\frac{Q_{\mathrm{abs}}}{Q_z}} \tag{36}$$

where $T(Q_z)$ is the Fresnel factor and has been added to account for enhanced surface scattering process, r_0 the Thomson radius (e^2/mc^2), Q_z the projection of $\mathbf{Q}$ in the direction normal to the surface, Γ the area per atom in each layer, d the layer spacing, and α the incident angle. The term $\mathrm{e}^{-Q_{\mathrm{abs}}/Q_z}$ accounts for the losses in the intensity due to the absorption of the window material and the thin layer of solvent.[52] This equation has been obtained for equal incident and exit angles. For specular CTR, i.e. when $\alpha = \theta$ or when $h = k = 0$, Equation (36) is converted to Equation (37):

$$R = T^4(Q_z) \frac{16\pi^2 r_0^2}{\Gamma^2 Q_z^2} |f(Q_z)|^2 \mathrm{e}^{-2\sigma^2 Q_z^2} \sum_{n=0}^{\infty} \mathrm{e}^{iQ_z dn} \, \mathrm{e}^{-\frac{Q_{\mathrm{abs}}}{Q_z}} \tag{37}$$

Solving the summation term of Equation (36) gives Equation (38):

$$\sum_{n=0}^{\infty} \mathrm{e}^{iQ_z dn} = \frac{1}{2\sin^2(\pi Q_z)} \tag{38}$$

Therefore, for an ideally truncated crystal, in which the topmost layer has the same lattice structure and d spacing as the bulk crystal, the CTR should show a dependence on $[\sin^2(\pi Q_z)]^{-1}$. However, crystals are not ideally truncated in most cases. The topmost layer of

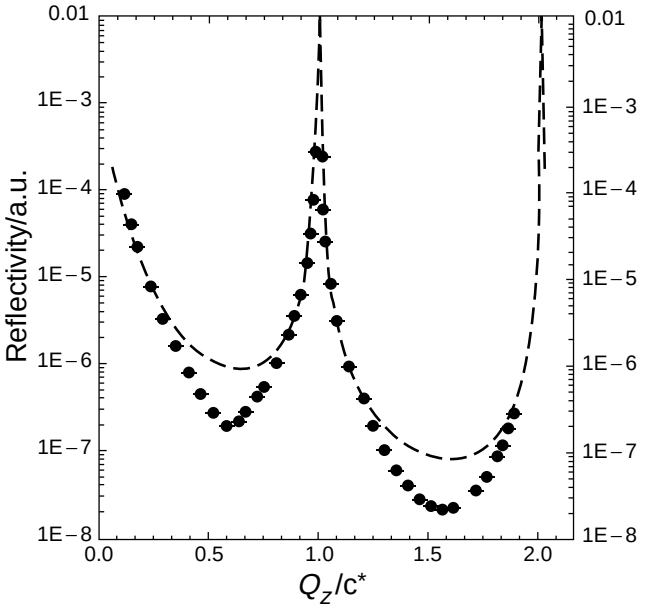

Figure 12 CTR measurement of a Au(111) electrode in 0.10 M H_2SO_4 containing 1.0 mM $Br^- + 1.0$ mM Cu^{2+} measured at E = +0.70 V (vs Ag/AgCl) (●). The dashed line is the calculated CTR curve from an ideally truncated Au(111) crystal.

the crystal may be reconstructed, the d spacing may be longer or shorter than in the bulk crystal, resulting in a truncation rod that does not follow the dependence on $[\sin^2(\pi Q_z)]^{-1}$. Moreover, the presence of overlayers on the surface greatly affects the reflectivity of the truncation rod, creating asymmetries in the CTR profile and valleys that have a lower intensity than those expected for an ideally truncated surface (Figure 12).

The presence of overlayers or reconstruction phenomena has a clear effect on the CTR profile. For those cases, Equation (36) has to be modified to account for the presence of such ''imperfections''. For instance, for a surface in which the topmost layer can be reconstructed and t different overlayers are present, the reflectivity can be calculated from Equation (39):

$$R = T^4(Q_z)\frac{4\pi^2 r_0^2}{\Gamma^2|\mathbf{k_0}|^2\sin^2\alpha}|S(Q_z)|^2 e^{-\frac{Q_{abs}}{Q_z}} \qquad (39)$$

where $S(Q_z)$ contains the sum over the atomic layers represented by Equation (40):

$$\begin{aligned} S(Q_z) = \sum_{n=1}^{\infty} & f_{bulk}(Q_z)\rho_{bulk} e^{-Q_z^2\sigma_{bulk}^2} e^{iQ_z n d_{bulk}} \\ & + f_s(Q_z)\rho_s e^{-Q_z^2\sigma_s^2} e^{-iQ_z(d_s - d_{bulk})} \\ & + \sum_{m=1}^{t} \left(f_m(Q_z)\rho_m e^{-Q_z^2\sigma_m^2} e^{-iQ_z\left(d_s - d_{bulk} + \sum_{j=1}^{m} d_j\right)} \right. \end{aligned} \qquad (40)$$

Each layer is defined by a Debye–Waller factor (σ), a layer distance from the layer below (d) and a relative atomic density (ρ) defined as the ratio of the number of atoms in the nth layer to the number of atoms in a layer of the bulk material (obviously $\rho_{bulk} = 1$). The first summation term corresponds to the different layers in the bulk material, the second term to the surface layer and the last one to the overlayers deposited on the surface. In Equation (40), only the topmost layer of the crystal is allowed to reconstruct, expand or contract with respect to the layer structure of the bulk material. Fitting this equation to the experimentally obtained CTR profile will allow determination of the different layer parameters.

CTR and diffraction measurements can be considered as complementary techniques. Whereas diffraction gives the in-plane structure, CTR provides information about the out-of-plane structure, allowing determination of the interlayer distances and of the nature of the species in the layer. The main disadvantage of the CTR measurements is that for complicated systems, the number of adjustable parameters is very high and several sets of parameters can be found that fit the experimental data. In such cases, the CTR measurements only allow some of the possible models for the interphase, proposed according to the diffraction measurements, to be discarded. Moreover, the error in the coverage values given by the CTR measurements is higher than that obtained from diffraction. On the other hand, CTR analysis is not restricted to the presence of ordered layers, unlike diffraction. Information on the coverage and the d spacing can be obtained for disordered overlayers or surfaces.

For ideal flat overlayers and surfaces, the intensity of the Bragg rods is independent of Q_z. However, real overlayers and surfaces can have atoms with different d spacing or can be constituted of two different species. In such cases the intensity (or reflectivity) is not constant and depends on Q_z. For such cases, the analysis of the Bragg rods is similar to that used for the CTR measurements, but with a structure factor containing only the terms related to the overlayers (or the surface). A different term has to be added for every different species appearing at a given d distance from the surface.

3.3 X-ray Standing Waves

Two different phenomena are combined in this technique: diffraction from a perfect crystal and interference. The interference between the incident and diffracted beams creates the standing waves that can be used to probe the position of the foreign atoms in the crystal, or adsorbed on the surface.

When the diffraction theory was considered, the intensity of the diffraction beam was calculated by using the kinematic approach. However, when perfect crystals are to be considered, this theory does not describe the intensity adequately and a dynamic approach has to be used.

In the vicinity of a Bragg diffraction peak, the incident and diffracted beams interfere to generate standing waves. These waves extend up to 1000 Å from the interface in both directions (towards the bulk and the interphase), allowing probing of the interphase and the bulk material. These waves have nodal and antinodal planes that are parallel to the diffracting planes with a nodal wavelength equal to the d spacing of the diffracting planes. As the angle of incidence is advanced through the Bragg reflection, the relative phase between the incident and diffracted beam changes by $-\pi$. Owing to this change, the nodal plane moves $d/2$ in the direction perpendicular to the surface (Figure 13).

If the energy of the incident beam is set slightly above the absorption edges of the atom we want to study in the interphase, the fluorescence yield of this species will exhibit a modulation as the angle is scanned through the Bragg peak. Analysis of such modulation will serve to determine the distance of the species relative to the surface. However, only one measurement will not be enough to determine the distance r, since two atoms separated by a distance equal to d would give the same modulation. Therefore, several Bragg diffractions or specular reflections have to be measured in order to unequivocally assign a distance. A complete review on the XSW analysis can be found in Zegenhagen.[56]

In some cases, layered synthetic microstructures (LSM) has been used instead of perfect crystals. LSM are periodic structures consisting in alternating layers of low and high electron density materials. These materials have several advantages:

- They can be manufactured with a great variety of d spacings and materials, allowing the study of the diffuse layer.
- The reflection curves compare well with the dynamic diffraction theory.
- They have a rather large energy band pass.[11,12]

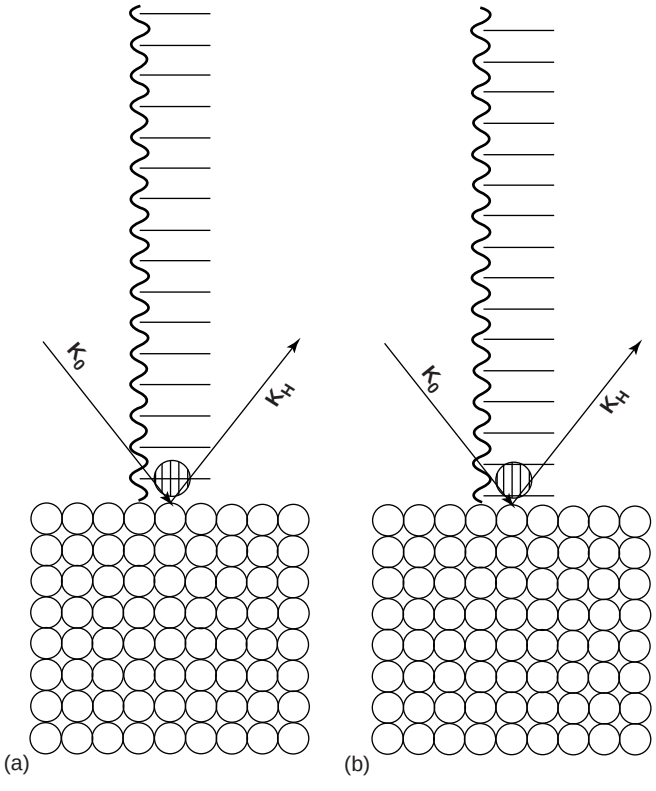

Figure 13 Schematic representation of the standing wave generated by the interference of the incident and the scattered beam before a Bragg diffraction peak (a) and after it (b).

3.4 Experimental Aspects

3.4.1 Four Circle Diffractometer

A four circle diffractometer (Figure 14) is used in order to search all the positions in the reciprocal space in scattering experiments. The sample has three rotational degrees of freedom: θ, χ and ϕ and the detector one, 2θ. Obviously, there is an indeterminacy in the position in the reciprocal space, since only three rotational degrees are required

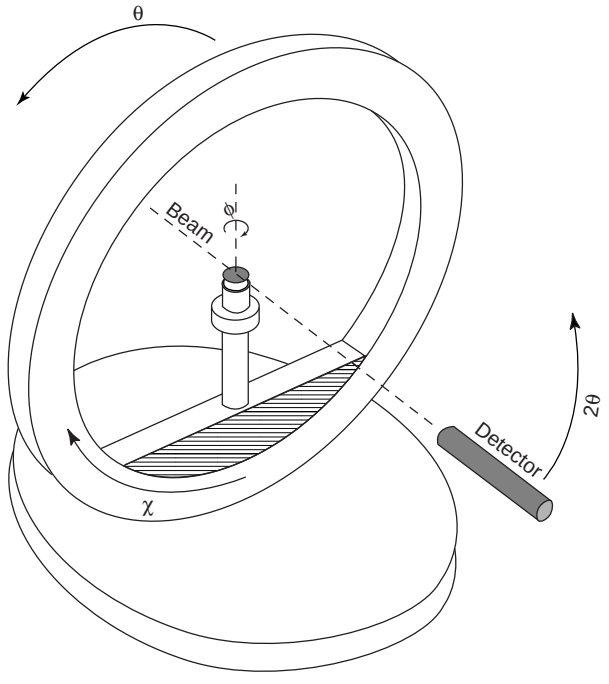

Figure 14 Sketch of a four-circle diffractometer.

to define a position. The degeneracy is solved using an additional constrain in the angle position. For normal operational conditions the 2θ angle is set to be equal to twice the value of θ. This is equivalent to have incident and scattered angles equal. CTR equations were derived in such conditions.

3.4.2 Electrochemical Cell

The cell used in scattering measurements is essentially the same as used in the EXAFS experiments for smooth electrodes (Figure 9). This cell presents the drawback of the thin layer configuration, which makes obtaining kinetic data very difficult. To solve that problem, different cells have been devised and used that do not require a thin layer configuration using a transmission geometry (Figure 15).[57–59] In these cells, a column of solution surrounded by a polymer film is maintained over the electrode. The counter and reference electrode are then positioned over the working electrode, allowing a much better current distribution over the working electrode. However, the thick column of electrolyte on top of the electrode surface increases the X-ray absorption by the electrolyte and requires the use of the brightest synchrotron sources.

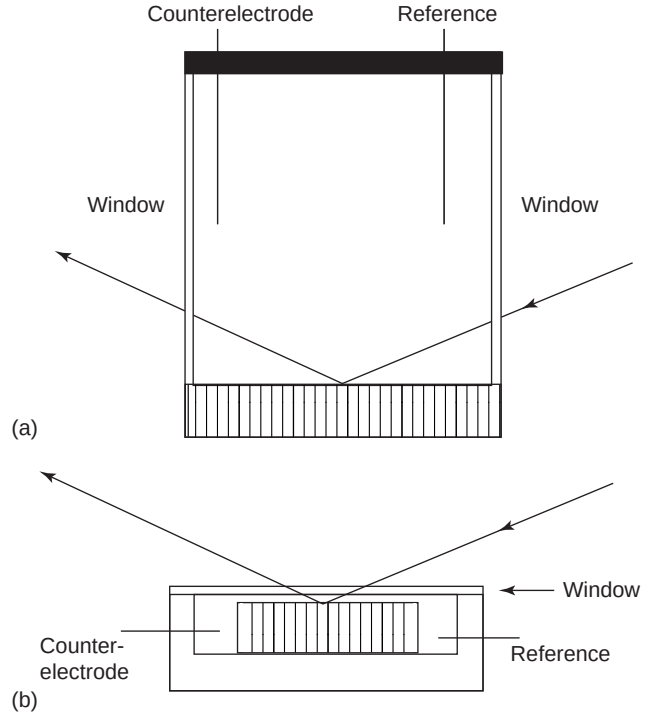

Figure 15 Scheme of the electrochemical cell used in X-ray scattering measurements using (a) transmission geometry and (b) reflection geometry.

3.4.3 Grazing Incident X-ray Diffraction Geometry

The determination of the in-plane structure of a surface requires the use of surface diffraction techniques. In order to limit the penetration of the X-rays into matter, the incident angle is kept small. With small incident angles, the scattering vector lies predominantly in the surface plane, probing the in-plane structure of the surface or interphase. Grazing incident geometry has several advantages over other possible geometries, such as an increase in the resolution of the surface peaks and a reduction in the background scattering from the bulk crystal.

3.4.4 Diffraction from Overlayers and Surfaces

The case of a crystal ideally terminated is seldom found in nature. The electronic deficiency caused in the topmost layer because of truncation favors the appearance of surface relaxation and reconstruction phenomena. In the simplest case, surface relaxation, the d spacing of the surface layer is altered from that found in the bulk material (maintaining the same lattice in the surface layer as in the bulk layers). Using grazing angle geometry (low Q_c), diffraction peaks for the surface can be found at the positions where the CTR are found. As explained in section 3.1, these peaks will correspond to positions where h and k are integers. The surface relaxation can be measured from CTR studies.

When there is a surface reconstruction process, the lattice of the surface layer will be different from that found in the rest of the crystal layers. For that reason, additional diffraction peaks will appear which correspond to the diffraction from the surface. These new diffraction peaks will appear at positions where h and k are not integers (known as noninteger peaks). For each additional noninteger peak, a new Bragg rod will be found, since the surface can be considered as an additional layer. As before, CTR measurements will give the d spacing of the surface layer, and will serve to confirm the diffraction structure proposed according to the diffraction measurements.

The presence of overlayers is similar to that found with reconstructed surfaces. The overlayer will give additional diffraction peaks, depending on its lattice. For each lattice peak, Bragg rods will appear. A nice example of the application of Bragg rods to the study of overlayers is found in the work by Toney et al. studying copper underpotential deposition (UPD) on Au(111) electrodes. The analysis of the intensity of Bragg rods served to discriminate the real structure of the Cu UPD from the other structures proposed (see section 4.2.2).[60] As in the previous cases, CTR measurements will give additional information of the surface and overlayer structure.

4 APPLICATION OF X-RAY METHODS TO ELECTROCHEMISTRY

4.1 Surface Structure of Metal Electrodes

A logical first step in the interphase studies deals with the surface characterization, especially when single crystal electrodes are used. For single crystal electrodes three different phenomena may take place when the electrode potential is changed: surface reconstruction, expansion or contraction of the topmost layer and surface roughening. For surface reconstruction studies, the best technique is GIXD, since it allows an unequivocal characterization. For reconstructed surfaces, is logical to expect an interlayer distance different to that found in bulk material, especially in the cases where the atomic surface density is different to that observed in the bulk layer. CTR studies can be used to measure the interlayer distance, jointly with the atomic density of the surface layer. CTR studies can also be used to measure surface roughness.

4.1.1 Au(111) Electrodes

Unlike most of the (111) planes of the face-centered cubic (fcc) crystals, the Au(111) surface reconstructs. The studies of this electrode surface can be used as a model of the information that can be obtained from the system using X-ray techniques. The first point to determine is the nature of the reconstruction and its stability range.

In UHV environments, the reconstruction leads to the formation of a $(23 \times \sqrt{3})$ structure (also called herringbone reconstruction) with an increase in the atomic surface density of 4.4%. In electrochemical environments, there is an additional variable, the electrode potential, and the reconstruction can be sensitive to changes of potential. In this particular case, the reconstruction is found to be potential dependent.[61–63] Below a critical potential, which is always more negative than the potential of zero change (p.z.c.), the structure found in three different electrolytes (0.01 M NaF, NaCl and NaBr) is the same as that found in the UHV environments: the $(23 \times \sqrt{3})$ surface structure (Figure 16). Above this potential, the reconstruction is partially lifted to give a $(p \times \sqrt{3})$ structure with $23 < p < 30$. At potentials positive to the p.z.c., the reconstruction disappears and the (1×1) surface structure is found. The phase transition between the $(23 \times \sqrt{3})$ and the (1×1) structure is reversible, that is, at a given potential the surface structure obtained is the same, irrespective of the initial conditions.

The kinetics of the phase transition can be obtained following the time evolution of the peak intensity of the reconstructed surface. It was shown that the transition to the (1×1) phase is much faster than the formation of the reconstructed phase. Dependence in the rate of the phase transition with the anion present in the electrolyte was also observed. Chloride and bromide anions accelerate the transition,[63] whereas the presence of pyridine, 2,2′-bypyridine and uracil slows down the phase transition.[64]

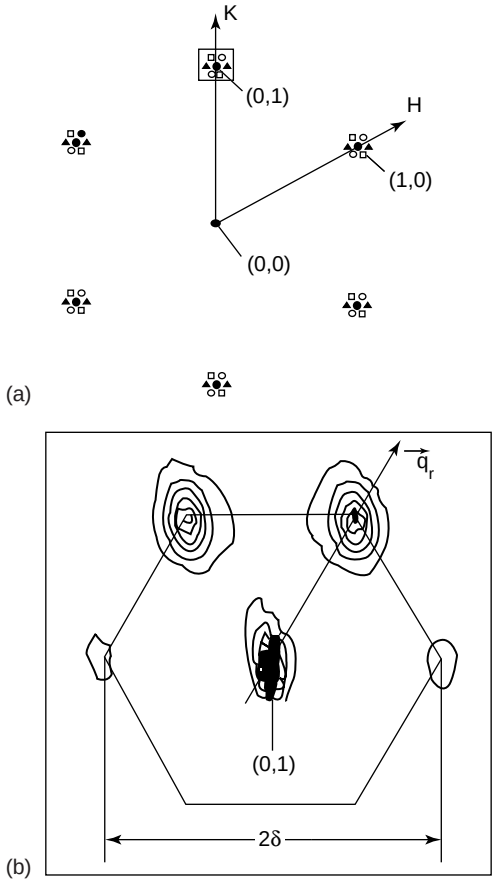

Figure 16 (a) In-plane diffraction pattern of the $(23 \times \sqrt{3})$ reconstruction in hexagonal coordinates. The solid circles are the periodicities from the underlying bulk substrate. (b) X-ray scattering equal-intensity contours in the vicinity of the (0,1) reflection in 0.01 M NaCl at −0.3 V vs Ag/AgCl. (Reproduced with permission from J. Wang, B.M. Ocko, A.J. Davenport, H.S. Isaacs, *Phys. Rev. B*, **46**, 10 321–10 338 (1992). © American Physical Society.)

Finally, the expansion or compression of the topmost layer was studied by CTR measurements. The presence of a reconstructed surface (which has a higher atomic density than the bulk layer) leads to an expansion of the interlayer distance for the reconstructed surface, i.e. the d spacing has increased 3.3% with respect to that found in the bulk material. No change in the distances was observed for the (1×1) structure.

4.1.2 Au(100) Electrodes

The general behavior of the Au(100) electrode is similar to that described for the Au(111) electrode.[65]

At negative potential with respect to the p.z.c., the surface layer exhibits an hexagonal reconstructed surface, whereas at positive potentials, the (1×1) structure is found. The hexagonal reconstruction has an atomic surface density 22% greater than that found in the bulk Au(100) layer. This reconstructed phase is the same that was found in UHV environments.[66,67] When the reconstruction is lifted at positive potentials these excess of atoms is segregated to an additional layer. This new layer has an atomic surface density that corresponds to 22% of the bulk layer. In perchloric acid, a partial reconstruction is observed at potentials negative to 0.4 V vs a standard calomel electrode (SCE).[57]

4.1.3 Pt(111) and Pt(100) Electrodes

No evidence of reconstruction has been found for these surfaces in electrochemical environments.[68–71] In these surfaces, the effect of the anion and hydrogen adsorption in the relaxation/contraction of the topmost layer and the effect of oxide formation has been studied.

4.2 Adsorbed Layers on Metal Surfaces

One of the most studied phenomena in surface electrochemistry in recent years is the UPD of metals. The UPD process is mainly controlled by the interaction between the depositing metal and the electrode surface. However, the anion present in the electrolyte also plays an important role in this process since the interaction of the anions with the electrode surface and with the depositing metal affects the whole UPD process. Therefore, a logical first step when studying UPD processes is to understand the process of anion adsorption on the bare surface.

4.2.1 Anion Adsorption on Single Crystal Electrodes

Studies of anion adsorption on single crystal electrodes have focused mainly on halide adsorption. These studies highlight the influence of the surface structure and nature on the electrode. A clear example of that is the adsorption of iodine on Au(111), Ag(111) and Pt(111) electrodes. For Au(111) and Ag(111) electrodes, iodine adsorption exhibits very complex behavior. The first structure observed is a $(\sqrt{3} \times \sqrt{3})$ R30° (only present in Ag(111) electrodes).[72] As the potential is made more positive, the structure changes to a uniaxial incommensurate structure, $c(p \times \sqrt{3})$, where p decreases with the potential.[72,73] Finally, the structure changes to a rotated incommensurate hexagonal phase. On the other hand, Pt(111) electrodes exhibit completely different behavior. Only two commensurate different phases are observed: (3×3) and a (7×7)R 21.8°.[72] Moreover, the I–I distance obtained is smaller in Pt that in Au or Ag, clearly the result of a stronger Pt–I interaction. The Au–I distance obtained by CTR measurements is in the range of 2.28 to 2.45 Å, depending on the iodine surface structure.[73] XSW studies performed with LSM materials showed changes in the diffuse layer of iodide.[15,74] At potentials for which the iodine adlayer was not saturated, the diffuse adlayer showed an accumulation of iodide associated to the adsorbed iodine. Upon saturation of the iodine adsorbed layer, there is a depletion of iodine in the diffuse layer.

Bromide adsorption on Au(111) only shows a rotated hexagonal structure that compresses when the potential is scanned in the positive direction,[72,75,76] similar to what is observed for chloride adsorption.[75] The structure of bromide on a Pt(111) electrode is also a hexagonal structure that compresses, but in this case the structure is aligned with the (1,0) surface direction.[77] Commensurate structures are found only on Ag(111) electrodes with a (7×7)R 21.8° structure.[72] Ordered structures and the phase transition between them are also observed on Au(100) and Ag(100) electrodes.[78–80]

4.2.2 Copper Underpotential Deposition on Gold Single Crystal Electrodes

The cyclic voltammogram of Cu UPD on Au(111) shows two well-defined and sharp pairs of peaks, corresponding to two different adsorption–desorption processes. The pair of peaks (deposition/stripping) at high underpotentials has a broad shoulder, whereas the deposition peak at low underpotentials splits in two on high-quality Au(111) single crystals.[81,82] For these reasons, two different surface structures have to be expected after each adsorption peak.

The first in situ measurements on this system were done by SEXAFS, which show the existence of a Cu-(1×1) structure on the electrode surface after the second deposition peak.[83,84] SEXAFS measurements provided an accurate determination of the Cu–Cu distance in this structure (2.92 ± 0.03 Å,[84] 2.89 ± 0.03 Å[85]). This value is the same as the Au–Au distance in the (111) plane (2.88 Å), implying that a Cu-(1×1) adlayer is commensurate with the Au(111) substrate.[84,85] In this structure, Cu adatoms probably sit on threefold hollow sites, as derived from the polarization dependence of the SEXAFS[84] or by estimating the number of gold nearest neighbors.[85] SEXAFS spectra also indicate the existence of oxygen (likely arising) from (bi)sulfate on the Cu adlayer in which Cu presents a +1 oxidation state, as suggested by XANES.[83,85]

The determination of the structure after the first UPD peak was more difficult and was, for some time,

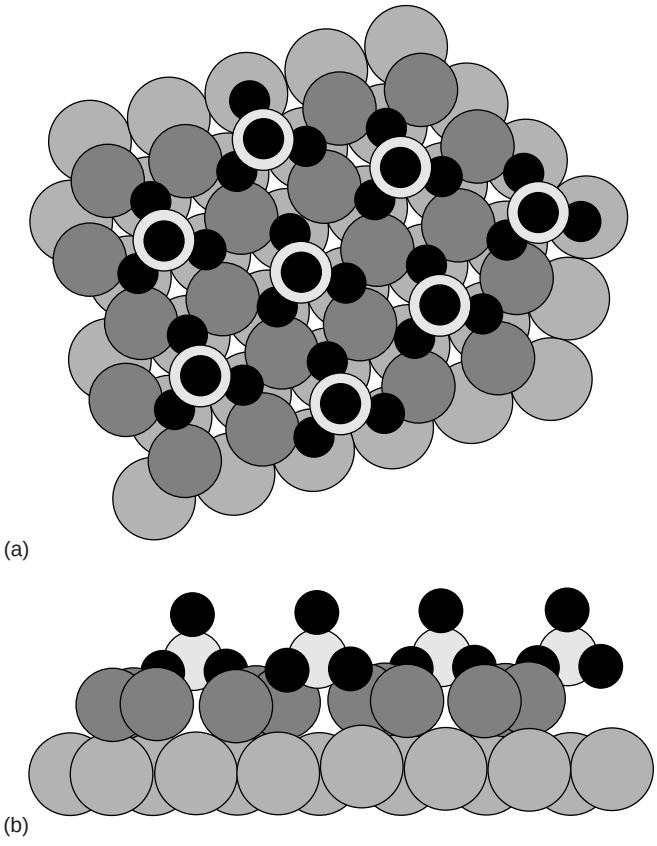

Figure 17 Surface structure of the $(\sqrt{3} \times \sqrt{3})$ honeycomb structure found for Cu UPD on Au(111) electrodes in sulfuric acid media. (a) Top view (b) side view. (Reproduced with permission from M.F. Toney, J.N. Howard, J. Richer, G.L. Borges, J.G. Gordon, O.R. Melroy, *Phys. Rev. Lett.*, **75**, 4772–4775 (1995). © American Physical Society.)

controversial. The first SPM[82,86–89] images showed a $(\sqrt{3} \times \sqrt{3})$R30° structure in which the maxima were interpreted as Cu adatoms. Based on this structure the total Cu coverage was 0.33. The $\Theta_{Cu} = 0.33$ was in clear disagreement with the results from chronocoulometry and quartz crystal microbalance (QCM), which indicated that the Cu coverage in the adlayer at these potentials was 0.67 and the (bi)sulfate coverage value was 0.33.[90–95] Also theoretical calculations predicted the formation of a honeycomb $(\sqrt{3} \times \sqrt{3})$R30° structure,[96] whose coverages were in agreement with those found with QCM,[94,95] and in UHV.[97,98] This apparent contradiction was reconciled by Toney et al. using X-ray CTR measurements.[60] They found that the Cu adlayer, at intermediate coverage values, had a honeycomb $(\sqrt{3} \times \sqrt{3})$R30° structure in which $\Theta_{Cu} = 0.67$ and with (bi)sulfate anions occupying the centers of the honeycomb ($\Theta_{anion} = 0.33$) in agreement with the coverage values found with chronocoulometry and QCM (Figure 17). This meant that the scanning probe techniques (STM, AFM) were not imaging the Cu adatoms but the (bi)sulfate anions that occupied the centers of the honeycomb and protruded well above the Cu plane.[60] Recent SEXAFS measurements also support the honeycomb model (Figure 18).[39] As was the case for the Cu-(1×1) structure, XANES clearly suggested that the oxidation state of the Cu adatoms after the first UPD peak is approximately +1,[99] again probably as a result of a polar bond or surface dipoles. For this structure, it has been proposed that the Cu adatoms occupy on-top positions.[100]

The voltammogram of Cu UPD in presence of chloride shows two pairs of peaks, as in the case of sulfuric acid, but the pair at high underpotentials is reversible and appears at potentials that are more positive than in sulfuric acid. The second pair of peaks almost overlaps with bulk Cu deposition. After the first peak, a bilayer structure was proposed after SEXAFS measurements, in which chloride is adsorbed on the copper ($\Theta_{Cu} = 0.62$) and the copper atoms are in registry with the top layer of chloride ions.[39,101] The distances in the proposed structure are quite similar to those found in solid CuCl.[102]

In the presence of bromide, the behavior is very similar to that found in the presence of chloride, with a predicted bilayer structure.[92,103,104] Recent X-ray scattering studies revealed the presence of an ordered hexagonal bromide adlayer at the onset of copper deposition. This adlayer undergoes a phase transition to form a (4×4) commensurate structure ($\Theta_{Br} = 0.56$) at the peak at 0.32 V vs Ag/AgCl. The bromide adlayer remains stable until bulk copper deposition.[104] Copper is deposited in-between the gold surface and the bromide layer.[104] As a general trend for all Cu–halide adlayers, the structure of these adlayers is governed by the halide–halide and the Cu–halide interactions in contrast to the Cu–(bi)sulfate adlayers in which the Cu–Au interaction is dominant. However, all interactions contribute to the UPD processes.

On Au(100) electrodes in sulfuric acid media, Cu UPD occurs in a relatively broad peak.[105] STM studies showed that the surface structure after the deposition peak is a pseudomorphic (1×1) structure in which the Cu atoms occupy the fourfold hollow site.[105] XSW measurements also indicated that the Cu adatoms occupy the fourfold hollow site.[106] However, SEXAFS measurements suggested a different picture of the adlayer at full coverage.[99] In the model proposed according to the experimental data (which is incompatible with STM and AFM images), the Cu–Cu distance is shorter than the Au–Au distance in the unreconstructed Au(100) surface and the Cu adatoms occupy the a-top sites.[99,107] This would imply that the topmost gold layer is rearranged or reconstructed. As on Au(111) electrodes, the Cu oxidation state is close to +1.[99,107]

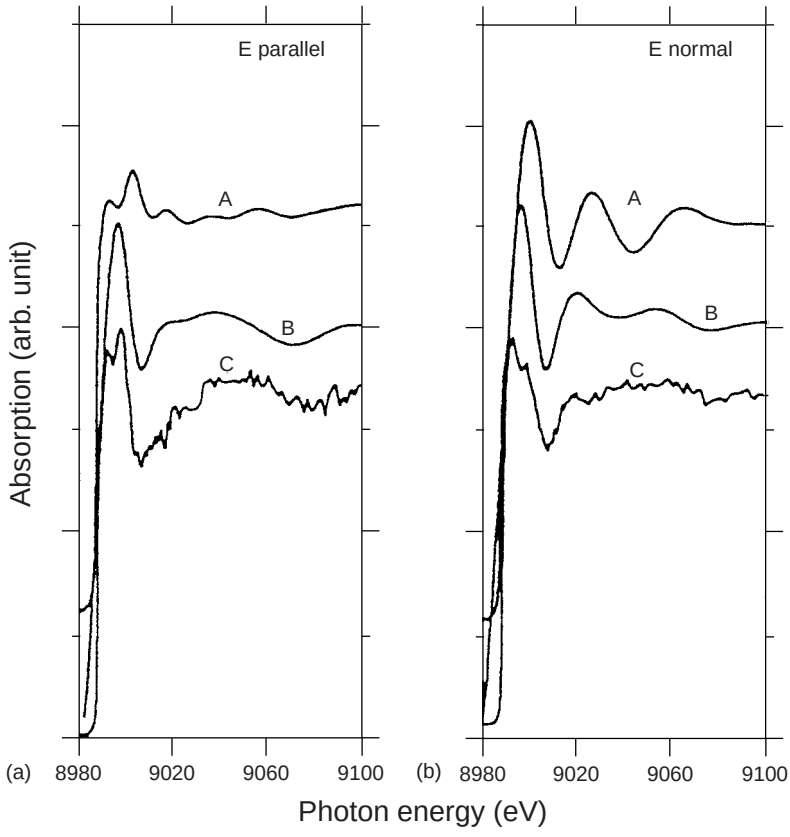

Figure 18 Comparison of (A) and (B) calculated FEFF6.0 and (C) experimental Cu XAFS spectra for the Cu-($\sqrt{3} \times \sqrt{3}$) honeycomb structure on Au(111) electrodes. The sulfate has one oxygen atom pointing to the surface in model (a) and three oxygen atoms in model (b). (Reproduced from S. Wu, J. Lipkowski, T. Tyliszczak, A.P. Hitchcock, 'Effect of Anion Adsorption on Early Stages of Copper Electrocrystallization at Au(111) Surface', *Prog. Surf. Sci.*, **50**, 227–236. © 1995 with permission from Elsevier Science.)

4.2.3 Lead Underpotential Deposition on Au(111) Electrodes

Lead and gold have very different atomic sizes (lead is ca. 20% larger than gold), which favors the formation of incommensurate adlayers. In fact, at potentials negative of the main UPD peaks, a hexagonal incommensurate structure has been found using X-ray diffraction techniques.[108,109] This structure was also observed by SPM.[110–113] In this structure, the lead–lead distances are compressed 0.7% with respect to bulk lead.[108,109] The lead adlayer is rotated with respect to the Au(111) plane with an angle which varies between 2.5° and 0° depending on the applied potential (Figure 19),[109] which is consistent with the change in lead–gold distance over the same potential region observed with another X-ray diffraction technique.[114]

4.2.4 Mercury Underpotential Deposition on Au(111) Electrodes

The behavior of the mercury UPD on Au(111) electrodes resembles in some cases that of copper UPD on the same electrodes. The voltammetric profile of mercury UPD on Au(111) electrodes in 0.1 M H_2SO_4 shows several pairs of peaks. The voltammogram shows a main deposition/stripping peak (centered at around +0.90 V vs Ag/AgCl) that splits into two. The initial stages of mercury UPD appear to trigger an order–disorder transition in the (bi)sulfate adlayer that gives rise to the first pair of peaks.[115] After the second pair of peaks an ordered surface structure is found. STM measurements have shown the existence of two different ordered structures after this peak:[116]

$$\begin{pmatrix} 2 & 0 \\ 3 & 2/3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & \bar{1} \\ 4 & 4 \end{pmatrix}$$

The first structure was later observed by in situ X-ray diffraction and identified as a ($\sqrt{3} \times \sqrt{19}$) surface structure.[117] CTR measurements indicated that the adlayer is probably constituted by Hg_2^{2+} cations and (bi)sulfate anions.[117] Based on the atomic distances derived from the CTR measurements, an adlayer structure similar (although distorted) to the honeycomb

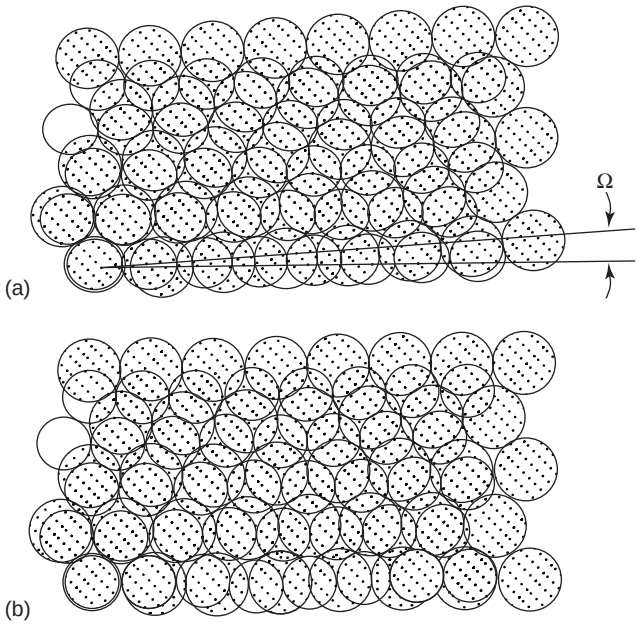

Figure 19 Surface structure of the Pb layer on Au(111) electrodes at (a) E < 0.13 V and (b) E > 0.16 V vs Ag/AgCl in 0.1 M $HClO_4$. (Reprinted with permission from M.F. Toney, J.G. Gordon, M.G. Sammant, G.L. Borges, O.R. Melroy, D. Yee, L.B. Sorensen, *J. Phys. Chem.*, **99**, 4733–4744 (1995).)

structure observed for copper UPD in the same medium was proposed.[117] At +0.82 V vs Ag/AgCl, the ordered structure disappears.[117]

Apart from the surface processes of mercury UPD, a process controlled by diffusion also appears in the voltammogram at potentials around +0.54 V vs Ag/AgCl. This redox reaction corresponds to the oxidation–reduction of mercury species in solution according to Equation (41):

$$2Hg^{2+} + 2e^- \longleftrightarrow Hg_2^{2+} \quad E^\circ = +0.698\,\text{V vs Ag/AgCl} \tag{41}$$

Coinciding with this diffusion-controlled process other additional ordered structures have been found. AFM measurements have identified a hexagonal structure at these potentials.[118,119] X-ray diffraction studies identified two different hexagonal adlayers in this region: one at potentials between +0.63 and +0.68 V (vs Ag/AgCl) with a Hg–Hg distance of 3.84 Å, and another at potentials below +0.63 V (vs Ag/AgCl) with a Hg–Hg distance of 3.33 Å.[120] The first hexagonal adlayer appears to be metastable and evolves, with time, to give the second adlayer.[120]

The effect of anions is also evident in the case of mercury UPD. Unlike copper UPD, the voltammetric profile for mercury UPD on Au(111) in the presence of chloride resembles that observed in sulfuric acid media alone.[115] This would indicate that in both chloride and sulfuric acid media the UPD process is governed by the mercury–substrate interactions. After the main UPD peaks, CTR measurements showed that a Hg_2Cl_2 bilayer is formed,[121] in which mercury is bonded to the gold surface and chloride is deposited on top of the mercury adatoms. In acetic acid media, hexagonal structures have been observed with X-ray diffraction after the main UPD peak.[121]

4.2.5 Copper Underpotential Deposition on Pt(111) Electrodes

EXAFS studies have shown that the copper layer in sulfuric acid media is not completely discharged on the surface at potentials after the UPD peak.[122,123] The charge of the copper species on the Pt(111) surface is close to +1 and not zero as expected based on complete charge transfer, while the platinum surface retains some negative charge, as in the case of Au(111) electrodes.[83,85] At full coverage, EXAFS results show that the copper layer has a close-packed structure with a copper–copper distance of 2.77 ± 0.03 Å. This structure corresponds to the copper residing in the threefold hollow sites of the Pt(111) surface.[122] GIXD and CTR measurement have determined that the UPD process of copper on Pt(111) electrodes is very similar to the one taking place on Au(111) electrodes.[124] The first step in the UPD process is the formation of the honeycomb structure, followed by the formation of a full copper monolayer with (bi)sulfate anions adsorbed on top of the copper layer.

Bromide and chloride anions also show a dramatic effect upon copper UPD onto platinum single crystals. In both media, the deposition and stripping again take place in two distinct steps on Pt(111),[125–128] in which copper retains a partial charge.[129] In chloride media, GIXD measurements showed that the adlayer presents a Bragg rod at the (0.765, 0, L) position after the first UPD peak, which corresponds to the formation of an incommensurate hexagonal CuCl adlayer (Figure 20).[125,126] After the second UPD peak, a commensurate (1 × 1) copper adlayer forms on the electrode surface, which, in turn, is covered by a disordered layer of chloride anions.[127] In this adlayer, SEXAFS findings also indicated that in the presence of chloride, the copper–copper bond distance of the deposited layer was close to that of bulk copper, unlike that of copper deposited in the absence of chloride.[130] The kinetics of the phase transitions between the incommensurate and the (1 × 1) structures were followed by GIXD.[131,132] These studies revealed that the transition from the (1 × 1) structure to the incommensurate structure occurs through a two-step mechanism. In the first step, the copper is desorbed which gives rise to the current response. On a much longer timescale, the reorganization of the layer takes place to give an ordered CuCl bilayer.

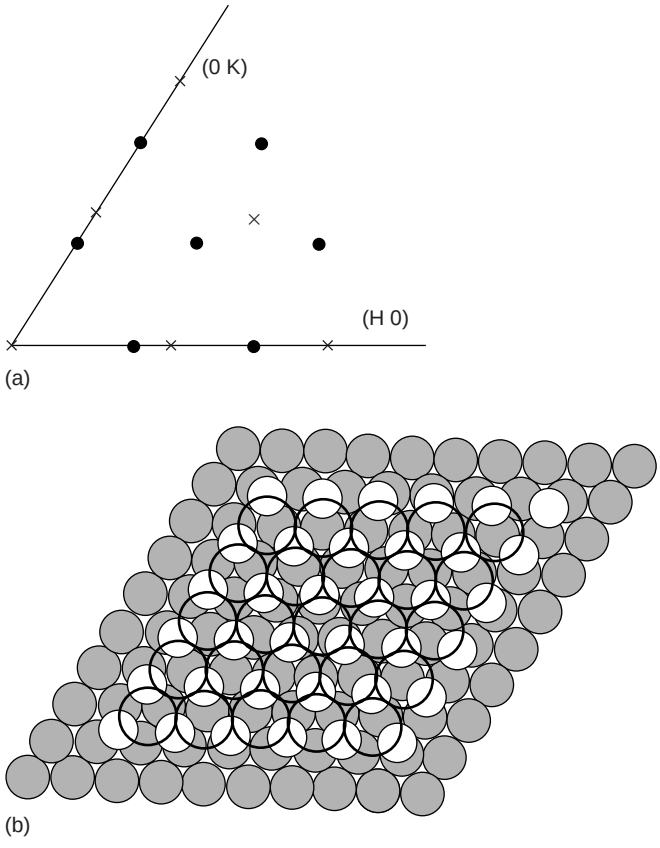

Figure 20 (a) Schematic representation of the in-plane reciprocal space in the potential region between the two main peaks in the Cu UPD in presence of chloride anions. (b) Surface structure of the adlayer. Pt atoms: shadowed circles; Cu atoms: white circles and Cl atoms: open circles. (Reproduced from N.M. Markovic, H.A. Gasteiger, C.A. Lucas, I.M. Tidswell, P.N. Ross, 'The Effect of Chloride on the Underpotential Deposition of Copper on Pt(111): AES, LEED, RRDE and X-ray Scattering Studies', *Surf. Sci.*, **335**, 91–100. © 1995 with permission from Elsevier Science.)

In bromide medium, an incommensurate hexagonal structure aligned along the (1,0) surface direction is observed by GIXD after the first UPD peak, corresponding to a CuBr layer with an interatomic distance of 3.74 Å.[125,126,133] At potentials negative of the second UPD peak, it has been proposed that copper forms a (1×1) structure, which is, in turn, covered by a disordered bromide layer.

4.2.6 Lead Underpotential Deposition on Ag(111) Electrodes

Attempts to elucidate the structure of lead adsorbed on Ag(111) have been done using surface X-ray scattering techniques.[108,109,134–137] Deposition from either perchloric acid or acetate media shows the presence of peaks in the azimuthal scans (varying ϕ angle) at $\phi = \pm 4.5°$.[109,134,135] This marks the presence of two equivalent domains of an incommensurate lead overlayer that are each $\sim 4.5°$ from the Ag[011] direction as confirmed in the STM measurements.[138–140] GIXD measurements have also shown that lead deposited onto Ag(111) at full monolayer coverage undergoes a compression of 1.4% relative to bulk lead.[134–136] The compression increases linearly with applied potential until the onset of bulk deposition where the compression is 2.8%. X-ray studies done employing surface differential diffraction have found that the distances measured for the lead layer were between 3.00 ± 0.05 and 3.2 ± 0.1 Å.[137] These distances can be ascribed to lead atoms being adsorbed between the a-top and the bridge sites.

In situ EXAFS has been performed in acetate media.[141] These studies have found that the deposited lead is in a zero-valent state and that the lead layer is deposited incommensurate to the silver substrate. There is also scattering observed from an oxygen atom, which is most likely coming from the water or acetate. This scattering implies that these molecules are adsorbed on the deposited lead layer.

4.2.7 Thallium Underpotential Deposition on Silver Single Crystal Electrodes

GIXD studies of thallium UPD on Ag(111) have shown that after complete deposition of one monolayer, there is an incommensurate, hexagonal structure on the surface.[109,142,143] This structure is compressed relative to bulk thallium by 1.4–3.0% and rotated from the Ag[01$\bar{1}$] direction by $\Omega = 4–5°$, depending on potential.[109,142,143] Vapor deposited thallium in UHV presents the same structure, suggesting that the solvent molecules do not affect the deposition structure.[143] Upon completion of deposition of the second monolayer, the thallium forms a bilayer that has also an incommensurate hexagonal structure.[109,142,143] The second layer is commensurate with the layer beneath it and the newly deposited thallium atoms sit in the threefold hollow sites of the bottom layer.[143] In this structure, the compression is 1.0% relative to bulk thallium and the rotation is 3.9°.[142,143]

Thallium deposition onto Ag(100) has also been studied by surface X-ray scattering.[144] A disordered phase is formed after the first voltammetric peak, whereas an ordered monolayer is obtained after the second voltammetric peak. This layer has a $c(p \times 2)$ close-packed structure which compresses uniaxially (p decreasing from 1.185 to 1.168) with decreasing potential.[144] With deposition of the second layer, it has been found that the first layer expands slightly and both layers form a $c(1.2 \times 2)$ bilayer.

4.3 Fuel Cell Electrodes

Most fuel cells use platinum as an electrode material (especially for the hydrogen fuel cell and the direct methanol fuel cell). The high cost of platinum makes the dispersion of the metal on a supporting material such as carbon advisable in practical applications. On some occasions, platinum has been alloyed in order to increase its catalytic activity towards the desired reaction.

The dispersion of platinum creates small particles that can have different electrocatalytic properties than the bulk material. The first studies carried out confirmed that the platinum oxide formed at positive potentials has a short-range structure similar to α-PtO_2.[145] Alloying the platinum resulted in a diminution in the *d*-band occupancy with respect to the platinum particles and a contraction in the Pt–Pt distances.[146] All the alloys studied showed an increase in the catalytic activity for the oxygen reduction except the PtMn/C alloy. A volcano type correlation was found between the Pt–Pt distance and the *d* band occupancy with the electrocatalytic activity for oxygen reduction. In contrast to the oxygen reduction, the electrocatalytic oxidation of hydrogen shows no significant differences between Pt and the Pt alloys, indicating that the structural parameters and *d* band occupancies have no effect on the reaction.[147] For Pt particles, structural changes are observed upon hydrogen adsorption, whereas the alloy particles exhibit no changes.[147,148] These changes in the Pt–Pt distance and in the Pt coordination number have been attributed to a reversible reconstruction that takes place upon hydrogen adsorption.[148,149]

The effect of the Pt particle size has been investigated in relation to methanol oxidation. The EXAFS and XANES results show that the reduced activity towards methanol oxidation is the consequence of a stronger CO and OH adsorption in the small particles.[149] The strong OH adsorption also results in a lower activity towards oxygen reduction. The effect of particle size has also been studied for Cu and Pb UPD on carbon-supported platinum.[150–152]

Pt–Ru particles exhibit an increase in the catalytic activity towards methanol oxidation due mainly to a bifunctional mechanism in which the Ru atoms facilitated the removal of the CO species. This effect has been confirmed by XANES, which shows no change in the oxidation state of the Ru atoms when CO is present in the solution for potentials at which the oxidation of ruthenium takes place in absence of CO.[153] The presence of Ru in the alloy diminishes the Pt–Pt distances and diminished the *d* band occupancy, a clear indication that the Ru has also an electronic effect on the platinum particles.[154] The formation of RuOH species in the alloy takes place at 0.24 V vs SCE.[154]

4.4 Battery and Oxide Electrodes

Oxidation of $Ni(OH)_2$ has been extensively studied by EXAFS and XANES.[155–161] The uncharged material (β-$Ni(OH)_2$) shows changes in the (0001) plane of the $Ni(OH)_2$ with charging–discharging cycles.[157] The charge material corresponds to a nickel oxide in an oxidation state of ca. +3.5, determined by using the edge position and the distances and coordination numbers of the oxygen atoms.[159] The structure of the charged material showed two different Ni–O interactions at 1.88 and 2.07 Å and with coordination numbers of 4.1 and 2.2 respectively, and two Ni–Ni interactions at 2.82 and 3.13 Å with coordination numbers of 4.7 and 1.0 respectively. The longer Ni–O interactions were interpreted as Ni–OH distances and the shorter were attributed to Ni–O in tetravalent nickel compounds. The same oxidation state has been obtained by simulating the near edge part of the spectrum with the FEFF6 code program.[160] Oxidation of α-$Ni(OH)_2$ also gives nickel oxides with an oxidation state of 3.5.[161]

Additional studies have been conducted with V_2O_5 electrodes,[162] CuO_2 electrodes,[163] PbO_2[164] and $Li_xMn_2O_4$.[165]

4.5 Corrosion

The nature of the passive film formed on iron electrodes has been investigated with in situ XANES in borate buffer solution. Comparison with model compounds indicates that the structure of the oxide passive film is γ-Fe_2O_3/Fe_3O_4.[50] Diffraction studies of the passive film have been used to discriminate between the two possible structures.[166] Calculating the theoretical structure factors for different structures and comparing them with those obtained from diffraction measurements, allowed the identification of the structure of the passive film. The film structure is based on the structure of Fe_3O_4, but with cation vacancies on the tetrahedral sites and octahedral sites and cations occupying the interstitial octahedral sites. The oxidation state of iron in the passive film is mainly Fe^{3+}, with a Fe^{2+} proportion that oscillates between 4 and 20% depending on the applied potential.[167] When the passive film was obtained at potentials below −0.6 V vs mercury sulfate electrode (MSE), dissolution is observed during the first stages of the film formation.[167] The cathodic dissolution of the passive films was studied using Fe_2O_3 and Fe_3O_4 films.[168] The reduction of the Fe_2O_3 film takes place in a two-step mechanism. In the first step, the Fe_2O_3 is reduced to Fe_3O_4 with partial dissolution. In the second step, the formed Fe_3O_4 film is reduced to Fe^{2+}. The dissolution of the passive film depends on the solubility of the Fe^{2+} species. Thus, in basic media no dissolution is observed, owing to the insolubility of Fe(II) species. In borate buffer, dissolution is only observed

during step 2. The anions also affect the dissolution rate of the passive film. During anodic dissolution, chloride anions accelerate the dissolution when compared to sulfate anions, as a result of the complexation of Fe cations by chloride anions.[169] Cathodic dissolution does not show any dependence on the anions. The effects of the presence of Cr and other additives to iron have also been studied.[170–172] These studies showed the presence of Cr(IV) at positive potentials.

The addition of chromium confers corrosion resistance to passive films. For that reason, it has been used extensively in aluminum alloys. The XANES studies indicate the possibility of oxidizing the initial Cr(III) to Cr(VI).[49,173,174] Depending on the experimental conditions, the Cr(VI) species are stable on the passive film and can be reduced to the initial oxidation state. The Cr(VI) species accumulate on the outer region of the passive film. Owing to the toxicity of chromium, cerium oxides have been used as a possible substitute in passive films. XANES studies indicate that cerium is in an oxidation state of +3.[175]

4.6 Other Examples

Several studies have been conducted on polymer electrodes modified with several salts[176–181] and modified electrodes.[182–185] These studies focus mainly in the determination by EXAFS of the short-range structure around the metallic atom in the polymer or modified electrodes.

ACKNOWLEDGMENTS

The author is indebted to Dr J.M. Feliu for critical review of the manuscript. This work is partially supported by the Ministry of Education and Culture (Spain), grant no. PB96-0409.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
CTR	Crystal Truncation Rod
EXAFS	Extended X-ray Absorption Fine Structure
fcc	Face-centered Cubic
FTIRS	Fourier Transform Infrared Spectroscopy
fwhm	Full Width at Half-maximum
GIXD	Grazing Incident X-ray Diffraction
LEED	Low-energy Electron Diffraction
LSM	Layered Synthetic Microstructures
MSE	Mercury Sulfate Electrode
p.z.c.	Potential of Zero Change
QCM	Quartz Crystal Microbalance
SCE	Standard Calomel Electrode
SEXAFS	Surface Extended X-ray Absorption Fine Structure
SPM	Scanning Probe Microscopy
STM	Scanning Tunneling Microscopy
UHV	Ultrahigh Vacuum
UPD	Underpotential Deposition
XANES	X-ray Absorption Near Edge Structure
XPS	X-ray Photoelectron Spectroscopy
XSW	X-ray Standing Waves

RELATED ARTICLES

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Infrared Spectroelectrochemistry • Microbalance, Electrochemical Quartz Crystal • Scanning Tunneling Microscopy, In Situ, Electrochemical • Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Absorption Techniques in X-ray Spectrometry • Structure Determination, X-ray Diffraction for

REFERENCES

1. J.O'M. Bockris, B.E. Conway, E. Yeager, *Comprehensive Treatise of Electrochemistry*, Plenum Press, New York, Vol. 1, 1980.
2. S. Pons, 'The Use of Fourier Transform Infrared Spectroscopy for In Situ Recording of Species in the Electrode–Electrolyte Interphase', *J. Electroanal. Chem.*, **150**, 495–504 (1983).
3. A. Bewick, 'In Situ Infrared Spectroscopy of the Electrode/Electrolyte Interphase', *J. Electroanal. Chem.*, **150**, 481–483 (1983).
4. M. Fleischmann, P.J. Hendra, A.J. McQuillan, 'Raman Spectra of Pyridine Adsorbed at a Silver Electrode', *Chem. Phys. Lett.*, **26**, 163–166 (1974).
5. M. Fleischmann, P.J. Hendra, A.J. McQuillan, R.L. Paul, E.S. Reid, 'Raman Spectroscopy at the Electrode–Electrolyte Interface', *J. Raman. Spectrosc.*, **4**, 269–274 (1976).
6. M. Fleischmann, P.J. Hendra, A.J. McQuillan, A. James, 'Raman Spectra from Electrode Surfaces', *J. Chem. Soc. Chem. Commun.*, 80–81 (1973).

7. D.J. Jeanmarie, R.P. Van Duyne, 'Surface Raman Spectrum Electrochemistry. Part I Heterocyclic, Aromatic and Aliphatic Amines Adsorbed on the Anodized Silver Electrode', *J. Electroanal. Chem.*, **84**, 1–20 (1977).
8. G. Richmond, 'Investigations of Electrochemical Interfaces by Nonlinear Optical Methods', in *Electrochemical Interfaces*, ed. H.D. Abruña, VCH Publishers, New York, 267–331, 1991.
9. G. Richmond, G.M. Robinson, V.L. Shannon, 'Second Harmonic Generation Studies of Interfacial Structure and Dynamics', *Prog. Surf. Sci.*, **28**, 1–20 (1988).
10. A.A. Gewirth, B.K. Niece, 'Electrochemical Applications of In Situ Scanning Probe Microscopy', *Chem. Rev.*, **97**, 1129–1162 (1997).
11. H.D. Abruña, 'X-rays as Probes of Electrochemical Interfaces', in *Modern Aspects of Electrochemistry*, eds. J. O'M. Bockris, R.E. White, B.E. Conway, Plenum Press, New York, 265–326, Vol. 20, 1989.
12. H.D. Abruña, 'Probing Electrochemical Interfaces with X-rays', *Adv. Chem. Phys.*, **77**, 255–335 (1990).
13. J. McBreen, 'In Situ Synchrotron Techniques in Electrochemistry', in *Physical Electrochemistry*, ed. I. Rubistein, Marcel Dekker, New York, 339–391, 1994.
14. L.E. Sharpe, W.R. Heineman, R.C. Elder, 'EXAFS Spectroelectrochemistry', *Chem. Rev.*, **90**, 705–722 (1990).
15. H.D. Abruña, G.M. Bommarito, D. Acevedo, 'The Study of Solid/Liquid Interfaces with X-ray Standing Waves', *Science*, **250**, 69–74 (1990).
16. M.F. Toney, O.R. Melroy, 'Surface X-ray Scattering', in *Electrochemical Interfaces: Modern Techniques for In Situ Interface Characterization*, ed. H.D. Abruña, VCH Verlag Chemical, Berlin, 55–129, 1991.
17. B.M. Ocko, 'X-ray Reflectivity and Surface Roughness', in *Spectroscopic and Diffraction Techniques in Interfacial Electrochemistry*, eds. C. Gutiérrez, C. Melendres, Kluwer Academic Publishers, Dordrecht, 343–359, 1990.
18. J. Robinson, 'X-ray Diffraction at the Electrode–Solution Interface', in *Spectroscopic and Diffraction Techniques in Interfacial Electrochemistry*, eds. C. Gutiérrez, C. Melendres, Kluwer Academic Publishers, Dordrecht, 313–341, 1990.
19. H. You, Z. Nagy, 'Application of X-ray Scattering Techniques for the Study of Electrochemical Interphases', *Current Topics in Electrochemistry*, **2**, 21–43 (1993).
20. K. Wilke, 'Synchrotron Radiation Sources', *Rep. Prog. Phys.*, **54**, 1005–1068 (1991).
21. J.A. Ibers, W.C. Hamilton, *International Tables for X-ray Crystallography*, Kynoch, Birmingham, 1974.
22. R. de L. Kronig, 'Theory of Fine Structure in the X-ray Absorption Spectra', *Z. Phys.*, **70**, 317–323 (1931).
23. R. de L. Kronig, 'Theory of Fine Structure in the X-ray Absorption Spectra-II', *Z. Phys.*, **75**, 191–210 (1932).
24. D.E. Sayers, E.A. Stern, F.W. Lytle, 'New Technique for Investigating Non-crystalline Structures: Fourier Analysis of the Extended X-ray Absorption Fine Structure', *Phys. Rev. Lett.*, **27**, 1204–1207 (1971).
25. E.A. Stern, 'Theory of the Extended X-ray Absorption Fine Structure', *Phys. Rev. B*, **10**, 3027–3037 (1974).
26. E.A. Stern, D.E. Sayers, F.W. Lytle, 'Extended X-ray Absorption-fine-structure Technique. II. Determination of Physical Parameters', *Phys. Rev. B*, **11**, 4836–4846 (1975).
27. C.A. Ashley, S. Doniach, 'Theory of Extended X-ray Absorption Edge Fine Structure Spectroscopy', *Phys. Rev. B*, **11**, 1279–1290 (1975).
28. P.A. Lee, J.B. Pendry, 'Theory of the Extended X-ray Absorption Fine Structure', *Phys. Rev. B*, **11**, 2795–2811 (1975).
29. P.A. Lee, G. Beni, 'New Method for the Calculation of Atomic Phase Shifts: Application to Extended X-ray Absorption Fine Structure (EXAFS) in Molecules and Crystals', *Phys. Rev. B*, **15**, 2862–2883 (1977).
30. A. Bianconi, 'XANES Spectroscopy', in *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*, John Wiley & Sons, New York, 573–625, 1988.
31. J. Stöhr, 'NEXAFS Spectroscopy', in *Springer Series in Surface Science*, eds. G. Ertl, R. Gomer, D.L. Mills, H.K.V. Lotsch, Springer, Berlin, Vol. 25, 1992.
32. B.K. Teo, *EXAFS: Basic Principles and Data Analysis*, Springer-Verlag, Berlin, 1986.
33. B.K. Teo, P.A. Lee, 'Ab Initio Calculation of Amplitude and Phase Functions for Extended X-ray Absorption Fine Structure Spectroscopy', *J. Am. Chem. Soc.*, **101**, 2815–2832 (1979).
34. J.J. Rehr, J. Mustre de Leon, S.I. Zabinsky, R.C. Albers, 'Theoretical X-ray Absorption Fine Structure Standards', *J. Am. Chem. Soc.*, **113**, 5135–5140 (1991).
35. J. Mustre de Leon, J.J. Rehr, S.I. Zabinsky, R.C. Albers, 'Ab Initio Curved-wave X-ray-absorption Fine Structure', *Phys. Rev. B*, **44**, 4146–4156 (1991).
36. J.J. Rehr, S.I. Zabinsky, A. Ankudinov, R.C. Albers, 'Atomic-XAFS and XANES', *Physica B.*, **209**, 23–26 (1995).
37. S.I. Zabinsky, J.J. Rehrm, A. Ankudinov, R.C. Albers, M.J. Ella, 'Multiple Scattering Calculations of X-ray Absorption Spectra', *Phys. Rev. B*, **52**, 2995–3009 (1995).
38. P.A. Lee, P.H. Citrin, P. Eisenberg, B.M. Kinkaid, 'Extended X-ray Absorption Fine Structure – Its Strength and Limitations as Structural Tool', *Rev. Mod. Phys.*, **53**, 769–806 (1981).
39. S. Wu, J. Lipkowski, T. Tyliszczak, A.P. Hitchcock, 'Effect of Anion Adsorption on Early Stages of Copper Electrocrystallization at Au(111) Surface', *Prog. Surf. Sci.*, **50**, 227–236 (1995).
40. E. Dartyge, L. Depautex, J.M. Dubisson, A. Fontaine, A. Jucha, P. Leboucher, G. Tourillon, 'X-ray Absorption in Dispersive Mode: a New Spectrometer and a Data Acquisition System for Fast Kinetics', *Nucl. Instrum. Methods*, **A246**, 452–460 (1986).
41. H. Tolentino, E. Dartyge, A. Fontaine, G. Tourillon, 'X-ray Absorption Spectroscopy in the Dispersive Mode

with Synchrotron Radiation: Optical Considerations', *J. Appl. Crys.*, **21**, 15–21 (1988).
42. F. D'Acapito, F. Boschenerini, A. Marcelli, S. Mobilio, 'Dispersive EXAFS Apparatus at Frascati', *Rev. Sci. Instrum.*, **63**, 899–901 (1992).
43. P.G. Allen, S.D. Conradson, J.E. Penner-Hahn, 'A 4-point Crystal Bender for Dispersive X-ray Diffraction', *J. Appl. Cryst.*, **26**, 172–179 (1993).
44. A.J. Dent, M.P. Wells, R.C. Farrow, C.A. Ramsdale, G.E. Derbyshire, G.N. Greaves, J.W. Couves, J.M. Thomas, 'Combined Energy Dispersive EXAFS and X-ray Diffraction', *Rev. Sci. Instrum.*, **63**, 902–905 (1992).
45. M.G. Samant, M.F. Toney, G.L. Borges, L. Blum, O.R. Melroy, 'Grazing Incident X-ray Diffraction of Lead Monolayers at a Silver(111) and a Gold(111) Electrode/Electrolyte Interface', *J. Phys. Chem.*, **92**, 220–225 (1988).
46. M.G. Samant, G.L. Borges, J.G. Gordon, O.R. Melroy, L. Blum, 'In Situ Surface Extended X-ray Absorption Fine Structure Spectroscopy of a Lead Monolayer at a Silver(111) Electrode/Electrolyte Interface', *J. Am. Chem. Soc.*, **109**, 5970–5974 (1987).
47. M.J. Albarrelli, J.H. White, M.G. Bommarito, M. McMillan, H.D. Abruña, 'In Situ EXAFS at Chemically Modified Electrodes', *J. Electroanal. Chem.*, **248**, 77–86 (1988).
48. J. McBreen, W.E. O'Grady, K.I. Pandya, R.W. Hoffman, D.E. Sayers, 'EXAFS Study of the Nickel Hydroxide Electrode', *Langmuir*, **3**, 428–433 (1987).
49. A.J. Davenport, H.S. Isaacs, G.S. Frankel, A.G. Schrott, C.V. Jahnes, M.A. Russak, 'In Situ X-ray Absorption Study of Chromium Valency Changes in Passive Oxides on Sputtered AlCr Thin Films Under Electrochemical Control', *J. Electrochem. Soc.*, **138**, 337–338 (1991).
50. A.J. Davenport, M. Sansone, 'High Resolution In Situ XANES Investigation of the Nature of the Passive Film on Iron in a pH 8.4 Borate Buffer', *J. Electrochem. Soc.*, **142**, 725–730 (1995).
51. A.R. Sandy, S.G.J. Mochrie, D.M. Zhener, K.G. Huang, D. Gibbs, 'Structure and Phases of the Au(111) Surface: X-ray Scattering Measurements', *Phys. Rev. B*, **43**, 4667–4687 (1991).
52. B.M. Ocko, J. Wang, A. Davenport, H. Isaacs, 'In Situ X-ray Reflectivity and Diffraction Studies of the Au(100) Reconstruction in an Electrochemical Cell', *Phys. Rev. Lett.*, **65**, 1466–1469 (1990).
53. I.K. Robinson, 'Crystal Truncation Rods and Surface Roughness', *Phys. Rev. B*, **33**, 3830–3836 (1986).
54. I.K. Robinson, 'Surface Crystallography', in *Handbook on Synchrotron Radiation*, eds. G. Brown, D.E. Moncton, Elsevier Science, Amsterdam, 221, Vol. 3, 1991.
55. D. Gibbs, B.M. Ocko, D.M. Zehner, S.G.J. Mochrie, 'Absolute X-ray Reflectivity Study of the Au(100) Surface', *Phys. Rev. B*, **38**, 7303–7310 (1988).
56. J. Zegenhagen, 'Surface Structure Determination with X-ray Standing Waves', *Surf. Sci. Reports*, **18**, 199–271 (1993).
57. K.M. Robinson, W.E. O'Grady, 'X-ray Diffraction and Electrochemical Study on the Oxidation of Flame-annealed Au(100) Single-crystal Surfaces', *J. Electroanal. Chem.*, **384**, 139–144 (1995).
58. H. You, C.A. Melendres, Z. Nagy, V.A. Maroni, W. Yun, R.M. Yonco, 'X-ray-reflectivity Study of the Copper–Water Interface in Transmission Geometry Under In Situ Electrochemical Control', *Phys. Rev. B*, **45**, 11 288–11 289 (1992).
59. F. Brossard, V.H. Etgens, A. Tadjeddine, 'In Situ X-ray Diffraction Using a New Electrochemical Cell Optimised for Third Generation Synchrotron Light Sources', *Nucl. Instrum. Meth. Phys. Res. B*, **129**, 419–422 (1997).
60. M.F. Toney, J.N. Howard, J. Richer, G.L. Borges, J.G. Gordon, O.R. Melroy, 'Electrochemical Deposition of Copper on a Gold Electrode in Sulfuric Acid – Resolution of the Interfacial Structure', *Phys. Rev. Lett.*, **75**, 4772–4775 (1995).
61. J. Wang, A.J. Davenport, H.S. Isaacs, B.M. Ocko, 'Surface Charge Induced Ordering of the Au(111) Surface', *Science*, **255**, 1416–1419 (1992).
62. B.M. Ocko, A. Gibaud, J. Wang, 'Surface X-ray Diffraction Study of the Au(111) Electrode in 0.01 M NaCl: Electrochemical Induced Surface Reconstruction', *J. Vac. Sci. Technol. A*, **10**, 3019–3025 (1992).
63. J. Wang, B.M. Ocko, A.J. Davenport, H.S. Isaacs, 'In Situ X-ray -Diffraction and -Reflectivity Studies of the Au(111)/Electrolyte Interface: Reconstruction and Anion Adsorption', *Phys. Rev. B*, **46**, 10 321–10 338 (1992).
64. S. Wu, J. Lipkowski, O.M. Magnussen, B.M. Ocko, Th. Wandlowski, 'The Driving Force for the $(p \times \sqrt{3}) \leftrightarrow (1 \times 1)$ Phase Transition of Au(111) in the Presence of Organic Adsorption: A Combined Chronocoulometric and Surface X-ray Scattering Study', *J. Electroanal. Chem.*, **446**, 67–77 (1998).
65. B.M. Ocko, J. Wang, A. Davenport, H.S. Isaacs, 'In Situ X-ray Reflectivity and Diffraction Studies of the Au(100) Reconstruction in an Electrochemical Cell', *Phys. Rev. Lett.*, **65**, 1466–1469 (1990).
66. D. Gibbs, B.M. Ocko, D.M. Zehner, S.G.J. Mochrie, 'Absolute X-ray Reflectivity Study of the Au(100) Surface', *Phys. Rev. B*, **38**, 7303–7310 (1988).
67. S.G.J. Mochrie, D.M. Zehner, B.M. Ocko, D. Gibbs, 'Structure and Phases of the Au(001) Surface: X-ray Scattering Measurements', *Phys. Rev. Lett.*, **64**, 2925–2928 (1990).
68. I.M. Tidswell, N.M. Markovic, P.N. Ross, 'Potential Dependent Surface Relaxation of the Pt(001)/Electrolyte Interface', *Phys. Rev. Lett.*, **71**, 1601–1604 (1993).
69. I.M. Tidswell, N.M. Markovic, P.N. Ross, 'Potential Dependent Surface Structure of the Pt(111)/Electrolyte Interface', *J. Electroanal. Chem.*, **376**, 119–126 (1994).

70. H. You, D.J. Zurawski, Z. Nagy, R.M. Yonco, 'In Situ X-ray Reflectivity Study of Incipient Oxidation of Pt(111) Surface in Electrolyte Solutions', *J. Chem. Phys*, **100**, 4699–4702 (1994).
71. H. You, Z. Nagy, 'Oxidation–Reduction-induced Roughening of Platinum(111) Surface', *Physica B*, **198**, 187–194 (1994).
72. B.M. Ocko, O.M. Magnussen, J.X. Wang, R.R. Adzic, Th. Wandlowski, 'The Structure and Phase Behavior of Electrodeposited Halides on Single-crystal Metal Surfaces', *Physica B*, **211**, 238–244 (1996).
73. B.M. Ocko, G.M. Watson, J. Wang, 'Structure and Electrocompression of Electrodeposited Iodine Monolayers on Au(111)', *J. Phys. Chem.*, **98**, 897–906 (1994).
74. G.M. Bommarito, J.H. White, H.D. Abruña, 'Electrodeposition of Iodide on Platinum: Packing Density and Potential-dependent Distributional Changes Observed In Situ with X-ray Standing Waves', *J. Phys. Chem.*, **94**, 8280–8288 (1990).
75. O.M. Magnussen, B.M. Ocko, R.R. Adzic, J.X. Wang, 'X-ray Diffraction Studies of Ordered Chloride and Bromide Monolayers at the Au(111)–Solution Interface', *Phys. Rev. B*, **51**, 5510–5513 (1995).
76. O.M. Magnussen, B.M. Ocko, J.X. Wang, R.R. Adzic, 'In Situ X-ray Diffraction and STM Studies of Bromide Adsorption on Au(111) Electrodes', *J. Phys. Chem.*, **100**, 5500–5508 (1996).
77. C.A. Lucas, N.M. Markovic, P.N. Ross, 'Observation of an Ordered Bromide Monolayer at the Pt(111)–Solution Interface by In Situ Surface X-ray Scattering', *Surf. Sci.*, **340**, L954–L959 (1995).
78. B.M. Ocko, O.M. Magnussen, J.X. Wang, Th. Wandlowski, 'One-dimensional Commensurate–Incommensurate Transition: Bromide on the Au(100) Electrode', *Phys. Rev. B*, **53**, R7654–7657 (1996).
79. Th. Wandlowski, J.X. Wang, O.M. Magnussen, B.M. Ocko, 'Structural and Kinetic Aspects of Bromide Adsorption on Au(100)', *J. Phys. Chem.*, **100**, 10277–10287 (1996).
80. B.M. Ocko, J.X. Wang, Th. Wandlowski, 'Bromide Adsorption on Ag(001): A Potential Induced Two Dimensionalising Order–Disorder Transition', *Phys. Rev. Lett*, **79**, 1511–1514 (1997).
81. M.H. Hölze, V. Zwing, D.M. Kolb, 'The Influence of Steps on the Deposition of Cu onto Au(111)', *Electrochim. Acta*, **40**, 1237–1247 (1995).
82. T. Hachiya, H. Honbo, K. Itaya, 'Detailed Underpotential Deposition of Copper on Gold(111) in Aqueous-solutions', *J. Electroanal. Chem.*, **315**, 275–291 (1991).
83. L. Blum, H.D. Abruña, J. White, J.G. Gordon II, G.L. Borges, M.G. Samant, O.R. Melroy, 'Study of the Underpotentially Deposited Copper on Gold by Fluorescence Detected Surface EXAFS', *J. Chem. Phys.*, **82**, 6732–6738 (1986).
84. O.R. Melroy, G.L. Borges, L. Blum, H.D. Abruña, M.J. Albarelli, M.G. Samant, M. McMillan, J.H. White, J.G. Gordon II, 'In-plane Structure of Underpotentially Deposited Copper on Gold(111) Determined by Surface EXAFS', *Langmuir*, **4**, 728–732 (1988).
85. A. Tadjeddine, G. Tourillon, D. Guay, 'Structural and Electronic Characterization of Underpotentially Deposited Copper on Gold Single-crystal Probed by In Situ X-ray Absorption-spectroscopy', *Electrochim. Acta*, **36**, 1859–1862 (1991).
86. O.M. Magnussen, J. Hotlos, R.J. Nichols, D.M. Kolb, R.J. Behm, 'Atomic-structure of Cu Adlayers on Au(100) and Au(111) Electrodes Observed by In Situ Scanning Tunneling Microscopy', *Phys. Rev. Lett.*, **64**, 2929–2932 (1990).
87. N. Batina, T. Will, D.M. Kolb, 'Study of the Initial-stages of Copper Deposition by In Situ Scanning-tunneling-microscopy', *Faraday Discuss.*, **94**, 93–100 (1992).
88. M.P. Green, K.P. Hanson, 'Copper Adlayer Formation on Au(111) from Sulfuric-acid Electrolyte', *J. Vac. Sci. Technol. A*, **10**, 3012–3018 (1992).
89. S. Manne, P.K. Hansma, J. Massie, V.B. Elings, A.A. Gewirth, 'Atomic-resolution Electrochemistry with the Atomic Force Microscope – Copper Deposition on Gold', *Science*, **251**, 183–186 (1991).
90. Z. Shi, J. Lipkowski, 'Investigations of SO_4^{2-} Adsorption at the Au(111) Electrode in the Presence of Underpotentially Deposited Copper Adatoms', *J. Electroanal. Chem.*, **364**, 289–294 (1994).
91. Z. Shi, J. Lipkowski, 'Coadsorption of Cu^{2+} and SO_4^{2-} at the Au(111) Electrode', *J. Electroanal. Chem.*, **365**, 303–309 (1994).
92. Z. Shi, S. Wu, J. Lipkowski, 'Coadsorption of Metal Atoms and Anions – Cu UPD in the Presence of SO_4^{2-}, Cl^- and Br^-', *J. Electrochim. Acta*, **40**, 9–15 (1995).
93. J.G. Gordon, O.R. Melroy, M.F. Toney, 'Structure of Metal Electrolyte Interfaces – Copper on Gold(111), Water on Silver(111)', *Electrochim. Acta*, **40**, 3–8 (1995).
94. G.L. Borges, K.K. Kanazawa, J.G. Gordon II, K. Ashley, J. Richer, 'An In Situ Electrochemical Quartz-crystal Microbalance Study of the Underpotential Deposition of Copper on Au(111) Electrodes', *J. Electroanal. Chem.*, **364**, 281–284 (1994).
95. M. Watanabe, H. Uchida, M. Miura, N. Ikeda, 'Electrochemical Quartz-crystal Microbalance Study of Copper Ad-atoms on Highly Ordered Au(111) Electrodes in Sulfuric Acid', *J. Electroanal. Chem.*, **384**, 191–195 (1995).
96. L. Blum, D.A. Huckaby, M. Legault, 'Phase Transitions at Electrode Interfaces', *Electrochim. Acta*, **41**, 2207–2227 (1996).
97. Y. Nakay, M.S. Zei, D.M. Kolb, G. Lehmpfuhl, 'A LEED and RHEED Investigation of Cu on Au(111) in the Underpotential Region', *Ber. Bunsenges. Phys. Chem.*, **88**, 340–345 (1984).
98. M.S. Zei, G. Qiao, G. Lehmpfuhl, D.M. Kolb, 'The Influence of Anions on the Structure of Underpotentially Deposited Cu on Au(111): A LEED, RHEED, and

AES Study', *Ber. Bunsenges. Phys. Chem.*, **91**, 349–353 (1987).
99. A. Tadjeddine, A. Lahrichi, G. Tourillon, 'In Situ X-ray-absorption Near-edge Structure – A Probe of the Oxidation-state of Underpotentially Adsorbed Metals', *J. Electroanal. Chem.*, **360**, 261–270 (1993).
100. E.D. Chabala, J. Cairns, T. Rayment, 'In Situ Real Time Study of an Electrode Process by Differential X-ray Diffraction. Part 2. Cu Underpotential Deposition on Au(111)', *J. Electroanal. Chem.*, **412**, 77–84 (1996).
101. S. Wu, Z. Shi, J. Lipkowski, A.P. Hitchcock, T. Tyliszczak, 'Early Stages of Copper Electrocrystallization: Electrochemical and In Situ X-ray Absorption Fine Structure Studies of Coadsorption of Copper and Chloride on the Au(111) Electrode Surface', *J. Phys. Chem. B*, **101**, 10 310–10 322 (1997).
102. R.G.W. Wyckoff, *Crystal Structures*, Interscience, New York, Vol. 1, 1960.
103. E. Herrero, S. Glazier, J.L. Buller, H.D. Abruña, 'X-ray and Electrochemical Studies of Cu UPD on Single Crystal Electrodes in the Presence of Bromide: Comparison Between Au(111) and Pt(111) Electrodes', *J. Electroanal. Chem.*, **461**, 121–130 (1999).
104. E. Herrero, S. Glazier, H.D. Abruña, 'X-ray and Electrochemical Studies of Cu UPD on Au(111) Single Crystal Electrodes in the Presence of Bromide', *J. Phys. Chem. B*, **102**, 9825–9833 (1998).
105. F.A. Möller, O.M. Magnussen, R.J. Behm, 'In Situ STM Studies of Cu Underpotential Deposition on Au(100) in the Presence of Sulfate and Chloride Anions', *Phys. Rev. B*, **51**, 2484–2490 (1995).
106. H.D. Abruña, T. Gog, G. Mateerlik, W. Uelhoff, 'X-ray Standing-wave Study of Copper Underpotentially Deposited on Au(100)', *J. Electroanal. Chem.*, **360**, 315–323 (1993).
107. G. Tourillon, D. Guay, A. Tadjeddine, 'In-plane Structural and Electronic Characteristics of Underpotentially Deposited Copper on Gold-(100) Probed by In Situ X-ray Absorption-spectroscopy', *J. Electroanal. Chem.*, **289**, 263–278 (1990).
108. M.G. Samant, O.R. Melroy, M.F. Toney, G.L. Borges, L. Blum, 'Grazing-incidence X-ray-diffraction of Lead Monolayers at a Silver(111) and Gold(111) Electrode/Electrolyte Interface', *J. Phys. Chem.*, **92**, 220–225 (1988).
109. M.F. Toney, J.G. Gordon, M.G. Sammant, G.L. Borges, O.R. Melroy, D. Yee, L.B. Sorensen, 'In Situ Atomic-structure of Underpotentially Deposited Monolayers of Pb and Tl on Au(111) and Ag(111) – A Surface X-ray-scattering Study', *J. Phys. Chem.*, **99**, 4733–4744 (1995).
110. M.P. Green, K.J. Hanson, R. Carr, I. Lindau, 'STM Observations of the Underpotential Deposition and Stripping of Pb on Au(111) Underpotential Sweep Conditions', *J. Electrochem. Soc.*, **137**, 3493–3498 (1990).
111. M.P. Green, K.J. Hanson, 'Alloy Formation in an Electrodeposited Monolayer', *Surf. Sci.*, **259**, L743–L749 (1991).
112. N.J. Tao, J. Pan, Y. Li, P.I. Oden, J.A. Derose, S.M. Lindsay, 'Initial-stage of Underpotential Deposition of Pb on Reconstructed and Unreconstructed Au(111)', *Surf. Sci.*, **271**, L338–L344 (1992).
113. C.-H. Chen, N. Washburn, A.A. Gewirth, 'In Situ Atomic-force Microscope Study of Pb Underpotential Deposition on Au(111) – Structural Properties of the Catalytically Active Phase', *J. Phys. Chem.*, **97**, 9754–9760 (1993).
114. E.D. Chabala, B.H. Harji, T. Rayment, M.D. Archer, 'In Situ X-ray-diffraction Study of Underpotential Deposition at the Gold(111) Surface', *Langmuir*, **8**, 2028–2033 (1992).
115. E. Herrero, H.D. Abruña, 'Underpotential Deposition of Mercury on Au(111): Electrochemical Studies and Comparison with Structural Investigations', *Langmuir*, **13**, 4446–4453 (1997).
116. J. Inukai, S. Sugita, K. Itaya, 'Underpotential Deposition of Mercury on Au(111) Investigated by In Situ Scanning-tunneling-microscopy', *J. Electroanal. Chem.*, **403**, 159–168 (1996).
117. J. Li, H.D. Abruña, 'Coadsorption of Sulfate/Bisulfate Anions with Hg Cations During Hg Underpotential Deposition on Au(111): An In Situ X-ray Diffraction Study', *J. Phys. Chem. B*, **101**, 244–252 (1997).
118. C.-H. Chen, A.A. Gewirth, 'AFM Study of the Structure of Underpotentially Deposited Ag and Hg on Au(111)', *Ultramicroscopy*, **42**, 437–444 (1992).
119. C.-H. Chen, A.A. Gewirth, 'In Situ Observation of Monolayer Structures of Underpotentially Deposited Hg on Au(111) with the Atomic Force Microscope', *Phys. Rev. Lett.*, **68**, 1571–1574 (1992).
120. J. Li, H.D. Abruña, 'Phases of Underpotentially Deposited Hg on Au(111): and In Situ Surface X-ray Diffraction Study', *J. Phys. Chem. B*, **101**, 2907–2916 (1997).
121. J. Li, E. Herrero, H.D. Abruña, 'The Effects of Anions in the Underpotential Deposition of Hg on Au(111): An Electrochemical and In Situ X-ray Diffraction Study', *Colloids Surf. A*, **134**, 113–131 (1998).
122. H.S. Yee, H.D. Abruña, 'In Situ X-ray Studies of the Underpotential Deposition of Copper on Platinum(111)', *J. Phys. Chem.*, **97**, 6278–6288 (1993).
123. H.S. Yee, H.D. Abruña, 'Ab-initio XAFS Calculations and In Situ XAFS Measurements of Copper Underpotential Deposition on Pt(111) – A Comparative Study', *J. Phys. Chem.*, **98**, 6552–6558 (1994).
124. C.A. Lucas, N.M. Markovic, P.N. Ross, 'Electrochemical Deposition of Copper onto Pt(111) in the Presence of (Bi)sulfate Anions', *Phys. Rev. B*, **56**, 3651–3654 (1997).
125. C.A. Lucas, N.M. Markovic, I.M. Tidswell, P.N. Ross, 'In Situ X-ray Scattering Study of the Pt(111)–Solution

Interface – Ordered Anion Structures and Their Influence on Copper Underpotential Deposition', *Physica B*, **221**, 245–250 (1996).
126. I.M. Tidswell, C.A. Lucas, N.M. Markovic, P.N. Ross, 'Surface Structure Determination Using Anomalous X-ray Scattering – Underpotential Deposition of Copper on Pt(111)', *Phys. Rev. B*, **51**, 10 205–10 208 (1995).
127. N.M. Markovic, H.A. Gasteiger, C.A. Lucas, I.M. Tidswell, P.N. Ross, 'The Effect of Chloride on the Underpotential Deposition of Copper on Pt(111): AES, LEED, RRDE and X-ray Scattering Studies', *Surf. Sci.*, **335**, 91–100 (1995).
128. R. Gómez, H.S. Yee, G.M. Bommarito, J.M. Feliu, H.D. Abruña, 'Anion Effects and the Mechanism of Cu UPD on Pt(111) – X-ray and Electrochemical Studies', *Surf. Sci.*, **335**, 101–109 (1995).
129. J.H. White, H.D. Abruña, 'Coadsorption of Copper with Anions on Platinum(111) – The Role of Surface Redox Chemistry in Determining the Stability of a Metal Monolayer', *J. Phys. Chem.*, **94**, 894–900 (1990).
130. H.S. Yee, H.D. Abruña, 'In Situ X-ray-absorption Spectroscopy Studies of Copper Underpotentially Deposited in the Absence and Presence of Chloride on Platinum(111)', *Langmuir*, **9**, 2460–2469 (1993).
131. A.C. Finnefrock, L.J. Buller, K.L. Ringland, J.D. Brock, H.D. Abruña, 'Time-resolved Surface X-ray Scattering Study of Surface Ordering of Electrodeposited Layers', *J. Am. Chem. Soc.*, **119**, 11 703–11 704 (1997).
132. H.D. Abruña, J.M. Feliu, J.D. Brock, L.J. Buller, E. Herrero, J. Li, R. Gómez, A. Finnefrock, 'Anion and Electrode Surface Structure Effects on the Deposition of Metal Monolayers: Electrochemical and Time-resolved Surface Diffraction Studies', *Electrochim. Acta*, **43**, 2899–2909 (1998).
133. N.M. Markovic, C.A. Lucas, H.A. Gasteiger, P.N. Ross, 'The Structure of Adsorbed Bromide Concurrent with the Underpotential Deposition (UPD) of Cu on Pt(111)', *Surf. Sci.*, **372**, 239–254 (1997).
134. M.G. Samant, M.F. Toney, G.L. Borges, L. Blum, O.R. Melroy, 'In Situ Grazing-incidence X-ray-diffraction Study of Electrochemically Deposited Pb Monolayers on Ag(111)', *Surf. Sci.*, **193**, L29–L36 (1988).
135. O.R. Melroy, M.F. Toney, G.L. Borges, M.G. Samant, J.B. Kortright, P.N. Ross, L. Blum, 'Two-dimensional Compressibility of Electrochemically Adsorbed Lead on Silver(111)', *Phys. Rev. B*, **38**, 10 962–10 965 (1988).
136. O.R. Melroy, M.F. Toney, G.L. Borges, M.G. Samant, J.B. Kortright, P.N. Ross, L. Blum, 'An In Situ Grazing-incidence X-ray Scattering Study of the Initial-stages of Electrochemical Growth of Lead on Silver(111)', *J. Electroanal. Chem.*, **258**, 403–414 (1989).
137. E.D. Chabala, T. Rayment, 'In Situ Surface Differential Diffraction Study of Metal Monolayer Formation by Underpotential Deposition on Silver(111) Oriented Surfaces', *Langmuir*, **10**, 4324–4329 (1994).
138. W. Obretenov, U. Schmidt, W.J. Lorenz, G. Staikov, E. Budevski, D. Carnal, U. Müller, H. Siegenthaler, E. Schmidt, 'Underpotential Deposition and Electrocrystallization of Metals – An In Situ Scanning-tunneling-microscopy Study with Lateral Atomic-resolution', *Faraday Discuss.*, **94**, 107–116 (1992).
139. W.J. Lorenz, L.M. Gassa, U. Schmidt, W. Obretenov, G. Staikov, V. Bostanov, E. Budevski, 'STM Studies in Underpotential Overpotential Metal-deposition', *Electrochim. Acta*, **37**, 2173–2178 (1992).
140. U. Müller, D. Carnal, H. Siegenthaler, E. Schmidt, W.J. Lorenz, W. Obretenov, U. Schmidt, G. Staikov, E. Budevski, 'Superstructures of Pb Monolayers Electrochemically Deposited on Ag(111)', *Phys. Rev. B*, **46**, 12 899–12 901 (1992).
141. M.G. Samant, G.L. Borges, J.G. Gordon, O.R. Melroy, L. Blum, 'In Situ Surface Extended X-ray Absorption Fine Structure Spectroscopy of a Lead Monolayer at a Silver(111) Electrode/Electrolyte Interface', *J. Am. Chem. Soc.*, **109**, 5970–5974 (1987).
142. M.F. Toney, J.G. Gordon, M.G. Samant, G.L. Borges, O.R. Melroy, L.-S. Kau, D.G. Wiesler, D. Yee, L.B. Sorensen, 'Surface X-ray-scattering Measurements of the Substrate-induced Spatial Modulation of an Incommensurate Adsorbed Monolayer', *Phys. Rev. B*, **42**, 5594–5603 (1990).
143. M.F. Toney, J.G. Gordon, M.G. Samant, G.L. Borges, O.R. Melroy, D. Yee, L.B. Sorensen, 'Underpotentially Deposited Thallium on Silver(111) by In Situ Surface X-ray-scattering', *Phys. Rev. B*, **45**, 9362–9374 (1992).
144. J.X. Wang, R.R. Adzic, O.M. Magnussen, B.M. Ocko, 'Structural Evolution During Electrocrystallization – Deposition of Tl on Ag(100) from Monolayer to Bilayer and to Bulk Crystallites', *Surf. Sci.*, **344**, 111–121 (1995).
145. M.E. Herron, S.E. Doyle, S. Pizzini, K.J. Roberts, J. Robinson, G. Hards, F.C. Walsh, 'In Situ Studies of a Dispersed Platinum on Carbon Electrode Using X-ray Absorption Spectroscopy', *J. Electroanal. Chem.*, **324**, 243–258 (1992).
146. S. Mukerjee, S. Srinivasan, M.P. Soriaga, J. McBreen, 'Role of Structural and Electronic Properties of Pt and Pt Alloys on Electrocatalysis of Oxygen Reduction: An In Situ XANES and EXAFS Investigation', *J. Electrochem. Soc.*, **142**, 1409–1422 (1995).
147. S. Mukerjee, J. McBreen, 'Hydrogen Electrocatalysis by Carbon Supported Pt and Pt Alloys: An In Situ X-ray Absorption Study', *J. Electrochem. Soc.*, **143**, 2285–2294 (1996).
148. P.G. Allen, S.D. Conradson, M.S. Wilson, S. Gottesfeld, I.D. Raistrick, J. Valerio, M. Lovato, 'In Situ Structural Characterization of a Platinum Electrocatalyst by Dispersive X-ray Absorption Spectroscopy', *Electrochim. Acta*, **39**, 2415–2418 (1994).
149. S. Mukerjee, J. McBreen, 'Effect of Particle Size on the Electrocatalysis by Carbon-supported Pt Electrocatalyst:

An In Situ XAS Investigation', *J. Electroanal. Chem.*, **448**, 163–171 (1998).
150. J. McBreen, W.E. O'Grady, G. Tourillon, E. Dartyge, A. Fontaine, 'XANES Study of Underpotential Deposited Copper on Carbon Supported Platinum', *J. Electroanal. Chem.*, **307**, 229–240 (1991).
151. J. McBreen, 'EXAFS Studies of Adsorbed Copper on Carbon Supported Platinum', *J. Electroanal. Chem.*, **357**, 373–386 (1993).
152. J. McBreen, M. Sansone, 'In Situ X-ray Absorption Spectroscopic Study of Adsorbed Pb on Carbon-supported Pt', *J. Electroanal. Chem.*, **373**, 227–233 (1994).
153. D. Aberdam, R. Durand, R. Faure, F. Gloaguen, L.J. Hazemann, E. Herrero, A. Kabbabi, O. Ulrich, 'X-ray Absorption Near Edge Structure Study of the Electro-oxidation of Co on $Pt_{50}Ru_{50}$ Nanoparticles', *J. Electroanal. Chem.*, **398**, 43–47 (1995).
154. J. McBreen, S. Mukerjee, 'In Situ X-ray Absorption Studies of a Pt–Ru Electrocatalyst', *J. Electrochem. Soc.*, **142**, 3399–3404 (1995).
155. J. McBreen, W.E. O'Grady, K.I. Pandya, R.W. Hoffman, D.E. Sayers, 'EXAFS Study of the Nickel Hydroxide Electrode', *Langmuir*, **3**, 428–433 (1987).
156. J. McBreen, W.E. O'Grady, G. Tourillon, E. Dartyge, A. Fontaine, K.I. Pandya, 'In Situ Time Resolved X-ray Absorption Near Edge Structure of the Nickel Oxide Electrode', *J. Phys. Chem.*, **93**, 6308–6311 (1989).
157. K.I. Pandya, R.W. Hoffman, J. McBreen, W.E. O'Grady, 'In Situ X-ray Absorption Spectroscopy Studies of Nickel Oxide Electrodes', *J. Electrochem. Soc.*, **137**, 383–388 (1990).
158. K.I. Pandya, W.E. O'Grady, D.A. Corrigan, J. McBreen, R.W. Hoffman, 'Extended X-ray Absorption Fine Structure Investigations of Nickel Hydroxides', *J. Phys. Chem.*, **94**, 21–26 (1990).
159. W.E. O'Grady, K.I. Pandya, K.E. Swider, D.A. Corrigan, 'In Situ X-ray Absorption Near Edge Structure Evidence for Quadrivalent Nickel in Nickel Battery Electrodes', *J. Electrochem. Soc.*, **143**, 1613–1616 (1996).
160. X. Quian, H. Sambe, D.E. Ramaker, K.I. Pandya, W.E. O'Grady, 'Quantitative Interpretation of the K-edge NEXAFS Data for Various Nickel Hydroxides and the Charged Nickel Electrode', *J. Phys. Chem. B*, **101**, 9441–9446 (1997).
161. T.W. Capehart, D.A. Corrigan, R.S. Conell, K.I. Pandya, R.W. Hoffman, 'In Situ Extended X-ray Absorption Fine Structure Spectroscopy of a Thin Film Nickel Hydroxide Electrode', *Appl. Phys. Lett.*, **58**, 865–867 (1991).
162. C. Cartier, A. Tranchant, M. Verdaguer, R. Messina, H. Dexpert, 'X-ray Diffraction and X-ray Absorption Studies of the Structural Modifications Induced by Electrochemical Lithium Intercalation into V_2O_5', *Electrochim. Acta*, **35**, 889–898 (1990).
163. P. Druska, H.H. Strehblow, 'In Situ Examination of Electrochemically Formed Cu_2O by EXAFS in Transmission', *J. Electroanal. Chem.*, **335**, 55–65 (1992).
164. M.E. Herron, D. Pletcher, F.C. Walsh, 'A Combined Electrochemical and In Situ X-ray Diffraction Study of the Cycling of Well-defined Lead Dioxide Layers on Platinum', *J. Electroanal. Chem.*, **332**, 183–197 (1992).
165. S. Mukerjee, T.R. Thurston, N.M. Jisrawi, X.Q. Yang, J. McBreen, M.L. Daroux, X.K. Xing, 'Structural Evolution of $Li_xMn_2O_4$ in Lithium-ion Battery Cells Measured In Situ Using Synchrotron X-ray Diffraction Techniques', *J. Electrochem. Soc.*, **145**, 466–472 (1998).
166. M.F. Toney, A.J. Davenport, L.J. Oblonsky, M.P. Ryan, C.M. Vitus, 'Atomic Structure of the Passive Oxide Film Formed on Iron', *Phys. Rev. Lett.*, **79**, 4282–4285 (1997).
167. L.J. Oblonsky, A.J. Davenport, M.P. Ryan, H.S. Isaacs, R.C. Newman, 'In Situ X-ray Absorption Near Edge Structure Study of the Potential Dependence of the Formation of the Passive Film on Iron Borate Buffer', *J. Electrochem. Soc.*, **144**, 2398–2404 (1997).
168. P. Schmuki, S. Virtanen, A.J. Davenport, C. Vitus, 'In Situ X-ray Absorption Near Edge Spectroscopic Study of the Cathodic Reduction of Artificial Iron Oxide Passive Films', *J. Electrochem. Soc.*, **143**, 574–582 (1996).
169. S. Virtanen, P. Schmuki, A.J. Davenport, C. Vitus, 'Dissolution of Thin Iron Oxide Films Used as Models for Iron Passive Films Studied by In Situ X-ray Absorption Near-edge Spectroscopy', *J. Electrochem. Soc.*, **144**, 198–204 (1997).
170. A.J. Davenport, M. Sansone, J.A. Bardwell, A.J. Aldykiewicz, M. Taube, C. Vitus, 'In Situ Multielement XANES Study of Formation and Reduction of the Oxide Film in Stainless Steel', *J. Electrochem. Soc.*, **141**, L6–L8 (1994).
171. J.A. Bardwell, G.I. Sproule, B. Macdougall, M.J. Graham, A.J. Davenport, H.S. Isaacs, 'In Situ XANES Detection of Cr(VI) in the Passive Film on Fe–^{26}Cr', *J. Electrochem. Soc.*, **139**, 371–373 (1992).
172. P. Schmuki, S. Virtanen, H.S. Isaacs, M.P. Ryan, A.J. Davenport, H. Böhni, T. Stenberg, 'Electrochemical Behavior of Cr_2O_3/Fe_2O_3 Artificial Passive Films Studied by In Situ XANES', *J. Electrochem. Soc.*, **143**, 791–801 (1998).
173. G.S. Frankel, A.J. Davenport, H.S. Isaacs, A.G. Schrott, C.V. Jahnes, M.A. Russak, 'X-ray Absorption Study of Electrochemically Grown Oxide Films an Al–Cr Sputtered Alloys: I. Ex situ Studies', *J. Electrochem. Soc.*, **139**, 1812–1820 (1992).
174. G.S. Frankel, A.G. Schrott, A.J. Davenport, H.S. Isaacs, C.V. Jahnes, M.A. Russak, 'X-ray Absorption Study of Electrochemically Grown Oxide Films an Al–Cr Sputtered Alloys: II. In Situ Studies', *J. Electrochem. Soc.*, **141**, 83–90 (1994).
175. A.J. Davenport, H.S. Isaacs, M.W. Kendig, 'X-ray Absorption Study of Cerium in the Passive Film on Aluminum', *J. Electrochem. Soc.*, **136**, 1837–1838 (1989).

176. D.A. Smith, M.J. Heeg, W.R. Heineman, R.C. Elder, 'Direct Determination of Fe–C Bond Lengths in Iron(II) and Iron(III) Cyanide Solutions Using EXAFS Spectroelectrochemistry', *J. Am. Chem. Soc.*, **106**, 3053–3054 (1984).
177. R.C. Elder, C.E. Lunte, A.F.M.N. Rahaman, J.R. Kirchoff, H.D. Dewald, W.R. Heineman, 'In Situ Observation of Copper Redox in a Polymer Modified Electrode Using EXAFS Spectroelectrochemistry', *J. Electroanal. Chem.*, **240**, 361–364 (1988).
178. J. McBreen, I.-C. Lin, 'X-ray Studies of PEO–$ZnBr_2$ Complexes', *J. Electrochem. Soc.*, **139**, 960–966 (1992).
179. J. McBreen, X.Q. Yang, H.S. Lee, Y. Okamoto, 'X-ray Absorption Studies of Mixed Salt Polymer Electrolytes', *J. Electrochem. Soc.*, **142**, 348–354 (1995).
180. J. McBreen, X.Q. Yang, H.S. Lee, Y. Okamoto, 'EXAFS Studies of Mixed Salt $CoBr_2$/LiBr–PEO Complexes', *Electrochim. Acta*, **40**, 2115–2118 (1995).
181. M. Aziz, R.J. Latham, R.G. Linford, W.S. Schlindwein, 'X-ray Spectroscopy of Polymer Electrolytes', *Electrochim. Acta*, **40**, 2119–2122 (1995).
182. M.J. Albarelli, J.H. White, M.G. Bommarito, M. McMillan, H.D. Abruña, 'In Situ Surface EXAFS of Chemically Modified Electrodes', *J. Electroanal. Chem.*, **248**, 77–86 (1988).
183. G. Tourillon, E. Dartyge, H. Dexpert, A. Fontaine, A. Jucha, P. Lagarde, D.E. Sayers, 'Electrochemical Inclusion of Metallic Clusters in Organic Conducting Polymers: An In Situ Dispersive X-ray Absorption Study', *J. Electroanal. Chem.*, **178**, 357–366 (1984).
184. G. Tourillon, E. Dartyge, H. Dexpert, A. Fontaine, A. Jucha, P. Lagarde, D.E. Sayers, 'Electrochemical Inclusion of Metallic Clusters in Organic Conducting Polymers: An In Situ Dispersive X-ray Absorption Study', *Surf. Sci.*, **156**, 536–547 (1985).
185. G. Tourillon, H. Dexpert, P. Lagarde, 'Charge and Size Displacement Effects on the Conductivity of Poly-3-methylthiophene: An EXAFS Study', *J. Electrochem. Soc.*, **134**, 327–331 (1987).

Electronic Absorption and Luminescence: Introduction

Matthew E. McCarroll and Isiah M. Warner
Louisiana State University, Baton Rouge, USA

Rezik A. Agbaria
Lawrence Livermore National Laboratory, Livermore, USA

1 Introduction **1**
1.1 History of Luminescence 1
2 Absorption of Light **2**
2.1 Interaction of Light and Matter 2
2.2 Principle Modes of Absorption 3
3 Luminescence Principles **4**
3.1 Multiplicity 4
3.2 Photoluminescence Transitions 5
3.3 Stokes' Law 6
3.4 Franck–Condon Principle and Mirror Imagery 6
3.5 Kinetics of Fluorescence 6
3.6 Linearity of Fluorescence 8
3.7 Fluorescence Lifetime 8
3.8 Fluorescence Quenching 9
3.9 Polarization 11
4 Basic Fluorescence Instrumentation **17**
4.1 General Instrumentation 17
4.2 Light Sources 17
4.3 Wavelength Selectors 18
4.4 Detectors 19
4.5 Spectral Correction 19
4.6 Measurement of Fluorescence Lifetimes 20
Abbreviations and Acronyms **22**
Further Reading **22**
References **22**

1 INTRODUCTION

The emission of light can be categorized as either incandescence or luminescence. Incandescence originates from hot bodies simply because they are at a high temperature. The term luminescence refers to all other forms of light emission and involves a radiative transition whereby a molecule lowers its energy by emission of a photon.

In order for the luminescence process to exist for any period of time, energy must be supplied to the system to maintain an excited state population. Different forms of luminescence have been classified based on the source of this energy. While several interesting forms of luminescence exist (e.g. electroluminescence, bioluminescence, radioluminescence, triboluminescence and sonoluminescence), this discussion will focus on the type of luminescence that is of analytical importance, i.e. photoluminescence. In photoluminescence, the excited state is produced by the absorption of light.

Luminescence processes can further be categorized as fluorescence or phosphorescence. This distinction is based on the multiplicity of the two energy levels involved in radiative deactivation from the excited state. If the electron spin states of the two energy levels are the same, the process is referred to as fluorescence. In contrast, if the spin states are not the same, the process is properly called phosphorescence. The distinction and consequences of these differences will be discussed in further detail in section 3.1.

Fluorescence spectroscopy has several analytical advantages over spectroscopic methods such as absorption spectrometry. The most significant are selectivity and sensitivity, which are inherent to the technique. The selectivity arises from the fact that both the absorption and emission occur at specific wavelength ranges for a given fluorophore. The sensitivity of the technique, when compared to absorption, arises from the fact that emission is a measure of low light levels above a theoretically zero background. In contrast, absorbance involves a measurement of the ratio of two large signals. The latter is a more difficult measurement. Therefore, the detection limit is typically 100 to 1000 times better in fluorimetry. These attributes make fluorescence spectroscopy a powerful tool for analytical measurements.

This text discusses fundamental principles involved in absorption and luminescence spectroscopies. Subsequent articles discuss specific topics and techniques in further detail (see **Phosphorescence Measurements, Applications of**, **Fluorescence Imaging Microscopy**, **Surface Measurements using Absorption/Luminescence**). Articles **Absorption and Luminescence Probes**, **Fluorescence in Organized Assemblies**, and **Indirect Detection Methods in Capillary Electrophoresis** discuss other techniques in detail.

1.1 History of Luminescence

Analysis by luminescence is a well-established technique with a history that dates back to the sixteenth and seventeenth centuries. The first documented observation of luminescence was by the Spanish doctor and scientist Nichola Monardes in 1577, in which he noted a mysterious

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

blue tinge in water stored in wood from the tree ''lignum nephriticum''. The phenomenon of fluorescence was also described by Isaac Newton and documented by Robert Boyle[1] in 1680. Throughout the 18th century, many new fluorescent materials were discovered, but an explanation of the phenomenon was not offered until 1833 when David Brewster[2] attempted to explain the red fluorescence observed from a material extracted from green leaves. Brewster mistakenly ascribed the effect to light scattering, rather than an emission process.

The next report of the fluorescence phenomenon was in 1845 by the astronomer John Herschel,[3] who examined the spectral properties of quinine solutions. Herschel noted that under certain observation angles, a ''celestial blue colour'' could be observed, although the solution appeared colorless under ordinary conditions. He noted the source of the blue emission was a well-defined area near the point of illumination and that even though the light was apparently transmitted through the solution, it lost the ability to produce the effect any further. He attributed the phenomenon to dispersion at the surface of the liquid and called the process epipolic. He referred to a beam of light that had passed though this solution as epipolized and considered it to be of a fundamentally different nature. In 1848, Brewster made note of Herschel's report and was able to show that, under appropriate conditions, the blue light could be produced at a considerable distance from the surface of the solution.[4] He contended that the observed phenomenon was simply a special case of internal dispersion. Both Herschel and Brewster noted that the light was qualitatively different after passing through the quinine solution, but could not explain the phenomenon and did not notice that the emitted light was of a longer wavelength. It is interesting and relevant to note that the concepts of absorption and emission had not yet been established.

In 1852, George Gabriel Stokes[4] published a comprehensive paper of monumental importance in the field of fluorescence spectroscopy. In this paper, Stokes offered an alternative explanation to that of Brewster and Herschel. Using a solution of quinine, a small amount of sunlight was allowed to pass through a pinhole in a paper shield around the tube. The solution was observed from above and the blue fluorescence could be seen extending a short distance from the pinhole. Remarkably, Stokes was able to carry out a crude excitation scan, by observing the emission when the solution was placed in different parts of spectrally dispersed sunlight which had been passed though a prism. It was noted that the emission was most prominent when the solution was illuminated by light ''a little short of the extreme violet'' (note that quinine has a maximum excitation of ~350 nm). These experiments showed that the ''epipolized'' light was not fundamentally different from the ordinary light, but simply lacked the shorter-wavelength components due to the filtering effect of the quinine absorption. Additionally, Stokes was able to show that the fluorescence emission was always of a longer wavelength than the absorption, a phenomenon we now refer to as the ''Stokes shift''. Furthermore, it was in this paper that Stokes presented the term fluorescence, named after the fluorescent mineral fluorspar.

Later work by Stokes established the link between the concentration of a fluorescent solution and the intensity of its fluorescence, as well as the quenching which is observed at high concentrations due to the inner filter effect.[5] Stokes is documented as being the first to suggest the analytical use of fluorescence, but Goppelsröder[6] performed the first fluorimetric analysis in 1867, in which determination of the morin chelate of Al(III) was demonstrated. Table 1 provides a brief chronology of the discovery of luminescence.

Table 1 Major historical events in early luminescence studies

1577	Nicolas Monardes – first documented observation of luminescence
1600s	Other observations were made, most notably by Robert Boyle and Isaac Newton
1833	David Brewster described fluorescence of a plant extract
1845	John Herschel completed a crude spectral analysis of the fluorescence of quinine solutions
1846	David Brewster attributed fluorescence to a scattering phenomenon
1852	George Gabriel Stokes published ''On the Change of Refrangability of Light'' in *Philisophical Transactions* and introduced the term ''fluorescence''
1864	George Gabriel Stokes proposed the use of fluorescence as an analytical tool
1867	Goppelsröder performed the first fluorimetric analysis

2 ABSORPTION OF LIGHT

2.1 Interaction of Light and Matter

Molecular spectroscopy involves the interaction of light and matter. This includes such processes as reflection, refraction, scattering, and absorption. Although electromagnetic radiation can be considered as both a wave and a particle, many phenomena are adequately described by a classical treatment that considers the wave properties of electromagnetic radiation. Other phenomena, however, can only be explained by evoking theory that considers the particle properties of electromagnetic radiation (quantum mechanical model). In fact, much of the confusion and conflict surrounding early investigations of

the fluorescence phenomenon (section 1.1) comes from the fact that classical wave theory could not adequately explain absorption and emission phenomena. It was not until nearly 30 years later with the discovery of the photoelectric effect that the particle properties of light could be used to describe absorption and emission processes. In this discussion, both theories will be used as necessary to convey concepts in the most efficient manner.

The wave model explains phenomena that do not involve absorption, e.g. diffraction, reflection, and refraction. The wave model depicts electromagnetic radiation as the propagation of mutually orthogonal magnetic and electric vectors, which oscillate at a characteristic frequency (ν) and wavelength (λ). These two parameters are related by $\lambda\nu = c$, where c is the speed of light in vacuum. A classical description of the interaction of light and matter can be considered as oscillations of the electrons and atoms that are induced by the interaction with the propagating wave. To explain absorption, quantum theory, which considers electromagnetic radiation as discrete packets of energy (photons), must be invoked. The energy of a photon is quantized, i.e. the energy of a particular photon is well defined. In addition, molecules are considered to exist in well-defined energy states. Since both absorption and emission processes are quantized, a transition can only occur when both the photon and the molecule are of a very specific energy. The energy (E) of a photon is related to its frequency (ν) by Planck's constant (h) as defined in Equation (1):

$$E = h\nu \tag{1}$$

Absorption of light by a molecule involves a transition of the molecule to a higher energy level. The energy of the absorbed photon is indicative of the energy difference between the two energy states, i.e. the ground state and the excited state. Therefore, absorption will only occur if the energy of the photon matches the energy difference between the two energy levels, as shown in Equation (2):

$$h\nu = E_2 - E_1 \tag{2}$$

where E_1 and E_2 are the energies of the two energy states involved in the transition. The simplified energy-level diagram shown in Figure 1 illustrates this dependence.

Molecular changes that occur upon absorption can vary, depending on the magnitude of the energy. Upon absorption, the molecule is elevated to a higher energy level and this process is referred to as a molecular transition. Most molecules eliminate this excess energy through heat dissipation with its surroundings. Under certain conditions, the reverse of absorption can occur, whereby a molecule emits a photon as it returns to a lower energy state. This emission process is governed by the same selection rules as absorption and is a quantized phenomenon.

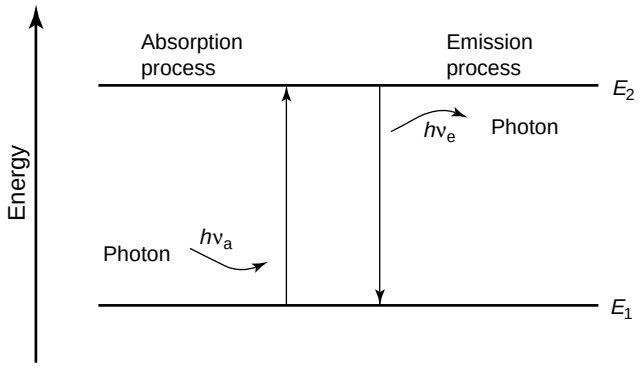

Figure 1 Simplified energy level diagram depicting absorption and emission processes.

2.2 Principle Modes of Absorption

A simple example of absorption is atomic absorption. In this case, the energy level of the atom is determined by the arrangement of electrons around the nucleus. Absorption signifies movement to a higher energy level through promotion of an electron to a higher orbital. Atomic emission is the opposite process, whereby the energy of the atom is lowered by the emission of a photon, the energy of which corresponds to the energy gap between the two atomic orbitals. In the case of molecular absorption, however, one must consider other forms of energy in addition to the electronic energy. Unlike atoms, molecules can undergo transitions to elevated energy levels through different vibrational and rotational modes.

The type of transition involved is dependent on the energy of the photon. Absorption of light in the ultraviolet and visible regions typically involves electronic transitions of the outer electrons. Absorption in the infrared region involves transitions to higher vibrational levels, while the far-infrared and microwave regions consist of rotational transitions. Absorption in the ultraviolet/visible (UV/VIS) region can involve one or more of these transitions. The total energy (E_{tot}) of the molecule is then the sum of the electronic (E_{elec}), vibrational (E_{vib}), and rotational (E_{rot}) energies, as shown in Equation (3):

$$E_{tot} = E_{elec} + E_{vib} + E_{rot} \tag{3}$$

The various electronic, vibrational, and rotational energy levels are illustrated in the energy level diagram shown in Figure 2.

The energy of the excited-state transition is the change in the total energy of the molecule and corresponds to the energy of the photon involved in the transition, as shown

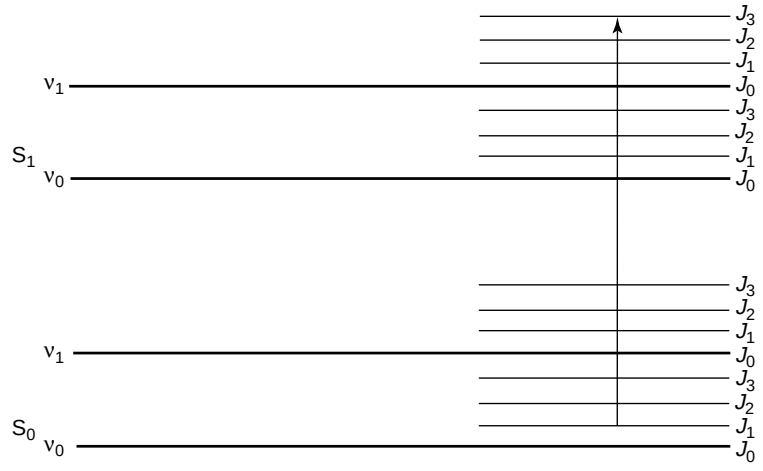

Figure 2 Energy level diagram illustrating electronic (S), vibrational (v), and rotational (J) energy levels.

in Equation (4).

$$\Delta E_{\text{tot}} = h\nu = \frac{hc}{\lambda} \tag{4}$$

The vertical transition shown in Figure 2 involves electronic, vibrational, and rotational transitions, as respectively indicated by the $S_0 \rightarrow S_1$, $v_0 \rightarrow v_1$ and $J_1 \rightarrow J_3$ transitions. An absorption spectrum is a measure of the absorbance as a function of wavelength. In the theoretical case of an isolated molecule and a perfect spectrometer, a line of finite width would be measured for each electronic, vibrational, and electronic transition. Atomic absorption spectra and some vapor-phase molecular absorption spectra come close to this prediction, with the limiting factor often being the resolving power of the spectrometer. A classic example of such a molecular absorption spectrum is that obtained for iodine vapor. It should be noted that an important aspect of an absorption spectrum is that the relative intensities of each line are indicative of the probability of a given transition, although instrumental artifacts can affect this.

Even though each transition is quantized, absorption spectra do not always show discrete rotational, vibrational, and electronic energy levels. Often, the interval between rotational and vibrational energy levels can be rather small. Moreover, in a sample containing many molecules, the energy level of each molecule can vary slightly due to localized variations, resulting in overlapping energy transitions. This effect is compounded in solution phase, resulting in relatively broad absorption bands. The degree of vibrational structure evident in the measured spectrum is a function of the solvent and structure of the molecule. Some rigid molecules, such as polyaromatic hydrocarbons, show vibrational structure. Most condensed-phase absorption spectra are, however, very broad and featureless.

Finally, it should be noted that the Beer–Lambert law describes the absorbance of a molecule, as shown in Equation (5)

$$A = \varepsilon bc \tag{5}$$

where ε is the molar extinction (probability of absorption), b is the cell pathlength and c is the molar concentration of absorber in the cell. It should be noted from Equation (6) that

$$A = \log \frac{I_o}{I_t} \tag{6}$$

where I_o is the intensity of the incident light and I_t is the intensity of the transmitted light. Rearranging this equation gives Equation (7):

$$A = \log I_o - \log I_t \tag{7}$$

Therefore, absorbance is a measure of the difference in the logarithm of the intensities and not the difference in intensities as one might intuitively expect. This is because absorption is an exponential function of cell pathlength and concentration of absorber.

A more thorough treatment of absorption spectrometry and associated techniques can be found in other articles within this section. (See **Near-infrared Absorption/Luminescence Measurements**, **Circular Dichroism and Linear Dichroism**).

3 LUMINESCENCE PRINCIPLES

3.1 Multiplicity

The multiplicity is defined as shown in Equation (8)

$$M = 2(S) + 1 \tag{8}$$

where M is the multiplicity and S is the spin quantum number of the molecule. Thus, the multiplicity is simply a representation of the number of unpaired electrons in the molecular orbital of the molecule. We shall restrict our discussion to the most common fluorescent molecules, i.e. organic molecules, which typically have an even number of electrons. Of course, free radicals are an exception. For an even number of electrons, $S = 0$ and the multiplicity is given by Equation (9):

$$M = 2(0) + 1 = 1 \tag{9}$$

and the molecule exists as a singlet state, i.e. $M = 1$. Therefore, the ground state in most organic molecules exists as a singlet energy state. It is possible, however, for the spin of an electron in an excited state to become reversed, resulting in two unpaired electrons. In this case, the spin is equal to 1 and the multiplicity is given by Equation (10):

$$M = 2(1) + 1 = 3 \tag{10}$$

This state is referred to as a triplet state (T). The multiplicities of the energy states involved in a luminescence transition are of great importance, as they designate the fundamental difference between fluorescence and phosphorescence, as noted in section 1.

3.2 Photoluminescence Transitions

Fluorescence is by far the most commonly employed luminescence technique and as such, it will be covered in greater detail. Fluorescence is a competitive process based on the formation of excited-state molecules and subsequent deactivation to a lower energy state. Molecular and environmental parameters affect the kinetics of this process. In fact, some of the most useful luminescence techniques are based on this dependence. Jablonski[7] first suggested explaining luminescence by use of the energy level diagram shown in Figure 3.

This diagram is an effective method of visualizing the different transitions involved in the processes of luminescence. The ground, first, and second excited states are depicted as S_0, S_1 and S_2, indicating singlet states of different energies. The term T_1 designates the triplet state. Within each electronic energy level, vibrational and rotational energy levels are designated as v_n and J_n, respectively. Figure 3 is separated into three parts illustrating the absorption, fluorescence, and phosphorescence phenomena. Vertical transitions denote transitions that occur on a very rapid timescale and can be considered as virtually instantaneous. Nonvertical lines are used to depict transitions occurring on a timescale slower than that of the vertical transitions. When a photon of an appropriate wavelength is absorbed by a molecule, the molecule is excited to a higher energy state. At room temperature, the vast majority of molecules will reside in the lowest vibrational level of the ground electronic state, i.e. the S_0 state for organic molecules. This is illustrated by considering the Boltzmann distribution, which is given by Equation (11)

$$R = e^{-\Delta E/kT} \tag{11}$$

where R is the ratio of molecules in a given state, ΔE is the energy difference between the two states, k is the Boltzmann constant and T is the temperature in Kelvin. Consider the distribution of molecules between the S_0v_0 and S_0v_1 states. Using a typical vibrational energy gap of

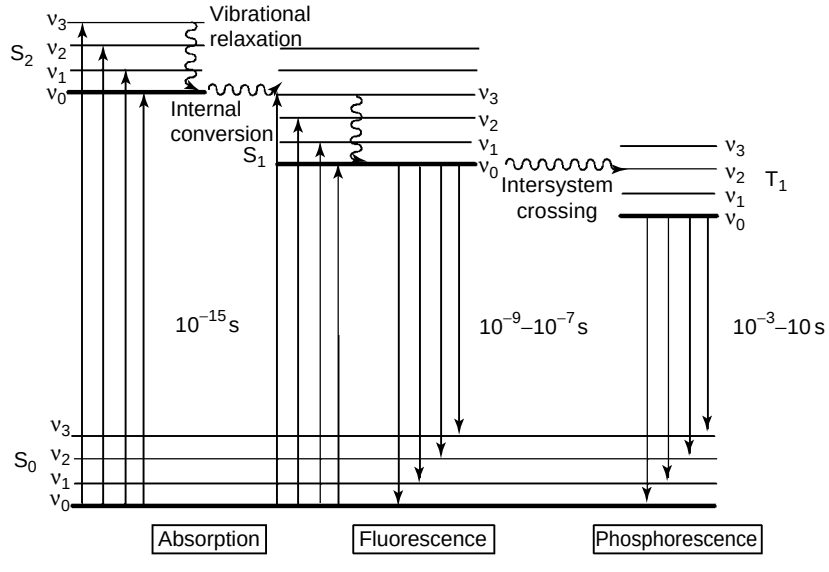

Figure 3 Jablonski energy level diagram illustrating photoluminescence processes.

$1500\,cm^{-1}$ and a temperature of 300 K, only 1% of the molecules will reside in the $S_0\nu_1$ state. The population of other excited states, such as higher electronic states, will be even smaller than this since ΔE is significantly greater. It should therefore be apparent that absorption will almost always involve excitation of a molecule from the lowest vibrational level of the ground state.

Upon absorption of photon energy, a molecule will act to minimize this energy by undergoing one of several transitions. The first and most rapid transition involves relaxation of the molecule to the lowest vibrational state of S_1. This is typically true, whether the absorption involved is to S_1 or S_2. This rapid relaxation process is referred to as internal conversion. From the lowest vibrational state of S_1, the excited molecule can further lower its energy by emission of a photon (fluorescence emission) or by conversion to the triplet state, T_1. The latter process is referred to as intersystem crossing and is a much slower process than internal conversion, since the multiplicity of the molecule must change during the transition. Processes that involve changes in multiplicity are termed ''forbidden processes'', i.e. have a low probability of occurring.

Once in the triplet state, the molecule can then lower its energy further by emission of a photon (phosphorescence). Note that the transition from T_1 to S_0 involves states of different multiplicities and is therefore a forbidden transition, occurring only under certain conditions, with low efficiency and at a slow rate. It should also be noted that since T_1 is lower in energy than S_1, phosphorescence occurs at longer wavelengths and is a much slower process with phosphorescence lifetimes ranging from microseconds to several seconds. This can be compared to fluorescence lifetimes, which are typically in the nanosecond range.

3.3 Stokes' Law

Internal conversion from excited states higher than S_1 typically occurs within 10^{-12} s. Since most fluorescence lifetimes are in the order of 10^{-9} s, this ensures that molecules will usually return to the lowest vibrational level of S_1 prior to emission. The energy gap between S_1 and S_0 is significantly larger, such that fluorescence becomes competitive with internal conversion for many molecules.

As noted in section 3.1, absorption usually occurs from the lowest vibrational level of the ground state. Thus, the fluorescence process typically involves absorption from the lowest vibrational level of S_0 and emission from the lowest vibrational level of S_1. As a consequence, fluorescence emission always occurs at a wavelength equal to, or longer than, that of the absorption. In fact, with the exception of some gas-phase molecules, emission is always of a longer wavelength than the absorption due to energy losses from solvent interactions. This is commonly referred to as the Stokes shift and was first noted by Stokes in 1852[4] (see section 1.1).

3.4 Franck–Condon Principle and Mirror Imagery

The absorption process is generally acknowledged to occur within one oscillation period of the absorbed light, which corresponds to a time period of 1×10^{-15} s for light in the ultraviolet region. The Franck–Condon principle assumes that the position and momentum of the molecule will not change during excitation. This means that the probability (Franck–Condon factor) of a transition is the same in the up (absorption) and down (emission) process. In other words, an intense band in the absorption spectrum will also be an intense band in the emission spectrum. A practical consequence of this is referred to as the mirror image rule (Figure 4).

There are many exceptions to the mirror image rule. These typically involve processes that affect the excited and ground states differently. For example, if the molecule becomes more rigid in the excited state, as is the case in the example of biphenyl shown in Figure 5, the emission may exhibit more defined vibrational structure than does the absorption spectrum.

3.5 Kinetics of Fluorescence

Fluorescence is a process controlled by a number of competing processes. For a given population of molecules in an excited state, not all will return to the ground state by a radiative process. The fraction of molecules in the excited state that do return to the ground state via fluorescence emission is referred to as the quantum yield efficiency, Φ_f. In its most fundamental representation, the

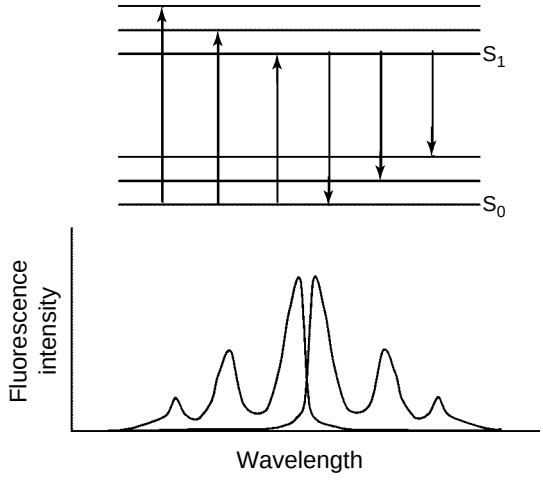

Figure 4 Idealized absorption and emission spectra (below) and the corresponding energy level diagram (above).

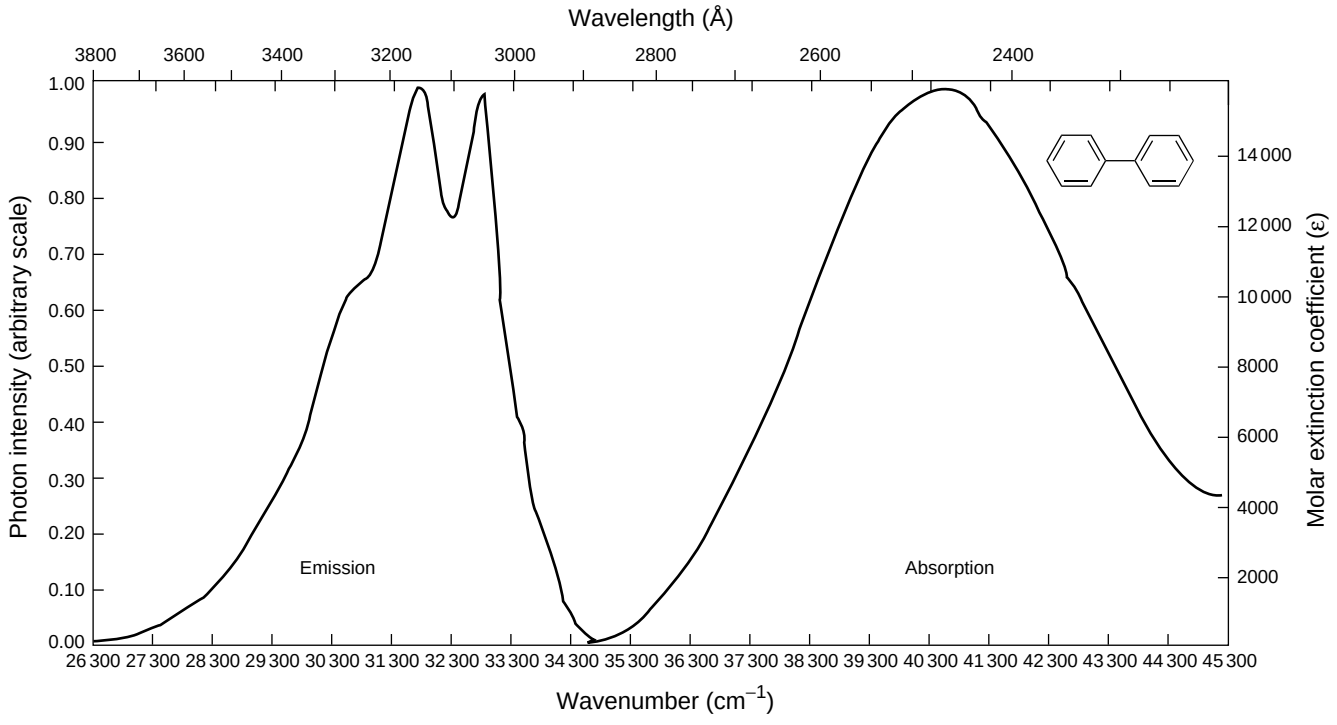

Figure 5 Absorption and emission spectra of biphenyl. (Reproduced with permission from Isadore B. Berlman, *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd edition, Academic Press, Orlando, 1971.)

Table 2 Quantum efficiencies of several fluorophores. (Reprinted from Ref. 8, pp. 15–16, by courtesy of Marcel Dekker, Inc.)

Compound	Solvent	Quantum efficiency
9-Aminoacridine	Water	0.98
	Ethanol	0.99
2-Methoxy-6-chloro-9-*N*-glycylacridine	Water	0.74
Anthracene	Benzene	0.29
	Hexane	0.33
	Ethanol	0.29
	Methyl methacrylate plastic	0.24
Fluorene	Ethanol	0.53
Fluorescein	0.1 M NaOH	0.92
Indole	Water, pH 7	0.65
Napthalene	Alcohol	0.12
	Hexane	0.10
Phenol	Water	0.22
Proflavin	0.05 M acetate, pH 4	0.34
Pyrene	Methyl methacrylate plastic	0.61
Quinine sulfate	0.5 M H_2SO_4	0.55
Riboflavin	Water, pH 7	0.26
Sodium salicylate	Water	0.28

quantum yield is simply the ratio of quanta of light emitted to quanta of light absorbed (Table 2). The quantum yield is a parameter that is affected by molecular properties as well as the solvent and environment surrounding the molecule. It has also been shown that temperature as well as excitation wavelength can have a marked effect on the quantum efficiency.[8]

Let us now consider the factors that affect fluorescence yield. The first of these is the process of absorption. The rate of formation of excited-state molecules by the absorption process is directly proportional to the molar absorptivity (ε) of the molecule, the intensity of the incident light (I_o), and the concentration of molecules in the ground state [S_0]. The intensity of the absorbed light (I_a) is given by Equation (12)

$$I_a = I_o - I_t \tag{12}$$

where I_t is the intensity of the transmitted light and can be considered the rate of absorption. If the sample is excited for a period of time that is long, compared to the rates of deactivation processes, the number of molecules in the excited state (S_1) will reach a steady state. The activation (absorption) rate will then be dependent on the rate of the deactivation processes, as this determines the population of molecules in the ground state. Therefore, under steady-state illumination the rate of absorption (Γ_a) is related to the rate of deactivation, i.e. the rate constants of the deactivation pathways, as shown in Equation (13):

$$\Gamma_a = (k_{vr} + k_{ic} + k_{isc} + k_f + k_Q[Q])[S_1] \tag{13}$$

Table 3 Representative equations and rate constants for processes involved in photoluminescence

Process	Equation	Rate constant
Absorption	$S_0 + h\nu \longrightarrow S_1$	k_a
Vibrational relaxation	$S_1\nu_3 \longrightarrow S_1\nu_0$	k_{vr}
Internal conversion	$S_2 \longrightarrow S_1$	k_{ic}
Fluorescence emission	$S_1 \longrightarrow S_0 + h\nu$	k_f
Intersystem crossing	$S_1 \longrightarrow T_1$	k_{isc}
Phosphorescence emission	$T_1 \longrightarrow S_0 + h\nu$	k_p
Collisional quenching	$S_1 + Q \longrightarrow S_0 + Q + \text{heat}$	k_Q

where k_{vr}, k_{ic}, k_{isc}, k_f, and k_Q are the rate constants for vibrational relaxation, internal conversion, intersystem crossing, fluorescence, and quenching, respectively. The term $[S_1]$ is the concentration of species that occupy the S_1 energy state under steady-state conditions and [Q] is the concentration of quencher molecules. These rate constants and the processes they represent are summarized in Table 3.

As previously mentioned, the quantum efficiency (Φ_f) is the fraction of absorbed photons that result in photon emission. The rate of fluorescence is therefore proportional to the rate of absorption as shown by Equation (14):

$$\Gamma_f = \Phi_f I_A \tag{14}$$

and Equation (15) gives:

$$\Gamma_f = k_f[S_1] = \left(k_{vr} + k_{ic} + k_{isc} + k_f + k_Q[Q]\right)[S_1] \quad \Phi_f \tag{15}$$

The fluorescence quantum efficiency is then given by Equation (16):

$$\Phi_f = \frac{k_f}{k_f + k_d} \tag{16}$$

where k_f is the rate of deactivation via fluorescence and k_d is the rate constant for all other deactivation processes. The rate constant k_d represents internal conversion (k_{ic}), vibrational relaxation (k_{vr}), intersystem crossing (k_{ic}), and collisional quenching (k_Q).

3.6 Linearity of Fluorescence

Much of the utility of fluorescence spectroscopy comes from the linearity of the fluorescence intensity as a function of concentration. This linearity can often extend over a relatively broad concentration range, especially for fluorophores with high quantum efficiencies. It is useful to consider the factors that affect the intensity of a fluorescent sample. As mentioned previously, under steady-state illumination the intensity of the fluorescence signal is equal to the rate of light absorption multiplied by the quantum efficiency, as shown in Equation (17):

$$I_f = \Gamma_a \Phi_f = (I_o - I_t)\Phi_f \tag{17}$$

From the Beer–Lambert law, the absorbed light can be related to the molar absorptivity (ε), cell pathlength (b), and concentration (c) as shown in Equation (18):

$$\frac{I_t}{I_o} = e^{-\varepsilon bc} \tag{18}$$

Substitution of Equation (18) into Equation (17) results in Equation (19):

$$I_f = I_o(1 - e^{\varepsilon bc})\Phi_f \tag{19}$$

A Taylor expansion can then be used to expand the exponential portion of Equation (18), resulting in Equation (20):

$$I_f = I_o \quad 1 - \frac{\varepsilon bc}{2} + \frac{\varepsilon bc^2}{6} - \cdots \quad \Phi_f \tag{20}$$

When the absorbance of the solution is low (≤ 0.01), Equation (19) can be reduced to Equation (21):

$$I_f = \Phi_f I_o \varepsilon bc \tag{21}$$

This shows that the rate of fluorescence is directly proportional to the quantum efficiency (Φ), the intensity of the exciting radiation (I_o), the molar absorptivity (ε), the pathlength (b), and the concentration of the fluorophore (c). Given that the quantum efficiency, molar absorptivity, intensity of the incident beam, and the pathlength will all be constant during a given experiment, the intensity of the fluorescence signal is then directly related to the concentration of the fluorophore by Equation (22)

$$I_f = kc \tag{22}$$

where k is a constant for a given experiment.

3.7 Fluorescence Lifetime

The fluorescence lifetime is another characteristic parameter of a fluorophore in a given environment. This parameter represents the probability of finding a molecule in the excited state at a given time. This fundamental parameter can be derived by letting n_0 represent the number of molecules in the ground state. We will also assume that fluorescence emission is the only path for the excited molecule to return to the ground state. Therefore, at time t, the number of molecules emitting per second is

proportional to n_t, the number of molecules in the excited state. This can be written as shown in Equation (23):

$$\frac{dn}{dt} = -k_f n_t \tag{23}$$

where k_f is the fluorescence rate constant. Integrating Equation (23) gives Equation (24):

$$\frac{dn}{n_t} = \quad -k_f\, dt \tag{24}$$

The resulting dependence of the excited-state population on time is shown in Equation (25).

$$n = n_0 e^{-k_f t} \tag{25}$$

This equation represents the average time that a molecule will reside in the excited state.

The intrinsic fluorescence lifetime (τ_0) is defined as the time required for all but 1/e of the excited state molecules to return to the ground state. This does not specify the path of relaxation to the ground state and encompasses all relaxation processes, both radiative and nonradiative. The radiative, or observed, lifetime (τ) refers specifically to the decay in fluorescence intensity following the cessation of steady-state illumination and is given by Equation (26):

$$I = I_o e^{-t/\tau} \tag{26}$$

where I is the fluorescence intensity, I_o is the fluorescence intensity at $t = 0$ and τ is the observed fluorescence lifetime. In the absence of any nonradiative relaxation mechanisms, the fluorescence lifetime and the fluorescence rate constant are related by Equation (27):

$$\tau = \frac{1}{k_f} \tag{27}$$

From this expression, the quantum yield can be expressed in terms of the intrinsic and observed lifetime as shown in Equation (28)

$$\Phi_f = \frac{\tau}{\tau_0} \tag{28}$$

making it possible to determine one unknown quantity, if the other two are known.

It is worthwhile to note that the spontaneous emission of light is proportional to the molar absorptivity. This is due to the fact that, in most cases, the excited molecule will reach thermal equilibrium prior to emission and the transition probabilities will be similar for both the emission and absorption processes. Therefore, there is an empirical link between absorptivity and the lifetime of a given fluorophore. An estimation of the radiative lifetime can be obtained using Equation (29)

$$\frac{1}{\tau_0} = \varepsilon_{max} \times 10^4 \tag{29}$$

where ε_{max} represents the maximum value of the extinction coefficient of the major absorption band.[9] Although this relationship is dependent on many assumptions, it is a useful way of emphasizing the dependence of the intrinsic lifetime on the molar absorptivity. It should be noted that lifetimes estimated in this fashion are only valid for the intrinsic lifetime. If any nonradiative relaxation mechanisms are present, significant errors will occur and the observed lifetime will be shorter than the calculated lifetime. There are more rigorous methods of estimating the lifetime,[9] but the intent of the current discussion is to simply note the dependence of the fluorescence lifetime on the molar absorptivity.

Recalling that the fluorescence and nonradiative rate constants are k_f and k_d, respectively, the fluorescence lifetime, τ, can be given by Equation (30):

$$\frac{1}{\tau} = k_f + k_d \tag{30}$$

A straightforward conclusion of the above relationship is that both processes determine the lifetime only when they are both of the same order of magnitude. If one process is much faster than the other, the relaxation rate will solely be determined by the faster process.

Fortunately, the decay time does not depend on the concentration of the fluorophore. This parameter can then be used by the experimentalist to examine situations under which the concentration of the fluorophore cannot be held constant, such as in the case of photobleaching. This scenario makes steady-state fluorescence measurements unreliable for quantitative studies. In such cases, the fluorescence decay time provides a more reliable measure of the fluorescence properties.

3.8 Fluorescence Quenching

Fluorescence quenching is one of the more analytically useful processes that compete with the fluorescence process. Thus many analytical techniques have been developed on the basis of fluorescence quenching. In the most general sense of the term, fluorescence quenching refers to any process that leads to a reduction in the fluorescence intensity of a given system. A more rigorous definition states that the reduction in intensity must involve nonradiative deactivation from the excited state. Based on these definitions, there are two basic categories of fluorescence quenching. The first is dynamic quenching, which involves nonradiative deactivation of the excited state upon collision with another molecule. The second is referred to as static quenching and refers to

a phenomenon where complex formation in the ground state precludes the formation of an excited-state species. While both types of quenching lead to a decrease in the observed fluorescence intensity, they are fundamentally different and thus will be discussed separately.

3.8.1 Dynamic Quenching

Dynamic quenching occurs when a molecule collides with the fluorophore during the lifetime of the excited state. As a result of the collision, the fluorophore relaxes to the ground state. This causes a decrease in the fluorescence lifetime of the species as well as a decrease in the fluorescence intensity. In the absence of quenching, the fluorescence quantum efficiency, Φ_o, is given by Equation (31).

$$\Phi_o = \frac{k_f}{k_{ic} + k_{ics} + k_f} \tag{31}$$

When quenching processes are present, the quenching rate and the concentration of quencher must be included, as shown in Equation (32).

$$\Phi = \frac{k_f}{k_{ic} + k_{ics} + k_f + k_Q[Q]} \tag{32}$$

Combining Equations (31) and (32), the ratio of the quantum efficiency in the absence and presence of quencher is given by Equation (33).

$$\frac{\Phi_o}{\Phi} = 1 + \frac{k_Q[Q]}{k_{ic} + k_{ics} + k_f} = 1 + \tau_0 k_Q[Q] \tag{33}$$

Alternatively, Equation (33) can be used to express the fluorescence efficiency in terms of fluorescence lifetime:

$$\frac{\tau_0}{\tau} = 1 + \tau_0 k_Q[Q] = 1 + K_{SV}[Q] \tag{34}$$

where τ_0 and τ are the fluorescence lifetimes in the absence and presence of quencher and K_{SV} is the Stern–Volmer quenching constant. Equations (33) and (34) were first used by Stern and Volmer[10] in 1919 and the Stern–Volmer equation is usually used in the form given in Equation (35):

$$\frac{F_o}{F} = K_{SV}[Q] + 1 \tag{35}$$

Equation (35) shows that a linear relationship exists between the ratio of the fluorescence intensity in the presence and absence of quencher (F_o/F) and the concentration of quencher. Therefore, a plot of the ratio F_o/F as a function of [Q] will yield a linear plot, the slope of which is equal to the Stern–Volmer constant. A typical Stern–Volmer plot is illustrated in Figure 6.

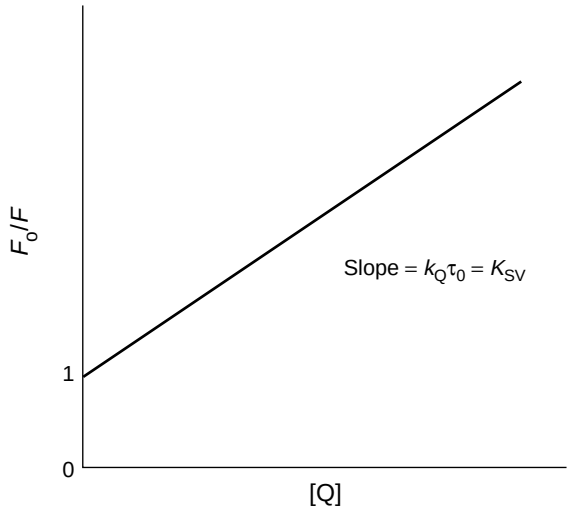

Figure 6 Idealized Stern–Volmer plot.

3.8.2 Static Quenching

Static quenching is a process that leads to a reduction in the observed fluorescence intensities, but not the quantum yield of fluorescence. Thus, the fundamental difference between dynamic and static quenching is that the quantum efficiency is not affected in static quenching. This is because static quenching leads to reduced fluorescence intensities by the formation of a ground-state complex. This ground-state complex precludes radiative deactivation to the ground state. The equilibrium of the complex can be given by Equation (36):

$$F + Q \longleftrightarrow F{-}Q \tag{36}$$

where F is the fluorescent species, Q is the quencher and F–Q is the complex. The equilibrium for this complex formation is given by Equation (37):

$$K_{eq} = \frac{[F{-}Q]}{[F][Q]} \tag{37}$$

where [F–Q] is the concentration of the complex. In the case where the complex is nonfluorescent, a useful relation can be developed. Let $[F_o]$ be the total concentration of fluorophore. The total fluorophore concentration is given by Equation (38):

$$[F_o] = [F] + [F{-}Q] \tag{38}$$

Substitution of Equation (38) into Equation (37) results in Equation (39):

$$K_{eq} = \frac{[F_o] - [F]}{[F][Q]} = \frac{[F_o]}{[F][Q]} - \frac{1}{[Q]} \tag{39}$$

which, upon rearrangement has the same form as the Stern–Volmer equation, as shown in Equation (40):

$$\frac{F_o}{F} = K_{eq}[Q] + 1 \tag{40}$$

However, in this form, the Stern–Volmer constant is replaced with K_{eq}, the equilibrium constant for the formation of the complex. This equation can be used to produce a plot analogous to that shown in Figure 6. This equation has led to the use of static quenching for the determination of formation constants.

In many experiments, it is important to determine whether the observed quenching is the result of dynamic or static quenching. There are several methods that can be used to distinguish between these two mechanisms of quenching. These methods are based on fundamental differences between dynamic and static quenching. The most conclusive method is to construct a Stern–Volmer plot using fluorescence lifetimes, rather than intensities. Since static quenching involves the formation of a ground-state complex which precludes emission from the excited state, no fluorescence is observed from the complexed fluorophore and the uncomplexed species are not affected. Thus, even though the intensities may change as a function of quencher concentration, the fluorescence lifetime will not. Hence, the Stern–Volmer plot will have a slope of zero in the case of static quenching. This is illustrated in Figure 7.

Another method that may be used to determine the form of quenching is to examine the temperature dependence of the quenching. With increased temperature, Brownian motion results in a higher rate of collisional encounters. This will have different effects on dynamic quenching and static quenching. In the case of dynamic quenching, increased collisional encounters lead to increased quenching. In the case of static quenching, the opposite effect is usually observed, since the stability of the complex will often decrease at higher temperatures. These effects are depicted in Figure 8.

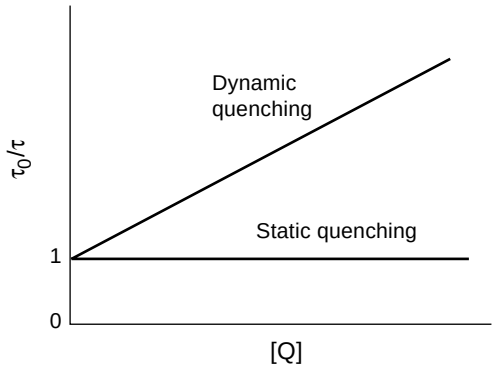

Figure 7 Idealized Stern–Volmer plots for dynamic and static quenching measured in terms of fluorescence lifetime.

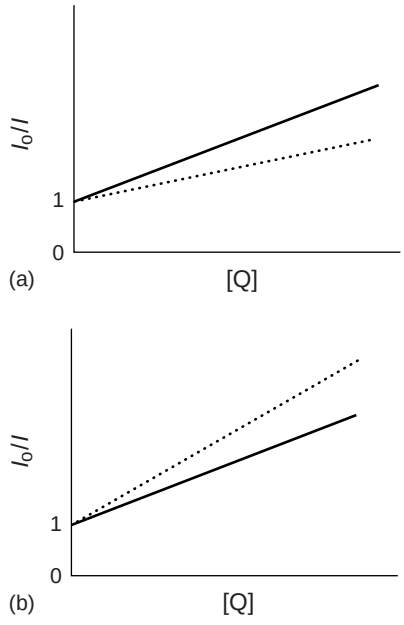

Figure 8 Change in Stern–Volmer plots for static (a) and dynamic (b) quenching in response to increased temperature (· · · ·).

3.9 Polarization

Fluorescence anisotropy is a technique that can elucidate the rotational motion of a fluorophore by measuring the polarization of the fluorescence emission. This technique can be utilized for analytical measurement in several ways, particularly since the size and shape of the fluorophore, the viscosity and temperature of the solvent environment, as well as binding interactions all influence the rotational diffusion of a molecule. This has led to the use of fluorescence anisotropy as a tool in the study of binding interactions,[11,12] microviscosities of membranes and micelles,[13–17] and conformational changes in DNA,[18] just to name a few. Most recently[19] fluorescence anisotropy has been used as a novel visual sensing device likely to find use in bioanalytical sensing applications.

When a fluorophore is excited with plane-polarized light, the subsequent emission from that molecule will be partially polarized. The degree of polarization depends on both intrinsic (molecular) and extrinsic (environmental) parameters. This dependence has led to the use of polarized fluorescence in the study of micellar and other microheterogeneous systems. Experimentally, the sample is excited with a plane-polarized beam of light, usually with vertical polarization. The intensity of the fluorescence emission is then measured after transmission of the light through a polarizer, which permits passage of vertically polarized light. This allows measurement of the parallel component (parallel to the exciting light) of the emission. The polarizer is then rotated 90° in order to

collect the perpendicular emission component. Together, these two measurements are used to represent the degree of polarization of the fluorescence emission.

The degree of polarization of light emitted from a fluorescent molecule can be described as either the fluorescence polarization (P), or the fluorescence anisotropy (r). These parameters are defined in Equations (41) and (42) respectively:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{41}$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{42}$$

where $I_{\parallel}$ and $I_{\perp}$ are the fluorescence emission intensities measured with polarization parallel and perpendicular to the polarization of the excitation radiation. For completely polarized light, both the polarization and anisotropy are equal to 1.0, and for nonpolarized light the anisotropy and polarization are equal to 0.0. At intermediate degrees of polarization, however, the anisotropy and polarization are not equal. As was first noted by Jablonski,[20] most theoretical expressions are much simpler when expressed in terms of anisotropy. A prime example of this is illustrated by considering the polarization measured from a mixture of fluorophores. The polarization of a mixture of fluorophores is given by Equation (43):

$$\left(\frac{1}{P_{\text{avg}}}\right)^{-1} = \sum_i \frac{f_i}{\left(\frac{1}{P_i} - \frac{1}{3}\right)} \tag{43}$$

where P_{avg} is the average polarization of the mixture, P_i is the polarization of component i, and f is the fractional contribution of the component to the overall intensity. This can be contrasted with the much simpler expression (Equation 44) obtained when the polarization is expressed as an anisotropy value.

$$r_{\text{avg}} = \sum_i f_i r_i \tag{44}$$

Similar simplification is found in most other expressions. The most significant observation is that the time-dependent anisotropy decays as a single exponential for spherical molecules, as shown in Equation (45):

$$r(t) = r_{\text{o}} \text{e}^{-t/\phi} \tag{45}$$

where r_{o} is the anisotropy at $t = 0$, and ϕ is the rotational correlation time for the molecule. Equation (45) will be discussed later in further detail.

For the reasons identified above, anisotropy will be used in this discussion, although polarization values are often encountered in the older literature. Importantly, the relations shown in Equations (46) and (47) can convert either value.

$$P = \frac{3r}{2 + r} \tag{46}$$

$$r = \frac{2P}{3 - P} \tag{47}$$

At this point, it is necessary to define a set of general coordinates which defines the discussion presented here. Fluorescence measurements are typically performed with an observation angle of 90°, and this configuration will be assumed throughout this discussion. With this configuration, the three primary instrumental components define the x,o,y-plane, where x is the direction of observation, o is the origin, or center of the irradiated volume, and y is the direction of the excitation beam (Figure 9).

These coordinates are used to define the molecular orientation and the polarization components of the excitation and emission with a common frame of reference. The fluorescence intensity measured at these coordinates is given by I_x, I_y, and I_z and the total fluorescence intensity, S, is given by Equation (48):

$$S = I_x + I_y + I_z \tag{48}$$

Equation (49) shows that when the excitation beam is vertically polarized along the z-axis,

$$I_x = I_y \tag{49}$$

and Equation (50) gives

$$S = I_z + 2I_x \tag{50}$$

Equation (51) shows that if the excitation beam is horizontally polarized along the x-axis, then

$$I_y = I_z \tag{51}$$

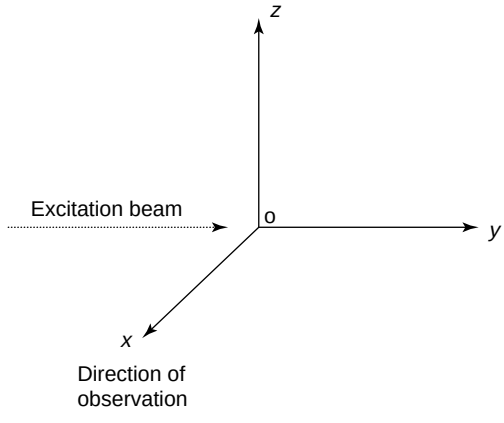

Figure 9 Operational coordinates used in fluorescence polarization studies.

and Equation (52) gives

$$S = 2I_z + I_x \tag{52}$$

Therefore, if the sample is excited with vertically polarized radiation, the emission component oriented along the z-axis corresponds to the parallel component ($I_\parallel$) and that in the direction of the y-axis corresponds to the perpendicular component ($I_\perp$). In order to selectively measure these two components of the emission, a polarizing device is placed in the optical path of the fluorescence emission. In this scenario, the emission intensity measured with the emission polarizer in a vertical position corresponds to the parallel component and is denoted by Equation (53)

$$I_{vv} = I_\parallel \tag{53}$$

where the ''v,v'' subscript denotes the vertical orientation of the excitation and emission polarizers. Measurement of the perpendicular component is denoted in a similar manner as shown in Equation (54)

$$I_{vh} = I_\perp \tag{54}$$

The fluorescence anisotropy is defined by Equation (55)

$$r = \frac{I_\parallel - I_\perp}{S} \tag{55}$$

which is sometimes represented by Equation (56)

$$r = \frac{D}{S} \tag{56}$$

where D simply denotes the difference between the parallel and perpendicular components of the emission. Most often, the fluorescence anisotropy is denoted by Equation (57)

$$r = \frac{I_\parallel - I_\perp}{I_\parallel + 2I_\perp} \tag{57}$$

illustrating that, fundamentally, two measurements must be made in order to determine the fluorescence anisotropy.

G-factor. In most cases, a correction called the G-factor must be made to anisotropy measurements. This is due to the fact that most fluorimeters and optics are biased toward a particular polarization of light. Additionally, even though the light source may produce nonpolarized light, some instrumental components (e.g. gratings) will impart some degree of polarization. The G-factor is simply a measure of the polarization bias of the instrument. This is typically determined by changing the excitation polarizer to a horizontal position. In this position, both horizontal and vertical polarizer positions are perpendicular to the polarization of the excitation beam. Therefore, any differences in the fluorescence intensity are due to instrumental artifacts. The G-factor is given by Equation (58)

$$G = \frac{I_{hv}}{I_{hh}} \tag{58}$$

and the corrected anisotropy is calculated using Equation (59).

$$r = \frac{I_\parallel - I_\perp G}{I_\parallel + 2I_\perp G} = \frac{I_{vv} - I_{vh}G}{I_{vv} + 2I_{vh}G} \tag{59}$$

3.9.1 Photoselection

A fundamental and critical aspect of the absorption of polarized light is the fact that only those molecules of a well-defined orientation will be excited. This is referred to as photoselection and is of utmost importance to understand fluorescence anisotropy. Recall that light consists of mutually orthogonal waves of electric and magnetic vectors propagating in the direction of the light path. Natural light that is unpolarized consists of photons whose electric field vectors are randomly orientated. When unpolarized light is passed through a polarizing element, only those photons with electric field vectors of one orientation are permitted to pass. Note that it is predominantly the oscillation of the electric field that has the greatest interaction with matter. Hence, when light is referred to as being vertically polarized, the orientation refers to the electric field vector. Absorption is greatest when the oscillation of the electric field is collinear with the orientation of the absorption dipole of the molecule. Therefore, those molecules which are properly aligned, relative to the polarization of the incident radiation, will preferentially absorb radiation and subsequently fluoresce. In other words, when the excitation is produced with polarized radiation, fluorescence emission results from those molecules which are ''photoselected'' during the excitation process.

3.9.2 Depolarization

3.9.2.1 Intrinsic Depolarization There are several mechanisms that lead to depolarization. These can be classified as intrinsic (molecular) and extrinsic (environmental) mechanisms. The two intrinsic mechanisms are photoselection and the angular separation between the excitation and emission dipoles. The major extrinsic form of depolarization is molecular rotation.

In dilute vitrified solutions, the extrinsic depolarization mechanisms (rotation and energy transfer) have been removed and any depolarization observed is due to intrinsic phenomena. Since fluorescence emission is only observed from molecules that have been ''photoselected'',

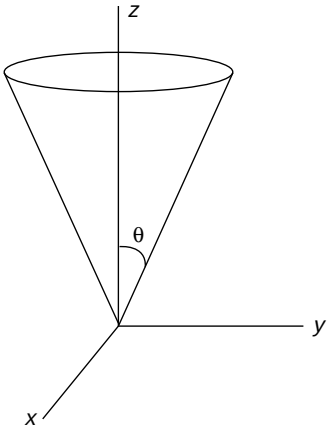

Figure 10 The process of photoselection.

one would expect the polarization to be preserved through the emission process, as in the case of light scattering. This is not observed, however, since the molecules need not be perfectly aligned in order to absorb radiation. A range of orientations is actually ''photoselected'' and this range is proportional to $\cos^2\theta$, where θ is the angle between the absorption dipole and the polarization of the exciting radiation (z-axis), as shown in Figure 10.

As a result, a cone of molecules randomly distributed about the z-axis will be selected. Thus, even in the absence of molecular rotation, emission from some molecules will not have completely vertical polarization and the emission will have components of both parallel and perpendicular orientations. The distribution of molecules excited is then given by Equation (60).

$$f(\theta)\,\mathrm{d}\theta = \cos^2\theta \sin\theta\,\mathrm{d}\theta \tag{60}$$

In the case of randomly distributed molecules that have collinear absorption and emission dipoles, the distribution suggested in Equation (60) can be reduced to Equation (61)

$$\overline{\cos^2\theta} = \frac{3}{5} \tag{61}$$

The dependence of the anisotropy on the orientation of the absorption dipole is then given by

$$r = \frac{3\overline{\cos^2\theta} - 1}{2} \tag{62}$$

where θ is the angle between the emission dipole and the z-axis. From Equations (61) and (62) it can be shown that the maximum anisotropy that can be obtained in an isotropic solution of fluorophores is 0.4. This value is observed when the absorption and emission dipoles are collinear. This scenario is rarely encountered and very few molecules have absorption and emission dipoles that are nearly collinear. The angular displacement between the absorption and emission dipoles is referred to as the α-angle and causes an additional decrease in the anisotropy.

Together, the α-angle and the process of photoselection define the intrinsic anisotropy, r_o, which is the anisotropy that would be measured in the absence of molecular rotation. The dependence of the intrinsic anisotropy on these processes is given by Equation (63).

$$r_o = \frac{2}{5}\left(\frac{3\cos^2\alpha - 1}{2}\right. \tag{63}$$

3.9.2.2 Extrinsic Depolarization Consider the case of a dilute isotropic solution of a fluorescent species. Once photoselection and excitation have occurred, the molecule will reside in the excited state for a period of time, prior to fluorescence emission. The amount of time that the molecule spends in the excited state is dependent on the fluorescence lifetime and the anisotropy will be reduced based on the degree of rotation that occurs during the excited state. From these considerations, it can be seen that there is a fundamental association between the measured anisotropy, the rotational rate of the molecule, and the fluorescence lifetime. These are related by the Perrin equation, as shown in Equation (64)

$$\frac{r_o}{r} = 1 + \frac{\tau}{\phi} \tag{64}$$

where r_o is the intrinsic anisotropy, r is the measured anisotropy, τ is the fluorescence lifetime and ϕ is the rotational correlation time of the molecule. By use of the Perrin equation, it is possible to determine the rotational diffusion rate of a molecule if the fluorescence lifetime and the intrinsic anisotropy are known. It is important to note that the measured rotational diffusion refers to that of the rotating species. In the case of simple fluorescent molecules, this corresponds to the rotation of the entire molecule. However, in the case of a large, rigid molecule containing a small fluorescent moiety, the rotational diffusion corresponds to the rotation of the entire assembly. More complicated scenarios are encountered when segmental motion exists between the fluorescent moiety and the rest of the molecule. These will be considered in more detail later.

Since the size of the molecule and the viscosity of the medium largely determine the rotational diffusion of a molecule, fluorescence anisotropy can be used to elucidate the molecular weight of fluorophores. Consider the case of a spherical molecule, where the rotary diffusion coefficient, D_o, is given by Equation (65)

$$D_o = \frac{kT}{6\eta V} \tag{65}$$

where k is the Boltzmann constant, T is the temperature in Kelvin, η is the viscosity of the solution and V is the volume of the rotating body, which is equal to the volume of the particle plus the volume of the hydration sphere. The rotation of molecules can be described by considering them as ellipsoids of revolution. This type of treatment is given elsewhere[21] and is beyond the scope of this discussion. The spherical particle can, however, be treated in a relatively simple manner. The depolarization can be ascribed to rotation about three primary axes, each with a rotational correlation time, ϕ_1, ϕ_2, and ϕ_3. The symmetry of the spherical particle dictates that all three rotational correlation times are identical and related to the rotational diffusion by Equation (66)

$$\phi_1 = \phi_2 = \phi_3 = \phi_o = \frac{1}{6D_o} \tag{66}$$

Combining Equations (65) and (66) shows that, in the case of spherical species, the rotational correlation time is given by Equation (67)

$$\phi_o = \frac{\eta V}{kT} \tag{67}$$

The discussion thus far has focused on a single molecule, but the previous equations can be applied to a quantity of molecules via Avogadro's number and the rotational correlation time is then given by Equation (68)

$$\phi = \frac{\eta V}{RT} \tag{68}$$

where R is the gas constant. The Perrin equation can then be written as Equation (69)

$$\frac{r_o}{r} = 1 + \frac{RT}{\eta V}\tau \tag{69}$$

More often, the Perrin equation is written as Equation (70)

$$r = r_o + \frac{\eta V}{RT}\tau r_o \tag{70}$$

which directly relates the measured anisotropy to the volume of the rotating species and the viscosity of the environment surrounding it.

3.9.3 Perrin Plot

A useful extension of the Perrin equation is based on the temperature/viscosity dependence of the anisotropy. The Perrin equation can be rearranged in the form of a linear equation as shown in Equation (71).

$$\frac{1}{r} = \frac{R\tau}{Vr_o}\frac{T}{\eta} + \frac{1}{r_o} \tag{71}$$

The Perrin plot is then constructed by plotting the reciprocal anisotropy as a function of the temperature and viscosity. The y-intercept of this plot, which represents anisotropy at infinite viscosity, yields the intrinsic anisotropy. Furthermore, the slope of this plot is given by Equation (72):

$$\text{slope} = \frac{R\tau}{Vr_o} \tag{72}$$

The Perrin plot is illustrated in Figure 11.

Experimentally, the Perrin plot is constructed either by varying the temperature of the solutions over a range of values, or by holding the temperature constant and varying the viscosity. In this manner, the intrinsic anisotropy and the molecular volume can be determined.

3.9.4 Association Reactions

Fluorescence anisotropy can be used to monitor association reactions. This is due to the dependence of the anisotropy on the size of the rotating species, as is evident from the volume term in Equation (68). Consider a system where a fluorophore may associate with a larger molecule, an example of which is the binding of proflavin to DNA. Upon binding, the rotational correlation time of the fluorophore will change substantially and be approximated by the rotational correlation of the much larger DNA molecule. This can be easily followed by use of fluorescence anisotropy. In fact, the degree of binding is directly related to the measured anisotropy. Equation (44) shows that anisotropy values are additive, where the observed anisotropy is an average of the individual components, weighted by the fractional intensity of each species. Hence

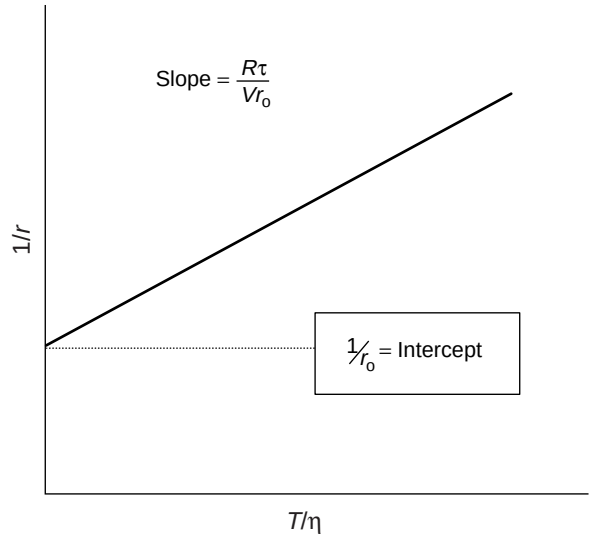

Figure 11 Idealized Perrin plot of a fluorescent molecule in a homogeneous solution.

Equation (73) shows that for a system which consists of a fluorophore that can bind to another system,

$$r_{obs} = f_f r_f + f_b r_b \tag{73}$$

where f_b and f_f represent the fractional intensity of the free and bound species, and r_b and r_f are the anisotropy of the two species. In many cases, the anisotropy of the free and bound fluorophore can be determined experimentally by examining the fluorophore alone and in the presence of an abundance of the species to which it binds. In such cases, the fraction of fluorophore that is bound can be given by Equation (74).

$$f_b = \frac{r_{obs} - r_f}{r_b - r_f} \tag{74}$$

Equation (74) assumes that the quantum efficiency does not change upon binding. If the quantum efficiency does change a more complicated expression that accounts for the quantum efficiency must be employed.

3.9.5 Time-dependent Anisotropy

When a sample is excited with a pulse of light, rather than steady-state illumination, the degree of polarization will decrease as a function of time. The decay of the fluorescence anisotropy is given by Equation (75)

$$r(t) = r_o e^{-t/\phi} \tag{75}$$

where the decay rate is dependent on the rotational correlation time and the maximum at $t = 0$ is equal to the intrinsic anisotropy. The potential utility of the technique should be apparent, as the intrinsic anisotropy and the rotational correlation time can be determined from a single experiment. Furthermore, a complex mixture consisting of either multiple fluorophores or a single fluorophore in multiple environments will have anisotropies that decay as a multiple exponential as shown in Equation (76)

$$r(t) = r_o \sum_i e^{-t/\phi_i} \tag{76}$$

where ϕ_i is the rotational correlation time of the ith component.

It should be noted that the anisotropy decay is not dependent on the fluorescence lifetime. This is a fundamental difference between time-dependent and steady-state anisotropy measurements. Recall from the Perrin equation (Equation (64)) the dependence of the steady-state anisotropy on the fluorescence lifetime. Consider the case of the fluorescent probe pyrene, which typically has a fluorescence lifetime in the order of 200 ns. Since the fluorescence lifetime is much longer than the rotational correlation time, the emission will be completely depolarized. In contrast, useful information can be obtained from such systems by measuring the anisotropy decay.

Experimentally, the anisotropy decay is obtained by collecting the parallel and perpendicular components as a function of time. Hence, the anisotropy is calculated using Equation (77).

$$r(t) = \frac{I_\parallel(t) - I_\perp(t)G}{I_\parallel(t) + 2I_\perp(t)G} \tag{77}$$

Figure 12 shows typical polarized fluorescence decays and the calculated anisotropy decay.

A major advantage of determining time-dependent anisotropy is that the anisotropy decay can reveal whether the fluorophore is able to rotate freely or whether its molecular motion is restricted. This can be the case when a fluorescent probe is dissolved in a highly ordered membrane system, or a larger molecule experiences segmental motion. In these cases, the anisotropy will decay to a nonzero value as shown in Figure 13.

The nonzero value to which the anisotropy decays is referred to as the limiting anisotropy (r_∞) and is described by Equation (78).

$$r(t) = (r_o - r_\infty)e^{-t/\phi_i} + r_\infty \tag{78}$$

Equation (78) assumes a simple hindered rotor that is symmetrical, but often satisfactorily explains more complex species. A useful equation, which relates the limiting anisotropy to the degree of rotational hindrance,

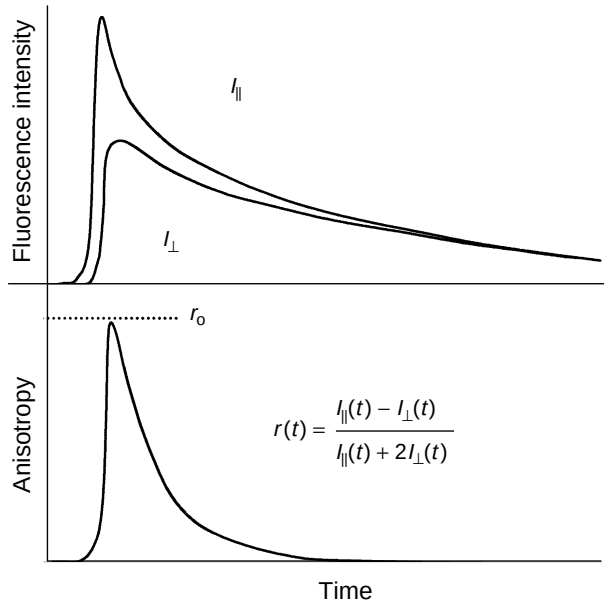

Figure 12 Measurement (above) and calculation (below) of time-resolved fluorescence anisotropy.

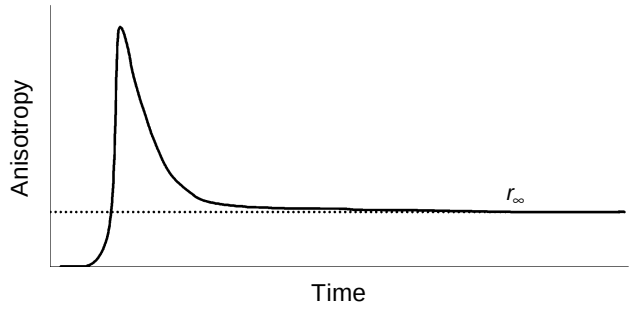

Figure 13 Time-resolved anisotropy of a system exhibiting a limiting anisotropy.

is shown in Equation (79)

$$\frac{r_\infty}{r_o} = \frac{3\cos^2\theta - 1}{2}^2 \quad (79)$$

where θ is referred to as the order parameter. See the references cited in 'Further Reading' for more sophisticated treatments that have been utilized.

4 BASIC FLUORESCENCE INSTRUMENTATION

4.1 General Instrumentation

In the simplest of cases, a fluorimeter consists of three components. These include a light source, a sample holder, and a detector. As an illustrative example, consider the apparatus Stokes used to examine quinine solutions. Sunlight served as the source, a glass tube held the solution of quinine and his eye served as the detector. It should be noted, however, that this was a special case where the fluorescence was easily visible because the excitation wavelength was in the ultraviolet and the fluorescence emission was easily visible as blue light. In most cases, a method of wavelength discrimination is employed to select both the excitation and observation wavelengths. Therefore, practically all fluorimeters will have at least five components, which include the source, two wavelength discriminators, a sample holder, and a detector (Figure 14).

While specific configurations may vary and new technology is continually incorporated in commercial instruments, the fundamental components are invariant and will be covered in this discussion. For fear of becoming dated, we have refrained from discussing specific commercial instruments and will instead introduce the basic components and list the various technologies that are commonly employed.

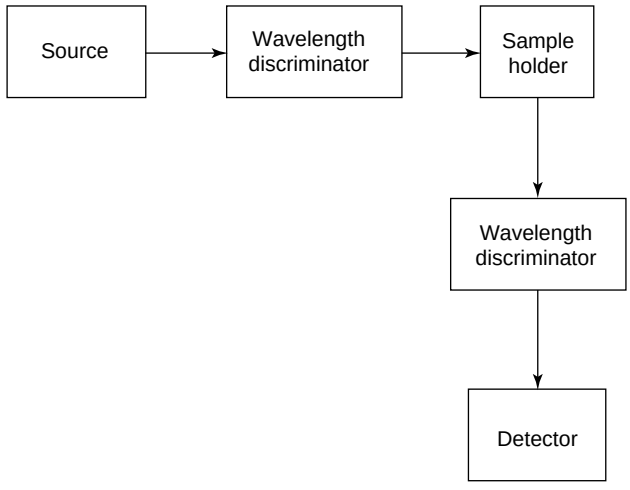

Figure 14 Schematic diagram of a basic fluorescence spectrophotometer.

4.2 Light Sources

There are two primary considerations when choosing a light source, the range of wavelengths it produces and the intensity at which it produces them. A source can produce a spectrum with characteristics of either a continuous or line emission. Most sources produce light with a broad continuous emission with some spectral lines superimposed, a prime example being the xenon lamp.

4.2.1 Xenon Lamp

The xenon lamp is probably the most commonly used source in fluorescence spectrophotometers. It produces a relatively continuous emission ranging from ~200 nm up to wavelengths in the infrared region, although the intensity diminishes significantly below ~270 nm. The spectrum of a xenon lamp is shown in Figure 15.

Figure 15 illustrates the wavelength dependence of the lamp intensity, which results in a distorted excitation

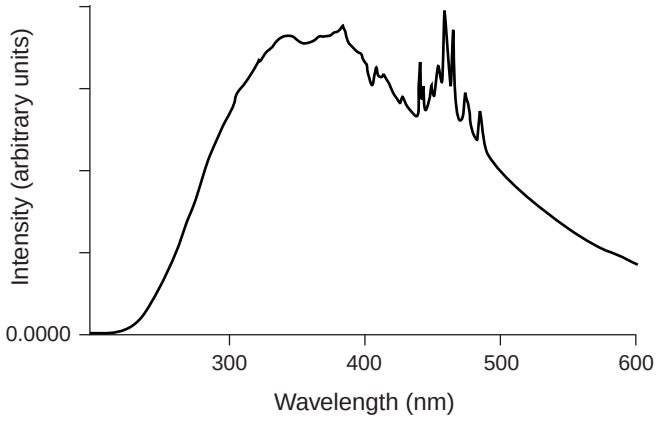

Figure 15 Emission spectrum of a typical xenon lamp.

spectrum. Corrections to account for this dependence will be discussed in section 4.5. Less expensive commercial instruments will often use a 150 W xenon lamp, while more expensive research grade instruments typically benefit from lamps rated from 450 to 750 W. The more powerful lamp directly affects the sensitivity of the instrument and can lead to extremely low limits of detection.

4.2.2 Mercury Lamps

Mercury lamps are typically more intense in the ultraviolet region than xenon lamps, but they only produce light at certain spectral lines. Therefore, these lamps can only be used in cases where the spectral output of the lamp is coincident with the excitation requirements of the fluorophore. Thus, mercury lamps typically find use in specialized applications, not in a general use fluorimeter. Another lamp, although not commonly employed, is the xenon–mercury lamp, which is a mercury lamp produced with amounts of xenon gas present. The ultraviolet output is greater than the xenon lamp, but the spectral output is not as smooth.

4.2.3 Lasers

Lasers are finding increased use as excitation sources in fluorescence instrumentation. While the cost of a laser is usually a significant consideration, the high power and spectral purity result in frequent use. A major drawback of a typical laser source is the limited range of wavelengths produced. Furthermore, the wavelengths produced are typically in the red or infrared region (Nd : yttrium aluminum garnet 1064 nm, Ti : sapphire 600–1000 nm). This is a region not useful for fluorescence excitation. The laser light is typically modified in the following manner. The fundamental wavelength of the laser is frequency doubled (or tripled) to produce light in the UV/VIS region. This is then used as an excitation source for a second dye laser, which will produce light of a longer wavelength possessing various degrees of tunability, depending on the laser dye that is used. Lastly, the output from the dye laser is frequency-doubled and tuned to the desired wavelength.

Even though much of the intensity is lost during the various conversions, the resulting excitation beam can offer significant advantages over conventional sources. A significant advantage of using a laser source is the fact that many lasers can operate in a pulsed mode, making them an ideal source for measurements in the time domain. Many laser systems are now routinely capable of producing pulses of a few picoseconds and some commercial systems can produce pulses as short as 100 fs. While a full discussion of laser sources is beyond the scope of this text, the reader should be aware of their potential applicability, especially since some relatively inexpensive dye and solid-state diode lasers are now becoming commercially available.

4.2.4 Flash Lamps

One of the most routinely used pulsed sources for fluorescence lifetime measurements is the flash lamp. A flash lamp consists of a lamp filled with gas and two electrodes that are used to produce an electrical discharge through the gas. Flash lamps typically produce lamp pulses of a few nanoseconds and the spectral output depends on the type of gas that is used. The most common gas used is nitrogen, although hydrogen is often used when shorter wavelengths are desired. The discharge of the electrodes is typically controlled with a thyratron tube, which discharges at a particular frequency independent of the capacitance of the lamp. This results in a flash that is very reproducible.

4.3 Wavelength Selectors

Wavelength selectors are used to selectively transmit a particular wavelength of light. The various technologies that are available can be categorized as either filters or monochromators. Filters have the ability to absorb certain wavelengths of light, but are fixed for a given filter. Monochromators have the ability to scan a region of the spectrum. Thus, monochromators must be used to obtain excitation or emission spectra. Both filters and monochromators have advantages for certain applications.

4.3.1 Filters

Optical filters are made from glass or quartz that has been produced with an amount of certain dyes that absorb certain wavelengths of light. Optical filters can be manufactured as either cut-on filters or cut-off filters. Cut-on filters absorb light up to a characteristic ''cut-on'' wavelength, above which they transmit radiation. Cut-off filters transmit light up to a certain wavelength and absorb wavelengths greater than the ''cut-off''. Figure 16 shows transmission curves for a series of filters that can be obtained from one of several manufacturers. Optical filters offer the significant advantages of high light throughput and low cost, but suffer from a lack of selectivity.

Interference filters consist of a coated dielectric material placed between two semitransparent films. Interference of the incident light will occur between the two films and only the wavelengths that interfere constructively will be transmitted through the filter. Interference filters offer the advantage of higher selectivity over optical filters, but still pass a relatively broad range of wavelengths and their transmission is poorer than that of optical filters.

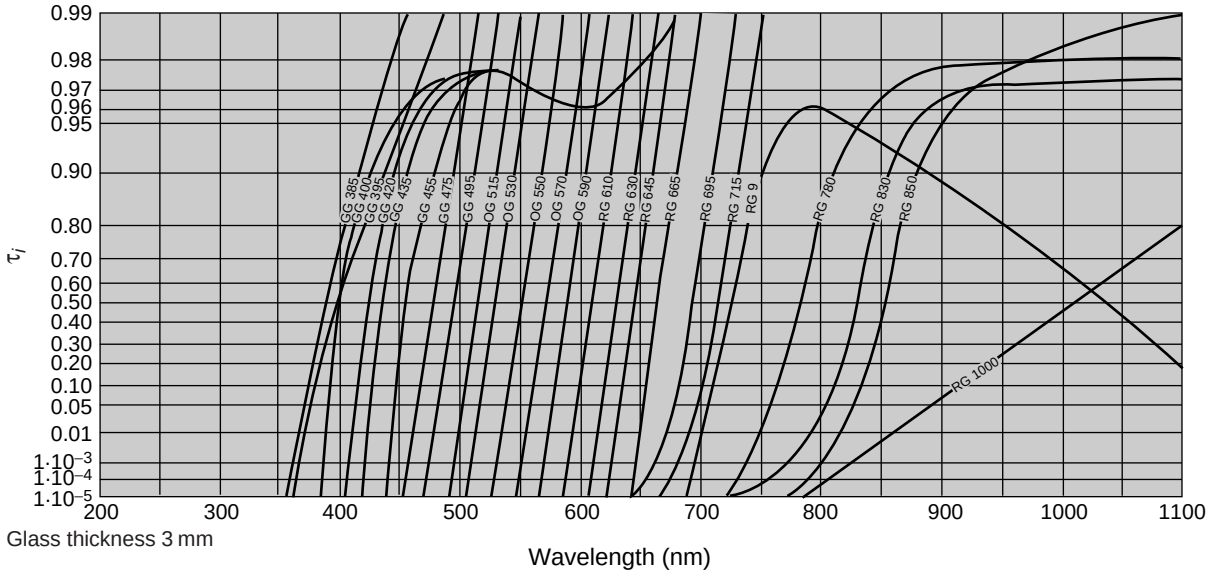

Figure 16 Transmittance curves of typical glass filters. (Reproduced with permission from Schott Glass Technologies Inc. product literature.)

4.3.2 Monochromators

The primary component of a monochromator is the dispersive element. Historically, a prism was used as the dispersive element, but today gratings are exclusively used. A monochromator consists of a grating and a set of slits to control the amount of light (and bandwidth) that enters and exits the monochromator. The type of grating and the geometry of the monochromator determine the spectral resolving power and the throughput of the monochromator. Even though the efficiency is usually higher in a system using optical filters, most commercial and research instruments are equipped with monochromators as wavelength-selective devices, due to their versatility.

4.4 Detectors

Although some systems are available that use multichannel detectors, such as photodiode arrays (PDAs) and charge-coupled devices (CCDs), most fluorescence spectrophotometers use photomultiplier tubes (PMTs) for detection. PMTs offer the advantage of inherent gain, where 10^6 to 10^9 electrons may be produced for each photon. The dark current of the PMT typically determines the detection limit of a fluorimeter. The dark current is the signal that is measured in the absence of light, due to a random nonphoton-induced cascade of electrons. Since the dark current is related to temperature, the PMT is often cooled, typically resulting in much lower limits of detection.

An additional consideration is the spectral response of the PMT, which is wavelength-dependent. Many types of PMTs are manufactured that are optimized for various wavelengths. Figure 17 shows the detector response curves for two commercial PMTs.

4.5 Spectral Correction

In an ideal fluorimeter, the measured excitation spectrum would be the same as the absorption spectrum. In reality

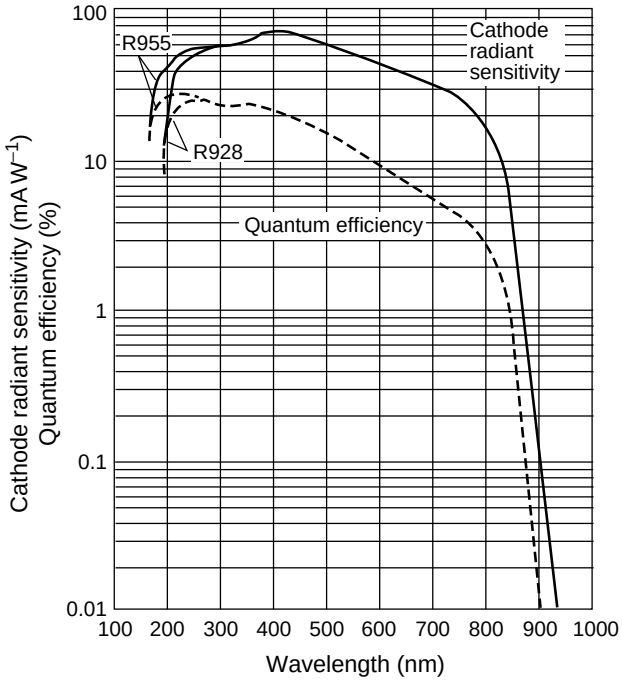

Figure 17 Spectral sensitivity and quantum efficiency of two commercially available PMTs. (Reproduced with permission from Hamamatsu Corporation product literature.)

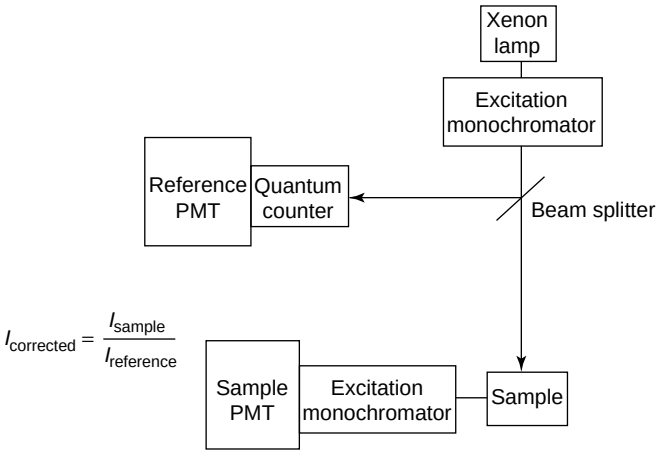

Figure 18 Schematic diagram of a fluorescence spectrophotometer utilizing a reference channel for spectral correction.

this is not the case, since most components of the fluorimeter are wavelength-dependent, particularly the source. The profile of the excitation spectrum will be significantly lower in the ultraviolet region, due to the decreased intensity of the xenon source. Other instrumental artifacts will also be present due to bias in the throughput of the monochromators and the spectral sensitivity of the detector. The most common type of correction that is applied uses a reference signal, as shown in Figure 18.

In this type of correction, a small fraction of the excitation beam is split prior to the sample. This light is then used to excite a fluorescent ''quantum counter''. The quantum counter is a material that produces a signal proportional to the intensity of the incident light. One of the most commonly used quantum counters is rhodamine B. A concentrated solution of rhodamine B in ethylene glycol absorbs virtually all light from 220 to 600 nm and gives a fluorescence signal that is proportional to the intensity of the excitation source. The excitation spectrum can then be recorded using a ratio of the sample PMT signal to that of the reference PMT signal, effectively correcting for the bias in the excitation source.

Schemes also exist for correcting the emission spectra, although they are less often employed. They typically are based on the measurement of a known emission source, such as a tungsten filament. The measured response can then be modified to produce the known spectrum, thus generating a correction factor that can be subsequently applied to a measured emission spectrum. It should be noted that spectral correction is typically only necessary for the determination of quantum efficiencies and overlap integrals in energy transfer studies.

4.6 Measurement of Fluorescence Lifetimes

While a complete discussion of the instrumentation involved in the measurement of fluorescence lifetime is beyond the scope of this text, we wish to present the reader with a basic overview of the two main types of experiments that are used to measure fluorescence decays. More detailed information can be found in **Fluorescence Lifetime Measurements, Applications of**.

4.6.1 Time-domain Experiments

One of the most common configurations for measurement of lifetimes involves exciting the sample with a short pulse of light and monitoring the fluorescence intensity of the sample as a function of time. If the pulse is sufficiently short (i.e. a δ-pulse) the fluorescence signal will decay exponential from a maximum value at time zero, as shown in Equation (80):

$$I(t) = I_o e^{-t/\tau} \tag{80}$$

where I_o is the fluorescence intensity at $t = 0$, and τ is the fluorescence lifetime. Measurements are typically completed in one of three configurations, each of which will be discussed briefly.

4.6.1.1 Time-correlated Single-photon Counting In the time-correlated single-photon counting technique, the sample is illuminated with a pulse of light of short duration. This ensures that a given molecule will only be excited once during the pulse. At the instant a population of molecules are excited, they will randomly begin returning to the ground state by emitting a photon, the probability of which is described by the fluorescence lifetime (see section 3.7). Figure 19 shows schematically the time-correlated single-photon counting technique.

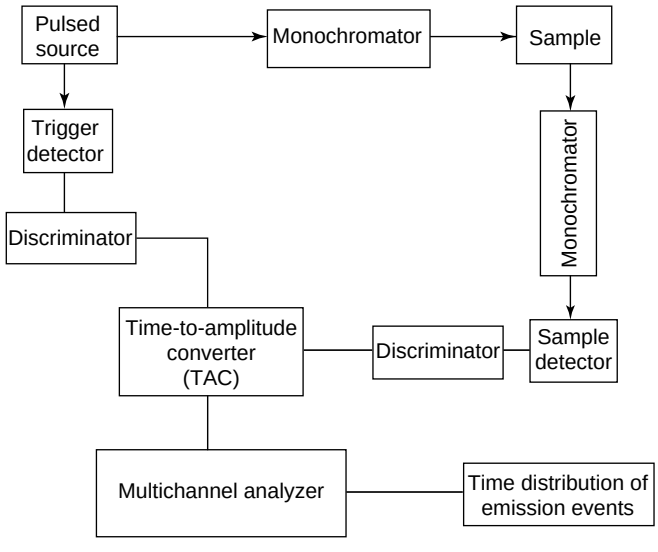

Figure 19 Schematic diagram of basic components involved in the time-correlated single-photon counting method.

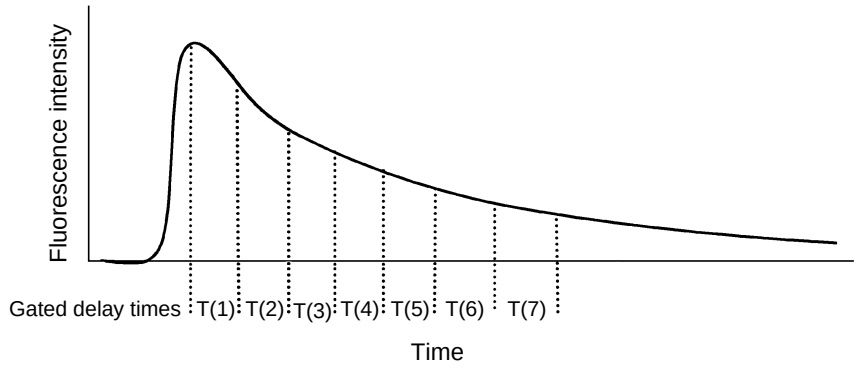

Figure 20 The pulse sampling method of collecting fluorescence decays.

When the source produces a pulse of light, the trigger detector sends a signal to start the time-to-amplitude converter (TAC). The TAC then linearly increases the voltage on a capacitor within the TAC until a ''stop'' signal is received from the sample detector. After the ''stop signal'' the voltage is measured on the capacitor, the magnitude of which correlates to the time of detection. The voltage is sent to a multi-channel analyzer, which stores the event in a histogram representing the delay time associated with the emission event. This process is repeated until a representative histogram is built up representing the time distribution of the emission events. The data can then be deconvolved (if necessary) and are typically analyzed with a least-squares treatment. Details of these procedures can be found elsewhere[22] and will not be covered in this discussion.

4.6.1.2 Pulse Sampling (Stroboscopic) Method In the pulse sampling method, the source is pulsed repetitively and the detector PMT is gated to operate for short periods of time, relative to the fluorescence lifetime. In this manner, the detector only measures fluorescence for a certain time-window during each cycle. The delay between the excitation pulse and the measurement window is varied, such that a profile showing the intensity as a function of time is collected. The process is illustrated in Figure 20.

An alternative method exists, where the PMT gain is held constant and the collection of data is gated. A sampling oscilloscope is used to measure the intensity of the PMT during very short time-windows and a multichannel analyzer is used to construct the delay curve. An advantage of this sampling technique is that much shorter sampling pulses can be measured (~30 ps).

4.6.2 Frequency-domain Experiments

In frequency-domain measurements, a steady-state source is used, but its intensity is varied in a sinusoidal manner. Since the intensity of fluorescence emission is directly related to the intensity of the source, the emission intensity will vary with the same frequency as the source. There is, however, a time lag between excitation and emission due the fluorescence lifetime. This time lag results in a phase shift (ϕ) between the oscillation of the source and the emission (Figure 21).

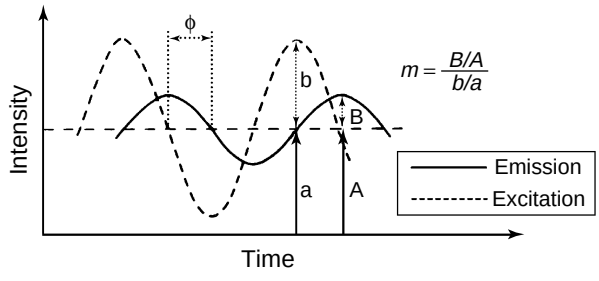

Figure 21 Measurement of phase-modulated fluorescence lifetime.

Additionally, the signal is demodulated (m) relative to the modulation of the excitation source. These parameters are related to the lifetime as shown in Equation (81)

$$\tan\phi = w\tau_{\mathrm{p}} \tag{81}$$

where w is the circular frequency of the modulation and τ_{p} is the phase lifetime. Additionally, the demodulation, m, is related to the lifetime as shown in Equation (82)

$$m = \left(1 + w^2\tau_{\mathrm{m}}^2\right)^{-1/2} \tag{82}$$

where τ_{m} is the demodulation lifetime. Equation (83) applies to single exponential decays.

$$\tau = \tau_{\mathrm{p}} = \tau_{\mathrm{m}} \tag{83}$$

It should be noted that more complicated treatments are needed for multiexponential decays, discussions of which can be found in several references.[23,24] See the article **Detectors, Absorption and Luminescence** for detailed discussion of instrumentation.

The use of luminescence in analytical chemistry has gained wide acceptance, due in part to the inherent sensitivity and selectivity of the technique.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
PDA	Photodiode Array
PMT	Photomultiplier Tube
TAC	Time-to-amplitude Converter
UV/VIS	Ultraviolet/Visible

FURTHER READING

G.G. Guilbault (ed.), *Practical Fluorescence*, 2nd edition, Modern Monographs in Analytical Chemistry, Marcel Dekker, New York, 1990.

J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum, New York, 1983.

J.R. Lakowicz (ed.), *Topics in Fluorescence Spectroscopy: Techniques*, Plenum, New York, Vol. 1, 1991.

J.R. Lakowicz, (ed.), *Topics in Fluorescence Spectroscopy: Principles*, Plenum, New York, Vol. 2, 1991.

C.A. Parker, *Photoluminescence of Solutions*, Elsevier, New York, 1968.

S.G. Schulman, *Fluorescence and Phosphorescence Spectroscopy: Physicochemical Principles and Practice*, Pergamon, Oxford, 1977.

REFERENCES

1. R. Boyle, *Experiments and Considerations Touching Colors*, London, 1680; Johnson Reprint Corp., New York, 1964.
2. D. Brewster, 'On the Colours of Natural Bodies', *Trans. R. Soc. Edinburgh*, **12**, 538 (1833).
3. J.F.W. Herschel, 'On a Case of Superficial Colour Presented by a Homogeneous Liquid Internally Colourless', *Phil. Trans.*, 143 (1845).
4. G.G. Stokes, 'On the Change of Refrangability of Light', *Phil. Trans.*, **142**, 463 (1852).
5. T.C. O'Haver, 'The Development of Luminescence Spectrometry as an Analytical Tool', *J. Chem. Educ.*, **55**, 423–428 (1978).
6. F. Goppelsröder, 'On the Application of Optical Properties to Detection and Determination of Organic Substances', *J. Prakt. Chem.*, (1867).
7. A. Jablonski, 'Uber den Mechanismus des Photolumineszenz von Farbstoffphosphoren', *Z. Phys.*, **94**, 38–46 (1935).
8. G.G. Guilbault (ed.), *Practical Fluorescence*, 2nd edition, Marcel Dekker, New York, 1990.
9. C.A. Parker, *Photoluminescence of Solutions*, Elsevier, Amsterdam, 1968.
10. O. Stern, M. Volmer, 'Über die Abklingungszeit der Fluoreszenz', *Phys. Z.*, **20**, 183–188 (1919).
11. A.B. Rawitch, G. Weber, 'The Reversible Association of Lysozome and Thyroglobulin', *J. Biol. Chem.*, **10**, 680–685 (1972).
12. S.A. Levison, 'Fluorescence Polarization Kinetic Studies of Macromolecular Reactions', in *Biochemical Fluorescence Concepts*, eds. R.F. Chen, H. Edelhoch, Marcel Dekker, New York, 375–408, Vol. 1, 1975.
13. M. Shinitzky, A.C. Dianoux, C. Gilter, G. Weber, 'Microviscosity and Order in the Hydrocarbon Region of Micelles and Membranes Determined with Fluorescent Probes. I. Synthetic Micelles', *Biochemistry*, **10**, 2107–2113 (1971).
14. R. Parthasarathy, M.M. Labes, 'Determination of Microviscosity and Micropolarity of Lyomesophases Utilizing a Fluorescent Probe', *Langmuir*, **6**, 542–547 (1990).
15. G. Komaromy-Hiller, R. von Wandruszka, 'Anisotropy Changes of a Fluorescent Probe during the Micellar Growth and Clouding of a Nonionic Detergent', *J. Colloid Interface Sci.*, **177**, 156–161 (1996).
16. R.E. Dale, L.A. Chen, S. Roth, L. Brand, 'Rotational Relaxation of the ''Microviscosity'' Probe Diphenylhexatriene in Paraffin Oil and Egg Lecithin Vesicles', *J. Biol. Chem.*, **252**, 7500–7510 (1977).
17. U. Cogan, M. Shinitzky, W. Weber, T. Nishida, 'Microviscosity and Order in the Hydrocarbon Region of Phospholipid and Phosphopholipid–Cholesterol Dispersions Determined with Fluorescent Probes', *Biochemistry*, **12**, 521–528 (1973).
18. N.F. Ellerton, I. Isenberg, 'Fluorescence Polarization Study of DNA–Proflavine Complexes', *Biopolymers*, **8**, 767–786 (1969).
19. I. Gryczynski, Z. Gryczynski, J.R. Lakowicz, 'Polarization Sensing with Visual Detection', *Anal. Chem.*, **71**, 1241–1251 (1999).
20. A. Jablonski, 'On the Notion of Emission Anisotropy', *Bull. Acad. Pol. Sci.*, **8**, 259–264 (1960).
21. R.F. Steiner, 'Fluorescence Anisotropy: Theory and Applications', in *Topics in Fluorescence Spectroscopy*, ed. J.R. Lakowicz, Plenum, New York, 1–52, Vol. 2, 1991.
22. D.J.S. Birch, R.E. Imhof, 'Time-domain Fluorescence Spectroscopy using Time-correlated Single-photon Counting', in *Topics in Fluorescence Spectroscopy*, ed. J.R. Lakowicz, Plenum, New York, 1–182, Vol. 1, 1991.
23. J.R. Lakowicz, I. Gryczynski, 'Frequency-domain Fluorescence Spectroscopy', in *Topics in Fluorescence Spectroscopy*, ed. J.R. Lakowicz, Plenum, New York, 293–336, Vol. 1, 1991.
24. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum, New York, 1983.

Absorption and Luminescence Probes

William E. Acree, Jr
University of North Texas, Denton, TX, USA

1	**Introduction**	1
2	**Ultraviolet/Visible Absorption Solvent Polarity Probes**	2
3	**Absorption and Luminescent Specific Ion and Molecular Recognition Probes**	5
4	**Ultraviolet/Visible Absorption and Fluorescent pH Probes**	7
5	**Fluorescent Wavelength-shift Solvent Polarity Probes**	9
6	**Fluorescent Twisted Intramolecular Charge Transfer Probes**	10
7	**Fluorescent Polycyclic Aromatic Hydrocarbon Solvent Polarity Probes**	12
8	**Fluorescent Anisotropic and Microviscosity Probes**	15
9	**Fluorescent Temperature Probes**	18
10	**Far-visible and Near-infrared Absorption and Luminescent Probes**	19
	Abbreviations and Acronyms	21
	Related Articles	22
	References	22

Absorption and luminescence probes are molecules or ions whose spectral properties depend upon the chemical nature of the solubilizing media in the immediate vicinity of the probe molecule. Probe character is generally evidenced by the appearance of new spectral bands, shifts in the absorption or emission wavelengths, or by changes in emission intensities. Probe molecules have been used to determine solvent polarity and solvent acidity/basicity, pH, microviscosity, temperature and/or presence of specific ions/molecules. When combined with fluorescence lifetime measurements, absorption and luminescence probes can provide additional information regarding diffusion coefficients in probe excimer formation, fluidity in lipid vesicles, and rates of intermolecular collisions in quenching experiments.

1 INTRODUCTION

Spectroscopic probe techniques are becoming increasingly popular for examining the surface microenvironment of chromatographic materials, surfactant micellization/adsorption, polymer/surfactant interactions, cyclodextrin cavities and inclusion complexes, sol–gel and inorganic glass formations, microemulsions and other molecularly organized solvent media. As typical examples, Street and Acree[1] estimated the effective dielectric constant of cyclodextrin (CD) on the basis of the measured fluorescence properties of pyrene-1-carboxyaldehyde [3029-19-4]. Stahlberg and Almgren[2] showed that the fluorescence properties of pyrene (Py) [129-00-0] could be used to investigate how the polarity of chemically modified silica surfaces varied with the surrounding solvent composition. For aqueous–methanol and aqueous–acetonitrile mixtures, the author noted large differences between the organic modifier interactions with RP-2 surfaces. Methanol decreased the polarity by hydrogen-bonding to the free silanol groups on the surface, whereas acetonitrile increased the polarity by absorbing to the alkyl chains and thus forming a layer of ''more or less freely moving'' organic molecules on the surface. Helburn et al.[3] subsequently studied the polarity and heterogeneity of reversed-phase chromatographic surfaces in equilibrium with aqueous–acetonitrile and aqueous–methanol solvent mixtures. Spectroscopic measurements using 2,6-dichloro-4-(2,4,6-triphenyl-*N*-pyridino)phenolate (ET-33) [121792-58-3] and *N*,*N*-diethyl-4-nitroaniline [2216-15-1] probe molecules indicated that the stationary phase is considerably more polar than bulk alkane solutions. Residual surface silanol groups strongly influenced the chemical and solvational properties of the interphase region.

Burrell and Hurtubise[4,5] employed solid-surface fluorescence methods and benzoquinoline probe molecules to investigate chromatographic interactions of polycyclic aromatic nitrogen heterocycles (PANHs) adsorbed onto water-deactivated silica gels. PANHs exist in both neutral and protonated forms, depending upon the acidity of the surrounding microenvironment. Protonation of a PANH molecule by a surface silanol results in the loss of fluorescence emission fine structure and a redshift in the emission wavelengths. From observed benzo[*h*]quinoline [230-27-3] emission spectra the authors were able to estimate the neutral to protonated form ratios of the various silica gel–water content compositions studied. Maximum localization of benzo[*h*]quinoline occurred near 5% water on surface sites that protonated the probe molecule.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The last three examples pertained to characterization of chromatographic stationary phases. The observed spectral response of the probe molecule was used to elucidate the chemical and physical properties of the solvated stationary phase. Spectroscopic studies have identified several hundred organic and organometallic compounds having absorption and fluorescent spectral properties that are highly sensitive to changes in solvent polarity, temperature, viscosity, solvent acidity/basicity and/or the presence of specific ions/molecules. Theoretically, any one of these compounds can serve as a spectroscopic probe molecule. From a practical standpoint, it is extremely time-consuming to fully characterize and understand a given molecule's spectral behavior. Consequently, only a relatively small number of compounds have actually been recommended as absorption and luminescence probe molecules. Reasonable efforts have been made to identify several suitable absorption probes and several suitable luminescence probes for each type of analytical application. Depending upon the sample being studied and instrumentation availability, a luminescence probe may be preferred to an absorption probe, and vice versa. A more in-depth discussion of probe methods, including the different types of measureable physical/structural parameters, can be found in a lengthy review article devoted to fluorescent probe molecules.[6]

2 ULTRAVIOLET/VISIBLE ABSORPTION SOLVENT POLARITY PROBES

Molecular interactions between a dissolved chromophore and surrounding solvent molecules often lead to noticeable changes in the chromophore's ultraviolet/visible (UV/VIS) absorption spectrum. Change in position, intensity or peak shape that occurs when a chromophore is transferred from gas phase to a solvent is referred to as solvatochromism. The most common form of solvatochromic behavior is the change in position of absorption band maxima whenever the chromophore is dissolved in different organic solvents or binary aqueous–organic mixtures having different composition. The direction of this wavelength shift generally depends upon relative magnitude of dipole moments of ground and excited states of the chromophore as shown in Figure 1.

Initially, the chromophore and surrounding solvent molecules are in the ground electronic level (S_0) and in the ground state configuration. Time averaged ground state configuration would have the dipole moments of solvational sphere solvent molecules aligned with the chromophore's dipole moment. Upon absorption of a photon, the chromophore is excited to the S_1 (or perhaps higher) electronic excited state, which is accompanied by

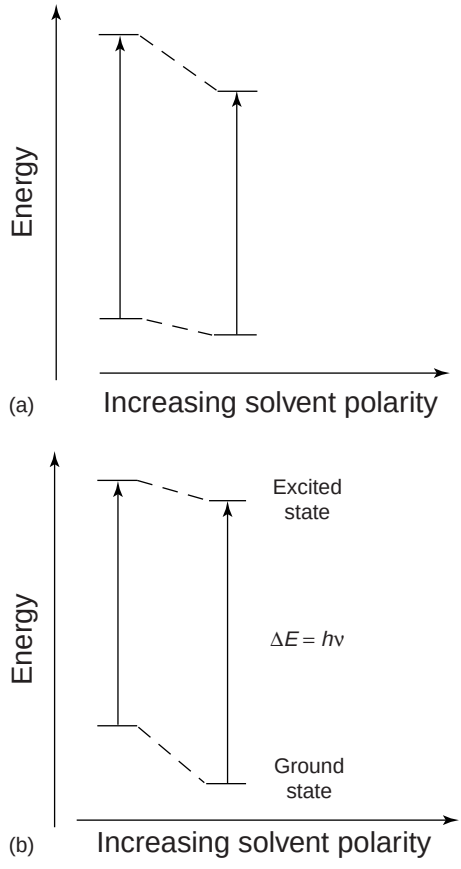

Figure 1 Schematic representation of solvent effects on the electronic transition of dipolar solutes. In the diagram (a) the excited state dipole moment is larger than the ground state dipole moment, $\mu^* > \mu$. In diagram (b) $\mu^* < \mu$. (Redrawn with permission from Lu and Rutan, *Anal. Chem.*, **68**, 1381–1386. Copyright (1996) American Chemical Society.)

a corresponding change in dipole moment. On the same time scale as the absorption process, there is a redistribution of electrons in solvational sphere molecules to better align with the chromophore's new dipole moment. There is insufficient time, however, for the physical reorientation of the solvational sphere molecules or for any change in solvational sphere composition. According to the Frank–Condon principle, electronic transitions are fast compared to nuclear motions and solvent molecule movements. Lack of molecular re-orientation for both absorption and emission spectral transitions leads to preferential solvation of ground or excited state, and a shift in the energy of the electronic transition. Positive solvatochromism is, by definition, a bathochromic shift (longer wavelength, lower energy) with increasing solvent polarity. Conversely, an hypochromic shift (shorter wavelength, higher energy) of absorption band with increasing solvent polarity is called negative solvatochromism.

At a more theoretical level, numerous mathematical expressions have been derived during the past 40 years

that relate the frequency of the absorption band maxima, ν_{max}, to the refractive index, n, and the dielectric constant, ε, of the solubilizing medium. Two examples of such treatments are the Bekarek and Bayliss models which are given by Equations (1) and (2), respectively.

$$\nu_{max} = A + \frac{n^2 - 1}{n^2} \quad \frac{\varepsilon - 1}{\varepsilon} \quad C \tag{1}$$

$$\nu_{max} = A + \left\{ n^2 - \frac{n^2 - 1}{2n^2 + 1} \right\} C \tag{2}$$

In real terms A is related to the position of the absorption maxima in the gas phase spectrum, and C incorporates several solute properties and physical constants. Intermolecular solute–solvent interactions are highly complicated in nature, and extremely difficult to understand. To date, no one single theoretical treatment satisfactorily explains the observed solvatochromic behavior that results whenever different chromophores are dissolved in various microenvironments.

Lack of a single, comprehensive theoretical expression for predicting solvatochromic wavelength shifts, combined with the inadequacy of defining solute–solvent interactions in terms of simple functions of solvent properties, led to the development of solvent polarity scales deduced from the spectral behavior of select chromophores. One of the most widely used empirical scales of solvent polarities is the $E_T(30)$ scale based upon Equation (3):

$$E_T(30)\ (\text{in kcal mol}^{-1}) = \frac{28\,591}{\lambda_{max}\ (\text{in nm})} \tag{3}$$

the charge-transfer absorption of the zwitterionic 2,6-diphenyl-4-(2,4,6-triphenyl-*N*-pyridinio)phenolate molecule (ET-30) [10081-39-7].[7,8] The zwitterionic compound, also known as the Dimroth and Reichardt's betaine, exhibits one of the larger observed solvatochromic effects of any known organic molecule. Wavelength shifts amount to several hundred nanometers in going from a very polar solvent ($\lambda_{max} \approx 453$ nm in water) to a nonpolar solvent ($\lambda_{max} \approx 925$ nm in hexane). In fact, the apparent color of solutions containing the ET-30 absorption probe molecule serves as a rough indicator of solvent polarity. Methanolic solutions of the dye molecule appear wine red ($\lambda_{max} \approx 515$ nm), whereas the apparent color changes to deep blue ($\lambda_{max} \approx 622$ nm) when the dye is dissolved in acetonitrile. Numerical values of the $E_T(30)$ solvent polarity parameter are calculated from the wavelengths of maximum absorbance using Equation (3). Similarly, the $E_T(33)$ scale is defined in terms of the spectral properties of the ET-33 molecule.

Other examples of spectroscopically deduced polarity measures are the Z and Z' scales suggested by Kosower.[9] These two scales are defined by Equation (4):

$$Z(\text{or } Z') = \frac{28\,591}{\lambda_{max}\ (\text{in nm})} \tag{4}$$

the charge-transfer absorption band of *N*-ethyl-4-(methoxycarbonyl)pyridinium iodide [1199-65-1] and 4-cyano-1-ethylpyridinium iodide [1622-95-3]. Practical applications of the Z scale are limited, however, to organic solvents and aqueous–organic mixtures having low water content. The position of the charge-transfer absorption peak becomes progressively more difficult to accurately evaluate in solvent mixtures with increasing water compositions. The charge-transfer peak begins to overlap with the pyridine ring absorption. *N*-Ethyl-4-methoxycarbonyl)pyridinium iodide is thus not a suitable spectroscopic probe molecule for characterizing the aqueous–organic mobile phase commonly used in reverse-phase high-performance liquid chromatography (HPLC).

The $E_T(30)$, $E_T(33)$, Z and Z' scales are known as single parameter scales in that the solvent is described with only a single experimentally determined parameter. Multiparameter approaches exist for expressing solvent polarity effects in terms of separate contributions from polarizability, and hydrogen-bond donor and acceptor interactions. For example, Kamlet, Taft and coworkers[10,11] introduced the π^* scale of solvent dipolarity/polarizability for representing solute–solvent interactions in the absence of strong forces such as hydrogen-bonding or ion–dipole interactions. Numerical values of π^* are calculated using the $\pi \rightarrow \pi^*$ absorption band of a series of solutes, such as 4-ethylnitrobenzene [100-12-9] or *N*,*N*-diethyl-4-nitroaniline. Table 1 lists several spectroscopic probes that have been recommended for π^* determinations, along with their corresponding mathematical equations. The α- and β- scales of hydrogen-bond acceptor and donor interactions have also been derived from solvatochromic measurements. A more complete list of UV/VIS absorption spectroscopic probes and solvent polarity scales is published elsewhere.[7,8]

Solvent polarity probes have been used in analytical chemistry to characterize chromatographic stationary and mobile phases, and to describe how solute retention times vary with organic modifier composition in the case of aqueous–organic mobile phase mixtures. Johnson et al.[13] reported that plots of logarithm of capacity factor, $\log k'_{solute}$, versus $E_T(30)$ parameter were better descriptors of solute retention in binary aqueous–acetonitrile and aqueous–methanol mobile phases than were $\log k'_{solute}$ versus percent organic modifier graphs. The authors' data analysis was summarized in the form of a frequency histogram of the squared linear correlation coefficient, r^2. The histogram for the $E_T(30)$ correlations encompasses approximately 98.2% of the

Table 1 Spectroscopic UV/VIS absorption probes used in π^* solvent polarity determinations

Probe[a]/π^* calculation[b]	Probe[a]/π^* calculation[b]
4-Nitroanisole $\pi^* = (34.12 - \nu_{max})/2.343$	4-Dimethylamino-β-nitrostyrene $\pi^* = (25.25 - \nu_{max})/3.354$
N,*N*-Diethyl-3-nitroaniline $\pi^* = (25.52 - \nu_{max})/2.214$	*N*,*N*-Dimethyl-4-nitroaniline $\pi^* = (28.10 - \nu_{max})/3.436$
4-Methoxy-β-nitrostyrene $\pi^* = (29.96 - \nu_{max})/2.250$	3,5-Dimethyl-4-nitroaniline $\pi^* = (28.68 - \nu_{max})/1.813$
1-Ethyl-4-nitrobenzene $\pi^* = (37.67 - \nu_{max})/2.259$	3-Methyl-4-nitroaniline $\pi^* = (31.40 - \nu_{max})/3.377$
N-Methyl-2-nitro-*p*-toluidine $\pi^* = (23.83 - \nu_{max})/1.632$	*N*,*N*-Dimethyl-2-nitroaniline $\pi^* = (25.30 - \nu_{max})/2.023$
Ethyl 4-nitroaminobenzoate $\pi^* = (33.31 - \nu_{max})/1.407$	*N*-(4-Nitrophenyl)aziridine $\pi^* = (32.11 - \nu_{max})/2.510$
4-Nitroaniline $\pi^* = (31.10 - \nu_{max})/3.138$	*N*-(4-Nitrophenyl)pyrrolidine $\pi^* = (27.56 - \nu_{max})/3.274$
N,*N*-Diethyl-4-nitroaniline $\pi^* = (27.52 - \nu_{max})/3.182$	*N*-(4-Nitrophenyl)piperdine $\pi^* = (27.93 - \nu_{max})/3.405$
4-Aminobenzophenone $\pi^* = (33.09 - \nu_{max})/1.682$	2-Nitroanisole $\pi^* = (32.56 - \nu_{max})/2.428$
3-Nitroaniline $\pi^* = (28.87 - \nu_{max})/1.664$	*N*-Methyl-2-nitroaniline $\pi^* = (24.59 - \nu_{max})/1.593$
Brooker's merocyanine $\pi^* = (17.74 - \nu_{max})/2.780$	*N*,*N*-Diethyl-3-methyl-4-nitroaniline $\pi^* = (27.69 - \nu_{max})/3.073$
Phenol blue $\pi^* = (18.12 - \nu_{max})/1.445$	*N*,*N*,3,5-Tetramethyl-4-nitroaniline $\pi^* = (27.36 - \nu_{max})/2.747$

[a] For a more complete listing of spectroscopic probes see Kamlet et al.[12]
[b] ν_{max} is the observed absorption maximum in wavenumbers ($\times 10^{-3}$ cm^{-1}).

332 data sets considered for $r^2 \geq 0.95$. Only 89.2% of the data sets are contained in the percent organic modifier histogram.

Carr and coworkers[14,15] applied the solvatochromic comparison method and linear solvation energy relationships in determining what solute chemical and physical characteristics govern retention in reversed-phase HPLC. The logarithm of the solute capacity factor was expressed as Equation (5):

$$\log k'_{solute} = \text{intercept} + M(\delta_s^2 - \delta_m^2)\frac{V_{solute}}{100} + S(\pi_s^* - \pi_m^*)\pi_{solute}^* + A(\beta_s - \beta_m)\alpha_{solute} + B(\alpha_s - \alpha_m)\beta_{solute} \quad (5)$$

where V_{solute} denotes that the molar volume of the solute, δ is the Scatchard–Hildebrand solubility parameter, π^* represents the polarizability/dipolarity parameter, and α and β refer to the hydrogen-bond acidity and basicity values. Each solute property is multiplied by a term that represents the difference in the complementary ''solvent'' properties for the mobile (m subscript) and stationary (s subscript) phases. The intercept and four curve-fit coefficients (M, S, A and B) are obtained through a multilinear least-squares regressional analysis. Numerical values of the coefficients determine which of the possible molecular interactions control solute retention. The more important contributions have the larger numerical coefficients. In the case of chromatographic retention behavior of a series of aromatic solutes in reversed-phase HPLC with a binary aqueous–acetonitrile mobile phase, Carr and coworkers found that solute size and hydrogen-bond basicity were the principle factors controlling retention. Multilinear correlations between pH and spectroscopically deduced solvent polarity parameters have been suggested for the standardization of potentiometric pH sensors in mixed aqueous–organic solvents.[16]

UV/VIS absorption probes have also been used to examine preferential solvation that occurs whenever a solute molecule is dissolved in a binary solvent mixture. Preferential solvation arises whenever the proportion of molecules of a given solvent component within the probe's solvational sphere is not equal to its bulk mole fraction composition. ''True preferential solvation'' is extremely difficult to theoretically model, and most published

approaches have assumed an idealized set of conditions in which the measured spectral response in the binary solvent mixture, $R_{\text{solute, AB}}$, is given by Equation (6):

$$R_{\text{solute,AB}} = X_A R_{\text{solute,A}} + X_B R_{\text{solute,B}} \quad (6)$$

a weighted mole fraction average of the probe's spectral response in the two neat solvents, $R_{\text{solute,A}}$ and $R_{\text{solute,B}}$. For UV/VIS absorption solvent polarity probes the spectral response is the reciprocal of the maximum absorbance wavelength, $1/\lambda_{max}$, and X_A and X_B refer to the mole fraction compositions of solvents A and B in the solvational sphere. Phillips and Brennecke[17] used Equation (6) to estimate the degree of preferential solvation around the solvatochromic probe phenol blue [2150-58-5]. Solvent mixtures investigated included acetone–cyclohexane, triethylamine–cyclohexane, ethyl butyrate–cyclohexane, cyclohexanone–cyclohexane, acetophenone–cyclohexane and acetophenone–toluene. In most cases spectroscopically obtained local compositions compared reasonably well with values calculated from thermodynamic models and measured solubility data. More sophisticated preferential solvation treatments take into account molecular size disparity and incorporate molar absorptivities for the different assumed solvational species.

3 ABSORPTION AND LUMINESCENT SPECIFIC ION AND MOLECULAR RECOGNITION PROBES

The idea of chemically modifying ionic or molecular analytes through complexation and/or derivatization in order to make the species more amenable to a particular spectroscopic analysis is not new. Over the past 50 years several hundred chelating ligands have been developed for the spectrophotometrical analysis of metal ions, and numerous derivatizing agents synthesized for spectrophotometric detection in conjunction with liquid chromatography and fluorescence microscopy. Probes have even been designed for select classes of biological molecules. For purposes of this review, such chelating agents and derivatizing agents are classified under the broad umbrella of absorption and fluorescent specific ion and molecular recognition probes.

Most chelating and derivatizing agents possess a chromophoric and/or fluorophoric moiety, which usually is an aromatic moiety. Such probes achieve their recognition specificity for the desired analyte either through selective complexation or chemical reaction. In the case of complexation, the ability of the probe to form molecular complex(es) is governed by the value of the formation equilibrium constant, which may depend upon analyte-probe molecular size/shape compatibility (i.e. crown ethers and cryptads). Derivatizing agents, on the other hand, achieve their selectivity through highly specific chemical reactions that occur between given functional groups on the analyte and probe molecule. A derivatizing agent thus reacts only with those molecules present in the mixture that contain the given functional group. Table 2 lists several of the common derivatizing agents that are used in HPLC to achieve enhanced absorption and fluorescence detection.

Absorption and luminescent recognition probes are often used in the construction of fiber-optic sensors for remote chemical detection. In the simplest version, the probe molecule is immobilized at the tip of an optical fiber. Light is transmitted from the radiation source to the sample through the fiber. The reflected radiation enters a second optical fiber and enables the changes in the spectral properties of the immobilized probe molecule in contact with the analyte solution to be recorded. The analyte concentration is calculated from the measured absorbance and/or fluorescent emission intensity. Specific ions that have been detected with fiber-optic chemical sensors include Cl^-, Br^- and I^- based upon fluorescence quenching of glass-immobilized quinoline [91-22-5], Be^{2+} and Al^{3+} using the immobilized fluorescent morin [480-16-0], and alkali metal ions using complexation-induced spectral changes of fluorescent crown ethers in poly(vinyl chloride) films.[21] Extensive compilations of complexing agents for the different metal ions can be found in published monographs[22–24] devoted to the subject.

Chemical sensors are also able to monitor the concentration levels of ammonia, carbon dioxide, oxygen, hydrogen cyanide and several other gaseous molecules. Chromophores employed in the absorbance-based NH_3 sensors include the acid–base indicators chlorophenol red [4430-20-0], bromocresol purple [115-40-2], bromocresol green [62625-32-5], bromothymol blue [76-59-5] and bromophenol blue [115-39-9]. Fluorescence intensity-based NH_3 sensors have contained acridine orange [10127-02-3], carboxy-4′,5′-dimethylfluorescein, 2,7-bis(2-carboxyethyl)-5-carboxyfluorescein [85138-49-4] and 8-hydroxypyrene-1,3,6-trisulfonic acid [27828-00-3] as probe molecules.[25] Similarly, cresol red [1733-12-6], phenol red [143-74-8], neutral red [553-24-2], bromothymol blue, brilliant yellow [3051-11-4], *m*-cresol purple [2303-01-7] and 3-nitrophenol [554-84-7] are used in CO_2 chemical sensors. The measured absorbance of the acid–base indicator is related to the CO_2 concentration in the sample.[26] Probes utilized in the NH_3 and CO_2 sensors actually respond to hydrogen ion, which is related to the analyte concentration through acid–base equilibria.

For example, in a typical ammonia gas sensor, the ammonia reacts with an acidic pH indicator according to

Table 2 Derivatizing agents[a] for enhanced absorption and fluorescence detection

Analyte	Derivatizing reagent	Detection
Enhanced UV/VIS absorption detection:		
Alcohols	Benzoyl chloride	UV
	p-Nitrobenzoyl chloride	UV
	p-Methoxybenzoyl chloride	UV
	3,5-Dinitrobenzoyl chloride	UV
Aldehydes and ketones	2,4-Dinitrophenylhydrazine	VIS
Amines	1-Fluoro-2,4-dinitrobenzene	UV
	Benzoyl chloride	UV
	p-Nitrobenzoyl chloride	UV
	p-Toluenesulfonyl chloride	UV
Amino acids	Phenylisothiocyanate	UV
Carboxylic acids	Phenacyl bromide	UV
	Benzyl bromide	UV
	O-4-Nitrobenzyl-*N*,*N*-diisopropylisourea	UV
	Aryl tolytriazines	UV
	2-Nitrophenylhydrazine hydrochloride	UV/VIS
Isocyanate monomers	4-Nitrobenzyl-*N*-propylamine	UV
Sugars	Phenyl isocyanate	UV
Enhanced fluorescence detection:		
Aldehydes and ketones	5,5-Dimethyl-1,3-cyclohexanedione	
	1,3-Cyclohexanedione	
	1-Dimethylaminonaphthalene-5-sulfonyl hydrazine	
1,2-, 1,3- and 1,4-Alkanediols	Phenanthrene boronic acid	
Amines	Dansyl chloride	
	o-Phthalaldehyde	
	NBD-Chloride	
	NBD-Fluoride	
	Sodium naphthoquinone sulfate	
	2-Methoxy-2,4-diphenyl-3(2*H*)furanone	
Tertiary amines	2-Naphthyl chloroformate	
Amino acids	Dansyl chloride	
	o-Phthalaldehyde	
Carboxylic acids	1-Bromoacetylpyrene	
	4-Bromomethyl-7-methoxycoumarin	
	9-(Chloromethyl)anthracene	
Sugars	Dansyl hydrazine	
Thiols	Dansylaziridine	
	N-Substituted maleimides	
	Halogenosulfonylbenzofurazans	

[a] For a more complete listing of derivatizing reagents see: Goto[18]; Lingeman and Underberg[19]; Meulendijk and Underberg[20].

the reaction in Equation (7):

$$HIn + NH_3 \rightleftharpoons In^- + NH_4^+ \quad (7)$$

The unprotonated form of the indicator is detected by either a fluorescence or an absorbance measurement. In the latter case, the measured absorbance, A, is related to the ammonia concentration in the sample solution by Equation (8):

$$A = \frac{\varepsilon b K_{eq} C_{In} [NH_3]_{sample}}{C_{NH_3} + K_{eq} [NH_3]_{sample}} \quad (8)$$

where ε denotes the chromophore's molar absorptivity, b is the effective pathlength at the tip of the fiber optic probe, K_{eq} represents the conditional equilibrium constant for proton transfer, $[NH_3]_{sample}$ is the ammonia concentration in the sample, and C_{In} and C_{NH_3} refer to the stoichiometric molar concentrations of the indicator and ammonia, respectively. Inspection of Equation (8) reveals that a nearly linear response occurs at low ammonia concentrations whenever $C_{NH_3} \gg K_{eq}\ [NH_3]_{sample}$. The calibration curve is linearized as a double reciprocal $1/A$ versus $1/[NH_3]_{sample}$. A similar expression holds in the case of a fluorescent probe molecule. Chemical sensors

for NH_3 and CO_2 with suitable modification permit measurement of pH directly.

Molecular oxygen can be conveniently detected with fluorosensors that rely on dynamic quenching of suitable probe molecules such as 1-pyrenebutyric acid [3443-45-6], perylene dibutyrate [79869-59-3] and 7-aminocoumarin dyes.[21] Immobilized room temperature luminescent metal chelates, such as aluminum(III) 8-hydroxy-7-iodo-5-quinolinesulfonate, zirconium(IV) 8-hydroxy-7-iodo-5-quinolinesulfonate, platinum dimer tetrakis(pyrophosphito)diplatinate(II) and $Ru(II)L_3^{2+}$ (L = 2, 2′-bipyridine, 1,10-phenanthroline and substituted derivatives) are also suitable probes for molecular oxygen sensors.[27–29] In the latter set of applications reduction in measured fluorescent and/or phosphorescent signals is mathematically related to molecular oxygen concentration, C_{oxygen}, through Stern–Volmer quenching Equation (9):

$$\frac{F_0}{F-1} = K_{oxygen}C_{oxygen} \tag{9}$$

where F_0 and F are the observed fluorescence intensity in the absence and presence of oxygen, respectively, and K is the quenching constant. Two-site Stern–Volmer models may be required at times to explain the quenching behavior of fluorophores immobilized in polymeric matrices. Microheterogeneity can cause the different solubilization sites to be quenched differently.

The newer chemical sensors are designed using fluorescent lifetime techniques. Analyte concentration is computed from the decay time of the fluorophore in contact with the sample being monitored. Lifetime-based chemical sensing is preferred over conventional intensity-based fluorescence methods because the measured lifetime is nearly independent of probe concentration. Lifetime measurements are thus unaffected by photobleaching or washout of the probe molecule. Probe molecules typically display lifetimes in the 1 to 10 nanosecond range. Molecular oxygen and ammonia have been detected based upon the spectral properties of select metal–ferron[27] and the sulforhodamine 101 [123333-78-8] + bromocresol green mixed probe resonance energy transfer system,[25] respectively. The preceding examples represent only a small fraction of the absorption and fluorescent probe-based chemical sensors that have been designed over the past two decades.

4 ULTRAVIOLET/VISIBLE ABSORPTION AND FLUORESCENT pH PROBES

Chemical sensors have been developed for pH determinations based upon both absorbance and fluorescence measurements. Probe molecules employed in the absorbance-based sensors include common acid–base indicators such as neutral red, cresol red, phenol red, bromothymol blue, brilliant yellow, bromocresol green, bromocresol purple, chlorophenol red and bromophenol blue. Extensive tabulations of acid–base indicators, and their spectroscopically determined equilibrium dissociation constants, K_{HIn}, have been published elsewhere.[30] Dissociation of indicator results in a significant wavelength shift in the maximum absorption band, which is generally accompanied by a visual change in indicator color. Protonated and unprotonated forms of the indicator have different colors. Mathematically, the hydrogen ion concentration is related to the molar concentrations of acid (HIn) and base (In^-) forms of the indicator probe molecule through the Henderson–Hasselbach Equation (10):

$$pH = pK'_{HIn} - \log\left\{\frac{[HIn]}{[In^-]}\right\} \tag{10}$$

where pK'_{HIn} refers to negative logarithm of the probe's apparent acid dissociation constant. For convenience, activity coefficients have been incorporated in the dissociation constant. Measurements are performed at fixed probe concentration, and over a narrow interval of 1–2 pH units near $pH \approx pK'_{HIn}$ a linear relationship between measured absorbance and solution pH is observed.

Mixed absorption probe systems have been used to extend the linear calibration curves by several pH units. Generally the indicator probes have approximately the same salt and same acid absorption spectrum (i.e. comparable wavelengths of maximum absorbance and nearly identical molar absorptivity ratios at the two wavelengths) with slightly different pK_a values. Concentrations are adjusted such that all probe molecules have the same measured absorbance in its salt (or acid) form at the analysis wavelength. Published studies have demonstrated that a linear absorbance versus pH calibration curve results whenever the differences in pK_as between each successive pair of absorption pH probes is 1.6 pH units. The calibration curve for a mixed probe system is more linear and extends over a considerably larger range of hydrogen ion concentrations than does the corresponding curve for single probe systems. For example, the measured absorbance is reported to be linear from pH = 1.0 to pH = 7.0 for the mixed probe system containing 25.4 mg of neutral red + 10.2 mg of methyl red + 7.8 mg of methyl yellow + 14.1 mg of thymol blue. A second four probe system, with orange IV replacing thymol blue, exhibits a linear response up to pH = 8.[31]

Fluorescence measurements also provide a convenient experimental means for pH determinations. In many applications involving chemical sensors, fluorescence-based measurements will be preferred to absorbance-based measurements because of their increased sensitivity,

greater latitude in geometric design, decreased background corrections, wider dynamic ranges, greater specificity through combined excitation–emission wavelength selection, and linear response at much lower spectroscopic probe concentrations. Generally, fluorescent pH probes are stronger acids in the excited state than in the ground state. For many pH probes fluorescence originates only from the excited state of the base form, and is observed at approximately 4–5 pH units below the ground-state pK_a. A few indicators, such as 8-hydroxypyrene-1,3,6-trisulfonic acid, are fluorescent in both basic and acidic forms. Solution pH can be monitored using hydroxypyrenetrisulfonic acid probe by simply measuring relative intensities of the molecule's basic and acidic forms, rather than by measuring the absolute intensity of the basic (or acidic) form. The calculated F_{base}/F_{acid} intensity ratio provides an internal reference point, which eliminates the necessity for repeated calibrations.

Figure 2 depicts the variation in the observed emission intensity ratios of CF, DiMCF, BCECF and HMCo as a function of solution pH. Fluorescence is measured at a fixed emission wavelength. Different excitation wavelengths are needed to excite the protonated (i.e. $\lambda_{ex} = 334$ nm in HMCo) and anionic (i.e. $\lambda_{ex} = 365$ nm in HMCo) forms of the four pH probes.[32] In the

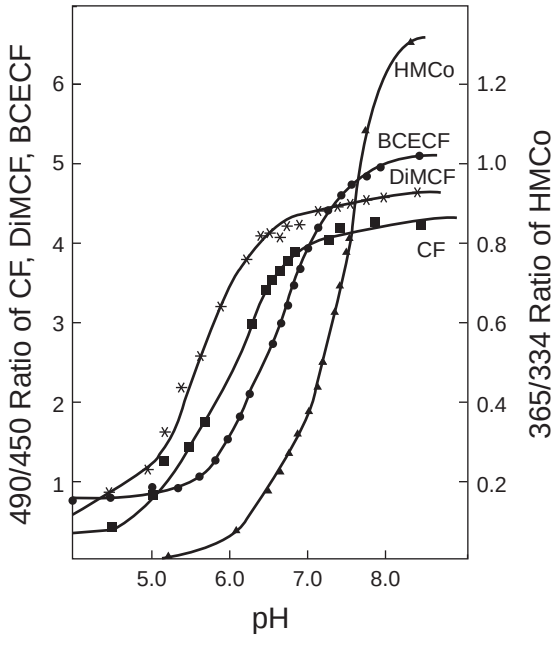

Figure 2 Titration curves in vitro of the fluorescence ratio versus pH for 10^{-7} molar 7-hydroxy-4-methylcoumarin (HMCo), , carboxyfluorescein (CF), ■, dimethylcarboxyfluorescein (DiMCF), ∗ and biscarboxyethylcarboxyfluorescein (BCECF), ●. (Redrawn with permission from M.L. Graber, D.C. DiLillo, B.D. Friedman, E. Pasotriza-Munoz, *Anal. Biochem.*, **156**, 202–212. Copyright (1986) Academic Press, Inc.)

case of fluorescein derivatives the emission spectrum is dominated by the dianion, with only small contributions from monoanion (protonated phenol group). Both the phenol and carboxylic acid function groups of fluorescein are almost totally ionized in aqueous solutions above a pH of 9. Near linear calculated F_{base}/F_{acid} versus pH calibration curves enable a fairly accurate determination of the hydrogen ion concentration over a 1–2 unit range of pH. Sensitivity of the analytical method is determined by difference between the quantum yields of fluorescent probe's acid and base forms. For many of the recently published fluorescence methods, estimated uncertainty was reported to be ±0.01–0.1 pH units. Table 3 lists several of the common fluorescent pH probes, along with their applicable pH range and select spectral properties. The pH ranges that are given pertain to steady-state fluorescence measurements.

Three mixed indicator systems, containing both a pH-sensitive absorption probe and pH-sensitive fluorescent probe molecule, are also included in the tabulation. These particular systems were designed to provide greater sensitivity by compressing the entire observed spectral change into a much narrower range of hydrogen ion concentrations. The method is based upon primary and secondary inner-filtering effects. In a typical application, the fluorescent probe molecule would have low absorption at its excitation wavelength in acidic solutions. A weak fluorescent signal would be observed at low pH. An absorption probe molecule is added to the solution to overlap the fluorescent probe's excitation spectrum. If necessary, a second absorption probe could be introduced to overlap with the emission spectrum as well. At low pH, the absorption probe has a high molar absorptivity. At the higher pHs, the fluorescent probe has a high absorption, whereas the absorption indicator is nearly optically transparent at both the excitation and emission wavelengths. Under these circumstances, the fluorescent probe's emission signal is significantly reduced in solutions of low pH because of strong absorption of the incoming excitation radiation by the second indicator. Very slight reduction in emission signal occurs at the upper end of the pH range. The net effect of the added absorption indicator is to compress the entire observed fluorescence emission change into a much narrower range of hydrogen ion concentrations, thereby increasing the analytical method's sensitivity.[35]

SNARF and SNAFL derivatives have been used as probe molecules in fluorescent lifetime-based pH determinations. Sensors containing the probes have employed phase-modulation fluorometry. Measured phase and modulation values depend upon both the excitation and emission wavelengths, as well as the hydrogen ion concentration. Through judicious wavelength selection, Szmacinski and Lakowicz[38] demonstrated that a given

Table 3 Representative fluorescent pH probes[a] and their spectral properties

Probe/property	Probe/property
Single indicator systems:	HMCo
8-Hydroxypyrene-1,3,6-trisulfonic acid	pH range: 7.0–8.0[24]
pH range: 7.0–8.0[24]	365/334 nm excitation ratio
450/405 nm excitation ratio	1,4-Dihydroxyphthalonitrile
Fluorescein and CF Derivatives	pH range: 6.0–8.0[34]
pH range: 6.0–7.2[24]	512/455 nm emission ratio
490/450 nm excitation ratio	2′,7′-Difluorofluorescein
SNARF Derivatives	(Oregon Green) Derivatives
pH range: 7.0–8.0[24]	pH range: 4.2–5.7[24]
580/640 nm emission ratio	510/450 nm excitation ratio or
SNAFL Derivatives	490/440 nm excitation ratio
pH range: 7.2–8.2[24]	Cl-NERF
490/540 nm excitation ratio	pH range: 2.5–4.5[36,37]
540/630 nm emission ratio	514/485 nm excitation ratio
Mixed indicator systems:	4-Hydroxycoumarin-4-acetic acid + *m*-cresol purple
Hydroxypyrenetrisulfonic acid + Neutral red	pH range: 7.5–9.0[35]
pH range: 7.0–7.5[35]	$\lambda_{ex} = 400$ nm, $\lambda_{em} = 458$ nm
$\lambda_{ex} = 469$ nm, $\lambda_{em} = 515$ nm	4-Hydroxycoumarin-4-acetic acid + Brilliant yellow
pH range: 7.5–8.0[35]	pH range: 7.0–7.5[35]
$\lambda_{ex} = 484$ nm, $\lambda_{em} = 515$ nm	$\lambda_{ex} = 390$ nm, $\lambda_{em} = 458$ nm

[a] A more complete listing of fluorescent pH probes appears in Dean,[30] and in Wolfbeis et al.[33]
SNARF, seminaphthorhodafluor; SNAFL, seminaphthofluorescein.

probe's useful pH range could be shifted by several pH units. The pH-sensitive range of carboxy SNARF-6 excited at $\lambda_{ex} = 543$ nm is from pH $\approx$ 5 to pH $\approx$ 10.5. Emission wavelengths of $\lambda_{em} = 660$ nm (pH $\approx$ 5–7.5) and $\lambda_{em} = 580$ nm (pH $\approx$ 7–10.5) are needed to cover the entire pH range. The wide pH range is possible because of the large wavelength shift in the emission spectra of the acid and base forms of carboxy SNARF-6.

5 FLUORESCENT WAVELENGTH-SHIFT SOLVENT POLARITY PROBES

Sufficient time exists between the excitation and emission process in fluorescence spectroscopy for physical re-orientation of solvational sphere molecules. Physical re-orientation occurs as solvent molecules move to align their respective dipole moments with the new dipole moment of the excited fluorophore. Alignment of fluorophore–solvent dipole moments leads to more effective molecular interactions, which is reflected in a lower excited state energy level relative to the ground electronic state. Energy lowering increases with increasing fluorophore–solvent interactions and solvent polarity as depicted in Figure 3.[39] Most fluorescent probes have a much larger excited state than ground state dipole moment, hence the energy level of the ground state is affected to a much lesser extent. As a general rule-of-thumb the larger redshifts in emission band

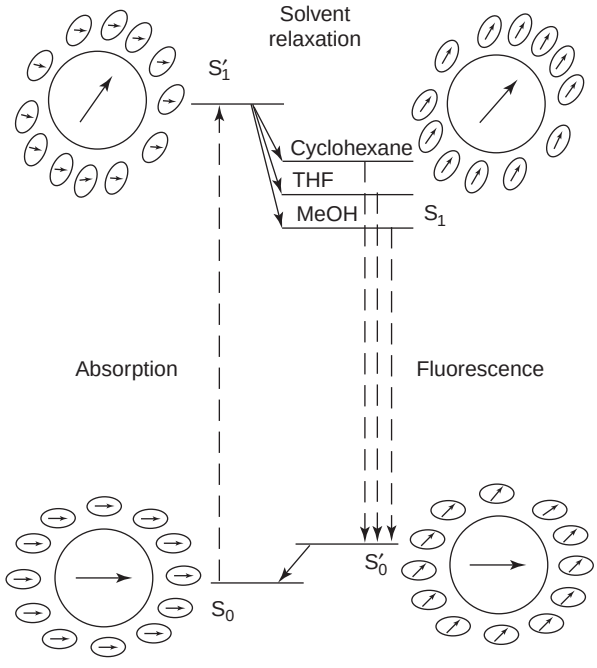

Figure 3 Jablonski diagram showing solvent relaxation process: (- - - -) radiative process; (———) nonradiative process. (Redrawn with permission from T.L. Cecil and S.C. Rutan, *Anal. Chem.*, **62**, 1998–2004. Copyright (1990) American Chemical Society.)

wavelength(s) are observed in the more polar solvents. Emission wavelength(s) of fluorescent solvent polarity

probe molecules provides a quantitative measure of the polarity of the solubilizing solvent media.

Theoretical interpretation of solvatochromic wavelength shifts at the molecular level is very difficult, often subject to interpretation, and as one may surmise there have been numerous competing mathematical expressions published in the spectroscopic literature to explain observed wavelength shifts associated with known solute probes. To a first approximation specific effects are ignored. The energy difference (in cm^{-1}) between solvated ground and excited state fluorophore can be described by the Lippert equation (Equation 11):[40]

$$\frac{1}{\lambda_{ex}} - \frac{1}{\lambda_{em}} \approx \frac{2}{hc}\frac{(\mu^* - \mu)^2}{a^3} \times \frac{\varepsilon - 1}{2\varepsilon + 1} - \frac{n^2 - 1}{2n^2 + 1} + \text{constant} \quad (11)$$

or modified forms such as Equation (12):

$$\frac{1}{\lambda_{ex}} - \frac{1}{\lambda_{em}} \approx \frac{2}{hc}\frac{(\mu^* - \mu)^2}{a^3} \times \frac{\varepsilon - 1}{2\varepsilon + 1} - \frac{\beta(n^2 - 1)}{(2n^2 + 1)} + \text{constant} \quad (12)$$

where μ and μ^* refer to the ground state and excited state dipole moment, respectively, ε and n are the dielectric constant and refractive index of the solvent, c is the speed of light, h denotes Planck's constant, and a is the radius of the solvent cavity wherein the fluorophore resides. Equations (11) and (12) both imply a linear relationship between the observed Stokes shift, $1/\lambda_{ex} - 1/\lambda_{em}$, and a specific function of the solvent's refractive index and dielectric constant. The excited state dipole moment can be estimated through a linear least-squares regressional analysis of experimental Stokes shift data in accordance with Equation (11) (or 12).

Most fluorescent compounds do show some degree of solvatochromism as a result of molecular interactions with surrounding solvent molecules. The $(\mu^* - \mu)^2/a^3$ multiplicative factor in Equations (11) and (12) governs to a large extent the magnitude of the Stokes shift that the fluorophore experiences in going from a nonpolar to polar solubilizing medium. For fluorescent solvent polarity probes there generally is a large difference between the ground and excited state dipole moments, which leads to sizeable shift in the excitation and/or emission wavelengths. Coumarin-153 [53518-18-6], 4-amino-*N*-methylphthalimide [2307-00-8], 6-propionyl-2-(dimethylamino)naphthalene (PRODAN) [70504-01-7], 1-phenyl-4-(2-naphthylmethylene) piperidine (PNMP) [91759-51-2], 4-phenyl-4-(4-cyano-1-naphthylmethylene)piperidine (PCNMP) [91759-53-3] and the other compounds listed in Table 4 exhibit fairly large wavelength shifts in their recorded fluorescence spectrum as a function of solvent polarity. Each molecule listed has been recommended as a solvent polarity probe. Caution should be exercised in selecting a probe molecule for a particular application. Several molecules listed in Table 4 do contain acidic and/or basic functional groups; and as a result their spectral response may not necessarily be just to solvent polarity. Proton or charge transfer may contribute to the wavelength shift(s) for select solvents. Catalan et al.[53] issued a similar precaution when using PRODAN as an indicator of the polarity of protein cavities, particularly if the environments being examined contain acidic sites. The authors had to introduce a new acidic term into the Lippert equation to satisfactorily explain the observed absorption and emission spectroscopic behavior of PRODAN in several solvents.

6 FLUORESCENT TWISTED INTRAMOLECULAR CHARGE TRANSFER PROBES

Solvatochromic probe character can also result from charge-transfer induced dual fluorescence. Molecular systems that are known to exhibit such behavior include 9,9′-bianthracene (BA) [1055-23-8], 4-(9-anthryl)-*N*,*N*-dimethylaniline [38474-09-8], 4-(*N*,*N*-dimethylamino)benzonitrile [1197-19-9], ethyl 4-(*N*,*N*-dimethylamino)benzoate [10287-53-3] and *N*-arylcarbazoles. Theoretical treatments often assume equilibrium between the locally excited (LE) and twisted intramolecular charge transfer (TICT) species, which for BA is represented as[54] (Equation 13)

$$\underset{\text{TICT}}{An^+{-}An^-} \rightleftharpoons \underset{\text{LE}}{An{-}An^*} \rightleftharpoons \underset{\text{TICT}}{An^-{-}An^+} \quad (13)$$

with full electron/charge transfer separation. The two anthracene rings (denoted as An) twist to a perpendicular configuration to accommodate the separated charges. Fluorescence emission originates from LE and TICT excited states. The TICT band appears at longer wavelengths, and is more affected by solvent polarity/polarizability than is the LE band (see Figures 4 and 5). Emission band intensities are dictated by the fluorescent quantum yield and relative concentration of the two presumed isomeric species. More polar solvents like dimethyl sulfoxide favor the TICT species, and recorded emission spectrum in polar solvents often shows just the single TICT band. A bilinear Lippert relationship is generally seen in the case of TICT formation, with separate lines for each emission process.

p-Toluidinonaphthalene sulfonate (TNS) is also listed as fluorescent TICT probe. In water, where the polarity is high, the TICT rates of TNS are very large and

Table 4 Spectral properties of select fluorescent wavelength-shift solvent polarity probes

Probe/spectral properties	Probe/spectral properties
Coumarin-153[8,41] Nonpolar: $\lambda_{ex}/\lambda_{em} \approx 395/454$ nm Polar: $\lambda_{ex}/\lambda_{em} \approx 426/550$ nm	PNMP[8,46] Nonpolar: $\lambda_{em} \approx 361$ nm Polar: $\lambda_{em} \approx 523$ nm
4-Amino-*N*-methylphthalimide[8] Nonpolar: $\lambda_{em} \approx 410$ nm Polar: $\lambda_{em} \approx 606$ nm	PCNMP[8,46] Nonpolar: $\lambda_{em} \approx 412$ nm Polar: $\lambda_{em} \approx 695$ nm
N-Phenyl-1-naphthylamine[42] 1,4-Dioxane: $\lambda_{em} \approx 400$ nm Water: $\lambda_{em} \approx 462$ nm	PRODAN[47] Nonpolar: $\lambda_{em} \approx 401$ nm Polar: $\lambda_{em} \approx 531$ nm
Pyrene-1-carboxyaldehyde[43] Nonpolar: $\lambda_{em} \approx 408$ nm Polar: $\lambda_{em} \approx 475$ nm Linear λ_{em} versus ε (for $\varepsilon > 10$)	8-Anilino-1-naphthalene-sulfonic acid[48,49] Ethanol: $\lambda_{em} \approx 464$ nm Water: $\lambda_{em} \approx 523$ nm 200-fold increase in ϕ_f in 1-octanol versus water
2,3-(Dihydro-1-methyl-4(1*H*)-quinolinylidene)propanedinitrile[8] Nonpolar: $\lambda_{em} \approx 500$ nm Polar: $\lambda_{em} \approx 588$ nm	Coumarin-311[50] Nonpolar: $\lambda_{em} \approx 373$ nm Polar: $\lambda_{em} \approx 444$ nm
1,9-Di(*N*-phenyl-*N*-methyl)-4,6-dimethylenenona-1,3,6,8-tetraen-5-one[44] Toluene: $\lambda_{ex}/\lambda_{em} \approx 452/491$ nm Water: $\lambda_{ex}/\lambda_{em} \approx 552/610$ nm	2-(*p*-Toluidino)naphthalene-6-sulfonate[51] Dioxane: $\lambda_{em} \approx 413$ nm Water: $\lambda_{em} \approx 500$ nm Large increase ϕ_f in organic solvents versus water
Coumarin-102[45] $\Delta(1/\lambda_{ex}) \approx 1.47 \times 10^3$ cm^{-1} $\Delta(1/\lambda_{em}) \approx 2.94 \times 10^3$ cm^{-1} Dimethyl sulfoxide–cyclohexane	Nile red[52] Nonpolar: $\lambda_{ex}/\lambda_{em} \approx 484/529$ nm Water: $\lambda_{ex}/\lambda_{em} \approx 591/657$ nm

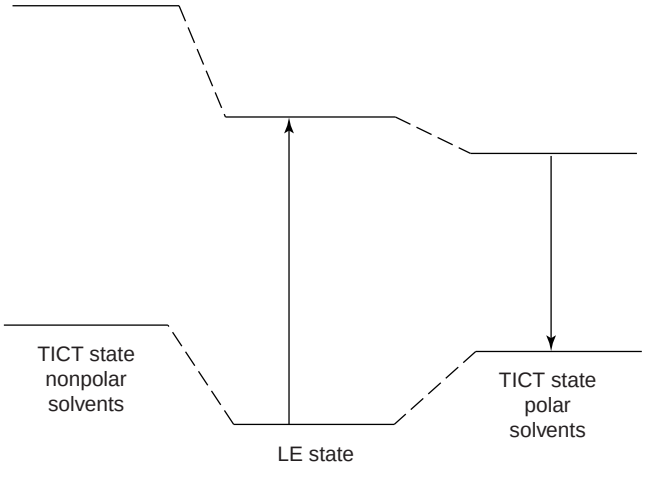

Figure 4 Potential energy diagram showing the local excited and charge transfer state energies as a function of solvent polarity for a typical TICT compound. (Redrawn with permission from R.D. Schulte, J.F. Kauffman, *Appl. Spectrosc.*, **49**, 31–39. Copyright (1995) Society for Applied Spectroscopy.)

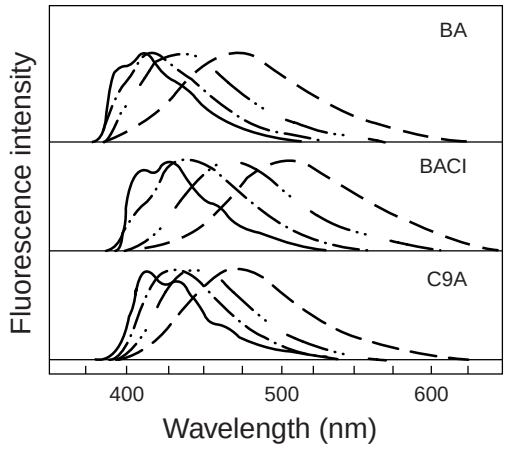

Figure 5 Corrected fluorescence spectra (room temperature) of BA, 10-chloro-9,9′-bianthracene (BACl) and *N*-(9-anthryl)-carbazole (C9A) in nonpolar and polar aprotic solvents: hexane (———), ethyl ether (- · - ·), tetrahydrofuran (- ·· - ··) and acetonitrile (- - - -). (Redrawn with permission from W. Rettig and M. Zander, *Ber. Bunsen-Ges. Phys. Chem.*, **87**, 1143–1149. Copyright 1983 Wiley-VCH.)

the fluorescent quantum yield is very small, i.e. $\phi_f \approx 0.001$. The TICT process in the case of this particular fluorescent probe is nonradiative. When TNS binds to the nonpolar, hydrophobic surfaces of micelles, proteins or apolar CD cavities, the TICT process is inhibited. This causes a dramatic increase in the fluorescent quantum yield. The fluorescence intensity enhancement

thus provides direct information about polarity of the microenvironment in the immediate vicinity of the TNS probe molecule.

The following examples illustrate the type of information that can be obtained with TICT probe molecules. Kajimoto and coworkers[55] used fluorescence emission from the TICT state of 4-(*N*,*N*-dimethylamino)benzonitrile to investigate solvent structure in the cybotactic region of supercritical trifluoromethane. Experimental measurements revealed that both the Stokes shift and calculated F_{TICT}/F_{LE} emission intensity ratio increased with increasing fluid density. Observed spectral wavelength shifts were compared to values predicted from theoretical models that assume a homogeneous dielectric solvent. Analysis of the experimental data indicated that near the critical point the local solvent density around the probe molecule was greater than the bulk solvent. The CHF_3 solvent molecules were not uniformly dispersed, but rather tended to aggregate around the fluorophore. At low densities, the number of solvent molecules within the fluorophore's solvational sphere depended linearly upon the bulk solvent density. More recently, TICT probes 4-(*N*,*N*-dimethylamino)benzonitrile and ethyl 4-(*N*,*N*-dimethylamino)benzoate were used to study interactions in supercritical trifluoromethane and carbon dioxide at different solvent compositions.[56]

TICT probes can examine structural features inside CD cavities. The most widely used CD molecules contain six, seven and eight glucose monomers arranged in toris shapes, and are named α-, β- and γ-CD, respectively. Coupling of the glucose molecules gives the CD a rigid, conical molecular structure with a hollow hydrophobic interior that enables the CDs to trap probe molecules based upon molecular shape/size compatibility. γ-CD is the larger of the commercially available CDs, having an internal diameter of approximately 9.5 Å and depth of roughly 7.8 Å. Published fluorescence measurements indicate that LE and TICT electronic states of 4-(*N*,*N*-dimethylamino)benzonitrile form 1 : 1 guest–host inclusion complexes with β-CD. The two electronic states reside in different microenvironments within the cavity.[57]

7 FLUORESCENT POLYCYCLIC AROMATIC HYDROCARBON SOLVENT POLARITY PROBES

Pertinent photophysical processes governing fluorescence behavior of polycyclic aromatic hydrocarbons (PAHs) dissolved in fluid solution are summarized below:

$$PAH + h\nu_0 \longrightarrow PAH^* \quad \text{(absorption/excitation)}$$
$$PAH^* \longrightarrow PAH + h\nu_1 \quad \text{(emission)}$$
$$PAH^* + PAH \longrightarrow (PAH)_2^* \quad \text{(excimer formation)}$$
$$(PAH)_2^* \longrightarrow 2PAH + h\nu_2 \quad \text{(excimer emission)}$$

Emission spectrum of PAH monomer consists of several major vibronic bands labeled I, II, III, etc., in progressive order, starting with the O–O vibronic band. Published studies[58–60] revealed that Py, benzo[*ghi*]perylene (BPe) [191-24-2], ovalene (Ov) [190-26-1], coronene (Co) [191-07-1], benzo[*a*]coronene (BCo) [190-7-5], naphtho-[2,3*a*]coronene (NCo) [190-74-9], dibenzo[*a*, *j*]coronene (DCo) [190-72-7], benzo[*e*]pyrene (BePy) [192-97-2], 3,4-dihydrobenzo[*ghi*]perylene (DHBPe) [16310-65-9], benzo[*rst*]pentaphene (BPP) [189-55-9], methylcoronene (MeCo) [13119-86-3], 1,2-dimethylcoronene (DiMeCo) [113523-79-8] and several other PAHs exhibit selective emission intensity of vibronic band I relative to band III (or band IV) in polar solvents (see Figure 6). Ratios of band I and III (or IV) serve as quantitative measure of solvent polarity and structure. Less than 25% of the PAH solutes studied to date exhibit probe character. Emission intensity ratios for the majority of PAHs remain essentially constant, irrespective of solvent polarity.

Tables 5 and 6 list experimental emission intensity ratios for Py, BPe, Co, BePy, DHBPe, Ov, MeCo and DiMeCo dissolved in select nonelectrolyte solvents. Careful examination of the entries in Table 7 reveals that, for the most part, the rank ordering of solvents by polarity is identical for the eight PAH probes. Linear least-squares analysis of the intensity ratio data shows that the polarity scales are highly correlated as indicated by the squared correlation coefficient, r^2.

Of the known PAHs that exhibit selective emission intensity enhancement, the eight molecules listed above are the ones recommended for solvent polarity

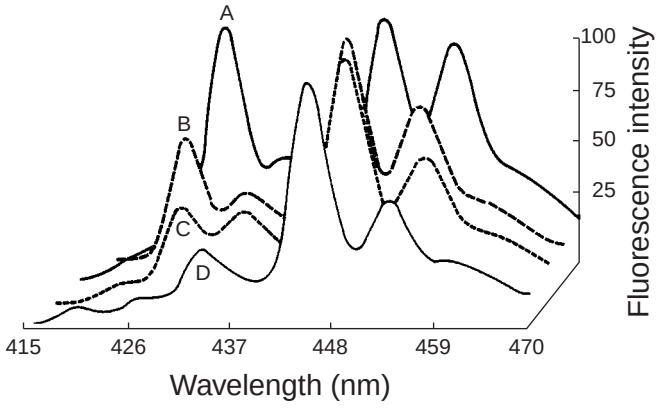

Figure 6 Fluorescence emission spectra for Co dissolved in dimethyl sulfoxide (A), methanol (B), carbon tetrachloride (C) and cyclohexane (D). For illustrational purposes, the four emission spectra have been normalized to the emission intensity of the III band. The Co solvent polarity scale is defined as the I(at 426–430 nm)/III(at 445–449 nm) emission intensity ratio.

Table 5 Ratios of fluorescence emissions of Py, BPe, Co and Ov in various organic solvents

Organic solvent	Py[a]	BPe[b]	Co[c]	Ov[d]
n-Hexane	0.58	0.38	0.12	0.30
n-Heptane	0.58	0.39	0.14	0.30
Cyclohexane	0.58	0.40	0.14	0.31
n-Hexadecane	0.60	0.42	0.13	0.36
Carbon tetrachloride	Rxn	0.63	0.41	0.82
Dibutyl ether	0.84	0.65	0.33	0.74
p-Xylene	0.95	0.79	0.44	1.10
o-Xylene	0.99	0.84	0.48	1.25
m-Xylene	1.01	0.81	0.45	1.19
Diethyl ether	1.02	0.83	0.45	1.11
Toluene	1.04	0.81	0.49	1.13
Benzene	1.05	0.83	0.51	1.19
1-Butanol	1.06	0.93	0.46	1.25
1-Propanol	1.09	0.96	0.49	NA
Cyclohexanol	1.09	0.94	0.48	NA
2-Propanol	1.09	0.92	0.52	1.18
Ethanol	1.18	1.05	0.57	1.45
Chloroform	1.25	0.96	0.59	1.30
Butyl acetate	1.35	0.97	0.61	1.42
Methanol	1.35	1.00	0.62	1.37
Tetrahydrofuran	1.35	1.03	0.64	1.52
Dichloromethane	1.35	1.06	0.63	1.46
Ethyl acetate	1.37	1.04	0.65	1.52
1,4-Dioxane	1.50	1.04	0.64	1.66
Acetone	1.64	1.19	0.76	1.80
Acetonitrile	1.79	1.23	0.79	1.75
N,*N*-Dimethylacetamide	1.79	1.31	0.90	2.16
N,*N*-Dimethylformamide	1.81	1.28	0.88	2.11
Dimethyl sulfoxide	1.92	1.36	0.98	2.38
Dynamic range	1.37	0.98	0.86	2.08

[a] Experimental Py values are taken in part from Dong and Winnik[58] and are based upon the properties of Py.
[b] Defined as the ratio of band I (≈405 nm)/band III (≈417 nm) for benzo[*ghi*]perylene.
[c] Defined as the ratio of band I (≈424 nm)/band III (≈445 nm) for Co.
[d] Defined as the ratio of band I (≈480–485 nm)/band III (≈500–505 nm) for Ov.

determinations. The eight PAHs have various shapes and different excitation wavelength(s). Depending upon the spectral characteristics of the organic solvent and/or molecular shape/size constraints imposed by the organized solvent media being examined, one PAH probe may be preferred from an experimental standpoint.

Py solvent polarity scale is defined as the F_1/F_3 emission intensity ratio, where band I corresponds to $S_1(v = 0) \rightarrow S_0(v = 0)$ transition and band III is a $S_1(v = 0) \rightarrow S_0(v = 1)$ transition. Karpovich and Blanchard[61] rationalized Py's solvatochromic behavior in terms of vibronic coupling between the weakly allowed Py first electronic excited singlet state ($\varepsilon_{max} \approx 200\,cm^{-1}\,M^{-1}$) and the strongly allowed second electronic singlet state. The observed Py absorption/excitation spectrum involves $S_0 \rightarrow S_2$ electronic transition. The nonprobe perylene molecule, on the other hand, exhibits the typical excitation–emission mirror image symmetry common to many PAHs. Perylene absorption/excitation spectrum corresponds to the ground state → first electronic excited state transition ($\varepsilon_{max} \approx 33\,000\,cm^{-1}\,M^{-1}$ for $S_0 \rightarrow S_1$).

Vibrationally mediated coupling between electronic S_1 and S_2 states can be mathematically described through both Herzberg–Teller (HT) and Born–Oppenheimer (BO) contributions. Equations derived by Karpovich and Blanchard indicate that the magnitude of nuclear displacement along the coupling coordinate is sensitive to the dipolar and dielectric properties of surrounding solvent media. Vibronic coupling contributions to the O–O absorption and emission cross-sections scale as HT + BO and HT − BO contributions, respectively. For a two level system, such as the O–O Py transition, cross-sections for the absorption and emission processes must be equal. Hence, $(HT + BO)_{O-O} = (HT - BO)_{O-O}$, and I band intensity depends solely upon the HT coupling between S_2 and S_1 electronic states. Band III corresponds to a $S_1(v = 0) \rightarrow S_0(v = 1)$ electronic transition, with its emission intensity being determined by the difference between

Table 6 Ratios of fluorescence emissions of BePy, DHBPe, MeCo and DiMeCo in various organic solvents

Organic solvent	BePy[a]	DHBPe[b]	MeCo[c]	DiMeCo[d]
n-Hexane	0.50	1.14	0.72	1.40
n-Heptane	0.52	1.16	0.74	1.44
Cyclohexane	0.51	1.12	0.72	1.50
n-Hexadecane	0.53	1.17	0.76	1.47
Carbon tetrachloride	1.11	1.47	1.03	1.37
Dibutyl ether	0.77	1.35	0.95	1.65
p-Xylene	0.98	1.44	1.13	1.79
o-Xylene	1.04	1.45	1.17	1.86
m-Xylene	0.98	1.43	1.15	1.81
Diethyl ether	0.95	1.47	1.08	1.79
Toluene	1.03	1.54	1.15	1.86
Benzene	1.08	1.59	1.22	1.90
1-Butanol	1.04	1.45	1.13	1.86
1-Propanol	1.08	1.48	1.20	1.90
Cyclohexanol	1.08	1.47	1.14	1.83
2-Propanol	1.05	1.55	1.14	1.85
Ethanol	1.19	1.57	1.23	2.02
Chloroform	1.42	1.78	1.33	2.05
Butyl acetate	1.26	1.66	1.30	1.97
Methanol	1.33	1.73	1.33	2.08
Tetrahydrofuran	1.28	1.73	1.37	2.07
Dichloromethane	1.49	1.73	1.41	2.25
Ethyl acetate	1.35	1.73	1.32	2.05
1,4-Dioxane	1.42	1.81	1.43	2.16
Acetone	1.70	1.89	1.50	NA
Acetonitrile	1.81	2.02	1.61	2.34
N,*N*-Dimethylacetamide	1.83	2.00	1.63	2.42
N,*N*-Dimethlyformamide	1.84	1.95	1.64	2.41
Dimethyl sulfoxide	2.05	2.10	1.80	2.57
Dynamic range	1.55	0.98	1.08	1.17

[a] Defined as ratio of band I (≈376 nm)/band IV (≈408 nm) for BePy.
[b] Defined as the ratio of band I (≈380 nm)/band II (≈390 nm) for DHBPe.
[c] Defined as the ratio of band I (426–429 nm)/band III (444–448 nm) for MeCo.
[d] Defined as the ratio of band I (428–432 nm)/band III (447–450 nm) for DiMeCo.

Table 7 Ordering of solvents by polarity

BPe = 0.674 Py + 0.109	($r^2 = 0.932$)
Co = 0.550 Py − 0.136	($r^2 = 0.956$)
Ov = 1.356 Py − 0.331	($r^2 = 0.941$)
BePy = 1.034 Py − 0.058	($r^2 = 0.963$)
DHBPe = 0.687 Py + 0.773	($r^2 = 0.956$)
MeCo = 0.692 Py + 0.393	($r^2 = 0.965$)
DiMeCo = 0.772 Py + 1.082	($r^2 = 0.936$)

the HT and BO contributions. This difference is less solvent dependent than each of the separate HT and BO contributions. The F_1/F_3 emission intensity ratio thus increases with increasing solvent polarity.

Fluorescent PAH solvent polarity probes have been used in applications ranging from the determination of critical micelle concentrations, to measurement of solvent polarity of gas–liquid chromatographic (GLC) and HPLC stationary phases to estimation of preferential solvation in binary solvent mixtures. In this latter application, most published approaches assume that a given fluorophore's solvational sphere contains just solvent molecules A. Other fluorophore molecules in the solution would be solvated completed by solvent component B. Each solvated fluorophore contributes to the observed fluorescence signal, F_{obs}, at each emission wavelength scanned. Emission intensities are additive at each wavelength. In the case of a PAH probe such as Py, the calculated F_1/F_3 emission intensity ratio is[62] given by the Equation (14):

$$\frac{F_1}{F_3} \approx [Y_A F_{1,\text{phase A}} + (1 - Y_A) F_{1,\text{phase B}}] + [Y_A F_{3,\text{phase A}} + (1 - Y_A) F_{3,\text{phase B}}] \quad (14)$$

Here, [Fluoro] is the total stoichiometric fluorophore molar concentration, and Y_A and $(1 - Y_A)$ represent the mole number fraction of each type of solvated fluorophore, i.e. $Y_A = [\text{Fluoro A}] \div [\text{Fluoro}]$ and $(1 - Y_A) = [\text{Fluoro B}]/[\text{Fluoro}]$. Solute concentration must be

maintained constant for all fluorescence measurements. Equation (14) does accurately reproduce fluorescence behavior of Py, DHBPe, Co and BePy dissolved in *n*-heptane + 1,4-dioxane and *n*-heptane + tetrahydrofuran mixtures. Calculated Y_A values indicated that the PAH solute probe was preferentially solvated by the nonpolar *n*-heptane cosolvent, with the degree of preferential solvation decreasing with increased PAH molecular size.[62]

Modified forms of the general spectral response equation have been used successfully in the calculation of equilibrium constants for PAH–CD complex formation, and in the calculation of binding constants for PAH–surfactant complexation. As a classical example, Xu et al.[63] examined interactions between Py and α- and β-CD and/or polymerically linked β-CDs. Experimental fluorescence emission intensity versus CD concentration curves were analyzed in terms of four existing literature models. Of the models considered only the one assuming sequential formation of 1:1 and 1:2 Py–CD complexes, with each complex being fluorescent and having different F_1/F_3 emission intensity ratios, was found to describe the monomeric CD data satisfactorily. At each emission wavelength scanned, the observed fluorescence signal, F_{obs}, is the sum of the emission intensities from the Py monomer and two fluorescent Py–CD complexes. The measured I and III band emission intensities are given by Equations (15) and (16):

$$F_1 = f_{Py}F_{1,Py} + f_{Py-CD}F_{1,Py-CD} + f_{CD-Py-CD}F_{1,CD-Py-CD} \quad (15)$$

$$F_3 = f_{Py}F_{3,Py} + f_{Py-CD}F_{3,Py-CD} + f_{CD-Py-CD}F_{3,CD-Py-CD} \quad (16)$$

where f is the fraction of Py existing as the uncomplexed Py, 1:1 and 1:2 Py–CD complexes. The four emission intensities involving the Py–CD complexes cannot be experimentally determined because each PAH–CD complex must be in equilibrium with the *uncomplexed* PAH monomer. To circumvent this problem, Equation (15) is recast as Equation (17):

$$F_1 = f_{Py}\left(\frac{F_1}{F_3}\right)_{Py} F_{3,Py} + f_{Py-CD}\left(\frac{F_1}{F_3}\right)_{Py-CD} F_{3,Py-CD} + f_{CD-Py-CD}\left(\frac{F_1}{F_3}\right)_{CD-Py-CD} F_{3,CD-Py-CD} \quad (17)$$

using F_1/F_3 emission intensity ratios, equilibrium constants and F_3 values which scale the calculation to the observed data. The various fractions were expressed in terms of the two equilibrium constants, uncomplexed Py and CD concentrations. Through a global fitting of F_1 and F_3 data simultaneously, optimized numerical values of the equilibrium constants were calculated.

The number of required input parameters (e.g. fluorescence emission properties) is reduced through simplifying approximations. Examination of Equation (14) reveals that the measured F_1/F_3 emission intensity ratio for a PAH probe dissolved in a binary solvent mixture (or alternatively, for a PAH probe dissolved in two different microenvironments) becomes Equation (18):

$$\frac{F_1}{F_3} = Y_A\left(\frac{F_1}{F_3}\right)_{\text{phase A}} + (1 - Y_A)\left(\frac{F_1}{F_3}\right)_{\text{phase B}} \quad (18)$$

a weighted average of the measured emission intensity ratios in each pure solvent, $(F_1/F_3)_{\text{phase A}}$ and $(F_1/F_3)_{\text{phase B}}$, provided that probe's measured III band intensity is identical in both neat solvents (or microenvironments/phases). Equation (18) requires knowledge of only the F_1/F_3 emission intensity ratio for each different solvational species, rather than individual emission intensities at two different wavelengths. This reduces the number of input parameters by a factor of two, and eliminates in part experimental artifacts that might result from each solution having a slightly different analytical concentration of the dissolved PAH probe molecule. Recent studies addressed influences that chiral modifiers, diols and amines have in regards to Py–CD complexation. Complexation between Py and derivatized-CDs and between Py and cationic surfactant headgroups have been studied by fluorescence probe methods, several of which were based upon Equation (18).

8 FLUORESCENT ANISOTROPIC AND MICROVISCOSITY PROBES

Fluorescence anisotropic measurements provide valuable information in regards to the fluorophore's orientational and rotational motions. Anisotropy is an angular correlation function that describes the mutual orientation of both the absorption and emission transition dipole moments at the times of excitation ($t = 0$) and emission ($t = t$). Fluorophores whose absorption dipoles are aligned with the electric vector of the excitation radiation have the highest probability of excitation when irradiated with polarized light. Photoselection rules dictate that the resulting fluorescence must also be polarized. Fluorescence does not occur immediately. There is sufficient time for the fluorophore to rotate during the excited state lifetime. Some depolarization occurs as a result of rotational motion. The magnitude of the depolarization depends upon the fluorophore's molecular size and shape, as well as upon the viscosity of the microenvironment in the fluorophore's immediate vicinity. Fluorescence

polarization measurements have been used to investigate ligand binding, protein denaturation and surfactant aggregation.

In steady-state fluorescence anisotropic studies, spectra are acquired by scanning wavelengths selected by an excitation monochromator, and then measuring the emission intensity parallel, $F_{\parallel}$, and perpendicular, $F_{\perp}$ to the polarization of the excitation radiation. Mathematically, the degree of polarization and the anisotropy are defined by Equations (19) and (20):

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}} \tag{19}$$

$$r = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + 2F_{\perp}} \tag{20}$$

The above definitions also hold for time-resolved fluorescence measurements. Emission intensities would now be time-dependent functions, $F_{\parallel}(t)$ and $F_{\perp}(t)$, as opposed to an actual numerical value. Correction terms can be added to eliminate effects from stray radiation.

It can be shown that the initial ($t = 0$) and maximum value of the anisotropy, r_0, is 2/5 for any macroscopically isotropic orientational distribution of fluorophore molecules. This is of course the theoretical limiting value that one expects to observe in systems where rotational motions become negligible. The limiting value is never observed, however, not even in systems where the temperature has been lowered to restrict rotational motions. The smaller experimental r_0 values have been explained in terms of torsional vibrations of the fluorescent molecules about their equilibrium orientation, combined with perhaps a small rotation of the absorption and emission dipoles due to coupling with vibrational modes. Since the theoretical value is never realized, the operational definition of the limiting fluorescence anisotropy is the maximum value that is experimentally measured.

Popular fluorescent anisotropic probe molecules include: perylene [198-55-0], PRODAN, 1,4-diphenylbuta-1,3-diene [886-65-7], 1,6-diphenylhexa-1,3,5-triene [1720-32-7], anthracene [120-12-7], 2-methylanthracene [613-12-7], 9-vinylanthracene [2444-68-0], potassium 2-(*p*-toluidino)naphthalene-6-sulfonate [32752-10-6], rhodamine B [81-88-9], rhodamine 6G [989-38-8], rhodamine 101 [41175-43-3], fluorescein [2321-07-5], 2,5,8,11-tetra-*tert*-butylperylene (TBP) [80663-92-9] and 9-(3-perylenoyl)nonanoic acid. Spectral properties of selected probes are listed in Table 8. Each molecule contains a fluorescent moiety. Functional groups are added to modify the molecule's rotational motion and to provide a means of judiciously controlling location of the probe within a molecularly-organized, biological and/or membrane system. Polar and ionic functional groups increase the probe's aqueous solubility.

Fluorescent anisotropic probe methods provide a convenient experimental technique for determining microviscosity of molecularly-organized structures. If rotational diffusion is the main cause of anisotropy loss, then observed anisotropy is related to the microviscosity, η, by the Perrin equation (Equation 21):

$$r = \frac{r_0}{1 + (\tau kT/\eta V)} \tag{21}$$

where k is the Boltzmann constant, T denotes the absolute Kelvin temperature, and V and τ refer to the effective molecular volume and fluorescence lifetime of the probe, respectively. Equation (21) applies to a spherical rotor, and in several published applications V was assumed to have a functional dependence upon r. Calibration curves of measured r_0/r values versus known $\tau T/\eta$ values are available for several probes.[66] More complex expressions have been derived for nonsymmetric species or molecules.

Microviscosities have been estimated using ratio of intramolecular excimer to monomer fluorescence emission intensities (i.e. $F_{\text{excimer}}/F_{\text{monomer}}$). The method involves construction of a calibration curve by measuring $F_{\text{excimer}}/F_{\text{monomer}}$ ratios in solutions of known viscosity since there is no theoretical relationship for how the emission intensity ratio should vary with η. As a cautionary note, calibration curves should be established under isothermal experimental conditions. Excimer formation is affected by temperature, and changes in measured $F_{\text{excimer}}/F_{\text{monomer}}$ ratio forms the basis for fluorescent thermometer development. 1,3-Di(α-naphthyl)propane[78,79] [14564-86-4] and bis(4-biphenylmethyl)ether[80] [68941-68-4] are recommended probe molecules for the $F_{\text{excimer}}/F_{\text{monomer}}$ method. Microviscosity determinations using intermolecular excimer formation are generally not as effective, particularly in molecularly organized solvent structures such as micelles. It is much easier to control micellar fluorophore distribution leading to excimer formation whenever the two chromophoric moieties are contained within the same molecule.

Perylene initially was identified as an anisotropic rotator in steady-state polarization measurements. Wavelength-dependent Perrin plots indicated the presence of at least two rotational motions, which were confirmed by subsequent time-resolved fluorescence techniques. Rate of rotation of perylene about its major symmetry axis (z-axis) is an order of magnitude greater than the rate of rotation perpendicular to the z-axis. Time-resolved fluorescence anisotropies of perylene are generally characterized by a double-exponential model. Calculated rotational decay times can be related to the diffusion about the in-plane and out-of-plane molecular axis.

Table 8 Spectral properties of select anisotropic and microviscosity probes

Probe/property	Probe/property
1,4-Diphenylbuta-1,3-diene $\lambda_{ex} = 295$ nm $\lambda_{em} = 420$ nm $V = 205$ Å^3[64]	PRODAN $\lambda_{ex} = 351$ nm; $r_0 = 0.336$[72] $\lambda_{em} = 385$ nm (solvent dependent)
2-(*p*-Toluidino)naphthalene-6-sulfonate $\lambda_{ex} = 295$ nm $\lambda_{em} = 420$ nm (solvent dependent) $V = 450$ Å^3[64]	2-Methylanthracene $\lambda_{ex} = 382$ nm $V = 105$ Å^3[73] V: Calibration curve[65]
Perylene $\lambda_{ex} = 252$ nm; $r_0 = -0.165$[65] $\lambda_{ex} = 314$ nm; $r_0 = 0.100$[65] $\lambda_{ex} = 410$ nm; $r_0 = 0.335$[65] $\lambda_{ex} = 413$ nm; $r_0 = 0.370$[66] $\lambda_{em} = 446$ nm V: Calibration curve[66]	9-Vinylanthracene $\lambda_{ex} = 392$ nm V: Calibration curve[65]
1,6-Diphenylhexa-1,3,5-triene $\lambda_{ex} = 360$ nm $\lambda_{em} = 430$ nm $r_0 = 0.362$ $kV = 8.6 \times 10^5$ P deg^{-1} s^{-1}[67]	Anthracene $\lambda_{ex} = 366$ nm; $r_0 = 0.370$[74] $V = 57.3$ Å^3[74]
Rhodamine 6G $\lambda_{ex} = 514.5$ nm, $r_0 = 0.380$[68] $\lambda_{em} = 560$ nm $r_0 = 0.372$[69] radius = 5.6 Å[68]	Rhodamine 101 $\lambda_{ex} = 514.5$ nm $\lambda_{em} = 580$ nm $r_0 = 0.374$[69]
Fluorescein $r_0 = 0.374$[69]	Rhodamine B $r_0 = 0.374$[69]
TBP $\lambda_{ex} = 272$ nm, $r_0 = -0.160$[70] $\lambda_{ex} = 416$ nm, $r_0 = 0.389$[70] $\lambda_{em} = 440-480$ nm	*N*,*N*-bis(2,5-*tert*-butylphenyl)-3,4,9,10-perylenetetracarboximide $\lambda_{ex} = 514.5$ nm; $r_0 = 0.382$[75] $\lambda_{ex} = 520$ nm; $r_0 = 0.37$[76] $\lambda_{em} = 580$ nm
4-Amino-4′-nitrostilbene $r_0 = 0.3125$[71] $V = 1533$ Å^3[71]	1-Dimethylaminonaphthalene-5-sulfoamide $\lambda_{ex} = 360$ nm; $r_0 = 0.33$[77]
	4-Dimethylamino-4′-nitrostilbene $r_0 = 0.3278$[71] $V = 1724$ Å^3[71]

Rotational decays become equal, however, if four *tert*-butyl groups are substituted symmetrically into perylene. Anisotropy decay of TBP is monoexponential.[70]

The following representative examples illustrate the type of information that anisotropic probes can provide. Jobe and Verrall[64] employed polarized fluorescence emission to determine microviscosity of ternary mixtures containing 2-butoxyethanol, cetyltrimethylammonium bromide (CTAB) and water. Diphenylbutadiene was used as anisotropic probe molecule. Strong columbic binding prevented the authors from using potassium 2-(*p*-toluidino)naphthalene-6-sulfonate. At fixed surfactant concentration, the microviscosity was found to decrease with increasing 2-butoxyethanol concentration. The polar alkoxyalcohol effectively penetrated into the palisade region causing the CTAB micelle to swell. As part of the investigation the microviscosity of aqueous CTAB, sodium dodecyl sulfate (SDS), tetradecyltrimethylammonium bromide (TTAB) and cetylpyridinium bromide (CPyB) were also determined. Calculated values ranged from $\eta \approx 14$ cP to $\eta \approx 29$ cP, and were in excellent agreement with published values based upon the polarized fluorescence behavior of anthracene and 2-methylanthracene.

Betts et al.[72] estimated solute–fluid cluster sizes in supercritical N_2O based upon the measured fluorescence anisotropic and lifetime data for the probe molecule PRODAN. Measured steady-state anisotropies in supercritical N_2O decreased with increasing pressure, whereas in liquid N_2O the observed anisotropy remained essentially constant. Two limiting cases were considered. In the first case, the authors assumed only minimal interaction between the probe PRODAN and surrounding N_2O solvent molecules. The probe was free to rotate within the locally enriched solvent cluster environment. Strong probe–solvent interactions were assumed in the second model. Probe plus cluster was now treated as a single rotating body. Analysis of experimental data within this framework yielded an average cluster size containing

between 25 and 103 N_2O molecules near the critical pressure.

More recently, time-resolved anisotropic measurements[81] have been utilized to examine rotational reorientation dynamics of Aerosol-OT reverse micelles formed in near-critical propane solvent. Cationic anisotropic probes rhodamine 6G and rhodamine 101 were solubilized within the reverse micelle. Experimental data were interpreted using a bi-exponential rate law expression, with the rotational correlation times describing both the independent lateral rotation diffusion of the fluorophore along the water–headgroup interfacial region within the reverse micelle and the global rotational motion of the entire micelle. Rotational correlation times for the micelle were 40 times slower than expected, and suggested the formation of micelle–micelle aggregates. The study complemented the authors' earlier work involving rotational motion of several xanthene dyes (rhodamine 6G, rhodamine B, rhodamine 101 and fluorescein). Here, probes were solubilized in Aerosol-OT reverse micelles formed in heptane.

9 FLUORESCENT TEMPERATURE PROBES

Fluorescence-based temperature sensing is becoming increasingly more popular for monitoring processes occurring in microscopic and hostile environments where conventional thermometry is either impractical or impossible. Temperature-sensitive fluorescent probes provide a relatively nonintrusive method to study biological cells. Published applications include temperature measurement in evaporating fuel sprays and burning flames, examination of temperature fluctuations on surfaces of turbine engines and monitoring dynamic heat transfer in chemical reactions. In each application, a systematic change was observed in spectral properties of the probe molecule with increasing/decreasing temperature. Observed spectral changes were generally manifested as either shifts in the emission wavelength(s) or variations in measured intensities of selected emission bands caused by thermal excitation of vibronic states in the electronic excited state (i.e. fluorescence hot-band emission), changes in the microenvironment's refractive index and dielectric constant, or by differences in the monomer/excimer (or monomer/exciplex) concentration ratio in systems where excimer (or exciplex) formation occurs.

Accurate temperature determination requires that measured fluorescence spectra be free of inherent fluctuations in the intensity of the excitation radiation source and other instrumental artifacts. Strategies that have proved successful in this regard have involved the design of probe molecules having two fluorescent moieties or the utilization of molecules that exhibit an emission intensity enhancement of select vibronic bands. The emission intensity of the non-enhanced band(s) would serve as an internal reference point, thus reducing the problems associated with measurements based upon a single emission process.

Of the probe molecules that have been suggested as fluorescent thermometers, *N*-(1-pyrenylmethyl)-1-pyrenebutamide (PMPBA) [187589-23-7] and *N*-(1-pyrenylmethyl)-1-pyreneacetamide (PMPAA) [187589-24-8] are perhaps the most versatile for applications between ambient room temperature and 100 °C.[82] The two pyrenyl fluorophoric centers give each molecule a high quantum yield and good sensitivity. As shown in Figure 7 the intramolecular excimer emission (which is centered around $\lambda_{em} \approx 448$ nm in the case of PMPBA) decreases with increasing temperature. Normalization of excimer emission signals to those of the monomer eliminates the effects caused by excitation source fluctuations and variations in fluorophore concentrations and optical pathlength. Measured natural logarithm of $F_{excimer}/F_{monomer}$ emission intensity ratio versus reciprocal of the absolute Kelvin temperature, $1/T$, is nearly linear over the entire recommended temperature range. Lou et al.[82] estimated the precision of temperature measurement to be within $\Delta T = \pm 0.7$ °C and $\Delta T = \pm 0.8$ °C for PMPBA and PMPAA, respectively (see Figure 8). 1,3-Bis-(1′-pyrenyl)propane (PyPyP) [61549-24-4] and 3-(4′-dimethylaminophenyl)-1-(1′-pyrenyl)propane (PyDMAP) [38764-41-9] can be used for measuring temperatures above 100 °C. The latter two molecules also exhibit temperature-dependent

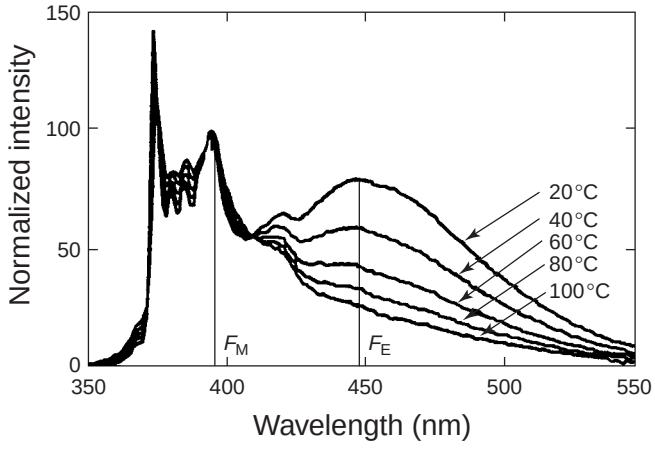

Figure 7 Observed fluorescence emission spectra of 0.33 μM PMPBA dissolved in dodecane as a function of temperature. Spectra have been normalized to the intensity of the monomer band, which occurs at about 396 nm. (Redrawn from J. Lou, T.A. Hatton, P.E. Laibinis, *Anal. Chem.*, **69**, 1262–1264. Copyright (1997) American Chemical Society.)

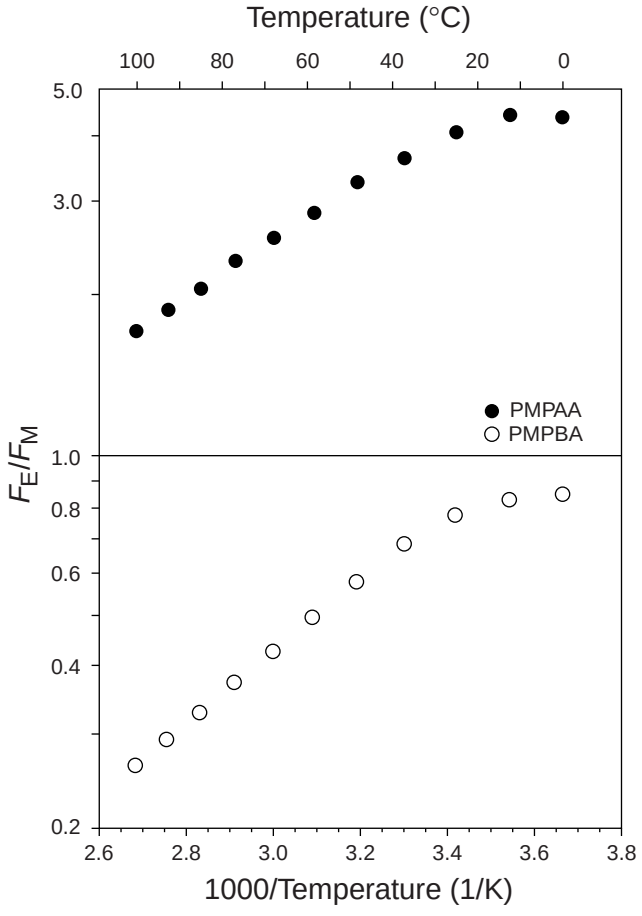

Figure 8 Relationship between F_E/F_M and temperature for 0.33 μM solutions of fluorophores PMPBA and PMPAA in dodecane. (Redrawn from J. Lou, T.A. Hatton, P.E. Laibinis, *Anal. Chem.*, **69**, 1262–1264. Copyright (1997) American Chemical Society.)

$F_{excimer}/F_{monomer}$ (PyPyP) and/or $F_{exciplex}/F_{monomer}$ (PyDMAP) fluorescent behavior. Both compounds are reported to be thermally stable for 10–20 minutes at temperatures of 300–400 °C.[83]

1,4-Di[2-(5-phenyloxazoyl)]benzene [1806-34-4], 2-(1-naphthyl)-5-phenyloxazole [846-63-9], *N*,*N*-bis(2,5-di-*tert*-butylphenyl)-3,4,9,10-perylene dicarboximide [83054-80-2] and BPe (to a much lesser extent) have been recommended as fluorescent temperature probes for measurements in the 25–170 °C range. Temperature is determined from nearly linear plots of $R_n(t)$ versus temperature [where $R_n(t)$ is the ratio of integrated emission intensity in the blue range to that in the red range at the system temperature T, normalized to the emission intensity ratio measured at ambient room temperature]. All four of the suggested probe molecules exhibit high fluorescent quantum yields and are insensitive to quenching by molecular oxygen. Expected temperature resolution is approximately ±4 °C.[84]

Fluorescent thermometers have been developed for measuring flame temperatures using anomalous $S_0 \leftarrow S_2$ fluorescence emission of isolated Py molecules.[85–87] Emission intensity ratio for the $S_0 \leftarrow S_2$ and $S_0 \leftarrow S_1$ spectral transitions is known to depend strongly upon both excitation and temperature. At low temperatures (425–625 K) and pressures (less than 50 torr) fluorescence emission is related directly to the internal vibrational temperature of the Py molecules. A plot of logarithm of S_2/S_1 emission intensity ratio versus the reciprocal of absolute Kelvin temperature is linear over the 425–625 K temperature region. Py is an organic molecule and it does undergo chemical reactions in the more oxidative flames, which limit its applicability to nonoxidative environments. Suggested systems that could be monitored with a Py probe molecule include high-temperature exhausts, diffusion flames and nonreactive flows of interest in heat transfer and fluid mechanics.

In recent years temperature measurements have also been performed with fluorescent probes that exhibit systematic shifts in emission wavelengths with temperature. Extreme caution should be exercised in using this type of probe molecule since it has not yet been conclusively established that the observed emission wavelength response depends solely upon temperature. Changes in solvent properties also contribute to the observed fluorescent behavior as discussed in the section devoted to fluorescent solvent polarity probes. It is quite conceivable that changes in the solvent's refractive index and/or dielectric constant may have contributed in part to observed probe character. Both dielectric constant and refractive index are temperature-dependent, and calibration curves established in one solvent media might not be transferable to another solvent media or molecularly organized system. Until this point is satisfactorily addressed it is recommended that the calibration curve relating the observed emission wavelength and temperature be made in a solvent whose physical properties closely mimic those of the intended application. This should help in minimizing any affects arising from changes in solvent properties brought about by temperature change.

10 FAR-VISIBLE AND NEAR-INFRARED ABSORPTION AND LUMINESCENT PROBES

Far-VIS and near-infrared absorption and luminescent probes[88,89] are currently used in biological and medical applications as in vivo sensors/tracers. Few biologically active molecules absorb in the 600–1000 nm spectral region, and attachment of a near-infrared probe molecule

Table 9 Spectral properties of select near-infrared probes

Probe/property	Probe/property
Noncovalent binding probes:	3,3′-Diethyloxatricarbocyanine iodide
cyanine iodide	$\lambda_{abs} = 677\,nm$[90]
$\lambda_{ex} = 746\,nm$	3,3′-Diethyloxadicarbocyanine iodide
$\lambda_{ex} = 775\,nm$[90]	$\lambda_{abs} = 577\,nm$[90]
3,3′-Diethythiadicarbocyanine iodide	Laser dye IR-125
$\lambda_{abs} = 646\,nm$[90]	$\lambda_{abs} = 780–786\,nm$[95]
Indocyanine Green	Squarilium dyes
$\lambda_{ex} = 780\,nm$	$\lambda_{abs} = 655–850\,nm$[89,96]
$\lambda_{ex} = 830\,nm$[91]	Croconium dyes
bis-Indole derivatives of	$\lambda_{abs} = 715–950\,nm$[89]
Heptamethine cyanines	NK-427 Dye
$\lambda_{abs} = 780–810\,nm$[92]	$\lambda_{ex} = 823\,nm$
bis-Benzindole derivatives of	$\lambda_{em} = 850\,nm$[97]
heptamethine cycanines	Transesterified derivatives of
$\lambda_{abs} = 750–830\,nm$[92]	laser dye IR-144
Covalent bonding probes:	$\lambda_{ex} = 755–796\,nm$
Laser dye IR-144	$\lambda_{ex} = 815–820\,nm$[93]
$\lambda_{ex} = 742\,nm$	BTPHB
$\lambda_{ex} = 819\,nm$[93]	$\lambda_{abs} = 698\,nm$ in water[95]
Isothiocyanato-analogs of	$\lambda_{abs} = 811\,nm$ in methanol[95]
Indocyanine Green	
$\lambda_{abs} = 780–790\,nm$[94]	

BTPHB, 2-[4′-chloro-7′(3″-ethyl-2″-benzothiazolinylidenyl)-3′,5′-(1‴,3‴-propanediyl)-1′,3′,5′-heptatriene-1′-yl]-3-ethylbenzothiazolium bromide.

provides a means to label the analyte molecule of interest. Near-infrared probes typically possess large molar absorptivities ($\varepsilon \approx 200\,000\,cm^{-1}\,M^{-1}$) and reasonably good quantum yields ($\phi_F \approx 0.8$), which facilitate spectrophotometric detection. Detection limit depends entirely upon spectral characteristics of the probe molecule and instrumentation capability. Background corrections for sample absorbance and/or fluorescence are negligible.

Molecules that exhibit long-wavelength absorption and emission generally have extensive delocalized double-bond conjugation. Two major families of molecules known to have long-wavelength absorption maxima are polymethine cyanine and phthalocyaninine dyes. These dyes are practically insoluble in nonpolar organic solvents because of their ionic character; however, they do dissolve in most polar organic solvents and binary aqueous–organic mixtures. Polar and ionic functional groups can be added to judiciously alter molecules' spectral properties and water solubility, or to provide a site for analyte attachment. Attachment involves either noncovalent binding or covalent bonding as in the case of a biological molecule, or complexation to a metal ion. For example, water solubility can be enhanced by the presence of a sulfonate group, whereas an added isothiocyanate group would allow the probe molecule to be selectively attached to the amino groups of proteins. Thiourea bond that would be formed is quite stable above a pH of 6.0. Several near-infrared probes are listed in Table 9, along with their spectral properties.

Far-VIS and near-infrared absorbing dyes have only recently been introduced into analytical chemistry, and their applications as spectroscopic probes have not yet been fully realized. Patonay and coworkers[95] explored the feasibility of using polymethine cyanine dyes as solvent polarity probes, based upon their aggregation tendency in binary aqueous–alcohol mixtures. Of the four dyes studied, only IR-144 [54849-69-3] and BTPHB showed any real noticeable spectral change with solvent composition. The effect was more dramatic for BTPHB in that over a 110 nm shift was observed in the absorption maxima wavelength between water ($\lambda_{abs} \approx 698\,nm$) and neat methanol ($\lambda_{abs} \approx 810\,nm$). Recorded absorption spectra contained two peaks at intermediate binary solvent compositions as shown in Figure 9. Peaks were well-resolved and required no deconvolution to accurately quantify. The peak at $\lambda_{abs} \approx 698\,nm$ was attributed to formation of dimers (or perhaps higher associated aggregates). The authors suggested that the calculated $A_{monomer}/(A_{monomer} + A_{dimer})$ absorbance ratio could serve as a useful measure of *solvent hydrophobicity*. Previously, it had been reported[97] that the fluorescent lifetime of the polymethine dye 3,3′-diethyl-2,2′-(4,5 : 4′,5′-dibenzo)dithiatricarbocyanine iodide (NK-427) [20682-18-2] was highly correlated with solvent dielectric constant. Further experimental research is needed,

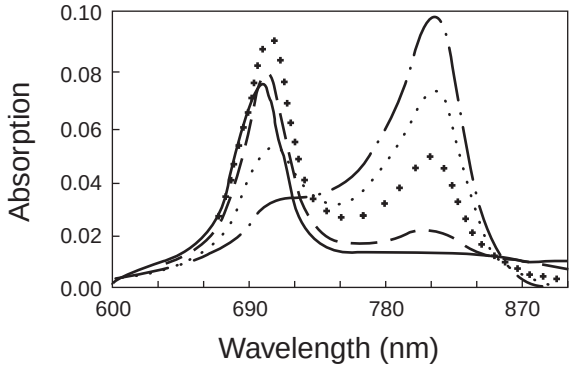

Figure 9 Representative spectra of BTPHB in water (———), and 20% (– – – –), 30% (++++), 40% (· · · ·) and 50% (– · —·) methanol. (Redrawn with permission from G. Patonay, M.D. Antoine, S. Devanthan, L. Strekowski, *Appl. Spectrosc.*, **45**, 457–461. Copyright (1991) Society for Applied Spectroscopy.)

however, to firmly establish the BTPHB and NK-427 scales as meaningful measures of solvent hydrophobicity.

To date, less than a dozen near-infrared dyes have been used as probe molecules in actual pH measurements. Published examples include pH determinations based upon measured emission intensities of three immobilized tetrasubstituted chloroaluminum 2,3-naphthocyanine dyes[98] [functional groups were $(5\text{-NC})_4-$, $(5\text{-O}_2\text{N})_4-$ and $(5\text{-HOOC})_4-$] and a bis(carboxylic acid) derivative of 2′{4′-chloro-7′-[2″-(1″-ethyl-3″,3″-dimethylindoleninylidene)]-3′,5′-(1‴,3‴-propanediyl)-1′,3′,5′-heptatrienyl-1′-yl}-1-ethyl-3,3-dimethylindoleinine bromide,[99] and upon hydrogen ion-induced changes in absorption spectra of aminodienone-carbocyanine dye system.[100] In the latter example, spectral change resulted from keto–enol tautomerism caused by protonation of oxygen atom. Generally, such spectral changes occur over a fairly narrow pH range of less than 2–3 pH units. Finally several far-VIS and near-infrared spectroscopic probes such as porphines, isoporphines and corroles, as well as select cyclic derivatives of phthalocyanines and naphthalocyanines, have served as complexing agents for metal ions. Specificity of probe for various metal ions in solution is determined by both ring size and analyte's affinity for the lone electron pairs on nitrogen ring atoms. Additional analytical applications will likely result in the next few years. New synthetic procedures are making probe molecules more readily available.

ABBREVIATIONS AND ACRONYMS

BA	9,9′-Bianthracene
BACl	10-Chloro-9,9′-bianthracene
BCECF	Biscarboxyethylcarboxyfluorescein
BCo	Benzo[*a*]coronene
BePy	Benzo[*e*]pyrene
BO	Born–Oppenheimer
BPe	Benzo[*ghi*]perylene
BPP	Benzo[*rst*]pentaphene
BTPHB	2-[4′-Chloro-7′(3″-ethyl-2″-benzothiazolinylidenyl)-3′,5′-(1‴,3‴-propanediyl)-1′,3′,5′-heptatriene-1′-yl]-3-ethylbenzothiazolium bromide
C9A	*N*-(9-Anthryl)carbazole
CD	Cyclodextrin
CF	Carboxyfluorescein
Co	Coronene
CPyB	Cetylpyridinium bromide
CTAB	Cetyltrimethylammonium bromide
DCo	Dibenzo[*a*, *j*]coronene
DHBPe	3,4-Dihydrobenzo[*ghi*]perylene
DiMCF	Dimethylcarboxyfluorescein
DiMeCo	1,2-Dimethylcoronene
ET-30	2,6-Diphenyl-4-(2,4,6-triphenyl-*N*-pyridinio)phenolate
ET-33	2,6-Dichloro-4-(2,4,6-triphenyl-*N*-pyridino)phenolate
GLC	Gas–Liquid Chromatographic
HMCo	7-Hydroxy-4-methylcoumarin
HPLC	High-performance Liquid Chromatography
HT	Herzberg–Teller
LE	Locally Excited
MeCo	Methylcoronene
NCo	Naphtho[2,3*a*]coronene
NK-427	3,3′-Diethyl-2,2′-(4,5 : 4′,5′-dibenzo)-dithiatricarbocyanine iodide
Ov	Ovalene
PAH	Polycyclic Aromatic Hydrocarbon
PANH	Polycyclic Aromatic Nitrogen Heterocycle
PCNMP	4-Phenyl-4-(4-cyano-1-naphthylmethylene)piperidine
PMPAA	*N*-(1-Pyrenylmethyl)-1-pyreneacetamide
PMPBA	*N*-(1-Pyrenylmethyl)-1-pyrenebutamide
PNMP	1-Phenyl-4-(2-naphthylmethylene)-piperidine
PRODAN	6-Propionyl-2-(dimethylamino)-naphthalene
Py	Pyrene
PyDMAP	3-(4′-Dimethylaminophenyl)-1-(1′-pyrenyl)propane
PyPyP	1,3-Bis-(1′-pyrenyl)propane
SDS	Sodium dodecyl sulfate
SNAFL	Seminaphthofluorescein
SNARF	Seminaphthorhodafluor
TBP	2,5,8,11-Tetra-*tert*-butylperylene

TICT	Twisted Intramolecular Charge Transfer
TNS	*p*-Toluidinonaphthalene sulfonate
TTAB	Tetradecyltrimethylammonium bromide
UV	Ultraviolet
VIS	Visible

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy • Fluorescence in Organized Assemblies • Fluorescence Lifetime Measurements, Applications of • Near-infrared Absorption/Luminescence Measurements • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

REFERENCES

1. K.W. Street, Jr, W.E. Acree, Jr, 'Estimation of the Effective Dielectric Constant of Cyclodextrin Cavities Based on the Fluorescence Properties of Pyrene-3-carboxyaldehyde', *Appl. Spectrosc.*, **42**, 1315–1318 (1988).
2. J. Stahlberg, M. Almgren, 'Polarity of Chemically Modified Silica Surfaces and Its Dependence on Mobile-phase Composition by Fluorescence Spectrometry', *Anal. Chem.*, **57**, 817–821 (1985).
3. R.S. Helburn, S.C. Rutan, J. Pompano, D. Mitchem, W.T. Patterson, 'Solvatochromic Studies of Solvated Chromatographic Stationary Phases', *Anal. Chem.*, **66**, 610–618 (1994).
4. G.J. Burrell, R.J. Hurtubise, 'Fluorescence Investigation of the Chromatographic Interactions of Nitrogen Heterocycles with Deactivated Silica', *Anal. Chem.*, **60**, 2178–2182 (1988).
5. G.J. Burrell, R.J. Hurtubise, 'Solid-surface Luminescence Interactions of Nitrogen Heterocycles Adsorbed on Silica Gel Chromatoplates Submerged in Chloroform/*n*-Hexane Solvents', *Anal. Chem.*, **60**, 564–568 (1988).
6. B. Valeur, 'Fluorescent Probes for Evaluation of Local Physical and Structural Parameters', in *Molecular Luminescence Spectroscopy, Part 3*, ed. S. Schulman, John Wiley & Sons, New York, 25–84, 1993.
7. C. Reichardt, 'Empirical Parameters of Solvent Polarity', *Angew. Chem.*, **77**, 30–40 (1965).
8. C. Reichardt, 'Solvatochromic Dyes as Solvent Polarity Indicators', *Chem. Rev.*, **94**, 2319–2358 (1994).
9. E.M. Kosower, 'The Effect of Solvent on Spectra. I. A New Empirical Measure of Solvent Polarity: Z-Values', *J. Am. Chem. Soc.*, **80**, 3253–3260 (1958).
10. M.J. Kamlet, J.M. Abboud, M.H. Abraham, R.W. Taft, 'Linear Solvation Energy Relationships. 23. A Comprehensive Collection of the Solvatochromic Parameters π^*, α and β, and Some Methods for Simplifying the Generalized Solvatochromic Equation', *J. Org. Chem.*, **48**, 2877–2887 (1983).
11. R.W. Taft, J.M. Abboud, M.J. Kamlet, M.H. Abraham, 'Linear Solvation Energy Relations', *J. Solution Chem.*, **14**, 153–186 (1985).
12. M.J. Kamlet, J.L. Abboud, R.W. Taft, 'The Solvatochromic Comparison Method. 6. The π^* Scale of Solvent Polarities', *J. Am. Chem. Soc.*, **99**, 6027–6038 (1977).
13. B.P. Johnson, M.G. Khaledi, J.G. Dorsey, 'Solvatochromic Solvent Polarity Measurements in Reversed-phase Liquid Chromatography', *Anal. Chem.*, **58**, 2354–2365 (1986).
14. P.W. Carr, R.M. Doherty, M.J. Kamlet, R.W. Taft, W. Melander, C. Horvath, 'Study of Temperature and Mobile-phase Effects in Reversed-phase High-performance Liquid Chromatography by the Use of the Solvatochromic Comparison Method', *Anal. Chem.*, **58**, 2674–2680 (1986).
15. P.C. Sadek, P.W. Carr, R.M. Doherty, M.J. Kamlet, R.W. Taft, M.H. Abraham, 'Study of Retention Processes in Reversed-phase High-performance Liquid Chromatography by Use of the Solvatochromic Comparison Method', *Anal. Chem.*, **57**, 2971–2978 (1985).
16. J. Barbosa, V. Sanz-Nebot, 'Preferential Solvation in Acetonitrile–Water Mixtures. Relationship Between Solvatochromic Parameters and Standard pH Values', *J. Chem. Soc., Faraday Trans.*, **90**, 3287–3292 (1994).
17. D.J. Phillips, J.F. Brennecke, 'Spectroscopic Measurement of Local Compositions in Binary Liquid Solvents and Comparison to the NRTL Equation', *I&EC Res.*, **32**, 943–951 (1993).
18. J. Goto, 'Fluorescence Derivatization', *Chromatogr. Sci.*, **48**, 323–358 (1990).
19. H. Lingeman, W.J.M. Underberg, 'Derivatization in Liquid Chromatography: Introduction', *Chromatogr. Sci.*, **48**, 1–50 (1990).
20. J.A.P. Meulendijk, W.J.M. Underberg, 'Ultraviolet/Visible Derivatization', *Chromatogr. Sci.*, **48**, 247–281 (1990).
21. O.S. Wolfbeis, 'Annual Chemical Congress: New Spectroscopic Sensors and Techniques', *Anal. Proc.*, **24**, 14–15 (1987).
22. E. Wänninen, 'Metal Indicators and Their Applications', in *Indicators*, eds. R. Belcher, H. Frieser, Pergamon Press, Oxford, 231–435, 1972.
23. H.A. Flaschka, A.J. Barnard, Jr (eds.), *Chelates in Analytical Chemistry*, Marcel Dekker, Inc., New York, Vol. 5, 1976.

24. R.P. Haugland (ed.), *Handbook of Fluorescent Probes and Research Chemicals*, 6th edition, Molecular Probes, Eugene, Oregon, 1996.
25. Q. Chang, J. Sipior, J.R. Lakowicz, G. Rao, 'A Lifetime-based Fluorescence Resonance Energy Transfer Sensor for Ammonia', *Anal. Biochem.*, **232**, 92–97 (1995).
26. A. Mills, Q. Chang, N. McMurray, 'Equilibrium Studies on Colorimetric Plastic Film Sensors for Carbon Dioxide', *Anal. Chem.*, **64**, 1383–1389 (1992).
27. Y.M. Lui, R. Pereiro-García, M.J. Valencia-González, M.E. Díaz-García, A. Sanz-Medel, 'Evaluation of Some Immobilized Room-temperature Phosphorescent Metal Chelates as Sensing Materials for Oxygen', *Anal. Chem.*, **66**, 836–840 (1994).
28. C. Preininger, I. Klimant, O.S. Wolfbeis, 'Optical Fiber Sensor for Biological Oxygen Demand', *Anal. Chem.*, **66**, 1841–1846 (1994).
29. W. Xu, R.C. McDonough, III, B. Langsdorf, J.N. Demas, B.A. DeGraff, 'Oxygen Sensors Based on Luminescence Quenching: Interactions of Metal Complexes with the Polymer Supports', *Anal. Chem.*, **66**, 4133–4141 (1994).
30. J.A. Dean, *Analytical Chemistry Handbook*, McGraw-Hill, 3.31–3.34, 1995.
31. S. Tucker, R. Robinson, C. Keane, M. Boff, M. Zenko, S. Batish, K.W. Street, Jr, 'Colorimetric Determination of pH', *J. Chem. Educ.*, **66**, 769–771 (1989).
32. M.L. Graber, D.C. DiLillo, B.D. Friedman, E. Pasotriza-Munoz, 'Characteristics of Fluoroprobes for Measuring Intracellular pH', *Anal. Biochem.*, **156**, 202–212 (1986).
33. O.S. Wolfbeis, E. Fuerlinger, H. Kroneis, H. Marsoner, 'Fluorimetric Analysis. 1. A Study on Flourescent Indicators for Measuring Near Neutral ('Physiological') pH Values', *Fresenius' Z. Anal. Chem.*, **314**, 119–124 (1983).
34. I. Kurtz, R.S. Balaban, 'Fluorescence Emission Spectroscopy of 1,4-Dihydroxyphthalonitrile: A Method for Determining Intracellular pH in Cultured Cells', *Biophys. J.*, **48**, 499–508 (1985).
35. G. Gabor, D.R. Walt, 'Sensitivity Enhancement of Fluorescent pH Indicators by Inner-filter Effects', *Anal. Chem.*, **63**, 793–796 (1991).
36. E.D. Niemeyer, F.V. Bright, 'The pH Within PFPE Reverse Micelles Formed in Supercritical CO_2', *J. Phys. Chem., B*, **102**, 1474–1478 (1998).
37. J.E. Whitaker, R.P. Haughland, D. Ryan, P.C. Hewitt, R.P. Haugland, F.G. Prendergrast, 'Fluorescent Rhodol Derivatives: Versatile, Photostable Labels and Tracers', *Anal. Biochem.*, **207**, 267–279 (1992).
38. H. Szmacinski, J.R. Lakowicz, 'Optical Measurement of pH Using Fluorescence Lifetime and Phase-modulation Fluorometry', *Anal. Chem.*, **65**, 1668–1674 (1993).
39. T.L. Cecil, S.C. Rutan, 'Correction for Fluorescence Response Shifts in Polyaromatic Hydrocarbon Mixtures with an Innovations-based Kalman Filter Method', *Anal. Chem.*, **62**, 1998–2004 (1990).
40. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum Press, New York, 1983.
41. M. Hof, P. Lianos, 'Structural Studies of Thin AOT Films by Using the Polarity Fluorescent Probe Coumarin-153', *Langmuir*, **13**, 290–294 (1997).
42. H. Träuble, P. Overath, 'The Structure of *Escherichia coli* Membranes Studied by Fluorescence Measurements of Lipid Phase Transitions', *Biochim. Biophys. Acta*, **307**, 491–512 (1973).
43. K. Kalyanasundaran, J.K. Thomas, 'Solvent-dependent Fluorescence of Pyrene-3-carboxaldehyde and Its Applications in the Estimation of Polarity at Micelle–Water Interfaces', *J. Phys. Chem.*, **81**, 2176–2180 (1977).
44. M.A. Kessler, O.S. Wolfbeis, 'New Highly Fluorescent Ketocyanine Polarity Probes', *Spectrochim. Acta*, **47A**, 187–192 (1991).
45. C.F. Chapman, M. Maroncelli, 'Fluorescence Studies of Solvation and Solvation Dynamics in Ionic Solutions', *J. Phys. Chem.*, **95**, 9095–9114 (1991).
46. G.F. Mes, B. de Jong, H.J. van Ramesdonk, J.W. Verhoeven, J.M. Warman, M.P. de Haas, L.E.W. Horsman-van den Dool, 'Excited-state Dipole Moment and Solvatochromism of Highly Fluorescent Rod-shaped Bichromophoric Molecules', *J. Am. Chem. Soc.*, **106**, 6524–6528 (1984).
47. G. Weber, F.J. Farris, 'Synthesis and Spectral Properties of a Hydrophobic Fluorescent Probe: 6-Propionyl-2-(dimethylamino)naphthalene', *Biochemistry*, **18**, 3075–3078 (1979).
48. G.W. Robinson, R.J. Robbins, G.R. Fleming, J.M. Morris, A.E.W. Knight, R.J.S. Morrison, 'Picosecond Studies of the Fluorescence Probe Molecule 8-Anilino-1-naphthalenesulfonic Acid', *J. Am. Chem. Soc.*, **100**, 7145–7150 (1978).
49. L. Stryer, 'The Interaction of a Naphthalene Dye With Apomyoglobin and Apohemoglobin: A Fluorescent Probe of Nonpolar Binding Sites', *J. Mol. Biol.*, **13**, 482–495 (1965).
50. L. Coosemans, F.C. De Schryver, A. Van Dormael, 'Influence of Solvent Polarities on Fluorescence Probes. 7-Amino- and 7-*N*-Dimethylamino-4-methylcoumarin', *Chem. Phys. Lett.*, **65**, 95–99 (1979).
51. W.O. McClure, G.M. Edelman, 'Fluorescent Probes for Conformational States of Proteins. I. Mechanism of Fluorescence of 2-*p*-Toluidinylnaphthalene-6-sulfonate, a Hydrophobic Probe', *Biochemistry*, **5**, 1908–1919 (1966).
52. P. Greenspan, S.D. Fowler, 'Spectrofluorometric Studies of the Lipid Probe, Nile Red', *J. Lipid Res.*, **26**, 781–789 (1985).
53. J. Catalan, P. Perez, J. Laynez, F. Garcia Blanco, 'Analysis of the Solvent Effect on the Photophysics Properties of 6-Propionyl-2-(dimethylamino)naphthalene (PRODAN)', *J. Fluoresc.*, **1**, 215–223 (1991).
54. W. Rettig, M. Zander, 'Broken Symmetry in Excited States of 9,9′-Bianthryl', *Ber. Bunsen-Ges. Phys. Chem.*, **87**, 1143–1149 (1983).

55. O. Kajimoto, M. Futakami, T. Kobayashi, K. Yamasaki, 'Charge-transfer-state Formation in Supercritical Fluid: (*N*,*N*-Dimethylamino)benzonitrile in CF_3H', *J. Phys. Chem.*, **92**, 1347–1353 (1988).
56. Y.-P. Sun, G. Bennett, K.P. Johnston, M.A. Fox, 'Quantitative Resolution of Dual Fluorescence Spectra in Molecules Forming Twisted Intramolecular Charge-transfer States. Toward Establishment of Molecular Probes for Medium Effects in Supercritical Fluids and Mixtures', *Anal. Chem.*, **64**, 1763–1768 (1992).
57. M.E. Baraka, R. García, E. Quiñones, 'A Study of the Inclusion Complexes of β-Cyclodextrin with Three Electronic States of 4-(*N*,*N*-Dimethylamino)benzonitrile', *J. Photochem. Photobiol. A: Chem.*, **79**, 181–187 (1994).
58. D.C. Dong, M.A. Winnik, 'The Py Scale of Solvent Polarities', *Can. J. Chem.*, **62**, 2560–2565 (1984).
59. W.E. Acree, Jr, S.A. Tucker, J.C. Fetzer, 'Fluorescence Emission Properties of Polycyclic Aromatic Compounds in Review', *Polycyclic Aromat. Compds.*, **2**, 75–105 (1991).
60. S.A. Tucker, W.E. Acree, Jr, J.C. Fetzer, 'Polycyclic Aromatic Hydrocarbon Solute Probes. Part XI. Unusual Solvent-modulated Fluorescence Emission Behavior of 1-Methylcoronene and Select Dimethylcoronenes in Nonelectrolyte Solvents', *Appl. Spectrosc.*, **49**, 8–14 (1995).
61. D.S. Karpovich, G.J. Blanchard, 'Relating the Polarity-dependent Fluorescence Response of Pyrene to Vibronic Coupling. Achieving a Fundamental Understanding of the Py Polarity Scale', *J. Phys. Chem.*, **99**, 3951–3958 (1995).
62. W.E. Acree, Jr, S.A. Tucker, D.C. Wilkins, 'Spectrochemical Investigations of Preferential Solvation. Fluorescence Emission Behavior of Select Polycyclic Aromatic Hydrocarbon Solute Probes Dissolved in Mixed Solvents', *J. Phys. Chem.*, **97**, 11 199–11 203 (1993).
63. W. Xu, J.N. Demas, B.A. DeGraff, M. Whaley, 'Interactions of Pyrene with Cyclodextrins and Polymeric Cyclodextrins', *J. Phys. Chem.*, **97**, 6546–6554 (1993).
64. D.J. Jobe, R.E. Verrall, 'Polarized Fluorescence Emission Measurements in Mixtures of 2-Butoxyethanol, Cetyltrimethylammonium Bromide, and Water', *Langmuir*, **6**, 1750–1751 (1990).
65. Measured in propylene glycol at −60 °C, J.R. Lakowitz, J.R. Knutson, 'Hindered Depolarizing Rotations of Perylene in Lipid Bilayers. Detection by Lifetime-resolved Fluorescence Anisotropy Measurements', *Biochemistry*, **19**, 905–911 (1980).
66. U. Cogan, M. Shinitzky, G. Weber, T. Nishida, 'Microviscosity and Order in the Hydrogen Region of Phospholipid and Phospholipid–Cholesterol Dispersions Determined with Fluorescent Probes', *Biochemistry*, **12**, 521–527 (1973).
67. M. Shinitzky, M. Inbar, 'Differences in Microviscosity Induced by Different Cholesterol Levels in the Surface Membrane Lipid Layer of Normal Lymbocytes Malignant Lymphoma Cells', *J. Mol. Biol.*, **85**, 603–615 (1974).
68. U. Narang, R. Wang, P.N. Prasad, F.V. Bright, 'Effects of Aging on the Dynamics of Rhodamine 6G in Tetramethyl Orthosilicate-derived Sol–Gels', *J. Phys. Chem.*, **98**, 17–22 (1994).
69. L.B.-Å Johansson, 'Limiting Fluorescence Anisotropies of Perylene and Xanthene Derivatives', *J. Chem. Soc., Faraday Trans.*, **86**, 2103–2107 (1990).
70. S. De Backer, M.R. Negri, S. De Feyter, G.B. Dutt, M. Ameloot, F.C. De Schryver, K. Müllen, F. Holtrug, 'Fluorescence Depolarization from 2,5,8,11-Tetra-*t*-butylperylene Globally Analysed Upon Excitation in S_1 and S_n, *Chem. Phys. Lett.*, **233**, 538–544 (1995).
71. A. Kawski, J. Kaminski, J. Kukielski, 'Thermal Motion of Prolate Luminescent Molecules in Solvents With Different Viscosities', *Z. Naturforsch.*, **34A**, 702–707 (1979).
72. T.J. Betts, J. Zagrobelny, F.V. Bright, 'Spectroscopic Determination of Solute–Fluid Cluster Size in Supercritical Nitrogen Oxide', *J. Am. Chem. Soc.*, **114**, 8163–8171 (1992).
73. M. Grätzel, J.K. Thomas, 'On the Dynamics of Pyrene Fluorescence Quenching in Aqueous Ionic Micellar Systems, Factors Affecting the Permeability of Micelles', *J. Am. Chem. Soc.*, **95**, 6885–6889 (1973).
74. R.K. Bauer, 'Emission Anisotropy of Fluorescence of Anthracene Solutions. I.', *Acta Phys. Pol.*, **35**, 101–116 (1969).
75. D. Ben-Amotz, J.M. Drake, 'The Solute Size Effect in Rotational Diffusion Experiments: A Test of Microscopic Friction Theories', *J. Chem. Phys.*, **89**, 1019–1029 (1988).
76. A.M. Williams, D. Ben-Amotz, 'Molecular-Optical Viscometer Based on Fluorescence Depolarization', *Anal. Chem.*, **64**, 700–703 (1992).
77. M. Shinitzky, 'Fluidity and Order in the Hydrocarbon–Water Interface of Synthetic and Biological Micelles as Determined by Fluorescence Polarization', *Israel J. Chem.*, **12**, 879–890 (1974).
78. N.J. Turro, M. Aikawa, A. Yekta, 'A Comparison of Intermolecular and Intramolecular Excimer Formation in Detergent Solutions. Temperature Effects and Microviscosity Measurements', *J. Am. Chem. Soc.*, **101**, 772–774 (1979).
79. N.J. Turro, C.-J. Chung, 'Photoluminescent Probes for Water-soluble Polymers. Pressure and Temperature Effects on a Polyol Surfactant', *Macromolecules*, **17**, 2123–2126 (1984).
80. J. Emert, C. Behrens, M. Goldenberg, 'Intramolecular Excimer-forming Probes of Aqueous Micelles', *J. Am. Chem. Soc.*, **101**, 771–772 (1979).
81. M.P. Heitz, F.V. Bright, 'Rotational Reorientation Dynamics of Aerosol-OT Reverse Micelles Formed in Nearcritical Propane', *Appl. Spectrosc.*, **50**, 732–739 (1996).

82. J. Lou, T.A. Hatton, P.E. Laibinis, 'Fluorescent Probes for Monitoring Temperature in Organic Solvents', *Anal. Chem.*, **69**, 1262–1264 (1997).
83. H.E. Gossage, L.A. Melton, 'Fluorescence Thermometers Using Intramolecular Exciplexes', *Appl. Opt.*, **26**, 2256–2259 (1987).
84. F. Bai, L.A. Melton, 'High Temperature, Oxygen Resistent Molecular Fluorescence Thermometers', *Appl. Spectrosc.*, **51**, 1276–1280 (1997).
85. D.L. Peterson, F.E. Lytle, N.M. Laurendeau, 'Flame Temperature Measurements Using the Anomalous Fluorescence of Pyrene', *Appl. Optics*, **27**, 2768–2775 (1988).
86. D.L. Peterson, F.E. Lytle, N.M. Laurendeau, 'Determination of Flame Temperature Using the Anomalous Fluorescence of Pyrene', *Opt. Lett.*, **11**, 345 (1986).
87. T. Deinum, C.J. Werkhoven, J. Langelaar, R.P.H. Rettschnick, J.D.W. Van Voorst, 'Sequence Congestion and Temperature Effects on the Fluorescence of Isolated Pyrene Molecules', *Chem. Phys. Lett.*, **27**, 206–209 (1974).
88. G. Patonay, M.D. Antoine, 'Near-infrared Fluorogenic Labels: New Approach to an Old Problem', *Anal. Chem.*, **63**, 321A-327A (1991).
89. J. Fabian, H. Nakazumi, M. Matsuoka, 'Near-infrared Absorbing Dyes', *Chem. Rev.*, **92**, 1197–1226 (1992).
90. D. Andrews-Wilberforce, G. Patonay, 'Investigation of Near-infrared Laser Dye Albumin Complexes', *Spectrochim. Acta*, **46A**, 1153–1162 (1990).
91. T. Imasaka, H. Nakagawa, T. Okazaki, N. Ishibashi, 'Enzyme Immunoassay of Insulin by Semiconductor Laser Fluorometry', *Anal. Chem.*, **62**, 2404–2405 (1990).
92. N. Narayanan, G. Patonay, 'A New Method for the Synthesis of Heptamethine Cyanine Dyes: Synthesis of New Near-infrared Fluorescent Labels', *J. Org. Chem.*, **60**, 2391–2395 (1995).
93. D.B. Shealy, R. Lohrmann, J.R. Ruth, N. Narayanan, S.L. Sutter, G.A. Casay, L. Evans III, G. Patonay, 'Spectral Characterization and Evaluation of Modified Near-infrared Laser Dyes with DNA Sequenching', *Appl. Spectrosc.*, **49**, 1815–1820 (1995).
94. R.J. Williams, M. Lipowska, G. Patonay, L. Strekowski, 'Comparison of Covalent and Noncovalent Labeling with Near-infrared Dyes for High-performance Liquid Chromatographic Determination of Human Serum Albumin', *Anal. Chem.*, **65**, 601–605 (1993).
95. G. Patonay, M.D. Antoine, S. Devanthan, L. Strekowski, 'Near-infrared Probe for Determination of Solvent Hydrophobicity', *Appl. Spectrosc.*, **45**, 457–461 (1991).
96. A.J. Sophianopoulos, J. Lipowski, N. Narayanan, G. Patonay, 'Association of Near-infrared Dyes with Bovine Serum Albumin', *Appl. Spectrosc.*, **51**, 1511–1515 (1997).
97. T. Imasaka, A. Yoshitake, K. Hirata, Y. Kawabata, N. Ishibashi, 'Pulsed Semiconductor Laser Fluorometry for Lifetime Measurements', *Anal. Chem.*, **57**, 947–949 (1985).
98. G.A. Casay, F. Meadows, N. Daniels, A. Roberson, G. Patonay, 'Fiber-optic Probe Applications Using Near-infrared Compounds: Determination of NaOH', *Spectrosc. Lett.*, **28**, 301–326 (1995).
99. J.-M. Zen, G. Patonay, 'Near-infrared Fluorescence Probe for pH Determinations', *Anal. Chem.*, **63**, 2934–2938 (1991).
100. G. Patonay, G.A. Casay, M. Lipowska, L. Strekowski, 'Spectroscopic Studies of a Near-infrared Absorbing pH Sensitive Aminodienone–Carbocyanine Dye System', *Talanta*, **40**, 935–942 (1993).

Circular Dichroism and Linear Dichroism

Alison Rodger
University of Warwick, Coventry, UK

1 Introduction to Circular Dichroism **1**
1.1 Circularly Polarized Light 2
1.2 Normal Absorption Spectroscopy 3
1.3 Circular Dichroism Spectroscopy 3
1.4 Units of Circular Dichroism Spectroscopy 4
1.5 Optical Activity (Optical Rotation) 4
1.6 Enantiomeric Purity and Enantiomeric Excess 5
2 Structure Analysis from Circular Dichroism **5**
2.1 Introduction 5
2.2 Qualitative Description of Circular Dichroism and Its Correlation with Structure for Some Systems 6
2.3 Interaction of Radiation with Matter and the Rosenfeld Equation for Circular Dichroism 9
2.4 Circular Dichroism of Magnetic Dipole Allowed Transitions: Why the Octant Rule Works for Carbonyls 12
2.5 Circular Dichroism of Electric Dipole Allowed Transitions 15
3 Linear Dichroism **17**
3.1 Introduction 17
3.2 Orientation Parameters 18
3.3 Molecular Alignment Techniques 21
3.4 Examples of Linear Dichroism Data 22
4 Instrumentation and the Dichroism Experiment **24**
4.1 Circular Dichrographs (Spectropolarimeters) 24
4.2 Design and Implementation of an Linear Dichroism Experiment 26
Acknowledgments **26**
Abbreviations and Acronyms **26**
Related Articles **26**
Further Reading **27**
General Circular Dichroism References 27
General Linear Dichroism References 27
References **27**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Circular dichroism (CD) is the difference in absorption, A, of left and right circularly *polarized light, Equation (1):*

$$CD = \Delta A = A_l - A_r \qquad (1)$$

For randomly oriented systems such as solutions of molecules, only chiral molecules will show any CD intensity corresponding to their absorption bands. Chiral molecules are those molecules that cannot be superposed on their mirror images.[1] Chiral *is derived from the Greek word* χειρ *meaning hand, hence the alternate term for chirality, ''handedness''. Two molecules that are mirror images of each other are often referred to as* enantiomers *and equimolar mixtures of two enatiomers form a racemic mixture which has no net CD intensity in solution.*

Linear dichroism (LD) is the difference in absorption of light linearly *polarized parallel and perpendicular to an orientation axis, Equation (68):*

$$LD = A_{\parallel} - A_{\perp} \qquad (68)$$

1 INTRODUCTION TO CIRCULAR DICHROISM

CD is the difference in absorption, A, of left and right *circularly* polarized light, Equation (1):

$$\mathrm{CD} = \Delta A = A_\mathrm{l} - A_\mathrm{r} \qquad (1)$$

For randomly oriented systems such as solutions of molecules, only chiral molecules will show any CD intensity corresponding to their absorption bands. Chiral molecules are those molecules that cannot be superposed on their mirror images.[1] *Chiral* is derived from the Greek word χειρ meaning hand, hence the alternate term for chirality, ''handedness''. Two molecules that are mirror images of each other are often referred to as *enantiomers* and equimolar mixtures of two enatiomers form a racemic mixture which has no net CD intensity in solution (Figure 1).

In order to understand when chirality is important it is helpful to consider the everyday problem of a pair of shoes, a left foot, and a shoe box. The shoes and the foot are chiral and the left foot only fits comfortably into the left shoe, whereas the achiral shoe box cannot distinguish between the two shoes. Chirality is important only where two chiral entities (usually molecules) interact. CD depends only on the asymmetry of the molecular system and so is uniquely sensitive to this aspect of it. As many of the questions we ask about chemical and

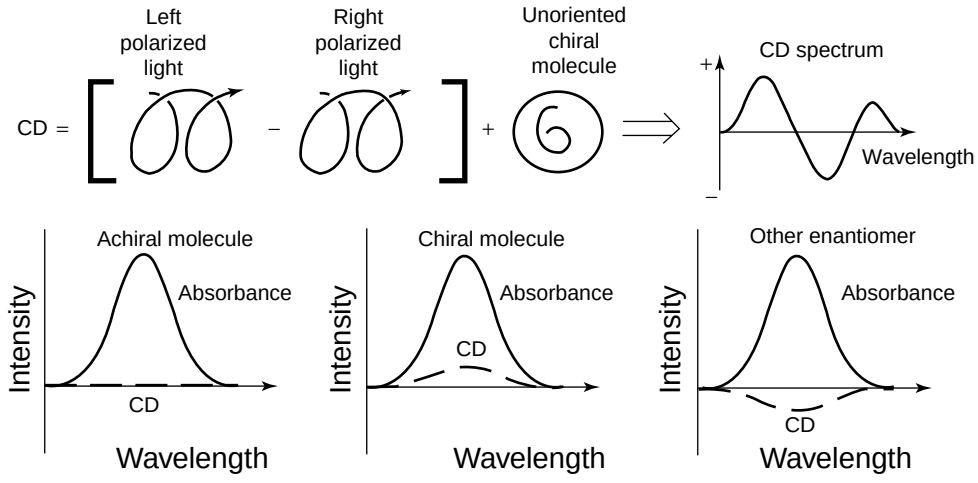

Figure 1 Schematic illustration of CD.

biological systems relate to the shapes of molecules and how this affects their properties, CD is a very useful tool for studying chiral molecules.

CD is now a routine tool in many laboratories. The most common applications are determining whether a chiral molecule has been synthesized or resolved into pure enantiomers and probing the structure of biological macromolecules, in particular determining the α-helical content of proteins.

1.1 Circularly Polarized Light

Although in principle CD can be measured for any kind of electromagnetic radiation absorbed by a sample, in practice experiments are usually limited to the visible or the easily accessible ultraviolet (UV) wavelength range (~170–800 nm). We often loosely refer to this electromagnetic radiation as "light". Vibrational circular dichroism (VCD) using infrared radiation is also used, but is technically significantly more challenging.[2] We shall concentrate on the ultraviolet/visible (UV/VIS) CD as this instrumentation is more readily available and significantly more widely used than VCD.

Electromagnetic radiation, as its name implies, has an electric and a magnetic field that oscillate at right angles to one another and to the propagation direction (Figure 2).[3–7] Light may be described by a transverse wave whose *polarization* is defined by the direction of its electric field, **E**. Electromagnetic radiation also has particle character: if we do an experiment that would probe wave-like behavior, it is apparent that light has it, and if we look for particle character we find that too. Thus we refer to the *wavelength*, λ, and *frequency*, ν, of radiation, but acknowledge that if a molecule absorbs energy it absorbs discrete units called *photons* whose energy are given by Planck's relationship,

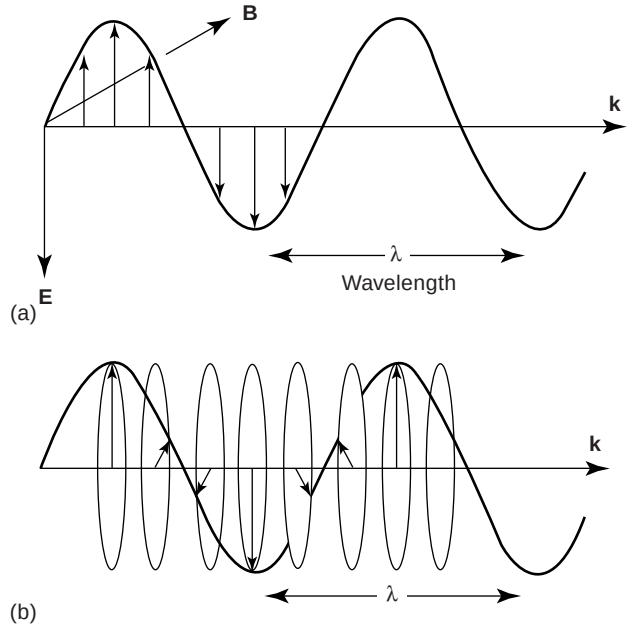

Figure 2 Schematic illustration of electromagnetic radiation illustrating the electric field, **E**, the magnetic field, **B**, and the propagation direction **k**. (a) Linearly polarized and (b) right circularly polarized electromagnetic radiation. Arrows denote direction of **E**.

Equation (2)[3–7]

$$E = h\nu = \frac{hw}{2\pi} = \hbar w \tag{2}$$

where h is Planck's constant and w is known as the angular frequency. It follows from the wave nature of light that a light beam can be linearly polarized with all photons having their electric field oscillating in the same plane (Figure 2). In a circularly polarized light beam the electric field vector retains constant magnitude

in time but traces out a helix about the propagation direction. Following the optics convention[6,8] we take the end of the electric field vector of right circularly polarized light to form a right-handed helix in space at any instant of time. At each point in space or time, the magnetic field, **B**, is perpendicular to the electric field such that **k**, the propagation direction, **E**, and **B** form a right-handed system as illustrated in Figure 2.

If we take X to be the direction of propagation of the light, then unit vectors describing the polarization (by convention the direction of the electric field component of the light) of two perpendicularly polarized light beams may be written, Equation (3)

$$\hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_Y = (0, 1, 0) \quad \hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_Z = (0, 0, 1) \tag{3}$$

If two equal magnitude beams are combined so that one lags $\pi/2$ or a quarter of a wavelength behind the other, then the electric field vector will trace out a circle giving circularly polarized light. The electric field polarization vector for right circularly polarized light (as defined above) may be written:[6–8]

$$\hat{\mathbf{e}}_\mathrm{r} = \frac{1}{\sqrt{2}} \left\{ (0, 1, -\mathrm{i}) \exp\left(\frac{2\pi \mathrm{i} X}{\lambda} - \mathrm{i}wt \right) \right\} \tag{4}$$

where i is the square root of -1. The oscillation of the electric field vector in time is described by the factor $\exp(-\mathrm{i}wt)$ where w is the angular frequency ($2\pi\nu$) and t is time; the magnitude of the electric field vector remains constant in time, but its direction rotates about X with frequency w. By taking the real part of Equation (4) we see that for an observer sitting at a fixed point, say $X = 0$, the electric field vector of right circularly polarized light rotates in time about X in a clockwise manner. At a fixed time, say $t = 0$, it forms a right-handed helix in space. In what follows we use the polarization vectors:

$$\hat{\mathbf{e}}_\mathrm{l} = \frac{1}{\sqrt{2}}(0, 1, \mathrm{i}) \quad \hat{\mathbf{e}}_\mathrm{r} = \frac{1}{\sqrt{2}}(0, 1, -\mathrm{i}) \tag{5}$$

for, respectively, left and right circularly polarized light, Equation (5).

1.2 Normal Absorption Spectroscopy

Most spectroscopic phenomena arise from the interaction of a molecule with one photon at a time. The molecule either absorbs, emits, or scatters the photon. The simplest form of spectroscopy is absorption, where we measure how much light of a given frequency is absorbed by a collection of molecules. If a molecule absorbs a photon of frequency ν, it increases its energy according to Planck's relationship (Equation 2). Absorbance is defined in terms of the intensity of incident, I_0, and transmitted, I, light, Equation (6):[3–5]

$$A = \log_{10}\left(\frac{I_0}{I} \right. \tag{6}$$

The Beer–Lambert law for the absorption, A, of light by a sample of concentration C is given by Equation (7)[3–5]

$$A = \varepsilon C l \tag{7}$$

where l is the length of the sample through which the light passes and ε is known as the extinction coefficient; if l is measured in cm and C in M (=mol dm^{-3}), then ε has units of mol^{-1} dm^3 cm^{-1}. The Beer–Lambert law is valid as long as the spectrometer can measure I (i.e. the concentration is not so large that essentially all photons are absorbed) and there are no concentration-dependent intermolecular interactions.

In practice, in a collection of molecules, the photons absorbed by different molecules will be of slightly different energies so what we measure is a curve like the absorbance curves of Figure 1, where the signal that is plotted is a measure of the probability that a transition will occur at that energy. Such a plot of the absorbance of light verses λ or ν is known as an *absorption spectrum*.

1.3 Circular Dichroism Spectroscopy

When a molecule absorbs UV and visible light its electronic distribution is altered. Now consider chiral molecules. Because a chiral molecule has no reflection plane, any rearrangement of its electrons will not have one either, so the electrons move in some kind of helix. As the electric field vectors also trace out helices (Figure 1) in circularly polarized light, the interaction between a chiral molecule and left- and right-handed photons will be different. This is the idea behind the definition of CD given in Equation (1). The CD version of the Beer–Lambert law (Equation 7) is Equation (8):

$$\Delta A = (\varepsilon_\mathrm{l} - \varepsilon_\mathrm{r}) C l = (\Delta\varepsilon) C l \tag{8}$$

Thus, a CD spectrum looks similar to an absorbance spectrum in that it occurs at the same wavelengths (Figure 1), but it can be either positive or negative in sign and if more than one transition lie close together, cancellation of intensity may occur (Figure 3). The challenging task is then to relate the helical motions of the electrons to the arrangement of the atoms and bonds in space.

As noted above, a CD spectrum is only expected for solutions of chiral molecules, though achiral molecules oriented skew to a light beam in an oriented sample, such

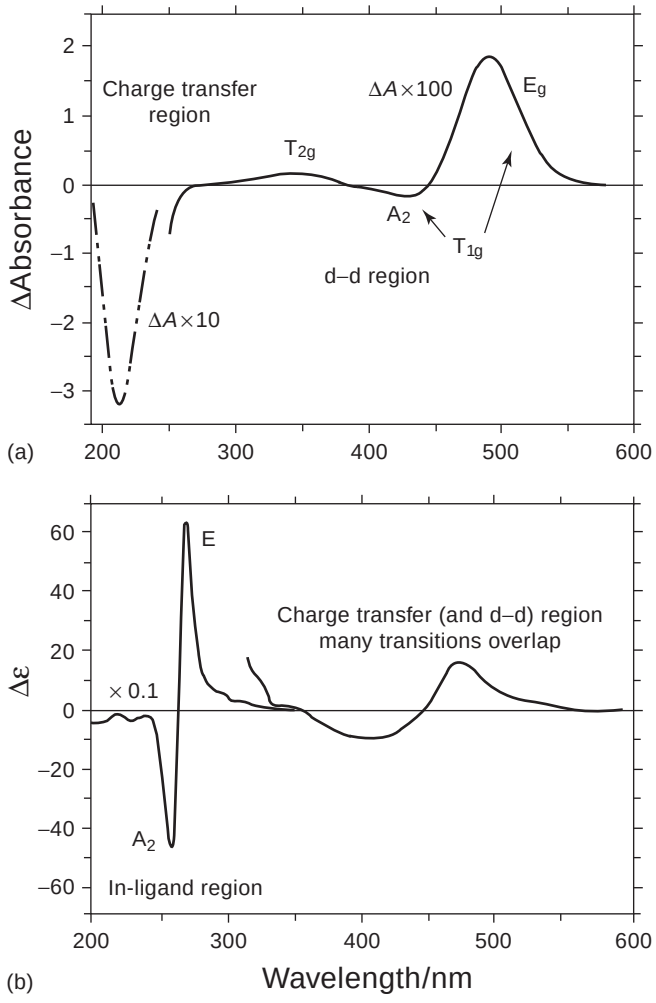

Figure 3 CD spectra of 0.1 M Λ-$[Co(ethylenediamine)_3]^{3+}$ in a 0.1 cm path length cell and $[Ru(1,10\text{-phenanthroline})_3]^{2+}$. Note cancellation of close lying bands.

as a crystal, will give a signal. The other requirement is that a molecule gives rise to an absorbance signal in the region of interest. If a molecule has no UV/VIS absorbance spectroscopy then no CD signal will be found. Sometimes, as with chiral HPLC (high-performance liquid chromatography) (see below) it is desirable to cause a nonabsorbing molecule to have a CD signal. This may be achieved by derivatizing the analyte molecule with a UV/VIS active chromophore, though care must be taken with the chemistry to ensure inversion of configuration or racemization does not occur.

1.4 Units of Circular Dichroism Spectroscopy

Most CD spectropolarimeters, although they measure differential absorbance, for historical reasons produce a CD spectrum in units of ellipticity, θ, in millidegrees, versus λ, rather than ΔA versus λ. The conversion between these two is given in Equation (9)[6,8]

$$\begin{aligned} \text{CD} = \Delta A(\text{absorbance units}) &= A_\text{l} - A_\text{r} \\ &= \frac{4\pi\theta(\text{degrees})}{180 \ln 10} = \frac{\theta(\text{millidegrees})}{32\,982} \end{aligned} \tag{9}$$

The CD analog of the Beer–Lambert law (Equation 8) then allows conversion to molar absorbance or molar ellipticity. In this context it is important to be clear what concentration is being measured. For example DNA concentrations are often given in terms of bases or phosphates, but may be in terms of base pairs, whereas protein concentrations are commonly in terms of amino acid residues and occasionally whole molecules.

1.5 Optical Activity (Optical Rotation)

Optical rotation (OR) is the difference in refractive indices of left and right circularly polarized light upon passing through the medium, Equation (10):[6]

$$\text{OR} = (n_\text{l} - n_\text{r})\frac{\pi d}{\lambda} \tag{10}$$

the units are radians if λ (wavelength) and d (pathlength) have the same units (usually dm are used). $(n_\text{l} - n_\text{r})$ is called the circular birefringence.

OR is usually measured by determining the rotation of linearly polarized light upon passing through the solution. If the linearly polarized light is rotated clockwise when viewed into the light source the OR is called a positive or right (dextro) OR. OR as a function of λ is called optical rotatory dispersion (ORD). ORD gives an S-shaped curve centered at the CD maximum (the so-called point of anomalous ORD since if a transition has zero band width the ORD would go to plus and minus infinity either side of λ_{max}). The ORD is also nonzero away from an absorption band, hence α_D values (the ORD at the sodium D line, 589 nm) may be used to characterize the enantiomeric excess of a solution (see below). The long-wavelength side of the ORD curve shows a positive ORD contribution for a positive CD band, but a negative ORD for a negative CD band. ORD is sometimes referred to as optical activity.

Since ORD and CD may be written respectively in terms of the dispersive and absorptive parts of the optical activity tensors, with all other factors the same, Kramers–Kronig[6] relations may be used to convert one to the other *if* the complete ORD spectrum or complete CD spectrum is available. The formalism for this process is fairly straightforward; however, in practice the problem is that the ''complete'' spectrum is not available. It is often possible to perform the conversion on a more limited wavelength range with an element of error. Except for single wavelength measurements to quantify enantiomeric excess, ORD is seldom used now as the more direct

relationship between absorbance spectroscopy and CD facilitates the interpretation of data in structural terms.

1.6 Enantiomeric Purity and Enantiomeric Excess

If a CD signal is observed for a solution, then there is a net excess of a chiral analyte; however, unless $\Delta\varepsilon$ for the compound is known, the enantiomeric purity of the sample cannot be determined. Alternative methods that measure the relative amounts of the two enantiomers directly are chiral HPLC[9] and NMR (nuclear magnetic resonance) with chiral shift reagents.[10] Simply choosing a chiral column or mobile phase additive or chiral shift reagent and then performing an experiment which gives only one peak does not prove that the solution is enantiomerically pure. It must be possible to show that if the mixture is racemic (or any percentage mixture of the two enantiomers) that the experimental method does indeed discriminate between the two species.

Enantiomeric purity is often given in terms of the percentage of one enantiomer present. Alternatively, we use enantiomeric excess. Enantiomeric excess relates more directly to the CD signal because if there is no enantiomeric excess then there is no CD signal; conversely the CD observed follows Equation (8) if the solution is enantiomerically pure. Enantiomeric excess is usually defined as the apparent analyte concentration determined from Equation (8) divided by the total analyte concentration, often converted to a percentage, Equation (11)

$$\mathrm{EE}\% = 100 \times \frac{(\text{Concentration of enantiomer 1}) - (\text{Concentration of enantiomer 2})}{(\text{Concentration of enantiomer 1}) + (\text{Concentration of enantiomer 2})} \tag{11}$$

2 STRUCTURE ANALYSIS FROM CIRCULAR DICHROISM

2.1 Introduction

The most obvious result to be deduced from the appearance of a CD spectrum is whether or not there is net asymmetry in the sample. It is not much of an overgeneralization to say that the only CD studies that have been interpreted beyond this stage are those where the transition being studied is essentially located in an achiral chromophore (or electronic subunit) of the system. Explicit calculations can, of course, be performed for intrinsically chiral chromophores (such as CHFClBr), however, this does not usually provide any information except that the input geometry for that molecule is correct. By way of contrast, carbonyl compounds containing the achiral chromophore (circled in Figure 4) have been extensively studied using CD spectroscopy and various structure–spectra correlation rules have been proposed and applied. It should be noted that any structure–spectrum correlation rule can only be applied with full confidence once its mechanistic origin is understood.

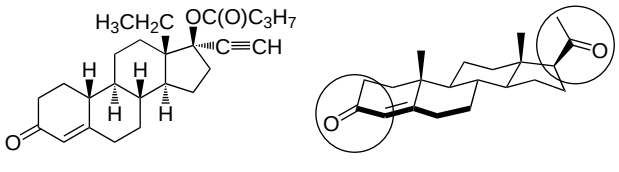

Figure 4 Progesterone with the carbonyl chromophores circled. Note progesterone has one carbonyl conjugated with a double bond so it is not strictly electronically independent of the rest of the molecule.

Given that the chromophores we are interested in are achiral, the electron redistribution that occurs during a transition occurs in a plane. This means that either the magnetic dipole transition moment, $\boldsymbol{m}$, (MDTM) (by which the transition interacts with the magnetic field of the radiation) or the electric dipole transition moment, $\boldsymbol{\mu}$, (EDTM) (by which the transition interacts with the electric field of the radiation) vanishes (or they are perpendicular to one another). The $\boldsymbol{m} = 0$, $\boldsymbol{\mu} \neq 0$ transitions are called electric dipole allowed (EDA), magnetic dipole forbidden (MDF). The $\boldsymbol{m} \neq 0$, $\boldsymbol{\mu} = 0$ transitions are called electric dipole forbidden (EDF), magnetic dipole allowed (MDA). The theoretical significance of this division is due to the way CD intensity is induced into the two different types of transitions. EDA transitions require the induction of a magnetic component, whereas MDA transitions require an induced electric component. The dependence of the different kinds of induced moments on the geometry of the system is very different, so when we wish to extract geometric information from CD we must be aware of which situation is in effect. The CD spectrum induced into EDA transitions is also different if it arises from the coupling of identical chromophores rather than from the coupling of nonidentical chromophores.

Experimentally the difference between EDA/MDF and EDF/MDA transitions is that the latter have low absorbance intensity (what they have is usually gained from vibronic coupling rather than from the MDTM) so the ratio of their CD signal to their absorbance is larger than that for EDA/MDF transitions. The dissymmetry factor, g, Equation (12)[6]

$$g = \frac{4R}{cD} \tag{12}$$

where R is the rotational CD strength (Equation 30), c is the speed of light, and D is the dipole strength (square of the EDTM) of the transition summarizes this. The

main historical consequence of this is that most of the early CD experiments were performed on EDF/MDA transitions such as the $n \rightarrow \pi^*$ carbonyl transitions and the T_{1g} tris-chelate transitions.

2.2 Qualitative Description of Circular Dichroism and Its Correlation with Structure for Some Systems

2.2.1 Carbonyl Circular Dichroism

Carbonyl chromophores, i.e. carbonyl functional groups that are not conjugated with any other double bond of the molecule, occur in many organic molecules of both natural and synthetic origin including steroids. That a simple reliable empirical rule relating carbonyl molecular geometry to observed CD spectrum was determined and to some extent rationalized in the 1960s was therefore extremely convenient. The lowest lying transition of a carbonyl is the $n \rightarrow \pi^*$ transition which involves rotation of electron density from the nonbonding oxygen lone pair into the π^* orbital of the C=O bond (Figures 4 and 5). The transition, which occurs somewhere between 350 nm and 240 nm, is weak with an extinction coefficient typically of 20–50 $mol^{-1}\,dm^3\,cm^{-1}$. When it is conjugated with a double bond as in progesterone (Figure 4) the absorbance intensity is increased and any rules to interpret the CD spectra must be applied with care. The octant rule of Moffitt et al.[11] is the unifying feature for much of the carbonyl CD work. In its original form, the rule stated that the CD of the $n \rightarrow \pi^*$ transition depends on the positions of other parts of the molecule relative to the carbonyl group according to Equation (13)

$$-\frac{xyz}{r^n} \tag{13}$$

where (x,y,z) is the unit vector along the lines of the position vector in the right-handed coordinate system of Figure 5 and r is the distance from the centre of the carbonyl bond to the chromophore. Höhn and Weigang[12] used perturbation theory to give a theoretical justification of the octant rule and found that $n = 4$, and that the relative importance of different parts of the molecule scaled with their polarizability. This was the first use of what has come to be called the independent systems/perturbation (ISP) approach.[13–15]

To apply the octant rule to hydrocarbon systems, one simply determines the cartesian coordinates of each carbon atom (assuming the hydrogens make little contribution), evaluates Equation (13) with $n = 4$ and performs a sum over all carbons in the molecule (or in practice over the net perturbers since atoms located symmetrically about a plane of the octant system will give equal magnitude and opposite sign contributions). An example is given in Figure 6. A simple Gaussian type curve with sign and relative magnitude determined by this sum should then correlate with the observed CD. For nonhydrocarbon systems, the sum must be weighted by the polarizabilities of the atoms or chromophores (see below) of the molecule.[16]

Although the octant rule has proved to be highly successful, there are sufficiently numerous exceptions to provide a warning against the blind use of empirical rules. Empirical modifications (see Rodger and Rodger and references therein),[15] while helpful for a given series of compounds, do not provide a global approach. Consideration of all terms in the ISP expansion (see below) is really required. The conclusion from this

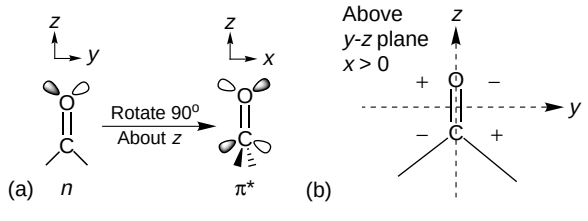

Figure 5 (a) Schematic illustration of the carbonyl $n \rightarrow \pi^*$ transition illustrating the (x,y,z) coordinate system used for the octant rule. (b) Octant rule CD signs. Invert the signs illustrated for the $x < 0$ region of space.

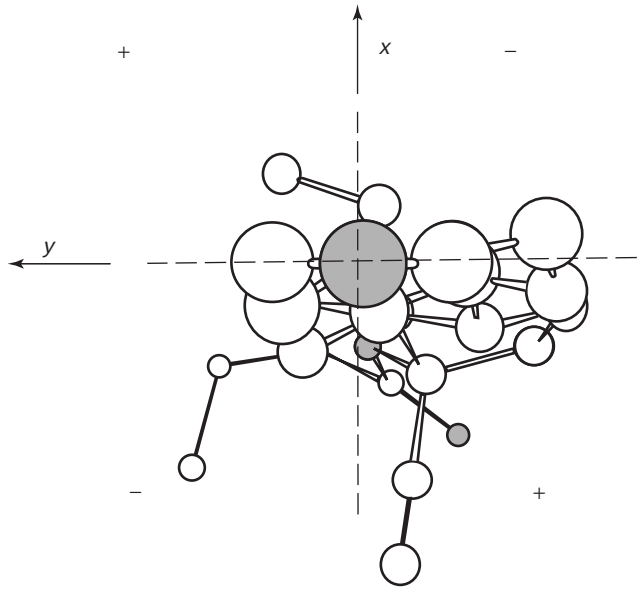

Figure 6 Projection onto the x–y octant plane of levonorgestrel butanoate viewed along the carbonyl bond. All atoms (except the carbonyl oxygen) have coordinates with $z < 0$. Octant signs are as indicated. The size of the carbon atoms indicates their proximity to the carbonyl group (large ones are close). The dominant octant rule net perturber lies in the $x < 0, y > 0, z < 0$ octant (although more distant atoms lie in the $x < 0, y < 0, z < 0$ octant) Oxygen atoms are shaded. Thus, a positive $n \rightarrow \pi^*$ CD signal is expected from this molecule. In fact $\Delta\varepsilon_{322\,nm} = 1.5\,mol^{-1}\,dm^3\,cm^{-1}$.[45]

analysis[15] is that the octant rule is not valid for polar and charged groups. It also requires the net perturbing part of the system not to lie on or near one of the octant planes. Further, no solvent cavities must be created by substituents on the molecules and planar zig-zags of bonds with the carbonyl bond must be absent for the CD magnitude to be given by the octant rule.[18]

2.2.2 Transition Metal Complexes

The spectroscopy of transition metal complexes can be subdivided into three categories: in-ligand transitions, which resemble those of the isolated ligands; charge-transfer transitions between the ligand and the metal; and d–d transitions of the chromophore defined by the central metal ion and its directly ligating atoms.

The in-ligand transitions need only be considered for aromatic ligands since they are the only ones for which the CD can be easily measured. This effectively restricts consideration to tris-chelate (and also *cis* bis-chelate) complexes with planar aromatic ligands. The ligands in such molecules are usually of C_{2v} symmetry (Figure 7). The in-ligand CD for these systems can usually be accurately predicted by the Kirkwood–Kuhn coupled oscillator mechanism (see below). As the complexes are based on an octahedral template but themselves have D_3 symmetry, the transitions all occur as couplets of A_2 (z, the three-fold axis, Figure 7) polarized and E_u (x,y) polarized transitions. The coupling of long axis ligand transition moments gives a bisignate CD signal the lower energy component being the out-of-phase E_u component that has a negative sign for the Δ enantiomer[19] and conversely for the Λ enantiomer (see below for the theoretical basis for this conclusion). The A_2 component (in the absence of coupling with other transitions) is equal in magnitude and opposite in sign from the E_u component as illustrated in Figure 3.

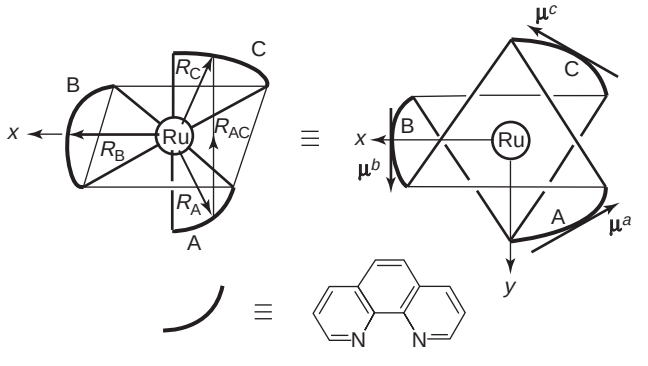

Figure 7 $[Ru(1,10\text{-phenanthroline})_3]^{2+}$, a tris-chelate transition metal complex of D_3 symmetry with C_{2v} symmetry planar aromatic ligands. The three-fold (z) axis is out of the plane of the page in the right-hand side of the figure.

Charge transfer CD has also been measured mainly for tris-chelate complexes with planar aromatic ligands. For such molecules the charge transfer bands often obscure the d–d transitions. The lowest energy EDA charge transfer transitions for tris-chelate complexes gives a bisignate signal (E_u/A_2) with the E_u component having the opposite sign from the lowest energy long-axis in-ligand E_u CD (see above). The sign difference can be used to assign the polarization of the charge transfer bands[20] whose energy ordering varies from molecule to molecule, in contrast to the situation for the in-ligand transitions.

Interpreting the CD for the MDA d–d transitions (CD for MDF/EDF is usually very small) is complicated by the fact that the dominant mechanism is determined by the symmetry of the d–d chromophore. One reliable empirical rule has been established over the years for tris-chelate complexes where transitions of E and A_2 polarization occur as close lying couplets: the E band CD signal is larger than the A_2 band signal.[19] The underlying reason for this is that if we take the d–d chromophore to have D_{3d} symmetry (Figure 8) then the E_g band gains its CD intensity from a r^{-4} quadrupolar mechanism (see below), whereas the A_2 band depends on an r^{-6} hexadecapolar mechanism as discussed below. In reality the d–d chromophore of tris-chelate complexes is not achiral but twisted. However, for solution phase work the approximation is satisfactory as the close energy of the A_2 and E components of any band means that any intrinsic CD cancels as it would for the octahedral parent geometry. It is only for crystal CD that the problem is noted.[21]

2.2.3 Macromolecules

The CD of macromolecules[22] differs from that of small molecules in that the spectroscopy of the component

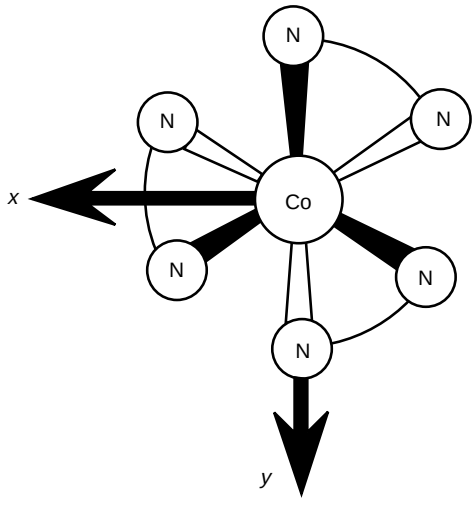

Figure 8 D_{3d} d–d chromophore of $[Co(ethylenediamine)_3]^{3+}$, ethylenediamine = $NH_2CH_2CH_2NH_2$.

parts, such as the amino acids, at least in principle, is usually understood and the form of the CD spectra result largely from the arrangement of the units in space. The CD of macromolecules is most commonly used in one of two ways: (i) to probe changes in the conformation of the macromolecule itself, and (ii) to probe its interaction with small molecules bound to it.

2.2.4 Proteins

At present the main use of CD in the study of proteins is as an empirical guage of protein structure and conformation. Proteins are long chains of amino acids, their primary structure being as illustrated in Figure 9.[23] If the side chain (R_i in Figure 9) is not H, then the ith tetrahedral carbon is a chiral center and we should expect an induced CD signal in transitions of the neighboring amide groups and side chains. If, however, there is free rotation about the bonds of the main chain, the observed amide $n \rightarrow \pi^*$ (occurring at ~220 nm), $\pi \rightarrow \pi^*$ (occurring at 207–190 nm), and $n \rightarrow \sigma^*$ (occurring as a shoulder at 175 nm) CD signals will be small since they will be an average of different geometry factors (see carbonyl CD above and coupled oscillator CD below). Thus a truly random or denatured protein will have only a small CD signal at energies accessible to most CD machines. The most stable conformation of a protein under physiological conditions, however, is not random but composed of well-defined structural units that give macrochiral units with significant CD intensities.

The basis of CD structural analysis of proteins is that the CD spectrum from 250 nm downwards is due largely to transitions of the main backbone chain, so it is independent of the nature of the side chains (this is not always true if there are a large number of well-ordered aromatic side chains or prosthetic groups such as a heme present). In any well-defined structural motif (such as the α-helix where the nth peptide unit forms hydrogen bonds between its C–O and the N–H of the $(n + 4)$th peptide and between its N–H and the $(n - 4)$th C–O; there is a 1.5 Å translation and 100° turn between two consecutive peptide units, giving 3.6 amino acid residues per turn), the chromophores giving rise to the CD are in well-defined positions relative to one another so have a well-defined CD spectrum. The net CD is then the sum of that due to the independent structural units. A wide range of fitting programs for protein CD data are available.[24–26] Some work in terms of the spectra of a set of identified structural

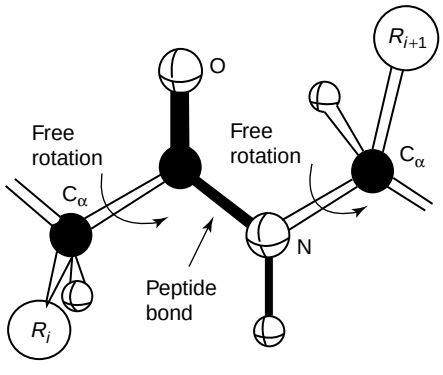

Figure 9 Protein primary structure. R_i denotes the side chain of the ith amino acid.

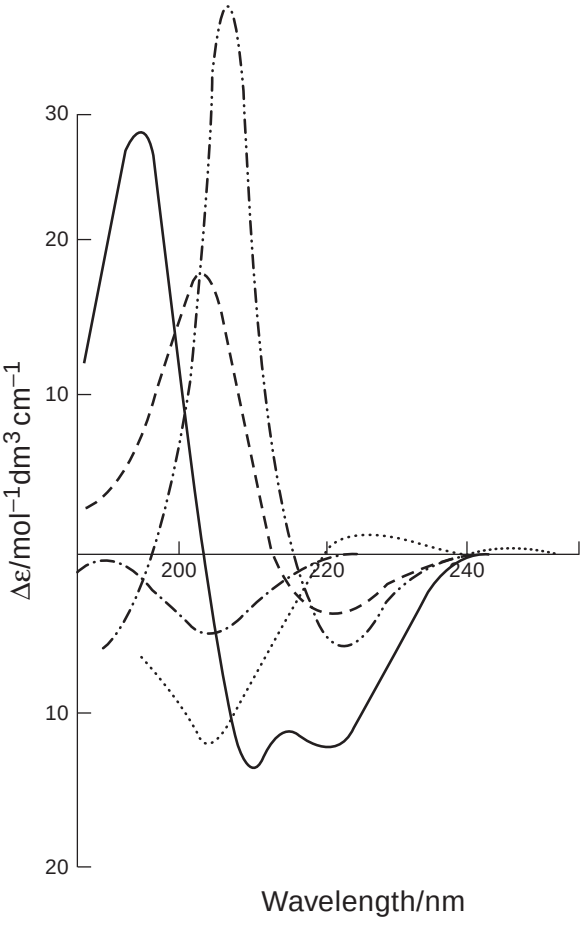

Figure 10 Typical standard protein motif CD spectra,[26,43] α-helix (—, average of poly(γ-methyl-L-glutamate) in F_6i-PrOH and poly(L-alanine) as a film), antiparallel β-sheet (- - - -, *tert*-butyloxycarbonyl (BOC) (L-alanine)$_7$OMe as a film), parallel β-sheet (– • • –, BOC (L-valine)$_7$OMe as a film), poly(L-proline) II-type helix (• • •, systemin in 10 mM phosphate buffer at 5 °C, pH = 7[43]), and random coil (– • –, collagen at 45 °C in 0.01 M phosphate buffer, pH = 3.5). The CD characteristics of α-helices are a negative signal with separate maxima of similar magnitude at 222 nm (the $n \rightarrow \pi^*$ transition) and 208 nm (part of the $\pi \rightarrow \pi^*$ transition that is polarized parallel to the helix axis); and a positive signal at ~195 nm (part of the $\pi \rightarrow \pi^*$ transition that is polarized perpendicular to the helix axis). The CD characteristics of β-sheets are a negative band at about 220 nm and a positive band of near 200 nm and the CD characteristics of the poly(L-proline) II-type helix are a positive band at ~225 nm and a negative band at about 200 nm.

motifs, others use a basis set of real protein spectra where the percentage of different structural units is known. The approach of Manalavan and Johnson,[25] where a large basis set of such proteins is used and a subset of these randomly chosen until a good fit is obtained probably gives the best estimate of the full range of structural motifs present in proteins. Almost all approaches will give a good α-helix content estimate as its profile is the most distinct and also largest in magnitude as illustrated in Figure 10.

2.2.5 DNA

CD experiments performed on DNAs generally fall into one of two categories: those used to probe the DNA and those used to probe a drug (or other molecule) bound to the DNA. DNA can be viewed as a more-or-less vertical spiral staircase where the steps are made up of pairs of planar aromatic molecules, called base pairs, and the sides are chains of alternating phosphate groups and ribose sugars.[21] The source of the chiral structure in DNA is the sugar backbone which imposes the twisted structure onto the whole molecule. Usually the twist is right-handed, though it appears that left-handed Z-DNA is intrinsically not much less stable than the right-handed forms.[27] The CD of a DNA molecule in the 180–300 nm region is due to the skewed orientation of transitions in the bases. Thus, if the DNA is untwisted there will be less CD and if it is twisted the other way the relative geometries of the coupling transition moments (see below) will change and the CD signal may even invert (as it does for Z-DNA relative to much of the B-DNA spectrum, Figure 11).

2.2.6 Induced Circular Dichroism of a Molecule Bound to DNA

If an achiral molecule is bound to DNA, its transitions gain an induced circular dichroism (ICD) due to perturbation by the chiral DNA. The sign and magnitude of this ICD is crucially dependent on the binding geometry, so in principle should yield information about that binding geometry. At its crudest level the appearance of the CD signal indicates that the drug molecule is bound. If the binding is in a single binding mode or a constant proportion of mixed modes then the ICD signal will be proportional to the concentration of bound drug and a titration series of data can be used to determine the binding constant of the drug to DNA.[8,28] An example is given in Table 1 and Figure 12.

2.3 Interaction of Radiation with Matter and the Rosenfeld Equation for Circular Dichroism

In classical mechanics, the energy of the interaction, H_{int}, between a molecule and the electric, **E**, and magnetic, **B**, fields of electromagnetic radiation is given

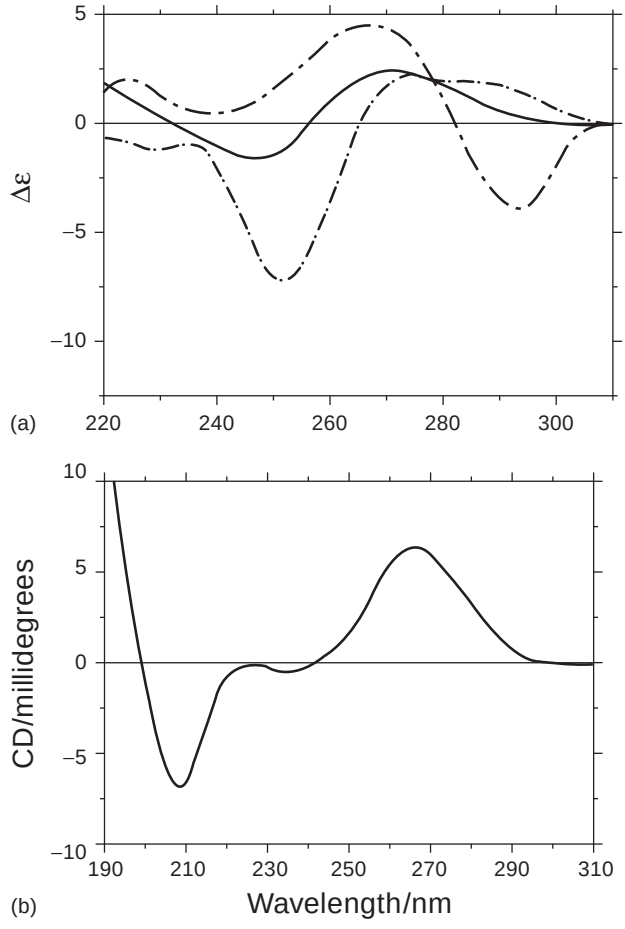

Figure 11 CD spectra of (a) —, B-form calf thymus DNA (pH = 6.8, 5 mM NaCl, 1 mM sodium cacodylate); – • – •, B-form poly[d(G–C)]$_2$ (pH = 6.8, 5 mM NaCl, 1 mM sodium cacodylate), and - — -, Z-form poly[d(G–C)]$_2$ (pH = 6.8, 5 mM NaCl, 1 mM sodium cacodylate, 50 μM $[Co(NH_3)_6]^{6+}$) (b) 372 nucleotide A-form mRNA (pH = 7).[44] The CD characteristics of these nucleic acid polymorphs are as follows. B-form: a positive band centered at 275 nm, a negative band at 240 nm, with the zero being around 258 nm. At 220 nm the CD signal either becomes positive or less negative; a small negative peak is then followed by a large positive peak from 180–190 nm. A-form: a positive CD band centered at approximately 260 nm that is larger than the corresponding B-form band, a fairly intense sharp negative band at 210 nm and a very intense positive band at 190 nm. The 250–230 nm region is also usually fairly flat though not necessarily zero. Z-form: a negative CD band at 290 nm, a positive band at approximately 260 nm, a large negative CD signal in the 195–200 nm region, and the CD passes through zero between 180 and 185 nm.

by Equation (14)[3–5,7]

$$H_{\text{int}} = -\boldsymbol{\mu}\mathbf{E} - \boldsymbol{m}\mathbf{B} + \text{higher order multipole terms} \quad (14)$$

where $\boldsymbol{\mu}$ is the electric dipole moment of the molecule and $\boldsymbol{m}$ is its magnetic dipole moment. $\boldsymbol{\mu}$ and $\boldsymbol{m}$ are determined by summing over the electron distribution of the molecule, Equation (15):[3–5,7]

Table 1 Induced CD signals at 215 nm (DNA signal subtracted out) determined for calf thymus DNA with additions of $[Co(NH_3)_6]^{3+}$

Total $[[Co(NH_3)_6]^{3+}](\mu M)$	$ICD_{215} = \rho$ (mdeg)	Concentration of bound ligand, $L_b(\mu M)$
15	1.2	11.4
30	2.13	19.8
50	2.98	27.7
100	3.74	34.7

Path length 1 cm, 100 mM in 10 mM NaCl, 1 mM sodium cacodylate, pH = 7.

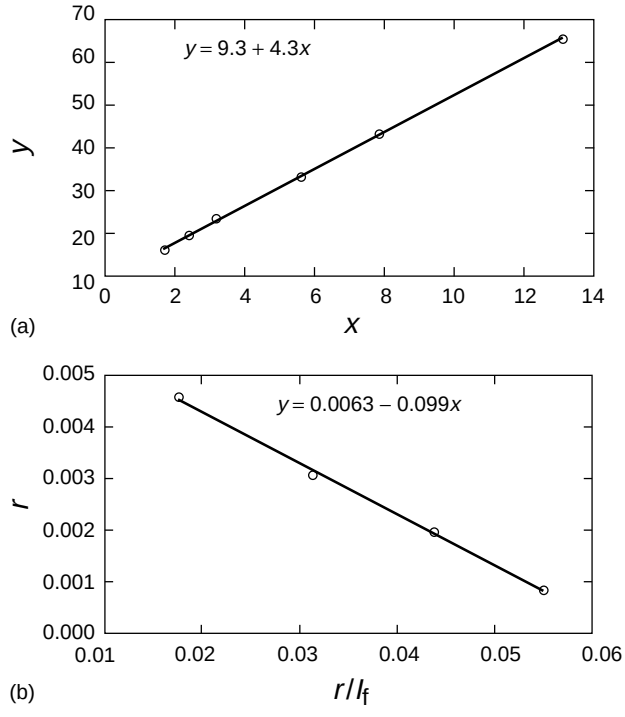

Figure 12 (a) Induced spectroscopy method plot[8,28] where $x = \left(\frac{L^k_{tot}}{\rho^k} - \frac{L^j_{tot}}{\rho^j}\right)$ $(\rho^k - \rho^j)$ and $y = L^k_{tot} - L^j_{tot}/(\rho^k - \rho^j)$ for ρ^i the ICD signal for the ith spectrum, and L^i_{tot} the ligand concentration for that spectrum. The intercept equals α, where α relates the bound ligand concentration to the ICD ρ: $L_b = \alpha\rho$, and n is the number of DNA bases per bound ligand. The slope is S_{tot}/α where S_{tot} is the total binding site concentration (i.e. the DNA concentration divided by n). From the graph: $\alpha = 9.3 \pm 1\,\mu M\,mdeg^{-1}$; $n = 16 \pm 2$. (b) Scatchard plot[8,28] using value of α determined from (a). r is binding ratio (concentration of ligand bound divided by DNA base concentration), I_f is the concentration of free ligand. The slope is $-K$, for K the equilibrium binding constant. The binding constant is therefore approximately $100\,000\,M^{-1}$.

$$\boldsymbol{\mu} = e\sum_i (\boldsymbol{r}_i)$$

$$\boldsymbol{m} = \frac{e}{2m_e}\sum_i \{\boldsymbol{r}_i \boldsymbol{p}_i\} \qquad (15)$$

where e is the (negative) unit charge on an electron, $\boldsymbol{r}_i$ denotes the position vector of the ith electron, $\boldsymbol{p}_i$ is the momentum vector for the ith electron, and m_e is the mass of an electron.

When we are interested in spectroscopic phenomena such as absorbance, CD, and LD (see below), we have to move beyond classical mechanics to quantum mechanics since the transitions we study involve discrete amounts of energy, i.e. the energy is quantized. Thus, instead of an expression for the classical energy of the interaction between the radiation and the molecule, we need the interaction Hamiltonian operator.[3–5,7] This follows from Equation (14) by converting $\boldsymbol{\mu}$ and $\boldsymbol{m}$ to operators and the interaction energy to the Hamiltonian operator. Curved 'hats' are used in what follows to denote operators, Equation (16):

$$\hat{H}_{int} = -\hat{\boldsymbol{\mu}}\mathbf{E} - \hat{\boldsymbol{m}}\mathbf{B} + \text{higher order terms} \qquad (16)$$

When we compare the relative strengths of different transitions experimentally we measure the intensity of the transition which is the area under the absorption band. The related theoretical quantity is the probability per unit time, P, that a transition will occur from the initial state, $|i\rangle$, (which is usually the ground state) to the final state, $|f\rangle$:[3–5,7]

$$\begin{aligned} P(|i\rangle \longrightarrow |f\rangle) &= k'|\langle f|\hat{H}_{int}|i\rangle|^2 \\ &= k'\langle f|\hat{H}_{int}|i\rangle^*\langle f|\hat{H}_{int}|i\rangle \\ &= k'\langle i|\hat{H}_{int}|f\rangle\langle f|\hat{H}_{int}|i\rangle \end{aligned} \qquad (17)$$

where the integral is over molecular position coordinates, and k' is a constant. Equation (17) is a form of the Fermi Golden Rule.[3–5,7] It is written in terms of bra–ket notation where the 'bra' $\langle 0|$ state is the complex conjugate of the 'ket' $|0\rangle$. When a bra and a ket are coupled together either directly or either side of an operator, such as $\hat{\boldsymbol{\mu}}$ in Equation (18) below, then we imply that an integral over all space is being performed. Round brackets $|\rangle$ in what follows denote unperturbed states of chromophores A and C (see below), and pointed ones $|\rangle$ denote the states of the whole system. Upon substituting the expression for $\hat{H}_{int}$ into Equation (17) we have Equation (18)

$$\begin{aligned} P(|i\rangle \longrightarrow |f\rangle) = k\{&\langle i|\hat{\boldsymbol{\mu}}|f\rangle\hat{\mathbf{e}}^* + \langle i|\hat{\boldsymbol{m}}|f\rangle\hat{\mathbf{b}}^*\} \\ &\times \{\langle f|\hat{\boldsymbol{\mu}}|i\rangle\hat{\mathbf{e}} + \langle f|\hat{\boldsymbol{m}}|i\rangle\hat{\mathbf{b}}\} \end{aligned} \qquad (18)$$

where $\hat{\mathbf{e}}$ is the unit vector along $\mathbf{E}$, $\hat{\mathbf{b}}$ is the unit vector along $\mathbf{B}$, the asterisk denotes the complex conjugate, $|i\rangle^* = \langle i|$, and the magnitudes of the electric and magnetic fields are absorbed into the constant k. $\langle f|\hat{\boldsymbol{\mu}}|i\rangle$ is the EDTM *from* state i *to* state f, and $\langle f|\hat{\boldsymbol{m}}|i\rangle$ the corresponding MDTM.

It follows that,[8] Equation (19)

$$P(|i\rangle \longrightarrow |f\rangle) = k[\boldsymbol{\mu}^{fi}\hat{\mathbf{e}}^*\boldsymbol{\mu}^{if}\hat{\mathbf{e}} + \boldsymbol{m}^{fi}\hat{\mathbf{b}}^*\boldsymbol{m}^{if}\hat{\mathbf{b}}]_{(\mathrm{I})} + k[\boldsymbol{\mu}^{if}\hat{\mathbf{e}}\boldsymbol{m}^{fi}\hat{\mathbf{b}}^* + \boldsymbol{\mu}^{fi}\hat{\mathbf{e}}^*\boldsymbol{m}^{if}\hat{\mathbf{b}}]_{(\mathrm{II})} \quad (19)$$

where $\boldsymbol{\mu}^{if} = \boldsymbol{\mu}^{i\to f} = \langle f|\hat{\boldsymbol{\mu}}|i\rangle$ and so on (note inversion of order of states in the notation).

2.3.1 *Interaction of a Molecule with Linearly Polarized Light: Absorbance and Linear Dichroism*

If no external magnetic field is present (except that due to the radiation) we can choose $|i\rangle$ and $|f\rangle$ to be real wavefunctions; thus $\boldsymbol{\mu}^{fi} = \boldsymbol{\mu}^{if}$ but $\boldsymbol{m}^{fi} = -\boldsymbol{m}^{if}$ since $\hat{\boldsymbol{m}}$ is imaginary. The polarization vectors for **E**, and hence **B**, are also real for linearly polarized light. It follows that the two parts of term (II) in Equation (19) are equal in magnitude but opposite in sign, so term II is zero. As magnetic terms are invariably small compared with electric ones, we can ignore the second term within term (I) and therefore write the probability of the transition occurring from $|i\rangle$ to $|f\rangle$ with light polarized parallel to the electric field (the absorbance, A_e, of the transition) to be:

$$A_\mathrm{e}(|i\rangle \longrightarrow |f\rangle) = k|\boldsymbol{\mu}^{if}\hat{\mathbf{e}}|^2 \quad (20)$$

For the usual situation where the system is a collection of randomly oriented molecules, the interaction averages to $\frac{1}{3}$ Equation (20) and becomes Equation (21):[3–5]

$$A_\mathrm{iso} = A = \frac{k}{3}|\boldsymbol{\mu}^{if}|^2 \quad (21)$$

where A_iso is the isotropic absorbance. LD, which forms the subject matter of the next section, is the difference in absorption of two in-phase linearly polarized light beams propagating in the same direction with perpendicular polarizations. Take the (‖) direction to be (0,0,1) and the perpendicular direction to be (0,1,0). Then Equation (22) follows from Equation (20)

$$\mathrm{LD}_\mathrm{molecule} = (A_\parallel - A_\perp)_\mathrm{molecule} = k\{|\mu_Z^{if}|^2 - |\mu_Y^{if}|^2\} \quad (22)$$

where μ_Y^{if} is the Y component of the $|i\rangle$ to $|f\rangle$ transition dipole moment, and so on. For a collection of N molecules we sum over the LD for each molecule, Equation (23)

$$\mathrm{LD}_\mathrm{total} = \sum_\mathrm{molecules}(A_\parallel - A_\perp)_\mathrm{molecule} = Nk\{\langle|\mu_Z^{if}|^2\rangle - \langle|\mu_Y^{if}|^2\rangle\} \quad (23)$$

where $\langle\rangle$ denotes average. The net LD thus depends on the method and extent of sample orientation since $\{X, Y, Z\}$ is a laboratory-fixed axis system about which the molecule knows nothing unless we tell it. In the absence of any molecular orientation the LD in Equation (23) vanishes because all molecular axes are equally likely to coincide with Z so, Equation (24)

$$\langle|\mu_Z^{if}|^2\rangle = \langle|\mu_Y^{if}|^2\rangle = \tfrac{1}{3}|\boldsymbol{\mu}^{if}|^2 \quad (24)$$

2.3.2 *Interaction of a Molecule with Circularly Polarized Light: Circular Dichroism*

When the radiation is circularly polarized, CD becomes possible. As long as there is no magnetic field (apart from that of the radiation) present, we can again choose to use real wavefunctions so $\boldsymbol{\mu}^{fi} = \boldsymbol{\mu}^{if}$ and $\boldsymbol{m}^{fi} = -\boldsymbol{m}^{if}$. To evaluate Equation (1) we substitute first the expression for left circularly polarized light into Equation (19) then subtract from it the result of substituting the expression for right circularly polarized light. We also make use of the equality $\hat{\mathbf{e}}_\mathrm{r} = \hat{\mathbf{e}}_\mathrm{l}^*$, and note that,[6–8] Equations (25) and (26)

$$\hat{\mathbf{b}} = \mathbf{k}\hat{\mathbf{e}} \quad (25)$$

so

$$\hat{\mathbf{b}}_\mathrm{l} = \tfrac{1}{\sqrt{2}}(1,0,0)\,(0,1,\mathrm{i}) = \tfrac{1}{\sqrt{2}}(0,-\mathrm{i},1) = \hat{\mathbf{b}}_\mathrm{r}^* \quad (26)$$

Thus[8]

$$\mathrm{CD} = A_\mathrm{l} - A_\mathrm{r} = k\underbrace{(\boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}^*\boldsymbol{\mu}^{if}\hat{\mathbf{e}}_\mathrm{l} + \boldsymbol{m}^{fi}\hat{\mathbf{b}}_\mathrm{l}^*\boldsymbol{m}^{if}\hat{\mathbf{b}}_\mathrm{l}) - (\boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}\boldsymbol{\mu}^{if}\hat{\mathbf{e}}_\mathrm{l}^* + \boldsymbol{m}^{fi}\hat{\mathbf{b}}_\mathrm{l}\boldsymbol{m}^{if}\hat{\mathbf{b}}_\mathrm{l}^*)}_{(\mathrm{I})} + k\underbrace{(\boldsymbol{\mu}^{if}\hat{\mathbf{e}}_\mathrm{l}\boldsymbol{m}^{fi}\hat{\mathbf{b}}_\mathrm{l}^* + \boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}^*\boldsymbol{m}^{if}\hat{\mathbf{b}}_\mathrm{l}) - (\boldsymbol{\mu}^{if}\hat{\mathbf{e}}_\mathrm{l}^*\boldsymbol{m}^{fi}\hat{\mathbf{b}}_\mathrm{l} + \boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}\boldsymbol{m}^{if}\hat{\mathbf{b}}_\mathrm{l}^*)}_{(\mathrm{II})} \quad (27)$$

Because $\boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}^*\boldsymbol{\mu}^{if}\hat{\mathbf{e}}_\mathrm{l} = \boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}\boldsymbol{\mu}^{if}\hat{\mathbf{e}}_\mathrm{l}^*$ and so on (the wavefunctions are real), the term (I) vanishes. Upon rearranging term (II) so that the EDTMs are all from $|f\rangle$ to $|i\rangle$ and the magnetic moments are the reverse, Equation (27) becomes Equation (28)[8]

$$\mathrm{CD} = A_\mathrm{l} - A_\mathrm{r} = 2k[\boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}^*\boldsymbol{m}^{if}\hat{\mathbf{b}}_\mathrm{l} - \boldsymbol{\mu}^{fi}\hat{\mathbf{e}}_\mathrm{l}\boldsymbol{m}^{if}\hat{\mathbf{b}}_\mathrm{l}^*] \quad (28)$$

Substitution of explicit forms for the electric and magnetic field (Equations 5 and 25) polarizations for the circularly polarized radiation, gives Equation (29)[8]

$$\mathrm{CD} = -2\mathrm{i}k[\mu_Y^{fi}m_Y^{if} + \mu_Z^{fi}m_Z^{if}] = 2k\mathrm{Im}[\mu_Y^{fi}m_Y^{if} + \mu_Z^{fi}m_Z^{if}] \quad (29)$$

where 'Im' denotes 'the imaginary part of'.

CD experiments are usually performed on collections of randomly oriented molecules. In that case the laboratory (radiation) defined axis system does not relate in any fixed way to a molecular axis system, so what

is measured is a rotational average of Equation (29) in Equation (30a):

$$\mathrm{CD} = \tfrac{4}{3}k\langle i|\tilde{\boldsymbol{\mu}}|f\rangle\langle f|\tilde{\boldsymbol{m}}|i\rangle = \tfrac{4}{3}k\mathrm{Im}[\boldsymbol{\mu}^{fi}\boldsymbol{m}^{if}] \qquad (30a)$$

The CD for the transition *from* state $|i\rangle$ *to* state $|f\rangle$ is thus given by the Rosenfeld equation (Equation 30b):[29]

$$\mathrm{CD} = \mathrm{Im}[\boldsymbol{\mu}^{fi}\boldsymbol{m}^{if}] \qquad (30b)$$

Recall that $\boldsymbol{\mu}^{fi}$ is the EDTM from state $|f\rangle$ to state $|i\rangle$.

If we limit our consideration to achiral molecules but impose an external magnetic field, then term (II) of Equation (27) vanishes, but term (I) does not (the states are complex rather than real). Further, we ignore the terms that involve squares of MDTMs as they will be much smaller than any EDTM terms. The magnetic circular dichroism (MCD) may then be written as Equation (31)

$$\begin{aligned}\mathrm{MCD} &= k[(\mu_X^{fi} - \mathrm{i}\mu_Z^{fi})(\mu_Y^{if} + \mathrm{i}\mu_Z^{if}) - (\mu_Y^{fi} + \mathrm{i}\mu_Z^{fi})(\mu_Y^{if} + \mathrm{i}\mu_Z^{if})] \\ &= 2\mathrm{i}k(\mu_Y^{fi}\mu_Z^{if} - \mu_Z^{fi}\mu_Y^{if}) \qquad (31)\end{aligned}$$

We shall not deal further with MCD in this article. See Rodger and Nordén[8,31] for more details.

2.4 Circular Dichroism of Magnetic Dipole Allowed Transitions: Why the Octant Rule Works for Carbonyls

In this section and the next we shall be working within the chromophoric or independent systems/perturbation (ISP) approach to CD theory,[8,12,13,15] and will limit our consideration to transitions in achiral chromophores, A, that acquire CD intensity by coupling to transitions in other chromophores C. C need not itself be chiral but it must be positioned so that {A + C} forms a chiral system. A *chromophore* is strictly a part of a system whose wavefunctions have *no* overlap with the rest of the system; electronic wavefunctions in a chromophore therefore have no electron exchange with the rest of the system. In practice, a chromophore is usually identified for a given electronic transition if that transition seems to be more or less dependent on only the identity of a subset of the molecule (such as the carbonyl functional group).

We also need to be able to write an equation for the interaction between A and C. Since there is no electron exchange between A and C (by definition), their interaction, V, is purely coulombic, Equation (32)

$$V = \sum_{a,c} \frac{q_\mathrm{a}q_\mathrm{c}}{|\boldsymbol{R}_\mathrm{AC} + \boldsymbol{r}_\mathrm{c} - \boldsymbol{r}_\mathrm{a}|} \qquad (32)$$

where q_a is the charge of a particle in A located at the end of the vector, $\boldsymbol{r}_\mathrm{a}$, that begins at the A origin, similarly q_c and $\boldsymbol{r}_\mathrm{c}$, and $\boldsymbol{R}_\mathrm{AC} = \boldsymbol{R}_\mathrm{C} - \boldsymbol{R}_\mathrm{A}$ is the vector from the origin of chromophore A to the origin of chromophore C (Figure 13).

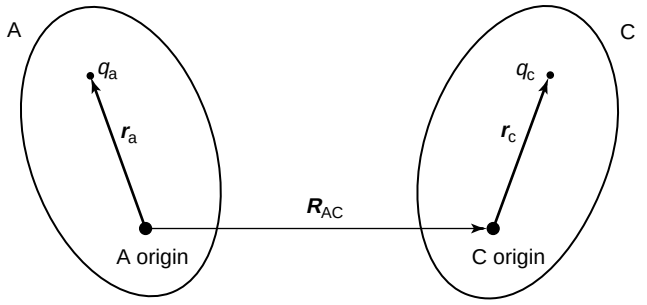

Figure 13 A and C chromophores.

The electric dipole moment operator, $\tilde{\boldsymbol{\mu}}$, where $\tilde{}$ is used here to indicate an operator for the whole system, may be written in terms of A and C operators as, Equation (33)

$$\tilde{\boldsymbol{\mu}} = e\left\{\sum_a(\hat{\boldsymbol{R}}_\mathrm{A} + \hat{\boldsymbol{r}}_\mathrm{a}) + \sum_c(\hat{\boldsymbol{R}}_\mathrm{C} + \hat{\boldsymbol{r}}_\mathrm{c})\right\} \qquad (33)$$

where $\hat{\boldsymbol{R}}_\mathrm{A} = \boldsymbol{R}_\mathrm{A}$ is the position operator for the A origin in the global coordinate system, and $\hat{\boldsymbol{r}}_\mathrm{a} = \boldsymbol{r}_\mathrm{a}$ is the position operator within A for particle a (Figure 13). We shall be considering only transition moments, so since $\boldsymbol{R}_\mathrm{A}$ is constant it follows that[8]

$$\tilde{\boldsymbol{\mu}} = e\left\{\sum_a(\hat{\boldsymbol{r}}_\mathrm{a}) + \sum_c(\hat{\boldsymbol{r}}_\mathrm{c})\right\} = \hat{\boldsymbol{\mu}}_\mathrm{A} + \hat{\boldsymbol{\mu}}_\mathrm{C} \qquad (34)$$

where $\hat{\boldsymbol{\mu}}_\mathrm{A}$ is the electric dipole moment operator for A and operates only on wavefunctions of A.

Similarly it follows from Equation (15) that the magnetic dipole moment operator for the whole system is,[8] Equation (35)

$$\begin{aligned}\tilde{\boldsymbol{m}} &= \breve{\boldsymbol{m}}_\mathrm{A} + \breve{\boldsymbol{m}}_\mathrm{C} \\ &= \frac{e}{2m_\mathrm{e}}\left\{\sum_a(\hat{\boldsymbol{R}}_\mathrm{A} + \hat{\boldsymbol{r}}_\mathrm{a})\hat{\boldsymbol{p}}_\mathrm{a} + \sum_c(\hat{\boldsymbol{R}}_\mathrm{C} + \hat{\boldsymbol{r}}_\mathrm{c})\hat{\boldsymbol{p}}_\mathrm{c}\right\} \\ &= \frac{e}{2m_\mathrm{e}}\hat{\boldsymbol{R}}_\mathrm{A}\sum_a\hat{\boldsymbol{p}}_\mathrm{a} + \hat{\boldsymbol{m}}_\mathrm{A} + \frac{e}{2m_\mathrm{e}}\hat{\boldsymbol{R}}_\mathrm{C}\sum_c\hat{\boldsymbol{p}}_\mathrm{c} + \hat{\boldsymbol{m}}_\mathrm{C} \\ &= \frac{e}{2m_\mathrm{e}}\hat{\boldsymbol{R}}_\mathrm{A}\hat{\boldsymbol{p}}_\mathrm{A} + \hat{\boldsymbol{m}}_\mathrm{A} + \frac{e}{2m_\mathrm{e}}\hat{\boldsymbol{R}}_\mathrm{C}\hat{\boldsymbol{p}}_\mathrm{C} + \hat{\boldsymbol{m}}_\mathrm{C} \qquad (35)\end{aligned}$$

where $\hat{\boldsymbol{m}}_\mathrm{A}$ is the total magnetic moment for chromophore A, $\hat{\boldsymbol{p}}_\mathrm{A}$ is the linear momentum operator for chromophore A, $\hat{\boldsymbol{m}}_\mathrm{A}$ is the intrinsic magnetic dipole moment operator within A, and so on. Thus the total magnetic moment for A includes both the intrinsic magnetic moment operator within the chromophore and any magnetic moment due to the linear motion of its charges about the system origin.

To evaluate the Rosenfeld equation (Equation 30b) for two interacting chromophores we need to know $\boldsymbol{\mu}^{01}$ and $\boldsymbol{m}^{01}$ for the whole system, i.e. for the $|0\rangle \rightarrow |1\rangle$ transition, which is the $|0) \rightarrow |1)$ transition on A in the presence of any perturbation due to C. We write the wavefunctions for the A/C system when their interaction is switched off as the product wavefunctions $|a)|c) = |ac)$ where $|a)$ is a wavefunction of A and $|c)$ is a wavefunction of C. The A function is always written first even in the ket form, thus, Equation (36)

$$\{|a)|c)\}^* = \{|ac)\}^* = (a|(c| = (ac| \tag{36}$$

A operators only operate on A wavefunctions, similarly C operators only operate on C wavefunctions. For example, Equation (37)

$$(a'c'|\widehat{\boldsymbol{\mu}}_{\mathrm{A}}|ac) = (a'|\widehat{\boldsymbol{\mu}}_{\mathrm{A}}|a)(c'|c) = \boldsymbol{\mu}_{\mathrm{A}}^{aa'}\delta_{cc'} \tag{37}$$

where $\delta_{cc'} = 0$ unless $c' = c$.

We are considering the situation when the $|0) \rightarrow |1)$ transition of A is MDA and EDF so $\boldsymbol{m}_{\mathrm{A}}^{01} \neq 0$ but $\boldsymbol{\mu}_{\mathrm{A}}^{10} = 0$. An EDF transition of an achiral chromophore, A, must 'borrow' some electric character in order to have a CD signal. If we also assume that *transition* moments give significantly stronger effects than *permanent* moments, then the 'dynamic coupling' model for the CD of MDA transitions follows.[12] Historically much more effort has been put into what is known as the 'static coupling' or 'one electron' mechanism (since one electron on A is assumed to be the only one to move) where the molecular framework provides a static perturbation via permanent moments. This mechanism certainly contributes to the net observed CD, however, its contribution is almost always smaller than that due to the dynamic coupling mechanism. More details about the static coupling model may be found in references in Schellman.[31]

To determine the wavefunctions of the interacting A/C system we first consider an A/C system where A and C do not interact. The Hamiltonian, $\widehat{H}$, of the noninteracting system is the sum, Equation (38)

$$\widehat{H} = \widehat{H}_{\mathrm{A}} + \widehat{H}_{\mathrm{C}} \tag{38}$$

where $\widehat{H}_{\mathrm{A}}$ is the Hamiltonian of isolated A, and similarly $\widehat{H}_{\mathrm{C}}$. This Hamiltonian is separable and the total energy of a state $|ac)$ is the energy of $|a)$ in A plus that of $|c)$ in C. So, for example, $|01)$ means A is in its ground state and C is in state $|1)$. The energy of the $|01)$ state of the combined system is therefore $0 + \varepsilon = \varepsilon$ (taking the zero point of energy to be when both A and C are in their ground states).

When the electrostatic interaction, V (Equation 32), is 'switched' on between A and C then the Hamiltonian of the interacting system is, Equation (39)

$$\widehat{H} = \widehat{H}_{\mathrm{A}} + \widehat{H}_{\mathrm{C}} + V \tag{39}$$

The wavefunctions for the combined system could be determined using the Equation (32) expression for V and the variation principle. We shall use perturbation theory to give expressions for the wavefunctions and later assume that A and C are far enough apart to expand V as a double Taylor series expansion to give terms dependent on monopoles, dipoles, quadrupoles and so on of A and C. This approach has the disadvantage of giving different final equations for degenerate and nondegenerate systems (and ignores near degenerate systems); however, it has the advantage of enabling us to write equations that can be used to relate system geometry to CD signal.

The perturbed wavefunctions $|i\rangle = |00\rangle$ and $|f\rangle = |10\rangle$ for the MDTM case where the perturbation comes from EDTMs are determined from nondegenerate perturbation theory to first order in V Equations (40) and (41):[3–5,32]

$$|00\rangle = |00) - \frac{(1c|V|00)}{\varepsilon_{\mathrm{c}} + \varepsilon_{\mathrm{a}}}|1c) \tag{40}$$

$$|10\rangle = |10) - \frac{(oc|V|10)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}}|0c) \tag{41}$$

from which it follows that, Equation (42)

$$\begin{aligned}\boldsymbol{\mu}^{10} &= \langle 00|\tilde{\boldsymbol{\mu}}|10\rangle = \langle 00|\widehat{\boldsymbol{\mu}}_{\mathrm{A}} + \widehat{\boldsymbol{\mu}}_{\mathrm{C}}|10\rangle \\ &= (00|\widehat{\boldsymbol{\mu}}_{\mathrm{A}}|10) - \frac{(00|V|1c)}{\varepsilon_{\mathrm{c}} + \varepsilon_{\mathrm{a}}}(1c|\widehat{\boldsymbol{\mu}}_{\mathrm{C}}|10) \\ &\quad - \frac{(0c|V|10)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}}(00|\widehat{\boldsymbol{\mu}}_{\mathrm{C}}|0c) \\ &= -\frac{(00|V|1c)}{\varepsilon_{\mathrm{a}} + \varepsilon_{\mathrm{c}}}\widehat{\boldsymbol{\mu}}_{\mathrm{C}}^{0c} - \frac{(0c|V|10)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}}\widehat{\boldsymbol{\mu}}_{\mathrm{C}}^{c0} = -\frac{2\varepsilon_{\mathrm{c}}V^{1c}}{\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2}\widehat{\boldsymbol{\mu}}_{\mathrm{C}}^{c0}\end{aligned} \tag{42}$$

since the A and C wavefunction are real. Similarly, Equation (43)

$$\begin{aligned}\boldsymbol{m}^{01} &= \langle 10|\tilde{\boldsymbol{m}}|00\rangle = \langle 10|\widehat{\boldsymbol{m}}_{\mathrm{A}} + \widehat{\boldsymbol{m}}_{\mathrm{C}}|00\rangle \\ &= (10|\widehat{\boldsymbol{m}}_{\mathrm{A}}|00) - \frac{(1c|V|00)}{\varepsilon_{\mathrm{a}} + \varepsilon_{\mathrm{c}}}(10|\widehat{\boldsymbol{m}}_{\mathrm{C}}|1c) \\ &\quad - \frac{(10|V|0c)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}}(0c|\widehat{\boldsymbol{m}}_{\mathrm{C}}|00) \\ &= \boldsymbol{m}_{\mathrm{A}}^{01} - \frac{(1c|V|00)}{\varepsilon_{\mathrm{a}} + \varepsilon_{\mathrm{c}}}\boldsymbol{m}_{\mathrm{C}}^{c0} - \frac{(10|V|0c)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}}\boldsymbol{m}_{\mathrm{C}}^{0c} \\ &= \boldsymbol{m}_{\mathrm{A}}^{01} - \frac{2\varepsilon_{\mathrm{a}}V^{1c}}{\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2}\boldsymbol{m}_{\mathrm{C}}^{0c}\end{aligned} \tag{43}$$

Substitution into the Rosenfeld equation (Equation 30b) gives the so-called dynamic coupling CD expressions to first order in V,[14,33] Equation (44)

$$R = \mathrm{Im}\left\{\frac{-2\varepsilon_c V^{1c}}{\varepsilon_c^2 - \varepsilon_a^2}\boldsymbol{\mu}_C^{c0}\boldsymbol{m}_A^{01}\right\} \tag{44}$$

The above equations are for systems composed of A plus one C. If there are two or more Cs then a sum over Cs is introduced.

In order to apply Equation (44) we need to have explicit forms for V^{1c} in terms of electric transition moments of A and C. It is now convenient to use a general form for the terms in the expansion of V using $i, j, k = \{x, y, z\}$ and the Kronecker delta, $\delta_{i,j}$, which is zero unless $i = j$. The dipole–dipole term in the expansion of V is[20] given in Equation (45)

$$\begin{aligned} V^{1c}(\boldsymbol{\mu}_A - \boldsymbol{\mu}_C) = V^{10,0c} &= (0c|V|10) \\ &= \frac{\boldsymbol{\mu}_A^{10}\boldsymbol{\mu}_C^{0c} - 3\widehat{\boldsymbol{R}}_{AC}\boldsymbol{\mu}_A^{10}\boldsymbol{\mu}_C^{0c}\widehat{\boldsymbol{R}}_{AC}}{R_{AC}^3} \\ &= \frac{1}{R_{AC}^3}(\mu_A^{10})_i(\mu_C^{0c})_j[\delta_{ij} - 3(\widehat{R}_{AC})_i(\widehat{R}_{AC})_j] \end{aligned} \tag{45}$$

where a sum over all repeated indices is implied. When substituted into Equation (44) the dipole–dipole term of V leads to a CD expression containing a $(\mu_A^{01})_i(m_A^{10})_k$ term. This moment product vanishes except for chromophores of C_s symmetry (i.e. those with only one reflection plane and no rotation axes) and x- and y- polarized transitions of C_{2v} chromophores. An MDA transition in such a chromophore is also EDA so the CD will probably be dominated by the coupled-oscillator CD of the EDTM which is dependent on $(R_{AC})^{-2}$ (see below).

For magnetically z-polarized transitions in A chromophores of C_{2v} symmetry such as the $n \rightarrow \pi^*$ transition of carbonyls (see above) the quadrupole–dipole term of V is the first nonvanishing one, Equation (46):[14]

$$\begin{aligned} V^{1c}(\boldsymbol{Q}_A - \boldsymbol{\mu}_C) = \frac{3}{2R_{AC}^4}&[(Q_A^{10})_{ij}(\mu_C^{0c})_k] \\ &\times \begin{array}{c}\delta_{ij}(\widehat{R}_{AC})_k + \delta_{ik}(\widehat{R}_{AC})_j + \delta_{jk}(\widehat{R}_{AC})_i \\ -5(\widehat{R}_{AC})_i(\widehat{R}_{AC})_j(\widehat{R}_{AC})_k\end{array} \end{aligned} \tag{46}$$

where $\bar{\boldsymbol{Q}}_A$ is the electric quadrupole moment operator of A. It is a second rank tensor whose components are, Equation (47)

$$(\bar{\boldsymbol{Q}}_A)_{ij} = e\sum_a (\boldsymbol{r}_i\boldsymbol{r}_j) \tag{47}$$

(cf. Equation 15 for notation). The quadrupole–dipole CD term is thus given by Equation (48)

$$\begin{aligned} R = -\frac{3}{R_{AC}^4}\mathrm{Im}&[(Q_A^{10})_{ij}(m_A^{01})_l] \quad \frac{\varepsilon_c(\mu_C^{c0})_l(\mu_C^{0c})_k}{(\varepsilon_c^2 - \varepsilon_a^2)} \\ &\times \begin{array}{c}\delta_{ij}(\widehat{R}_{AC})_k + \delta_{ik}(\widehat{R}_{AC})_j + \delta_{jk}(\widehat{R}_{AC})_i \\ -5(\widehat{R}_{AC})_i(\widehat{R}_{AC})_j(\widehat{R}_{AC})_k\end{array} \end{aligned} \tag{48}$$

Unless $\{i, j\}$ equals $\{x, y\}$ or $\{y, x\}$ the A-moment product (first square brackets in Equation 48) vanishes. Thus the CD strength for MDA z-polarized transitions of C_{2v} chromophores such as the $n \rightarrow \pi^*$ transition of carbonyls (Figure 4) is, Equation (49):[15]

$$\begin{aligned} R(C_{2v}, z) = -\frac{6}{R_{AC}^4}\mathrm{Im}&\quad (Q_A^{01})_{xy}(m_A^{10})_z \quad \frac{\varepsilon_c(\mu_C^{0c})_z}{(\varepsilon_c^2 - \varepsilon_a^2)} \\ &\times \begin{array}{c}(\mu_C^{c0})_x(\widehat{R}_{AC})_y + (\mu_C^{c0})_y(\widehat{R}_{AC})_x \\ -5(\mu_C^{c0})_k(\widehat{R}_{AC})_x(\widehat{R}_{AC})_y(\widehat{R}_{AC})_k\end{array} \end{aligned} \tag{49}$$

The octant rule follows from Equation (49) when we reduced our consideration to the term that has an isotropic polarizability component (the octant rule term), i.e. set $k = z$ and ignore the first two terms in the final brackets. As noted above, there are occasions when the octant rule is not appropriate because other terms in Equation (49) (or even higher order terms in the V expansion) become dominant.

Equation (49) after appropriate symmetry averaging also leads to the CD expressions for degenerately x,y-polarized transitions of molecules with D_{3d} A (Figures 7 and 8), since the A moment product of Equation (49) does not vanish for these polarizations under this symmetry. Take z to be the three-fold rotation axis and x to align with one of the two-fold rotation axes of A. By symmetry, the only nonvanishing A moment products are those in Equation (50) so the E-polarized tris chelate transition metal complex CD is:[20]

$$\begin{aligned} R\left(D_{3d}, \frac{x}{y}\right) = -\mathrm{Im}&\frac{3\varepsilon_c}{4R_{AC}^4(\varepsilon_c^2 - \varepsilon_a^2)} \\ &\begin{array}{c}(Q_A^{10})_{xx}(m_A^{01})_x - (Q_A^{10})_{yy}(m_A^{01})_x \\ -2(Q_A^{10})_{xy}(m_A^{01})_y\end{array} \\ &\times \begin{array}{c}2(\mu_C^{c0})_x(\mu_C^{0c})_x(\widehat{R}_{AC})_x - 5(\mu_C^{c0})_k \\ \times (\mu_C^{0c})_x(\widehat{R}_{AC})_x(\widehat{R}_{AC})_x(\widehat{R}_{AC})_k\end{array} \\ &\times \begin{array}{c}-4(\mu_C^{c0})_x(\mu_C^{0c})_y(\widehat{R}_{AC})_y + 5(\mu_C^{c0})_k \\ \times (\mu_C^{0c})_x(\widehat{R}_{AC})_y(\widehat{R}_{AC})_y(\widehat{R}_{AC})_k\end{array} \\ &\times \begin{array}{c}-2(\mu_C^{c0})_y(\mu_C^{0c})_y(\widehat{R}_{AC})_x + 10(\mu_C^{c0})_k \\ \times (\mu_C^{0c})_x(\widehat{R}_{AC})_x(\widehat{R}_{AC})_y(\widehat{R}_{AC})_k\end{array} \\ &-4[(Q_A^{10})_{yz}(m_A^{01})_x - (Q_A^{10})_{xz}(m_A^{01})_y] \\ &\times \begin{array}{c}(\mu_C^{c0})_y(\mu_C^{0c})_z(\widehat{R}_{AC})_x + 5(\mu_C^{c0})_k \\ \times (\mu_C^{0c})_x(\widehat{R}_{AC})_y(\widehat{R}_{AC})_z(\widehat{R}_{AC})_k\end{array} \\ &\times \begin{array}{c}-(\mu_C^{c0})_z(\mu_C^{0c})_x(\widehat{R}_{AC})_y - 5(\mu_C^{c0})_k \\ \times (\mu_C^{0c})_y(\widehat{R}_{AC})_x(\widehat{R}_{AC})_z(\widehat{R}_{AC})_k\end{array} \\ &\begin{array}{c}+(\mu_C^{c0})_x(\mu_C^{0c})_y(\widehat{R}_{AC})_z - (\mu_C^{c0})_y \\ \times (\mu_C^{0c})_x(\widehat{R}_{AC})_z\end{array} \end{aligned} \tag{50}$$

The isotropic polarizability part of Equation (50) is given in Equation (51)

$$R(D_{3d})_{x,y} = \frac{15\varepsilon_c\alpha_C(\varepsilon_a)}{4R_{AC}^4(\varepsilon_c^2-\varepsilon_a^2)} \quad \mathrm{Im}[(Q_A^{10})_{xx}(m_A^{01})_x - (Q_A^{10})_{yy}(m_A^{01})_x - 2(Q_A^{10})_{xy}(m_A^{01})_y] \times(\widehat{R}_{AC})_x[(\widehat{R}_{AC})_x^2 - 3(\widehat{R}_{AC})_y^2] \tag{51}$$

Equation (51) is the basis of the above discussion of E_g polarized d–d transitions of tris-chelate metal complexes.

When A has higher than two-fold rotational symmetry, then the quadrupole–dipole term of Equation (50) is zero for transitions polarized along the high symmetry rotation axes (e.g. for A_2 polarized transitions of tris-chelates). Since the octupole term in the V expansion also vanishes under these circumstances,[20] the term in the V expansion for A_2 polarized transitions is the hexadecapole–dipole term which has an R_{AC}^{-6} dependence so is smaller than the E_g R_{AC}^{-4} term; hence the dominant E band rule discussed above.

2.5 Circular Dichroism of Electric Dipole Allowed Transitions

For EDA transitions, the independent systems approach leads to mechanisms that are usually referred to as coupled-oscillator models. Because we choose to use perturbation theory to give expressions for wavefunctions (as discussed above), we need to consider identical and nonidentical {A,C} systems separately.

2.5.1 Identical A and C: the Degenerate Coupled Oscillator Model

Most transitions for which we measure UV/VIS absorbance spectra are EDA. If an EDA transition, with EDTM $\boldsymbol{\mu}_A^{01}$ and transition energy ε, of an achiral chromophore, A, has a measurable CD spectrum, then it must have "borrowed" some magnetic character to give a net helical electron rearrangement. The degenerate coupled-oscillator model describes the situation where A and C are identical and the magnetic character comes from the coupling of $\boldsymbol{\mu}_A^{01}$ with the same energy EDTM in C, $\boldsymbol{\mu}_C^{01}$. $\boldsymbol{\mu}_C^{01}$ must be oriented skew relative to $\boldsymbol{\mu}_A^{01}$ for a CD signal to be observed near ε. $\boldsymbol{\mu}_C^{01}$ induces a magnetic effect in A since, although it is a linear motion of charge within C, when viewed from A it moves (circles) around A's origin ($\boldsymbol{\mu}_A^{01}$ simultaneously induces a magnetic component into the C transition $\boldsymbol{\mu}_C^{01}$).

In this case, where A and C are identical, by symmetry (or degenerate perturbation theory) the two states that result from the mixing of $|10\rangle$ (A in $|1\rangle$ and C in the ground state) and $|01\rangle$ are given by Equation (52)[3–5,32]

$$|10_\pm\rangle = \tfrac{1}{\sqrt{2}}\{|10\rangle \pm |01\rangle\} \tag{52}$$

where the $|10_+\rangle$ state takes the upper sign of the $\pm$ or $\mp$ terms in this equation and those that follow. As above, we use real normalized wavefunctions. If we also assume that permanent moments are much smaller than transition moments, then the energies of these two states are, Equation (53)[3–7,32]

$$\begin{aligned}\varepsilon^\pm &= \langle 10_\pm|\widehat{H}_A + \widehat{H}_C \pm V|10_\pm\rangle \\ &= \tfrac{1}{2}\{(10| \pm (01|\}|\widehat{H}_A + \widehat{H}_C + V|\{|10\rangle \pm |01\rangle\} \\ &= \tfrac{1}{2}\{(1|\widehat{H}_A|1) + (0|\widehat{H}_C|0) \pm (10|V|01) \\ &\quad + (0|\widehat{H}_A|0) + (1|\widehat{H}_C|1) \pm (01|V|10)\} \\ &= \varepsilon^0 + \varepsilon^1 \pm V^{10,01} = \varepsilon \pm V^{11}\end{aligned} \tag{53}$$

where, for example, Equation (54)

$$\begin{aligned}V^{ba,dc} &= (ac|V|bd) \\ &= (\boldsymbol{\mu}_A^{ba}\boldsymbol{\mu}_C^{dc} - 3\widehat{\boldsymbol{R}}_{AC}\boldsymbol{\mu}_A^{ba}\boldsymbol{\mu}_C^{dc}\widehat{\boldsymbol{R}}_{AC} \\ &\quad + \text{higher order terms})\, R_{AC}^{-3}\end{aligned} \tag{54}$$

Because all wavefunctions are real, when one state on each of A and C is the ground state we further simplify the notation as indicated in the last line of Equation (53). V^{11} is much smaller than ε so the two new perturbed states are close in energy.

The EDTMs for the transitions from the two excited states of Equation (52) to the ground state $|00\rangle$ are, Equation (55):

$$\begin{aligned}\langle 00|\tilde{\boldsymbol{\mu}}|10_\pm\rangle &= \tfrac{1}{\sqrt{2}}\{(00|\widehat{\boldsymbol{\mu}}_A + \widehat{\boldsymbol{\mu}}_C[|10\rangle \pm |01\rangle]\} \\ &= \tfrac{1}{\sqrt{2}}\{\boldsymbol{\mu}_A^{10} \pm \boldsymbol{\mu}_C^{10}\}\end{aligned} \tag{55}$$

The $|0\rangle \rightarrow |1\rangle$ transitions in both A and C are MDF, and so have no intrinsic magnetic moment. We therefore set $\widehat{\boldsymbol{m}}_A = \widehat{\boldsymbol{m}}_C = 0$ in Equation (35). The momenta MDTM terms are more conveniently dealt with when re-expressed in terms of EDTMs for the appropriate transition exploiting a very useful relationship,[32] Equation (56)

$$(k|\widehat{\boldsymbol{p}}_C|j) = \frac{\mathrm{i}m_e}{e\hbar}(\varepsilon_k - \varepsilon_j)\boldsymbol{\mu}_C^{jk} \tag{56}$$

from which follows Equation (57)

$$\begin{aligned}\boldsymbol{m}_C^{jk} = (k|\tilde{\boldsymbol{m}}_C|j) &= \frac{e}{2m_e}(k|\widehat{\boldsymbol{R}}_C\widehat{\boldsymbol{p}}_C|j) \\ &= \frac{i}{2\hbar}\boldsymbol{R}_C(\varepsilon_k - \varepsilon_j)\boldsymbol{\mu}_C^{jk}\end{aligned} \tag{57}$$

This is the magnetic moment at the origin (0,0,0) created by the tangential EDA transition moment $\boldsymbol{\mu}_C^{jk}$. The

MDTMs for the return transition from to the ground state to the perturbed excited states are then, Equation (58)

$$(10_{\pm}|\tilde{\boldsymbol{m}}|00) = \frac{\mathrm{i}\varepsilon}{\sqrt{8}\hbar}\left\{\boldsymbol{R}_{\mathrm{A}}\mu_{\mathrm{A}}^{01} \pm \boldsymbol{R}_{\mathrm{C}}\mu_{\mathrm{C}}^{01}\right\} \tag{58}$$

Substituting Equations (55) and (58) into the Rosenfeld equation (Equation 30b) gives the CD strengths of bands centered at ε^{+} and ε^{-} (Equation 53) arising from the coupling of degenerate $|0) \rightarrow |1)$ transitions on A and C to be,[8] Equation (59)

$$\begin{aligned} R_{\pm}^{01} &= \pm\frac{\varepsilon}{4\hbar}\left\{\mu_{\mathrm{C}}^{01}\mu_{\mathrm{A}}^{01}(R_{\mathrm{C}} - \boldsymbol{R}_{\mathrm{A}})\right\} \\ &= \pm\frac{\varepsilon}{4\hbar}\left\{\boldsymbol{\mu}_{\mathrm{C}}^{01}\boldsymbol{\mu}_{\mathrm{A}}^{01}\boldsymbol{R}_{\mathrm{AC}}\right\} \end{aligned} \tag{59}$$

where the '+' transition is not necessarily higher in energy than the '−' transition.

The physical interpretation and applications of Equation (59) for solving geometric and spectroscopic problems is most easily seen by choosing the coordinate system illustrated in Figure 14 in which case Equation (59) may be written, Equation (60):

$$R^{\pm} = \pm\frac{\varepsilon\mu^2 R_{\mathrm{AC}}}{4\hbar}\sin\alpha\sin\gamma\sin\tau \tag{60}$$

and the energies of the two bands are, Equation (61)

$$\varepsilon^{\pm} = \varepsilon \pm \frac{\mu^2(\sin\alpha\sin\gamma\cos\tau + 2\cos\alpha\cos\gamma)}{R_{\mathrm{AC}}^3} \tag{61}$$

where μ is the magnitude of μ_{A} and μ_{C}.

The expected CD spectrum for two degenerate chromophores as a function of τ is illustrated in Figure 15. Because independent systems theory is identical to the theory of exciton particles, these degenerate couplets are often referred to as exciton CD spectra.

2.5.2 *Different A and C: Nondegenerate Coupled Oscillator Circular Dichroism*

When A and C are not identical then the wavefunctions of the A/C system follow from nondegenerate perturbation theory.[3–5,32] We may write Equations (62) and (63)

$$|00\rangle = |00) - \frac{(1c|V|00)}{\varepsilon_{\mathrm{c}} + \varepsilon_{\mathrm{a}}}|1c) \tag{62}$$

$$|10\rangle = |10) - \frac{(0c|V|10)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}}|0c) \tag{63}$$

The EDTM of the perturbed $|10\rangle \rightarrow |00\rangle$ transition is given by Equation (64)

$$\begin{aligned} \boldsymbol{\mu}^{10} &= \langle 00|\tilde{\boldsymbol{\mu}}|10\rangle = \langle 00|\hat{\boldsymbol{\mu}}_{\mathrm{A}} + \hat{\boldsymbol{\mu}}_{\mathrm{C}}|10\rangle \\ &= (00|\hat{\boldsymbol{\mu}}_{\mathrm{A}}|10) - (00|\hat{\boldsymbol{\mu}}_{\mathrm{C}}|0c)\frac{(0c|V|10)}{\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{a}}} \\ &\quad - (1c|\hat{\boldsymbol{\mu}}_{\mathrm{C}}|10)\frac{(00|V|1c)}{\varepsilon_{\mathrm{c}} + \varepsilon_{\mathrm{a}}} = \boldsymbol{\mu}_{\mathrm{A}}^{10} - \frac{2\varepsilon_{\mathrm{c}}V^{1c}}{\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2}\boldsymbol{\mu}_{\mathrm{C}}^{c0} \end{aligned} \tag{64}$$

Similarly,[8] Equation (65)

$$\begin{aligned} \boldsymbol{m}^{01} &= \langle 10|\tilde{\boldsymbol{m}}|00\rangle = \frac{e}{2m_{\mathrm{e}}}\left\{\boldsymbol{R}_{\mathrm{A}}\boldsymbol{p}_{\mathrm{A}}^{01} - \frac{2\varepsilon_{\mathrm{a}}V^{1c}}{\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2}\boldsymbol{R}_{\mathrm{C}}\boldsymbol{p}_{\mathrm{C}}^{0c}\right\} \\ &= \frac{\mathrm{i}}{2\hbar}\left\{\varepsilon_{\mathrm{a}}\boldsymbol{R}_{\mathrm{A}}\boldsymbol{\mu}_{\mathrm{A}}^{01} - \frac{2\varepsilon_{\mathrm{a}}\varepsilon_{\mathrm{c}}V^{1c}}{\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2}\boldsymbol{R}_{\mathrm{C}}\boldsymbol{\mu}_{\mathrm{C}}^{0c}\right\} \end{aligned} \tag{65}$$

since the transition has no intrinsic magnetic moment within A as in the case of the degenerate coupled oscillator model.

The CD strength is therefore (Equation 31)[8]

$$\begin{aligned} R(|00\rangle \longrightarrow |10\rangle) &= \frac{-\varepsilon_{\mathrm{a}}\varepsilon_{\mathrm{c}}V^{1c}}{\hbar(\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2)} \\ &\quad \times \{\boldsymbol{\mu}_{\mathrm{A}}^{10}\boldsymbol{R}_{\mathrm{C}}\boldsymbol{\mu}_{\mathrm{C}}^{0c} + \boldsymbol{\mu}_{\mathrm{C}}^{c0}\boldsymbol{R}_{\mathrm{A}}\boldsymbol{\mu}_{\mathrm{A}}^{01}\} \\ &= \frac{-\varepsilon_{\mathrm{a}}\varepsilon_{\mathrm{c}}V^{1c}}{\hbar(\varepsilon_{\mathrm{c}}^2 - \varepsilon_{\mathrm{a}}^2)}\{\boldsymbol{\mu}_{\mathrm{C}}^{0c}\boldsymbol{\mu}_{\mathrm{A}}^{01}\boldsymbol{R}_{\mathrm{AC}}\} \end{aligned} \tag{66}$$

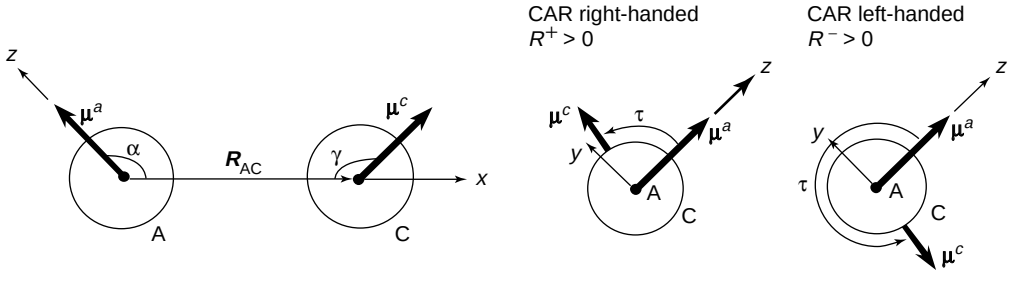

Figure 14 Diagram illustrating the geometry and coordinates for an A/C system. Note that τ is the angle taken in the anticlockwise direction between the *projections* of the EDTMs onto the y–z plane when the observer is looking down the x-axis. $0^{\circ} < \tau < 180^{\circ}$ if $\boldsymbol{\mu}^c \times \boldsymbol{\mu}^a\boldsymbol{R}_{\mathrm{AC}} > 0$ (i.e. if the three vectors form a right-handed parallelepiped), and $180^{\circ} < \tau < 360^{\circ}$ if $\boldsymbol{\mu}^c \times \boldsymbol{\mu}^a\boldsymbol{R}_{\mathrm{AC}} < 0$. CAR right-handed summarizes a mnemonic for the situation where $\boldsymbol{\mu}^c$, $\boldsymbol{\mu}^a$, and $\boldsymbol{R}_{\mathrm{AC}}$ form a right-handed axis system with $0^{\circ} < \tau < 180^{\circ}$.

Again using the coordinate system defined in Figure 14 we may write Equation (67)

$$R(\varepsilon_a) = \frac{-\varepsilon_a \varepsilon_c (\mu^a \mu^c)^2}{\hbar(\varepsilon_c^2 - \varepsilon_a^2) R_{AC}^2} \{\sin\alpha \sin\gamma \cos\tau + 2\cos\alpha \cos\gamma\} \times \sin\alpha \sin\gamma \sin\tau \quad (67)$$

which is usually easier to implement than Equation (66).

3 LINEAR DICHROISM

3.1 Introduction

LD is the difference in absorption of light linearly polarized parallel and perpendicular to an orientation axis, Equation (68):

$$\text{LD} = A_{\parallel} - A_{\perp} \quad (68)$$

LD is related to CD in that both require the difference between the absorbances of different polarized light beams to be measured and CD spectropolarimeters can be adapted to produce the required alternating beams of polarized light for LD. However, LD measurements are performed on systems that are either intrinsically oriented or are oriented during the experiment (Figure 16). The general LD references given below may be consulted for more details on the definitions and equations that follow. Rodger and Nordén[8] use the formalism adopted here.

The equations for LD were given in Equations (22) and (23) in the context of the interaction of radiation with matter. In order to understand LD more fully, consider what happens when a molecule absorbs a photon: absorption can be pictorially viewed as either the electric field or the magnetic field (or both) of the radiation pushing the electron density from a starting arrangement to a higher energy final one. The direction of net linear displacement of charge is known as the polarization of the transition. The polarization and intensity of a transition are characterized by the EDTM (see Equations 18 and 19) which may be regarded as the antenna by which the molecule absorbs light. Each transition thus has its own antenna and the maximum probability of absorbing light is obtained when the antenna and the electric field of the light are parallel.

Now imagine a linearly polarized light beam (Figure 16) and a sample of molecules all oriented in exactly the same way. If the normal absorption is first measured with the light polarized so that it is parallel to the direction of orientation of the sample and then measured when the light is polarized perpendicular to this direction, the difference between these two spectra is the LD spectrum. The two extreme situations are:

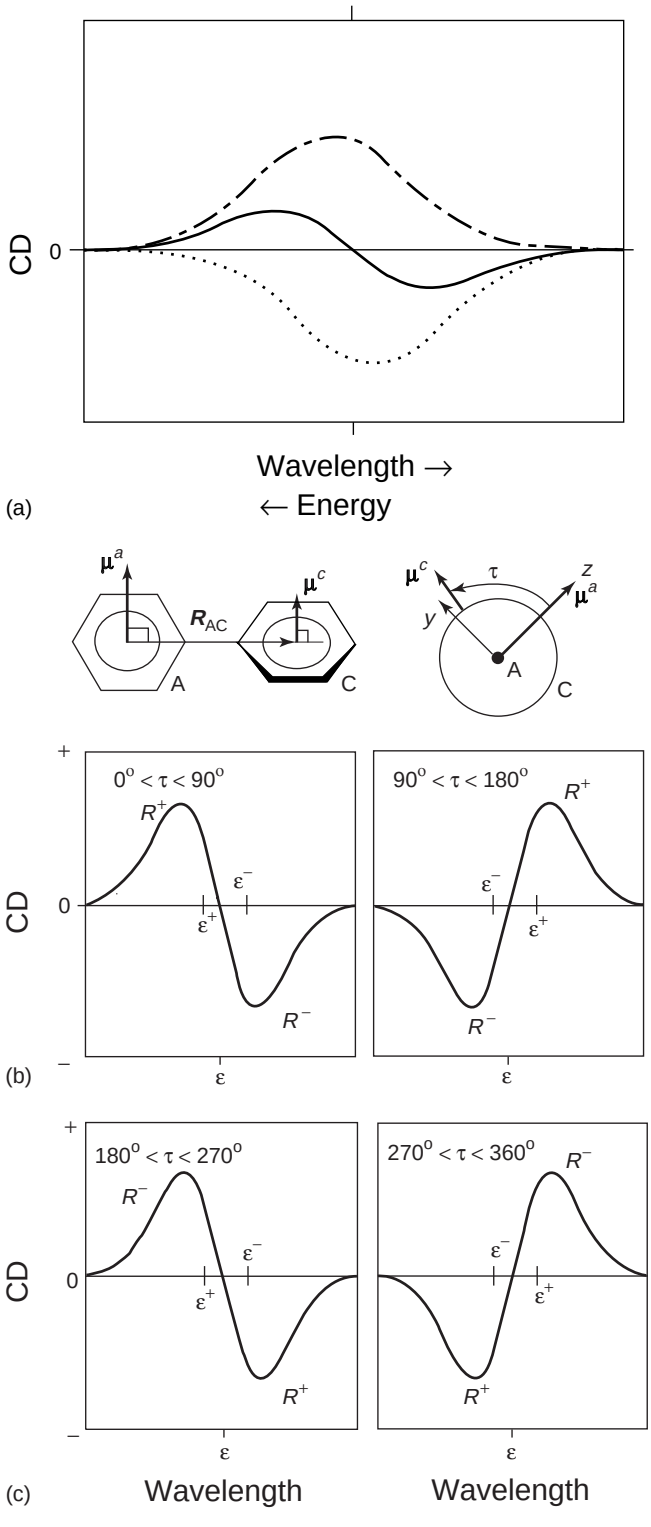

Figure 15 Exciton CD spectra for coupling 'short-axis' polarized transitions. (a) Illustrates characteristic form that results from cancellation of overlapping bands; - - - -, higher energy component; — – —, lower energy component; —, observed CD spectrum. (b) Illustrates the geometry of the system. (c) Shows the net spectra as a function of τ (assuming Gaussian band shapes for the two transitions).

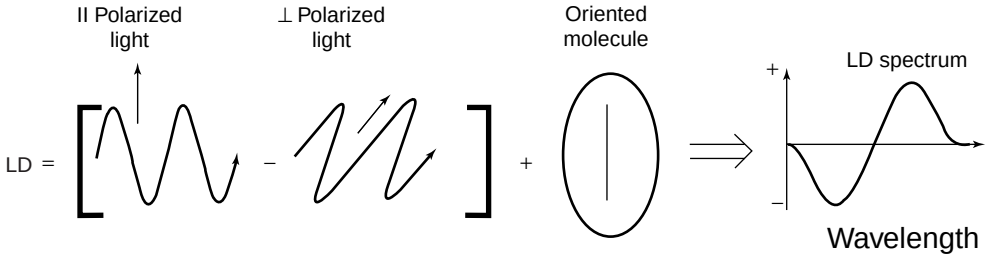

Figure 16 Schematic illustration of LD.

1. If the polarization of the transition that is being probed is perfectly parallel to the orientation direction, then Equation (69) follows

$$\mathrm{LD} = A_{\parallel} - A_{\perp} = A_{\parallel} > 0 \tag{69}$$

2. If the polarization of the transition that is being probed is perpendicular to the orientation direction, then Equation (70) follows

$$\mathrm{LD} = A_{\parallel} - A_{\perp} = -A_{\perp} < 0 \tag{70}$$

For intermediate polarizations, the LD is between these cases. Thus, the polarization for a given transition can be determined from its LD spectrum if the orientation of the molecule is known; conversely, LD can be used as a probe of molecular orientation if the polarization of a transition moment within the molecule is known. In practice, one type of experiment is often performed, followed by the other, using different orientation methods.

The two extreme cases of Equations (69) and (70) are both illustrated in the simplified anthracene spectrum of Figure 17.[8] The spectrum is simplified in two ways:

1. perfect orientation of the molecules has been assumed, and
2. the long wavelength band has been assumed to be of pure short-axis polarization.

In reality perfect orientation is never achieved and there is a significant long-axis polarized component at ~320 nm due to coupling with the 250 nm band.

3.2 Orientation Parameters

A variety of different orientation methods for LD spectroscopy are outlined below. In practice, with the possible exception of a crystalline environment, the orientation is never perfect and we need to use the concept of an orientation distribution. In many cases, either the averaging inherent in the orientation method or the nature of the sample means that (at least) local uniaxial orientation can be assumed (Figure 18) and the simple orientation

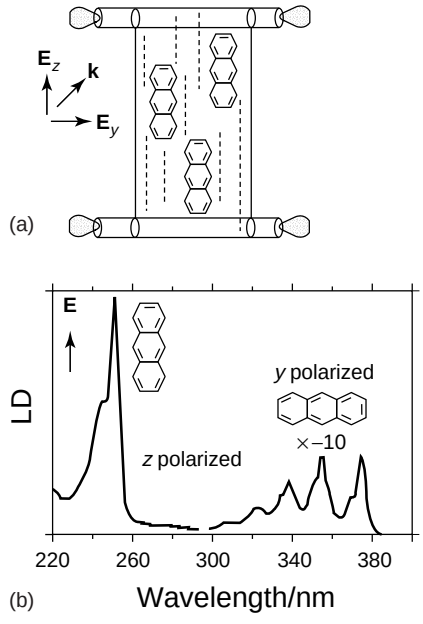

Figure 17 (a) Stretched film orientation showing propagation direction, **k**, and electric field polarization, **E**, of the electromagnetic radiation, and (b) LD spectrum of long (z) and short (y) axis polarized transitions of an idealized (see text for assumptions) anthracene sample.

parameter S can be used. Equation (71) is then valid

$$\mathrm{LD}^{\mathrm{r}} = \frac{\mathrm{LD}}{A} = \frac{A_{\parallel} - A_{\perp}}{A} = \frac{3}{2}S(3\cos^2\alpha - 1) \tag{71}$$

LD^{r} is the so-called reduced LD, S may be viewed as a scaling factor defining the efficiency of macroscopic orientation, and α is an angle that specifies the orientation of the transition moment that is responsible for the absorption of light at the particular wavelength (if several transitions absorb at the same wavelength an average is obtained) with respect to the orientation axis (Figure 19). $S = 1$ for perfect orientation and $S = 0$ for random (i.e. no) orientation. In Equation (71) A is the absorption of the LD sample under isotropic (unoriented) conditions so $A_{\parallel}$, $A_{\perp}$, and A all have Cl as a common factor, hence when using LD^{r} it is not necessary to know the sample concentration or path length. If the sample is macroscopically as

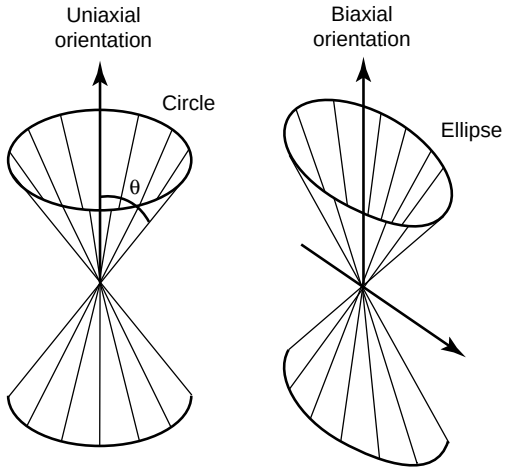

Figure 18 Uniaxial and biaxial orientation. Uniaxial orientation requires all orientations on a cone about the orientation axis to be statistically equally probable. This may be due to static or dynamic factors.

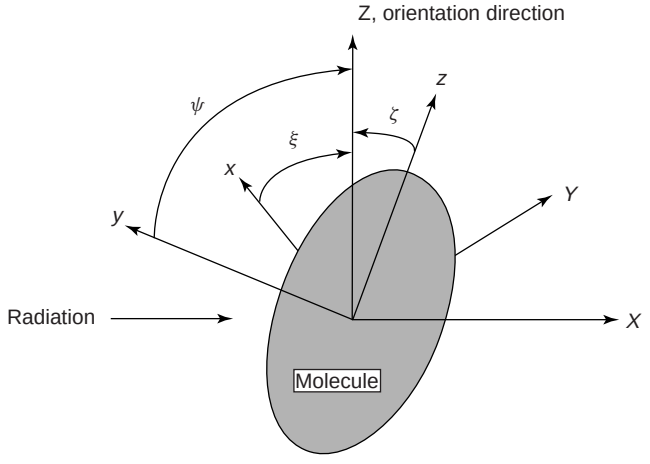

Figure 19 Macroscopic and molecular axes systems and angles relating them.

well as molecularly uniaxial, such as molecules oriented in a polymer film drawn in one direction or polar molecules in an electric field, there is a simple relation between $A_{\parallel}$, $A_{\perp}$, and A which makes it unneccessary to measure all three quantities, Equation (72):

$$A = \tfrac{1}{3}(A_{\parallel} + 2A_{\perp}) \tag{72}$$

When the orientation method used in an LD experiment provides a uniaxial sample (Figure 18) and the sample molecules are also characterized by a unique axis around which all orientations are equally probable (so the molecule is 'rod-like'), we may then write, Equation (73):

$$S = \tfrac{1}{2}(3\langle\cos^2\theta\rangle - 1) \tag{73}$$

where θ is the angle between the macroscopic orientation direction Z and the molecular orientation axis z (Figure 19). θ is used instead of ζ, to which the same definition is given below, as a reminder that Equation (73) only holds for rod-like molecules in uniaxial samples. It is important to note that the averages of ($\cos^2$ angles) are appearing in the above equations. These are not the same as the ($\cos^2$) of (average angles), though in practice we usually have to ignore what can be a considerable difference.

More generally, instead of a single orientation parameter we need to have three since in some LD experiments, including rectangular squeezing of a gel or when the light is incident radially upon a cylindrical couette flow cell (Figure 20), the system behaves as if the orientation were biaxial (Figure 18) for which the constant probability contour for a given molecular axis traces out an ellipse. (Note: flow-oriented DNA, see below, is uniaxially oriented due to the helical nature of the molecule.) A laboratory (macroscopic) axis system $\{X, Y, Z\}$ is defined first, such that X is the direction of propagation of the incident radiation and Z is the macroscopic direction of orientation. Y then completes the right-handed axis system. The more general definition of LD is then, Equation (74)

$$\mathrm{LD} = A_Z - A_Y \tag{74}$$

(cf. Equation 23). Similarly, we write Equation (75)

$$\mathrm{LD}^{\mathrm{r}} = 3\left(\frac{A_Z - A_Y}{A_X + A_Y + A_Z} \right. = 3(\langle\hat{\mu}_Z^2\rangle - \langle\hat{\mu}_Y^2\rangle) \tag{75}$$

When studying LD it is also convenient to define the molecular axis system, $\{x, y, z\}$. We take z to be the molecular orientation axis, i.e. the axis in the molecule which has maximum value shown by Equation (76) of

$$\langle\cos^2\zeta\rangle \tag{76}$$

where ζ is the angle between z and Z. x is the axis perpendicular to z that has the smallest value shown by Equation (77) of

$$\langle\cos^2\xi\rangle \tag{77}$$

where ξ is the angle between x and Z. ψ is similarly the angle between y and Z. z is often, but not always, the highest order symmetry determined axis.

The general definitions of the orientation parameters are given by Equation (78)

$$\boldsymbol{S} = \begin{matrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{matrix}$$

$$= \begin{matrix} \tfrac{1}{2}(3\langle\cos^2\xi\rangle - 1) & 0 & 0 \\ 0 & \tfrac{1}{2}(3\langle\cos^2\psi\rangle - 1) & 0 \\ 0 & 0 & \tfrac{1}{2}(3\langle\cos^2\zeta\rangle - 1) \end{matrix} \tag{78}$$

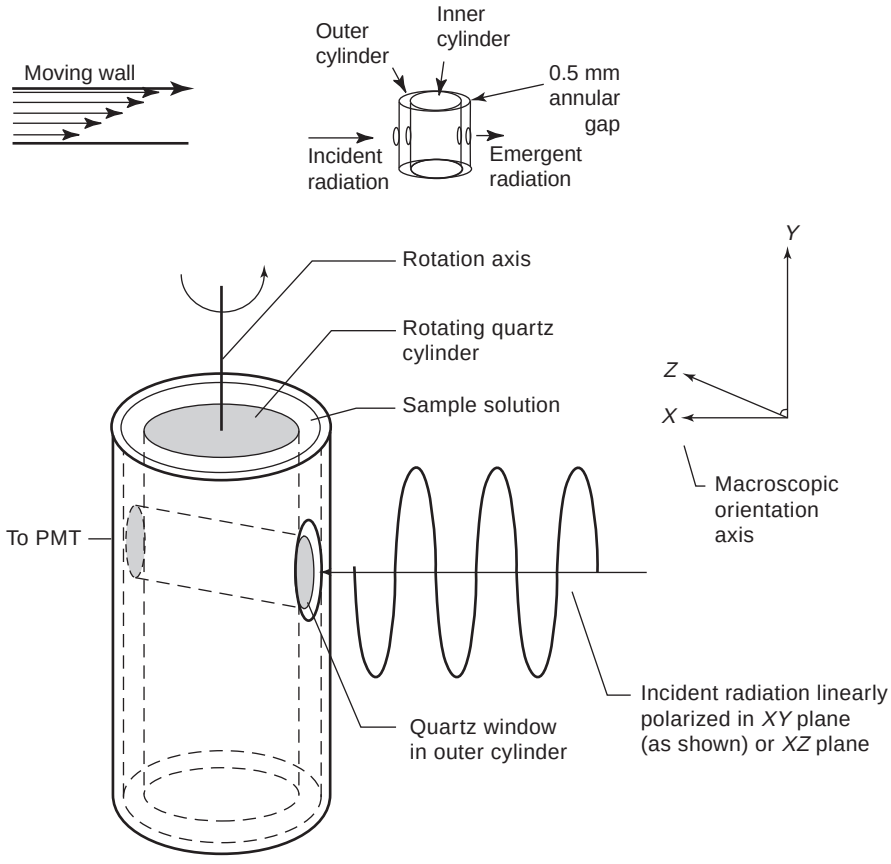

Figure 20 Flow orientation in a coaxial (couette) flow cell with radial incident light. The total light path is typically 1 mm. PMT = Photomultiplier tube.

Since we defined z to be the most oriented molecular axis, and x to be the least oriented Equation (79) follows:

$$S_{zz} \geq S_{yy} \geq S_{xx} \tag{79}$$

Further, the trace of the orientation tensor is zero so that, Equation (80):

$$S_{xx} + S_{yy} + S_{zz} = 0 \tag{80}$$

For rod-like molecules $S_{xx} = S_{yy}$ and S for the uniaxial case is S_{zz}.

In solving any particular problem, the positions of $\{x, y, z\}$ are seldom known, so it is usually convenient first to choose a temporary molecular axis system $\{x', y', z'\}$ which probably bears no relationship to the orientation axis system but is convenient for our analysis (see example below). Analogously to the definitions we used with the 'true' molecular axis system, the orientation of the macroscopic orientation axis Z in the new coordinate system may then be written, Equation (81), as the vector

$$(\cos\xi', \cos\psi', \cos\zeta')_{x',y',z'} \tag{81}$$

also, Equation (82)

$$\boldsymbol{S}' = \begin{matrix} S_{x'x'} & S_{x'y'} & S_{x'z'} \\ S_{y'x'} & S_{y'y'} & S_{y'z'} \\ S_{z'x'} & S_{z'y'} & S_{z'z'} \end{matrix} = \begin{matrix} \frac{1}{2}(3\langle\cos^2\xi'\rangle - 1) & 3\langle\cos\xi'\cos\psi'\rangle & 3\langle\cos\xi'\cos\zeta'\rangle \\ 3\langle\cos\xi'\cos\psi'\rangle & \frac{1}{2}(3\langle\cos^2\psi'\rangle - 1) & 3\langle\cos\psi'\cos\zeta'\rangle \\ 3\langle\cos\xi'\cos\zeta'\rangle & 3\langle\cos\psi'\cos\zeta'\rangle & \frac{1}{2}(3\langle\cos^2\zeta'\rangle - 1) \end{matrix} \tag{82}$$

For planar molecules in uniaxially oriented systems, $x' = x$ so $\xi' = \xi$. It follows that,[8,34]

$$\mathrm{LD}^{\mathrm{r}} = 3(S_{y'y'}\sin^2\alpha' + S_{z'z'}\cos^2\alpha' + S_{y'z'}\sin\alpha'\cos\alpha') \tag{83}$$

where α' is the angle between the transition dipole moment and axis z'. The three orientation parameters in Equation (83) are three unknowns, so three different transitions of known polarization are needed to determine them. The position of z is almost determined from

Equation (84)[8,34]

$$\tan(2\beta) = \frac{S_{y'z'}}{S_{z'z'} - S_{y'y'}} \quad (84)$$

where β is the angle between z' and z. Either chemical intuition or another experiment such as fluorescence polarization anisotropy[35] is required to determine on which side of z', z lies. It is often apparent upon inspection.

3.3 Molecular Alignment Techniques

In addition to the requirement of linearly polarized light for LD, LD experiments are made technically demanding because of the need to orient the sample. The orientation method used for this depends on the sample. Long polymers may be oriented by shear flow whereas small molecules require a stronger orienting force. A selection of orientation methods is described below. More details may be found in the general LD references.

3.3.1 Stretched Polymers

Small molecules can often be absorbed into polymer films, usually polyethylene or poly(vinyl alcohol) (both can be used well into the UV even down to 200 nm) depending on the polarity of the analyte. When the film is mechanically stretched, either before or after the small molecules are added, the absorbed molecules align their long axes preferentially along the stretch direction (Figure 17). There has been no evidence that a solute is better oriented when introduced into the film before stretching compared with it being added after stretching; with poly(vinyl alcohol); however, the film is usually cast containing the analyte so it cannot be added after stretching.

3.3.2 Flow Orientation

Long polymers may be oriented by the viscous drag caused when a solution is flowed between narrow walls.[36] Depending on the cell design, the light is then propagated either along the flow direction or perpendicular to it. This technique is commonly used for LD studies of DNA (the DNA needs to be at least of the order of 1000 base pairs in length to get significant orientation). The most successful flow cell has proved to be a cylindrical couette flow cell where the solution containing the DNA is subjected to a constant gradient over the annular gap between two coaxial cylinders one of which is rotating (Figure 20).

3.3.3 Electric Field Orientation

Effective uniaxial orientation for polar or polarizable molecules (including DNA since its ionic environment is polarizable) may be achieved by an electric field between two parallel plates.[37] Electric orientation is conceptually simpler than flow orientation, and the data can be fairly easily extrapolated to infinite orientation which contrasts with the situation for flow orientation. Heating effects may be circumvented by using pulsed field techniques, which additionally enable the study of relaxation phenomena. However, the pulsed nature of the field generally requires each electric dichroism experiment to be performed at a fixed wavelength or else a diode array system to be used; it is also usual to measure $A_{\parallel}$ and A (the unoriented absorbance) rather than $A_{\perp}$ (as the orientation is uniaxial, only two of the three absorbances need be determined independently as discussed above).

3.3.4 Squeezed Gel Orientation

The method of orientation that has proved most successful for membrane proteins, particularly photosynthetic ones, is to embed the protein in a gel (usually polyacrylamide) and then physically squeeze the gel either unidirectionally or bidirectionally (Figure 21).[28] The main shortcomings of the gel orientation technique are its limitation to wavelengths longer than 250 nm, and the limited dynamic range of deformation.

3.3.5 Migrative Orientation

While migrating in an electrophoresis gel, DNA is macroscopically oriented and gives an LD signal.

3.3.6 Magnetic Field Orientation

Magnetic fields may be used to orient molecules in much the same way as electric fields, however, the effect is small and this method is not widely used unless the particles are large (such as chloroplasts) and carry substantial magnetic dipole moments.

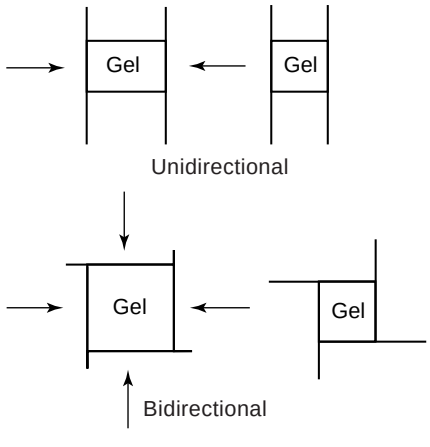

Figure 21 Principles of biaxial and uniaxial squeezing of gels.

3.3.7 Crystalline Samples

If the absorbance is not too large for the spectrometer to measure, crystals may be used for LD studies.

3.3.8 Liquid Crystal Orientation

The molecules within a liquid crystal are oriented with respect to one another and by sandwiching the liquid crystal between two quartz plates, macroscopic orientation may be achieved.[38] The sample must be tilted with respect to the light beam to get an LD signal. Molecules dissolved in a liquid crystal may also have a preferential orientation.

3.3.9 Orientation by Evaporation, for Example onto Quartz

This method may be appropriate for some samples, particularly planar aromatic molecules that readily adsorb onto quartz. The quartz plate has to be tilted for LD measurement.

3.4 Examples of Linear Dichroism Data

3.4.1 DNA

The absorption, and hence LD, spectra of nucleic acids in the easily accessible UV region of the spectrum (down to 180–190 nm) are dominated by the $\pi \rightarrow \pi^*$ transitions of the purine and pyrimidine bases which, by symmetry, are all polarized in the plane of the bases. If DNA is in the idealized B-DNA form with the DNA bases perpendicular to the helix axis[39] then a flow oriented DNA sample (Figure 20) should have its LD^r across the whole spectrum constant and negative. In practice, there is some variation in magnitude (Figure 22) which indicates that the bases are tilted.

Normal B-DNA, which according to the classical fiber structure has $\theta_X = -2.1^\circ$ and $\theta_Y = 4.0^\circ$ (Figure 23, though fiber B-DNA has 10 base pairs per turn of the helix rather than the 10.4–10.6 of solution DNA), should display only small variations in LD^r in the wavelength region 230–300 nm with $LD^r/S = -1.47 \pm 0.02$. A-form DNA, with $\theta_X = 19.3^\circ$ and $\theta_Y = -3.2^\circ$, is, however, expected to exhibit a sloping LD^r, with LD^r/S varying from about -1.2 at 230 nm to about -1.4 at 300 nm, depending on the sequence of the DNA.[40] Z-DNA is longer and stiffer and so is expected to have a significantly larger LD signal than the same sequence of B-DNA as shown in Figure 22.

LD data has been analyzed to determine the average values for base roll, tilt, twist, and slide for the different DNA bases in solution B-DNA; a value of $\alpha = 73^\circ$ is appropriate for purines, whereas $\alpha = 65^\circ$ for pyrimidines.[40] These results are dependent on the accuracy of the assignment of base transition

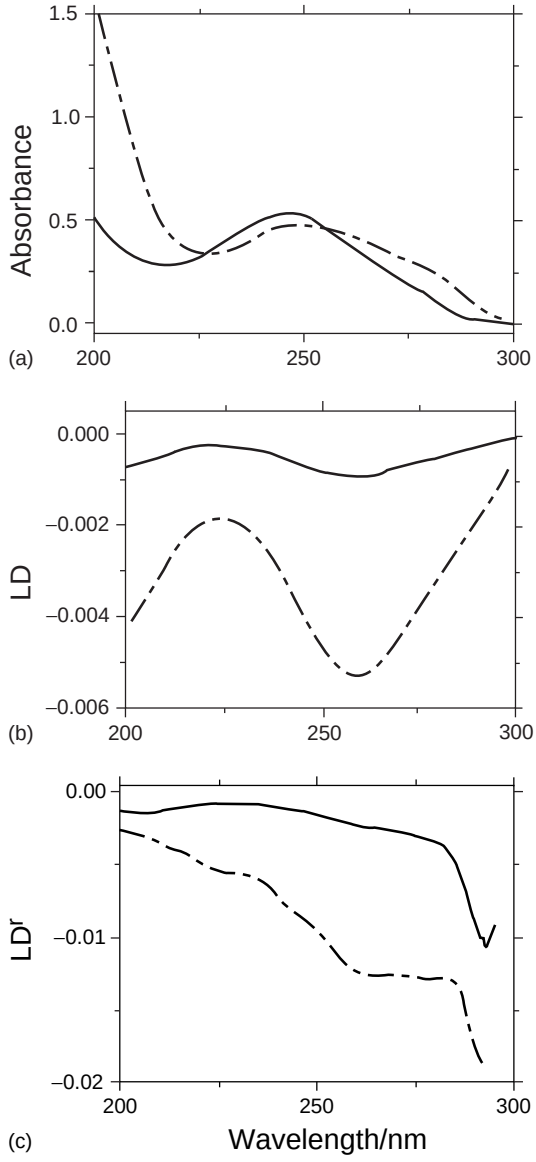

Figure 22 Absorbance (a), LD (b), and LD^r (c) spectra of B-form and Z-form poly[d(G–C)]$_2$ (60 μM in 10 mM NaCl, 1 mM cacodylate buffer, pH = 7). The Z-form is induced by adding 40 μM $[Co(NH_3)]^{3+}$. The DNA has been flow oriented. As all $\pi \rightarrow \pi^*$ transitions of the bases are polarized approximately perpendicular to the helix axis, the LD signals are negative. —, B-DNA; - – -, Z-DNA.

polarizations, intensities, and bandwidths. An effective average value of 84° for α at 260 nm is often assumed for B-DNA based on the fiber diffraction structure and an estimate of contributions from different transitions to the total LD.[42]

3.4.2 Linear Dichroism of DNA-bound Ligands

Once the orientation of DNA itself is understood a next step is to use LD to probe the binding and orientation

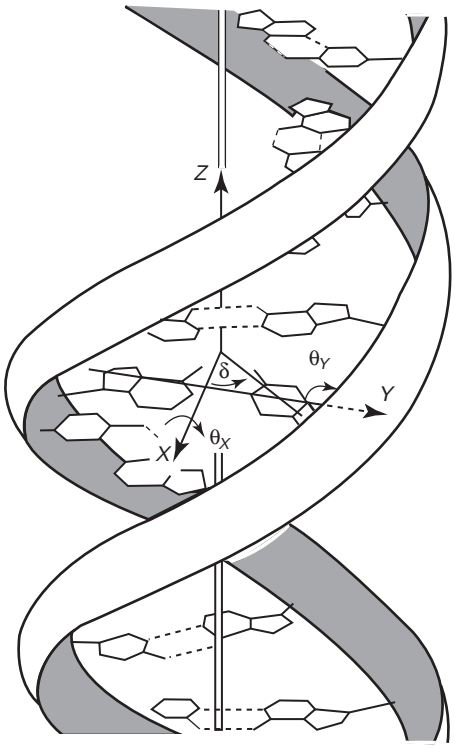

Figure 23 Geometry parameters for DNA as used in the text.

on DNA of small ligand molecules and even proteins. A key feature of such LD studies is that if there is no specific binding, then there will be no LD of the ligand transitions. Conversely, the presence of an LD signal from an absorption band of the putative ligand immediately implies that it is bound. Although LD cannot tell us where a ligand binds to DNA, it can be used to determine the orientation of the ligand on the DNA (if the ligand's transition moments have been assigned). This often provides significant clues as to the binding mode. A ligand may bind to DNA in a number of different types of site, as well as with sequence and orientation preferences. Broadly speaking a ligand may bind in the following ways.

1. *Externally* bound to the phosphate backbone: this is usually orientationally fairly nonspecific and the induced LD (and CD) is correspondingly small.
2. *Intercalated* between DNA bases: this mode requires planar aromatic molecules and the DNA to unwind and open up a slot between adjacent base-pairs so the ligand may be sandwiched between them. The DNA is stiffer and locally better oriented near the intercalation site, so the LD^r of an intercalator is usually slightly larger (more negative) at 260 nm than that of the average DNA base. If the bases are tilted, then the intercalator follows that tilt and this may be apparent in the LD^r. For example, the short axis of 9-hydroxyellipticine aligns with the long axis of the base-pair pocket and has a slightly smaller LD^r signal than the long axis of the molecule which is perpendicular to the pocket and also more perpendicular to the helix axis than the base pairs (Figure 25).
3. In the *minor groove*: this mode is frequently adopted by aromatic molecules containing internuclear bonds with some rotational freedom. For example, long molecules such as netropsin, distamycin, Hoechst 33258, and 4′,6-diamidino-2-phenylindole (Figure 24), which can fit snugly into the minor groove with the molecule following the curvature of the groove but are too large to fit into an intercalation site, favor this mode of binding. The long axis of

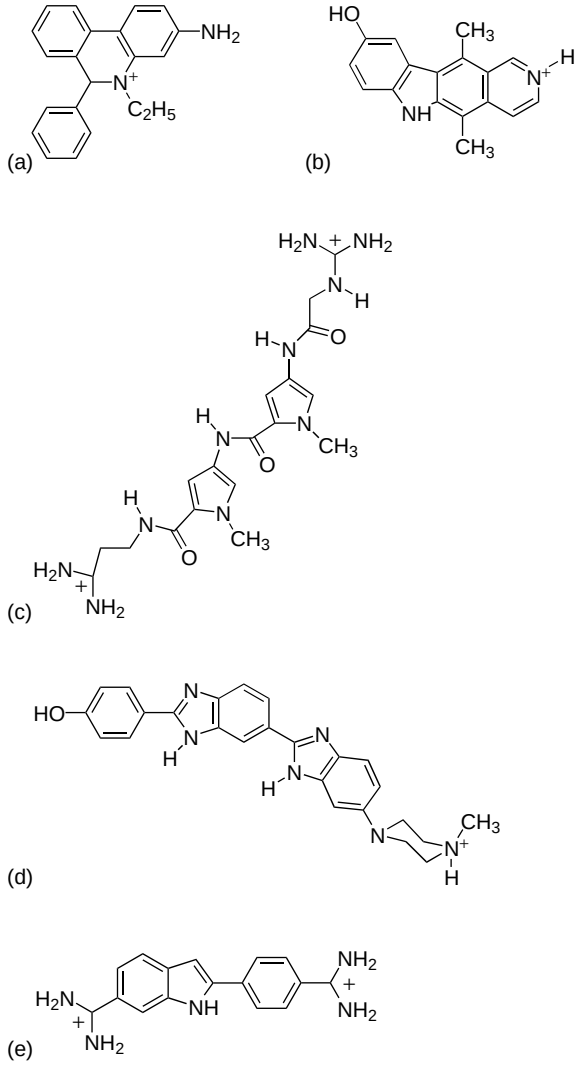

Figure 24 The intercalators (a) ethidium bromide and (b) 9-hydroxyellipticine, and some minor groove binders: (c) netropsin, (d) Hoechst 33258, and (e) 4′,6-diamidino-2-phenylindole.

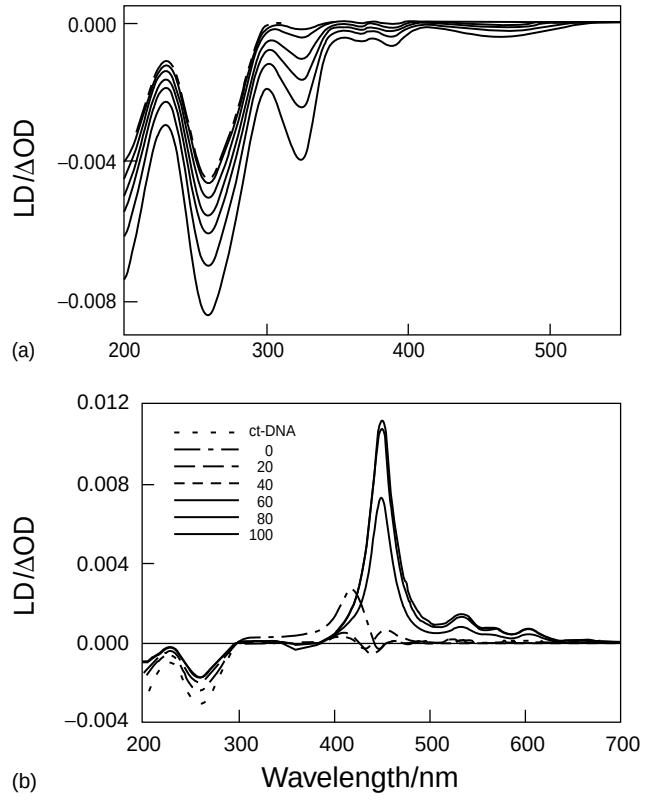

Figure 25 Some examples of LD spectra for DNA bound ligands. (a) 9-Hydroxyellipticine (1–10 μM in 2.5 μM increments, signal increases with ligand concentration) with 100 μM calf thymus DNA in NaCl (20 mm) and phosphate buffer (1 mM, pH 7);[41,45,46] (b) *trans*-bis-(4-*N*-methylpyridiniumyl)diphenylporphyrin (5 μM) with 40 μM calf thymus DNA in phosphate buffer (1 mM, pH 7) and varying NaCl concentrations (see figure insert) showing the change in binding orientation as a function of ionic strength. Signal increases with ligand concentration.

such molecules is oriented at about 45° to the helix axis and the short axis polarized transition perpendicular to it (approximately parallel to the base transitions). This mode is selective (or even specific) for adenine–thymine-rich DNAs as the guanine amino group sterically hinders the minor groove of guanine-containing sequences.

4. In the *major groove*: this type of binding is found for several regulatory proteins and there is enough space to accommodate most smaller ligands in a variety of orientations.

Some examples of LD spectra for DNA bound ligands are given in Figure 25.

3.4.3 Determination of the Molecular Orientation Axis of a Molecule in a Stretched Film

Infrared LD is particularly useful for determining orientations, particularly of small molecules in stretched films, because the polarization of vibrations are often known simply from their position in the spectrum. For example, riboflavin has three infrared bond stretching vibrations.[35] For convenience we choose the primed axis system as illustrated in Figure 26. The values of the LD^r for each transition are also given in Figure 26.

For the three transitions for which the data is given in Figure 26, Equation (83) leads to Equation (85)

$$\begin{aligned} +0.070 &= 3(S_{z'z'}) \\ +0.580 &= 3(0.030S_{z'z'} + 0.970S_{y'y'} - 0.174S_{y'z'}) \\ +0.110 &= 3(0.250S_{z'z'} + 0.750S_{y'y'} - 0.433S_{y'z'}) \end{aligned} \quad (85)$$

From which Equation (86) follows

$$S_{z'z'} = 0.024 \quad S_{y'y'} = 0.256 \quad S_{y'z'} = 0.373 \quad (86)$$

Thus, β (Equation 84) either equals 61° or 151°. By inspection of Figure 26, we conclude that $\beta = 61°$ is correct and the dashed line of Figure 26 is the molecular orientation axis z.

4 INSTRUMENTATION AND THE DICHROISM EXPERIMENT

4.1 Circular Dichrographs (Spectropolarimeters)

The essential features of a CD spectropolarimeter are a source of (more-or-less) monochromatic left and right circularly polarized light and a means of detecting the difference in absorbance of the two polarizations of light. The normal method of achieving these requirements, because the CD of molecules is generally quite weak compared with the total absorbance signal, is to implement a polarization phase-modulation technique. A photoelastic modulator (PEM) (in older instruments a Pockels cell) produces alternatively right and left circularly polarized

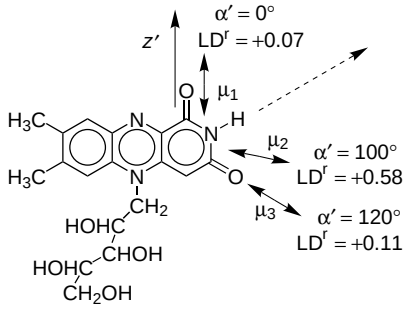

Figure 26 Riboflavin indicating the temporary axis system used in the calculation and the polarizations and LD^r values for the three transitions used in the text. Data are from Matsuoka and Nordén.[34]

light. The light intensity is constant, but upon passage through a sample exhibiting CD an intensity fluctuation (corresponding to the different absorptions of left and right circularly polarized light) that is in phase with the modulator frequency appears. The unabsorbed photons hit a photomultiplier tube which produces a current whose magnitude depends on the number of incident photons.

The light source in most CD instruments is a xenon arc lamp for UV and visible CD measurements. This means that the instrument has most sensitivity in the 300–400 nm region of the spectrum. The fact that a lamp is getting old becomes apparent by decreased light intensity in the 200 nm and 700 nm regions of the spectrum. The optics of a typical CD machine are illustrated in Figure 27. A series of mirrors and prisms and slits are used to produce collimated monochromatic radiation (in reality it is not monochromatic but has a well-defined wavelength range, typically of 0.5–2 nm). This light is then linearly polarized and subsequently circularly polarized. The conversion of linearly polarized light into circularly polarized light is achieved by the PEM.

The PEM consists of a piece of crystal quartz mechanically coupled (glued) to a piece of isotropic (silica) quartz, the light passing through the latter. The quartz plate is oriented so that light polarized (say) vertically travels through it at one speed and light polarized horizontally travels more slowly. If the linearly polarized light is incident on the plate at a 45° angle, then it may be considered as split into equal magnitude vertical and horizontal components. The PEM exhibits birefringence so that $(n_Z - n_Y) \neq 0$, i.e. the refractive indices n_Z and n_Y for vertical and horizontal polarizations are different. During passage through the PEM the phase difference between the vertical and horizontal light beam components will amount to, Equation (87):

$$\delta = \frac{2\pi D}{\lambda} = \frac{2\pi d(n_Z - n_Y)}{\lambda} \tag{87}$$

where d is the thickness of the PEM, λ is the wavelength of the light, and D is the path difference between the light components. At a certain value of $(n_Z - n_Y)$, δ will be equal to $\pi/2$ and D equal to $\lambda/4$, making the so-called 'quarter wave' plate required to produce circularly polarized light.

An ac (alternating current) voltage is applied to the crystal part, causing the whole PEM assembly to oscillate (typically at 50 kHz). By adjusting the voltage amplitude so that the birefringence amplitude corresponds to the quarter wave condition at each λ, the time dependence of $(n_Z - n_Y)$ as it oscillates between $+|n_Z - n_Y|$ and $-|n_Z - n_Y|$ gives light that is alternately left and right circularly polarized. The polarized light thus produced then passes through the sample compartment (and the sample) and what is not absorbed is detected by the photomultiplier tube. If the sample has no CD the photomultiplier current will be a constant, direct current (dc) with a magnitude determined by the normal absorbance of the sample. If the sample exhibits CD

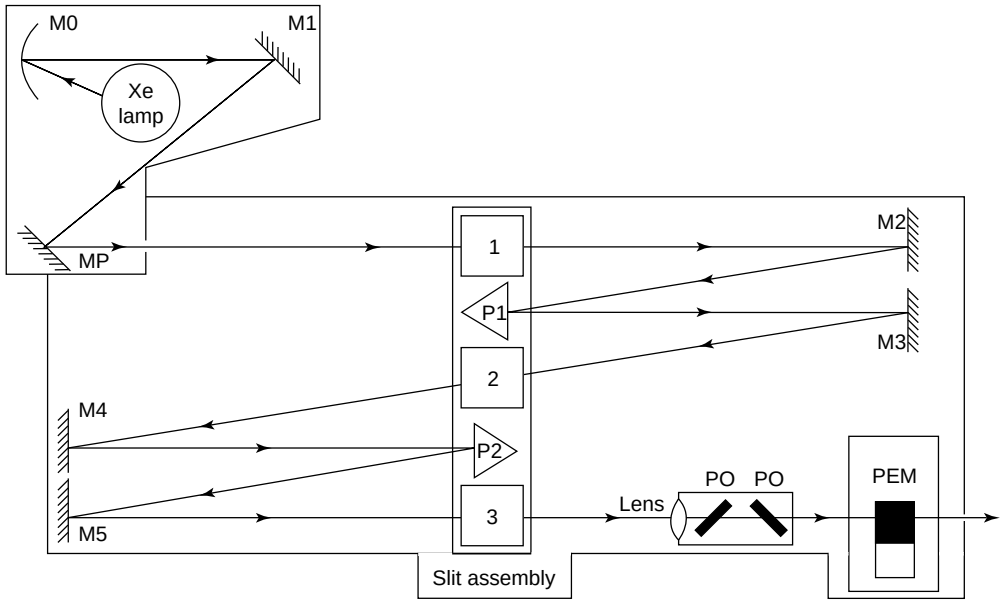

Figure 27 Schematic diagram of the optics of a CD spectropolarimeter. M denotes mirror, P denotes prism, and PO denotes polarizer. As illustrated the monochromator prisms (P1 and P2) are also the polarizers; to improve the polarization further a series of tilted quartz plates are inserted in the light path preceding the PEM. Sample and photomultiplier tube are attached after the PEM.

the photomultiplier current will also show an oscillating component (ac). The CD is obtained as the ratio between the ac and dc components. Its sign is determined from the phase of the ac component using a lock-in amplifier that has the ac voltage of the PEM as a time reference.

With both DNA and protein CD experiments it is important to think carefully about the buffer to be used. CD instruments are single beam spectrometers and so there is no automatic compensation mechanism for buffer absorbance as there is for normal absorbance spectroscopy where the buffer can be placed in a reference cuvette. Many CD experiments fail or give nonsense results as all the light is absorbed by the nonchiral parts of the solution.

4.2 Design and Implementation of an Linear Dichroism Experiment

The components of an LD experiment are a source of linearly polarized light, a means of detecting how much light is absorbed, a method of orienting the sample (as discussed above), and a way to change the relative orientations of sample and light beam. There are two main methods for measuring LD spectra. The one requiring less specialized equipment is the two-spectra method. Using a double beam spectrophotometer equipped with a polarizer (e.g. of Glan type) which is not sensitive to beam divergence, the sample is oriented parallel to the polarization direction of the polarizer to obtain $A_{\parallel}$. It is then oriented perpendicular to the polarization direction of the polarizer to obtain $A_{\perp}$. (Alternatively, the polarizer may be rotated, in which case any effect of internal polarization of the light by the optics of the spectrometer must be considered.)

The other method, the differential method, is much easier to implement and a wider range of sample orientation techniques may be used. However, much more sophisticated instrumentation is required as the phase-modulation technique of a circular dichrometer is used, either by supplementing the latter with a quarter wave device (e.g. a Fresnel rhomb or an Oxley prism to achieve achromacy[3–7]) or by increasing the driving voltage of the PEM and doubling the beat frequency of the lock-in amplifier. A factor of two increase in the voltage across the quarter wave PEM that turns linearly polarized light into alternating left and right circularly polarized light for CD measurements, turns the PEM into a half-wave plate that produces alternating pulses of orthogonal beams of linearly polarized light for LD spectroscopy. The differential method has the advantage of being extremely sensitive, needing only very low sample orientation or concentration.

ACKNOWLEDGMENTS

Professor Curtis Johnson's contribution to this article is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

ac	Alternating Current
CD	Circular Dichroism
dc	Direct Current
EDA	Electric Dipole Allowed
EDF	Electric Dipole Forbidden
EDTM	Electric Dipole Transition Moment
HPLC	High-performance Liquid Chromatography
ICD	Induced Circular Dichroism
ISP	Independent Systems/Perturbation
LD	Linear Dichroism
MCD	Magnetic Circular Dichroism
MDA	Magnetic Dipole Allowed
MDF	Magnetic Dipole Forbidden
MDTM	Magnetic Dipole Transition Moment
NMR	Nuclear Magnetic Resonance
OR	Optical Rotation
ORD	Optical Rotatory Dispersion
PEM	Photoelastic Modulator
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VCD	Vibrational Circular Dichroism

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Photodynamic Therapy

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Circular Dichroism in Analysis of Biomolecules • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Clinical Chemistry **(Volume 2)**
Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Forensic Science **(Volume 5)**
Chiroptical Spectroscopy in Drug Analysis

Nucleic Acids Structure and Mapping **(Volume 6)**
Nucleic Acids Structure and Mapping: Introduction • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • PNA and Its Applications • RNA Tertiary Structure

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction • Chiral Purity in Drug Analysis • Proteins and Peptides Purification in Pharmaceuticals Analysis • Steroid Analysis • Vibrational Spectroscopy in Drug Discovery, Development and Production

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Biopolymer Chromatography • Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Chemical Shifts in Nuclear Magnetic Resonance

FURTHER READING

General Circular Dichroism References

L.D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.

D.P. Craig, T. Thirunamachandran, *Molecular Quantum Electrodynamics: An Introduction to Radiation–Molecule Interaction*, Academic Press, London, 1984.

N. Harada, K. Nakanishi, *Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry*, University Science Books, California, 1983.

S.F. Mason, *Molecular Optical Activity and the Chiral Discrimination*, Cambridge University Press, Cambridge, 1982.

J. Michl, E.W. Thulstrup, *Spectroscopy with Polarized Light*, VCH, New York, 1986.

K. Nakanishi, N. Berova, R.W. Woody, (eds.) *Circular Dichroism: Principles and Applications*, VCH, New York, 1994.

F.S. Richardson, 'Theory of Optical Activity in the Ligand-field Transitions of Chiral Transition Metal Complexes', *Chem. Rev.*, **79**, 17–36 (1979).

A. Rodger, B. Nordén, Circular Dichroism and Linear Dichroism Oxford University Press, Oxford, 1997.

General Linear Dichroism References

J. Michl, E.W. Thulstrup, *Spectroscopy with Polarized Light*, VCH, New York, 1986.

B. Nordén, 'Applications of Linear Dichroism Spectroscopy', *Appl. Spectrosc. Rev.*, **14**, 157–248 (1978).

B. Nordén, C. Elvingson, M. Jonsson, B. Åkerman, 'Microscopic Behaviour of DNA During Electrophoresis: Electrophoretic Orientation', *Quart. Rev. Biophys.*, **24**, 103–164 (1991).

B. Nordén, M. Kubista, T. Kuruscev, 'Linear Dichroism Spectroscopy of Nucleic Acids', *Q. Rev. Biophys.*, **25**, 51–170 (1992).

A. Rodger, B. Nordén, *Circular Dichroism and Linear Dichroism*, Oxford University Press, Oxford, 1997.

J. Schellman, H.P. Jensen, 'Optical Spectroscopy of Oriented Molecules', *Chem. Rev.*, **87**, 1359–1399 (1987).

A. Rodger, 'Linear dichroism', *Methods Enzymol.*, **226**, 232–258 (1993).

REFERENCES

1. R.E. Allen, (ed.) *The Concise Oxford Dictionary*, Clarendon Press, Oxford, 1990.
2. P.J. Stephens, M.A. Lowe, *Ann. Rev. Phys. Chem.*, **36**, 213–241 (1985).
3. P.W. Atkins, *Molecular Quantum Mechanics*, Oxford University Press, Oxford, 1983.
4. P.W. Atkins, *Physical Chemistry*, 4th edition, Oxford University Press, Oxford, 1991.
5. J.M. Hollas, *Modern Spectroscopy*, 2nd edition, John Wiley and Sons, Chichester, 1992.
6. L.D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1982.
7. D.P. Craig, T. Thirunamachandran, *Molecular Quantum Electrodynamics, An Introduction to Radiation–Molecule Interaction*, Academic Press, London, 1984.
8. A. Rodger, B. Nordén, *Circular Dichroism and Linear Dichroism*, Oxford University Press, Oxford, 1997.
9. S. Allenmark, *Chromatographic Enantiomer Separation*, 2nd edition, Ellis Horwood Limited, Chichester, 1991.
10. W.H. Pirkle, A. Tsipouras, '3,5-Dinitrobenzoyl Aminoacid Esters – Broadly Applicable Chiral Solvating Agents for NMR Determination of Enantiomeric Purity', *Tetrahedron Lett.*, **26**, 3989–3992 (1985).
11. W. Moffitt, R.B. Woodward, A. Moscowitz, W. Klyne, C. Djerassi, 'Structure and the Optical Rotatory Dispersion of Saturated Ketones', *J. Am. Chem. Soc.*, **83**, 4013–4018 (1961).
12. E.G. Höhn, O.E. Weigang, 'Electron Correlation Models for Optical Activity', *J. Chem. Phys.*, **48**, 1127–1137 (1968).
13. F.S. Richardson, 'Theory of Optical Activity in the Ligand-field Transitions of Chiral Transition Metal Complexes', *Chem. Rev.*, **79**, 17–36 (1979).

14. P.E. Schipper, A. Rodger, 'Symmetry Rules for the Determination of the Intercalation Geometry of Host/Guest Systems Using Circular Dichroism: A Symmetry Adapted Coupled-Oscillator Model', *J. Am. Chem. Soc.*, **105**, 4541–4550 (1983).
15. A. Rodger, P.M. Rodger, 'The Circular Dichroism of the Carbonyl n-π^* Transition: An Independent Systems/Perturbation Approach', *J. Am. Chem. Soc.*, **110**, 2361–2368 (1988).
16. A. Rodger, M. Moloney, 'n-π^* Circular Dichroism of Planar Zig-Zag Carbonyl Compounds', *J. Chem. Soc., Perkin II*, 919–925 (1991).
17. A. Toumadje, W.C. Johnson, Jr, 'Systemin has the Characteristics of a Poly(L-proline)-II Type Helix', *J. Am. Chem. Soc.*, **117**, 7023–7024 (1995).
18. D.A. Lightner, T.D. Bouman, W.M.D. Wijekoon, A. Hansen, 'Mechanism of Ketone $n \rightarrow \pi^*$ Optical Activity. Experimental and Computed Chiroptical Properties of 4-axial and 4-equatorial alkyladamantan-2-ones', *J. Am. Chem. Soc.*, **108**, 4484–4497 (1986).
19. B. Bosnich, 'The Application of Exciton Theory to the Determination of the Absolute Configurations of Inorganic Complexes', *Acc. Chem. Res.*, **2**, 266–273 (1969).
20. P.E. Schipper, A. Rodger, 'Generalized Selection Rules for Circular Dichroism: A Symmetry Adapted Perturbation Model for Magnetic Dipole Allowed Transitions', *Chem. Phys.*, **109**, 173–193 (1986) and references therein.
21. J.A. Schellman, H.P. Jensen, 'Optical Spectroscopy of Oriented Molecules', *Chem. Rev.*, **87**, 1359–1399 (1987).
22. L. Stryer, *Biochemistry*, 3rd edition, W.H. Freeman and Company, New York, 1988.
23. K. Nakanishi, N. Berova, R.W. Woody, (eds.) *Circular Dichroism: Principles and Applications*, VCH, New York, 1994.
24. W.C. Johnson, 'Analyzing Protein Circular Dichroism Spectra for Accurate Secondary Structures', *Proteins: Struct. Funct. Genet.*, **7**, 307–312 (1999).
25. P. Manavalan, W.C. Johnson, 'Variable Selection Method Improves the Prediction of Protein Secondary Structure Using a Simple Matrix Multiplication', *Anal. Biochem.*, **167**, 76–85 (1987).
26. W.C. Johnson, Jr, 'Secondary Structure of Proteins Through Circular Dichroism Spectroscopy', *Ann. Rev. Biophys. Biophys. Chem.*, **17**, 145–166 (1988).
27. H. Votavova, R. Matlova, J. Sponar, 'B–Z Transition of Synthetic Oligonucleotides Containing Non-Z Forming Elements', *J. Biomol. Struct. Dyn.*, **12**, 163–172 (1994).
28. A. Rodger, 'Linear Dichroism', *Methods Enzymol.*, **226**, 232–258 (1993).
29. L. Rosenfeld, 'Quantenmechanische Theorie Der Natürlichen Optischen Aktivität Von Flüssigkeiten Und Gasen', *Z. Phys.*, **52**, 161–174 (1928).
30. A. Rodger, B. Nordén, 'A Coupled-oscillator Type Approach to Magnetic Circular Dichroism', *Enantiomer*, **3**, 409–421 (1998).
31. J.A. Schellman, 'Symmetry Rules for Optical Rotation', *Acc. Chem. Res.*, **1**, 144–151 (1968).
32. H. Eyring, J. Walter, G.E. Kimball, *Quantum Chemistry*, John Wiley and Sons, New York, 1944.
33. J.G. Kirkwood, 'On the Theory of Optical Rotatory Power', *J. Chem. Phys.*, **5**, 479–491 (1937).
34. Y. Matsuoka, B. Nordén, 'Linear Dichroism Studies of Nucleic Acid Bases in Stretched Poly(vinyl alcohol) Film. Molecular Orientation and Electronic Transition Moment Directions', *J. Phys. Chem.*, **86**, 1378–1386 (1982).
35. J.R. Lacowitz, *Principles of Fluorescence Spectroscopy*, Plenum Press, New York, 1983.
36. A. Wada, 'Dichroic Spectra of Biopolymers Oriented by Flow', *Appl. Spectros. Rev.*, **6**, 1–30 (1972).
37. E. Fredericq, C. Houssier, *Electric Dichroism and Electric Birefringence*, Clarendon Press, Oxford, 1973.
38. B. Nordén, G. Lindblom, I. Jonás, 'Linear Dichroism Spectroscopy as a Tool for Studying Molecular Orientation in Model Membrane Systems', *J. Phys. Chem.*, **81**, 2086–2093 (1977).
39. R.R. Sinden, *DNA Structure and Function*, Academic Press Ltd., San Diego, 1994.
40. P.-J Chou, W.C. Johnson, Jr, 'Base Inclinations in Natural and Synthetic DNAs', *J. Am. Chem. Soc.*, **115**, 1205–1214 (1993) and references therein.
41. A. Rodger, H.C. Latham, P. Wormell, A. Parkinson, M. Ismail, K.J. Sanders, 'DNA-drug Systems: how Circular Dichroism Data is Complemented by other Spectroscopic Techniques', *Enantiomer*, **3**, 393–408 (1998).
42. Y. Matsuoka, B. Nordén, 'Linear Dichroism Studies of Nucleic Acids. II Calculation of Reduced Dichroism Curves of A- and B-form DNA', *Biopolymers*, **21**, 2433–2452 (1982).
43. W.C. Johnson, Jr, 'Circular Dichroism and Its Empirical Application to Biopolymers', *Methods Biochem. Anal.*, **31**, 61–163 (1985).
44. S.F. Newbury, J.A. McClellan, A. Rodger, 'Spectroscopic and Thermodynamic Studies of Conformational Changes in Long, Natural mRNA Molecules', *Anal. Commun.*, **33**, 117–122 (1996).
45. K.J. Sanders, Spectroscopic Investigations of Interactions of Chiral Molecules in Biological Systems, PhD thesis, University of Warwick, 1998.
46. M.A. Ismail, DNA–Ligand Interactions: A Biophysical Study of 9-Hydroxyellipticine, Hoechst 33258 and a Meso-substituted Porphyrin Derivative Binding to DNA, PhD thesis, University of Warwick, 1998.

Detectors, Absorption and Luminescence

William E. Acree, Jr
University of North Texas, Denton, TX, USA

1 General Detector Requirements and Performance Characteristics 1
2 Photocathode and Photosensitive Materials 3
3 Photovoltaic or Barrier-layer Cells 4
4 Vacuum Phototubes 5
5 Photomultiplier Tubes and Microchannel Plate Photomultiplier Tubes 6
6 Photodiode and Photodiode Array Detectors 8
7 Charge Transfer Devices 12
8 Streak Tubes and Streak Cameras 15
9 Thermal Detectors 15
Abbreviations and Acronyms 17
Related Articles 17
References 17

Absorption and luminescence detectors convert radiant energy directly into an electrical signal or into another physical quantity, such as heat or resistance, which is subsequently converted into an electrical signal. Detectors are subdivided into three main categories depending upon their mode of operation. Thermal detectors sense the change in temperature that results from the absorption of incoming radiation. Photon detectors, on the other hand, respond to incident arrival rates rather than to photon energies. Multichannel detectors are arrays of multiple thermal detectors or arrays of multiple semiconductor detectors. Detectors of the third category permit simultaneous detection of dispersed radiation over a broad spectral wavelength region. Absorption and luminescent detectors are components in UV (ultraviolet)-visible, infrared (IR) and atomic absorption spectrophotometers and in fluorescence, phosphorescence, Raman and atomic emission spectrometers.

1 GENERAL DETECTOR REQUIREMENTS AND PERFORMANCE CHARACTERISTICS

Commercial absorption and luminescence detectors can vary widely in terms of their sensitivity, spectral response, response time, linearity and signal-to-noise ratios (S/Ns). Specific analytical applications often dictate the minimum performance characteristics that the radiation detector can possess. For example, in the case of a photomultiplier tube (PMT) the minimum requirements for measuring fluorescence lifetimes by both pulse and phase-modulated methods include: (1) fast time-response of typically 1.5 ns risetime, (2) electron gains exceeding 5 million, (3) low noise of less than 100 counts s^{-1}, (4) low transit-time dispersion with point of illumination on the cathode (i.e. <20 ps mm^{-1}), and (5) low afterpulsing from ion or photon feedback to the photocathode and low repulsing from photoelectron generation at the dynodes. The characteristics given below are commonly employed in evaluating the different absorption and luminescence detectors for a particular analytical application.

Response time is the measure of the detector's ability to detect rapid changes in the incident radiant power. Quantitatively, the response time is evaluated in terms of the time constant $\tau = 1/(2\pi f_c)$, where f_c is the frequency at which the incident radiant power has fallen to 0.707 of its maximum value when a sinusoidal input of frequency f_c is incident on the detector. The risetime is defined as the time required for the output to increase for 10–90% of its final value as a result of an instantaneous increase in the radiant power reaching the detector.

Linearity is a statement of whether the electrical signal produced by the detector, $S(\lambda)$, is directly proportional to the incident radiant power, $R(\lambda)$, Equation (1)

$$S(\lambda) = Q(\lambda)\, R(\lambda) \tag{1}$$

The sensitivity, $Q(\lambda)$, is defined as the slope of a plot of the electrical output of the detector versus incident radiant power, $R(\lambda)$. For photon detectors, and for many thermal detectors, both $Q(\lambda)$ and $R(\lambda)$ are wavelength dependent. The sensitivity of a given detector is thus specified at a particular wavelength. Usually, radiation detectors exhibit linearity over a limited range of incident radiant power. The total range, expressed in powers of 10, for which the electrical signal is linear is called the linear dynamic range. The dynamic range is defined as the total range of incident radiant power over which the detector is responsive.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Many real detectors exhibit a small, constant response even in the absence of radiation. The small electrical output is known as the dark current. Photoemissive detectors (a type of photon detector) exhibit dark currents due to thermal emission of electrons from a photosensitive material. Dark currents in photoconductive detectors (a second kind of photon detector) occur from the elevation of electrons from the valence band to the conduction band through thermal processes. In such cases, the electrical signal contains a contribution from the dark current, S_{dark}, Equation (2)

$$S(\lambda) = Q(\lambda)\, R(\lambda) + S_{\text{dark}} \tag{2}$$

Dark currents can in principle be subtracted from the total electrical signal, $S(\lambda)$. Dark current measurements are made with the entrance optics of the spectrometer closed. Over short measurement periods, S_{dark} is usually constant. Many of the commercial spectrophotometers that are equipped with detectors that produce dark currents also contain compensating circuits to minimize the magnitude of S_{dark} when measurements are made.

Gain is an important consideration when comparing the performance of different PMTs. Briefly, electrons are emitted from the photocathode surface of the PMT when it is exposed to radiation. The photoejected electrons are captured by an electrical potential and accelerated toward an electrode (dynode) coated with a compound (e.g. CuBe, BeO, GaP or CsSb) that ejects several electrons when struck by a high-energy electron. These new electrons are in turn accelerated toward another dynode (operated at a potential more positive than the preceding dynode), where the process is repeated and further multiplication occurs. Ideally, the electron gain of the PMT, G, is expressed as $G = (\Gamma)^n$, where Γ is the average number of electrons ejected at each dynode per impinging electron, and n refers to the number of dynodes inside the PMT. The exact numerical value of Γ depends upon the imposed potential difference between two successive dynodes and upon the chemical nature of the material that coats the dynode. For most commercial PMTs, Γ ranges between $\Gamma = 3$ and $\Gamma = 10$; however, several of the newer GaP coatings can achieve values as large as $\Gamma = 50$. Typical gains associated with PMTs used in commercial spectrophotometers vary from a lower limit of approximately $\Gamma \approx 10^3$ to an upper limit of about $\Gamma \approx 10^8$.

Electrical signals from the detector in a spectrometer almost always vary with time, even when the power of the incoming incident radiation is constant. Noise is defined as the unwanted fluctuation in the desired signal which obscures its accurate quantification. Noise is associated with each individual component of a spectrometer. The noise that is finally observed is a complex composite of the contributions from the different recognizable kinds of instrument noise that include thermal (or Jonson) noise, shot noise (also called Shottky or quantum noise) and flicker noise.

Thermal noise is caused by the thermal agitation of electrons or other charge carriers in resistors, capacitors, radiation detectors and other resistive elements in the spectrometer. Charge inhomogeneities are periodically created as a result of the random thermal agitation of charged particles. The charge inhomogeneities lead to voltage fluctuations which appear in the readout as unwanted noise. The magnitude of thermal noise is given by[1]

$$v_{\text{rms}} = (4\,kTR\Delta f)^{0.5} \tag{3}$$

where v_{rms} is the root-mean-square noise voltage in frequency bandwidth Δf. In Equation (3) k is the Boltzmann constant, T denotes the absolute Kelvin temperature, and R is the resistance in ohms of the resistive element. Examination of Equation (3) reveals that thermal noise can be reduced significantly by lowering the temperature of the component(s) and by lowering the circuit resistance. In practice, detector noise is often reduced by cooling. For example, lowering the temperature of a UV-visible photodiode array (PDA) from ambient room temperature to 77 K with liquid nitrogen would lead to a 50% reduction in the thermal noise.

Shot noise is encountered whenever electrons or other charged particles cross a junction. In radiation detectors such junctions are found at the pn interfaces of semiconductors, and in the evacuated space between the anode and cathode of vacuum tubes and photocells. The currents are composed of a series of quantized events, that is, the transfer of individual electrons across the junction. Under these conditions, the Poisson probability distribution applies and the statistical fluctuations in the average direct current, I, are given by[1]

$$i_{\text{rms}} = (2Ie\Delta f)^{0.5} \tag{4}$$

where i_{rms} is the root-mean-square current fluctuation. In Equation (4) e is the charge on the electron and Δf is the bandwidth of the frequency being considered. The sources of noise, as well as their mathematical treatment, are discussed in greater detail elsewhere.[1]

The S/N is mathematically defined as

$$\text{S/N} = \frac{\text{Signal}}{N_{\text{rms}}} \tag{5}$$

the ratio of the analytical signal, S, and the root-mean-square noise in the analytical signal, N_{rms}. In many spectrophotometric measurements, the overall noise level remains fairly constant and is independent of the measured electrical signal. Noise contributions thus become increasingly more important as the detector

signal decreases. The S/N provides a more meaningful measure of instrument performance than does the noise level alone. A S/N of 1000 corresponds to a relatively high precision measurement.

Multichannel charge-coupled devices (CCDs) and PDA detectors can improve the S/N for spectrophotometric measurements in two ways. First, an array detector benefits from the Fellgett advantage. The array detector monitors several wavelength regions simultaneously. The entire analysis time is spent accumulating experimental absorption/luminescence data at all wavelengths. Conventional scanning spectrophotometers, on the other hand, spend only a small fraction of the total analysis time at any one wavelength. The total analysis time is generally divided into equal time intervals. Only one time interval is spent at any given wavelength. Simultaneous wavelength detection increases the S/N by a factor of $N^{0.5}$, where N is the number of spectral regions viewed (or alternatively the number of time intervals used) during the analysis time, which is held fixed. Secondly, CCD array detectors lead to an increased S/N because the individual pixels frequently outperform PMTs. This is not the case, however, with the photodiodes. The high quantum efficiency and low read noise of today's modern solid state CCD detectors, combined with the quiet-amplification of solid state electronics, yield S/Ns similar to (or better than) those obtainable using PMTs. Moreover, both photodiode and CCD spectrophotometers have fewer movable parts, which greatly improves both wavelength accuracy and precision.[2]

Two additional quantities have been defined to describe the noise characteristics of absorption and luminescence detectors. The noise equivalent power, Φ_{noise}, is the root-mean-square power in watts that would be needed by a sinusoidally modulated input incident on the radiation detector to cause an electrical signal equal to the root-mean-square dark noise, σ_{dark}, in a 1-Hz bandwidth, i.e. $\sigma_{dark} = \Phi_{noise}(\lambda)R(\lambda)$. As indicated by the mathematical relationship, Φ_{noise} is wavelength dependent. A meaningful statement of the noise equivalent power requires that the wavelength of measurement be specified. The modulation frequency, detector area and electrical bandwidth should also be stated. Associated with noise equivalent power is the detectivity, D, which is defined as the reciprocal of Φ_{noise}. The detectivity (expressed in 1/watts) is a measure of minimum detectability.[1]

2 PHOTOCATHODE AND PHOTOSENSITIVE MATERIALS

All photon detectors use a photosensitive material to absorb the incoming incident radiation. For photoemissive photon detectors the absorbed energy leads to the emission of electrons and the subsequent development of a photocurrent that is proportional to the amount of absorbed radiation (see Equation 1). In the case of photoconductive photon detectors the absorbed radiation promotes electrons into conduction bands, with detection now based upon the resulting enhanced conductivity.

The photocathode is an important component of a photoemissive detector. Currently, two types of photocathode are used in commercial detectors. The opaque photocathode consists of a thick photoemissive material which emits electrons from the side struck by the incoming incident radiation. The transparent photocathode has the photoemissive material deposited on planar or spherical-section windows such that electrons are emitted from the backside of the photocathode opposite the incoming radiation. The spectral response of the photocathode is governed by the photoemissive material, and also by the window material in the case of the transparent photocathode. The short-wavelength limit of the transparent photocathode is determined to a large extent by the window material. Window inserts of fused silica or quartz extend the range into the far UV.

The photoemissive process responsible for converting the radiant energy into an electrical signal involves three steps. First, the photon is absorbed by the photosensitive material of the photocathode and energy is transferred from the photon to the electron. The transferred energy causes the electron to move toward the material–vacuum interface. If the electron has sufficient energy it escapes over the potential barrier at the surface into the vacuum of the tube. The energy required to release the electron from the surface corresponds to the photocathode work function, $E_{cathode}$. The energy can be used to calculate the threshold wavelength, λ_{thresh}, through the relationship λ_{thresh} (in nm) $= hc/E_{cathode} = 1240/E_{cathode}$, where h is Planck's constant and c is the speed of light. The work function represents the energy that must be given to a Fermi level electron of the metal to raise it up to the potential energy of the metal–vacuum interface. Typical work functions for pure materials range from 2–5 eV. Semiconductor materials can have substantially lower work functions.

Figure 1 depicts the radiant sensitivity (in milliamps per watt) of several types of photoemitter as a function of wavelength. Careful examination of the various curves reveals that the most sensitive photocathode compositions are the bialkali types (117 and 115), which have nominal compositions of K/Cs/Sb. Red-sensitive materials are the multialkali types (e.g. S-20 and 112 with $Cs/Na_2/K/Sb$) or Ag/O/Cs formulations (denoted as S-1). The latter photoemitter is useable up to a wavelength of about $\lambda \approx 1000$ nm. The UV end of the spectrum can be reached by coupling virtually any type of photoemitter with an optically transparent fused silica

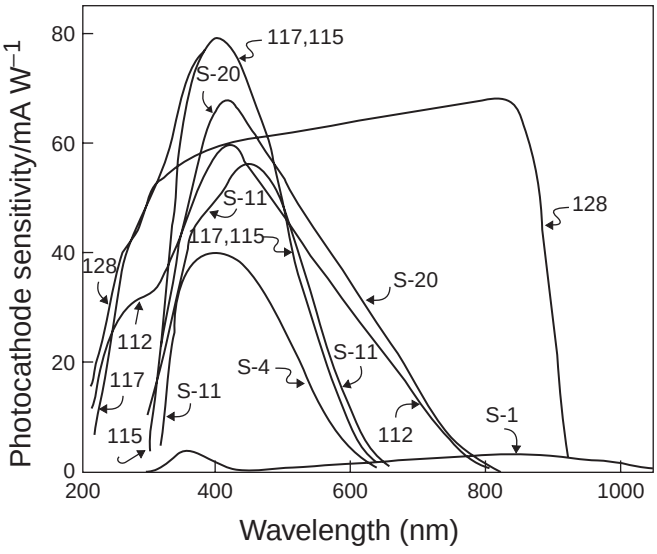

Figure 1 Spectral response curves of select photoemissive materials. Nominal compositions of the various materials are: S-1 (Ag/O/Cs), S-11 (Cs_3/Sb), S-4 (Cs_3/Sb), 117 (K/Cs/Sb); 115 (K/Cs/Sb); and 128 (Ga/As). (Redrawn with permission from F.E. Lytle, *Anal. Chem.*, **46**, 545A–557A (1974). Copyright American Chemical Society.[3])

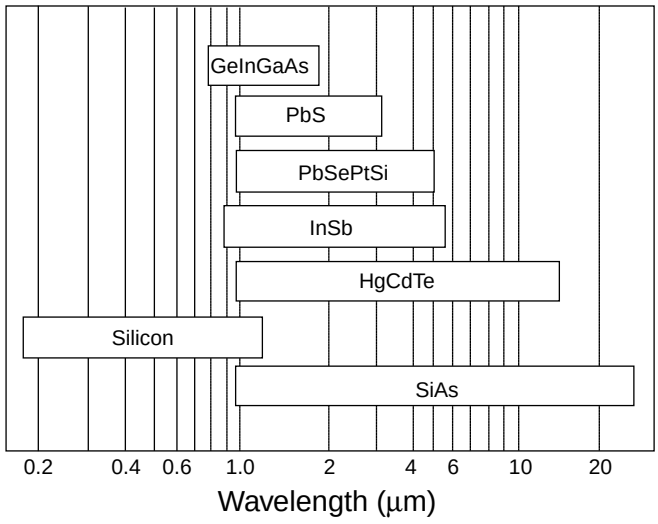

Figure 2 Approximate spectral ranges of various semiconductors used in photoconductive detectors. The actual spectral range is dependent upon temperature, detector design and semiconductor position. (Redrawn from J.M. Harnly, R.F. Fields, *Appl. Spectrosc.*, **51**, 334A–351A (1997). Copyright Society for Applied Spectroscopy.[2])

window. The Ga/As photoemitters exhibit a fairly flat spectral response from $\lambda \approx 440$ nm to $\lambda \approx 880$ nm, and can be used in spectrometers to measure radiant energy in the 200–940 nm spectral region.

Photoconductive detectors operate differently than the photoemissive detectors that were just discussed. A pn-semiconductor junction (a diode) is reversed-biased leading to the formation of a depletion layer. No current flows in the absence of electromagnetic radiation. When photons interact with the detector, electrons are promoted from the valence band to the conduction band where they act as charge carriers. In the photovoltaic mode of operation, an external circuit monitors the small potential change that develops across the semiconductor as electrons are promoted into the conduction band. The current generated by the electron–hole pairs is measured in the photoconductive mode of operation. The measured signal in both cases is proportional to the instantaneous photon flux at the pn-semiconductor junction. Figure 2 depicts the approximate spectral range of the various semiconductors used in commercial absorption and luminescence detectors. At present silicon semiconductors are used for applications in the UV and visible spectral regions. Recent technological advances in fabricating materials based upon the group 13 and group 15 nitrides may in future years lead to new semiconductor materials for the UV spectral region.[4–6]

Although most semiconductor diodes operate as photoconductive detectors, a Schottky diode functions in a similar fashion to a photoemissive detector. Schottky diodes are fabricated on silicon substrates with deposition of either iridium or platinum. The absorbed radiation causes electrons to be emitted from a metal layer having a low work function, rather than creating electron–hole pairs that would be expected in a semiconductor-type material. The emitted electrons are collected in the potential wells of neighboring associated structures.

3 PHOTOVOLTAIC OR BARRIER-LAYER CELLS

The photovoltaic cell (also called a barrier-layer cell) is a very simple absorption and luminescence detector for measuring radiation in the visible spectral region. The device consists of a flat iron or copper electrode upon which has been deposited a very thin layer of semiconducting material (e.g. selenium). A very thin transparent metallic film of silver or gold is sputtered over the semiconducting material. The metallic film acts as the collector electrode and the entire assembly is protected by a transparent lacquer coating. A typical photovoltaic cell is depicted in Figure 3.

A photovoltaic cell operates as follows. Electrons at the selenium–silver surface are excited whenever radiation of sufficient energy falls upon the semiconductor surface. Electron excitation leads to the formation of conduction electron–hole pairs. The electrons then migrate toward the deposited metallic silver (or gold) film, while the holes travel toward the iron (or copper) electrode. The liberated electrons subsequently pass through the external circuit

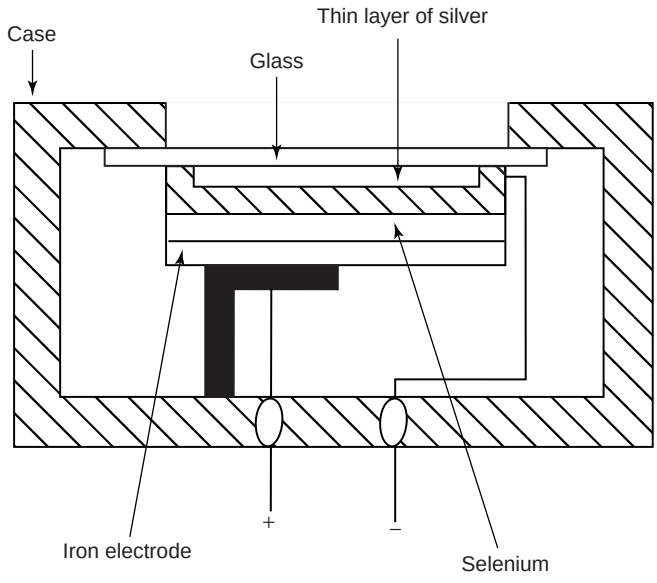

Figure 3 Schematic design of a typical photovoltaic (barrier-layer) cell.

to interact with the holes. An electrical current develops from the electron–hole movement. Photovoltaic cells typically generate currents on the order of 10–100 mA, which are large enough to be measured with a microammeter. The measured photocurrent is directly proportional to the number of photons that strike the semiconductor surface, provided that the resistance of the external microammeter circuit is sufficiently small ($<400\,\Omega$).

A photovoltaic cell (with selenium semiconductor) is primarily used in filter photometers to monitor radiation in the visible spectral region. The greatest sensitivity of the cell is for the green through yellow wavelengths. The current output of the cell is not easily amplified and this restricts the use of the cell to those applications where fairly high levels of illumination exist. High illumination levels eliminate the need to amplify the electrical signal.

The photovoltaic cell does provide an economical, low-cost means for measuring radiant power in the visible spectral region, particularly at high levels of illumination. The cell does suffer from lack of sensitivity at low levels of illumination and amplification of the electrical signal is not easily accomplished. Photovoltaic cells also experience fatigue with prolonged use. The photocurrent output decreases gradually during continued operation. Improved circuit design has helped to minimize fatigue effects.

4 VACUUM PHOTOTUBES

The vacuum phototube is an inexpensive photoemissive detector consisting of a semicylindrical cathode and a wire anode sealed inside an evacuated transparent tube. The concave surface of the cathode supports a photoemissive material (see Figure 1), which absorbs the incoming incident radiation. An electron may be released from the photoemissive material if the absorbed energy exceeds the photocathode work function. The emitted electron is directed towards the wire anode by the potential that is applied across the two electrodes. A current is generated by the resulting electron flow.

It should be noted that only a fraction of the photons striking the photoemissive material actually yield photoelectrons having sufficient kinetic energy to leave the photocathode surface. This is true even if the incoming photons have energies greater than the photocathode work function. The ratio of the number of photoelectrons ejected to the number of incident photons is referred to as the quantum efficiency, $Q_{\text{eff}}(\lambda)$. The quantum efficiency is wavelength dependent and typical values for most photoemissive materials range between $Q_{\text{eff}}(\lambda) = 0$ and $Q_{\text{eff}}(\lambda) = 0.50$. The rate at which photoelectrons are emitted from the photocathode, r_{cathode} (in s^{-1}), is found by multiplying the quantum efficiency by the photon flux of the incident radiation, $R_{\text{flux}}(\lambda)$, Equation (6)

$$r_{\text{cathode}} = Q_{\text{eff}}(\lambda)\, R_{\text{flux}}(\lambda) \tag{6}$$

Similarly, not all of the photoelectrons emitted are collected at the anode. The fraction that is collected is called the collection efficiency, CE_{anode}. The arrival rate of photoelectrons at the anode, r_{anode}, and the resulting anodic photocurrent, i_{anode}, are given by Equations (7) and (8), respectively

$$r_{\text{anode}} = CE_{\text{anode}}\, r_{\text{cathode}} = CE_{\text{anode}}\, Q_{\text{eff}}(\lambda)\, R_{\text{flux}}(\lambda) \tag{7}$$

and

$$i_{\text{anode}} = CE_{\text{anode}}\, r_{\text{cathode}}\, e = CE_{\text{anode}}\, Q_{\text{eff}}(\lambda)\, R_{\text{flux}}(\lambda) e \tag{8}$$

where e is the electron charge, $e = 1.6 \times 10^{-19}$ C. Both expressions contain wavelength-dependent quantities. If $Q_{\text{eff}}(\lambda)$ and $R_{\text{flux}}(\lambda)$ are not constant over the wavelength range incident upon the photocathode, then Equations (7) and (8) become integrals over the wavelength range. For most vacuum phototubes the anode current must be maintained lower than 10^{-5}–10^{-4} A in order to obtain a linear response.

Finally, vacuum phototubes do experience a small anodic dark current that typically ranges between 10^{-14} and 10^{-12} A. The dark current primarily results from thermal emission of electrons at the photocathode, which can occur at moderate bias voltages. Natural radioactivity from ^{40}K in the glass housing, ionization of residual gas, ohmic leakage, and field emission are additional sources of dark current in certain instances.

The theoretical treatment of vacuum phototubes is described in greater detail in a monograph devoted to spectrochemical analysis.[1]

From an operational standpoint vacuum phototubes are rarely used as fast detectors, except in those special instances that require the monitoring of low-repetition high-intensity sources. A nonexistent multiplier gain, combined with an average anode current slaved to the low cathode value, make accurate quantification of fast optical signals difficult.

5 PHOTOMULTIPLIER TUBES AND MICROCHANNEL PLATE PHOTOMULTIPLIER TUBES

The PMT is perhaps the most widely used of the different absorption and luminescence detectors. Not too many years ago PMTs were in almost every commercial UV-visible scanning and atomic absorption spectrophotometer, phosphorescence and flame emission spectrometer, and spectrofluorometer sold. Like the vacuum phototube, the PMT is a photoemissive detector which converts radiant energy directly into an electrical signal. Electrons are emitted from the photoemissive cathode surface when it is exposed to radiation. The tube contains additional electrodes (see Figure 4) called dynodes. The first dynode is maintained at a potential of 90 V or so more positive than the cathode. Electrons emitted from the photocathode are captured by the electrical potential and accelerated toward the first dynode (second electrode). Upon striking the dynode, each photoelectron causes emission of several additional electrons from the CuBe (or MgO, GaP, CsSb) alloy coating of the secondary electrode. These electrons in turn are accelerated toward the second dynode, which is at a 90 V potential more positive than the first dynode. Again, several electrons are emitted for each electron that strikes the dynode surface. The process is repeated at each of the n dynodes in the PMT. By the time that the electrons reach the anode, the original signal will have been multiplied by a factor of 10^3 (or more). Each incident photon ultimately results in 10^3 (or more) electrons being collected at the anode. Typical gains associated with PMTs used in commercial spectrophotometers vary from a lower limit of about $G \approx 10^3$ to an upper value of approximately $G \approx 10^8$.

In using PMTs, either the average current that is produced as a direct result of the arrival of the avalanched electrons or the number of anodic pulses per unit time can be measured. The latter measurement is referred to as photon counting even though charge packets, and not photons, are actually counted. The associated electron

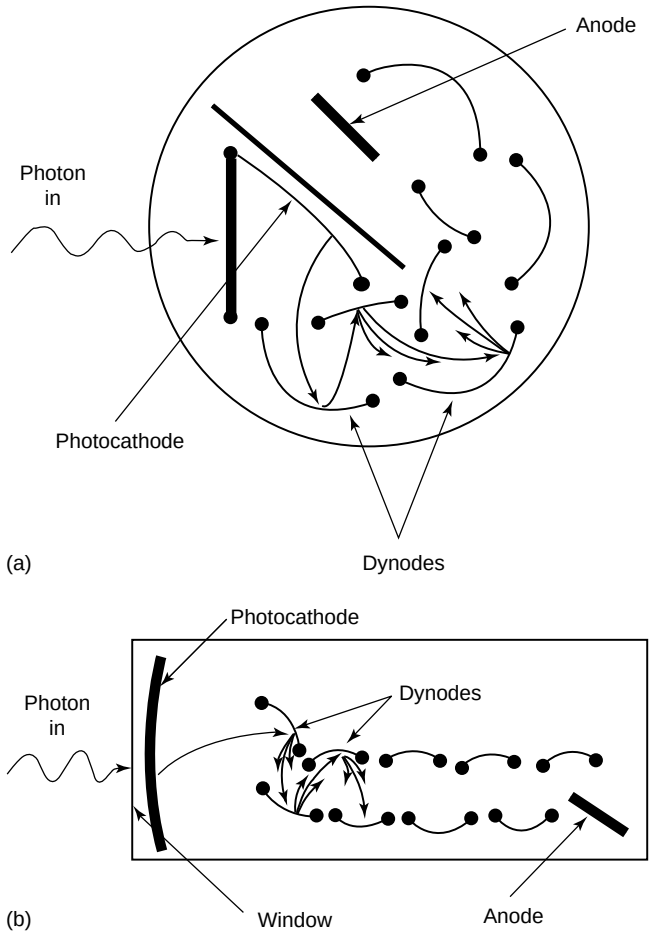

Figure 4 Structures of the side window-type (a) and linear focused head-on-type (b) PMTs. For illustrational purposes, the threefold amplification is shown for only the first few diodes.

circuits in the photon counting mode measure the arrival rate of electrons at the anode, which is given by the mathematical expression in Equation (9)

$$r_{\text{anode}} = CE_{\text{fir dyn}} r_{\text{cathode}}\, G = CE_{\text{fir dyn}}\, G \int_0^{\infty} Q_{\text{eff}}(\lambda) R_{\text{flux}}(\lambda)\, \mathrm{d}\lambda \qquad (9)$$

where $CE_{\text{fir dyn}}$ is the fraction of the photoelectrons from the photocathode collected by the first dynode giving rise to the secondary electron emission. The PMT gain is denoted by G, $Q_{\text{eff}}(\lambda)$ refers to the quantum efficiency of the cathode's photoemissive surface, and $R_{\text{flux}}(\lambda)$ is the radiation flux incident upon the photocathode. For monochromatic (or nearly monochromatic) radiation, Equation (9) reduces to Equation (7) for the vacuum phototube, except for the fact that each anodic pulse now contains G electrons. The anodic pulse of a vacuum phototube contains only a single electron as the gain is unity (i.e. $G = 1$).

Figure 4 shows the structure of both the side window circular cage-type and linear focused head-on-type PMT. Both types of PMT are used in absorbance and luminescence measurements. A lower cost tends to make the side window tubes the detector of choice for routine absorbance and for steady state fluorescence measurements. For lifetime measurements there is no clearcut choice. Side window PMTs have a much faster instrumental response than commercially available linear focused PMTs. Kinoshita and Kushida[7] obtained an instrument response of 160 ps full width at half-maximum (fwhm) to a 10 ps mode-locked laser excitation. Bebelaar[8] reported an instrument response of 235 ps fwhm with a linear focused PMT. This is one of the faster instrument responses measured on a linear focused PMT. At this width it should be possible to determine fluorescence lifetimes experimentally down to $\tau_{fluor} \approx 30$ ps. Side window PMTs, with their much faster response times, are able to measure smaller fluorescence lifetimes. It should be noted that in order to obtain the faster pulses, it is necessary to irradiate only a very small area of the side window PMT's photocathode. This requirement does restrict the sensitivity of the measurements to a level which can only be tolerated when using mode-locked laser excitation.

If instrument response time were the only consideration, then a side window PMT might be the detector of choice for time-resolved fluorescence measurements. One of the advantages afforded by linear focused PMTs is a slower transit time dependence on the point of photocathode illumination. Commercial linear focused PMTs have a transit time dispersion of about 13 ps mm^{-1}.[9] With a linear focused PMT there is no need to sacrifice sensitivity by reducing the illuminated area in order to obtain the maximum time resolution. The spectral ranges of side window and linear focused PMTs are comparable.

Determination of fluorescence lifetimes requires a minimum fast time response of 1.5 ns (or less) risetime. In PMTs the risetime originates from the distribution in anode arrival times of the avalanched electrons produced by simultaneously triggered photocathode events. A distribution in arrival times results because the ejected electrons have different energies and the electrons take slightly different pathways in traveling to the next anode. A single electron emitted at the photocathode will yield a several nanosecond distributed current pulse by the time the event is finally detected at the anode.

Manufacturers have been successful in reducing the risetime by replacing the early unfocused box-and-grid and venetian blind photocathode–anode configurations by the circular cage and in-line designs shown in Figure 4. Planar cathodes, which can have edge-to-center transit time differences as large as 10 ns, have been replaced by curved photosensitive surfaces. The two aforementioned modifications have helped to reduce the risetime of the PMT to 1–2 ns. The risetime can be further reduced by utilizing dynode materials such as GaP which emit more secondary electrons when impacted by a high-energy electron. The new GaP coatings can emit as many as 50 electrons per each electron impact, i.e. $\Gamma = 50$. This allows the total number of dynodes to be reduced, while still maintaining the high gains necessary to detect low levels of incident radiation. Several commercial PMTs employ a GaP coating for only the first dynode. For such PMTs, the gain achieved by the detector is given by $G = \Gamma_{GaP}\ (\Gamma_{other})^{n-1}$, where $n - 1$ represents the number of non-GaP dynodes in the PMT.

In addition, modern PMTs impose high electric field strengths at the dynode surfaces and use crossed electric and magnetic field strengths to control the pathway the electrons take to reach the next dynode better. The above modifications have enabled construction of PMTs having ~150 ps risetimes. From an operational point of view, the detector should have a risetime that is short compared to the variation in the signal being measured. As a rule-of-thumb, the risetime should be one-tenth (or less) of the variation in the measured signal.[3]

Fast instrument responses and fast risetimes are important considerations in detector selection for time-resolved studies. In certain instances involving single photon events it might be possible to reference measurements to some fixed point on the resultant output pulse. A fast risetime is no longer as important if a reference point is used. Detector selection then becomes governed more by the time jitter of the anodic pulse maximum for repeated equivalent events. The variation in transit time follows roughly a Gaussian distribution, with the width being one-quarter to one-third that of a delta-function response pulse. Anode jitter time can be extremely sensitive to the extent of optical coverage afforded by the photocathode.[3]

Dark currents in PMTs result from the thermal emission of electrons at the photocathode surface, ionization of residual gas, natural radioactivity from ^{40}K in the glass, ohmic leakage and thermal emission of electrons from the dynodes. The first four sources were identified previously as causes of dark current in vacuum phototubes. Dark currents in PMTs typically are in the range of 10^{-11}–10^{-7} A. Thermal emission of electrons from the photocathode (or early dynodes) are multiplied by the full gains of the PMT. The amplification leads to the much larger dark current of the PMT, compared with the 10^{-14}–10^{-12} A dark current for vacuum phototubes.

In early analytical applications PMTs were used to monitor radiation in both the UV and visible spectral regions. The majority of photoemissive materials listed in Figure 1 have their maximum sensitivity somewhere in the 200–600 nm wavelength region. In the vacuum UV, from about $\lambda \approx 30$ nm to $\lambda \approx 190$ nm, the sensitivities of the photoemissive materials in Figure 1 are generally quite

low. Radiation can be successfully detected, however, by adding a fluorescing plate (sodium salicylate on a Pyrex window) as a converter in front of the PMT. Whenever radiation strikes the added window, the compound fluoresces at about 400 nm, leading to a reasonably linear PMT signal across the cited spectral range. For many years the best IR sensitive photocathode material was Ag/O/Cs, denoted as S-1 in Figure 1. The modest quantum efficiency, $Q_{\text{eff}}(\lambda)$, of Ag/O/Cs may explain why very few red and near-IR (600–1000 nm) fluorescence measurements were made until just recently. Most photoemissive detectors lacked sensitivity in this spectral region. Red and near-IR fluorescence measurements are now possible using PMTs equipped with a In/Ga/As (Cs) photocathode, which offers improved performance out to $\lambda = 1010$ nm and a dark current 100 times smaller than the S-1 photocathode.[10]

Microchannel plate photomultiplier tubes (MCPPMTs) have been around for several years. It was not until recently, however, that the detector made an impact in fluorescence lifetime studies. Electron magnification occurs as electrons travel between thin glass plates through the many microscopic channels. Each channel is lined with a secondary emitting surface and functions as an individual electron multiplier. The glass surfaces of each plate between channels are coated with a thin nichrome (or other conducting material) layer and a large voltage (~1000 V) is applied across the thickness of the plate to direct the flow of electrons toward the anode. Figure 5 depicts a MCPPMT with three plates.

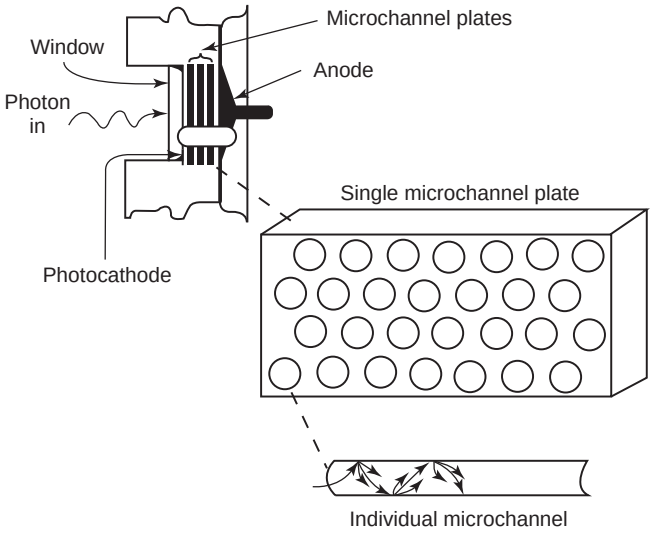

Figure 5 Schematic diagram of a MCPPMT detector (top), depicting the hexagonal arrangement of the plate's microchannels. The bottom drawing shows the electron amplification which takes place in the individual microchannels when electrons strike the secondary emitting surface.

The MCPPMT operates in the following fashion. A photoelectron emitted from the photoemissive surface of the photocathode when exposed to incident radiation is accelerated a short distance (approximately 3 mm) to the first microchannel plate (MCP). The photoelectron enters one of the channels of the plate where it eventually strikes the wall producing secondary electrons. These electrons in turn are accelerated further down the channel and additional collisions occur. On the average between one and three secondary electrons are released per collision. Upon exiting the channel, the electrons spread out and enter a number of adjacent channels in the second MCP (see Figure 5). Further multiplication occurs in the second and in each succeeding MCP. Spatial information is maintained since the secondary electrons are confined within the microchannels. After the electron cascade passes through the last plate it is finally collected at the anode as a pulse of current.

MCPPMTs are more compact than conventional PMTs, and as a result electrons travel much shorter distances. The more compact design leads to both faster time responses and less fluctuation in the length of time between the release of the first photoemissive photoelectron and its subsequent detection as an anodic current pulse. Risetimes of 150 ps and a transit-time jitter (i.e. impulse response) of approximately 25 ps fwhm at 200 counts s^{-1} noise at ambient room temperature have been recorded with a commercial 6-μm channel MCPPMT.[11] The compact design of the MCPPMT makes it amenable to cooling. Each MCP can provide, depending upon length and secondary emitting surface, electrons gains of 10^3–10^4 over an active diameter of more than 30 mm with an electron equivalent input noise of 10^{-16}–10^{-15} A cm^{-2}. When placed in tandem, as would be the case with a MCPPMT, combined MCP can give gains of $G = 10^6$ to $G = 10^7$, which are comparable to those obtained with conventional PMTs.[12]

6 PHOTODIODE AND PHOTODIODE ARRAY DETECTORS

Photodiode absorbance detectors operate differently from the vacuum phototubes and PMTs that were just discussed. In a photodiode the incoming radiation promotes electrons from the valence band to the conduction band of a semiconducting material, as opposed to releasing electrons from a photoemissive surface. A semiconductor is a crystalline material having a conductivity between that of an insulator and a conductor. Crystalline silicon and germanium are the two semiconducting materials that have found widest application in the construction of electronic devices. Both elements are in group 14 of the periodic table and thus have four valence electrons for

bond formation. All four valence electrons are localized in the silicon/germanium crystal through covalent bonding with neighboring atoms. In theory, silicon and germanium should behave as insulators since there are no free electrons in the crystalline lattice. Sufficient thermal agitation occurs at ambient room temperature, however, to liberate an occasional covalently bonded electron. The liberated electron is free to move about the crystalline lattice to conduct electricity. Thermal excitation also leaves a positively charged region where the liberated electron once resided. The positively charged region, referred to as a hole, is also mobile in that a bound electron from a neighboring Si (or Ge) atom jumps to the electron deficient region and in turn leaves behind a hole. Conduction in a semiconductor involves movement of thermal electrons in one direction and holes in the other direction.

The conductivity of a Si (or Ge) crystal can be enhanced by judiciously adding a small controlled amount of group 13 or group 15 impurity. Replacement of a silicon atom with a group 15 element (As or Sb) introduces one unbound valence electron into the structure (see Figure 6). Very little thermal energy is needed to free this extra electron for conduction of electricity. A Si (or Ge) crystal that has been doped with a group 15 element is referred to as an n-type semiconductor. Electrons are the majority charge carriers. Holes still exist as the result of thermal agitation of silicon atoms; however, their number is small compared to the number of electrons. Similarly, a p-type semiconductor is formed whenever a Si (or Ge) crystal is doped with a trivalent group 13 (In or Ga) element. An electron deficiency exists at each place where In (or Ga) is inserted. Holes are created in the crystalline lattice whenever covalently bonded electrons from neighboring Si (or Ge) atoms move to the vacant orbital associated with the group 13 impurity atom. Movement of holes from one Si (or Ge) atom to another, as described above, leads to a current in which the majority carrier is positive. A p-type semiconductor is inherently less conductive than a n-type semiconductor. Holes are less mobile than free electrons.

A photodiode is formed by having adjacent p-type and n-type regions within a single silicon or germanium crystal. Close proximity of n-type and p-type materials permits movement of holes from the p-region into the n-region and movement of electrons in the opposite direction. When the diode is reversed-biased (see Figure 7) the majority carriers in each region drift away from the junction to form a depletion layer, which contains few charges. Essentially no current flows in the absence of electromagnetic radiation. When photons strike the semiconductor, electrons are promoted from the valence band to the conduction band where they act as charge carriers. The current that is produced is proportional to the incident light intensity.

Photodiodes are used as absorbance detectors to monitor radiation in the UV-visible and near-IR spectral regions. The spectral response of most photodiodes

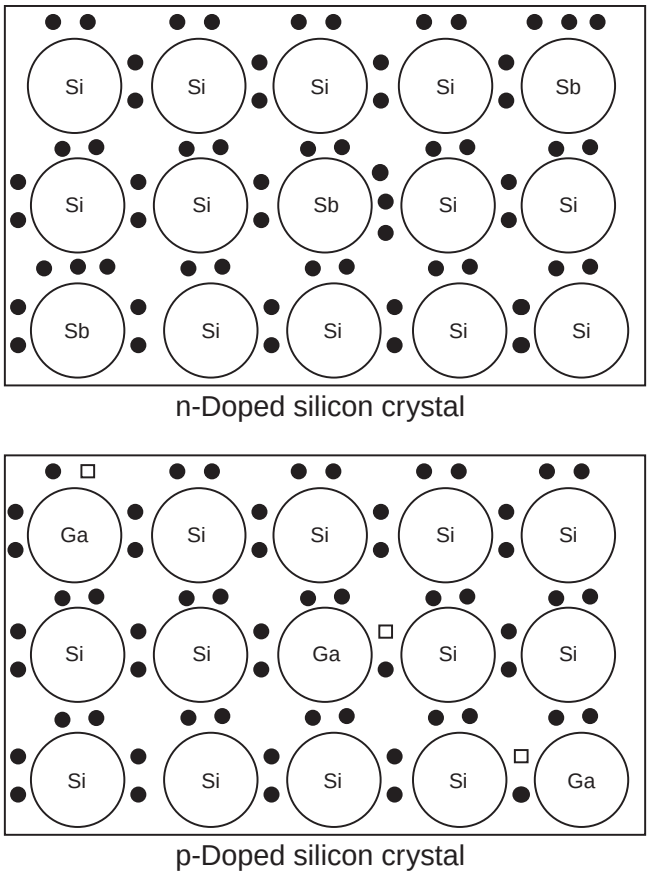

Figure 6 Representation of a n-doped (top) and a p-doped (bottom) silicon crystal. Valence electrons and holes are denoted by the solid circles (●) and open squares (□), respectively.

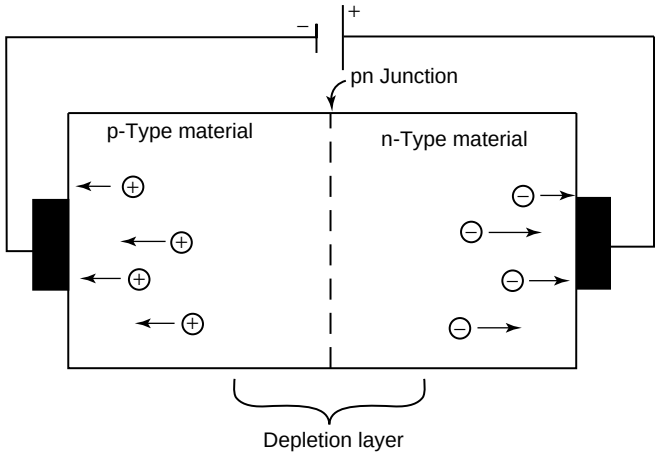

Figure 7 Schematic diagram of a silicon diode operated under reversed bias. Movement of electrons and holes away from the pn junction leads to the formation of a depletion layer.

reaches a maximum in the near-IR spectral region ($\lambda = 800-1000$ nm). These devices exhibit excellent linearity over six to seven decades of incident radiation power. Photodiodes have subnanosecond response times. Design simplicity and small size, combined with the performance characteristics listed above, have made photodiodes a very popular detector for analytical applications where light levels are relatively high.

MCPs have been used in several instances to intensify signals prior to detection by a photodiode detector. These devices operate in similar fashion to the MCPPMTs, except that the electrons upon exiting the MCP are now accelerated towards a phosphor screen. The resulting photons are coupled with a fiber-optic bundle to the photodiode. The photodiode monitors the photon bursts at the phosphor screen. MCPs are designed to provide a 10^3-10^4 gain prior to detection by a silicon photodiode. Bormett and Asher[13] reported that diffraction can pose problems when using MCP intensifiers. Diffraction redirects a fraction of the spectrophotometer-dispersed radiation back into the instrument, where it may be reflected or scattered from other optical components and later reimaged onto the detector.

Avalanche photodiodes (APDs) are used in analytical applications, such as single-photon timing and phase fluorometry,[14–24] which require both a fast response and high sensitivity. These devices are operated in the reversed-biased breakdown region of the pn junction and provide an internal gain of up to $G = 1000$. The electron–hole pairs produced by the impinging incident radiation create additional pairs in an avalanche effect. Careful control of the bias voltage is required in order to obtain the stable current gains needed for accurate measurements. Response times do depend upon the internal gain. Very fast instrumental response widths as low as 70 ps fwhm have been reported for single-photon timing experiments.

Silicon single-photon APDs provide excellent responsivity down to wavelengths of $\lambda = 1000$ nm. Germanium (Ge) and InGaAs photodiodes are used to monitor longer wavelengths (see Figure 2). These semiconductors have relatively high quantum yields; however, they do give fairly noisy signals at ambient room temperature. This is particularly true in the case of Ge semiconductors. PDA detectors used for near-IR studies are often cooled to below room temperature in order to reduce the noise. The noise problem is further addressed by minimizing the photodiode active area as much as possible. Noise is proportional to the active surface area. Areas of less than 1 mm^2 are required to achieve subpicowatt noise levels. In the case of single photon APDs, the surface areas are typically in the neighborhood of 10 μm in diameter. The small active surface area significantly reduces the detector's sensitivity, to the point where a microscope and/or high-powered laser source may be needed for accurate quantification. It should also be noted that APDs can exhibit a wavelength-dependent tail in their response due to charge carriers generated in the neutral p-region beneath the depletion layer. The tail becomes more pronounced at longer wavelengths and can hinder reconvolution analysis of single-photon timing experiments.[14]

Since the late 1970s rapid advances in semiconductor-based technology and the related development of powerful microprocessor systems led to the evolution of numerous multichannel absorption and luminescence detectors. Multichannel detectors enable the analyst to acquire a large amount of spectrally and/or spatially resolved data in a single observation. The linear photodiode array (LPDA) detector consists of an array of photodiodes (also called elements, channels and pixels). Each photodiode acts as a separate photon-to-charge transducer and a charge storage device. Linear arrays designed for commercial UV-visible spectrophotometers are spaced on 25-μm centers and have apertures of 2.5 mm. This particular design gives an aspect ratio of 100 : 1, which corresponds to the typical aspect ratio of a conventional UV-visible scanning spectrophotometer or polychromator slit. A reversed-biased pn semiconductor diode serves as each photosensitive element. Readout is performed by a shift register operating in real-time, variable integration, diode group or "skipping-diode" mode. Commercial units read out and digitize the diode information at rates from 4–28 ms per photodiode. The various readout modes are reviewed in a two part series devoted to PDA detectors in UV-visible spectroscopy.[25,26]

Figure 8 depicts a LPDA being operated under reversed bias. The PDA consists of p-type bars diffused

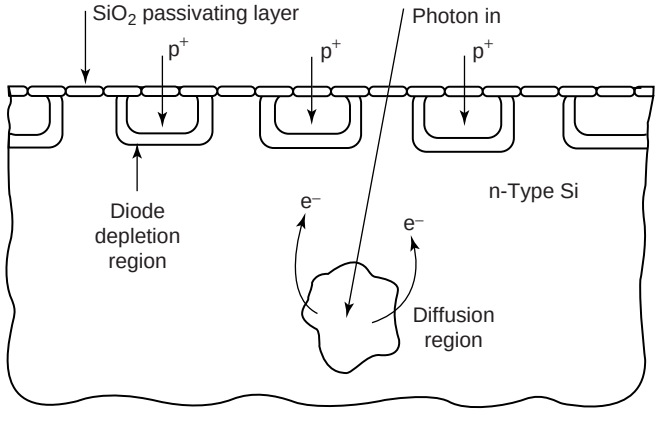

Figure 8 Diagram of the LPDA showing reversed bias pn junction with diffused p-type bars in the n-type silicon substrate. (Redrawn with permission from D.G. Jones, *Anal. Chem.*, **57**, 1057A–1073A (1985). Copyright American Chemical Society.[25])

into a n-type silicon substrate. Both semiconductor types are photosensitive. The capacitive diode is fully charged at the beginning of each measurement cycle. Electron–hole pairs form upon irradiation. Charges generated from light incident on the areas between diodes (i.e. between p-regions) divide proportionally between adjacent diodes. Photogenerated electron–hole pairs separate in the electric field, with the electrons and holes drifting to the positively charged cathode and negatively charged anode, respectively. A leakage current flows if the diode remains connected to the voltage source. This current is proportional to the photon flux incident on the diode. If the diode were to be disconnected from the voltage source, then the electron–hole pairs will discharge the potential previously stored. The photon flux incident upon the diode is found by measuring the current needed to recharge the diode to its original value, or by measuring the resulting change in potential across the diode.

PDAs are available commercially in both linear and area formats. Linear PDA are useful for monitoring radiation in the 200–1000-nm spectral region. Two-dimensional PDAs have found application in analytical chemistry for measuring radiation in the near- and mid-IR regions. Long-wavelength materials, such as HgCdTe, InSb and InGaAs (see Figure 2) allow measurements to be made down to a wavelength of $\lambda = 1800$ nm (or longer). Two-dimensional PDAs can either be used for direct wavelength detection on the focal plane of a spectrograph, or used in conjunction with a modulated radiation source to generate chemical specific maps of a sample.[2]

Parameters that govern the overall performance of PDA spectrophotometers include dynamic range and precision, spectral resolution, S/N, level of stray radiation and sensitivity. These parameters are determined to a large extent by the photodiode size and spacing. As a general rule-of-thumb, the dynamic range increases with increasing diode size, whereas greater resolution is obtained with the narrower photodiodes. The photodiode dimensions that match the spectrophotometer slit offer the best compromise for maximum resolution and dynamic range. Linear PDAs inside of commercial UV-visible scanning spectrophotometers have a spectral resolution of ± 1 nm in the UV spectral region (200–400 nm) and ± 2 nm in the visible range (400–800 nm). PDA detectors designed for spectroscopic applications have a dynamic range of about 10^4 for single integration time,[27] which is over two orders of magnitude smaller than can be obtained with a PMT. The linear dynamic range of a LPDA is in general lower than that of a single photodiode. (A single photodiode exhibits linearity over six to seven decades of incident radiation power.) The lower dynamic range results from internally reflected light from the window, the silicon wafer and spectrophotometer's metallic parts. Finally, PDAs do benefit from the Fellgett advantage in that the entire analysis time is spent making measurements at all wavelengths. The largest signal is obtained whenever the time between readouts is increased until the signal is just below saturation (i.e. integrate the signal on-chip). On-chip integration increases the S/N proportionally to the exposure time, as opposed to the square root of the accumulated exposure time in the case of conventional ensemble average.[25]

The quantum efficiency, Q_{eff}, of a PDA is quite high; however, the recharge process is quite noisy and may result in a readout noise of greater than 1000 electrons. A large readout noise significantly degrades S/Ns for weak signals. In analytical chemistry PDAs are generally used in applications, such as absorption spectroscopy, where illumination levels are high. PDA detectors (without intensifiers) are virtually useless in Raman spectroscopy because the readout noise overwhelms the weak scattering signal. Also, the dark current signal can be fairly large due to capacitor leakage, which would preclude S/N improvements based upon longer integration times. To apply the PDA to Raman spectroscopy successfully, the signal must be amplified significantly above both the readout and dark noise. MCPs have been used to boost weak Raman scattering signals for subsequent PDA detection, with only slight degradation of spatial resolution being observed. The MCP is placed prior to the PDA. Electrons exiting the MCP hit a phosphor screen and the PDA monitors the resulting photon bursts. MCP itensified PDAs have a major advantage when time resolution is required. The MCP can be switched on and off quickly. Although the PDA is integrating the signal during the entire time between readouts, the intensifier gain is so large that only the photons arriving when the MCP is on are counted. The ''contrast ratio'' between gate on and gate off signals is typically 10^6 (or larger).

Linear PDAs have completely revolutionized absorbance detection in high-performance liquid chromatographic (HPLC) separations. Analysts now obtain multiwavelength chromatograms from a single injection, which decreases the analysis time and facilitates both solute identification and quantification. In cases where the solutes are not resolved by the column but where spectral overlap is minimal, it is possible to determine the concentration of all components within a single elution profile by judiciously monitoring each component at wavelength(s) that are free of interference. One can select between the wavelength of maximum absorbance to achieve greater sensitivity, or the wavelength giving the optimum signal relative to possible interferences if greater selectivity is desired. Prior to LPDAs, several injections would be needed to obtain the same information. The output intensity from several discrete photodiodes can be summed to increase the detectability by a factor of

five-fold to 500-fold. Moreover, simultaneous data acquisition at different wavelengths also ensures the detection of most mixture components. The applicability of LPDAs in HPLC is described in several review articles.[26,28,29]

7 CHARGE TRANSFER DEVICES

PDAs do not offer the sensitivity, dynamic range, linearity and noise performance to be competitive with PMTs in many analytical applications. PDAs are used primarily in those applications where a broad spectral region needs to be monitored simultaneously (e.g. kinetic studies, HPLC detection). Here, the multichannel advantage outweighs the noise, cross talk and limited sensitivity and dynamic range shortcomings of PDAs. Recent advances in solid state and semiconductor technology have led to new multichannel detectors incorporating charge transfer devices (CTDs), which have performance characteristics comparable to commercial PMTs. Hanley et al.[30] derived mathematical formulas describing CTD performance with regard to minimum and maximum detectable absorbance, maximum S/N and maximum dynamic range for absorbance and luminescence measurements.

CTD is a generic term given to two types of related solid state radiation detectors, the CCD and the charge injection device (CID). The CCD was co-invented by Willard Boyle and George Smith in 1969. These two individuals, who were employees of Bell Laboratories, wanted to develop a silicon diode for a picturephone. They devised a method for passing charge from one capacitor to another by changing an applied voltage in a coordinated fashion. This idea led to the CCD, which was subsequently constructed and tested at Bell Laboratories. Three years later Gerry Michon and Hugh Burke, employees at General Electric Company's corporate research center, built the first CID by placing two adjacent metal-oxide semiconductor (MOS) capacitors on the edge of a wafer of an experimental device. This construction was found to be light sensitive. The historical development of both CCDs and CIDs is reviewed in greater detail elsewhere.[31]

The MOS capacitor is an important component of CTDs. A MOS capacitor consists of a gate electrode overlying insulating, epitaxy and substrate layers. The substrate is crystalline silicon (~0.5 mm thick for mechanical support) which has been highly doped with either a group 13 (p-type, CCDs) or a group 15 (n-type, CIDs) element for electrical conductivity. The gate electrodes are made of a semitransparent, highly polycrystalline silicon material. The epitaxy is a thin lightly doped region that forms the photoactive region of the MOS capacitor. The epitaxy is p-doped for CCDs and n-doped for CIDs. Insulating silicon oxide and nitride layers separate the epitaxy from the gate electrode structures. The gate and epitaxy form the two capacitor plates.

CTDs operate much like a photographic film in that they integrate signal information when irradiated. The basic operation of a CTD detector can be illustrated using Figure 9, which depicts the cross-section of a single pixel of a CTD in the charge integration mode. In this case, the pixel consists of two conductive electrodes that overlie an insulating layer of silicon oxide. The silicon oxide layer separates the electrodes from a region of n-doped silicon. Electrons are the majority charge carrier for this example. When the two electrodes are negatively charged with respect to the silicon, a charge inversion layer is created under each electrode. The mobile holes created by the absorption of photons migrate and collect under the electrodes. This region, called a potential well, is capable of holding up to 10^5–10^6 charges before overflow to an adjacent pixel (referred to as blooming). As shown in Figure 9 the right electrode is maintained at the more negative potential. Positive charge thus accumulates under the right electrode. The amount of charge in a CTD detector during exposure to radiation can be measured in one of two ways. First, in a CCD the charge is moved to a charge-sensing amplifier for measurement. Alternatively, the voltage change that results from movement of the charge from the region under the right electrode to the region under the left electrode can be measured. The latter measuring mode is used in CIDs.

Figure 10 shows how cycling of gate potential sequentially shifts a charge packet from gate to gate to move it across the CCD. Note a CCD collects photogenerated electrons, rather than holes, under the electrodes as the

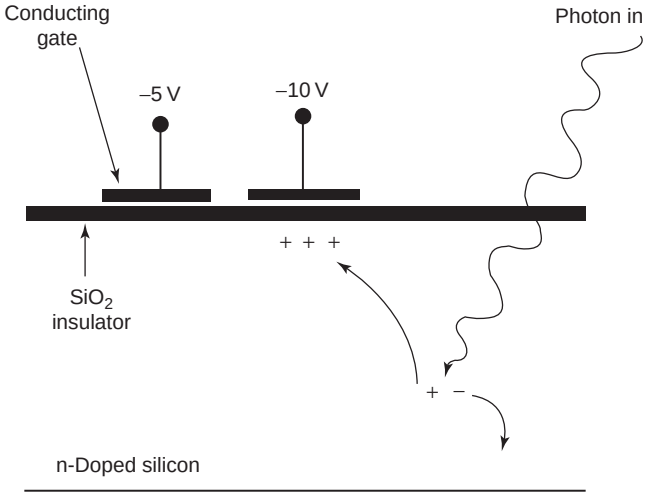

Figure 9 Cross-section of a CTD detector element when the electrodes are biased for charge integration. (Redrawn with permission from J.V. Sweedler, R.B. Bilhorn, P.M. Epperson, G.D. Sims, M.B. Denton, *Anal. Chem.*, **60**, 282A–291A (1988). Copyright American Chemical Society.[32])

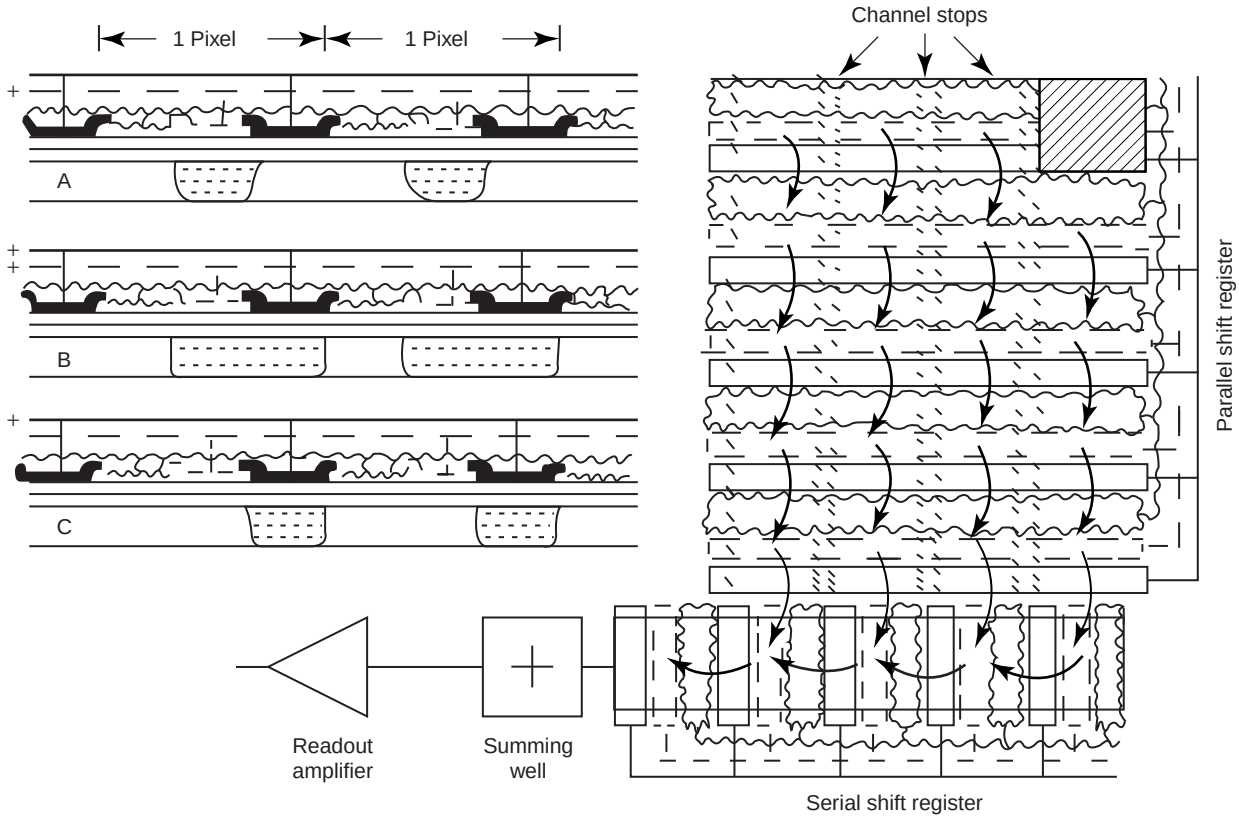

Figure 10 Pictorial diagram of a CCD. The left frame illustrates how cycling of gate potential sequentially transfers a charge packet from one gate to another. Three such transfers are needed to move the charge across a single pixel. The right frame shows how charge is transferred from pixel to pixel in the parallel register into the serial register and then to the readout amplifier. The shaded area in the upper right hand corner corresponds to one pixel. (Redrawn from J.M. Harnly, R.F. Fields, *Appl. Spectrosc.*, **51**, 334A–351A (1997). Copyright Society for Applied Spectroscopy.[2])

silicon substrate is doped with a p-type impurity. The device depicted in Figure 10 has three-gate pixels and a three-phase architecture, i.e. three transfers are required to move the charge across a single pixel. First, the charges in the vertical registers (referred to as the parallel shift register) are shifted downwards one register. The charges that were shifted into the horizontal register (also called the serial shift register) are then transferred sequentially into a summing well and then to a readout amplifier. Once the serial register is emptied, the entire process begins again with another vertical shift, and continues until all of the pixels are emptied.

Charge blooming is the leakage of accumulated charge from an illuminated detector element to adjacent elements. The onset of blooming in a CCD occurs whenever the maximum number of photogenerated charge carriers that can be held in an element is exceeded. Channel stops in the parallel shift register prevent charge blooming across columns and as a result the bloomed charge spills down the column. Manufacturers now offer CCDs with antiblooming drains, which accumulate and transport away the excess photogenerated charges. These CCDs generally exhibit slightly lower quantum efficiencies. Part of the area is taken up by the antiblooming drain.

CCDs are currently being used in UV-visible absorption, fluorescence and Raman spectroscopies, where measurements might require the long integration of low intensity radiation. Epperson et al.[33] reported a detection limit of anthracene of 1 ppt at S/N = 2 based upon fluorescence measurements with a relatively weak 35-W deuterium lamp excitation source. The reported detection is an order of magnitude lower than that determined using a high-powered tunable dye laser excitation source.[34] Murray and Dierker[35] studied monolayers of organic sorbates of different substrates by unenhanced Raman spectroscopy. The authors reported a S/N = 15 when measuring the monolayer coverage of cadmium stearate films deposited on glass and silicon surfaces with a CCD. Zhao and McCreery[36] successfully demonstrated the applicability of a CCD as a radiation for Fourier transform (FT) Raman spectroscopy in the near-IR region. Measurements revealed that the common path interferometer-CCD spectrometric system possessed many of the advantages inherent to FT

spectroscopy (e.g. high throughput, excellent frequency precision and wide spectral coverage) and CCD detection (shot-noise limited operation). The observed S/N of the multichannel FT system was comparable to that for a dispersive spectrometer for the same laser power and integration times.

Modified CCDs have been used in conjunction with an echelle grating inductively coupled plasma emission spectrometer to study samples containing up to 72 different elements. The detector, called a segmented array charge-coupled device (SCD), consisted of 224 linear CCDs imbedded into a monolithic silicon block. The light beam originating from the torch is split by a Schmidt cross-disperser. Cross-dispersed radiation is focused onto an SCD for intensity measurements in the UV spectral region (160–375 nm). The undispersed radiation is passed through a prism for cross-dispersion and subsequently focused onto a second SCD for intensity quantification in the visible (375–782 nm). Each array segment can be addressed separately and the charge integration time can be varied over a wide enough range to provide dynamic ranges of 10^5 or greater.[37,38]

CIDs operate in a similar fashion, except for the method of measuring the charge accumulated under the electrode. Each pixel of a CID is defined by the intersection of a row and column gate electrodes as shown in Figure 11. At the beginning of a measurement period negative potentials are applied to both electrodes, which leads to the formation of potential wells that collect and store holes formed in the n-doped silicon substrate by absorption of photons. Holes preferentially migrate to the collection gate because of its more negative potential. Excess charge rarely spills over into adjacent elements. After integration (or the end of the measurement time), each pixel is read using correlated double sampling. The potential gate for the sense gate is disconnected and the potential read (step B in Figure 11). The holes that have accumulated under the collection gate are then transferred to the sense gate by changing the potential applied to the collection gate from negative to positive. The new potential of the sense gate is measured. The difference between the two potential readings is proportional to the charge collected in the well. After the pixel is read, the charge may be either injected into the silicon substrate by raising the gate potential and collapsing the well (step D in Figure 11), or returned back to the collection gate by restoring its original negative potential in a nondestructive readout (NDRO). The NDRO is often used in CIDs to reduce random noise and to extend the dynamic range of the sensor. A major advantage that CIDs have over CCDs is that successive measurements can be made while integration is still taking place.

Published articles[39–41] have also reported that two-dimensional CID array detectors can be used in an echelle

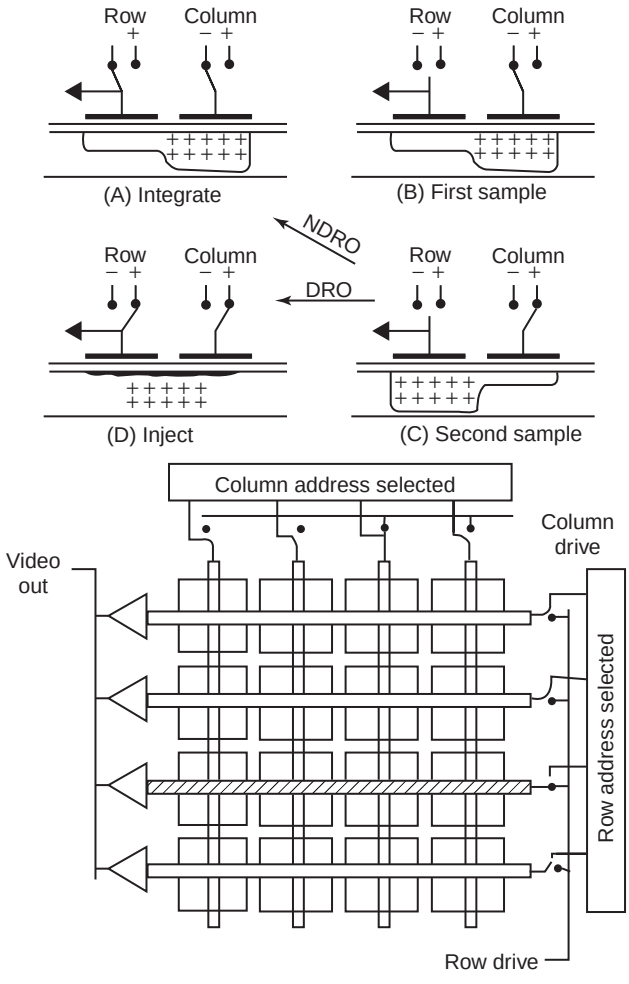

Figure 11 Pictorial representation illustrating the operation of a CID. Top frames A–C show how the sense and collection gates accumulate charge for both a destructive readout (DRO) and NDRO. The bottom frame shows the random access of pixels with column and row gates. (Redrawn from J.M. Harnly, R.F. Fields, *Appl. Spectrosc.*, **51**, 334A–351A (1997). Copyright Society for Applied Spectroscopy.[2])

spectrometer for direct current and inductively coupled plasma atomic emission spectroscopy. Dark currents in the detector elements are minimized by housing the unit in a liquid nitrogen cryostat maintained at a temperature of 135 K. Two-dimensional CID arrays possess the necessary number of resolution elements for measuring hundreds of emission lines and surrounding background. Analysis of complex samples requires that several emission lines and associated background regions be measured for each element in order to reduce errors arising from matrix effects. The preceding examples describe only a few of the many published applications where either a CCD or CID was used. CTDs have been used as absorbance and luminescence detectors in nearly all areas of analytical chemistry including molecular luminescence,[33,42] atomic

emission spectroscopy,[37–41] Raman spectroscopy,[35,36] X-ray diffraction,[43,44] microscopy[45] and chemical separations.[46]

8 STREAK TUBES AND STREAK CAMERAS

The streak camera is a device that converts time information from a luminous event into spatial information. Streak cameras are capable of measuring fast optical signals. There are two general types of streak cameras, the mechanical and the image convertor. In the mechanical version the signal is swept quickly across a piece of photographic film by a high-speed rotating mirror. Typical performance characteristics would be 2–100-ns resolution and a 15–800-ms time window.[3] In the image convertor (Figure 12) electrons emitted from the photocathode upon irradiation are extracted by the grid and accelerated at high velocity toward the anode. After passing through a hole in the anode the electrons are deflected by the voltage ramp applied to the deflection plate. The voltage ramp increases in strength with increasing time. A MCP intensifier may be used to multiply the number of electrons prior to detection. Upon exiting the MCP, the electrons strike a phosphor screen. The result of the aforementioned processes is that the time dependence of the radiation intensity is translated into the distance dependence of electron impact upon the phosphor screen. The shape of the voltage ramp used to deflect the electrons determines the relationship between time and impact distance. Photon bursts from the phosphor screen can be photographed with a normal camera, or sent to a CCD for quantification. Image convertors have time resolutions of 0.1–20 ps. The main advantage that streak cameras have enjoyed over photon counting methods has been time resolution. Streak camera time resolutions of 0.5 ps fwhm (or less) are obtainable. The best time-resolved single-photon counting methods, employing fast MCPPMTs, are about two orders of magnitude longer in fwhm.

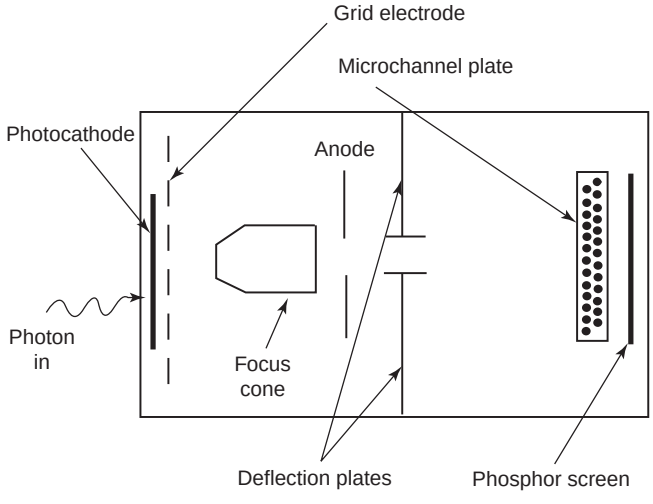

Figure 12 Diagram of a streak tube, which is the heart of the streak camera system. The MCP (not required) multiplies the number of electrons prior to their impact on the phosphor screen. A streak camera system would also include the camera or CCD for recording the photon bursts.

Many of the early streak cameras were plagued by time and intensity nonlinearity, flyback artifacts and a limited dynamic range. Manufacturers have now overcome many of these problems with improved designs. Streak cameras were designed to measure the time dependence of light emitted by a sample. Several of the processes that can be monitored include: protein binding reactions (0.1 ps to 1 ms); energy transfer in proteins (<1 ps to 100 ps); rotational diffusion (<1 ps to >1 ns); excited state proton transfer (<10 ps to >100 ps); electronic relaxation (<1 ps to 100 ns); and exciplex formation (<10 ps to 10 ns).[47,48] The corresponding timescales, picoseconds (ps) to milliseconds (ms), are given in parentheses. Several of these processes can also be monitored with absorption spectroscopic methods.

9 THERMAL DETECTORS

Thermal detectors, which are used in the IR region, sense the change in temperature that results from the absorption of incoming incident radiation. Thermal detectors are made of low heat capacity materials as the radiant power level from a typical IR beam is small (in the range of 10^{-9}–10^{-7} W). The size and thickness of the absorbing element is minimized as much as possible and every effort is made to focus the entire IR beam onto the detector surface. Thermal detectors are housed in a vacuum and carefully shielded inside the spectrometer. The beam from the radiation source is generally chopped to minimize further the undesired efforts of extraneous heat. The analyte signal would then have the frequency of the chopper and could easily be distinguished from extraneous noise, which would vary slowly with time.

The first of the thermal detectors to be considered is the thermocouple, which consists of a pair of junctions formed whenever two identical pieces of metal (e.g. Bi) are fused to each end of a dissimilar metal (e.g. Sb). One junction serves as a reference of "fixed" temperature, while the second junction becomes the sensor for IR radiation. (Note, the reference junction does not need to be maintained at constant temperature if the analytical signal is chopped.) The sensing transducer junction is normally blackened to improve its heat-absorbing capacity and sealed in an evacuated chamber

having a window made from a material that is optically transparent to IR radiation. The reference junction is designed to possess a relatively large heat capacity and is carefully shielded from the incident radiation. A potential develops between the two junctions that varies with the difference in temperature of the junctions. A well-designed thermocouple IR detector is capable of responding to temperature differences as small as 10^{-6} K. The thermocouple detector has uniform spectral response in the 1–40-μm wavelength region, excellent linearity and reasonable sensitivity. The detector's sensitivity can be further enhanced by connecting several thermocouples in series. Now, the output potential is the sum of the potentials across the individual thermocouples. Multiple thermocouple devices are referred to as thermopiles. Thermocouple detectors are used in dispersive mid-IR spectrometers.

IR radiation can also be detected with thermistors. These detectors are constructed from intrinsic semiconductor materials that exhibit relatively large changes in resistance as a function of temperature. The responsive element is again kept small and blackened to absorb radiation. The thermistor is generally placed in a bridge circuit (e.g. Wheatstone bridge) with a reference thermistor that is not irradiated. The resistance can be measured by a null-comparison technique, or the out-of-balance voltage of the bridge can be monitored. Thermistors exhibit moderate sensitivity and a wide linear range.

Thermocouples and thermistors respond considerably slower than PMTs, CCDs, CIDs and photodiodes because time is required for the temperature of the element to change. Typical response times for thermocouples and thermistors are in the range of 0.01–0.1 s. Such response times are much too slow for FT spectroscopy. Commercial Fourier transform infrared (FTIR) spectrometers use pyroelectric or Golay detectors, which exhibit responses that are fast enough to track the changes in the time-domain signal from an interferometer.

Pyroelectric detectors are constructed from single crystalline wafers of pyroelectric materials, such as triglycine sulfate (usually deuterated with a fraction of the glycines replaced with alanine), lithium niobate ($LiNbO_3$) or lithium tantalate ($LiTaO_3$). Pyroelectric substances exhibit a large spontaneous electrical polarization at temperatures below their Curie point, which in the case of triglycine sulfate is 47 °C. ($LiNbO_3$ and $LiTaO_3$ have higher Curie points and do not exhibit hygroscopicity, which is sometimes a problem encountered with triglycine sulfate.) If the temperature of a pyroelectric material is changed, then the degree of polarization is changed. The change in polarization can be quantified by placing the pyroelectric crystal between two electrodes to form a capacitor. When the polarization changes, the charges induced on the electrodes create a measurable current in an external electric circuit that connects the two sides of the capacitor. The output current is proportional to the surface area of the crystal and to its rate of change of polarization with temperature. Pyroelectric detectors have fairly fast response of less than 1 ms over most of the IR region encountered in analytical chemistry. For wavelengths below 2 μm, the triglycine sulfate must be blackened to enhance photon absorption. This can slow the detector's response. Linear arrays of pyroelectric detectors are also available.

It should also be noted that thermocouples, thermopiles and pyroelectric devices have also been utilized in conjunction with electrical substitution radiometers in absolute photon flux measurements.[49] The detector is a thermal sensor in contact with a good absorber of radiation and an electric heater. Photons are absorbed by the thermal sensor and the detector output, $T_{\text{radiation}}$, is read. The radiation source is then shuttered to prevent light from reaching the detector. Once the sensor returns to its original temperature electric power is applied to the heater until the detector's output reaches $T_{\text{radiation}}$. Measurement of the electric power when $T_{\text{electric}} = T_{\text{radiation}}$ yields the equivalent of the radiant power absorbed. Sensitivities of electrically calibrated pyroelectric radiometers range from 100 mW down to 100 μW full scale.

Golay detectors can be used in FTIR spectrometers having slow-scanning and/or stepped-scan interferometers. Figure 13 depicts a Golay cell, which is essentially an airtight pneumatic chamber filled with a gas of low thermal conductivity (e.g. xenon). The chamber is closed at one end by a thin rigid blackened metal plate and by a flexible silvered diaphragm at the other end. Radiation passes through a small IR-transmitting window and is absorbed by the blackened plate. Heat, conducted to the gas, causes it to expand and distort the flexible

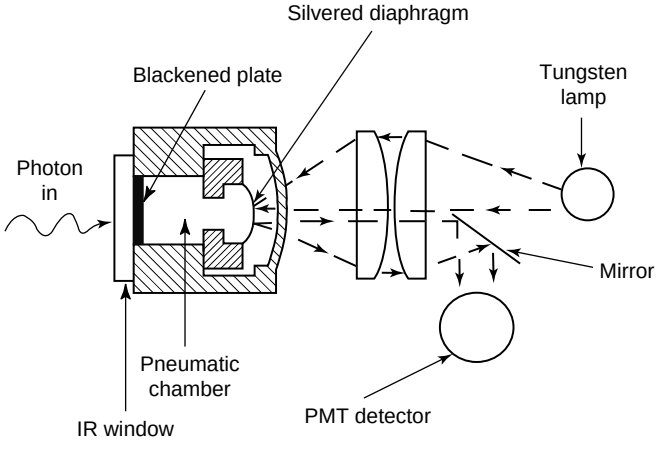

Figure 13 Schematic diagram of a Golay cell.

diaphragm. A fine leak connects the detection chamber to a ballasting reservoir of gas to prevent ambient room temperature changes from affecting the measurement. Light from a tungsten filament lamp is focused upon the silvered diaphragm and is reflected back to a PMT (or other suitable radiation detector). Distortion of the silvered diaphragm moves the light beam across the PMT surface and changes its electrical output. Golay detectors are considerably more sensitive than thermocouples. Response times are typically on the order of 20 ms. Golay detectors may be used only with slow-scanning or stepped-scan interferometers, and where the IR radiation is modulated by an external chopper. Radiation modulated in the audiofrequency regime must be detected by either pyroelectric detector or semiconductor detector (e.g. PbS, PbSe and HgCdTe).[50] As noted previously, semiconductor detectors can be operated in either the photovoltaic or photoconductive model. The approximate spectral ranges of the various semiconductors used in commercial absorption and luminescent detectors are given in Figure 2. This list will undoubtedly grow in number in future years as new semiconducting materials are produced and tested.

ABBREVIATIONS AND ACRONYMS

APD	Avalanche Photodiode
CCD	Charge-coupled Device
CID	Charge Injection Device
CTD	Charge Transfer Device
FT	Fourier Transform
FTIR	Fourier Transform Infrared
fwhm	full width at half-maximum
HPLC	High-performance Liquid Chromatographic
IR	Infrared
LPDA	Linear Photodiode Array
MCP	Microchannel Plate
MCPPMT	Microchannel Plate Photomultiplier Tube
MOS	Metal-oxide Semiconductor
NDRO	Nondestructive Readout
PDA	Photodiode Array
PMT	Photomultiplier Tube
SCD	Segmented Array Charge-coupled Device
S/N	Signal-to-noise Ratio
UV	Ultraviolet

RELATED ARTICLES

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Lifetime Measurements, Applications of • Near-infrared Absorption/Luminescence Measurements • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

Raman Spectroscopy **(Volume 15)**
Dispersive Raman Spectroscopy, Current Instrumental Designs

REFERENCES

1. J.D. Ingle, Jr, S.R. Crouch, *Spectrochemical Analysis*, Prentice Hall, Englewood Cliffs, New Jersey, 1988.
2. J.M. Harnly, R.E. Fields, 'Solid-state Array Detectors for Analytical Spectrometry', *Appl. Spectrosc.*, **51**, 334A–351A (1997).
3. F.E. Lytle, 'Measuring Fast Optical Signals: Detectors', *Anal. Chem.*, **46**, 545A–557A (1974).
4. M.A. Khan, J.N. Kuznia, D.T. Olson, J.M. Van Hove, M. Blasingame, L.F. Reitz, 'High-responsivity Photoconductive Ultraviolet Sensors Based on Insulating Single-crystal GaN Epilayers', *Appl. Phys. Lett.*, **60**, 2917–2919 (1992).
5. Q. Chen, M.A. Khan, C.J. Sun, J.W. Yang, 'Visible-blind Ultraviolet Photodetectors Based on GaN p-n Junctions', *Electron. Lett.*, **31**, 1781–1782 (1995).
6. M.S. Shur, M.A. Khan, 'Optoelectronic GaN-based Field Effect Transistors', *Proc. SPIE*, **2397**, 294–303 (1995).
7. S. Kinoshita, T. Kushida, 'High-performance, Time-correlated Single Photon Counting Apparatus Using a Side-on Type Multiplier', *Rev. Sci. Instrum.*, **53**, 469–472 (1982).
8. D. Bebelaar, 'Time Response of Various Photomultipliers and its Wavelength Dependence in Time-correlated Single-photon Counting with an Ultimate Resolution of 47 ps fwhm', *Rev. Sci. Instrum.*, **57**, 1116–1125 (1986).
9. F. Calligaris, P. Ciuti, I. Gobrielli, R. Giacomich, R. Mosetti, 'Wavelength Dependence of Timing Properties of the XP2020 Photomultiplier', *Nucl. Instrum. Methods*, **157**, 611–613 (1978).
10. R.B. Thompson, 'Red and Near-infrared Fluorometry', in Topics *in Fluorescence Spectroscopy: Volume 4: Probe Design and Chemical Sensing*, ed. J.R. Lakowicz, Plenum Press, New York, 151–181, 1994.
11. H. Kume, T. Taguchi, K. Nakatsugawa, K. Ozawa, S. Suzuki, R. Samuel, Y. Nishimura, I. Yamazaki, 'Compact Ultrafast Microchannel Plate Photomultiplier Tubes', in *Time-resolved Laser Spectroscopy in Biochemistry III*, ed. J.R. Lakowicz, *Proc. SPIE*, **1640**, 440–447, 1992.
12. Y. Talmi, 'TV-Type Multichannel Detectors', *Anal. Chem.*, **47**, 699A–709A (1975).
13. R.W. Bormett, S.A. Asher, '2-D Light Diffraction from CCD and Intensified Reticon Multichannel Detectors

Causes Spectrometer Stray Light Problems', *Appl. Spectrosc.*, **48**, 1–6 (1994).
14. S. Cova, R. Ripamonti, A. Lacaiti, 'Avalanche Semiconductor Detector for Single Optical Photons with a Time Resolution of 60 ps', *Nucl. Instrum. Methods Phys. Res.*, **A253**, 482–487 (1987).
15. S. Cova, A. Longoni, A. Andreoni, R. Cubeddu, 'A Semiconductor Detector for Measuring Ultraweak Fluorescence Decays with 70 ps fwhm Resolution', *IEEE J. Quantum Electron.*, **QE-19**, 630–634 (1983).
16. R.G.W. Brown, K.D. Ridley, J.R. Rarity, 'Characterization of Silicon Avalanche Photodiodes for Photon Correlation Measurements. 1: Passive Quenching', *Appl. Opt.*, **25**, 4122–4126 (1986).
17. S. Cova, A. Longoni, G. Ripamonti, 'Active-quenching and Grating Circuits for Single-photon Avalanche Diodes (SPADS)', *IEEE Trans. Nucl. Sci.*, **NS-29**, 599–601 (1982).
18. R.G.W. Brown, R. Jones, J.R. Rarity, K.D. Ridley, 'Characterization of Silicon Avalanche Photodiodes for Photon Correlation Measurements. 2: Active Quenching', *Appl. Opt.*, **26**, 2383–2389 (1987).
19. H. Dautet, P. Deschamps, B. Dion, A.D. MacGregor, D. MacSween, R.J. McIntyre, C. Trottier, P.P. Webb, 'Photon Counting Techniques with Silicon Avalanche Photodiodes', *Appl. Opt.*, **32**, 3894–3900 (1993).
20. S. Cova, A. Lacaita, M. Ghioni, G. Ripamonti, T.A. Louis, '20ps Timing Resolution with Single-photon Avalanche Diodes', *Rev. Sci. Instrum.*, **60**, 1104–1110 (1989).
21. T.A. Louis, G. Ripamonti, A. Lacaita, 'Photoluminescence Lifetime Microscope Spectrometer Based on Time-correlated Single-photon Counting with an Avalanche Diode Detector', *Rev. Sci. Instrum.*, **61**, 11–22 (1990).
22. A. Lacaita, S. Cova, M. Ghioni, 'Four-hundred-picosecond Single Photon Timing with Commercially Available Avalanche Photodiodes', *Rev. Sci. Instrum.*, **59**, 1115–1121 (1988).
23. W. Fichtner, W. Hacker, 'Time-resolution of Ge Avalanche Photodiodes Operating as Photon Counters in Delayed Coincidence', *Rev. Sci. Instrum.*, **47**, 374–377 (1976).
24. A. Lacaita, S. Cova, F. Zappa, P.A. Francese, 'Subnanosecond Single-photon Timing with Commercially Available Germanium Photodiodes', *Opt. Lett.*, **18**, 75–77 (1993).
25. D.G. Jones, 'Photodiode Array Detectors in UV-Vis Spectroscopy: Part I', *Anal. Chem.*, **57**, 1057A–1073A (1985).
26. D.G. Jones, 'Photodiode Array Detectors in UV-Vis Spectroscopy: Part II', *Anal. Chem.*, **57**, 1207A–1214A (1985).
27. Y. Talmi, R.W. Simpson, 'Self-scanned Photodiode Array: A Multichannel Spectrometric Detector', *Appl. Opt.*, **19**, 1401–1414 (1980).
28. R.L. Stevenson, 'UV-Vis Absorption Detectors for HPLC', *Chromatogr. Sci.*, **23**, 23–86 (1983).
29. H. Oppe, 'Detectors Based on the Measurement of Optical and Electrical Properties of the Mobile Phase', *J. Chromatog. Lib.*, **13**, 113–129 (1978).
30. Q.S. Hanley, C.W. Earle, F.M. Pennebaker, S.P. Madden, M.B. Denton, 'Charge-transfer Devices in Analytical Chemistry', *Anal. Chem.*, **68**, 661A–667A (1996).
31. G.R. Sims, 'A Brief History of Charge-transfer Devices', in *Charge-Transfer Devices in Spectroscopy*, eds. J.V. Sweedler, K.L. Ratzlaff, M.B. Denton, VCH, New York, 1–7, 1994.
32. J.V. Sweedler, R.B. Bilhorn, P.M. Epperson, G.R. Sims, M.B. Denton, 'High-performance Charge Transfer Device Detectors', *Anal. Chem.*, **60**, 282A–291A (1988).
33. P.M. Epperson, J.V. Sweedler, R.B. Bilhorn, G.R. Sims, M.B. Denton, 'Applications of Charge Transfer Devices in Spectroscopy', *Anal. Chem.*, **60**, 327A–335A (1988).
34. J.H. Richardson, M.E. Ando, 'Sub-part-per-trillion Detection of Polycyclic Aromatic Hydrocarbons by Laser Induced Molecular Fluorescence', *Anal. Chem.*, **49**, 955–959 (1977).
35. C.A. Murray, S.B. Dierker, 'Use of an Unintensified Charge-coupled Device Detector for Low-light-level Raman Spectroscopy', *J. Opt. Soc. Am. A*, **3**, 2151–2159 (1986).
36. J. Zhao, R.L. McCreery, 'Multichannel Fourier Transform Raman Spectroscopy: Combining the Advantages of CCDs with Interferometry', *Appl. Spectrosc.*, 1209–1214 (1996).
37. T.W. Barnhard, M.I. Crockett, J.C. Ivaldi, P.L. Lundberg, 'Design and Evaluation of an Echelle Grating Optical System for ICP-OES', *Anal. Chem.*, **65**, 1225–1230 (1993).
38. T.W. Barnhard, M.I. Crockett, J.C. Ivaldi, P.L. Lundberg, D.A. Yates, P.A. Levin, D.J. Sauer, 'Solid-state Detector for ICP-EOS', *Anal. Chem.*, **65**, 1231–1239 (1993).
39. M.J. Pilon, M.B. Denton, R.G. Schleicher, P.M. Moran, S.B. Smith, Jr, 'Evaluation of a New Array Detector Atomic Emission Spectrometer for Inductively Coupled Plasma Atomic Emission', *Appl. Spectrosc.*, **44**, 1613–1620 (1990).
40. G.R. Sims, M.B. Denton, 'Multielement Emission Spectrometry Using a Charge-injection Device Detector', in *Multichannel Image Detectors*, ed. Y. Talmi, ACS Symposium Series 236, American Chemical Society, Washington DC, 117–132, 1983.
41. J.V. Sweedler, 'Charge Transfer Device Detectors and Their Application to Chemical Analysis', *Crit. Rev. Anal. Chem.*, **24**, 59–98 (1993).
42. C.E. Hooper, R.E. Ansorge, J.G. Rushbrooke, 'Low-light Imaging Technology in the Life Sciences', *J. Biolumin. Chemilumin.*, **9**, 113–122 (1994).

43. W. Phillips, M. Stanton, D. O'Mara, Y. Li, I. Naday, E. Westbrook, 'CCD-based Detector for Crystallographic Applications Using Laboratory X-ray Sources', *Proc. SPIE-Int. Soc. Opt. Eng.*, **2009**, 113–138 (1993).
44. I. Naday, E.M. Westbrook, M.L. Westbrook, D.J. Travis, M. Stanton, W.C. Phillips, D. O'Mara, J. Xie, 'Characterization and Data Collection on a Direct-coupled CCD X-ray Detector', *Nucl. Instrum. Methods Phys. Res.*, **A348**, 635–640 (1994).
45. T.W.J. Gadella, T.M. Jovin, R.M. Clegg, 'Fluorescence Lifetime Imaging Microscopy (FLIM): Spatial Resolution of Microstructures on the Nanosecond Time Scale', *Biophys. Chem.*, **48**, 221–239 (1993).
46. L.B. Koutny, E. Yeung, 'On-line Detection of Proteins in Gel Electrophoresis by Ultraviolet Absorption and by Native Fluorescence Utilizing a Charge-coupled Device Imaging System', *Anal. Chem.*, **65**, 183–187 (1993).
47. T.M. Nordland, 'Streak Camera for Time-domain Fluorescence', in *Topics in Fluorescence Spectroscopy: Volume 1: Techniques*, ed. J.R. Lakowicz, Plenum Press, New York, 183–260, 1991.
48. Y. Tsuchiya, 'Advances in Streak Camera Instrumentation for the Study of Biological and Physical Processes', *IEEE J. Quantum. Electron.*, **QE-20**, 1516–1528 (1984).
49. E.F. Zalewski, 'Radiometry: New Developments Relating to Spectrofluorometry and Photon Yield Measurements', in *Optical Radiation Measurements*: Academic Press, New York, 195–248, Vol. 3, 1982.
50. P.R. Griffiths, J.A. de Haseth, *Fourier-transform Infrared Spectrometry*, John Wiley and Sons, New York, 209–217, 1986.

Fluorescence Imaging Microscopy

John J. Lemasters
University of North Carolina at Chapel Hill, Chapel Hill, USA

1 Introduction to Optical Microscopy **1**
1.1 Lateral Resolution 1
1.2 Specimen Preparation for Brightfield Microscopy 1
2 Fluorescence Optical Microscopy **2**
2.1 Fluorescence and Phosphorescence 2
2.2 Epi-illumination 2
2.3 Confocal Fluorescence Microscopy 2
2.4 Multiphoton Excitation Fluorescence Microscopy 4
3 Fluorescence Labeling **5**
3.1 Autofluorescence 5
3.2 Immunofluorescence 5
3.3 Parameter-indicating Fluorophores in Living Cells and Tissues 6
4 Conclusions **11**
Abbreviations and Acronyms **11**
Related Articles **12**
References **12**

In fluorescence microscopy, an optical microscope creates a magnified image of the fluorescence and phosphorescence of specimens excited by light. Autofluorescence is the intrinsic fluorescence of a specimen, which is often weak. To enhance fluorescence, tags are added, such as fluorescently labeled antibodies which precisely localize specific molecular components. Living cells and tissues may be labeled with a variety of parameter-indicating fluorophores to measure ion concentrations, electrical potentials and specific chemical species. Thick living specimens create problems of superimposition and haze from out-of-focus fluorescence, but advances in confocal fluorescence microscopy and multiphoton fluorescence microscopy now permit successful fluorescence imaging of thick living specimens with submicrometer three-dimensional resolution.

1 INTRODUCTION TO OPTICAL MICROSCOPY

1.1 Lateral Resolution

Optical microscopy provides useful magnification to about 2000 times. The diameter of the smallest object resolvable by an optical microscope, r, is given by the Abbe equation [Equation 1]:

$$r = \frac{0.6\lambda}{n \sin \alpha} \quad (1)$$

where λ is the wavelength of light, n is the refractive index of the object medium and α is the half-angle of light collection by the objective lens of the microscope. The quantity $n \sin \alpha$ is called the numerical aperture (*N.A.*) and characterizes the ability of objective lenses to resolve small objects. The *N.A.* of high-power oil immersion objective lenses can be as high as 1.4. As an example, for green light (0.54 μm), the resolution of an *N.A.* 1.4 objective lens is 0.23 μm.

1.2 Specimen Preparation for Brightfield Microscopy

For useful imaging near the limits of resolution, biological and nonbiological specimens must be prepared as thin films or slices, called sections. Sections of most materials of thickness less than 20 μm are translucent and nearly invisible. Even under the microscope, lack of contrast makes observation of microscopic structure almost impossible unless special optics are used, such as phase contrast microscopy and differential interference contrast (DIC) microscopy. Without such special optics, sections must be heavily stained by addition of strongly absorbing dyes. In medicine and biology, the dyes hematoxylin and eosin (H&E) are most commonly used to stain microscope sections. Hematoxylin is a basic (cationic) dye that binds electrostatically to acidic (anionic) sites of nucleic acids (DNA and RNA). Eosin is an acidic anionic dye that binds to proteins, which become cationic when the staining procedure is performed at pH 4–5. Under the microscope, structures rich in nucleic acids, such as cell nuclei, have the blue color of hematoxylin, whereas protein-rich regions of the cytoplasm display the red staining of eosin.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 FLUORESCENCE OPTICAL MICROSCOPY

2.1 Fluorescence and Phosphorescence

Staining must be very intense to visualize specimens by brightfield microscopy. This limits sensitivity of brightfield microscopy to observe components of low abundance in a specimen. To improve sensitivity, specimens may be labeled with fluorescent tags. Absorbed light excites these tags to emit photons of longer wavelength. Fluorescence lifetime ranges between tenths of a nanosecond to about 100 ns and reflects the average time between absorbance of the excitation photon and release of the emitted photon. Quenching is any phenomenon that causes decay of the excited state without emission of a photon. Quenching commonly occurs when the energy of the excited state is transferred to solvent molecules and is dissipated as thermal energy. Light absorbance excites some molecules to a triplet state, which persists for hundreds of microseconds or longer before decaying to the ground state. For such molecules, release of a photon upon return to the ground state is called phosphorescence. Specialized fluorescence lifetime microscopes can image fluorescence and phosphorescence based both on lifetime and intensity. Thus, short- and long-lived emitters can be distinguished even if the wavelength of emission is identical.[1]

2.2 Epi-illumination

In fluorescence (and phosphorescence) microscopy, light of a specific wavelength illuminates the specimen to excite a specific absorbing species. Fluorescence emission is then magnified and imaged by the same basic optics as a conventional brightfield microscope. Special optical filters must be employed both to select the color of the excitation light and to distinguish fluoresced light from the excitation light. In a typical optical arrangement called epi-illumination, a dichroic mirror reflects the excitation light from a bright light source, such as a heated tungsten filament or mercury vapor lamp, into the microscope objective on to the specimen (Figure 1). Excitation light first passes through a barrier filter, which selects the desired wavelength of excitation. For example, a bandpass interference filter transmitting blue light between 470 and 500 nm might be employed to excite the green fluorescence of many fluorophores. Neutral density filters are also often needed to attenuate the excitation light to decrease photodamage and photobleaching. Fluorescence from excited fluorophores in the specimen passes back through the objective lens. Because dichroic mirrors can be constructed to reflect light of shorter wavelengths and transmit longer wavelength light, the dichroic mirror transmits only the fluoresced light, separating it from backscattered excitation light. After passing through another barrier filter, often a long-pass filter that transmits all light longer than a specified wavelength, the fluoresced light is formed into an image to be observed by the eye or recorded by photographic film or video camera. Epi-illumination fluorescence microscopes are very efficient at eliminating stray excitation light. As a result, fluorescence is imaged against a background of almost complete blackness. This accounts for the remarkable sensitivity of fluorescence microscopy to visualize fluorescently tagged specimens.

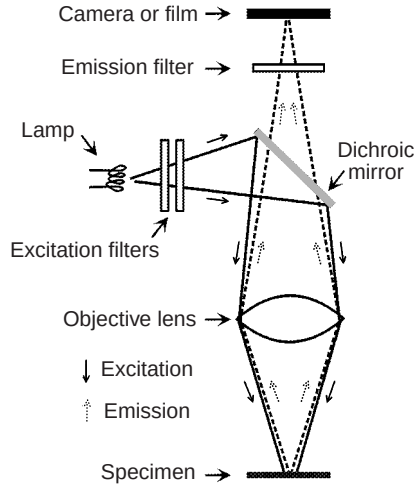

Figure 1 Scheme illustrating epi-illumination fluorescence microscopy. Light from a lamp passes through neutral density and interference barrier filters, reflects off a dichroic mirror and is projected on the specimen by the objective lens. Fluorescence excited by this illumination passes back through the objective lens, the dichroic mirror and an excitation barrier filter to be focused as an image on photographic film or video camera.

2.3 Confocal Fluorescence Microscopy

2.3.1 Problem of Imaging Fine Detail in Thick Specimens

Depth of field is the axial distance in the specimen space that is in focus in a single image. For conventional brightfield and fluorescence microscopes, the effective depth of field at high power is 2–3 μm. Since the lateral resolving power of optical microscopy is much greater than this depth of field, superimposition of detail within this field obscures structural detail that might otherwise be resolved. For specimens thicker than a few micrometers, light from out-of-focus planes is also superimposed, further degrading the image. In fluorescence microscopy, this out-of-focus fluorescence creates diffuse halos around in-focus objects under observation. In recent years, the technique of confocal microscopy has evolved to visualize fluorescence arising from only a very thin optical plane through the specimen.[2] Confocal images achieve submicrometer depths of field with virtually complete

rejection of light from out-of-focus planes. This smaller depth of field and rejection of out-of-focus light together produce images of remarkable detail and quality.

2.3.2 Pinhole Principle

The thin optical slices of confocal microscopy are obtained by illuminating one point at a time with a spot of light scanned across the specimen. The microscope objective focuses the illuminating beam, which is typically from a laser, to a tiny spot within the specimen (Figure 2). The diameter of these spots depends upon the wavelength of light and the *N.A.* of the objective lens (see Equation 1). Fluoresced light returns through the objective lens, and a dichroic mirror again separates the returning light from the illuminating light. After passing through an additional barrier filter, fluorescence is projected by the objective on to a small pinhole. Light originating from the in-focus cross-over point of the illuminating light beam is focused to a small spot precisely at the pinhole. This in-focus fluorescence passes through the pinhole to a light detector beyond, typically a photomultiplier. Fluorescence from above and below the focal plane is projected to spots behind and in front of the pinhole, respectively. As a result, out-of-focus fluorescence is spread out when it reaches the pinhole and does not traverse the pinhole aperture (Figure 2). In this way, the pinhole selectively transmits in-focus light and rejects out-of-focus light.

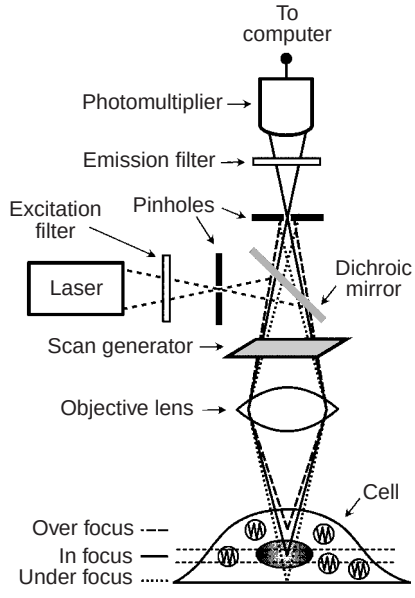

Figure 2 Scheme of a laser scanning confocal microscope. A laser beam passes through a barrier filter and is focused on a pinhole, which acts as a point source of light. A dichroic mirror reflects the laser beam into an $x-y$ scan generator and through to the objective lens. The scan generator together with the objective lens scans a diffraction-limited spot of light across the specimen. Fluoresced light passes back through the objective, scan generator, dichroic mirror and emission barrier filter to be focused on a pinhole aperture placed in the conjugate image plane. Light originating from the specimen plane of focus passes through the pinhole to a photomultiplier. Light from above and below the focal plane strikes the walls of the pinhole aperture and does not pass through to the photomultiplier. Photomultiplier output is fed to a computer.

To create a two-dimensional image, vibrating mirrors in the light path scan the illuminating light beam across the specimen.[3] The mirrors also ''descan'', returning fluorescence in order to focus it on the stationary pinhole. Computer memory then stores light detected by the photomultiplier, and a video monitor displays detected fluorescence in register with the movement of the light beam across the specimen.

2.3.3 Axial Resolution of Confocal Microscopy

The thickness of a confocal optical section decreases as the pinhole diameter decreases. The limit of resolution occurs when the pinhole becomes smaller than the diffraction limited spot (Airy disk) of in-focus light that is projected on it. Further decreases in pinhole diameter produce no further improvement of axial resolution (section thickness). Axial resolution (Z) at this limit is given by Equation (2):[4]

$$Z = \frac{2\lambda n}{(N.A.)^2} \tag{2}$$

Assuming green light (0.54 μm), oil immersion ($n = 1.55$) and an *N.A.* of the objective lens of 1.4, the axial resolution is 0.85 μm.

The better axial resolution of confocal microscopy improves the discrimination of individual small objects within a thick specimen, as compared with widefield microscopy. Confocal and nonconfocal images of cultured hepatocytes labeled with Rhodamine 123 illustrate the improvement (Figure 3a and b). Rhodamine 123 is a green-fluorescing cationic fluorophore that accumulates electrophoretically into the negatively charged mitochondria. In hepatocytes, mitochondria are round, oval and sometimes snake-like subcellular organelles of about 1 μm in diameter. When imaged by conventional epi-illumination fluorescence microscopy, individual mitochondria are easily identified at the cell periphery where the cytoplasm is the thinnest, but in thicker central regions superimposition of detail and out-of-focus fluorescence obscure the definition of the individual organelles (Figure 3a). Rather, an irregular mass of fluorescence occupies the cell center, suggesting falsely that the mitochondria are fused into a larger structure. By contrast, confocal microscopy shows that the mitochondria are well delineated individual structures, and the improvement of useful resolution is immediately obvious (Figure 3b).

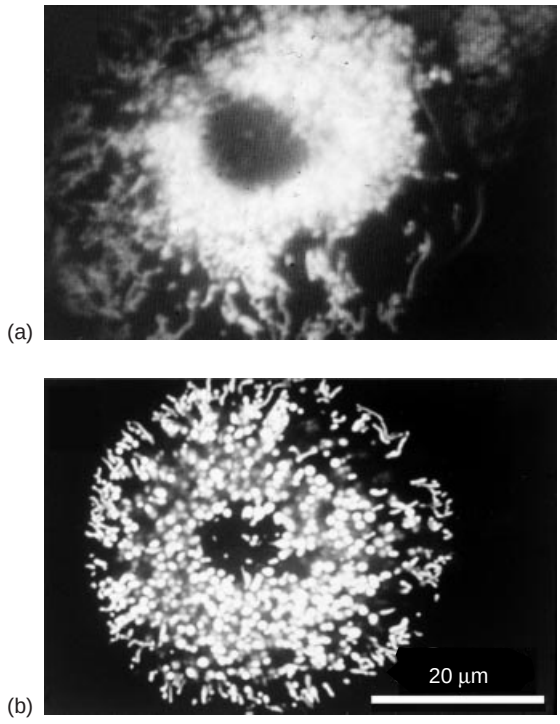

Figure 3 Comparison of conventional and confocal fluorescence microscopy. (a) Conventional widefield fluorescence micrograph of rhodamine 123-labeled mitochondria in a cultured hepatocyte recorded by a low-light video camera. (b) Confocal fluorescence micrograph of another rhodamine 123-labeled hepatocyte. Magnification is approximately the same. (Adapted from Lemasters et al.[5])

2.3.4 Excitation Sources for Confocal Microscopy

In laser scanning confocal microscopy, the type of laser determines the wavelengths available to excite fluorescence: 488 and 514 nm for argon lasers, 543 nm for helium–neon lasers, 488, 568 and 647 nm for argon–krypton lasers and 351–365 nm for ultraviolet (UV)–argon lasers.[6] Some confocal microscope designs employ a white light source. With interference filters, virtually any excitation wavelength can then be selected, but the fluorescence images are generally dimmer than with laser illumination. The wavelength of excitation must be chosen to match the fluorescence excitation spectrum of the fluorophore under study, e.g. 488 nm for fluorescein and 543 or 568 nm for rhodamine. Similarly, dichroic mirrors and barrier filters must be appropriate for the emission spectrum. Sometimes a second dichroic filter is placed in the emission light path so that fluorescence of different wavelengths may be separated and detected. With multiline lasers or a combination of different lasers, specimens tagged with two or more fluorophores may be excited simultaneously. The different colors of fluorescence emission from these fluorophores can then be detected simultaneously using different photomultipliers. Simultaneous detection of multiple fluorophores is particularly advantageous for comparing dynamic changes of parameters of interest in living specimens.

2.4 Multiphoton Excitation Fluorescence Microscopy

2.4.1 Multiphoton Fluorescence Excitation

If two photons impinge on a fluorophore virtually at the same time, their energy can be absorbed simultaneously to produce fluorescence excitation. The energy yield from absorbance of two photons is equivalent to that of one photon of twice the energy (i.e. half the wavelength). Hence two photons of red light can excite a fluorophore that normally requires UV illumination. Importantly, the rate of two-photon excitation varies with the square of the light intensity, just as the rate of a chemical reaction, $A + A \longrightarrow B$, varies with the square of the concentration of A. Similarly, three-photon and higher order fluorescence excitation can be achieved, which varies with the cube and higher powers of intensity.

2.4.2 Multiphoton Fluorescence Imaging

The nonlinear dependence of multiphoton fluorescence excitation has a very important implication for laser scanning microscopy. As illustrated in Figure 4, the light intensity of a beam of laser light focused by the objective lens is greatest at the cross-over point. As light travels through this cross-over point, the beam spreads out again. Because two-photon excitation is dependent on the square of light intensity, the zone of fluorescence excitation is effectly confined only to the point of highest light flux, namely at the cross-over point. As the beam spreads out, fluorescence excitation drops off rapidly. Thus, as the laser beam is scanned across a specimen, fluorescence arises only from the plane traversed by the cross-over point of the laser beam. The result is a thin optical section, but without the necessity for detection through a pinhole.[7,8]

A further difference between multiphoton microscopy and pinhole confocal microscopy is that fluorescence need not be imaged through the optics of the microscope.

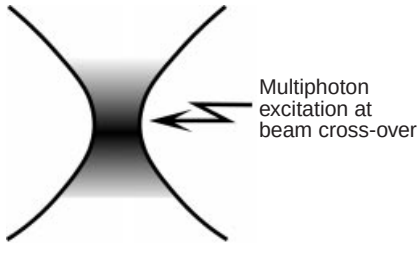

Figure 4 Scheme illustrating multiphoton excitation at cross-over of the focused laser beam.

Since the region excited by the two-photon phenomenon is intrinsically restricted to the plane of cross-over illumination, light may be collected directly by an external photomultiplier placed near the specimen. This detection configuration avoids light losses that occur when fluorescence is conducted back through the imaging system and scanner.

2.4.3 Laser Sources for Multiphoton Microscopy

To achieve efficient two- and three-photon fluorescence excitation, the instantaneous light intensity must be extremely high. Continuous illumination at such high power would almost instantly destroy specimens. Instead, pulsed lasers are employed to produce brilliant ultrashort 100–300-fs bursts of illumination. The instantaneous light intensity is sufficient to produce multiphoton excitation but the average power is much less because the individual pulses are separated by 10–13-ns intervals (75–100-MHz repetition rate). Currently, the most widely employed laser for multiphoton microscope systems is the pulsed Ti : sapphire laser, which is tunable between 680 and 1000 nm. When tuned to 700 nm, pulsed Ti : sapphire lasers produce two-photon excitation of UV-absorbing fluorophores. Longer wavelengths produce two-photon excitation of fluorescein- and rhodamine-like fluorophores. At 900 nm or greater, the Ti : sapphire laser can also achieve three-photon excitation of UV-absorbing fluorophores with simultaneous two-photon excitation of visible wavelength fluorophores. Modulation of the laser pulse width influences the relative strengths of two- and three-photon excitation. Shorter pulses favor three-photon excitation, whereas longer pulses produce relatively stronger two-photon excitation.[7]

2.4.4 Advantages of Multiphoton Microscopy Over Confocal Microscopy

Since photons are only absorbed by the target fluorophores at the cross-over point of the laser beam, essentially no photobleaching and photodamage occur above and below the image plane. Moreover, most biological tissue has little intrinsic absorbance of red and infrared (IR) light. Thus, at power levels required for both two- and three-photon excitation, collateral damage is virtually absent even after prolonged exposure. Because light scattering decreases with increasing wavelength and because intrinsic red- and IR-absorbing chromophores are in low abundance, biological tissues transmit red and especially near-IR light very well. Thus, the laser beam for multiphoton excitation penetrates deep into tissues. For both UV and visible wavelength fluorophores, fluorescence imaging deep into tissues by two- and three-photon microscopy becomes possible. This would be simply impossible by conventional one-photon fluorescence microscopy.

Unlike conventional fluorescence microscopy, the wavelength of emitted fluorescence in multiphoton microscopy is shorter than the excitation wavelength. Thus, emitted light scatters more than the excitation light. However, by using a photodetector placed near the specimen outside the microscope optics (nondescanned external detection) or by removing the pinhole from the detection light path, even scattered light can be captured with high efficiency. Such scattered light would not be collected by a conventional confocal microscope.

3 FLUORESCENCE LABELING

3.1 Autofluorescence

The most direct approach in fluorescence microscopy is observation of the intrinsic fluorescence or autofluorescence of specimens. In biology, oxidized flavoproteins, reduced forms of pyridine nucleotides [nicotinamide adenine dinucleotide (reduced form) (NADH) and nicotinamide adenine dinucleotide phosphate (reduced form) (NADPH)] and α-tocopherol (vitamin E) are autofluorescent and can provide useful images. Autofluorescence can also be an undesirable artifact that obscures observation of specific fluorescent tags. Certain preparative techniques produce autofluorescence in biological specimens, particularly chemical cross-linking with glutaraldehyde and other fixatives.

3.2 Immunofluorescence

Much more commonly, specimens are labeled with specific fluorescent tags. In immunocytochemistry, a primary antibody is prepared against a specific cellular constituent, such as a protein or complex carbohydrate. In direct immunocytochemistry, this antibody is reacted with a fluorescent molecule, such as fluorescein or rhodamine isothiocyanate. After treatment of the specimen with the fluorescently labeled antibody, fluorescence microscopy then shows the cellular and subcellular distribution of the constituent of interest (Figure 5a). In indirect immunocytochemistry, the tissue section is first treated with unlabeled primary antibody, such as rabbit immunoglobulin G (IgG). Subsequently, the section is exposed to a fluorescently labeled secondary antibody against the primary antibody, such as goat anti-rabbit antibody (Figure 5b). In general, indirect immunocytochemistry produces brighter fluorescent labeling than direct immunocytochemistry, because several fluorescently tagged secondary antibodies bind to each primary antibody. Other high affinity binding reactions are also

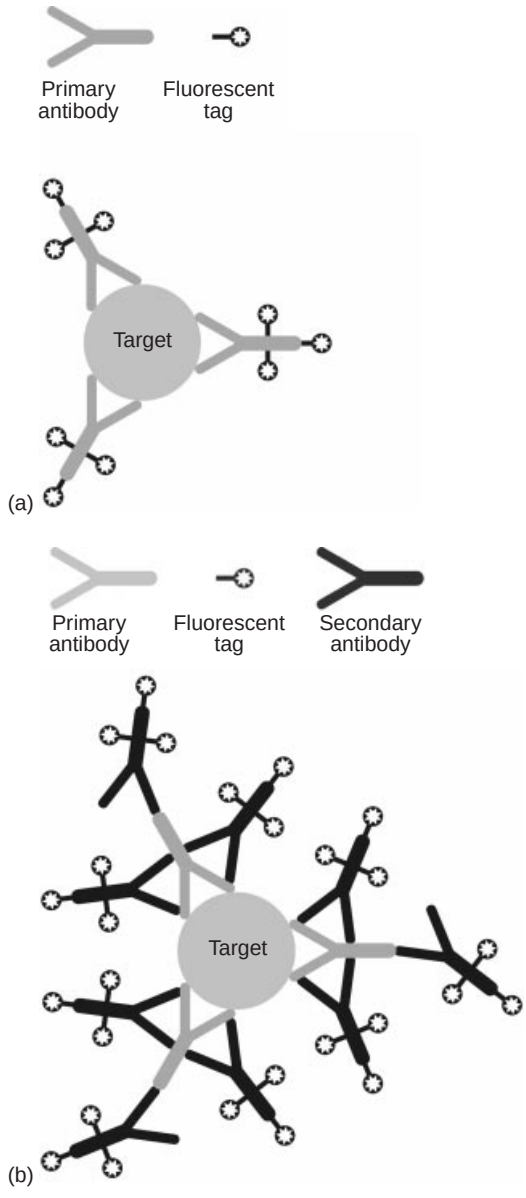

Figure 5 (a) Direct and (b) indirect immunocytochemistry. In direct immunocytochemistry, a fluorescently labeled primary antibody binds directly to its target. In the indirect method, an unlabeled antibody binds the target. Subsequently, a fluorescently labeled secondary antibody binds the primary antibody.

useful to label biological specimens, such as the binding of fluorescently tagged phalloidin, a mushroom toxin derivative, to actin microfilaments.

3.3 Parameter-indicating Fluorophores in Living Cells and Tissues

Fluorescence microscopy is becoming an indispensible tool to study cells and tissues in their living state. Although autofluorescence sometimes provides useful signals, in general living specimens must first be labeled with one or more parameter-indicating fluorescent probes. These fluorophores constitute four broad categories: (1) probes of general cell function and integrity, such as cell viability, cell shape and volume, electrical potential, and membrane permeability; (2) markers of specific intracellular structures, such as mitochondria, lysosomes and endoplasmic reticulum; (3) probes of specific chemical species, such as ions, oxygen radicals, and thiols; and (4) fluorescent molecular markers of specific cellular proteins. Table 1 shows a partial list of parameter-indicating fluorophores useful in fluorescence microscopy of living cells.[9–36] A much more comprehensive listing can be found in Haugland.[20]

3.3.1 Probes of General Cell Function

3.3.1.1 Cell Viability and Integrity Cells are normally impermeable to large hydrophilic molecules unless a specific membrane transporter system exists. When cell viability is lost, this permeability barrier breaks down, which allows otherwise impermeant molecules to cross freely into and out of the cell interior.[9–11] For example, viable cells exclude the cationic fluorophore propidium iodide. When viability is lost, propidium iodide enters cells and binds to nuclear DNA. This binding enhances the red fluorescence of propidium iodide, which is readily visualized by fluorescence microscopy.

Other fluorophores, such as fluorescein and calcein, are retained by viable cells and are released only after loss of cell viability. These fluorophores are loaded into cells as their neutral, membrane-permeant ester derivatives. After the neutral esters have crossed into cells, endogenous esterases release the membrane-impermeant free acid form of the fluorophores, which becomes trapped in the cytoplasm. When cell viability is lost, the trapped dye is released. In theory, cells should retain these fluorophores indefinitely. However, many cells possess transporters that allow these anionic fluorophores to cross the plasma membrane and gradually leak from the cytoplasm. Thus, dye retention may not be as reliable as dye exclusion as an indicator of cell viability.

The fluorophore calcein labels the cytosol when cells are incubated with calcein AM ester at 37 °C. In some cell types, calcein uptake is almost exclusively in the cytosol and the nucleus. Mitochondria exclude calcein and appear by negative contrast as round dark voids in the fluorescence. These voids disappear after mitochondria injury causing increased mitochondrial membrane permeability and uncoupling of oxidative phosphorylation. In this way, redistribution of intracellular calcein can reveal increases of permeability of intracellular membranes.[12]

3.3.1.2 Cell Volume and Topography Cell shape, volume and surface topography are essential features

Table 1 Fluorophores for living cells and tissues

Cellular parameter	Fluorophore
General cell function	
Cell viability	Propidium iodide[9]
	Ethidium homodimer[10]
	Calcein AM ester[11]
Membrane permeability	Calcein[12]
	Fluorescein- and rhodamine–dextran[13,14]
Cell volume/topography	Calcein[15]
Membrane potentials	Rhodamine 123[16]
	TMRM[17]
	Tetramethylrhodamine ethyl ester[17]
	JC-1[18]
	Di-4-ANEPPS[19]
Subcellular compartments	
Mitochondria	Rhodamine 123[16]
	TMRM[17]
	Tetramethylrhodamine ethyl ester[16]
	JC-1[18]
	MitoTracker Green® and MitoTracker Red®[20]
Endosomes	Rhodamine microspheres[21]
Lysosomes	LysoTracker Red®[20]
	Fluorescein- and rhodamine–dextran[21]
Golgi	BODIPY-ceramides[20,22]
Endoplasmic reticulum	$DiOC_6(3)$[23]
	ER-Tracker Blue-White DPX®[20]
Specific chemical species	
Ca^{2+}	Fura-2[24]
	Indo-1[24]
	Fluo-3[25]
	Fluo-4[20]
	Rhod-2[25]
pH	BCECF[26]
	SNARF-1[15,20]
Mg^{2+}	Mag-Fura[27]
Na^+	SBFI[28]
K^+	PBFI[28]
Glutathione	Monochlorobimane[29,30]
Protein thiols	Monobromobimane[29,30]
Oxygen radicals	Dichlorofluorescein[31]
	Dihydrorhodamine 123[32]
	Dihydroethidium[33]
Molecular tags and indicators	
Green fluorescent proteins	GFP[34]
	BFP
	CFP
	YFP
Bioengineered Ca^{2+} indicators	Cameleons[35,36]

AM, acetoxymethyl; TMRM, tetramethylrhodamine methyl ester; JC-1, tetrachlorotetraethylbenzimidazolylcarbocyanine; Di-4-ANEPPS, 4-dibutylaminonaphthalenylethenylsulfopropyl pyridium; BODIPY, boron dipyrromethene; $DiOC_6(3)$, dihexyloxacarbocyanine; BCECF, 2′,7′-biscarboxyethylcarboxyfluorescein; SNARF, seminaphthorhodafluor; BFP, blue fluorescent protein; CFP, cyan fluorescent protein; GFP, green fluorescent protein; SBFI, sodium-binding benzofuran isophthalate; PBFI, potassium-binding benzofuran isophthalate; YFP, yellow fluorescent protein.

(a)

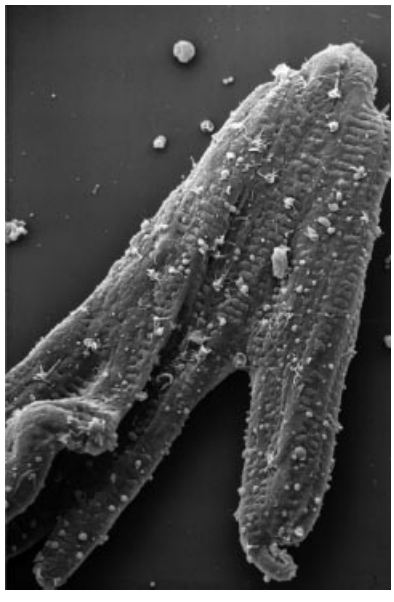

(b)

Figure 6 Confocal volume rendering and scanning electron micrograph of cardiac myocytes. (a) Computer-generated volume rendering of a calcein-loaded cardiac myocyte calculated from serial confocal images collected through the entire thickness of the cell. (b) Scanning electron micrograph of another cardiac myocyte at about the same magnification. (Adapted from Chacon et al.[15])

of cell structure. By labeling the intracellular space of cells with a fluorophore such as calcein, serial confocal images through the entire thickness of individual cells can be used to reconstruct cell volume and shape. Such reconstructions, or volume renderings, employ contour-dependent shading to depict cellular shape and surface detail, as illustrated in Figure 6(a) for a single cardiac myocyte in culture. Such volume renderings rival scanning electron micrographs in the depiction of topographical detail, although electron microscopy achieves superior lateral spatial resolution (Figure 6b). However, confocal fluorescence microscopy permits the repeated imaging of living cells, unlike scanning electron microscopy, which requires that specimens be chemically fixed, dried and coated with metal. Repeated volume renderings also quantify changes of cell volume over time.[15]

3.3.1.3 Electrical Potential Both the plasma membrane and mitochondria maintain a negative transmembrane electrical potential ($\Delta\Psi$). As a consequence, lipophilic cationic fluorophores such as rhodamine 123 and TMRM accumulate electrophoretically in response to these membrane potentials.[16,17] Uptake is related to $\Delta\Psi$ by the Nernst equation [Equation 3]:

$$\Delta\Psi = -60 \log \frac{F_{\text{in}}}{F_{\text{out}}} \tag{3}$$

where F_{in} and F_{out} represent the concentrations of the monocationic fluorophore inside and outside the membrane, respectively. From Equation (3), uptake ratios of 10, 100 and 1000 : 1 signify $\Delta\Psi$ values of -60, -120 and -180 mV, respectively. $\Delta\Psi$ of the plasma membrane ordinarily ranges between -30 and -90 mV, depending on cell type, and mitochondrial $\Delta\Psi$ is -100 to -190 mV. These $\Delta\Psi$ values are additive, such that mitochondria are as much as 240 mV more negative than the extracellular space, corresponding to a fluorophore concentration ratio of 10 000 : 1 inside mitochondria relative to the outside of the cell.

From confocal and multiphoton fluorescence images, average extracellular fluorescence intensity of a potential-indicating fluorophore can be divided into intracellular fluorescence for each picture element or pixel.[15] By applying the Nernst equation to each pixel, electrical potential of every point inside the cell relative to the extracellular space can be calculated and displayed as a pseudocolor map. Differences in electrical potential across membranes then represent the corresponding transmembrane potentials.

Figure 7 illustrates the intracellular electrical potential measured in a cultured rat hepatocyte. Under the cell surface and in the nucleus, the electrical potential is about -40 mV. Since extracellular electrical potential represents the ground (0 mV), $\Delta\Psi$ is -40 mV across the plasma membrane. Distributed throughout the cytoplasm are also more electronegative mitochondria with potentials as great as -160 mV. Since the electrical potential of the surrounding cytosol is -40 mV, the difference, -120 mV, represents mitochondrial $\Delta\Psi$. Mitochondria appear to show heterogeneity of electrical potential. This heterogeneity is due, at least in part, to the fact that not

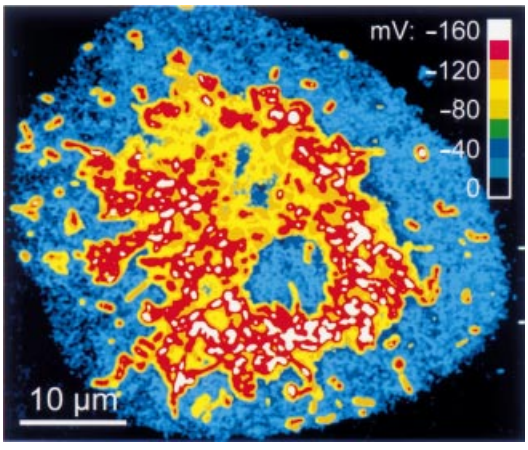

Figure 7 Distribution of electrical potential in a hepatocyte. Red fluorescence of a TMRM-loaded hepatocyte was imaged by laser scanning confocal microscopy. The distribution of intracellular electrical potential was determined by the application of Equation (3) to each pixel in the image. (From E. Chacon, G. Zahrebelski, J.M. Reece, A.-L. Nieminen and J.J. Lemasters, unpublished work.)

all mitochondria extend completely through the confocal slice, causing electrical potential to be underestimated.

3.3.2 *Markers of Specific Intracellular Structures*

Many fluorophores tag specific intracellular structures. As discussed above, cationic fluorophores such as rhodamine 123 and TMRM brightly label mitochondria (Figures 3 and 7). Carbocyanine dyes label endoplasmic reticulum[23] and fluorescent ceramide conjugates label the Golgi apparatus.[20–22] Weak bases, such as LysoTracker Red®,[20] accumulate into acidic lysosomal compartments (Table 1).

3.3.3 *Probes of Specific Chemical Species*

3.3.3.1 Ion Indicators and Ratio Imaging Many fluorophores measure ion concentration in individual living cells. Typically, these fluorophores are loaded as their neutral ester derivatives, as described above for calcein. Inside cells, the fluorescence of ion-indicating fluorophores depends on both the ion concentration and the amount of fluorophore present in the light path. When ion binding to a fluorophore causes a red or blue shift of the excitation or emission spectrum, a ratioing procedure may be used to correct for changes in fluorophore concentration (Figure 8).[37] Images are acquired at two different excitation or emission wavelengths: one wavelength that changes with changing ion concentration, and another wavelength that either does not change or changes in the opposite direction. After background subtraction to remove any contribution of nonspecific scattered light

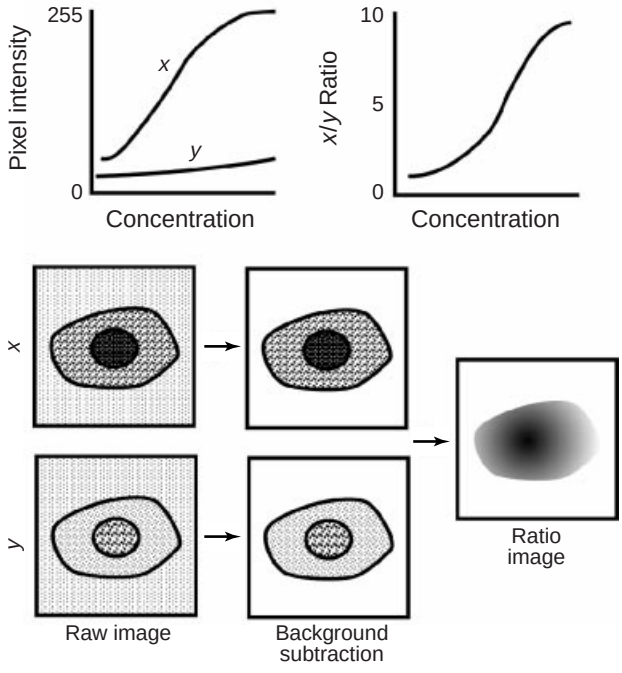

Figure 8 Principle of ratio imaging. See text for details.

and instrumental dark current, the fluorescence image at the first wavelength is divided by the image at the second wavelength on a pixel-by-pixel basis. The result is a ratio image. For example, with 568-nm excitation, the fluorescence of carboxy SNARF-1 emitted at 640 nm increases as the pH increases, but emission at 585 nm does not change. Consequently, the ratio of fluorescence at the two wavelengths is a function of pH but is independent of SNARF-1 concentration.[15] Ratioing eliminates variations in image brightness due to differences of regional fluorophore concentration, dye leakage over time and photobleaching. Using a calibration curve, ratio images are converted to maps of intracellular ion distribution. Not all ion-indicating fluorophores are suitable for this ratioing procedure. Only indicators that have ion-induced spectral shifts, such as Indo-1 and Fura-2 for Ca^{2+}, SNARF-1 and BCECF for pH, SBFI for Na^+, PBFI for K^+ and Mag-Fura for Mg^{2+} (Table 1). Unfortunately, all these dyes except SNARF-1 and BCECF require UV excitation.

Ion Indicators and Nonratiometric Indicators Green-fluorescing Fluo-3 and Fluo-4 and red-fluorescing Rhod-2 are useful visible wavelength indicators of free Ca^{2+} (Table 1). Their binding affinity for Ca^{2+} is in the range 400–500 nM and their fluorescence increases 30–80-fold after Ca^{2+} binding. However, Ca^{2+} binding does not produce a spectral shift. Hence ratio imaging cannot be applied. Nonetheless, Fluo-3, Fluo-4 and Rhod-2 are useful to measure relative changes of free Ca^{2+}, especially

in confocal and multiphoton microscopy which produce uniform optical slices through cells of interest.

Compartmental Loading of Ion Indicators Cation-indicating fluorophores are hydrophilic multivalent carboxylic acids that do not cross cellular membranes. To load these indicators, cells are incubated with their uncharged acetate or AM ester. The esters cross the plasma membrane, become de-esterified and are thus trapped in the cytoplasm. Since esterases reside in several different cellular compartments, ion-indicating fluorophores can distribute into several different compartments after ester loading, often in a heterogeneous fashion.

In many cell types, the temperature of ester loading strongly influences the intracellular distribution of various fluorophores.[12] Warm ester loading at 37 °C favors cytosolic uptake, whereas cold ester loading at 4 °C promotes fluorophore entry into both the cytosol and membranous organelles, in particular the mitochondria. Apparently during warm loading, cytosolic esterases hydrolyze the fluorophore esters before they have the chance to diffuse into mitochondria. By contrast, cold temperature slows esterase activity such that esterified probes are able to diffuse across the cytosol and into mitochondria before hydrolysis occurs. The temperature dependence of ester loading varies with cell type and also with the particular fluorophore used. In cardiac myocytes, for example, calcein and Fluo-3, a green-fluorescing Ca^{2+} fluorophore, localize to the cytosol after warm ester loading, whereas Indo-1, a blue-fluorescing Ca^{2+} indicator, and SNARF 1 load into mitochondria even at 37 °C.[12,38]

Anionic fluorophores in the cytosol gradually leak across the plasma membrane through an organic anion carrier,[39] but release of fluorophores localized to mitochondria and other organelles is much slower. Selective release from the cytosol can be used to achieve selective mitochondrial loading of ion-indicating fluorophores. First, cold loading is used to distribute a fluorophore into both the cytosol and the mitochondria. Subsequently, the loaded cells are incubated at 37 °C for several hours. During the warm incubation, fluorophore in the cytosol leaks across the plasma membrane and is lost to the extracellular medium, but fluorophore in the mitochondria and other organelles (chiefly lysosomes) remains. Subsequently, warm ester loading can be used to load a second fluorophore into the cytosol. In this way, one fluorophore enters the cytosol and another fluorophore enters mitochondria and organelles (Figure 9a and b). By using two-color fluorescence microscopy, the fluorescence of the two fluorophores can be independently measured to compare, for example, changes of cytosolic and mitochondrial Ca^{2+} directly.[40–42]

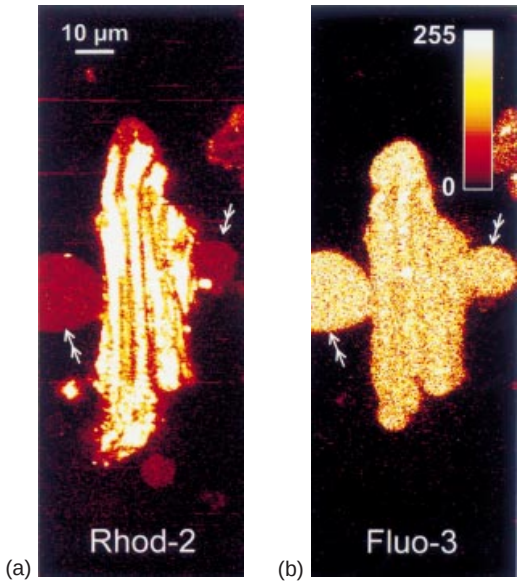

Figure 9 Simultaneous loading of fluorophores into different subcellular compartments. The Ca^{2+} indicator Rhod-2 was localized to the mitochondria of a cardiac myocyte by a cold loading–warm incubation protocol. Subsequently, Fluo-2, another Ca^{2+} indicator, was loaded into the cytosol by warm ester loading, them low Na^{+} buffer was added to increase intracellular Ca^{2+} and saturate the Ca^{2+}-dependent fluorescence of both indicators. Rhod 2 fluorescence (a) was present in mitochondria, but not in toxic blebs (double arrows), whereas Fluo 3 fluorescence (b) was diffusely present in both the cytosol and the blebs. (Adapted from Trollinger et al.[40])

3.3.3.2 Reactive Oxygen Species and Cellular Thiols Incomplete reduction of oxygen causes formation of reactive oxygen species (ROS), including superoxide, hydrogen peroxide and hydroxyl radical. ROS are formed during normal metabolism and as a consequence of various pathological processes. Certain nonfluorescent compounds, such as dichlorofluorescin, dihydrorhodamine 123 and hydroethidium, react with ROS to form highly fluorescent derivatives (Table 1).[31–33] Fluorescence microscopy can then be used to image the rate and subcellular location of ROS formation in single cells.[31] Glutathione and protein thiols are important targets of oxidation by ROS. Monochlorobimane reacts with glutathione and monobromobimane reacts with both glutathione and protein thiols to form fluorescent adducts. These adducts may also be imaged by fluorescence microscopy to quantify the intracellular distribution of protein and nonprotein thiols in single cells.[29,30]

The oxidation–reduction status of pyridine nucleotides [nicotinamide adenine dinucleotide (oxidized form) (NAD^{+}) and nicotinamide adenine dinucleotide phosphate (oxidized form) ($NADP^{+}$)] can be conveniently monitored by imaging of autofluorescence. NADH and NADPH, the reduced forms of NAD^{+} and $NADP^{+}$, are

excited to emit blue light (450 nm) by near-UV light (340 nm). NADH and NADPH fluorescence is highly quenched in the cytosol. Consequently, nearly all NADH and NADPH fluorescence arises from mitochondria.[31] The green fluorescence of oxidized flavin can also be used to monitor the oxidation–reduction of flavoproteins in mitochondria and elsewhere.[43]

3.3.4 Molecular Markers

Recently, the GFP of the jelly fish *Aequorea victoria* has emerged as an important fluorescent reporter of gene expression and protein localization in living cells.[34] GFP is a 27 000-Da monomer consisting of 238 amino acids in an 11-stranded β-barrel surrounding a coaxial α-helix containing the chromophore. Wild-type GFP absorbs near-UV and blue light and emits green light with a maximum at 509 nm. Several red- and blue-shifted GFP mutants have been developed to increase brightness and yield emissions of blue, cyan and yellow in addition to green.

To use GFP, fusion constructs are made of the cDNA for GFP and a cDNA of a target protein. After gene transfer by transfection or other means (e.g. viral vectors), living cells use this DNA to synthesize a fusion protein comprised of the target protein and its GFP tag. Using fluorescence microscopy, the subcellular localization and intracellular trafficking of the GFP fusion protein can then be followed in the living state (Figure 10a and b). Fluorescence resonance energy transfer occurs between the different-colored GFP mutants and is the basis for the development of unique new chimeric indicators. For example, Ca^{2+}-indicating fluorophores have been developed by fusing cyan-emitting fluorescent protein (CFP) with calmodulin and yellow-emitting fluorescent protein (YFP). Ca^{2+} binding then alters fluorescence energy transfer between CFP and YFP. These proteins are called cameleons and can be imaged by fluorescence microscopy.[35,36] Mutagenesis and molecular engineering of GFP should continue to provide useful new molecular probes for fluorescence microscopy.

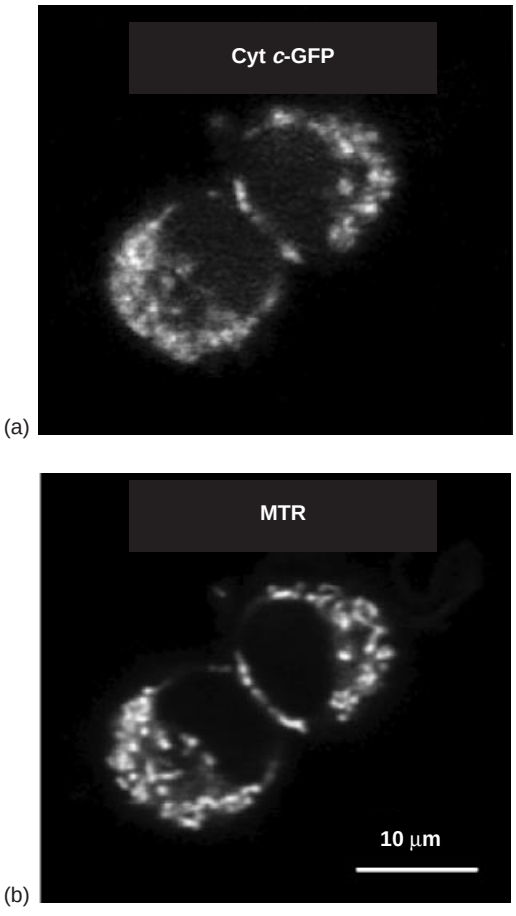

Figure 10 Confocal fluorescence microscopy of cytochrome *c* GFP. Pheochromocytoma cells were transfected with cytochrome *c*-GFP and co-loaded with MitoTracker Red®. The green fluorescence of cytochrome *c*-GFP (a) co-localizes with the red fluorescence of MitoTracker Red® (b). The punctate structures in both images are individual mitochondria. (Adapted from Heiskanen et al.[44] Courtesy of A.L. Nieminen.)

4 CONCLUSIONS

Fluorescence microscopy is a powerful tool to identify and localize specific molecular components with lateral resolution of 0.2 μm. New techniques of confocal and multiphoton microscopy also permit an axial resolution approaching 0.8 μm even in thick specimens. Unlike electron microscopy, fluorescence optical microscopy allows nondestructive observation of living cells and tissues. The number of specific parameter-indicating fluorophores useful for live-cell imaging continues to increase. These include markers of cell shape and volume, organelles, membrane potentials, ions and other chemical constitutents. With the advent of GFP and its many mutants, virtually any cellular protein can be fluorescently tagged in the living state. These developments make fluorescence microscopy an indispensible tool for the molecular biologist, cell biologist, physiologist and others, both inside and outside the life sciences.

ABBREVIATIONS AND ACRONYMS

AM	Acetoxymethyl
BCECF	2′,7′-Biscarboxyethylcarboxy-fluorescein
BFP	Blue Fluorescent Protein
BODIPY	Boron Dipyrromethene

CFP	Cyan Fluorescent Protein
Di-4-ANEPPS	4-Dibutylaminonaphthalenyl-ethenylsulfopropyl Pyridium
DIC	Differential Interference Contrast
$DiOC_6(3)$	Dihexyloxacarbocyanine
GFP	Green Fluorescent Protein
H&E	Hematoxylin and Eosin
IgG	Immunoglobulin G
IR	Infrared
JC-1	Tetrachlorotetraethylbenzimida-zolylcarbocyanine
NAD^+	Nicotinamide Adenine Dinucleotide (Oxidized Form)
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate (Oxidized Form)
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
PBFI	Potassium-binding Benzofuran Isophthalate
ROS	Reactive Oxygen Species
SBFI	Sodium-binding Benzofuran Isophthalate
SNARF	Seminaphthorhodafluor
TMRM	Tetramethylrhodamine Methyl Ester
UV	Ultraviolet
YFP	Yellow Fluorescent Protein

RELATED ARTICLES

Surfaces **(Volume 10)**
Scanning Probe Microscopy, Industrial Applications of

Electronic Absorption and Luminescence **(Volume 12)**
Detectors, Absorption and Luminescence • Fluorescence Lifetime Measurements, Applications of • Phosphorescence Measurements, Applications of • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

REFERENCES

1. E.J. Hennink, R. de Haas, N.P. Verwoerd, H.J. Tanke, 'Evaluation of a Time-resolved Fluorescence Microscope Using a Phosphorescent Pt–Porphine Model System', *Cytometry*, **24**, 312–320 (1996).
2. C.J.R. Sheppard, D.M. Shotton, *Confocal Laser Scanning Microscopy*, Springer, New York, 1997.
3. J.G. White, W.B. Amos, M. Fordham, 'An Evaluation of Confocal Versus Conventional Imaging of Biological Structures by Fluorescence Light Microscopy', *J. Cell Biol.*, **105**, 41–48 (1987).
4. S. Inoué, 'Foundations of Confocal Scanned Imaging in Light Microscopy', in *Handbook of Biological Confocal Microscopy*, 2nd edition, ed. J.B. Pawley, Plenum Press, New York, 1–17, 1995.
5. J.J. Lemasters, E. Chacon, G. Zahrebelski, J.M. Reece, A.-L. Nieminen, 'Laser Scanning Confocal Microscopy of Living Cells', in *Optical Microscopy: Emerging Methods and Applications*, eds. B. Herman, J.J. Lemasters, Academic Press, New York, 339–354, 1993.
6. E. Gratton, M.J. vandeVen, 'Laser Sources for Confocal Microscopy', in *Handbook of Biological Confocal Microscopy*, 2nd edition, ed. J.B. Pawley, Plenum Press, New York, 69–97, 1995.
7. W. Denk, J.H. Strickler, W.W. Webb, 'Two-photon Laser Scanning Fluorescence Microscopy', *Science*, **248**, 73–66 (1990).
8. W. Denk, D.W. Piston, W.W. Webb, 'Two-photon Molecular Excitation in Laser-scanning Microscopy', in *Handbook of Biological Confocal Microscopy*, 2nd edition, ed. J.B. Pawley, Plenum Press, New York, 445–458, 1995.
9. J.J. Lemasters, J. DiGuiseppi, A.-L. Nieminen, B. Herman, 'Blebbing, Free Ca^{++} and Mitochondrial Membrane Potential Preceding Cell Death in Hepatocytes', *Nature*, **325**, 78–81 (1987).
10. M. Ankarcrona, J.M. Dypbukt, E. Bonfoco, B. Zhivotovsky, S. Orrenius, S.A. Lipton, P. Nicotera, 'Glutamate-induced Neuronal Death: A Succession of Necrosis or Apoptosis Depending on Mitochondrial Function', *Neuron*, **15**, 961–973 (1995).
11. K.H. Jones, J.A. Senft, 'An Improved Method to Determine Cell Viability by Simultaneous Staining with Fluorescein Diacetate–Propidium Iodide', *J. Histochem. Cytochem.*, **33**, 77–79 (1985).
12. A.-L. Nieminen, A.K. Saylor, S.A. Tesfai, B. Herman, J.J. Lemasters, 'Contribution of the Mitochondrial Permeability Transition to Lethal Injury After Exposure of Hepatocytes to *t*-Butyl Hydroperoxide', *Biochem. J.*, **307**, 99–106 (1995).
13. Y. Nishimura, J.J. Lemasters, 'Plasma Membrane Permeability Changes in Cultured Sinusoidal Endothelial Cells Exposed to Cyanide Monitored by Negative Contrast Confocal Microscopy', *Cell Vision – J. Anal. Morphol.*, **4**, 174–175 (1997).
14. Z. Dong, Y. Patel, P. Saikumar, J.M. Weinberg, M.A. Venkatachalam, 'Development of Porous Defects in Plasma Membranes of Adenosine Triphosphate-depleted Madin–Darby Canine Kidney Cells and Its Inhibition by Glycine', *Lab. Invest.*, **78**, 657–668 (1998).
15. E. Chacon, J.M. Reece, A.-L. Nieminen, G. Zahrebelski, B. Herman, J.J. Lemasters, 'Distribution of Electrical Potential, pH, Free Ca^{2+}, and Volume Inside Cultured Adult Rabbit Cardiac Myocytes During Chemical Hypoxia: A Multiparameter Digitized Confocal Microscopic Study', *Biophys. J.*, **66**, 942–952 (1994).

16. L.V. Johnson, M.L. Walsh, B.J. Bockus, L.B. Chen, 'Monitoring of Relative Mitochondrial Membrane Potential in Living Cells by Fluorescence Microscopy', *J. Cell Biol.*, **88**, 526–535 (1981).
17. D.L. Farkas, M.D. Wei, P. Febbroriello, J.H. Carson, L.M. Loew, 'Simultaneous Imaging of Cell and Mitochondrial Membrane Potentials', *Biophys. J.*, **56**, 1053–1069 (1989).
18. M. Reers, S.T. Smiley, C. Mottola-Hartshorn, A. Chen, M. Lin, L.B. Chen, 'Mitochondrial Membrane Potential Monitored by JC-1 Dye', *Methods Enzymol.*, **260**, 406–417 (1995).
19. L.M. Loew, L.B. Cohen, J. Dix, E.N. Fluhler, V. Montana, G. Salama, W. Jian-Young, 'A Naphthyl Analog of the Aminostyryl Pyridinium Class of Potentiometric Membrane Dyes Shows Consistent Sensitivity in a Variety of Tissue, Cell, and Model Membrane Preparations', *J. Membr. Biol.*, **130**, 1–10 (1992).
20. R.P. Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, 7th edition, Molecular Probes, Eugene, OR, 1999.
21. G.J. Gores, A.-L. Nieminen, B.E. Wray, B. Herman, J.J. Lemasters, 'Intracellular pH During ''Chemical Hypoxia'' in Cultured Rat Hepatocytes: Protection by Intracellular Acidosis Against the Onset of Cell Death', *J. Clin. Invest.*, **83**, 386–396 (1989).
22. R.E. Pagano, O.D. Martin, H.C. Kang, R.P. Haugland, 'A Novel Fluorescent Ceramide Analog for Studying Membrane Traffic in Animal Cells: Accumulation at the Golgi Apparatus Results in Altered Spectral Properties of the Sphingolipid Precursor', *J. Cell Biol.*, **113**, 1267–1279 (1991).
23. C. Lee, L.B. Chen, 'Dynamic Behavior of Endoplasmic Reticulum in Living Cells', *Cell*, **54**, 37–46 (1988).
24. G. Grynkiewicz, M. Poenie, R.Y. Tsien, 'A New Generation of Ca^{2+} Indicators with Greatly Improved Fluorescence Properties', *J. Biol. Chem.*, **260**, 3440–3450 (1985).
25. A. Minta, J.P.Y. Kao, R.Y. Tsien, 'Fluorescent Indicators for Cytosolic Calcium Based on Rhodamine and Fluorescein Chromophores', *J. Biol. Chem.*, **264**, 8171–8178 (1989).
26. T.J. Rink, R.Y. Tsien, T. Pozzan, 'Cytoplasmic pH and Free Mg^{2+} in Lymphocytes', *J. Cell Biol.*, **95**, 189–196 (1982).
27. B. Raju, E. Murphy, L.A. Levy, R.D. Hall, R.E. London, 'A Fluorescent Indicator for Measuring Cytosolic Free Magnesium', *Am. J. Physiol.*, **256**, C540–C548 (1989).
28. A. Minta, R.Y. Tsien, 'Fluorescent Indicators for Cytosolic Sodium', *J. Biol. Chem.*, **264**, 19 449–19 457 (1989).
29. A.-L. Nieminen, G.J. Gores, T.L. Dawson, B. Herman, J.J. Lemasters, 'Mechanisms of Toxic Injury by $HgCl_2$ in Rat Hepatocytes Studied by Multiparameter Digitized Video Microscopy', in *Optical Microscopy for Biology*, eds. B. Herman, K. Jacobson, Alan R. Liss, New York, 323–335, 1990.
30. G. Bellomo, M. Vairetti, L. Stivala, F. Mirabelli, P. Richelmi, S. Orrenius, 'Demonstration of Nuclear Compartmentalization of Glutathione in Hepatocytes', *Proc. Natl. Acad. Sci. USA*, **89**, 4412–4416 (1992).
31. A.-L. Nieminen, A.M. Byrne, B. Herman, J.J. Lemasters, 'Mitochondrial Permeability Transition in Hepatocytes Induced by *t*-BuOOH: AND(P)H and Reactive Oxygen Species', *Am. J. Physiol.*, **272**, C1286–C1294 (1997).
32. S. Perticarari, G. Presani, M.A. Mangiarotti, E. Banfi, 'Simultaneous Flow Cytometric Method to Measure Phagocytosis and Oxidative Products by Neutrophils', *Cytometry*, **12**, 687–693 (1991).
33. G. Rothe, A. Oser, G. Valet, 'Dihydrorhodamine 123: A New Flow Cytometric Indicator for Respiratory Burst Activity in Neutrophil Granulocytes', *Naturwissenschaften*, **75**, 354–355 (1988).
34. R.Y. Tsien, 'The Green Fluorescent Protein', *Annu. Rev. Biochem.*, **67**, 509–544 (1998).
35. A. Miyawaki, J. Llopis, R. Heim, J.M. McCaffery, J.A. Adams, M. Ikura, R.Y. Tsien, 'Fluorescent Indicators for Ca^{2+} Based on Green Fluorescent Proteins and Calmodulin', *Nature*, **388**, 882–887 (1997).
36. A. Miyawaki, O. Griesbeck, R. Heim, R.Y. Tsien, 'Dynamic and Quantitative Ca^{2+} Measurements Using Improved Cameleons', *Proc. Natl. Acad. Sci. USA*, **96**, 2135–2140 (1999).
37. L. Tanasugarn, P. McNeil, G.T. Reynolds, D.L. Taylor, 'Microspectrofluorometry by Digital Image Processing: Measurement of Cytoplasmic pH', *J. Cell Biol.*, **98**, 717–724 (1984).
38. H. Ohata, E. Chacon, S.A. Tesfai, I.S. Harper, B. Herman, J.J. Lemasters, 'Mitochondrial Ca^{2+} Transients in Cardiac Myocytes During the Excitation–Contraction Cycle: Effects of Pacing and Hormonal Stimulation', *J. Bioenerg. Biomembr.*, **30**, 207–222 (1998).
39. E.D. Wieder, H. Hang, M.H. Fox, 'Measurement of Intracellular pH Using Flow Cytometry with Carboxy-SNARF-1', *Cytometry*, **14**, 916–921 (1993).
40. D.R. Trollinger, W.E. Cascio, J.J. Lemasters, 'Dual Loading of Rhod 2 and Fluo 3 Shows Ruthenium Red Blocks Mitochondrial But Not Cytosolic Ca^{2+} Transients in Myocytes', *J. Mol. Cell. Cardiol.*, **30**, A245 (1998).
41. D.R. Trollinger, W.E. Cascio, J.J. Lemasters, 'Selective Loading of Rhod 2 into Mitochondria Shows Mitochondrial Ca^{2+} Transients During the Contractile Cycle in Adult Rabbit Cardiac Myocytes', *Biochem. Biophys. Res. Commun.*, **236**, 738–742 (1997).
42. J.J. Lemasters, D.R. Trollinger, T. Qian, W.E. Cascio, H. Ohata, 'Confocal Imaging of Ca^{2+}, pH, Electrical Potential and Membrane Permeability in Single Living Cells', *Methods Enzymol.*, **302**, 341–358 (1999).
43. G. Hajnoczky, L.D. Robb-Gaspers, M.B. Seitz, A.P. Thomas, 'Decoding of Cytosolic Calcium Oscillations in the Mitochondria', *Cell*, **82**, 415–424 (1995).
44. K.M. Heiskanen, M.B. Bhat, H.W. Wang, J. Ma, A.-L. Nieminen, 'Mitochondrial Depolarization Accompanies Cytochrome *c* Release During Apoptosis in PC6 Cells', *J. Biol. Chem.*, **274**, 5654–5658 (1999).

Fluorescence in Organized Assemblies

Najma Memon, Aamna Balouch and Willie L. Hinze
Wake Forest University, Winston-Salem, NC, USA

1 Introduction **1**
2 Brief Description of Organized Assemblies **2**
2.1 Survey of Surfactant Aggregate Systems 2
2.2 Cyclodextrin Systems 6
2.3 Summary of Selected Properties of Organized Assemblies 7
3 Organized Assemblies in Analytical Fluorescence Measurements **26**
3.1 Enhancement of Fluorescence Intensity in Organized Media 26
3.2 Utilization of Organized Assemblies to Improve Fluorescence Methods for Inorganic and Organic Analytes 39
4 Some General Experimental Considerations **55**
5 Conclusions **58**
Acknowledgments **58**
Abbreviations and Acronyms **59**
Related Articles **60**
References **60**

This article provides an overview of the general properties of organized assembly (ordered media) systems, such as aqueous surfactant and bile salt micelles, lipid (liposomes) and surfactant vesicles, and cyclodextrins (CDs), and summarizes their utilization to enhance the performance of analytical fluorescence measurements. In many instances, organic molecules and metal complex species when included within a CD cavity or solubilized and bound to surfactant aggregates exhibit enhanced fluorescence, improving detectability of such analytes. The altered microenvironment within the organized medium can impede the interfering action of other species (inorganic or organic) present in the sample matrix, which often improves the selectivity of the analytical method. These benefits of improved sensitivity and selectivity arise from the compartmentalization, isolation, and shielding of the excited singlet state of the guest analyte from quenching and nonradiative decay processes, as well as preventing side reactions that otherwise can occur in bulk solution (or sample matrix). Organic solvents or time-consuming extraction steps can also be avoided, owing to the increased solubility of nonpolar organic or inorganic reagents and/or analyte molecules in water in the presence of the organized medium, allowing the use of an aqueous medium to perform the procedure. The possibility of conducting reactions and forming fluorescent organic or metal chelates in micellar (or other organized) media that are not observed in a bulk homogeneous solvent system serves to expand the scope of possible chemistries for design/development of new, unique, and improved fluorescent assays. Examples of fluorescent methods for determination of both organic and inorganic analytes are provided, which serve to illustrate the advantages and benefits accrued from the use of the micelles, vesicles, liposomes, or CDs in such procedures, along with experimental considerations and cautions in utilizing organized media.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

1 INTRODUCTION

The past 35 years has witnessed ever-increasing interest in and utilization of the effects of organized assemblies (also termed *ordered media*) on photophysical and photochemical processes and reactions.[1–8] Reactants accommodated in such assemblies, i.e. normal and reversed micelles, vesicles and liposomes, CDs, etc., often achieve a greater degree of organization and altered microenvironment compared to that in bulk homogeneous solution, which may promote unique general properties and reaction features, characteristics which may mimic processes in biosystems and have wide applications in chemical analysis and separation science.[9–16]

This article focuses on several types of aqueous-based organized assembly (also referred to as *ordered media* or *membrane mimetic*) systems and their utilization to enhance analytical luminescence measurements. Included for discussion are the aqueous (normal) surfactant and bile salt micelle, aqueous synthetic surfactant vesicle and liposome, and aqueous CD systems. Following a brief description of these different organized assemblies and their basic properties, examples of applications involving such media in the area of analytical fluorescence methodology are surveyed. The goal is to present a general overview of the different possible beneficial roles and uses of such media in analytical luminescence spectroscopic measurements. The overview is intended to be more applied and illustrative in nature rather than being a comprehensive discussion of all the voluminous reported work in this area. For more fundamental information and details on the fluorescence technique in general, the reader is referred to a monograph on this topic[17]

(*see* **Pervaporation, Analytical**; **Electronic Absorption and Luminescence: Introduction**).

2 BRIEF DESCRIPTION OF ORGANIZED ASSEMBLIES

Of the many different organized assemblies and supramolecular systems reported in the literature, this article restricts itself to those surfactant aggregate systems and CD hosts that exist in aqueous solution. This choice was dictated by the fact that most of the reported applications in analytical fluorescence measurements have utilized such media. Detailed discussion of all aspects and salient features of the chemistry and properties of such surfactant and CD ordered media is beyond the scope of this article. More fundamental information on these systems can be obtained by consulting the cited monographs, reviews, and research articles on desired topics mentioned herein.

2.1 Survey of Surfactant Aggregate Systems

Surfactant (detergent) molecules, which contain an alkyl hydrophobic chain (tail) and a polar head group, are capable of forming aggregate supramolecular assemblies that exhibit properties distinctly different from those of the individual monomers prior to aggregation. Depending on the lipophilic–hydrophilic balance, a variety of aggregate morphologies, such as normal micelles, rods, disks, planar bilayers, vesicles, liposomes, and hexagonal phases, among others, have all been observed. With increasing surfactant concentration, morphologies ranging from micelles (L_1), through a cubic array of micelles (I_1), hexagonally packed rods (H_1), cubic bicontinuous spheres (V_1), to lamellae (L_3) where phase inversion takes place, and inverted morphologies develop.[11,12,18,19] Among these, this overview focuses on just the normal aqueous surfactant and bile salt micelle, liposome, and synthetic vesicle systems. A monograph addresses the use of reverse micelles in chemical analysis, including their utilization in fluorescence measurements.[19]

2.1.1 Normal Surfactant Micelles

In terms of actual analytical applications, aqueous (or normal) surfactant micelle systems have probably been the most frequently employed form of surfactant-based organized media in chemical analysis applications. Normal micelles are formed in aqueous solution by the aggregation of amphiphilic surfactant molecules. In simplest terms, the micellar entity is considered to be a globular structure in which the alkyl hydrocarbon core region is surrounded by a hydrophilic outer region comprising the surfactant head group (which can be cationic, anionic, zwitterionic, or nonionic), counterions, and water molecules (Figure 1).[2,11,12,18,20,21] The minimum surfactant concentration required for micelle aggregate formation is referred to as the *critical micelle concentration* (CMC). At concentrations below this value in solution, the surfactant molecules exist as solvated monomers. The number of individual surfactant molecules present in a micelle aggregate is referred to as the *aggregation number* (N).[11,12,18] Typically, such aggregates contain 60–150 surfactant molecules. It should be stressed that the micellar species is not a permanent (bonded) entity, but a dynamic structure that exists in equilibrium with its monomers, with the monomeric

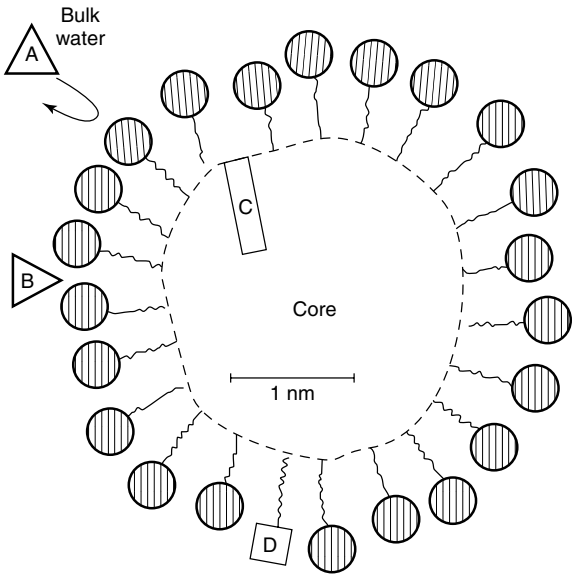

Figure 1 Artistic representation of an idealized cross section of an aqueous normal micelle. The polar headgroups (◯) and hydrophobic tails (–) are schematically indicated to show only their relative locations but not their actual numbers, distributions, or configurations (the counterions, which would be associated with an ionic micelle, are not shown). The dotted line encompasses most of the hydrophobic core region of the micelle. Analytes can bind to the micellar aggregate at different locations. For example, a charged analyte (A) would be electrostatically repelled from the micelle surface if it were of the same charge type as the ionic micelle. An oppositely charged analyte ion (B) would be electrostatically attracted to the micellar surface and ion exchange with the micelle counterion. A neutral hydrophobic analyte molecule (C) would partition towards the nonpolar core region of the micelle depending upon its specific hydrophobicity. A neutral amphiphilic solute (D) might attempt to align itself with the surfactant molecule so as to maximize the electrostatic, hydrogen-bonding, and/or hydrophobic interactions possible between itself and the micelle surfactant molecules. (Adapted with permission from Hinze et al.[20])

surfactant molecules both entering and exiting the micelle.[11,12,18,21,22]

The CMC values for surfactant micelles have been determined from breaks in plots of the luminescence intensity of an organic probe molecule as a function of surfactant concentration as a consequence of alteration of the microenvironment experienced by the probe upon formation of and binding to the micelle aggregate.[22–92] Table 1 includes the CMC values for some classic micelle-forming surfactants. CMC values can be altered owing to the presence of the analyte and any other components required to adjust the pH and/or ionic strength.[93–99] For instance, the CMC of sodium dodecyl sulfate (SDS) in water was determined as 0.1 g per 100 mL, whereas a value of 0.2 g per 100 mL was observed in the presence of 120 ppm of the analyte, fenproporex.[96] The CMC of SDS was likewise altered from 8.1 mM in water to 1.1 mM in the presence of 0.125 M NaCl;[22] whereas the CMC of cetyltrimethylammonium chloride (CTAC) (hexadecyltrimethylammonium chloride) was 1.3 mM in water, it was only 0.28 mM in an aqueous 0.010 M NaOH solution. For cetyltrimethylammonium bromide (CTAB) (hexadecyltrimethylammonium bromide), the CMC in water is 0.92 mM, whereas it is 0.31 mM in the presence of 1.0 M sulfuric acid.[100] The CMC of SDS increased from 8.1 to 15.0 mM in the presence of 6.0 M urea whereas that of Triton X-100 changed from 0.20 to 0.60 mM in the presence of 3.0 M urea.[73] In

Table 1 Summary of some pertinent properties of bulk solvent and organized assembly systems, including CMC values for micelle-forming surfactants/bile salts

	Medium	CMC (mM)	ε[a]	$E_T(30)$[b]	η (cP)[c]	Py^*[d]
Bulk solvents	Bulk water (H_2O)	–	78.54	63.1	0.89	1.79
	Bulk water (D_2O)	–	–	–	1.096	–
	Bulk dimethyl sulfoxide	–	46.68	45.1	1.10	1.95
	Bulk acetonitrile	–	37.50	45.6	0.38	1.79
	Bulk methanol	–	32.7	55.4	0.55	1.35
	Bulk ethanol	–	24.55	51.9	1.08	1.10
	Bulk 1-butanol	–	17.5	50.2	2.65	1.10
	Bulk benzene	–	2.28	34.3	–	1.05
	Bulk *n*-hexane	–	1.89	–	0.31	0.85
	Bulk ethylene glycol	–	38.7	–	18.5	1.64
Surfactant micellar systems	NaL	24.0	39	–	–	–
	SDS	8.1	32, 41.6, 45, >40, 48, 51–55, 56, 62	53.8, 57.5, 40, 53–61	3.6, 4, 6.9, 9–12, 15–37, 31, 17–50, 193	1.22
	Hexadecyltrimethyl-ammonium bromide (CTAB)	0.92	16, 33, 37	53.2	6.6, 17, 19, 24, 30, 39	1.42
	Hexadecyltrimethyl-ammonium chloride (CTAC)	1.3	28–34, 39	53–64	27, 31	–
	TDAB	3.6	28–32	53.3	22	1.54
	DTAC	15.0	24, 30–40, >40	54–64	–	1.55
	Brij-35	0.06–0.09	27–29	52.8	–	1.30
	Triton X-100	0.3	5.5–7.9, ≤15, 27–30	50.5, 53.0	28 ± 8, 60, 90	1.32
	Poly(ethylene glycol) (7.5) *n*-nonylphenyl ether	–	–	–	–	1.40
	Poly(ethylene glycol) (8.0) dodecyl ether ($C_{12}E_8$)	–	27–29	52.8	–	–
	Poly(ethylene glycol) (6.0) dodecyl ether ($C_{12}E_6$)	0.09	–	–	57 ± 7	–

(*continued overleaf*)

Table 1 (*continued*)

	Medium	CMC (mM)	ε[a]	$E_T(30)$[b]	η (cP)[c]	Py^*[d]
	Igepal CO-630	0.055	$\leq$12	–	–	1.18
	Sulfobetaine-12[e]	3.3	32	–	–	–
	Octyl β-D-gluco-pyranoside	25.0	20–30, 40	–	–	–
Bile salt micelles	Sodium taurodeoxy-cholate	2.3–4.0	–	–	–	0.72
	NaC	5.2–12.0	–	–	–	0.69–0.79
	NaDC	0.9–3.6	–	–	30–110	0.62–0.67
	NaTC	2.7–7.0	–	–	–	1.00–1.38
Vesicle/ liposome systems	P-LYSO[f]	–	39	–	–	–
	DPPC	–	2.4–2.9, 3, 5.5, 7, 15, 16, 17.9,43	43, 49.8[g], 47–50	18–30[h], 80, 70–120[h], 1.1–74[h], 94–940[h]	–
	DMPC	–	3.2–3.9, 16, 27	42, 49.8, 55.4	13–22[h], 53–80[h],	1.09
	PC[i]	–	6.5, 17.5, 25.6, 26–33, 28, 29–36, 43	–	4, 27, 59.2, 69–110[h]	1.18–1.32
	PC[i] – cholesterol (2 : 1 ratio)	–	28	–	40	–
	PC[i] – cholesterol (various ratios)	–	5.5	–	60–120[h]	–
	DHP	–	45, 51	–	–	–
	DODAB	–	45	–	–	–
CD systems	α-CD	–	10, 47.5	45, 57	2.6	1.78
	β-CD	–	41–72[j], 48, 49	57, 58	2.3	0.93
	γ-CD	–	55, 70, 74	–	–	0.94

[a] Dielectric constant (ε) for the micelle–water interface; data taken from the literature.[32–34,40,41,43,45,50,52,53,56,58,61,62,64,68,71,75,77,90]
[b] Refers to Dimroth and Reichardt's empirical solvent polarity scale; data taken from the literature. [23,28,29,37,48,52,66,67,69,71,89]
[c] Refers to the microscopic viscosity at 20–25 °C; data taken from the literature.[23,26,31,39,42,44,50–52,57,64,72,75,78,79,81–85,87,89] The microviscosity is temperature dependent; for instance, for SDS η varies from 6 cP at 30 °C to 13 cP at 10 °C and 17 cP at 0 °C; similarly, the values for CTAC are 18, 49, and 70 cP, respectively, for the same temperatures.[79]
[d] Refers to the Dong–Winnik Py^* solvent polarity scale based on the photophysical properties of pyrene. Values taken from the literature.[47,54,55,57,76,77,80,86]
[e] Refers to zwitterionic *N*-dodecyl-*N,N*-dimethyl-3-ammonio-1-propanesulfonate.
[f] Refers to γ-palmitoyl L-α-lysolecithin.
[g] Temperature was 50 °C.
[h] Value depends upon the temperature, with η decreasing with an increase in temperature, vesicle charge, nature and binding site of probe molecule, and whether small (sonicated) unilamellar vesicles (SUVs) or multilamellar vesicles (MLVs) are present.
[i] Refers to egg PC or egg lecithin.
[j] Depends upon the actual CD concentration; $\varepsilon = 72$ at [β-CD] = 1 mM and 41 at [β-CD] = 10 mM.
NaL, sodium laurate; SDS, sodium dodecyl sulfate; TDAB, tetradecyltrimethylammonium bromide; DTAC, dodecyltrimethylammonium chloride; NaC, sodium cholate; NaDC, sodium deoxycholate; NaTC, sodium taurocholate; DPPC, dipalmitoylphosphatidylcholine; DMPC, dimyristoylphosphatidylcholine; PC, phosphatidylcholine; DHP, dihexadecyl phosphate; DODAB, dioctadecyldimethylammonium bromide.

some cases, the presence of another hydrophobic species, particularly dye molecules, can induce micellization of surfactants.[101] Such aggregated species are sometimes referred to as *dye-rich induced micelles*.[102] Addition of CDs can also induce aggregation of ionic surfactants at concentrations below their normal CMC.[103] The CMC and aggregation number of aqueous normal micelles were found to be almost independent of pressure.[81] Typically, there is a slight increase in the CMC of ionic surfactants with increased temperature above ca. 30 °C.[94,97,104,105] For instance, the CMCs for CTAB, CTAC, and SDS, which are 0.92, 1.3, and 8.1 mM, respectively, at 25 °C, increased to 1.5, 2.0, and 10.4 mM, respectively, at 65 °C.[94,97,104] The addition of organic solvents to aqueous solutions can preclude the formation of surfactant micellar aggregates.

Typically, the presence of 20–40% (v/v) of organic cosolvents (such as methanol, ethanol, and acetonitrile) in water precludes micelle formation. In fact, this has been observed using fluorescent compounds, i.e. drastic reductions in fluorescence intensity (FI) in this range of added cosolvent.[106–113] Dilution of micelle-containing solutions with water can also result in destruction of the micelles if the final surfactant concentration is below the CMC for that surfactant.

A number of monographs provide a more complete description of surfactant and micellar systems and their properties and salient features.[6,11,12,18,21,94,114–116]

2.1.2 *Bile Salt Surfactant Aggregate Systems*

In comparison with the classical linear aliphatic surfactants, the bile salts have a different molecular structure and stereochemistry. The bile salts are rigid, facial amphiphilic molecules and do not contain the usual polar head group and apolar tail of the classic micelle-forming surfactants. Instead, they possess a steroidal ring structure and are crescent shaped, with the hydrophobic convex side containing the hydrocarbon elements of the steroid ring along with the angular methyl groups, while the more polar concave side bears the hydroxy substituents.[117,118] When bile salts are dissolved in water, they also form aggregate species termed micelles above the CMC, as do classical linear anionic surfactants. However, their aggregation behavior is more complex, showing more gradual (and successive) changes in aggregate size with concentration. Bile salt aggregates in aqueous solution consist of transient species (i.e. monomers, dimers, tetramers, ..., higher oligomers) in dynamic equilibrium with one another. The relative population of these structures is very sensitive to the conditions of the solution, being influenced by the concentration of the bile salt, the absence or presence of additives in the system, and the temperature.[118,119] Table 1 also provides the CMC values for some bile salt surfactant micelle systems.

Bile salt aggregation is viewed as stepwise formation of globular primary bile salt micelles consisting of 2–10 monomers held together by hydrophobic interactions between the nonpolar backs of the bile salt molecules which constitute the interior region with the more polar hydroxyl and carboxylate groups projecting outward and accessible to water molecules according to the Small model. In the next stage, at high bile salt concentrations (and/or added salt), larger secondary bile salt aggregates are formed by mutual association of the primary micelles as a consequence of intermolecular hydrogen bonding between the primary micelle's bile salt hydroxyl and side-chain groups.[117] In addition to this model, a helical model for bile salt aggregates has also been proposed.[119]

In general, it is thought that bile salt aggregate systems are more hydrophilic than conventional SDS micelles.[120] The bile salt micelle systems are less surface active than conventional aqueous normal micelles.[119] As a consequence, bile salt micelle solutions do not form suds/foam as much; this is very useful in the preparation and manipulation of such solutions, especially if gas purging is required as in some fluorescence measurements.[121] The micellar properties of bile salts, such as sodium taurocholate (NaTC), are relatively insensitive to experimental conditions.[122] Thus, compared with the normal aqueous micellar systems, the bile salt media are also much more versatile and flexible in terms of their tolerances of pH, ionic strength, temperature, and counterion concentrations and are less likely to precipitate.[121,123] Bile salt solutions can tolerate the presence of greater concentrations of additives, such as organic solvents, compared with conventional micelles such as SDS, without disruption of their aggregate structure.[119,124] They are therefore more robust aggregate systems than traditional aqueous normal micelles.[122] In addition, bile salts and their micellar aggregates are chiral entities, so they can function as a chiral selector and differentially bind the enantiomers of some organic compounds.

A monograph provides additional information on bile salt/acid systems and their utilization in chemical analysis applications.[125]

2.1.3 *Liposome/Synthetic Vesicle Systems*

Amphiphiles with long alkyl double chains such as the phospholipids or some synthetic amphiphiles form aqueous compartments enclosed by a lipid–surfactant bilayer, known as *vesicle* or *liposome* aggregate systems, when dispersed in aqueous solution. The hydrocarbon tails of the lipid or surfactant molecules are sequestered inside the bilayer, and the polar head groups stick out into the aqueous phase – inside and outside the compartment.[126] A macroscopically isotropic small unilamellar vesicle (SUV) can be formed via ultrasound sonication of a lipid or surfactant solution above its phase transition temperature.[127,128] Alternatively, SUVs can be prepared by injecting/infusing a small aliquot of an ethanolic solution of the lipid through a thin needle into a rapidly stirred buffer solution at a temperature 15 °C higher then the transition temperature of the phosphatidylcholine (PC) lipid employed.[129,130] The lamellar phase transition temperature (T_M) of dimyristoylphosphatidylcholine (DMPC),

dipalmitoylphosphatidylcholine (DPPC) and dimyristoylphosphatidylserine (DMPS) are 20–24, 34–42, and 31–39 °C, respectively.[71,129–134] The latter recipe produces SUVs having an average diameter of 25–50 nm.[129] A large unilamellar vesicle (LUV) can also be prepared from such a lipid by variation of the preparation conditions to yield vesicles (liposomes) with a radius in the range 0.05–5 μm.[18,130,135–137] Such monolamellar vesicles can also be prepared by the so-called detergent method.[138–140] Several reports provide recipes for the different methods of vesicle/liposome preparation, the type(s) of liposomes formed by each, and their advantages/disadvantages.[18,140–144]

In addition to the multilamellar vesicle (MLV) (multicompartment, onionlike structure) type, whose diameters are in the range 1000–8000 Å, the unilamellar systems can also be prepared from lipids such as egg PC, bovine phosphatidylserine (PS), and egg phosphatidylglycerol (PG).[18,132,145] Some synthetic surfactants can also produce bilayer aggregates.[18] In particular, double-chain ammonium amphiphiles such as didodecyldimethylammonium bromide (DDAB), dihexadecyl phosphate (DHP), and dioctadecyldimethylammonium chloride (DODAC) or dioctadecyldimethylammonium bromide (DODAB) also belong to a representative class of such bilayer-forming compounds.[146–150] The phase transition temperature for DODAC vesicles is 37 °C. For vesicle systems formed by dispersing lipid or synthetic surfactant molecules in aqueous medium, there are two types of water molecules present in the system – those in the bulk and those enclosed within the water pool of the vesicle (liposome).[131] In general, the water pool of vesicles is much larger in comparison with that in reverse micelles. Vesicles/liposomes are dynamic aggregate systems since they can exhibit phase transition and fusion behavior and are osmotically active.[18] For most analytical applications, the single-compartment bilayer liposomes (or vesicles) formed by the sonication method are probably ''best'' since they possess a more uniform size distribution and can be more reproducibly prepared. Purification of liposomes from the free monomeric lipid, detergent, and unencapsulated solutes can be accomplished either by gel filtration chromatography or dialysis.[137,141]

Although there are many similarities in properties between surfactant vesicles/liposomes and micelles, there are some important and notable differences.[18,137,151] Structurally, the interior volume of vesicle/liposome aggregates consists of an encapsulated aqueous solution in contrast to that of micelles whose interior comprises the hydrophobic alkyl chains of the surfactant. The surfactant vesicles are much larger, more rigid (less fluid), and more static aggregate entities, particularly at temperatures below their phase-transition temperature.[151] Although vesicle aggregate structures are in equilibrium with their monomer, in most instances, the monomer concentration is fairly low (in range 10^{-5}–10^{-10} M).[152] Thus, the rate of exchange of monomers between the bilayer and surrounding solution is slow (minutes to hours).[18,152] Once formed, the vesicle aggregates cannot be destroyed by dilution, whereas micelles are destroyed. Vesicles, owing to their dimensions, can solubilize a larger number of analyte molecules per aggregate entity compared to micelle systems. In addition, molecules/analyte species may be entrapped within the inner water pool of vesicles for extended periods of time (days and weeks).[147] Of course, one disadvantage of vesicle/liposome systems is that their preparation is more complicated and less convenient compared with that of micelles. In addition, one must be aware of the possibility of aging effects due to vesicle–vesicle fusion, etc.[149,151,153]

Several monographs are devoted to the topic of liposomes, their preparation, characterization, and medical-related applications.[141,154,155] In addition, several articles discuss, compare and contrast the features of vesicle-organized media relative to surfactant micelles and CDs with respect to their usefulness in luminescence analysis.[156,157]

2.2 Cyclodextrin Systems

CDs (also referred to as *cycloamyloses*) are naturally occurring macrocyclic oligomers. They consist of six or more linked glycopyranose moieties and are doughnut-shaped molecules. The CDs differ in the number of glucose units they possess, with the designation alpha- (α-CD), beta- (β-CD), and gamma-CD (γ-CD) containing six, seven, and eight such units, respectively. The internal diameters and depths of the CDs are 4.5–6.0 and 4.5 Å for α-CD, 6.0–8.0 and 7.0 Å for β-CD, and 8.0–10.0 and 7.0 Å for γ-CD. The interior of the torsos of CDs is relatively apolar relative to water and is therefore capable of including a variety of guest analyte molecules.[13,18] The interior volumes are 176, 346, and 510 $Å^3$ for α-, β-, and γ-CD, respectively.[13,158] For comparison, the micellar hydrophobic volume for SDS, CTAC, and Triton X-100 are 21 000, 50 400, and 73 600 $Å^3$, respectively.[159] Also, in contrast to bile salt and surfactant micellar and vesicle aggregate systems, the CDs are chemically and conformationally stable, permanent molecular entities. Thus, CDs are more robust and durable in comparison and can be employed in a variety of solvent systems since there is no monomer ⇔ aggregate equilibrium involved as in the formation of micellar aggregates. These cyclic oligosaccharides are well known to be able to include a wide variety of guest molecules in their toroidal cavity and alter the physicochemical

properties of such guest species.[13,18,160–162] During the past decade, a variety of modified CDs have been prepared, many of which exhibit enhanced (or more selective) binding ability including enantioselective binding.[163–165]

Several monographs provide much more detailed information on CDs along with their properties and applications.[13,160–162,165]

Table 2 compares and contrasts these different organized micelle, vesicle, and CD assembly systems in terms of some of their properties and features.[18,166–168] Additional information on and comparisons among these different membrane mimetic agents can be found in an excellent monograph by Fendler.[18]

2.3 Summary of Selected Properties of Organized Assemblies

The ability of organized assemblies to solubilize solutes, to alter their microenvironment and effective concentrations, and hence to affect their reactivities, product distributions, stereochemistries, ionization constants, and redox potentials has been exploited in a number of chemical analysis applications.[9–14,94,125,166] The influence of the medium upon the energy of electronic states and solvatochromic shifts has been the topic of several reviews.[1–8,167–170] Organized assembly systems solubilize and bind various types of organic molecules in aqueous solutions. Fluorescence probe techniques (*see* **Absorption and Luminescence Probes**) have provided much valuable information about the microenvironment of ordered media (i.e. their apparent viscosity, polarity, etc.), which in turn allows conclusions to be made concerning the structure, location, and dynamics of molecules solubilized in such organized systems.[1–8,18,171–176] This section will summarize some of these more important general features, as they pertain to enhancing the performance of analytical fluorescence measurements.

2.3.1 Solubilization Ability

As illustrated in Table 3, the aqueous solubility of sparingly water-soluble organic compounds can be increased by the addition of organized assemblies to the aqueous solution.[42,94,177–185] The solubility of hydrophobic compounds in water can be dramatically enhanced by the addition of a surfactant to

Table 2 General comparison of the organized assembly systems[a]

Characteristic feature/property	Aqueous normal micelles	Bile salt micelles	Vesicles/liposomes	CDs
Constituents	Various synthetic surfactants	Bile salt surfactants	Lipids and surfactants	Various CD molecules
Method of preparation	Dissolution of surfactant in water (the surfactant concentration must exceed the CMC)	Dissolution of bile salt in water (bile salt concentration must exceed the CMC)	Ultrasonication, injecting thin lipid films into water, injecting ether or alcohol solutions of lipids into water, gel filtration of lipid–detergent micelles	Dissolution of the CD in water
Weight-averaged molecular mass	2000–6000	–	$>10^7$	970–1300
Diameter (Å)	30–60	20–40	300–5000	4.7–8.3
General stability	Weeks, months	Weeks, months	Days, weeks	Months
Kinetic stability (exit rate of monomer, s)	ca. 10^{-5}	–	>Seconds	Not applicable, bonded entity
Effect of dilution with water	Micelle destroyed if the surfactant concentration is less than the CMC	Micelle destroyed if bile salt concentration is less than the CMC	Unaltered	Unaltered
Number of solute molecules taken up per assembly	Few	Few	Large	Typically 1 or 2
Solute solubilization site(s)	Surface, Stern layer, in vicinity of polar head groups	In vicinity of either polar or nonpolar areas, polar head group	Inner aqueous pool, either or both sides of the polar surface, within the bilayer	Interior cavity of CD or near hydroxyl groups of the outer rim of the CD molecule

[a] Data taken from the literature.[18,166–168]

Table 3 Selected solubility data for solutes in different organized assembly systems

Analyte molecule	Solubility in indicated medium	References
Coumarin 153	2.0×10^{-6} M (water)	42
	7.0×10^{-5} M (9 mM CTAB micelles)	
	1.0×10^{-3} M (40 mM SDS micelles)	
Coumarin C-120 and C-1[a]	$(1–2) \times 10^{-4}$ M (water)	177
	1.0×10^{-3} M (SDS micelles)	
Biphenyl	4.1×10^{-5} M (water)	178
	0.72 M (dodecane)	
	0.21 M (0.05 M SDS micelles)	
	1.00 M (0.05 M CTAB micelles)	
Argon gas (at 297.15 K and 101.325 kPa)	0.6161 cm^3 Ar/mol (water)	179
	23.4 cm^3 Ar/mol (micellar SDS)	
4-Cyano-4′-octaoxybiphenyl	2.7×10^{-7} M (water)	180
	5.29×10^{-2} M (SDS micelles)	
	1.01×10^{-1} M (CTAB micelles)	
	9.8×10^{-3} M (100 mM Igepal CO-730 micelles)	
Xanthione	2.5×10^{-7} M (water)	181
	4.9×10^{-6} M (0.01 M SDS micelles)	
	2.5×10^{-5} M (0.02 M SDS micelles)	
	6.7×10^{-5} M (0.04 M SDS micelles)	
	1.9×10^{-4} M (0.1 M SDS micelles)	
Naphthalene	$(2.2–2.6) \times 10^{-4}$ M (water)	182
	1.3×10^{-2} M (0.20 M NaC micelles, pH 8–9)	
	0.38 M (0.04 M SDS micelles)	178
	1.11 M (0.02 M CTAB micelles)	
Pyrene	$4.0–8.0 \times 10^{-7}$ M (water)	178
	7.1×10^{-2} M (hexane)	
	7.0×10^{-2} M (0.06 M SDS micelles)	
	4.1×10^{-1} M (0.04 M CTAB micelles)	
	9.8×10^{-5} M (3 mM CPC[b] micelles)	183
	2.8×10^{-4} M (10 mM CPC[b] micelle)	
	2.2×10^{-3} M (0.5 M potassium dodecanoate micelles)	94
Folic acid	9.7×10^{-5} M (water)	184
	1.8×10^{-2} M (cyclohexane)	
	4.2×10^{-4} M (10 mM SDS micelles)	
	9.8×10^{-4} M (100 mM SDS micelles)	
	2.1×10^{-2} M (100 mM CTAB micelles)	
	9.8×10^{-3} M (100 mM Igepal CO-730 micelles)	
Ferrocene	4.25×10^{-5} M (water)	185
	2.5×10^{-4} M (β-CD)	

[a] Coumarin C-120 refers to 4-methyl-7-aminocoumarin and coumarin C-1 refers to 4-methyl-7-dimethylaminocoumarin.
[b] CPC refers to cetylpyridinium chloride.

the aqueous solution.[21,94,177–203] Such enhanced solubilization is due to the formation of micelles by the surfactant, because below the CMC little increase in solubility is observed, yet above the CMC the increase in solubility is typically directly proportional to the surfactant concentration.[94,178,190] The absolute solubility of a solute in a micellar system depends upon the specific experimental conditions (surfactant concentration, presence and concentration of other additives, pH, temperature, and nature of the solute).[94,192] In addition, the solubilization power of surfactant aggregates depends upon the shape of that micelle entity (spherical, rodlike, etc.), with the solubilization capacity being proportional to the respective micellar volumes.[193] The uptake of

solutes into the micelle can change the physical state of the micelle and result in changes in the physical properties of the system.[94] As in the situation with the surfactant monomer, the solubilized solute is not rigidly fixed in the micelle. Instead, it can move about within the micelle aggregate and also is in constant dynamic equilibrium with the bulk aqueous phase. Almgren et al.[178] have described the dynamic and static aspects of solubilization of neutral solutes in ionic surfactant micellar solutions. In contrast to the solubility of organic hydrocarbons in hydrocarbon surfactant micelles, their solubility in perfluorocarbon micelles appears to be somewhat limited.[194]

As illustrated by the data in Table 3, the presence of micelle assemblies can enhance the solubility of sparingly soluble solutes by several orders of magnitude. As for some other examples, the solubility of phthalimides increases by approximately an order of magnitude in 0.10 M SDS micellar solutions compared with the solubility in pure water.[195] The aqueous solubility of α-terthiophene is $<10^{-7}$ M. However, aqueous solutions containing 38 μM α-terthiophene could be prepared in the presence of surfactant CTAC, dodecyltrimethylammonium chloride (DTAC), SDS, or Triton X-100 micellar media.[159] The presence of cationic CTAB micelles increases the solubility of folic acid in water by a factor of 216.[184] Most metal ion complexes of oxine (8-hydroxyquinoline) and its derivatives are insoluble in aqueous solution; however, these can be solubilized in surfactant micellar media.[196] Cholesterol monohydrate's solubility in aqueous micellar NaTC is ca. 90 times greater than that in bulk aqueous solution.[119,197]

Likewise, surfactant vesicle/liposome systems can enhance the solubility of organic compounds. For instance, the aqueous solubility of bilirubin-IXα is 9×10^{-8} M. However, concentrations up to 1.0×10^{-5} M were prepared in aqueous solutions containing 200 μM PC vesicles.[198] Bilirubin's solubility in the bile salt sodium cholate (NaC) micellar medium was reported as 30 mM.[125,199] Although the flavonoid baicalein as well as coomassie brilliant blue dye–CTAB detergent complexes are insoluble in water, they can be solubilized in phospholipid vesicle membrane media.[200,201] Nile red is sparingly soluble in water ($<1\,\mu g\,mL^{-1}$), but very soluble in aqueous micelle or vesicle solutions and organic solvents.[202] Fluorescence spectroscopy has been employed to study the solubilization of ketoconazole and griseofulvin in different bile salt micelles.[203]

The presence of CDs also generally increases the solubility of organic compounds in water, but this depends on the solute and the CD.[13,160,161] For instance, the solubilities of benzoic and aliphatic carboxylic acids were increased in the presence of α- and β-CDs, whereas the solubilities of carboxylic acids containing bulky groups were not appreciably affected.[161] The solubility of the flavin, lumichrome, and carbazole alkaloids in water increases with an increase in the concentration of β-CD.[204,205] The presence of 10 mM β-CD enhanced the aqueous solubility of *syn*-bimanes by ca. 10%.[206] In the presence of 3 mM β-CD, the solubility of lumichrome is threefold greater than in bulk water alone. However, the solubility of hydrophobic molecules in CDs is poor compared with other organized assemblies, such as aqueous micelles, and this factor is considered by some to be a serious disadvantage of CD systems.[207] In some cases, analyte-included CD complexes can precipitate from solution, especially from more concentrated stock solutions.[185,208] The solubility product and solubility of CD inclusion complex precipitates as well as requirements for formation of such precipitates have been reported.[209,210] CDs have been employed in preparation of pharmaceutical formulations not only to enhance solubility, but also to increase the permeation and bioavailability of drugs in biological systems.[211–213] However, the utilization of derivatized CDs, which are themselves more water soluble, generally improves their solubilization capacity. For instance, complexation of an indolinonic nitroxide radical with randomnly methylated β-CD served to improve not only the nitroxide's aqueous solubility, but also its photostability.[214]

In some situations, organized assemblies have been employed merely for solubilization purposes in fluorescence assays. For example, the surfactant Triton X-100 was employed in the isolation of membrane-bound proteins owing to its ability to disrupt membranes and solubilize proteins without denaturation. The subsequent assay of the protein was achieved by reacting it with fluorescamine and monitoring the FI (excitation and emission wavelengths, $\lambda_{ex} = 390$ nm, $\lambda_{em} = 475$ nm) of the product.[215] The presence of the Triton X-100 solubilization agent did not affect the fluorimetric measurements.

A sensitive fluorescent assay for lipase activity utilized egg PC liposomes or a mixed bile salt/liposome system to solubilize the nonfluorescent substrate 4-methylumbelliferyl butyrate.[216] Triacylglycerol lipase was incubated in this medium, and after termination of the analytical reaction the FI measured ($\lambda_{ex} = 365$ nm, $\lambda_{em} = 450$ nm). The detection limit of lipolytic activity was $10\,\mu M\,h^{-1}$ in the mentioned organized assembly medium.[216]

A simple, sensitive, and rapid assay for vitamin K_1 was developed which uses the photo-oxidation of glucose sensitized by vitamin K_1.[217] The hydrogen peroxide generated in the photochemical reaction was determined through the iron(II)-catalyzed oxidation of benzoic acid

to form hydroxybenzoic acid isomers, which fluoresce. The use of Triton X-100 micelles was employed in order to overcome the problem of the poor water solubility of the phylloquinone. In addition, in the absence of TX-100, the ferrous ion is rapidly oxidized to ferric ion, which is unstable in basic aqueous media and slowly precipitates (owing to the formation of the oxide and/or hydroxide salts), which causes signal instability and clogs the flow-injection analysis (FIA) tubing. The presence of the Triton X-100 micellar medium successfully avoids these deleterious effects and problems.[217] CTAC micelles were incorporated into a fluorescent method for the simultaneous determination of fat-soluble (vitamins A and E) and water-soluble vitamins (B_1, B_2, and B_6)[187] (*see* **Vitamins: Fat and Water Soluble, Analysis of**).

Hence, the solubilization property of organized assemblies allows one to work directly in an aqueous phase without the need for organic or mixed aqueous–organic solvents. In fact, the utilization of organized assembly systems and their solubilization ability can aid analyte sample preservation and storage[218] (*see* **Sampling and Sample Preparation in Process Analysis**; **Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices**). Two of the more frequent and serious difficulties encountered when working with real-world samples concern analyte loss due to its adsorption on the surfaces (walls) of the storage container or to its interaction with organic particulate matter or soil sediments (humic materials, bentonite, etc.). The use of surfactant micellar media can obviate both problems. For example, one study found that the presence of Brij-35 surfactant in the sample solution was as effective as 40% (v/v) acetonitrile in preventing loss of polycyclic aromatic compounds (PACs) due to adsorption on the surface of borosilicate glass, white polyethylene, or polytetrafluoroethylene (PTFE) containers.[219,220] The addition of nonionic Triton X-114 was also found to enhance PAC recovery during sample preconcentration.[221] In a Brij-35-modified procedure for determination of 13 polycyclic aromatic hydrocarbons (PAHs), it was reported that the Brij-35 micellar solutions containing these PAHs were stable for at least 24 h, except for those of naphthalene, anthracene, and acenaphthene, whose FI decreased ca. 15, 17, and 12%, respectively, over 24 h.[222] Enhanced stability of PAHs in the presence of the nonionic surfactant polyoxyethylene-10-lauryl ethers has also been reported.[223] With biological/clinical samples, it was noted that the addition of SDS was effective in preventing the sticking of protein to the tubes and pipet tips, leading to more reproducible results in a protein fluorescence quantitation based upon reaction with *o*-phthaldialdehyde.[224] The nonionic surfactant polyethylene (20) sorbitan monolaurate was demonstrated to prevent DNA adhesion to the surface of the containers.[225] Likewise, the use of micellar sodium dodecylbenzenesulfonate (SDDBS) prevented the adsorption of different metal–porphine chelates on glassware surfaces.[226]

It has been reported that the presence of CDs can prevent (or reverse) the effects of sample adsorption of analytes to the walls of the container. For example, addition of β-CD led to the dissolution of pyrene that had been adsorbed on the walls of the glassware.[227] Addition of solutions of CDs also successfully sequestered (redissolved) organic molecules from the walls of glass tubes on which they had been deposited into the aqueous solutions.[228] Memory effects due to adsorption of metal ions, such as Eu^{3+}, on the walls of glass containers during storage of stock solutions have been reported.[229] Such effects might be diminished or eliminated if surfactant micelles or CDs were added and present in the stock solution.

2.3.2 *Solute Incorporation and Binding to the Organized Assembly*

Solute binding to organized assemblies has resulted in widespread applications of such systems in pharmaceutical chemistry, catalysis, chemical analysis and separation science, and food technology, among others.[9–13,18,94,160,230]

The binding association of a solute molecule with a micelle may be divided into several stages corresponding to the different sites of possible interaction: the bulk solution, the micelle outer surface region, and the more hydrophobic micelle interior (refer to possible binding sites as depicted in Figure 1). In the initial stage, the rate of interaction is the bulk solution encounter rate, which is typically near the diffusion-controlled limit (i.e. ca. $3 \times 10^9\ dm^3\ mol^{-1}\ s^{-1}$),[178,231–233] particularly when the solute and micelle species are of opposite charge. Solute protonation or deprotonation may occur in the second step since the micelle surface region pH is typically different, sometimes by ca. 2 pH units, from that of the bulk aqueous solution phase.[233,234] The rate of proton transfer in micelle systems is also typically diffusion controlled, so that this is not ordinarily the rate-determining step.[233,235] The final stage of incorporation represents the complex set of interactions resulting from hydrophobic and/or electrostatic forces as the solute occupies its preferential averaged-binding site, either in the micellar core or surface regions or in some intermediate position. This last step is typically the rate-limiting one for solute incorporation. For instance, when neutral dye molecules, such as methyl red (MR), methyl orange (MO), or pyridine-2-azo-*p*-dimethylaniline (PADA), are incorporated into SDS or CTAB micelles, the rate constants are in the

10^6 dm^3 mol^{-1} s^{-1} range.[233] However, the rate constants increase to the diffusion-controlled limit in the cases where these dyes or organic molecules and the micelles involved are oppositely charged. This was rationalized in terms of the electrostatic forces becoming dominant in the final interaction step. Similarly, the rate can be increased by increasing the hydrophobicity of the solute molecule or the micelle.[233] As noted by Quina, ''small molecules and ions can migrate rapidly in and out of the aggregates, while those incorporated in the microaggregate can migrate within the aggregate.''[176] Solute incorporation rate constants (entrance rates) are typically in the range of 10^8–10^{10} M^{-1} s^{-1}, while the solute rate constants for exit from the micelle (exit rates) are in the range 10^2–10^8 s^{-1} depending upon the specific solute–surfactant micelle combination.[236] Since the entrance rate constants are generally near diffusion controlled, it is the exit rate constant that typically dictates the magnitude of the degree of solute binding to the micelle.[236,237]

Solute complexation with CDs is also a dynamic process.[13,18] The entry and exit rate constants for the complexation of naphthylethanols with β-CD have been determined.[238] Rate constants for entry and exit of guest molecules into such CD hosts are typically of the order of 10^8–10^9 M^{-1} s^{-1} and 10^5–10^7 s^{-1}, respectively (for 1 : 1 complexation stoichiometry).[18,238] To a first approximation, the strength of binding correlates with the fit of the solute molecule in the CD cavity.[13,18] The interior of the doughnut-shaped CD molecule is lined with oxygen and CH groups, providing a relatively hydrophobic binding site for molecules of appropriate size that can fit the particular CD's cavity. One exterior side of the torus contains the primary hydroxyl groups, while the other side contains a number of secondary hydroxyl groups on the C-2 and C-3 atoms of the glucose moieties. Thus, these hydroxyl groups offer hydrogen-bonding sites to the portion of a solute molecule that possesses an appropriate hydrogen-bonding moiety.[18] Since CDs (and also the bile salt micelles) are chiral entities, differences can be observed in the binding constants for an enantiomeric pair of analytes for these two organized assembly systems.

In synthetic surfactant vesicle or natural liposome aggregate systems, there are several possible analyte binding sites that exhibit a range of polarities. At the water-pool–head-group interfacial regions, the sites are relatively hydrophilic, whereas more hydrophobic (and ordered) sites are available within the hydrocarbon interior of the bilayers.[18] Of course, between these two extremes lie sites of intermediate hydrophobicity. Figure 2 depicts some of these potential binding sites for a liposome system, illustrating the binding of the relatively

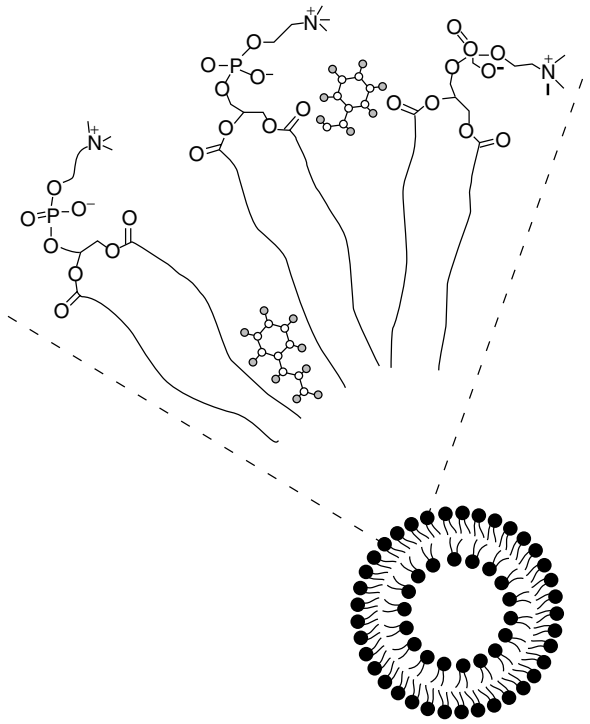

Figure 2 Artistic representation showing the bilayer structure of an entire small unilamellar vesicle (liposome) of egg yolk PC, along with an upper expanded view of possible solute binding sites. A more polar solute, such as benzyl alcohol, would be located near the surface phospho head group areas, whereas the less polar *n*-propylbenzene prefers to bind to the more hydrophobic interior region. (Adapted with permission from E. Okamura, M. Nakahara, *J. Phys. Chem.*, **103**, 3505–3509 (1999). Copyright 1999, American Chemical Society.)

polar benzyl alcohol and less polar *n*-propylbenzene molecules to an egg yolk PC SUV.[239] Consequently, it should not be surprising that, depending on the luminescent probe employed and its specific binding site, a number of possibilities for interaction of solutes of differing hydrophobicity (or polarity) within the vesicle interior exist.[240–243]

The fluorescence technique has proven to be a very useful means by which to determine the binding constants for equilibrium interaction of an analyte, A, with a particular organized assembly, OA, system Equation (1):[13,173,244–249]

$$A_W + OA \xleftrightarrow{K_B} A_{OA} \tag{1}$$

The equilibrium or binding constant, K_B (M^{-1}), is given by Equation (2):

$$K_B = \frac{[A_W]}{[A_{OA}]}[OA] \tag{2}$$

where $[A_W]$ and $[A_{OA}]$ denote, respectively, the analyte substrate molar concentration in the aqueous and organized assembly phases, and $[OA]$ is the concentration of the organized assembly under consideration. For micellar systems, the concentration of micellized surfactant is equal to the total surfactant concentration minus that surfactant's CMC. For micellar media, this solute–micelle binding constant, K_B, can be related to the partition (or distribution) coefficient, P, for distribution of the solute between the micellar and bulk aqueous phases, by Equation (3):

$$K_B = (P - 1)v \quad (3)$$

where v is the partial molar volume of the surfactant.[12,248,250]

The literature contains a plethora of data regarding the interaction and binding of solute (analyte) species to organized assembly systems.[249–317] Some representative data (in the form of binding constants or partition coefficients) for the interaction of ions and organic species with different organized media is presented in Table 4. As can be observed, the same analyte can bind to greatly differing extents to different organized assembly systems. Orders of magnitude differences in binding constants can exist for the same general type of organized assembly system. Solute binding constants and residence times in aqueous micelles are of the order of 10^3–$10^6\,M^{-1}$ and 10^3–$10^5\,s^{-1}$, respectively.[18] For example, the binding of pyrene to different surfactant/bile salt micelles ranges from ca. $1.5 \times 10^5\,M^{-1}$ for NaC to $1.0 \times 10^7\,M^{-1}$ for

Table 4 Summary of binding constants (or partition coefficients) for interaction of selected analytes with various organized assemblies

Analyte molecule	Organized medium	Binding constant, K_B (or P^a)	References
5-Aminoindazole	Aqueous. SDS micelles, pH 9.0, 25 °C	$24\,M^{-1}$	249
	Aqueous. CTAB micelles, pH 9.0, 25 °C	$30\,M^{-1}$	
5-Aminoindazole anion	Aqueous. CTAB micelles, pH 9.0, 25 °C	$75\,M^{-1}$	249
Xanthione	Aqueous. SDS micelles	$5.8 \times 10^5\,M^{-1}$	181
4-Aminobenzoic acid	α-CD	$631\,M^{-1}$	253
Benzoylacetone (keto form)	β-CD	$40\,M^{-1}$	252
Benzoylacetone (enol form)	β-CD	$314\,M^{-1}$	
Benzoylacetone (enol form)	SDS micelles	$235\,M^{-1}$	
Naphthalene	α-CD	$83\,M^{-1}$	251
	β-CD	$630–900\,M^{-1}$	207, 251, 254
	β-CD + 20 mM sodium butylsulfate	$320\,M^{-1}$	207
	γ-CD	$130\,M^{-1}$	251
	SDS micelles	$350–400; 2500\,M^{-1}$	178, 255
	CTAB micelles	$1500–9100\,M^{-1}$	
	NaC micelles	$2000–4800\,M^{-1}$	256
2-Methylnaphthoate	α-CD	$200\,M^{-1}$	59
	β-CD	$1965\,M^{-1}$	
	γ-CD	$213\,M^{-1}$	
1-Aminonaphthalene	CTAB micelles	$628\,M^{-1}$	74
Oxygen	SDS micelles	$300\,M^{-1}$	18
Fluorene	β-CD	$1490\,M^{-1}$	254
	γ-CD	$178\,M^{-1}$	
3-Aminofluoranthene	CTAB micelles	$6300\,M^{-1}$	74
Pyrene	β-CD	$160–277\,M^{-1}$	254, 257
	β-CD + 1% 2-propanol	$22\,700\,M^{-1}$	
	β-CD	$44\,M^{-1}$	207
	β-CD + 20 mM sodium butylsulfate	$3000\,M^{-1}$	
	γ-CD	$300–560\,M^{-1}$	254, 257
	γ-CD + 1% 1-propanol	$4500\,M^{-1}$	
	γ-CD + 10% t-butanol	$6760\,M^{-1}$	183
	CTAB micelles	$1.02 \times 10^7\,M^{-1}$	178
	SDS micelles	1.7×10^5–$2 \times 10^6\,M^{-1}$	18, 178
	CPC micelles	$5.0 \times 10^5\,M^{-1}$	183
	DHP vesicles	$(3.4 \times 10^5)^a$	258
	n-β-Octylglucoside micelles	$(1.5 \times 10^4)^a$	40
	NaC micelles	$(1.5–4.7) \times 10^5\,M^{-1}$	256

Table 4 (*continued*)

Analyte molecule	Organized medium	Binding constant, K_B (or P^a)	References
9-Aminophenanthrene	CTAB micelles	$8000\ M^{-1}$	244
3-Nitroanisole	β-CD	$99\ M^{-1}$	259
	CTAC micelles	$80\ M^{-1}$	260
2-Bromo-5-nitroanisole	CTAC micelles	$280\ M^{-1}$	260
p-Aminobenzoic acid	α-CD	$503\ M^{-1}$	261
	β-CD	$183\ M^{-1}$	
Sodium salicylate	β-CD	$100\ M^{-1}$	262
Methyl salicylate	α-CD	$20\ M^{-1}$	263
	β-CD	$156\ M^{-1}$	
	γ-CD	$47\ M^{-1}$	
Sodium decyl sulfate	β-CD	$3446–8750\ M^{-1}$	264, 265
SDS	β-CD	$8150–25\,600\ M^{-1}$	
TDSS	β-CD	$14\,561–48\,200\ M^{-1}$	
Iodide ion	CTAB micelles	$(368)^a$	266
	α-CD	$12.4–32\ M^{-1}$	319
	β-CD	$6.7–18\ M^{-1}$	
Phenol	CTAB micelles	$68–200\ M^{-1}$ $(420)^a$	267
	SDS micelles	$9–12\ M^{-1}$	
Phenol	Brij-35 micelles	$208\ M^{-1}$	
	β-CD, 28 °C	$40–118\ M^{-1}$	255, 268, 269
	SDS micelles	$(52.4)^a$	270
3-Nitrophenoxide	CTAC micelles	$1500\ M^{-1}$	260
1-Naphthol	Triton X-100 micelles, 10 mM NaOH	$41\ M^{-1}$	245
	β-CD	$2310\ M^{-1}$	271
	β-CD in 1 : 1 MeOH–water	$859\ M^{-1}$	272
	β-CD in 1 : 1 EtOH–water	$171\ M^{-1}$	
	SDS micelles	$104\ M^{-1}$	255
2-Naphthol	β-CD, pH 6.2	$526\ M^{-1}$	269
	α-CD, pH 6.2	$126\ M^{-1}$	
	γ-CD, pH 6.2	$53\ M^{-1}$	
	SDS micelles	$102–240\ M^{-1}$	255
	DTAB micelles	$120\ M^{-1}$	
4-Cyano-4′-octaoxybiphenyl	SDS micelles	$3.15 \times 10^6\ M^{-1}$	180
	CTAB micelles	$6.60 \times 10^6\ M^{-1}$	
Ferrocene	α-CD	$1.39 \times 10^2\ M^{-1}$	180
	β-CD	$1.65 \times 10^4\ M^{-1}$	
	γ-CD	$9.04 \times 10^2\ M^{-1}$	
$Ru(bipy)_3^{2+}$ [b]	SDS micelles	$3.5 \times 10^3\ M^{-1}$	271
$Ru(phen)_3^{2+}$	Triton X-100 micelles	ca. $200\ M^{-1}$	272
$Ru(phen)(CN)_4^{2-}$	CTAB micelles	$3.4 \times 10^5\ M^{-1}$	273
$Ru(phen)_2(CN)_2$	CTAB micelles	$1.3 \times 10^3\ M^{-1}$	
$Ru(Me_4phen)_3^{2+}$	CTAB micelles	$3.4 \times 10^2\ M^{-1}$	
	Triton X-100 micelles	$1.0 \times 10^5\ M^{-1}$	273
Sodium cation	SDS micelles	$1.0\ M^{-1}$	274
Terbium(III) cation	SDS micelles	$500\ M^{-1}$	275
	NaTC micelles	$157\ M^{-1}$	276
	α-CD, 0.10 M NaCl	$602\ M^{-1}$	277
	α-CD, pH 12.3	$17\,400\ M^{-1}$	
Eu^{3+}	α-CD, 0.10 M NaCl	$400\ M^{-1}$	277
La^{3+}	α-CD, 0.10 M tetra-methyl-ammonium chloride	$7080\ M^{-1}$	277
	β-CD, 0.10 M tetra-methyl-ammonium chloride	$5900\ M^{-1}$	
ANS[c]	PC Liposomes[d]	$(17–25) \times 10^4\ M^{-1}$	278
	α-CD	$5\ M^{-1}$	279

(*continued overleaf*)

Table 4 (*continued*)

Analyte molecule	Organized medium	Binding constant, K_B (or P^a)	References
	β-CD	$65–110\ M^{-1}$	279, 280
	γ-CD	$1260\ M^{-1}$	279
	SDS micelles	$8410\ M^{-1}$	25
	Sodium decyl sulfate micelles	$7100\ M^{-1}$	
	DPPC vesicles	$1.4 \times 10^5\ M^{-1}$	281
Cardiotoxin	PS vesicles	$2 \times 10^7\ M^{-1}$	282
N-Methylaniline	β-CD, 28 °C	$47.6\ M^{-1}$	268
N,N-DMA	β-CD, 25/28 °C	$230–272\ M^{-1}$	268, 283
	α-CD, 25 °C	$103\ M^{-1}$	283
	γ-CD, 25 °C	$64\ M^{-1}$	
	Egg PC vesicles	(20)[a]	284
N,N-Diethylaniline	β-CD, 28 °C	$960\ M^{-1}$	284
	Phospholipid liposomes	(383)[a]	285
	DMPC vesicles	(ca. 160)[a]	
Aniline	β-CD, 28 °C	$50\ M^{-1}$	255
Doxorubicin	β-CD, pH 7.1	$130–210\ M^{-1}$	285
	γ-CD, pH 7.1	$235–350\ M^{-1}$	
Quinine	β-CD, 20 °C	$1200\ M^{-1}$	286
	NaTC micelles, pH 7.4	(1.8)[a]	287
	NaGDC micelles, pH 7.4	(8.82)[a]	
Methyl-*p*-dimethylamino-benzoic acid	β-CD	$580\ M^{-1}$	288
o-Hydroxybenzoic acid (neutral)	β-CD, 15 °C	$980\ M^{-1}$	289
	β-CD, 25 °C	$740\ M^{-1}$	
	β-CD, 40 °C	$440\ M^{-1}$	
(anion form)	β-CD, 25 °C	$90\ M^{-1}$	
	β-CD, 40 °C	$37\ M^{-1}$	
Erythrosin B	α-CD	$10–20\ M^{-1}$	290
	β-CD	$20–40\ M^{-1}$	
	γ-CD	$50–110\ M^{-1}$	
Rose Bengal	γ-CD	$40–100\ M^{-1}$	290
	CTAB micelles	$\geq 10^4\ M^{-1}$	291
	SDS micelles	$\geq 10^2\ M^{-1}$	
Merocyanine 540	SDS micelles	$1.4 \times 10^5\ M^{-1}$	292
4-Aminophthalimide (4AP)	β-CD	$208\ M^{-1}$	293
	SDS micelles	$3400\ M^{-1}$	294
	CTAB micelles	$4500\ M^{-1}$	
	Triton X-100 micelles	$5600\ M^{-1}$	
11-(4AP)undecanoic acid	Triton X-100 micelles	$8500\ M^{-1}$	294
	CTAB micelles	$8000\ M^{-1}$	
	SDS micelles	$4200\ M^{-1}$	
7-Ethoxycoumarin	SDS micelles	$2.8 \times 10^4\ M^{-1}$	61
	CTAC micelles	$3.5 \times 10^4\ M^{-1}$	
Benzene	SDS micelles	$19–33\ M^{-1}$	178, 255
	CTAB micelles	$36–40\ M^{-1}$	
	Brij-35 micelles	$34–41\ M^{-1}$	255
	NaC micelles	$62–140\ M^{-1}$	271
	α-CD	$17\ M^{-1}$	210
	β-CD	$120\ M^{-1}$	
	γ-CD	$12\ M^{-1}$	
	PC, DMPC or DPPC vesicles	(3000–4000)[a]	295
Adriamycin	DMPC vesicles, 37 °C	$200\ M^{-1}$	296, 297
	Dimyristoylphosphatidylglycerol vesicles, 37 °C	$3400\ M^{-1}$	

(*continued overleaf*)

Table 4 (*continued*)

Analyte molecule	Organized medium	Binding constant, K_B (or P^a)	References
Daunomycin	DMPC vesicles, 1 M NaCl	$2.5 \times 10^3\ M^{-1}$	297
	DPPC vesicles, 1 M NaCl	$6.3 \times 10^3\ M^{-1}$	
	DMPC vesicles, 35 °C	$(5.5)^a$	298
	DMPC vesicles, 37 °C	$1400\ M^{-1}$	297
	DMPG vesicles, 37 °C	$9000\ M^{-1}$	
Carvedilol	DMPC–DMPG mixed vesicles, pH 7.4	$9200\ M^{-1}$	299
Naproxen	β-CD	$480\ M^{-1}$	300
Carbocyanin dye, diS-C_3-(5)	Egg lecithin vesicles, pH 7.5	$(2.4 \times 10^6)^a$	138
DPH	DPPC vesicles	$2.85 \times 10^6\ M^{-1}$	281
Carbon tetrachloride	CTAC micelles	$(3.5 \times 10^3)^a$	146
	DODAC vesicles (SUV)	$(4.7 \times 10^3)^a$	
	SDS micelles	$14–50\ M^{-1}$	268
	CTAB micelles	$100\ M^{-1}$	
Heptanol	DODAC vesicles (SUV)	$(8.5 \times 10^3)^a$	146
	Egg yolk lecithin liposomes	$(171.4)^a$	270
	α-CD	$2.46 \times 10^3\ M^{-1}$	301
	β-CD	$985\ M^{-1}$	
	γ-CD	$37\ M^{-1}$	
Heptanol	NaDC micelles, pH 8.2	$(2000)^a$	302
	SDS micelles	$88–147\ M^{-1}$	268
	DTAB micelles	$72\ M^{-1}$	
Octanol	DODAC vesicles (SUV)	$(17 \times 10^3)^a$	146
	α-CD	$4.82 \times 10^3\ M^{-1}$	301
	β-CD	$1.91 \times 10^3\ M^{-1}$	
	γ-CD	$67\ M^{-1}$	
	SDS micelles	$274\ M^{-1}$	255
1-Pentanol	SDS micelles	$(46.6)^a$	270
	Egg yolk lecithin liposomes	$(19.5)^a$	
	β-CD	$63.1\ M^{-1}$	303
4-Methyl-7-aminocoumarin	SDS micelles	$2 \times 10^4\ M^{-1}$	177
Indole	β-CD	$184\ M^{-1}$	304
	β-CD + 36% $CaCl_2$	$6085\ M^{-1}$	
3-Methylindole	Brij-35 micelles	$2.6 \times 10^4\ M^{-1}$	305
DM3HI[e] (neutral form)	SDS micelles, pH 9.5	$1010\ M^{-1}$	306
(monocation)	SDS micelles, pH 1.0	$12\,900\ M^{-1}$	
(neutral form)	CTAB micelles, pH 9.5	$2670\ M^{-1}$	
Indole-2-carboxylic acid	CTAB micelles, pH 10	$1.71 \times 10^5\ M^{-1}$	307
Indole-3-carboxylic acid		$4.4 \times 10^4\ M^{-1}$	
Indole-4-carboxylic acid		$4.0 \times 10^4\ M^{-1}$	
Indole-5-carboxylic acid		$6.6 \times 10^3\ M^{-1}$	
Auramine-O	β-CD	$197\ M^{-1}$	308
	α-CD	$5000\ M^{-1}$	
	SDS micelles	$1260\ M^{-1}$	
	Brij-78 micelles	$1000\ M^{-1}$	
Estradiol	β-CD, 35 °C	$3700\ M^{-1}$	208
	γ-CD, 35 °C	$6350\ M^{-1}$	
Procaine	β-CD	$137\ M^{-1}$	309
Dibucaine	CTAC micelles	$2000\ M^{-1}$	310
(neutral form)	SDS micelles	$2300\ M^{-1}$	
(protonated cationic form)	CTAC micelles	$150\ M^{-1}$	
	SDS micelles	$>25\,000\ M^{-1}$	
Triflupromazine	PC vesicles	$(2.1 \times 10^5)^a$	128
1,4-Dihydroxyanthraquinone	β-CD, pH 3.4	$300\ M^{-1}$	85
	β-CD, pH 7.1	$200\ M^{-1}$	
Curcumin	Triton X-100 micelles	$4.9 \times 10^3\ M^{-1}$	311
Hemoglobin	PC liposomes + 10 mol% cardiolipin	$2.5 \times 10^5\ M^{-1}$	136

(*continued overleaf*)

Table 4 (*continued*)

Analyte molecule	Organized medium	Binding constant, K_B (or P[a])	References
Benzimidazole (neutral form)	CTAB micelles	$53\,M^{-1}$	98
(neutral form)	SDS micelles	$55\,M^{-1}$	
(anion)	CTAB micelles	$3.6 \times 10^3\,M^{-1}$	
(cation)	SDS micelles	$3.7 \times 10^2\,M^{-1}$	
Tetracycline	Brij-35 micelles, pH 3.0	(8.64)[a]	94
	SDS micelles, pH 3.0	(2690)[a]	
	DTAC micelles, pH 3.0	(13)[a]	
	Brij-35 micelles, pH 3.9	(6.31)[a]	
	SDS micelles, pH 3.9	(1130)[a]	
	DTAC micelles, pH 3.9	(15)[a]	
2-(5-Bromo-2-pyridylazo)-5-diethylaminophenol	β-CD, pH 9.0	$93.0\,M^{-1}$	312
	β-CD, pH 9.0, 20% ethanol	$92.6\,M^{-1}$	
	β-CD, pH 9.0, 40% ethanol	$22.8\,M^{-1}$	
Phenolphthalein	β-CD, pH 10.5, 10 °C	$7.44 \times 10^4\,M^{-1}$	313
	β-CD, pH 10.5, 50 °C	$0.76 \times 10^4\,M^{-1}$	
Phenolphthalein	β-CD, pH 10.5, 10 °C, + 37 mM tetrahydrofuran	$1.64 \times 10^4\,M^{-1}$	
	β-CD, pH 10.5, 50 °C, + 37 mM tetrahydrofuran	$0.29 \times 10^4\,M^{-1}$	
Acridine orange	β-CD, pH 7.0	$825–926\,M^{-1}$	314
MB	β-CD, pH 7.0	$621–714\,M^{-1}$	314
	β-CD, pH 7.2	$4.50 \times 10^3\,M^{-1}$	315
Cyclooctatetraene	SDS micelles	$1500\,M^{-1}$	316
Xanthone	α-CD	$50\,M^{-1}$	317
	β-CD	$1100\,M^{-1}$	
	γ-CD	$200\,M^{-1}$	
Rhodamine 123	TX-100	$65\,000\,M^{-1}$	237
	Brij-35	$19\,000\,M^{-1}$	
1-Naphthylacetate	β-CD, 25 °C	$252\,M^{-1}$	317
2-Naphthylacetate	β-CD, 25 °C	$926\,M^{-1}$	
B-Lapachone	β-CD	$1100\,M^{-1}$	213

[a] P refers to the partition (or distribution) coefficient, with values given in parentheses.
[b] Refers to tris(2,2′-bipyridinium)ruthenium(II) cation.
[c] Refers to 1-anilonaphthalene-8-sulfonate.
[d] Refers to egg PC.
[e] Refers to 3,3-dimethyl-2-phenyl-3*H*-indole.
DPH, 1,6-diphenyl-1,3,5-hexatriene; DTAB, dodecyltrimethylammonium bromide; DMA, dimethylaniline; SB-12, sulfobetain-12; phen, 1,10-phenanthroline; ANS, 8-anilino-1-naphthalenesulfonate, MB, methylene blue; TDSS, sodium tetradecyl sulfate.

binding to CTAB micelles (Table 4). A comprehensive compilation of solute–micelle binding constants (partition coefficients) is available.[318]

In the case of micellar media, linear solvation free energy relationships of the form shown in Equation (4) have been reported:

$$\log K_B = c + a\sum\alpha + b\sum\beta + s\pi_2 + rR_2 + v\left(\frac{V_x}{100}\right) \quad (4)$$

in which the binding of solutes to a specific surfactant micelle is related to the medium-independent parameters for solute hydrogen bond acidity $\left(\sum\alpha\right)$ and basicity $\left(\sum\beta\right)$, excess molar refraction (R_2), dipolarity (π_2), and volume (V_x).[255] For the cationic surfactants DTAB and CTAB, nonionic surfactant Brij-35, and anionic SDS, solute incorporation and binding to the micelle is dominated by the V_x term (positive, reflecting the hydrophobic effect) and the $\sum\beta$ terms (negative, implying that bulk water is a better hydrogen-bond donor than the micellar solubilization site).[255] The contributions of $\sum\alpha_2$ and R_2, although smaller, were reported to vary in a chemically satisfying manner with the surfactant charge and structure.[255] Such linear salvation free energy relationship (LSER) relationships allow the determination of the relative importance of these factors in binding to a specific surfactant micelle and also allow the estimation of the actual solute binding constant for a specific surfactant micelle provided that the solute's

required parameters are known. Such an approach should also be applicable to synthetic surfactant vesicle or natural liposome aggregate systems. A similar approach utilizing the transition energy of fluorescent probes in CTAB, SDS, and Triton X-100 was conducted, and the aforementioned solvatochromic parameters determined for these three micellar systems.[320] However, the values were determined using a rather limited number of probe molecules.

Table 4 also includes solute–CD binding constants for formation of 1 : 1 inclusion complexes. Several monographs and reviews summarize methods utilized to determine solute–CD binding constants and also provide binding constant data for a wide range of solutes.[13,253,296,319,321,322] On the basis of the examination of 569 such complex systems, Equation (5) was found to provide reasonable estimates of the magnitude of the CD binding interaction:

$$\log K_B = -1.74 - [Z] + 0.032(\Delta A) \qquad (5)$$

where $[Z]$ incorporates solvent–solute (solvation) and solute–solute interactions and ΔA is the decrease in nonpolar surface area (in $Å^2$ per molecule) of the substrate that is exposed to the solvent upon binding to the CD cavity.[319] A phenomenological theory has been proposed for estimation/prediction of the magnitude of the binding constants for formation of 1 : 1 α-CD–solute inclusion complexes in aqueous solution at 25 °C.[289,319]

In addition to the selected data presented in Table 4, binding constants for the inclusion complexation of 56 mono- and 1,4-disubstituted benzenes with α- and β-CD have been reported.[261] Binding constants for the interaction of naphthalene and some of its mono- and dimethyl derivatives with α-, β-, and γ-CD are available.[251] The binding constants of 13 flavonols to β-CD under acidic and basic conditions have been reported.[323] It is evident from the data in Table 4 that the three different CDs can have large differences in the magnitude of the binding constant with respect to the same solute (as a consequence of the steric requirements and size of the solute in relation to the size of the specific CD cavity). It is important to note that the same solute can bind differently (positionally) to the three CDs.[263,324] For instance, methyl salicylate binds to α- and β-CD in different geometries.[263] In the former case it was postulated that the ester and hydroxyl groups were buried in the interior cavity, whereas in the latter case the phenyl end of the methyl salicylate molecule binds in the cavity. The magnitude of solute–CD binding constants typically decreases with an increase in temperature.[289] For instance, the binding constant for interaction of 1-naphthyl acetate with β-CD is 463 M^{-1} at 15 °C and only 134 M^{-1} at 40 °C.[325]

The $CD : solute$ stoichiometry and magnitude of binding interaction can also be impacted by the presence of other organic additives, depending on the additive's hydrophobicity and spatial volume.[173] For example, it was noted that the binding constant for interaction of 1-naphthol with β-CD decreased if methanol or ethanol was added to the aqueous solution.[272] However, Nakajima[326] demonstrated an increased binding constant of pyrene and β-CD in the presence of ethanol. Warner et al. have also reported that the presence of small amounts of other additives can impact the magnitude of CD–aromatic solute binding constants. For example, the addition of 1.0% of alcohol greatly enhances the binding of pyrene to CDs.[173,257] Dramatic enhancements of fluorescence quantum yield with coumarin laser dyes have been observed upon inclusion by β-CD in aqueous solution, but only when small amounts of organic cosolvents were also included.[327] This was interpreted in terms of the fact that the small organic molecules displace water (which has a pronounced fluorescence quenching ability) from the CD cavity and act as space fillers to enhance binding. Likewise, addition of cyclohexanol as cosolvent to aqueous solutions of γ-CD containing 2-naphthyloxyacetic acid (2-NAA) resulted in an increase in the FI of 2-NAA.[328] As 1-butanol is added to aqueous solution containing the γ-CD–pyrene 1:2 complex, it displaces one of the pyrene molecules from the CD cavity to form the 1 : 1 : 1 CD–pyrene–butanol complex.[329] The binding interaction between sodium salicylate and β-CD was also increased owing to the presence of various alcohols.[262] However, in this instance, the addition of the alcohol to the system resulted in a decrease in the FI observed from salicylate.

The nature of the bulk solvent as well as charge on the CD (or solute) also impacts their solute binding ability. For instance, the interaction of ketocyanine dyes with β-CD dissolved in aqueous and dimethylformamide (DMF) as bulk solvent is different. The equilibrium binding constants were found to be greater for ketocyanine binding to β-CD that was dissolved in the nonaqueous media. An investigation of the binding of piroxicam to neutral β-CD and a charged β-CD revealed that the charged form interacted more strongly with this analyte.[330,331] The neutral form of the drug celecoxib was found to bind to β-CD approximately four times better compared to its ionized form.[332]

In the case of CDs, buffer and salt components employed to adjust the pH and/or ionic strength influence the equilibrium constant for the particular CD–solute inclusion complexation process.[13,333] For instance, the binding constant for interaction of 3-hydroxy-2-naphthoic acid with β-CD at pH 5 and 25 °C ranged from 264 M^{-1} in the presence of sodium perchlorate to 1260 M^{-1} in

the presence of sodium chloride (ionic strength 0.1 M) (the binding constant in the absence of any salt was 1083 M^{-1}).[333] Thus, differences in the strength of analyte binding to CDs could also impact and influence the magnitude of fluorescence observed. Hence it is important to compare only similar systems and, in analytical measurements, to ensure that such variables are held constant.

Likewise, the addition of additives can alter the FI of analytes bound to vesicle/liposome assemblies. For example, the FI of 1,6-diphenyl-1,3,5-hexatriene (DPH) in DPPC vesicles is reduced by over 50% due to the presence of added ethanol.[334] The ethanol causes the DPPC aggregates to undergo the L_β to $L_\beta I$ phase transition which alters the location and binding environment of the DPH solute with a resultant decreased fluorescence efficiency. Alternatively, surfactants can be added to produce similar effects. Alterations in temperature can also have dramatic effects on the magnitude of solute partitioning into vesicle/liposome systems.[295,335] Differences in the vesicle size (which varies with the preparation method employed) can affect solute binding. For instance, the magnitude of the binding constants for interaction of *n*-heptanethiol with DODAC vesicles were 500, 1200, and 2800 M^{-1} for vesicles whose hydrodynamic diameters (size) were 2850, 520, and 220 Å, respectively.[336] Hence all of these factors that influence analyte binding to vesicle/liposome aggregate systems can in turn alter the maximum fluorescence emission intensity observed from those solutes as well.[337]

With respect to solute binding to a particular organized assembly, it should be recognized that there is a distribution of locations instead of a unique specific binding location in the assembly (i.e. a heterogeneity of binding sites exists, each site being characterized by a different microenvironment). In addition, there is the possibility of organized assembly heterogeneity (a distribution of micelles or vesicles of different aggregation numbers, etc.). Ordered lipid systems are thought to be heterogeneous in the sense that fluorescent solutes, when embedded in such ordered lipid media, are situated within different sites of different ordering and mobilities.[338] Hence this multitude of possible interactions during the excited-state lifetime of the solute in such an organized system results in a multitude of deactivation channels of the excited state, leading to distributions in such fluorescence properties as lifetimes, quantum yields, etc.[339–343] Huang and Bright[342] noted that for micellar media, there could be several contributions to such observed distributions including (i) water gradients, (ii) micelle polydispersity, and (iii) compositional diversity within the micelles.

All of the binding constant data reported for the CD–solute combinations in Table 4 were for the 1 : 1 complexes (one solute molecule bound per CD molecule). However, in some situations, higher-order complexation can occur, with 1 : 2, 2 : 1 or 2 : 2 CD:solute complex stoichiometry being reported.[265,296,317,321,344,345] Coumarin 153 forms both 1 : 1 and 1 : 2 complexes with methylated β-CD.[346] The size of the particular CD cavity serves to impose possible steric restrictions with respect to the number of analyte molecules which can be bound. For example, the larger-diameter γ-CD can accommodate two pyrene molecules, whereas in the cases of binding to α- or β-CD, only one pyrene is bound.[227] Such results indicate that an important factor in solute binding to CDs in aqueous solutions is the size of the analyte in relation to the cavity dimensions of the particular CD.

In the case of micelles (or vesicle systems), an useful means to assess the degree of solute occupancy is to assume that the analytes distribute among the micelle/vesicle aggregates according to Poisson statistics.[1,5,7,8] The probability of a micelle containing x analyte molecules is given by Equation (6):

$$P(x) = \mu^x \frac{e^{-\mu}}{x!} \tag{6}$$

where μ is the stoichiometric mean number of analyte molecules per micelle (i.e. [analyte]/[micelle]) and $P(x)$ the probability of a micelle with an occupation of x analyte molecules[26,316,347] (for a brief discussion of the distribution of solubilized molecules among micelles, see Miller[348]). The concentration of micelles [micelles] is given by Equation (7):

$$[\text{micelle}] = \frac{C - \text{CMC}}{N} \tag{7}$$

where C is the surfactant concentration and N is the micelle aggregation number. As an example, for a solution of an analyte whose concentration is 1 mM in a 0.10 M SDS micelle medium, the average number of analyte molecules per SDS micelle, μ, would equal 0.6 (assuming an aggregation number of 60 for SDS). For this example, $P(0) = 0.549$ (54.9% of the micelles contain no analyte molecule) and $P(1) = 0.329$ (32.9% of the SDS micelles contain one analyte molecule), and 12.2% of the micelles would contain two or more analyte molecules.[347] By contrast, if the analyte concentration is only 1.0×10^{-5} M, then only 0.3% of the occupied micelles (0.6%) would contain two or more analyte molecules.[89]

As another example, for a micelle concentration of 1.0 mM and an oxygen concentration of ca. 0.10 mM, $\langle O_2 \rangle = 0.1$; thus 90% of the micelles will be free of oxygen.[7] As a consequence, analyte molecules bound to the micelles would be "protected" from oxygen quenching. Consequently, manipulation of the concentration of micelles in relation to that of the analyte

(or analytical reagents/impurities present) is possible so as to prevent multiple analyte/solute/impurity occupation of the micelles. At any given analyte concentration, an increase of concentration of micelles will serve to decrease μ. A relatively high surfactant micelle to analyte concentration ratio will serve to ensure that the probability of any one micelle containing two or more analyte (or analyte–solute or analyte–impurity, etc.) molecules is very low. This in turn eliminates (or greatly reduces) the possibility of formation of the dimeric form of the analyte molecule (which typically are not or weakly luminescent) or diminishes other solute/impurity quenching of that analyte.

In addition, for situations in which both analyte and quencher bind to the micellar aggregate, the quenching efficiency typically decreases with increasing surfactant concentration.[266] This is a consequence of the fact that increasing surfactant concentrations result in formation of a greater number of micelle aggregate species; thus, the probability that any one micelle has both the analyte and quencher bound to it decreases. Thus, impurity quenching (of analytes) can be suppressed in micellar solutions provided that the fluorescent analyte and the impurity molecules are likely to be solubilized in different micelles. In fact, in one report it was noted that the fluorescence quantum yield for 9,10-dimethylanthracene in ethanol was lower compared to that in the literature and that this was likely due to impurities present in the chemical.[349] However, in micellar solutions, the anticipated value was observed despite the presence of that impurity. This clearly demonstrates that impurity quenching can be suppressed in micellar solutions.

2.3.3 *Effect of Organized Assembly Systems upon the Reaction Rate and Path*

Aqueous organized assembly systems often induce changes in a wide variety of ground- and excited-state reactions. Reactions and the position of chemical equilibrium in such systems often proceed in a manner very different from that observed in homogeneous bulk solvents owing to the altered microenvironment and concentration effects in the organized medium.[7,18,161,350,351] It is well known that surfactant micellar systems or CDs can impact (either accelerate or inhibit) many types of chemical reactions.[18,94,160,161,350–353] An excellent recent review article discusses reactions in micellar systems[350] and an exhaustive compilation of altered reactivities in different ordered assembly systems is provided in monographs by Fendler.[18,353] A model (pseudophase ion exchange (PPIE)) has been developed for ionic micellar solutions, which is capable of describing and predicting most ionic micellar effects on reaction rate and equilibria over a wide range of experimental conditions (pH, detergent, added salt, etc.).[354] A review article summarizes and discusses the use of CDs to mediate a wide range of different types of organic reactions.[355]

As examples of ordered media effects, reactions involving the cleavage of phenyl esters, decarboxylation of cyanoacetates, and oxidation of α-hydroxyketones were accelerated by factors of 300, 44.2, and 3.3, respectively, in the presence of CDs compared with that in their absence.[161] Decarboxylation reactions were also enhanced in the presence of micellar systems. For instance, the presence of CTAB, Triton X-100, and SDS enhanced the rate of decarboxylation of 5-amino-1,3,4-thiadiazole-2-carboxylic acid by factors of 5.0, 5.7, and 1.3, respectively.[18] Cationic CTAX micelles caused an increase in the rates of reduction of the carbonyl group of steroids by factors of 2.2–7.8, depending upon the steroid's structure.[94] In addition to alteration of the rates of reactions, the product distribution (including stereochemistry) can be affected by the presence of organized assemblies such as CDs or micelles.[94,161] The photodecarboxylation reaction of ketoprofen occurs with a decreased rate in the presence of β-CD compared to that in aqueous solution. Also, in the β-CD medium, an additional photoreaction was observed.

Review articles discuss the factors that affect the kinetics of photoreactions in organized assemblies and the topics of photochemistry in surfactant solutions and the fate of excited-state probes in micelles.[2–8,167–169,356–360]

The type and concentration of organized media can also have a profound effect upon different photochemical processes.[6,7,94] For instance, the quantum yield of trans $\rightarrow$ cis photoisomerization of (dibutylamino)stilbazolium butanesulfonate was found to be almost 2 orders of magnitude greater in micelle and vesicle ordered media compared to that in bulk water.[32] In contrast, the yield for the trans $\rightarrow$ cis photoisomerization of azobenzene in the presence of CDs was roughly only half of that observed in 80:20 (v/v) water/methanol solution.[356] The rate constants for hydrolysis of cytotoxic pyronins is altered compared to that observed in a bulk alkaline medium.[357] For example, the presence of cationic CTAB micelles resulted in a threefold higher rate constant compared to that in aqueous solution, whereas the hydrolysis reaction was completely inhibited by anionic SDS micelles.

Vesicle/liposome systems can also alter reactivity. The second-order rate constants for the alkaline hydrolysis of *p*-nitrophenyl octanoate in the presence of DODAC or DODAB surfactant vesicle systems were 28–61 times larger than the rate observed in water in the absence of the amphiphile at the same pH.[336] The bimolecular

rate constant for the photooxidation reaction of 1,3-diphenylisobenzofuran with singlet molecular oxygen in the presence of zinc(II) phthalocyanine is roughly 2 orders of magnitude smaller in the presence of DPPC liposomes compared with that in ethanol solution.[361]

The rate and quantum yield of electron exchange between excited pyrene, P and dimethylaniline (DMA) (Equation 8):

$$DMA + P^* \longrightarrow P^- + DMA^+ \qquad (8)$$

to yield the pyrene anion and DMA cation has been examined in different media.[362] A continual decrease in the yield of the pyrene anion (ion yield of 1.0 in the presence of aqueous CTAB micelles) is observed on going from aqueous micelle to aqueous vesicle to swollen micelle to microemulsion systems, with no yield of ions observed in reverse micelles.[362]

The presence of β-CD reduced the initial rate of product formation in the photohydroxylation of 3-nitroanisole compared to that observed in homogeneous aqueous buffer solution.[259] Many organic molecules exhibit a greater degree of photostability in organized assembly media compared to the general behavior observed in organic solvents.[1–8,18,42,214] The observed rate constant for the second-order decay of mesitylthiyl radicals was 3000 times less in a 0.05 M SDS micellar medium than that observed in heptane.[363]

Quenching of excited states of molecules can be dramatically altered in the presence of organized media. Depending on the nature (size, hydrophobicity, or charge type if an ion) of the quencher and luminescent molecules in relation to the type (micelle, CD, vesicle, etc.), nature (neutral or ionic), and concentration of the organized media, either decreases or increases in quenching can be observed. In aerated solutions, the fluorescence lifetime of pyrene in pure water is shorter (126 ns) than in aqueous solutions of CTAC or SDS micelles (157–158 ns).[364] This was initially rationalized in terms of either a (i) lesser solubility of oxygen in the micelle relative to water, (ii) rate constant for oxygen-quenching of solubilized pyrene slower than diffusion controlled in bulk solution, or (iii) quenching of pyrene fluorescence by water.[364] Reasons (i) and (iii) were subsequently ruled out.

The main factor responsible for enhanced lifetimes in organized media thus appears to be decreased quenching constants. For example, the excited state of pyrene is quenched by oxygen molecules in bulk solvents. The bimolecular quenching rate constant is $1.1 \times 10^{10}\,M^{-1}\,s^{-1}$ in aqueous solution.[362] However, it is about an order of magnitude less in the presence of β-CD ($1.2 \times 10^{9}\,M^{-1}\,s^{-1}$). In the presence of a mixed organized system containing either SDS or CTAB and the β-CD, the quenching rate constant is further reduced to $6.0 \times 10^{8}\,M^{-1}\,s^{-1}$.[362] These results were rationalized as follows: The pyrene can bind to the cavity of the β-CD molecule, wherein it is somewhat protected from quenching by other molecules, such as oxygen, following its excitation. The addition of either SDS or CTAB results in the surfactant molecule also binding to the CD cavity, which leads to a more hydrophobic microenvironment for pyrene in addition to a greater binding constant and greater protection from quencher molecules. Similar results concerning resistance to oxygen-quenching of fluorescence for other organic solutes bound to organized assemblies have been reported.[42] Pyrene is also considerably protected from oxygen if present in DODAC vesicles.[146] In a comparative study of several micellar systems, the effectiveness of the protection of pyrene from oxygen-quenching (as given by the lifetimes of excited pyrene) was Brij-35 > SDS > surface-active crown ether decyl-18-crown-6.[365] The conclusions of another study indicated that nonionic surfactant micelles were more effective in protecting excited states of organic molecules from oxygen-quenching compared to ionic micelles.[366]

Likewise, 1 to 2 orders of magnitude reductions in the bimolecular quenching rate constants were observed for pyrene with the quenchers nitromethane, thallium(I) cation, copper(II) cation, and methyl viologen cation in the presence of these organized assemblies.[364] Thus, via the use of CDs, with or without added surfactants, it is possible to shield pyrene to various degrees from quenching interactions due to external species.[207] A nearly fourfold decrease in the quenching efficiency of iodide ion toward 2-naphthol was observed when the naphthol was bound to α-CD.[324] The presence of β-CD, which served to bind the methyl-2-aminobenzoate (MAB) solute, dramatically protected MAB against quenching by iodate ion. The fluorescence quenching of pyrene by *N,N*-DMA was diminished by a factor of ca. 7.5 in the presence of β-CD compared with that observed in the 50 : 50 (v/v) dimethylsulfoxide–water bulk solvent in its absence.[283] Therefore, the use of organized media can minimize or obviate detrimental analyte–solute (or solvent) interactions that otherwise might lead to quenching deactivation of the analyte's excited state[20]

However, if both fluorescent solute and quencher molecules are of appropriate size so that they both bond to one CD cavity, then the quenching interaction can be enhanced. For example, in the presence of β-CD, the fluorescence quenching of naphthalene, 1-methylnaphthalene, and acenaphthene by trimethylamine was greater by factors of 17, 91, and 280, respectively, compared to that observed in water at 25 °C.[367]

Metal cations are attracted as counterions to the surface of anionic micelles or vesicles and act as more

efficient quenchers, whereas anionic species are repelled and ineffective. The opposite would be true for cationic micelles or vesicles for which anions would be more efficient quenchers.[347] Hence, the use of appropriate micellar media can be utilized to achieve some selectivity against charged quenchers. For example, metal (or other) cationic species are repelled from cationic micelles and their quenching of micellar-bound solutes is thus ineffective. Likewise, anionic species would be repelled from anionic micelles and hence bind solutes protected from such quenchers. Consequently, through judicious choice of micelle (or other organized media), charge type (and concentration) in relation to the analyte, and quencher molecules, it is possible to either decrease or increase the quenching efficiency as desired.

The quenching effects of neutral species can also be diminished when luminescent species are bound to micelle systems. For example, the quenching of several aminobenzoic acid derivatives of aminoacylmonomethylamides by acrylamide in pH 7.4 phosphate buffer was decreased roughly by a factor of two in the presence of SDS micelles.[368] The lowering in quenching indicated a diminution in the rate of collisions between the aminoacylmonomethylamide compounds and the quencher acrylamide. Likewise, the quenching effects of iodide ion toward lumichrome was decreased about fourfold in the presence of β-CD.[204]

The utilization of organized media for examination of the mechanism of resonance energy transfer has been the focus of many studies. This is due to the importance of such energy transfer in solar energy conversions, photosynthesis, etc. The addition of different organized assemblies has been shown to enhance the efficiency of such energy transfer steps in a number of systems. For instance, efficient energy transfer from the excited state of acridine orange-10-dodecyl bromide (AOD) to methylene blue (MB) has been observed in the presence of SDS micelles.[369] In contrast, only very little sensitized fluorescence of MB was observed in ethanol as the bulk solvent. Enhancement of the efficiency of energy transfer is possible if the particular organized assembly binds the two species involved so that they are in close proximity. For some systems, it has been shown that vesicle or liposome systems are superior to micelles in this regard. The energy transfer from a donor oxacyanine dye molecule to the acceptor thiacyanine dye is 20–100 times more efficient in a cationic vesicle system than in CTAB micelles.[148]

The rate of energy transfer from excited $Tb(pdc)_3^{3-}$ (pyridine-2,6-dicarboxylate, pdc) to $Ln(pdc)_3^{3-}$ is very slow in bulk aqueous solution; however, association of these complex ions with CTAB results in rapid and efficient energy transfer.[370] The apparent second-order rate constant of energy transfer from 2-acetylnaphthalene to europium(III) ions in SDS micelles is 1 to 2 orders of magnitude larger than the value in homogeneous solution in the absence of SDS, depending upon the specific SDS and Eu(III) concentrations.[371]

In addition, enzyme-generated triplet species can transfer their energy to suitable acceptors and elicit sensitized emission in organized assemblies. For instance, the electronically excited state generated during the peroxidase-catalyzed aerobic oxidation of appropriate substrates efficiently led to fluorescence emission from acceptors like 9,10-dibromoanthracene or chlorophyll solubilized in CTAB, SDS, and Triton X-100 or Brij-35 micellar media.[372] It was noted that ''micelle-solubilized chlorophyll is an excellent detector of enzyme-generated triplet carbonyl species and the use of micelles make it possible to extend (this approach to study) other photobiologically important yet water-insoluble acceptors''.[372]

Hence it is possible to exploit such unique kinetic effects due to the presence of appropriate organized assembly systems in order to manipulate the rates of reaction for the purpose of enhancing the fluorescence performance; that is, the use of ordered media can catalyze otherwise slow analytical fluorescent derivatization reactions. Alternatively, one can employ such organized systems to inhibit or slow down undesired side or decomposition reactions in order to stabilize the desired fluorescent species.

For example, in a fluorescence assay for cyanide ion based on its reaction with 1,4-naphthaquinone-2-sulfonic acid to form a fluorescent product, the reaction time in the presence of CTAB micelles is only ca. 5 min compared to a reaction time of roughly 90 min in its absence.[373] The rate of phenolic oxidative coupling of morphine and related opiates to yield fluorescent products, utilized for postcolumn high-performance liquid chromatography (HPLC) detection, was increased in the presence of the nonionic surfactant Triton X-100.[374] In addition, the micellar medium protected the fluorescent product from further oxidation. The presence of the micellar system thus improved the stability of the system and detection limits possible relative to that observed in bulk solvent with no surfactant present. There are also other examples of micellar catalysis to speed up fluorescent reactions. In a method for determination of laccase activity, the apparent rate of oxidation of *o*-phenylenediame catalyzed by laccase was enhanced by Brij-35 micelles.[375] The detection limit was $0.07\ U\ L^{-1}$ laccase. The fluorescence reaction between *n*-(4′-hydroxyphenyl)-*N*-(4-methylquinolinyl)amine, hydrogen peroxide, and cobalt(II) was remarkably enhanced in the presence of cationic surfactant media.[376] The fluorescence output was about five times greater in the presence of trimethylstearylammonium chloride (STAC) than in its absence. This reaction was employed to determine

hydrogen peroxide in the 0–250 ng mL^{-1} concentration range.[376] Brij-35 was found to be effective in increasing the rate (roughly sixfold) of the hemin-catalyzed fluorogenic reaction between hydrogen peroxide and *N,N′*-dicyanomethyl-*o*-phenylenediamine.[377] The subsequent method for peroxide was more sensitive (by a factor of ca. 5) in the surfactant medium compared to that in aqueous solution in absence of the Brij-35.

Thiols react with 4-nitro-*N*-*n*-butyl-1,8-naphthalimide (4-NBN) to form highly fluorescent products that can be employed for the fluorimetric determination of thiols.[378] The rate of this reaction was greatly enhanced (by a factor of ca. 10^5) by the presence of cationic CTAC micelles. This increased rate was attributed to the micellar concentration of the reactants and the micellar effect on the extent of the acid–base ionization that generates the reactive thiolate anion which is the nucleophile.[378]

A very popular fluorimetric method for the determination of primary amino compounds is based on their reaction with *o*-phthalaldehyde in the presence of thiols (such as 2-mercaptoethanol) under basic conditions to form the fluorescent 1-(2′-hydroxyethylthio)-2-alkylisoindole product.[379] The isoindole product formed from many analytes is unstable owing to subsequent decomposition and oxidative reactions. The presence of different micellar and CD organized media can slow such undesired reactions involving the fluorophore.[379,380] For an overview of the effect of different CD media upon Roth's method and the fluorescence output from over 45 different aromatic organic compounds, the reader is referred to Frankewich et al.[380]

2.3.4 *Organized Assembly Effects upon Chemical Equilibrium*

Surfactant micelles and vesicles as well as CD media can shift the position of ground-state chemical equilibria. Perhaps the most important and frequent analytical application of such equilibrium shift media has been to alter the position of acid–base equilibrium.[11,12,94,101,152,381–408] Acid–base ionization constants (and other equilibrium constants) can be altered, some by several orders of magnitude, in the presence of organized assemblies compared to those observed in bulk aqueous solution (selected data presented in Table 5). A quantitative method for the evaluation of the influence of surfactant micellar media upon the protolytic equilibrium of various organic reagents has been published.[399] Micelle-induced pK_a shifts are due to "medium" (altered microenvironment) and electrostatic effects.[152] Interfacial water is known to be less polar than bulk water and is akin to an electrolyte salt solution (the ionic strength in the Stern layer of ionic surfactant micelles can be in the range 1.0–3.0 M). The degree of dissociation of the charged

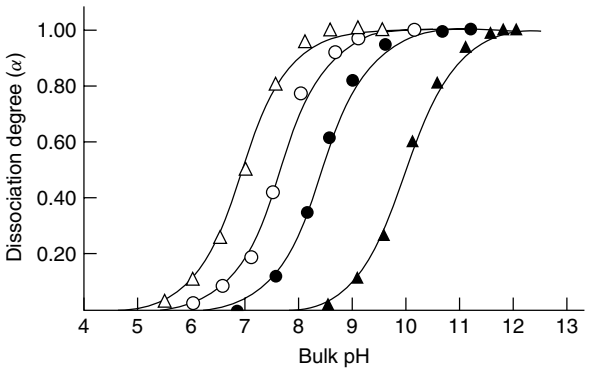

Figure 3 Plots of the degree of dissociation of tetracaine as a function of pH in aqueous solution (●) and in the presence of the surfactant micelles CTAB (△), Triton X-100 (○), and SDS (▲). (Reproduced with permission from Garcia-Soto and Fernandez.[388])

aliphatic amino group of tetracaine as a function of pH in the absence and presence of surfactant-forming micelles is shown in Figure 3. As can be observed, compared to the curve in bulk aqueous solution, the presence of surfactant micelles can shift the ionization curve to either higher or lower pH values.[388] For example, while the ground-state pK_a (and excited-state pK_a) of 4-methyl-7-hydroxyflavylium cation in water is 4.4 (ca. −1.0), values of 2.25 (0.05), 3.3 (−1.2), and 6.5 (0.3), respectively, were reported in the presence of CTAB, TX-100, and SDS micellar media.[409] Hence organized media can function as pH shift reagents.

Likewise, a shift of the pK value of one unit was reported for tetracaine in the presence of egg yolk PC liposomes and as large as 1–3.5 units for 5-doxylstearic acid and 16-doxylstearic acid in PC liposomes, respectively.[400–402] Cationic surfactant media typically enhance the acid dissociation (decrease pK_a's) of organic compounds, whereas anionic surfactant media have the opposite effect (see Figure 3).

The pK_a of acids in the presence of CDs is typically larger than (or equal to) that observed in the absence of the CD, whereas the pK_a of phenolic compounds is generally smaller than (or equal to) that in the absence of the CD.[160,161] The pK_a's of seven flavonols were found to increase slightly with increases in the β-CD concentration.[323] Typically, the magnitude in the shifts of the pK values in CDs is not as great as that observed in surfactant vesicle or micellar systems.

The kinetics of the excited-state protolytic dissociation of 1- and 2-naphthols in aqueous solution, micelles, and liposomes has been examined.[347,384,406,407] Many analyte molecules, such as 2-naphthol, when excited in neutral aqueous solution, undergo spontaneous deprotonation as a result of the fact that the pK_a of the excited state can be 6–7 units lower than that of the ground state. For

Table 5 Effect of organized assemblies on some selected equilibrium and excited-state quenching processes

System/process	Equilibrium constants		References
	In bulk solvent	In organized medium	
Ionization of the monocation of 5-aminoindazole	$pK_{a(water)} = 5.27$	$pK_{a(0.002\,M\ SDS)} = 4.4$	249
		$pK_{a(0.06\,M\ SDS)} = 6.2$ $pK_{a(0.10\,M\ SDS)} = 6.6$ $pK_{a(0.06\,M\ CTAB)} = 5.05$	
	$pK_{a(water)^*}$[a] $= 2.5$	$pK^*_{a(0.006\,M\ SDS)} = 3.0$ $pK^*_{a(0.10\,M\ SDS)} = 5.4$ $pK^*_{a(0.06\,M\ CTAB)} = 2.2$	249
Ionization of dipyridamole	$pK_{a(water)} = 5.8$	$pK_{a(SDS\ micelles)} = 7.3$ $pK_{a(CTAC\ micelles)} = 4.3$	267
Keto–enol equilibrium of benzoylacetone	$K_{E(water)} = 0.62$	$K_{E(SDS\ micelles)} = 150$	252
Ionization of 1-naphthol	$pK_{a(water)} = 9.23–9.45$	$K_{E(\beta-CD)} = 5$ $pK_{a(\beta-CD)} = 10.20$ $pK_{a(\alpha-CD)} = 9.45$ $pK_{a(CTAB\ micelles)} = 8.53$ $pK_{a(SDS\ micelles)} = 10.04$ $pK_{a(C16E16\ micelles)} = 10.30$	384,385 160 385
Ionizations of PAR[b]	$pK_{a_{1(water)}} = 3.02$ $pK_{a_{2(water)}} = 5.56$	$pK_{a_1(CPB\ micelles)}$[c] $= 2.58$ $pK_{a(CPB\ micelles)} = 4.48$	386
Ionization of α-tocopherol		$pK_{a(CTAC\ micelles)} = 11.7$ $pK_{a(DDAB\ vesicles)}$[d] $= 11.0$ $pK_{a(DMPC\ vesicles)}$[e] > 14.5 $pK_{a(DPB\ micelles)}$[f] $= 13.1$	387
Ionization of tetracaine	$pK_{2a(water)} = 8.26$ $pK_{2a(45:55\ dioxane-water)} = 7.25$ $pK_{2a(82:12\ dioxane-water)} = 7.09$	$pK_{2a(TritonX-100\ micelles)} = 7.58$ $pK_{2a(SDS\ micelles)} = 9.92$ $pK_{2a(CTAB\ micelles)} = 6.88$	388
Quenching of DIP by acrylamide[g]	$K_{sv(water,\ pH\ 5)} = 15.0\,M^{-1}$ $K_{sv(water,\ pH\ 7)} = 30.0\,M^{-1}$	$K_{sv(SDS\ micelles)} = 0.2\,M^{-1}$ $K_{sv(CTAC\ micelles)} = 3.6\,M^{-1}$	383
Ionizations of PAN	$pK_{a(water)} = 2.8$ and 11.0	$pK_{a(CPC\ micelles)}$[h] ≈ 1.3 and 8.4	389
Quenching of 2-naphthol by iodide ion[i]	$K_{sv(water)} = 20\,M^{-1}$	$K_{sv(\alpha-CD)} = 5.8\,M^{-1}$ $K_{sv(\beta-CD)} = 10.0\,M^{-1}$ $K_{sv(\gamma-CD)} = 19.0\,M^{-1}$	227
Ionizations of APDMCI[j]	$pK_{1(water)} = 4.0$ $pK_{2(water)} = 13.7$	$pK_{1(2\,mM\ \beta-CD)} = 3.6$ $pK_{2(2\,mM\ \beta-CD)} = 11.4$	65
Ionization of RB[k]	$pK_{a(water)} = 8.63$	$pK_{a(CTACmicelles)} = 6.93$	66
Ionization of pyranine	$pK_{a(water)} = 7.22$	$pK_{a(DPPC\ vesicles,\ outer\ surface)} = 6.00$ $pK_{a(DPPC\ vesicles,\ outer/inner\ surface)} = 6.90$	132
Ionization of PNA[l]	$pK_{a(water)} = 5.0$	$pK_{a(DMPC\ vesicles,\ interfacial\ region,\ 37\,^{\circ}C)} = 3.7$ $pK_{a(DMPC\ vesicles,\ membrane\ bound,\ 37\,^{\circ}C)} = 5.7$	273
Quenching of oxacyanine dye by a thiacyanine dye[i]	–	$K_{sv(4\,mM\ CTAB\ micelles)} = 7.5 \times 10^4\,M^{-1}$ $K_{sv(1\,mM\ 2C_{12}-L-Glu-ph-C_4N\ \textit{vesicles})} = 9.5 \times 10^6\,M^{-1}$	148
Ionization of long-chain (C_{15}) hydroxycoumarin	$pK_{a(water)} = 7.75$	$pK_{a(DHP\ vesicles)} = 10.5$ $pK_{a(DODAB\ vesicles)} = 5.70$ $pK_{a(DPPC\ vesicles)} = 10.5$	58
Ionization of excited state 1-aminonaphthalene	$pK_{a(water)^*} = 13.5$	$pK_{a(CTAB\ micelles)^*} = 11.9$	74
Ionizations of 3,3-dimethyl-2-phenyl-3*H*-indole	$pK_{a(water)} = 3.25$ $pK_{a(water)^*} = 9.34$	$pK_{a(SDS\ micelles)} = 4.75$ $pK_{a(SDS\ micelles)^*} = 11.79$	306

(*continued overleaf*)

Table 5 (*continued*)

System/process	Equilibrium constants		References
	In bulk solvent	In organized medium	
Formation constant for charge-transfer complex formed between anthracene and *p*-methylbenzenediazonium-tetrafluoroborate	$K_{C(acetonitrile)} = 0.6\,M^{-1}$	$K_{C(SDS\ micelles)} = 210\,M^{-1}$	391
Formation constant for charge-transfer complex formed between pyrene and methylviologen	$K_{C(methanol)} = 3.4\,M^{-1}$	$K_{C(SDS\ micelles)} = 704\,M^{-1}$	392
Quenching of pyrene by cesium chloride[l]	$K_{SV(water)} = 11\,M^{-1}$	$K_{SV(SDS\ micelles)} = 305\,M^{-1}$ $K_{SV(Brij-35\ micelles)} = 1.3\,M^{-1}$	327
Ionization of 2,6-di-*tert*-butyl-4-nitrophenol	$pK_{a(water)} = 6.92$	$pK_{a(CTAB\ micelles)} = 5.13$	18,393
Ionization of coumarin dye	$pK_{a(water)} = 7.75$	$pK_{a(CTABmicelles)} = 6.35$ $pK_{a(TritonX-100\ micelles)} = 8.85$ $pK_{a(SDS\ micelles)} = 11.15$	394
Excited-state ionization of piroxicam	$pK^*_{a(water)} = 2.70$	$pK_{a(\beta-CD)^*} = 5.58$	395
Ionizations of morin	$pK_{a_1(aqueous10\%ethanol)} = 4.8$	$pK_{a_1(CTAB\ micelles)} = 3.5$	396
Ionization of quercetin	$pK_{a_1(water)} = 7.18$	$pK_{a_1(Ethonium\ micelles)}$[m] $= 6.03$	397
Ionization of lumogallion	$pK_{a_2(water)} = 5.6$	$pK_{a_2(CPC\ micelles)} = 4.77$	398
Ionization of bromocresol green	$pK_{a(water)} = 4.87$	$pK_{a(0.1\ M\ Septonex\ micelles)}$[n] $= 3.36$	399
Ionization of pyronin Y	$pK_{a(water)} = 11.4$	$pK_{a(CTAB\ micelles)} = 9.3$ $pK_{a(SDS\ micelles)} = 12.2$	357
Ionization of β-naphthol	$pK_{a(water)} = 9.5$	$pK_{a(\beta-CD)} = 9.9$	400
Ionization of benzoic acid	$pK_{a(water)} = 4.11$	$pK_{a(\alpha-CD)} = 5.20$	161
Ionization of *p*-nitrophenol	$pK_{a(water)} = 7.09$	$pK_{a(\alpha-CD)} = 6.15$	161
		$pK_{a(0.058\ g\ mL^{-1}\ SDS\ micelles)} = 7.92$	94
		$pK_{a(CTAB\ micelles)} = 6.35$	310,393
Ionization of dibucaine	$pK_{a(water)} = 9.00$	$pK_{a(CTAC\ micelles)} = 7.29$ $pK_{a(SDS\ micelles)} = 10.52$	310
Ionization of 2-naphthol	$pK_{a(water)} = 9.5$	$pK_{(\beta-CD)} = 9.67$	398

[a] The asterisk denotes the singlet excited-state ionization constant.
[b] Refers to 4-(2-pyridylazoresorcinol).
[c] Refers to cetylpyridinium bromide (CPB).
[d] DDAB refers to didodecyldimethylammonium bromide.
[e] DMPC refers to dimyristoylphosphatidylcholine.
[f] DPB refers to dodecyldimethyl propiobetaine.
[g] Refers to the Stern–Volmer constant, K_{sv}, for the quenching of dipyridamole (DIP) by acrylamide in buffer at the indicated pH.(383)
[h] Refers to 1-(2-pyridylazo)-2-naphthol (PAN).
[i] Refers to the Stern–Volmer constant, K_{sv}, for the quenching of the indicated fluorescent molecule by quencher molecule or ion in aqueous solution or organized assembly system.
[j] Refers to 2-(*p*-aminophenyl)-3,3-dimethyl-5-cyano-3*H*-indole.
[k] Refers to 2,6-diphenyl-4-(2,4,6-triphenyl-1-pyridinio)-1-phenolate.
[l] Refers to 9-(3-pyrenyl)nonanoic acid.
[m] CMC of ethonium surfactant micelles is 2.15 mM.(397)
[n] Refers to cationic surfactant 1-ethoxycarbonylpentadecyltrimethylammonium bromide.

example, 2-naphthol has a pK_a (in water) of 9.4–9.5, but in the first excited singlet state, the pK^*_a drops to 2.8; it is a much stronger acid in its excited state.(403,404) The presence of organized media can impact both those values. In the case of benzimidazole, the $pK_{a(water)}$ is ca. 13.1, whereas the $pK^*_{a(water)}$ is 12.1.(98) However, in the presence of 0.10 M CTAB, these values are 11.6 and 11.5, respectively.(98)

As a consequence of such solute pK shifts, the use of an appropriate organized assembly system alters the pH requirements for the fluorescence assay, functioning as pH shift reagents.(404) For instance, in the

determination of ethylenediaminetetraacetate (EDTA), the optimum pH for formation of the fluorescent EDTA–Zr(IV)–Alizarin red X ternary complex is shifted toward a lower range (pH 4.5–5.5) in the presence of CTAB micelles compared to that seen (pH 5.3–6.3) in the absence of micelles.[405] The CTAB micelle medium allows the formation of the complex at more acidic conditions, which leads to an increase in the stability constant of the complex.

Likewise, the optimum pH required for the horseradish peroxidase (HRP)-catalyzed fluorogenic reaction between hydrogen peroxide and *p*-hydroxyphenylpropionic acid (HPPA) was shifted from 8.0 in aqueous solution to ca. 7.5 in aqueous solution containing CTAB micelles.[410] The lower pH requirement in the presence of CTAB leads to less deactivation of the enzyme, HRP, compared with that observed in bulk aqueous solution.

The effect of micellar and CD organized media upon the position of the keto–enol equilibria of different 1,3-dicarbonyl compounds, as well as on their pK values and reactivity, was the topic of a comprehensive study by Iglesias.[411]

Thus, in some situations, it is possible to utilize organized assembly systems as ''pH shift'' (or other ''equilibrium shift'') reagents to facilitate desired fluorogenic reactions in chemical analysis.

2.3.5 *Bulk Solvent and Microscopic Properties of Organized Assemblies*

Micellar, vesicular, and CD organized assemblies are fluid, transparent to ultraviolet light, stable, and macroscopically homogeneous. They are particularly attractive for use in analytical spectroscopy in general because they are commercially available, relatively nontoxic, and environmentally benign, and their solutions easy to prepare. Also, because of these characteristics they have been employed in numerous photochemical and photophysical investigations.[1–8,18,360,364]

In particular, the fluorescence properties of aromatic probe molecules have proved invaluable as a tool to provide information on micellar systems, such as determination of micellar parameters (CMC and N) and estimates of microscopic polarity and viscosity, water content, partitioning between water and micelle phase, permeability for quenchers, etc.[1–8,410–415] A mini review nicely summarizes the influence of the solvent medium on the energy of electronic states of organic molecules.[416] As a consequence of such medium solvatochromatic effects, fluorescence measurements often exhibit differences between fluorescence parameters (such as emission wavelength maximum, λ_{em}; excited-state lifetime, τ; quantum yield; φ; FI; etc.) obtained in an aqueous solution relative to that obtained in aqueous organized assembly solutions.

A number of different techniques have been employed to determine the apparent effective microscopic environment experienced by bound solutes in organized systems. Table 1 compares some of the properties of bulk solvents to the microscopic environmental properties experienced by solutes when bound to different organized assembly media. The term *micropolarity* typically refers to the effective polarity experienced by the solute when bound to a particular organized assembly and is a reflection of the porosity or extent of water penetration into the hydrocarbon portion of that organized medium.[417] A compilation of the effective dielectric constants in organized surfactant media is available.[30] Such results indicate that the apparent dielectric constant of the interfacial region of micelles, bilayers, and microemulsions is substantially lower than that of bulk water (Table 1). In general, the polarity of micellar interfaces decreases with increasing surfactant chain length and surfactant concentration. It decreases with temperature. It is interesting to note that the static dielectric constant of aqueous salt solutions (in the absence of any organized assembly) is considerably lower than with water alone. For example, all salts show a lowering of the dielectric constant; at 21 °C, the dielectric constant of aqueous sodium chloride solutions ranges from 73.9 (0.5 N NaCl) to 48.4 (5.0 N NaCl solution).[37] For cationic and anionic surfactant systems, the concentration of ions (counterions) in the micelle Stern layer region has been estimated to be about 3.0 M.[5,418]

For PC vesicles and multibilayer systems, a lower polarity than that of aqueous normal micelles is typically observed.[90] Microemulsion systems exhibit a polarity similar to that of aqueous micelle media. The effective polarity of the CD cavity has been reported to be equivalent to the polarity of an 80:20 (v/v) methanol–water or ethanol mixture.[65,419] The range of values for the effective polarity of CDs has been thought to be due to the fact that (i) some probe molecules might not be entirely included within the cavity of the CD, (ii) the number of included water molecules present in the complex might be dependent upon the nature of the solute, and (iii) the probes most frequently employed respond primarily to the bulk polarity of the region and their sensitivity to hydrogen-bonding interactions might be very weak.[65] More important, the polarity or microenvironment sensed by a specific probe depends on the position that the probe occupies within the CD cavity.[36]

The term *microviscosity* or *microfluidity* is typically a measure of the extent of hindrance experienced by a solubilized molecule for translational motion within a particular micelle, vesicle, or CD organized system.[417] Micellar microviscosity values are sensitive to the reporter molecule's position in the micelle aggregate. For instance, using the probe dipyrenylpropane as a reporter molecule, the microviscosity of SDS micelles

was estimated to be ca. 19 cP at 20 °C.[89] However, when diphenylpropane is used as the reporter probe molecule, a considerably smaller value for the microfluidity was obtained, namely $\eta \approx 4$ cP. Use of *N*, *N*, *N*′, *N*′-tetramethyldiaminodiphenylketoimine hydrochloride as the probe molecule yielded an estimate of 5 cP for the viscosity in SDS.[420] A value of 4.3 cP has been determined from examination of the excited-state torsional relaxation of the probe 1,1′-dihexyl-3,3,3′, 3′-tetramethylindocarbocyanine iodide.[421] Such differences in estimates of microenvironmental parameters can be due to different solubilization sites of the probe molecules, and/or reflect nonisotropic fluidity (or environmental property) in some organized assembly systems. Such influence of the size of the probe molecule upon the observed microfluidity of lipid bilayer/vesicle and bile salt systems has also been reported.[422,423] An excellent discussion of the fluidity parameters of the lipid regions of micelles and liposomes as determined by fluorescence polarization measurements is available.[422]

In most cases, the variation in the ε, η, and $E_{T(30)}$ values for a particular organized assembly system can be rationalized in terms of the different time-averaged locations for the various probe molecules employed to probe that specific system and in the appropriate choice of the probe and empirical scales employed.[71,80,240,242] It has also been pointed out that in many cases the reported values do not reflect the microenvironment of the micelle-bound probe since they were determined at surfactant concentrations <50 mM;[66] hence there are different degrees of interaction of the probes with their environments.[71] Thus, comparisons between systems should be made only if the same probe molecule was employed using the same technique for determination of the microenvironment parameter under consideration. Also, mixed micelles or organized assembly systems have different binding sites and microenvironments compared to that of their respective individual components.[423] The presence of additives can dramatically impact some of these microscopic parameters as well. For instance, the apparent microviscosity of CTAB was estimated at ~40 cP, whereas the values were 200–300 cP or ~400 cP for CTAB with added cetyl alcohol or 1-hexadecanesulfonate, respectively.[83]

The reduced fluidity of analytes in surfactant micelles has recently been utilized for analytical advantage. Strong rotational depolarization of fluorescence is typically observed for small molecules in bulk homogeneous solutions. However, when bound to organized assemblies, such small molecules experience a highly viscous, less fluid microenvironment which results in their fluorescence being partially polarized. Yan and Myrick have taken advantage of this and shown that it is possible to differentiate between different xanthene dyes on the basis of their micelle-enhanced steady-state fluorescence polarization (SSFP) even though the dyes' regular fluorescent emission spectra were almost identical.[424]

Demas et al.[425,426] have described a novel fluorescence technique which allows determination of the solvent environment around solutes, specifically, the degree of solvent (water) accessibility of a micelle-bound solute. Novaki and El Seoud[66] have reported that the effective water concentrations at the solubilization sites of different solvatochromic probes in cationic micellar media range between 22.1 and 47.8 M, depending upon the specific probe, surfactant micelle, and reference solvent mixture employed. Another report estimated water concentrations of 3.0, 54.0, and 14 M at the solubilization site of 9-anthroate in micelles of Triton X-100, SDS, and DTAC, respectively.[427]

It has been estimated that for ionic surfactant micelle systems the concentration of counterions in the micellar Stern layer is of the order of 3 M.[5,418] The refractive index for aqueous micellar solutions has been reported to be in the range 1.3–1.5, depending on the specific surfactant, its concentration, and the presence or absence of other additives.[428–431] Of course, the effects of the refractive index of the solvent system can influence some of the solution-phase photophysical properties of molecules (such as their fluorescence decay times).[430,432]

3 ORGANIZED ASSEMBLIES IN ANALYTICAL FLUORESCENCE MEASUREMENTS

A number of specific reviews on the applications of surfactants,[14,433–438] micelles,[9,14,20,433,434,438–444] microemulsions,[440,445,446] reversed micelles,[446] CDs[13,173,435–439,447–449], and related media in analytical spectroscopy have been published, each of which typically has a section dealing with fluorescence measurements in such media. Excellent overviews on the utilization of organized assemblies to facilitate luminescence analysis of environmental species have recently been published.[14,173] A review concerning the use of organized assemblies to enhance analytical chemiluminescence measurements is also available[166] (*see* **Chemiluminescence, Electrogenerated**).

3.1 Enhancement of Fluorescence Intensity in Organized Media

3.1.1 Effect of Organized Media Upon a Solute's Fluorescence Properties

Upon transfer from a bulk solvent, such as water, to an organized assembly phase, the luminescence properties (such as excitation and emission wavelengths, λ_{ex} and

λ_{em}, lifetime, τ, and emission intensity) of many solutes can be altered, sometimes dramatically. [14,18,20,409,438,439] For instance, the inclusion of organic compounds in CDs frequently leads to fluorescence enhancements and also a shift of the emission maximum toward shorter wavelengths.[438,439,450,451] Figure 4 illustrates the effects of increasing concentrations of different surfactants upon the maximum fluorescence emission wavelength of 4-cyano-4′-propoxybiphenyl.[180] With increases in surfactant concentration, the respective micelles of these surfactants form and biphenyl binds to them, resulting in an alteration (shift to shorter wavelength) of the emission wavelength due to a change in the effective microenvironment experienced by biphenyl. The polarity of the respective micelle–biphenyl binding sites is similar to that observed in bulk solvents that are less polar than water, such as DMF or ethanol.[180] The different concentration ranges over which the emission wavelength changes merely reflect the fact that these four micelle-forming surfactants have different CMC values (ranging from 6.1×10^{-5} M for NPh-EO9 to 1.2×10^{-4} M for NPh-EO30 and 8.0×10^{-3} M for SDS).[180]

Numerous other such examples are available.[452–458] Upon addition of surfactant micelles, the fluorescence wavelength observed for the Al–lumogallion complex shifts to 552 nm, compared to 584 nm in bulk aqueous solution.[453] Likewise, a 15-nm blue shift of the emission wavelength from 350 to 335 nm is observed when PS vesicles are added to aqueous buffered solutions

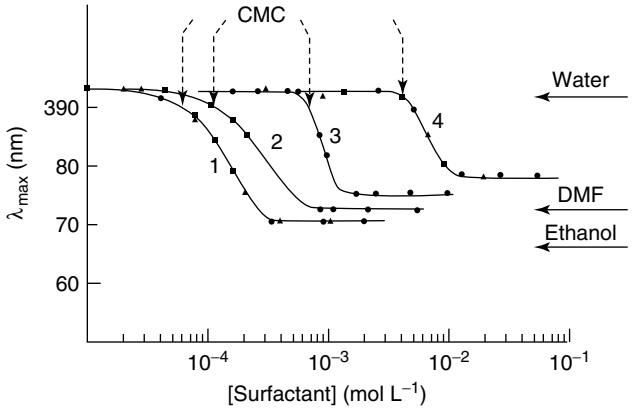

Figure 4 Plots of the maximum fluorescence emission wavelength, λ_{em}, of 4-cyano-4′-propoxybiphenyl as a function of increasing concentrations of the surfactants: (1) nonylphenol polyoxyethylene glycol (9–10) [NPh-EO9]; (2) nonylphenol polyoxyethylene glycol (30) [NPh-EO30]; (3) SDS; and (4) CTAB. The arrows to the right indicate the λ_{ex} maximum obtained in the bulk solvents water, dimethylformamide (DMF), and ethanol in the absence of any surfactant. The descending dotted lines indicate the CMC values for these four surfactants. (Reproduced with permission of Wiley-VCH from C. David, E. Szalai, D. Baeyens-Volant, *Ber. Bunsen-Ges. Phys. Chem.*, **86**, 710–716 (1982).)

containing cardiotoxin.[282] A blue shift of the 8-anilino-1-naphthalenesulfonate (ANS) emission maximum from 515 nm in aqueous solution to 480 nm in DMPC vesicles was reported.[133] The fluorescence emission wavelength of 4-aminophthalimide is shifted from 540 nm in water to 513 nm in 2 mM β-CD.[293] In the presence of SDS micelles, the emission maxima for 2-anilinonaphthalene-6-sulfonic acid and *N*-phenylnaphthylamine were found to be blue-shifted by 28 and 41 nm, respectively, compared to that in bulk aqueous solution.[342,343] The maximum fluorescence emission wavelength observed for the niobium(V)–lumogallion-tartaric acid complex shifted from 630 nm in aqueous solution to 610 nm in the presence of Triton X-100 micellar media.[454] For some analytes, red shifts are observed. For instance, the emission wavelength of [2-chloro-3-(2,2-dicyanoethenyl)-2-cyclopenten-1-ylidene]methylpyridinium salt shifted from 615 nm in pure water to 626 nm in an aqueous solution of Triton X-100.[46] Similar shifts were observed for other carbocyanine-type probe molecules.

More important than the shifts in emission wavelengths are the intensity enhancements observed for many fluorescent species. Many fundamental photophysical studies present data that clearly demonstrate that the fluorescence quantum yields and FI of organic molecules and inorganic metal complexes can be enhanced in the presence of ordered media in comparison to that observed in bulk solvent systems.[14,20,148,173,282,307,438,448,455–589] Some representative data is presented in Table 6. It should be cautioned that not all the results noted for a particular compound were obtained under the same standard or optimized conditions, and hence direct quantitative comparisons are not possible in such cases. However, the bottom line is that addition of surfactant and bile salt micelles, vesicle/liposomes, and/or CDs can elicit enhanced fluorescence from a range of potential analytes that in many cases can be translated into improved sensitivity in fluorescence assays for those species.

In addition, the fluorescence intensities of a number of other organic molecules were reported to be intensified, but no quantitative data were provided as to the magnitude of such enhancements. Most such reports were in the context of physical, photochemical, or binding studies rather than analytical applications. For example, the luminescence intensity of the following organic compounds was reportedly enhanced by the indicated organized media: doxorubicin (in the presence of β- or γ-CD)[285]; ANS and 2-*p*-toluidinylnaphthalene-6-sulfonate (by SDS and Triton X-100 micelles or β-CD)[38,73]; perylene (by CTAB)[60]; 2-[*p*-(dimethylamino)phenyl]-3,3-dimethyl-5-cyano-3*H*-indole (by β-CD)[65]; pyrene

Table 6 Summary of fluorescence enhancements of drugs, vitamins, dyes, and other organic substances in different organized assemblies

Compound(s)	Organized system	Enhancement factor[a]	Comment	References
Aromatic carboxylic acids	CTAB micelles	2–3[c]	[b]	458
Salicylic acid	CTAB micelles, pH > 4.0	~2.0[c]	–	459
	β-CD, pH 9.0	1.35	–	380
Sodium salicylate	β-CD	~1.5[c]	[b]	262
Methyl salicylate	α-CD	~3.2[c]	–	263
	β-CD	~1.6–2.0[c]	–	263,380
	γ-CD	~3.0[c]	–	263
Bilirubin	Dodecyldimethylamine oxide micelles	13[c]	–	460
Hematoporphyrin	Lecithin vesicles	6[c]	[d]	461
α-Tocopherol	Liposomes	[c,e]	[f]	462
	Brij-35 micelles	7–10[c]	–	463,464
	CTAB micelles	3.3[c]	–	586
Dansylated amino acids	Various micellar systems	4–19	–	463
Apomorphine	CTAC micelles	7[c]	–	465
N-*n*-Propylnorapomorphine	CTAC micelles	6[c]	–	465
Ribonuclease	NaLS micelles	3.0[c]	–	466
	β-CD	3.3[c]	–	380
Biphenyl	α- or β-CD	1.4[h]	–	467
Substituted biphenyls:				
4-Hydroxy-	CTAC or TDSS micelles	1.2–8.9[c]	[g]	465
2-Hydroxy-	CTAC or TDSS micelles	1.3–1.9[c]	[g]	
3,4-Dihydroxy-	CTAC or TDSS micelles	1.9–10.1[c]	[g]	
4-Chloro-	α- or β-CD	2.3[h]	–	467
3-Chloro-	β-CD	1.25[h]	–	
2,3,5,6-Tetrachloro-	α- or β-CD	1.03–1.27[h]	–	
3,3′, 4, 4′-Tetrachloro-	CTAB micelles	2.7–6.5[c]	–	107,108
	SDS micelles	2.2–3.7[c]	–	
	$C_{12}E_{10}$ micelles	11.1[c]	–	
3,3′, 4, 4′,5-Pentachloro-	SDS micelles	1.3[c]	–	107
	$C_{12}E_{10}$ micelles	2.1[c]	–	
9,10-Dimethylanthracene	NaLS micelles	1.55[i]	–	468
9,10-Di-*n*-propylanthracene	NaLS micelles	1.41[c]	–	
α-Naphthyl acetate	β-CD	1.2[c]	[j]	469
	γ-CD	2.1[c]	[j]	
Insulin	CDEA[k] micelles	1.7[c]	–	466
N,N-Dimethyltryptamine	α- and β-CD	1.2–1.4[c]	–	470
Mescaline hemisulfate	α- and β-CD	2.2–2.5[c]	–	
Ibogaine	α- and β-CD	1.2–1.8[c]	–	
Anthraceneundecanoic acid	Brij-35 micelles	10.30[c]	–	471
	SDS micelles	2.19	–	
	CTAC micelles	6.39[c]	–	
	SB-12 micelles	8.37[c]	–	
2-Aminochrysene	CTAB micelles	3.7[c]	–	74
3-Aminofluoranthene	CTAB micelles	3.0[c]	–	
1-Aminonaphthalene	CTAB micelles	1.5[c]	–	
Perylene	Lecithin or lecithin–cholesterol liposomes	1.3–1.5[c]	–	472
	CTAB or CTAC micelles	1	–	473
	SDS or Triton X-100 micelles	1	–	474
Pyrene	SDS micelles	1.1–2[c]	–	475–477
	CTAC micelles	10.2[m]	–	476
	DTAC micelles	34.5[m]	–	
	β- or γ-CD	1.35[c]	–	
Pyrene	Mixed lecithin liposome/Brij-35 system	2.8[c]	–	478
	Mixed lecithin liposome/CTAB system	1.5[c]	–	
	Mixed lecithin liposome/NaLS system	1.7[c]	–	
	Brij-35 micelles	13.8[x]	–	222
	Triton X-100 micelles	3.3[c]	–	477

(*continued overleaf*)

Table 6 (*continued*)

Compound(s)	Organized system	Enhancement factor[a]	Comment	References
Chrysene	SDS micelles	3.6[c]	–	474
	CTAB micelles	12.1[c]	–	
	Triton X-100 micelles	11.8[c]	–	
	Brij-35 micelles	2.6[x]	–	222
Benzo[*a*]pyrene	PS[n] vesicles	3.0[c]	[o]	479
	Brij-35 micelles	3.5[x]	–	222
	CTAC micelles	4.5[x]	–	480
	NaLS micelles	3.9[x]	–	
	SB-12 micelles	4.2[x]	–	
	$C_{12}E_{10}$ micelles	8.0[c]	–	481
Benzo[*e*]pyrene	Brij-35 micelles	3.5[x]	–	222
	$C_{12}E_{10}$ micelles	3.0[c]	–	256
	SDS micelles	2.5[c]	–	257
Dibenz[*a*, *c*]anthracene	SDS micelles	3.3[c]	–	248
	CTAB micelles	14.2[c]	–	
	Brij-96 micelles	15.0[c]	–	
Dibenz[*a*, *h*]anthracene	$C_{12}E_{10}$ micelles	30.0[c]	–	256
Coronene	SDS micelles	1.6[c]	–	248
	CTAB micelles	29.2[c]	–	
	Triton X-100 micelles	25.7[c]	–	
	$C_{12}E_{10}$ micelles	6.0[c]	–	256
1,2-Benzanthracene	SDS micelles	2.0[c]	–	257
Benzo[*ghi*]perylene	Brij-35 micelles	6.7[x]	–	222
	$C_{12}E_{10}$ micelles	20.0[c]	–	481
Indeno[1,2,3-*cd*]pyrene	$C_{12}E_{10}$ micelles	4.0[c]	–	481
Benzene derivatives	β-CD	L	–	483
N-Phenyl-1-naphthylamine	Different aqueous micelles	~10.0[c]	–	484
	β-CD	6.3[c]	–	380
	NaCDC micelles, pH 8.93	~10.0[c]	–	485
	100 mM SDS micelles	~13.7	–	342
	CTAC micelles	20.3[c]	–	480
	SB-12 micelles	22.4[c]	–	
	Brij-35 micelles	32.4[c]	–	
Cardiotoxin II (snake venom toxin)	PS[n] vesicles	2.0[c]	[p]	282
Auramine[q]	SDS micelles[l]	5.2[c]	–	420
	SDS micelles + 5 M urea	4.6[c]	–	
	SDS micelles + 0.75 M NaCl	6.5[c]	–	
Auromine-O	SDS micelles	2.0[c]	[f]	308
	α-CD	1.4[c]	[f]	
	γ-CD	7.5[c]	[f]	
	Brij-78 micelles	5.0[c]	[f]	
1,3,5-Hexatriene	DMPC vesicles[r]	Several hundred-fold[c]	–	486
	Triton X-100 micelles	estimated (Est.) 50–100[c]	–	486
Abz–Xaa–$NHCH_3$[s]	SDS micelles	1.03–2.45[c]	[t]	368
DMA3HF[u]	SDS micelles	~6.0[c]	–	29
N-Methyl-2,3-naphthalimide	β-CD	1.3[c]	–	487
3-Aminophthalimide	SDS micelles	1.4[c]	–	195
3-Monomethylamino-*N*-methylphthalimide	SDS micelles	2.7[c]	–	195
4-Aminophthalimide (4-AP)	β-CD	11.9[c]	–	380
	β-CD	4.7[c]	[cc]	293
	SDS micelles	4.0[c]	–	195
4-Dimethylamino-*N*-methylphthalimide	SDS micelles	10.0[c]	–	195
11-(4AP)-undecanoic acid	SDS micelles	2.9[c]	–	294
	CTAB micelles	2.8[c]	–	
	Triton X-100 micelles	3.0[c]	–	
N-Alkyl(11)-4-aminophthalimide	SDS micelles	7.0[c]	[z]	488

(*continued overleaf*)

Table 6 (*continued*)

Compound(s)	Organized system	Enhancement factor[a]	Comment	References
Fenproporex	SDS micelles	2.6[c]	[b]	96
	β-CD	1.9[c]	–	380
PAS-S[v]	SDS micelles, 1 M HCl	5.5[c]	–	245
	CTAB micelles, 1 M HCl	16.0[c]	–	
	CTAB micelles, 10 mM NaOH	6.5[c]	–	
	Triton X-100 micelles, 1 M HCl	14.5[c]	–	
	Triton X-100 micelles, 10 mM NaOH	3.0[c]	–	
1-Naphthol	Triton X-100 micelles, 10 mM NaOH	2.80[c]	–	245
	SDS micelles, pH 12	1.8[c]	–	480
	Brij-35 micelles, pH 12	1.2[c]	–	480
	CTAC micelles, pH 12	8.4[c]	–	
	β-CD	21.1–27.0[c]	–	271,272,280
	γ-CD	1.9[c]	–	271
	α-CD	1.5[c]	–	
	DPPC vesicles, pH 6.0	1.2[c]	–	489
	dioctyldecyldimethylammonium bromide (DOAB) or DDAB vesicles, pH 6.0	2.8[c]	–	
2-Naphthol	β-CD, pH 6.2	~1.2–1.5[c]	–	227,271
	γ-CD	1.1[c]	–	271
	α-CD	1.1[c]	–	
	SDS micelles, pH 6.0	1.6[c]	–	490
Fluorescein-labeled phenobarbital	CTAC micelles, pH 6.0	2.64[c]	–	491
	γ-CD	1.57[c]	–	
	β-CD	2.05[c]	–	
	N,N-dimethyldodecylamine-N-oxide (LDAO) micelles[w], pH 7.1	1.05[c]	–	
Fluorescein	1% SDS micelles	5.8[c]	[f/ww]	492,493
	2% sodium dodecylbenzene sulfonate micelles	9.6[c]	[f/ww]	
	2% CTAB micelles	~11[c]	[f/ww]	
o-Hydroxybenzaldazine (or 2,3-diazabutadiene)	CTAB micelles	2.85[x]	[y]	494
Acenaphthene	Brij-35 micelles	2.0[x]	–	222
Naphthalene	Brij-35 micelles	2.0[x]	–	222
	SDS micelles	1.38[c]	–	490
	Triton X-100 micelles	2.6[c]	–	
Anthracene	CTAC micelles	1.85[c]; 0.93[x]	–	349,471
	CTAB micelles	0.28[x]	–	349
	Brij-35 micelles	2.10[c]	–	471
	SDS micelles	1.91[c]	–	
	SB-12 micelles	2.21[c]	–	
9-Methylanthracene	CTAC micelles	2.00[x]	–	349
Fluorene	CTAC micelles	1.06[x]	–	349
	Brij-35 micelles	1.00[x]; 1.07[c]	–	222,480
	SDS micelles	1.2[c]	–	480
	SB-12 micelles	1.17[c]	–	
Fluoranthene	Brij-35 micelles	1.8[x]	–	222
	SDS micelles	1.5[c]	–	482
Phenanthrene	Brij-35 micelles	2.4[x]	–	222
Dipyridamole	CTAC micelles, pH 8.0	~1.3[c]	–	383
	SDS micelles, pH 8.0	~0.7[c]	–	
Quinine	β-CD	~3.0[c]	–	286
	SDS micelles	1.3[c]	–	495
Retinol	Triton X-100 micelles	4.2[c]	[z]	496
	Ovolecithin vesicles	2.5[c]	[z]	
	Dipalmitoyl lecithin vesicles	3.0[c]	[z]	
	β-CD	1.1[c]	–	380
	NaC micelles, pH 7.6	~14[c]	–	497

(*continued overleaf*)

Table 6 (*continued*)

Compound(s)	Organized system	Enhancement factor[a]	Comment	References
	NaDC, pH 7.6	10.0[c]	–	
	CTAB miclles	7.0[c]	–	586
Procaine	β-CD	7.9[c]	–	380
	α-CD	~3.3[c]	[b]	309
	β-CD	~10.0–12[c]	[b]	309,498
	SDS micelles	~1.7[c]	–	498
Cyanine dyes	SDS micelles	15[c]	[aa]	39
3,3′-Diethylthiacarbocyanine iodide	SDS micelles	4.5[c]	[f]	499
Cardiotoxin[bb]	PS[n] vesicles	3.2[c]	–	282
N-Methylaniline	β-CD	6.0[c]	[cc]	268
Aniline	β-CD	4.0[c]	[cc]	
Phenol	β-CD	1.1–1.7[c]	[cc]	268,269
3,5-Dimethylphenol	β-CD	2.2[c]	–	269
3,4,5-Trimethylphenol	β-CD	3.6[c]	–	
2-Isopropoxyphenol	CTAB micelles	1.8[c]	[b]	500
2-Isopropoxyphenolate	CTAB micelles	19.5[c]	[b]	
APDMCI[dd]	β-CD, pH 9.5	6.0[c]	–	380
Erythrosin B	γ-CD	2.7[c]	–	290
	α-CD	1.8[c]	–	
	β-CD	2.1[c]	–	
	TX-100 micelles	4.2[c]	[f]	501
	Brij-35 micelles	3.6[c]	[f]	
	Tween 20, 40, or 80 micelles	4.9; 4.6; 4.8[c]	[f]	
	CTAB micelles	4.9[c]	[f]	
	CPB micelles	4.7[c]	[f]	
	SDS micelles	1.0[c]	–	
Rose Bengal	γ-CD	4.0[c]	–	290
Eosin Y	0.10 M CTAB	2.2[c]	–	502
Acridine orange	Sodium dodecyl sulfonate micelles	2.3[c]	–	503
Acridine yellow	SDS micelles	2.3[c]	–	504
	Sodium dodecyl sulfonate micelles	2.3[c]	–	503
Phenosafranin	SDS micelles	5.7[c]	–	504
	Sodium dodecyl sulfonate	17.5[c]	–	
Safranin T	SDS micelles	5.3[c]	–	504
	Sodium dodecyl sulfonate micelles	15.9[c]	–	
Trypaflavin	SDS micelles	1.5[c]	–	504
TNS[ee]	14 mM β-CD	~52[c]	[ff]	505
	14 mM SDS micelles	~38[c]	[ff]	
	α-CD, pH 5.3	44[c]	–	47
	β-CD, pH 5.3	27[c]	–	
	γ-CD, pH 5.3	4.4[c]	–	
Chalcones[gg]	10 mM SDS micelles	~29–150	[hh]	506
Carbocyanine dye[ii]	Triton X-100 micelles	~14.0	–	46
	CTAB micelles	~7.5	–	
Merocyanine 540	DMPC[r] liposomes	3.4–4.0[x]	–	135,507
	250 μM PC[jj] vesicles	~1.9[c]	–	508
	DPPC[ll] liposomes	4.3[x]	[o]	507
	DTAB micelles	2.3[x]	–	
ANS	DMPC[r] vesicles	100[c]	–	133
	DPPC[ll] vesicles	>50[c]	–	281
	NaCD micelles, pH 8.2	~100[c]	–	283
	NaC micelles, pH 8.2	~70[c]	–	
	SDS micelles	5.0[c]	–	495
	CTAC micelles	45–66[c]	–	
	α-CD	1.3[c]	–	271
	β-CD	3.1–11.4[c]	–	265,271
	γ-CD	4.3[c]	–	271
2-Anilinonaphthalene-6-sulfonic acid	SDS micelles	~12.3	–	342

(*continued overleaf*)

Table 6 (*continued*)

Compound(s)	Organized system	Enhancement factor[a]	Comment	References
Bombesin	Lysolecithin micelles	1.4–1.7[c]	–	510
	DMPS[kk] vesicles	~1.2[c]	–	
	β-CD (+urea)	1.15–1.35[c]	–	380
Triazinylaniline dyes	DMPC[r]/DPPC[ll]	>30[c]	–	511
Nystatin	DMPC[r] vesicles	1.6–26[c,mm]	–	511
	β-CD and its derivatives	2.4–7.1[c]	–	380
Thiacyanine dye	CTAC micelles	~14[c]	–	148
	DHDMAB[nn] vesicles	~25[c]		
	$2C_{12}$-L-Glu-ph-C_4NBr vesicles	~260[c]	–	148
EDKC[oo]	PC[pp] liposomes	28.0[c]	–	513
p-Aminobenzoic acid	β-CD	3.0[c]	[b]	514
Methyl *p*-dimethyl-aminobenzoic acid	β-CD	~5.0[c]	–	288
o-Aminobenzoate anesthetics	PC[ji] liposomes	3–4[c]	–	515
Rhodamine B	SDS micelles	1.6[c]	[f]	516
	Sulfapol[qq] micelles	1.8[c]	[f]	
MB	10 mM SDS micelles	2.0[c]	[f]	517
	β-CD, pH 7.2	1.5[c]	–	518
Indole	CTAB micelles	93–133[c,g]	–	307
	β-CD	1.26–1.4[c]	–	304,380
	β-CD plus 36% $CaCl_2$	2.27[c]	–	304
DM3HI[rr]	SDS micelles, pH 9.5	2.0[c]	–	306
	SDS micelles, pH 1.0	4.3[c]	–	
	CTAB micelles, pH 9.5	3.0[c]	–	
Indolecarboxylic acids	CTAB micelles	30–878[c]	–	307
HSMQB[ss]	SDS micelles	3.2	–	563
	CTAB micelles	1.9	–	
	Brij-35 micelles	2.0	–	
DBASBS[tt]	SDS micelles	~20	–	32
	DTAB[uu] micelles	~20	–	
	Egg lecithin vesicles	~35	–	
Tetrakis(sulfonatophenyl) porphyrin	Triton X-100 micelles	1.5[c]	[aa]	520
Chlorophyll *a*	Brij-35 micelles	~330[vv]	–	518
Chlorophyll *b*	Brij-35 micelles	~610[vv]	–	
APDMCI[xx]	SDS micelles, pH 9.5	6.5[c]	–	521
	CTAB micelles, pH 9.5	6.5[c]	–	
1,5-Diphenyl-3-vinyl-Δ^2-pyrazoline	20 mM SDS micelles	2.5[yy]	–	522
Benzodiazepines (7)	Anionic surfactant micelles	1.2–6.5[c]	–	523,524
Prazepam	β-CD	1.5[c]	–	380
Naproxen	8 mM β-CD	1.4[c]	–	300
Carbendazim	SDS micelles, 0.1 M HCl	2.6[c]	–	99
Carbofuran	Anionic and nonionic surfactant micelles	4.1–22.9[c]	–	525
TTCDF	SDS micelles	8.0[c]	–	526
	CTAB micelles	12.0[c]	–	
	Brij-96 micelles	16.0[c]	–	
	$C_{12}E_{10}$ micelles	39.7[c]	–	
	α-CD	1.25[c]	–	
Curcumin	Triton X-100 micelles	~7.0[c]	–	311
THPC[zz]	α-CD	17.0[c]	–	527
	β-CD	110[c]	–	
	γ-CD	200[c]	–	
Benzimidazole (monoanion)	CTAB micelles, pH 14	~11[c]	–	98
Amiloride	SDS micelles, pH 3.0	7.4[c]	–	528
	SDS micelles, pH 3.0, 8% 1-propanol	4.3[c]	–	
Mazindol	SDS micelles, 0.004% H_2SO_4	4.2[c]	[aa]	529
Bisacodile[aaa]	Brij-98 micelles	~15[c]	–	530

(*continued overleaf*)

Table 6 (*continued*)

Compound(s)	Organized system	Enhancement factor[a]	Comment	References
Furosemide	β-CD, pH 7.0	1.25[c]	–	380
	SDS micelles, pH 3.0	2.7[c]	–	528
	SDS micelles, pH 3.0, 8% 1-propanol	2.0[c]	–	
	0.05 M SDS micelles	1.7[c]	–	531
Methyl paraben	α-CD, pH 4.5	~2.6	–	532
	β-CD, pH 4.5	~12	–	
Estradiol	β-CD, 35 °C	2.0[bbb]	–	208
Ethinylestradiol	β-CD, 35 °C	1.7[bbb]	–	208
	γ-CD, 35 °C	~1.3[bbb]	–	
Dimethylaminobenzonitrile[ccc]	γ-CD	4.0[c]	–	28,533
	β-CD	3.5–10[c]	–	534
	α-CD	23.2[c]	–	28
Diethylaminobenzonitrile[ddd]	α-CD	4.9[c]	–	28
	β-CD	4.2[c]	–	
4-(Dimethylamino)chalcone	CTAB micelles	64.9[c]	[eee]	535
	CPB micelles	28.5[c]	[eee]	
	SDS micelles	15.1[c]	[eee]	
	NaL micelles	23.8[c]	[eee]	
	Tween-20 micelles	66.6[c]	[eee]	
Dansylcadaverine	NaDC micelles, pH 8.0	~50[c]	–	509
	NaC, pH 8.2	~50[c]	–	
	β-CD	8.2[c]	–	380
Propranolol	0.05 M SDS micelles	1.6[c]	–	531
Bumetamide	0.05 M SDS micelles	1.6[c]	–	531
Carvedilol	Mixed DMPC–MDPG vesicle system, pH 7.4	2.19[c]	–	299
13-substituted flavonols	β-CD, 0.05 M NaOH	1.55–4.59[c]	–	323
	β-CD, 0.01 M HCl	0.69–2.19[c]	–	
5-Methoxypsoralen	β-CD	6.0[c]	–	536
14-dansylamino acids	0.014 M β-CD	2.1–24.2[c]	–	380
	0.10 M β-CD in 4.0 M urea	3.3–34.7[c]	–	
2-Methylhydro-1,4-naphthoquinone	β-CD, pH 3.55	2.6[c]	–	537
p-Sulfonated calix[4,6]arene	CTAB	9.0[c]	–	543

[a] Fluorescence enhancement factor refers to the ratio of the FI of the indicated analyte molecule in the presence of the organized medium relative to that observed in a homogeneous bulk solvent under otherwise identical experimental conditions.
[b] Fluorescence enhancement attributed to protection of the analyte from collisional deactivation due to the presence of the organized assembly.
[c] Value relative to bulk water; for ionizable analytes, the magnitude of the enhancement value often depends on the concentration of the particular organized assembly system and upon the solution pH.
[d] Enhancement attributed to increased microviscosity in the lipid bilayer, which diminished competing, nonradiative processes.
[e] α-Tocopherol fluorescence enhanced in liposomes.
[f] Enhancement attributed to a shift in the position of the solute monomer–dimer–multimer equilibrium; from the multimeric or dimeric form (which is usually weakly or nonfluorescing) in the bulk solvent to its monomeric form (which is fluorescent) in the presence of the indicated organized medium.
[g] Enhancement factors are pH dependent.
[h] Value relative to that in dioxane as bulk solvent.
[i] Value relative to that in toluene as bulk solvent.
[j] Values are dependent upon the CD concentration.
[k] Refers to cetyldimethylethylammonium bromide.
[l] Increases in the fluorescence lifetime and emission intensity of the solute observed in the organized system.
[m] Values relative to that in hexane as the bulk solvent.
[n] Refers to PS.
[o] FI is increased in the vesicle system because of the more structured and ordered microenvironment.
[p] Enhancement attributed to micropolarity effect, i.e. the tryptophan moiety in the toxin, which fluoresces, experiences a less polar environment in the vesicle system, and is also less accessible to quenching from water molecules.
[q] Auramine refers to *N*, *N*, *N′*, *N′*-tetramethyldiaminodiphenylketoimine hydrochloride.
[r] Refers to DMPC vesicles.
[s] Refers to aminoacyl monomethylamide derivatives of *o*-aminobenzoic acid, where X_{aa} = Arg, Phe, Leu, and Glu.[368]
[t] Enhancement attributed to the less polar and more rigid micellar microenvironment which restricts the movement of the analyte molecule.
[u] Refers to 4′-*N,N*-dimethylamino-3-hydroxyflavone.

Table 6 (*footnote continued*)

[v] Refers to 2-(4-alkylamino-2-hydroxyphenyl)benzothiazole.
[w] Refers to *N,N*-dimethyldodecylamine-*N*-oxide.
[x] Relative to ethanol or methanol as the reference bulk solvent.
[y] Enhancement due to binding to micelle which restricts the rotational freedom of the azine molecule.
[z] Enhancement due to deaggregation of indicated solute molecules in presence of micelle or lipid systems.
[aa] Enhancement attributed to inhibition of the radiationless decay due to the greater rigidity (greater microviscosity) experienced by the organic compound when bound to the micellar aggregates.
[bb] Obtained from *Naja mossambica*.
[cc] Enhancements attributed to formation of inclusion complexes which results in a decrease in the solutes' rotational freedom and elimination of water molecules in the vicinity of the fluorescent molecules.[268,293]
[dd] Refers to 2-(*p*-aminophenyl)-3,3-dimethyl-5-cyano-3*H*-indole.
[ee] Refers to 2-(*p*-toluidinyl)naphthalene-6-sulfonate.
[ff] Enhancement due to the less polar microenvironment experienced upon binding to the indicated ordered medium.
[gg] Refers to 4-[2-(2-pyridyl)ethenyl]- and 4-[2-(4-pyridyl)ethenyl]-chalcones.
[hh] Enhancement attributed to increased medium viscosity accompanying solute binding to the SDS micelles.
[ii] Refers to 2-chloro-3-(((2,2-dicyanoethenyl)-2-cyclopenten-1-ylidene)methyl)pyridinium chloride.
[jj] Refers to egg PC vesicles.
[kk] Refers to DMPS.
[ll] Refers to DPPC.
[mm] Value depends on the temperature.
[nn] Refers to dihexadecyldimethylammonium bromide.
[oo] Refers to *N,N′*-bis(2-ethyl-1,3-dioxolane)kryptocyanine.
[pp] Refers to saturated PC.
[qq] Refers to the surfactant $C_{12}H_{25}C_6H_4SO_3Na$; CMC = 2.0 mM.[516]
[rr] Refers to 3,3-dimethyl-2-phenyl-3*H*-indole.
[ss] Refers to 2-(*p*-hydroxy)styryl-3-methylquinolizinium bromide.
[tt] Refers to (dibutylamino)stilbazolium butylsulfonate.
[uu] Refers to DTAB.
[vv] Relative to a 1 : 99 (v/v) acetone–water system.
[ww] Enhancement attributed to micellar effects upon the nonquinoid lactone–quinoid tautomeric equilibrium of the dye molecule, so that a greater portion of the molecules are in the nonaggregated, quinoidal form.
[xx] Refers to 2-[(*p*-amino)phenyl]-3,3-dimethyl-5-cyano-3*H*-indole.
[yy] Relative to methanol as the bulk solvent.
[zz] Refers to 5,10,15,20-tetra(*m*-hydroxyphenyl)chlorin.
[aaa] Refers to 4,4′-(2-pyridylmethylene)diphenol diacetate.
[bbb] Relative to 20:80 (v/v) methanol–water solution as bulk solvent system.
[ccc] Refers to 4-(*N,N*-dimethylamino)benzonitrile.
[ddd] Refers to 4-(*N,N*-diethylamino)benzonitrile.
[eee] Enhancement in micellar media reported to be due to the restriction exerted by the micelles on the formation of the nonemissive biradical state P^*.[535]

NaLS, sodium lauryl sulfate; CDEA, cetyldimethylethyltrimethylammonium bromide; TNS, 2-*p*-toluidinonaphthalene-6-sulfonate; TTCDF, 2,3,7,8-tetrachlorodibenzofuran.

(by γ-CD)[283]; 1-anilinonaphthalene-8-sulfonate (by γ-CD)[545]; carbazole alkaloid derivatives (by βCD and derivatized β-CDs)[205]; naphthalene (by β-CD and β-CD containing added sodium tetradecyl sulfate, TDSS)[207]; 1-dimethylaminonaphthalene 5-sulfonamide (by CTAB micelles and egg lecithin liposomes)[75]; diphenylpolyenes (DMPC, DPPC, and DODAC vesicles)[104]; anthracene derivatives (by CTAB, CTAC, SDS, Brij-35, and SB-12 micelles[471]; acridine orange (by SDS and other micellar systems)[70,546,547]; fluorescamine-derivatized peptides (by β-CD)[548]; 9,10-dimethylanthracene and *N*-methyldiphenylamine (by CTAC and CTAB micelles)[349]; coumarin 540A (by α- and β-CD)[549]; 7-(diethylamino) coumarin laser dyes (by γ-CD)[550]; *p*-dimethylaminochalcone (by β-CD)[103]; rhodamine 6F and oxazine 725 dyes (by SDS)[551]; some isomeric estrogens (by β-CD)[552]; eosin (by various nonionic, anionic, and cationic surfactant micelles)[553]; the tripeptide lysyltryptophyllysine (by DMPS)[134]; benzo[*a*]pyrene (by PC vesicles)[126]; hepatitis A virus peptide (by DPPC liposomes)[554]; DPH (by 1-*O*-octadecyl-2-*O*-methyl-DL-glycero-3-phosphorylcholine vesicles and Triton X-100 micelles)[555]; Oxonol VI (by dioleoylphosphatidylcholine vesicles)[307]; 9-aminoacridine and Atabrine (by lipid bilayers)[308]; rhodamine 6 Zh, rhodamine B, rhodamine 6G, Nile blue, and oxazine 1 (by various surfactant micelle media and PC liposomes)[102,136,516,556–558]; rhodamine B-labeled thyroxine (by SDS micelles)[558]; Nile red, also referred to as *Nile blue A oxazone* (by micelles and lipids/liposomes)[202,559,560]; oxonol V (by dioleoylphosphatidylcholine vesicles)[561]; 2-aminonaphthalene (by CTAB micelles)[74]; dihematoporphyrin ether or ester (by CTAB micelles)[562]; styryl derivatives of quinolizinium salts (by CTAB, SDS, and Brij-35 micelles)[519];

14 quinolizinium salts (by SDS and Brij-35 micelles)[563]; 11 *trans*-stilbene derivatives (in SDS and CTAC micelles and DPPC and dicetyl phosphate vesicles)[242]; pinacyanol (by SDS micelles)[102]; acrylodan-labeled bovine serum albumin (by SDS, CTAB, and several other cationic micellar media)[564]; amphiphilic ruthenium complex (by $C_{12}E_6$ micelles)[565]; cresyl violet (by SDS micelles)[566]; Triton X-405 (by Brij-35 micelles)[567]; pyronin Y (by SDS and CTAB micelles)[357]; trimethinethiacyanine dye (by cationic vesicles)[148]; phenolphthalein, diphenylamine, phenylfluorone, *o*-dianisidine, 2,2′-dipyridylketone azine, phenyl-2-pyridylketone hydrazone, 1,3,5-triphenyl-Δ^2-pyrazoline, and 2,3-diphenyl-hexahydro-$\Delta^{1,7}$-isoindazole (by various nonionic surfactant micelles)[530]; naphthylethanols (by β-CD)[568]; *N,N*-diethylaminoethyl (2,*N*-methylamino)benzoate (by egg PC vesicles)[515]; fluorenone (by α-CD)[317]; 1,4-diphenyl-1,3-butadiene (by γ-CD)[569]; diflunisal (by β-CD and γ-CD)[568]; 2,6-dimethylphenol and 2,4,6-trimethylphenol (by β-CD)[269]; triazinylaniline derivatives (by DMPC and DPPC vesicles)[511]; aflatoxins (by α-, β-, and γ-CD)[570]; cholestatrienol (by a mixed 1-palmitoyl-2-oleoylphosphatidyl-choline–cholesterol vesicle system)[571]; various coumarin derivatives (by β-CD)[572,573]; benzo[*a*]pyrene (by liposomes)[103]; triazinylaniline derivatives (by DMPC and DPPC vesicles)[511]; 3-hydroxy-2-naphthoic acid (by β-CD)[574]; cinalukast and montelukast (by α-CD and modified β-CD)[14]; ochratoxin A (by β-CD)[322]; 4-dimethylamino-chalcone (by TX-100 micelles)[575]; and fenoprofen (by β-CD).[576] Thus, organized media can produce an increase in FI of a wide variety of organic compounds compared with that obtained in an aqueous (or organic) medium under otherwise identical instrumental conditions. In most instances, such increased fluorescence facilitates analytical determination of those solutes due to enhanced sensitivity, as will be illustrated in a later section.[99,433,447,500,536]

For the following analytes, the presence of organized assemblies has no appreciable effect (or only a very modest enhancement) upon the FI: coumarin-6-sulfonyl chloride (in SDS or CTAB micelles)[577]; coumarin 480 (by CTAB, SDS, and Triton X-100 micelles)[578]; indole (by α-CD)[579]; uranium–*o*-hydroxyhydroquinonephthalein complex (by SDS and Tween 20 micelles)[580]; 1-amino-4-hydroxyanthraquinone, 1,5-dihydroxyanthraquinone, oxine, and 2,2′-dipyridylketone hydrazone (by various nonionic surfactant micelles)[530]; bimanes (by β-CD)[206]; fluorenone (by β-CD)[207]; testosterone enanthate, caffeine, ephedrine, chlorthalidone, dihydrochlorothiazide, and spironolactone (by SDS micelles)[531]; 8-methoxypsoralen (by β-CD)[536]; and *p*-cresol (by β-CD).[269]

The differences in the fluorescence enhancement factor observed (Table 6) for the same solute in different organized assemblies are a consequence of differences in binding interaction between that solute and the specific organized medium and the specific microenvironmental properties experienced by the bound solute. As a result, the fluorescence enhancements depend upon the specific experimental conditions (i.e. type and concentration of the organized assembly, pH, presence of other additives and cosolvents, etc.). Figure 5(a) illustrates the variation of the FI of chlorophyll *a* and *b* as a function of the concentration of Triton X-100.[518] The intensity increases with Triton X-100 concentration, reflecting the effect

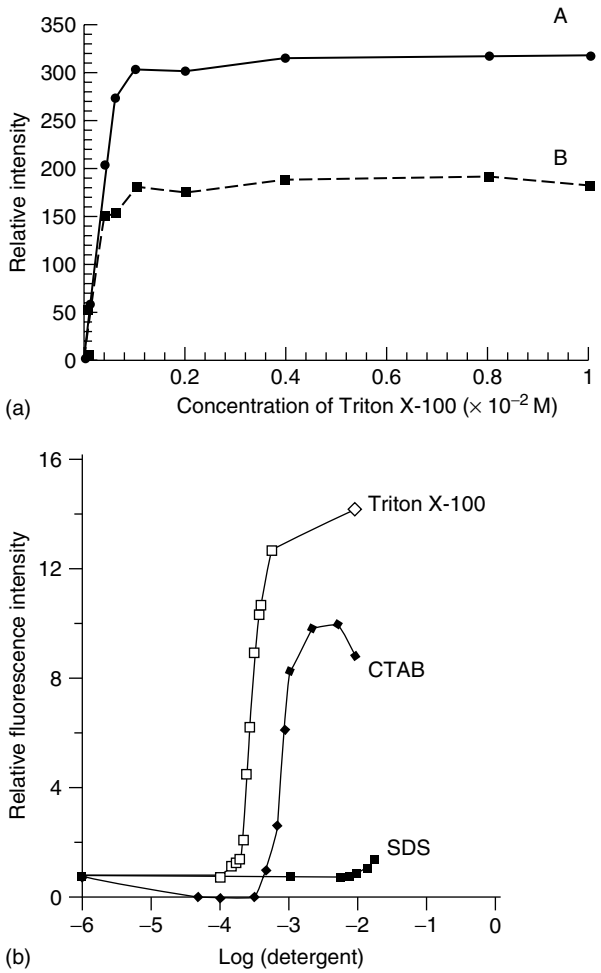

Figure 5 (a) Dependence of the FI of chlorophyll *a* (A) and chlorophyll *b* (B) as a function of increasing concentration of the nonionic surfactant Triton X-100. (Reproduced with permission from Santana et al.[518]) (b) Plots of the relative FI of [2-chloro-3-(2,2-dicyanoethenyl)-2-cyclopenten-1-ylidene]–methyl pyridinium salt versus the logarithm of the surfactant SDS, CTAB, or Triton X-100 concentration. The fluorescence measurements were made at the maximum emission wavelength for each surfactant system, i.e. 626 nm in Triton X-100, 624 nm in CTAB, and 620 nm in SDS. (Reproduced with permission of Wiley-VCH from M.A. Kessler, O.S. Wolfbeis, *Ber. Bunsen-Ges. Phys. Chem.*, **93**, 927–931 (1989).[46])

of micelle formation and binding of chlorophyll to the micellar aggregates, after which the intensity levels off. Figure 5(b) illustrates the effect of the concentration of three different charge-type surfactants, nonionic Triton X-100, cationic CTAB, and anionic SDS, upon the FI of ((2-chloro-3-(2,2-dicyanoethenyl)-2-cyclopenten-1-ylidene)methyl)pyridinium salt.[46] In the case of Triton X-100 and SDS, the FI remains unaffected as the surfactant concentration is increased until the CMC is reached. At and above the CMC, an equilibrium exists between analyte molecules bound to the micellar aggregate and analyte molecules in the bulk solution. As the surfactant concentration is increased further, the number (and sometimes size) of the micelles increases such that the ratio of micellar bound to free analyte molecules becomes progressively larger until saturation is reached.[46] The greater FI in the presence of the neutral surfactant Triton X-100 compared to anionic SDS reflects the fact that the analyte binds to the former much better (probably very small binding to anionic SDS by the anionic carbocyanine solute). In the case of CTAB, the fluorescence decreased with CTAB even below CMC levels (probably owing to quenching by bromide counterions).[46] Once CTAB micelles form, the analyte is incorporated into (and protected by) the micellar aggregate with a resultant increase in its fluorescence emission. The subsequent decrease in fluorescence signal at much greater CTAB concentrations is probably due to changes in the micelle aggregate structure (change in shape from spherical to rodlike), quenching from excess CTAB, and/or quenching from impurities present in the surfactant preparation. Analyte FI versus surfactant concentration profiles similar to those shown in Figure 5(a) and (b) have been reported for many other analyte–organized media combinations.[96,99,103,108,110,187,195,271,287,525,581–586]

The fluorescence emission spectra of methyl *p*-dimethylaminobenzoate as a function of the concentration of β-CD is shown in Figure 6. The emission intensity is observed to increase with increase in the CD concentration because a greater proportion of the analyte molecules present bind to the protective CD cavity upon increasing the CD concentration.[288] In many instances, the limited water solubility of β-CD restricts one's ability to attain the maximum fluorescence signal from a particular analyte.[380]

As is evident from the data in Table 6, the magnitude of the fluorescence enhancement depends on the organized assembly system employed. For instance, the enhancement factor for the analyte auromine-O ranges from 1.4 in the presence of α-CD to 5.0 in Brij-78 micelles and 7.5 in γ-CD.[308] Nitrobenzoxadiazolyl (NBD)-labeled alkylamines exhibited greater fluorescent quantum yields in the presence of nonionic TX-100 micelles compared

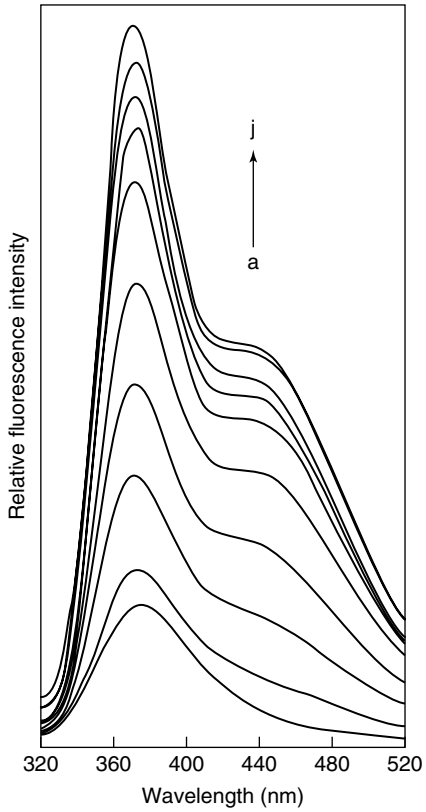

Figure 6 Dependence of the fluorescence emission spectra of methyl *p*-dimethylaminobenzoate as a function of the β-CD concentration. The β-CD concentrations were (a) 0.00, (b) 0.2, (c) 0.4, (d) 0.8, (e) 1.5, (f) 2.5, (g) 3.5, (h) 4.5, (i) 6.0, and (j) 8.0 mM. (Reproduced with permission from Jiang.[288])

to that observed with α-CD or β-CD.[587] Coumarin-6-sulfonyl chloride–derivatized amino acids in a β-CD medium exhibited greater fluorescence emission than in the presence of α-CD.[584] This is presumably due to the fact that the coumarin-6-sulfonyl chloride amino acid derivatives formed stronger inclusion complexes with β-CD relative to α-CD. For the iron(III)–2-pyridinecarbaldehyde-5-nitropyridylhydrazone complex, the enhancement factors are 1.5, 2.1, 2.4, and 4.1 in the presence of micellar cetylpyridinium bromide (CPB), Triton X-100, DTAB, and hexadecyltrimethylammonium bromide, respectively.[588] In addition, even for the same type of organized assembly system, the FI can depend upon the specific chemical nature of the ordered medium.[56,95,100–103,110,418,453,529,563,583,589] For example, depending upon the specific anionic charge-type surfactant micellar system employed, the enhancement of the fluorescence signal observed for diazepam ranged from 6.2 in SDS to 2.1 in sodium α-alkenesulfonate as surfactant.[523] The FI of the aluminum(III)–lumogallion complex in the presence of nonionic surfactant micelles varies by a factor of ca. 4.5, depending on the

specific structure and nature of the nonionic surfactant employed.[95] For metal ion complexes, the metal ion–ligand concentration mole ratio can also have substantial effects on the fluorescence enhancements observed in organized assemblies. Once a given organic or metal chelate analyte is identified for analysis, the most important factor governing the observed fluorescence emission is the nature of the surfactant employed, namely, the charge type and chemical structure of the surfactant, which are the leading properties required for determining the final degree of interaction between the analyte micelles.[100,101] In general, the better the binding to the organized assembly entity, the greater will be the effect on the solute's luminescence parameters.[454] An important factor is the charge of the ordered medium relative to the actual charge of the organic analyte (or metal chelate).[100,444] For ionizable species, the pH will obviously also impact on the magnitude of binding observed and, hence, the fluorescence properties.

The degree of fluorescence enhancement is also dependent upon the amount of organic cosolvent present in the organized assembly system.[66,110,173,271,590] Figure 7 illustrates the effect of added ethanol upon the relative FI of 3,4,4′-trichlorobiphenyl (TRCB) in an aqueous nonionic polyoxyethylene(10)lauryl ether surfactant micellar solution.[107,108] The fluorescence signal in the micellar system remains essentially constant with the percentage of ethanol up to ca. 30%. At this amount of added alcohol, the micellar aggregates disintegrate and the value dramatically drops, increasing slowly again with ethanol concentrations above ca. 40%. This latter increase is practically identical to

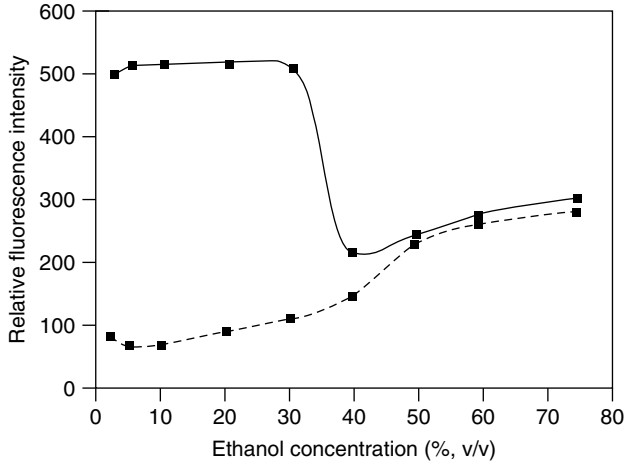

Figure 7 Plot of the relative fluorescence of an aqueous solution of 3,4,4′-trichlorobiphenyl (TRCB) as a function of added ethanol cosolvent in the absence (dashed line) and the presence of the nonionic surfactant polyoxyethylene(10)lauryl ether. (Reproduced with permission from Hernandez Garcia et al.[108])

that obtained in an aqueous ethanol solution in the absence of surfactant.[108] Other such examples have been reported in the literature. For instance, with 1-naphthol as the fluorophore, a 50% reduction in the fluorescence enhancement factor in the presence of CD was observed if ca. 30% (v/v) methanol, ca. 20% ethanol, or ca. 10% acetonitrile was also present in the solution.[271,272] Likewise, in the presence of a 0.05 M SDS micellar system, the fluorescence enhancement observed for the diuretic amiloride was 7.4-fold, whereas in the same system with 8% (v/v) added 1-propanol, the factor was only 4.3.[528]

In the case of CDs, added organic solvents (alcohols or nitriles) can result in either an additional enhancement or diminution of an analyte's fluorescence emission. If the additive can form a 1 : 1 : 1 ternary complex (both analyte and additive included within the CD cavity), then in most instances an additional increase in the FI is observed. For example, addition of 1-butanol further enhanced the fluorescence observed from fluorene in the presence of β-CD by a factor of ca. 1.7.[303] Likewise, addition of ethanol or acetonitrile enhanced the fluorescence emission of some coumarin dyes in β- and γ-CD.[327] It was thought that the small organic solvent molecules are also included in the CD cavity along with the coumarin molecule and that such solvent molecules displace water from the cavity which would otherwise quench the coumarin fluorescence.

However, in many situations with other fluorescent species, added cosolvents lead to diminished fluorescence from analyte–CD systems.[13] This is most often the result of the fact that a rapid decrease in the solute–CD binding constant is observed with increases in the percentage of added organic cosolvents, which can be ascribed to competition between the added cosolvent molecule and the solute molecule for inclusion in the CD cavity. In addition, an increase in the percentage of organic cosolvent means that the bulk solvent system is becoming more hydrophobic and more like the microenvironment of the CD cavity; hence the solute partitions to CD to a lesser extent. The net result of such diminished binding is that less protection is afforded to the analyte molecule and its excited state is more easily deactivated via nonradiative or quenching processes, which results in diminished fluorescence.

The pH can also impact on the magnitude of any fluorescence enhancement of ionizable analytes. For example, in the case of different flavonols, the enhancement factors observed in the presence of β-CD were different under acidic conditions (for the fully protonated form) compared to that under basic conditions (for the fully deprotonated form).[289,290] In almost all cases, a greater fluorescence enhancement factor was observed for the fully deprotonated flavonol. In the case of 1-naphthol, in the presence of micellar Triton X-100,

SDS, and CTAB, the fluorescence is enhanced by a factor of 90, 66, and 20, respectively, relative to that seen in aqueous solution in the absence of surfactant.[591] Under more basic conditions, the naphthol anion emission is also enhanced in the presence of CTAB and Triton X-100, whereas in SDS its intensity decreases compared with that in water. As noted in an earlier section, the presence of organized media can alter the pK values and function as pH shift reagents.

Of course, the binding of analytes to different organized assemblies does not always result in an enhancement in the luminescence signal. In fact, for some analytes, a diminution in intensity is observed.[592] For example, the fluorescence of the following compounds (or species) reportedly decreased upon binding to the indicated ordered media relative that is observed in the bulk aqueous medium: lumichrome (by β-CD)[204]; 2-(*p*-toluidinyl)naphthalene-6-sulfonate (by CTAB)[505]; 7-ethoxycoumarin and 7-ethoxy-4-methylcoumarin (by SDS and CTAC micelles)[61]; acridine (by CTAC, DTAC or SDS)[88,593]; calcein (by bile salt, Triton X-100, and zwitterionic sulfobetaine micelles)[594]; 4,6,8(14)-triene-3-one steroids (by β- and γ-CD)[595]; coumarin (by SDS, Triton X-100, and CTAB)[596]; benzo- and methylquinolizinium salts (by CTAB micelles)[563]; fluorescein (by Triton X-100, Tween 80, and Tween 20 micelles)[492,493]; PAHs (by CPB micelles, CMC = 0.64–0.73 mM)[597]; cresyl violet (by Triton X-100 micelles)[566]; 6-propionyl-2-(dimethylamino)naphthalene, termed *Prodan* (by SDS and DTAB micelles)[598]; niobium complex of 8-hydroxyquinoline-5-sulfonic acid (HQS) (by CPB micelles)[93]; oxandrolone and methyl testosterone (by SDS micelles)[531]; fluorescein-labeled gentamycin (by SDS micelles)[558]; and uranium-*o*-hydroxyhydroquinonephthalein chelate complex (by CTAC).[580]

Most of the negative results for CTAB or CPB micelles noted above stem from the fact that the bromide counterion and/or the pyridinium ion quench the luminescence of the indicated analytes.[74,349,563,588,597] Substitution of other cationic surfactants that do not possess these moieties might result in enhanced luminescence from some of these compounds/complexes. In addition, some organic molecules exhibit weak or no luminescence in organic solvents. Thus, transfer of such compounds from bulk water to an organized assembly phase (which is relatively less polar) results in decreased fluorescence. As an example, the fluorescence of acridine in micellar CTAC, DTAC, and SDS is only 8.0, 10.0, and 29%, respectively, relative to that observed in bulk water.[593] This stems from the fact that acridine fluoresces in water but not in organic solvents. Thus, transfer from the bulk aqueous phase to the less polar microenvironment of the micellar systems results in diminished fluorescence. Obviously, in such situations the use of aqueous-based organized assemblies would not prove advantageous unless one wanted to reduce background luminescence from such species.

3.1.2 *Mechanism of Fluorescence Enhancements in Organized Assemblies*

As illustrated by the data in Table 6, a variety of organic molecules (and also metal complexes) exhibit enhanced fluorescence upon binding to surfactant micellar or vesicle/liposome aggregates or CD molecules. Upon such incorporation, a number of potential mechanisms/factors are thought to be responsible for intensified luminescence, all of which are due to increases in the rate of the radiative process relative to the various competing nonradiative processes. These factors include shielding of the analyte species from quenching by bulk water, increased local viscosity in the binding site with concomitant reduction by oxygen quenching, reduced accessibility to solvent-borne quenchers and, perhaps, reduced quenching (rates) due to the more rigid local microenvironment of the particular organized assembly. As a consequence, the fluorescence quantum yield (and singlet excited-state lifetime) of the analyte fluorophore is often greater when it is ''bound'' to an organized assembly system relative to that observed in aqueous solution.[95,454,590,599–602]

In particular, isolation of analyte molecules which contain hydrogen-bonding moieties from water or other protic solvents possessing O–H or N–H bonds eliminates quenching from those solvent molecules and gives rise to enhanced fluorescence quantum yields (and intensity).[317,380,522,601,603] Thus, binding of such analytes possessing amino or hydroxyl moieties to organized assemblies would be expected to offer the prospects of enhanced fluorescence. In addition, it was reported that enhanced fluorescence in the presence of surfactant micelles is expected for organic analytes with nonrigid structures and/or that possess several phenyl groups in close proximity.[530] For acyclic azine molecules, transfer from a bulk solvent to a host micelle is thought to limit the rotational freedom of the azine molecule, leading to greater fluorescence efficiency.[494] In some instances, particularly in the case of metal chelate complexes, a portion of the intensified fluorescence can result from an increase in the molar absorptivity of that complex in the organized medium relative to that in a bulk solvent.[381,491,527] In the case of some metal complex species, enhanced intermolecular (or intramolecular) energy transfer efficiency also contributes to the intensified fluorescence observed in such systems.[275,590,604]

Additionally, the presence of organized species eliminates (or minimizes) alternative structures (dimers, aggregates, acid–base forms, etc., which have different

fluorescent and dynamic properties) for the analyte.[35] For some analytes, particularly dyes and some drug molecules, the addition of organized media causes increases in fluorescence as a consequence of their ability to alter the degree of aggregation of the analyte molecules.[100,501,516,553,556,605–613] Typically, the aggregated (dimer, tetramer, and multimer) forms of organic species exhibit weak or no fluorescence; hence deaggregation of the multimeric form due to the presence of the organized media and binding of the monomeric form (which exhibits greater fluorescence efficiency) leads to intensified luminescence.

For instance, the dimerization constants for the dyes MB, acridine orange, MO, and rhodamine R6G in water were reported as ca. 2300, 5200, 9200, and 5600 M^{-1}, respectively.[606,614] ''Dipole–dipole interactions within the resonating dimeric structure and/or a twisted molecular structure in the dimeric molecule can result in significant fluorescence quenching.''[547] The presence of appropriate organized media (i.e. surfactant present well above the CMC) can shift the monomer $\overleftarrow{\rightarrow}$ dimer equilibrium to the left with resultant enhanced fluorescence from the monomeric form of the dye.[547] The addition of Triton X-100 to an aqueous solution of rhodamine 6G prevents such dye aggregation, which permitted development of the first continuous-wave laser.[607] The addition of CD to relatively concentrated rhodamine-containing aqueous solutions also resulted in enhanced fluorescence (due to a shift in the dimer $\overleftarrow{\rightarrow}$ monomer equilibrium).[605] However, if CDs were added to less concentrated rhodamine solutions, then the fluorescence due to the rhodamine was somewhat quenched.

CDs were also utilized to control the aggregation equilibria and state of the dyes oxazine 1 perchlorate and oxazine 170 perchlorate.[610] Also, in water, cyanine dyes slowly form dimers and H aggregates such that the fluorescence of the monomer decreases. However, in the presence of SDS micelles, no dye aggregation occurs and only monomer fluorescence is observed.[39] Likewise, Triton X-100 or CTAB micelles solubilize chlorophyll in its monomer form, whereas in water, chlorophyll aggregates to different extents.[615]

3.2 Utilization of Organized Assemblies to Improve Fluorescence Methods for Inorganic and Organic Analytes

In general, new limits of sensitivity, and often improved selectivity, can be achieved in fluorimetric methods by analytically exploiting the unique properties offered by organized assembly systems. The use of micellar, vesicle, and CD media to enhance the FI of analyte molecules (species) has led to the development of fluorimetric methods that exhibit improved sensitivity [96,99–106,113,205–207,229,307,395,405,454,465,470,500,532,581,599,616–627] and also aids in the solubilization of the organic analyte or metal chelate,[107,196,518,599,628,629] eliminating the need to use organic solvents.[187,449,630,631] Other advantages reported include better selectivity,[307,405,532,599,616,619,622,632] wider calibration range (linear dynamic range),[589,599] greater accuracy and reproducibility,[599] greater flexibility in the choice of optimum pH[619,629] and method convenience,[518,599] greater stability of the luminescent species,[581,590] and, in some cases, eliminating the need for a prior extraction, cleanup, and/or chromatographic separation step.[518,583,624,633,634] Specific examples of the utilization of organized assemblies to improve the analytical performance of fluorescence methods for a variety of organic and inorganic analytes are outlined in the following sections.

3.2.1 Determination of Organic Analytes

Various types of organized assembly systems have been employed to enhance or improve fluorimetric methods for the determination of organic compounds.[14,173,635–677] Table 7 summarizes the conditions, analytical parameters and benefits derived from use of different organized media for the improvement of fluorimetric assays for some selected organic analytes. In most instances, the primary benefit of the organized medium is enhanced sensitivity. For instance, analytical fluorescence calibration data were determined for 13 PAHs in the presence of Brij-35 micelles.[222] Detection limits ranged from 0.005 ng mL^{-1} for chrysene to 0.2 ng mL^{-1} for naphthalene. The sensitivity and selectivity of PAH determinations in this medium are better than those obtained in methanol. The Brij-35 modified fluorescence method was evaluated for the determination of pyrene, benzo[*a*]pyrene, and benzo[*ghi*]perylene in particulate air samples.[222] A method for the resolution of multicomponent PAHs in micellar media via use of linear variable angle fluorescence and chemometric techniques has been reported.[481]

Derivative synchronous fluorescence spectrometry using polyoxyethylene(10)lauryl ether ($C_{12}E_{10}$) micelles was employed to determine simultaneously 3,3′, 4, 4′-tetrachlorobiphenyl (TTCB) and 2,3,7,8-tetrachlorodibenzofuran (TTCDF), with detection limits of 5.3 and 1.3–2.7 ng mL^{-1}, respectively.[639] The presence of $C_{12}E_{10}$ enhanced the FI of TTCB and TTCDF by factors of ca. 11 and 40, respectively, compared to that observed in an aqueous 2% (v/v) ethanol solution. Likewise, a simultaneous synchronous fluorimetric method for determination of benzo[*a*]pyrene and perylene in Triton X-100 micelles was developed, with detection limits of 0.27 and 0.30 ng mL^{-1}, respectively.[640] Binary mixtures of dibenz[*a*, *c*]anthracene and coronene were determined

Table 7 Summary of experimental and analytical parameters for the determination of organic analytes in the presence of different organized assembly media

Analyte	System conditions	Analytical parameters[a]	Comments	References
Stilbestrol	Triton X-100, pH 8.80–10.20	$\lambda_{ex} = 318$ nm, $\lambda_{em} = 415$ nm, LDR = 0.0–3.0 μg mL^{-1}, $D_L = 0.03$ μg mL^{-1}	TX-100 enhanced FI by factor of 9.0 and eliminated solubility problems	599
Fenproporex[b]	SDS, 20 °C	$\lambda_{ex} = 260$ nm, $\lambda_{em} = 560$ nm, $D_L = 0.8$ ppm	SDS or other anionic micelles enhanced FI by factors of 1.6–2.7	96
Procaine	β-CD	$\lambda_{ex} = 293$ nm, $\lambda_{em} = 330$ nm, LDR = 10–680 ng mL^{-1}, $D_L = 2.5$ ng mL^{-1}	β-CD enhanced FI by factor of ca. 12	498
Chlorophyll *a*	Triton X-100	$\lambda_{ex} = 431$ nm, $\lambda_{em} = 673$ nm LDR = 0–100 ng mL^{-1}, $D_L = 0.04$ ng mL^{-1}	TX-100 enhanced FI by factor of ca. 2000; no prior separation step required	518
Chlorophyll *b*	Triton X-100	$\lambda_{ex} = 462$ nm, $\lambda_{em} = 654$ nm $D_L = 0.09$ ng mL^{-1}	TX-100 enhanced FI by factor of ca. 3000	518
Indole-4-carboxylic acid	CTAB, pH 13	Synchronous maximum $\lambda^{\circ}_{s,em} = 288$ nm, $\Delta\lambda = 100$ nm, LDR = 0–350 ppb, $D_L = 1.0$ ppb	CTAB enhanced FI by factor of ca. 341	307
Propoxur[c]	CTAB	$\lambda_{ex} = 270$ nm, $\lambda_{ex} = 295$ nm, LDR = 0–4.8×10^{-6} M, $D_L = 5 \times 10^{-8}$ M	CTAB enhanced FI 1.8-fold	106
Carbendazim[d]	CTAB, 0.10 M NaOH	$\lambda_{ex} = 306$ nm, $\lambda_{ex} = 327$ nm, $D_L = 1.3 \times 10^{-8}$ M	CTAB enhanced FI by factor of 11.6	99
Carbaryl[e]	Nemol K1030[f]	$\lambda_{ex} = 336$ nm, $\lambda_{ex} = 450$ nm, $D_L = 1.4$ ng mL^{-1}	Nonionic Nemol micelles enhanced FI ca. 1.7-fold[g]	583
TTCB	POLE[h]	$\lambda_{ex} = 264$ nm, $\lambda_{ex} = 330$ nm, LDR = 0.5–10 ng mL^{-1}, $D_L = 1.4$ ng mL^{-1}	POLE micelles enhanced FI by factor of 6.8	107
TRCB	POLE[h]	$\lambda_{ex} = 260$ nm, $\lambda_{ex} = 328$ nm, LDR = 1.0–50.0 ng mL^{-1}, $D_L = 4.9$ ng mL^{-1}	POLE micelles enhanced FI by factor of 6.6	108
2-Naphthol	SDS	$\lambda_{ex} = 322$ nm, $\lambda_{em} = 355$ nm, LDR = 0.00–0.9 ppm, $D_L = 0.015$ ppm	SDS micelles enhanced FI by factor of 1.6	490
Kinetin[i]	SDS, pH 1.4–2.4	$\lambda_{ex} = 325$ nm, $\lambda_{em} = 500$ nm, LDR = 0.21–4000 ng mL^{-1}, $D_L = 0.21$ ng mL^{-1}	SDS micelles enhanced FI by factor of 9.5	589
Piroxicam[j]	β-CD, pH 3.5	$\lambda_{ex} = 320$ nm, $\lambda_{ex} = 440$–450 nm, LDR = 0.02–1.0 μg mL^{-1}, $D_L = 0.02$ μg mL^{-1}	β-CD enhanced FI by factor of ca. 9.4	395
Nalidixic[k]	γ-CD	$\lambda_{ex} = 314$ nm, $\lambda_{ex} = 357$ nm, LDR = 0.1–2.0 μg mL^{-1}, $D_L = 0.038$ μg mL^{-1}	γ-CD enhanced FI ca. two fold	582
7-Hydroxymethyl nalidixic acid	γ-CD, pH 2.0	$\lambda_{ex} = 258$ nm, $\lambda_{em} = 363$ nm, LDR = 0.10–4.00 ppm, $D_L = 0.02$ μg mL^{-1}	γ-CD enhanced FI by factor of ca. 2.0	635
Quercetin	Ethonium dichloride micelles[l]	$\lambda_{ex} = 436$ nm, $\lambda_{em} = 540$ nm, $D_L = 5 \times 10^{-6}$ M	Micellar medium enhanced FI by factor of 1.5	397
Anthracene and derivatives	SDS and γ-CD, pH 9.3	$\lambda_{ex} = 325$ nm, absolute detection limits estimated at subfemtogram levels	Micelles and CDs enhanced solubility; detection following capillary electrophoresis	636
Diazepam	SDS, 0.5 M H_2SO_4	$\lambda_{em} = 500$ nm, $D_L = 0.04$ ppm	SDS enhanced FI by factor of 5.0	620
Piroxicam	β-CD, pH 3.5			

(*continued overleaf*)

Table 7 (*continued*)

Analyte	System conditions	Analytical parameters[a]	Comments	References
Diuretics (7)	SDS, pH 3.0	$\lambda_{ex} = 270$ nm, $\lambda_{em} = 430$ nm, D_L range from 1.4 ng mL^{-1} (for bumetanide) to 54 ng mL^{-1} (for piretanide)	Fluorescence detection using SDS micellar mobile phase in HPLC separation	528
Methyl paraben[m]	α-CD, pH 4.5	$\lambda_{ex} = 259$ nm, $\lambda_{em} = 334$ nm, LDR = 0.04–3.0 ppm, $D_L = 11$ ng mL^{-1}	α-CD enhanced FI by factor of 12	532
Trovafloxacin	SDS, pH 5.5	$\lambda_{ex} = 270$ nm, $\lambda_{em} = 410$ nm	Method about 10x more sensitive in SDS medium as compared to water	553
Moxifloxacin	SDS, pH 4.0	$\lambda_{ex} = 294$ nm, $\lambda_{em} = 503$ nm	SDS enhanced FI by factor of 2.1 and minimized background signals from biological fluids	585
Trazodone	SDS, pH 7.0	$\lambda_{ex} = 326$ nm, $\lambda_{em} = 435$ nm, DL = 1.3 nM	SDS enhanced FI by factor of ~2	554
Danofloxacin(D) and Mabrofloxacin (M)	SDS, pH 3.0	$\lambda_{ex} = 300$ nm, $\lambda_{em} = 446$ nm, DL = 0.5 ng mL^{-1} (D), DL = 5.0 ng mL^{-1} (M)	HPLC with micelle as mobile phase	555
11-Methyl benzopheno-thiazine	β-CD	DL = 1 ng mL^{-1}	–	540
Bromodiolone	β-CD CTAB	$\lambda_{ex} = 310$ nm, $\lambda_{em} = 392$ nm, DL = 23 ng mL^{-1} (β-CD), DL = 37 ng mL^{-1} (CTAB)	β-CD and CTAB enhanced quantum yields by factor of 10	544
Histamine (as its fluorescamine derivative)	Brij-700 or SDS (pH 10.0) or CTAC (pH 7.0)	$\lambda_{ex} = 380$ nm, $\lambda_{em} = 460$ nm, DL = 0.5–33 ng mL^{-1} (depends on surfactant)	Micellar media enhanced derivative stability and sensitivity of method	556
Curcumin	Mixed micelle of CTAB and sodium dodecylsulfate	$\lambda_{ex} = 423$ nm, $\lambda_{em} = 492$ nm, DL = 0.017 ng mL^{-1}	Mixed micelles enhanced the FI	557

[a] LDR refers to linear dynamic range and D_L is the detection limit.
[b] Refers to (±)-3-(α-methylphenethylamino)propionitrile.
[c] Refers to 2-(methylethoxy)phenylmethyl carbamate.
[d] Refers to methylbenzimidazol-2-yl carbamate.
[e] Refers to 1-naphthyl *N*-methylcarbamate.
[f] Refers to nonionic surfactant nonylphenol ethylene oxide (11.7); CMC = 2.0×10^{-4} M.[107,108]
[g] Carbaryl in this basic micellar medium was instantaneously hydrolyzed to 1-naphthol, with Nemon 1030 having an effect comparable to that previously reported for an aqueous 10% (v/v) ethanol solution.
[h] Refers to polyoxyethylene(10)lauryl ether, $C_{12}E_{10}$.
[i] Refers to N^6-furfuryladenine.
[j] Refers to 4-hydroxy-2-methyl-*N*-2-pyridinyl-2*H*-1,2-benzothiazine-3-carboxamide 1,1-dioxide.
[k] Refers to 1-ethyl-1,4-dihydro-7-methyl-4-oxy-1,8-naphthyridine-3-carboxylic acid.
[l] Refers to ethylene-1,2-bis(*N*,*N*-dimethyl)carbodecoxymethyl dichloride.[626]
[m] Refers to 4-hydroxybenzoic acid methyl ester.
TTCB, 3,3′,4,4′-Tetrachlorobiphenyl.

in CTAB micelles by synchronous fluorescence spectrometry, with detection limits of 0.20 and 0.22 ng mL^{-1}, respectively.[474] The simultaneous determination of chrysene, perylene, and coronene was also reported, with detection limits of ca. 0.14 ng mL^{-1} for each of these PAHs. Constant-energy synchronous luminescence spectrometry (CESLS) in conjunction with SDS micelles allowed for the simultaneous identification and determination of benzo[a]pyrene, perylene, and pyrene in a mixture.[641] The presence of the micelles obviated the need to use an organic solvent and greatly enhanced the fluorescence signals, allowing detection of these PAHs in the low nanogram per milliliter levels (limits of detection ranged from 0.34 to 0.70 ng mL^{-1}).[641]

The utilization of micellar benzyldimethylhexadecylammonium chloride allowed for the determination of indole-4-carboxylic acid with a detection limit of 1.0 ng mL^{-1}.[307] CDs were employed to enhance the fluorescence of the pesticides bendiocarb and promecarb and detection limits of 0.57 and 0.091 μg mL^{-1}, respectively, were achieved.[539] The fat-soluble vitamins; A and E, can be easily solubilized in aqueous micellar solutions,

which allowed their fluorimetric quantification in milk samples[586] with detection limits of 54.0 and 55.2 ng L^{-1}, respectively, reported with Brij-35 as the micelle.

HRP-catalyzed fluorogenic reactions involving hydrogen peroxide with substrates such as HPPA has frequently been utilized in clinical and biochemical assays. In the conduct of such assays, a sequential approach or compromise of conditions is required since the enzyme activity, reaction rate, and product FI are not at maximum at the same pH values.[630] However, if the reaction is conducted in the presence of cationic surfactants, such as CTAB or CTAC, this difficulty can be overcome, as the reaction rate is 2.4–2.6 times greater than that in an aqueous solution. A CTAB micellar modified FIA method for the determination of HRP was developed with a linear dynamic range of 0.029–290 mg L^{-1} HRP ($\lambda_{ex} = 300$ nm, $\lambda_{em} = 405$ nm).[630] The micellar rate and pH effects on this reaction sequence should also be applicable for the enhancement of the fluorimetric determination of hydrogen peroxide.

The FI of an adduct of arginine with 2,3-naphthalenedicarbaldehyde ($\lambda_{ex} = 462$ nm, $\lambda_{em} = 520$ nm) in a weakly alkaline solution was enhanced ca. 12-fold by the addition of β-CD.[642] The reaction is highly specific for arginine, so no prior separation step is required in the procedure.

HRP catalyzes the reaction of acetaminophen and hydrogen peroxide with the formation of several unknown fluorescent products. The addition of Triton X-100 results in a fivefold increase in the observed fluorescence from this system.[643] This system is thought to be amenable to coupling to other reaction systems for the determination of substrates (or enzymatic activity). As an example, a linear calibration curve was obtained for glucose in the presence of the above system and glucose oxidase.

Since aqueous (or mixed aqueous–organic) solutions of organized assemblies have been utilized as chromatographic mobile phases or electrophoretic run buffers (*see* **Capillary Electrophoresis**; **Micellar Electrokinetic Chromatography**), their benefits with respect to fluorescence detection can sometimes be achieved with these separation methods.[318,449,463,482,548,644–647] The utilization of aqueous solutions of SDS as the mobile phase facilitated the separation and direct synchronous fluorescence detection of PAHs.[645] It was noted that the micellar matrix produced an enhancement effect on the fluorescent signals. In some instances, detection limits were poorer, owing to the presence of large background signals due to the organized assemblies.[648] For instance, the detection limits for PAHs separated by capillary zone electrophoresis (CZE) using 80:20 (v/v) water–DMF containing NaDC micelles in the run buffer were not as good as that achieved using CDs in the run buffer.[646] With CDs in the CZE run buffer, significant increases in the measured FI for 3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde (CBQCA)-tagged peptides were observed.[548] For instance, the relative FI of CBQCA–Des-Asp^1-angiotensin increased by almost a factor of 10.0 in the presence of 20 mM β-CD. The presence of α-CD also enhanced the fluorescence observed from the CBQCA derivatives of smaller peptides.[548] The HPLC detection of some PAHs was enhanced via the use of SDS micelles in the chromatographic mobile phase with fluorescence detection.[482]

Several articles review the enhanced fluorescence effects observed for dansylated amino acids due to the micellar mobile phases utilized for their HPLC separation,[14,463] as well as enhancements in the detection of environmental species separated by HPLC (or CE) that utilized micelles (or CDs) in the mobile phase (or run buffer).[14,449] A reversed phase HPLC method utilizing CDs in the mobile phase to enhance fluorescence detection was employed to determine aflatoxins in food and agricultural products.[649] Capillary electrophoresis utilizing a mixed β-CD/sodium deoxycholate run buffer system was found to resolve the optical isomers of D/L-serine (as their naphthalene-2,3-dicarboxaldehyde (NDA) derivatives) with laser-induced fluorescence as a means of detection (limit of detection was 3.0×10^{-8} M).[650] Alanine, glutamate, and γ-aminobutyric acid were also determined using fluorescence detection following their CE separation with a mixed micellar (SDS and sodium cholate) run buffer (detection limits in the range of 11–40 nM).[647] In addition, catecholamines and amino acids (derivatized with 4-chloro-7-nitro-2,1,3-benzoxadiazole, NBD) were separated and determined by CE via use of SDS micelles in the run buffer. The SDS served to improve the fluorescence of the NBD-labeled catecholamines.[456] The CE analysis of the mycotoxin zearalenone with CD enhanced fluorescence (limit of quantitation = 5 ng g^{-1} for maize sample) has been reported.[538]

The effect of SDS on the fluorescence observed from histone–ANS complexes has been reported.[651] The enhancement of ANS fluorescence caused by different groups of histones in the presence of SDS is roughly the same. This allowed the development of an ANS staining method for histones that is rapid, economical and simpler than the existing method for such staining of histones separated via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Enhanced fluorescence due to the presence of organized assemblies can also sometimes be employed to determine the concentration of that organized component. For example, a very sensitive fluorescence assay for phospholipid vesicles using DPH has been reported.[486,555] The FI of DPH is proportional to the phospholipid concentration, with a detection limit of

0.70 nmol reported.[486] This optimized method compared favorably with the classical, conventional methods for analysis of such vesicle systems in terms of sensitivity, accuracy, and convenience.

An HPLC method in which the fluorescence enhancement of probe molecules was utilized as the detection mode has been reported for the determination of CDs.[271,652,653] Addition of the fluorescent probe molecule to the mobile phase results in an *in situ* complexation (with resultant enhanced FI), making the use of a postcolumn reaction detection system unnecessary. With 1-naphthol as the probe molecule in the mobile phase, a detection limit of 90 pmol was reported for β-CD.[271] Use of 2-*p*-toluidinonaphthalene-6-sulfonate (TNS) as the probe molecule resulted in HPLC mass detection limits of 3.7, 0.050, and 0.14 μg for α-, β-, and γ-CD, respectively.[653] Derivatives of β-CD can be determined in an analogous manner. In similar fashion, TNS or ANS was utilized in the run buffer to enhance the fluorescence detection of nonabsorbing analytes (anionic surfactants such as SDS) in CZE and micellar enhanced/electrokinetic capillary chromatography (MECC or MEKC).[471,495] Detection limits of 0.1–0.4 mM SDS were reported, although the reproducibility was relatively poor. Also, any analyte that partitions into the micelle (present as part of the run buffer) will displace some of the added fluorescent probe molecule with a concomitant decrease in the fluorescence signal, which allows indirect detection of nonabsorbing solutes in MECC.[495] For instance, using ANS as the probe in conjunction with SDS in the run buffer allowed indirect fluorescence detection of the analytes phenol, 3,5-dimethylphenol, and naphthalene. In addition, a flow-injection titration method has recently been reported for determination of surfactants (SDS, SB-12 (*N*-dodecyl-*N,N*-dimethyl-3-ammonio-1-propanesulfonate), Triton X-100, CTAB) based upon the enhanced fluorescence observed for TNS in their presence.[653]

There are some reports concerning the use of dipping or spray reagents containing organized surfactant assemblies to enhance fluorescence detection of organic analytes following their separation by thin-layer chromatography (TLC). For example, the use of Triton X-100 dissolved in chloroform or benzene as a spray reagent enhanced the fluorescence signal of the analyte ethoxyquin by a factor of >200![654] After application, the organic solvent was evaporated, leaving the pure, viscous Triton X-100 on the silica gel TLC sheet, which resulted in the enhanced emission signal. Likewise, detection of 7-fluoro-2,1,3-benzoxadiazole-4-sulfonate (SBD) derivatized gluthathione was improved by a factor of ca. 2.0 via the use of organized assemblies in a fluorescent enhancer dipping reagent following separation via high-performance TLC on silica gel.[655] For this analyte, the fluorescence enhancement ratios obtained were 1.80, 1.85, 1.86, and 2.16 for 40% (v/v) Triton X-100 in toluene and 20 mM aqueous solutions of α-, γ-, and β-CD, respectively.[655] The detection limit achieved was 17 pg of gluthathione per spot.[656] A fiber-optic remote sensor for *in situ* fluorimetric quantification in TLC has also been reported.[657] The fluorescence signal was again enhanced by dipping the silica gel plates in a mixture of Triton X-100 and chloroform. Detection limits of 5.3 and 6.4 μM were reported for the analytes acepromazine and propiomazine, respectively.[657] Significant enhancements in fluorescence for dansylated amino acids and aromatic hydrocarbons were observed on silica gel and alumina TLC plates when using CTAC or NaC micellar or β-CD spray reagents.[658]

Some of the same surfactant micelle systems (nonionic, zwitterionic) that can be employed to enhance the FI from organic analytes can also be utilized above their cloud point (CP) for extractive preconcentration (*see* **Waste Extraction Procedures**; **Cyanogen Chloride and Cyanogen Bromide Analysis in Drinking Water**). Although nonionic surfactants dissolve in aqueous solution to form micellar aggregates, above a temperature referred to as that surfactant's CP, which is characteristic of the surfactant employed and the conditions, the surfactant phase separates from the bulk aqueous phase akin to a conventional liquid–liquid extraction process[94] (for a review, see Hinze et al.[659]). Thus, via the use of appropriate surfactants and conditions, it is possible to combine this CP extraction with fluorescence detection to achieve high concentration factors (from the CP extraction step) and also fluorescence enhancements due to the presence of the surfactant-rich aggregate phase, both of which serve to improve analyte detectability. As examples of this technology, the nonionic surfactant Genapol X 80 was employed in the CP extraction and fluorescence detection of the pesticides napropamide and thiabendazole from water and soil samples.[660] The fluorescence enhancement factors obtained in the surfactant-rich phase were 1.2–2.0 for these two pesticides. Detection limits below 0.2 μg L^{-1} were achieved with recovery rates of up to 95%. In addition, the acid-induced CP extraction using SDS served as a preconcentraion step prior to the fluorimetric determination of a number of PAHs.[661]

The presence of micellar media can often enhance the photostability of organic molecules. For instance, while irradiation of a cyanine dye in aqueous solution with a 450 W Xe lamp through a 400 nm cut-off filter resulted in a complete depletion of the dye within 30 min, no fading whatsoever was detected after 24-h irradiation in the presence of an SDS micellar solution.[39] Likewise, rhodamine 6G was found to be much more stable in SDS micellar media than in bulk methanol or ethanol

when submitted to near-ultraviolet/visible (UV/VIS) illumination from a continuous xenon arc lamp. This is a consequence of the fact that the photodegradation reaction(s) is inhibited in the presence of micellar media.

Pyridoxal is oxidized in the presence of cyanide ion to form 4-pyridoxolactone as the fluorescent product. The use of cationic CTAB resulted in an enhancement of both the reaction kinetics (by a factor of ca. 2) and also the quantum yield of the product.[662] The limit of detection of pyridoxal using the micellar-enhanced fluorescent FIA approach was 0.17 ng, which is three times lower than that achieved in an aqueous solution in the absence of surfactant.

Micellar media have also been employed to enhance some fluorescence immunoassay techniques for determination of biological/clinical molecules. Fluorescent dyes are typically employed to label analyte molecules in ligand binding assays. The assay requires that there be a fairly large difference in the fluorescence properties of the bound compared with the free labeled analyte. Halfman et al.[558] reported that SDS micelles quenched the emission of free fluorescein-labeled gentamycin without impacting the emission intensity of labeled gentamycin bound to gentamycin antibody, thus fulfilling the requirement for a large fluorescence difference between the bound and free labeled species. The micellar quenching effect thus allowed the development of a convenient homogeneous fluorescence immunoassay for gentamycin based only on FI measurements. Guidelines for the design of similar assays for other analytes and the charge-type surfactant micelles that would be required were also discussed. More recently, specially derivatized CDs have been developed that allow for the quenchofluorimetric determination of organic species, including enantiomeric compositions.[162–165,663,664] For instance, a β-CD modified to contain the dansyl fluorophore was synthesized and shown to serve as an enantioselective fluorescent sensor for amino acids. In the absence of enantiomeric analytes, the dansyl group, although chemically bound the CD, is included in the CD cavity and exhibits intense fluorescence. In the presence of an analyte, the analyte can displace the dansyl group, which when in contact with bulk solvent exhibits much diminished fluorescence. The decrease in dansyl fluorescence is proportional to the analyte concentration. Such approach has been employed as a means of sensing amino acids as well as endocrine-disrupting chemicals. Other fluorescent moieties have been appended to cyclodextrins for application as sensors (*see* **Fluorescence-based Biosensors**; **Biosensor Design and Fabrication**) for other analyte groups.

The quenching of fluorescence from acridine orange sensitized rhodamine 6G (or B) system in sodium dedecylbenzenesulfonate micelles upon the addition of analytes (such as erythromycin and vitamin B_{12}) has been employed to quantitate those analytes.[637,638]

In closing this section, it should be noted that more information and examples concerning the use of CDs in analytical fluorimetry and diagnostics can be found in a book chapter[439] and a recent review article.[173] In addition, a review by McGown et al.[665] focuses on luminescence measurements in bile salt surfactant systems.

3.2.2 *Photochemically Induced Fluorescence (PIF) Detection*

For weakly fluorescent and nonfluorescent analytes, derivatization to convert the analyte into a more detectable product (i.e. better fluorescing) constitutes a means of enhancing that analyte's detectability. In addition to chemical reactions, an alternative approach is to use a photochemical reaction.[14,666,667] In this case, intense UV/vis radiation converts nonfluorescent or weakly fluorescent analytes of interest into highly fluorescent products. Such a PIF approach has been employed as a postcolumn derivatization technique for detection in chromatographic systems.[449,668]

A flow-injection fluorimetric method for determination of vitamin K_1 based on a photochemical reaction in the presence of Triton X-100 micelles has been reported.[217,669] As previously noted, the primary function of the micellar media was to enhance the stability and solubility of different reactive species formed during the photochemical reaction. This method permits the determination of vitamin K_1 in the 10^{-6}–10^{-4} M concentration range. The anionic surfactant SDS was employed in the HPCL-PIF determination of K vitamins.[631] A FIA-PIF method for the determination of menadione in the presence of SDS with a detection limit of 0.18 mg mL^{-1} has been developed by Perez-Ruiz.[669] The determination of phylloquinone in fruits and vegetables by FIA with SDS micelle–enhanced PIF (detection limit 0.05 μg mL^{-1}) has been reported.[541]

A rapid micellar-enhanced PIF-FIA method for the determination of nonfluorescent sulfonylurea herbicides in water has been developed.[670] In this application, the primary purpose of incorporation of surfactant micellar SDS or CTAC in the carrier stream was to improve the solubility of the herbicides in an aqueous medium. In addition, the photoinduced fluorescence signals were greater by a factor of 1.1–4.1 for the four herbicides examined in the presence of the micelles compared with those in aqueous solution in the absence of surfactant micelles. Detection limits of 0.2–0.5, 0.1–0.2, 0.2–0.8, and 1.0–6.0 ng mL^{-1} were reported for the herbicides chlorsulfuron, methsulfuron-Me, 3-rimsulfuron, and sulfometuron-Me, respectively.[670,671] In similar fashion, a photochemical–spectrofluorimetric method

for determination of two pyrethroid insecticides and four phenylurea herbicides using an anionic and/or cationic surfactant micellar medium and CDs has been reported.[672–674] Likewise, such micellar-enhanced photochemically induced fluorescence approaches have been employed for the determination of other herbicides (e.g. mecoprop, chlorphenoxyacid, phenylurea residues, etc.) using time-resolved and/or flow-injection PIF techniques[449,667,675,676] (*see* **Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of**).

Upon irradiation, mazindol, an anorectic agent, forms a photoproduct whose fluorescence is much greater than that of maxindol itself. Irradiation of mazindol solutions in the presence of micellar SDS provokes a further enhancement in their fluorescence and the maximum FI is reached faster than in the absence of the surfactant.[529] After a 5-min irradiation time, the FI is roughly 42 times greater than that observed in aqueous solutions. This micellar-enhanced photochemical method for mazindol has a detection limit of 3.6 ng mL^{-1} (in the absence of SDS the detection limit is 65 ng mL^{-1}).[529]

The effects of different CD media upon the photolysis of dissolved anthracene by fluorometry has been reported. Derivatized CDs promoted the photolysis reaction to a greater extent relative to the native, underivatized CDs.[191] In view of the effect of organized assemblies upon the rate of photophysical processes, it is thought that this general type of application deserves much greater attention and will lead to the development of very sensitive assays for nonfluorescent or weakly fluorescent analytes. In particular, the use of this general PIF approach as a means of HPLC detection[677] in conjunction with micellar or CD mobile phases (or run buffers) in liquid chromatography (or capillary electrophoresis or electrochromatography) should prove to be very fruitful.

3.2.3 *Fluorimetric Determination of Inorganic Ions and Organometallic Species in Organized Media*

The use of surfactant micellar systems and, to a much lesser extent, CDs can increase the stability and solubility of many organic ligands and metal complexes in water and, when such metal complexes are fluorescent, a marked enhancement in the fluorescence signal is frequently observed. Consequently, the use of such organized assemblies can improve many existing spectrofluorimetric methods by increasing their sensitivity and/or eliminating the need for extraction procedures or the use of mixed aqueous–organic solvents.[449,577,588] In addition, as has been previously noted, the possibility of conducting reactions and forming metal chelates in micellar (or other organized) media that are not observed in a bulk homogeneous solvent system can lead to new improved fluorescence methods for select inorganic species.[678–721] Some representative data in this regard is summarized in Table 8.

The fluorescence of different metal ion–morin complexes has been enhanced by the presence of surfactant micelles.[381,680–684] For example, cationic surfactants such as CTAB enhanced the intensity of Nb(V) and Ta(V) complexes, that of the former by ca. 80–400-fold.[100,628] The quantum yields of the morin–Nb(V) complex (pH 3.0) were reported as ca. 0, 0.292, 0.30, and 0.32 in water, ethanol, aqueous cetylpyridinium chloride (CPC) (hexadecylpyridinium chloride) micelles, and aqueous CTAB micelles, respectively.[381] The molar absorptivity values for the morin–Nb (V) complex at the excitation wavelength in the micellar media were greater relative to that seen in water or ethanol as bulk solvents.[381] Nonionic surfactants, such as Genapol PR-20, enhanced the intensity of the Al(III), Pb(II), and Zn(IV) complexes of morin by factors of 8.0, 9.0, and 75, respectively.[622] Triton X-100 micellar media enhanced the fluorescence observed from the morin–Ta(V) and morin–Nb(V) complexes by factors of 85 and 49, respectively.[100] The FI of the morin complexes of Zr(IV), Hf(IV), Al(III), Ga(III), In(III), and Sb(V) were enhanced by factors of ca. 2–13 in the presence of anionic SDS micellar media.[681,685]

In addition to enhanced sensitivity, the presence of organized assemblies in certain instances can serve to reduce the number of potential interferences. For example, in the fluorescence determination of Al(III) via the use of the complexing agent morin, in nonionic surfactant micelles, interference due to Cr^{3+}, Mg^{2+}, NH_4^+, and PO_4^{3-} was eliminated.[622] In fact, in the absence of micelles, phosphate ion seriously interferes with this method. However, this interference is completely eliminated owing to the presence of the micellar medium.

Dasgupta et al. conducted a comprehensive study of the fluorescence properties of the metal complexes of HQS in aqueous, mixed aqueous (water–DMF), and aqueous surfactant CTAC media.[686,687] Modest to significant increases in fluorescence intensities (enhancements in the range 1.2–10.2-fold) were observed in the presence of the surfactant CTAC medium for the metals Mg(II), Ca(II), Sr(II), Ba(II), Al(III), Cd(II), and Zn(II). The optimum pH for complex formation (and FI) was shifted (reduced) by 1–2 units in the CTAC system.[686] In addition, the pH range for maximum fluorescence was extended in this system. It was also noted that the presence of iron(III) quenches the fluorescence of these metal–HQS complexes and that the presence of the CTAC medium accentuates this quenching effect. The simultaneous determination of Al^{3+} and Zn^{2+} by first- and second-derivative synchronous fluorimetry via the use of HQS in the presence of CTAB has also been reported.[688]

Table 8 Summary of data concerning the fluorimetric determination of selected inorganic species in different organized assembly systems

Analyte	System conditions	Analytical parameters[a]	Comments	References
Al(III)	8-Quinolinol, SDS, pH 7.0	$\lambda_{ex} = 370\,nm$, $\lambda_{em} = 504\,nm$, D_L[a] $= 1.0$ ppb	Reversed phase HPLC separation with SDS micellar mobile phase; SDS enhanced FI by factor of ca. 4.3	695
	Ferron, didodecyldimethy-lammonium bromide vesicles, pH 5.8	$\lambda_{ex} = 386\,nm$, $\lambda_{em} = 493\,nm$, $D_L = 4\,ng\,mL^{-1}$	DODAB vesicles enhanced FI	151
	Acid chrome blue–black, Triton X-100, (or SDS) micelles, pH 5.5	$\lambda_{em} = 585\,nm$, $D_L = 0.16–0.40\,ng\,mL^{-1}$	Triton X-100 enhanced FI by factor of ca. 8.0	503
	HQS, CTAC, pH 6.0	$\lambda_{ex} = 380\,nm$, $\lambda_{em} = 490\,nm$, LDR $= 10–500\,\mu g\,mL^{-1}$, DL $0.50\,\mu g\,mL^{-1}$	CTAC, FIA with multisyringe based system	693
	HQS, CTAB, pH 5.5	Synchronous fluorescence, $\Delta\lambda = 107\,nm$, LDR $= 5–50\,ng\,mL^{-1}$, $D_L = 1.1\,ng\,mL^{-1}$	First and second derivative synchronous mode; CTAB enhanced FI by factor of 7.0	621
	HQS, DODAB[b] vesicles, pH 6.0	$\lambda_{ex} = 380\,nm$, $\lambda_{em} = 495\,nm$, LDR[a] $= 1–100\,\mu g\,L^{-1}$, $D_L = 1\,\mu g\,L^{-1}$	DODAB vesicles enhanced FI by factor of ca. 10	704
	Lumogallion, TX-100, pH 5.0	$\lambda_{ex} = 450\,nm$, $\lambda_{em} = 530\,nm$, DL $= 0.70$	Method validated by NIST standard and interlaboratory assessments	697
	Lumogallion, Triton X-100	$\lambda_{ex} = 522\,nm$, $\lambda_{em} = 566\,nm$, LDR[b] $= 0–2$ ppb, $D_L = 0.05$ ppb	Triton X-110 CP extraction conducted; TX-100 enhanced FI by factor of ca. 2.3	700
	Lumogallion, Brij-35, pH 5.46	$\lambda_{ex} = 484\,nm$, $\lambda_{em} = 552\,nm$, $D_L \approx 0.15$ nm	FIA method; Brij-35 enhanced FI by factor of 5.1	453
	Lumogallion, pH 5.0, Antarox CO 890[c]	$\lambda_{ex} = 365\,nm$, $\lambda_{em} = 558\,nm$, $D_L = 5 \times 10^{-10}\,g\,mL^{-1}$	Nonionic surfactant micelles enhanced FI by factor of 6.1	659
	Lumogallion, ethonium dichloride micelles, pH 6.0	$\lambda_{em} = 546\,nm$, $D_L = 0.03\,ng\,mL^{-1}$	Micelle enhanced FI by factor of 4.0	626
	Morin, pH 3.8, Genapol PF-20[d]	$\lambda_{ex} = 430\,nm$, $\lambda_{em} = 495\,nm$, $D_L = 0.2$ ppb	Nonionic surfactant micelles enhanced FI 10-fold	622
Be(II)	Morin, pH 6.0, Triton X-100	$\lambda_{ex} = 440\,nm$, $\lambda_{em} = 530\,nm$, $D_L = 0.04\,ng\,mL^{-1}$	TX-100 enhanced FI by factor of 5.0	113
Cd(II)	TSQ[e], pH 8.5, SDS	$\Delta\lambda = 135\,nm$, LDR = $15–250\,\mu g\,mL^{-1}$	Derivative synchronous mode; SDS (and Brij-35) micelles enhanced FI	634
Fe(III)	MPAQ[f], CTAB	$\lambda_{ex} = 316\,nm$, $\lambda_{em} = 528\,nm$, LDR $= 10–700\,ng\,mL^{-1}$, $D_L = 10.0\,ng\,mL^{-1}$	CTAB enhanced FI by factor of 3.0	616
	PCNPH[g], CTAB, pH 7.0	$\lambda_{ex} = 300\,nm$, $\lambda_{em} = 420\,nm$, LDR $= 0.20–1.45\,\mu g\,mL^{-1}$, $D_L = 0.028\,\mu g\,mL^{-1}$	CTAB enhanced FI by factor of 4.1	588
	HHQP[h], pH 3.5, Brij-58	$\lambda_{ex} = 470\,nm$, $\lambda_{em} = 525\,nm$, LDR $= 0–300\,ng\,mL^{-1}$, $D_L = 5.0\,ng\,mL^{-1}$	Quenching method, Brij-58 enhanced the quenching effect by factor of 1.5	659
	5-(4-methoxy-phenylazo)-8-(4-toluenesulfonamido) quinoline (MPATSQ), CTAB, (HCl) = 0.014 M	$\lambda_{ex} = 317\,nm$, $\lambda_{em} = 534\,nm$, LDR $= 2.1$ nM–$3.0\,\mu M$, DL $= 2.1$ nM	–	694
Ga(III)	Lumogallion, pH 2.0, ethonium dichloride[i]	$\lambda_{em} = 546\,nm$, $D_L = 1.0\,ng\,mL^{-1}$	Micelles enhanced FI by factor of 7–8	626

(*continued overleaf*)

Table 8 (*continued*)

Analyte	System conditions	Analytical parameters[a]	Comments	References
	PAN[j], SDS, pH 4.0	$\lambda_{ex} = 546$ nm, $\lambda_{em} = 573$ nm LDR = 5–400 ppb, $D_L = 0.8$ ppb	SDS enhanced FI by factor of 10	106
	DCMHQ[k], Triton X-100 or SDS micelles	LDR = 5–50 ng mL^{-1}, $D_{L(SDS)} = 1.6$ ng mL^{-1}, $D_{L(TX-100)} = 1.5$ ng mL^{-1}	Presence of micellar media enhanced FI	396
Hf(IV)	Morin, 0.25 M H_2SO_4, CPC	$\lambda_{ex} = 436$ nm, $\lambda_{em} = 512$ nm, $D_L = 0.01$ μg mL^{-1}	CPC micellar medium enhanced FI by factor of 2.0	707
In(III)	Lumogallion, pH 5.0, ethonium dichloride	$\lambda_{em} = 546$ nm, $D_L = 4.0$ ng mL^{-1}	Micelles enhanced FI by factor of 7–8	626
Mg(II)	HQS, pH 9.5, CTAB	$\lambda_{ex} = 393$ nm, $\lambda_{em} = 505$ nm, $D_L = 0.005$ μg mL^{-1}	Method applied for determination of Mg(II) in inorganic salts	689
Mo(VI)	Alizarin red S, pH 4.3, CTAB	$\Delta\lambda = 110$ nm, LDR = 32–4000 ng mL^{-1}, $D_L = 9.5$ ng mL^{-1}	Derivative synchronous mode; CTAB enhanced FI of the complex	623
Nb(V)	Lumogallion, pH 1.0, Triton X-100	$\lambda_{ex} = 495$ nm, $\lambda_{em} = 600$ nm $D_L = 7.77$ ppb	Triton X-100 enhanced FI by factor of 19.5	454
	Morin, CTAB, 1 M H_2SO_4	$\lambda_{ex} = 433$ nm, $\lambda_{em} = 503$ nm, $D_L = 1.0$ ppb	CTAB enhanced FI by factor of 80	100,628
	Morin, CPC, 3 M H_2SO_4	$\lambda_{ex} = 436$ nm, $\lambda_{em} = 512$ nm, $D_L = 0.5$ ng mL^{-1}	No fluorescence observed in water; in CPC micelles, complex is fluorescent	381
	Quercetin, CTAB, 0.3 M H_2SO_4	$\lambda_{ex} = 435$ nm, $\lambda_{em} = 497$ nm, $D_L = 3.0$ ppb	CTAB enhanced FI by factor of 65	100
Pb(II)	Morin, pH 3.3, Genapol PR-20[d]	$\lambda_{ex} = 420$ nm, $\lambda_{em} = 495$ nm, $D_L = 0.06$ μg mL^{-1}	Genapol PF-20 enhanced FI by factor of ca. 9.0	669
Sc(III)	Quinizarin[l], pH > 12, β-CD	$\lambda_{ex} = 545$ nm, $\lambda_{em} = 600$ nm, LDR = 0.5–400 ng mL^{-1}, $D_L = 0.15$ ng mL^{-1}	β-CD aids in the dissolution of the chelate and enhances its stability	672
Tributyltin ion	Morin, Triton X-100, pH 5.7	$\lambda_{ex} = 408$ nm, $\lambda_{em} = 534$ nm, LDR = 0.03–1.0 μg mL^{-1} $D_L = 16$ ng mL^{-1} (as Sn)	Postcolumn fluorescent detection mode; the Triton X-100 micelles solubilized the organotin and morin, enhanced FI and increased flexibility of pH requirements	629
Sn(II)	PAN, mixed TX-100 and bis(2-ethylhexyl) sulfosuccinate (AOT)	$\lambda_{ex} = 300$, $\lambda_{em} = 360$, DL = 2 ng mL^{-1}	Mixed micelles served to solubilize the tin complex and enhanced FI	674
Ta(V)	Morin, CTAB, 0.2 M H_2SO_4	$\lambda_{ex} = 422$ nm, $\lambda_{em} = 496$ nm, $D_L = 0.03$ ppm	CTAB enhanced FI by factor of 400	100
	Quercetin, CTAB, 0.05 M H_2SO_4	$\lambda_{ex} = 430$ nm, $\lambda_{em} = 485$ nm, $D_L = 0.05$ ppm	CTAB enhanced FI by factor of 37	100
Tb(III)	BPMPHD[m], Gd^{3+}, CTAB, pH 5.5	$\lambda_{ex} = 305$ nm, $\lambda_{em} = 545$ nm, $D_L = 2.0 \times 10^{-12}$ M, LDR = 2.0×10^{-11}–8.0×10^{-6}	CTAB enhanced FI by factor of 20	590
W(VI)	Alizarin red S, pH 4.3, CTAB	$\Delta\lambda = 110$ nm, LDR = 49–2500 ng mL^{-1}, $D_L = 14.7$ ng mL^{-1}	Derivative synchronous mode; CTAB enhanced FI of the complex	623
Zn(II)	TSQ[e], Brij-35, pH 7.0–8.7	$\lambda_{ex} = 367$ nm, $\lambda_{em} = 515$ nm, LDR = 8.0–550 ng mL^{-1}, $D_L = 3.0$ ng mL^{-1}	Brij-35 solubilized the Zn–TSQ complex and enhanced FI by factor of ca. 3.0	633
	DCMHQ[k], pH 6.0, CTAB or Brij-35	$\lambda_{ex} = 397$ nm, $\lambda_{em} = 534$ nm, LDR = 3–100 ng mL^{-1}, $D_L = 3.0$ ng mL^{-1}	CTAB (or Brij-35) enhanced FI by factor of ca. 25.0	624
Zn(II)	Morin, pH 4.7, Genapol PF-20[d]	$\lambda_{ex} = 433$ nm, $\lambda_{em} = 503$ nm, $D_L = 3.0$ ng mL^{-1}	Nonionic micelle enhanced FI by factor of 75.0	109
	Morin, 0.12 M H_2SO_4, CTAB	$\lambda_{ex} = 424$ nm, $\lambda_{em} = 508$ nm	CTAB enhanced FI by factor of ca. 31.0 and lowered the pK_a of morin	675

(*continued overleaf*)

Table 8 (*continued*)

Analyte	System conditions	Analytical parameters[a]	Comments	References
	Morin, 0.25 M H_2SO_4, CPC	$\lambda_{ex} = 436$ nm, $\lambda_{em} = 512$ nm, $D_L = 0.01\,\mu g\,mL^{-1}$	CPC enhanced FI by factor of 2.0	707
Zr(IV) and Ga(III)	QNPH[n], TTAC[o], pH 4.5–5.3	$\lambda_{ex} = 345$ nm, $\lambda_{ex} = 543$ nm, LDR (Ga) = 0.0–28.0 ng mL^{-1}, LDR (Zr) = 0.0–36.0 ng mL^{-1}	TTAC enhanced FI of reagent QNPH and the degree of quenching observed with added Zr(IV) or Ga(III) cations	677
Zr(IV)	HNA[p]	$\lambda_{ex} = 368$ nm, $\lambda_{em} = 468$ nm, DL = 17 ng mL^{-1}	–	721
EDTA[q] and its salts	Zr(IV) – Alizarin red S, CTAB, pH 4.7	$\lambda_{ex} = 478$ nm, $\lambda_{em} = 600$ nm, $D_L = 3.4$ ng mL^{-1}	CTAB micelles	405

[a] D_L refers to detection limit and LDR to linear dynamic range.
[b] Refers to DDAB.
[c] Refers to nonionic surfactant nonylphenoxypoly(ethyleneoxy)ethanol.
[d] Refers to nonionic surfactant, ethylene oxide–propylene oxide condensate in which the propylene oxide : ethylene oxide ratio is 1 : 6.
[e] Refers to 8-(*p*-toluenesulfonamido)quinoline.
[f] Refers to 5-(4-methylphenylazo)-8-aminoquinoline.
[g] Refers to 2-pyridinecarbaldehyde-5-nitropyridylhydrazone.
[h] Refers to *o*-hydroxyhydroquinonephthalein.
[i] Refers to nonionic surfactant polyethylene glycol monolauryl ether.
[j] Refers to 1-(2-pyridylazo)-2-naphthol (PAN).
[k] Refers to 5,7-dichloro-2-methyl-8-hydroxyquinoline.
[l] Refers to 1,4-dihydroxyanthraquinone.
[m] Refers to 1,6-bi(1′-phenyl-3′-methyl-5′-pyrazolone)hexanedione (diacylated pyrazolone).
[n] Refers to *o*-hydroxyhydroquinonephthalein.
[o] Refers to tetradecyltrimethylammonium chloride.
[p] Refers to 3-hydroxy-2-naphthoic acid.
[q] Refers to ethylenediaminetetraacetic acid.

Fluorimetric methods for Al^{3+}, Mg^{2+}, and Zn^{2+} based upon the formation of their 8-hydroxy-7-iodoquinoline-5-sulfonic acid chelates in the presence of cationic surfactants such as CTAB have also been reported by other groups.[689–694] CTAB micelles enhanced the fluorescence observed for some of these systems, allowing better detection limits. For instance, the sensitivity of the fluorimetric procedure for Al(III) improved by a factor of 8.0 in the presence of CTAB.[690]

The use of SDS micelles in the mobile phase allowed the reversed-phase HPLC separation and fluorescence detection of Al(III) as its 8-quinolinolato complex.[695] The FI was about four times greater than in bulk homogeneous solution, allowing a limit of detection of 1.0 ppb. In another report, micellar liquid chromatography with a CTAB micellar mobile phase allowed for the determination of $10\,\mu g\,mL^{-1}$ Al(III) as its HQS complex.[696] The effect of CPB, CTAB, and Triton X-100 micellar media upon the Ta(V), Ti(V), Al(III), and Nb(V) chelates with lumogallion, ferron, quercetin (3,3′,4′,5,7-pentahydroxyflavone), morin (3,5,7,2′, 4′-pentahydroxyflavone), oxine-5-sulfonic acid (HQS), and Alizarin red S has been reported.[101,678,696,697] In most instances, the presence of these surfactant micelles dramatically enhanced the FI. Mechanistic aspects of these systems and factors impacting on the fluorescence enhancements were discussed.[101,678] The fluorescence characteristics of the gallium(III) chelate formed with lumogallion, hexyl-lumogallion, and dodecyl-lumogallion in the presence of surfactant micellar systems have been reported.[698,699] A 10-fold increase in the quantum efficiency of the Ga(III)–hexyl-lumogallion chelate was observed in the presence of the nonionic surfactant polyethylene glycol monolauryl ether.[698] Gallium was easily determined in the parts per billion (ppb) range using this micellar-modified fluorimetric procedure.

A Brij-35-enhanced FIA/fluorimetric method for the determination of aluminum(III) based on the formation of the Al–lumogallion complex has been developed.[453] The detection limit was 0.15 nM, making it one of the most sensitive assays for Al(III). The method is very sensitive, easy to use, and rapid (cycle time 3.0 min).

Triton X-100 micellar medium was utilized as the extraction medium (CP extraction procedure) to extract the Al(III)–lumogallion complex from water (extraction efficiency 72%).[700] The surfactant-rich medium also resulted in a 2.3-fold increase in the FI compared with that observed in bulk aqueous solution. The detection limit attained was 0.05 ppb Al(III).

The sensitivity of the spectrofluorimetric method for Ga(III) with 1-(2-pyridylazo)-2-naphthol (PAN) in the presence of SDS micelles is ca. 20 times better than

that obtained in a 20:80 (v/v) ethanol–water solvent system.[106]

The fluorescence reactions of Ti(IV) and Ta(V) ions with pyrogallol red, bromopyrogallol red, Alizarin red S, morin, and 8-quinolinol derivatives have been evaluated as the basis for the fluorimetric determination of these metal ions in the presence of surfactant micellar media.[701]

The effects of β-CD and hexadecylpyridinium bromide, CTAB, and SDS surfactant micellar media upon the fluorimetric properties of the metal chelate complexes formed between 2-(2-hydroxyphenyl)benzoxazole and Al(III), Be(II), Ga(III), and Zn(II) have been reported.[702]

The use of flavones (flavonol, chrysin, fisetin, quercetin, morin, etc.) as fluorescent labeling reagents for the postcolumn detection of organotin compounds has been reported.[608,703,704] The FI observed after reaction of such flavones with organotin chlorides, i.e. triphenyltin, diphenyltin, monophenyltin, tributyltin, dibutyltin, trimethyltin, and tricyclohexyltin chloride, in the presence of micellar Triton X-100 was determined. The results indicated that morin and fisetin are the most general labeling agents for organotin species, with morin being the best reagent for mono-organotin compounds and fisetin the most sensitive for triphenyltin.[608,704] The fluorescence of the flavonol complexes of mono- and diorganotin compounds was found to be greater in hexane as solvent than in the Triton X-100 system. However, the micellar system proved superior to hexane as a medium for the triorganotin complexes.[608] The system has been employed for the fluorimetric detection of organotin compounds following their HPLC separation.[703]

The fluorimetric method for determination of selenium based upon its reduction to selenite and subsequent reaction with 2,3-diaminonaphthalene to form the fluorescent 4,5-benzopiaselenol complex is enhanced in the presence of organized assemblies. For instance, the detection limit for Se in the presence of a CTAB micellar medium is 23.7 times lower than that observed in aqueous solution.[705] Another report indicated that the FI of 4,5-benzopiaselenol was enhanced by a factor of 5–6 in the presence of either SDS micelles or β-CD.[706] This enhancement factor increased to ca. 30 when the fluorescence was measured in a mixed SDS micelle–β-CD system relative to that observed in bulk water.

A few studies concerning the formation and fluorescent properties of metal chelates have been conducted in vesicle or liposome systems. In one study, the features of the Al^{3+}–8-quinolinol-5-sulfonic acid chelate in the presence of cationic DODAB vesicles were evaluated.[147] The fluorescence enhancement observed was similar to that seen in the presence of cationic CTAB micelles. However, in the case of the Al^{3+}–ferron chelate, the fluorescence signal observed in the presence of DODAB vesicles was greater than that in CTAB, Brij-35, or zwitterionic surfactant micellar media.[151] No fluorescence enhancements above that seen in aqueous solution were noted in the presence of CDs or anionic SDS micelles for this chelate. The proposed vesicle-enhanced method for Al(III) was validated by determining the aluminum content of different infant milk powders, with a detection limit of $4.0\,ng\,mL^{-1}$.[151] These results demonstrated that in some situations, a vesicle medium is superior to that of micelles or CDs and that vesicle aggregate systems can be utilized as a medium for developing more sensitive analytical methods for the determination of certain metal ions.

In addition to the determination of metal ion concentrations, different organic chelating agents have been employed for the determination of the concentration of cationic detergents, surfactants, or surface-active agents.[707,708] For instance, the use of lumogallion has been reported for the fluorimetric determination of the cationic surfactants CPB and ethonium, with detection limits of 0.027 and $0.024\,\mu g\,mL^{-1}$, respectively.[707] The method is rapid and simple and does not require the use of an extraction step to remove the excess reagent as required in other spectral methods.

Several reports have considered the general mechanism and primary factors that influence the enhancement of the fluorescence of metal chelates by surfactant micelles.[503,678,679,709,710] Solubilization and binding equilibria for reagents and metal chelates between the micelles and the aqueous phase seemed to be the most important factor governing the fluorescence enhancements observed for a number of metal ion–ligand chelates.[503,679] The presence of the micellar surfactant can also alter the coordination number of the metal ion complex,[711] which in some instances was postulated to be the reason for enhanced luminescence. There is a continuing need for comprehensive, mechanistic studies in this area. Perhaps studies reported in the physical chemistry literature could aid in this endeavor. For example, Demas and DeGraff reported on the binding interactions of a series of luminescent α-diimine complexes of ruthenium(II) and osmium(II) with SDS, Triton, and Brij micelles as studied by luminescence spectroscopy and quenching techniques.[712–714] The strength of binding and degree of solvent shielding of the complexes by the micelles were found to correlate systematically with the charge and polarity of the complexes and the hydrophobicity of the coordinated ligands. The factors responsible for the enhancement of luminescence efficiency were also identified.

Micellar (or CD) enhancement of fluorescence quenching has been employed to determine some environmentally important metal ions. For example, in the

presence of the anionic micelle, NaLS, the fluorescence of the probe molecule naphthalene was efficiently quenched by metal ions thus allowing their quantification via Stern–Volmer calibration plots (with detection limits in the range of 1–10 μM).[715] Likewise, quenching of 2,6-bis(benzoxazolyl)pyridine in a β-CD medium allowed the selective determination of the iron(III) ion.[716] Anions can also be determined via micellar-amplified quenching. For example, as shown in Figure 8, the slope of the Stern–Volmer plot for the quenching of *N*-methylacridone by iodide ion in the presence of a cationic DTAC micelle medium is much greater (factor of ca. 11) than that observed in bulk aqueous solution alone, which translates into a lower limit of detection for the iodide ion.[717]

A Triton X-100 micelle system containing the fluorescene pyrene molecule and the chelating agent dodecyldioxo-2,3,2 has been proposed as a sensor for transition-metal ions.[718] The sensor action is based on the transition metal ion's ability to quench the pyrene probe fluorescence.

3.2.4 *Lanthanide- (or Ligand-) Sensitized Fluorescence in Organized Assemblies*

The sensitization of the fluorescence of lanthanide and actinide ions by organic ligands is well known and has been extensively utilized during the past two decades in various analytical applications, including immunoassays, quantification of organic compounds, chromatographic detection schemes, etc.[719–723] In particular, europium, terbium, samarium, and curium are known to form highly fluorescent chelates with a variety of organic ligands, often in the presence of a so-called synergistic agent. The intense sensitized fluorescence results from excitation of the organic ligand donor molecule, the energy of which is in part transferred (intramolecular energy transfer step) from its excited triplet (or singlet) state to the emitting level of the chelated lanthanide (or actinide) ion acceptor, resulting in ion-specific emission, and is characterized

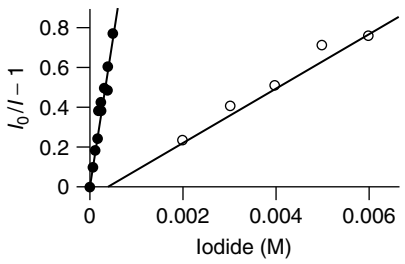

Figure 8 Stern–Volmer plots for quenching of *N*-methylacridon's fluorescence by iodide ion in aqueous solution (open circles) and in an aqueous solution containing DTAC surfactant micelles (dark circles). Conditions: [*N*-methylacridone] = 5.0×10^{-6} M; $\lambda_{ex} = 384$ nm; $\lambda_{em} = 422$ nm; [DTAC] = 0.030 M.

by large Stokes shifts, narrow (line-type) fluorescence bands, and long lifetimes. These properties allow spectral discrimination of the analytical signal and/or temporal time discrimination that eliminates (or greatly diminishes) background fluorescence and scattering signals and allows time-resolved signal accumulation such that very sensitive and selective fluorimetric methods can be developed. For a more detailed, quantitative discussion of this topic, the interested reader is referred to published review articles.[719,720]

In some cases, an intermolecular energy transfer from the triplet level of an organic sensitizer directly to the unbound lanthanide ion has been reported.[604,724] Such an intermolecular energy transfer process in a homogeneous solution is typically inefficient. However, intermolecular energy transfer from organic species to lanthanide ions in aqueous micellar solution has been observed.[275,604,725] For instance, efficient energy transfer has been observed from SDS micellar solubilized naphthalene to terbium(III).[275] Likewise, very efficient energy transfer from the 2-acetylnaphthalene triplet to Eu^{3+} was observed in SDS micellar solution.[371] Such intermolecular energy transfer may be observed only by the use of organized media to bring the donor and acceptor molecules into close proximity and substantially increase their effective local concentrations.[275,567,604,726] Since micelles solubilize the ternary complexes and can increase the efficiency of intermolecular energy transfer and also favorably impact luminescence parameters as previously described, they have been employed as enhancers in lanthanide-sensitized fluorescence assays. In particular, lanthanide ions exhibit low emission quantum yields in water. However, when complexed with appropriate ligands and solubilized in micellar media, intense, sharp emission bands are observed since the added agents help to "remove" water from the lanthanide ion's primary solvation shell, thus preventing quenching of the excited state by water molecules.[720] In addition, the fluorescence lifetimes of the lanthanide complexes are greater.[727,728] For example, the lifetime of the europium chelate of thenoyltrifluoroacetone (TTA) in ethanol was 180 μs, whereas it is 780 μs in the presence of Triton X-100 micelles and trioctylphosphine oxide (TOPO).

A comparative study of the fluorescence efficiency in the Eu^{3+}–TTA–TOPO chelate in ethanolic and aqueous Triton X-100 media has been conducted.[728] It should be noted that large enhancements in the yield of intramolecular charge-transfer emission from organic molecules have also been observed in CD media.[550] A comparative investigation on the effect of cationic micelles upon the fluorescence observed from the Eu^{3+} (and Tb^{3+}) complexes with Tiron, salicylic acid, acetylacetone, and 4-sulfonyl salicylic acid has been reported.[729] Such lanthanide ion–organic

ligand–synergistic agent–micellar systems have been employed to determine the concentration of either the lanthanide–actinide ions or organic species that can act as ligands, as well as for time-resolved fluoroimmunoassays of a number of biological/clinical compounds.[719,730,731] In general, for determination of lanthanide or actinide ions, a sensitizer organic ligand that can form a tris-chelate to the ion, such as a diketone (e.g. TTA), is employed along with a strong donor ligand (synergistic agent), such as TOPO (or other organic phosphates, e.g. tributyl phosphate or sulfoxides), which can compete with water for inner coordination sites, and surfactant micelle.[732] This approach for the determination of lanthanides or actinides is termed *ligand-sensitized fluorescence*. The micelle system typically serves to solubilize the syngeristic agents (as TOPO) and the ternary metal complexes in aqueous solution and enhances the fluorescence.[733–740] In addition, it has been postulated that the presence of the micellar medium alters the enol–ketone equilibrium ratio (ratio of ca. 0.016 in water and 20.0 in benzene) of the diketone (TTA) reagent so that the chelating properties of the enol form, which favors formation of the lanthanide–actinide chelate, predominates.[741] β-CD can also shift such keto–enol equilibria toward the enol tautomeric form[252]; however, its use in such lanthanide-sensitized systems has apparently not yet been reported. The partition behavior of some Eu^{3+}–diketone–TOPO adducts between nonionic surfactant micellar and bulk aqueous phases has been investigated.[742] The use of time-resolved laser-induced spectrofluorimetry has been employed for the study of uranium–surfactant micelle[743] and Eu^{3+}–phospholipid vesicle[744] interactions. The fluorescence properties of terbium(III)–phospholipid complexes have been reported.[745] A thermodynamic study of the inclusion complexes formed between trivalent lanthanide ions and native CDs has been conducted.[277] The luminescent properties of the tris(dibenzoylmethane)europium(III) dihydrate–β-CD complex in the solid state have been reported.[746] It might prove interesting to examine the luminescent behavior of lanthanide ions and their diketone complexes in the presence of crown ether–based surfactant micelles. For example, the binding of Gd^{3+} to 1,4,7,10-tetraazacyclododecane-1-[1′-carboxy-1′-dodecyl(methyl)aminooxoethyl]-4,7,10-triacetic acid (DOTA) (CMC = 0.35 mM) has been examined in the context of its ability to function as a magnetic resonance imaging (MRI) contrast agent.[747] This system might prove to be interesting from the analytical luminescent viewpoint as well.

As for examples of the general utilization of this sensitized fluorescence technique, the determination of europium, terbium, and samarium ions was reported via use of TTA (or pivaloyltrifluoroacetone (PTA)) and TOPO in aqueous solution containing the nonionic surfactant nonaoxyethylene dodecyl ether ($C_{12}E_9$).[748–752] The stoichiometry and complex equilibrium involved with this system have been determined.[733] Europium(III) ($\lambda_{ex} = 355$ nm, $\lambda_{em} = 615$ nm) was determined in the 10^{-11}–10^{-5} M range with a detection limit of 6.0×10^{-12} M via formation of the $Eu(TTA)_3(TOPO)_2$ complex in an aqueous Triton X-100 micellar medium.[734] For this system, the effectiveness of different nonionic surfactant micelles with respect to increasing the FI observed from this complex (or that of the corresponding Tb^{3+} complex) followed the order Triton X-100 > octylphenol ethoxylate (Triton) surfactants (OP) > Tween 80 systems.[740] Instead of the use of TTA or TOPO as the organic chelating and synergistic agents, a number of other organic ligands/agents have been proposed and evaluated.[732,753] Examples of such systems include TTA–phen–Triton X-100 (or Polysorbate 20),[754–756] 2-naphthoyltrifluoroacetone (NTFA)–TOPO–Triton X-100,[757] sodium benzoate–TOPO–Triton X-100,[735] 2-benzoylindan-1,3-dione (BID)–TOPO–Triton X-100,[736] TTA–diphenylguanidine–Triton X-100,[110] PTA–2,2′-bipyridine (BP)–Triton X-100,[758] hexafluoroacetylacetone (HFAC)–TOPO–Triton X-100[759] trimesic acid (benzene-1,3,5-tricarboxylic acid)–TOPO–Triton X-100,[760] and benzoic acid derivatives–TOPO (or EDTA)–CTAC[761] among others. The fluorescence observed from Tb, Dy, or Eu using benzoate (or trimesic acid) as the chelating ligand in the presence of TOPO and Triton X-100 micelles was enhanced by almost 4 (or 3) orders of magnitude compared to the use of diketone ligands.[735,760] With the lanthanide–benzoate–TOPO–Triton X-100 system, the detection limits for lanthanides (Dy, Tb, Eu) were in the 10^{-9}–10^{-10} M range.[735] A detection limit of 4.0×10^{-9} M Eu^{3+} was reported for the BID–TOPO–Triton X-100 system.[736] In addition to use of Triton X-100, the nonionic surfactant Tween 20 has also been employed as the micellar entity. A detection limit of 50 ppt Dy^{3+} was reported for the TTA–TOPO–Tween 20 system.[762] Dysprosium(III) ion was determined at concentrations as low as 1.0×10^{-9} M using the EDTA–Tiron–CTAB system.[763] The fluorescence signal observed in this system was ca. 12.3 times greater in the CTAB (or cetyldimethylbenzylammonium chloride) micellar systems than that observed for Triton X-100.[763] A detection limit of 20 pM Tb^{3+} was reported via use of an unoptimized 1,1,1-trifluoro-2,4-tridecanedione (TFTD) or 1,1,1-trifluoro-5,5-dimethyl-2,4-hexanedione (TFDMH)–TOPO–Triton X-100 system.[732] Terbium (III) has also been determined via use of the PTA–TOPO–$C_{12}E_9$ system in the flow injection mode with a detection limit of 16 ng mL^{-1}.[764] Tb^{3+} was also determined down to the 0.003 ng mL^{-1} level

using the system Tiron–imidodiacetic acid–CTAB.[739] Determination of curium at the $0.10\,ng\,mL^{-1}$ level in a micellar medium by this approach has been reported.[741] Thulium(III) and yttrium(III) were determined down to 1.0×10^{-7} M and $1.8\,ng\,mL^{-1}$, respectively, using the system bis(1′-phenyl-3′-methyl-5′-pyrazol-4′-one)hexanedione–CTAB.[581,765] The presence of the CTAB enhanced the FI of these Th^{3+} and Y^{3+} chelates by a factor of ca. 25 and 8, respectively. Curium(III) has been determined at the $0.10\,ng\,L^{-1}$ (5×10^{-13} M) concentration level using the TTA–TOPO–Triton X-100 system.[629] The fluorescence observed from the ytterbium(III)–4-(2-pyridylazo)resorcinol–CTAB (or CPB or chloride) system was 2.8–4.5 times greater than in the absence of the cationic CTAB or cetylpyridinium salt surfactant media.[749] This enhancement was attributed to the greater rigidity of the micellar bound complex.

Hence these assays are superior to the traditional solvent extraction methods that have been employed for determination of these lanthanide/actinide ions, in terms of not only their simplicity and speed, but also their sensitivity.[748]

It has been reported that the presence of certain other ions, such as La, Gd, Lu, or Y, can lead to additional enhancements of the fluorescence observed from the chelates of Tb, Eu, and Dy. This is referred to as the *columinescence* (or *cofluorescence*) *effect* (Xu et al.[766] have reviewed this topic). This effect is based on an intermolecular energy transfer from the chelates of the enhancing ion (such as Gd^{3+} or Tb^{3+}) to the chelates of the emitting ion (such as Eu^{3+} or Sm^{3+}).[766] Fluorescence enhancement of rare earths due to columinescence has been reported using β-diketone ligands as the energy transfer agent along with TOPO or BP as synergistic ligands and Triton X-100 micelles.[758] Detection limits for Eu^{3+}, Tb^{3+}, and Sm^{3+} were reported as 0.019, 0.27, and 3.8 pM, respectively for the PTA–BP–Triton X-100 system containing added yttrium(III) cation.[758] These detection limits were slightly better than those reported for the PTA–phen–Triton X-100 system with added Y^{3+} for these three metal ions.[767] The addition of gadolinium(III) to the system containing samarium or europium(III) ions with TTA–phen–Triton X-100 resulted in 1–2 orders of magnitude better sensitivity for the determination of Sm^{3+} or Eu^{3+} compared to that observed in its absence.[768] The specific detection limits reported were $(5.0–7.5) \times 10^{-10}$ M for samarium and 7.5×10^{-12} M for europium.[769,770] Substitution of 4,7-diphenyl-1,10-phenanthroline for phen in this general system led to a detection level of $0.6\,pg\,mL^{-1}$ for Eu^{3+}.[771] A detection limit of 10 pM for Eu^{3+} was reported for the system TTA–phen–Triton X-100 in the presence of Lu(III).[772] Use of the TTA–TOPO–CTAB system with added gadolinium(III) allowed the determination of europium(III) with a detection limit of 8.0×10^{-11} M.[737] A slightly more sensitive method (detection limit 2.6×10^{-11} M) for europium(III) was achieved using the system dibenzoylmethane (DBM)–CPC in the presence of Gd^{3+} and triethanolamine.[773] A detection limit of 3.0×10^{-12} M Eu^{3+} was reported for the diphacinone [2-(diphenylacetyl)indan-1,3-dione]–ammonia–Triton X-100 system containing La^{3+} ion.[616] Using the TTA–CTAB–Triton X-100 system in conjunction with Gd^{3+} also allowed the determination of Eu(III) with a reported detection limit of 2.0×10^{-12} M.[738] Detection limits of 4–15 fM were reported for the TTA (or benzoyltrifluoroacetone (BTA))–phen–Triton X-100 system containing added yttrium(III) ion.[774] In the determination of europium(III) via use of the TTA–phen–surfactant system containing terbium(III), the relative luminescence intensity as a function of the specific surfactant micelle system followed the order Triton X-100 > Tween 20 > polyoxyethylene lauryl ether > SDDBS > CTAB > CPC.[754] Under optimized conditions using Triton X-100 micellar media, a detection limit of 1.0×10^{-13} M Eu(III) was reported.[754] The use of mixed ligand complexes of TTA and 1,10-diaza-15-crown-5 in the presence of CTAB and Gd^{3+} (or other lanthanides) also allowed the determination of ppb levels of Eu(III).[775] The reagents Y^{3+}–TTA–CTAB and Y^{3+}–sparfloxacin–SDS were utilized to determine Eu^{3+}, with a detection limit of 1.0×10^{-11} M[776] and 1.0×10^{-13} M,[777] respectively. A similar detection limit for Eu(III) was reported for the reagent system composed of Gd^{3+}–ofloxacin–sodium dodecylbenzenesulfate micelles.[730] The presence of the Triton X-100 or CTAB micelles in these reagent combinations was thought to enhance the interchelate energy transfer, protect the chelates against the quenching effects of water molecules, and stabilize the fluorescence in such systems.[737,754,758,766,768] Cationic surfactants can function as a cation in the ternary complex that forms in some instances.[738,778] Typically, there is an optimum surfactant concentration beyond which further increases in concentration lead to decreased fluorescence. This is the result of dilution of the system with increasing micelle concentration so that the probability that both donor chelate species and acceptor fluorescent species are bound to the *same* micelle is decreased.[766] Separation of the donor and acceptor in different micelles obviously decreases the efficiency of the intermolecular energy transfer step with a resultant diminution in the FI.

To the best of our knowledge, there are no reports concerning the examination of such systems for the determination of lanthanides/actinides in the presence of CD or vesicle organized assemblies. However, enhanced fluorescence was observed for Eu^{3+}–TTA–phen in the

presence of Tb^{3+} (or Gd^{3+})–TTA–phen when confined in Langmuir– Blodgett films.[779] In addition, there is one report concerning the synthesis of polymerized fluorescent liposomes that incorporate lanthanide ions.[780]

Energy transfer from an organic compound to the lanthanide ion can also be utilized to improve the fluorimetric determination of organic analytes that can chelate the metal ion. This approach for assaying organic components is referred to as *lanthanide-sensitized luminescence* (for a review see Georges[719]). A number of reports indicate that the presence of micelles can further intensify the luminescence intensity of such systems and improve the sensitivity. Methyl anthranilate in food samples, after hydrolysis to *o*-aminobenzoic acid, was determined by the sensitized luminescence obtained from the Tb^{3+}–TOPO–aminobenzoic complex in the presence of TX-100. This approach was applied with success for analysis of methyl anthranilate and adenosine triphosphate (ATP) in a variety of food and drink samples.[781,782] The antibiotics ciprofloxacin, enrofloxacin, and levofloxacin have also been determined via terbium-sensitized luminescence in the presence of SDS micelles[783,784] or sensitized fluorescence of europium in the presence of 1,10-phenanthroline in SDDBS micelles.[752] Tetracyclines (such as doxycycline, oxytetracycline, demeclocycline, methacycline, and chlortetracycline) have been determined using europium-sensitized fluorescence with EDTA as a coligand in the presence of CTAC micelles.[731,785]

Table 9 summarizes other representative applications of the lanthanide-sensitized luminescence for the determination of organic compounds in the presence of

Table 9 Summary of micellar-enhanced lanthanide-sensitized determination of organic species

Analyte species	Lanthanide ion	Micelle system/conditions	Comments	References
Tetracyclines (9)	Eu(III)	Triton X-100 or CPC, pH 8.0, $\lambda_{ex} = 394$ nm, $\lambda_{em} = 618$ nm, $D_L \approx 2 \times 10^{-7}$ M	Micelles enhanced luminescence intensity by factor of up to 34	617
Tetracycline	Eu(III)	Triton X-100, pH 7–8.5, $\lambda_{ex} = 390$ nm, $\lambda_{em} = 613$ nm, $D_L = 3.0 \times 10^{-9}$ M	TX-100 enhanced FI by factor of 5 (at pH 9) to 12.5 (at pH 7)	719
Diphacinone	Eu(III)	Triton X-100, pH 8.5, $\lambda_{ex} = 336$ nm, $\lambda_{em} = 619$ nm, LDR = 0.5–5.0 μg mL^{-1} D_L = 0.20 μg mL^{-1}	No fluorescence observed in absence of surfactant	780
Pindone	Eu(III)	CTAB, Triton X-100, TTA, pH 6.4, $\lambda_{ex} = 273$ nm, $\lambda_{em} = 614$ nm, LDR = 0.1–10.0 μg mL^{-1}, $D_L = 0.04$ μg mL^{-1}	Indirect kinetic method, CTAB enhanced intensity of fluorescent reagent complex	778
Benzoic acid (BA) and saccharin (S)	Tb(III)	Triton X-100, TOPO, pH 6.0, $\lambda_{ex} = 284$ nm, $\lambda_{em} = 545$ nm, $D_{L(BA)} = 0.04$ μg mL^{-1}, $D_{L(S)} = 1.1$ μg mL^{-1}	Triton X-100 noticeably enhanced the luminescence signal, kinetic method	602
Salicylic, fluorosalicylic and aminosalicylic acids	Eu(III)	CTAC, EDTA, pH 12, $\lambda_{ex(SA)} = 330$ nm, $\lambda_{ex(FSA)} = 335$ nm, $\lambda_{ex(ASA)} = 310$ nm, $\lambda_{em} = 545$ nm, $D_L = (2–4) \times 10^{-10}$ M	CTAC micelles enhanced FI 10-fold	618
Propyl gallate	Tb(III)	SDS, Triton X-100, pH 6.5, $\lambda_{ex} = 310$ nm, $\lambda_{em} = 545$ nm, LDR = 0.08–3.0 μg mL^{-1}, $D_L = 0.02$ μg mL^{-1}	Triton X-100 solubilized the Tb^{3+}–SDS chelate and enhanced the initial rate of this kinetic method	112
Orotic acid	Tb(III)	CTAB, pH 6.0, $\lambda_{ex} = 325$ nm, $\lambda_{em} = 544$ nm, LDR = 10^{-7}–5×10^{-5} M, $D_L = 5 \times 10^{-8}$ M	CTAB enhanced FI and lowered the second pK_a of analyte, analyte detected following MEKC[a] with CTAB	619
Steroids (6)	Tb(III)	SDS, $\lambda_{ex} = 247$ nm, $\lambda_{em} = 547$ nm, $D_L = 100$ pg mL^{-1} (for bolasterone) to 3000 pg mL^{-1} (for cortisone)	SDS enhanced FI of steroidal–Tb(III) chelates, analytes detected following MLC[b] with SDS	790
	Tb(III)	SDS, $\lambda_{ex} = 238$ nm, $\lambda_{em} = 520$ nm, pH 7.0	Demonstrated use of SDS in run buffer for enhanced energy transfer and CZE separation	619

[a] Also referred to as MECC, *micellar enhanced capillary chromatography*.
[b] Micellar liquid chromatography, which employed aqueous solutions of surfactant micelles as the mobile phase.

surfactant micellar media.[112,602,617–619,719,778–791] Also, the addition of lanthanide ions, such as Tb^{3+}, to PAHs solubilized in NaTC bile salt micelles generally resulted in enhanced fluorescence.[121,786] Detection limits for the determination of the PAHs anthracene, perylene, benzo[*k*]fluoranthene, pyrene, and benzo[*ghi*]perylene were in the range 0.021–0.70 nM in the Tb^{3+}–NaTC system.[121] As noted by Georges, the application of this general method is not limited to the examples just mentioned or shown in Table 9, but is also potentially valid for many other organic species, provided that they exhibit large molar absorptivities and have the ability to function as energy donors with respect to a lanthanide ion.[719,787]

More recently, quenching of lanthanide complexes in micelles has been utilized to determine the quencher concentration. Addition of ATP to the Eu(III)/Gd(III) complex with TTA in nonionic Brij-35 micelles diminished the observed luminescence fom this complex and allowed the determination of ATP (detection limit 8.0×10^{-8} M) in fruit juice samples.[782]

Amino- and mercapto-containing compounds have been detected following HPLC separation using lanthanide-sensitized fluorescence.[111] In this detection scheme, a precolumn derivatization method was employed in which the amino compounds were reacted with the Eu^{3+} chelate of *N*-(*p*-isothiocyanatobenzyl) diethylenetriaminetetraacetic acid (IDTTA) (or Eu^{3+} chelate of 1-[*p*-([(5-maleimidopen-tyl)carbonyl]amino} benzyl)ethylenediaminetetraacetic acid (MEDTA) in the case of thiols) and resolved via reversed-phase HPLC. Next, a postcolumn derivatization step is conducted in which the labeled Eu–IDTTA–amino compounds (or Eu–MEDTA–thiols) are reacted with the reagent NTFA–TOPO–Triton X-100, which results in formation of the highly fluorescent Eu^{3+}–NTFA–TOPO chelate. For thiols, the detection limits reported ranged from 13.2 to 18.8 fmol.[111] The presence of Triton X-100 micelles was necessary for observation of enhanced fluorescence. Terbium-sensitized fluorescence was utilized as a detection mode in HPLC for 14 different steroidal hormones which were eluted with an SDS-containing mobile phase (detection limits ranged from 0.08 to 0.55 $\mu g\,mL^{-1}$).[788] Use of Eu^{3+}–TOPO–Triton X-100 (pH 10.5) as a postcolumn derivatization reagent allowed the sensitive time-resolved sensitized luminescence detection of tetracyclines following their HPLC separation.[789] Use of Tb^{3+}–CTAB allowed the sensitive fluorimetric determination of 2,6-dihydroxypyrimidine-4-carboxylic acid following capillary electrophoretic separation using CTAB in the run buffer.[619] This method required no sample pretreatment and was free of any known interference in the urine samples examined. Micellar HPLC with SDS mobile phases in conjunction with sensitized terbium fluorescence detection allowed the development of a superior analytical method for semiquinoid steroids.[790] The method was sensitive, selective and simple, and allowed the direct injection of urine on to the column without the need for any sample preparation step.

Numerous reports concerning the dissociated enhanced lanthanide fluoroimmunoassay (DELFIA) technique for determination of clinical/biological analytes have appeared in the clinical/medical literature (reviews are available).[791–796] The DELFIA technique was based on the work of Hemmila et al. DELFIA involves the use of an immunocomponent (antibody, etc.) labeled with a lanthanide metal (i.e. Eu^{3+} or Sm^{3+}) chelate which is nonfluorescent. Typically, the binding of the lanthanide(III) ion to an immunoreactive component is accomplished via use of polycarboxylic derivatives (EDTA, glycol-bis(2-aminoethylether)-*N,N,N′,N′*-tetraacetic acid (EGTA), diethylenetriaminepentaacetic acid (DTPA)) that contain a suitable functional group required for covalent coupling. Such polycarboxylates have very large lanthanide ion–binding constants, but are strongly pH dependent. The immunoassay can be conducted as either a noncompetitive or competitive reaction in a solid-phase system. The amount of immunoconjugate is determined by dissociation of the lanthanide ion from the immunocomponent on the solid phase. This is accomplished by lowering the pH (pH < 4.0, which lowers the stability of the chelate complex) with the dissociated lanthanide ion going into an aqueous solution that contains the ''enhancement agents'' (i.e. β-diketone or other sensitizer, synergistic agents such as TOPO, and micelle-forming surfactant). The enhancement agents greatly intensify the fluorescence as previously noted, and in DELFIA, time-resolved intensity measurements are made to eliminate noise and increase sensitivity.[720,791] In such systems, the label may be detected in concentrations as low as 10^{-17} mol of Eu^{3+} per microtitration well.[797] The general system has been successfully applied for the assay of the following analytes in human serum: cortisol, thyroxine, α-fetoprotein, ferritin, thyrotropin, prolactin, lutropin, choriogonadotropin, follitropin, prolactin, chorionic gonadotropin, hepatitis B surface antigen, luteinizing hormone, somatotropin, digoxin, thyroxine-binding globulin, thyroid-stimulating hormone, follicle-stimulating hormone, immunoreactive trypsin, creatine kinase, 17-α-hydroxyprogesterone, progesterone, carcinoembryonic antigen, and rubella IgG and IgM antibodies,[690,720,758,774,794,791,798–801] among many others. Sensitive cytotoxicity assays using DELFIA with Eu–DTPA have been developed.[801] The utilization of chelated terbium as a label in fluorescence immunoassays has also been reported.[802,803] The cofluorescence enhancement systems are also suited for use in double (or multi) label time-resolved fluorimetric

immunoassays.[720,758,766,767,774] A review of the cofluorescence effect as it pertains to time-resolved fluoroimmunoassays is available.[766]

More recently, liposomes have been employed in steady-state and time-resolved fluoroimmunoassays for clinical/biological analytes.[804–808] Liposomes can incorporate fluorescent labels (including fluorescent lanthanide chelates), either encapsulated within their inner aqueous region (polar species) or embedded in the lipid membrane (nonpolar or amphiphilic species). In an assay dubbed ''complement immunolysis'', recognition of an analyte by an antibody (or other species) triggers the release of thousands of reporter molecules which are subsequently detected by fluorescence (or other means).[804] The large number of molecules released thus provides substantial amplification of the original antibody recognition event.[805] Such encapsulated contents can also be released upon lysis of the liposomes and the assays developed in the usual manner as in DELFIA protocols. Haptens have been determined via this approach.[808]

4 SOME GENERAL EXPERIMENTAL CONSIDERATIONS

The chemicals required for the preparation of the organized assembly systems are all commercially available. Cognizance should be taken of the purity of the surfactant or CD preparations utilized to prepare the solutions required for analytical luminescence applications. Although highly purified CDs do not fluoresce, some less purified formulations sometimes contain fluorescence impurities due to the preparation scheme employed in their purification. This is probably the cause of some reports in the analytical literature which note that CDs are strongly fluorescent.[107,108] or quench fluorescent solutes.[809] In addition, the presence of fluorescent impurities in commercial nonionic and some cationic surfactant preparations has been noted.[455,686] Weak fluorescence from NaDC solutions has also been reported when a He–Cd laser was employed as the excitation source.[646] Quenching impurities in other commercial preparations have also been reported.[178] In particular, peroxides can be present in nonionic surfactant preparations.[660] The presence of quenching impurities in the preparations can have severe detrimental effects upon the desired fluorescence measurement. Purification procedures for various surfactants and CDs are available in the literature.[195,660,810] In addition, high-purity commercial preparations are now available for many of the surfactants or CDs required in fluorimetric applications. However, there are still some recent reports that note that while CDs and surfactant micelles elicited strong enhancing effects on analyte fluorescence emission, the relative standard deviations were quite high ''likely the result of intense signals given by the [organized assembly] blanks'' which precluded their use in the desired fluorescent method.[811]

Also, for most applications, care should be taken to avoid the use of organized assembly materials that are themselves inherently fluorescent. For example, surfactants containing an aromatic ring, such as the Tritons, Igepals, Nonidets, alkylpyridinium salts, alkylbenzenesulfonates (e.g. SDDBS), and benzyldimethylhexadecylammonium halides, should probably be avoided since their use could give rise to high background fluorescence at the desired analytical emission wavelengths.[107,453,812] In addition, the pyridinium surfactants can act as potent quenchers toward many organic analytes.[107,535] The reduced forms (nonaromatic ring system) of some of these materials, such as reduced Triton X-100, are commercially available, but are more expensive.

Some surfactant micelles, such as CTAB, SDS, and Tergitol XD, present solution preparation difficulties.[220,453] That is, the preparation of the more concentrated stock solutions of such surfactants requires warming/heating the solution in order to ensure complete dissolution of the solid surfactant material.[220,404,628,679,813] Surfactants such as CTAB, TDSS, and sodium hexadecylsulfate have Krafft points that are well above room temperature. For instance, the Krafft point of SDS is ca. 16–23 °C.[180,220] Thus, when using such surfactant micelle systems, the luminescence measurements have to be conducted at an appropriate temperature (i.e. experimental temperature above the Krafft temperature) in order to prevent precipitation of the surfactant. Alternatively, it is often more convenient to utilize other micelle-forming surfactants to produce the same desired effects. For instance, CTAB has a Krafft temperature of ca. 17–20 °C, whereas that of CTAC is ca. 11 °C.[180,618] Hence CTAC is preferred to CTAB (which also has the bromide ion which can quench excited states of some organic species).

It should be mentioned that heating the aqueous solutions of some nonionic surfactant solutions can lead to phase separation (CP) behavior, where the surfactant-rich phase separates from the bulk aqueous (surfactant-depleted) phase.[659,660] The CP temperatures for typical nonionic surfactant solutions have been compiled.[659] In order to avoid complications from such phase separation, as has been reported in the literature,[814] one should utilize such surfactant micelle solutions at a temperature well below their usual CP temperature. Materials that are toxic, such as the surfactant myristylammonium bromide, should be avoided.[453] Surfactants containing heavy-atom counterions (such as Br or I) or other moieties (such as pyridinium ion) could quench certain organic molecules' excited states. Hence this fact should

be taken into account for surfactants such as CTAB, tetradecyltrimethylammonium bromide (TDAB), and alkylpyridinium salts.[249,588,813]

Whereas the solubilities of α- and γ-CD are ca. 14.5 and 23.2 g per 100 mL in water, respectively, that of β-CD is only 1.85 g per 100 mL.[161] Thus, when employing β-CD in fluorescence measurements, it is sometimes not possible to achieve the proper (high) concentrations required to produce maximum benefits.[380] In such cases, the use of β-CD derivatives that are much more soluble in water, such as hydroxypropyl or methylated β-CDs, or aqueous urea solutions, which serve to solubilize the β-CD, is recommended.[380,815] Alternatively, water–cosolvent mixtures can be employed in order to increase the β-CD concentration.[816] When utilizing CD media in analytical derivatization reactions under basic conditions, cognizance should be made of the ability of CDs to catalyze hydrolysis reactions which might interfere with the desired analytical reaction.[817] Degradation of organized assemblies on storage can lead to unexpected behavior in use, and perhaps give rise to misleading interpretation of results.[94] In this regard, it should be recognized that some organized assembly media, such as the CDs and anionic alkylsulfate surfactants, are susceptible to acid-catalyzed hydrolysis, particularly when heated or subjected to long-term storage under very acidic conditions.[94,160,161,395,818] Hence stock solutions of such organized media should not be stored for a long time under very acidic conditions. Under very basic conditions, CDs can ionize (the pK of their hydroxyl groups is in the range 12.0–12.4).[160,277] Nonionic surfactants can form peroxides during storage. Some commercial formulations have added oxidizing agents present,[94] which might also interfere with fluorescence measurements.

Experimental protocols for the preparation of solutions required for the determination of fluorescence enhancement factors in the presence of an organized assembly system relative to that in a bulk solvent have been proposed.[206,819] Differences in some of the fluorescence parameters have been noted depending upon how the solute solutions are prepared (i.e. mixing via ultrasonication vs stirring of the solutions).[180] For fluorescence measurements in all organized assembly systems, appropriate background (reference) solutions should be prepared without the analyte molecule present in order to correct for possible contributions of light scattering to the observed analytical signal.[571,820] In some cases, the appropriate use of filters can minimize the effect of scattered light.[820] However, for reproducible results, it is probably best if the precautions that have been recommended in papers concerning basic investigations of such systems also be employed in their use in chemical analysis applications. For instance, vesicle solutions should be prepared and used within 24–48 h of their preparation, with storage in the dark under N_2 at 4 °C when not in use.[198,295,821] Care should be taken during all vesicle preparation steps to avoid undue exposure to light.[198,339,508,821] In many instances, proper "aging" (equilibration) of vesicle analyte–containing solutions is required in order to obtain reproducible spectral data.[148] Depending on the conditions and additives present, some liposome-containing systems can undergo time-dependent fusion with resultant formation of larger vesicle membranes (see Figure 9).[150,153,822,823] Deliberately keeping the amphiphile concentration as low as possible minimizes such fusion.[824] Fluorescence techniques are available to monitor the time course of vesicle fusion.[150,822,825] Thus, the best reproducibility is typically obtained in analytical applications by using vesicle/liposome systems of the same "age". Greater fluorescence enhancements are typically observed in vesicle/liposome media when the measurements are conducted at a temperature below the phase transition temperature (T_M) (also referred to as the *crystal-to-liquid crystal temperature*).[148] Likewise, some aging effects have been reported for micellar solutions.[190,826–828] Hence, in many studies, aqueous stock solutions were always prepared fresh daily or utilized within a specified time frame in order to avoid errors due to such aging and allow time for equilibration of reagents within the organized assembly system.[95,423,454,813,826]

In many instances, it is necessary to bubble nitrogen through the organized assembly solutions in order to displace oxygen so as to avoid its quenching effect upon some analytes[75] or the possibility of undesired photooxidation reactions.[385] For example, the FI of the analytes benzo[*a*]pyrene, chrysene, benzo[*ghi*]perylene and pyrene in oxygen-free solutions are 2.9, 2.7, 6.4, and

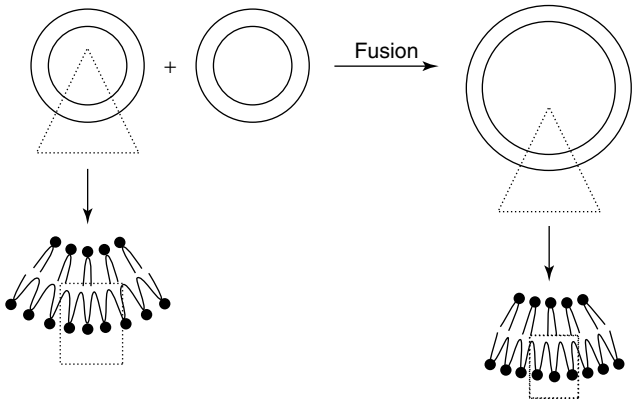

Figure 9 Oversimplified diagram (cartoon) illustrating the vesicle–vesicle fusion process which results in the formation of a larger aggregate entity with decreased membrane curvature and an increase in the hydrophobic to hydrophilic surface ratio. (Adapted with permission of the Royal Society of Chemistry from B.C.R. Guillaume, D. Yogev, J.H. Fendler, *J. Chem. Soc., Faraday Trans.*, **88**, 1281–1284 (1992).)

9.0 times greater, respectively, than those observed in oxygenated solution.[829] Oxygen is ca. 20 times more soluble in air-saturated alkane solvents than in water.[364] In fact, the solubility of oxygen in water is only ca. 1% (by mass) of that observed in hydrocarbon solvents.[713] The oxygen concentration at 25 °C in micellar CTAB, SDS, and Tweens was reported to be ca. 0.13, 0.12, and 0.13–0.23 mM, respectively,[830] whereas that in solutions of sodium octanoate ranged from 1.28 to 1.57 mol O_2 per atmosphere in 1000 g of H_2O.[831] Another report indicated the oxygen solubility in water as 1.41 mM, whereas that in SDS was 1.47 mM; in another, the water solubility of oxygen was reported as 0.265 mM and in SUV vesicles as 0.85 mM.[713] The solubility of atmospheric oxygen in water in the presence of several anionic surfactants at different temperatures and surfactant concentrations has been reported.[832] The solubility of oxygen in common solvents (such as water, alcohols or cyclohexane) at 25 °C and 1 atm has been compiled.[833]

One problem with the use of surfactant or micellar solutions arises when deoxygenation by purging with an inert gas is required, that is, the problem of foam formation and frothing, which makes such degassing problematic and cumbersome. Several approaches have been employed to overcome this difficulty. For instance, a simple gas bubbling system for removal of dissolved oxygen in surfactant and micellar solutions has been described.[834] This cell allows a continuous, rapid flow of a purge gas through such solutions without loss of fluid in spite of severe foaming. Using this cell, the oxygen concentration was reduced to a negligible value ($<10^{-5}$ M) in about 13 min.[834] A very similar cell system has also been described by other workers to minimize foam losses during deaeration.[835]

Alternatively, *in situ* chemical deoxygenation can be employed.[836] For instance, sulfite ion has been utilized as an oxygen scavenger in anionic micellar[836] or CD[837] solutions. This method is based on the redox reaction shown in Equation (9):

$$O_2(g) + 2SO_3^{2-}(aq) \Longleftrightarrow 2SO_4^{2-}(aq) \qquad (9)$$

This approach is rapid (requiring about 12–15-min reaction time), convenient, precise, and superior to the solution nitrogen purging technique.[836] In particular, this deoxygenation procedure has been employed in the context of micelle- (or CD)-stabilized room-temperature phosphorescence[836,837] (*see* **Phosphorescence Measurements, Applications of**) as well as in applications of terbium-sensitized luminescence applications.[783] However, it has been cautioned that addition of sulfite to some micelle systems, such as SDS containing thallium(I), can result in the formation of a metal–sulfite complex which absorbs radiation below 275 nm.[838] Therefore, in applications involving the determination of metal ions in which this deoxygenation scheme is employed, experiments should first be conducted to ensure that such a metal–sulfite complex does not form or, if formed, that the resulting complex does not absorb appreciably at the required fluorescence excitation wavelength.

Another effective means of degassing solutions of organized media is through multiple freeze–pump–thaw cycles on a high-vacuum line to avoid the effects of traces of oxygen.[839,840] However, such vacuum degassing is not very convenient and requires a vacuum line system, etc. In addition, vacuum degassing micelle solutions are not significantly better than nitrogen purging as a means of deoxygenation as gauged from fluorescence probe studies.[840] However, there are some reports that note that the nitrogen bubbling technique for removal of oxygen from solutions does not eliminate all traces of the $O_2(g)$, with an estimated 2 μM residual oxygen gas still present following such purging.[502,686]

Lastly, the use of an automated sample deoxygenation system might also be applicable to solutions containing some of the organized assembly systems.[840] Two reports provide a nice survey of current methods employed for deoxygenation of solutions along with their advantages, limitations, and analytical applications.[841,842]

It should be noted that freshly prepared aqueous solutions of some organic solutes in the absence and presence of CDs appear to be unstable when the solutions are exposed to the atmosphere and/or bubbled by air or nitrogen gas.[843] Repeated bubbling–freezing–thawing cycles enhance the effect.

The preparation of required concentrated stock solutions of analytes in aqueous organized assembly solutions is sometimes problematic owing to extremely low solute solubilities or slow dissolution rates.[363,844] In many such cases, the following general protocol has been employed: (i) initially prepare a concentrated analyte stock solution in a water-miscible organic solvent (alcohol, acetone, tetrahydrofuran, etc.) in which it is soluble, (ii) then add a small aliquot of this concentrated solution to the aqueous organized assembly solution (typically strive to keep the percentage of the organic cosolvent under 0.5% (v/v)), (iii) next, purge the resulting organized assembly solution with nitrogen or argon in order to remove/displace the small amount of organic solvent present, and (iv) lastly, add an appropriate amount of water to this solution to compensate for evaporation.[194,363,844] Alternatively, the purge step can be skipped provided all analytical fluorescence measurements are conducted on standard and unknown solutions that contain the same amount of organic cosolvent. While this procedure is successful in most instances, for some solutes it results in the

formation of microcrystals of the analyte, and these can be a source of artifacts, since their resulting solutions can scatter light and/or give rise to strong excimer emission.[194,844]

Some of the preparation recipes for formation of SUV systems require the presence of small amounts of an organic solvent (typically chloroform, which is a potential fluorescence quencher as well). Thus, nitrogen purging is sometimes also utilized to remove traces of such organic solvents.[511,515]

Recall that the presence of organic cosolvents can preclude (or alter) surfactant micelle formation. Thus, if such cosolvents are utilized in the preparation of solutions, they should be present at low concentrations (final percentage <1.0% (v/v)).[385] Such a low organic cosolvent concentration serves to minimize the perturbation to the surfactant aggregate structure. As previously noted, bile salt surfactant systems can tolerate greater amounts of such organic additives than the classic surfactant micelle-forming systems.

As noted, when manipulating (pipetting, shaking to mix, etc.) aqueous surfactant micelle solutions, bubble (foam) formation can cause difficulties (such as when trying to add water to dilute to the mark of volumetric glassware). It has been noted that when working with such surfactant solutions in the laboratory, they can be freed of bubbles by placing the container in an ultrasonic bath filled with water. Apparently, after only a few seconds of exposure to ultrasound, clear, bubble-free solutions are obtained.[845] In one reported application, 15 min of this type of ultrasonic ''stirring'' was employed to remove air bubbles.[670]

As has been cautioned in the literature by others, care should be exercised with respect to the type of electrodes employed and the meaning of any pH measurements made in micellar- (or vesicle)-containing solutions.[846]

The interested reader is referred to any number of published applications for more detailed experimental protocols and the different experimental variables that can impact the fluorescence signal when utilizing different organized assemblies in fluorescence methods.[711,847–878]

5 CONCLUSIONS

In contrast to the situation in the mid-1980s, when the use of organized assemblies in analytical fluorimetry was still largely unexplored,[20] this is now an established practice in chemical analysis. The reported advantages in terms of the ability to control solubility, reactivity, sensitivity, and, to a lesser degree, selectivity by the judicious choice and use of an organized assembly system is the reason for the popularity and success of organized assembly–enhanced fluorescence measurements. Future research should be directed toward increased use of these systems in rate/kinetic methods of analysis, development of functionalized ordered media which target a specific analyte of interest, and different means by which to immobilize the surfactant micelle/vesicle systems. Development of such materials that mimic their properties will serve to extend the scope of possible applications. Some trends in this field have been noted in a review article.[879] In particular, the utilization of ''designer'' surfactant and CD media for application as specific fluorescent sensors is becoming very popular.[164,165,173,880–882] The use of surfactant-mediated CP extraction (for preconcentration of the analyte) combined with fluorescence enhancement should lead to the development of ultratrace methods. An overlooked but very beneficial application concerns the utilization of these aqueous organized media to enhance sample preservation and long-term storage of environmental and biological samples, in the context of not only fluorescence measurements but chemical analysis in general.

ACKNOWLEDGMENTS

This chapter is dedicated to late Prof. Janos H. Fendler for his pioneering research and development of the general field of organized assemblies and membrane mimetic chemistry. We thank our coworkers, whose names appear in the references listed, for their dedicated and enthusiastic contributions to our work cited in this review. Many thanks are due to Prof. Frank H. Quina (Instituto de Quimica, Universidad de Sao Paulo, Brazil) for his careful reading of the proofs and helpful comments in the preparation of this article. In addition, we thank an anonymous referee for helpful comments regarding the improvement of this manuscript as well as for providing additional useful citations. Acknowledgment is also made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, Pharmatec, Inc. (Alachua, FL) and Wake Forest University (through a Research and Publication Fund Grant) for providing financial support for our own work in this research area. This article was written while one of us (W. L. Hinze) was supported by a Wake Forest University Reynolds Research Leave. Additionally, N. Memon and A. Balouch gratefully acknowledge the Higher Education Commission of Pakistan and NCEAC (University of Sindh, Jamshoro, Pakistan) for their support of this work and granting of a study leave during their time at Wake Forest University.

ABBREVIATIONS AND ACRONYMS

ANS	8-Anilino-1-naphthalenesulfonate
AOD	Acridine orange-10-dodecyl Bromide
ATP	Adenosine Triphosphate
BA	Benzoic Acid
BID	2-Benzoylindan-1,3-dione
BP	2,2′-Bipyridine
BTA	Benzoyltrifluoroacetone
CBQCA	3-(4-Carboxybenzoyl)-2-quinolinecarboxaldehyde
CD	Cyclodextrin
CDEA	Cetyldimethylethyltrimethyl-ammonium bromide
CESLS	Constant-energy Synchronous Luminescence Spectrometry
CMC	Critical Micelle Concentration
CP	Cloud Point
CPB	Cetylpyridinium Bromide
CPC	Cetylpyridinium Chloride
CTAB	Cetyltrimethylammonium Bromide
CTAC	Cetyltrimethylammonium Chloride
CZE	Capillary Zone Electrophoresis
DBM	Dibenzoylmethane
DDAB	Didodecyldimethylammonium Bromide
DELFIA	Dissociated Enhanced Lanthanide Fluoroimmunoassay
DHP	Dihexadecyl Phosphate
DMA	Dimethylaniline
DMF	Dimethylformamide
DMPC	Dimyristoylphosphatidylcholine
DMPS	Dimyristoylphosphatidylserine
DODAB	Dioctadecyldimethylammonium Bromide
DODAC	Dioctadecyldimethylammonium Chloride
DOTA	1,4,7,10-Tetraazacyclododecane-1-[1′-carboxy-1′-dodecyl(methyl)-aminooxoethyl]-4,7,10-triacetic acid
DPH	1,6-Diphenyl-1,3,5-hexatriene
DPPC	Dipalmitoylphosphatidylcholine
DTAB	Dodecyltrimethylammonium Bromide
DTAC	Dodecyltrimethylammonium Chloride
EDTA	Ethylenediaminetetraacetate
FI	Fluorescence Intensity
FIA	Flow-injection Analysis
HFAC	Hexafluoroacetylacetone
HPLC	High-performance Liquid Chromatography
HPPA	*p*-Hydroxyphenylpropionic Acid
HQS	8-Hydroxyquinoline-5-sulfonic Acid
HRP	Horseradish Peroxidase
IDTTA	*N*-(*p*-Isothiocyanatobenzyl)-diethylenetriaminetetraacetic Acid
LUV	Large unilamellar Vesicle
MAB	Methyl 2-aminobenzoate
MB	Methylene Blue
MECC or MEKC	Micellar Enhanced/Electrokinetic Capillary Chromatography
MEDTA	1-[*p*-([(5-maleimidopentyl)carbonyl]amino}benzyl)ethylenediaminetetraacetic acid
MLV	Multilamellar Vesicle
MO	Methyl Orange
MR	Methyl Red
MRI	Magnetic Resonance Imaging
2-NAA	2-Naphthyloxyacetic Acid
NaC	Sodium Cholate
NaDC	Sodium Deoxycholate
NaL	Sodium Laurate
NaLS	Sodium Lauryl Sulfate
NaTC	Sodium Taurocholate
NBD	Nitrobenzoxadiazolyl
NDA	Naphthalene-2,3-dicarboxaldehyde
4-NBN	4-Nitro-*N*-*n*-butyl-1,8-naphthalimide
NTFA	2-Naphthoyltrifluoroacetone
PAC	Polycyclic Aromatic Compound
PADA	Pyridine-2-azo-*p*-dimethylaniline
PAH	Polycyclic Aromatic Hydrocarbon
PAN	1-(2-Pyridylazo)-2-naphthol
PC	Phosphatidylcholine
PG	Phosphatidylglycerol
phen	1,10-Phenanthroline
PIF	Photochemically Induced Fluorescence
PPIE	Pseudophase Ion Exchange
ppb	Parts Per Billion
PS	Phosphatidylserine
PTA	Pivaloyltrifluoroacetone
PTFE	Polytetrafluoroethylene
SB-12	Sulfobetain-12
SBD	7-Fluoro-2,1,3-benzoxadiazole-4-sulfonate
SDDBS	Sodium Dodecylbenzenesulfonate
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SSFP	Steady-state Fluorescence Polarization
STAC	Trimethylstearylammonium Chloride
SUV	Small (Sonicated) Unilamellar Vesicle
TDAB	Tetradecyltrimethylammonium Bromide
TDSS	Sodium Tetradecyl Sulfate

TFDMH	1,1,1-Trifluoro-5,5-dimethyl-2,4-hexanedione
TFTD	1,1,1-Trifluoro-2,4-tridecanedione
TLC	Thin-layer Chromatography
TNS	2-*p*-Toluidinonaphthalene-6-sulfonate
TOPO	Trioctylphosphine Oxide
TRCB	3,4,4′-Trichlorobiphenyl
TTA	Thenoyltrifluoroacetone
TTCB	3,3′, 4, 4′-Tetrachlorobiphenyl
TTCDF	2,3,7,8-Tetrachlorodibenzofuran
UV/vis	Ultraviolet/visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry • Drugs of Abuse, Analysis of • Lipid Analysis for Important Clinical Conditions • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Cyanogen Chloride and Cyanogen Bromide Analysis in Drinking Water • Dioxin-like Compounds, Screening Assays • Dyes, Environmental Analysis of

Environment: Water and Waste **(Volume 3)**
Inorganic Analysis in Environmental Samples by Capillary Electrophoresis

Environment: Water and Waste **(Volume 4)**
Phenols Analysis in Environmental Samples • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Analysis in Environmental Samples Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
Polycyclic Aromatic Compounds Mapping

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis of Peptides • Fluorescence Spectroscopy in Peptide and Protein Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy • Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Micellar Electrokinetic Chromatography • Thin-layer Chromatography

REFERENCES

1. E. Blatt, W.H. Sawyer, 'Depth-Dependent Flourescent Quenching in Micelles and Membranes', *Biochim. Biophys. Acta*, **822**, 43–62 (1985).
2. M.H. Gehlen, F.C. De Schryver, 'Time-Resolved Fluorescence Quenching in Micellar Assemblies', *Chem. Rev.*, **93**, 199–221 (1993).
3. M. Gratzel, J.K. Thomas, 'The Application of Fluorescence Techniques to the Study of Micellar Solution', in *Modern Fluorescence Spectroscopy*, ed. E. Wehry, Plenum Press, New York, 169–216, Chapter 4, Vol. 2, 1976.
4. K. Honda (ed.), *Photochemical Processes in Organized Molecular Systems*, North Holland, London, 1991.
5. K. Kalyanasundaram, 'Photophysics of Molecules in Micelle-Forming Surfactant Solutions', *Chem. Soc. Rev.*, **7**, 453–472 (1978).
6. K. Kalyanasundaram, *Photochemistry in Microheterogeneous Systems*, Academic Press, New York, 1987.
7. N.J. Turro, M. Gratzel, A.M. Braun, 'Photophysical and Photochemical Processes in Micellar Systems', *Agnew. Chem., Int. Ed. Engl.*, **19**, 675–696 (1980).
8. R. Zana, 'Luminescence Probing Method', in *Surfactant Solutions: New Methods of Investigation*, ed. R. Zana, Marcel Dekker, New York, 241–294, Chapter 5, 1987.
9. L.J. Cline Love, J.G. Dorsey, J.G. Habarta, 'The Micelle–Analytical Chemistry Interface', *Anal. Chem.*, **56**, 1132A–1139A (1984).
10. E. Pelizzetti, E. Pramauro, 'Analytical Applications of Organized Molecular Assemblies', *Anal. Chim. Acta*, **169**, 1–29 (1985).
11. U. Pfuller, *Mizellen, Vesikel, Mikroemulsionen Tensidassoziate und ihre Anwendung in Analytik und Biochemie*, VEB Verlag, Berlin, 1986.
12. E. Pramauro, E. Pelizzetti, *Surfactants in Analytical Chemistry: Applications of Organized Amphiphilic Media*, Elsevier, Amsterdam, 1996, and references therein.

13. J. Szejtli, T. Osa (eds.), *Comprehensive Supramolecular Chemistry: Cyclodextrins*, Elsevier, Oxford, Vol. 3, 1996.

14. J.J. Santana Ridriguez, R. Halko, J.R.B. Rodriguez, J.J. Aaron, 'Environmental Analysis Based on Luminescence in Organized Supramolecular Systems', *Anal. Bioanal. Chem.*, **385**, 525–545 (2006).

15. G.L. McIntire, 'Micelles in Analytical Chemistry', *Crit. Rev. Anal. Chem.*, **21**, 257–278 (1990).

16. S.N. Shtykov, 'Chemical Analysis in Nanoreactors: Main Concepts and Applications', *J. Anal. Chem.*, **57**, 859–868 (2002).

17. J.R. Lakowicz (ed.), *Principles of Fluorescence Spectroscopy*, 3rd edition, Springer-Verlag, New York, 2006.

18. J.H. Fendler, *Membrane Mimetic Chemistry*, Wiley-Interscience, New York, 1982.

19. W.L. Hinze (ed.), *Organized Assemblies in Chemical Analysis*, *Vol. 1 – Reverse Micelles*, JAI Press, Greenwich, CT, 1994.

20. W.L. Hinze, H.N. Singh, Y. Baba, N.G. Harvey, 'Micellar Enhanced Analytical Fluorimetry', *Trends Anal. Chem.*, **3**, 193–199 (1984).

21. Y. Mori, *Micelles: Theoretical and Applied Aspects*, Plenum Press, New York, 1992.

22. K. Goodling, K. Johnson, L. Lefkowitz, B.W. Williams, 'Luminescent Characterization of Sodium Dodecylsulfate Micellar Solution Properties', *J. Chem. Educ.*, **71**, A8–A12 (1994).

23. K.R. Acharya, S.C. Bhattacharyya, S.P. Moulik, 'The Surfactant Concentration-Dependent Behaviour of Safranine T in Tween (20, 40, 60, 80) and Triton X-100 Micellar Media', *J. Photochem. Photobiol.*, **109**, 29–34 (1997).

24. Z. Bi, 'Study of Absorption and Fluorescence Spectra of Fluorescein in Aqueous Solutions and its Possible Applications', *Huaxue Shiji*, **8**, 324–327 (1986).

25. K.S. Birdi, H.N. Singh, S.U. Dalsager, 'Interaction of Ionic Micelles with the Hydrophobic Fluorescent Probe 1-Anilino-8-Naphthalenesulfonate', *J. Phys. Chem.*, **83**, 2733–2737 (1979).

26. E. Blatt, 'Fluorescence Probe Studies of Anthracene and 9-Methylanthracene in Detergent Micelles', *Aust. J. Chem.*, **40**, 851–872 (1987).

27. A. Chattopadhyay, E. London, 'Fluorimetric Determination of Critical Micelle Concentration Avoiding Interference from Detergent Charge', *Anal. Biochem.*, **139**, 408–412 (1984).

28. G.S. Cox, P.J. Hauptman, N.J. Turro, 'Dialkylaminobenzonitriles as Fluorescence Polarity Probes for Aqueous-Solutions of Cyclodextrins', *Photochem. Photobiol.*, **39**, 597–601 (1984).

29. S.M. Dennison, J. Guharay, P.K. Sengupta, 'Intramolecular Excited-State Proton Transfer and Charge Transfer Fluorescence of a 3-Hydroxyflavone Derivative in Micellar Media', *Spectrochim. Acta, Part A*, **55**, 903–909 (1999).

30. C.J. Drummond, F. Grieser, T.W. Healy, 'A Single Spectroscopic Probe for the Determination of Both the Interfacial Solvent Properties and Electrostatic Surface-Potential of Model Lipid-Membranes', *Faraday Discuss.*, 95–106 (1986).

31. J. Emert, C. Behrens, M. Goldenberg, 'Intramolecular Excimer-Forming Probes of Aqueous Micelles', *J. Am. Chem. Soc.*, **101**, 771–772 (1979).

32. H. Ephardt, P. Fromherz, 'Fluorescence and Photoisomerization of an Amphiphilic Aminostilbazolium Dye as Controlled by the Sensitivity of Radiationless Deactivation to Polarity and Viscosity', *J. Phys. Chem.*, **93**, 7717–7725 (1989).

33. M. Fragata, F. Bellemare, 'Model of Singlet Oxygen Scavenging by [Alpha]-Tocopherol in Biomembranes', *Chem. Phys. Lipids*, **27**, 93–99 (1980).

34. M. Fragata, F. Bellemare, 'Micropolarities of Lipid Bilayers and Micelles: Ii. Application Limits of Dielectric Constant Determinations with the Polarity Probe α-Tocopherol', *J. Colloid Interface Sci.*, **107**, 553–559 (1985).

35. P. Goon, C. Manohar, V.V. Kumar, 'Determination of Critical Micelle Concentration of Anionic Surfactants: Comparison of Internal and External Fluorescent Probes', *J. Colloid Interface Sci.*, **189**, 177–180 (1997).

36. S. Hamai, 'Association Modes of a 1 : 1 Inclusion Compound of β-Cyclodextrin with 1-Cyanonaphthalene in Aqueous Solutions: Self-Association, Association with Alcohols, and Association with a 1 : 1 β-Cyclodextrin-Anisole Inclusion Compound', *J. Phys. Chem.*, **94**, 2595–2600 (1990).

37. J.B. Hasted, D.M. Ritson, C.H. Collie, 'Dielectric Properties of Aqueous Ionic Solutions. Parts I and II', *J. Chem. Phys.*, **16**, 1–21 (1948).

38. P. Horowitz, 'A Comparison between 8-Anilinonaphthalene-1-Suifonate and 2-p-Toluidinylnaphthalene-6-Sulfonate as Fluorescent Indicators of the Critical Micelle Concentration of Sodium Dodecyl Sulfate', *J. Colloid Interface Sci.*, **61**, 197–198 (1977).

39. R. Humphry-Baker, M. Graetzel, R. Steiger, 'Drastic Fluorescence Enhancement and Photochemical Stabilization of Cyanine Dyes through Micellar Systems', *J. Am. Chem. Soc.*, **102**, 847–848 (1980).

40. H. Itoh, S. Ishido, M. Nomura, T. Hayakawa, S. Mitaku, 'Estimation of the Hydrophobicity in Microenvironments by Pyrene Fluorescence Measurements: N-β-Octylglucoside Micelles', *J. Phys. Chem.*, **100**, 9047–9053 (1996).

41. K. Iwamoto, J. Sunamoto, 'Liposomal Membranes. 9. Fluorescence Depolarization Studies on N-Dansylhexadecylamine in Liposomal Bilayers', *Bull. Chem. Soc. Jpn.*, **54**, 399–403 (1981).

42. G. Jones, W.R. Jackson, S. Kanoktanaporn, W.R. Bergmark, 'Photophysical and Photochemical Properties of Coumarin Dyes in Amphiphilic Media', *Photochem. Photobiol.*, **42**, 477–483 (1985).

43. K. Kalyanasundaram, J.K. Thomas, 'Environmental Effects on Vibronic Band Intensities in Pyrene Monomer Fluorescence and Their Application in Studies of Micellar Systems', *J. Am. Chem. Soc.*, **99**, 2039–2044 (1977).

44. K. Kano, J.H. Fendler, 'Dynamic Fluorescence Investigations of the Effect of Osmotic Shocks on the Microenvironments of Charged and Uncharged Dipalmitoyl-D, L-α-Phosphatidylcholine Liposomes', *Chem. Phys. Lipids*, **23**, 189–200 (1979).

45. K. Kano, H. Goto, T. Ogawa, 'Solvent Effects on Fluorescence of Intramolecular Heteroexcimer System and its Use for Investigating Polarity of Microscopic Environment', *Chem. Lett.*, 653–656 (1981).

46. M.A. Kessler, O.S. Wolfbeis, 'A New HeNe-Laser Excitable Fluorescent Surfactant Probe', *Ber. Bunsen-Ges. Phys. Chem.*, **93**, 927–931 (1989).

47. H. Kondo, H. Nakatani, K. Hiromi, 'Interaction of Cyclodextrins with Fluorescent Probes and its Application to Kinetic Studies of Amylase', *J. Biochem. (Tokyo)*, **79**, 393–405 (1976).

48. I. Koppel, J. Koppel, 'E_T Parameters of Some Surfactants and Their Solutions', *Org. React.*, **26**, 78–91 (1989).

49. M. Kriechbaum, O.S. Wolfbeis, E. Koller, 'Evaluation of Critical Micelle Concentrations of Non-Ionic Detergents Using New Superpolar Lipid Probes', *Chem. Phys. Lipids*, **44**, 19–29 (1987).

50. C.E. Kung, J.K. Reed, 'Microviscosity Measurements of Phospholipid Bilayers Using Fluorescent Dyes that Undergo Torsional Relaxation', *Biochemistry*, **25**, 6114–6121 (1986).

51. K.Y. Law, 'Fluorescence Probe for Microenvironments: Anomalous Viscosity Dependence of the Fluorescence Quantum Yield of p-N,N-Dialkylaminobenzylidenemalononitrile in 1-Alkanols', *Chem. Phys. Lett.*, **75**, 545–549 (1980).

52. K.Y. Law, 'Fluorescence Probe for Micro-Environments – a New Probe for Micelle Solvent Parameters and Premicellar Aggregates', *Photochem. Photobiol.*, **33**, 799–806 (1981).

53. J.G. Lessard, M. Fragata, 'Micropolarities of Lipid Bilayers and Micelles. 3. Effect of Monovalent Ions on the Dielectric Constant of the Water Membrane Interface of Unilamellar Phosphatidylcholine Vesicles', *J. Phys. Chem.*, **90**, 811–817 (1986).

54. G.P. L'Heureux, M. Fragata, 'Micropolarities of Lipid Bilayers and Micelles: 4. Dielectric Constant Determinations of Unilamellar Phosphatidylcholine Vesicles with the Probes Pyrene and 16-(1-Pyrenyl)Hexadecanoic Acid', *J. Colloid Interface Sci.*, **117**, 513–522 (1987).

55. G.P. L'Heureux, M. Fragata, 'Micropolarities of Lipid Bilayers and Micelles: 5. Localization of Pyrene in Small Unilamellar Phosphatidylcholine Vesicles', *Biophys. Chem.*, **30**, 293–301 (1988).

56. G.P. L'Heureux, M. Fragata, 'Fluorescence Characteristics of Pyrene and Phosphatidylethanolamine-Bound Pyrene Incorporated into Lipid Vesicles Solubilized in Media of Differing NaCl', *Biophys. Chem.*, **34**, 163–168 (1989).

57. P. Lianos, R. Zana, 'Micelles of Tetradecyltrialkylammonium Bromides with Fluorescent Probes', *J. Colloid Interface Sci.*, **88**, 594–598 (1982).

58. S. Lukac, 'Surface Potential at Surfactant and Phospholipid Vesicles as Determined by Amphiphilic pH Indicators', *J. Phys. Chem.*, **87**, 5045–5050 (1983).

59. J.M. Madrid, F. Mendicuti, W.L. Mattice, 'Inclusion Complexes of 2-Methylnaphthoate and α-Cyclodextrin: Experimental Thermodynamics and Molecular Mechanics Calculations', *J. Phys. Chem. B*, **102**, 2037–2044 (1998).

60. R.C. Mast, L.V. Haynes, 'The Use of the Fluorescent Probes Perylene and Magnesium 8-Anilinonaphthalene-1-Sulfonate to Determine the Critical Micelle Concentration of Surfactants in Aqueous Solution', *J. Colloid Interface Sci.*, **53**, 35–41 (1975).

61. T. Moriya, 'The Solubilization Equilibrium of 7-Ethoxycoumarins in Ionic Micelles', *Bull. Chem. Soc. Jpn.*, **61**, 585–587 (1988).

62. P. Mukerjee, C. Ramachandran, R.A. Pyter, 'Solvent Effects on the Visible Spectra of Nitroxides and Relation to Nitrogen Hyperfine Splitting Constants. Nonempirical Polarity Scales for Aprotic and Hydroxylic Solvents', *J. Phys. Chem.*, **86**, 3189–3197 (1982).

63. M. Nakagaki, H. Komatsu, T. Handa, 'Estimation of Critical Micelle Concentrations of Lysolecithins with Fluorescent-Probes', *Chem. Pharm. Bull.*, **34**, 4479–4485 (1986).

64. M. Nakagaki, H. Komatsu, T. Handa, 'Microenvironment around Thiacarbocyanine Dyes in Lysolecithin Micelles, Surfactant Micelles and Lecithin Liposomal Membranes', *Chem. Pharm. Bull.*, **34**, 4486–4493 (1986).

65. S. Nigam, G. Durocher, 'Spectral and Photophysical Studies of Inclusion Complexes of Some Neutral 3H-Indoles and Their Cations and Anions with β-Cyclodextrin', *J. Phys. Chem.*, **100**, 7135–7142 (1996).

66. L.P. Novaki, O.A. El Seoud, 'Solvatochromism in Aqueous Micellar Solutions: Effects of the Molecular

Structures of Solvatochromic Probes and Cationic Surfactants', *Phys. Chem. Chem. Phys.*, **1**, 1957–1964 (1999).

67. L.P. Novaki, O.A. El Seoud, 'Solvatochromism in Pure Solvents: Effects of the Molecular Structure of the Probe', *Ber. Bunsen-Ges. Phys. Chem.*, **100**, 648–655 (1996).
68. C. Ramachandran, R.A. Pyter, P. Mukerjee, 'Microenvironmental Effects on Transition Energies and Effective Polarities of Nitroxides Solubilized in Micelles of Different Charge Types and the Effect of Electrolytes on the Visible Spectra of Nitroxides in Aqueous Solutions', *J. Phys. Chem.*, **86**, 3198–3205 (1982).
69. R. Romero, J. Sunamoto, J.H. Fendler, 'Steady State and Nanosecond Resolved Fluorescence of Dansyl-N-Octadecylamine in Bilayer Liposomes', *J. Colloid Interface Sci.*, **5**, 111–118 (1976).
70. M. Rujimethabhas, P. Wilairat, 'Determination of Critical Micelle Concentration Using Acridine Orange Dye Probe – Undergraduate Experiment', *J. Chem. Educ.*, **55**, 342 (1978).
71. A. Safarzadeh-Amiri, M. Thompson, U.J. Krull, 'An Investigation of the Physical Properties of Synthetic Phospholipid Bilayers Using a Fluorescent Quinoline-Based Probe', *J. Photochem. Photobiol.*, **49**, 151–165 (1989).
72. A. Safarzadeh-Amiri, M. Thompson, U.J. Krull, 'Trans-4-Dimethylamino-4'-(1-Oxobutyl)Stilbene: A New Fluorescent Probe of the Bilayer Lipid Membrane', *J. Photochem. Photobiol.*, **47**, 299–308 (1989).
73. N. Sarkar, K. Bhattacharyya, 'Effect of Urea on Micelles: Fluorescence of p-Toluidino Naphthalene Sulphonate', *Chem. Phys. Lett.*, **180**, 283–286 (1991).
74. R.S. Sarpal, S.K. Dogra, 'Absorption and Fluorescence Spectral Characteristics of Aromatic Amines in Cetyltrimethylammonium Bromide', *J. Chem. Soc., Faraday Trans.*, **88**, 2725–2731 (1992).
75. M. Shinitzky, 'Fluidity and Order in Hydrocarbon-Water Interface of Synthetic and Biological Micelles as Determined by Fluorescence Polarization', *Isr. J. Chem.*, **12**, 879–890 (1974).
76. K.W. Street, 'Cyclodextrin Cavity Polarity and Chromatographic Implications', *J. Liq. Chromatogr. Related Technol.*, **10**, 655–662 (1987).
77. K.W. Street, W.E. Acree, 'Estimation of the Effective Dielectric-Constant of Cyclodextrin Cavities Based on the Fluorescence Properties of Pyrene-3-Carboxaldehyde', *Appl. Spectrosc.*, **42**, 1315–1318 (1988).
78. W.D. Turley, H.W. Offen, 'Lipid Microviscosity of DMPC Vesicles at High Pressures. Dipyrenylpropane Excimer Fluorescence', *J. Phys. Chem.*, **90**, 1967–1970 (1986).
79. N.J. Turro, M. Aikawa, A. Yekta, 'A Comparison of Intermolecular and Intramolecular Excimer Formation in Detergent Solutions. Temperature Effects and Microviscosity Measurements', *J. Am. Chem. Soc.*, **101**, 772–774 (1979).
80. N.J. Turro, P.L. Kuo, 'Fluorescence Probes for Aqueous Solutions of Nonionic Micelles', *Langmuir*, **1**, 170–172 (1985).
81. N.J. Turro, M. Okamoto, P.L. Kuo, 'Pressure Effects on the Fluorescence Decay of Pyrene in Micellar Solutions', *J. Phys. Chem.*, **91**, 1819–1823 (1987).
82. N.J. Turro, T. Okubo, 'Micellar Microviscosity of Ionic Surfactants under High Pressure', *J. Am. Chem. Soc.*, **103**, 7224–7228 (1981).
83. N.J. Turro, T. Okubo, 'Microscopic Environment of Poly(Styrenesulfonate) Macroanions. Emission and Absorption Spectra, Lifetime, and Depolarization Measurements of a Cationic Fluorescence Probe Under High Pressure', *J. Am. Chem. Soc.*, **104**, 2985–2988 (1982).
84. E.N. Viktorova, I.V. Kovaleva, R.I. Kurtsinovskaya, M.I. Snegov, 'Luminescence Polarization of Dyes in Aqueous Solutions with Additions of Detergents', *Opt. Spectrosc. (USSR)*, **66**, 201–203 (1989).
85. Z.N. Volovik, A.P. Demchenko, A.A. Ishchenko, Y.L. Slominsky, A.I. Tolmachev, 'Nonsymmetrical Polymethine Dyes as Fluorescence Probes in Studies of Membrane Phospholipid-Bilayer Microviscosity', *Ukr. Biokhim. Zh.*, **60**, 64–70 (1988).
86. R. Waris, K.W. Street, W.E. Acree, J.C. Fetzer, 'Polycyclic Aromatic Hydrocarbon Solute Probes. Part IV: Effect of Solvent Polarity on the Fluorescence Emission Fine Structures of Select Pyrene and Pentaphene Derivatives', *Appl. Spectrosc.*, **43**, 845–850 (1989).
87. G. Weber, M. Shinitzky, A.C. Dianoux, C. Gitler, 'Microviscosity and Order in the Hydrocarbon Region of Micelles and Membranes Determined with Fluorescent Probes. I. Synthetic Micelles', *Biochemistry*, **10**, 2106–2113 (1971).
88. J.E. Wong, T.M. Duchscherer, G. Pietraru, D.T. Cramb, 'Novel Fluorescence Spectral Deconvolution Method for Determination of Critical Micelle Concentrations Using the Fluorescence Probe Prodan', *Langmuir*, **15**, 6181–6186 (1999).
89. K.A. Zachariasse, 'Intramolecular Excimer Formation with Diarylalkanes as a Microfluidity Probe for Sodium Dodecyl Sulphate Micelles', *Chem. Phys. Lett.*, **57**, 429–432 (1978).
90. K.A. Zachariasse, P. Nguyen Van, B. Kozankiewicz, 'Investigation of Micelles, Microemulsions, and Phospholipid Bilayers with the Pyridinium-N-Phenolbetaine $E_T(30)$, a Polarity Probe for Aqueous Interfaces', *J. Phys. Chem.*, **85**, 2676–2683 (1981).
91. L.L. Yu, M.Y. Tan, B. Ho, J.L. Ding, T. Wohland, 'Determination of Critical Micelle Concentrations and

Aggregation Numbers by Fluorescence Correlation Spectroscopy: Aggregation of a Lipopolysaccharide', *Anal. Chim. Acta*, **556**, 216–225 (2006).

92. Y. Nakahara, T. Kida, Y. Nakatsuji, M. Akashi, 'New Fluorescence Method for the Determination of the Critical Micelle Concentration by Photosensitive Monoazacryptand Derivatives', *Langmuir*, **21**, 6688–6695 (2005).
93. J.I.G. Alonso, M.E.D. Garcia, A.S. Medel, 'The Surfactant-Sensitized Analytical Reaction of Niobium with 8-Hydroxyquinoline-5-Sulphonic Acid', *Talanta*, **31**, 361–366 (1984).
94. D. Attwood, A.T. Florence, *Surfactant Systems: Their Chemistry, Pharmacy and Biology*, Chapman & Hall, New York, 1983.
95. J.L. Carrion Dominguez, M. de la Guardia Cirugeda, 'Spectroscopic Study of the Aluminium/Lumogallion System in the Presence of Non-Ionic Surfactants', *Anal. Chim. Acta*, **198**, 53–61 (1987).
96. M.V. Galdu, M. Delaguardia, L. Braco, 'Influence of Anionic Surfactants on the Sensitization of the Fluorometric-Determination of Fenproporex', *Analyst*, **112**, 1047–1050 (1987).
97. E. Martinez de la Ossa, V. Flores, 'Effect of Temperature and Electrolytes on CMC-Combined Effect of Temperature and Electrolytes on Critical Micelle Concentration', *Tenside Surfactants Deterg.*, **24**, 38–41 (1987).
98. S. Nigam, S.K. Dogra, 'Proton Transfer Reactions of Benzimidazole in Ionic Micelles in the Excited Singlet State', *J. Photochem. Photobiol., A*, **54**, 219–230 (1990).
99. J. Sancenon, M. de la Guardia, 'Micellar Enhanced Fluorimetric Determination of Carbendazim in Natural Waters', *Anal. Chim. Acta*, **287**, 49–57 (1994).
100. A. Sanz-Medel, J.I. Garcia Alonso, E. Blanco Gonzalez, 'Metal Chelate Fluorescence Enhancement in Micellar Media and its Applications to Niobium and Tantalum Ultratrace Determinations', *Anal. Chem.*, **57**, 1681–1687 (1985).
101. M.E.D. Garcia, A. Sanz-Medel, 'Dye-Surfactant Interactions: A Review', *Talanta*, **33**, 255–264 (1986).
102. H. Sato, M. Kawasaki, K. Kasatani, 'Energy Transfer between Rhodamine 6G and Pinacyanol Enhanced with Sodium Dodecyl Sulfate in the Premicellar Region. Formation of Dye-Rich Induced Micelles', *J. Phys. Chem.*, **87**, 3759–3769 (1983).
103. Y.-B. Jiang, X.-J. Wang, 'Direct Evidence for Beta-Cyclodextrin-Induced Aggregation of Ionic Surfactant Below Critical Micelle Concentration', *Appl. Spectrosc.*, **48**, 1428–1431 (1994).
104. T. Wolff, B. Klaussner, G. Von Bunau, 'Multiplicity Selective Photoisomerization of 5,5-Diphenyl-1,3-Hexadiene in Aqueous Micellar Solutions Containing Heavy Counter-Ions', *J. Photochem. Photobiol., A*, **47**, 345–351 (1989).
105. A.A. Fouda, L.H. Madkour, D.F. Evans, 'Critical Micelle Concentration of Hexadecyltrimethylammonium Chloride in Water over the Temperature Range 25 to 160 °C, *Indian J. Chem., Sect A*, **25A**, 1102–1104 (1986).
106. L.M.P. Diez, M.W. Kabbani, J.S.D. Alegria, 'Spectrofluorimetric Determination of Trace Amounts of Ga(III) in Aluminum and Biological Samples with 1-(2-Pyridylazo)-2-Naphthol in Sodium Dodecyl-Sulfate Micellar Medium', *Fresenius J. Anal. Chem.*, **334**, 41–44 (1989).
107. J.H. Garcia, Z.S. Ferrera, A.J. Bermejo, A.J.B. Martinlazaro, J.J.S. Rodriguez, 'Fluorometric Study of PCBs and Aroclors in Micellar Media – Analytical Applications', *Anal. Lett.*, **27**, 1355–1382 (1994).
108. J. Hernandez Garcia, J.R. Betancort Rodriguez, A.J. Bermejo Martin-Lazaro, J.J. Santana Rodriguez, 'Sensitive Spectrofluorimetric Determination of 3,3′,4,4′-Tetrachlorobiphenyl Using a Neutral Micellar Medium', *Anal. Chim. Acta*, **290**, 146–153 (1994).
109. F. Hernandez Hernandez, J. Medina Escriche, M.T. Gasco Andreu, 'Enhancement of the Fluorescence of the Zinc – Morin Complex by a Non-Ionic Surfactant', *Talanta*, **33**, 537–540 (1986).
110. G. Jinzhang, W. Xiwen, H. Jiang, B. Guangbi, 'Studies on Derivative Fluorimetry. II. Simultaneous Determination of Traces of Samarium, Europium and Terbium through their Ternary Complexes with 2-Thenoyltrifluoroacetone and Diphenylguanidine in the Presence of Triton X-100', *Inorg. Chim. Acta*, **140**, 273–275 (1987).
111. Y. Okabayashi, T. Kitagawa, 'High-Performance Liquid Chromatography for Amino Compounds and Thiol Compounds Derivatized with Europium Chelate', *Anal. Chem.*, **66**, 1448–1453 (1994).
112. S. Panadero, A. Gomezhens, D. Perezbendito, 'Kinetic Determination of Propyl Gallate in Edible and Cosmetic Oils with Sensitized Terbium(III) Luminescence Detection', *Analyst*, **120**, 125–128 (1995).
113. F. Wei, E. Ten, Z. Wu, 'Enhancement of the Fluorescence of the Beryllium – Morin Complex by Non-Ionic Surfactants', *Talanta*, **37**, 947–950 (1990).
114. M.J. Rosen, *Surfactants and Interfacial Phenomena*, Wiley-Interscience, New York, 1978.
115. S.E. Fribert, B. Lindman (eds.), *Organized Solutions*, *Surfactant Science Series, Vol. 44*, Marcel Dekker, New York, 1992.
116. R. Zana (ed.), *Dynamics of Surfactant Self-Assemblies: Micelles, Microemulsions, Vesicles, and Lyotropic Phases*, *Surfactant Science Series, Vol. 125*, Taylor & Francis, New York, 2005.

117. D.M. Small, 'The Physical Chemistry of Cholanic Acid', in *The Bile Acids*, eds. P.P. Nair, D. Kritchevsky, Plenum Press, New York, 249–403, Vol. 1, 1971.

118. M.C. Carry, 'Bile Acids and Bile Salts', in *Steroids and Bile Acids*, eds. H. Danielesson, J. Sjovall, Elsevier, New York, 345–403, 1985.

119. W.L. Hinze, W. Hu, F.H. Quina, I. Ullah, 'Bile Acid/Salt Surfactant Systems: General Properties and Survey of Analytical Application', in *Bile Acid/Salt Surfactant Systems*, *Organized Assemblies in Chemical Analysis: Vol. 2*, ed. W.L. Hinze, JAI Press, Stamford, CT, 1–70, Ch. 1, 2000.

120. H. Kunieda, K.-I. Ohyama, 'Three-Phase Behavior and HLB Numbers of Bile Salts and Lecithin in a Water-Oil System', *J. Colloid Interface Sci.*, **136**, 432–439 (1990).

121. K. Nithipatikom, L.B. McGown, 'Effects of Metal Cations on the Fluorescence Intensity of Polycyclic Aromatic Hydrocarbons in Sodium Taurocholate Micellar Solutions', *Anal. Chem.*, **60**, 1043–1045 (1988).

122. L.B. McGown, K. Nithipatikom, 'Sodium Taurocholate Micelles in Fluorometric Analysis', *J. Res. Natl. Bur. Stand.*, **93**, 443–444 (1988).

123. L.B. McGown, S.M. Meyerhoffer, P.R. Hertz, G. Li, 'Luminescence in Bile Salt Micelle', in *Bile Acid/Salt Surfactant Systems*, *Organized Assemblies in Chemical Analysis: Vol. 2*, ed. W.L. Hinze, JAI Press, Stamford, CT, 167–186, Chapter 5, 2000.

124. R.O. Cole, M.J. Sepaniak, W.L. Hinze, J. Gorse, K. Oldiges, 'Bile-Salt Surfactants in Micellar Electrokinetic Capillary Chromatography – Application to Hydrophobic Molecule Separations', *J. Chromatogr.*, **557**, 113–123 (1991).

125. W.L. Hinze (ed.), *Bile Acid/Salt Surfactant Systems*, JAI Press, Stamford, CT, Vol. 2, 2000.

126. R. Song, K.-P. Li, 'Fluorometry of Carcinogenic Polycyclic Aromatic Hydrocarbons in Biological Systems', *Appl. Spectrosc.*, **47**, 1604–1608 (1993).

127. T.G. Burke, T.R. Tritton, 'Location and Dynamics of Anthracyclines Bound to Unilamellar Phosphatidylcholine Vesicles', *Biochemistry*, **24**, 5972–5980 (1985).

128. K. Kitamura, M. Yamamoto, S. Takegami, M. Sugiura, 'Determination of the Partition Coefficient of Triflupromazine between Phosphatidylcholine Vesicles and Water by ^{19}F Nuclear Magnetic Resonance Spin-Lattice Relaxation Time Measurement', *Talanta*, **49**, 261–265 (1999).

129. E.H.W. Pap, J.J. Ter Horst, A. van Hoek, A.J.W.G. Visser, 'Fluorescence Dynamics of Diphenyl-1,3,5-Hexatriene-Labeled Phospholipids in Bilayer Membranes', *Biophys. Chem.*, **48**, 337–351 (1994).

130. J. Sunamoto, K. Akiyoshi, T. Kihara, M. Endo, 'A Water-Soluble Spin Probe Newly Developed for Liposomal Studies', *Bull. Chem. Soc. Jpn.*, **65**, 1041–1046 (1992).

131. A. Datta, S.K. Pal, D. Mandal, K. Bhattacharyya, 'Solvation Dynamics of Coumarin 480 in Vesicles', *J. Phys. Chem. B*, **102**, 6114–6117 (1998).

132. K. Kano, J.H. Fendler, 'Pyranine as a Sensitive Ph Probe for Liposome Interiors and Surfaces – pH Gradients Across Phospholipid Vesicles', *Biochim. Biophys. Acta*, **509**, 289–299 (1978).

133. T.Y. Tsong, 'Effect of Phase Transition on the Kinetics of Dye Transport in Phospholipid Bilayer Structures', *Biochemistry*, **14**, 5409–5414 (1975).

134. S. Yamashita, A.G. Szabo, D.T. Krajcarski, N. Yamasaki, 'Fluorescence Studies on the Phase Dependence of Interactions between a Tripeptide, Lys-Trp-Lys, and Dimyristoylphosphatidylserine', *Bull. Chem. Soc. Jpn.*, **62**, 3075–3080 (1989).

135. P.F. Aramendia, M. Krieg, C. Nitsch, E. Bittersmann, S.E. Braslavsky, 'The Photophysics of Merocyanine 540 – a Comparative-Study in Ethanol and in Liposomes', *Photochem. Photobiol.*, **48**, 187–194 (1988).

136. G.P. Gorbenko, 'Resonance Energy Transfer Study of Hemoglobin Complexes with Model Phospholipid Membranes', *Biophys. Chem.*, **81**, 93–105 (1999).

137. H.V. Jakubowski, M. Penas, K. Saunders, 'The Study of Lipid Aggregates in Aqueous-Solution – Formation and Properties of Liposomes with an Encapsulated Metallochromic Dye', *J. Chem. Educ.*, **71**, 347–349 (1994).

138. N.P. Matylevich, M.N. Ivkova, V.A. Pechatnikov, V.G. Ivkov, 'Interaction of the Fluorescent-Probe $C_{3-(5)}$ with Lecithin-Cholesterol Membranes', *Biofizika*, **31**, 997–1000 (1986).

139. M. Rotenberg, D. Lichtenberg, 'What Determines the Size of Phospholipid-Vesicles Made by Detergent-Removal Techniques', *J. Colloid Interface Sci.*, **144**, 591–594 (1991).

140. F. Szoka, D. Papahadjopoulos, 'Comparative Properties and Methods of Preparation of Lipid Vesicles (Liposomes)', *Annu. Rev. Biophys. Bioeng.*, **9**, 467–508 (1980).

141. G. Gregoriadis (ed.), *Liposome Technology: Preparation of Liposomes*, CRC Press, Boca Raton, FL, 1984.

142. I.F. Uchegbu, *Synthetic Surfactant Vesicles*, CRC Press, Boco Raton, FL, 2000.

143. G. Gregoriadis (ed.), *Liposome Technology: Liposome Preparation and Related Techniques*, Taylor & Francis, New York, 2006.

144. V.P. Torchilin, V. Weissig (eds.), *Liposomes: A Practical Approach*, Oxford University Press, London, 2003.

145. A.P. Winiski, M. Eisenberg, M. Langner, S. McLaughlin, 'Fluorescent Probes of Electrostatic Potential 1 nm from the Membrane Surface', *Biochemistry*, **27**, 386–392 (1988).

146. E. Abuin, E. Lissi, D. Aravena, A. Zanocco, M. Macuer, 'A Fluorescence Probe Study of the Effect of Size on the Properties of Dioctadecyldimethylammonium Chloride Vesicles', *J. Colloid Interface Sci.*, **122**, 201–208 (1988).

147. M.J. Gonzalez Alvarez, M.E. Diaz Garcia, A. Sanz-Medel, 'Exploitation of Synthetic Surfactant Vesicles for Enhanced Fluorescence of Metal Chelates', *Anal. Chim. Acta*, **234**, 181–186 (1990).

148. N. Nakashima, R. Ando, T. Kunitake, 'Fluorescence Behavior and Energy-Transfer of Cyanine Dyes Bound to Bilayer Membranes of Double-Chain Ammonium Amphiphiles', *Bull. Chem. Soc. Jpn.*, **60**, 1967–1973 (1987).

149. A.M. O'Reilly, J.D. Winefordner, 'The Effects of Magnitite Particles Encapsulated in Dihexyldecylphosphate Vesicles on the Luminescence Properties of Molecules', *Spectrochim. Acta, Part A*, **44**, 1395–1401 (1988).

150. L.A.M. Rupert, J. Engberts, D. Hoekstra, 'Role of Membrane Hydration and Membrane Fluidity in the Mechanism of Anion-Induced Fusion of Didodecyldimethylammonium Bromide Vesicles', *J. Am. Chem. Soc.*, **108**, 3920–3925 (1986).

151. M.R.F. de la Campa, Y.M. Liu, M.E.D. Garcia, A. Sanz-Medel, 'Fluorescence and Liquid Room-Temperature Phosphorescence of the Aluminium-Ferron Chelate in Vesicles', *Anal. Chim. Acta*, **238**, 297–305 (1990).

152. F. Grieser, C.J. Drummond, 'The Physicochemical Properties of Self-Assembled Surfactant Aggregates as Determined by Some Molecular Spectroscopic Probe Techniques', *J. Phys. Chem.*, **92**, 5580–5593 (1988).

153. D.E. Bennett, D.F. O'Brien, 'Photoinduced Fusion of Liposomes', *J. Am. Chem. Soc.*, **116**, 7933–7934 (1994).

154. G. Gregoriadis (ed.), *Liposome Technology: Incorporation of Drugs, Proteins, and Genetic Materials*, CRC Press, Boca Raton, FL, Vol. II, 1984.

155. G. Gregoriadis (ed.), *Liposome Technology: Targeted Drug Delivery and Biological Interaction*, CRC Press, Boca Raton, FL, Vol. III, 1984.

156. S.L. Neal, M.M. Villegas, 'Comparing Organized Media Types Using Fluorescence Probe Characterization Methods', *Anal. Chim. Acta*, **307**, 419–430 (1995).

157. S.L. Neal, M.M. Villegas, 'Phospholipid Vesicles as Organized Media in Spectroluminescence Analysis', *Anal. Chem.*, **67**, 2659–2665 (1995).

158. R.J. Bergeron, 'Cycloamylose-Substrate Binding', in *Inclusion Compounds*, eds. J.L. Atwood, J.E.D. Davies, D.D. MacNicol, Academic Press, New York, 391, Chapter 12, Vol. 3, 1984.

159. J. van Stam, F. Imans, L. Viaene, F.C. De Schryver, C.H. Evans, α-Terthiophene in Micellar Solutions: Influence of Micellar Size and Charge on the Terthiophene Photophysics and Photochemical Reactivity', *J. Phys. Chem. B*, **103**, 5160–5166 (1999).

160. J. Szejtli, *Cyclodextrin Technology*, Kluwer Academic Publishers, Dordrecht, 1988.

161. M.L. Bender, M. Komiyama, *Cyclodextrin Chemistry*, Springer, New York, 1978.

162. H. Dodziuk, *Cyclodextrins and Their Complexes*, Wiley-VCH, Weinheim, 2006.

163. Y. Liu, B.H. Han, H.Y. Zhang, 'Spectroscopic Studies on Molecular Recognition of Modified Cyclodextrins', *Curr. Org. Chem.*, **8**, 35–46 (2004).

164. Y. Liu, Y. Chen, 'Cooperative Binding and Multiple Recognition by Bridged Bis(Beta-Cyclodextrin) with Functional Linkers', *Acc. Chem. Res.*, **39**, 681–691 (2006).

165. C.J. Easton, S.F. Lincoln, *Modified Cyclodextrins*, Imperial College Press, London, 1999.

166. W.L. Hinze, N. Srinivasan, T.K. Smith, S. Igarashi, H. Hoshino, 'Organized Assemblies in Analytical Chemiluminescence Spectroscopy: An Overview', in *Advances in Multidimensional Luminescence*, eds. L.B. McGown, I.M. Warner, JAI Press, Greenwich, CT, 149–206, Vol. 1, 1990.

167. G. Bunau, T. Wolff, 'Photochemistry in Surfactant Solutions', *Adv. Photochem.*, **14**, 273–331 (1988).

168. J.H. Fendler, 'Microemulsions, Micelles, and Vesicles as Media for Membrane Mimetic Photochemistry', *J. Phys. Chem.*, **84**, 1485–1491 (1980).

169. M.G. Kuzmin, N.K. Zaitsev, 'Kinetics of Photochemical Charge Separation in Micellar Solutions', *Interface Struct. Electrochem. Processes Boundary Two Immiscible Liq.*, 207–244 (1987).

170. R. Zana, 'Micelles and Microemulsions – Review of Recent Structural and Dynamic Results in Relation with Reactivity', *J. Chim. Phys. Phys.- Chim. Biol.*, **83**, 603–612 (1986).

171. G.B. Behera, B.K. Mishra, P.K. Behera, M. Panda, 'Fluorescent Probes for Structural and Distance Effect Studies in Micelles, Reversed Micelles and Microemulsions', *Adv. Colloid Interface Sci.*, **82**, 1–42 (1999).

172. N.E. Geacintov, 'Principles and Applications of Fluorescence Techniques in Biophysical Chemistry', *Photochem. Photobiol.*, **45**, 547–553 (1987).

173. S.O. Fakayode, M. Lowry, K.A. Fletcher, X. Huang, A.M. Powe, I.M. Warner, 'Cyclodextrin Host – Guest Chemistry in Analytical and Environmental Chemistry', *Curr. Anal. Chem.*, **3**, 171–181 (2007), and references therein.

174. A. Malliaris, 'Recent Advances in Single Photon Counting in Using Fluorescence Spectroscopy for the Study of Molecular Aggregates', *Appl. Spectrosc. Rev.*, **27**, 51–66 (1991).

175. M.J.P. Gallego, C. Bravo-Diaz, E. Gonzalez-Romero, 'Fluorimetric Determination of Structural Parameters of

BuOH/SDS/H_2O Reverse Micelles', *Colloids Surf., A*, **249**, 25–28 (2004).

176. F.H. Quina, E.A. Lissi, 'Photoprocesses in Microaggregates', *Acc. Chem. Res.*, **37**, 703–710 (2004).
177. E.I. Akimova, M.I. Snegov, 'Photoxidation of Aminocoumarines in Aqueous Micellar Solutions', *Opt. Spectros. (USSR)*, **63**, 41–42 (1987).
178. M. Almgren, F. Grieser, J.K. Thomas, 'Dynamic and Static Aspects of Solubilization of Neutral Arenes in Ionic Micellar Solutions', *J. Am. Chem. Soc.*, **101**, 279–291 (1979).
179. M. Celeste, C. Serra, J.A.P. Coelho, J.C.G. Calado, A.M.F. Palavra, 'Solubility of Argon in Micellar Aqueous Solutions of Sodium Dodecyl Sulfate', *J. Colloid Interface Sci.*, **173**, 278–283 (1995).
180. C. David, E. Szalai, D. Baeyensvolant, 'Photophysical Processes in Cyanobiphenyl Derivatives: Cyanobiphenyl Derivatives as Fluorescent-Probes in Micellar Environment', *Ber. Bunsen-Ges. Phys. Chem.*, **86**, 710–716 (1982).
181. A. Maciejewski, A. Jakubowska, E. Dutkiewicz, W. Augustyniak, 'Spectral and Photophysical Properties of Aromatic Thiones in Micellar Systems: Xanthione as Convenient Absorption and Emission Probes', *J. Colloid Interface Sci.*, **177**, 528–541 (1996).
182. P. Mukerjee, J.R. Cardinal, 'Solubilization as a Method for Studying Self-Association – Solubility of Naphthalene in Bile-Salt Sodium Cholate and Complex Pattern of its Aggregation', *J. Pharm. Sci.*, **65**, 882–886 (1976).
183. G. Nelson, G. Patonay, I.M. Warner, 'Effects of Selected Alcohols on Cyclodextrin Inclusion Complexes of Pyrene Using Fluorescence Lifetime Measurements', *Anal. Chem.*, **60**, 274–279 (1988).
184. D. Rivera, E. Morales, L. Mercado, G.A. Infante, 'Solubilization of Folic Acid by Micelle-Forming Surfactants', *Science-Ciencia*, **6**, 26–27 (1978).
185. J.S. Wu, K. Toda, A. Tanaka, I. Sanemasa, 'Association Constants of Ferrocene with Cyclodextrins in Aqueous Medium Determined by Solubility Measurements of Ferrocene', *Bull. Chem. Soc. Jpn.*, **71**, 1615–1618 (1998).
186. K.H. Shin, K.W. Kim, E.A. Seagren, 'Combined Effects of pH and Biosurfactant Addition on Solubilization and Biodegradation of Phenanthrene', *Appl. Microbiol. Biotechnol.*, **65**, 336–343 (2004).
187. V. Leon-Ruiz, S. Vera, M.P. San Andres, 'Validation of a Screening Method for the Simultaneous Identification of Fat-Soluble and Water-Soluble Vitamins (A, E, B-1, B-2 and B-6) in an Aqueous Micellar Medium of Hexadecyltrimethylammonium Chloride', *Anal. Bioanal. Chem.*, **381**, 1568–1575 (2005).
188. H. Hoiland, A.M. Blokhus, 'Solubilization in Aqueous Surfactant Systems', in *Handbook of Surface and Colloid Chemistry*, ed. K.S. Birdi, CRC Press, Boca Raton, FL, Chapter 8, 1997.
189. R.Y. Krasnoshchekova, M.Y. Gubergrits, 'Solubilization of Polycyclic Arenes in the Micelles of Ionic Surfactants', *Colloid J. (USSR)*, **45**, 300–303 (1983).
190. S.H. Yalkowsky, *Solubility and Solubilization in Aqueous Media*, American Chemical Society, Washington, DC, Chapter 7, 1999.
191. M.J.Z.Y. Li Chunmao, 'Study on Effects of Cyclodextrins on the Photolysis of Dissolved Anthracene by Fluorometry', *Luminescence*, **20**, 261–265 (2005).
192. W.S. Higazy, F.Z. Mahmoud, 'Effects on the Micellar Solubilization of Organic Compounds by Surfactant Micelles. II. Nature and Position of Substituent Groups in Aromatic Acids', *Pol. J. Chem.*, **73**, 1555–1562 (1999).
193. S. Ozeki, S. Ikeda, 'The Difference in Solubilization Power between Spherical and Rodlike Micelles of Dodecyldimethylammonium Chloride in Aqueous Solutions', *J. Phys. Chem.*, **89**, 5088–5093 (1985).
194. K. Kalyanasundaram, 'Pyrene Fluorescence as a Probe of Fluorocarbon Micelles and their Mixed Micelles with Hydrocarbon Surfactants', *Langmuir*, **4**, 942–945 (1988).
195. E.N. Viktorova, T.V. Veselova, M.I. Snegov, A.S. Cherkasov, 'Spectral-Luminescent Properties of Pthalimides in Aqueous Micellar Solutions of Sodium Dodecylsulfate', *Opt. Spectros. (USSR)*, **53**, 252–257 (1982).
196. H. Hoshino, T. Yotsuyanagi, 'Reversed-Phase Partition High-Performance Liquid-Chromatography of Trace Amount of Aluminum with '8-Quinolinol-Diethyldithiocarbamate System', *Bunseki Kagaku*, **29**, 807–808 (1980).
197. K. Matsuoka, M. Suzuki, C. Honda, K. Endo, Y. Moroi, 'Micellization of Conjugated Chenodeoxy-and Ursodeoxycholates and Solubilization of Cholesterol into their Micelles: Comparison with Other Four Conjugated Bile Salts Species', *Chem. Phys. Lipids*, **139**, 1–10 (2006).
198. S.D. Zucker, J. Storch, M.L. Zeidel, J.L. Gollan, 'Mechanism of the Spontaneous Transfer of Unconjugated Bilirubin between Small Unilamellar Phosphatidylcholine Vesicles', *Biochemistry*, **31**, 3184–3192 (1992).
199. K.L. Puranam, S. Raghothama, P. Balaram, 'A H^1NMR Study of Bilirubin IXα Solubilization by Cholate Micelles – Application of Nuclear Overhauser Effects', *Biochim. Biophys. Acta*, **922**, 67–77 (1987).
200. T. Hino, Z. Yang, H. Takeuchi, T. Niwa, T. Tanaka, Y. Kawashima, H. Kojima, 'Molecular Interaction of Baicalein and Phosphatidylcholine in Liposomal Membrane', *J. Colloid Interface Sci.*, **160**, 483–486 (1993).
201. R. Almog, R.A. Saulsbery, 'The Solubility of a Dye-Detergent Complex in Phospholipid Vesicles', *J. Colloid Interface Sci.*, **159**, 328–334 (1993).

202. P. Greenspan, S.D. Fowler, 'Spectrofluorometric Studies of the Lipid Probe, Nile Red', *J. Lipid Res.*, **26**, 781–789 (1985).

203. N.R. Calafato, G. Pico, 'Griseofulvin and Ketoconazole Solubilization by Bile Salts Studied Using Fluorescence Spectroscopy', *Colloids Surf., B*, **47**, 198–204 (2006).

204. B. Sarkar, U. Das, S. Bhattacharyya, S.K. Bose, 'The Quenching of Lumichrome Fluorescence by β-Cyclodextrin – Evidence for Inclusion Complex', *Bull. Chem. Soc. Jpn.*, **68**, 1807–1809 (1995).

205. M. Sbai, S.A. Lyazidi, D.A. Lerner, B. del Castillo, M.A. Martin, 'Modified β-Cylclodextrins as Enhancers of Fluorescence Emission of Carbazole Alkaloid Derivatives', *Anal. Chim. Acta*, **303**, 47–55 (1995).

206. I.R. Politzer, K.T. Crago, D.L. Kiel, T. Hampton, 'The Effects of β-Cyclodxtrins as Enhancers of Fluorescence, UV Absorption and Solubility of Selected Bimanes in Aqueous Solutions', *Anal. Lett.*, **22**, 1567–1580 (1989).

207. S. Hashimoto, J.K. Thomas, 'Fluorescence Study of Pyrene and Naphthalene in Cyclodextrin-Amphiphile Complex Systems', *J. Am. Chem. Soc.*, **107**, 4655–4662 (1985).

208. N. Sadlej-Sosnowska, 'Fluorimetric Determination of Association Constants of Three Estrogens with Cyclodextrins', *J. Fluoresc.*, **7**, 195–200 (1997).

209. I. Sanemasa, J.S. Wu, K. Toda, 'Cyclodextrin Precipitates with Normal Alcohols in an Aqueous Medium and their Stability on Being Left Standing in Air', *Bull. Chem. Soc. Jpn.*, **70**, 371–375 (1997).

210. I. Sanemasa, J.S. Wu, K. Toda, 'Solubility Product and Solubility of Cyclodextrin Inclusion Complex Precipitates in an Aqueous Medium', *Bull. Chem. Soc. Jpn.*, **70**, 365–369 (1997).

211. M.D. Veiga, M. Merino, M. Cirri, F. Maestrelli, P. Mura, 'Comparative Study on Triclosan Interactions in Solution and in the Solid State with Natural and Chemically Modified Cyclodextrins', *J. Inclusion Phenom. Macrocyclic Chem.*, **53**, 77–83 (2005).

212. N. Micali, V. Villari, A. Mazzaglia, L.M. Scolaro, A. Valerio, A. Rencurosi, L. Lay, 'Cyclodextrin Nanoaggregates and their Assembly with Protein: A Spectroscopic Investigation', *Nanotechnology*, **17**, 3239–3244 (2006).

213. N. Nasongkla, A.F. Wiedmann, A. Bruening, M. Beman, D. Ray, W.G. Bornmann, D.A. Boothman, J.M. Gao, 'Enhancement of Solubility and Bioavailability of Beta-Lapachone Using Cyclodextrin Inclusion Complexes', *Pharm. Res.*, **20**, 1626–1633 (2003).

214. E. Damiani, R. Tursilli, A. Casolari, P. Astolfi, L. Greci, S. Scalia, 'Effect of Complexation with Randomly Methylated β-Cyclodextrin on the Aqueous Solubility, Photostability and Antioxidant Activity of an Indolinonic Nitroxide Radical', *Free Radic. Res.*, **39**, 41–49 (2005).

215. N.R. Sims, P.R. Carnegie, 'Use of Fluorescamine for the Estimation of Protein in the Presence of the Detergent Triton X-100', *Anal. Biochem.*, **65**, 578–580 (1975).

216. I. Roberts, 'Hydrolysis of 4-Methylumbelliferyl Butyrate: A Convenient and Sensitive Fluorescent Assay for Lipase Activity', *Lipids*, **20**, 243–247 (1985).

217. T. Perez-Ruiz, C. Martinez-Lozano, V. Tomas, J. Martin, 'Flow-Injection Fluorimetric Determination of Vitamin K_1 Based on a Photochemical Reaction', *Talanta*, **50**, 49–56 (1999).

218. W.L. Hinze, H.N. Singh, Z.S. Fu, R.W. Williams, D.W. Kippenberger, M.D. Morris, F.S. Sadek, 'Micelle Mediated Methodologies for the Preconcentration and Separation of Polycyclic Aromatic Compounds', in *Chemical Analysis of Polycyclic Aromatic Compounds*, ed. T. Vo-Dinh, John Wiley & Sons, New York, 151–169, Chapter 5, 1989.

219. A. Lopez Garcia, E. Blanco Gonzalez, J.I. Garcia Alonso, A. Sanz-Medel, 'Potential of Micelle-Mediate Procedures in the Sample Preparation Steps for the Determination of Polynuclear Aromatic Hydrocarbons in Waters', *Anal. Chim. Acta*, **264**, 241–248 (1992).

220. E.R. Brouwer, A.N.J. Hermans, H. Lingeman, U.A.T. Brinkman, 'Determination of Polycyclic Aromatic Hydrocarbons in Surface Water by Column Liquid Chromatography with Fluorescence Detection Using On-Line Micelle-Mediated Sample Preparation', *J. Chromatogr. A*, **669**, 45–57 (1994).

221. C. Garcia Pinto, J.L. Perwez Pavon, B. Moreno Cordero, 'Cloud Point Preconcentration and High Performance Liquid Chromatographic Determination of Polycyclic Aromatic Hydrocarbons with Fluorescence Detection', *Anal. Chem.*, **66**, 874–881 (1994).

222. S. Rubio-Barroso, M. Nour Kayali, L.M. Polo-Diez, 'Fluorimetric Study of Polycyclic Aromatic Hydrocarbons in Brij-35 Micellar Solution. Evaluation of Polycyclic Aromatic Hydrocarbons in Air Samples', *Anal. Chim. Acta*, **283**, 304–308 (1993).

223. V. Pino, J.H. Ayala, A.M. Afonso, V. Gonzalez, 'Determination of Polycyclic Aromatic Hydrocarbons in Seawater by High-Performance Liquid Chromatography with Fluorescence Detection Following Micelle-Mediated Preconcentration', *J. Chromatogr. A*, **949**, 291–299 (2002).

224. V.A. Fried, M.E. Ando, A.J. Bell, 'Protein Quantitation at the Picomole Level: An o-Phthaldialdehyde-PreTSK Column Derivatization Assay', *Anal. Biochem.*, **146**, 271–276 (1985).

225. M. Fujita, H. Kanbara, K. Murakawa, 'Non-Ionic Surfactant for Preventing DNA Adsorption to Containers', Jpn. Kokai Tokkyo Hoho, JP 02 268682, 1989; *Chem. Abstr*, **115**, 2520m.

226. M. Kimura, T. Mitani, H. Watanabe, 'Prevention of Adsorption of α, β, γ, δ-Tetrakis(1-Methylpyridinium-4-yl)Porphine and its Cadmium(II) Chelate on Glassware Surface', *Bunseki Kagaku*, **35**, 556–558 (1986).

227. T. Yorozu, M. Hoshino, M. Imamura, 'Fluorescence Studies of Pyrene Inclusion Complexes With α,β, and γ-Cyclodextrins in Aqueous Solutions. Evidence for Formation of Pyrene Dimer in γ-Cyclodextrin Cavity', *J. Phys. Chem.*, **86**, 4426–4429 (1982).

228. R.A. Agbaria, E. Roberts, I.M. Warner, 'Excimer Fluorescence of Trans-Stilbene in Ternary Aqueous Solutions of γ-Cyclodextrin', *J. Phys. Chem.*, **99**, 10056–10060 (1995).

229. J. Georges, S. Ghazarian, 'Study of Europium-Sensitized Fluorescence of Tetracycline in a Micellar Solution of Triton X-100 by Fluorescence and Thermal Lens Spectrometry', *Anal. Chim. Acta*, **276**, 401–409 (1993).

230. E.M.S. Castanheira, A.M.R. Pinto, M. Queiroz, 'Fluorescence of a Benzothienopyridopyrimidone in Solution and in Lipid Vesicles', *J. Fluoresc.*, **16**, 251–257 (2006).

231. M. Almgren, F. Grieser, J.K. Thomas, 'Rate of Exchange of Surfactant Monomer Radicals and Long-Chain Alcohols between Micelles and Aqueous Solutions', *J. Chem. Soc., Faraday Trans. I*, **75**, 1674–1687 (1979).

232. C. Ju, C. Bohne, 'Dynamics of Probe Complexation to Bile Salt Aggregates', *J. Phys. Chem.*, **100**, 3847–3854 (1996).

233. V.C. Reinsborough, J.F. Holzwarth, 'Kinetics of the Interactions between Dyes and Micelles', *Can. J. Chem.*, **64**, 955–959 (1986).

234. A.D. James, B.H. Robinson, 'Micellar Catalysis of Metal Complex Formation – Kinetics of Reaction between N(II) and Pyridine-2-Azo-p-Dimethylaniline (PADA) in Presence of Sodium Dodecylsulfate Micelles – Model System for Study of Metal-Ion Reactivity at Charged Interfaces', *J. Chem. Soc., Faraday Trans. I*, **74**, 10–21 (1978).

235. H. Bruhn, J.F. Holzwarth, 'Catalytic Influence of the Environment on Outer-Sphere Electron-Transfer Reactions in Aqueous Solutions', *Faraday Discuss. Chem. Soc.*, **74**, 129–140 (1982).

236. R. Zana, 'Dynamics of Surfactant Self-Assemblies', in *Dynamics in Micellar Solutions of Surfactants*, *Surfactant Science Series, Vol. 125*, ed. R. Zana, Taylor & Francis, New York, 75–160, 2005.

237. M. Novo, S. Felekyan, C.A.M. Seidel, W. Al-Soufi, 'Dye-Exchange Dynamics in Micellar Solutions Studied by Fluorescence Correlation Spectroscopy', *J. Phys. Chem. B*, **111**, 3614–3624 (2007).

238. T.C. Barros, K. Stefaniak, J.F. Holzwarth, C. Bohne, 'Complexation of Naphthylethanols with β-Cyclodextrin', *J. Phys. Chem. A*, **102**, 5639–5651 (1998).

239. E. Okamura, M. Nakahara, 'NMR Study Directly Determining Drug Delivery Sites in Phospholipid Bilayer Membranes', *J. Phys. Chem. A*, **103**, 3505–3509 (1999).

240. D.M. Shin, D.G. Whitten, 'Solvatochromic Behavior of Intramolecular Charge-Transfer Diphenylpolyenes in Homogeneous Solution and Microheterogeneous Media', *J. Phys. Chem.*, **92**, 2945–2956 (1988).

241. M. Takezaki, T. Tominaga, 'Fluorescence Quenching Reaction of Porphyrins in Micelles: Ionic Porphyrins Quenched by Nitrobenzene in Ionic Micelles', *J. Photochem. Photobiol., A*, **174**, 113–118 (2005).

242. B.R. Suddaby, P.E. Brown, J.C. Russell, D.G. Whitten, 'Photochemical Reactions in Organized Assemblies. 44. Surfactant and Hydrophobic Derivatives of Trans-Stilbenes as Probes of Vesicle and Micelle Solubilization Sites: Studies Using Fluorescence and Photoisomerization as Probes', *J. Am. Chem. Soc.*, **107**, 5609–5617 (1985).

243. L. Martin, A.I. Olives, B. del Castillo, M.A. Martin, 'Fluorescence Quenching of β-Carboline Alkaloids in Micellar Media. A Study to Select the Adequate Surfactant to Use in Analytical Techniques', *Luminescence*, **20**, 152–161 (2005).

244. Y. Jiang, X. Huang, 'Fluorescence Spectroscopic Determination of Formation Constant of Inclusion Complexes of Cyclodextrin', *Huaxue Tongbao*, **9**, 46–49 (1990).

245. M. de La Guardia, E. Peris-Cardells, J. Sancenon, J.L. Carrion, E. Pramauro, 'Fluorimetric Determination of Binding Constants between Micelles and Chemical Systems', *Microchem. J.*, **44**, 193–200 (1991).

246. C. Wang, J. Chen, X. Zhang, G. Xu, Y. Tang, 'Distribution of Molecules in Micelle Solutions', *Wuli Huaxue Xuebao*, **3**, 440–444 (1987).

247. Y. Jiang, X. Huang, G. Chen, 'Determination of Dissociation Constants of Cyclodextrin Inclusion Complex with Spectrofluorometry', *Xiamen Dauxue Xuebao Ziran Kexueban*, **29**, 188–192 (1990).

248. E. Pramauro, A.B. Prevot, 'Solubilization in Micellar Systems – Analytical and Environmental Applications', *Pure Appl. Chem.*, **67**, 551–559 (1995).

249. S. Pandey, R.S. Sarpal, S.K. Dogra, 'Luminescence Characteristics of 5-Aminoindazole in Cationic and Anionic Micelles at Various pH', *J. Colloid Interface Sci.*, **172**, 407–414 (1995).

250. I.V. Berezin, K. Martinek, A.K. Yatsimirskii, 'Physicochemical Foundations of Micellar Catalysis', *Russ. Chem. Rev.*, **42**, 787 (1973).

251. M. Fujiki, T. Deguchi, I. Sanemasa, 'Association of Naphthalene and its Methyl-Derivatives with Cyclodextrins in Aqueous Medium', *Bull. Chem. Soc. Jpn.*, **61**, 1163–1167 (1988).

252. E. Iglesias, V. Ojea-Cao, L. Garcia-Rio, J.R. Leis, 'Effects of β-Cyclodextrin on the Keto-Enol Equilibrium of Benzoylacetone and on Enol Reactivity', *J. Org. Chem.*, **64**, 3954–3963 (1999).

253. M.V. Rekharsky, Y. Inoue, 'Complexation Thermodynamics of Cyclodextrins', *Chem. Rev.*, **98**, 1875–1918 (1998).

254. Y. Wu, K. Ishibashi, T. Beguchi, I. Sanemasa, 'The Solubility and Cyclodextrin Complexations of Fluorene, Fluoranthene and Pyrene in Aqueous Medium', *Bull. Chem. Soc. Jpn.*, **63**, 3450–3455 (1990).

255. F.H. Quina, E.O. Alonso, J.P.S. Farah, 'Incorporation of Nonionic Solutes into Aqueous Micelles: A Linear Solvation Free Energy Relationship Analysis', *J. Phys. Chem.*, **99**, 11708–11714 (1995).

256. H. Sugioka, Y. Moroi, 'Micelle Formation of Sodium Cholate and Solubilization into Micelle', *Biochim. Biophys. Acta*, **1394**, 9–110 (1998).

257. G. Nelson, G. Patonay, I.M. Warner, 'The Utility of Time-Resolved Emission Spectroscopy in the Study of Cyclodextrin-Pyrene Inclusion Complexes', *Talanta*, **36**, 199–203 (1989).

258. M. Almgren, 'Migration and Partitioning of Pyrene and Perylene between Lipid Vesicles in Aqueous Solution Studied with a Fluorescence Stopped-Flow Technique', *J. Am. Chem. Soc.*, **102**, 7882–7887 (1980).

259. C.H. Evans, T. Gunnlaugsson, 'Modification of an Intermolecular Photoreaction by Cyclodextrin: The Photohydroxylation of 3-Nitroanisole', *J. Photochem. Photobiol., A: Chem.*, **78**, 57–62 (1994).

260. A.C. Tedesco, R.M.Z.G. Naal, J.C. Carreiro, J.B.S. Bonilha, 'Kinetic Studies of the Nucleophilic Quenching of Photoexcited Nitroaryl Ethers (Nare) by Bromide Ions in Cationic Micellar Solution', *Tetrahedron*, **50**, 3071–3080 (1994).

261. L. Liu, X.-Q. Guo, 'Novel Prediction for the Driving Force and Guest Orientation in the Complexation of α- and β-Cyclodextrin with Benzene Derivatives', *J. Phys. Chem.*, **103**, 3461–3467 (1999).

262. G.M. Escandar, 'Spectroscopic Study of Salicylate-Cyclodextrin Systems in the Presence and Absence of Alcohols', *Spectrochim. Acta, Part A*, **55**, 1743–1752 (1999).

263. G.S. Cox, N.J. Turro, 'Methyl Salicylate Fluorescence as a Probe of the Geometry of Complexation to Cyclodextrins', *Photochem. Photobiol.*, **40**, 185–188 (1984).

264. W. Eli, W. Chen, Q. Xue, 'The Association of Anionic Surfactants with β-Cyclodextrin. An Isothermal Titration Calorimeter Study', *J. Chem. Thermodyn.*, **31**, 1283–1296 (1999).

265. J.W. Park, H.J. Song, 'Association of Anionic Surfactants With β-Cyclodextrin: Fluorescence Probe Studies on 1 : 1 and 1 : 2 Complexation', *J. Phys. Chem.*, **93**, 6454–6458 (1989).

266. E. Blatt, 'The Association Properties of Iodide with Cetyltrimethylammonium Bromide Micelles as Revealed by Steady-State Fluorescence Quenching Measurements', *Aust. J. Chem.*, **40**, 201–207 (1987).

267. E. Lissi, E. Abuin, A.M. Rocha, 'Partitioning of Aromatic Alcohols Between a Water Solution and Cationic Micelles', *J. Phys. Chem.*, **84**, 2406–2408 (1980).

268. M. Hoshino, M. Imamura, K. Ikehara, Y. Hama, 'Fluorescence Enhancement of Benzene Derivatives by Forming Inclusion Complexes With β-Cyclodextrin in Aqueous Solutions', *J. Phys. Chem.*, **85**, 1820–1823 (1981).

269. S. Monti, G. Koehler, G. Grabner, 'Photophysics and Photochemistry of Methylated Phenols in β-Cyclodextrin Inclusion Complexes', *J. Phys. Chem.*, **97**, 13011–13016 (1993).

270. H. Fujiwara, Y.Z. Da, K. Ito, 'The Energy Aspect of Oil-Water Partition and its Application to the Analysis of Quantitative Structure – Activity Relationships. Aliphatic Alcohols in the Liposome Water/Partition System', *Bull. Chem. Soc. Jpn.*, **64**, 3707–3712 (1991).

271. H.J.E.M. Reeuwijk, H. Irth, U.R. Tjaden, F.W.H.M. Merkus, J. van der Greef, 'Liquid Chromatographic Determination of β-Cyclodextrin Derivatives Based on Fluorescence Enhancement after Inclusion Complexation', *J. Chromatogr.*, **614**, 95–101 (1993).

272. D. Wu, R.J. Hurtubise, 'Fluorescence Properties of 1-Naphthol, 2-Naphthol and 1,2,3,4-Tetrahydronaphthol in Aqueous Alcohol Solvents with and without β-Cyclodextrin', *Talanta*, **40**, 901–907 (1993).

273. S.W. Snyder, J.N. Demas, B.A. Degraff, 'Solvent Accessibility of Ruthenium(II) Photosensitizers with Cetyltrimethylammonium Bromide', *Chem. Phys. Lett.*, **145**, 434–438 (1988).

274. C. Yuan, C.H. Hung, C.T. Jafvert, 'Groundwater Remediation by Anionic Surfactant Micelles – An Innovative Double Layer Model Applied to Na^+ and Mg^{2+} Association with Dodecylsulfate Micelles', *Water Sci. Technol.*, **38**, 99–106 (1998).

275. J.R. Escabi-Perez, F. Nome, J.H. Fendler, 'Energy Transfer in Micellar Systems. Steady State and Time Resolved Luminescence of Aqueous Micelle Solubilized Naphthalene and Terbium Chloride', *J. Am. Chem. Soc.*, **99**, 7749–7754 (1977).

276. S.M. Meyerhoffer, T.J. Wenzel, L.B. McGown, 'Lanthanide Shift NMR Studies of Bile Salt Aggregates', *J. Phys. Chem.*, **96**, 1961–1967 (1992).

277. N. Fatin-Rouge, J.-C.G. Bunzli, 'Thermodynamic and Structural Study of Inclusion Complexes between Trivalent Lanthanide Ions and Native Cyclodextrins', *Inorg. Chim. Acta*, **293**, 53–60 (1999).

278. N. Kamo, T. Aiuchi, K. Kurihara, Y. Kobatake, 'The Quantitative Analysis of the Interaction of 1-Anilinonaphthalene-8-Sulfonate with Liposomes: Fluorescence Intensity and Zeta-Potential', *Colloid Polym. Sci.*, **256**, 31–36 (1978).

279. H.J. Schneider, T. Blatter, S. Simova, 'Host Guest Chemistry. 26. NMR and Fluorescence Studies of Cyclodextrin Complexes with Guest Molecules Containing Both Phenyl and Naphthyl Units', *J. Am. Chem. Soc.*, **113**, 1996–2000 (1991).

280. G.C. Catena, F.V. Bright, 'Thermodynamic Study on the Effects of β-Cyclodextrin Inclusion with Anilinonaphthalenesulfonates', *Anal. Chem.*, **61**, 905–909 (1989).

281. J.Y. Ma, J.K. Ma, K.C. Weber, 'Fluorescence Studies of the Binding of Amphiphilic Amines with Phospholipids', *J. Lipid Res.*, **26**, 735–744 (1985).

282. J. Dufourcq, J.F. Faucon, 'Specific Binding of a Cardiotoxin from Naja Mossambica Mossambica to Charged Phospholipids Detected by Intrinsic Fluorescence', *Biochemistry*, **17**, 1170–1176 (1978).

283. K. Kano, I. Takenoshita, T. Ogawa, 'Fluorescence Quenching of Pyrene and Naphthalene in Aqueous Cyclodextrin Solutions. Evidence of Three-Component Complex Formation', *J. Phys. Chem.*, **86**, 1833–1838 (1982).

284. E. Perochon, A. Lopez, J.F. Tocanne, 'Polarity of Lipid Bilayers. A Fluorescence Investigation', *Biochemistry*, **31**, 7672–7682 (1992).

285. N. Husain, T.T. Ndou, A.M. Delapena, I.M. Warner, 'Complexation of Doxorubicin with β-Cyclodextrins and γ-Cyclodextrins', *Appl. Spectrosc.*, **46**, 652–658 (1992).

286. X.-M. Wang, H.Y. Chen, 'Investigation of the β-Cyclodextrin-Quinine Inclusion Complex in Aqueous Solution by Spectroscopic Study', *Spectrochim. Acta, Part A*, **51**, 333–339 (1995).

287. M.A. Schwarz, R.H.H. Neubert, H.H. Ruttinger, 'Application of Capillary Electrophoresis for Characterizing Interactions between Drugs and Bile Salts. Part I', *J. Chromatogr. A*, **745**, 135–143 (1996).

288. Y.-B. Jiang, 'Twisted Intramolecular Charge Transfer of Methyl p-Dimethylaminobenzoate in Aqueous β-Cyclodextrin Solution', *Spectrochim. Acta, Part A*, **51**, 275–282 (1995).

289. E. Junquera, V.G. Baonza, E. Aicart, 'Energetics of the Encapsulation of o-, m-, and p-Hydroxybenzoic Acids by β-Cyclodextrin and its Methylated and Hydroxypropylated Derivatives in Aqueous Solution', *Can. J. Chem.*, **77**, 348–355 (1999).

290. L. Flamigni, 'Inclusion of Fluorescein and Halogenated Derivatives in α-, β-, and γ-Cyclodextrins: A Steady-State and Picosecond Time-Resolved Study', *J. Phys. Chem.*, **97**, 9566–9572 (1993).

291. W. Reed, M.J. Politi, J.H. Fendler, 'Rotational Diffusion of Rose Bengal in Aqueous Micelles: Evidence for Extensive Exposure of the Hydrocarbon Chains', *J. Am. Chem. Soc.*, **103**, 4591–4593 (1981).

292. E.L. Quitevis, A.H. Marcus, M.D. Fayer, 'Dynamics of Ionic Lipophilic Probes in Micelles: Picosecond Fluorescence Depolarization Measurements', *J. Phys. Chem.*, **97**, 5762–5769 (1993).

293. T. Soujanya, T.S.R. Krishna, A. Samanta, 'Effect of β-Cyclodextrin on Intramolecular Charge-Transfer Emission of 4-Aminophthalimide', *J. Photochem. Photobiol., A*, **66**, 185–192 (1992).

294. G. Saroja, A. Samanta, 'Photophysical Studies on a Fluorescence Probe Labelled Fatty Acid: Chain Folding in a Micellar Environment', *J. Chem. Soc., Faraday Trans.*, **92**, 2697–2701 (1996).

295. L.R. De Young, K.A. Dill, 'Solute Partitioning into Lipid Bilayer Membranes', *Biochemistry*, **27**, 5281–5289 (1988).

296. K.A. Connors, 'Measurement of Cyclodextrin Stability Constants', in *Comprehensive Supramolecular Chemistry*, *Cyclodextrins, Vol. 3*, eds. J. Szejtli, T. Osa, Pergamon Press, New York, 205–241, Chapter 6, 1996.

297. T.G. Burke, M. Israel, R. Seshadri, J.H. Doroshow, 'A Fluorescence Study Examining How 14-Valerate Side Chain Substitution Modulates Anthracycline Binding to Small Unilamellar Phospholipid Vesicles', *Biochim. Biophys. Acta*, **982**, 123–130 (1989).

298. A. Pezeshk, V. Pezeshk, J. Wojas, W.K. Subczynski, W.K. Subczynski, 'Partitional and Motional Properties of a Spin-Labeled Daunomycin in Lipid Bilayers. An ESR Study', *J. Inorg. Biochem.*, **46**, 67–76 (1992).

299. H.Y. Cheng, C.S. Randall, W.W. Holl, P.P. Constantinides, T.L. Yue, G.Z. Feuerstein, 'Carvedilol-Liposome Interaction: Evidence for Strong Association with the Hydrophobic Region of the Lipid Bilayer', *Biochim. Biophys. Acta*, **1284**, 20–28 (1996).

300. J. Wang, I.M. Warner, 'Studies of the Naproxen: β-Cyclodextrin Inclusion Complex', *Microchem. J.*, **48**, 229–239 (1993).

301. J.S. Wu, J.Z. Zheng, K. Toda, I. Sanemasa, 'Association of Alcohol-Cyclodextrin in Aqueous Medium Determined by Headspace Gas Chromatography', *Anal. Sci.*, **15**, 701–703 (1999).

302. C.H. Spink, S. Colgan, 'Thermodynamics of Alcohol Binding to Deoxycholate Micelles and the Hydrophobic Effect', *J. Phys. Chem.*, **87**, 888–894 (1983).

303. S. Hamai, '1-1-1 Inclusion-Compounds of Beta-Cyclodextrin with Fluorene and Alcohols or Nitriles in Aqueous-Solution', *Bull. Chem. Soc. Jpn.*, **62**, 2763–2767 (1989).

304. A. Orstan, J.B.A. Ross, 'Investigation of The β-Cyclodextrin-Indole Inclusion Complex by Absorption

and Fluorescence Spectroscopies', *J. Phys. Chem.*, **91**, 2739–2745 (1987).

305. K.D. Ashby, K. Das, J.W. Petrich, 'The Effect of Micelles on the Steady-State and Time-Resolved Fluorescence of Indole, 1-Methylindole, and 3-Methylindole in Aqueous Media', *Anal. Chem.*, **69**, 1925–1930 (1997).

306. R.S. Sarpal, M. Belletete, G. Durocher, 'Fluorescence Probing and Proton-Transfer Equilibrium Reactions in Water, SDS, and CTAB Using 3,3-Dimethyl-2-Phenyl-3H-Indole', *J. Phys. Chem.*, **97**, 5007–5013 (1993).

307. A. González, R.M. Linares, J.H. Ayala, A.M. Afonso, V. González, 'Effects of Cationic Micelles on Fluorescence of Indole and Indolecarboxylic Acids. Analytical Determinations', *Microchim. Acta*, **118**, 153–162 (1995).

308. A.G. Mwalupindi, A. Rideau, R.A. Agbaria, I.M. Warner, 'Influence of Organized Media on the Absorption and Fluorescence Spectra of Auramine-O Dye', *Talanta*, **41**, 599–609 (1994).

309. J. Gu, J. Pan, 'Determination of the Cyclodextrin Inclusion Constant with the Constant Current Coulometric Titration Method', *Talanta*, **50**, 35–39 (1999).

310. S.R.W. Louro, O.R. Nascimento, M. Tabak, 'Charge and pH-Dependant Binding Sites for Dibucaine in Ionic Micelles: A Fluorescence Study', *Biochim. Biophys. Acta* **1190**, 319–328 (1994).

311. M. Khopde, Sujata Priyadarsini, Indira K. Venkatesan, M.N.A.P. Rao, 'Free Radical Scavenging Ability and Antioxidant Efficiency of Curcumin and its Substituted Analogue', *Biophys. Chem.*, **80**, 83–89 (1999).

312. Z.-T. Jiang, R. Li, J.-B. Xi, B.Q. Yi, 'Determination of Trace Amounts of Manganese by β-Cyclodextrin Polymer Solid Phase Spectrophotometry Using 2-(5-Bromo-2-Pyridylazo)-5-Diethylaminophenol', *Anal. Chim. Acta*, **392**, 247–253 (1999).

313. P.K. Zarzycki, H. Lamparczyk, 'The Equilibrium Constant of β-Cyclodextrin-Phenolphtalein Complex; Influence of Temperature and Tetrahydrofuran Addition', *J. Pharm. Biomed. Anal.*, **18**, 165–170 (1998).

314. Z. Yuan, M. Zhu, S. Han, 'Supramolecular Inclusion Complex Formation and Application of β-Cyclodextrin with Heteroanthracene Ring Cationic Dyes', *Anal. Chim. Acta*, **389**, 291–298 (1999).

315. G.-C. Zhao, J.-J. Zhu, H.-Y. Chen, 'Spectroscopic Studies of the Interactive Model of Methylene Blue with DNA by Means of β-Cyclodextrin', *Spectrochim. Acta, Part A*, **55**, 1109–1117 (1999).

316. M.B. Levin, M.I. Snegov, A.S. Cherokasov, 'Determination of the Rate Constants of Molecular Processes Regulating the Level of Induced Absorption in a Laser Based on an Aqueous Micellar Solution of Rhodamine 6G with Lamp Pumping', *Opt. Spectros. (USSR)*, **62**, 341–344 (1987).

317. R.S. Murphy, T.C. Barros, J. Barnes, B. Mayer, G. Marconi, C. Bohne, 'Complexation of Fluorenone and Xanthone to Cyclodextrins: Comparison of Theoretical and Experimental Studies', *J. Phys. Chem. A*, **103**, 137–146 (1999).

318. A. Berthod, C.G. Alvarez-Coque, *Micellar Liquid Chromatography*, Marcel Dekker, New York, 2000, 503–590.

319. K.A. Connors, 'Prediction of Binding Constants of α-Cyclodextrin Complexes', *J. Pharm. Sci.*, **85**, 796–802 (1996).

320. M. Shannigrahi, S. Bagchi, 'Use of Fluorescence Probes for Characterization of Solvation Properties of Micelles: A Linear Solvation Energy Relationship Study', *J. Phys. Chem. B*, **108**, 17703–17708 (2004).

321. K.A. Connors, 'Population Characteristics of Cyclodextrin Complex Stabilities in Aqueous Solution', *J. Pharm. Sci.*, **84**, 843–848 (1995).

322. R. Verrone, L. Catucci, P. Cosma, P. Fini, A. Agostiano, V. Lippolis, M. Pascale, 'Effect of Beta-Cyclodextrin on Spectroscopic Properties of Ochratoxin A in Aqueous Solution', *J. Inclusion Phenom. Macrocyclic*, **57**, 475–479 (2007).

323. S. Yamada, K. Shibata, A. Murata, 'Spectrofluorometric Study on the Inclusion Reaction of Flavonols with β-Cyclodextrin', *Anal. Sci.*, **9**, 467–471 (1993).

324. T. Yorozu, M. Hoshino, M. Imamura, H. Shizuka, 'Photoexcited Inclusion Complexes of β-Naphthol With α-, β-, and γ-Cyclodextrins in Aqueous Solutions', *J. Phys. Chem.*, **86**, 4422–4426 (1982).

325. I.X. Garcia-Zubiri, G. Gonzalez-Gaitano, J.R. Isasi, 'Application of Automated Docking to the Binding of Naphthalenes to β-Cyclodextrin in Water: Correlation with Spectrofluorimetric Data', *J. Inclusion Phenom. Macrocyclic*, **57**, 265–270 (2007).

326. A. Nakajima, 'A Study of the System of Pyrene and Beta-Cyclodextrin in Aqueous-Solution Utilizing the Intensity Enhancement Phenomenon', *Spectrochim. Acta, Part A*, **39**, 913–915 (1983).

327. W.R. Bergmark, A. Davis, C. York, A. Macintosh, G. Jones, 'Dramatic Fluorescence Effects for Coumarin Laser Dyes Coincluded with Organic Solvents in Cyclodextrins', *J. Phys. Chem.*, **94**, 5020–5022 (1990).

328. A. Ueno, K. Takahashi, Y. Hino, T. Osa, 'Fluorescence Enhancement of α-Naphthyloxyacetic Acid in the Cavity of γ-Cyclodextrin, Assisted by a Space-Regulating Molecule', *J. Chem. Soc., Chem. Commun.*, 194–195 (1981).

329. K. Kano, I. Takenoshita, T. Ogawa, 'γ-Cyclodextrin Enhanced Excimer Fluorescence of Pyrene and Effect of Normal-Butyl Alcohol', *Chem. Lett.*, 321–324 (1982).

330. Y.H. Zhou, G.M. Zhang, Z.R. Wang, H.Y. Wang, C. Dong, S.M. Shuang, 'The Interaction of Piroxicam with Neutral (Hydroxypropyl) and Anionically Charged (Sulfated) Beta-Cyclodextrin', *J. Inclusion Phenom. Macrocyclic Chem.*, **56**, 215–220 (2006).

331. M. Shannigrahi, S. Bagchi, 'Time Resolved Fluorescence Study of Ketocyanine Dye-Beta Cyclodextrin Interactions in Aqueous and Non-Aqueous Media', *Chem. Phys. Lett.*, **403**, 55–61 (2005).

332. M.M. Al Omari, M.B. Zughul, J.E.D. Davies, A.A. Badwan, 'Effect of Buffer Species on the Inclusion Complexation of Acidic Drug Celecoxib with Cyclodextrin in Solution', *J. Inclusion Phenom. Macrocyclic Chem.*, **55**, 247–254 (2006).

333. Z.P. Yi, C.C. Zhao, Z.Z. Huang, H.L. Chen, J.S. Yu, 'Investigation of Buffer-Cyclodextrin Systems', *Phys. Chem. Chem. Phys.*, **1**, 441–444 (1999).

334. P. Nambi, E.S. Rowe, T.J. McIntosh, 'Studies of the Ethanol-Induced Interdigitated Gel Phase in Phosphatidylcholines Using the Fluorophore 1,6-Diphenyl-1,3,5-Hexatriene', *Biochemistry*, **27**, 9175–9182 (1988).

335. M. Langner, S.W. Hui, 'Merocyanine Interaction with Phosphatidylcholine Bilayers', *Biochim. Biophys. Acta*, **1149**, 175–179 (1993).

336. M.K. Kawamuro, H. Chaimovich, E.B. Abuin, E.A. Lissi, I.M. Cuccovia, 'Evidence That the Effects of Synthetic Amphiphile Vesicles on Reaction Rates Depend on Vesicle Size', *J. Phys. Chem.*, **95**, 1458–1463 (1991).

337. P.R. Sainz-Rozas, J.R. Isasi, G. Gonzalez-Gaitano, 'Binding of Dibenzofuran and its Derivatives to Water-Soluble Beta-Cyclodextrin Polymers', *J. Photochem. Photobiol., A*, **173**, 248–257 (2005).

338. J.J. Fisz, 'Fluorescence Depolarization in Macroscopically Ordered Uniaxial Molecular Samples. II. The Ground-and Excited-State Interactions of Fluorophores with Ordered Environment. A Discussion of the Published Experimental Data', *Chem. Phys.*, **132**, 303–314 (1989).

339. E. Prenner, A. Hermetter, G. Landl, H. Stuetz, H.F. Kauffmann, A.J. Kungl, 'Fluorescence Lifetime Distributions of Various Phospholipids Labeled with Covalently Bound Diphenylhexatriene in the Erythrocyte Ghost Membrane', *J. Phys. Chem.*, **97**, 2788–2792 (1993).

340. L. Davenport, J.R. Knutson, L. Brand, 'Studies of Membrane Heterogeneity Using Fluorescence Associative Techniques', *Faraday Discuss.*, 81–94 (1986).

341. D.R. James, W.R. Ware, 'A Fallacy in the Interpretation of Fluorescence Decay Parameters', *Chem. Phys. Lett.*, **120**, 455–459 (1985).

342. J. Huang, F.V. Bright, 'Microheterogeneity of Sodium Dodecylsulfate Micelles Probed by Frequency-Domain Fluorometry', *Appl. Spectrosc.*, **46**, 329–339 (1992).

343. G.C. Catena, F.V. Bright, 'Contineous Lifetime Distributions of β–Cyclodextrin Anilinonaphthalene Sulfonic Acid Inclusion Complexes', *J. Fluoresc.*, **122**, 377–381 (1991).

344. Y.L. Loukas, 'Multiple Complex Formation of Unstable Compounds with Cyclodextrins: Efficient Determination and Evaluation of the Binding Constant with Improved Kinetic Studies', *Analyst*, **122**, 377–381 (1997).

345. J.B. Chao, H.B. Tong, D.S. Liu, S.P. Huang, 'Preparation and Characterization of Inclusion Complexes of Pefloxacin Mesylate with Three Kinds of Cyclodextrins', *Spectrochim. Acta, Part A*, **64**, 166–170 (2006).

346. P. Sen, D. Roy, S.K. Mondal, K. Sahu, S. Ghosh, K. Bhattacharyya, 'Fluorescence Anisotropy Decay and Solvation Dynamics in a Nanocavity: Coumarin 153 in Methyl Beta-Cyclodextrins', *J. Phys. Chem. A*, **109**, 9716–9722 (2005).

347. U. Khuanga, B.K. Selinger, R. McDonald, 'Study of Surfactant Micelles with a Fluorescent Probe', *Aust. J. Chem.*, **29**, 1–12 (1976).

348. D.J. Miller, 'Distribution of Solubilized Molecules Among Micelles', *J. Chem. Educ.*, **55**, 776–777 (1978).

349. T. Wolff, 'External Heavy-Atom Effects of Counterions on Excited Aromatic-Molecules in Aqueous Micellar Solutions of Cetyltrimethylammonium Salts', *Ber.Bunsen-Ges. Phys. Chem.*, **86**, 1132–1134 (1982).

350. T. Dwars, E. Paetzold, G. Oehme, 'Reactions in Micellar Systems', *Angew. Chem., Int. Ed. Engl.*, **44**, 7174–7199 (2005).

351. M.K. Khan, *Micellar Catalysis*, Taylor & Francis, New York, 2007.

352. C.A. Bunton, 'Micelles: Chemical Kinetics', *Encyclopedia of Surface and Colloid Science*, 2nd edition, Taylor & Francis, New York, 3825–3839, 2006.

353. J.H. Fendler, E.J. Fendler, *Catalysis in Micellar and Macromolecular Systems*, Academic Press, New York, 1975.

354. F.H. Quina, H. Chaimovich, 'Ion Exchange in Micellar Solutions. 1. Conceptual Framework for Ion Exchange in Micellar Solutions', *J. Phys. Chem.*, **83**, 1844–1850 (1979).

355. K. Takahashi, 'Organic Reactions Mediated by Cyclodextrins', *Chem. Rev.*, **98**, 2013–2034 (1998).

356. P. Bortolus, G. Grabner, G. Kohler, S. Monti, 'Photochemistry of Cyclodextrin Host-Guest Complexes', *Coord. Chem. Rev.*, **125**, 261–268 (1993).

357. M. El Baraka, M. Deumie, P. Viallet, 'Kinetics of the Hydrolysis of Cytotoxic Pyronins in Micellar Systems', *J. Photochem. Photobiol., A*, **67**, 75–90 (1992).

358. I. Capek, 'Fate of Excited Probes in Micellar Systems', *Adv. Colloid Interface Sci.*, **97**, 91–149 (2002).

359. N.J. Turro, M. Gratzel, A.M. Braun, 'Photophysical and Photochemical Processes in Micellar Systems', *Angew. Chem., Int. Ed. Engl.*, **19**, 675–696 (1980).

360. M. Almgren, P. Hansson, 'Micelles: Fluorescence Studies', *Encyclopedia of Surface and Colloid Science*, 2nd edition, Taylor & Francis, New York, 3840–3853, 2006.

361. G. Valduga, S. Nonell, E. Reddi, G. Jori, S.E. Braslavsky, 'The Production of Singlet Molecular Oxygen by Zinc(II) Phthalocyanine in Ethanol and in Unilamellar Vesicles – Chemical Quenching and Phosphorescence Studies', *Photochem. Photobiol.*, **48**, 1–5 (1988).

362. J.K. Thomas, S. Hashimoto, 'Photochemistry in Organized and Constrained Media', *New J. Chem.*, **11**, 145–149 (1987).

363. T.J. Burkey, D. Griller, 'Micellar Systems as Devices for Enhancing the Lifetimes and Concentrations of Free-Radicals', *J. Am. Chem. Soc.*, **107**, 246–249 (1985).

364. D. Mobius, 'Photochemistry in Organized Media', *New J. Chem.*, **11**, 203–205 (1987).

365. I.R. Gould, P.L. Kuo, N.J. Turro, 'Quenching of Pyrene Fluorescence by Cesium Ions in Micellar Systems. Protection by Surface-Active Crown Ethers', *J. Phys. Chem.*, **89**, 3030–3034 (1985).

366. E. Lissi, A. Dattoli, E. Abuin, 'Deactivation of Pyrene by Indole in Micellar Solution', *Bol. Soc. Chil. Quim.*, **30**, 37 (1985).

367. K. Kano, S. Hashimoto, A. Imami, T. Ogawa, 'Three Component Complexes of Cyclodextrin: Exciplex Formation in Cyclodextrin Cavity', *J. Inclusion Phenom. Mol. Recognit. Chem.*, **2**, 737–746 (1984).

368. R.F. Turchiello, M.T. Lamy-Freund, I.Y. Hirata, L. Juliano, A.S. Ito, 'Ortho-Aminobenzoic Acid as a Fluorescent Probe for the Interaction Between Peptides and Micelles', *Biophys. Chem.*, **73**, 217–225 (1998).

369. Y. Usui, A. Gotou, 'Migration of Singlet Excitation-Energy from Acridine Orange-10-Dodecyl Bromide to Methylene-Blue in Micelles', *Photochem. Photobiol.*, **29**, 165–167 (1979).

370. W.J. Dong, C.D. Flint, 'Intermolecular Energy Transfer between Lanthanide Complex-Ions in Micellar Solution. 2. Energy-Transfer from $[Tb(Acac)_3phen]$ to Ln^{3+} Ions', *J. Chem. Soc., Faraday Trans.*, **88**, 3435–3439 (1992).

371. F. Tanaka, S. Ogura, S. Yamashita, 'Energy Transfer from the Triplet State of 2-Acetylnaphthalene to Eu^{3+} in Micellar Solutions', *J. Photochem. Photobiol., A*, **54**, 213–217 (1990).

372. I.L. Brunetti, G. Cilento, L. Nassi, 'Energy-Transfer from Enzymically-Generated Triplet Species to Acceptors in Micelles', *Photochem. Photobiol.*, **38**, 511–519 (1983).

373. S. Chang, S. Spurlin, W.L. Hinze, 'Practical Applications of Micelle Systems in Analytical Chemistry', *Science-Ciencia*, **6**, 54–56 (1978).

374. P.E. Nelson, 'High-Performance Liquid Chromatography Detection of Morphine by Fluorescence after Post-Column Derivatisation: II. The Effect of Micelle Formation', *J. Chromatogr. A*, **298**, 59–65 (1984).

375. H. Huang, R. Cai, Y. Du, Z. Lin, Y. Zeng, 'Micelle Enhanced Spectrofluorimetric Assay for Laccase Activity by a Flow-Injection Stopped-Flow Technique', *Anal. Chim. Acta*, **318**, 63–69 (1995).

376. I. Mori, Y. Fujita, M. Toyoda, K. Kato, N. Yoshida, M. Akagi, 'Fluorimetric Determination of Hydrogen Peroxide by Use of the Fluorescence Reaction Between N-(4'-Hydroxyphenyl)-N-(4-Methylquinolinyl)Amine and Cobalt(II) in the Presence of Trimethylstearylammonium Chloride', *Talanta*, **38**, 683–686 (1991).

377. L. Zhu, C. Chen, Y.Z. Li, Y.X. Ci, 'Enhancement Effect of Brij-35 on the Hemin-Catalyzed Fluorogenic Reaction between Hydrogen Peroxide and N,N'-Dicyanomethyl-o-Phenylenediamine', *Anal. Chim. Acta*, **369**, 205–213 (1998).

378. E.R. Triboni, M.J. Politi, I.M. Cuccovia, H. Chaimovich, P. Berci, 'Rate-Limiting Step and Micellar Catalysis of the Non-Classical Nitro Group Nucleophilic Substitution by Thiols in 4-Nitro-N-Butyl-1,8-Naphthalimide', *J. Phys. Org. Chem.*, **16**, 311–317 (2003).

379. F. Dai, V.P. Burkert, H.N. Singh, W.L. Hinze, 'Update and Evaluation of the Effectiveness of Different Thiols and Micellar Media in Roth's Fluorimetric Method for the Determination of Primary Amino Compounds', *Microchem. J.*, **57**, 166–198 (1997).

380. R.P. Frankewich, K.N. Thimmaiah, W.L. Hinze, 'Evaluation of the Relative Effectiveness of Different Water-Soluble β-Cyclodextrin Media to Function as Fluorescence Enhancement Agents', *Anal. Chem.*, **63**, 2924–2933 (1991).

381. A.T. Philipenko, T.A. Vasil'chuk, A.I. Volkova, 'Fluorescence of Niobium-Morin Complex in Micellar Solutions of Cationic Surfactants and in Aqueous-Ethanolic (Acetone) Solutions', *Zh. Anal. Khim.*, **38**, 855–861 (1983).

382. N.O. McHedlov-Petrossyan, N.A. Vodolazkaya, Y.N. Surov, D.V. Samoylov, '2,4,5,7-Tetranitrofluorescein in Solutions: Novel Type of Tautomerism in Hydroxyxanthene Series as Detected by Various Spectral Methods', *Spectrochim. Acta, Part A*, **61**, 2747–2760 (2005).

383. C.P.F. Borges, S. Honda, R.G.S. Berlinck, H. Imasato, P. Berci Filho, M. Tabak, 'Synthesis, Characterization and Interaction with Ionic Micelles of Tetraacetylated Dipyridamole', *Spectrochim. Acta, Part A*, **51**, 2575–2584 (1995).

384. Y. Zheng, M. Wu, 'Effect of Media Properties on the Dissociation and Fluorescence of 1-Naphthaol', *Fenxi Huaxue*, **20**, 1013–1018 (1992).

385. C.M. Harris, B.K. Selinger, 'Excited-State Processes of Naphthols in Aqueous Surfactant Solutions', *Z. Phys. Chem., NF*, **134**, 65–92 (1983).

386. J. Bily, L. Cermakova, 'A Spectrophotometric Study of the Effect of Cationic Tensides (CPB,CPTB) on the Dissociation-Constants of 4-(2-Pyridylazoresorcinol): The Effect of a Strong Electrolyte (NaCl)', *Anal. Lett.*, **19**, 747–753 (1986).

387. C.J. Drummond, F. Grieser, 'The Ionization Behavior of DL-α-Tocopherol (Vitamin E) in Model Membranes – Micelles and Vesicles', *Biochim. Biophys. Acta*, **836**, 275–278 (1985).

388. J. Garcia-Soto, M.S. Fernandez, 'The Effect of Neutral and Charged Micelles on the Acid-Base Dissociation of the Local Anesthetic Tetracaine', *Biochim. Biophys. Acta*, **731**, 275–281 (1983).

389. L.A. Al'bota, 'Effects of Surfactants on 1-(2-Pyridylazo)-2-Naphthol', *Zh. Anal. Khim.*, **40**, 978–984 (1985).

390. M.C. Doody, H.J. Pownall, Y.J. Kao, L.C. Smith, 'Mechanism and Kinetics of Transfer of a Fluorescent Fatty Acid between Single-Walled Phosphatidylcholine Vesicles', *Biochemistry*, **19**, 108–116 (1980).

391. H.-J. Timpe, G. Israel, H.G.O. Becker, I.R. Gould, N.J. Turro, 'Photoinduced Electron Transfer between Aromatic Hydrocarbons and Arene Diazonium Salts in Micellar Solutions', *Chem. Phys. Lett.*, **99**, 275–280 (1983).

392. F.M. Martens, J.W. Verhoeven, 'Charge-Transfer Complexation in Micellar Solutions. Water Penetrability of Micelles', *J. Phys. Chem.*, **85**, 1773–1777 (1981).

393. M.J. Minch, M. Giaccio, R. Wolff, 'Effect of Cationic Micelles on the Acidity of Carbon Acids and Phenols. Electronic and Proton Nuclear Magnetic Resonance Spectral Studies of Nitro Carbanions in Micelles', *J. Am. Chem. Soc.*, **97**, 3766–3772 (1975).

394. M.S. Fernandez, P. Fromherz, 'Lipoid pH Indicators as Probes of Electrical Potential and Polarity in Micelles', *J. Phys. Chem.*, **81**, 1755–1761 (1977).

395. G.M. Escandar, 'Spectrofluorimetric Determination of Piroxicam in the Presence and Absence of Beta-Cyclodextrin', *Analyst*, **124**, 587–591 (1999).

396. R. Compano, A. Izquierdo, M.D. Prat, 'Determination of Gallium by Fluorescence Spectroscopy in Micellar Medium', *Quim. Anal. (Barcelona)*, **10**, 31–40 (1991).

397. T.A. Vasil'chuk, A. Pilipekno, A.I. Volkova, 'Ion Association of Morin and Quarecetin with Cationc Surfactant Active Materials (CSAM) and Their Uses in Analysis', *Ukr. Khim. Zh.*, **51**, 278–283 (1985).

398. R. Boyer, G. Deckey, C. Marzzacco, M. Mulvaney, C. Schwab, A.M. Halpern, 'The Photophysical Properties of 2-Naphthol: A Physical-Chemistry Experiment', *J. Chem. Educ.*, **62**, 630–632 (1985).

399. L.A. Al'bota, 'Influence of Solubilization on the Protolytic Equilibira of Organic Reagents', *Zh. Anal. Khim.*, **41**, 1780–1787 (1986).

400. A. Kusumi, J.S. Hyde, 'Effects of pH on Electron-Spin-Resonance Spectra of Stearic Acid Spin Labels in Membranes: Probing the Membrane-Surface', *Fed. Proc.*, **41**, 1394 (1982).

401. S. Schreier, W.A. Frezzatti, P.S. Aranjo, H. Chaimovich, I.M. Cuccovia, 'Efect of Lipid Membranes on the Apparent pK of Local Anesthetic Tetracaine: Spin Label and Titration Studies', *Biochim. Biophys. Acta*, **769**, 231–237 (1984).

402. A. Sanson, M. Ptak, J.L. Rigaud, C.M. Garybobo, 'ESR Study of Anchoring of Spin-Labeled Stearic Acid in Lecithin Multilayers', *Chem. Phys. Lipids*, **17**, 435–444 (1976).

403. U. Khuanga, R. McDonald, B.K. Selinger, 'Excited State Processes in Surfactant Micelles', *Z. Phys. Chem., NF*, **101**, 209–224 (1976).

404. C.J. Mazzacco, 'The Effect of Solvent and Micelles on the Rate of Excited-State Deprotonation', *J. Chem. Educ.*, **73**, 254–255 (1996).

405. A.M. Campana, F.A. Barrero, M.R. Ceba, 'Sensitive Spectrofluorimetric Method for the Determination of Ethylenediaminetetraacetic Acid and its Salts in Foods with Zirconium Ions and Alizarin Red S in a Micellar Medium', *Anal. Chim. Acta*, **329**, 319–325 (1996).

406. Y.V. Il'ichev, A.B. Demyashkevich, M.G. Kuzmin, H. Lemmetyinen, 'Temperature Effects on Excited State Protolytic Dissociation in Aqueous Solutions, Micelles and Liposomes: Influence of Phase Transitions in Lipid Bilayers', *J. Photochem. Photobiol., A*, **74**, 51–63 (1993).

407. K.M. Solntsev, Y.V. Il'ichev, A.B. Demyashkevich, M.G. Kuzmin, 'Excited-State, Proton-Transfer Reactions of Substituted Naphthols in Micelles. Comparative Study of Reactions of 2-Naphthol and its Long-Chain Alkyl Derivatives in Micellar Solutions of Cetyltrimethylammonium Bromide', *J. Photochem. Photobiol.*, **78**, 39–48 (1994).

408. A.A. Freitas, L. Paulo, A.L. Macanita, F.H. Quina, 'Acid-Base Equilibria and Dynamics in Sodium Dodecyl Sulfate Micelles: Geminate Recombination and Effect of Charge Stabilization', *Langmuir*, **22**, 7986–7993 (2006).

409. F. Pina, M.J. Melo, S. Alves, R. Ballardini, M. Maestri, P. Passaniti, 'Micelle Effect on Ground and Excited State Proton Transfer Reactions Involving the 4-Methyl-7-Hydroxyflavylium Cation', *New J. Chem.*, **25**, 747–752 (2001).

410. A. Malliaris, 'Fluorescence Probing in Aqueous Micellar Systems: An Overview', *Int. Rev. Phys. Chem.*, **7**, 95–115 (1988).

411. E. Iglesias, 'Application of Organized Microstructures to Study Keto-Enol Equilibrium of Beta-Dicarbonyl Compounds', *Curr. Org. Chem.*, **8**, 1–24 (2004).

412. R. Zana (ed.), *Surfactant Solutions: New Methods of Investigation*, Marcel Dekker, New York, 1987.

413. T. Shanmugapriya, C. Selvaraju, P. Ramamurthy, 'Anomalous Association and Fluorophore Influence on the Position of Dimethylaniline in Micelles: Fluorescence Quenching of 1,8-Acridinedione', *Spectrochim. Acta, Part A*, **66**, 761–767 (2007).

414. K.M. Glenn, R.M. Palepu, 'Fluorescence Probing of Aerosol OT Based Reverse Micelles and Microemulsions in n-Alkanes (C-6 - C-16) and Quenching of Safranine-T in These Systems', *J. Photochem. Photobiol., A*, **179**, 283–288 (2006).

415. R. Ranganathan, C. Vautier-Giongo, M.S. Bakshi, B.L. Bales, J. Hajdu, 'Phospholipid Containing Mixed Micelles – Characterization of Diheptanoyl Phosphatidylcholine (DHPC) and Sodium Dodecylsulfate and DHPC and Dodecyltrimethylammonium Bromide', *Chem. Phys. Lipids*, **135**, 93–104 (2005).

416. P. Suppan, 'Solvatochromic Shifts – The Influence of the Medium on the Energy of Electronic States', *J. Photochem. Photobiol., A*, **50**, 293–330 (1990).

417. R. Varadaraj, P. Valint, J. Bock, S. Zushma, N. Brons, 'Micellar Microstructure of Linear and Guerbet Micelles – Fluorescence Probe Study', *J. Colloid Interface Sci.*, **144**, 340–344 (1991).

418. P. Mukerjee, 'The Nature of the Binding of Counterions on Charged Colloids and Macromolecules', *J. Phys. Chem.*, **66**, 943–945 (1962).

419. A. Heredia, G. Requena, F.G. Sanchez, 'An Approach for the Estimation of the Polarity of the β-Cyclodextrin Internal Cavity', *J Chem. Soc., Chem. Commun.*, 1814–1815 (1985).

420. S. Miyagishi, T. Asakawa, M. Nishida, 'Influence of External Environment on Microviscosity in Micelles', *J. Colloid Interface Sci.*, **115**, 199–205 (1987).

421. F. Grieser, M. Lay, P.J. Thistlethwaite, 'Excited-State Torsional Relaxation in 1,1'-Dihexyl-3,3,3',3'-Tetramethylindocarbocyanine Iodide: Application to the Probing of Micelle Structure', *J. Phys. Chem.*, **89**, 2065–2070 (1985).

422. M. Shinitzky, Y. Barenholz, 'Fluidity Parameters of Lipid Regions Determined by Fluorescence Polarization', *Biochim. Biophys. Acta*, **515**, 367–394 (1978).

423. K. Nithipatikom, L.B. McGown, 'Studies of Sodium Taurocholate Micelles Using Fluorescent Probe Molecules', *Photochem. Photobiol.*, **47**, 797–802 (1988).

424. Y. Yan, M.L. Myrick, 'Quantitative Measurement and Discrimination of Isochromatic Fluorophores Based on Micelle-Enhanced Steady-State Fluorescence Polarization in Fluid Solution', *Anal. Chim. Acta*, **441**, 87–93 (2001).

425. W.J. Dressick, B.L. Hauenstein, T.B. Gilbert, J.N. Demas, B.A. DeGraff, 'New Probe of Solvent Accessibility of Bound Photosensitizers. 2. Ruthenium(II) and Osmium(II) Photosensitizers in Triton X-100 Micelles', *J. Phys. Chem.*, **88**, 3337–3340 (1984).

426. B.L. Hauenstein, W.J. Dressick, S.L. Buell, J.N. Demas, B.A. DeGraff, 'A New Probe of Solvent Accessibility of Bound Photosensitizers. 1. Ruthenium(II) and Osmium(II) Photosensitizers in Sodium Lauryl Sulfate Micelles', *J. Am. Chem. Soc.*, **105**, 4251–4255 (1983).

427. A.L. Macanita, F.P. Costa, S.M.B. Costa, E.C. Melo, H. Santos, 'The 9-Anthroate Chromophore as a Fluorescent Probe for Water', *J. Phys. Chem.*, **93**, 336–343 (1989).

428. A.E. Beg, 'A Modified Technique for Refrective Index Increment Measurements for Cetyltrimethylammonium Bromide Solutions Using Interference Refrectrometer', *Pak. J. Sci. Ind. Res.*, **30**, 476–479 (1987).

429. K.J. Mysels, L.H. Princen, 'Light Scattering by Some Laurylsulfate Solutions', *J. Phys. Chem.*, **63**, 1696–1700 (1959).

430. R.A. Lampert, S.R. Meech, J. Metcalfe, D. Phillips, A.P. Schaap, 'The Refractive Index Correction to the Radiative Rate Constant in Fluorescence Lifetime Measurements', *Chem. Phys. Lett.*, **94**, 137–140 (1983).

431. R. Kondepudi, S. Srinivasan, 'Optical Studies on Some Dyes for Liquid Solar Concentrators', *Sol. Energy Mater.*, **20**, 257–263 (1990).

432. M.M.G. Krishna, N. Periasamy, 'Fluorescence of Organic Dyes in Lipid Membranes: Site of Solubilization and Effects of Viscosity and Refractive Index on Lifetimes', *J. Fluoresc.*, **8**, 81–91 (1998).

433. X. Geng, 'Applications of Surfactants in Analytical Chemistry', *Fenxi Shiyanshi*, **5**, 39–43 (1986).

434. J.L.C. Dominguez, M.D.L.G. Cirugeda, 'Contributions to Microscopy in Micellar Media', *Opt. Pura. Appl.*, **21**, 85–93 (1988).

435. W.R.G. Baeyens, L.B. Lin, P. De Moerloose, B. Del Castillo, C. De Jonge, 'Cyclodextrins: Spectroscopic, Pharmaceutical and Chromatographic Application', *Ann. Rev. Acad. Farm.*, **54**, 698–714 (1988).

436. L.J.C. Love, R. Weinberger, 'Recent Advances and Future Prospects in Fluorescence and Phosphorescence Spectroscopy', *Spectrochim. Acta, Part B*, **38**, 1421–1433 (1983).

437. S. Scypinski, L.J.C. Love, 'Cyclodextrin Enhanced Luminescence Spectroscopy', *Am. Lab.*, **16**, 55–59 (1984).

438. M. Vosmanska, Z. Holzbecher, 'Tensides and Cyclodextrins in Fluorometry', *Chem. Listy*, **81**, 1170–1181 (1987).

439. W.L. Hinze, F. Dai, R.P. Frankewich, K.N. Thimmaiah, J. Szejtli, 'Cyclodextrins as Reagents in Analytical Chemistry and Diagnostics', in *Comprehensive Supramolecular Chemistry*, *Cyclodextrins, Vol. 3*, eds. J. Szejtli, T. Osa, Pergamon Press, New York, 587–614, Chapter 20, 1996.

440. J. Georges, 'Molecular Fluorescence in Micelles and Microemulsions: Applications', *Analysis*, **17**, 231–251 (1989).

441. J. Georges, 'Molecular Fluorescence in Micelles and Microemulsions – Micellar Effects and Analytical Applications', *Spectrochim. Acta Rev.*, **13**, 27–45 (1990).

442. L. Li, Y. Zheng, 'Fluorescence and Phosphorescence Phenomena in Micellar Solutions-A New Area of Trace Analysis', *Fenxi Huaxue*, **13**, 548–555 (1985).

443. R. Vonwandruszka, 'Luminescence of Micellar Solutions', *Crit. Rev. Anal. Chem.*, **23**, 187–215 (1992).

444. Z. Li, 'Surfactants as Enhancing Agents and Their Analytical Applications in Analytical Chemistry: State-of-Art of Enhanced Analysis', *Huaxue Shiji*, **10**, 96–103 (1988).

445. M.H. Memon, P.J. Worsfold, 'Analytical Applications of Microemulsions', *Anal. Proc.*, **23**, 418–419 (1986).

446. G.R. Ramos, M.C.G. Alvarez-Coque, A. Berthod, J.D. Winefordner, 'Fluorescence in Microemulsions and Reversed Micelles : A Review and New Results', *Anal. Chim. Acta*, **208**, 1–19 (1988).

447. G. Chen, D. Zhang, 'Applications of Cyclodextrins in Analytical Chemistry', *Huaxue Tongbao*, 1–5 (1988).

448. R.E. Galian, A.G. Bracamonte, A.V. Veglia, 'Hydroxypropyl-β-Cyclodextrin Effect on the Fluorescence of Auxin and Skatole and on the Simultaneous Determination of Binary Mixtures of Indole Compounds in Urine by First Derivative Spectrofluorimetry', *Anal. Chim. Acta*, **540**, 393–401 (2005).

449. I.D. Meras, A. Espinosa-Mansilla, D.A. Rodriquez, 'Complexation Study of Cinalukast and Montelukast with Cylodextrins', *J. Pharm. Biomed. Anal.*, **43**, 1025–1032 (2007).

450. A. Banerjee, P.K. Sengupta, 'Encapsulation of 3-Hydroxyflavone and Fisetin in Beta-Cyclodextrins: Excited State Proton Transfer Fluorescence and Molecular Mechanics Studies', *Chem. Phys. Lett.*, **424**, 379–386 (2006).

451. Y.V. Fedorov, O.A. Fedorova, E.N. Andryukhina, S.P. Gromov, M.V. Alfimov, J.J. Aaron, 'Guest-Host Interactions Between Crown-Containing 2-Styrylbenzothiazole and Hydroxypropyl-β-Cyclodextrin', *J. Inclusion Phenom. Macrocyclic Chem.*, **49**, 283–289 (2004).

452. D.H. Lee, W.F. Koo, 'Solvents and Micellar Effects on the Absorption and Emmision Spectra of Indole and Methylindoles', *Nonmunjip-Han Nam Taehakkyo, Chayon Kwahak Pyon*, **16**, 139–152 (1986).

453. J.A. Resing, C.I. Measures, 'Fluorometric Determination of Al in Seawater by Flow Injection Analysis with in-Line Preconcentration', *Anal. Chem.*, **66**, 4105–4111 (1994).

454. A. Sanz-Medel, M.M. Fernandez Perez, M. De la Guardia Cirugeda, J.L.C. Dominguez, 'Metal Chelate Fluorescence Enhancement by Nonionic Micelles: Surfactant and Auxiliary Ligand Nature Influence on the Niobium-Lumogallion Complex', *Anal. Chem.*, **58**, 2161–2166 (1986).

455. J. Lasovsky, M. Rypka, F. Grambal, E. Ruzicka, 'The Enhancement of Fluorescence in Micellar Solutions of Ionic Surfactants', *Acta Univ. Palacki. Olomuc., Fac. Rerum Nat.*, **85**, 47–54 (1986).

456. L. Zhou, S.M. Wang, K. Tian, Y.M. Dong, Z.D. Hu, 'Simultaneous Determination of Catecholamines and Amino Acids by Micellar Electrokinetic Chromatography with Laser-Induced Fluorescence Detection', *J. Sep. Sci.*, **30**, 110–117 (2007).

457. M.O. Iwunze, 'Media Influence on the Enhancement of the Fluorescence of Berberine Hydrochloride', *Monatsh. Chem.*, **131**, 429–435 (2000).

458. G.A. Clarke, J. Burton, 'Spectral Study of Aromatic Acids in Micellar Environment', in *Solution Behaviour of Surfactants*, eds. K.L. Mittal, Y. Nitzan, Plenum Press, New York, 1047, Vol. 2, 1982.

459. G.A. Clarke, Personal Communication (1981).

460. A.A. Lamola, J. Eisinger, W.E. Blumberg, S.C. Patel, J. Flores, 'Fluorometric Study of the Partition of Bilirubin Among Blood Components: Basis for Rapid Microassays of Bilirubin and Bilirubin Binding Capacity in Whole Blood', *Anal. Biochem.*, **100**, 25–42 (1979).

461. B. Ehrenberg, Z. Malik, Y. Nitzan, 'Fluorescence Spectral Changes of Hematoporphyrin Derivative upon Binding to Lipid Vesicles, *Staphylococcus Aureus* and *Escherichia Coli* Cells', *Photochem. Photobiol.*, **41**, 429–435 (1985).

462. J. Jung, D.H. Kim, 'α–Tocopherol as a Fluorescence Probe. I. Its Spectroscopic Characterization and Use as a Probe of Liposomes', *Hanguk Nonghwa Hakkoe Chi*, **25**, 224 (1982).

463. T. Takeuchi, 'Micelles for Signal Enhancement and Novel Selectivity of Dansyl Amino Acids', *J. Chromatogr., A*, **780**, 219–228 (1997).

464. N. Ishibashi, K. Kina, Personal Communication (1987).

465. M.R. de Moreno, R.V. Smith, 'Fluorescence Behavior of Phenolic Compounds in Solutions Containing Micelles', *Anal. Lett.*, **16**, 1633–1645 (1983).

466. R.G. Cowgill, 'Fluorescence and the Structure of Proteins. III. Effects of Denaturation on Fluorescence

of Insulin and Ribonuclease', *Arch. Biochem. Biophys.*, **104**, 84–92 (1964).

467. R. Femia, S. Scypinski, L.J.C. Love, 'Fluorescence Characteristics of Polychlorinated Biphenyl Isomers in Cyclodextrin Media', *Environ. Sci. Technol.*, **19**, 155–159 (1985).

468. I.E. Obyknovennaya, M.I. Snegov, A.S. Cherkasov, 'Quenching of Fluorescence of Anthracene in Aqueous Micellar Solvent', *Opt. Spectrosc.*, **57**, 604–609 (1984).

469. A. Ueno, K. Takahashi, T. Osa, 'One Host – Two Guests Complexation between γ-Cyclodextrin and Sodium α-Naphthylacetate as Shown by Excimer Fluorescence', *J. Chem. Soc., Chem. Commun.*, 921–922 (1980).

470. O. Jules, S. Scypinski, L.J.C. Love, 'Fluorescence Enhancement of Several Hallucinogenic Drugs in Cyclodextrin Media', *Anal. Chim. Acta*, **169**, 355–360 (1985).

471. C.D. Tran, T.A. Van Fleet, 'Micellar Induced Simultaneous Enhancement of Fluorescence and Thermal Lensing', *Anal. Chem.*, **60**, 2478–2482 (1988).

472. V.G. Bieri, D.F.H. Wallach, 'Fluorescence Quenching in Lecithin and Lecithin-Cholesterol Liposomes by Paramagnetic Lipid Analogues – Introduction of a New Probe Approach', *Biochim. Biophys. Acta*, **389**, 413–427 (1975).

473. L.K. Patterson, E. Vieil, 'Comparison of Micellar Effects on Singlet Excited States of Anthracene and Perylene', *J. Phys. Chem.*, **77**, 1191–1192 (1973).

474. J.J.S. Rodriguez, J.H. Garcia, M.M.B. Suarez, A.B. Martinlazaro, 'Analysis of Mixtures of Polycyclic Aromatic-Hydrocarbons in Sea Water by Synchronous Fluorescence Spectrometry in Organized Media', *Analyst*, **118**, 917–921 (1993).

475. P.K.F. Koglin, D.J. Miller, J. Steinwandel, M. Hauser, 'Determination of Micelle Aggregation Numbers by Energy Transfer', *J. Phys. Chem.*, **85**, 2363–2366 (1981).

476. B.B. Craig, J. Kirk, M.A.J. Rodgers, 'Fluorescence Decay Parameters for Pyrene in Micellar and Homogeneous Liquid Dispersions', *Chem. Phys. Lett.*, **49**, 437–440 (1977).

477. H. Singh, W.L. Hinze, 'Micellar Enhanced Spectrofluorimetric Methods: Application to the Determination of Pyrene', *Anal. Lett.*, **15**, 221–243 (1982).

478. H.E. Edwards, J.C. Allen, 'An Investigation of the Interactions between Lecithin Liposomes and Surfactant Molecules Using Pyrene Fluorescence', *J. Chem. Res.-S*, 179 (1982).

479. T. Vo Dinh, K.P. Li, J.D. Winefordner, 'Fluorescence Studies of Benzo[a]pyrene in Liposome Membrane Systems', *Biochem. Biophys. Res. Commun.*, **73**, 187–192 (1976).

480. H.N. Singh, Y. Baba, W.L. Hinze, 'Enhancement Factors Determined from Comparison of Calibration Plots from at Least Five Different Analyte Concentration Levels (Air-Saturated Solutions) in Indicated Surfactant Micelle System Relative to that Obtained in Water under Otherwise Identical Experimental Conditions', (Unpublished data).

481. J. Amador-Hernandez, A. Cladera, J.M. Estela, P.L. Lopez-de-Alba, V. Cerda, 'Resolution of a Multicomponent Polycyclic Aromatic Hydrocarbon System in Micellar Media by Linear Variable Angle Fluorescence Applying Distinct Chemometric Techniques', *Analyst*, **123**, 2235–2241 (1998).

482. D.W. Armstrong, W.L. Hinze, K.H. Bui, H.N. Singh, 'Enhanced Fluorescence and Room Temperature Liquid Phosphorescence Detection in Pseudophase Liquid Chromatography (PLC)', *Anal. Lett.*, **14**, 1659–1667 (1981).

483. M. Hoshino, M. Imamura, K. Ikehara, Y. Hama, 'Fluorescence Enhancement of Benzene Derivatives in Aqueous β–Cyclodextrin Solutions', *Kokagaku Toronkai Koen Yoshishu*, **92**, (1979).

484. R.M.M. Brito, W.L.C. Vaz, 'Determination of the Critical Micelle Concentration of Surfactants Using the Fluorescent Probe N-Phenyl-1-Naphthylamine', *Anal. Biochem.*, **152**, 250–255 (1986).

485. L. Fisher, D. Oakenfull, 'Environment of Solubilized Molecules in Bile-Salt Micelles', *Aust. J. Chem.*, **32**, 31–39 (1979).

486. E. London, G.W. Feigenson, 'A Convenient and Sensitive Fluorescence Assay for Phospholipid Vesicles Using Diphenylhexatriene', *Anal. Biochem.*, **88**, 203–211 (1978).

487. L.F. Vieira Ferreira, M.J. Lemos, V. Wintgens, J.C. Netto-Ferreira, 'Influence of the Alkyl Chain Length on the Fluorescence Properties of N-Alkyl-2,3-Naphthalimides Included in β-Cyclodextrin', *Spectrochim. Acta, Part A*, **55**, 1219–1227 (1999).

488. G. Saroja, A. Samanta, 'Hydrophobicity-Induced Aggregation of N-Alkyl-4-Aminophthalimides in Aqueous Media Probed by Solvatochromic Fluorescence', *J. Chem. Soc., Faraday Trans.*, **94**, 3141–3145 (1998).

489. Y.V. Ilichev, K.M. Solntsev, M.G. Kuzmin, H. Lemmetyinen, 'Effects of Surfactant Charge and Structure on Excited State Protolytic Dissociation of 1-Naphthol in Vesicles', *J. Chem. Soc., Faraday Trans.*, **90**, 2717–2724 (1994).

490. L. Bruzzone, 'Spectrofluorimetric Determination of 2-Naphthol in a Pharmaceutical Shampoo Using Anionic SDS Micellar Media', *Anal. Lett.*, **29**, 2141–2152 (1996).

491. T.L. Keimig, L.B. McGown, 'Micellar Modification of the Spectral, Intensity and Lifetime Characteristics of the Fluorescence of Fluorescein-Labelled Phenobarbital', *Talanta*, **33**, 653–656 (1986).

492. A.K. Mathur, C. Agarwal, B.S. Pangtey, A. Singh, B.N. Gupta, 'Surfactant-Induced Fluorescence Changes in Fluorescein Dye', *Int. J. Cosmet. Sci.*, **10**, 213–218 (1988).

493. V.N. Mishra, N. Datt, 'Study of the Tautomeric Behavior of Fluorescein Dye Due to Surfactant Interaction', *Indian J. Chem., Sect. A*, **24**, 597–600 (1985).

494. M.M. Habashy, M.S. Antonious, B.A. Elsayed, M.S.A. Abdelmottaleb, 'Fluorescence Properties of Some Substituted Acyclic Azines', *J. Lumin.*, **36**, 173–176 (1986).

495. D.F. Swaile, C.L. Copper, M.J. Sepaniak, D.E. Burton, L.L. Powell, 'The Use of Fluorescent Probe Molecules for Detection in Capillary Electrokinetic Separations', *Talanta*, **41**, 1499–1505 (1994).

496. G.K. Redd, D.S. Smith, 'Retinol: A Fluorescent Probe for Membrane Lipids', *FEBS Lett.*, **9**, 287–289 (1970).

497. R. Paul, K. Sreekrishna, 'Interaction of Vitamin A with Biliary and Model Micelles', *Indian J. Biochem. Biophys.*, **16**, 257–258 (1979).

498. F. Garcia Sanchez, A.L. Ramos Rubio, C. Cruces Blanco, M. Hernandez Lopez, J.C. Marquez Gomez, C. Carnero, 'Enhanced Fluorimetric Determination of Procaine in Pharmaceutical Preparations by Aqueous β-Cyclodextrin Inclusion and Non-Aqueous Competitive Action', *Anal. Chim. Acta*, **205**, 139–147 (1988).

499. H. Sato, M. Kawasaki, K. Kasatani, N. Nakashima, K. Yoshihara, 'Interaction of Cationic Dye and Anionic Detergent Above and Below the Critical Micelle Concentration as Revealed by Fluorescence Characteristics', *Bull. Chem. Soc. Jpn.*, **56**, 3588–3594 (1983).

500. K.D. Khalaf, J. Sancenon, M. Delaguardia, 'Study of the Interaction of Propoxur and its Metabolites with Surfactants', *Microchem. J.*, **48**, 200–209 (1993).

501. R.C. Kapoor, V.N. Mishra, 'Influence of Surfactants on the Fluorescence and Absorption Spectra of Erythrosin in Aqueous Solution', *J. Lumin.*, **17**, 175–183 (1978).

502. A. Seret, E. Gandin, A. Vandevorst, 'Flash-Photolysis of Eosin in Aqueous Micellar Dispersions', *Chem. Phys. Lett.*, **135**, 427–431 (1987).

503. A.T. Pilipekno, A.I. Volkova, G.N. Pshinko, 'Mechanism of Effects of Surfactants of Fluorescent Reactions', *Zh. Anal. Khim.*, **135**, 2095–2105 (1981).

504. J. Lasovsky, F. Grambal, 'Increasing the Intensity of Radiation Emitted by Ionic or Ampholytic Dyes', Czech. Pat. CS 210294, 1983; *Chem. Abstr.*, **98**, 188819z.

505. D.J. Jobe, R.E. Verrall, R. Palepu, V.C. Reinsborough, 'Fluorescence and Conductometric Studies of Potassium 2-(p-Toluidinyl)Naphthalene-6-Sulfonate/Cyclodextrin/ Surfactant Systems', *J. Phys. Chem.*, **92**, 3582–3586 (1988).

506. E.Z.M. Ebeid, M.H. Abdelkader, R.M. Issa, S.A. Eldaly, 'Viscosity and Medium Effects on the Fluorescence and Photochemical Behavior of Some Aryl Chalcones', *Chem. Phys. Lett.*, **146**, 331–336 (1988).

507. M. Krieg, 'Singlet Oxygen Production and Fluorescence Yields of Merocyanine-540 – A Comparative Study in Solution and Model Membrane Systems', *Biochim. Biophys. Acta*, **1105**, 333–335 (1992).

508. A.S. Verkman, 'Mechanism and Kinetics of Merocyanine 540 Binding to Phospholipid Membranes', *Biochemistry*, **26**, 4050–4056 (1987).

509. R. Paul, M.K. Mathew, R. Narayanan, P. Balaram, 'Fluorescent Probe and NMR Studies of the Aggregation of Bile-Salts in Aqueous Solution', *Chem. Phys. Lipids*, **25**, 345–356 (1979).

510. P. Cavatorta, A. Spisni, A.G. Szabo, G. Farruggia, L. Franzoni, L. Masotti, 'Conformation of Bombesin in Buffer and in the Presence of Lysolecithin Micelles: NMR, CD and Fluorescence Studies', *Biopolymers*, **28**, 441–463 (1989).

511. D.J. Cowley, R.S.J. Todd, 'Triazinylaniline (TA) Derivatives as Fluorescence Probes. 2. Steady-State and Time-Resolved Fluorescence Anisotropy Studies of TA Probes in Alkanols and in Small Unilamellar Vesicles of Phospholipids', *J. Chem. Soc., Perkin Trans. 2*, 299–306 (1995).

512. N.O. Petersen, R. Gratton, E.M. Pisters, 'Fluorescence Properties of Polyene Antibiotics in Phospholipid Bilayer Membranes', *Can. J. Chem.*, **65**, 238–244 (1987).

513. O. Valdesaguilera, L. Cincotta, J. Foley, I.E. Kochevar, 'Photobleaching of a Cyanine Dye in Solution and in Membranes', *Photochem. Photobiol.*, **45**, 337–344 (1987).

514. J.H. Turnbull, R.J. Walker, 'The Effects of β-Cyclodextrin on the Luminescence of *para*-Aminoacids', *Nuovo Cimento Soc. Ital. Fis., B*, **63B**, 441–445 (1981).

515. W. Birmachu, J.K. Reed, 'The Fluorescence Properties of Ortho Aminobenzoate Anesthetics in Defined Solvents and Phospholipid Vesicles', *Photochem. Photobiol.*, **47**, 675–683 (1988).

516. Z. Konefal, 'The Influence of Detergents on the Luminescence Properties of Rhodamine B and 6G in Aqueous Solutions', *Acta Phys. Chem.*, **30**, 135–140 (1984).

517. M.K. Carroll, G.M. Hieftje, 'A Novel Fiber-Optic-Based Pump-Probe Instrument for the Acquisition of Kinetic Information', *Appl. Spectrosc.*, **46**, 317–321 (1992).

518. J.J. Santana, M. Gunshefski, J.D. Winefordner, 'Micellar Enhanced Spectrofluorometric Determination of Chlorophyll a and Chlorophyll b in Fresh Waters', *Talanta*, **39**, 195–200 (1992).

519. P. Martin, M.A. Martin, B. Del Castillo, I. Cayre, 'Polarity Effect on Fluorescence of Styryl Derivatives of Quinolizinium Salts in Micellar Media', *Anal. Chim. Acta*, **205**, 129–137 (1988).

520. S. Basu, 'Fluorescence Enhancement of Tetrakis(Sulfonatophenyl) Porphyrin in Homogeneous and Micellar Solutions', *J. Photochem. Photobiol., A*, **56**, 339–347 (1991).

521. S. Nigam, M. Belletete, R.S. Sarpal, G. Durocher, 'Solubilization of 2-Cyano-3H-Indole Molecular Probes in SDS and CTAB Micelles. Spectral and Photophysical Characterization', *J. Chem. Soc., Faraday Trans.*, **91**, 2133–2139 (1995).

522. M.D. Premakumara de Costa, A. Prasana de Salva, S.T. Pathirana, 'Two Dimentional Fluorescent Sensors: The Different Dependences of the Fluorescent Band Position and the Fluorescence Quantum Yield of 1,5-Diphenyl-3-Vinyl-Δ^2-Pyrazoline Upon Solvent Dipolarity and Hydrogen Bond Activity', *Can. J. Chem.*, **65**, 1416–1419 (1987).

523. M. De la Guardia Cirugeda, F. Rodilla Soriano, 'Micelle Enhancement of Benzodiazepine Fluoresence', *Analyst*, **114**, 77–82 (1989).

524. M. de la Guardia, J.L. Carrión, M.V. Galdú, 'The Use of Topological Models in Analytical Chemistry', *J. Chemom.*, **3**, 193–207 (1989).

525. M. Delaguardia, M.L. Hernandez, S. Sancenon, J.L. Carrion, 'Micellar Fluorescence Enhancement of Carbamate Pesticides', *Colloids Surf.*, **48**, 57–64 (1990).

526. J.J.S. Rodriguez, Z.S. Ferrera, J.H. Garcia, A.J.B. Martinlazaro, 'Analysis of Binary-Mixtures of 3,3',4,4'-Tetrachlorobiphenyl and 2,3,7,8-Tetrachlorodibenzofuran by Derivative Synchronous Fluorescence Spectrometry in Organized Media', *Analyst*, **119**, 2241–2246 (1994).

527. D. Demore, A. Kasselouri, O. Bourdon, J. Blais, G. Mahuzier, P. Prognon, 'Enhancement of 5,10,15,20-Tetra(m-Hydroxyphenyl)chlorin Fluorescence Emission by Inclusion in Natural and Modified Cyclodextrins', *Appl. Spectrosc.*, **53**, 523–527 (1999).

528. S. Carda-Broch, J.S. Esteve-Romero, M.C. Garcia-Alvarez-Coque, 'Chromatographic Determination of Diuretics in Urine Samples Using Hybrid Micellar Mobile Phases with Fluorimetric Detection', *Anal. Chim. Acta*, **375**, 143–154 (1998).

529. M. Delaguardia, M.V. Galdu, 'Fluorescence Enhancement of Mazindol in the Presence of Anionic Surfactants', *Spectrochim. Acta, Part A*, **44**, 403–407 (1988).

530. F. Grases, C. Genestar, J.J. Gil, 'Effect of Surfactants on the Fluorescence Intensity of Organic Reagents', *Microchem. J.*, **31**, 44–49 (1985).

531. A. Berthod, J.M. Asensio, J.J. Laserna, 'Micellar Liquid Chromatography for Rapid Screening of Illegal Drugs in Sport', *J. Liq. Chromatogr.*, **12**, 2621–2634 (1989).

532. F.G. Sanchez, A.N. Diaz, L.S. Feria, 'α-Cyclodextrin Enhanced Fluorometric Determination of the Foodstuff Preservative Methyl Paraben', *Anal. Lett.*, **27**, 2171–2180 (1994).

533. A. Nag, K. Bhattacharyya, 'Dual Luminescence of Dimethylaminobenzonitrile in γ-Cyclodextrin. Environmental Effects on Twisted Intramolecular Charge-Transfer Phenomenon', *J. Chem. Soc., Faraday Trans.*, **86**, 53–54 (1990).

534. A. Nag, R. Dutta, N. Chattopadhyay, K. Bhattacharyya, 'Effect of Cyclodextrine Cavity Size on Twisted Intramolecular Charge Transfer Emission: Dimethylaminobenzonitrile in β-Cyclodextrin', *Chem. Phys. Lett.*, **157**, 83–86 (1989).

535. Y.-B. Jiang, X.-J. Wang, L. Lin, 'Fluorescent Probing of the Restriction by Aqueous Micelles of the Formation of the Photoinduced Biradical State of 4-(Dimethylamino)chalcone', *J. Phys. Chem.*, **98**, 12367–12372 (1994).

536. J. Blais, P. Prognon, G. Mahuzier, P. Vigny, 'Relative Affinity of 5-Methoxypsoralen and 8-Methoxypsoralen toward β-Cyclodextrin: A Fluorescence, Circular Dichroism and Chromatographic Study', *J. Photochem. Photobiol., B*, **2**, 455–465 (1988).

537. I. Tabushi, K. Yamamura, K. Fujita, H. Kawakubo, 'Specific Inclusion Catalysis By β-Cyclodextrin in the One-Step Preparation of Vitamin K_1 or K_2 Analogs', *J. Am. Chem. Soc.*, **101**, 1019–1026 (1979).

538. C.M. Maragos, M. Appell, 'Capillary Electrophoresis of the Mycotoxin Zearalenone Using Cyclodextrin Enhanced Fluorescence', *J. Chromatogr. A*, **1143**, 252–257 (2007).

539. N.L. Pacioni, A.V. Veglia, 'Determination of Poorly Fluorescent Carbamate Pesticides in Water, Bendiocarb and Promecarb, Using Cyclodextrin Nanocavities and Related Media', *Anal. Chim. Acta*, **583**, 63–71 (2007).

540. M.D.G. Seye, C. Parrain, J.J. Aaron, N. Motohashi, 'Spectrofluorimetric Study of an Inclusion Complex of 2-Hydroxypropyl-β-Cyclodextrin with the Antitumour 11-Methyl Benzophenothiazine. Analytical Application to Urine', *J. Inclusion Phenom. Macrocyclic*, **55**, 271–277 (2006).

541. T. Perez-Ruiz, C. Martinez-Lozano, J. Martin, M.D. Garcia, 'Automatic Determination of Phylloquinone in Vegetables and Fruits Using on-Line Photochemical Reduction and Fluorescence Detection Via Solid Phase Extraction and Flow Injection', *Anal. Bioanal. Chem.*, **384**, 280–285 (2006).

542. S.K. Ghosh, A. Pal, S. Kundu, M. Mandal, S. Nath, T. Pal, 'Emission Behavior of 1-Methylaminopyrene in Aqueous Solution of Anionic Surfactants', *Langmuir*, **20**, 5209–5213 (2004).

543. Y.Y. Zhou, C. Liu, H.W. Xu, H.P. Yu, Q. Lu, L. Wang, 'Study on the Fluorescence Behavior of p-Sulfonatedcalix[4,6]arene in Cationic Cetyltrimethylammonium Bromide Solution and its Analytical Application', *Spectrochim. Acta, Part A*, **66**, 919–923 (2007).

544. D. Subbiah, S. Kala, A.K. Mishra, 'Study on the Fluorescence Characteristics of Bromadiolone in Aqueous and Organized Media and Application in Analysis', *Chemosphere*, **61**, 1580–1586 (2005).

545. V. Crescenzi, A. Gamini, A. Palleschi, R. Rizzo, 'Physicochemical Aspects of Inclusion Complex Formation between Cycloamyloses and Aromatic Guest Molecules in Water', *Gazz. Chim. Ital.*, **116**, 435–440 (1986).

546. T. Ban, K. Kasatani, M. Kawasaki, H. Sato, 'Fluorescence Decay of the Acridine Orange Sodium Dodecyl-Sulfate System – Formation of Dye-Rich Induced Micelles in the Premicellar Region', *Photochem. Photobiol.*, **37**, 131–139 (1983).

547. X.-Q. Guo, Z.-L. Zhang, Y.-B. Zhao, D.-Y. Wang, J.-G. Xu, 'DNA-Dye Fluorescence Enhancement Based on Shifting the Dimer-Monomer Equilibrium of Fluorescent Dye', *Appl. Spectrosc.*, **51**, 1002–1007 (1997).

548. J. Liu, K.A. Cobb, M. Novotny, 'Capillary Electrophoretic Separations of Peptides Using Micelle-Forming Compounds and Cyclodextrins as Additives', *J. Chromatogr., A*, **519**, 189–197 (1990).

549. S. Scypinski, J.M. Drake, 'Photophysics of Coumarin Inclusion Complexes with Cyclodextrin. Evidence for Normal and Inverted Complex Formation', *J. Phys. Chem.*, **89**, 2432–2435 (1985).

550. A. Nag, T. Chakrabarty, K. Bhattacharyya, 'Effect of γ-Cyclodextrin on the Intramolecular Charge Transfer Processes in Aminocoumarin Laser Dyes', *J. Phys. Chem.*, **94**, 4203–4206 (1990).

551. M. Kaschke, O. Kittelmann, K. Vogler, A. Graness, 'Picosecond Absorption Studies of Intermolecular Electronic Energy Transfer in Micellar Systems. 1. Inhomogeneous Spatial Distribution and Indirect Donor-Donor Interaction', *J. Phys. Chem.*, **92**, 5998–6003 (1988).

552. K. Shimada, T. Masue, K. Toyoda, M. Takani, T. Nambara, 'The Utility of Cyclodextrin in Mobile Phase for High Performance Liquid-Chromatographic Separation of Isomeric Estrogens', *J. Liq. Chromatogr.*, **11**, 1475–1484 (1988).

553. V.N. Mishra, 'Influence of Surfactants on the Fluorescence and Absorption Spectra of Eosin in Aqueous Solution', *Acta Chim. Hung.*, **116**, 5–12 (1984).

554. J.A. Perez, I. Haro, I. Martin, M.A. Alsina, F. Reig, 'Surface and Polarization Fluorescence Studies on the Interaction of an Rgd Sequence Containing a *Hepatitis a* Virus Peptide with Phospholipids', *Anal. Chim. Acta*, **303**, 65–72 (1995).

555. J. Coene, E. Van Den Eeckhout, P. Sandra, 'Fluorescence Behaviour of an Alkyl Lysophospholipid in the Presence of 1,6-Dipheylhexatriene', *Anal. Chim. Acta*, **227**, 303–308 (1989).

556. Y.P. Morozova, E.B. Zhigalova, 'Spectral Properties of Cationic Dyes in Aqueous-Surface Active-Substance Mixtures', *Zh. Fiz. Khim.*, **56**, 2491–2494 (1982); *Chem. Abstr.*, **98**, 9696n.

557. G. Nowicka, W. Nowicki, 'Interaction of Rhodamine Dye with Anionic Surfactants', *J. Chim. Phys.*, **91**, 247–257 (1994).

558. C.J. Halfman, F.C.L. Wong, D.W. Jay, 'Solvent Perturbation Fluorescence Immunoassay Technique', *Anal. Chem.*, **57**, 1928–1930 (1985).

559. S.D.F.P. Greenspan, 'Biological Applications of the Fluorescent Lipid Probe, Nile Red', *Kodak Lab. Chem. Bull.*, **56**, 1–4 (1985).

560. J.-R. Daban, M. Samso, S. Bartolome, 'Use of Nile Red as a Fluorescent Probe for the Study of the Hydrophobic Properties of Protein-Sodium Dodecyl Sulfate Complexes in Solution', *Anal. Biochem.*, **199**, 162–168 (1991).

561. R.J. Clarke, 'Binding and Diffusion Kinetics of the Interaction of a Hydrophobic Potential-Sensitive Dye with Lipid Vesicles', *Biophys. Chem.*, **39**, 91–106 (1991).

562. R. Cubeddu, W.F. Keir, R. Ramponi, T.G. Truscott, 'Photophysical Properties of Porphyrin Chlorin Systems in the Presence of Surfactants', *Photochem. Photobiol.*, **46**, 633–638 (1987).

563. M.A. Martin, B. del Castillo, D.A. Lerner, 'Influence of Micellar Media on the Fluorescence of Various Benzo- and Methyl-Quinolizinium Salts', *Anal. Chim. Acta*, **205**, 117–127 (1988).

564. J.S. Lundgren, F.V. Bright, 'Effects of Surfactants on the Dynamical Behavior of Acrylodan-Labeled Bovine Serum Albumin', *J. Phys. Chem.*, **100**, 8580–8586 (1996).

565. R. Knoesel, D. Markovitsi, J. Simon, G. Duportail, 'Annelides. 8. Luminescence Properties of Amphiphilic Complexes of Ruthenium in Micellar Phases', *J. Photochem.*, **22**, 275–283 (1983).

566. S.J. Isak, E.M. Eyring, 'Cresyl Violet Chemistry and Photophysics in Various Solvents and Micelles', *J. Photochem. Photobiol., A*, **64**, 343–358 (1992).

567. T.T. Ndou, R. Vonwandruszka, 'Energy Transfer in Triton X-405 Micelles: The Effects of Temperature and Brij-35', *Photochem. Photobiol.*, **50**, 547–551 (1989).

568. A. Zornoza, M. Sánchez, I. Vélaz, L. Fernández, 'Diflunisal and its Complexation with Cyclodextrins. A Fluorimetric Study', *Biomed. Chromatogr.*, **13**, 111–112 (1999).

569. A.M. de la Pena, R.A. Agbaria, M. Sanchez Pena, I.M. Warner, 'Spectroscopic Studies of the Interaction

of 1,4-Diphenyl-1,3- Butadiene with α-, β-, and γ-Cyclodextrins', *Appl. Spectrosc.*, **51**, 153–159 (1997).

570. M.L. Vazquez, A. Cepeda, P. Prognon, G. Mahuzier, J. Blais, 'Cyclodextrins as Modifiers of the Luminescence Characteristics of Aflatoxins', *Anal. Chim. Acta*, **255**, 343–350 (1991).
571. F. Schroeder, G. Nemecz, E. Gratton, Y. Barenholz, T.E. Thompson, 'Fluorescence Properties of Cholestatrienol in Phosphatidylcholine Bilayer Vesicles', *Biophys. Chem.*, **32**, 57–72 (1988).
572. A. Takadate, H. Fujino, S. Goya, F. Hirayama, M. Otagiri, K. Uekama, H. Yamaguchi, 'Fluorescence Behavior of 7-Substituted Coumarin Derivatives by Inclusion Complexation with β-Cyclodextrin', *Yakugaku Zasshi*, **103**, 193–197 (1983).
573. X. Zhang, J. Chen, C. Wan, L. Liu, G. Xu, Y. Tang, 'Fluorescence Investigation on the Complexation Process of Coumarin-Cyclodextrin System', *Wuli Huaxue Xuebao*, **2**, 493–498 (1986).
574. Z.P. Yi, Z.Z. Huang, J.S. Yu, H.L. Chen, 'Be Careful When You Want to Use Buffers in Aqueous Solutions of Cyclodextrins. Effect of Buffers on the Formation Constants of 3-Hydroxy-2-Naphthoic Acid with β-Cyclodextrin', *Chem. Lett.*, 1161–1162 (1998).
575. N.A. Burmistrova, S.P. Mushtakova, P.E. Kuznetsov, 'Fluorescence of 4-Dimethylaminochalcone and Interaction in the System Aromatic Amines – Surfactants', *J. Anal. Chem.*, **61**, 1183–1186 (2006).
576. J. Sadecka, M. Cakrt, P. Majek, 'Spectrofluorimetric Study of Fenoprofen and its Complexes with Cyclodextrins', *Chem. Anal. (Warsaw)*, **50**, 1087–1098 (2005).
577. S.M.Z. Al-Kindy, S.A. El-Sherbini, M.H. Abdel-Kader, 'Uv-Visile Absorption and Fluorescence Characteristics of the Luminescent Label Coumarin-6-Sulphonyl Chloride in Homogeneous and Micellar Solutions', *Anal. Chim. Acta*, **285**, 329–333 (1994).
578. N. Sarkar, A. Datta, S. Das, K. Bhattacharyya, 'Solvation Dynamics of Coumarin 480 in Micelles', *J. Phys. Chem.*, **100**, 15483–15486 (1996).
579. R.W. Ricci, 'Inclusion Complexes of Cyclohexaamylose with Indole', *Carbohydr. Res.*, **129**, 278–286 (1984).
580. I. Mori, Y. Fujita, K. Ikuta, Y. Nakahashi, E. Kakimi, K. Kato, 'Fluorometric Determination of Uranium and Tungsten with o-Hydroxyhydroquinonephthalein in the Presence of Non-Ionic Surfactant', *Talanta*, **36**, 688–690 (1989).
581. J.H. Yang, H.M. Ge, N. Jie, X.Z. Ren, C.L. Tong, J.Z. Wang, 'Study of the Fluorescence System Thulium-Bis(1'-Phenyl-3'-methyl-5'-pyrazol-4'-one)Hexanedione-Cetyltrimetrilammonium Bromide and its Analytical Application', *Analyst*, **120**, 1705–1708 (1995).
582. I. Duranmeras, A.M. Delapena, F. Salinas, I.R. Caceres, 'Spectrofluorometric Determination of Nalidixic-Acid Based on Host-Guest Complexation with γ-Cyclodextrin', *Analyst*, **119**, 1215–1219 (1994).
583. J. Sancenon, J.L. Carrion, M. de la Guardia, 'Fluorometric Determination of Carbaryl in Micellar Media', *Talanta*, **36**, 1165–1169 (1989).
584. S.M.Z. Al-Kindy, F.E.O. Suliman, A.A. Al-Hamadi, 'Fluorescence Enhancement of Coumarin-6-Sulfonyl Chloride Amino Acid Derivatives in Cyclodextrin Media', *Anal. Sci.*, **17**, 539–543 (2001).
585. J.A. Ocana, F.J. Barragan, M. Callejon, 'Spectrofluorimetric Determination of Moxifloxacin in Tablets, Human Urine and Serum', *Analyst*, **125**, 2322–2325 (2000).
586. P. Ramos-Lledo, S. Vera, M.P. San Andres, 'Determination of Vitamins A and E in Milk Samples by Fluorescence in Micellar Media', *Fresenius J. Anal. Chem.*, **369**, 91–95 (2001).
587. F. Bertorelle, W. Dondon, S. Fery-Forgues, 'Compared Behavior of Hydrophobic Fluorescent NBD Probes in Micelles and in Cyclodextrins', *J. Fluoresc.*, **12**, 205–207 (2002).
588. K.-W. Cha, C.-I. Park, 'Spectrofluorimetric Determination of Iron (III) with 2-Pyridinecarbaldhyde-5-Nitro-Pyridylhydrazone in the Presence of Hexadecyltri- methylammonium Bromide Surfactant', *Talanta*, **43**, 1335–1340 (1996).
589. H. Wu, H. Wang, 'Studies of the Influence of the Surfactant Sodium Dodecyl Sulfate on the Fluorescence Properties of Kinetin', *Anal. Chim. Acta*, **329**, 161–164 (1996).
590. J. Yang, H. Ge, N. Jie, X. Ren, N. Wang, H. Zou, 'Study on the Co-Luminescence System of Tb-Gd-BPMPHD-CTAB and its Analytical Application', *Spectrochim. Acta, Part A*, **51**, 185–195 (1995).
591. D. Mandal, S.K. Pal, K. Bhattacharyya, 'Excited-State Proton Transfer of 1-Naphthol in Micelles', *J. Phys. Chem. A*, **102**, 9710–9714 (1998).
592. P. Kaewsuya, N.D. Danielson, D. Ekhterae, 'Fluorescent Determination of Cardiolipin Using 10-N-Nonyl Acridine Orange', *Anal. Bioanal. Chem.*, **387**, 2775–2782 (2007).
593. T. Wolff, 'The Solvent Dependent Fluorescence Quantum Yield of Acridine as a Probe for Water in Micelles and for the Preferred Location of Acridine in Micellar Solutions', *Ber. Bunsen-Gen. Phys. Chem.*, **85**, 145–148 (1981).
594. D.A. Kendall, R.C. MacDonald, 'Characterization of a Fluorescence Assay to Monitor Changes in the Aqueous Volume of Lipid Vesicles', *Anal. Biochem.*, **134**, 26–33 (1983).
595. M.A. Kempfle, R.F. Muller, R. Palluk, H.A. Winkler, 'The Binding of Fluorescent 4,6,8(14)-Triene-3-one

Steroids to Cyclodextrins as a Model for Steroid – Protein Interactions', *Biochim. Biophys. Acta*, **923**, 83–87 (1987).

596. K. Muthuramu, V.R. Murthy, 'Photodimerization of Coumarin in Aqueous and Micellar Media', *J. Org. Chem.*, **47**, 3976–3979 (1982).

597. J.H. Ayala, A.M. Afonso, V. Gonzalez, 'Effects of Cetylpyridinium Bromide Micelles on the Spectrofluorimetric Characteristics of Polycyclic Aromatic Hydrocarbons', *Talanta*, **44**, 257–267 (1997).

598. K.K. Karukstis, S.W. Suljak, P.J. Waller, J.A. Whiles, E.H.Z. Thompson, 'Fluorescence Analysis of Single and Mixed Micelle Systems of SDS and DTAB', *J. Phys. Chem.*, **100**, 11125–11132 (1996).

599. W. Hui, W. Huaigong, 'A Fluorescence Synergistic Method for the Determination of Stilboestrol', *Anal. Lett.*, **28**, 1793–1798 (1995).

600. F. Garcia Sanchez, J.C. Marques Gomez, M. Hernandez Lopez, 'Spectrofluorimetric Determination of Zirconium Based on the Inclusion Complex of 8-Hydroxy-7-Iodoquinolone Sulfonic Acid with α- and β-Cyclodextrin as Ligand', *Anal. Lett.*, **23**, 923–933 (1990).

601. J. Yang, H. Ge, N. Jie, X. Ren, H. Zou, 'The Luminescence System of Yttrium(III)-BPMPHD – CTAB and the Determination of Yttrium(III)', *Talanta*, **41**, 2055–2060 (1994).

602. M.P. Aguilar-Caballos, A. Gomez-Hens, D. Perez-Bendito, 'Simultaneous Determination of Benzoic Acid and Saccharin in Soft Drinks by Using Lanthanide-Sensitized Luminescence', *Analyst*, **124**, 1079–1084 (1999).

603. V.A. Petrov, G.E. Dobretsov, B.O. Glotov, Y.A. Vladimirov, '4-Dimethylaminochalcone and 3-Methoxybenzanthrone as Fluorescent Probes to Study Biomembranes. 4. Effect of Cholesterol on the State of Water Bound with Membrane', *Stud. Biophys.*, **89**, 171–178 (1982).

604. J.R. Darwent, W.J. Dong, C.D. Flint, N.W. Sharpe, 'Intermolecular Energy-Transfer Between Phenanthrene and Lanthanide Ions in Aqueous Micellar Solution', *J. Chem. Soc., Faraday Trans.*, **89**, 873–880 (1993).

605. I.R. Politzer, K.T. Crago, T. Hampton, J. Joseph, J.H. Boyer, M. Shah, 'Effect of β-Cyclodextrin on the Fluorescence, Absorption and Lasing of Rhodamine 6G, Rhodamine B and Fluorescein Disodium Salt in Aqueous Solutions', *Chem. Phys. Lett.*, **159**, 258–262 (1989).

606. I.U. Aguirresacona, F.L. Arbeloa, I.L. Arbeloa, 'Spectroscopic and Thermodynamic Study on the Aggregation of Rhodamines in Solution – Solvent Effect', *J. Chem. Educ.*, **66**, 866–869 (1989).

607. O. Valdes-Aguilera, D.C. Neckers, 'Aggregation Phenomena in Xanthene Dyes', *Acc. Chem. Res.*, **22**, 171–177 (1989).

608. C. Leal, M. Granados, M.D. Prat, R. Compano, 'Labelling of Organotin Compounds for Fluorimetric Detection', *Talanta*, **42**, 1165–1170 (1995).

609. H.L.Z. Wang, 'Fluorimetric Determination of Silver with Eosin B – 1,10-Phenanthroline and Nonionic Surfactant', *Fenxi Huaxue*, **14**, 773–775 (1986).

610. W.G. Herkstroeter, P.A. Martic, S. Farid, 'Inclusion by Cyclodextrins to Control Dye Aggregation Equilibria in Aqueous Solution', *J. Am. Chem. Soc.*, **112**, 3583–3589 (1990).

611. J. Jung, D.H. Kim, 'α-Tocopherol as a Fluorescent Probe. 1. Its Spectroscopic Characterization and Use as Probe of Liposome', *Hanguk Nonghwa Hakhoe Chi.*, **25**, 224–231 (1982).

612. M.I. Snegov, A.S. Cherkasov, 'Effect of Concentration of the Components of Aqueous Micellar Solutions of Rhodamine 6G on Their Spectral Luminescent Properties', *Opt. Spektrosk.*, **49**, 67–71 (1980).

613. C. Xu, Z. Tong, 'Hydrophobic Effect on Photochemical and Photophysical Processes. X. Aggregation and Excimer Formation of 2,2-Binaphthyl, a Nonpolar Molecule, in Aggregating Solvent', *Ganguang Kexue Yu Kuang Huaxue*, **1**, 72–76 (1990).

614. K.L. Kendrick, W.R. Gilkerson, 'The State of Aggregation of Methyl Orange in Water', *J. Solution Chem.*, **16**, 257–267 (1987).

615. O.A. Galutva, L.I. Nekrasov, 'Physicochemical Properties of Chlorophyll in Model Systems Containing Surface-Active Substances', *Russ. J. Phys. Chem.*, **56**, 267–277 (1982).

616. G. Zhu, Z. Si, W. Jiang, W. Zhu, B. Wang, 'Fluorescence Enhancement of the Europium – Lanthanum – Diphacinone Ammonia System', *Anal. Lett.*, **24**, 2237–2244 (1991).

617. R.D. Jee, 'Study of Micellar Solutions to Enhance the Europium-Sensitized Luminescence of Tetracyclines', *Analyst*, **120**, 2867–2872 (1995).

618. N. Arnaud, J. Georges, 'Improved Detection of Salicylic Acids Using Terbium-Sensitized Luminescence in Aqueous Micellar Solutions of Cetyltrimethylammonium Chloride', *Analyst*, **124**, 1075–1078 (1999).

619. R. Milofsky, S. Spaeth, 'Determination of Orotic Acid in Aqueous Solutions by Micellar Electrokinetic Capillary Chromatography Using Sensitized Lanthanide-Ion Luminescence Detection', *Chromatographia*, **42**, 12–16 (1996).

620. M. De La Guardia, F. Rodilla, 'Interaction of Diazepam with Surfactants. Spectrophotometric and Spectrofluorometric Study', *J. Mol. Struct.*, **143**, 493–496 (1986).

621. F. Salinas, A.M. de la Pena, M.S. Duran, 'Determination of Aluminium with 8-Hydroxyquinoline-5-Sulfonic Acid in Presence of a Cationic Surfactant by First and Second

Derivative Synchronous Fluorimetry', *Anal. Lett.*, **21**, 1457–1468 (1988).

622. J.M. Escriche, M.D. Cirugeda, F.H. Hernandez, 'Increase in the Sensitivity of the Fluorescent Reaction of the Complexing of Aluminum with Morin Using Surfactant Agents', *Analyst*, **108**, 1386–1391 (1983).

623. C.C. Blanco, A.M.G. Campana, F.A. Barrero, M.R. Ceba, 'Micellar Medium for the Analysis of Complex Mixtures of Molybdenum and Tungsten by Derivative Synchronous Spectrofluorimetry in Steels', *Talanta*, **42**, 1037–1044 (1995).

624. R. Compano, A. Grima, A. Izquierdo, M.D. Prat, 'Determination of Zinc in Foods by Fluorescence Spectroscopy in Micellar Media', *Anal. Chim. Acta*, **227**, 219–226 (1989).

625. K. Kina, K. Tamura, N. Ishibashi, 'Effect of Longchain Quaternary Ammonium Salt on the Fluorimetric Determination of Zinc Using Oxine-5-Sulfonic Acid', *Bunseki Kagaku*, **23**, 1404–1406 (1974).

626. A.T. Pilipenko, A.I. Volkova, G.I. Pshenko, V.P. Denisenko, 'Increase of Sensitivity of Fluorescent Reactions of Aluminum, Gallium and Indium Complexing with Lumogallion Upea by Means of Cationic SAS', *Ukr. Khim. Zh.*, **46**, 200–204 (1980).

627. A. Adenier, J.J. Aaron, 'A Spectroscopic Study of the Fluorescence Quenching Interactions between Biomedically Important Salts and the Fluorescent Probe Merocyanine 540', *Spectrochim. Acta, Part A*, **58**, 543–551 (2002).

628. A. Sanz-Medel, J.I. Garcia Alonso, 'Spectrofluorimetric Determination of Niobium with Morin Enhanced by Cetyltrimethylammonium Bromide Micelles', *Anal. Chim. Acta*, **165**, 159–169 (1984).

629. L. Ebdon, J.I.G. Alonso, 'Determination of Tributyltin Ions in Estuarine Waters by High-Performance Liquid-Chromatography with Fluorometric Detection Using Morin in a Micellar Solution', *Analyst*, **112**, 1551–1554 (1987).

630. Q.G. Li, J.G. Xu, X.Z. Huang, G.Z. Chen, 'The Effects of Media Properties on the Horseradish Peroxidase-Catalyzed Fluorogenic Reaction', *Talanta*, **41**, 2049–2054 (1994).

631. T. Perez-Ruiz, C. Martinez-Lozano, M.D. Garcia, J. Martin, 'High Performance Liquid Chromatography-Photochemical Reduction in Aerobic Conditions for Determination of K Vitamins Using Fluorescence Detection', *J. Chromatogr. A*, **1141**, 67–72 (2007).

632. G. Sanchez, F. Marquez Gomez, J.C.M. Hernandez Lopez, 'Spectrofluorimetric Determination of Zirconium Based on the Inclusion Complex of 8-Hydroxy-7-Iodo Quinoline Sulfonic Acid with β-Cyclodextrin as Ligand', *Anal. Lett.*, **23**, 923–933 (1990).

633. M. Falcon, J. Guiteras, A. Izquierdo, M.D. Prat, 'Spectrofluorimetric Determination of Zinc in Foods with 8-(p-Toluenesulphonamido)Quinoline in Micellar Medium', *Talanta*, **40**, 17–20 (1993).

634. M.D. Prat, J. Guiteras, R. Compañó, J.L. Beltrán, 'Spectrofluorimetric Method for the Simultaneous Determination of Zinc and Cadmium', *J. Fluoresc.*, **1**, 267–272 (1991).

635. I. Duran-Meras, A.M.D.L. Pena, F. Salinas, I.R. Caceres, 'Spectrofluorimetric Study of the Inclusion Complex of 7-Hydroxymethylnalidixic Acid with γ-Cyclodextrin in Aqueous Solution', *Appl. Spectrosc.*, **51**, 684–688 (1997).

636. T. Imasaka, K. Nishitani, N. Ishibashi, 'Determination of Anthracene Derivatives by Cyclodextrin-modified Micellar Electrokinetic Chromatography Combined with Helium–Cadmium Laser Fluorimetry', *Anal. Chim. Acta*, **256**, 3–7 (1992).

637. B.S. Liu, Z.C. Liu, G. Jing, 'Fluorescence Resonance Energy Transfer between Acridine Orange and Rhodamine B and Analytical Application on Determination of Vitamin B-12', *Anal. Lett.*, **38**, 1367–1377 (2005).

638. B.S. Liu, Z.C. Liu, Z.Z. Cao, 'Fluorescence Resonance Energy Transfer between Acridine Orange and Rhodamine 6G and Analytical Application in Micelles of Dodecyl Benzene Sodium Sulfonate', *J. Lumin.*, **118**, 99–105 (2006).

639. J. Hernández García, Z. Sosa Ferrera, A.J. Bermejo Martín-Lázaro, J.J. Santana Rodríguez, 'Fluorescence Behaviour of Polychlorinated Dibenzofurans in Organized Molecular Systems. Analytical Applications', *Microchim. Acta*, **118**, 185–196 (1995).

640. J.J. Santana Rodriguez, Z. Sosa Ferrera, A. Afonso Perera, V. Gonzales Diaz, 'Simultaneous Synchronous Flourimetric Determination of Benzo[a]Pyrene and Perylene in Micellar Media', *Anal. Chim. Acta*, **255**, 107–111 (1991).

641. L.F. He, D.L. Lin, Y.Q. Li, 'Micelle-Sensitized Constant-Energy Synchronous Fluorescence Spectrometry for the Simultaneous Determination of Pyrene, Benzo[a]Pyrene and Perylene', *Anal. Sci.*, **21**, 641–645 (2005).

642. T. Miura, M. Kashiwamura, M. Kimura, 'A Fluorometric Method for the Specific Determination of Serum Arginine with 2,3-Naphthalenedicarbaldehyde', *Anal. Biochem.*, **139**, 432–437 (1984).

643. C.P. McCormick, Z.K. Shihabi, 'HPLC of Fluorescent Products of Acetaminophen Reaction with Peroxidase', *J. Liq. Chromatogr. Related Technol.*, **13**, 1159–1171 (1990).

644. R.O. Cole, M.J. Sepaniak, 'Bile Salt Surfactants in MEKC-Application to Hydrophobic Molecule Separation', *LC-GC Mag.*, **10**, 380–387 (1989).

645. L.A. Spino, D.W. Armstrong, T. Vo-Dinh, 'Direct Synchronous Luminescence Detection of Co-Eluting Solutes in Pseudophase Liquid Chromatography', *J. Chromatogr. A*, **409**, 147–154 (1987).

646. T. Kaneta, T. Yamashita, T. Imasaka, 'Separation of Polycyclic Aromatic Hydrocarbons by Micellar Electrokinetic Chromatography with Laser Fluorescence Detection', *Anal. Chim. Acta*, **299**, 371–375 (1995).

647. L.Y. Zhang, M.X. Sun, 'Selective Determination of γ-Aminobutyric Acid, Glutamate and Alanine by Mixed Micellar Electrokinetic Chromatography and Fluorescence Detection', *J. Chromatogr. A*, **1095**, 185–188 (2005).

648. E. Orejuela, M. Silva, 'Rapid Determination of Aniline Metabolites of Chlorpropham in Potatoes by Micellar Electrokinetic Chromatography Using Negative-Charged Mixed Micelles and Laser-Induced Fluorescence Detection', *Electrophoresis*, **26**, 2991–2998 (2005).

649. E. Chiavaro, C. Dall'Asta, G. Galaverna, A. Biancardi, E. Gambarelli, A. Dossena, R. Marchelli, 'New Reversed-Phase Liquid Chromatographic Method to Detect Aflatoxins in Food and Feed with Cyclodextrins as Fluorescence Enhancers Added to the Eluent', *J. Chromatogr. A*, **937**, 31–40 (2001).

650. S.L. Zhao, Y.R. Song, Y.M. Liu, 'A Novel Capillary Electrophoresis Method for the Determination of D-Serine in Neural Samples', *Talanta*, **67**, 212–216 (2005).

651. J.R. Daban, A.M. Aragay, 'Rapid Fluorescent Staining of Histones in Sodium Dodecyl Sulfate-Polyacrylamide Gels', *Anal. Biochem.*, **138**, 223–228 (1984).

652. T. Takeuchi, T. Miwa, 'Visualization or Single Enhancement Via Inclusion Complexation with Cyclodextrin in Liquid Chromatography', *Kuromatogurafi*, **14**, 58–59 (1993).

653. T. Takeuchi, T. Miwa, 'Indirect Fluorimetric Detection of Cyclodextrins Via Postcolumn Mixing with 2-p-Toluidinyl-6-Naphthalenesulfonate in Liquid Chromatography', *J. High Resolut. Chromatogr.*, **22**, 609–612 (1999).

654. S. Uchiyama, M. Uchiyama, 'Thin-Layer Chromatography-Fluorometry of Ethoxyquin Using Triton X-100', *J. Chromatogr. A*, **262**, 340–345 (1983).

655. W. Baeyens, B. Ling, K. Stragier, 'HPTLC Fluorescence Enhancement of Glutathione-SBD: Effect of Cyclodextrins, Triton and Alkanes', *Chromatographia*, **29**, 243–247 (1990).

656. B.L. Ling, W.R.G. Baeyens, H. Marysael, K. Stragier, P. Demoerloose, 'Influence of Stationary Phases and Fluorescence Dipping Enhancers on the HPTLC Densitometry of SBD-Thiol Derivatives', *J. Liq. Chromatogr.*, **12**, 3135–3149 (1989).

657. A. Navas Diaz, 'Fibre-Optic Remote Sensor for in Situ Flourimetric Quantification in Thin-Layer Chromatography', *Anal. Chim. Acta*, **255**, 297–303 (1991).

658. A. Alak, E. Heilweil, W.L. Hinze, H. Oh, D.W. Armstrong, 'Effects of Different Stationary Phases and Surfactant or Cyclodextrin Spray Reagents on the Fluorescence Densitometry of Polycyclic Aromatic-Hydrocarbons and Dansylated Amino-Acids', *J. Liq. Chromatogr.*, **7**, 1273–1288 (1984).

659. F.H. Quina, W.L. Hinze, 'Surfactant-Mediated Cloud Point Extractions: An Environmentally Benign Alternative Separation Approach', *Ind. Eng. Chem. Res.*, **38**, 4150–4168 (1999).

660. G. Stangl, R. Niessner, 'Cloud Point Extraction of Napropamide and Thiabendazole from Water and Soil', *Microchim. Acta*, **113**, 1–8 (1994).

661. I.Y. Goryacheva, S.N. Shtykov, A.S. Loginov, I.V. Panteleeva, 'Preconcentration and Fluorimetric Determination of Polycyclic Aromatic Hydrocarbons Based on the Acid-Induced Cloud-Point Extraction with Sodium Dodecylsulfate', *Anal. Bioanal. Chem.*, **382**, 1413–1418 (2005).

662. M.A. Hernandez Torres, M.G. Khaledi, J.G. Dorsey, 'Micellar-Catalyzed Reactions for Flow-Injection Systems: Determination of Pyridoxal', *Anal. Chim. Acta*, **201**, 67–76 (1987).

663. F. Delattre, P. Woisel, G. Surpateanu, F. Cazier, P. Blach, '1-(4-Nitrophenoxycarbonyl)-7-Pyridin-4-yl Indolizine: A New Versatile Fluorescent Building Block. Application to the Synthesis of a Series of Fluorescent β-Cyclodextrins', *Tetrahedron*, **61**, 3939–3945 (2005).

664. R. Corradini, C. Paganuzzi, R. Marchelli, S. Pagliari, S. Sforza, A. Dossena, G. Galaverna, A. Duchateau, 'Fast Parallel Enantiomeric Analysis of Unmodified Amino Acids by Sensing with Fluorescent β-Cyclodextrins', *J. Mater. Chem.*, **15**, 2741–2746 (2005).

665. L.B. McGown, S.M. Meyerhoffer, P.R. Hertz, G. Li, 'Luminescence in Bile Salt Micelle', in *Bile Acid/Salt Surfactant Systems*, *Organized Assemblies in Chemical Analysis, Vol. 2*, ed. W.L. Hinze, JAI Press, Stamford, CT, 5, 167–186, 2000.

666. A.H. Wu, X.H. Shen, H.C. Gao, 'Investigation on Photophysical Properties of a Substituted 3H-Indole-Modified β-Cyclodextrin II. Efficient Photoinduced Energy Transfer with Naphthalene and its Derivatives', *J. Photochem. Photobiol., A*, **185**, 144–149 (2007).

667. E.M.A. Lopez, A.M. Garcia-Campana, J.J. Aaron, L.C. Rodriguez, 'Simultaneous Quantification of Chlorophenoxyacid Herbicides Based on Time-Resolved Photochemical Derivatization to Induce Fluorescence in Micellar Medium', *Talanta*, **60**, 355–367 (2003).

668. U.A.T. Brinkman, G.J. Dejong, C. Gooijer, 'Use of Luminescence Techniques for Sensitive and Selective

Determinations in HPLC (High-Performance Liquid-Chromatography)', *Pure Appl. Chem.*, **59**, 625–637 (1987).

669. T. Perez-Ruiz, C. Martinez-Lozano, V. Tomas, J. Martin, 'Flow-Injection Fluorimetric Determination of Menadione Using On-Line Photo-Reduction in Micellar Media', *Anal. Chim. Acta*, **514**, 259–264 (2004).

670. A. Coly, J.-J. Aaron, 'Sensitive and Rapid Flow Injection Analysis of Sulfonylurea Herbicides in Water with Micellar-Enhanced Photochemically Induced Fluorescence Detection', *Anal. Chim. Acta*, **392**, 255–264 (1999).

671. A. Coly, J.-J. Aaron, 'Photochemically-Induced Fluorescence Determination of Sulfonylurea Herbicides Using Micellar Media', *Talanta*, **49**, 107–117 (1999).

672. J.J. Aaron, A. Coly, 'Photochemical-Spectrofluorimetric Determination of Two Pyrethroid Insecticides Using an Anionic Micellar Medium', *Analyst*, **121**, 1545–1549 (1996).

673. A. Bautista, J.-J. Aaron, M.C. Mahedero, A. Muñoz de la Peña, 'Usefulness of Micellar Media for the Quantitative Analysis of Phenylurea Herbicides in Water by Photochemically-Induced Fluorescence', *Analusis*, **27**, 857–863 (1999).

674. M.C. Mahedero, A.M. De La Pena, A. Bautista, J.J. Aaron, 'An Investigation of Inclusion Complexes of Cyclodextrins with Phenylurea Herbicides by Photochemically-Induced Fluorescence. Analytical Applications', *J. Incl. Phenom. Macro*, **42**, 61–70 (2002).

675. A.M. Garcia-Campana, J.J. Aaron, J.M. Bosque-Sendra, 'Micellar Enhanced Photochemically Induced Fluorescence Detection of Chlorophenoxyacid Herbicides. Flow Injection Analysis of Mecoprop and 2,4-Dichlorophenoxyacetic Acid', *Talanta*, **55**, 531–539 (2001).

676. S. Irace-Guigand, E. Leverend, M.D.G. Seye, J.J. Aaron, 'A New On-Line Micellar-Enhanced Photochemically-Induced Fluorescence Method for Determination of Phenylurea Herbicide Residues in Water', *Luminescence*, **20**, 138–142 (2005).

677. J.R. Poulsen, J.W. Birks, 'Photochemical Reaction Detection in HPLC', in *Chemiluminescence and Photochemical Reaction Detection in Chromatography*, ed. J.W. Birks, VCH Publishers, New York, 149–230, Chapter 6, 1989.

678. E.B. Gonzalez, A.S. Medel, 'Surfactant Sensitized Colorimetric and Fluorimetric Reactions – New Analytical Reactions for Ti(IV) and Ta(V)', *An. Quim. Ser. B*, **81**(1), 60–68 (1985).

679. A. Sanzmedel, R.F. Delacampa, J.I.G. Alonso, 'Metal Chelate Fluorescence Enhancement in Micellar Media – Mechanisms of Surfactant Action', *Analyst*, **112**, 493–497 (1987).

680. V.A. Pilipenko, T.A. Vasilchuk, 'Fluorescence Reactions of Zirconium and Hafnium with Morin in Micellar Solutions of Surface-Active Substances and in Aqueous-Organic Media', *J. Anal. Chem.*, **41**, 1103–1109 (1986).

681. W. Cui, Z. Han, H. Shi, 'Sensitization by Surfactants on the Fluorescence Reactions of Metals with Morin', *Gaodeng Xuexiao Huaxue Huebao*, **7**, 672–676 (1986).

682. K.J. Kim, B. Kim, B.-G. Lee, J.W. Park, 'Effects of Surfactants on the Fluorescence of Beryllium-Morin Complex', *Bull. Korean Chem. Soc.*, **9**, 106–108 (1988).

683. S. Can, S. Zhang, S. Shou, 'Spectrofluorimetric Determination of Trace Niobium with Morin Enhanced by Hexadecyltrimethylammonium Bromide Micelles', *Fenxi Huaxue*, **16**, 448–450 (1988).

684. A.T. Pilipenko, A.I. Vlkova, 'Fluorescence Reaction of Zirconium and Hafnium with Morin in Micallar Solutions of Surfactants and Aqueous-Organic Media', *Zn. Anal. Khim.*, **41**, 1608–1614 (1986).

685. W.C. Cui, Z.J. Han, H. Shi, 'Sensitization by Surfactants on the Fluorescence Reactions of Metals with Morin', *Gaodeng Xuexiao Huaxue Huebao*, **7**, 672–676 (1986).

686. K. Soroka, R.S. Vithanage, D.A. Phillips, B. Walker, P.K. Dasgupta, 'Fluorescence Properties of Metal Complexes of 8-Hydroxyquinoline-5-Sulfonic Acid and Chromatographic Applications', *Anal. Chem.*, **59**, 629–636 (1987).

687. D.A. Phillips, K. Soroka, R.S. Vithanage, P.K. Dasgupta, 'Enhancement and Quenching of Fluorescence of Metal Chelates of 8-Hydroxyquinoline-5-Sulfonic Acid', *Microchim. Acta*, **88**, 207–220 (1986).

688. A.M. Delapena, F. Salinas, M.E. Sanchez, J.A. Murillo, 'Simultaneous Determination of Aluminum and Zinc by 1st Derivative and 2nd Derivative Synchronous Fluorimetry', *Analyst*, **113**, 1435–1438 (1988).

689. W. Cui, L. Mi, H. Shi, 'Sensitization Effect of Surfactants on the Fluorescence Reaction of Magnesium 8-Hydroxyquinoline-5-Sulfonic Acid and its Applications', *Huaxue Shiji*, **7**, 125–128 (1985).

690. W. Cui, L. Kong, H. Shi, 'Study of Fluorescence Reaction of Aluminum with 8-Hydoxy-7-Indoquinoline-5-Sulfonic Acid and Surfactants and its Applications', *Yejin Fenxi*, **12**, 1–4 (1992).

691. S. Jiang, N. Jie, H. Shi, 'Fluorimetry of the Complex of Zinc with 8-Hydroxy-7-Iodoquinoline-5-Sulfonic Acid and Surfactant and its Application to Determination of Zinc in Alloys', *Fenxi Huaxue*, **14**, 934–936 (1986).

692. K. Kina, K. Shiraishi, N. Ishibashi, 'A New Azo-Compound Containing Long Alkyle Chain for Fluorometric Determination of Trace Gallium', *Bunseki Kagaku*, **28**, 501–505 (1976).

693. G. de Armas, M. Miro, J.M. Estela, V. Cerda, 'Multisyringe Flow Injection Spectrofluorimetric

Determination of Warfarin at Trace Levels with on-Line Solid-Phase Preconcentration', *Anal. Chim. Acta*, **467**, 13–23 (2002).

694. Z.T. Zeng, R.A. Jewsbury, 'Fluorimetric Determination of Iron Using 5-(4-Methoxyphenylazo)-8-(4-Toluenesulfonamido)Quinoline', *Analyst*, **125**, 1661–1665 (2000).
695. M. Sato, H. Yoshimura, H. Obi, S. Hatakeyama, E. Kaneko, H. Hoshino, T. Yotsuyanagi, 'Determination of Aluminium in Human Serum by Kinetic Differentiation Mode Reversed Phase Micellar HPLC with Fluorometric Detection', *Chem. Lett.*, 203–204 (1996).
696. N. Memon, M.I. Bhanger, 'Micellar Liquid Chromatographic Determination of Aluminum as its Complex with 8-Hydroxyquinoline-5-Sulfonic Acid', *Acta Chromatographica*, **14**, 172–179 (2004).
697. J.L. Ren, J. Zhang, J.Q. Luo, X.K. Pei, Z.X. Jiang, 'Improved Fluorimetric Determination of Dissolved Aluminium by Micelle-Enhanced Lumogallion Complex in Natural Waters', *Analyst*, **126**, 698–702 (2001).
698. S.K. Kina, N. Ishabashi, 'A New Azo-Compound Containing a Long Alkyl Chain for Fluorometric-Determination of Trace Gallium', *Bunseki Kagaku*, **25**, 501–505 (1976).
699. K. Kina, K. Hirokata, N. Ishabashi, 'Preparation and Application of a Surface-Active Azo Compound for Fluorometric-Determination of Trace Gallium', *Bunseki Kagaku*, **26**, 246–252 (1977).
700. K. Watanabe, Y. Tomizawa, M. Itagaki, 'Fluorescence Properties of Aluminium(III)-Lumogallion Complex in Nonionic Surfactant Micelle', *Bunseki Kagaku*, **45**, 845–850 (1996).
701. E.B. Gonzalez, S.A. Medel, 'Enhancement of Metal Chelate Fluorescence in Surfactant System', *An. Quim. A-Fis. Tech.*, **81**, 60–65 (1985).
702. M. Vosmanska, Z. Holzbecher, 'Fluorimetric Determination of Aluminium, Beryllium, Gallium and Zinc with 2-(2-Hydroxyphenyl) Benzoxazole', *Sb. Vys. Sk. Chem. Techn*, **H22**, 85–101 (1988).
703. C.C. Kalambaheti, P. Wilairat, N. Sakulpaisith, P. Visoottviseth, 'Determination of Phenyltin Compounds by Reversed Phase Liquid Chromatography', *Appl. Organomet. Chem.*, **12**, 1–12 (1998).
704. W. Kleibohmer, K. Cammann, 'Fluorescence Enhancement in Micellar Systems – Fluorimetric Determination of Organotin Compounds with Morine', *Fresenius Z. Anal. Chem.*, **335**, 775–779 (1989).
705. E.M.R. Rodriguez, M.S. Alaejos, C.D. Romero, 'Fluorimetric Determination of Selenium in Biological Samples', *Anal. Lett.*, **32**, 1699–1721 (1999).
706. Y.-X. Zheng, D.-H. Lu, 'Synergistic Enhancement of the Fluorescence Intensity of 4,5-Benzopiaselenol by Surfactants and β-Cyclodextrin in Aqueous Solution', *Microchim. Acta*, **106**, 3–9 (1992).
707. A.T. Pilipenko, T.L. Shevchenko, A.I. Volkova, 'Fluorescent Ionic Associates of IREA Lumogallion with Cetylpyridinium Chloride and Ethonium', *Ukr. Khim. Zh.*, **52**, 163–167 (1986).
708. A.T. Pilipenko, G.N. Pshinko, A.I. Volkova, 'Fluorescent Method for Determination of Anionic Surfactants in Water', *Khim. Tekhnol. Vody*, **1**, 35–39 (1979).
709. J.L.C. Dominguez, M.D. Cirugeda, 'Some Observations on the Fluorometric-Determination of Metallic Elements in Micellar Media', *Microchem. J.*, **39**, 50–58 (1989).
710. D.Y. Sasaki, B.E. Padilla, 'Dithioamide Metal Ion Receptors on Fluorescent Lipid Bilayers for the Selective Optical Detection of Mercuric Ion', *Chem. Commun.*, 1581–1582 (1998).
711. S. Jiang, N. Jie, C. Cheng, 'Flurorimetric Determination of Micro Amounts of Scandium with 8-Hydoxy-7-Iodoquinoline-5-Sulphonic Acid and Hexadecyltrimethylammonium Bromide', *Fenxi Huaxue*, **16**, 236–238 (1988).
712. W.J. Dressick, J.N. Demas, B.A. Degraff, 'Luminescence Intensity as a Probe of the Interaction of Transition-Metal Photosensitizers with Micelles', *J. Photochem.*, **24**, 45–52 (1984).
713. M.F. Sonnenschein, 'Influence of Curvature on the Concentration of Oxygen in $DC_{23}PC$ Bilayers', *J. Colloid Interface Sci.*, **138**, 583–586 (1990).
714. K. Mandal, B.L. Hauenstein, J.N. Demas, B.A. DeGraff, 'Interactions of Ruthenium(II) Photosensitizers with Triton X-100', *J. Phys. Chem.*, **87**, 328–331 (1983).
715. L.V. Vargas, J. Sand, T.A.S. Brandao, H.D. Fiedler, F.H. Quina, F. Nome, 'Determination of Environmentally Important Metal Ions by Fluorescence Quenching in Anionic Micellar Solution', *Analyst*, **130**, 242–246 (2005).
716. L.H. Feng, Z.B. Chen, D.S. Wang, 'Selective Sensing of Fe^{3+} Based on Fluorescence Quenching by 2,6-Bis(Benzoxazolyl)Pyridine with β-Cyclodextrin in Neutral Aqueous Solution', *Spectrochim. Acta, Part A*, **66**, 599–603 (2007).
717. A. Balouch, M.I. Bhanger, K.N. Thimmaiah, W.L. Hinze, 'Sensitivity Enhancement of Quenchofluorometric Methods for the Determination of Ionic and Neutral Analytes Via Micellar Amplification of Quenching', *Journal of Luminescence*, (to be submitted).
718. Y. Diaz-Fernandez, A. Perez-Gramatges, S. Rodriguez-Calvo, C. Mangano, P. Pallavicini, 'Structure and Dynamics of Micelle-Based Fluorescent Sensor for Transition Metals', *Chem. Phys. Lett.*, **398**, 245–249 (2004).
719. J. Georges, 'Lanthanide-Sensitized Luminescence and Applications to the Determination of Organic Analytes: A Review', *Analyst*, **118**, 1481–1486 (1993).

720. P.G. Sammes, G. Yahioglu, 'Modern Bioassays Using Metal Chelates as Luminescent Probes', *Natl. Prod. Rep.*, **13**, 1–28 (1996).

721. F. Canada-Canada, M.I. Rodriguez-Caceres, 'Spectrofluorimetric Determination of 3-Hydroxy-2-Naphthoic Acid by Use of its Ternary Complex with Zirconium (IV) and β-Cyclodextrin: Application to Determination in River Water', *J. Fluoresc.*, **17**, 23–28 (2007).

722. I. Hemmila, V. Laitala, 'Progress in Lanthanides as Luminescent Probes', *J. Fluoresc.*, **15**, 529–542 (2005).

723. P.R. Selvin, 'Principles and Biophysical Applications of Lanthanide-Based Probes', *Annu. Rev. Biophys. Biomol. Struct.*, **31**, 275–302 (2002).

724. J.R. Darwent, C.D. Flint, N.W. Sharpe, 'Intermolecular Energy Transfer from Phenanthrene to Europium in Aqueous Micellar Solution', *J. Chem. Soc., Chem. Commun.*, 747–748 (1988).

725. J.R. Darwent, C.D. Flint, N.W. Sharpe, 'Intermolecular Energy-Transfer from Phenanthrene to Europium in Aqueous Micellar Solution', *J. Chem. Soc., Chem. Commun.*, 747–748 (1988).

726. W. Jiang, Y.D. Ma, W.Y. Zhao, Y. Feng, N.X. Wang, Z.K. Si, 'Fluorimetric Determination of Europium Based on the Co-Fluorescence System Eu-Y-Sparfloxacin-SDS', *Microchim. Acta*, **144**, 29–34 (2004).

727. J. Georges, J.M. Mermet, 'Simultaneous Time-Resolved Fluorescence and Thermal Lens Measurements: Application to Energy Transfer Studies in Europium Chelates', *Spectrochim. Acta, Part A*, **49**, 397–404 (1993).

728. J. Georges, 'Investigation of Fluorescence Efficiency in the Europium-Thenoyltrifluoroacetone Chelate in Aqueous and Ethanolic Solutions by Laser-Induced Fluorescence and Photothermal Spectroscopic Methods', *Anal. Chim. Acta*, **317**, 343–351 (1995).

729. X.Z. Du, J.U. Hou, H.L. Deng, J.Z. Gao, J.W. Kang, 'Comparative Study on Fluorescence Enhancement and Quenching of Europium and Terbium Chelate Anions in Cationic Micelles', *Spectrochim. Acta, Part A*, **59**, 271–277 (2003).

730. F. Wang, W. Huang, Y. Hou, Z.N. Xu, 'The Co-Luminescence Effect of Eu-Gd-Ofloxacin-SDBS System and its Analytical Application', *J. Fluoresc.*, **17**, 105–111 (2007).

731. M.C. Mahedero, D. Bohoyo, F. Salinas, T. Ardila, D. Airado, B. Roldan, 'Improvements in Europium Sensitized Fluorimetric Determination of Demeclocycline and Methacycline', *J. Pharm. Biomed. Anal.*, **37**, 1101–1104 (2005).

732. I. Hemmila, 'Time-Resolved Fluorometric Determination of Terbium in Aqueous Solution', *Anal. Chem.*, **57**, 1676–1681 (1985).

733. T. Taketatsu, 'Complex-Formation of Europium with Thenoyltrifluoroacetone and Tri-Normal-Octylphosphine Oxide in Micellar Solution of Non-Ionic Surfactant', *Chem. Lett.*, 1057–1058 (1981).

734. N. Arnaud, J. Georges, 'Fluorimetric Determination of Europium over a Large Dynamic Range Using its Ternary Complex with Thenoyltrifluoroacetone and Trioctylphosphine Oxide in a Micellar Solution of Triton X-100', *Analyst*, **122**, 143–146 (1997).

735. S. Peter, B.S. Panigrahi, K.S. Viswanathan, C.K. Mathews, 'Fluorescence Enhancement of Dysprosium, Europium and Terbium Using Sodium Benzoate-Trioctylphosphine Oxide-Triton X-100', *Anal. Chim. Acta*, **260**, 135–141 (1992).

736. Z. Si, W. Jiang, J. Hu, Q. Jin, 'Study on the Europium-2-Benzoyl-Indan-1,3-Dione-Trioctyl-Phosphine Oxide-Triton X-100 Fluorescence System and its Analytical Application', *Anal. Lett.*, **29**, 303–311 (1996).

737. Y.X. Ci, Z.H. Lan, 'Fluorescence Enhancement of the Europium(III) – Thenoyltrifluoroacetone – Trioctylphosphine Oxide Ternary Complex by Gadolinium(III) and its Application to the Determination of Europium(III)', *Analyst*, **113**, 1453–1457 (1988).

738. G. Zhu, Z. Si, P. Liu, W. Jiang, 'Study of the Fluorescence Enhancement System Europium-Gadolinium-Thenoyltrifluoroacetone-Cetyltrimethylammonium Bromide-Triton X-100 and its Application', *Anal. Chim. Acta*, **247**, 37–43 (1991).

739. G. Zhu, J. Yang, S. Xu, Y. Niu, 'Terbium-Iminodiacetic Acid-Tiron-Surfactant Fluorescent System', *Gaodeng Xuexiao Huaxue Huebao*, **7**, 121–124 (1986).

740. S. Sun, M. Lin, 'Fluorescence Characteristics of Ternary Complexes of Europium (Terbium)-Trifluoroacetylacetone-TOPO-Nonionic Surfactant System', *Zhongguo Xitu Xuebao*, **6**, 93–94 (1988).

741. C. Moulin, P. Decambox, P. Mauchien, 'Determination of Curium at 0.1 ng l^{-1} Levels in a Micellar Medium by Time-Resolved Laser-Induced Spectrofluorimetry', *Anal. Chim. Acta*, **254**, 145–151 (1991).

742. T. Taketatsu, 'Partition of Europium (III) with β-Diketones and Neutral Additives between Micellar and Bulk Phases in Aqueous Nonionic Surfactant Solutions', *Anal. Chim. Acta*, **174**, 323–326 (1985).

743. C. Moulin, P. Reiller, C. Beaucaire, D. Lemordant, 'Time-Resolved Laser-Induced Spectrofluorometry for the Study of Uranium-Anionic Surfactant Micelle Interactions', *J. Colloid Interface. Sci.*, **157**, 411–417 (1993).

744. T.R. Herrmann, A.R. Jayaweera, A.E. Shamoo, 'Interaction of Europium(III) with Phospholipid Vesicles as Monitored by Laser-Excited Europium(III) Luminescence', *Biochemistry*, **25**, 5834–5838 (1986).

745. H.N. Halladay, M. Petersheim, 'Optical Properties of Terbium(III)-Phospholipid Complexes and Their Relation to Structure', *Biochemistry*, **27**, 2120–2126 (1988).

746. H.F. Brito, C.A.A. Carvalho, O.L. Malta, J.J. Passos, J.F.S. Menezes, R.D. Sinisterra, 'Spectroscopic Study of the Inclusion Compound of β-Cyclodextrin and Tris(Dibenzoylmethane)Europium(III) Dihydrate', *Spectrochim. Acta, Part A*, **55**, 2403–2410 (1999).

747. J.P. Andre, E. Toth, H. Fischer, A. Seelig, H.R. Macke, A.E. Merbach, 'High Relaxivity for Monomeric Gd(DOTA)-Based MRI Contrast Agents, Thanks to Micellar Self-Organization', *Chem. Eur. J.*, **5**, 2977–2983 (1999).

748. T. Taketatsu, A. Sato, 'Spectrofluorimetric Determination of Europium and Samarium with Mixtures of 2-Thenoyltrifluoroacetone and Tri-N-Octylphosphine Oxide in Aqueous Solutions Containing Triton X-100', *Anal. Chim. Acta*, **108**, 429–432 (1979).

749. N.S. Poluektov, S.B. Meshkova, Y.V. Korovin, 'Luminescence of Ytterbium(III) Complexes with Certain Heterocyclic Azo Compounds and Surfactants', *Ukr. Khim. Zh.*, **53**, 844–847 (1987).

750. T. Taketatsu, 'Spectrophotofluorimetric Determination of Terbium, Europium and Samarium with Pivaloyltrifluoroacetone and Tri-N-Octylphosphine Oxide in Micellar Solution of Nona-Oxyethylene Dodecyl Ether', *Talanta*, **29**, 397–400 (1982).

751. M. Aihara, M. Arai, T. Taketatsu, 'Flow-Injection Spectrofluorimetric Determination of Europium(III) Based on Solubilizing its Ternary Complex with Thenoyltrifluoroacetone and Trioctylphosphine Oxide in Micellar Solution', *Analyst*, **111**, 641–643 (1986).

752. S.N. Shtykov, T.D. Smirnova, Y.G. Bylinkin, N.V. Kalashnikova, D.A. Zhemerichkin, 'Determination of Ciprofloxacin and Enrofloxacin by the Sensitized Fluorescence of Europium in the Presence of the Second Ligand and Micelles of Anionic Surfactants', *J. Anal. Chem.*, **62**, 136–140 (2007).

753. H. Hoshino, S.Y. Utsumi, T. Yotsuyanagi, 'Aromatic Polyaminocarboxylate Ligands for Energy Transfer Luminescence Measurements of Lanthanide Ions in Aqueous Solutions', *Talanta*, **41**, 93–98 (1994).

754. Y. Jing He, Z. Gui-yun, W. Bo, 'Enhanced Luminescence of the Europium/Terbium/Thenoyltrifluoroacetone/1,10-Phenanthroline/Surfactant System, and its Analytical Application', *Anal. Chim. Acta*, **198**, 287–292 (1987).

755. W. Li, W. Li, G. Yu, Q. Wang, Y. Jin, 'Enhanced Luminescence and Energy Transfer of Eu(III) by Tb(III) in Chelates in Micelle Solutions', *J. Alloys Compd.*, **191**, 107–110 (1993).

756. H.G. Huang, X.X. Geng, 'Fluorometric-Determination of Samarium and Europium with 2-Thenoyltrifluoroacetone and 1,10-Phenanthroline Micellar System', *Bunseki Kagaku*, **35**, 579–584 (1986).

757. T. Berthoud, P. Decambox, B. Kirsch, P. Mauchien, C. Moulin, 'Direct Determination of Traces of Lanthanide Ions in Aqueous Solutions by Laser-Induced Time-Resolved Spectrofluorimetry', *Anal. Chim. Acta*, **220**, 235–241 (1989).

758. Y.-Y. Xu, I.A. Hemmila, 'Analytical Application of the Co-Fluorescence Effect in Detection of Europium, Terbium, Samarium and Dysprosium with Time-Resolved Fluorimetry', *Talanta*, **39**, 759–763 (1992).

759. G. Zhu, Z. Si, J. Yang, J. Ding, 'Simultaneous Spectrofluorimetric Determination of Terbium, Samarium and Europium with Hexafluoroacetylacetone-Trioctylphosphine Oxide and Triton X-100', *Anal. Chim. Acta*, **231**, 157–159 (1990).

760. B.S. Panigrahi, S. Peter, K.S. Viswanathan, C.K. Mathews, 'Fluorescence Enhancement and Cofluorescence in Complexes of Terbium, Dysprosium and Europium with Trimesic Acid', *Anal. Chim. Acta*, **282**, 117–124 (1993).

761. N. Arnaud, J. Georges, 'Influence of pH, Surfactant and Synergic Agent on the Luminescent Properties of Terbium Chelated with Benzoic Acid Derivatives in Aqueous Solutions', *Analyst*, **125**, 1487–1490 (2000).

762. J. Hu, G. Chen, Y. Zeng, 'Laser-Induced Fluorimetry. V. Dysprosium-Trifluoroacetylacetone-Topo-Tween-20 Fluorescent System and its Analytical Application', *Yingyong Huaxue*, **7**, 96–98 (1990).

763. G. Zhu, J. Yang, Z. Si, 'Study of the Fluorescent Dysprosium/EDTA/Tiron/Surfactant System', *Anal. Chim. Acta*, **215**, 331–335 (1988).

764. M. Aihara, M. Arai, T. Tomitsugu, 'Flow Injection Spectrofluorimetric Determination of Terbium (III) Based on Solubilizing its Ternary Complex in Micellar Solution', *Anal. Lett.*, **19**, 1907–1916 (1986).

765. I. Mori, T. Matsuo, M. Toyoda, Y. Fujita, C. Matoba, K. Akagiri, 'Application of o-Hydroxyhydroquinonephthalein for the Fluorophotometric Determination of Zirconium(IV) and Gallium(III) in the Presence of Tetradecyltrimethylammonium Chloride', *Anal. Lett.*, **28**, 649–662 (1995).

766. Y.Y. Xu, I.A. Hemmila, T.N.E. Lovgren, 'Co-Fluorescence Effect in Time-Resolved Fluoroimmunoassays: A Review', *Analyst*, **117**, 1061–1069 (1992).

767. Y.-Y. Xu, I.A. Hemmila, 'Co-Fluorescence Enhancement System Based on Pivaloyltrifluoroacetone and Yttrium for the Simultaneous Detection of Europium, Terbium, Samarium and Dysprosium', *Anal. Chim. Acta*, **256**, 9–16 (1992).

768. J.H. Yang, G.Y. Zhu, H. Wang, 'Application of the Columinescence Effect of Rare-Earths: Simultaneous

Determination of Trace Amounts of Samarium and Europium in Solution', *Analyst*, **114**, 1417–1419 (1989).

769. G. Zhu, J. Yang, L. Wang, 'Co-Luminescence Effect of Rare Earths: The Samarium-Gadolinium-TTA [2-Thenoyltrifluoroacetone]-Phen [1,10-Phenanthroline]-Triton X-100 Fluorescence System', *Shandong Daxue Xuebao, Ziran Kexueban*, **23**, 56–64 (1988).

770. J. Yang, G. Zhu, B. Zhang, 'Co-Luminescence Effect of Rare Earths: The Co-Luminescence Effect of Europium-Gadolinium-2-Thenoyltrifluoroacetone-Phenanthroline-Triton X-100 System and its Application', *Fenxi Huaxue*, **16**, 29–33 (1988).

771. Y.-X. Ci, Z.-H. Lan, 'Enhanced Fluorimetric Determination of Europium(III) with Thenoyltrifluoroacetone and 4,7-Diphenyl-1,10-Phenanthroline by Gadolinium(III)', *Anal. Lett.*, **21**, 1499–1513 (1988).

772. J. Yang, G. Zhu, D. Huang, 'Studies on Columinescence Effect of Rare Earth Metals-Co-Luminescence Effect of Europium-Lutetium-Thenoyltrifluoroacetone-Phenanthroline-Triton X-100 System and its Analytical Application', *Huaxue Shiji*, **10**, 249–250 (1988).

773. L. Hu, J. Li, G. Song, H. Zhang, 'Enhanced Luminescence of Europium by Gadolinium and its Application', *Anal. Lett.*, **30**, 945–951 (1997).

774. Y.Y. Xu, I. Hemmila, V.M. Mukkala, S. Holttinen, T. Lovgren, 'Co-Fluorescence of Europium and Samarium in Time-Resolved Fluorometric Immunoassays', *Analyst*, **116**, 1155–1158 (1991).

775. E.V. Malinka, S.V. Bel'tyukova, N.A. Nazarenko, 'Preparation, Properties and Application of Lanthanoid Thenoyltrifluoroacetonate Adducts with Diaza-15-Crown-5. Fluorescence Determination of Europium', *Zh. Anal. Khim.*, **43**, 1782–1786 (1988).

776. Z. Si, G. Zhu, J. Hu, B. Zhang, W. Jiang, W. Ai, 'Fluorescence Enhancement of Rare Earths by Yttrium and its Application', *Anal. Lett.*, **27**, 1183–1191 (1994).

777. J. Wei, M. Yudao, Z. Wenyuan, F. Yu, W. Naixing, S. Zhikun, 'Fluorimetric Determination of Europium Based on the Co-Fluorescence System Eu-Y-Sparfloxacin-SDS', *Microchim. Acta*, **144**, 29–34 (2004).

778. B. Sendra, S. Panadero, A. Gomez-Hens, 'Determination of Pindone in Baits by Using Time-Resolved Lanthanide-Sensitized Luminescence and Kinetic Methodology', *Anal. Lett.*, **32**, 1835–1846 (1999).

779. G.L. Zhong, Y. Feng, K.Z. Yang, G.Y. Zhu, 'Enhanced Luminescence of Eu(III) by La(III), Gd(III), Tb(III) and Y(III) in Langmuir-Blodgett Films of Mixed Rare Earth Complexes', *Chem. Lett.*, 775–776 (1996).

780. B.C. Roy, M.A. Fazal, A. Arruda, S. Mallik, A.D. Campiglia, 'Polymerized Fluorescent Liposomes Incorporating Lanthanide Ions', *Org. Lett.*, **2**, 3067–3070 (2000).

781. M. Galtayries, M.P. Aguilar-Caballos, A. Gomez-Hens, 'Determination of Methyl Anthranilate in Food Samples by Coupling Stopped-Flow Mixing Technique and Time-Resolved Luminescence Detection', *Luminescence*, **15**, 363–369 (2000).

782. S.N. Shtykov, T.D. Smirnova, Y.G. Bylinkin, 'Determination of Adenosine Triphosphoric Acid and its Effect on the Quenching of the Fluorescence of Europium (III) Diketonate in Micelles of Brij-35', *J. Anal. Chem.*, **59**, 438–441 (2004).

783. J.A. Ocana, M. Callejon, F.J. Barragan, 'Terbium-Sensitized Luminescence Determination of Levofloxacin in Tablets and Human Urine and Serum', *Analyst*, **125**, 1851–1854 (2000).

784. J.A. Hernandez-Arteseros, R. Compano, R. Ferrer, M.D. Prat, 'Application of Principal Component Regression to Luminescence Data for the Screening of Ciprofloxacin and Enrofloxacin in Animal Tissues', *Analyst*, **125**, 1155–1158 (2000).

785. N. Arnaud, J. Georges, 'Sensitive Detection of Tetracyclines Using Europium-Sensitized Fluorescence with EDTA as Co-Ligand and Cetyltrimethylammonium Chloride as Surfactant', *Analyst*, **126**, 694–697 (2001).

786. S.M. Meyerhoffer, L.B. McGown, 'Fluorescent Probe Studies of Metal Salt Effects on Bile Salt Aggregation', *J. Am. Chem. Soc.*, **113**, 2146–2149 (1991).

787. S.M.Z. Al-Kindy, F.E.O. Suliman, 'Determination of Ibuprofen in Pharmaceutical Formulations Using Time-Resolved Terbium-Sensitized Luminescence', *Luminescence*, **22**, 294–301 (2007).

788. S. Torres-Cartas, R.M. Villanueva-Camanas, M.C. Garcia-Alvarez-Coque, 'Sensitized Lanthanide Fluorescence Detection of Steroidal Hormones', *J. Liq. Chromatogr. Related Technol.*, **23**, 1171–1186 (2000).

789. J.X. Duggan, 'Phosphorimetric Detection in HPLC Via Trivalent Lanthanides – High-Sensitivity Time-Resolved Luminescence Detection of Tetracyclines Using Eu^{3+} in a Micellar Post Column Reagent', *J. Liq. Chromatogr.*, **14**, 2499–2525 (1991).

790. M. Amin, K. Harrington, R. von Wandruszka, 'Determination of Steroids in Urine by Micellar HPLC with Detection by Sensitized Terbium Fluorescence', *Anal. Chem.*, **65**, 2346–2351 (1993).

791. *LKB News*, 'Autumn', LKB Wallac Oy, Turku, 2–7, 1984.

792. T. Stahlberg, E. Markela, H. Mikola, P. Mottram, I. Hemmila, 'Europium and Samarium in Time-Resolved Fluoroimmunoassys', *Am. Lab.*, **25**, 15–20 (1993).

793. E.P. Diamandis, R.A. Evangelista, A. Pollak, E.F. Templeton, J.A. Lowden, 'Time-Resolved Fluoroimmunoassays with Europium Chelates as Labels', *Am. Clin. Lab.*, **8**, 26–28 (1989).

794. H. Alfthan, U.H. Stenman, 'Time-Resolved Fluoroimmunoassays', *Am. Biotechnol. Lab.*, **6**, 8–13 (1988).

795. E.P. Diamandis, T.K. Christopoulos, 'Europium Chelate Labels in Time-Resolved Fluorescence Immunoassays and DNA Hybridization Assays', *Anal. Chem.*, **62**, A1149–1157A (1990).

796. T. Lovgren, 'Europium Flourescence and its Application in Immunochemistry', *Kem.-Kemi*, **12**, 418–421 (1985).

797. P. Mottram, P. Soderback, 'Application of Lanthanide Chelates for Time-Resolved Fluoroimmunoassay', *Am. Clin. Lab.*, **9**, 34–38 (1990).

798. M.U. Suonpaa, J.T. Lavi, I.A. Hemmila, T.N.E. Lovgren, 'A New Sensitive Assay of Human α-Fetoprotein Using Time-Resolved Fluorescence and Monoclonal Antibodies', *Clin. Chim. Acta*, **145**, 341–348 (1985).

799. H. Dechaud, R. Bador, F. Claustrat, C. Desuzinges, R. Mallein, 'New Approach to Competitive Lanthanide Immunoassay: Time-Resolved Fluoroimmunoassay of Progesterone with Labeled Analyte', *Clin. Chem.*, **34**, 501–504 (1988).

800. T. Lovgren, I. Hemmila, K. Pettersson, J.U. Eskola, E. Bertoft, 'Determination of Hormones by Time-Resolved Fluoroimmunoassay', *Talanta*, **31**, 909–916 (1984).

801. *Brochure on Delfia Labels for Research Applications, No.1244-1115-02*, Pharmacia Wallac, Turku.

802. M.P. Bailey, B.F. Rocks, C. Riley, 'Chelated Terbium as a Label in Fluorescence Immunoassay', in *Nonisotopic Immunoassay*, ed. T.T. Ngo, Plenum Press, New York, 187–197, 1988.

803. *A Novel Cell-Mediated Cytotoxicity Assay Based on Time-Resolved Fluorescence*, Wallac Oy, Turku, LKB Wallac Application Note 516, 1985.

804. S.C. Kinsky, *Methods in Enzymology*, Academic Press, New York, 1974.

805. R.B. Thompson, B.P. Gaber, 'Improved Fluorescence Assay of Liposome Lysis', *Anal. Lett.*, **18**, 1847–1863 (1985).

806. A. Orellana, M.-L. Laukkanenm, K. Keinanen, 'Europium Chelate-Loaded Liposomes: A Tool for the Study of Binding and Integrity of Liposomes', *Biochim. Biophys. Acta*, **1284**, 29–34 (1996).

807. Y. Ishimori, M. Notsuke, M. Koyama, 'Reagents for Immunochemical Analysis', Jpn. Kokai Tokkyo Koho, JP 60159652 1985.

808. A. Ius, M.A. Bacigalupo, G. Meroni, 'A Homogeneous Time-Resolved Fluoroimmunoassay for Haptens Utilizing Liposomes', *Anal. Biochem.*, **238**, 208–211 (1996).

809. G. Patonay, A. Shapira, P. Diamond, I.M. Warner, 'A Systematic Study of Pyrene Inclusion Complexes with α-, β-, and γ-Cyclodextrins', *J. Phys. Chem.*, **90**, 1963–1966 (1986).

810. A.J. Sophianopoulos, I.M. Warner, 'Purification of β-Cyclodextrin', *Anal. Chem.*, **64**, 2652–2654 (1992).

811. I. Cydzik, J.R. Albert-Garcia, J.M. Calatayud, 'Photo-Induced Fluorescence of Fluometuron in a Continuous-Flow Multicommutation Assembly', *J. Fluoresc.*, **17**, 29–36 (2007).

812. J.H. Garcia, J.R. Betancort Rodriguez, A.J. Bermejo Martin-Lazaro, J.J. Santana Rodriguez, 'Sensitive Spectrofluorimetric Determination of 3,3',4,4'-Tetrachlorobiphenyl Using a Neutral Micellar Medium', *Anal. Chim. Acta*, **290**, 146–153 (1994).

813. H.D. Burrows, S.J. Formosinho, M. Fernanda, J.R. Paiva, E.J. Rasburn, 'Halide Ion Induced Quenching and Enhancement of the Fluorescence of Fluoranthene Solubilized in Cetyltrimethylammonium Bromide (CTAB) Micelles', *J. Chem. Soc., Faraday Trans. 2*, **76**, 685–692 (1980).

814. A.G. Howard, A.J. Coxhead, I.A. Potter, A.P. Watt, 'Determination of Dissolved Aluminum by the Micelle-Enhanced Fluorescence of its Lumogallion Complex', *Analyst*, **111**, 1379–1382 (1986).

815. D.Y. Pharr, Z.S. Fu, T.K. Smith, W.L. Hinze, 'Solubilization of Cyclodextrins for Analytical Applications', *Anal. Chem.*, **61**, 275–279 (1989).

816. A.K. Chatjigakis, C. Donze, A.W. Coleman, P. Cardot, 'Solubility Behavior of β-Cyclodextrin in Water/Cosolvent Mixtures', *Anal. Chem.*, **64**, 1632–1634 (1992).

817. A.I. Olives, E.E. Romero-Ale, M.A. Martin, B. del Castillo, P. Lopez-Alvarado, J.C. Menendez, 'Changes in the Reactivity of the Fluorescent Reagents Carbazole-9-Carbonyl Chloride and 9-Carbazolylacetic Acid in the Presence of Cyclodextrins', *J. Inclusion Phenom. Macrocyclic*, **57**, 553–559 (2007).

818. R.L. Van Etten, J.F. Sebastian, G.A. Clowes, M.L. Bender, 'Acceleration of Phenyl Ester Cleavage by Cycloamyloses. A Model for Enzymic Specificity', *J. Am. Chem. Soc.*, **89**, 3242–3253 (1967).

819. G. Patonay, M.E. Rollie, I.M. Warner, 'Sample Preparation Considerations for Measurement of Fluorescence Enhancement Using Cyclodextrins and Micelles', *Anal. Chem.*, **57**, 569–571 (1985).

820. M. Shinitzky, G. Weber, A.C. Dianoux, C. Gitler, 'Microviscosity and Order in the Hydrocarbon Region of Micelles and Membranes Determined with Fluorescent Probes. I. Synthetic Micelles', *Biochemistry*, **10**, 2106–2113 (1971).

821. B.W. Williams, C.D. Stubbs, 'Properties Influencing Fluorophore Lifetime Distributions in Lipid Bilayers', *Biochemistry*, **27**, 7994–7999 (1988).

822. J. Sunamoto, K. Iwamoto, H. Iwamoto, H. Kondo, K. Kano, T. Ogawa, 'Liposome Fusion as Monitored by a Fluorescence Quenching Technique', *Chem. Lett.*, 15–18 (1981).

823. B.C.R. Guillaume, D. Yogev, J.H. Fendler, 'Long-Range Intermolecular Vibrational Coupling of Water Molecules in Contact with Surfactant Vesicles', *J. Chem. Soc., Faraday Trans.*, **88**, 1281–1284 (1992).

824. A.M. Carmona-Ribeir, B.R. Midmore, 'Surface Potential in Charged Synthetic Amphiphile Vesicles', *J. Phys. Chem.*, **96**, 3542–3547 (1992).

825. D.A. Kendall, R.C. MacDonald, 'A Fluorescence Assay to Monitor Vesicle Fusion and Lysis', *J. Biol. Chem.*, **257**, 13892–13895 (1982).

826. J. Havel, I. Buresovajancarova, V. Kuban, 'Quantitative Description of the Changes in the Acid-Base Properties of Bromocresol Green in the Presence of Submicelle and Micelle Concentrations of Cationic Tenside Septonex', *Collect. Czech. Chem. Commun. C.*, **48**, 1290–1304 (1983).

827. M. Gratzel, J.K. Thomas, 'On the Dynamics of Pyrene Fluorescence Quenching in Aqueous Ionic Micellar Systems. Factors Affecting the Permeability of Micelles', *J. Am. Chem. Soc.*, **95**, 6885–6889 (1973).

828. Z. Djuric, M. Jovanovic, 'Effect of pH and Storage on Critical Micellar Concentration Values of Sodium Lauryl Sulfate', *Arh. Farm.*, **35**, 197–205 (1985).

829. W.A. MacCrehan, W.E. May, 'Oxygen Removal in Liquid Chromatography with a Zinc Oxygen-Scrubber Column', *Anal. Chem.*, **56**, 625–628 (1984).

830. E. Morales, L. Mercado, D. Rivera, G.A. Infante, 'Solublization of Petroleum Oil by Micelle Forming Surfactants', *Science-Ciencia*, **6**, 23–25 (1978).

831. W. Prapaitrakul, A. Shwikhat, A.D. King, 'The Influence of pH on Gas Solubilities in Aqueous Solutions of Sodium Octanoate at 25°C', *J. Colloid Interface. Sci.*, **115**, 443–449 (1987).

832. M.S. Ryabova, 'Solubility of Atmospheric Oxygen in Water in the Presence of Anionic Surfactants', Deposited Document, VINITI, 4542–4584, 1984; *Chem. Abstr.*, **102**, 226775t.

833. M.E. Rollie, G. Patonay, I.M. Warner, 'Automated Sample Deoxygenation for Improved Luminescence Measurements', *Gov, Rep. Announce. Index (U.S.)*, **87**(6), 435(1987), Abstr. No. 709 395; *Chem. Abstr.*, **106**, 187928f.

834. S.L. Buell, J.N. Demas, 'Simple Sample Cell for Deoxygenation of Surfactant Solutions', *Rev. Sci. Instrum.*, **53**(8), 1298–1299 (1982).

835. J.C. Selwyn, J.C. Scaiano, 'Static and Dynamic Phenomena in the Quenching of Phenanthrene Triplets by Conjugated Dienes in Anionic Micelles', *Can. J. Chem.*, **59**, 663–668 (1981).

836. M.E. Diaz Garcia, A. Sanz-Medel, 'Facile Chemical Deoxygenation of Micellar Solutions for Room Temperature Phosphorescence', *Anal. Chem.*, **58**, 1436–1440 (1986).

837. Y. Wei, C. Liu, 'Chemical Deoxygenation for Cyclodextrin-Induced Room Temperature Phosphorescence', *Fenxi Huaxue*, **18**, 1001–1005 (1990).

838. N.E. Nugara, A.D. King, 'Light Absorption and Mixed Micelle Composition as Factors in Determining Intensities of Room-Temperature Phosphorescence', *Anal. Chem.*, **61**, 1431–1435 (1989).

839. J.R. Darwent, C.D. Flint, N.W. Sharpe, 'Intermolecular Energy Transfer from Phenanthrene to Europium in Aqueous Micellar Solution', *J. Chem. Soc., Chem. Commun.*, 747–748 (1988).

840. M.W. Geiger, N.J. Turro, 'Pyrene Fluorescence Lifetime as a Probe for Oxygen Penetration of Micelles', *Photochem. Photobiol.*, **22**, 273–276 (1975).

841. M.E. Rollie, G. Patonay, I.M. Warner, 'Deoxygenation of Solutions and its Analytical Applications', *Ind. Eng. Chem. Res.*, **26**, 1–6 (1987).

842. I.B. Butler, M.A.A. Schoonen, D.T. Rickard, 'Removal of Dissolved Oxygen from Water: A Comparison of Four Common Techniques', *Talanta*, **41**, 211–215 (1994).

843. K. Kano, B. Zhou, M. Sakaguchi, 'Effects of Aeration and Freezing and Thawing on Self-Aggregation of Naphthalene in Aqueous β-Cyclodextrin Solutions', *Sci. Eng. Rev. Doshisha Univ.*, **25**, 253–259 (1985).

844. F.M. Winnik, M.A. Winnik, H. Ringsdorf, J. Venzmer, 'Bis(1-Pyrenylmethyl) Ether as an Excimer-Forming Probe of Hydrophobically Modified Poly(N-Isopropyl-acrylamides) in Water', *J. Phys. Chem.*, **95**, 2583–2587 (1991).

845. R.E. Bilbo, 'Letter in Lab Notes Section', *Aldrichimica Acta*, **20**, 62 (1987).

846. A. Berthod, C. Saliba, 'pH Measurements in Micellar Solutions', *Analusis*, **13**, 437–442 (1985).

847. N. Arnaud, S. Drevet, J. Georges, 'Effect of Surfactants in the Determination of Orotic-Acid in Aqueous-Solutions by Sensitized Terbium Fluorescence', *Analyst*, **119**, 2453–2456 (1994).

848. H. Ferreira, M. Lucio, B. de Castro, P. Gameiro, J. Lima, S. Reis, 'Partition and Location of Nimesulide in EPC Liposomes: A Spectrophotometric and Fluorescence Study', *Anal. Bioanal. Chem.*, **377**, 293–298 (2003).

849. J.L. Vilchez, J. Taoufiki, O. Ballesteros, A. Navalon, 'Micelle Enhanced Spectrofluorimetric Method for the Determination of Antibacterial Trovafloxacin in Human Urine and Serum', *Microchim. Acta*, **150**, 247–252 (2005).

850. G.J. Yang, P. Liu, X.L. Qu, S. Ming, C.Y. Wang, Q.S. Qu, X.Y. Hu, Z.Z. Leng, 'Micellar-Enhanced Spectrofluorimetric Determination of Trazodone Hydrochloride in Human Urine and Serum', *Anal. Lett.*, **40**, 151–162 (2007).

851. J.L. Vilchez, A. Navalon, L. Araujo, A. Prieto, 'Determination of Danofloxacin and Marbofloxacin in Milk

Samples by Micellar Liquid Chromatography with Fluorescence Detection', *Anal. Lett.*, **40**, 601–613 (2007).

852. R. Adamou, A. Coly, S.E. Douabale, M. Saleck, M.D. Gaye-Seye, A. Tine, 'Fluorimetric Determination of Histamine in Fish Using Micellar Media and Fluorescamine as Labelling Reagent', *J. Fluoresc.*, **15**, 679–688 (2005).

853. F. Wang, X. Wu, F. Wang, S.F. Liu, Z. Jia, J.H. Yang, 'The Sensitive Fluorimetric Method for the Determination of Curcumin Using the Enhancement of Mixed Micelle', *J. Fluoresc.*, **16**, 53–59 (2006).

854. A.T. Pilipenko, G.N. Pishinko, A.I. Volkova, 'Reaction of Aluminum with Acid Chrome Blue Black in the Presence of Surfactants', *Ukr. Khim. Zh.*, **48**, 957–962 (1982).

855. J.I.G. Alonso, A.L. Garcia, A. Sanzmedel, E.B. Gonzales, L. Ebdon, P. Jones, 'Flow-Injection and Liquid Chromatographic Determination of Aluminum Based on its Fluorimetric Reaction with 8-Hydroxyquinoline-5-Sulphonic Acid in a Micellar Medium', *Anal. Chim. Acta*, **225**, 339–350 (1989).

856. N. Ishibashi, K. Kina, 'Sensitivity Enhancement of the Fluorometric Determination of Aluminum by the Use of Surfactant', *Anal. Lett.*, **5**, 637–641 (1972).

857. H. Wang, L.Z. Liu, 'Fluorimetric Determination of Silver with Eosin B – 1,10-Phenanthroline and Nonionic Surfactant', *Fenxi Huaxue*, **14**, 773–775 (1986).

858. F.G. Sanchez, M.H. Lopez, A. Heredia, 'Spectrofluorimetric Determination of Beryllium Based on the Inclusion Complex of 1-Amino-4-Hydroxyanthraquinone with β-Cyclodextrin as Ligand', *Anal. Chim. Acta*, **187**, 147–155 (1986).

859. Q. Lin, Z. Si, J. Xu, 'Fluorimetric Determination of Cadmium with Ferron[Sodium Salt] and Hexadecyltrimethylammonium Bromide', *Huaxue Shiji*, **10**, 106–108 (1988).

860. I. Mori, Y. Fujita, K. Ikuta, Y. Nakahashi, K. Kato, N. Niwa, 'Fluorometric Determination of Iron(III) with o-Hydroxyhydroquinonephthalein in the Presence of Brij 58', *Anal. Lett.*, **22**, 1969–1979 (1989).

861. N. Ishibashi, K. Kina, Y. Goto, 'Selective Determination of Gallium by Flow Injection Fluorimetry', *Anal. Chim. Acta*, **114**, 325–328 (1980).

862. K. Kina, N. Ishibashi, 'Effect of the Nonionic Surfactant on the Fluorometric Determination of Gallium Using Lumogallion', *Microchem. J.*, **19**, 26–31 (1974).

863. M.W. Kabbani, S. Rubio-Borroso, L.M. Polo-Diez, 'Determination of Traces of Gallium in the Presence of Aluminium by Synchronous Derivative Spectrofluorimetry', *Anal. Chim. Acta*, **218**, 167–171 (1989).

864. F.G. Sanchez, M.H. Lopez, J.C.M. Gomez, 'Fluorometric-Determination of Gallium by Means of a Cyclodextrin Inclusion Compound as Ligand', *Fresenius Z. Anal. Chem.*, **328**, 499–500 (1987).

865. S.B. Meshkova, N.V. Rusakova, N.S. Poluektov, 'Luminescence Reaction of Lanthanum with 8-Hydroxyquinoline-5-Sulfonic Acid and Cetyltrimethylammonium Salt', *Z. Anal. Khim.*, **37**, 1988 (1982).

866. J. Medina, F. Hernandez, R. Marin, F.J. Lopez, 'Study of the Fluorescence of the Lead Morin System in the Presence of Nonionic Surfactants', *Analyst*, **111**, 235–237 (1986).

867. S. Li, D. Zeng, M. Ma, P. Li, 'Fluorimetric Determination of Ruthenium by Use of the Ruthenium-Quinolin-8-o-l-Hexadecyltrimethylammonium Bromide System', *Yankuang Ceshi*, **7**, 130–132 (1988).

868. S. Li, D. Zeng, P. Li, M. Ma, 'Fluorimetric Reaction of Ruthenium with 1,10-Phenanthroline and Sodium Dodecylesulfate', *Yankuang Ceshi*, **7**, 221–223 (1988).

869. F.G. Sanchez, M.H. Lopez, J.C.M. Gomez, 'Fluorometric-Determination of Scandium Using the Cyclodextrin 1,4-Dihydroxyanthraquinone Inclusion Complex', *Analyst*, **112**, 1037–1040 (1987).

870. J.L. Manzoori, M. Amjadi, D. Abolhasani, 'Spectrofluorimetric Determination of Tin in Canned Foods', *J. Hazard. Mater.*, **137**, 1631–1635 (2006).

871. Y.-X. Fan, Y.-X. Zheng, 'Effect of Cationic Micelles on the Fluorescence of the Zirconium–Morin Complex', *Anal. Chim. Acta*, **281**, 353–360 (1993).

872. F.G. Sanchez, J.C.M. Gomez, M.H. Lopez, 'Spectrofluorimetric Determination of Zirconium Based on the Inclusion Complex of 8-Hydroxy-7-Iodo Quinoline Sulfonic-Acid with Beta-Cyclodextrin as Ligand', *Anal. Lett.*, **23**, 923–933 (1990).

873. I. Mori, T. Matsuo, M. Toyoda, Y. Fujita, C. Matoba, K. Akagiri, 'Application of o-Hydroxyhydroquinonephthalein for the Fluorophotometric Determination of Zirconium(IV) and Gallium(III) in the Presence of Tetradecyltrimethylammonium Chloride', *Anal. Lett.*, **28**, 649–662 (1995).

874. J.A. Munoz, A.M.G. Campana, F.A. Barrero, 'Effect of Cationic Micelles on the Formation of the Complex Oxalate-Alizarin Red S-Zr(IV): Application to the Sensitive Fluorescence Determination of Oxalate Ion', *Talanta*, **47**, 387–399 (1998).

875. P. Izquierdo, A. Gomez-Hens, D. Perez-Bendito, 'Study of the Eu(III)-Tetracycline–Thenoyltrifluoroacetone System by Using the Stopped-Flow Mixing Technique: Determination of Tetracycline in Serum', *Anal. Chim. Acta*, **292**, 133–139 (1994).

876. S. Panadero, A. Gomez-Hens, D. Prez-Bendito, 'Stopped-Flow Determination of Diphacinone Based on Lanthanide-Sensitized Luminescence', *Anal. Chim. Acta*, **280**, 163–168 (1993).

877. K. Kano, 'Fluorescent Study of Molecular Aggregates', *Hyomen*, **26**, 243–253 (1988).

878. M.A.C. de la Torre, A. Gomez-Hens, 'Evaluation of the Terbium(III)-Sensitized Luminescence with Benzenepolycarboxylic Acids – Determination of Terephthalic Acid in Drink Samples', *Anal. Chim. Acta*, **407**, 53–60 (2000).

879. D.A. Lerner, M.A. Martin, 'Luminescence in Organized Media and Supramolecular Interactions: Physicochemical Aspects and Luminescence Spectroscopy – Applications and Recent Trends', *Analusis*, **28**, 649–663 (2000).

880. J.J. Wang, A. Ueno, 'Naphthol-Modified Cyclodextrins as Fluorescent Sensors for Detecting Contaminants in Drinking Water', *Macromol. Rapid Commun.*, **21**, 887–890 (2000).

881. H. Nakashima, N. Yoshida, 'Fluorescent Molecular Sensing for Various Types of Surfactants using Amino-β-Cyclodextrins Bearing Naphthalene Chromophore', *Chem. Lett.*, **35**, 1168–1169 (2006).

882. S. Pagliari, R. Corradini, G. Galaverna, S. Sforza, A. Dossena, M. Montallti, L. Prodi, N. Zaccheroni, R. Marchelli, 'Enantioselective Fluorescence Sensing of Amino Acids by Modified Cyclodextrins: Role of the Cavity and Sensing Mechanism', *Chem. Eur. J.*, **10**, 2749–2758 (2004).

Fluorescence Lifetime Measurements, Applications of

Frank V. Bright, Siddharth Pandey, and Gary A. Baker
State University of New York at Buffalo, Buffalo, New York, USA

1 Introduction 1
1.1 Time Correlated Single Photon Counting 3
1.2 Multifrequency Phase and Modulation Fluorometry 3
1.3 Pump–Probe Spectroscopy 5
1.4 Up-conversion 6
2 Instrumentation 6
2.1 Time Domain 6
2.2 Frequency Domain 7
3 Solvation Dynamics 8
4 Applications in Biochemistry and Biophysics 9
5 Polymer Photophysics 12
6 Applications in the Separation Sciences 13
7 Characterization of Organized Media 15
8 Fluorescence Lifetime Imaging Microscopy 17
9 Fluorescence Lifetime-based Sensing 18
10 Multiphoton Excited Fluorescence 19
11 Summary and Future Prospects 21
Abbreviations and Acronyms 22
Related Articles 22
References 23

Over the years, the emissive characteristics (spectral, temporal, and polarization) of fluorophores have been widely used to probe the local environments of fluorescent centers. Fluorescence lifetime and rotational reorientation time measurements, in particular, offer a means to elucidate subtle features when examining, for example, complex systems involving environmental heterogeneity, energy transfer, quenching pathways, or excited state reactions. Further, because fluorescence occurs on the nanosecond (10^{-9} s) timescale, competing or perturbing kinetic processes like collisional quenching, solvent relaxation, energy transfer, and rotational reorientation can affect the fluorescence. Thus, a carefully chosen fluorophore can serve as an internal nanosecond or faster stopwatch of environmental fluctuations.

This article is divided into a brief introduction, a short discussion on instrumentation, and several sections on various applications of excited state fluorescence measurements.

1 INTRODUCTION

During the last few decades, fluorescence spectroscopy has developed into an important and widely used technique in the biological, chemical, and physical sciences. Multidimensional in nature, fluorescence is characterized by several parameters, most notably the excitation and emission wavelengths, the steady state polarization or anisotropy, the excited state lifetime, and the rotational reorientation time.[1] Of these, the latter two reflect the time course of the emission process; therefore, it is often of interest to determine such time-dependent emission from samples. This arises because the fluorescence anisotropy and/or intensity decay kinetics are usually characteristic of the system under investigation.[1–16] In condensed phase systems, these two processes typically occur in 10^{-8}–10^{-13} s. As a result, any process which occurs on a similar timescale can potentially perturb the fluorescence. For example, phenomena like collisional quenching, energy transfer, solvent relaxation, and rotational reorientation all occur on the fluorescence timescale. Hence an excited state fluorophore can be used to report on its local environment as well as temporal changes in the environment. Consequently, this information provides insight into important processes occurring on the nanosecond and subnanosecond timescale. For example, the presence of two fluorescent tryptophan residues within a protein, each of which resides in a distinct physicochemical environment, may manifest itself in the recovery of two apparent decay rates, as has been observed by Ross et al.[2] for liver alcohol dehydrogenase. Alternatively, the dynamical behavior of proteins can result in intrinsically multiexponential fluorescence anisotropy decays, where the more rapidly relaxing components reflect the amino acid residue segmental motions.[3] Further, many aspects of the changes at the molecular level such as energy transfer,[4] conformational distributions,[5] and transient effects in collisional quenching[6,7] display complex emission kinetics that are accessible from dynamical measurements.

In principle, resolution of the time-dependent decays of anisotropy and/or intensity can be equivalently accomplished using either time- or frequency-domain measurements.[8–16] In the time domain (impulse response approach), the temporal emission from the sample

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

following pulsed excitation (ideally a Dirac δ-function) is measured using detectors with high temporal resolution (e.g. streak cameras) or time-sampling circuitry (e.g. time-correlated single photon counting approach). In the frequency domain (harmonic response approach), the emission phase angle lag and modulation attenuation relative to the sinusoidally modulated incident light are measured.

The first analytical paper on time-resolved spectroscopy, published in 1957 by Keirs, Britt, and Wentworth, was based on measuring phosphorescence lifetimes (see O'Connor and Phillips[10]). Some 42 years later, the improvements in optical technology and continued advances in digital electronics and computing power have lead us to the era of ultrafast spectroscopy.[16] Before we discuss these methods that are actually used to measure the anisotropy and/or intensity decay kinetics, it is instructive to look at the general expressions describing fluorescence intensity and anisotropy decay kinetics.

For a system containing n noninteracting fluorescence species (components), following excitation with an optically short Dirac δ-function, the time-dependent emission decay can be described by a multiexponential series of the form, Equation (1):[8–15]

$$I(t) = \sum_{i=1}^{n} \alpha_i \exp\left(-\frac{t}{\tau_i}\right. \tag{1}$$

This expression is strictly true for discrete lifetime components. Here, α_i is a pre-exponential factor that represents the fractional contribution of the ith component, possessing excited state lifetime τ_i, to the total time-resolved intensity decay. The fractional contribution to the total intensity decay (f_i) of each species is then given by Equation (2):

$$f_i = \frac{\alpha_i \tau_i}{\sum_{i=1}^{n} \alpha_i \tau_i} \tag{2}$$

Of course, nonexponential or complex exponential decays are often observed in what are assumed to be simple systems because of competing processes (e.g. solvent relaxation, energy migration) or inherent heterogeneity. In cases where single or multiple fluorophores are located simultaneously in or interconverting between an ensemble of distinct environments, a distribution of excited state lifetimes reflecting the environmental distributions would be expected a priori. From Equations (1) and (2) it can be shown that:

$$I(t) = \sum_{i=1}^{n} f_i \tau_i^{-1} \exp\left(-\frac{t}{\tau_i}\right. \tag{3}$$

For the limiting case where the number of emitting centers is large, Equation (3) can be expressed as an integral of the form, Equation (4):

$$I(t) = \int_0^\infty f(\tau)\tau^{-1} \exp\left(-\frac{t}{\tau}\right. \mathrm{d}\tau \tag{4}$$

Most approaches to the study of excited state lifetime distributions assume that $f(\tau)$ is given by one of several continuous distribution functions, Equations (5–7):

$$\text{uniform:} \quad f(\tau) = A \quad \text{from } (\tau - W/2) \text{ to } (\tau + W/2) \tag{5}$$
$$f(\tau) = 0 \quad \text{at all other values}$$

or

$$\text{Gaussian:} \quad f(\tau) = A\left\{\frac{\exp[-(\tau - \tau')^2/2\sigma^2]}{\sigma\sqrt{2\pi}}\right\} \tag{6}$$

or

$$\text{Lorentzian:} \quad f(\tau) = \frac{A}{\{1 + [(\tau - \tau')/(W/2)]^2\}} \tag{7}$$

where τ' is the central or mean lifetime value, σ is the standard deviation of the Gaussian distribution, W is the full width at half-maximum (fwhm) for the Lorentzian and uniform distributions, τ is the lifetime, and A is an amplitude constant determined from the normalization condition, Equation (8):

$$f(\tau)\,\mathrm{d}\tau = 1 \tag{8}$$

In a time-resolved anisotropy decay experiment, the sample is excited with a brief pulse of vertically polarized light and the time dependence of the parallel $[I_\parallel(t)]$ and perpendicular $[I_\perp(t)]$ components of the fluorescence provide the time-resolved decay of anisotropy $r(t)$, Equation (9):

$$r(t) = \frac{I_\parallel(t) - I_\perp(t)}{I_\parallel(t) + 2I(t)_\perp} \tag{9}$$

For a simple isotropic rotor, $r(t)$ decays with a single rotational reorientation time, ϕ, Equation (10):

$$r(t) = r_0 \exp\left(-\frac{t}{\phi}\right. \tag{10}$$

where r_0 is the so-called limiting anisotropy measured in the absence of any rotational reorientation. For more complicated systems, $r(t)$ may take the form of a string of exponentials, Equation (11):

$$r(t) = r_0 \sum_{i=1}^{n} \beta_i \exp\left(-\frac{t}{\phi_i}\right. \tag{11}$$

where β_i and ϕ_i are the fractional contribution of the total decay of anisotropy and the rotational correlation time

attributed to the *i*th apparent rotational reorientation event, respectively.

There are several methods used to obtain excited state intensity decay $I(t)$ and/or anisotropy decay $r(t)$. Some of the methods will be discussed here briefly. For a more elaborate description, the reader is directed to the specific literature references provided.

1.1 Time Correlated Single Photon Counting

Time correlated single photon counting (TCSPC)[10–12] relies on the concept that the probability distribution for emission of a single photon after an excitation event yields the actual intensity versus time distribution of all the photons emitted as a result of the excitation. By statistically sampling the single photon emission following a large number of excitation events, the experiment constructs this probability distribution.[10]

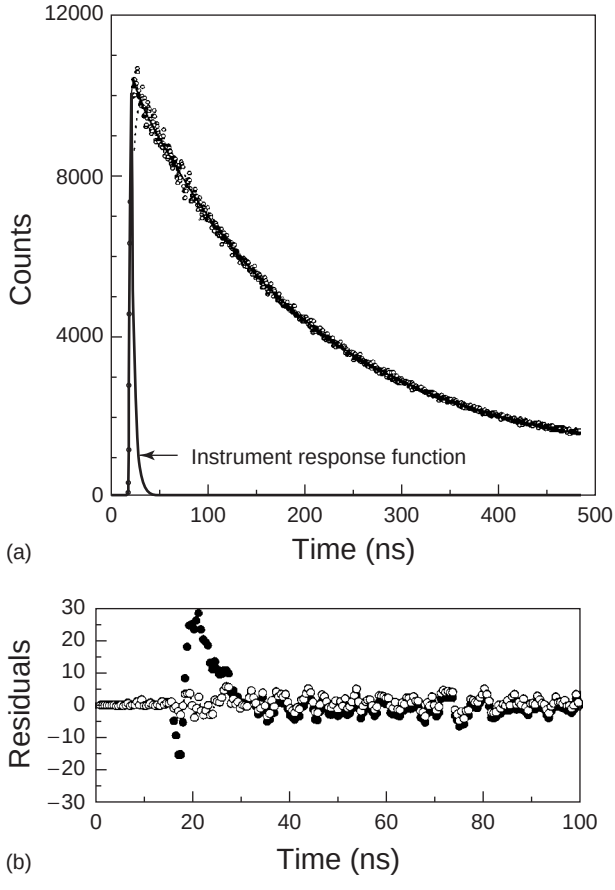

Figure 1 Typical TCSPC excited state intensity decay data for dilute 1-ethylpyrene dissolved in liquid toluene. The best fits for mono- (- - - -) and bi-exponential (———) decay functions are shown through the experimental data points (a). $\tau_1 = 170\,\text{ns}$ ($f_1 = 0.92$), $\tau_2 = 1.68\,\text{ns}$ ($f_2 = 0.08$), $\chi^2_{\text{single}} = 9.68$, $\chi^2_{\text{double}} = 1.73$. The recovered residuals for single (○) and double (●) exponential fits are provided in (b).

The TCSPC technique was first used by Bollinger and Thomas in 1961 (see O'Connor and Phillips[10]). To measure the luminescence decay characteristics of scintillators excited by α-, β-, and γ-radiation. Koechlin (see O'Connor and Phillips[10]), in the same year, used TCSPC to determine the shapes of scintillator emission curves with resolution in the nanosecond range. By way of illustration, Figure 1(a) shows a TCSPC fluorescence decay curve for a dilute solution of 1-ethylpyrene dissolved in liquid toluene. From the fits and corresponding residuals for single and double exponential decay laws, associated χ^2 values and residuals (Figure 1b), it is clear that a double exponential decay law is required to describe the overall decay kinetics adequately.

In TCSPC, as the excited state intensity decay kinetics approach the timescale of the so-called instrument response function (Figure 1a), the discrimination between the two becomes more and more critical to recovering accurate decay constants. The typical way to recover the decay parameters from the decay trace and the instrument response function is to solve the convolution integral, Equation (12):

$$F(t) = \int_0^t P(t')I(t-t')\,\mathrm{d}t' \tag{12}$$

In this expression, $P(t')$ is the instrumental response, $I(t-t')$ is the impulse response from the sample which describes the kinetic model, and $F(t)$ represents the decay curve. If the measured excited state decay is represented by $Y(t)$, the goodness of fit (χ^2) between $F(t)$ and $Y(t)$ can be optimized by iterating the parameters in $I(t)$ via assessing using normalized chi-squared values (χ^2) where, Equation (13):

$$\chi^2 = \frac{1}{(N-n)} \sum_{i=1}^{N} \frac{(Y_i - F_i)^2}{Y_i} \tag{13}$$

for N data points and n fitting parameters. For a good fit, χ^2 should be "close to unity" (see Figure 1).[10]

1.2 Multifrequency Phase and Modulation Fluorometry

In the frequency domain, the sample under study is excited with sinusoidally, amplitude-modulated light.[13–15] Typically, these frequencies span the mega- to gigahertz regions and have been generated by using acousto-optical, electro-optical or piezo-optical devices, continuous wave (CW) or mode-locked lasers, and synchrotron sources.[13] The resulting time-dependent fluorescence is slave to the frequency of the excitation waveform, but is phase shifted (θ) and amplitude demodulated (M) to an extent dependent on the fluorescence lifetime(s) inherent to the sample. For any

decay process the calculated (denoted by subscript c) values of θ and M are given by[13] Equations (14) and (15)

$$\theta_c(w) = \tan^{-1} \frac{S(w)}{C(w)} \tag{14}$$

$$M_c(w) = \{S(w)\}^2 + \{C(w)\}^2 \,^{1/2} \tag{15}$$

where w is the angular frequency and, for a single exponential decay, Equations (16) and (17):

$$\frac{S(w)}{C(w)} = w\tau \tag{16}$$

and

$$[S(w)]^2 + [C(w)]^2 = \left(1 + w^2\tau^2\right)^{-1} \tag{17}$$

For a general decay model expressed by Equation (1), the frequency-dependent sine $S(w)$ and cosine $C(w)$ Fourier transforms are given by Equations (18) and (19):

$$S(w) = \int_0^\infty I(t) \sin wt \, dt \tag{18}$$

$$C(w) = \int_0^\infty I(t) \cos wt \, dt \tag{19}$$

The decay times are recovered by experimentally measuring $\theta_m(w)$, the frequency-dependent phase shift, and $M_m(w)$, the frequency-dependent modulation, values while probing the sample at several angular modulation frequencies ($w = 2\pi f$, where f is the linear modulation frequency in hertz), followed by analysis using nonlinear least squares fitting algorithms. In this approach, the sum of the squared differences between the calculated ($\theta_c(w)$ and $M_c(w)$) and measured ($\theta_m(w)$ and $M_m(w)$) values are minimized using the χ^2 as a measure of the goodness-of-fit, Equation (20):

$$\chi^2 = \frac{1}{D}\sum_w \frac{(\theta_m(w) - \theta_c(w))}{v_\theta^2}^2 + \frac{1}{D}\sum_w \frac{(M_m(w) - M_c(w))}{v_M^2}^2 \tag{20}$$

Here, v_θ and v_M are the variances in the measured phase angle and demodulation, respectively, and D is the number of degrees of freedom. When v_θ and v_M reflect accurately the uncertainties in the θ (phase angle) and M (demodulation) experimental measurements, the fit quality is judged by the closeness of χ^2 to unity, the randomness of the phase angle and demodulation residuals about zero, and the physical significance of the model. Figure 2 shows a typical phase-modulation data set for calcium green dissolved in pH 7.2 phosphate saline buffer (10 mM in Ca(II)). In Figure 2(a),

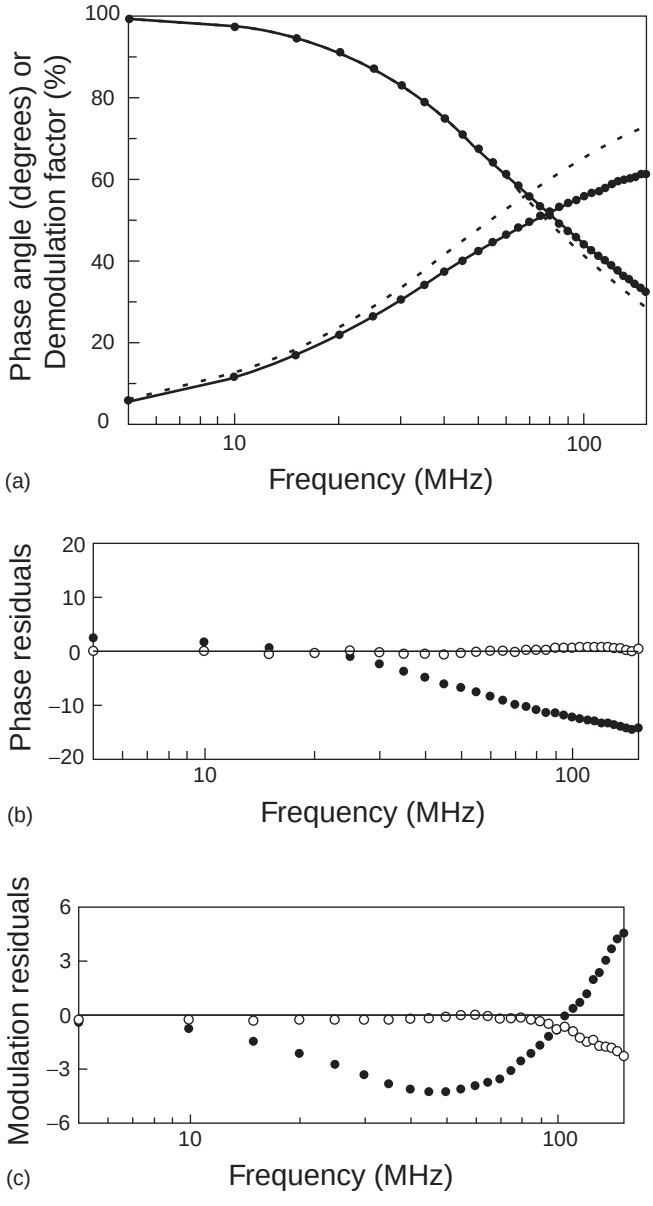

Figure 2 Typical MHF (multiharmonic Fourier transform) phase-modulation data (●) for calcium green dissolved in buffer with 10 mM Ca(II) at 40 °C (a). The best single (- - - -) and double (———) exponential fits to the data are shown as well as the associated χ^2 values: $\chi^2_{single} = 98$, $\chi^2_{double} = 0.86$, $\tau_1 = 3.70$ ns ($f_1 = 0.86$), $\tau_2 = 0.70$ ns ($f_2 = 0.14$). Frequency-dependent phase (b) and modulation (c) residuals are also provided for single (●) and double (○) exponential fits.

the points denote the experimental data and the fit to a single and double exponential decay law are shown. The fit of these data to a double exponential decay law is far superior to the fit to a single exponential decay law. Frequency-dependent phase and demodulation residuals (Figures 2b and 2c) demonstrate the difference between the fit to a single and double exponential decay law.

1.3 Pump–Probe Spectroscopy

For measurement of very fast relaxation processes which demand a time resolution below 10^{-10} s, most detectors are simply not fast enough (with the notable exception of streak cameras which, besides being very expensive, are restricted in their application by lower sensitivity, lower dynamic range, and the necessity of careful calibration and alignment). In this case, the pump–probe[16,17] technique offers the best alternative. In this technique, the species under investigation are excited by a fast laser pulse (i.e. pump) (Figure 3a). A second laser pulse, with a variable delay time Δt relative to the pump pulse, probes the time evolution of the excited state population density (Figure 3b). The signal intensity is proportional to the fluorescence intensity, provided that the probe pulse is not overly intense. The probe pulse arrival time is controlled by physically changing the probe beam optical path (Figure 3c). Incremental variation in the arrival time (i.e. probe pulse delay) allows the temporal decay function to be constructed.

In pump–probe spectroscopy, the time resolution is limited by the fundamental laser pulse widths and not by the detector intrinsic response time. The ability to generate ultrafast laser pulses thus allows pump–probe spectroscopy to be performed on a subpicosecond

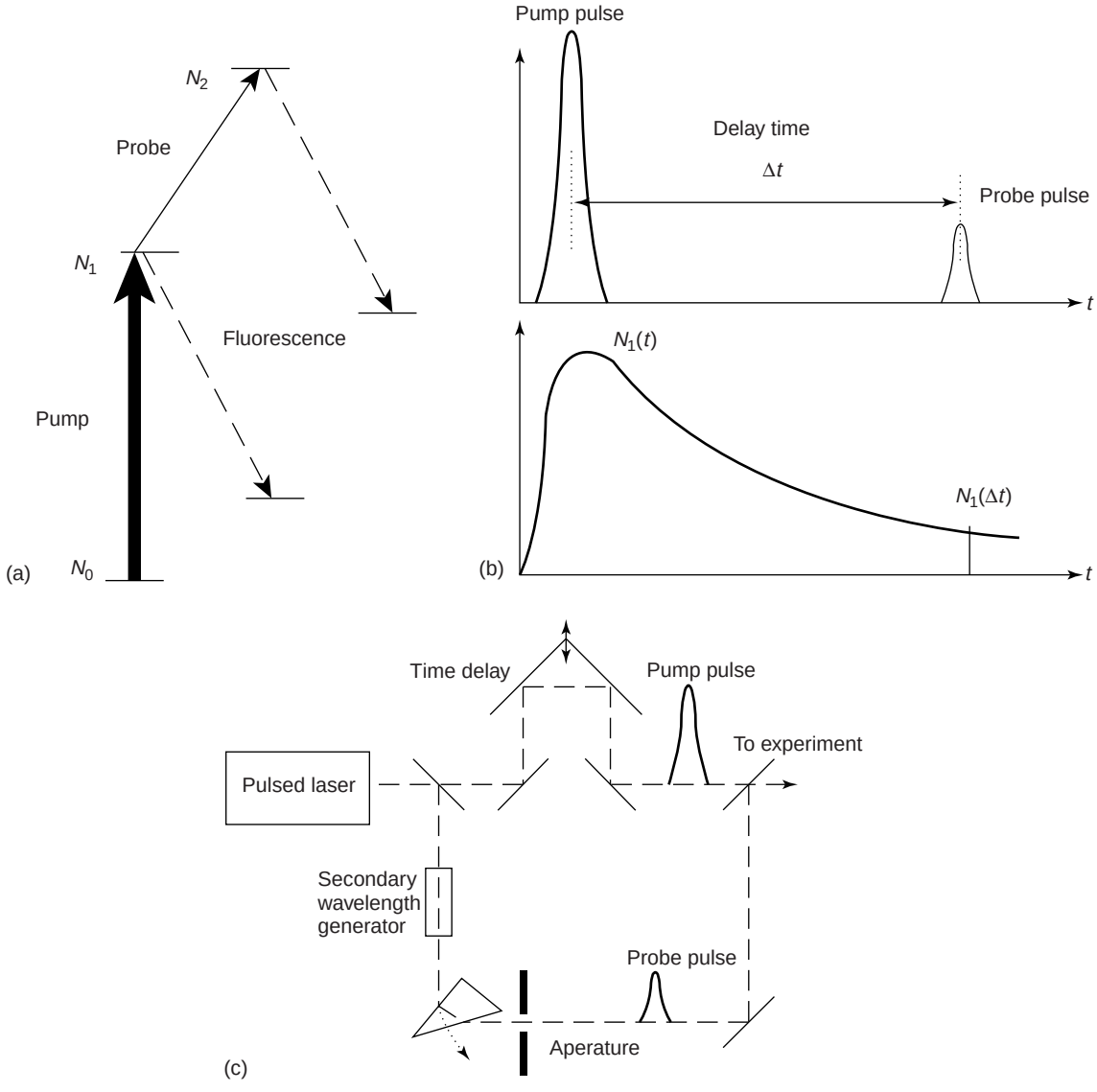

Figure 3 (a) In pump–probe spectroscopy, molecules under investigation are excited by a fast laser pulse on the $|0\rangle \rightarrow |1\rangle$ transition. (b) A probe pulse with a variable time delay Δt against the pump pulse probes the time evolution of the population density $N_1(t)$. (c) Schematic of a simplified instrumental set-up used in pump–probe spectroscopy for the measurement of an ultrafast relaxation processes. (Adapted from Demtroder.)[17]

timescale. Moreover, a broader range of applications is opened up by using two independently tunable ultrafast lasers that use the same pump laser to synchronize the pump and probe pulses.

1.4 Up-conversion

Another method that is well suited for the determination of excited state fluorescence anisotropy and intensity decay kinetics is termed up-conversion. This methodology subpicosecond fluorescence anisotropy and intensity decays to be determined.[16] Up-conversion refers to the nonlinear phenomenon of sum-frequency generation within anisotropic crystals. Figure 4 shows a general schematic of an up-conversion instrument. This method relies upon the generation of a signal of frequency w_3, which is the sum frequency for fluorescence w_2 and the exciting laser frequency w_1. In this case, w_3 can be distinguished spatially as well as by its frequency. When used to measure fluorescence decays, up-conversion offers several advantages over TCSPC. As well as vastly improved time resolution, it permits detection of fluorescence in the far red and infrared without the use of red-sensitive detectors. For example, fluorescence at 900 nm can be mixed with laser light at 600 nm to generate a sum frequency at 360 nm. The primary disadvantage, however, is that up-conversion of fluorescence toward the blue end of the spectrum is more difficult owing to the absence of efficient mixing crystals operating in the far ultraviolet. The ultimate time resolution of an up-conversion experiment is set by the excitation pulse duration.

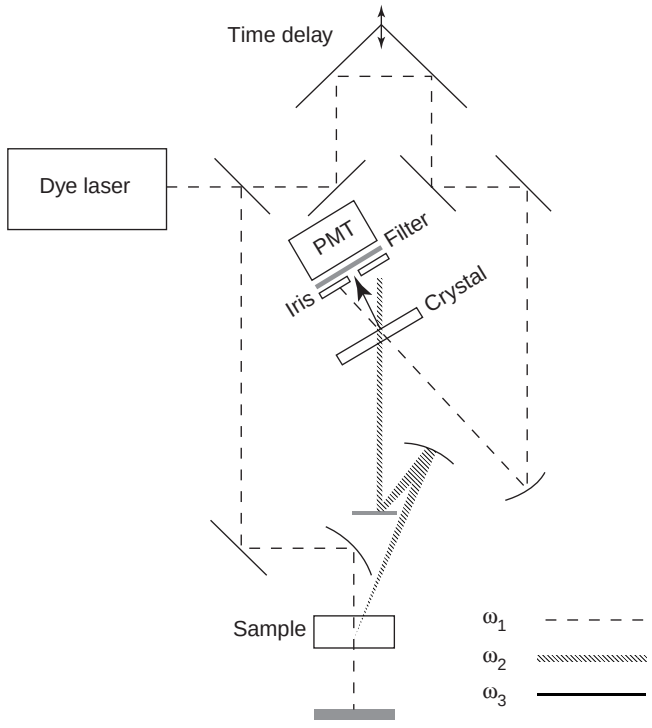

Figure 4 Simplified block diagram of the instrumental set-up used for an up-conversion experiment. A fraction of the laser output is used to excite the sample while the remainder is variably delayed for use in the up-conversion process. Fluorescence is collected via an off-axis curved mirror and directed into the up-conversion crystal. The resulting ultraviolet output intensity of the crystal is proportional to the intensity of the fluorescence at a particular delay time photomultiplier tube (PMT). (Adapted from Wirth.)[16]

2 INSTRUMENTATION

Since the early 1980s, significant advances have been made in instrumentation for time domain and frequency domain fluorescence spectroscopy. With present time domain instrumentation and optical up-shifting techniques, femtosecond processes are now being routinely explored.

2.1 Time Domain[10,12]

The additional specialized equipment needed to implement TCSPC includes a pulsed light source such as a laser or spark discharge, a fast response detector, and a time-to-amplitude convertor (TAC). Figure 5 shows a simplified schematic of a TCSPC instrument. It is important to mention here that several additional instrumental refinements are required but are of secondary importance and beyond the scope of this article. The reader is directed to O'Connor and Phillips[10] for a more detailed discussion. The TAC, the most critical device in the time correlation process, measures the elapsed time between the initial rise in intensity of the pulsed light source and the detection of an emitted photon.[12] In operation, an electrical pulse is generated at a time exactly correlated with the time of generation of the optical pulse by the trigger (e.g. a PMT output, an antenna pick-up, or a logical synchronized pulse). After receiving the trigger pulse, the TAC initiates the charging of a capacitor plate which is routed to the TAC start input via a discriminator. While the same optical pulse that triggered the TAC excites the fluorescent sample, an aperture is adjusted on the emission arm so that only one photon is detected for every 100–200 excitation events. The signal resulting from this photon stops the charging ramp in the TAC, which transmits a pulse. The amplitude of this pulse is proportional to the charge in the capacitor, and hence to the time difference between the start and stop pulses. The TAC output pulse is given a numerical value within the analog-to-digital converter (ADC) and a count is stored in the data storage device (a multichannel pulse height analyzer; MCPHA) in an address corresponding to

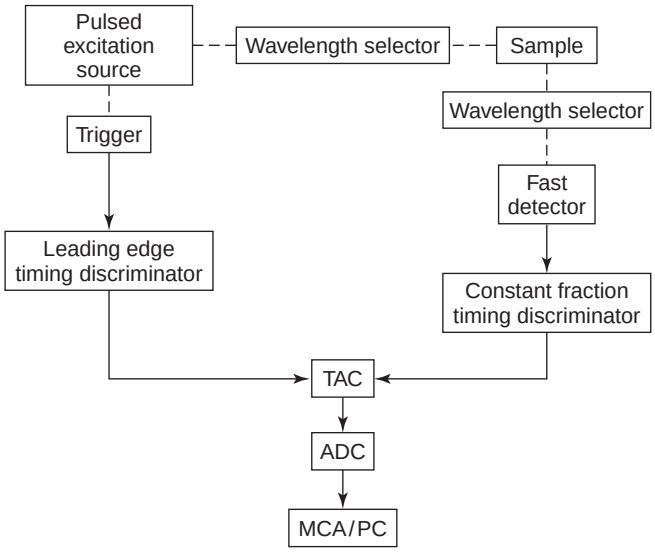

Figure 5 Simplified block diagram of a conventional TCSPC apparatus (Adapted from O'Connor and Phillips[10] and Cline-Love and Shaver.[12]) Please refer to the text for further experimental description. MCA/PC, Multichannel Analyzer Personal Computer.

that number. Excitation and data storage are repeated in this way until the histogram of number of counts against address number in the storage device represents, to some required precision, the decay curve of the sample.[10] A typical TCSPC trace from a flash lamp-based instrument is illustrated in Figure 1. With current pulsed laser sources and microchannel plate PMT detectors, it is possible to determine excited state processes routinely on the low picosecond timescale.

2.2 Frequency Domain

In the frequency domain,[13] the latest instrumentation has made it possible to measure picosecond decay kinetics routinely. These advances have been made possible by the use of excitation sources with large bandwidths and detectors with high gain and wide bandwidths.[13]

A simplified schematic for a multifrequency, cross-correlation[14] phase and modulation instrument that is capable of parallel acquisition of phase and modulation data is presented in Figure 6. In a frequency domain measurement, the excitation source can be an intensity modulated CW light source or a high repetition rate pulsed laser. The choice of source usually depends on the required wavelengths, the power level needed, and the experience and other constraints of the researchers. Laser sources can be easily modulated using one of several modulation schemes.[13] In most modern frequency domain instruments, the harmonic content of pulsed lasers produces modulation to many gigahertz such that the bandwidth of frequency domain instruments is now limited only by the detector response time.[14]

To illustrate the range of light sources used in phase modulation fluorescence, we have selected two extreme examples. Sipior et al.[18] demonstrated the use of an inexpensive, commercially available blue light emitting diode (LED) to produce amplitude-modulated near-ultraviolet light as a source for nanosecond phase-modulation fluorescence lifetime measurements. The most recent advance in multifrequency phase and modulation fluorescence has been the development of instrumentation for the parallel acquisition of phase and modulation data.[19–21] In this design, the phase and modulation across the entire

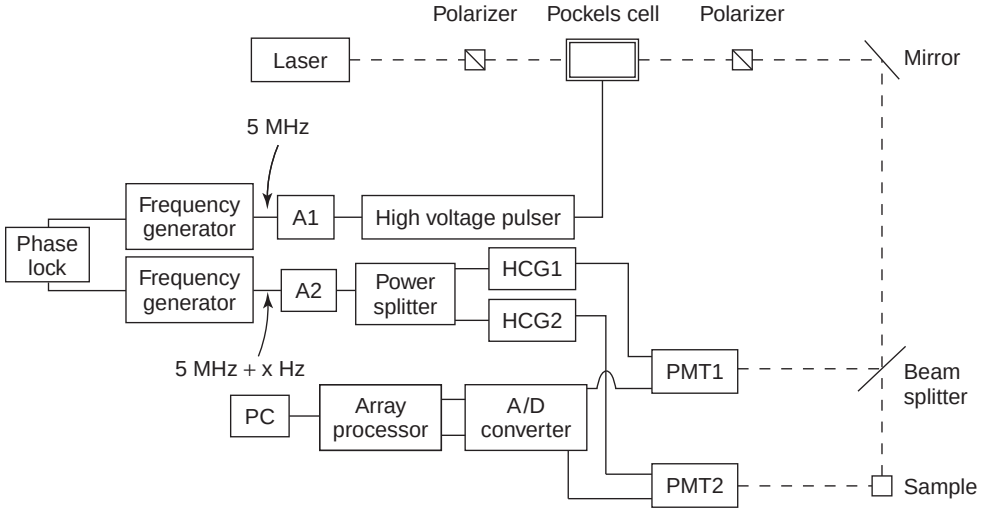

Figure 6 Simplified schematic of a multiharmonic Fourier phase and modulation fluorometer. The dashed lines denote optical paths while the solid lines indicate electrical connections. Abbreviations represent power amplifiers (A1, A2); PMT detectors (PMT1, PMT2); and harmonic comb generators (HCG1, HCG2). (Adapted from Bright et al.)[13]

frequency spectrum can be collected simultaneously and experiments that took about 1 h can now be completed in <10 s.[13] More recently, a parallel multiharmonic frequency domain fluorometer was designed by Bright et al. that can rapidly measure excited state fluorescence anisotropy and intensity decay kinetics following one-, two-, or three-photon excitation.[21]

3 SOLVATION DYNAMICS

The time evolution of an emission spectrum provides insight into how the local solvent environment affects the emission process. Time-resolved fluorescence experiments have been used to study the energetics of the solvation process, because fluorophores provide an internal clock from which to observe processes occurring on the fluorescence timescale. That is, if a secondary process (a reaction) occurs on the same timescale as fluorescence, it can often perturb the fluorescence intensity decay kinetics. Thus, the kinetics of the secondary process can be inferred via its impact on the time-resolved fluorescence decay. Bakshiev, Mazurenko, and co-workers pioneered the use of time-resolved fluorescence for transient solvation studies (see Ravichandran and Bagchi,[22] Simon,[23] Nagarajan et al.[24]). Initial experiments in this field were limited by the nanosecond time resolution of early instrumentation. As a consequence, the majority of these experiments were carried out in viscous solvents, at relatively low temperatures, and/or on samples dissolved in nonpolar (noninteracting) solvents modified (perturbed) by the addition of small amounts (<1% by volume) of a polar cosolvent. In this way, these investigators were able to bring the solvation process ''down to'' the time regime of their instrumentation. Recent advances in instrumentation have now made it possible to perform studies in neat liquid solvents like acetonitrile, water, methanol, and other alcohols. Subpicosecond spectroscopy has also been used by to study the microscopic solvation dynamics of less viscous liquids.[24]

In order to understand the time evolution of spectral shifts due to universal excited solute–solvent interactions, it is useful to invoke the Franck–Condon principle. Upon excitation, a molecule will retain the particular solvent configuration which is favored in the ground state, because solvent molecules cannot reorient during the excitation event. If the excited state, formed directly by light absorption or indirectly via a fast subnanosecond reaction, has a new dipole moment, and if the solvent relaxation time is less than the excited state fluorescence lifetime, the solvent dielectric medium surrounding the fluorophore will be polarized differently from in the ground state. Thus, a new reaction field developed by this orientation strain is created by the fluorophore and the solvent molecules surrounding this new dipole will relax and reorient to accommodate the change in dipole moment of the excited fluorophore until equilibrium has been restored. When this solvent relaxation time becomes comparable to the excited state fluorescence lifetime, time-resolved spectral shifts in addition to electronic relaxation will be observable.

Pierce and Boxer reported the time-dependent solvation for a fluorophore within a protein by following the time-dependent fluorescence Stokes shift.[25] Specifically, these authors reported on the complex formed between sperm whale apomyoglobin and 2′-(*N*,*N*-dimethylamino)-6-naphthoyl-4-*trans*-cyclohexanoic acid at temperatures between 243 and 298 K. The decay of the mean emission energy was not single exponential and there were decay components spanning and probably exceeding their experimentally observable time window (20 ps to 20 ns). These results suggested that either the activation energies of the rate-limiting motions in the relaxation are dependent on substrate conformation, or that different types of protein motion with different characteristic frequencies participate in the relaxation.

Mertz et al.,[26] while investigating how the Stokes shift of proflavin was affected by solvent dipolarity, reported that the dependence of the solvent reorganization free energy on the solvent dielectric properties can be explained if two effects are taken into account. The first effect is the distortion of dielectric properties of a uniform solvent by embedding a bulk solute molecule in the solvent. The second effect is the influence of the solvent dipolarity on the electronic density redistribution upon an electronic transition event. These particular data also supported the notion that the commonly adapted equation for Stokes shift that establishes equality between Stokes shift and the doubled reorganization energy is not rigorously true, in general. Indeed, these authors showed that the measured experimental data are consistent with a more rigorous and appropriate approximation of Stokes shift that takes into account the quantum nature of the solute local modes.

Rempel et al., using temperature-dependent time-resolved Stokes shift measurements of Coumarin 153 (C153), explored the solvation dynamics in binary mixtures of alcohols and alkanes and concluded that, depending upon the actual alcohol concentration, the Stokes shift takes place on a timescale ranging from 300 ps to several nanoseconds.[27] These characteristic times were two orders of magnitude longer than the relaxation times typically observed for C153 dissolved in neat alcohols. This effect was attributed to solvent/cosolvent rotational reorientation. These authors also observed that the temporal behavior of the Stokes shift was monoexponential and the time constants depended

linearly on the alkane viscosity. They concluded that their data strongly support a diffusion-controlled process of solvation in these mixtures.

Glasbeek et al. used fluorescence up-conversion of 1-phenyl-4-[(4-cyano-1-naphthyl)methylene]piperidine (PCNMP) in its lowest excited charge transfer state when dissolved in a series of solvents.[28] These experiments were performed at room temperature with 300 fs time resolution. The observed Stokes shift was attributed to the solvent response to the large photoinduced PCNMP dipole moment ($\sim$30 D). The solvation dynamics were reported to be governed by the rotational diffusional motions of the ethereal solvent molecules.

Solvent relaxation of indole dissolved in polar protic solvents and analysis of the fluorescence lifetime distributions and time-dependent spectral shifts were reported by Demchenko et al.[29] These authors used the maximum entropy method (MEM) and obtained relatively narrow peaks for the fluorescence lifetime distribution, which originate first from the main emission decay component and second from additional short components which are due to spectral kinetic processes. These short-lived components were extremely sensitive to temperature. The authors were also able to use time-resolved emission spectroscopy to reveal relaxation times over the same range recovered for the observed lifetime distributions. Results of excited state stabilization energy measurements, combined with time-dependent spectral shifts, were found to be consistent with a mechanism of general dielectric solvent relaxation rather than formation of binary excited state complexes in the time range studied.

Wendt and Richert recently reported the Stokes shift and Stokes shift dynamics for naphthalene (change in dipole moment upon excitation $\sim$0 D) and quinoxaline ($\sim$1.3 D) dissolved in *n*-propanol and other glass-forming solvents.[30] The time dependence of the Stokes shift for naphthalene was reported governed by the structural (or shear stress) relaxation time of *n*-propanol, without any signature from the strong dielectric relaxation. For this solvent, structural and dipolar contributions can be distinguished because the timescale for dipolar reorientation is 25-fold slower than the shear stress relaxation time. These authors concluded that the naphthalene solvation reflects excitation-induced changes in the van der Waals interactions, which makes it an ideal probe for assessing shear stress or mechanical relaxations on microscopic spatial scales near the glass transition temperature (T_g).

In a recent review, Fleming and Cho present experimental and theoretical studies of polar solvation dynamics and spectral line broadening.[31] Spectral broadening, arising from interactions with the environment in the time domain (the dephasing timescale), is an effect arising from the averaging over the ensemble that is intrinsic to linear spectroscopic measurements. An important feature of time-resolved fluorescence and ultrafast spectroscopy is that they can be used to undo the ensemble average and allow molecular dynamics to be observed directly in condensed phases.

Bright et al., during their investigation of the behavior of acrylodan-labeled bovine serum albumin and acrylodan-labeled human serum albumins (BSA-Ac and HSA-Acs) sequestered within a tetramethylorthosilane-derived xerogel, used the dipolar relaxation kinetics to quantify the relaxation of the Ac fluorophore cybotactic region following optical excitation.[32] These results suggested that there was a substantial amount of nanosecond and subnanosecond dipolar relaxation within the local environment surrounding cysteine-34 in both proteins, even when they were sequestered within ''dry'' xerogels. The solvation dynamics of BSA-Ac and HSA-Ac differ with the initial xerogel preparation, suggesting that the slight differences in amino acid sequences do indeed affect the cybotactic region surrounding the Ac residues within these proteins.

4 APPLICATIONS IN BIOCHEMISTRY AND BIOPHYSICS

Time-resolved fluorimetry is probably second only to NMR (nuclear magnetic resonance) in its ability to solve diverse and significant problems in the biochemical and biophysical sciences.[33,34] In particular cases, stemming from signal inadequacy, sample scarcity or special constraints, fluorescence spectroscopy is uniquely suited to tackle such problems. Areas of recent study include immunoassay, enzymatics, protein–ligand binding, biosensor development and diagnostics, denaturation kinetics, nucleic acid determination, oligonucleotide hybridization, and protein self-association.

Time-resolved fluorescence is an invaluable tool for studying processes involving primary energy and charge transfer in photosynthesis and photosensitization. For example, our understanding of photosynthesis was largely achieved by performing dynamic fluorescence experiments over several timescales.[33,35] As a result of such efforts, a contemporary view of photosynthetic kinetics has been realized. Although many aspects of the process are still veiled in mystery, the most widely accepted model of the photosynthetic membrane includes four features essential to the complex electron-transfer mechanics: (1) photosystem I (PS I) and II (PS II) reaction centers which operate in series to effect charge separation, (2) a pigment-binding protein light-harvesting assembly, (3) a membrane to isolate products physically, and (4) requisite redox and assembly enzymes.

In photosynthetic pathways, appropriately energetic trapped photons promote electrons into higher energy excited states. Electrons in these elevated states, however, tend to decay to ground singlet states on a nanosecond timescale, wasting potential energy in the form of light and heat. In order to prevent this scenario, chromophores must donate excited electrons to quinone acceptors faster than the rate of excited state decay (i.e. recombination). The resulting charge separation at the reaction center chemically preserves some of the incident photon energy in the form of a stable transmembrane electrical potential.

The role of protein–chromophore interactions in the mechanism of energy transfer was probed by de Paula et al.[35] by using time-resolved fluorescence to study chlorophyll a bound to the 47 kDa antenna protein (CP47) of spinach PS II. These authors used picosecond fluorescence spectroscopy to characterize CP47 and distinguish between functional (intact or perturbed) and nonfunctional (denatured) centers in various formulations. The structural heterogeneity (e.g. donor–acceptor distance, vibronic coupling) gives rise to multiple recombination rates which are manifest in the multiexponential decays observed.

The field of photosensitization is another expanding area of application for dynamic fluorimetry. For example, photodynamic therapy (PDT) involves the administration and photoactivation of a suitable photosensitizing agent as a medical means of treating solid, localized tumors.[33,36] In PDT, it is generally accepted that energy transfer takes place from a luminophore excited triplet state to ground state triplet oxygen to produce the cytotoxic singlet oxygen. Although numerous dyes such as the acridines, coumarins, and porphyrins have been studied for their tumor-localizing properties and effectiveness as potential PDT agents, research has increasingly centered on hematoporphyrin derivatives and Photofrin II® (a proprietary drug marketed by Quadra Logic Technologies Inc.). Unfortunately, Photofrin II® has been found to be a complex mixture of monomers, ''dimers'' (dihematoporphyrins formed by ether or ester linkages), and aggregates. Further, the most tumor-localized components are not necessarily the best photosensitizers making pharmacokinetic assessments difficult. Time-resolved fluorescence has been used to differentiate between these various components providing useful information about the uptake, retention, and aggregation dynamics of the drug forms in tumors and tumor-free tissues. In addition, new or unexplored molecules can be characterized as potential lead compounds for PDT based in part on their excited state luminescence lifetimes. Chlorins and bacteriochlorins are currently being investigated by Pandey and others because these agents exhibit high absorbances at higher wavelengths allowing them to be more effectively photoexcited when they are loaded more deeply into cancerous tissue making the treatment of large deep-seated tumors more effective.[33,36]

Our understanding of the complexity of genomes in various cell types has grown considerably with the development of recombinant DNA techniques and sensitive methods for analyzing the organization of specific genes. Such techniques rely on specific nucleic acid hybridization probes which have become irreplaceable in contemporary molecular biology for the detection of specific, complementary nucleic acid sequences. However, stability, detection sensitivity, safety, and disposal problems common to use of radioisotopes (^{32}P, ^{3}H, ^{13}C) have stimulated interest in nonisotopic alternative labels. The high specific activity and simplicity associated with the use of luminescent probes (e.g. acridine), when measured by time-resolved fluorometry, provide powerful nonradioactive substitutes in the field of nucleic acid hybridization. Popular among these alternatives, lanthanide labels covalently bound to nucleic acids have shown unique promise in areas of diagnostic medicine. To date, there have been applications in areas like genetic predisposition to disease including inherited disease, viral infection, bacterial identification, and antibiotic sensitivity testing.[33,34]

A novel technique for screening point mutations was developed for diagnosis of familial defective apolipoprotein B-100 (FDB) by coupling polymerase chain reaction (PCR) performed on the appropriate region of the apoB gene with time-resolved fluorometric detection.[37] Like its cousin familial hypercholesterolemia (FH), FDB affects the interaction between cellular low density lipoprotein receptors and the proper ligands on plasma lipoproteins apoB-100 and apoE. However, unlike FH where a large number of various mutations occur, FDB involves a single point mutation corresponding to amino acid position 3500. In practice, the PCR amplification products are hybridized in situ with Eu^{3+}-labeled oligonucleotides complementary to either the wild type or the mutant genome. This allows the presence or absence of the apoB-100 mutation to be monitored based on the europium chelate excited state lifetimes. When a sample population of 127 Swedish patients clinically diagnosed with heterozygous FH were screened by this method, two were correctly identified as actually suffering from FDB.

By following a similar strategy, a method for simultaneous multicolor analysis of PCR products for any of seven amplified human papilloma virus (HPV) types was reported.[38] HPV infection has been associated with multiple cancer types (primarily anogenital) and identification is, as such, important to determine the risk of cervical carcinoma in women. In this approach, the authors perform a seminested PCR in a single tube using a biotinylated inner primer. Amplification products were captured on streptavidin-functionalized 96-pronged manifold solid supports, and then denatured to render

single-stranded molecules serving as targets in subsequent hybridization reactions. These targets were then probed with a mix of seven type-specific (HPV types 16, 18, 31, 33, 35, 39, and 45 associated with cervical cancer), differentially labeled nucleotides containing 10 or 20 lanthanide chelates at the 5′ ends with seven distinct combinations of Eu^{3+}, Tb^{3+}, and Sm^{3+} ions. After a washing step in microtiter wells, the lanthanide ions were released from the bound probes. This allowed the seven viral strains to be correctly identified by time-resolved fluorescence measurements.

Proteins are dynamic in nature and their function and behavior are intimately related to their structure, conformation, and dynamics. Thus, if one aims to understand and exploit events that rely on protein conformational changes (e.g. protein–drug/ligand association, antigen/hapten–antibody binding, enzyme–substrate interactions, protein–membrane insertion, protein self-association), it becomes necessary to track, quantify, and ultimately understand the behavioral aspects of proteins as they undergo net conformational changes (measurements at or near equilibrium) and particularly as conformations change in near real time. Advances in protein physical chemistry have been made as a result of combining novel fluorescent probes and coupling chemistries with the information rich nanosecond and subnanosecond time-resolved optical techniques. In contrast to techniques like NMR and X-ray crystallography, subnanomolar protein concentrations can routinely be studied by fluorescence allowing assessment of association parameters under true equilibrium conditions. Recent results by Ingersoll et al.[39] illustrate the potential of on-the-fly phase-modulation fluorescence to track the nanosecond and subnanosecond dynamics of BSA-Ac. In these experiments, the authors determined the temporal evolution of fluorescence anisotropy and intensity decay kinetics of BSA-Ac when it was challenged with trypsin or β-mercaptoethanol (BME) (Figure 7). Both the BME-induced disulfide interchange and the trypsin-mediated protein digestion were successfully tracked on-the-fly by multiharmonic Fourier fluorescence suggesting the potential of dynamic optical techniques for real time reaction monitoring.

Molecular behavior at interfaces is of critical importance in areas ranging from energy conversion, chromatography, and catalysis to toxicity, adhesion and transport properties, and chemical sensing. By using time-resolved fluorescence in concert with interfacially selective detection (generally by exploiting some total internal reflection configuration), the underlying kinetics and mechanisms associated with surface bound molecules can be accessed even in aqueous media.[40] Such information is relevant to the study of biosensor interfacial dynamics where the ultimate performance is governed by

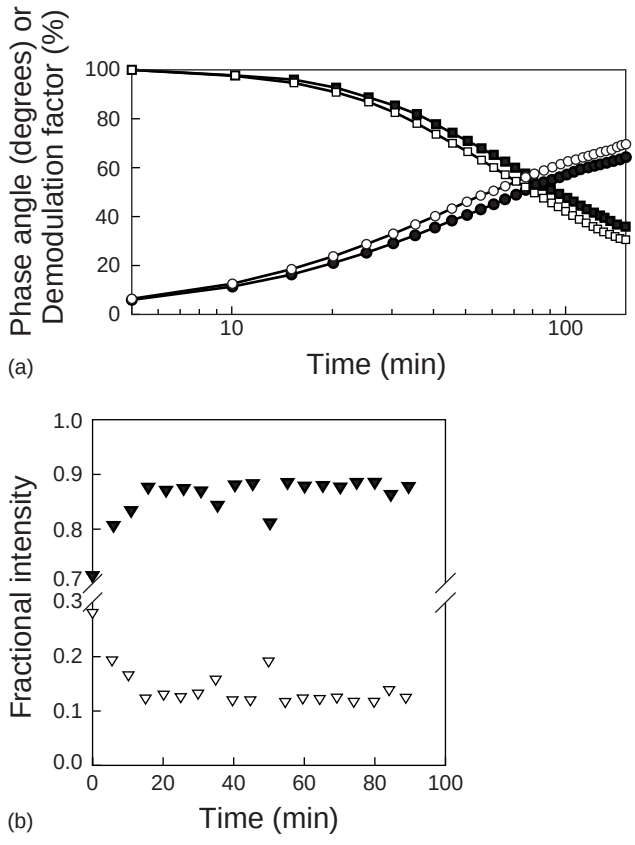

Figure 7 On-the-fly frequency domain MHF intensity decay data for BSA-Ac reacting with BME. (a) Multifrequency phase angle (●, ○) and demodulation factor (■,) data before and after reacting for 90 min with BME. Filled points, 0 min, open points, 90 min. (b) Residuals, frequency-dependent phase ( ) and modulation (Δ) plots for frequency domain data sets acquired on-the-fly at the times indicated. The fractional intensities associated with τ_1 and τ_2 are shown. The apparent excited state fluorescence lifetimes associated with f_1( ) and f_2 (▽) are 3.97 ± 0.08 ns and 1.21 ± 0.04 ns, respectively and they are independent of reaction time. (Reproduced with permission from Ingersoll et al., *Applied Spectroscopy*.)[39]

the complex interplay among the biorecognition element (BRE) structure and the conformation at the interface, the ease with which the target analyte can reach and interact with the BRE, and the homogeneity, stability, and mobility of the BRE at the interface. In this context, the Bright group developed multifrequency phase modulation total internal reflection fluorescence (MPM/TIRF) to study in situ the interfacial distribution of surface immobilized BRE dynamics.[41] Initial work showed that an immunosurface composed of F(ab′) antifluorescein polyclonal antibodies at a fused silica surface on interacting with free fluorescein hapten consisted of a broad distribution of microdomains. In contrast, the F(ab′)–fluorescein system in solution was described by a single discrete microdomain (i.e. excited state fluorescence lifetime).

That is, an ensemble of F(ab′) fragments having slightly different binding site accessibilites, surface-induced conformational changes, and/or actual affinity constants exists at and is modulated by the surface. Also, these species change/drift with storage time causing loss of immunosurface integrity over time.

A powerful hybrid technique called phase-resolved evanescent wave induced fluorescence (PREWIF) was reported by Lundgren et al.[42] for studying interfaces by combining the interfacial selectivity inherent to TIRF (total internal reflection fluorescence) with the temporal and spectral resolution of phase-resolved fluorescence. The potential of PREWIF for in situ studies of heterogenous interfaces under ambient conditions was demonstrated by resolving individual components within thin sol–gel-derived glass films doped with TNS/PRODAN (toluidinyl naphthalene sulfonic acid/6-propionyl-2-(*N*,*N*-dimethylamino)naphthalene), rhodamine 6G (R6G)/eosin Y, or R6G monomer/R6G. These researchers were also able to resolve the static emission from an ~100 Å thick immunosurface composed of protein G-antifluorescein-bound fluorescein even in the presence of excess free fluorescein in the bulk. PREWIF was also used to recover the relaxed and unrelaxed spectra contributing to the overall fluorescence from BSA-Ac physisorbed onto a quartz surface.

5 POLYMER PHOTOPHYSICS

Processes involved in polymer chain dynamics span a wide range of time and lengthscales. The large scale motions of polymer chains have historically been analyzed in terms of normal modes where higher order modes correspond to motions on faster timescales and on shorter lengthscales. Polymer chain dynamics on sufficiently short lengthscales depend greatly on the chemical structure of the polymer chain itself.[43–50] Polymer photophysics, both in its historical development as well as its current practice, can be divided into a few regimes of study: excimer formation, fluorescence anisotropy, fluorescence quenching, and excited state energy migration. Time-resolved optical spectroscopy provides a uniquely detailed picture of local segmental motion during polymer chain dynamics.[43] This is accomplished by observing the time dependence of the polymer orientation autocorrelation function directly. A common feature of time-resolved fluorescence methods is the use of nonfluorescent polymer chains with an extrinsic chromophore label attached. Alternatively, the emission from an intrinsic polymer-bound fluorophore can be used to investigate the energy migration as well as polymer.[43,44]

A typical time-resolved fluorescence experiment measures the rotational reorientation of a fluorophore. If the said fluorophore is part of a polymer chain or in registry with the chain, then polymer chain motions can be inferred by observing the chromophore rotational reorientation. The detection limits of fluorescence allow such measurements to be made even at a level of one chromophore per 1000 oligomeric repeat units (i.e. a chain concentration of 0.1%). When a chromophore with a strongly polarized electronic transition is excited by polarized light, those chromophores with absorbance transition moments aligned in the direction of the excitation polarization are preferentially excited. If these chromophores emit before they significantly reorient or depolarize, the fluorescence will also be preferentially polarized (i.e. anisotropic) along this direction. The fluorescence polarization depends upon the interrelation of two physical parameters: the excited state fluorescence lifetime and the molecular rotational reorientation time.[44]

Most of the work using time-resolved fluorescence to investigate local polymer dynamics has been performed using dilute solutions of fluorescently labeled polymers.[44] These studies provide an opportunity to understand local chain motions without considering the complicating effects of chain–chain interactions present in more concentrated systems. It is important to mention here that the timescales for local dynamics in solution are comparable to the fluorescence timescale. Recently, a charge transfer fluorescence molecule was used as a probe to study the structure of polystyrene latices and polystyrene–glycidylmethacrylate core–shell latices.[45] Information on the composition of the latex was obtained by using steady-state and time-resolved fluorescence measurements. The fluorescent probe, PCNMP, that was used in these studies has a dipolar excited state which is populated by excitation in the ultraviolet range by an intramolecular electron transfer process. As is typical for charge transfer luminescence, however, the emission spectra are rather broad, limiting static fluorescence to qualitative observations because the broad emission spectra make it difficult to distinguish between several coinciding species. By using excited state intensity decay information as an additional experimental observable, Hofstraat et al.[45] were able to provide a clearer picture of such a complicated system.

In another study, time-resolved fluorescence spectra of poly(*N*-vinylcarbazole) in solution were analyzed by using a principal multivariate spectral estimation method.[46] The data clearly indicated that the polymer emission spectra were composed of precisely three emitting species (monomer, partial overlap excimer, and sandwich excimer). Intensity rise as well as decay curves of the fluorescence of each component are complex and depend on the polymer tacticity. These authors determined that the partial overlap excimer was formed not only by ultrafast trapping in the pre-existing site

but also through dynamic processes such as energy migration. It was further reported that the excited state monomer and the partial overlap excimer were produced via sandwich excimer dissociation and reactivation. In organized molecular systems, structural inhomogeneity, giving rise to various excited state species is ubiquitous, suggests the broad applicability of such an approach.

Time-resolved fluorescence anisotropy measurements are particularly attractive because they allow work to take place under dilute conditions and they provide an intrinsic molecular ruler (i.e. the dimensions of the fluorophore). Time-resolved fluorescence anisotropy techniques have been used to investigate fluorophore–polymer interactions in the bulk, to follow subunit dynamics, and to track the motion of the elastic cross-link junctions within various polymers. According to Soutar et al., in studies of macromolecular dynamics, emission anisotropy measurements are the most informative of the luminescence-based approaches that can be adopted.[47,48] Fluorophores simply dispersed in the medium of interest can furnish information on the local free volume accessible to the probe whereas covalently bound fluorophores can reveal details regarding the kinetic behavior of chain segments, termini, and substituent groups. Bolstering their claims, Soutar et al. reported on the intramolecular segmental relaxation behavior of poly(methyl methacrylate) (PMMA) and poly(methyl acrylate) (PMA) at low concentration dissolved in dichloromethane by using time-resolved fluorescence anisotropy.[47] The fluorescence anisotropy decay kinetics of each polymer, over the temperature range of 230–310 K, were adequately described by single exponential models for acenaphthylene- and 1-vinylnaphthalene-based labels. The recovered rotational reorientation times for segmental motion exhibited an Arrhenius dependence in the temperature range studied, giving rise to activation energies of 14 and 11 kJ mol^{-1} for PMMA and PMA, respectively. These values are considerably lower than those reported for either polymer in other common solvents. In fact, the magnitude of the differences in activation energies are too large to be explained on the assumption that the solvents function solely to provide frictional resistance to the polymer dynamics. These authors suggested that: (1) PMMA and PMA exhibit specific interactions with dichloromethane and/or other solvents, such as toluene or (2) the naphthyl labels may actually experience specific interactions with dichloromethane which distort the apparent polymer behavior.

Stein, Hoffman, Frank, and Fayer have used time-resolved fluorescence anisotropy decays to investigate the reorientational dynamics of a cross-link junction in poly(dimethylsiloxane) networks from a chromophore attached to the cross-link between $T_g + 75$ K to $T_g + 150$ K.[49] At the junction, the probe chromophore, 1-dimethylamino-5-sulfonylnaphthalene amide (dansyl amide), lies pendant to a silicon that acts as a trifunctional cross-linking atom. In cyclohexanol, the fluorescence anisotropy decays for the chromophore are in agreement with the Debye–Stokes–Einstein hydrodynamic theory of rotational diffusion, demonstrating that the cross-linker can be used as a probe of reorientational relaxation. The fluorescence anisotropy decay kinetics are remarkably fast for a cross-linked poly(dimethylsiloxane) network, reflecting fast rotational reorientation of the cross-link junction. This reorientation appears diffusive and has a temperature dependence in accord with the Williams–Landel–Ferry equation.[49] A model was proposed suggesting that reorientational and translational motion of the cross-link occur simultaneously and both are coupled to fluctuations of the polymer chain termini.

Niemeyer and Bright used time-resolved fluorescence anisotropy and reported on the effects of polymer molecular weight, temperature, and near- and supercritical CO_2 gas sorption on the rotational reorientation dynamics of a model solute, *N*,*N*′-bis-(2,5-di-tert-butylphenyl)-3,4,9,10-perylenedicarboximide (BTBP), dissolved in poly(dimethylsiloxane).[50] These authors showed that there is a linear correlation between the BTBP rotational reorientation time and the poly(dimethylsiloxane) polymer bulk density. Temperature-dependent studies of the BTBP/poly(dimethylsiloxane) system showed that the BTBP rotational reorientation dynamics are accurately described by an Arrhenius activation model where the recovered activation energies for the BTBP rotational reorientation are statistically equivalent to those for poly(dimethylsiloxane) viscous flow. The addition of CO_2 to the BTBP/poly(dimethylsiloxane) system led to appreciably decreased BTBP rotational reorientation times with increased CO_2 pressure. These results illustrate how polymer dilation by CO_2 can be used to tailor the dynamics within polymer systems.

6 APPLICATIONS IN THE SEPARATION SCIENCES

Because the chromatographic retention time and excited state fluorescence lifetime of a solute are independent phenomena, they contain orthogonal information about the identity of a particular analyte. Time-resolved fluorescence techniques have been applied to reverse-phase HPLC (high-performance liquid chromatography), capillary electrophoresis (CE) and other separation techniques.[51–56]

Loss in selectivity due to overlapping peaks is a major problem encountered in the detection process based on retention times alone. Further, detection at only one or a

few emission/excitation wavelengths may be inadequate to resolve overlapping ''chromatographic'' peaks, while the use of an array detector to collect an entire spectrum can decrease sensitivity as a result of dispersion. Thus, if selectivity is not substantial, it may be impossible to detect the presence of minor components or matrix effects.[53] Lytle et al. pioneered using time-resolved fluorescence to identify and resolve overlapping chromatographic peaks as well as to indicate the presence of impurities and matrix effects.[51] In the Lytle method, the eluent excited state fluorescence lifetime is determined on-the-fly during liquid chromatographic separation. The lifetimes provided the additional selectivity needed to confirm peak assignments based on retention time. Lytle and his associates emphasized that a combination of coarse excited state fluorescence lifetime measurement and retention time yielded more information about the identity of an unknown than extremely refined measurement of either one alone. The Lytle measurements were made in the time domain by using a pulsed laser for excitation and a PMT detector. The PMT anode current was split in half and one portion was delayed by 10 ns before being sent to one input of a two-channel oscilloscope. Two intensity chromatograms were thus obtained representing two different delay times separated by 10 ns. A ratio of the two chromatograms provided a ratiogram in which each lifetime was identified by a unique concentration-independent ratio, whereby a changing ratio across a peak was indicative of coelution. Results of experiments with polycyclic aromatic hydrocarbon (PAH) standards of known identity were used to illustrate the methodology. Further, 15 PAHs were positively identified in combustion products, illustrating the applicability of these lifetime chromatography measurements to real samples.

Frequency domain fluorescence has been used to detect and resolve overlapping peaks on-the-fly in reverse-phase HPLC.[52–54] Direct, simultaneous measurements of fluorescence intensity, phase-shift, and demodulation were made at 1 s time intervals for PAHs as they eluted from the chromatograph.[52] During early investigations, these experiments were hindered by limitations of the frequency domain spectrofluorometers, which could only collect data at one modulation frequency per chromatogram thereby necessitating multiple sample injections. The resulting chromatograms had to be perfectly overlaid and the data combined at each point along the chromatographic peaks.[52] Smalley et al.[53,54] have investigated on-the-fly fluorescence lifetime detection in HPLC using a MHF phase-modulation spectrofluorometer,[19,20] which eliminates the problem of multiple injection and offers other advantages as well. These authors showed that the MHF data acquired on-the-fly during a single chromatographic run contained all the information needed to determine the eluent excited state fluorescence intensity decay kinetics and to indicate and resolve overlapping peaks at intervals as short as several milliseconds. This allowed essentially continuous monitoring of the intensity decay kinetics and static emission intensity during chromatographic elution. This work demonstrated accurate lifetime detection and peak resolution for simple mixtures of PAHs by HPLC. This technique can be used for component identification, indication of coeluting peaks, matrix effects or impurities, and resolution of overlapping peaks. This technique also requires no a priori assumptions about spectral features, chromatographic peak shapes, or fluorescence lifetimes.

Soper et al. investigated the dynamic (on-line) measurement of excited state fluorescence lifetimes for several chromophores separated by free solution CE using time-correlated single-photon counting with NIR (near-infrared) excitation.[55] The chromophores used in this work were tricarbocyanine dyes that show absorption and emission properties in the NIR and possess excited state lifetimes between 500–1000 ps. The instrument used for this study consisted of a solid-state, passively mode-locked Ti-sapphire laser for excitation and a single photon avalanche diode detector operated in a Geiger mode (Figure 8). The instrument response function of this device was 165 ps fwhm, appropriate for making subnanosecond lifetime measurements in CE applications. The authors demonstrated electrophoretic separation of two NIR dyes, DTTCl (cationic) and IR-125 (anionic), in a 95 : 5 methanol : water running buffer (pH = 9.5). The excited state fluorescence lifetimes were calculated using maximum likelihood estimators (MLE) methods.[55] At a loading level of 1.42 zmol for IR-125 and 49 zmol for DTTCl, the excited state fluorescence lifetime values were determined to be 482 ± 14 ps for IR-125 and 943 ± 23 ps for DTTCl, which agreed favorably with the values determined for these dyes at high concentrations in static solutions. To demonstrate the feasibility of making excited state fluorescence lifetime determinations in capillary gel electrophoresis, where the gel can produce high scattering backgrounds, the excited state fluorescence lifetimes of C-terminated fragments (produced from the M13 mp 18 template labeled at the 5′ end of a universal M13 sequencing primer with a NIR fluorescent tag) were determined. The mean excited state fluorescence lifetime for 30 different peaks in the electropherogram was 581 ps and the standard deviation was 9 ps.

On-the-fly frequency domain fluorescence lifetime detection in CE was demonstrated by Li and McGown.[56] Virtually continuous detection was achieved by interfacing a commercial CE instrument with a commercial MHF instrument. These authors modified the CE capillary cartridge to allow the capillary to pass through

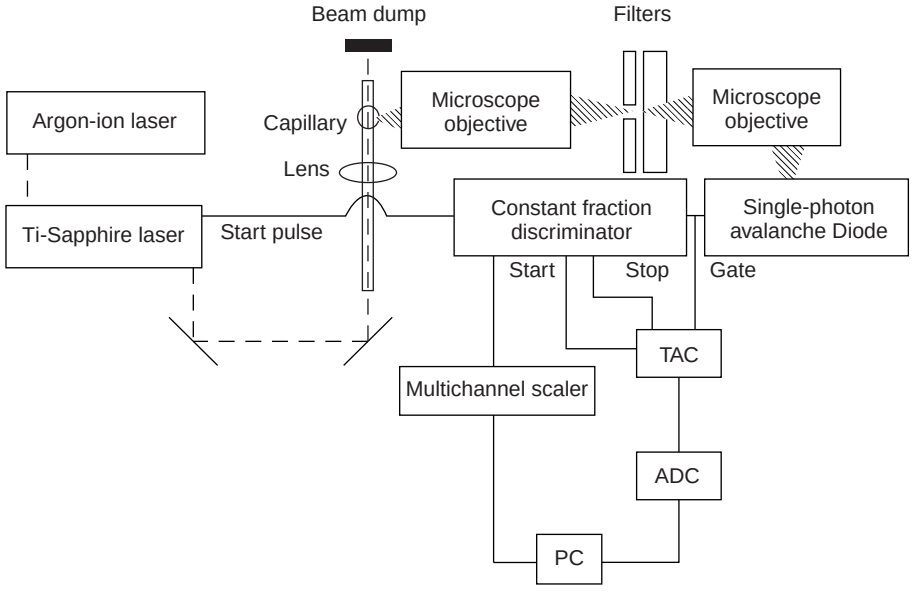

Figure 8 NIR time-correlated single-photon counting instrument used for excited state fluorescence lifetime determinations in a CE experiment. (Adapted from Smalley et al.)[53]

a specially constructed capillary column mount capable of being micropositioned within the MHF sample chamber. Fluorescence intensity and intensity decay data were recovered from the dynamic MHF data. These data were analyzed using conventional nonlinear least squares or the MEM. The latter method allowed these authors to recover the intensity decay kinetics without a priori knowledge of the system decay function.

7 CHARACTERIZATION OF ORGANIZED MEDIA

Microheterogeniety has been observed in a variety of chemically and biologically organized systems, such as normal and reverse micelles, lipid vesicles, ion-exchange membranes, and cyclodextrins. Applications of time-resolved fluorescence spectroscopy toward the investigation of such systems are widespread.[57–70]

Huang and Bright examined the microenvironments within sodium dodecyl sulfate (SDS) micelles by using two fluorescent probes, 2-anilinonaphthalene-6-sulfonic acid (2,6-ANS) and *N*-phenylnaphthylamine (1-AN).[58] The excited state intensity decay kinetics were recovered by multifrequency phase and modulation measurements and analyzed using a global analysis protocol. The results showed that the fluorescence intensity decay kinetics of 2,6-ANS, which probes the outer core region (i.e. the palisade layer) of SDS micelles, was characterized by a unimodal Lorentzian lifetime distribution. In contrast, a single discrete excited state fluorescence lifetime was observed for 1-AN, which positioned itself in the micelle inner core region. The intensity decay kinetics of these probes were investigated as functions of temperature, counter-ion (Na^+ and Mg^{+2}) concentration, and linear alcohol cosolvents (*n*-butanol, *n*-pentanol, *n*-hexanol, and *n*-heptanol). Their collective results confirmed that the outer-core region of SDS micelles is microheterogeneous while the inner-core is essentially homogeneous. In addition, the intensity decay kinetics and the outer-core probe, 2,6-ANS, partitioning appeared to be more sensitive to variations in temperature and counter-ions in comparison to those of the inner core probe, 1-AN. The authors proposed that this observation was a result of different degrees of water penetration in the outer-core region. In the SDS system, the effects of micelle polydispersity and compositional diversity on the environmental microheterogeniety of the fluorescent probe seem to be minimal in comparison to water gradient effects.

Water immobilization at surfactant interfaces in reverse micelles has been investigated extensively.[59–64] For example, several research groups have examined the dynamical response of water in aerosol OT (AOT) reverse micellar systems.[59–64] Zhang and Bright investigated the reorganization of water within AOT reverse micelles on the nanosecond timescale using fluorescence spectroscopy of 1,8-anilino-8-naphthalenesulfonic acid (ANS).[60] For water loading $w_0 < 2.5$ ($w_0 =$ [water]/[AOT]), they observed two relaxation processes with time constants of 0.5–1.7 ns and 3.5–11.8 ns that they attributed to ''free'' and ''bound'' water reorganization,

respectively. For $w_0 > 2.5$, however, a single reorganization process with a time constant of 0.8–1.2 ns was reported. In general, they found that the observed rate constants decreased with increasing w_0. Cho et al.[61] have also probed the AOT interior using ~50 ps time-resolved fluorescence of ANS and suggested that the water molecules near the surfactant interface were ice-like owing to the strong interactions with the polar headgroup surface of high curvature and that the molecules appear bulk-like ~6–18 Å from the interface. Sarkar et al.[62,63] performed time-resolved fluorescence measurements of the probe coumarin 480 in micellar solutions and observed fluorescence decays with time constants ranging from 1.7–12 ns. They also attributed these dynamical changes to relaxation processes of water molecules in various water pool environments. Mittleman et al.[64] used ultrafast spectroscopy to probe the dielectric relaxation of water in AOT reverse micelles. These authors found that the timescale and relaxation amplitude are smaller than those of bulk water. This result was explained in terms of reduced long-range collective behavior resulting from the small water pool in the reverse micelles. Levinger et al. investigated the mobility of water in water/AOT/isooctane reverse micelles by using polar solvation dynamics measurements using ultrafast, time-resolved fluorescence up-conversion spectroscopy.[59] The major results of their work revealed that interactions of water with the surfactant headgroups in AOT reverse micelles effectively eliminates bulk-like solvent dynamics. This was most evident for very small reverse micelles, $w_0 < 5$. As the water pool volume increased in the AOT reverse micelles, more and more water molecules became free to move. This was manifested in a dynamic relaxation event on a timescale comparable to that of bulk water. However, although the solvation dynamics in the largest reverse micelles revealed a considerable fast relaxation component, the overall dynamics were still slower than that observed for bulk water. This suggests that solvent motion within a reverse micellar environment is affected by the surfactant interface.

Characterization of short-lived intermediates in microheterogeneous systems has been carried out by Whitten et al. using time-resolved spectroscopic techniques.[65] The intensity decay kinetics for a series of *trans*-stilbene-derivatized amphiphiles, which readily form bilayer systems in aqueous media, were analyzed by using a lifetime distribution analysis. The results indicated that the observed fluorescence originates from different types of excited state species that consist of two or more *trans*-stilbene units; one emission feature was attributed to the excited state of a ground state aggregate while the other was assigned to an excimer that may arise from a ''defect'' in the bilayer. The complex exponential nature of the decay kinetics was attributed to distributions of environments experienced by the fluorescing species.

Dimer formation of naphthalene in Langmuir–Blodgett (LB) monolayer films was probed by time-resolved fluorescence spectroscopy by Yamazaki et al.[66] Naphthalene incorporated in LB monolayer films exhibited two types of fluorescence emission due to monomer and dimer species. The dimer fluorescence corresponded well to that observed in the naphthalene crystal under extraordinarily high pressure and in naphthalene paracyclophanes. These researchers also reported on the excitation energy relaxation in LB films with rhodamine B, oxacyanine, and pyrene by means of picosecond time-resolved fluorescence spectroscopy. They found that the LB film is characterized by: (1) an inhomogeneous distribution of guest chromophores with fractal or fractal-like structures, (2) the formation of dimers and/or aggregates of dyes in molecular cages compressed within the fatty acid matrix, and (3) excimer formation from the preformed dimers. Their experimental results indicated that naphthalene chromophores were distributed within the LB monolayer films randomly as expected from statistical calculations. Further, naphthalene in LB films of low and medium concentrations (<10 mol%) was dispersed uniformly and randomly as monomer much more than for pyrene or ionic dyes like oxacyanine and rhodamine B, in which cases even low concentrations (1 mol%) resulted in dimer and/or excimer formation.

Cyclodextrins have drawn significant attention over the last few decades primarily because of their simplicity and the utility of these systems as models for protein–ligand and enzyme–substrate binding. In earlier work, Bright et al. presented nanosecond and picosecond time-resolved fluorometric studies of fluorophore-β-cyclodextrin inclusion complexes.[67] They found that the recovered intensity decay results for various anilinonaphthalenesulfonate ANS probes (2,6-ANS, 2,7-ANS, 2,8-ANS, and 1,8-ANS) with β-cyclodextrin are not well described by single or double exponential decay models. Their experimental data demonstrated that the fluorescence intensity decay kinetics of ANS-β-cyclodextrin inclusion complexes are modeled most accurately by Gaussian or Lorentzian lifetime distributions. In turn, this indicates that there is an ensemble of similarly formed ANS-β-cyclodextrin complexes in coexistence. This, the authors proposed, arose from rapid exchange between free and cyclodextrin-bound ANS complexes.

Wirth and Burbage used frequency domain fluorescence anisotropy decays to characterize the adsorbate reorientation and dynamics at a water/(octadecylsilyl)silica interface.[68] These authors showed that one can

independently study the fluorophore rotational reorientation in the surface plane can be studied independently as well as rotations out of the surface plane. Acridine orange, the fluorescent probe used in this work, exhibited hindered rotational reorientation for its out-of-plane rotation and it rotated slowly in the surface plane. The authors suggested that acridine orange resides at the interface of water and octadecylsilane and it senses the local morphological changes at this interface.

Montgomery and Wirth also investigated the effects of SDS surfactant concentration, above and below the SDS critical micelle concentration (CMC), on the rotational reorientation dynamics of a hydrophobic probe, 1,4-bis[*o*-methylstyryl]benzene, in a monolayer of covalently bonded dimethyloctadecylsiloxane chains.[69] The results showed that for any level of added SDS, the probe reorientation was less hindered than that measured for a pure water mobile phase. The net effect of SDS was to make the hydrophobic environment on the surface more fluid. However, near the SDS CMC, unusual behavior was observed and the probe rotational motion was even less hindered than it was above or below the CMC.

The adsorption of premicellar SDS onto methylated (C_1) silica surfaces was also studied to explore the possibility of monomer physisorption at an interface.[70] Methylation provided a hydrophobic surface precluding SDS chain interpenetration into the methyl monolayer. The premicellar adsorption maxima were found to be smaller for the C_1 surface than for the *n*-octadecylsiloxane (C_{18}) surface. This implicates interpenetration as an important factor in premicellar adsorption maxima. Surface roughness was also found to increase the premicellar adsorption maximum on the C_1 surface. The rotational reorientation behavior of acridine orange on these surfaces points to a structural transition in the adsorbed SDS layer as the origin of the large premicellar adsorption maximum of SDS on C_{18}.

8 FLUORESCENCE LIFETIME IMAGING MICROSCOPY

Fluorescence microscopy is an indispensable tool in the biosciences because of the excellent detection limits and the environmental responsiveness of selected probes and stains. However, steady-state microfluorimetry is inherently limited because it is more difficult to deconvolve the contributions from multiple ''emissive'' species, including autoluminescence from intrinsic cellular components, Raman scatter, and quantification problems exacerbated by difficulties in measuring quantum yield profiles across a substrate. In contrast, fluorescence lifetime imaging microscopy (FLIM) does not suffer from the inherent drawbacks of steady-state techniques and provides a means of acquiring information on the organization, structure, and dynamics of complex tissues, cells, and subcellular structures.[71]

In FLIM, the spatial (two- or three-dimensional) distribution of fluorescence decay times are determined directly at every pixel of an image simultaneously. The high temporal and spatial resolution is extremely important because it allows dynamic environmental parameters to be monitored in a spatially defined manner in the specimen under study (e.g. a single living cell). The intrinsic molecular information provided by dynamic measurements also minimizes background concerns and, as the excited state fluorescence lifetime is the only physical attribute observable, the resulting image is, in effect, simplified and filtered to include only the most analytically useful information. Judicious use of fluorescent reporter groups which exhibit excited state lifetime changes when challenged by specific intracellular factors allows one to quantitate such things as environmental membrane potential, temperature, dipolarity, and metabolic chemistries (H^+, K^+, Ca^{2+}, Zn^{2+}, O_2, glucose). The fluorophore/macromolecule mobility can also be assessed and information obtained at the nanometer level using resonance energy transfer techniques.

Since it was first demonstrated in 1990 by Wang et al.[72] that spatial variations in fluorescence lifetimes could be used as a source of image contrast, FLIM has received considerable attention as a powerful noninvasive method for obtaining spatially resolved information at the subcellular and even the molecular level. With state-of-the-art instrumentation, the total data acquisition time required to obtain a FLIM image with the diffraction-limited spatial resolution is typically ~30 s allowing the acquisition, processing, and display of fluorescence lifetime-resolved images in quasi-real time. Although this timescale is rapid enough for many applications, efforts are currently underway to complete this task in less time. Rapid, real time data acquisition would create new possibilities in areas including lifetime imaging endoscopy, large scale medical prescreening and diagnosis, pharmaceutical and medical microplate assays, combinatorial library evaluation, tracking biochemical kinetics, and tracing metabolic pathways.

Great advances in the ability to study dynamic events by FLIM were made by Schneider and Clegg.[73] These authors reported a new instrument capable of acquiring and displaying a time-resolved image in under 200 ms. Gratton et al. have also reported on a scanning fluorescence microscope using an asynchronous pump–probe approach which provides improved spatial

resolution compared to conventional one-photon fluorescence microscopy.[74] Lakowicz et al.[75] demonstrated the suitability of phase supression techniques to FLIM for minimizing the contribution from a particular decay time (i.e. species). For example, in the phase-resolved difference image, acrylamide-quenched Y_t-base emission was nulled leaving only a residual signal from Y_t-base present in inaccessible (unquenchable) environments. Lakowicz et al. were also the first to produce a fluorescence lifetime-based cellular chemical profile by tracking intracellular Ca^{2+} concentration in COS cell lines using the probe Quin-2.[76]

Although in its infancy, advances such as these in concert with rapid improvements in laser sources, electro-optic modulators, intensified CCD (charge coupled device) array detectors, computer processing algorithms, and bioreagents (monoclonal antibodies, nucleic acid probes, fluorescent labels) are sure to make FLIM a more accessible, popular, and routine technique in the analyst's arsenal.

9 FLUORESCENCE LIFETIME-BASED SENSING

Fluorescence-based sensing can rely on intensity, intensity ratio, or excited state intensity decay kinetics measurements. Although simple in concept and instrumentation, fluorescence intensity based sensing can be confounded by several factors. Frequent recalibration and other corrections are thus required because the fluorescence intensity depends on excitation intensity, light losses in optical fibers, photobleaching, and changes in the light scattering and/or sample absorption characteristics.[77–84] Wavelength-ratiometric fluorescent probes avoid some of these problems, however, lack of known and systematic methods to create such probes as well as limited commercial availability has hindered the widespread use of this approach. Because excited state fluorescence lifetime measurements are insensitive to probe concentration, photobleaching, washout and excitation instabilities, lifetime-based methods offer new opportunities for chemical sensing. Additionally, application of fluorescence lifetime-based sensing to remote monitoring with fiber optics renders recalibrations and corrections needed with intensity measurements unnecessary.[77–85]

Fluorescence lifetime-based sensing can be based on various molecular level physical or chemical processes and approaches such as collisional quenching, energy transfer, lifetime-based sensing using intrinsic versus conjugated probes, and so on.[78] For example, Draxler and Lippitsch have developed an optical pH sensor based on different fluorescence decay times of acid and base forms of the indicator diethylaminomethylpyrene.[79] The fact that the amino group is decoupled from the chromophoric system, excited state electron transfer probability from the amino group to pyrene influences the fluorescence decay times. This pyrene-based indicator was sequestered within a polyurethane-based hydrogel matrix providing the aqueous environment necessary for the acid–base reactions. The sensor has a favorable pK_a value (viable over the entire physiological pH range), an acceptable response time, and low cross-sensitivity against other cations. According to the researchers, reproducibility and long-term stability are the exceptional properties of this sensor.

Szmacinski and Lakowicz measured the pH-dependent fluorescence decay times of the seminaphthofluoresceins (SNAFL) and seminaphthorhodafluors (SNARF) using frequency domain fluorometry and an inexpensive 543 nm He–Ne laser light source.[80] They found that phase and modulation values were strongly dependent on pH in the physiological pH range as well as on the excitation and emission wavelengths. This dependence allows the range of pH sensitivity to be chosen by selection of wavelength(s) and enables increased precision of pH measurements by use of phase and/or modulation measurements at several wavelengths.

Recently, Chang et al.[81] developed a fluorescence lifetime-based sensor for water, by immobilizing a solvent-polarity-sensitive fluorescent metal–ligand compound, dipyridol[3,2-*a*:2′,3′-*c*]phenazine, di[*cis*-1,2-bis(diphenylphosphino)ethylene]osmium(II) hexafluorophosphate by ionic bonding onto a carboxymethyl cellulose cation-exchange resin, sandwiched between a thin sol–gel-processed layer and a glass substrate. When this solid film sensor is inserted from a water-free into a water-containing organic solvent, a decrease in fluorescence lifetime was observed due to fluorescence quenching. Because of the long decay time of this compound the frequency domain measurements could be performed using an inexpensive amplitude-modulated commercial blue LED. For example, the sensor exhibited a 39.6° phase angle decrease at a 2 MHz modulation frequency when placed from anhydrous acetone into a mixture of 20% (v/v) water in acetone. Similar observations were noticed in five other organic solvents upon addition of water. For this sensor, the response and recovery were both complete within seconds and the detection limit was solvent dependent. With ethyl acetate, the detection limit was as low as 0.02% (v/v) of water. This sensor also displayed excellent long-term stability with little drift in performance after a period of two months. Based on similar principles and instrumentation, these researchers designed a fluorescence lifetime-based sensor for methanol.[82]

The solvatochromic dye used in this case was tris(4,4′-dicarboxy-2,2′-bipyridine)ruthenium(II) hexafluorophosphate which was dissolved in a PMMA solution to form a thin film in a similar manner as in the case of the water sensor.

Birch et al. have demonstrated an optical method for the selective detection of Cu(II) ions in water using time-resolved fluorescence resonance energy transfer from the xanthene dye rhodamine 800 encapsulated in a Nafion-based sensor platform.[83] They showed that fluorescence quenching of the probe rhodamine 800 by other potentially interfering metal ions such as cobalt, nickel, and chromium were negligible in comparison to copper. They used a 670 nm picosecond diode laser as the excitation source and time-resolved single-photon counting for detection. They were able to determine Cu(II) between 5 and 50 mM. They estimated that the lower limit of detection for the sensor was 10 ppb.

Fluorescence lifetime-based biosensing of zinc was demonstrated by Thompson and Patchan by using frequency domain fluorescence and 5,1-dansylamide as the probe.[84] Briefly, the fluorescence lifetime of the bound dansylamide inhibitor increased approximately sevenfold and the emission was blue shifted 100 nm relative to the corresponding values for the free probe. The carbonic anhydrase/dansylamide indicator system permitted measurement of zinc concentration range over more than three decades.

Draxler and Lippitsch investigated the influence of the microenvironment on the fluorescence behavior of indicator molecules during lifetime-based sensing.[85] The two examples discussed were for a pH sensor using a pyrene-based fluorophore sequestered within a hydrogel and a ruthenium complex for oxygen sensing embedded in a polystyrene membrane. The interactions of a luminophore with a polymer matrix influence its luminescence decay properties. In many cases, deviations from a single-exponential decay could be observed, which may be due to spatial nonuniformity in the matrix refractive index, variations in the chromophore steric stabilization by the polymer environment, or nonresonant energy transfer from the luminophore to acceptor sites within the matrix. These authors also showed that this nonresonant energy transfer leads, at least in the case of a homogeneous distance distribution between luminophores and quenching sites, to a simple mathematical description of the luminescence decay which needs only one additional parameter. They proved the suitability of this approach in experimental assessment with these two different luminophores each in a different polymer. For the examples given, the proposed model gave a physically reasonable and more tractable procedure for evaluating luminescence decay kinetics when compared to multiple-exponential or rate-distribution models.

10 MULTIPHOTON EXCITED FLUORESCENCE

The high peak powers available from picosecond and femtosecond lasers can result in multiphoton excitation, wherein the fluorophore simultaneously absorbs two or more long wavelength photons to yield an excited singlet state. Multiphoton excitation requires high peak powers to increase the probability that two or more photons are simultaneously available within a given molecular volume for excitation. Because of the interactions of two or more photons with the fluorophore, the selection rules governing these transitions are different from those for one-photon excited fluorescence.[86] The interest in multiphoton excitation has increased dramatically as the result of its applications to time-resolved fluorescence and fluorescence microscopy. Several research groups have demonstrated time-resolved multiphoton excited fluorescence in the time and frequency domains as a tool for determining excited state anisotropy and intensity decay kinetics.[21,86–92] Two-photon induced time-resolved fluorescence has been observed for a wide variety of biochemical fluorophores, including membranes labeled with 1,6-diphenyl-3,4,5-hexatriene, labeled DNA, and proteins.[86,87] Lakowicz et al. have reported extensive investigation regarding multiphoton excited luminescence from biochemical fluorophores using the frequency domain technique.[86,87] The primary advantages associated with the frequency domain methodology are its high accuracy and the speed with which the phase angle and demodulation factor can be measured.[86,87]

Bright et al. reported on the performance of a new multiharmonic frequency domain instrument that uses the high harmonic content of a passively mode-locked pulse-picked femtosecond Ti-sapphire laser as the excitation source for the determination of one-, two-, or three-photon excited time-resolved fluorescence anisotropy and intensity decay kinetics.[21] The new instrument uses a parallel data acquisition strategy[19,20] which allows complete frequency domain data sets to be acquired in as little as 30 s. They demonstrated the potential of this new instrument by measuring the excited state fluorescence lifetimes and rotational reorientation times for fluorophores and fluorophore mixtures that exhibit nanosecond and subnanosecond single exponential and mutiexponential decay kinetics under one-, two-, or three-photon excitation. The instrument was able to recover the fluorescence decay kinetics accurately and offered a measurement precision on the order of 5–10 ps. The

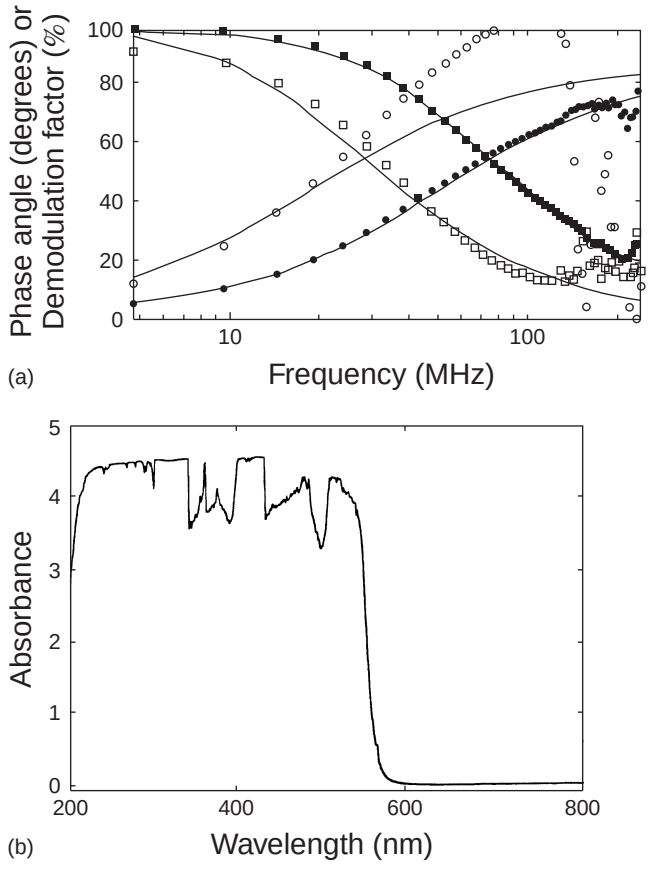

Figure 9 One- and two-photon excited dynamical measurements in an optically dense aqueous sample containing 10 μM aqueous R6G and excess bromocresol green. One- (open symbols) and two-photon excited (closed symbols) phase (circles)–modulation (squares) data sets acquired with the new instrument and fits to single-exponential decay models (a). Absorbance spectrum for the optically dense solutions (b). One-photon excitation at 450 nm. Two-photon excitation at 900 nm. (Reproduced with permission from Watkins et al.[21] Copyright (1998) American Chemical Society.)

instrument's capabilities were further tested by performing time-resolved intensity decay measurements on dilute fluorophore solutions in optically dense samples such as bromocresol green (shown in Figure 9) and undiluted whole blood under two-photon excitation conditions. The results of these experiments showed that the new instrument can recover the true fluorophore decay kinetics accurately in these optically dense samples. They also demonstrated the speed of the new instrument by recording the time-resolved anisotropy decay kinetics associated with fluorescein-labeled bovine serum albumin (BSA-FL) under two-photon excitation conditions on-the-fly as it was digested by the enzyme trypsin (Figure 10). Complete, 100 frequency data sets were acquired on-the-fly every 30 s. The results provided insights into the complex events that occur as BSA-FL is digested by trypsin. The time resolution of the new instrument is currently limited by the PMT detectors which set the practical upper modulation frequency limit to between 250–300 MHz.

Birch et al. have recently developed a time-correlated single-photon timing nonlinear fluorometer for recording the fluorescence decay times and rotational reorientation times of molecular probes using 120 fs regeneratively amplified Ti-sapphire laser excitation via simultaneous nonresonant absorption of two or more NIR photons.[88] A microchannel plate photomultiplier giving 70 ps impulse response is used for detection. Studies on 1,6-diphenylhexatriene, R6G, and *p*-terphenyl in propylene glycol demonstrated two- and three-photon induced fluorescence characteristics. Time-resolved detection of

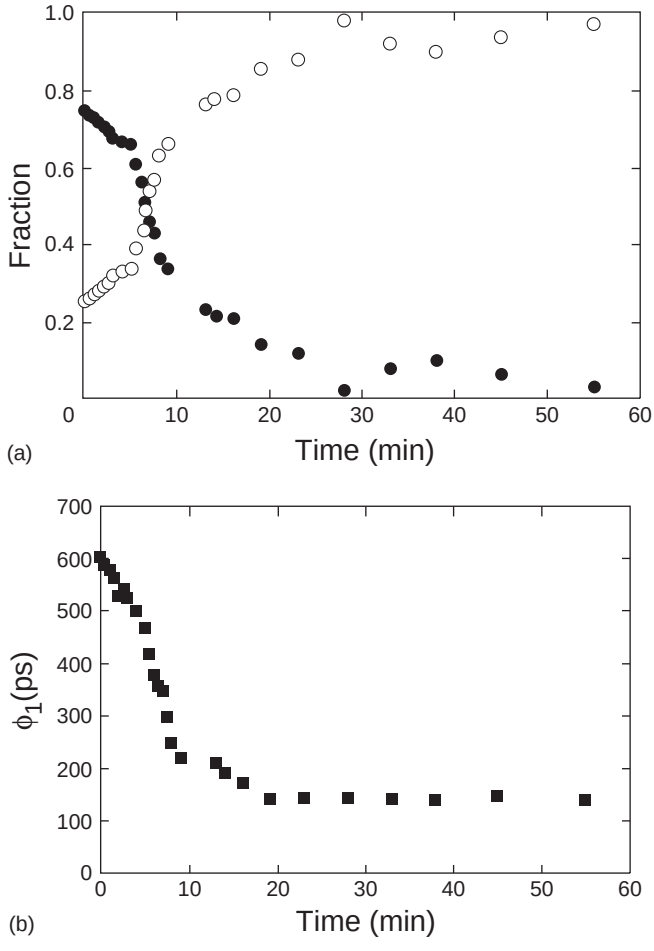

Figure 10 Summary of on-the-fly, two-photon excited fluorescence anisotropy decay kinetics of BSA-FL reacting with trypsin. Fractional contributions to rotational reorientation times 2 (●) and 1 (○) as a function of reaction time (a). Rotational reorientation times for ϕ_1 as a function of reaction time. ϕ_2 is independent of reaction time, 35 ± 0.5 ns (b). (Reproduced with permission from Watkins et al.[21] Copyright (1998) American Chemical Society.)

multiphoton induced anisotropic fluorescence using TCSPC techniques confirm that a higher degree of molecular orientation for two- and three-photon excitation is obtained in comparison to one-photon excitation. Multiphoton excitation, therefore, suggests itself as a more sensitive means of acquiring fluorescence anisotropy decays when probing the dynamics and structure of nanoscale microheterogeneous systems such as biomembranes, porous sol–gel glasses, and molecular matrices by observing molecular rotation of the fluorophore. Moreover, time-resolved multiphoton induced fluorescence studies using TCSPC offer potential in other applications, such as nonlinear fluorescence microscopy.

Saturated hydrocarbons, including linear, branched, and cyclic alkanes, are known to fluoresce when excited in the vacuum ultraviolet near 140–160 nm. These saturated hydrocarbons display emission from 200–240 nm, which could readily be transmitted by quartz optics. Vacuum ultraviolet excitation conditions are incompatible with biochemical and biophysical experiments because it is necessary to exclude oxygen which strongly absorbs below 250 nm. They also require the use of special optics which transmit the vacuum ultraviolet excitation. Lakowicz et al. have demonstrated that saturated alkanes can be excited with two-photon excitation allowing frequency domain fluorescence spectroscopic measurements.[86,89–91] Two-photon excitation of alkanes was accomplished by using the pulse train from a cavity-dumped frequency-doubled R6G dye laser at 298–300 nm, with the light focused on a $\sim$10 μm diameter spot. They were able to recover two-photon excited fluorescence intensity decay traces from saturated alkanes such as the cyclic alkanes cyclohexane and methylcyclohexane,[89] the linear alkane tetradecane,[90] and fully saturated cholestane.[91] In these studies, they also showed that the alkane emission was quenched by alcohols, water, and oxygen.[86,89–91] By using the frequency domain technique to recover the excited state fluorescence intensity decay kinetics, they were able to suggest that the quenching of tetradecane by *n*-propanol is most likely to be collisional with a Stern–Volmer quenching constant of $37\,M^{-1}$, and an unquenched lifetime of 4.4 ns to result in a bimolecular quenching constant of $8.4 \times 10^9\,M^{-1}\,s^{-1}$. In the absence of *n*-propanol, the intensity decay was found to be a single exponential. The presence of *n*-propanol in tetradecane, however, resulted in a shift in the frequency response to higher frequencies, which indicates a decrease in the excited state fluorescence intensity decay time of tetradecane. An indication of the dynamic nature of the quenching observed from time-resolved fluorescence data was the modest multi- or complex exponential decays in the presence of *n*-propanol.[90] Such effects occur due to transit effects in diffusional quenching. Also, the value of the bimolecular quenching constant is near that expected for 100% efficient collisional quenching in a solvent of tetradecane's viscosity (2.18 cP). Additional quenching studies, using the known solubility of oxygen in liquid alkanes, yielded a bimolecular quenching constant for tetradecane and O_2 of $4.8 \times 10^{10}\,M^{-1}\,s^{-1}$, which was slightly above the diffusion-controlled limit.[86,90] The researchers attribute this value of bimolecular quenching constant to the fact that there could be absorption by oxygen at 200 nm, which could in effect increase the interaction distance for quenching due to energy transfer. An important implication of the sensitivity of alkane fluorescence to dissolved O_2 is that alkane fluorescence can provide an intrinsic probe of O_2 transport in membranes. Choletane and tetradecane, as well as the cyclic alkanes, were also found to be collisionally quenched by water.

Recently, Lakowicz et al.[92] observed fluorescence emission from 2,5-diphenyl-1,3,4-oxadiazole (PPD) resulting from two-photon excitation with two different wavelengths near 380 and 760 nm. The two-color two-photon (2C2P) induced emission was observed when the PPD sample was illuminated with both wavelengths, but only when the picosecond laser pulses were spatially and temporally overlapped. For 2C2P excitation, the time zero anisotropy was larger than possible for single-photon excitation and was consistent with collinear electronic transitions for both wavelengths. The intensity depended on the polarization of each beam in a manner consistent with collinear transitions. Their results demonstrate that 2C2P excitation can be readily observed with modern laser sources. This phenomenon is suggested to have numerous applications in the chemical and biomedical sciences, as a method for spatial localization of the measured volume.

11 SUMMARY AND FUTURE PROSPECTS

In this article, we have highlighted the applications of excited state fluorescence anisotropy and intensity decay measurements in several areas that we view as being of interest to the analytical community. Continued technological advances in excitation sources and detection electronics coupled with the development of long-lived red-NIR fluorescent dye chemistries are crucial for the practical application of time-resolved fluorescence to further applications as well as in facilitating transfer of this powerful technique to routine analysis (e.g. environmental, toxicological, clinical, and chemical process monitoring). Eventually, such advances in time-resolved fluorescence may result in a new generation of commercially viable clinical and medical devices (e.g. sensors) for use in the clinical laboratory or during office visits, at the patient's bedside, and even in home health care.

By way of example, measurement of excited state fluorescence intensity decay kinetics of saturated linear and cyclic alkanes following two- or multiphoton excitation opens up new avenues to probing molecular systems. Alkyl groups can be regarded as intrinsic fluorophores present in all macromolecules and emission from alkyl groups can be expected whenever they are in a nonpolar environment shielded from contact with the dipolar media. For instance, such aliphatic side chains of detergents and lipids may serve as intrinsic fluorescent probes of micelles and bilayers.

Finally, it should be noted that this article is by no means comprehensive. Within subsections, references were chosen to demonstrate particular recent areas of application or to highlight novel achievements. As such, they are in an academic sense representative of the types of research that have been conducted in this field in the 1990s.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-Digital Converter
ANS	1,8-Anilino-8-naphthalenesulfonic acid
AOT	Aerosol OT
BME	β-Mercaptoethanol
BRE	Biorecognition Element
BSA-Ac	Acrylodan-labeled Bovine Serum Albumin
BSA-FL	Fluorescein-labeled Bovine Serum Albumin
BTBP	*N*,*N*′-bis-(2,5-Di-tert-butylphenyl)-3,4,9,10-perylenedicarboximide
CCD	Charge Coupled Device
CE	Capillary Electrophoresis
CMC	Critical Micelle Concentration
CW	Continuous Wave
FDB	Familial Defective Apolipoprotein B-100
FH	Familial Hypercholesterolemia
FLIM	Fluorescence Lifetime Imaging Microscopy
fwhm	Full Width At Half-maximum
HPLC	High-performance Liquid Chromatography
HPV	Human Papilloma Virus
HSA-Ac	Acrylodan-labeled Human Serum Albumin
LB	Langmuir–Blodgett
LED	Light Emitting Diode
MCA/PC	Multichannel Analyzer Personal Computer
MCPHA	Multichannel Pulse Height Analyzer
MEM	Maximum Entropy Method
MHF	Multiharmonic Fourier Transform
MLE	Maximum Likelihood Estimators
MPM/TIRF	Multifrequency Phase Modulation Total Internal Reflection Fluorescence
NIR	Near-infrared
NMR	Nuclear Magnetic Resonance
PAH	Polycyclic Aromatic Hydrocarbon
PCNMP	1-Phenyl-4-[(4-cyano-1-naphthyl)-methylene]piperidine
PCR	Polymerase Chain Reaction
PDT	Photodynamic Therapy
PMA	Poly(methyl acrylate)
PMMA	Poly(methyl methacrylate)
PMT	Photomultiplier Tube
PPD	2,5-Diphenyl-1,3,4-oxadiazole
PREWIF	Phase-resolved Evanescent Wave Induced Fluorescence
PS	Photosystem
R6G	Rhodamine 6G
SDS	Sodium Dodecyl Sulfate
SNAFL	Seminaphthofluoresceins
SNARF	Seminaphthorhodafluors
TAC	Time-to-Amplitude Convertor
TCSPC	Time Correlated Single Photon Counting
TIRF	Total Internal Reflection Fluorescence
TNS/PRODAN	Toluidinyl Naphthalene Sulfonic Acid/6-Propionyl-2-(*N*,*N*-dimethyl-amino)naphthalene
1-AN	*N*-phenylnaphthylamine
2,6-ANS	2-Anilinonaphthalene-6-sulfonic acid
2C2P	Two-color Two-photon

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Luminescence in Environmental Analysis

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
Fluorescence In Situ Hybridization

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Absorption and Luminescence Probes • Fluorescence Imaging Microscopy • Fluorescence in Organized Assemblies • Near-infrared Absorption/Luminescence Measurements • Phosphorescence Measurements, Applications of

Kinetic Determinations **(Volume 13)**
Luminescence-based Kinetic Determinations

REFERENCES

1. R.D. Fugate, 'Fluorescence Duration Gives New Dimension to Chemical Assays', *Res. Dev.*, April, 120–124 (1985).
2. J.B.A. Ross, K.W. Roussling, L. Brand, 'Time-resolved Fluorescence of the Two Trytophans in Horse Liver Alcohol Dehydrogenase', *Biochemistry*, **20**, 4369–4377 (1981).
3. I. Munro, I. Pecht, L. Stryer, 'Subnanosecond Motions of Tryptophan Residues in Proteins', *Proc. Natl. Acad. Sci. USA.*, **76**, 56–60 (1979).
4. J.R. Lakowicz, I. Gryczynski, H.C. Cheung, C. Wang, M.L. Johnson, 'Distance Distributions in Native and Random-coil Troponin I from Frequency-domain Measurements of Fluorescence Energy Transfer', *Biopolymers*, **27**, 821–830 (1988).
5. I. Gryczynski, M. Eftink, J.R. Lakowicz, 'Conformation Heterogeneity in Proteins as Origin of Heterogeneous Fluorescence Decays, Illustrated by Native and Denatured Ribonuclease T_1', *Biochim. Biophys. Acta*, **954**, 244–252 (1988).
6. J.R. Lakowicz, M.L. Johnson, I. Gryczynski, N. Joshi, G. Laczko, 'Transient Effects in Fluorescence Quenching Measured by 2 GHz Frequency-domain Fluorometry', *J. Phys. Chem.*, **91**, 3277–3285 (1987).
7. J.R. Lakowicz, N.B. Joshi, M.L. Johnson, H. Szmacinski, I. Gryczynski, 'Diffusion Coefficients of Quenchers in Proteins from Transient Effects in the Intensity Decays', *J. Biol. Chem.*, **262**, 10907–10910 (1987).
8. F.V. Bright, 'Modern Molecular Fluorescence Spectroscopy', *Appl. Spectrosc.*, **49**, 14A–19A (1995).
9. L.B. McGown, 'Fluorescence Lifetime Selectivity in Chemical Analysis', *Prog. Anal. Spectrosc.*, **11**, 383–415 (1988).
10. D.V. O'Connor, D. Phillips, *Time-Correlated Single Photon Counting*, Academic Press, London, 1984.
11. G. Nelson, G. Patonay, I.M. Warner, 'Multidimensional Time-correlated Single Photon Counting Fluorescence-Lifetime Measurements', *Adv. Multidimen. Luminesc.*, **1**, 73–95 (1991).
12. L.J. Cline-Love, L.A. Shaver, 'Time-Correlated Single Photon Technique: Fluorescence Lifetime Measurements', *Anal. Chem.*, **48**, 364A–371A (1976).
13. F.V. Bright, T.A. Betts, K.S. Litwiler, 'Advances in Multifrequency Phase and Modulation Fluorescence Analysis', *CRC Crit. Rev. Anal. Chem.*, **21**, 389–405 (1990).
14. J.R. Lakowicz, I. Gryczynski, 'Frequency-domain Fluorescence Spectroscopy', Topics in *Fluorescence Spectroscopy*, Volume 1: Techniques, ed. J.R. Lakowicz, Plenum Press, New York, Chapter 5, 1991.
15. D.M. Jameson, E. Gratton, R.D. Hall, 'The Measurement and Analysis of Heterogeneous Emission by Multifrequency Phase and Modulation Fluorometry', *Appl. Spectrosc. Rev.*, **20**, 55–106 (1984).
16. M.J. Wirth, 'Ultrafast Spectroscopy', *Anal. Chem.*, **62**, 270A–277A (1990).
17. W. Demtroder, *Laser Spectroscopy: Basic Concepts and Instrumentation*, Springer-Verlag, New York, 1996.
18. J. Sipior, G.M. Carter, J.R. Lakowicz, G. Rao, 'Blue Light-emitting Diode Demonstrated as an Ultraviolet Excitation Source for Nanosecond Phase-modulation Fluorescence Lifetime Measurements', *Rev. Sci. Instrum.*, **68**, 2666–2670 (1997).
19. B.A. Feddersen, D.W. Piston, E. Gratton, 'Digital Parallel Acquisition in Frequency Domain Fluorimetry', *Rev. Sci. Instrum.*, **60**, 2929–2936 (1989).
20. G.W. Mitchell, 'Picosecond Multiharmonic Fourier Fluorometer', US Patent 4,937, 457, 1990.
21. A.N. Watkins, C.M. Ingersoll, G.A. Baker, F.V. Bright, 'A Parallel Multiharmonic Frequency-domain Fluorometer for Measuring Excited-state Decay Kinetics Following One-, Two-, and Three-photon Excitation', *Anal. Chem.*, **70**, 3384–3396 (1998).
22. S. Ravichandran, B. Bagchi, 'Orientational Relaxation in Dipolar Systems: How Much do we Understand the Role of Correlations?', *Int. Rev. Phys. Chem.*, **14**, 271–314 (1995).
23. J.D. Simon, 'Time-resolved Studies of Solvation in Polar Media', *Acc. Chem. Res.*, **21**, 128–134 (1988).
24. V. Nagarajan, A.M. Brearley, T.-J. Kang, P.F. Barbara, 'Time-resolved Spectroscopic Measurements on Microscopic Solvation Dynamics', *J. Phys. Chem.*, **86**, 3183–3196 (1987).
25. D.W. Pierce, S.G. Boxer, 'Dielectric Relaxation in a Protein Matrix', *J. Phys. Chem.*, **96**, 5560–5566 (1992).

26. E.L. Mertz, V.A. Tikhomirov, L.I. Krishtalik, 'Stokes Shifts as a Tool for Probing the Solvent Reorganization Energy', *J. Phys. Chem. A*, **101**, 3433–3442 (1997).
27. F. Cichos, A. Willert, U. Rempel, C. von Borczyskowski, 'Solvation Dynamics in Mixtures of Polar and Nonpolar Solvents', *J. Phys. Chem. A*, **101**, 8179–8185 (1997).
28. E.R. Middelhoek, H. Zhang, J.W. Verhoeven, M. Glasbeek, 'Subpicosecond Studies of the Solvation Dynamics of Fluoroprobe in Liquid Solution', *Chem. Phys.*, **211**, 489–497 (1996).
29. M. Vincent, J. Gallay, A.P. Demchenko, 'Solvent Relaxation Around the Excited State of Indole: Analysis of Fluorescence Lifetime Distributions and Time-dependence Spectral Shifts', *J. Phys. Chem.*, **99**, 14931–14941 (1995).
30. H. Wendt, R. Richert, 'Purely Mechanical Solvation Dynamics in Supercooled Liquids: The $S_0 \leftarrow T_1(0\text{-}0)$ Transition of Naphthalene', *J. Phys. Chem. A*, **102**, 5775–5781 (1998).
31. G.R. Fleming, M. Cho, 'Chromophore–Solvent Dynamics', *Annu. Rev. Phys. Chem.*, **47**, 109–134 (1996).
32. J.D. Jordan, R.A. Dunbar, F.V. Bright, 'Dynamics of Acrylodan-labeled Bovine and Human Serum Albumin Entrapped in a Sol–Gel-derived Biogel', *Anal. Chem.*, **67**, 2436–2443 (1995).
33. H. Schneckenburger, H.K. Seidlitz, J. Eberz, 'New Trends in Photobiology: Time Resolved Fluorescence in Photobiology', *J. Photochem. Photobiol.*, **2**, 1–19 (1988).
34. S.B. Bambot, J.R. Lakowicz, G. Rao, 'Potential Applications of Lifetime-based, Phase-modulation Fluorimetry in Bioprocess and Clinical Monitoring', *Trends Biotechnol.*, **13**, 106–115 (1995).
35. J.C. de Paula, A. Liefshitz, S. Hinsley, W. Lin, V. Chopra, K. Long, S.A. Williams, S. Betts, C.F. Yocum, 'Structure–Function Relationships in the 47-kDa Antenna Protein and its Complex with the Photosystem II Reaction Center Core: Insights from Picosecond Fluorescence Decay Kinetics and Resonance Raman Spectroscopy', *Biochemistry*, **33**, 1455–1466 (1994).
36. M.M. Ressler, R.K. Pandey, 'Creating New Photosensitizers for Cancer Therapy', *CHEMTECH*, **28**, 39–45 (1998).
37. G. Eggertsen, M. Eriksson, O. Wiklund, A. Iitiä, S.-O. Olofsson, B. Angelin, L. Berglund, 'Time-resolved Fluorometry in the Genetic Diagnosis of Familial Defective Apolipoprotein B-100', *J. Lipid Res.*, **35**, 1505–1509 (1994).
38. M. Samiotaki, M. Kwiatkowski, N. Ylitalo, U. Landegren, 'Seven-color Time-resolved Fluorescence Hybridization Analysis of Human Papilloma Virus Types', *Anal. Biochem.*, **253**, 156–161 (1997).
39. C.M. Ingersoll, A.N. Watkins, G.A. Baker, F.V. Bright, 'Tracking Nanosecond and Subnanosecond Protein Dynamics On-the-fly Using Frequency-domain Fluorescence', *Appl. Spectrosc.*, **52**, 933–942 (1998).
40. C.M. Ingersoll, F.V. Bright, 'Using Fluorescence to Probe Biosensor Interfacial Dynamics', *Anal. Chem.*, **69**, 403A–408A (1997).
41. F.V. Bright, R. Wang, M. Li, R.A. Dunbar, 'Probing the Dynamics of Surface-immobilized Bioreceptors using Picosecond Time-resolved Spectroscopy', *Immunomethods*, **3**, 104–111 (1993).
42. J.S. Lundgren, E.J. Bekos, R. Wang, F.V. Bright, 'Phase-resolved Evanescent Wave Induced Fluorescence. An in situ Tool for Studying Heterogeneous Interfaces', *Anal. Chem.*, **66**, 2433–2440 (1994).
43. C.E. Hoyle, J.M. Torkelson, (eds.), *Photophysics of Polymers*, ACS Symposium Series 358, American Chemical Society, Washington, DC, 1987.
44. M.D. Ediger, 'Time-resolved Optical Studies of Local Polymer Dynamics', *Annu. Rev. Phys. Chem.*, **42**, 225–250 (1991).
45. J.W. Hofstraat, H.J. Verhey, J.W. Verhoeven, M.U. Kumke, G. Li, S.L. Hemmingsen, L.B. McGown, 'Fluorescence Lifetime Studies of Labelled Polystyrene Latices', *Polymer*, **38**, 2899–2906 (1997).
46. H. Sakai, A. Itaya, H. Masuhara, K. Sasaki, S. Kawata, 'Fluorescence Dynamics of Poly(*N*-vinylcarbazole) in Fluid Solution. Multivariate Analysis of Time-resolved Fluorescence Spectra', *Chem. Phys. Lett.*, **208**, 283–289 (1993).
47. I. Soutar, L. Swanson, R.L. Christensen, R.C. Drake, D. Phillips, 'Time-resolved Luminescence Anisotropy Studies of the Relaxation Behavior of Polymers. 1. Intramolecular Segmental Relaxation of Poly(methyl methacrylate) and Poly(methyl acrylate) in Dilute Solutions in Dichloromethane', *Macromolecules*, **29**, 4931–4936 (1996).
48. I. Soutar, C. Jones, D.M. Lucas, L. Swanson, 'Applications of Time-resolved Luminescence Anisotropy Measurements to the Study of Polymer Dynamics', *J. Photochem. Photobiol., A*, **102**, 87–92 (1996).
49. A.D. Stein, D.A. Hoffman, C.W. Frank, M.D. Fayer, 'Reorientational Motion of a Cross-link Junction in a Poly(dimethylsiloxane) Network Measured by Time-resolved Fluorescence Depolarization', *J. Chem. Phys.*, **96**, 3269–3278 (1992).
50. E.D. Niemeyer, F.V. Bright, 'Effects of CO_2 Sorption on the Rotational Reorientation Dynamics of a Model Solute Dissolved in Molten Poly(dimethylsiloxane)', *Macromolecules*, **31**, 77–85 (1998).
51. D.J. Desilets, P.T. Kissinger, F.E. Lytle, 'Measurement of Fluorescence Lifetimes During Liquid Chromatography', *Anal. Chem.*, **59**, 1830–1834 (1987).
52. W.T. Cobb, L.B. McGown, 'Multifrequency Phase-modulation Fluorescence Lifetime Determinations On-the-fly in HPLC', *Appl. Spectrosc.*, **43**, 1363–1367 (1989).
53. M.B. Smalley, J.M. Shaver, L.B. McGown, 'On-the-fly Fluorescence Lifetime Detection in HPLC Using a Mutiharmonic Fourier Transform Phase-modulation Spectrofluorometer', *Anal. Chem.*, **65**, 3466–3472 (1993).
54. M.B. Smalley, L.B. McGown, 'Limits of Detection and Resolution for On-the-fly Fluorescence Lifetime Detection in HPLC', *Anal. Chem.*, **67**, 1371–1376 (1995).

55. S.A. Soper, B.L. Legendre, Jr, D.C. Williams, 'On-line Fluorescence Lifetime Determinations in Capillary Electrophoresis', *Anal. Chem.*, **67**, 4358–4365 (1995).
56. L.-C. Li, L.B. McGown, 'On-the-fly Frequency-domain Fluorescence Lifetime Detection in Capillary Electrophoresis', *Anal. Chem.*, **68**, 2737–2743 (1996).
57. M.H. Gehlen, F.C. DeSchryver, 'Time-resolved Fluorescence Quenching in Micellar Assemblies', *Chem. Rev.*, **93**, 199–221 (1993).
58. J. Huang, F.V. Bright, 'Microheterogeneity of Sodium Dodecyl Sulfate Micelles Probed by Frequency-domain Fluorometry', *Appl. Spectrosc.*, **46**, 329–339 (1992).
59. R.E. Riter, D.M. Willard, N.E. Levinger, 'Water Immobilization at Surfactant Interfaces in Reverse Micelles', *J. Phys. Chem. B*, **102**, 2705–2714 (1998).
60. J. Zhang, F.V. Bright, 'Nanoscale Reorganization of Water Within the Interior of Reveresed Micelles Revealed by Frequency-domain Fluorescence Spectroscopy', *J. Phys. Chem.*, **95**, 7900–7907 (1991).
61. C.H. Cho, M. Chung, J. Lee, T. Nguyen, S. Singh, M. Vedamuthu, S.H. Yao, J.B. Zhu, G.W. Robinson, 'Time- and Space-resolved Studies of the Physics and Chemistry of Liquid Water Near a Biologically Relevant Interface', *J. Phys. Chem.*, **99**, 7806–7812 (1995).
62. N. Sarkar, K. Das, A. Dutta, S. Das, K. Bhattacharyya, 'Solvation Dynamics of Coumarin 480 in Reverse Micelles. Slow Relaxation of Water Molecules', *J. Phys. Chem.*, **100**, 10523–10527 (1996).
63. N. Sarkar, A. Dutta, S. Das, K. Bhattacharyya, 'Solvation Dynamics of Coumarin 480 in Micelles', *J. Phys. Chem.*, **100**, 15483–15486 (1996).
64. D.M. Mittleman, M.C. Nuss, V.L. Colvin, 'Terahertz Spectroscopy of Water in Inverse Micelles', *Chem. Phys. Lett.*, **275**, 332–338 (1997).
65. D.G. Whitten, M.S. Farahat, E.R. Gaillard, 'Symposium in Print: Time-resolved Fluorescence and Transient Spectroscopy in Determining Photochemical and Photophysical Channels in Reacting Systems in Solutions and Microheterogeneous Media', *Photochem. Photobiol.*, **65**, 23–32 (1997).
66. K. Abe, S. Suzuki, I. Mukai, S. Akimoto, N. Ohta, I. Yamazaki, 'Dimer Formation of Naphthalene in Langmuir–Blodgett Monolayer Films Probed by Time-resolved Fluorescence Spectroscopy', *Chem. Lett.*, 479–480 (1996).
67. F.V. Bright, G.C. Catena, J. Huang, 'Evidence for Lifetime Distributions in Cyclodextrin Inclusion Complexes', *J. Am. Chem. Soc.*, **112**, 1343–1346 (1990).
68. M.J. Wirth, J.D. Burbage, 'Adsorbate Reorientation at a Water/(Octadecylsilyl)silica Interface', *Anal. Chem.*, **63**, 1311–1317 (1991).
69. M.E. Montgomery, Jr, M.J. Wirth, 'Effect of Sodium Dodecyl Sulfate on the Orientational Behavior of a Hydrophobic Probe in a C_{18} Monolayer Bonded to Silica', *Anal. Chem.*, **64**, 2566–2569 (1992).
70. M.J. Wirth, D.A. Piasecki-Coleman, M.E. Montgomery, Jr, 'Spectroscopic Investigation of Sodium Dodecyl Sulfate Adsorbed to a Trimethylsiloxane Monolayer on Silica', *Langmuir*, **11**, 990–995 (1995).
71. X.F. Wang, A. Periasamy, B. Herman, 'Fluorescence Lifetime Imaging Microscopy (FLIM): Instrumentation and Applications', *CRC Crit. Rev. Anal. Chem.*, **23**, 369–395 (1992).
72. X.F. Wang, S. Kitajima, T. Uchida, D.M. Coleman, S. Minami, 'Time-resolved Fluorescence Microscopy Using Multichannel Photon Counting', *Appl. Spectrosc.*, **44**, 25–30 (1990).
73. P.C. Schneider, R.M. Clegg, 'Rapid Acquisition, Analysis, and Display of Fluorescence Lifetime-resolved Images for Real-time Applications', *Rev. Sci. Instrum.*, **68**, 4107-4119 (1997).
74. C.Y. Dong, P.T.C. So, T. French, E. Gratton, 'Fluorescence Lifetime Imaging by Asynchronous Pump–Probe Microscopy', *Biophys. J.*, **69**, 2234–2242 (1995).
75. J.R. Lakowicz, H. Szmacinski, K. Nowaczyk, K.W. Berndt, M.L. Johnson, 'Fluorescence Lifetime Imaging', *Anal. Biochem.*, **202**, 316–330 (1992).
76. J.R. Lakowicz, H. Szmacinski, K. Nowaczyk, W.J. Lederer, M.S. Kirby, M.L. Johnson, 'Fluorescence Lifetime Imaging of Intracellular Calcium in COS Cells using Quin-2', *Cell Calcium*, **15**, 7–27 (1994).
77. F.V. Bright, 'A New Fiber-optic-based Multifrequency Phase-modulation Fluorometer', *Appl. Spectrosc.*, **42**, 1531–1537 (1988).
78. H. Szmacinski, J.R. Lakowicz, 'Lifetime-based Sensing', Topics in *Fluorescence Spectroscopy*, Volume 4: Probe Design and Chemical Sensing, ed. J.R. Lakowicz, Plenum Press, New York, Chapter 10, 1994.
79. S. Draxler, M.E. Lippitsch, 'pH Sensors Using Fluorescence Decay Time', *Sensors and Actuators B*, **29**, 199–203 (1995).
80. H. Szmacinski, J.R. Lakowicz, 'Optical Measurements of pH Using Fluorescence Lifetimes and Phase-modulation Fluorometry', *Anal. Chem.*, **65**, 1668–1674 (1993).
81. Q. Chang, Z. Murtaza, J.R. Lakowicz, G. Rao, 'A Fluorescence Lifetime-based Solid Sensor for Water', *Anal. Chim. Acta*, **350**, 97–104 (1997).
82. Q. Chang, J.R. Lakowicz, G. Rao, 'Fluorescence Lifetime-based Sensing of Methanol', *Analyst*, **122**, 173–177 (1997).
83. D.J.S. Birch, O.J. Rolinski, D. Hatrick, 'Fluorescence Lifetime Sensor of Copper Ions in Water', *Rev. Sci. Instrum.*, **67**, 2732–2737 (1996).
84. R.B. Thompson, M.W. Patchan, 'Fluorescence Lifetime-based Biosensing of Zinc: Origin of the Broad Dynamic Range', *J. Fluoresc.*, **5**, 123–130 (1995).
85. S. Draxler, M.E. Lippitsch, 'Lifetime-based Sensing: Influence of the Microenvironment', *Anal. Chem.*, **68**, 753–757 (1996).

86. J.R. Lakowicz, I. Gryczynski, 'Multiphoton Excitation of Biochemical Fluorophores', Topics in *Fluorescence Spectroscopy*, Volume 5: Nonlinear and Two- Photon-Induced Fluorescence, ed. J.R. Lakowicz, Plenum Press, New York, Chapter 3, 1997.
87. B. Kierdaszuk, I. Gryczynski, J.R. Lakowicz, 'Two-photon Induced Fluorescence of Proteins', Topics in *Fluorescence Spectroscopy*, Volume 5: Nonlinear and Two-photon-induced Fluorescence, ed. J.R. Lakowicz, Plenum Press, New York, Chapter 5, 1997.
88. A. Volkmer, D.A. Hatrick, D.J.S. Birch, 'Time-resolved Nonlinear Fluorescence Spectroscopy Using Femtosecond Multiphoton Excitation and Single-Photon Timing Detection', *Meas. Sci. Technol.*, 1339–1349 (1997).
89. J.R. Lakowicz, I. Gryczynski, 'Fluorescence Intensity Decays of Cyclohexane and Methylcyclohexane with Two-photon Excitation from a High Repetition Rate Frequency-doubled Dye Laser', *Biospectroscopy*, **1**, 3–8 (1995).
90. I. Gryczynski, A. Razynska, J.R. Lakowicz, 'Two-photon Induced Fluorescence of Linear Alkanes; a Possible Intrinsic Lipid Probe', *Biophys. Chem.*, **57**, 291–295 (1996).
91. R.A. Agbaria, I. Gryczynski, H. Malak, J.R. Lakowicz, 'Two-photon-induced Fluorescence of Cholestane', *Biospectroscopy*, **2**, 219–224 (1996).
92. J.R. Lakowicz, I. Gryczynski, H. Malak, Z. Gryczynski, 'Fluorescence Spectral Properties of 2,5-Diphenyl-1,3,4-Oxadiazole with Two-color Two-photon Excitation', *J. Phys. Chem.*, **100**, 19406–19411 (1996).

Indirect Detection Methods in Capillary Electrophoresis

Shahab A. Shamsi
Georgia State University, Atlanta, USA

1 Introduction 2
1.1 Indirect Photometric Detection in Capillary Zone Electrophoresis 2
1.2 Theory and Operating Principle 2
1.3 Detector Response Equation 3
1.4 The Transfer Ratio and Kohlrausch's Regulating Function 3

2 Factors Influencing Indirect Photometric Detection Sensitivities 4
2.1 Molar Absorptivity of the Indirect Photometric Detection Reagent 4
2.2 Mobility of the Indirect Photometric Detection Reagent 5
2.3 Concentration of the Indirect Photometric Detection Reagent 7
2.4 Type and Concentration of the Nonabsorbing Buffer 7
2.5 pH of the Indirect Photometric Detection Reagent 8
2.6 Effect of Detection Wavelength 11
2.7 Effect of Background Noise 12
2.8 Effect of Absorption Pathlength 12

3 Determination of Anionic and Cationic Compounds 12
3.1 Indirect Photometric Detection of Inorganic Anions 12
3.2 Indirect Photometric Detection of Inorganic Cations 15
3.3 Indirect Photometric Detection of Short-chain Organic Acids and Amines 19
3.4 Indirect Photometric Detection of Long-chain Fatty Acids 22
3.5 Indirect Photometric Detection of Phosphates and Phosphonates 22
3.6 Indirect Photometric Detection of Aliphatic Anionic and Cationic Surfactants 28
3.7 Indirect Photometric Detection of Carbohydrates 28
3.8 Indirect Photometric Detection of Amino Acids and Phospholipids 31

4 Simultaneous Separation and Indirect Detection of Cations and Anions 33
4.1 Selection of Suitable Cationic and Anionic Indirect Photometric Detection Reagents 33
4.2 Selection of a Suitable pH 34
4.3 Effect of Indirect Photometric Detection Binary Electrolyte Concentration 34
4.4 Simultaneous Separation of 22 Anions and Cations in a Single Run 35
4.5 Selected Applications 36

5 Indirect Fluorescence Detection in Capillary Electrophoresis 36
5.1 Factors Influencing Indirect Fluorescence Detection Response 37
5.2 Applications of Indirect Fluorescence Detection for the Analysis of Ionic Compounds 39
5.3 Applications of Indirect Fluorescence Detection for the Analysis of Neutral Compounds 42

6 Other Indirect Detection Modes 43
6.1 Indirect Chemiluminescence Detection 43
6.2 Indirect Electrochemical Detection 44

Acknowledgments 46
Abbreviations and Acronyms 46
Related Articles 47
References 47

An overview is given of existing indirect detection methods of inorganic and organic anions in capillary zone electrophoresis (CZE). Kohlrausch's regulating function (KRF), used by several researchers to predict the detection sensitivity of ultraviolet (UV)-inactive analytes in CZE, is reviewed. Sensitive indirect photometric detection (IPD) of ionic compounds in CZE is largely governed by the value of the change in absorbance in a solute zone. This in turn depends on a transfer ratio (TR), ratio of the charge, and ratio of the electrophoretic mobility of the analyte and IPD reagent. The influence of other experimental parameters, such as pH, type and concentration of the absorbing electrolyte, detection wavelength, and absorption pathlength, on detection sensitivity are also critically discussed. An extensive part of this review deals with the separation of various classes of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

ionic analytes using capillary zone electrophoresis/indirect photometric detection (CZE/IPD). Different approaches for simultaneous separation and indirect detection of nonabsorbing anions and cations are evaluated. From this evaluation, some rules of thumb as to when these approaches will be most effective are drawn. In addition, indirect fluorescence detection (IFD) in CZE and micellar electrokinetic chromatography (MEKC) are reviewed and possible application areas are highlighted. Some recent developments in other indirect detection modes, such as indirect chemiluminescence detection (ICLD) and indirect conductivity detection (ICD), are outlined.

1 INTRODUCTION

Direct detection employed in capillary electrophoresis (CE) refers to a situation where the eluent or electrolyte has a lower value of the measured physical property (i.e. absorbance, fluorescence, chemiluminescence, conductivity, amperometry), whereas indirect detection refers to a reverse condition. These two situations are shown schematically in Figure 1 for photometric detection. Indirect detection methods have been used in both ion chromatography (IC) and CE for mostly ionic species whose structure lacks the measured physical property for direct detection. These detection methods are universal in applicability and can be used with all detectors that are normally used to measure the analyte signal in a direct detection mode. Thus, indirect detection is applied to a wide variety of compounds without the need for pre- and postcolumn derivatization.

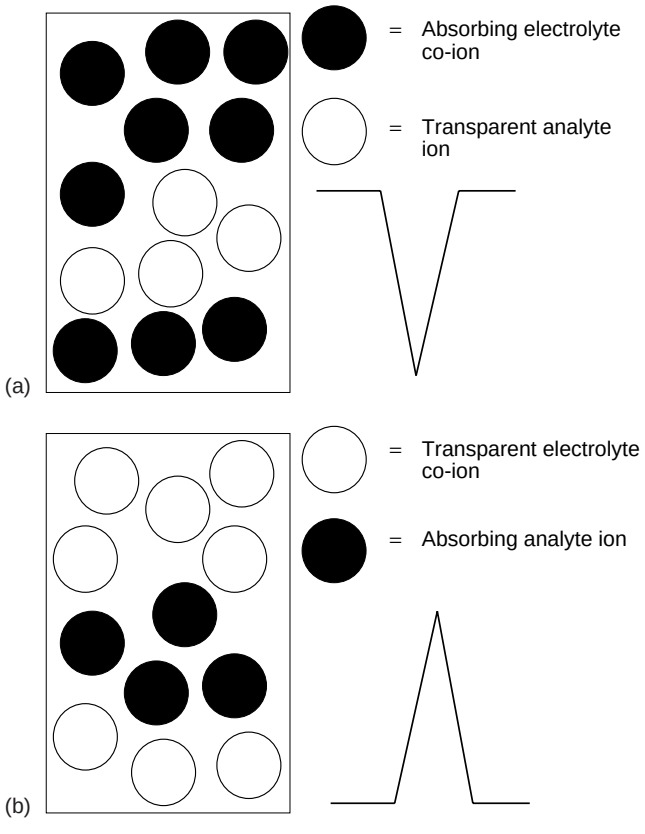

Figure 1 Comparsion of IPD (a) and direct photometric detection (b) in CE.

The first report dealing with the use of indirect detection in CZE for the analysis of ionic species was demonstrated more than 10 years ago by Hjerten et al.[1] This was followed by a report by Foret et al.[2] in 1989. Currently, it is an active area of research in CE for a variety of ionic compounds. Accordingly, this article provides a critical review dealing with the IPD and IFD methods of ionic species separated by CZE. All publications that have appeared in abstracted journals up until April 1999 are summarized. In addition, IFD of neutral species after their separation by MEKC is discussed. Only three communications on indirect chemiluminescence and three on indirect electrochemical (amperometry, conductivity) detection have been published, to my knowledge, since the beginning of 1990; therefore these indirect detection modes are discussed briefly.

1.1 Indirect Photometric Detection in Capillary Zone Electrophoresis

The real interest in the use of CZE with IPD, especially for inorganic ion analysis, started with the work of Jones et al. in the early 1990s.[3–8] At present, this is the most popular indirect detection mode in CZE analysis of a variety of inorganic and organic ions, and is discussed comprehensively in recent review articles.[9–13]

1.2 Theory and Operating Principle

In CZE, indirect absorbance detection (also known as IPD) has been used for ionic species whose structure lacks the absorbance properties for direct detection. The key element for IPD in CE is to maintain a large continuous background absorbance signal at the UV detector by employing a chromophoric ionic species (an ion with the same charge as the analyte) in the carrier electrolyte. If the concentration of this light-absorbing species remains constant in the electrical double layer in the CE column, a steady level of background absorbance translated as a stable baseline is displayed on the electropheretogram. When a non-UV-detectable ionic species passes the detection window, charge displacement of the chromophoric ion leads to a quantifiable decrease in the background absorbance. This translates to a negative peak; however, leads on the data acquisition system can be

switched to record signals as a positive peak. Similarly to IC, sensitive IPD of ionic species is possible in CZE mode because of the charge displacement that occurs in the electrical double layer of the bare fused-silica capillary. However, in CZE this charge displacement is enforced not by the ion-exchange groups as in IC, but by a fixed and oppositely migrating flux of counterions and by the requirement of electroneutrality.

1.3 Detector Response Equation

The concentration limit of detection (C_{LOD}), in moles per liter, for a nonabsorbing analyte detected using IPD can be expressed as a function of the concentration of the UV-absorbing species (C_M), the TR, and the dynamic reserve (DR), according to Equation (1):[14,15]

$$C_{LOD} = \frac{C_M}{DR \bullet TR} = \frac{N_{BL}}{TR \bullet \varepsilon l} \quad (1)$$

where N_{BL} is the baseline noise, ε is the molar absorptivity of the IPD reagent, and l is the pathlength of the detection cell. The DR is defined as the ability to measure a small change on the top of a large signal, and is equal to the ratio of the background absorbance to noise. The TR is defined as the number of moles of the IPD reagent displaced by one mole of the analyte. For example, TR values of 1 and 2 are possible when respective ions Cl^- and SO_4^{2-} are displaced by naphthalenemonosulfonate (NMS). Briefly, the TR predicts that, everything else being the same, the molar sensitivity of divalent ions is going to be twice that for monovalent ions. However, in reality a mobility match in CZE can offset such effects. Equation (1) also predicts that C_{LOD} can be improved by reducing C_M or by increasing the DR. Because the DR is also related to C_M, decreasing the concentration of the absorbing species (IPD reagents) will not necessarily improve the limit of detection (LOD). This is because the DR is also decreased, and electrophoretic dispersion in CZE can result in a poor LOD for nonabsorbing analytes. Hence, improving C_{LOD} often requires maximizing the DR by either reducing the background noise or increasing the background absorbance of the IPD reagent. Nevertheless, it can be concluded that all three parameters (C_M, DR and TR) of Equation (1) are not necessarily independent and an operator needs to optimize them to achieve the best possible detection limits. It is worth mentioning that C_{LOD} values for IPD of around 10^{-4}–10^{-5} M are routinely achieved in CZE.[16] The best C_{LOD} in IPD is obtained using a double-beam laser-induced diode detection, for which the C_{LOD} is about 10^{-7} M.[17] These systems typically reduce the noise level by an order of magnitude.

1.4 The Transfer Ratio and Kohlrausch's Regulating Function

As discussed earlier, the extent of displacement of the light-absorbing co-ion by the analyte is known as the TR. According to Equation (1), C_{LOD} is inversely proportional to the TR. Thus, a higher value of TR results in a large analyte peak area and a decrease in the C_{LOD} term. In addition, Equation (1) predicts that a TR between a singly charged analyte and the IPD co-ion would be unity. This means that peak areas for analyte ions of the same charge and concentration should be similar, which enables one to perform a universal calibration. However, Ackerman et al. called attention to the fact that a nonlinear relationship existed between peak areas and effective mobilities.[18] These researchers explained their results by considering KRF, also abbreviated as "w". This function w is generally written according to Equation (2):

$$w = \frac{C_i Z_i}{\mu_i} = \text{Constant} \quad (2)$$

where Z_i represents the absolute charges, μ_i the absolute values of effective mobility, and C_i the ionic concentrations of all ionic species. In addition, note that negative ions have negative μ and z values, whereas positive ions have positive μ and z values.

After the sample ion B is injected and the voltage is applied, a volume element of the capillary, which is originally filled with IPD reagent AC (A = UV-absorbing co-ion; C = non-UV-absorbing counterion) at a concentration C_A, will consist of a mixture of both co-ion A and the solute ion B for a while. However, the original situation will be restored very soon. Ackerman et al.[18] gave the expression shown in Equation (3) for the migrating sample ions in this case:

$$C_A^C = C_A^S + C_B^S k_B \quad (3)$$

The superscripts C and S in Equation (3) refer to the concentrations of the IPD electrolyte zone and the solute zone respectively. According to Nielen,[19] Cousins et al.,[20] Buchberger et al.,[12] and Klampfl et al.,[21] Equation (2) can be verified from Kohlrausch theory according to Equation (4):

$$k_B = TR = \frac{Z_B \mu_A (\mu_B + \mu_C)}{Z_A \mu_B (\mu_A + \mu_C)} \quad (4)$$

where k_B is referred to as the TR (also called the response factor); Z_A and Z_B represent absolute charges on analyte and the IPD co-ion A, respectively; μ_A, μ_B, and μ_C refer to the effective electrophoretic mobility of the IPD co-ion (i.e. ions of the same charge as that of the analyte) A, analyte ion B, and the nonabsorbing counterion (i.e.

ions of opposite charge to that of the analyte ion) C. The absorbances of the IPD ion (A^C) and the sample ion (A^S) can be written in accordance with Beer's law, as shown by Equations (5) and (6), respectively:

$$A^C = \varepsilon_A C_A^C l \quad (5)$$

$$A^S = \varepsilon_A C_A^S l + \varepsilon_B C_B^S l \quad (6)$$

From Equation (3), the UV signal generated as a change in background absorbance (ΔA) can be shown according to Equation (7):

$$\Delta A = A^C - A^S = C_B^S l(\varepsilon_A k_B - \varepsilon_B) \quad (7)$$

For non-UV-absorbing sample ions, $\varepsilon_B = 0$. Substituting for the TR from Equation (4), Equation (8) was derived:[12]

$$\Delta A = C_B^S l \varepsilon_A \frac{Z_B \mu_A(\mu_B + \mu_C)}{Z_A \mu_B(\mu_A + \mu_C)} \quad (8)$$

An important characteristic of Equation (8) indicates that, in an IPD system, the value of ΔA in a solute zone is proportional to the value of the TR (also known as the response factor, k_B), the ratio of the charge of the solute and the light-absorbing electrolyte ion, as well as the concentration of the solute ion. Hence, the values of ΔA (i.e. the detection sensitivities) for solutes with identical charges and concentrations can be determined by the values of the response factors. The value of ΔA is negative (i.e. analytes detected as negative peaks) if the running buffer with co-ion A is being used as a light-absorbing ion. In contrast, ΔA can be either positive or negative depending on the relative mobility of a solute to the nonabsorbing counterion in a system; and absorbing co-ion as the IPD reagent. Additionally, Equation (8) also indicates that when the mobility of the counterion (μ_C) is higher than the mobility of the solute (μ_B), ΔA will acquire a positive value; and a negative value when μ_C is lower than μ_B. Furthermore, when a solute ion has the same mobility as the co-ion (which is also a light-absorbing ion), equivalent-per-equivalent exchange can be expected. This charge displacement process improves the peak shape and increases the TR, which in turn provides a better LOD.

A literature survey indicates that several researchers have tested the validity of Equation (8). Nielen[19] first applied this equation to calculate the response factors of alkylsulfates using veronal as the IPD co-ion. He found the response factors of alkyl sulfates to be consistent with the trend predicted in Equation (8). Cousin et al.[20] recently determined the TR values for several inorganic anions using chromate, pyromellitate, trimellitate, phthalate and benzoate as IPD reagents. Comparison of TR values for these five reagents indicated that chromate is the best IPD reagent for high-mobility inorganic anions. However, the fit of the experimental data to the calculated data using Equation (8) was poor. Very recently, Doble et al.[22] observed that the agreement between experimental and theoretical values of TR can be improved if the IPD reagent is free of any interfering co-ions. For example, reformulating the electroosmotic flow modifier (i.e. addition of a tetradecyl-ammonium hydroxide instead of tetradecylammonium bromide) to the IPD electrolyte is beneficial because the hydroxide co-anions are utilized by the buffer.

2 FACTORS INFLUENCING INDIRECT PHOTOMETRIC DETECTION SENSITIVITIES

Theoretical considerations suggest that optimal sensitivity in IPD can be achieved if several conditions are fulfilled. First, the light-absorbing reagent should have a large molar absorptivity and an effective mobility that closely match those of the analyte ions. Second, the concentration of the IPD reagent and type and concentration of the nonabsorbing background electrolyte should be carefully selected. This is because high sensitivity is obtained only if the visualization co-ion is displaced by the analyte ion. Third, the pH of the light-absorbing electrolyte should be such that it minimizes the contribution of non-UV-absorbing H^+ and OH^- ions during the charge displacement process. For example, at very high pH (pH > 12.1), which is required for the ionization of weakly acidic compounds (e.g. sugars), the signal-to-noise ratio (S/N) in IPD drops sharply.[23] Fourth, choice of appropriate detection wavelength, low background noise and a long absorption pathlength are also desirable for sensitive detection.

2.1 Molar Absorptivity of the Indirect Photometric Detection Reagent

According to Equation (1), increasing the molar absorptivity ε of the IPD reagent increases the DR. The advantages of a high-ε reagent become apparent as one tries to decrease the C_M, resulting in improved detection limits. On the other hand, to obtain full benefit of a high-ε reagent care must be taken to match its mobility to the sample ion. A series of publications dealing with IPD in CZE has considered a high-ε reagent for optimum sensitivity.[2,5,20,24–31] For example, using sorbate (a high-ε reagent) instead of a benzoate (a low-ε reagent), Foret et al.[2] achieved a significantly (ca. 50 times) lower LOD. Shamsi and Danielson[24] investigated a series of

ribonucleotide reagents for the separation of polyphosphates (P_n) and polyphosphonates (RP_n). Best detection limits for these phosphorus-containing anions were obtained with adenosine 5′-monophosphate (AMP). This reagent provided the highest ε and closest mobility match for P_n and RP_n. Similar results were obtained by Beck and Engelhardt[25] and Weston et al.[26] when comparing a series of cationic IPD reagents for the separation and IPD of inorganic and organic cations. To extend the versatility of IPD reagents in the visible region, Mala et al.[27] investigated methyl green and chlorophenol red as IPD electrolytes for analysis of cations and anions, respectively. These dyes possess ε values an order of magnitude higher than conventional light-absorbing reagents such as chromate or phthalate. However, the reported LOD was not significantly better than those obtained with conventional reagents. Probably the presence of nonabsorbing co-ions competed with absorbing co-ions in the charge displacement, resulting in a less pronounced effect on the sensitivity of IPD.

2.2 Mobility of the Indirect Photometric Detection Reagent

As the sensitivity in IPD is governed by the ε of the carrier electrolyte, so too are its charge and electrophoretic mobility. Mikkers et al.[32,33] were the first to discuss the rules governing the effect of electrophoretic migration on the analyte zone concentration distribution using a nondiffusional model derived from KRF. According to this rule, symmetrical peaks are obtained only when the electrolyte co-ions have a mobility similar to that of the analytes. Electrolyte ions that have a slower mobility than the analytes will exhibit fronting peaks. In contrast, electrolyte co-ions with faster mobility than the analytes result in tailing peaks. Examples of the aforementioned rule were illustrated by Jones and Jandik[34] using three IPD electrolytes: chromate (high mobility), phthalate (intermediate mobility), and *p*-hydroxybenzoate (low mobility). Figure 2 shows the peak shapes of anions using these three IPD electrolytes. Note that the mobility for each IPD electrolyte is indicated by an arrow and represents the region that results in the most symmetrical peak shapes for analyte ions. Thus, the peak shapes in Figure 2 indicate that chromate has a mobility close to inorganic anions (peaks 1 and 2), phthalate has a mobility close to the carboxylates (peaks 3–5), and *p*-hydroxybenzoate has a mobility close to short-chain linear alkanesulfonates (peaks 6 and 7).

Shamsi and Danielson[35] used naphthalenesulfonates as IPD electrolytes and investigated the peak shapes for inorganic anions, organic acids and long-chain aliphatic sulfonates and sulfates. Figure 3 shows electropherograms of a standard mixture of seven common inorganic anions with singly charged NMS, doubly charged

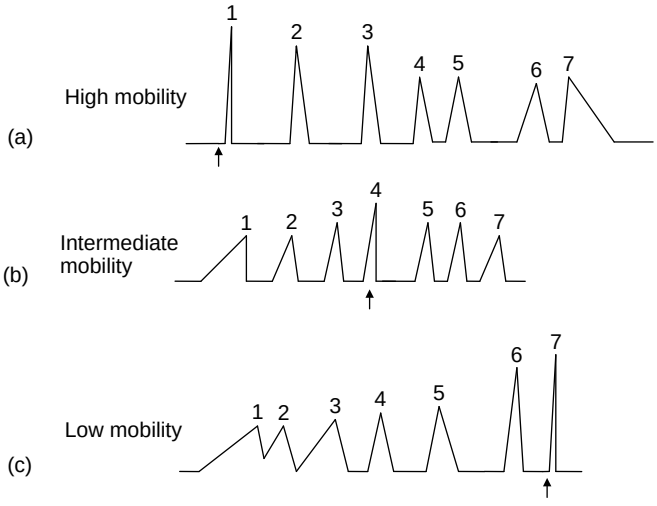

Figure 2 Comparison of the peak symmetry for high-mobility (a), intermediate-mobility (b) and low-mobility (c) electrolytes for CZE analysis of anions. Peaks 1 and 2 symbolize inorganic anions; peaks 3–5 represent short-chain carboxylates; and peaks 6 and 7 represent short-chain alkanesulfonates. Chromate (high mobility), phthalate (intermediate mobility), and *p*-hydroxybenzoate (low mobility) are examples of the IPD co-ions.

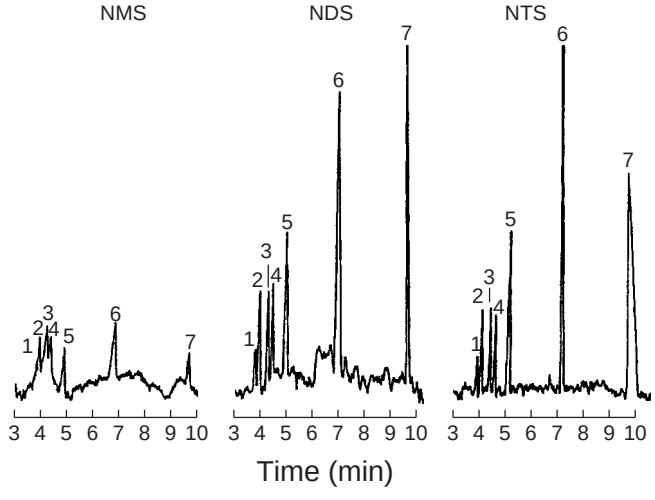

Figure 3 Comparison of naphthalenesulfonate electrolytes for the separation of seven common inorganic anions. The electrolyte consisted of 4 mM NMS, NDS or NTS in 100 mM H_3BO_3, 5 mM $Na_2B_4O_7$, 2 mM diethylenetriamine (DETA), pH 8 buffer. Peak identification: 10 ppm each of (1) bromide, (2) chloride, (3) nitrite, (4) nitrate, (5) sulfate, (6) fluoride, and (7) orthophosphate. Vacuum injection for 3 s, −30 kV applied for separation, and current = 16–18 μA. IPD at 274, 288, and 284 nm, respectively. (Reproduced with permission from Shamsi and Danielson.[35] Copyright 1994 American Chemical Society.)

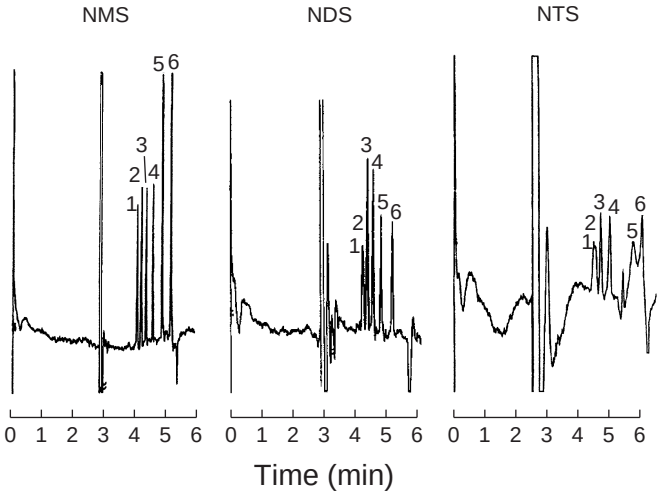

Figure 4 Comparison of naphthalenesulfonate electrolytes for the separation of alkyl sulfates. Peak identification: 20 ppm each of (1) C_{14} SO_4^-, (2) C_{12} SO_4^-, (3) C_{10} SO_4^-, (4) C_8 SO_4^-, (5) C_6 SO_4^-, and (6) C_4 SO_4^-. The electrolyte consisted of 5 mM NMS, NDS or NTS in 100 mM H_3BO_3, 5 mM $Na_2B_4O_7$, pH 8 buffer. Vacuum injection: 4 s for NMS and 5 s for NDS and NTS; +30 kV applied for separation; current = 8–10 μA. (Reproduced with permission from Shamsi and Danielson.[35] Copyright 1994 American Chemical Society.)

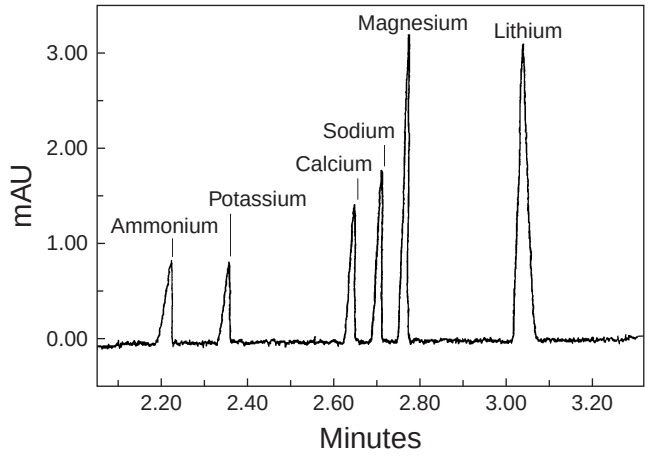

Figure 5 Separation of inorganic cations with imidazole/α-hydroxybutyric acid (α-HIBA) electrolyte. Conditions: 5 mM imidazole/6.5 mM α-HIBA/2 mM 18-crown-6, pH 4.5; 75 mm i.d. × 60 cm capillary; 30-s injection; 20 kV; UV detection at 214 nm. Sample 1 μg mL^{-1} of each cation. (Reproduced with permission from Mazzeo.[36])

naphthalenedisulfonate (NDS), and triply charged naphthalenetrisulfonate (NTS) electrolytes. Although all three electropherograms were run under similar conditions, the analyte peak shapes were very different. With NMS, poor sensitivity and peak fronting were evident, whereas NDS and NTS electrolytes provided almost equally good peak shapes for all seven anions. Note that for the NDS electrolyte, the orthophosphate peak becomes very symmetrical, whereas fluoride gives an excellent peak shape with NTS. These differences in peak shapes for orthophosphate and fluoride are consistent with their very close mobility match with NDS and NTS respectively. The authors clearly showed that although NDS and NTS provided good electrophoretic separations, NMS, having a much lower electrophoretic mobility than the inorganic anions, was not a suitable electrolyte for this mixture. In contrast, the reverse was found to be true when the same group of electrolytes were compared for the separation of aliphatic anionic surfactants (C_4–C_{14} SO_4^-). For C_4–C_{14} SO_4^-, NMS was found to be the best electrolyte, whereas NTS, because of its very high mobility, gives an inadequate separation for this particular mixture (Figure 4). In general, NDS showed much wider applicability (i.e. reasonable peak shapes for high-mobility inorganic anions, intermediate-mobility organic acids, and low-mobility long-chain aliphatic anionic surfactants) compared with NMS or NTS. Nevertheless, these studies clearly indicate that the selection of a UV-absorbing electrolyte in IPD must include the consideration of achieving not only high molar absorptivity, but one with a close mobility match to the sample components.

As in the case of anions, several studies have shown that mobility matching is also important for the IPD of cationic solutes.[36] An electropherogram is shown in Figure 5 for the separation of five inorganic cations and ammonium ions using imidazole as the IPD reagent. In this electropherogram, ammonium and potassium exhibit fronting whereas lithium exhibits tailing. Additionally, calcium, sodium, and magnesium show no peak asymmetry, indicating that the imidazole has a mobility close to these cations.

Peak shape is also related to the ratio of the conductivities of the analyte and the running electrolyte. According to Hjerten,[37] the absolute difference in conductivity depends on two factors: the difference in mobility between an electrolyte and the analyte ion; and the total concentration of the analyte. This dependence is shown by Equation (9):

$$\Delta\kappa = \frac{c_{\text{analyte}}}{\mu_{\text{analyte}}}(\mu_{\text{co-ion}} - \mu_{\text{analyte}})(\mu_{\text{counterion}} - \mu_{\text{analyte}}) \quad (9)$$

where $\Delta\kappa$ is described as the conductivity difference between a sample zone and the carrier electrolyte, c_{analyte} and μ_{analyte} are the concentration and mobility of the analyte respectively, and $\mu_{\text{co-ion}}$ and $\mu_{\text{counterion}}$ are the respective electrophoretic mobilities of the co-ion and the counterion. It is worth noting that the electrophoretic mobility of a cation usually has a positive sign, whereas the electrophoretic mobility of anions has a negative sign.

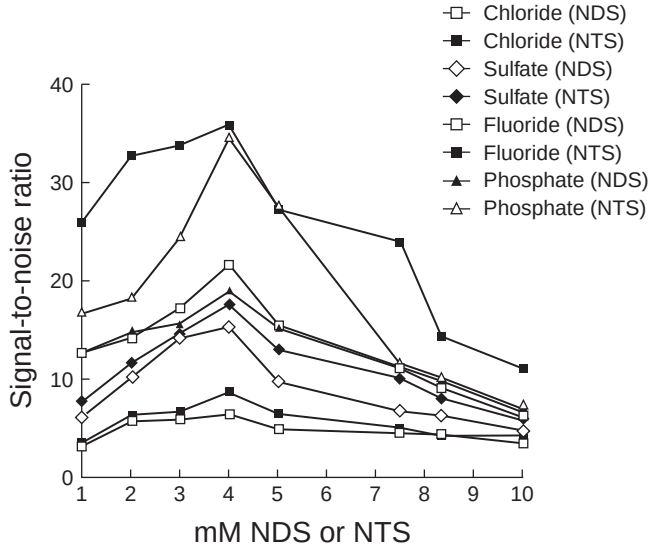

Figure 6 S/Ns of some inorganic anions as a function of UV-absorbing electrolyte (NDS or NTS) concentration prepared in a 100 mM H_3BO_3, 5 mM $Na_2B_4O_7$ 2 mM diethylenetriamine, pH 8 buffer. (Reproduced with permission from Shamsi.[52])

2.3 Concentration of the Indirect Photometric Detection Reagent

The concentration of the IPD electrolyte co-ion is also an important consideration, for two major reasons. First, it is the ratio of the conductivities of the sample and the electrolyte that will control the peak shape; in general, peak efficiency and resolution increase with increasing ratio of the light-absorbing carrier electrolyte concentration to sample concentration. Second, the absorbance detector used in IPD mode shows a linear response range up to 0.20 AU, therefore the maximum concentration of the IPD electrolyte that can be used is the concentration that will give a background absorbance of <0.20 AU. Recent studies by Shamsi et al. indicated that the background absorbance inside the fused-silica capillary can be linear up to ca. 0.5 AU, a value dependent on both the type of IPD electrolyte and the detector optics.[24] The same authors have also studied the effect of increasing the concentration of NDS or NTS reagent on the S/N of some inorganic anions. As shown in Figure 6, sensitivity first tends to increase with the increase in NDS or NTS concentration between 1 and 4 mM, providing sharper peaks and better resolution. With a further increase in concentration, i.e. >4 mM NDS or NTS, a downward trend in S/N is evident. This decrease in S/N is obviously due to an increase in the magnitude of noise, which increases proportionally with increasing concentration of UV-absorbing electrolytes. In addition, the linear response range for the nonchromophoric analytes is strongly influenced by the IPD electrolyte concentration. The use of a higher concentration of IPD reagents will extend the upper linear range of the analyte ion, although at a high IPD electrolyte concentration the background noise and Joule heating also increase, thus decreasing the sensitivity of IPD. In general, most electrolytes used in IPD have concentrations in the range 1–10 mM. This means three orders of magnitude in linearity for the analyte ion. It should be noted that in IPD the running electrolyte concentrations are ca. 10 times less than those employed in CZE with direct UV detection. Therefore, peak distortion due to a mobility mismatch is more critical with indirect than with direct photometric detection.

2.4 Type and Concentration of the Nonabsorbing Buffer

One approach that optimizes electrophoretic separation and IPD of inorganic anions is the use of nonabsorbing co-anionic buffers such as borate,[24,35,38–40] acetate,[41] phosphate,[42] and carbonate.[43] Obvious advantages of these buffer co-anions are high reproducibility, sharper peaks, and rugged separations. In practice, baseline noise at higher ionic strength of the co-anionic buffer results in reduced S/N due to dilution of the zones of the light-absorbing reagent by nonabsorbing co-anions. Moreover, the presence of these co-anionic buffers leads to system peaks that may interfere in the analysis of some inorganic anions.[24,35]

The migration time and S/N of some inorganic anions as a function of borate ($H_3BO_3/Na_2B_4O_7$) concentration from 20 to 200 mM has little effect on the migration of anions. However, increasing the concentration of the same buffer anions improved the detection limits by a factor of 2 for low-mobility inorganic anions (e.g. phosphate, fluoride) (Figure 7). In contrast, little effect on

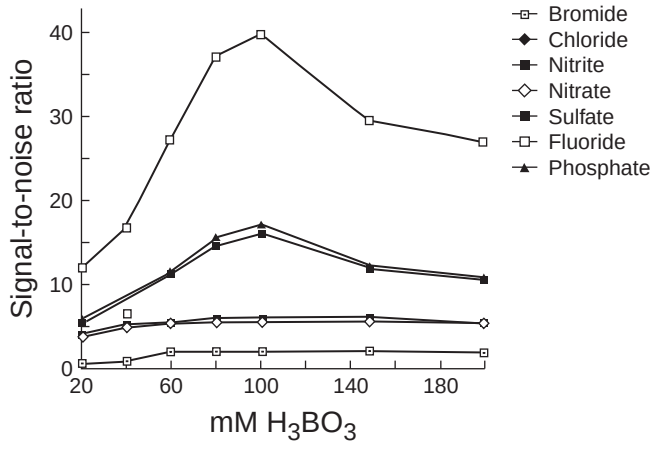

Figure 7 S/Ns of some inorganic anions as a function of borate concentration prepared in a 3 mM NTS, 2 mM diethylenetriamine, pH 8 buffer. (Reproduced with permission from Shamsi.[52])

the S/N of high-mobility inorganic anions (e.g. bromide, chloride, nitrite, and nitrate) was seen. Probably, the mobility of borate anion matches the former more closely than the latter group of anions.

A second approach entails the use of nonabsorbing countercations such as Tris[44–47] and triethanolamine.[48–50] The running CZE buffers are generally prepared by titration of the acid form of IPD electrolyte with the buffering base. Because these running CZE buffers consist of a single co-anion, the charge displacement process is simplified and the S/N remains unaffected.

2.5 pH of the Indirect Photometric Detection Reagent

Similarly to ion-exchange chromatography, IPD in CZE is also sensitive to pH extremes. The CZE separations of inorganic cations are mostly performed over the range 3–5. In contrast, anionic CZE separations are rarely obtained at a pH of <7.0. The reason for this is that these inorganic cations and anions, as well as the capillary surface, have different properties at different pH values. For example, at high pH, CZE separations of some metal cations (e.g. transition metals, alkaline earths) are not feasible, due to precipitate formation and interaction of these cations with the negative charge of the fused-silica capillary. Moreover, aromatic bases that are commonly used as IPD reagents for cation separations have pK_a values in the acidic pH range. This means that they are not appreciably ionized at neutral to basic pH values. In view of this, the pH of the running electrolyte should be sufficiently low so that IPD reagent exists as cations with sufficiently high mobility to match the mobility of fast-moving inorganic cations. On the other hand, the selection of pH for the CZE separation of inorganic anions is not as critical. This is because the majority of the inorganic anions are strong acids and hence are completely ionized in neutral to alkaline pH conditions. One exception is the weakly ionized solute (e.g. carbohydrates), which requires a pH approaching 12 to have any appreciable fraction in the anionic form. When choosing pH > 12, the concentration of the hydroxide ions is no longer negligible relative to the concentration of the IPD electrolyte. Consequently, a decrease in TR generally reduces the IPD signal.

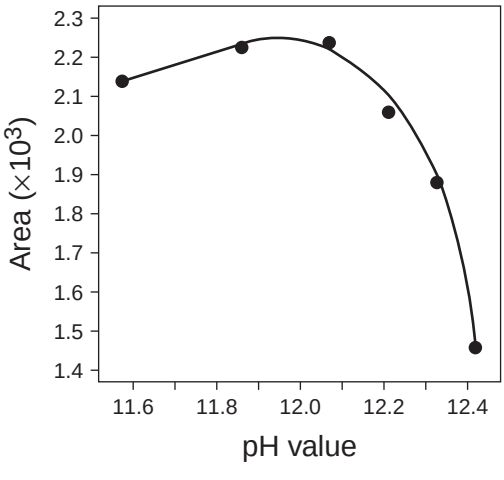

Figure 8 Impact of pH value on sensitivity. Electrolyte is 6 mM sorbate; fused-silica capillary = 122 × 100 cm, 50 μm i.d; current is 6, 8, 13, 17, 20, and 24 μA at pH 11.58, 11.87, 12.08, 12.22, 12.33, and 12.42, respectively; voltage is +28 kV; separation temperature = 30 °C; UV detection at 256 nm; sampling by vacuum injection for 1 s; sample is 12.5 mM mannose. (Reproduced with permission from Jandik and Bonn.[16])

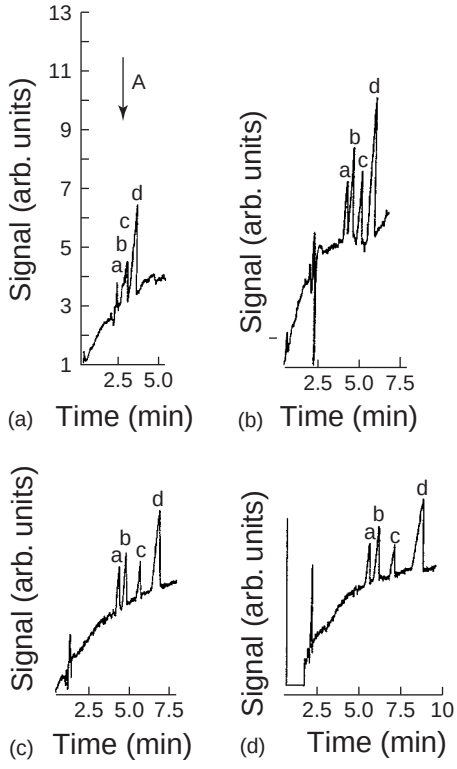

Figure 9 The influence of pH of the background electrolyte solution: 10 mM phthalic acid. The pH values tested were: (a) pH 4.3; (b) pH 6.5; (c) pH 7.7; (d) pH 9.2. Injection time was 2 s. The concentration of the anions was 50 μM each. Peak identification: a = bromide; b = chloride; c = nitrate; d = sulfate. (Reproduced with permission from Ma and Zhang.[31])

Vorndran et al.[51] first described the use of sorbate as the IPD reagent for the analysis of carbohydrates. A rather simple electrolyte system containing 6 mM sorbate adjusted to pH 12.1 with sodium hydroxide permits the separation of 11 carbohydrates in ca. 20 min. Although resolution of carbohydrates improved with rising pH, the sensitivity of detection decreased considerably above pH 12.1 (Figure 8). Ma et al.[31] investigated the influ-

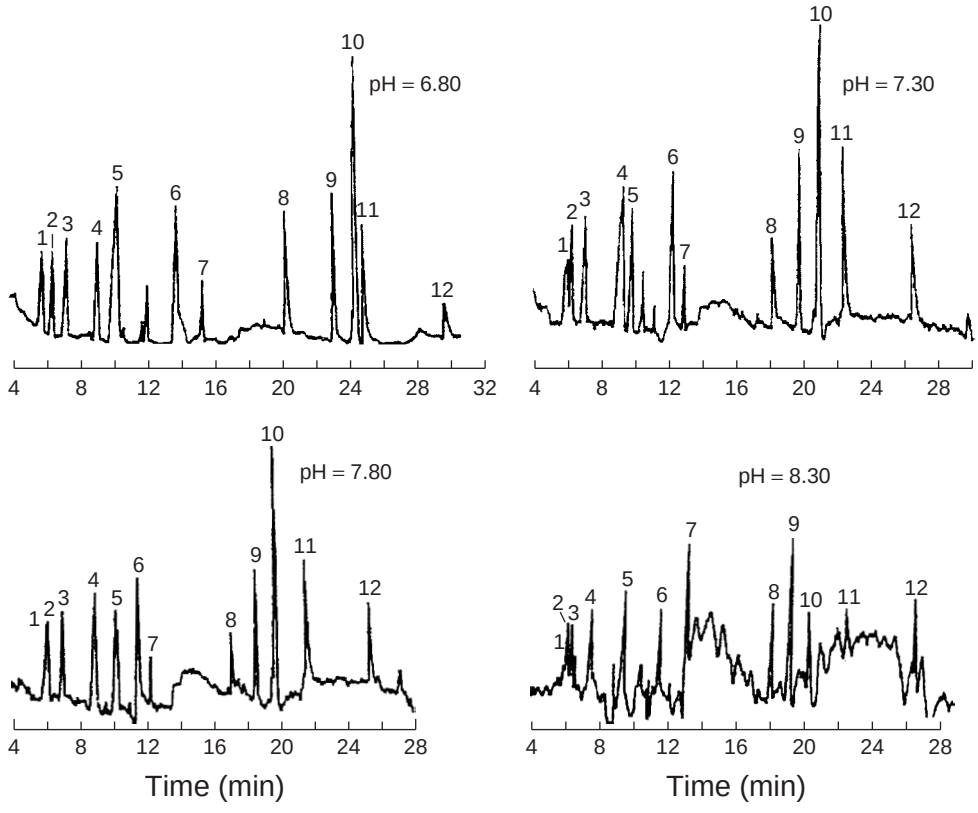

Figure 10 Effect of electrolyte pH on the separation of 12 P_n and RP_n. Electrolyte consisted of 5 mM AMP and 2 mM DETA adjusted to various pH values with 1 mM NaOH. Vacuum injection for 15 s; −30 kV applied for separation; current = 6–13 μA. Peak identification: 5–10 mg L^{-1} each of: (1) 21 μmol L^{-1} cyclic P_3; (2) 51 μmol L^{-1} P_1 fluorophosphate; (3) 62 μmol L^{-1} P_1 phosphite; (4) 52 μmol L^{-1} P_1 phosphate; (5) 78 μmol L^{-1} P_1 hypophosphite; (6) 49 μmol L^{-1} 1-hydroxyethylidene-1,1-diphosphonic acid (HEDP); (7) 21 μmol L^{-1} cyclic P_6; (8) 57 μmol L^{-1} P_2 phosphate; (9) 21 μmol L^{-1} hexamethylenediaminetetra(methylenephosphonic acid) (HDTMP); (10) 34 μmol L^{-1} aminotri(methylenephosphonic acid) (ATMP); (11) 23 μmol L^{-1} ethylenediaminetetra(methylenephosphonic acid) (EDTMP); (12) 18 μmol L^{-1} diethylenetriaminepenta(methylenephosphonic acid) (DETPMP). (Reproduced with permission from Shamsi and Danielson.[24] Copyright 1995 American Chemical Society.)

ence of the pH of the phthalate electrolyte on the separation and IPD of four inorganic anions (bromide, chloride, nitrate, and sulfate). At low pH (i.e. pH 4.3), phthalate electrolyte being singly charged has lower electrophoretic mobility, resulting in poor S/N for the high-mobility inorganic anions (Figure 9, spectrum a). Increasing the pH to 6.5 fully ionized the phthalate electrolyte and improved the S/N (Figure 9, spectrum b). Further increase in pH value enhanced the overall resolution of the anions (Figure 9, spectra c and d), but at the expense of longer migration times and peak broadening.

Shamsi and Danielson[24] studied the effect of pH on the detection limits of phosphorus-containing anions using AMP as the IPD reagent. The four electropherograms shown in Figure 10 illustrate that both the S/N and the migration time of P_n and RP_n decrease as the pH is increased from 6.30 to 8.30. However, the overall increase in mobility of these anions is not as much as one would have expected. This is because electroosmotic flow and electrophoretic vectors oppose each other. The increase in baseline noise at pH 8.30 was found to be caused by dilution of the AMP zones by the nonabsorbing co-anions (hydroxide and borate anions). Excessive heating due to an increase in current from 6 μA at a pH of 6.80 to 13 μA at a pH of 8.30 resulted in greater mobility of hydroxide ions, and consequently an increase in the baseline noise.

Shamsi and Danielson also compared the theoretical and experimentally determined detection limits of 12 P_n and RP_n. The experimentally determined TR (ΔT) was compared with the TR expected on the basis of an equivalent-to-equivalent exchange between the electrolyte and analyte ions. It was shown that there can be situations in which a large TR value for some analytes is less important than a small ΔT value with respect to the experimental detection limits. Several examples of this situation are shown in Table 1. At a pH of 7.80, P_2 phosphate, HDTMP, ATMP, EDTMP, and DETPMP have more favorable ΔT and TR values than P_1 phosphite,

Table 1 Comparison of theoretical and experimentally determined detection limits using AMP as electrolyte (effective charges of −1.8 and −2.0 at pH 6.8 and 7.8, respectively, calculated using equation from Haddad and Jackson[198])

Analyte	Effective charge		ΔT[a,b]		TR[c]		Calc. LOD × 10^6 (M)[d]		Exp. LOD × 10^6 (M)	
	pH 6.80	pH 7.80	pH 6.80	pH 7.80	pH 6.80	pH 7.80	pH 6.80	pH 7.80	pH 6.80	pH 7.80
P_1 fluorophosphate	–	–	+3.2	+2.8	–	–	–	–	2.0	10.0
Cyclic P_3	–	–	+3.0	+2.8	–	–	–	–	0.84	4.2
P_1 phosphite	−1.6	−1.9	+2.8	+2.6	0.86	0.97	9.5	8.5	2.4	8.3
P_1 phosphate	−1.3	−1.8	+2.3	+2.1	0.70	0.91	11.0	8.9	2.5	8.4
P_1 hypophosphite	−1.0	−1.0	+2.0	+1.8	0.55	0.51	15.0	16.0	3.1	11.0
HEDP (Dequest 2010)	−2.4	−2.9	+1.2	+1.4	1.3	1.5	6.2	5.6	1.2	3.0
Cyclic P_6	–	–	+0.8	+1.4	–	–	–	–	3.0	3.2
P_2 phosphate	−2.6	−3.0	−0.5	−0.2	1.4	1.5	5.7	5.4	5.6	5.7
HDTMP (Dequest 2054)	−3.3	−3.9	−1.2	−0.1	1.8	2.0	4.5	4.2	1.0	1.2
ATMP (Dequest 2006)	−3.6	−3.9	−1.6	−0.5	1.9	2.0	4.2	4.1	4.4	4.5
EDTMP (Dequest 2041)	−2.6	−3.7	−1.7	−0.9	1.4	1.9	5.7	4.4	5.8	4.6
DETPMP (Dequest 2060)	−4.0	−5.1	−3.0	−1.7	2.2	2.6	3.7	3.1	4.8	2.4

[a] Measured with respect to nitrate.
[b] ΔT = Relative migration time of AMP − Relative migration time of analyte.
[c] TR = Effective charge of analyte/Effective charge of AMP.[24]
[d] Calculated using Equation (1).

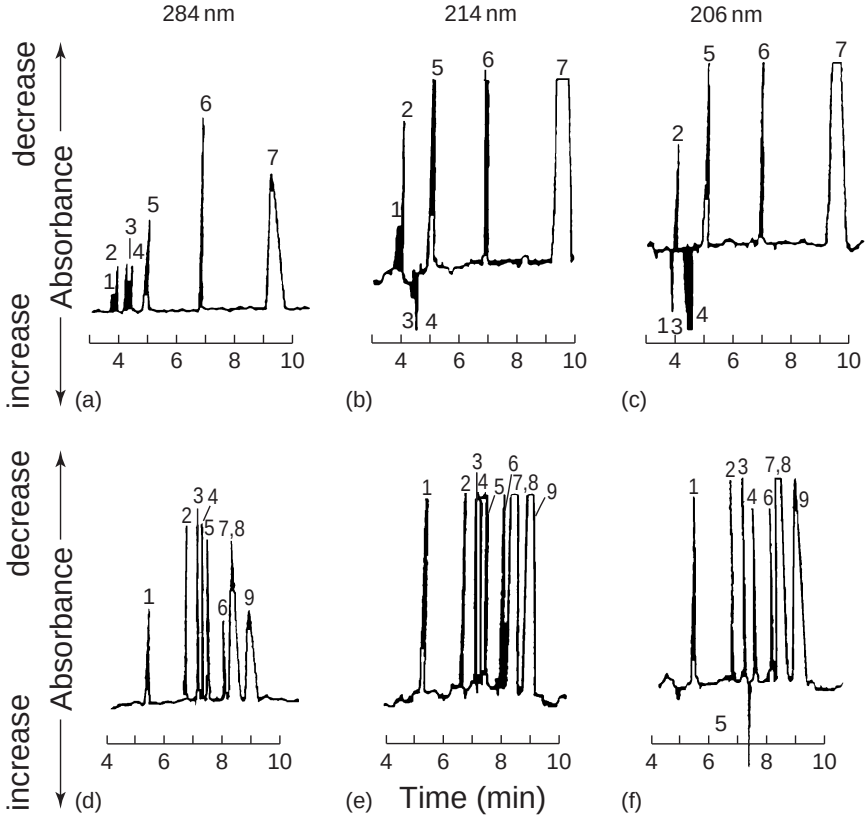

Figure 11 Effect of wavelength on sensitivity and direction of photometric signal for (a–c) inorganic anions and (d–f) organic acids. Peak identification: (a–c) as Figure 3; (d–f) 1 = oxalate, 2 = malonate, 3 = formate, 4 = fumarate, 5 = maleate, 6 = succinate, 7 = malate, 8 = citrate, and 9 = tartrate. (Reproduced with permission from Shamsi.[52])

P_1 phosphate, or P_1 hypophosphite. For this reason, the detection limits of the former group of analytes were 2–5 times better then the latter. However, at a lower pH of 6.80 the differences in ΔT and the experimental detection limits between these two groups of analytes were not great. In general, the authors found that a low pH improves the sensitivity but at the same time increases the ΔT values for P_n and RP_n. Because of these two competing factors, the detection limits of some phosphorus anions, such as P_1 phosphate or P_1 hypophosphite, were improved whereas the detection limits of DETPMP became worse. On the other hand, there was not much change in the detectability of HDTMP, ATMP, and EDTMP at pH 6.80 compared to pH 7.80.

2.6 Effect of Detection Wavelength

Another method of optimizing sensitivity is the choice of an appropriate detection wavelength. The largest signal generated for IPD is found where the light-absorbing reagent absorbs most strongly and the analytes absorb the least. In general, most inorganic cations have no UV absorbance at any wavelength. Hence, IPD reagents for inorganic cations can be selected that have strong UV absorbance at wavelengths as low as 185 nm. Figure 11 shows the effect of wavelength on sensitivity and direction of the photometric signal for seven common inorganic anions and organic acids at three different detector wavelengths using NTS as the IPD reagent. Although shorter wavelengths of 214 and 206 nm can improve sensitivity for some nonabsorbing analytes, a more universal detection (i.e. all peaks in one direction) is obtained at a longer wavelength of 284 nm for both inorganic anions and organic acids.[52] Although NTS does have another wavelength maximum at 234 nm, this wavelength was not utilized because the absorbance was outside the Beer's law working range. Analytes such as nitrite and nitrate, which are UV-absorbing at a wavelength of 214 nm, showed positive peaks but nitrite detectability was not improved because the absorption difference between nitrite and NTS is not great at this wavelength. In contrast, the sensitivity of detection of bromide, nitrite and nitrate can be improved by direct UV detection at a wavelength of 206 nm. In addition, organic acids such as maleate can be detected directly at 206 nm.

Optimizing the sensitivity for inorganic anions using chromate as an IPD electrolyte at different wavelengths has also been studied by Jones in detail.[53] This

Table 2 Peak height ratio with respect to 254 nm for nine different wavelengths using a mixture of 11 inorganic anions and detection limits defined as 3 × noise (in ppm)[a]

No.	Wavelength (nm)								
	185	214	229	254	313	365	405	436	546
Peak height (with respect to 254 nm)									
1 Thiosulfate	3.55	−0.58	−0.61	1.00	dl	1.67	0.57	0.14	np
2 Bromide	−0.18	−0.08	0.21	1.00	dl	1.56	0.54	0.14	np
3 Chloride	2.71	0.51	0.20	1.00	dl	1.52	0.52	0.12	np
4 Sulfate	4.04	0.53	0.20	1.00	dl	1.51	0.52	0.13	np
5 Nitrite	2.87	−2.29	−0.92	1.00	dl	1.55	0.53	0.11	np
6 Nitrate	2.26	−2.94	−0.38	1.00	dl	1.49	0.53	0.14	np
7 Molybdate	5.10	−2.24	−1.74	1.00	dl	2.19	0.75	0.19	np
8 Tungstate	2.64	0.12	0.16	1.00	dl	1.53	0.54	0.12	np
9 Fluoride	4.03	0.52	0.21	1.00	dl	1.57	0.51	0.15	np
10 Phosphate	3.79	0.60	0.25	1.00	dl	1.44	0.49	0.12	np
11 Carbonate	3.93	0.49	0.17	1.00	dl	1.55	0.58	0.12	np
Detection limit (3 × noise, in ppm)									
1 Thiosulfate	0.71	−1.32	−1.39	0.25	6.00	0.70	2.68	3.05	np
2 Bromide	−10.17	−6.84	2.62	0.18	6.00	0.55	2.05	2.21	np
3 Chloride	0.38	0.60	1.52	0.10	3.00	0.31	1.17	1.40	np
4 Sulfate	0.30	0.69	1.84	0.12	6.00	0.38	1.42	1.63	np
5 Nitrite	0.51	−0.19	−0.48	0.15	6.00	0.44	1.65	2.15	np
6 Nitrate	0.95	−0.22	−1.72	0.22	6.00	0.67	2.42	2.48	np
7 Molybdate	1.13	−0.71	−1.00	0.58	15.00	1.23	4.64	5.15	np
8 Tungstate	2.31	15.71	11.25	0.61	15.00	1.85	6.74	8.52	np
9 Fluoride	0.19	0.43	1.09	0.08	1.50	0.22	0.88	0.86	np
10 Phosphate	0.78	1.49	3.55	0.30	6.00	0.96	3.63	4.29	np
11 Carbonate	0.84	2.02	5.74	0.33	6.00	1.00	3.41	4.72	np

[a] Reproduced by permission from Jones.[53] (Copyright 1993 Elsevier Science N.V.) A negative number indicates that the analyte is more absorbing than the electrolyte. The terms dl and np stand for detection limit and no peak, respectively.

electrolyte provides the strongest background signal at 185 nm, followed by the second and third strongest background signals at 365 nm and 254 nm respectively. As shown in Table 2, 185 nm provided the largest peak height ratio for 9 of 11 inorganic anions, whereas 365 nm had the second and 254 nm had the third largest peak height ratio for at least seven inorganic anions. However, the largest peak height does not necessarily provide the lowest detection limits. This is because the lamp energy can also influence the detector performance. Higher-energy lamps generate more photons per unit time, thus reducing the shot noise. A comparison of lamp output energy at 185, 254, and 365 nm indicated that the mercury lamp operated at 254 nm provided the highest photodiode current (ca. 1.55 nA) and thus the best detection limits for 11 inorganic anions (Table 2).

2.7 Effect of Background Noise

In general, the noise level observed in IPD is higher than in direct detection using the same instrument. Wang and Hartwick[47] were the first to study in detail the factors that influence the background noise in IPD. The authors found that the noise in IPD is generated from at least two different sources: the IPD reagent, and the detector electronics. Although the mechanism of noise produced from the IPD reagent was not clear, it appeared to be dependent on the concentration of the IPD reagent, the applied voltage, and the surface modification of the capillary. To improve the C_{LOD} in IPD, light-emitting diodes have been developed that provide excellent baseline stability and reduced noise level up to one order of magnitude.[54–56] However, these diodes require the use of IPD reagents that absorbs light in the visible region. Characterization of a suitable IPD electrolyte in the visible region is problematic for two major reasons: first, chromophores that have intense absorption in the visible region are large organic molecules with lower electrophoretic mobilities that do not match those of the analytes (e.g. high-mobility inorganic anions); and second, the adsorptive interactions of such molecules with the capillary cause baseline fluctuations. Thus, the actual C_{LOD} in the visible region with light-emitting diodes is no better than the conventional IPD in the UV region.

2.8 Effect of Absorption Pathlength

The influence of detection cell length, l, with the aim of increasing the background noise, has been investigated by several authors. Ma and Zhang.[31] studied the effect of increasing pathlength by comparing capillaries having different diameters. They observed that the detection limits were about the same when a capillary internal diameter was increased from 25 μm to 75 μm, but Joule heating increased with increasing capillary diameter, resulting in an increase in the background noise. This result is in agreement with the data published by Anderson et al.,[57] but is in contrast to the report by Steinar et al.[58] in which the S/N increased with an increase in capillary diameter from 10 μm to 1000 μm. Probably in the latter study the background noise level was not measured with the voltage on, and therefore the contribution of Joule heating to noise was insignificant. Weston et al. reported improvement in the IPD signal by a factor of approximately two using a capillary of extended pathlength.[26] A 75-μm capillary was extended to 300 μm by blowing a bubble at the detection window. This procedure is very useful because the contribution of Joule heating to the background noise is negligible due to the capillary diameter being unchanged for almost the entire capillary length.

3 DETERMINATION OF ANIONIC AND CATIONIC COMPOUNDS

Detection of nonchromophoric anions and cations in CZE is usually accomplished using anionic and cationic chromophoric reagents, respectively. Tables 3–10 list typical carrier electrolytes that are compatible for IPD of anionic and cationic solutes. These compounds can be classified into groups as shown and discussed below:

1. IPD of inorganic anions (Table 3);
2. IPD of inorganic cations (Table 4);
3. IPD of short-chain organic acids and amines (Table 5);
4. IPD of long-chain fatty acids (Table 6);
5. IPD of inorganic and organic phosphates (Table 7);
6. IPD of long-chain aliphatic anionic and cationic surfactants (Table 8);
7. IPD of carbohydrates (Table 9);
8. IPD of amino acids and phospholipids (Table 10).

3.1 Indirect Photometric Detection of Inorganic Anions

As discussed earlier, IPD was probably used for the first time in CZE by Hjerten et al. in 1987[1] and Foret et al. in 1989.[2] Although these publications addressed the CZE determination of some low-molecular-weight anionic species, a real breakthrough in the development of IPD in CZE was a consequence of the work of Jones and Jandik.[3–8] These researchers were the first to report the analysis of 36 inorganic and organic anions in ca. 3.0 min using a chromate as the IPD electrolyte and tetradecyltrimethylammonium bromide (TTAB) to reverse the electroosmotic flow (Figure 12).[8] Since that report, chromate has been the most widely used carrier

electrolyte for the analysis of inorganic anions (Table 3), for several important reasons. One is that chromate is a high-mobility anion that provides very symmetrical peak shapes for analysis of fast-moving inorganic solutes (e.g. bromide, chloride, nitrite, nitrate, sulfate, fluoride, and phosphate). In addition, this electrolyte possesses strong UV absorption over a broad (200–400 nm) range. The second reason is that detection using chromate is performed at 254 nm, which maximizes the largest possible number of anions to be detected universally, including those that are themselves UV-absorbing at low wavelengths. This is beneficial because most commercial CE instruments have a wide DR with low noise at 254 nm.

Table 3 Some typical examples of IPD in CE separation of inorganic anions

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Refs.
Thiosulfate, bromide, chloride, sulfate, nitrite, nitrate, azide, molybdate, chlorite, tungstate, phosphite, monofluorophosphate, chlorate, carbonate, phosphate, fluoride	Standard mixture; Kraft black liquor; urine	Chromate	5 mM Sodium chromate, 0.4 mM OFM Anion-BT adjusted to pH 8	254	3, 6–8
Thiosulfate, bromide, chloride, sulfate, nitrite, nitrate, phosphate, fluoride	Standard mixture	Chromate	5 mM Sodium chromate, 0.5 mM TTAB, 5 mM boric acid, pH 8	254	49
Chloride, bromide, nitrite, sulfate, nitrate, fluoride, phosphate, carbonate	Standard mixture	Chromate	5 mM Sodium chromate, 3.5 mM tetradecyltrimethylammonium hydroxide, 10 mM CHES, pH 9	254	36
Bromide, chloride, nitrite, nitrate, sulfate, fluoride, phosphate	Standard mixture	Chromate	5 mM Sodium chromate, 0.14% (w/v) PDDPiCr	254	192
Thiosulfate, chloride, sulfate, sulfite, carbonate	Standard mixture	Chromate	5 mM Sodium chromate, 0.001% hexadimethrine bromide, 20% acetonitrile, pH 11	185	193
Arsenate, bromate, bromide, chlorate, chloride, fluoride, molybdate, nitrate, nitrite, perchlorate, sulfate, phosphate, tetrafluoroborate, thiocyanate, thiosulfate, tungstate	Standard mixture; tapwater; ground and surface water; steam flow; swimming-pool water	Pyromellitate	2.25 mM Pyromellitate, 6.5 mM sodium hydroxide, 1.6 mM triethylamine (TEA), 0.75 mM hexamethonium bromide or hydroxide	250	20, 29, 30, 49, 59, 196
Bromide, chloride, nitrate, fluoride, sulfate, phosphate		Pyromellitate	2.25 mM Pyromellitate, 6.5 mM sodium hydroxide, 1.6 mM TEA	250	194
Bromide, chloride, nitrite, nitrate, sulfate, fluoride, phosphate, dithionate, tetrathionate, thiocyanate, iodide, chlorite, bromate, chlorate, iodate, perchlorate, cyclic P_3	Standard mixture	NDS; NTS	4–8.3 mM NDS or NTS, 100 mM H_3BO_3/5 mM $Na_2B_4O_7$, 2 mM DETA, pH 8	288, 284, 214	35

(*continued overleaf*)

Table 3 (*continued*)

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Refs.
Bromide, chloride, nitrate, nitrite, sulfate, fluoride, phosphate, phosphite	Drinking and wastewaters	Trimellitate	5 mM Trimellitate, 1 mM TTAB	220, 240, 265	62
Fluoride, chloride, sulfate, phosphate	Standard mixture	Trimellitate	5 mM Trimellitate, 1 mM TTAB	240	20
Bromide, chloride, nitrate, sulfate	Standard mixture	*O*-Benzylbenzoic acid	20 mM *O*-Benzylbenzoic acid, pH 6.5	228	31
Chloride, fluoride, nitrite, nitrate, sulfate, phosphate, phosphite	Standard mixture; fermentation broths	2-Nitroso-1-naphthol-6-sulfonate; 1-nitroso-2-naphthol-3,6-disulfonate	0.5 mM 2-Nitroso-1-naphthol-6-disulfonate or 1-nitroso-2-naphthol-3,6-disulfonate, pH 8	254	63
Bromide chloride, nitrite, nitrate, sulfate, thiosulfate	Standard mixture	Vanadate	7.5 mM Sodium vanadate, pH 8	254, 226, 270	197
Chlorate, chloride, fluoride, nitrate, sulfate	Standard mixture	Permanganate	20 mM Permanganate, pH 7	635, 670	55
Bromide, chloride, nitrate, sulfate	Standard mixture	Chlorophenol red; indigo carmine	0.5 mM Chlorophenol red or indigo carmine, 5 mM Tris, pH 6 adjusted with acetic acid	635, 670	27
Sulfate, sulfide, tetrathionate, thiosulfate	Standard mixture	1,3-Benzene-disulfonate	3 mM 1,3-Benzenedisulfonic acid, 10 mM Tris, 0.5 mM DETA, pH 8.15	254, 214	29
Bromide, chloride, nitrate, sulfate	Standard mixture	2-Sulfobenzoic acid	20 mM 2-Sulfobenzoic acid, pH 6.5	228	31

CHES, 2-(*N*-cyclohexylamino)ethanesulfonic acid; PDDPiCr, poly(1,1-dimethyl-3,5-dimethylenepiperidinium) chromate.

Although the first recipe of chromate/TTAB electrolyte was very useful for IPD of many inorganic anions, this recipe has suffered from several problems. First, the electroosmotic flow modifier TTAB used in the chromate electrolyte was a bromide salt. Samples containing trace bromide levels induced a negative peak where bromide (generated from TTAB) normally migrates in the electropherogram. This negative peak not only decreased the bromide sensitivity but also caused problems with chloride integration. Second, the elution order of some inorganic anions (thiosulfate < bromide < chloride) was not very beneficial. A favorable elution order would be one in which chloride migrated first so that it could be used as a reference component. The benefit of this technique is easier peak identification based on the migration time ratio (migration time of analyte peak/migration time of the chloride peak). Additionally, referencing can also minimize drifts in the migration time due to changes in electroosmotic flow. Third, this recipe using the chromate electrolyte involved the electrolyte being prepared from the sodium salt and therefore unbuffered. It is well known that the use of an unbuffered electrolyte in CZE can cause electrolysis of water at the inlet and outlet vials.

To utilize better the benefits of chromate in IPD of inorganic anions, Mazzeo[36] has proposed reformulation of the chromate electrolyte. The author has reported the use of tetradecyltrimethylammonium hydroxide instead of TTAB. The conversion of TTAB to its hydroxide was accomplished by passing the former through an anion-exchange cartridge in the hydroxide form. The main problem with this procedure is the large excess of hydroxide ions introduced, resulting in a high-pH chromate. In such cases, it becomes necessary to reduce the pH with various organic acids (e.g. acetic acid or boric acid). However, a negative peak still appears in the electropherogram that corresponds to the migration time of the acid co-anion. Several reports have appeared describing the addition of low-mobility co-anionic buffers such as carbonate,[43] borate,[40,49,57] and CHES.[58] Because of the low mobility of the aforementioned co-anionic buffers, a negative system peak is observed in the electropherogram but is well removed from the high-mobility inorganic anions. Nevertheless, the addition of

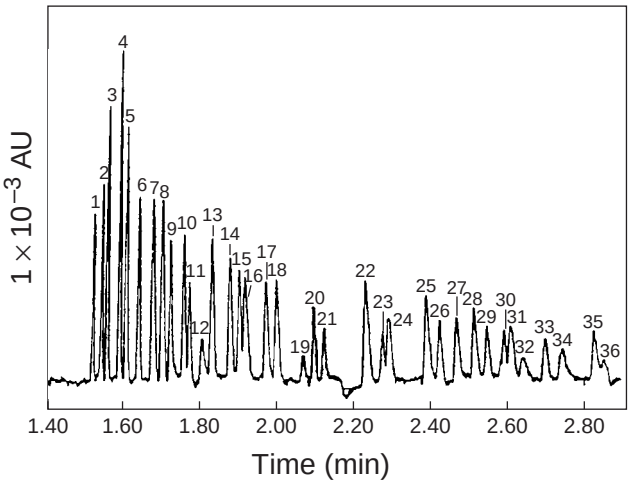

Figure 12 Electropherogram (83-s section) of 36 anions. Peak identification and concentrations (ppm): 1 = thiosulfate (1.3); 2 = bromide (1.3); 3 = chloride (0.7); 4 = sulfate (1.3); 5 = nitrite (1.3); 6 = nitrate (1.3); 7 = molybdate (3.3); 8 = azide (1.3); 9 = tungstate (3.3); 10 = monofluorophosphate (1.3); 11 = chlorate (1.3); 12 = citrate (0.7); 13 = fluoride (0.3); 14 = formate (0.7); 15 = phosphate (1.3); 16 = phosphite (1.3); 17 = chlorite (1.3); 18 = glutarate (1.7); 19 = *o*-phthalate (0.7); 20 = galactarate (1.3); 21 = carbonate (1.3); 22 = acetate (1.3); 23 = chloroacetate (0.7); 24 = ethanesulfonate (1.3); 25 = propionate (1.3); 26 = propanesulfonate (1.3); 27 = D,L-aspartate (1.3); 28 = crotonate (1.3); 29 = butyrate (1.3); 30 = butanesulfonate (1.3); 31 = valerate (1.3); 32 = benzoate (1.3); 33 = L-glutamate (1.3); 34 = pentanesulfonate (1.7); 35 = D-gluconate (1.7); 36 = D-galacturonate (1.7). The electrolyte is 5 mM chromate and 0.4 mM OFM Anion-BT adjusted to pH 8.0. Applied potential is 30 kV (negative polarity); capillary dimensions are 60 cm (52 cm to detector) × 50 μm i.d. fused silica; indirect UV detection; injection by electropherogram at 1 kV for 15 s. (Reproduced with permission from Jones and Jandik.[8])

such buffering agents requires a careful pH optimization to prevent the interference of system peaks with analytes of interest. As suggested by Haddad,[10] a more attractive alternative to buffer the chromate electrolyte is to use a buffering countercation such as Tris or diethanolamine in the chromate electrolyte.

Pyromellitic acid (PMA; or 1,2,4,5-benzenetetracarboxylic acid) is another suitable IPD reagent for the analysis of high-mobility inorganic anions.[20,29,30,49,59] Although PMA offers higher molar absorptivity than chromate, the latter is still a preferred IPD reagent because of its excellent mobility match for the common inorganic anions. Mazzeo[36] has also reported improved resolution and better peak shapes with shorter analysis time for inorganic anions using chromate instead of PMA.

Recently, naphthalenesulfonate reagents have been characterized by Shamsi and Danielson for IPD of a large number of inorganic anions.[35] A comparison of NMS, NDS, and NTS revealed that NMS was the best electrolyte for aliphatic anionic surfactants, whereas NDS and NTS caused more efficient separation of inorganic anions (Figures 3 and 4). NDS, in particular, was found to be applicable for efficient CZE separation for a much wider analyte range. For example, NDS electrolyte provided successful IPD of all three classes of anions, namely inorganic anions, organic acids, and aliphatic anionic surfactants.

Several other IPD reagents, including phthalate,[20,29,31,39] trimellitate,[20,47,60–62] *o*-benzylbenzoic acid,[31] sulfobenzoic acid,[31] and 1-nitroso-2-naphthol-3,6-disulfonate,[63] have also been examined. However, these reagents provide sensitive IPD of only a few inorganic anions. As a general observation, it can be said that currently chromate, PMA, and NTS are the three most suitable IPD reagents for the CZE analysis of high-mobility inorganic anions.

3.2 Indirect Photometric Detection of Inorganic Cations

As evidenced by recent reviews, CZE with IPD has been widely used in the analysis of inorganic cations (Table 4) in various samples.[64,65] Unlike anion analysis, CZE separation of cations is easily obtained in the low pH range of 4–5 without modifying the capillary surface and without reversing the electroosmotic flow. Owing to small differences in the electrophoretic mobility of a metal cation, it has become quite common to include a neutral or anionic complexing agent in the IPD buffer to enhance cation selectivity. Short-chain mono-, di-, tri- and tetracarboxylic acids are the most popular complexing agents. Several authors have studied the use of α-HIBA, glycolic acid, lactic acid, tartaric acid, succinic acid, citric acid, ethylenediaminetetraacetic acid, and tropolone as complexants for the separation of metal cations. In addition, to resolve the potassium and ammonium peaks from one another, 18-crown-6 is generally added to the IPD electrolyte.

The use of imidazole as carrier electrolyte with IPD was proposed initially by Beck and Engelhardt[25] for both high-mobility alkali and alkaline earth metals, and has since been used by a number of authors.[66–68] The electrolyte imidazole is a high-mobility cation that shows sufficient UV absorbance to permit sensitive IPD at 214 nm. An example of the separation that can be achieved with 5 mM imidazole adjusted to pH 4.5 with sulfuric acid is shown in Figure 13. It is apparent that the elution order of the solute ions (e.g. Ba^{2+}, Ca^{2+}, Mg^{2+}) is influenced by their ability to complex sulfate. Detection limits (ca. 0.1 ppm) and linearity of response (0.5–10 ppm) were established and the method was applied for routine analysis of inorganic cations in apple vinegar. Lin et al. have determined the relative merits of using different complexant (α-HIBA, acetate, lactate, malonate, succinate, citrate) with 5 mM imidazole for CZE separation of alkali and alkaline earth

metals.[66] The authors noted that monoprotic acids (acetate, lactate, α-HIBA) showed the weakest complexation with divalent cationic solute, whereas the triprotic acid citrate showed the strongest. Yang[67] and Francois et al.[68] have also investigated imidazole as an IPD electrolyte but used 18-crown-6 as the complexing agent. Both groups of researchers obtained maximum resolution of ammonium and potassium cations using this complexant. Quang and Khaledi optimized the concentration of complexant α-HIBA using a simple univariate approach.[69] The researchers developed a migration model to predict the elution time of 14 metal cations

Table 4 IPD in CE separation of inorganic cations

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Lithium, sodium, potassium, calcium, magnesium, barium	Standard mixture; apple vinegar	Imidazole	5 mM Imidazole, pH 4.5	214	25
Potassium, barium, calcium, sodium, magnesium, lithium	Standard mixture	Imidazole	5 mM imidazole and (a) 2.1 mM acetic acid, pH 6 (b) 4.5 mM glycolic acid, pH 4 (c) 5 mM lactic acid, pH 4 (d) 6.4 mM α-HIBA, pH 4 (e) 1.9 mM oxalic acid, pH 4 (f) 0.1 mM citric acid, pH 4 (g) 3.6 mM malic acid, pH 4	214	66
Ammonium, potassium, sodium, lithium, magnesium, strontium, barium, manganese, nickel, zinc, copper	Standard mixture; Chinese tea infusion	Imidazole	5 mM Imidazole, 6.5 mM α-HIBA–20% methanol, 0.53 mM 18-crown-6	214	67
Ammonium, potassium, calcium, sodium, magnesium, lithium	Standard mixture	Imidazole	10 mM Imidazole, 2.5 mM 18-crown-6, pH 4.5	214	68
Potassium, barium, strontium, calcium, sodium, magnesium, manganese, cadmium, iron, cobalt, lithium, nickel, zinc, copper	Standard mixture; tapwater	Imidazole; benzylamine	6 mM Imidazole (pH 3.95) or 6 mM benzylamine (pH 4), 12 mM α-HIBA	214 210	69
Potassium, calcium, sodium, magnesium, copper	Standard mixture; fruit juice; acid etching bath	4-Methylbenzylamine (MBA)	5 mM MBA, 6.5 mM α-HIBA, pH 4.4	214	70
Potassium, barium, strontium, calcium, sodium, magnesium, manganese, cadmium, iron, cobalt, lead, nickel, lithium, zinc, copper	Standard mixture; cough syrup; fermentation bath	MBA	5 mM MBA, 6.5 mM α-HIBA, pH 4.4	214	71
Rubidium, potassium, calcium, sodium, magnesium, lithium, lanthanum, cerium, praseodymium, samarium, neodymium, europium, gadolinium, terbium, dysprosium, holmium, erbium, thulium, ytterbium	Standard mixture	MBA	5 mM MBA, 4.0 mM α-HIBA, pH 4.4	214	71

Table 4 (*continued*)

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Potassium, barium, calcium, sodium, magnesium, manganese, iron, cobalt, nickel, copper, zinc, lithium, lanthanum, cerium, praseodymium, samarium, neodymium, europium, gadolinium, terbium, dysprosium, holmium, erbium, thulium, ytterbium	Standard mixture	*N*, *N*-Dimethyl-benzylamine (DBA)	6 mM DBA, 4 mM α-HIBA, pH 4.6 adjusted with acetic acid	214	73
Cesium, potassium, calcium, sodium, barium, strontium, magnesium, manganese, chromium, iron, cadmium, lithium, cobalt, nickel, lead, zinc, copper	Standard mixture	(a) Imidazole (b) Pyridine (c) Benzylamine	(a) 10 mM Imidazole, 13 mM glycolic acid, pH 4.0 (b) 10 mM Pyridine, 12 mM glycolic acid, pH 4.0 (c) 10 mM Benzylamine, 16 mM glycolic acid, pH 4.0	(a) 210 (b) 254 (c) 210	74
Potassium, sodium, barium, lithium, strontium, magnesium, calcium	Standard mixture	Pyridine	10 mM Pyridine, 0.8 mM ethylenediaminetetraacetic acid, pH 5	254	75
Ammonium, potassium, sodium, calcium, barium, magnesium, strontium, manganese, cadmium, iron, cobalt, nickel, zinc, lead	Standard mixture	MBA	7.5 mM MBA, 11 mM lactic acid, 2.6 mM 18-crown-6, 8% methanol, pH 4.3	214	76
Potassium, barium, strontium, calcium, magnesium, sodium, aluminum, copper, lithium, vanadium	Standard mixture	Nicotinamide	8 mM Nicotinamide, 0.6 mM 18-crown-6, pH 3.2 adjusted with formic acid	254	76
Potassium, sodium, calcium, magnesium	Orange juice; apple juice	Imidazole	5 mM Imidazole–sulfuric acid, pH 3 and pH 4.5	214	77
Hydronium, lithium, potassium, calcium, zinc, chromium, copper, aluminum	Standard mixture	Ephedrine	5.2 mM Ephedrine, 4.7 mM α-HIBA, pH 2.8	204	78
Potassium, barium, strontium, lithium, calcium, magnesium, sodium, rubidium, ammonium, cesium	Standard mixture; drinking water; rainwater; mineral water	Benzimadazole; 2,4,6-collidine	5 mM Benzimadazole or 5 mM collidine, 40 mM crown ether, 0.1% (w/v) hydroxyethylcellulose, 0.1% (w/v) hydroxymethylethyl-cellulose, pH 5.2	214	79

(*continued overleaf*)

Table 4 (*continued*)

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Ammonium, barium, potassium, sodium, calcium, magnesium, lithium	Standard mixture; drinking water	(a) $CuSO_4$ (b) Dimethyldiphenyl-phosphonium ion (DDP)	(a) 4 mM $CuSO_4$, 4 mM 18-crown-6, pH 3 (b) 5 mM DDP, 4 mM 18-crown-6, 5 mM morpholinoethanesulfonic acid, pH 6	215	80 81
Potassium, barium, lithium, calcium, sodium, magnesium, zinc, cadmium, chromium	Standard mixture	2-Aminopyridine (2-AP)	15 mM 2-AP, acetate buffer, pH 5	214	82
Potassium, calcium, ammonium, magnesium, manganese	Ultrex-grade hydrogen peroxide	MBA UV-Cat(II)	(a) 5 mM MBA, 6.5 mM α-HIBA, pH 4.4 (b) 1.2 mM UV-Cat(II), 6.5 mM α-HIBA, 3 mM tropolone, pH 4.4	185	83

as a function of pH and α-HIBA concentrations. The computer simulation predicted the optimal conditions to be 12 mM α-HIBA and pH 3.5. Again, 6 mM imidazole was used at the optimal pH and complexant concentration that provided a fast separation of metal cations in <4 min.

Monofunctional aromatic bases such as 4-methylbenzylamine,[70,72] *N*,*N*-dimethylbenzylamine,[73] benzylamine, and pyridine[73–75] have been shown to provide good CZE separation of mono-, di-, and trivalent cations. Weston et al.[70,71] showed that the simultaneous CZE separation of alkali, alkaline earth, and transition metals, as well as lanthanides, is possible using 5 mM MBA electrolyte and 6.5 mM α-HIBA at pH 4.4. Using this IPD electrolyte, they reported ca. 100 μg L^{-1} detection limits

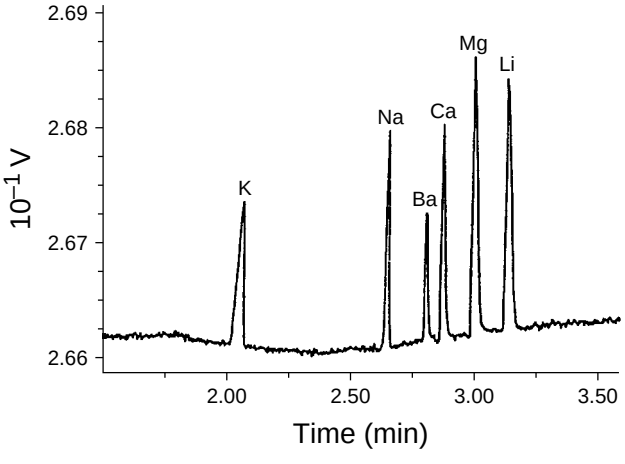

Figure 13 Separation of alkali and alkaline earth metal ions under optimum conditions: 5 mM imidazole, pH 4.5; 25 kV, 9 μA; sample concentration 1 ppm. (Reproduced with permission from Beck and Engelhardt.[25])

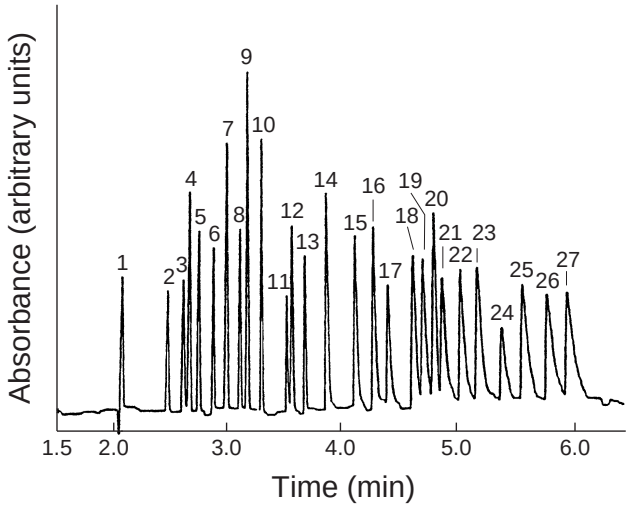

Figure 14 Separation of 27 inorganic cations with 4-methylbenzylamine/lactate electrolyte. Conditions: 8 mM 4-methylbenzylamine/15 mM lactic acid/5% methanol, pH 4.25; 75 μm i.d. × 60 cm capillary; 30-s injection; +30 kV; UV detection at 214 nm. Peaks: (1) potassium; (2) barium; (3) strontium; (4) sodium; (5) calcium; (6) magnesium; (7) manganese(II); (8) cadmium(II); (9) lithium; (10) cobalt(II); (11) lead(II); (12) nickel(II); (13) zinc(II); (14) lanthanum(III); (15) cerium(III); (16) praseodymium(III); (17) neodymium(III); (18) samarium(III); (19) gadolinium(III); (20) copper(II); (21) terbium(III); (22) dysprosium(III); (23) holmium(III); (24) erbium(III); (25) thulium(III); (26) ytterbium(III); (27) lutetium(III). Sample: 1–5 μg mL^{-1} of each cation. (Reproduced with permission from Shi and Fritz.[72])

for 28 metals. The peak capacity for cation separations in CZE is not as high as for anions. However, indications of improved peak capacity have been reported by Shi and Fritz.[72] Using 8 mM MBA, 15 mM lactic acid, and 5% methanol at a pH of 4.25, resolution of 27 metal cations

in a single run was achieved in only 6 min (Figure 14). In a separate paper Shi and Fritz were able to resolve potassium and ammonium ions when crown ether was added to the same IPD reagent.[76] Further addition of crown ether caused Sr^{2+} to move ahead of Ba^{2+} and allowed the resolution of these ions. The authors also suggested the use of 8% methanol with low applied voltage and low injection size as a possible solution to the problem of determining a trace amount of metal ions (e.g. 1 ppm Ca^{2+}, Mg^{2+}, Sr^{2+}, or Ba^{2+}) in the presence of a very high concentration of another metal ion (e.g. 1000 ppm of Na^{+}). On the other hand, Yang et al.[77] have advocated the use of an imidazole–sulfuric acid buffer system rather than MBA as a means of improving the peak shapes of sodium when injected at high concentration. Chen and Cassidy[73] optimized the CZE separation of metal cations in terms of the nature of the IPD reagent, the concentration of the complexant, the pH, and the surfactant type and concentration. Optimal conditions were 6 mM *N*,*N*-dimethylbenzylamine, 4.2 mM HIBA, and 0.2 mM Triton X-100 at pH 5.0 (adjusted with acetic acid). Under these conditions CZE separations of 26 metal cations was attained in <11 min. The detection limits were in the range 0.6–1 $\mu g\,mL^{-1}$ for lanthanide, transition and alkaline earth metal ions, and in the range 0.1–0.8 $\mu g\,mL^{-1}$ for alkali metal ions.

Several other IPD reagents have been described for CZE separation of inorganic cations. These include ephedrine,[74,78] benzimidazole,[79] copper(II),[80] diphenylphosphonium hydroxide,[81] 2-AP,[82] and guanidine · HCl.[83] In general, they have all shown good CZE separation of metal cations but not exceptionally high peak capacity. Of the various cationic IPD electrolytes described in this section, imidazole and MBA seem to provide the best overall performance in terms of selectivity, resolution, and peak capacity of inorganic cations. However, more work is still needed to improve the peak capacity for cation separations up to the level achieved for IPD of anions.

3.3 Indirect Photometric Detection of Short-chain Organic Acids and Amines

IPD is normally required for the analysis of most short-chain polar anions of organic acids (e.g. oxalate, malonate, formate, acetate, and propionate). This is because many of these organic acids do not fluoresce and have negligible absorbance at useful wavelengths in the

Table 5 IPD in CE separation of short-chain organic acids and amines

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Short-chain organic acids					
Formate, carbonate, methanesulfonate, acetate, chloroacetate, dichloroacetate, propionate, succinate, butyrate, benzoate, fumarate, glutarate, adipate, pimelate, suberate, azelate, sebacate, formate, oxalate	Standard mixture; air extracts (solid and liquid)	2,6-NDC	2 mM NDC, 5 mM NaOH, 0.5 mM TTAB, pH 11.0	280	84
Methanesulfonate, ethanesulfonate, propanesulfonate, butanesulfonate, pentanesulfonate, hexanesulfonate	Standard mixture	NDC	2 mM NDC, 5 mM NaOH, 0.5 mM TTAB, pH 8.0	280	84
Mesoxalic acid, pyruvic acid, glyoxylic acid, citraconic acid, mesaconic acid, citric acid, glutaconic acid, itaconic acid, 2-hydroxybutyric acid, acrylic acid, glutaric acid, methacrylic acid, acetic acid, crotonic acid, butyric acid	Standard mixture	Phthalate	5 mM Phthalate, 0.001% (w/v) polybrene, pH 3.85	185	88
Acetate, propionate, butyrate	Standard mixture	NDC	1 mM NDC, 3 mM β-cyclodextrin, pH 9.0	240	85

(*continued overleaf*)

Table 5 (*continued*)

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Oxalate, citrate, malonate, tartrate, methylmalonate, methylglutarate, lactate, ethylmalonate, adipate, methylsuccinate, pyruvate, succinate, glutarate	Standard mixture; urine	Phthalate	(a) 5 mM Carbonate, 1.5 mM phthalate, pH 7.0 (polyacrylamide-coated capillary) (b) 5 mM Carbonate, 1.5 mM phthalate, pH 9.0, 0.15 mM MTAB (c) 5 mM Carbonate, 1.5 mM phthalate, pH 8.9 (uncoated capillary)	230	86
Oxalate, formate, acetate, propionate, butyrate, benzoate, methanesulfonate	Standard mixture	(a) Chromate	(a) 5 mM Chromate, 0.5 mM TTAB, pH 8.0	(a) 254	87
		(b) PMA	(b) 2.25 mM PMA, 0.75 mM NaOH, 1.6 mM triethylamine, pH 7.7	(b) 254	
		(c) Phthalate	(c) 5 mM Phthalate or NDC, 0.5 mM TTAB, 1.0 mM H_3BO_3, pH 6	(c) 254	
		(d) NDC	(d) 2 mM NDC, 0.5 mM TTAB, 5 mM NaOH, pH 11	(d) 280	
Formate, malonate, citrate, succinate, pyruvate, acetate, lactate, pyroglutamate	Standard mixture; beverages	2,6-Pyridinedicarboxylic acid (PDA)	5 mM PDA, 0.5 mM CTAB, pH 5.6	200	89
Propionic acid, acetic acid, ascorbic acid, lactic acid, formic acid, citric acid, malic acid, tartaric acid	Standard mixture	8-Hydroxyquinoline-5-sulfonic acid (8-HQSA); 1,2,4,5-benzenetetracarboxylic acid (BTA)	5 mM HQSA or 3 mM BTA, pH 3	290	90
Oxalate, malonate, formate, succinate, maleate, tartrate, fumarate, malate, citrate	Standard mixture	NDS; NTS	4 mM NDS or NTS, 100 mM H_3BO_3, 5 mM $Na_2B_4O_7$, 2 mM DETA, pH 8	288, 284	35
Short-chain organic amines					
Diethylamine, triethylamine, diethanolamine, triethanolamine	Standard mixture	Imidazole	5 mM Imidazole, pH 4.25	214	25
Putrescine, L-histidine, spermidine, spermine, L-lysine, L-arginine	Standard mixture; tumor cells	Quinine sulfate	8 mM Quinine sulfate, 20% ethanol, pH 5.9	236	92
Ammonia, methylamine, 1,3-diaminopropane, putrescine, histamine, cadaverine, agmatine, ethylamine, ethanolamine, spermidine, propylamine, morpholine, isopropylamine, diethylamine, butylamine, spermine, isobutylamine, 1-methylbutylamine, hexylamine, tyramine, phenethylamine	Standard mixture; synthetic samples	$CuSO_4$	4 mM $CuSO_4$, 4 mM formic acid, 4 mM 18-crown-6	210	93

CTAB, cetyltrimethylammonium bromide; MTAB, myristoyltrimethylammonium bromide.

UV region. Various IPD electrolytes have been used in the analysis of organic acids. Again, the choice of the IPD reagent is governed by the need to obtain symmetrical peak shapes by closely matching their charge and electrophoretic mobility with those of the organic acids. Wu et al. evaluated chromate, pyromellitate, trimellitate, phthalate, terphthalate, and benzoate as IPD reagents for the determination of 10 organic acids.[61] Trimellitate was found to be the most suitable reagent because it provided the closest mobility match to the analyte with the highest molar absorptivity. In addition, various IPD reagents such as naphthalenedicarboxylic acid (NDC),[48,84,85] trimellitate,[47,60–62] phthalate,[20,39,86–88] NDS and NTS[35] all have suitable molar absorptivity and mobility matches that enable them to be of use for IPD of organic acids.

IPD has proved to be reliable for the quantitation of organic acids in real samples (e.g. urine,[86] beverages,[89] Bayer liquor[90]). However, the IPD method requires appropriate capillary conditioning procedures. For example, Soga and Ross found that when phthalate or trimellitate was employed as IPD reagent for the determination of organic acids, citrate and oxalate were not observed at concentrations below 50 mg L^{-1}.[89] Additionally, the S/N of malate deteriorated. The authors concluded that the decrease in S/N was due to adsorption of these organic acids onto the capillary wall. A procedure for rinsing the capillary with phosphoric acid before each run was proposed. However, a large phosphate peak appeared, which masked the analyte peaks. The use of a pyridinium chloride electrolyte was proposed as an alternative. This IPD electrolyte has the potential to complex with the metal impurities on the capillary wall, hence reducing the adsorption of the organic acids.

Table 5 lists the range of organic acids that have been analyzed employing CZE with IPD. Generally, the determinations of organic acids using the IPD method have been successful. Figure 15 shows an electropherogram obtained for the analysis of 18 organic acids in <7 min.[91] Although sensitivity in CZE is undoubtedly a major limitation, many IPD reagents (e.g. NDC, NDS) can be employed to lower the LOD.[84,85] The utility of CZE for the simultaneous measurement of short-chain organic acids and inorganic anions can be appreciated by considering the separation power and time as benefits of this technique. It should be noted that neither ion-exchange high-performance liquid chromatography (HPLC) nor ion-exclusion HPLC provides a complete separation of these compounds in IPD mode.

Even though there are a great number of samples that contain organic bases such as biogenic polyamines, food samples are the most common applications. Problems are often encountered with conductometric cation-exchange

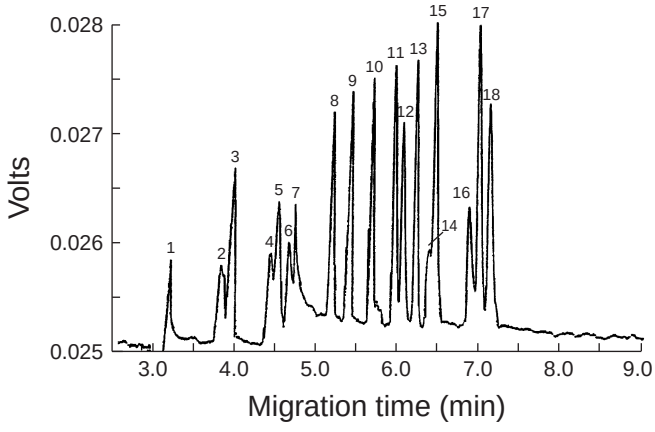

Figure 15 Capillary electropherogram of an 18-component organic acid mixture: (1) oxalic acid, (2) formic acid, (3) fumaric acid, (4) pyruvic acid, (5) maleic acid, (6) L-malic acid, (7) citric acid, (8) lactic acid, (9) succinic acid, (10) D,L-aspartic acid, (11) D-glucoronic acid, (12) D-gluconic acid, (13) acetic acid, (14) L-ascorbic acid, (15) shikimic acid, (16) gallic acid, (17) propionic acid, and (18) *n*-butyric acid. (Reproduced with permission from Craston and Saeed.[91] Copyright 1997 Elsevier Science N.V.)

HPLC for the analysis of short-chain polyamines because these polyprotonated species are highly charged in typical acidic eluents. Unless the eluent pH is high, the elution of polyamines using cation-exchange chromatography is problematic. For these reasons, CZE with IPD is attractive because it affords a simple means of facilitating both separation and detection. The potential of CZE with IPD for the determination of biogenic amines has been recognized by several authors. One of the early reports in the IPD of amines is by Beck and Engelhardt.[25] Simultaneous separation of metal ions, amines, and amino alcohols was obtained using 5 mM imidazole as the IPD electrolyte at a pH of 4.5. A complete CZE separation was achieved in ca. 4 min, with a C_{LOD} of around 0.4 ppm. Some polyamines (e.g. putrescine, spermine, and spermidine) along with inorganic cations and basic amino acids in tumor cells were determined simultaneously in <10 min by Ma et al.[92] The system employed quinine sulfate as IPD reagent at a pH of 5.9. Mass detection limits attained by this method were in the femtomole range with a linear detection response of two orders of magnitude. Extensive optimizations by Arce et al.[93] have shown that cupric sulfate is an ideal IPD electrolyte for the detection of biogenic amines. A rapid separation of 21 amines using this optimized electrolyte system is shown in Figure 16. A typical detection limit attained by this method for biogenic amines was 0.05 ppm. Although the linear concentration range (0.05–10 ppm) was rather narrow, the recoveries (90–110%) obtained for the representative amines were reasonable.

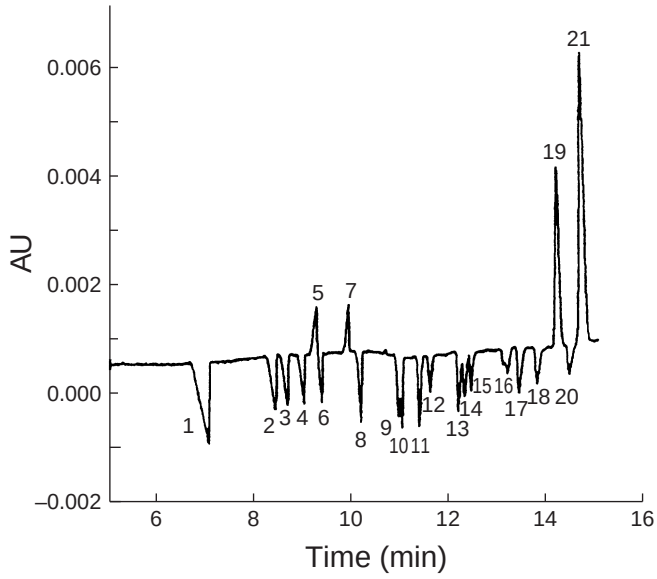

Figure 16 Electropherogram of a standard mixture of amines: (1) ammonia; (2) methylamine; (3) 1,3-diaminopropane; (4) putrescine; (5) histamine; (6) cadaverine; (7) agmatine; (8) ethylamine; (9) ethanolamine; (10) spermidine; (11) propylamine and morpholine; (12) isopropylamine; (13) diethylamine; (14) butylamine; (15) spermine and isobutylamine; (16) amylamine and isoamylamine; (17) 1-methylbutylamine; (18) hexylamine; (19) phenylethylamine; (20) heptylamine; (21) tyramine. The electrolyte comprised 4 mM cupric sulfate, 4 mM formic acid and 4 mM 18-crown-6 at pH 3.0. The separation was performed at 10 kV. (Reproduced with permission from Arce et al.[93] Copyright 1997 Friedr. Vieweg & Sohn.)

3.4 Indirect Photometric Detection of Long-chain Fatty Acids

Long-chain free fatty acids (FFAs) are an important class of naturally occurring compounds that can be found in living cells. Because these compounds contain an acidic hydrogen, they predominantly exist as anions in basic solution. Considering the differences in mass-to-charge ratios, both saturated and unsaturated FFAs can be separated by CZE. However, saturated FFAs exhibit only a weak absorption band in the region of 200 nm. Therefore, CZE with direct UV detection is problematic and leads to limited sensitivity. Currently, CZE with IPD is of considerable interest for the determination of long-chain saturated FFAs (Table 6).[45,94–98]

Probably the first attempt to separate long-chain FFAs in CZE with IPD was reported in 1994 by Gutnikow et al.[94] Both normal and reversed electroosmotic flow in CZE was developed using TNBSA as the IPD reagent. Although, a rapid separation in <7 min of FFAs up to C_{10} was achieved in reversed electroosmotic flow mode using 30% acetone, peak capacity was much higher in the normal mode. For example, in the normal mode, analysis of C_3-C_{18} organic acids was feasible in <20 min using 60% acetonitrile. Despite the use of low-mobility TNBSA as the IPD reagent, the electropherogram showed asymmetric peaks for $C_{14}-C_{18}$ FFAs. Buchberger et al.[95] resolved both short- and long-chain FFAs in a reversed CZE mode with reversed polarity. The solubility parameters and adsorption of FFAs on capillary walls were overcome with the use of 70% ethyleneglycol monomethyl ether and trimethylammonium propanesulfonate in the IPD buffer. The FFAs in butter were profiled; however, one of the limitations of this method was its inability to resolve isomers of FFAs. Nevertheless, the simplicity of the sample preparation could be beneficial for routine analysis.

Roldan-Assad and Gariel developed a CZE method for the determination of C_2-C_{14} linear FFAs.[45] Because long-chain FFAs ($\geq C_{14}$) have limited water solubility and their tendency to form micelles decreases the separation selectivity between successive homologs, IPD electrolytes containing cyclodextrins and 60% methanol were used to enhance both solubility and selectivity. However, this occurred at the expense of a slightly longer analysis time (>16 min). The C_{LOD} values of FFAs were approximately 0.2 ppm (1×10^{-6} mol L^{-1}) and the detection linearity was extended over three orders of magnitude (10^{-6}–10^{-3} mol L^{-1}). Collet and Gariel[96] later studied the effect of cyclodextrin type and concentration for the separation and indirect detection of saturated and unsaturated FFAs ranging between C_{14} and C_{18}. The authors concluded that the CZE method developed using cyclodextrin was not suitable for the separation and IPD of very-long-chain ($C_{19}-C_{31}$) FFAs.

Drange et al.[97] first used nonaqueous electrolyte for CZE separation of very-long-chain FFAs. The authors showed that the separation of $C_{14}-C_{26}$ acids could be accomplished conveniently in 15 min using ANTs in *N*-methylformamide and dioxane. Recently, Haddadian et al. investigated the CZE separation of long-chain saturated $C_{12}-C_{31}$ FFAs differing by only one carbon atom.[98] Figure 17 shows an impressive separation of $C_{12}-C_{31}$ FFAs under the optimum conditions using AMP as IPD reagent. Because AMP was found to be soluble in aqueous and partially aqueous, as well as nonaqueous media, it appears to be an excellent IPD electrolyte for the separation and IPD of complicated mixtures of both saturated and unsaturated FFA isomers (Figure 18).

3.5 Indirect Photometric Detection of Phosphates and Phosphonates

CZE with IPD has recently emerged as a valid alternative to the more traditional postcolumn IC method for the determination of both phosphates and phosphonates (Table 7).[24,99–101] Advantages of CZE over IC for these

Table 6 IPD in CE separation of long-chain free fatty acids

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
C_1-C_{10} FFAs	Standard mixture; fat hydrolysate, butter, palm oil	3,5-Dinitro-benzenesulfonic acid (DNBSA)	(a) 10 mM DNBSA, 0.5 mM CTAB, 30% acetone, pH 9.0	214	94
C_3-C_{18} FFAs	Standard mixture	Trinitrobenzene-sulfonic acid (TNBSA)	(b) 10 mM TNBSA, 60% acetonitrile, pH 9.0	254	94
C_2-C_{18} FFAs	Standard mixture; butter	Diethylbarbiturate	5 mM Diethylbarbiturate, 0.5 M *Z*-1 methyl, 70% Ethyleneglycol monomethyl ether	240	95
C_2-C_{14} linear FFAs	Standard mixture; coco oil extract	(a) *p*-Anisate	(a) 10 mM *p*-Anisate, 20 mM Tris, 0.75 mM trimethyl-β-cyclodextrin, pH 8.2	264	45
C_7-C_{18} FFAs	Standard mixture; coco oil extract	(b) *p*-Anisate	(b) 10 mM *p*-Anisate, 20 mM Tris, 1 mM trimethyl-β-cyclodextrin, 60% methanol	264	45
$C_{14}-C_{26}$ FFAs	Standard mixture; hydrogenated fish oil	Anthraquinone-2-carboxylic acid (ANT)	(a) 2.5 mM ANT, 40 mM Tris, 3 : 1 (v/v) NMF–dioxane	264	97
$C_{12}-C_{31}$ FFAs	Standard mixture	(a) AMP	(a) 2.5 mM AMP, 40 mM Tris, 3 : 1 v/v NMF–dioxane	259	98
Isomers of FFAs: $C_{14:1}$, $C_{16:1}$, $C_{18:1}$, $C_{18:2}$, $C_{18:3}$, $C_{20:1}$, $C_{20:2}$, $C_{20:3}$, $C_{22:1}$, $C_{22:2}$, $C_{22:3}$	Standard mixture	(b) AMP	(b) 2.5 mM AMP, 40 mM Tris, 5 : 4 : 1 (by vol.) NMF–dioxane–water	259	98

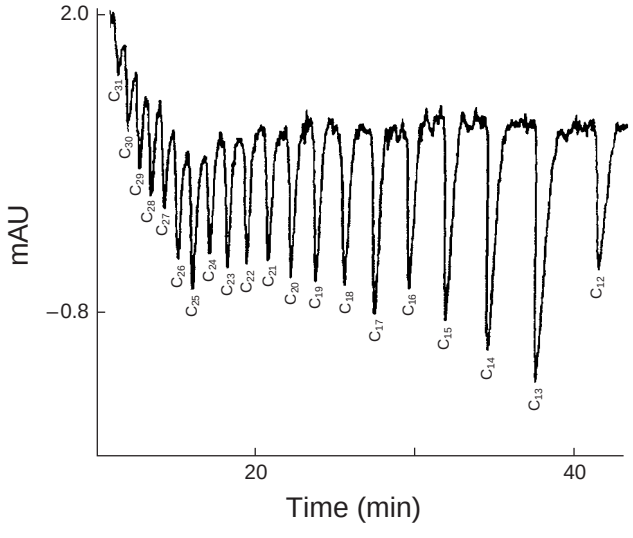

Figure 17 Electropherogram of saturated FFAs ($C_{12}-C_{31}$). Electrolyte: 40 mM Tris, 3 : 3 NMF–dioxane, 2.5 mM AMP, 0.5% (w/v) Brij. Hydrodynamic injection for 1 s; +20 kV applied voltage; IPD at 259 nm; capillary temperature 40 °C; analyte concentration is 30–35 ppm. (Reproduced with permission from Haddadian et al.[98])

compounds are large peak capacity, short analysis time, better efficiency, and minimal sample and reagent consumption. Despite the fact that Jones and Jandik demonstrated chromate to be the best IPD electrolyte for the separation of 30 inorganic anions and short-chain organic acids in ca. 3.0 min, the utility of chromate for the separation of relatively high-molecular-weight and less-mobile phosphates and phosphonates is poor. For example, peak tailing and poor migration time and area reproducibility of orthophosphate were observed by a number of researchers using a chromate electrolyte.[102–105] Theoretical considerations suggest that reagents that can be useful for indirect detection of phosphate and phosponate derivatives should have medium to low mobility, a relatively high ε, and a favorable DR.

Linear polymers of phosphoric acid, all of which contain phosphorus–oxygen linkages, are termed linear P_n. The first report on CZE/IPD applied to the separation of P_1-P_3 phosphates using a phthalate electrolyte at a pH of 4.0 was described by Stover et al.[99] However, the system was not fully optimized and the method was reported to have marginal baseline stability and

Table 7 IPD in CE separation of phosphates and phosphonates

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
P_1-P_3 phosphates	Standard mixture; potato bath	Phthalate	5 mM Phthalate, 0.5 mM DTAB, pH 4.2	250	99
P_1-P_4 phosphates	Standard mixture; toothpaste; soap	AMP	5 mM AMP, 100 mM H_3BO_3/2 mM $MgCl_2$, pH 7.10	259	24
P_1 phosphate, cyclic P_3, P_2 phosphate, P_1 phosphite, ATMP, P_1 hypophospite, P_1 fluorophosphate, cyclic P_6, DETPMP, EDTMP, HDTMP	Standard mixture	AMP, GMP, CMP, UMP	5 mM AMP, 100 mM H_3BO_3/5 mM $Na_2B_4O_7$, 2 mM DETA, pH 7.8	259	24
P_1-P_3 phosphate, P_1 fluorophosphate, cyclic P_3	Standard mixture; toothpaste	ATP	5 mM ATP, 0.02 mM CTAB, pH 3.5	260	105
Longer-chain P_n, chain length 1–30	Standard mixture	PMA	2.25 mM PMA, 6.5 mM NaOH, 0.75 mM HMOH, 1.6 mM triethanolamine pH 7.7	254	106
P_2-P_4 phosphate, cyclic P_3, glass phosphates	Standard mixture	PMA	5.0 mM PMA, 20 mM Tris, pH 7.2, polyacrylamide-coated capillary	254	107
Inositol mono-, bis-, tris-, hexakisphosphates	Standard mixture	Chromate	2.5 mM K_2CrO_4, 0.5 mM TTAB, 5 mM H_3BO_3, pH 7.3	270	108
Inositol 1- and 2-monophosphates	Standard mixture	Phthalate	5 mM Phthalate, 0.5 mM TTAB, 5 mM $Na_2B_4O_7$, pH 7.3	254	108
C_1-C_6 mono- and diesters of phosphates	Standard mixture	AMP	5 mM AMP, 100 mM H_3BO_3, pH 7.2	259	109
C_1-C_4 phosphonates	Standard mixture	PPA	10 mM PPA, 100 mM borate buffer, pH 6.0	254	110
Glyphosate, AMPA	Standard mixture; herbicide	AMP	5 mM AMP, 100 mM H_3BO_3, pH 7.10	259	24
Glyphosate, AMPA	Standard mixture	Phthalate	10 mM Phthalate, 0.5 mM TTAB, pH 7.5	240	111
O-(2-Ethylhexyl)-methylphosphonic acid, methylphosphonate, isopropylmethylphosphonate, isobutylmethylphosphonate, *O*-isopropylethylphosphonic acid, cyclohexylmethylphosphonate, pinacolylmethylphosphonate	Standard mixture	PPA	10 mM PPA, 200 mM borate, 0.35 mM dodecyldimethylammonium bromide, 0.03 wt.% Triton X-100, 40 °C, pH 4.0	210	101

CMP, cytidine 5′-monophosphate; GMP, guanosine 5′-monophosphate; UMP, uridine 5′-monophosphate; ATP, adenosine 5′-triphosphate; PPA, phenylphosphonic acid; AMPA, aminomethylphosphonic acid; DTAB, dodecyltrimethylammonium bromide; HMOH, hexamethonium hydroxide.

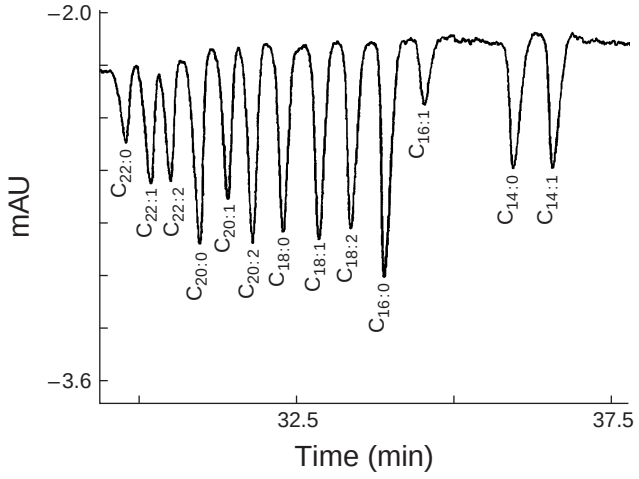

Figure 18 Electropherogram of saturated and unsaturated FFAs ($C_{14}-C_{22}$). Electrolyte: 40 mM Tris, 5:4:1 NMF–dioxane–water, 2.5 mM AMP. Hydrodynamic injection for 1 s; +20 kV applied voltage; IPD at 259 nm; capillary temperature 20 °C; analyte concentration is 30–45 ppm. (Reproduced with permission from Haddadian et al.[98])

reproducibility. When the calibration plots of normalized peak area vs. concentration were drawn for P_1, P_2 and P_3 phosphates, approximately the same slopes but different intercepts were observed. For example, P_2 and P_3 phosphates provided nearly equal slope values but large negative intercepts. The authors concluded that the reason for the low detection sensitivity of P_2 and P_3 phosphates might include interaction with the surfactant (e.g. dodecyltrimethylammonium bromide used as a modifier for reverse electroosmotic flow) and/or on-line hydrolysis. In an attempt to improve the reproducibility, Shamsi and Danielson first evaluated the use of ribonucleotides, namely AMP, CMP, GMP, and UMP, as IPD electrolytes for the simultaneous separation of 12 P_n and RP_n in <30 min.[24] These reagents contain a phosphate group in their structure, thus preventing any irreversible binding of P_n and RP_n with the silanol groups. Even though all of the ribonucleotides were able to separate a mixture of P_n and RP_n with reasonable selectivity, AMP was found to be the best electrolyte because of the its high ε, large DR and favorable TR required for sensitive IPD. The quality of separation of some P_n (particularly P_3 and P_4 phosphates) was dramatically improved by incorporating Mg^{2+} in the running buffer with positive polarity. As a result, the reproducibility of migration times is 0.5–1.0% relative standard deviation (RSD). One practical application reported was the determination of phosphates and phosphonates in soap and various brands of toothpaste. Later Wang and Li[100,105] reported the use of ATP at pH 3.5 for the CZE separation and IPD of cyclic P_3 phosphate, monofluorophosphate and P_1-P_3 phosphates. Although the value of DR for ATP was 2.8×10^3, which is much greater than that obtained by Shamsi et al.[24] using AMP (DR = 612), the detection limits for P_n were not improved. This implies that the greater DR obtained for ATP was compensated by a loss in TR.

The CZE separation of P_n with chain length ≥ 4 is more challenging because of their similar electrophoretic mobility. Again, Stover first explored the buffer containing PMA with hexamethonium hydroxide for high-efficiency separation and IPD of P_n.[106] Glassy sodium P_n with average chain lengths of 5–44 were analyzed in less than 20 min. Stover thought that the cationic electroosmotic flow modifier hexamethonium would ion-pair with the anionic P_n leading to improved CZE separation. After this success in the CZE mode, Wang and Li recently reported the separation of higher-chain-length P_n using capillary gel electrophoresis (CGE).[107] Presumably, the sieving effect of the gel provided good resolution of P_n. Wang and Li used the same PMA electrolyte as Stover used,[106] but without the use of hexamethonium hydroxide. The migration order of linear P_n was found to be based on charge-to-size ratios. For example, migration time first increased from P_2 to P_4 phosphate, but P_1 phosphate eluted later, followed by all linear P_n. The authors reported that quantitative analysis of P_n was difficult due to the bias generated by the electrokinetic injection. Currently, systematic approaches to the design of a novel buffer system and IPD reagents that offer enhanced selectivity and sensitivity for long-chain P_n are still being developed. As yet, they are not at the point where buffer recipes can be employed conveniently for routine separations of these long-chain P_n in CZE or CGE mode. However, there is no doubt that CE methods for P_n will substitute for IC in specific applications.

IPD of inositol phosphates after their separation by CZE was studied by Henshall et al.[39] at low pH using a dichromate and phthalate electrolyte. Although the use of chromate provided separation of all four inositol (mono-, bis-, tris-, and hexakis-) phosphates in less than 10 min, the separation window was rather narrow (0.7 min) and therefore shifts in the migration time resulted in peak misidentification. On the other hand, phthalate when buffered at pH 5.9 provided acceptable results with excellent sensitivity for the inositol mono- and bis-phosphate, but this electrolyte does not seem to be useful for the separation of all four inositol phosphates in a single run. The CZE/IPD methodology for the determination of six inositol phosphates in fermentation broth was also investigated by Buscher et al.[108] Using 1-naphthol-3,6-disulfonic acid (NDSA) as the IPD reagent, the authors reported the baseline resolution of inositol mono-, bis-, tris-, and tetrakisphosphates, whereas inositol pentakis- and hexakisphosphates remain unresolved.

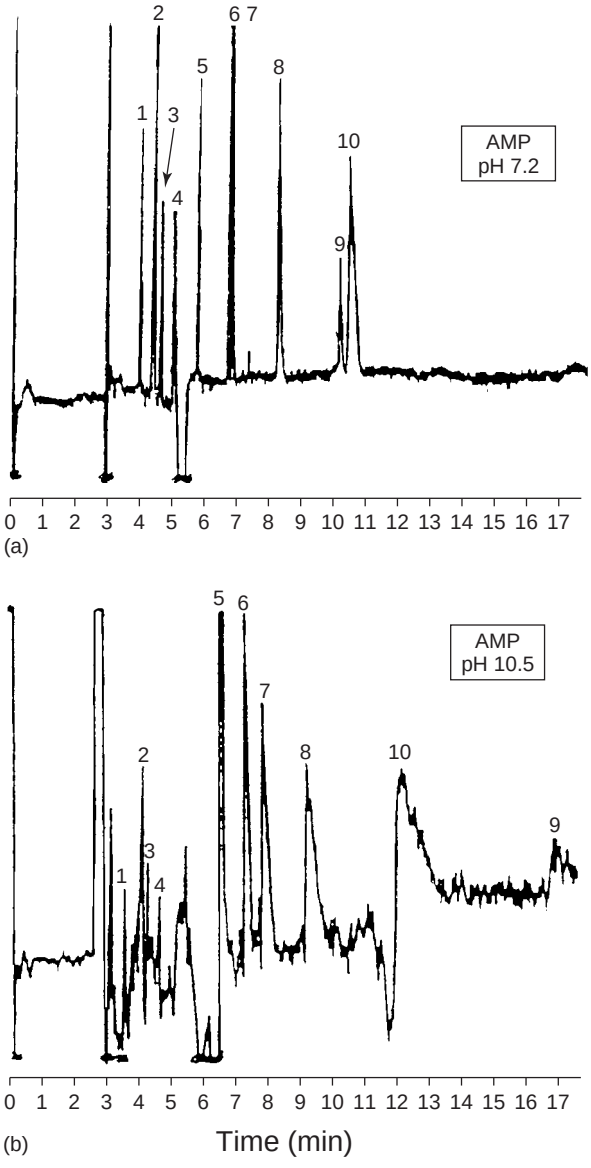

Figure 19 Effect of pH on the separation of a standard mixture of 10 mono- and diesters of phosphates: (a) AMP at pH 7.2; (b) AMP at pH 10.5. All electrolytes contained 100 mM H_3BO_3. Peak identification: 100 mg mL^{-1} each of (1) diethylhexylphosphate; (2) dibutylphosphate; (3) diisopropylphosphate; (4) diethylphosphate; (5) monoethylhexylphosphate; (6) monobutylphosphate; (7) monoisopropylphosphate; (8) monoethylphosphate; (9) monomethylphosphate; (10) dimethylphosphate. Vacuum injection for 2 s; +30 kV applied for separation; current is 8 μA and 70 μA at pH 7.2 and 10.5, respectively; IPD at 259 nm. (Reproduced with permission from Shamsi and Danielson.[109])

Inositol trisphosphate was the only compound that provided a closer mobility match to NDSA, resulting in symmetrical peak shape. Hence, this approach was extended further by the same authors for the analysis of inositol trisphosphate in plasma.[41]

Phosphorus-containing surfactants, namely mono- and diesters of aliphatic organophosphates, are used in a wide variety of applications. The determination of these compounds is of relevance not only in the detergent industry, owing to their surface-active property, but also in medicinal research, as well as the nuclear and fuel processing industry. Recently, Shamsi et al.[109] developed a CZE/IPD method for analysis of alkyl phosphates. An example electropherogram showed that the separation of 10 C_1-C_6 mono- and diesters of phosphate is possible in ca. 11 min using AMP electrolyte at pH 7.2 (Figure 19a). Although the resolutions of some phosphate esters were improved at a higher pH value of 10.5, the S/N decreased significantly under such conditions (Figure 19b). Hence, a pH of 7.2 was recommended for the analysis of these aliphatic organophosphates.

Alkylphosphonic acids have considerable agricultural and biological importance due to their biocidal potency and their presence at trace levels in various antibacterial (fosfomycin) or antiviral (foscarnet) formulations. Pianetti et al. have utilized PPA at 200 nm for the IPD of C_1-C_4 phosphonates in less than 10 min.[110] The authors reported that an increase in borate concentration resulted in a better S/N. In contrast, S/N decreased with increasing concentration of PPA. The calibration plots based on peak height for C_1-C_4 phosphonates showed a linear range of 5–50 μg mL^{-1} with an RSD of less than 3.8%. Moreover, the mass LOD of alkylphosphonic acids was

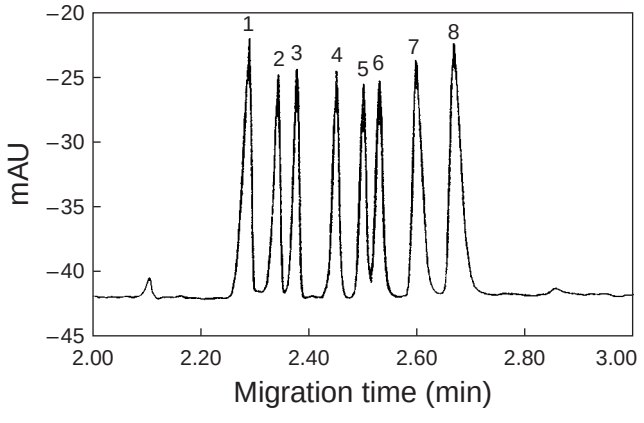

Figure 20 Separation of alkylphosphonic acids. Experimental conditions: buffer is 200 mM borate and 10 mM PPA; 75 μm (i.d.) × 56 cm fused-silica capillary; peaks detected by indirect UV detection at 210 nm; injection for 6 s at 50 mbar; temperature = 40 °C; 0.35 mM didodecyldimethylammonium hydroxide and 0.03 wt.% Triton X-100, pH 4.0; voltage = 30 kV (negative polarity). Anions: (1) *O*-(2-ethylhexyl)methylphosphonic acid; (2) methylphosphonate; (3) ethylmethylphosphonate; (4) isopropylmethylphosphonate; (5) isobutylmethylphosphonate; (6) *O*-isopropylethylphosphonic acid; (7) cyclohexylmethylphosphonate; (8) pinacolylmethylphosphonate. (Reproduced with permission from Nassar et al.[101] Copyright 1998 American Chemical Society.)

ca. 0.21 pmol. Another application of phosphonate compounds is glyphosate, with its major metabolite AMPA. Shamsi et al.[24] also developed a normal-polarity CE method using AMP electrolyte for quantitation of these two analytes in a commercial herbicide solution. Likewise, Cikalo et al.[111] also reported the analysis of glyphosate and AMPA using CZE/IPD. However, these employed a reverse-polarity CE with phthalate as the IPD reagent. Although the use of electrokinetic injection provided improved detection limits, the application of the method to wheat samples was unsuccessful. This is because of the high conductivity of the extracted sample, which prevented injection in the electrokinetic mode. Very recently, Nassar et al.[101,112] in a series of two publications used PPA as the IPD electrolyte for the separation and indirect detection of short-chain alkylphosphonates (also known as chemical warfare agent degradation products). The phosphonates studied included methylphosphonic acid, monoisopropylphosphonic acid, and the monoalkyl esters of methylphosphonic acid. Using a mixed surfactant system (0.35 mM didodecyltrimethyl ammonium hydroxide, 0.03 wt.% Triton X-100) to reverse the electroosmotic flow, and PPA as the IPD electrolyte, complete separation of phosphonates was accomplished in less than

Table 8 IPD in CE separation of anionic and cationic surfactants

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
C_4-C_{12} SO_4^-, C_1-C_{14} SO_3^-, alkyl ether sulfates	Standard mixture	NMS, NDS	5 mM NMS or NDS, 100 mM H_3BO_3, 5 mM $Na_2B_4O_7$, pH 8	274, 288	35
C_4-C_{12} SO_3^-	Standard mixture; shampoo	NMS	10 mM NMS, 30% acetonitrile, pH 10	274	7
C_4-C_{18} SO_4^-, C_4-C_{18} SO_3^-	Standard mixture	NMS; PTS	5 mM NMS or PTS, 100 mM H_3BO_3, 50% methanol, pH 6.0	206, 221	113
C_4-C_{18} SO_3^-, C_6-C_{18} sarcosines	Standard mixture; commercial formulations	ATP	5 mM ATP, 100 mM H_3BO_3, 50% methanol, pH 6.0	206	114
C_4-C_{12} SO_3^-	Standard mixture	Salicylate	5 mM Salicylate, 5 mM phosphate, 1.0 mM $MgCl_2$, pH 7.0	230	116
C_2-C_{12} SO_4^-	Standard mixture; commercial formulations	Veronal	12 mM Veronal buffer, pH 8.6	240	19
C_4-C_{18} SO_3^-, C_9-C_{13} SO_4^-, $C_{12}-C_{14}$ alkyl ethoxysulfate, $n = 2^\circ$ of ethoxylation	Standard mixture; household detergents	NMS; octylbenzene-sulfonate (OBS); dodecylbenzene-sulfonate (DBS)	10 mM NMS, OBS or DBS, 100% methanol	214	117
C_2-C_{16} SO_3^-, C_8-C_{18} SO_4^-	Standard mixture	*p*-Toluenesulfonic acid (PTSA)	5 mM PTSA	214	118
TEA, ClETMA, TBA, THA, DDMA	Standard mixture	Benzylamine; ephedrine; tetrazolium violet	5 mM Benzylamine, ephedrine or tetrazolium violet, 100 mM H_3BO_3, 50% methanol, pH 6.0	204, 300	113, 114
$C_{12}-C_{18}$ alkyldimethyl-ammonium quaternary compounds	Standard mixture; disinfectant product	Benzyltrimethyl-ammonium chloride (BTMACl)	3 mM BTMACl, 8 mM monobasic sodium phosphate, 3 mM SDS, tetrahydrofuran–water (57.5 : 42.5)	210	120
$C_{12}-C_{18}$ trimethylammonium chlorides	Standard mixture	BTMACl	5 mM BTMACl, 50% acetonitrile, 5 mM BTMACl, 3 mM SDS, 50 mM phosphate buffer, pH 6.8	214, 214	117, 121

PTS, *p*-toluene sulfonate; SDS, sodium dodecylsulfate; TEA, tetraethylammonium; ClETMA, chloroethyltrimethylammonium; TBA, tetrabutylammonium; THA, tetrahexylammonium; DDMA, didodecyldimethylammonium.

3 min (Figure 20).[101] The authors found this electrolyte system to be stable for at least three months. The detection limit reported was about 100 μg L^{-1}. Moreover, the method was applied to the aqueous leachates of soil, wipes of surfaces, and vegetation samples known to have been exposed to nerve agents. Nassar et al.[112] in the same laboratory expanded this work with quantitative analysis of chemical warfare degradation products in reaction masses. They found that the CZE/IPD methodology can provide a limit of quantitation of 500 μg L^{-1} and an LOD of 100 μg L^{-1} for both isopropylmethylphosphonic acid and pinacolylmethylphosphonic acid, with a linear response over the range 0.5–100 μg L^{-1}.

3.6 Indirect Photometric Detection of Aliphatic Anionic and Cationic Surfactants

Aliphatic anionic and cationic surfactants are important additives that are widely used as cleaning agents, dispersants, emulsifiers, solubilizers, and stabilizers. Structurally, many of these compounds consist of an aliphatic hydrocarbon tail with a polar head group and nonchromophoric substituent. Recent reports on the separation of nonchromophoric anionic surfactants indicate that CE is a complementary technique to reversed-phase IC with respect to retention order.[35,113,114] In addition, minimal organic solvent consumption and inexpensive column replacement are practical benefits of CE besides its inherent good peak capacity.

In the analysis of anionic and cationic surfactants (Table 8), a medium- to low-mobility IPD co-ion must be chosen. For the determination of nonchromophoric anionic surfactants such as alkyl sulfates, alkane sulfonates, alkylether sulfates, and alkylsulfosuccinates, several IPD reagents have been explored. Numerous IPD electrolytes used for anionic surfactants include veronal, NMS,[7,35,113–115] salicylate,[116] dihydroxybenzoic acid,[117] PTS,[113,118] dinitrobenzoic acid,[119] AMP,[114] and octyl- and dodecylbenzene sulfonate.[115] In contrast, there are only a few reports on the CZE separation of non-UV-active cationic surfactants. The strong adsorption of these substances at the inner wall of a fused-silica capillary is clearly a matter of concern. However, recent reports have indicated that a large volume of organic solvents can help to overcome this problem.[120,121] By far the more prominent IPD reagents for cationic surfactants include benzyldimethyldodecylammonium,[120] benzylamine,[114] ephedrine,[113,114] benzyltrimethylammonium,[121] and tetrazolium violet.[114]

The separation of complex mixtures of alkyl sulfates and alkane sulfonates of identical chain length is a challenging problem in CZE because of their very similar electrophoretic mobilities. Figure 21 is a typical electropherogram obtained using the best anionic IPD reagent, such as NMS with 50% methanol, to enable the baseline separation of C_6–C_{18} SO_4^-/SO_3^- pairs in 30 min. A similar electropherogram but with lower sensitivity could be generated with PTS as the IPD reagent.

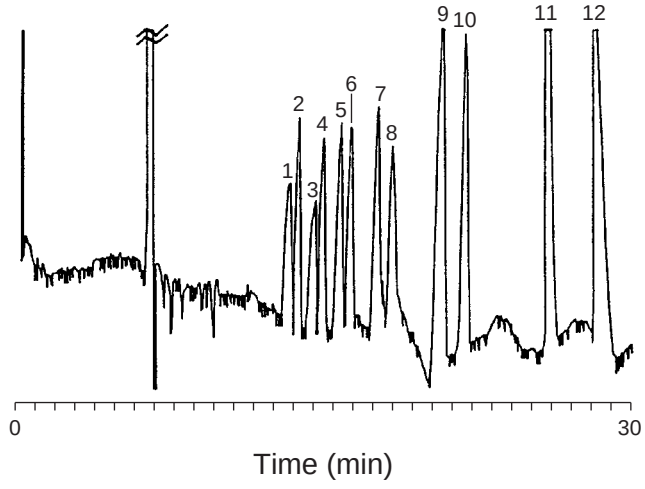

Figure 21 Separation of a standard mixture of anionic sulfonated and sulfated surfactants. Electrolyte consisted of 5 mM NMS with 100 mM H_3BO_3 (pH 6.0) and 50% methanol. Peak identification: 100 mg mL^{-1} each of (1) C_{18} SO_3^-, (2) C_{18} SO_4^-, (3) C_{14} SO_3^-, (4) C_{14} SO_4^-, (5) C_{12} SO_3^-, (6) C_{12} SO_4^-, (7) C_{10} SO_3^-, (8) C_{10} SO_4^-, (9) C_8 SO_3^-, (10) C_8 SO_4^-, (11) C_6 SO_3^-, and (12) C_6 SO_4^-. Vacuum injection for 3.5 s; +30 kV applied for separation; current = 2–3 μA; IPD at 206 nm. (Reproduced with permission from Shamsi and Danielson.[113] Copyright 1995 American Chemical Society.)

3.7 Indirect Photometric Detection of Carbohydrates

The determination of carbohydrates (Table 9) provides a unique challenge to a separation chemist. This is because most carbohydrates do not possess strong chromophores in their structure and are therefore not accessible to direct detection in the UV. For these reasons, carbohydrates are generally derivatized with a suitable chromophore such as 2-AP[122] or *p*-aminobenzoic acid[123] for direct UV detection, or by indirect detection[23,123] using a suitable light-absorbing reagent. It is well known that carbohydrates are very weak acids; therefore very-high-pH eluents have been used for their separation via anion-exchange chromatography.[124,125] A number of researchers have also utilized high-pH (pH ~ 12) electrolytes for the CZE separation of carbohydrates in the IPD mode.

Vorndran et al.[23,51] have shown that sorbic acid is a very useful IPD reagent for several important reasons. First, sorbic acid has high molar absorptivity (ε = 27 800 at 256 nm). Second, it carries a single charge, which ensures a favorable TR value. Third, it does not interact with the

Table 9 IPD in CE separation of carbohydrates

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Raffinose, 2-deoxy-D-ribose, galactose, glucose, rhamnose, mannose, *N*-acetylneuraminic acid, gluconic acid, galacturonic acid, mannuronic acid	Standard mixture	Sorbate	6 mM Sorbate, 30 °C, pH 12.1	256	123
Fucose, galactose, glucose, *N*-acetylgalactosamine, *N*-acetylneuraminic acid	Standard mixture; fruit juices	Sorbate	6 mM Sorbate, 15 °C, pH 12.2	256	126
Sucrose, lactose, maltose, glucose, rhamnose, fructose, ribose, gluconic acid	Standard mixture	NAA	2 mM NAA, pH 12.2, 25 °C	222	127
Sucrose, maltose, glucose, fructose	Standard mixture; culture medium	Sorbate; riboflavin; dimethoxycinnamic acid	12 mM Sorbate, riboflavin or dimethoxycinnamic acid, 63 mM NaOH or LiOH	256, 267, 310	129
Glucose, rhamnose, mannose, fructose	Standard mixture	Sorbate; riboflavin	12 mM Sorbate or riboflavin, 5% diethylamine, pH 12.6	256, 267	130
Sucrose, trehalose, raffinose, maltotriose	Standard mixture	Riboflavin	12 mM Riboflavin, pH 11.8	267	130
Glucose-6-phosphate, fructose-1,6-biphosphate, dihydroxyacetone phosphate, fructose-6-phosphate, 2-phosphoglycerate, glyceraldehyde-3-phosphate, ribose-5-phosphate, sucrose-6-phosphate	Standard mixture of glycolytic pathway	AMP; NDS	5 mM AMP or NDS, 100 mM H_3BO_3, pH 7.2	259, 288	136
Heparin, disaccharides, glycosaminoglycans	Standard mixture	5 mM Tricarboxybenzoic acid; sulfosalicylic acid	5 mM Tricarboxybenzoic acid or sulfosalicylic acid, pH 3.5	214	131
α-Cyclodextrin, β-cyclodextrin, γ-cyclodextrin	Standard mixture	Benzoic acid	30 mM Benzoic acid titrated with Tris to pH 6.2	214	132

NAA, 1-naphthylacetic acid.

capillary surface or with the analyte. Fourth, it has an effective mobility that matches well with the mobilities of the ionized sugars. Vorndran et al.[23] found increased resolution of carbohydrates as the pH was increased, but a drastic drop in IPD sensitivity was observed as the pH increased past 12.2. They postulated that the decrease in IPD sensitivity is probably due to the displacement of nonabsorbing hydroxide ion, rather than sorbate ions, by the analytes. In addition they determined the best compromise pH for sensitive detection of carbohydrates to be 12.1. At this pH, C_{LOD} was about 0.5 mM. In a similar study, Vorndran et al.[51] investigated the effect of sorbate concentration on the IPD signal. The researchers reported that a lower concentration of sorbate adversely affects the TR value due to competition with hydroxide ion. On the other hand, at high sorbate concentrations S/N again deteriorates because less light is able to reach the photodiode. Likewise, Klockow et al.[126] used sorbate as an IPD reagent for the determination of carbohydrates in a variety of fruit juices. Again, the optimum CZE conditions were found with a buffer of pH 12.2–12.3.

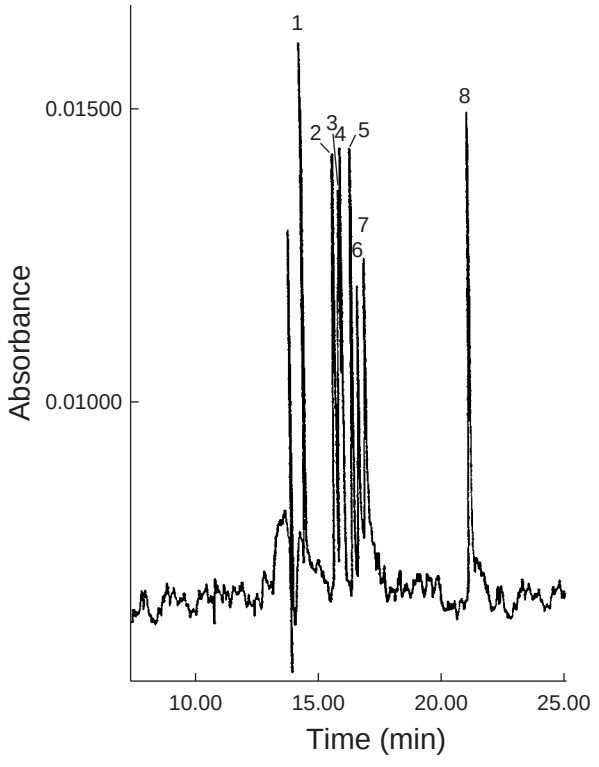

Figure 22 Electropherogram of a mixture of eight selected monosaccharides (1 mM each) in 2 mM NAA, pH 12.2. CE conditions: capillary, 120 cm (113 cm to the detector) × 50 μm i.d.; voltage = +25 kV; temperature = 25 °C; indirect absorbance detection at 222 nm; hydrodynamic injection for 3 s. Peaks: 1 = sucrose; 2 = lactose; 3 = maltose; 4 = glucose; 5 = rhamnose; 6 = fructose; 7 = ribose; 8 = gluconic acid. (Reproduced with permission from Lee and Lin.[127])

Lee and Lin[127] have evaluated the potential of six IPD reagents for the determination of carbohydrates: NAA, 2-naphthalenesulfonic acid, 1,3-dihydroxynaphthalene, phenylacetic acid, *p*-cresol, and sorbic acid. They noted that NAA is the best IPD reagent for CZE separation of eight sugars (sucrose, lactose, maltose, glucose, rhamnose, ribose, gluconic acid) (Figure 22). In addition, NAA also provided a three- to sixfold increase in separation efficiency and about a fivefold improvement in the detection limit (0.1 mM) when compared with sorbate. Lu and Westerlund[128] reported the utility of tryptophan as another IPD electrolyte for the determination of carbohydrates, which provided at least a twofold improvement in detection limits over sorbate.

CZE/IPD of underivatized carbohydrate based on sorbate or riboflavin electrolytes has also been exploited by several researchers.[129,130] Xu et al.[129] have used narrow (25 μm) capillaries and low voltages at high pH to improve the C_{LOD} of underivatized sugars. After optimization of the CZE conditions, the pH range for the separation was extended to 13. The C_{LOD} of sugars at this high pH was of the order of 0.05–1.0 mM. The developed method was applied to the analysis of culture medium samples. Liu et al.[130] reported a three- to fivefold improvement in separation efficiency with a stable baseline using diethylamine as an electrolyte additive, when compared to sodium hydroxide, at a pH as high as 12.6. It was hypothesized that low-conductivity diethylamine forms ion-pairs with anionic carbohydrates, and also suppresses electroosmotic flow by masking the silanol group at the capillary surface. The authors also reported improved separation of di- and trisaccharides with a borate buffer at pH 9.5, in place of sodium hydroxide or diethylamine, keeping the type of IPD reagent (sorbate, riboflavin) constant for both variations.

CZE/IPD for the qualitative and quantitative analysis of low-molecular-weight glycosaminoglycans (heparins) was reported by Damm et al.[131] Although heparin fragments are also amenable to direct UV detection at 232 or 200 nm, the use of 5-sulfosalicylic acid or 1,2,4-tricarboxybenzoic acid in IPD mode provides an order of magnitude higher sensitivity than direct UV detection.

Cyclodextrins belong to another class of nonchromophoric carbohydrates that can be resolved by CZE provided that the IPD reagent binds selectively with the central hydrophobic core of the cyclodextrin molecules. Nardi et al.[132] separated α, β, γ-, and δ-cyclodextrins at a pH of 6.2 using 30 mM benzoic acid as the IPD reagent. The authors demonstrated that formation of an inclusion complex of benzoate with the cyclodextrins can be beneficial for quantitative analysis of the latter in drug formulation. Various charged and uncharged cyclodextrins have also been shown to be detected in the IPD mode after their CZE separation using a range of aromatic anions.[133–135] However, it is postulated that the absorbance difference caused by cyclodextrins is probably due to displacement of the IPD reagent in the sample zone rather than a specific absorption change caused by inclusion complex formation.

Very recently a mixture of six different phosphorylated sugars and other organophosphorus compounds of biochemical interest (e.g. glucose-6-phosphate, fructose-6-phosphate, fructose-1,6-biphosphate, dihydroxyacetone phosphate, glyceraldehyde-3-phosphate, 2-phosphoglycerate) found in the glycolytic pathway were resolved under CZE/IPD conditions by Schaeper et al.[136] They found optimum analysis with 5 mM AMP and 100 mM boric acid at a pH of 7.2. Although the authors reported successful separation of five monophosphorylated sugars using the same conditions as for the glycolytic pathway, the separation of bisphosphorylated sugars could not be achieved unless magnesium was added to the AMP buffer. On the other hand, when NDS was used as the IPD electrolyte, the separation of biphosphorylated sugars was

Table 10 IPD in CE separation of amino acids and phospholipids

Analytes	Sample	IPD reagent	Electrolyte system	Detection λ (nm)	Ref.
Amino acids					
Lysine, tryptophan, leucine, isoleucine, phenylalanine, valine, histidine, methionine, glutamine, alanine, threonine, asparagine, serine, glycine, tyrosine, cysteine, glutamic acid, aspartic acid, arginine	Standard mixture	*p*-Amino-sulfonate; DMAB	(a) 10 mM *p*-Aminosulfonate or DMAB, pH 11.0 (b) 10 mM *p*-Aminosulfonate or DMAB, pH 11.0, 20 mM α-Cyclodextrin	266, 288 266, 288	137 138
Amino acids					
Lysine, tryptophan, leucine, isoleucine, phenylalanine, valine, histidine, methionine, glutamine, alanine, threonine, asparagine, serine, glycine, tyrosine, cysteine, glutamic acid, aspartic acid, arginine, uric acid, citrulline, hydroxyproline	Standard mixture	Salicylate	2 mM Salicylate, 5 mM carbonate, 0.15 mM myristoyltrimethylammonium chloride, pH 10.7	230	139
Phosphonoamino acids					
Phosphotyrosine, phosphothreonine, phosphoserine, glutamic acid, aspartic acid	Standard mixture; phosvitin (phosphoserine-containing protein)	AMP	5 mM AMP, 100 mM H_3BO_3, 10 mM HEPES, 2.5% Brij, pH 7.5	259	142
Phospholipids					
$C_{14}-C_{18}$ phosphatidic acid, phosphatidylserine, phosphatidylinositol, phosphoglycerol, cardiolipin	Standard mixture; soybean extract; bovine brain extract	AMP	(a) 5 mM AMP, 100 mM H_3BO_3, 90% methanol, 10% water (b) 5 mM AMP, 100 mM H_3BO_3, 80% methanol, 10% acetonitrile, 10% water	259	145

DMAB, 4-(*N*,*N*′-dimethylamino)benzoate.

possible without the addition of magnesium. With a 5-s vacuum injection into the capillary the authors reported a detection limit of 0.5–1 ppm (S/N = 3).

3.8 Indirect Photometric Detection of Amino Acids and Phospholipids

A number of publications have described the successful CZE separation of underivatized amino acids with IPD (Table 10).[137–141] Lee et al.[137] have evaluated nine IPD reagents for the analysis of a mixture of 20 common amino acids. The IPD reagents evaluated were sorbate, benzoate, phthalate, nicotinate, salicylate, *p*-aminosalicylate (PAS), *p*-aminobenzoate, DMAB, and 7-amino-4-hydroxy-2-naphthalenesulfonate (AHNS). These authors found that both type and concentration of the IPD reagents influence the separation behavior and detectability of amino acids. Two IPD reagents (PAS and DMAB) were recommended for the routine analysis of amino acids because they have effective mobilities that match closely the mobilities of most amino acids at high pH. For example, Figure 23 shows that the resolution of 17–19 amino acids can be achieved in 35 and 50 min, respectively, using PAS or DMAB, respectively, at pH 11 with normal polarity.[138] This was possible upon addition of 20 mM α-cyclodextrin to PAS or DMAB at pH 11.0 (Figure 23). In a systematic approach to separate amino acids, Chen et al.[139] used CZE with IPD for the analysis of amino acids in urine. Resolution was achieved by adjusting pH, buffer concentration, and

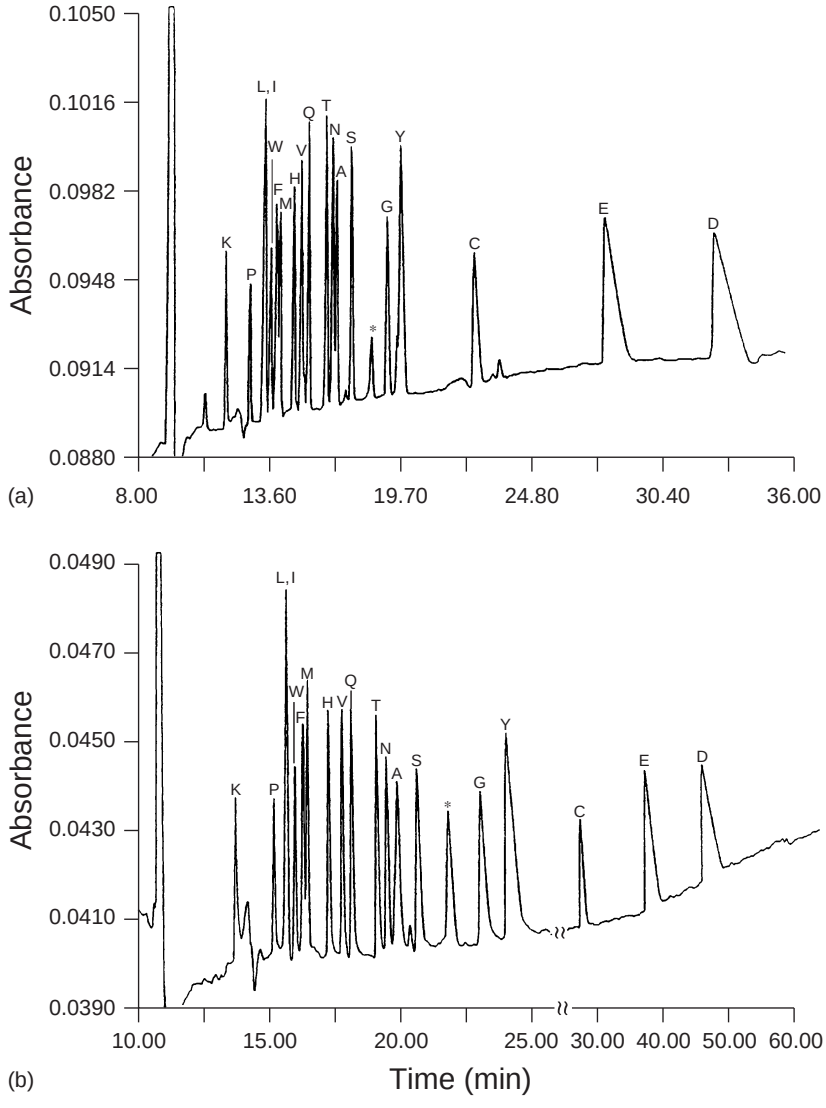

Figure 23 Electropherogram of 20 common amino acids in 10 mM PAS (a) and 10 mM DMAB (b) in the presence of 20 mM α-cyclodextrin (Sigma) at pH 11.0. Concentration of amino acids, 0.5 mM each. Peak identification: K = Lys, P = Pro, L = Leu, I = Ile (not resolved), W = Trp, F = Phe, M = Met, H = His, V = Val, Q = Gln, T = Thr, N = Asn, A = Ala, S = Ser, G = Gly, Y = Tyr, C = Cys, E = Glu, D = Asp. Arg is merged with the system peak (the first peak) and $*$ is an unknown peak. (Reproduced with permission from Lee and Lin.[138])

the type of IPD reagent (benzoate, salicylate). Similarly Hong et al.[140] have also optimized the separation of amino acids with IPD. They indicated that the choice of IPD reagent is critical for the efficiency of separation in CZE. Using the optimized conditions, baseline separation of 13 amino acids was reported. One recent report[142] demonstrated CZE separation of phosphoamino acids using AMP as the IPD reagent. Baseline resolution of three phosphoamino acids (phosphoserine, phosphothreonine, phosphotyrosine) without co-migration of two native amino acids (glutamic acid, aspartic acid) was accomplished in less than 20 min with detection limits in the micromolar range. An interesting application of the IPD method for the determination of phosphoserine in the protein phosvitin after acid hydrolysis was also described in this report.[142]

Currently, there are only a few reports on the CZE separation of anionic phospholipids (Table 10).[143–145] In the separation of these very hydrophobic anions, MEKC with direct UV detection is preferred. However, the C_{LOD} using direct detection is rather poor even at a wavelength of 200 nm. CZE/IPD was recently reported by Haddadian et al.[145] for the determination of $C_{12}-C_{18}$ phosphatidic acids, phosphatidylserine, phosphatidylinositol, phosphatidylglycerol, and cardiolipin, giving LOD values in the low-ppm range for these anionic phospholipids. The

authors demonstrated the versatility of AMP by utilizing it as the IPD reagent for the analysis of phospholipids in soybean and brain extracts.

4 SIMULTANEOUS SEPARATION AND INDIRECT DETECTION OF CATIONS AND ANIONS

The potential of CZE for the simultaneous separation of cations and anions with IPD holds particular promise for ion analysis for several important reasons. First, combined separation and detection is very convenient compared to individual separations because more information about sample cations and anions as well as the chemical processes can be obtained in a single run. Second, it is obvious that this approach is cost-effective with less work and a shorter analysis time. Although, at the present time, only few research groups worldwide are actively engaged in the development of these analytical methods, the method has the potential to broaden the scope of CZE applications.

Despite the merits discussed above, simultaneous separation and IPD of cations and anions is problematic in CZE because when the voltage is applied cations and anions tend to migrate inherently toward their respective electrodes. Because a detector is generally located on one side of the capillary, proper control of both migration direction and migration velocity is essential, so that the two classes of ions of opposite charges can move in the same direction. There are four major approaches that have been developed to cope with the aforementioned problem. The first approach utilizes Ce^{3+} and *p*-hydroxybenzoate counterions for IFD of some small inorganic cations and chloroanions respectively.[146] However, dual detectors (one at each end of the capillary) are required, making this approach difficult to adopt in a commercial CE system. The second approach, first reported and applied in IPD by Shamsi and Danielson[113] and later by Xiong and Li,[147,148] is based on the principle that if the velocity of electroosmotic flow is high enough, then both cations and anions can be determined in one run provided that they have a suitable mobility match with their respective indirect photometric co-ions. The third approach is described by Kuban et al.,[149] in which the detection window is placed approximately in the center of the capillary and the sample is injected from both ends of the capillary. Finally, the last approach reported recently[150] describes the metal cations being transformed into anionic chelates after their complexation with ethylenediaminetetraacetic acid. Hence, cations are separated together with anions under the co-electroosmotic flow conditions in CZE. Unfortunately, with this approach alkali metals and ammonium cations cannot be determined because they do not form stable complexes with ethylenediaminetetraacetic acid.

4.1 Selection of Suitable Cationic and Anionic Indirect Photometric Detection Reagents

In evaluating light-absorbing co-cations and co-anions for simultaneous separation and IPD of the nonchromophoric cations and anionic analytes, several factors should be considered:[113,147,148]

1. A two-component IPD solution must be chromophoric with high ε and DR.
2. The IPD reagent must provide a suitable mobility match, i.e. small ΔT value (ΔT = migration time of the IPD reagent minus the migration time of analyte ion).
3. IPD reagents should preferably be strong bases or acids to ensure that high fractions of both are ionized. For example, cationic IPD reagents should have a high pK_a value so that they can be displaced effectively by the nonchromophoric cationic analytes even at higher pH. In addition, anionic IPD reagents should be strong acids so that they could not only adjust the solution pH but also provide charge displacement with good sensitivity at low pH.
4. Contributions from the UV-absorbing counterions must be minimized. For this reason, mixing of free aromatic bases and acids (instead of salts) to generate a two-component IPD reagent is important. This procedure minimizes the dilution of UV-absorbing components by nonabsorbing extraneous spectator ions. Thus, a pH range of 6–7 appears to be favorable because this range minimizes the contribution of H^+ and OH^- ions (generated from the buffer), which in turn improves the detectability.
5. The light-absorbing reagents should not contain spectator ions that need to be detected in the sample. For example, the bromide salt of cationic IPD reagents will exhibit a negative system peak where bromide normally migrates in the electrolyte. This negative peak not only decreases the sensitivity for bromide but may also cause difficulties with integration of analytes migrating closer to bromide.[36]
6. The absorbance match of the IPD reagents should be as close as possible (i.e. ΔA should be minimum) so that neither the cation nor the anion is masking the absorbance of the other ion.
7. The total background absorbance should be additive and be within the linear range of the UV absorbance detector.

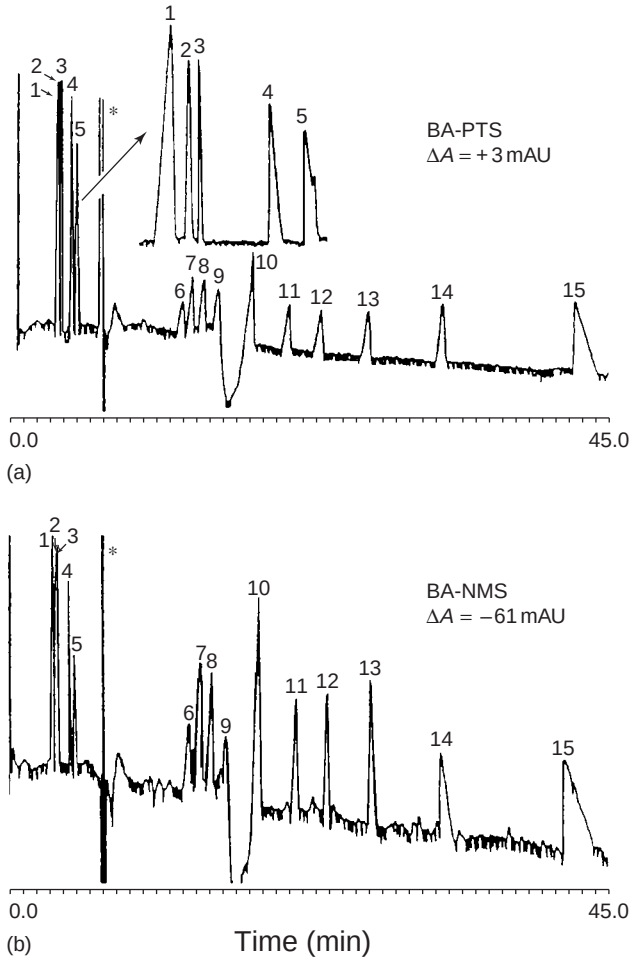

Figure 24 Comparison of benzylamine–PTS (BA-PTS) (a) and benzylamine–NMS (BA-NMS) (b) for the simultaneous separation of a standard mixture of cationic and anionic surfactants. Electrolyte consisted of 5 mM BA-PTS or BA-NMS with 100 mM H_3BO_3 (pH 6.0) and 50% methanol. Peak identification: 50 mg L^{-1} of (1) triethylamine; 25 mg L^{-1} each of (2) chloroethyltrimethylamine, (3) tetrabutylamine, (4) tetrahexylamine, and (5) didodecyldimethylamine; 50 mg L^{-1} each of (6) C_{18} SO_3^-, (7) C_{16} SO_3^-, (8) C_{14} SO_3^-, (9) C_{12} SO_3^-, (10) C_{10} SO_3^-, (11) C_8 SO_3^-, (12) C_7 SO_3^-, (13) C_6 SO_3^-, (14) C_5 SO_3^-, and (15) C_{18} SO_3^-. * is the water peak. Vacuum injection for 3 s; +30 kV applied for separation; current = 4–5 μA; IPD at 208 nm with BA-PTS and at 200 nm with BA-NMS. (Reproduced with permission from Shamsi and Danielson.[113] Copyright 1995 American Chemical Society.)

Some of the above-mentioned criteria were first evaluated by Shamsi and Danielson.[113] Benzylamine or pyridine bases were mixed with NMS or PTS to generate four different combinations of two-component IPD reagents. For the selection of a suitable wavelength for both IPD reagents (point 4), the background absorbance of 5 mM each of benzylamine, pyridine, PTS, or NMS was measured from the readout of the CE detector by flushing the capillary with light-absorbing solutions. The cation – anion background absorbance (ΔA) values for two combinations of electrolytes (benzylamine–PTS and benzylamine–NMS) were determined as a function of wavelength in 1–2 nm increments over the range 200–220 nm. The authors found the ΔA values to be +3 mAU and −61 mAU using benzylamine–PTS and benzylamine–NMS electrolytes at its minimum interference wavelength of 200 nm. Using this approach the authors compared the electropherogram obtained for the simultaneous separation and IPD of a 15-component mixture of cationic and anionic surfactants. As shown, the benzylamine–NMS electrolyte improved the detection of anionic surfactants because of the negative ΔA value (Figure 24b), but benzylamine–PTS detected cationic surfactants better, as predicted by the positive ΔA value (Figure 24a).

4.2 Selection of a Suitable pH

As discussed earlier (point 4, section 4.1) simultaneous separation and indirect detection should be optimized by a judicious selection of electrolyte pH. This is because not only the detection sensitivity but also the mobility of cationic and anionic IPD reagents and solute ions, and the magnitude of the electroosmotic flow, are affected by a change in the pH of the IPD solution. Xiang and Li verified experimentally the effect of pH using a seven-component model mixture of inorganic cations and anions.[148] A series of electropherograms obtained using pH values of 4.52, 5.00, 5.42, 6.00, 6.52, 7.04, 7.53, and 8.03 are shown in Figure 25 on the same sensitivity scale. As the pH value is increased, the peak efficiency of anionic solutes is increased. It is apparent that only two out of five anionic solutes could be detected when the pH was less than 6.00 (Figure 25, spectra a–c). On the other hand, resolutions for both cations and anions were not feasible at pH > 7.50 (Figure 25, spectra g and h). Thus, as predicted earlier by Shamsi et al., and verified experimentally by Xiang et al., the use of pH 6–7 (Figure 25, spectra e and f) seems to be the best pH range for the joint separation and IPD of anions and cations.

4.3 Effect of Indirect Photometric Detection Binary Electrolyte Concentration

Xiong et al.[148] studied the effect of a two-component IPD electrolyte (imidazole and sulfosalicylic acid) concentration at constant pH on the selectivity of cation and anion separation. They determined that by increasing the total IPD electrolyte concentration, selectivity for the separation of cations and anions (within their individual groups) increased at pH 6.52 and 7.04. However, it was also clear that the enhancement in cation selectivity is not as significant as that for anion separation. Moreover, it was found that when the total IPD electrolyte

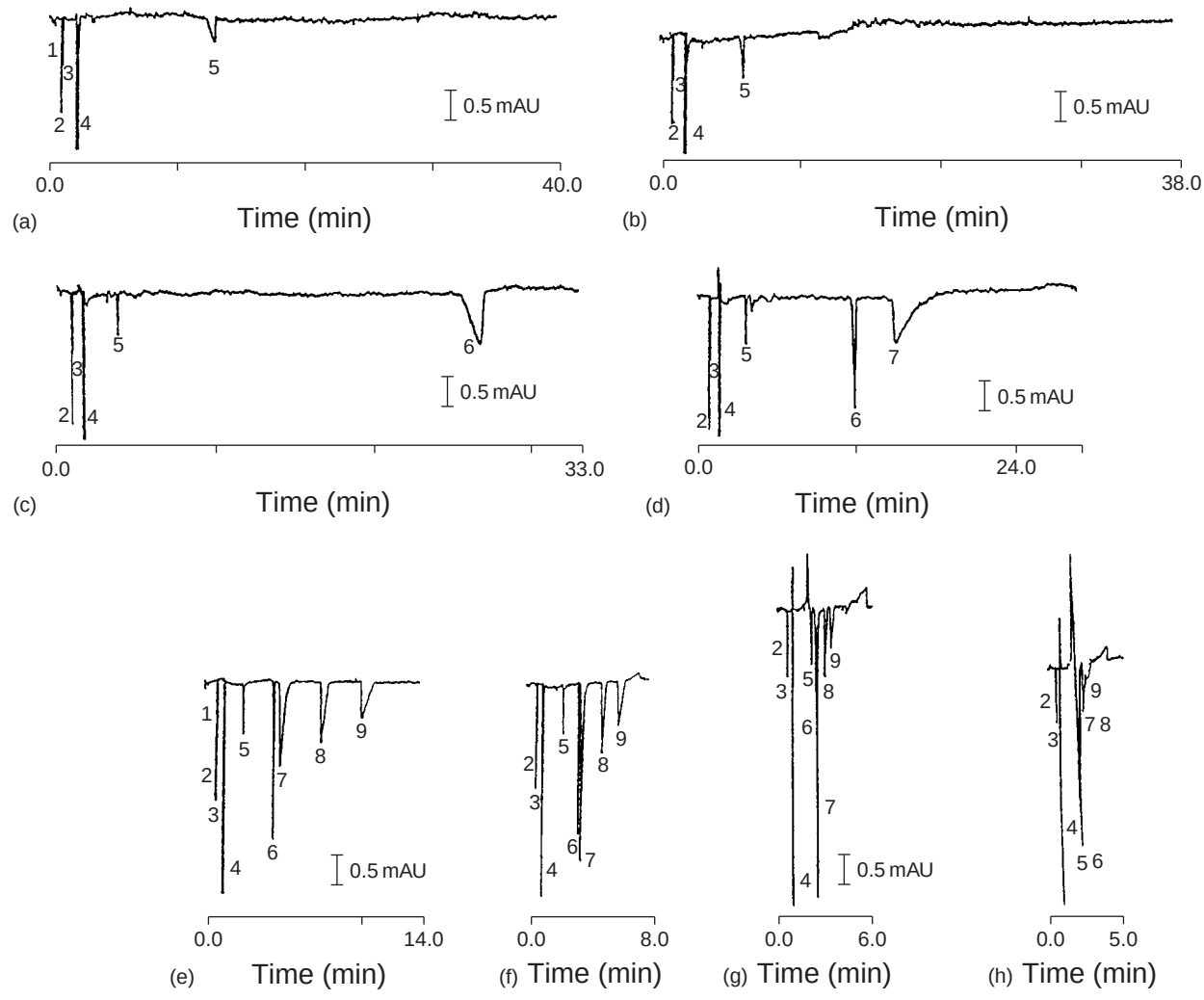

Figure 25 Electropherograms of the model sample mixture in the binary background electrolyte at different pH. Conditions: 3.06 mM imidazole and 0.1–1.5 mM sulfosalicylic acid; applied voltage = +15 kV; ambient temperature; injection for 15 s; current = 0.5–3 mA. The pH values tested were: (a) 4.52, (b) 5.00, (c) 5.42, (d) 6.00, (e) 6.52, (f) 7.04, (g) 7.53, (h) 8.03. Peaks (in ppm): 1, K^+ (5.9); 2, Na^+ (16.8); 3, Li^+ (8.73); 4, water dip; 5, HPO_4^{2-} (4.75); 6, F^- (9.92); 7, $HCOO^-$ (7.42); 8, ClO_3^- (9.96); 9, ClO_4^- (8.73). (Reproduced from Xiong and Li[148] by permission of Wiley–VCH.)

concentration was greater than 5.0 mM (with imidazole concentration higher than 4 mM) the migration times of more slowly migrating anions (e.g. ClO_3^- and ClO_4^-) increased rapidly. The authors concluded that a higher pH is advantageous for anion separations, but higher total electrolyte is deleterious to the same separations. Xiong et al.[148] further elaborated on this issue by utilizing an imidazole–benzenesulfonic acid system at two different ionic strengths (Figure 26). Although the electroosmotic flow generated under these two conditions was roughly the same, i.e. 79.5×10^{-5} and 75.4×10^{-5} $cm^2\,V^{-1}\,s^{-1}$ at pH 6.54 and 7.03, respectively, the migration times for anions were significantly increased at higher total IPD electrolyte concentration. Perhaps the counteracting effect of pH and total IPD electrolyte concentration should be considered when a simultaneous separation is performed.

4.4 Simultaneous Separation of 22 Anions and Cations in a Single Run

Using the third approach discussed early in section 4, Kuban and Karlberg[149] reported an impressive simultaneous CZE separation of cations and anions with IPD in less than 5 min (Figure 27). It is noteworthy that this is the highest peak capacity reported so far for the combined separation and IPD of these fast-moving ions. Unfortunately, using the same buffer recipe, transition metals and some alkaline earth metal ions (e.g. Ba^{2+}) cannot be analyzed because the cations tend to precipitate as

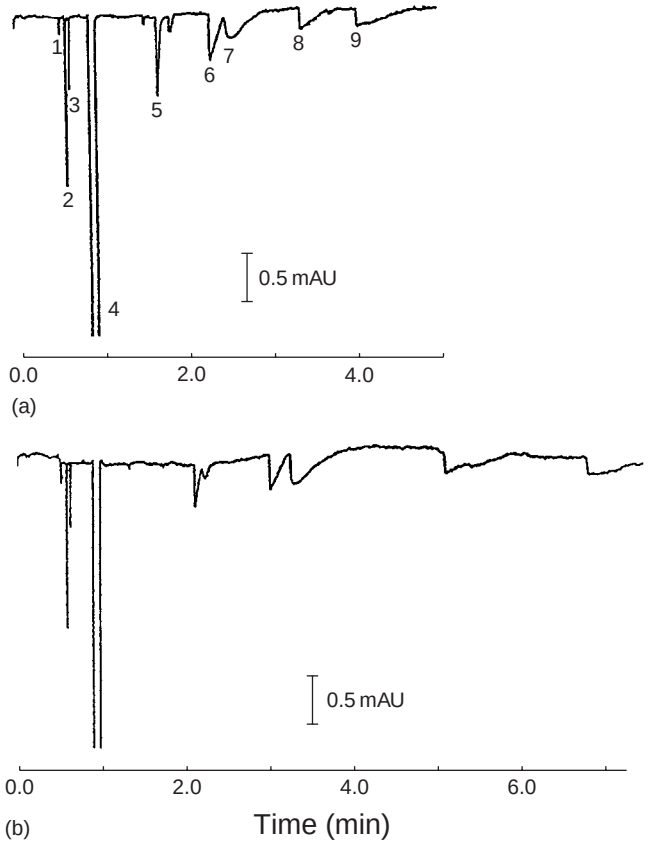

Figure 26 Separation results of the model sample mixture in imidazole–benzenesulfonic acid background electrolyte at different pHs and total electrolyte concentrations: (a) pH 6.54, 4.23 mM total electrolyte concentration (2.77 mM imidazole; 1.46 mM benzenesulfonic acid), 2.0 μA; (b) pH 7.03, 7.65 mM total electrolyte concentration (5.70 mM imidazole; 1.95 mM benzenesulfonic acid). Current = 3.5 μA; voltage = +29.0 kV; ambient temperature; detection at 214 nm; injection for 30 s; capillary = 40 cm (32 cm effective length) × 50 μm i.d. Peaks (in ppm): 1, K^+ (5.9); 2, Na^+ (16.8); 3, Li^+ (4.3); 4, water dip; 5, HPO_4^{2-} (4.4); 6, F^- (3.7); 7, $HCOO^-$ (4.3); 8, ClO_3^- (5.9); 9, ClO_4^- (4.3). (Reproduced from Xiong and Li[148] by permission of Wiley–VCH.)

hydroxides or react with the chromate ion. Alternative electrolytes therefore need to be developed for such ionic separations.

4.5 Selected Applications

Simultaneous determination of small cations and anions appears to be an efficient analytical method for ''total ion analysis'' in a variety of aqueous samples. Electrophoretic analysis of cations and anions in mud and milk samples was described by Kuban and Karlberg.[149] The dialyzed milk and mud samples were injected into the CE separation system and the total number of ions were analyzed in less than 5 min with RSD < 0.3% for migration times. In addition, good correlation was obtained

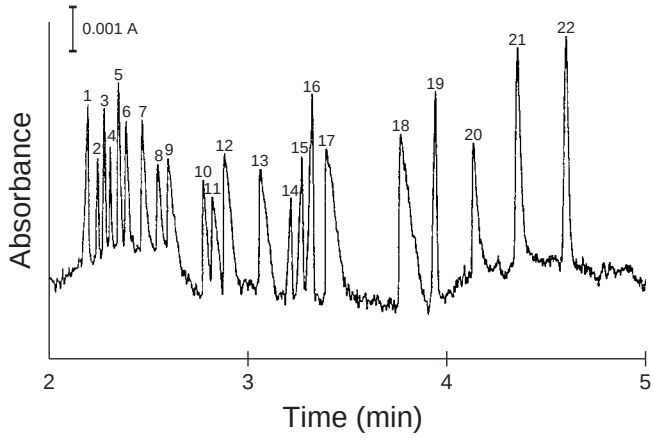

Figure 27 Separation of 22 anions and cations. Electrolyte: 6 mM 4-aminopyridine, 2.7 mM H_2CrO_4, 30 μM cetyltrimethylammonium bromide, pH 8. Indirect UV at 262 nm, with peak reversal. The linear baseline drift observed during the first 2 min was corrected for. Peak identification: (1) $S_2O_3^{2-}$, (2) Br^-, (3) Cl^-, (4) SO_4^{2-}, (5) NO_2^-, (6) NO_3^-, (7) WO_4^{2-}, (8) MoO_4^{2-}, (9) citrate, (10) maleate, (11) fumarate, (12) F^-, (13) HPO_4^{2-}, (14) Cs^+, (15) K^+, (16) NH_4^+, (17) HCO_3^-, (18) acetate, (19) Na^+, (20) Ca^{2+}, (21) Mg^{2+}, and (22) Li^+. (Reproduced with permission from Kuban and Karlberg.[149] Copyright 1998 American Chemical Society.)

between IC and capillary ion analysis data for simultaneous analysis of cations (K^+, Na^+, NH_4^+, Ca^{2+}, Mg^{2+}) and anions (Cl^-, SO_4^{2-}, NO_3^-, HCO_3^-) in tapwater and rainwater. Cations and organic acids in soft drinks (apple, orange, and grape juices) have also been determined using 1,2-dimethylimidazole–trimellitic acid with 6 mM 18-crown-6 as binary IPD reagent.[147] However, the matrix effect strongly influenced the migration of some organic acid (e.g. citrate). The researchers concluded that this is probably due to differences in the viscosity of the injected samples, which influenced the electroosmotic flow.

5 INDIRECT FLUORESCENCE DETECTION IN CAPILLARY ELECTROPHORESIS

Although the discussion thus far has concentrated on CZE/IPD, IFD also offers a means to detect analytes, even when they do not provide a signal at a given detector. In this type of detection, fluorescing ions rather than absorbing ions are used as the CZE electrolyte, and hence a large fluorescence background is always present when the excitation laser interacts with the solution flowing through the detection region. When the analytes appear in the detection region, displacement of the fluorophore ion by an analyte leads to a quantifiable change in the background fluorescence intensity that

translates to a negative peak. A simple form of fluorescence displacement is based on the displacement of charged analyte ions by fluorescing ions with the same charge on a regular equivalent-per-equivalent basis due to the requirement of electroneutrality. Other alternative displacement processes also exist, but charge interaction appears to be the most efficient. Similarly to indirect absorbance detection, IFD also can be considered as a universal detection scheme for CZE. It is pertinent to comment that the technique of capillary zone electrophoresis/indirect fluorescence detection (CZE/IFD) has been discussed in detail and in depth, in terms of basic principle and instrumentation, in some excellent monographs.[151–154]

5.1 Factors Influencing Indirect Fluorescence Detection Response

The sensitivity attainable with IFD depends on instrumental considerations. Although lamp-based fluorometers have been reported for IFD of alkali and alkaline earths ions,[155] inorganic anions,[146,156] and triorganotin compounds,[157] a more spectacular detection limit is observed when lasers are used as the excitation source.[151–154] However, unlike an absorption detector (where the background signal is very stable compared to the measured signal), background fluorescence signals in IFD fluctuate due to the inherent instability of lasers. In addition, fluorophores bleaching upon irradiation with the laser beam is a problem. Despite the high cost and complex operation of the lasers, a C_{LOD} that is an order of magnitude better (i.e. 10^{-7} M) than IPD is disappointing.

In addition to the instrumental considerations discussed above, detector response in IFD also depends on some chemical considerations. These include: concentration and type of the fluorescent reagent (also known as the IFD reagent); pH of the electrophoretic medium; and ionic strength of the buffer. According to Yeung and Kuhr,[151] IFD in CZE appeared to be more beneficial than the IPD method at a lower concentration of the visualization reagent. This is because the DR for IPD will decrease with a decrease in the following parameters: chromophore concentration; molar absorptivity; and pathlength (see Equation 1, section 1.3). Although the magnitude of fluorescence intensity also drops with a reduction in the aforementioned parameters, a high value of the DR over a wide range of concentration is still maintained in IFD due to the large signal generated from laser excitation. The observation of Yeung and Kuhr is consistent with recent reports published by a number of workers. For example, Andersson et al.[57] found that IFD is more attractive than IPD when the concentration of the fluorophore (e.g. salicylate) is less than 2 mM. Desbène et al.[158] also studied the response sensitivity of IFD by varying the concentration of fluorescein over the range 10^{-4}–10^{-8} M. Potassium *N*-butylbenzenesulfonate was used as a test sample because its electrophoretic mobility was closer to the mobility of fluorescein. The researchers reported the best compromise concentration for IFD sensitivity to be 10^{-5} M fluorescein sodium salt.

As outlined by Garner and Yeung,[154,159] there are five important considerations in the selection of IFD reagents:

1. The fluorophore that should serve as a detection reagent should have a large molar absorptivity for the excitation wavelength available. This is critical if a laser rather than a lamp is used as an excitation source.
2. The quantum efficiency of the fluorophore should be as high as possible.
3. The fluorophore should be soluble and inert in the running CZE solvent system.
4. For the optimum value of TR (Equation 1, section 1.3) the fluorophore should be fully charged with a preferable charge of unity, and present at as low a concentration as possible.
5. The fluorophore should have minimum adsorption on the walls of the capillary because any adsorbed material will not participate in the charge displacement process.

Based on these considerations, the authors first studied fluorescein as the IFD reagent for the detection of aspartic acid and glutamic acid[161] after their separation by CZE. Even without stabilizing the laser source, mass limits of detection (ca. 2 fmol) were reported for the two amino acids.

The sensitivity of IFD is also dependent on the stability of the fluorescence reagent at a given pH; therefore, it is important to select a pH range where the fluorescent intensity of the IFD reagent is constant. Garner et al.[159] first studied the effect of pH on the fluorescence signal of fluorescein. They reported that an increase in pH from 7 to 9 resulted in a very unstable baseline, presumably due to the interaction of fluorescein with the capillary wall. The researchers then switched to Coumarin 343, a fluorophore with good water solubility, high molar absorptivity (2×10^4 at an excitation wavelength of 242 nm) and high quantum yield. Unlike fluorescein, a relatively high pH stability of this coumarin was found, with some degradation above pH 11.5. The authors demonstrated excellent mass sensitivity by detecting fructose at the 2 fmol mass level. This was accomplished using 5-μm i.d. capillaries, which is a very difficult task with lamp sources. In the same report, Garner et al.[159] described why IPD and IFD analyses of carbohydrates

are limited by Equation (10):

$$TR = \frac{\alpha[\text{sugar}]}{[FL^-] + [OH^-]} \quad (10)$$

According to Equation (10), at a constant concentration of chromophore or fluorophore [FL^-] and carbohydrate [sugar], the TR is affected by the values of α (the fraction of carbohydrate in ionized form) and [OH^-] (the concentration of hydroxide in solution). Thus, α in the numerator and [OH^-] in the denominator are competing functions of pH. Consequently, the total TR (TR_{total}) goes through a maximum if it is plotted as a function of pH. The maximum in this plot is the optimal pH for IPD or IFD. However, carbohydrates that have a strong negative charge at lower pH (e.g. phosphorylated and sulfated sugars) are best optimized at lower pH where the effect of hydroxide ion in Equation (10) is minimized. Desbène et al.[158] reported that the fluorescence intensity of fluorescein first increases as the pH is raised from 4 to 8, and plateaus only in the highly basic pH range of 8–10. For optimum detection sensitivity the authors recommend a pH of 9.2, where the fluorescence intensity of fluorescein was a maximum and the electroosmotic flow was stabilized. This result obtained by Desbène et al.[158] is in disagreement with the earlier report of Garner et al.,[159] who found an erratic baseline when

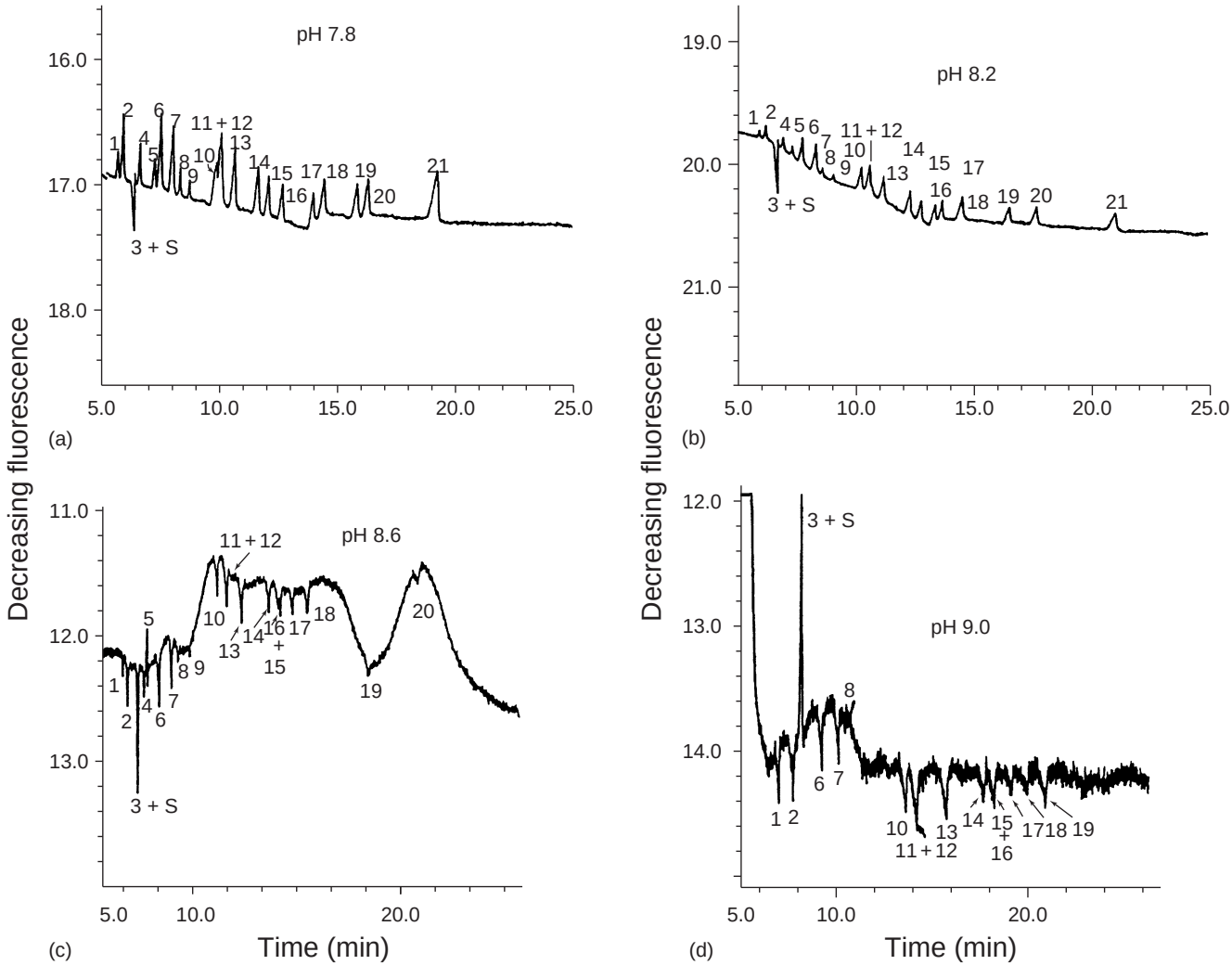

Figure 28 Effect of pH on the separation, sensitivity, and direction of the fluorimetric signal of a mixture of 21 inorganic anions and organic acids. Electrolyte consisted of 20 μM FMN in 100 mM H_3BO_3 and 2 mM DETA, adjusted to various pH values with 1 mM NaOH. All anion peak concentrations are 2 mg L^{-1} for inorganic anions and 5 mg L^{-1} for organic acids. Peaks: (1) bromide, (2) chloride, (3) nitrite, (4) nitrate, (5) chromate, (6) sulfate, (7) oxalate, (8) molybdate, (9) tungstate, (10) malonate, (11) fluoride, (12) fumarate, (13) formate, (14) succinate, (15) malate, (16) citrate, (17) tartrate, (18) phosphate, (19) hypophosphate, (20) phthalate, (21) carbonate. Pressure injection for 4 s at pH 7.8–8.6 and for 8 s at pH 9.0; voltage of −15 kV applied for separation; current varied from 7.0 to 41 μA. (Reproduced with permission from Shamsi et al.[160])

fluorescein was used at pH 9.0. The discrepancy may be due to the fact that a temperature-controlled commercial CE system equipped with a stabilized laser source was employed in the former study. Very recently, Shamsi et al.[160] reported a loss in S/N for inorganic anions using flavin mononucleotide (FMN) fluorophore at pH values greater than 8.0. Because FMN shows a stable response in the pH range of 3.5–8.0,[161] one would expect a reduction in fluorescence response at the weakly alkaline pH values. The situation was a bit more complicated, however, when peak direction was switched, i.e. positive peaks were observed at pH values such as 8.6 or 9.0. Probably the replacement of OH^- as well as FMN by the sample anions could cause a temporary localized decrease in pH. At pH 8.6 or 9.0, this may lead to a fluorescence increase, whereas at pH 7.8 or 8.2 the fluorescence remains quite constant. Therefore, positive peaks (decreased fluorescence) of inorganic anions and organic acids were observed at pH 7.8 and 8.2 and the baseline noise was quite low (Figure 28a,b). In contrast, negative peaks (increased fluorescence) were observed at pH values of 8.6 and 9.0 (Figure 28c,d) due to the more important decrease in OH^-. The optimum pH for the best S/N for anions was found to be around 8.0 or lower, where preliminary replacement of FMN by the sample anions causes the signal change.

Finally, the ionic strength of the CZE buffer was also shown to play a major role in evaluating the sensitivity of IFD.[158] In fact, buffers of low ionic strength can be very beneficial because the Joule heating induced by the low operating currents is minimized. However, distortion of the analyte zone with buffers of very low ionic strength is possible. Figure 29 illustrates the reduction in S/N of Li^+ and K^+ as a function of the ionic strength of borate buffered at pH 9.2. It is clear that IFD sensitivity was a maximum at a sodium borate concentration of 10 mM. A 10-fold improvement in sensitivity of the cations was observed when the sodium borate concentration was further decrease from 10 to 2.5 mM.

5.2 Applications of Indirect Fluorescence Detection for the Analysis of Ionic Compounds

Although a wide variety of UV-absorbing chromophores are now available for both cation and anion analysis by CZE/IPD, little work has been done to characterize IFD electrolytes for CZE of ionic analytes. This is for two major reasons: the number of fluorophores that can serve as a detection reagent is smaller than the number of chromophores, and fluorophores that have strong emission are large organic molecules with low mobility, resulting in poor selectivity for high-mobility inorganic anions.

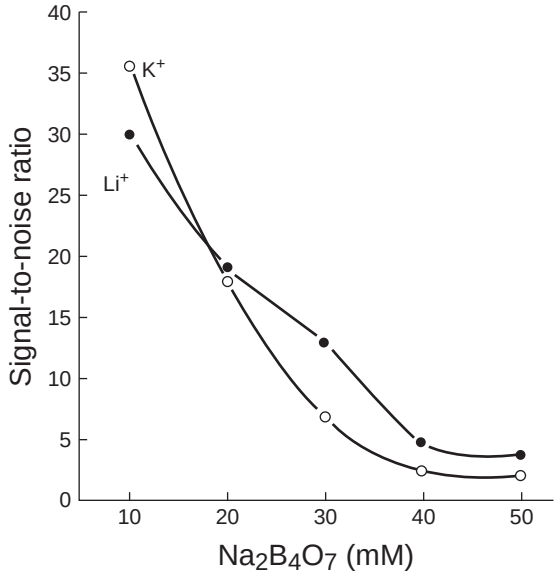

Figure 29 Evolution of S/N, in the hydrodynamic injection mode, as a function of the ionic strength of the electrolyte in the mobile phase. Operating conditions: fused-silica capillary 57 cm × 75 µm i.d.; applied voltage = 10 kV; temperature = 30 °C; sample mixture of $LiNO_3$ (10^{-4} M) and KNO_3 (10^{-4} M); electrolyte, $Na_2B_4O_7$ (pH 9.2); fluorescein concentration, 10^{-5} M; fluorimetric detection ($\lambda_{exc.} = 488$ nm, $\lambda_{detect.} = 520$ nm). (Reproduced with permission from Desbène et al.[158])

5.2.1 *Capillary Zone Electrophoresis/Indirect Fluorescence Detection of Anionic Compounds*

In 1988, Kuhr et al. in Yeung's laboratory first introduced IFD for the CZE analysis of underivatized amino acids, proteins, and nucleosides.[162,163] Later, Yeung's laboratory further evaluated IFD for an extensively large number of analytes. The anionic compounds for which successful CZE/IPD was developed include inorganic anions,[164] nucleotides,[165] sugars[159] and tryptic digests,[165] lactate and pyruvate,[166] and more recently free cyanide and related compounds.[167]

Salicylate was the IFD reagent reported for determination of seven common inorganic anions in less than 5 min.[164] Using a 0.25 mM equimolar mixture of salicylic acid and sodium salicylate at a pH of 4.0, detection limits in the submicromolar range (1×10^{-7} M) were obtained. The authors found that the same IFD reagent is also useful for the separation of 12 nucleotides in ca. 6 min.[164] However, in order to reduce the electroosmotic flow, the salicylate concentration was increased to 1.0 mM and the pH was decreased to 3.5. In a related paper, Garner and Yeung[159] were able to resolve three sugars (glucose, sucrose, fructose) and seven underivatized amino acids (phenylalanine, alanine, leucine, serine, glutamic acid, aspartic acid, proline) using 1 mM coumarin at pH 11.5

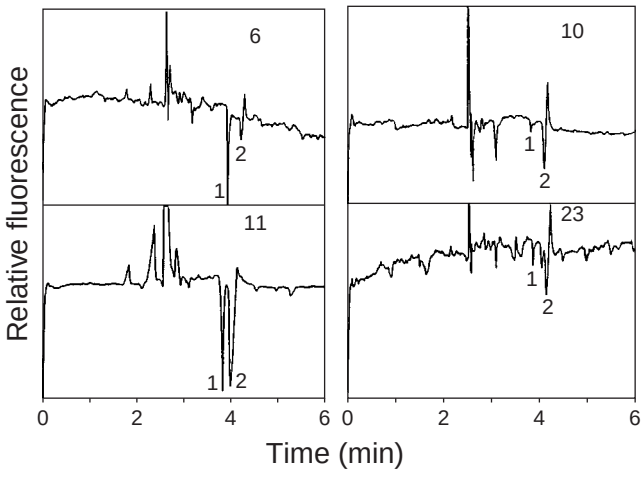

Figure 30 Electropherograms for the analysis of red blood cells. The numbers 6, 11, 10, and 23 refer to the chronological order of the experiments. Peaks 1 and 2 refer to lactate and pyruvate, respectively. (Reproduced with permission from Xue and Yeung.[166])

and 9.5 respectively. Although the authors determined that the optimum detection pH for the CZE separation of carbohydrates would be 11.65 (based on Equation 10), still a pH of 11.5 was used to generate the electropherogram of sugars. This is because they found significant degradation of coumarin as they increased the pH toward the optimum value. A more sophisticated approach was described later by Xue et al. in Yeung's laboratory.[166] The authors applied IFD for the quantitation of the major anions (lactate and pyruvate) within the single erythrocytes at the femtomole level (Figure 30). Unfortunately, it was difficult to obtain the migration time precisely (i.e. the RSD for migration time change was up to 20%). Consequently, a migration time ratio method was used that provided much improved precision (i.e. 3% RSD). Marti et al. in collaboration with Yeung's group[167] have developed an IFD method after CZE separation of free cyanide (CN^-) and related inorganic anions (CNO^-, SCN^-, and NO_3^-). Using 1 mM fluorescein at pH 11, the CZE separation and IFD of CN^- was achieved in less than 4 min without any interferences from common inorganic anions (Cl^-, SO_4^{2-}, and HCO_3^-). The C_{LOD} of CN^- was determined to be 5×10^{-6} M, which was close to the expected theoretical value (Equation 1, section 1.3).

Fluorescein as an IFD electrolyte was also utilized by Chao et al. for the separation and IFD of 11 priority phenols.[168] Nearly baseline separation of 11 phenols was achieved in ca. 14 min at pH 9.9, with 15 mM borate containing 1 mM fluorescein. The method was applied successfully to the analysis of a National Institute of Standards and Technology standard reference material (NIST SRM 184) and priority phenols in industrial wastewaters.

Although the use of lasers for IFD began as early as 1988, substantial interest has developed recently due to the commercial availability of a laser-induced fluorescence detector for CE. Desbène et al.[158,169] in a series of two papers have advocated the use of fluorescein for this commercial system. In the first paper,[158] the authors analyzed a mixture containing three transition metals (Fe^{3+}, Zn^{2+}, and Cu^{2+}), three alkaline earths (Ba^{2+}, Ca^{2+}, and Mg^{2+}), and two alkali metals (Li^+ and K^+). Because the use of fluorescein provides maximum sensitivity at basic pH, a complexing agent such as ethylenediaminetetraacetic acid was added to transform transition metal and alkaline earth metal cations into anionic chelates, so that their interaction with the capillary could be minimized at basic pH. As expected, Li^+ and K^+ eluted before the electroosmotic flow, whereas the six metal chelates eluted after the electroosmotic flow under the optimized conditions of 10 μM fluorescein and pH 7.5. They rationalized the observations based on the treatment developed by Ackerman et al.,[18] who have shown that if the absolute mobility of the analyte is smaller than the mobility of the counterion associated with the chromophore or fluorophore, then negative peaks will be observed, and vice versa. Thus, the observed positive peak for K^+ and negative peak for Li^+ agree with the theory of Ackerman et al. However, it was not very clear as to why the Fe^{3+} chelate provided a negative fluorescent peak whereas Ba^{2+}, Ca^{2+}, Mg^{2+}, Zn^{2+}, and Cu^{2+} chelates showed positive peaks.

In a second paper Desbène et al.[169] separated two mixtures – alkanesulfonates (C_1-C_{10} SO_3^-) and alkyl carboxylates (C_1-C_{12} COO^-) – under aqueous conditions using 10 μM fluorescein and 5 mM $Na_2B_4O_7$ at pH 9.2 in about 4 and 7 min, respectively. However, the analysis of fatty acids longer than C_{12} required the use of organic solvents (2-propanol, ethanol, or acetonitrile). The authors found no decrease in background fluorescence, irrespective of the solvent used, until the percentage of organic solvent was lower than 40% (v/v). Beyond this limit, baseline fluctuations and noise were increased substantially. They were able to achieve detection limits as low as 700–800 ppb, and the system was used to establish calibration curves and response coefficients for IFD.

Several other groups have continued to extend the utility of fluorescein. For instance, Wang et al. evaluated the feasibility of using nonaqueous solvents in CZE with IFD.[170] C_6-C_{24} FFAs differing by only one carbon atom were separated and detected in a highly alkaline methanol–acetonitrile system in combination with fluorescein as the IFD reagent. Similarly, Huang et al.[171] used 1.5 mM fluorescein at pH 9.8 for effective separation and detection of five arsenic compounds (arsenite, arsenate, monomethylarsonate, dimethylarsonate, and phenylarsonate) using CZE with IFD in 8 min. Detection limits in the 0.04–0.16 mg mL^{-1} range were

reported. Unfortunately, the developed method could not be applied for the analysis of arsenic in groundwater, due to matrix effects.

The CZE/IFD method using salicylate reagent was compared to CZE with direct and indirect UV detection for the determination of isoprenyl pyrophosphate.[57] The researchers reported that IFD was the most suitable detection method that provided the lowest C_{LOD} (0.5 μM) and highest separation efficiency (600 000) compared to direct or indirect UV detection. However, the C_{LOD} was still not low enough to allow analysis of dilute samples of rat liver.

Very recently, the novel IFD reagent FMN, also known as riboflavin-5′-phosphate, was developed in our laboratories for indirect laser-induced fluorescence detection (ILIFD) of a wide spectrum of anionic compounds.[160] Unlike fluorescein, FMN provided strong fluorescence at both acidic and basic pH values. A CZE separation of 21 inorganic and organic anions in about 20 min was possible for the first time with ILIFD. Figure 31 shows the separation of 21 inorganic anions and organic acids at 5 μM and 30 μM FMN. At 100 μM baseline drift was substantial and the S/N decreased (electropherogram not shown). The optimum concentration of FMN was found to be 30 μM because of a better S/N and a reasonable baseline drift. In addition, the use of methanol provided improved resolution and facilitated simultaneous separation and universal IFD of aliphatic and aromatic anionic surfactants. Using pressure injection, the C_{LOD} was 20–30 μg L^{-1} for inorganic anions, 40–50 μg L^{-1} for organic acids, 250–500 μg L^{-1} for P_n, and 500–1000 μg L^{-1} for alkyl sulfates. C_{LOD} was optimized further by at least twofold with electrokinetic injection, and the developed CZE/IFD was applied to the simultaneous analysis of ethoxylated lauryl sulfate and lauryl sulfate in a commercial shampoo in less than 10 min.

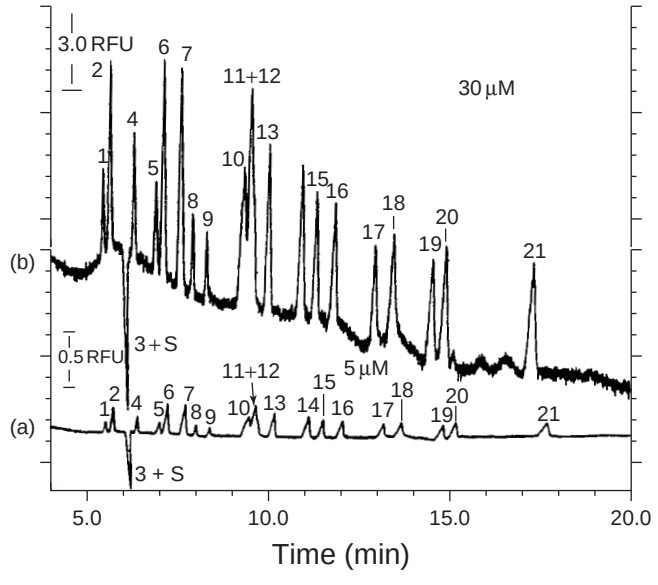

Figure 31 Effect of FMN concentration on separation, selectivity, and baseline stability of a mixture of 21 inorganic anions and organic acids. Electrolyte consisted of 5 μM (a) and 30 μM (b) FMN in 100 mM H_3BO_3, 2 mM DETA, pH 7.8 buffer; pressure injection for 4 s; voltage of −15 kV applied for separation. Current varied from 6.5 to 8.0 μA. Peak identification is the same as in Figure 28. RFU = relative fluorescence units. (Reproduced with permission from Shamsi et al.[160])

5.2.2 *Capillary Zone Electrophoresis/Indirect Fluorescence Detection of Cationic Compounds*

Unlike anionic separations, IFD of cations by CZE is largely unexplored. A literature survey indicated that the cationic compounds for which successful IFD has been reported include only a few organic cations (alkylamines, alkanolamines, tetraalkylammonium, and alkyltins).[165,175] Quinine sulfate excited with an argon ion laser at 350 nm was the first cationic fluorophore studied for the analysis of some alkali and alkaline earth metal ions and several substituted amines.[165] The optimized system used included 0.38 mM quinine sulfate as IFD reagent and 0.58 mN H_2SO_4 to create a pH of 3.7. Because quinine sulfate is a relatively low-mobility cation, the later-eluting peaks (diethylamine, diethanolamine, tetraethylammonium, and trishydroxyaminomethane) were taller and sharper than those for early-eluting ions (ammonium or dimethylammonium). This pattern of peak profiles in IFD was consistent with the theory developed on peak shapes for CZE/IPD.

In a recent paper, Church et al.[172] have also demonstrated the potential of quinine sulfate in combination with α-HIBA for the separation and ILIFD of lanthanide ions, including La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+}, Sm^{3+}, and Eu^{3+}, in the low-ppb range (6–11 μM). An impressive C_{LOD} was possible due to the combination of electrokinetic injection and transient isotacophoresis, which enabled the metal cations to stack into more concentrated zones.

Underivatized amino acids as positively charged ions were investigated using a semiconductor laser with an excitation source of 670 nm.[173] Three potential dyes (methylene blue, rhodamine, and oxazine 750) that can be excited at 670 nm were evaluated as IFD electrolytes. Among the three dyes studied, methylene blue was found to be the most suitable due to its lesser adsorption on the capillary surface. At a high pH value of 11 the adsorption of methylene blue was reduced further, but the fluorescence intensity was substantially decreased at this pH. Therefore, an acidic pH of 3.80 was used to separate and detect the two amino acids (glycine and proline) via CZE/IFD in ca. 15 min.

A small cationic fluorophore, 2-AP, was investigated for ILIFD of sodium and potassium at attomole levels. Li

and Yeung[174] reported that the difficulty associated with the determination of these inorganic cations at such low levels is due to their contamination in the environment. After detailed studies, it was realized that particulate contamination was the primary source of interference for the cations. This is because one dust particle contains femtomoles of sodium and potassium because as the particle is inadvertently injected false peaks are obtained. After optimizing the procedure for cell samples they reported the mass LOD of potassium and sodium to be 110 and 70 amol respectively.

The first successful use of a lamp-based source for IFD in CZE was accomplished by Bächmann et al.[155] for the determination of ammonium and alkali and alkaline earth metals by means of an inexpensive fluorescence detector. This accomplishment was carried out using cerium(III) fluorophore excited at 251 nm, with emission collected at a 345-nm cut-off filter. The authors found that in cerium(III) sulfate solution the effective positive charge of Ce^{3+} is smaller than that in solutions of cerium(III) nitrate and chloride. Their observation was in agreement with the lower values of effective mobilities observed for cations separated by CZE using cerium(III) sulfate compared to cerium(III) nitrate. In addition, the authors recommended cerium(III) nitrate and chloride salts as suitable IFD electrolytes for the measurement of fast-moving cations (e.g. Cs^+, NH_4^+, K^+), whereas cerium(III) sulfate was suggested as a better electrolyte for slow-moving cations (e.g. Li^+). Some metal cations in the cola beverage were profiled, with C_{LOD} reported in the low micromolar range (0.1–0.3 μM).

Finally, Lee and Whang have recently introduced 6-aminoquinoline as the IFD reagent for trialkyltin compounds.[175] A CZE separation of trimethyl-, triethyl-, tripropyl-, tributyl-, and triphenyltin with IFD using a nonlaser-based fluorometer was achieved in less than 16 min. A C_{LOD} in the range 8–18 μM that corresponds to a mass LOD of 80–180 fmol was obtained. The authors pointed out that the lamp-based IFD is simple and cost-effective, with the added advantage of a wide range of wavelengths available for excitation compared to the laser-based IFD.

5.3 Applications of Indirect Fluorescence Detection for the Analysis of Neutral Compounds

Although MEKC can be used easily to separate neutral compounds, indirect detection of these compounds is somewhat problematic. This is because the charge displacement cannot be used to displace the absorbing or fluorescent ion from the analyte zone. One of the most elegant solutions to the problem was provided by Amankawa and Kuhr,[176] who introduced the use of anionic SDS as complexing agent to the positively charged

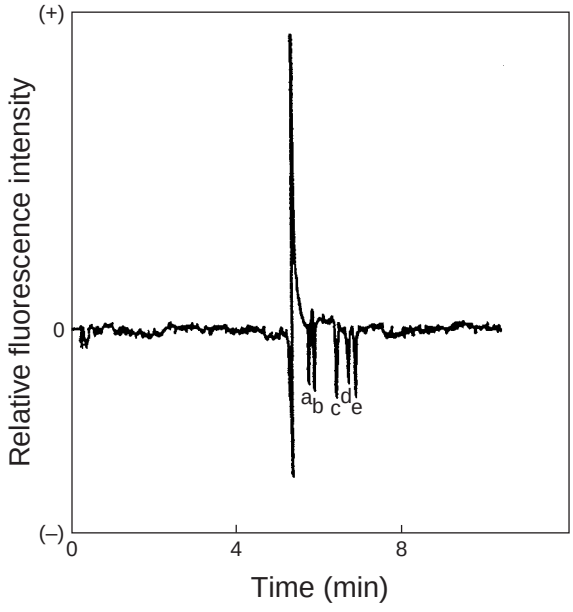

Figure 32 Electropherogram showing the separation of a mixture of alchols by MEKC and IFD. Buffer was composed of 0.50 mM quinine sulfate in 100.0 mM SDS aqueous solution, pH 6.8. Instrumental parameters were as follows: capillary was 80 cm fused-silica (60 cm to detector), 50 μm i.d. and 360 μm o.d., excitation wavelength = 325 nm; voltage = 25 kV (current = 26 μA); temperature = 30 °C; sample was composed of 0.06% v/v (7.0 mM) of each component and was injected for 1.0 s by applying a 5-in. Hg vacuum to the cathodic compartment. Peak identity is as follows: (a) 2-propanol, (b) 1-propanol, (c) 2-butanol, (d) 1-butanol; (e) 2-methyl-1-propanol. (Reproduced with permission from Amankawa and Kuhr.[176] Copyright 1991 American Chemical Society.)

quinine fluorophore. The authors observed enhanced fluorescence of the quinine ion because of its ion-pairing with the micelle, leading to a large fluorescent background signal. Upon injection, the nonionic analyte enters the micelle and can disturb the micelle–fluorophore complex. Thus, the background fluorescence is decreased, resulting in a negative peak. Despite the enhancement of the background fluorescence intensity of quinine sulfate, no significant shift in the maximum of the quinine emission was observed in the presence of SDS. A pH of 6.80 with 0.50 mM quinine sulfate and 100 mM SDS provided the separation of a mixture of five aliphatic alcohols in less than 8 min (Figure 32). The authors were able to detect phenols with much higher sensitivity (about five times) than aliphatic alchols, even though the latter group of solutes eluted much earlier than the former. This indicated that the phenolic compounds have a relatively stronger quenching effect and can disturb the micelle–fluorophore complex more efficiently.

Recently, Smith et al.[177] evaluated three IFD reagents (quinine sulfate, 8-anilinonaphthalene-1-sulfonic acid and *N*-phenylnaphthylamine (PNA)) in combination with a

nonionic surfactant (decanoyl-*N*-methylglucamide) and borate. The authors observed that PNA, the most hydrophobic fluorophore, provided the best sensitivity for IFD. Similarly to the observation of Amankawa et al.,[176] the results obtained by Smith et al. also showed an increase in IFD response with an increase in capacity factor of the solute. Figure 33 shows a typical electropherogram of an alkyl phenyl ketone homologous series as evidence for this observation. In addition, Smith et al. demonstrated a linear relationship between peak height and capacity factors, in contrast to a nonlinear behavior found by Amankawa et al.

Kennedy et al.[178] have applied MEKC for the IFD of high explosives. Fluorescein and rhodamine B were evaluated as potential fluorophores. In general, the authors found rhodamine B to be a better IFD reagent than fluorescein because the former resulted in a stable baseline and greater sensitivity. Despite the fact that some explosives could be detected at concentrations as low as 6×10^{-6} M, overall detection sensitivity was poor. The researchers attributed this lack of sensitivity to the laser instability, and suggested the use of a laser power stabilizer as a possible solution to laser noise and baseline drift.

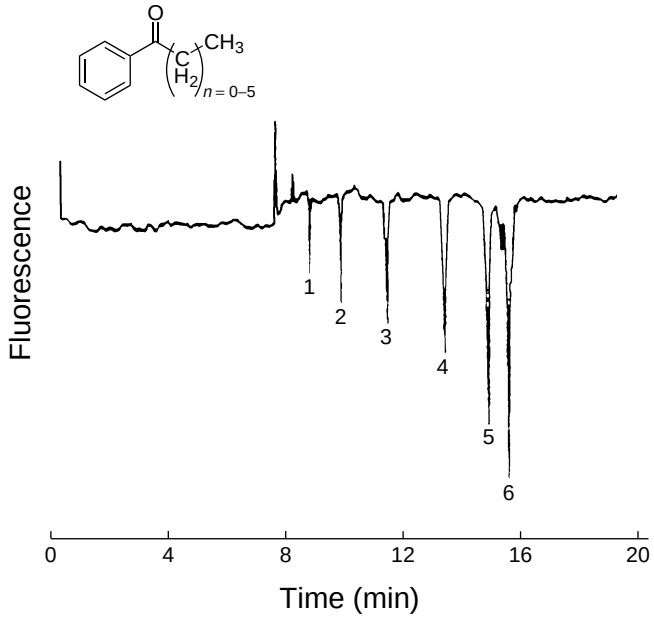

Figure 33 Electropherogram of alkyl phenyl ketone homologus series. Experimental conditions: running electrolyte, 100 mM borate containing 50 mM decanoyl-*N*-methylglucamide (MEGA 10) and 0.5 mM PNA, pH 10.0; capillary, 50-cm (detection point) untreated fused-silica, 80 cm (total length) × 50 mm i.d.; voltage = 15 kV; excitation wavelength = 365 nm; emission wavelength > 400 nm. Analytes: (1) acetophenone ($n = 0$), (2) propionophenone ($n = 1$), (3) butyrophenone ($n = 2$), (4) valerophenone ($n = 3$), (5) hexanophenone ($n = 4$), and (6) heptanophenone ($n = 5$). Concentration injected: 1 mM of each. (Reproduced with permission from Smith and Rassi.[177])

Fuchigami and Imasaka accomplished[179] IFD of flavin adenine dinucleotide and deoxyadenosine monophosphate following their MEKC separation. Methylene blue was the IFD reagent and a semiconductor laser emitting at 670 nm was used as the excitation source. This indirect fluorescence detection/micellar electrokinetic chromatography (IFD/MEKC) method also provided lower mass detection limits (fmol), by several orders of magnitude, than those reported for an HPLC-based method using the same semiconductor laser fluorescence detection.

6 OTHER INDIRECT DETECTION MODES

6.1 Indirect Chemiluminescence Detection

The general indirect route to chemiluminescence detection is to use a nonchemiluminescence analyte that will interfere or suppress a chemiluminescence reaction. Thus, inverted or negative peaks are observed if the analytes are detected indirectly due to a decrease from a constant chemiluminescence background. Liao et al. were probably the first to demonstrate the successful use of ICLD in CE.[180] It is well known that the chemiluminescence reaction of luminol with hydrogen peroxide is catalyzed by Cu(II), and the chemiluminescence response is proportional to the concentration of free Cu(II). Liao et al. showed that in the absence of analytes a high and constant chemiluminescence background is maintained when Cu(II) is continuously mixed with the luminol and H_2O_2 in the reaction/detection capillary at pH 10.0. After the CE separation, analytes (e.g. amino acids) enter the reaction zone and complex with Cu(II), decreasing the concentration, and catalytic activity of the free Cu(II). As a result of the postcapillary formation of Cu(II)–amino acid complexes, chemiluminescence intensity is decreased, resulting in a negative peak.

In a subsequent paper,[181] the same authors examined a number of parameters of the ICLD system, including buffer concentration, chemiluminescence reagent concentration, catalyst concentration, α-cyclodextrin concentration and pH. It was found that the luminol chemiluminescence reaction needs an alkaline pH in the range 9–11. A comparison of high-pH buffers (borate, dibasic sodium phosphate, and carbonate) indicated that 15 mM carbonate at pH 10.0 provides the optimum chemiluminescence emission, leading to the best sensitivity for ICLD of amino acids. They suggested that the luminol/H_2O_2 concentration should not exceed 5 mM : 25 mM. This is because a higher concentration of luminol decreases the chemiluminescence intensity. On the other hand, an increase in the concentration of H_2O_2 increases the chemiluminescence intensity. However, bubble formation in the buffer reservoir was a problem due to unsteady

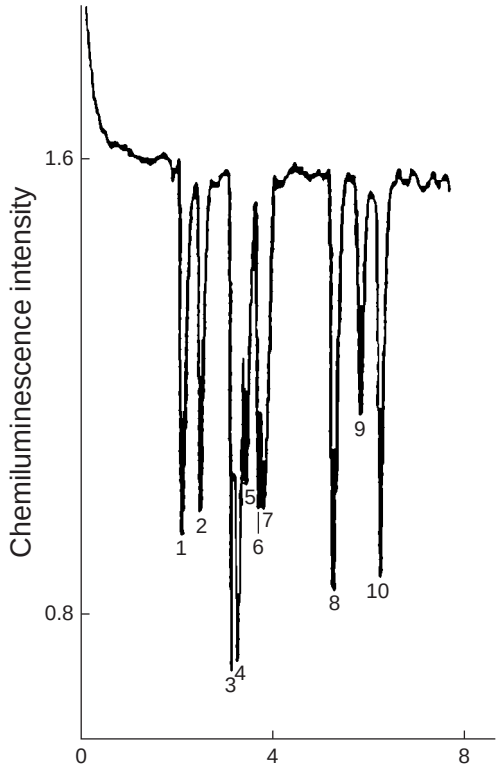

Figure 34 Electropherogram of 10 amino acids with ICLD. Electrophoretic buffer: 15 mM carbonate–5 mM luminol–25 mM H_2O_2, pH 10.0. Separation capillary, 50 μm i.d., 50 cm long; sample injection for 2 s at 21 kV; sample concentration, 1 mM each of analyte. Peaks: 1 = arginine, 2 = proline, 3 = valine, 4 = leucine, 5 = glutamine, 6 = asparagine, 7 = serine, 8 = cysteine, 9 = glutamic acid, 10 = aspartic acid. (Reproduced with permission from Liao et al.[180])

electrophoretic current and a substantial increase in baseline noise. They also observed that the maximum S/N for ICLD is ca. 30 μM $CuSO_4$. Tartaric acid, which prevented the formation of $CuCO_3$ or CuOH in alkaline pH was added as an auxiliary reagent to Cu(II). The optimization procedure discussed above resulted in CZE separation with ICLD of 10 amino acids (Figure 34). However, a complete baseline separation of 10 amino acids was not successful. Use of 50 mM α-cyclodextrin and 5% methanol led to improved resolution only at the expense of detection sensitivity. One of the major limitations of ICLD is the sigmoidal nature of the calibration curve, which limits the dynamic range to only about two orders of magnitude. Although, the mass LODs of amino acids was in the 100–140 fmol range, it depended mainly on the formation constants of amino acids with Cu(II). Application of the developed ICLD method was demonstrated by analyzing glutamate in a commercial drink.

A similar ICLD method for catecholamines and neurotransmitters has been studied recently by Zhang et al.[182] The authors found that Co(II) is a better catalyst for chemiluminescence reaction compared to Cu(II). The rationale was based on the fact that the Co(II) possesses a shorter half-life (time required to obtain maximum chemiluminescence intensity), with a fast kinetic reaction leading to a stronger chemiluminescence intensity. On the other hand, the use of Cu(II) with a longer half-life may result in weaker chemiluminescence intensity. Zhang et al. were able to resolve and detect four catecholamines (catechol, epinephrine, norepinephrine, and dopamide) by employing 10 mM SDS, 1 mM luminol and 5 mM borate (as separation electrolytes), with 5 mM H_2O_2–5 μM Co(II) as the chemiluminescence reagent. In addition, mass LODs in the range 22–87 fmol for catecholamines and 9–250 fmol for six amino acids were reported. These detection limits were about two or three orders of magnitude lower than those obtained in HPLC with ICLD.

In summary, the use of ICLD in CE can provide femtomole mass detection limits but it suffers from two major shortcomings: the method of ICLD is useful only for a few organic compounds, therefore one area of key importance is the development of appropriate recipes for electrophoresis buffers and chemiluminescence reagents to separate and detect a variety of non-chemiluminescence inorganic and organic species (metal ions, peptides, proteins and cyclic nucleotides); and in the present design of the CE/chemiluminescence interface it takes ca. 7 s for the solution to flow from the tee to the reaction cell. As a result, the chemiluminescence catalyzed by metal ions decays before it can be used to provide a sufficient background chemiluminescence signal for indirect detection. Hence, the design of a new and improved CE/chemiluminescence interface should help to decrease the half-life required for maximum chemiluminescence signals, leading to lower LODs.

6.2 Indirect Electrochemical Detection

Electrochemical principles appropriate for indirect detection in CE include potentiometry, amperometry, and conductivity. In recent years, only the latter two detection methods have been implemented successfully in the indirect mode in CE.

6.2.1 Indirect Amperometric Detection

The general indirect route to amperometric detection is to use an electrophore-competing ion that is itself electroactive, and to monitor the changes in the concentration of this electroactive species as analyte ions are eluted from the capillary. The earliest and perhaps the only example reported so far in the literature is likely to be the work of Olefirowicz and Ewing,[183] who used 3,4-dihydroxybenzylamine (DHBA) as an electrophore

to monitor the emergence of sample ions by registering a decrease in the background current. The authors used 0.1 mM DHBA and a working electrode held at a constant potential of 0.7 V vs. SSCE, taking advantage of the pH-dependent oxidation behavior of DHBA to orthoquinone as shown in Scheme 1:

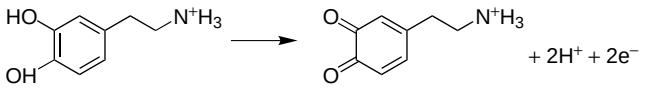

Scheme 1

A high and constant background current was obtained due to continous oxidation and a steady flow of DHBA. After the analytes were separated by CE and passed through a detector region, a lower level of oxidizable species was observed in the detection zone. The background current thus decreases, resulting in a negative peak. It is noteworthy that the electrophore needs to be present at a very low concentration only, and may therefore be used in conjunction with a buffered system to achieve reproducible migration times. Figure 35 shows the separation of three nonelectroactive cationic amino acids and three nonelectroactive cationic dipeptides using 2-morpholinoethanesulfonic acid buffered at pH 5.65 containing 0.01 mM DHBA, giving a mass LOD as low as 380 amol for the amino acid arginine.

6.2.2 Indirect Conductivity Detection

Although CE with direct conductivity detection in both suppressed[186,187] and nonsuppressed[188,189] mode is documented, little work has been done with ICD. Similarly to other indirect methods, ICD in CZE is achieved when the carrier electrolyte has a higher level of background conductivity than the analytes. This means that in order to achieve a sensitive signal for ICD, the conductivities of the carrier and electrolyte co-ions have to differ significantly. On the other hand, CZE separation requires that the mobilities of the analyte and carrier ions should be close to avoid excessive peak spreading. These two contradicting requirements can be compensated partly by increasing the ionic strength of the carrier electrolyte.[188] In general, for sensitive detection of high-mobility inorganic anions and cations a low-mobility carrier electrolyte with direct conductivity detection is preferred. In contrast, sensitivities of large cations and anions with low electrophoretic mobilities are much better with ICD.

One of the early reports of ICD in CE is by Gallagher and Danielson.[189] To detect low-mobility anionic surfactants (C_6–C_{12} SO_3^-) and cationic surfactants (tetraethyl-, tetrabutyl-, 2-chloroethytrimethyl-, and tetrahexylammonium), a wide range of inorganic cations and anions were

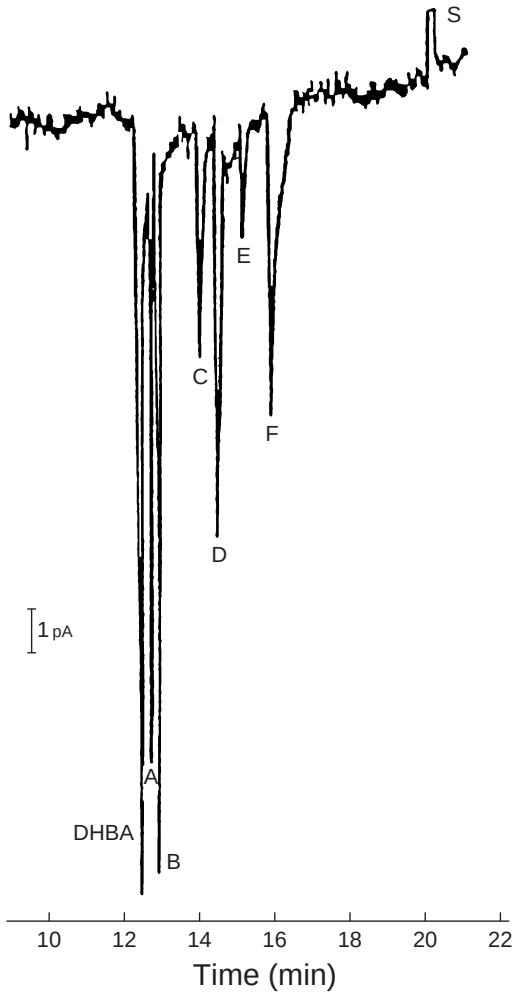

Figure 35 Electropherogram of amino acids and peptides with indirect amperometric detection on a 9-mm i.d. capillary; buffer, 0.01 mM dihydroxybenzylamine–0.025 M morpholinoethanesulfonic acid (pH 5.65); 97-cm separation capillary; 1.0-cm detection capillary; injection for 1 s at 20 kV; separation voltage = 20 kV. Peaks: A = 8.5 fmol Lys; B = 8.2 fmol Arg; C = 7.4 fmol His; D = 7.0 fmol Arg-Leu; E = 6.4 fmol His-Gly; F = 5.8 fmol His-Phe; S = system peak. (Reproduced with permission from Olefirowicz and Ewing.[183])

evaluated as ICD electrolytes. The authors found that the S/N for anionic surfactants using F^- as an ICD electrolyte was about five times larger than those obtained with SCN^-, Br^- or Cl^-. Moreover, the S/N for cationic surfactants using NaCl was higher than when using LiCl. Hence, NaF was found to be the best ICD electrolyte. The simultaneous separation and ICD of both cationic and anionic surfactants using the optimized concentration of 30 mM NaF is shown in Figure 36. For three replicate injections, the RSD of the peak height for cationic surfactant ranged from 0.6 to 2.8%, whereas for anionic surfactants it ranged from 0.07 to 1.1%. A typical detection limit attained by this method was ca. 6 mg L^{-1}

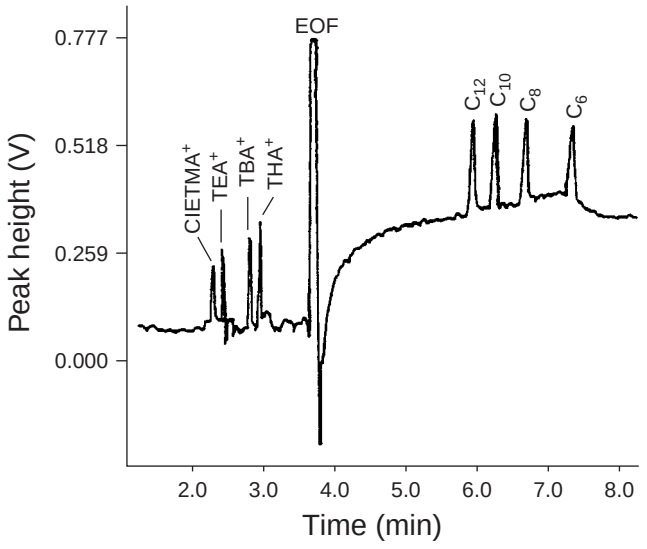

Figure 36 Electropherogram of a 37.5 ppm anionic and cationic surfactant mixture. Electrolyte: 10 mM NaF, 1 mM TEA, 90% water, and 10% methanol. ClETMA, chloroethyltrimethylammonium; TBA, tetrabutylammonium; THA, tetrahexylammonium; EOF, electroosmotic flow. (Reproduced with permission from Gallagher and Danielson.[189])

(31.9 μM for C_6 SO_3^-), which was not as low as that reported for IPD.

Lucy and Wu[190] recently confirmed that ICD is most sensitive for low to medium-mobility analytes. For example, using bromide, acetate, and formate counterions, ICD after CZE separation of C_6–C_{12} SO_3^- was possible. Lucy and Wu also showed that the response of conductivity detection is linear for more than two orders of magnitude (5×10^{-6}–5×10^{-4}), with C_{LOD} as low as 1 μM for C_6 SO_3^-. Owing to similarities in conductivity and electrophoretic mobility, a response equation was developed and applied to show that the possibility exists for a universal calibration in conductivity detection.

Very recently, Fracassi da Silva and do Lago designed an oscillometric conductivity detector for CE.[191] Using this detector, ICD of quaternary ammonium salts was accomplished with potassium acetate as the CZE buffer. The peak sensitivity was found to be proportional to the migration time, e.g. tetramethylammonium ion eluted first and tetrabutylammonium ion eluted last, providing C_{LOD} values of 10.4 and 4.2 μM, respectively.

ACKNOWLEDGMENTS

This work was supported by the Georgia State University Research Program Starter Grant for New Faculty.

ABBREVIATIONS AND ACRONYMS

AHNS	7-Amino-4-hydroxy-2-naphthalenesulfonate
AMP	Adenosine 5′-Monophosphate
AMPA	Aminomethylphosphonic Acid
ANT	Anthraquinone-2-carboxylic Acid
ATMP	Aminotri(methylenephosphonic Acid)
ATP	Adenosine 5′-Triphosphate
BTA	1,2,4,5-Benzenetetracarboxylic Acid
BTMACl	Benzyltrimethylammonium chloride
C_{LOD}	Concentration Limit of Detection
CE	Capillary Electrophoresis
CGE	Capillary Gel Electrophoresis
CHES	2-(*N*-Cyclohexylamino)ethanesulfonic Acid
CMP	Cytidine 5′-Monophosphate
CTAB	Cetyltrimethylammonium bromide
CZE	Capillary Zone Electrophoresis
CZE/IFD	Capillary Zone Electrophoresis/Indirect Fluorescence Detection
CZE/IPD	Capillary Zone Electrophoresis/Indirect Photometric Detection
DBA	*N*, *N*-Dimethylbenzylamine
DBS	Dodecylbenzenesulfonate
DDP	Dimethyldiphenylphosphonium Ion
DETA	Diethylenetriamine
DETPMP	Diethylenetriaminepenta(methylenephosphonic Acid)
DHBA	3,4-Dihydroxybenzylamine
DMAB	4-(*N*,*N*′-Dimethylamino)benzoate
DNBSA	3,5-Dinitrobenzenesulfonic acid
DR	Dynamic Reserve
EDTMP	Ethylenediaminetetra(methylenephosphonic Acid)
FMN	Flavin Mononucleotide
GMP	Guanosine 5′-Monophosphate
HDTMP	Hexamethylenediaminetetra(methylenephosphonic Acid)
HEDP	1-Hydroxyethylidene-1,1-diphosphonic Acid
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICD	Indirect Conductivity Detection
ICLD	Indirect Chemiluminescence Detection
IFD	Indirect Fluorescence Detection
IFD/MEKC	Indirect Fluorescence Detection/ Micellar Electrokinetic Chromatography
ILIFD	Indirect Laser-induced Fluorescence Detection
IPD	Indirect Photometric Detection
KRF	Kohlrausch's Regulating Function

LOD	Limit of Detection
MBA	4-Methylbenzylamine
MEKC	Micellar Electrokinetic Chromatography
MTAB	Myristoyltrimethylammonium bromide
NAA	1-Naphthylacetic Acid
NDC	Naphthalenedicarboxylic Acid
NDS	Naphthalenedisulfonate
NDSA	1-Naphthol-3,6-disulfonic Acid
NMS	Naphthalenemonosulfonate
NTS	Naphthalenetrisulfonate
OBS	Octylbenzenesulfonate
P_n	Polyphosphates
PAS	*p*-Aminosalicylate
PDA	2,6-Pyridinedicarboxylic Acid
PMA	Pyromellitic Acid
PNA	*N*-phenylnaphthylamine
PPA	Phenylphosphonic Acid
PTS	*p*-Toluene Sulfonate
PTSA	*p*-Toluenesulfonic Acid
RP_n	Polyphosphonates
RSD	Relative Standard Deviation
SDS	Sodium Dodecylsulfate
S/N	Signal-to-noise Ratio
TEA	Tetraethylammonium
TNBSA	Trinitrobenzenesulfonic Acid
TR	Transfer Ratio
TTAB	Tetradecyltrimethylammonium bromide
UMP	Uridine 5′-Monophosphate
UV	Ultraviolet
2-AP	2-Aminopyridine
8-HQSA	8-Hydroxyquinoline-5-sulfonic Acid
α-HIBA	α-Hydroxybutyric Acid

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of • Organic Analysis in Environmental Samples by Capillary Electrophoresis

Food **(Volume 5)**
Lipid Analyses in Food

Forensic Science **(Volume 5)**
Capillary Ion Electrophoresis in Forensic Science

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Ion Chromatography

REFERENCES

1. S. Hjerten, K. Elenbring, F. Kilar, J. Liao, A.J.C. Chen, C.J. Siebert, M. Zhu, 'Carrier-free Zone Electrophoresis, Displacement Electrophoresis and Isoelectric Focusing in a High-performance Electrophoresis Apparatus', *J. Chromatogr.*, **403**, 47–61 (1987).
2. F. Foret, S. Fanali, L. Ossicini, 'Indirect Photometric Detection in Capillary Zone Electrophoresis', *J. Chromatogr.*, **470**, 299–308 (1989).
3. W.R. Jones, P. Jandik, 'New Methods for Chromatographic Separations of Anions', *Am. Lab.*, **22**, 51–64 (1990).
4. W.R. Jones, P. Jandik, 'Controlled Changes of Selectivity in the Separation of Ions by Capillary Electrophoresis', *J. Chromatogr.*, **546**, 445–458 (1991).
5. P. Jandik, W.R. Jones, 'Optimization of Detection Sensitivity in the Capillary Electrophoresis of Inorganic Anions', *J. Chromatogr.*, **546**, 431–443 (1991).
6. P. Jandik, W.R. Jones, A. Weston, P.R. Brown, 'Electrophoretic Capillary Ion Analysis: Origins, Principles, and Applications', *LC-GC Mag. Sep. Sci.*, **9**, 634–646 (1991).
7. J. Romano, P. Jandik, W.R. Jones, P.E. Jackson, 'Optimization of Inorganic Capillary Electrophoresis for the Analysis of Anionic Solutes in Real Samples', *J.Chromatogr.*, **546**, 411–421 (1991).
8. W.R. Jones, P. Jandik, 'Various Approaches to Analysis of Difficult Sample Matrices of Anions using Capillary Ion Electrophoresis', *J. Chromatogr.*, **608**, 385–393 (1992).
9. D. Kaniansky, M. Masar, J. Marak, R. Bodor, 'Capillary Electrophoresis of Inorganic Anions', *J. Chromatogr., A*, **834**, 133–178 (1999).
10. P. Doble, P.R. Haddad, 'Indirect Photometric Detection of Anions in Capillary Electrophoresis', *J. Chromatogr., A*, **834**, 189–212 (1999).
11. V. Pacakova, P. Coufal, K. Stulik, 'Capillary Electrophoresis of Inorganic Cations', *J. Chromatogr., A*, **834**, 257–275 (1999).
12. W. Buchberger, S.M. Cousins, P.R. Haddad, 'Optimization of Indirect UV Detection in Capillary Zone Electrophoresis of Low Molecular-mass Anions', *Trends Anal. Chem.*, **13**, 313–319 (1994).
13. Z. El Rassi, 'Capillary Electrophoresis of Carbohydrates', in *Advances in Chromatography*, eds. P.R.

Brown, E. Gruska, Marcel Dekker, New York, 177–250, Vol. 34, 1994.
14. E.S. Yeung, 'Indirect Detection Methods: Looking for What is Not There', *Acc. Chem. Res.*, **22**, 125–130 (1989).
15. M. Macka, P.R. Haddad, 'Determination of Metal Ions by Capillary Electrophoresis', *Electrophoresis*, **18**, 2482–2501 (1997).
16. P. Jandik, G. Bonn, 'Instrumentation for Capillary Electrophoresis', in *Capillary Electrophoresis of Small Molecules and Ions*, eds. P. Jandik, G. Bonn, VCH Publishers, New York, 236–237, 140–143, 1993.
17. Y. Xue, E.S. Yeung, 'Double-beam Laser Indirect Absorption Detection in Capillary Electrophoresis', *Anal. Chem.*, **65**, 2923–2927 (1993).
18. M.T. Ackerman, F.M. Everaerts, J.L. Beckers, 'Quantitative Analysis in Capillary Zone Electrophoresis with Conductivity and Indirect UV Detection', *J. Chromatogr.*, **549**, 345–355 (1991).
19. M.W.F. Nielen, 'Quantitative Aspect of Indirect UV Detection in Capillary Zone Electrophoresis', *J. Chromatogr.*, **588**, 321–326 (1991).
20. S.M. Cousins, P.R. Haddad, W. Buchberger, 'Evaluation of Carrier Electrolytes for Capillary Zone Electrophoresis of Low Molecular-mass Anions with Indirect UV Detection', *J. Chromatogr., A*, **671**, 397–402 (1994).
21. C.W. Klampfl, M.U. Katzmayr, W. Buchberger, 'Separation of Inorganic and Organic Anions by Capillary Zone Electrophoresis with Simultaneous Indirect UV and Conductivity Detection', *Electrophoresis*, **19**, 2459–2464 (1998).
22. P. Doble, M. Macka, P.R. Haddad, 'Buffered Chromate Electrolyte for Separation and Indirect Absorbance Detection of Inorganic Anions in Capillary Electrophoresis', *Anal. Commun.*, **34**, 351–353 (1997).
23. A.E. Vorndran, P.J. Oefner, H. Scherz, G.K. Bonn, 'Indirect UV Detection of Carbohydrates in Capillary Zone Electrophoresis', *Chromatographia*, **33**, 163–168 (1992).
24. S.A. Shamsi, N.D. Danielson, 'Ribonucleotide Electrolytes for Capillary Electrophoresis of Polyphosphates and Polyphosphonates with Indirect Photometric Detection', *Anal. Chem.*, **67**, 1845–1852 (1995).
25. W. Beck, H. Engelhardt, 'Capillary Electrophoresis of Organic and Inorganic Cations with Indirect UV-detection', *Chromatographia*, **33**, 313–316 (1992).
26. A. Weston, P.R. Brown, P. Jandik, A.L. Heckenberg, W.R. Jones, 'Optimization of Detection Sensitivity in the Analysis of Organic Cations by Capillary Ion Electrophoresis', *J. Chromatogr.*, **608**, 395–402 (1992).
27. Z. Mala, R. Vespalec, P. Bocek, 'Capillary Zone Electrophoresis with Indirect Photometric Detection in the Visible Range', *Electrophoresis*, **15**, 1526–1530 (1994).
28. F. Steiner, W. Beck, H. Engelhardt, 'Optimization of Indirect Ultraviolet Detection in High Performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr.*, **738**, 11–23 (1996).
29. S. Motellier, K. Gurdale, H. Pitsch, 'Sulfur Speciation by Capillary Electrophoresis with Indirect Spectrophotometric Detection: In Search of a Suitable Carrier Electrolyte to Maximize Sensitivity', *J. Chromatogr., A*, **770**, 311–319 (1997).
30. M.P. Harold, M.J. Wojtusik, J. Riviello, P. Henson, 'Parameters Influencing Separation and Detection of Anions by Capillary Electrophoresis', *J. Chromatogr.*, **640**, 463–471 (1993).
31. Y.F. Ma, R.L. Zhang, 'Optimization of Indirect Photometric Detection of Anions in High-performance Capillary Electrophoresis', *J. Chromatogr.*, **625**, 341–348 (1992).
32. F.E.P. Mikkers, F.M. Everaerts, T.P.E.M. Verheggen, 'Concentration Distribution in Zone Electrophoresis', *J. Chromatogr.*, **169**, 1–10 (1979).
33. F.E.P. Mikkers, F.M. Everaerts, T.P.E.M. Verheggen, 'High Performance Zone Electrophoresis', *J. Chromatogr.*, **169**, 11–20 (1979).
34. W.R. Jones, 'Electrophoretic Capillary Ion Analysis', in *Handbook of Capillary Electrophoresis*, ed. J.P. Landers, CRC Press, Boca Raton, FL, 216–217.
35. S.A. Shamsi, N.D. Danielson, 'Naphthalenesulfonates as Electrolytes for Capillary Electrophoresis of Inorganic Anions, Organic Acids and Aliphatic Anionic Surfactants', *Anal. Chem.*, **66**, 3757–3764 (1994).
36. J.R. Mazzeo, 'Capillary Electrophoresis of Inorganic Ions', in *High Performance Capillary Electrophoresis*, ed. M.J. Khaledi, Wiley Interscience, New York, 825–862, 1998.
37. S. Hjerten, 'Zone Broadening in Electrophoresis with Special Reference to High-performance Capillary Electrophoresis in Capillaries: An Interplay between Theory and Practice', *Electrophoresis*, **11**, 665–690 (1990).
38. T. Hirokawa, S. Kobayashi, Y. Kiso, 'Complex Forming Equilibria in Isotachophoresis: Evaluation of the Ion-pair Formation Constants of Phosphorus Oxo Acids with Histidine and Assessment of their Separability', *J. Chromatogr.*, **410**, 279–285 (1987).
39. A. Henshall, M.P. Harold, J.M.Y. Tso, 'Separation of Inositol Phosphates by Capillary Electrophoresis', *J. Chromatogr., A*, **608**, 413–419 (1992).
40. L.K. Goebel, H.M. McNair, H.T. Rasmussen, B.P. McPherson, 'Separation of Ethoxylated Alcohol Sulfates by Capillary Electrophoresis using Indirect UV-detection', *J. Microcol. Sep.*, **5**, 47–50 (1993).
41. B.A.P. Buscher, U.R. Tjaden, H. Irth, E.M. Andersson, J. Van der Greef, 'Determination of 1,2,6-Inositol Triphosphate Derivatives in Plasma using Iron(III)-loaded Adsorbents and Capillary Zone Electrophoresis–Indirect UV Detection', *J. Chromatogr., A*, **718**, 413–419 (1995).

42. R. Neubert, K. Raith, J. Schiewe, 'Capillary Zone Electrophoresis in Skin Fatty Acids Analysis', *Pharmazie*, **52**, 212–215 (1997).
43. S.A. Oehrle, P.C. Bossle, 'Analysis of Nerve Agent Degradation Products using Capillary Ion Electrophoresis', *J. Chromatogr., A*, **692**, 247–252 (1995).
44. T. Wang, R.A. Hartwick, 'Binary Buffers for Indirect Absorption Detection in Capillary Zone Electrophoresis', *J. Chromatogr., A*, **589**, 307–313 (1992).
45. R. Roldan-Assad, P. Gariel, 'Capillary Zone Electrophoretic Determination of C_2-C_{18} Linear Saturated Free Fatty Acids with Indirect Absorbance Detection', *J. Chromatogr., A*, **708**, 339–350 (1995).
46. E.V. Dose, G.A. Guichon, 'High Resolution Modeling of Capillary Zone Electrophoresis and Isotacophoresis', *Anal. Chem.*, **63**, 1063–1072 (1991).
47. T. Wang, R.A. Hartwick, 'Noise and Detection Limits of Indirect Absorption Detection in Capillary Zone Electrophoresis', *J. Chromatogr.*, **607**, 119–125 (1992).
48. T. Ehmann, K. Backmann, L. Fabry, H. Reifer, S. Pahlke, L. Kotz, 'Optimization of Electrokinetic Sample Introduction in Capillary Electrophoresis for the Ultratrace Determination of Anions on Silicon Wafer Surfaces', *Chromatographia*, **45**, 301–311 (1997).
49. M.M. Rhemrev Boom, 'Determination of Anions with Capillary Electrophoresis and Indirect Ultraviolet Detection', *J. Chromatogr., A*, **680**, 675–684 (1994).
50. M.J. Wojtusik, M.P. Harrold, 'Factors Influencing Trace Ion Analysis with Preconcentration by Electrostacking', *J. Chromatogr., A*, **671**, 411–417 (1994).
51. A.E. Vorndran, E. Grill, C. Huber, P.J. Oefner, G.K. Bonn, 'Capillary Zone Electrophoretic Analysis of Carbohydrates by Direct and Indirect UV Detection', *Chromatographia*, **34**, 308–316 (1992).
52. S.A. Shamsi, 'Reversed-phase Ion Chromatography and Capillary Electrophoresis of Ionic Compounds', PhD Dissertation, Miami University, 1995.
53. W.R. Jones, 'Method Development Approaches for Capillary Ion Electrophoresis', *J. Chromatogr.*, **640**, 387–395 (1993).
54. Y. Xue, E.S. Yeung, 'On-column Double-beam Laser Absorption Detection for Capillary Electrophoresis', *Anal. Chem.*, **65**, 1988–1993 (1993).
55. W. Tong, E.S. Yeung, 'Simple Double Beam Absorption Detection System for Capillary Electrophoresis', *J. Chromatogr., A*, **718**, 177–185 (1995).
56. S.J. Williams, E.T. Bergström, D.M. Goodall, H. Kawazumi, K.P. Evans, 'Diode Laser-based Indirect Absorbance Detection for Anions in Capillary Electrophoresis', *J. Chromatogr.*, **636**, 39–45 (1995).
57. P.E. Andersson, W.D. Pfeffer, L.G. Blomberg, 'Indirect Detection in Capillary Electrophoresis: Comparison between Indirect UV and Indirect Laser-induced Fluorescence Detection for the Determination of Isoprenyl Pyrophosphates', *J. Chromatogr., A*, **699**, 323–330 (1995).
58. F. Steiner, W. Beck, H. Engelhardt, 'Optimization of Indirect Ultraviolet Detection in High Performance Liquid Chromatography and Capillary Electrophoresis', *J. Chromatogr., A*, **738**, 11–23 (1996).
59. P. Doble, P.R. Haddad, 'Indirect Photometric Detection of Anions in Capillary Electrophoresis', *J. Chromatogr., A*, **834**, 189–212 (1999).
60. W. Buchberger, P.R. Haddad, 'Separation of Metallocyanide Complexes by Capillary Electrophoresis', *J. Chromatogr., A*, **687**, 343–349 (1994).
61. C.H. Wu, Y.S. Lo, Y.-H. Lee, T.-I. Lin, 'Capillary Electrophoretic Determination of Organic Acids with Indirect Detection', *J. Chromatogr., A*, **716**, 291–301 (1995).
62. M. Van Holderbeke, H. Vanhoe, L. Moens, R. Dams, 'Determination of Ten Inorganic Anions in Drinking and Wastewaters with Capillary Zone Electrophoresis via Indirect UV Detection', *Biomed. Chromatogr.*, **9**, 281–282 (1995).
63. H. Siren, A. Maattanen, M.L. Riekola, 'Determination of Small Anions by Capillary Electrophoresis using Indirect UV Detection with Sulfonated Nitrosonaphthol Dyes', *J. Chromatogr., A*, **767**, 293–301 (1997).
64. A.R. Timerbaev, 'Strategies for Selectivity Control in Capillary Electrophoresis of Metal Species', *J. Chromatogr., A*, **792**, 495–518 (1997).
65. E. Dabek-Zlotorzynska, E.P.C. Lai, A.R. Timberbaev, 'Capillary Electrophoresis: The State-of-the-art in Metal Speciation System', *Anal. Chim. Acta*, **359**, 1–26 (1998).
66. T.-I. Lin, Y.-H. Lee, Y.-C. Chen, 'Capillary Electrophoretic Analysis of Inorganic Cations', *J. Chromatogr., A*, **654**, 167–176 (1993).
67. Q. Yang, J. Smeyers-Verbeke, W. Wen, M.S. Khots, D.L. Massart, 'Simultaneous Separation of Ammonium and Alkali, Alkaline Earth and Transition Metal Ions in Aqueous–Organic Media by Capillary Ion Analysis', *J.Chromatogr., A*, **688**, 339–349 (1994).
68. C. Francois, Ph. Morin, M. Dreux, 'Effect of the Concentration of 18-Crown-6 Added to the Electrolyte upon the Separation of Ammonium, Alkali and Alkaline-earth Cations by Capillary Electrophoresis', *J. Chromatogr., A*, **706**, 535–553 (1995).
69. C. Quang, M.G. Khaledi, 'Prediction and Optimization of the Separation of Inorganic Metal Cations by Capillary Ion Electrophoresis', *J. Chromatogr.*, **659**, 459–466 (1994).
70. A. Weston, P.R. Brown, A.L. Heckenberg, P. Jandik, W.R. Jones, 'Effect of Electrolyte Composition on the Separation of Inorganic Metal Cations by Capillary Ion Electrophoresis', *J. Chromatogr.*, **602**, 249–256 (1992).
71. A. Weston, P.R. Brown, P. Jandik, W.R. Jones, A.L. Heckenberg, 'Factors Affecting the Separation of Inorganic Metal Cations by Capillary Electrophoresis', *J. Chromatogr.*, **593**, 289–295 (1992).
72. Y. Shi, J.S. Fritz, 'Separation of Metal Ions by Capillary Electrophoresis with a Complexing Agent', *J. Chromatogr., A*, **640**, 473–479 (1993).

73. M. Chen, R.M. Cassidy, 'Separation of Metal Ions by Capillary Electrophoresis', *J. Chromatogr., A*, **640**, 425–431 (1995).
74. Y.-H. Lee, T.-I. Lin, 'Determination of Metal Cations by Capillary Electrophoresis: Effect of Background Carrier and Complexing Agents', *J. Chromatogr., A*, **675**, 227–236 (1994).
75. T. Wang, S.F.Y. Li, 'Migration Behavior of Alkali and Alkaline-earth Metal Ion–EDTA Complexes and Quantitative Analysis of Magnesium in Real Samples by Capillary Electrophoresis with Indirect Ultraviolet Detection', *J. Chromatogr., A*, **707**, 343–353 (1995).
76. Y. Shi, J.S. Fritz, 'New Electrolyte Systems for the Determination of Metal Cations by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **671**, 429–435 (1994).
77. Q. Yang, M. Jimidar, T.P. Hamoir, J. Smeyers-Verbeke, D.L. Massart, 'Determination of Alkali and Alkaline Earth Metals in Real Samples by Capillary Ion Electrophoresis', *J. Chromatogr.*, **673**, 275–285 (1994).
78. W.R. Barger, R.L. Mowery, J.R. Wyatt, 'Separation and Indirect Detection by Capillary Zone Electrophoresis of ppb (w/w) Levels of Aluminum Ions in Solutions of Multiple Cations', *J. Chromatogr., A*, **680**, 659–665 (1994).
79. E. Simunicova, D. Kaniansky, K. Loksikova, 'Separation of Alkali and Alkaline Earth Metal and Ammonium Cations by Capillary Zone Electrophoresis with Indirect UV Absorbance Detection', *J. Chromatogr., A*, **665**, 203–209 (1994).
80. J.M. Riviello, M.P. Harrold, 'Capillary Electrophoresis of Inorganic Cations and Low-molecular Mass Amines using a Copper-based Electrolyte with Indirect UV Detection', *J. Chromatogr., A*, **652**, 385–392 (1993).
81. M.J. Wojtusik, M.P. Harrold, 'Factors Influencing Trace Ion Analysis with Preconcentration by Electrostacking', *J. Chromatogr., A*, **671**, 411–417 (1994).
82. K. Chang, F.R. Nordmeyer, J.D. Lamb, 'A Novel Buffer System for Separation of Metal Cations by Capillary Electrophoresis with Indirect UV-detection', *J. Capill. Electrophor.*, **2**(6), 279–285 (1995).
83. R.A. Carpio, P. Jandik, E. Fallon, 'Capillary Electrophoresis Methods for the Trace Cation Analysis of Semiconductor Grades of Hydrogen Peroxide', *J. Chromatogr., A*, **657**, 185–191 (1993).
84. E. Dabek-Zlotorzynska, J.F. Dlouhy, 'Capillary Zone Electrophoresis with Indirect UV Detection of Organic Anions using 2,6-Naphthalenedicarboxylate', *J. Chromatogr., A*, **685**, 145–153 (1994).
85. G.W. Tindall, D.R. Wilder, R.I. Perry, 'Optimizing Dynamic Range for the Analysis of Small Ions by Capillary Zone Electrophoresis', *J. Chromatogr.*, **641**, 163–167 (1993).
86. H. Chen, Y. Xu, F. Van Lente, M.P.C. Ip, 'Indirect Ultraviolet Detection of Biologically Relevant Organic Acids by Capillary Electrophoresis', *J. Chromatogr., B*, **679**, 49–59 (1996).
87. E. Dabek-Zlotorzynska, J.F. Dlouky, 'Application of Capillary Electrophoresis in Atmospheric Aerosols Analysis: Determination of Inorganic and Organic Anions', *J. Chromatogr., A*, **671**, 389–395 (1994).
88. D. Volgger, A.J. Zemann, G.K. Bonn, M.J. Antal, 'High Speed Separation of Carboxylic Acids by Co-electroosmotic Capillary Electrophoresis with Direct and Indirect UV Detection', *J. Chromatogr., A*, **758**, 263–276 (1997).
89. T. Soga, G.A. Ross, 'Capillary Electrophoretic Determination of Inorganic and Organic Anions using 2,6-Pyridinedicarboxylic Acid: Effect of Electrolyte's Complexing Ability', *J. Chromatogr., A*, **767**, 223–230 (1997).
90. H.-T. Chang, H.-S. Chen, R. Lee, 'Dynamic Control and Indirect Absorption Detection for High-speed Capillary Electrophoretic Separation of Organic Acids', *J. Chromatogr., A*, **800**, 339–344 (1998).
91. D.H. Craston, M. Saeed, 'Analysis of Carboxylic Acid and Related Acids in the Environment by Capillary Electrophoretic Techniques', *J. Chromatogr., A*, **827**, 1–12 (1998).
92. Y. Ma, R. Zhang, L.C. Cooper, 'Indirect Photometric Detection of Polyamines in Biological Samples Separated by High-performance Capillary Electrophoresis', *J. Chromatogr., A*, **608**, 93–96 (1992).
93. L. Arce, A.M. Rios, M. Valcárel, 'Selective and Rapid Determination of Biogenic Amines by Capillary Zone Electrophoresis', *Chromatographia*, **46**, 170–176 (1997).
94. G. Gutnikov, W. Beck, H. Engelhardt, 'Separation of Homologous Fatty Acids by Capillary Electrophoresis', *J. Microcolumn Sep.*, **6**, 565–570 (1994).
95. W. Buchberger, K. Winna, 'Determination of Free Fatty Acids by Capillary Zone Electrophoresis', *Mikrochim. Acta*, **122**, 45–52 (1996).
96. J. Collet, P. Gariel, 'Capillary Zone Electrophoresis Separation of C_{14}–C_{18} Linear Saturated and Unsaturated Free Fatty Acids with Indirect UV-detection', *J. Capill. Electrophor.*, **3**, 77–82 (1996).
97. E. Drange, E. Lundanes, 'Determination of Long Chain Fatty Acids Using Non-aqueous Capillary Electrophoresis with Indirect UV Detection', *J. Chromatogr., A*, **771**, 301–309 (1997).
98. F. Haddadian, S.A. Shamsi, I.M. Warner, 'Separation of Saturated and Unsaturated Free Fatty Acids using Capillary Electrophoresis', *J. Chromatogr. Sci.*, **37**, 100–107 (1999).
99. F.S. Stover, S.S. Keffer, 'Separation of Condensed Phosphates using Capillary Zone Electrophoresis with Indirect UV Detection', *J. Chromatogr., A*, **657**, 450–454 (1993).
100. T. Wang, S.F.Y. Li, 'Separation of Polyphosphates and Polycarboxylates by Capillary Electrophoresis in a Carrier Electrolyte containing Adenosine-5′-triphosphate and Cetyltrimethylammmonium Bromide with Indirect UV Detection', *J. Chromatogr., A*, **723**, 197–205 (1996).

101. A.F. Nassar, S.V. Lucas, W.R. Jones, L.D. Hoffland, 'Separation of Chemical Warfare Agent Degradation Products by the Reversal of the Electroosmotic Flow in Capillary Electrophoresis', *Anal. Chem.*, **70**, 1085–1091 (1998).
102. M.M. Rhemrev-Boom, 'Determination of Anions with Capillary Electrophoresis and Indirect Ultraviolet Detection', *J. Chromatogr., A*, **680**, 675–684 (1994).
103. M.T. Galceran, L. Puigneou, M. Diez, 'Comparison of Different Electroosmotic Flow Modifiers in the Analysis of Inorganic Anions by Capillary Electrophoresis', *J. Chromatogr., A*, **732**, 167–174 (1996).
104. S.A. Oehrle, 'Analysis of Anions in Drinking Water by Capillary Ion Electrophoresis', *J. Chromatogr., A*, **733**, 101–104 (1996).
105. T. Wang, S.F.Y. Li, 'Development of a Capillary Electrophoresis Method with Indirect Photometric Detection for the Determination of Anions Related to Anticaries', *J. Chromatogr., A*, **781**, 457–466 (1997).
106. F.S. Stover, 'Capillary Electrophoresis of Long-chain Polyphosphates', *J. Chromatogr., A*, **769**, 349–351 (1997).
107. T. Wang, S.F.Y. Li, 'Separation of Synthetic Inorganic Polymers of Condensed Phosphates by Capillary Gel Electrophoresis with Indirect Photometric Detection', *J. Chromatogr., A*, **802**, 159–165 (1998).
108. B.A.P. Buscher, H. Irth, E. Andersson, U.R. Tjaden, J. Van der Greef, 'Determination of Inositol Phosphates in Fermentation Broth using Capillary Zone Electrophoresis with Indirect UV-detection', *J. Chromatogr., A*, **678**, 145–150 (1994).
109. S.A. Shamsi, N.D. Danielson, 'Capillary Zone Electrophoresis of Phosphate Ester Surfactants with Adenosine Monophosphate and Indirect Photometric Detection', *J. Chromatogr., A*, **737**, 315–324 (1996).
110. G.A. Pianetti, M. Taverna, A. Baillet, G. Mahuzier, D. Baylocq Ferrier, 'Determination of Alkyl Phosphonic Acids by Capillary Zone Electrophoresis using Indirect Detection', *J. Chromatogr., A*, **630**, 371–377 (1993).
111. M.G. Cikalo, D.M. Goodall, W. Matthew, 'Analysis of Glyphosate using Capillary Electrophoresis with Indirect Detection', *J. Chromatogr., A*, **745**, 189–200 (1996).
112. A.F. Nassar, S.V. Lucas, C.A. Myles, W.R. Jones, M. Campisano, L.D. Hoffland, 'Quantitative Analysis of Chemical Warfare Agent Degradation Products in Reaction Masses using Capillary Electrophoresis', *Anal. Chem.*, **70**, 3598–3604 (1998).
113. S.A. Shamsi, N.D. Danielson, 'Individual and Simultaneous Class Separations of Cationic and Anionic Surfactants using Capillary Electrophoresis with Indirect Photometric Detection', *Anal. Chem.*, **67**, 4210–4216 (1995).
114. S.A. Shamsi, N.D. Danielson, 'Capillary Electrophoresis of Cationic Surfactants with Tetrazolium Violet and Anionic Surfactants with Adenosine Monophosphate and Indirect Photometric Detection', *J. Chromatogr., A*, **739**, 405–412 (1996).
115. K. Heineg, C. Vogt, G. Werner, 'Separation of Anionic Surfactants using Aqueous and Nonaqueous Capillary Electrophoresis', *J. Capill. Electrophor.*, **3**, 261–270 (1996).
116. S. Chen, D.J. Pietrzyk, 'Separation of Sulfonated and Sulfated Surfactants by Capillary Electrophoresis: Effect of Buffer Cation', *Anal. Chem.*, **56**, 2770–2775 (1993).
117. J.M. Gibbons, S.H. Hoke, 'Capillary Electrophoresis with Indirect UV-detection Determination of Sodium Dodecylsulfate in Simulated Stream Water', *J. High Resolut. Chromatogr.*, **17**, 665–667 (1994).
118. H. Salimi-Moosavi, R.M. Cassidy, 'Application of Nonaqueous Capillary Electrophoresis to the Separation of Long-chain Surfactants', *Anal. Chem.*, **68**, 293–299 (1996).
119. A. Monclus, M. Rodriguez, J. Masana, 'The Analysis of Anionic Surfactants in Toiletries using Capillary Electrophoresis', *LC/GC Int.*, **9**, 79–87 (1996).
120. C.S. Weiss, J.S. Hazlett, H.M. Datta, M.H. Danzer, 'Determination of Quaternary Ammonium Compounds by Capillary Electrophoresis using Direct and Indirect UV Detection', *J. Chromatogr., A*, **608**, 325–332 (1992).
121. K. Heinig, C. Vogt, G. Werner, 'Separation of Ionic and Neutral Surfactants by Capillary Electrophoresis and High-performance Liquid Chromatography', *J. Chromatogr., A*, **745**, 281–292 (1996).
122. S. Honda, S. Iwase, A. Makino, S. Fujiwara, 'Simultaneous Determination of Reducing Monosaccharides by Capillary Zone Electrophoresis as the Borate Complexes of *N*-2-Pyridylglycamines', *Anal. Biochem.*, **176**, 72–77 (1989).
123. P.J. Oefner, A.E. Vorndran, E. Grill, C. Huber, G.K. Bonn, 'Capillary Zone Electrophoresis Analysis of Carbohydrates by Direct and Indirect UV-detection', *Chromatographia*, **34**, 308–316 (1992).
124. L.J. Basa, M.W. Spellman, 'Analysis of Glycoprotein Derived Oligosaccharides by High-pH Anion Exchange Chromatography', *J. Chromatogr.*, **499**, 205–220 (1990).
125. R.R. Townsend, M.R. Hardy, D.A. Cumming, J.P. Carver, P. Bendiak, 'Separation of Branched Sialylated Oligosaccharides using High-pH Anion Exchange Chromatography with Pulsed Amperometric Detection', *Anal. Biochem.*, **182**, 1–11 (1989).
126. A. Klockow, A. Paulus, V. Figueiredo, R. Amado, H.M. Widmer, 'Determination of Carbohydrates in Fruit Juices by Capillary Electrophoresis and High-performance Liquid Chromatography', *J. Chromatogr., A*, **680**, 187–200 (1994).
127. Y.-H. Lee, T.-I. Lin, 'Determination of Carbohydrates by High-performance Capillary Electrophoresis with Indirect Absorbance Detection', *J. Chromatogr., B*, **681**, 87–97 (1996).

128. B. Lu, D. Westerlund, 'Indirect UV Detection of Carbohydrates in Capillary Zone Electrophoresis by using Tryptophan as a Marker', *Electrophoresis*, **17**, 325–332 (1996).
129. X. Xu, W.Th. Kok, H. Poppe, 'Sensitive Determination of Sugars by Capillary Zone Electrophoresis with Indirect UV Detection under Highly Alkaline Conditions', *J. Chromatogr., A*, **716**, 231–240 (1995).
130. Y. Liu, C. Shu, J.D. Lamb, 'High Performance Capillary Electrophoretic Separation of Carbohydrates with Indirect UV Detection using Diethylamine and Borate as Electrolyte Additives', *J. Capill. Electrophor.*, **4**, 97–103 (1997).
131. J.B.L. Damm, G.T. Overklift, 'Indirect UV Detection as a Non-selective Detection Method in the Qualitative and Quantitative Analysis of Heparin Fragments by High-performance Capillary Electrophoresis', *J. Chromatogr., A*, **678**, 151–165 (1994).
132. A. Nardi, S. Fanali, F. Foret, 'Capillary Zone Electrophoresis of Cyclodextrin with Indirect-UV Detection', *Electrophoresis*, **11**, 774–779 (1990).
133. R.J. Tait, D.J. Skanchy, D.P. Thompson, N.C. Chetwyn, D.A. Dunshee, R.A. Rajewski, V.J. Stella, J.F. Stobaugh, 'Characterization of Sulfoalkyl Ether Derivatives of β-CD by Capillary Electrophoresis with Indirect UV Detection', *J. Pharm. Biomed. Anal.*, **10**, 615–622 (1992).
134. K.L. Larsen, F. Mathiesen, W. Zimmerman, 'Separation and Analysis of Cyclodextrins by Capillary Zone Electrophoresis', *Carbohydr. Res.*, **298**, 59–63 (1997).
135. Y.-H. Lee, T.-I. Lin, 'Capillary Electrophoretic Analysis of Cyclodextrins and Determination of Formation Constants for Inclusion Complexes', *Electrophoresis*, **17**, 333–340 (1996).
136. J.P. Schaeper, S.A. Shamsi, N.D. Danielson, 'Separation of Phosphorylated Sugars using Capillary Electrophoresis with Indirect Photometric Detection', *J. Capill. Electrophor.*, **3**, 215–221 (1996).
137. Y.-H. Lee, T.-I. Lin, 'Capillary Electrophoretic Determination of Amino Acids with Indirect Absorbance Detection', *J. Chromatogr., A*, **680**, 287–297 (1994).
138. Y.-H. Lee, T.-I. Lin, 'Capillary Electrophoretic Determination of Amino Acids: Improvement by Cyclodextrin Additives', *J. Chromatogr., A*, **716**, 335–346 (1995).
139. H. Chen, Y. Xu, M.P.C. Ip, 'Determination of Amino Acids in Urine by Capillary Electrophoresis with Indirect UV Detection', *J. Liq. Chromatogr. Relat. Technol.*, **20**, 2475–2493 (1997).
140. W. Hong, G. Junling, F. Ruonong, H. Hanfang, D. Tiankei, 'Capillary Zone Electrophoresis of Native Amino Acids with Indirect Ultraviolet Detection', *J. Beijing Inst. Technol. (Engl. Ed.)*, **7**, 43–48 (1998).
141. S.P.D. Lalljie, G. Murugan, H.J. Dumont, P. Sandra, 'Determination of Ions, Amino Acids/Amines in the Hemolymph of Fairy Shrimp by Capillary Electrophoresis with Indirect UV and Laser-induced Fluorescence Detection', *Hydrobiologia*, **319**, 103–109 (1996).
142. M.W. Crowder, A. Numan, F. Haddadian, M.A. Weitzel, N.D. Danielson, 'Capillary Electrophoresis of Phosphonoamino Acids with Indirect Photometric Detection', *Anal. Chim. Acta*, **384**, 127–133 (1999).
143. L. Ingvardsen, L.S. Michaelsen, H. Soresen, 'Analysis of Individual Phospholipids by High Performance Capillary Electrophoresis', *J. Am. Oil Chem. Soc.*, **71**, 183–187 (1994).
144. R. Szucs, K. Verleysen, G.S.M.J.E. Duchateau, P. Sandra, B.G.M. Vandeginste, 'Analysis of Phospholipids in Lecithins: Comparison between Micellar Electrokinetic Chromatography and High Performance Liquid Chromatography', *J. Chromatogr., A*, **738**, 25–29 (1996).
145. F. Haddadian, S.A. Shamsi, J.P. Schaeper, N.D. Danielson, 'Capillary Electrophoresis of Phospholipids with Indirect Photometric Detection', *J. Chromatogr. Sci.*, **36**, 395–400 (1998).
146. K. Baechmann, I. Haumann, T. Groh, 'Simultaneous Determination of Inorganic Cations and Anions in Capillary Zone Electrophoresis with Indirect Fluorescence Detection', *Fresenius' J. Anal. Chem.*, **343**, 901–902 (1992).
147. X. Xiong, S.F.Y. Li, 'Selection and Optimization of Background Electrolytes for Simultaneous Detection of Small Cations and Organic Acids by Capillary Electrophoresis with Indirect Photometry', *J. Chromatogr., A*, **822**, 125–136 (1998).
148. X. Xiong, S.F.Y. Li, 'Dual UV-Absorbing Background Electrolytes for Simultaneous Separation and Detection of Small Cations and Anions by Capillary Zone Electrophoresis', *Electrophoresis*, **19**, 2243–2251 (1998).
149. P. Kuban, B. Karlberg, 'Simultaneous Determination of Small Cations and Anions by Capillary Electrophoresis', *Anal. Chem.*, **70**, 360–365 (1998).
150. P. Kubáň, P. Kubáň, V. Kubáň, 'Simultaneous Capillary Electrophoretic Separation of Small Anions and Cations after Complexation with Ethylenediaminetetraacetic Acid', *J. Chromatogr., A*, **836**, 75–80 (1998).
151. E.S. Yeung, W.G. Kuhr, 'Indirect Detection Methods for Capillary Separations', *Anal. Chem.*, **63**, 275–283 (1991).
152. E.S. Yeung, 'Optical Detectors for Capillary Electrophoresis', in *Advances in Chromatography*, eds. P.R. Brown, E. Grushka, Marcel Dekker, New York, 1–51, Vol. 35, 1995.
153. L.N. Amankawa, M. Albin, W.G. Kuhr, 'Fluorescence Detection in Capillary Electrophoresis', *Trends Anal. Chem.*, **11**, 114–120 (1992).
154. E.S. Yeung, 'Optical Detection Schemes for Capillary Electrophoresis', in *Capillary Electrophoresis Technology*, Chromatogr. Sci. Ser., ed. N.A. Guzman, Marcel Dekker, New York, 587–600, Vol. 64, 1993.

155. K. Bächmann, J. Boden, I. Haumann, 'Indirect Fluorometric Detection of Alkali and Alkaline Earth Metal Ions in Capillary Zone Electrophoresis with Cerium(III) as Carrier Electrolyte', *J. Chromatogr., A*, **626**, 259–265 (1992).
156. T. Groh, K. Bächmann, 'Selective Influencing of the Migration Time of SO_4^{2-} by Forming Neutral Species with Pb^{2+} in Capillary Zone Electrophoresis', *Electrophoresis*, **13**, 458–461 (1992).
157. Y.-T. Lee, C.-W. Whang, 'Capillary Electrophoresis of Triorganotin Compounds with Indirect Fluorescence Detection', *J. Chromatogr., A*, **746**, 269–275 (1996).
158. P.L. Desbène, C.J. Morin, A.M. Desbène Monvernay, R.S. Groult, 'Utilization of Fluorescein Sodium Salt in Laser-induced Fluorimetric Detection of Ions Separated by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **689**, 135–148 (1995).
159. T.W. Garner, E.S. Yeung, 'Indirect Fluorescence Detection of Sugars Separated by Capillary Zone Electrophoresis with Visible Laser Excitation', *J. Chromatogr., A*, **515**, 639–644 (1990).
160. S.A. Shamsi, N.D. Danielson, I.M. Warner, 'Flavin Mononucleotide for Indirect Laser-induced Fluorescence Detection of Anions Separated by Capillary Zone Electrophoresis', *J. Chromatogr., A*, **835**, 159–168 (1999).
161. P. Cerletti, 'Properties of Riboflavin Phosphates', *Anal. Chim. Acta*, **20**, 243–250 (1959).
162. W.G. Kuhr, E.S. Yeung, 'Indirect Fluorescence Detection of Native Amino Acids in Capillary Zone Electrophoresis', *Anal. Chem.*, **60**, 1832–1834 (1988).
163. W.G. Kuhr, E.S. Yeung, 'Optimization of Sensitivity and Separation in Capillary Zone Electrophoresis with Indirect Fluorescence Detection', *Anal. Chem.*, **60**, 2642–2646 (1988).
164. L. Gross, E.S. Yeung, 'Indirect Fluorimetric Detection and Quantification in Capillary Zone Electrophoresis of Inorganic Anions and Nucleotides', *J. Chromatogr.*, **480**, 169–178 (1989).
165. L. Gross, E.S. Yeung, 'Indirect Fluorimetric Detection of Cations in Capillary Zone Electrophoresis', *Anal. Chem.*, **62**, 427–431 (1990).
166. Q. Xue, E.S. Yeung, 'Indirect Fluorescence Determination of Lactate and Pyruvate in Single Erythrocytes by Capillary Electrophoresis', *J. Chromatogr., A*, **661**, 287–295 (1994).
167. V. Martí, M. Anguilar, E.S. Yeung, 'Indirect Fluorescence Detection of Free Cyanide and Related Compounds by Capillary Electrophoresis', *J. Chromatogr., A*, **709**, 367–374 (1995).
168. Y.-C. Chao, C.-W. Whang, 'Capillary Zone Electrophoresis of Eleven Priority Phenols with Indirect Fluorescence Detection', *J. Chromatogr., A*, **663**, 229–237 (1994).
169. A.M. Desbène, C.J. Morin, N.L. Mofaddel, R.S. Groult, 'Utilization of Fluorescein Sodium Salt in Laser-induced Indirect Fluorimetric Detection. II. Application to Inorganic Anions', *J. Chromatogr., A*, **716**, 279–290 (1995).
170. T.L. Wang, P.H. Wei, S.F.Y. Li, 'Nonaqueous Capillary Zone Electrophoresis for Separation of Free Fatty Acids with Indirect Fluorescence Detection', *Electrophoresis*, **19**, 2187–2192 (1998).
171. Yu-M. Huang, C.-W. Whang, 'Capillary Electrophoresis of Arsenic Compounds with Indirect Fluorescence Detection', *Electrophoresis*, **19**, 2140–2144 (1998).
172. M.N. Church, J.D. Spear, R.E. Russo, G.L. Klunder, P.M. Grant, B.D. Andersen, 'Transient Isotachophoretic–Electrophoretic Separations of Lanthanides with Indirect Laser-induced Fluorescence Detection', *Anal. Chem.*, **70**, 2475–2480 (1998).
173. T. Higashijima, T. Fuchigami, T. Imasaka, N. Ishibashi, 'Determination of Amino Acids by Capillary Zone Electrophoresis based on Semiconductor Laser Fluorescence Detection', *Anal. Chem.*, **64**, 711–714 (1992).
174. Q. Li, E.S. Yeung, 'Contamination Control in Capillary Electrophoresis and Quantitative Measurement of Potassium and Sodium in Single Human Erythrocytes', *J. Capill. Electrophor.*, **1**(1), 55–61 (1994).
175. Y.-T. Lee, C.-W. Whang, 'Capillary Electrophoresis of Triorganotin Compounds with Indirect Fluorescence Detection', *J. Chromatogr., A*, **746**, 269–275 (1996).
176. L.N. Amankawa, W.G. Kuhr, 'Indirect Fluorescence Detection in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **63**, 1733–1737 (1991).
177. J.T. Smith, Z. El Rassi, 'Micellar Electrokinetic Capillary Chromatography with In Situ Charged Micelles: Evaluation of Background Fluorophores for Indirect Fluorescence Detection', *J. Capill. Electrophor.*, **1**(2), 136–143 (1994).
178. S. Kennedy, B. Caddy, J.M.F. Douse, 'Micellar Electrokinetic Capillary Chromatography of High Explosives Utilising Indirect Fluorescence Detection', *J. Chromatogr., A*, **726**, 211–222 (1996).
179. T. Fuchigami, T. Imasaka, 'Capillary Micellar Electrokinetic Chromatography based on Indirect Semiconductor Laser-induced Fluorescence Detection', *Anal. Chim. Acta*, **291**, 183–188 (1994).
180. S.-Y. Liao, Y.-C. Chao, C.-W. Whang, 'Indirect Chemiluminescence Detection in Capillary Electrophoresis', *J. High Resolut. Chromatogr.*, **18**, 667–669 (1995).
181. S.-Y. Liao, C.-W. Whang, 'Indirect Chemiluminescence Detection of Amino Acids Separated by Capillary Electrophoresis', *J. Chromatogr., A*, **736**, 247–254 (1996).
182. Y. Zhang, B. Huang, J.-K. Cheng, 'On-line Indirect Chemiluminescence Detection in Capillary Electrophoresis', *Anal. Chim Acta*, 157–163 (1998).
183. T.M. Olefirowicz, A.G. Ewing, 'Capillary Electrophoresis with Indirect Amperometric Detection', *J. Chromatogr.*, **499**, 713–719 (1990).
184. P.K. Dasgupta, L. Bao, 'Suppressed Conductometric Capillary Electrophoresis Separation Systems', *Anal. Chem.*, **65**, 1003–1008 (1993).

185. S. Kar, P.K. Dasgupta, H. Liu, H. Hwang, 'Computer-interfaced Bipolar Pulsed Conductivity Detector for Capillary Systems', *Anal. Chem.*, **66**, 2537–2543 (1994).
186. C. Haber, W.R. Jones, J. Soglia, S.M. McGlynn, A. Caplan, J.R. Reineck, C. Krstanovic, 'Conductivity Detection in Capillary Electrophoresis – a Powerful Tool in Ion Analysis', *J. Capill. Electrophor.*, **3**, 1–11 (1996).
187. C. Haber, R.J. Van Suan, W.R. Jones, 'Quantitative Analysis of Anions at ppb/ppt Levels with Capillary Electrophoresis and Conductivity Detection: Enhancement of System Linearity and Precision using an Internal Standard', *Anal. Chem.*, **70**, 2261–2267 (1998).
188. C.W. Klampfl, M.V. Katzmayr, W. Buchberger, 'Separation of Inorganic and Organic Anions by Capillary Zone Electrophoresis with Simultaneous Indirect UV and Conductivity Detection', *Electrophoresis*, **19**, 2459–2464 (1998).
189. P. Gallagher, N.D. Danielson, 'Capillary Electrophoresis of Cationic and Anionic Surfactants with Indirect Conductivity Detection', *J. Chromatogr., A*, **781**, 533–540 (1997).
190. C.A. Lucy, Q. Wu, 'Characteristics and Calibration of Conductivity Detection in Capillary Electrophoresis', *J. Chromatogr. Sci.*, **36**, 33–36 (1998).
191. J.A. Fracassi da Silva, C.L. do Lago, 'An Oscillometric Detector for Capillary Electrophoresis', *Anal. Chem.*, **70**, 4339–4343 (1998).
192. C. Stathakis, R.M. Cassidy, 'Cationic Polymers for Selectivity Control in Capillary Electrophoretic Separations of Inorganic Anions', *Anal. Chem.*, 2110–2115 (1994).
193. A.J. Zemann, 'Sub-minute Separations of Organic and Inorganic Anions with a Co-electrosomotic Capillary Electrophoresis', *J. Chromatogr., A*, **787**, 243–251 (1997).
194. H. Burt, D.M. Lewis, K.N. Tapley, 'Resin Coating for Capillaries giving a Net Positive Charge and Great Potential for Customized Modification of Surface Properties', *J. Chromatogr., A*, **739**, 367–371 (1996).
195. B.J. Wildman, P.E. Jackson, W.R. Jones, P.J. Alden, 'Analysis of Anionic Constituents of Urine by Inorganic Capillary Electrophoresis', *J. Chromatogr.*, **546**, 459–466 (1991).
196. M.A.G.T. Van den Hoop, R.F.M.J. Cleven, J.J. van Staden, J. Neele, 'Analysis of Fluoride in Rain Water: Comparison of Capillary Electrophoresis with Ion Chromatography and Ion-selective Potentiometry', *J. Chromatogr., A*, **739**, 241–248 (1996).
197. G.W. Tindall, R.L. Perry, 'Separation of Fast Anions without Flow Reversal', *J. Chromatogr., A*, **696**, 349–352 (1995).
198. P.R. Haddad, P.E. Jackson, *Ion Chromatography, Principles and Applications*, Elsevier Science Publishers, Amsterdam, 86–89, Vol. 46, 1990.

Near-infrared Absorption/Luminescence Measurements

A.R. Swamy, J.C. Mason, H. Lee, F. Meadows, M. Baars, L. Strekowski, and G. Patonay
Georgia State University, Atlanta, USA

1 Introduction 1
2 Near-infrared Chromophores and Light Absorption Properties 3
3 Chemistry of Near-infrared Dyes 4
3.1 Indocarbocyanine Dyes 4
3.2 Squarylium Dyes 7
3.3 Phthalocyanines and Naphthalocyanines 7
4 Analytical Applications of Near-infrared Fluorescence 8
4.1 Applications of Non-covalent Labeling with Near-infrared Dyes 9
4.2 Bioanalytical Applications of Covalent Labeling with Near-infrared Dyes 9
4.3 Environmental Applications of Near-infrared Fluorescence 21
4.4 Other Applications of Near-infrared Fluorescence 24
5 Conclusions 29
Abbreviations and Acronyms 29
Related Articles 30
References 30

The near-infrared (NIR) region (600–1100 nm) offers several advantages in comparison with the conventional ultraviolet/visible (UV/VIS) region for spectroscopic measurements and detection. The lack of commercial instrumentation impeded the development of NIR techniques for years. The advent of inexpensive photodiodes and diode lasers used widely by the telecommunication industry gave the necessary impetus for development of analytical NIR techniques. This article describes the fundamental properties of NIR fluorophores and absorbers and provides a comparison with conventional UV/VIS dyes. The various analytical applications developed in the past decade that have utilized the advantages offered by this region are described with a particular emphasis on bioanalytical uses.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The current interest in development of new techniques for analysis is largely attributed to limitations of the more conventional instrumental methods. Classical biomolecule identification usually involves separation of a complex mixture of biological molecules followed by tests for identification of the separated fractions. No single test provides definitive identification of an unknown biomolecule, hence a complex series of tests are required. Such processes are time-consuming and cannot be carried on a reasonable timescale needed in clinical laboratories. These obstacles and the desire to attain adaptability to primitive field test conditions have prompted many researchers to explore modern alternatives to classical instrumental procedures. Most of these modern analytical tools are characterized by rapid data acquisition and data reproducibility which is due to coupling with computer-aided instrument control, data recording and interpretation. A number of modern instrumental techniques have been applied to the identification of biomolecules. These well-established methods include high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), circular dichroism (CD), gas chromatography (GC), and mass spectrometry (MS). Many of these techniques have also been adapted for field applications. Most of these techniques conventionally have been applied in the UV/VIS region. However, the NIR region offers several advantages over the conventional UV/VIS region.

Sir William Herschel, the English astronomer, discovered the NIR region (600–1100 nm) at the beginning of the 19th century[1] by using a prism and a thermometer; however, this region of the electromagnetic spectrum did not receive much attention until the advent and recent availability of inexpensive light sources, such as diode lasers and detectors such as photodiode and avalanche photodiode (APD) detectors. The 600–1100-nm region corresponds to an energy range of 48–26 kcal mol^{-1}. Atomic and molecular transitions in this long-wavelength region are processes that require relatively low-energy photons, because the ground- and excited-state species are close in energy. As a result, several classes of molecules, such as polymethine and phthalocyanine (Pc) dyes, and certain elements, such as ruthenium and osmium, require a moderately low energy input to produce spectroscopically measurable electronic transitions.

Owing to its sensitivity and selectivity, fluorescence is used as a major analytical tool in the identification of target molecules of interest. Typically, this involves using fluorophores as reporter molecules or labels. Background fluorescence from components other than the

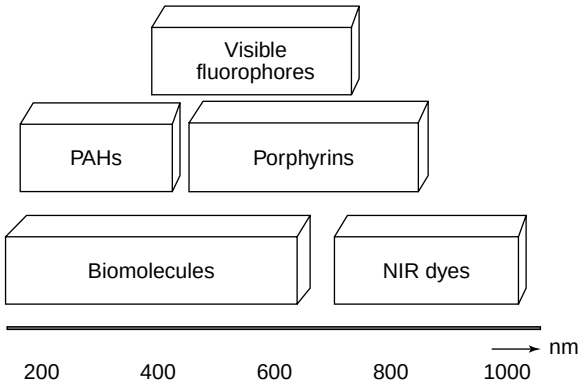

Figure 1 Visible to NIR region and possible interferences from native fluorescence.

fluorophore of interest decreases the sensitivity of detection in solution. In biological systems, this background fluorescence is typically from the autofluorescence of certain biological components. Similar problems are encountered with environmental samples. Typical background fluorescence occurs at all wavelengths in the visible region and at various intensities depending on the concentration of interfering molecules present in the sample (Figure 1). Elimination of background fluorescence is essential where the undesired components are not separated and fluorescence must be detected in the presence of intervening molecules. In fluorescence analysis, excitation light can be scattered from interaction with various types of molecules at the surfaces of containers. This scattered light effect is present in all types of fluorescence detection. Scattered light contributes to a significant portion of background noise, especially in biological samples, and it can be due to either Rayleigh, Raman, or Tyndall effects. These interferences can be best minimized by using fluorophores with relatively high Stokes shifts (typically >40–50 nm).

Fluorescence is more sensitive and selective than absorbance as a spectroscopic tool, not only because it is measured against a zero background but also because the magnitude of fluorescence signal, F, at low dye concentrations is expressed by Equation (1):

$$F = 2.303\Theta_f I_0 \varepsilon b C \quad (1)$$

where I_0 is the excitation power, ε is the molar absorptivity at the excitation wavelength, Θ_f is the quantum yield, b is pathlength, and C is the dye concentration. It can be seen that the limits of detection can be improved by a stronger excitation source. Laser-induced fluorescence (LIF) provides a superior approach to improve the sensitivity of fluorescence techniques. However, one must keep in mind that the limit of detection increases only as the inverse square root of excitation power and a strong excitation source can cause photo bleaching of the fluorophore. The limitations of conventional lasers as excitation sources include their high price, size, maintenance costs, and their limited wavelength selection. LIF in the NIR region (600–1100 nm) offers several advantages, and the recent advances in semiconductor laser technology have made the use of lasers more practical. In addition, the extensive use of NIR-emitting laser diodes in the telecommunications industry has made them more readily available. These types of lasers are inexpensive (typically <$150), small (~1 cm), and have a long operating lifetime (>100 000 h). A comparison of selected visible and NIR laser excitation sources is provided in Table 1. The gallium–aluminum–arsenide (GaAlAs) laser diode has drawn much interest because its emission wavelength of 785 nm is compatible with several classes of polymethine cyanine and naphthalocyanine (NPc) dyes that exhibit NIR fluorescence.[2]

Table 1 Comparison of NIR and visible laser excitation sources

Parameter	Laser diode	Argon laser
Wavelength (nm)	785	488
Lifespan (h)	100 000+	3000
Power output (W)	0.02	1–5
Power consumption (W)	0.15	1800
Replacement ($)	<150	>5000

In addition to the availability of diode lasers used for excitation, fluorescence detection in the NIR region has advantages since noise resulting from scatter is related to wavelength of detection by the factor of $1/\lambda^4$. Detection at 820 versus 500 nm results in more than a sixfold reduction in scatter noise. The low background interference in the NIR spectral region allows NIR fluorophores to be used as ideal probes in both biological and environmental applications. The advantages offered by the NIR region are summarized in Table 2. Detection in the NIR allows the commonly used photomultiplier tube (PMT) to be replaced with the more efficient photodiode and APD type. The APDs have excellent quantum efficiency in the NIR region. Table 3 illustrates the advantages offered by APDs in comparison with PMTs. These benefits include their low cost, compact structure, and their durability. In addition, they have low internal noise and very low power consumption. All these features allow NIR fluorescence

Table 2 Comparison of NIR and visible regions

Noise source	NIR region	Visible region
Detector	Low	High
Scatter (Rayleigh/Raman)	Reduced	6× greater at 520 nm than 820 nm
Background/auto-fluorescence	Mostly absent	Autofluorescence of biomolecules

Table 3 Comparison of APDs and PMTs

Parameter	APD	PMT
Replacement cost ($)	<50	~500
Lifetime (h)	>100 000	10 000, sensitive to overexposure
Quantum efficiency (at 820 nm) (%)	80	0.3
Internal amplification	High	Low
Size	Very small (nm)	Small (cm)
Power consumption	Very low	Low

instrumentation to be highly versatile and amenable to miniaturization. These features aid in the development of a portable compact and rugged instrument for field applications.

The recent progress in solid-state diode laser technology, the availability of commercial NIR absorbing dyes, and the development of diverse synthetic routes to functionalized NIR absorbing dyes have allowed NIR spectroscopy to be used in many diverse applications. These specific uses rely on the response of the chromophore after its excitation by NIR radiation. A simple absorber can be used in spectrophotometric analysis,[3] as molecular probes in immunoassays (IAs), or in applications of security printing, such as laser-readable bar codes.[4] The absorber can re-emit the input energy via fluorescence or phosphorescence. Also, the NIR-absorbing chromophore can convert the excited-state energy into other forms that can be utilized for specific applications. After absorption of laser energy, the chromophore can produce a local heating effect in the medium in which it resides.[3,4] This technique is used in the optical data recording technologies, such as WORM (write once read many) and DRAW (direct read after writing),[4] full color laser imaging, and in medicinal uses such as tissue welding.[3] Conversion of the NIR radiation into electrical energy by the chromophore is the fundamental basis for photovoltaic cells and electrophotography, including laser printing.[3,4] Furthermore, transfer of the chromophore's excited-state energy to endogenous ground-state molecular oxygen is the foundation for the selective destruction of neoplastic tissues in photodynamic therapies.[4,5] In particular, the NIR region provides a convenient spectral range for photodynamic therapies and in vivo imaging owing to the multicentimeter penetration of biological tissues.[5] Taken together, these benefits make NIR analysis an operator-safe, nondestructive, sensitive, and versatile technique that is taking the place of conventional methods, such as radiolabeling. Moreover, dilute samples can be analyzed safely and relatively fast compared with other techniques. Several applications have been developed in the past decade utilizing the various advantages offered by this novel methodology ranging from DNA sequencing, pH and hydrophobicity determination, metal ion detection, antibody (Ab) labeling, HPLC, and high-performance CE.

2 NEAR-INFRARED CHROMOPHORES AND LIGHT ABSORPTION PROPERTIES

To take full advantage of the high-technology NIR instruments, new NIR chromophores with specific properties are required. The comprehension of the color–structure relationship of near-infrared dyes (NIRDs) can provide useful information for their development since the light-absorption property of an organic molecule is correlated with its structural features. A highly conjugated system in either a linear or a cyclic arrangement in the molecule is responsible for the absorption maxima in the NIR region.

In principle, the absorption property of a chromophore is the characteristic of the energy that is absorbed to cause the electronic transition. NIR-absorbing chromophores require relatively low energy for the transition, and this corresponds to the longer wavelength of the electromagnetic spectrum in comparison with UV/VIS absorption. Since the transition energy is the energy required for a single electron to be excited from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO), the closeness between the HOMO and LUMO orbital is primarily responsible for the amount of energy required to induce this electronic transition. Therefore, the shift of the absorption maxima (λ_{max}) into the NIR region can be induced if the gap between the HOMO and the LUMO orbitals is brought close enough to give the transition energy in the range 48–26 kcal mol^{-1}. In practice, effective structural modification of the chromophore can cause the desired bathochromic shift. Griffiths summarized in detail the general strategies to develop new NIR-absorbing dyes using this approach.[3] These strategies include (i) extending the conjugation of a chromophore, (ii) increasing the interaction between electron donor and acceptor groups within a chromophore, (iii) altering the electronegativity of atoms within a chromophore, (iv) introducing specific branching or bridging within a chromophore, (v) metal complexation with a chromophore, (vi) intermolecular charge-transfer complex formation, and (vii) formation of a free radical that is part of the chromophore.

The Pariser–Parr–Pople molecular orbital (PPPMO) method[6] and XNDO/S method[7] are useful in the estimation of the absorption position of the chromophore that needs to be constructed. The perturbational molecular orbital (PMO) theory in the form of Dewar's rule

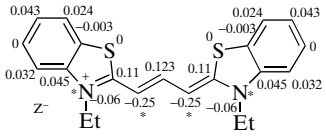

(1)

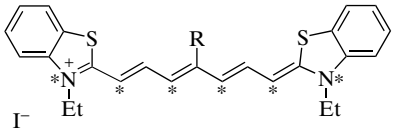

(2)

R = H, λ_{max} = 758 nm
R = CN, λ_{max} = 860 nm ($\Delta\lambda$ = 102 nm)
R = OMe, λ_{max} = 730 nm ($\Delta\lambda$ = –28 nm)

can also aid in the prediction of the light-absorption properties of chromophores with structural changes.[8] This rule is particularly useful in the development of new NIRDs since one can correlate the spectroscopic effects with a structural change at a specific position. In the basic analysis of an odd-alternate system of the conjugated chromophore, each alternate atom starting from a terminal donor is starred, as exemplified in **(1)**. Dewar's rule implies that an increase in the electronegativity at a starred carbon results in a hypsochromic shift. On the other hand, a bathochromic shift will be induced if the electronegativity of an atom at unstarred position is increased. In particular, a bathochromic shift of a cyanine chromophore is induced if the electron donor groups at starred positions or electron-withdrawing groups at unstarred positions are introduced.[6] This is further illustrated by the studies of substituted trimethine cyanines **(1)** on the basis of the PPPMO theory by Yasui et al., which showed the decrease in the electron densities at the starred positions and increase at the unstarred positions of the polymethine chain accompanying the first excitation of **(1)**.[9] An additional example of the application of Dewar's rule is given in **(2)** for three heptamethine cyanine dyes of a general structure **(2)**. As can be seen, the electron-withdrawing substituent (CN) at the central unstarred position of the chromophore causes a bathochromic shift and a hypsochromic shift is observed for the electron-donating group (OMe) relative to absorption of the parent unsubstituted dye R = H).[10]

3 CHEMISTRY OF NEAR-INFRARED DYES

So far, the theoretical correlation between structure and absorption properties has been discussed. These general rules can also be applied in the practical synthetic design of NIR-absorbing dyes. However, it should be noted that not only the light absorption properties of NIR chromophores but also other characteristics for different analytical applications are desired in the design of these systems. NIR chromophores with proper functional groups, organic and/or water solubility, high molar absorptivity or fluorescence maximum in the NIR region are required for most applications, including medical and optical purposes, dye lasers, and photographic sensitizers. These characteristics can be generally introduced in two ways, namely (i) by the structural modification of commercially available dyes and (ii) by direct synthesis utilizing appropriate precursors. Several examples of direct synthetic routes to NIRDs with specific applications are discussed below.

3.1 Indocarbocyanine Dyes

Functionalized NIR absorbing chromophores are widely used for labeling purposes. However, the commercial availability of such systems is somewhat limited. The first NIR-absorbing dyes containing $-N{=}C{=}S$ and $-C(O)CH_2I$ functionalities for selective coupling with amino and thiol groups, respectively, of biomolecules were synthesized by Waggoner's group.[11] These dyes, synthesized by using classical chemistry, contained the reactive group at one of the terminal heterocyclic units, and were extremely difficult to purify. Subsequently, Strekowski et al.[12,13] discovered a facile functionalization of indolium heptamethine cyanine dyes at the central *meso* position of the chromophore, and the synthesis of **(8a–c)** is given in Scheme 1 as an example. These dyes are substituted with an isothiocyanato or *N*-succinimidyl ester function for selective labeling at the amino group of proteins or bioconjugates such as amino-functionalized nucleotides.[13,14] The structural design of **(8)** includes the presence of sulfonic acid groups for excellent solubility in water and the overall symmetry of the molecule to contribute to the facile purification of the dye by simple crystallization. The symmetry is responsible for a high molar absorptivity of the chromophore. An important structural feature is the rigid trimethylene bridge in the center of the molecule that enhances the fluorescence quantum yield by decreasing conformational freedom of the NIR fluorophore in comparison with that of the fluorophore without the bridge.

The key synthetic step to **(8)** is a facile nucleophilic displacement of the chloro substituent in the intermediate product **(7)** which, apparently, involves an $S_{NR}1$ pathway.[12] In the preparation of **(7)** the starting material **(6)** is derived from dialdehyde **(11)** (for structure, see Scheme 2) by the reaction with aniline. The second component **(5)** is prepared efficiently by the reaction of **(3)** and **(4)**, as shown in Scheme 1.

This approach, as discussed above, was utilized by Flanagan et al.[15] in DNA sequencing involving capillary

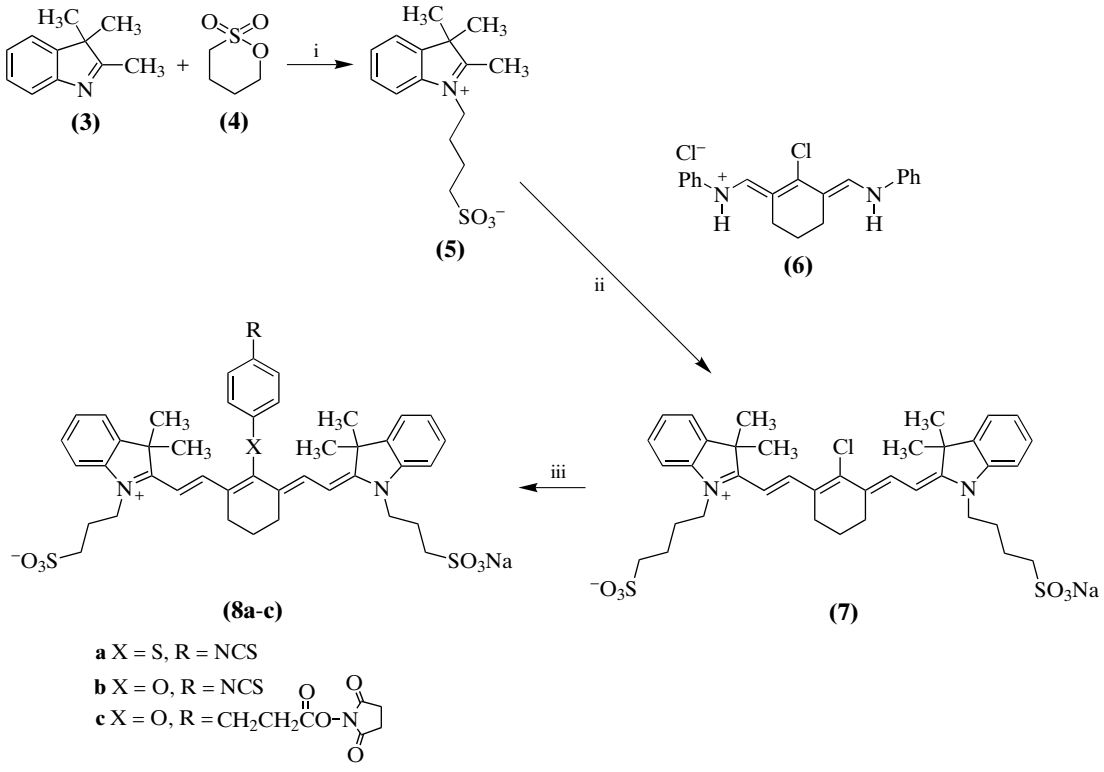

Scheme 1 Synthetic route to heptamethine dyes. (i) Reflux in 1,2-dichlorobenzene. (ii) Reflux in EtOH in the presence of sodium acetate. (iii) Substituted sodium phenolate or thiophenol in DMF, 23 °C. [Strekowski et al.[13] approach.]

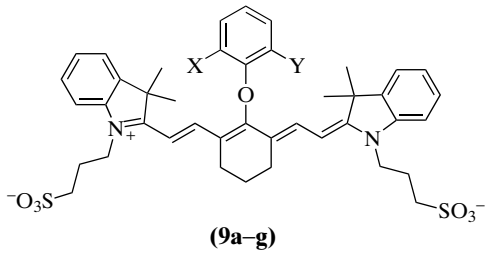

(9a–g)

X, Y	X, Y
a H, I	**e** Br, Br
b H, Br	**f** Cl, Cl
c H, Cl	**g** H, H
d H, F	

gel electrophoresis (CGE). Their dyes **(9a–g)** contain different atoms incorporated in the periphery of the chromophore, and have been prepared by substitution of the *meso*-chlorine atom in the intermediate dye with 2,5-disubstituted phenolates. The introduction of fluorine or heavy atoms such as iodine, bromine, and chlorine into the molecular framework has been found to affect the fluorescence lifetime without affecting the spectral properties of the chromophore. It was reasoned that introduction of heavy atoms induces spin–orbit coupling. As a result, the intersystem crossing (ISC) rates are enhanced producing a reduction in the excited-state lifetimes associated with the singlet state.

Recently, several commercial NIR-absorbing cyanine dyes have also been made available with amine-reactive groups for covalent coupling with synthetically altered nucleotides or oligonucleotides. The dyes are available under the Cy™ and IRD™ trade names, and have been synthesized by Mujumdar et al.[16] and LI-COR, Inc.,[17] respectively. The LI-COR IRD™ dyes, such as the nonsymmetric heptamethine dye IRD800™ phosphoramidite [**(18)**, Scheme 2], can be prepared by using the methodology described by Narayanan et al.[18] The synthetic route involves heating the bisaldehyde **(11)** and the hydroxy-functionalized quatenary indolium salt **(10)** with azeotropic removal of water (Scheme 2) to give the intermediate 'half' cyanine carboxaldehyde **(12)**. This intermediate generated in situ is further allowed to react with indolium alkyl sulfonate **(5)**, which furnishes the cyanine dye **(13)**. The absorption maximum of the NIR chromophore in **(13)** can then be fine tuned by addition of an electron-donating or electron-withdrawing group at the *meso*-position, as already discussed. Thus, the reaction of **(13)** with sodium phenolate **(14)** furnishes a phenoxy derivative **(15)**. Further treatment of the *meso*-substituted cyanine dye **(15)** with 1*H*-tetrazole **(16)** and 2-cyanoethyl tetraisopropylphosphorodiamidite

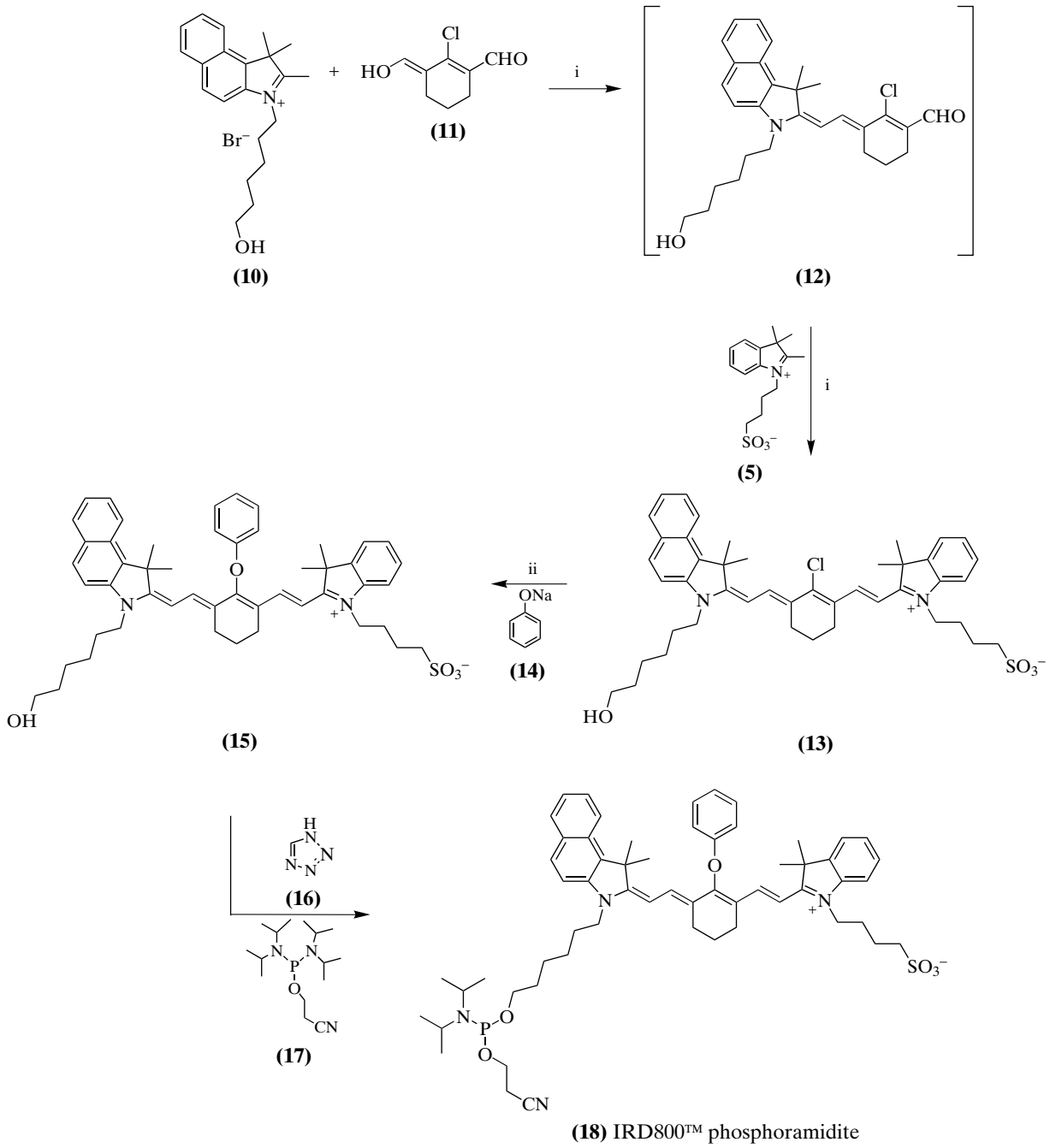

Scheme 2 Synthetic route to IRD800™ phosphoramidite. (i) Reflux in a mixture of benzene and *n*-butanol with azeotropic removal of water. (ii) Stirring in DMF at 23 °C.

(17) produces the final dye **(18)** (IRD800™) that contains a reactive phosphoramidite functionality for automated DNA synthesis applications. The spectral properties of IRD800™ are retained after coupling with the oligonucleotide. The unlabeled chromophore **(18)** exhibits an absorption maximum at 787 nm in methanol with a 25-nm Stokes shift in fluorescence and a 15.0% quantum yield, while the oligonucleotide-labeled chromophore **(18)** has an absorption maximum at 796 nm in water with a 23-nm Stokes shift and a 14.7% fluorescence quantum yield.

Using similar synthetic strategies, nonsymmetric pentamethine cyanine dyes have been developed by Mujumdar et al.[16] The dyes, such as Cy5-dUTP™ **(19)**, can be covalently linked to biomolecules including nucleotides and the conjugates used in enzymatic polymerization

processes. The dye-labeled dUTP analogs can be incorporated into DNA probes by nick translation, random priming, and polymerase chain reaction (PCR).

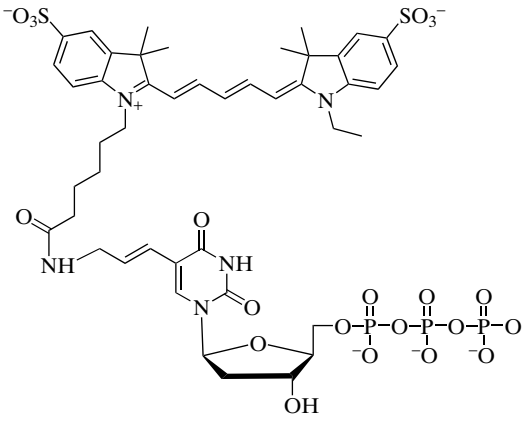

(19)

Cy5-dUTP™

3.2 Squarylium Dyes

In addition to the indocarbocyanine dyes, synthetic routes to indolenine–squaraine dyes, such as the *N*-succinimidyl ester derivative **(23)** (Scheme 3) for conjugation to proteins, are being developed. The squarylium dyes exhibit high photostability and quite long fluorescent lifetimes (nanosecond range) in aqueous solutions. Further benefits of using NIR-absorbing squaraine dyes for biological applications include the fact that their spectral properties (absorbance, emission, and lifetime) are independent of pH under physiological conditions (pH 6–9).[19] As shown in Scheme 3, Terpetschnig et al.[19] synthesized a water-soluble NIR-absorbing dye containing an activated ester moiety. The reaction of **(20)** and **(21)**[20] produced carboxy-substituted dye **(22)**. Following purification by preparative thin-layer chromatography (TLC), the nonsymmetric squaraine **(22)** was coupled to *N*-hydroxysuccinimide to produce the active ester **(23)**.

3.3 Phthalocyanines and Naphthalocyanines

The NIR-absorbing dyes such as Pc and NPc type are used as industrial colorants and in special applications of electronics and optics.[21] They efficiently complex various metal ions, and approximately 70 central metal ion species of Pcs are known. The cation is held in the planar 18-electron cavity and the metal cation can strongly influence the physical and spectral properties of the dye, including absorbance, emission, and fluorescence lifetime.[21] These NIRDs are sensitive to environmental changes such as pH and metal concentration. Classical synthetic routes to metal ion-containing Pc dyes [metal ion phthalocyanine (MPc) dyes] are illustrated in Scheme 4 where the metal ion acts as a central template for cyclotetramerization. As can be seen, MPc **(28)** is obtained directly from diiminoisoindolenine **(24)**, phthalic anhydride **(25)**, phthalonitrile **(26)**, or phthalimide **(27)** in the presence of a metal salt.[21] These dyes, though highly insoluble, are extremely stable and can be obtained by precipitation and further purified by sublimation. Metal ion-containing

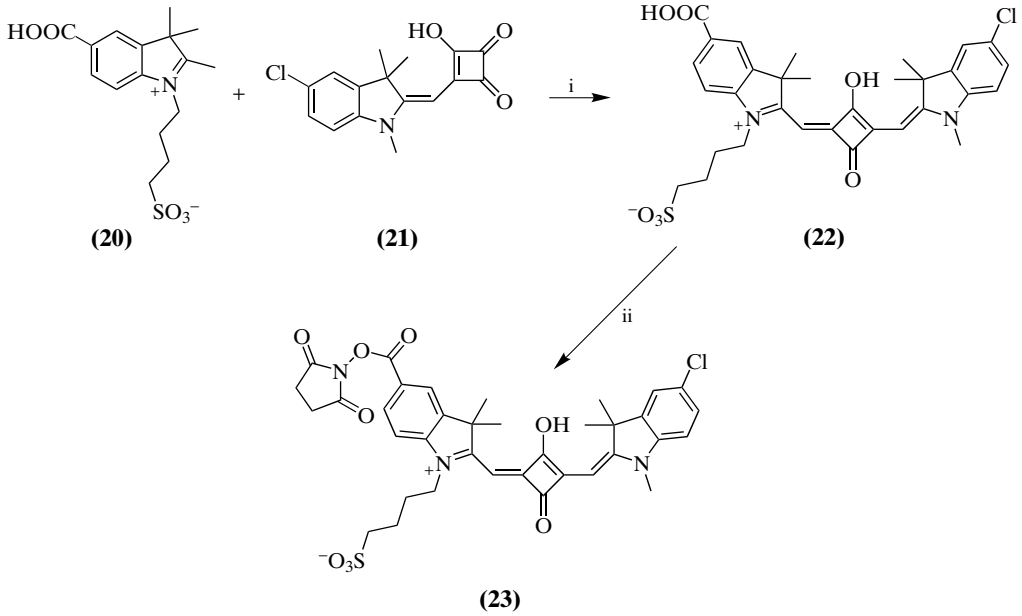

(20) **(21)** **(22)** **(23)**

Scheme 3 Synthetic route to nonsymmetric squarine dye with *N*-succinimidyl moiety. (i) Reflux in *n*-butanol and toluene with azeotropic removal of water. (ii) Stirring with *N*-hydroxysuccinimide and DCC in dimethoxyethane at 23 °C.[19]

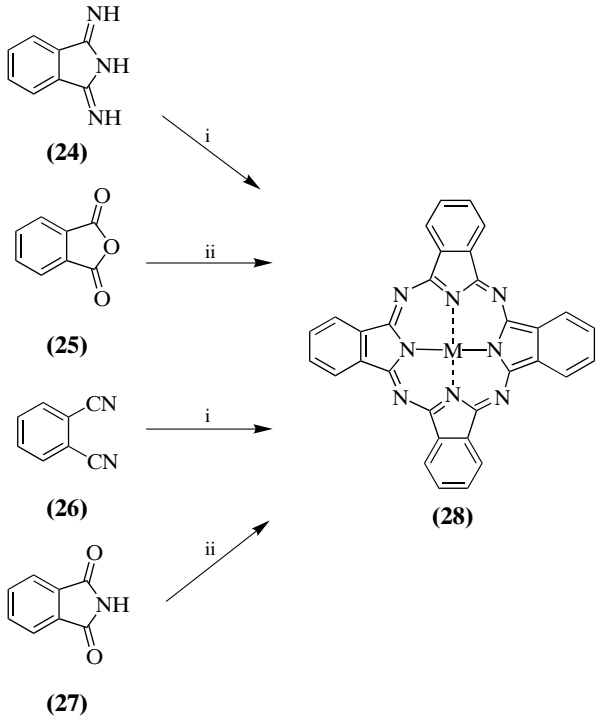

Scheme 4 Classical synthetic routes to metal ion NPcs. (i) Heat in a high-boiling solvent (e.g. quinoline) with metal salt. (ii) Heat in a high-boiling solvent with urea and metal salt.

NPcs, such as **(31)** in Scheme 5, are obtained in a similar fashion.

The inherent sensitivity to metal and hydrogen ion concentration makes Pcs and NPcs ideal candidates for use in pH or metal determination. However, the poor aqueous solubility is a shortcoming that has to be overcome by structural modification or functionalization. One approach to overcome the insolubility is to incorporate sulfonate groups into the periphery of the existing Pcs or NPcs (Scheme 5). Patonay et al. at Georgia State University[22] have synthesized a tetrasulfonylated aluminum NPc **(31)** that can be incorporated into an optical probe for NIR metal determinations. Once incorporated into the fiber, the four sulfonate groups at the periphery of the NIR fluorophore interact with metal ions, such as K^+, to produce spectral changes in their absorption maximum (λ_{max}) and/or fluorescence intensity. The NIR probe studied by Patonay et al. provides metal detection over the range 1×10^{-8}–0.5×10^{-1} M.

4 ANALYTICAL APPLICATIONS OF NEAR-INFRARED FLUORESCENCE

LIF offers tremendous advantages over conventional analytical methods. The NIR region which has no background interference when coupled with LIF can attain sensitivities obtained by radiolabeling. This method also provides a step towards the maximum achievable sensitivity, that is, single-molecule detection. Many reviews of NIR fluorescence that describe dyes, instrumentation, and applications are available. An authoritative review, though dated, is provided by Warner et al.[23] A recent book summarizing the results of a North Atlantic Treaty Organization (NATO) Advanced Research Workshop covers the synthesis, optical properties and applications of selected NIRDs in high-technology fields.[24] In the next section, representative examples are reviewed to emphasize and acknowledge further the advantages and usefulness of NIR fluorescence techniques.

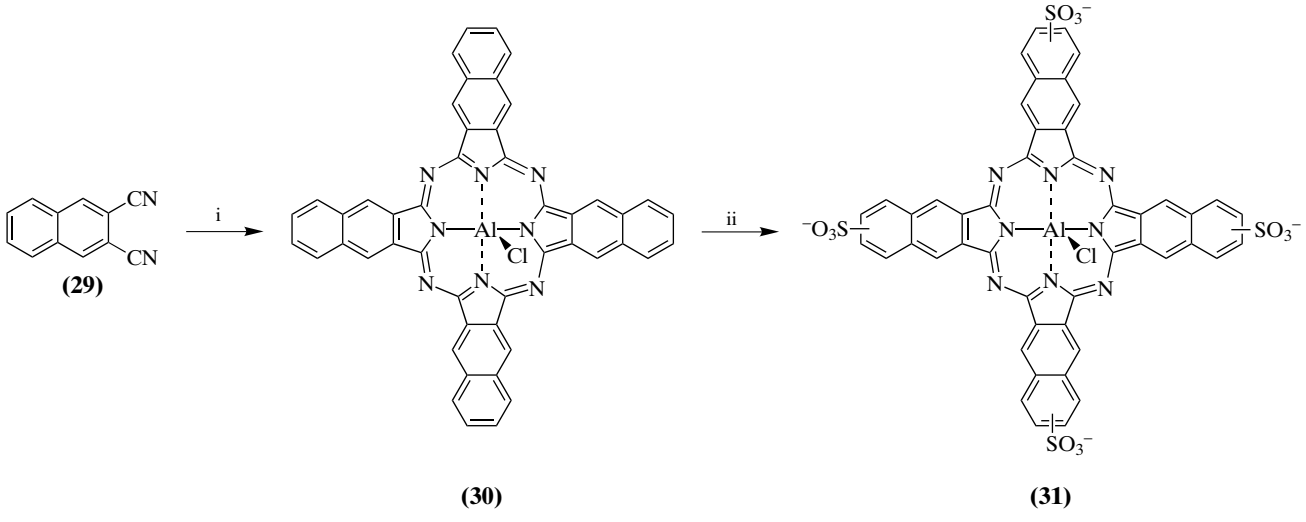

Scheme 5 Synthesis of tetrasulfonylated aluminum NPcs. (i) Reflux in 1,2-dichlorobenzene containing 1 M $AlCl_3$. (ii) Heat in oleum (15% of free SO_3).[22]

4.1 Applications of Non-covalent Labeling with Near-infrared Dyes

The NIRDs used as reporter molecules can be attached to the biomolecule of interest by two methods, via (i) covalent bond and (ii) noncovalent hydrophobic and/or ionic interactions. Theoretically, the hydrophobic backbone of many NIRDs allows for noncovalent labeling of large biomolecules. However, only a limited number of practical applications have been tested.

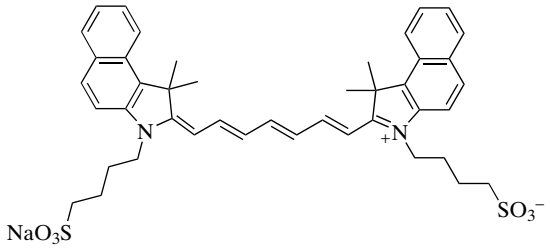

(32)

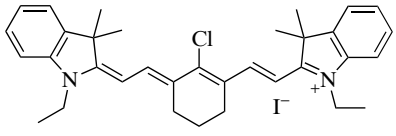

(33)

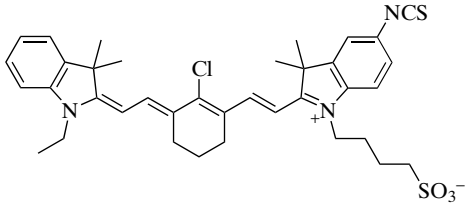

(34)

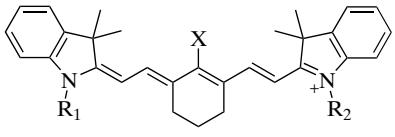

(35)

X = OPh-4(NCS), SPh-4(NCS); R_1 = Et, $(CH_3)_4SO_3Na$; R_2 = Et, $(CH_3)_4SO_3^-$

One of the earliest applications studied was the binding of indocyanine green (ICG) **(32)** to serum albumins by Kamisaka et al. in 1974.[25] Similar studies by Sauda et al.[26] using semiconductor laser fluorimetry illustrated that human serum albumin labeled with ICG provided picomolar detection limits. A comparison of noncovalent [**(33)** and **(34)**] and covalent dyes **(35)** labeling used in HPLC determination was investigated by Williams et al.[27] in terms of stability and specificity. Their results indicated that noncovalent labeling occurs rapidly at a physiological pH range. However, this interaction was found to be nonspecific and less stable than covalent binding.

4.2 Bioanalytical Applications of Covalent Labeling with Near-infrared Dyes

Many bioanalytical applications which rely on covalent labeling have been developed. Techniques that utilize this methodology include IAs, CE, DNA sequencing, and the development of gene probes.[28] This section details some of these applications with their associated merits and disadvantages.

4.2.1 Application of Near-infrared Fluorescence in Capillary Electrophoresis

Analytical separations of intact proteins using HPLC and CE have been shown to have excellent resolving power.[29] CE is particularly advantageous for protein analysis because of its higher chromatographic efficiency, faster separations time, and ability to use small amounts of material. The low molecular diffusion rates of proteins help achieve theoretical plate numbers as high as 10^6.

Many excellent reference books on CE have appeared in recent years,[30–32] and the reader is encouraged to refer to them for additional information not covered below. The *Handbook of CE*[31] is very thorough, while the CE primer series offered by Beckman Instruments[30] provides a brief overview of the principles of different separation techniques for a variety of applications.

Five main modes of electrophoresis have been developed over the years. These include capillary zone electrophoresis (CZE), isoelectric focusing (IEF), CGE, isotachophoresis (ITP), and micellar electrokinetic chromatography (MEKC). These different modes account for the wide applicability of CE to different analytes. However, one of the challenges in CE applications is the detection of analytes at low concentrations. This is a result of the physical dimensions of the capillary used for separation. Typical injection volumes in the nanoliter range combined with small detector windows of approximately 100–200 mm require highly sensitive detection systems.

UV/VIS absorbance is the most commonly used detection method for CE. The advantage is simplicity, low cost, and ease of use. Detection limits are in the range of 10^{-13}–10^{-16} mol for direct absorption (a factor of 10 less for indirect methods). Other methods employed include MS, electrochemical, refractive index and radiometric detection.[28] Fluorescence detection can greatly enhance the sensitivity of CE methods. For example, direct fluorescence provides detection limits in the range of

10^{-15}–10^{-17} mol (a factor of 10 less for indirect methods). However, the number of compounds that fluoresce naturally is limited and derivatization of analytes with a fluorescent tag is frequently required. Fluorescence emission is measured against a zero background signal, resulting in improved detection limits in comparison with absorption methods. Furthermore, LIF coupled to CE provides a sensitivity of approximately 10^{-18}–10^{-23} mol. This improved sensitivity of fluorescence detection is attributed to concentrating high power monochromatic light into a very small area, which provides an ideal excitation source for CE fluorescence detection.

LIF is one of the most sensitive methods of detection available for use with CE. Although the number of publications on the subject has been rising steadily each year, the practical use of LIF with CE has been limited owing to scatter from incident light reflected by the capillary walls with on-column LIF detection. Scatter can be greatly reduced by utilizing postcolumn detection using a sheath-flow arrangement, first described by Chen and Dovichi.[33] However, this set-up adds a significant amount of complexity to the method, requires additional equipment such as a low-flow HPLC pump, and is not commercially available. Another deterrent is the high cost associated with visible lasers which are suitable for excitation of the more popular fluorescent labels including fluorescein isothiocyanate (FITC) and 3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde (CBQCA).

Several authors have reported on the advantages of using semiconductor lasers with CE. Most investigations to date have involved the use of far-red fluorescent dyes. Williams et al.[34] described a diode laser-based indirect absorbance detector for the analysis of a series of tetraalkylammonium ions. In this study, rhodamine 700 ($\lambda_{abs} = 642$ nm, $\lambda_{em} = 668$ nm) was used as a background absorber, and detection limits were found to be of the order of 20 mM. Another study utilizing a time-resolved photon counting fluorimetry CE method was described by Song et al.[35] Four fluorescent dyes were evaluated using an LIF detector with a 74-ps pulsed semiconductor laser emitting at 655 nm. The dyes investigated included methylene blue ($\lambda_{abs} = 665$ nm, $\lambda_{em} = 685$ nm), oxazine 725 ($\lambda_{abs} = 655$ nm, $\lambda_{em} = 679$ nm), 1,1′,3,3,3′,3′-hexamethyl dicarbocyanine (HIDC) iodide ($\lambda_{abs} = 636$ nm, $\lambda_{em} = 667$ nm), and rhodamine 700. The lowest limit of detection was reported to be 2 fmol, which was obtained for oxazine 725 at a concentration of 0.2×10^{-5} M. Fuchagami et al. synthesized a novel far-red fluorescent labeling dye for use with a semiconductor LIF detector.[36] Several amino acids were derivatized with 9-cyano-*N*,*N*,*N*′-triethyl-*N*′-(5′-succinimidyloxycarbonylpentyl)pyronine chloride ($\lambda_{abs} = 663$ nm, $\lambda_{em} = 685$ nm) and were separated and detected by CE. The experimental system used a 2-mW semiconductor laser emitting at 660 nm and a PMT detector. Detection was performed off-column using a sheath flow cell to reduce Rayleigh scatter, achieving a limit of detection [signal-to-noise ratio (SNR) = 2] of 800 zmol.

Table 4 Comparison of visible (R6G) and NIR (IR132) single-molecule detection. [Adapted from Soper et al.[38]]

Parameter	Visible	NIR
Excitation wavelength (nm)	532	800
Observed wavelength (nm)	570	840
Instrument response (fwhm) (ps)	100	250
Photon detection efficiency	0.0007	0.007
Probe volume (pL)	1.1	0.8
Transit time, τ_t (ms)	24	10
Background rate in time window (counts s^{-1})	225	145 (±12)
Average photons/molecule	39	18 (±4)
Detection efficiency (%)	78	97 (±7)
Probability of error	0.3	0.04

The advantages of using APD CE detectors at wavelengths above 650 nm have been reported. Kawazumi et al.[37] compared PMT and APD detection using oxazine 725 excited by a 29-mW semiconductor laser emitting at 655 nm. The APD was operated in a near-Geiger mode with time-gated detection, resulting in a 47-amol limit of detection. The APD limit of detection was 98 times greater than that with a PMT. As already mentioned, detection in the NIR region offers considerable advantages over working in the conventional UV/VIS region.

A significant amount of work with NIR fluorescent dyes including their applications to CE has been conducted at Louisiana State University by Soper et al.[38–40] They observed bursts of photons from single NIR fluorescent molecules in a flowing stream using a Ti : sapphire laser (12 mW) and an APD detector. The dye infrared (IR)-132 ($\lambda_{abs} = 805$ nm, $\lambda_{em} = 847$ nm) at a concentration of 25 fM in methanol was used in these experiments with time-gated off-column detection.[38] A comparison of a single-molecule detection of visible (R6G) dye and the NIRD IR-132 (Table 4) shows a significant reduction in background fluorescence, resulting in a lower discriminator threshold and error probability, and thereby provided a higher single-molecule detection efficiency in the latter case. A set-up similar to the photon burst experiment, using continuous excitation and CE with on-column detection, was utilized for an investigation of binary solvent effects in CZE with several NIRDs and dye-labeled amino acids.[39] One of the dyes substituted with an isothiocyanate group **(37)**, was synthesized from dye **(36)** by Soper et al. and covalently bound to amino acids. The results showed that the detectability, efficiency, and resolution of the

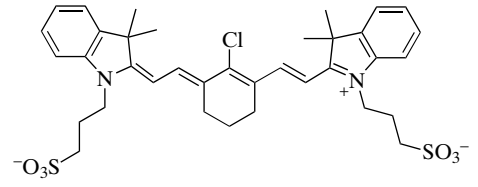

(36)

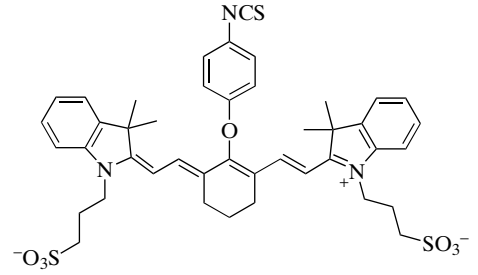

(37)

Table 5 SNR, theoretical plates per meter and mass detection limits for cationic dyes in mixed methanol–aqueous borate buffers (pH 9.1). [Adapted from Flanaghan et al.[39]]

Dye	SNR		Plates m^{-1}		Mass detection limit (ymol)	
	95:5	60:40	95:5	60:40	95:5	60:40
DTTCI	1478		65 500		5460	
IR-140	1713	146	64 500	63 600	446	20 280
IR-132	1636	247	64 300	55 700	429	24 960
HDITCP	1858		57 800		351	

HDITCP, 1,1′,3,3,3′,3′-hexamethyl-4,4′,5,5′-dibenzo-2,2′-indotricarbocyanine perchlorate.

NIR methodology can be significantly improved with the use of mixed organic–aqueous running buffers (Table 5). However, cationic NIRDs were found to be problematic at low concentrations owing to capillary wall adsorption. The on-column limit of detection was found to be 0.4 zmol using the dye HDITCP ($\lambda_{abs} = 780$ nm, $\lambda_{em} = 825$ nm) at a concentration of 2×10^{-10} M in 95% methanol, and the limit of detection of a dye-labeled amino acid was 21 zmol, using a 1×10^{-9} M solution in 9:1 methanol–water buffer. A comparable CE system with a 10-mW Ti:sapphire laser emitting at 795 nm was used to evaluate NIR fluorescence detection for CGE and DNA sequencing applications. A comparison between detection for NIRD-labeled primer **(38)** and fluorescein labeled primer **(39)** is shown in Table 6. Under similar conditions, the significantly lower Raman contributions and background fluorescence resulted in a background of 10 000 counts for NIR compared with >200 000 counts for visible excitation in contrast to the larger detection efficiency at 800 nm in comparison with 550 nm. Although the NIRD exhibits a lower quantum efficiency, a much lower limit of detection of 34 zmol was achieved for an IRD41™ labeled M13 primer ($\lambda_{abs} = 786$ nm, $\lambda_{em} = 812$ nm) in comparison with 1.5 amol for the visible primer primarily owing to the significantly smaller background in the NIR region.[40]

Until recently, all near-infrared laser-induced fluorescence (NIRLIF) work completed to date involved the use of laboratory-made CE systems and detectors. However, two commercial LIF detectors are currently available for use with CE, manufactured by Beckman and Zeta Technology. The Beckman LIF detector is specifically designed for use with their P/ACE™ CE instrument, while the Zeta LIF detector is a general-purpose detector also applicable to HPLC use. Both detectors are optimized for use in the visible region and utilize PMT detectors, with laser excitation via a fiber-optic cable. Recently, Patonay et al. at Georgia State University, actively involved in research in other areas of NIR fluorescence, developed a simple interface between a commercial APD-based NIR-LIF detector and a CE instrument. The system retained the fully automated injection, separation, and data collection capabilities of the commercial CE instrument and was optimized for detection around 820 nm with laser excitation at 787 nm, where the detector sensitivity was improved about 400-fold in comparison with the conventional PMT-based LIF detector.

To illustrate the challenges in design of developing a NIR fluorescence detector interface to exploit fully the advantages offered by this region, a more detailed description of the work done by Baars et al.[28] is presented in the following section. A Beckman LIF detector with the P/ACE™ 5000 CE instrument was modified to accommodate the diode laser excitation and NIR fluorescence detector. The back mounting plate and a portion of the self-aligning beam probe assembly of the Beckman LIF detector was used in the interface. The flat mirror, filter holder, PMT and detector electronics were removed resulting in a collimated fluorescence output from the LIF cartridge, passing through the beam block and beam probe, to allow interfacing with the NIR detector. A proprietary microscope and laser assembly manufactured by LI-COR, Inc. (Lincoln, NE, USA) was used as the NIRLIF detector. This detector system is comparable to components used in the commercially available automated DNA sequencing instrument[41] (LI-COR Model 4000), with the exception of modified focal length optics and an extra band-pass filter. The laser assembly for excitation contained a GaAlAs laser diode emitting around 787 nm (20 mW peak power, modulated with a 50% duty cycle) and a focusing lens (focal length $f = 46$ mm). The detector consisted of a three-stage Peltier-cooled APD. The detector assembly contained a plano-convex lens ($f = 31$ mm) to collect the fluorescence image, three identical band-pass filters (825 ± 15 nm)

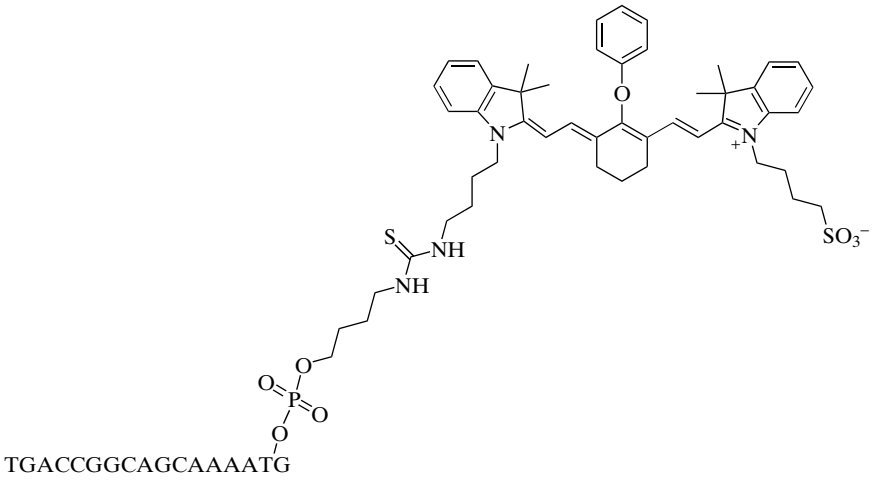

(38)

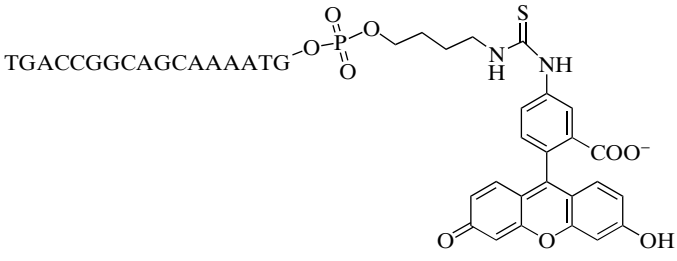

(39)

to reduce background noise from laser scattered light (Rayleigh), and a second plano-convex lens ($f = 31$ mm) to refocus the signal on to the APD photoactive area (0.5 mm diameter). The interface between the LIF cartridge and the NIR fluorescence detector comprised an aspheric condenser lens mounted immediately past the beam probe assembly. An 800-μm circular aperture was installed at the focal point of the condenser lens to reduce stray light. The NIR detector was fixed to an $x-y-z$ micrometer stage installed on the mounting plate and focused on the image produced by the interface lens. The diode laser was disconnected from the LI-COR microscope assembly and was mounted on another $x-y-z$ micrometer stage mounted on the side of the CE base. The laser signal was focused directly on the excitation fiber, resulting in a 4 mW average excitation power at the capillary interface. The optical path of the complete system is shown in Figure 2.[28]

Fluorescence emission was collected by the mirror and reflected as a collimated beam at 180° from the angle of excitation and the collimated beam was then focused by the aspheric condenser lens to the appropriate image size, at the focal length of the detector. The detector optics filter the fluorescence signal through three band-pass filters prior to focusing the signal on the window of the APD detector. Instrument control and data collection of the P/ACE™ 5000 CE system were fully automated through the Beckman System Gold chromatographic software (version 8.01), run on a 486-33 personal computer. The detector signal was interfaced with the chromatographic software. This was accomplished by connecting the analog detector output to a Beckman 406 analog-to-digital convertor (ADC), whose output can be collected by the chromatographic software. In their set-up, the P/ACE™ 5000 CE unit was used to control all CE functions such as capillary rinses and injection and separation parameters and in addition, it was

Table 6 Electrophoretic parameters and detection limits for NIRD-labeled primer and fluorescein-labeled primer. [Adapted from Williams and Soper[40]]

Parameter	NIR primer	FITC primer
Migration time (s)	2522	660
Apparent mobility ($cm^2\ V^{-1}\ s^{-1}$)	5.6×10^{-5}	2.1×10^{-4}
Injection volume (L)	2.9×10^{-9}	1.1×10^{-8}
Amount injected (mol)	3.8×10^{-19}	5.8×10^{-17}
Net signal (counts s^{-1})	33 500	15 920
Background (counts s^{-1})	10 000	19 490
SNR	335	114
Detection limit (mol) (SNR = 3)	3.4×10^{-20}	1.5×10^{-18}
Quantum yield of tag	0.07	0.90

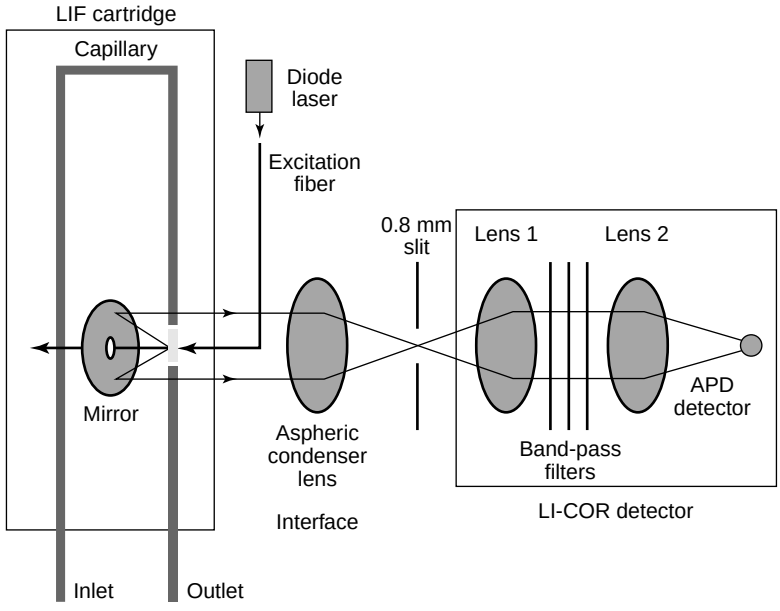

Figure 2 Optical path of NIR/LIF interface with CE instrument.

used for data collection and chromatographic integration, reporting, and plotting functions. The maximum detector output was matched to the 2-V input limit of the ADC using a simple voltage divider circuit. A block diagram of the main instrument components is shown in Figure 3.

The image size produced by the aspheric condenser lens was evaluated for proper matching of the image size to the photodiode active surface area. The experimentally determined mean diameter is 0.575 mm, indicated a slight overfilling of the 0.5-mm detector window. The rapid change in signal upon minor detector position was a good indication of an acceptable image size match. A three-dimensional representation of the image produced by the aspheric condenser lens is shown in Figure 4.

The sensitivity of the system was evaluated by injecting serial dilutions of the labeling dye NN382 **(40)**, (λ_{abs} = 776 nm, λ_{em} = 796 nm, H_2O) which is suitable for use with the systems excitation and detection wavelengths (Figure 5). A linear detector response ($r = 0.9999$) was observed over more than a 250-fold concentration range between 1.4×10^{-9} and 5.5×10^{-12} M (Figure 6). A SNR, signal/root mean square noise) of 15 was observed for a 70-nL injection of the 5.5×10^{-12} M solution, which indicated a limit of detection (SNR = 3) of approximately 80 zmol. The electropherogram of a 70-nL injection of the 5.5 pM solution (385 zmol injected) is provided in Figure 7.

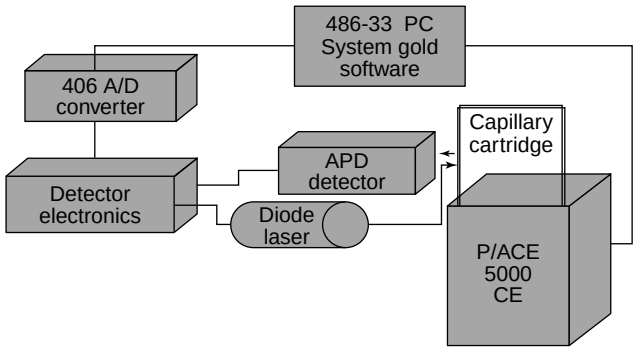

Figure 3 Experimental instrument component diagram.

The sensitivities obtained were under 100% aqueous conditions, and additional improvements could be expected with the use of organic modifiers, cyclodextrins, or micellar additives above their critical micelle concentration in the run buffer, from an improved fluorescence quantum yield when the dye is exposed to a more hydrophobic environment. Comparison of the photophysical properties of the dye in methanol suggested that a fourfold improvement in signal could be obtained when using the dye NN382 in a more hydrophobic run buffer, containing a cosolvent such as methanol. Additional improvements in detection sensitivity could be attained using an NIRD with an emission wavelength better matched to the band-pass filters in the detector. The current system utilizes three 825-nm band-pass filters that significantly attenuate the fluorescence signal emitted at 800 nm. The relative fluorescence signal reaching the detector is only a fraction of the fluorescence produced.

Light collection efficiency and lens and filter transmission efficiency influence the amount of light that will

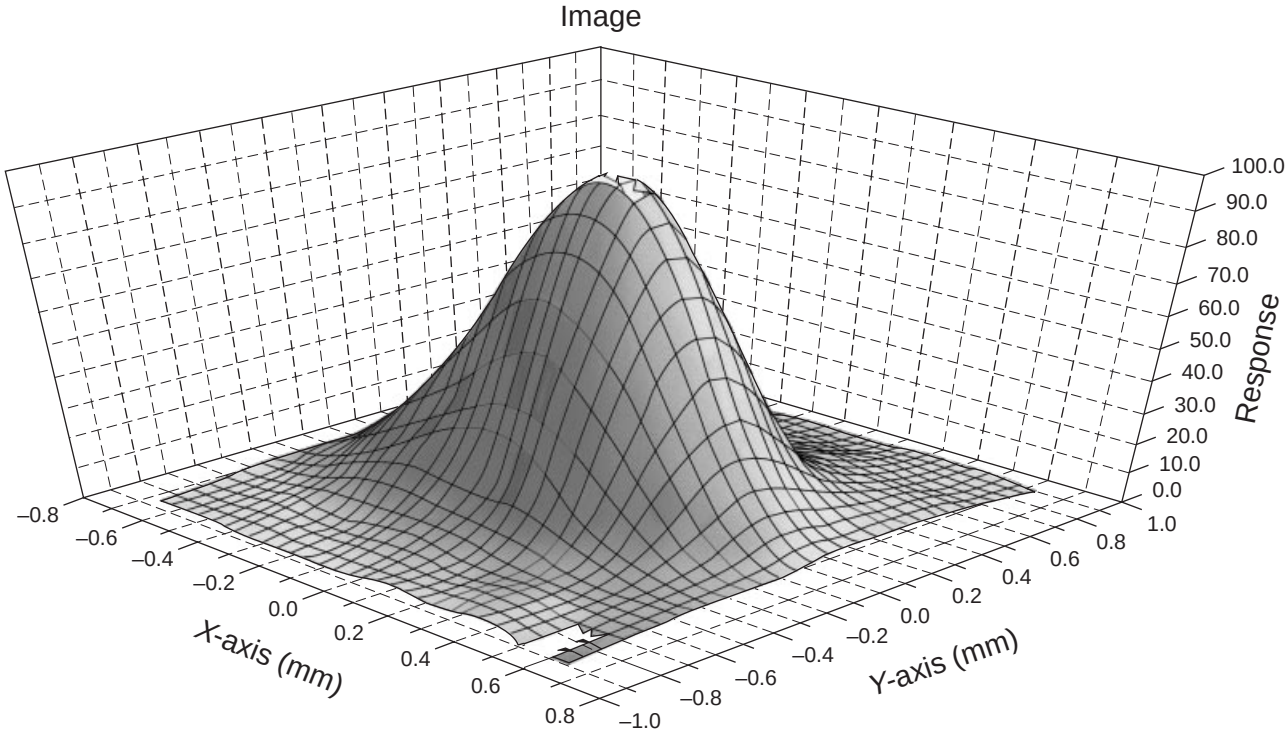

Figure 4 Three-dimensional representation of the image produced by the aspheric condenser lens.

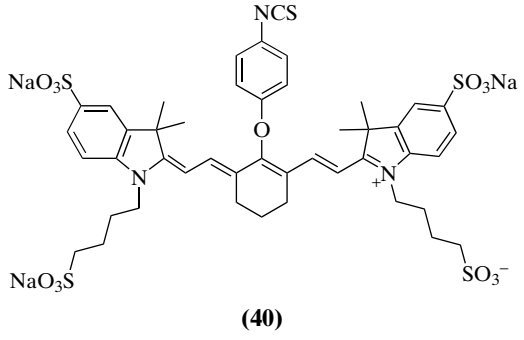

(40)

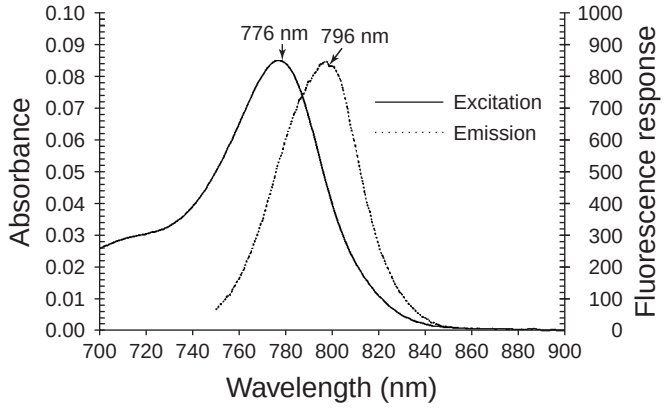

Figure 5 Absorbance and fluorescence spectra of NN382 (in phosphate buffer, pH 7.2).

actually reach the detector. Thus, the maximum efficiency of the optical path in the experimental set-up can be estimated as follows:

- light collection efficiency of mirror: 10%;
- transmission efficiency of aspheric condenser lens: 98%;
- transmission efficiency of first detector lens: 98%;
- maximum transmission efficiency of each band-pass filter: 70%;
- transmission efficiency of second detector lens: 98%.

The overall result is that only about 3.2% of the fluorescence signal reaches the detector.

The quantum efficiency of the detector (even though relatively high for an APD at 800 nm, ~80%) further reduces the number of photons that actually produce a detector response. The improvement that can be attained when better matching band-pass filters with the dye emission profile, or dye selection with the system band-pass filters, can be estimated. A comparison made between the area of the NN382 **(40)** emission curve from 810 to 840 nm and the maximum area that could be attained from 781 to 811 nm (±15 nm from the emission maximum) shows that maximum area was 70% larger than the area of the curve corresponding to the band-pass filter wavelengths. Using a dye with an emission wavelength

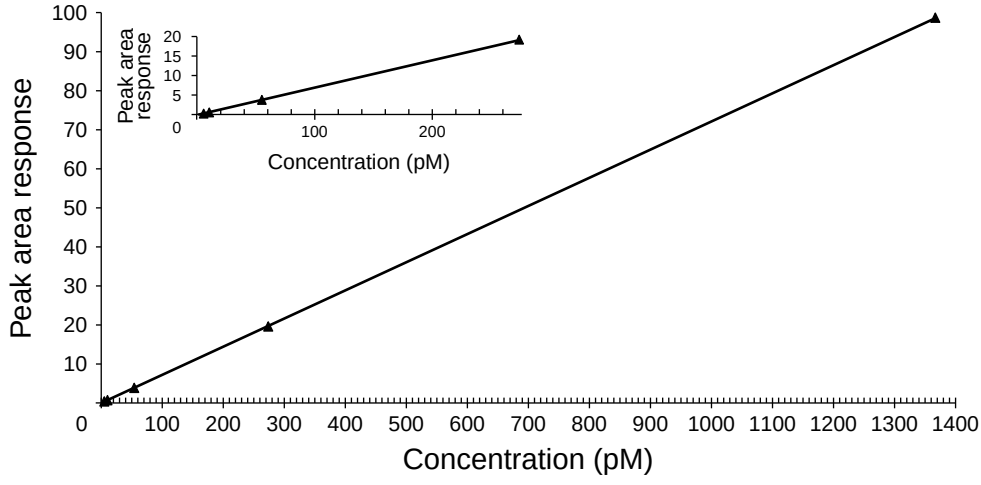

Figure 6 NN382 response curve (384–98 zmol on-column).

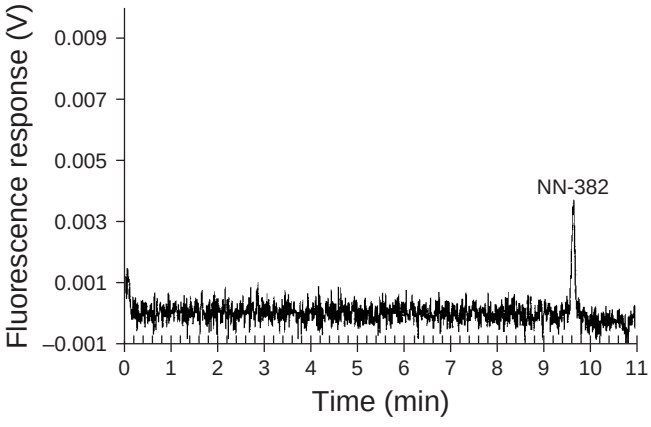

Figure 7 Electropherogram of 70-nL injection of 5.5×10^{-12} M NN382.

around 825 nm could therefore result in about a 1.7-fold increase in signal, assuming a comparable fluorescence emission profile. The noise would remain unchanged, suggesting that an additional 1.7-fold improvement in SNR could be attained. All the above-mentioned factors demonstrate important aspects of LIF instrumentation design. To optimize sensitivity, laser excitation should occur below the absorption maximum when a dye has a relatively short Stokes shift. Band-pass filters should be optimized for the fluorescence emission maximum of the dye. This design will effectively attenuate the Rayleigh scatter from the laser while optimizing fluorescence signal collection.

The automated NIRLIF system developed has sensitivity in line with state of the art results obtained by several investigators utilizing highly optimized laboratory-made detection systems. The latter systems utilize high-quality microscope optics for fluorescence emission collection and tightly focused laser excitation beams. Few investigators have reported limits of detection below the zeptomole range. Sub zeptomole detection limits have been achieved using CE by investigators including Soper et al.[38] (0.4 zmol) and Dovichi et al.[33] (0.050 zmol). Impressive sensitivities as low as single-molecule detection have also been reported for molecules in a flowing stream, demonstrating the possibilities of LIF when used with CE.[42,43] In order to attain these sub zeptomole levels, an off-column sheath-flow detection arrangement was generally utilized to reduce the amount of background scatter. Neat dye solutions were used under conditions that optimized the quantum yield of the dye (Φ), not necessarily desirable for real CE applications. High laser powers, time gating and data smoothing have also been utilized to achieve these results.

The data presented here are for an NIR labeling dye used under 100% aqueous conditions in a fully automated instrument, suitable for routine use. Additional improvements in sensitivity could be achieved by increasing laser excitation power, improving the dye fluorescence quantum yield, and optimizing the dye emission profile with the band-pass filters. The NIRLIF system uses more compact, rugged, and less expensive components with an expected lifetime of >100 000 h. The fully automated injection, separation, and data collection capabilities of the commercial CE system were preserved.[28]

The fully automated system developed was then used to evaluate the suitability of the NIRD NN382 **(40)** as a peptide labeling agent. Six angiotensin-I (Ang-I) variants were selected as model peptides for derivatization and separation studies.[44] The calculated charge-to-mass ratios of the derivatives and excess dye (at pH 7.2) are given in Table 7, listed in expected order of elution in a normal polarity CZE separation. The smallest charge-to-mass ratio difference is between the human and Val-5 forms, where the substitution of valine for isoleucine

Table 7 Sequence of labeled Ang-I peptides listed in order of decreasing charge to mass ratio and expected elution order in a normal polarity CZE system (amino acid differences from the human form are bold and underlined)

Ang-I variant	Sequence	Charge/mass
Goosefish	(NN382)-**Asn**-Arg-Val-Tyr-**Val**-His-Pro-Phe-His-Leu	−0.001237
Elasmobranch	(NN382)-**Asn**-Arg-**Pro**-Tyr-Ile-His-Pro-Phe-**Gln**-Leu	−0.001274
Salmon	(NN382)-**Asn**-Arg-Val-Tyr-**Val**-His-Pro-Phe-**Asn**-Leu	−0.001289
Human	(NN382)-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu	−0.001664
Val-5 human	(NN382)-Asp-Arg-Val-Tyr-**Val**-His-Pro-Phe-His-Leu	−0.001674
Bullfrog	(NN382)-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-**Asn**-Leu	−0.001732
NN382		−0.003006

represents a difference of a single methyl group for the 2295 (formula weight) labeled peptide.

The closely related decapeptides were labeled with the NIRD, separated using CE, and detected by NIRLIF. Derivatization of the peptides was achieved under aqueous conditions using 2.5–500 pmol of Ang-I in a 50-μL sample (5×10^{-8}–1×10^{-5} M). The fluorescence response was linear over a 200-fold range ($r > 0.9986$) and the SNR of a 1.3-amol injection of the Ang-I variants was >15, indicating a limit of detection (SNR = 3) of <240 zmol. Four of six peptides were resolved from each other and excess dye using CZE. Two pairs of co-eluting peptides were successfully resolved using MEKC. Two neutral NIRDs were identified as suitable markers for measuring electroosmotic and micellar flow.[44]

Another application evaluated with the NIRLIF system was in CGE. The suitability of NIRLIF detection with CGE was investigated. The purity of NIR-labeled primers was assessed, and detection linearity and sensitivity using CGE were evaluated. Slab gel electrophoresis has been used for many years for the analysis of proteins, oligonucleotides, DNA, and other biomolecules. The applications of this technique are very powerful and widespread, but also slow, labor intensive and prone to poor reproducibility.[45] CGE was developed to automate and simplify traditional gel electrophoresis methods. The advantages of using CGE include the ability to automate the process fully, provide on-line quantitative detection, and use smaller sample amounts.[30–32,45] The capillaries are superior at dissipating heat, allowing the use of higher voltages, resulting in improved efficiencies and shorter run times.

LIF detection has been successfully applied to DNA sequencing methods. Swerdlow et al.[46] reported sequencing rates of up to 1000 bases per hour using fluorescent-labeled primers. Another scheme uses fluorescent intercalating dyes, such as those marketed by Beckman, for the analysis of ds DNA with LIF detection. Although the total sample throughput is still limited using CE compared with slab gel methods, this limitation is likely to be overcome. Through the use of one-channel, four-dye detection schemes and the development of multicapillary instruments, this limitation can be overcome. Beckman has recently introduced an eight-capillary CE instrument for DNA analysis and Ueno and Yeung reported on the use of a 100 capillary bundle.[47]

CGE with LIF detection was utilized to assess the purity of NIR-labeled oligonucleotides. The CGE method showed great sensitivity, with limits of detection around 90 zmol and a 1000-fold linear dynamic range. The number of theoretical plates obtained for the primers exceeded 300 000. The excellent sensitivity and high separation efficiency observed for the NIR-labeled primers indicate their suitability for many other CGE applications.[28] Sensitivity results are comparable to those reported by Williams and Soper[40] for a similar NIR-labeled primer. The excellent sensitivity and separation efficiencies obtained for the NIR-labeled primers suggest they may be suitable for many of the other CGE applications described earlier.

The use of fluorescent-labeled primers in binding studies of synthetic oligonucleotides, such as antisense therapeutics, has been demonstrated. Vilenchik et al.[48] monitored the hybridization of a phosphorothioate target with a complementary labeled DNA probe. Protein to DNA binding studies were shown by Xian et al.[49] using a CE mobility shift assay with CGE. An assay to determine the activity of a transcription factor (SpP3A2), obtained from the nuclei of a single sea urchin egg, was based on the difference in mobility of a free or bound fluorescent labeled DNA probe.

4.2.2 Application of Near-infrared Fluorescence in Immunoassays

IA is a method of analysis that relies on specific interactions between an Ab and an antigen (Ag) to measure a variety of substances, ranging from complex viruses and microorganisms to simple pesticide molecules and industrial pollutants. To observe and measure this reaction, a label is introduced via a second Ab. Conventionally, this label consists of a radioactive isotope (radioimmunoasay (RIA)), an enzyme (enzyme-linked immunosorbent assay (ELISA)) or a fluorescent molecule (fluorescence immunoassay (FIA)). The use of Abs as analytical regents was first reported in 1959 when Berson and Yalow successfully demonstrated the measurement of picogram levels of human insulin in samples of body

fluids by RIA.[50] Since then, various IAs for detecting hundreds of molecules of endogenous and exogenous origin have been described. This technique proved to be reliable, fast, and very sensitive; many other RIAs have been developed for clinical and medical tests since then. The use of immunochemical techniques in the environmental field was first proposed in 1971 by Ercegovich,[51] who suggested the use of immunological screening methods for the rapid detection of pesticide residues and for confirming results of conventional analysis. An RIA for the insecticides aldrin and dieldrin was the first reported IA for an environmental contaminant.[52] Although a few RIAs still exist in the medical field, they are seldom used in environmental and food analysis because of the need for special handling and disposal of the radioactive materials. Radiolabels were gradually replaced with enzyme labels owing to the hazards associated with radioactive materials. ELISA, which was first introduced by Engvall and Perlman in 1971,[53] has become perhaps the most popular IA format used in laboratories today. The modern diagnosis of many diseases, especially infectious diseases, is almost completely dependent on these assays. In diseases of global importance such as acquired immunodeficiency syndrome (AIDS),[54] cysticercosis,[55] malaria,[56] filariasis, and schistosomiasis,[57,58] IAs play a key role in screening and diagnosis.

Utilizing the advantages of fluorescence over absorbance, conventional fluorophores can be theoretically capable of detecting fewer than 10^6 molecules L^{-1}[59] in conjugation with the highly specific activity provided by IAs. However, in practice, the high background from the biological samples autofluorescence, or sample reagents and scattering and quenching effects of solvents, limit the detection to about 10^{-10} molecules L^{-1}. LIF provides an alternative to improve the sensitivity of fluorescence in IAs. The use of LIF in IAs has been reviewed.[60–62] The limitations of conventional lasers as excitation sources are their high price, size, maintenance costs and their limited wavelength selection. LIF in the NIR region (600–1100 nm) offers several advantages. Recent advances in semiconductor laser technology have made the use of lasers more practical. The widespread use of NIR-emitting laser diodes in the telecommunications industry has made them more readily available. This type of laser is inexpensive (typically <$150) and small (~1 cm) and has a much longer operating lifetime (>100 000 h).[63]

The GaAlAs laser diode has drawn much interest because its emission wavelength of approximately 800 nm is compatible with several classes of polymethine cyanine dyes, which exhibit NIR fluorescence. Detection in the NIR region allows one to replace commonly used photodiodes by APD. The APDs have excellent quantum efficiency in the NIR region. Some of the advantages of APDs are that they are much cheaper, more compact, and longer lasting. In addition, they have low internal noise and very low power consumption. All these features make the NIR fluorescence IA highly amenable to miniaturization and can aid in the development of a portable, compact and rugged instrument for field application.

Heptamethine cyanine dyes are a class of NIR fluorophores that have been used for DNA sequencing, pH and hydrophobicity determination, metal ion detection, and Ab labeling.[63] They are ideal for labeling Abs involving conjugation chemistry, e.g. the reaction of the isothiocyanato group (NCS) on the NIRD with the primary amine groups on Abs. These dyes have high molar absorptivities (ca. 10^5 L mol^{-1} cm^{-1}), high quantum yields (20–40%), relatively short fluorescence lifetimes (500–1000 ps) and are small (~1000 Da). The small size of these dyes in comparison with the target molecule allows for a high number of labels per Ab without compromising Ag/Ab interactions. This implies higher selectivity and signal coupled with low background interference, in addition to the previously discussed advantages offered by the NIR region. Additionally, the use of a solid matrix generates a stronger signal by concentrating the fluorescent molecules and thereby reducing quenching effects of the solvent.[2]

Boyer et al. were the first to demonstrate the feasibility using diode laser detection in the NIR region.[64] They compared the efficacy of the NIR assay method with the conventional ELISA method. Williams[65] constructed an NIR fluorescence detector for detecting fluorescence in an IA format. The set-up developed is illustrated in Figure 8. The instrumentation comprised an excitation source coupled with a fiber-optic cable, a silicon photodiode as detector, a sample-holding apparatus that could be adapted to hold various forms of solid support matrices coupled with a motor drive, and a data acquisition interface. The instrument could detect concentrations of 100 pmol of NIRDs. Using this set-up, the near-infrared

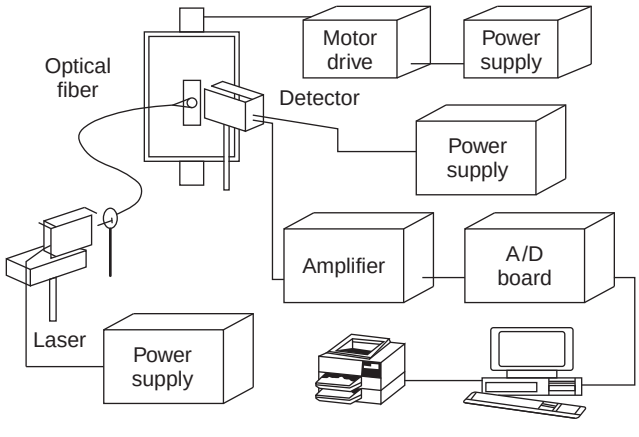

Figure 8 Block diagram of NIR fluorescence detector.

fluorescence immunoassay (NIRFIA) on a nitrocellulose matrix was capable of detecting 5×10^{-10} M human immunoglobulin G (IgG). The overall assay could be performed in less than 2.5 h in the absence of the substrate development step which is employed in the conventional ELISA method. This NIRFIA allowed the detection of approximately 470 00 labeled Ab molecules.

This method developed, however, had its disadvantages. A high degree of scatter generated by the membrane, problems with nonspecific binding of the conjugate, and its lack of compatibility with the most common format of modern IAs, the microtiter plate assay format, limit its practical use. In continuation of the development of solid-phase NIRFIA, the abovementioned issues and others were addressed in a study by Swamy.[2] An NIRFIA was developed based on the heptamethine cyanine dye NN382 **(40)**. The dye used in this study had a high molar absorptivity ($\varepsilon = 180\,000\,\mathrm{L\,mol^{-1}\,cm^{-1}}$) and quantum yield ($\Phi = 0.59$, for dye–Ab conjugate). The isothiocyanate functionality reacts selectively with the amino group of the Abs to form a stable thiourea bond. The presence of the sulfonated groups makes the dye highly water soluble, making it ideal for labeling Abs. Another advantage of the negatively charged sulfonate groups on the dye is that it minimizes nonspecific binding to the solid matrix (polystyrene (PS)).

A solid-phase NIRFIA was systematically developed in the following steps. A LI-COR 4200 prototype fluorescence microscope[41] was coupled with an orthogonal scanner and was evaluated for optimum depth of view. Next, the optimum scanner operating conditions were determined in terms of gain, offset, and scan speeds to optimize settings for maximum dynamic range and SNR and ensure reproducible scans. Appropriate modifications to the mounting of the scanner microscope and scanning platform were made to allow the examination of microtiter plates (Figure 9). The linear response of the detector was verified with increasing gain for a given voltage setting. A number of microtiter plates (solid matrix) were then evaluated in order to provide the one with minimum background noise. Once the reader had been configured and the solid support had been augmented, the procedure for covalent conjugation of NIRD NN 382 **(40)** to goat antihuman immunoglobulin G (GAHG) was systematically optimized in terms of pH, temperature, time of reaction, and reactant molar ratio. The conditions that provided the maximum results were a pH of 10, an initial dye : Ab ratio of 1 : 125, a temperature of 25 °C and a coupling time of 1 h. These conditions are all critical parameters since any deviation results in under- or overlabeling of the Ab, which, in turn, results in a low specific activity and detection ability. The absolute concentration of reactants was found to be critical for the rate of the coupling reactions; therefore, in the optimization procedure, stock solutions of dye ($10\,\mathrm{mg\,mL^{-1}}$) and Ab ($1\text{–}3\,\mathrm{mg\,mL^{-1}}$) concentrations were used. Although the conditions for coupling other species of Ab were not optimized, other conjugates were successfully prepared with high specific activity. The optimum conjugation procedure was defined by the conditions that produced an NIRD-labeled Ab with the highest specific activity in an immunosorbent assay for normal human immunoglobulin G (NHIgG). The best conjugate was used to determine detection limits for the assay and was able to detect 2×10^{-11} M of NHIgG. This provided a sensitivity about 10 times greater than that achieved by conventional labels in similar assays. Finally, the assay was validated by 100% agreement with clinically diagnosed samples for schistosomiasis and its cross-reactors.[2]

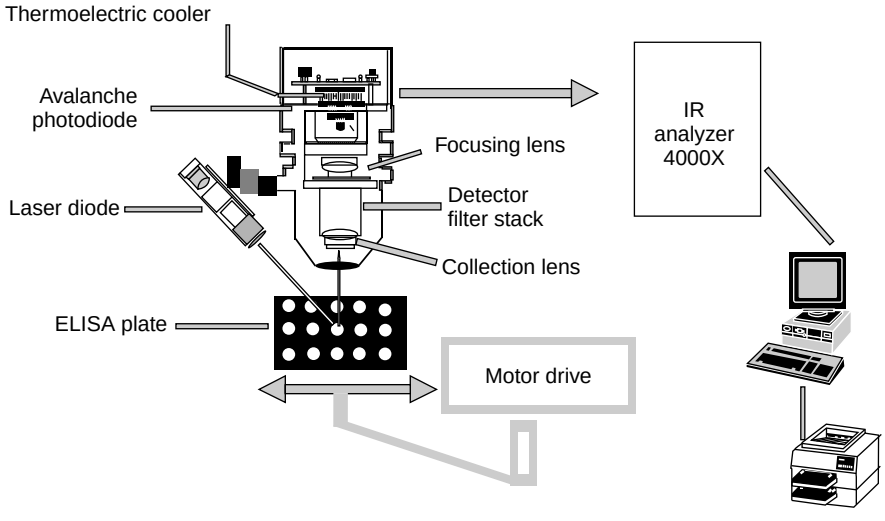

Figure 9 Model assembly to scan microtiter plates.

One major disadvantage of this method is the lack of commercial instrumentation for the assay. Quantification of the signal obtained in an image format is problematic. However, commercial software for quantifying on line signals from DNA sequences in the NIR region is available (LI-COR). Adaptation of similar software for microtiter plate scans would be immensely helpful. The method developed does have many advantages over conventional methods and the versatile nature of the scan bed does not limit this assay to the microtiter format. Technically, the scanner could scan any matrix, limited only by the physical dimensions of the scanner (ca. 23 cm and 61 cm long). This could allow for scanning multiple assays and 500–1000-well plates. The assay is very quick, requiring less than 20 min for the assay itself, and requires fewer steps than ELISA. The ability to detect lower concentrations of Abs would help in earlier diagnosis and an opportunity to test drugs against opportunistic infectious agents at an earlier stage. The labeling of Abs with NIRDs provides an important advancement to IAs and the comparatively small size of the NIRD allows for a higher molar ratio of NIRD per Ab. This in turn yields a higher signal and allows for lower detection limits. The assay has also been validated for clinical samples and could be further evaluated in the development of a diagnostic test.

In principle, various bioanalytical methods can be developed using the NIR immunochemistry described in the preceding section. To demonstrate the versatility of the method, Swamy et al. developed two bioanalytical applications utilizing NIRFIA in collaboration with the State University of New York (SUNY) at Brooklyn and the Department of Entomology and Environment Toxicology at the University of California, Davis, CA, USA. The first application involved an NIR-labeled Ab in the detection of extracellular Ag expressed on cells. This demonstrated the feasibility of using NIR-labeled Abs in clinical imaging applications such as NIR optical tomography. The second application uses the NIRFIA as an analytical tool in environmental applications in the quantitative analysis of two pesticides, bromacil and fenvalerate.

In the evaluation of the NIRD-labeled Ab for imaging applications, the NIRD NN82 was coupled to monoclonal Ab E48 (MAb E48) and the conjugate was evaluated in a direct assay to detect a 22-kD Ag expressed on the surface of human squamous cell carcinoma line (HuSCC) A431. The specificity of binding was confirmed in a competitive assay. The preliminary results showed good sensitivity and specificity of the NIRD-labeled Ab. In the environmental application, the assay was evaluated in a rapid tracer format competitive assay for the detection of the pesticides bromacil and pyrethroid. In this assay, the specific Ab captured on the surface of microtiter plate well and the tracer with fluorescent label compete with the analyte for the Ab. The results obtained showed that the NIRFIA was at least as sensitive as ELISA, with both assays detecting pesticides in the micrograms per liter (parts per billion) range.[2]

The chemistry and the instrumentation of the NIRFIA technique for environmental applications have been developed, but are amenable to further miniaturization. Further optimizations would allow for the future construction of compact instrumentation for analysis which could conduct fast and reliable assays at remote locations. In continuation of this approach, preliminary results obtained in our laboratory demonstrated the feasibility of using the NIRFIA approach for an NIR/fiber optic immunosensor (FFOI). This approach is particularly useful for measuring small amounts of analyte and can be automated and carried out in remote locations. Several fluorescent immunosensors have been reported; however, lack of commercial instrumentation and labels limit the use of these techniques and most applications utilize UV/VIS dyes. The applications of the visible fluorescent immunosensors described are all susceptible to interference from biomolecules such as bilirubin and porphyrins. A comprehensive review of immunosensors has been published by Robinson et al.[66] Danesvar[67] developed a NIR/FFOI for detection of trace amounts of human IgG and *Legionella pneumophilia*. They optimized the assay on different solid phases such as poly(methyl methacrylate) (PMMA) and PS. The PS coating method eliminated the activation process required with PMMA, thereby reducing the preparation time by 18 h; in addition, it also eliminated the protein G step of the previous assay. The assay was carried out in a sandwich format (Figure 10). The set-up used to measure the signal (Figure 11) allowed concentrations of 10–11 M for IgG (SD = 0.13) and 0.5 ng mL^{-1} for *Legionella pneumophilia* sera type 1 (LPS1)[67] to be detected. These results were comparable to those attained by conventional ELISA. Another application using a similar format was developed by Evans[68] and provided a competitive assay for the pesticide bromacil. A detection limit of 5 ppb was obtained, which was slightly higher than that reported for ELISA (0.1 ppb). However, the method had advantages since it did not require extensive sample preparation, the assay time was substantially reduced, and the assay could be adapted for remote site analysis.[68]

The NIR/FFOI technique in principle is a flexible methodology, which can be theoretically adapted to develop assays for any compound for which Abs are available, including infectious agents, serum analytes, and environmental pollutants. This technique provides a real-time analysis with automated readout capabilities, eliminating the need for operator intervention or

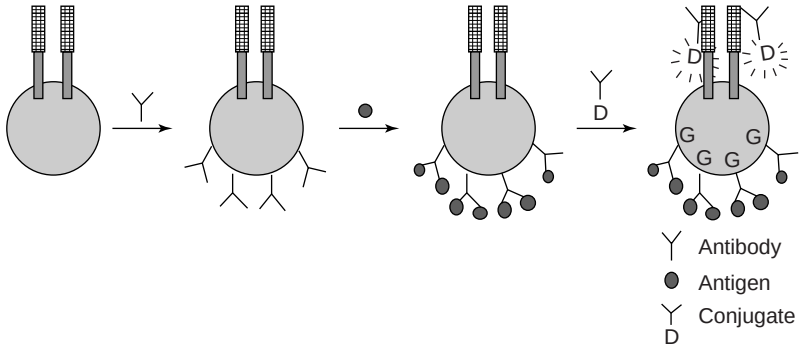

Figure 10 Sandwich assay on fiber-optic probe.

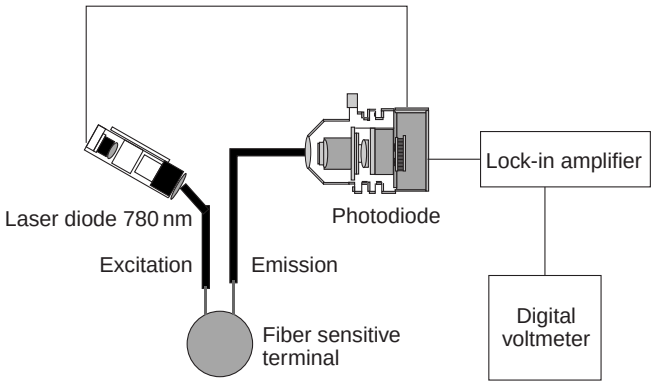

Figure 11 Fiber-optic immunosensor set-up.

complicated sample preparation. Although the detection optics for the assay were originally constructed on a 2×3 ft optical table, the set-up can be reduced in size. The preliminary sensitivity, selectivity, and simplicity of the assay are encouraging and further design refinement and other applications should aid the development of versatile NIR/FFOI.

4.2.3 Application of Near-infrared Fluorescence in High-performance Liquid Chromatography

Another bioanalytical application developed using the advantages of NIR fluorescence is HPLC detectors. In recent years, the ability to measure ultralow levels of pharmaceuticals in biological matrices has presented significant challenges. Using fluorescence as a tool for detection, sensitivities of picograms per milliliter have been achieved. Rahavendran and Karnes have described the advantages of LIF detection in HPLC.[69] There have not been many publications dealing directly with HPLC detection in the NIR region. Until recently, two groups worked on the development of applications of NIR fluorescence in HPLC, Winefordner's group at the University of Florida and Ishibashi's group in Fukuoka, Japan.

One of the first NIR fluorescence-based HPLC detectors was developed by Ishibashi's group. They detected 0.3 pg of a carbocyanine dye,[70] with instrumentation comprising a 3-mW diode laser emitting at 780 nm with a PMT cooled to $-20\,^{\circ}$C for detection. They reported a detection limit of 1.9 pg at room temperature. Sternberg et al.[71] of Beckman Instruments used a 2.5-mW diode laser coupled with a commercial diode-array detector and achieved a detection limit of approximately 10^{-10} M for a carboxyl cyanine-labeled oligonucleotide. Winefordner et al. explored the detection of cyanine dyes in different instrumental configurations,[72] using a diode laser for excitation with a PMT for detection. They achieved detection of 46 000 molecules in a liquid jet fluorescence spectrometer. Karnes et al.[73] used diode laser-induced NIR fluorescence for detection of ICG in plasma. They demonstrated a detectability of more than two orders of magnitude over absorbance.

Another NIR HPLC application was reported by Kuklenyik,[74] who developed and compared two detectors for NIR fluorescence detection based on a

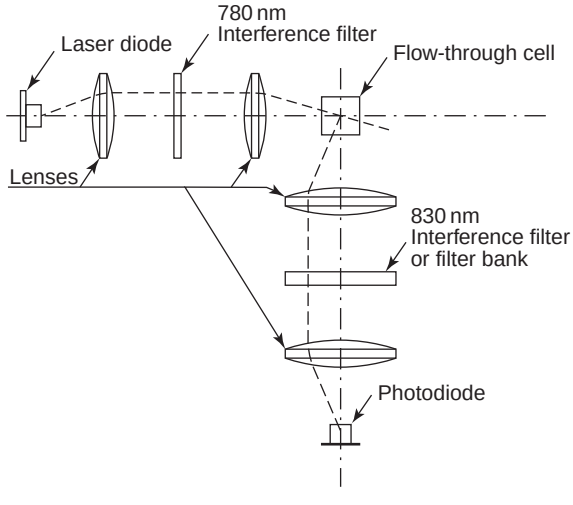

Figure 12 Optical path of NIR/HPLC fluorescence detector.

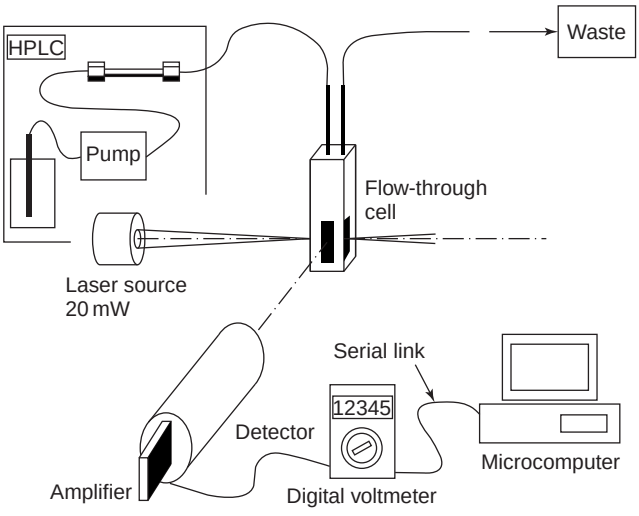

Figure 13 Experimental set-up of NIR/HPLC fluorescence detector.

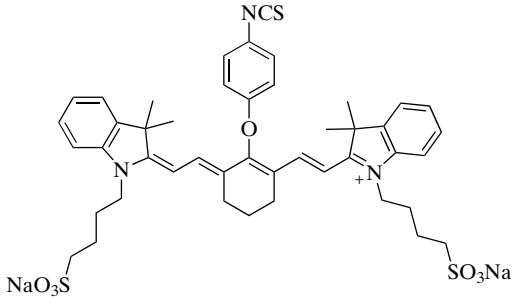

(41)

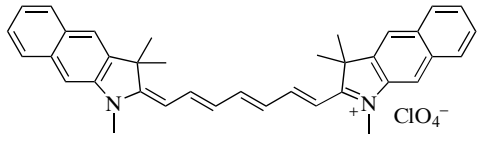

(42)

silicone photodiode and an APD. The merits of this type of detector were also compared with those of conventional PMT detectors. An HPLC system (SSI gradient system) with a high-pressure pump capable of providing 500 psi pressure controlled by an IBM computer was coupled with an in-house developed excitation and detection system. The experimental set-up of the excitation and detection and the final set-up are shown in Figures 12 and 13. It comprised a silicon photodiode detector coupled with a 24-μL HPLC flow cell shown in Figure 12 and, in the set-up with an APD, a modified Model 4200 LI-COR microscope[41] was used instead of the silicon photodiode. Detector performance was evaluated with two NIRDs, IRD40™ **(41)** and 1,1′,3,3,3′,3′-hexamethyl-4,4′,5,5′-dibenzoindotricarbocyanine perchlorate **(42)**, first without an HPLC column followed by subsequent

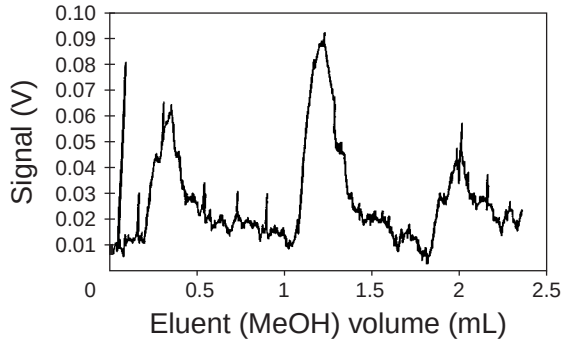

Figure 14 Detection of three consecutive injections of IRD40™ on a C_{18} column.

runs with C_{18} HPLC columns. Detector signals for three consecutive injections of IRD40™ at 10^{-10} M concentration are shown in Figure 14. Detection limits of 1.22×10^{-11} and 5.89×10^{-12} M were achieved for the two dyes and the numbers of dye molecules detected were calculated to be 36 500 and 17 600, respectively.[74] The overall results showed that the less expensive, more efficient photodiode was sensitive for detection in the NIR region. The detector was designed to collect signals at 820 nm, and a better match of the dye emission with the detector and better light collection efficiency should help improve the sensitivity much more. Even in the early stage of development in this direction, the NIR fluorescence method shows sensitivity comparable to or better than those of conventional methods. Further evaluations and modifications could help achieve much greater sensitivities.

4.3 Environmental Applications of Near-infrared Fluorescence

The fundamental principle involving the analytical applications of dyes to study the microenvironment essentially involves the changes induced by the analyte that can be directly measured by spectral changes. These changes depend on the concentration of the analyte and the interaction between the analyte of interest and the dye. Typically these probes involve a cation-selective receptor either as an integral part of the chromophore's π-system or covalently attached to the fluorophore via an alkyl tether.[75–77] The coordination of the cation affects the spectral position of absorption and emission bands, molar absorptivity, and, to a smaller extent, the fluorescence quantum yield. The interaction of a metal ion with an organic reagent might result in the enhancement of the fluorescence or the quenching of the fluorescence of the organic reagent in the presence of an analyte. A statistical study showed that about 89% of luminescent methods are based on the enhancement of fluorescence and about 11% are based on quenching reactions.

The investigation of toxic metal ions in the environment has been of growing interest in the past few decades and the decreasing permissible limits for most of these metal ions set by major health organizations such as the World Health Organization (WHO) and the United States Environmental Protection Agency (USEPA) have provided a tremendous thrust for the development of sensitive analytical techniques. Most of the publications devoted to the detection of metal ions by means of spectroscopic techniques used UV/VIS probe molecules such as rhodamine, fluorescein, 8-hydroxyquinoline, and their derivatives.[78–83] Table 8 provides a list of some of the different fluorescent sensors that have been used for the determination of the most commonly studied metal ions, namely Al(III), Be(II), Co(II), Fe(III), and Li^+. Unfortunately, spectral interference is significant in this region. The use of NIRDs is a better alternative.[84–90] The low background interference observed for other molecules in the NIR spectral region together with the longer Raman shift offers better SNRs. Casay et al.[91] investigated tetrasubstituted NPc NIRDs as potential probes for the determination of toxic metal ions. They developed the first NIR optical probe that could detect lead, lithium, and cadmium at parts per billion levels[91] and potassium at 0.566 ppm.[92] More recently, Tarazi[93] reported the detection of metal ions using three novel NIRDs, namely TG 170 **(43)**, NN525 **(44)**, and JCM-15C5 **(45)**. The spectral characteristics of the three dyes are listed in Tables 9–11. They evaluated TG 170 for the detection of Al(III) and Be(II), NN525 for the detection of Fe(III) and Co(II) ions, and JCM-15C5 for the detection of Li^+ in the presence of other interfering ions.

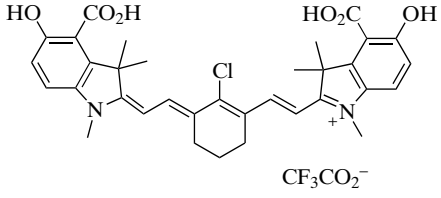

(43)

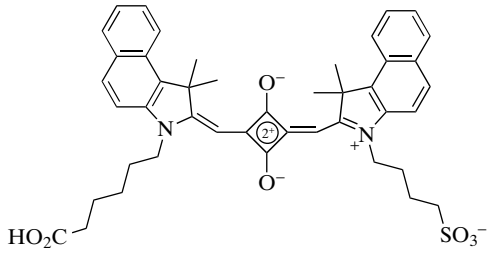

(44)

Table 8 Different fluorescent methods currently employed for detection of various metal ions

Metal ion	Reagent	Method[a]	Detection limit (ppm)
Al(III)	Alizarin Red S	E	8×10^{-2}
	Lumogallion	E	3×10^{-2}–3.2×10^{-1}
	8-Hydroxyquinoline	E	5×10^{-3}
	Morin	E	2.5×10^{-4}–5×10^{-3}
	N-Salicylidene-2-hydroxy-5-sulfoaniline	E	8×10^{-5}–8×10^{-2}
	Superchrome Carnet Y	E	8×10^{-4}–1.6×10^{-1}
	2-Hydroxy-3-naphthoic acid	E	1.8×10^{-4}–1.8×10^{-3}
Be(II)	2-Ethyl-5-hydroxy-7-methoxyisoflavone	E	4×10^{-4}–1.2×10^{-2}
	2-Hydroxy-3-naphthoic acid	E	1.8×10^{-4}–1.8×10^{-3}
	Morin	E	4×10^{-4}–1.6×10^{-3}
Co(II)	Al(III)–Pontachrome BBC	E	10^{-3}
	Al(III)–Superchrome Blue Black	Q	10^{-3}–2×10^{-2}
	2-[2-Hydroxy-1-naphthyl]dithiocarbazic acid-4-chlorobenzyl ester	Q	10^{-1}–1.1
	1-(2-Pyridylazo)-2-naphthol	E	5.9×10^{-2}
	Al(III)–Pontachrome BBR	Q	2×10^{-2}–2×10^{-1}
Fe(III)	4′-(4-Methoxyphenyl)-2,2′,2″-terpyridyl	Q	10^{-2}–10^{-1}
	Rhodamine B	Q	1
	2,2′,2″-Terpyridyl	Q	10^{-2}–5×10^{-1}
Li^+	Dibenzothiazolylmethane	E	5×10^{-2}
	5,7-Dibromo-8-hydroxyquinoline	E	2.5×10^{-1}–2.5
	1,4-Dihydroanthraquinone	E	5×10^{-2}–5×10^{-1}
	1,5-Dihydroanthraquinone	E	5×10^{-2}–5×10^{-1}
	1,8-Dihydroanthraquinone	E	5×10^{-2}–7×10^{-1}
	8-Hydroxyquinoline	E	4×10^{-2}–1

[a] E = enhancement of fluorescence; Q = quenching of fluorescence.

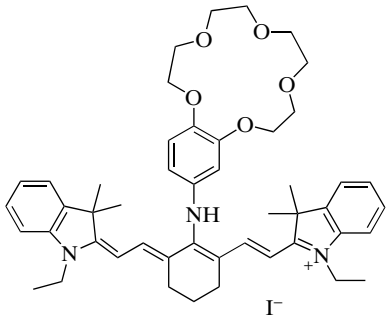

(45)

Table 9 Spectroscopic characteristics of TG170 in various solvents

Solvent	λ_{max} (nm)	λ_{em} (nm)	Φ_f (%)	Log ε ($L\,mol^{-1}\,cm^{-1}$)
Dimethyl sulfoxide	819	822	0.167	4.96
Methanol	805	812	1.30	5.40
Buffer (pH 5.9)	791	803	2.48	5.99

Table 10 Spectroscopic characteristics of NN525 in various solvents

Solvent	λ_{max} (nm)	λ_{em} (nm)	Φ_f (%)	Log ε ($L\,mol^{-1}\,cm^{-1}$)
Acetonitrile	664	673	0.86	5.64
Dimethylacetamide	672	681	0.92	5.74
Dimethylformamide	672	681	0.81	4.67
Dimethyl sulfoxide	676	688	0.73	4.69
Methanol	663	670	0.87	5.72
Tetrahydrofuran	669	676	0.91	5.76
Water	661	666	0.03	5.38

Table 11 Spectroscopic characteristics of KVA22 in various solvents

Solvent	λ_{max} (nm)	λ_{em} (nm)	Φ_f (%)	Log ε ($L\,mol^{-1}\,cm^{-1}$)
Acetone	780	799	3.02	4.12
Acetonitrile	776	799	1.58	4.83
Dimethyl sulfoxide	793	809	2.03	4.66
Methanol	777	780	1.52	4.81
Methylene chloride	787	803	2.74	4.96

The hydroxy carboxy functionality of TG 170 **(43)** has a known selective complexation with Al(II) and Be(II) ions. Using this dye, they reported a detection limit of 52 ppb for Al(III) and 24.6 ppb for Be(II) ion. For the NIRD NN525 **(44)**, only Fe(II), Fe(III), and Co(II) ions resulted in quenching of fluorescence. Detection limits of 3.54 ppb for Fe(III) and 1.55 ppb for Co(II) were obtained. The calibration curves obtained for the

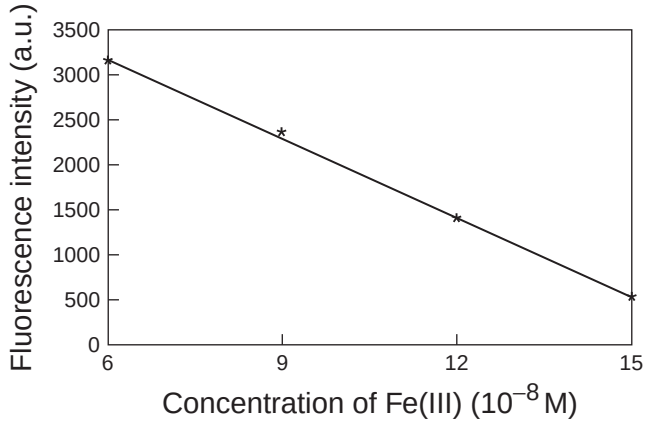

Figure 15 Calibration curve for detection of Fe(III) with NN525.

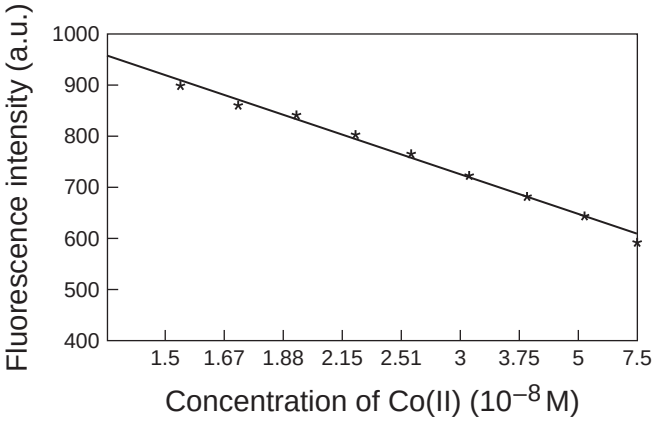

Figure 16 Calibration curve for detection of Co(II) with NN525.

two metal ions are illustrated in Figures 15 and 16. The crown ether functionalized cyanine dye JCM-15C5 **(45)** was used for detection of Li^+ ions in the presence of other interfering ions. They obtained a linear plot of fluorescence intensity for Li^+ amounts ranging from 0.0347 to 0.222 ppb (Figure 17). A detection limit of 0.0743 ppb was reported. The stability constant for the JCM-15C5 complex was higher than those achieved in other studies. The method development represents a significant improvement over conventional methods. Owing to the minimal interferences encountered in the NIR region, it eliminates the need for extensive sample preparation, and additionally it allows for the development of tests that can be carried out at remote sites by rugged compact instrumentation.

Another novel application developed with NIRDs for the detection of Ca^{2+} was reported by Akkaya and Turkyilmaz.[94] They developed a squaraine-based fluorescent sensor for calcium [Scheme 6, **(47)**]. The probe developed was very sensitive to Ca^{2+} concentrations in the

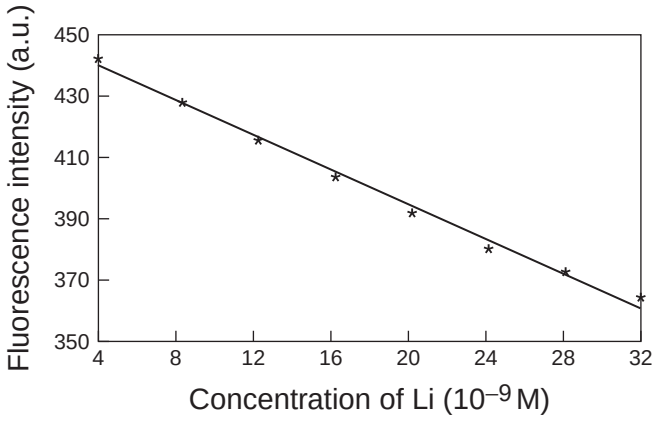

Figure 17 Calibration curve for detection of Li^+ with JCM-15C5 **(45)**.

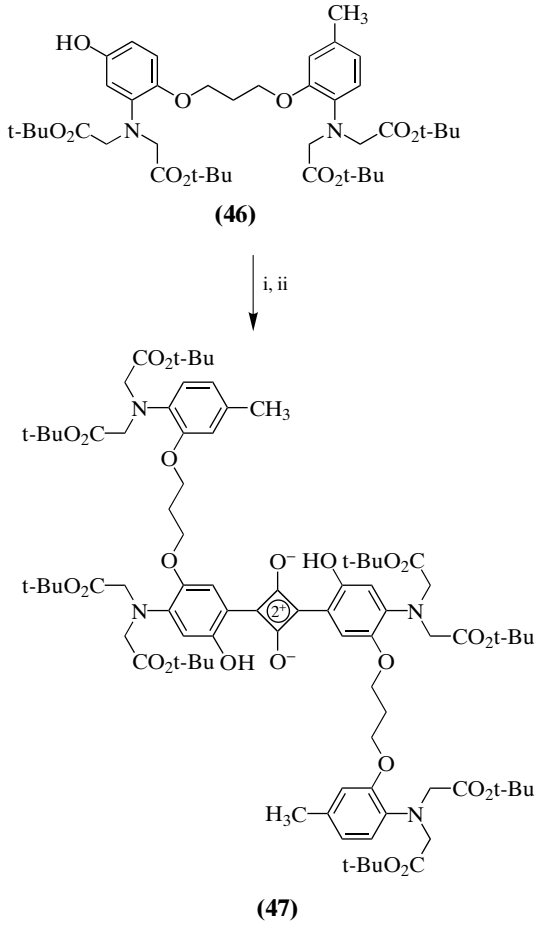

Scheme 6 Synthesis of Ca^{2+} fluorescent chemosensor [Akkaya and Turkyilmaz[94]].

micromolar range (Figure 18). Signal changes resulting from the loss of the donor alkylamine functions on Ca^{2+} chelation, decreasing the molar absorptivity from 300 000 to 50 000. The λ_{max} of the dye at 698 nm permits the use of an inexpensive commercial laser diode emitting at 690 nm for excitation and a photodiode for detection, opening avenues for the development of compact, rugged systems free from background interferences for cellular ion-flux studies and real-time imaging applications.[94]

Another unique environmental application developed was an NIR phase-resolved fluorescence spectroscopy-based fiber-optic corrosion sensor for the detection of aluminum and iron and their corrosion by-products by Chin et al.[95] The method essentially involves an NIRD that is embedded in a fiber-optic cladding, coupled with appropriate excitation and detectors and integrated with an in-house phase-resolved fluorescence spectroscopy board. A block diagram of the set-up is shown in Figure 19. They studied three in-house synthesized NIRDs, TG1 **(48)**, TG2 **(49)**, and TG3 **(50)** for complexing Al(III) ions. The method was more effective than conventional fluorescence techniques because it measures the lifetime changes of the dye which are independent of the surrounding environment as opposed to fluorescence. Table 12 lists the fluorescence lifetimes for the three dyes in the presence of Al(III) ions. This method developed was very versatile as it could be integrated into airframe and wings of aircraft or other critical failure locations (fasteners, bolts, composite joints, welds) for measuring corrosion. Thereby, strain and structural fatigue can be monitored using NIR technology. The preliminary results obtained were very encouraging and future potential uses of this system could include monitoring composite life cycles and corrosion and structural fatigue of launch pads, rockets, space structures, and advanced aircraft.[96,97]

4.4 Other Applications of Near-infrared Fluorescence

In this section, we describe other analytical applications developed with NIR fluorescence. The first application discussed is the determination of hydrogen ion concentration (pH), which is an important part of many

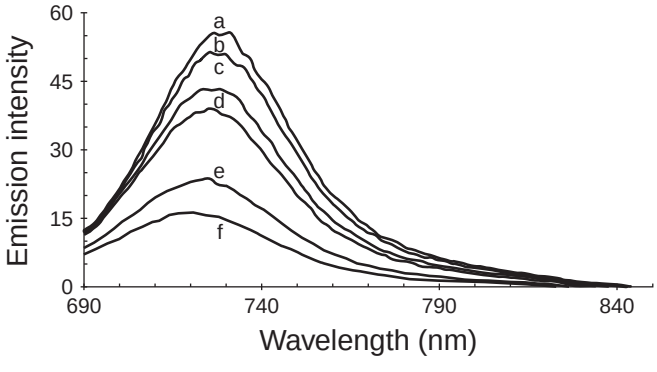

Figure 18 Emission spectrum of fluorescent chemosensor as a function of Ca^{2+} concentration.

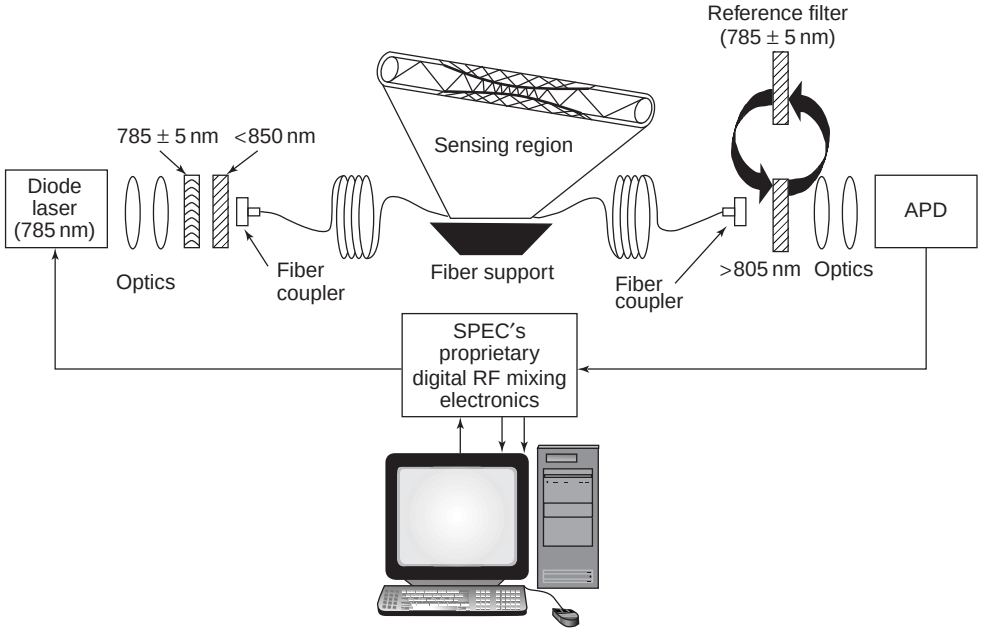

Figure 19 Fiber-optic set-up for phase-modulated fluorescence detection of metal ions.

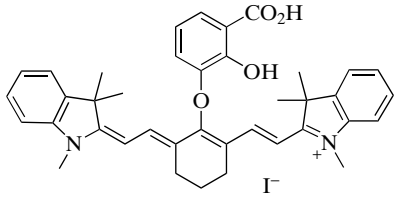

(48)

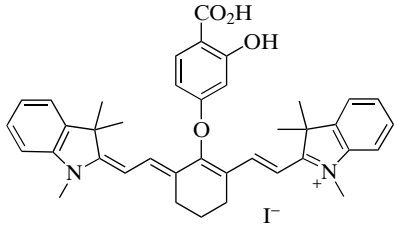

(49)

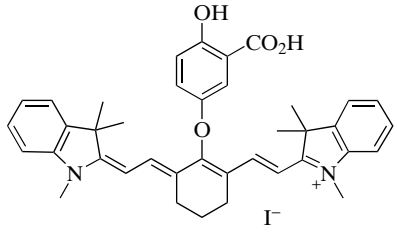

(50)

analytical procedures. The use of fluorescent dyes and indicators for pH measurement is well documented in the literature. A number of indicators with spectral characteristics in the UV/VIS region have been reported,[98–101] but only a few have been reported for pH measurements in the NIR region. Several classes of NIR pH-sensitive dyes have been developed in our laboratories at Georgia State University in collaboration with Strekowski et al.[102–105] Of these pH-sensitive dyes, the most useful include the bis(aminodiene)ones that undergo transformation from the keto form to the cyanine enol form. A pH-sensitive dye **(51)** (Scheme 7) showed distinct bands under acidic and basic conditions (Figure 20). The absorbance maximum at 531 nm undergoes a strong bathochromic shift in acidic conditions with a concomitant change in fluorescence spectra with an absorbance band at 709 nm. The change in absorbance maximum is attributed to protonation of the oxygen atom which induces a shift to the cationic enol form of the dye and restores the cyanine chromophore to give a bathochromatic shift to the NIR region. Another pH-sensitive NIRD **(53)** was evaluated. Representative spectra of the NIRD trapped in Nafion

Table 12 Lifetime data for NIRDs TG1, TG2 and TG3

Dye	Dye concentration (M)	Al(III) concentration (M)	τ (ns)
TG1	3.71×10^{-6}	0.00	0.794
TG1	3.71×10^{-6}	6.32×10^{-3}	0.523
TG2	2.33×10^{-6}	0.00	0.792
TG2	2.33×10^{-6}	6.32×10^{-3}	0.561
TG3	6.59×10^{-6}	0.00	0.387
TG3	6.59×10^{-6}	6.32×10^{-3}	0.455

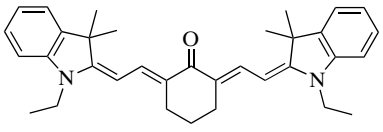

(51)

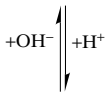

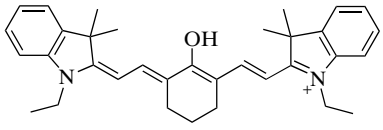

(52)

Scheme 7 Keto-enol transformation of pH-sensitive NIRD **(51)**.

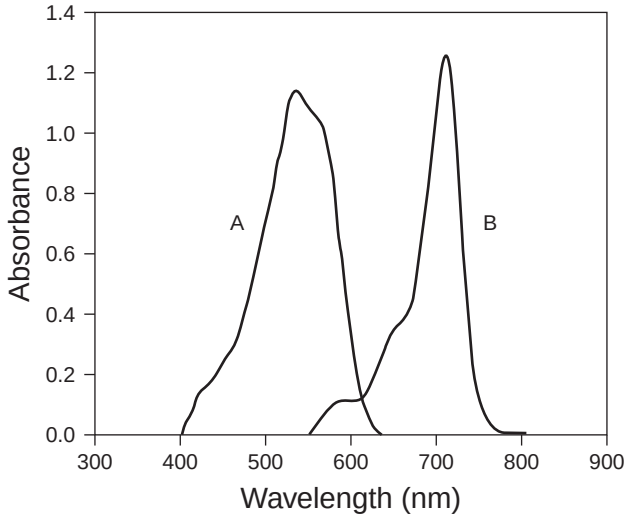

Figure 20 Absorption spectrum of pH-sensitive NIRD **(51)** in (A) acidic and (B) basic conditions.

matrix at different pHs are shown in Figure 21. The changes in the spectra in the pH range 11–12.4 were reproducible. Owing to the extremely acidic environment of the Nafion membrane, the dye provided a useful pH range, which deviated from the pK_a value of the dye alone. This study demonstrated the use of Nafion as a substrate for immobilizing NIRDs and the feasibility of developing an NIR pH sensor.

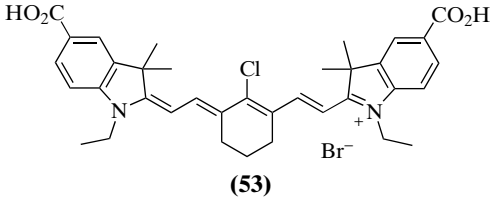

(53)

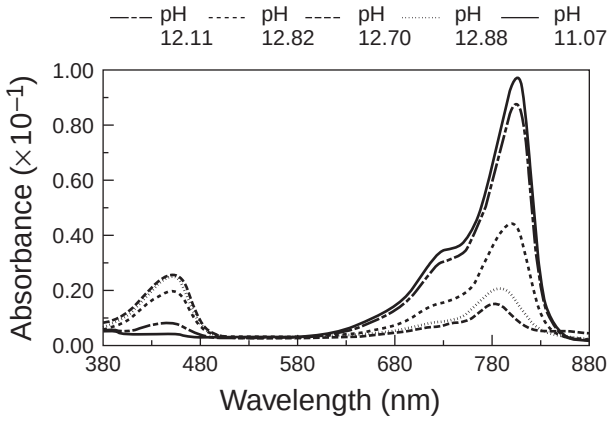

Figure 21 Absorption spectra of NIRD **(53)** entrapped in Nafion at different pH.

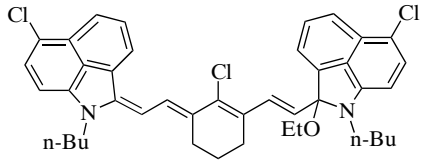

(54)

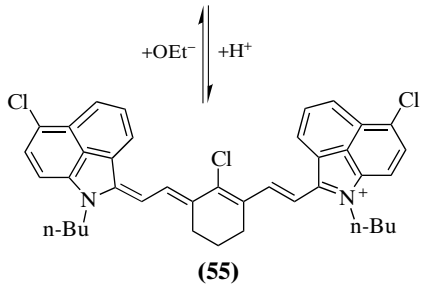

(55)

Scheme 8 Novel pH-sensitive NIR ethoxy adduct of IR-1048 **(54)**.

Mason et al.[105] recently prepared a number of C2-derivatized indolenine NIR chromophores that exhibit unique spectral properties as a function of pH. Of this series, JCM-1048-OEt **(54)** (Scheme 8) was studied by Tarazi.[93] The molecular structure of the dye contains an ethoxy moiety covalently linked to the C2-position of the 3*H*-indole. This addition results in the interference of the cationic delocalization that is typical of cyanine dyes and produces a complex absorption spectrum with a λ_{max} at 511 nm. However, addition of acid regenerates the cationic chromophore IR-1048 **(55)** to give a dramatic red shift to 1050 nm. The spectral characteristics of the dye in different pHs are shown in Figure 22. As can be seen, the dye exhibits two characteristic absorptions, one in the NIR and the other in the UV/VIS region. The dye exhibited a linear range for the NIR peak over the pH range 6–9 in acetone, pH 5–9 in acetonitrile and pH 6–12

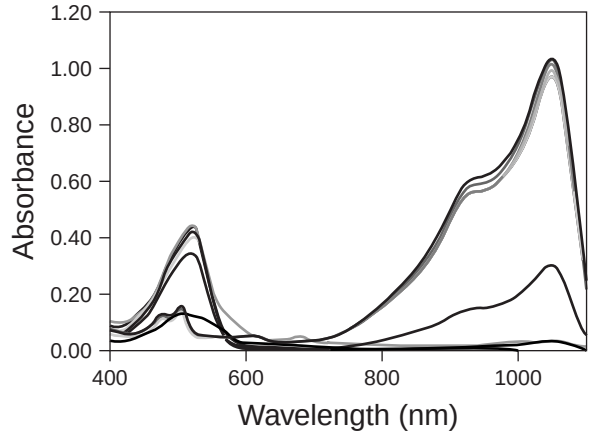

Figure 22 Absorption of JCM 1048-OEt **(54)** at different pHs.

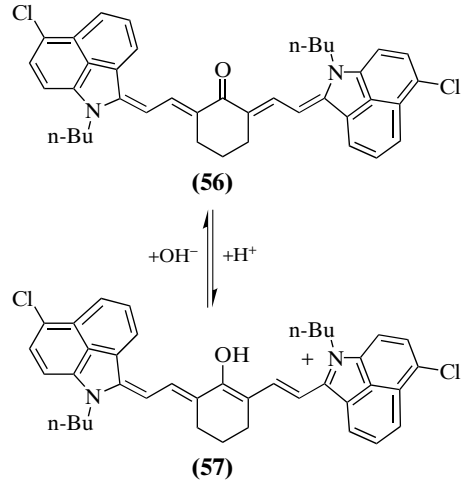

Scheme 9 Structure of pH-sensitive NIRD JCM 648/932 **(56/57)**.

in water as the solvent. In addition to the C2-derivatized NIRD, Mason et al.[106] reported the synthesis of a bis(aminodiene)one, JCM-648/932 **(56)** (Scheme 9) that exhibits pH-dependent absorptions in the NIR region. In neutral or basic conditions the equilibrium favors the keto form with an absorbance maximum of 648 nm in methanol. However, addition of acid results in the protonation of the carbonyl to restore the cyanine system. This acidic species **(57)** shows electronic absorption at 932 nm. The transformation from the keto to the enol form is fully reversible and can be carried out repeatedly.

Another application developed using NIR fluorophores is fluorescence detected circular dichroism (FDCD). This method is similar to absorption CD, but fluorescence intensity as opposed to absorbance is the measure of molecular ellipticity. Absorption CD is a valuable tool for examining the molecular chirality of proteins in the ground state. However, owing to the limited number of chiral centers, CD is often limited to qualitative measurements and global conformational changes instead of optical activity about specific centers. Tinoco and Turner reported the ability to circumvent conventional problems with CD measurements by the FDCD technique.[107,108] FDCD can be used to measure the CD of a single contributing chiral fluorophore while avoiding nonfluorescent chiral absorbers and achiral fluorophores.[108,109] For example, fluorescence from a single tryptophan residue has been measured in the midst of other chiral absorbing residues present in proteins. Structural studies of proteins with and without ligand binding have also been reported.[100–112] Similarly to fluorescence spectroscopy, FDCD makes use of a 90° angle relative to the excitation source for detecting the fluorescence. Unlike fluorescence, where plane polarized light is used for excitation, circularly polarized light is used in FDCD. When in the absence of energy transfer and only a single chiral fluorophore exists at the analytical wavelength, the detected FDCD is for a particular population of fluorophores. This situation may be inconceivable in many applications where many proteins may be present in a mixture. In order to avoid complications with matrix interferences, chromatographic techniques have been used in conjunction with FDCD.

Yeung et al.[113,114] reported the use of LIF with FDCD coupled to HPLC and CE. They reported detection limits of 170 and 0.07 pg for riboflavin by HPLC/FDCD and CE/FDCD, respectively. Thomas et al. reported the use of multidimensional FDCD for resolving mixtures without physical separation.[111,112] This method utilized the differences in the excitation and emission bands for resolving the fluorescence spectra of optically active fluorophores. Resolving mixtures of optically active fluorophores based on their fluorescence lifetimes has also been reported.[115,116] Since CD is a weighted-average measurement,[115] many chromophores present in biomolecules contribute to the CD signals. This presents a quantitative limitation in CD measurements. Although the same rules apply to FDCD, intrinsic fluorescence from optically active fluorophores limits the measured CD to a few possible molecules such as tryptophan, tyrosine, and phenylalanine in proteins. However, when chiral nonfluorescent molecules absorb in the ultraviolet (UV) region, the radiation intensity incident on the fluorophore of interest can be weakened. Consequently, the emission intensity may diminish. Considering the lower quantum yields associated with fluorescent protein residues and the number of potential interfering transitions, CD and FDCD at shorter wavelengths are limited.

Long-wavelength alternatives to UV CD measurements have been reported. Keiderling discussed the use of vibrational circular dichroism (VCD), a chiroptical

analog of IR spectroscopy.[117] This technique measures the intrinsic vibrational modes of proteins which avoids some of the limitations of conventional absorbance CD. Sophianopoulos et al. used extrinsic NIR chromophores as structural probes in near-infrared circular dichroism (NIRCD) measurements.[118] Since synthetic probes possessing various properties are readily available, photophysical limitations of intrinsic chromophores can be avoided. Although VCD and NIRCD have the advantage of avoiding biological background interferences, they lack the sensitivity gained in FDCD.

Since there are very few proteins that exhibit long wavelength fluorescence, greater sensitivity, selectivity, and spectral resolution can be gained through the use of near-infrared fluorescence-detected circular dichroism (NIRFDCD) of extrinsic fluorophores. Many biomolecules typically absorb only at lower wavelengths, and thereby complex samples can be potentially analyzed without physically separating matrix components. Typically, NIRDs are optically inactive, low molecular weight (MW $\approx 1000\,\mathrm{g\,mol^{-1}}$) symmetrical compounds that do not yield FDCD. When placed in an asymmetrically perturbed environment, as in protein binding sites, chirality may be induced, resulting in Cotton effects.[118] Therefore, these compounds must meet two criteria in order to exhibit FDCD: first, the molecule must be fluorescent in the NIR region, and second, it must possess induced chirality through noncovalent interactions with the macromolecule clearly indicating that binding has occurred. In our laboratories NIRFDCD was used in order to gain a better understanding of the structure of NIRDs that bind proteins. An understanding of dye binding to albumin may relay important information concerning these dyes as labels or probes in analytical techniques such as the separation of proteins with HPLC, gel permeation chromatography (GPC) or in gel electrophoresis. Meadows et al.[119] recently investigated the NIR squarylium dye NN525 **(44)** as a non-covalent probe for protein structural determinations. This dye, owing to its rigid squarate residue, has greater photostability relative to its cyanine counterpart. They found that it forms a stable complex with serum albumins with binding constant in excess of $10^6\,\mathrm{L\,mol^{-1}}$ compared with $10^5\,\mathrm{L\,mol^{-1}}$ for many other conventional dyes.[120,121] This in turn added to the ease of detection. Also, the high quantum yield and molar absorptivity permit the use of smaller amounts of dye relative to visible counterparts. Preliminary results showed that NIRFDCD of squarylium fluorophores can be measured using 10–100-fold lower dye concentrations in comparison with normal absorbance CD.[119] The advantage of using lower concentrations of dyes simplifies the assessment of competing equilibria associated with high concentrations such as dye or protein aggregation.

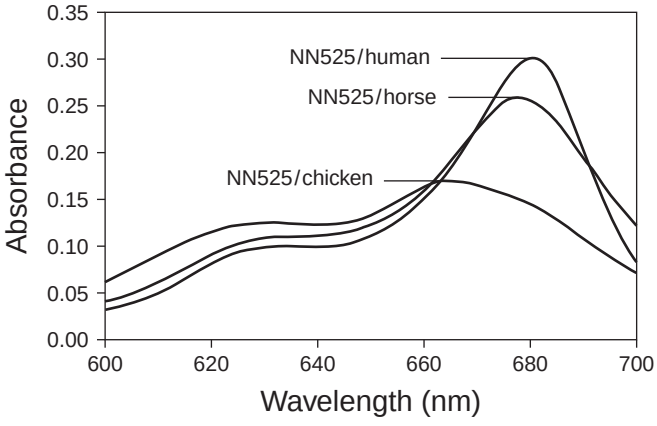

Figure 23 Absorption spectra of NN525 **(43)** bound to different proteins.

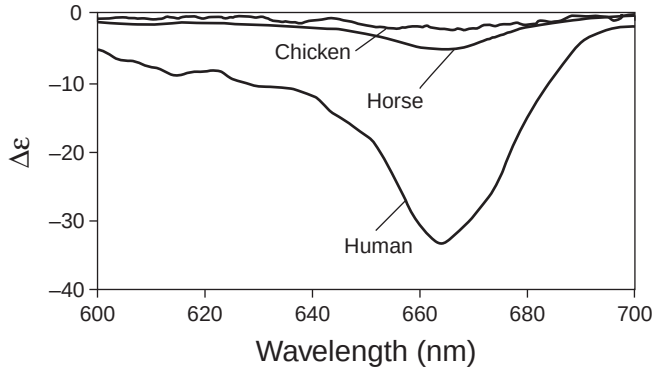

Figure 24 NIRFDCD spectra of NN525 **(43)** bound to different proteins.

Dye aggregation may in some instances result in the absorption of circularly polarized light.[122]

A common finding for many of the NIRDs previously studied is the enhancement of photophysical properties upon binding to hydrophobic environments (Figure 23).[118,123–125] An increase in these properties permits the use of even lower probe concentrations. This can be advantageous since lower ratios of bound fluorophores can leave proteins virtually unperturbed structurally.

Previous reports by Sophianopoulos et al.[118] support the binding of NIR squarylium compounds to similar, if not identical, binding sites on bovine serum albumin (BSA) with immeasurable conformational changes. Many of the symmetrical NIR compounds do not have measurable optical activity in the protein-unbound state. When these dyes bind in asymmetric binding sites of proteins, induced Cotton effects can occur. As can be seen in Figure 24, the relative FDCD intensities found for the three different albumin–dye complexes is in accordance with the absorbance spectra. This indicates

that those transitions occur through asymmetric perturbations in the fluorophore. Structural differences in the binding sites occupied by NN525 cause intensities to differ. Although not confirmed, previous data supported a common binding site near positively charged amino acid residues.[118] Homology in this site among different species may account for the common negative Cotton effects observed for different albumins. Differences in the hydrophobicities and/or electrostatic interactions at these sites may explain the intensity variations exhibited for the different species. Changes in the extent of the association or weak induced chirality can result from weaker noncovalent interactions. The squarylium dyes had large molar ellipticities (10^5 ° cm^2 $dmol^{-1}$) when bound to BSA. Although induced CD may occur with binding to macromolecules, optical activity due to intramolecular forces, such as dye aggregation[123,126] in aqueous solutions, is less likely at lower concentrations used in FDCD. In many cases these complexes quench fluorescence, but absorb circularly polarized light. Therefore, the CD intensities may falsely indicate asymmetric interactions in the protein-bound dye. The lower concentrations (10^{-7} M) needed in these NIRFDCD studies help prevent solubility limitations that can cause the formation of aggregates, avoiding extraneous Cotton effects. Hence the probability of FDCD data containing artifacts from aggregates is small relative to CD. The use of extrinsic fluorophores in NIRFDCD can provide an abundance of complementary information about conformational changes near the fluorophore. The main mode of binding of NIR squarylium dyes to BSA was determined to be hydrophobic, although electrostatic forces may also contribute. Van der Waals contacts near the dye binding site may enhance their association to proteins. The use of CD in conjunction with tryptophan FDCD in proteins and NIRFDCD of noncovalently bound dyes can provide a more detailed description of protein structure and interactions of proteins with ligands. A further advantage of NIRFDCD, the need for lower concentrations, prevents the occurrence of probe and protein intramolecular interactions, which simplifies data interpretation.

5 CONCLUSIONS

The discussions in this article demonstrate the various advantages of working with the novel emerging technology of NIR fluorescence. Reduction in background interference offered by this method is the greatest advantage. No other analytical method known thus far can compete with the highest possible sensitivity achievable, i.e. single-molecule detection, offered by LIF. NIR detection provides advantages in complex biological and environmental systems where background interference is often a major concern. The availability of commercial NIRDs and NIR instrumentation, although limited, has provided a tremendous advance in the development of applications in this region. The commercial dyes or instrumentation can be tailored to specific applications and aid in interfaces with other commercial instrumentation readily available.

The photostability of NIRDs is a cause for concern. However, recent studies have focused on the development of functionalized photostable NIRDs to overcome this problem. The combination of diode laser excitation sources with photodiodes for detection and coupling of these to suitable fiber-optic interfaces aid in miniaturization and allow for development of remote sensing applications. Time-resolved measurements for discrimination between similar species have made the method very useful both in biological applications such as DNA sequencing, cytology, immunoasssays, and intracellular measurements and in environmental applications in determining metal ions, pH and microenvironment determination. Other areas that utilize the advantages include photodynamic therapy, thermal transfer printing, optical recording, transparent bar codes, and forgery prevention. Progress continues to be made with more applications being developed using the emerging novel NIR fluorescence technology.

ABBREVIATIONS AND ACRONYMS

Ab	Antibody
ADC	Analog-to-digital Convertor
Ag	Antigen
AIDS	Acquired Immunodeficiency Syndrome
Ang-I	Angiotensin-I
APD	Avalanche Photodiode
BSA	Bovine Serum Albumin
CBQCA	3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde
CD	Circular Dichroism
CE	Capillary Electrophoresis
CGE	Capillary Gel Electrophoresis
CZE	Capillary Zone Electrophoresis
DRAW	Direct Read After Writing
ELISA	Enzyme-linked Immunosorbent Assay
FDCD	Fluorescence Detected Circular Dichroism
FFOI	Fiber Optic Immunosensor
FIA	Fluorescence Immunoassay
FITC	Fluorescein Isothiocyanate
GaAlAs	Gallium–Aluminum–Arsenide
GAHG	Goat Antihuman Immunoglobulin G

GC	Gas Chromatography
GPC	Gel Permeation Chromatography
HDITCP	1,1′,3,3,3′,3′-Hexamethyl-4,4′,5,5′-dibenzo-2,2′-indotricarbocyanine Perchlorate
HIDC	1,1′,3,3,3′,3′-Hexamethyl Dicarbocyanine
HOMO	Highest Occupied Molecular Orbital
HPLC	High-performance Liquid Chromatography
HuSCC	Human Squamous Cell Carcinoma Line
IA	Immunoassay
ICG	Indocyanine Green
IEF	Isoelectric Focusing
IgG	Immunoglobulin G
IR	Infrared
ISC	Intersystem Crossing
ITP	Isotachophoresis
LIF	Laser-induced Fluorescence
LPS1	*Legionella Pneumophilia* Sera Type 1
LUMO	Lowest Unoccupied Molecular Orbital
MEKC	Micellar Electrokinetic Chromatography
MPc	Metal Ion Phthalocyanine
MS	Mass Spectrometry
NATO	North Atlantic Treaty Organization
NHIgG	Normal Human Immunoglobulin G
NIR	Near-infrared
NIRCD	Near-infrared Circular Dichroism
NIRD	Near-infrared Dye
NIRFDCD	Near-infrared Fluorescence-detected Circular Dichroism
NIRFIA	Near-infrared Fluorescence Immunoassay
NIRLIF	Near-infrared Laser-induced Fluorescence
NPc	Naphthalocyanine
Pc	Phthalocyanine
PCR	Polymerase Chain Reaction
PMMA	Poly(methyl methacrylate)
PMO	Perturbational Molecular Orbital
PMT	Photomultiplier Tube
PPPMO	Pariser–Parr–Pople Molecular Orbital
PS	Polystyrene
RIA	Radioimmunoasay
SNR	Signal-to-noise Ratio
TLC	Thin-layer Chromatography
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VCD	Vibrational Circular Dichroism
WHO	World Health Organization
WORM	Write Once Read Many

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Circular Dichroism in Analysis of Biomolecules • Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Fluorescence Spectroscopy in Peptide and Protein Analysis

Electronic Absorption and Luminescence **(Volume 12)**
Absorption and Luminescence Probes • Fluorescence in Organized Assemblies • Fluorescence Lifetime Measurements, Applications of • Indirect Detection Methods in Capillary Electrophoresis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis

REFERENCES

1. W. Herschel, *Philos. Trans. R. Soc. London*, **90**, 225 (1800).
2. A.R. Swamy, 'Development and Optimization of NIR Fluorescence Immunoassays for Immunodiagnostic Applications', PhD Dissertation, Georgia State University, Atlanta, GA, 1999.
3. J. Griffiths, 'Approaches to the Design and Synthesis of Near-infrared Absorbing Dyes', in *Modern Colorants: Synthesis and Structure*, eds. A.T. Peters, H.S. Freeman, Blackie Academic and Professional, Glasgow, 41–62, 1995.
4. J. Fabian, H. Nakazumi, M. Matsuoka, 'Near-infrared Absorbing Dyes', *Chem. Rev.*, **92**, 1197–1226 (1992).
5. E.D. Sternberg, D. Dolphin, C. Brückner, 'Porphyrin-based Photosensitizers for Use in Photodynamic Therapy', *Tetrahedron*, **54**, 4151–4202 (1998).

6. M. Matsuoka, 'Synthetic Design of Infrared Absorbing Dyes', in *Infrared Absorbing Dyes*, eds. A.R. Katritzky, G.J. Sabongi, Plenum Press, New York, 1–17, 1990.
7. R. Zahradnik, 'Electron Spectra of Conjugated Compounds in the Ultraviolet and Visible Regions', *Fortschr. Chem. Forsch.*, **10**, 1–58 (1968).
8. M.J.S. Dewar, 'Color and Constitution', *J. Chem. Soc.*, 2329–2334 (1950).
9. S. Yasui, M. Matsuoka, T. Kitao, 'Color and Constitution of Trimethine Cyanine Dyes', *Shikizai Kyokaishi*, **61**, 375–380 (1988).
10. A.I. Tolmachev, L.Y. Slominskii, A.A. Ishchenko, 'New Cyanine Dyes Absorbing in the NIR Region', in *Near Infrared Dyes for High Technology Applications*, eds. S. Daehne, U. Resch-Genger, O.S. Wolfbeis, NATO ASI Series, Kluwer, Dordrecht, 385–415, 1998.
11. R.B. Mujumdar, L.A. Ernst, S.R. Mujumdar, A.S. Waggoner, 'Cyanine Dye Labeling Reagents Containing Isothiocyanate Groups', *Cytometry*, **10**, 11–19 (1989).
12. L. Strekowski, M. Lipowska, G. Patonay, 'Substitution Reactions of Nucleofungal Group in Heptamethine Cyanine Dyes. Synthesis of a Isothiocyanato Derivative for Labeling of Proteins with a Near-infrared Chromophore', *J. Org. Chem.*, **57**, 4578–4580 (1992).
13. M. Lipowska, G. Patonay, L. Strekowski, 'New Near-infrared Cyanine Dyes for Labeling of Proteins', *Synth. Commun.*, **23**, 3087–3094 (1993).
14. J.H. Flanagan, A.H. Khan, S. Menchen, S.A. Soper, R.P. Hammer, 'Functionalized Tricarbocyanine Dyes as Near-infrared Fluorescent Probes for Biomolecules', *Bioconj. Chem.*, **8**, 751–756 (1997).
15. J.H. Flanagan, S.E. Romero, B.L. Legendre, R.P. Hammer, S.A. Soper, 'Heavy-atom Modified Near-IR Fluorescent Dyes for DNA Sequencing Applications: Synthesis and Photophysical Characterization', *Proc. SPIE*, **2980**, 328–337 (1997).
16. R.B. Mujumdar, L.A. Ernst, S.R. Mujumdar, C.J. Lewis, A.S. Waggoner, 'Cyanine Dye Labeling Reagents: Sulfoindocyanine Succinimidyl Esters', *Bioconj. Chem.*, **4**, 105–111 (1993).
17. N. Narayanan, G. Little, R. Raghavachari, J. Gibson, A. Lugade, C. Prescott, K. Reiman, S. Roemer, D. Steffens, S. Sutter, D. Draney, 'New NIR Dyes: Synthesis, Spectral Properties and Applications in DNA Analysis', in *Near Infrared Dyes for High Technology Applications*, eds. S. Daehne, U. Resch-Genger, O.S. Wolfbeis, NATO ASI Series, Kluwer, Dordrecht, 141–159, 1998.
18. N. Narayanan, G. Patonay, 'A New Method for the Synthesis of Heptamethine Cyanine Dyes: Synthesis of New Near-infrared Labels', *J. Org. Chem.*, **60**, 2391–2395 (1995).
19. E. Terpetschnig, H. Szmacinski, A. Ozinskas, J.R. Lakowicz, 'Synthesis of Squarine-*N*-hydroxysuccinimide Esters and Their Biological Application as Long-wavelength Fluorescent Probes', *Anal. Biochem.*, **217**, 197–204 (1994).
20. E. Terpetschnig, J.R. Lakowicz, 'Synthesis and Characterization of Unsymmetrical Squarines: A New Class of Cyanine Dyes', *Dyes Pigments*, **21**, 227–234 (1993).
21. N.B. Mckeown, 'Phthalocyanine Synthesis', in *Phthalocyanine Materials: Synthesis, Structure and Function*, eds. B. Dunn, J.W. Goody, A.R. West, Cambridge University Press, Cambridge, 3–31, 1998.
22. G. Casay, T. Czuppon, N. Narayanan, J. Lipowski, G. Patonay, 'Near-infrared Fluorophores and Semiconductor Lasers in Analytical Applications', *Process Control Qual.*, **5**, 35–44 (1993).
23. I.M. Warner, S.A. Soper, L.B. McGown, 'Molecular Fluorescence, Phosphoresence, and Chemiluminescence Spectrometry', *Anal. Chem.*, **68**, 73R–91R (1996).
24. S. Daehne, U. Resch-Genger, O.S. Wolfbeis (eds.), *Near Infrared Dyes for High Technology Applications*, NATO ASI Series, Kluwer, Dordrecht, 1998.
25. K. Kamisaka, I. Listowski, J.J. Betheil, I.M. Arias, 'Competitive Binding of Bilirubin, Sulfobromophthalein, Indocyanine Green, and Other Organic Anions to Human and Bovine Serum Albumin', *Biochim. Biophys. Acta*, **365**, 169–180 (1974).
26. K. Sauda, T. Imasaka, N. Ishibashi, 'Determination of Protein in Human Serum by High-performance Liquid Chromatography with Semiconductor Laser Fluorometric Detection', *Anal. Chem.*, **58**, 2649–2653 (1986).
27. R.J. Williams, M. Lipowska, L. Strekowski, G. Patonay, 'Comparison of Covalent and Non-covalent Labeling with NIR Dyes for High Performance Liquid Chromatographic Determination of Human Serum Albumin', *Anal. Chem.*, **65**, 601–605 (1993).
28. M. Baars, 'Near-infrared Laser Induced Fluorescence as an Ultra-sensitive Method of Detection for Capillary Electrophoresis', PhD Dissertation, Georgia State University, Atlanta, GA, 1998.
29. D.A. Skoog, J.J. Leary, *Principles of Instrumental Analysis*, Harcourt Brace College Publishers, Philadelphia, PA, 1992.
30. *Introduction to Capillary Electrophoresis*, Primer 360643, Beckman Instruments, Fullerton, CA, Vol. I, 1991.
31. J.P. Landers, *Handbook of Capillary Electrophoresis*, CRC Press, Boca Raton, FL, 1997.
32. M.L. Susan, M.R. Donna, *Pharmaceutical and Biomedical Applications of Capillary Electrophoresis*, Elsevier Science, Amsterdam, 1996.
33. D. Chen, N. Dovichi, 'Yoctomole Detection Limit by Laser-induced Fluorescence in Capillary Electrophoresis', *J. Chromatogr.*, **657**, 265–269 (1994).
34. S.J. Williams, E.T. Bergstrom, D.M. Goodall, H. Kawazumi, K.P. Evans, 'Diode Laser Based Indirect Absorbance Detector for Capillary Electrophoresis', *J. Chromatogr.*, **636**, 39–45 (1993).

35. J.M. Song, H. Kawazumi, T. Inoue, T. Ogawa, 'Time Resolved Photon Counting Fluorimetry with a Semiconductor Laser for Capillary Electrophoresis', *J. Chromatogr., A*, **727**, 330–333 (1996).
36. T. Fuchigama, T. Imasaka, M. Shiga, 'Subattomole Detection of Amino Acids by Capillary Electrophoresis Based on Semiconductor Laser Fluorescence Detection', *Anal. Chim. Acta*, **282**, 209–213 (1993).
37. H. Kawazumi, J.M. Song, T. Inoue, T. Ogawa, 'Application of Avalanche Photodiode in a Near-Geiger Operation as Fluorescence Detector for Capillary Electrophoresis', *J. Chromatogr., A*, **744**, 31–36 (1996).
38. S.A. Soper, Q.L. Mattingly, P. Vengunta, 'Photon Burst Detection of Single NIR Fluorescent Molecules', *Anal. Chem.*, **65**, 740–747 (1993).
39. J.H. Flanaghan, Jr, B.L. Legendre, Jr, R.P. Hammer, S.A. Soper, 'Binary Solvent Effects in CZE with NIR Fluorescence Detection of Related Tricarbocyanine Dyes and Labeled Amino Acids', *Anal. Chem.*, **67**, 341–347 (1995).
40. D.C. Williams, S.A. Soper, 'Ultra Sensitive NIR Fluorescence Detection for CGE and DNA Sequencing Application', *Anal. Chem.*, **67**, 3427–3432 (1995).
41. L. Middendorf, J. Amen, R. Bruce, D. Draney, D. DeGraff, J. Gewecke, D. Grone, P. Humphrey, N. Narayanan, R. Raghavachari, A. Lugade, A. Oomen, H. Osterman, R. Peterson, J. Rada, G. Little, S. Roemer, 'Near-infrared Fluorescence Instrumentation for DNA Analysis', in *Near Infrared Dyes for High Technology Applications*, eds. S. Daehne, U. Resch-Genger, O.S. Wolfbeis, NATO ASI Series, Kluwer, Dordrecht, 141–159, 1998.
42. S. Lehotay, P. Johnson, T. Barber, J. Winefordner, 'Ultralow Detection Limit of a Near-infrared Dye by Diode-laser-induced Fluorescence in a Flowing Stream', *Appl. Spectrosc.*, **44**, 1577–1579 (1990).
43. S.A. Soper, E. Brooks, J. Martin, J. Jett, J. Hahn, H. Nutter, R. Keller, 'Single-molecule Detection of Rhodamine 6G in Ethanolic Solutions Using Continuous Wave Laser Excitation', *Anal. Chem.*, **63**, 432–437 (1991).
44. M.J. Baars, G. Patonay, 'Ultra Sensitive Detection of Closely Related Peptides using CE with NIR LIF Detection', *Anal. Chem.*, **71**, 667–671 (1998).
45. H. Schwartz, A. Guttman, *Separation of DNA by Capillary Electrophoresis*, Primer 607397, Beckman Instruments, Fullerton, CA, 1995.
46. H. Swerdlow, J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S. Wu, M.J. Dovichi, C. Fuller, 'Three DNA Sequencing Methods Using Capillary Gel Electrophoresis and Laser-induced Fluoresence', *Anal. Chem.*, **63**(24), 2835–2841 (1991).
47. K. Ueno, E.S. Yeung, 'Simultaneous Monitoring of DNA Fragments Separated by Electrophoresis in a Multiplexed Array of 100 Capillaries', *Anal. Chem.*, **66**(9), 1424–1431 (1994).
48. M. Vilenchik, A. Belenky, A. Cohen, 'Monitoring and Analysis of Antisense DNA by High-performance Capillary Gel Electrophoresis', *J. Chromatogr.*, **663**, 105–113 (1994).
49. J. Xian, M. Harrington, E. Davidson, 'DNA–Protein Binding Assays from a Single Sea Urchin Egg: A High-sensitivity Capillary Electrophoresis Method', *Proc. Natl. Acad. Sci. USA*, **93**, 86–90 (1996).
50. S.A. Berson, R.S. Yalow, 'Quantitative Aspects of the Reaction Between Insulin and Insulin-binding Antibody', *J. Clin. Invest.*, **38**, 1996–2016 (1959).
51. C.D. Ercegovich, *Analysis of Pesticide Residues*, American Chemical Society, Washington, DC, 162–177, 1971.
52. J.L. Langone, H. Van Vunakis, 'Radioimmunoassay for Dieldrin and Aldrin', *Res. Commun. Pathol. Pharmacol.*, **10**, 163–171 (1975).
53. E. Engvall, P. Perlman, 'Enzyme-linked Immunosorbent Assay (ELISA). Quantitative Assay of Immunoglobulin G', *Immunochemistry*, **8**, 871–874 (1971).
54. V.C.W. Tsang, K. Hancock, M. Wilson, 'Enzyme-linked Immunotransfer Blot Technique for Human T-lymphotropic Virus Type III/Lymphoadenopathy-associated Virus (HTLV-III/LAV) Antibodies', Monograph: Immunology Series No. 15. Procedural Guide, Centers for Disease Control, Atlanta, GA, 1986.
55. V.C.W. Tsang, J.A. Brand, A.E. Boyer, 'An Enzyme-linked Immunoelectrotransfer Blot Assay and Glycoprotein Antigens for Diagnosing Human Cysticercosis (*Taenia solium*)', *J. Infect. Dis.*, **159**, 50–59 (1989).
56. R.S. Jotte, J. Scott, 'Malaria: Review of Features Pertinent to the Emergency Physician', *J. Emerg. Med.*, **11**, 729–736 (1993).
57. V.C.W. Tsang, K.R. Tsang, K. Hancock, M.A. Kelly, B.C. Wilson, S.E. Maddison, '*Schistosoma mansoni* Adult Microsomal Antigens, a Serologic Reagent. I. Systematic Fractionation, Quantitation, and Characterization of Antigenic Components', *J. Immunol.*, **130**, 1359–1365 (1983).
58. V.C.W. Tsang, K. Hancock, S.E. Maddison, 'Quantitative Capacities of Glutaraldehyde- and Sodium Metaperiodate-coupled Peroxidase–Anti-human IgG Conjugates in Enzyme-linked Immunoassays', *J. Immunol. Methods*, **70**, 91–100 (1984).
59. I.A. Hemmila, *Applications of Fluorescence in Immunoassays*, John Wiley & Sons, New York, 1991.
60. T. Imasaka, A. Yoshitake, N. Ishibashi, 'Semiconductor Laser Fluorimetry in NIR Region', *Anal. Chem.*, **56**, 1077–1079 (1984).
61. T. Imasaka, N. Ishibashi, 'Diode Lasers and Practical Trace Analysis', *Anal. Chem.*, **62**, 363A–371A (1990).
62. T. Imasaka, H. Nagakawa, N. Okazaki, N. Ishibashi, 'Enzyme Immunoassay of Insulin by Semiconductor Laser Fluorimetry', *Anal. Chem.*, **62**, 2404–2405 (1990).
63. G.A. Casay, D.B. Shealy, G. Patonay, 'Near-infrared Fluorescent Probes', *Top. Fluoresc. Spectrosc.*, **4**, 183–222 (1994).

64. A.E. Boyer, M. Lipowska, J.M. Zen, G. Patonay, V.C.W. Tsang, 'Evaluation of NIR Dyes as Labels for Immunoassay Utilizing Laser Diode Detection', *Anal. Lett.*, **25**(3), 415–428 (1992).
65. R.J. Williams, 'Development of a Solid Phase Immunoassay for Human Immunoglobulin G Detection: A Novel Bio-analytical Application of Semiconductor Laser Induced NIR Fluorescence', PhD Dissertation, Georgia State University, Atlanta, GA, 1996.
66. G.A. Robinson, V.M. Cole, S.J. Rattla, G.C. Forrest, 'Bioelectrochemical Immunoassay for Human Chorionic Gonadotropin in Serum Using an Electrode-immobilized Capture Antibody', *Biosensors*, **2**, 45–57 (1986).
67. M.I. Danesvar, 'Design and Development of a Fiber Optic Immunosensor Utilizing NIR Fluorophores', PhD Dissertation, Georgia State University, Atlanta, GA, 1996.
68. L. Evans, III, 'Bio-analytical Applications of Polymethine Dyes Covalently Bound to Immunoglobulin G', PhD Dissertation, Georgia State University, Atlanta, GA, 1998.
69. S.V. Rahavendran, H.T. Karnes, 'Solid State Diode Laser Induced Fluorescence Detection in High Performance Liquid Chromatography', *Pharm. Res.*, **10**(3), 328–334 (1993).
70. K. Sauda, T. Imasaka, N. Ishibashi, 'High-performance Liquid Chromatographic Detector Based on Near-infrared Semiconductor Laser Fluorometry', *Anal. Chim. Acta*, **187**, 353–356 (1986).
71. F.A. Chen, A, Tusak, S. Pentoney, K. Konrad, C. Lew, J. Sternberg, 'Semiconductor Laser Induced Fluorescence Detection in Capillary Electrophoresis using a Cyanine Dye', *J. Chromatogr.*, **652**, 355–360 (1993).
72. P.A. Johnson, T.E. Barber, B.W. Smith, J.D. Winefordner, 'Ultra-low Detection Limits for an Organic Dye Determined by Fluorescence Spectroscopy with Diode Laser Excitation', *Anal. Chem.*, **61**, 861–863 (1989).
73. M. Gui, S.N. Sadayappan, V. Rahavendran, H.T. Karnes, 'Diode Laser Induced NIR Fluorescence Detection for High Performance Liquid Chromatography Analysis of Indocyanine Green in Plasma', *Anal. Chim. Acta*, **342**, 145–149 (1997).
74. P. Kuklenyik, 'Simple HPLC Detector for the Detection of NIR Dyes', *Instrum. Sci. Technol.*, **23**(2), 113–122 (1995).
75. W. Rettig, R. Lapouyade, 'Fluorescent Probes Based Upon Twisted Intermolecular Charge Transfer TITC States and Other Adiabatic Photoreactions', *Top. Fluoresc. Spectrosc.*, **4**, 109–149 (1994).
76. H.G. Lohr, F. Vogtle, 'Chromophores and Fluoroionophores. A New Class of Dye Reagent', *Acc. Chem. Res.*, **18**, 65–72 (1985).
77. A.W. Czarnik, 'Chemical Communication in Water Using Fluorescent Chemosensors', *Acc. Chem. Res.*, **97**, 1515–1566 (1997).
78. L.A. Saari, W.R. Seitz, 'Immobilized Morin as Fluorescence Sensor for Determination of Al(III)', *Anal. Chem.*, **55**, 667–670 (1983).
79. Z. Zhujun, W.R. Seitz, 'A Fluorescent Sensor for Aluminum(III), Magnesium(II), Zinc(II) and Cadmium(II) Based on Electrostatically Immobilized Quinolin-8-ol Sulfonate', *Anal. Chim. Acta*, **171**, 251–258 (1985).
80. Y. Kawabata, R. Tahara, T. Kamishika, T. Imasaka, N. Ishibashi, 'Ion Selective Optrode Using Hexadecyl–Acridine Orange Attached on Poly(vinyl chloride) Membrane', *Anal. Chem.*, **62**, 2054–2055 (1990).
81. K.S. Litwiter, P.M. Kluczynski, F.V. Bright, 'Determination of the Transduction Mechanism for Optical Sensors Based on Rhodamine 6G Impregnated Persulfonate Films Using Steady State Frequency Domain Fluorescence', *Anal. Chem.*, **63**, 797–802 (1991).
82. M.J. Quintele, M. Gallego, M. Valcárcel, 'Flow Injection Spectrophotometric Method for the Speciation of Aluminum in River and Tap Waters', *Analyst*, **118**, 1199–1203 (1993).
83. F. Sanchez Rojas, E. Cristofol Alcaraz, J.M. Cano Pavon, 'Determination of Aluminum in Water by Flow Injection with Fluorimetric Detection by Using Salicylaldehyde Carbohydrazone as Reagent in a Micellar Medium', *Analyst*, **119**, 1221–1223 (1994).
84. T. Imasaka, T. Kamikubo, Y. Kawabata, N. Ishibashi, 'Ultratrace Photometric Determination of Phosphate with a Solid State Emitter as a Light Source', *Anal. Chim. Acta*, **153**, 261–263 (1983).
85. J. Hicks, G. Patonay, 'A Simple Driver for Dynamic or Static Operation of a Laser Diode Intracavity Spectrophotometer', *Anal. Instrum.*, **8**, 213–226 (1989).
86. E. Unger, G. Patonay, 'NIR Laser Diode Intracavity Absorption Spectrometry', *Anal. Chem.*, **61**, 1425–1427 (1989).
87. J. Hicks, G. Patonay, 'Linearity Considerations for NIR Laser Diode Intracavity Absorption Spectrometer', *Anal. Chem.*, **62**, 1543–1545 (1990).
88. J.M. Barrero, M.C. Moreno-Bondi, M.C. Peréz-Conda, C. Cámara, 'A Biosensor for Ferric Ion', *Talanta*, **40**, 1619–1623 (1993).
89. R.B. Thompson, Z. Ge, M. Patchan, C.-C. Huang, C. Fierka, 'Fiber Optic Biosensor for Co(II) and Cu(II) Based on Fluorescence Energy Transfer with an Enzyme Transducer', *Biosens. Bioelectron.*, **11**, 557–564 (1996).
90. F. Fages, B. Bodenant, T. Weil, 'Fluorescent Siderophore-based Chelators. Design and Synthesis of a Trispyrenyl Trishydroxamate Ligand, an Intramolecular Excimer-forming Sensing Molecule Which Responds to Iron(III) and Gallium(III) Metal Cations', *J. Org. Chem.*, **61**, 3956–3961 (1996).
91. G.A. Casay, T. Czuppon, L. Evans, III, G. Patonay, 'Detection of Toxic Metal Ions with Near-infrared Compounds', *SPIE Proc.*, **2293**, 42–53 (1994).

92. G.A. Casay, T. Czuppon, C. Rucker, G. Patonay, 'Optical Fiber Metal Probes Using Semiconductor Laser Diodes and Near-infrared Compounds', *Instrum. Technol.*, **22**, 157–177 (1994).
93. L. Tarazi, 'Analytical Application of Near-infrared Molecular Probes', PhD Dissertation, Georgia State University, Atlanta, GA, 1998.
94. E.U. Akkaya, S. Turkyilmaz, 'A Squarine-based NIR Fluorescent Chemosensor for Calcium', *Tetrahedron Lett.*, **l38**(25), 4513–4516 (1997).
95. R. Chin, W. Hallidy, A. Cruce, N. Calazar, K. Jamison, G. Patonay, L. Strekowski, M. Maklad, 'NIR–PRFS Fiber Optic Corrosion Sensor', *Proc. SPIE*, **2682**, 275–286 (1996).
96. R. Rogowski, M. Hoblen, J. Heyman, 'A Method for Monitoring Strain in Large Structures: Optical and Radio Frequency Devices', *Rev. Progr. Quant. Nondestruct. Eval.*, **7A**, 559 (1988).
97. P.D. Dean, R.O. Claus, D.A. Martin, D. Trites, 'Fiber Optic Sensor Systems for Smart Aerospace Structures', *Proc. SPIE*, **996**, 55 (1988).
98. L.A. Saari, W.R. Seitz, 'pH Sensor Based on Immobilized Fluoresceinamine', *Anal. Chem.*, **54**, 821–823 (1982).
99. Y. Kawabata, T. Kamichika, T. Imasaka, N. Ishibashi, 'Fiber-optic Sensor for CO_2 with pH Indicator Dispersed in Poly(ethylene glycol) Membrane', *Anal. Chim. Acta*, **219**, 119–223 (1989).
100. L. Brown, P.J. Halling, G.A. Johnson, C.J. Suckling, 'Water Soluble Indicators for Measurement of pH in Water Immiscible Solvents', *Tetrahedron Lett.*, **31**, 5799–5802 (1990).
101. M.A. Arnold, 'Fiber Optic Chemical Sensors', *Anal. Chem.*, **64**, 1015A–1025A (1992).
102. J.M. Zen, G. Patonay, 'Near-infrared Fluorescence Probe for pH Determination', *Talanta*, **63**, 2934–2938 (1991).
103. G. Patonay, G.A. Casay, M. Lipowska, L. Strekowski, 'Spectroscopic Studies of a Near-infrared Absorbing pH Sensitive Aminodienone–Carbocyanine Dye System', *Talanta*, **40**, 935–942 (1993).
104. L. Strekowski, M. Lipowska, G. Patonay, 'A New pH Sensitive NIR Chromophore', *Synth. Commun.*, **22**(17), 2593–2598 (1992).
105. J.C. Mason, J.E. Britton, H. Lee, K. Van Aken, G. Patonay, L. Strekowski, 'The Addition Reaction of Hydroxide or Ethoxide Ion with Benzindolium Heptamethine Cyanine Dyes', *Dyes and Pigments*, in press.
106. J.C. Mason, G. Patonay, L. Strekowski, 'A New pH Sensitive Chromophore', *Heterocycl. Commun.*, **3**(5), 409–411 (1997).
107. D.H. Turner, I. Tinoco, 'Fluorescence Detected Circular Dichroism', *J. Am. Chem. Soc.*, **96**, 4340–4342 (1974).
108. I. Tinoco, D.H. Turner, 'Fluorescence Detected Circular Dichroism. Theory', *J. Am. Chem. Soc.*, **98**, 6453–6456 (1976).
109. I. Tinoco, B. Ehrenberg, I.Z. Steinberg, 'Fluorescence Detected Circular Dichroism and Circular Polarization of Luminescence in Rigid Media: Direction Dependent Optical Activity Obtained by Photo Selection', *J. Chem. Phys.*, **66**, 916–920 (1977).
110. M.P. Thomas, G. Patonay, I.M. Warner, 'Fluorescence-detected Circular Dichroism in Biochemical Analysis', *Pract. Spectrosc.*, **12**, 421–452 (1991).
111. M.P. Thomas, G. Patonay, I.M. Warner, 'Fluorescence-detected Circular Dichroism Studies of Serum Albumins', *Anal. Biochem.*, **164**, 466–473 (1987).
112. M.P. Thomas, G. Patonay, I.M. Warner, 'Selective Monitoring of Stern–Volmer Quenching of Chiral Molecules via Fluorescence Detected Circular Dichroism', *Anal. Lett.*, **20**, 717–730 (1987).
113. R.E. Synovec, E.S. Yeung, 'Fluorescence Detected Circular Dichroism as Detection Principle in High Performance Liquid Chromatography', *J. Chromatogr.*, **368**, 85–93 (1986).
114. P.L. Christensen, E.S. Yeung, 'Fluorescence Detected Circular Dichroism for On-column Detection in Capillary Electrophoresis', *Anal. Chem.*, **61**, 1344–1347 (1989).
115. K. Wu, L. Geng, M.J. Joseph, L.B. McGown, 'Steady State and Lifetime Resolved Fluorescence Detected Circular Dichroism Using a Modified Phase Modulation Spectrofluorimeter', *Anal. Chem.*, **65**, 2339–2345 (1993).
116. L. Geng, L.B. McGown, 'Chirality Distributions Through Lifetime Resolution in Fluorescence Detected Circular Dichroism and Circularly Polarized Luminescence: A Theoretical Treatment', *Anal. Chem.*, **64**, 68–74 (1992).
117. T.A. Keiderling, 'Vibrational Circular Dichroism of Peptides and Proteins', eds. K. Nakanishi, N. Berova, R.W. Woody, VCH, New York, 497–521, 1994.
118. A.J. Sophianopoulos, J. Lipowski, N. Narayanan, G. Patonay, 'Association of NIR Dyes with Bovine Serum Albumin', *Appl. Spectrosc.*, **51**, 1511–1515 (1997).
119. F. Meadows, 'Investigation of Noncovalent Interaction of NIR Fluorophores for Bioanalytical Applications', PhD Dissertation, Georgia State University, Atlanta, GA, 1999.
120. F.L. Rodkey, 'Binding of Phenol Red by Human Serum Albumin', *Arch. Biochem. Biophys.*, **94**, 526–531 (1961).
121. F.L. Rodkey, 'Binding of Bromocresol Green by Human Serum Albumin', *Arch. Biochem. Biophys.*, **108**, 510–513 (1964).
122. H. Chen, K.Y.H. Law, J. Perlstein, D.G. Whitten, 'Amphiphilic Squaraine Dye Aggregates: Evidence for a Cyclic Chiral Structure as a General Supramolecular Structure for Aggregates of Dyes and Aromatic Molecules', *J. Am. Chem. Soc.*, **117**, 7257–7258 (1995).
123. Y.Y. Davidson, B.M. Gunn, S.A. Soper, 'Spectroscopic and Binding Properties of NIR Tricarbocyanine Dyes to Double Stranded DNA', *Appl. Spectrosc.*, **50**, 211–221 (1996).

124. S.A. Soper, B.L. Legendre, 'Binding Properties of NIR Dyes to Proteins and Separation of the Dye/Protein Complexes Using Capillary Electrophoresis with Laser Induced Fluorescence Detection', *Appl. Spectrosc.*, **50**, 1196–1202 (1996).

125. D.A. Wilberforce, G. Patonay, 'Investigation of NIR Laser Dye Albumin Complexes', *Spectrochim. Acta, Part A*, **46**, 1153–1162 (1990).

126. H. Chen, M.S. Farahat, K.Y. Law, D.G. Whitten, 'Aggregation of Surfactant Squaraine Dyes in Aqueous Solution and Microheterogeneous Media: Correlation of Aggregation Behavior with Molecular Structure', *J. Am. Chem. Soc.*, **118**, 2584–2594 (1996).

Phosphorescence Measurements, Applications of

Marta E. Díaz García and Rosana Badía
University of Oviedo, Spain

1 Introduction **1**
2 Theory **2**
2.1 Fluorescence and Phosphorescence 2
2.2 Kinetics 2
2.3 Quenching and the Stern–Volmer Relationship 3
3 Instrumentation **4**
4 Operating Principles and Applications **5**
4.1 Low-temperature Phosphorimetry 5
4.2 Solid-surface Room Temperature Phosphorimetry 6
4.3 Room Temperature Phosphorescence in Ordered Media 7
4.4 Cyclodextrin Room Temperature Phosphorescence 9
4.5 Sensitized Room Temperature Phosphorescence in Liquid Solutions 9
5 Techniques of Lifetime Measurements **11**
5.1 Time-resolved Measurement 11
5.2 Phase-resolved Measurement 11
6 Conclusions **12**
Abbreviations and Acronyms **12**
Related Articles **12**
References **12**

Luminescence is a term used for both fluorescence and phosphorescence. These are emission processes following the absorption of light, and the distinction between them is based upon the process of the deactivation mechanism. Phosphorescence has been identified as the light emitted from the triplet excited state to the ground state, a spin-forbidden transition. Fluorescence is a radiational transition between electronic states of the same multiplicity.

The first analytical uses of phosphorescence were published in the 1950s, but the technique was not widely accepted due to the practical aspect of measuring the signal at cryogenic temperatures (77 K).

Developments in room temperature phosphorescence (RTP) have given rise to fundamental and practical advances which have stimulated interest in phosphorimetry. Solid-surface room temperature phosphorescence (SSRTP) has been generally observed by immobilizing the lumiphor on a solid support such as filter paper or polymeric resins. SSRTP has been shown to be a sensitive and selective approach for the analysis of trace components in biological, pharmaceutical, and environmental samples. Techniques such as SSRTP combined with flow injection analysis provide new avenues for optical sensing transduction and improvement of SSRTP techniques. Applications of SSRTP in flowing systems include the sensing of cations, anions, oxygen, glucose, ethanol, antibiotics, and moisture.

A number of developments in RTP have focused on sensitized phosphorescence (SP) and on the use of ordered media (micelles and cyclodextrins (CDs)), allowing the observation of RTP in fluid solution. In SP the triplet energy of a donor (analyte) is transferred to an acceptor that phosphoresces in fluid solution and the emission of the acceptor is measured. Therefore, SP can only provide quantitative information. In ordered media, a micelle (or a CD) protects the long-lived triplet state from the quenching processes that normally take place in solution, allowing both quantitative and qualitative measurements. SP and ordered media RTP provide the limits of detection for many lumiphors in the nmol^{-1} range and calibration graphs have a wide linear range.

As new instrumental and basic insights emerge in SSRTP, Micelle Stabilized Room Temperature Phosphorescence (MSRTP), and solution-sensitized RTP, these techniques will find much greater use in the trace and ultratrace analysis of dissolved, volatile, and gaseous substances in areas such as pharmaceutical analysis, environmental science, and clinical chemistry.

1 INTRODUCTION

Phosphorescence emission has been observed from a variety of compounds and is differentiated from fluorescence by the long-lived emission light after cut off of the excitation source. However, the technique has not became a popular analytical method due to problems arising from the practical aspect of measuring the signal at cryogenic temperatures.

Recent developments in RTP have given rise to both fundamental and practical advances which have stimulated interest in phosphorimetry. Promising experimental designs involve creative composite techniques such as high-performance liquid chromatography/MSRTP detection, SSRTP, RTP optosensing, RTP fiber optic chemical and biochemical sensors and RTP immunoassays, among other possibilities.

Phosphorescence is a complementary technique to fluorescence, allowing a wider range of molecules to be

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

studied. The sensitivity of phosphorescence is comparable to that of fluorescence; its selectivity is enhanced through the use of excitation and emission wavelengths and by using time discrimination.

The introduction of microelectronics-based phosphorimeters using a pulsed xenon lamp affords easier operation and more sophisticated data treatment using chemometrics.

2 THEORY

2.1 Fluorescence and Phosphorescence

A diagram of the energy levels of a molecule, known as a Jablonski diagram, is shown in Figure 1. Absorption of radiation by a molecule in the lowest vibrational level of the ground state (S_0) results in its excitation to one of the higher energy levels (the lowest, S_1, is called the first excited singlet state). In about 10^{-12}–10^{-14} s, the electronically excited molecule will decay by loss of thermal energy to the lowest vibrational level of the excited states, a nonradiative process known as vibrational relaxation. As the energy gap between excited states is generally small, energy can be transferred from the lowest vibrational levels of S_2 to the highest vibrational levels of S_1. This nonradiative process is known as internal conversion. Eventually, the first excited state (S_1) can be reached by all the excited molecules.

From the lowest vibrational level of S_1, the energy difference to S_0 is usually so great that internal conversion to S_0 has a low probability and is a relatively slow process. In this case, luminescence is observed. Luminescence is the term used for both fluorescence and phosphorescence. In addition to radiative (fluorescence) and nonradiative (internal conversion, vibrational relaxation) pathways, an excited molecule in S_1 can return to the ground state by intersystem crossing to the lowest excited triplet state (T_1). From this level the molecule can again undergo vibrational relaxation to the lowest vibrational level.

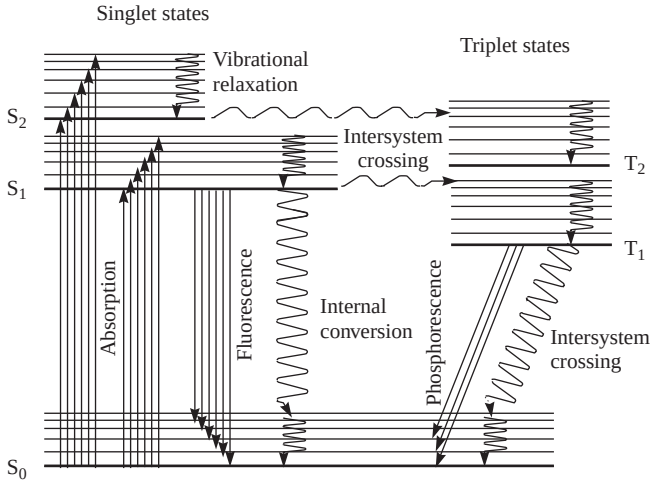

Figure 1 Jablonsky diagram.

The triplet state is lower in energy than the singlet state and a molecule in this triplet state can return to the ground singlet state by different processes, namely emission of radiation (phosphorescence) and nonradiative loss. From a theoretical point of view the main difference between fluorescence and phosphorescence is the nature of the electronic transition involved. Fluorescence is a short-lived emission from the lowest vibrational level of the lowest excited singlet state to the ground state, and does not invoke a change of the electronic spin state. This transition is spin allowed and occurs with a high degree of probability. Phosphorescence involves a change in electronic spin, and thus is called a 'forbidden' transition and is highly improbable. As a result, the radiative lifetimes of fluorescence and phosphorescence are very different – for fluorescence it ranges from 10^{-9} to 10^{-7} s (i.e. the same order as the lifetime of an excited singlet state) and for phosphorescence from 10^{-3} to 10 s. Owing to the relative energies of the excited states, phosphorescence occurs at longer wavelengths than fluorescence.

Given the fact that molecules in the triplet state have long lifetimes, the probability of collisional transfer of energy with other molecules (e.g. solvent molecules) is greatly increased. This process is very efficient in solution at room temperature, being the main pathway for loss of triplet state energy. Phosphorescence requires special experimental conditions under which deactivation processes (intra- and intermolecular) are diminished as much as possible. Fluorescence does not require such special conditions.

A more thorough account of the fundamental radiative and nonradiative processes that are encountered in molecular luminescence spectroscopy is given in standard textbooks and monographs.[1–7]

2.2 Kinetics

Fluorescence and phosphorescence both obey a first-order kinetic rate law; however, the kinetics of the emission as phosphorescence is somewhat more complex than that of fluorescence because an intersystem crossing process to the triplet level is involved.[8,9] As only those molecules in the triplet state emit phosphorescence, the emission process may be written as in Equation (1)

$$T_1 \longrightarrow S_0 + h\nu_p \qquad (1)$$

where h is Planck's constant and ν is the phosphorescent wave frequency.

The kinetic rate equation describing the total deactivation of T_1 is given by Equation (2)

$$\text{rate}_p = \frac{d[T_1]}{dt} = -\left(k_p + k_{is}^{ts}\right)[T_1] \quad (2)$$

where $[T_1]$ is the concentration of molecules in T_1 at any time, k_p is the first-order rate constant for phosphorescence emission, and k_{is}^{ts} the rate constant for intersystem crossing (triplet to singlet). The terms k_p and k_{is}^{ts} have the dimensions of time^{-1}. The rate constant k_p is smaller than the corresponding k_f for fluorescence emission and ranges from $0.033\,s^{-1}$ (for aromatic hydrocarbons) to $10^4\,s^{-1}$. Typical k_f values range from $10^5\,s^{-1}$ to $10^9\,s^{-1}$. Therefore, phosphorescence is a relatively slow process and the fraction of excited molecules that emit phosphorescence is small; there are nonradiative competing processes which occur at comparable or greater rates than phosphorescence.

For a first-order process, the lifetime τ is defined as the time at which the emission intensity has decreased to 1/e (or to 36.8%) of its initial value following termination of excitation, and can be expressed as the reciprocal of its rate constant. Thus, the phosphorescence lifetime is (Equation 3)

$$\tau_p = \frac{1}{k_p} \quad (3)$$

and the observed lifetime of the triplet state is (Equation 4)

$$\tau_t = \frac{1}{\left(k_p + k_{is}^{ts}\right)} \quad (4)$$

The phosphorescence quantum yield Φ_p is defined as the ratio of the number of photons emitted to the number absorbed (Equation 5):

$$\Phi_p = \frac{\text{photons emitted as phosphorescence}}{\text{photons absorbed}} \quad (5)$$

In its simplest form, the kinetics definition of Φ_p is (Equation 6)

$$\Phi_p = \frac{\text{rate of emission}}{\text{rate of absorption}} = \frac{k_p[T_1]}{k_{abs}} \quad (6)$$

where k_{abs} is the rate of light absorption. Assuming that no chemical reactions or quenching take place, three processes can deactivate S_1: fluorescence, internal conversion, and intersystem crossing (singlet to triplet), with rate constants k_f, k_{ic}, and k_{is}^{st}, respectively. So, in the steady state, the rate of absorption equals to the total rate of deactivation of S_1 (Equation 7):

$$k_{deact} = \left(k_f + k_{ic} + k_{is}^{st}\right)[S_1] \quad \text{and}$$
$$k_{abs} = \left(k_f + k_{ic} + k_{is}^{st}\right)[S_1]. \quad (7)$$

Substituting this value in Equation (6) gives Equation (8):

$$\Phi_p = \frac{k_p[T_1]}{\left(k_f + k_{ic} + k_{is}^{st}\right)[S_1]} \quad (8)$$

In the absence of quenching, there is a loss of absorbed energy through intersystem crossing (triplet to singlet) and, at the steady state, Equation (9) holds:

$$k_{is}^{st}[S_1] = k_p[T_1] + k_{is}^{ts}[T_1] \quad (9)$$

where k_{is}^{ts} is the rate constant for intersystem crossing (triplet to singlet).

Now, multiplying the right-hand side of Equation (8) by $k_{is}^{st}[S_1]/k_{is}^{st}[S_1]$ gives Equation (10):

$$\Phi_p = \frac{k_{is}^{st}[S_1]}{\left(k_f + k_{ic} + k_{is}^{st}\right)[S_1]} \frac{k_p[T_1]}{k_{is}^{st}[S_1]} \quad (10)$$

Then, substituting Equation (9) and cancellation of $[S_1]$ and $[T_1]$ gives Equation (11),

$$\Phi_p = \frac{k_{is}^{st}}{k_f + k_{ic} + k_{is}^{st}} \frac{k_p}{k_{is}^{ts} + k_p} \quad (11)$$

which is the complete kinetic definition of Φ_p. This expression is the product of the fraction of excited molecules initially in the S_1 state that cross over to the T_1 state (the triplet quantum yield, Φ_T) and the fraction of molecules in the T_1 state that emit a photon (the phosphorescence quantum efficiency, θ_p; Equation 12):

$$\Phi_p = \Phi_T \theta_p \quad (12)$$

2.3 Quenching and the Stern–Volmer Relationship

Another nonradiative pathway of deactivation of an excited singlet or triplet state is quenching.[10–12] Luminescence is said to be quenched when the lumiphor and another species (quencher) interact, accelerating the decay of an excited state to a lower excited state or to the ground state. This interaction may take place in the excited state of the lumiphor and be collisional in nature. In this situation the process is referred to as dynamic quenching. The excitation energy is transferred to the quencher and is nonradiatively dissipated. Alternatively, the lumiphor and the quencher may associate in the ground state, forming a species that does not display the original luminescence characteristics. This process is known as static quenching.

In a phosphorescence measurement, if quenching is included in Equation (3), the rate of deactivation from the triplet state T_1 is (Equation 13)

$$\text{rate}_p = -\left(k_p + k_{is}^{ts} + k_q[Q]\right)[T_1] \quad (13)$$

where $[Q]$ is the quencher concentration and k_q is the bimolecular quenching rate. Thus, the phosphorescence quantum yield in presence of a quencher is (Equation 14)

$$\Phi_p^q = \frac{k_p}{\left(k_p + k_{is}^{ts} + k_q[Q]\right)} \frac{k_{is}^{st}}{\left(k_f + k_{ic} + k_{is}^{st}\right)} \quad (14)$$

Dividing Equation (11) by Equation (14), the ratio of quantum yields in the absence of quencher (Φ_p) and in the presence of quencher (Φ_p^q) yields Equation (15):

$$\frac{\Phi_p}{\Phi_p^q} = \frac{k_p + k_{is}^{ts} + k_q[Q]}{k_p + k_{is}^{ts}} = 1 + \frac{k_q}{k_p + k_{is}^{ts}}[Q] \quad (15)$$

Using Equation (4),

$$\frac{\Phi_p}{\Phi_p^q} = 1 + k_q\tau_t[Q] = 1 + K_q[Q] \quad (16)$$

Equation (16) is known as the Stern–Volmer relationship, in which K_q is the Stern–Volmer quenching constant and τ_t is the observed triplet lifetime in the absence of quencher. This equation states that plots of Φ_p/Φ_p^q versus $[Q]$ should be linear with a slope equal to K_q. In the case of dynamic quenching, the slope of the plot, K_q, is $\tau_p k_q$. For static quenching K_q is the binding constant of the ground-state complex between the phosphor and the quencher.

Generally, the quencher concentration is always much larger than the concentration of excited molecules being quenched. Then, the product $k_q[Q]$ may be considered a constant as $[Q]$ is not substantially modified during the quenching process. In this case, the product $k_q[Q]$ may be defined as a ''pseudo-first-order rate constant'' and the triplet state lifetime in the presence of quencher τ_p^q is (Equation 17)

$$\tau_p^q = \frac{1}{\left(k_p + k_{is}^{ts} + k_q[Q]\right)} \quad (17)$$

The oxygen molecule in its stable ground state is paramagnetic, having two unpaired electrons in different orbitals. Such an electronic configuration, called a triplet, is unusual for ground-state molecules. Oxygen in this state is a potent quencher, especially of the triplet states, quenching almost all organic phosphors.

Oxygen quenching is diffusion limited and may be described by a modified Stern–Volmer relationship which holds for both fluorescence and phosphorescence and for both lifetimes and intensities (Equation 18):

$$\frac{I_0}{I} = \frac{\tau^0}{\tau} = 1 + k_q\tau^0[O_2] \quad (18)$$

in which I_0, τ^0 and I, τ are the intensities and lifetimes in the absence of oxygen and at an oxygen concentration $[O_2]$, respectively. In the most simplistic terms, Equation (18) means that when O_2 interacts with a lumiphor it decreases the luminescence intensity and shortens its lifetime.

3 INSTRUMENTATION

The equipment required for measuring phosphorescence intensities, lifetimes, excitation and emission spectra are commercially available. Those instruments may be used in low-temperature phosphorescence, MSRTP, and SSRTP, as well as in solution-sensitized RTP work, with no or minor instrumental modifications.[7,12,13]

The major components of fluorescence and phosphorescence instrumentation include a source for exciting radiation, an excitation wavelength selector system, a sample cell, an emission wavelength selector system, and a detector (Figure 2).

Taking advantage of its long lifetime, phosphorescence can be measured in the absence of fluorescence by using a phosphoroscope, a device which allows excitation light to be gated in time.[14] Therefore, if incident radiation is cut off, prompt fluorescence and direct scattering cease almost immediately. However, phosphorescence remains and can be detected free of interference from fluorescence for some time afterwards. Two types of shutter systems are commonly employed:

- The Aminco–Keirs phosphoroscope, in which the sample cell is surrounded by a rotating cylinder with two diametrically opposite apertures. The emitted phosphorescence is collected at right angles to the excitation radiation. The delay between excitation and emission is determined by the speed of rotation and the length of each observation which is determined by the size of the window.
- The Becquerel phosphoroscope, in which the sample cell is placed between two circular rotating discs with opposing 90° sectors cut out of each disk. The duration of each burst of radiation and the delay between them are determined by the speed of rotation. Unlike

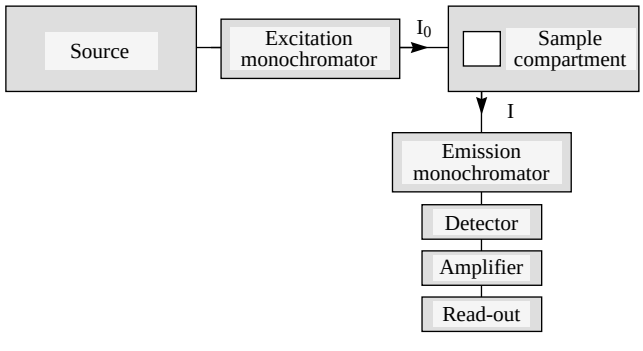

Figure 2 Layout of a typical phosphorescence spectrometer.

the Aminco–Keirs phosphoroscope, this shutter system requires 180° geometry and is not so extensively used. Earlier phosphorimeters used these mechanical chopping systems with a continuous excitation source (e.g. xenon arc lamp, a high-pressure mercury lamp or a tungsten incandescent lamp).

Instrumentation for improved acquisition and analysis of phosphorescence lifetime decays have come to the fore with the increasing use of microprocessors and pulsed sources in luminescence works.

Figure 3 illustrates three cycles of sample excitation and observation for a pulsed-source-gated-detector phosphorimeter. With an initial pulse of source energy, with a duration t_f (typically about 10 μs), the phosphorescence intensity reaches a maximum value and then decays exponentially. At a delay time t_d, after which the source flash has decayed, the photomultiplier detector is turned on, and the integrated phosphorescence signal is measured during the on time (t_g). The photomultiplier detector is then turned off and the sequence repeated. Pulsed source-time resolved phosphorimetry was first proposed by Fisher and Winefordner in 1972.[15]

The advantages of pulsed-source-time-resolved phosphorimeters over conventional instruments with mechanical modulation are:

- A higher source peak intensity can be produced during the pulse without overheating the lamp as the thermal energy is dissipated during the dark periods between the pulses.
- The possibility of improving selectivity for a short-lived phosphor compared to a long-lived phosphor.
- The possibility of measuring phosphors with short lifetimes (0.1–50 ms).[16] The performances of mechanical systems are limited by the maximum speed of rotation available.
- Control over the timing parameters allows the analytical potential of phosphorimetry to be extended, such as the possibility of analyzing mixtures of phosphorescent compounds having different lifetimes.

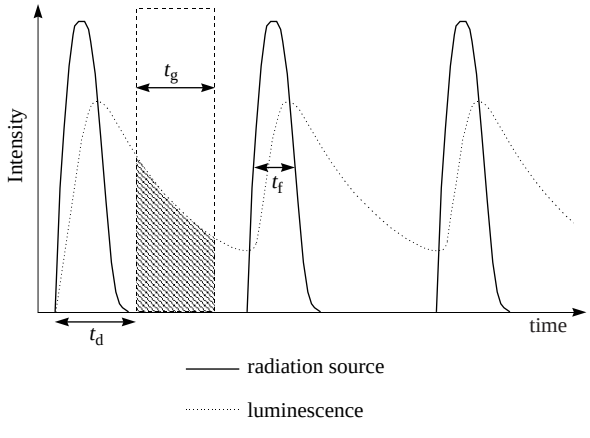

Figure 3 Timing sequence of events occurring during three cycles of sample excitation in a phosphorimeter employing a pulsed radiation source and t_f: width of pulse at half height, t_g: width and t_d delay time between start of pulse and observation.

4 OPERATING PRINCIPLES AND APPLICATIONS

4.1 Low-temperature Phosphorimetry

Low-temperature phosphorimetry from solutions was considered for years as the way of obtaining analytically useful phosphorescence signals. However, low-temperature phosphorescence has not become a popular analytical method because of the need for cryogenic cooling, the irreproducibility of the formation of the cooled matrix, and the problems related to sample introduction into the phosphorimetric system.[17]

Up to the present time, the immersion method has been widely used for cooling the sample, which is lowered into a quartz Dewar flask filled with a coolant. Among the possible coolants (liquid nitrogen, oxygen, rare gases, air) only liquid nitrogen is safe, inexpensive and pure enough to be non-luminescent and transparent at all wavelengths between 200 nm and 800 nm.

Depending on the rate of sample cell cooling and the chemical nature and composition of the solvent system, the cooled matrix can be a clear glass, a cracked glass, or a snow. When the capillary tube makes contact with liquid nitrogen, the sample cools very quickly, and cracked glasses or snows result for most solvents. This problem can be avoided if the capillary tube is gradually cooled over the liquid nitrogen before contact is made.

Only a few solvents form clear glasses at low temperature, the most notable being ethanol, isopentane, and *n*-propanol. However, several solvents that are not useful alone do form good mixed solvents. So, ethanol has been combined with a variety of other solvents, the most popular of the mixed solvents being EPA, a 5:5:2 (by volume) mixture of ether, isopentane, and ethanol.

The above practical disadvantages can be minimized by using a device based on conduction cooling of a short capillary sample cell in a copper block cooled by liquid nitrogen. These devices allow improved sample turnover time to about one sample per minute, it is not necessary to handle Dewar flasks, and the amount of liquid nitrogen is reduced. Unfortunately, it is difficult to obtain good thermal contact between the sample cell and the copper block. Also, it is not easy to thermally insulate the cooling device, which results in difficult quantitative analysis due to fogging of the view area.

4.2 Solid-surface Room Temperature Phosphorimetry

Roth[18] in 1967, first listed the RTP detection limits of 18 organic compounds adsorbed on filter paper. Since this report, the field of analytical SSRTP has expanded considerably.[19–23] Only certain materials are useful for inducing RTP from adsorbed phosphors, among which filter paper is the most widely used. Nearly all of the solid materials employed to develop RTP give phosphorescence background signals. Only sodium acetate has a low background emission. Thus one criterion for selection of a solid surface is to measure the relative RTP signal of the sample compared to the background RTP of the solid surface. Vo-Dinh[13] compared about 20 different kinds of commercial filter papers using pyrene as a model analyte. It was found that the difference in performance between papers was substantially less than an order of magnitude.

In RTP work, the solvent used to deposit the phosphor on the solid substrate can influence the final RTP signal. Acidic, alkaline and neutral solutions have been used in a number of applications. Also, ethanol–water mixtures have been extensively employed. In order to produce spots yielding rapid and reproducible results, a microsyringe or a micropipet is used to deliver the sample to the surface. Once the phosphor is spotted on the solid surface, it is necessary to dry the adsorbed sample prior to the RTP measurement step. In fact, moisture is one of the main factors that drastically diminish the RTP signal. There are different ways to dry the sample: blowing hot air onto the sample, placing the sample inside a desiccator, heating the sample inside an oven, or placing the sample under an infrared heating lamp. The use of an infrared heating lamp has been found to be very practical as the method is rapid (3–5 min compared with 1–5 h in a desiccator), convenient and, usually, nondestructive. The dry sample is kept dry during RTP measurement by passing a flow of air, nitrogen, or argon through the sample compartment. The main steps of an RTP assay are shown in Figure 4 for the method using filter paper.

There is no general model that explains the interactions required to develop RTP from compounds spotted on solid supports. It has been proposed that hydrogen bonding of polar or ionic organic compounds to the hydroxyl groups of the paper (or to the surface silanol groups of silica gel) is the main mechanism that provides the rigidity required to restrict collisional quenching of phosphor molecules in the triplet state. In this single working model, the moisture effect is explained by assuming that water molecules disrupt the hydrogen bonding, thus facilitating oxygen accessibility to the phosphor in the sample matrix.

So far, the analytical potential of SSRTP has been hampered by its discontinuous character, the necessary sample manipulation, support background, and water/oxygen quenching, which limit conventional practical work. Attempts have been made to develop an automatic instrument for SSRTP based on a continuous filter paper on which the samples were sprayed using a conventional atomic absorption nebulizer (Figure 5).

Promising uses for SSRTP include the development of fiber optical chemical sensors. Campiglia and Vo-Dinh[24] have reported the development of a universal RTP sensor adaptable to different analytical situations involving phosphorescent compounds. Filter paper

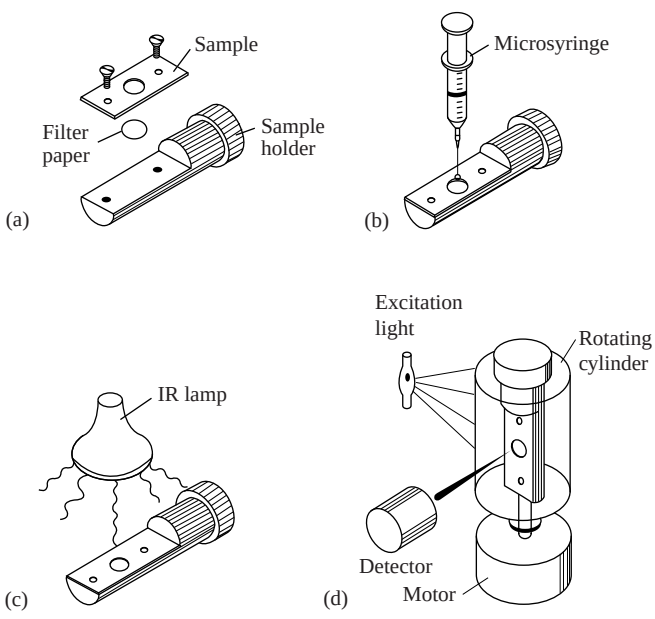

Figure 4 The procedure for RTP analysis using a filter paper substrate: (a) sample holder preparation; (b) sample spotting; (c) sample drying with an infrared lamp; (d) RTP measurement. (Reproduced by permission of John Wiley & Sons from T. Vo-Dinh, *Room Temperature Phosphorimetry for Chemical Analysis*, 99 (1984).)

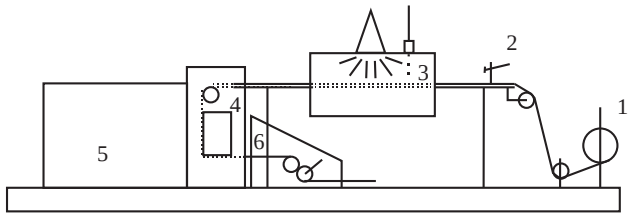

Figure 5 Side view diagram of an AutoAnalyzer continuous filter coupled to a Perkin Elmer LS-5 luminescence spectrofluorimeter: (1) filter paper roll; (2) atomic absorption nebulizer; (3) drying chamber; (4) spectrofluorimeter; (5) modified sample compartment; (6) Technicon AutoAnalyzer continuous filter (Technicon Instruments, Terrytown, NY, USA). (Reproduced by permission of Society for Applied Spectroscopy from A.D. Campiglia, L.M. Perry, J.D. Winefordner, *Appl. Spectrosc.*, **43**, 1431–1443 (1989).)

pretreated with thallium(I) acetate was used as a solid support to increase phosphorescence emission from several model polycyclic aromatic hydrocarbons (PAHs) and nitrogen heterocyclic compounds. Advantage was taken of coupling fiber optics into a pulsed nitrogen laser excitation source to measure RTP emission from compound imbibed on the filter paper. Detection limits at the nanogram per liter level were estimated for all the compounds studied. Reproducibility, highly dependent on drying conditions, ranged from 6.1% for chrysene to 11.9% for 7,8-benzoquinoline. Linear dynamic ranges extended over 2–3 orders of magnitude.

In spite of the generally accepted idea that water adversely affects SSRTP, it has been demonstrated that analytically useful SSRTP can be observed in a continuous aqueous flow provided that the solid support is adequately selected for a given phosphor. The solid support is conveniently packed into a flow cell and the analyte introduced into the flow system, either continuously or by means of an injection valve. The approach involves the integration of the basic reaction/retention/detection processes at the solid support in the flow cell. This detection principle in flow analysis is known as *optosensing* at active surfaces.

RTP optosensing is an important new development in phosphorimetry. Not only can conventional organic phosphors be determined, but also metal ions, anions, surfactants, antibiotics, gases, and vapors (e.g. oxygen, humidity). The favorable analytical performance of RTP optosensing systems for oxygen, allows its use as an optical transducer for enzymatic biosensing in oxygen-mediated reactions, both in solution (e.g. glucose determination) or in a gaseous environment (e.g. ethanol determination; Table 1). It is clear that the combination of flow analysis and SSRTP offer promising new analytical possibilities to phosphorimetry, particularly in the field of optical sensing.[36]

4.3 Room Temperature Phosphorescence in Ordered Media

4.3.1 Organized Surfactant Assemblies: Micelle Stabilized Room Temperature Phosphorescence

Surface active agents (surfactants) are amphiphilic molecules in which a polar head group is attached to a long nonpolar chain. The nonpolar moiety is usually a hydrocarbon tail and the polar region, the head, can be either an ionic or a polar neutral group. These amphiphilic molecules associate dynamically in solution to form relatively well-defined assemblies termed micelles. Micelle formation starts at a characteristic surfactant concentration, the critical micelle concentration (CMC). In aqueous solution, individual surfactant molecules aggregate in such a way that the hydrophobic tails are packed together, forming a nonpolar core, while the hydrophilic heads face the aqueous solution, forming a boundary surface between the core and the isotropic aqueous solution.

An understanding of the organized assembly effect, together with the knowledge of the degree of interaction between the organized medium and the luminescence system being examined, are the keys in developing new analytical luminescence methodologies. In fact, the degree and type of binding between any lumiphor-organized system combination depend on the hydrophobic–coulombic net interactions. These are determined by the relative chemical nature of the luminescent probe and that of the organized medium considered. In this sense, localization of a lumiphore in a given micellar medium may be in any or all of several regions of the aggregate. An amphiphilic solute may align itself with the polar portion in the surface layer of the micelle with the tail directed inward, towards the core of the micelle. Ionic species with an opposite charge to the micellar aggregate may be bound to the micellar surface by electrostatic attraction. Nonpolar species may penetrate to a particular depth into the surface layer.

Table 1 RTP chemical and biochemical optosensing

Analyte	Solid support	Phosphor	LOD	RSD (%)	Ref.
Aluminum	Dowex 1 × 2-100®	Al-ferron	$2\,\mu g\,L^{-1}$	3.2	25
Terbium	Cation exchanger	Tb-pyrazolone	3×10^{-9} M	–	26
Gadolinium	Chelex-100®	Gd-pyrazolone	5×10^{-6} M	2	27
Iodine	Dowex 1 × 2-100®	8-Quinolinol-5-sulphonic	$10\,\mu g\,mL^{-1}$	2.7	28
Anthracyclines	Amberlite XAD-2®	Eu-anthracycline	$6–9\,ng\,mL^{-1}$	1	29
Tetracyclines	Amberlite XAD-2®	Eu-tetracycline	$0.25–0.40\,ng\,mL^{-1}$	1	30
Anionic surfactant	Dowex 1 × 2-200®	Al-ferron	5×10^{-3} mM	0.7	31
Oxygen	Bondesil®	Erythrosine B	0.000 06% (0.6 ppm)	0.2	32
Glucose	Dowex 1 × 2-200®	Al-ferron	0.1 mM	–	33
Ethanol	Amberlite XAD-2®	Erythrosine B	8 mM	1	34
Moisture	Sol–gel matrix	Al-ferron	0.09% RH	2	35

LOD, limit of detection; RH, relative humidity.

Normal micelles have been the organized media more commonly used to develop analytically useful RTP in fluid solutions. In a conventional experiment, the luminescent molecule is dissolved in a sodium lauryl sulfate (SLS) micellar solution and then diluted with thallium lauryl sulfate (TlLS)/SLS solution to obtain a total detergent concentration of $0.10\,mol\,L^{-1}$ and a 30 : 70 ratio of Tl/Na. The solution is finally deoxygenated by purging with nitrogen for 15–20 min and analyzed using time discrimination.

4.3.2 Principles and Applications of Micelle Stabilized Room Temperature Phosphorescence

MSRTP was first reported by Kalyanasundaram et al.[37] Cline Love and Weinberger[38] showed that three fundamental requirements are necessary to develop analytically useful MSRTP for most lumiphors: securing the presence of micellar aggregates (using a surfactant concentration above its CMC), the presence of a heavy atom or a heavy species, and securing oxygen removal.

RTP can be observed from many molecules in fluid solution by incorporating the phosphor into a micellar system. The protective effect of the micellar microenvironment greatly minimizes collisional deactivation of triplet species by restricting their diffusive motions. Only at surfactant concentrations well above the CMC is the protective effect on phosphorescence observed (micellar effect). An additional factor that favorably influences the RTP observed is the relative micelle/phosphor polarity and/or charge. So, no MSRTP is observed in anionic micellar media for negatively charged molecules. However, these negatively charged phosphors experience, in a cationic micellar medium, attractive electrostatic interactions acting concurrently with the hydrophobic ones to provide the necessary immobilization for the observed strong RTP signals.

However, micelles can be functionalized with a heavy atom; for example, the countercation of anionic micelles can be replaced by a heavy element such as silver or thallium. The local effective concentration of the heavy atom in the micelle is then larger than its actual molar concentration. The intrinsic proximity between the phosphor solubilized in the micelle and the heavy counterions results in a heavy atom effect, thus enhancing the rate of intersystem crossing. It is known that the presence of heavy atom perturbers results in shorter phosphorescence lifetimes and causes a considerable reduction of fluorescence emission. Adequate choice of surfactant assemblies and heavy atoms allows one to observe analytically useful MSRTP from a variety of compounds. For example, MSRTP has been observed from polynuclear aromatic and heterocyclic compounds solubilized in SLS anionic micelles and in the presence of thallium, silver, or lead as heavy atom perturbers.[38]

Oxygen is a potent quencher of the excited triplet state. Micellar media provide a microenvironment that can protect a solubilized phosphor from interactions with oxygen allowing for MSRTP in some solutions that have not been deoxygenated. In order to enhance MSRTP, oxygen removal is usually carried out by using an inert gas (e.g. nitrogen). Diaz-Garcia and Sanz Medel[39] studied in detail the use of sodium sulfite as a chemical oxygen scavenger for MSRTP (Equation 19).

$$2SO_3^{2-} + O_2 \longrightarrow 2SO_4^{2-} \qquad (19)$$

The method is simple and reliable. The analytical performance of MSRTP with sulfite deoxygenation was tested for determination of arenes,[40] metals (Table 2), nucleic acids[45] as well as for immunoassays.[46] This simple technique can be extended to chemical deoxygenation of other organized media such as vesicles and CDs to develop RTP in solution,[47] as well as in flowing systems.[48] MSRTP provides limits of detection for many aromatic molecules in the nanomole per liter range and calibration graphs have a wide linear range. The overall precision is about 5%. It is worth mentioning the high sensitivity observed if one considers the high probability of nonradiative pathways that may occur in solution at room temperature.

Cline Love and Weinberger[38] have demonstrated that MSRTP techniques can be applied for detection of phosphors in liquid chromatography in two different approaches:

- Using the micellar solution as a mobile phase. A liquid chromatography/MSRTP procedure for

Table 2 Analytical figures for metal ion MSRTP determination

Metal ion	Reagent medium	λ_{exc} (nm)	λ_{em} (nm)	pH	LOD ($ng\,mL^{-1}$)	RSD (%)	Ref.
Nb^{V}	Ferron/CTAB	363	572	5.7	4	2	41
Al^{III}	Ferron/CTAB	386	586	5.5	5.4	4.5	42
Ga^{III}	Ferron/CTAB	396	615	5.5	5	4	43
Pd^{II}	Coproporphyrin IV/Triton X-100®	395	666	1.0	0.18	2	44

CTAB, cetyl trimethyl ammonium bromide.

different PAHs was described in detail using an SLS/TlLS (70%/30%) micellar mobile phase and a propylcyano-bonded column as stationary phase. Detection limits were in the nanogram range for PAHs investigated in coal liquids and wastewater products. Working ranges covered more than three orders of magnitude.

- In a postcolumn addition mode, the micellar phase is merged with the mobile phase after chromatographic separation, as demonstrated for aromatic compounds.

Detergents have also been used in paper-substrate RTP, leading to signal enhancements in the twofold to ninefold range for PAHs adsorbed on filter paper impregnated with Tl/SLS. The main interactions proposed in this system do not involve the presence of micelles but the enhancement is due to improved heavy atom/phosphor proximity and to an increased migration of the species through the cellulose pores of the paper.

4.4 Cyclodextrin Room Temperature Phosphorescence

CDs are cyclic oligosaccharides composed of α-(1 $\rightarrow$ 4)-linked D-glucopyranose units forming thus a truncated cone-shaped structure with a central cavity. The best-known members of this class of natural products are α-, β-, and γ-CDs, as they are commonly referred to, composed of six, seven, and eight glucose units, respectively. The hydrophilic hydroxyl groups of the glucose units occupy both rims of the cone, rendering the CDs soluble in water, whereas the cavity, lined by hydrogen atoms and the glycosidic oxygen bridges, is relatively hydrophobic compared to water. Thus, CDs provide a microheterogeneous environment in water. The unique structural features of the CD cavity explain some of the unusual characteristics of these compounds.

Because of these structural properties, CDs can form inclusion complexes with guest molecules that fit into the cavity, even if only partially. Complexation is based on a combination of several intermolecular interactions, including hydrophobic, electrostatic and dipole–dipole interactions, dispersive forces, steric fit, and hydrogen bonding.

As a consequence of inclusion-complex formation, the microenvironment around the guest molecule in the CD cavity is different from that in the bulk media. In fact, inclusion of a lumiphor in a CD cavity results in important changes in the photophysical and/or photochemical properties of the guest. Analytical luminescence benefits that result from complexation with CDs include:

- Protection of excited states from collisional quenching and other nonradiative pathways. The cavity of the CD is a finite space in that, once it has been occupied by a lumiphor, other molecules are excluded from occupying the space at the same time. This prevents collisional encounters. Also, constraints imposed on the mobility of the guest reduce the rate of the radiationless transition of excited singlet and/or triplet states.
- Solubility enhancement of poorly water-soluble lumiphors.
- Partial protection of phosphor guests against quenching by dissolved oxygen in solution.

Although cyclodextrin room temperature phosphorescence (CDRTP) has not been widely used as an analytical technique, CDs have significant potential as ordered media to develop new RTP methodologies in solution. As in MSRTP, a heavy atom (external or intramolecular) or a heavy structure (e.g. 2-bromoethane) should be present to populate the triplet state. Sample deoxygenation, either using a chemical scavenger as sodium sulfite or bubbling an inert gas, results in increased phosphorescence intensity.

In a conventional experiment for CDRTP analysis of PAHs an aliquot of the analyte of interest is added to a flask and the solvent is removed by gently heating on a hotplate. An aliquot of 1,2-dibromomethane (heavy atom perturber) is then added, followed by dilution with 0.01 mol L^{-1} aqueous CD solution. After shaking vigorously by hand, a cloudiness in the solution can appear due to precipitation of excess 1,2-dibromomethane complexed by CD. This turbidity has no influence on the reproducibility and quality of RTP spectra. Finally, the sample is de-areated for about 15 min with nitrogen and placed in a standard fluorescence cell with a Teflon® stopper for RTP measurement. Following this method, sample preparation time is about 20 min.

4.5 Sensitized Room Temperature Phosphorescence in Liquid Solutions

Direct solution room temperature intensities are generally too low to be analytically useful. Donkerbroek et al.[49] applied the technique of SP in liquid solutions as a useful alternative to direct solution RTP.

In SP the analyte (a weakly or nonphosphorescent species) is excited by means of light absorption. Subsequently, the analyte transfers its triplet energy to an acceptor molecule, thus producing an acceptor triplet with a high phosphorescent yield (Figure 6).

Sensitized RTP is observed under the following conditions:

- When the triplet state energy of the acceptor is lower than that of the donor and when the energy gap

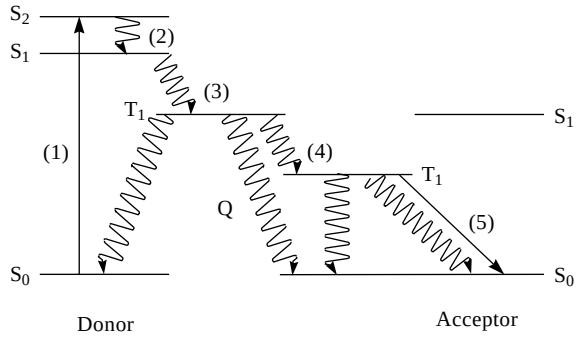

Figure 6 Diagram showing the SP process. The donor (analyte) is excited to the S_2 electronic level (1); then, internal conversion (2) and intersystem crossing (3) can take place. Subsequently, energy transfer to a suitable acceptor having a triplet energy lower than that of the donor occurs (4) and, eventually, phosphorescence of the acceptor is observed. Straight arrows refer to radiative processes and wavy arrows to nonradiative ones.

between the donor and acceptor triplet states is sufficiently high to prevent significant reverse transfer mechanisms.

- When the lifetime of the donor excited state is longer than the energy transfer process in the solvent under consideration. The energy-transfer efficiency $\theta_t^{D,A}$ is given by Equation (20):[49]

$$\theta_t^{D,A} = \frac{K_t[A]}{\left(1/\tau_t^D + K_t[A]\right)} \tag{20}$$

where K_t is the rate constant of the bimolecular energy-transfer reaction, τ_t^D is the triplet lifetime of the donor (D) in the absence of the acceptor A, and [A] is the acceptor concentration. Therefore, appreciable sensitized RTP will occur if $K_t[A] > 1/\tau_t^D$. This condition can be met if solvent deoxygenation is carried out.

- If absorption of excitation radiation by the acceptor is negligible.
- If there is an appreciable population of the donor's triplet state.

Sensitized RTP has been observed for a variety of organic species (drugs, polynuclear aromatics, etc.) and compounds that do not fluoresce in liquid solutions. However, only about 10–15% of the molecules that meet the energy requirements are found to undergo SP. The potential acceptors found useful in sensitized RTP are limited, with biacetyl and 1,4-dibromonaphthalene being the acceptors of choice in many applications.

Sensitized RTP has been successfully applied as a detection method for both chromatographic and continuous-flow systems, and the approach can be considered as complementary to fluorescence detection. Spectroscopically, a drawback of sensitized RTP is that all spectral information about the donor is lost as the emission spectrum is that of the acceptor. The advantages of using SP are:

- The emission wavelength is constant because, for all donors, the same emitting species is employed.
- Analytes are detected at a relatively long emission wavelength of the acceptor, which results in minimizing the effects of background fluorescence and stray light.
- In principle, the approach could be useful for those compounds that phosphoresce in cryogenic conditions, because the primary condition to be fulfilled is an efficient triplet formation.

A complimentary technique with less requirements is quenched RTP in liquid solutions, in which quenching of biacetyl phosphorescence by the analyte acceptors is used to determine the latter.

For sensitized RTP a suitable wavelength is used for analyte excitation and biacetyl phosphorescence is detected. However, in quenched RTP a wavelength is selected where the molar absorptivity of biacetyl is relatively high and it is the only species excited. Thus, the analyte (acting as an acceptor) leads to a decrease of the biacetyl RTP.

Donkerbroek et al.[50] derived Equation (21),

$$\frac{1}{I'_{dir}} = \frac{1}{I_{dir}} + \frac{k_q[Q]}{I_{abs}^B \theta_{isc}^B k_p^B} \tag{21}$$

where I_{dir} is the direct RTP intensity of biacetyl in the absence of quencher (analyte), I'_{dir} is the RTP of biacetyl in the presence of quencher, k_q is the bimolecular rate constant of the quenching reaction, [Q] is the quencher concentration, I_{abs}^B is the rate of light absorption by biacetyl, θ_{isc}^B is the triplet formation efficiency of biacetyl, and k_p^B is the rate constant of the phosphorescence process.

As indicated by this equation, a straight line with a slope proportional to k_q and an intercept equal to the inverted intensity of the nonquenched signal is obtained by plotting $1/I'_{dir}$ versus [Q]. Therefore, for a particular analyte the sensitivity of the quenched RTP is proportional to the k_q value. Donkerbroek et al.[50] estimated the limit of detection (LOD) for various compound classes by quenched RTP using Equation (22):

$$\mathrm{LOD(M)} = \frac{10}{k_q(\mathrm{M^{-1}\,s^{-1}})} \tag{22}$$

Therefore, determination of k_q for a number of analytes is of importance to evaluate the analytical potential of the approach. Quenched RTP in solution

holds considerable potential for use in continuous-flow detector systems. It can also be applicable to many compounds that may have inherently poor detection properties by other techniques. For example, it could be used for nonultraviolet-absorbing species detection in ion chromatography (nitrite, sulfite, *cis*-platin, etc.).

5 TECHNIQUES OF LIFETIME MEASUREMENTS

The commonly used approaches to measure luminescence lifetimes are time-resolved and phase-resolved measurements.[7]

5.1 Time-resolved Measurement

In this method a pulsed radiation source excites the sample and the decay profile is monitored. If the excitation pulse is of very short duration (10 μs) and assuming that only a single decaying component is present, the emission is a monoexponential decay (Equation 23)

$$I_{(t)} = I_0 e^{-t/\tau} \tag{23}$$

where $I_{(t)}$ is the phosphorescence intensity of the compound at a delay time t, and I_0 is the phosphorescence intensity at time $t = 0$. The luminescence lifetime τ is the mean excited state lifetime.

For a binary system of components A and B, the theoretical expressions for pulsed-source phosphorimetry are (Equation 24)

$$\begin{aligned} I_t^{(A)} &= I_0^{(A)} e^{-t/\tau_A} \\ I_t^{(B)} &= I_0^{(B)} e^{-t/\tau_B} \\ I_t^{(A+B)} &= I_t^{(A)} + I_t^{(B)} \end{aligned} \tag{24}$$

where $I^{(A+B)}$ is the total phosphorescence intensity of the mixture. For multicomponent systems similar expressions may be derived.

This technique has important advantages over simple intensity measurements as it involves comparing the luminescence intensity at various times, it is self-referencing, and it compensates for factors such as changes in the efficiency of the source, detector, or electronics.

The three methods generally used to determine phosphor concentrations by time-resolved phosphorimetry are the multiple analytical curve method, the exponential method, and the logarithmic decay time method.[12]

Time-resolved phosphorescence for organic compounds is now well established and was first suggested by Winefordner in 1969.[51] Application of chemometric techniques to the interpretation of RTP emission spectrum–decay data and to simultaneous determination of organic[52] and inorganic[53] species has also been reported recently.

5.2 Phase-resolved Measurement

This method allows phase resolution of the phosphorescence signal from phosphors with different lifetimes. The sample is excited using a source (usually a continuous-wave laser) whose light intensity is modulated at a frequency f.

Table 3 summarizes the equations derived by Aaron and Winefordner, describing the phase and frequency characteristics of luminescence.[54]

Phase-resolved phosphorimetry is especially useful for strong phosphors or concentrated solutions of weak phosphors if fluorescence or stray light interferences are not present. Also, it is adequate for short time analysis (<1 ns) of molecular decays, as the relative error decreases slightly with the faster decay times of the sample under study.

Table 3 Theoretical expressions for phase-resolved phosphorimetry

Total luminescence intensity	dc term	ac term
	$I_L = k_L I_0' + m_L k_L I_0'' \cos(wt - \theta_L)$	
Individual intensities		
Fluorescence	$I_F = k_F I_0' + m_F k_F I_0'' \cos(wt - \theta_F)$	
Phosphorescence	$I_P = k_P I_0' + m_P k_P I_0'' \cos(wt - \theta_P)$	
Frequency functions of the ac term of phosphorescence	Degree modulation	Phase-shift angle
	$m_P = (1 + 4 - \pi^2 f^2 \tau_P^2)^{-1/2}$	$\theta_P = \tan^{-1}(2\pi f \tau_P)$

k_L, k_F and k_P are factors taking into account the quantum efficiency and concentration factors for total luminescence, fluorescence and phosphorescence, respectively; I_0' and I_0'' are the constant-intensity term and sinusoidally varying intensity term of the exciting light function $I_0(t)$, respectively; m_L, m_F, and m_P are the degrees of modulation of total luminescence, fluorescence, and phosphorescence respectively; θ_L, θ_F, and θ_P are the phase-shift angles (in degrees) of the total luminescence, fluorescence, and phosphorescence respectively; τ_P phosphorescence lifetime (seconds); w = angular frequency (hertz); f = linear frequency (hertz). ac, alternating current; dc, direct current. (Reproduced by permission of Pergamon Press from J.J. Aaron, J.D. Winefordner, *Talanta*, **22**, 707–715 (1975).)

6 CONCLUSIONS

In spite of the fact that analytical applications of phosphorescence have been somewhat scarce in the past, the examples quoted in this article indicate the variety of possibilities for developing new highly selective, sensitive, simple and rapid phosphorimetric methodologies. With the recent advances in RTP, the introduction of new instrumentation, and the breadth of potential applications, the use of analytical phosphorimetric approaches is likely to increase.

The majority of effort in the future should go into the process of translating research findings into practical applications and/or commercially viable systems. Areas such as miniaturization and development of portable phosphorimeters, and the design and construction of phosphorimetric sensors to monitor analytes in a continuous way, should also deserve special attention, particularly in clinical chemistry and pharmaceutical and environmental fields.

ABBREVIATIONS AND ACRONYMS

ac	Alternating Current
CD	Cyclodextrin
CDRTP	Cyclodextrin Room Temperature Phosphorescence
CMC	Critical Micelle Concentration
CTAB	Cetyl Trimethyl Ammonium Bromide
dc	Direct Current
EPA	Mixture of Ether, Isopentane, and Ethanol
LOD	Limit of Detection
MSRTP	Micelle Stabilized Room Temperature Phosphorescence
PAHs	Polycyclic Aromatic Hydrocarbons
RH	Relative Humidity
RTP	Room Temperature Phosphorescence
SLS	Sodium Lauryl Sulfate
SP	Sensitized Phosphorescence
SSRTP	Solid-surface Room Temperature Phosphorescence
TlLS	Thallium Lauryl Sulfate

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Fluorescence in Organized Assemblies • Fluorescence Lifetime Measurements, Applications of

REFERENCES

1. H.A. Strobel, W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, 3rd edition, Wiley, Toronto, 1989.
2. H.H. Willard, L.L. Merritt, Jr, J.A. Dean, F.A. Settle, Jr, *Instrumental Methods of Analysis*, 7th edition, Wadsworth Publishing Co., California, 1988.
3. D.A. Skoog, J.J. Leary; *Principles of Instrumental Analysis*, 4th edition, McGraw-Hill, London, 1995.
4. E.L. Wehry, *Modern Fluorescence Spectroscopy*, Plenum Press, 1976, Vols. 1, 3 and 4, 1981.
5. C.A. Parker, *Photoluminescence of Solutions*, Elsevier, Amsterdam, 1968.
6. S.G. Schulman, *Molecular Luminescence Spectroscopy. Methods and Applications*, Wiley, New York, 1985.
7. J.D. Ingle, Jr, S.R. Crouch, *Spectrochemical Analysis*, Prentice-Hall, NJ, 1988.
8. D.O. Cowan, R.L. Drisko, *Elements of Organic Photochemistry*, Plenum Press, New York, 1976.
9. D. Roberts, *Light, Chemical Change and Life: A Source Book in Photochemistry, Block 4: Physical Properties of Excited States*, Open University Press, Milton Keynes, 1982.
10. O. Stern, M. Volmer, 'Über die Abklingungszeit der Fluoreszenz', *Physik. Zeitschr.*, **20**, 183–188 (1919).
11. N.J. Turro, *Modern Molecular Photochemistry*, The Benjamin/Cummings Publishing Co., Menlo Park, CA, 1978.
12. R.J. Hurtubise, *Phosphorimetry: Theory, Instrumentation, and Application*, VCH, Weinheim, 1990.
13. T. Vo-Dinh, *Room Temperature Phosphorimetry for Chemical Analysis*, Wiley, New York, 1984.
14. G.G. Guilbault, *Comprehensive Analytical Chemistry*, Elsevier, Oxford, Vol. III, 1977.
15. R.P. Fisher, J.D. Winefordner, 'Pulsed Source-time Resolved Phosphorimetry', *Anal. Chem.*, **44**, 948–956 (1972).
16. K.F. Harbaugh, C.M. O'Donnell, C.M. Winefordner, 'Pulsed Source, Time Resolved Phosphorimetry Determination of Phosphorescence Lifetimes', *Anal. Chem.*, **45**, 381–382 (1973).
17. G.G. Guilbault, in *Wilson and Wilson's Comprehensive Analytical Chemistry*, ed. G. Svehla, Elsevier, Oxford, Vol. III, 1977.
18. M. Roth, 'Ambient Temperature Phosphorescence. A Selective and Nondestructive Method for Detection of Some Aromatic Compounds on Paper Chromatograms

and on Cellulose Layers', *J. Chromatogr.*, **30**(1), 276–278 (1967).
19. R.T. Parker, R.S. Freedlander, R.B. Dunlap, 'The Development of Room Temperature Phosphorescence into a New Technique for Chemical Determination. Part 1. Physical Aspects of Room Temperature Phosphorescence', *Anal. Chim. Acta*, **119**, 189–205 (1980).
20. R.J. Hurtubise, *Molecular Luminescence Spectroscopy. Methods and Application*, ed. S.G. Schulman, Wiley-Interscience, New York, Part II, 1988.
21. R.T. Parker, R.S. Freedlander, R.B. Dunlap, 'The Development of Room Temperature Phosphorescence into a New Technique for Chemical Determination. Part 2. Analytical Consideration of Room Temperature Phosphorimetry', *Anal. Chim. Acta*, **120**, 1–17 (1980).
22. M. Gunshefski, J.J. Santana, J. Stephenson, J.D. Winefordner, 'Solid-surface Room Temperature Phosphorescence', *Appl. Spectrosc. Rev.*, **27**, 143–192 (1992).
23. R.J. Hurtubise, 'Solid-matrix Luminescence Analysis: Photophysics, Physicochemical Interactions and Applications', *Anal. Chim. Acta*, **351**, 1–22 (1997).
24. A.D. Campiglia, T. Vo-Dinh, 'Fiber Optic Sensors for Laser-induced Room Temperature Phosphorescence Detection of Polycyclic Aromatic Compounds', *Talanta*, **43**, 1805–1814 (1996).
25. R. Pereiro García, Y.M. Liu, M.E. Díaz García, A. Sanz Medel, 'Solid-surface Room Temperature Phosphorescence Optosensing in Continuous Flow Systems: An Approach for Ultratrace Metal Ion Determination', *Anal. Chem.*, **63**, 1759–1763 (1991).
26. Z.L. Gong, Z.L. Zhang, A.Y. Zhang, 'Room Temperature Phosphorescence Optosensor for Terbium', *Anal. Lett.*, **29**(4), 515–527 (1996).
27. Z.L. Gong, Z.L. Zhang, 'Room Temperature Phosphorescence Optosensing for Gadolinium', *Mikrochim. Acta*, **126**(1–2), 117–121 (1997).
28. Y.M. Liu, R. Pereiro García, M.E. Díaz García, A. Sanz Medel, 'Room Temperature Luminescence Optosensing Based on Immobilized Metal Chelates: Application to Iodine Determination', *Anal. Chim. Acta*, **255**, 245–251 (1991).
29. F. Alava Moreno, M.J. Valencia González, M.E. Díaz García, 'Room Temperature Phosphorescence Optosensor for Anthracyclines', *Analyst*, **123**, 151–154 (1998).
30. F. Alava Moreno, M.E. Díaz García, A. Sanz-Medel, 'Room Temperature Phosphorescence Optosensor for Tetracyclines', *Anal. Chim. Acta*, **281**, 637–644 (1993).
31. R. Badía, M.E. Díaz García, 'Room Temperature Phosphorescence Flow Through Biosensing of Anionic Surfactants', *Anal. Chim. Acta*, **371**, 73–80 (1998).
32. R. Badía, M.E. Díaz García, A. Sanz Medel, 'A Sensitive Probe for Oxygen Sensing in Gas Mixtures, Based on Room-temperature Phosphorescence Quenching', *Mikrochim. Acta*, **121**, 51–61 (1995).
33. M.J. Valencia González, Y.M. Liu, M.E. Díaz García, 'Optosensing of D-Glucose Oxidase Minireactor and an Oxygen Room Temperature Phosphorescence Transducer', *Anal. Chim. Acta*, **283**, 439–446 (1993).
34. M.J. Valencia González, N. Velasco García, M.E. Díaz García, 'Optical Biosensing Using Enzymes in Organic and Gas Phase', *Ciencia*, **4**(1), 41–46 (1996).
35. J.M. Costa Fernández, M.E. Díaz García, A. Sanz-Medel, 'A Critical Comparison of Different Solid Supports to Develop Room Temperature Phosphorescence Sensing Phases of Air Moisture', *Sensors and Actuators B*, **38–39**, 103–109 (1997).
36. A. Sanz-Medel, 'Solid Surface Photoluminescence and Flow Analysis: A Happy Marriage', *Anal. Chim. Acta*, **283**, 367–378 (1993).
37. K. Kalyanasundaram, F. Grieser, J.K. Thomas, 'Room Temperature Phosphorescence of Aromatic Hydrocarbons in Aqueous Micellar Solutions', *Chem. Phys. Lett.*, **51**, 501–505 (1977).
38. L.J. Cline Love, R. Weinberger, 'Recent and Future Prospects in Fluorescence and Phosphorescence Spectroscopy', *Spectrochim. Acta*, **38B**, 1421–1433 (1983).
39. M.E. Díaz García, A. Sanz Medel, 'Facile Chemical Deoxygenation of Micellar Solutions for Room Temperature Phosphorescence', *Anal. Chem.*, **58**, 1436–1440 (1986).
40. J. Wei Jun, L. Chang Song, 'Study on Five Polycyclic Aromatic Hydrocarbons by Chemical Deoxygenation Micelle-stabilized Room Temperature Phosphorimetry', *Microchem. J.*, **48**, 94–103 (1993).
41. A. Sanz Medel, P.L. Martínez García, M.E. Díaz García, 'Micelle Stabilized Room Temperature Liquid Phosphorimetry of Metal Chelates and its Application to Niobium Determination', *Anal. Chem.*, **59**, 774–778 (1987).
42. M.R. Fernández de la Campa, M.E. Díaz García, A. Sanz-Medel, 'Room Temperature Liquid Phosphorimetry of Aluminium–Ferron Chelate in Micellar Media', *Anal. Chim. Acta*, **212**, 235–243 (1988).
43. Y.M. Liu, M.R. Fernández de la Campa, M.E. Díaz García, A. Sanz Medel, 'Micelle-stabilized Liquid Room Temperature Phosphorimetry for Metal: The Micellar Reaction of Gallium with 7-Iodo-Hydroxyquinoline-5-Sulfonic Acid and its Application to the Metal Determination', *Mikrochim. Acta*, **I**, 53–64 (1991).
44. S. Igarashi, K. Endo, 'Room Temperature Phosphorescence of Palladium (II) Coproporphyrin III Complex in a Precipitated Nonionic Surfactant Micelle Phase: Determination of Traces of Palladium (II)', *Anal. Chim. Acta*, **320**, 133–138 (1996).
45. M. Fernández Roza, M.E. Díaz García, 'Room Temperature Phosphorescent Palladium–Porphine Probe for DNA Determination', *Anal. Chem.*, **69**, 2406–2410 (1997).
46. E.Y. Mantrova, M.V. Demcheva, A.P. Savitsky, 'Universal Phosphorescence Immunoassay', *Anal. Biochem.*, **219**, 109–114 (1994).

47. M.R. Fernández de la Campa, Y.-M. Liu, M.E. Díaz-García, A. Sanz-Medel, 'Fluorescence and Liquid Room Temperature Phosphorescence of the Aluminium–Ferron Chelate in Vesicles', *Anal. Chim. Acta*, **238**, 297–305 (1990).
48. Y.M. Liu, M.R. Fernández de la Campa, M.E. Díaz García, A. Sanz Medel, 'Phosphorescence Detection in Flowing Systems: Selective Determination of Aluminum by Flow-injection Liquid Room Temperature Phosphorimetry', *Mikrochim. Acta*, **234**, 233–238 (1990).
49. J.J. Donkerbroek, N.J.R. van Eikema Hommes, C. Gooijer, N.H. Velthorst, R.W. Frei, 'Sensitised Room Temperature Phosphorescence for Detection in Continuous Flow and Chromatographic Systems', *Chromatographia*, **15**, 218–222 (1982).
50. J.J. Donkerbroek, A.C. Veltkamp, C. Gooijer, N.H. Velthorst, R.W. Frei, 'Quenched Room Temperature Phosphorescence Detection for Flow Injection Analysis and Liquid Chromatography', *Anal. Chem.*, **55**, 1886–1893 (1983).
51. J.D. Winefordner, 'Time-resolved Phosphorimetry', *Acc. Chem. Rec.*, **2**, 361–367 (1969).
52. F. Alava-Moreno, Y.-M. Liu, M.E. Díaz-García, A. Sanz-Medel, 'Kalman Filtering Aided Time-resolved Solid-surface Room Temperature Phosphorimetry for Simultaneous Determination of Tetracyclines in Solution', *Mikrochim. Acta*, **112**, 47–54 (1993).
53. Y.-M. Liu, J.I. García-Alonso, R. Fernández de la Campa, M.E. Díaz-García, A. Sanz-Medel, 'Time-resolved Micelle- stabilized Room-temperature Phosphorimetry for Simultaneous Determination of Gallium and Indium', *Mikrochim. Acta*, **1**, 199–207 (1991).
54. J.J. Aaron, J.D. Winefordner, 'Phosphorimetry, a Spectrochemical Method of Analysis', *Talanta*, **22**, 707–715 (1975).

Surface Measurements using Absorption/Luminescence

Philip B. Oldham and Alexander N. Asanov
Mississippi State University, USA

Vinay M. Rangnekar
MedPharmx, Inc., Pomona, CA, USA

1 Introduction 1
2 Theory of Absorption and Luminescence 2
2.1 Ultraviolet/Visible Absorption 2
2.2 Luminescence 2
3 Analytical Information 3
4 Fabrication of Thin Films 5
4.1 Physical Deposition 5
4.2 Chemically Modified Surfaces 5
5 Experimental 6
5.1 Attenuated Total Reflectance 6
5.2 Total Internal Reflection Fluorescence 8
Acknowledgments 13
Abbreviations and Acronyms 13
Related Articles 13
References 14

The physical and chemical characteristics of surfaces along with the structure and dynamics of molecules near surfaces or interfaces have recently become critically important. Surface measurements have provided important information concerning the properties of the surfaces and the surface/adsorbate interactions. The physical, chemical and biological sciences have benefited tremendously from these studies, with significant developments being made in the areas of diagnostic devices, optical devices, electronic devices and sensors.

The specific behavior of macromolecules at or near surfaces, interfaces, and membranes is currently of primary interest in the biological sciences. Important applications include: adsorption of blood proteins on biomaterials in thrombogenesis research; the binding to and triggering of living cells by hormones, neurotransmitters and antigens; cell adhesion to various surfaces; the mechanism of electron transport in mitochondrial and photosynthetic membranes; and also reaction rate enhancement with membrane receptors by nonspecific adsorption and surface diffusion of ligands. However, investigation of biomolecular behavior at surfaces and interfaces remains a difficult task. Most of the common analytical methods available for investigation of surfaces either lack the extent of surface selectivity required or demand relatively harsh sample handling that severely limits the biological relevance of any results obtained. This article specifically describes ultraviolet/visible (UV/VIS) absorption and photoluminescence techniques applicable to analysis of surfaces or thin films.

1 INTRODUCTION

The physical and chemical characteristics of surfaces along with the structure and dynamics of molecules near surfaces or interfaces have recently become critically important. Surface measurements have provided important information concerning the properties of the surfaces and the surface/adsorbate interactions. The physical, chemical and biological sciences have benefited tremendously from these studies with significant developments being made in the areas of diagnostic devices, optical devices, electronic devices and sensors.

The specific behavior of macromolecules at or near surfaces, interfaces, and membranes is currently of primary interest in the biological sciences. Important applications include: adsorption of blood proteins on biomaterials in thrombogenesis research; the binding to and triggering of living cells by hormones, neurotransmitters and antigens; cell adhesion to various surfaces; the mechanism of electron transport in mitochondrial and photosynthetic membranes; and also reaction rate enhancement with membrane receptors by nonspecific adsorption and surface diffusion of ligands. However, investigation of biomolecular behavior at surfaces and interfaces remains a difficult task. Most of the common analytical methods available for investigation of surfaces either lack the extent of surface selectivity required or demand relatively harsh sample handling that severely limits the biological relevance of any results obtained.

The two general areas of electron and optical spectroscopies are both powerful tools for examining surfaces and studying interfaces. Electron spectroscopy mainly includes techniques such as X-ray photoelectron spectroscopy (XPS), Auger spectroscopy and secondary-ion mass spectrometry (SIMS). The main optical spectroscopy techniques are attenuated total reflectance (ATR), total internal reflection fluorescence (TIRF), total internal reflection Raman (TIRR), surface-enhanced Raman spectroscopy (SERS) and ellipsometry. This article is mainly concerned with the application of ATR and TIRF spectroscopy in the ultraviolet (UV) and visible (VIS) wavelength range for surface studies of molecular species. The most comprehensive treatment of the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

fundamental optics that support both ATR and TIRF can be found in *Internal Reflection Spectroscopy*.[1]

2 THEORY OF ABSORPTION AND LUMINESCENCE

Absorption and luminescence are photophysical processes observed in both atomic and molecular systems. Since excitation of the chemical system by absorption is a prerequisite to any subsequent luminescence, we will begin with a discussion of optical absorbance. Also, the discussion will be limited to molecular systems.

2.1 Ultraviolet/Visible Absorption

The interaction of light with matter can be described by combining the classical and quantum theories of light with the quantum concepts of molecular structure. Light is a form of electromagnetic radiation and exhibits a well-known dual nature in that it behaves both as a classical wave and as a particle. The propagation of light is described as a wave phenomenon and is therefore characterized by a wavelength (λ) and a frequency (ν) which are related as shown in Equation (1)

$$\lambda = \frac{c}{\nu} \tag{1}$$

where c is the speed of light in a vacuum.

Three primary processes are possible whenever light impinges upon matter. The light can either be transmitted without loss, scattered or absorbed by the medium. The last case involves a transfer of energy to the medium composed of molecules characterized by their respective electronic energy configurations. When light is absorbed, the electronic configuration is altered and the molecule is promoted to a higher energy level. Each molecule possesses a series of discrete energy levels. To move from a lower energy level to a higher one, the molecule must absorb an integral quantum of light that is equal to the difference in the two molecular energy levels (Equation 2).

$$\Delta E = E_2 - E_1 \tag{2}$$

A quantum of light is called a ''photon'' and is described by Equation (3)

$$E = h\nu = \frac{hc}{\lambda} \tag{3}$$

where h is Planck's constant. Therefore, the electronic transitions associated with the absorption of light can be characterized by the wavelength of light absorbed. The energy necessary to produce electronic transitions in most molecules corresponds to the 200–800 nm wavelength region, or the UV (200–350 nm) and VIS (350–800 nm) ranges. Particular structural features that lead to absorption are called ''chromophores''. A more thorough discussion of structural effects on absorption can be found in the literature.[2]

2.2 Luminescence

Although there are numerous types of luminescence, for the purposes of this article we are primarily concerned with two: fluorescence and phosphorescence. Of these two, fluorescence is by far the most popular for analytical methodologies. The analytical sensitivity and selectivity of molecular fluorescence is confirmed by the number and diversity of analytical applications published each year.[3] Molecular fluorescence has proven especially valuable for the analysis of biological samples in which many of the molecules of interest exhibit fluorescence. The nature of fluorescence permits the examination of intracellular mechanisms and the detection of biologically active compounds present at very low concentrations.

The processes of molecular excitation and relaxation are illustrated graphically in Figure 1. The electronic energy levels represented by S_0, S_1 and S_2 are the ground, first excited and second excited singlet states, respectively. The excited triplet state is represented by T_1. A number of more closely spaced vibrational energy levels are associated with each electronic energy level. As shown in Equation (4), the excitation of a molecule to an excited singlet can occur owing to the absorption of a sufficiently energetic photon ($h\nu_a$, where h is Planck's constant and ν_a is the frequency of the absorbed photon).

$$S_0 + h\nu_a \longrightarrow S_1 \tag{4}$$

Excitation can originate from any of the ground state vibration levels. However, at room temperature most of the molecules populate the lowest vibrational level of the ground state. Therefore, the energy needed to promote a molecule to the S_1 state corresponds to the energy differences between the first vibrational level of S_0 and any of the various vibrational levels of S_1. This is a rapid process occurring within 10^{-15} s.

There are numerous mechanisms of relaxation available to excited-state molecules. They are often classified as either radiative or nonradiative depending on whether the excess energy is released as a photon or as some other form of energy. For most fluorescent molecules the various nonradiative processes generally combine to relax the excited state molecule down to the lowest vibrational level of S_1. Once the excited molecule relaxes to S_1, ν_0, the molecule can reach the ground state, S_0, by releasing a photon. This radiative transition between states of the same multiplicity is called fluorescence. Deexcitation to

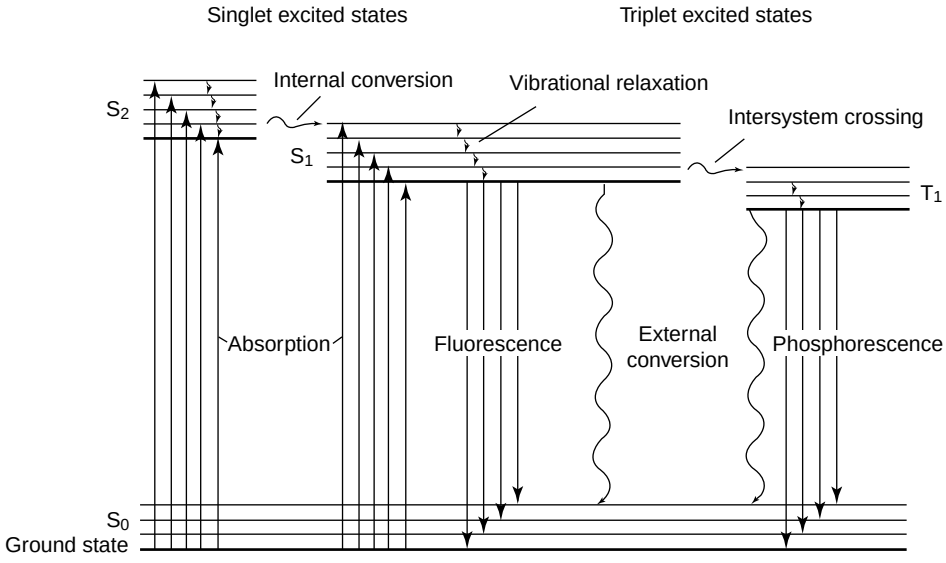

Figure 1 Jablonski energy diagram.

any of the vibrational levels of S_0 is possible and usually occurs in 10^{-9}–10^{-6} s. This, of course, is not possible in molecules where a more competitive (faster) mechanism of relaxation is available. The direct excitation from the lowest vibrational level of S_0 to the lowest vibrational level of S_1 and the subsequent relaxation to the lowest vibrational level of S_0 is called the "0–0" transition. This transition is seldom observed except in the gaseous or crystalline state. In solution, the photon $h\nu_f$ (Equation 5) emitted as fluorescence is generally of longer wavelength (less energy) than the photon $h\nu_a$ which is absorbed (Equation 4).

$$S_1, \nu_0 \longrightarrow S_0, \nu_x + h\nu_f \tag{5}$$

Following fluorescence, further relaxation to S_0, ν_0 is possible by vibrational relaxation. The energy difference between the absorbed photon and the emitted photon is called the Stokes shift and is observed owing to the fact that fluorescence generally occurs from the lowest vibrational level of S_1 to some upper vibrational level of S_0. Furthermore, fluorophores can lose excitation energy owing to solvent effects in solution.

It is also possible for a molecule to undergo intersystem crossing from S_1 to T_1, provided that the respective vibrational levels of the two electronic states are energetically similar. Once this happens the molecule is free to relax to the lowest vibrational level of T_1. At this point, a radiative deactivation is once again competitive (Equation 6).

$$T_1, \nu_0 \longrightarrow S_0, \nu_x + h\nu_p \tag{6}$$

This is called phosphorescence and differs from fluorescence emission in two distinct ways. First, T_1 is generally lower in energy than S_1 so that phosphorescence emission is longer in wavelength than fluorescence. Also, since the occurrence of a radiative transition between states of different multiplicity has a low probability, the lifetime of phosphorescence (10^{-4}–10 s) is significantly longer than that of fluorescence. For this reason phosphorescence is more susceptible to quenching than fluorescence and is seldom observed in solution at room temperature.

3 ANALYTICAL INFORMATION

The total energy of the molecule is the sum of electronic, vibrational and rotational energies such that for each electronic state several vibrational and rotational sublevels are possible. This results in relatively broad absorbance spectra and minimal qualitative information content useful for sample identification purposes. However, quantitative applications are well documented based on the relationship commonly called Beer's Law (Equation 7).

$$A = -\log T = abc \tag{7}$$

In this equation, A is absorbance, T is transmittance, a is the absorptivity constant, b is optical pathlength, and c is sample concentration. Linear behavior is generally observed at concentrations below 0.01 M. Assuming a good absorber, a typical commercially available absorbance spectrophotometer can provide detection limits in the low nanomolar range. A typical commercial double-beam absorbance spectrophotometer is shown in Figure 2.

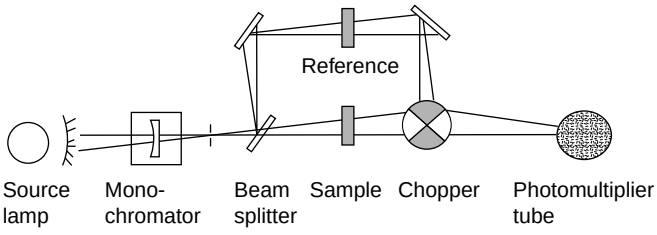

Figure 2 Double-beam absorbance spectrophotometer.

A detailed evaluation of the kinetics of fluorescence provides several analytically useful quantities that are necessary for the derivation of the quantitative aspects of the fluorescence measurement. Fluorescence is a first-order process with the rate of fluorescence depending on the concentration of excited molecules in S_1, as shown in Equation (8).

$$\frac{d(h\nu_f)}{dt} = k_f[S_1] \tag{8}$$

The rate of production of $[S_1]$ is equal to the rate of light absorption, I_a, as shown in Equation (9).

$$\frac{d[S_1]}{dt} = I_a \tag{9}$$

If illumination is maintained over several fluorescence lifetimes, a steady-state relationship between the rate of absorption, I_a, and the rate of fluorescence, $k_f([S_1])$, is achieved (Equation 10).

$$I_a = k_f[S_1] \tag{10}$$

However, Equation (10) is valid only in cases where there are no other processes competing with fluorescence for the deactivation of the excited state. A more realistic approach, shown in Equation (11), considers all of the possible relation processes.

$$I_a = (k_f + k_{IC} + k_{ISC} + k_Q[Q])([S_1]) \tag{11}$$

Each k represents the first-order rate constant for the respective processes described previously in Figure 1. If Equation (11) is rearranged and substituted for $[S_1]$ in Equation (8) it yields the relationship shown in Equation (12), in which the rate, or intensity, of fluorescence is equal to the rate of absorption multiplied by the quantum efficiency, ϕ_f.

$$I_f = \frac{d(h\nu)_f}{dt} = I_a\phi_f \tag{12}$$

The quantum efficiency, ϕ_f, is the ratio of the rate constant for fluorescence without any competition and with competition (Equation 13).

$$\phi_f = \frac{k_f}{k_f + k_{IC} + k_{ISC} + k_Q[Q]} \tag{13}$$

In simpler terms, ϕ_f is given by Equation (14).

$$\phi_f = \frac{\text{Number of photons emitted}}{\text{Number of photons absorbed}} \tag{14}$$

One divided by the denominator in Equation (13) is called the natural fluorescence lifetime, t, and is the time required for the fluorescence intensity to decay to 1/e times its initial value.

The preceding discussion (Equation 12) demonstrated that the intensity of fluorescence I_f is equal to the rate of absorption I_a multiplied by the fluorescence quantum efficiency ϕ_f. According to the Beer–Lambert law the amount of light absorbed by matter can be described by the differential Equation (15)

$$-\frac{dI}{dx} = kIc \tag{15}$$

where I is the intensity of light, x is the distance through the medium which it passes, k is a first-order rate constant and c is the concentration of the absorbing species. When this is integrated over a distance, b, the popular form of absorbance Equation (16) is produced.

$$A = \log(I_T I_0) = 2.3abc \tag{16}$$

In this form, A is the absorbance, I_T is the transmitted light, I_0 is the incident light, and a is the absorptivity. The amount of the incident light that is absorbed by the fluorescent sample is given by Equation (17).

$$I_T - I_0 = I_0 = I_0(1 - 10^{-2.3abc}) \tag{17}$$

This can be substituted for I_a in Equation (12) and expanded as a Taylor series to remove the exponential term. For dilute solutions in which the total absorbance is low (<0.01 AU), the Taylor expansion can be truncated. This produces a linear relationship between I_f and the fluorophore concentration c, as shown in Equation (18).

$$I_f = 2.3\, I_0\, \phi_f\, abc \tag{18}$$

At higher absorbances, the equation is no longer linear owing to a significant attenuation of the incident light as it passes through the sample.

The sensitivity of a given technique is defined as the amount of signal produced per unit concentration (or weight) of the sample. This is determined by the slope of a standard calibration curve. Equation (18) indicates the slope of a fluorescence calibration curve to be equal to $2.3\, I_0\, \phi_f\, ab$. Therefore, the sensitivity of the fluorescence measurement can be regulated by adjusting the intensity

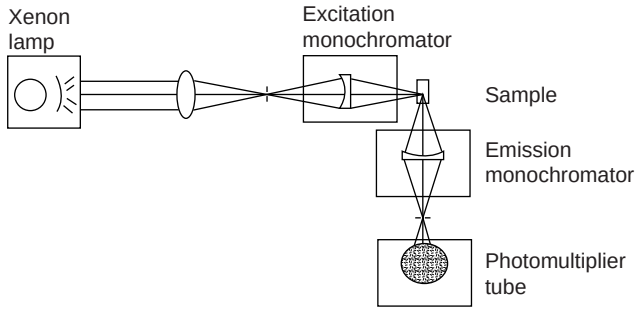

Figure 3 Fluorescence spectrophotometer.

of the exciting light (I_0). This provides experimental versatility with fluorescence that is unavailable from absorption spectroscopy.

In the literature, sensitivity is often confused with detection limit. The detection limit is the minimum detectable quantity of an analyte. This value is more often determined by background interferences and detector noise levels than by the sensitivity. The low detection limits obtainable by fluorescence are well documented[4,5] in the literature and are often 1000 times below those of UV/VIS absorption. In fact, single molecule detection by fluorescence has been well documented in the literature and is now becoming somewhat routine.[6]

Fluorescence measurements are much more selective than most other analytical techniques. Since not all molecules fluoresce, the mere detection of fluorescence is selective. Also, since fluorescence generally originates from S_1, ν_0, the fluorescence spectrum of a pure one-component sample is independent of the excitation spectrum. No matter what excitation wavelength is selected, the shape of the emission spectrum will be the same. This fact provides the theoretical basis for the analysis of multicomponent samples. An optical diagram of a typical fluorescence spectrophotometer is shown in Figure 3.

4 FABRICATION OF THIN FILMS

4.1 Physical Deposition

The two main methods for producing thin, inorganic, surface coating films are vacuum sputtering and chemical vapor deposition (CVD). These techniques typically give reproducible high-quality films whose properties can be controlled within a certain range by variation of sputtering or chemical deposition parameters. Different modifications of vacuum sputtering and CVD are widely used in the optical industry to manufacture metal, metal oxide and metal halide coatings.

Thin polymer layers on solid supports are prepared by casting from polymer solution. Spinning of the substrate more often provides uniform thickness of the film. To manufacture large volumes of surface-supported films, the method of casting onto a linearly moving substrate is applied. To provide better adhesion of the polymer film, the substrate surface can be chemically modified. For example, simple treatment of glass with siloxanes dramatically enhances the adhesion of many polymers to originally hydrophilic glass surfaces.

Another powerful technique for thin organic layer preparation involves the Langmuir–Blodgett method of monolayer formation at the air–water interface. In principle, this technique can be applied for fabrication of thin films of a vast variety of organic compounds. However, the Langmuir–Blodgett method works better for amphiphilic molecules which comprise both hydrophilic and hydrophobic parts. Surface density and orientation of amphiphilic molecules can be controlled using a Langmuir–Blodgett balance. Since 1990, a renaissance of research on molecular organization stimulated the development of sophisticated Langmuir–Blodgett balances with capabilities of automated well-controlled deposition of organic monolayers and multilayers. A very complex multilayer molecular architecture of a sandwich type, which consists of a sequence of different compound monolayers, can be built at a solid surface.

4.2 Chemically Modified Surfaces

Surface chemistry is one of the parameters that determines the main features of the interfacial behavior of an adsorbate. The surface of a substrate can be modified by casting a thin polymer film, by vacuum deposition or CVD of an inorganic film, or by irreversible adsorption of macromolecules such as synthetic and natural biopolymers. Chemical modification of a thin monomolecular layer of a substrate is one of the most efficient ways to obtain a well-characterized solid surface with the desired properties. Numerous methods for chemical modification of inorganic and organic surfaces have been reported. This section describes selected methods of chemical modification of inorganic substrates such as glass, quartz and metal oxides, and focuses on procedures designed for covalent attachment of biological molecules.

Silane chemistry is one of the most widely used technologies to covalently attach a reactive group to the surface of an inorganic substrate. Different derivatives of trialkoxysilane compounds tipped with a variety of reactive groups are available commercially and can be used as coupling agents for covalent binding of proteins, DNA and other biomolecules of interest. These reagents bind to oxide surfaces via silane bonds. The opposite, free end of the silane reagent, that can carry such groups

as $-NH_2$, $-COOH$, or other reactive groups, is used to functionalize the surface by subsequent coupling of biomolecules. Triethoxysilanes are the most widely used compounds that react with the free hydroxyl groups of metal oxides by S_N2 exchange of the silane ethoxy groups for the silanol oxygen, with loss of ethanol. The reaction is catalyzed in the presence of a base. If the triethoxysilane derivative contains a primary amine group, the reaction is self-catalyzed, making external catalysis by a base unnecessary. In the catalyzed reaction, the base substitutes one ethoxy group of triethoxysilane, and a pentacoordinate silane intermediate is formed when a surface hydroxyl group replaces the base.

The purpose of the chemical modification is to obtain a monolayer or a submonolayer coating with active reactive sites at the tip of the attached molecule. However, in the case of the silanization, a multilayer silane film tends to occur instead of a mono- or submonolayer coating. This disadvantage can be avoided by proper selection of the silanization procedure. The silanization reaction dramatically depends on reaction conditions, and post-reaction ''curing''. Since the formation of siloxane bonds is possible between two surface-bound molecules (horizontal polymerization) and between triethoxysilane molecules in solution and at the surface, multilayer siloxane films tend to be obtained at excessive concentrations and time of reaction. A monolayer siloxane coating is formed during several minutes of exposure of a glass or silica surface to a trialkoxysilane solution in a dry solvent such as toluene or xylene at room temperature. Thus, a monolayer coating can be obtained by reducing the reaction time and/or concentration of the reagent.

Silanization procedures utilizing aqueous solutions of trialkoxysilanes have also been reported. However, these coatings typically exhibit poor stability, and functionalized coatings can be desorbed by washing. One reason for this behavior is related to the electrostatic interactions between oxide surfaces and triethoxysilane compounds containing amines. Since glass and silica surfaces are negatively charged over a wide pH range, aminated trialkoxysilane derivatives tend to be adsorbed owing to electrostatic interactions between positively charged amine groups and a negatively charged surface. Subsequent curing of the surface with an adsorbed amine derivative of triethoxysilane at 100–120 °C is required to stimulate transformation of the adsorbed compound to covalent binding via a silane bond. The functional activity of immobilized biomolecules depends on the chemistry and the length of spacer units between the triethoxysilane groups and the reactive group involved in the coupling of biomolecules. Relatively long spacer groups of 4–8 carbon atoms inserted between the reactive group ($-NH_2$, $-COOH$) and the surface-attaching triethoxysilane group allow better access of dissolved counterparts to the surface-attached functional group. In the case of short nonaliphatic spacers, surface-immobilized biomolecules frequently demonstrate reduction of their functional activity, which deteriorates significantly with storage time.

5 EXPERIMENTAL

5.1 Attenuated Total Reflectance

The total internal reflection (TIR) phenomenon can be explained with the aid of Figures 4 and 5. ATR is a term used with absorbance spectroscopy based on a common optical phenomenon called TIR. Application of TIR for ATR measurements has been most popular for vibrational absorbance with Fourier transform infrared (FTIR) spectroscopy. However, the fundamental principles are the same for ATR with the electronic absorbance applications of interest in this article. The principles of TIR were initially described by Newton in the early 18th century and are well documented in the more recent literature.[1,7,8] In brief, when a beam of light propagating within a medium of refractive index (η_1) encounters an interface with a medium of lower refractive index (η_2), it can undergo TIR for incidence angles (θ_i) greater than the critical angle (θ_c). Although the incident light totally reflects at the interface, a portion of the electromagnetic radiation penetrates the interface into the less dense medium. The intensity of this interfacial field, typically called the ''evanescent wave'', decays exponentially with distance from the interface (Equation 19).

$$\theta_c = \sin^{-1}\left(\frac{\eta_2}{\eta_1}\right. \tag{19}$$

The penetration depth (d_p) of the evanescent wave in the less dense medium is a function of incidence angle, refractive index ratio and incident light wavelength, λ_i (Equation 20). The penetration depth (pathlength) of the evanescent wave can be conveniently altered by changing incidence angles. The extremely short pathlength of the evanescent wave (in the order of the wavelength of light) excites a very small sample volume and thereby minimizes background absorption effects.

$$d_p = \frac{\lambda_i}{2\pi\eta_1\sqrt{\sin^2\Theta_i - (\eta_2/\eta_1)^2}} \tag{20}$$

Although light is efficiently propagated through the optical waveguide, the small proportion of energy present in the evanescent wave can be absorbed by surface components according to Beer's law. The total light loss in the waveguide is the sum of losses due to scattering from the bulk sample and surface, respectively, and from surface absorbance. With appropriate experimental

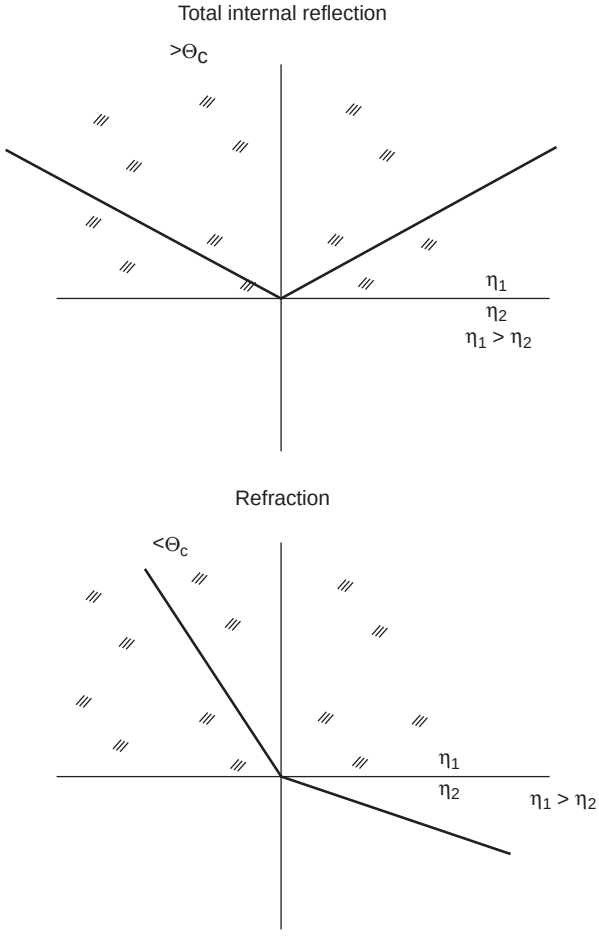

Figure 4 Conditions for internal reflection and refraction.

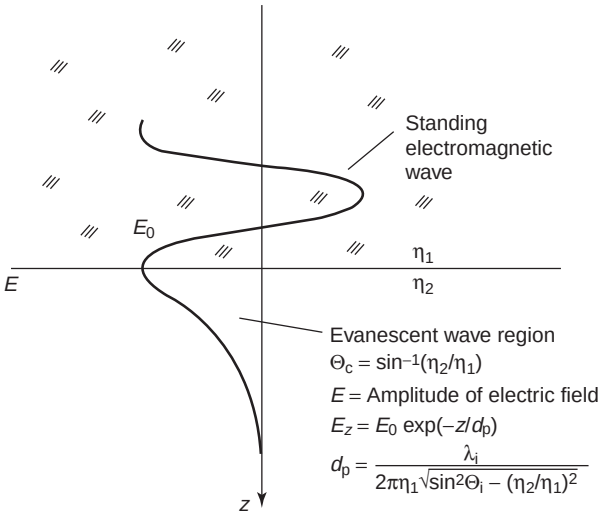

Figure 5 Graph of electric field amplitude and evanescent wave.

care,[9] the attenuation of the evanescent wave by absorbance can be accurately calibrated for a quantitative measure of components contained within the penetration depth of the evanescent field.

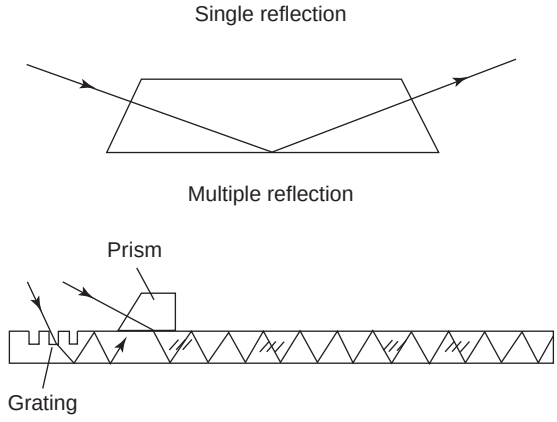

Figure 6 Generalized optical scheme for single and multiple reflection waveguides.

5.1.1 Instrumentation

It is difficult to describe any general instrumentation scheme for waveguide detection because it can take many different forms depending on the particular application (Figure 6). It has been properly noted that "the true genius behind waveguide detectors lies in the reagent chemistry and geometry".[10] However, in general, the typical ATR waveguide device must provide for multiple reflections owing to the relatively poor sensitivity of absorbance measurements with extremely short pathlengths. The exact number of reflections can vary from $<10\,cm^{-1}$ with planar slab waveguides to several thousand per centimeter with integrated optical waveguides (IOW). However, it should be noted that, to date, IOWs have not become popularly applied owing to fabrication problems, high scattering losses and fluorescence background problems.[11]

Associated instrumentation must provide for efficient optical coupling of light into the waveguide which is typically accomplished by either a prism or a grating coupler. Depending on the design of the waveguide, either broadband or monochromatic light can be used. Efficient detection in ATR also requires effective output coupling of the propagated beam into the appropriate detection device. Use of most of the common optical detectors (e.g. photodiodes, photomultiplier tubes, charge-coupled devices (CCDs)) has been reported in the literature in connection with ATR waveguide devices.

5.1.2 Applications

Initial applications of ATR for electronic spectroscopy of surfaces were primarily interested in increasing detection

sensitivity for surface adsorbates as compared with bulk transmission measurements. This was accomplished by use of multiple internal reflection waveguides which effectively increase the optical pathlength for the absorbance measurement while containing the optical beam parallel to the analytical surface. For example, Stephens and Bohn[9] observed better than a 1000-fold enhancement in the absorbance detection of a chemisorbed *p*-nitrobenzoic acid monolayer on a 150 μm glass waveguide compared with a traditional transmission experiment. They were also able to collect the absorbance spectrum of a 130 nm thick physisorbed film of phthalocyanine. Numerous other literature reports have cited similar sensitivity enhancements.

Closely paralleling those initial applications of ATR there appeared a large number of ATR-based sensor devices for various environmental and biological applications. For example, Yang and Saavedra[12] reported development of a sol–gel-derived planar IOW for Pb(II) and pH. The sol–gel layer served as a porous yet stable support for the indicator phase. The thin (<200 nm) sol–gel layer provided rapid response times. Several other similar sensors using various waveguide geometries have been reviewed by Dessy.[10]

Several reports of the combination of waveguide optics with electrochemical techniques have also appeared in the literature.[13–18] Dunphy et al.[13] demonstrated a 10^4 signal enhancement during the reduction of a submonolayer of surface-adsorbed methylene blue with a single-mode IOW coated with indium tin oxide as compared with a traditional transmission experiment. They estimated detectabilities of less than 1% of a monolayer for most cases. It has been noted that a fundamental problem that has limited popular use of IOW/ATR devices concerns the difficulty in detecting the broadband spectrum, since both the mode propagation angle and the coupling angle are functions of wavelength. Therefore, the IOW is typically only efficient over a narrow wavelength band. Mendes et al.[19] described a creative IOW design which combines a broadband Xe lamp with a prism input coupler and a series of two integrated gratings to compensate for wavelength dispersion in the IOW. A third output grating is then used to couple the dispersed ATR spectrum onto a CCD detector to provide simultaneous detection of the entire absorbance spectrum.

Some more recent reports[20,21] have described the use of IOW/ATR for measurement of absorption linear dichroism. The combination of absorption linear dichroism with emission anisotropy measurements allows molecular orientation of surface species to be determined. This is a difficult question to resolve but highly relevant to many areas of biotechnology. Edminston et al.[20] looked at the molecular orientation of cytochrome *c* on a variety of surface types, and Wood et al.[21] studied the orientation of site-directed immobilization of cytochrome *c* on thiol-based self-assembled monolayers.

5.2 Total Internal Reflection Fluorescence

Hirschfeld[22] first introduced TIRF in 1965, to examine solid/liquid interfaces. Since then many researchers have applied TIRF to examine various molecules at solid/solid, solid/liquid and liquid/liquid interfaces. TIRF is a surface-selective technique which is capable of detecting fluorescent probe molecules within a few hundred nanometers of the interface. The fluorescent probe molecules assist in investigating the interfacial chemical structure and environment. The probe molecules provide information about their surroundings by their spectral response, including spectral wavelength shifts, excited state lifetime and emission polarization.

TIRF spectroscopy has proven to be a very powerful and versatile technique for the study of surface and/or interfacial behavior of biological molecules and their aggregates.[23,24] TIRF has been successfully applied to numerous studies associated with solute adsorption, orientation and rotational mobility associated with conformational changes. Surface selectivity is achieved in TIRF by detecting only the evanescent-wave-excited fluorescence signals which originate within approximately the first 100 nm from the waveguide surface. This exceptionally short optical pathlength allows investigation of surface behavior even in the presence of highly concentrated solutions. In short, TIRF provides in situ, real-time, nondestructive, and highly sensitive detection suitable for studies on expensive biological materials available only in microliter quantities. Reliable detection of a few macromolecules per square micrometer has been documented with the utilization of commercial fluorescence spectrophotometers,[25] while custom-designed systems allowed single-molecule detection.[26]

The principles explained in section 5.1 for ATR apply equally to TIRF in respect to generation and behavior of the evanescent wave at the optically active surface. It is also important to understand the polarization properties of the evanescent wave, particularly for studies involving molecular orientation at the surface. The polarization properties of the transmitted field (evanescent wave) depend on the polarization properties of the incident light.[27] For instance, if the incident electric field is s-polarized its electric field vector is perpendicular to the plane of incidence and the transmitted field has the same polarization as the incident field. However, if the incident electric field is p-polarized its electric field vector lies in the plane of incidence and the transmitted field is elliptically polarized. As a result, the s-polarization case is amenable to fluorescence anisotropy measurements

of the interfacial region, whereas the p-polarization is considerably more complex. Additional properties of the evanescent wave and the underlying principles of TIRF are available in the literature.[1,27–29] The evanescent wave is primarily responsible for the electronic excitation of the fluorophore present in the lower refractive index medium. The intensity of the fluorescence excited (Equation 21) in the evanescent volume is the integrated product of the intensity of the evanescent wave and the fluorophore concentration (C_z).[28] In Equation (21), k_{ex} is a proportionality constant.

$$I_{ex} = k_{ex}\varepsilon\phi_f I_i \int_0^\infty C_z e^{-z/d_p}\, dz \qquad (21)$$

An important aspect of TIRF involves the reverse coupling of the emitted fluorescence back across the interface into the optical waveguide. The observed fluorescence depends on the observation angle (θ_o). By the principle of reciprocity, the effect of the observation angle on this reverse coupling is similar to the effect of the incidence angle on the evanescent wave penetration depth. Therefore, at θ_o greater than θ_{co} the fluorescence intensity contained in the evanescent observation volume can be described by Equation (22).

$$I_{em} = k_{em} \int_0^\infty e^{-z/d_p}\, dz \qquad (22)$$

θ_{co} differs from θ_c (Equation 19) owing to the dependence of the refractive index on the fluorescence wavelength, λ_f. In Equation (22), k_{em} is a proportionality constant and d_p is calculated using Equation (20) with λ_i and θ_i replaced by λ_f and θ_o, respectively. Subsequently, the TIRF intensity reaching the detector (for fixed $\theta_i > \theta_{ci}$ and fixed $\theta_o > \theta_{co}$) is the product of the two integrals in Equations (21) and (22) and can be represented by Equation (23), which is also referred to as a Laplace transform of the concentration–depth profile.

$$I_f = k_{ex}k_{em}\varepsilon\phi I_i \int_0^\infty C_z e^{-z/(d_{pi}+d_{po})}\, dz \qquad (23)$$

The TIRF signal can also be measured at a fixed observation angle for varying incidence angles which range from normal to grazing. Reciprocally, TIRF measurements can be made at a fixed incidence angle for varying observation angles which range from normal to grazing. (Both the incidence and observation angles are defined with respect to a normal to the interface.) This variable-angle technique is called variable-angle total internal reflection fluorescence (VATIRF). Generally, VATIRF data are represented as the variable-angle fluorescence intensity data curve. The data curve is a convolution of the fluorophore concentration–depth profile and the decaying evanescent wave. The complete concentration profile can be potentially extracted from the data by performing an inverse Laplace transform of Equation (23).

5.2.1 Instrumentation

Since the introduction of TIRF in the 1960s, waveguides and the associated instrumentation have generally been home-built and designed for specific applications. In 1995, the first general-purpose commercial TIRF cell was introduced. The design of the TIRF flow cell made the normally difficult TIRF experiment become routine. Figure 7 illustrates the basic optical design. The TIRF flow cell was prealigned for a specific commercial fluorometer and supplied with an easily assembled flow system to study kinetics of various surface interactions. This design provides high reproducibility of TIRF measurements by ensuring the exact positioning of optical elements against the excitation beam and emission axis. The TIRF cell differed from similar noncommercial TIRF cells

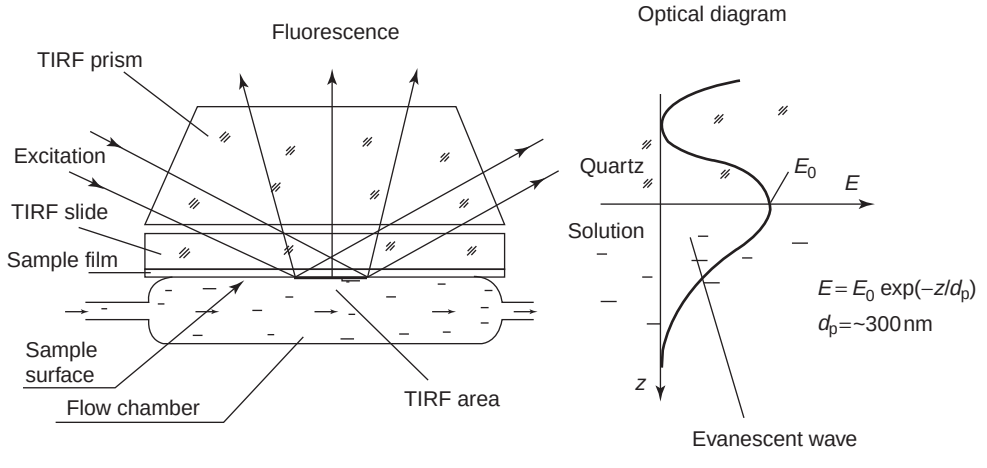

Figure 7 Commercial version of TIRF cell.

in the simple and fast assembly of the sandwich cell. The transparent gasket which forms the flow chamber (approx. 20 μl) and the transparent back plate facilitate easy visualization of the surface and allow acquisition of microscopic pictures by a long-focus objective.

The cell was equipped with a UV-quartz dovetail prism and optically coupled cover slide, as shown in Figure 7. The cover slide provides an easily interchangeable working surface and minimizes wear on the TIRF prism. Cover slides with hydrophilic, hydrophobic and electroconductive thin films, or specific soluble protein docking films could be obtained depending on the application. The TIRF cell could be used with most of the popular, commercially available research-grade spectrofluorometers.

5.2.2 Applications

Although an extensive review of these areas is beyond the scope of this article, the following sections will attempt to highlight some particular applications in the areas of protein adsorption, polymer adsorption, biosensors and chromatographic surfaces by way of example.

5.2.2.1 Protein Adsorption Using TIRF, protein adsorption at interfaces has been examined by intrinsic as well as extrinsic probes. External and internal standards have been used to quantitate the TIRF signal. Harrick and Leob[30] examined the adsorption of dansyl-labeled bovine serum albumin (BSA) onto sapphire and quartz surfaces. Adsorbed films were produced by immersing the internal reflection element (quartz or sapphire) into the solution of labeled protein. Multiple internal reflection elements with different geometries, a hemicylindrical prism and plates with 45° and 60° bevel angles with and without cavities, were used in the study. Van Wagenen et al.[31] used the intrinsic fluorescence of proteins to monitor the adsorption of unlabeled proteins on quartz surfaces. This intrinsic method avoids the need to label the protein thus eliminating the possibility of altering the physicochemical properties of the protein. Rockhold et al.[32] developed a quantitative method to determine the amount of protein adsorbed on hydrophilic glass and quartz surfaces. The protein adsorption studies were performed by measuring the intrinsic protein fluorescence as well as extrinsic fluorescence from fluorescein-labeled protein.

Hlady et al.[33–36] studied protein adsorption on non-gradient and gradient surfaces. They examined the adsorption of BSA and human immunoglobulin at non-gradient and gradient silica/electrolyte interfaces.[33,34] In a separate study they compared the adsorption of human high-density lipoproteins and low-density lipoproteins on hydrophilic and hydrophobic surfaces, respectively. The initial stage lipoprotein adsorption was of most interest. The adsorption of the two lipoproteins onto the two surfaces was found to be different. The lipoprotein adsorption on the hydrophilic surface was described quantitatively using a transport-limited adsorption model and assuming that the lipoprotein fluorescence quantum yield decreased upon adsorption.

TIRF spectroscopy was employed by Golander et al.[37] to quantify the adsorption of fluorescein-labeled lysozyme on bare mica and mica hydrophobized with a monolayer of 50% eicosyl amine ($C_{20}H_{41}NH_2$) and 50% eicosyl alcohol ($C_{20}H_{41}OH$). The steady-state binding of a fluorogenic lysozyme substrate to an adsorbed lysozyme layer, and subsequent steady-state enzymatic cleavage of the substrate were also quantitatively studied using a TIRF cell. The relative binding and cleavage rates were found to be larger on the hydrophobic mica than on the bare mica. Based on the lysozyme structure and on measurements of enzyme activity, possible models were proposed for the orientations of lysozyme on bare and hydrophobic mica.

Lok et al.[38] examined the adsorption of proteins (BSA) from flowing solutions on polymer (cross-linked polymethylsiloxane) films. The results of the study showed that at high solution concentrations, the evanescent-wave-excited bulk solution fluorescence and fluorescence excited by scattered light contributed a significant fraction of the total fluorescence signal. The contribution of surface-adsorbed protein to the total fluorescence was determined by exchanging the labeled protein solution in the flow cell with an unlabeled solution. The observed initial fast decrease in fluorescence intensity on exchange with unlabeled solution was attributed entirely to the displacement of bulk-solution-labeled protein, and not to the removal of loosely adsorbed protein.

TIRF spectroscopy cannot independently provide information about the exchange kinetics of adsorbed protein molecules in chemical equilibrium with the bulk protein solution. To access the kinetics of exchange, a combination of TIRF with fluorescence recovery after photobleaching (FPR) or fluorescence correlation spectroscopy (FCS) was introduced by Thompson et al.[39] A theoretical basis was developed for measuring equilibrium adsorption/desorption kinetics and surface diffusion of fluorescent-labeled solute molecules at solid surfaces. In the TIRF/FPR technique, adsorbed molecules are irreversibly photobleached by a flash from a totally internally reflected laser beam. Subsequent fluorescence recovery is monitored by an evanescent field, as bleached molecules exchange with unbleached molecules from solution and the surrounding nonilluminated regions of the surface. In the TIRF/FCS technique, the evanescent intensity is maintained at a constant and fairly dim level all through the experiment. The spontaneous fluorescence fluctuations due to individual molecules entering and leaving a

well-defined portion of the evanescent field are autocorrelated. Under appropriate experimental conditions, the rate constants and diffusion coefficient can be obtained from TIRF/FPR and TIRF/FCS curves.

Burghardt and Axelrod[40] applied the TIRF/FPR technique to study the dynamics of BSA adsorbed at a quartz glass/aqueous buffer interface. The technique allowed the measurements of desorption rates of rhodamine-labeled BSA and the detection of surface diffusion on a system in equilibrium without the need for any intrinsic spectroscopic change between the bound and unbound states. The data obtained revealed the presence of both irreversibly bound states and a multiplicity of reversibly bound states. The relative amount of reversible to irreversible adsorption was found to increase with increasing bulk protein concentration.

Burghardt and Axelrod[41] studied the conformational and dynamical properties of adsorbed proteins by combining TIRF with conventional fluorescence spectroscopic examination of singlet–singlet energy transfer and fluorescence polarization. They constructed a TIRF apparatus which easily fitted inside the sample chamber of a commercial spectrofluorometer. BSA was multiply labeled with two donor–acceptor pairs of 4-chloro-7-nitro-2,1,3-benzoxadiazole–rhodamine and dansyl–eosin. The labeled BSA molecules were adsorbed at a quartz glass/aqueous buffer interface. The adsorbed protein was in equilibrium with the bulk solution. The spectral data exhibited significantly less effective singlet–singlet transfer from the donor to the acceptor in surface-adsorbed BSA relative to bulk-solution-dissolved BSA. With certain assumptions, this energy transfer change was interpreted as a conformational change of BSA upon adsorption.

Morrison and Weber[42] employed TIRF polarization measurements to selectively examine fluorescent molecules adsorbed to the interface region between two immiscible liquids. A fluorescent compound bis-ANS (4,4′-bis-1-phenylamino-8-naphthalenesulfonate), was examined at a decalin/water interface. The adsorbed fluorophore displayed an apparently hindered rotation in the plane of the interface with rotational diffusion coefficients 3- to 12-fold lower than that expected for bis-ANS in solution. As a better approximation of a biological membrane, bis-ANS binding to the decalin/water interface was examined in the presence of a nonfluorescent surfactant, cetyltrimethylammonium bromide (CTAB). The addition of CTAB increased the affinity of bis-ANS for the interface. Also the addition of BSA was found to cause desorption of bis-ANS from the interface owing to competitive binding.

5.2.2.2 Polymer Adsorption Masuhara et al.[43] examined the interfacial characteristics of a polymer film on a sapphire substrate by time-resolved TIRF spectroscopy. In the study polystyrene films were doped with fluorescent *p*-bis(2-(5-phenylene-oxazolyl)benzene). Fluorescent decay curves were investigated as a function of incident angle. Monitoring the fluorescence helped to elucidate the photophysical processes and molecular motions in the nanosecond to picosecond time domains. As a further improvement of the technique, Itaya et al. performed the TIRF measurement under a microscope.[44] This enabled them to obtain time- and space-resolved data on fluorescence characteristics.

Itaya et al.[45] studied poly(methyl methacrylate) (PMMA) films doped with pyrene by TIRF. They examined the distribution of the dopant molecules in the polymer film and the microenvironment around the dopant molecules. The results were compared with fluorescence data obtained under normal excitation conditions. In these studies sapphire was used as an internal reflection element. Polymer films were both cast and film-coated onto the sapphire substrate. The concentration of the dopant molecule (pyrene) was varied in the experiments. The results indicated that micropolarity around pyrene molecules located at or near the sapphire/polymer film interface was higher than the bulk. Based on the pyrene monomer and excimer fluorescence data they concluded that the aggregate state of pyrene was significantly different at the interface and the bulk.

5.2.2.3 Biosensors The development of convenient surface chemistries suitable for the attachment of biospecific molecules to waveguide surfaces has provided tremendous opportunities for the design of very sensitive and selective biosensor devices. The development and optimization of such devices is highly dependent on a thorough understanding of the molecular dynamics at the surface or interface. Both steady-state and lifetime-based TIRF techniques have been very useful for investigation of surface and interfacial behavior. An excellent review[46] of lifetime-based TIRF applications for biosensors appeared in 1997.

Most of the biosensor development activity so far has centered on immunoassay reaction chemistries due to their inherent biospecificity and commercial availability. A large number and variety of TIRF biosensors has been reported. A thorough review[47] appeared in 1991. Plowman et al.[48] were the first to demonstrate femtomolar sensitivity for a model immunoassay system. This system consisted of a dual channel, 1-μm thick, siliconoxynitride IOW deposited on a 10-cm quartz wafer. A He–Ne laser was used for excitation and the emitted fluorescence was dispersed with a single grating spectrograph onto a CCD camera for simultaneous detection of both channels. This arrangement conveniently provided comparable sample

and reference spectra. More channels could be added for multianalyte applications.

One problem associated with TIRF biosensors is that a sensor with adequate sensitivity generally requires a biospecific interaction with high binding affinity. The affinity constants for most antibody–antigen interactions are determined largely by the dissociation constants, k_d, with little variation observed in rates of association. Additionally, surface immobilization typically results in a reduced k_d. In this case, the sensor binds the analyte kinetically, irreversibly preventing response to changes in analyte concentration and reuse. Regeneration of the sensor surface is difficult at best. On the other hand, a higher dissociation rate, which would lend itself to a linear and reusable sensor, results in lower affinity and poor sensitivity. Consequently, most biosensors are disposable devices and quantitation is obtained using multiple single-use sensors. Recently, a new reusable biosensor platform which provides simultaneous fluorescence detection and electrochemical control of biospecific binding has been developed.[25] Biotin was covalently attached to a transparent indium tin oxide electrode, which also served as an integral part of a TIRF flow cell. TIRF was used to monitor biospecific interactions while electrochemical polarization was employed to control the interactions. Antibody–antigen and avidin–biotin interactions both exhibited reversibility over 30 reuse cycles. A generalized TIRF biosensor scheme known as a ''sensogram'' is shown in Figure 8.

5.2.2.4 Chromatographic Surfaces The application of TIRF has been further expanded for the characterization of chromatographic surfaces to obtain a better understanding of the solute separation mechanism on a molecular level. The microenvironment polarity of chromatographic surfaces has been investigated by TIRF.[49,50] Hartner et al.[49] derivatized a fused silica plate with octadecyltrichlorosilane to obtain a C_{18} surface layer. The derivatized plate was optically coupled, with an index matching fluid, to an internal reflection quartz prism. Rangnekar and Oldham[50] chemically modified a multiple internal reflection element (single-pass parallelepiped 60° bevel angle UV-quartz plate) with octadecylchlorosilane and AquaSil™. Their TIRF optical configuration provided multiple internal reflections (number of reflections = 7) which increased the sensitivity. In both studies, the interfacial molecular environment was probed by pyrene physisorbed onto the modified surface. Pyrene TIRF emission spectra were collected in the presence of overlaying methanol/water mixtures. TIRF provided sufficient selectivity and sensitivity to allow the environment of the interface to be probed by pyrene which was adsorbed from the bulk solution. The percentage of the TIRF signal from the surface-adsorbed pyrene and the bulk-solution pyrene depends on the partition ratio under the conditions and the depth of penetration of the evanescent wave beyond the interface. The vibronic band intensities of pyrene, which are quite sensitive to the surrounding polarity, provided an insight to the polarity of the interior of the C_{18} chains and the chain configuration in the presence of different polar solvents.

TIRF anisotropy measurements have led to estimations of the microviscosity of the chromatographic stationary phase[51] and the orientational dynamics of solutes in a chromatographic stationary phase.[52,53] The basis of fluorescence anisotropy measurements is that,

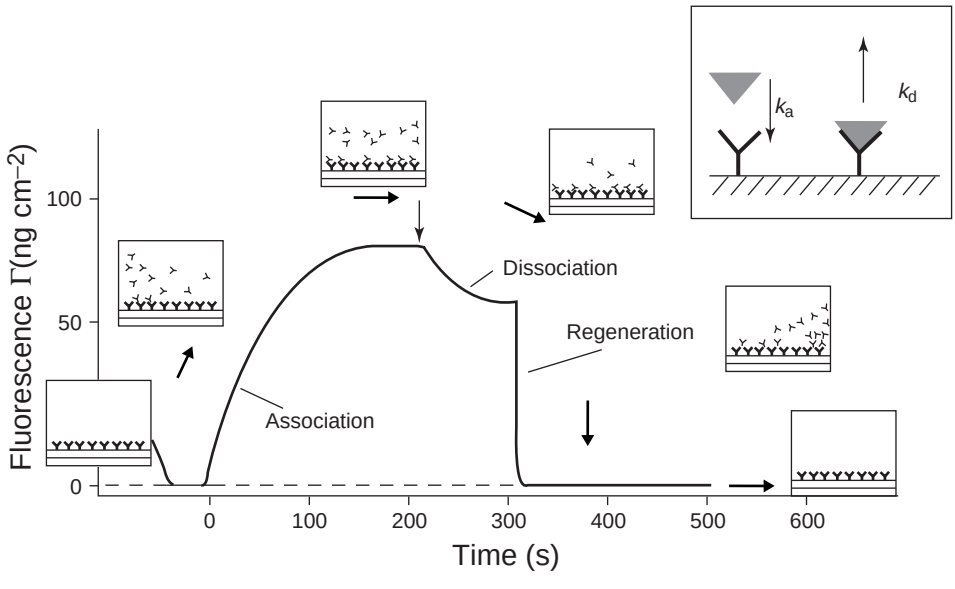

Figure 8 Generalized sensogram for a TIRF biosensor.

upon excitation of the sample with polarized light, the polarization of the emission intensity is related to the molecular orientation. Since the molecule rotates within its excited state lifetime, the emitted fluorescence is depolarized. The fluorescence depolarization (or anisotropy) measurements disclose information concerning rotational mobility, which in turn is associated with the size and shape of the fluorophore and the viscosity of the surrounding medium.

Rangnekar et al.[51] chemically modified a UV-quartz surface with different silylating reagents to obtain either C_1-, C_3-, C_8- or C_{18}-modified silica surfaces. TIRF anisotropy data of 1,6-diphenyl-hexatriene were collected on unmodified and modified surfaces in the presence of overlaying solvents covering a range of viscosities. Their results indicate that the adsorption and partitioning mechanisms depend at least partially on the carbon number of the alkyl chain. They estimated the C_{18} surface microviscosity to be approximately 18 cP.

Wirth and Burbage[52] have derived mathematical relationships for TIRF anisotropy studies of adsorbate reorientation at surface/liquid interfaces. The group examined the reorientation of acridine orange adsorbed on an (octadecylsilyl)silica surface. In the study, an octadecylsilane (ODS)-derivatized fused silica plate was optically coupled, with a silica index matching fluid, to a trapezoidal internal reflection prism. The thermodynamic equilibrium between the surface-adsorbed acridine orange and aqueous acridine orange was maintained by continuously flowing the solution through a flow cell. The surface acridine orange was excited with light polarized either in-plane or out-of-plane by using an evanescent wave. Fluorescence-domain spectroscopy was used to characterize the fluorescence anisotropy decays of acridine orange. The experiments revealed that the out-of-plane rotations of acridine orange at the ODS/water interface were strongly hindered, while the in-plane rotations gave rise to a double exponential decay of unknown origin. The interpretation was that acridine orange resides at the ODS/water interface. Wirth's group[53] has also examined the effect of wetting by alcohols on the orientations of C_{18} chains. The geometry of the probe solute bis-*p*-(*o*-methyl)styrylbenzene used in the study favored preferential alignment with the C_{18} chains. It was assumed that any changes in the orientational distribution and dynamics of the C_{18} chains caused related changes in the orientational distribution and dynamics of the probe solute. The results indicate that the C_{18} chains lie flat on the surface when water is the mobile phase. A small amount of alcohol added to the water causes the C_{18} chains to become extended towards the surface normal and the chains become random in orientation. But it is likely that wetting has a significantly smaller effect upon the chain orientation.

ACKNOWLEDGMENTS

The authors wish to thank Ms Fang Qian for assistance with figure preparation. Sections 2 and 3 of this article are reprinted (with minor changes) from P.B. Oldham, 'Multidimensional Fluorescence Identification of Phytoplankton', *Modern Techniques for Rapid Microbiological Analysis*, ed. W.H. Nelson, VCH-Wiley, New York, Chapter 7, 1991, by permission of John Wiley & Sons, Inc.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflectance
bis-ANS	4,4′-Bis-1-phenylamino-8-naphthalene-sulfonate
BSA	Bovine Serum Albumin
CCD	Charge-coupled Device
CTAB	Cetyltrimethylammonium Bromide
CVD	Chemical Vapor Deposition
FCS	Fluorescence Correlation Spectroscopy
FPR	Fluorescence Recovery After Photobleaching
FTIR	Fourier Transform Infrared
IOW	Integrated Optical Waveguides
ODS	Octadecylsilane
PMMA	Poly(methyl methacrylate)
SERS	Surface-enhanced Raman Spectroscopy
SIMS	Secondary-ion Mass Spectrometry
TIR	Total Internal Reflection
TIRF	Total Internal Reflection Fluorescence
TIRR	Total Internal Reflection Raman
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VATIRF	Variable-angle Total Internal Reflection Fluorescence
VIS	Visible
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Environment: Water and Waste **(Volume 4)**
Luminescence in Environmental Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Surfaces **(Volume 10)**
Photoluminescence in Analysis of Surfaces and Interfaces

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Detectors, Absorption and Luminescence • Fluorescence Lifetime Measurements, Applications of • Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

REFERENCES

1. N.J. Harrick, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, New York, 1979.
2. R.M. Silverstein, F.X. Webster, *Spectrometric Identification of Organic Compounds*, 6th edition, J. Wiley & Sons, New York, 1997.
3. S.A. Soper, I.M. Warner, L.B. McGown, 'Molecular Fluorescence, Phosphorescence, and Chemiluminescence Spectrometry', *Anal. Chem.*, **70**(12), 477R–494R (1998).
4. C.I. Miles, G.H. Schenk, 'Fluorescence and Phosphorescence of Phenylethylamines and Barbiturates. Analysis of Amphetamine and Barbiturate Preparations', *Anal. Chem.*, **45**(1), 130–136 (1973).
5. J.C. Wright, F.J. Gustafson, 'Ultratrace Inorganic Ion Determination by Laser Excited Fluorescence', *Anal. Chem.*, **50**(12), 1147A–1160A (1978).
6. Shimon Weiss, 'Fluorescence Spectroscopy of Single Biomolecules', *Science*, **283**(5408), 1676–1683 (1999).
7. P.C. Clemmow, *The Plane Wave Spectrum Representation of Electromagnetic Fields*, Pergamon Press, New York, 1966.
8. P.A. Suci, W.M. Reichert, 'Determination of Fluorescence Density Profiles of Langmuir–Blodgett-deposited Films by Analysis of Variable-angle Fluorescence Data Curves', *Langmuir*, **4**, 1131–1141 (1988).
9. D.A. Stephens, P.W. Bohn, 'Absorption Spectrometry of Bound Monolayers on Integrated Optical Structures', *Anal. Chem.*, **61**(5), 386–390 (1989).
10. R.E. Dessy, 'Waveguides as Chemical Sensors', *Anal. Chem.*, **61**(19), 1079A–1094A (1989).
11. T.E. Plowman, M.D. Garrison, D.S. Walker, W.M. Reichert, 'Surface Sensitivity of SiON Integrated Optical Waveguides (IOWs) Examined by IOW-attenuated Total Reflection Spectrometry and IOW-Raman Spectroscopy', *Thin Solid Films*, **243**, 610–615 (1994).
12. L. Yang, S.S. Saavedra, 'Chemical Sensing Using Sol–Gel Derived Planar Waveguides and Indicator Phases', *Anal. Chem.*, **67**(8), 1307–1314 (1995).
13. D.R. Dunphy, S.B. Mendes, S.S. Saavedra, N.R. Armstrong, 'The Electroactive Integrated Optical Waveguide: Ultrasensitive Spectroelectrochemistry of Submonolayer Absorbates', *Anal. Chem.*, **69**(15), 3086–3094 (1997).
14. N. Winograd, T. Kuwana, 'Characteristics of the Electrode–Solution Interface Under Faradaic and Non-Faradaic Conditions as Observed by Internal Reflection Spectroscopy', *J. Electroanal. Chem.*, **23**, 333–342 (1969).
15. K. Itoh, A. Fujishima, 'An Application of Optical Waveguides to Electrochemistry: Construction of Optical Waveguide Electrodes', *J. Phys. Chem.*, **92**, 7043–7045 (1988).
16. K. Itoh, A. Fujishima, *Electrochemistry in Transition*, eds. O.J. Murphy, S. Srinivasan, B.E. Conway, Plenum Press, New York, 219–225, 1992.
17. C. Piraud, E. Mwarania, G. Wylangowski, J. Wilkinson, K. O'Dwyer, D.J. Schiffrin, 'An Optoelectrochemical Thin-film Chlorine Sensor Employing Evanescent Fields on Planar Optical Waveguides', *Anal. Chem.*, **64**(6), 651–655 (1992).
18. C. Piraud, E. Mwarania, G. Wylangowski, J. Yao, D.J. Schiffrin, J.S. Wilkinson, 'Optoelectrochemical Transduction on Planar Optical Waveguides', *J. Lightwave Technol.*, **10**(5), 693–699 (1992).
19. S.B. Mendes, L. Li, J.J. Burke, J.E. Lee, D.R. Dunphy, S.S. Saavedra, 'Broad-band Attenuated Total Reflection Spectroscopy of a Hydrated Protein Film on a Single Mode Planar Waveguide', *Langmuir*, **12**, 3374–3376 (1996).
20. P.L. Edminston, J.E. Lee, S.-S. Cheng, S.S. Saavedra, 'Molecular Orientation Distributions in Protein Films. 1. Cytochrome *c* Adsorbed to Substrates of Variable Surface Chemistry', *J. Am. Chem. Soc.*, **119**, 560–570 (1997).
21. L.L. Wood, S.-S. Cheng, P.L. Edminston, S.S. Saavedra, 'Molecular Orientation Distributions in Protein Films. 2. Site-directed Immobilization of Yeast Cytochrome *c* on Thiol-capped, Self-assembled Monolayers', *J. Am. Chem. Soc.*, **119**, 571–576 (1997).
22. T. Hirschfeld, 'Total Reflection Fluorescence', *Can. Spectrosc.*, **10**(5), 128 (1965).
23. J.D. Andrade (ed.), *Protein Adsorption*, Plenum Press, New York, 1985.
24. V. Hlady, 'Spectroscopic and Other Techniques for Studying Adsorption of Bioproducts at Interfaces', in *Interfacial Phenomena and Bioproducts*, eds. J.L. Brash, P.W. Wojciechowski, Marcel Dekker, New York, 1996.
25. A.N. Asanov, W.W. Wilson, P.B. Oldham, 'Regenerable Biosensor Platform: A Total Internal Reflection Fluorescence Cell with Electrochemical Control', *Anal. Chem.*, **70**(6), 1156–1163 (1998).
26. P.W. Ambrose, P.M. Goodwin, J.P. Nolan, 'Single-molecule Detection with Total Internal Reflection Excitation: Comparing Signal-to-background and Total Signals in Different Geometries', *Cytometry*, **36**(3), 224–231 (1999).
27. A.I. Mahan, C.V. Bitterli, 'Total Internal Reflection: A Deeper Look', *Appl. Opt.*, **17**, 509–519 (1978).

28. W.M. Reichert, P.A. Suci, J.T. Ives, J.D. Andrade, 'Evanescent Detection of Adsorbed Protein Concentration–Distance Profiles: Fit of Simple Models to Variable-angle Total Internal Reflection Fluorescence Data', *Appl. Spectrosc.*, **41**(3), 503–508 (1987).
29. E. Lee, R.E. Benner, J.B. Fenn, R.K. Chang, 'Angular Distribution of Fluorescence from Liquids and Monodispersed Spheres by Evanescent Wave Excitation', *Appl. Opt.*, **18**(6), 862–868 (1979).
30. N.J. Harrick, G.I. Loeb, 'Multiple Internal Reflection Fluorescence Spectrometry', *Anal. Chem.*, **45**(4), 687–691 (1973).
31. R.A. Van Wagenen, S. Rockhold, J.D. Andrade, 'Probing Protein Adsorption: Total Internal Reflection Intrinsic Fluorescence', *Adv. Chem. Ser.*, **199**, 351–370 (1982).
32. S.A. Rockhold, R.D. Quinn, R.A. Van Wagenen, J.D. Andrade, M. Reichert, 'Total Internal Reflection Fluorescence (TIRF) as a Quantitative Probe of Protein Adsorption', *J. Electroanal. Chem.*, **150**(1–2), 261–275 (1983).
33. V. Hlady, D.R. Reinecke, J.D. Andrade, 'Fluorescence of Adsorbed Protein Layers. 1. Quantitation of Total Internal Reflection Fluorescence', *J. Colloid Interface Sci.*, **111**(2), 555–569 (1986).
34. V. Hlady, C. Golander, J.D. Andrade, 'Hydrophobicity Gradient on Silica Surfaces: A Study Using Total Internal Reflection Fluorescence Spectroscopy', *Colloids Surf.*, **33**(1–2), 185–190 (1988).
35. V. Hlady, J. Rickel, J.D. Andrade, 'Fluorescence of Adsorbed Protein Layers. II. Adsorption of Human Lipoproteins Studied by Total Internal Reflection Intrinsic Fluorescence', *Colloids Surf.*, **34**(2), 171–183 (1988).
36. V. Hlady, J.D. Andrade, 'Fluorescence Emission from Adsorbed Bovine Serum Albumin and Albumin-bound 1-Anilinonaphthalene-8-sulfonate Studied by TIRF', *Colloids Surf.*, **32**(3–4), 359–369 (1988).
37. C.G. Golander, V. Hlady, K. Caldwell, J.D. Andrade, 'Adsorption of Human Lysozyme and Adsorbate Enzyme Activity as Quantified by Means of Total Internal Reflection Fluorescence, Iodine-125 Labeling and ESCA', *Colloids Surf.*, **50**, 113–130 (1990).
38. B.K. Lok, Y. Cheng, C.R. Robertson, 'Protein Adsorption on Crosslinked Poly(dimethylsiloxane) Using Total Internal Reflection Fluorescence', *J. Colloid Interface Sci.*, **91**(1), 104–116 (1983).
39. N.L. Thompson, T.P. Burghardt, D. Axelrod, 'Measuring Surface Dynamics of Biomolecules by Total Internal Reflection Fluorescence with Photobleaching Recovery or Correlation Spectroscopy', *Biophys. J.*, **33**(3), 435–454 (1981).
40. T.P. Burghardt, D. Axelrod, 'Total Internal Reflection Fluorescence Photobleaching Recovery Study of Serum Albumin Adsorption Dynamics', *Biophys. J.*, **33**(3), 455–467 (1981).
41. T.P. Burghardt, D. Axelrod, 'Total Internal Reflection Fluorescence Study of Energy Transfer in Surface-adsorbed and Dissolved Bovine Serum Albumin', *Biochemistry*, **22**(4), 979–985 (1983).
42. L.E. Morrison, G. Weber, 'Biological Membrane Modeling with a Liquid/Liquid Interface. Probing Mobility and Environment with Total Internal Reflection Excited Fluorescence', *Biophys. J.*, **52**(3), 367–379 (1987).
43. H. Masuhara, S. Tazuke, N. Tamai, I. Yamazaki, 'Time-resolved Total Internal Reflection Fluorescence Spectroscopy for Surface Photophysics Studies', *J. Phys. Chem.*, **90**(22), 5830–5835 (1986).
44. A. Itaya, A. Kurahashi, H. Masuhara, N. Tamai, I. Yamazaki, 'Dynamic Fluorescence Microprobe Method Utilizing a Total Internal Reflection Phenomenon', *Chem. Lett.*, **6**, 1079–1082 (1987).
45. A. Itaya, T. Yamada, K. Tokuda, H. Masuhara, 'Interfacial Characteristics of Poly(meththyl methacrylate) Film: Aggregation of Pyrene and Micropolarity Revealed by Time-resolved Total Internal Reflection Fluorescence Spectroscopy', *Polymer. J.*, **22**(8), 697–704 (1990).
46. C.M. Ingersoll, F.V. Bright, 'Using Fluorescence to Probe Biosensor Interfacial Dynamics', *Anal. Chem.*, **69**(13), 403A–408A (1997).
47. D.L. Wise, L.B. Wingard (eds.), *Biosensors with Fiberoptics*, Humana Press, Clifton, NJ, 1991.
48. T.E. Plowman, W.M. Reichert, C.R. Peters, H.K. Wang, D.A. Christensen, J.N. Herron, 'Femtomolar Sensitivity Using a Channel-etched Thin Film Waveguide Fluoroimmunosensor', *Biosens. Bioelectron.*, **11**, 149–160 (1996).
49. K.C. Hartner, J.W. Carr, J.M. Harris, 'Total Internal Reflection Fluorescence for Adsorbed Probe Molecule Studies of Liquid/Solid Interfacial Environments', *Appl. Spectrosc.*, **43**(1), 81–87 (1989).
50. V.M. Rangnekar, P.B. Oldham, 'Investigation of the Microenvironment Polarity of a Chromatographic Surface Using Total Internal Reflection Fluorescence', *Anal. Chem.*, **62**(11), 1144–1147 (1990).
51. V.M. Rangnekar, J.T. Foley, P.B. Oldham, 'Investigation of Chromatographic Surface Viscosity Using Total Internal Reflection Fluorescence', *Appl. Spectrosc.*, **46**(5), 827–831 (1992).
52. M.J. Wirth, J.D. Burbage, 'Adsorbate Reorientation at a Water/(Octadecylsilyl) Silica Interface', *Anal. Chem.*, **63**(13), 1311–1317 (1991).
53. M.E. Montgomery, Jr, M.A. Green, M.J. Wirth, 'Orientational Dynamics of a Hydrophobic Guest in a Chromatographic Stationary Phase: Effect of Wetting by Alcohol', *Anal. Chem.*, **64**(10), 1170–1175 (1992).

Ultraviolet and Visible Molecular Absorption and Fluorescence Data Analysis

Sharon L. Neal
University of Delaware, Newark, USA

1 Introduction **1**
2 Principles of Ultraviolet/Visible Molecular Absorbance and Fluorescence Measurements **3**
2.1 Ultraviolet/Visible Absorbance Data 3
2.2 Fluorescence Intensity Measurements 4
2.3 Anisotropy Measurements 9
2.4 Quenching Measurements 10
2.5 Energy Transfer Measurements 11
2.6 Multidimensional Measurements 12
3 Instrumentation for Ultraviolet/Visible Molecular Fluorescence Measurements **13**
3.1 Absorbance Spectrophotometer 13
3.2 Time-resolved Spectrofluorometer 13
3.3 Frequency-domain Spectrofluorometer 14
4 Data Analysis of Ultraviolet/Visible Molecular Absorbance and Fluorescence Measurements **14**
4.1 Classification of Data Analysis Methods 14
4.2 Regression Methods 15
4.3 Decay Analysis Methods 18
4.4 Matrix Factorization Methods 23
5 Error Analysis and Result Validation **27**
5.1 Parameter Confidence Intervals 27
5.2 Rank Estimation 28
5.3 Residual Analysis 28
5.4 Stability 29
6 Software **29**
7 Future Development **29**
Acknowledgments **29**
List of Symbols **29**
Abbreviations and Acronyms **32**
Related Articles **32**
Appendix 1: Matrix Algebra Basics **33**
Notation 33
Definitions 33
Special Matrices 33
Operations 33
Structure 34
References **34**

This article is a survey of the data analysis methods that have been applied to ultraviolet (UV) and visible molecular absorbance and fluorescence measurements. The section after the Introduction reviews the principles of the most widely used types of absorbance and fluorescence measurements. These include steady-state, dynamic and multidimensional measurements. After a brief description of absorption and fluorescence instruments, three classes of data analysis methods for molecular absorbance and fluorescence are described: regression methods, decay analysis and matrix factorization methods. Result validation, software alternatives and future developments are discussed in short sections at the end of the article. A short matrix algebra primer is provided as an appendix to the article.

1 INTRODUCTION

Ultraviolet/visible (UV/VIS) molecular absorbance measurements are among the most widely used spectrochemical methods because they provide nearly universal detection and require modest investments in equipment and user training, except in the case of ultrafast dynamic measurements. UV/VIS molecular fluorescence measurements are less widely used than absorbance methods; relatively few molecules are intrinsically good fluorescers and equipment, maintenance and training for dynamic experiments can be expensive. Despite this, fluorescence measurements are critical in several areas of important research. First, the detection limits of fluorescence are unparalleled by any other method for molecules in solution. Single molecule studies routinely capitalize on this feature.[1,2] Second, fluorescence emission occurs on the timescale of molecular motions and is thus sensitive to molecular interactions and dynamics.[3] Dynamic fluorescence measurements can be used to measure excited state reaction rates,[4,5] translational[6] and rotational diffusion coefficients,[7] intramolecular distances, solute–solvent interactions and related parameters.[8] Several fluorophores emit microenvironment sensitive fluorescence which can be correlated to local solvent properties such as polarity, viscosity, pH,[9] and potential.[10] These capabilities are the basis of the wide use of fluorescence methods in photochemical, biophysical, and photobiological studies. (Of course, this feature has a dark side. It is a primary source of fluorescence sensitivity

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

to matrix effects.) Third, fluorescence is a function of many experimental variables.[11] Experimental variables can be adjusted to avoid known interference or the fluorescence measurement can be made as a function of simultaneous variation in several variables in order to take advantage of numerical methods for analysis in the presence of interference.[12]

This article is a survey of the data analysis methods that have been applied to UV/VIS molecular absorbance and fluorescence measurements. While the name of this Series implies an exhaustive reference that includes the most obscure data analysis procedures, this report is limited to methods that are in wide use or enable analysts to take advantage of advances in instrument technology. Most of the material is presented in two of the seven report sections. The first, section 2, reviews the principles of the most widely used types of absorbance and fluorescence measurements. This section introduces mathematical expressions that describe the data acquired in each type of measurement for which analysis methods will be discussed. These expressions are important because they define the information content of each measurement and provide the basis for deciding which analysis procedure(s) is (are) best suited to recovering the desired information from the data. The reviews also provide brief overviews of the potential applications of each measurement. The second major section is section 4, which organizes data analysis methods for molecular absorbance and fluorescence into three classes. Section 4 also describes the principles and application of the most prominent data analysis methods in each of the three classes. Section 5 is the discussion of result validation for the various analysis classes. It is not as large as sections 2 and 4, but the topic is very important. Despite the best efforts of generations of scientists and statisticians, there is generally an element of subjectivity in data analysis. This is particularly true of measurements comprised of combinations of exponentials, such as dynamic absorbance and fluorescence. It is imperative that data analysis results be evaluated for statistical relevance and stability and bound by confidence intervals whenever possible.

Even the simplest measurement, for example a single sample absorbance and the absorbances of the calibration standards, can be organized into a data set using matrix algebra. Topics such as multicomponent analysis, multichannel (multiple variable) measurements, and curve fitting which involve more complicated data arrangements and operations, would be impossible to discuss without matrix algebra. There are many good tutorials on matrix algebra that will help the reader master the algebra used in this article.[13] The most important ideas to remember are: (1) addition, subtraction and scalar multiplication for vectors and matrices are directly analogous to their scalar counterparts; (2) multiplying by an inverse matrix is analogous to division; (3) most single variable operations, for example differentiation, integration, expansion, have direct analogies to vectors and matrices.

A detailed description of the notation used in this report is given at the beginning of the matrix algebra basics in Appendix 1. It suffices here to say that conventional matrix notation and variable names are used wherever possible. For example, boldface Roman and Greek letters represent matrices and vectors. Unfortunately, a uniform and consistent notation scheme is incompatible with some commonly used variable names. For example, the capitalized variable I typically represents light intensity, a scalar quantity. In this article, underlining and italics will distinguish scalar variables and vectors. For example, the intensity of fluorescence is represented by the italicized variable I_f; a single underline under an italicized variable indicates a related vector quantity, $\underline{I_f}$; and a matrix is indicated by double underlining, $\underline{\underline{I_f}}$. (It is important to remember that a series of italicized variables and their unitalicized counterparts, e.g. S, $\underline{S}$, $\underline{\underline{S}}$ and s, **s**, **S**, are unrelated.) In section 2, boldface also will be applied to the independent variables to help emphasize the order of the measurement. So, $I_f(\lambda_{ex}, \lambda_{em}, \varphi_{ex}, t)$ or I_f is a single value: the intensity observed after excitation at λ_{ex} at one λ_{em} through polarizer set to angle φ_{ex} at t seconds after excitation. $\underline{I_f}(\boldsymbol{\lambda}_{ex}, \boldsymbol{\lambda}_{em}, \varphi_{ex}, t)$ or $\underline{I_f}(\boldsymbol{\lambda}_{ex})$ is a vector: the fluorescence emission spectrum produced by excitation at λ_{ex}, observed through polarizer set at φ_{ex} at time t after excitation. $\underline{\underline{I_f}}(\lambda_{ex}, \boldsymbol{\lambda}_{em}, \varphi_{ex}, \mathbf{t})$ or $\underline{\underline{I_f}}(\boldsymbol{\lambda}_{em}, \mathbf{t})$ is a matrix: the fluorescence intensity produced by excitation at λ_{ex}, monitored as a function of λ_{em} and t through a polarizer set at φ_{ex}. This can go on, but I hope the point is made. Data dispersed in more than three dimensions and N-way matrices are germane to the topic, but the requisite data analysis methods are beyond the scope of this article. There is only a brief discussion of three-way fluorescence data in section 4.4.3. The other cases of unconventional notation arise when a single variable name is commonly used to define two entities. Some common variable names have been changed in order to avoid multiple definitions for variables wherever possible. For example, the variable r commonly represents both the distance between donor–acceptor pairs in energy transfer measurements and the fluorescence anisotropy. In this article, d denotes the interatomic distance to distinguish it from the fluorescence anisotropy. Despite these steps, multiple variable definitions could not be completely avoided. The most glaring example is the use of the letters i, j, k, l, m, n, p and q to represent element labels in vectors and matrices and counters in sums and products in addition to their use as variables.

2 PRINCIPLES OF ULTRAVIOLET/VISIBLE MOLECULAR ABSORBANCE AND FLUORESCENCE MEASUREMENTS

2.1 Ultraviolet/Visible Absorbance Data

2.1.1 Steady-state Ultraviolet/Visible Absorbance Data

The energy level spacing of many molecules, especially large organic molecules, is such that they absorb light in the UV to visible wavelength range, that is ~200–800 nm. This is illustrated in the Jablonski diagram depicted in Figure 1. The intensity of light absorbed by a sample is an extensive property that depends on the electronic energy level spacing in the chromophore and the number of chromophores the light beam encounters. Specifically, Beer–Lambert's law relates the intensity of light absorbed to the concentration of solution species, as shown in Equations (1) and (2):

$$I_T(\lambda_{ex}) = I_I(\lambda_{ex})e^{-\varepsilon(\lambda_{ex})bc} \quad (1)$$

$$A(\lambda_{ex}) = \log\frac{I_I(\lambda_{ex})}{I_T(\lambda_{ex})} = \varepsilon(\lambda_{ex})bc \quad (2)$$

where I_I is the incident light intensity, I_T is the transmitted light intensity, $\varepsilon(\lambda_{ex})$ is the molar absorptivity at a specific wavelength, b is the beam pathlength through the sample, c is the concentration of the chromophore and A is the absorbance. The absorbance spectrum, $\underline{A}(\boldsymbol{\lambda}_{ex})$, is a plot of the absorbance as a function of excitation wavelength. The absorption wavelength reflects the energy of the excited state. The spectrum band structure reflects the vibronic band spacing of the chromophore excited state. Figure 2 depicts the absorbance spectra of two forms of phytochrome, a photoreceptor that regulates plant growth and development.[14] The biologically active far-red-light absorbing form is reversibly photogenerated from the red-light absorbing form.

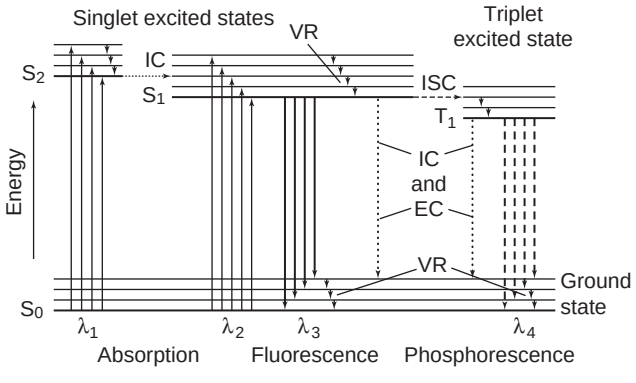

Figure 1 Jablonski energy level diagram. S_0, ground state singlet; S_1, excited singlet state; T_1, excited triplet state; λ_1–λ_4, wavelength ranges; IC, internal conversion; VR, vibrational relaxation; ISC, intersystem crossing; EC, external conversion.

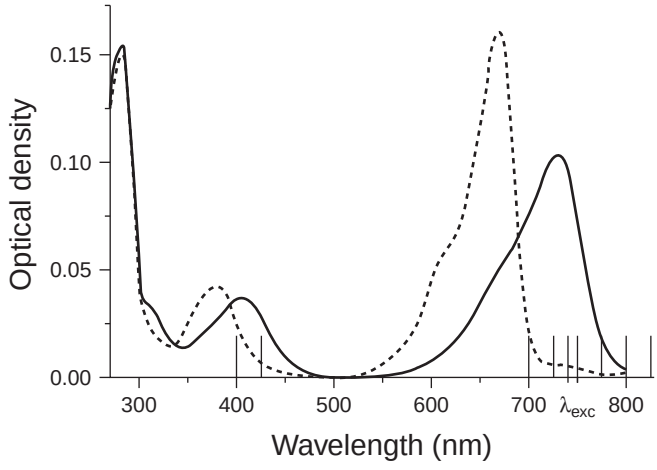

Figure 2 Steady-state absorbance spectrum of the red-light and far-red-light absorbing phytochrome in 10 mM phosphate buffer, pH 7.8 at 278 K. (Reproduced by permission of the American Chemical Society from Bischoff et al.[14])

As a consequence of Beer's law, the majority of absorbance data is acquired as an indicator of concentration. Hence, the goal of most absorbance data analysis is to use the dependence of the absorbance on the analyte concentration for quantitation. UV/VIS molecular absorbance spectra of molecules in liquid solution near room temperature are typically broad. Spectral overlap, absorption by other sample components and matrix effects that produce deviations from Beer's law all complicate quantitation.

2.1.2 Dynamic Ultraviolet/Visible Absorbance Data

Several experimental procedures generate dynamic absorbance data. As in the case of steady-state data, the goal of such analyses is to quantify the components of a system, but with respect to time. These experiments monitor reaction progress, so the goal of the data analysis is isolation of individual component spectra and evaluation of system kinetics. Kinetic analysis generally consists of identifying a model that describes the reaction and determining the rates of the various processes that occur within the system.

In the simplest versions of these experiments, absorbances at a single wavelength are measured as a function of time. Absorbance spectrometers based on multichannel detectors are available commercially at modest prices, making spectra measured as a function of time a common alternative data format.[15] Multichannel spectrometers can be used to monitor chromatographic and electrophoretic separations or, more commonly, to study kinetic behavior. The mixing apparatus, source pulse rate and detector rise time define the measurement timescale.

A spectrometer based on a deuterium lamp and photodiode array can monitor reactions exhibiting millisecond or slower kinetics. Stopped-flow methods, xenon-flash lamps and intensified photodiode arrays can monitor microsecond kinetics. Pulsed-laser techniques such as differential transient absorption can monitor the time-dependent behavior of faster processes. For instance, picosecond dynamics can be monitored by measuring the difference in the absorbance of the unexcited sample and the sample at some time after excitation by a pump laser pulse.[16] Monitoring this difference as a function of the delay between the pump beam and probe beam produces a record of the changes in the absorption as a function of time on the scale of the delay. Figure 3 depicts the transient absorption spectra of the visual pigment rhodopsin.

In multichannel kinetic measurements, the data set acquired is an absorbance–time matrix. Beer's law defines the components of this matrix. Assuming unit pathlength, the matrix elements are the sums of the products of the molar absorptivities of the sample components at the observed wavelengths and concentrations of the components at the observation times, as shown in Equation (3).

$$\underline{\underline{A}} = \sum_{k}^{K} \boldsymbol{\varepsilon}_{\cdot k} c_{\cdot k}^{\mathrm{T}} = \mathbf{E}\mathbf{C}^{\mathrm{T}} \tag{3}$$

Each column of $\underline{\underline{A}}$ is the absorbance spectrum of the sample at a single observation time. Each row is the time course of the absorbance at a single wavelength. Each column of **E** consists of the molar absorptivities of one sample constituent at all the wavelengths monitored during the measurement. Each column of **C** stores the ground-state concentrations of one sample constituent as a function of the observation times. In the case of transient absorbance measurements, the columns of **C** are the excited-state concentration profiles. The goal of kinetic analysis is to isolate the time profiles of the sample constituents, **C**, and determine the kinetic rates that describe the time dependence of the concentrations.

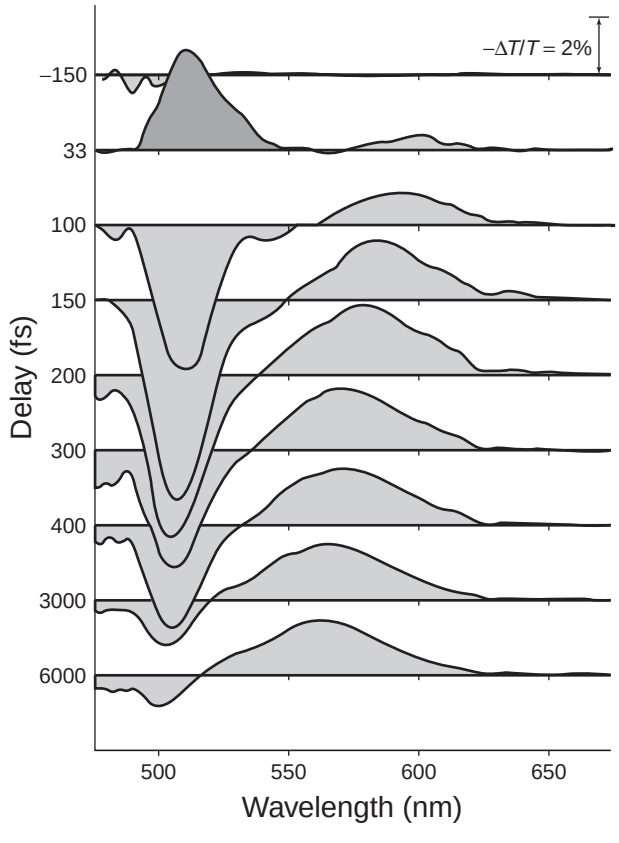

Figure 3 Transient differential absorption spectra of rhodopsin at various probe delays (~10 fs probe) following a 35 fs pump pulse at 500 nm. (Reproduced by permission of Blackwell Science from Mathies.[16])

2.2 Fluorescence Intensity Measurements

2.2.1 Steady-state Intensity Measurements

When rigid molecules that have extended conjugated π-electron systems are photoexcited, many of the vibrational and rotational deactivation mechanisms that depopulate excited states are not available. Such molecules relax via photon emission. In conjugated organic molecules, fluorescence is usually observed during transitions from the first excited to ground singlet states, as depicted by bold solid lines in Figure 1.

The intensity of fluorescence emission is a function of several measurement variables and conditions; the most crucial are the excitation source properties. First, the wavelength of the excitation must correspond to the electronic energy level spacing of the fluorophore (Figure 1) in order to produce the excited state. Fluorescence emission is a random process that is driven by photoexcitation. A differential equation describes the rate of fluorophore excitation and relaxation characterizing the time dependence of the number of excited states in a homogeneous population of fluorophores, as shown in Equation (4)

$$\frac{\mathrm{dz}}{\mathrm{dt}} = \mathrm{e} - \mathrm{k_f z} \tag{4}$$

where z represents the instantaneous concentration of excited fluorophores, e is the rate of excitation, and $\mathrm{k_f}$ is the first-order rate constant describing excited state deactivation. The relaxation rate is actually a combination of the rates of all processes that deactivate the excited state, $\mathrm{k_f} = \mathrm{k_f^0} + \mathrm{k_{nr}}$ where the nonradiative processes (see Figure 1) have been collected into a single term, $\mathrm{k_{nr}}$. The

time dependence of the emission also depends on the time course of the excitation beam, $\mathbf{e} = \mathbf{I}_I(\lambda_{ex}, \mathbf{t})$. The observed intensity is the convolution of the excitation beam and the impulse response, i.e. system response to excitation by an infinitely narrow pulse. In the case of single fluorophores in dilute isotropic solution the impulse response function is generally a monoexponential decay. Equation (5) shows:

$$I_f(\lambda_{ex}, \lambda_{em}, \mathbf{t}) = k_f V \mathbf{z} = k_f V[I_I(\lambda_{ex}, \mathbf{t}) \circ z_0(\lambda_{ex}, \lambda_{em})\, e^{-\mathbf{t}/\tau_f}] \quad (5)$$

[where V is the volume of solution from which fluorescence is collected, $\circ$ denotes convolution, and $z_0(\lambda_{ex}, \lambda_{em})$ is the initial ($t = 0$) excited state concentration.] The decay constant in the impulse response is the inverse of the average lifetime of a fluorophore in the excited state, τ_f. It is the time required for the excited state population to decrease by 63%.

In steady-state measurements, the excitation function is constant, $\mathbf{e}(\mathbf{t}) = I_I(\mathbf{t})k_A c$. The ground state fluorophore concentration, c, is unchanged by the production of excited states at rate k_A. The constant (except for noise) source intensity produces a steady pool of excited fluorophores. The number of excited fluorophores reaches a constant value as a dynamic equilibrium between excited state production and deactivation is reached. The solution of Equation (4), shown in Equation (6), in this case is proportional to

$$z(\mathbf{t}) = \frac{k_A c}{k_f + k_{nr}} \quad (6)$$

Another expression for the steady-state fluorescence can be derived from the intensity of the light absorbed by the molecule, as shown in Equation (7)

$$I_A = I_I - I_T = I_I(1 - e^{-\varepsilon bc}) \quad (7)$$

The fluorescence quantum yield, ϕ_f, is the number of photons emitted per photon absorbed. Equation (8) shows:

$$I_f(\lambda_{ex}) = \phi_f I_A(\lambda_{ex}) \approx 2.303 \phi_f I_I(\lambda_{ex}) \varepsilon(\lambda_{ex}) bc \quad (8)$$

The intensity of fluorescence, I_f, is proportional to the incident light intensity at low absorbances ($A < 0.050$). The excitation spectrum $\underline{I_f}(\boldsymbol{\lambda}_{ex})$ illustrates the dependence of the emission intensity on excitation wavelength. The excitation spectrum has the same profile as the absorbance spectrum when it is corrected for instrument response and measurement conditions, yet it is measured at concentrations well below the detection limits of absorption spectrometers. The excitation spectrum of perylene is pictured in Figure 4.

The steady-state fluorescence intensity is a function of several measurement parameters other than the source intensity and excitation wavelength, λ_{ex}. These include the emission wavelength λ_{em}, excitation light polarization φ_{ex} and quencher or energy transfer acceptor concentration, [Q] or [A]. The emission wavelength dependence of the fluorescence intensity is given by Equation (9)

$$I_f(\lambda_{em}) = \phi_f I_I(\lambda_{ex}) \varepsilon(\lambda_{ex}) \zeta(\lambda_{em}) bc \quad (9)$$

where ζ is the probability that a photon will be emitted at a specific wavelength and the other variable designations are the same as in Equation (8). The emission spectrum is a record of the wavelength dependence of the fluorescence intensity. The emission spectrum of perylene is shown in Figure 5. The symmetry between the perylene excitation and emission spectra is observed in the spectra of many fluorophores. It is attributed to similarities in the energy level spacing and transition probabilities in the ground and excited states (the Mirror Image law).

As Figures 4 and 5 illustrate, fluorescence excitation and emission spectra of molecules in solution are quite broad. Bandwidths over 50 nm are common.

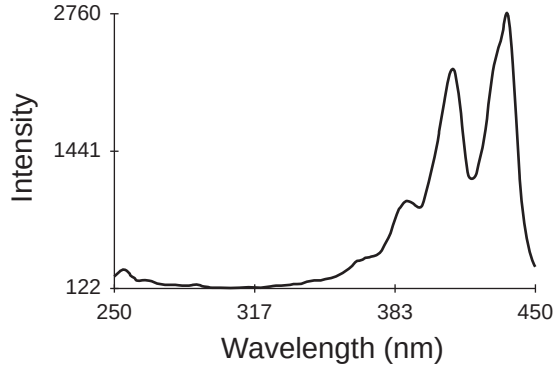

Figure 4 Uncorrected fluorescence excitation spectrum of perylene in ethanol monitored at 437 nm.

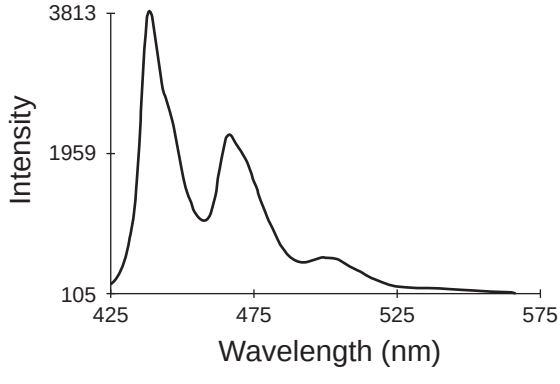

Figure 5 Uncorrected fluorescence emission spectrum of perylene in ethanol excited at 415 nm.

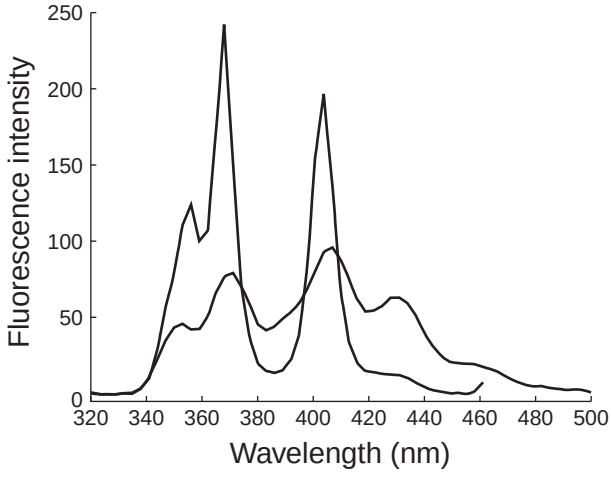

Figure 6 Fluorescence emission spectra (excitation wavelength fixed at 335 nm) and fluorescence synchronous spectra (excitation wavelength scanned from 300 nm to 475 nm) of an anthracene/perylene mixture in ethanol.

Consequently, spectral overlap is a primary limitation to accurate fluorescence analysis of multicomponent samples. In synchronous spectra, the fluorescence intensity is monitored as both the excitation and emission wavelength (monochromator) settings are changed.[17] The resulting spectra can have reduced bandwidths and improved selectivity, as the synchronous spectrum in Figure 6 illustrates. Derivative spectra[18] can also exhibit narrowed bandwidths. However, the resolution enhancement can be offset by the concomitant decrease in the signal-to-noise ratio.

As in the case of absorbance measurements, analyte quantitation is the goal of most steady-state fluorescence intensity analysis. When the species of interest is not intrinsically fluorescent, fluorescent groups can be attached covalently or noncovalently, then used to locate and/or quantitate the analyte. In most routine analyses, the analyte concentration is proportional to a single fluorescence intensity value. The ratio of the fluorescence measured at two wavelengths also can be correlated to analyte concentration.[19] This approach is not as widely used as the single channel method, but has the additional benefit of reducing measurement vulnerability to source fluctuations and matrix effects. Calibration and quantitation using the entire excitation or emission spectrum, which can permit simultaneous multicomponent analysis and contaminant detection, is gaining acceptance but is not widely practiced.[12,20]

2.2.2 Time-resolved Intensity Measurements

The kinetics of photoinduced processes monitor dynamic processes that occur on the timescale of the fluorophore lifetime or the effect of the solvent on fluorophore dynamic properties. Probe methods are widely used to characterize macromolecules such as surfactants, polymers,[9] proteins,[21] nucleic acids[22] and lipids.[23]

In time-correlated single photon counting[24] and other time-resolved emission decay measurements, the emission intensity observed at the detector is the convolution of the excitation pulse and impulse response, as shown in

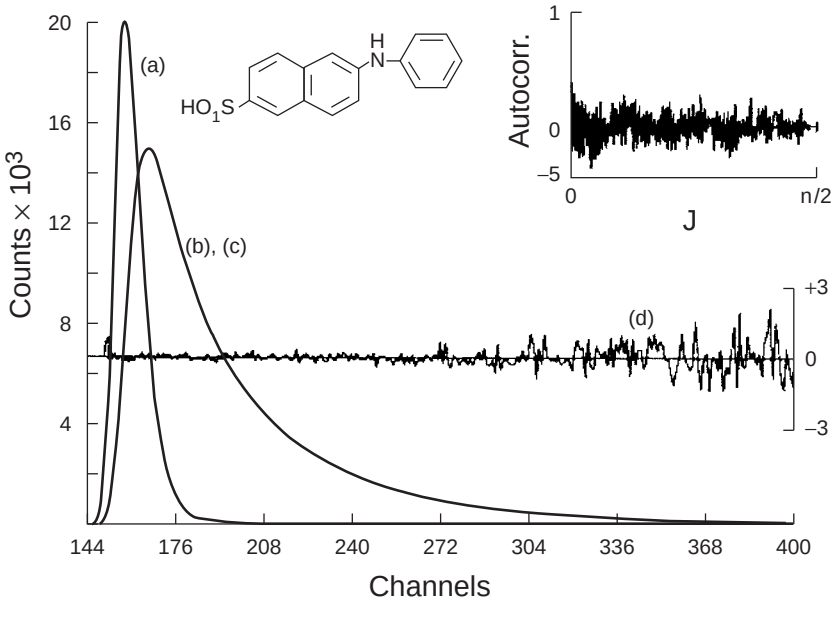

Figure 7 Time-resolved emission decay of 2,6-toludinylnaphthalene sulfonate in phospholipid vesicles: (a) excitation profile; (b) emission decay; (c) simulated decay; (d) residuals. (Reproduced by permission of the Biophysical Society from Easter et al.[25])

Equation (10)

$$I_f(\lambda_{ex}, \lambda_{em}, \mathbf{t}) = k_f V[I_I(\lambda_{ex}, \mathbf{t}) \circ z_0(\lambda_{ex})\zeta(\lambda_{em})\, e^{-\mathbf{t}/\tau_f}] = k_f V \sum_s I_I(\lambda_{ex}, s) z_0(\lambda_{ex})\zeta(\lambda_{em})\, e^{-(\mathbf{t}-s)/\tau_f} \quad (10)$$

where z_0 now equals 2.303 A (λ_{ex}). Figure 7 illustrates the time-dependence of 2,6-toludinylnaphthalene sulfonate in phospholipid vesicles after pulsed excitation.[25]

2.2.3 Frequency-domain Intensity Measurements

The frequency-domain intensity decay measurement produces the Fourier transform of the time-domain measurement using a sinusoidally modulated excitation beam.[26] The signal observed at the detector in a frequency-domain measurement of a single fluorophore in ideal solution is the convolution of the impulse response and excitation function. Specifically, Equation (11) shows

$$\underline{I_f}(\lambda_{ex}, \lambda_{em}, \mathbf{t}) = k_f V[(I_I(\lambda_{ex}) + p_{ex} \sin w\mathbf{t}) \times \circ\, z_0(\lambda_{ex})\zeta(\lambda_{em})\, e^{-\mathbf{t}/\tau_f}] = \overline{I_f}(\lambda_{ex}, \lambda_{em}) + p_{em} \sin w\mathrm{t} + q_{em} \cos w\mathrm{t} = \overline{I_f}(\lambda_{ex}, \lambda_{em})[1 + m_{em} \sin(w\mathrm{t} - \varphi_{em})] \quad (11)$$

at each frequency, w. The emission oscillates at frequency w, but is demodulated and lags behind the oscillating excitation beam (Figure 8a). The phase lag, φ_{em}, and relative modulation, m_{em}/m_{ex}, of the emission at several frequencies can be related to the impulse response through the Fourier transform. Equation (12) shows

$$\mathbf{z}(t) = z_0(\lambda_{ex})\zeta(\lambda_{em})\, e^{-\mathbf{t}/\tau_f} \Longleftrightarrow z_0(\lambda_{ex})\zeta(\lambda_{em}) \times \frac{1/\tau - iw}{1/\tau^2 + w^2} = z_0(\lambda_{ex})\zeta(\lambda_{em}) m_{em}\, e^{-i\boldsymbol{\varphi}_{em}} = \mathbf{q} + i\mathbf{p} = \tilde{\mathbf{z}}(w) \quad (12)$$

The vectors $\boldsymbol{\varphi}_{em}$ and $\mathbf{m}_{em}$ represent, respectively, the phase lags and demodulation factors of the decay at the excitation modulation frequencies, $\boldsymbol{\omega}$, used to acquire the data. The cosine and sine transforms of the decay at the measurement frequencies are the elements of $\mathbf{p}$ and $\mathbf{q}$, respectively. Figure 8(b) and 8(c) illustrates two representations of the frequency-dependent decay of perylene emission. The first is a plot of the phase and modulation versus the modulation frequency. The second is a plot of the Fourier coefficients, $\mathbf{q}$ and $\mathbf{p}$, versus modulation frequency.

2.2.4 Multiexponential Decays

Before leaving this brief introduction to dynamic measurements, a few words about multiple fluorophores and fluorophores in nonideal solutions are in order here. When a fluorophore is solvated by a nonideal solution, the number of processes affecting the deactivation of the

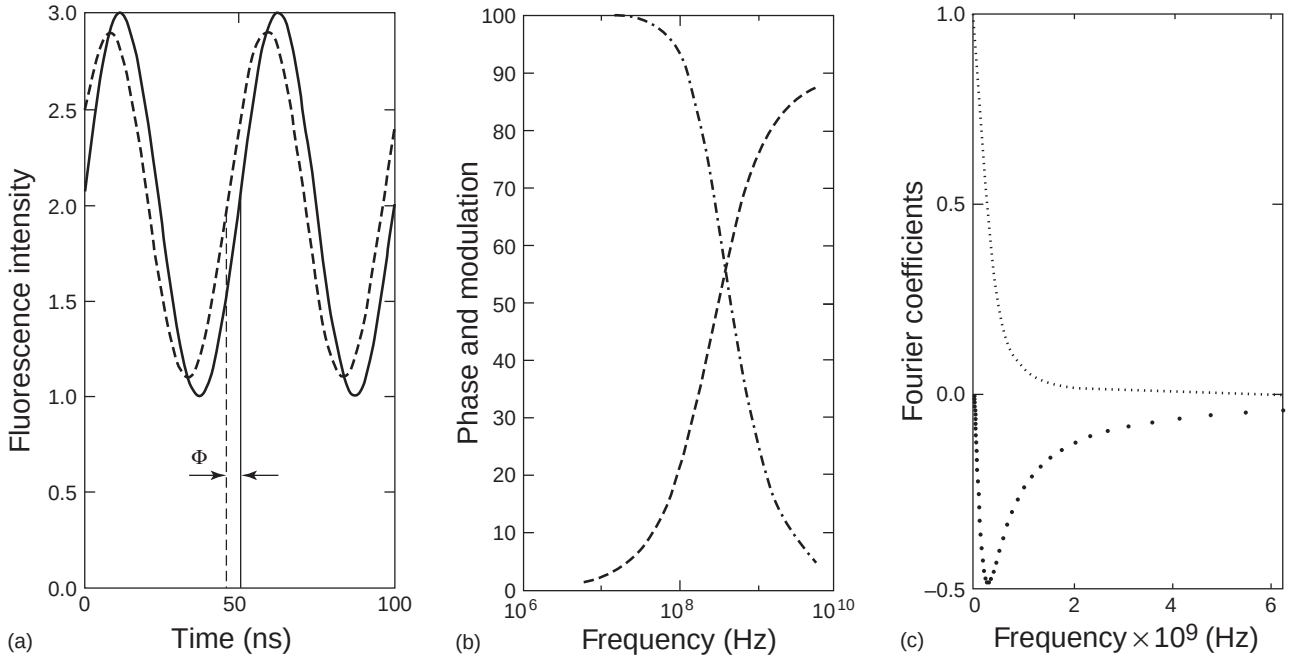

Figure 8 Frequency-domain measurement of perylene emission. (a) Oscillating excitation and demodulated, phase (Φ) retarded perylene emission. (b) Phase and modulation of perylene emission as a function of excitation modulation frequency. (c) Fourier coefficients of decay as a function of excitation modulation frequency.

excited state increases. If the fluorophore is solubilized in a microheterogeneous environment, e.g. in a biomembrane or surfactant aggregate, the fluorophore may decay at a range of rates if the variations in the fluorophore environments lead to differences in relaxation rates. In this case, a distribution of lifetimes will describe the emission decay as shown in Equation (13)

$$\underline{I_{\mathrm{f}}(\mathbf{t})} = \sum_{\tau}^{\mathrm{T}} \alpha(\tau)\mathrm{e}^{-\mathbf{t}/\tau} \tag{13}$$

where the pre-exponential factor, $\alpha(\tau)$, constitutes the lifetime distribution. Figure 9 illustrates the lifetime distribution and time-dependence of the emission of dilute concentrations of pyrene in microheterogeneous media such as lipids or surfactants after pulsed excitation.[27] The monoexponential decay that has the same average decay constant is shown for comparison.

It is often the case that first-order kinetics describe the decay and interactions of the fluorescent components of a sample. If a sample contains n components, Equation (4) becomes Equation (14)

$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{vmatrix} z_1 \\ \vdots \\ z_n \end{vmatrix} = \begin{vmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_{nn} \end{vmatrix} \begin{vmatrix} z_1 \\ \vdots \\ z_n \end{vmatrix} + \mathrm{e}\begin{vmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{vmatrix} = \mathbf{Kz} + \mathrm{e}\mathbf{I} \tag{14}$$

The elements of $\mathbf{z}$ are the instantaneous concentrations of the excited fluorophores. The diagonal elements of $\mathbf{K}$ (the transfer matrix) are the decay constants of the sample components, the off-diagonals are the rates of conversion between the excited sample components. The vector function $\mathrm{e}\mathbf{I}$ is the driving or input function that produces the excited-state populations, i.e. the laser pulse in time-resolved measurements.

For time-domain experiments, assuming excitation powers well below saturating intensities, the solution of this system of equations is a time-dependent vector of excited state concentrations. Equation (15) shows

$$\mathbf{z}(t) = \mathbf{W}\mathrm{e}^{\mathbf{\Lambda}\mathrm{t}}\mathbf{W}^{-1}\mathbf{z}_0 \tag{15}$$

where $\mathbf{W}$ is a matrix whose columns are the eigenvectors of the matrix $\mathbf{K}$ (i.e. $\mathbf{KW} = \mathbf{W\Lambda}$), $\mathrm{e}^{\mathbf{\Lambda}t}$ is a matrix that has the exponential of the product of the observation time and decay constants (eigenvalues of $\mathbf{K}$) on the diagonal, and $\mathbf{z}_0$ is a vector of the initial concentrations of the emitting excited states. Each component of this solution is a combination of the exponentials that decay at rates given by elements of $\mathbf{\Lambda}$. Figure 10 is a simulation of the time-dependence of pyrene monomer and excimer emission after pulsed excitation.[28]

Writing this solution at all observation times in matrix format yields Equation (16):

$$\mathbf{Z}^{\mathrm{T}} = \mathbf{W\Gamma L}^{\mathrm{T}} \tag{16}$$

where the columns of $\mathbf{Z}$ and $\mathbf{L}$ are the decay profiles of the sample components and their constituent exponentials, respectively. The diagonal elements of $\mathbf{\Gamma}$ are concentration factors. When the fluorophores do not interact in the excited state, the $\mathbf{K}$ is diagonal and $\mathbf{W}$ is an identity matrix. Each component decays monoexponentially with a decay rate equal to the inverse excited state lifetime. When there is an excited state reaction, $\mathbf{K}$ is not diagonal nor is $\mathbf{W}$ an identity matrix. Each is a combination of n

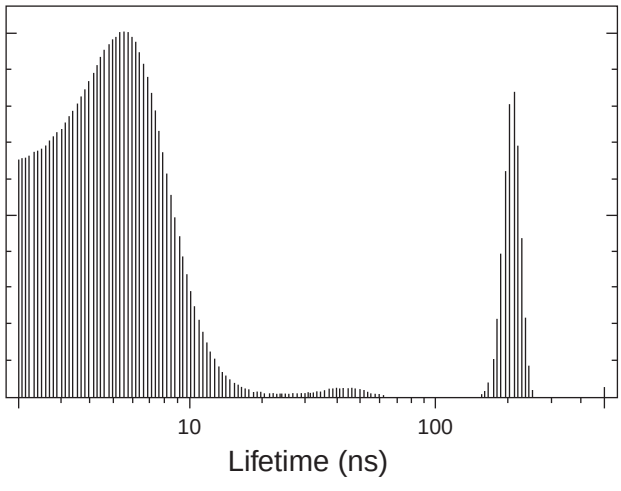

Figure 9 Lifetime distribution of pyrene solubilized in sodium dodecyl sulfate micelles. (Reproduced by permission of the American Chemical Society from Siemiarczuk and Ware.[27])

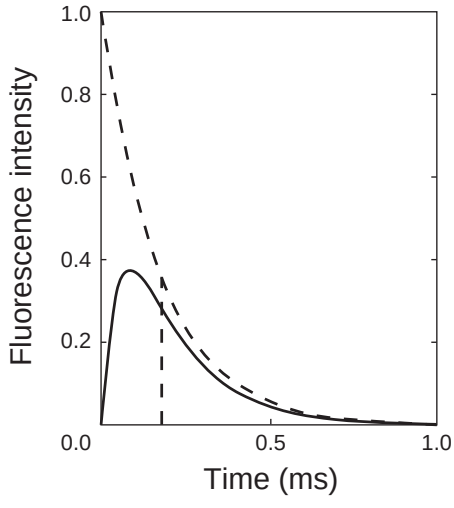

Figure 10 Simulated time-resolved emission decay of pyrene monomer (- - - - -) and diffusion-controlled excimer (———) in isotropic solution. (Rates based on Birks et al.[28])

exponentials that have decay constants that are the eigenvalues of **K** rather than the inverses of the component excited state lifetimes.

2.3 Anisotropy Measurements

2.3.1 Steady-state Anisotropy Measurements

The steady-state fluorescence intensity also depends on the polarization of the excitation. When randomly distributed molecules frozen in rigid solution are excited by a pulse of linearly polarized light, molecules that have their absorption dipole aligned with the electric vector of the excitation have the largest probability of absorbing the light (photoselection). If these molecules fluoresce, the emitted light also is polarized because the positions of the molecules are fixed in the frozen solution. If the solution is not frozen, the emission polarization (or its corollary, the anisotropy) decays with time as Brownian rotations and other molecular motions randomize the orientation of the excited molecules in the solution. The fluorescence anisotropy decay is a function of many sample properties including the fluorophore shape and size, the solvent viscosity and temperature and fluorophore dynamics such as rotations, torsional motion and conformational changes. Consequently, time-dependent anisotropy measurements are a source of information regarding molecular structure and photophysical processes that affect molecular structure.

The emission polarization can be measured using the anisotropy, r, or the polarization, P. These parameters are defined in Equations (17) and (18):

$$\mathrm{r} = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} = \frac{I_{\parallel} - I_{\perp}}{I_{\mathrm{f}}} \tag{17}$$

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{18}$$

where $I_{\parallel}$ and $I_{\perp}$ are the intensities of fluorescence observed through parallel and perpendicular polarizers, respectively. The two measurements are completely analogous but it is more convenient to use the anisotropy in the analysis of multicomponent systems because r is a simple weighted average of the component anisotropies, as shown in Equation (19)

$$\mathrm{r} = \sum_{k}^{K} \alpha_k \mathrm{r}_k \tag{19}$$

where α_k is the fraction of the steady-state intensity emitted by the kth sample component.

When the fluorophore position is immobilized, the anisotropy is a measure of the average angular displacement of the emission dipole from the absorption dipole.[29] When the fluorophore is in motion, the polarization of the emission depends on the rate at which the molecules rotate away from their initial configuration. The steady-state anisotropy is the time-average of the time-dependent anisotropy (see the next section). In the case of spherical molecules, the steady-state anisotropy is given by Equation (20)

$$\mathrm{r} = \frac{\mathrm{r}_0}{1 + \tau_{\mathrm{f}}/\tau_{\mathrm{c}}} = \frac{\mathrm{r}_0}{1 + 6D\tau_{\mathrm{f}}} \tag{20}$$

In this expression, r_0 is the limiting or fundamental anisotropy (the anisotropy observed in the absence of any depolarization), τ_{f} is the fluorescence lifetime (the average lifetime of the excited state after photoexcitation), τ_{c} is the rotational correlation time (the inverse of the fluorophore rotational rate multiplied by six). D is called the fluorophore rotational diffusion coefficient. The rotational diffusion coefficient can be related to the solvent viscosity and molecular volume by the Stokes–Einstein equation, as shown in Equation (21):

$$6D = \frac{RT}{\eta V} \tag{21}$$

where R is the gas constant, T is the temperature, η is the macroscopic solvent viscosity and V is the molar volume of the fluorophore.

2.3.2 Time-resolved Anisotropy Measurements

Time-resolved anisotropy measurements depend on the fluorophore volume, solvent viscosity and temperature just as steady-state measurements do. Investigators have used dynamic anisotropy measurements in a variety of studies. Some examples include measuring solute rotation in biomembranes[30] and polymers, monitoring protein folding,[31] studies of ligand binding to macromolecules[32] and solute interaction with chromatographic surfaces.[33]

In the case of spherical molecules, the anisotropy decay is monoexponential as shown in Equation (22)

$$\mathbf{r(t)} = \mathrm{r}_0 \mathrm{e}^{-\mathbf{t}/\tau_{\mathrm{c}}} \tag{22}$$

The decay constant is the inverse rotational correlation time which is the inverse of six times the rotational diffusion coefficient (see Equations 20 and 21). Since the depolarization depends on the shape of the molecule, substantial efforts were made to extend the theoretical descriptions to include nonspherical molecules so that shape and orientational information could be recovered from polarization measurements. The most complete theoretical treatments accomplish this by solving a generalized rotational diffusion equation and calculating the intensity of fluorescence emission in the directions parallel and perpendicular to the excitation electric vector.[34] These treatments predict that the fluorescence anisotropy decay of an asymmetric molecule is comprised of five exponential components. The generalized

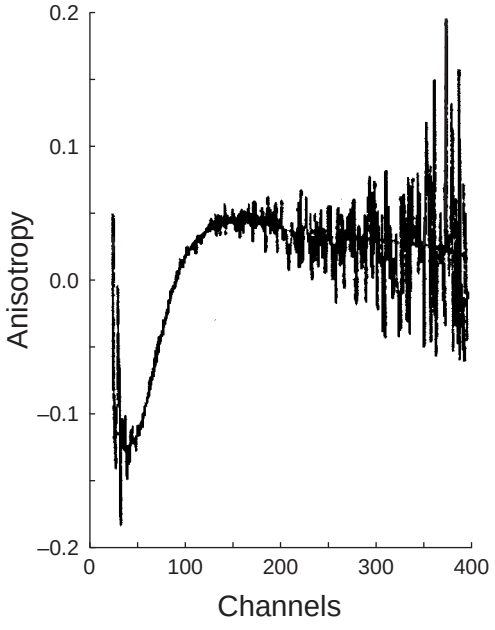

Figure 11 Time-resolved anisotropy decay of perylene in methanol. (Reproduced by permission of the American Institute of Physics from Barkley et al.[36])

time-dependent anisotropy is given by Equation (23)

$$\begin{aligned}\mathbf{r(t)} = 0.3[&4\varsigma_{12}e^{-3(D_3+D)\mathbf{t}} + 4\varsigma_{23}e^{-3(D_1+D)\mathbf{t}} \\ &+ 4\varsigma_{31}e^{-3(D_2+D)\mathbf{t}} + (\Phi+\Psi)e^{-(6D+2\Delta)\mathbf{t}} \\ &+ (\Phi-\Psi)e^{-(6D-2\Delta)\mathbf{t}}]\end{aligned} \tag{23}$$

where D_i, D and Δ are functions of the rotation tensors (rotational coefficients around the molecular axes), ς_{ij}, Φ and Ψ are functions of the fluorophore absorption and emission transition moments. Here, the difficult task of analyzing complex exponentials is compounded by the fact that two pairs of the five decay constants are typically close and the fact that individual fluorophores within a single macromolecule can decay and/or rotate independently.[35] Figure 11 illustrates the time-dependence of perylene anisotropy.[36] Perylene is a symmetric disk-shaped molecule such that $D_1 = D_\parallel$ and $D_2 = D_3 = D_\perp$. Equation (23) reduces to a triple exponential decay.

2.4 Quenching Measurements

2.4.1 Steady-state Quenching Measurements

Quenchers are solutes that attenuate fluorophore emission. Quenching measurements reveal the accessibility of fluorophores to the quencher and thus information about the structure of the medium around the fluorophore.[8] This approach can used to determine if a chromophore is buried deeply within a macrocycle or extended near the surface.[37] Quenching measurements can also be used to measure how well a surfactant or other delivery system sequesters molecules from the bulk solvent.[38]

Quenching mechanisms fall into two primary classes. Static quenchers prevent fluorophores from being excited; dynamic quenchers increase the rate of excited state deactivation through collisional processes. Combinations of both mechanisms can be observed simultaneously in complex media. The dependence of the steady-state fluorescence intensity on quencher concentration in either case is given by Equation (24)

$$\frac{I_f(0)}{I_f([Q])} = 1 - K_Q[Q] \tag{24}$$

$I_f(0)$ is the intensity of fluorescence at zero quencher concentration, K_Q is the quencher constant. In static quenching, K_Q is the equilibrium constant for the complexation reaction between the fluorophore and quencher. In dynamic quenching, K_Q is the product of the fluorophore lifetime and the biomolecular quenching constant. Static and collisional quenching can be distinguished by dynamic measurements, which will be discussed in the next section.

2.4.2 Time-resolved Quenching Measurements

Lifetime measurements distinguish static and dynamic quenching. In the case of static quenching, the duration of the excited state is unchanged. Only the number of fluorophores being excited is reduced, so the ratio of unquenched to quenched lifetimes is always one. Dynamic quenchers reduce the excited state lifetime and, proportionately, the steady-state intensity. Therefore, the quenching constant indicated by the

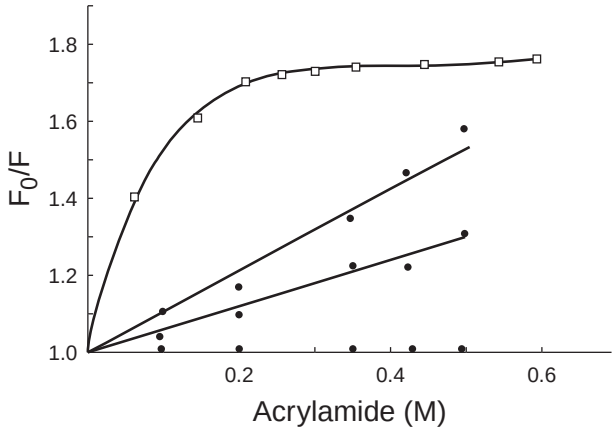

Figure 12 Stern–Volmer plots of dansylated creatine kinase quenching by acrylamide. F_0/F is the ratio of fluorescence intensity in the absence (F_0) and presence (F) of quencher. (Reproduced by permission of the Biophysical Society from Grossman.[39])

ratio of unquenched to quenched lifetimes and the ratio of unquenched to quenched intensities is the same for dynamic quenching. Specifically, Equation (25) shows

$$\frac{I_f(0)}{I_f([Q])} = 1 + k_q\tau_0[Q] = \frac{\tau_0}{\tau} \tag{25}$$

where $I_f(0)$ and τ_0 are the intensity of fluorescence and fluorophore lifetime at zero quencher concentration, respectively. Figure 12 illustrates the difference in the intensity and lifetime ratios. Figure 12 shows the Stern–Volmer plots of the quenching of labelled creatine kinase by acrylamide.[39] The lifetime ratios fall on straight lines, indicating collisional quenching. The intensity ratio exhibits the negative curvature, indicating that some of the fluorescent labels are partially sequestered from the acrylamide in the solution.

2.5 Energy Transfer Measurements

2.5.1 Steady-state Energy Transfer Measurements

Excited state energy can be transferred from one molecule (the donor) to another (the acceptor) without photon emission by the donor. This phenomenon occurs through dipole–dipole interactions when the donor emission spectrum overlaps the acceptor excitation spectrum and the molecules are spatially close ($\sim\leq$50 Å). Specifically, the rate of Forster energy transfer[40] is given by Equation (26)

$$k_{ET} = (8.71 \times 10^{23})\frac{\kappa^2\phi_D J}{d^6 n^4}k_D \tag{26}$$

The variables ϕ_D and k_D represent the fluorescence quantum yield and decay time (inverse lifetime) of the donor, respectively. The variables d and κ^2 represent the distance between and relative orientation of the donor–acceptor pair, respectively. The overlap integral J measures the overlap between the donor emission spectrum and acceptor excitation spectrum. The solvent refractive index is n.

The energy transfer efficiency, E, which is directly related to the donor–acceptor distance, can be measured from steady-state intensities as shown in Equation (27)

$$E = \frac{D_F^6}{D_F^6 + d^6} = 1 - \frac{I_D([A])}{I_D(0)} \tag{27}$$

The Forster distance D_F is the distance at which the energy transfer rate is equal to the donor decay rate so that 50% of donor molecules relax via energy transfer. $I_D(0)$ is the fluorescence emission intensity from the donor in the absence of acceptor, $I_D([A])$ represents donor intensities in the presence of acceptor. The decrease in the donor fluorescence and concomitant increase in acceptor fluorescence as the acceptor concentration is increased is shown in Figure 13.

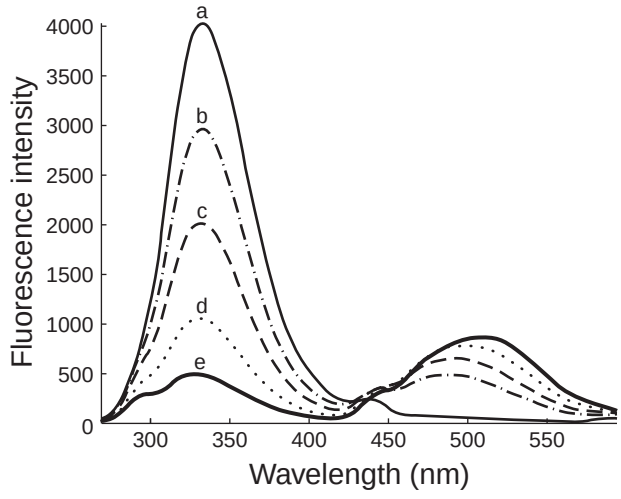

Figure 13 Energy transfer between tryptophan in human serum albumin (HSA) (excited at 295 nm) and bound 3-hydroxyflavone. (a) 40 μM HSA; (b) 40 μM HSA/20 μM 3HF; (c) 40 μM HSA/40 μM 3HF; (d) 40 μM HSA/60 μM 3HF; (e) 40 μM HSA/40 μM 3HF. All in pH 7 HEPES.

2.5.2 Dynamic Energy Transfer Measurements

When an energy transfer donor decays in the presence of an acceptor, the decay of the donor is given by Equation (28)

$$\underline{I_D}([A], \mathbf{t}) \propto \sum_d^{D_{max}} \alpha_d e^{-\mathbf{t}/\tau_0 + k_{ET}\mathbf{t}} = \sum_d^{D_{max}} \alpha_d e^{-\mathbf{t}/\tau_0[1+(D_F/d)^6]} \tag{28}$$

where D_{max} is the maximum donor–acceptor distance (see section 2.5.1 for other definitions). If the transfer occurs in viscous solvent (so that the donor-to-acceptor distance is fixed during the excited state lifetime) and the concentration is low (to prevent intermolecular interactions), the vector α_d represents the distribution distances between the donor and acceptor. The width of the distribution reflects the range of donor-to-acceptor distances in the system. If donor and acceptor are bound to fixed sites, e.g. on a protein, then $\boldsymbol{\alpha}_d$ represents a series of discrete values.

The energy transfer rate (Equation 26) has been called the spectroscopic ruler.[41] Determination of the distance distribution from the donor decay yields mean distances and conformational information regarding the system being monitored. The steady-state and dynamic energy transfer measurements are used to measure the distance between sites in macromolecular systems.[42] They are an important source of information regarding

the structure of systems that have not been crystallized or are not amenable to multidimensional nuclear magnetic resonance (NMR) studies.

2.6 Multidimensional Measurements

2.6.1 Steady-state Multidimensional Measurements

A multidimensional data set is produced when the fluorescence emitted by a sample is monitored as a function of the simultaneous variation of two or more measurement variables. The combination of independent measurement variables produces a selectivity increase that is inversely proportional to the correlation in the constituent responses to the individual variables. In many cases, the selectivity increase is sufficient to enable the analyst to resolve the sample constituents, or monitor an analyte in the presence of contaminants.

The most well-known type of steady-state multichannel fluorescence measurement is the intensity measured as a function of the excitation and emission wavelengths. The resulting map has been called the excitation–emission matrix (EEM)[11] and the total fluorescence spectrum.[43] The EEM of a single component sample is the outer product of the excitation and emission spectra of the fluorophore, as shown in Equation (29)

$$\mathbf{M}(\boldsymbol{\lambda}_{ex}, \boldsymbol{\lambda}_{em}) = \gamma \mathbf{x}(\boldsymbol{\lambda}_{ex})\mathbf{y}(\boldsymbol{\lambda}_{em})^{T} \tag{29}$$

where **M** represents the EEM, **x** the fluorophore excitation spectrum and **y** the fluorophore emission spectrum. The scalar γ is a concentration dependent proportionality constant.

The EEM of a dilute multicomponent sample is the sum of the K EEMs, $\mathbf{M}_k$, of the individual fluorophores that comprise the sample. Equation (30) shows

$$\mathbf{M} = \sum_{k}^{K} \mathbf{M}_k = \sum_{k}^{K} \gamma_k \mathbf{x}_{\cdot k} \mathbf{y}_{\cdot k}^{T} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{Y}^{T} \tag{30}$$

The excitation and emission spectra of the sample components are stored in the columns of matrices **X** and **Y**, respectively. An isometric plot of a mixture of anthracene and perylene is shown in Figure 14.

2.6.2 Time-resolved Multidimensional Intensity Measurements

Multichannel acquisition of dynamic data also produces selectivity increases. All of the experimental variables discussed here, λ_{ex}, λ_{em}, φ_{ex}, [Q], and [A] can be combined with time, or equivalently excitation modulation frequency, to produce multidimensional data sets.

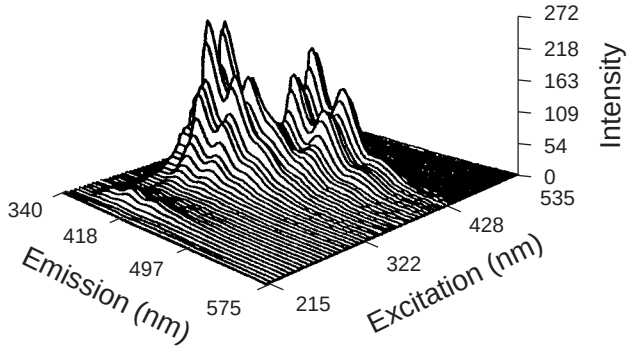

Figure 14 EEM of perylene and anthracene mixture in ethanol.

For example, the emission-time resolved decay matrix, **D**, is the sum of the outer products of the component emission spectra, $\mathbf{y}_l$, and intensity decay profiles, $\mathbf{z}_l$, as shown in Equation (31)

$$\mathbf{D} = \sum_{l}^{L} \mathbf{y}_l \mathbf{z}_l'^{T} = \mathbf{Y}\mathbf{Z}^{T} \tag{31}$$

In dilute solutions of fluorophores that do not interact (**K** in Equation 14 is diagonal), the decay profiles are pure exponentials, $\mathbf{l}_i$, scaled by a concentration factor, γ_u. The decay constants of these exponentials are the fluorophore decay rates. Equation (32) shows

$$\mathbf{D} = \mathbf{Y}\boldsymbol{\Gamma}\mathbf{L}^{T} \tag{32}$$

where the concentration factors, i.e. diagonal elements of $\boldsymbol{\Gamma}$, are the initial excited state concentrations of the sample components. When the sample fluorophores participate in excited state reactions (see section 2.2.4), the decay of each fluorophore is affected by the decay of the other components. In this case, the decay is a combination of exponentials, reflecting the structure of the photokinetic transfer matrix describing the excited state reaction. Equation (33) shows

$$\mathbf{D} = \mathbf{Y}\mathbf{W}\boldsymbol{\Gamma}\mathbf{L}^{T} \tag{33}$$

The columns of the matrix **W** are the eigenvectors of the photokinetic matrix, **K**; the diagonal elements of $\boldsymbol{\Gamma}$ are concentration factors given by $\mathbf{W}^{-1}\mathbf{z}_0$ where $\mathbf{z}_0$ are initial intensities of the emission decays. The decay constants of the exponentials are the eigenvalues of **K**. The time-resolved emission-decay matrix of a mimetic photosynthetic porphyrin[44] is shown in Figure 15.

Multidimensional data acquisition facilitates analysis of multicomponent samples without separation. This is especially useful in kinetic studies[45] because analyses can be performed in the presence of influential sample components. Calibration methods for multidimensional

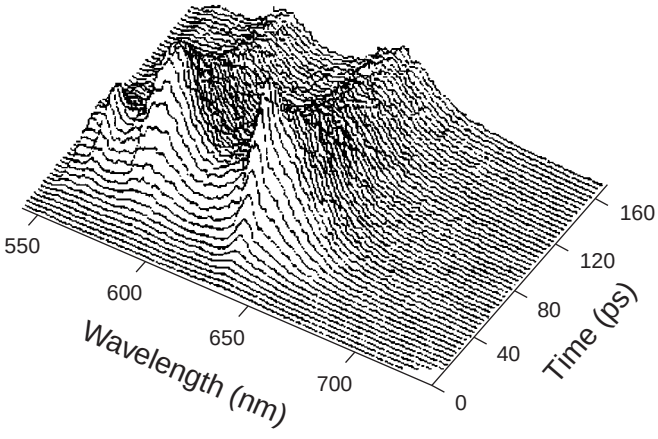

Figure 15 Time-resolved emission decay matrix of photosynthetic porphyrin complex excited at 532 nm. (Reproduced by permission from Osuka et al.[44])

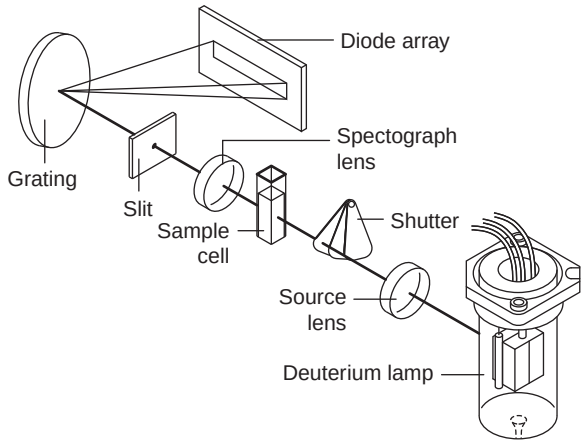

Figure 16 Diagram of UV/VIS diode array spectrophotometer, Hewlett-Packard 8452A. (Reproduced by permission of Agilent Technologies.)

data sets can provide robust quantitation in spite of the presence of sample contaminants.[12]

3 INSTRUMENTATION FOR ULTRAVIOLET/VISIBLE MOLECULAR FLUORESCENCE MEASUREMENTS

3.1 Absorbance Spectrophotometer

Absorbance spectrophotometers are available in several configurations. All of them include a light source, wavelength selection device, sample holder and photon detector. Figure 16 is a schematic of a typical apparatus for measuring steady-state absorbance data and monitoring millisecond kinetics at multiple wavelengths. A deuterium lamp provides broadband excitation from 200 to 800 nm. The light transmitted by the sample is collected by optics, dispersed by a spectrograph and focused onto the surface of a photodiode array. Instrument noise in most absorbance spectrophotometers is a combination of detector dark current (zero percent transmittance (0% T) noise), source flicker (frequency dependent (1/f) noise) and signal shot noise. Error from these sources is normally distributed.

3.2 Time-resolved Spectrofluorometer

Spectrofluorometers are constructed using many of the same components used to construct absorbance spectrophotometers, but the detector–sample axis is 90° to the source–sample axis to minimize collection of scattered light. The classical time-resolved method is time correlated single photon counting. A high repetition rate laser can acquire an entire decay profile in a few seconds. Figure 17 is a schematic of an instrument

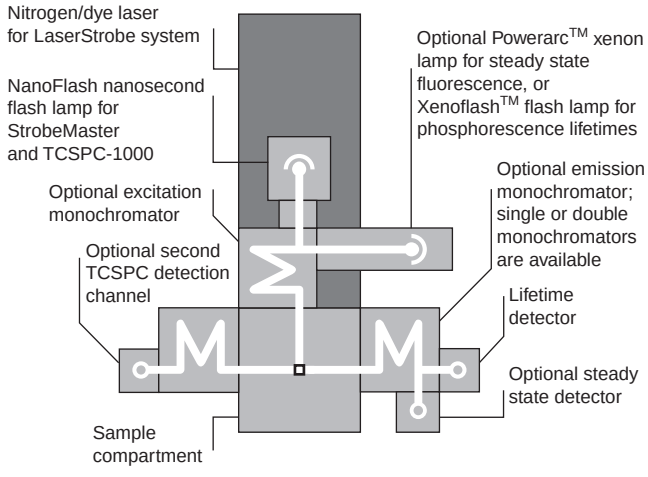

Figure 17 Diagram of general-purpose time-resolved spectrofluorimeter, Photon Technologies International, Inc. TimeMaster. (Reproduced by permission of Photon Technology, Int., Inc.)

equipped to measure steady-state and time-resolved fluorescence. The source depicted is a flash lamp, but pulsed lasers are commonly used in this application. The decay is detected by a gated photomultiplier tube. The monochromator in the detection channel permits the measurement of steady-state emission spectra and wavelength dependent decays. The spectrometer has two emission channels equipped with polarizers (T-format) to enable simultaneous measurement of parallel and perpendicular polarized emission. In this type of instrument, measurement errors are produced by detector dark current and analyte shot noise and source flicker. These noise sources are normally distributed. This contrasts the Poisson distributed errors of time-correlated single photon counting measurements.

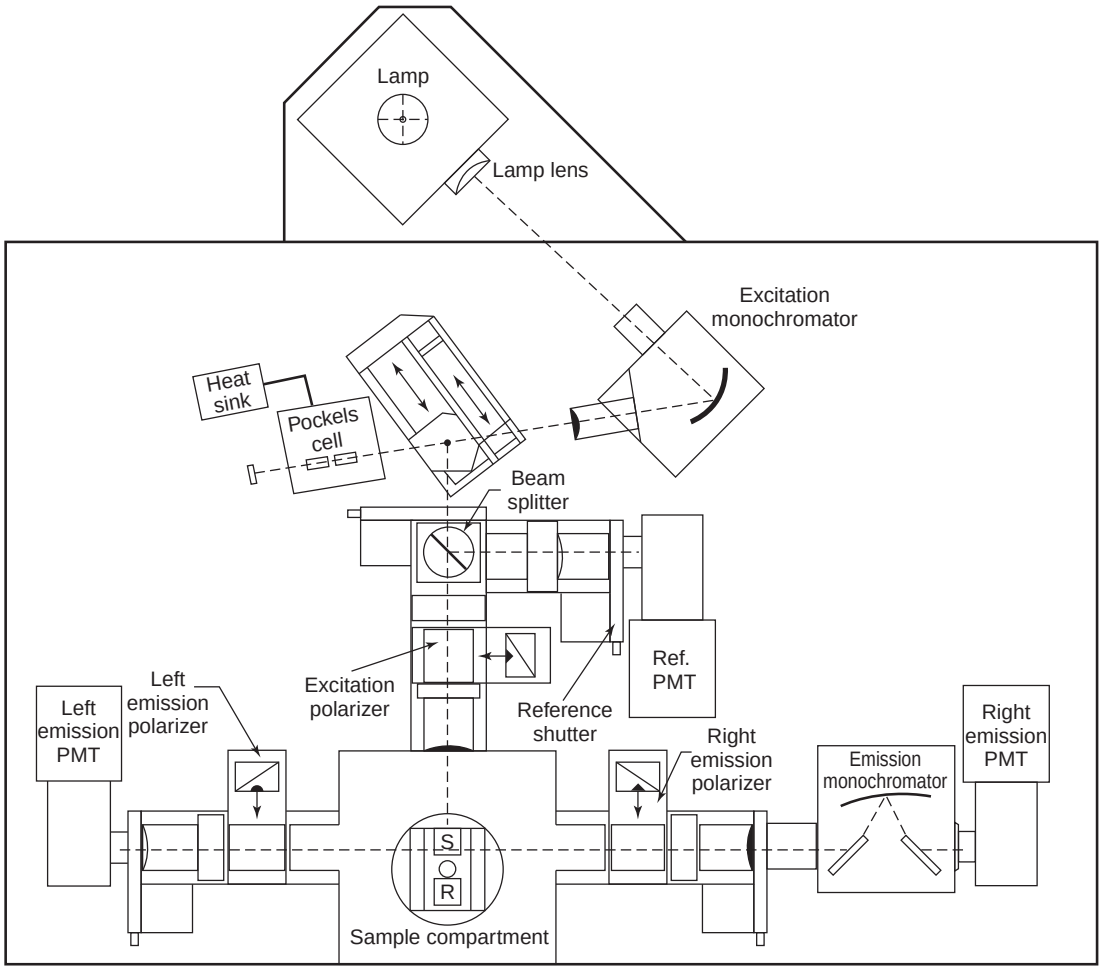

Figure 18 Diagram of general-purpose frequency-domain spectrofluorimeter, ISS, K2. (Reproduced by permission of ISS, IL, USA.)

3.3 Frequency-domain Spectrofluorometer

Figure 18 is a partial schematic of a versatile instrument for measuring steady-state and frequency-domain fluorescence. The source, a xenon arc lamp, is modulated at rates up to 250 MHz by a Pockels cell. The reference excitation beam and sample emission are detected by photomultiplier tubes. These signals are acquired by modulating the gain of both detectors at a frequency close, but not equal to, the source modulation frequency. When this cross-correlation frequency mixes with the signal, the signal carrier frequency is reduced to the difference in the source and detector frequencies. The monochromator in the detection channel permits the measurement of steady-state emission spectra and wavelength dependent decays. The spectrometer has two emission channels equipped with polarizers (T-format) to enable simultaneous measurement of parallel and perpendicular polarized emission. This instrument measures the modulation ratio and phase lag of the emission relative to those of the excitation. Source flicker, radiofrequency pick-up (from synthesizer harmonic noise), synthesizer phase noise and detector dark current are the significant sources of error in this instrument.[46] These sources are normally distributed.

4 DATA ANALYSIS OF ULTRAVIOLET/VISIBLE MOLECULAR ABSORBANCE AND FLUORESCENCE MEASUREMENTS

4.1 Classification of Data Analysis Methods

There are many similarities in the data types and mathematical expressions outlined in section 2. Consequently, three classes of procedure provide most analysis for absorbance and fluorescence data. To recover sample information from steady-state data, we need programs that model the relationship between the spectral property

(I_f, r, etc.) and the sample property of interest ([c], V, etc.). The model then can be used to predict the properties of unknown samples. This is the function of regression methods, which are discussed in section 4.2. In dynamic measurements, the sample properties of interest are related to the decay constants or pre-exponential factors. Therefore, decay analysis methods, discussed in section 4.3, are needed to analyze dynamic data. The multivariate data sets, steady-state and dynamic, can be factored into the responses of the sample components. In some cases, the resolved responses will require additional analysis, e.g. decay analysis. Matrix factorization methods are discussed in section 4.4. Prototypical procedures from each of these classes and the primary concerns of the analyst applying them to absorbance and fluorescence data will be discussed in the following sections.

4.2 Regression Methods

4.2.1 Linear Least Squares Regression

When the ideal (noise-free) instrument response of a sample is a known (or suspected) to be a linear function of a single experimental (independent) variable, the parameter values that define the response model can be determined by linear least squares regression (LLSR). The measured response, **g**, to changes in the experimental variable, **f**, is the combination of a simulated response, **s**, and measurement error **ξ**. A single coefficient, β, defines the contributions of the variable to the simulated response. In general terms, this relationship is given by Equation (34)

$$\mathbf{g} = \mathbf{s}(\beta, \mathbf{f}) + \boldsymbol{\xi} \tag{34}$$

where **g**, **s** and **ξ** are N element vectors. The values of the independent variable are the N elements of the vector, **f**, and β is a scalar. For the sake of a more concrete discussion, let us consider Equation (34) a model of a static quenching study (cf. Equation 24). The quencher concentration is the independent variable, **f**. The ratio of fluorescence intensities in the absence and presence of quencher is the response, **g**. The quenching constant is β.

There are several equivalent formalisms for describing the principles of least-squares methods. The discussions that follow are close to those given by Scales.[47] The simulation, **s**, fits the instrument response, **g**, in the least-squares sense when the parameters, β, minimize the residuals. The residuals are measured by the sum of the squared differences between the response and the simulation, i.e. $\rho = \sum_n^N \xi_n^2$. This sum is minimized when its first derivative with respect to β is zero, as shown in Equation (35)

$$\frac{\partial \rho}{\partial \beta} = \frac{\partial}{\partial \beta}[\mathbf{g} - \mathbf{s}(\mathbf{f}, \beta)]^{\mathrm{T}}[\mathbf{g} - \mathbf{s}(\mathbf{f}, \beta)] = 0 \tag{35}$$

Least squares parameters that meet this criterion have the highest probability of being the correct value, i.e. producing a simulation that fits the instrument response. The distribution of simulated responses around the experimental response reveals the inverse relationship between the experimental error and coefficient probability. Notice that $p(\beta)$ increases as the residuals decrease, as shown in Equation (36)

$$p(\beta) \propto \exp\left\{ \frac{1}{2} \sum_n^N \left(\frac{g_n - s_n(\mathbf{f}, \beta)}{\sigma_n} \right)^2 \right\} \tag{36}$$

The variable σ_n denotes the standard error in the nth (of N) measurement of **g**, i.e. the intensity ratio in our quenching example. Equation (36) reveals why it is so important to ensure that residuals are drawn from a normal distribution when applying least-squares methods. If the residuals are not normally distributed, least-squares parameters will not have the highest probability of representing the data accurately.

The parameter, e.g. quenching constant, which minimizes the residuals, is determined using Equations (37–39):

$$\rho(\beta) = \sum_n^N \xi_n^2 = \boldsymbol{\xi}'\boldsymbol{\xi} = (\mathbf{g} - \mathbf{f}\beta)^{\mathrm{T}}(\mathbf{g} - \mathbf{f}\beta)$$
$$= \mathbf{g}^{\mathrm{T}}\mathbf{g} - 2\beta\mathbf{f}^{\mathrm{T}}\mathbf{g} - \beta^2\mathbf{f}^{\mathrm{T}}\mathbf{f} \tag{37}$$

$$\frac{\partial \rho}{\partial \beta} = -2(\mathbf{f}^{\mathrm{T}}\mathbf{g} + \mathbf{f}^{\mathrm{T}}\mathbf{f}\beta) = 0 \tag{38}$$

$$\hat{\beta} = (\mathbf{f}^{\mathrm{T}}\mathbf{f})^{-1}\mathbf{f}^{\mathrm{T}}\mathbf{g} \tag{39}$$

The circumflex (ˆ) indicates that $\hat{\beta}$ is an estimate. Remember that $\mathbf{f}^{\mathrm{T}}\mathbf{f}$ and $\mathbf{f}^{\mathrm{T}}\mathbf{g}$ are scalar values. They are more familiar than they look. If **g** and **f** were written as mean-centered vectors, i.e. replaced by $\mathbf{g} - \bar{\mathbf{g}}$ and $\mathbf{f} - \bar{\mathbf{f}}$, respectively, and expanded as sums rather than dot products, Equation (39) would be identical to Equation (40), the formula taught to undergraduates for calculating the slope of a line.[48]

$$\hat{\beta} = \frac{N\sum_n^N f_n g_n - \sum_n^N f_n \sum_i^N g_n}{N\sum_n^N f_n^2 - \left(\sum_n^N f_n\right)^2} \tag{40}$$

4.2.2 Multiple Linear Regression

Multiple linear regression (MLR)[20] is the extension of LLSR to the analysis of the responses of several samples to several independent variables. The regression determines the values of parameters that model the dependence of the response on experimental values. The

procedure is based on the idea that the observed sample response is a linear combination of the sample property of interest. For example, the steady-state fluorescence intensity of a K component mixture usually can be modeled as a linear combination of the intensities and concentrations of the K components (Beer's law). At least K simultaneous equations are required to determine the concentrations. If the sample fluorescence is measured at K or more emission or excitation wavelengths, the concentrations of the components can be recovered. In the following section, measured response, **G**, the experimental variable, **F**, the coefficients, **B**, and measurement error Ξ are the matrix counterparts of these quantities in section 4.2.1. Think of **G**, **S** and Ξ as being constructed by arranging the N element responses for M samples into a table (matrix). **F** must be an $N \times K$ matrix and **B** must be a $K \times M$ matrix in order to estimate **G**.

MLR consists of two steps: calibration and quantitation. Calibration consists of measuring the response of standards or samples that can be assayed reliably using alternative methods. In matrix format the calibration model is given by Equation (41)

$$\mathbf{G}_{\text{std}} = \mathbf{F}\mathbf{B}_{\text{std}} + \boldsymbol{\Xi} \tag{41}$$

In the steady-state fluorescence example, **G** is the matrix of fluorescence intensities at **K** or more wavelengths, **B** is the matrix of concentrations and **F** a matrix of spectral profiles.

In the calibration step, the predictor matrix, $\mathbf{F}^{-}_{\text{MLR}}$, is calculated from $\mathbf{G}_{\text{std}}$ and $\mathbf{B}_{\text{std}}$, as shown in Equation (42)

$$\mathbf{F}^{-}_{\text{MLR}} = \mathbf{B}_{\text{std}}\mathbf{G}^{\text{T}}_{\text{std}}(\mathbf{G}_{\text{std}}\mathbf{G}^{\text{T}}_{\text{std}})^{-1} \tag{42}$$

The predictor matrix is written as a special inverse (superscript minus) to emphasize the fact that solving for **B** in Equation (41) is the inverse of regression problems traditionally discussed in statistics tutorials. The values of $\mathbf{F}^{-}_{\text{MLR}}$ calculated by Equation (42) minimize the sum of squares of the elements of Ξ.

In the quantitation step, the properties of a sample response, e.g. component concentrations, are determined using $\mathbf{F}^{-}_{\text{MLR}}$. Equation (43) shows

$$\widehat{\mathbf{B}}_{\text{sample}} = \mathbf{F}^{-}_{\text{MLR}}\mathbf{G}_{\text{sample}} \tag{43}$$

The validity of the model embodied in $\mathbf{F}^{-}_{\text{MLR}}$ is measured by the sum of squares of the prediction error, σ^2_{PRESS}. The prediction error depends on how well $\mathbf{F}^{-}_{\text{MLR}}$ reproduces the values of the calibration set, as shown in Equations (44) and (45)

$$\widehat{\mathbf{B}}_{\text{std}} = \mathbf{F}^{-}_{\text{MLR}}\mathbf{G}_{\text{std}} \tag{44}$$

$$\sigma_{\text{PRESS}} = \sum_{m}^{M}\sum_{k}^{K}[(\mathbf{B}_{\text{std}})_{\text{km}} - (\widehat{\mathbf{B}}_{\text{std}})_{\text{km}}]^2 \tag{45}$$

MLR calibrations perform well under ideal conditions. For multivariate spectral measurements a calibration (standard) set is ideal when it is free of interfering contaminants, matrix effects, interactions between standards, severely overlapped standard spectra and has a good signal-to-noise ratio. The method can be used for simultaneous determination of sample components, but it is usually applied with other regression methods[49] because it is less effective when the component spectra are strongly overlapped.

4.2.3 Principle Components Regression and Partial Least Squares Projection

MLR is called a 'hard' modeling technique because it establishes a correlation between experimental responses and sample properties, i.e. physically observable quantities. MLR incorporates all the information in the calibration set, including measurement error and contaminants, into the calibration matrix. Soft modeling methods such as principal components regression (PCR) and partial least squares (PLS)[20] minimize these limitations of MLR by transferring information to the prediction matrix, $\mathbf{F}^{-}$, more selectively.

All data matrices can be expressed as combinations of orthonormal basis vectors or factors, as shown in Equation (46)

$$\mathbf{G} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\text{T}} = \mathbf{Q}\mathbf{V}^{\text{T}} \tag{46}$$

The columns of **V** summarize the variance, i.e. collect the shapes, distributed along the rows of **G**. The first column of **G** summarizes all the variance of **G**, so it is similar to the average of the rows of **G**. The second column of **V** summarizes all the variance not captured by the first column of **V**. Consequently, it is orthogonal (completely uncorrelated) to the first. This successive extraction of uncorrelated vectors continues until all the variance of **G** is captured. The rows of **Q** are called scores because they reflect how important each column of **V** is to each row of **G**. The number of columns of **V** required to describe the rows of **G** is the rank of **G**. Ideally, the rank of a spectral matrix is K, the number of components absorbing or emitting photons. Experimentally, the rank of a matrix will equal the smaller matrix dimension because of measurement error. In this case, the number of components is the pseudorank of the matrix. The first K columns of **V** (primary principal components) describe the spectral response while the rest (secondary principal components) describe the measurement error. Eliminating the contribution of factors associated with measurement error in Equation (46) reduces the measurement error in the matrix.

In PCR, the response matrix, **G**, is factored into orthogonal basis vectors and scores and the secondary principal components are discarded. The product of the

remaining components is an estimate of **G** that contains significantly lower amounts of measurement error, as shown in Equation (47)

$$\widehat{\mathbf{G}}_{\text{std}} = \breve{\mathbf{Q}}\breve{\mathbf{V}}^{\text{T}} \tag{47}$$

where the superpositioned curve indicates that the secondary columns of **Q** and **V** were discarded prior to the calculation of $\widehat{\mathbf{G}}$. Procedures that estimate the pseudorank are discussed in section 5.2. When all vectors **V** are used to reconstruct $\mathbf{G}_{\text{std}}$, the PCR and MLR results are identical.

The PCR calibration is based on the same linear model used in MLR, but the predictor matrix is calculated from the scores of the response matrix. The calibration model is shown in Equation (48)

$$\mathbf{G}_{\text{std}} = \mathbf{Q}\mathbf{V}^{\text{T}} = \mathbf{F}\mathbf{B}_{\text{std}} + \boldsymbol{\Xi} \tag{48}$$

The predictor matrix is calculated from the primary factors of $\mathbf{G}_{\text{std}}$, as shown in Equations (49) and (50)

$$\mathbf{F}^{-}_{\text{PCR}} = \mathbf{B}_{\text{std}}\breve{\mathbf{V}}^{\text{T}}\breve{\mathbf{Q}}(\breve{\mathbf{Q}}\breve{\mathbf{Q}}^{\text{T}})^{-1} \tag{49}$$

$$\widehat{\mathbf{B}}_{\text{sample}} = \mathbf{F}^{-}_{\text{PCR}}\mathbf{G}_{\text{sample}} \tag{50}$$

If the calibration standards of the steady-state fluorescence intensity example were collected using very dilute solutions, for example, the PCR prediction error would be lower than the MLR error. The predictor matrix estimated by MLR is an unbiased estimator of $\mathbf{G}_{\text{std}}$ but its elements can have large variances, leading to imprecise predictions.[50] In PCR, the predictor is biased (the values of $\mathbf{F}^{-}_{\text{PCR}}$ do not minimize the sum of squares of Ξ), but has a smaller variance and lower prediction error.

PCR constructs the prediction matrix using orthonormal factors of **G** calculated to capture the maximum variance of $\mathbf{G}_{\text{std}}$. These factors are calculated without regard to the calibration variables in $\mathbf{B}_{\text{std}}$. PLS methods factor $\mathbf{G}_{\text{std}}$ into orthogonal factors that are maximally correlated to the calibration variables.[50] This approach minimizes the large errors associated with highly correlated calibration variables. Theoretically, the PLS factors are calculated to minimize the prediction error rather than the residuals of $\mathbf{G}_{\text{std}}$, as shown in Equation (51)

$$\begin{aligned}\widehat{\mathbf{G}}^{\text{T}}_{\text{std}} &= \breve{\mathbf{T}}\breve{\mathbf{C}}^{\text{T}} \\ \text{B}^{\text{T}}_{\text{std}} &= \breve{\mathbf{T}}\breve{\mathbf{P}}^{\text{T}}\end{aligned} \tag{51}$$

Keep this straightforward principle in mind because the PLS algorithm is more difficult to understand than the MLR or PCR algorithms. The connection between the form of the predictor matrix and the calibration results is not obvious.

The following algorithm is called the PLS kernel algorithm. It is especially efficient when the number of responses is large and much larger than the pseudorank (number of variables).[51] After rank estimation, the calculation of **T**, **C** and **P** consists of three steps that are repeated until the factors needed to reconstruct $\mathbf{G}_{\text{std}}$ and $\mathbf{B}_{\text{std}}$ have been determined. (Note: it was impossible to record the algorithm here without using variables that have other meanings elsewhere in the article. The variables **T**, **P**, **C**, **W** and **R** have no connection to their meanings elsewhere in the article.)

1. Calculate the basis vector, $\mathbf{t}_{\cdot 1}$, of $\mathbf{G}_{\text{std}}$ that is maximally correlated to $\mathbf{B}_{\text{std}}$. It is the most prominent eigenvector of $\mathbf{B}^{\text{T}}\mathbf{B}\mathbf{G}^{\text{T}}\mathbf{G}$ (subscripts dropped), given by Equation (52)

$$\mathbf{B}^{\text{T}}\mathbf{B}\mathbf{G}^{\text{T}}\mathbf{G}\mathbf{t}_{\cdot 1} = \mathbf{t}_{\cdot 1}\lambda_{11} \tag{52}$$

   where λ_{11} is the eigenvalue of $\mathbf{G}^{\text{T}}\mathbf{G}\mathbf{B}^{\text{T}}\mathbf{B}$, called the PLS kernel matrix.
2. Calculate $\mathbf{p}_{\cdot 1}$ and $\mathbf{c}_{\cdot 1}$, shown in Equations (53) and (54), which are the scores of the rows of $\mathbf{B}_{\text{std}}$ and $\mathbf{G}_{\text{std}}$ with respect to $\mathbf{t}_{\cdot 1}$.

$$\mathbf{p}_{\cdot 1} = (\mathbf{t}^{\text{T}}_{\cdot 1}\mathbf{B}^{\text{T}})(\mathbf{t}^{\text{T}}_{\cdot 1}\mathbf{t}_{\cdot 1})^{-1} \tag{53}$$

$$\mathbf{c}_{\cdot 1} = (\mathbf{t}^{\text{T}}_{\cdot 1}\mathbf{G}^{\text{T}})(\mathbf{t}^{\text{T}}_{\cdot 1}\mathbf{t}_{\cdot 1})^{-1} \tag{54}$$

   Calculate $\mathbf{w}_{\cdot 1}$, the weight of $\mathbf{t}_{\cdot 1}$ on the part of the rows of $\mathbf{G}_{\text{std}}$ that are correlated to the columns of $\mathbf{B}_{\text{std}}$. Equation (55) shows:

$$\mathbf{w}_{\cdot 1} = \mathbf{B}\mathbf{G}^{\text{T}}\mathbf{G}\mathbf{t}_{\cdot 1} \tag{55}$$

3. Subtract contributions of $\mathbf{t}_{\cdot 1}$ from $\mathbf{G}_{\text{std}}$ and $\mathbf{B}_{\text{std}}$ and return to step 1. Equation (56) shows

$$\begin{aligned}\mathbf{R}_{-1} &= \mathbf{I} - \mathbf{t}_{\cdot 1}\mathbf{t}^{\text{T}}_{\cdot 1} \\ \mathbf{G}^{\text{T}}_{-1}\mathbf{G}_{-1} &= \mathbf{R}_{-1}\mathbf{G}^{\text{T}}\mathbf{G}\mathbf{R}_{-1} \\ \mathbf{B}^{\text{T}}_{-1}\mathbf{B}_{-1} &= \mathbf{R}_{-1}\mathbf{B}^{\text{T}}\mathbf{B}\mathbf{R}_{-1}\end{aligned} \tag{56}$$

   $\mathbf{R}_{-1}$ is a projection matrix that removes contributions of $\mathbf{t}_{\cdot 1}$ from $\mathbf{G}^{\text{T}}\mathbf{G}$ and $\mathbf{B}^{\text{T}}\mathbf{B}$. Subtracting the contributions of $\mathbf{t}_{\cdot 1}$ insures that the subsequent columns of **T** will be orthogonal to $\mathbf{t}_{\cdot 1}$ and each other. The remaining columns (2 through K) of **T**, **P**, **C** and **W** are determined by repeating steps 1–3 until the number of factors indicated by rank estimation have been determined.

The predictor matrix is calculated from the scores of **B** and weights of **T** on **G** and **B**, as shown in Equation (57)

$$\mathbf{F}^{-}_{\text{PLS}} = \mathbf{W}(\mathbf{P}^{\text{T}}\mathbf{W})^{-1}\mathbf{C}^{\text{T}} \tag{57}$$

The correlation between **G** and **B** reduces the error of prediction in PLS regression compared to PCR and MLR. Equations (58) and (59) show

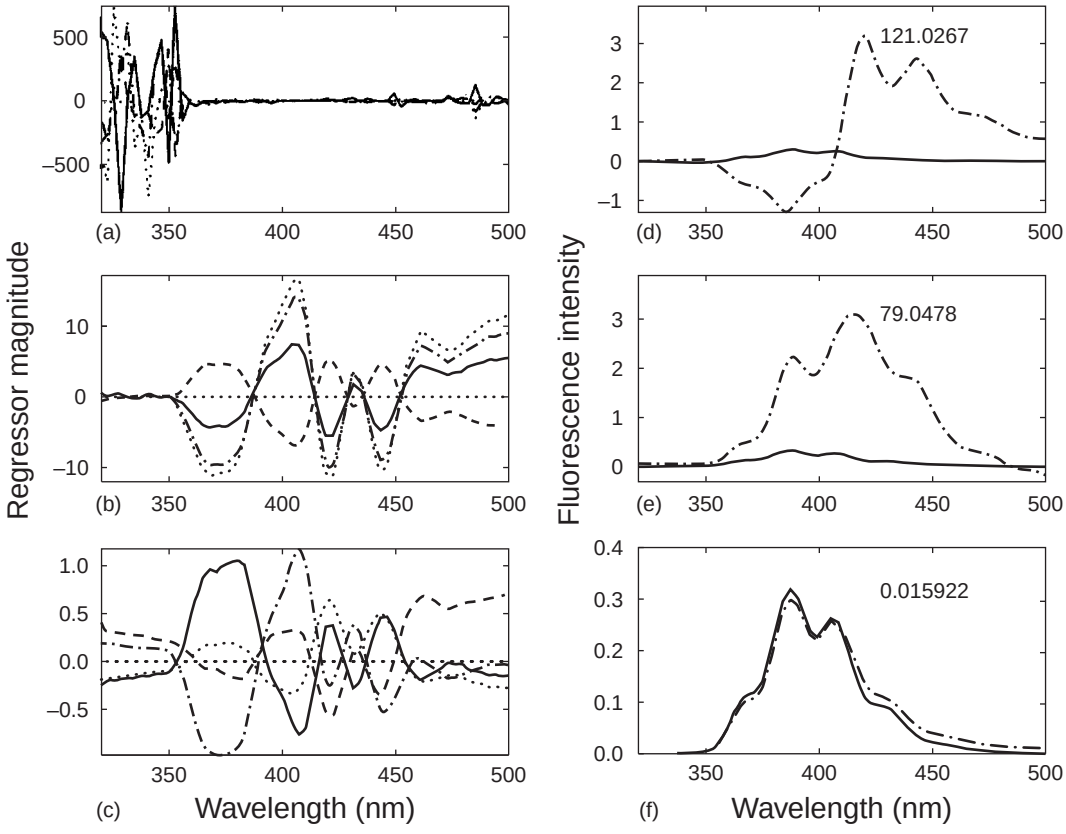

Figure 19 Regression vectors calculated by (a) MLR, (b) PCR, and (c) PLS from a calibration set consisting of four mixture emission spectra distorted by random noise and simulated baseline drifts. Spectra (d–e) estimated by regressors in (a–c) magnitude differences and PRESS values (inserts) show superiority of PLS in cases of collinear variables and instrument drift.

$$\widehat{\mathbf{B}}_{\text{std}} = \mathbf{F}_{\text{PLS}}^{-}\mathbf{G}_{\text{std}} \tag{58}$$

$$\sigma_{\text{PRESS}} = \sum_{m}^{M}\sum_{k}^{K} (\mathbf{B}_{\text{std}})_{\text{km}} - (\widehat{\mathbf{B}}_{\text{std}})_{\text{km}}{}^{2} \tag{59}$$

In the quantitation step, the properties of the sample responses are calculated using $\mathbf{F}_{\text{PLS}}^{-}$, as shown in Equation (60)

$$\mathbf{B}_{\text{sample}} = \mathbf{F}_{\text{PLS}}^{-}\mathbf{G}_{\text{sample}} \tag{60}$$

PLS improves quantitation when the calibration standards deviate from Beer's law because of matrix effects, instrument variations and contaminants. It is also more robust when the calibration standards are substantially correlated. All of these deviations are likely to be observed in the steady-state fluorescence intensity.

4.2.4 Application of Regression to Absorbance and Fluorescence Data

Regression methods can be used to recover any sample property that correlates the observed response to the experimental variable. For example, the correlation between the carcinogen metabolism and the habits (e.g. smoking) of workers exposed to polynuclear aromatic hydrocarbons has been modeled using MLR.[52] Most applications of regression methods in the literature involve simultaneous determination of multiple sample components using multiple wavelength absorbance and fluorescence intensity measurements. Figure 19 depicts the MLR, PCR and PLS regression vectors used to correlate a hypothetical set of mixture spectra to component concentrations. Regression methods have not been widely used by the biochemical/biophysical fluorescence community, though they could be used to calculate apparent viscosity from steady-state anisotropy of spherical probes, quenching constants and donor–acceptor distances. A list of selected applications of regression methods is provided in Table 1.[49,52–67]

4.3 Decay Analysis Methods

The development, application and interpretation of decay analysis methods is an active area of research and discussion. The field is dominated by nonlinear least squares and global target analysis. These methods are described in two of the following sections. The maximum

Table 1 Selected regression applications

	Reference
Correlation between metabolite fluorescence and the habits (e.g. smoking) of workers exposed to polynuclear aromatic hydrocarbons	Sithisarankul et al.[52]
Molecular weights from absorbance data	Dahlen et al.[53]
Protein mixture composition from derivative spectroscopy	Arteaga et al.[54]
Spectrofluorimeter slit width from fluorescence spectra	Norgaard[55]
Resolve effects of mixed solvent constituents on fluorescent dyes	Banerjee et al.[56]
Rapid screening of plutonium solutions	Carey and Wangen[57]
Process stream monitoring	Andrew and Worsfold[58]
Optical purity from absorbance and circular dichroism	Erskine et al.[59]
Simultaneous determinations	
Polynuclear aromatic hydrocarbons in water	Amador-Hernandez et al.,[60] Guiteras et al.,[61] Chang et al.[62]
Metal ion mixtures	Blanco et al.,[49] Rodriguez et al.[63]
Acetylsalicylic acid metabolites in urine	de La Pena et al.[64]
Salicylic acid and diflunisal in human serum	de La Pena et al.[65]
Pesticides in groundwater and soils	Meras et al.[66]
Paracetamol in pharmaceutical preparations	Bouhsain et al.[67]

entropy method (MEM) and multivariate frequency-domain decay analysis (MFDA) method are described because they are model-independent in that neither requires the analyst to select a model to describe the decay prior to the analysis.

4.3.1 Nonlinear Least-squares Fitting

Nonlinear least-squares fitting is a very widely used numerical approach to estimating the contribution of user-specified functions or models to experimental data. There are two types of least squares algorithms for determining parameter values when the model is not a linear function of the experimental variables. Gradient methods use the derivatives of the sum of the squared residuals (ρ) with respect to the parameters ($\boldsymbol{\beta}$) to identify the minimum in ρ. Function comparison (or direct search) methods minimize ρ by systematically searching the $\rho(\beta)$ response surface for the minimum value of ρ. These contrasting approaches are illustrated by the Levenberg–Marquardt[68] modification of the Gauss–Newton method (gradient) and the Nelder and Meade[69] simplex algorithm (function comparison). In general, gradient methods converge more quickly than function comparison methods. On the other hand, function comparison methods are easier to encode because they do not involve derivatives or matrix inverses. In the specific case of nonlinear least squares analysis, the derivatives of ρ are well known, therefore many programs are based on the faster convergence of gradient methods such as Levenberg–Marquardt.

In the Levenberg–Marquardt algorithm, the goal is to adjust the parameter values so that the residuals get smaller at each iteration, as shown in Equation (61)

$$\rho({}^{k+1}\boldsymbol{\beta}) = \rho({}^{k}\boldsymbol{\beta} + {}^{k}\mu{}^{k}\boldsymbol{\nu}) < \rho({}^{k}\boldsymbol{\beta}) \tag{61}$$

where $\rho({}^{k+1}\boldsymbol{\beta})$ is shorthand for $\rho[\mathbf{f}({}^{k+1}\boldsymbol{\beta})]$ and means the $(k+1)$st (next) iterate of ρ. As shown in Equation (62) each new parameter estimate, ${}^{k+1}\boldsymbol{\beta}$, is a combination of the previous iterate, ${}^{k}\boldsymbol{\beta}$, and a search vector, $\boldsymbol{\nu}$.

$${}^{k+1}\boldsymbol{\beta} = {}^{k}\boldsymbol{\beta} + \Delta\boldsymbol{\beta} = {}^{k}\boldsymbol{\beta} + {}^{k}\mu{}^{k}\boldsymbol{\nu} \tag{62}$$

where Δ denotes 'a change in parameter value'. The keys to the reduction in ρ are the length, ${}^{k}\mu$, and direction of the search vector, ${}^{k}\boldsymbol{\nu}$. In the neighborhood of a minimum any function can be expressed as a quadratic. This means that the minimum in ρ can be approximated by its Taylor expansion using the first two derivatives, as shown in Equation (63)

$$\rho({}^{k+1}\boldsymbol{\beta}) \cong \rho({}^{k}\boldsymbol{\beta}) + {}^{k}\mu(\underline{\nabla\rho})^{\mathrm{T}}{}^{k}\boldsymbol{\nu} + \tfrac{1}{2}{}^{k}\mu^{2k}\boldsymbol{\nu}^{\mathrm{T}}(\underline{\underline{\nabla^2\rho}})^{k}\boldsymbol{\nu} \tag{63}$$

where $\nabla\rho$ is the gradient (first derivative of ρ with respect to the parameters, $\boldsymbol{\beta}$) and $\underline{\underline{\nabla^2\rho}}$ is the Hessian, the matrix of second derivatives with respect to the parameters. As the minimization proceeds, the gradient, $\underline{\nabla\rho}$, approaches zero. (This is a central feature in Newton-type methods.) The updated gradient of ρ, ${}^{k+1}\underline{\nabla\rho}$, is given by Equation (64)

$${}^{k+1}\underline{\nabla\rho} \cong {}^{k}\underline{\nabla\rho} + {}^{k}\boldsymbol{\mu}{}^{k}\underline{\underline{\nabla^2\rho}}{}^{k}\boldsymbol{\nu} \tag{64}$$

Since the form of ρ is known explicitly, the gradient and Hessian can be written in terms of the residuals, $\boldsymbol{\xi}$, and parameters, $\boldsymbol{\beta}$. They are shown in Equations (65–67):

$$\rho = (\mathbf{f}-\mathbf{g})^{\mathrm{T}}(\mathbf{f}-\mathbf{g}) = \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}, \tag{65}$$

$$(\nabla\rho)_j = 2\sum_i^n \xi_i \frac{\partial\xi_i}{\partial\beta_j} = 2\mathbf{J}^{\mathrm{T}}\boldsymbol{\xi}, \tag{66}$$

$$(\nabla^2\rho)_{kj} = 2\sum_i^n \left\{ \frac{\partial\xi_i}{\partial\beta_k}\frac{\partial\xi_i}{\partial\beta_j} + \xi_i\frac{\partial^2\xi_i}{\partial\beta_k\partial\beta_j} \right\} = 2\mathbf{J}^{\mathrm{T}}\mathbf{J} + 2\boldsymbol{\xi}\circ\mathbf{H} \tag{67}$$

In these expressions, $\mathbf{J}$ and $\mathbf{H}$ are matrices describing the first and second partial derivatives of the residuals (elements of $\boldsymbol{\xi}$ not ρ) with respect to the parameter

vector elements. Substituting these expressions into Equation (64) and setting to zero yields Equation (68)

$$({}^{k}\mathbf{J}^{\mathrm{T}k}\mathbf{J} + {}^{k}\boldsymbol{\xi} \circ {}^{k}\mathbf{H})^{k}\boldsymbol{\nu} = -{}^{k}\mathbf{J}^{\mathrm{T}k}\boldsymbol{\xi} \tag{68}$$

The search direction $\boldsymbol{\nu}$ is determined by solving this system of equations. The Levenberg–Marquardt algorithm solves for $\boldsymbol{\nu}$ by ignoring the second derivatives of the residuals ($\boldsymbol{\xi}\cdot\mathbf{H}$) and adding a small diagonal increment $\mathbf{J}'\mathbf{J}$ to guard $\boldsymbol{\nu}$ from the consequences of strong correlations in the variables. The search direction is given by Equation (69)

$${}^{k}\boldsymbol{\nu} = ({}^{k}\mathbf{J}^{\mathrm{T}k}\mathbf{J} + {}^{k}\vartheta\mathbf{I})^{-1}(-{}^{k}\mathbf{J}^{\mathrm{T}}\boldsymbol{\xi}) \tag{69}$$

In practice, these equations are not solved by matrix inversion. Gaussian elimination is more reliable and robust. The increment ${}^{k}\vartheta$ must be large enough to insure that $\boldsymbol{\nu}$ leads to a reduction in ρ. Later in the optimization, ${}^{k}\vartheta$ must grow smaller at each iteration in order for the algorithm to converge on the optimal parameters. The size of this increment is determined by a linear search. The resulting descent direction is added to the previous parameter estimate, as shown in Equation (70)

$${}^{k+1}\boldsymbol{\beta} = {}^{k}\boldsymbol{\beta} + {}^{k}\boldsymbol{\nu} \tag{70}$$

The descent direction moves the estimate toward a zero-valued gradient. This process is repeated until $\mathbf{J}^{\mathrm{T}}\boldsymbol{\xi}$ meets some user-specified tolerance close to zero. The associated parameters simulate the experimental data with minimized residuals. If the parameters model decay data, the model is given by Equation (71)

$$\mathbf{g} = \beta_1 \mathrm{e}^{-\beta_2 \mathbf{t}} + \beta_3 \mathrm{e}^{-\beta_4 \mathbf{t}} + \cdots + \boldsymbol{\xi} \tag{71}$$

4.3.2 Global Target Analysis

Nonlinear least-squares algorithms are used in the analysis of a large number of data types, attesting to the power of the method. On the other hand, it is difficult to analyze complex decays accurately and unambiguously. Global target analysis has evolved from efforts to address this limitation.[4,70] The central feature of global analysis is the simultaneous analysis of several related experimental responses. The combined responses are analyzed as a group using a single model. The parameter values for each response are not only consistent with that particular response, but with all the responses in the set. In most cases, the confidence intervals around the parameter values are significantly reduced when compared to parameters recovered from the analysis of single responses. Consequently, parameters determined using global analysis have a higher probability of accuracy than their single response counterparts. Nonlinear least-squares fitting is the method of choice in most global analyses, but this method can be implemented using the optimization method of the analyst's choice. In fact, one of the earliest global analyses[71] used the method of moments[72] to determine the fluorescence lifetime from fluorescence decays.

The reduction in the parameter confidence intervals is a significant advance in itself. The enhanced sensitivity in the 'goodness of fit' observed in global analysis has a second important consequence. Physical model evaluation has become an integral part of many global analysis programs. Parameters can fit experimental responses well, yet have no physically meaningful interpretation. In order to avoid physically meaningless parameter values, global target analysis defines the simulated response in terms of physical invariants. Using this approach, a data set should be analyzed using several alternative models. The applicability of the models that were tested is quantified via error (i.e. residual) analysis. As long as the residuals are consistent with the spectrometer measurement error, the data can be associated with a model. It is important to remember that regardless of the goodness of fit, data analysis can only determine that a model is compatible with the data, not prove that a specific process is producing the emission. In general, parameters (pre-exponentials and decay constants) recovered from complex ($\geq$3 components) exponential decays are not unique. It is possible to discover a model that fits the data well but does not describe the decay process accurately. Bajzer and Prendergast[73] illustrate this fact by showing that multiexponential intensity decays of some single-tryptophan proteins can be described just as well by considering the tryptophan emission quenched by unspecified energy transfer acceptors 'present in or close to the protein matrix' as they are by conformational changes or lifetime distributions.

4.3.3 Maximum Entropy Method

When the decay of fluorophores in locally heterogeneous media are measured, the goal of objective, unambiguous decay analysis is more difficult to achieve. Analyzing such measurements by fitting the data to models selected by the analyst can be risky. If structural nuances are associated with small features in the measured response, those features could be overlooked during fitting to standard distribution functions. The MEM[74,75] is a powerful method for recovering the value and shape of distributed parameters from dynamic spectral data without a priori assumptions about the nature of the parameters.

A linear matrix transformation models the measurement process, as shown in Equation (72)

$$\sum_{p}^{P} \mathrm{h}_{np}\beta_p = \mathrm{g}_n - \xi_n \qquad \text{or} \qquad \mathbf{H}\boldsymbol{\beta} = \mathbf{g} - \boldsymbol{\xi} \tag{72}$$

where $\boldsymbol{\beta}$ is the parameter distribution, e.g. lifetime distribution (Equation 13), $\mathbf{g}$ is the measured response, e.g. donor decay, $\boldsymbol{\xi}$ is the residual vector and $\mathbf{H}$ is the transformation matrix relating the parameter to the measurement. Theoretically, the parameter distribution, $\boldsymbol{\beta}$, could be estimated by solving Equation (72) using matrix inversion methods, but this is impractical for several reasons. First, the sizes of the matrices $\mathbf{H}$, $\boldsymbol{\beta}$ and $\mathbf{g}$ usually prohibit direct solutions. Second, a direct solution would incorporate the influence of the measurement error, $\boldsymbol{\xi}$, into the results. In decay analysis, the transformation, $\mathbf{H}$, is an inverse Laplace transformation. The transformation is ill-conditioned which means that small errors in $\boldsymbol{\xi}$ can lead to large errors in the resulting distribution.

A direct solution also defies the nature of the problem. Theoretically, an infinite number of solutions reproduce the measured response within the limits of the measurement error. The MEM solves Equation (72) by minimizing the information content of the parameter estimate, $S(\boldsymbol{\beta})$, subject to the constraint that the parameter estimate reconstructs the measured data within the limits provided by the measurement error. Equation (73) shows in mathematical notation,

$$\begin{aligned} &\text{maximize} \quad S(\hat{\boldsymbol{\beta}}) = -\sum_{p}^{P} \frac{\hat{\beta}_p}{\beta_0} \log\left(M\frac{\hat{\beta}_p}{\beta_0} - 1\right) \\ &\text{subject to} \quad C(\hat{\boldsymbol{\beta}}) = \sum_{n}^{N} \left(\frac{\mathrm{f}_n - \mathrm{g}_n}{\sigma_n}\right)^2 \end{aligned} \tag{73}$$

where S is the entropy of the distributed parameter, β_0 is the integral of the distributed parameter, σ_n^2 is the variance associated with measuring the nth component of the response and reconstructed data (simulation) vectors, i.e. f_n and g_n, respectively. The simulation is given by Equation (74)

$$\hat{\mathbf{f}} = \mathbf{H}\hat{\boldsymbol{\beta}} \tag{74}$$

The entropy function measures the information content of the parameter vector. Maximizing S, subject to the chi-squared constraint, $C(\beta)$, on the simulation yields the ''simplest'' distribution, i.e. the one that has the lowest information content (highest entropy), from the family of distributions that represent the data within the fitting criterion.

The MEM algorithm published by Reiter[76] uses constrained nonlinear optimization to solve for $\boldsymbol{\beta}$. The algorithm minimizes the entropy iteratively by solving successive quadratic approximations to the problem. This is similar to the approach used by the Levenberg–Marquardt nonlinear least-squares fittings. The difference here is that the solution at each iteration (step) must satisfy the constraints as well as minimize the objective function. The residuals, which form the constraints, are measured by the chi-squared function rather than sum of squares. The algorithm begins with a quadratic approximation to the entropy function and linear approximation to the chi-squared function in terms of the parameters, $\boldsymbol{\beta}$. Equations (75) and (76) show

$$S(\boldsymbol{\beta} + \Delta\boldsymbol{\beta}) \approx S(\boldsymbol{\beta}) + \Delta\boldsymbol{\beta}^{\mathrm{T}}\underline{\nabla S} + \tfrac{1}{2}\Delta\boldsymbol{\beta}^{\mathrm{T}}\underline{\underline{\nabla^2 S}}\Delta\boldsymbol{\beta} \tag{75}$$

$$C(\boldsymbol{\beta} + \Delta\boldsymbol{\beta}) \approx C(\boldsymbol{\beta}) + \Delta\boldsymbol{\beta}^{\mathrm{T}}\underline{\nabla C} \tag{76}$$

where S is the entropy function, $\boldsymbol{\beta} + \Delta\boldsymbol{\beta}$ is the updated parameter vector, $\underline{\nabla S}$ is the entropy gradient (derivative with respect to the parameter elements) and $\underline{\underline{\nabla^2 S}}$ is the entropy Hessian. The Hessian is a matrix comprised of the second derivatives of S with respect to β on the diagonal and the partial derivatives with respect to β_i and β_j on the off-diagonals. C and $\underline{\nabla C}$ represent the value and derivative of the chi-squared function respectively. To minimize S subject to constraints, C, we combine S and C via a Lagrange multiplier, q. Equations (77) and (78) show

$$L = S - qC \tag{77}$$

$$L(\boldsymbol{\beta} + \Delta\boldsymbol{\beta}) \approx L(\boldsymbol{\beta}) + \Delta\boldsymbol{\beta}^{\mathrm{T}}\underline{\nabla L} + \Delta\boldsymbol{\beta}^{\mathrm{T}}\underline{\underline{\nabla^2 L}}\Delta\boldsymbol{\beta} \tag{78}$$

S is maximized and C satisfied when $\partial L/\partial\beta$ and $\partial L/\partial q$, defined in Equations (79) and (80), are zero.

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \underline{\nabla L} + \underline{\underline{\nabla^2 L}}\Delta\boldsymbol{\beta} = \underline{\nabla S} - q\underline{\nabla C} + \underline{\underline{\nabla^2 L}}\Delta\boldsymbol{\beta} = 0 \tag{79}$$

$$\frac{\partial L}{\partial q} = C + \Delta\boldsymbol{\beta}^{\mathrm{T}}\underline{\nabla C} \tag{80}$$

Combining these two conditions yields a system of simultaneous linear equations, as shown in Equation (81)

$$\begin{matrix} \underline{\underline{\nabla^2 L}} & -\underline{\nabla C} & \Delta\boldsymbol{\beta} \\ -\underline{\nabla C^{\mathrm{T}}} & 0 & \mathrm{q} \end{matrix} = \begin{matrix} -\underline{\nabla \mathrm{S}} \\ C \end{matrix} \tag{81}$$

The Hessian matrix, $\underline{\underline{\nabla^2 L}}$, and gradients, $\underline{\nabla C}$ and $\underline{\nabla S}$, can be very large. They can be condensed by re-expressing these quantities in terms of some convenient coordinate axes, e.g. the products of the derivatives and the parameter distribution, so that $\Delta\boldsymbol{\beta}$ can be determined using standard methods for solving simultaneous equations. The parameter vector is updated using the increment shown in Equation (82)

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \Delta\boldsymbol{\beta} \tag{82}$$

In some cases, the increment will be so large, the updated parameter will not be within the region prescribed by the entropy and constraint approximations, i.e. Equations (75) and (76). In such cases, a smaller step replaces the solution of Equation (81). This process is repeated until the convergence criteria are met.

Hypothetically, the entropy and chi-squared gradients are perpendicular at convergence. In practice, this condition may be difficult to achieve. Convergence criteria based on the chi-squared function can produce satisfactory results. The response model for the decay is given by Equation (83)

$$\mathbf{g} = \sum_{p}^{P} \beta_p \mathrm{e}^{-\lambda_p \mathbf{t}} \tag{83}$$

where λ_p is the decay rate associated with the pth element of the parameter distribution, β_p.

4.3.4 Multivariate Frequency-domain Decay Analysis

When the decay of a mixture of fluorophores is governed by first-order kinetics, the emission response to excitation at modulation frequency w is given by the Fourier transform of the time domain solution (see Equation 15). Specifically, as shown by Equation (84)

$$\mathbf{z}(\mathrm{t}) = \mathbf{W}\mathbf{O}\mathrm{e}^{-i(w\mathrm{t}\mathbf{I}+\mathbf{\Phi})}\mathbf{W}^{-1}\mathbf{z}_0 \tag{84}$$

where $\mathbf{W}$, $\mathbf{W}^{-1}$ and $\mathbf{z}_0$ have the same definitions as in the time domain solution, $\mathbf{O}$ is a diagonal matrix of signal amplitudes, $\mathrm{m}_{ii} = \mathrm{m}_{\mathrm{ex}}/(w^2 + \lambda_i^2)^{1/2}$, and $\mathrm{e}^{i\mathbf{\Phi}}$ is a matrix that has the complex phases of the observed signals on the diagonal. Each element of $\mathbf{z}$ is an oscillation that is demodulated and phase shifted relative to the excitation function. Writing this solution as the Fourier transform of the impulse response we have Equation (85)

$$\mathbf{z}(\mathrm{t}) = \tilde{\mathbf{z}}(w)\mathrm{e}^{iwt} \tag{85}$$

where $\tilde{\mathbf{z}}$ represents the complex Fourier transform coefficients of $\mathbf{z}(\mathrm{t})$ at frequency w.

The Fourier transform of impulse response also is prominent in the Fourier transform of the kinetic equation. It is the product of the decay transform and the complex frequency adjusted for the initial conditions, as shown in Equation (86)

$$\frac{\mathrm{d}\tilde{\mathbf{z}}}{\mathrm{dt}} = \mathbf{K}\tilde{\mathbf{z}} = iw\tilde{\mathbf{z}} - \mathbf{z}_0 \tag{86}$$

This means that at each modulation frequency, the frequency-domain decay coefficients, $\tilde{\mathbf{z}}$, are linear functions of the photokinetics transfer matrix. Sugar[77] showed that when the initial conditions are available, $\mathbf{K}$ could be fit directly to experimental frequency-domain decay data.

When Equation (86) is expanded to multiple modulation frequencies, it becomes a system of linear equations, as shown in Equation (87)

$$\mathbf{K}\mathbf{Z}^{\mathrm{T}} = \mathbf{Z}^{\mathrm{T}}\mathbf{\Omega} - \mathbf{Z}_0^{\mathrm{T}} \tag{87}$$

where $\mathbf{Z}$ and $\mathbf{Z}_0$ are matrices constructed from the values of $\tilde{\mathbf{z}}$ and $\mathbf{z}_0$ measured at all the modulation frequencies (the rows of $\mathbf{Z}_0$ are identical). $\mathbf{\Omega}$ is a diagonal matrix of complex modulation frequencies, $\imath w_l$ $(l = 1 : L)$. As long as the number of frequencies, L, is greater than the number of emitting sample components, $\mathbf{K}$ becomes the unknown in an overdetermined system of equations constructed from the complex part of Equation (87). Neal[78] showed that $\mathbf{K}$ is easily calculated from Equation (88) without fitting.

$$\widehat{\mathbf{K}} = (\mathbf{Z}^{\mathrm{T}}\Omega)_{\mathrm{imag}}\mathbf{Z}^{+}_{\mathrm{imag}} \tag{88}$$

The initial intensities can be calculated by using $\mathbf{K}$ in a rearrangement of Equation (87), to give Equation (89)

$$\mathbf{Z}_0^{\mathrm{T}} = \mathbf{Z}^{\mathrm{T}}\mathbf{\Omega} - \mathbf{K}\mathbf{Z}^{\mathrm{T}} \tag{89}$$

4.3.5 Application of Decay Analysis Methods

The global analysis approach is so general that it can be applied with equal facility to dynamic absorbance data, fluorescence lifetime determinations, determination of excited state kinetics and rotational diffusion coefficients. The primary limitations of global analysis are the fact that the analyst must select the photokinetic model to be used in the fit and the limited number of distribution functions that can be used to model the decay.

The MEM has become the standard complement to global/nonlinear least-squares analysis. The MEM recovers a lifetime distribution from fluorescence decays without input from the analyst regarding the number of components or shape of the distribution. In this sense, the method is model independent. It is important to remember that the generalized exponential model embodied by the MEM (Equation 74) does not describe all types of emission decay. For example, fluorophores in polymers and other viscous, anisotropic media emit decays better characterized by stretched exponentials.[79] Decays of fluorophores subjected to excited state reactions are not always amenable to MEM. Decay components can be characterized by negative pre-exponentials (from negative values of $\mathbf{W}$ in Equation 15). The classical MEM program cannot recover distributions that have negative values because the entropy function does not exist in the domain of negative numbers. Newer versions that address this limitation have been described,[80] but are not as widely distributed as the classic version. This limitation also applies to the analysis of anisotropy decays using the MEM. MEM can reveal unusual features in rotational correlation time distributions and improve the 'goodness of fit' criteria. The analyst must use some other method to determine if the anisotropy decay components are associated with negative pre-exponentials.

The MFDA method also is model-independent in the sense that the analyst need not select the number of

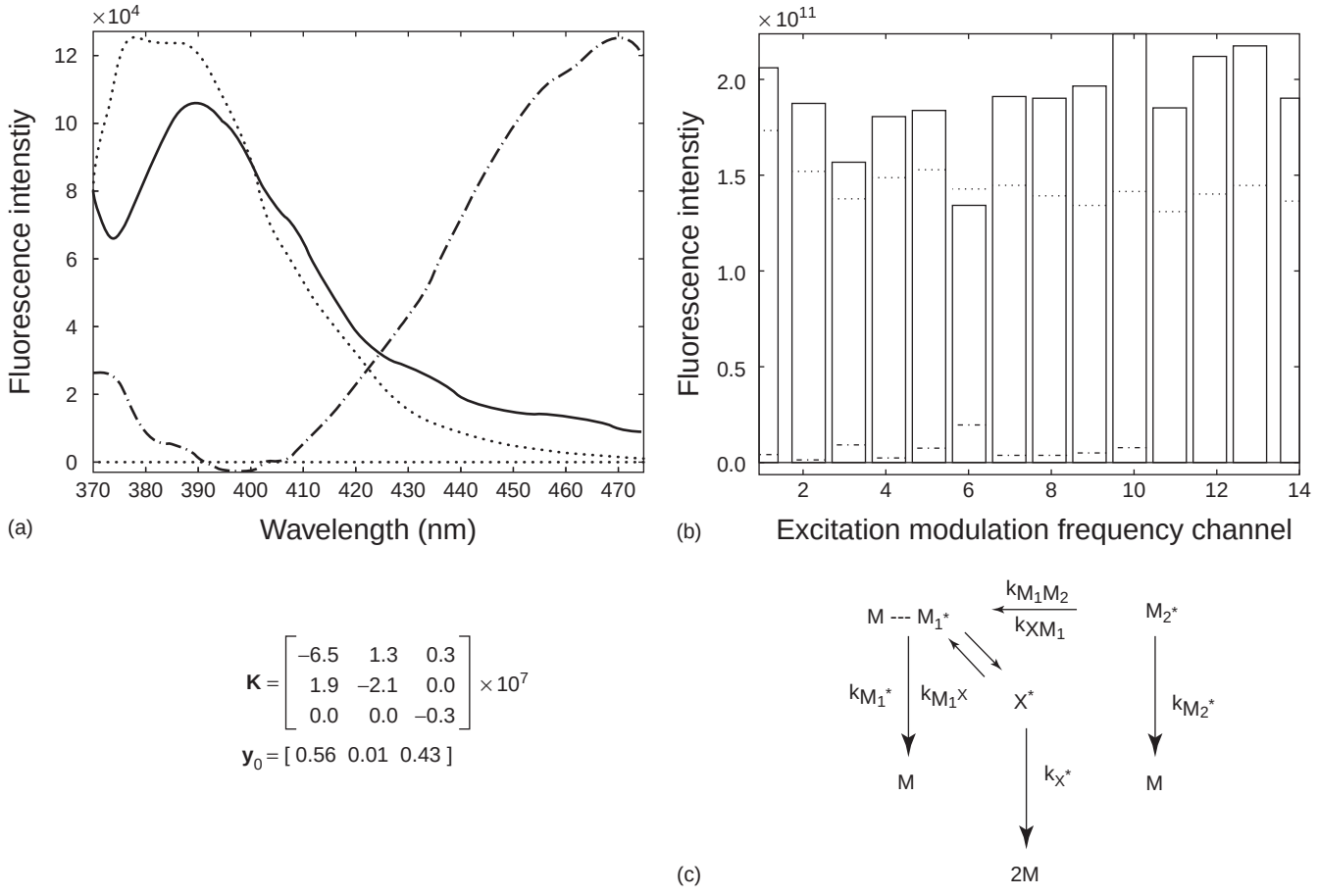

Figure 20 MFDA of emission of pyrene in SDS micelles. The spectra (a) and frequency-domain decays (not shown) of three components: self-associated monomers, $\mathbf{M}_1$ (solid lines); sequestered monomers, $\mathbf{M}_2$ (dotted line); and excimers, X (dash-dotted lines) were resolved from the emission decay matrix (not shown). The photokinetic transfer matrix, **K**, is calculated from the frequency domain decays. The nonzero off-diagonals of **K** indicate a photokinetic mechanism (c). The relative initial intensity vector, $\mathbf{y}_0$, is the scaled average of the initial fluorescence intensity at each modulation frequency (b). (Reproduced by permission of the American Chemical Society from Neal.[78])

system parameters to be determined. MFDA results are not subjective because they are calculated from the measured data without reference to simulated data. Here again, it is important to recognize that there are model limitations inherent in the formalism on which the algorithm is based. MFDA treats decay data as if it were produced by first-order processes, whether they were or not. The MFDA formalism also includes the assumption that a pure decay for each component has been isolated prior to the analysis. The results of analysis of pyrene and pyrene excimer fluorescence decay in surfactant micelles by MFDA are illustrated in Figure 20.

The details of measurement protocols such as the signal-to-noise ratio, sampling rate and sampling cut-off can have significant consequences on all decay analysis methods. If the decay is inadequately sampled or acquired with an insufficient signal-to-noise ratio, the accuracy as well as the precision of the analysis may be unacceptable. James and Ware[81] have studied the signal-to-noise ratio required to distinguish lifetime distributions from closely spaced discrete lifetimes using time-correlated single photon counting. Shaver and McGown[82,83] have studied the impact of measurement conditions on MEM in frequency-domain measurements. Clearly, special attention should be given to the details of the decay acquisition when analysis results are interpreted.[84,85] A list of selected decay analysis applications is provided in Table 2.[14,15,22,31,42,86–98]

4.4 Matrix Factorization Methods

4.4.1 Self-modeling Curve Resolution

Self-modeling curve resolution (SMCR)[99,100] is a numerical procedure for determining the responses of individual components from matrix-formatted data. SMCR resolves

Table 2 Selected decay analysis applications

	Reference
Determination of reaction kinetics	
Olefin epoxidation	Stultz et al.[15]
Chromium(VI) reduction by L-cysteine	Lay and Levina[86]
Phytochrome photoreception	Bischoff et al.[14]
Protein structure and conformation	
Human recombinant interferon α_2	Vincent et al.[87]
Aspartate transcarbamylase	Fetler et al.[88]
Human epidermal growth factor	Gallay et al.[89]
Interdomain distance distributions in phosphoglycerate kinase	Haran et al.[42]
Analysis of protein folding	
Ribonuclease	Plaza del Pino et al.[90]
Unstable mutant of staphylococcal nuclease A	Eftink and Ramsay[31]
Photochemistry and photophysics of fluorophore–biomolecule complexes	
Ethidium bromide–DNA Complexes	Byrne and de Mello[22]
1-Cyanopyrene–1,2-dimethylindole exciplex	Khalil and Boens[91]
Cholesterol and phospholipid bilayers	Bernsdorff et al.[92]
Electrophoresis detection	Li and McGown[93]
Properties of microheterogeneous media	
Micelle size distributions	Siemiarczuk et al.[94]
Micelle–vesicle transition	Seras et al.[95]
Silica adsorption sites	Liu and Ware[96]
Polymer micelle formation	Wilhelm et al.[97]
Reverse micelles	Wittouck et al.[98]

component responses from bilinear matrices, i.e. matrices that are sums of the outer products of the component responses to experimental variables. For example, the steady-state fluorescence EEM is the sum of the outer products of the excitation and emission spectra of the sample components, so it is an example of a bilinear matrix. The matrix can be written as Equation (90)

$$\mathbf{M} = \sum_{k}^{K} \mathbf{M}_k = \sum_{k}^{K} \gamma_k \mathbf{x}_k \mathbf{y}_k^{\mathrm{T}} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{Y}^{\mathrm{T}} \tag{90}$$

where K is the number of sample components, $\mathbf{M}_k$ is the EEM of the kth component and the columns of $\mathbf{X}$ and $\mathbf{Y}$ are the component excitation and emission spectra, respectively. Consequently, each column and row of this matrix is a linear combination of the excitation or emission spectra of the sample components.

The columns and rows of every matrix $\mathbf{M}$ can be expressed as combinations of orthogonal vectors, as shown in Equation (91)

$$\mathbf{M} = \sum_{k}^{K} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}} \tag{91}$$

where the columns of the $N \times P$ matrix $\mathbf{U}$ are orthonormal (i.e. $\mathbf{U}'\mathbf{U} = \mathbf{I}$, the identity matrix) basis vectors for the column space of $\mathbf{M}$, the columns of the $M \times P$ matrix $\mathbf{V}$ are orthonormal basis vectors for the row space of $\mathbf{M}$ and P is the smaller of the two matrix dimensions N and M. The matrix $\boldsymbol{\Sigma}$ is diagonal; the magnitude of each nonzero element, σ_k, reflects the contribution of the product $\mathbf{u}_k\mathbf{v}_k^{\mathrm{T}}$ to the variance of $\mathbf{M}$. The factorization of $\mathbf{M}$ into orthogonal vectors is called the singular value decomposition. The number of basis vectors is equal to P (the smaller matrix dimension) rather than K (the number of fluorophores) because of the superposition of measurement noise on the instrument response. However, only K of these basis vectors describe variance associated with the spectral signal. The remaining $P-K$ vectors describe noise and should be eliminated from the matrices $\mathbf{U}$ and $\mathbf{V}$ prior to further analysis. When the number of fluorescent components is not known prior to the analysis, K can be estimated using a number of statistical and signal processing techniques.[101,102] The determination of K is called rank estimation. This topic is discussed in more detail in section 5.2. From this point, $\mathbf{U}$ and $\mathbf{V}$ represent truncated $N \times K$ and $M \times K$ matrices, respectively.

The data matrix $\mathbf{M}$ is the product of $\mathbf{U}$, $\mathbf{S}$, and $\mathbf{V}$, but $\mathbf{U}$ and $\mathbf{V}$ are inconsistent with physically meaningful spectral responses. Physically meaningful responses (e.g. spectra and concentrations) must be non-negative. These conditions rarely are met by the columns of $\mathbf{U}$ and $\mathbf{V}$. However, since the columns of $\mathbf{M}$ are combinations of both the columns of $\mathbf{U}$ and the columns of $\mathbf{X}$, $\mathbf{U}$ and $\mathbf{X}$ must be linear combinations of each other. Similarly, $\mathbf{V}$ and $\mathbf{Y}$ are linear combinations of each other. Consequently, it is possible to reconstruct $\mathbf{X}$ and $\mathbf{Y}$ from $\mathbf{U}$ and $\mathbf{V}$, respectively. The column vectors $\mathbf{X}$ are constructed by combining multiples of the columns of $\mathbf{U}$ so that they form vectors that meet the mathematical conditions on physically realistic responses. The matrix $\mathbf{Y}$ is similarly reconstructed from $\mathbf{V}$, but with the additional constraint that the product of $\mathbf{X}$, $\boldsymbol{\Gamma}$ and $\mathbf{Y}$ continue to equal $\mathbf{M}$. In many cases, the non-negativity and other constraints do not define unique values for $\mathbf{X}$ and $\mathbf{Y}$. Neal et al.[100] described a variant of the original SMCR algorithm which factors $\mathbf{M}$ into a single solution for $\mathbf{X}$ and $\mathbf{Y}$. This factorization of $\mathbf{M}$ is based on the idea that if the rows and columns are transformed independently, the diagonality of concentration factor matrix, $\boldsymbol{\Gamma}$, reflects the quality of the solution. Specifically, Equation (92) shows

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Pi}\boldsymbol{\Pi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Theta}^{-\mathrm{T}}\boldsymbol{\Theta}\mathbf{V}^{\mathrm{T}} = \hat{\mathbf{X}}\hat{\boldsymbol{\Gamma}}\hat{\mathbf{Y}}^{\mathrm{T}} \tag{92}$$

where the circumflexes indicate estimated values. The steps of the factorization are:

- Calculate **U**, **Σ** and **V** from **M** by singular value decomposition.
- Construct non-negative estimates of **X** and **Y** using transformation matrices, **Π** and **Θ**: $\widehat{\mathbf{X}} = \mathbf{U\Pi}$ and $\widehat{\mathbf{Y}} = \mathbf{V}\Theta$.
- Estimate **Γ** from the estimates of **X** and **Y**: $\widehat{\mathbf{\Gamma}} = \mathbf{\Pi}^{-1}\mathbf{\Sigma\Theta}^{-T}$.
- Minimize the off-diagonal elements of $\widehat{\mathbf{\Gamma}}$.

The values of **Π** and **Θ** that diagonalize the concentration factor matrix, $\widehat{\mathbf{\Gamma}}$ are determined by constrained minimization of the sum of the squared off-diagonals of $\widehat{\mathbf{\Gamma}}$. As this sum is reduced, the **U** and **V** are reconciled to the mathematical constraints on spectral responses, transforming them to $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$. The accuracy of the transformation depends on several factors including the matrix signal-to-noise ratio and the similarity of the component spectral responses that constitute **X** and **Y**. High signal-to-noise ratio and dissimilar component responses promote accurate transformations.

4.4.2 Evolving Factor Analysis

Evolving factor analysis (EFA)[103] is a numerical procedure for partitioning matrices when some features of the profile of one of the responses are known. For example, when the eluent from a chromatographic or electrophoretic separation is monitored spectroscopically, it is known that each system component will have a single maximum in its elution profile. EFA-type algorithms operate on the principle that when the part of the profile associated with a single component is deleted from the matrix the rank of the matrix is reduced by one. In an early EFA algorithm,[103] the eigenvalues of sub-matrices constructed by successively adding rows starting from the beginning of the acquisition are compared with eigenvalues of sub-matrices constructed by successively including rows starting from the end of the acquisition. Plots of the eigenvalues were combined to produce crude elution profiles.

Today, more efficient algorithms are used. For example,[104] the basic structure for the data collected when separations or kinetics are monitored spectroscopically is given by Equation (93)

$$\underline{\underline{A}} = \sum_{k}^{K} \varepsilon_{\cdot k} c_{\cdot k}^{\mathrm{T}} = \mathbf{EC}^{\mathrm{T}} \tag{93}$$

where the columns of **E** represent the absorbance spectra (molar absorptivities times pathlength), and the columns of **C** represent the time evolution of the sample components, respectively. When the rows associated with the kth component are set aside, principal components analysis of the remaining matrix is given by Equation (94)

$$\underline{\underline{A}}^{\circ} = \mathbf{Q}^{\circ}(\mathbf{V}^{\circ})^{\mathrm{T}} \tag{94}$$

The columns of $\mathbf{V}^{\circ}$ are the singular vectors ($\mathbf{Q} = \mathbf{U\Sigma}$) of the time profiles of the remaining spectra. Since the time evolution of the kth compound is not described by $\mathbf{V}^{\circ}$, any part of the rows of **A** that is not described by $\mathbf{V}^{\circ}$ must be the part of the kth component profile. The response of the deleted compound is given by Equation (95)

$$\underline{\underline{A}}(\mathbf{I} - \mathbf{V}^{\circ}(\mathbf{V}^{\circ})^{\mathrm{T}}) = \underline{\underline{A}}_{k} = \boldsymbol{\varepsilon}_{\cdot k}^{+}\mathbf{c}_{\cdot k}^{\mathrm{T}} \tag{95}$$

where $\boldsymbol{\varepsilon}_{\cdot k}^{+}$ represents the part of the kth component spectrum that is uncorrelated with the other spectra and $\mathbf{c}_{\cdot k}$ is the time profile of the kth component. This procedure is repeated for all K components of the sample and the recovered time profiles are used to calculate the spectra of the sample components. Equation (96) shows

$$\mathbf{E} = \underline{\underline{A}}\mathbf{C}^{+} \tag{96}$$

where $\mathbf{C}^{+}$ is the pseudoinverse of **C**.

4.4.3 Generalized Rank Annihilation Method

The generalized rank annihilation method (GRAM) simultaneously factors two matrices, e.g. sample and calibration matrix, without the potential ambiguity of SMCR or the EFA requirement for unique maxima in the component profiles. The sample and calibration standard matrices, **M** and **N**, given in Equations (97) and (98), are combinations of the same row and column basis vectors, e.g. fluorescence excitation and emission spectra.

$$\mathbf{M} = \mathbf{X}\mathbf{\Gamma}_{\mathrm{M}}\mathbf{Y}^{\mathrm{T}} \tag{97}$$

$$\mathbf{N} = \mathbf{X}\mathbf{\Gamma}_{\mathrm{N}}\mathbf{Y}^{\mathrm{T}} \tag{98}$$

$\mathbf{\Gamma}_{\mathrm{M}}$ and $\mathbf{\Gamma}_{\mathrm{N}}$ are diagonal concentration factor matrices. Lorber[105] realized that the component responses, **X** and **Y**, could be determined by combining Equations (97) and (98) into a single generalized eigenvalue problem giving Equation (99)

$$\mathbf{MZ\Lambda} = \mathbf{NZ} \tag{99}$$

where $\mathbf{Z} = \mathbf{Y}^{+}$ and $\mathbf{\Lambda} = \mathbf{\Gamma}_{\mathrm{N}}\mathbf{\Gamma}_{\mathrm{M}}^{-1}$. The eigenvectors of $\mathbf{N}^{-1}\mathbf{M}$ are the columns of the pseudoinverse of the matrix of component responses, **Y**. The column response can be calculated directly from $\mathbf{Y}^{+}$ and **Λ**, as shown in Equation (100)

$$\mathbf{X} = \mathbf{MZ\Lambda} \tag{100}$$

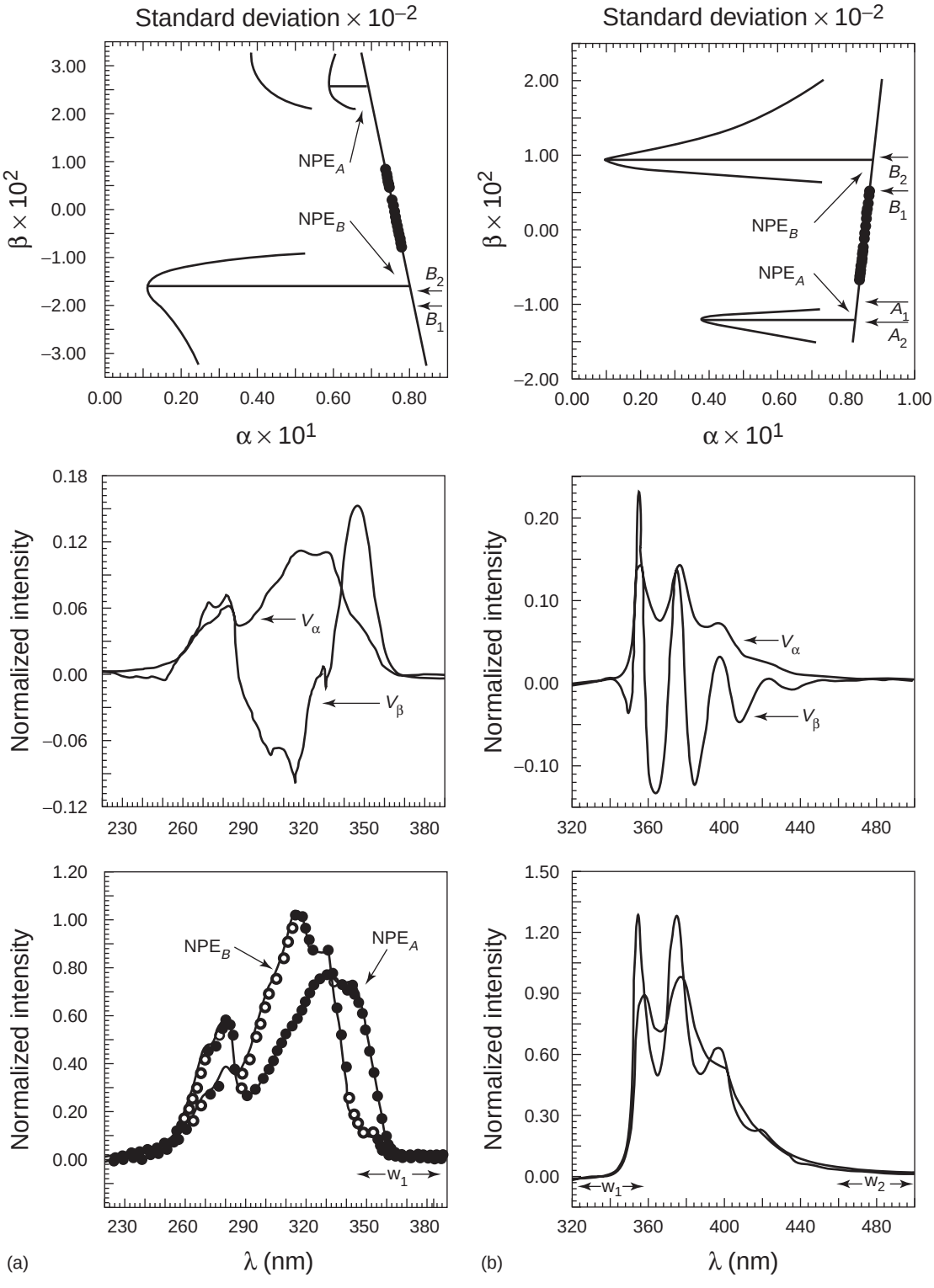

Figure 21 Resolution of pure excitation (a) and emission (b) spectra of *trans*-1-(2-naphthyl)-2-phenylethene conformers resolved from mixture spectra by SMCR. Principal component loadings (top) of mixture spectra fall on single lines indicating two components. The principal components (middle) are linear combinations of the component spectra. The spectra (bottom) are reconstructed from combinations of the singular vectors. (Reproduced by permission of the American Chemcial Society from Saltiel et al.[108])

The original algorithm has been revised to accommodate rectangular matrices and calibration matrices that do not include all the components in the sample matrix.[106]

The acquisition need not stop with two matrices. For example, the emission-decay matrix could be acquired as a function of excitation wavelength. The result is a three-way matrix given by Equations (101) and (102)

$$d_{nmp} = \sum_{k}^{K} x_{nk} y_{mk} z_{pk} + e_{nmp} \tag{101}$$

$$\mathbf{D} = \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \tag{102}$$

The generalized multiplication operator, $\otimes$, combines three column matrices into the three-way matrix. If the number of sample components that absorb and emit are the same, the spectra and decays of all K components can be recovered from the three-way matrix. The model embodied in Equation (101) corresponds to a system in which all the components respond independently to all three sets of experimental variables, e.g. excitation wavelength, emission wavelength, and time. Direct trilinear decomposition (DTD),[107] a variant of GRAM, reduces the P matrix slices to a single pair and calculates $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ without ambiguity using the approach described above.

4.4.4 Application of Matrix Factorization

Spectrometric data matrices are acquired regularly, if not routinely. Multiple wavelength kinetic measurements are used to study dynamic processes in multicomponent systems from photosynthetic reaction centers to polymers. In most cases, these matrices are analyzed using globals or a principal components based variant that fits the decay factors to user specified decay models. SMCR, GRAM and DTD have not been used as widely. GRAM and DTD are projection methods and they are not always robust. On the other hand, GRAM factors matrices without the rotational ambiguity of SMCR. Therefore, it is useful to acquire a second matrix under related conditions when possible. Figure 21 illustrates the *trans*-1-(2-naphthyl)-2-phenyethene conformer spectra resolved by SMCR of quenched mixture spectra arranged in matrix format.[108] Moderate measurement errors can cause dramatic oscillations in the results. Alternating trilinear decomposition[109] is an optimization routine that implements the DTD resolution less stringently and with more stable results. The requirement that the analyst has two samples containing different relative ratios of the same components also may be limiting the use of this method. A second matrix has been generated instrumentally using phase-resolved detection[110] and using perpendicular emission polarizer settings.[111] Theoretically, it should also be possible to produce the second matrix by adding a selective quencher to the sample. A short list of matrix factorization applications is provided in Table 3.[104,111–126]

Table 3 Selected matrix factorization applications

	Reference
Reaction Kinetics	
Photoinitiated diphenylketyl radical reactions	Shield and Harris[112]
Energy migration in allophycocyanin-B trimer	Yamazaki et al.[113]
Coumarin solvation in organic solvents	Yip et al.[114]
Electron and energy-transfer in reaction-center of photosystem II	Holzwarth et al.[115]
Resolution of complex equilibria	
Protonation equilibria of cCMP and polynucleotides	Gargallo et al.[116]
Acid–base equilibria of excited toluidinylnaphthalene sulfonic acid	Karukstis et al.[117]
Acid–base and Copper(II) complexation equilibria of polynucleotides	Gargallo et al.[118]
Mixture Resolution	
Metabolites of acetylsalicylic acid	da Silva and Novais[119]
Carbamate pesticides in high-performance liquid chromatography eluant	Tauler et al.[104]
Resolution of nucleic acid mixtures	Phillips and Georghiou[111]
Hydrocarbons in ocean water	Booksh et al.[120]
Pattern recognition/sample fingerprinting	
Dissolved organic matter in fresh, coastal and marine water	Coble[121]
Characterization of sugar beet leaves, mesophyll and chloroplasts	Cerovic et al.[122]
Humic acids, etc., from corals	Matthews et al.[123]
Medium characterization	
Polarity of phospholipid bilayers	Perochon et al.[124]
Conformational states and changes	
Neuromedin B and delta sleep-inducing peptide in lipid membranes	Polverini et al.[125]
Polyelectrolyte and conformational effects in homopolyribonucleotides	Izquierdo-Ridorsa et al.[126]

5 ERROR ANALYSIS AND RESULT VALIDATION

5.1 Parameter Confidence Intervals

The precision of linear parameters recovered by regression methods can be determined from the variance of the independent variables.[127] For a linear regression Equation (103) shows

$$\hat{\mathbf{B}}_{\text{sample}} = \mathbf{F}^{-}\mathbf{G}_{\text{sample}} \tag{103}$$

The variances of the elements on each column of the parameter matrix **B** are given by the diagonal elements of the following matrix, as shown in Equation (104)

$$_{b_{\cdot i}}\boldsymbol{\Psi} = \mathrm{var}(\mathbf{F}^{-}\mathbf{g}_{\cdot i}) = \mathbf{F}^{-}\mathbf{g}_{\cdot i}(\mathbf{F}^{-}\mathbf{g}_{\cdot i})^{\mathrm{T}} = \mathbf{F}^{-}\mathbf{g}_{\cdot i}\mathbf{g}_{\cdot i}^{\mathrm{T}}(\mathbf{F}^{-})^{\mathrm{T}} \quad (104)$$

In other words, the variance of parameter β_{21} would be $_{b_{\cdot 1}}\boldsymbol{\Psi}_{22}$. It is important to mean center **F** and **G**, that is subtract the column average from every element on a column, prior to calculating the variance matrices. The off-diagonal elements of $\boldsymbol{\Psi}$, i.e. ψ_{ij}, are the covariances of the parameters which reflect the impact of variation in parameter *i* on the variance of parameter *j*.

The precision of parameters recovered by nonlinear least-squares also are functions of the variances of the parameters. Rather than mean center **F** and **G** to calculate $\boldsymbol{\Psi}$, the variance–covariance matrix of the derivatives of the weighted residuals at the minimum is a convenient estimate of the variance in the parameters. The variance–covariance matrix is given by Equation (105)

$$_{\mathbf{J}}\boldsymbol{\Psi} = (\mathbf{J}^{\mathrm{T}}\mathbf{J})^{-1} \quad (105)$$

The square roots of the diagonal elements of $_{\mathbf{J}}\boldsymbol{\Psi}$ provide an estimate for the standard error of the parameters. For many applications of nonlinear least-squares these estimates are sufficient, but they are imperfect. They are the variances in the linear approximation of the parameters. (Remember, there is no contribution from the second derivatives in Equation 64.) The estimate also assumes that the parameters are completely uncorrelated, i.e. orthogonal. Unless all the off-diagonals of $\boldsymbol{\Psi}$ are zero, this criterion is not satisfied. Error estimates without these assumptions can be calculated, for example, using Monte Carlo methods, but they are computationally expensive and are generally reserved for exceptional circumstances.

5.2 Rank Estimation

The first step in the analysis of matrix-formatted spectral data, that is PCR, PLS, MFDA, SMCR, EFA, GRAM and DTD, is the determination of the number of chromophores or fluorophores that give rise to the observed matrix. Measurement errors superimposed on the fluorescence signal inflate the rank of experimental matrices to the smaller of the matrix dimensions. The singular vectors and singular values are ordered by magnitude, so that the first matrix component, i.e. $\mathbf{u}_{\cdot 1}s_{11}\mathbf{v}_{\cdot 1}^{\mathrm{T}}$, describes the largest fraction of the matrix variance and the last component, $\mathbf{u}_{\cdot P}s_{PP}\mathbf{v}_{\cdot P}^{\mathrm{T}}$, describes the smallest. The largest matrix components describe the variance associated with the spectral response, the remaining components describe the measurement error. In the case of high signal-to-noise ratios and moderate spectral overlap, the difference between primary (spectral) and secondary (noise) matrix components is clear in the magnitudes of the singular values. In the case of low signal-to-noise ratios and severely overlapped component spectra, determination of the number of spectral components, i.e. the pseudorank, can be a difficult problem. Determination of the optimal number of matrix factors is critical to accurate analysis. If the number of factors is too low, the model will be inaccurate because responses produced by the sample of interest end up in the residuals. If the number of factors is too high, noise is incorporated into the model, increasing the error of prediction or in the parameters.

There are two types of rank indicators. Some methods compare the variance of the matrix components, i.e. the relative magnitudes of the σ_{ii}. Other methods compare the frequency content of the singular vectors, capitalizing on the predominance of high frequency signal components in measurement noise. The wisest course is to use the consensus result of several methods. The most rigorous method for determining the number of factors contributing to a model is cross-validation.[101] The optimal number of factors is the number that produces the smallest average prediction error when the matrix is reconstructed using column factors generated by systematically removing the columns of the matrix one at a time. Other variance-based methods include reduced eigenvectors, the indicator function and F-tests.[128] The autocorrelation function[129] (see discussion below) and the Fourier coefficient method[102] estimate the rank by determining the presence of non-random variance (low frequency signals) in the singular vectors.

5.3 Residual Analysis

Other measures of fit quality are based on the residuals. The chi-squared function measures the residuals relative to the measurement precision. The reduced chi-squared is given by Equation (106)

$$\chi^2 = \frac{1}{\nu}\sum_{i}^{n}\left(\frac{\mathrm{f}_i - \mathrm{g}(\alpha_i)}{\sigma_i}\right)^2 \quad (106)$$

where ν is the number of degrees of freedom associated with the residuals, i.e. the number of parameters subtracted from the number of all but one of the measured values. The reduced chi-squared is close to unity when the data have been fitted adequately.

The size of the residuals is not the only indicator of fit quality. The residuals also should distribute randomly around the origin when the measurement error is normally distributed; systematic errors indicate that the data, **f**, are not well modeled by the simulation, **g**. The autocorrelation of the residuals is a simple measure of the presence of structure in the residuals. It is

given by the sum of residual covariances, as shown in Equation (107)

$$\zeta_k = \sum_{i=1}^{N-k} \xi_i \xi_{i+k} \tag{107}$$

The analyst chooses the lag, k. Autocorrelations calculated using k equal to 1 and 2 are common. Visual inspection of plots of ξ_i vs ξ_{i+k} also show indications of trends in residuals. The autocorrelation function can also be used for rank estimation by determining the presence of systematic variance in matrix factors.

5.4 Stability

The stability of a calculation is primarily a product of the algorithm used to perform the calculation. Sometimes, a procedure that is conceptually straightforward inflates computer round-off errors to an unacceptably high level. The best insurance against this is to use software based on well-designed and tested algorithms. The stability of a transformation can be determined from the condition number of the associated matrix.[130] For Equation (103), the condition number is given by Equation (108)

$$\text{cond}(\mathbf{F}) = \frac{\sigma_{11}}{\sigma_{KK}} \tag{108}$$

where σ_{KK} and σ_{11} are the last and first singular values of **F**. When the condition number is small, e.g close to unity, the matrix is well-conditioned. Transformations of well-conditioned matrices are stable and not subject to error inflation. If it is found that a data set is ill-conditioned, the analyst should attempt to use variable settings which will improve the condition number. A complete discussion of this topic is beyond the scope of this article, and can be found in references that discuss multivariate calibration in detail.

6 SOFTWARE

Encoding data analysis algorithms can be time-consuming. Fortunately, programs for most of the methods described in this article are available commercially. The names and distributors of the programs are given in the referenced citation. One alternative to buying the programs is to purchase a programming environment that has basic mathematical and statistical operations encoded already. This will reduce the time and effort to encode these algorithms significantly. Several software packages meet this criterion. They vary substantially in price and function. MATLAB (The MathWorks, Natick, MA) is a general-purpose programming language with graphics that is optimized for matrix computations. Extensive toolboxes that have specialized functions for various applications, e.g. signal processing and image analysis, are available. GRAMs/32 (Galactic Industries, Salem, NH) is a package designed specifically for acquisition and analysis of spectral data. The GRAMS environment includes graphics and many pre-programmed functions to simplify data analysis.

7 FUTURE DEVELOPMENT

As technological advances facilitate the acquisition of spectrometric data, the amount and complexity of spectral data will continue. The trend toward increasing dimensionality also will continue. Some of the impetus behind these trends comes from the interest in resolving the components of complex interactions in situ. The tendency for fluorophore emission to respond to the experimental variables independently leads to the increased selectivity of multichannel measurements. It is clear that many of the measurements described in this work will be acquired with spatial resolution using microscopes. This means that the data analysis methods described here will be combined with and complicated by image analysis and reconstruction.

The data analysis methods, specifically generalized rank annihilation and DTD which recover component responses without unambiguity, will continue to be developed and eventually find wide use. Neural networks have not been widely used to analyze absorbance or fluorescence, but they hold promise as multivariate calibration and pattern recognition devices. In pattern recognition devices, neural networks will be used to identify characteristic signals (fingerprints) in complex data sets. In calibration, neural networks can be expected to compensate for strongly correlated variables and matrix effects and may eventually facilitate calibration transfer across spectrometers.

ACKNOWLEDGMENTS

The author's sincere thanks go to Brad Rowe and Linda Staib for their excellent and timely assistance in the preparation of this manuscript.

LIST OF SYMBOLS

The following list is not exhaustive, but contains the main quantities discussed in this article.

$A, \underline{A}, \underline{\underline{A}}$	absorbance	log of ratio of transmitted to incident light, Scalar quantity commonly designated by capital
α_f	response parameter	fluorescence intensity decay pre-exponential
α_c	response parameter	fluorescence anisotropy decay pre-exponential
α_d	response parameter	fluorescence energy transfer decay pre-exponential
b	optical pathlength	distance traveled by light during measurement
β $\boldsymbol{\beta}$, $\mathbf{B}$	response parameters	e.g. sample component concentrations
c, $\mathbf{c}$, $\mathbf{C}$	molar concentration	moles of substance per liter of solution
γ, $\boldsymbol{\gamma}$, $\boldsymbol{\Gamma}$	concentration factor	relative intensity of sample components
d	distance	donor–acceptor distance. Typically represented by r
δ	Dirac function	infinitely narrow pulse
$\mathbf{D}$	data matrix	wavelength–time matrix
D	diffusion coefficient	measure of fluorophore rotation. Scalar quantity commonly designated by capital
D_F	Forster distance	distance of acceptor from donor at which emissive rate and energy transfer rate are equal. Typically represented by R_0. Scalar quantity commonly designated by capital
e	excitation function	e.g. laser pulse for time-resolved measurements
ε, $\boldsymbol{\varepsilon}$, $\mathbf{E}$	molar absorptivity	nominal absorbance of 1 molar solution through 1 cm pathlength
ξ, $\boldsymbol{\xi}$, $\boldsymbol{\Xi}$	residuals	difference between experimental data and simulation
f, $\mathbf{f}$, $\mathbf{F}$	experimental variable	e.g. concentration, time
$\mathbf{g}$	generic instrument model	e.g. fluorescence intensity decay
$\mathbf{H}$	Hessian matrix	matrix of second derivatives
		linear transformation matrix
$I, \underline{I}, \underline{\underline{I}}$	intensity	light intensity. Scalar quantity commonly designated by capital
$I_f, \underline{I_f}, \underline{\underline{I_f}}$	fluorescence intensity	intensity of fluorescence emitted. Scalar quantity commonly designated by capital
$I_\parallel$	parallel polarized intensity	intensity of parallel polarized light. Scalar quantity commonly designated by capital
$I_\perp$	perpendicular polarized intensity	intensity of perpendicular polarized light. Scalar quantity commonly designated by capital
I_I	incident intensity	intensity of incident light. Scalar quantity commonly designated by capital
I_D	donor intensity	intensity of fluorescence emitted by energy transfer donor in the absence of acceptor. Scalar quantity commonly designated by capital
I_{DA}	energy transfer intensity	intensity of fluorescence emitted by energy transfer donor in the presence of acceptor. Scalar quantity commonly designated by capital
I_T	transmitted intensity	intensity of transmitted light. Scalar quantity commonly designated by capital
I_Q	quenched intensity	fluorescence intensity in the presence of quencher. Scalar quantity commonly designated by capital
J	overlap integral	measure of overlap in donor emission and acceptor excitation spectra

$\mathbf{J}$	Jacobian matrix	matrix of first derivatives
k_A	absorption rate	rate of photon absorption
k_f	emissive rate	rate of fluorescence photon emission
k_{ET}	energy transfer rate	rate of excitation energy transfer from donor to acceptor
k_q	bimolecular quenching constant	
κ		orientation of energy transfer donor–acceptor pair
K_Q	generic quenching constant	
$\mathbf{K}$	photokinetic transfer matrix	
$L, \underline{L}$	Lagrange function	objective function in constrained optimization
$\mathbf{L}$	exponential matrix	
λ	wavelength	distance (nm) between maxima in light waves
λ_{ex}	excitation wavelength	distance (nm) between maxima in light waves
λ_{em}	emission wavelength	distance (nm) between maxima in light waves
$\mathbf{\Lambda}$	photokinetic eigenvalues	
m_{ex}	excitation modulation	ratio of oscillating to steady-state component in modulated exciting beam
m_{em}	emission modulation	ratio of oscillating to steady-state component in modulated emission beam
$\mathbf{M}$	excitation–emission matrix	usually represents samples
$\mathbf{N}$	excitation–emission matrix	usually represents standards
n	refractive index	
η	viscosity	
ϑ	degrees of freedom	
$\mathbf{O}$	modulation matrix	ratio of emission and excitation modulations on diagonal
$w, \boldsymbol{\omega}, \mathbf{\Omega}$	angular frequency	$2\pi \times$ frequency of modulated radiation (frequencies on matrix diagonal)
p, $\mathbf{p}$, $\mathbf{P}$	decay sine transform	Imaginary Fourier coefficients of decay
P	polarization	difference in intensity of fluorescence monitored, through parallel and perpendicular polarizers, scaled by sum of polarized intensities. Scalar quantity commonly designated by capital.
ϕ_f	quantum yield	number of fluorescence photons emitted per photon absorbed
φ	emission phase	phase lag of emission produced by sinusoidally modulated excitation
Φ	transition function	length of absorption dipole vector
Ψ	transition function	length of emission dipole vector
$\psi, \boldsymbol{\psi}, \mathbf{\Psi}$	covariance matrix	
q	Lagrange multiplier	modifies constraints in optimization problems
q, $\mathbf{q}$, $\mathbf{Q}$	decay cosine transform	Real Fourier coefficients of decay
r, $\mathbf{r}$, $\mathbf{R}$	anisotropy	difference in intensity of fluorescence monitored, through parallel and perpendicular polarizers, scaled by total emission intensity
R	gas constant	
ρ	residual sum of squares	
s, $\mathbf{s}$, $\mathbf{S}$	simulated response	
$S, \underline{S}, \underline{\underline{S}}$	entropy	regularizing function indicating information content

σ, $\boldsymbol{\sigma}$, $\boldsymbol{\Sigma}$	standard error	diagonal elements of singular value matrix equal square roots of the matrix component variances
ς, $\boldsymbol{\varsigma}$	transition function	cross product of absorption and emission dipoles
t, **t**	time	typically in ns
τ_c	rotational correlation time	average time for rotating fluorophore to trace
τ_D	donor lifetime	average lifetime of energy transfer donor in the absence of acceptor
τ_f	fluorescence lifetime	average lifetime of excited fluorophore
T	transformation	
u, **u**, **U**	column singular vectors	
V	molar volume	
v, **v**, **V**	row singular vectors	
w, **w**, **W**	eigenvectors of **K**	eigenvectors of photokinetic matrix
x, **x**, **X**	excitation spectrum matrix	excitation spectra of components stored on columns
y, **y**, **Y**	emission spectrum matrix	emission spectra of components stored on columns
z, **z**, **Z**	emission decay matrix	decay profiles of components stored on columns
ζ	emission profile	

ABBREVIATIONS AND ACRONYMS

DTD	Direct Trilinear Decomposition
EEM	Excitation–Emission Matrix
EFA	Evolving Factor Analysis
GRAM	Generalized Rank Annihilation Method
LLSR	Linear Least Squares Regression
MEM	Maximum Entropy Method
MFDA	Multivariate Frequency-domain Decay Analysis
MLR	Multiple Linear Regression
NMR	Nuclear Magnetic Resonance
PCR	Principal Components Regression
PLS	Partial Least Squares
SMCR	Self-modeling Curve Resolution
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo • Optical Coherence Tomography

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Clinical Chemistry **(Volume 2)**
Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Luminescence in Environmental Analysis • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
Fluorescence in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
Fluorescence In Situ Hybridization

Peptides and Proteins **(Volume 7)**
Fluorescence Spectroscopy in Peptide and Protein Analysis

Process Instrumental Methods **(Volume 9)**
Ultraviolet/Visible Spectroscopy in Process Analyses

Surfaces **(Volume 10)**
Photoluminescence in Analysis of Surfaces and Interfaces

Chemometrics **(Volume 11)**
Chemometrics

Electronic Absorption and Luminescence **(Volume 12)**
Electronic Absorption and Luminescence: Introduction • Absorption and Luminescence Probes • Fluorescence Lifetime Measurements, Applications of

Kinetic Determinations **(Volume 13)**
Luminescence-based Kinetic Determinations

APPENDIX 1: MATRIX ALGEBRA BASICS

Notation

Matrices are denoted by uppercase, boldface Roman and Greek letters, vectors are denoted by lowercase, boldface Roman and Greek letters. Vectors that are columns of a matrix are denoted using the subscripted, lower case boldface matrix variable, i.e. $\mathbf{m}_{\cdot j}$ denotes the jth column of $\mathbf{M}$. Similarly, vectors that are rows of the matrix $\mathbf{M}$, are denoted $\mathbf{m}_{i\cdot}$, where i represents the ith row. The matrix element on row i in column j is denoted m_{ij}. The variable m is a generic reference to any single element of $\mathbf{M}$. Scalar variables that are not drawn from matrices and vectors are denoted by uppercase and lowercase italic Roman and Greek letters. The uppercase variables may be subscripted for identification purposes. For example, I_f is the fluorescence intensity, where as I_0 is the intensity of the incident excitation beam. Complex variables are indicated using a superpositioned tilde, i.e. $\tilde{\mathbf{M}}$. The superpositioned circumflex, i.e. $\hat{\mathbf{M}}$, indicates an estimated value. As usual, superscript -1, e.g. M^{-1}, denotes the matrix inverse, and the superscript capital T denotes the matrix transpose, $\mathbf{M}^T$. The superscript $+$ denotes the pseudoinverse of a matrix, e.g. M^+. The pseudoinverse[14] is a limited inverse, $\mathbf{M}^+\mathbf{M} = \mathbf{I}$, that can be used to solve linear systems when a matrix is has no classical inverse. Left superscripts, e.g. ${}^k\mathbf{m}$, indicate the iteration number in repetitive procedures.

Definitions

1. scalar: single number m
2. vector: sequence of numbers arranged in a column

$$\mathbf{m} = \begin{matrix} m_1 \\ \vdots \\ m_n \end{matrix}$$

3. matrix: sequence of numbers arranged in a table:
   (a) sequence of vectors
   (b) a linear operator, representation of a linear multivariable function

$$\mathbf{M} = \begin{matrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{m1} & \cdots & m_{mn} \end{matrix} = \begin{matrix} \vdots & \vdots & \vdots \\ \mathbf{m_{\cdot 1}} & \mathbf{m_{\cdot 2}} & \mathbf{m_{\cdot 3}} \\ \vdots & \vdots & \vdots \end{matrix}$$

Special Matrices

1. identity matrix:

$$\mathbf{I} = \begin{matrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{matrix} \qquad \mathbf{Ib} = \mathbf{b}; \quad \mathbf{AI} = \mathbf{A}$$

2. symmetric matrix:

$$\mathbf{A} = \mathbf{A}^T \quad a_{ij} = a_{ji} \quad \text{for all} \quad i, j$$

3. inverse matrix:*

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{matrix} \mathbf{c_{11}} & \cdots & \mathbf{c_{1n}} \\ \vdots & \ddots & \vdots \\ \mathbf{c_{n1}} & \cdots & \mathbf{c_{nn}} \end{matrix}^T \qquad \mathbf{A^{-1}A} = \mathbf{AA^{-1}} = \mathbf{I}$$

* square matrices only
c_{ij} represents the cofactor of a_{ij}
cofactor of a_{ij} is determinant of $\mathbf{A}$ after row i and column j are deleted

4. pseudoinverse: $\mathbf{A}^+ = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T$ (see singular value decomposition below)

$$\mathbf{A}^+\mathbf{A} = \mathbf{I}$$

Operations

1. addition:

$$\begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix} + \begin{matrix} b_1 \\ b_2 \\ b_3 \end{matrix} = \begin{matrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{matrix} = \begin{matrix} c_1 \\ c_2 \\ c_3 \end{matrix}$$

$$\mathbf{a + b = c} \quad \mathbf{A + B = C}$$

$$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} + \begin{matrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{matrix} = \begin{matrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{matrix} = \begin{matrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{matrix}$$

2. subtraction:

$$\begin{matrix} a_1 \\ \vdots \\ a_n \end{matrix} - \begin{matrix} b_1 \\ \vdots \\ b_n \end{matrix} = \begin{matrix} a_1 - b_1 \\ \vdots \\ a_n - b_n \end{matrix} = \begin{matrix} c_1 \\ \vdots \\ c_n \end{matrix}$$

$$\mathbf{a - b = c} \quad \mathbf{A - B = C}$$

3. multiplication:
   (a) scalar multiplication:

$$c \begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix} = \begin{matrix} ca_1 \\ ca_2 \\ ca_3 \end{matrix} = c\mathbf{a}$$

$$c \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} = \begin{matrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{matrix} = c\mathbf{A}$$

   (b) dot (scalar, inner) product:

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{matrix} b_1 \\ b_2 \\ b_3 \end{matrix} = \sum_{k=1}^{3} a_k b_k = \mathbf{a}^T\mathbf{b}$$

$$(1 \times 3)(3 \times 1) \Longrightarrow 1 \times 1$$

if $\mathbf{a}^T\mathbf{b} = 0$, $\mathbf{a}$, $\mathbf{b}$ are orthogonal
$(\mathbf{a}^T\mathbf{a})^{1/2} = \|\mathbf{a}\|$, vector length

(c) outer product:

$$\begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix} \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} = \mathbf{ab}^T = \begin{matrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{matrix}$$

$$= \begin{matrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{matrix} = \mathbf{C}$$

$$(3 \times 1)(1 \times 3) \Longrightarrow (3 \times 3)$$

(d) matrix multiplication:

$$\mathbf{AB} = \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \begin{matrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{matrix}$$

$$= \begin{matrix} \sum_{k=1}^{2} a_{1k}b_{k1} & \sum_{k=1}^{2} a_{1k}b_{k2} \\ \sum_{k=1}^{2} a_{2k}b_{k1} & \sum_{k=1}^{2} a_{2k}b_{k2} \end{matrix}$$

$$= \begin{matrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{matrix} = \mathbf{C}$$

4. Division:

$$\mathbf{A}^{-1}\mathbf{b} = \begin{matrix} a_{11}^{-1} & \cdots & a_{1j}^{-1} & \cdots & a_{1n}^{-1} \\ \vdots & & & & \vdots \\ a_{i1}^{-1} & & \ddots & & a_{in}^{-1} \\ \vdots & & & & \vdots \\ a_{n1}^{-1} & \cdots & a_{nj}^{-1} & \cdots & a_{nn}^{-1} \end{matrix} \begin{matrix} b_1 \\ \vdots \\ b_n \end{matrix}$$

$$= \begin{matrix} \sum_{j}^{n} a_{1j}^{-1}b_j \\ \vdots \\ \sum_{j}^{n} a_{ij}^{-1}b_j \\ \vdots \\ \sum_{j}^{n} a_{nj}^{-1}b_j \end{matrix} = \begin{matrix} c_1 \\ \vdots \\ c_n \end{matrix} = \mathbf{c}$$

or $\mathbf{A}^{+}\mathbf{b} = \mathbf{c}$ if $\mathbf{A}$ is not square

5. matrix transpose:

$$\mathbf{A}^T = \begin{matrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{matrix}$$

6. matrix exponential:

$$e^{\mathbf{A}} = \begin{matrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{matrix} \begin{matrix} e^{\lambda_{11}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{\lambda_{nn}} \end{matrix} \begin{matrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{matrix}^{-1}$$

where $\mathbf{AV} = \mathbf{V\Lambda}$ (see below)
for diagonal $\mathbf{A}$, $\mathbf{V} = \mathbf{I}$

Structure

1. Singular value decomposition:

$$\mathbf{M} = \begin{matrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mk} \end{matrix} \begin{matrix} s_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_{kk} \end{matrix} \times \begin{matrix} v_{11} & \cdots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nk} \end{matrix}^{T} = \mathbf{USV}^T$$

$\mathbf{UU}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_m \qquad \mathbf{VV}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$
$\mathbf{U}$ and $\mathbf{V}$ are orthonormal (orthogonal and normalized to unit length)
if $\mathbf{M}$ is square: $\mathbf{M}^{-1} = \mathbf{VS}^{-1}\mathbf{U}^T$
if $\mathbf{M}$ is not square: $\mathbf{M}^{+} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M} = \mathbf{VS}^{-1}\mathbf{U}^T$

2. Eigenanalysis:

$$\mathbf{MV} = \begin{matrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{m1} & \cdots & m_{mn} \end{matrix} \begin{matrix} v_{11} & \cdots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nk} \end{matrix}$$

$$= \begin{matrix} \lambda_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{kk} \end{matrix} \begin{matrix} v_{11} & \cdots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nk} \end{matrix} = \mathbf{\Lambda V}$$

$$\mathbf{M} = \mathbf{VSV}^{-1}$$

REFERENCES

1. S.A. Soper, B.L. Legendre, 'Single-molecule Detection in the Near-IR Using Continuous-wave Diode Laser Excitation with an Avalanche Photon Detector', *Appl. Spectrosc.*, **52**(1), 1–6 (1998).
2. M.J. Wirth, D.J. Swinton, 'Single-molecule Probing of Mixed-mode Adsorption at a Chromatographic Interface', *Anal. Chem.*, **70**(24), 5264–5271 (1998).
3. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum Publishing Co., New York, 1983.
4. J.M. Beechem, M. Ameloot, L. Brand, 'Global Analysis of Fluorescence Decay Surfaces: Excited-state Reactions', *Chem. Phys. Lett.*, **120**, 466–472 (1985).
5. M. Ameloot, N. Boens, 'Non a priori Analysis of Fluorescence Decay Surfaces of Excited-state Processes. 1. Theory', *J. Phys. Chem.*, **95**, 2041–2047 (1991).
6. J.M. Kovaleski, M.J. Wirth, 'Applications of Fluorescence Recovery After Photobleaching', *Anal. Chem.*, **69**(19), 600A–605A (1997).
7. A. Kawski, 'Fluorescence Anisotropy – Theory and Applications of Rotational Polarization', *Crit. Rev. Anal. Chem.*, **23**(6), 459–529 (1993).

8. P. Lianos, 'Time-resolved Fluorescence Quenching Models in Restricted Geometries', *Heterogen. Chem. Rev.*, **3**(1), 53–63 (1996).
9. F.M. Winnik, S.T.A. Regismond, 'Fluorescence Methods in the Study of the Interactions of Surfactants with Polymers', *Colloid Surface A*, **118**(1–2), 1–39 (1996).
10. N. Ikemoto, B. Antoniu, J.J. Kang, 'Characterization of Depolarization Induced Calcium Release from Sarcoplasmic-reticulum In Vitro with the Use of Membrane-potential Probe', *Biochem. Biophys. Res. Commun.*, **184**(1), 538–543 (1992).
11. I.M. Warner, G. Patonay, M.P. Thomas, 'Multidimensional Luminescence Measurements', *Anal. Chem.*, **57**, 463A–483A (1985).
12. K.S. Booksh, B.R. Kowalski, 'Theory of Analytical Chemistry', *Anal. Chem.*, **66**(15), 782A–791A (1994).
13. D.W. Pierce, S.G. Boxer, 'Dielectric Relaxation in a Protein Matrix', *J. Phys. Chem.*, **96**, 5560–5566 (1992).
14. M. Bischoff, G. Hermann, S. Rentsch, D. Strehlow, 'Ultrashort Processes of Native Phytochrome: Femtosecond Kinetics of the Far-red-absorbing form Pfr', *J. Phys. Chem. A*, **102**, 4399–4404 (1998).
15. L.K. Stultz, R.A. Binstead, M.S. Reynolds, T.J. Meyer, 'Epoxidation of Olefins by [RuIV(bpy)2(py)O]2^+ in Acetonitrile Solution. A Global Kinetic Analysis of the Epoxidation of Trans-stilbene', *J. Am. Chem. Soc.*, **117**, 2520–2532 (1995).
16. R.A. Mathies, 'The Femtosecond Cis–Trans Isomerization in Vision: A Classic Barrierless Photochemical Reaction', in *Ultrafast Processes in Chemistry and Photobiology*, eds. M.A. El-Sayed, I. Tanaka, Y. Molin, Basil Blackwell, Oxford, for IUPAC, 215–224, 1995.
17. C.L. Stevenson, T. VoDihn, 'Analysis of Polynuclear Aromatic Compounds Using Laser Excited Synchronous Fluorescence', *Anal. Chim. Acta*, **303**, 247–253 (1995).
18. R. Compano, R. Ferrer, J. Guiteras, M.D. Prat, 'Flow Injection Method for the Fluorimetric Determination of Zn with 8-(Benzenesulphonamido)quinoline', *Mikrochim. Acta*, **124**(1–2), 73–79 (1996).
19. E.V. Davies, M.B. Hallett, 'Near Membrane Ca^{2+} Changes Resulting from Store Release in Neutrophils: Detection by FFP-18', *Cell Calcium*, **19**(4), 355–362 (1996).
20. K.R. Beebe, B.R. Kowalski, 'An Introduction to Multivariate Calibration and Analysis', *Anal. Chem.*, **59**(17), 1007A (1987).
21. A. Sytnik, I. Litvinyuk, 'Energy Transfer to a Proton-transfer Fluorescence Probe: Tryptophan to a Flavonol in Human Serum Albumin', *Proc. Natl. Acad. Sci. USA*, **93**(23), 12 959–12 963 (1996).
22. C.D. Byrne, A.J. de Mello, 'Photophysics of Ethidium Bromide Complexed to ct-DNA: A Maximum Entropy Study', *Biophys. Chem.*, **70**, 173–184 (1998).
23. Y. Barenholz, T. Cohen, R. Korenstein, M. Ottolenghi, 'Organization and Dynamics of Pyrene and Pyrene Lipids in Intact Lipid Bilayers. Photo-induced Charge Transfer Processes', *Biophys. J.*, **59**, 110–124 (1991).
24. L.M. Bollinger, G.E. Thomas, *Rev. Sci. Instrum.*, **32**, 1044 (1961).
25. J.M. Easter, R.P. DeToma, L. Brand, 'Nanosecond Time-resolved Emission Spectroscopy of a Fluorescence Probe Adsorbed to L-α-egg Lecithin Vesicles', *Biophys. J.*, **16**, 571–583 (1976).
26. E. Gratton, J.R. Lakowicz, B.P. Maliwal, 'Recent Developments in Frequency-domain Fluorometry', *Anal. Instrum.*, **14**(3&4), 193–223 (1985).
27. A. Siemiarczuk, W.R. Ware, 'Pyrene Lifetime Broadening in SDS Micelles', *Chem. Phys. Lett.*, **167**(3), 263–268 (1990).
28. J.B. Birks, D.J. Dyson, I.H. Munro, ' ''Excimer'' Fluorescence II. Lifetime Studies of Pyrene Solutions', *Proc. R. Soc. A*, **275**, 575–588 (1963).
29. F. Perrin, *J. Phys. Radium*, **5**, 497 (1934).
30. C.R. Mateo, M.P. Lillo, J.C. Brochon, M. Martinez-Ripoll, J. Sanz-Aparicio, A.U. Acuna, 'Rotational Dynamics of 1,6-Diphenyl-1,3,5-diphenylhexatriene and Derivatives from Fluorescence Depolarization', *J. Phys. Chem.*, **97**, 3486–3491 (1993).
31. M.R. Eftink, G.D. Ramsay, 'Studies of the Unfolding of an Unstable Mutant of Staphylococcal Nuclease: Evidence for Low Temperature Unfolding and Compactness of the High Temperature Unfolded State', *Proteins*, **28**, 227–240 (1997).
32. L.B. McGown, S.M. Meyerhoffer, 'Critical Micelle Concentration Behavior of Sodium Taurocholate in Water', *Langmuir*, **6**, 187–191 (1990).
33. M.J. Wirth, M.E. Montgomery, Jr, M.A. Green, 'Orientational Dynamics of a Hydrophobic Guest in a Chromatographic Stationary Phase: Effect of Wetting by Alcohol', *Anal. Chem.*, **64**, 1170–1175 (1992).
34. T.J. Chuang, K.B. Eisenthal, *J. Phys. Chem.*, **57**, 5094 (1972).
35. C.N. Bialik, B. Wolf, E.L. Rachofsky, J.B.A. Ross, W.R. Laws, 'Dynamics of Biomolecules: Assignment of Local Motions by Fluorescence Anisotropy Decay', *Biophys. J.*, **75**, 2564–2573 (1998).
36. M.D. Barkley, A.A. Kowalczyk, L. Brand, 'Fluorescence Decay Studies of Anisotropic Rotations of Small Molecules', *J. Chem. Phys.*, **75**(7), 3581–3593 (1981).
37. A.J. Kungl, A.J.W.G. Visser, H.F. Kauffmann, M. Breitenbach, 'Time-resolved Fluorescence Studies of Dityrosine in the Outer Layer of Intact Yeast Ascospores', *Biophys. J.*, **67**, 309–317 (1994).
38. S.L. Neal, M.M. Villegas, 'Phospholipid Vesicles as Organized Media in Spectroluminescence Analysis', *Anal. Chem.*, **67**(15), 2659–2665 (1995).
39. S.H. Grossman, 'Conformational Heterogeneity of Creatine Kinase Determined from Phase Resolved Fluorometry', *Biophys. J.*, **59**, 590–597 (1991).
40. T. Forster, 'Zwischenmolekulare Energiewanderung und Fluoreszenz', *Ann. Phys. (Leipzig)*, **2**, 55–75 (1948).

41. G. Liu, 'Determination of Intramolecular Distance Distribution Functions Using the ''Spectroscopic Ruler''. 1. Theoretical Feasibility', *Macromolecules*, **26**, 1144–1151 (1993).
42. G. Haran, E. Haas, B.K. Szpikowska, M.T. Mass, 'Domain Motions in Phosphoglycerate Kinase: Determination of Interdomain Distance Distributions by Site-specific Labeling and Time-resolved Fluorescence Energy Transfer', *Proc. Natl. Acad. Sci. USA*, **89**, 11 764–11 768 (1992).
43. L.B. McGown, R.D. Stevens, M.S. Cooter, 'Fluorescence Lifetime-enhanced Spectral Characterization of Human Serum', *J. Fluorescence*, **1**(4), 235–248 (1991).
44. A. Osuka, F. Kobayashi, S. Nakajima, K. Maruyama, I. Yamazaki, 'Synthesis of 1,2-Phenylene-bridged Diporphyrin Linked with Porphyrin Monomer and Its Intramolecular Singlet–Singlet Excitation-energy Transfer', *Chem. Lett.*, 161–164 (1993).
45. J.C.I. Fister, J.M. Harris, 'Data Analysis Techniques for Monitoring Excited State Kinetics', *Anal. Chim. Acta*, **348**, 311–325 (1997).
46. B. Barbieri, F. DePiccoli, M. van de Ven, E. Gratton, 'What Determines the Uncertainty of Phase and Modulation Measurements in Frequency Domain Fluorometry?', *SPIE: Time-resolved Laser Spectroscopy in Biochemistry II*, **1204**, 158–160 (1990).
47. L.E. Scales, *Introduction to Non-linear Optimization*, Springer Verlag, New York, 1985.
48. D.A. Skoog, D.M. West, F.J. Holler, *Analytical Chemistry. An Introduction*, 6th edition, Saunders College Publishing, Philadelphia, 1994.
49. M. Blanco, J. Coello, F. Gonzalez, H. Iturriaga, S. Maspoch, A.R. Puigdomenech, 'Application of Partial Least-squares Regression to the Resolution of Highly Correlated Spectra. Simultaneous Spectrofluorimetric Determination of Al^{3+}, Ga^{3+} and In^{3+}', *Talanta*, **43**(9), 1489–1496 (1996).
50. A. Phatak, S. de Jong, 'The Geometry of Partial Least Squares', *J. Chemometrics*, **11**(4), 339–366 (1997).
51. S. Rannar, F. Lundgren, P. Geladi, S. Wold, 'A PLS Kernel Algorithm for Data Set with Many Variables and Fewer Objects. Part 1: Theory and Algorithm', *J. Chemometrics*, **8**(2), 111–125 (1994).
52. P. Sithisarankul, P. Vineis, D. Kang, N. Rothman, N. Caporaso, P. Strickland, 'Association of 1-Hydroxypyrene-glucuronide in Human Urine with Cigarette Smoking and Broiled or Roasted Meat Consumption', *Biomarkers*, **2**(4), 217–221 (1997).
53. J. Dahlen, M. Backstrom, J. Ephraim, H. Boren, B. Allard, 'Determination of the Molecular Weight of Fulvic Acids by UV/VIS Spectroscopy', *Chemosphere*, **38**(4), 783–794 (1999).
54. G.E. Arteaga, Y. Horimoto, E. Lichan, S. Nakai, 'Partial Least-squares Regression of 4th-derivative Ultraviolet Absorbency Spectra Predicts Composition of Protein Mixtures – Application to Bovine Caseins', *J. Agr. Food Chem.*, **42**(9), 1938–1942 (1994).
55. L. Norgaard, 'Spectral Resolution and Prediction of Slit Widths in Fluorescence Spectroscopy by Two- and Three-way Methods', *J. Chemometrics*, **10**(5–6), 615–630 (1996).
56. D. Banerjee, A.K. Laha, S. Bagchi, 'Studies of Solvation in Homogeneous Media by a Spectroscopic Method – A Ketocyanine Dye in Neat and Binary Solvents', *J. Photochem. Photobiol. A*, **85**(1–2), 153–159 (1995).
57. W.P. Carey, L.E. Wangen, 'Determining Chemical Characteristics of Plutonium Solutions Using Visible Spectrometry and Multivariate Chemometric Methods', *Chemom. Intell. Lab. Sys.*, **10**(1–2), 245–257 (1991).
58. K.N. Andrew, P.J. Worsfold, 'Comparison of Multivariate Calibration Techniques for the Quantification of Model Process Streams Using Diode-array Spectrophotometry', *Analyst*, **119**(7), 1541–1546 (1994).
59. S.R. Erskine, B.M. Quencer, K.R. Beebe, 'Rapid Optical Purity Determination Using Chiral Spectroscopy, Achiral Spectroscopy and Multivariate Analysis', *Appl. Spectrosc.*, **49**(11), 1682–1691 (1995).
60. J. Amador-Hernandez, A. Cladera, J.M. Estela, P.L. Lopez-de-Alba, V. Cerda, 'Resolution of a Multicomponent Polycyclic Aromatic Hydrocarbon System in Micellar Media by Linear Variable Angle Fluorescence Applying Distinct Chemometric Techniques', *Analyst*, **123**(11), 2235–2241 (1998).
61. J. Guiteras, J.L. Beltran, R. Ferrer, 'Quantitative Multicomponent Analysis of Polycyclic Aromatic Hydrocarbons in Water Samples', *Anal. Chim. Acta*, **361**(3), 233–240 (1998).
62. K.Y. Chang, C.W. Brown, R.K. Force, 'Laser-induced Fluorescence Multicomponent Analysis in Full Spectra and Their Fourier Transforms by Principal Component Regression', *Appl. Spectrosc.*, **47**(11), 1801–1807 (1993).
63. A.M.G. Rodriguez, A.G. de Torres, J.M.C. Pavon, C.B. Ojeda, 'Simultaneous Determination of Iron, Cobalt, Nickel and Copper by UV/Visible Spectrophotometry with Multivariate Calibration', *Talanta*, **47**(2), 463–470 (1998).
64. A.M. de La Pena, I. Duran-Meras, M.D. Moreno, F. Salinas, M.M. Galera, 'Resolution of Ternary Mixtures of Salicylic, Salicyluric and Gentisic Acids by Partial Least Squares and Principal Component Regression – Optimization of the Scanning Path in the Excitation–Emission Matrices', *Fresenius J. Anal. Chem.*, **351**(6), 571–576 (1995).
65. A.M. de La Pena, M.D. Moreno, I. Duran-Meras, F. Salinas, 'Synchronous Fluorimetric Determination of Salicylic Acid and Diflunisal in Human Serum Using Partial Least-squares Calibration', *Talanta*, **43**(8), 1349–1356 (1996).
66. I.D. Meras, A.M. de La Pena, M.I.R. Caceres, F.S. Lopez, 'Simultaneous Fluorometric Determination of Nalidixic Acid and 7-Hydroxymethylnalidixic Acid

by Partial Least Squares Calibration', *Talanta*, **45**(5), 899–907 (1998).
67. Z. Bouhsain, S. Garrigues, M. delaGuardia, 'PLS–UV Spectrophotometric Method for the Simultaneous Determination of Paracetamol, Acetylsalicylic Acid and Caffeine in Pharmaceutical Formulations', *Fresenius. J. Anal. Chem.*, **357**(7), 973–976 (1997).
68. D.W. Marquardt, *J. Soc. Ind. Appl. Math.*, **11**, 431 (1963).
69. J.A. Nelder, R. Meade, 'A Simplex Method for Function Minimization', *Comput. J.*, **7**, 308–313 (1965).
70. J.M. Beechem, E. Gratton, M. Ameloot, J.R. Knutson, L. Brand, 'The Global Analysis of Fluorescence Intensity and Anisotropy Decay Data: Second-generation Theory and Programs', in *Topics in Fluorescence Spectroscopy*, ed. J.R. Lakowicz, Plenum, New York, 241–305, Vol. 2, 1991.
71. J. Eisenfeld, C.C. Ford, 'A Systems-theory Approach to the Analysis of Multiexponential Fluorescence Decay', *Biophys. J.*, **26**, 73–84 (1979).
72. I. Isenberg, R.D. Dyson, 'The Analysis of Fluorescence Decay by a Method of Moments', *Biophys. J.*, **9**, 1337–1350 (1969).
73. Z. Bajzer, F.G. Prendergast, 'A Model for Multiexponential Tryptophan Fluorescence Intensity Decay in Proteins', *Biophys. J.*, **65**, 2313–2323 (1993).
74. J. Skilling, R.K. Bryan, 'Maximum Entropy Image Reconstruction: General Algorithm', *Mon. Not. R. Astron. Soc.*, **211**, 111–124 (1984).
75. J.-C. Brochon, A.K. Livesey, J. Pouget, B. Valeur, 'Data Analysis in Frequency-domain Fluorometry by the Maximum Entropy Method – Recovery of Fluorescence Lifetime Distributions', *Chem. Phys. Lett.*, **174**, 517–522 (1990).
76. J. Reiter, 'An Algorithm for Deconvolution by the Maximum Entropy Method with Astronomical Applications', *J. Comp. Phys.*, **103**, 169–183 (1992).
77. I. Sugar, 'Use of Fourier Transforms in the Analysis of Fluorescence Data. 1. A General Method for Finding Explicit Relationships Between Photophysical Models and Fluorescence Parameters', *J. Phys. Chem.*, **95**, 7508–7515 (1991).
78. S.L. Neal, 'Model-independent Estimation of the Photokinetic Transfer Matrix from Matrix-formatted Frequency Domain Fluorescence', *J. Phys. Chem.*, **101**(37), 6883–6889 (1997).
79. P. Lianos, G. Duportail, 'Reactions in Lipid Vesicles. Pyrene Excimer Formation in Restricted Geometries. Effect of Temperature and Concentration', *Eur. Biophys. J.*, **21**, 29–34 (1992).
80. W.I.F. David, 'Extending the Power of Powder Diffraction for Structure Determination', *Nature*, **346**, 731–734 (1990).
81. D.R. James, W.R. Ware, 'Recovery of Underlying Distributions of Lifetimes from Fluorescence Decay Data', *Chem. Phys. Lett.*, **126**, 7–10 (1986).
82. J.M. Shaver, L.B. McGown, 'Maximum Entropy Method for Frequency Domain Fluorescence Lifetime Analysis. 1. Effects of Frequency Range and Random Noise', *Anal. Chem.*, **68**, 9–17 (1996).
83. J.M. Shaver, L.B. McGown, 'Maximum Entropy Method for Frequency-domain Fluorescence Lifetime Analysis. 2. Timing, Mismatched Intensity, and Reference Lifetime Errors', *Anal. Chem.*, **68**, 611–620 (1996).
84. K.A. Zachariasse, W. Kuhnle, U. Leinhos, P. Reynders, G. Striker, 'Time-resolved Monomer and Excimer Fluorescence of 1,3-Di(1-pyrenyl)propane at Different Temperatures. No Evidence for Distributions from Picosecond Laser Experiments with Nanosecond Time Resolution', *J. Phys. Chem.*, **95**, 5476–5488 (1991).
85. A. Siemiarczuk, W.R. Ware, 'Temperature Dependence of Fluorescence Lifetime Distributions in 1,3-Di(1-pyrenyl)propane with the Maximum Entropy Method', *J. Phys. Chem.*, **93**, 7609–7618 (1989).
86. P.A. Lay, A. Levina, 'Kinetics and Mechanism of Chromium (VI) Reduction to Chromium (III) by L-cysteine in Neutral Aqueous Solutions', *Inorg. Chem.*, **35**, 7709–7717 (1996).
87. M. Vincent, I.M. Li De La Sierra, M.N. Berberan-Santos, A. Diaz, M. Diaz, G. Padron, J. Gallay, 'Time-resolved Fluorescence Study of Human Recombinant Interferon Alpha-2-association State of the Protein, Spatial Proximity of the Two Tryptophan Residues', *Eur. J. Biochem.*, **210**, 953–961 (1992).
88. L. Fetler, P. Tauc, G. Herve, M.M. Ladjimi, J.C. Brochon, 'The Tryptophan Residues of Aspartate Transcarbamylase: Site-directed Mutagenesis and Time-resolved Fluorescence Spectroscopy', *Biochemistry*, **31**, 12 504–12 513 (1992).
89. J. Gallay, M. Vincent, I.M.L. De La Sierra, J. Alvarez, R. Ubieta, J. Madrazo, G. Padron, 'Protein Flexibility and Aggregation State of Human Epidermal Growth Factor. A Time-resolved Fluorescence Study of the Native Protein and Engineered Single-tryptophan Mutants', *Eur. J. Biochem.*, **211**, 213–219 (1993).
90. I.M. Plaza del Pino, A. Parody-Morreale, J.M. Sanchez-Ruiz, 'Maximum Entropy Analysis of Kinetic Processes Involving Chemical and Folding–Unfolding Changes in Proteins', *Anal. Biochem.*, **244**, 239–255 (1997).
91. M.M.H. Khalil, N. Boens, 'Global Compartmental Analysis of the Fluorescence Decay Surfaces of the Excimer of 1-Cyanopyrene and the Exciplex of 1-Cyanopyrene with 1,2-Dimethylindole in Toluene', *J. Phys. Chem.*, **97**, 3111–3122 (1993).
92. C. Bernsdorff, R. Winter, T.L. Hazlett, E. Gratton, 'Influence of Cholesterol and β-Sitosterol on the Dynamic Behavior of DPPC as Detected by TMA-DPH and PyrPC Fluorescence', *Ber. Bunsenges. Phys. Chem.*, **99**, 1479–1488 (1995).
93. L.-C. Li, L.B. McGown, 'On-the-fly Frequency-domain Fluorescence Lifetime Detection in Capillary Electrophoresis', *Anal. Chem.*, **68**, 2737–2743 (1996).

94. A. Siemiarczuk, W.R. Ware, Y.S. Liu, 'A Novel Method for Determining Size Distributions in Polydisperse Micelle Systems Based on the Recovery of Fluorescence Lifetime Distributions', *J. Phys. Chem.*, **97**, 8082–8091 (1993).
95. M. Seras, J. Gallay, M. Vincent, M. Ollivon, S. Lesieur, 'Micelle-vesicle Transition of Nonionic Surfactant–Cholesterol Assemblies Induced by Octyl Glucoside: A Time-resolved Fluorescence Study of Dehydroergosterol', *J. Colloid Interface Sci.*, **167**, 159–171 (1994).
96. Y.S. Liu, W.R. Ware, 'Photophysics of Polycyclic Aromatic Hydrocarbons Adsorbed on Silica Gel Surfaces. 1. Fluorescence Lifetime Distribution Analysis: An Ill-conditioned Problem', *J. Phys. Chem.*, **97**(22), 5980–5986 (1993).
97. M. Wilhelm, C.L. Zhao, Y. Wang, R. Xu, M.A. Winnik, J.-L. Mura, G. Reiss, M.D. Croucher, 'Poly(styrene-ethylene oxide) Block Copolymer Micelle Formation in Water: A Fluorescence Probe Study', *Macromolecules*, **24**, 1033–1040 (1991).
98. N. Wittouck, R.M. Negri, M. Amelott, F.C. De Schryver, 'AOT Reversed Micelles Investigated by Fluorescence Anisotropy of Cresyl Violet', *J. Am. Chem. Soc.*, **116**, 10 601–10 611 (1994).
99. W.H. Lawton, E.A. Sylvestre, 'Self Modeling Curve Resolution', *Technometrics*, **13**(3), 617–633 (1971).
100. S.L. Neal, E.R. Davidson, I.M. Warner, 'Resolution of Severely Overlapped Spectra from Matrix-formatted Spectral Data Using Constrained Nonlinear Optimization', *Anal. Chem.*, **62**, 658–664 (1990).
101. S. Wold, 'Cross-validatory Estimation of the Number of Components in Factor and Principal Components Models', *Technometrics*, **20**(4), 397–405 (1978).
102. T.M. Rossi, I.M. Warner, 'Rank Estimation of Excitation–Emission Matrices Using Frequency Analysis of Eigenvectors', *Anal. Chem.*, **58**, 810–815 (1986).
103. H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuehler, 'Quantification of a Known Component in an Unknown Mixture', *Anal. Chim. Acta*, **193**, 287–293 (1982).
104. R. Tauler, G. Durand, D. Barcelo, 'Deconvolution and Quantitation of Unresolved Mixtures in Liquid-chromatographic Diode-array Detection Using Evolving Factor Analysis', *Chromatographia*, **33**(5–6), 244–254 (1992).
105. A. Lorber, 'Generalized Rank Annihilation Factor Analysis', *Anal. Chem.*, **58**, 496–499 (1986).
106. B.E. Wilson, E. Sanchez, B.R. Kowalski, 'An Improved Algorithm for the Generalized Rank Annihilation Method', *J. Chemometrics*, **3**, 493–498 (1989).
107. E. Sanchez, B.R. Kowalski, 'Tensorial Resolution', *J. Chemometrics*, **4**(1), 29–46 (1990).
108. J. Saltiel, D.F. Sears, Jr, J.-O. Choi, Y.-P. Sun, D.W. Eaker, 'Fluorescence, Fluorescence–Excitation, and Ultraviolet Absorption Spectra of Trans-1-(2-naphthyl)-2-phenylethene Conformers', *J. Phys. Chem.*, **98**(1), 35–46 (1994).
109. H.L. Wu, M. Shibukawa, K. Oguma, 'An Alternating Trilinear Decomposition with Application to Calibration of HPLC-ADA for Simultaneous Determination of Overlapped Chlorinated Aromatic Hydrocarbons', *J. Chemometrics*, **12**(1), 1–26 (1998).
110. L. McGown, D.W. Millican, 'Fluorescence Lifetime Resolution of Spectra in the Frequency Domain Using Multiway Analysis', *Anal. Chem.*, **62**, 2242–2247 (1990).
111. G.R. Phillips, S. Georghiou, 'Global Analysis of Steady-state Polarized Fluorescence Spectra Using Trilinear Curve Resolution', *Biophys. J.*, **65**, 918–926 (1993).
112. S.R. Shield, J.M. Harris, 'Time-resolved Monitoring of Aromatic Radicals in Photoinitiated Processes', *Anal. Chem.*, **70**(13), 2576–2583 (1998).
113. T. Yamazaki, Y. Nishimura, I. Yamazaki, M. Hirano, K. Matsuura, K. Shimada, M. Mimuro, 'Energy Migration in Allophycocyanin-B-trimer with a Linker Polypeptide – Analysis by Principal Multicomponent Spectral Estimation Method', *FEBS Lett.*, **353**(1), 43–47 (1994).
114. R.W. Yip, Y.X. Wen, A.G. Szabo, 'Decay Associated Fluorescence Spectra of Coumarin-1 and Coumarin-102. Evidence for a 2-state Solvation Kinetics in Organic Solvents', *J. Phys. Chem.*, **97**(40), 10 458–10 462 (1993).
115. A.R. Holzwarth, M.G. Muller, G. Gatzen, M. Hucke, K. Griebnow, 'Ultrafast Spectroscopy of the Primary Electron and Energy-transfer Processes in the Reaction-center of Photosystem II', *J. Luminescence*, **60–61**, 497–502 (1994).
116. R. Gargallo, F.C. Sanchez, A. Izquierdo-Ridorsa, D.L. Massart, 'Application of Eigenstructure Tracking Analysis and SIMPLISMA to the Study of the Protonation Equilibria of cCMP and Several Polynucleotides', *Anal. Chem.*, **68**(13), 2241–2247 (1996).
117. K.K. Karukstis, D.A. Krekel, D.A. Weinberger, R.A. Bittker, N.R. Naito, S.H. Bloch, 'Resolution of the Excited States of the Fluorescence Probe TNS Using a Trilinear Analysis Technique', *J. Phys. Chem.*, **99**, 449–453 (1995).
118. R. Gargallo, R. Tauler, A. Izquierdo-Ridorsa, 'Acid–Base and Copper(II) Complexation Equilibria of Poly(inosinic)–Poly(cytidylic) Acid', *Biopolymers*, **42**(3), 271–283 (1997).
119. J.C.G.E. da Silva, S.A.G. Novais, 'Trilinear PARAFAC Decomposition of Synchronous Fluorescence Spectra of Mixtures of the Major Metabolites of Acetylsalicylic Acid', *Analyst*, **123**(10), 2067–2070 (1998).
120. K.S. Booksh, A.R. Muroski, M.L. Myrick, 'Single-measurement Excitation/Emission Matrix Spectrofluorometer for Determination of Hydrocarbons in Ocean Water. 2. Calibration and Quantitation of Naphthalene and Styrene', *Anal. Chem.*, **68**(20), 3539–3544 (1996).
121. P.G. Coble, 'Characterization of Marine and Terrestrial DOM in Seawater Using Excitation Emission Matrix Spectroscopy', *Mar. Chem.*, **51**(4), 325–346 (1996).
122. Z.G. Cerovic, F. Morales, I. Moya, 'Time-resolved Spectral Studies of Blue-green Fluorescence of Leaves,

Mesophyll and Chloroplasts of Sugar-beet (*β-Vulgaris L.*)', *Biochim. Biophys. Acta*, **1188**(1–2), 58–68 (1994).
123. B.J.H. Matthews, A.C. Jones, N.K. Theodorou, A.W. Tudhope, 'Excitation–Emission-matrix Fluorescence Spectroscopy Applied to Humic Acid Bands in Coral Reefs', *Mar. Chem.*, **1996**(3–4), 317–332 (1996).
124. E. Perochon, A. Lopez, J.F. Tocanne, 'Polarity of Lipid Bilayers. A Fluorescence Investigation', *Biochemistry*, **31**, 7672–7682 (1992).
125. E. Polverini, R. Casadio, P. Neyroz, L. Masotti, 'Conformational Changes of Neuromedin B and Delta Sleep-inducing Peptide Induced by Their Interaction with Lipid Membranes as Revealed by Spectroscopic Techniques and Molecular Dynamics Simulation', *Arch. Biochem. Biophys.*, **349**(2), 225–235 (1998).
126. A. Izquierdo-Ridorsa, E. Casassas, R. Gargallo, I. Marques, R. Tauler, 'A Comparative Study of Polyelectrolyte Effects and Conformational Changes in Several Purine and Pyrimidine Homopolyribonucleotides', *React. Funct. Polym.*, **28**(2), 127–137 (1996).
127. S.R. Searle, *Matrix Algebra Useful for Statistics*, Wiley & Sons, New York, 1982.
128. E.R. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, Wiley & Sons, New York, 1991.
129. R.I. Shrager, R.W. Hendler, 'Titration of Individual Components in a Mixture with Resolution of Difference Spectra, pKs, and Redox Transitions', *Anal. Chem.*, **54**, 1147–1152 (1982).
130. G.H. Golub, C.H. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1983.

Gas Chromatography: Introduction

G.A. Eiceman
New Mexico State University, Las Cruces, USA

1 Background 1
2 Historical Review of Gas Chromatography Developments 1
3 Principles of Gas Chromatography 2
4 Overview of Contributions 3
Acknowledgment 4
Abbreviations and Acronyms 4
References 4

1 BACKGROUND

Gas chromatography (GC; sometimes called gas–liquid chromatography or GLC) has developed from a single successful application for separating volatile carboxylic acids in the early 1950s[1] into a universally accepted chemical measurement tool spanning the disciplines of chemistry, biochemistry, forensics, toxicology, environmental sciences, and others. The development of GC arose in the context of surging interests in electronics and analytical instrumentation in the post-World War II era. A growing reliance then and now upon physical methods for chemical analyses, such as mass spectrometry (MS) or infrared spectrometry, meant that instrumental characterizations of complex mixtures would be difficult to interpret without prefractionating a sample into individual constituents. GC met this requirement for volatile and semivolatile organic compounds. These compounds constitute only a fraction of all organic substances; however, their relative importance in foodstuffs, cosmetics, and medicines, or in some instances as persistent and toxic pollutants in the environment, guaranteed a role for GC in modern analytical methods.

The speed and effectiveness of GC for separating petroleum materials and natural products such as fragrances were recognized early on and led to a widespread demand for commercial instrumentation. Fundamental developments in descriptions of the chromatographic processes and improved methods of operation provided a foundation for the extension of GC into further applications. This resulted in vigorous activity surrounding GC and the number of abstracts exceeded 1800 per annum after only a decade of development.[2] This pace of publication is comparable to that of today, where about 2000 citations per annum can be seen in results from abstracting services.[3] Careful examination of the literature reveals that much of the essential direction and technology had been described by the early 1960s,[4] although refinements and improvements in the technology have continued. During the 1990s, such advances in the principles of the technique have slowed, although new applications or adaptations of GC abound annually.

2 HISTORICAL REVIEW OF GAS CHROMATOGRAPHY DEVELOPMENTS

In the 1950s the use of stationary liquid phases on solid supports for separating substances within a chemical class rapidly became a generally applicable tool for separating complex mixtures of volatile organic compounds, in contrast to the previous use of an uncoated adsorbent.[5] This was particularly effective in the resolution of constituents in natural fragrances and petroleum samples, and the growth of demand for these separations fostered a nascent industry for the commercialization of GC. Consequently, the availability and convenience of GC instruments for scientists promoted worldwide advances in GC practices and in the description of fundamental principles. Early and nearly complete descriptions of the separation processes were made by the late 1950s, using mathematicales models adapted from existing separation methods. Although packed columns were nearly standard items, the advantages of capillary columns were understood and had been tested by 1958.[6]

During the 1960s commercial instrumentation was developed further, assisted by the availability of programmable ovens, an expanded selection of detectors, and other refinements. The four main detectors in use today (flame ionization, electron capture, thermal conductivity, and MS) were described in this period and most detectors were at a comparatively advanced, stable stage of design by the late 1960s. Discussion groups, abstracting services, and journals such as the *Journal of Gas Chromatography* (later, *Journal of Chromatographic Science*) and the *Journal of Chromatography* aided in the dissemination of discoveries. Other resources became available in the Chromatographic Science Series[7] and in compilations from the International Symposium,[8] which were drawn from the *Journal of Chromatography*. Column technology, which is central to a separation, was advanced with the development of numerous liquid phases and protocols to make possible interlaboratory comparisons of retention.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The fundamentals for separations with packed columns became organized and concise[9] at this time. In short, the essential methods for operating GC instruments, optimizing columns and controlling experimental parameters were developed during the 1960s, even though the technology was rudimentary. The refined technologies of the 1990s, when compared to developments in the 1960s, will be seen as improvements to previously described concepts and principles. Theoretical descriptions were provided with remarkable, though not complete, detail in this decade.

In the 1970s, three main themes arose in GC: the utility for GC methods in the burgeoning environmental and biomedical fields; the growing availability of comparatively inexpensive mass spectrometers as a detector in GC; and the commercial availability of capillary columns. In combination, these three influences radically transformed GC as a measurement science and pushed the method to the state of maturity that is seen today. Capillary columns were known to offer improved resolution over packed columns,[10] and capillary columns gradually supplanted packed columns throughout the 1980s. As a consequence, an entire line of investigation, namely, molecular interactions between substances and various stationary phases, and the efforts to create specialized stationary phases, was eliminated. In place of a vast number of stationary phases with specialized tools to evaluate and select an appropriate phase, only a small number of column types with superb resolution now meet most, if not all, analytical needs. Advances in optimization of chromatographic parameters[11] in terms of temperatures and flows made high resolution possible and sensible. Moreover, the optimization of column construction with respect to mixed stationary phases (packed columns) or lengths of columns in series (capillary columns) was made possible by the discovery of a linear relationship between partition coefficients and fractional composition.[12]

Throughout the 1990s, GC has experienced a period of enormous activity, although fundamental advances have remained relatively stagnant. Any basic advances, when made, occurred in incremental steps. In contrast, data acquisition became facile with inexpensive desktop computers, and comparatively inexpensive GC/MS became commonplace. Whereas in the previous generations, only large institutions could afford GC/MS instruments, the mass-selective detectors and ion traps of the 1990s placed GC/MS within the practical reach of even small laboratories.

3 PRINCIPLES OF GAS CHROMATOGRAPHY

The short summary given below can be supplemented by several extensive descriptions of chromatography, with special attention to GC.[13,14] In a GC experiment, a vapor sample is moved with a flowing gas (the mobile phase, generally either nitrogen or helium) through a glass or metal column containing a phase that is immobilized (the stationary phase). The stationary phase is typically a low-vapor-pressure liquid polymer, and is either coated or chemically bonded to a stationary support (either an inert solid or simply the inner wall surface of an open tube, i.e. a capillary column). As the mobile or gas phase is forced, under pressure, through the column, the sample components are also carried toward the detector at a speed dependent upon the chemical structure of the sample components, the characteristics of the stationary phase, the column temperature, and specifics of the column such as gas flow rate and the amount of stationary phase. Differences in the time of passage through the column (the elution time) are described by the extent to which a substance is dissolved into the stationary phase, i.e. the partition coefficient K which is equal to $[\text{analyte}]_S/[\text{analyte}]_M$, or the ratio of concentrations in the stationary and mobile phases. Thus, the separation process is founded in differences in the partition or solubility of various analytes in the stationary phase. It is the time spent in the stationary phase that prescribes the retention volume; all compounds spend the same amount of time in the mobile phase.

As the molecules are carried through the column by the carrier gas, an increased length of column should permit increased separation between different molecules with different solubilities. Unfortunately, other factors such as diffusion and resistance to mass transport cause broadening of the concentration profiles which lead to losses in resolution. Of all the terms that may be altered, resistance to mass transfer has offered the best opportunity for improvements and improvements have been attained with thin, evenly applied stationary phases; this is best exemplified with capillary columns. In general, control over diffusion is unwieldy in GC because the principal control is found in the selection of the carrier gas, where only two practical selections exist, nitrogen or helium.

Regardless of column design or stationary phase or sample, in GC the critical event is the partition of a substance between the gas phase and the stationary or liquid phase. The partition coefficient K is proportional to $(P^o)^{-1}$, where P^o is the vapor pressure of the pure substance. As vapor pressure is proportional to temperature, GC may be considered approximately as a type of distillation method and this general pattern can be seen in the separation of a homologous series of any chemical such as the *n*-alkanes. However, the subtle advantage of GC resides in the molecular interactions between the stationary phase and the sample, in which molecular interactions exert distinct influences on the solubility or partition

coefficient. This is expressed in the activity coefficient γ which contains both a thermal[15] (molecular interaction) and an athermal (size)[16] term. Thus, $K \propto (\gamma P^{o})^{-1}$ and resolution is controlled by more than vapor pressure alone; consequently, substances and even isomers within a chemical family can be separated by GC. In this respect, gas–liquid chromatography is complementary to gas–solid chromatography, in which separations occur through surface or adsorption interactions, which afford class separations through the large differences between functional groups or moieties.

A major theme in the development of the principles for GC separation was the creation of liquid phases that offer the best possible activity coefficients for a certain solute or sample. This was deemed necessary owing to the limited resolution of packed columns and led to the creation of a large number of stationary phases with tailored or specific intermolecular interactions. By the late 1970s, over 400 stationary phases were commercially available and confronted practitioners of GC with a problematic choice of the best stationary phase. As noted in the section below on **Column Technology in Gas Chromatography**, capillary columns afford resolution and separating capabilities that were impossible in previous generations with packed columns. Consequently, some inefficiency in separation could be tolerated if only an approximately good liquid phase was used, and so the number of stationary phases fell to under a dozen during the 1990s.

The discussion above suggests that GC is a fairly mature analytical method that has been well described for nearly two decades. Consequently, progress in fundamentals has slowed and the questions have become refined and focused. This pattern may be seen in biennial reviews which are made through the American Chemical Society in the section entitled "*Gas Chromatography*" in a special issue, entitled Fundamental Reviews, of *Analytical Chemistry*.[17–19]

4 OVERVIEW OF CONTRIBUTIONS

The discussions in the following articles amplify the themes given above and will reflect the current development of GC as a measurement technology. The reviews have been made broadly readable for those interested in GC, and are current for those who have an interest in recent advances in GC. The general utility of GC for substances that are vital to human health and environmental control is so widespread and effective that it is inconceivable that a dramatically improved technique will arise and suddenly replace GC methods. Certainly, there are no replacements plainly seen among other analytical methods, and GC should be a valued method so long as chemical determinations rely upon physical methods of characterization.

As noted above, one of the key revolutions in GC during the 1990s was the availability and consistency of bonded-phase capillary columns, and these columns have provided unprecedented consistency and resolution in GC methods. The foundations of these advances are described in **Column Technology in Gas Chromatography**. Another area where advances have occurred in the last decade and where future improvements may be anticipated is data processing. Chemical information can now be gleaned from chromatograms, using advanced methods of pattern recognition or using artificial intelligence to analyze GC findings. The most promising approaches are described along with the essential tools for interpreting chromatograms in **Data Reduction in Gas Chromatography**. As noted in the introduction, physical methods of chemical characterization were enhanced dramatically with a preseparation method such as GC, and these are described in **Hyphenated Gas Chromatography**. The performance and precision of GC measurements has been associated with the quality of instrumentation, and these are presented in the contribution entitled **Instrumentation of Gas Chromatography**. Although the number of liquid phases have been reduced in number, as described above, some knowledge of the chemical and physical properties of liquid phases is advantageous with GC measurements. The contribution entitled **Liquid Phases for Gas Chromatography** should provide the necessary background with an update on current trends. Perhaps one of the most exciting developments in GC is the addition of separating power through the use of two or more columns with different stationary phases during a single analysis. This approach has undergone a rediscovery for analytical measurements as described in **Multidimensional Gas Chromatography**. Although early developments in GC were invigorated through the capabilities to resolve and characterize complex mixtures of volatile organic compounds from natural products, useful information can be taken from a GC measurement only with high reliability for sample preparation. This is not trivial for complex, ultratrace or nonvolatile samples. Also, innovative methods of sample handling have extended the use of GC into previously difficult measurements with a wide range of substances. Consequently, any foundational evaluation of GC must include the methods for handling samples, and these are included in **Sample Preparation for Gas Chromatography**.

These categories include the core areas where GC continues to undergo changes or where recent changes have caused pivotal changes in the method. The sections, it is hoped, will provide a reference both for those who are unfamiliar with GC and for those who wish an evaluation

of the current standards and possible future trends in the method.

ACKNOWLEDGMENT

The proofing assistance of David Young, Jonathan Bergloff and Kara Douglas is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

GC Gas Chromatography
MS Mass Spectrometry

REFERENCES

1. A.T. James, A.J.P. Martin, 'Gas–Liquid Partition Chromatography: The Separation and Micro-estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid', *Biochem. J.*, **50**, 679 (1952).
2. M. Gill, S.T. Preston, Jr, 'Bibliography and Index of the Literature on Gas Chromatography 1964', *J. Gas Chrom.*, **2**(12), 391–428 (1964).
3. See CASELECTS PLUS, Chemical Abstracts Service, Columbus, OH, also see http://caselects.cas.org.
4. H.A. Szymanski (ed.), *Lectures on Gas Chromatography 1962*, Plenum Press, NY, 1963.
5. A good review of early developments in gas–solid chromatography can be found in U. Wintermeyer, 'Historical Review', in *Packings and Stationary Phases in Chromatographic Techniques*, ed. K.K. Unger, Chromatographic Science Series, Marcel Dekker, New York, Vol. 17, 1990.
6. M.J.E. Golay, 'Theory and Practice of Gas Liquid Partition Chromatography with Coated Capillaries' in *Gas Chromatography*, eds. V.J. Coate, H.J. Noebels, I.S. Fagerson, Academic Press, NY, 1–13, 1958.
7. J.A. Perry (ed.), *Introduction to Analytical Gas Chromatography: History, Principles, and Practice*, Marcel Dekker, New York, Vol. 14.
8. A. Zlatkis (ed.), *Advances in Chromatography 1976*, Chromatography Symposium, Houston, TX, 1976.
9. W.R. Supina, *The Packed Column in Gas Chromatography*, Supelco, Bellefonte, PA, 1974.
10. W. Jennings, *Gas Chromatography with Glass Capillary Columns*, Academic Press, New York, 1978.
11. S. Morgan, S. Deming, 'Optimization Strategies for the Development of Gas–Liquid Chromatography Methods', *J. Chromatogr.*, **112**, 267 (1975).
12. R. Laub, R. Purnel, 'Optimization of Gas Chromatographic Analyses of Complex Mixtures of Unknown Composition', *Anal. Chem.*, **48**, 1720 (1976).
13. B.L. Karger, L.R. Snyder, C. Horvath, *An Introduction to Separation Science*, Wiley, New York, 212–246, 1973.
14. J.C. Giddings, *Unified Separation Science*, Wiley, New York, 1991.
15. G.R. Luckhurst, D.E. Martire, 'Perturbation Approach to the Free Energy of Solution', *Trans. Faraday Soc.*, **65**, 1248 (1969). Also, see a simplified model in Karger et al., Ibid., 218.
16. M.L. Huggins, 'Thermodynamic Properties of Solutions of Long-chain Compounds', *Ann. N.Y. Acad. Sci.*, **43**, 1 (1942).
17. G.A. Eiceman, H.H. Hill, Jr, J. Gardea-Torresday, 'Gas Chromatography', *Analytical Chemistry Fundamental Reviews*, **70**, 321R–339R (1998).
18. G.A. Eiceman, H.H. Hill, Jr, B. Davani, J. Gardea-Torresday, 'Gas Chromatography', *Analytical Chemistry Fundamental Reviews*, **68**, 291R–308R (1996).
19. G.A. Eiceman, H.H. Hill, Jr, B. Davani, 'Gas Chromatography', *Analytical Chemistry Fundamental Reviews*, **66**, 621R–633R (1994).

Column Technology in Gas Chromatography

Paul Larson
Agilent Technologies, Wilmington, USA

1 Introduction 1
2 Theory and Operating Principles 1
3 Column Construction 3
3.1 Capillary Columns 3
3.2 Packed Columns 7
4 Present and Future Directions in Column Technology 8
List of Symbols 8
Abbreviations and Acronyms 9
Related Articles 9
References 9

Column technology has fostered the development of gas chromatography (GC). As the manufacture of columns, especially capillary columns, has become more reproducible, the instruments have also improved. Chromatographic theory helps to guide the development of both the column and instrument in GC. Aspects of the theory directly related to the construction of columns are summarized. While packed columns can still be made in the local laboratory, the construction of capillary columns has moved to commercial vendors with the introduction of thin-walled fused-silica capillary columns. The technology required for the construction of columns is described.

1 INTRODUCTION

Over the last 40 years, GC has evolved into a widely used technique. Martin first proposed GC in 1951 and used packed columns in his investigations. Almost simultaneously with the commercialization of GC in 1957, Golay purposed capillary columns as a means of separation.[1] By the time of the 1961 conference on GC, Desty reported significant progress in applying capillary columns to petroleum-based samples.[2,3] The early investigators were practical and pragmatic, using materials readily available, such as metal tubing for capillary columns and industrial materials for supports.

The invention of fused silica in 1979 was the result of a systematic study of column materials for capillary columns.[4] At about the same time, a group in Stanford's Electrical Engineering Department reported on the design of a silicon micromachined gas chromatograph, where the gas sampling valve, column, and thermistor were fabricated on a silicon wafer.[5]

The invention of fused silica was significant in several ways. It provided a much improved inertness because of its low metal content, and it influenced stationary phase development because of the smoothness and wetting characteristics of the surface. The roughness of the packings in the packed columns and the metal capillary columns had allowed different phases to be used in column preparation and allowed the individual analyst to prepare columns for his or her own use. Investigators had to develop a better understanding of the interaction between the surface and the polymer used as the stationary phase in order to coat a wide range of stationary phases on to the fused silica. Fused silica also influenced instrument development because it became possible to manufacture very reproducible and stable columns. This article will focus on the manufacture of fused-silica columns and compare the design and manufacturing process with that of packed columns where appropriate.

2 THEORY AND OPERATING PRINCIPLES

The purpose of this article is to discuss GC column technology, not to go into great detail on GC theory. However, theory can be used as a guide for column selection and design, especially for capillary columns. Most commercially available capillary columns are described in terms of coating efficiency, retention index and capacity factor, k', in addition to their physical dimensions [length, internal diameter (ID) and film thickness, d_f]. Packed columns are described in terms of the support material, its size, and the amount of stationary phase, plus the column dimensions.

The purpose of the GC column is to provide the necessary separation for the compounds of interest. The capacity factor is a relative measure of how much a solute is retained and is defined in Equation (1) as

$$k' = t_r - t_m \quad (1)$$

where k' is the capacity factor, t_r is the retention time of the solute, and t_m is the retention time of an unretained peak (the minimum time required for a peak to pass through the column). The capacity factor is related to the partition coefficient (at a particular temperature) by

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Equation (2):

$$K_D = \beta k' \tag{2}$$

where K_D is the partition coefficient of the solute at a given temperature and β is the phase ratio of the column, i.e. the ratio of the gas-phase volume to the liquid-phase volume.

Another parameter can be related to the retention time of the solute. The Kováts retention index gives a retention scale relative to a homologous series of compounds, usually the *n*-alkanes. The retention index for a particular isothermal temperature is given by Equation (3):

$$I = 100n\left\{\frac{\log(t_{rx} - t_m) - \log(t_{rN} - t_m)}{\log[t_{r(N+n)} - t_m] - \log(t_{rN} - t_m)}\right\} + 100N \tag{3}$$

where N is the number of methylene units in the least retained *n*-alkane and $(N + n)$ is the number of methylene units in the more retained *n*-alkane. Retention indices allow intercomparisons of columns in terms of their relative retention of different classes of compounds. For the siloxane phases, an aromatic compound such as acenaphthylene is used to show the relative differences in the stationary phase retention.

While the separation of components is primarily accomplished by the stationary phase selectivity in packed columns, capillary columns offer greater efficiency. This is considered in column selection along with the stationary phase selectivity. The column efficiency is experimentally determined by measuring the retention time, t_r, and the peak of interest's width at half-height, $w_{1/2}$. The efficiency can then be calculated using Equation (4):

$$n = 5.545\left(\frac{t_r}{w_{1/2}}\right)^2 \tag{4}$$

This gives the total plates for the column. For capillary columns, the column efficiency is normalized to column length, L, and is then given by Equation (5):

$$N = \frac{5.545(t_r/w_{1/2})^2}{L} \tag{5}$$

and is specified in plates per meter. An additional measure used with capillary columns is the concept of utilization of theoretical plates (%*UTE*) described by Equation (6):

$$\%UTE = \frac{N}{H_{min}} \times 100 \tag{6}$$

H_{min} is the theoretical minimum plate height and is derived from the Van Deemter equation, which is given in short form in Equation (7) as

$$H = A + \frac{B}{u} + Cu \tag{7}$$

where $A = 2\lambda d_p$, λ is a packing factor and is related to the particle shape and uniformity of packing, d_p is the particle diameter, $A = 0$ for capillary columns, $B = 2\gamma D_g/u$, γ is a correction term for the packing ($\gamma = 1$ for capillary columns), u is the linear velocity, D_g is the gas diffusion coefficient, and $C = 8(\pi k d_f^2 u)/[(1 + k')^2 D_L]$, where k' is the capacity factor, d_f is the film thickness and D_L is the diffusion coefficient of the solute in the stationary phase. As can be seen from Equation (7), the packed column efficiency is dependent upon the packing diameter, the uniformity of the packing material, and the uniformity of the packing process. The equation is simplified for the capillary column case. For capillary columns, the Golay equations are used.

Golay presented and subsequently published his seminal paper, "Theory of Chromatography in Open and Coated Tubular Columns with Round and Rectangular Cross-sections", at the International Gas Chromatography Symposium in 1957.[1] In his paper he derived the relationships for both circular and rectangular cross-section columns. For a circular cross-section column with radius r, the relationship is given as Equations (8) and (9):

$$du = \left(\frac{2D}{u_0} + \frac{1 + 6k + 11k^2}{24(1 + k)^2}\frac{u_0 r_0^2}{D}\right) dx_1 \tag{8}$$

$$\frac{du}{dx_1} = \frac{2D}{u_0} + \frac{1 + 6k + 11k^2}{24(1 + k)^2}\frac{u_0 r_0^2}{D} \tag{9}$$

These are now in the form of the Van Deemter equation, given as Equation (7). Note that there is no A term for the Golay equations. The A term deals with the support material in packed columns. For open-tubular columns, the equation is reduced to Equation (10):

$$H = \frac{B}{u} + Cu \tag{10}$$

The optimum velocity is derived by finding the minimum and is given in Equation (11):

$$\left(\frac{B}{C}\right)^{1/2} = u_{opt} \tag{11}$$

Solving the Golay equations for the minimum gives Equation (12) for the height equivalent theoretical plate H_{min}:

$$H_{min} = \frac{2r_0}{1 + k}\left(\frac{1 + 6k + 11k^2}{24}\right)^{1/2} \tag{12}$$

From these equations, we can see that for the circular cross-section columns the column diameter is a controlling factor for the column efficiency.

Two more concepts are important in the discussion of columns. The first is the concept of resolution, which is a measure of the ability to separate two different solutes, and the second is sample capacity (Equations 13 and 14):

$$R = \left(\frac{N^{1/2}}{4} \quad \frac{(\alpha - 1)}{\alpha} \quad \frac{k}{(1+k)}\right) \tag{13}$$

$$R = \text{column efficiency} \times \text{selectivity} \times \text{column capacity} \tag{14}$$

Column efficiency is primarily a function of the column diameter for capillary columns and particle diameter for packed columns. The selectivity is a function of the stationary phase and the third term is determined by the eluting compound's interaction between the stationary phase and the mobile phase.

Sample capacity also contributes to the resolving power. Sample capacity influences a column's capability for separation because if peaks are overloaded they will broaden and cause a loss of separation. Pretorius and Smuts[6] derived an equation for sample capacity (Equation 15):

$$M = \tfrac{1}{2}\sqrt{H_{\min} L}\pi r^2 (1+k) C_{\mathrm{M}}^{\ddagger} \tag{15}$$

where M is the weight of sample injected and distributed between the mobile and stationary phases and $C_{\mathrm{M}}^{\ddagger}$ is the maximum concentration of the component. The concentration of the solute in the stationary phase is dependent upon the nature of the solute and not the geometry of the column. Therefore, any difference in the weight of the sample in the stationary phase should be due to the stationary phase volume, which is dependent upon the geometry of the column. (This argument can also be extended to the mobile phase, where the volume determines the sample capacity.)

The best way to illustrate the effect of geometry on sample capacity is to consider two columns of equal length, coated with the same stationary phase, and with the same phase ratio and therefore the same capacity factor, k'. One column has an ID of 0.530 mm and the other column has an ID of 0.100 mm. Using the same stationary phase defines the maximum concentration in the stationary phase and is only dependent upon the choice of the stationary phase. Selecting columns with the same phase ratio also defines the value for the capacity factor and should be the same for the two columns. These assumptions allow the ratio of the expressions for the two different columns to be taken and is shown in Equation (16):

$$\frac{M_1}{M_2} = \frac{1/2\sqrt{H_{1\,\min} L}\,\pi r_1^2 (1+k_1) C_{\mathrm{M}}^{\ddagger}}{1/2\sqrt{H_{2\,\min} L}\,\pi r_2^2 (1+k_2) C_{\mathrm{M}}^{\ddagger}} \tag{16}$$

With $k_1 = k_2$ and the concentration maximum the same, Equation (16) reduces to Equation (17):

$$\frac{M_1}{M_2} = \frac{\sqrt{H_{1\,\min}}\, r_1^2}{\sqrt{H_{2\,\min}}\, r_2^2} \tag{17}$$

The ratio can be further reduced by substituting from Equation (12) that $H_{\min}$ is proportional to r. The ratio is reduced to the ratio of the column radius (shown in Equation 18) as an estimate of the effect of the column dimensions on sample capacity:

$$\frac{M_1}{M_2} = \frac{r_1^{5/2}}{r_2^{5/2}} \tag{18}$$

The comparison will become more complex as the lengths, phase ratios, and stationary phases are changed. Table 1 shows the relative sample capacity for different column diameters, but with the limiting assumptions made above. As can be seen from the comparison, as the diameter is decreased the sample capacity is considerably reduced. One development that has circumvented this limitation is to make multiple parallel capillary columns. The sample capacity directly increases as the number of columns increases. The sample capacity decreases at a faster rate than the column efficiency. There are many instances where the sample capacity can be the determining factor in the choice of a column.

The sample capacity of packed columns is higher. A 6 ft × 2 mm 5% methylsilicone Chromosorb® W 100–120-mesh packed column has about five times the sample capacity of the 30 m × 0.530 mm × 2.65 µm methylsilicone capillary column. The 6-ft packed column can only generate 3000–5000 total plates whereas the 30-m column can generate 60 000–75 000 total plates when operating near the optimum velocity. This capillary column provides nearly the same sample capacity but improved separation due to column efficiency and better column inertness.

Table 1 Comparison of relative sample capacity based on column diameter for columns of different column diameters with the column length and phase ratio the same for both columns

Column diameter (µm)	Sample capacity relative to 0.530-mm ID column
750	2.382
530	1.000
450	0.664
320	0.283
250	0.153
200	0.087
180	0.067
100	0.015
50	0.004

3 COLUMN CONSTRUCTION

3.1 Capillary Columns

The manufacture of capillary columns requires multiple steps. These steps will be briefly outlined to give an overview of the process before discussing the individual steps. To prepare a capillary column requires the choice of the column tubing material and the manufacture of the column tubing, a decision on pretreatments and deactivations, the choice of the stationary phase, the choice of the coating technique and coating solvent, if the stationary phase is to be cross-linked then the choice of the cross-linking initiator, if the column is to be rinsed then the rinse solvent, and the conditioning prior to use.[7]

A variety of tubing materials have been used for the construction of capillary columns. Initially, commercially available metal tubing was used. This had some inherent advantages for the early investigators in that the surface roughness allowed different stationary phases to be tried and used. The major limitation was the activity of the tubing. This motivated investigators to develop the means for drawing glass capillary tubing. Much of the work used either soda-lime glass (called this because of the Na_2O in the glass) or a borosilicate type of glass. The presence of the Na_2O lowers the melting temperature of the soda-lime glass, which makes it easy to work with. Similarly, borosilicate glass (which has B_2O_3 as a flux in the glass) has a lower melting point than SiO_2 (see Table 2 for glass composition). Silicon dioxide glass is available as either quartz or fused silica. Quartz is a naturally occurring material obtained through mining and can contain trace metal contaminants, sometimes at significant enough levels to affect the chromatographic performance of the column. Fused silica is a synthetic material prepared by oxidizing tetrachlorosilane to silicon dioxide. This process leads to a very pure material, with very low metal content and a low concentration of chloride. Soda-lime and Pyrex glasses were drawn using a relatively small machine (roughly 20 in by 30 in) that would automatically coil the drawn capillary. This tubing was thick walled and required the ends of the tubing to be straightened. While column end straighteners were made to help with this task, the inexperienced chromatographer still found that this required straightening before installation into the gas chromatograph was difficult to master. This caused reluctance to use capillary columns. Up until the invention of the thin-walled fused-silica capillary columns, the manufacture of capillary columns was still the domain of chromatographers in private laboratories and with limited commercialization. Inexpensive drawing machines and devices for straightening the column ends were available to these laboratories.

The purity of the fused-silica material is important because it allows the tubing to be drawn as thin-walled tubing. The primary benefit of the thin-walled tubing is that it is inherently straight. For smaller diameter columns the outside diameter (OD)/ID ratio is increased to stiffen the material. If these columns were drawn with the same OD/ID ratio as the larger diameter columns, such as the 530-mm columns, the tubing would be limp and harder to handle. Thus it is more difficult to maintain the integrity of the sealing surface by requiring smaller diameter ferrules or excessively extruding graphite ferrule material. It is also harder to make the ferrules with the smaller OD columns. Although it is possible to draw other glasses into thin-walled tubing, it is not practical. The presence of the fluxes (Na_2O, CaO, etc.) in these glasses considerably reduces the inherent strength of the silicon dioxide. Another mechanism, sodium devitrification, also affects the strength of the silicon dioxide. The presence of sodium eventually causes the glass to form microcracks. With the thicker walled tubing this not a problem, but the thin-walled fused-silica tubing requires a protective sheath. The protective sheath is necessary because sodium is picked up very easily during the handling of the column. The protective sheath is normally polyimide with a temperature limit in the 350–375 °C range. The temperature limit will be determined by exactly which polyimide is used (above this temperature the polyimide begins to oxidize). The polyimide used at the introduction of fused-silica columns in 1979 could be completely removed by exposure to 400 °C for 1 week. Aluminum has also been used as a sheath material to extend the time of exposure to high temperatures (400–450 °C). When coating the protective sheath on to the thin-walled fused-silica tubing, care must be taken not to trap particles between the polyimide and the fused-silica tubing. Particles can be a source of abrasion and sodium, both of which can cause microcracks and eventually breakage of the column. Therefore, the drawing process now requires a very clean environment and a much larger area. The drawing machine has now become a clean room with enough room for the drawing tower (this requires a ceiling

Table 2 Composition of different glasses (%)

Glass type	SiO_2	Al_2O_3	Na_2O	CaO	MgO	B_2O_3	BaO
Soda-lime	68	3	15	6	4	2	2
Borosilicate	81	2	4	–	–	13	–
Fused silica	100	<1 ppm total metals					

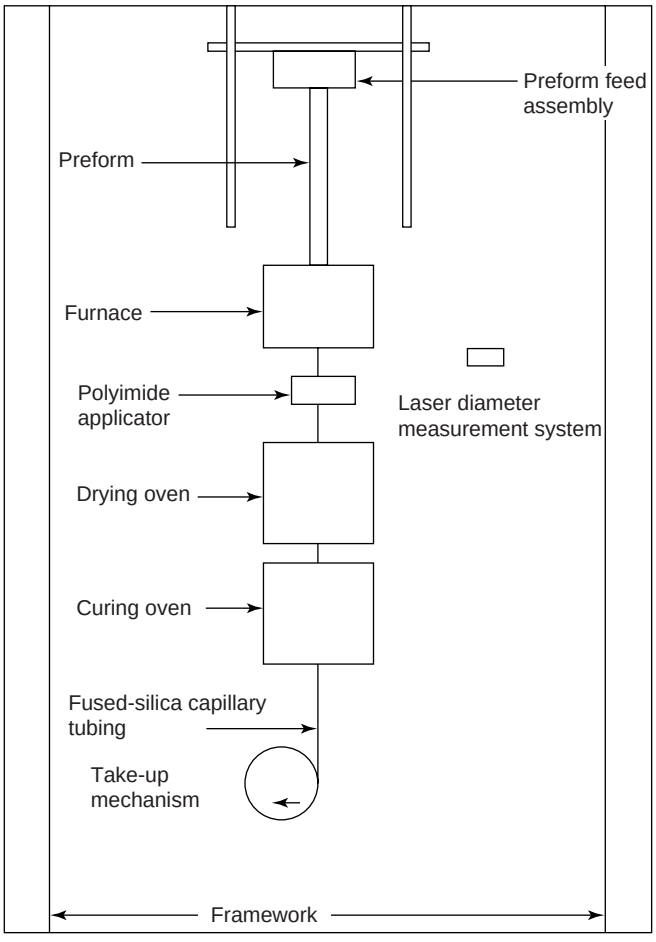

Figure 1 Schematic diagram of a fused-silica capillary column drawing tower.

height of 20–30 ft or higher). The laminar flow for the air greatly reduces the number of particles on the tubing prior to application of the protective sheath. The drawing tower (as shown in Figure 1) consists of the framework for the mounting of the different equipment, the feed mechanism for the preform (typically 1 m or greater in length), the graphite furnace where the preform is drawn down to the required OD and ID, the polyimide applicator, and a length of drying ovens to cure the polyimide before the tubing reaches the take-up reel. If the polyimide is not cured, the drawn tubing will bond together. The cost of preparing the thin-walled fused-silica tubing in private laboratories is essentially prohibitive and fused-silica tubing is therefore commercially prepared almost completely. The preparation of the columns has also shifted to the column vendors.

As the column preparation continues, the next steps will prepare the tubing to be coated with the stationary phase. After the tubing has been drawn, the desired length is cut and, for the fused-silica tubing mounted on a basket, the column can be pretreated. For the softer glasses, several pretreatments were used, including leaching with dry hydrogen chloride gas and growing barium carbonate crystals on the surface. These techniques were carried out to improve the wettability of the surface by increasing the roughness of the surface. The fused-silica material is not as easily leached, for several reasons. The thickness of the residual water layer is less than that of the softer glasses, and for the softer glasses the presence of the leachable metals allow the leaching to go deeper and to occur more quickly. These metals on the atomic scale have large hydrodynamic radii, which helps the leaching of the glass. In fused silica these metals are absent. While the pretreatment methods tend to be gas-phase-based for the softer glasses, pretreatment methods for fused silica tend to be liquid-based. For the fused silica, this first step may be as simple as a rinse with distilled water or other dilute aqueous solutions. This pretreatment can be used to increase the number of surface silanol groups. Silane, siloxane, and silazane regeants can then be reacted with these groups to complete the pretreatment or deactivation.

Deactivation can also be done prior to the coating of the columns. Carbowax 20M™ is a polyethylene glycol that can be used both as a stationary phase and a deactivation agent. When it is used for deactivation, a plug of Carbowax™ solution is passed through the column, the column is sealed and then heated to 280 °C for a period of time (this can be as long as overnight). At the completion of the deactivation time, the column is cooled and unsealed. It is then rinsed with a rinsing solvent to remove the residue and then purged to remove the residual solvent. Other types of deactivation include siloxane-based deactivations. One of the earlier of these deactivations is the D_4 deactivation, where D_4 represents cyclic octamethyltetrasiloxane. The plug of D_4 is passed through the column and the column ends are sealed. The column is then heated to 400 °C. At 400 °C, the cyclic siloxane slowly opens and will react with both the fused-silica surface at the surface silanols and with itself to produce a thin deactivated layer on the interior surface of the column. Other deactivations are based on more reactive reagents, such as the disilazanes, and react under milder conditions. Still other proprietary compounds and techniques are used by the column manufacturers.

The next step is the coating of the column with the stationary phase. The coating step can be carried out on untreated material, but many of the stationary phases require pretreatment and/or deactivation steps for efficient columns to be prepared. To illustrate how deactivation affects this process, the coating of a phenylmethylsiloxane of molecular weight 3000 was carried out on Carbowax™-deactivated, D_4-deactivated and undeactivated fused-silica columns. The only column of these three that was an efficient (i.e. well coated) column

was the column with the Carbowax™ deactivation. The phase will bead up on the other two columns. This low-molecular-weight stationary phase is a liquid. To coat the fused silica properly required a higher molecular weight phenylmethylsiloxane (a gum rather than a liquid) to be developed. While these gum phases were readily available for polydimethylsiloxane and 5% phenylmethylsiloxane copolymer at the introduction of fused-silica columns, the other phases were not. Much of the research into capillary columns during the 1980s was devoted to optimizing stationary phase development. Much progress was made in the development of stationary phases with higher molecular weight, hence with better wetting of the column surface. The better wetting of the surface is necessary for the production of efficient columns. These phases now include phenylmethylsiloxane (also denoted 50% phenylmethylsiloxane) and cyanopropylmethylsiloxane, which extend the polarity of columns available to the chromatographer. Simply increasing the molecular weight does not necessarily produce a viable phase. For example, to coat the phenylmethylsiloxane column efficiently on material that has not been deactivated requires a gum phase of molecular weight ~100 000. Grassie et al. reported that with polymers built with the phenylmethyl monomer, the longer the polymer chain length, the more it will react with itself.[8] This manifests itself with a choppy baseline as the column is temperature programmed to 300 °C. Either the polymer chain must be interrupted or a different starting material for the polymer must be used. If diphenyldimethylsiloxane is used as the chain repeating unit, this problem may be avoided.

The gum phases used with fused-silica columns have affected the production of the columns in a different way. With the higher molecular weight stationary phases, the coating solution becomes very viscous as the concentration of the stationary phase increases. This limits the technique of coating the column to primarily the static coating technique. For lower molecular weight phases, the dynamic coating technique can also be used. The dynamic technique will be discussed first. To carry out dynamic coating requires the preparation of a concentrated stationary phase solution (usually in the 10–20% range). A volume of 2–4 mL of this solution is transferred to a coating station consisting of a fitting that includes a connection to an inert purge gas source and a reservoir for the coating solution. A plug of solution is then introduced into the column and pushed along by the gas. The speed at which this plug moves determines the film thickness. This technique will work better with the wider bore capillary columns, such as 530-μm ID columns, because the pressure drop is not as great and therefore the film thickness will not change as much. With longer, smaller diameter columns such as 200-μm ID columns, the film thickness will vary more because the pressure drop is higher. This causes the velocity to increase rapidly as the plug nears the exit of the column. The column to column repeatability is not as good as with the static coating technique.

The static coating technique is carried out by filling the column with the coating solution. Once the column has been filled, the exit end of the column is sealed. The column is then attached to a vacuum source and the solvent is slowly evaporated, leaving the stationary phase behind. Once the column has been evacuated, further reaction and conditioning of the stationary phase can occur in the column. The concentration of the stationary phase can be calculated using Equation (19):

$$\beta = \frac{\text{volume mobile phase}}{\text{volume stationary phase}} = \frac{V_m}{V_s} = \frac{\text{volume of the volumetric flask}}{\text{density} \times \text{weight of the stationary phase}} \quad (19)$$

The density for the polydimethylsiloxane phases is 1.0 g mL^{-1} and the solution phase ratio is easily calculated. For other stationary phases it is necessary to include the density to calculate the film thickness of the stationary phase properly. Thus, to prepare a coating solution of dimethylpolydimethylsiloxane for a phase ratio of 250 would require dissolving 1.000 g of the phase in a 250-mL volumetric flask. Once the phase has dissolved in the solvent, the solution can be used for coating columns.

As seen from the brief description of the static coating process, the choice of the solvent for the stationary phase is important and helps to determine the length of time required to coat the column and the range of film thickness that can be coated with the solvent–stationary phase combination. Solvents with high volatility (low boiling points) such as pentane (36.05 °C), methylene chloride (39.85 °C), and acetone (56.1 °C) have been used for static coating of columns. The polydimethylsiloxane phases are readily soluble in pentane. Whereas the low-molecular weight polyphenylmethylsiloxanes (liquid) are appreciably soluble in pentane, the higher molecular weight gums of the phenylmethylsiloxanes are not and require a different solvent, such as methylene chloride. Since the column length can be between 10 and 150 m, the filled column is a very effective thermometer which responds to small changes in the temperature of the environment around the column while the column is being evacuated. During the evacuation process, the column must be shielded from the temperature swings in the room and, if coating baths are used, from the temperature swings of the heaters. The high-volatility solvents tend to have the higher coefficients of thermal expansion. As the coating solution expands or contracts it will wash the already deposited phase and change the coverage of the column. To a lesser extent, the concentration of the solution changes and will deposit less at higher temperatures or more at lower

temperatures. Sudden cooling of the solution will cause the coating solution to contract and can leave uncoated areas along the column. The thicker the coating solution, the less prone it is to this behavior. If the coating solution is not degassed it may cause cavitation problems when pumped with a high-pressure liquid chromatograph pump. The trapped gas can also cause problems with the evacuation process by expanding and pushing the coating solution along as it expands. This will happen when the pressure drop is high enough to cause the bubble to expand. When this occurs, the whole or part of the column can be washed out. Again, thicker stationary phase solutions tend to reduce this tendency.

The static coating technique requires a high-pressure pump, such as a liquid chromatograph pump, for the higher viscosity solutions to be filled. The columns must then be sealed without trapping air at the end. A vacuum system must be available to evacuate the columns and a means (such as a buffer volume) must be provided to bring up the vacuum slowly. Care must be taken to hold the temperature as constant as possible. Quick temperature swings around the columns should be avoided.

After the columns have been evacuated, the next step in the process is the cross-linking step.[9] The cross-linking agent must have several characteristics to be a good choice for the column-making process:

1. It must dissolve in the coating solution at the required concentration.
2. It cannot be volatile at the column evacuation temperature.
3. It must react at a temperature less than the maximum recommended operating temperature for the phase and preferentially at the lowest temperature as possible (this allows the stationary phase to remain as viscous as possible).
4. It should not degrade or change the stationary phase characteristics. Usually, this is accomplished by the cross-linking occurring at vinyl end caps, methyl groups, or OH groups remaining in the stationary phase or on the fused silica. Cross-linking via the methyl groups on a dimethylpolysiloxane would proceed as follows:

   (a) Free radical formation:

   $$R{-}R \longrightarrow 2R^{*}$$

   (b) Abstraction of hydrogen:

   $$R^{*} + CH_3(SiO)CH_3 \longrightarrow RH + {}^{*}CH_2(SiO)\,CH_3$$

   (c) CH_3 recombination of radicals to cross-link dimethylpolysiloxane:

   $$2{}^{*}CH_2(SiO)\,CH_3 \longrightarrow CH_3(SiO)CH_2{-}CH_2(SiO)$$

5. The decomposition products should be easily removed with purging and/or solvent rinsing.

The cross-linking is carried out by dissolving an initiator such as dicumyl peroxide **(1)** in the stationary phase solution. Once the solvent has been evaporated, the column contains the stationary phase and the initiator. The column is then slowly purged with an inert gas as the temperature is increased. Dicumyl peroxide decomposes to the peroxide radical for the cross-linking reaction to occur. After the cross-linking step is completed, the column is usually rinsed and then carefully conditioned under flow. The choice of solvent and thermal conditioning profile will be determined by the stationary phase.

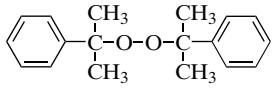

(1)

The final step in the process is to test the column. Several different parameters are checked using chromatographically well behaved compounds such as the *n*-alkanes on the siloxane-based stationary phases:

- the capacity factor, k'
- the retention index, *RI*
- the coating efficiency.

Other compounds are typically chosen to show the inertness level of the column (using acidic and basic compounds) and the resolving power of the column with closely eluting pairs of compounds, such as the xylenes. Application-specific columns are tested with representative compounds for the application.

3.2 Packed Columns

The manufacture of packed columns requires fewer steps than capillary columns. These steps will be briefly outlined to give an overview of the process before discussing the individual steps. To prepare a packed column requires the choice of the column tubing material, the choice of the support material, a decision on pretreatments and deactivations of the support material, the choice of the stationary phase, the choice of the coating solvent, and the packing technique.[10–13]

Several different materials have been used for the tubing material for packed columns, including metal tubing, glass tubing, and polymeric tubing. The choice of tubing will depend upon how the packed column will be used. Glass or glass-lined tubing is used when activity levels need to be reduced. If glass tubing is used, it must be preconfigured to fit the gas chromatograph in which it will be used. The metal and polymeric tubing are also coiled, if necessary. The nature of the packing material may require that the coiling of the metal tubing be done prior to loading of the packing.

The support material is typically a porous material, although glass beads have been used. The materials used for the support include diatomaceous earth, zeolites, porous polymers, and carbon-based material, such as charcoal. For the diatomaceous earth supports, the source of the diatomaceous earth, the treatments, whether acid washed and treated further, and any deactivation steps contribute to the inertness of the column along with the column tubing.

Once the appropriate support material has been selected, the stationary phase is applied to the support by dissolving the phase in an appropriate solvent. Since packed column loadings typically range from 1 to 20% (w/w), the phases tend to have lower molecular weights than comparable phases used with capillary columns. This is because of the limited solubilities of the higher molecular weight gums used for the manufacture of the capillary columns. The stationary phase solution is usually prepared with a higher boiling solvent than would be used with the capillary columns. The stationary phase solution is added to the packing and mechanically mixed, usually by rotating, and the solvent is evaporated.

When the solvent has been removed, the coated support material is then loaded into the tubing using a funnel to direct the material into the tubing and is mechanically agitated, usually by tapping to disperse the material throughout the length of the column. Plugs are placed at the ends of the columns to hold the material in place.

Comparison of the range of stationary phase loading for packed columns with the phase ratios of capillary columns requires the phase ratio of the packed column to be estimated. The phase ratio of a 6 ft $\times$ 2 mm 5% polymethylsiloxane on Chromosorb W®, 110–120 mesh, is estimated to be about 58. The 1–20% range of column loading corresponds to a phase ratio range of 15–300, while the phase ratio range for capillary columns is typically 25–500. To increase the column efficiency, the particle diameter must be reduced, which increases the pressure drop across the column. As with capillary columns, there is a trade-off with increased column efficiency and increased pressure required.

4 PRESENT AND FUTURE DIRECTIONS IN COLUMN TECHNOLOGY

Although fused silica has provided a very stable and inert support, there are applications, such as high-temperature simulated distillation, which will rapidly degrade a methylsiloxane column, both on the inside and outside of the column if the column is polyimide coated. The polyimide has a threshold of $\sim$350 °C for its oxidation. Continuous exposure at 400 °C for a period of about 1 week will strip the column of its protective coating. The methylsiloxane column will also show significant reduction in its ability to partition the solute. These losses have led to other approaches, such as depositing a layer of silica on the interior of the metal capillary tubing and the development of arylene-based stationary phases. The arylene-based stationary phases have arylene units introduced into the polymer backbone.

There has also been a movement towards miniaturization in analytical techniques and instrumentation. Over the past 20 years, researchers have investigated the use of silicon-based materials for capillary columns. This requires the application of microlithographic techniques to column manufacture. One desirable feature of this approach is the fact that the silicon material presents a thin silica layer. This layer is known as the native oxide layer. In the processing of silicon wafers, there are several different methods of growing or depositing an oxide layer on the wafer. The thermal oxidation of the silicon grows the highest density oxide layer. This thermal oxide is much like the drawn fused silica in its density. A second technique for depositing the oxide is chemical vapor deposition (CVD), which can give a more porous silicon dioxide layer. The inherent problem with columns developed using microlithographic techniques is that microlithography is essentially carried out in two-dimensional (planar) space. The channels for the columns are prepared by using controlled etch solutions. The final step in preparing the columns is to seal a second plate over the etched channel and make connections to the plate. Coating these plate columns with stationary phase presents a different set of problems. This approach will require these problems to be solved in order for it to be viable.

Molecular sieves represent another area where some interesting work is being done. UOP mat/sen (El Dorado Hills, CA) has developed several different molecular sieves that are very specific for different gases. While these specialty molecular sieves have applications as traps, there may also be ways to utilize these materials for porous layer open-tubular (PLOT) columns and micropacked columns.

As can be seen, column technology has utilized a wide range of materials and techniques. Further advances will be made through the systematic application of materials

science, polymer chemistry, and a variety of analytical techniques.

LIST OF SYMBOLS

α	selectivity
β	phase ratio of the column, i.e. the ratio of the gas-phase volume to the liquid-phase volume
γ	correction term for the packing
λ	packing factor related to the particle shape and uniformity of packing
$C_M^{\ddagger}$	maximum concentration of the component
d_f	film thickness
d_p	particle diameter
D_g	gas diffusion coefficient
D_L	diffusion coefficient of the solute in the stationary phase
H_{min}	theoretical minimum plate height
k'	capacity factor
K_D	partition coefficient of the solute at a given temperature
L	column length
M	weight of sample injected and distributed between the mobile and stationary phases
n	column efficiency
N	column efficiency normalized to column length, usually plates per meter
R	resolution
t_m	retention time of an unretained peak
t_r	retention time of the solute
u	carrier gas linear velocity
$\%UTE$	utilization of theoretical plates
$w_{1/2}$	width at half-height

ABBREVIATIONS AND ACRONYMS

CVD	Chemical Vapor Deposition
GC	Gas Chromatography
ID	Internal Diameter
OD	Outside Diameter
PLOT	Porous Layer Open-tubular

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Atomic Emission Detection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Polymers and Rubbers **(Volume 9)**
Gas Chromatography in Analysis of Polymers and Rubbers

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography

REFERENCES

1. M.J.E. Golay, 'Theory of Chromatography in Open and Coated Tubular Columns with Round and Rectangular Cross-sections', in *Gas Chromatography*, ed. D.H. Desty, Butterworths, London, 36–55, 1958.
2. D.H. Desty, A. Goldup, 'Coated Capillary Columns – an Investigation of Operating Conditions, in *Gas Chromatography*, ed. D.H. Desty, Butterworths, London, 162–183, 1960.
3. D.H. Desty, 'The History and Potentiality of Capillary Columns in Gas Chromatography', *J. Chromatogr. Sci.*, **25**, 552–563 (1987).
4. R. Dandeneau, E. Zerenner, 'An Investigation of Glasses for Capillary Chromatography ', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**, 351–356 (1979).
5. S.C. Terry, J.B. Angell, ' A Column Gas Chromatography System on a Single Wafer of Silicon', in *Theory, Design, and Biomedical Applications of Solid State Chemical Sensors*, eds. P.W. Cheung, D.G. Fleming, M.R. Neuman, W.H. Ko, CRC Press, West Palm Beach, FL, 207–218, 1978.
6. V. Pretorius, T.W. Smuts, 'Sample Capacity in Open Tubular and Micro-packed Columns for GC', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**, 444–446 (1979).

7. K. Grob, *Making and Manipulating Capillary Columns for Gas Chromatography*, Hüethig, Heidelberg, 1986.
8. N. Grassie, I.G. MacFarlane, K.F. Francey, 'The Thermal Degradation of Polysiloxanes – II. Poly(methylphenylsiloxane), *Eur. Polym. J.*, **15**, 415–422 (1979).
9. W. Noll, *Chemistry and Technology of Silicones*, Academic Press, Orlando, FL, 1968.
10. J.A. Yancey, 'Liquid Phases Used in Packed Gas Chromatographic Columns. Part III. McReynolds' Constants, Preferred Liquid Phases, and General Precautions', *J. Chromatogr. Sci.*, **24**, 117–124 (1986).
11. D.M. Ottenstein, 'The Chromatographic Support in Gas Chromatography', *J. Chromatogr. Sci.*, **11**, 136–144 (1973).
12. D.M. Ottenstein, 'Column Support Materials for Use in Gas Chromatography', *J. Chromatogr. Sci.*, **25**, 536–546 (1987).
13. D.M. Ottenstein, 'The Chromatographic Support', *Adv. Chromatogr.*, **3**, 137–196 (1966).

Data Reduction in Gas Chromatography

G.A. Eiceman
New Mexico State University, Las Cruces, USA
J. Gardea-Torresdey
University of Texas, El Paso, USA

1 Introduction 1
2 Essentials of Chromatographic Data 2
2.1 General Facets of a Chromatogram 2
2.2 Quantitative Aspects of Gas Chromatography Data 2
2.3 Qualitative Aspects of Gas Chromatography Data 3
2.4 The Chromatogram as Diagnostic Tool 5
3 Advanced Data Reduction Methods in Gas Chromatography 5
3.1 Multivariate Data Analysis in Gas Chromatography 5
3.2 Neural Network Analysis in Gas Chromatography 7
4 Future Developments 7
List of Symbols 7
Abbreviations and Acronyms 7
Related Articles 8
References 8

The data obtained from a gas chromatographic measurement provide the user with details on both the quantitative and the qualitative composition of the sample. This information is displayed in the chromatogram from which information also can be gleaned concerning the performance and quality of the measurement. Quantitative determinations require attention to the relative response of the detector. The assignment of identity of individual constituents is possible in a formal sense only with authentic standards and mass spectra though retention times or indices can be indicators of identity. These aspects of data reduction have been understood for several decades and chemical information can be extracted from a chromatogram with ease and convenience.

In recent years, advanced chemometrics methods have been applied to chromatographic data with intrinsic advantages for probing patterns and trends too subtle or too complex for simple inspection of chromatograms. These methods are rooted in tools such as multivariate analysis and neural network analysis. Such methods are being developed in the area of natural products analysis.

1 INTRODUCTION

A chromatogram is the immediate result obtained from any analytical measurement by gas chromatography (GC) and discloses, for most detectors, all the chemical information about a sample that will be provided by the GC experiment. In addition, details about the overall performance of the chromatograph can be found in the chromatogram and diagnostics are best made when a known sample such as a standard mixture is analyzed. A chromatogram is a plot of detector response (usually in current or voltage) versus time (usually in minutes) following the introduction of a sample into the chromatograph. Most analyses by GC conclude with an understanding of the concentrations or amounts of one or more constituents in a sample and some indication of the chemical identity of such constituents. These results are readily available from the chromatogram and this accounts for the wide acceptance of GC as an analytical tool. At an elementary level, the information and the methods to extract this information have been unchanged during the last several decades and can be found in specific monographs on GC[1,2] or in general analytical chemistry texts.[3,4] These will be reviewed and illustrated concisely in the next section.

In addition to the routine use of chromatographic data to answer questions such as 'is there benzene in my sample and if so how much is there?', there may be interest in using the chromatogram for comparing with results from other samples. This might be motivated by interests in the origins of the sample or to learn if subtle changes in sample composition had occurred during storage, treatment, or processing of the sample. Normally, human skills of comparison are eclipsed when samples are highly complex or numerous. For example, observing the differences or similarities between chromatograms containing more than 20 peaks can be daunting especially when there may be only minor differences in concentrations of several components. Small differences in chromatograms from GC analysis of volatile organic compounds in blood or urine or breath might be essential in determining an illness, but easily overlooked without computation aids.

While comparisons of complex chromatograms are impractical through manual methods of data reduction, advanced computational methods can provide the necessary tools for such comparisons. These methods have been refined rapidly during the last few years with the wide

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

availability of affordable, powerful and small computers and the release of specialized, commercial software for advanced data reduction. In the third section below, two of the promising methods for advanced data reduction are described.

2 ESSENTIALS OF CHROMATOGRAPHIC DATA

2.1 General Facets of a Chromatogram

In Figure 1, a gas chromatogram is shown from the analysis of a standard mixture of normal alkanes by gas chromatography/mass spectrometry (GC/MS) and the results are useful for illustrating the information that can be gleaned from an analysis by GC. The important terms for the analyst include the number of constituents (seen ideally as separated or well-resolved peaks) and the relative abundances of each constituent (i.e. the relative heights or areas of each peak, as a rough approximation). Each peak can be defined by a few terms including: the retention time (t_r), peak area (or height), peak width (w), peak shape distortion or asymmetry factor (η), and the retention time for an unretained compound (t_m). These all assist in expressing the ultimate goal of a chromatographic experiment, namely, the satisfactory separation or resolution of constituents in a mixture. Other terms such as column length, carrier gas flow rate, inlet pressure, ambient pressure, and column temperature cannot be obtained from the chromatogram and are acquired from independent measurements. Such terms are especially useful when studying the details of chromatographic behavior but are not essential for analytical purposes. All these terms and their meaning have been standardized over the last 30 years[1–4] and are helpful in determining or documenting the composition of the sample or in exploring separation processes.

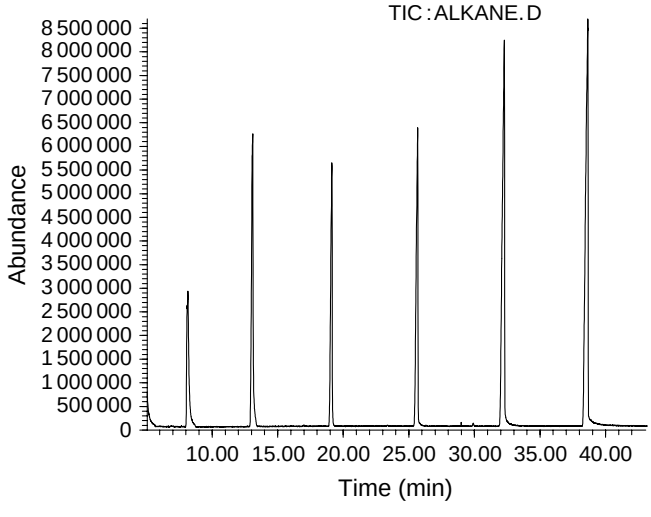

Figure 1 Gas chromatogram of *n*-alkanes taken with temperature programmed analysis. The analysis was made using a capillary column in a gas chromatograph mass spectrometer. Thus, the solvent peak was eluted and vented before data acquisition by the mass spectrometer was started. The chromatogram shows the general appearance of well-resolved peaks in a mixture.

The retention time is the timescale of the chromatographic measurement as shown in the *x*-axis of the chromatogram in Figure 1 and is measured from the point of injection (not seen in the figure). Each peak is assigned a retention time measured at the peak maximum for a substance and this is shown in Table 1. For example, the first peak has a retention time (t_r) of 8.154 min and the second peak shown has a t_r of 13.084 min and these supply information regarding the chemical identity of that peak as described below. In addition, the peak heights or areas contain information about the amounts of substances in the sample and this too is described below. A salient feature of the chromatogram before detailed discussion is the information so often sought: this sample shows the presence of six constituents and the amounts of each, as a first approximation, appear comparable in order of magnitude.

2.2 Quantitative Aspects of Gas Chromatography Data

The size of any peak is a measure of the amount of that chemical in the sample and there is a proportional relationship between peak area (or peak height) and amount of chemical (number of molecules, moles, mass, or concentration). However, all detectors in GC are operated on principles where there is some discrimination, based upon molecular structure, in the sensitivity of response (Δ response/Δ amount). Consequently, any direct comparison of peak heights of various substances without consideration of differences in detector response will be erroneous. Quantitative determinations for GC will be accurate only when the amount or concentration of each analyte is determined using individual calibration curves made for that specific analyte or when relative response factors can be used to normalize a response. Highly selective detectors such as electron capture detector (ECD)

Table 1 Results from analysis of chromatogram in Figure 1

Peak number (see Figure 1)	Retention time (min)	Width (min)	Area	Start time (min)	End time (min)
1	8.154	0.156	330 211 451	7.841	8.912
2	13.084	0.140	555 572 742	12.523	13.453
3	19.125	0.131	444 008 379	18.702	19.485
4	25.666	0.132	576 691 495	25.229	26.189
5	32.272	0.161	990 352 553	31.727	33.174
6	38.637	0.180	1 152 937 448	37.906	39.546

Table 2 Differences in response for various chemicals in two GC detectors. Absorption coefficients and normalization factors are measures of relative response for the same amount of chemical used in the original literature. Exact definitions can be found in refs. 5 and 6

ECD			
Chemical	Absorption coefficient	Chemical	Absorption coefficient
Methylene chloride	1	Chlorobenzene	1
1,3,5-Trichlorobenzene	60	*o*-Dichlorobenzene	42
Chloroform	500	*p*-Dichlorobenzene	11
Hexachlorobenzene	1100	Vinylchloride	0.1
FID			
Chemical	Normalization factor	Chemical	Normalization factor
n-Hexane	1.00	1-Butanol	0.66
Cyclopentane	0.98	Acetic acid	0.24
Benzene	0.97	Ethylacetate	0.38
2,2,4-Trimethylpentane	0.98	Acetone	0.49
1-Methyl-3-ethylbenzene	1.03	Ethyl phenyl ketone	0.80

FID, flame ionization detector.

and nitrogen-phosphorus detectors exhibit differences in response factors by orders of magnitude as shown in Table 2.[5] Other detectors such as the FID show a small range of responses within a chemical group such as alkanes though response factors may differ widely between chemical families.[6] In practice, this means that calibration curves should be made for each chemical. Creating individual calibration curves is not burdensome in GC methods since the separation capabilities allow standard solutions to be made as mixtures and calibrations to be made simultaneously for many constituents with a few analyses. The results in Figure 1 were taken from a mass spectrometer detector and the peaks are all alkanes with comparable response factors. Not all peaks have same intensity or height (or area), so differing concentrations between the constituents may be reasonably concluded (Table 1).

If chromatographic performance between samples and standards is identical or very similar, peak heights are suitable for quantitative work or peak areas may be used and are especially helpful when slight variations exist in chromatographic performance. These terms are generated automatically with modern gas chromatographs which are manufactured with data analysis options of varying degrees of sophistication. Difficulties arise in the automated determination of areas for partially resolved components. In the instance when best chromatographic performance is obtained and peaks are still not totally resolved, deconvolution of the chromatogram may be helpful. At present, commercial software packages allow deconvolution of peaks as shown in Figure 2 and these afford refined methods for obtaining peak areas.[7] The additional time and expense in data reduction for these deconvolutions is returned in analytical information that may not be obtainable otherwise.

2.3 Qualitative Aspects of Gas Chromatography Data

A second aspect of data reduction is the determination of the chemical identity for a peak and there is no true substitute for a mass spectrum of a peak as shown in Figure 1. The combination of prefractionation by GC with chemical identification by mass spectrometry (MS) explains, in part, the explosive growth of GC/MS instruments and methodologies. This principle about chemical identity cannot be diminished and is especially critical for samples where there is little or no understanding of possible constituents. In some instances where the number of constituents is limited and the sample is drawn from a well-characterized and fairly stable source, retention time of a peak (t_r) is a de facto measure of chemical identity. In these instances, this measure will suffice as all the proof necessary for an assignment of identity. In other instances, nothing short of a mass spectrum for a peak in a chromatogram and a confirmation of the analyte retention time with an authentic sample will be satisfactory for identifying a constituent in a chromatogram. Regardless, the retention time (t_r) is and has been an essential term in GC analyses and is best reported as the corrected retention time t'_r where the time is corrected and normalized to the time for an unretained solute (t_m) as shown in Equation (1). While selection of an unretained solute is more difficult than may be expected, methane is commonly used when column temperatures are high and the detector is a FID (hence the term t_m). In mass spectrometers, where

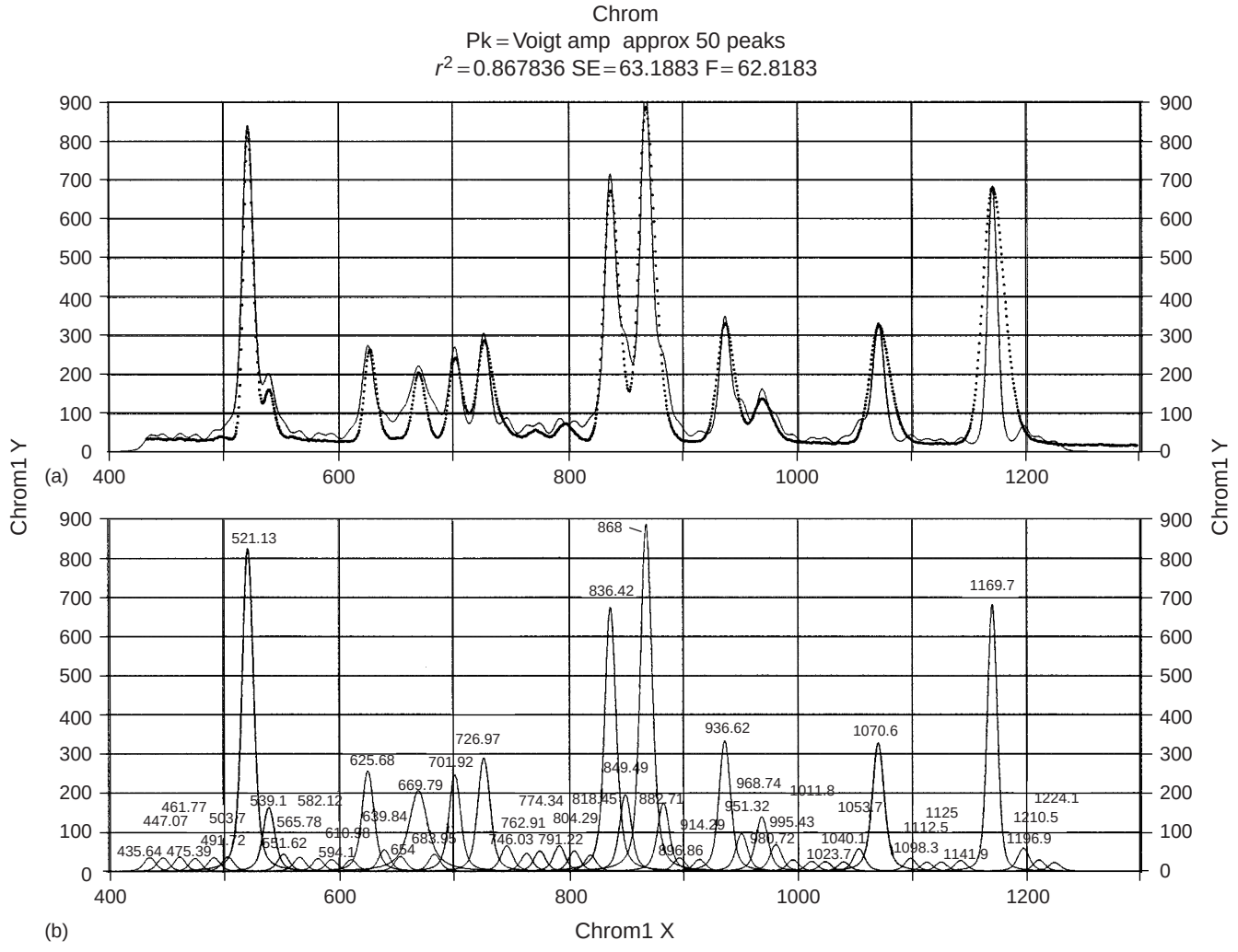

Figure 2 The use of deconvolution in obtaining peak areas for a chromatogram (a) in which there is incomplete chromatographic resolution. (b) The individual peaks as seen from deconvolution are shown. The results were generated using PeakFit 4.0.

the filament is usually started after the solvent or air peak has eluted, a reliable determination of t_m is inconvenient and must be made in a separate determination.

$$t'_r = t_r - t_m \tag{1}$$

The great facility of a corrected retention time is that there is a direct reference to the partition coefficient (K) of the analyte via the phase ratio for the column (β) as shown also in Equation (2). The value K is dependent upon stationary phase,

$$\frac{t'_r}{t_m} = \frac{K}{\beta} \tag{2}$$

chemical compound, and temperature. So, if stationary phase and temperature are controlled, K (and thus t'_r) will be characteristic for a given chemical.

Unfortunately, phase ratios are difficult to obtain or modify so practical chemical measurements require a means to compare results between laboratories. One means used in recent decades to allow comparison of chromatographic results on a given stationary phase is the use of chemical standards to calibrate retention times. A common standard comprised a homologous series of normal hydrocarbons and was called the Kovats index.[8,9] In this method, the retention of a chemical was expressed in carbon numbers, rather then retention time, by referencing retention for constituents to the retention of a set of *n*-alkane standards. Any differences in physical parameters of a column (flow, length, temperature, and phase ratio) would be compensated by this calibration of retention to chemical standards. Such methods were particularly useful since many chromatographers used self-made columns where experimental variables ranged widely. With the advent of commercial capillary columns manufactured with stringent quality control, retention times are matched via software to stored library values

and adjustments using carrier gas flows are made to meet standard t_r values. In this way, retention times are directly intercomparable between instruments or laboratory using retention timescales though retention on a column, in a sense, is still being calibrated against a chemical standard.

2.4 The Chromatogram as Diagnostic Tool

A final aspect of the chromatogram is that which provides a full measure of the performance of the gas chromatograph including injection port, column, and detector. A full discussion can be seen in sections on trouble shooting[1] and a brief discussion is given here.

Peaks that exhibit tailing are usually an indication of adsorptive interactions between the sample and surfaces in the inlet, column, or detector, poor flow connections, or overloading of the stationary phase. Active sites, that is highly polar or reactant surfaces in the inlet, column, or detector, are usually remedied through chemical deactivation of these sites. Alternatively, improved materials or new columns or improved designs may be needed.

Peaks that show front skewing are indicative of excessive sample loading on the column and the partition coefficient has been pushed into a region where the isotherm of C_l versus C_g is nonlinear where C_l and C_g are the concentrations of the sample in the liquid and gas phases respectively. This is shown in Figure 3.

When a hydrocarbon sample shows tailing on a fresh, nonpolar or semipolar column, the responsibility is usually poor connections or fitting and is shown in Figure 3. Drifting baselines and more are treated in a recent monograph that experimentalists will find useful.[1]

Band-broadening is a normal part of a chromatogram[10–12] and many of the advances in high resolution chromatography have occurred through improved management of band-broadening.[13] Additional band-broadening can arise through the injection of sample and other extra column contributions. These and baseline changes are treated in a series of practical guides.[14–16]

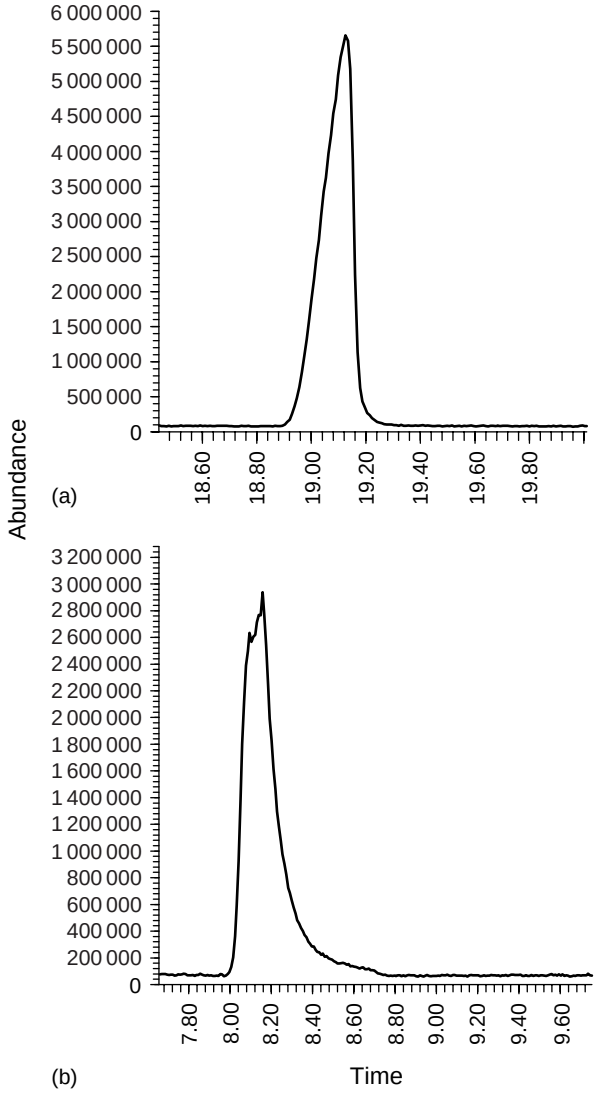

Figure 3 Two examples of peak distortion. (a) shows overloading of the liquid phase and (b) shows tailing indicative of a poor connection. The asymmetry factor is a ratio of distances from a center line.

3 ADVANCED DATA REDUCTION METHODS IN GAS CHROMATOGRAPHY

When chromatograms are simple and information regarding a single (or a few substances) is sought for a measurement, interpretation of the chromatogram can be uncomplicated and rapid. However, when chromatograms contain large numbers of peaks and the patterns defy simple interpretation, advanced methods of data reduction can be useful. There have been two approaches, broadly speaking, to coping with complex samples and these include principal component analysis or cluster analysis and neural network methods. Both methods are computationally intensive and have blossomed with the rise of personal computers in the 1980s and 1990s.

3.1 Multivariate Data Analysis in Gas Chromatography

In multivariate data analysis, data are manipulated so that subtle relationships between constituents in the data can be resolved or clarified. This is done by mathematical methods of recreating the reference points in the information contained in the data and the details of these methods can be found in the articles referenced below. The goal in multivariate analysis is to amplify similarities or differences in data for various samples and generally minimize the number of variables that

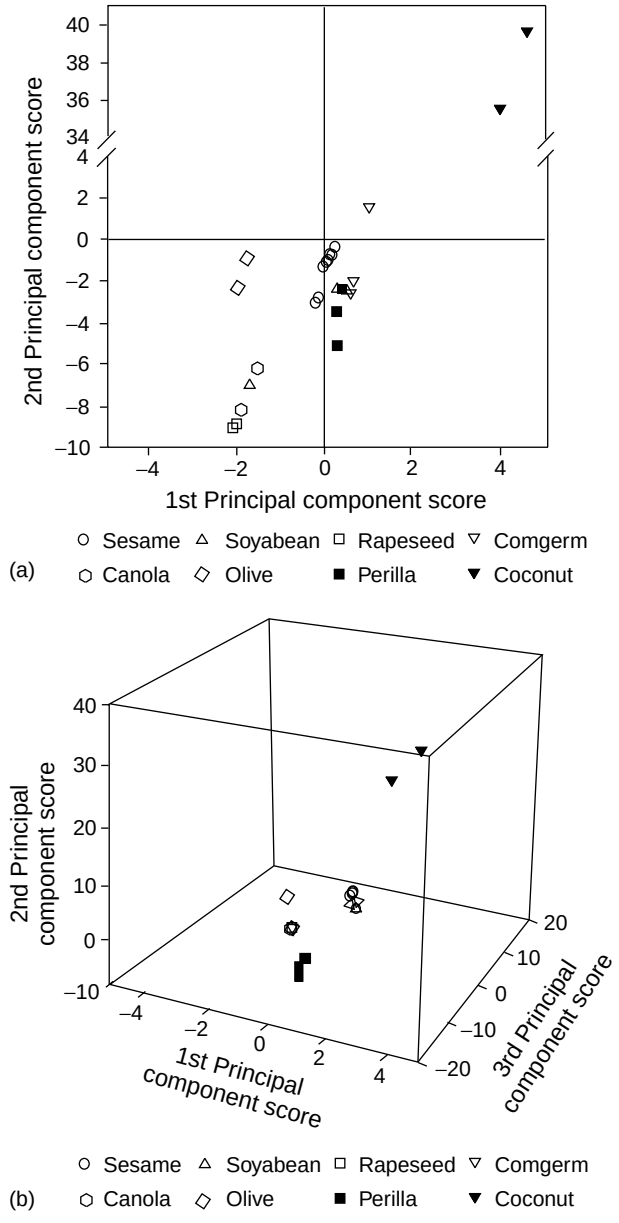

Figure 4 Principal components plot for fatty acid content of vegetable oils in 2-D (a) and 3-D (b) plots. (Drawn from the article by Lee et al.,[17] used with permission.)

describe a sample. The tools and rules to identify or create the terms for comparisons (principal components in principal component analysis) are beyond the scope of this article. It is sufficient to say that certain peaks or regions of a chromatogram may be compared with other peaks as a measure of similarities. When done for several such comparisons, computational methods allow rapid analysis of results. The end goal is the assignment of a sample as a single data point in two-dimensional (2-D) or three-dimensional (3-D) plots or in multidimensional space. The presence of other measurements in proximity to the sample constitutes evidence of a common or unique composition or identity.

An example of one use of these tools is principal component analysis of chromatographic data from characterizing fatty acids in vegetable oils.[17] Though the chromatograms were comparatively simple consisting of only ~10 fatty acids, the number of sample types were numerous including olive, coconut, sesame, and others. The results are shown in Figure 4 where the various types of samples are resolved in 2-D and 3-D.

Another example of a chemometric approach to processing gas chromatographic results can be seen in the highly complex patterns from pyrolysis gas chromatograms of polymers.[18] Polymers were identified or recognized in poor resolution GC through the presence of functional groups such as olefin, aromatic, amide bonds, carbonyls, and others. Such methods of data analysis may be expected to be beneficial in biomedical and medical applications of GC. One example of a successful application of chemometrics in biochemical measurements was demonstrated for urinary organic acid profiles and uterine cervical cancer.[19] In this work, a total of 50 organic acids from urine samples (0.25 mL) were resolved by GC and discriminant (cluster) analysis was performed on the GC data using 16 acids as the variables for discriminating between the two groupings. Discriminant analysis applied to these 16 variables correctly classified 26 urine samples into two separate clusters according to tumor types in the canonical plot.

Advanced data handling of gas chromatographic data has been used in microbiology where the volatile fatty acids in bacteria, long a reliable method to type bacteria through chemotaxonomy,[20] were characterized by GC. The data, when treated by chemometrics allowed subgrouping of 169 coagulase-negative staphylococci collected during an outbreak of nosocomial sepsis in a hematologic unit.[21] Their findings corroborated the results from conventional typing methods showing

Table 3 Examples of multivariate analysis methods and applications with GC data

Sample	Refs.
Foods and beverages	
Blueberries	22
Orange juice	23
Microbial defects in milk	24, 25
Fresh cabbage	26
Almond cultivars	27
Wines	28
Roots of echinacea	29
Non-foodstuffs	
Sulfur in coal by pyrolysis GC	30
Plastics in recycling	31

similarities in strain characteristics and, therefore, the same epidemiological origin.

Multivariate analysis methods then, are powerful tools to extract from gas chromatograms details that otherwise elude the unaided eye simply owing to the complexity or subtlety of the data. This is especially so with natural materials such as food odors or vapors. Several more examples are listed in Table 3.[22–31]

3.2 Neural Network Analysis in Gas Chromatography

Another approach to comparing complex chromatograms through what may be nonsubjective computational tools is that of neural network analysis. In all neural networks, the software is used to mimic the human ability to discern patterns. In contrast to humans, computers can operate without fatigue and can survey large amounts of data without the time restraints that humans experience. In neural network methods the network learns to recognize data from known samples. In practice, the chromatogram or the intensity versus retention time pattern of the chromatogram are entered into the input layer after which hidden neuron layers establish weight factors that form a correct answer for the known sample. Other known samples can be trained so that other answers are obtained and an overall weight factor for each neuron in each layer is created, providing that the network properly learns the differences between samples. Thus, unknown samples should be rapidly categorized when chromatographic results for an unknown sample are presented to the trained network. This is especially helpful when the chromatograms are complex and defy reasonable comparisons by humans.

Neural networks have been used in situations where the samples are complex and differences might be due to nonlinear causes. This is especially important in instances where uncertainty exists when two complex mixtures are combined and is illustrated nicely with the recognition of Arochlors in environmental samples.[32] In this approach, a simple architecture was developed in which ~10 peaks were used to classify samples on the basis of the level of Arochlor contamination in environmental samples.

In another example, neural networks were used to classify vegetable oils after GC analysis of samples for methyl ester derivatives of fatty acids[33] and to assign botanical origins of vinegar samples.[34] Polyalcohols, rather than volatile chemicals, were found useful in recognizing the country of origin of various vinegar samples. A similar use of neural networks and GC data from pyrolysis analysis of hashish was found successful in discriminating between hashish obtained from different sources.[35] Lastly, defects in milk were screened using neural network analysis with GC results.[36]

Unlike the traditional methods for interpreting chromatograms, these advanced methods can be expected to undergo further refinements in the years ahead. They do, however, establish a basis from which automated data reduction can be successfully completed and used as a powerful tool in the utilization of GC data.

4 FUTURE DEVELOPMENTS

The current standards for data processing in chromatographic systems afford a high level of convenience and precision in qualitative and quantitative determinations; these meet or exceed the analytical needs for resolving power of contemporary columns. Nevertheless, the technology is rooted in principles that have been long a part of GC separations and these may be expected to continue well into the future. In contrast, the methods necessary to compare analytical findings for a large number of samples are in nascent stages of application. The methods are developed but not well advanced in daily applications. These chemometric methods may be expected to undergo refinement and general adoption during the next decade. This may be especially so in view of the growing and now major emphasis placed on biomedical chemistry.

LIST OF SYMBOLS

t_r	retention time
η	asymmetry factor
t_m	retention time of unretained solute
K	partition coefficient
β	phase ratio for a column
C_g	concentration of sample in gas phase
C_l	concentration of sample in liquid phase
w	peak width

ABBREVIATIONS AND ACRONYMS

ECD	Electron Capture Detector
FID	Flame Ionization Detector
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
MS	Mass Spectrometry
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Chemometrics **(Volume 11)**
Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Soft Modeling of Analytical Data

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. H.M. McNair, J.M. Miller, *Basic Gas Chromatography*, John Wiley & Sons, New York, 29–54, 126–141, 1998.
2. B.L. Karger, L.R. Snyder, C. Horvath, *An Introduction to Separation Science*, John Wiley & Sons, New York, 212–246, 1973.
3. D.C. Harris, *Quantitative Chemical Analysis*, 5th edition, W.H. Freeman & Co., New York, 675–712, 1999.
4. G.D. Christian, *Analytical Chemistry*, 5th edition, John Wiley & Sons, 505–514, 1997.
5. J.E. Lovelock, M.L. Gregory, in *Gas Chromatography*, eds. N. Brenner, J.E. Callen, M.D. Weiss, Academic Press, New York, 1962.
6. W.A. Dietz, 'Response Factors for Gas Chromatographic Analyses', *J. Gas Chrom.*, **5**(2), 68 (1967).
7. PEAKFIT 5.0
8. E. Kovats, 'Gas-chromatographische Charakterisierung Organischer Verbindungen, Tiel 1 Retention Sindices Aliphatischer Hologenide, Alkohole, Aldehyde and Ketone', *Helv. Chim. Acta.*, **206**, 1915 (1958).
9. L.S. Ettre, 'The Kovats Retention Index System', *Anal. Chem.*, **36**, 31A (1964).
10. J.J. van Deemter, F.J. Zuiderweg, A. Klinkenberg, 'Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Non-ideality in Chromatography', *Chem Eng. Sci.*, **5**, 271 (1956).
11. H.W. Moody, 'Evaluation of the Parameters in the Van Deemter Equation', *J. Chem. Educ.*, (1982).
12. R.H. Perrett, J.H. Purnell, 'Contribution of Diffusion and Mass Transfer Processes to Efficiency of Gas Liquid Chromatography Columns', *Anal. Chem.*, **35**, 430 (1963).
13. K.J. Hyver, R. Sandra, *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard Corporation, Part No. 5950-3562, 1989.
14. J. Hinshaw, 'GC Troubleshooting – A Systematic Approach', *LC-GC*, **5**, 790–794 (1987).
15. J. Hinshaw, 'GC Troubleshooting – Inlets', *LC-GC*, **5**, 954–960 (1987).
16. J. Hinshaw, 'GC Troubleshooting – Columns', *LC-GC*, **6**, 24–29, 228–231, 794–798.
17. D.-S. Lee, B.-S. Noh, S.-Y. Bae, K. Kim, 'Characterization of Fatty Acids Composition in Vegetable Oils by Gas Chromatography and Chemometrics', *Anal. Chim. Acta.*, **358**, 163–175 (1998).
18. E. Kullik, M. Kaljurand, M. Koel, 'Analysis of Pyrolysis Gas Chromatograms Using Pattern Recognition Techniques', *J. Chromatogr.*, **126**, 249–256 (1976).
19. K.R. Kim, H.G. Park, M.J. Paik, H.S. Ryu, K.S. Oh, S.W. Myung, H.M. Liebich, 'Gas Chromatographic Profiling and Pattern Recognition Analysis of Urinary Organic Acids from Uterine Myoma Patients and Cervical Cancer Patients', *J. Chromatogr., B*, **712**(1–2), 11–22 (1998).
20. G.A. Eiceman, W. Windig, A.P. Snyder, 'Gas Chromatography of Microorganisms', in *Gas Chromatography: Biochemical, Biomedical, and Clinical Applications*, ed. R.E. Clement, John Wiley & Sons, New York, 327–347, 1990.
21. P. Kotilainen, P. Huovinen, E. Eerola, 'Use of Gas-liquid-chromatography for Subgrouping Coagulase-negative Staphylococci During a Nosocomial Sepsis Outbreak', *Eur. J. Clin. Microbiol. Infect. Dis.*, **14**(5), 412–420 (1995).
22. D.W. Baloga, N. Vorsa, L. Lawter, 'Dynamic Headspace Gas Chromatography-Mass Spectrometry Analysis of Volatile Flavor Compounds from Wild Diploid Blueberry Species', *ACS Symp. Ser.*, **596**, 235–247 (1995).
23. P.E. Shaw, M.G. Moshonas, B.S. Buslig, 'Multivariate-analysis for Classification of Commercial Orange Juice Products by Volatile Constituents Using Headspace Gas-chromatography', *ACS Symp. Ser.*, **596**, 33–47 (1995).
24. Y. Horimoto, K. Lee, S. Nakai, 'Classification of Microbial Defects in Milk Using a Dynamic Headspace Gas Chromatograph and Computer-aided Data Processing. 1. Principal Component Similarity Analysis', *J. Agric. Food Chem.*, **45**(3), 733–742 (1997).
25. Y. Horimoto, K. Lee, S. Nakai, 'Classification of Microbial Defects in Milk Using a Dynamic Headspace Gas Chromatograph and Computer-aided Data Processing. 2. Artificial Neural Networks, Partial Least-squares Regression Analysis, and Principal Component Regression Analysis', *J. Agric. Food Chem.*, **45**(3), 743–747 (1997).
26. H. Siegl, A. Hanke, W.H. Schnitzler, 'Cultivar Differences of Volatile Aroma Compounds in Fresh Chinese Cabbage', *Gartenbauwissenshaft*, **62**(1), 17–22 (1997).
27. M.L.M. Carratala, C. Garcia-Lopez, V. Berenguerna-Varro, N. Graneteruel, 'New Contribution to the Chemometric Characterization of Almond Cultivars on the Basis of their Fatty Acid Profiles', *J. Agric. Food Chem.*, **45**(3), 743–747 (1997).
28. J. Weber, M. Beeg, C. Barzsch, K.H. Feller, D.D. Garcia, M. Reichenbacher, K. Danzer, 'Improvement of the Chemometric Variety Characterization of Wines by Improving the Detection Limit for Aroma Compounds', *HRC-J. High. Res. Chrom.*, **22**(6), 322–326 (1999).
29. D. Lienert, E. Anklam, U. Panne, 'Gas Chromatography Mass Spectral Analysis of Roots of Echinacea Species and

Classification by Multivariate Data Analysis', *Phytochem. Anal.*, **9**(2), 88–89 (1998).
30. P. Selsbo, P. Almen, I. Ericsson, 'Quantitative Analysis of Sulfur in Coal by Pyrolysis Gas Chromatography and Multivariate Data Analysis', *Energy Fuels*, **10**(3), 751–756 (1996).
31. P.J. Tandr, J.A. Butcher, H. Tao, P.D. Harrington, 'Analysis of Plastic Recycling Products by Expert-systems', *Anal. Chim. Acta.*, **312**(3), 231–244 (1995).
32. G.R. Magelssen, J.W. Elling, 'Chromatography Pattern Recognition of Aroclors Using Iterative Probabilistic Neural Networks', *J. Chromatogr., A*, **775**, 231–242 (1997).
33. R.A. Francelin, F.A.C. Gomide, F.M. Lancas, 'Use of Artificial Neural Networks for the Classification of Vegetable-oils after GC analysis', *Chromatographia*, **35**(3–4), 160–166 (1993).
34. V. Gerbi, G. Zeppa, R. Beltramo, A. Carnacini, A. Antonelli, 'Characterisation of White Vinegars of Different Sources with Artificial Neural Networks', *J. Sci. Food Agric.*, **78**(3), 417–422 (1998).
35. M. Hida, T. Mitsui, Y. Minami, Y. Fujimura, 'Classification of Hashish by Pyrolysis-gas Chromatography', *J. Anal. Appl. Pyrolysis*, **32**, 197–204 (1995).
36. Y. Horimoto, K. Lee, S. Nakai, 'Classification of Microbial Defects in Milk Using a Dynamic Headspace Gas Chromatograph and Computer-aided Data Processing. 2. Artificial Neural Networks, Partial Least-squares Regression Analysis, and Principal Component Regression Analysis', *J. Agric. Food Chem.*, **45**(3), 743–747 (1997).

Hyphenated Gas Chromatography

W.M. Coleman, III and Bert M. Gordon
R.J. Reynolds Tobacco Company Winston-Salem, USA

1 Introduction 1

2 Pre-gas Chromatograph Automated On-line Sample Preparation Techniques 2
- 2.1 Solid-phase Microextraction 2
- 2.2 Solid-phase Microextraction/Gas Chromatography 4
- 2.3 Automated Solid-phase Microextraction/Gas Chromatography Application 4
- 2.4 Headspace/Gas Chromatography 4
- 2.5 Static Headspace/Gas Chromatography 5
- 2.6 Dynamic Headspace (Purge and Trap)/Gas Chromatography 5
- 2.7 Applications of Static and Dynamic Headspace Gas Chromatography 8
- 2.8 Large-volume Injection/Gas Chromatography 8
- 2.9 Applications of Large-volume Injection/Gas Chromatography 9
- 2.10 Multidimensional Gas Chromatography 10
- 2.11 Applications of Multidimensional Gas Chromatography 11

3 Post-gas Chromatograph Sample Analysis Techniques 12
- 3.1 Gas Chromatography/Mass Spectrometry 12
- 3.2 Applications of Gas Chromatography/Mass Spectrometry 15
- 3.3 Gas Chromatography/Fourier Transform Infrared Spectroscopy 15
- 3.4 Gas Chromatography/Fourier Transform Vapor Phase Infrared Spectroscopy 15
- 3.5 Gas Chromatography/Fourier Transform Matrix Isolation Infrared Spectroscopy 16
- 3.6 Applications of Gas Chromatography/Fourier Transform Matrix Isolation Infrared Spectroscopy 19

4 Conclusions 19

Abbreviations and Acronyms 19

Related Articles 20

References 21

Hyphenated gas chromatography (GC) refers to not only the coupling of a GC to information-rich detectors but also the coupling of the gas chromatograph to automated sample preparation techniques. Examples of information-rich detectors include mass and infrared spectrometers, whereas automated sample preparation techniques include solid-phase microextraction (SPME), large-volume injection (LVI), purge and trap (PT), headspace (HS), and gas chromatographic approaches. The gas chromatographic approach involves the wedding of two gas chromatographs and is commonly referred to as multidimensional gas chromatography (MDGC). The union of the automated sample preparation techniques with MDGC and information-rich detectors affords the capability for quantitative and qualitative analysis of a wide variety of sample matrices for analytes at parts per billion (ppb) concentrations. However, not every hyphenated configuration is worth considering, and careful matching of the hyphenated approaches is essential to obtain the maximum analytical benefits.

1 INTRODUCTION

Hyphenated GC has for the most part traditionally referred to the coupling of a powerful separation technique, namely capillary GC, with powerful information-rich detectors such as mass spectrometry (MS) or infrared spectrometry (IR). The term ''hyphenation'' was first popularized by Hirschfeld in 1980.[1] However, more properly, hyphenated GC should refer to not only the coupling of GC with information-rich detectors but also the coupling of GC to novel automated on-line sample preparation techniques such as PT, static HS, LVI, and SPME. Likewise, the wedding of two gas chromatographs, referred to as MDGC can easily be classified as a hyphenated gas chromatographic technique. In this case, the separation power of at least two gas chromatographs with columns of different separation capabilities is brought to bear. It now becomes readily apparent that the merger of advanced sample preparation techniques with potent information-rich detectors in conjunction with the separation capabilities of capillary GC provides for a powerful analytical approach for the analysis of volatile and semivolatile compounds.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The union, for example, of LVI with MDGC and MS can provide an analytical technique that has detection limits from ppb to parts per trillion (ppt), with typical precision of ±10%. In a similar fashion, the union of MDGC with IR and MS can provide both qualitative (structural) and quantitative (amounts) of target compounds in very complex mixtures such as natural product extracts. The union of SPME with GC and MS can afford the capability for the analysis of ppb levels of target compounds in environmental samples. The underlying theory to hyphenated GC, that is the union of GC with information-rich detectors and automated on-line sample preparation techniques, rests with the fact that the information gained from the data generated by a GC hyphenated approach possesses the potential to be several orders of magnitude more powerful and useful than that obtained from employment of the techniques separately. This does not mean that every hyphenated technique is worth using or even considering. For example, coupling a GC fitted with a column having inadequate separation capabilities with an MS defeats the purpose of the technology. Thus, adequate thought as to the distinct advantages of each component of the marriage must be carefully considered before confirming the union. In other words, thoughtful and careful matching of hyphenated approaches is essential to the realization of the desired benefits.

The combinations possible using only a few mergers of GC separation with on-line sample preparation and detectors are almost endless, particularly when additional separation techniques such as liquid chromatography (LC), supercritical fluid chromatography (SFC), and capillary electrophoresis (CE), as well as additional detectors such as inductively coupled plasma mass spectrometry (ICP/MS), are coupled to a GC. For the purposes of this account, the sample preparation techniques (i.e. pre-GC) discussed are SPME, static HS, PT, LVI, and MDGC. The information-rich detectors (i.e. post-GC) discussed are MS (unit mass and high resolution) and IR (vapor phase (VP) and matrix isolation (MI)).

2 PRE-GAS CHROMATOGRAPH AUTOMATED ON-LINE SAMPLE PREPARATION TECHNIQUES

Sample-enrichment techniques in chromatographic analyses are generally employed because the concentration of the analyte of interest is below the detection limit of the analytical instrument. In some other cases the analyte must be removed from its matrix because introduction of the sample matrix would not be compatible with the desired chromatographic technique. In either case, the analyte must be accumulated in some fashion. This is usually accomplished by transferring the analytes from the parent matrix into or onto another matrix having a well-defined character which is compatible with the chromatographic procedure.

2.1 Solid-phase Microextraction

The information presented for SPME theory and practice is a concise summary of a comprehensive treatise of the topic covered by Pawliszyn.[2] SPME is fundamentally a solvent-free sample preparation/enrichment technique having qualitative and quantitative potential. A relatively thin film extracting phase of very small volume, $\leq 1\,\mu L$, is firmly coated and bound to a fused silica fiber which in turn can be exposed to a sample matrix. The sample matrix can be room air or aqueous solutions and in some limited cases organic solvents. The extracting phase bound to the fiber is, in a large number of cases, very similar to the phases employed in capillary GC technology. Polar or nonpolar phases such as poly(dimethylsiloxane) (PDMS) and carbowaxdivinylbenzene (CWDVB) and combinations of each can be employed.

The most convenient configuration of SPME technology is depicted in Figure 1. The thin fragile fiber is housed when not in use inside a tube-like structure, which is reminiscent of a GC syringe needle. Exposure of the fiber to the sample matrix of interest is effected by gently sliding the fiber outside of the housing and into the matrix. This exposure can be performed manually or automatically and both devices are available from Supelco, Inc. (Bellefonte, PA, USA). The partition coefficient of the analyte of interest between the sample matrix and the fiber coating material governs the amount of analyte extracted by the fiber coating. Obviously then, an analyte equilibrium can eventually be established between the sample matrix and the fiber. This behavior renders the SPME approach ideal for quantitative analysis. Furthermore, the presence of a very small amount of coating, which in turn removes very small amounts of analyte, is not likely to alter the sample matrix to any significant extent. SPME can also be used in qualitative analyses by employing a variety of coatings varying in polarity.

From a quantitative perspective, the mass of analyte extracted by the coating is linked to the equilibrium of the analyte in a three-phase system defined by the sample matrix (such as water or air), the fiber coating, and the HS of the sample (Figure 1). Assuming control of the volume of the fiber and its coating, the HS, and the sample matrix, placement of the fiber is not critical. Thus, extracts of the HS above a sample of interest or extracts of the matrix itself can be obtained.

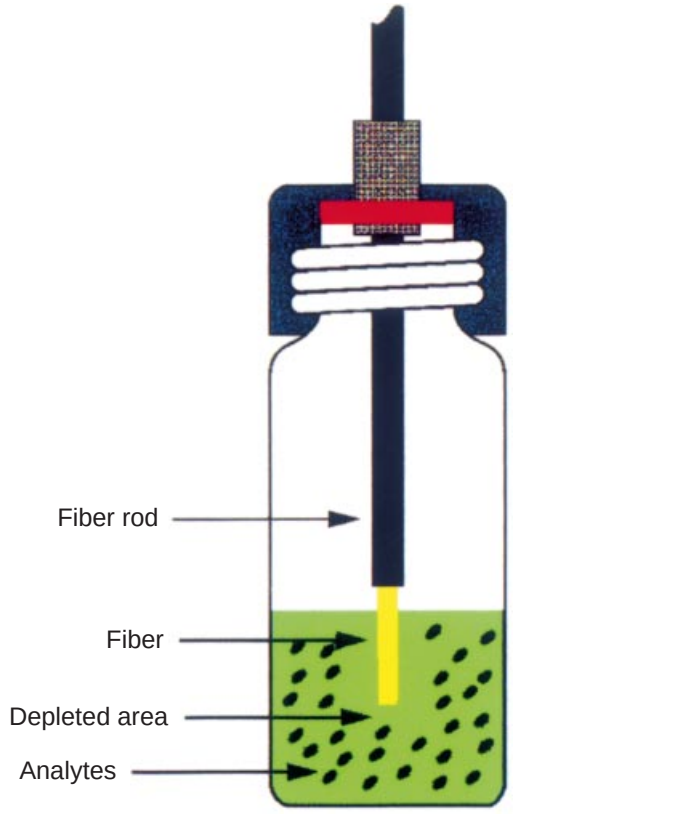

Figure 1 Description of an SPME extraction device.

Consequently, direct contact of an SPME fiber with a well-defined sample configuration can lead to quantitative results.

Several extraction parameters have been documented as having meaningful impacts on qualitative and quantitative aspects of the amount of analyte extracted by the fiber. These include fiber conditioning, fiber type, exposure time and location (liquid or HS), temperature, salting (sample ionic strength), pH, and agitation.[2–6] There are, thus, some practical steps that can be taken to optimize each of these parameters to the benefit of the desired end result. To a first approximation, selection of the fiber type can be governed by the age-old adage that ''like dissolves like''. For example, if polar analytes such as flavor compounds are of interest, then a polar material such as CWDVB would be a logical first choice. A combination of temperature with exposure location can be used to an advantage. For example, sampling of the HS above a particularly complex matrix at increased temperature could possibly reduce the complexity of the materials extracted by the fiber. Adjustment of the pH of an aqueous sample can alter the nature of the species in solution and thus alter the extraction performance of the fiber. Adjustment of the ionic strength of an aqueous sample with the addition of simple ions from salts such as NaCl can bring about the salting-out phenomenon, thereby improving the performance of the fiber towards the analyte of interest. Agitation of the sample through sample stirring in liquid cases improves the performance of the SPME fiber; likewise, use of fiber vibration also improves the performance of the fibers, particularly as it relates to required sampling times. It should be mentioned here that control of the sampling time is very critical in obtaining reproducible results. Autosamplers can control sampling times to the second, thus excellent precision can be obtained even though equilibrium has not yet been reached.

The successful implementation of SPME technology rests in large part on a properly conditioned SPME fiber. Failure to properly condition an SPME fiber can result in unacceptable accuracy and precision. Fortunately, proper conditioning parameters have been well established for all of the commercially available SPME fibers, and these should be followed to the letter to ensure acceptable performance of the fibers. During use, fibers should be periodically examined closely, usually employing a microscope to ensure that no loss in phase coating has occurred through use.

2.2 Solid-phase Microextraction/Gas Chromatography

Having now established basic theories surrounding SPME, there are some operating principles associated with effective coupling of SPME to GC that must be addressed. Most SPME applications have been developed for GC, with a few applications for high-performance liquid chromatography (HPLC). The successful interfacing of SPME to GC, to produce the hyphenated technique SPME/GC, rests with the theory that analytes captured on the nonvolatile thin film can be efficiently thermally desorbed in a reproducible fashion. This process can be viewed as the reverse of extraction from a sample matrix. To accomplish desorption effectively and reproducibility, a high carrier gas linear velocity coupled with an increased temperature must be consistently established. Furthermore, the time for desorption must be carefully controlled even though theory would indicate that desorption times of seconds at 200 °C should be sufficient for most volatile or semivolatile analytes.[2]

Standard GC injection ports, such as split/splitless, can be effectively applied to the reproducible thermal desorption of analytes from an SPME fiber. One slight modification of the inert injection port liner is required to ensure the necessary linear velocity around the fiber. In general this involves changing the standard split/splitless liner of 3–5 mm internal diameter (i.d.), for a liner having an i.d. of approximately 0.8 mm and operating the injection port in the splitless mode. This small change results in an injection port volume change in approximately 3 s for the 0.8 mm linear compared with 44 s for the 3 mm liner. Such a dramatic change positively influences the desorption process, resulting in analytes being effectively transferred to the GC column in a narrow band. Additional ways to produce analytes in a narrow band include using a capillary GC column with a thick film stationary phase or cryofocusing the analytes at the front of the GC column using subambient GC oven temperatures. Rapid injection of the needle and simultaneous exposure of the fiber also facilitate the concentration of analytes in a narrow band. The formation of a focused narrow band of analytes directly effects the performance of the separation, either maintaining or reducing the maximum resolution of the GC column.

2.3 Automated Solid-phase Microextraction/Gas Chromatography Application

The enantiomeric distributions of volatile components in essential oils can be accurately and precisely determined using automated SPME/GC.[2,7] Employing a PDMS SPME fiber of 7 μm film thickness, a GC fitted with a column capable of separating optical isomers and an MS as detector, the optical isomer distributions in a grapefruit and a lavender essential oil were determined (Figure 2). The approach was relatively simple, involving automated sampling of the HS above the essential oil samples contained within a sealed 1.8 mL sample vial. The results obtained from the SPME approach were very precise having an average percentage relative standard deviation (RSD) of less than 6. The results were also accurate, agreeing very well with earlier findings using a heartcutting GC approach. The accurate and precise results were obtained without establishing equilibrium conditions, because such essential sampling parameters as fiber exposure time, sample HS volume, fiber desorption temperature, and sample temperature were carefully controlled. Likewise, evaluations of selected fiber types were conducted until the nonpolar fiber PDMS was found to yield reproducible results.

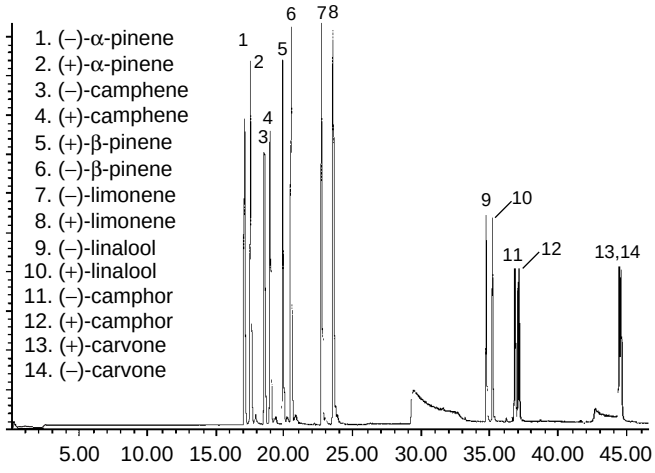

Figure 2 Auto SPME/chiral GC/MSD (mass selective detector) of the HS above selected standard essential oil components. (Reproduced courtesy of Preston Publications.)

2.4 Headspace/Gas Chromatography

HS analyses using GC can be subdivided into three general categories.[8–12]

1. The sample for chromatographic analysis is obtained from a closed vessel where the material under investigation is caused to come into equilibrium with its vapor at a specified temperature. Once equilibrium is established the concentration of the analyte does not change with time if the temperature is held constant. When a sample is taken from the ''static HS'' of an isothermally sealed vessel the equilibrium is altered. Thus, the procedures associated with taking a sample must be reproducible.
2. The sample for chromatographic analysis is obtained from a gas stream which has been purged through the sample matrix, usually water. The purge gas, now

laden with the analytes, is usually passed through a trap which is inert to the purge gas but capable of efficiently capturing the analytes. Once trapped, the analytes are usually thermally desorbed onto the GC column. This PT approach is limited by some extent to the capacity of the trapping matrix, thus breakthrough volumes must be carefully established.

3. The sample for chromatographic analysis is obtained from a gas stream that has been passed over the sample matrix, such as soil or over the top of a water sample. As expected, the gas stream in this case is much less efficient when compared to the approach described in category 2 above.

2.5 Static Headspace/Gas Chromatography

Static (equilibrium) HS analysis is based on the theory that an equilibrium between a condensed phase and a gaseous phase can be reproducibly established for the analytes of interest and that the gaseous phase containing the analytes can be sampled reproducibly. Naturally, this approach will not be very successful when the analytes of interest possess very low vapor pressures because very little of the analyte will be found in the gaseous phase. Elevated temperatures, changes in pH, and the presence of additional electrolytes can in certain cases be employed to increase the analyte vapor pressures.

When coupling the static HS sample-enrichment approach to a GC, several key details and parameters must be carefully attended to. The combined instrumentation usually consists of two individual units, the static HS unit and the GC with appropriate detector. The two units are connected via a thermally controlled inert transfer line that serves as a source of carrier gas for the GC as well as a mode for the transfer of the analytes from the static HS unit to the GC column. In some cases, the two units are also linked and synchronized for automated sample preparation and analysis by software programs available from the manufacturers. Flow control and purity of the carrier gas are essential. Flow is normally controlled in the HS unit employing flow control valves. As the basic premise of static HS experiments rests with the presence of a thermally stable sealed environment, providing a leak-free thermally stable instrument configuration is fundamental to the successful operation of the system. All commercially available static HS systems have detailed descriptions of processes necessary to ensure a leak-free environment. Accurate and durable thermocouples are also provided.

The operation of the instrumentation usually follows the following sequence of events. Appropriate carrier gas flows and temperature zones are established. Additional carrier gas flow is initiated to sweep the lines, sample loop, and needle. A sample in a sealed vial is allowed to equilibrate at an elevated temperature for a specified length of time (Figure 3). The sealed sample vial is raised onto a needle that punctures a septum and pressurization gas fills the vial to a predetermined level. The vial is allowed to equilibrate for a relatively short time to ensure complete diffusion of the pressurization gas with the sealed sample vial's atmosphere. A vent valve is opened and the pressurized contents of the sealed sample vial exit the system through a thermally controlled sample loop of previously selected volume, usually $\geq$1 mL. The vent valve is then closed and the contents of the loop allowed to equilibrate for a specified time. Next a multiple port valve is activated, placing the sample loop in the carrier gas stream. The carrier gas then sweeps the contents of the loop through the heated transfer line and into the GC. Usually upon initiation of sample transfer to the GC, instrumentation software is employed to automatically begin the chromatographic separation and data collection. Automated multiple sampling configurations can be employed that will perform repetitive analysis of multiple vials, essentially repeating the steps outlined above.

From the description of the operation and sequence of events outlined above for the static HS/GC experiment, an appreciation of the number of possible variables can be obtained. Thus, successive samples undergoing incremental changes in a number of static HS parameters could be used to optimized the analytical procedure. The parameters that are manipulated in order to optimize the procedure include times, temperatures, and pressures associated with sample heating, sample equilibration, loop fill, loop equilibration, and sample transfer. In some cases, manufacturers will provide software that allows for automated systematic changes in these and other key parameters, leading to an optimized approach.

2.6 Dynamic Headspace (Purge and Trap)/Gas Chromatography

The fundamental underlying tenet governing dynamic HS (i.e. PT) analysis, is that the change in the mass of a volatile or semivolatile analyte with time can be expressed in terms of the volumetric flow rate of a stripping inert gas.[9] That is, it is physically possible to strip volatile and semivolatile compounds from solid or liquid matrices by passing an inert gas over or through them and that the amount of material stripped can be related to the amount of gas passed through the matrices. Coupling this physical behavior with the capacity to trap the stripped materials on an inert trapping media affords an excellent opportunity for sample enrichment. Several important assumptions underlie the acceptable performance of PT:

- the concentration of the analyte in the stripping gas is equilibrated with the concentration in the condensed phase;

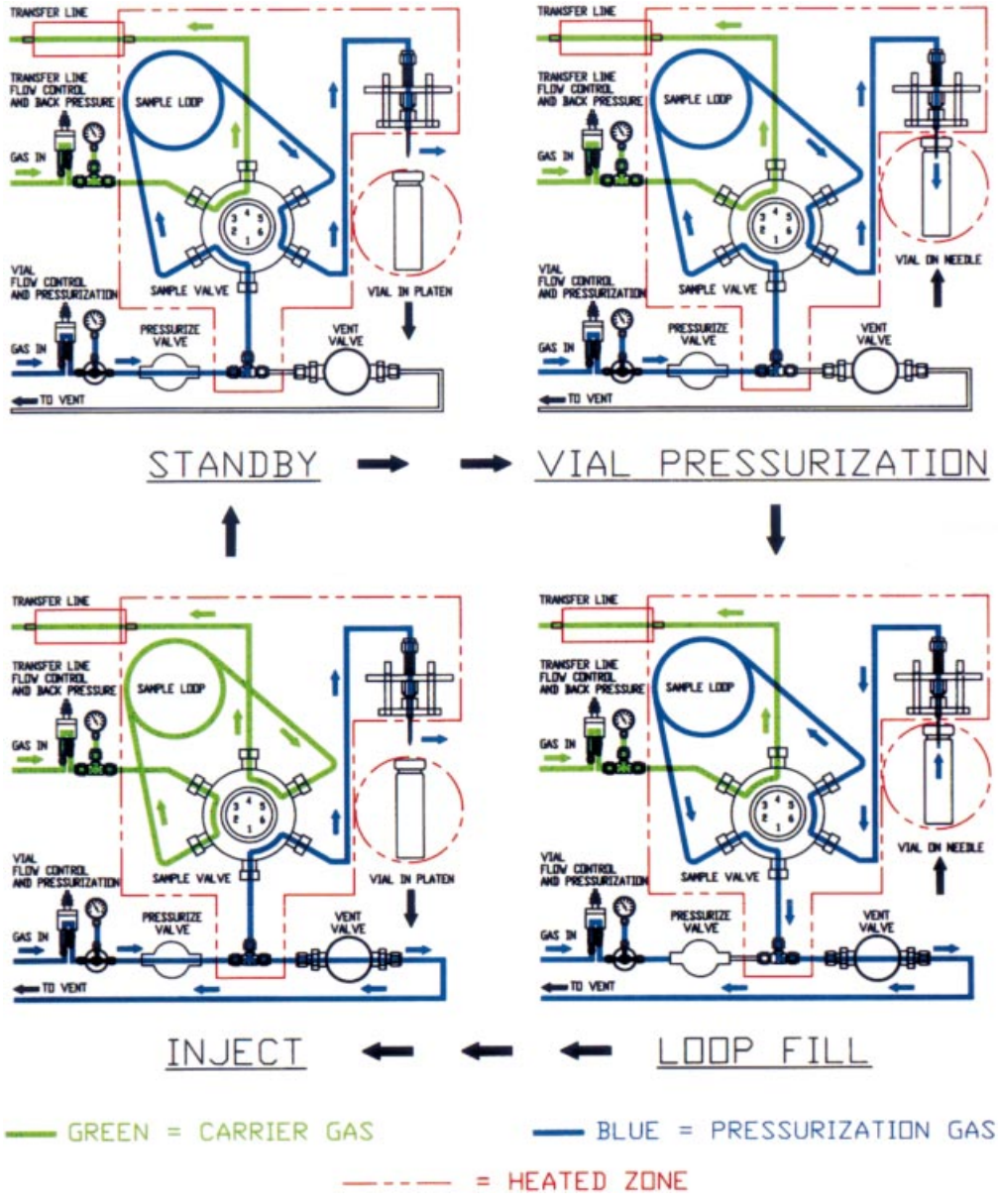

Figure 3 A typical sequence of events in static HS analysis. (Reproduced courtesy of Tekmar Instrument Company.)

- the liquid or solid matrix is essentially nonvolatile;
- the distribution constant of the analyte is not dependent on the analyte concentration.

The issues associated with coupling a PT unit to a GC are very similar to those described above for the coupling of a static HS unit to a GC. The two units are connected via a thermally controlled inert transfer line that can serve as a source of carrier gas for the GC as well as a mode of transfer for the analytes from the PT unit to the GC. Most often the PT and GC units are linked and synchronized for automated sample preparation and analysis. A leak-free environment is critical with PT as well as thermal stability.

The manipulation of the instrumentation follows the general sequence of events described below. Carrier gas flow compatible with the GC is established and temperature zones are equilibrated. In PT techniques, the system is not sealed as in the static approach, thus the HS above the sample need not come to equilibrium. Instead, the sample is placed into a chamber, at a preselected temperature, that is sparged with carrier gas at a specified rate and time (Figure 4). The sparging may take place by allowing the carrier gas to pass through a liquid or solid or over the HS above the liquid or solid. The sweeping carrier gas removes the analytes from the matrix and transports them under a thermally controlled

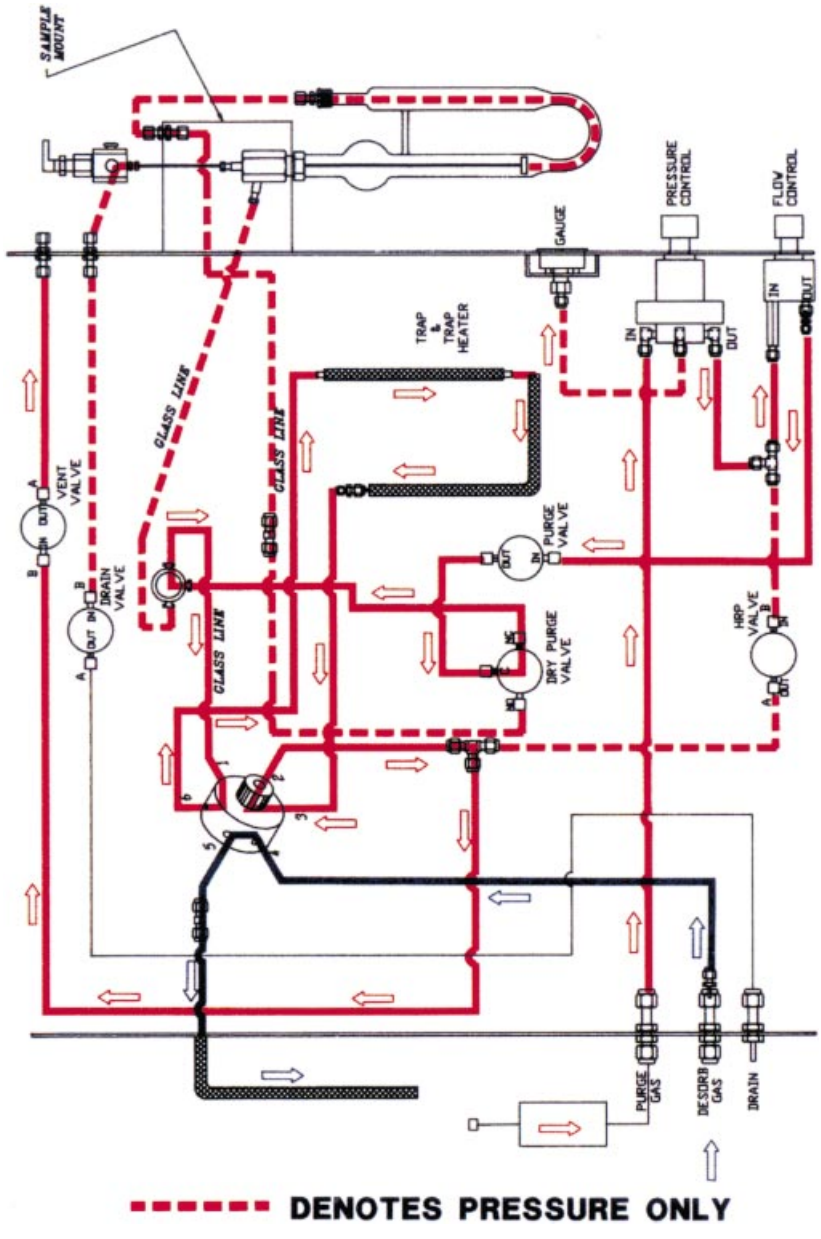

Figure 4 Typical instrument configuration for PT. (Reproduced courtesy of Tekmar Instrument Company.)

environment to a trap. The trap is usually allowed to come close to room temperature prior to transfer of the analytes. After a specified time has elapsed the sparging is stopped. It is important to note here that the characteristics of the solid sorbent material contained within the trap can vary over a wide range. The amount and type of trapping material is usually preselected, based on the types of analytes under consideration. Once the analytes have been stripped from the matrix to the trap, a multiple port valve is activated and the trap is heated and backflushed, thereby desorbing the analytes from the sorbent material and transferring the analytes to the GC through the heated transfer line. The GC column can be cryocooled below ambient temperatures to facilitate the capturing of the analytes in a narrow band at the head of the GC column. Once the transfer of the analytes to the GC is complete, computer software programs can be employed to begin the GC analysis. During the chromatographic separation process the temperature of the trapping material is normally raised to a value slightly higher than the desorption value. This increase in temperature cleans the sorbent material of residual material and renders the trap ready for the next sample.

To optimize the analytical procedure for target analytes, most instrumentation is equipped with the capacity to systematically increment the salient PT parameters.

The parameters that may be manipulated in order to optimize the procedure include: sample purge time, carrier gas purge flow rate, and sample temperature. Although not under the control of the instrumentation, a number of other parameters have significant impacts on the PT experiment. Among these are the sample size and the choice of the trapping media. In general, increasing the amount of material to be sparged will increase the amount of analyte captured on the trap. There are some practical limits to this approach, most of which are dependent on the character of the sample matrix. For example, too much solid material could possibly retard the efficient flow of stripping gas. Likewise, increasing the amount of a liquid sample that possesses some foaming characteristics may result in contamination of the system with undesirable sample matrix components. The most popular adsorbent for the enrichment of trace volatiles and semivolatiles from different matrices is Tenax™ GC. Tenax™ GC is an organic polymeric material, poly(2,6-diphenyl-*p*-phenylene oxide). The retention of unsaturated and polar material on the Tenax™ GC is such that the material has found wide applications in environmental analyses, flavor analyses, biological fluids, and human volatiles. The thermal stability of Tenax™ GC at temperatures of approximately 350 °C coupled with its low affinity for water further substantiate it as a very acceptable and usually the preferred trapping medium for analysis of volatile and semivolatile organic compounds by PT techniques.

Thus, by way of comparison, dynamic HS techniques/gas chromatographic techniques afford powerful applications in the areas of the qualitative and quantitative analyses of trace levels of volatile and semivolatile materials. Conversely, static HS/GC techniques afford powerful applications in the areas of qualitative and quantitative analyses of minor volatile and semivolatile components.

2.7 Applications of Static and Dynamic Headspace Gas Chromatography

The chemical composition of essential oils has been shown to be related to a variety of factors including age, genotype, and geographical growing locations.[13,14] In an application of static and dynamic (i.e. PT) HS techniques, quantitative and qualitative differences in the distribution of components in an essential oil were documented (Table 1). Through the use of an information-rich detector, MS coupled to the GC, qualitative information on two sources of coriander oil was obtained. Quantitative percentage distribution values for the components of the two were gathered using a flame ionization detector (FID) coupled to a GC. Both the static and dynamic HS approaches were successful in detecting statistically significant differences in the two coriander oils. In addition, both approaches agreed on the major components of each oil, as well as the relative distribution of each component. The precision of the static HS approach was found to be slightly better than that of the dynamic approach. The use of the MS and FID detectors allowed for the determination of sample 1 as an aged sample due to the presence of increased amounts of the linalool oxide, known to increase in aged essential oil samples.

Table 1 Quantitative comparison by component percentage distribution of two coriander oils by static and dynamic HS analyses[13]

Component	Dynamic		Static	
	Sample 1	Sample 2	Sample 1	Sample 2
α-Pinene	4.52	0.43	5.64	0.81
β-Pinene	0.52	1.77	0.20	1.29
Myrcene	0.26	2.76	0.49	1.19
γ-Terpinene	0.00	3.96	0.00	2.14
cis-Linalool oxide	3.14	0.29	0.20	0.22
trans-Linalool oxide	2.77	0.22	3.54	0.35
Linalool	81.70	77.55	84.35	80.95
Camphor	3.91	4.56	4.31	6.22
Geranyl acetate	0.25	2.25	0.19	1.47
β-Caryophylene	0.53	0.59	0.96	1.23

2.8 Large-volume Injection/Gas Chromatography

The injection system of a gas chromatographic system is designed to introduce accurately and reproducibly a small representative amount of the sample being investigated into the column without degrading the maximum separation efficiency of the column.[15,16] The most common injection techniques are split, splitless, and cold on-column. Split injection is normally employed for minor component analysis where the solute concentration in the sample is present at concentrations greater than 50 ppm (50 ng μL^{-1}). The splitless injection technique, if properly implemented, is most frequently used for trace component analysis in the range from 0.1 ppm to 200 ppm. Cold on-column injection allows the sample to be deposited directly onto the front of the column and has an applicable concentration range similar to the splitless technique. The major advantages of the cold on-column technique are twofold. First, the sample is not subjected to the high temperatures required for split and splitless injections that could cause analyte decomposition or artifact formation. Second, as an intermediate flash vaporization step is not present, discrimination of higher-boiling components is minimized. A further

limitation of these three most common injection techniques is that the sample volume is limited to less than 2 μL. Vogt et al.[17] developed another technique, programmed temperature sample introduction (PTV) that overcomes the 2 μL constraint and allows a larger aliquot of the sample analytes to be deposited onto the column while simultaneously eliminating the solvent. This injection technique is now more commonly referred to as LVI.

Figure 5 is a diagram of a PTV inlet. The PTV inlet closely resembles the classical split/splitless inlet as the sample is deposited in a liner. The carrier gas enters the liner at the top and exits at the bottom, and a solenoid valve allows for switching between split and splitless operation. A functional difference between the PTV inlet and a split/splitless inlet lies in the temperature control. Conventional split/splitless inlets normally operate at elevated temperatures (>200 °C) where the sample is flash vaporized before entering the column. Although the PTV inlet can operate under these elevated isothermal temperatures, it is also capable of rapidly heating or cooling the liner. Heating rates in excess of $10\,^{\circ}C\,s^{-1}$ and subambient cooling are common. Operationally, the sample is injected onto the inlet liner, usually at a temperature below the boiling point of the solvent. The inlet is then rapidly heated to transfer the sample from the liner to the front of the column in a very narrow band, thus preserving the separating power of the column. The capability of introducing the sample into the inlet at lower temperatures affords numerous advantages and are similar to those described for the on-column inlet.

The most important feature of the PTV is the capability for ultratrace analysis (i.e. sample concentrations below 1 ppt) via the introduction of sample aliquots up to 1 mL. This is accomplished by operating the PTV in the solvent venting mode. The preferred implementation of this is referred to as SVSF. The chromatographic events are divided into five time zones and are described below.

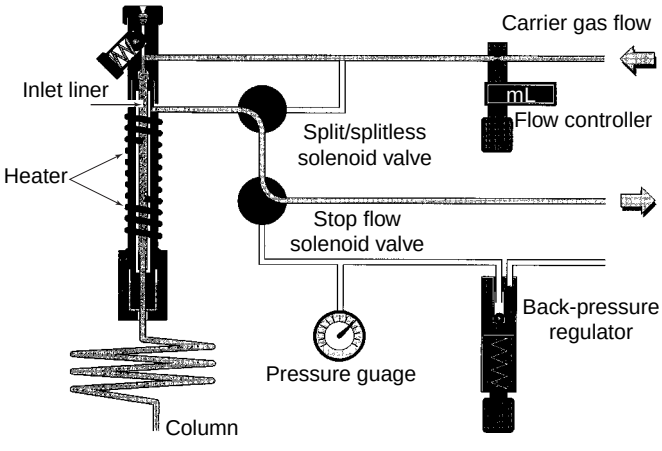

Figure 5 Diagram of a PTV inlet system. Flow paths are shown for the solvent venting–stop flow (SVSF) configuration. (Reproduced courtesy of Jerstel GmbH and Co.)

1. *Pre-injection* The split valve is open, GC and inlet temperatures are at initial values and the column head pressure is at normal operating pressure.
2. *Sample injection* Immediately before the sample injection the column head pressure is reduced to ambient (stop flow). The sample is injected onto the inlet liner at a slow constant rate defined by the solvent boiling point, solute vapor pressures, inlet temperature, and gas flow rates. Injection speeds of $50\,\mu L\,min^{-1}$ are not uncommon, resulting in 10 min to inject 500 μL. During this time the inlet conditions are such that the solvent vaporizes and is swept from the inlet liner. Solute molecules, possessing lower vapor pressures (higher boiling points), remain in the liner as a liquid or solid.
3. *Post-injection* The syringe needle is withdrawn from the inlet and the run-time clock starts (time zero). The inlet conditions of stop flow remain in effect, allowing further purging of the remaining solvent.
4. *Sample transfer* The column head pressure is restored to normal operating pressure and the split valve is in the splitless mode. The solutes are transferred onto the column by rapidly heating the inlet liner.
5. *Sample analysis* The GC temperature program starts and, after appropriate times, the inlet returns to the split mode and then starts its cool down to the initial temperature.

A prerequisite for proper quantitative analysis via LVI/GC requires the vapor pressures of the solutes being analyzed to be significantly higher than that of the solvent. Failure to meet these conditions will result in a nonreproducible loss of solute and poor precision. Optimizing the injection speed relative to temperature, flow rates, and vapor pressures cannot be underestimated if precise quantitation is required. Also, a constant slow injection at a definable rate requires the addition of a computer-controlled injector. These injectors are commercially available.

2.9 Applications of Large-volume Injection/Gas Chromatography

GC is routinely used for the analysis of many environmentally significant assays.[17,18] The determination of trace components in water usually includes sample preparation steps that enrich the analytes and remove the matrix. The sample enrichment is frequently a time-consuming and labor intensive operation. Elimination of the sample enrichment process would not only decrease analysis

costs and increase sample throughput (via automation), but would also increase the precision of the method. LVI successfully addresses these concerns by both directly injecting the water without preconcentration and by incorporating an in-vial extraction of the water into an organic solvent. Other categories, such as the analysis of beverages, have found significant assay improvements using LVI techniques.

2.10 Multidimensional Gas Chromatography

High-resolution GC incorporating capillary columns is capable of separating and quantifying hundreds of compounds in a single run and has been used for the analysis of complex mixtures such as natural products, biological fluids, and environmental samples since the early 1960s.[19–22] Depending upon the analytes of interest, even the optimum GC system cannot resolve all of the components in a simple mixture of less than 10 compounds. In addition, as GC detectors such as FID and MS have improved by decreasing detection limits, analysts have come to realize that almost every peak in a GC chromatogram of a complex mixture is most probably a combination of more than one component. MDGC was developed to address this situation. In its simplest form MDGC is defined as a GC system of two or more columns of different selectivity and a device that enables the selective transfer of a portion of a chromatographic run from one column to the second column (a process known as heartcutting). The incorporation of columns of different selectivity is imperative as columns of identical selectivity (stationary phase) would add little value toward increased resolution of analytes. In addition, the hardware used to effect the transfer should be inert with respect to the analytes of interest so as to maintain the maximum efficiency of the separating columns.

Numerous different hardware implementations of capillary-to-capillary heartcutting MDGC systems are commercially available. Most are based on pneumatic flow control first introduced by Deans[20] in 1968. A comparison of three different implementations of MDGC was presented by Gordon et al.[21] A diagram of one of these systems in a dual-oven configuration is shown in Figure 6. The operation of a typical MDGC system for heartcutting includes three basic operations:

1. An initial separation is performed on the first column (i.e. precolumn), without transfer of eluent to the second, or analytical, column.
2. The retention time windows (the run time immediately before to the time immediately after) for the analytes of interest are determined from retention times obtained from the precolumn separation. The sample is injected a second time onto the precolumn and, at the appropriate times, the solutes are heartcut to the analytical column.
3. Final separation occurs on the column under temperature program conditions.

Precise balancing of the pressures and volumetric flows of the MDGC system is paramount. Key components essential to the proper functioning of an MDGC system are a fixed restrictor and a midpoint restrictor. The fixed restrictor is normally a short length of narrow-bore (100–150 μm) uncoated deactivated fused silica column. The combined pressure drop of the fixed restrictor and the mid-point restrictor should be equal to that of the analytical column at normal operating conditions of

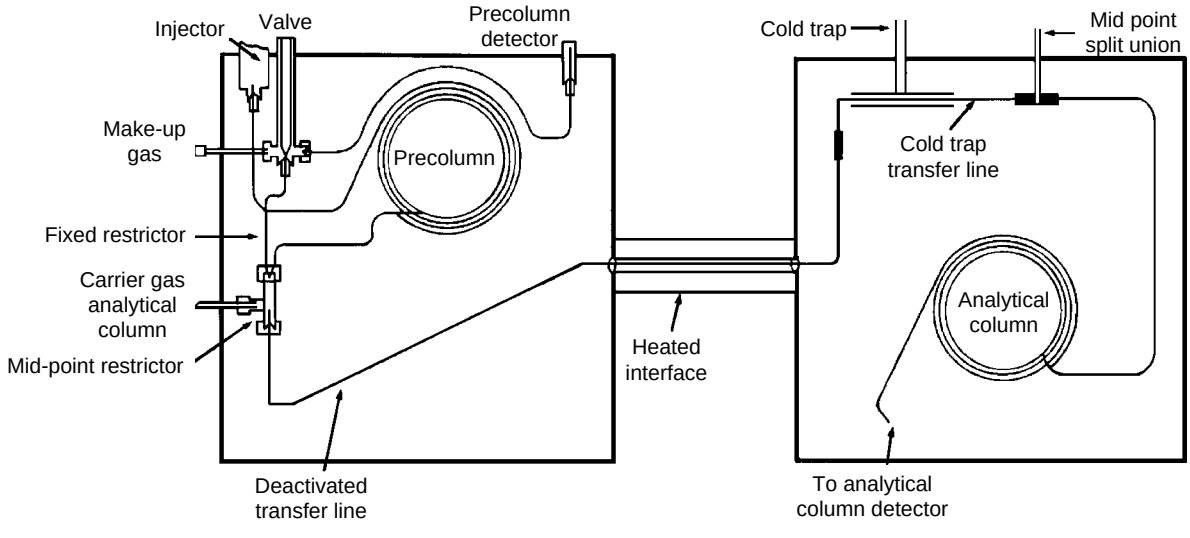

Figure 6 The flow paths of an MDGC system. (Reproduced courtesy of Marcel Dekker, Inc.)

temperature and flow. Failure to balance the pressures in this fashion will result in changes in the retention times on the precolumn during heartcutting, when compared to the initial separation performed without heartcutting, resulting in either incomplete transfer of the analytes of interest or transferring more of the precolumn eluent than desired. A cold trap situated in front of the analytical column serves to trap the heartcut material eluting from the precolumn as a relatively broad band into a sharp band, thus preserving the resolving power of the analytical column. Evaporating carbon dioxide or liquid nitrogen under pressure can be used as cryogenic materials.

As mentioned earlier, the selectivity of the two columns in the MDGC system should differ. For example, if a polar column such as polyethylene glycol is used as the pre-column, then a dimethylpolysiloxane-based analytical column is usually a good choice for the analytical column. Such differences in separation fundamentals multiply the effectiveness of the MDGC separation.

Computer control of the valve sequences necessary to perform an MDGC analysis is required for reproducibility. If the heartcut starts to early, other analytes could be transferred to the analytical column and potentially co-elute with the compounds of interest and, conversely, if the heartcut is initiated late, the compounds of interest may not be transferred to the analytical column. Precise control of numerous hardware operations is required to obtain reproducible results. These include:

- the timing of the heartcut valve;
- turning on and off the cold trap;
- the exact flow rate and opening and closing of the mid-point split valve;
- the start of the temperature programming of the analytical column oven.

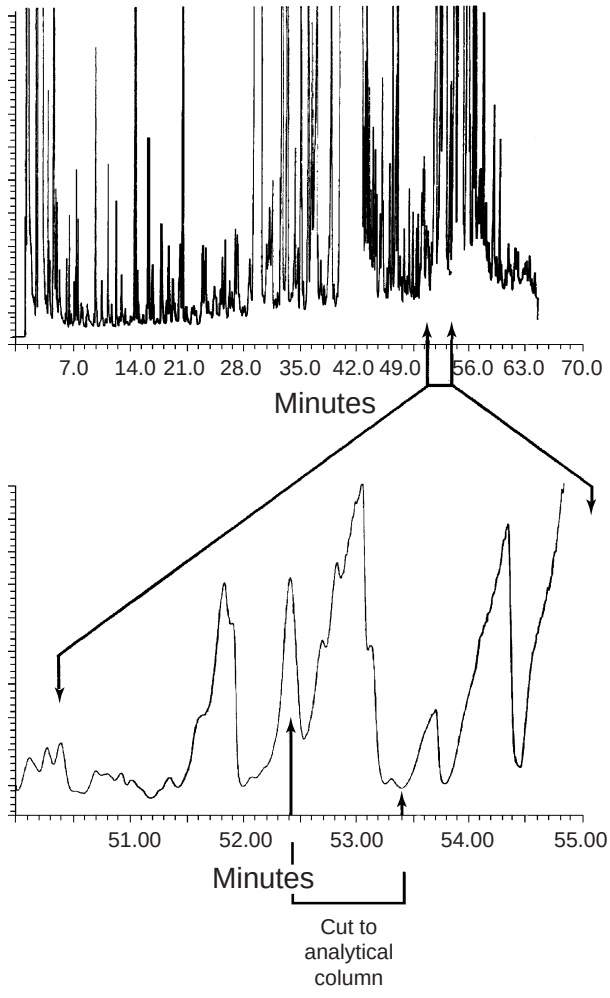

Figure 7 Chromatogram of a tobacco smoke sample: (above) full precolumn (single-column) chromatogram, (below) an expanded view of the section indicated on the top chromatogram. (Reproduced courtesy of Preston Publications.)

2.11 Applications of Multidimensional Gas Chromatography

Most successful MDGC applications address many issues associated with both sample preparation and analysis.[23–25] Among the issues are:

- target compound analysis;
- use of precolumn to essentially replace classical fractionation of the sample;
- decreases in sample loss and artifact formation due to classical sample fractionation;
- total sample analysis via sequential heartcutting.

The application of MDGC techniques to target compound analyses in complex mixtures employs and addresses many of the issues listed above. For example, MDGC techniques were applied to the analysis of a target compound in rice and to tobacco smoke. The compound 2-acetyl-1-pyrroline has been determined in a crude complex rice extract. Sample preparation involved preparing a methylene chloride extract of the rice followed by injection of the crude extract into an MDGC system. Thus, no classical fractionation was required, eliminating any loss or artifact formation. Furthermore, total extract analysis by sequential heartcutting was available but not necessary for this application. Tobacco and tobacco smoke, like most complex mixtures, are known to contain hundreds of semivolatile components. A conventional single-column capillary gas chromatogram of a tobacco smoke sample is shown in the top of Figure 7. The bottom of this figure shows an expanded section of the full chromatogram. The section from approximately 52.5 min to 53.5 min appears to contain 2–4 peaks and mass spectral scans were able to identify an ethyl-substituted pyrazine as well as numerous unidentifiable

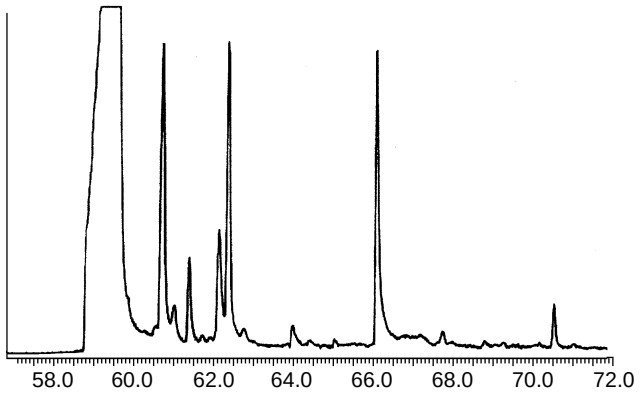

Figure 8 Analytical column chromatogram generated from the heartcut indicated on Figure 7. (Reproduced courtesy of Preston Publications.)

components. This same region of the separation was subsequently heartcut to a second capillary column resulting in the chromatogram shown in Figure 8. Mass spectral scans were able to confirm the presence of nine identifiable components.

3 POST-GAS CHROMATOGRAPH SAMPLE ANALYSIS TECHNIQUES

The underlying principle governing sample detection and analysis by GC is governed by well-defined consistent responses of a particular sensor (i.e. a detector) to the presence of an analyte within the detector's environment.[9–26] In GC, the presence of the analyte within the detector's environment is accomplished by the appropriate interfacing of the end of the GC column with the sensor of choice. Each detector will present its own set of requirements for successful interfacing to a GC.

The output of the detector is called a response, and there is a wide array of relationships between the characteristics of a detector and the characteristics of analytes. Generally, the amount of analyte within the zone of the detector governs the amount of the detector response. However, the chemistry of the analyte including, for example, its structure and stability, dramatically affects the ability of the detector to register a response.

In order of importance in a chromatographic system, the detector ranks a very close second behind the GC column. Scott[26] has listed seven major detector specifications that are important: detector linearity, linear dynamic range, detector noise level, detector sensitivity, pressure sensitivity, flow sensitivity, and temperature sensitivity.

Detectors are not universal linear devices. That is, detectors provide a linear response to changes in concentration over a definable range. This linear response usually takes the form of a linear plot of concentration on the x axis versus detector output on the y axis. In some cases the detectors for gas chromatographs can provide a linear response over a concentration range of more than two orders of magnitude. Once this linear range has been defined for the analyte of interest, then the linear dynamic range of the detector has been established for that analyte.

Detectors have a net response, that is, random changes in signal levels, even in the absence of an analyte. This random response is termed detector noise. Minimization of detector noise is critical to the performance of any system. When detector noise is minimized, detector sensitivity is improved. Detector sensitivity is defined as the minimum concentration of analyte that can be discerned from the random noise. Thus, if the noise is consistently low, sensitivity can be maximized. The degree of detector sensitivity to changes or fluctuations in pressure, flow, and temperature vary greatly depending on the detector. For mass spectrometers all three of these variables have a significant impact on detector performance. For FID, provided that the temperature is high enough to prevent analyte condensation, temperature changes are not as significant to detector performance.

A diverse array of detectors are now currently being used in GC applications. In a number of cases these detectors can be described as information rich. That is, not only do these detectors provide evidence for the presence of an analyte, but they also furnish structural information concerning the analyte. The two most popular detection systems fulfilling the requirements as information rich are MS and IR. The following sections will discuss the theory and instrumentation surrounding these two systems.

3.1 Gas Chromatography/Mass Spectrometry

Many of the basic principles governing the performance of a GC/MS system are those associated with gas flow and vacuum parameters.[27–33] The GC must operate over a wide range of gas flows and pressures whereas the MS must operate under reduced pressures. Calculations surrounding the optimization of these parameters associated with each component of the GC/MS system can be made; however, approximations provide for adequate values and detailed calculations are not required. Assuming that the mean free path of ions (i.e. the molecular flow) within the MS is sufficient and that the vacuum pumping speed is adequate to handle the incoming gas flow from the GC, then the successful wedding of GC with MS will be possible. There is a very practical reason for maintaining molecular flow. When molecular flow is established, molecule–molecule and ion–molecule reactions and interactions are virtually eliminated. A vacuum

in the range of 1×10^{-5} torr is normally required for most MS experiments. These types of reactions are undesirable in the GC/MS experiment because they have the potential for the production of additional molecules and ions that will complicate the molecular fragmentation patterns of the resulting data. Such complications will severely hinder the ability to interpret the resulting mass spectra.

Although not basic to the performance of a GC/MS system, the computer is a necessary component with which to realize the full potential of the GC/MS system. Thus, when speaking of current commercially available GC/MS systems, it is assumed that the actual system consists of a GC, an MS, and a computer (Figure 9). In most current configurations, the computer controls both the GC and the MS. Furthermore, the data collected from a GC/MS experiment is stored and can be evaluated employing software programs commonly available on most commercial systems.

The GC/MS interface provides an acceptable means by which the GC carrier gas with accompanying analytes eluting from the end of the GC column may enter the MS ion chamber under vacuum. When capillary columns are employed, the flow rate through the column is most often less than $3\,\text{mL}\,\text{min}^{-1}$. Thus, with current vacuum technologies all of the GC capillary column flow can enter the MS ion chamber with no sacrifice in the vacuum requirements. This capability is very advantageous in that all of the analyte present in the GC column can find its way into the MS ion chamber, thereby affording the potential for maximum sensitivity.

Two types of GC/MS interfaces, both with advantages and disadvantages, dominate those currently employed: direct and open split. With the direct interface, all of the analyte exiting the GC column enters the MS, thereby improving/maximizing sensitivity. From a practical hardware aspect, however, use of the direct interface requires that the GC/MS be vented should changes in columns be required. Venting of the MS requires exposure of the MS ion chamber to the room atmosphere. To prevent significant damage to the components in the ion chamber, which are very sensitive to air and water at elevated temperatures, this venting must be controlled and done only at room temperature. Failure to follow these procedures can drastically reduce the lifetime of many of the MS components to unacceptable values.

When open-split interfaces are used, changing a GC column does not require venting of the MS. In the open split configuration ample make-up carrier gas is supplied to the MS, thereby excluding the entry of air and water. Conversely, with the open-split interface, not all of the analyte exiting the GC column enters the MS ion chamber. This results in reduced sensitivity. Depending on the application this reduction in sensitivity may not present an issue for the desired analyte. Thus, use of a direct or open-split interface for GC/MS experiments should be evaluated on a case-by-case basis.

One of the fundamentally critical aspects of optimum performance of a GC/MS system rests with the proper installation of the GC column in the MS. Having properly conditioned the GC capillary column and established the desired carrier gas linear velocity, insertion of the end of the column directly into the MS ion chamber is possible. Very close attention must be paid to the proper insertion depth into the MS chamber. Manufacturers of GC/MS systems are very clear on the necessity of performing this step accurately. Improper installation of the GC column can lead to a complete absence of instrument performance.

To establish the necessary vacuum within the MS ion chamber receiving the carrier gas flow from the GC will often require the presence of two vacuum pumps.

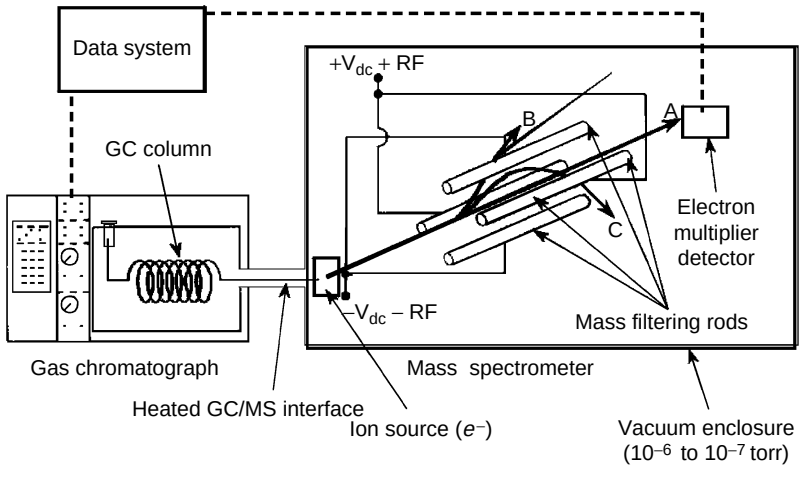

Figure 9 Diagram of a GC/MS system.

A roughing pump and an ion chamber pump (usually a turbo or diffusion pump) comprise the dual pumps. Continuous monitoring the vacuum in a GC/MS system is important in assessing proper functions. With a properly installed capillary GC column having the dimensions 30 m long, 0.25 mm i.d., 0.25 μm film thickness, and a linear velocity of approximately 30 cm s^{-1}, the pressure in the MS ion chamber is approximately 5×10^{-5} mm Hg. This vacuum reading is adequate for proper functioning of most commercial systems. When installing a GC column in the GC/MS system, 12–24 h are normally required to achieve equilibrated thermal and pressure parameters.

After having established stable operating temperatures and pressures, mass spectrometers must be properly calibrated. Conveniently, GC/MS–computer manufacturers provide user-friendly software controlled procedures for calibrating the MS. Many quadrupole mass spectrometers are calibrated using perfluorotributylamine (PFTBA). Under computer control the MS is autotuned to target parameters associated with the mass spectrum of PFTBA. Manual tuning of the MS is possible, but is very time-consuming. The computer output from consecutive autotunes can be used to diagnose the performance and status of the GC/MS system. For example, systematic increases in the amount of multiplier voltage can indicate that the ion filaments are deteriorating. The presence of ions at $m/z = 18$ (water), 28 (nitrogen), 32 (oxygen), and 40 (argon) strongly indicates a leak in the system. The strong indication of a leak requires immediate attention or the GC/MS system performance will dramatically and rapidly deteriorate.

In the normal course of events, venting or shutting down of the GC/MS system will become necessary. Proper attention to manufacturer recommendations for shutting down the MS will prevent abnormal deterioration of the system. Depending on the type of ion-chamber pump employed, decreases of temperature or spinning speed will be required as part of the venting process. When and only when these requirements have been met is it safe to open the MS to the atmosphere and conduct the necessary repairs or changes. Should repairs extend for a period of time, it is important to protect the MS from exposure to dust and moisture. Once repairs or changes have been made then the MS may be pumped down and operations resume.

From a data acquisition perspective the GC/MS–computer system set-up affords many diverse options. In particular, the type of mass-spectral data gathered can be controlled through instrumentation settings. For example, the initial set-up for the analysis of an unknown mixture would most probably involve collection of the mass spectral data in the scan mode. Data collection in this manner provides a total ion chromatogram consisting of a plot of total ion response as a function of time. Data gathered in this way has qualitative implications. The response can be related to the number of ions present, this in turn having quantitative implications. In a second mode of data collection, data from only selected ions may be collected, yielding a selected ion chromatogram (SIM) consisting of a plot of the selected ion response as a function of time. This approach also has quantitative aspects.

From a qualitative perspective, a GC/MS–computer system represents a very powerful physical probe into the structure of analytes. Using the normal, essentially standard, instrumental system configuration settings for gathering total ion chromatograms in the scan mode affords the potential for searching mass spectral databases having more than 100 000 mass spectra gathered under the same standard conditions. Thus, should the GC separation provide for good resolution of the components of interest, automated searching of mass spectral databases (libraries) can furnish valuable data concerning the structure of the analyte. The mass spectra are generated through the fragmentation of the analyte under a well-defined set of conditions yielding ions. These ions usually have a charge of −1, thus the obtained plot of ion abundance versus mass to charge ratio, m/z, can be viewed as the mass spectrum of the component. The search can readily be accomplished by automatically submitting, via instrumental software programs, the mass spectrum of the unknown analyte for comparison against the mass spectra contained in the database. Most automated search routines will provide a match quality ranging from 0 to 100, with 100 being a perfect mass-spectral match. The instrumentation can also provide a graphical representation of the mass spectrum of the unknown analyte alongside that of the library spectrum. Very close examination of the output from the library search routines is mandatory to ensure accurate structural assignments. At this point, combination of the separation parameters from the GC with the results from the MS represents a very powerful analytical combination. That is, a combined assessment of both the analyte retention time and the analyte mass spectrum gives more confidence in the structural assignments. In addition, as the amount of ions present can be related to the amount of analyte present, quantitative analysis by total ion current (TIC) can also be realized.

Should quantitative analysis of a known component be the main thrust of the GC/MS experiment, then the instrumentation can be configured in the SIM mode. In this mode, only ions associated with the analyte of interest will be collected by the MS–computer system. This approach provides for increases in specificity and sensitivity. The specificity is improved because only ions associated with the analyte of interest are collected. For

the same reason, the sensitivity is increased because only a few ions, i.e. two or three, are being collected versus approximately 300–500 being collected in the TIC mode. Both internal standard and external standard calibration methods are possible with GC/MS instrumentation. Thus, GC/MS–computer instrumentation set to collect data in the TIC mode has powerful qualitative implications, whereas instrumental settings in the SIM mode have powerful quantitative implications.

All of the issues associated with applications associated with GC/MS are also valid when analyses are pursued with GC/high-resolution MS. Low-resolution instruments have been defined arbitrarily as those instruments that separate unit masses up to $m/z = 2000$. An instrument that can resolve two ions differing in mass by at least one part in $10–15 \times 10^3$ is generally considered a high-resolution mass spectrometer. For example, a mass spectrometer that can resolve an ion mass of 600.00 from one of mass 599.95 would have a resolution of 10 000. With this mass resolution capability the data from high-resolution mass spectrometers can be used to calculate the ion atomic composition. The possibility for high mass resolution is realized by the introduction of an electrostatic field after or before a magnetic field along the path of the ions. Ions are generated in the source and they are accelerated toward the analyzer. As the ions enter the magnetic field they are directionally focused. Upon entering the electric sector, which is perpendicular to the flow of the ions, the ions are again curved. This application of double focusing is the fundamental theory supporting the resolution capability of high-resolution MS.

3.2 Applications of Gas Chromatography/Mass Spectrometry

GC/MS currently finds a very wide array of applications including analyses of biological tissue, petroleum residues, pesticides, essential oils, and pharmaceuticals.[34,35] In one specific case, analysis for and speciation of street drugs has been accomplished using a GC/MS approach. Following derivatization with pentafluoropropionic anhydride (PFPA), GC separation with MS detection of isomers of the street drug MBDB was accomplished in 9 min. Interpretation of the MS spectral data assisted in determining the structures of the isomer. In some cases GC/high-resolution MS is employed to provide accurate mass confirmation of the presence of drugs of abuse.

Polychlorinated biphenyls (PCBs) are widespread toxic organopollutants. A method for qualitative and quantitative analysis using GC/MS for PCBs in soil has been developed. Advantage of the properties of MS is taken through the use of a ^{13}C-substituted PCB as an internal standard. The presence of the ^{13}C label in the PCB did not significantly affect the retention time of the PCB. Furthermore, the presence of the ^{13}C atoms in the PCB produced a unique mass spectrum having ions attributable to the heavier ^{13}C. These unique ions allowed for quantitation of the PCB levels in selected soil samples.

3.3 Gas Chromatography/Fourier Transform Infrared Spectroscopy

Two distinct analytical approaches that can be classified as gas chromatography/Fourier transform infrared spectroscopy (GC/FTIR) have dominated the recent literature. The techniques and instrumentation requirements for each approach are so significantly distinct as to require separate descriptions of the fundamental theory and hardware governing the successful application of each approach. In this light, the two GC/FTIR approaches are considered in independent sections.

3.4 Gas Chromatography/Fourier Transform Vapor Phase Infrared Spectroscopy

The fundamental theory underpinning the successful functioning of gas chromatography/Fourier transform vapor phase infrared spectroscopy (GC/FTVPIR) systems relies in large part on the ability to rapidly acquire multiple entire IR spectra in a very short time, i.e. four spectra per second.[36–42] This is possible because modern spectrometers employ a rapid scanning Michelson interferometer for measuring the IR spectrum. The rapid acquisition of spectral data is also a fundamental requirement because very low levels of components, nanograms, will pass through the IR beam in a few seconds with properly functioning capillary GC.

Following collection of the spectral data, Fourier transform mathematics converts the time-domain data, an interferogram, into frequency-domain data, an IR spectrum. This calculation is accomplished using a computer, due to the significant amount of data required to adequately represent the IR spectrum. Gram–Schmidt mathematics can also be performed on the spectral data to yield a total response chromatogram (TRC). A TRC is a plot of IR response versus time that is very similar in presentation to the plots obtained from MS response versus time, a TIC.

A simplified diagram of a GC/FTIR system can be found in Figure 10. The light from the IR source passes through the interferometer and then through a flow cell, i.e. a light pipe. Any light that passes through the flow cell is focused on a detector where the light is converted into an electrical signal. The electronic signal is ultimately converted, via computer-driven mathematics, into an IR spectrum. One end of the light pipe is inserted into the end of a capillary GC column. Thus, three operations are

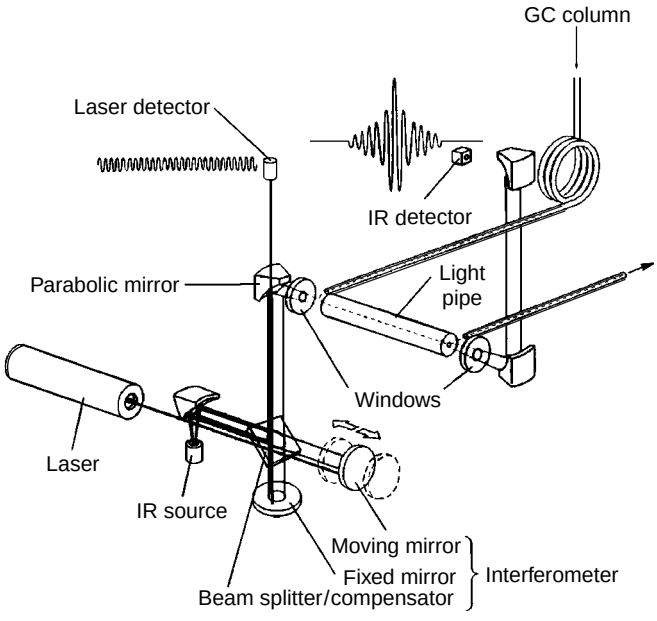

Figure 10 Diagram of a GC/FTIR system. (Reproduced courtesy of Bio-Rad Company, Inc.)

occurring, simultaneously: IR light is continually passing through the light pipe, effluents from the GC column are entering and exiting the light pipe, and the detector is collecting the IR spectral data. At selected times only GC carrier gas is present in the light pipe, essentially when no analyte is exiting the GC column. Under these conditions an IR spectrum is taken which serves as a blank spectrum. A blank spectrum is essential for the mathematical calculations involved in the conversion of the interferogram to the IR spectrum.

From a hardware perspective, GC/FTVPIR has some unique requirements. A constant inert and dry purge gas, such as nitrogen, must continuously purge the optical system. Among the reasons for this obligation are:

- Exclusion of air, particularly carbon dioxide, from the optical path is essential due to its intense IR adsorption around 2250 cm^{-1}.
- Exclusion of water is essential because of its intense IR adsorption around 3500 and 1600 cm^{-1}, as well as water's ability to degrade the IR transmission properties of the KBr windows attached to the ends of the light pipe. Water will react with the KBr windows, resulting in the windows acquiring a frosted look, which severely hinders the transmission of IR light.
- Dry air will result in premature oxidation of the IR source.

The temperature of the optical path must be kept constant, $\geq$200 °C. Variations in temperature cause changes in sensitivity and changes in flow rates, not to mention subtle variations in the characteristics of the IR spectrum itself.

The light pipe with its KBr windows is the heart of the flow cell assembly. The inside of the glass light pipe is coated with gold to prevent the loss of IR light and provide as inert a surface as possible. The internal volume of the light pipe is approximately 100 µL with an i.d. of approximately 1 mm. Material eluting from the GC column enters one end of the pipe, passes through and exits the other end. As the material passes through the pipe, portions of the IR light are adsorbed, this phenomenon being captured by the detector.

Due to the relatively small diameter of the light pipe (1 mm), extreme care must be exercised in the proper alignment of the pipe itself, the parabolic mirrors of the IR light path, and the detector. All of these parameters can be optimized for maximum IR light throughput by monitoring the output of the detector while these objects are slightly moved. Maximum voltage output from the detector translates into more IR light throughput. Sequential adjustments of the objects will yield a maximum voltage indicating maximum IR light throughput. Once alignment is achieved, little to no further adjustments should be necessary.

Installation of the end of the GC column into the light pipe at the ideal position is essential for optimum performance of the GC/FTVPIR system. When the system is up and operational, IR light is continually passing through the pipe and causing the detector to output a certain voltage. Slow insertion of the end of the GC column into the light pipe from a side port can be made until there comes a point where the end of the GC column breaks the plane of the IR light beam, causing a change (drop) in the output voltage of the detector. The optimum position is, thus, the point of closest approach of the GC column end to the IR light beam without any interference. This location allows the GC column effluent optimum interaction with the IR light beam, thereby providing maximum sensitivity.

To obtain the necessary sensitivity from the IR detector, the mercury cadmium telluride IR detector must be cooled with liquid nitrogen. Most commercial systems provide IR detectors having Dewars capable of housing enough liquid nitrogen for 4 h. Thus, periodic checking of the detector liquid-nitrogen level is essential.

3.5 Gas Chromatography/Fourier Transform Matrix Isolation Infrared Spectroscopy

The successful operation of a gas chromatography/Fourier transform matrix isolation infrared spectroscopy (GC/FTMIIR) set-up depends on complete isolation of the effluent of a GC column at temperatures approaching

10 K in an inert noble gas matrix.[39,42–50] Complete isolation of the molecule in the matrix virtually eliminates all intermolecular interactions. In addition, molecular rotation is severely reduced at the very low temperature. Once captured at 10 K, usually in argon, maintenance of the temperature is fundamental to the success of the approach because the Fourier transform infrared spectroscopy (FTIR) spectra are gathered post-deposition. The availability of argon in which the analytes are captured is made possible by employing helium GC carrier gas containing approximately 5% argon. The presence of this small amount of argon in the GC carrier gas does not have any significant impact on the performance of the chromatography. At 10 K helium does not freeze and is hence pumped away via a diffusion pump–roughing pump configuration.

Figure 11 represents a simplified schematic of a GC/FTMIIR system. A description of the physical characteristics of the instrumentation is essential to a full understanding of the operation of the technology. The vacuum chamber denoted by the rectangle is fitted on top with a Pyrex™ window/O-ring for viewing the collection disk. The vacuum is provided by a diffusion pump and a roughing pump configuration. IR light from a Fourier transform spectrometer passes through specially designed windows as it reflects off a pair of dual parabolic mirrors, the collection disk, and a focusing mirror onto the surface of a mercury cadmium telluride detector. The detector must be cooled with liquid nitrogen. Positioning of the mirrors is very critical to the successful functioning of the GC/FTMIIR system. The optimum position of the multiple mirrors is accomplished in a way similar to the VP system by maximizing the voltage output of the detector under normal operating conditions. The optimization must take place under normal operating conditions because temperatures in the area of 10 K cause subtle changes in the position of the collection disk.

The collection disk is constructed of oxygen-free high-conductivity copper plated with hardened gold. The curved surface of the collection disk is highly polished to yield a uniform mirror finish. The surface of the collection disk serves as a platform for the collection of the effluent from the GC column from the GC transfer line. The effluent is the carrier gas, usually helium, containing about 5% argon as well as any analytes that may be present. When the effluent strikes the surface of the collection disk at 10 K the analytes and argon freeze on the disk and the helium is pumped away. Simultaneously, while the effluent is exiting from the transfer line, the collection disk is rotated at a predetermined rate. Thus, as the disk rotates the resultant physical phenomenon is a thin layer of argon on the surface of the collection disk. Close examination of this layer of argon reveals the presence of what appears to be a very fine transparent thread on the surface of the collection disk. Imbedded within the thread of argon are the analytes that have been separated by the GC column.

The position and characteristics of the GC transfer line relative to the surface of the collection disk are very important for the optimum deposition of the argon. The capillary GC transfer line protruding slightly from the heated nozzle (Figure 12) is made of deactivated fused silica of approximately 150 μm in diameter. The end of the GC transfer is positioned within approximately 150 μm of the surface of the collection disk. Incorrect positioning of the tip of the GC transfer line will result in an uneven/broken appearing thread of argon on the surface of the collection disk. More importantly, if the

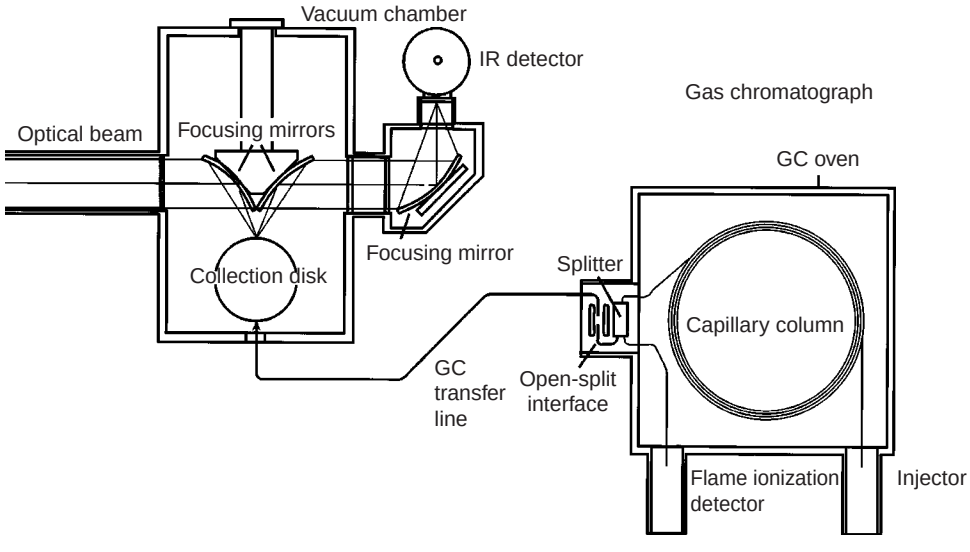

Figure 11 Diagram of a GC/FTMIIR system. (Reproduced courtesy of Marcel Dekker, Inc.)

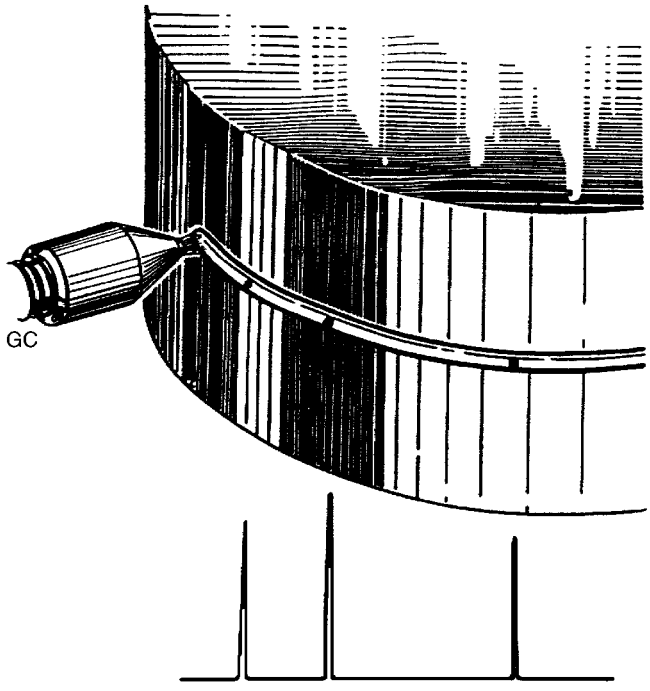

Figure 12 Drawing of the collection disk used in the GC/FTMIIR experiment showing the end of the GC transfer line protruding from the heated nozzle. Also illustrated is the argon thread. The dark marks in the thread indicated the position of solutes and correspond to the chromatogram at the bottom of the figure. (Reproduced courtesy of the American Chemical Society.)

tip of the GC transfer line is too close to the collection disk, contact between the two could result, scratching the surface of the collection disk. A scratched surface renders the collection of IR data impossible.

As the positioning of the collection disk is under computer control, the position of the disk relative to the IR light beam is always known. Thus, after a GC separation has been completed, exact positioning of the collection disk relative to the incoming IR light beam is possible. Analysis of the argon thread by FTIR takes place in the following way. The initial portion of the thread is scanned to establish a background. Then the disk is rotated at a predetermined rate while FTIR interferograms are rapidly taken and stored on the computer. After the entire run has passed under the IR beam, the data are converted into a plot of IR adsorption versus time. This plot is very similar to the Gram–Schmidt plot obtained from the Fourier transform vapor phase infrared spectroscopy (FTVPIR) experiment. The main difference is that in the FTVPIR experiment the data are collected in real time, whereas in the Fourier transform matrix isolation infrared spectroscopy (FTMIIR) experiment the data are collected post-run. To provide some assistance in determining the exact location of the argon matrix-trapped analyte on the collection disk, the instrumentation is fitted with an open split interface that allows a portion of the GC column effluent to be directed to a FID. The output from the FID is captured by the GC software, producing a trace of FID response versus time. The FID is a universal type of detector responding to most volatile and semivolatile organic compounds. Thus, the potential for correlating a FID response with an FTIR response is available with the instrumentation. The presence of retention times from the FID data also facilitates positioning of the collection disk directly under the IR beam such that the IR beam is focusing directly on the analyte of interest. For example, the response of hydrocarbons from the FID detector is normally relatively intense, at least with respect to the IR response that would be expected from a very low concentration of a hydrocarbon. Consequently, the FID would assist in locating components which may not possess strong IR-adsorbing chromophores.

Having captured the analytes in an inert noble gas matrix, coupled with the ability to hold them there for an extended time (hours), offers a distinct advantage to the matrix isolation (MI) experiment over the VP experiment. The VP experiment is a flow-through approach in which the IR data is gathered on the fly. In marked contrast, in the MI experiment the captured analytes can be examined over extended time. This capability translates directly into a tremendous improvement in sensitivity. Imagine possessing the capability of gathering IR spectra for as long as one wants. That is, the capacity for co-adding multiple numbers of scans of the analyte together with multiple scans of background spectra and subtracting the two. In practical terms this capability has afforded the capacity to collect FTMIIR spectra having excellent signal-to-noise ratios on as little as picogram quantities of material in a few minutes. Also, additional increases in sensitivity are realized because the lower temperature of the MI experiment versus the VP approach reduces the decomposition of thermally unstable molecules.

The lack of intermolecular interaction coupled with the virtual absence of molecular rotation, leads to IR absorption spectra from the MI experiment having significant and meaningful differences from IR spectra taken in other matrices. For example, narrow lines, frequently below 0.5 cm^{-1} for relatively small molecules, can be routinely obtained under MI conditions. Spectra such as these have been very useful in studying reactive molecules such as free radicals and reaction intermediates. Thus, the characteristics of the IR spectra obtained from an MI environment at extremely low temperatures would be expected to be very different from those obtained at elevated temperatures, such as VP experiments or even at room temperature. In the MI experiments, the ratio

of inert gas molecules to analyte molecules dictates the characteristics of the IR spectrum. This behavior is not unlike that observed as a function of concentration in solution room-temperature IR spectra. Acceptable ratios of about 5000 argon atoms to one analyte atom have produced excellent FTMIIR spectra using the GC/FTMIIR instrumentation.

Through a systematic investigation of a series of organic compounds possessing different functional groups, a relationship between the energy of selected band maxima as a function of matrix has been established. That is, the matrix in which the IR spectrum is taken significantly influences the energy and position of band maxima. For example, in the case of methyldecanoate carbonyl, absorption bands were observed at 1754, 1738, and 1769 cm^{-1} for the compound in an argon matrix, in the solid state, and in the VP, respectively. In general, the IR absorption maxima for compounds captured in an argon matrix were found to be intermediate between high values established by the VP experiments and low values established by IR spectra taken in the solid state. Likewise, statistically significant differences were also noted in IR absorption band maxima for organic compounds captured in a xenon or argon matrix. For example, the amide I band for diethylacetamide was found at 1661 cm^{-1} in a xenon matrix and at 1673 cm^{-1} in an argon matrix.

The virtual absence of molecular interaction and rotation in MI species afforded the capability of detecting a substantial difference in the IR spectra of *cis* and *trans* isomers in selected compounds. The elevated temperature of the VP experiment essentially eliminated any possibility of detecting this difference due to the relatively large widths of the absorption bands in the VP experiments. For example, the FTMIIR spectra of the *cis* and *trans* isomers of pinane were readily distinguishable one from the other.

3.6 Applications of Gas Chromatography/Fourier Transform Matrix Isolation Infrared Spectroscopy

Essential oils represent one of the more complex natural products having significant economic implications.[42] Determination of the structure of the components in these oils is paramount to effective assessment of their potential. For example (−)menthol, the major component of peppermint oil, is responsible for the major cooling effect known for peppermint. The other isomers of menthol, neomenthol, isomenthol, and neoisomenthol are structurally very similar to menthol but do not possess the cooling effect. GC/FTMIIR has been employed successfully to characterize the IR spectra of these compounds affording the capability to detect their presence in essential oils.

4 CONCLUSIONS

Within approximately two decades of the first coining of the term "hy-phen-ated",[1] hyphenated capillary GC techniques have become the methods of choice for the qualitative and quantitative analysis of complex mixtures.[51] Development efforts in modern GC-based instrumentation, depending on the integration of theories of GC, computers, and powerful information-rich techniques such as MS and FTIR detectors, have produced a wide array of very potent systems. Orchestration of combinations, such as on-line sample introduction, SPME, with GC, and MS, SPME/GC/MSD, has led to hyphenated systems with applications in a variety of areas. The enormous data sets generated from the information-rich detectors have been handled relatively easily with the advent of fast personal computers.

Future analytical hyphenated systems will most probably encompass the use of high-speed GC with accompanying high-speed mass spectrometers, infrared detectors and other specific detectors. Advances in hardware associated with on-line sample introduction approaches, such as SPME, LVI, MDGC, and HS techniques (dynamic and static) will assist in lowering further both the qualitative and quantitative limits of detection. Miniaturization of components should lead to additional systems being very compact and field deployable, with data transmission links directly to laboratory-based data analysis platforms. These advances should facilitate the collection of data followed by the rapid conversion of that data into information, and this new information into knowledge.

ABBREVIATIONS AND ACRONYMS

CE	Capillary Electrophoresis
CWDVB	Carbowaxdivinylbenzene
FID	Flame Ionization Detector
FTIR	Fourier Transform Infrared Spectroscopy
FTMIIR	Fourier Transform Matrix Isolation Infrared Spectroscopy
FTVPIR	Fourier Transform Vapor Phase Infrared Spectroscopy
GC	Gas Chromatography
GC/FTIR	Gas Chromatography/Fourier Transform Infrared Spectroscopy
GC/FTMIIR	Gas Chromatography/Fourier Transform Matrix Isolation Infrared Spectroscopy

GC/FTVPIR	Gas Chromatography/Fourier Transform Vapor Phase Infrared Spectroscopy
HPLC	High-performance Liquid Chromatography
HS	Headspace
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
i.d.	Internal Diameter
IR	Infrared Spectrometry
LC	Liquid Chromatography
LVI	Large-volume Injection
MDGC	Multidimensional Gas Chromatography
MI	Matrix Isolation
MS	Mass Spectrometry
MSD	Mass Selective Detector
PCB	Polychlorinated Biphenyl
PDMS	Poly(dimethylsiloxane)
PFPA	Pentafluoropropionic Anhydride
PFTBA	Perfluorotributylamine
ppb	Parts Per Billion
ppt	Parts Per Trillion
PT	Purge and Trap
PTV	Programmed Temperature Sample Introduction
RSD	Relative Standard Deviation
SFC	Supercritical Fluid Chromatography
SIM	Selected Ion Chromatogram
SPME	Solid-phase Microextraction
SVSF	Solvent Venting–Stop Flow
TIC	Total Ion Current
TRC	Total Response Chromatogram
VP	Vapor Phase

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Sample Preparation for Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment • Formaldehyde, Environmental Analysis of • Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Atomic Emission Detection in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Food **(Volume 5)**
Sample Preparation, Headspace Techniques

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Pulp and Paper **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. T. Hirschfeld, 'The Hy-phen-ated Methods', *Anal Chem.*, **52**(2), 297A–312A (1980).
2. J. Pawliszyn, *Solid Phase Microextraction Theory and Practice*, Wiley-VCH, New York, 1997.
3. W.M. Coleman, III, 'A Study of the Behavior of Polar and Nonpolar Solid-phase Microextraction Fibers for the Extraction of Maillard Reaction Products', *J. Chromatogr. Sci.*, **35**, 245–258 (1997).
4. W.M. Coleman, III, 'A Study of the Behavior of Maillard Reaction Products Analyzed by Solid-phase Microextraction–Gas Chromatography–Mass Selective Detection', *J. Chromatogr. Sci.*, **34**, 213–218 (1996).
5. W.M. Coleman, III, B.M. Lawrence, 'A Comparison of Selected Analytical Approaches to the Analysis of an Essential Oil', *Flav. Frag. J.*, **12**, 1–8 (1997).
6. J. Ai, 'Solid Phase Microextraction for Quantitative Analysis in Nonequilibrium Situations', *Anal. Chem.*, **69**, 1230–1236 (1997).
7. W.M. Coleman, III, T.A. Perfetti, B.M. Lawrence, 'Automatic Injection Solid-phase Microextraction–Mass Selective Detection Analyses of Essential Oils', *J. Chromatogr. Sci.*, **36**, 575–578 (1998).
8. A.J. Nunez, L.F. Gonzalez, J. Janak, 'Pre-concentration of Headspace Volatiles for Trace Organic Analysis by Gas Chromatography', *J. Chromatogr.*, **300**, 127–162 (1984).
9. J. Novak, P.A. Leclercq, *Quantitative Analysis by Gas Chromatography, Second Edition, Revised and Expanded*, Marcel Dekker, Inc., New York, 1988.
10. R.G. Westendorf, 'A Quantitative Method for Dynamic Headspace Analysis Using Multiple Runs', *J. Chromatogr. Sci.*, **23**, 521–523 (1985).
11. W.M. Coleman, III, J.L. White, T.A. Perfetti, 'A Hyphenated GC-based Quantitative Analysis of Volatile Materials from Natural Products', *J. Chromatogr. Sci.*, **32**, 323–327 (1994).
12. S.M. Abeel, A.K. Vickers, Daron Decker, 'Trends in Purge and Trap', *J. Chromatogr. Sci.*, **32**, 328–338 (1994).
13. W.M. Coleman, III, B.M. Lawrence, 'Comparative Automated Static and Dynamic Quantitative Headspace Analyses of Coriander Oil', *J. Chromatogr. Sci.*, **30**, 396–398 (1992).
14. B. Kolb, L.S. Ettre (eds.), *Static Headspace–Gas Chromatography: Theory and Practice*, Wiley-VCH, New York, 1997.
15. W. Vogt, K. Jacob, H.W. Obwexer, 'Sampling Method in Capillary Column GLC Allowing Injections up to 250 µL', *J. Chromatogr.*, **174**, 437–439 (1979).
16. W. Vogt, K. Jacob, A.B. Ohnesorge, H.W. Obwexer, 'Capillary GC Injection System for Large Sample Volumes', *J. Chromatogr.*, **186**, 197–205 (1980).
17. J. Teske, J. Efer, W. Engewald, 'Large Volume PTV Injection: Comparison of Direct Water Injection and In-vial Extraction for GC Analysis of Triazines', *Chromatographia*, **47**, 35–41 (1998).
18. K. MacNamara, R. Leardi, A. Hoffmann, 'Direct Injection of Distilled Spirits with PTV Matrix Removal: The Perfect Splitless Injection?', Presented *at Twentieth International Symposium on Capillary Chromatography*, eds. P. Sandra, A.J. Rackstraw, Riva del Garda, Italy, 1998.
19. D.R. Deans, 'A New Technique in Heartcutting in Gas Chromatography', *Chromatographia*, **1**, 18–22 (1968).
20 B.M. Gordon, C.E. Rix, M.F. Borgerding, 'Comparison of State-of-the-art Column Switching Techniques in High Resolution Gas Chromatography', *J. Chromatogr. Sci.*, **23**, 1–10 (1985).
21. J.F. Elder, B.M. Gordon, M.S. Uhrig, 'Complex Mixture Analysis by Capillary to Capillary Column Heartcutting GC/MS', *J. Chromatogr. Sci.*, **24**, 26–33 (1986).
22. B.M. Gordon, M.S. Uhrig, M.F. Borgerding, H.L. Chung, W.M. Coleman, III, J.F. Elder, J.A. Giles, D.S. Moore, C.E. Rix, E.L. White, 'Analysis of Flue-cured Tobacco Essential Oil by Hyphenated Analytical Techniques', *J. Chromatogr. Sci.*, **26**, 174–180 (1988).
23. E.L. White, M.S. Uhrig, T.J. Johnson, B.M. Gordon, R.D. Hicks, M.F. Borgerding, W.M. Coleman, III, J.F. Elder, Jr, 'Quantitative Determination of Selected Compounds in a Kentucky 1R4F Reference Cigarette Smoke By Multidimensional Gas Chromatography and Selected Ion Monitoring–Mass Spectrometry', *J. Chromatogr. Sci.*, **28**, 393–399 (1990).
24. H.-J. deGeus, J. deBoer, U.A. Th. Brinkman, 'Multidimensionality in Gas Chromatography', *Trends Anal. Chem.*, **15**, 398–408 (1996).
25. D.W. Wright, 'Applications of Multidimensional Gas Chromatography Techniques to Aroma Analysis', *Food Sci. Technol.*, **79**, 113–141 (1997).
26. R.P.W. Scott, *Introduction to Analytical Gas Chromatography, Second Edition, Revised and Expanded*, Marcel Dekker, Inc., New York, 1998.
27. R.E. Finnigan, 'Quadrupole Mass Spectrometers, from development to Commercialization', *Anal. Chem.*, **66**, 969A–975A (1994).
28. A.N. Eaton, 'Mass Detection is now a Viable Tool for the Chromatographer Thanks to the Simplification and ''Ruggedization'' of MS Techniques', *Today's Chemist at Work*, October, 34–37 (1994).
29. S. Meyerson, 'From Black Magic to Chemistry, The Metamorphosis of Organic Mass Spectrometry', *Anal Chem.*, **66**, 960A–964A (1994).
30. F. Wach, 'Benchtop GC/MS Instruments, Performance Comes in Smaller Packages', *Anal. Chem.*, **66**, 927A–930A (1994).
31. F.W. Karasek, R.E. Clement, *Basic Gas Chromatography–Mass Spectrometry, Principles and Techniques*, Elsevier, Amsterdam, 1988.
32. F.G. Kitson, B.S. Larson, C.N. McEwen (eds.), *Gas Chromatography and Mass Spectrometry*, Academic Press, San Diego, 1996.

33. J.Q. Walker (ed.), *Chromatography Fundamentals, Applications and Troubleshooting*, Preston Publications, Niles, IL, USA, 1996.
34. C.R. Clark, J. DeRuiter, F.T. Noggle, 'Chromatographic and Mass Spectrometric Methods for the Differentiation of *N*-Methyl-1-(3,4-methylenedioxyphenyl)-2-butamine from Regioisomeric Derivatives', *J. Chromatogr. Sci.*, **34**, 230–237 (1996).
35. R. Smits (ed.), '4th International Symposium on Hyphenated Techniques in Chromatography and Hyphenated Chromatographic Analyzers', *J. Chromatogr.*, **750**, 1–411 (1996).
36. R.A. Nyquist, *The Interpretation of Vapor-phase Infrared Spectra, Group Frequency Data*, Sadtler Research Laboratories, Philadelphia, Vol. 1, 1984.
37. P.R. Griffiths, J.A. deHaseth, *Fourier Transform Infrared Spectrometry*, John Wiley and Sons, New York, 1986.
38. P.R. Griffiths, J.A. deHaseth, L.V. Azarraga, 'Capillary GC/FTIR', *Anal. Chem.*, **55**, 1361A–1387A (1983).
39. J.F. Schneider, J.C. Demirgian, J.C. Stickler, 'A Comparison of GC/IR Interfaces: The Light Pipe Versus Matrix Isolation', *J. Chromatogr. Sci.*, **24**, 330–335 (1986).
40. J.C. Demirgian, 'Gas Chromatography–Fourier Transform Infrared Spectroscopy–Mass Spectrometry, A Powerful Tool for Component Identification in Complex Organic Mixtures', *Trends Anal. Chem.*, **6**, 58–64 (1987).
41. B. Lacroix, J.P. Huvenne, M. Deveaux, 'Gas Chromatography with Fourier Transform Infrared Spectrometry for Biomedical Applications', *J. Chromatogr.*, **492**, 109–136 (1989).
42. W.M. Coleman, III, B.M. Gordon, 'Analysis of Natural Products by Gas Chromatography/Matrix Isolation/Infrared Spectrometry', *Adv. Chromatogr.*, **34**, 57–108 (1994).
43. G.T. Reedy, D.G. Ettinger, J.F. Schneider, S. Bourne, 'High-resolution Gas Chromatography/Matrix Isolation Infrared Spectrometry', *Anal. Chem.*, **57**, 1602–1609 (1985).
44. W.M. Coleman, III, B.M. Gordon, 'Gas Chromatography–Matrix Isolation–Fourier Transform–Infrared Spectrometry of Natural Products', *J. Chromatogr. Sci.*, **29**, 371–376 (1991).
45. A.J. Barnes, W.J. Orville-Thomas, A. Muller, R. Gaufres (eds.), *Matrix Isolation Spectroscopy*, D. Reidel, Dordrecht: Holland, 1981.
46. E.L. Wehry, G. Mamantov, 'Matrix Isolation Molecular Spectrometry in Chemical Analysis', *Prog. Analyt. Spectrosc.*, **10**, 507–527 (1987).
47. R.N. Perutz, 'Matrix Isolation', *Annu. Rep. Progr. Chem.*, **82**, 157–191 (1986).
48. T.T. Holloway, B.J. Fairless, C.E. Freidline, H.E. Kimball, R.D. Kloepfer, C.J. Wurrey, L.A. Jonooby, H.G. Palmer, 'Performance of a Gas Chromatographic–Matrix Isolation–Fourier Transform Infrared Spectrometer', *Appl. Spectrosc.*, **42**, 359–369 (1988).
49. S. Jagannathan, J.R. Cooper, C.L. Wilkins, 'Matrix Effects in Matrix Isolation Infrared Spectroscopy', *Appl. Spectrosc.*, **43**, 781–786 (1989).
50. W.M. Coleman, III, B.M. Gordon, 'Examinations of the Matrix Isolation Fourier Transform Infrared Spectra of Organic Compounds: Part XVII', *Appl. Spectrosc.*, **43**, 1424–1427 (1989).
51. N. Ragunathan, K.A. Krock, C. Klawun, T.A. Sasaki, C.L. Wilkins, 'Multispectral Detection for Gas Chromatography', *J. Chromatogr. A*, **703**, 335–382 (1995).

Instrumentation of Gas Chromatography

G.A. Eiceman
New Mexico State University, Las Cruces, USA

1 Introduction 1
2 Carrier Gas 1
2.1 Selection of Gases 2
2.2 Control of Flow 2
2.3 Gas Sources and Purity 2
3 Sample Inlets 3
3.1 Syringes and Switching Valves 3
3.2 Pyrolysis 3
3.3 Other Topics 3
4 Ovens 4
4.1 Conventional Designs 4
4.2 Other Designs for Control of Column Temperature 4
5 Detectors and Data Systems 4
5.1 Detectors 5
5.2 Data Systems 5
6 Miniaturized, High-speed, and Portable Gas Chromatographs 5
6.1 Instrument Designs 6
6.2 Detectors 6
Acknowledgments 7
List of Symbols 7
Abbreviations and Acronyms 7
Related Articles 7
References 7

Instrumentation for gas chromatography (GC) comprises well-defined components, each of which contributes to the overall chromatographic performance. Most of these components have reached a mature level of technical development after nearly 50 years of development. Nonetheless, advances are occurring in an evolutionary, not revolutionary, manner with sample handling and in refinements of detectors. One area of dramatic advance in GC instrumentation is the development of small, sophisticated, portable gas chromatographs. Each component of a gas chromatograph (columns excepted) is discussed below where principles and recent developments are emphasized.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The key parts of a gas chromatograph include: a source of gas as the mobile phase, an inlet to deliver sample to a column, the column where separations occur, an oven as a thermostat for the column, a detector to register the presence of a chemical in the column effluent, and a data system to record and display the chromatogram. In addition, some facility is needed so that temperatures of various components can be accurately known and controlled. These parts of a gas chromatograph have been unchanged in function or purpose for over the last 40 years although technology has been ever improving in design, materials, and methodology. In particular, analog electronics for control of temperature zones and data acquisition were replaced with digital electronics and interfaced with computers in the 1970s and 1980s. The arrangement of these components is shown in a block diagram in Figure 1 and this arrangement is common to virtually all gas chromatographs regardless of age, model or manufacturer. A modern gas chromatograph is shown in Figure 2. In the discussion below, the general function of each component is provided with comments on the status of the technology. Most descriptions of GC will include a cursory description of instrumentation; few will provide a detailed treatment of the instrumentation or technical details. Some of the best discussions of hardware can be found in publications released by instrument manufacturers.[1,2] Unfortunately, these may not be found routinely in libraries but the reward for efforts to obtain them is found in the useful details for optimizing an analysis or practical help for maintaining the instrument.

The column may arguably be considered the key component of a gas chromatograph and accordingly has been treated separately under another heading. However, the total variance of a separation (σ_T) will conform to principles of error propagation and be a sum of variances from the injector (σ_I), column (σ_c), detector (σ_d), and data system (σ_{ds}), i.e. $\sigma_T = \sqrt{(\sigma_i + \sigma_c + \sigma_d + \sigma_{ds})}$. Thus, each of these components contributes to the overall efficiency of a GC separation and merits individual attention.

2 CARRIER GAS

The carrier gas or mobile phase in GC is an essential, but limiting, facet in separations. Carrier gas is the means to move constituents of a sample through the column and yet the choice of possible gases is restricted. Moreover, the carrier gas has properties that sometimes can complicate an analysis. Unlike liquid chromatography (where a wide selection of mobile phase compositions may be possible),

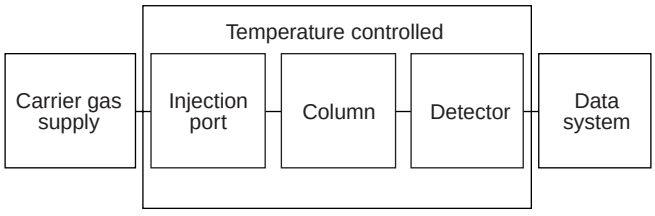

Figure 1 Block diagram of gas chromatograph.

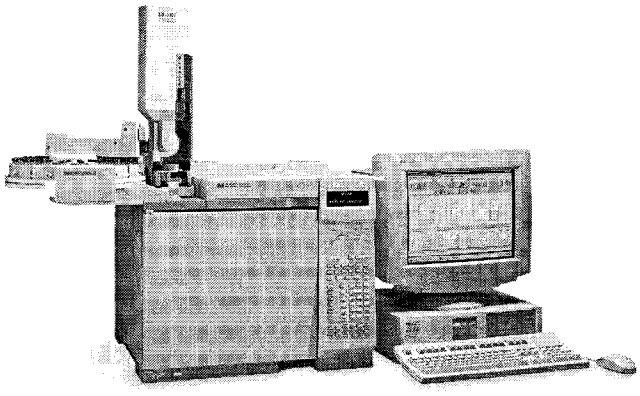

Figure 2 Photograph of modern gas chromatograph.

very little can be gained in separations through altering the mobile phase composition to influence the partition coefficient (k) or separation factor (α) in GC.

2.1 Selection of Gases

The choice of a practical carrier gas is simple: nitrogen or helium. Air may be used as a carrier gas under certain conditions with portable or on-site chromatographs but this is uncommon with laboratory-scale instruments. The choice of nitrogen or helium is made, in part, on the principles of separation and, in part, on economics: ~$20 for a nitrogen cylinder versus ~$50 for a helium cylinder. However, the selection is more complex than the prices of gas cylinders alone. Column efficiency in GC contains a term for contributions to longitudinal broadening in the carrier gas and this is given by the D_g term in the van Deemter equation.[3] This term is proportional to the square root of molar mass for the carrier gas, and nitrogen or argon would be preferred over helium based on D_g only. This effect can be seen in Figure 3, where nitrogen provides better performance than helium and has the lower contribution to plate height. However, the shape of the curve for height equivalent to a theoretical plate (HETP) versus flow rate (as linear velocity) for helium shows a reasonably good efficiency at high flow rates.[1,4] (HETP is equal to L/N, where L is the column length and N is the number of theoretical

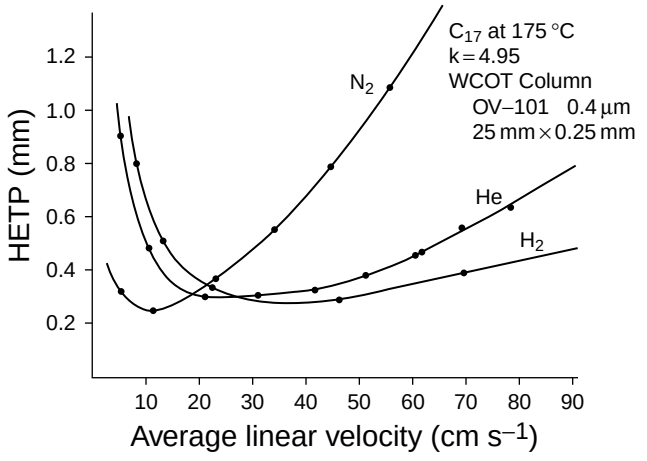

Figure 3 Plot of separation efficiency versus column flow rate.

plates in a column.) In contrast, the van Deemter curve for nitrogen is comparatively narrow. Consequently, a GC separation using nitrogen at 10 cm s^{-1} can be accomplished with comparable separating efficiency using helium at 50–60 cm s^{-1}. The practical consequence of this is that costs for using helium, on a per sample basis, might be lower than those for nitrogen when the speed of analysis is factored into the calculations.

2.2 Control of Flow

One difficulty in GC is the compressibility of the carrier gas and subsequent influence on separating performance. This was recognized in the first paper on GC where correction factors for gas flow rates were described.[5] The implications for isothermal methods are significant but will be critical with temperature programmed GC when column temperatures may span 200 °C or more.[6,7] When temperature is increased for a column with constant pressure on the inlet, the average flow rate in the column will decrease owing to increased viscosity of the gas mobile phase in a proportional but nonlinear manner. Under such conditions, flow rates may slow at high temperature and both separation speed and efficiency may suffer. Flow may be kept constant through mass flow meters that have inlet and outlet orifices, adjustable based upon pressure differences.[2] Constant flow can be delivered across a range of pressure drops that may arise due to changes in temperature but cannot compensate for changes in barometric pressure. An advance in instrumentation during the past decade has been the commercialization of flow programming so that flows may be made highly reproducible.

2.3 Gas Sources and Purity

A common gas source for nitrogen or helium is the pressurized cylinder or bottled gas supply, readily supplied

as a steel tank with a two-stage pressure regulator. This is still a common gas source though gas generators for nitrogen (air and hydrogen too) can be commercially competitive with bottled gas and have advantages in safety.

Regardless of the gas source, special attention must be given to the purity of tubing used to connect the source and the gas chromatograph and to impurities in the gas supply. Most columns do not tolerate moisture and oxygen well when operated at temperatures over 100 °C. Best results for column longevity and chromatographic reproducibility occur when the carrier gas is cleaned over molecular sieve beds (to reduce moisture). In addition, specialized traps can be purchased to reduce or remove hydrocarbons and oxygen in the carrier gas.

3 SAMPLE INLETS

The chromatographic process begins when sample is introduced into the column, ideally without disrupting flows in the column. The chromatographic results will be reproducible inasmuch as this is accomplished with a minimum of change in pressure or flow of the carrier gas or mobile phase. Also, the injection step establishes the initial (and best possible) peak width for the GC measurement. Thus, delivery of sample into the column should be controlled, reproducible, and rapid.

3.1 Syringes and Switching Valves

A common method for placing samples on a GC column is to use the microliter syringe with a needle to penetrate a plastic membrane. In this method a gas-tight seal is maintained and sample is deposited into a heated zone. If liquid or solid, sample is volatilized and swept to the column and this can be accomplished by manual injections in ~1 s. Syringe injection is a convenient and generally effective method though the thermoplastic septum develops leaks after repeated injections. Fatigue of the plastic septum limits the number of injections to ~30 before the septum must be replaced. A second difficulty arises with impurities from off-gassing or decomposition of the septum and these are seen as so-called ghost peaks or peaks in control blanks.[8]

Advances with capillary columns introduced unprecedented precision and accuracy to GC measurements and limitations with syringes became apparent.[9] Discrimination toward high boiling point components was seen with syringe injections and techniques to remedy the failings have been developed.[9] Sometimes thermal volatilization may lead to decomposition of samples so efforts to remove the discrimination and decomposition motivated the use of so-called on-column injections where sample is deposited directly from the syringe into the column. Another complication with syringe injections is the introduction of particulate and reactive materials into columns. Protection is afforded by precolumns. Further information on syringe injections and the range of options for injection methods can be found in excellent reference sources.[10,11]

Gas samples can be injected into the column using gas-tight syringes or using rotary gas switching valves that offer enormous flexibility for GC instruments.[12] Precision gas switching valves allow a gas sample to be measured with a precise volume and introduced into carrier gas flow without interrupting column flow. Sample is loaded into a loop and then, with a change in the valve position, is swept into the column under flow of the gas source. Heated switching valves such as those made by VICI, Inc.[13] are also useful in the analysis of sorbent traps. When traps are heated and switched in-series with the analytical column, constituents will be thermally desorbed for GC separations. Switching valves can be automated via electronic actuators and can be incorporated into purge-and-trap methods that are useful for characterizing aqueous samples for volatile organic constituents.

3.2 Pyrolysis

Another inlet option which is now routine in certain specific applications of material sciences is that of sample pyrolysis where solid samples are rapidly heated to a point of thermal decomposition in a reproducible manner. At temperatures in excess of 600 °C, substances such as natural or synthetic polymers thermally decompose to small molecular weight, stable substances that provide a chromatographic profile which is unique to certain materials. Such an injector enlarges the application of GC to solid samples that would not normally be considered suitable for GC characterization, and pyrolysis methods have become standardized for some applications such as assaying plastics. A journal now exists for analytical applications of pyrolysis, the *Journal of Analytical and Applied Pyrolysis*. Attachments to inlets are commercially available and serve to extend GC in forensic and industrial applications, as shown in Table 1.

3.3 Other Topics

In purge-and-trap methods, a small amount of water is treated vigorously with an inert gas flow sufficient to sweep volatile organic compounds (VOCs) into the gas phase from the water phase. The gas flow, including the VOCs, is passed through an adsorbent trap where VOCs are retained and concentrated while water vapor and

Table 1 Examples of applications of pyrolysis GC

Example	Ref.
Lignin by pyrolysis methylation	14
Synthetic polymers	
Fast GC	15
Bibliography	16
Rosin glycerin esters in paper	17
Chlorinated polyethylene structure	18
Coating materials: bibliography	19
Proteinaceous binders in paints	20

Table 2 Examples of applications of SPME methods with GC analyses

Example	Ref.
Dimethyl sulfide in beer	23
Diacetyl in wine	24
Organochlorine compounds in water	25
Wine headspace compounds	26
Ecstasy and amphetamine in confiscated samples	27
Parathion in biological samples	28
Trimethylamine in urine	29
Volatile compounds in sunflower oil	30

gas are unretained and released to vent. After a short time of purging, the trap is placed in series with the gas chromatograph via a gas switching valve and VOCs are transferred to the analytical column. This is done through a rapid increase in temperature of the sorbent trap and the vapors are carried to the analytical column under carrier gas flow. Purge-and-trap methods are the basis for an expansive use of GC for monitoring the VOCs at ppb levels in aqueous samples as specified[21,22] by the United States Environmental Protection Agency (USEPA). Such instrumentation and methods are available today in commercial instrumentation and can be fully automated. An advantage in this method of sample preparation is that nonvolatile constituents, which could potentially foul the column, remain in the sample container and do not enter the column.

Another adaptation of GC for trace organic analysis was the development of solid-phase microextraction (SPME) with a syringe design. In this, a stationary phase is bonded on a silica fiber that can be introduced to a gas or aqueous sample. Organic chemicals partition into the coated fiber during a time of contact and the fiber is withdrawn into a syringe needle. In an injection port the fiber is pushed from the protection of the needle and sample is thermally desorbed into the GC column. This method has advantages of convenience and simplicity and the growth of use of SPME injections has been dramatic during the 1990s. A list of some representative applications is given in Table 2.

4 OVENS

4.1 Conventional Designs

Liquids or solids must be converted to vapor state and maintained as a vapor throughout the GC separation. Therefore, most gas chromatographs are equipped with ovens to keep the column at temperatures from 40 to 350 °C. Exceptions are those chromatographs that are used in separating simple gases such as light hydrocarbons or permanent gases. Early gas chromatographs were equipped with isothermal ovens. Today, temperature-programmed ovens allow separations of chemicals spanning a range of vapor pressures in a single analysis.

Conventional ovens, unchanged in decades, consist of a resistive wire coil that radiates into the inner volume of the oven. Heat from the resistive wire source is spread, ideally in an even manner, throughout the oven volume using a fan attached to an electric motor. A thermistor or thermocouple inside the oven is part of regulating the oven temperature via the amount of heat released by the heating element. This is controlled by the power delivered to the element and a feedback circuit to control and program the oven temperature. Efforts to create isothermal conditions, i.e. no thermal gradients inside the oven volume, are essential for reproducible chromatography and are criteria in evaluating good oven designs. Gradients in excess of a few degrees between various regions of an oven are practical in the best of oven designs and can be more than a few degrees in poorly designed ovens. One of the only systematic evaluations of GC ovens was given by Welsh[31] and his discussion provides measures for characterizing GC ovens.

4.2 Other Designs for Control of Column Temperature

Several alternatives to conventional ovens have been devised and may be especially helpful for short columns or instances where little space is available for a bulky, heated air oven. Two approaches have been used and include small thermal ovens[32] and innovative column heating arrangements.[33] Column heating based on resistive heating is compact, uses minimal power, and can decrease analysis times.[34] These methods are based upon application of heat directly to the column or a base upon which the column is crafted or attached. The approach is unlikely to become a laboratory standard but is being explored for use in miniature or portable gas chromatographs.[35]

5 DETECTORS AND DATA SYSTEMS

The subject of detectors in GC is a pivotal theme since the separation processes will have been wasted if the analyte

cannot be detected. Excellent primers on GC detectors are available[36,37] and any general text on instrumental analysis will have introductory material on the common detectors. A biennial review contains an extensive section on developments of GC detectors and can serve as a guide to primary literature.[38]

5.1 Detectors

Effluent from the column enters a detector where the composition of the carrier gas stream is characterized through one of several possible chemical or physical properties of molecules. The mainstays in GC have been the flame ionization detector (FID), the thermal conductivity detector (TCD) and the electron capture detector (ECD). Other commercially available detectors include the photoionization detector (PID), the nitrogen–phosphorus detector and the atomic emission detector, though these have been less prevalent historically than the FID, TCD, and ECD. Other detectors have been introduced through the years but have never become widely used in GC methods. The FID relies upon the formation of gaseous ions from organic molecules combusted in a hydrogen–air flame; the TCD is based upon changes in the heat absorbing properties of the gas effluent when the carrier gas is altered with analyte; the ECD response is governed by the ability of some molecules to attract and remove thermalized electrons. Despite long-standing conventions for the design and operation of these detectors, advances still occur.

Examples of evolutionary changes include the small FID designs[39] and designs where gas mixing is arranged to provide optimum response.[40–42] A recurring theme in advances in ECD has been a nonradioactive alternative to the normal source, 10 mCi of ^{63}Ni. Despite promising discoveries, the radioactive source is still the favored choice. The applications of ECD illustrate the advantages of selective detectors where analyte can be found in the presence of potentially interfering matrix.

Examples where the ECD was chosen to detect a specific chemical family over interfering backgrounds include: halocarbons in air for oceanographic tracer studies;[43] chlorobutanol in mouse tissues and fluids;[44] organochlorine compounds in milk products;[45] pesticides and other organochlorides in water;[46] organochlorine pesticides in edible oils and fats.[47]

In the last two decades, inexpensive mass spectrometers or mass-selective detectors (MSDs) have dramatically transformed the practice of GC. Once the purview of laboratories able to sustain the high cost of mass spectrometers and the high level of maintenance, instrument manufacturers made mass spectrometers both robust and inexpensive. This development, when combined with the appreciation that analytical confidence is highest with a mass spectrometer as the detector, has resulted in a near general availability of gas chromatography/mass spectrometry (GC/MS) instrumentation. In a GC/MS analysis, a mass spectrum can be obtained continuously at fixed intervals of ~0.1 s throughout the analysis. Consequently, a mass spectrum can be obtained for each chromatographic peak and the shoulders and baselines in the chromatogram. No other detector can provide the richness of information available in such results. Detection limits can be enhanced through the use of single ion monitoring where the mass spectrometer is used for detecting the intensity of one or a few ions. This can provide the specificity of a mass spectral pattern for response without losses in detection limit associated with scanning over unused m/z space. One revolution in the past decade has been the application of powerful desktop computers to control instrumentation and especially to control data acquisition and handling.

5.2 Data Systems

At a fundamental level, acquisition of chromatographic results has been little changed since the early days of GC, though the digital revolution has meant that strip chart recorders, once the mainstay of collecting chromatograms, cannot be found today and only electronic recording-integrators or microcomputers are used. Signal from the detector amplifier is digitized and stored to disk allowing enormous convenience in retrieving and replaying results. This means that peak retention times, peak areas, etc. are automatically reported and have been since the mid-1970s. In addition, software allows the results to be displayed in an automated manner, i.e. reports can be generated according to standard reporting formats. All this can be economically integrated into the total instrument control and management through computers and is an option on all chromatographs and a standard feature on most instruments.

6 MINIATURIZED, HIGH-SPEED, AND PORTABLE GAS CHROMATOGRAPHS

One area of GC that has shown vibrancy with advances during the 1990s is that of small, fast, and portable GC instruments. Though process gas chromatographs were amongst the first sophisticated analyzers placed into industrial on-site measurements, the subject has taken new significance following the burgeoning environmental movement. Making measurements where a sample is located rather than relocating samples to a centralized laboratory underlies this trend. Recently, a new journal

has appeared to support these efforts, *Field Analytical Chemistry and Technology*, which includes portable GC advances.

In high-speed GC, retention times can be pushed under a few minutes or seconds with short, narrow bore columns or high flow rates. Part of the challenge in fast GC is the compressibility of the carrier gas and the necessary speed (low time constants) for subcomponents such as injectors and detectors for high-speed separations.

6.1 Instrument Designs

An example of the size possible for small gas chromatographs is an ultimate miniature gas chromatograph created using silicon micromachining and integrated circuit processing techniques.[48] This GC analyzer contains a 0.9 m long × 300 μm wide × 10 μm high rectangular column coated with a 0.2-μm thick liquid phase. The injector is a 10-μm-long sampling loop with the same cross-section as the column. Dual detectors based upon a coated chemiresistor and on thermal conductivity are used. The complete system is packaged in less than 23 cm^2 and is 2.5 mm high. Although limited in scope to the detection of ammonia and nitrogen dioxide, this miniature chromatograph offers exciting possibilities for future field instruments.

One trade-off for high-speed GC is the loss of capacity due to the smaller diameter and shorter columns. Application of packed capillary columns in high-speed GC has been shown to improve capacity and selectivity[49] while obtaining high-speed separation for light hydrocarbons.[50] An alternative is the multicapillary column which improved capacity while maintaining the efficiency obtained with small internal diameter columns.[51]

Injection techniques for high-speed GC must provide narrow bandwidths due to fast analysis time requirements without compromises in resolution.[52] One means to accomplish this is through cryogenic inlets which provide narrow bandwidths[53] and in some instances injection times can be shorter than 10 ms.[54]

6.2 Detectors

Mass spectrometers might be considered ideal detectors and adaptations to fieldable gas chromatographs have been made. One design is the size of a standard size suitcase, weighs ~31 kg and uses ~600 W under peak loads.[55] Naturally, field portability is a goal and man-portability has been demonstrated.[56] A cart-portable unit has been described and was capable of screening 1000 samples an hour at a speed of 20–32 km an hour.[57] Portable GC/MS instruments were also used for industrial hygiene applications.[58] As the need for field measurements increases and mass spectrometry (MS) becomes miniaturized, the number of field portable GC/MS systems can be expected to increase dramatically.

Another sophisticated detector for GC is the ion mobility spectrometer and recently it too has been coupled to gas chromatographs for field analysis.[59,60] The utility of this hand-held gas chromatograph/ion mobility spectrometer analyzer was demonstrated by rapid screening (~20 s) with chemical warfare agents such as dimethyl methylphosphonate.[61] One of the most impressive demonstrations was that from the US Army where a portable pyrolysis-gas chromatograph/ion mobility spectrometer was developed for screening airborne pathogens. This unit is shown in Figure 4.

This part of GC instrumentation may be expected to undergo continued development as interest in speed of analysis pushes dimensions and hardware. Therefore, it represents one of the few areas of growth in GC

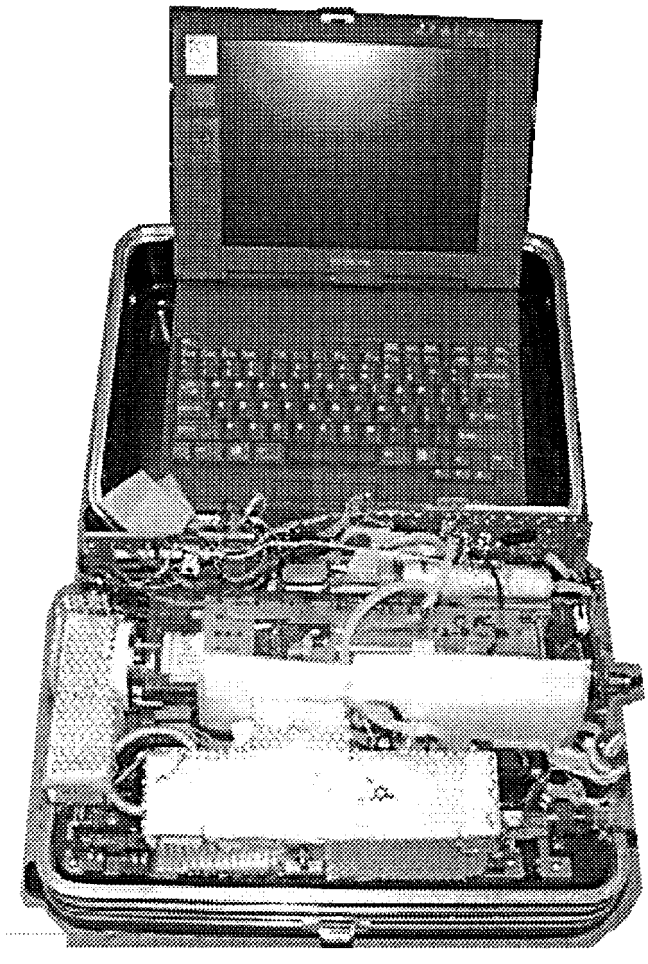

Figure 4 Gas chromatograph/ion mobility spectrometer for bacterial analysis.

instrumentation that can be recognized in the absence of a revolutionary change in GC methods.

ACKNOWLEDGMENTS

The assistance of David Young, Jonathan Bergloff, and Kara Douglas in checking the manuscript is gratefully acknowledged.

LIST OF SYMBOLS

α	Separation factor
D_g	Diffusivity of solute in carrier gas
k	Partition coefficient

ABBREVIATIONS AND ACRONYMS

ECD	Electron Capture Detector
FID	Flame Ionization Detector
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HETP	Height Equivalent to a Theoretical Plate
MS	Mass Spectrometry
MSD	Mass-selective Detector
PID	Photoionization Detector
SPME	Solid-phase Microextraction
TCD	Thermal Conductivity Detector
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound

RELATED ARTICLES

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

REFERENCES

1. K.J. Hyver, P. Sandra, *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard Corporation, HP Part No. 5950-3562, 1989.
2. B. Thompson, *Fundamentals of Gas Analysis by Gas Chromatography*, Varian Associates, Inc., Palo Alto, CA, 1977.
3. H.W. Moody, 'Evaluation of the Parameters in the Van Deemter Equation', *J. Chem. Educ.*, **59**(4), 290 (1982).
4. R.H. Perrett, J.H. Purnell, 'Contribution of Diffusion and Mass Transfer Processes to Efficiency of Gas Liquid Chromatography Columns', *Anal. Chem.*, **35**, 430 (1963).
5. A.T. James, A.J.P. Martin, 'Gas–Liquid Partition Chromatography: the Separation and Micro-estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid', *Biochem. J.*, **50**, 679 (1952).
6. W. Harris, H. Habgood, 'Some Temperature Effects in Gas Chromatography', *Talanta*, **11**, 115 (1964).
7. W. Harris, H. Habgood, *Programmed Temperature Gas Chromatography*, J. Wiley & Sons, New York, 1966.
8. V.M. Olsavicky, 'A Comparison of High Temperature Septa for Gas Chromatography', *J. Chromatogr. Sci.*, 197–200 (1978).
9. K. Grob, Jr, H.P. Neukom, 'The Influence of Syringe Needle on the Precision and Accuracy of Vaporizing GC Injections', *J. High Res. Chrom. Chrom. Commun.*, **2**, 15–21 (1979).
10. K. Grob, 'Classical Split and Splitless Injection', in *Capillary GC*, A. Huethig, Heidelberg, 1987.
11. K. Grob, 'On-column Injection', in *Capillary GC*, A. Huethig, Heidelberg, 1987.
12. G.L. Pratt, J.H. Purnell, 'Sampling Valve for Use in Gas Chromatographic Analysis of the Products of Gaseous Reactions', *Anal. Chem.*, **32**, 1213 (1960).
13. See http://www.vici.com/
14. K. Kuroda, A. Izumi, 'Analysis of Lignin by Pyrolysis–Methylation: Pyrolysis Products Derived from Beta-O-4, Beta-beta, Beta-5, and Beta-1 Substructures', *Mokuzai Gakkaishi*, **43**(1), 112–114 (1997).
15. F.C.Y. Wang, A.D. Burleson, 'The Development of Pyrolysis Fast Gas Chromatography for Analysis of Synthetic Polymers', *J. Chromatogr. A*, **833**(1), 111–119 (1999).
16. J.K. Haken, 'Pyrolysis Gas Chromatography of Synthetic Polymers: a Bibliography', *J. Chromatogr. A*, **825**(2), 171–187 (1998).
17. L. Wang, Y. Ishida, H. Ohtani, S. Tsuge, 'Determination of Fortified Rosin Glycerin Ester Sizing Agents in Paper by Reactive Pyrolysis Gas Chromatography in the Presence of an Organic Alkali', *Anal. Sci.*, **14**(2), 431–434 (1998).
18. F.C.Y. Wang, P.B. Smith, 'Composition and Microstructure Analysis of Chlorinated Polyethylene by Pyrolysis Gas Chromatography Mass Spectrometry', *Anal. Chem.*, **69**(4), 618–622 (1997).
19. G. Chivair, N. Gandini, P. Russo, D. Fabbri, 'Characterisation of Standard Tempera Painting Layers Containing Proteinaceous Binders by Pyrolysis (/Methylation) Gas Chromatography Mass Spectrometry', *Chromatographia*, **47**(7–8), 420–426 (1998).
20. J.K. Haken, 'Pyrolysis Gas Chromatography of Coating Materials: a Bibliography', *Prog. Org. Coat.*, **36**(1–2), 1–10 (1999).

21. T.A. Bellar, 'Method 502.1 Volatile Halogenated Organic Compounds in Water by Purge and Trap Gas Chromatography', Environmental Monitoring Systems Laboratory, US Environmental Protection Agency Office of Research and Development, Cincinnati, Ohio, 1989.
22. United States Environmental Protection Agency (USEPA), 'SW-846 Method 5030A Purge-and-Trap-Sample Preparation and Extraction for Volatile Organics', Test Methods for Evaluating Solid Waste, Vol. 1B, Physical/Chemical Methods, SW-846, 3rd edition, USEPA Office of Solid Waste and Emergency Response. Washington, DC, 1995.
23. C.J. Scarlata, S.E. Ebeler, 'Headspace Solid-phase Microextraction for the Analysis of Dimethyl Sulfide in Beer', *J. Agric. Food Chem.*, **47**(7), 2505–2508 (1999).
24. Y. Hayasaka, E.J. Bartowsky, 'Analysis of Diacetyl in Wine Using Solid-phase Microextraction Combined with Gas Chromatography Mass Spectrometry', *J. Agric.*, **47**(2), 612–617 (1999).
25. D.N. Zhang, Z.P. Zhou, Y.Z. Tang, C.Y. Wu, W. Zhan, Y. Xu, 'Analysis of Organchlorine Compounds in Water by Solid Phase Microextraction and Gas Chromatography', *Chin. J. Anal. Chem.*, **27**(7), 768–772 (1999).
26. G.Y. Vas, K. Koteleky, M. Farkas, A. Dobo, K. Vekey, 'Fast Screening Method for Wine Headspace Compounds Using Solid-phase Microextraction (SPME) and Capillary GC Technique', *Am. J. Enol. Vitic.*, **49**(1), 100–104 (1998).
27. K.E. Kongshaug, S. PedersenBjergaard, K.E. Rasmussen, M. Krogh, 'Solid-phase Microextraction/Capillary Gas Chromatography for the Profiling of Confiscated Ecstacy and Amphetamine', *Chromatographia*, **50**(3–4), 247–252 (1999).
28. F. Musshoff, H. Junker, B. Madea, 'Rapid Analysis of Parathion in Biological Samples Using Headspace Solid-phase Micro-extraction (HS-SPME) and Gas Chromatography/Mass Spectrometry (GC/MS)' *Clin. Chem. Lab. Med.*, **37**(6), 639–642 (1999).
29. G.A. Mills, V. Walker, H. Mughal, 'Quantitative Determination of Trimethylamine in Urine by Solid-phase Microextraction and Gas Chromatography Mass Spectrometry', *J Chromatogr. B*, **723**(1–2), 281–285 (1999).
30. A. Keszler, K. Heberger, M. Gude, 'Identification of Volatile Compounds in Sunflower Oil by Headspace SPME and Ion-trap GC/MS', *HRC-J. High Res. Chrom.*, **21**(6), 368–370 (1998).
31. P. Walsh, 'Evaluation of Commercial Gas Chromatograph Ovens', *Am. Lab.*, **9**, 35 (1977).
32. A.P. Snyder, C.S. Harden, A.H. Brittain, M.G. Kim, N.S. Arnold, H.L.C. Meuzelaar, 'Portable Hand-held Gas Chromatography Ion Mobility Spectrometry Device', *Anal. Chem.*, **65**, 299–306 (1993).
33. V. Jain, J.B. Phillips, 'Fast Temperature Programming on Fused-silica Open-tubular Capillary Columns by Direct Resistive Heating', *J. Chromatogr. Sci.*, **33**(1), 55–59 (1995).
34. E.U. Ehrmann, H.P. Dharmasena, K. Carney, E.B. Overton, 'Novel Column Heater for Fast Capillary Gas Chromatography', *J. Chromatogr. Sci.*, **34**(12), 533–539 (1996).
35. M. vanDeursen, J. Beens, C.A. Cramers, H.G. Janssen, 'Possibilities and Limitations of Fast Temperature Programming as a Route Towards Fast GC', *HRC-J. High Res. Chrom.*, **22**(9), 509–513 (1999).
36. R. Buffington, M.K. Wilson, *Detectors for Gas Chromatography – a Practical Primer*, Hewlett-Packard Corporation, Part No. 5958-9433, 1987.
37. F.W. Karasek, L.R. Field, 'Ionization Detectors in Gas Chromatography', *Res./Dev.*, **28**, 42–44 (1977).
38. G.A. Eiceman, H.H. Hill, Jr, J. Gardea-Torresday, 'Gas Chromatography', *Anal. Chem. Fundam. Rev.*, **70**, 321R–339R (1998).
39. C.K. Lo, Patent, US Appl. 375,330, 17 Jan, 1995.
40. L.T. Mathews, J.L. Medici, US Appl. 294,269, 23 Aug, 1994.
41. T. Yoshihara, M. Kamoshita, *Jpn. Kokai Tokkyo Koho JP*, 07,218,490, 18 Aug, 1995.
42. K. Hirai, *Jpn. Kokai Tokkyo Koho JP*, 07,103,957, 21 Apr, 1995.
43. S.M. Boswell, D. Smythe-Wright, 'Dual-detector System for the Shipboard Analysis of Halocarbons in Sea-water and Air for Oceanographic Tracer Studies', *Analyst (Cambridge, U.K.)*, **121**(4), 505–509 (1996).
44. T.J. Amiss, I.W. Smoak, 'Determination of Chlorobutanol in Mouse Serum, Urine and Embryos by Capillary Gas-Chromatography with Electron-capture Detection', *J. Chromatogr. B*, **673**(1), 59–66 (1995).
45. G. Filek, M. Bergamini, W. Lindner, 'Steam Distillation-solvent Extraction, a Selective Sample Enrichment Technique for the Gas-chromatographic Electron-capture Detection of Organochlorine Compounds in Milk Powder and Other Milk-products', *J. Chromatogr. A*, **712**(2), 355–364 (1995).
46. C. Aguilar, F. Borrull, R.M. Marce, 'Determination of Pesticides in Environmental Waters by Solid-phase Extraction and Gas Chromatography with Electron-capture and Mass-spectrometry Detection', *J. Chromatogr., A*, **771**(1–2), 221–231 (1997).
47. M.W. Arend, K. Ballschmiter, 'A New Sample Preparation Technique for Organochlorines in Cod Liver Oil Combining SPE and NP-HPLC with HRGC-ECD', *Fresenius J. Anal. Chem.*, **366**(4), 324–328 (2000).
48. R.D. Mortimer, J.B. Shields, 'Total, Free, and Glucose-conjugated Acid Metabolites from Permethrin and Cypermethrin in Foods by Gas-Chromatography Electron-capture Detection', *J. Assoc. Off. Anal. Chem. Int.*, **78**(5), 102–107 (1995).
49. H. Yun, M.L. Lee, 'Fast Gas Chromatography of Light Hydrocarbons and Permanent Gases on Porous-layer Open-tubular Columns', *Field Anal. Chem. Technol.*, **1**(1), 60–64 (1996).

50. Y. Shen, M.L. Lee, 'High Speed Solbatin Gas Chromatography Using Packed Capillaries Containing Sub-5 μm Particles', *J. Chromatogr. A*, **778**(1+2), 31–42 (1997).
51. J.J. Baumbach, G.A. Eiceman, D. Klockow, S. Sielemann, A.V. Irmer, 'Exploration of a Multicapillary Column for Use in Elevated Speed Gas Chromatography', *Int. J. Environ. Anal. Chem.*, **66**, 225–239 (1997).
52. M.A. Klemp, A. Peters, R.D. Sacks, 'High-speed GC Analysis of VOCs: Sample Collection and Inlet Systems', *Environ. Sci. Technol.*, **28**(8), A369–A376 (1994).
53. M.A. Klemp, M.L. Akard, R.D. Sacks, 'Cryofusing Inlet with Reverse Flow Sample Collection for Gas-chromatography', *Anal. Chem.*, **65**(18), 2516–2521 (1993).
54. A.J. Borgerding, C.W. Wilkerson, 'Cryogenically Cooled Microloop System for Sampling and Injection in Fast GC', *J. Anal. Chem.*, **68**(4), 701–707 (1996).
55. B.D. Andresen, J.D. Eckels, J.F. Kimmons, D.W. Myers, 'Portable Gas Chromatograph–Mass Spectrometer', US Appl. 5,525,799, 11 June, 1996.
56. R.J. Matejczyk, 'Proceedings of Micro- and Nanofabricated Structures and Devices for Biomedical Environmental Applications', *Proc. SPIE-Int. Soc. Opt. Eng.*, **2941**, 140–146 (1997).
57. H.L.C. Meuzelaar, W.H. McClennen, J.P. Dworzanski, N.S. Arnold, W.-H. Du, 'Measurement and Data Management of Air Pollutants', *Proc. Annu. Meet.-Air Waste Manage. Assoc.*, **88**(10), 95-MP17.01 (1995).
58. H.V. Piltingsrud, 'A Field Deployable Gas Chromatograph Mass Spectrometer for Industrial Hygiene Applications', *Am. Ind. Hyg. Assoc. J.*, **58**(8), 564–577 (1997).
59. D.A. Atkinson, 'Evaluation of Gas Chromatography Coupled with Ion Mobility Spectrometry for Use as a Field Screening Tool', *Diss. Abstr. Int., B*, **56**(6), 3161 (1995).
60. A.G. Loudon, J. Kolbe-Boelke, J. Adler, J. Stach, 'The Use of On-site Analysis for the Protection of Personnel', *Analusis*, **23**(2), M22–M24 (1995).
61. J.P. Dwarzanski, W.H. McClennen, P.A. Cole, S.N. Thornton, H.L.C. Meuzelaar, N.S. Arnold, A.P. Snyder, 'Field-portable, Automated Pyrolysis-GC/IMS System for Rapid Biomarker Detection in Aerosols: a Feasibility Study', *Field Anal. Chem. Tech.*, **1**(5), 295–306 (1997).

Liquid Phases for Gas Chromatography

Nicholas H. Snow
Seton Hall University, South Orange, USA

1 Introduction 1
2 The Role of the Stationary Phase 2
2.1 Solute–Stationary Phase Interactions 2
2.2 Important Separation Variables 3
2.3 Packed Columns 5
2.4 Capillary Columns 5
2.5 Comparison of Packed and Capillary Column Separations 5
3 Packed Columns 6
3.1 Tubing Materials and Dimensions 6
3.2 Solid Supports 6
3.3 Packing and Preparing Packed Columns 7
3.4 Instrumentation Requirements 8
3.5 Choosing a Stationary Phase 8
3.6 Guide to Packed Column Gas Chromatography Liquid Phases 9
4 Capillary Columns 9
4.1 Design and Manufacture 10
4.2 Instrumentation Requirements 10
4.3 Choosing a Liquid Phase 11
4.4 Guide to Stationary Phases for Capillary Gas Chromatography 11
5 Care and Maintenance of Gas Chromatography Columns 12
5.1 Packed Columns 12
5.2 Capillary Columns 12
5.3 Quality Testing 13
6 Recent and Future Developments 14
6.1 Liquid Phases for Specific Applications 14
6.2 Chiral Stationary Phases 14
6.3 Organic Polymer Stationary Phases 15
7 Conclusions 15
Acknowledgments 15
List of Symbols 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

In gas chromatography (GC) the liquid phase, also called the stationary phase, provides the medium in which the separation occurs. The magnitude of separation is given by the relative strengths of intermolecular interactions between analyte and stationary phase molecules, thus numerous liquids have been used over the years. The stationary phase can be coated onto a solid support and packed into a tube (packed column GC), or coated onto the wall of a capillary (capillary GC). Today, most GC analyses are performed using capillary columns.

1 INTRODUCTION

In GC, the liquid phase, which is coated onto the inside wall of a capillary column, or onto solid particles in a packed column, is commonly referred to as the "heart" of the system, as this is where the entire separation occurs. Since the inception of GC, the design of columns and of liquid phases has undergone significant changes, but columns may still be classified into several distinct categories. If the stationary phase is a solid, the technique is termed gas–solid chromatography (GSC); if it is a liquid, then the technique is called gas–liquid chromatography (GLC). Packed columns arise from packing a tube, typically 6.3 or 3.2 mm (1/4 or 1/8 inch) in outside diameter with stationary phase particles that may or may not be coated with a liquid phase. Capillary or open tubular columns come in a variety of dimensions, from 0.05 mm to 0.53 mm in inside diameter and have a liquid or solid stationary phase coated onto the inside walls of the capillary. Capillary columns are generally used for applications that demand high resolution, trace analysis or high temperatures, while packed columns are generally used for applications requiring high sample capacity, ease of use and low temperatures. This article addresses liquid phase coatings for packed and capillary columns.

In any GC column, packed or capillary, the ability to achieve a separation depends upon specific interactions between the stationary phase material and the specific analytes, and upon the vapor pressure of the analyte. As a rule of thumb, GC can analyze any compound for which vapor pressure or boiling point data can be obtained by other methods. The specific interactions include dispersion forces, hydrogen bonding and weak electrostatic interactions. It is the need to modulate these interactions to produce a desired selectivity and the need for stable inert liquid phase materials that drives the development of new stationary phases. Column dimensions, such as length, tubing inside diameter, liquid phase loading and particle diameter for packed columns and length, tubing inside diameter, and liquid coating film thickness for capillary columns also have strong effects on separation capability.

Table 1 Time line of important developments in liquid phases for GC

Date	Event
1952	GC invented by Martin and James; Packed columns developed
1958	Theory of capillary column performance described
1958	RI proposed
1959	Patent on capillary columns by Perkin Elmer; inlet splitter
1960	Glass drawing machine developed by Desty
1966	Rohrschneider introduces phase constants
1979	Fused silica capillary columns invented
1983	Introduction of megabore columns as alternative to packed columns
1980s	Development of quality testing and deactivation procedures; cross-linked polymeric phases; interface to spectroscopic detectors
1990s	Programmed temperature vaporizer; electronic pressure control

RI, kovats retention index.

Packed columns were introduced in 1952 by Martin and James along with the first gas chromatograph.[1] The vast majority of GC work in the 1950s–1970s was performed using packed columns, as capillary columns were prohibitively difficult to use for routine work at the time. Since the plate numbers of packed columns are limited, separation power was generated by novel stationary phase design, leading to the availability of hundreds of stationary phases and column designs. Today, use of packed columns is more limited, as capillary columns, with their high plate numbers and improved analytical sensitivity, have become routine.

Capillary, or open tubular columns, introduced in the late 1950s, were originally made of glass or stainless steel and were very difficult to use. In 1956, Golay[2] described the tremendous potential benefit of the open tubular arrangement. However, capillary columns saw relatively little routine use over the next 25 years, as the glass columns commonly in use were very difficult to make, install into instruments and easy to break. In 1979, Dandeneau and colleagues invented the fused silica capillary columns in use today.[3] Since then, capillary column manufacture and use has increased dramatically, capturing approximately 80% of the GC column market.[4] Immediate benefits were seen, as routine high-resolution trace analysis became possible. Table 1 provides a brief time line of major developments in GC liquid phase technology.

2 THE ROLE OF THE STATIONARY PHASE

In contrast to liquid phase chromatographic separations, in which separating power is determined by both the mobile and stationary phases, in GC, only the stationary phase–analyte interactions determine the separation capability of a column. The mobile phase (carrier gas) acts only as a carrier of vapor phase molecules. Thus, in developing GC methods, the nature of the stationary phase is a primary consideration. Both the chemical nature of the liquid phase and the dimensions of the column make strong contributions to the capability to develop a separation. Because the mass of liquid phase present in a capillary column is about 1000 times less than in a packed column, and because packed columns are practically limited to about 0.3 m (12 feet) in length, while capillary columns may be 100 m long or more, there are significant differences in approach to method development, so these are treated separately. These dimensional differences lead to the many different types of separations that can be performed on capillary columns.

2.1 Solute–Stationary Phase Interactions

Analyte vapor pressure and the nonbonded chemical interactions that influence solubility are the chief contributors to retention in GC. Thus, if high selectivity is the goal of choosing a stationary phase, these must be considered carefully. The major forces involved in determining the strength of retention on a given column are:

- London forces, which are based on dispersion between molecules and generate the effective boiling point of the analyte sorbed onto the stationary phase;
- Keesom forces, which are dipole–dipole interactions between analyte and stationary phase molecules;
- Debye forces, which are induced dipole-induced dipole interactions between analyte and stationary phase molecules; and hydrogen bonding in which polar functional groups on the analyte molecules interact with similar groups on the stationary phase.

Of these, hydrogen bonding and London forces will provide the strongest interactions, while Keesom and Debye forces will be secondary. Thus, in choosing a stationary phase, structural similarity between the analyte and stationary phase, which for nonpolar compounds results in London forces and for polar compounds results in hydrogen bonding, should be the major consideration in determining the retention power of the stationary phase for a given analyte, or the selectivity for a given group of analytes.

The basic characteristics required of all GC stationary phases can therefore be summarized. First, the analyte must be soluble in the stationary phase material. In selecting a phase, ''like dissolves like'' is often sufficient for effective choices. Otherwise, considering hydrogen

bonding and other intermolecular interactions between analyte and stationary phase is required. Second, if more than one component is present in the mixture, the differential solubility, which determines selectivity, of the analytes in the stationary phase must be high. Physically, the stationary phase should exhibit low vapor pressure, so that it can be operated at high temperatures and should have low viscosity at lower temperatures. These variables will be addressed further in sections 3.5 and 4.3, for packed and capillary columns, respectively.

In the 1960s, much work focused on the characterization of the many stationary phases then available for packed columns. The classic work performed by Kovats, Rohrschneider and McReynolds provided much of the basis for future developments in column chemistry. Also, in the 1970s and 1980s, quality tests developed by Grob, Grob and Grob[5,6] and others, continued to provide insight into column chemistry and performance. The Grob tests are addressed further in section 5.3. In order to compare chemistry among diverse stationary phases, a common means for recording retention times that accounted for variations on operating conditions was required. The Kovats retention index (RI)[7] was the first of these. Retention times of the *n*-alkanes are used as standards and the retention times of other compounds are calculated by Equation (1):

$$\mathrm{RI} = 100N + 100 \ \frac{\log t'_{\mathrm{R}}(\mathrm{an}) - \log t'_{\mathrm{R}}(N)}{\log t'_{\mathrm{R}}(N+1) - \log t'_{\mathrm{R}}(N)} \qquad (1)$$

where RI refers to the retention index of compound "an", N refers to the number of carbon atoms in the normal alkane eluting immediately before the analyte and $N+1$ refers to the normal alkane eluting immediately after the analyte; t'_{R} refers to the adjusted retention time. This was modified for linear temperature programs (Equation 2):

$$\mathrm{RI} = 100N + 100 \ \frac{t_{\mathrm{R}}(\mathrm{an}) - t_{\mathrm{R}}(N)}{t_{\mathrm{R}}(N+1) - t_{\mathrm{R}}(N)} \qquad (2)$$

This form of relative retention calculation results in a retention index that allows retention to be reported free of minor instrumental variations and variation in column dimensions. The Kovats indices also provide retention values that are directly proportional with the free energy of the bulk solute–stationary interactions. Thus, direct retention comparisons can be made between columns with differing dimensions and liquid phases. Tables of retention indexes have been published.[8] Several other authors[9–12] have developed alternative retention index systems, but the Kovats system remains most widely used.

Using the retention indexes of selected compounds (test probes) Rohrschneider[13] and McReynolds[14] developed schemes for assessing the chemical characteristics of GC stationary phases. A summary of the test probes used by McReynolds and Rohrschneider is shown in Table 2, along with the chemistry being measured by each test probe. Both McReynolds and Rohrschneider chose squalane, the liver oil of sharks, a nonpolar mixture of long chain, highly branched hydrocarbons, and one of the classical GC liquid phases, as a standard liquid phase. First, they measured the RI of each test probe on squalane. Next, for other phases, they measured the RI under identical conditions. The McReynolds constant for the new phase is given by the difference in RI for the test probe on the new phase versus squalane. For example, a high McReynolds constant for benzene indicates a column that will strongly retain compounds that are likely to undergo induced dipole interactions, such as other aromatics and olefins. Table 3 shows McReynolds constants for some common stationary phases. Today, column vendors and others who report stationary phase characterization data still report Rohrschneider–McReynolds constants.[15] The structures and chemistries of these stationary phases are defined in more detail in sections 3.6 for packed columns and 4.4 for capillary columns.

Table 2 Summary of phase constants used by McReynolds (M) and Rohrschneider (R)

Compound	Purpose
Benzene (M, R)	Measures induced dipole attraction (aromatics and olefins)
n-Butanol (M, R)	Measures electron attracting groups (alcohols, nitriles, acids, halogens)
2-Pentanone (M, R)	Measures electron repelling groups (ketones, ethers, aldehydes, esters)
Nitropropane (M, R)	Measures ability to form complexes
Pyridine (M, R)	Measures ability to form complexes
2-Methyl-2-pentanol (M)	Substituted alcohol – interactions similar to *n*-butanol
Iodobutane (M)	Polar alkane interactions
2-Octyne (M)	Unsaturated hydrocarbon interaction similar to benzene
1,4-Dioxane (M)	Proton acceptor
Cis-hydrindane (M)	Dispersion

2.2 Important Separation Variables

As can be noted from the above discussion of McReynolds constants, there are many variables related to retention and flow in columns that are important in any discussion of stationary phase characteristics. Thermodynamics of the phase transfer between mobile and stationary phases mostly determines the retention time of any analyte. This is most easily expressed as the standard free energy change, ΔG° of the partitioning process between the mobile and stationary phases. The partition coefficient,

Table 3 McReynolds constants for common stationary phases. (Reprinted with permission from H. McNair and J. Miller, *Basic Gas Chromatography*, New York, John Wiley and Sons, 1998, 64.)

Liquid phase	Benzene	*n*-butanol	2-pentanone	Nitropropane	Pyridine
Squalane	0	0	0	0	0
OV-1	16	55	44	65	42
OV-101	17	57	45	67	43
Dexsil-300	41	83	117	154	126
OV-17	119	158	162	243	202
QF-1	144	233	355	463	305
OV-210	146	238	358	468	310
OV-225	228	369	338	492	386
Carbowax 20M®	322	536	368	572	510
DEGS	492	733	581	833	791
OV-275	629	872	763	1106	849

DEGS, diethyl glycol succinate.

K_p is given by Equation (3):

$$K_p = e^{-\Delta G^\circ/RT} \tag{3}$$

Which leads to the retention factor (Equation 4):

$$k = \frac{K}{\beta} \tag{4}$$

And to the retention time (Equation 5):

$$t_R = t_m(1 + k) \tag{5}$$

The value of the standard free energy is given by the specific stationary phase–analyte combination. This derivation is reviewed in more detail by Snow[16] and forms the basis for the many efforts to predict retention times from thermodynamic data.[17–19] These thermodynamic arguments are the same for both packed and capillary columns, with the major difference being that phase ratios (β), the ratio of gas phase volume to liquid phase volume within the column, for capillary columns are generally 1–2 orders of magnitude larger then for packed columns.

Not only does the identity of the stationary phase determine the retention time, but the dimensions of the column have a profound effect on the retention time and the peak width, and ultimately the chromatographic resolution of closely eluting peaks. A summary of the important variables and their effects is given in Table 4. The master resolution Equation (6), as described by Purnell[20] also provides a useful summary, as high resolution is the goal of most chromatographic method development.

$$R_s = \left(\frac{k}{1+k}\right)\left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\sqrt{N}}{4}\right) \tag{6}$$

Terms relating to retention $k/(1 + k)$, selectivity $(\alpha - 1/\alpha)$ and system efficiency $(\sqrt{N}/4)$ are seen to each contribute to the overall resolution. Two peaks with $R_s = 1.5$ are said to be ''baseline resolved''. According to this equation, for the largest resolution to be achieved in the shortest time, k, which is highly temperature dependent, should vary in the relatively tight range of 2–20, with little return seen in raising k above 20. N and α will vary greatly according to the column type, dimensions and stationary phase. Using the master resolution equation, van Deemter and Golay equations, packed and capillary columns can be examined separately and compared.

Table 4 Summary of important column-related variables and their effects on GC retention times and peak widths

Variable	Effect on retention	Effect on peak width
Column length	Retention times longer; peaks are separated further	Peaks broaden with length, but N higher
Column internal diameter	None if linear gas velocity the same	Narrower column; sharper peaks
Particle diameter (packed columns)	Pressure drop increased with smaller particles; retention time longer at same inlet pressure	Smaller particles; sharper peaks
Liquid phase film thickness	Thicker film; longer retention	Thinner film; sharper peaks
Column temperature	Increased temperature; decreased retention time	Increased temperature; sharper peaks; effect on N difficult to define; check maximum and minimum temperature before using column

2.3 Packed Columns

As discussed in section 3 of this chapter, the length of a packed column is limited by the inlet pressure that can be practically be generated by the gas chromatograph, leaving typical lengths in the 6–12 foot range. Limiting the column length also limits the number of theoretical plates, with typical values for a 6.3 mm (1/4 inch) outside-diameter column being 6000–10 000 theoretical plates. The master resolution equation (Equation 6) can be used to describe the requirements for a column. If a desired resolution of 1.5 is assumed and that "normal" chromatographic conditions include k from 1–20 and N from 1000 to 10 000, the selectivity required to achieve a resolution of 1.5 is easily calculated by rearranging the master resolution equation. Table 5 shows the results of such a calculation using the master resolution equation, solved for selectivity. It is seen that selectivity greater than 1.064 is needed to achieve baseline resolution in the best case for packed columns, while for most packed columns, the necessary selectivity is higher. Since selectivity in GC is entirely based upon the nature of the solute–stationary phase interactions, the choice of liquid phase is critical in the development of nearly every method. This has led to the development of a very wide variety of liquid phases for packed columns, which are treated in more detail in sections 3.5 and 3.6.

Peak widths in packed columns are related to the height equivalent to a theoretical plate (HETP or H), which is described by the van Deemter equation (Equation 7):[21]

$$H = 2\lambda d_p + \frac{2\gamma D_M}{\mu} + \frac{qkd_f^2\mu}{(k+1)^2 D_s} \qquad (7)$$

The variables are described in the list of symbols at the end of this article. Many of the terms in this equation are related to the nature and dimensions of the liquid phase. The diameter of the stationary phase particles affects eddy diffusion in the first term, with larger particles generating larger H. Mass transfer in the stationary phase is affected by the diffusivity of the solute, indicating that less viscous liquid phases will provide lower H, as will a thinner film of liquid (smaller loading). The film thickness has the most dramatic effect, as it is squared.

Table 5 Selectivity requirements for packed and capillary GC columns, calculated from the master resolution equation

k	1000	2000	5000	10 000	50 000	100 000	250 000
1	1.612	1.367	1.204	1.136	1.057	1.039	1.025
2	1.398	1.252	1.146	1.079	1.042	1.029	1.018
5	1.295	1.192	1.113	1.078	1.033	1.023	1.015
10	1.263	1.137	1.103	1.071	1.030	1.021	1.013
15	1.254	1.132	1.100	1.068	1.029	1.021	1.013
20	1.250	1.130	1.098	1.067	1.029	1.020	1.013
∞	1.234	1.154	1.093	1.064	1.027	1.019	1.012

2.4 Capillary Columns

Due to their length (10–100 m), capillary columns have much higher plate numbers than packed columns. In terms of generating resolution, therefore, the necessary selectivity for capillary columns is significantly less than for packed columns. Table 5, above, also shows the calculation of the necessary selectivity required to achieve a resolution of 1.5, for capillary columns, assuming retention factors between 1 and 20 and plate numbers between 10 000 and 500 000, typical values for capillary columns. It is seen that lower selectivity values, 1.012 (compared with 1.064 for the packed column described above) for $k = 20$ and $N = 250\,000$, are needed than for packed columns. Thus, in the initial developments, there were far fewer liquid phases available for capillary columns than for packed columns, as stationary phase selectivity is less critical in the achievement of separation.

The Golay equation[2] (Equation 8) can be used to describe HETP for capillary columns.

$$H = 2D_M\mu + \frac{1 + 6k + 11k^2}{24(1+k)^2}\frac{r_c^2}{D_M} + \frac{2kd_f 2\mu}{3(1+k)^2 D_s} \qquad (8)$$

It differs from the van Deemter equation in that, since a capillary column is an open tube, there is no term for eddy diffusion and since there is a much greater relative amount of mobile phase in a capillary, there is an additional mass transfer term for the mobile phase. The major variables that are affected by the liquid phase are the diffusivity of the solute in the liquid phase, indicating that low viscosity liquid phases are desirable, as are thin films. Again, the film thickness has the most dramatic liquid phase-related effect on H and therefore, on the peak width.

2.5 Comparison of Packed and Capillary Column Separations

Chromatograms showing packed and capillary column separations of a polychlorinated biphenyl standard that demonstrate the characteristics of each resulting from the above theory are shown in Figures 1 and 2.[22] About 15 peaks are shown in the packed column separation, which demonstrates about 1500 theoretical plates. The capillary column separation shows about 60 peaks and about 100 000 theoretical plates. Both separations were conducted isothermally, with the same liquid phase. Thus, with the same selectivity, the capillary column provides sharper peaks with much greater resolution in a similar analysis time. The greater mass of stationary phase in the packed column also requires higher elution temperatures and a slower analysis time, with much greater consumption of carrier gas and lower sensitivity.

The selectivity factors, as described above and calculated in Table 5 have a dramatic effect on the separation power of a column. For example, if $k = 5$ and

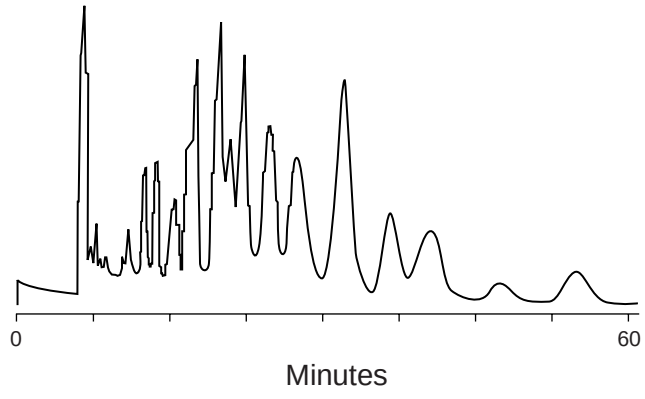

Figure 1 Packed column GC separation of a polychlorinated biphenyl standard Arochlor® 1260; isothermal at 210 °C; 1500 theoretical plates. (Reprinted with permission from H. McNair and J. Miller, *Basic Gas Chromatography*, New York, John Wiley and Sons, 1998, 94.)

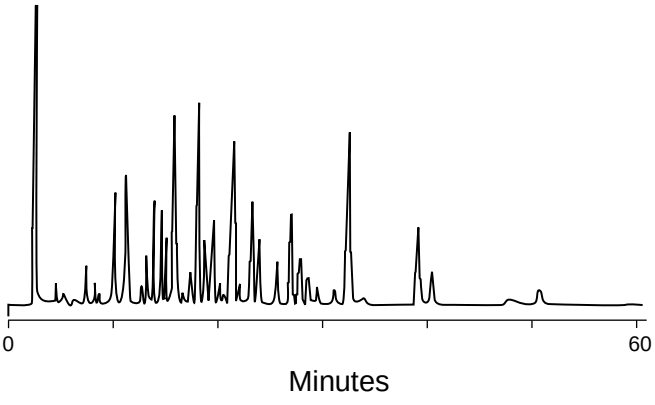

Figure 2 Capillary column GC separation of a polychlorinated biphenyl standard Arochlor® 1260; split 30 to 1; isothermal at 210 °C. (Reprinted with permission from H. McNair and J. Miller, *Basic Gas Chromatography*, New York, John Wiley and Sons, 1998, 94.)

$t_R = 10.0$ min is assumed for a typical peak, then the next eluting peak must have a retention time of at least 10.616 min for baseline resolution to be achieved on a packed column having 10 000 theoretical plates. For a capillary column having 250 000 theoretical plates, under the same conditions, the required retention time for the next peak is 10.120 min. This indicates that the capillary column can separate much more complex mixtures than a packed column, as many more peaks can fit within the separation time, for example 20 min. Peak capacity of columns is described in more detail by Giddings.[23]

3 PACKED COLUMNS

Most of the early work in GC was performed on packed columns. Following the invention of the capillary column, they proved difficult to make and use, so work in GC continued with packed columns. Today, about 80% of all work in GC is done with capillary columns;[4] however, packed columns are still the method of choice for those analyses requiring the relatively large volume of liquid phase present in packed columns, such as fixed gases, solvents and other analyses in which partition coefficients are low, yet strong retention is required.

3.1 Tubing Materials and Dimensions

Typically, packed columns are 6.3 or 3.2 mm (1/4 or 1/8 inch) in outside diameter and 0.6–3.6 m (2–12 feet) long. Glass and stainless steel are the most common tubing materials, although nickel, Teflon®, copper, Teflon®-lined stainless steel and aluminum are also used. Copper and aluminum columns can be easily made in the laboratory from straight tubes, which can be bent to fit the instrument configuration, while the other materials must be purchased in the correct configuration for the instrument into which they will be installed. Figure 3 provides a summary of some common configurations used by major instrument manufacturers.

3.2 Solid Supports

In a packed column, the liquid phase must be coated onto a solid support material. Ideally, the support material should be capable of withstanding the physical stresses in GC, including high temperatures (up to 350 °C, generally) and moderate pressures (up to 5 atm), should be inert and not react with sample components or liquid phase materials and should be readily available in a variety of mesh sizes. The important physical properties include particle size, porosity, surface area and packing density. Smaller particles will provide more plates, as shown by the van Deemter equation, yet will require a higher

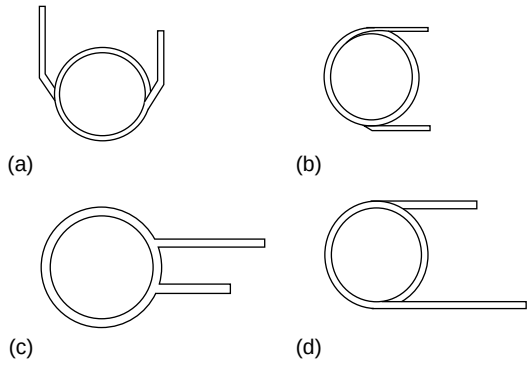

Figure 3 Packed column tubing configurations (a) Hewlett Packard 5890, 5880, 6890; (b) Perkin Elmer 8300, 8400, 8500, 8600, 8700, auto system; (c) Shimadzu GC14A, GC15A, GC16A; (d) Varian 3300, 3400, 3700, Vista Series.

pressure drop, necessitating a shorter column. A larger surface area, which is determined by the porosity, will require more liquid phase to achieve a complete coating. Diatomite supports, graphitized carbon blacks and, to a lesser extent, Teflon® are commonly used. Each of these has advantages and limitations, which are governed by their physical properties and by their surface chemistry.

Diatomite supports are all derived from the various forms of diatomaceous earth. Firebrick supports are obtained by heating, grinding and pulverizing bricks; filter aid supports are obtained by adding sodium carbonate to diatomaceous earth and heating to obtain an aggregate that is generally more inert than firebrick; and Chromosorb G is a filter aid support that has been additionally heat treated to generate more durable particles. Each of these may be further treated with a silylating reagent such as dichlorodimethyl silane (DMDCS), to cover active silanol functionalities on the surface, or may be acid or base washed to remove any active metal contaminants. It should be noted that silanized supports are very sensitive to hydrolysis by water, so care should be taken to eliminate water from any methods where they are used. Acid or base washing is used to eliminate reactive metals such as iron from the surface. These can act as Lewis acids and participate in the decomposition of liquid phases and analytes. Table 6[24] provides a reference to the common solid supports used for GLC.

The firebrick supports Chromosorb P and Gas Chrom R are pink in color and are very strong particulates, with dimensions as shown in Table 6. These have very high surface area and can accommodate liquid phase loading of up to 30%. As their surfaces contain active silanols, they are typically reserved for the analysis of nonpolar analytes, although they can be deactivated for analysis of polar compounds. The white calcinated supports, Chromosorb W, Gas Chrom Q and Suplecoport, for example, are preferred for polar analytes, although they are more fragile and accept slightly lower liquid loading (ca. 25%). Chromosorb G is denser and more durable than Chromosorb W, but can only accept about 5% loading of liquid phase.

Other materials, such as carbon black and Teflon® have been used as solid supports for gas liquid chromatography. Teflon® is typically used for the analysis of small polar molecules, such as water, acids, amines, alcohols, and other compounds that may be corrosive to diatomite supports. Teflon® has a relatively low upper temperature limit of about 250 °C and may develop static charges, making packing difficult. Carbon black is used as both an adsorbent and as a solid support. As a solid support, it allows for a combination of separation due to the liquid phase and the surface properties of the support, which separate according to the size and shape of the molecule. These are typically used for difficult separations of small molecules.

3.3 Packing and Preparing Packed Columns

Packed columns may either be purchased pre-made from a vendor, or they may be prepared in the laboratory from the appropriate length and diameter of tubing and the bulk packing material. Most common liquid phases are available precoated onto the solid support. When coating a liquid phase onto a solid support, first, the appropriate mass must be determined from the desired loading. There are three common methods for coating liquid phases onto solid supports:

1. Rotating Evaporator Method. The appropriate mass of liquid phase is dissolved in a suitable solvent in a round-bottomed flask and the weighed solid support is added. The solvent is then evaporated by rotary evaporation at reduced pressure.
2. Pan Coating Method. The liquid phase is dissolved in just enough solvent to wet the solid support and the

Table 6 Characteristics of selected solid supports for packed column GLC. (Reprinted with permission from E. Barry, *Modern Practice of Gas Chromatography*, ed. R. Grob, 2nd edition, New York, John Wiley and Sons, 131, 1995.)

Source	Support	Packing density ($g\,mL^{-1}$)	Surface area ($m^2\,g^{-1}$)	Pore volume ($mL\,g^{-1}$)	Maximum liquid phase loading (%)
Firebrick	Chromosorb P, NAW, AW/DMDCS, AW	0.32–0.38	4–6	1.60	30
Celite Filter Aid	Chromosorb W AW	0.21–0.27	1.0–3.5	3.56	15
Celite Filter Aid	Chromosorb W HP	0.23	0.6–1.3	3.56	15
Celite Filter Aid	Chromosorb G NAW, AW/DMDCS, HP	0.49	0.5, 0.4(HP)	0.92	5

AW, acid washed; NAW, not acid washed.

solution is added to the solid in a pan. The solvent is allowed to evaporate with gentle agitation and heat.

3. Funnel Coating Method. 20 g of solid support is added to 100 mL of liquid phase solution in a filter flask. The flask is placed under vacuum for a few minutes. The pressure is released and the mixture is placed in a sintered glass funnel and vacuum is applied for several minutes. The resulting solid is allowed to air dry.

More details on the coating of solid supports can be found in Barry.[24]

Glass and stainless steel tubing comes pre-shaped to fit into the instrument's fittings; copper and aluminum tubing can be bent into the appropriate shape. The pre-shaped tubing is available for a wide variety of vendors' instruments and column lengths. To pack the stationary phase into a 6.3 or 3.2 mm (1/4 or 1/8 inch) outside-diameter tube, a funnel, with a short piece of Tygon® tubing, a small amount of glass wool and an electric engraving tool are needed. Plug one end of the tubing loosely with glass wool, and using the funnel, connected to the other end by the Tygon® tubing, add the stationary phase. Vibrate the tubing and add stationary phase until no more can be added. Plug the open end of the column with a small piece of glass wool, leaving the appropriate length for the GC inlet and syringe needle unpacked. Following packing, the column should be installed into the instrument, carrier gas flow established and should be conditioned for several hours at a temperature 25 °C above the maximum temperature at which it will be used, but below the maximum temperature of the liquid phase.

3.4 Instrumentation Requirements

There are several instrumental considerations in the use of packed columns. First, the flow systems, inlets and detectors should be able to accommodate high carrier gas flow of 50–100 mL min^{-1}. On-column, direct, or gas valve injection are most commonly performed on packed columns, as syringe needles will easily fit within 6.3 or 3.2 mm (1/4 or 1/8 inch) tubing, and valve pneumatics are easily connected. This does provide the advantage of quantitative sample transfer into the column, without the use of splitters or special syringes, which are required for capillary columns. Packed columns have high thermal mass, requiring a very powerful oven heater and cooler if temperature programming is desired. This high thermal mass makes isothermal operation most common with packed columns. Due to the high flow rates, detector cells also are large and have high thermal mass, to maintain temperature equilibrium. In packed column gas chromatography/mass spectrometry (GC/MS), splitting devices must be used between the column and the detector, in order to reduce the flow entering the mass spectrometer sufficiently. Instrumentation for packed column GC is described in detail in **Instrumentation of Gas Chromatography** and in McNair[25] and Miller.[26]

3.5 Choosing a Stationary Phase

As shown in section 2 of this article, the relatively low plate counts of packed columns make the choice of liquid phase critical in packed column gas liquid chromatography. As a result, hundreds of liquid phases have been used over the years. A 1968 survey[27] indicated 179 commonly used liquid phases. A 1997 vendor catalog[28] lists 138, so a dizzying array of stationary phases remains available. Practically, it is not advisable to attempt to choose between all of these liquid phases; it is desirable to have a few choices that will solve most separation problems. The old maxim ''like dissolves like'' can provide a good start, as, from the discussion of stationary phase characterization presented above, it is difficult to develop a systematic scheme for choosing packed column liquid phases. Consulting the literature or the vendors' application notes is also a good place to start. Recently, Supelco has made a compendium containing over 200 of the packed column separations available.[29]

If consulting the literature does not provide a reasonable choice, then the McReynolds constants provide a starting point, as they measure the likely selectivity of the liquid phase for compounds similar to the test probes. For example, a liquid phase with a high value for the constant for *n*-butanol, would be a good choice for separating alcohols or other compounds expected to undergo hydrogen bonding. McNair and Miller[30] have provided a general summary of recommended stationary phases compiled by several workers, which is shown in Table 7. It is seen that these cover a wide range of polarities, and with appropriate temperature control, can achieve most separations. A nonpolar column is always included, as are several of moderate polarity, along with a highly polar silicone and

Table 7 Recommended stationary phases for method development and general use in packed and capillary column GC. (Reprinted with permission from H. McNair and J. Miller, *Basic Gas Chromatography*, New York, John Wiley and Sons, 75, 1998.)

Hawkes[31]	Yancey[32]	McNair[33]
OV-101	OV-101	OV-1
OV-17	OV-17	OV-17
Carbowax® >= 4000	Carbowax 20M®	Carbowax 20M®
OV-210	OV-202	OV-210
DEGS	OV-225	OV-275
Silar 10C		

Table 8 Recommended stationary phases and applications for packed and capillary column GC. (Reprinted with permission from E. Barry, *Modern Practice of Gas Chromatography*, ed. R. Grob, 2nd edition, New York, John Wiley and Sons, 213, 1995.)

Stationary phase	Temperature range	Application
100% methyl polysiloxane	−60 °C to 350 °C	Alkaloids, amines, drugs, FAME, hydrocarbons, petroleum products, phenols, solvents, waxes, general purposes
5% phenyl–95% dimethyl polysiloxane	−60 °C to 350 °C	Alcohols, alkaloids, aromatic hydrocarbons, flavors, fuels, halogenates, herbicides, pesticides, petroleum products, solvents, waxes, general purposes
50% phenyl–50% methyl polysiloxane	−60 °C to 350 °C	Alcohols, drugs, herbicides, pesticides, phenols, steroids, sugars
14% cyanopropylmethyl–86% dimethyl polysiloxane	0 °C to 250 °C	Alcohols, arochlors, alcohol acetates, drugs fragrances, pesticides
50% cyanopropylmethyl–50% phenyl polysiloxane	0 °C to 250 °C	Carbohydrates, FAME
Trifluoropropyl polysiloxane	0 °C to 275 °C	Drugs, environmental samples, ketones, nitro-aromatics
Polyethylene glycol	60 °C to 225 °C	Alcohols, flavors, fragrances, FAME, amines, acids

FAME, fatty acid methyl esters.

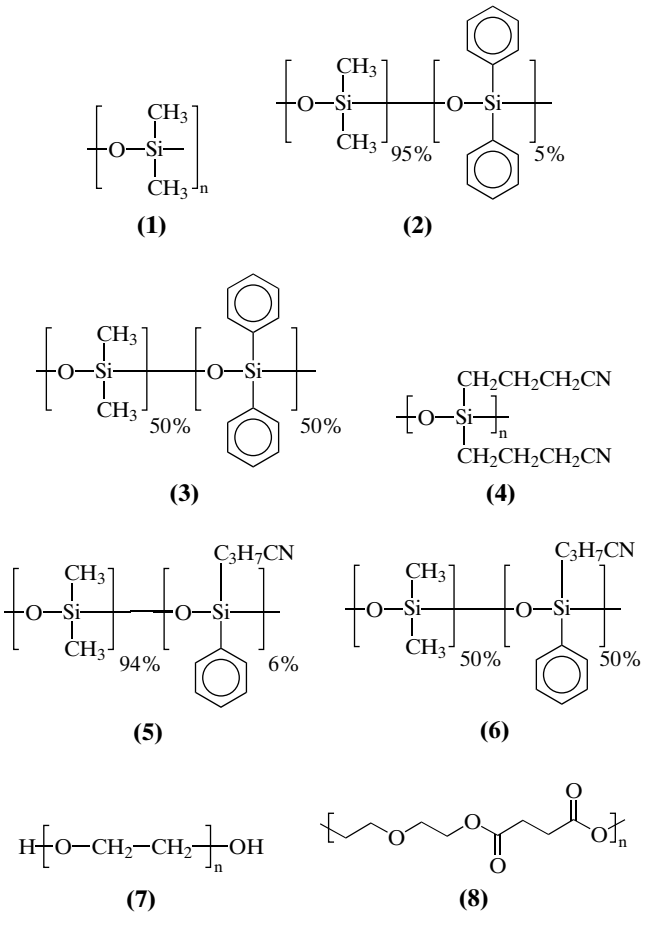

a polyglycol. Structures 1–8 show the following liquid phases: 100% polydimethyl siloxane **(1)**; 5% phenyl, 95% polydimethyl siloxane **(2)**; 50% phenyl, 50% polydimethyl siloxane **(3)**; 100% cyanopropyl polysiloxane **(4)**; 6% cyanopropyl, 94% dimethyl polysiloxane **(5)**; 50% cyanopropyl, 50% dimethyl polysiloxane **(6)**; polyethylene glycol **(7)**; diethyl glycol succinate (DEGS) **(8)**.

Finally, once a potential liquid phase is chosen, the upper and lower temperature limits must be considered. At the upper temperature, the stationary phase itself has significant vapor pressure and will bleed from the column, reducing efficiency and shortening column lifetime. At the lower temperature, which may be the freezing point or glass transition temperature of the polymer, separation is also lost, as analytes no longer penetrate into the solid stationary phase.

3.6 Guide to Packed Column Gas Chromatography Liquid Phases

A few of the most commonly used packed column stationary phases are described above. There are, however, myriad stationary phases available. Table 8,[24] while not all-inclusive, provides a more extensive listing of common stationary phase materials and applications.

4 CAPILLARY COLUMNS

Capillary columns were first developed by Golay in 1957,[2] shortly after the initial development of GC. Theory predicts that the narrow, open-tubular format should provide very efficient transport of analyte molecules along the column length, and therefore, very sharp peaks. The initial glass capillary columns saw little routine use, however, as the glass columns had to be hand-drawn from larger tubing using an expensive drawing apparatus, and once made, were very fragile. In order to install a typical, helical glass column into an instrument, the glass first had to be straightened by heating, so that the appropriate compression fittings could be installed. Capillary GC has flourished since the development of the fused silica capillary column by Dandeneau and colleagues in 1979.[3] This development provided capillary columns that were

easy to handle and use. Other materials, such as stainless steel, aluminum cladded fused silica and Teflon® have also been used to make capillary columns for specific applications. The specific properties of these columns, as they relate to the liquid phase contained within them and to the development of separations, are discussed below. This section will also address the selection of appropriate capillary columns and stationary phases, along with a guide to stationary phase selection.

4.1 Design and Manufacture

Fused silica capillary columns are manufactured using optical capillaries with internal diameters commonly ranging from 0.100 mm to 0.530 mm, and with column lengths from 10–100 m. Smaller internal diameters are used for a few high-speed applications. The tubing is usually prepared by drawing a long length (hundreds of meters) using a drawing machine, which is then cut to the desired length after coating with the stationary phase. The fused silica tubing is coated on the outside with polyimide that provides flexibility and tensile strength, along with the characteristic brown color. Users of capillary columns should be warned to wear safety glasses at all times when handling columns that are coiled, as the coils can unwind with such force that a column end can damage the eyes. Fused silica capillary tubing is prepared by extruding larger diameter tubing in a furnace, with the drawing process controlled by an infared laser beam. Following drawing, the tubing is coated with a polyimide that provides flexibility, protection against water degradation and corrosion. The polyimide coating is stable to approximately 400 °C. For applications requiring temperatures higher than 400 °C, aluminum clad columns, in which the outer polyimide coating is replaced with a thin (ca. 20 μm) coating of aluminum can be used. These are commonly used in the separations of waxes, oils and triglycerides. Stainless steel has also been used as an outer coating for high-temperature work. The main advantage of stainless steel coating is that the column can be coiled very tightly, allowing it to be used in a small space.

Once the capillary tubing has been manufactured, it must then be prepared for coating, and coated with a thin (0.1–5 micron) film of liquid phase. The main requirements are the same as for packed columns: the column wall should be inert to the analytes and should allow the liquid phase to coat evenly. Classical coating techniques have involved the simple coating of polymeric liquids to the inside walls of the capillaries. More recently, the polymer has been chemically bound to the capillary wall. Prior to coating, the fused silica tubing must be deactivated, typically by reaction with silylating reagents. In 1985, Blomberg and Markides[34] presented a thorough discussion of organo-silicon chemistry for capillary columns, including a discussion of pretreatments. The techniques include D4-deactivation, in which the capillary is coated with octamethylcyclotetrasiloxane (D4), sealed and heated to 400 °C,[35] and the capillary heating with an appropriate polysiloxane,[36] high temperature silylation using hexamethyl disiloxane (HMDS),[37] and reaction with disiloxanes.[38]

Once the capillary inside surface has been deactivated, it must then be coated with the liquid phase polymer. A stable, thin (0.1–5 micron) and uniform coating is needed. Coating procedures were reviewed by Jennings,[39] with the most common method being static coating. In static coating, first described by Bouche and Verzele,[40] a solution of known concentration of the liquid phase is prepared in a volatile solvent. One end of the column is sealed and a vacuum is applied to the other end. Evaporation of the solvent leads to the deposition of a uniform coating of the less volatile liquid phase on the column wall. This method allows an accurate determination of the volume of stationary phase added to the column, and therefore the film thickness. Newly prepared capillary columns are then coiled into a circular metal cage to protect the column and to simplify mounting into the instrument oven. The chromatographic performance of a newly produced column can be evaluated using a column test mixture, as described later in this article.

4.2 Instrumentation Requirements

The specifics of GC instrumentation are discussed elsewhere in this Encyclopedia, however, there are several considerations that are directly related to the low carrier gas flow rates, small quantity of stationary phase, and open tubular nature of capillary columns. First, as shown by the Golay equation, the carrier gas flow rate is critical to the optimum performance of capillary columns, and should be adjusted carefully. Second, the small quantity of stationary phase, requires that analyte mass also be small, typically of the order of nanograms–picograms. Higher analyte quantities can overload the stationary phase, causing multiple retention mechanisms and distorted peaks. The open tubular nature of capillary columns indicates that the inlet–outlet pressure drop will be significantly less than for a packed column of the same length, allowing for very long (100 m is common) columns and high plate numbers. Column oven temperature control is also critical to optimum performance, as capillary columns, due to their low thermal mass compared to packed columns, are much more easily and rapidly temperature programmed. Thus the column oven itself should be of low thermal mass and be thermally insulated from other heated portions of the instrument. Most modern GC

ovens employ forced air convection for rapid heating and cooling of the capillary column. More details on instrumentation for capillary columns are provided in **Instrumentation of Gas Chromatography** and in several references.[41,42]

4.3 Choosing a Liquid Phase

Selecting a stationary phase for capillary GC is done using a similar thought process as for packed columns, but there are far fewer liquid phase choices readily available. First, there is a wide variety of literature available in both the research literature and in vendors' publications; this is an excellent place to start. One vendor provides a compact disk[43] compilation of over 1300 application notes, while several have application notes posted to searchable pages on the World Wide Web.[44–48] If the literature fails to produce a satisfactory result, then the ''like dissolves like'' principle may apply to the stationary phase choice.

In capillary GC there are several liquid phases that are most commonly used, plus myriad phases for specialty applications. Barry[49] provides a summary of general applications for the most common stationary phases, which is shown in Table 8. Chemical structures for the common stationary phases used in capillary GC are given in section 3.5. Since capillary columns generally have much higher plate counts than packed columns, the stationary phase choice can be much simpler. For most applications, a nonpolar phase such as polydimethyl siloxane (OV-1) or 5% phenyl polydimethyl siloxane (SE-54) can be used. More polar columns are used for specific applications, as shown in Table 8.

As the initial tool in a method development scheme, a moderate length (15–30 m) capillary column with a nonpolar stationary phase, such as polydimethyl siloxane or 5% phenyl polydimethyl siloxane, with internal diameter 0.2 or 0.25 mm and film thickness of 0.2–0.5 micron is recommended. This represents a compromise between analysis speed, high resolution and sample capacity. Once the temperature and flow conditions for a separation have been optimized as much as possible on this column, an alternative stationary phase can be selected, if the separation is not satisfactory. Klee[50] and Klee and Kaiser[51] provide systematic methods for selecting and optimizing capillary columns.

4.4 Guide to Stationary Phases for Capillary Gas Chromatography

Perhaps the most confusing part of stationary phase selection for capillary GC is that the many vendors use different designations for stationary phases that are essentially similar. A listing of the most common capillary GC stationary phases, obtained from a survey of vendors' catalogs is given in Table 9, along with a cross-reference of major vendors' designations.

Table 9 Cross-reference of common stationary phases for GLC (courtesy of Professor Harold M. McNair)

Stationary phase	Restek	J & W	Supelco	HP	SGE	Chrompack	Quadrex
100% Polydimethyl siloxane	Rtx-1	DB-1	SPB-1, SP-2100	HP-1, Ultra-1	BP-1	CP-Sil 5 CB	007-1
95% dimethyl–5% phenyl polysiloxane	Rtx-5, XTI-5	DB-5	SPB-5	HP-5, Ultra-2	BP-5	CP Sil 8 CB	007-2
80% dimethyl–20% phenyl polysiloxane	Rtx-20		SPB-20				007-7
65% dimethyl–35% phenyl polysiloxane	Rtx-35						007-11
14% cyanopropyl phenyl–86% dimethyl polysiloxane	Rtx-1701	DB-1701			BP-10	CP Sil 19 CB	007-1701
50% methyl–50% phenyl polysiloxane	Rtx-50	DB-17	SP-2250	HP-17		CP Sil 43 CB	007-17
Trifluoropropyl-methyl polysiloxane	Rtx-200	DB-210					
50% cyanopropyl phenyl–50% dimethyl polysiloxane	Rtx-225	DB-225	SP-2300	HP-225	BP-225		007-225
Carbowax® PEG	STABIL WAX	DB-WAX	SUPELCO WAX-10	HP-20M	BP-20	CP Wax 52 CB	007-CW
Carbowax® PEG for amines	STABIL WAX-DB	CAM					
Carbowax® PEG for acids	STABIL WAX-DA	DB-FFAP	NUKOL, SP-1000	HP-FFAP	BP-21	CP Wax 58 CB	FFAP

5 CARE AND MAINTENANCE OF GAS CHROMATOGRAPHY COLUMNS

In both packed and capillary column GC, increasing tailing of one or more of the peaks in an analytical run, over several separations, most likely indicates degraded column performance. Depending on the types of samples being analyzed, a column may last for as little as 50–100 injections, or for several thousand. In any case, there are some general considerations in extending column lifetime, which are discussed here. For a more thorough treatment, see the monograph by Rood[52] for capillary GC and see the troubleshooting guide presented by Supelco[53] for packed columns.

5.1 Packed Columns

In general, packed columns are very rugged and easy to maintain. This is a result of the large mass of liquid phase that is present. Packed columns are more resistant to dirty solvents, or corrosive solvents, such as water or wet samples. Packed columns should be stored with the ends capped, to prevent oxidative damage over time. When they are installed into the GC, care should be taken not to overtighten the compression fittings, or leaks can result. When installed into the instrument, a gentle flow of carrier gas should be maintained at all times, to avoid oxidative damage to the stationary phase. When the liquid phase is becoming degraded from use, tailing peaks usually result. It is good practice to maintain a regular schedule of analyzing a known standard of an active analyte and to keep a record of the results. When this peak shows significant distortion over a period of several trials, column maintenance is necessary. In a glass column, discoloration of the first few centimeters of the stationary phase can sometimes be seen.

Usually, contaminants, such as nonvolatile material or contaminants irreversibly adsorbed onto the liquid phase, are found within the first few inches of the column. It is acceptable to carefully remove the glass wool plug, pour out the first few centimeters of stationary phase and to replace it with fresh packing. Tap the column gently on the benchtop or use a vibrating engraving tool to ensure even packing. Replace the glass wool plug with fresh glass wool, as this is a common source of contamination, and condition the column, as described for a new column. If the peak distortion persists, then the entire column packing needs replacement.

5.2 Capillary Columns

Capillary columns require more careful handling than packed columns, as there is a much smaller quantity of stationary phase present. In general, they should be stored with the ends sealed, to avoid oxidation of the stationary phase. Inserting the ends into a used septum makes an excellent seal, or they can be flame sealed. *Wear safety glasses at all times when handling capillary columns, as the glass capillary ends can spring out and damage the eyes upon contact*. Carrier gases should be free of organic contaminants and water, as these can also foul and contaminate the column. Typically, this requires a minimum of a molecular sieve trap to remove water and low molecular weight organic compounds and an activated carbon trap to remove higher molecular weight organic compounds. Heated gas purifier systems, available from several vendors, can also be used. The trapping material should be replaced or regenerated on a regular schedule. Carrier gases should also be high purity to avoid column and instrument contamination. Carrier gas flow should be maintained at all times that the column is installed into the gas chromatograph, to avoid oxidative damage to the stationary phase. It is good practice to leave the system running continuously, with carrier gas flowing and all zones heated to above ambient temperature. Most modern gas chromatographs have a ''gas saver'' feature to prevent excessive gas usage while samples are not running.

Tailing of peaks that were not previously tailing indicates column degradation. This diagnosis is more difficult in a capillary GC system than in a packed column system, as the column is typically separated from the inlet, where contamination is likely to occur. Thus, inlet contamination should be ruled out as the cause of tailing or distorted peaks, before column contamination is considered. As with packed columns, the contamination typically occurs within the first few centimeters of the capillary. If column contamination is suspected, the easiest treatment is to cut the first few centimeters from the column and discard it, and then to re-install the column into the gas chromatograph. Be careful to follow the manufacturer's directions for cutting and re-installing the column, or the contamination can be made worse. Usually, this cutting procedure can be applied several times, before too much cutting reduces column performance. When installing the column, take care to avoid sharp bends in the tubing and contact with the oven wall or wire cage hanger. These can cause stress points or hot spots that can lead to breakage, leaks or stationary phase degradation.

Before using a column, the minimum and maximum operating temperatures, which are usually listed in the accompanying literature, should be checked and the column should not be operated below the minimum or above the maximum temperature. Below the minimum temperature, the polymer may solidify or become glassy, preventing absorption of analyte molecules and adversely impacting retention. As the maximum temperature

is approached, the stationary phase itself begins to decompose, leading to vaporization of the decomposition products. These appear as a rise in the baseline, known as column bleed. At temperatures exceeding the maximum temperature, the column bleed becomes excessive and can damage the column or detector. Approximate maximum temperatures for common stationary phases are provided in Table 8.

The brown polyimide coating also deserves mention, as it gives the capillary column its tensile strength and flexibility; however, it, too can degrade over time. Usually, a darkening of the brown color indicates this. Beware that the column may become brittle and easily break at continuously heated zones, such as the inlet or detector. Special care should be taken if the column is placed into long transfer lines such as those used for GC/MS.

5.3 Quality Testing

There are several general methods for evaluating the performance of capillary columns. For both packed and capillary columns, perhaps the best method is to monitor one's own samples. Look for changes in the peak shape and area over time for a known standard. A column which may show poor performance in one application, such as acid analysis (acids are strongly adsorbed on capillary columns in which the deactivation of the silica tubing was poor) may be acceptable for another such as hydrocarbons. Today, most column vendors provide a test chromatogram with every column that provides some indication of the column's supposed ''quality''. These should not be used to compare columns from different vendors, as they are determined under widely varying temperature, flow and injection conditions, which can have dramatic effects on the peak shapes. Figure 4 shows a test chromatogram and data from a typical column vendor.[54]

In 1979, K. Grob et al. developed a standard quality test method for capillary columns, with an improved method described in 1981.[5,6] According to Grob, a quality test should contain enough components to provide the needed information, should provide complete information from a single chromatographic run, should be general for all stationary phases, provide some quantitative information and be standardized to allow column comparison. The Grob test is temperature programmed, and evaluates separation efficiency, adsorption of several types of components, and the actual column film thickness.

Separation efficiency can be measured by evaluating the Trennzahl[55] number (Equation 9):

$$\mathrm{TZ} = \frac{t_{\mathrm{R}}(2) - t_{\mathrm{R}}(1)}{W_{\mathrm{h}}(1) + W_{\mathrm{h}}(2)} - 1 \qquad (9)$$

for two homologs. This is essentially the difference between the retention times divided by the sum of peak

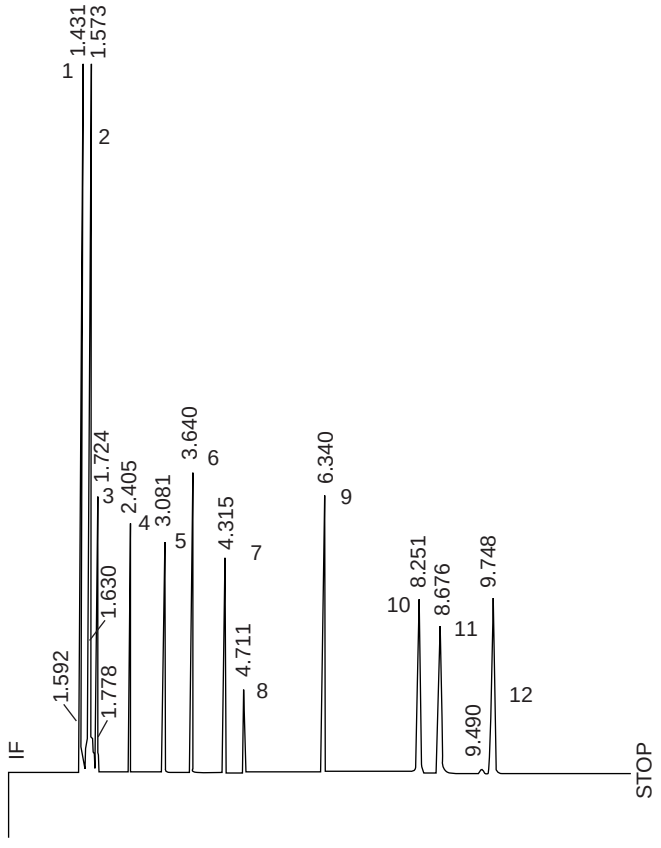

Figure 4 Typical column quality test chromatogram 1: pentane; 2: undecane; 3: 4-chlorophenol; 4: 1-decylamine; 5: tridecane; 6: methyl caprate; 7: tetradecane; 8: acenaphthylene; 9: 1-dodecanol; 10: pentadecane.

widths and it expresses roughly the number of peaks that could fit between two adjacent peaks on a chromatogram. A larger Treunzahl number indicates a more efficient column. It is noted that the classical efficiency measure, N, the number of theoretical plates is not easily applied to temperature programmed separations. The Trennzahl number is a measure of the number of peaks of width similar to the homologs that can fit between the homolog peaks.

There are several indications of excessive adsorption or mixed retention mechanisms. These include broadened Gaussian peaks, tailing peaks, reduced peak area and misshapen peaks with long retention. Usually, all of these result in reduced peak height, as compared to a normal Gaussian peak. An abnormally low peak area for a given component usually indicates irreversible adsorption somewhere in the system. Hydrogen bonding is characterized by peak tailing and can be examined using a primary alcohol. An aldehyde can provide an indication of excessive nonhydrogen bonding adsorption, while acidic and basic components can be used to study those interactions. Using these considerations, Grob developed

Table 10 Components of the Grob test mixtures for assessing the quality of capillary GC columns[5,6]

Compound	Purpose
Methyl decanoate	Determination of separation efficiency; second standard for 100% peak height used in adsorption calculations
Methyl undecanoate	
Methyl dodecanoate	
n-Decane	Standard for retention index and 100% peak height adsorption calculations
n-Undecane	
n-Dodecane	
1-Octanol	Hydrogen bonding adsorption on silicones
Nonanal	Non-hydrogen bonding adsorption
2,3-Butanediol	Hydrogen bonding adsorption on polyethylene glycols
2,6-Dimethylaniline	Acid base interactions
2,6-Dimethylphenol	Acid base interactions
Dicyclohexylamine	Acid base interactions (sterically hindered)
2-Ethylhexanoic acid	Acid base interactions

two test mixtures, for polar and nonpolar columns. The components of these mixtures and their purpose are given in Table 10. The introduction of the Grob test led to many future developments in column deactivation procedures. The specific procedures for using the Grob test mixtures for column evaluation and comparison are given in Grob.[5,6] The test mixtures used by the column vendors are typically variations of the Grob mixture.

6 RECENT AND FUTURE DEVELOPMENTS

Packed columns currently are used for a fairly stable 20% of GC analyses. Today, they are used for specific applications, such as gas analysis, solvent analysis, process control and education that benefit from the large volume of stationary phase, lower resolution and simple instrumentation. Majors,[56] who polled both the readership of ''LC-GC'' magazine and several recognized experts, recently surveyed this. Thus, much of the recent developments in GC stationary phases have been in the improvement of capillary column technology.

With the explosion in capillary GC use that occurred in the 1980s, and the development of quality tests, such as the Grob test mixture, many of the developments have led to better deactivation of the silica tubing and stationary phase materials themselves, improved temperature stability and to a much wider variety of stationary phases, many designed for specific applications, such as United States Environmental Protection Agency (USEPA) approved methods. A recent fundamental survey of GC[57] shows 76 references in liquid phase developments and 70 articles describing research on solid adsorbents and supports for the period 1994–1997. Much of the research on solid materials has been in the study of surface chemistry and deactivation procedures. Work on liquid phases has been on the development of new phases for specific difficult applications, such as the analysis of light hydrocarbons on capillary columns, and the development of high selectivity for specific compound classes, such as polycyclic aromatic hydrocarbons, substituted benzenes, and chiral separations. Developments in chiral stationary phases, organic polymers and inorganic liquid phases are described here.

6.1 Liquid Phases for Specific Applications

The column vendors have been introducing liquid phases tailored to specific applications and methods for several years. A survey of the major vendors shows liquid phases available for a variety of specialty applications, ranging from specific compendial methods to classes of analytes. Table 11 provides a summary of some applications for which specially formulated liquid phases are available.

6.2 Chiral Stationary Phases

The development of chiral stationary phases for GC has been an active research area in GC for many years. Cyclodextrins and derivatized cyclodextrins are predominantly used for chiral GC applications and have been recently reviewed in the text by Konig.[58] Due to the

Table 11 Selected commercially available specialty liquid phases and their applications

Application	Description
Thermally stable modified form of common phases; low bleed; highly inert	For GC/MS, electron capture detection and other highly sensitive analyses
35% Phenyl polysiloxane	Conformational analysis
Bonded poly(ethylene glycol)	FAME
Bis(cyanopropyl) polysiloxane	Positional and geometric isomers of polysiloxanes
Base modified polysiloxanes	Amines; basic analytes
Carbowax® amine	Primary, secondary, tertiary amines
6% Cyanopropyl phenyl, 94% polydimethyl siloxane	United States Pharmacopeia and European Pharmacopeia volatile organic contaminants methods

The major vendors provide proprietary bonded phases specifically designed for many compendial methods such as those for the USEPA and United States Pharmacopeia.

Table 12 Chiral GC applications using cyclodextrin-based liquid phases[58]

Application	Refs.
3,4-Diphenyl cyclopentene	64
Methyl 2-chloro propionate	65
Volatile compounds in oils	66–68
Polychlorinated biphenyls	69
Chemical warfare agents	70
Xylenes	71
FAME	72
Furanoids	73
Aromatic alcohols	74
Di- and trisubstituted benzenes	75, 76
Phenols	77
Amino alcohols	78
DDT	79

DDT, dichlorodiphenyl trichloroethane.

complexity of chiral recognition, many forces are involved in generating separations on cyclodextrin columns. Typically, multiple interactions are required, which may be any or all of the following possibilities. These include hydrogen bonding,[59] van der Waals interactions,[60] size of the inclusion cavity,[61] polar interactions[62] and steric hindrance.[63] Table 12 provides a listing of applications recently described using cyclodextrin columns.

6.3 Organic Polymer Stationary Phases

As described above, most modern GC liquid phases are composed of organic polymers that are coated into capillaries or onto solid supports. Most of the recent research and development of new polymeric substances can be divided into several areas: classification of currently available materials, liquid crystals, crown ethers, substituted silicones and chemically bonded cyclic organo-siloxanes (silica gels). Research on the specific interactions that determine retention on GC liquid phases continues today, with Santiuste and Takacks describing the interaction capacity of 26 common liquid phases.[80] Rohrschneider recently reviewed work in the general area of phase constants for liquid phase characterization.[15] Several workers have employed liquid crystals as GC stationary phases. Liquid crystals have recently been employed for the separation of hydrocarbons,[81] dioxins,[82] phthalic acids[83] and polycyclic aromatic hydrocarbons.[84] Liquid crystals show promise for the separation of positional isomers of both aliphatic and aromatic compounds.[85,86] Crown ethers have been bonded to polysiloxane liquid phases to provide special selectivity for short chain alcohols, ethyl esters of carboxylic acids and substituted benzene and phenol isomers.[87,88] Sol–gel technology has recently been used to make stationary phases for capillary columns with very high surface area and thermal stability, improving both capacity and ruggedness.[89]

7 CONCLUSIONS

The liquid phase is the heart of the GC separation system. It is here where all separation occurs. Development of new liquid phase technologies began along with the invention of GC and continues to be active. The renaissance in GC development that began with the introduction of fused silica capillary columns in 1979 also strongly affected stationary phase development. Today's capillary columns are relatively inexpensive, rugged, reproducible, inert and are available in a large variety of dimensions and liquid phases. In the future, new liquid phases that are more inert, have lower bleed, or address specific chemistries will continue to be developed. Stationary phase development continues to be a very active area of GC research and development.

ACKNOWLEDGMENTS

This article is respectfully dedicated to the memory of Professor Mathew Petersheim (1954–1998), friend, mentor and chair of the Seton Hall University Department of Chemistry from 1994–1996, who was taken from us far too soon.

LIST OF SYMBOLS

α	selectivity
β	column phase ratio
d_p	packing particle diameter
d_f	stationary phase film thickness
D_M	diffusion coefficient of an analyte molecule in the mobile phase
D_S	diffusion coefficient of an analyte in the stationary phase
ΔG°	standard free energy change
γ	tortuosity factor in the van Deemter Equation – related to the column packing
H	height equivalent to a theoretical plate
K	partition coefficient
k	retention factor
λ	packing factor – a constant in the van Deemter Equation related to the packing efficiency of the particles
μ	average linear gas velocity
N	number of theoretical plates
N	carbon number of the normal alkane eluting immediately before an analyte
$N+1$	carbon number of the normal alkane eluting immediately after an analyte

R	gas constant
r_c	radius of a capillary column
R_s	resolution
t_R	retention time
t'_R	adjusted retention time ($t_R - t_M$)
t_M	gas hold-up time
T	temperature
TZ	Treunzahl Number

ABBREVIATIONS AND ACRONYMS

AW	Acid Washed
DDT	Dichlorodiphenyl Trichloroethane
DEGS	Diethyl Glycol Succinate
DMDCS	Dichlorodimethyl Silane
FAME	Fatty Acid Methyl Esters
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GLC	Gas–Liquid Chromatography
GSC	Gas–Solid Chromatography
HETP	Height Equivalent to a Theoretical Plate
HMDS	Hexamethyl Disiloxane
NAW	Not Acid Washed
RI	Kovats Retention Index
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography in Screening of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of

Coatings **(Volume 2)**
Gas Chromatography in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Gas Chromatography by Direct Aqueous Injection in Environmental Analysis • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Solid-phase Microextraction in Environmental Analysis

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Full Range Crudes, Analytical Methodology of • Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Inverse Gas Chromatography in Analysis of Polymers

Pulp and Paper **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Multidimensional Gas Chromatography • Sample Preparation for Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. A.T. James, A.J.P. Martin, 'Gas–Liquid Partition Chromatography: The Separation and Microestimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid', *Biochem. J.*, **50**, 679 (1952).
2. M.J.E. Golay, in *Gas Chromatography 1957* (East Lansing Symposium), eds., V.J. Coates, H.J. Noebels, I.S. Ferguson, Academic Press, New York, 1–13, 1958.

3. R. Dandeneau, E.H. Zerenner 'An Investigation of Glasses for Capillary Chromatography', *J. High. Resolut. Chromatogr. Chromatogr. Commun.*, **2**, 351 (1979).
4. R.E. Majors, 'Response to GC Users' Survey: Current Trends in Column Use and Techniques', *LC–GC*, **8**, 444 (1990).
5. K. Grob, Jr., G. Grob, K. Grob, 'Comprehensive Standardized Quality Test for Glass Capillary Columns', *J. Chromatogr.*, **156**, 1–20 (1978).
6. K. Grob, G. Grob, K. Grob, Jr. 'Testing Capillary Gas Chromatographic Columns', *J. Chromatogr.*, **219**, 13–20 (1981).
7. E. Kovats, 'Gas-chromatographische Charakteriserung Organisher Verbindungen Teil 1: Retentionsindices Aliphaitscher Halssende, Alkohole, Aldehyde und Ketone', *Helv. Chim. Acta*, **41**, 1915–1932 (1958).
8. The Sadtler Standard Gas Chromatography Retention Index Library, Volumes I–IV, Sadtler Research Laboratories, Philadelphia, PA, 1985.
9. M.B. Evans, J.K. Haken, 'Recent Developments in the Gas Chromatographic Retention Index Scheme', *J. Chromatogr.*, **472**, 93–127 (1989).
10. H.F. Yin, Y.L. Sun, 'The Achievement of Reproducible Temperature Programmed Retention Indexes in Gas Chromatography Using Different Columns and Detectors', *Chromatographia*, **29**(1–2), 39–43 (1990).
11. T. Wang, B. Zhong, M. Chen, Y. Sun, 'Definitions and Methods of Calculation of the Temperature Programmed Retention Index, ITP, III. A Simplified Calculation Method Based on the Extended Kovats Definition', *J. Chromatogr.*, **390**(2), 275–283 (1987).
12. J.M. Takacks 'Contribution to the Retention Index System in Gas–Liquid Chromatography', *J. Chromatogr. Sci.*, **29**, 382–399 (1991).
13. L. Rohrschneider, 'Method of Characterization for Gas Chromatographic Separation of Liquids', *J. Chromatogr.*, **22**, 6–22 (1966).
14. W.O. McReynolds, 'Characterization of Some Liquid Phases', *J. Chromatogr. Sci.*, **8**, 685–691 (1970).
15. L. Rohrschneider, 'Development and Use of Phase Constants for the Characterization of Stationary Phase Selectivity in Gas Chromatography', *LC–GC*, **11**(10), 730–742 (1993).
16. N.H. Snow, 'Determination of Free Energy Relationships Using Capillary Gas Chromatography', *J. Chem. Educ.*, **73**, 592–597 (1996).
17. E. Dose, 'Prediction of Gas Chromatographic Retention Times Using Thermodynamic Retention Indices', *Anal. Chem.*, **59**, 2414 (1987).
18. J. Curvers, J. Rijks, C. Carmers, P. Knauss, P. Larson, 'Temperature Programmed Retention Indices: Calculation from Isothermal Data', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **8**, 607–617 (1985).
19. DryLab GC (Software for GC Method Development), Walnut Creek, CA, LC Resources, 1998.
20. J.H. Purnell, 'Correlation of Separating Power and Efficiency of Gas Chromatographic Columns', *J. Chem. Soc.*, 1268–1274 (1960).
21. J.J. van Deemter, F.J. Zuiderweg, A. Klinkenberg, 'Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Non-ideality on Chromatography', *Chem. Eng. Sci.*, **5**, 271–289 (1956).
22. H. McNair, J. Miller, *Basic Gas Chromatography*, John Wiley and Sons, New York, 91, 1998.
23. J. Giddings, *Unified Separation Science*, John Wiley and Sons, New York, 131–137, 1992.
24. E.F. Barry, 'Columns: Packed and Capillary/Column Selection in Gas Chromatography' in *Modern Practice of Gas Chromatography*, ed. R. Grob, John Wiley and Sons, New York, 123–234, 1995.
25. H.M. McNair, E.J. Bonelli, *Basic Gas Chromatography*, Varian Instrument Division, Palo Alto, CA, 9–20, 81–122, 1969.
26. J.M. Miller, *Chromatography, Concepts and Contrasts*, John Wiley and Sons, New York, 109–151, 1988.
27. H.M. McNair, E.J. Bonelli, *Basic Gas Chromatography*, Varian Instrument Division, Palo Alto, CA, 37–80, 1969.
28. 1997 Catalog, Supelco, Incorporated, Bellfonte, PA, 1997.
29. 'Packed Column Applications Guide', Supelco, Bellfonte, PA, 1998.
30. H.M. McNair, J.M. Miller, *Basic Gas Chromatography*, John Wiley, New York, 75, 1998.
31. S. Hawkes, D. Grossman, A. Hartkopf, T. Isenhour, J. Leary, J. Parcher, 'Preferred Stationary Liquids for Gas Chromatography', *J. Chromatogr. Sci.*, **13**, 115–117 (1975).
32. J.A. Yancey, 'Liquid Phases Used in Packed Gas Chromatographic Columns. Part III. McReynolds Constants, Preferred Liquid Phases and General Precautions', *J. Chromatogr. Sci.*, **24**, 117–124 (1986).
33. H.M. McNair, 'Gas Chromatography' Training Manual for American Chemical Society Short Courses, 1994.
34. L.G. Blomberg, K.E. Markides, 'The Role of Organosilicon Chemistry in the Preparation of Capillary Columns for Gas Chromatography', *J. HRC & CC*, **8**, 632–650 (1985).
35. T.J. Stark, R.D. Dandeneau, L. Mering, The Pittsburgh Conference 1980 Abstract 002, Atlantic City, NJ, 1980.
36. G. Schomburg, H. Husman, H. Borwitsky, 'Alkylpolysiloxane Glass Capillary Columns Combining High Temperature Stability of the Stationary Liquid and Deactivation of the Surface. Thermal Treatment of Dealkalinized Glass Surfaces by the Stationary Liquid Itself', *Chromatographia*, **12**, 651 (1979).
37. K. Grob, G. Grob, K. Grob Jr., 'Deactivation of Glass Capillary Columns by Silylation Part 1: Principles and Basic Technique', *J. HRC&CC*, **2**, 31 (1979).
38. R.C. Kong, C.L. Wooley, S.M. Fields, M.L. Lee, 'Deactivation of Small Diameter Fused Silica Capillary Columns for Gas and Supercritical Fluid Chromatography', *Chromatographia*, **18**, 362 (1984).

39. W. Jennings, *GC with Glass Capillary Columns*, 2nd edition, Academic Press, NY, 1980.
40. J. Bouche, M. Verzele, 'Static Coating Procedure for Glass Capillary Columns', *J. Gas Chromatogr.*, **6**, 501 (1968).
41. J.V. Hinshaw, L.S. Ettre, Introduction to Open Tubular Column Gas Chromatography, Cleveland, OH, Advanstar Publications, 15–26, 127–162, 1994.
42. K. Hyver (ed.), *High Resolution Gas Chromatography*, 3rd edition, Hewlett-Packard, Wilmington, DE, Chapter 4, 1989.
43. CP-SCANVIEW (Compact Disk containing Over 1400 Application Notes and Searching Software), Chrompack, Raritan, NJ, 1998.
44. *http://www.chrompack.com* (searched on October 29, 1998), Chrompack, Raritan, NJ, 1998.
45. *http://chem.external.hp.com/cag/main.html* (searched on October 29, 1998), Hewlett-Packard, Wilmington, DE, 1998.
46. *http://www.sigald.sial.com/supelco/lit/lit.htm* (searched on October 29, 1998), Supelco, Bellfonte, PA, 1998.
47. *http://www.jandw.com* (searched on October 29, 1998), J & W Scientific, Folsom, CA, 1998.
48. *http://www.restekcorp.com* (searched on October 29, 1998), Restek Corporation, Bellfonte, PA, 1998.
49. E.F. Barry 'Columns: Packed and Capillary Column Selection in Gas Chromatography', in *Modern Practice of Gas Chromatography*, ed. R. Grob, John Wiley and Sons, New York, 213, 1995.
50. M. Klee, 'Optimizing Separations in Gas Chromatography' in *Modern Practice of Gas Chromatography*, ed. R. Grob, 3rd edition, John Wiley & Sons, New York, 225–264, 1995.
51. M. Klee, M. Kaiser, K. Laughlin, 'Systematic Approach to Stationary Phase Selection in Gas Chromatography', *J. Chromatogr.*, **279**, 681–688 (1983).
52. D. Rood, *A Practical Guide to the Care, Maintenance and Troubleshooting of Capillary Gas Chromatography Systems*, Huthig, New York, 1995.
53. 'Troubleshooting Guide – How to Locate Gas Chromatography Problems and Solve Them' Supelco, Bellfonte, PA, 1998.
54. Column Test Chromatogram Packaged with HP-5MS Column Serial Number 57324021, Hewlett-Packard, 1997.
55. M. Lee, F. Yang, K. Bartle, *Open Tubular Column Gas Chromatography: Theory and Practice*, Wiley, New York, 26–27, 1984.
56. R. Majors, 'The Future of Gas Chromatography Column Technology: What the Experts Are Saying', *LC–GC*, **7**, 888 (1989).
57. G.A. Eicemann, H.H. Hill, Jr., J. Gardea-Torresdey, 'Gas Chromatography', *Anal. Chem.*, **70**(12), 321R–340R (1998).
58. W.A. Konig, *Gas Chromatographic Enantiomer Separation with Modified Cyclodextrins*, Huthig, Heidelberg, 1992.
59. T. Takeichi, T. Hishashi, S. Shigeru, T. Yuzi, M. Masami, 'Cyclodextrin Carbamates as a Novel Chiral Stationary Phase for Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **18**(3), 179–189 (1995).
60. Z. Juvancz, J. Petersson, 'Enantioselective Gas Chromatography', *J. Microcolumn Sep.*, **8**(2), 99–114 (1996).
61. A. Shitangkoon, G. Vigh, 'Systematic Modification of the Separation Selectivity of Cyclodextrin Based Gas Chromatographic Stationary Phases by Varying the Size of the 6-O Substituents', *J. Chromatogr. A*, **738**(1), 31–42 (1996).
62. A. Berthod, Y.Z. Eve, L. Kang, W. Daniel, 'Determination and Use of the Rohrschneider–McReynolds Constants for Chiral Stationary Phases Used in Gas Chromatography', *Anal. Chem.*, **67**(5), 849–857 (1995).
63. B.E. Kim, K.P. Lee, K.S. Park, S.H. Lee, J.H. Park, 'Enantioselectivity of 6-O-Alkyldimethylsilyl-2,3-Di-O-Methyl-beta-Cyclodextrins as Chiral Stationary Phases in Capillary GC', *Chromatographia*, **46**(3/4), 145–150 (1997).
64. L.A. Asuncion, J.E. Baldwin, 'Quantitative Analysis of the Four Isomers of 3,4-diphenyl Cyclopentane by Chiral Gas Chromatography', *J. Org. Chem.*, **60**(18), 5778–5784 (1995).
65. D.U. Staerk, A. Shitangkoon, G. Vigh, 'Preparative Gas Chromatographic Separation of the Enantiomers of Methyl 2-chloropropionate Using a Cyclodextrin-based Stationary Phase', *J. Chromatogr. A*, **702**(1,2), 251–257 (1995).
66. C. Bicchi, A. D'Amato, V. Manzin, A. Galli, M. Galli, 'Cyclodextrin Derivatives for the Gas Chromatographic Separation of Racemic Mixtures of Volatile Compounds Part VIII: 2,6-Di-O-methyl-3-o-pentyl-γ-cyclodextrin and 2,6,-Di-O-methyl-3-o-(4-oxo-pentyl)- and 2,6-Di-O-pentyl-3-o-(4-oxopentyl)-β-and γ-cyclodextrin', *J. High Resolut. Chromatogr.*, **18**(5), 295–298 (1995).
67. T.J. Betts, 'Indications of the Chemical Structure of Oxygen-containing Constituents of Volatile Oils by Capillary Gas Chromatography Using One or Two Modified β and α-Cyclodextrin Phases', *J. Chromatogr. A*, **724**(1,2), 403–410 (1996).
68. T.J. Betts, 'Use of a Trio of Modified Cyclodextrin GC Phases to Provide Structural Information on Some Constituents of Volatile Oils', *J. Chromatogr. A*, **707**(2), 390–395 (1995).
69. J. Krupcik, M. Majekova, P. Majek, J. Hrouzek, E. Benicka, F. Onuska, P. Sandra, J. deZeeuw, 'On the Interconversion Energy Barriers Obtained for Atropisomers of Some Polychlorinated Biphenyls by AM1 Semi-empirical Quantum Chemistry and Gas Chromatography on a Modified Cyclodextrin Stationary Phase', *Fresenius' J. Anal. Chem.*, **352**(7–8), 696–698 (1995).
70. J.R. Smith, J. Schlager, 'Gas Chromatographic Separation of the Stereoisomers of Organophosphorous Chemical Warfare Agents Using Cyclodextrin Capillary Columns', *J. High Resolut. Chromatogr.*, **19**(3), 151–154 (1996).

71. J. Shi, Y. Shao, G. Zou, 'Preparation of Phenyl Isocyanate and Study of its Chromatographic Behavior as a Stationary Phase', *Sepu*, **13**(6), 444–446 (1995).
72. R. Stritzel, B. Dobner, F. Bringezu, P. Nuhn, 'Separation of Racemic 2-Alkyl Branched Fatty Acid Methyl Esters by Gas Chromatography on a Commercially Available Chiral Phase', *J. High Resolut. Chromatogr.*, **19**(2), 121–123 (1996).
73. P. Kreis, A. Dietrich, A. Mosandl, 'Elution Order of the Furnaoid Linalol Oxides on Common Gas Chromatographic Phases and Modified Cyclodextrin Phases', *J. Essential Oil Res.*, **8**(4), 339–341 (1996).
74. L. Wu, Z. Li, A. Mi, Y. Jiang, 'Enantiomeric Analysis in Gas Chromatography II: Separation and Quantitation of Chiral Aromatic Alcohols', *Sepu*, **14**(2), 81–85 (1996).
75. H.B. Zhang, Y. Ling, R. Fu, Y.X. Wen, J.L. Gu, 'A New Gas Chromatographic Stationary Phase: Polysiloxane with Beta-cyclodextrin Side Chain for the Separation of Chiral and Positional Isomers', *Chromatographia*, **46**(1/2), 40–48 (1997).
76. D.Q. Xiao, Y. Ling, R. Fu, J.J. Gu, A. Luo, 'Cyclodextrins With Heterocyclic Substitution as GC Stationary Phases', *Chromatographia*, **46**(1/2), 85–91 (1997).
77. A. Steinborn, R. Reinhardt, W. Engewald, K. Wyssuwa, K. Schulze, 'Enantiomer Separation of a-Campholene and Fencholene Derivatives by Capillary Gas Chromatography on Permethylated Cyclodextrins', *J. Chromatogr. A*, **697**(1,2), 485–494 (1995).
78. E. Kuesters, P. Andrea, 'Enantiomeric Separation of Amino Acids by Gas Chromatography on a Chiral Stationary Phase: Influence of the Perfluoroacylating Reagent on the Separation', *J. High Resolut. Chromatogr.*, **17**(9), 639–642 (1995).
79. H.R. Buser, M.D. Mueller, 'Isomer Selective and Enantioselective Determination of DDT and Related Compounds Using Chiral High Resolution Gas Chromatography/Mass Spectrometry and Chiral High Performance Liquid Chromatography', *Anal. Chem.*, **67**, 2691–2698 (1995).
80. J.M. Santiuste, J.M. Takacks, 'Substance Specific and Average Polarity Factors for Characterization of Stationary Phases Used in Gas–Liquid Chromatography', *J. Chromatogr. Sci.*, **35**(10), 495–501 (1997).
81. Z.P. Vetrova, I.A. Ianova, N.T. Karabanour, O. Akipova, 'Liquid Crystals in Gas Chromatography', *Izv. Akad. Nauk. Ser. Fiz.*, **59**(3), 154–157 (1995).
82. K.P. Naikwandi, I.A. Albrecht, F.W. Karasek, H. Gohda, 'Development of Capillary Columns for Environmental Analyses', *Organohalogen Compd.*, **19**, 139–142 (1994).
83. V.A. Evteeva, O.B. Akopova, 'Application of Liquid Crystalline Phases for Separation of Isomeric Phthalic Acids by Gas–Liquid Chromatography', *Izv. Akad. Nauk. Ser. Fiz.*, **59**(3), 158–160 (1995).
84. S. Blokhina, A. Krestov, M. Oklhovich, A. Lokhanova, 'Thermodynamics of the Sorption of *n*-Alkanes and Alcohols on a Mesomorphic Nickel Complex', *Zh. Fiz. Khim.*, **69**(5), 901–904 (1995).
85. F. Perez, P. Berdague, J. Cortieu, J.P. Bayle, S. Boudah, M.H. Guermouche, 'Comparison of Two Azobenzene Liquid Crystal Stationary Phases in Open Tubular Column Gas Chromatography', *J. Chromatogr. A*, **746**(2), 247–254 (1996).
86. F. Perez, P. Berdague, J. Cortieu, J.P. Bayle, S. Boudah, M.H. Guermouche, 'Investigation of the Retention Properties of Three Laterally Substituted Liquid Crystal Stationary Phases by Gas Chromatography', *Bull. Soc. Chim. Fr.*, **133**(4), 427–433 (1996).
87. X. Zhou, Y. Hui, W. Caiying, C. Yuanyin, 'Two New Chiral Polysiloxanes Containing Chiral Crown Ethers for Capillary Gas Chromatography (CGC) Separation of Enantiomers', *Sepu*, **12**(6), 404–405 (1994).
88. Z. Zhou, Z. Yongchang, X. Minggui, C. Ye, 'GC Analysis of Optically Active Amyl Alcohol in Fusel Oil', *Sichuan Daxue Xuebao Ziran Kexueban*, **32**(1), 74–77 (1995).
89. D. Wang, S.L. Chang, A. Malik, 'Sol–Gel Column Technology for Single Step Deactivation, Coating and Stationary Phase Immobilization in High Resolution Gas Chromatography', *Anal. Chem.*, **69**(22), 4566–4576 (1997).

Multidimensional Gas Chromatography

Wolfgang Bertsch
The University of Alabama, Tuscaloosa, USA

1 Introduction and Historical Perspective **1**
2 Peak Capacity and Overlap in Gas Chromatography **3**
3 Conventional Two-dimensional Gas Chromatography **6**
3.1 Instrumentation and Procedures 6
3.2 Practical Aspects 9
3.3 Selected Applications 9
4 Comprehensive Two-dimensional Gas Chromatography **14**
4.1 Operational Aspects and Future Developments 19
Acknowledgments **20**
List of Symbols **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

Two-dimensional gas chromatography (2-D GC) is a logical extension of gas chromatography (GC) carried out on a single column. It provides physical resolution of components which are difficult to analyze by other means. Two different arrangements are possible in 2-D GC. In the conventional mode, effluent fractions from the primary column are switched onto a secondary column of different polarity. In the comprehensive mode, the entire effluent from the primary column is examined by the secondary column. The major advantage of 2-D GC is the enhancement of peak capacity. Statistical methods of overlap theories predict severe limitations of the resolution of components in multicomponent mixtures. The problems are exacerbated when components form clusters, such as in the analysis of isomers. The determination of target analytes at trace levels also poses unique challenges. Selective detectors, in particular mass spectrometry (MS), can provide answers when merged components are capable of producing distinctive spectra. MS can compensate for poor chromatographic resolution in many cases but it cannot be applied for the deconvolution of isomers and closely related substances. In this situation, 2-D GC is the only choice.

Instrumentation in conventional 2-D GC is based on one of two different effluent switching technologies. A valve can be used to divert the effluent from the primary column to a secondary column of different selectivity. In this arrangement, the sample inevitably comes into contact with the metal of the valve body. The other design is based on fluidic switching by application of a difference in pressure in the interface between the columns. This technology, generally referred to as Deans switching, can avoid metals in the sample flow path but it is technically more difficult than valve-based technologies.

Comprehensive 2-D GC involves a focusing step. The effluent from the primary column is periodically injected into a high speed column in the form of narrow chemical pulses. The separation on the secondary column must be complete (or nearly complete) before the next pulse can be injected. The secondary column is polar, i.e. substances that have similar boiling points but differ in polarity can be resolved in the second dimension. The separations process is continuous and resembles data acquisition in routine gas chromatography/mass spectrometry (GC/MS). One of the true advantages of comprehensive 2-D GC is that the separations mechanisms on both columns can be decoupled, leading to true orthogonality. Components with similar properties fall into clearly defined data spaces. The chromatograms show a measure of order which is useful in substance identification and group type analysis.

Comprehensive 2-D GC is still in its infancy. There is much room for improvements, including in the data acquisition/display area. Conventional 2-D GC, however, is mature. Because of the complexity of 2-D GC instrumentation, selective detection, i.e. GC/MS, is usually preferred where it is applicable. Many applications based on 2-D GC have been published. The analysis of flavors/fragrances, petrochemicals and persistent organohalogens in the environment by 2-D GC offers distinctive advantages over competing technologies.

1 INTRODUCTION AND HISTORICAL PERSPECTIVE

GC is undoubtedly the most widely used technique for the separation of volatile components. It is amenable to the analysis of a wide range of organic and inorganic materials. Some of the strongest aspects of GC come from its operational characteristics. It is relatively easy and inexpensive to achieve high separation power and short turnaround times. The wide array of excellent general and specific detectors available for GC is perhaps one of its biggest attractions. There is no other instrument in chromatography that can offer this level of performance. GC occupies an important position in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the array of separation techniques. Although it is not the oldest form of chromatography, it was the first one to mature as an instrumental technique. Relatively sophisticated instrumentation based on packed column technology and flame ionization detection or electron capture detection became commercially available in the late 1950s and early 1960s. The separation power and versatility of capillary column GC was demonstrated early on. Efficient metal capillary columns with apolar stationary phases were relatively easy to produce but problems due to active surfaces could not easily be conquered.[1] The introduction of glass as a column material opened an avenue to overcome this major hurdle.[2] Glass drawing machines became commercially available and aided in the development of this column type. Many recipes were offered for the production of truly inert surfaces and stationary phase coating techniques were perfected.[3] Immobilized stationary phases, introduced in the early 1970s, were another major development.

Many scientists have contributed to the advancement of capillary column technology during these years. History records many individual inventions and advancements. No individual was as influential in the overall development of glass capillary column technology as Grob.[4] One might have assumed that chromatographers would jump at this new development and substitute packed column applications demanding high resolving power with capillary column based methodology. Unfortunately, little progress was made in this respect. Glass proved to be too difficult to handle on a routine basis and capillary column dedicated instruments were not built because of patent infringement issues. The situation again changed dramatically in 1979 when a scientist working for the Hewlett Packard Company adapted glass fiber optics technology to GC.[5] Flexible fused silica capillary columns quickly replaced glass capillary column technology. This column type was not only easier to handle from a practical perspective but a wide variety of commercial columns were introduced onto the market shortly after the basic patent expired. Capillary column GC finally took off. It is not surprising that many scientists pushed the exploration of the principles of GC to its limits. The introduction of an entirely new family of affordable spectrometers as on-line detectors made a major impact. The practical limits of GC as a tool and as a technique were approached in more than one area. Although GC inherently has high resolving power, it has limitations in many areas, including the resolution of highly complex samples.

2-D GC is one of the most effective techniques to deal with multicomponent samples, especially those consisting of substances having a wide range of polarities. The principle is simple: components that experience similar retention characteristics on one particular stationary phase are subjected to an additional separation step on a second column of different selectivity. Two components of approximately the same boiling point which elute on an apolar phase as a single peak may have different functional groups. Differences in polarity, polarizability, etc., can then be exploited. This process can be carried out off-line but this technology is very difficult for trace components. On-line techniques rely on some type of column switching hardware that allows the transfer of an effluent fraction between the columns. This type of manipulation, generally referred to as heartcutting, GC/GC, or 2-D GC, has been used since the mid-1950s, primarily in process control. Simmons and Snyder[6] are credited with the first use of 2-D GC in general analysis. In principle it is possible to resolve component(s) from a mixture as long as a set of stationary phases is available that differentiates between the target compound(s) and its interferences. Full characterization of a complex sample requires the transfer of a large number of individual cuts. The process becomes laborious and time-consuming. In practice, there are other hurdles to overcome as well. As long as only a single heartcut is being made, it is possible to correlate a secondary chromatogram to a transfer from the primary column. This advantage is lost when multiple heartcuts are made in the course of a single run. Since stationary phases act independently, components from individual transfers tend to mingle and can thus no longer be correlated.

A technical solution to carry out multiheartcut 2-D GC has been developed. Narrow slices of effluent are periodically injected from a primary column into a short, high-speed secondary column. Components which are not resolved in the first dimension undergo a second separation step. The process is analogous to routine GC/MS and is known as comprehensive 2-D GC. In both processes, the entire sample is sliced into narrow packets for further examination. The practical implementation of comprehensive 2-D GC is the brainchild of Phillips who invented a thermal modulator as a sample introduction device. Comprehensive 2-D GC is less than 10 years old.[7] As with any new technology, there are still many obstacles to be overcome. The technical side is moving along nicely. New ground had to be broken in several areas, notably in data acquisition and display. The situation can be compared to the early days of GC/MS when analog to digital interfaces and computers were slow and of limited capacity. Looking at the steady progress in GC/MS, one has to be optimistic that current stumbling blocks will eventually fall by the wayside.

The inventors and proponents of comprehensive 2-D GC have not done a very good job of ''selling'' their invention. Many chromatographers who are quite knowledgeable remain skeptical because the fundamental principles, technical approaches and strengths/weaknesses of

this new tool have not been adequately explained. There is also a lack of general applications. At present only one small firm[8] has secured the rights to exploit the patents issued for the basic design.[9,10] No more than around a dozen laboratories in the world are actively pursuing research in a major way in this area.

2-D GC is a multifaceted technique. Its major use is in the physical separation of multicomponent mixtures but there are other areas of application as well. The preparative scale isolation of trace components in a complex mixture is an example. Hardware differs for conventional heartcutting-type operations and comprehensive 2-D GC differs significantly. It is described in separate paragraphs.

2 PEAK CAPACITY AND OVERLAP IN GAS CHROMATOGRAPHY

Two-dimensional separation procedures become especially important when complex samples consisting of hundreds of components must be analyzed. Resolution in GC depends on several factors. For isothermal conditions, Equation (1) applies:

$$R_s = \frac{1}{4}\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k'}{k' + 1}\right)\sqrt{\frac{L}{H}} \tag{1}$$

where R_s is resolution, α selectivity, k' solute retention factor, L column length, and H theoretical plate height. Since the number of theoretical plates is proportional to column length (Equation 2)

$$N = \frac{L}{H} \tag{2}$$

where N is the number of theoretical plates, resolution varies with the square root of column length. Retention, and thus time of analysis, is however directly proportional to column length. In practice, resolution also depends on other factors such as the relative magnitude of the peaks being resolved. Merged peaks of approximately equal height can easily be recognized and quantitated with good accuracy at less than unit resolution. As peak ratios continue to deviate, merged peaks become less distinctive. In practice, it is not uncommon to quantitate adjacent components in a chromatogram that differ in concentration by several orders of magnitude.

Even though Equations (1) and (2) are useful in describing fundamental relationships in GC, they are very limited in practice. The vast majority of GC separations are carried out under temperature programmed conditions. Unfortunately, the effects of temperature programming on resolution are complex and difficult to describe. An alternative way to evaluate the separation power of a column under both isothermal and temperature programmed conditions is the use of the peak capacity concept. Peak capacity is defined as the maximum number of component peaks that can be packed side by side into the available separation space at a given resolution.[11] The frequency of component overlap is thus a function of peak capacity. For isothermal runs, peak capacity n_c can be related to the number N of theoretical plates as shown in Equation (3)[10]

$$n_c = \theta N^{1/2} \tag{3}$$

where n_c is peak capacity, and θ is a variable that depends on retention time range. The determination of peak capacity under conditions of temperature programming is complex.[12] In conventional capillary column GC, typical peak capacity values are in the range of a few hundred. It would appear that peak capacities of this magnitude would be adequate for all but the most complex samples. Unfortunately, this is not the case. Davis and Giddings[13] developed a concept based on the random distribution of component peaks in chromatographic separation space. The mathematical treatment which is based on Poisson statistics is now commonly referred to as the statistical model of overlap (SMO) theory. It has been refined and expanded for the last 20 years by Davis,[14] a former Giddings student, and his group. Alternative concepts for the evaluation of statistical peak overlap have also been offered.[15,16] Regardless of the model used, the relationships are complex and the findings are not easily applicable in practice. Simple calculations show, however, that single column systems are not capable of producing separation numbers which are adequate even for samples of only moderate complexity.

SMO theory can be used to predict resolution as a function of several variables. The number of peaks p observed in a chromatogram and the expected component number m in a sample are related to column peak capacity n_c as shown in Equation (4)

$$p = m\mathrm{e}^{-m/n_c} = m\mathrm{e}^{-\alpha} \tag{4}$$

where α is the saturation factor. In the special case where the expected number of components is equal to the peak capacity of the column at hand, the function reaches a maximum. With random spacing, only 37% of the peaks can theoretically be resolved. Resolution becomes an exponential function of peak capacity, as shown in Equation (5)

$$\left(\frac{p}{n_c}\right)_{\max} = \mathrm{e}^{-1} = 0.3679 \tag{5}$$

In practical chromatography, the results are even worse since factors such as peak distortion, peaks differing in magnitude, instrument noise and similar disturbances

have not been considered in this model. The number of single component peaks s, shown in Equation (6), is smaller than p,[14]

$$s = m\mathrm{e}^{-2m/n_c} = m\mathrm{e}^{-2\alpha} \tag{6}$$

The number of peaks representing single components cannot exceed 18% of the available peak capacity n_c. The ratio s/m represents the fraction of all components isolated as single component peaks. It can be expressed as probability P_1, as shown in Equation (7):

$$P_1 = \frac{s}{m} = \mathrm{e}^{-2\alpha} \tag{7}$$

As expected, the occurrence of single component peaks is linked to saturation.

Some examples can illustrate the consequences of statistical component overlap and point toward potential remedies. Consider a sample consisting of 50 components which needs to be analyzed on a column having a peak capacity n_c of 100. It would appear that most components should be adequately resolved, since twice the necessary peak capacity is available. Only 18 substances or 36% of the sample will however be represented as single component peaks.[13] This is a clear disappointment. Decreasing the level of saturation improves the chance of obtaining single component peaks but the gains are painfully slow. At $\alpha = 0.5$ and $\alpha = 0.25$, P_1 increases to 37% and 61%, respectively. To elevate P_1 to 82%, saturation α must be lowered to 0.1, i.e. a 10-fold excess of peak capacity must be available. Equations (4) and (7) show that the 100-component sample requires a column peak capacity n_c of 288 for a relatively modest probability P_1 of 50%. Under favorable circumstances, this separation can be achieved on a column producing 100 000 theoretical plates. To increase the probability to 90% and 99%, respectively, columns of 4 000 000 and 500 000 000 theoretical plates are required. This type of performance is clearly outside the range of most GC instruments. The current limit of single column GC is around 4 000 000 theoretical plates, corresponding to a peak capacity n_c of approximately 1000.[17] Gasoline, a sample of moderate complexity, was analyzed on such a system. The chromatogram produced 970 distinguishable signals, mostly in the form of merged peaks. The results from this analysis are impressive but it took 15 h to complete the separation.

The achievement of high efficiency carries another price tag. Since plate height H is approximately inversely proportional to the inverse of column diameter, small bore columns have to be used. Columns with small internal diameter not only produce a high pressure drop but the amount of stationary phase available also becomes a limiting factor.[18] Columns of 50–100 μm internal diameter have low sample capacity and are difficult to operate.[19] Recycle chromatography is an alternative to overcome both limitations. This technique is based on the use of two relatively short conventional capillary columns connected via a low dead volume valve. The effluent fraction which requires further resolution is alternatively routed through each of the columns by turning the valve at the appropriate time. A recycle unit consisting of two 20-m segments produced over 2 000 000 theoretical plates for a run consisting of 20 cycles.[20] Large peak capacities can easily be obtained. The resolution of isomer pairs such as pristine and phytane which requires very high efficiencies on any stationary phase can be accomplished. Recycle operations are quite amenable to 2-D GC because they are already based on the use of valves. There is a problem however with the timing of valve switching. A very low dead volume in-line detector is necessary or some of the effluent has to be diverted to an off-line detector. The prospects of recycle GC are exciting but these problems must first be overcome to make it truly competitive.[21]

2-D GC can provide an alternative to single column GC when highly efficient systems are required. There are two conditions that must be fulfilled: (1) the components must be subjected to two separation steps (mechanisms); and (2) once separation has been achieved in any single step, the substances must remain separated.[22] The power of 2-D GC comes from the use of complementary stationary phase selectivities and the utilization of the "empty space" in the second column. Figure 1 shows a basic diagram of the separation concept in a coupled column system. It is obvious that narrow heartcuts reduce the chances of coincidental overlap in the second column. From a practical point of view, one needs to consider additional factors.

- Components which have undergone peak broadening in the primary column continue to do so in the secondary column. Final peak widths may become unacceptable.

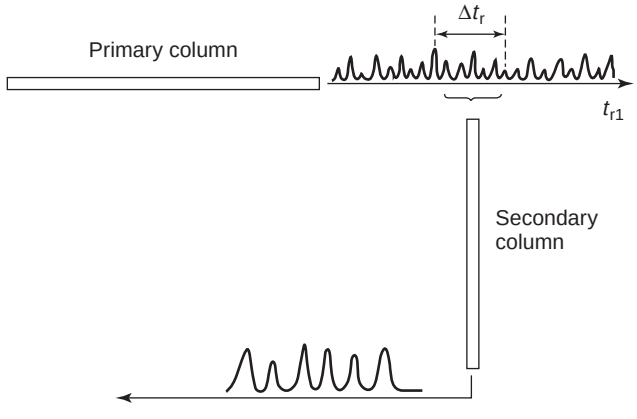

Figure 1 Schematic diagram of idealized 2-D GC showing a single heartcut.

- Peaks that are already partially separated may merge in the secondary column due to ''a mixed phase mechanism''.
- It may be difficult to correlate the retention times of components on both columns unless one can keep track of their relative movements.

The installation of an intermediate trap between the columns is a way to overcome some of the hurdles. Solutes can be refocused and reintroduced as narrow pulses. In this manner, relatively broad cuts can be taken. As the width of the cut is widened, an increasing number of substances is introduced and a point of diminishing return is eventually reached. The other extreme is the transfer of effluent on the timescale of the width of a single peak. Narrow heartcuts are likely to produce only a few components in secondary chromatograms. There is an obvious balance between the number of transfers, complexity of secondary chromatograms, and time requirements.

The major use of heartcutting in 2-D GC is the physical separation of a few trace target compounds in the presence of major interferences. The complete two-dimensional characterization of a sample requires a different approach. Figure 2 shows an arrangement where adjacent heartcuts are performed within the same run. The maximum peak capacity $n_{c(max)}$ of such a system is approximately given by Equation (8):

$$n_{c(max)} = \bar{n}_c \times \text{number of columns} \qquad (8)$$

where $\bar{n}_c$ is the average peak capacity of the columns used in the system. Two independent secondary columns allow the separate development of individual heartcuts. Components eluting at the interface between the individual transfers will be found in both secondary chromatograms. All other substances will appear in only one of the chromatograms. The components which have been resolved on the first column thus remain separated. It is obvious that the complete characterization of a sample in an assembly of multiple secondary columns is not practical. As an alternative, an arrangement using multiple parallel traps can be used.[23,24] Effluent from the primary column is essentially stored in these traps. This approach differs from the conventional heartcut-type operation indicated in Figure 1. The contents of each trap can be injected into the secondary column, one at a time. The ambiguity introduced by multiple heartcutting is thus eliminated. There are some practical problems, as expected. The apparatus is complicated and the adsorbents used in the traps may have a tendency to interact with some of the more delicate sample components. The use of recently introduced polysiloxane-based trapping materials can overcome activity problems.[25]

The most effective way to improve the total peak capacity of a system is the addition of a large number of secondary columns. Figure 3 shows the equivalent of a two-dimensional separation in planar chromatography. The hardware giving access to this type of dispersion in GC is actually quite simple, at least in principle. The effluent is focused as it emerges from the end of the primary column. It is then periodically injected as a series of sharp pulses into a single secondary column. Elution time from the secondary column is synchronized to the frequency of injection from the primary column, i.e. the last component must have left the secondary column before the next pulse is injected. This type of operation is referred to as *comprehensive* 2-D GC because the entire sample is processed in a comprehensive manner.

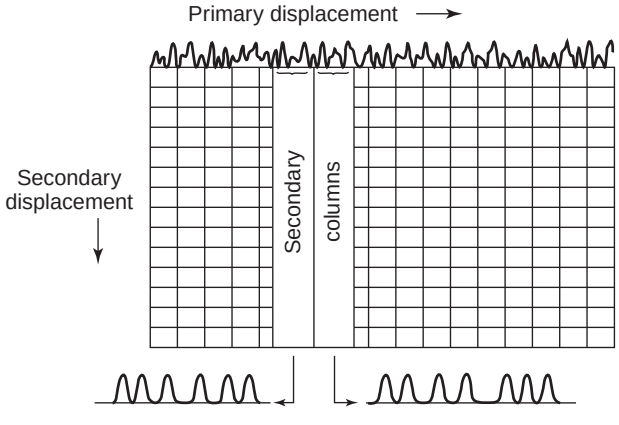

Figure 2 Superposition of two adjacent secondary columns of a coupled column system.

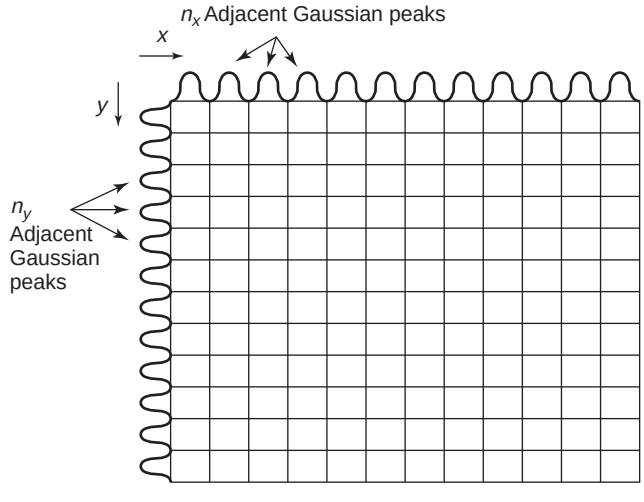

Figure 3 Idealized presentation of the peak capacity of a comprehensive two-dimensional system. Individual peak capacities are represented by the number of adjacent profiles.

Table 1 Comparison of conventional (single column) GC with various modes of 2-D GC[a,b]

	Conventional single column GC	Heartcut 2-D GC	Parallel trap 2-D GC	Heartcut recycle 2-D GC	Comprehensive 2-D GC
Simple mixture	++	+	+	+	+
Complex mixtures					
One or a few compounds of interest	+	++	++	+	+
Many compounds of interest	−	+	++	−	++
"All" compounds of interest	−−	+/−	+	−−	++
Resolution	+/−	+	+	++	++
Complexity	++	+	+/−	−	−
Flexibility	++	+/−	+/−	−	−
Robustness	++	+	+/−	+/−	+/−

[a] Symbols (++, +, +/−, − and −−) indicate suitability.
[b] Reproduced from H.-J. de Geus et al., 'Two-dimensional Gas Chromatography for Environmental Analysis', *Trends Anal. Chem.*, **15**, 398–408, © 1996, with permission from Elsevier Science.

The maximum peak capacity of such a system can be approximated by Equation (9):

$$n_{c(max)} = n_{c(1)} \times n_{c(2)} \quad (9)$$

where $n_{c(1)}$ and $n_{c(2)}$ are the peak capacities of the primary and secondary columns, respectively.[26] The term GC × GC which is used interchangeably with the term comprehensive 2-D GC emphasizes the multiplicative nature of individual column peak capacities in the combined system. Although there are some restrictions on the optimal use of this technology, it is clearly the only practical route toward the achievement of very high peak capacities. At the present time, peak capacities n_c approaching 54 000 have already been demonstrated.[27]

Instrumentation and methodologies of conventional 2-D GC and comprehensive 2-D GC differ in many aspects. There are distinct advantages and disadvantages to each type. It is difficult to make generalizations because the suitability of a particular type of 2-D GC depends on details of the particular instrument design and on the problem at hand. Table 1 is a general overview of advantages and disadvantages of the basic instrument types available.[28] Aspects of conventional 2-D GC and comprehensive 2-D GC are presented separately.

3 CONVENTIONAL TWO-DIMENSIONAL GAS CHROMATOGRAPHY

3.1 Instrumentation and Procedures

2-D GC instrumentation consists of an arrangement of two columns where one of the columns acts as inlet for the other one. Functional instruments require a great deal of attention to the interaction between sample components and surfaces and to unswept volumes in interfaces. Contact of sample components with unsuitable surfaces has the potential of causing undesirable reactions or tailing due to adsorption. Excessive transit times are a source for experimental dilution and band broadening. Interestingly, little information has appeared in the recent literature on new developments and refinements of 2-D GC hardware. Most of the developments in 2-D GC date back to the 1970s and 1980s. This is in stark contrast to many other areas of GC where improved technology has driven instrument design to new heights.

The introduction of low cost spectrometric detectors, in particular the mass spectrometer, has diverted much interest away from 2-D GC. The emergence of routine MS in conjunction with GC is the main reason why 2-D GC has not flourished. These techniques are in competition with each other. Table 2 summarizes trends in both disciplines. The number of publications in GC has held relatively steady, as opposed to liquid chromatography which experiences slow but steady growth. The use of GC/MS is still on the rise although it has slowed somewhat. By contrast, 2-D GC in all its forms is insignificant. It has stagnated over the last 10 years.

Benchtop-type GC/MS is relatively inexpensive, versatile and has a user interface conducive to the production of large volumes of data. In contrast, 2-D GC requires fine-tuning and is relatively difficult to carry out. In many

Table 2 Number of publications in the *Chemical Abstracts* database, according to keyword index, reported in 4-year increments

Year	GC	GC/MS	2-D GC
1978	4497	894	11
1982	4761	994	16
1986	4643	1237	43
1990	4049	1542	57
1994	4056	2258	56
1998	3992	2541	62

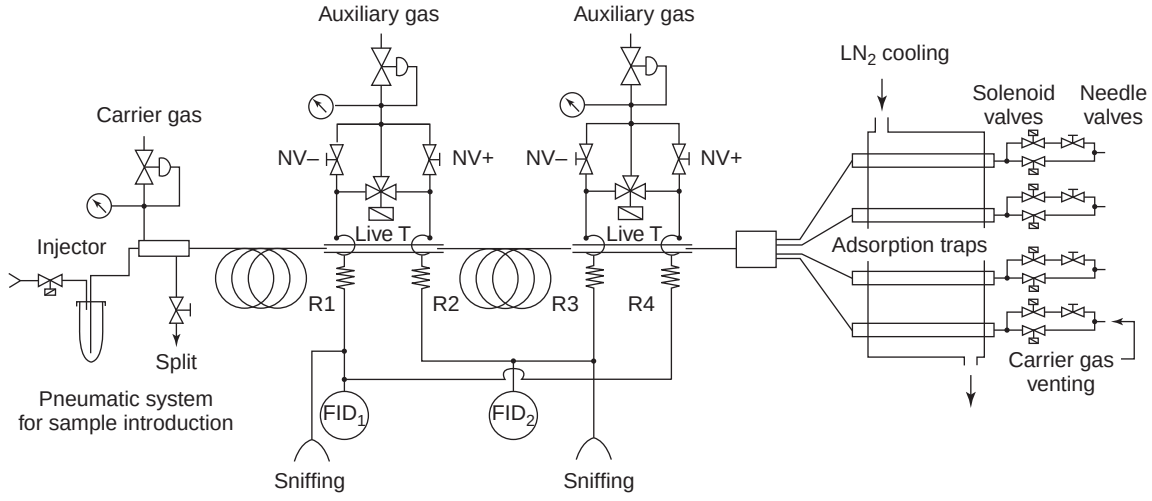

Figure 4 Schematic diagram of dynamic headspace equipment combined with 2-D GC. The system is based on Deans switches (Siemens Sichromat). The micropreparative module can be connected to a detector such as a mass spectrometer. NV, needle valves; R, restrictors; sniffing, mask for organoleptic evaluation. (Reproduced from S. Nitz et al., 'Determination of Sensorial Active Trace Compounds by Multidimensional Gas Chromatography Combined with Different Enrichment Techniques', *J. Chromatogr.*, **471**, 173–185, © 1989, with permission from Elsevier Science.)

situations, mass spectra of target compounds are sufficiently different from the interfering background to permit characterization and quantitation with little or no chromatography. Mass spectral selectivity is often substituted for chromatographic resolution. Unfortunately, there are many pitfalls to this approach. A great deal of unreliable data may be generated from the inappropriate use of MS. There are many situations where the physical separation of target compounds is the only viable option. MS can generate highly reproducible data which may suffer from the same interferences. High precision is often taken as an indicator for accuracy. Interferences in MS are very common because related substances produce common ions. Strict quality control measures must be put in place to safeguard against poor data quality.

Conventional 2-D GC instrumentation offers a wide variety of options. In the most simple case, the columns are connected via a microvalve and are placed in a single oven. Complex instruments feature sophisticated pressure balanced pneumatic switches under conditions of mass flow control. This principle is often referred to as a Deans switch, after its original proponent.[29] A secondary inlet for the injection of calibration standards into the second column and a monitor detector between the two columns are useful, almost necessary, additions. Other features include dual ovens, a computer controlled interface responding to events in real time, an intermediate trap and a sophisticated selective detector, such as a mass spectrometer, infrared spectrometer, or atomic emission spectrometer. In special cases where the on-column enrichment of trace components is required, a parallel trap accessory may be a part of the instrument. Figure 4 shows the design of such an instrument. This particular design is geared toward the on-line evaluation of flavors and fragrances and features two Deans-type switches.[30] This instrument can be equipped with a variety of detectors, including a sophisticated isotope ratio mass spectrometer.[31] Isotope ratio MS is an important tool to uncover adulteration in the food industry and study details of biosynthesis and metabolism.[32] Valve-based instruments can also be quite sophisticated. Figure 5 shows the design of such an instrument.[33] Only one of the five valves actually comes into contact with the sample. The other units are used to carefully balance pressures and turn on/off the flows in the rest of the instrument. Vacuum can also be used to switch effluent between columns. The effluent is virtually sucked into an interface, as it emerges from the primary column.[34–36]

Contact of the sample with surfaces other than well deactivated glass or fused silica is of much concern. It is obvious that valve-based systems must contain metal that may cause problems with labile compounds. Instruments equipped with Deans-type switches can totally avoid metals in the effluent flow path.[37] Arrangements based on fused silica column technology permit the use of friction-type connectors, tees and crosses which are made of glass and can thus be well deactivated.[38] Metal, in particular stainless steel, has perhaps an unjustified bad reputation as a poor chromatographic surface. Instruments equipped with both valves and Deans-type switches have been compared for potential artifacts. The results are inconclusive.[39] Oxygenated compounds, in particular free acids, anhydrides, and aldehydes, are usually good indicators of residual activity.[40,41] It

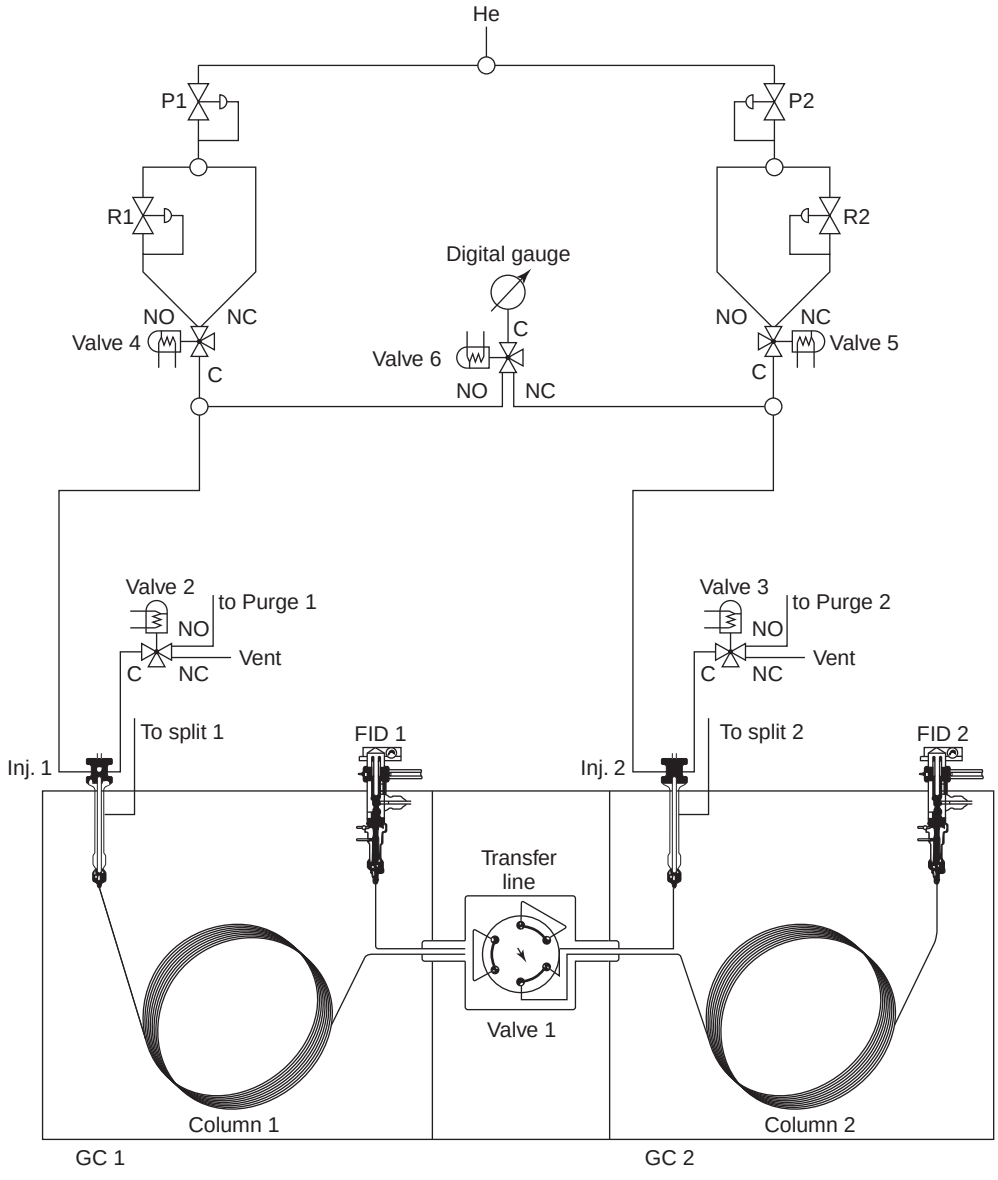

Figure 5 Pneumatic layout of a valve-based double-oven two-dimensional system. Note that the instruments can be operated separately or as a single unit. P, pressure controller; R, restrictor or needle valve; NO and NC, normally open and normally closed position, respectively. Only valve 1 (heated) is in the effluent flow path. (Reproduced by permission from Mondello et al.[33])

appears that noble metals such as Pt/Ir may not be as inert as many users imagine. There is at least one report on the tautomerization of α-ketols on a commercial 2-D GC instrument that uses a small piece of Pt/Ir in its interface.[42] Significant advances have been made in the deactivation of metal surfaces within the last 5 years. Columns are now available that can compete with fused silica in every aspect, including inertness. The controversies point out that each situation should be evaluated on a case-by-case basis.

Compared to other developments in GC, relatively little progress has been made in the advancement of 2-D GC instrumentation over the last 20 years. Commercial aspects have also been rather subdued. Most instrument manufacturers seem to be willing to modify or adapt instruments for 2-D GC. A dedicated instrument however is available from only one source, the Siemens Co. This instrument is based on a Deans switch and has a dual oven configuration. The design is largely based on the work of Schomburg and his group who have pioneered many developments in the design of 2-D GC instrumentation.[43–45] Figure 4 shows some details of the heart of this instrument, referred to as a live switch. There are several suppliers and specialty companies that

provide components for the conversion of conventional instruments to 2-D GC use. The DANI, SGE, AC Controls and Gerstel companies provide accessories and instrument modification based on the Deans principle. These firms are headquartered in Italy, Australia, Holland and Germany but have outlets in the USA and many other countries. The switching components include external computer controlled solenoid valves. Mass flow control is a preferred way to ensure the stability of switching under conditions of temperature programming.[46] The timing of effluent switching can be a very critical issue. It is sometimes important to divert a single peak, such as an unresolved pair of enantiomers, into a secondary column. One of the most elegant ways to ensure precise cutting is to use an internal standard. This principle has been implemented. The system determines the accurate position of a cut by calculation of retention indices and makes adjustments to compensate for changes.[47] The pressure–flow relationships in Deans-type switches are now well understood.[48]

The use of an intermediate trap between the columns raises challenges but also provides for additional flexibility. A trap must be capable of the retention of the target analyte and, more importantly, allow rapid reinjection. A variety of designs have been offered.[49,50] Details of the performance of individual components and instrumentation are beyond the scope of this discussion. The reader is referred to the original literature. Several reviews are available on the subject.[44,51–56] There is also extensive coverage of 2-D GC in general textbooks.[57,58] It is interesting to note that only a single review has appeared within the last 8 years.[59]

3.2 Practical Aspects

Even though the major use of 2-D GC is for the resolution of target components in complex mixtures, it is also useful in other areas. A general lack of sample capacity is a major problem associated with the use of narrow bore thin film columns. Multiple injections can be used to build up the targeted trace component(s) to a level where adequate detection or preparative scale isolation is feasible. Instruments with intermediate traps or parallel traps are particularly suitable for this purpose.[60,61] Analysis time is another major issue. Column flow can be reversed after the heartcut has taken place. Backflushing techniques can be incorporated into heartcut operations. This approach saves much time in routine analysis.

The complete characterization of a complex sample is perhaps the most difficult of all situations. The analysis of tobacco smoke is the classical example of a truly complex sample. Tobacco smoke condensate consists of thousands of substances at the parts per million level and above. The substances appear in a wide range of concentrations, are often multifunctional and may have both acidic and basic properties. Figure 6(a) shows a chromatogram of a tobacco essential oil on the first column of a two-dimensional system.[62] The resolving power of the column is modest but it is apparent that the separation, especially of minor components, would still be inadequate on a more efficient column. Figure 6(b) presents chromatograms derived from two of the heartcuts. Approximately 1500 peaks were observed and 300 were identified using mass spectra and retention indices. It is likely that secondary chromatograms still contain unresolved components. This is a situation where another separation step may be necessary.

3.3 Selected Applications

2-D GC applications have been described for many analytical problems amenable to GC. There are three distinctive areas where 2-D GC has really made an impact and is thus in routine use. These are in certain aspects of petroleum chemistry, the analysis of persistent chlorinated substances in the environment and the food/fragrance industry. The classification of crude oils and petroleum-based distillates is the oldest and most traditional application of conventional 2-D GC. Petroleum contains mainly hydrocarbons and heteroatom compounds. Both groups are of interest. Heteroatom compounds include the elements sulfur, nitrogen and oxygen. Substances of this type are usually undesirable because they are potent catalyst poisons and potential air pollutants. Petroleum-type samples are often subjected to group specific reactions or are selectively determined by spectroscopic methods. The American Society for Testing and Materials (ASTM) has issued a large number of individual methods. Hydrocarbon analysis makes use of many disciplines of chemistry. Hyphenated techniques such as liquid chromatography/gas chromatography (LC/GC) and GC/MS are particularly useful. One of the most common types of analysis involves the speciation and quantitation of paraffins, isoparaffins, olefins, naphthenes and aromatics, abbreviated as PIONA. The routine determination of petroleum products in terms of these substance classes is very important. As an example, high olefin content is undesirable because these compounds are unstable. The refinery engineer requires reliable data to optimize the blending of process streams and for many other critical operations.

Hydrocarbons in petroleum are extremely complex. There are some 4×10^7 potential paraffin isomers in the C_{10} to C_{25} carbon range alone.[63] PIONA analysis by 2-D GC is carried out in instruments that consist of a series of valves, traps, absorbers and reactors. Effluent is switched back and forth among as many as five columns. Each column is optimized to deal with

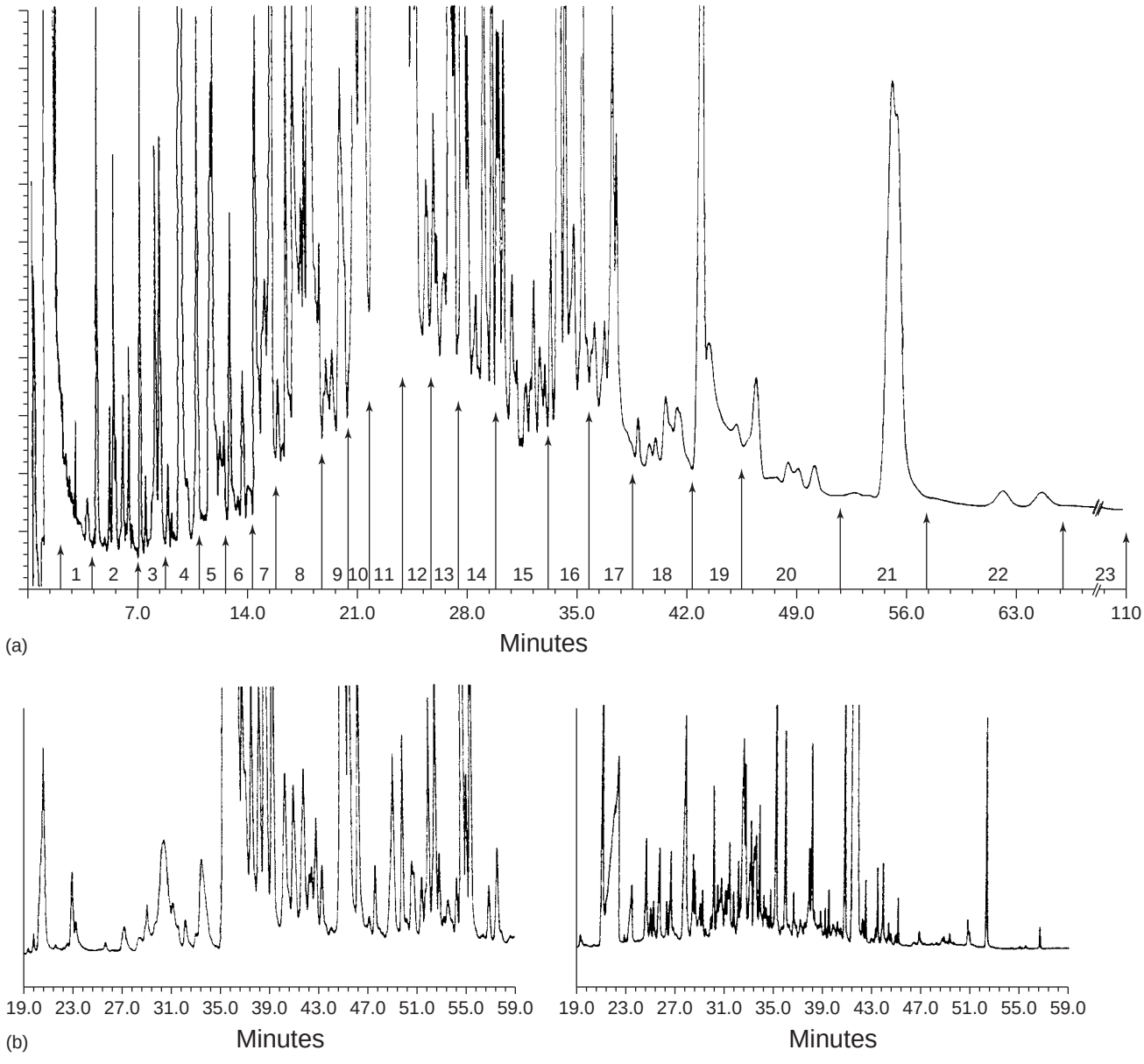

Figure 6 (a) Chromatogram of a tobacco smoke essential oil on the primary column (column 15 m × 0.53 mm × 1.0 µm, DB-Wax), indicating the position of heartcuts. (b) Secondary chromatograms (column 30 m × 0.25 mm × 0.25 µm, DB-1701) of two adjacent heartcuts 7 and 8, as indicated in (a). (Reproduced from B.M. Gordon et al., *J. Chromatogr. Sci.*, **26**, 174–180 (1988), by permission of Preston Publications, a division of Preston Industries, Inc.)

a particular range of compounds. A dedicated PIONA analyzer was originally introduced in 1971 when packed columns were widely used.[64] The basic instrument has undergone considerable refinement over the years but its basic design has changed little.[65,66] With the introduction of reliable capillary columns, packed columns are being phased out.[67–72] MS is increasingly used as an additional tool. Figure 7 shows the outcome of a typical two-dimensional separation.[73] A mass spectrometer was able to deconvolute overlapping peaks. The analysis of highly volatile materials, which has been traditionally done by gas solid chromatography[74] on packed columns, can now easily be conducted on the new generation of porous layer open tubular (PLOT) type capillary columns.[75,76]

2-D GC is also very useful in geochemistry and oil exploration.[76] The analysis of alkylnaphthalenes in crude oil is a good example to illustrate the value of 2-D GC as an indispensable tool. Alkylnaphthalenes are important indicators of geochemical parameters such as the thermal history (maturity) of sedimentary organic matter.[77] The ratio of alkylnaphthalene isomers and other fused aromatic hydrocarbons is an important indicator. These

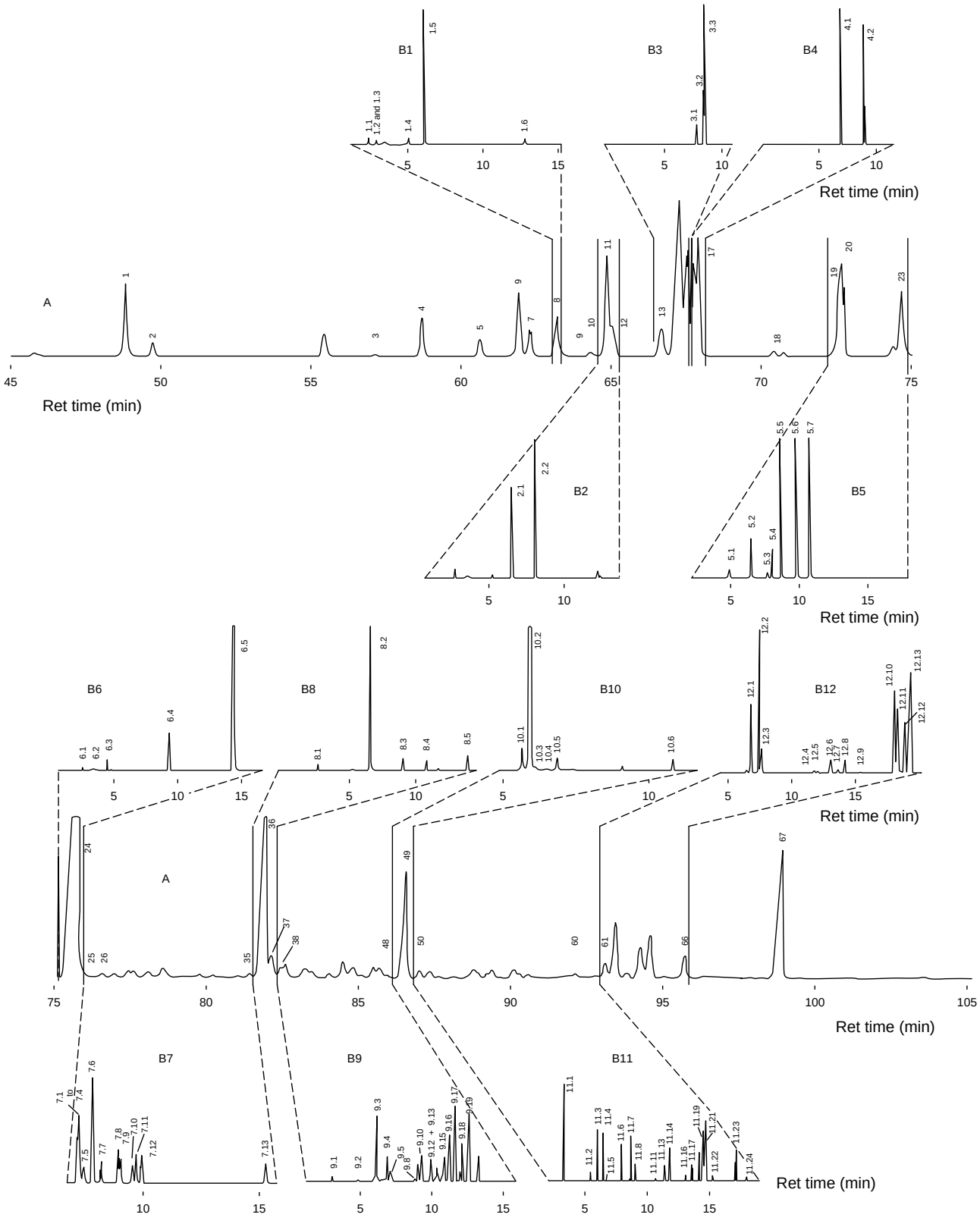

Figure 7 Chromatograms of a petroleum distillate (monoaromatic hydrocarbons). Chromatograms A (column 100 m × 0.25 mm × 1.2 µm, 5% phenyl 95% methyl polysiloxane) and B (column 25 m × 0.32 mm × 1.0 µm, free fatty acid phase) correspond to primary and secondary separations respectively. (Reproduced by permission from Hendrickx and Ramaekers.[73])

substances cannot be determined by MS because the mass spectra of the isomers are virtually identical. Only 2-D GC can achieve complete resolution.[78,79] Recently, there have been some exciting developments in comprehensive 2-D GC where some peak overlap can be accommodated without the use of valves or other switching devices. Details will be provided in a later paragraph.

The second area where 2-D GC has made a major impact is in the analysis of persistent organohalogens in the environment. The major culprits are chlorinated biphenyls, toxaphenes, dioxins, furans and some pesticides. In many cases, one has to deal with mixtures of isomers in complex matrices. The compounds are usually present at very low levels and detection by selective detectors, in particular the electron capture detector, is necessary. Analysis requires speciation, i.e. the composition of individual isomers must be determined because toxicity and biodegradation between congeners may vary by orders of magnitude. Polychlorinated biphenyls (PCBs) may contain up to 209 distinctive chlorine substituted biphenyl structures.[80] Much effort has been devoted to finding a selective stationary phase capable of resolving the majority of the 150 congeners that may be present in significant quantities.[81–83] A perfect phase has not been found. The sample is simply too complex. 2-D GC, however, can solve the problem.[84,85]

The analysis of PCBs which display axial chirality in their nonplanar conformations is an interesting and challenging problem. It is obvious that the atropisomers must be resolved on a stationary phase with chiral selectivity. Modified cyclodextrins are particularly useful for this purpose.[86] The typical two-dimensional gas chromatograph contains an apolar primary column and a chiral secondary column. Data from environmental

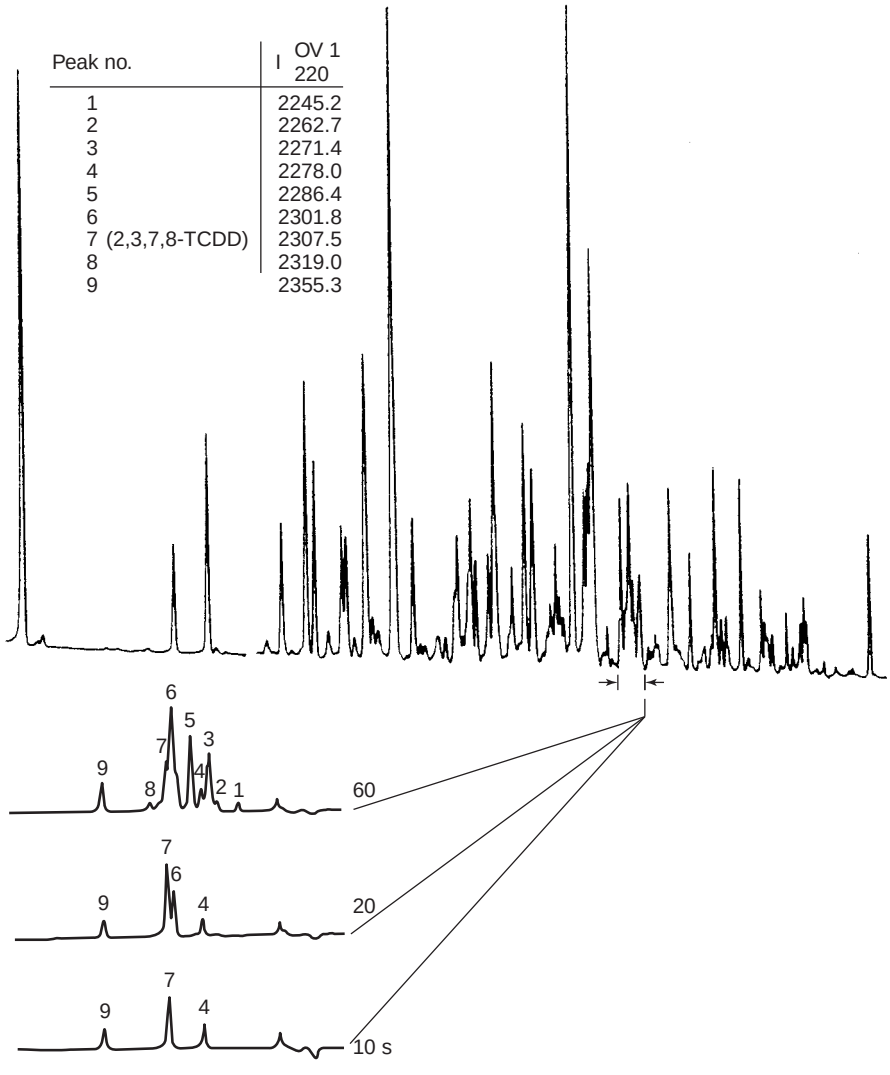

Figure 8 Influence of width of heartcut on the number of compounds isolated from a mixture of polychlorinated dioxins and furans. (Reproduced by permission from Schomburg et al.[90])

samples show that there are indeed differences in the biodegradation of different atropisomers.[87,88] The analysis of toxaphenes which are composed of borane and camphene derivatives provides similar challenges. 2-D GC is amenable to the determination of the various isomers.[89] The determination of chlorinated dioxins and benzofurans is perhaps one of the best known applications of 2-D GC. These classes include some of the most toxic substances ever recorded. Figure 8 shows an application from Schomburg's laboratory dating to the mid-1980s.[90] It is apparent that the width of the window selected for heartcutting has a major influence on the outcome of the analysis. The example shows that the heartcut containing the substance of interest, the 2,3,4,8-congener, must be chosen with precision. Photodegradation studies also rely on high resolution because the fate of individual congeners must be followed to establish correlations between structure and photodegradation rates.[91] High-resolution MS and multisector MS are useful alternatives for dioxin analysis but this technology comes at considerable cost.[92,93]

Pesticide analysis is another important area of analysis where 2-D GC is useful. The trade in foods between different countries requires the routine monitoring of samples for any of the 300 compounds that have been introduced. Most pesticides contain a combination of halogen, nitrogen, phosphorus and sulfur and are thus amenable to selective detection. A combination of electron capture detector, flame photometric detector and nitrogen/phosphorus detector is often used. Identification is usually based on a combination of retention time and detector response ratios.[94] MS is also useful.[95] In critical cases, 2-D GC has to be used in conjunction with selective detectors.[96]

The third area where 2-D GC is widely used is in the analysis of flavors and fragrances. The human nose is a very selective detector that requires a pure substance for organoleptic evaluation. Odor thresholds vary by orders of magnitude. Enrichment may thus become necessary.[97–99] The analysis of essential oils is of considerable interest.[100–102] In practice it is often also necessary to determine the source of ''off-flavors'' in a product that has been rejected by the consumer. Quite often, the substance(s) responsible for the objectionable flavor or taste is present in minute concentration(s) and 2-D GC is the only workable solution.[30] Figure 9 presents an application of this type.[104] The defective coffee sample produced a 10-fold increase of one component. It should be noted that the primary column was intentionally overloaded to provide an adequate amount for mass spectral identification.

Many compounds produced in nature are optically active. The determination of enantiomeric ratios is important.[104–107] It is routinely carried out for a variety of fruits and other natural products.[108–110] Enantioselective 2-D GC is a powerful tool to detect adulteration in foods.[111] Nature usually produces one of the enantiomers in excess whereas synthetic products are racemates. The origin specific analysis of flavors is also possible. Under conditions of identical metabolic pathways, isotope ratios should be indicators of a specific environment. The measurement of carbon-12 to carbon-13 ratios is particularly useful. Isotope ratios provide conclusive evidence. 2-D GC can be combined with on-line isotope MS in the authentication of foods.[31,112,113] The measurement of isotopes requires high precision. Data from isotope measurements are increasingly integrated in many areas where fingerprint-type comparisons are necessary.[114] Regardless of the problem at hand, analysis

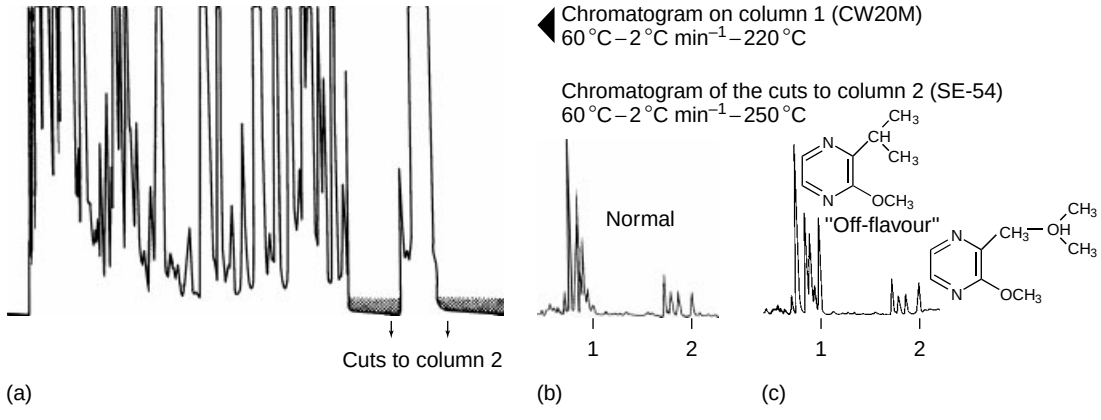

Figure 9 (a) Headspace chromatogram of defective coffee beans on primary column (CW20M), (b) selected cut of a normal coffee bean, and (c) of a defective coffee bean on secondary column (SE-54). Compound 1 was responsible for the ''peasy'' off-flavor. (Reproduced from S. Nitz et al., 'Determination of Sensorial Active Trace Compounds by Multidimensional Gas Chromatography Combined with Different Enrichment Techniques', *J. Chromatogr.*, **471**, 173–185, © 1989, with permission from Elsevier Science.)

of enantiomers requires the use of standards or retention time libraries because mass spectra of enantiomers are identical.[115]

A considerable body of data has been assembled in the three areas for which some examples have been presented. Much of the information is published in specialty journals. The reader may want to contact the appropriate original sources for additional details.

4 COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY

Conventional 2-D GC has inherent limitations. Single heartcuts yield unambiguous results. Narrow heartcuts obviously provide the highest resolution potential because the transferred effluent is likely to contain only a few components. As the window of the heartcut widens, the probability of overlap in the second dimension also increases. The basic tradeoffs have been demonstrated in Figure 8. Addition of an intermediate trap does not fundamentally change the situation. The primary advantage of an intermediate trap is its ability to reverse the effects of band broadening from the first column. To save time, users often cut multiple effluent fractions into the intermediate trap. This practice is usually acceptable if the different heartcuts are far apart. The chances of the mixing of compounds from different heartcuts on the secondary column depends primarily on the spacing of the cuts and the difference in stationary phase polarities. The chance of interferences is minimal if only widely distanced heartcuts are taken. Accidental overlap from adjacent fractions can be eliminated by diversion of the effluent into a series of parallel traps or into separate secondary columns. Both of these operational modes have problems. Multitrap arrangements require much time because analysis is sequential. Elution from the secondary column must be complete before the contents of the next trap are introduced. Parallel columns do not have this time penalty but the cost and complexity of the system increase.

The complete characterization of a wide boiling range sample can be accomplished by comprehensive 2-D GC. The term ''comprehensive'' implies that the entire sample is subjected to periodic heartcuts in a fashion analogous to routine GC/MS. The principle of the operation has been illustrated in Figure 3. In essence, effluent from the first column is injected at a fixed frequency into a single secondary column. The frequency of transfer is analogous to the scan rate in GC/MS. Dispersion in 2-D GC occurs according to stationary phase interaction as opposed to ''mass dispersion'' in GC/MS. Both processes are quite similar in principle. The implementation of comprehensive 2-D GC however is quite difficult. The major challenge comes from the need to inject the effluent into the second column as a series of narrow pulses, followed by rapid elution. Narrow pulses can be generated by several methods. The most straightforward approach is to use a valve, fluidic switch or similar device.[116] Diaphragm-type valves can produce precisely defined pulses at a high rate and are quite suitable for this purpose. Unfortunately, most of the sample is lost.[117,118] A more promising approach is the use of a modulator which is capable of both focusing and reinjection of the analyte. Figure 10 shows the basic arrangement of a comprehensive two-dimensional gas chromatograph based on this principle.[119] The modulator accessory is commercially available.[120] Components eluting from the first column are focused on a short section of a fused silica column which is coated with a thick layer of stationary phase. This device, called a modulator tube, is inserted between the columns. The components are released from the modulator tube and injected into the secondary column by application of heat transferred from a mechanical sweeper. The rate of movement of the sweeper is limited by the heater temperature and the heating time constant of the column. The primary column is usually a conventional capillary column which is programmed slowly. The secondary column typically is a narrow-bore, high-speed column capable of elution within a few seconds or tens of seconds. High speed is necessary because elution from the secondary column must be complete or nearly complete before the next injection can be made. Figure 11 presents some details on the formation of these pulses.[121] The process can be divided into four segments.

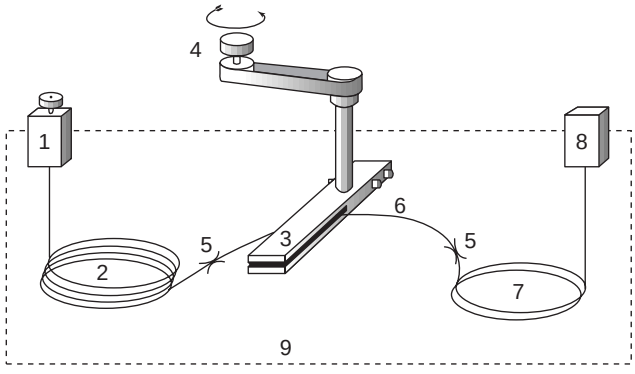

Figure 10 Schematic diagram of basic comprehensive 2-D GC instrumentation based on a mechanical sweeper: 1, inlet; 2, primary column; 3, heated slit of sweeper; 4, motor drive of sweeper; 5, column connectors; 6, modulator tube; 7, secondary column; 8, detector; 9, GC oven.

1. Accumulation. The analyte moving into the modulator tube forms a sharp leading edge due to stationary phase focusing.

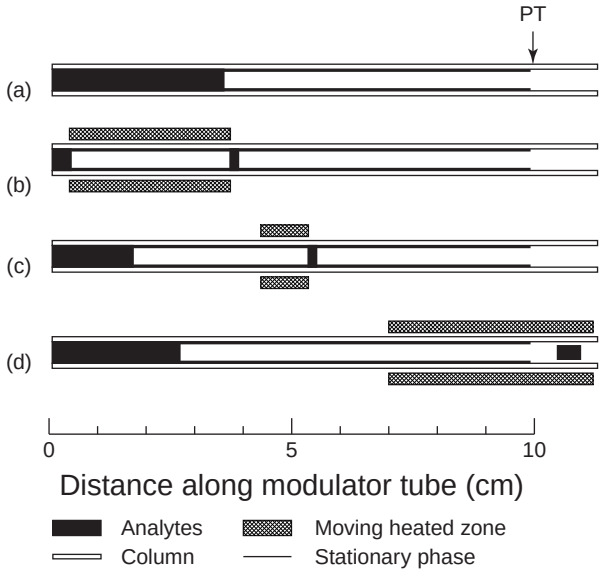

Figure 11 Detailed view of chemical pulse formation: (a) accumulation; (b) cut; (c) focus; and (d) launch. PT, phase termination. (Reproduced by permission from Phillips et al.[121])

2. Cut. The slotted heater moves over the modulator tube holding the adsorbed sample and brings the substances dissolved in the stationary phase back into the gas phase.
3. Focus. The front of the ''plug'' moves slightly ahead of the steep temperature gradient formed by the slotted heater. It is self ''sharpening''.
4. Launch. The ''plug'' eventually reaches a section of uncoated tubing or the secondary column. The phase ratio and carrier gas velocity in this section are high. The substances start to move at high speed.

All of these processes overlap and are coordinated. Several critical steps must be harmonized to make this complicated procedure a success. The components in the modulator tube must be heated rapidly, i.e. the thermal response of the system must be fast. This requires that heat transfer is rapid. In essence, it is necessary to heat and cool the modulator tube in a fraction of a second. Fortunately, fused silica has low thermal mass. Time constants of the order of 100 ms can be achieved, depending on factors such as column dimension and the temperature of the slotted heater.[121] Effective heat transfer requires that the distance between the surfaces of the slotted heater and the modulator tube is minimized. Factors such as the dimension of the connectors between the columns and the modulator tube become critical.

Prior to the invention of the rotating heater, a metal coated electrically heated column was used for GC × GC modulation. Electrical heating of metal coated modulator tubes seems to be an attractive alternative to mechanical devices. Thermal modulation was the first approach that was used to produce narrow pulses.[122–125] Electrical heating is attractive because the amount of transmitted energy can easily be controlled. It is, however, difficult to match the thermal requirements of the fused silica with the thermal mass of the surrounding film of metal. Relatively high temperatures are required for the formation of narrow pulses. This places considerable stress on the fragile components and causes frequent burnout. Metal lined tubing and miniature wire coil heaters are more stable but heat transfer characteristics become unsatisfactory because of high thermal mass.[126]

Another approach for the formation of pulses by mechanical means has also been offered. It is based on cryogenic focusing and uses a movable trap.[127] The accumulation and focusing phases are combined.

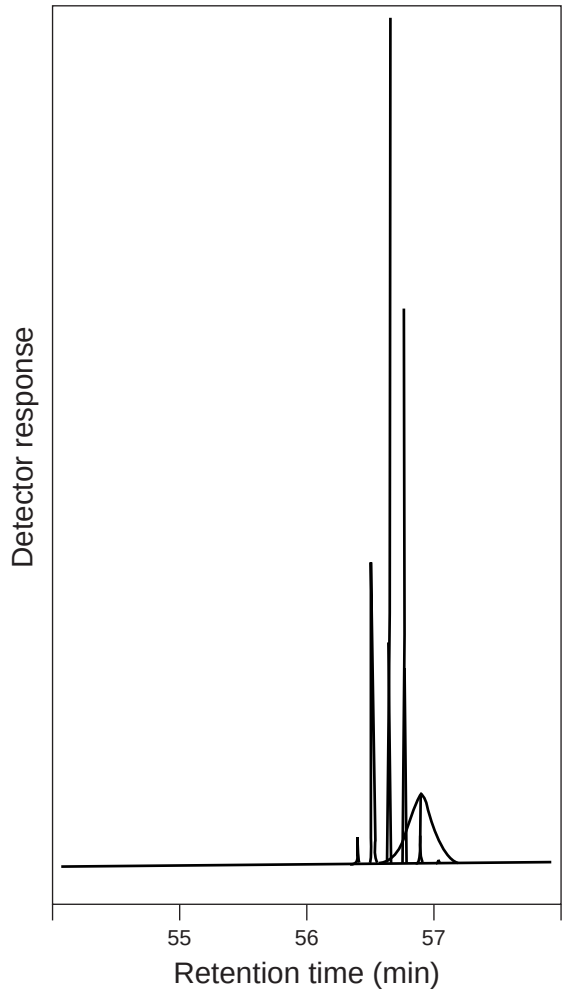

Figure 12 Effect of thermal modulation or peak widths and peak intensities. The low and broad peak is not modulated. The high and sharp peaks are the result of modulation of the same compound. (Reproduced by permission from Beens et al.[132])

The substances are moving toward a gradient of low temperature and are retained until the cryotrap is displaced by mechanical movement. The substances are then launched by being exposed to the temperature of the environment in the GC oven. An external source of heat is not required. This system is referred to as a longitudinally modulating cryogenic system (LMCS). It seems to be versatile and effective.[128–130] A number of advantages are claimed for this approach.[131] The method seems to be particularly suitable for highly volatile compounds which may not experience the necessary degree of stationary phase focusing on a thick film modulator tube at ambient temperature. Data available at this time are insufficient to predict which of the two types of mechanical modulators will eventually become more successful.

Modulation provides some unexpected benefits. It is possible to reverse the effects of peak dispersion caused by

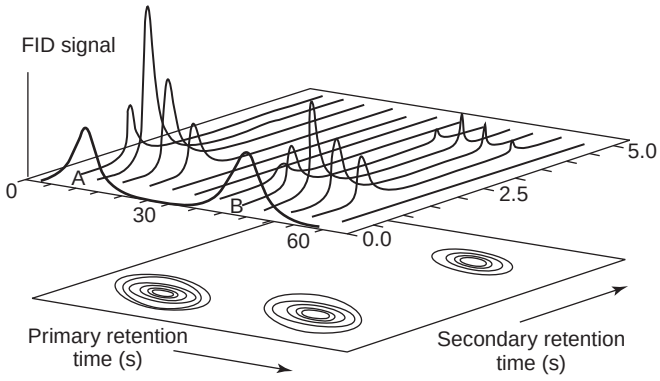

Figure 13 Presentation of data in comprehensive 2-D GC. The one-dimensional (collapsed) GC trace is displayed on the front plane. Sequential GC × GC traces are shown on the top; contour plots are shown on the bottom plane.[135] FID, flame ionization detection.

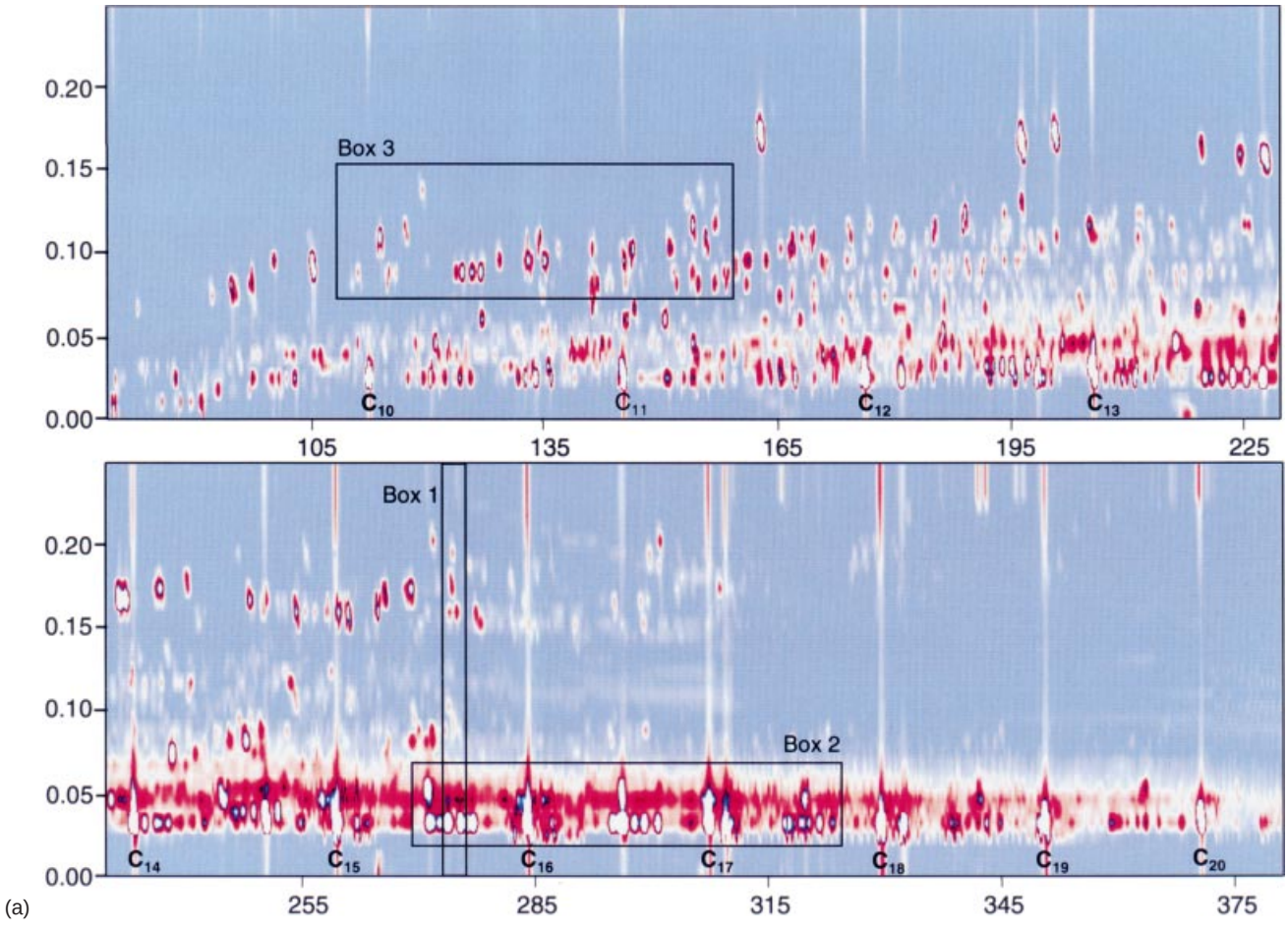

Figure 14 (a) Total ion GC × GC/MS of a marine diesel fuel sample. The *x*-axis (min) reflects volatility based retention on the apolar primary column (13 m × 0.10 mm × 1.3 µm, dimethylpolysiloxane); the *y*-axis (min) reflects polarity based retention on a polar secondary column (2 m × 0.10 mm × 0.10 µm, cyanopropylmethylpolysiloxane). Both columns programmed 30–250 °C at 0.5 °C min^{-1}. Run time 440 min. Thermal modulator programmed at $\Delta T = 100$ °C above oven temperature, velocity 0.2 rev s^{-1}, period 14.4 s. MS conditions: scan 45–350 amu, 2.43 scans s^{-1}. (b) Enlargement of box 1. (c) Enlargement of box 2. (d) Enlargement of box 3. The peaks in the chromatogram are represented in the bottom window by shapes corresponding to the molecular ion of each peak. (Reproduced by permission from Frysinger et al.[140])

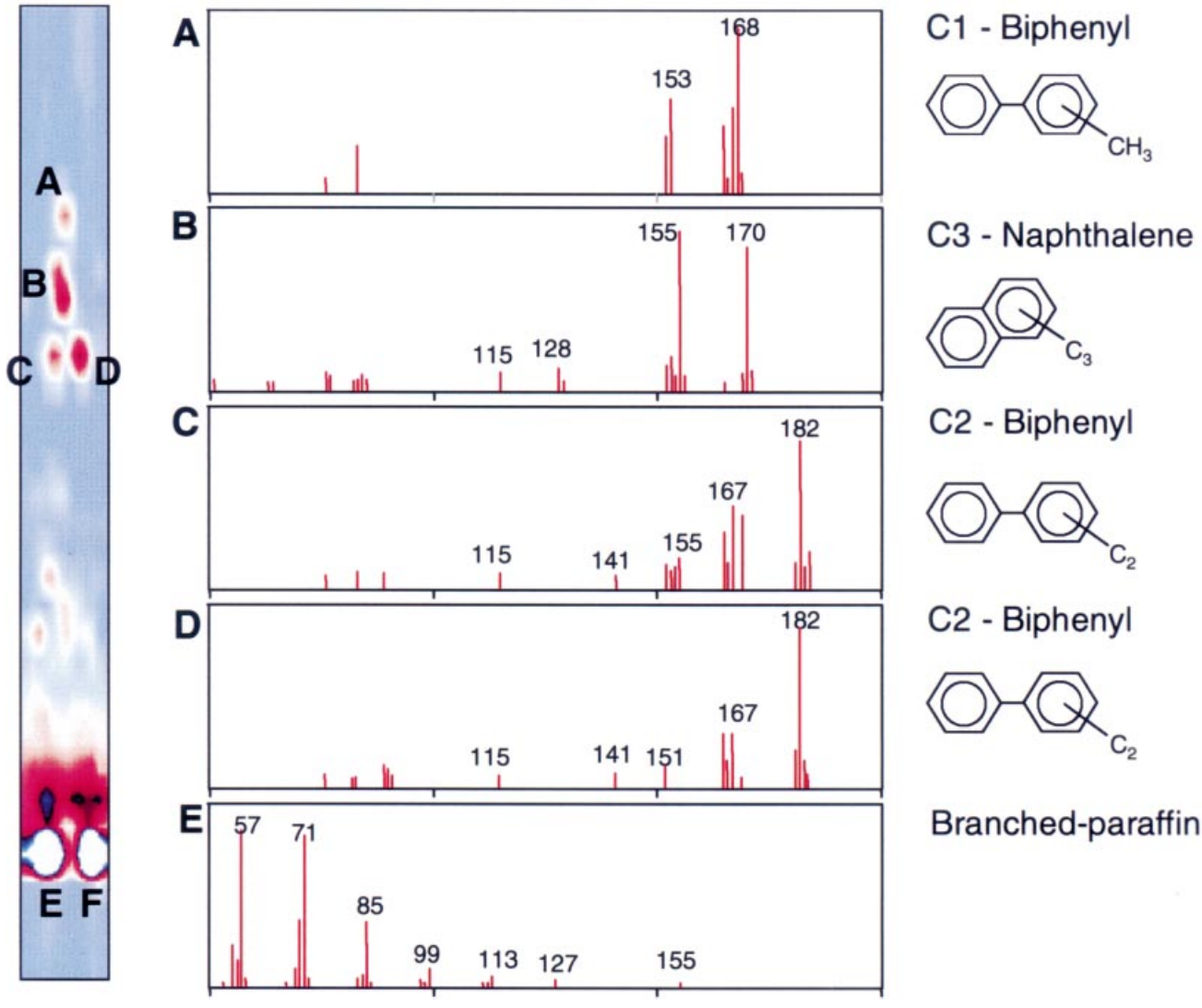

(b)

Figure 14 (*Continued*)

migration in the primary column. In essence, a wide peak can be sharpened into a series of narrow peaks. Figure 12 shows an overlay of peaks from two experiments.[132] A wide peak with a peak width of approximately 30 s is focused and periodically reinjected into a secondary column in the form of narrow pulses. Because of the law of mass conservation, peak areas must be the same. The slices from the individual focus/reinjection cycles must be integrated and summed for quantitation. The benefit of the process is a better signal-to-noise ratio and thus lower detection limit.[133]

At the present time, data acquisition and presentation of signals of the type shown in Figure 12 still present problems. One must keep in mind that an enormous volume of data is generated at a very rapid rate. The situation may be comparable to the early days of GC/MS. Most electrometers on currently available gas chromatographs do not allow proper filtering and noise rejection.[134] In conventional GC instrumentation, the analog signal from the detector is translated into digital pulses which are subjected to a low frequency pass filter below being reconverted into the final output signal.[132] This process produces large time constants and conventional equipment may not be capable of following rapidly changing signals. To overcome the problem, the signal has to be extracted from the electrometer board before the band pass filter. This results in increased noise. Detection limits are improved but the gain is not as significant as one may expect by looking at Figure 12. Nevertheless, this problem can be solved, in principle, by the introduction of fast amplifiers. A new generation of fast electronics is currently being introduced to the market.

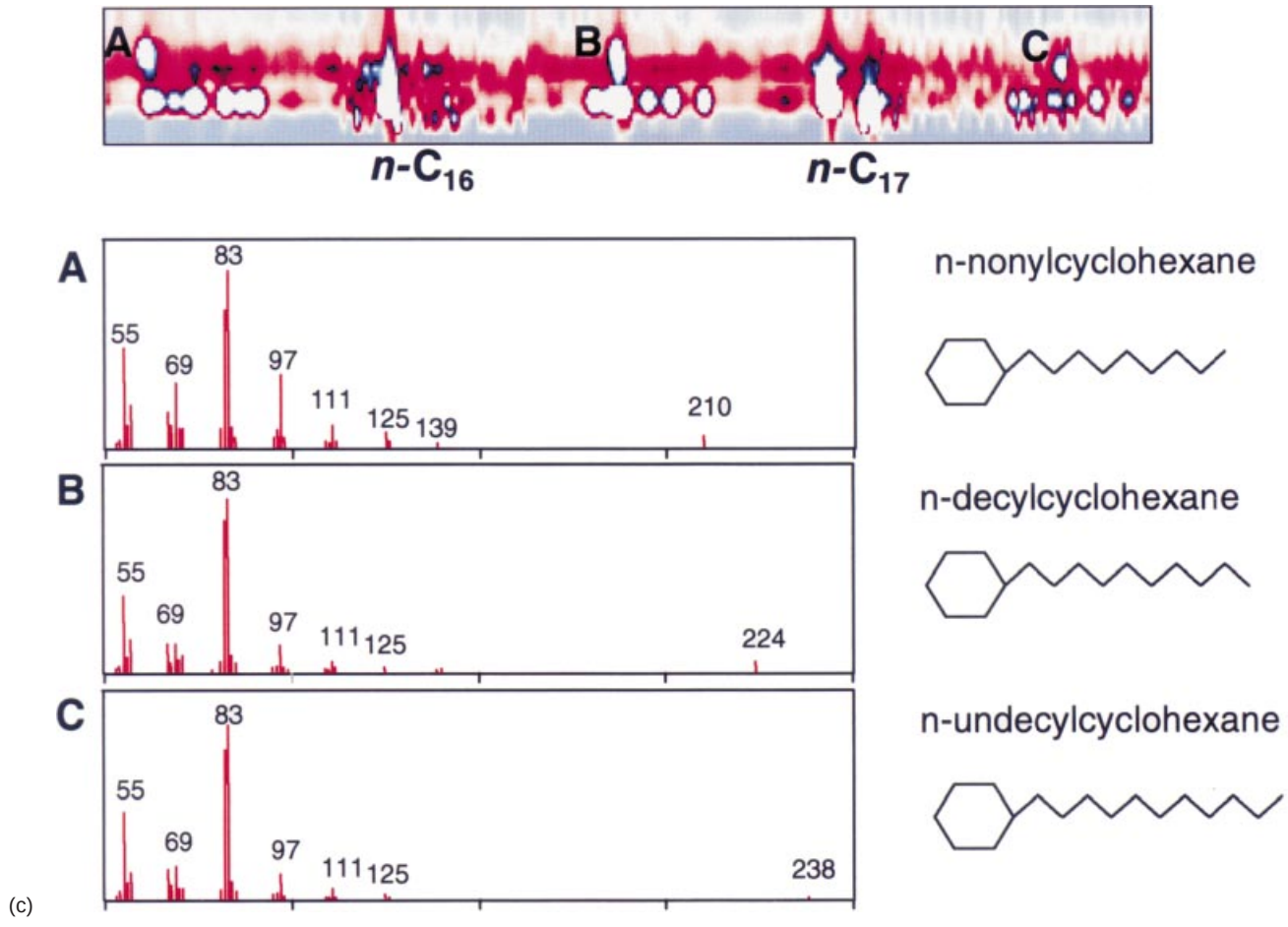

Figure 14 (*Continued*)

Quantitation and data treatment in comprehensive 2-D GC require a different approach. Figure 13 shows an excerpt from a hypothetical chromatogram.[135] Each peak is cut into slices which are then subjected to individual analysis in the second dimension. Signal A consists of a single component whereas signal B contains two individual substances. The data can be displayed as three-dimensional plots or as contour plots.[136] Integration requires summation of individual slices.

Groups of components with similar properties tend to fall into well defined planes within the chromatogram. This is a consequence of intrinsic sample dimensionally, a property predicted by Giddings.[137] Groups of components with similar structural features can thus be integrated by delineation of boundary conditions.[138] The chromatograms produce order.[139] Figure 14(a) shows an excerpt of a comprehensive 2-D GC chromatogram reflecting the C_{10} to C_{20} range of a petroleum distillate.[140] This chromatogram is particularly interesting because it represents the first application of a mass spectrometer as a detector in comprehensive 2-D GC. The *x*-axis and *y*-axis represent retention in the primary and in the secondary dimension, respectively. Components at the bottom of the chromatogram experience the lowest retention on the polar secondary column and are thus the least polar substances in the mixture. They represent alkanes. Figure 14(b) shows an enlargement from the chromatogram section delineated by box 2. Peaks A, B, and C are situated slightly above the group of substances at the bottom. They represent a homolog series of alkylcyclohexanes which are slightly more polar than *n*-alkanes and most of the branched alkanes. Figure 14(c) shows a vertical slice (box 1). Peaks E and F have the lowest and peak A has the highest retention on the polar stationary phase. These compounds, representing different structures, have been separated according to polarity. They have similar boiling points and are thus unresolved on the apolar stationary phase. Figure 14(d) corresponds to box 3 which is displaced in the *y*-direction. These substances are obviously quite polar. The schematic presentation at the bottom of the figure shows that compounds sharing similar features fall into well defined spaces. Mass spectra confirm the assignments. The order produced in the process can be a valuable aid in

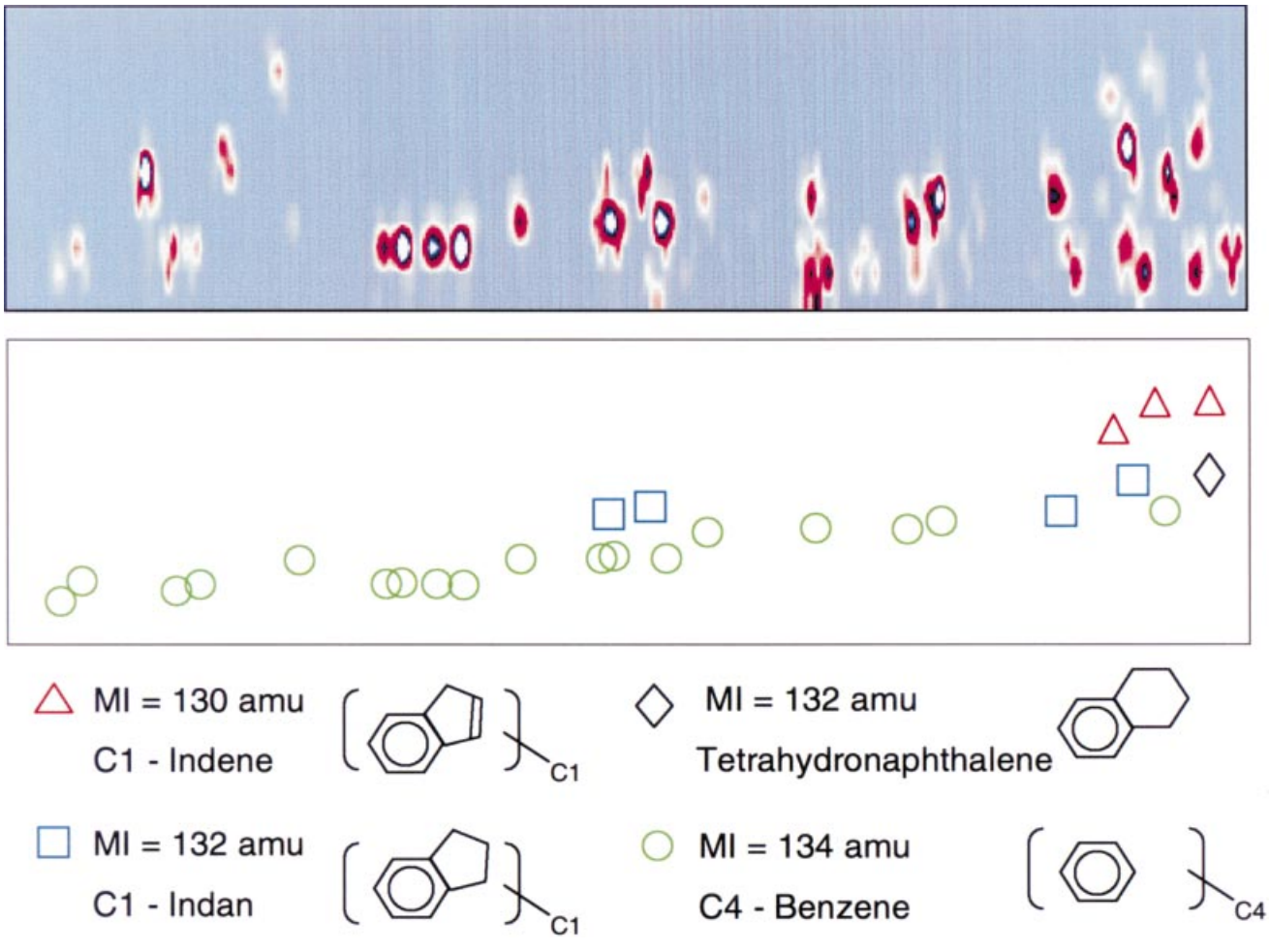

(d)

Figure 14 (*Continued*)

identification. Software for display and integration of flame ionization detection based data has now been developed. It includes options such as interactive zoom in/zoom out features, visual clues for peak start/peak stop, peak area integration routines, etc.[138] The distribution of components in two-dimensional space can be predicted.[136]

4.1 Operational Aspects and Future Developments

At the present time, comprehensive 2-D GC is still in its early stages of development. The practicality of MS detection has been demonstrated but the scan rates of conventional mass spectrometers are too slow to deal with the high rate at which the data are being generated. A new generation of rapid scan time-of-flight (TOF) type mass spectrometers can solve this problem.[141] These instruments are capable of generating up to 100 scans per second. Flame ionization detection is the most practical approach at this time. It is necessary to adapt and miniaturize other detectors, in particular the electron capture detector, to the environment of comprehensive 2-D GC. Comprehensive two-dimensional gas chromatograms of complex mixtures such as PCBs are ordered.[142] Components are thus more easily identified than in chromatograms where the order of elution is not structured. Chemometric methods of data treatment such as the Generalized Rank Annihilation Method (GRAM) can be used to deconvolute reduced resolution chromatograms.[117,118] The required run-to-run retention time precision, signal intensity and similar factors can be met by comprehensive 2-D GC. Independent temperature programming of the individual columns can provide optimal use of space and provide truly orthogonal information.[143]

Many chromatographers are not familiar with comprehensive 2-D GC. Applications have been limited to demonstrations in the fields of petroleum analysis. Two short reviews have been published in the mid-1990s

on the subject matter.[28,144] A thorough up-to-date review emphasizing fundamental aspects of 2-D GC has appeared after original submission of this article.[145] It was initiated by J.B. Phillips who has since passed away. J. Beens consequently finished the review. The future development of comprehensive 2-D GC looks very promising.

ACKNOWLEDGMENTS

I would like to than E.B. Ledford and G.S. Frysinger for valuable input and help with some of the figures and references. The late J.B. Phillips freely shared his experiences and provided unpublished data. J.H. Sandy provided the data for Table 2.

LIST OF SYMBOLS

α	Selectivity and saturation
H	Height equivalent of a theoretical plate
k'	Capacity factor
L	Column length
m	Number of expected components
N	Number of theoretical plates
n_c	Peak capacity
p	Number of observable peaks
P_1	Probability of obtaining a single component
R_s	Resolution
s	Number of peaks containing a single component
θ	Variable, determined by retention time range

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GRAM	Generalized Rank Annihilation Method
LC/GC	Liquid Chromatography/Gas Chromatography
LMCS	Longitudinally Modulating Cryogenic System
MS	Mass Spectrometry
PCBs	Polychlorinated Biphenyls
PIONA	Paraffins, Isoparaffins, Olefins, Naphthenes and Aromatics
PLOT	Porous Layer Open Tubular
SMO	Statistical Model of Overlap
TOF	Time-of-flight
2-D GC	Two-dimensional Gas Chromatography

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Dioxin-like Compounds, Screening Assays • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Food **(Volume 5)**
Flavor Analysis in Food

Pesticides **(Volume 7)**
Multiclass, Multiresidue Analysis of Pesticides, Strategies for

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. M.J.E. Golay, 'Vapor Phase Chromatography and the Telegrapher's Equation', *Anal. Chem.*, **29**, 928–934 (1957).
2. D.H. Desty, J.N. Haresnip, B.H.F. Whyman, 'Glass Capillary Columns for Gas Chromatography', *J. Inst. Petrol.*, **48**, 287–298 (1959).
3. K. Grob, 'The Glass Capillary Column in Gas Chromatography. A Tool and a Technique', *Chromatographia*, **8**(89), 423–432 (1975).
4. K. Grob, *Making and Manipulating Capillary Columns for Gas Chromatography*, Hüthig, Heidelberg, 1986.
5. R.D. Dandenau, E.H. Zerenner, 'An Investigation Into Glasses for Capillary Chromatography', *J. High Resolut. Chromatogr.*, **2**, 351–356 (1979).
6. M.C. Simmons, L.R. Snyder, 'Two Stage Gas–Liquid Chromatography', *Anal. Chem.*, **30**, 32–35 (1958).
7. Z. Liu, J.B. Phillips, 'Comprehensive Two-dimensional Gas Chromatography Using an On-column Thermal Modulator Interface', *J. Chromatogr. Sci.*, **29**, 227–231 (1991).
8. ZEOX Corp., Lincoln, NE, USA.

9. J.B. Phillips, Z. Liu, 'Apparatus and Method of Multidimensional Chemical Separation', US Patent 5 196 039, 1993.
10. J.B. Phillips, Z. Liu, 'Chromatographic Technique and Apparatus', US Patent 5 135 549, 1992.
11. J.C. Giddings, 'Maximum Number of Components Resolvable by Gel Filtration and Other Elution Chromatographic Methods', *Anal. Chem.*, **39**, 1027–1028 (1967).
12. J. Krupcik, J. Garaj, P. Cellar, G. Guiochon, 'Calculation of the Peak Capacity in Capillary Gas Chromatography', *J. Chromatogr.*, **312**, 1–10 (1984).
13. J.M. Davis, J.C. Giddings, 'Statistical Theory of Component Overlap in Multicomponent Chromatograms', *Anal. Chem.*, **55**, 418–424 (1983).
14. J.M. Davis, 'Statistical Theories of Peak Overlap in Chromatography', in *Advances in Chromatography*, eds. P.R. Brown, E. Grushka, 109–176, Vol. 34, 1994.
15. F. Dondi, A. Bassi, A. Cavazzini, M.C. Pietrogrande, 'A Quantitative Theory of the Statistical Degree of Peak Overlapping in Chromatography', *Anal. Chem.*, **70**, 766–773 (1998).
16. P.J. Sloneker, X. Li, T.H. Ridgway, J.G. Dorsey, 'Information Orthogonality of Two-dimensional Chromatographic Separations', *Anal. Chem.*, **68**, 682–689 (1996).
17. T.A. Berger, 'Separation of a Gasoline on an Open Tubular Column with 1.3 Million Effective Plates', *Chromatographia*, **42**(1/2), 63–71 (1996).
18. C.A. Cramers, P.A. Leclercq, 'Strategies for Speed Optimization in Gas Chromatography: an Overview', *J. Chromatogr.*, **842**, 3–13 (1999).
19. M. Proot, F. David, P. Sandra, M. Verzele, 'Temperature Dependence of the Resolution of Diastereomeric Isoprenoids in Capillary GC Experiments with 50 µm, 100 µm, and 250 µm i.d. Columns', *J. High Resolut. Chromatogr. Chrom. Commun.*, **8**, 426–430 (1985).
20. W. Jennings, J.A. Settlage, R.J. Miller, O.G. Raable, 'New Approach to Chromatographic Optimization in Glass Capillary Gas Chromatography', *J. Chromatogr.*, **186**, 189–196 (1979).
21. D. Roberts, W. Bertsch, 'Determination of Extra Column Effects in Recycle Gas Chromatography with Capillary Columns', *J. High Resolut. Chromatogr. Chrom. Commun.*, **11**, 783–787 (1988).
22. J.C. Giddings, 'Concepts and Comparisons in Multidimensional Separation', *J. High Resolut. Chromatogr. Chrom. Commun.*, **10**, 319–323 (1987).
23. N. Ragunathan, T.A. Sasaki, K.A. Krock, C.L. Wilkins, 'Multidimensional Fast Gas Chromatography with Matrix Isolation Infrared Detection', *Anal. Chem.*, **66**, 3751–3756 (1994).
24. S. Nitz, F. Drawert, M. Albrecht, 'Micropreparative System for Enrichment of Capillary GC Effluents', *J. High Resolut. Chromatogr.*, **11**, 322–327 (1988).
25. E.K. Ortner, E.R. Rohwer, 'Trace Determination of Organic Compounds in Water by Direct Enrichment in Multichannel Thick Film Traps with Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **22**, 521–526 (1999).
26. J.C. Giddings, 'Use of Multiple Dimensions in Analytical Separations', in *Multidimensional Chromatography, Techniques and Application*, ed. H.J. Cortes, Marcel Dekker, New York, Vol. 50, 1990.
27. J.B. Phillips, E.B. Ledford, 1998, personal communication.
28. H.-J. de Geus, J. de Boer, U.A.T. Brinkman, 'Two-dimensional Gas Chromatography for Environmental Analysis', *Trends Anal. Chem.*, **15**, 398–408 (1996).
29. D.R. Deans, 'A New Technique for Heart Cutting in Gas Chromatography', *J. Gas Chromatogr.*, **1**, 18–22 (1968).
30. S. Nitz, H. Kollmannsberger, F. Drawert, 'Determination of Sensorial Active Trace Compounds by Multidimensional Gas Chromatography Combined with Different Enrichment Techniques', *J. Chromatogr.*, **471**, 173–185 (1989).
31. S. Nitz, B. Weinreich, F. Drawert, 'Multidimensional Gas Chromatography – Isotope Ratio Mass Spectrometry (MDGC-IRMS). Part A: System Description and Technical Requirements', *J. High Resolut. Chromatogr.*, **15**, 387–391 (1992).
32. D. Juchelka, T. Beck, U. Hener, F. Dettmar, A. Mosandl, 'Multidimensional Gas Chromatography Coupled On-line with Isotope Ratio Mass Spectrometry (MDGC-IRMS): Progress in the Analytical Authentication of Genuine Flavor Components', *J. High Resolut. Chromatogr.*, **21**, 145–151 (1998).
33. L. Mondello, M. Catalfamo, A. Cotroneo, G. Dugo, H. McNair, 'Multidimensional Capillary GC/GC for the Analysis of Real Complex Samples', *J. High Resolut. Chromatogr.*, **22**, 350–356 (1999).
34. L.V. Ligon, R.J. May, 'Isomer Specific Analysis of Selected Chlorodibenzofurans', *J. Chromatogr.*, **294**, 87–94 (1984).
35. W.V. Ligon, R.J. May, 'Target Compound Analysis by Two-dimensional Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **294**, 77–86 (1984).
36. W. Bertsch, F. Hsu, A. Zlatkis, 'Heartcutting Techniques in High Resolution Gas Chromatography Applied to Sulfur Compounds in Cigarette Smoke', *Anal. Chem.*, **48**, 928–931 (1995).
37. R.E. Kaiser, R.I. Rieder, 'Native Ethylene Glycol in Wine: Application of a Dead Volume Free, Very Fast ''Deans Heart-cut'' System with Multi-chromatography', *J. High Resolut. Chromatogr.*, **10**, 240–243 (1987).
38. W. Bretschneider, P. Werkhoff, 'Progress in All-glass Stream Splitting Systems in Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **11**, 543–546 (1988).
39. B.M. Gordon, C.E. Rix, M.F. Borgerding, 'Comparison of State-of-the-art Column Switching Techniques in

High Resolution Gas Chromatography', *J. Chromatogr. Sci.*, **23**, 1–10 (1985).
40. P.L. Mills, W.E. Guise, 'A Multidimensional Gas Chromatographic Method for Analysis of *n*-Butane Oxidation Reaction Products', *J. Chromatogr. Sci.*, **34**, 431–459 (1996).
41. S.T. Adam, 'Quality Test of a Mechanical Switching Valve for Two-dimensional Open Tubular Gas Chromatography', *J. High Resolut. Chromatogr. Chrom. Commun.*, **11**, 85–89 (1988).
42. D. Häring, T. König, B. Withopf, M. Herderich, P. Schreier, 'Enantio-differentiation of α-Ketols in Sherry by One- and Two-dimensional HRGC Techniques', *J. High Resolut. Chromatogr.*, **20**, 351–354 (1997).
43. G. Schomburg, F. Weeke, 'Multidimensional Gas Chromatography (MDC) in Capillary Columns Using Double Oven Instruments and a Newly Designed Coupling Piece for Monitoring Detection After Pre-separation', *Chromatographia*, **16**, 87–91 (1982).
44. G. Schomburg, 'Multidimensional Gas Chromatography as Sampling Technique', in *Sample Introduction in Capillary Gas Chromatography*, eds. W. Bertsch, W.G. Jennings, R.E. Kaiser, Hüthig, Heidelberg, 1985.
45. G. Schomburg, H. Husmann, L. Podmaniczky, F. Weeke, 'Analysis of Volatiles Methods and Applications', in *Proceedings International Workshop on Flavor Chemistry*, ed. P. Schreier, Würzburg, 1983.
46. J.P.E.M. Rijks, A.R. Rijks, 'Programmed Cold Sample Introduction and Multidimensional Preparative Capillary Gas Chromatography. Part I: Introduction, Design and Operation of a New Mass Flow Controlled Multidimensional GC System', *J. High Resolut. Chromatogr.*, **13**, 261–266 (1990).
47. K. Himberg, E. Sippola, 'Real-time Control of Column Switching in Multi-dimensional Gas Chromatography', *J. High Resolut. Chromatogr.*, **16**, 645–650 (1993).
48. J. Blomberg, U.A.T. Brinkman, 'Practical and Theoretical Aspects of Designing a Flame-ionization Detector/Mass Spectrometer Deans Switch. Pressure–flow Relations in Gas Chromatograph–Detector Interfaces Using Vacuum-outlet Conditions', *J. Chromatogr.*, **831**, 257–265 (1999).
49. E.L. Anderson, M.M. Thomason, H.T. Mayfield, W. Bertsch, 'Advances in Two-dimensional Gas Chromatography with Glass Capillary Columns', *J. High Resolut. Chromatogr. Chrom. Commun.*, **2**, 335–342 (l979).
50. J.A. Settlage, W.G. Jennings, 'Inexpensive Method for Rapid Heating of Cold Trap', *J. High Resolut. Chromatogr. Chrom. Commun.*, **3**, 146 (1980).
51. W. Bertsch, 'Methods in High Resolution Gas Chromatography. Two-dimensional Techniques', *J. High Resolut. Chromatogr. Chrom. Commun.*, **1**, 85–91, 187–198, 289–300 (1978).
52. G. Schomburg, 'Multidimensional Capillary Gas Chromatography', *LC-GC*, **5**(4), 304–316 (1986).
53. D.R. Deans, 'Use of Heart Cutting in Gas Chromatography: a Review', *J. Chromatogr.*, **203**, 19–28 (1981).
54. W. Bertsch, 'Multidimensional Gas Chromatography', in *Multidimensional Chromatography*, ed. H.J. Cortes, Marcel Dekker, New York, 74–144, 1990.
55. U.K. Goekeler, F. Müller, 'Process Multidimensional Gas Chromatography', in *Multidimensional Chromatography*, ed. H.J. Cortes, Marcel Dekker, New York, 191–217, 1990.
56. K. Himberg, E. Sippola, M.L. Riekkola, 'Multidimensional Gas Chromatography: State of the Art', *J. Microcol. Sep.*, **1**, 271–277 (1989).
57. F.I. Onuska, F.W. Karasek, *Open Tubular Column Gas Chromatography in Environmental Sciences*, Plenum Press, New York, 1984.
58. W.V. Ligon, 'Multidimensional Gas Chromatography: Techniques and Applications', in *Gas Chromatography: Biochemical, Biomedical and Clinical Applications*, ed. R.E. Clement, Wiley, New York, 55–85, 1990.
59. G. Schomburg, 'Two-dimensional Gas Chromatography: Principles, Instrumentation, Methods', *J. Chromatogr.*, **703**, 309–325 (1995).
60. C. Bicchi, A. D'Amato, C. Frattini, G.M. Nano, A. Pisciotta, 'Some Applications of Multidimensional Gas Chromatography to the Enrichment of Minor Components in Essential Oil Analysis', *J. High Resolut. Chromatogr.*, **12**, 705–708 (1989).
61. K.D. Kempfert, 'Evaluation of Apparent Sensitivity Enhancement in GC/FTIR Using Multidimensional GC Techniques', *J. Chromatogr. Sci.*, **27**, 63–70 (1989).
62. B.M. Gordon, M.S. Uhrig, M.F. Borgerding, H.L. Chung, W.M. Coleman, J.F. Elder, J.A. Giles, D.S. Moore, C.E. Rix, E.L. White, 'Analysis of Flue-cured Tobacco Essential Oil by Hyphenated Analytical Techniques', *J. Chromatogr. Sci.*, **26**, 174–180 (1988).
63. H.-J. Neumann, B. Paczynska-Lahme, D. Severin, *Composition and Properties of Petroleum*, Ferdinand Enke, Stuttgart, 1981.
64. H. Boer, P. van Arkel, 'Multidimensional Gas Chromatography in the Petroleum Industry', *Chromatographia*, **4**, 300–306 (1971).
65. H. Boer, *Exercises in Polydimensional Gas Chromatography*, ed. S.G. Perry, Applied Science, Barking, 1973.
66. P. van Arkel, J. Beens, J. Spaans, D. Grutterink, R. Verbeck, 'An Improved Multidimensional Gas Chromatograph for PONA Analysis', *J. Chromatogr. Sci.*, **26**, 267–275 (1988).
67. T.J. Lechner-Fish, S.L. Ryder, 'Analysis of Simulated Petroleum Wellhead Fluids Using Multidimensional Gas Chromatography', *Am. Lab.*, **2**, 33X–33EE (1997).
68. J.J. Szakasits, R.E. Robinson, 'Hydrocarbon Type Determination of Naphthas and Catalytically Reformed Products by Automated Multidimensional Gas Chromatography', *Anal. Chem.*, **63**, 114–120 (1991).
69. F.P. Di Sanzo, J.L. Lane, R.E. Yoder, 'Application of State-of-the-art Multidimensional High Resolution Gas

Chromatography for Individual Component Analysis of Gasoline Range Hydrocarbons', *J. Chromatogr. Sci.*, **26**, 206–209 (1988).
70. F.P. Di Sanzo, V.J. Giarrocco, 'Analysis of Pressurized Gasoline-range Liquid Hydrocarbon Samples by Capillary Column and PIONA Analyzer Gas Chromatography', *J. Chromatogr. Sci.*, **26**, 258–266 (1988).
71. J. Curvers, P. van der Sluys, 'Automated Multidimensional Gas Chromatographic PONA Analysis of Hydrocarbon-type Samples. Optimization with Respect to Speed of Analysis', *J. Chromatogr. Sci.*, **26**, 267–270 (1988).
72. S.T. Teng, A.D. Williams, K. Urdal, 'Detailed Hydrocarbon Analysis of Gasoline by GC/MS (SI-PIONA)', *J. High Resolut. Chromatogr.*, **17**, 469–475 (1994).
73. H.J.W. Hendrickx, J.J.M. Ramaekers, 'Analysis of a C_9–C_{10} Aromatic Hydrocarbon Pyrolysis Distillate by Multidimensional Capillary GC and Multidimensional Capillary GC/MS', *J. High Resolut. Chromatogr.*, **17**, 407–410 (1994).
74. P. Coleman, L.S. Ettre, 'Analysis of Gases Containing Inorganic and Organic Compounds Using a Combination of a Thick-film Capillary Column and a Packed Adsorption Column', *J. High Resolut. Chromatogr. Chrom. Commun.*, **8**, 112–118 (1985).
75. H. Tani, M. Furuno, 'Rapid Analysis of Hydrocarbons and Inert Gases by Multidimensional Gas Chromatography', *J. High Resolut. Chromatogr. Chrom. Commun.*, **9**, 712–716 (1986).
76. S. Wu, W.H. Chatham, S.O. Farwell, 'Multidimensional HRGC for Sample Components with a Wide Range of Volatilities and Polarities', *J. High Resolut. Chromatogr.*, **13**, 229–233 (1990).
77. M. Radke, J. Brooks, D.H. Welte, *Advances in Petroleum Geochemistry*, Academic Press, London, Vol. 2, 1987.
78. R.G. Schäefer, Z.J. Höltkemeier, 'Direkte Analyse von Dimethylnaphthalinen in Erdölen mittels Zweidimensionaler Kapillar-gas-chromatographie', *Anal. Chim. Acta*, **260**, 107–112 (1992).
79. R.G. Schäfer, J. Höltkemeier, 'Analysis of Alkylnaphthalenes in Crude Oils by Two-dimensional Gas Chromatography', *Chromatographia*, **26**, 311–315 (1988).
80. M.D. Erickson, *Analytical Chemistry of PCBs*, 2nd edition, CRC Lewis Publishers, Boca Raton, FL, 1997.
81. G. Frame, 'Congener-specific. Comprehensive Standard Sets Combined with Improved Capillary GC Columns and Detectors Bring Complete Characterization of PCB Mixtures within Reach', *Anal. Chem.*, **69**, 468A–475A (1997).
82. P. Hess, J. de Boer, W.P. Cofino, P.E.G. Leonards, D.E. Wells, 'Analysis of Polychlorinated Biphenyls by Gas Chromatography', *J. Chromatogr.*, **703**, 417–465 (1995).
83. B.R. Larsen, 'HRC Separation of PCB Congeners', *J. High Resolut. Chromatogr.*, **18**, 141–151 (1995).
84. G.M. Frame, 'Analysis of PCBs by Multidimensional Gas Chromatography', *Fresenius Z. Anal. Chem.*, **357**, 701–713 (1997).
85. R.M. Kinghorn, P.J. Marriott, M. Chambers, 'Multidimensional Capillary Gas Chromatography of Polychlorinated Biphenyl Marker Compounds', *J. High Resolut. Chromatogr.*, **19**, 622–626 (1996).
86. W.A. König, *Gas Chromatographic Enantiomer Separation with Modified Cyclodextrins*, Hüthig, Heidelberg, 1992.
87. E. Benika, R. Novakovsky, J. Hrouzek, J. Krupcik, P. Sandra, J. de Zeeuw, 'Multidimensional Gas Chromatographic Separation of Selected PCB Atropisomers in Technical Formulations and Sediments', *J. High Resolut. Chromatogr.*, **19**, 95–98 (1996).
88. G.P. Blanch, A. Glausch, V. Schurig, R. Serrano, M.J. Gonzalez, 'Quantification and Determination of Enatiomeric Ratios of Chiral PCB 95, PCB 132, and PCB 149 in Shark Liver Samples (C. *Coelopesis*) from the Atlantic Ocean', *J. High Resolut. Chromatogr.*, **19**, 392–396 (1996).
89. H.-J. de Geus, R. Baycan-Keller, M. Oehme, J. de Boer, U.A.T. Brinkman, 'Determination of Enantiomer Ratios of Borane Congeners in Biological Samples Using Heart-cut Multidimensional Gas Chromatography', *J. High Resolut. Chromatogr.*, **21**, 39–46 (1998).
90. G. Schomburg, H. Husmann, B. Hubinger, *J. High Resolut. Chromatogr. Chrom. Commun.*, **8**, 395–400 (1985).
91. L.D. Sivils, S. Kapila, Q. Yan, A.A. Elsewi, 'Application of a Two-dimensional Chromatography System for Gas-phase Photodegradation Studies of Polychlorinated Dibenzo-*p*-dioxins', *J. Chromatogr.*, **688**, 2221–2230 (1994).
92. W.V. Ligon, R.J. May, 'Target Compound Analysis by Two-dimensional Gas Chromatography/Mass Spectrometry', *J. Chromatogr.*, **294**, 77–86 (1984).
93. J. Iida, T. Takeda, T. Takasuga, J. Moncur, P. Ireland, B. Wright, 'Approaches to Dioxin Analysis by HRGC-HRMS and Hybrid HRGC-MS/MS', *J. High Resolut. Chromatogr.*, **14**, 103–109 (1991).
94. P. Klaffenbach, H.-J. Stan, 'Automated Screening of Food and Water Samples for Pesticide Residues by Means of Capillary GC-MSD Using a Library of Pesticide Spectra and Macro Programming', *J. High Resolut. Chromatogr.*, **14**, 754–761 (1991).
95. C. Bicchi, A. D'Amato, A. Binello, 'Identification of Pesticide Residues in Real Matrices by Combining Retention Indices with Specific Multidetector Responses', *J. High Resolut. Chromatogr.*, **19**, 80–84 (1996).
96. H.J. Stan, S. Heil, 'Two-dimensional Capillary Gas Chromatography with Three Selective Detectors as a Valuable Tool in Residue Analysis – State-of-the-art', *Fresenius Z. Anal. Chem.*, **339**, 34–39 (1991).
97. S. Nitz, H. Kollmannsberger, M. Albrecht, F. Drawert, 'Thermal Desorption and Sniffing Mass Spectrometric Monitoring of Enriched Trace Compounds by Means of

a ''Live Total Transfer System'' ', *J. Chromatogr.*, **547**, 516–522 (1991).
98. O. Nishimura, 'Application of a Thermal Desorption Cold Trap Injector to Multidimensional GC and GC/MS', *J. High Resolut. Chromatogr.*, **18**, 699–704 (1995).
99. C. Bicchi, A. D'Amato, C. Frattini, G.M. Nano, A. Pisciotta, 'Some Applications of Multidimensional Gas Chromatography to the Enrichment of Minor Components in Essential Oil Analysis', *J. High Resolut. Chromatogr.*, **12**, 705–708 (1989).
100. P.A. Rodriguez, C.L. Eddy, 'Use of Two-dimensional Gas Chromatography in the Organoleptic Evaluation of an Orange Extract', *J. Chromatogr. Sci.*, **24**, 18–21 (1996).
101. D.W. Wright, K.O. Mahler, L.B. Ballard, 'The Application of an Expanded Multidimensional GC System to Complex Fragrance Evaluations', *J. Chromatogr. Sci.*, **24**, 60–65 (1986).
102. M. Herraiz, G. Reglero, T. Herraiz, E. Loyola, 'Analysis of Wine Distillates Made from Muscat Grapes (Pisco) by Multidimensional Gas Chromatography and Mass Spectrometry', *J. Agric. Food Chem.*, **38**, 1540–1543 (1990).
103. R.H.M. van Ingen, L.M. Nijssen, 'Determination of Diethylene Glycol Monoethyl Ether in Flavors by Two-dimensional Capillary Gas Chromatography', *J. High Resolut. Chromatogr.*, **12**, 484–485 (1989).
104. C. Bicchi, A. Pisciotta, 'Use of Two-dimensional Gas Chromatography in the Direct Enantiomer Separation of Chiral Essential Oil Components', *J. Chromatogr.*, **508**, 341–348 (1990).
105. C. Wang, H. Frank, G. Wang, L. Zhou, E. Bayer, P. Lu, 'Determination of Amino Acid Enantiomers by Two-dimensional Gas Chromatography with Valveless Column Switching', *J. Chromatogr.*, **262**, 352–359 (1983).
106. X. Wang, C. Jia, H. Wan, 'The Direct Chiral Separation of Some Optically Active Compounds in Essential Oils by Multidimensional Gas Chromatography', *J. Chromatogr. Sci.*, **33**, 22–25 (1995).
107. S. Nitz, H. Kollmannsberger, B. Weinreich, F. Drawert, 'Enantiomeric Distribution and $^{13}C/^{12}C$ Isotope Ratio Determination of γ-Lactones: Appropriate Methods for the Differentation Between Natural and Non-natural Flavors', *J. Chromatogr.*, **557**, 187–197 (1991).
108. A. Mosandl, H. Hener, Hagerauer-Hener, A. Kustermann, 'Direct Enantiomer Separation of Chiral δ-Lactones from Food and Beverages by Multidimensional Gas Chromatography', *J. High Resolut. Chromatogr. Chrom. Commun.*, **12**, 532–536 (1989).
109. W.A. König, B. Gehrke, D. Ischeln, P. Evers, I. Donnecke, W. Wang, 'New Selectively Substituted Cyclodextrins as Stationary Phases for the Analysis of Chiral Constituents of Essential Oils', *J. High Resolut. Chromatogr.*, **15**, 367–372 (1992).
110. C. Askari, U. Hener, H.G. Schmarr, A. Rapp, A. Mosandl, 'Stereodifferentiation of Some Chiral Monoterpenes Using Multidimensional Gas Chromatography', *Fresenius Z. Anal. Chem.*, **340**, 768–772 (1991).
111. A. Mosandl, K. Fischer, U. Hener, P. Kreis, K. Rettinger, V. Schubert, H.G. Schmarr, 'Stereoisomeric Flavor Compounds. Chirospecific Analysis of Natural Flavors and Essential Oils Using Multidimensional Gas Chromatography', *J. Agric. Food Chem.*, **39**, 1131–1134 (1991).
112. D. Juchelka, T. Beck, U. Hener, F. Dettmar, A. Mosandl, 'Multidimensional Gas Chromatography Coupled On-line with Isotope Ratio Mass Spectrometry (MDGC-IRMS): Progress in the Analytical Authentication of Genuine Flavor Components', *J. High Resolut. Chromatogr.*, **21**, 145–151 (1998).
113. W. Meier-Augenstein, 'On-line Recording of $^{13}C/^{12}C$ Ratios and Mass Spectra in One Gas Chromatographic Analysis', *J. High Resolut. Chromatogr.*, **18**, 28–32 (1995).
114. J.L. Brazier, 'Use of Isotope Ratios in Forensic Analysis', in *Forensic Applications of Mass Spectrometery*, ed. J. Yinon, CRC Press, Boca Raton, FL, 1995.
115. G. Full, P. Winterhalter, G. Schmidt, P. Herion, P. Schreier, 'MDGC-MS: a Powerful Tool for Enantioselective Flavor Analysis', *J. High Resolut. Chromatogr.*, **16**, 642–644 (1993).
116. M.C. Harvey, R.E. Robinson, M.C. Harvey, S.D. Stearns, 'Further Studies on High Speed Switching of Chromatographic Valves', *J. Chromatogr. Sci.*, **32**, 190–194 (1994).
117. C.A. Bruckner, B.J. Prazen, R.E. Synovec, 'Comprehensive Two-dimensional High-speed Gas Chromatography with Chemometric Analysis', *Anal. Chem.*, **70**, 2796–2804 (1998).
118. B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, 'Second-order Chemometric Standardization for High-speed Hyphenated Gas Chromatography: Analysis of GC/MS and Comprehensive GC × GC Data', *J. Microcol. Sep.*, **11**, 97–107 (1991).
119. J. Beens, 'Chromatographic Couplings for Unraveling Oil Fractions', Dissertation, University of Amsterdam, Amsterdam, 1998.
120. E.B. Ledford, J.B. Phillips, 'Technical Report: High Speed Two-dimensional Gas Chromatograph', Zoex Corporation, Lincoln, NE.
121. J.B. Phillips, R.B. Gaines, J. Blomberg, F.W.M. van der Wielen, J.-M. Dimandja, V. Green, J. Granger, D. Patterson, L. Racovalis, H.-J. de Geus, J. de Boer, P. Haglund, J. Lipsky, V. Sinha, E.B. Ledford, 'A Robust Thermal Modulator for Comprehensive Two-dimensional Gas Chromatography', *J. High Resolut. Chromatogr.*, **22**, 3–10 (1999).
122. J.B. Phillips, D. Luu, J.B. Pawliszyn, G.C. Carle, 'Multiplex Gas Chromatography by Thermal Modulation of

a Fused Silica Column', *Anal. Chem.*, **57**, 2779–2787 (1985).
123. Z. Liu, S.R. Sirimanne, D.G. Patterson, L.L. Needham, J.B. Phillips, 'Comprehensive Two-dimensional Gas Chromatography for the Fast Separation and Determination of Pesticides Extracted from Human Serum', *Anal. Chem.*, **66**, 3086–3092 (1994).
124. Z. Liu, J.B. Phillips, 'Comprehensive Two-dimensional Gas Chromatography Using an On-column Thermal Modulator Interface', *J. Chromatogr. Sci.*, **29**, 227–231 (1991).
125. C.J. Venkatramani, J.B. Phillipps, 'Comprehensive Two-dimensional Gas Chromatography Applied to the Analysis of Complex Mixtures', *J. Microcol. Sep.*, **5**, 511–516 (1993).
126. H.J. de Jeus, J. deBoer, U.A.T. Brinkman, 'Development of a Thermal Desorption Modulator for Gas Chromatography', *J. Chromatogr.*, **767**, 137–151 (1997).
127. P.J. Marriott, R.M. Kinghorn, 'Longitudinally Modulated Cryogenic System. A Generally Applicable Approach to Solute Trapping and Mobilization in Gas Chromatography', *Anal. Chem.*, **69**, 2582–2588 (1997).
128. P.J. Marriott, R.M. Kinghorn, 'Studies on Cryogenic Trapping of Solutes During Chromatographic Elution in Capillary Gas Chromatography', *J. High Resolut. Chromatogr*, **19**, 403–408 (1996).
129. R.M. Kinghorn, P.J. Marriott, 'Enhancement of Signal-to-noise Ratios in Capillary Gas Chromatography by Using a Longitudinally Modulated Cryogenic System', *J. High Resolut. Chromatogr.*, **21**, 32–38 (1998).
130. R.M. Kinghorn, P.J. Marriott, 'Comprehensive Two-dimensional Gas Chromatography Using a Modulating Cryogenic Trap', *J. High Resolut. Chromatogr.*, **21**, 620–622 (1998).
131. R.M. Kinghorn, P.J. Marriott, 'High Speed Cryogenic Modulation – a Technology Enabling Comprehensive Multidimensional Gas Chromatography', *J. High Resolut. Chromatogr.*, **22**, 235–238 (1999).
132. J. Beens, H. Boelens, R. Tijssen, J. Blomberg, 'Quantitative Aspects of Comprehensive Two-dimensional Gas Chromatography (GC × GC)', *J. High Resolut. Chromatogr.*, **21**, 47–54 (1998).
133. H.J. de Geus, J. deBoer, J.B. Phillips, E.B. Ledford, U.A.T. Brinkmann, 'Increased Signal Amplitude Due to Mass Conservation in a Thermal Desorption Modulator', *J. High Resolut. Chromatogr.*, **21**, 411–413 (1998).
134. R. Annino, 'Practical Limits of High Speed Chromatography: Evaluation with a Computer Based Theoretical Model of Several Commercially Available Gas Chromatographs', *J. High Resolut. Chromatogr.*, **19**, 285–290 (1996).
135. R.B. Gaines, G.S. Frysinger, M.S. Hendrick-Smith, J.D. Stuart, 'Oil Spills Source Identification by Comprehensive Two-dimensional Gas Chromatography', *Environ. Sci. Technol.*, **33**, 2106–2112 (1999).
136. J. Beens, R. Tijssen, J. Blomberg, 'Prediction of Two-dimensional Gas Chromatographic Separations. A Theoretical and Practical Exercise', *J. Chromatogr.*, **822**, 233–251 (1998).
137. J.C. Giddings, 'Sample Dimensionality: a Predictor of Order – Disorder in Component Peak Distribution in Multidimensional Separation', *J. Chromatogr.*, **703**, 3–15 (1995).
138. G.S. Frysinger, R.B. Gaines, E.B. Ledford, 'Quantitative Determination of BTEX and Total Aromatic Compounds in Gasoline by Comprehensive Two-dimensional Gas Chromatography (GC × GC)', *J. High Resolut. Chromatogr.*, **22**, 195–200 (1999).
139. E.B. Ledford, Jr, J.B. Phillips, J. Xu, R.B. Gaines, J. Blomberg, 'Ordered Chromatograms: A Powerful Methodology in Gas Chromatography', *Am. Lab.*, **28**, 22–25 (1996).
140. G.S. Frysinger, R.B. Gaines, 'Comprehensive Two-dimensional Gas Chromatography with Mass Spectrometric Detection (GC × GC/MS). Applied to the Analysis of Petroleum', *J. High Resolut. Chromatogr.*, **22**, 251–255 (1999).
141. J. Dimandja, 1999, personal communication.
142. J. Xu, 'Comprehensive Two-dimensional Gas Chromatography: Development and Applications', Dissertation, Southern Illinois University, Carbondale, IL, 1997.
143. C.J. Venkatramani, J. Xu, J.B. Phillips, 'Separation Orthogonality in Temperature-programmed Comprehensive Two-dimensional Gas Chromatography', *Anal. Chem.*, **68**, 1486–1492 (1996).
144. J.B. Phillips, J. Xu, 'Comprehensive Multi-dimensional Gas Chromatography', *J. Chromatogr.*, **703**, 327–334 (1995).
145. J.B. Phillips, J. Beens, 'Comprehensive Two-dimensional Gas Chromatography: a Hyphenated Method with Strong Coupling Between the Two Dimensions', *J. Chromatogr.*, **856**, 331–347 (1999).

Sample Preparation for Gas Chromatography

Kornél Torkos
Eötvös Loránd University, Budapest, Hungary

1 Introduction 1
2 Gas Sample Preparation 1
2.1 Monitoring Airborne Contaminants in the Atmosphere 1
3 Liquid Sample Preparation 2
3.1 Liquid–Liquid Extraction 2
3.2 Purge-and-trap Technique 3
3.3 Headspace Extraction 3
3.4 Solid-phase Extraction 3
3.5 Solid-phase Microextraction 3
4 Solid Sample Preparation 4
4.1 Soxhlet Extraction 4
4.2 Supercritical Fluid Extraction 4
4.3 Ultrasonic-assisted Solid–Liquid Extraction 5
4.4 Matrix Solid-phase Dispersion 5
5 Derivatization Methods for Gas Chromatography 6
5.1 Acylation 6
5.2 Alkylation 6
5.3 Silylation 6
Acknowledgments 6
Abbreviations and Acronyms 6
Related Articles 7
References 7

The success of a gas chromatographic analysis mostly depends on the sample preparation procedure. Biological and environmental samples are not suitable for direct analysis. Organic analytes can be separated from matrixes by extraction methods such as gas–solid, liquid–liquid, solid–liquid, purge-and-trap, headspace, solid-phase (SPE), solid-phase microextraction (SPME), supercritical fluid (SFE) and Soxhlet extractions. These extraction methods are called phase separation methods. Matrix solid-phase dispersion (MSPD) is a special highly efficient phase separation procedure for biological matrixes such as meat, liver, muscle, vegetables, etc. The aim of the extraction procedure is not only separation of organic analytes from matrixes but also concentration in another phase.

If the organic analytes contain protic (active) hydrogen (alcohols, carboxylic acids, amines, oxo compounds, etc.) some derivatization reactions, such as acylation, alkylation and silylation, need to be carried out first. The aim of the derivatization methods is to increase the sensitivity of detection and thermal stability of the organic analytes. The other aspect of derivatization is to reduce the polarity of the compounds.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The most important part of gas chromatographic analysis is the preliminary sample preparation process. The sample preparation procedures for gas chromatography (GC) are divided into two main types:

- extraction methods
- derivatization methods.

The original samples are often too diluted, contaminated, complex, etc. and are thus not suitable for injection into the gas chromatograph.

In addition to the chemical properties of the samples, their physical characteristics also influence the preparation procedures. To analyze water or soil samples, some kind of extraction method is needed. Often, in order to increase the sensitivity of the analytical method, a derivatization step is included, particularly after extraction. The aim of the extraction is the separation of organic analytes from matrixes and their concentration in another phase. Derivatization methods need to be carried out first if the compounds contain protic (active) hydrogens (alcohols, oxo compounds, carboxylic acids, phenols, amines, etc.) as these compounds are not suitable for direct gas chromatographic analysis.

The most relevant extraction and derivatization methods are summarized in this article.

2 GAS SAMPLE PREPARATION

2.1 Monitoring Airborne Contaminants in the Atmosphere

Airborne organic contaminants can be collected on the surface of different types of adsorbent materials, such as activated charcoals, polymer adsorbents, molecular sieves, etc. or absorbed in suitable solvents. The adsorption of organic vapours on solids is used as a preconcentration method. The recommended collecting devices are adsorbent tubes, filter cassettes, and impingers. The sampling process is carried out with an oil-free sampling pump. The pumping time of the air sample

depends on the characteristics of the trapped compounds. The trapped compounds can be desorbed from the adsorbents with a solvent or by thermal desorption methods before analysis. These methods of monitoring organic vapours are used for occupational hygiene control.[1]

3 LIQUID SAMPLE PREPARATION

3.1 Liquid–Liquid Extraction

This method is used to analyze water samples. Although the instrument (a separation funnel) is simple, the method can be expensive due to the high cost of ultrapure solvents. A given volume of the water sample is shaken with a given volume of a suitable organic solvent so that organic micropollutants are transferred mainly into the organic phase (Figure 1). To increase the efficiency of the extraction and to avoid foam developing, sodium chloride is added to the sample before extraction. After separating the phases, most of the organic solvent is distilled off. This concentrating process is carried out in a Kuderna–Danish apparatus (Figure 2). The degree of concentration is calculated from the ratio of the initial volume of the sample to the volume of the concentrated sample. For example, for a 1 L water sample which is finally concentrated to 1 mL, the degree of concentration is 1000 for both micropollutants and impurities of the solvent. After a large number of extractions, storage and handling of solvent wastes can be environmentally hazardous. The Kuderna–Danish apparatus is used for the extraction of nonvolatile or less volatile components, because its operating temperature is higher than the boiling point of the solvent. Liquid–liquid extraction is losing its importance as nonsolvent extraction methods become more and more popular.[2]

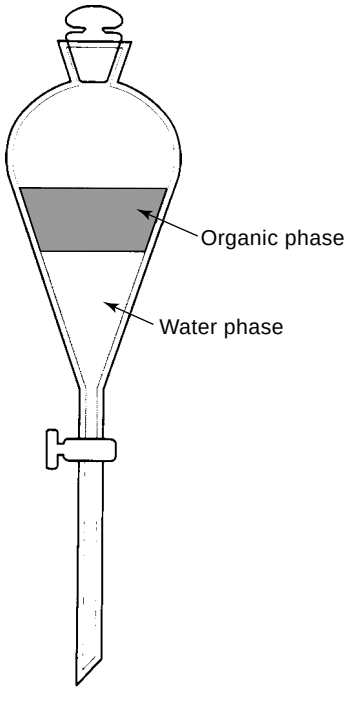

Figure 1 Separation funnel.

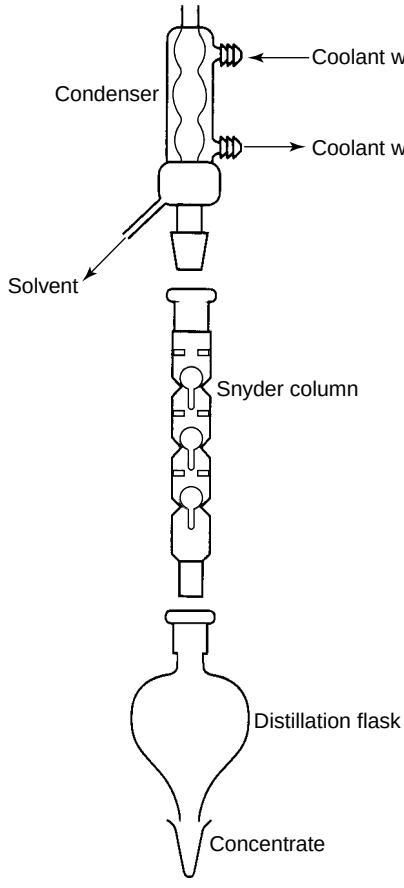

Figure 2 Kuderna–Danish apparatus.

3.2 Purge-and-trap Technique

This method is excellent for extracting volatile organic components from water. Ultrapure gas is bubbled through a 5–20 mL water sample. Volatile organic compounds (VOCs) move with the gas bubbles and the flow is driven into a heatable adsorber column, where components adsorb at room temperature. Then, by heating the adsorber column, these components are forced into a cooled capillary column, where they condense. This process is called cryofocusing. Finally, by heating the capillary column to 200–300 °C, the components pass into the gas chromatograph. Benzene homologs benzene, toluene, ethylbenzene, and xylenes (BTEX) and volatile halogenides are analyzed with this method according

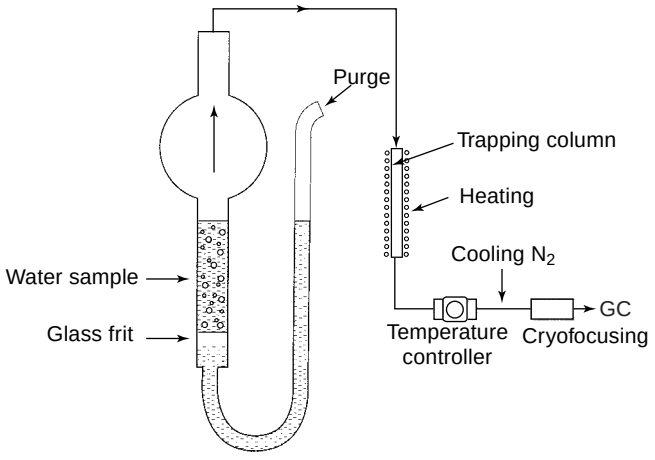

Figure 3 Scheme of purge-and-trap technique.

to the United States Environmental Protection Agency (USEPA) (Figure 3).

Since the whole sample can be analyzed, this extraction method is highly sensitive, plus it has the advantage of being a nonsolvent technique. The detection limit is nanograms per liter in the case of a few milliliter water sample. Commercial instruments are automatic.[3]

3.3 Headspace Extraction

This method is used for extracting volatile compounds. The water sample is placed into a vial, which is put into a thermostat. A sample is then extracted from the headspace at a given temperature, between 40 and 60 °C. Two techniques exist, known as static and dynamic (Figure 4). Headspace extraction is a nonsolvent and automatic method. In certain cases the purge-and-trap technique can be replaced by this method, because of its high sensitivity.

3.4 Solid-phase Extraction

With the exception of volatile compounds, the SPE method is widely used. The sample is passed through a short, packed column, where organic components adsorb or absorb (Figure 5). Then they are eluted with a small amount (2–10 mL) of solvent. Three types of packing material exist:

1. normal phase (mainly SiO_2, Al_2O_3, silicates and other inorganic adsorbents)
2. reverse phase, which contain chemically bonded liquid film on an inorganic, solid support
3. ion-exchange packing.

Depending on the chemical properties of the film, a preseparation (clean-up) phase may be possible.

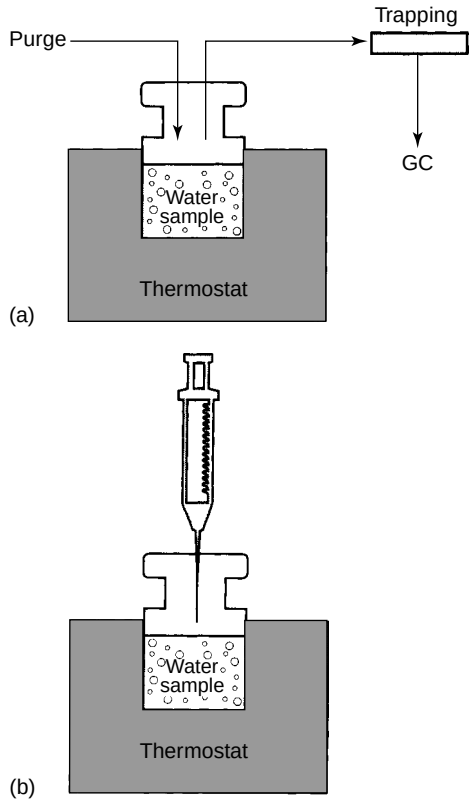

Figure 4 Scheme of headspace extraction: (a) dynamic headspace; (b) static headspace.

SPE has many advantages over more traditional sample preparation techniques such as liquid–liquid extraction. The main benefits are

- reduced need for organic solvent
- ease of automation
- concentration of analytes
- high recoveries
- highly purified extracts, etc.

Use of the SPE technique has increased dramatically in the last ten years.[4]

3.5 Solid-phase Microextraction

Solid-phase microextraction (SPME) is the most rapidly developing field of environmental analysis. SPME techniques were originally used for extracting water samples, but the method is also suitable for examining the headspace of solid samples. This method is simple, fast and quite efficient. Only 5–20 mL of sample is necessary, while no solvent is needed. The degree of concentration is 10^2–10^5, depending on the circumstances. The essential part of the instrument is a thin, fused silica fiber with chemically bonded organic film on its surface (Figure 6). When the fiber is immersed into the water sample or kept

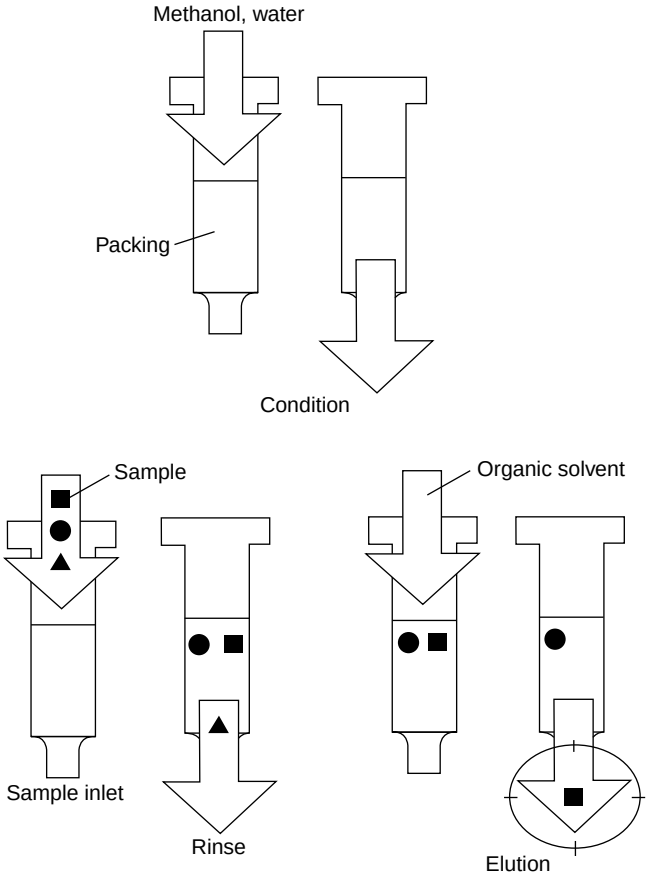

Figure 5 The steps involved in SPE. ■, Sample; ●, , matrix components.

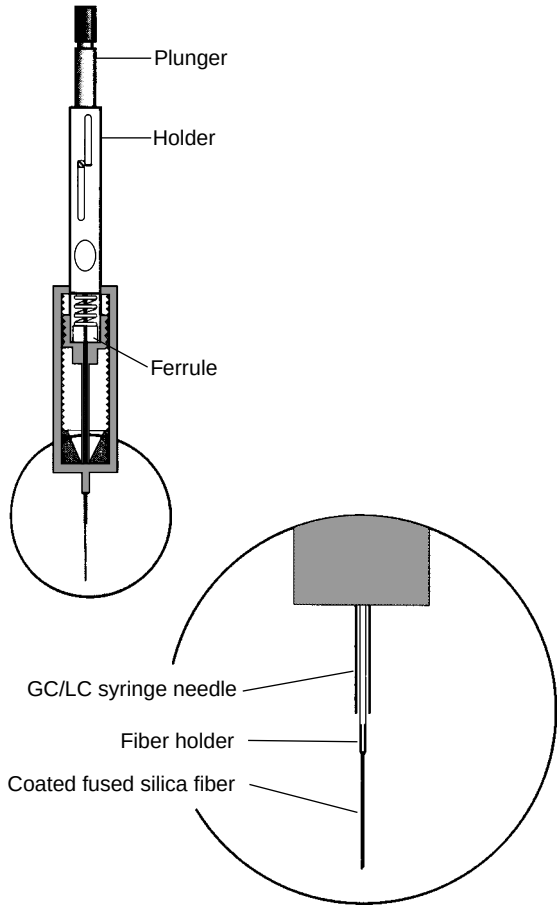

Figure 6 SPME manifold.

in the sample headspace for a given time (2–30 min) with vigorous stirring, organic micropollutants absorb in the film, depending on their distribution coefficients. Finally, using the holder, the fiber is placed into the injector of a GC instrument, where components desorb at high temperature (200–300 °C).[5]

4 SOLID SAMPLE PREPARATION

4.1 Soxhlet Extraction

This is a classical extraction method used for solid samples. Vapor from the solvent moves up from a flask, which is heated. The vapor condenses on a condenser, and drops of the solvent fall back onto solid sample particles. Once the side-vessel is full, the solvent runs back to the bottom flask (Figure 7). This method is a so-called exhaustive extraction, but is very time-consuming (8–48 h). The method can be automated, and in commercial instruments it is possible to extract more than one sample at the same time.

4.2 Supercritical Fluid Extraction

In a phase diagram of a one-component system we can see the temperature and pressure intervals, where different phases are thermodynamically stable (Figure 8). The values at which two certain phases can be in equilibrium are given by the barrier lines. Above a given temperature (T_c, critical temperature) the two phases become one. Under special circumstances, when both temperature and pressure are higher than the critical values, materials are said to be in a supercritical state. Supercritical fluids have properties of both liquids and gases. Their diffusion and viscosity constants are similar to those of gases, while their densities are closer to those of liquids (see Table 1). Supercritical fluids are excellent solvents, because it is easy to manipulate their solvation properties by changing the temperature and pressure. A scheme of the supercritical fluid extraction (SFE) process can be seen in Figure 9.

In the field of solid sample analysis, SFE is becoming more and more popular. Its efficiency is similar to that of Soxhlet extraction, while the process lasts only 20–30 min.

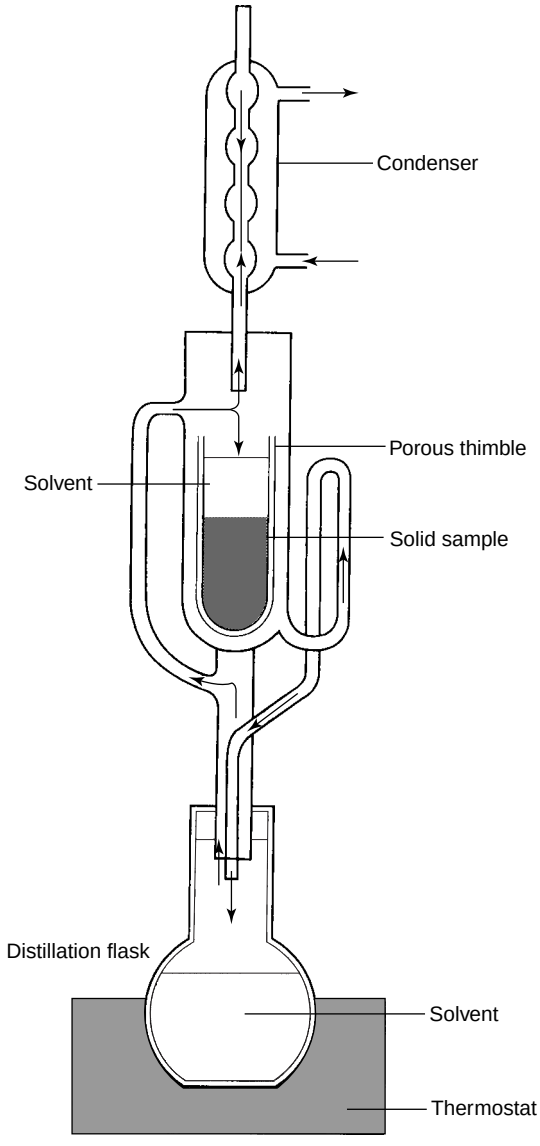

Figure 7 Soxhlet apparatus.

Supercritical CO_2 is the most common solvent. The method can be automated.[6,7]

4.3 Ultrasonic-assisted Solid–Liquid Extraction

This sample preparation method is used for extracting compounds from contaminated soil samples. The process is based on cavitation in liquid media. The bubble generation caused by cavitation is similar to boiling and it contributes to the dispersion of the solid sample. Usually the ultrasonic source is immersed into the mixture. Since the moisture content of the solid sample is rarely high, it is recommended that the ultrasonic-assisted extraction is begun with a polar organic solvent (such as acetone) and continued with a nonpolar one (such as *n*-pentane).

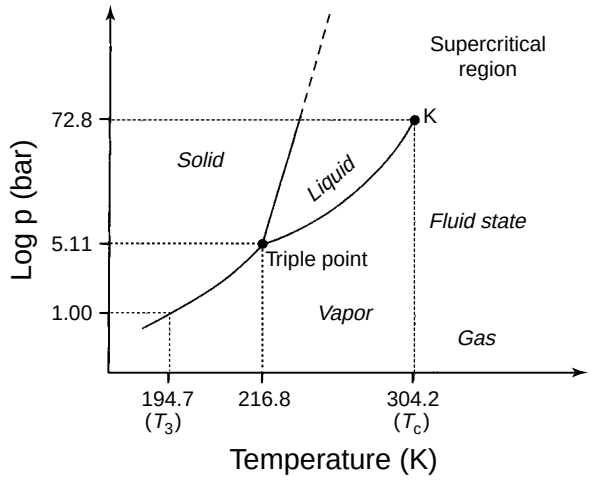

Figure 8 Phase diagram of CO_2.

Table 1 Characteristics of gaseous, liquid, and supercritical states

	Gas	Liquid	Supercritical fluid
Density (d, g cm^{-3})	10^{-3}	1	0.3
Diffusion coefficient (D, cm^2 s^{-1})	10^{-1}	5×10^{-6}	10^{-3}
Viscosity (η, g cm^{-1} s^{-1})	10^{-4}	10^{-2}	10^{-4}

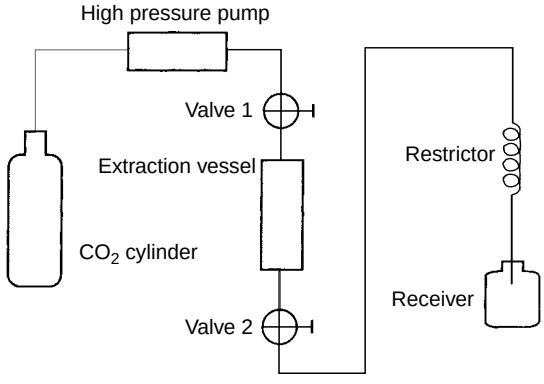

Figure 9 Scheme of SFE technique.

Fields of application are with total petroleum hydrocarbons (TPH), GC, BTEX, chlorinated VOCs, polycyclic aromatic hydrocarbons (PAH) and polychlorinated biphenyls (PCB).

4.4 Matrix Solid-phase Dispersion

Matrix solid-phase dispersion (MSPD) is a highly efficient sample preparation process for use with solid matrixes. It is a microscale extraction technique, which usually needs 0.5–1 g of sample. This method was developed to extract

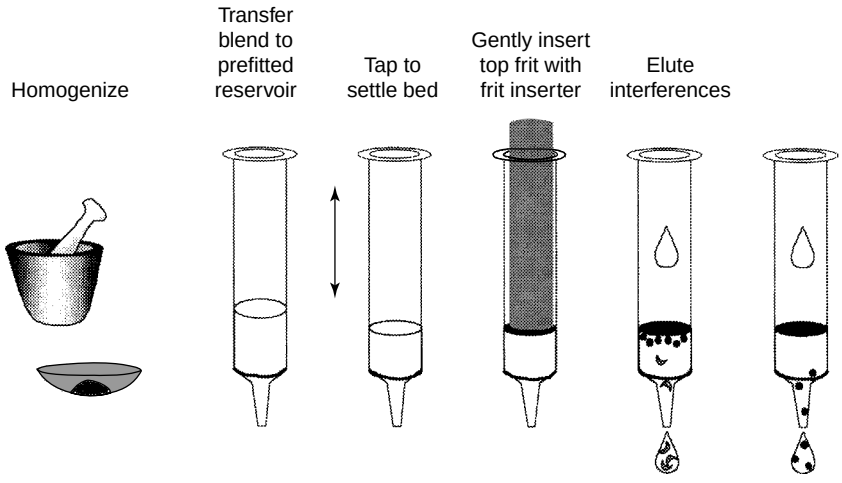

Figure 10 MSPD procedure.

additive residues from biological matrixes such as meat, liver, muscle, vegetable, etc.

Homogenization of the sample is carried out with a bulk-bonded silica-based sorbent in a mortar, and mechanical shearing is used to disrupt the structure of the tissue (Figure 10). The homogenization step is followed by tedious liquid–liquid extraction procedures. Using a suitable solvent this method is highly selective.[8]

5 DERIVATIZATION METHODS FOR GAS CHROMATOGRAPHY

The aim of derivatization methods for GC is the substitution of active functional groups to increase or decrease volatility, to improve detectability and thermal stability, and to reduce the polarity of molecules.[9,10] There are three main methods in this field:

- acylation
- alkylation
- silylation.

5.1 Acylation

The acylation method is used for compounds that contain OH, SH and NH groups, and reduces the polarity of these molecules. General acylation reagents are perfluoro acid anhydrides, such as trifluoroacetic anhydride and pentafluoropropionic anhydride, and fluoroacetamides, such as *N*-methylbis (trifluoroacetamide). The detectability is increased by electron capture detection (ECD).

5.2 Alkylation

The alkylation procedure is used to modify acidic hydrogens, such as phenols and carboxylic acids. Perfluoroalkyl reagents improve the detectability while other alkylation reagents, such as dimethylformamide, reduce the polarity of compounds.

5.3 Silylation

Silylation is substitution of an active hydrogen for a trimethylsilyl group. The silylated derivatives are more volatile and more stable. This method reduces the polarity of the compound and reduces the possibility of hydrogen bonding. Many hydroxy, carboxy and amino compounds have been successfully measured after silylation. General silylation reagents include dimethyldichlorosilane, trimethylsilyl-*N*,*N*-dimethylcarbamate, trimethylchlorosilane, trimethylsilylimidazole, etc.

ACKNOWLEDGMENTS

The author would like to thank his colleagues, Katalin Böröczky and Peter Horvatovich for useful suggestions and for helping to edit this compilation.

ABBREVIATIONS AND ACRONYMS

BTEX	Benzene, Toluene, Ethylbenzene, and Xylenes
ECD	Electron Capture Detection
GC	Gas Chromatography
MSPD	Matrix Solid-phase Dispersion
PAH	Polycyclic Aromatic Hydrocarbons
PCB	Polychlorinated Biphenyls
SFE	Supercritical Fluid Extraction
SPE	Solid-phase Extraction

SPME	Solid-phase Microextraction
TPH	Total Petroleum Hydrocarbons
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Airborne Instrumentation for Aerosol Measurements

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Supercritical Fluid Extraction of Organics in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Trihalomethanes in Water, Analysis of • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Portable Instrumentation: Introduction

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General • Sample Preparation, Headspace Techniques

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Industrial Hygiene cont'd **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation • Pesticides in Water: Sampling, Sample Preparation, Preservation

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

REFERENCES

1. J.W. Eichelberger, W.L. Buddle, EPA Method 524.2 Rev. 3, US Environmental Protection Agency, Cincinnati, OH, 1989.
2. W.G. Jennings, A. Rapp, *Sample Preparation for Gas Chromatographic Analysis*, Alfred Hüthig Verlag, Heidelberg, 1983.
3. V.L. Avila, R. Wood, M. Flanagen, R. Scott, 'Determination of Volatile Priority Pollutants in Water by Purge and Trap Capillary GC/MS', *J. Chromatogr. Sci.*, **25**, 286–291 (1987).
4. Y. Picû, J.C. Moltû, J. Manes, G. Forrl, 'Solid Phase Techniques in the Extraction of Pesticides and Related Compounds from Foods and Soils', *J. Microcolumn Sep.*, **6/4**, 331–359 (1994).
5. J. Pawliszyn, *Solid Phase Microextraction. Theory and Practice*, Wiley-VCH, New York, 1997.
6. L.T. Taylor, *Supercritical Fluid Extraction*, John Wiley and Sons, Inc., New York, 1995.
7. M.D. Luque de Castro, M. Valcarel, M.T. Tena, *Analytical Supercritical Fluid Extraction*, Springer-Verlag, Berlin, 1994.
8. M.D. Crouch, S.A. Barker, 'Analysis of Toxic Wastes in Tissues from Aquatic Species. Application of Matrix Solid Phase Dispersion', *J. Chromatogr., A*, **774**, 287–309 (1997).
9. K. Blau, G. King, *Handbook of Derivatives for Chromatography*, Heyden and Son Ltd, London, 1978.
10. D. Knausz, A. Meszticzky, L. Szakàcs, B. Csàkvàri, K. Ujszàszi, 'Silylating Agent for Alcohols, Phenols and Carboxylic Acids', *J. Organomet. Chem.*, **256**, 11 (1983).

Infrared Spectroscopy: Introduction

Marianne L. McKelvy
The Dow Chemical Company, Midland, USA

Infrared (IR) spectroscopy provides one of the most versatile and useful techniques for the qualitative and quantitative characterization of most types of materials found in the solid, liquid or gaseous state. The general utility of this technique is reflected in the fact that many sections of the Encyclopedia contain articles dealing with the application of IR spectroscopy to that particular area, for example, the characterization of polymeric materials. This section, then, focuses on basic theory of IR spectroscopy and spectral techniques that may be applied to a broad range of chemical systems.

The origins of IR spectroscopy lie in the vibrational motions of atoms within a molecule. The sharply defined frequencies of these vibrations fall within the IR region of the electromagnetic spectrum, from 14 000 to 20 cm^{-1}. The most useful range for general problem solving lies between 4000 and 400 cm^{-1}, the mid-IR range. Specific applications in the near-IR, the region above 4000 cm^{-1}, are discussed in other sections of this Encyclopedia, most notably in articles titled **Near-infrared Spectroscopy, In Vivo Tissue Analysis by; Near-infrared Spectroscopy in Food Analysis; Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels; Near-infrared Spectroscopy of Polymers and Rubbers** and **Infrared Spectroscopy in Process Analysis**.

When a sample is irradiated with IR energy, that energy will coincide with vibrational frequencies of functional groups within the molecule. When these vibrational motions include a change in the dipole moment of the molecule, IR absorptions occur. The pattern of these absorptions for a molecule, the IR spectrum, is generally unique for a particular compound. Comparison of the IR spectrum of an unknown material with that of known standards is a powerful identification tool. In the absence of standard spectra, the fact that many functional groups (C=O, C≡N, −OH, C=C, etc.) possess absorption bands that are characteristic of that moiety can be used to identify the material. The article **Interpretation of Infrared Spectra, A Practical Approach** discusses these concepts in detail.

Since the intensity of spectral absorbance is directly proportional to the amount of sample present (Beer's Law), IR spectra can also be used to make quantitative determinations of the amount of individual components in a sample mixture. These procedures can be as simple as a calibration curve that measures the intensity of a band in a sample with known path length to complex determinations of multiple components in a matrix. However involved the analysis, the existence of well-defined standards of the materials to be characterized is critical for good quantitative determination. The Encyclopedia's section on *Chemometrics* discusses these quantitative approaches in detail.

Modern instrumentation for acquisition of spectral data in the mid-IR largely consists of Fourier transform infrared (FTIR) instruments. The improved signal-to-noise and enhanced data acquisition speed of these devices has resulted in the replacement of wavelength-dispersive instruments with FTIR instruments. Additional technological improvements over the last decade have significantly lowered the cost of these instruments, making them highly affordable.

This improvement in signal-to-noise has allowed for the introduction of sampling accessories and coupled devices that further expand the analytical capability of the IR technique. Reflectance accessories have improved the ability to analyze surfaces at various depths of penetration. The IR microscope allows the routine collection of IR spectra of samples as small as ten micrometers. Reflectance accessories for the microscope extend this capability to surface analysis as well. Coupling of gas and liquid chromatographic systems to IR spectrophotometers has extended the range of sample types available for analysis. In addition to the articles in this section, applications of gas chromatography/infrared (GC/IR) and liquid chromatography/infrared (LC/IR) to the polymer and food industries are discussed elsewhere in the Encyclopedia (**Liquid Chromatography in Food Analysis; Gas Chromatography in Analysis of Polymers and Rubbers; Inverse Gas Chromatography in Analysis of Polymers; Coupled Liquid Chromatographic Techniques in Molecular Characterization**).

Spectral Data, Modern Classification Methods for discusses the difficulty of classifying spectral data, where there are a large number of variables (i.e. wavelengths) represented in each spectral case. Two strategies for dealing with this problem are presented: the use of high dimensional classifiers and reduction of the number of variables using feature extraction methods.

The article **Cavity Ringdown Laser Absorption Spectroscopy**, although not a traditional IR technique, is an introduction to this developing variant on absorption spectroscopy. This new technique promises to bring improved sensitivity to the measurement of chemical species in a broad range of environments, from molecular beams to plasmas.

Much of this section deals with the application of particular instrumental techniques to broad systems of chemical materials. In **Emission Spectroscopy, Infrared**, the theory of IR emission spectroscopy, spectrometer design and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

experimental considerations for measurements of solids, liquids and gases are discussed.

Gas Chromatography/Infrared Spectroscopy presents light pipe, matrix isolation and direct deposition interfaces for GC/IR. Applications to several industrial problems are discussed. The importance of spectral searching and databases is considered. Finally, the future of GC/IR and double-hyphenated techniques is presented.

Infrared Reflection–Absorption Spectroscopy provides an overview of this technique and the analysis of material surfaces. Theory, experimental and instrumental considerations are discussed. Finally, this technique is compared to other surface techniques that provide vibrational spectroscopic information.

Interpretation of Infrared Spectra, A Practical Approach begins with detailed discussion of the origins of IR spectra and then examines the various functional group frequencies in detail. The third portion of the article is devoted to a discussion of the importance of sample context – history and chemistry – in the analysis. The paper concludes with some simple guidelines and thought processes to assist with spectral interpretation.

Theory of Infrared Spectroscopy presents a discussion of the theoretical basis of IR spectroscopy, group theory, and discussion of solvent effects and the impact of hydrogen bonding, crystallinity and chain length of polymers on IR spectra.

Liquid Chromatography/Infrared Spectroscopy discusses techniques for both flow cell and solvent elimination approaches to LC/IR. Applications of both types of systems, including gel permeation chromatography, on-line measurements and flow injection analysis, are reviewed. Discussion of future directions for this technique is included.

The theory and design of the IR microscope are presented in **Microspectroscopy**. A discussion of sampling techniques useful in the preparation of specimens for the microscope is included. Reflection measurements are discussed in great detail. IR imaging, including a discussion of current array detector technology as well as the outlook for future development, is included.

The article **Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials** pays particular attention to sampling techniques for these materials. Spectral assignments for metal oxides and salts, hydroxides and silicates are discussed in detail. Numerous references to historic/archaeological and planetary applications of IR spectroscopy are also cited in this paper.

Spectral Databases, Infrared presents the history and application of spectral databases to the interpretation of IR spectra. Many references to published spectral collections and digitized spectral libraries are included. Various algorithms for spectral identification are discussed in detail.

ABBREVIATIONS AND ACRONYMS

FTIR	Fourier Transform Infrared
GC/IR	Gas Chromatography/Infrared
IR	Infrared
LC/IR	Liquid Chromatography/Infrared

Cavity Ringdown Laser Absorption Spectroscopy

Chuji Wang
Mississippi State University, Starkville, MS, USA

George P. Miller
University of Tulsa, Tulsa, OK USA

Christopher B. Winstead
University of Southern Mississippi, Hattiesburg, MS USA

1 Introduction **1**
2 Theory and Operating Principles **2**
2.1 Cavity Ringdown Spectroscopy 2
2.2 Sensitivity 3
2.3 Cavity Design and Cavity Ringdown Spectroscopy Implementation 4
3 Instrumentation **5**
3.1 Ringdown Cavities 5
3.2 Laser Sources 5
3.3 Electronics 6
3.4 Data Analysis 7
4 Molecular Applications of Cavity Ringdown **7**
4.1 Introduction 7
4.2 Applications 8
5 Atomic Applications of Cavity Ringdown Spectroscopy **12**
5.1 Introduction 12
5.2 Principle of Plasma Cavity Ringdown Spectroscopy 12
5.3 Plasma Cavity Ringdown Spectroscopy for Elemental Measurements 12
5.4 Plasma Cavity Ringdown Spectroscopy for Isotopic Measurements 13
5.5 Electrothermal Atomization Cavity Ringdown Spectroscopy 15
6 Medical Application of CRDS–Breath Analysis **16**
6.1 Introduction 16
6.2 Breath Analysis 16
7 Fiber Loop Ringdown Techniques **17**
7.1 Introduction 17
7.2 Fiber Loop Ringdown Chemical Sensors 18
7.3 Fiber Loop Ringdown Physical Sensors 18
8 Ringdown Instruments **20**
9 Perspective and Future Developments **20**
Acknowledgments **20**
Abbreviations and Acronyms **21**
References **21**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

Cavity ringdown spectroscopy (CRDS) has developed rapidly during the last two decades and has been implemented for a variety of applications ranging from fundamental spectroscopic studies, trace chemical detection for environmental monitoring, combustion flame chemistry and plasma diagnostics, elemental and isotopic measurements, breath gas analysis, and fiber ringdown chemical and physical sensor development. A host of chemical species, including atoms, ions, molecules, clusters, and free radicals, have been studied by CRDS. The experimental ringdown scheme has also evolved from an original mirror-based format to various new forms for detection and analysis of analytes in either gas phase or liquid solutions. Recent research effort has also extended the CRDS approach for fiber optical sensor development. CRDS has become a mature spectrometric technique for a range of prototype and commercial instrumentation. This article provides a review of the latest developments in CRDS applications, specifically emphasizing new trends in elemental and isotopic analysis, plasma diagnostics, breath gas analysis, and new fiber ringdown techniques for sensor development. The introduction of cutting-edge laser sources into CRDS methods, the combination of CRDS with conventional analytical tools, and current CRDS instrumentation are also briefly discussed.

1 INTRODUCTION

In the original implementation of CRDS, a laser pulse was trapped in a high-finesse optical cavity formed by two high-reflectivity mirrors. The many round trips of the pulse within the cavity enhanced the absorption path length for a sample gas in the cavity by several orders of magnitude. The decay time for light in the cavity was used to determine optical absorption rather than the change in intensity of the light. Following its introduction by O'Keefe and Deacon in 1988,[1] CRDS has been implemented for the spectroscopic study of numerous species, including molecules, free radicals, clusters, atoms, and ions in a variety of

environments, such as chemical reaction cells, supersonic beams, high-temperature plasmas/flames, and material processing chambers. Research and applications relevant to CRDS have expanded the original experimental scheme to the use of novel optical cavities,[2–6] light sources,[7–10] methods of cavity excitation and shutoff,[11–17] and combinations with other traditional instruments.[18–27] Although ringdown-based commercial instrumentation has been realized, thus far, only in semiconductor manufacturing,[28] this technology has already shown much promise in a wide variety of application areas, such as detection of trace species in environmental monitoring,[29–34] plasma and flame chemistry diagnostics,[35–45] high-sensitivity elemental and isotopic measurements,[46–53] breath gas analysis for medical diagnostics,[54–56] and fiber ringdown techniques for sensor development.[57–65] CRDS has been applied in spectral regions ranging from the ultraviolet to the infrared by using different laser sources. CRDS coupled with a frequency comb technique demonstrated the capability of measuring numerous chemical species simultaneously.[66,67]

More than 10 review papers on CRDS have been published to date. The first historical overview of CRDS was published in a book edited by Busch and Busch.[68] This book contains many of the early publications on various aspects of CRDS. Berden et al. have given a comprehensive review on experimental schemes and applications of CRDS.[69] Scherer et al. reviewed early publications in fundamental spectroscopic studies using CRDS.[70] Miller and Winstead have reviewed atomic and analytical applications of CRDS.[71] Wagner et al. have reviewed early CRDS studies in the infrared spectral region.[72] The application of CRDS for kinetics studies has been reviewed by Atkinson.[73] The technological innovations of cavity ringdown have been discussed in Vallance's review.[74] Paldus et al. have given a historical review of CRDS development.[75] Recently, Mazurenka et al. have reviewed the powerful application of diode lasers in CRDS and its variant, cavity enhanced absorption spectroscopy (CEAS), also widely known as an integrated cavity output spectroscopy (ICOS).[76] Very recently, Wang has reviewed plasma cavity ringdown spectroscopy (P-CRDS) for elemental and isotopic measurements (Figure 1).[77]

The purpose of this article is to provide an update of the previous review[71] on the application of CRDS in analytical and atomic spectroscopy by Miller and Winstead in 2000. The discussion attempts to cover a wide range of applications of CRDS with a new emphasis in this article on environmental monitoring, the latest developments in P-CRDS for elemental and isotopic measurement, and combustion flame chemistry and plasma diagnostics. Emerging medical applications of CRDS and new fiber ringdown techniques for optical sensor development are new additions to the article. Finally, a brief update of the current status of ringdown commercialization/instrumentation is also presented.

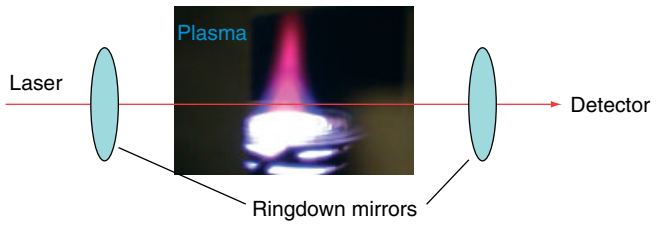

Figure 1 Concept of a plasma-CRDS system. (From ref. 77. Reproduced by permission of The Royal Society of Chemistry.)

2 THEORY AND OPERATING PRINCIPLES

2.1 Cavity Ringdown Spectroscopy

The number of experimental methods that are essentially based on the CRDS principle continues to grow. However, in its original form, CRDS is implemented by injecting and trapping a laser pulse in a stable optical cavity, which is formed by two highly reflective mirrors. The intensity of the light in the cavity decays exponentially with time at a rate determined by the round-trip losses experienced by the laser pulse. This intensity decay is monitored using a photomultiplier tube or a photodiode placed behind the second mirror, and the ringdown time is usually determined by fitting the observed waveform to a single exponential function. A schematic diagram of a typical experimental system is given in Figure 2. For the simplest case, when the dominant losses are the mirror reflectivity and absorption from a sample gas in the cavity, the time constant for the exponential decay is given approximately by

$$\tau = \frac{d}{c[(1-R)+\alpha l_s]} \tag{1}$$

where d is the cavity length, R is the reflectivity of the cavity mirrors, α is the familiar Beer's law absorption coefficient of a sample in the cavity, l_s is the length of the optical path through the sample, and c is the speed of light. For a gas filling the entire cavity, $d = l_s$. As the laser wavelength is varied, the ringdown time will decrease upon tuning to an absorption wavelength of the sample (e.g. when α increases). Alternatively, when the laser wavelength is fixed on resonance with an absorption wavelength of the sample, variations in α caused by changes in the analyte concentration are reflected in the ringdown time.

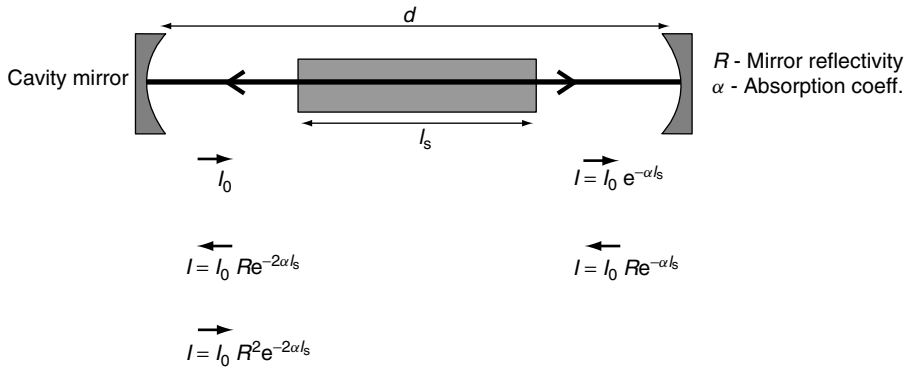

Figure 2 Basic optical configuration for cavity ringdown spectroscopy.

A "back-of-the-envelope" understanding of the time behavior from Equation (1) can be gained from an examination of Figure 2, where the first few passes of a laser pulse through the cavity are depicted. When a laser pulse is incident on the first cavity mirror, most of the energy is reflected away from the cavity. Assuming that the fraction of the original pulse that enters the cavity has intensity I_0, after one pass through the sample gas, the intensity will be reduced according to Beer's law to $I = I_0 e^{-\alpha l_s}$. Following reflection from the second cavity mirror, the intensity will be reduced by the finite mirror reflectivity to $I = I_0 R e^{-\alpha l_s}$. After n round trips through the cavity and substituting $R^{2n} = e^{2n \ln R}$, the intensity will be

$$I = I_0 R^{2n} e^{-2n\alpha l_s} = I_0 e^{-2n(-\ln R + \alpha l_s)} \tag{2}$$

If the cavity is long enough that the pulse never overlaps itself (e.g. longer than one-half of the physical length of the laser pulse), a series of separated pulses decaying in time will be detected according to Equation (2).[70] In many cases, due to electronic response time or the use of shorter cavities, separated pulses are not detected, but rather a continuous exponential decay. The peak envelope of the separated pulses or the time constant for a continuous exponential decay can be found by converting from the discrete variable n to a continuous variable t such that $n = t/T$, where $T = 2d/c$ is the pulse round-trip time in the cavity. In this case, we have

$$I = I_0 e^{-tc(-\ln R + \alpha l_s)/d} \tag{3}$$

Of course, the absolute intensities detected by the Photomultiplier tube (PMT) are reduced by an additional factor of the mirror transmission, but this does not affect the time behavior of the exponential decay. From Equation (3), we see that the characteristic time for the laser pulse decay is given by Equation (4):

$$\tau = \frac{d}{c(-\ln R + \alpha l_s)} \approx \frac{d}{c[(1 - R) + \alpha l_s]} \tag{4}$$

as given in Equation (1). While this analysis does not provide a complete development of the technique, it captures the essential elements that yield the excellent sensitivity offered by CRDS. The performance and some limitations of CRDS and this simple model are discussed. Note that if the ringdown time is measured in the absence of a sample gas in the cavity (e.g. $\alpha = 0$), the reflectivity of the mirrors can be determined. Once the mirror reflectivity has been measured and a sample introduced, CRDS provides an absolute measure of the absorbance αl_s, from which the absolute density n of the absorber is determined by

$$\alpha l_s = \sigma n l_s \tag{5}$$

where σ is the absorption cross section of the absorber in the laser wavelength. This self-calibrating feature can serve as a great benefit in comparison to other high-sensitivity techniques, such as laser-induced fluorescence (LIF) or resonantly enhanced multiphoton ionization (REMPI).

2.2 Sensitivity

The high sensitivity of the CRDS technique stems significantly from the large number of passes the light pulse makes through the sample in the cavity. For mirrors of 99.99% reflectivity in a 1-m-long cavity, the ringdown time is approximately 33.4 μs for an empty cavity. This is equivalent to a 10-km path length traveled during the first time constant. For analytical applications, the detection sensitivity of CRDS is normally characterized by the detection limit, which is determined by a minimum detectable absorbance,[78]

$$\text{Absorbance} = \sigma n l_s = (1 - R)\frac{\Delta\tau}{\tau} \tag{6}$$

where $\Delta\tau$ is the difference between τ_0 (the decay time with no analyte) and τ (the decay time with analyte present). Equation (6) indicates that for a given mirror

reflectivity, the measured absorbance depends upon the quantity, $\frac{\Delta\tau}{\tau}$. When τ is replaced by τ_0 and $\Delta\tau$ replaced by σ_τ (one standard deviation of the measured ringdown time), then the absorbance in Equation (6) represents the minimum detectable absorbance based on a $1\sigma_\tau$ criterion. Note that in most cases the detection limit is derived on the basis of the standard $3\sigma_\tau$ criteria, namely,

$$A_{\min} = (1 - R)\frac{3\sigma_\tau}{\tau_0} \tag{7}$$

Clearly, Equation (7) demonstrates that increased mirror reflectivity and a more stable baseline (no analyte) ringdown time will improve $A_{\min}$. Here, σ_τ and τ_0 may be the values determined after averaging a number of individual ringdown measurements, further improving $A_{\min}$. Alternatively, sensitivity can be also characterized by a detection limit that explicitly includes the assumption that the single decay measurement uncertainty is improved by averaging over N ringdown events,[79] as given in Equation (8):

$$A_{\min} = (1 - R)\frac{\Delta\tau}{\tau\sqrt{N}} \tag{8}$$

If the measuring time is restricted to 1 s, then N is replaced by data collection frequency, f. Thus, the unit of $A_{\min}$ is $\mathrm{Hz}^{-1/2}$. This is also called *noise-equivalent minimum detectable absorbance*.

Examining Equations (1) and (5), one can see that the detection sensitivity is determined by a ringdown system's physical parameters, such as cavity length, sample path length, mirror reflectivity, and by the system electronic noise. Higher mirror reflectivity, longer sample path length, and lower electronic noise will improve detection sensitivity. The combined effect of the electronic noise with the optical alignment of the ringdown system is often experimentally characterized by $\sigma_\tau/\overline{\tau}$, which is typically referred to as the *ringdown baseline noise*. For a pulsed-CRDS system, the baseline noise is typically larger than 0.3%. However, for a continuous wave cavity ringdown spectroscopy (CW-CRDS) system, the baseline noise can be much lower. The lowest baseline noise reported thus far is 2.7×10^{-4} obtained in a near infrared (NIR) CW-CRDS system with an average over 300 ringdown events.[54] Detection sensitivity on the order of 1×10^{-8} (absorbance) demonstrated in the early studies[80,81] can now be readily achieved with the availability of ultrahigh reflectivity mirrors.

2.3 Cavity Design and Cavity Ringdown Spectroscopy Implementation

The simple model outlined in Section 2.1 neglects a number of other potentially important considerations. First, to achieve true single exponential behavior, the conditions for Beer's law behavior must be fulfilled.[78] These conditions include requiring that the frequency linewidth of the laser source be narrower than the absorption linewidth. If the laser source linewidth is broader than the absorption, then only that part of the laser frequency resonant with the absorption will be attenuated. The nonresonant frequencies of the laser pulse will continue to propagate in the cavity without being absorbed, leading to distinctly nonsingle exponential behavior. In the case where the absorption and laser linewidths are nearly the same, CRDS can still be used to extract quantitative information so long as the absorption constitutes a small portion of the total cavity loss (e.g. the absorption is small compared to losses at the mirrors).

The more detailed response of CRDS to the effects of cavity design and implementation was the subject of much discussion in the early CRDS literature.[49–51,78,82,83] A ringdown cavity is a high-finesse etalon, and, as such, possesses a discrete mode structure dependent upon the design of the cavity. For a cavity constructed from two spherical surface mirrors of radius R_1 and R_2 separated by a distance L, the frequencies of the stable cavity modes are given by the equation[52,84,85]:

$$\upsilon = \frac{c}{2L}\left\{n + (1 + m + l)\frac{\cos^{-1}\sqrt{(g_1 g_2)}}{\pi}\right\} \tag{9}$$

where $g_1 = 1 - \frac{L}{R_1}$ and $g_2 = 1 - \frac{L}{R_2}$. The parameters g_1 and g_2 are dimensionless quantities that can be used to determine whether a given cavity design is stable. For a stable cavity, the condition $0 < g_1 g_2 < 1$ must hold. In Equation (9), n, l, and m are integer quantities used to index the modes of the cavity. The parameter n represents the longitudinal mode number and is approximately of the order of $2L/\lambda$, l and m represent the transverse mode numbers. For the lowest order mode, commonly referred to as TEM_{00}, l and m are zero. Here, we see that for a cavity constructed from plane parallel mirrors where R_1 and R_2 are infinite, we recover the well-known frequency transmission of the Fabry–Perot type of etalon

$$\upsilon = \frac{nc}{2L} \tag{10}$$

The frequency structure of the cavity modes can be visualized as consisting of TEM_{00} longitudinal modes whose frequency separation is given by Equation (10) as $c/2L$. The frequency space between these TEM_{00} modes will be filled to a greater or lesser degree by higher order transverse TEM_{lm} modes, depending upon the exact design of the cavity. When a cavity is subjected to excitation by a very narrow-linewidth continuous wave (CW) laser, not all laser frequencies are transmitted

through the cavity.[86] Only those frequencies that correspond to stable modes of the cavity are transmitted. The behavior of such a cavity under pulsed excitation was also the subject of much discussion.[49–51,78,82,83] Arguments have been presented that the short coherence length of pulsed lasers, in general, precludes mode effects.[82] Other concerns have been expressed that for the absorption features narrower than the mode spacing of the cavity, it is possible that absorption features could be missed (since only light at the mode frequencies rings down in the cavity).[87] In practice, however, for pulsed laser excitation, such effects have not been problematic. Cavities can be easily designed to possess a near continuum of transverse modes.[88] Typical pulsed laser linewidths are also wider than even the normal spacing between the TEM_{00} longitudinal modes, so that light from the pulsed laser is injected into the cavity regardless of laser frequency or transverse mode structure. Finally, it has been noted that while mode effects may modulate the total intensity of the signal exiting the cavity, the signal usually never drops to zero.[89] Since the ringdown time and not the intensity is monitored in CRDS, a small modulation does not greatly hinder the ringdown measurement.

3 INSTRUMENTATION

3.1 Ringdown Cavities

A ringdown cavity in its original form consists of two high-reflectivity mirrors with a mirror spacing that meets the stable cavity conditions discussed in the preceding section. Nowadays, a ringdown cavity can have various forms, such as a cavity consisting of more than two high-reflectivity mirrors,[90] a pair of superpolished prisms,[3] a single piece of prism,[2] a section of single-mode optical fiber with two identical fiber Bragg gratings inscribed on it,[4] a fiber loop,[91] or a cavity with a plasma[18] or a flame in it.[86] Although there are many forms of ringdown cavities, the fundamental characteristics remain the same: laser light is trapped in a high-finesse cavity to experience many round trips to enhance the measurement sensitivity. Owing to different geometries and/or different cavity configurations, experimentally observed ringdown times can range from nanoseconds to hundreds of microseconds. Berden demonstrated an extremely short ringdown cavity (3 cm) in 1998.[92] Le Grand and Le Floch reported a very long cavity (270 cm) in 1990.[93] A cavity used to demonstrate a compact acetone detection device was only 13 cm in length (Figure 3), which yielded a ringdown time of 700 ns in the deep ultraviolet (UV) region.[55] For liquid detection, cavity lengths can even be on the order of μm. Conversely, using ultrahigh reflectivity mirrors in the infrared (IR) region, e.g. 99.9985%, a 43-cm cavity had a ringdown time of up to 100 μs,[28] and a ringdown time of several hundreds of microseconds has been reported in a fiber loop ringdown system.[94]

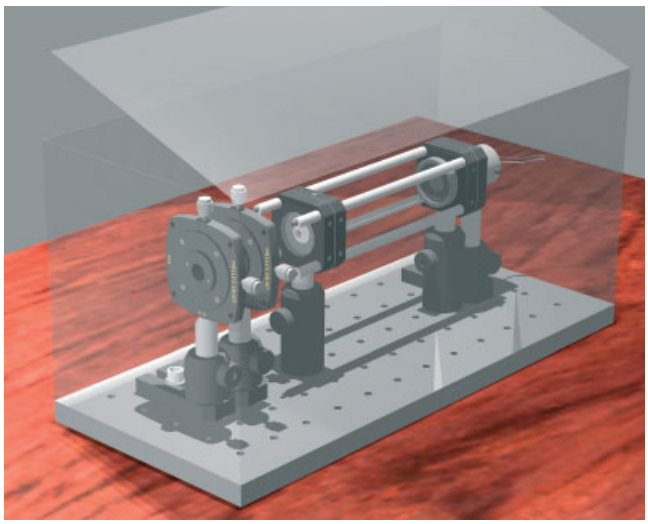

Figure 3 A compact ringdown cavity for CRDS prototyping. (*Appl. Spectrosc.*, 2004, **58**, 784–791 by Wang et al., Figure 2 on page 786.)

Often a mirror-based ringdown cavity can only be used for measurement of a very limited number of species because of the relatively narrow bandwidth of ultrahigh reflectivity mirrors. However, a broadband prism cavity, first introduced by Lehmann and Rabinowitz in 1999, can have a spectral response ranging from the UV to IR.[3] This type of prism cavity can potentially measure spectral fingerprints of various species in different spectral regions. A new type of fiber loop (cavity) for detection of a small volume of liquid is discussed in Section 7.

3.2 Laser Sources

In principle, there is no restriction on the laser sources to be used for a ringdown system. Both pulsed lasers and CW lasers with wavelengths ranging from the ultraviolet to the IR can be used, given availability of ringdown mirrors coated for the appropriate wavelength region. Commonly employed pulsed lasers include pulsed dye lasers pumped by Nd : YAG or excimer lasers, or widely tunable optical parametric oscillator (OPO) lasers. CW ring dye lasers pumped by Ar^+ lasers and quantum cascade lasers (QCLs) have been also widely used for ringdown studies, as have mid-IR laser sources based on difference frequency generation (DFG) or Raman shifting techniques.[95] Inexpensive telecommunications diode lasers have been increasingly used in ringdown research and instrumentation because of low cost, small footprint, and high resolution. Broadband laser

sources have also been explored for ringdown studies. Engeln and Meijer injected a broadband dye laser pulse into a ringdown cavity and used a Fourier transform spectrometer to separate the frequency components of the pulse exiting the cavity.[96] Scherer also utilized a broadband laser beam for the development of a ringdown spectrometer.[97] Fiedler et al. measured the C–H overtone of liquid benzene using incoherent broadband CEAS.[98] Perhaps the most nonstandard laser system used for ringdown was reported by Engeln et al., who implemented CRDS in the IR using a free electron laser system to obtain a spectrum of a thin film of C_{60}.[99,100] Very recently, a broad-bandwidth optical frequency comb has been coherently coupled to a ringdown cavity for real-time, quantitative measurements of trace concentrations, transition strengths and linewidths, and population redistributions for molecules such as C_2H_2, O_2, H_2O, and NH_3.[66] The same group has demonstrated an improved frequency comb-based cavity ringdown spectroscopy system based on a mode-locked erbium-doped fiber laser source centered at 1.5 μm, resulting in an absorption sensitivity of $2 \times 10^{-8}\,cm^{-1}\,Hz^{-1/2}$ per detection channel for CO, NH_3, and C_2H_2 in a very large spectral bandwidth (1.45–1.65 μm).[67]

A critical issue when employing an absorption spectroscopy method such as CRDS is the laser linewidth. If the laser used in CRDS has a large linewidth, two measurement problems may arise. First, hyperfine structures of an optical transition or an isotopic shift cannot be captured owing to the lack of laser resolution. This is particularly true for atomic transitions and isotopic shift. Second, if the laser linewidth is too broad, accurate measurements for high-resolution spectroscopy and atomic absorption cannot be achieved (mostly, the overall absorption signal can be undermeasured). The effects of the laser linewidth on the accuracy of ringdown measurements have been experimentally observed and theoretically analyzed.[49–51,78,82,83,101–103] Even some pulsed lasers are suitable for moderate to high-resolution spectroscopy. For instance, a pulsed dye laser with double-grating configuration can have a linewidth of $0.08\,cm^{-1}$ at 590 nm (e.g. NarrowScan) and $0.03\,cm^{-1}$ at 570 nm (e.g. Sirah). A widely tunable OPO system (e.g. the MOPO-HF), can have a linewidth of $0.075\,cm^{-1}$. In contrast, CW lasers such as external cavity diode lasers (ECDL) or distributed feedback (DFB) diode lasers typically have a laser linewidth of kilohertz to megahertz levels ($1\,cm^{-1} = 30\,GHz$). Thus, as with other very high resolution absorption spectroscopy methods, CW laser CRDS methods are preferred for very high resolution studies.

Laser beam propagation behavior in a ringdown cavity depends on the cavity geometry and the nature of the medium in the cavity. For example, different gases filling the cavity have different refractive indices, thus generating different scattering losses. Refractive index variations or the existence of a large refraction gradient can increase ringdown baseline noise, create beam steering effects, or even prevent establishment of a ringdown event. For example, beam-steering effects have been observed in a ringdown cavity containing atmospheric-pressure plasma.[53,104] Severe beam deflections due to gas turbulence have been observed in plasma confined in tubing, preventing the establishment of a ringdown event.[48] A careful design of the ringdown cavity geometry or an optimized plasma gas flow pattern can minimize beam-steering effects.

3.3 Electronics

For pulsed-CRDS, the bandwidth of the laser is typically several times larger than the longitudinal mode spacing of the ringdown cavity. Therefore, the light on each laser pulse is coupled into the cavity and no additional electronic parts are required to achieve cavity excitation. When a narrow-linewidth diode laser is used as the light source, an additional electronic control is needed. The linewidth of a diode laser, e.g. an external cavity diode laser or a DFB diode laser, is much narrower than the cavity mode spacing, implying that the laser must be tuned into resonance with a cavity mode in order to excite the cavity. Engeln et al. first reported the use of an intensity-modulated, narrow-linewidth CW ring dye laser for CRDS.[105] This work utilized an unstabilized ringdown cavity and depended upon random fluctuations (such as variations in the cavity length due to vibration) to couple the laser into the cavity modes. To increase the coupling efficiency, a cavity with a dense mode structure was utilized. The phase shift of the intensity-modulated beam exiting the cavity was used to obtain a high-sensitivity spectrum. Later on, Romanini et al. described a method of actively coupling narrow-linewidth CW lasers (such as ring dye and diode lasers).[11,12] In these experiments, the length of the cavity was dithered by mounting one cavity mirror on a piezoelectric translator. By modulating the length of the cavity, the mode structure of the cavity was swept through the laser frequency. When the laser and cavity mode frequencies overlapped, light was efficiently injected into the cavity mode, rapidly increasing the intensity of the light in the cavity (i.e. the cavity was excited). The schematic of the method is shown in Figure 4. After these initial reports, several early studies using CW-CRDS were reported.[106–112]

CW-CRDS has matured into a standard ringdown technique. The methods of achieving the cavity excitation can be simply classified into three dominant categories[68–77]: (i) cavity length modulation to vary the cavity mode spacing, (ii) current modulation to vary the laser frequency, and (iii) a combination of the two

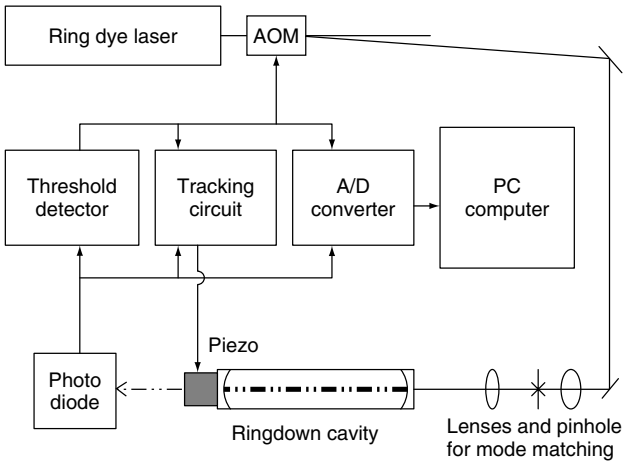

Figure 4 Simplified scheme of a CW CRDS setup. AOM, acousto-optic modulator. (Reprinted from ref. 81, with permission from Elsevier Science.)

methods. The first and the third methods have been almost abandoned owing to the complexity and the high cost of the required instrumentation. Currently, most instrumentation and research using CW-CRDS implements the current modulation scheme. In CW-CRDS, once the cavity is excited, shutoff of the excitation is needed to observe the ringdown decay behavior. Typically, an acousto-optical modulator (AOM) is used to deflect the laser beam away from the optical axis of the ringdown cavity to shut off the cavity buildup, thus allowing the ringdown waveform to be subsequently recorded.[68–77] This method requires an expensive AOM device and a trigger circuit to control the procedure. Another major disadvantage of this method includes a loss of laser power through the AOM. When a laser beam passes through the crystal of an AOM, losses in the observed laser power are typically 15–20%. This imposes a distinct disadvantage when a low-power DFB diode laser is used as the light source in a CW-CRDS system, since lower signal intensity means lower signal-to-noise ratio. Although many CW-CRDS-based research efforts are still using an AOM, CRDS-based commercial instruments have avoided implementing this device.[15]

An alternative approach to shut off the cavity excitation without using an AOM in CW-CRDS is controlling the laser current. Once the cavity is excited, the electrical signal from the detector is used to trigger a circuit, which drives the laser diode current below the threshold of the driving current. This procedure typically requires a specially designed electric circuit. This circuitry is a trivial engineering effort, but it is not readily available for research laboratories. Recently, a simple and effective method has been introduced to achieve cavity excitation and shutoff, using a standard off-the-shelf power combiner.[17,46] With this approach, the power combiner couples the output of a function generator with the output of a pulse generator to create a special voltage waveform to manipulate the diode laser current. This eliminates the need to use special electronic modulation circuits to achieve current modulation. This method has been successfully used in several ringdown research studies and applications.[17,46,57,58,62,64,113,114]

3.4 Data Analysis

Ringdown signals from a PMT or photodiode detector are typically digitized using a digital oscilloscope, transient digitizer, or fast A/D converter interfaced to a computer. Ringdown software is typically in-house developed, but is also commercially available. To determine the ringdown time, the resulting data is fitted as a single exponential function. In practice, averaging of multiple waveforms is often employed to smooth out shot-to-shot variations in the ringdown signals and improve the signal-to-noise ratio. Strictly speaking, such averaging is only valid if the waveforms are normalized in intensity. Alternatively, the ringdown time of each individual transient can be computed, followed by averaging a number of the computed time constants. Some early ringdown experiments used a dual-channel boxcar integrator to compare two different regions of the ringdown signal and extract a quantity proportional to the ringdown time.[115] This approach yielded results comparable to data fitting methods. However, the software method has the added advantage of providing a quick method of establishing if the ringdown waveform is a true single exponential decay. It is a simple matter to check the observed ringdown time in different regions of the transient to determine if the waveform is truly a single exponential function. In practice, e.g. if the difference in the ringdown times obtained by fitting different ranges in the decay curve is less than about 1%, fitting a single exponential decay is justified. Cavity misalignments, linewidth effects, and index of refraction variations in the cavity can all lead to nonexponential or multiple exponential waveforms.[116] Scherer et al. have considered, in some detail, the effect of various data-acquisition schemes and have concluded that while the temporal resolution of a digitizer system is important for determining the number of points sampled, the vertical resolution can be critical to the performance of a CRDS system.[70]

4 MOLECULAR APPLICATIONS OF CAVITY RINGDOWN

4.1 Introduction

At the time of this writing (2007), CRDS and its variants have yielded more than 700 publications worldwide

(based on a search using SciFinder). Of these publications, the majority are using CRDS in the area of molecular spectroscopy. Whether for fundamental molecular structure studies, or for applications in areas such as environmental monitoring, combustion studies, or plasma processing, CRDS has proven its capability to detect weak molecular absorptions or to measure trace molecular species. Molecular applications of CRDS continue to be one of the most active fields using CRDS. It also is the field that clearly demonstrates the extreme sensitivity of the cavity ringdown technique. A complete review of molecular spectroscopy using CRDS is beyond the scope of this article. However, reviews are available for further information concerning molecular spectroscopy using CRDS.[69,76,80,117–119]

4.2 Applications

4.2.1 Fundamental Molecular Spectroscopy

The measurement of a doubly forbidden transition in molecular oxygen by O'Keefe and Deacon provided the first demonstration of the ringdown technique.[1] Since then, molecular spectroscopists have taken advantage of the high sensitivity of CRDS and utilized CRDS in a wide variety of environments. Saykally and coworkers, in one of the earliest studies, demonstrated the usefulness of the technique in supersonic expansion applications by combining CRDS with a pulsed vaporization cluster source to record absorption spectra of Cu_2 and Cu_3 in a supersonic molecular beam.[120] This same group has continued the application of CRDS to cluster studies, eventually expanding their efforts to include a number of metal and mixed species clusters[121] as well as water[95] and methanol[122] clusters. The ultrasensitivity of CRDS also lends itself naturally to studies of molecular predissociation.[123–125] Figure 5 depicts a comparison of CRDS and LIF absorption spectra for HNO that clearly reveal features missing in the LIF spectrum due to predissociation of the molecular excited state.[89] Radical spectroscopy has also been the subject of ringdown studies, as demonstrated by early investigations of OH in flames[86,123,126,127] and the SH radical in an absorption cell.[125] Miller's group have observed weak electronic transitions of large alkoxy radicals in the NIR (1.2–1.3 μm) using CW-CRDS and the radicals are generated by laser photolysis of selected precursors. The absorption cross sections of these transitions are typically only on the order of $10^{-21}\,cm^{-2}$. With the high sensitivity of ringdown spectroscopy, abundant rotational and vibrational data have been obtained for methyl peroxy (CH_3O_2),[128] trifluoromethyl peroxy (CF_3O_2),[129] acetyl peroxy ($CH_3C(O)O_2$),[130] propyl peroxy ($C_3H_7O_2$),[131] phenyl peroxy ($C_6H_5O_2$),[132] and ethyl peroxy ($C_2H_5O_2$)[133] radicals, whose spectra in UV or visible (VIS) are typically overlapped so that little spectroscopic information can be obtained. An early demonstration of the extreme sensitivity of

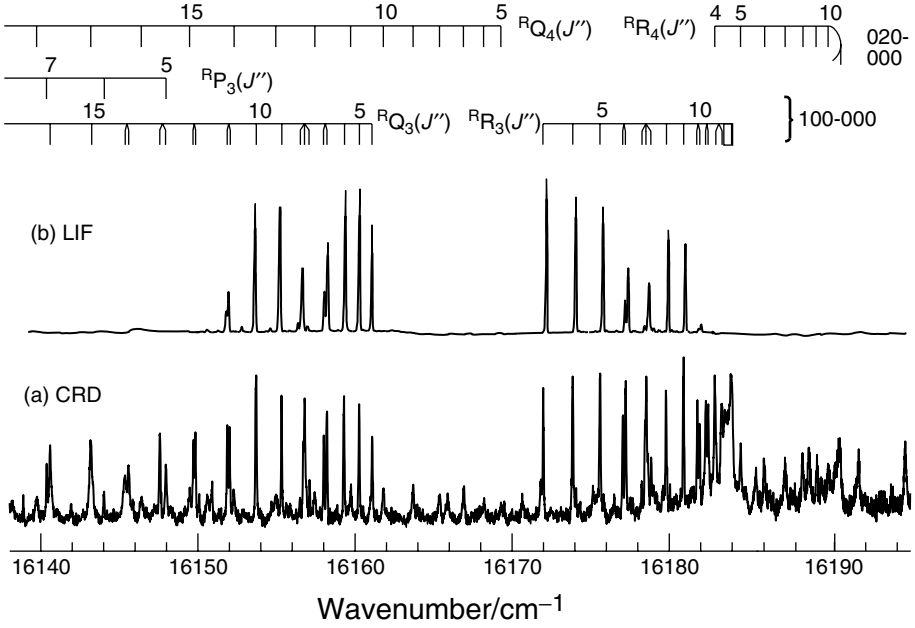

Figure 5 Cavity ringdown spectra (a) and LIF (b) of HNO A^1A''–X^1A' transition (100)-(000) $K' = 4$-$K'' = 3$ and (020)-(000) $K' = 5$-$K'' = 4$ subbands. Line assignments are shown by the combs above the spectra. Much of the structure present in the CRDS spectrum is absent in the LIF spectrum because of predissociation of the A^1A'' state. (From ref. 89. Reproduced by permission of the American Institute of Physics.)

the technique was the observation of weak overtone transitions. Romanini and Lehmann used cavity ringdown to investigate very weak overtone transitions in HCN.[80] In so doing, they demonstrated per pass absorbance detection limits on the order of 10^{-7}. In addition to molecular structure studies, CRDS has also been used in kinetics experiments.[134–137] By monitoring the temporal densities of chemical reaction products and/or reactants, reaction kinetic constants and reaction mechanisms are investigated.[134–138] It should be noted that fundamental molecular spectroscopic studies using CRDS have been extended beyond the few areas briefly mentioned above and the publications continue to increase.

4.2.2 *Combustion Flame Chemistry and Plasma Diagnostics*

A consistent challenge facing the combustion community has been to measure concentrations for various reactant, intermediate, and product molecules as a function of flame conditions and position within the flame. Direct sampling from such an environment is inaccurate, particularly for radicals, since reactions continue to proceed as the gas is sampled and transferred to an analysis system. The distribution of species in the gas sample may bear little resemblance to the original gas at the sampling position in the flame. Thus, accurate in situ methods for combustion analysis have long been desired. However, the hostile environment and intense background emission of such systems can frustrate attempts to use traditional optical analytical methods for in situ measurements. Emission spectroscopy is only capable of detecting excited state species, greatly limiting its usefulness for diagnostic purposes. Other techniques such as LIF,[117] coherent anti-Stokes Raman spectroscopy (CARS),[139] and degenerate four-wave mixing (DFWM)[117] have all yielded excellent results, but can be hampered by difficulties in calibrating for absolute measurements. A highly sensitive absorption spectroscopy technique, such as CRDS, provides an ideal complement to other analytical diagnostic methods.

The use of CRDS for flame analysis was first reported in 1994 by Meijer et al.[86] This work measured the absolute absorbance for OH radicals (near 308 nm) generated in a Bunsen burner flame. Since that time, a number of CRDS studies have investigated radical concentrations in flames including HCO,[140] CH_3,[117] CH_2,[141] and CH.[142] Also, studies have continued to investigate the behavior of the OH radical under various combustion conditions.[82,86,126,127,143] Mercier et al. carried out extensive measurements comparing CRDS, direct absorption spectroscopy, and LIF for quantitative analysis of OH radical concentrations and spatial profiles in flames.[143] Zalicki et al. demonstrated the detection of methyl radicals generated when a mixture of methane and hydrogen flowed over a heated filament.[144] Absorbances and spatial profiles were measured for methyl radical column densities far below those required for standard multipass absorption spectroscopy. Schocker et al. measured absolute concentrations of OH, HCO, and CH_2 in a propene flame near the soot-formation limit.[145] Both temperature and radical concentration profiles were discussed in that study.

Direct applications of CRDS for detection of species in a plasma have been somewhat more limited in number. However, due to the dual facts that plasma technology is very important in semiconductor industries, material processing, and analytical applications, and that CRDS has the unique capability of measuring absolute concentration of chemical species in plasmas, this number has grown rapidly during the last few years. In a very early development, O'Keefe used the technique to measure N_2^+ ions in a glow discharge.[146] Kotterer, Conceicao, and Maier, studied N_2^+ ions in a hollow cathode source.[147] Negative ion concentrations were reported by Quandt, et al., who used CRDS for quantitative measurements of H^- ions in a magnetic multipole source.[148,149] In this case, rather than traditional absorption, the optical losses associated with the electron photodetachment process were monitored. Later, Romanini et al. investigated the spectra of naphthalene cations created in a slit jet/electronic discharge source.[150] This work resulted in the first gas-phase electronic spectrum for an ionized polycyclic aromatic hydrocarbon. Wang et al.[151] measured OH density profiles in an atmospheric argon inductively coupled plasma (ICP) using a pulsed-CRDS system. They also measured the density of SrOH radicals in the plasma using an external cavity diode laser at 679 nm.[46] Recently, Oever et al. measured NH and NH_2 radicals using CRDS in an extensive characterization of the Ar-NH_3 expanding thermal plasma source, which is widely used for silicon nitride deposition in industrial processes.[152]

Of perhaps the most practical interest for plasma analysis is the use of CRDS for measurements in reactive plasmas for semiconductor growth and plasmas employed in analytical applications. Monitoring the species present during semiconductor growth can provide important information for controlling the growth process, detecting impurities, and optimizing plasma operation conditions. In related research, Campargue et al. used CRDS to determine the concentrations of SiH_2 radicals in a dc discharge plasma containing a mixture of silane and argon.[153] Booth, et al. have employed CRDS to measure radicals and etch products in a capacitively coupled radio-frequency plasma.[154] SiF_2 was measured as a product of

the etch process when a silicon wafer was placed in the system.

Interest in plasma diagnostics using CRDS continues to increase. Key plasma parameters, which include electron density and plasma temperatures, can be characterized through ringdown measurement of species in the plasma. Owing to the large gradients of plasma temperature and heterogeneity of plasmas, plasma temperature profile and electron density profile are spatially dependent. Linewidth broadening mechanisms in plasma, including Doppler broadening, Stark broadening, and collisional broadening, are complicated.[155] Which mechanism is dominant under a particular plasma condition is dependent on the plasma temperature(s), electron density, location in the plasma, etc. Detailed discussions on linewidth broadening mechanisms under plasma conditions can be found elsewhere.[20,155–157] A recent review on plasma diagnostics using laser techniques including CRDS was given by Ganguly et al.[158]

Yarlin et al. measured the spatial profiles of N_2^+ concentration in an atmospheric pressure N glow discharge by CRDS.[53] They also measured the temporally resolved N_2^+ concentration in a pulsed atmospheric pressure N plasma.[104] Wang et al.[20] measured concentration and lineshapes of atomic lead transitions at 283 nm in different vertical and lateral positions of an atmospheric argon ICP. Fitting a lineshape into a Voigt lineshape yields Doppler and Lorentzian components, from which plasma gas kinetic temperature and electron density were estimated. In this way, a plasma gas kinetic temperature profile and electron density profile in the plasma were mapped out. Using similar method, the lineshapes of the rotational transition line $S_{21}(1)$ in the OH $A^2\sum^+ - X^2\prod$ (0–0) band were measured and plasma kinetic temperature profile was also derived.[151] Through simultaneous measurements of ringdown absorption spectra and emission spectra of OH radicals in the plasma, the local thermoequilibrium approximation (LTE) was evaluated.[151] One disadvantage of the ringdown measurement in a circular plasma is the line-of-sight measurement, which yields an integrated measurement along the laser beam path. Although Abel inversion can convert the line-of-sight measurements to radial distributions, the conversion process requires a relatively large number of data points in order to achieve an accurate measurement. However, practically, this line-of-sight measurement gives an acceptable measurement uncertainty of plasma gas temperature and electron density, e.g. 10–15%, for plasma diagnostics.[20,151,156]

Another relatively new CRDS-related application is sputtering plasma monitoring. Ion beam sputtering plasma is created when an energetic ion beam bombards the surface of a metallic target. In the process, chemical species, such as small molecules, free radicals, and atoms are sputtered from the target and form a sputtering plasma. Real-time in situ monitoring of sputter erosion is important in the operation of electric propulsion devices[159] and the development of nuclear magnetic fusion.[160] Compared with other optical techniques that are used for sputtering monitoring, such as optical emission spectrometry (OES),[161,162] LIF,[163,164] and resonance-enhanced multiphoton ionization time-of-flight mass spectrometry (REMPI-TOF-MS),[165,166] CRDS is advantageous in the areas of real-time response, high sensitivity, and absolute measurement of number densities. In an earlier work,[35] CRDS was used to measure the number density of silicon (Si) atoms generated by radio discharge plasma sputtering. Recently, CRDS has been used to study sputter erosion of a metallic target.[36,37] In those studies, the ion beam source and the target were housed in a vacuum chamber with a background pressure of 10^{-6}–10^{-7} Torr. Energetic argon ions bombarded a metallic target and elements were generated around the surface of the target. CRDS with a UV laser beam (375–400 nm) from an OPO system was used to detect the densities of the sputtered elements, such as ion (Fe), aluminum (Al), molybdenum (Mo), and titanium (Ti). Detection sensitivities of these elements were in the 10^6–10^7 atoms/cm^3 levels, which were lower than typical number densities of sputtered particles in electric propulsion devices. The analytical applications of CRDS associated with a plasma source are discussed in Section 5.

4.2.3 *Environmental Applications*

Trace chemical detection in environmental applications is another important area in which CRDS is playing an increasingly important role. Soon after the publication of the original CRDS paper, O'Keefe and Lee reported the detection of NO_2 at subpart per billion levels using CRDS.[87] Brown,[167] in 2002, measured NO_3 using narrowband pulsed dye lasers and Ball,[168] in 2003, also measured NO_3 using a broadband pulsed dye laser. Mazurenka et al.[169] demonstrated fast Fourier transform cavity ringdown spectroscopy (FFT-CRDS) for detection of NO_2 with a detection sensitivity of 150 ppt. Simpson developed a CW-CRDS spectrometer to monitor NO_3 and N_2O_5 at 662 nm with a detection limit of 2.4 ppt for N_2O_5 in a 25-s average.[170] Meijer's research group demonstrated the detection of NH_3 at the 10 ppb level, and a detection limit of less than 1 ppt was reported for mercury in room air.[101]

Results such as the above have sparked a keen interest in the use of CRDS for difficult environmental monitoring problems. Vasudev et al. explored the use of CRDS for applications in monitoring offgases from

thermal treatment systems such as hazardous waste incinerators.[29] The real-time detection of low vapor pressure chlorinated organic species, and in particular species such as the dioxins and furans, is highly desirable and extremely difficult. Other ultrasensitive techniques such as LIF and REMPI suffer a reduction in sensitivity for such species due to a strong intersystem crossing between excited state singlet and triplet levels. CRDS, being an absorption method, is much less affected by the fate of the excited state and should possess the sensitivity to implement such measurements. The Orr-Ewing research group reported measurements of trace volatile organic compounds (VOCs), such as ethyne, ethene, 1,3-butadiene, 1-butadiene, 2-methylpropene using two DFB lasers and one ECDL in the NIR.[171] The detection limits of these VOCs are in the subparts per billion to a few parts per million levels. The same group combined a preconcentration technique typically used in gas chromatography (GC) with CW-CRDS and demonstrated the capability of measuring ethane in urban air at a mixture ratio level of 6 ppb.[172] Winstead et al. measured absorption cross sections of the first C−H overtone of several VOCs, such as benzene, toluene, chlorobenzene, and dichlorobenzene using an NIR external cavity diode lasers cavity ringdown spectroscopy (ECDL-CRDS) system.[173,174] More recently, Cias reported measurement of absorption cross sections of the first C−H overtone of 1,3-butadiene, 2-methyl-1,3-butadiene (isoprene), and 2,3-dimethyl-1,3-butadiene using CW-CRDS at 1650 nm.[113] The overlap of the overtone spectra of many large VOCs challenges the spectral selectivity. One possible solution is to measure the fundamental transition in the middle IR using an IR laser source.

With the increasing concern about global warming and more stringent standards for reduction of the greenhouse gas (CO_2 and CH_4) emissions, interest in measurement and monitoring of atmospheric greenhouse gases remains high. CRDS is receiving growing recognition of being able to play an important role in this field. For example, the use of CRDS as an intelligent control system for advanced coal combustion system to monitor typical emission gases (e.g. CO_2, CO, NO, NO_2, N_2O, SO_2) together with total mercury.[175] Early measurements of important atmospheric/environmental molecules, such as CO_2, NH_3, CH_4, and HCHO were performed in a laboratory low-pressure system. Fawcett reported the first measurement of tropospheric methane using CW-CRDS at 1650 nm with a detection limit of 52 ppb.[31] Awtry and Miller developed sensors for air quality monitoring in spacecraft capsules using CW-CRDS at 1550 nm.[32] They demonstrated detection limits for CO and CO_2 at single digit parts per million levels and for NH_3, HCN, and C_2H_2 at a few parts per billion levels. Similar atmospheric molecules, such as, CO, CH_4, NH_3, and C_2H_2, were also measured by CRDS's variant, CEAS or ICOS in the early studies.[176]

One of the important applications in greenhouse gas monitoring is to identify the source of greenhouse gases. The isotopic abundance ratio of C13 to C12 is a reliable signature for source identification. Diode laser CRDS offers a high-resolution and high-sensitivity measurement of the C13 isotope. Crosson et al. developed a sensitive NIR CW-CRDS system for measurement of the isotopic ratio of C13 to C12 in the range of 6261–6264 cm^{-1}.[54] The low-pressure measurement at 70 Torr yields a high accuracy of 0.22 per mil, which is comparable with the results obtained using isotope ratio mass spectroscopy (IRMS). Castrillo et al. also developed a spectrometer for measurement of C13/C12 in CO_2 at 2 μm.[177] The spectrometer has an accuracy of 0.5 per mil for C13. Wang et al. recently developed a battery-powered, remotely accessible, portable ringdown spectrometer for detection of CO_2, CH_4, and C13.[114] The instrument is being tested at the US Department of Energy carbon sequestration sites. Two laser diodes at 1650 and 1596 nm were multiplexed through a fiber splitter for simultaneous measurements of CO_2, CH_4, and C13 in two different wavelength regions. Ding et al. combined 31 DFB laser diodes to measure absorption spectra of $^{13}CO_2$ in the 6130–6750 cm^{-1} and numerous new transitions were added to the CO_2 spectral database.[178] ICOS and DFG-based IR absorption have also been used to measure C13 isotope in the IR spectral regions. Owing to the larger line strength of the fundamental transition in the IR,[179] these methods have also achieved high sensitivity.

Recently, CRDS has been employed for studies of optical extinction by aerosol particles. CRDS measures the total optical extinction resulting from the sum of scattering and absorption losses.[90,180–182] The losses depend on the number density, size distribution, and index of refraction of the particles. Different laser sources including UV and NIR laser have been used to measure extinction caused by particles such as nonabsorbing polymer beads and absorbing smoke particles. Thompson et al. investigated the extinction coefficient of ambient air as a sensitive indicator of micrometer-sized airborne particles through CRDS at two wavelengths.[183] Very recently, Timothy et al. demonstrated the measurement of single aerosol particle extinction using optical feedback cavity ringdown spectroscopy (OF-CRDS) method.[184,185] The OF-CRDS technique allows high repetition rate measurement up to 1.25 KHz, which allows multiple measurements of the aerosol extinction of a single particle.

5 ATOMIC APPLICATIONS OF CAVITY RINGDOWN SPECTROSCOPY

5.1 Introduction

Application of CRDS for atomic absorption measurements was first reported by Jongma et al.[101] who reported measurement of elemental mercury in air using CRDS with a detection limit of less than 1 ppt. Subsequently, Miller and Winstead[18] demonstrated the first P-CRDS measurements for analytical atomic spectrometry by introducing an ICP source into a CRDS system and measuring lead concentrations using a pulsed laser at 283 nm. In analytical P-CRDS, the plasma is used as an atomization source to generate atoms and ions and CRDS serves as a sensitive detector. Several publications have reported continuing development of analytical P-CRDS. P-CRDS has been demonstrated to be a powerful technique not only for high-sensitivity analysis of a variety of elements but also for trace isotopic measurements. During the last 10 years, several elements and isotopes have been measured using the P-CRDS technique. The use of different plasma and laser sources has been demonstrated and detection limits comparable to inductively coupled plasma mass spectrometry (ICP-MS) have been reported. A review of P-CRDS for elemental and isotopic measurements has recently been published.[21,77] This section incorporates an update on the development of the P-CRDS technique since the last review. Emphasis is placed on the improvement of the P-CRDS technique and isotopic measurements.

5.2 Principle of Plasma Cavity Ringdown Spectroscopy

P-CRDS follows the same measurement principles as CRDS. Including losses resulting from the high-temperature atmospheric plasma and the open-air cavity, the observed ringdown time in a P-CRDS system can be expressed as[20] in Equation (11):

$$\tau = \frac{d}{c(1 - R + \beta_{\text{plasma}} l_s + \beta_{\text{air}}(d - l_s) + \text{absorbance})} \tag{11}$$

where d is the distance between the two cavity mirrors, R is the mirror reflectivity, β_{plasma} and β_{air} are the broadband scattering coefficients in the plasma and open-air cavity, respectively, c is the speed of light, and l_s is the laser beam path length through the plasma. In practice, because the scattering coefficients vary slowly over the bandwidth of the cavity mirrors, such losses are simply incorporated into an effective reflectivity R_{eff}. The ringdown time equation can be expressed as in Equation (12):

$$\tau = \frac{d}{c(1 - R_{\text{eff}} + \text{absorbance})} \tag{12}$$

where the absorbance is defined in Equation (5). In the case of plasmas, the density of absorbers is position dependent. As a result of the complicated linewidth broadening mechanisms under atmospheric plasma conditions, the lineshape of the absorption line is of a Voigt profile and the absorption cross section can be represented by Equation (13)[52,186]:

$$\sigma_{ij}(\nu) = \frac{g_j}{g_i} \frac{\lambda^4}{4\pi^2 c} \frac{A_{ji}}{\Delta\lambda_D} V(a, 0) \sqrt{(\pi \ln 2)} \tag{13}$$

where g_j and g_i are the upper and lower state degeneracies respectively, λ is the transition wavelength, c is the speed of light, and A_{ji} is the spontaneous emission transition rate (also known as *transition strength* or *transition Einstein A coefficient*). For most atomic/ionic transitions, these parameters are well documented in the literature. $\Delta\lambda_D$ is the Gaussian component of the broadened linewidth due to Doppler broadening. $V(a, 0)$ is the Voigt function determined by linewidth broadening mechanisms and the a parameter is determined by the ratio of Lorentzian and Gaussian components.[157] In practice, the measured lineshape is typically the spatially averaged and integrated lineshape along the particular line of sight; therefore, an averaged value of the cross section is obtained.

Using the experimental lineshape data to determine the average cross section allows for the measurement of the average density of atoms or ions in the lower state of the transition to be probed. For the purpose of plasma diagnostics, such as measuring the distribution profile of electron density, gas kinetic temperature, and analyte/species density, the line-of-sight measurement may generate a large deviation[20,151,156,187,188]; Abel inversion may be used to convert the lateral measurements to the radial distributions.[20,189] However, for analytical applications in which the detection sensitivity is the quantity of concern, the line-of-sight measurement is appropriate.

5.3 Plasma Cavity Ringdown Spectroscopy for Elemental Measurements

5.3.1 Inductively Coupled Plasma Cavity Ringdown Spectroscopy

The ICP is a robust plasma source and has widely been used as an excitation source for emission spectroscopy study of analytes. Miller and Winstead first coupled an inductively coupled plasma cavity ringdown spectroscopy (ICP-CRDS) to demonstrate general analytical atomic applications of CRDS.[18] The ICP-CRDS system was initially constructed using a pulsed laser source and an ICP source to demonstrate the technical feasibility

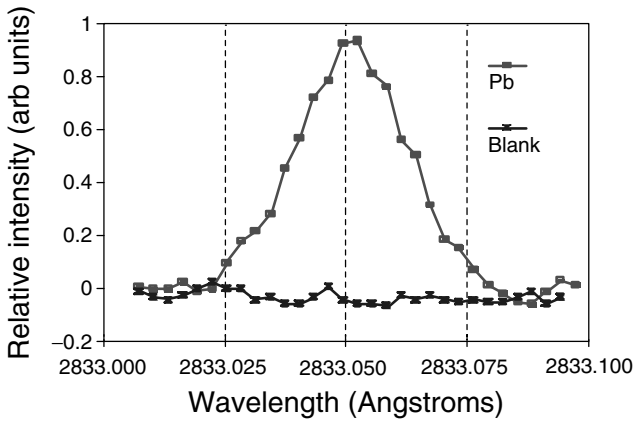

Figure 6 An ICPCRDS wavelength scan recorded 70 mm above the ICP load coil for blank and 100-ppm Pb solutions. (From ref. 18. Reproduced by permission of The Royal Society of Chemistry.)

and to explore the detection sensitivity of the system. A standard lead (Pb) solution was diluted at different concentrations and injected into an ultrasonic nebulizer to generate sample gas mist, which was carried by argon into an atmospheric-pressure argon ICP. The plasma used a standard ICP quartz torch coupled with a 27.12 MHz ICP power supply.[18] The detection limit of Pb at 283-nm transition line was in the microgram per microliter level. Figure 6 shows the first elemental measurement of Pb using the ICP-CRDS technique.[18]

Following the initial introduction of ICP-CRDS,[18] extensive P-CRDS optimization has been investigated by optimizing each operating parameter individually and, subsequently, in combination with the other parameters. The detection limit of the system has been improved to 300 pg mL^{-1} for Pb,[20] which is several orders of magnitude lower than the initially demonstrated detection limit. This detection limit of Pb is much better than the detection limit obtained by ICP-OES, slightly better than inductively coupled plasma laser-induced fluorescence (ICP-LIF), and comparable with that from ICP-MS.[20] It should be noted that the results in that work were obtained with a modified ICP plasma torch, which allowed the ICP to operate at an extremely low power of 200 W. With this modified ICP-CRDS system, Pb, Hg, and Mn were also measured and the detection limits were in the nanogram per microliter to picogram per microliter levels.[20–23,46–48]

5.3.2 Microwave Induced Plasma Cavity Ringdown Spectroscopy

The measurement of Pb by the low-power ICP-CRDS system inspired the exploration into different plasma sources, especially those capable of operating at low plasma powers. Microwave induced plasma (MIP) is a powerful alternative source for atomic excitation in elemental analysis and has been extensively used in analytical atomic spectrometry.[190,191] Compared with other types of plasma sources, an MIP offers some attractive characteristics, such as low power, low gas flow rates, and low cost for instrumentation and maintenance, making it a desirable source for absorption measurements.[192] Duan et al. have investigated coupling an microwave induced plasma cavity ringdown spectroscopy (MIP-CRDS).[23] In that study, lead was chosen as a typical element for the system demonstration so that the results could be directly compared to those from an ICP-CRDS system. A detection limit of 0.8 ng mL^{-1} for Pb was obtained in that initial feasibility study.

Physically, an MIP-CRDS system has a smaller instrument footprint, lower weight, and lower operating costs than an ICP-CRDS system. MIP-CRDS also has one unique analytical merit that is not offered by ICP-CRDS: the baseline noise, σ/τ, is not affected by the presence of the plasma. In the ICP-CRDS system, the ringdown baseline noise was three to four times larger when the plasma was on than when the plasma was off.[20] However, it was found that the MIP used in the MIP-CRDS system did not generate additional baseline noise to the system. Whether the plasma was on or off had no effect on the baseline noise, which was the same at about 0.3% in both on and off conditions. These results indicate that the baseline noise is improved by a factor of 3–4 by using an MIP in place of an ICP. It should be noted that in addition to the plasma associated baseline noise, the laser beam position in the plasma and the plasma gas flow rates also influence the detection sensitivity. Very recently, a low-power miniature microwave plasma torch (MPT) has been employed with CRDS to measure elemental mercury with a plasma operating power of only 15 W.[48]

Diode lasers have also been employed for atomic spectrometry in P-CRDS. An external cavity diode laser operating at 679 nm was used to measure strontium (Sr).[46] In that study, diluted standard solutions of strontium (Sr) were introduced into the plasma by an in-house fabricated sampling device combined with an ultrasonic nebulizer. SrOH radicals[193–195] were generated in the plasma, and the detection of Sr was made by measuring SrOH radicals in the plasma. This study demonstrated the potential of P-CRDS instrumentation using a compact CW laser source.

5.4 Plasma Cavity Ringdown Spectroscopy for Isotopic Measurements

Isotopic analysis is an important field in analytic atomic spectroscopy. For example, in the field of environmental monitoring, isotopic abundance is often

used as spectral fingerprint to study transform, fate, and chemical processes of analytes. Most isotopic analyses are performed by a laboratory ICP-MS system. However, these systems require tedious sample preparation and sophisticated operation and maintenance, and they are not portable for real-time, on-site analysis. The ICP-MS system can also suffer from isobaric effects. For instance, an ICP-MS system cannot readily distinguish ^{238}U from ^{238}Pu. In early studies, an ICP was combined with the LIF technique (ICP-LIF) to measure U isotopes at 286 nm. Uranium has a strong transition at 286 nm and the isotopic shift is of 0.6 cm^{-1}.[196] With the ICP-LIF technique, the measurements of ^{235}U and ^{238}U in different matrices were conducted with a detection limit of 1 $\mu g\,mL^{-1}$.[122] LIF combined with laser ablation was also implemented to measure isotope ratios of ^{235}U transition at 682.67 nm and ^{238}U at 682.69 nm under low-pressure conditions and achieved a detection limit of 0.6 $\mu g\,mL^{-1}$.[197]

The ICP-CRDS system was applied to isotopic measurements of uranium at three different wavelengths: 286, 358, and 409 nm.[21] The U transitions at 286 and 409 nm are ionic transitions, but the U transition at 358 nm is an atomic transition. With the optimized ICP-CRDS system, the detection limits of U and its isotopes were in the nanogram per microliter levels.[21] Figure 7 shows typical isotopic measurements of uranium with a 50%/50% mixture of ^{235}U and ^{238}U. ^{235}U existing in ^{238}U with the natural abundance ratio of 0.79% was also measured.[21] This research demonstrated high-resolution isotopic analysis with high detection sensitivity using the P-CRDS technique.

The Hg transition at 254 nm has seven stable isotopic components with well-documented abundance ratios.[198–200] An early ICP-CRDS system was used

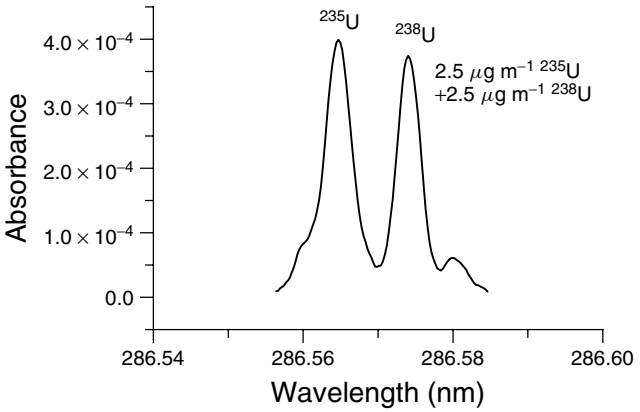

Figure 7 Uranium isotopes at 286 nm measured by using ICP-CRDS. *(Appl. Spectrosc.*, 2003, **57**, 1167–1172 by Wang et al., Figure 3 on page 1169.)

to measure a low-resolution hyperfine contour of this transition.[201] The isotopic abundance ratio of the stable isotopic components was barely quantified. The linewidth broadening under the high-temperature ICP ultimately limited the instrument's capability of measuring high-resolution hyperfine structures of the Hg transition at 254 nm. With the lower plasma temperature in the MIP, the Hg hyperfine structures were more clearly resolved.[47] With a newly developed tube-shaped MIP torch (MPT), which allowed the plasma to operate at powers as low as 15 W, high-resolution Hg hyperfine structures were also recorded as shown in Figure 8.[48]

When the three different plasma sources (ICP, MIP, and MPT) were applied to the P-CRDS system for Hg measurements, interesting spectral background information was obtained.[47] With the ICP-CRDS system, no significant spectral interference was observed. However,

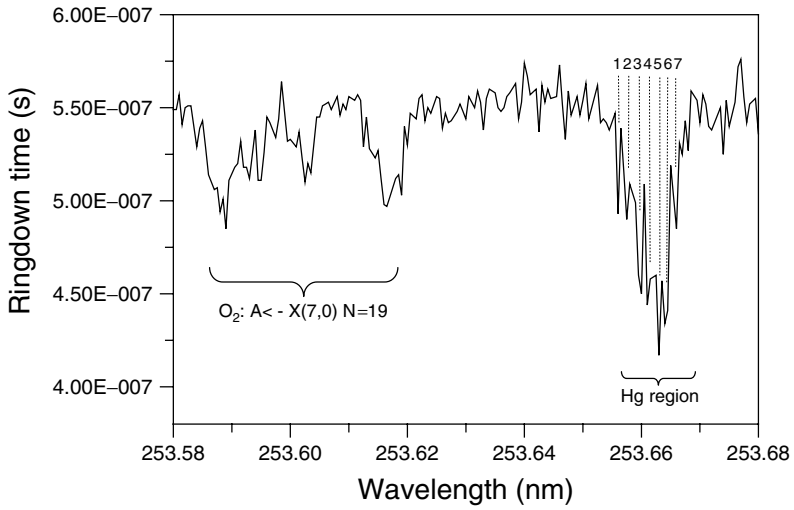

Figure 8 Hyperfine structures of mercury transition at 254 nm measured by using MIP-CRDS. (Reprinted with permission from ref. 48. © 2005 American Chemical Society.)

with the MIP-CRDS system, a strong OH transition located at 253.65 nm generated a significant spectral interference with the absorption peak of Hg at 254 nm. This OH transition was attributed to the overlap of two broadened rovibrational transitions $R_{21}(21)$ and $P_1(15)$ of the OH A-X (3–0) band.[47] The spectral interference limited the detection sensitivity of the Hg. This OH interference observed when using the MIP was remarkably eliminated by using a tube-shaped MIP torch, which was constructed in a long quartz tube.[48] The disappearance of the OH-spectral interference in the microwave plasma torch cavity ringdown spectroscopy (MPT-CRDS) is partially due to the fact that the MPT has a very small cross-sectional area at the end of the tube interacting with the plasma surroundings.

5.5 Electrothermal Atomization Cavity Ringdown Spectroscopy

An atomization technique that would also appear to lend itself well to CRDS is electrothermal atomization (ETA). As with other conventional absorption techniques, electrothermal atomization atomic absorption spectroscopy (ETA-AAS) attempts to measure a small change in the total transmitted intensity of a light source through an absorbing medium. For a commercial ETA-AAS system, the typical absorbance detection may approach 10^{-4} at best. Comparing this value to the 10^{-5}–10^{-8} absorbance for CRDS, the potential of coupling the two techniques together is obvious. An estimate of potential electrothermal atomization cavity ringdown spectroscopy (ETA-CRDS) performance can be obtained by comparison with modern commercial ETA-AAS systems, where a Pb characteristic mass of approximately 5 pg is typical.[202] The characteristic mass implies an absorbance of 4.4×10^{-3}. By again assuming an approximately 10^{-5} absorbance detection and scaling the mass linearly with the absorbance, the CRDS detection limit is expected to be in tens of femtogram range.

Winstead, Miller, et al. successfully demonstrated the first use of ETA-CRDS for measurement of elemental mercury.[19] The introduction of the graphite furnace within the cavity requires simply that the furnace be stable and that the laser beam pass through the center of the graphite tube. In this initial study, the furnace housing windows were removed to reduce optical loss and small holes at the ends of the graphite furnace were introduced to pass the laser beam. Some evidence of the presence of oxygen within the system, especially at high temperatures, was noted. Mounting the mirrors directly on the furnace housing with an argon gas purge would eliminate this problem. Once again, the primary loss mechanisms governing the decay time of the cavity are mirror reflectivity losses, atomic absorption in the graphite furnace, plus Rayleigh scattering from air in the cavity. As CRDS is based on measuring an intensity decay rate, it discriminates effectively against the furnace blackbody radiation.

Owing to the transient nature of the atomic absorption signal in a graphite furnace, a fast, high-resolution data-acquisition system is required so that integrated absorbance values can be obtained for each atomization cycle. The PMT signal was digitized by a 12-bit, 100 MHz sample rate analog-to-digital converter with the timing for data acquisition determined by the software controlling the graphite furnace. Eighty waveforms were saved over a period of 4 s for each atomization cycle, from which the peak and integrated absorbance values were determined. Figure 9 depicts the response of $1/\tau$ for lead (283 nm) as a function of the data-acquisition time following the start of the atomization step. Using a system having effective mirror reflectivity of 99.86% and a baseline ringdown time variation of 2–3% yielded detection limits of 1 and 2.7 pg for lead and mercury, respectively.[19] These results are an order of magnitude better for mercury, but only slightly better for lead, than the typical detection limit obtained by ETA-AAS.[202] Additional enhancements can come from further optimization of the laser linewidth, increased cavity stability during the furnace atomization sequence, a faster duty cycle for data acquisition, and optimization of the furnace operation itself. For example, the atomization process within a graphite furnace is generally complete in under 1 s while an Nd : YAG laser repetition rate is only 10–100 Hz (20 Hz in the example above). This implies that significant benefit would accrue from the use of higher repetition rate lasers.

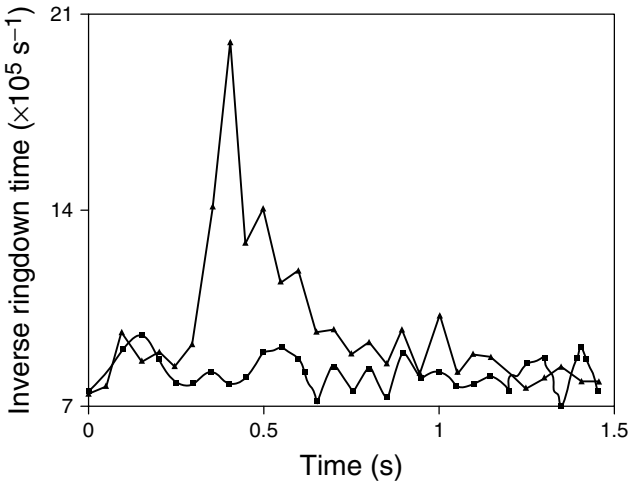

Figure 9 Analytical shape of lead atomization peak obtained from cavity ringdown time measurements by injecting 7.5 μl of 2 ng ml^{-1} lead standard solution for each heating cycle. (From ref. 19. Reproduced by permission of The Royal Society of Chemistry.)

Recent work has focused on developing a CRDS mercury emission monitor. Such an instrument requires, among other things, a determination and elimination of any background absorptions or interferences caused by other constituents of flue gas emissions from coal-fired power plants. For example, it is well known that sulfur dioxide (SO_2) absorbs UV radiation in the same spectral region as mercury (253 nm) resulting in a significant problem for most mercury detection equipment. Indeed, presently used commercial Continuous Emission Monitors (CEMs) based on atomic emission, atomic fluorescence, or atomic absorption require the removal of SO_2, and other flue gas constituents, before an accurate determination of the mercury concentration can be made. Recent research using CRDS to monitor mercury emissions from coal-powered utilities used thermal decomposition to convert mercury compounds to elemental mercury to provide a total mercury measurement. It has been noted[20,47,201] that the presence of isotopic structure during a scan of the mercury 253-nm line raised the possibility of separating these two absorbance components. Using a diode-seeded alexandrite laser, this possibility was confirmed. That is, the characteristic Hg isotopic peaks could be used to provide a unique identifier for mercury in exhaust gases.[175,203] For example, using a gas mixture containing 1.2 ppb mercury and 2175 ppm SO_2 respectively, the five mercury peaks obtained using CRDS are easily resolved above the SO_2 background absorption, and it was possible to accurately extract both the Hg and SO_2 concentrations. In 2004, Miller performed a field test of this instrument at a Department of Energy (DOE) Pittsburgh facility.[175] The results obtained support the conclusion that CRDS is a very promising technique for monitoring total mercury. However, practical success of such instruments dependent on the development of compact and low-cost UV laser sources remains a challenge.

6 MEDICAL APPLICATION OF CRDS–BREATH ANALYSIS

6.1 Introduction

Normal human breath contains hundreds of VOCs in very low concentrations ranging from part per trillion to part per billion levels.[204] Some VOCs have been identified as biomarkers for specific diseases. For instance, alkanes are present in the case of lung cancer and formaldehyde in the case of breast cancer[205]; the presence of isoprene in human breath is related to blood cholesterol levels[206]; and patients with Type 1 diabetes (T1D) or on ketogenic diets have acetone in their breath.[207,208] Such knowledge suggests that breath analysis is useful for human disease diagnostics. However, owing to the low concentration and the large number of chemical species in exhaled breath, breath analysis requires highly sensitive and selective instruments to detect and identify the abnormal concentrations of specific biomarkers. One major instrument presently employed for breath VOCs analysis is gas chromatograph mass spectrometry (GC-MS).[209–213]

Owing to its unique features of high sensitivity and absolute measurement of gas concentration of species of interest, CRDS has been well recognized to be an excellent technique for real-time, noninvasive breath gas analysis. However, its application in breath gas analysis is still in its infancy. At the time of this article being written, there have been only a few publications reporting on breath analysis using CRDS[54,55,214] or using its variant – cavity leak-out spectroscopy (CALOS),[215–217]/ICOS.[218] However, the application of CRDS to breath analysis is expected to increase. Recent reviews on breath analysis are available.[56,219–221]

6.2 Breath Analysis

Monitoring of C13 isotopic ratio in breath gas can help diagnose the presence of bacterium *Helicobacter pylori*. Crosson et al.[54] measured carbon isotopic ratio of C13 to C12 in breath CO_2 using CW-CRDS in the NIR spectral region. For the original form of CRDS, this is likely the first report on medical application of CRDS. They used a DFB laser diode at 1657 nm and performed measurements at low pressure (70 Torr). The detection limit of δ C13 was 0.22 per mil and was compared with the results from a commercial isotopic spectrometer. Wahl et al.[222] reported recent progress on C13 isotope measurement using CW-CRDS, also in the NIR spectral region. Using a preconcentration process, they demonstrated the capability of measuring the C13 isotope in both high-concentration CO_2 (e.g. 4%) and low-concentration CO_2 (e.g. 350 ppm) with an accuracy of 0.2 per mil. Wang et al.[55] developed a compact ringdown device using a miniature 266-nm laser source for detection of acetone, a biomarker of diabetes in breath. The instrument has sufficient detection sensitivity to detect abnormal breath acetone. The instrument was further improved and evaluated in the areas of detection sensitivity and response time with acetone sample solutions.[223] They demonstrated that a single measurement of acetone can be done within 46 s using the prototype acetone breath analyzer. The detection limit of acetone was 0.49 ppm. Very recently, this acetone detection device has been reconfigured to be a portable unit and tested with 59 human subjects including type 1 and type 2 diabetic outpatients in a clinic.[224]

Instead of using CRDS, several studies on breath biomarkers using the CRDS variants CALOS, CEAS,

and ICOS, have been reported. Ethane and pentane in human breath gas are biomarkers for lipid peroxidation and oxidative damage of human tissue. Murtz et al.[215] first reported on the measurement of breath ethane and pentane using mid-IR CALOS. The detection limit of ethane is 100 ppt. The same group also demonstrated on-line monitoring of breath ethane using the CALOS at 3.3 μm,[216] which was generated by periodically poled lithium niobate (PPLN) pumped by an Nd:YAG laser. The detection sensitivity was later improved to 6 ppt by using a mid-IR OPO laser source, and multiple breath compounds in human breath gas such as methane, ethane, and water were simultaneously analyzed.[217] Halmer et al.[225] recently reported on a mid-IR CALOS spectrometer and demonstrated the detection of breath carbonyl sulfide (OCS) with a detection limit of 7 ppt. In that study, multiple species in a single breath, such as ethane, CO_2, and O_2 were simultaneously measured. The temporal resolution of the breath-ethane concentration was clearly resolved within the first 500 s. In comparison, the GC method used in that research required 30–60 min for a measurement and the CALOS needed less than 1 min, as shown in Figure 10. The method demonstrated in this work is expected to be extended to the analysis of other small molecules such as NO, CO, OCS, and other hydrocarbons. Bakhirkin et al.[226] have demonstrated an ICOS-based system using a QCL at 5.2 μm to measure breath NO for diagnostics of lung disease. Very recently, McCurdy et al. have demonstrated real-time simultaneous measurement of NO and CO_2 in a single breath cycle using ICOS with two QCLs operating at 5.45 and 5.22 μm. They demonstrated the detection limit of NO at 0.4 ppb with a 1 s integration time.[227]

The Ye research group has recently introduced a novel cavity-based technique for analysis of multiple breath species in a broad spectral range. They reported human breath analysis using an optical frequency comb fiber laser combined with CEAS over a range of 200 nm in the NIR. The measured mixture gas includes CO, CO_2, CH_4, H_2O, and NH_3.[228]

Human breath gas analysis dates back to more than three decades ago. However, real-time, on-line breath gas analysis has been made possible only recently by CRDS or its variants, CALOS, CEAS, and ICOS. The promising application of CRDS in breath analysis for disease diagnostics and metabolic status monitoring can be seen in recent reviews and related publications.[56,219–221,227,229] A current challenge in breath analysis using CRDS and its variants is to eliminate the spectral interference from many other species present in human breath. To address this issue requires more studies on the spectral fingerprints of human breath compounds. Breath gas analysis by CRDS is in its infancy, but we believe that this will be one of the fastest growing areas of CRDS application. Ringdown breath analyzers are expected to become a commercial reality and replace the time-consuming, expensive, and nonportable GC-MS for point-of-care (POC) breath analysis in some situations in the near future.

7 FIBER LOOP RINGDOWN TECHNIQUES

7.1 Introduction

One of the essential characteristics of CRDS is enhanced detection afforded by the long path/multiple interaction approach. This feature has been recently implemented by

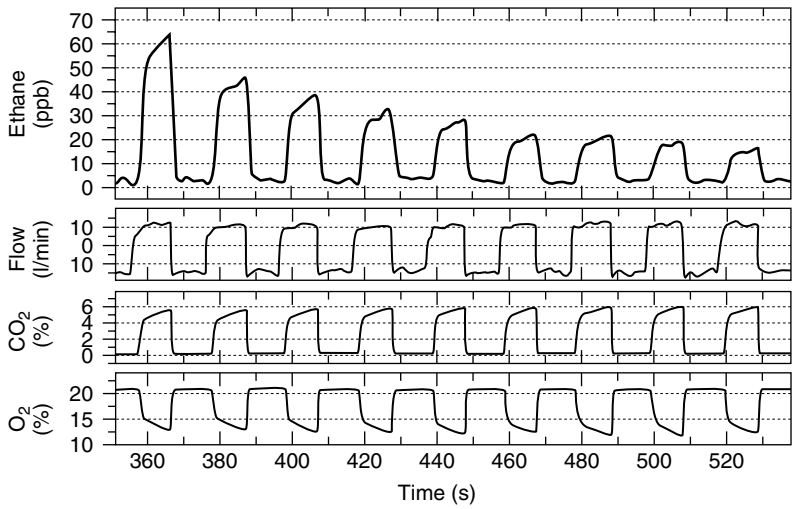

Figure 10 Typical measurement of single-breath concentrations of ethane, CO2, and O2. Each single expiration is analyzed separately. Only the first 500 s are displayed. The complete measurement took 42 min. (From ref. 216. Reproduced by permission of Springer. © 2006.)

using a ''conceptual cavity'' – a fiber loop formed by a section of single-mode fiber, thus being generally referred to as *fiber loop ringdown*.[91] This method is modeled after the ringdown concept. However, a conceptually new approach is used, eliminating dependence on an ultrahigh reflectivity cavity. This new fiber ringdown technique utilizes an optical resonator – an optical fiber loop – as the ringdown ''cavity''. Light is coupled into the fiber loop; and when the light source is rapidly shut off, the resultant light rings down inside the fiber loop for many round trips. In each round trip, a small fraction of light leaks to a photodetector through a fiber coupler; and the remaining light rings in the fiber experience internal fiber transmission losses. The signal intensity observed by the detector follows an exponential decay. The lower the losses of the light in the fiber, the longer the decay time constant (ringdown time). When an external action, such as an absorption or evanescent wavefield interaction is applied to a small section of the fiber, the observed ringdown time decreases. This type of fiber ringdown technique functionally resembles the standard high-reflectivity CRDS for absorbance measurements, but without the requirements of high-reflectivity parts.

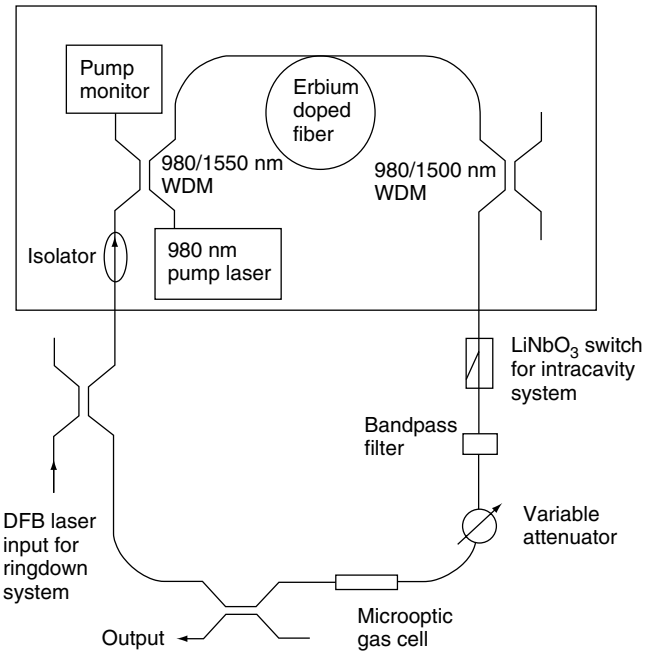

Figure 11 Schematic of a typical fiber loop ringdown experimental system. (Reproduced from ref. 91 by permission of IOP Publishing Ltd.)

7.2 Fiber Loop Ringdown Chemical Sensors

Culshaw et al. first introduced an optical fiber loop with a length of several tens of meters with a 5-cm open path microoptical gas cell for gas-phase absorption measurement.[91] Figure 11 shows a schematic of the fiber loop ringdown system. Owing to the large insertion loss from the gas cell, an additional Erbium-doped fiber loop was included in the fiber loop to operate just below the threshold for net amplification of the light in the loop. Owing to the excessive optical loss of the laser power in the sensor head (the cell) and large optical insertion losses in the fiber connection/coupling points in the fiber loop, the device showed a short ringdown time, ~100 ns and poor detection sensitivity. On the basis of the same instrument scheme, the same group recently demonstrated an improved version of the device, which yielded ringdown times of up to hundreds of microseconds.[94] Loock's group reported on a simplified fiber loop ringdown technique by introducing a micro air gap into a section of fiber in the fiber loop for detection of a small volume of dye solutions.[6,230] They demonstrated the measurement of 7×10^{-15} mol dye solution using both a CW laser and a pulsed laser. In this study, the laser pulse was coupled into the fiber loop through a direct illumination of a bend in the fiber loop and the detection of the laser pulse leaking out of the fiber was achieved through another bend in the fiber loop. The same group further advanced the fiber loop ringdown technique by introducing phase-shift measurement, which greatly improved the data-acquisition rate to close to real time (10–100 ms).[60] This technique has been demonstrated to be suitable for low-cost, real-time, and on-line detection of capillary electrophoresis with a detection limit at micromolar concentration levels. Using flow injections, the device can detect a series of solution samples at different concentrations. The demonstrated detection limit is 5.3×10^{-12} mol sample in a 530 p-liter volume.[231] Very recently, a minimum fractional absorption of 6 cm^{-1} has been demonstrated by using the fiber loop ringdown technique with a fast gain switch diode laser.[232] In an early publication, Tarsa et al. reported their study on an optical fiber resonator for CRDS, in which the sensor head was made of a section of tapered fiber in the loop and evanescent field absorption was detected.[5,59] A small volume of liquid sample was detected by this type of fiber loop ringdown technique. They also demonstrated the detection of a single-cell adsorption event.[61] Vogler et al. developed a similar fiber loop ringdown device and demonstrated measurement of the diffusion coefficient of hydrogen on silicon through monitoring the OH absorption in the NIR region.[233]

7.3 Fiber Loop Ringdown Physical Sensors

In addition to chemical sensor development using the fiber loop ringdown technique, physical sensors using the fiber loop ringdown concept have been recently demonstrated. Wang and Scherrer extended the fiber loop ringdown

technique for development of low-cost, fast-response, and large-measuring-range pressure sensors.[57,58,234] A section of bare single-mode fiber with a length of 1 cm was used as the sensor head. The sensing principle is primarily based on the fact that micromechanical deformation of the fiber drastically increases optical loss in the fiber loop.[235] The sensor showed repeatable response to pressure, as shown in Figure 12. By using a different baffle layer outside of the bare fiber, different sensing ranges of pressure and/or force can be achieved by the fiber loop ringdown pressure/force sensor. This technique can be further developed for detection and monitoring of mechanical fatigue of materials in an inaccessible and/or harsh environments.

Using a similar fiber loop ringdown approach, Wang and Mbi introduced an optical fiber Bragg grating loop for temperature sensor development.[62,65,236] The Bragg wavelength is temperature dependent. Change of the temperature in the sensor head fiber Bragg grating (FBG) shifts the relative position of the laser wavelength in the FBG wavelength-dependent bandwidth curve, and thus the temperature change results in different transmission rates of the laser beam through the FBG. Different optical losses due to the shift of the FBG curve resulting from the temperature change in the FBG are detected by measuring the ringdown time. One advantage of the fiber FBG loop ringdown temperature sensors is high temperature accuracy. In that work,[65] an accuracy of 0.06 °C was demonstrated in the temperature range of 45–78 °C. Another advantage of the FBG loop ringdown temperature sensor is the low cost compared with the current FBG-optical spectral analyzer (FBG-OSA) temperature sensor. In the FBG loop ringdown temperature sensor, an inexpensive photodiode replaces the expensive optical spectral analyzer (OSA) as the detector. A detailed discussion on the advantages and disadvantages of the FBG loop ringdown temperature sensors compared with the FBG-OSA temperature sensors can be found elsewhere.[65] A schematic of fiber grating loop ringdown sensor is shown in Figure 13.

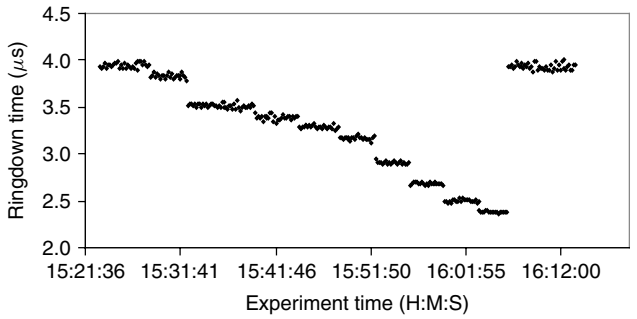

Figure 12 Fiber loop ringdown pressure sensor's response. (From ref. 57. Reproduced by permission of Optical Society of America.)

By using a long period grating (LPG) as the sensor head, the fiber grating loop ringdown temperature sensor demonstrated a large temperature measuring range, from −169 to 950 °C.[62,65] More recently, a double loop dual sensor head configuration has been developed for multiple location sensing.[237] Using high-temperature composite fiber materials to fabricate a sensor head, such as an FBG or an LPG, high-temperature, fast-response,

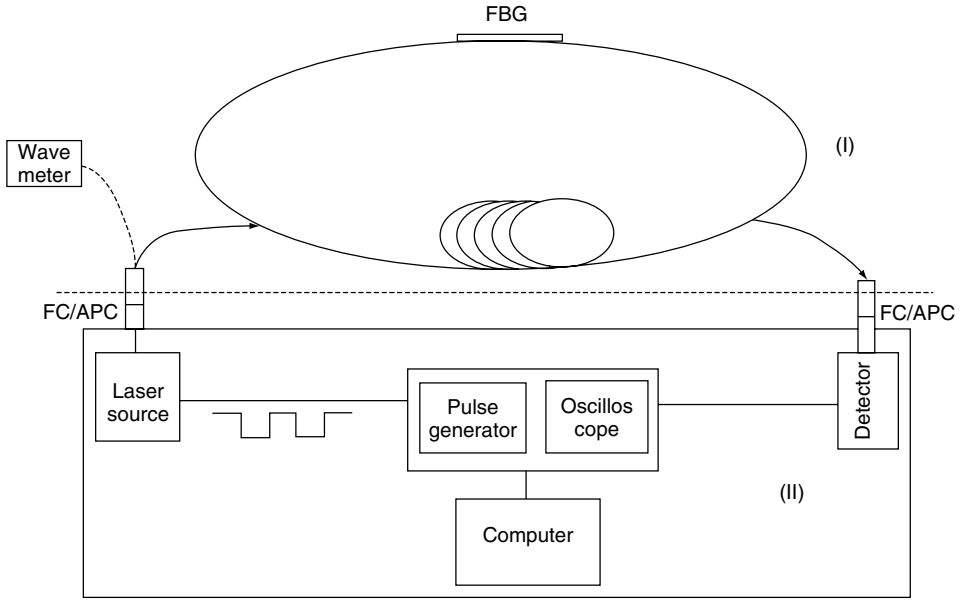

Figure 13 Schematic of a typical fiber grating loop ringdown temperature sensor. Section I: the fiber grating loop with an FBG as the sensing element; Section II: the sensor's control portion. (Reproduced from ref. 65 by permission of IOP Publishing Ltd.)

low-cost, fiber grating loop ringdown temperature sensors can be developed for a variety of applications in harsh environments.

8 RINGDOWN INSTRUMENTS

It has been two decades since the first publication of the CRDS technique. Although CRDS has become a mature technique, manufacturing and sale of commercial ringdown instruments for a nonresearch market is a relatively recent event. Owing to the intrinsic nature of the ringdown approach, (e.g. extreme sensitivity to optical and mechanical stability) transitioning the CRDS technique from a research laboratory to a commercial reality is a significant science and engineering achievement. This stride was first successfully made by scientists and engineers at Tiger Optics in 2002[238] with the introduction of a moisture analyzer, the MTO-1000-H_2O. A product line (e.g., H_2O, CH_4, and O_2 ringdown analyzers and tracers) has been subsequently created. Instruments for measuring low concentrations of water and methane in high-purity semiconductor manufacturing gases or in specialty gases are being distributed all over the world. In addition to engineering challenges, a common challenge in ringdown instrumentation is to reduce the instrument cost. Cost is mainly determined by the price of laser sources and high-reflectivity mirrors. Laser cost may not be an issue for instrumentation using inexpensive telecommunications diode lasers; however, for an instrument needing a UV laser source, the availability of compact and low-cost UV laser sources remains a challenge. The effect of mirror cost is diminished when mass production of a ringdown product is implemented. Coating 40 identical high-reflectivity mirrors at a specific wavelength does not cost much more than coating one pair.

9 PERSPECTIVE AND FUTURE DEVELOPMENTS

During the last two decades, CRDS has experienced a full technological cycle, from the introduction of the concept, laboratory fundamental research, to full commercialization of the ringdown instrument. CRDS has had a relatively short cycle through the invention–development–commercialization process. From the analytical spectroscopy point of view, CRDS has brought about a revolutionary change in terms of ultrahigh sensitivity, simplicity, and absolute measurement of trace chemical species in close to real time. The development of this powerful analytical technique coincides with the rapid expansion of new semiconductor materials and telecommunications technologies, which promote rapid advancement of new laser sources and optical fiber components, such as compact, inexpensive, fiber pigtailed laser diodes. The new generation of laser sources, optical coating/polishing technologies, and advanced fiber components allows for continued aggressive development of ringdown instrumentation and applications. CRDS has been demonstrated for trace analysis, environmental monitoring, material processes, medical applications, remote sensing, etc. CRDS has been extensively used to study and detect gas-phase species, and its applications in liquid detection are being aggressively pursued. Theoretical considerations for a total internal reflection ring cavity for thin films and liquids were investigated by Pipino et al. in the early days of CRDS development.[2,239] Engeln et al. have also demonstrated the spectroscopy of thin films by inserting a coated optical flat into a standard ringdown cavity.[100] Loock's group demonstrated promising real-time analytical applications in microfluids using the fiber loop ringdown technique.[63] The detection of a single-cell absorption event represents a start of the ringdown applications in a new field.[61] Snyder and Zare first demonstrated the novel combination of CRDS with high-performance liquid chromatography (HPLC) in 2003.[25] They designed a flow cell with wedged angles that minimized optical losses through all cell surfaces at Brewster's angle; this flow cell was coupled to the output of an HPLC separation for the detection of separated analytes by their absorption properties using CRDS with a pulsed laser. The technique has been extended to use a CW laser and has demonstrated the potential of liquid-phase CW-CRDS to be used as an absorption detector for HPLC measurements.[25] Very recently, Sneppen et al. have demonstrated liquid chromatography (LC)–CRDS using an OPO system to achieve measurements in UV region.[240] Being a powerful measurement technique, the technology associated with the ringdown concept is sure to continue to find new areas of application, such as security screening, food safety inspection, plenary exploration, clean energy study, disease diagnostics, and remote sensing.

ACKNOWLEDGMENTS

The authors wish to acknowledge the following support: National Science Foundation Grants CTS-0626302, ATM-0352926 (Wang). Department of Energy Grants DE-FG26-06NT42686, DE-PS26-01NT41048, and Sensor Research and Development Corporation (Miller). Department of Energy Environmental Management Science Program Grants DE-FG07-97ER62517 (Miller

and Winstead), DE-FG07-02ER63515 (Winstead and Wang).

ABBREVIATIONS AND ACRONYMS

AOM	Acousto-optical Modulator
CALOS	Cavity Leakout Spectroscopy
CARS	Coherent Anti-Stokes Raman Spectroscopy
CEAS	Cavity Enhanced Absorption Spectroscopy
CRDS	Cavity Ringdown Spectroscopy
CW	Continuous Wave
CW-CRDS	Continuous Wave Cavity Ringdown Spectroscopy
DFB	Distributed Feedback
DFG	Difference Frequency Generation
DFWM	Degenerate Four-wave Mixing
ECDL	External Cavity Diode Lasers
ECDL-CRDS	External Cavity Diode Lasers Cavity Ringdown Spectroscopy
ETA	Electrothermal Atomization
ETA-AAS	Electrothermal Atomization Atomic Absorption Spectroscopy
ETA-CRDS	Electrothermal Atomization Cavity Ringdown Spectroscopy
FBG-OSA	Fbg-optical Spectral Analyzer
FFT-CRDS	Fast Fourier Transform Cavity Ringdown Spectroscopy
GC	Gas Chromatography
GC-MS	Gas Chromatograph Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ICOS	Integrated Cavity Output Spectroscopy
ICP	Inductively Coupled Plasma
ICP-CRDS	Inductively Coupled Plasma Cavity Ringdown Spectroscopy
ICP-LIF	Inductively Coupled Plasma Laser-induced Fluorescence
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
IRMS	Isotope Ratio Mass Spectroscopy
LC	Liquid Chromatography
LIF	Laser-Induced Fluorescence
LPG	Long Period Grating
LTE	Local Thermoequilibrium Approximation
MIP	Microwave Induced Plasma
MIP-CRDS	Microwave Induced Plasma Cavity Ringdown Spectroscopy
MPT	Microwave Plasma Torch
MPT-CRDS	Microwave Plasma Torch Cavity Ringdown Spectroscopy
OES	Optical Emission Spectrometry
OF-CRDS	Optical Feedback Cavity Ringdown Spectroscopy
OPO	Optical Parametric Oscillator
OSA	Optical Spectral Analyzer
P-CRDS	Plasma Cavity Ringdown Spectroscopy
POC	Point-of-care
PPLN	Periodically Poled Lithium Niobate
QCLs	Quantum Cascade Lasers
REMPI	Resonantly Enhanced Multiphoton Ionization
REMPI-TOF-MS	Resonance-Enhanced Multiphoton Ionization Time-of-flight Mass Spectrometry
T1D	Type 1 Diabetes
VOCs	Volatile Organic Compounds
UV	Ultraviolet

REFERENCES

1. A. O'Keefe, D.A.G. Deacon, "Cavity Ring-Down Optical Spectrometer for Absorption Measurements Using Pulsed Laser Sources", *Rev. Sci. Instrum.*, **59**, 2544–2551 (1988).
2. A.C.R. Pipino, J.W. Hudgens, R.E. Huie, "Evanescent Cavity Ring-Down Spectroscopy with a Total-Internal-Reflection Minicavity", *Rev. Sci. Instrum.*, **68**, 2978–2989 (1997).
3. K.K. Lehmann, P. Rabinowitz, 'High-Finesse Optical Resonator for Cavity Ring-Down Spectroscopy Based Upon Brewster's Angle Prism Retroreflectors', US Patent No. 6,097,555, August 2000.
4. M. Gupta, H. Jiao, A. O'Keefe, "Cavity-Enhanced Spectroscopy in Optical Fibers", *Opt. Lett.*, **27**, 1878–1880 (2002).
5. P.B. Tarasa, P. Rabinowitz, K.K. Lehmann, "Passive Optical Fiber Resonator for Cavity Ringdown Spectroscopy", *224 ACS National Meeting*, Boston, 18–22 August 2002.
6. R.S. Brown, L. Kozin, Z. Tong, R.D. Oleschuk, H.-P. Loock, "Fiber-Loop Ring-Down Spectroscopy", *J. Chem. Phys.*, **117**, 10444–10447 (2002).
7. J.J. Scherer, D. Voelkel, D.J. Rakestraw, J.B. Paul, C.P. Collier, R.J. Saykally, A. O'Keefe, "Cavity Ringdown Laser Absorption Spectroscopy and Time-of-Flight Mass Spectroscopy of Jet-Cooled Copper Silicides", *Chem. Phys. Lett.*, **245**, 273–280 (1995).

8. A.J. Hallock, E.S.F. Berman, R.N. Zare, ''Use of Broadband, Continuous-Wave Diode Lasers in Cavity Ring-Down Spectroscopy for Liquid Samples'', *Appl. Spectrosc.*, **57**, 571–573 (2003).
9. B.A. Paldus, C.C. Harb, T.G. Spence, R.N. Zare, C. Gmachl, F. Capasso, D.L. Sivco, J.N. Baillargeon, A.L. Hutchinson, A.Y. Cho, ''Cavity Ringdown Spectroscopy Using Midinfrared Quantum-Cascade Lasers'', *Opt. Lett.*, **25**, 666–668 (2000).
10. J.B. Paul, J.J. Scherer, A. O'Keefe, L. Lapson, J.G. Anderson, C. Gmachl, F. Capasso, A.Y. Cho, ''Infrared Cavity Ringdown and Integrated Cavity Output Spectroscopy for Trace Species Monitoring'', *Proc. SPIE: Int. Soc. Opt. Eng.*, **4577**, 1–11 (2001).
11. D. Romanini, J. Gambogi, K.K. Lehmann, *Proceedings of the 50th International Symposium on Molecular Spectroscopy*, ed. T.A. Miller, Department of Chemistry, Ohio State University, Columbus, 284, 1995.
12. D. Romanini, A.A. Kachanov, F. Stoeckel, ''Cavity Ring-Down Spectroscopy: Broad Band Absolute Absorption Measurements'', *Chem. Phys. Lett.*, **270**, 538–545 (1997).
13. Y. He, B.J. Orr, ''Rapid-Swept, Continuous-Wave Cavity Ringdown Spectroscopy with Optical Heterodyne Detection: Single- and Multi-Wavelength Sensing of Gases'', *Appl. Phys. B*, **75**, 267–280 (2002).
14. R.N. Zare, J. Martin, B.A. Paldus, 'Deflecting Light into Resonant Cavities for Spectroscopy', US Patent No. 5,815,277, September 1998.
15. R. Augstine, C.R. Krusen, C. Wang, W.B. Yan, 'System and Method for Controlling a Light Source for Cavity Ringdown Spectroscopy', US Patent NO. 7,277,177, October 2007.
16. S.P. Koirala, *'Plasma Cavity Ringdown Spectroscopy – A Powerful Technique for Elemental Measurements and Plasma Diagnostics'*, M.S. Thesis, Department of Physics and Astronomy, Mississippi State University, December 2003.
17. C. Wang, S.T. Scherrer, C.B. Winstead, *A Simple Method and Device for Control of Cavity Energy Buildup and Shutoff in CW-Cavity Ringdown Spectroscopy: Application for Ringdown Measurements of Atmospheric CH_4, CO_2, and H_2O at 1.65 μm*, (to be published).
18. G.P. Miller, C.B. Winstead, ''Inductively Coupled Plasma Cavity Ringdown Spectrometry'', *J. Anal. At. Spectrom.*, **12**, 907–912 (1997).
19. C.B. Winstead, F.J. Mazzotti, J. Mierzwa, G.P. Miller, ''Preliminary Results for Electrothermal Atomization-Cavity Ringdown Spectroscopy (ETA-CRDS)'', *Anal. Commun.*, **36**, 277–279 (1999).
20. C. Wang, F.J. Mazzotti, G.P. Miller, C.B. Winstead, 'Cavity Ringdown Spectroscopy for Diagnostic and Analytical Measurements in an Inductively Coupled Plasma', *Appl. Spectrosc.*, **56**, 386–397 (2002).
21. C. Wang, F.J. Mazzotti, G.P. Miller, C.B. Winstead, ''Isotopic Measurements of Uranium Using Inductively Coupled Plasma Cavity Ringdown Spectroscopy'', *Appl. Spectrosc.*, **57**, 1167–1172 (2003).
22. C. Wang, F.J. Mazzotti, J. Mierzwa, G.P. Miller, Y. Duan, C.B. Winstead, "Cavity Ringdown Spectroscopy in Atmospheric Pressure Plasma Applications", *The 53rd Pittsburgh Conference*, New Orleans, 17–22 March 2002.
23. Y. Duan, C. Wang, C.B. Winstead, ''Exploration of Microwave Plasma Source Cavity Ring-Down Spectroscopy for Elemental Measurements'', *Anal. Chem.*, **75**, 2105–2111 (2003).
24. K.L. Bechtel, R.N. Zare, A.A. Kachanov, S.S. Sanders, B.A. Paldus, ''Moving Beyond Traditional UV-Visible Absorption Detection: Cavity Ring-Down Spectroscopy for HPLC'', *Anal. Chem.*, **77**, 1177–1182 (2005).
25. K.L. Snyder, R.N. Zare, ''Cavity Ring-Down Spectroscopy as a Detector for Liquid Chromatography'', *Anal. Chem.*, **75**, 3086–3091 (2003).
26. L. van der Sneppen, A.E. Wiskerke, F. Ariese, C. Gooijer, W. Ubachs, ''Improving the Sensitivity of HPLC Absorption Detection by Cavity Ring-Down Spectroscopy in a Liquid-Only Cavity'', *Anal. Chim. Acta*, **558**, 2–6 (2006).
27. L. van der Sneppen, A.E. Wiskerke, F. Ariese, C. Gooijer, ''Cavity Ring-Down Spectroscopy for Detection in Liquid Chromatography: Extension to Tunable Sources and Ultraviolet Wavelengths'', *Appl. Spectrosc.*, **60**, 931–935 (2006).
28. J.B. Dudek, P.B. Tarsa, A. Velasquez, M. Wladyslawski, P. Rabinowitz, K.K. Lehmann, ''Trace Moisture Detection Using Continuous-Wave Cavity Ring-Down Spectroscopy'', *Anal. Chem.*, **75**, 4599–4605 (2003).
29. R. Vasudev, A. Usachev, W.R. Dunsford, ''Detection of Toxic Compounds by Cavity Ring-Down Spectroscopy'', *Environ. Sci. Technol.*, **33**, 1936–1939 (1999).
30. M.I. Mazurenka, B.L. Fawcett, J.M.F. Elks, D.E. Shallcross, A.J. Orr-Ewing, ''410-nm Diode Laser Cavity Ring-Down Spectroscopy for Trace Detection of NO_2'', *Chem. Phys. Lett.*, **367**, 1–9 (2003).
31. B.L. Fawcett, A.M. Parkes, D.E. Shallcross, A.J. Orr-Ewing, ''Trace Detection of Methane Using Continuous Wave Cavity Ring down Spectroscopy at 1.65μm'', *Phys. Chem. Chem. Phys.*, **4**, 5960–5965 (2002).
32. A.R. Awtry, J.H. Miller, ''Development of a CW -Laser-Based Cavity-Ringdown Sensor Aboard a Spacecraft for Trace Air Constituents'', *Appl. Phys. B*, **75**, 255–260 (2002).
33. Y.A. Bakhirkin, A.A. Kosterev, R. Curl, F.K. Tittel, D.A. Yarekha, L. Hvozdara, M. Giovannini, J. Faist,

"Sub-ppbv Nitric Oxide Concentration Measurements Using CW Thermoelectrically Cooled Quantum Cascade Laser-Based Integrated Cavity Output Spectroscopy", *Appl. Phys. B*, **82**, 149–154 (2006).

34. A.A. Kosterev, A.A. Malinovsky, F.K. Tittel, C. Gmachl, F. Capasso, D.L. Sivco, J.N. Baillargeon, A.Y. Cho "Cavity Ringdown Spectroscopic Detection of Nitric Oxide with a Continuous-Wave Quantum-Cascade Laser", *Appl. Opt.*, **40**, 5522–5529 (2001).
35. A. Schwabedissen, A. Brockhaus, A. Georg, J. Engemann, "Determination of the Gas-Phase Si Atom Density in Radio Frequency Discharges by Means of Cavity Ring-Down Spectroscopy", *J. Phys. D: Appl. Phys.*, **34**, 1116–1121 (2001).
36. A.P. Yalin, V. Surla, M. Butweiller, J.D. Williams, "Detection of Sputtered Metals with Cavity Ring-Down Spectroscopy", *Appl. Opt.*, **44**, 6496–6505 (2005).
37. V. Surla, P.J. Wilbur, J.D. Williams, M. Johnson, A.P. Yalin, "Sputter Erosion Measurements of Titanium and Molybdenum by Cavity Ringdown Spectroscopy", *Rev. Sci. Instrum.*, **75**, 3025–3030 (2004).
38. J.P. Booth, G. Cunge, L. Biennier, D. Romanini, A. Kachanov, "Ultraviolet Cavity Ring-Down Spectroscopy of Free Radicals in Etching Plasmas", *Chem. Phys. Lett.*, **317**, 631–636 (2000).
39. J.B. Wills, J.A. Smith, W.E. Boxford, J.M.F. Elks, M.N.R. Ashfold, A.J. Orr-Ewing, "Measurements of C_2 and CH Concentrations and Temperatures in a DC Arc Jet Using Cavity Ring-Down Spectroscopy", *J. Appl. Phys.*, **92**, 4213–4222 (2002).
40. E.H. Wahl, T.G. Owano, C.H. Kruger, P. Zalicki, Y. Ma, R.N. Zare, "Measurement of Absolute CH_3 Concentration in a Hot-Filament Reactor Using Cavity Ring-Down Spectroscopy", *Diamond Relat. Mater.*, **5**, 373–377 (1996).
41. R. Evertsen, R.L. Stolk, J.J. ter Meulen, "Investigations of Cavity Ring Down Spectroscopy Applied to the Detection of CH in Atmospheric Flames", *Combust. Sci. Technol.*, **157**, 341–342 (2000).
42. P. Zalicki, Y. Ma, R.N. Zare, E.H. Wahl, T.G. Owano, C.H. Kruger, "Measurement of the Methyl Radical Concentration Profile in a Hot-Filament Reactor", *Appl. Phys. Lett.*, **67**, 144–146 (1995).
43. W.M.M. Kessels, A. Leroux, M.G.H. Boogaarts, J.P.M. Hoefhagels, M.C.M. van de Sanden, D.C. Schram, "Cavity Ring Down Detection of SiH_3 in a Remote SiH_4 Plasma and Comparison with Model Calculations and Mass Spectrometry", *J. Vac. Sci. Technol., A*, **19**, 467–476 (2001).
44. J.A. Smith, J.B. Wills, H.S. Moores, A.J. Orr-Ewing, M.N.R. Ashfold, Y.A. Mankelevich, N.V. Suetin, "Effects of NH_3 and N_2 Additions to Hot Filament Activated CH_4/H_2 Gas Mixtures", *J. Appl. Phys.*, **92**, 672–681 (2002).
45. B. Bakowski, G. Hancock, R. Peverall, G.A.D. Ritchie, L.J. Thornton, "Characterization of an Inductively Coupled N_2 Plasma Using Sensitive Diode Laser Spectroscopy", *J. Phys. D: Appl. Phys.*, **37**, 2064–2072 (2004).
46. C. Wang, S.P. Koirala, S.T. Scherrer, Y. Duan, C.B. Winstead, "Diode Laser Microwave Induced Plasma Cavity Ringdown Spectrometer: Performance and Perspective", *Rev. Sci. Instrum.*, **75**, 1305–1313 (2004).
47. C. Wang, S.T. Scherrer, Y. Duan, C.B. Winstead, "Cavity Ringdown Measurements of Mercury and its Hyperfine Structures at 254 nm in an Atmospheric Microwave Plasma: Spectral Interference and Analytical Performance", *J. Anal. At. Spectrom.*, **20**, 638–644 (2005).
48. Y. Duan, C. Wang, S.T. Scherrer, C.B. Winstead, "Development of Alternative Plasma Sources For Cavity Ring-Down Measurements of Mercury", *Anal. Chem.*, **77**, 4883–4889 (2005).
49. J.T. Hodges, J.P. Looney, R.D. van Zee, "Laser Bandwidth Effects in Quantitative Cavity Ring-Down Spectroscopy", *Appl. Opt.*, **35**, 4112–4116 (1996).
50. K.K. Lehmann, D. Romanini, "The Superposition Principle and Cavity Ring-Down Spectroscopy", *J. Chem. Phys.*, **105**, 10263–10277 (1996).
51. J.T. Hodges, J.P. Looney, R.D. van Zee, "Response of a Ring-Down Cavity to an Arbitrary Excitation", *J. Chem. Phys.*, **105**, 10278–10288 (1996).
52. A.E. Siegman, *Lasers*, University Science Books, Mill Valley, 1986.
53. A.P. Yalin, R.N. Zare, C.O. Laux, C.H. Kruger, "Spatial Profiles of N_{2+} Concentration in an Atmospheric Pressure Nitrogen Glow Discharge", *Plasma Sources Sci. Technol.*, **11**, 248–253 (2002).
54. E.R. Crosson, K.N. Ricci, B.A. Richman, F.C. Chilese, T.G. Owano, R.A. Provencal, M.W. Todd, J. Glasser, A.A. Kachanov, B.A. Paldus, T.G. Spence, R.N. Zare, "Stable Isotope Ratios Using Cavity Ring-Down Spectroscopy: Determination of $_{13}C/_{12}C$ for Carbon Dioxide in Human Breath", *Anal. Chem.*, **74**, 2003–2007 (2002).
55. C. Wang, S.T. Scherrer, D. Hossain, "Measurements of Cavity Ringdown Spectroscopy of Acetone in the Ultraviolet and Near-Infrared Spectral Regions: Potential for Development of a Breath Analyzer", *Appl. Spectrosc.*, **58**, 784–791 (2004).
56. M. Mürtz, H. Dahnke, M. Horstjann, D. Kleine, S. Thelen, P. Hering, "Ultra Sensitive Trace Gas Detection for Biomedical Applications", *Spectrochim. Acta, Part A*, **63**, 963–969 (2006).
57. C. Wang, S.T. Scherrer, "Fiber Ringdown Pressure Sensors", *Opt. Lett.*, **29**, 352–354 (2004).

58. C. Wang, S.T. Scherrer, "Fiber Loop Ringdown for Physical Sensor Development: Pressure Sensor", *Appl. Opt.*, **43**, 6458–6464 (2004).

59. P.B. Tarsa, A.D. Wist, P. Rabinowitz, K.K. Lehmann, "Single-Cell Detection by Cavity Ringdown Spectroscopy", *Appl. Phys. Lett.*, **85**, 4523–4525 (2004).

60. Z. Tong, A. Wright, T. McCormick, R. Li, R.D. Oleschuk, H.-P. Loock, "Phase-Shift Fiber-Loop Ring-Down Spectroscopy", *Anal. Chem.*, **76**, 6594–6599 (2004).

61. P.B. Tarsa, P. Rabinowitz, K.K. Lehmann, "Evanescent Field Absorption in a Passive Optical Fiber Resonator Using Continuous-wave Cavity Ring-Down Spectroscopy", *Chem. Phys. Lett.*, **383**, 297–303 (2004).

62. C. Wang, "Fiber Ringdown Temperature Sensors", *Opt. Eng. Lett.*, **44**, 030503-1 (2005).

63. H.-P. Loock, "Ring-Down Absorption Spectroscopy for Analytical Microdevices", *Trends Anal. Chem.*, **25**, 655–664 (2006).

64. C. Wang, A. Mbi, "Optical Superposition in Double Fiber Loop Ringdown", *Proc. SPIE*, **6377**, 6377021–6377028 (2006).

65. C. Wang, A. Mbi, "An Alternative Method to Develop Fibre Grating Temperature Sensors Using the Fibre Loop Ringdown Scheme", *Meas. Sci. Technol.*, **17**, 1741–1751 (2006).

66. J.T. Michael, K.D. Moll, R.J. Jones, B. Safdi, J. Ye, "Broadband Cavity Ringdown Spectroscopy for Sensitive and Rapid Molecular Detection", *Science*, **311**, 1595–1599 (2006).

67. M.J. Thorpe, D.D. Hudson, K.D. Moll, J. Lasri, J. Ye, "Cavity-Ringdown Molecular Spectroscopy Based on an Optical Frequency Comb at 1.45-1.65 μm", *Opt. Lett.*, **32**, 307–309 (2007).

68. K.W. Busch, M.A. Busch, eds, *Cavity-Ringdown Spectroscopy: An Ultratrace-Absorption Measurement Technique*, *ACS Symposium Series 720*, Oxford University Press, Oxford University, 1999.

69. G. Berden, R. Peeters, G. Meijer, "Cavity Ring-Down Spectroscopy: Experimental. Schemes and Applications", *Int. Rev. Phys. Chem.*, **19**, 565–607 (2000).

70. J.J. Scherer, J.B. Paul, A. O'Keefe, R.J. Saykally, "Cavity Ringdown Laser Absorption Spectroscopy: History, Development, and Application to Pulsed Molecular Beams", *Chem. Rev.*, **97**, 25–51 (1997).

71. G.P. Miller, C.B. Winstead, 'Cavity Ringdown Laser Absorption Spectroscopy', in *Encyclopedia of Analytical Chemistry: Instrumentation and Applications*, ed. R.A. Meyers, John Wiley & Sons, Chichester, 10734, 2000.

72. D.R. Wagner, G.P. Miller, C.B. Winstead, 'Infrared Cavity Ringdown Spectroscopy', in *Handbook of Vibrational Spectroscopy*, eds. J.M. Chalmers, P.R. Griffiths, John Wiley & Sons, Chichester, 2002.

73. D.B. Atkinson, "Solving Chemical Problems of Environmental Importance Using Cavity Ring-Down Spectroscopy", *Analyst*, **128**, 117–125 (2003).

74. C. Vallance, "Innovations in Cavity Ringdown Spectroscopy", *New J. Chem.*, **97**, 867–874 (2005).

75. B.A. Paldus, A.A. Kachanov, "An Historical Overview of Cavity-Enhanced Methods", *Can. J. Phys.*, **83**, 975–999 (2005).

76. M. Mazurenka, A.J. Orr-Ewing, R. Peverall, G.A.D. Ritchie, "Cavity Ring-Down and Cavity Enhanced Spectroscopy Using Diode Lasers", *Annu. Rep. Prog. Chem. Sect. C: Phys. Chem.*, **101**, 100–142 (2005).

77. C. Wang, "Plasma-Cavity Ringdown Spectroscopy (P-CRDS) for Elemental and Isotopic Measurements", *J. Anal. At. Spectrom.*, **22**, 1347–1363 (2007).

78. P. Zalicki, R.N. Zare, "Cavity Ring-Down Spectroscopy for Quantitative Absorption Measurements", *J. Chem. Phys.*, **102**, 2708–2717 (1995).

79. R.D. van Zee, J.T. Hodges, J.P. Looney, "Pulsed, Single-Mode Cavity Ringdown Spectroscopy", *Appl. Opt.*, **38**, 3951 (1999).

80. D. Romanini, K.K. Lehmann, "Cavity Ringdown Overtone Spectroscopy of HCN, $H_{13}CN$ and $HC_{15}N$", *Chem. Phys.*, **102**, 633–642 (1995).

81. D. Romanini, A.A. Kachanov, N. Sadeghi, F. Stoeckel, "CW Cavity Ring Down Spectroscopy", *Chem. Phys. Lett.*, **264**, 316–322 (1997).

82. J.J. Scherer, D. Voelkel, D.J. Rakestraw, J.B. Paul, C.P. Collier, R.J. Saykally, A. O'Keefe, "Infrared Cavity Ring-Down Laser Absorption Spectroscopy (IR-CRLAS)", *Chem. Phys. Lett.*, **245**, 273–280 (1995).

83. J. Martin, B.A. Paldus, P. Zalicki, E.H. Wahl, T.G. Owano, J.S. Harris Jr, C.H. Kruger, R.N. Zare, "Cavity Ring-Down Spectroscopy with Fourier-Transform-Limited Light Pulses", *Chem.Phys. Lett.*, **258**, 63–70 (1996).

84. O. Svelto, *Principles of Lasers*, Plenum Press, New York, 1989.

85. D.C. O'Shea, W.R. Callen, W.T. Rhodes, *Introduction to Lasers and Their Applications*, Addison-Wesley, Reading, 1978.

86. G. Meijer, M.G.H. Boogaarts, R.T. Jongma, D.H. Parker, A.M. Wodtke, "Coherent Cavity Ring Down Spectroscopy", *Chem. Phys. Lett.*, **217**, 112–116 (1994).

87. A. O'Keefe, O. Lee, "Trace Gas Analysis by Pulsed Laser Absorption Spectroscopy", *Am. Lab.*, **21**, 19 (1989).

88. J.P. Koplow, D.A.V. Kliner, L. Goldberg, "Development of a Narrow-Band, Tunable, Frequency-Quadrupled Diode Laser for UV Absorption Spectroscopy", *Appl. Opt.*, **37**, 3954–3960 (1998).

89. J. Pearson, A.J. Orr-Ewing, M.N.R. Ashfold, R.N. Dixon, "Spectroscopy and Predissociation Dynamics of the A^1A'' state of HNO", *J. Chem. Phys.*, **106**, 5850–5873 (1997).

90. B.A. Richman, A.A. Kachanov, B.A. Paldus, "Novel Detection of Aerosols: Combined Cavity Ring-Down and Fluorescence Spectroscopy", *Opt. Express*, **13**, 3376–3387 (2005).

91. G. Stewart, K. Atherton, H. Yu, B. Culshaw, "Investigation of an Optical Fibre Amplifier Loop for Intracavity and Cavity Ring-Down Loss Measurements", *Meas. Sci. Technol.*, **12**, 843–849 (2001).

92. G. Berden, R. Engeln, P.C.M. Christianen, J.C. Maan, G. Meijer, "Cavity Ring Down Spectroscopy on the Oxygen A Band in Magnetic Fields up to 20 Tesla", *Phys. Rev. A*, **58**, 3114–3123 (1998).

93. Y. Le Grand, A. Le Floch, "Sensitive Dichroism Measurements Using Eigenstate Decay Times", *Appl. Opt.*, **29**, 1244–1246 (1990).

94. G. Stewart, P. Shields, B. Culshaw, "Development of Fibre Laser Systems for Ring-Down and Intracavity Gas Spectroscopy in the Near-IR", *Meas. Sci. Technol.*, **15**, 1621–1628 (2004).

95. J.B. Paul, C.P. Collier, R.J. Saykally, J.J. Scherer, A. O'Keefe, "Direct Measurement of Water Cluster Concentrations by Infrared Cavity Ring-Down Laser Absorption Spectroscopy", *J. Phys. Chem.*, **101**, 5211–5214 (1997).

96. R. Engeln, G. Meijer, "A Fourier Transform Cavity Ring Down Spectrometer", *Rev. Sci. Instrum.*, **67**, 2708–2713 (1996).

97. J.J. Scherer, "Ringdown Spectral Photography", *Chem. Phys. Lett.*, **292**, 143–153 (1998).

98. S.E. Fiedler, A. Hese, A.R. Ruth, "Incoherent Broad-Band Cavity-Enhanced Absorption Spectroscopy of Liquids", *Rev. Sci. Instrum.*, **76**, 23107–23113 (2005).

99. R. Engeln, E. van den Berg, G. Meijer, L. Lin, G.M.H. Knippels, A.F.G. van der Meer, "Cavity Ring down Spectroscopy with a Free-Electron Laser", *Chem. Phys. Lett.*, **269**, 293–297 (1997).

100. R. Engeln, G. Von Helden, A.J.A. van Roij, G. Meijer, "Cavity Ring down Spectroscopy on Solid C_{60}", *J. Chem. Phys.*, **110**, 2732–2733 (1999).

101. R.T. Jongma, M.G.H. Boogaarts, I. Holleman, G. Meijer, "Trace Gas Detection with Cavity Ring down Spectroscopy", *Rev. Sci. Instrum.*, **66**, 2821–2828 (1995).

102. J.T. Hodges, J.P. Looney, R.D. van Zee, "Far-Field Scattering of an Axisymmetric Laser Beam of Arbitrary Profile by an On-Axis Spherical Particle", *Appl. Opt.*, **35**, 4283–4290 (1996).

103. A.P. Yalin, R.N. Zare, "Effect of Laser Lineshape on The Quantitative Analysis of Cavity Ring-Down Signals", *Laser Phys.*, **12**, 1065–1072 (2002).

104. A.P. Yalin, R.N. Zare, C.O. Laux, C.H. Kruger, "Temporally Resolved Cavity Ring-Down Spectroscopy in a Pulsed Nitrogen Plasma", *Appl. Phys. Lett.*, **81**, 1408–1410 (2002).

105. R. Engeln, G. Helden, G. Berden, G. Meijer, "Phase Shift Cavity Ring Down Absorpion Spectroscopy", *Chem. Phys. Lett.*, **262**, 105–109 (1996).

106. Y. He, M. Hippler, M. Quack, "High-Resolution Cavity Ring-Down Absorption Spectroscopy of Nitrous Oxide and Chloroform Using a Near-Infrared CW Diode Laser", *Chem. Phys. Lett.*, **289**, 527–534 (1998).

107. D. Romanini, P. Dupre, R. Jost, "Non-Linear Effects by Continuous Wave Cavity Ringdown Spectroscopy in Jet-Cooled NO_2", *Vib. Spectrosc.*, **19**, 93–106 (1999).

108. B.A. Paldus, J.S. Harris Jr, J. Martin, X. Kie, R.N. Zare, "Laser Diode Cavity Ring-Down Spectroscopy Using Acousto-Optic Modulator Stabilization", *J. Appl. Phys.*, **82**, 3199–3204 (1997).

109. B.A. Paldus, C.C. Harb, T.G. Spence, B. Wilke, J. Xie, J.S. Harris, R.N. Zare, "Cavity-Locked Ring-Down Spectroscopy", *J. Appl. Phys.*, **83**, 3991–3997 (1998).

110. M.D. Levenson, B.A. Paldus, T.G. Spence, C.C. Harb, J.S. Harris Jr, R.N. Zare, "Optical Heterodyne Detection in Cavity Ring-Down Spectroscopy", *Chem. Phys. Lett.*, **290**, 335–340 (1998).

111. K.J. Schulz, W.R. Simpson, "Frequency-Matched Cavity Ring-Down Spectroscopy", *Chem. Phys. Lett.*, **297**, 523–529 (1998).

112. J.W. Hahn, Y.S. Yoo, J.Y. Lee, J.W. Kim, H.W. Lee, "Cavity Ringdown Spectroscopy with a Continuous-Wave Laser: Calculation of Coupling Efficiency and a new Spectrometer Design", *Appl. Opt.*, **38**, 1859–1866 (1999).

113. P. Cias, C. Wang, T.S. Dibble, "Absorption Cross-Sections of the C-H Overtone of Volatile Organic Compounds: 2 methyl-1,3-butadiene (Isoprene), 1,3-butadiene, and 2,3-dimethyl-1,3-butadiene", *Appl. Spectrosc.*, **61**, 230–236 (2007).

114. C. Wang, N. Srivastava, J. Cambre, B.A. Jones, R.B. Reese, 'Development of a Portable Ringdown Spectrometer for Greenhouse Gases and Carbon Isotope', *The 34th FACSS Annual Meeting*, Memphis, 14–18 October 2007.

115. D. Romanini, K.K. Lehmann, "Ring-Down Cavity Absorption Spectroscopy of the Very Weak HCN Overtone Bands with Six, Seven, and Eight Stretching Quanta", *J. Chem. Phys.*, **99**, 6287–6301 (1993).

116. H. Naus, I.H.M. van Stokkum, W. Hogervorst, W. Ubachs, "Quantitative Analysis of Decay Transients Applied to a Multimode Pulsed Cavity Ringdown Experiment", *Appl. Opt.*, **40**, 4416–4426 (2001).

117. M.D. Wheeler, S.M. Newman, A.J. Orr-Ewing, M.N.R. Ashfold, "Cavity Ring-Down Spectroscopy", *J. Chem. Soc., Faraday Trans.*, **94**, 337–351 (1998).

118. J.B. Paul, R.J. Saykally, "Cavity Ringdown Laser Absorption Spectroscopy", *Anal. Chem. News Features*, **69**, 287A–292A (1997).

119. R.A. Provencal, J.B. Paul, C.N. Chapo, R.J. Saykally, "Cavity Ringdown Laser Absorption Spectroscopy", *Spectroscopy*, **14**, 24–32 (1999).

120. A. O'Keefe, J.J. Scherer, A.L. Cooksy, R. Sheeks, J. Heath, R.J. Saykally, "Cavity Ring Down Dye Laser Spectroscopy of Jet-Cooled Metal Clusters: Cu_2 Cu_3", *Chem. Phys. Lett.*, **172**, 214–218 (1990).

121. J.J. Scherer, J.B. Paul, C.P. Collier, R.J. Saykally, "Cavity Ringdown Laser Absorption Spectroscopy and Time-Of-Flight Mass Spectroscopy of Jet-Cooled Copper Silicides", *J. Chem. Phys.*, **102**, 5190–5199 (1995).

122. R.A. Provencal, J.B. Paul, K. Roth, C. Chapo, R.N. Casaes, R.J. Saykally, G.S. Tschumper, H.F. Schaefer III, "Infrared Cavity Ringdown Spectroscopy of Methanol Clusters: Single Donor Hydrogen Bonding", *J. Chem. Phys.*, **110**, 4258–4267 (1999).

123. J.J.L. Spaanjaars, J.J. ter Meulen, G. Meijer, "Relative Predissociation Rates of OH $(A^2\sum^+ .v'' = 3)$ from Combined Cavity Ringdown–Laser Induced Fluorescence Measurements", *J. Chem. Phys.*, **107**, 2242–2248 (1997).

124. J. Luque, J.B. Jeffries, G.P. Smith, D.R. Crosley, "Predissociation of CH $B^2\sum^+ v' = 0,1$ Levels Studied by Cavity Ring-Down Absorption Spectroscopy", *Chem. Phys. Lett.*, **346**, 209–216 (2001).

125. M.D. Wheeler, A.J. Orr-Ewing, M.N.R. Ashfold, T. Ishiwata, "Predissociation Lifetimes of the $A^2\sum^+ V = 1$ State of the SH Radical Determined by Cavity Ring-Down Spectroscopy", *Chem. Phys. Lett.*, **268**, 421–428 (1997).

126. V.A. Lozovsky, I. Derzy, S. Cheskis, "Nonequilibrium Concentrations of the Vibrationally Excited OH Radical in a Methane Flame Measured by Cavity Ring-Down Spectroscopy", *Chem. Phys. Lett.*, **284**, 407–411 (1998).

127. S. Cheskis, I. Derzy, V.A. Lozovsky, A. Kachanov, D. Romanini, "Cavity Ring-Down Spectroscopy of OH Radicals in Low Pressure Flame", *Appl. Phys. B*, **66**, 377–381 (1998).

128. M.B. Pushkarsky, S.J. Zalyubovsky, T.A. Miller, "Detection and Characterization of Alkyl Peroxy Radicals Using Cavity Ringdown Spectroscopy", *J. Chem. Phys.*, **112**, 10695–10698 (2000).

129. S.J. Zalyubovsky, D. Wang, T.A. Miller, "Observation of the $\tilde{A}-\tilde{X}$ Electronic Transition of the CF^3O^2 Radical", *Chem. Phys. Lett.*, **335**, 298–304 (2001).

130. S.J. Zalyubovsky, B.G. Glover, T.A. Miller, "Rotationally Resolved Electronic Spectra of the $\tilde{B}-\tilde{X}$ Transition in Multiple Conformers of 1-Butoxy and 1-Pentoxy Radicals", *J. Phys. Chem.*, **107**, 7704–7712 (2003).

131. S.J. Zalyubovsky, B.G. Glover, T.A. Miller, C. Hayes, J.K. Merle, C.M. Hadad, "Near IR Cavity Ringdown Spectroscopy and Kinetics of the Isomers and Conformers of the Butyl Peroxy Radical", *J. Phys. Chem.*, **109**, 1308–1315 (2005).

132. M.P.G. Just, E.N. Sharp, S.J. Zalyubovsky, T.A. Miller, "Cavity Ringdown Spectroscopy of the $\tilde{A}-\tilde{X}$ Electronic Transition of the Phenyl Peroxy Radical", *Chem. Phys. Lett.*, **417**, 378–382 (2006).

133. P. Rupper, E.N. Sharp, G. Tarczay, T.A. Miller, "Investigation of Ethyl Peroxy Radical Conformers via Cavity Ringdown Spectroscopy of the $\tilde{A}-\tilde{X}$ Electronic Transition", *J. Phys. Chem. A*, **111**, 832–840 (2007).

134. T. Yu, M.C. Lin, "Kinetics of Phenyl Reactions Studied by the Cavity Ring-Down Method", *J. Am. Chem. Soc.*, **115**, 4371–4372 (1993).

135. T. Yu, M.C. Lin, "Kinetics of the $C_6H^5 + CCl_4$ Reaction in the Gas Phase: Comparison with Liquid-Phase Data", *J. Phys. Chem.*, **98**, 9697–9699 (1994).

136. J. Park, D. Chakraborty, D.M. Bhusari, M.C. Lin, "Kinetics of C_6H_5 Radical Reactions with Toluene and Xylenes by Cavity Ringdown Spectrometry", *J. Phys. Chem.*, **103**, 4002–4008 (1999).

137. D.B. Atkinson, J.W. Hudgens, "Rate Coefficients for the Propargyl Radical Self-Reaction and Oxygen Addition Reaction Measured Using Ultraviolet Cavity Ring-Down Spectroscopy", *J. Phys. Chem.*, **103**, 4242–4252 (1999).

138. J. Thiebaud, C. Fittschen, "Near Infrared CW-CRDS Coupled to Laser Photolysis: Spectroscopy and Kinetics of the HO_2 Radical", *Appl. Phys. B*, **85**, 383–389 (2006).

139. J.J. Scherer, K.W. Aniolke, N.P. Cernansky, D.J. Rakestraw, "Determination of Methyl Radical Concentrations in a Methane/Air Flame by Infrared Cavity Ringdown Laser Absorption Spectroscopy", *J. Chem. Phys.*, **107**, 6196–6203 (1997).

140. J.J. Scherer, D.J. Rakestraw, "Cavity Ringdown Laser Absorption Spectroscopy Detection of Formyl (HCO) Radical in a Low Pressure Flame", *Chem. Phys. Lett.*, **265**, 169–176 (1997).

141. A. McIlroy, "Direct Measurement of CH_2 in Flames by Cavity Ringdown Laser Absorption Spectroscopy", *Chem. Phys. Lett.*, **296**, 151–158 (1998).

142. X. Mercier, P. Jamette, J.F. Pauwels, P. Desgroux, "Absolute CH Concentration Measurements by Cavity Ring-Down Spectroscopy in an Atomospheric Diffusion Flame", *Chem. Phys. Lett.*, **305**, 334–342 (1999).

143. X. Mercier, E. Therssen, J.F. Pauwels, P. Desgroux, "Cavity Ringdown Measurements of OH Radical in Atmospheric Premixed and Diffusion Flames. A Comparison with Laser Induced Fluorescence and Direct Laser Absorption", *Chem. Phys. Lett.*, **299**, 75–83 (1999).

144. P. Zalicki, Y. Ma, R.N. Zare, E.H. Wahl, J.R. Dadamio, T.G. Owano, C.H. Kruger, "Methyl Radical Measurement by Cavity Ring-Down Spectroscopy", *Chem. Phys. Lett.*, **234**, 269–274 (1995).

145. A. Schocker, K.K. Hoinghaus, A. Brockhinke, "Quantitative Determination of Combustion Intermediates with Cavity Ring-Down Spectroscopy: Systematic Study in Propene Flames Near the Soot-Formation Limit", *Appl. Opt.*, **44**, 6660–6672 (2005).

146. A. O'Keefe, unpublished research.

147. M. Kotterer, J. Conceicao, J.P. Maier, "Cavity Ringdown Spectroscopy of Molecular Ions: $A\ ^2\prod_u \leftarrow X^2\sum_g^+$(6-0) Transition of N_2^+", *Chem. Phys. Lett.*, **259**, 233–236 (1996).

148. E. Quandt, I. Kraemer, H.F. Dobele, "Measurements of Negative-Ion Densities by Cavity Ringdown Spectroscopy", *Europhys. Lett.*, **45**, 32–37 (1998).

149. E. Quandt, H.F. Dobele, W.G. Graham, "Measurements of Negative Ion Densities by Absorption Spectroscopy", *Appl. Phys. Lett.*, **72**, 2394–2396 (1998).

150. D. Romanini, L. Biennier, F. Salama, A. Kachanov, L.J. Allamandola, F. Stoeckel, "Jet-Discharge Cavity Ring-Down Spectroscopy of Ionized Polycyclic Aromatic Hydrocarbons: Progress in Testing the PAH Hypothesis for the Diffuse Interstellar Band Problem", *Chem. Phys. Lett.*, **303**, 165–170 (1999).

151. C. Wang, F.J. Mazzotti, S.P. Koirala, C.B. Winstead, G.P. Miller, "Measurements of OH Radicals in a Low-Power Atmospheric Inductively Coupled Plasma by Cavity Ringdown Spectroscopy", *Appl. Spectrosc.*, **58**, 734–740 (2004).

152. P.J. van den Oever, J.L. van Hemmen, J.H. Helden, D.C. Schram, R. Engeln, M.C.M. van de Sanden, W.M.M. Kessels, "Downstream Ion and Radical Densities in an Ar-NH_3 Plasma Generated by the Expanding Thermal Plasma Technique", *Plasma Sources Sci. Technol.*, **15**, 546–555 (2006).

153. A. Campargue, D. Romanini, N. Sadeghi, "Measurement of SiH_2 Density in a Discharge by Intracavity Laser Absorption Spectroscopy and CW Cavity Ring-Down Spectroscopy", *J. Phys. D: Appl. Phys.*, **31**, 1168–1175 (1998).

154. J.P. Booth, G. Cunge, L. Bienneir, D. Romanini, A. Katchanov, 'Absolute Concentrations of Reactive Species in Etching Plasmas by High-Sensitivity UV Absorption Spectroscopy', *International Conference on Phenomena in Ionised Gases XXIV*, Warsaw, 1999.

155. H.V. Eckert, *Physics and Technology of Low-Temperature Plasma*, English edition, The Iowa State University Press, Ames, 1977.

156. C.O. Laux, T.G. Spence, C.H. Kruger, R.N. Zare, "Optical Diagnostics of Atmospheric Pressure Air Plasmas", *Plasma Sources Sci. Technol.*, **12**, 125–138 (2003).

157. R.G. Breene, *The Shift and Shape of Spectral Lines*, Pergamon Press, Oxford, 1961.

158. B.N. Ganguly, W.R. Lempert, K. Akhtar, J.E. Scharer, F. Leipold, C.O. Laux, R.N. Zare, A.P. Yalin, 'Plasma Diagnostics', in *Non-Equilibrium Air Plasmas at Atmospheric Pressure*, ed. K.H. Becker, Institute of Physics Publishing, Bristol, 446–536, 2005.

159. J.E. Polk, 'An Overview of the Results from an 8200 Hour Wear Test of the NSTAR Ion Thruster', paper AIAA-99-2446, presented at the *Joint Propulsion Conference*, Los Angeles, June 1999.

160. M.J. Pellin, R.B. Wright, D.M. Gruen, "Laser Fluorescence Spectroscopy of Sputtered Zirconium Atoms", *J. Chem. Phys.*, **74**, 6448–6457 (1981).

161. N. Andersen, B. Andersen, E. Veje, "Atomic Excitations in Sputtering Processes", *Radiat. Eff.*, **60**, 119–127 (1982).

162. R.P. Doerner, D.G. Whyte, D.M. Goebel, "Sputtering Yield Measurements During Low Energy Xenon Plasma Bombardment", *J. Appl. Phys.*, **93**, 5816–5823 (2003).

163. E. Dullni, "Laser Fluorescence Measurements of the Flux Density of Titanium Sputtered from an Oxygen Covered Surface", *Appl. Phys. A*, **38**, 131–138 (1985).

164. R.B. Wright, M.J. Pellin, D.M. Gruen, C.E. Young, "Laser Fluorescence Spectroscopy of Sputtered Uranium Atoms", *Nucl. Instrum. Methods*, **170**, 295–302 (1980).

165. G. Nicolussi, W. Husinsky, D. Gruber, G. Betz, "Calculations of the Geometry and Optical Properties of F_{Mg} Centers and Dimer (F_2-type) Centers in Corundum Crystals", *Phys. Rev. B*, **51**, 8770–8788 (1995).

166. C. Staudt, A. Wucher, "Sputtering of Ag Atoms into Metastable Excited States", *Phys. Rev. B*, **66**, 085415-1-085415-12 (2002).

167. S.S. Brown, H. Stark, A.R. Ravishankara, "Cavity Ring-Down Spectroscopy for Atmospheric Trace Gas Detection: Application to the Nitrate Radical (NO3)", *Appl. Phys. B*, **75**, 173–182 (2002).

168. S.M. Ball, R. Jones, "Broad-Band Cavity Ring-Down Spectroscopy", *Chem. Rev.*, **103**, 5239–5262 (2003).

169. M. Mazurenka, R. Wada, A.J.L. Shillings, T.J.A. Butler, J.M. Beames, A.J. Orr-Ewing, "Fast Fourier Transform Analysis in Cavity Ring-Down Spectroscopy: Application to an Optical Detector for Atmospheric NO_2", *Appl. Phys. B*, **81**, 135–141 (2005).

170. W.R. Simpson, "Continuous Wave Cavity Ring-Down Spectroscopy Applied to In Situ Detection of Dinitrogen Pentoxide (N_2O_5)", *Rev. Sci. Instrum.*, **74**, 3442–3452 (2003).

171. A.M. Parkes, B.L. Fawcett, R.E. Austin, S. Nakamichi, D.E. Shallcross, A.J. Orr-Ewing, "Trace Detection of Volatile Organic Compounds by Diode Laser Cavity Ring-Down Spectroscopy", *Analyst*, **128**, 960–965 (2003).

172. A.M. Parkes, R.E. Lindley, A.J. Orr-Ewing, "Combining Pre-Concentration of Air Samples with Cavity Ring-Down Spectroscopy for Detection of Trace Volatile Organic Compounds in the Atmosphere", *Anal. Chem.*, **76**, 7329–7335 (2004).

173. C.B. Winstead, S. Scherrer, S.C. Foster, C. Wang, 'Near-Infrared Cavity Ringdown Measurement of C-H Stretching Overtones in Selected Volatile Organic Compounds', *232nd ACS National Meeting*, San Francisco, 10–14 September 2006.

174. S.T. Scherrer, C. Wang, C.B. Winstead, 'Near Infrared Measurements of Volatile Organic Compounds Using Diode Laser cavity Ringdown Spectroscopy', *60th International Symposium on Molecular Spectroscopy*, Columbus, 20–24 June 2005.

175. G.P. Miller, *Pilot-Scale Test Report of a Cavity Ringdown Spectrometer*, Final Report, Sensor Research and Development Corporation, Orono, 2004.

176. D.S. Baer, J.B. Paul, M. Gupta, A. O'Keefe, "Sensitive Absorption Measurements in the Near-Infrared Region Using Off-Axis Integrated-Cavity-Output Spectroscopy", *Appl. Phys. B*, **75**, 261–265 (2002).

177. A. Castrillo, G. Casa, E. Kerstel, L. Gianfrani, "Diode Laser Absorption Spectrometry for $^{13}CO_2/^{12}CO_2$ Isotope Ratio Analysis : Investigation on Precision and Accuracy Levels", *Appl. Phys. B*, **81**, 863–869 (2005).

178. Y. Ding, A. Campargue, E. Bertseva, S. Tashkun, V.I. Perevalov, "Highly Sensitive Absorption Spectroscopy of Carbon Dioxide by ICLAS-VeCSEL between 8800 and 9530 cm^{-1}", *J. Mol. Spectrosc.*, **231**, 117–123 (2005).

179. F.K. Tittel, D. Richter, A. Fried, "Mid-Infrared Laser Applications in Spectroscopy", *Top. Appl. Phys. (Solid-State Mid-Infrared Laser Sources)*, **89**, 445–510 (2003).

180. A.D. Sappey, E.S. Hill, T. Settersten, M.A. Linne, "Fixed-Frequency Cavity Ringdown Diagnostic for Atmospheric Particulate Matter", *Opt. Lett.*, **23**, 954–956 (1998).

181. J.D. Smith, D.B. Atkinson, "A Portable Pulsed Cavity Ring-Down Transmissometer for Measurement of the Optical Extinction of the Atmospheric Aerosol", *Analyst*, **126**, 1216–1220 (2001).

182. A.W. Strawa, T. Owano, R. Castaneda, D.S. Baer, B.A. Paldus, "The Measurement of Aerosol Optical Properties Using Continuous Wave Cavity Ring-Down Techniques", *J. Atmos. Oceanic Technol.*, **20**, 454–465 (2003).

183. J.E. Thompson, H.D. Nasajpour, B.W. Smith, J.D. Winefordner, "Atmospheric Aerosol Measurements by Cavity Ringdown Turbidimetry", *Aerosol Sci. Technol.*, **37**, 221–230 (2003).

184. T.J.A. Butler, J.L. Miller, A.J. Orr-Ewing, "Cavity Ring-Down Spectroscopy Measurements of Single Aerosol Particle Extinction. I. The Effect of Position of a Particle within the Laser Beam on Extinction", *J. Chem. Phys.*, **126**, 174302–174309 (2007).

185. J.L. Miller, A.J. Orr-Ewing, "Cavity Ring-Down Spectroscopy Measurement of Single Aerosol Particle Extinction. II. Extinction of Light by an Aerosol Particle in an Optical Cavity Excited by a CW Laser", *J. Chem. Phys.*, **126**, 174303–174310 (2007).

186. G.F. Kirkbright, M. Sargent, *Atomic Absorption and Fluorescence Spectroscopy*, Academic Press, London, 1974.

187. D.S. Baer, R.K. Hanson, "Tunable Diode Laser Absorption Diagnostics for Atmospheric Pressure Plasmas", *J. Quant. Spectrosc. Radiat. Transfer*, **47**, 455–475 (1992).

188. J.A.M. van der Mullen, "On the Atomic State Distribution Function in Inductively Coupled Plasmas-I. Thermodynamic Equilibrium Considered on the Elementary Level", *Spectrochim. Acta, Part B*, **44**, 1067–1080 (1989).

189. P.Y. Yang, R.M. Barnes, "Modification of Abel Inversion Technique for Line Width Calculation", *Spectrochim. Acta, Part B*, **44**, 561–570 (1989).

190. Y. Duan, Y. Su, Z. Jin, S.P. Abeln, "Design and Development of a Highly Sensitive, Field Portable Plasma Source Instrument for On-Line Liquid Stream Monitoring and Real-Time Sample Analysis", *Rev. Sci. Instrum.*, **71**, 1557–1563 (2000).

191. Y. Duan, Y. Su, Z. Jin, S.P. Abeln, "A Field-Portable Plasma Source Monitor for Real-Time Air Particulate Monitoring", *Anal. Chem.*, **72**, 1672–1679 (2000).

192. Y. Duan, X. Du, Y. Li, Q. Jin, "Characterization of a Modified, Low-Power Argon Microwave Plasma Torch (MPT) as an Atomization Cell for Atomic Fluorescence Spectrometry", *Appl. Spectrosc.*, **49**, 1079–1085 (1995).

193. P.I. Presunka, J.A. Coxon, "Laser Spectroscopy of the $\tilde{A}^2\prod - \tilde{X}^2\sum^+$ Transition of SrOH: Deperturbation Analysis of K-Resonance in the $\nu_2 = 1$ Level of the $\tilde{A}^2\prod$ State", *J. Chem. Phys.*, **101**, 201–222 (1994).

194. M.S. Beardah, A.M. Ellis, "Gas Phase Dynamics and Structure: Spectroscopy, Molecular Interactions, Scattering, and Photochemistry–Observation of Several New Electronic Transitions of the SrOH Free Radical", *J. Chem. Phys.*, **110**, 11244–11254 (1999).

195. R. Mavrodineanu, H. Boiteux, *Flame Spectroscopy*, John Wiley & Sons, New York, London, Sydney, 380, 1965.

196. J. Blaise, L.J. Radziemski, "Energy Levels of Neutral Atomic Uranium (U_I)", *J. Opt. Soc. Am.*, **66**, 644–659 (1976).

197. B.W. Smith, A. Quentmeier, M. Bolshov, K. Niemax, "Measurement of Uranium Isotope Ratios in Solid Samples Using Laser Ablation and Diode Laser-Excited Atomic Fluorescence Spectrometry", *Spectrochim. Acta, Part B*, **54**, 943–958 (1999).

198. V. Kaufman, ''Wavelengths, Energy Levels, and Pressure Shifts. in Mercury-198'', *J. Opt. Soc. Am.*, **52**, 866–870 (1962).

199. D. Foucher and H. Hintelmann, ''High-Precision Measurement of Mercury Isotope Ratios in Sediments Using Cold-Vapor Generation Multi-Collector Inductively Coupled Plasma Mass Spectrometry'', *Anal. Bioanalytical Chem.* **384**, 1470–1478 (2006).

200. Y. Nishimura, T. Fujimoto, ''$\lambda = 2537$ ÅLine from a Low-Pressure Mercury Discharge Lamp Emission Profile and Line Absorption by a Gas Containing a Mercury Vapor'', *Appl. Phys. B*, **38**, 91–98 (1985).

201. F.J. Mazzotti, *'Investigation of Inductively Coupled Plasma as an Atomization Source for Analytical and Fundamental Measurements Using Cavity Ringdown Spectroscopy'*, PhD Dissertation, Mississippi State University, December 2005.

202. Varian GTA-100 Graphite Furnace System, *Manufacturer Specifications*, Varian, Palo Alto, 1997.

203. G.P. Miller, *A Cavity Ringdown Spectroscopy Mercury Continuous Emission Monitor*, Final Technical Report, DE-FC26-01FT41221, Sensor research and Development Corporation, Orono, 2004.

204. L. Pauling, A.B. Robinson, R. Teranishi, P. Cary, ''Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography'', *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 2374–2376 (1971).

205. S.M. Gordon, J.P. Szidon, B.K. Krotoszynski, R.D. Gibbons, H.J. O'Neill, ''Volatile Organic Compounds in Exhaled Air from Patients with Lung Cancer'', *Clin. Chem.*, **31**, 1278–1282 (1985).

206. T. Karl, P. Prazeller, D. Mayr, A. Jordan, J. Rieder, R. Fall, W. Lindinger, ''Human Breath Isoprene and its Relation to Blood Cholesterol Levels: New Measurements And Modeling'', *J. Appl. Physiol.*, **91**, 762–770 (2001).

207. A. Manolis, ''The Diagnostic Potential of Breath Analysis'', *Clin. Chem.*, **29**, 5–15 (1983).

208. K. Musa-Veloso, E. Rarama, F. Comeau, R. Curtis, S. Cunnane, ''Ketonemia and Seizures: Metabolic and Anticonvulsant Effects of Two Ketogenic Diets in Childhood Epilepsy'', *Pediatr. Res.*, **52**, 443–448 (2002).

209. D.P. Joachim, B. Andrew, ''Lindstrom Exhaled Human Breath Measurement Method for Assessing Exposure to Halogenated Volatile Organic Compounds'', *Clin. Chem.*, **43**, 723–730 (1997).

210. J.H. Raymer, K.W. Thomas, S.D. Cooper, D.A. Whitaker, E.D. Pellizzari, ''A Device for Sampling of Human Alveolar Breath for the Measurement of Expired Volatile Organic Compounds'', *J. Anal. Toxicol.*, **14**, 6337–6344 (2001).

211. S.-K. Rossana, D. Kevin, ''Cashman Potential Applications of Breath Isoprene as a Biomarker in Modern Medicine: a Concise Overview'', *Wien. Klin. Wochenschr.*, **117**, 180–186 (2005).

212. M. Phillips, R.N. Cataneo, J. Greenberg, M.I. Munawar, S. Nachnani, S. Samtani, ''Pilot Study of a Breath Test for Volatile Organic Compounds Associated with Oral Malodor: Evidence for the Role of Oxidative Stress'', *Oral Dis.*, **11**, 32–34 (2005).

213. C. Deng, J. Zhang, X. Yu, W. Zhang, X. Zhang, ''Determination of Acetone in Human Breath by Gas Chromatography-Mass Spectrometry and Solid-Phase Microextraction with On-Fiber Derivatization'', *J. Chromatogr.*, **810**, 269–275 (2004).

214. C. Wang, 'Biomedical Applications of Cavity Ringdown Spectroscopy: Present and Perspective Abstracts', *57th Southeast/61st Southwest Joint Regional Meeting of the American Chemical Society*, Memphis, 1–4 November 2005.

215. H. Dahnke, D. Kleine, P. Hering, M. Murtz, ''Real-Time Monitoring of Ethane in Human Breath Using Mid-Infrared Cavity Leak-Out Spectroscopy'', *Appl. Phys. B*, **72**, 971–975 (2001).

216. D. Halmer, S. Thelen, P. Hering, M. Muertz, ''Online Monitoring of Ethane Traces in Exhaled Breath with a Difference Frequency Generation Spectrometer'', *Appl. Phys. B*, **85**, 437–443 (2006).

217. G. von Basum, D. Halmer, P. Hering, M. Muertz, S. Schiller, F. Mueller, A. Popp, F. Kuehnemann, ''Parts Per Trillion Sensitivity for Ethane in Air with an Optical Parametric Oscillator Cavity Leak-Out Spectrometer'', *Opt. Lett.*, **29**, 797–799 (2004).

218. M.L. Silva, D.M. Sonnenfroh, D.I. Rosen, M.G. Allen, A. O'Keefe, ''Integrated Cavity Output Spectroscopy Measurements of NO Levels in Breath with a Pulsed Room-Temperature QCL'', *Appl. Phys. B*, **81**, 705–710 (2005).

219. M. Murtz, 'Breath Diagnostics Using Laser Spectroscopy', *Opt. Photonics News*, **16**, 30–35 (2005).

220. T.H. Risby, S.F. Solga, ''Current Status of Clinical Breath Analysis'', *Appl. Phys. B*, **85**, 421–426 (2006).

221. W. Cao, Y. Duan, ''Current Status of Methods and Techniques for Breath Analysis'', *Crit. Rev. Anal. Chem.*, **37**, 3–13 (2007).

222. H. Wahl, B. Fidric, C.W. Rella, S. Koulikov, B. Kharlamov, S. Tan, A.A. Kachanov, B.A. Richman, E.R. Crosson, B.A. Paldus, S. Kalaskar, D.R. Bowling, ''Applications of Cavity Ring-Down Spectroscopy to High Precision Isotope Ratio Measurement of 13C/12C in Carbon dioxide'', *Isotopes Environ. Health Stud.*, **42**, 21–35 (2006).

223. C. Wang, A. Mbi, ''A New Acetone Detection Device Using Cavity Ringdown Spectroscopy at 266 nm: Evaluation of the Instrument Performance Using

Acetone Sample Solutions", *Meas. Sci. Technol.*, **18**, 2731–2741 (2007).

224. To be published.

225. D. Halmer, G. von Basum, P. Hering, M. Murtz, "Mid-Infrared Cavity Leak-Out Spectroscopy for Ultrasensitive Detection of Carbonyl Sulfide", *Opt. Lett.*, **30**, 2314–2316 (2005).

226. Y.A. Bakhirkin, A.A. Kosterev, C. Roller, R.F. Curl, F.K. Tittle, "Mid-Infrared Quantum Cascade Laser Based Off-Axis Integrated Cavity Output Spectroscopy for Biogenic Nitric Oxide Detection", *Appl. Opt.*, **43**, 2257–2266 (2004).

227. M.M. McCurdy, Y. Bakhirkin, G. Wysocki, R. Lewicki, F.K. Tittle, "Recent Advances of Laser-Spectroscopy-Based Techniques for Applications in Breath Analysis", *J. Breath Res.*, **1**, R1–R12 (2007); 014001.

228. M.J. Thorpe, D. Balslev-Clausen, M.S. Kirchner, Jun Ye, "Cavity Enhanced Optical Frequency Comb Spectroscopy: Application to Human Breath Analysis,"Opt. Express, **16**, 2387–2397 (2008).

229. R. Mukhopadhyay, "Don't Waste your Breath", *Anal. Chem.*, **76**, 273A–276A (2004).

230. Z. Tong, M. Jakubinek, A. Wright, A. Gillies, H.-P. Loock, "Fiber-Loop Ring-Down Spectroscopy: A Sensitive Absorption Technique for Small Liquid Samples", *Rev. Sci. Instrum.*, **74**, 4818–4826 (2003).

231. R. Li, H.-P. Loock, R.D. Oleschuk, "Capillary Electrophoresis Absorption Detection Using Fiber-Loop Ring-Down Spectroscopy", *Anal. Chem.*, **78**, 5685–5692 (2006).

232. M. Andachi, T. Nakayama, M. Kawasaki, S. Kurokawa, H.-P. Loock, "Fiber-Optic Ring-Down Spectroscopy Using a Tunable Picosecond Gain-Switched Diode Laser", *Appl. Phys. B*, **88**, 131–135 (2007).

233. D.E. Vogler, M.G. Muller, M.W. Sigrist, "Fiber-Optic Cavity Sensing of Hydrogen Diffusion", *Appl. Opt.*, **42**, 5413–5417 (2003).

234. C. Wang, 'Fiber Ringdown Pressure/Force Sensors', US Patent NO. 7,241,986, July 2007.

235. T. von Lerber, M.W. Sigrist, "Cavity Ring-Down Principle for Fiber Optic Resonators: Experimental Realization of Bending Loss and Evanescent-Field Sensing", *Appl. Opt.*, **41**, 3567–3575 (2002).

236. C. Wang, 'Fiber Bragg Grating Loop Ringdown Method and Apparatus', US Patent Pending # 60/588,035, August 2004.

237. A. Mbi, *'Novel Fiber Optic Temperature Sensors: Fiber Grating Loop Ringdown'*, M.S. thesis, Mississippi State University, May 2006.

238. www.Tigeroptics.com, 2007.

239. A.C.R. Pipino, J.W. Hudgens, R.E. Huie, "Evanescent Wave Cavity Ring-Down Spectroscopy as a Probe of Surface Processes", *Chem. Phys. Lett.*, **280**, 104–112 (1997).

240. L. van der Sneppen, F. Ariese, C. Gooijer, W. Ubachs, "Cavity Ring-Down Spectroscopy for Detection in Liquid Chromatography at UV Wavelengths Using Standard Cuvettes in a Normal Incidence Geometry", *J. Chromatogr. A*, **1148**, 184–188 (2007).

Emission Spectroscopy, Infrared

Gábor Keresztury
Hungarian Academy of Sciences, Budapest, Hungary

1 Introduction **1**
2 Theoretical Background **3**
2.1 Basic Terms and Principles 3
2.2 The Laws of Blackbody Radiation 5
2.3 Thermal Radiation of Nonblackbodies 7
2.4 Practical Aspects of Theory 8
3 Instrumentation: Spectrometers and Detectors **15**
3.1 Dispersive Instruments 16
3.2 Fourier Transform Infrared Instruments 16
3.3 Open-path Fourier Transform Infrared Instruments 18
3.4 Choice of Detectors 19
3.5 Custom-built Instruments for Special Applications 19
4 Laboratory Methods **20**
4.1 Conventional Methods of Thermal Excitation 20
4.2 Advanced Methods for Solids 22
5 Collection, Handling and Presentation of Spectral Data **24**
5.1 Data Acquisition and Reduction 24
5.2 Presentation of Emission Spectra: Intensity Scales 24
5.3 Quantitative Aspects 25
6 Conclusion and Future Perspective **25**
Abbreviations and Acronyms **26**
Related Articles **26**
References **26**

Infrared emission spectroscopy (IRES) is a branch of experimental infrared (IR) spectroscopy that can be considered as an alternative to the more commonly used IR absorption or reflection techniques for observing vibrational transitions in molecular systems. While these methods provide spectral data that are similar regarding their information content (structural, analytical), IRES may have definite advantages (or drawbacks) depending on the object to be studied.

This article aims at giving an up-to-date account of the theory, instrumentation and laboratory practice of IRES. Following an introduction to the principles and nature of thermal radiation, the governing laws of IRES are outlined and analyzed in some depth, with the aim of defining the conditions for obtaining analytically meaningful IR emission spectra. Spectral distortions concomitant with the measurement of condensed-phase samples and other sources of error are pointed out along with the procedures devised for their elimination. Various methods of excitation of thermal radiation and the instrumentation most often used in laboratory measurements of IR emission spectra are described, with emphasis on modern Fourier transform (FT) techniques and with an outlook to special applications like remote sensing and process monitoring. Attention is given to data manipulation and presentation of emission spectra in relation to quantitative evaluation of spectral data. Recent advances in sample handling techniques, for example allowing the measurement of optically thick samples or thin surface layers, are also reviewed, while reference is given to the original sources of literature in the field.

1 INTRODUCTION

IRES is concerned with the analysis of energy distribution, as a function of frequency (wavenumber) or wavelength, of the electromagnetic radiation emitted by objects (samples) in the IR spectral region. The IR region of the electromagnetic spectrum is usually considered to be from 4000 to 400 cm^{-1} (2.5 to 250 μm), or including also the near-infrared (NIR) and far-infrared (FIR) regions – from 14 000 to 10 cm^{-1} (0.7–1000 μm). Emission encountered in this spectral region is associated with vibrational, vibrational-rotational or pure rotational motions of molecules. Since these quantized motions may be more or less excited at temperatures higher than absolute zero (0 K = −273.15 °C = −459.67 °F), their relaxation leads to emission of *thermal* (mostly IR) radiation. Thermal excitation is the most common excitation mechanism of IR emission of gases, liquids and solids, and we will focus here on spectroscopic aspects of this phenomenon. Other excitation mechanisms such as chemiluminescence or photoluminescence will not be dealt with in detail as they are governed by somewhat different laws.

IR emission spectra of various objects can be continuous or discontinuous. The spectral distribution of energy in the continuous (or blackbody-like) emission spectra are characteristic only of the temperature of the emitter, while discontinuous spectra containing narrow emission bands or lines are characteristic of its material properties as

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

well. The goal of IR emission spectroscopic measurement is the same as that of IR spectroscopy in general: to get an insight, through the observable spectral features of the objects, into the chemical composition, molecular structure, configuration and conformation of the molecules, the nature of chemical bonding and intermolecular interactions within the sample. If the spectral features can be assigned to known chemical species, the measured emission spectral data can be utilized in qualitative or quantitative analysis, in much the same way as in the more commonly used infrared absorption spectroscopy (IRAS). Depending on the samples and conditions of measurement (temperature, physical state, energy conditions) one or the other IR spectroscopic observation technique may be preferred. A special topic of IRES of great theoretical interest is the investigation of highly excited vibrational and vibrational-rotational states, not conveniently accessible to absorption studies.

In laboratory practice, the instruments used for the measurement of IR emission spectra are essentially the same as those used in IRAS, with a difference in sample arrangement: in IRES the emitting sample replaces the standard light source or it is put in an optically equivalent position (often in an auxiliary sample compartment outside the instrument cover). In principle, any type of IR spectrometer is capable of measuring IR emission spectra, but since the radiation to be detected may be rather weak (when the sample temperature is relatively low), high-performance FT (interferometric) spectrometers are preferred to dispersive instruments owing to their higher optical throughput and other known advantages leading to higher sensitivity.

Thus, a typical Fourier transform infrared (FTIR) instrument used for the measurement of IR emission spectra consists of an emission port (receiving the radiation from the auxiliary emission accessory), a modulator (interferometer) and a broad-range IR detector. The IR radiation emitted by the sample enters the interferometer in the form of a collimated beam where it gets modulated, then is focused on the detector. The analog signal of the interferogram is measured at equidistant data points and is converted to digital data, by means of an analog-to-digital converter (ADC). As a rule, the interferogram is measured repeatedly many times and the data are accumulated (averaged) to increase the signal-to-noise ratio (S/N). The so-called single-beam emission spectrum (the energy spectrum) is obtained from the interferogram via Fourier transformation by means of a PC or dedicated microcomputer. Double-beam emittance spectra (corresponding to transmittance spectra in absorption spectroscopy) are obtained by ratioing the single-beam emission spectrum of the sample to that of a laboratory blackbody, used as reference, measured under the same conditions (same temperature) as the sample.

The applicability and potentials of IRES as a method of measurement to solve analytical problems must be considered in comparison with those of IR absorption and reflection techniques. One of the main advantages of IRES stems from its single-ended measurement geometry, which means that the sample need not be put inside the instrument: it should be accessible only from one side or is seen from a distance. This makes IRES one of the basic methods for remote sensing, e.g. of natural sources of thermal radiation such as hot gases, and explains its widespread application in environmental, atmospheric, extraterrestrial or astrophysical research. The same applies to in situ measurements for process control, e.g. in the chemical industry. In laboratory conditions IRES is preferred when very hot or chemically aggressive samples have to be examined which cannot be or would be risky to place into the sample compartment of a spectrometer. The IR emission technique is equally applicable to smooth and rough surfaces, and can be used with transparent and nontransparent substrates such as metals. Thin overlayers on metal surfaces (e.g. lubricants on moving machine parts) are ideal samples for IRES without any sample preparation, and chemisorbed molecules on metal single crystal surfaces can be detected even at monomolecular coverage when using special detection techniques.

The principal drawback to IRES is that for IR emission to occur, a molecule must go from an excited vibrational state back to the ground state; and in order to have a detectable flux of such radiation, there must be a significant proportion of molecules in the excited state. To achieve such a population of vibrational modes whose transitions are observed in the mid-IR region, the sample must be heated, which may not only be inconvenient but also injurious to many substances, especially those of organic origin. A further difficulty is introduced by background emission: the emission of the hot surroundings of the sample, which may overwhelm the signal from the sample of interest. Thus, for practical applications of IRES, a restricting condition is that during measurement the sample must be kept at a constant temperature differing from that of the spectrometer (detector).

Furthermore, with condensed-phase samples, IRES is limited to the examination of thin layers not exceeding a few micrometers in thickness – just as in transmission. Measurement of the surfaces of much thicker samples or thin overlayers on top of opaque substrates is not a simple task; it requires special sample handling methods such as transient heating or cooling to obtain meaningful emission spectra. Since IR emission is often accompanied by undesirable absorption and reflection phenomena and heat transfer that may disturb the observation or distort the emission spectrum, a good understanding of the

underlying laws and principles is needed to set up the right measurement conditions.

Several earlier reviews on IRES are available in the literature reflecting the state of the art and the field of interest of the time of their publication.[1–6] Our presentation will necessarily overlap with these, while illuminating some new topics of present-day interest.

2 THEORETICAL BACKGROUND

2.1 Basic Terms and Principles

2.1.1 Infrared Radiation

Just like visible light, IR radiation is part of the spectrum of electromagnetic waves. At the same time it may also be considered as consisting of massless energy parcels called photons. In fact, both descriptions are needed to account for all radiative and optical phenomena that have been observed.

An electromagnetic wave may be identified by its wavelength (λ), frequency (ν), or wavenumber ($\tilde{\nu}$). These three quantities are related to each other through Equation (1):

$$\tilde{\nu} = \frac{\nu}{c} = \frac{1}{\lambda} \tag{1}$$

where c is known as the speed of light (actually, it refers to the velocity of all kinds of electromagnetic radiation); in vacuum $c = c_0 = 2.99792458 \times 10^8\,\mathrm{m\,s^{-1}}$. Note that the speed of light depends on the medium in which the wave travels, and may be related to the speed of light in vacuum, c_0, by Equation (2):

$$c = \frac{c_0}{n} \tag{2}$$

where n is the refractive index of the medium. By definition, the refractive index of vacuum is unity ($n \equiv 1$), while for any other media it is greater: for most gases it is very close to 1 (for air $n = 1.0003$ in the visible part of the spectrum), but most dielectric media have n values as great as 1.4 to 4.

The wavelength, λ (expressed in $\mu\mathrm{m} = 10^{-6}\,\mathrm{m}$), is a quantity that can be measured experimentally (e.g. with an interferometer). The wavelength scale was introduced in spectroscopy in 1868 and has been used since, mainly in the ultraviolet (UV) and visible regions. The frequency, ν, is defined as the number of waves that pass a given point in one second (measured in cycles/s $= \mathrm{s^{-1}} = \mathrm{Hz}$). However, in the IR region it is more practical to divide the actual frequency value with the speed of light and replace it with the much smaller value obtained, the wavenumber, $\tilde{\nu}$, which gives the number of waves in one centimeter of vacuum, leading to its units as $\mathrm{cm^{-1}}$. It is this spectral variable and measurement unit that are almost exclusively used in IR spectroscopy (although the wavenumber values are often referred to as ''frequencies'' in the spectroscopists' slang).

An electromagnetic wave or photon carries with itself an amount of energy, E, proportional to its frequency, (Equation 3):

$$E = h\nu = hc\,\tilde{\nu} \tag{3}$$

where $h = 6.62608 \times 10^{-34}\,\mathrm{J\,s}$ is known as the Planck constant. The unit of energy most often used in connection with photons or electromagnetic waves is the electron volt, eV ($1\,\mathrm{eV} = 1.6022 \times 10^{-19}\,\mathrm{J}$). The energy of the photon is conserved as light passes from one medium to another, which means that the frequency of light does not change at the interface. Note, however, that in accordance with Equations (1) and (2), wavelengths and wavenumbers do change when light enters another medium with a different refractive index. (However, in everyday applications of spectroscopy this dependence is generally neglected, i.e. wavenumber values pertaining to vacuum are used.)

The arrangement and limiting values of the different regions of the electromagnetic wave spectrum are shown in Figure 1 in terms of the above mentioned scales. The terms IR emission and thermal radiation are often used as synonyms; in fact, thermal radiation, defined as electromagnetic waves that are emitted by a medium owing solely to its temperature, may extend from IR through the visible spectrum into the near UV region in case of very hot emitters (e.g. incandescent light sources, or the sun).

2.1.2 Principles of Thermal Radiation

All media continuously emit electromagnetic radiation, randomly in all directions, depending on their

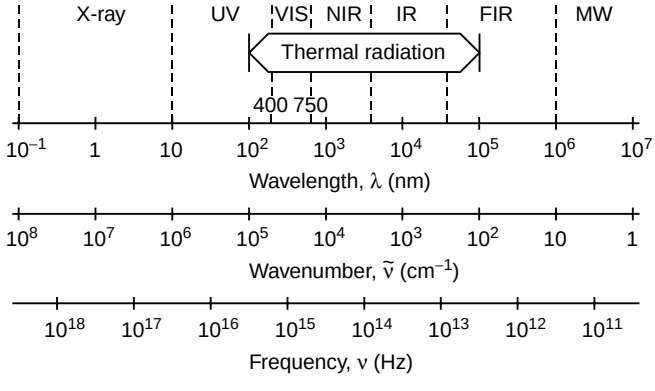

Figure 1 Thermal radiation in the spectrum of electromagnetic radiation.

temperature and on the properties of their material. Radiation emitted by a homogeneous medium is also homogeneous, unpolarized and isotropic. Before discussing the quantitative laws of thermal radiation that are fundamental to IR emission spectroscopy of gases, liquids and solids, a general acquaintance with the nature of thermal radiation and thermal equilibrium may be appropriate.

Any emission of IR or thermal radiation is concomitant with some loss of energy of the emitter. Since objects at a temperature of absolute zero (0 K) are at their lowest possible energy level, they are obviously unable to emit any energy in that state. But at higher temperatures a certain proportion of molecules are rotationally and vibrationally excited (electronic excitation takes place at much higher than room temperature), and these molecules spontaneously emit energy in the form of thermal radiation while relaxing to lower-lying energy levels. At the same time the surrounding molecules or objects may absorb some of the emitted photons and get excited. Thus there is a continuous radiative exchange of energy leading to redistribution and equalization of temperature within a thermally isolated system.

Let us consider an evacuated enclosure with totally reflecting inner walls, thermally isolated from the outside, containing two objects of similar mass and different temperature. In accordance with the general experience, the two objects will, in time, have the same temperature. Since heat conduction and convection is excluded in vacuum, the only interaction between the two objects leading to this result is thermal radiation. As was discovered at the beginning of the 19th century (C. Prévost, 1809), any object emits energy depending on its own temperature, irrespective of the temperature of the surroundings. Thus in our example the hotter object initially emits more energy than it absorbs and gets cooler, while the cooler object absorbs more energy than it emits and gets warmer, until thermal equilibrium sets in. At thermal equilibrium the objects emit just as much radiant energy per second as they absorb, thus there is also radiation equilibrium, i.e. a uniform radiation field inside the enclosure.

The outcome of our fictional experiment will be the same if one of the objects in the isolated enclosure is absolutely black throughout the spectrum (usually referred to as blackbody) and the other is not. But since the blackbody absorbs all radiation falling on it, at thermal equilibrium it must also emit the same amount of energy. The other object may absorb less (and reflect or transmit the rest) but then it will also emit less than the black object. By this reasoning we discover a one-to-one correspondence between absorptance (α) and emittance (ε) of an object, an important law discussed in more detail in section 2.3.

2.1.3 Radiant Energy

The radiation field is characterized by its radiant energy density, H, defined in Equation (4) as

$$H = \frac{Q}{V} \tag{4}$$

where Q is the total radiant energy and V is the volume. The radiation field is composed of radiation of various frequencies or wavelengths, which can be represented by Equation (5):

$$H = \int_0^\infty H_\nu \times \mathrm{d}\nu = \int_0^\infty H_{\tilde{\nu}} \times \mathrm{d}\tilde{\nu} = \int_0^\infty H_\lambda \times \mathrm{d}\lambda \tag{5}$$

where H_ν, $H_{\tilde{\nu}}$, and H_λ are the spectral radiant energy densities expressed in terms of frequency, wavenumber and wavelength, respectively. (''Spectral'' in these terms indicates that the radiant energy density is a function of the actual spectral variable, and is considered within the spectral intervals (ν, $\nu + \mathrm{d}\nu$), ($\tilde{\nu}$, $\tilde{\nu} + \mathrm{d}\tilde{\nu}$) or ($\lambda$, $\lambda + \mathrm{d}\lambda$) respectively.)

2.1.4 Energy Levels, Transitions and Emission Probability

The total energy of a molecular system is composed of the energy of electronic, vibrational and rotational motions and, according to quantum theory, it can assume discrete values depending on quantum numbers that specify the level of excitation of each type of motion. Any absorption or emission of IR radiation is associated with transitions between vibrational or vibrational-rotational energy levels (Figure 2), whereas the amount of energy absorbed or released, ΔE, is proportional to the frequency, ν, or wavenumber, $\tilde{\nu}$, of the radiation (Equation 6):

$$\Delta E = E_j - E_i = h\nu = hc\,\tilde{\nu} \tag{6}$$

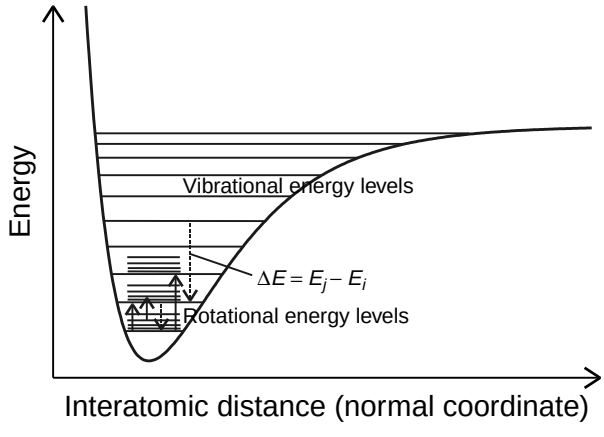

Figure 2 Energy levels in the potential energy curve of a diatomic molecule; arrows drawn with full line signify absorption, arrows with dotted lines signify emission.

where E_j is the energy of the upper level, E_i is that of the lower level. Transition from a higher to a lower energy state (designated as $E_j \rightarrow E_i$) is accompanied by emission of energy, while transition from lower to higher energy states ($E_j \leftarrow E_i$) is concomitant with absorption of energy. The energy levels and transitions between them are often characterized, instead of the energies, by the so-called term values obtained when both sides of Equation (6) are divided by hc, leading to quantities given in units of cm^{-1}. The frequent use of Equation (6) justifies that in IR spectroscopy priority is given to the wavenumber scale and the cm^{-1} units.

A molecule consisting of N atoms has, in general, $3N - 6$ vibrational degrees of freedom. Depending on the atomic masses, the geometrical arrangement of atoms within the molecule and the forces acting between them, an equal number of so-called normal vibrations will be formed, each of them having a well-defined fundamental vibrational frequency, ν_k, associated with a specific form or mode of vibration, and a series of quantized energy levels. According to the theory of molecular vibrations,[7] at a temperature of absolute zero no vibrational modes are excited (all normal modes assume their zero-point motions), and the vibrational energy of the system is at its minimum. However, at higher temperatures, when the molecules are in thermal equilibrium with their surroundings, a certain portion of them is always at thermally excited vibrational and rotational levels (excitation of electronic energy takes place at much higher temperatures than normal room temperature). The excited molecules spontaneously emit energy in the form of electromagnetic radiation. The intensity of an emission line belonging to transition $j \rightarrow i$, defined as the amount of energy emitted by the sample per second, is determined by Equation (7):

$$I_{ij} = N_j hc\,\tilde{\nu} A_{ij} \tag{7}$$

where N_j is the number of molecules in the excited state j, and A_{ij} is the transition probability (also known as the Einstein coefficient of spontaneous emission). The population of an excited state, N_j, relative to that of the initial (lower) state, N_i, as a function of temperature is governed by Boltzmann statistics (Equation 8):

$$\frac{N_j}{N_i} = \exp\ \frac{-(E_j - E_i)}{kT}\ = \exp\ \frac{-hc\,\tilde{\nu}}{kT} \tag{8}$$

where $k = 1.380658 \times 10^{-23}\,J\,K^{-1}$ is the Boltzmann constant and T is the absolute temperature. Thus the fraction of excited molecules decays exponentially with the excitation energy but increases with temperature. For instance, at 300 K for a vibration at $200\,cm^{-1}$ the value of N_j/N_i is expected to be 0.38, while for a $1000\,cm^{-1}$ vibration it is only 0.0008; at 500 K these values become about 0.56 and 0.06, respectively. The transition probability of spontaneous emission is defined as the fraction of molecules undergoing transition from the excited state j to the lower state i per second, and can be calculated as in Equation (9):

$$A_{ij} = \frac{64\pi^4 \tilde{\nu}_{ij}^3}{3h} |\vec{M}_{ij}|^2 \tag{9}$$

where the transition moment (roughly: the change of the electric dipole moment vector, $\vec{\mathbf{M}}$, during the vibrational or rotational transition), $\vec{M}_{ij} = \int \psi_j^* \hat{\mathbf{M}} \psi_i d\tau$, with ψ_i and ψ_j being the eigenfunctions of the initial and final states, respectively, and $\hat{\mathbf{M}}$ the electric dipole moment operator.

The above relations provide the general theoretical basis for prediction of band intensities in the emission spectrum. It has been recognized, however, that the limiting values of these intensities are those obtained for an absolute blackbody, and it is more practical to relate the emission spectrum of a sample to that of the blackbody.

2.2 The Laws of Blackbody Radiation

2.2.1 Stefan–Boltzmann Law

Much of the early classical research on thermal radiation was devoted to the search for the correct form of the spectral radiant energy density of blackbody, $H_{\tilde{\nu}}$, as a function of frequency (or wavelength) and temperature. But before that problem could be solved, the temperature dependence of the total radiant energy density of the radiation emitted by a blackbody, H, was established by Stefan experimentally (1878) and confirmed by Boltzmann theoretically (1884). The relation they obtained is known as the Stefan–Boltzmann law (Equation 10):

$$H = \sigma \times T^4 \tag{10}$$

with $\sigma = 2\pi^5 k^4/15c^2h^3 = 5.6705 \times 10^{-8}\,W\,m^{-3}\,K^{-4}$ being the Stefan–Boltzmann constant (where $k = 1.380658 \times 10^{-23}\,J\,K^{-1}$ is the Boltzmann constant).

2.2.2 Wien's Displacement Law

Two partially successful approximations to the spectral distribution of the blackbody radiation were published around the end of the 19th century, the Rayleigh–Jeans law, acceptable only for low frequencies, and Wien's radiation law, quite accurate for the high-frequency region; now they are only of historical value. Another empirical law discovered prior to the right form of the blackbody radiation function is Wien's displacement law

(Equation 11) relating the frequency of the maximum of blackbody radiation to the absolute temperature:

$$\tilde{\nu}_{\max} = \text{const} \times T \tag{11}$$

2.2.3 The Planck Function

The correct analytical form of the spectral distribution of blackbody radiation was found by M. Planck in 1900 as an interpolation formula between the above-mentioned two approximations; soon afterwards he established his finding theoretically as well by extending Boltzmann statistics with the quantization condition for the emitted energy. The distribution function he obtained this way had integrated in itself all radiation laws established earlier and is known as Planck's law of blackbody radiation or the Planck function. It is also in agreement with all observed data. The actual form of the Planck function depends on whether it is expressed in terms of frequency, wavenumber or wavelength. In the frequency representation (Equation 12), the density of radiant energy emitted from a unit area of the blackbody source per solid angle in spectral element $d\nu$ is:

$$H_\nu(T)\,d\nu = \frac{2hc^{-2}\nu^3}{e^{h\nu/kT} - 1}\,d\nu \tag{12}$$

(Variations of the Planck function containing a factor of 8π instead of 2 in the above equation refer to radiance considered in the whole solid angle.)

The wavenumber and wavelength representations can be obtained from Equation (12) through Equation (13):

$$\nu = c\tilde{\nu} = \frac{c}{\lambda} \quad \text{and} \quad d\nu = c\,d\tilde{\nu} = -\frac{c}{\lambda^2}\,d\lambda \tag{13}$$

and using the equivalence of the three representations expressed in Equation (5), to give Equations (14) and (15):

$$H_{\tilde{\nu}}(T)\,d\tilde{\nu} = \frac{2hc^2\tilde{\nu}^3}{e^{hc\tilde{\nu}/kT} - 1}\,d\tilde{\nu} \tag{14}$$

and

$$H_\lambda(T)\,d\lambda = \frac{2hc^2}{\lambda^5(e^{hc/\lambda kT} - 1)}\,d\lambda \tag{15}$$

It is customary to introduce the so-called first and second radiation constants, C_1 and C_2, having the values given in Equations (16) and (17):

$$C_1 = 2hc^2 = 1.1911 \times 10^{-12}\,\text{J cm}^2\,\text{s}^{-1} \tag{16}$$

and

$$C_2 = \frac{hc}{k} = 1.4388\,\text{cm K} \tag{17}$$

and write the Planck function in the wavenumber representation simply as Equation (18):

$$H_{\tilde{\nu}}(T) = \frac{C_1\tilde{\nu}^3}{e^{C_2\tilde{\nu}/T} - 1} \tag{18}$$

which is more practical to use when only the relative energy distribution is important (C_1 may be adjusted to fit the actual application).

Note that Equations (12) and (14) are the same within a constant factor, while Equation (15) describes curves of essentially different shapes, with their maxima shifted considerably toward higher wavenumbers. Thus there are two basically different representations of the Planck function, as shown in Figure 3 (in the same linear wavenumber scale, for three different temperatures). Correspondingly, also Wien's displacement law (Equation 11) has two versions, Equations (19) and (20):

$$\tilde{\nu}_{\max}[\text{cm}^{-1}] = 1.93 \times T[\text{K}] \tag{19}$$

for Equation (10), and

$$\lambda_{\max}[\text{cm}] = \frac{0.283}{T[\text{K}]} \tag{20}$$

for Equation (15), and these relations give different maxima for the same temperatures. One has to have a clear idea which of these forms are to be used when applied to IR emission spectroscopy. Seeking the connection with the measurement of blackbody emission spectra, the curves shown in Figure 3 in full line (wavenumber version, Equation 14) can be obtained if the spectrum is scanned with a spectral slit constant in wavenumber (or frequency), while those shown in broken line would be obtained if the spectrum were

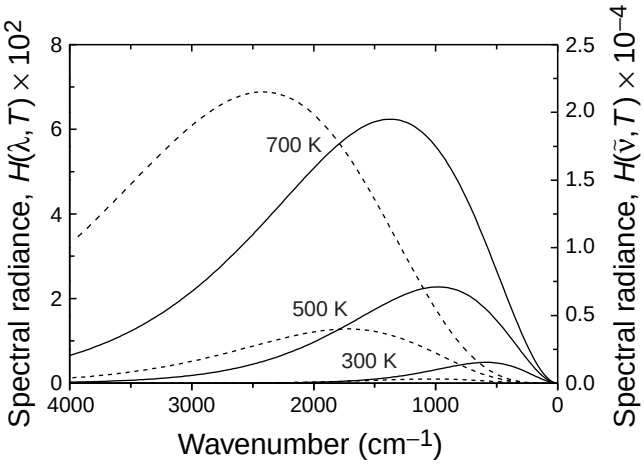

Figure 3 Theoretical blackbody radiation curves (Planck functions) for three different temperatures in the wavenumber representation (full lines) and in the wavelength representation (broken lines).

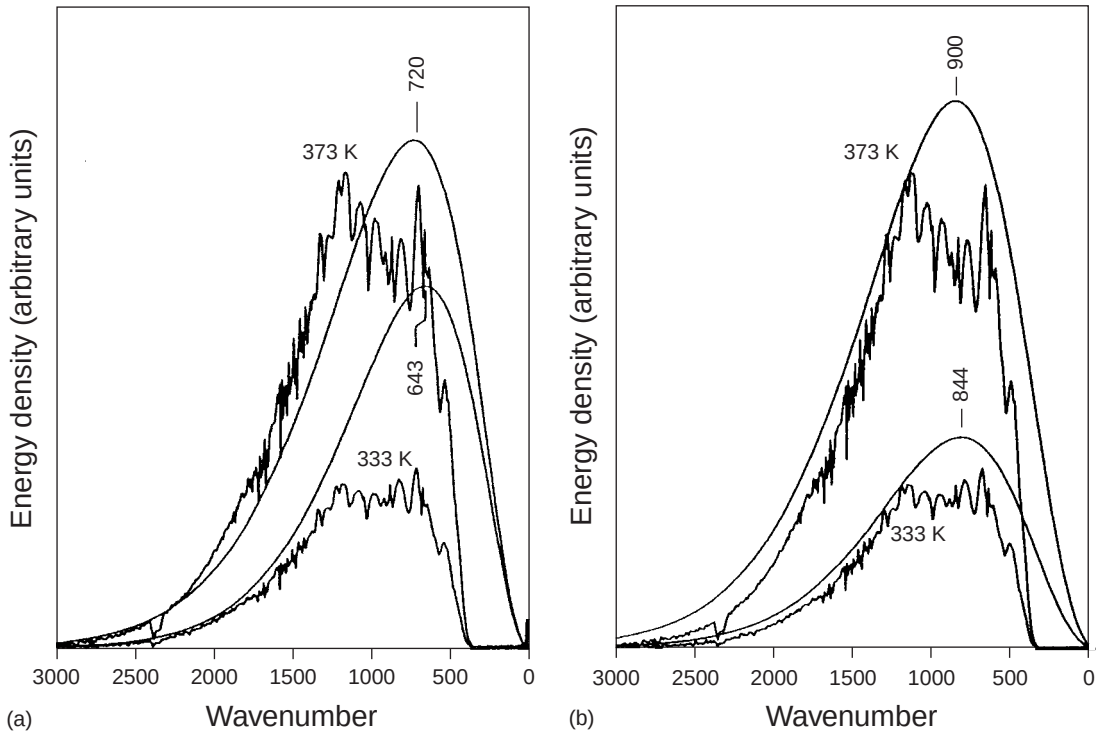

Figure 4 Comparison of blackbody emission spectra measured on a Digilab FTS-20C spectrometer at 60 and 100 °C (with deuterated triglycine sulfate (DTGS) detector at 29 °C) with uncorrected Planck functions (a), and with Planck functions corrected for detector temperature (b). (Reproduced by permission of the Society for Applied Spectroscopy from Keresztury, Mink, *Appl. Spectrosc.*, **46**, 1747–1749 (1992).[9])

scanned with a spectral slit constant in wavelength. Since modern interferometric (FT) IR spectrometers yield spectra with constant resolution in the wavenumber scale in the whole working region, it is imperative that with FTIR instruments the wavenumber version of the Planck function be used (i.e. Equation 14 or 18) whenever a theoretical blackbody reference spectrum is needed.

Another pitfall that should be avoided when theoretical and experimental blackbody emission spectra are compared (e.g. in order to determine the emitter's temperature or to find the instrument response function) is to assume that all radiation collected is also detected by the instrument. From the principles of thermal radiation we know that the detector (which is supposed to be as nearly a blackbody as possible) also emits thermal radiation at temperatures higher than absolute zero. Since IR detectors can respond only to deviations from radiation equilibrium, the detector does not 'see' all radiation falling on it, only the difference of the blackbody-like emissions from the sample and from the detector (no matter which of them is at higher temperature).[8] Thus, if a room-temperature detector is used and the sample temperature is not very high, the measured spectrum is related to the difference of the corresponding Planck functions.[9] This is illustrated in Figure 4.

2.3 Thermal Radiation of Nonblackbodies

When a monochromatic beam of light of unit intensity falls on a slab of condensed-phase material (see Figure 5a), it is divided into three fractions: one may be reflected off the surface, another is absorbed inside the layer and the remainder is transmitted, so that (if scattering and luminescence are neglected) these three fractions add up to unity (Equation 21):

$$\alpha + \rho + \tau = 1 \tag{21}$$

where α, ρ and τ are known as the absorptance (or absorption factor), reflectance (reflection factor) and transmittance (transmission factor) of the sample, respectively. (The same quantities are sometimes erroneously

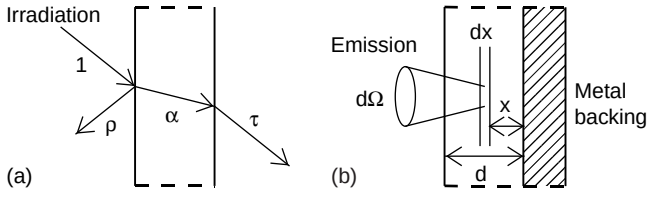

Figure 5 Schematics of (a) irradiation of a semitransparent slab, and (b) observation of emission by a supported slab.

called absorptivity, reflectivity and transmissivity in the literature. The terms with the ending -ance refer to properties of a body (with a given composition, shape, etc.), while the terms with the ending -ivity should refer to properties of a substance measured under standard conditions.)

By definition, a blackbody absorbs all radiation falling on it, thus the absorptance of a blackbody is unity at all frequencies, (Equations 22–24):

$$\alpha = 1 \quad \text{and} \quad \rho = \tau = 0 \tag{22}$$

At another extreme, for a shiny, reflecting metal surface

$$\rho \approx 1 \quad \text{and} \quad \alpha \approx \tau \approx 0 \tag{23}$$

while for an ideal spectroscopic window material

$$\tau \approx 1 \quad \text{and} \quad \alpha \approx \rho \approx 0 \tag{24}$$

However, for most of the common materials, α, ρ and τ are not only between 0 and 1 but strongly wavenumber-dependent as well. They may have narrow absorption lines or broader absorption bands at specific frequencies, and the reflectance may also change in the vicinity of the peaks of strong absorption bands. Those objects for which α, ρ and τ have intermediate values between 0 and 1 throughout the whole spectrum are called graybodies. The ideal laboratory blackbody does not exist; blackbody samples used in emission measurements as reference are, in fact, graybodies that have absorptance or emittance above 0.90 (some may come close to 0.99) in a very broad spectral interval.

2.3.1 Kirchhoff's Law

One of the basic principles of IRES is the principle of equivalence of absorptance and emittance (discovered by G.W. Kirchhoff in 1860). In its original form Kirchhoff's law (Equation 25) states that the ratio of the spectral luminance (or radiant flux) of any sample, L_s, at a given temperature, T_i, to its absorptance, α_s, equals the luminance of the blackbody, L_{bb}, at the same temperature.

$$\frac{L_s(\tilde{\nu}, T_i)}{\alpha_s(\tilde{\nu})} = L_{bb}(\tilde{\nu}, T_i) \tag{25}$$

which implies that $\alpha_{bb} = 1$. At the same time the emittance of a sample, $\varepsilon_s(\tilde{\nu})$, is defined (Equation 26) as the ratio of the radiant flux emitted by the sample (L_s) to the flux emitted by a blackbody (L_{bb}) at the same temperature:

$$\varepsilon_s(\tilde{\nu}) = \frac{L_s(\tilde{\nu}, T_i)}{L_{bb}(\tilde{\nu}, T_i)} \tag{26}$$

L_s may be regarded as the observed luminance of the sample or, as it is often referred to, its single-beam emission spectrum (see also section 2.4). Comparison of Equations (25) and (26) yields Equation (27):

$$\varepsilon_s(\tilde{\nu}) \equiv \alpha_s(\tilde{\nu}) \tag{27}$$

which is the essence of Kirchhoff's law.

According to this, if a sample absorbs light of a certain frequency, it will also emit at the same frequency. Moreover, this relation means that not only the frequencies but also the so-called selection rules that determine the band intensities will be the same for absorption and emission processes involving the same energy levels of the molecule.

Thus, substituting emittance for absorptance in Equation (21) one gets the basic relation of optical emission spectroscopy (Equation 28):

$$\varepsilon(\tilde{\nu}) + \rho(\tilde{\nu}) + \tau(\tilde{\nu}) = 1 \tag{28}$$

This relation has to be taken into account when choosing the materials for building an emission cell or when interpreting IR emission spectra of condensed materials.

2.4 Practical Aspects of Theory

It follows from Equations (22) and (27) that a blackbody (with $\alpha = 1$) is an ideal emitter (Equation 29):

$$\varepsilon_{bb} = 1 \quad \text{and} \quad \rho_{bb} = \tau_{bb} = 0 \tag{29}$$

Since these conditions are valid for all parts of the spectrum, no material can emit more energy at any given frequency and at any given temperature than a blackbody.

For other types of materials with properties described by Equations (23) and (24) analogous relations can be obtained by substituting emittance for absorptance. Thus, a shiny metal surface is a very poor emitter ($\varepsilon = \alpha \approx 0$), so it makes an excellent back-plate for the sample. In contrast, IR-transmitting cell windows are rarely used as back-plates in IRES since they transmit background radiation.

Nonblackbodies or graybodies in general have emittance

$$\varepsilon(\tilde{\nu}) = 1 - \tau(\tilde{\nu}) - \rho(\tilde{\nu}) \tag{30}$$

in the range of 0 to 1, which changes with wavenumber (Equation 30).

An opaque body's transmittance is zero ($\tau_{op} = 0$) but its reflectance may be significant in some spectral regions ($\rho(\tilde{\nu}) \neq 0$), therefore (Equation 31):

$$\varepsilon_{op}(\tilde{\nu}) = \alpha_{op}(\tilde{\nu}) = 1 - \rho(\tilde{\nu}) \tag{31}$$

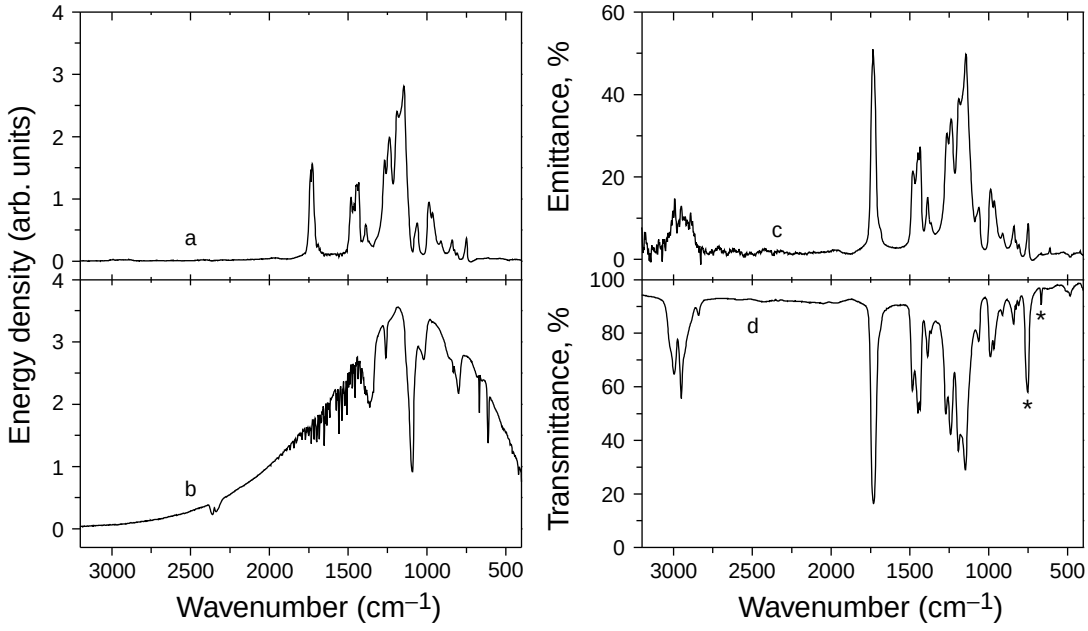

Figure 6 Comparison of emission and absorption spectra of a poly(methyl methacrylate) (PMMA) film: (a) Single-beam emission spectrum of PMMA film on an Al foil at 120 °C; (b) blackbody reference emission spectrum at 120 °C; (c) double-beam emittance spectrum of PMMA obtained as a ratio of A to B; (d) transmittance spectrum of the same PMMA film measured in reflection–absorption mode (* indicates chloroform inclusion bands).

In molecular spectroscopy one is usually interested in internal properties of the sample, when surface effects may be neglected ($\rho \approx 0$); in this case (Equation 32):

$$\varepsilon(\tilde{\nu}) \approx 1 - \tau(\tilde{\nu}) \tag{32}$$

which shows the complementary nature of transmittance and emittance spectra. The spectra in Figure 6 illustrate this relation: the emittance spectrum of a PMMA film on an Al foil (Figure 6c) was obtained by ratioing the single-beam emission spectrum of the sample (Figure 6a) to that of a blackbody reference (Figure 6b); the transmittance spectrum shown for comparison (Figure 6d) was measured in reflection–transmission mode.

2.4.1 Single- and Double-beam Emission Spectra and Background Effects

Rearranging Equation (26) one obtains $L_s(\tilde{\nu})$, the single-beam emission spectrum of the sample (Equation 33):

$$L_s(\tilde{\nu}, T_i) = \varepsilon_s(\tilde{\nu}) \times L_{bb}(\tilde{\nu}, T_i) \tag{33}$$

This relation shows that the so-called single-beam emission spectrum of any sample at a certain temperature can be viewed as its emittance, $\varepsilon_s(\tilde{\nu})$, modulated by the blackbody radiation spectrum corresponding to the same temperature.

In the ideal case, if the detector received radiation only from the sample and responded to it perfectly, $L_{bb}(\tilde{\nu}, T_i)$ could be replaced by the corresponding Planck function, $H(\tilde{\nu}, T_i)$, yielding Equation (34):

$$L_s(\tilde{\nu}, T_i) = \varepsilon_s(\tilde{\nu}) \times H(\tilde{\nu}, T_i) \tag{34}$$

However, the actually measured spectra are affected by background radiation, $B(\tilde{\nu})$, and by imperfections of the instrument used. To account for these effects one can write Equation (35):[8]

$$L_s(\tilde{\nu}, T_i) = R(\tilde{\nu}) \times [\varepsilon_s(\tilde{\nu}) \times H(\tilde{\nu}, T_i) + B(\tilde{\nu})] \tag{35}$$

where $R(\tilde{\nu})$ is the instrument response function (in the ideal case $R(\tilde{\nu}) = 1$). $B(\tilde{\nu})$ represents background radiation of different origin often written as the sum of two terms, $B(\tilde{\nu}) + I(\tilde{\nu}) \times \rho(\tilde{\nu})$, to distinguish between self-radiation of the instrument, and outside radiation reflected off the sample. Equation (35) assumes a detector kept at or near 0 K, which is rarely the case. With liquid-nitrogen-cooled detectors ($T_d = 77$ K) Equation (35) is a good approximation. However, if a room-temperature detector is used which emits a fair amount of energy, the above equation must be extended to Equation (36), with one more term corresponding to its radiation:[10]

$$L_s(\tilde{\nu}, T_i, T_d) = R(\tilde{\nu}) \times [\varepsilon_s(\tilde{\nu}) \times H(\tilde{\nu}, T_i) - \varepsilon_d(\tilde{\nu}) \times H(\tilde{\nu}, T_d) + B(\tilde{\nu})] \tag{36}$$

where ε_d is the emittance of the detector and T_d is its temperature (note the negative sign of the added term, meaning opposite direction of that radiation flux).

A similar relation can be written (Equation 37) for the measured single-beam emission spectrum of the laboratory blackbody reference:

$$L_{bb}(\tilde{\nu}, T_i, T_d) = R(\tilde{\nu}) \times [H(\tilde{\nu}, T_i) - H(\tilde{\nu}, T_d) + B(\tilde{\nu})] \quad (37)$$

where it has been assumed that $\varepsilon_{bb}(\tilde{\nu}) = \varepsilon_d(\tilde{\nu}) = 1$.

Now the emittance of the sample, $\varepsilon_s(\tilde{\nu})$, cannot be obtained as the ratio of L_s and L_{bb}, as has been defined in Equation (26), unless the last two terms in Equation (36) are eliminated. The standard procedure to do it involves making four measurements:[11] the emission spectra of both the sample and the blackbody reference are measured at two temperatures; then, assuming that the sample emittance is independent of temperature over the operating range, the emittance of the sample can be calculated as the ratio of the increase in sample luminance to that of the blackbody luminance (Equation 38):

$$\varepsilon_s(\tilde{\nu}) = \frac{L_s(\tilde{\nu}, T_2) - L_s(\tilde{\nu}, T_1)}{L_{bb}(\tilde{\nu}, T_2) - L_{bb}(\tilde{\nu}, T_1)} \quad (38)$$

where $T_2 > T_1$. Substituting the full expression for luminances from Equations (36) and (37) into Equation (38) it can be seen that using this procedure, which is referred to as background correction,[8] all terms that do not depend on sample temperature are, in principle, eliminated.

When a low-temperature detector is used, $B(\tilde{\nu})$ includes radiation emitted by the surface of the beamsplitter in the interferometer (oxide species formed on the surface of the semi-transparent Ge layer) mainly responsible for the so-called instrument self-emission. It has been found[11] that subtraction of the interferograms rather than the single-beam spectra in Equation (38) leads to better compensation of the beamsplitter emission, since this avoids some phase-correction problems. Radiation originating in the beamsplitter suffers a different number of reflections in the interferometer than radiation coming from the sample, which creates different phase relations for this part of the signal, not correctly accounted for in the process of phase correction that follows Fourier transformation.[12] Figure 7 shows a typical example of background correction in the case of a cooled MCT detector.[8] One can see that the spectral features can be identified in the uncorrected spectrum (broken line) as well, but the relative intensities are strongly altered. Hence, in quantitative analysis based on emission spectra background correction is obligatory.

2.4.2 *Role of Instrument Response*

Equations (35–37) show that the instrument response function, $R(\tilde{\nu})$, has an impact on the appearance of single-beam emission spectra. Consequently, $R(\tilde{\nu})$ can be determined experimentally by comparing the emission spectrum of a good laboratory blackbody to

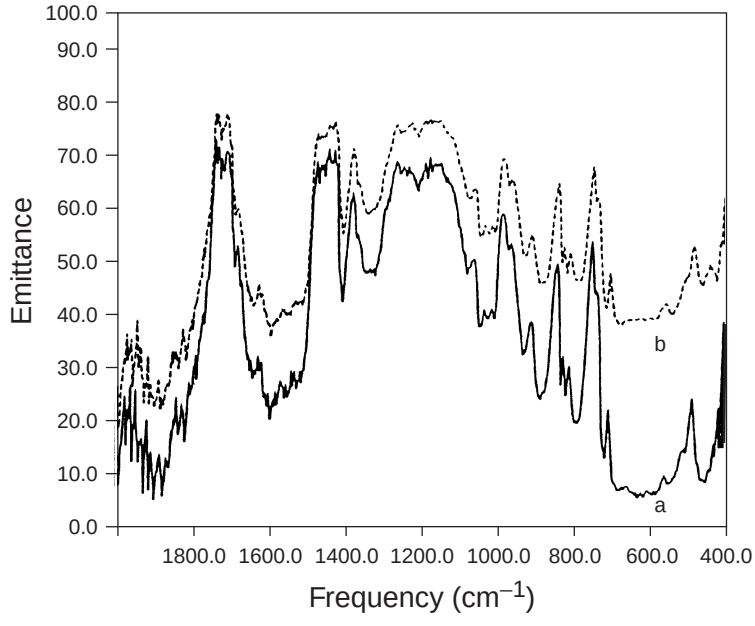

Figure 7 Emittance spectra of PMMA film using an mercury cadmium telluride (MCT) detector: (a) with background correction, and (b) without background correction. (Reproduced by permission of the Society of Applied Spectroscopy from Chase, *Appl. Spectrosc.*, **35**, 77–81 (1981).[8])

the appropriate Planck functions. For instance, using a room-temperature detector, stray radiation of the room-temperature surroundings (background) can be neglected, and $R(\tilde{\nu})$ can be obtained from Equation (37) as follows, from Equation (39):

$$R(\tilde{\nu}) = \frac{L_{bb}(\tilde{\nu}, T_i, T_d)}{[H(\tilde{\nu}, T_i) - H(\tilde{\nu}, T_d)]} \quad (39)$$

i.e. as the ratio of the measured single-beam blackbody spectrum to the difference of two theoretical blackbody emission curves corresponding to the temperature of the sample and the detector, respectively. $R(\tilde{\nu})$ is expected to be independent of sample temperature but the precision of its determination is the higher, the greater the temperature difference.

The explicit knowledge of $R(\tilde{\nu})$ is not necessary when working with double-beam emittance spectra, but it may prove very useful in several other applications:

- Blackbody reference spectra can be simulated for any desired temperature by multiplying $R(\tilde{\nu})$ with the corresponding Planck function, which can save time by eliminating a great number of reference measurements;
- Single-beam emission spectra can be corrected for instrumental distortions if they are divided by $R(\tilde{\nu})$;
- If the temperature of a blackbody radiator is to be determined, before fitting the Planck functions to the measured single-beam spectra they must be corrected for detector temperature and instrument response.[9]

2.4.3 *Effect of Sample Temperature*

In order to be detected, the radiating sample may be warmer or cooler than the detector (Figure 8), i.e. in the absence of a temperature difference there is no signal (no interferogram in case of an FTIR measurement). The signal appears when $\Delta T = T_s - T_d > 0$ and it grows with increasing ΔT. (The sign of the temperature difference is not reflected in the single-beam spectrum, but the interferogram gets inverted when ΔT is negative.)

The temperature dependence of single-beam spectra originates from the temperature dependence of the Planck function (see Equation 18 and Figure 3), which is reflected also in the Stefan–Boltzmann law (Equation 10). It is anticipated that owing to this very strong temperature dependence of the intensity of single-beam emission spectra, the sample temperature must be strictly controlled and held constant during data acquisition.

Band intensities in double-beam emission spectra do not show temperature dependence (see Figure 8), unless chemical or structural modifications take place.[8] This follows from theory: when ratioing the single-beam emission spectrum of the sample to that of the blackbody (measured at the sample temperature), the factors depending on temperature will be eliminated. Nonetheless, any temperature mismatch between sample and reference (blackbody) will show up in the spectrum

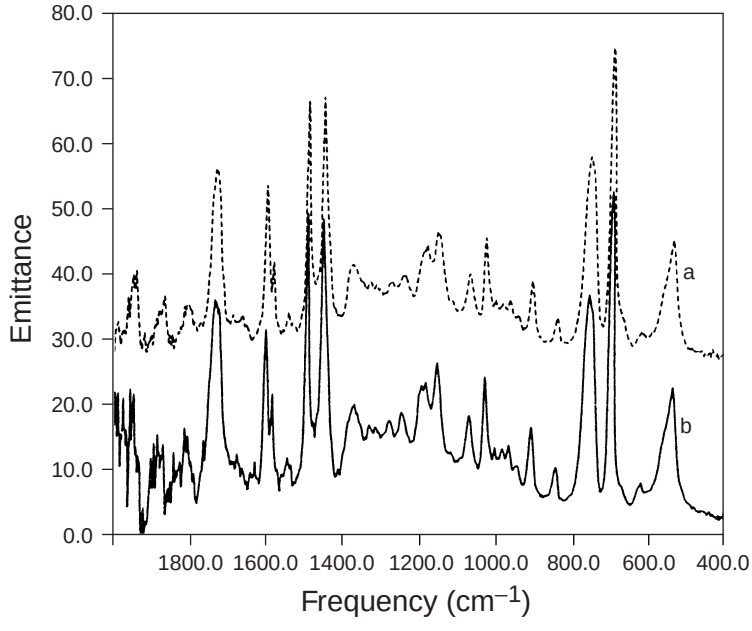

Figure 8 Emittance spectra of a polystyrene film: (a) at 40 °C and (b) at 10 °C (with the detector at 25 °C). (Reproduced by permission of the Society of Applied Spectroscopy from Chase, *Appl. Spectrosc.*, **35**, 77–81 (1981).[8])

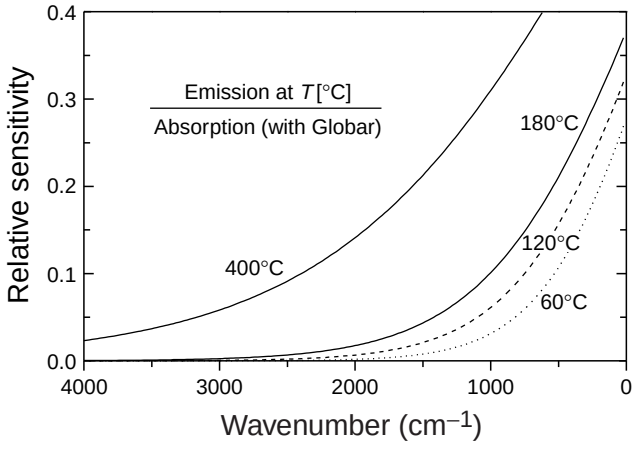

Figure 9 Relative sensitivity of FTIR emission measurements done with sample temperatures 60, 120, 180 and 400 °C compared to an absorption FTIR measurement.

in the form of curved backgrounds and altered relative band intensities.

Note also that an emittance spectrum measured at a higher temperature will have a higher S/N (meaning higher sensitivity) due to the greater amount of energy available for the two single-beam measurements involved. In this context it is interesting to compare the sensitivity of the IR emission measurement to that of IR absorption. In theory, the sensitivity of the two techniques (with the same instrument) depends on the temperatures of the emitters of the blackbody source used for emittance measurements and the Globar source used for absorption measurement.[12] The relative sensitivity curves shown in Figure 9 were obtained with Planck functions with a Globar source temperature taken as 1500 K and blackbody sample temperatures 60, 120, and 180 °C, commonly used in laboratory measurements. The curves clearly show the strong decrease in sensitivity of IRES toward higher wavenumbers, which manifests itself as increased noise level in the region above 3000 cm^{-1} (this is seen in Figure 6c). In order to have good S/N in this region, the sample temperature should be increased to about 400 °C.

2.4.4 Effect of Sample Thickness

The thickness of condensed-phase samples is a critical parameter in IRES when conventional methods of thermal excitation are used (samples in thermal equilibrium and no temperature gradient across the sample). Theory and practice have shown that the ideal sample thickness for IR emission spectroscopy is the same as for IR absorption spectroscopy, i.e. 1–5 µm. With the increase of sample thickness, the intensities of emission bands will increase and broaden gradually, so that spectral contrast gets worse and worse (see the lower

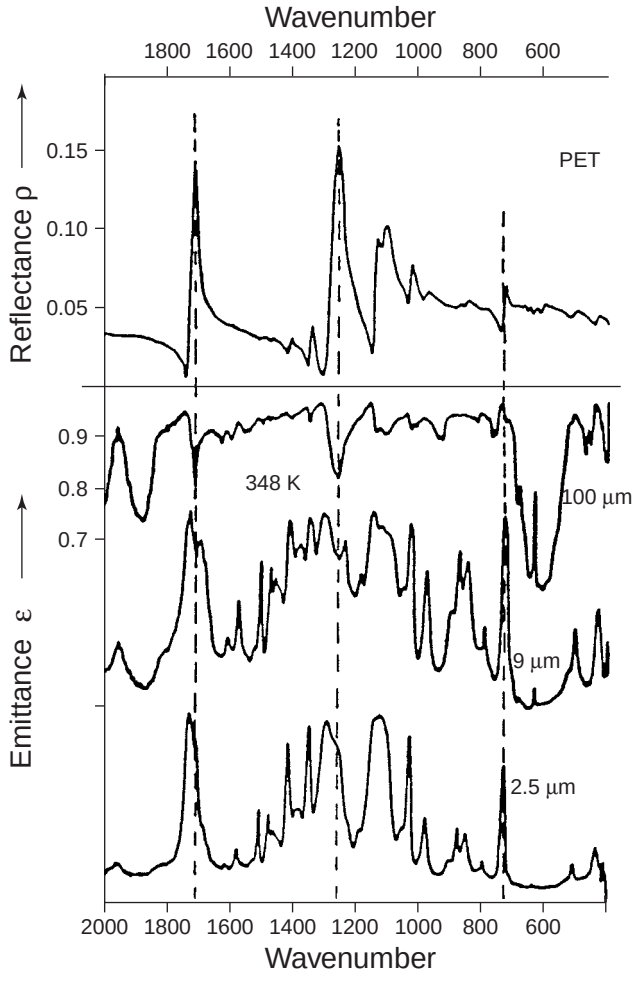

Figure 10 Emittance spectra of 2.5, 9 and 100 µm thick poly(ethylene terephthalate) (PET) films compared to the reflectance spectrum of PET. Note the correspondence between reduced emission of the 100 µm film and the reflectance features. (Reproduced by permission of the Society of Photo-Optical Instrumentation Engineers from Willis, Sheppard and Chalmers.[18])

three traces in Figure 10). Excessive thickness makes the sample opaque (nontransparent) when the single-beam emission spectrum of a semitransparent material becomes continuous like that of the blackbody. In this case the emission spectrum is no longer characteristic of the material and depends mainly on the temperature of the sample.

The quantitative relationship (Equation 40) between thickness and sample emittance is given by the emission spectroscopic equivalent of the Lambert–Beer law (valid if Equation 32 holds):

$$\varepsilon(\tilde{\nu}) = 1 - \tau(\tilde{\nu}) = 1 - 10^{-K(\tilde{\nu})d} \qquad (40)$$

where $K(\tilde{\nu})$ is the linear decadic absorption coefficient of the material and d is the layer thickness; for solutions or gas mixtures $K(\tilde{\nu})$ is replaced by $b(\tilde{\nu}) \times c$, where b stands

for the molar decadic absorption coefficient (usually also denoted by ε), and c is concentration. The connection to the absorbance of the sample, $A(\tilde{\nu})$, which is the linear intensity scale used in absorption spectrometry, is given by Equation (41):

$$A(\tilde{\nu}) = K(\tilde{\nu}) \times d = -\log[1 - \varepsilon(\tilde{\nu})] \tag{41}$$

2.4.5 *Problems Expected with Condensed-phase Samples*

An emission experiment, in fact, differs from the simple transmission–reflection model depicted in Figure 5(a) in that the emitted radiation originates within the sample and goes in every direction in space, but we observe only radiation emerging from a given direction within a certain solid angle, as shown in Figure 5(b). Several papers have dealt with theoretical models and suggested approximate expressions for the emittance of semitransparent condensed-phase samples under various sampling conditions.[13–17] Single and multiple layers of samples have been considered in different arrangements, e.g. between IR transmitting windows, between a reflecting metal back-plate and a window, or on a back-plate but without a window, etc. These treatments are based on ray tracing and energy conservation at all interfaces while considering the optical properties of the materials: internal transmittance and possible reflections at the interfaces.

Perhaps the most comprehensive treatment can be found in the publications by Hvistendahl et al.[15] and Rytter[16] who have also pointed out some errors in earlier publications. For a simple case that is often used in practice, an uncovered layer of liquid or solid sample on a perfectly reflecting back-plate support, the emittance can be expressed (Equation 42) through the internal transmittance of the sample (τ), and its reflectance at the front surface (ρ) as

$$\varepsilon = \frac{(1 - \rho)(1 - \tau^2)}{1 - \rho\tau^2} \tag{42}$$

which refers to double pass of the emitted radiation through the layer (compare with the basic Equation 30). For detailed results concerning multilayered samples or other methods of observation, the original papers should be consulted.[15,16]

2.4.6 *Reduced Emission Phenomena: Surface Reflectivity and Self-absorption*

In case of measurements of solids and liquids, irregular or distorted band shapes are often observed, especially for strong bands near their peaks. This is usually referred to as reduced emission phenomena.[17] In addition to reduced relative band intensities as compared to those observed in the absorption (transmittance) spectrum, some part of the band peak seems to be missing, leading to inversion of the peak and apparent splitting of the band. This can be seen on the spectra of the thick samples in Figure 10.[18] Two explanations have been put forward to explain this phenomenon: surface reflectivity and self-absorption.

It has been shown,[15,16] that surface reflectivity appearing in the expressions of sample emittance (e.g. in Equation 30) may be responsible for such distortions because of variations (sudden increase and decrease) of reflectivity in the vicinity of strong absorption bands, which cannot be neglected. These authors have used the damped harmonic oscillator model to simulate this phenomenon (see Figure 11a and b), and have shown that surface reflectivity leads to asymmetric distortion of strong emission bands.

A simple and effective practical method has been proposed to eliminate this kind of distortion:[15] to use, instead of a laboratory blackbody as reference, a sufficiently thick, opaque sample (for which $\tau = 0$) of the same material as the sample. The quantity obtained this way (Equation 43):

$$\varepsilon_s^* = \frac{L_s}{L_{op}} = \frac{\varepsilon_s}{\varepsilon_{op}} \tag{43}$$

is referred to as emittance with an opaque sample as reference. Since the thin and thick samples have approximately the same surface reflectivity, this procedure diminishes its influence and distortion-free emittance spectra can be obtained. (See an example in Figure 11c and d.)

Griffiths[12] attributed some reduced emission phenomena to self-absorption, i.e. to reabsorption of radiation originating in the bulk of the sample by the colder surface layer. This can also be verified theoretically using similar spectral simulations[10,16] as above leading to Equations (44) and (45):

$$\varepsilon(\tilde{\nu}) = (1 - \tau_b)\tau_s \tag{44}$$

or

$$\varepsilon(\tilde{\nu}) = (1 - \tau_b^2)\tau_s \tag{45}$$

corresponding to emitting samples without or with a reflecting back-plate, respectively, where τ_b and τ_s are transmittances of the bulk and surface layer, respectively, depending on the respective thicknesses (d) as Equation (46):

$$\tau(\tilde{\nu}) = 10^{-K(\tilde{\nu})d} \tag{46}$$

Calculations show (see Figure 12) that self-absorption can lead to band splitting for strong absorptions of thin layers and also for weaker absorptions of thick

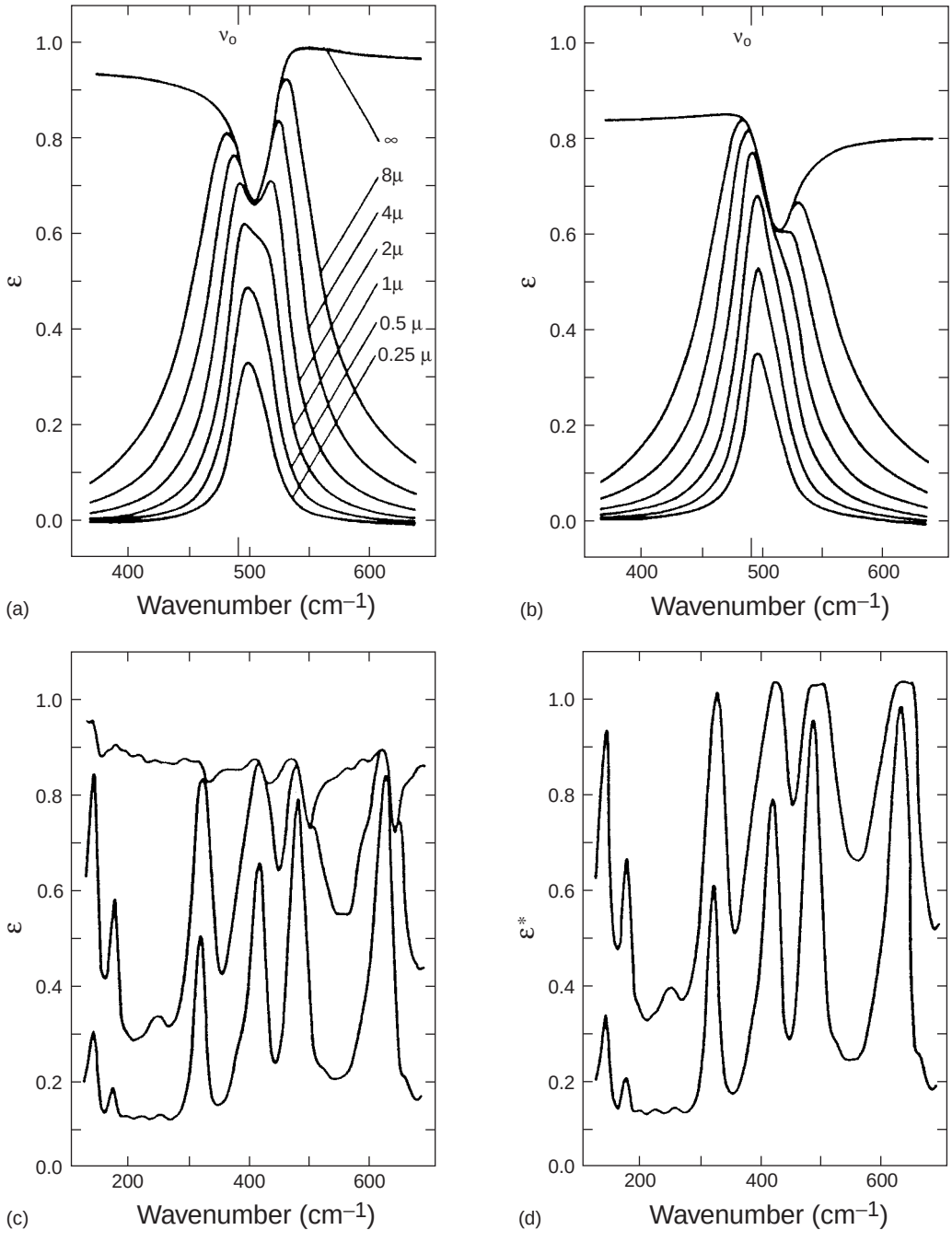

Figure 11 Effect of surface reflectivity on emission bands at increasing sample thickness: (a) Calculated band profiles for the case of reflecting backplate and sample without window, and (b) the same with a transparent window over the sample. (c) Emittance spectra of very thin (lower trace), intermediate (middle trace) and very thick, opaque (upper trace) Al_2Cl_6 samples measured at 200 °C with blackbody reference, and (d) emittance spectra of the same very thin (lower trace) and intermediate (upper trace) Al_2Cl_6 samples with the use of the opaque sample as reference. (Reproduced by permission of the Society for Applied Spectroscopy from Hvistendahl, Rytter, Oye, *Appl. Spectrosc.*, **37**, 182–187 (1983).[15])

layers, and this effect leads to symmetric splitting of strong absorption bands, which allows it to be distinguished from surface reflectivity.[16] Heating the sample from the front is expected to diminish self-absorption.

2.4.7 Sample Orientation and Metal Surface Effects

In the preceding discussion the emitted radiation has been assumed to be homogeneous and unpolarized under any viewing angle, which is correct for gas

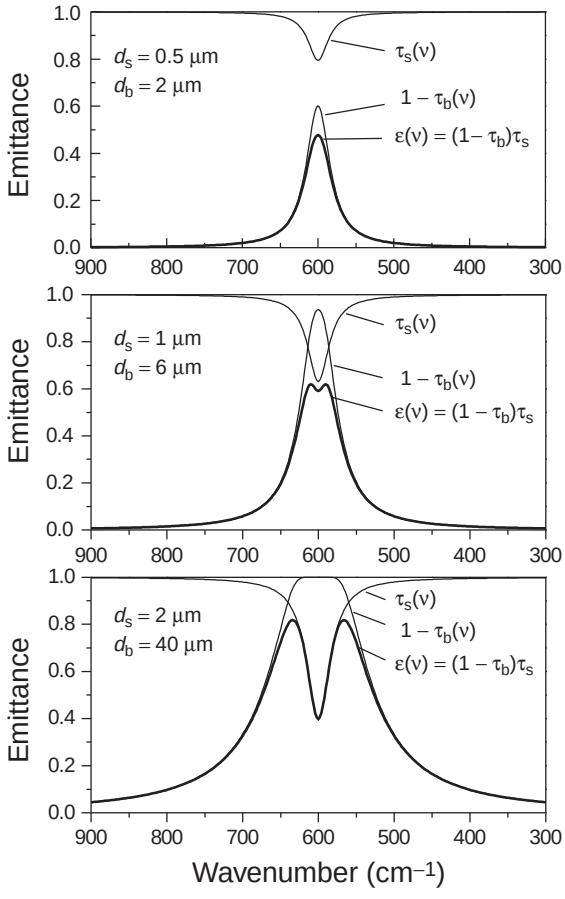

Figure 12 Simulation of the effect of self-absorption by a cooler surface layer (d_b and d_s are thicknesses of the emitting bulk and absorbing surface layers, respectively). (Reprinted with permission from Keresztury, Mink, Kristóf, *Anal. Chem.*, **67**(20), 3782–3787 (1995).[10] © 1995 American Chemical Society.)

phase work and for most condensed-phase samples. However, if the sample shows macroscopic order such that the molecules are oriented or aligned about a distinguished direction (which occurs, for instance, in oriented crystalline layers or stretched polymer films), the transition dipole moments associated with vibrational transitions also assume preferred directions in space. In this case the radiation is strongest in the plane perpendicular to the transition dipole, and is polarized in the plane defined by the propagation direction and the transition dipole.

Another practically important case is when the emitting molecules are located near a metal surface. When the distance between the surface and the molecules is very small compared to the wavelength of the emitted radiation (e.g. as in the case of adsorbed molecules), special selection rules apply, known as metal surface selection rules. Greenler[19] has calculated the angular intensity distribution of light emitted by radiating dipoles that

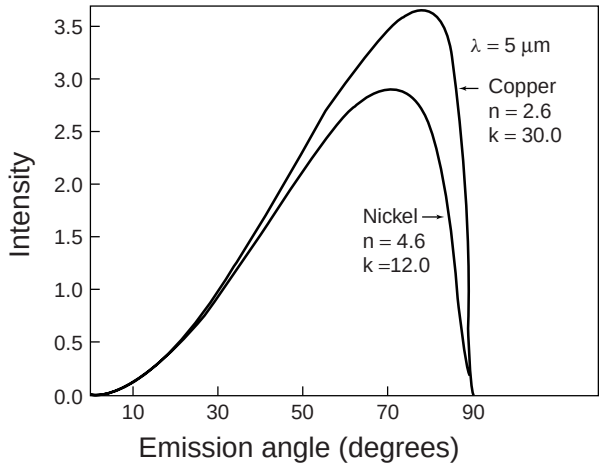

Figure 13 Calculated angular intensity distribution of emission by a dipole oriented perpendicularly to the metal surface (the emission angle is measured from the normal to the surface). (Reprinted from Greenler, 'Light Emitted from Molecules Adsorbed on a Metal Surface', *Surface Science*, **69**, 647–652 © (1977) with permission from Elsevier Science.[19])

are parallel and perpendicular to the metal surface. His results indicate that

1. one should not expect to see IR emission from molecular vibrations whose transition moments are parallel to the surface;
2. for vibrations with transition moments perpendicular to the metal surface, the best viewing angle is 70° to 80° from the surface normal and the radiation is polarized parallel to the plane of incidence, while in the direction perpendicular to the metal surface the emission intensity is zero (see Figure 13).

These selection rules help in determining the orientation of molecules on the surface and have to be taken into account when setting up such an experiment. For instance, part of the unpolarized background radiation can be eliminated and the S/N can be substantially increased by inserting an IR polarizer into the beam or by using a polarization-modulation technique.[20]

3 INSTRUMENTATION: SPECTROMETERS AND DETECTORS

Laboratory measurements in IRES can be performed with commercial IR spectrometers designed for absorption measurements to be done in the transmission regime. Some of these instruments have a provision to switch over to emission mode, or at least have an emission port on the instrument housing that can accept an auxiliary emission accessory.

Commercial IR spectrometers can be divided into two broad classes: dispersive (prism or grating) instruments and FT (or interferometric) instruments. Dispersive instruments dominated IR spectroscopy from the early days of its history until the end of the 1970s, when FTIR instruments started proliferating at a great pace, together with fast digital data handling facilities.

Both dispersive and FTIR instruments are mechanical-optical devices that have the following main components:

- incandescent IR radiation source (switched off or replaced by the sample in IRES);
- a dispersion element (in the dispersive instruments) or an interferometer (in the FTIR instruments);
- an IR detector (with amplifier electronics);
- a device for recording (plotting or storing) the measured data.

The first three components are connected by optical elements (mostly mirrors) that direct, collimate or focus the light as appropriate. The radiation source and the detector are in the two terminal focal points of the spectrometer optics. The peculiarity of emission measurements is that the sample itself serves as the source of radiation; thus the sample replaces the usual light source or it is put at an optically equivalent location inside or outside the spectrometer cover (by an auxiliary port), and the conventional sample compartment remains empty.

3.1 Dispersive Instruments

Conceptually, the dispersive instruments are much simpler, since they measure the spectrum directly. The IR beam to be analyzed is dispersed by a prism or more commonly by a diffraction grating to spatially separate the different wavenumber components, and the optical image of the spectrum is scanned through with a narrow slit permitting only energy from a narrow spectral interval to fall on the detector. During the scan, the intensity of the radiation is recorded as a function of wavenumber.

To suit the requirements of the much more commonly used transmission spectrometry, most dispersive IR spectrometers are built as double-beam instruments with optical-null or ratio-recording detection[21] (shown schematically in Figure 14). Their working principle is as

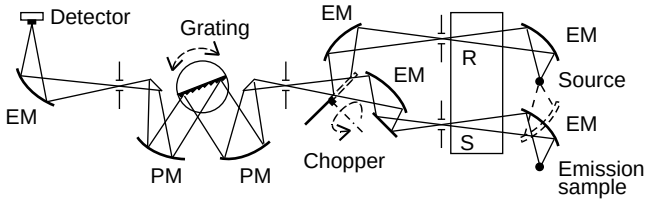

Figure 14 Optical scheme of a double-beam dispersive IR spectrometer. (PM, paraboloid mirror; EM, ellipsoid mirror.)

follows: the radiation emitted by the IR source is divided by symmetrically arranged mirrors into two equivalent beams, the sample beam and the reference beam, both passing through the sample compartment next to each other. The beams are focused on the entrance slits of the monochromator and are combined by means of a rotating chopper into one common path, letting the alternating sample and reference beams fall on the dispersing element, then through the exit slit of the monochromator on the detector. The spectrum is recorded in real time (during scanning) as the ratio of the sample beam intensity (I) to the reference beam intensity (I_0).

In a transmission measurement the emitter is an incandescent light source operating at above 1000 K (such as a Globar or a nichrome filament); from its radiation the sample is supposed to absorb more energy than the reference, thus the recorded intensity ratio gives the transmittance of the sample relative to that of the reference: $\tau(\tilde{\nu}) = I(\tilde{\nu})/I_0(\tilde{\nu})$.

In contrast, in an emission measurement the standard light source of the instrument is switched off or removed and the emitting sample and a reference emitter (a laboratory blackbody) are put in its place or at optically equivalent positions. Usually some modification of the source optics is needed to accommodate both emitters in the focus of the respective paths on a joint, heated (temperature controlled) sample stage. In principle, both emitters can be put in the sample compartment as well, but then special light-gathering optics are needed to increase the throughput of the emitted radiation to the detector. Various double-beam or semi-double-beam emission source arrangements have been discussed in Huong's review.[4] The recorded intensity ratio, in this case, directly gives the emittance of the sample, $\varepsilon_s(\tilde{\nu})$, i.e. its emission intensity relative to that of the blackbody at the same temperature in accordance with Equation (26).

Although the emittance spectrum can be directly measured with a double-beam grating spectrometer, these instruments are rarely used today in emission spectroscopy, partly because of the overwhelming technical advantages of FTIR instruments for detection of weak signals, and also because two costly emission attachments are needed.

3.2 Fourier Transform Infrared Instruments

FTIR spectrometers measure single-beam spectra (or rather interferograms corresponding to single-beam spectra) directly, thus the method by which emittance spectra are obtained is quite different from the way they are measured on a double-beam grating spectrometer. The usual procedure of measuring $\varepsilon_s(\tilde{\nu})$ on an FTIR spectrometer is to measure the reference (e.g. blackbody) interferogram first, store the

measured data in the memory of the computer and calculate the single-beam reference spectrum; the sample is then inserted in the emission accessory and the same operations are done with the sample to obtain its single-beam spectrum. Finally, the desired double-beam spectrum (the emittance spectrum) is calculated as a point-by-point ratio of the two single-beam spectra.

The schematic optical layout of an FTIR spectrometer equipped with an auxiliary emission accessory is shown in Figure 15. The heart of an FTIR spectrometer is the Michelson interferometer consisting of a beamsplitter and two mutually perpendicular mirrors, one fixed and the other moving back and forth at a constant velocity. The collimated IR radiation enters the interferometer and falls on the beamsplitter (a semitransparent, semireflecting material) at 45° where it is divided into two equal parts and travels towards the two mirrors. The reflected beams returning from the fixed and moving mirrors with changing path difference are combined together at the beamsplitter where they interfere with one another. This interference causes variations in the instantaneous intensity of the beam that is focused on the detector. The measured intensity, $I(\delta)$, of the modulated beam as a function of optical path difference, δ, in the two arms of the interferometer is called the interferogram; it contains multiplex information from all spectral elements in each data point. The interferogram, $I(\delta)$, and the single-beam spectrum, $L(\tilde{\nu})$, are FT pairs, thus the spectrum is obtained from the digitized interferogram by Fourier transformation (Equation 47):

$$L(\tilde{\nu}) = \int_{-\infty}^{\infty} I(\delta) \times \cos 2\pi\tilde{\nu}\delta \times \mathrm{d}\delta \qquad (47)$$

For a most comprehensive description of the theory and a wide range of practical aspects of FTIR spectrometry the reader is referred to the monograph written by Griffiths and deHaseth.[12]

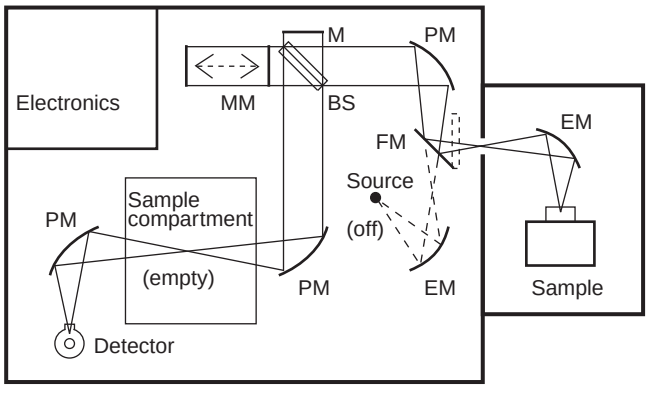

Figure 15 Optical scheme of an FTIR spectrometer equipped with an auxiliary emission compartment. (PM, paraboloid mirror; MM, moving mirror; M, fixed mirror; EM, ellipsoid mirror; BS, beamsplitter.)

With an FTIR instrument, other optical arrangements are also possible, e.g. the radiating sample could replace the detector when an auxiliary detector is used, but this is rarely a practical solution. For versatility it is essential that the instrument be equipped with an emission port easily accessible through a flip mirror to accept light from an auxiliary emission attachment.

3.1.1 Advantages of the Fourier Transform Infrared Technique

For all the complexity of its operation principle, the interferometric measurement technique offers several advantages compared with conventional dispersion IR spectroscopy.

Higher Signal-to-noise Ratio Under conditions of equal measurement time, FTIR spectra show higher S/N. This is the consequence of two factors:

1. the higher optical throughput of the interferometer than that of a grating monochromator (a factor of 50 to 200, due to elimination of the narrow optical slits);
2. the so-called multiplex advantage arising from the fact that in FTIR all of the spectral elements are observed for the whole measurement time, while in a grating instrument the measurement time has to be divided by the number of resolution elements in the whole spectrum. For the whole IR region at a resolution of 4 to $1\,\mathrm{cm}^{-1}$ this gives a gain factor of 30 to 60 in S/N, and even more at higher resolutions. Higher S/N means higher sensitivity, which is essential in IRES measurements, especially when the temperature difference between sample and detector is relatively small and low radiation levels have to be measured.

Greater Frequency Accuracy The interferometers of FTIR instruments have a built-in visible laser reference to control the mirror movement, which helps determine the wavenumber scale of the spectrum with very high accuracy (no additional wavenumber calibration is needed except for high-resolution spectra).

Computerized Data Handling Computerized data handling is not limited to FTIR systems but because of the need to store the measured data in digital form and to perform the Fourier transformation and other necessary calculations before the spectrum is obtained, FTIR spectrometers are ab ovo equipped with dedicated microcomputers or are connected to PCs and work under computer control. The instrument manufacturers supply

proprietary software with the instrument for data acquisition and analysis, which may incorporate interactive tutorials and integrated diagnostics packages as well.

Based on the above advantages, FTIR spectrometers as of today are preferred to dispersive machines in most IRES applications. There are more than half a dozen well-known instrument manufacturers in the market offering good-performance FTIR instruments in a broad price range (from $30 000 to $150 000) that are meant basically for transmission work. The cheaper models may be adequate for routine laboratory measurements to obtain qualitative data at low spectral resolution (max. 1 cm^{-1}), and are usually limited to the 4000 to 400 cm^{-1} spectral range. The more expensive ones are supposed to have a higher-quality interferometer (featuring dynamic alignment, possible step-scan operation) suitable for higher-resolution work (at least 0.1 cm^{-1}), and be more versatile, offering a choice of built-in interchangeable components, e.g. different detectors, beamsplitters and sources to cover the entire spectrum from the FIR to the NIR. While faster scanning instruments with high optical throughput are always preferred for IRES, high resolution is needed only for gas-phase work, for sure identification of unexpected species with resolvable rotational fine structure. Lower-resolution spectra with higher S/N are more useful in most quantitative applications than highly resolved but noisy spectra.

3.3 Open-path Fourier Transform Infrared Instruments

One of the natural applications of IRES, remote sensing, stems from its single-ended measurement geometry allowing the observation of spectra of thermal emitters located at a distance from the instrument. Since atmospheric or environmental studies of emission sources most often involve field measurements, the instruments used must be mobile, i.e. mounted on a vehicle, or be compact and lightweight to be carried more easily. For this purpose dedicated open-path Fourier transform infrared (OP/FTIR) spectrometers are used that are equipped with a telescope to collect the radiation and matched transfer optics to pass a collimated beam into the interferometer.

One of the most popular designs consists of a 25-cm Newtonian telescope and a standard Michelson interferometer with a liquid-nitrogen-cooled MCT detector. A simplified optical diagram of such an instrument is shown in Figure 16. The spectrometer is mounted on the top of the telescope barrel making a very compact structure that can be set up on a tripod stand to point it conveniently toward the remote source. The instrument can run on battery power and may be controlled by a portable laptop computer in remote locations. Other OP/FTIR spectrometers may incorporate a Cassegrain-type telescope, a more

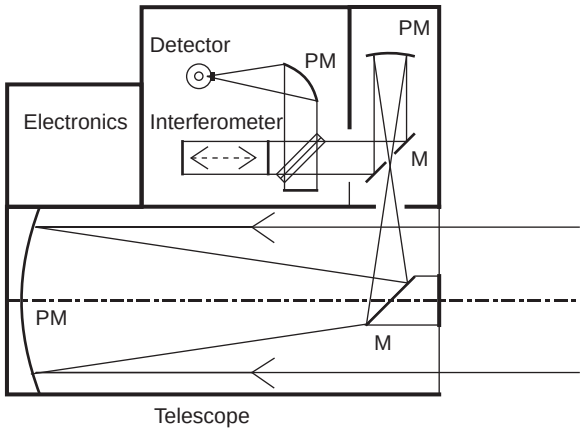

Figure 16 Optical scheme of an open-path FTIR spectrometer with a Newtonian receiver telescope. (PM, paraboloid mirror; M, mirror.)

sensitive liquid-helium-cooled detector (especially for the FIR region, in high altitude and extraterrestrial measurements), and provision for measurement of blackbody reference spectra or calibration standards for quantitative analysis. With the addition of a portable artificial IR source (equipped with a source telescope) or a remote cube-corner array retro-reflector, these instruments can also work in transmission mode. A famous early mobile remote sensing FTIR instrument built for the United States Environmental Protection Agency (USEPA) in the 1970s around a commercial, moderately high resolution FTIR system configured to fit into a van, the so-called Remote Optical Sensing of Emissions (ROSE) system, has been described together with some of the measurement results.[22]

Obtaining the appropriate blackbody reference spectrum may pose some problem, since in the case of remote natural sources of radiation the temperature of the source may not a priori be known. Knowledge of the instrument response in this case may prove useful, since it allows easy simulation of single-beam blackbody reference spectra from Planck functions for various temperatures (see section 2.4). After dividing the sample spectrum by the trial reference spectra, the one giving a double-beam (emittance) spectrum with the flattest background will correspond to the right temperature.

A serious limiting factor of the applicability of IRES in environmental or remote atmospheric measurements (e.g. the measurement of the plume of a power plant smokestack or hot gases from combustion jets) is atmospheric absorption. Since there is a great absorbing pathlength (typically from several meters to about a kilometer) between the remote source and the spectrometer, atmospheric water vapor and CO_2 may completely block certain spectral regions, prohibiting the detection of some pollutants.

3.4 Choice of Detectors

The description of IR detectors is beyond the scope of this article, but one should be aware of how an emission measurement is affected by the type of detector used, and choose the detector or treat the data accordingly.

There are two basic types of IR detectors: room temperature pyroelectric detectors and cooled quantum detectors. To a first approximation a detector can be characterized by its operating range and its specific detectivity, D^*,[12] which depends on wavenumber. Broad-range detectors that can be used in the whole IR region are less sensitive than narrow-range detectors. The most commonly used room-temperature detectors are triglycine sulfate (TGS) and DTGS, the latter being several times more sensitive in the whole IR region. The most popular cooled, broad-range detector is the photoconductive MCT detector (MCT, HgCdTe) operating at liquid-nitrogen temperature (77 K). The MCT detector is by more than an order of magnitude more sensitive than DTGS and its response to the changing signal is also incomparably faster, allowing the measurement rate to be increased to 5–10 interferogram scans per second (depending on resolution). Higher sensitivity detectors are available for narrower spectral regions (some may require cooling below 20 K with liquid helium); for example the InSb detector has nearly unit quantum efficiency but it does not work below 1800 cm^{-1}.

Based on the above properties, the cooled MCT detector could be preferred in IRES to compensate for the usually low signal level. In addition, heat radiation leaving the detector at 77 K is negligible, which means that in this way practically the full sample emission is detected (see Equations 35 and 36 and the discussion in section 2.4), even if it is not heated. One should not forget, however, that an MCT or other low-temperature detector will detect, in addition to sample emission, thermal radiation of all room-temperature objects in its field of view, including parts of the instrument itself, (e.g. the beamsplitter).[11] Thus, with the choice of an MCT detector comes the obligation of making a correction for background radiation with the extra measurements involved, according to Equation (38).

In contrast, with the use of a DTGS detector, the sample must be kept at a temperature higher (or lower) than room temperature to get appreciable radiation flux. But although the single-beam spectrum may not be as intense (the S/N may not be as high) as with an MCT detector, the spectra obtained are usually free from spurious bands. For this reason, a room-temperature DTGS detector may be a better choice for routine IRES applications.

3.5 Custom-built Instruments for Special Applications

While commercial instruments with matched sample optics are suitable for most IRES experiments, important research fields such as the study of adsorbed species on metal single crystal surfaces or IR chemiluminescence have more stringent requirements owing to the extremely low emitted signal levels. Because of the latter, liquid-helium-cooled detectors have to be used when the sample is kept at room temperature, but then precautions must be taken to minimize background radiation.

Durana et al.[23] have solved this problem by utilizing a cryogenically cooled FTIR spectrometer, by immersing a Michelson interferometer designed to operate at 77 K into a cold helium-purged chamber and cooling the whole optical collection system to 77 K. The equipment shown schematically in Figure 17 was equipped with a liquid-helium-cooled photoconductive Ge/Cu detector and with a gold/KRS-5 wire grid polarizer inserted into the collection beam. The samples (e.g. molecular monolayers adsorbed on metal single crystal surfaces) were maintained at a temperature between 80 and 500 ± 1 K and tilted to give an angle of incidence of 50–55° to approach the optimum

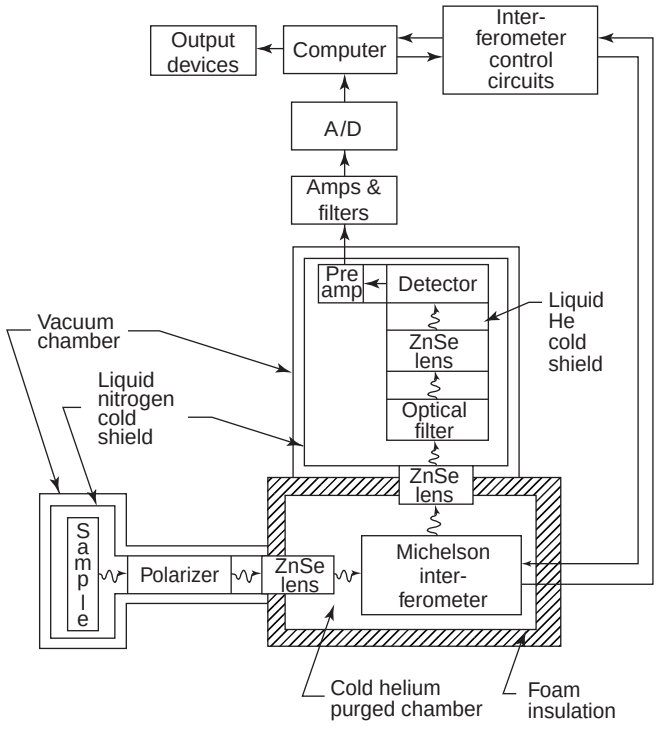

Figure 17 Experimental schematic of a cryogenically cooled FTIR spectrometer that can be used to study chemiluminescence. (Reproduced from Allara, Teicher, Durana, 'Fourier Transform Infrared Emission Spectrum of a Molecular Monolayer at 300 K', *Chem. Phys. Lett.*, **84**(1), 20–24 © (1981) with permission from Elsevier Science.)

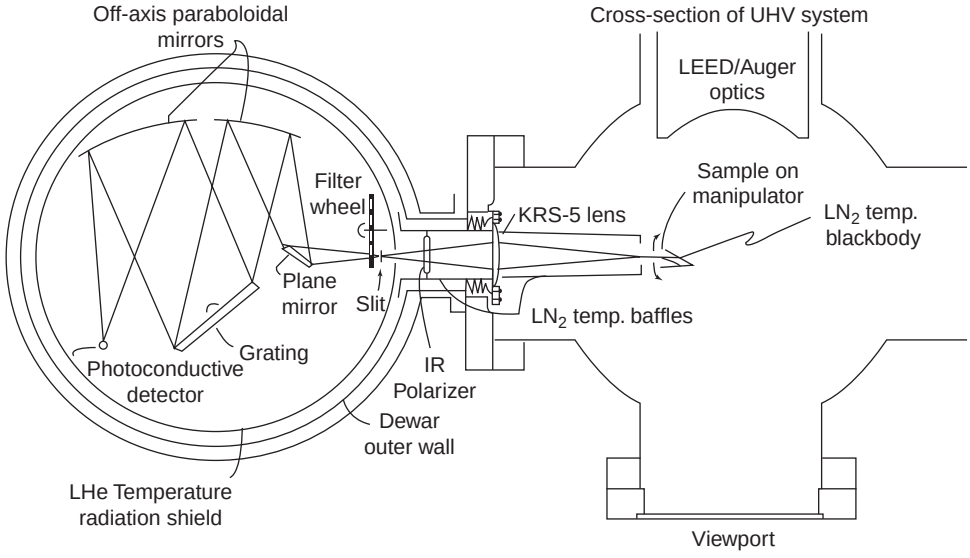

Figure 18 Schematic of a cryogenically cooled grating spectrometer used for studying adsorption on single crystal surfaces. (Reprinted with permission from S. Chiang, R.G. Tobin, P.L. Richards, *J. Vac. Sci. Technol. A*, **2**(2), 1069–1074 © 1984 American Vacuum Society.[25])

experimental conditions predicted by Greenler[19] (see section 2.4). A similar type of cryogenic interferometer system has been used previously to carry out high-sensitivity gas-phase studies of IR chemiluminescence.[24] The working ranges of these instruments were 2000–1250 and 2000–720 cm^{-1}.

Another cryogenic spectrometer based on Czerny–Turner-type grating spectrometer (see Figure 18) was described by Richards et al.[25] and used successfully in a series of studies of weak low-frequency adsorbate–substrate vibrations on clean, metal single crystal samples. In this case the whole spectrometer was cooled with liquid helium to 4.2 K, and was connected to an ultrahigh vacuum system sharing the sample compartment with other surface analysis methods. One advantage of this instrument over the cooled FTIR systems introduced above is the broader useful spectral range extending down to 330 cm^{-1}. The authors claim that the S/N attainable with this instrument in this particular application is comparable to that of an FTIR instrument, and the grating instrument is easier to cool. Anyhow, one tends to share the opinion[8] that this instrumental modification is really not viable for routine analytical applications.

4 LABORATORY METHODS

4.1 Conventional Methods of Thermal Excitation

As it has been mentioned above, IR detectors respond to any deviation from radiation balance between sample and detector. This means that the sample (or more precisely, the portion of sample in the field of view of the detector) must be either at a higher or at a lower temperature than the detector.

In conventional thermal emission spectroscopy the samples are heated to above room temperature and are kept at constant temperature (in thermal equilibrium) during the measurement. Various types of high-temperature or variable-temperature cells have been designed for IR transmission work,[26] and in principle all these can also be used in IRES. Genuine emission cells, however, have a metal body (made of highly reflecting, low-emissivity material) and are backed with a reflector plate which, in addition to blocking the outside radiation, brings a twofold gain in emission intensity. The shiny metal backplate (polished stainless steel or gold-covered copper block) often serves also as a heater for liquid and solid films. Temperature feedback to the controller is provided by a thermocouple mounted in a hole drilled in the heater block.

4.1.1 Methods for Gases

The measurement of IR emission spectra of hot gases poses little problem experimentally. The temperature in this case may be high enough to produce intense spectral features, and the bands show rotational structure consisting of series of very narrow lines that are easy to identify using high resolution (at least 0.1 cm^{-1}). Also, there are no surface effects that could distort the spectral features. The pathlength requirements of the gas samples are not so stringent either as with condensed-phase

samples: pathlengths of a few millimeters to hundreds of meters or even kilometers (in case of open-path instruments) may be acceptable depending on the partial pressure of the gas.

Laboratory measurements of emission spectra of gases often serve for calibration purposes, as reference for remote sensing by IRES. These calibration spectra must be measured at known (controlled) pressure, temperature and pathlength, but may be obtained either in emission or in transmission mode, then converted to absorbance scale. For IRES measurements, the standard single-pass or multipass transmission gas cells[26] surrounded with a cylindrical oven can be used (e.g. a 10-cm cylindrical Pyrex cell with KBr windows), although in this case higher background emission is expected than with a metal tube surface. Limiting the field of view with cold (water-cooled) apertures may help improve the S/N.

The potential of FTIR emission spectroscopy for the study of gas-phase molecules has been reviewed recently by Bernath.[27]

4.1.2 Methods for Liquids

Again, conventional heated liquid transmission cells can be easily converted to emission cells by replacing one of the two parallel windows by the heatable reflecting back plate. As a horizontal liquid sample holder, flat-bottomed shallow metal cups can also be used, in good thermal contact with a heater stage (Figure 19). In horizontal liquid sample holders the top window may be omitted, although a windowed cell has some advantages:[16] the sample may be compressed to a thin film between the window and the support, evaporation is minimized and a uniform sample temperature is assured with a window of high thermal conductivity, which may also minimize cooling of the surface and occurrence of self-absorption.

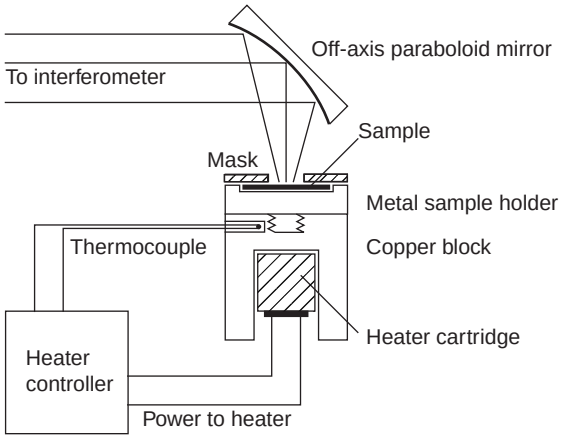

Figure 19 Cross-sectional view of a typical emission accessory for solids and liquids with exchangeable sample holder and collection optics.

4.1.3 Methods for Solids

The most widely proliferated method is to put the sample in direct contact with an electrically heated metal block when the sample is heated from the reverse side (not facing the spectrometer). A circular, heatable metal stage of about 10–15 mm diameter may serve as a good solid sample holder, e.g. in an emission accessory shown in Figure 19. This configuration has the disadvantage that the front surface of the sample may be somewhat cooler than the bulk and reabsorption of the emitted radiation may occur, leading to distortions and splitting of emission bands as described earlier.

Horizontal sample stages are more practical in case of some solid samples, as well: they hold polycrystalline or powdered samples more easily than vertical or tilted sample holders. Good adhesion to the support for better heat transfer may be facilitated by depositing the sample on the support from solution or suspension by slow evaporation of the solvent. The optimum amount of sample to deposit is around $0.1\,\mathrm{mg\,cm^{-2}}$. Unless the sample is too thick, a useful emission spectrum can be obtained from uneven layers or chunky materials as well, but the spectral contrast may deteriorate considerably compared to the spectra of even and smooth emitting layers. For polymer films or other smooth layers of neat materials the ideal sample thickness is in the range of 1–5 μm.

Microreactor-type sample cells have also been built, for example for in situ studies of heterogeneous catalytic reactions.[28] They are based on quartz, Pyrex glass or stainless steel chambers that withstand high temperatures, can be evacuated or placed under pressure, and purged with gases as necessary and are equipped with the usual heating and temperature controlling facilities. A typical catalytic microreactor design is shown in Figure 20.

4.1.4 Samples at Ambient or Low Temperatures

There are several ways to avoid cooling of the front surface. Naturally, it should not occur with the use of a

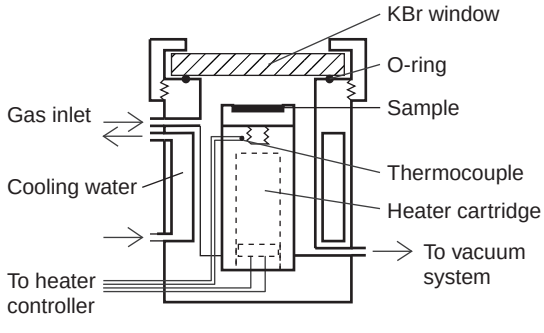

Figure 20 Cross-section of a catalytic microreactor-type emission accessory.

cooled detector if the sample is kept at ambient temperature, or with the use of a room-temperature detector if the sample is cooled. Under these circumstances, however, the measurement is practically restricted to the spectral region below 2000 cm^{-1} because the maximum of available energy will be that of a 300 K blackbody (see Figure 3), and the signal, if any, would hardly exceed noise level at higher frequencies.

4.1.5 Laboratory Blackbody Sources

There is a great variety of so-called laboratory blackbody emission sources used to provide a reference or calibration standard for thermal emission measurements. A heated black-walled cavity with an aperture makes a good approximation to an ideal thermal emitter (with emittance close to 0.99), and commercial blackbody calibration sources are of this type. These devices have precise temperature control (some of them up to 3000 K) but since they are usually too large to fit into most emission accessories designed for microsamples, they are used mostly in gas-phase work and remote sensing.

With condensed-phase samples, it is desirable also that exactly the same optical arrangement be used with the same heater and temperature controller settings in both sample and reference measurements. Thus, perhaps the most convenient choice in practice is to use a metal plate covered with a thick layer of heat-resistant matt-black paint; but there are some other suitable black materials such as a sooted metal disc or aluminum foil, a graphite disc or a layer of carbon black on the standard sample stage. One should keep in mind, however, that the emittance of these laboratory blackbodies may not exceed 0.9.

4.1.6 Emission Accessories

An emission accessory consists of a controlled-temperature sample cell placed in the focus of a light collection optics (normally an off-axis ellipsoid or off-axis paraboloid mirror) that sends the radiation into the spectrometer. Ideally, the cell is mounted on a three-dimensional (3-D) translation stage to facilitate sample alignment. Solid sample holders may require a tilting facility. A sketch of a typical emission accessory is shown in Figure 11. A version of this accessory incorporates a microscope mounted in the collecting mirror to help locate the area of interest on the sample surface.[29] Further modifications have been described to improve the sample temperature stability to permit the acquisition of quantitative IR emission data.[30]

4.2 Advanced Methods for Solids

The greatest challenge of IRES is to obtain meaningful spectra (having vibrational structure) of thick, opaque solid samples. We have seen that conventional thermal excitation by heating the sample from the rear is unable to solve this problem. However, heating of the sample can be accomplished by various means from the front, as well. In addition, by simultaneous heating and cooling of different points of the sample, temperature gradients can be maintained or generated temporarily or periodically along the thickness of the sample, which creates favorable conditions for observation of emission from a thin surface layer. Thus, irradiation of the front surface with laser beams, both visible and NIR, pulsed or continuous wave, and also blowing the surface with a stream of hot air have been tested.

4.2.1 Laser-induced Thermal Emission

One of the recent developments is laser-induced thermal emission spectroscopy (LITES).[31,32] Results of the preliminary investigations in LITES have shown that it is a feasible approach because heat generation via absorption of laser light by the sample is more or less localized on its surface, regardless of the thickness, shape and composition of the sample or the support material. This means that sample thickness is no longer a limiting factor, i.e. emission spectra of the surface layers of thick samples or objects of any dimensions can be measured. However, the efficiency of this method depends on the absorptivity of the sample at the laser wavelength. Thus, for some samples (e.g. red paint) the 514.5 nm green line of an Ar-laser may be quite efficient[32] while in general a NIR neodymium yttrium aluminum garnet (Nd/YAG) laser is thought to be a better choice.[31]

4.2.2 Methods of Transient Heating

Excitation of thermal emission of the surface layer by a continuous-wave laser is not without difficulties because owing to thermal diffusion the thickness of the heated emitting layer increases with time, which causes a gradual deterioration of spectral contrast. This effect can be reduced by various regimes of transient or modulated heating. Fully successful LITES experiments have been described[33] with the use of the experimental set-up shown in Figure 21 where heat build-up in the sample is prevented with a cooling gas jet directed next to the heated spot on a rotating sample. Under such conditions a transiently heated thin surface layer is generated in the sample, and it is only the emission of this surface layer that is detected. If instead of a continuous laser a pulsed laser is used in conjunction with a rotating sample, the cooling gas jet can be omitted.[34]

A much more economical but similarly successful version of the transient-heating method uses a hot-gas jet in place of a laser to heat the surface of the sample (Figure 22).[35] This heat source combined with a rotating

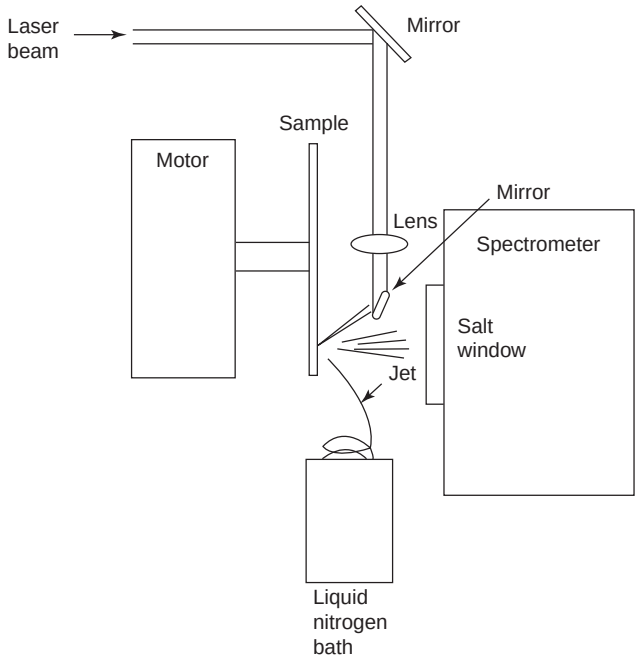

Figure 21 Experimental arrangement used for observing transient infrared emission spectroscopy (TIRES) spectra with laser excitation. (Reprinted with permission from Jones, McClelland, *Anal. Chem.*, **61**, 650–656 (1989).[33] © 1989 American Chemical Society.)

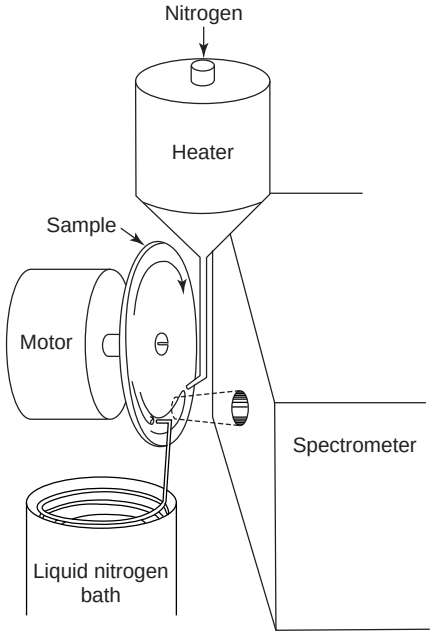

Figure 22 Experimental set-up used for observing TIRES spectra with hot gas jet excitation. (Reprinted with permission from Jones, McClelland, *Anal. Chem.*, **62**, 2074–2079 (1990).[35] © 1990 American Chemical Society.)

sample and cooling gas jet allows the study of materials that do not absorb strongly at common laser wavelengths.

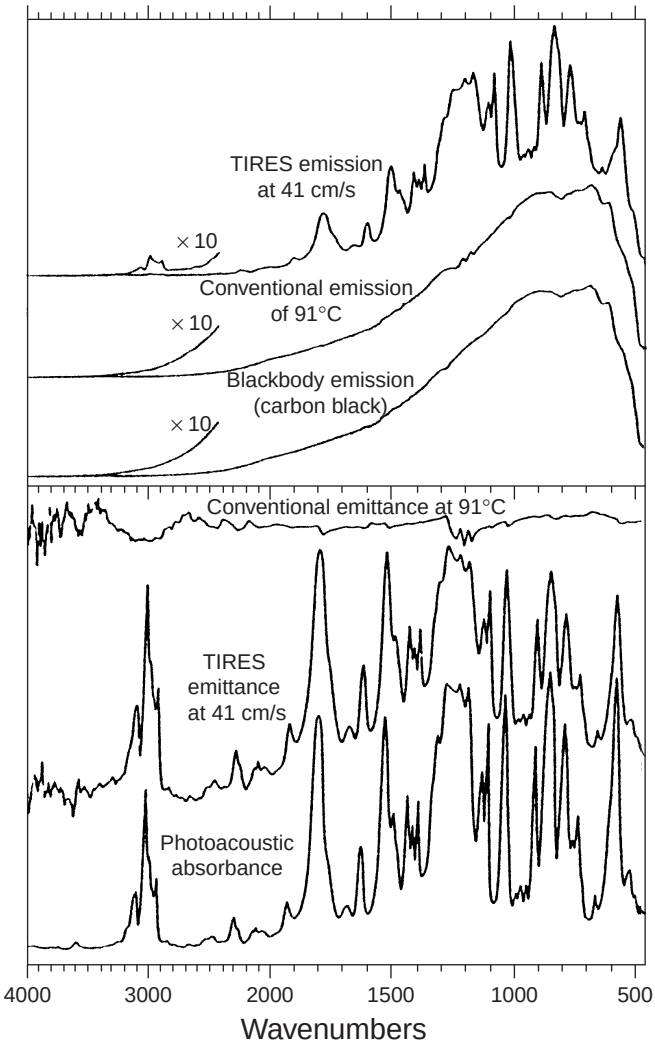

Figure 23 Comparison of the performance of the TIRES method with that of conventional IRES on an opaque polycarbonate sample. Upper part: single-beam emission spectra; lower part: double-beam emittance spectra compared to photoacoustic absorbance spectra. (Reprinted with permission from Jones, McClelland, *Anal. Chem.*, **62**, 2074–2079 (1990).[35] © 1990 American Chemical Society.)

With these methods, termed TIRES, analytically useful IR emission spectra can be obtained from optically thick samples as well (see Figure 23).

4.2.3 Modulated Heating

Another new method devised for the study of high-opacity, refractory materials is named modulated emission spectroscopy (MES).[36] It uses an argon ion laser as a pump source modulated by a wheel chopper, while the sample sits on a metallic rod immersed in liquid nitrogen. Note that modulated heating (as with pulsed lasers) requires more complex and costly electronics ensuring

the synchronization of signal sampling with the frequency of modulation or laser firing.

The main advantage of the methods outlined above (LITES, TIRES, MES) is the ability to measure emission spectra of thick samples. It is also possible to perform depth probing by changing the heating-and-cooling regime and hence the thermal diffusion length in the sample.

Beside the indisputable advantages of TIRES, it should be mentioned that some uncertainty arises when converting the single-beam emission spectra to the double-beam emittance scale. The problem is that the heating regimes are not exactly identical between different samples or between sample and blackbody reference. The temperature of the emitting layer is not exactly known and hard to reproduce, thus the comparison of emission intensities and the quantitative evaluation of spectra obtained with the TIRES method is not so straightforward.

4.2.4 Transient Cooling and Emission–Absorption Spectra

Self-absorption is usually considered an unwanted phenomenon. There are instances, however, such as multilayer samples or formation of surface species, when weak negative spectral features appear on the background of continuous, blackbody-like emission of the bulk at characteristic frequencies of some surface species. It has been recognized that these features are not always due to surface reflection, but may be due to appearance of absorption bands on the emission continuum of the bulk.[5] In such cases reabsorption of the bulk radiation by the surface species may be a chance to detect surface species on opaque substrates. If the specific application makes it desirable to exploit reabsorption, it can be fostered by cooling the front surface of the hot sample by a stream of cold air or inert gas. This method of measurement, first reported by Jones and McClelland[37] and later augmented with unsteady-state heat-flow calculations and harmonic oscillator modeling,[38] has been dubbed transient infrared transmission spectroscopy (TIRTS) or infrared emission with transient cooling (IRE/TC). It can be considered as the opposite of transient heating TIRES; not surprisingly, the two methods perform very similarly.

There are other instances when the emitting lower layer is not thick enough to be opaque, but a chemically different surface layer appears to absorb at its characteristic frequencies from the emission spectrum of the lower layer in regions of spectral overlap. Such spectra, containing positive and negative bands alike, are hard to interpret by conventional methods. They may be called emission–absorption spectra, and require spectral simulation and modeling to be fully understood.[10]

5 COLLECTION, HANDLING AND PRESENTATION OF SPECTRAL DATA

5.1 Data Acquisition and Reduction

In view of the relatively low signal level of IR emission spectra, high-speed, high-sensitivity FTIR instruments are commonly used in IRES. With these instruments 1 to 10 interferograms can be measured in a second, so in order to increase the S/N it is feasible to accumulate and average several hundred or even several thousand scans (S/N increases as the square-root of the number of scans). Fourier transformation of the final interferogram using the fast Fourier transform (FFT) algorithm[12] of Cooley and Tukey is just a matter of seconds on PCs for spectra at medium resolution ($1-2\,cm^{-1}$). With condensed-phase samples there is usually no need to use higher than $4\,cm^{-1}$ resolution, since the natural band widths are generally much greater, and higher resolution is concomitant with higher noise.

The advanced data-handling software accompanying most commercial FTIR spectrometers allows the smoothing of noisy spectra, the increase of the apparent spectral resolution by deconvolution techniques, the decomposition of overlapping bands, and the plotting of the spectra in any desired format, etc. These possibilities do not exist without a computer being interfaced with the instrument. Emission spectroscopy may require a few special data-processing operations that are not necessarily found in the otherwise user-friendly, menu-driven software of some commercial systems (e.g. the transformation from emittance to absorbance, and the generation of theoretical blackbody emission spectra). Thus it may prove very handy to have access to independent multipurpose spectroscopy software (such as SpectraCalc or GRAMS/32 of Galactic Industries Corporation) or to general purpose data-handling programs (such as the Microcal Origin of Microcal Software, or the Mathlab of The MathLab).

5.2 Presentation of Emission Spectra: Intensity Scales

Sometimes the single-beam spectra are presented as the final result of the measurement. Alternatively they can be converted to double-beam spectra by dividing them by a reference spectrum measured at the same temperature, with the aim of obtaining a uniform intensity scale throughout the whole spectral region of the measurement and of eliminating the instrumental distortion effects. In the IRES literature three kinds of double-beam emission spectra are encountered:[10]

1. relative emittance, ε_s^{rel}, obtained as $L_s(T)/L_0(T)$, i.e. when the sample emission spectrum is divided by the emission spectrum of a reference material other

than a blackbody, for example that of the empty heater stage or backplate, or when the spectrum of the sample, treated in some way, is divided by that of the untreated sample (all measured at the sample temperature);

2. emittance spectrum, ε_s, obtained as $L_s(T)/L_{bb}(T)$ or according to Equation (38);
3. reflection-corrected emittance, ε_s^*, obtained as $L_s(T)/L_{op}(T)$ according to Equation (43).

Of these, relative emittance (1) has the serious disadvantage that it is in an arbitrary intensity scale that can reach very high values with increased noise level. This follows from the fact that the reference spectrum is sometimes very weak, and division with near zero values leads to emittance values with high relative noise.

Emittance spectra (2) or reflection-corrected emittance spectra (3), on the contrary, are standardized in the sense that they have intensity values between 0 and 1 (or 0 and 100%), so they are similar and complementary to transmittance spectra. It is important to understand that none of the above intensity scales are linearly proportional to sample thickness or analyte concentration. Thus spectral manipulations requiring a linear intensity scale (such as subtraction of spectra, multiplication by a constant, etc.) should not be performed with emittance spectra since they lead to spectral distortions (e.g. altered relative intensities, spurious shoulders and splittings) and may falsify the quantitative information contained in them. Transformation to the linear intensity scale, absorbance, can be performed according to Equation (41).

5.3 Quantitative Aspects

As we have seen above, Equations (40) and (41) provide the connection between emittance (ε), transmittance (τ) and the linear intensity scale, absorbance (A). Although Equation (40) containing the decadic (base 10) absorption coefficients is the most frequently used form of the Lambert–Beer law, for the sake of completeness we give here all versions including those written in terms of the napierian (base e) absorption coefficients (Equations 48 and 49):

$$\varepsilon(\tilde{\nu}) = 1 - 10^{-K(\nu)d} = 1 - 10^{-b(\nu)cd} \tag{48}$$

$$\varepsilon(\tilde{\nu}) = 1 - \tau(\tilde{\nu}) = 1 - \exp[-\alpha(\tilde{\nu})d] = 1 - \exp[-\kappa(\tilde{\nu})cd] \tag{49}$$

where $K(\tilde{\nu})$ and $b(\tilde{\nu})$ are the linear and molar decadic absorption coefficients, while $\alpha(\tilde{\nu})$ and $\kappa(\tilde{\nu})$ are the linear and molar napierian absorption coefficients, respectively; d is the sample thickness or pass length, and c is the concentration. The quantities called linear are used in case of neat, condensed-phase samples, and those called molar are used in case of solutions, mixtures or gases.

In case of multicomponent samples the absorbance spectrum can be expressed through the emittance spectra of the components as follows (Equation 50):

$$A(\tilde{\nu}) = -\log[1 - \varepsilon(\tilde{\nu})] = \sum_i^n \varepsilon_i(\tilde{\nu})c_i d \tag{50}$$

where n is the number of components.

Another method increasingly used for qualitative evaluation of single-beam emission or double-beam emittance data directly (without transforming to absorbance scale) is partial least squares (PLS) multivariate calibration.[39,40] With this method in some cases even better precision can be achieved than is possible with the use of a more traditional univariate analysis, especially when one has to deal with overlapping bands belonging to different chemical species.

6 CONCLUSION AND FUTURE PERSPECTIVE

As we have seen, the fundamental principles and governing laws of IRES with thermal excitation are firmly established; the various sources of distorting effects observed in the spectra are well understood and correction methods have been developed. Compared with IRAS, the S/N in an emission experiment is from 0.2 to 2 orders of magnitude smaller; but in principle, IRES should be more sensitive than IRAS, because it measures the neat sample signal instead of the changes in a continuous, high-intensity background created by the light source used in absorption spectroscopy.[27] IRES may live up to these expectations as soon as an IR avalanche detector (the IR analog of a photomultiplier tube) becomes available for the measurement of very low radiation levels.

Nevertheless, the present level of IR technology already makes IRES an attractive and very useful tool that offers solutions to problems in many branches of chemistry and technology. A recent literature survey[41] shows continued interest in application of IRES to very diverse areas such as: remote sensing and open-path atmospheric monitoring, IRES studies of chemical reactions through highly vibrationally excited reaction products, in situ and real-time monitoring of chemical reactions, deposition processes, etc. the study of vibration-rotation bands of stable and transient molecules, photoluminescence and chemiluminescence. An emerging new field of application of the transient IR spectroscopic methods is the determination of chemical and physical properties of solids and viscous liquids in a moving stream.[42] With further advancement this technique can lead to real-time, on-line

process control, e.g. in certain production plants or in waste processors.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-Digital Converter
DTGS	Deuterated Triglycine Sulfate
FFT	Fast Fourier Transform
FIR	Far-infrared
FT	Fourier Transform
FTIR	Fourier Transform Infrared
IR	Infrared
IRAS	Infrared Absorption Spectroscopy
IRES	Infrared Emission Spectroscopy
IRE/TC	Infrared Emission with Transient Cooling
LITES	Laser-Induced Thermal Emission Spectroscopy
MCT	Mercury Cadmium Telluride
MES	Modulated Emission Spectroscopy
Nd/YAG	Neodymium Yttrium Aluminum Garnet
NIR	Near-infrared
OP/FTIR	Open-Path Fourier Transform Infrared
PET	Poly(ethylene terephthalate)
PLS	Partial Least Squares
PMMA	Poly(methyl methacrylate)
ROSE	Remote Optical Sensing of Emissions
S/N	Signal-to-noise Ratio
TGS	Triglycine Sulfate
TIRES	Transient Infrared Emission Spectroscopy
TIRTS	Transient Infrared Transmission Spectroscopy
USEPA	United States Environmental Protection Agency
UV	Ultraviolet
3-D	Three-dimensional

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Infrared Spectroscopy **(Volume 12)**
Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy • Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

REFERENCES

1. W.C. Price, 'Infra-red Emission Spectra', in *Infra-red Spectroscopy and Molecular Structure. An Outline of the Principles*, ed. M. Davies, Elsevier, Amsterdam, 441–458, 1963.
2. P.R. Griffiths, 'Infrared Emission Spectroscopy. I. Basic Considerations', *Appl. Spectrosc.*, **26**(1), 73–76 (1972).
3. J.B. Bates, 'Infrared Emission Spectroscopy', in *Fourier Transform Infrared Spectroscopy. Applications to Chemical Systems*, eds. J.R. Ferraro, L.J. Basile, Academic Press, New York, 99–142, Vol. 1, 1978.
4. P.V. Huong, 'Infrared Emission Spectroscopy', in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Heyden, London, 85–107, Vol. 4, 1978.
5. J.M. Chalmers, M.W. Mackenzie, 'Solid Sampling Techniques', in *Advances in Applied Fourier Transform Infrared Spectroscopy*, ed. M.W. Mackenzie, Wiley, New York, 106–188, 1988.
6. F.J. DeBlase, S. Compton, 'Infrared Emission Spectroscopy: A Theoretical and Experimental Review', *Appl. Spectrosc.*, **45**(4), 611–618 (1991).
7. G. Herzberg, *Molecular Spectra and Molecular Structure II. Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand Reinhold, New York, 1945.
8. D.B. Chase, 'The Sensitivity and Limitations of Condensed Phase Infrared Emission Spectroscopy', *Appl. Spectrosc.*, **35**(1), 77–81 (1981).
9. G. Keresztury, J. Mink, 'Significance of Correction for Detector Temperature in Infrared Emission Spectroscopy', *Appl. Spectrosc.*, **46**(11), 1747–1749 (1992).
10. G. Keresztury, J. Mink, J. Kristóf, 'Quantitative Aspects of FTIR Emission Spectroscopy and Simulation of Emission–Absorption Spectra', *Anal. Chem.*, **67**(20), 3782–3787 (1995).
11. D. Kember, D.H. Chenery, N. Sheppard, J. Fell, 'Fourier-transform i.r. Emission Studies of Weakly Emitting Overlayers on Metal Surfaces; Experimental and Spectral-ratioing Procedures and Comparative Use of Room Temperature TGS and Low Temperature MCT Detectors', *Spectrochim. Acta*, **35A**(5), 455–459 (1979).

12. P.R. Griffiths, J. deHaseth, 'Fourier Transform Infrared Spectrometry', *Chemical Analysis Series*, Vol. 83, Wiley, New York, 1986.
13. J.K. Wilmshurst, 'Infrared Spectra of Molten Salts', *J. Chem. Phys.*, **39**(10), 2545–2548 (1963).
14. T.R. Kozlowski, 'Application of High Temperature Infrared Emission Spectroscopy to Molten Salts', *Appl. Opt.*, **7**(5), 795–800 (1968).
15. J. Hvistendahl, E. Rytter, H.A. Oye, 'IR Emission Spectroscopy of Molten Salts and Other Liquids Using Thick Samples as Reference', *Appl. Spectrosc.*, **37**(2), 182–187 (1983).
16. E. Rytter, 'A Comment on Emission, Reflection and Transmission Spectroscopy of Thin Films', *Spectrochim. Acta*, **43A**(4), 523–529 (1987).
17. D. Kember, N. Sheppard, 'The Use of Ratio-recording Interferometry for the Measurement of Infrared Emission Spectra: Applications to Oxide Films on Copper Surfaces', *Appl. Spectrosc.*, **29**(6), 496–500 (1975).
18. H.A. Willis, N. Sheppard, J. Chalmers, 'Precision in Infrared Quantitative Analysis; Absorptance is the Key', 7th International Conference on Fourier Transform Spectroscopy, ed. D.G. Cameron, Proc. SPIE, Vol. 1145, 431–432, 1989.
19. R.G. Greenler, 'Light Emitted from Molecules Adsorbed on a Metal Surface', *Surf. Sci.*, **69**, 647–652 (1977).
20. K. Tochigi, H. Momose, Y. Misawa, T. Suzuki, 'Polarization-modulation Emission FTIR Measurement of Thin Organic Films on Metal Surfaces', *Appl. Spectrosc.*, **46**(1), 156–158 (1992).
21. J.E. Stuart, *Infrared Spectroscopy. Experimental Methods and Techniques*, Marcel Dekker, New York, 1970.
22. W.F. Herget, 'Air Pollution: Ground-based Sensing of Source Emissions', in *Fourier Transform Infrared Spectroscopy. Applications to Chemical Systems*, eds. J.R. Ferraro, L.J. Basile, Academic Press, New York, 111–127, Vol. 2, 1979.
23. D.L. Allara, D. Teicher, J.F. Durana, 'Fourier Transform Infrared Emission Spectrum of a Molecular Monolayer at 300 K', *Chem. Phys. Lett.*, **84**(1), 20–24 (1981).
24. J.G. Moehlmann, J.T. Gleaves, J.W. Hudgens, J.D. McDonald, 'Infrared Chemiluminescence Studies of the Reaction of Fluorine Atoms with Monosubstituted Ethylene Compounds', *J. Chem. Phys.*, **60**(12), 4790–4799 (1974).
25. S. Chiang, R.G. Tobin, P.L. Richards, 'Vibrational Spectroscopy of Chemisorbed Molecules by Infrared Emission', *J. Vac. Sci. Technol. A*, **2**(2), 1069–1074 (1984).
26. D.M. Adams, M.A. Hooper, 'Methods of Obtaining Spectra at High and Low Temperatures', in *Laboratory Methods in Infrared Spectroscopy*, eds. R.G.J. Miller, B.C. Stace, Heyden, London, 129–148, 1972.
27. P.F. Bernath, 'Infrared Fourier Transform Emission Spectroscopy', *Chem. Soc. Reviews*, **25**(2), 111–115 (1996).
28. P.C.M. van Woerkom, 'FTIR Studies of Catalyst Systems under Realistic Processing Conditions', in *Advances in Applied Fourier Transform Spectroscopy*, ed. M.W. Mackenzie, John Wiley & Sons, New York, 323–347, 1988.
29. M. Handke, N.J. Harrick, 'A New Accessory for Infrared Emission Spectroscopy Measurements', *Appl. Spectrosc.*, **40**(3), 401–405 (1986).
30. B. Wangmaneerat, T.M. Niemczyk, D.M. Haaland, 'Modifications to a Commercial Infrared Emission Apparatus to Permit Quantitative Applications', *Appl. Spectrosc.*, **46**(1), 178–180 (1992).
31. L.T. Lin, D.D. Archibald, D.E. Honigs, 'Preliminary Studies of Laser-induced Thermal Emission Spectroscopy of Condensed Phases', *Appl. Spectrosc.*, **42**, 477–483 (1988).
32. A. Tsuge, Y. Uwamino, T. Ishizuka, 'Applications of Laser-induced Thermal Emission Spectroscopy to Various Samples', *Appl. Spectrosc.*, **43**, 1145–1149 (1989).
33. R.W. Jones, J.F. McClelland, 'Transient Infrared Emission Spectroscopy', *Anal. Chem.*, **61**, 650–656 (1989).
34. R.W. Jones, J.F. McClelland, 'Transient Infrared Emission Spectroscopy by Pulsed Laser Excitation', *Anal. Chem.*, **61**, 1810–1815 (1989).
35. R.W. Jones, J.F. McClelland, 'Quantitative Analysis of Solids in Motion by Transient Infrared Emission Spectroscopy Using Hot-gas Jet Excitation', *Anal. Chem.*, **62**, 2074–2079 (1990).
36. O. Guillois, I. Nenner, R. Papoular, C. Reynaud, 'IR Emission Spectrum of Solid Materials under Periodic Heating: Modulated Emission Spectroscopy (MES)', *Appl. Spectrosc.*, **48**(3), 297–306 (1994).
37. R.W. Jones, J.F. McClelland, 'Transient Infrared Transmission Spectroscopy', *Anal. Chem.*, **62**, 2247–2251 (1990).
38. R.J. Pell, Ch.E. Miller, B.R. Kowalski, J.B. Callis, 'Infrared Emission Spectroscopy with Transient Cooling', *Appl. Spectrosc.*, **47**, 2064–2071 (1993).
39. R.J. Pell, B.C. Erickson, R.W. Hannah, J.B. Callis, B.R. Kowalski, 'Quantitative Infrared Emission Spectroscopy Using Multivariate Calibration', *Anal. Chem.*, **60**, 2824–2827 (1988).
40. B. Wangmaneerat, J.A. McGuire, T.M. Niemczyk, D.M. Haaland, J.H. Linn, 'Quantitative Infrared Emission Spectroscopy of Phosphosilicate Glass on Silicon Wafers Using Multivariate Calibration', *Appl. Spectrosc.*, **46**(2), 340–344 (1992).
41. M.L. McKelvy, T.R. Britt, B.L. Davis, J.K. Gillie, F.B. Graves, L.A. Lentz, 'Infrared Spectroscopy', *Anal. Chem.*, **70**(12), 119R–177R (1998).
42. S.J. Weeks, J.F. McClelland, S.L. Wright, R.W. Jones, 'Transient Infrared Spectroscopy for On-line Analysis of Solids and Viscous Liquids', *Mikrochim. Acta*, [Suppl.] **14**, 801–802 (1997).

Gas Chromatography/Infrared Spectroscopy

Jean-Luc Le Quéré
Institut National de la Recherche Agronomique, Dijon, France

1 Introduction **1**
2 Theory and Operating Principles **2**
2.1 Fourier Transform Infrared Spectrometry 2
2.2 Gas Chromatography Requirements 3
2.3 Gas Chromatography/Fourier Transform Infrared Spectrometry Coupling 4
3 The Gas Chromatography/Fourier Transform Infrared Interfaces **6**
3.1 The Light Pipe 7
3.2 The Matrix Isolation Technique 8
3.3 The Direct Deposition Interface 10
3.4 Gas Chromatography/Fourier Transform Infrared Spectra and Sensitivity 12
4 Some Applications of Gas Chromatography/Fourier Transform Infrared Spectrometry **15**
4.1 Environmental 15
4.2 Flavor and Fragrances 15
4.3 Pheromones 15
4.4 Fatty Acids 16
5 Perspective **16**
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Gas chromatography/infrared (GC/IR) spectroscopy coupling combines the separation power of high-resolution gas chromatography (HRGC) to the highly specific identification technique, infrared (IR) spectroscopy, based on the absorption of IR radiation by the molecules. Relying on different molecular properties, IR is an excellent complement to mass spectrometry (MS) in structural analyses. For volatile compounds, the coupling of gas chromatography/mass spectrometry (GC/MS) has been the universal basic technique for a long time. The advantages of Fourier transform infrared spectrometry (FTIRS) allowed the coupling of gas chromatography (GC) to an FTIR (Fourier transform infrared) spectrometer, ending with a powerful tool, gas chromatography/Fourier transform infrared spectrometry (GC/FTIRS). The first device developed to interface a gas chromatograph to an FTIR spectrometer was a flow-through gas cell, known as the light pipe (LP), delivering vapor-phase IR spectra of eluting solutes. Two other cryogenic devices which allowed mobile-phase elimination by trapping the compounds eluting from the GC at sub-ambient temperature have been developed: the matrix isolation (MI) interface operating at 11 K and the direct deposition (DD) window operating at 77 K. Each of these three types of interface has advantages and drawbacks that are discussed in terms of sensitivity and specificity. Some selected applications in selected domains of analytical chemistry, i.e. environment, flavor and fragrances, pheromones and fatty acids, are presented and discussed again in terms of sensitivity and specificity according to the interface type used. Finally some concluding remarks on technique developments for GC/FTIRS and related techniques combining chromatography with IR spectroscopy are given.

1 INTRODUCTION

Structural analysis of individual components in a complex mixture requires the coupling of separative techniques to identification techniques, mainly spectroscopic. For volatile samples capillary GC allows the separation of most samples with high resolution, thanks to the many stationary phases commercially available. For many years, the only coupled technique available for volatile samples was GC/MS. It is one of the most widely used tools for structure determination and many areas of analytical chemistry rely upon capillary GC/MS for both qualitative and quantitative analyses. The determination of the structure of a compound from its mass spectrum can be made in several ways, the initial approach generally consisting in comparison of the experimental spectrum with existing library spectra. When the spectrum is not in any library, complete interpretation of the spectrum, sometimes helped with chemical ionization now available even on very basic benchtop instruments, is more time-consuming, and not always successful. Moreover, MS alone does not solve all the structural problems. Particularly, MS is not able to differentiate some isomers. Generations of chemists have used IR spectroscopy as a powerful tool in the analysis of molecular structures. The informative content of an IR spectrum is very high, with clear identification of chemical functions and substructures. Each molecule possesses its own characteristic IR fingerprint, and the possibility

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of identifying compounds with very high accuracy by simple examination and matching of their IR spectra has been realized. Moreover, IR is particularly suited to differentiate conformational and structural isomers. Coupling GC with conventional grating IR instruments, when a spectrum takes some minutes to be acquired, was not practically feasible, despite trials essentially in the effluent stopped-flow mode.[1,2]

Taking into consideration the numerous advantages of FTIRS over conventional dispersive IR spectroscopy, among which is the speed of analysis, the first attempts to interface a gas chromatograph to an FTIR spectrometer were made in the late 1960s,[1] even though at that time, the instruments were sometimes used in the stopped-flow mode using high-capacity GC columns,[1,3] and not with on-line detection in the on-flow mode.

Since then, FTIR instrumentation has evolved considerably and compulsory computers provided extreme on-line computing power for less and less cost. State-of-the-art FTIR spectrometers are able to interface with capillary GC and to acquire IR data fast enough to accommodate the narrowest peaks obtained with high-resolution capillary GC. GC/FTIRS appears as a complementary tool to GC/MS for volatile compound identification purpose, but still represents a challenge in terms of sensitivity limits compared to this last ubiquitous technique.

Optimization of GC/FTIRS as a combined technique requires careful attention to all the component parts: GC, transfer lines to the measuring interface, the interface itself, and the IR spectrometer. Each part will be addressed in the following, but not at the same level. Only those parts specific to the technique will be developed.

HRGC using narrow-bore fused-silica wall-coated capillary columns is now standard in most analytical laboratories. Availability of stationary phases as well as capillary column dimensions adapted to a wide range of application, has made the technique widespread. Only specific comments related to the coupling to IR spectrometry will be presented.

It is beyond the scope of this article to present the basic principles of FTIRS. These principles may be found in another article, and in dedicated treatises.[4,5] Only those advantages of FTIRS over conventional grating IR spectroscopy which are relevant to the interfacing of GC will be developed.

The operating principles of GC/FTIRS coupling will be presented, with special emphasis on IR reconstructed chromatograms, IR spectra computation and the spectral library search. Finally, the main part will be devoted to the GC/IR interface. Three types are now commercially available: LP; MI; and DD. Each type has specific advantages and drawbacks, with clear impact on the technique sensitivity.

Some applications in various fields of analytical chemistry will be finally presented in order to illustrate the various interesting possibilities of the method.

2 THEORY AND OPERATING PRINCIPLES

2.1 Fourier Transform Infrared Spectrometry

An FTIR instrument used for coupling to GC is basically a single-beam FTIR spectrometer where the modulated IR beam is delivered by a rapid-scanning two-beam interferometer generally based on the original Michelson design.[4] The Michelson interferometer is a device that can divide a beam of IR radiation into two paths using a beam-splitter and then recombine the two beams within a single modulated beam after a path difference has been introduced by a moving mirror (Figure 1). The interference pattern thus generated by the interferometer, the interferogram, reflects the intensity variations of the beam and can be measured as a function of path difference by an IR detector. This interferogram is the cosine Fourier transform of the spectrum of the incident radiation. As the Fourier transform is mathematically reciprocal, the IR spectrum is calculated by computing the cosine Fourier transform of the interferogram.

All rapid-scanning interferometers usually operate by signal-averaging successive scans. For precise signal-averaging, each interferogram must be sampled at exactly the same retardation delay for every scan. Therefore, very precise matching of the movable mirror is required to control the change in optical path difference. This control is obtained by the use of the interference pattern of the monochromatic light of a He–Ne laser. The sine interference pattern of the laser source, obtained either with the main interferometer or with a dedicated

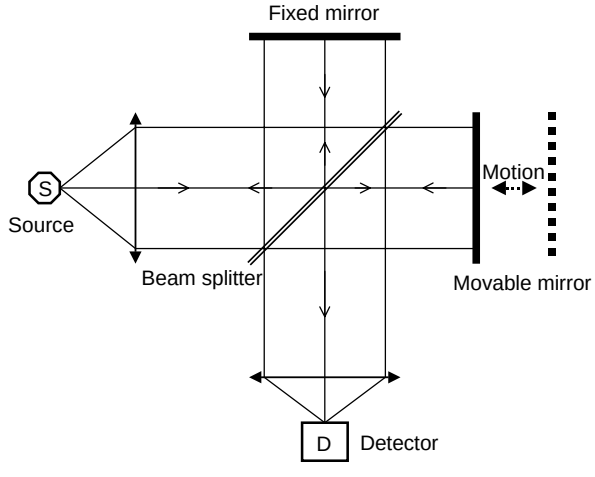

Figure 1 Schematic representation of a Michelson interferometer.

synchronized separate reference interferometer, allows precise digitization of the IR interferogram at the zero crossings of the laser sinusoidal interferogram. Moreover, with this He–Ne laser, FTIR spectrometers have a built-in wavenumber calibration of very high precision (ca. 0.01 cm^{-1}). This is an important advantage of FTIR instruments over grating spectrophotometers known as the Connes advantage.

Other advantages with clear impacts on coupling with GC are specific to FTIRS. With FTIRS, all the spectral elements emanating from the IR source and transmitted through the sample are simultaneously received at the detector. Contrarily, with grating spectrometers where a monochromator is used, the spectral frequencies reach the detector successively. This multiplex or Felgett advantage allows a better signal-to-noise (S/N) ratio to be achieved, only proportional, at a given spectral resolution, to the square root of the measurement time, and independent of the number of measured spectral elements. The throughput or Jacquinot advantage concerns the energy range of the transmitted IR beam, which is not limited in FTIRS by the linear slits used in grating instruments to reach the desired spectral resolution. This advantage is clearly important when GC/IR interfaces may drastically attenuate the radiation throughput.

Finally, one of the main advantages of FTIR spectrometers in coupling with GC is their rapid-scanning capacity. This allows acquisition of full spectra in fractions of a second, which is mandatory for capillary GC when the duration of a GC peak is typically a few seconds. Using signal-averaging, the sensitivity is also increased as the S/N ratio, dependent on the square root of the measurement time, increases with the square root of the number of scans being signal-averaged.

The last advantage often previously cited[6] for FTIR instruments was the presence of a computer, compulsory for Fourier transform operations. Nowadays, with the availability of compact and inexpensive microcomputers in every analytical laboratory, this advantage can hardly be considered as specific to FTIR instrumentation.

Most of the commercially available GC interfaces are specific accessories to conventional FTIR spectrometers (Figure 2). The collimated modulated IR beam from the interferometer, normally focused at the sample point in the sample compartment, is simply directed by a removable mirror to an external port where the GC interface is connected. The only exception is the IR detector initially commercialized by Hewlett Packard as the IRD™, which is a dedicated GC/FTIR instrument with optimized optics using a vapor-phase LP interface.

2.2 Gas Chromatography Requirements

The separation of volatile components of a complex mixture by GC is widely used by analytical chemists in various fields for both qualitative and quantitative purposes.

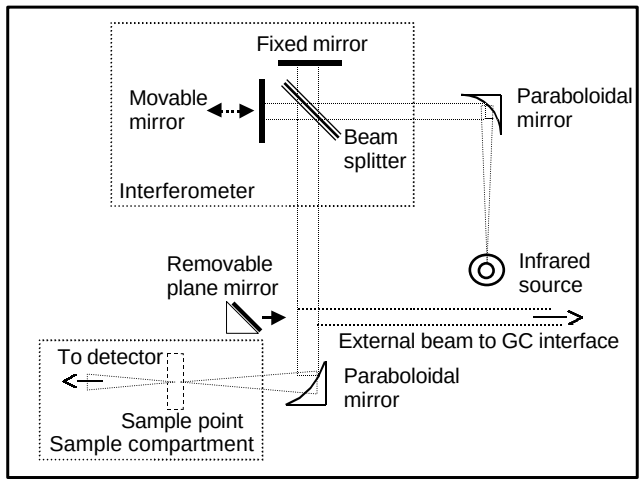

Figure 2 A typical FTIR spectrometer used to generate a modulated IR beam.

GC is basically a dynamic technique in which the individual components, the eluates, elute in a few seconds from the column. The dynamic range of component concentration within the sample is generally very high, the components ranging from a few tens of micrograms to a few nanograms or less. Rapid-scanning interferometers capable of acquiring a few spectra within a single second with signal averaging may accommodate the high resolution obtained with capillary GC using conventional wall-coated open tubular (WCOT) columns (0.25 or 0.32 mm internal diameter). Hence, collection of IR spectra of samples in submicrogram quantities is possible, subject to the use of an interface adapted to vapor-phase IR spectra recording.

However, since the IR response is directly related to the concentration of the sample within the GC peak, and to the presence of strong IR-absorbing chemical functions in the molecules, the best S/N ratios will be obtained for strong absorbers focused within a narrow GC peak. Optimization of the GC separation is therefore all the more necessary since the dynamic range of compound concentration within the analyzed mixture is generally very high. This implies that for any GC/IR interface using either on-the-fly detection or eluate trapping very low dead volumes must be used in order to be compatible with HRGC and to avoid peak broadening or even eluates re-mixing.

Finally, the GC column to be used in GC/FTIR applications has to be chosen in terms of sample capacity and achievable resolution.[5] For standard applications with commercial interfaces, conventional WCOT columns of 0.25 or 0.32 mm internal diameter with thin film

thickness are adequate. Chemically bonded and cross-linked stationary phases are advisable to avoid bleeding which may pollute the interface. The dynamic range of compound concentration for GC/FTIRS analysis is determined by column capacity at the upper end and by IR S/N ratio at the lower end.[7]

2.3 Gas Chromatography/Fourier Transform Infrared Spectrometry Coupling

The potential of an FTIR spectrometer interfaced to a GC was realized soon after rapid-scanning interferometers were marketed.[6] It was clear that one application for which the advantages of FTIRS should be beneficial was the on-line measurement of IR spectra of eluates from a gas chromatograph. Since then, the layout of instruments has not really changed (Figure 3), even though the cost for any individual part of this technique has dramatically dropped (this is particularly true for computers) and in the meantime their performances have been considerably improved. The GC/FTIR interface ''box'' will be developed later.

Passing from collected interferograms during a GC run to IR spectra of each individual GC peak requires the steps shown in Figure 4. The first step is to transform the set of collected interferograms into a reconstructed IR chromatogram. To save space on storing media, it is advisable to store only those interferograms which contain relevant spectral information. However, modern storage media may accommodate complete runs.

Basically, and theoretically, to locate the interferograms corresponding effectively to spectra of GC eluates, it is necessary to transform all the collected interferograms of the GC run into their respective spectra and then discriminate solute spectra from background. Clearly this is not of practical use for complex mixtures of several dozen peaks. Alternatively, as the presence of any IR-active

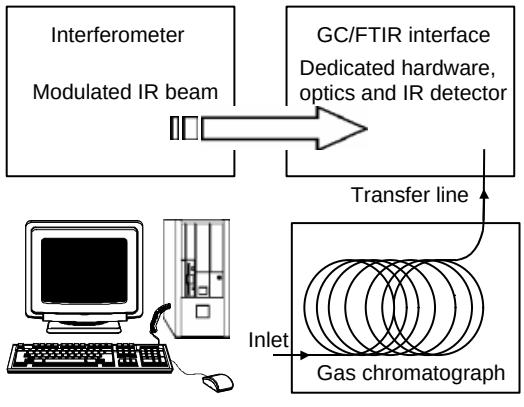

Figure 3 Schematic layout of a GC/FTIR system with the essential parts: interferometer, interface, gas chromatograph, and computer.

chemical in the IR beam should reduce the interferogram intensity, one method consists of calculating the overall absorption power of each interferogram. In this method, analogous to the total ion current (TIC) in GC/MS, however, no reliable results are obtained, because most GC/IR absorbances are well below the practical interferogram intensity difference threshold.[8] Several methods to construct or reconstruct chromatograms from interferometric data have been used to overcome these drawbacks, and some of them have been compared in terms of optimal S/N ratio and computation time.[9] Among the many reconstruction methods proposed so far, two of the early described methods that construct chromatograms directly from the interferometric data are currently used.

The first of the two methods relies on the fast Fourier transform (FFT) algorithm. A spectrum is computed in real time and ratioed against a stored reference spectrum acquired before the GC run and computed at the same resolution. The integrated IR absorbance is calculated in discrete spectral regions, generally corresponding to characteristic absorptions of selected functional groups, to produce spectral window chromatograms. In the original work[6,10] for evident computational limitations, five spectral windows could be defined using $32\,cm^{-1}$ resolution spectra computed in real time from 512 data points around the centerburst of the interferograms collected as single scans at a resolution of $8\,cm^{-1}$. With modern computers it is possible to produce several separate spectral regions or ''windows'' chromatograms calculated from spectra signal-averaged and computed in real time at the spectral resolution used to acquire the interferograms. These functional group chromatograms, that can be compared to single-ion monitoring (SIM) in GC/MS, are useful for identifying GC fractions belonging to a particular chemical class in complex mixtures. The S/N ratio of the functional group chromatograms may be optimized in a straightforward manner if the spectral window is defined precisely around the maxima of the absorption bands to be monitored. This is not always feasible when handling unknown samples, where it is not possible to know the most characteristic wavenumbers in advance.[5] To find suitable spectral ranges for reconstructing window chromatograms, an elegant approach was developed that used spectral data representation as a contour plot.[5,11] In this representation, a routine technique in two-dimensional nuclear magnetic resonance (NMR) spectrometry, the complete spectral information is plotted as contours of absorbance bands in a two-dimensional map of wavenumbers against time coordinates. Using this kind of data representation, even spectral information of overlapping peaks may be used to reconstruct very precise functional group chromatograms.[5,11] The total absorbance across

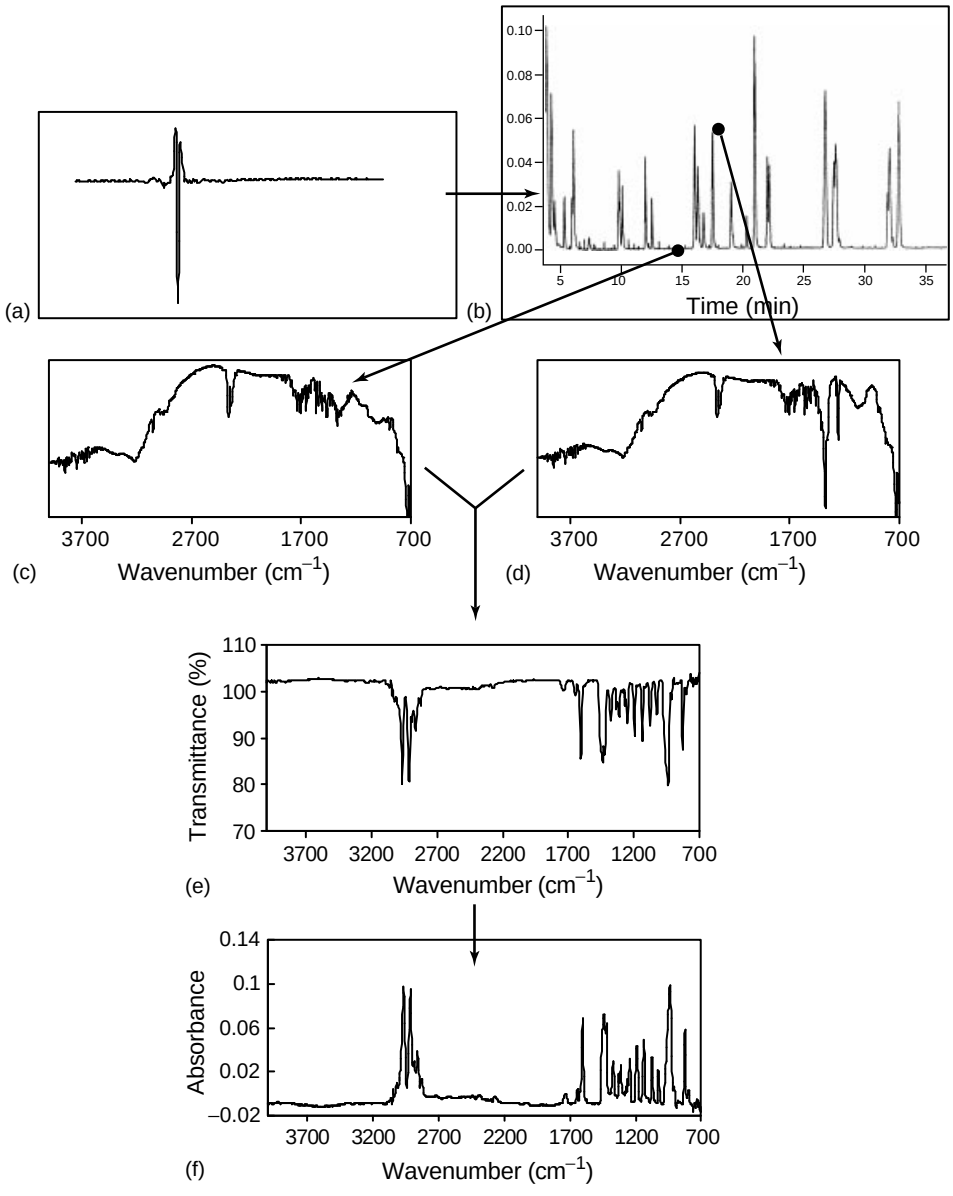

Figure 4 Procedure used to generate a GC/FTIR transmittance or absorbance spectrum: (a) interferogram; (b) reconstructed IR chromatogram; (c) single-beam reference spectrum measured when no GC peak is eluting; (d) single-beam sample spectrum of eluting compound; (e) transmittance spectrum, ratio of (d/c); (f) absorbance spectrum.

the entire spectral region can be calculated to give a reconstructed chromatogram independent of the window regions that have to be chosen by the operator and where all the peaks may be observed, but at the cost of the S/N ratio.[6]

The other method that may be used for real-time reconstruction of chromatograms directly from interferometric data is the Gram–Schmidt (GS) vector orthogonalization method developed by de Haseth and Isenhour.[8] The GS method does not involve any Fourier transform as it relies on the multiplex nature of the interferogram: a single interferogram contains complete spectral information on absorbing samples. Therefore, a portion of the interferogram contains sufficient information to determine whether an eluate is present in the IR beam or not. In this technique, a reference set of x interferograms is collected immediately before the GC run. A small segment of each reference interferogram near the centerburst is selected. Consecutive values of the interferogram portion are treated as a vector. If n data points are chosen, they are treated as an n-dimensional vector $\mathbf{I}$ in hyperspace as shown in Equation (1).

$$\mathbf{I} = (i_1, i_2, i_3, \ldots, i_n) \tag{1}$$

where $i_1, i_2, i_3, \ldots, i_n$ are the n consecutive values of the interferogram portion, or the coordinates of the vector in the hyperspace. Using the set of x reference interferograms, a set of x orthogonal basis vectors is constructed. This set of basis vectors forms a multidimensional reference hyperspace and represents the background conditions, i.e. the optics characteristics when no sample eluting from the GC column is present in the light path. Each single-scan interferogram in a subsequent GC/IR run is treated in the same way as a sample vector. This sample vector can be orthogonalized and compared to the existing basis, and the absolute length of the resulting scalar from each interferogram measures the total IR absorbance of any IR active sample in the IR beam,[8] i.e. refers to the quantity of eluate in the light path. To a certain extent, GS reconstructed chromatograms can be compared to TIC chromatograms in GC/MS, since each absorption band contributes to the signal. However, this must be relativized, as the signal is also dependent on the molar absorptivities of bands which tend to be much greater in the IR spectra of polar molecules than in those of nonpolar molecules.[6] Optimization of GS parameters (offset from the centerburst of the interferogram, number of interferogram data points used to build the vectors, number of basis vectors in the reference hyperspace) has been the subject of specific work.[5] A good compromise seems to deliver an optimum signal giving acceptable S/N ratio for the reconstructed chromatogram when 100–150 consecutive data points are selected from the interferogram in the region starting ca. 30–50 data points from the centerburst, and when approximately 10 reference vectors are included in the basis set.[4,5] The algorithm being very rapid, GS reconstruction can be done in real time simultaneously with data collection with uncompared sensitivity. It has also been demonstrated that GS chromatograms, the magnitude of which being dependent on the absorptance (1 minus transmittance) of the absorbing material, may be used for quantitative studies, as at the low concentrations typically used for GC/FTIR, absorptance varies linearly with compound concentration.[5,7]

Despite some tentative chromatogram reconstructions with digitally filtered interferograms, leading to the selectivity of the FFT method at the computational speed of the GS method,[12] and with claimed advantages on combined selectivity, S/N ratio and computational speed, the GS method remains the standard method to reconstruct total IR chromatograms. Once a reconstructed IR chromatogram has been computed from interferometric data using any reconstruction method, it is necessary to produce an interpretable IR spectrum for each peak in the mixture. This is generally done as illustrated in Figure 4. A single-beam background spectrum showing characteristics of optics, interface and detector is calculated from a region of the chromatogram where no compound is eluting, preferably not too far from a peak of interest. A single-beam sample spectrum is obtained from the interferometric data corresponding to an eluate, generally by co-adding several interferograms within the chromatographic peak, before applying FFT. Each saved interferogram is the result of the co-addition of a few (four to five) single interferograms, for signal-averaging and S/N ratio purposes.

The final transmittance or absorbance (−log transmittance) spectrum is calculated by ratioing the single-beam spectra. If the database spectral search is to be carried out on experimental spectra, the data must be expressed as absorbance spectra.

2.3.1 Spectral Resolution

The spectral resolution in FTIRS depends on retardation, or optical pathlength difference. Namely, to resolve two spectral lines separated by a distance $\Delta\nu\,\mathrm{cm}^{-1}$, the retardation $\delta\,\mathrm{cm}$ should at least be $\delta = (\Delta\nu)^{-1}$, and if the maximum retardation of an interferometer is $\Delta\,\mathrm{max}$, the best spectral resolution achievable with this interferometer is $\Delta\nu = (\Delta\,\mathrm{max})^{-1}$. Increasing the resolution, i.e. increasing the retardation, results in increasing measurement time. Clearly, in GC/FTIR coupling, the measurement time cannot be extended too much and the optimum one is determined by the elution duration of a GC peak for on-the-fly detection, taking into account signal-averaging constraints. To improve the precision of band maximum determination and photometric accuracy, interpolated points may be added in the spectrum by extending the interferogram data points with zeros before computing. This zero filling procedure (one level of zero filling doubles the file size for Fourier transformation) increases the digital resolution of the spectrum without introducing any error itself, as the raw data and the instrumental line shape are not affected. For most of the GC/IR analytical problems, a spectral resolution of $8\,\mathrm{cm}^{-1}$ and one level of zero filling before FFT seem an acceptable working compromise.[5] For specific needs, a higher spectral resolution may be used (i.e. $4\,\mathrm{cm}^{-1}$) but at the cost of decreased S/N ratio when the measurement time cannot be increased.[5]

3 THE GAS CHROMATOGRAPHY/FOURIER TRANSFORM INFRARED INTERFACES

An important breakthrough for FTIRS, which partly explains the rapid developments of GC/FTIR coupling, was the introduction of the narrow-range mercury cadmium telluride (MCT) photodetector cooled to liquid

nitrogen temperature. The specific detectivity of this detector is more than an order of magnitude greater than that of the standard deuterated triglycine sulfate (DTGS) pyroelectric bolometer supplied with most commercial FTIR spectrometers. Moreover, the MCT photodetector accommodates the high modulation frequencies necessary for GC/FTIRS, when the DTGS response decreases with increasing interferogram modulation frequency.[4,5] The first demonstration of GC/FTIRS was effective soon after the commercial achievement of rapid-scanning interferometers. Thus Low and Freeman[13] reported on-the-fly IR detection of GC eluates from a packed column using a small flow-through gas cell (ca. 0.6 mL volume) placed between the source and the interferometer.[13] Soon after, the first GC/FTIR accessory, consisting of a gas cell that fitted inside the spectrometer sample compartment, collected interferograms for packed column chromatography using either flow-through or trapped (stop-flow) modes.[4] In 1974 Azarraga and McCall (cited by Griffiths and de Haseth[4]) showed that capillary GC/FTIRS could be performed using a support-coated open tubular column with on-the-fly sampling using the gas cell (or LP) GC/IR accessory. A major breakthrough was the development of gold-coated borosilicate glass LPs by Azarraga.[14] These LPs were constructed to have a very high transmittance and they allowed the coupling of WCOT columns.[15] Today, most FTIR vendors propose optimized LP accessories for GC/FTIR coupling to fused-silica WCOT GC columns.

3.1 The Light Pipe

A typical arrangement for a GC/FTIR system using a LP interface is represented schematically in Figure 5. Clearly, the interface has to be optimized from both chromatographic and spectrometric points of view. In a typical GC/IR measurement the effluent from the GC column is passed through a heated transfer line continuously to a heated measuring gas cell, the LP. Meanwhile, the modulated IR beam collimated through an external port of the interferometer is focused to an IR-transparent window at the entrance of the LP. The radiation containing information of absorbing material within the effluent exiting from the LP is focused onto the MCT detector element. The effluent getting out of the LP may be directed, through another heated transfer line, to a conventional GC detector, as IR detection is nondestructive.

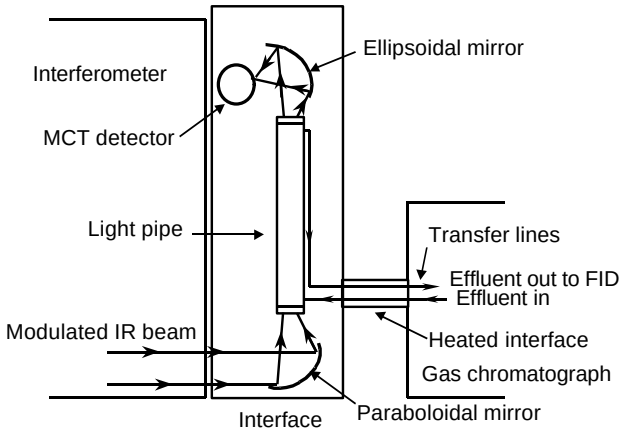

Figure 5 A typical GC/FTIR LP interface arrangement.

The LP (Figure 6) is essentially a gold-coated glass tube. The gold coating allows the IR beam to undergo multiple reflections in order to enhance the overall optical throughput of the cell. This material is assumed to be chemically inert, remarkably resistant and stable, an important advantage for a material that must be necessarily heated and that should be chromatographically inert. The high temperature of the LP is however a problem as a large decrease in signal intensity occurs

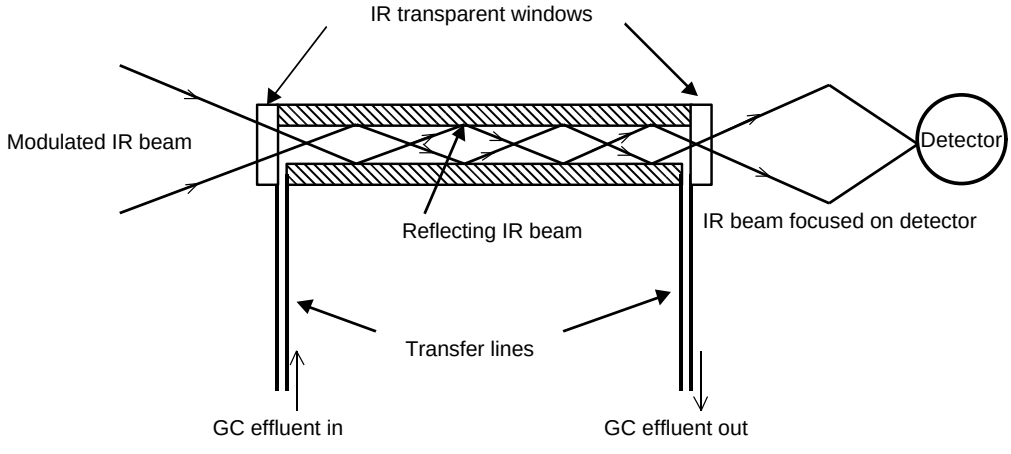

Figure 6 Schematic LP interface.

with increasing temperature. This effect results from high levels of unmodulated IR radiation emitted from the hot LP, this being an IR radiator itself, radiation which saturates the detector or its associated preamplifier.[16] Using a cold shield positioned at the end of the LP eliminates the emission from the hot LP.[16] However, it was then demonstrated that the same improvement may be obtained by using a simple optimized optical configuration where the IR beam exiting the LP is precisely focused on the detector element.[17] Practically, the temperature of the LP will be normally dictated by the upper temperature used in the GC analysis.[5]

From the GC point of view, the LP and transfer lines are dead volumes, especially when using narrow-bore GC columns. Generally, the column or fused-silica tubings are fed through the heated transfer lines up to the entrance and the exit of the LP in order to have a virtually zero dead volume. The size and geometry of the LP have been the subject of theoretical studies.[4] Clearly the length, inner diameter and length-to-diameter ratio are critical factors in achieving good S/N ratios in the resulting spectra. Ideally, the volume of the LP, V_{cell}, should not exceed the volume of carrier gas between the half-height points, $V_{1/2}$, of the narrowest peak in the chromatogram[4,6,18] to avoid a severe degradation of the GC resolution. Obviously this requirement cannot be fulfilled in every circumstance, and the practical LP volume should be a compromise between optical throughput and chromatographic considerations. A reasonable choice will depend on the most frequently used column type.[5] For narrow-bore fused-silica columns cell volumes between 50 and 150 μL are typical, and adding make-up gas could be advantageous when the volume of the LP exceeds the volume of the GC peaks. A comprehensive review of the hardware from transfer lines to optics, including sampling considerations, may be found in a specialized treatise.[4] Some practical information may be found in a monograph by Herres.[5] The sensitivity of the LP interface in terms of detection limits is in the nanogram range. Even though some studies report detection limits at the subnanogram level, it should be assumed that it was achieved only for exceptionally strong absorbing analytes. Moreover, as already outlined in 1984 by Cooper and Taylor,[19] published detection limits do not reflect the amount of material required for a positive identification, especially when spectral search identification is being used. The ''practical'' limit of detection/identification is in the tens of nanograms range for compounds that are good IR absorbers and fairly volatile. This limit is clearly higher for poor absorbers and less volatile compounds. This limit seems the threshold limit obtainable with an LP system, even with carefully optimized optics,[17] as no significant improvement has been made since the development of LP interfaces adapted to narrow-bore fused-silica WCOT GC columns.[7,20–22] A typical vapor-phase IR spectrum is presented in Figure 7. This spectrum was obtained with ca. 1 μg of 3-hydroxy-5-methyl-2-hexanone after on-column injection of an aromatic extract of water buffalo milk.[23] On-the-fly IR detection of eluates from GC separation within the LP makes this technique a true real-time system.

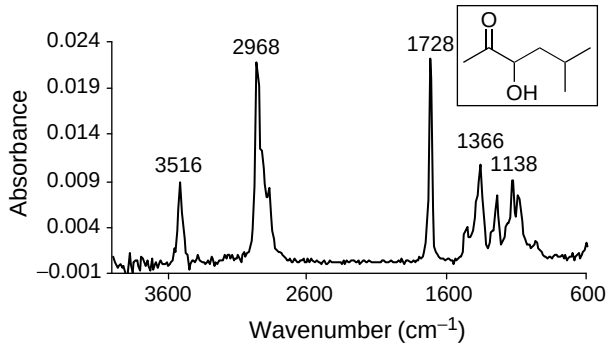

Figure 7 Vapor-phase IR spectrum of 3-hydroxy-5-methyl-2-hexanone.

3.2 The Matrix Isolation Technique

As outlined above, even an optimized LP system is not sensitive enough to solve all the problems amenable to HRGC and HRGC/MS. One possible way to gain substantial sensitivity, that is to enhance the S/N ratio, is to raise the sample absorbance and to lengthen the observation time per sample. This could be achieved by means of cold trapping the components using a suitable trap capable of freezing out highly volatile compounds, followed by off-line IR measurement with extensive signal-averaging if necessary. This can be done by using a technique known as MI.

The MI technique has been extensively used to study detailed spectrometric properties of a large variety of molecules, free radicals and ions.[4] In this technique, the molecules under investigation are co-condensed on a spectroscopically transparent window or an IR-reflecting surface with a large excess of an inert and IR-transparent gas, such as argon, at cryogenic temperatures (below 20 K) under vacuum. Trapped and diluted in a solid matrix at these temperatures, the individual molecules do not interact with each other and the molecular rotational motion is absent. As a consequence, IR absorption bands sharpen significantly, and even small shifts and band splitting may occur.

Early work demonstrated the feasibility of coupling GC to MI/FTIR,[24–26] although these original interfaces suffered from the same problem of insufficient focusing of the GC eluate on the cryogenic surface, precluding any significant gain in sensitivity. An optimized version

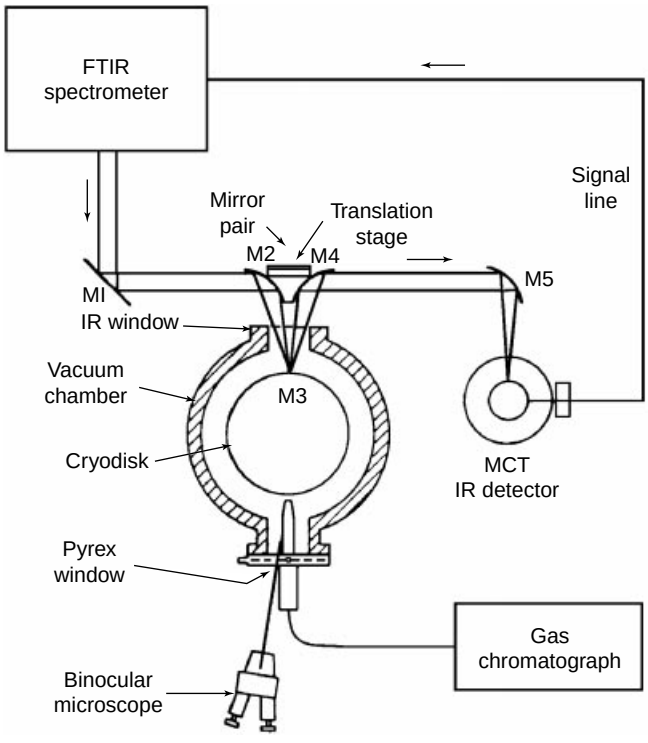

Figure 8 Original layout of the GC/MI/FTIR interface. (Reproduced by permission of the American Chemical Society from Reedy et al.[27])

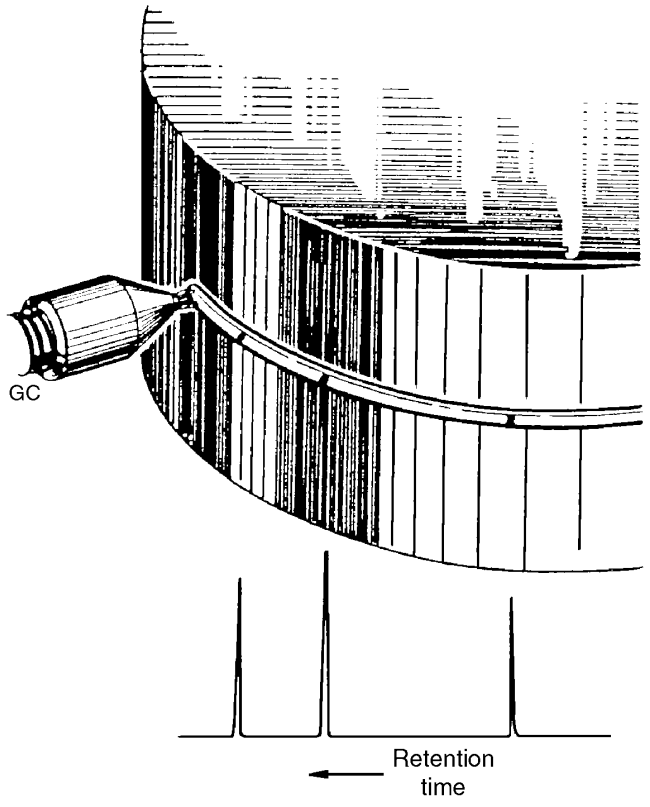

Figure 9 Enlarged view of the sample collection surface and deposition tip of the MI interface. (Reproduced by permission of the American Chemical Society from Reedy et al.[27])

was described in 1985[27] with claimed sensitivity in the 100 pg range, giving rise to a commercial GC/MI/FTIR instrument, known as the Cryolect™, the interface being developed by the Cryolect™ Scientific Corp., and available at that time from Mattson Instruments. In this system, presented schematically in Figure 8, the cryogenic surface was the gold-plated mirrored surface of a rotating cryogenic disk or ''cryodisk''. The cryodisk was placed in a vacuum chamber and maintained at a temperature of 11 K by a closed cycle helium refrigerator which did not require the use of liquid helium. The sample inlet port and the IR beam port were found on opposite sides, facing the curved surface of the cryodisk. The sample inlet port was a capillary tube passing through a copper block and a metal nozzle sealed into a Pyrex window. The end of the capillary tube protruded slightly from the nozzle. The Pyrex window and attached binocular microscope allowed observation of the nozzle and capillary tip, so that the position of the latter could be precisely adjusted with respect to the surface of the cryodisk. For IR data collection, a pair of off-axis paraboloid mirrors (M2 and M4 in Figure 8) were mounted outside the vacuum chamber. The first one focused a collimated IR beam coming from a FTIR spectrometer onto the mirrored surface (M3) of the cryodisk through a thin fused KCl window. In reflecting from the mirrored surface M3, the IR beam passed twice through condensed material and was redirected and collimated by mirror M4 towards a MCT IR detector. Figure 9 displays an enlarged view of the sample collection surface. Fused-silica capillary tubing, 150 μm in diameter and separated by 150 μm from the collection surface, protruding slightly beyond the heated nozzle, directed the GC effluent against the mirrored surface of the cryodisk. The helium carrier gas, noncondensable at the working temperature and eliminated in vacuo, was balanced with 2% argon. The argon and sample molecules formed a solid layer on the cold surface. The eluates from the chromatograph were retained in sequence in the track of frozen argon on the surface of the rotating cryodisk. Each collected compound was contained within an area typically 0.3 mm in diameter yielding a high level of spectral absorbance per mass unit of sample. Measurement of IR spectra by specular reflectance from the metallic surface was made by rotating the cryodisk after deposition was completed so that the portion of the solid argon matrix that contained components of interest was placed at the focus of the IR beam. Computer-controlled disk motion facilitated this operation, as did retention time data obtained with the flame ionization detector (FID) of the GC by dividing the

GC effluent in two parts, but GC/MI/FTIRS using this interface could not be considered as a real-time technique, as post-GC run operation was necessary. However, once a compound was deposited, extensive signal-averaging could be done if necessary to enhance the S/N ratio of spectra obtained for diluted samples or poorly absorbing molecules. The deposited compounds were removed from the cryogenic sample collector by warming it up after completion of the data collection.

These results demonstrated the achievement of MI of molecules eluting from a high-resolution gas chromatograph without any loss of separation power.[27] The gain in sensitivity was estimated to attain a factor between 10 and 100 compared with LP GC/FTIRS.[20] The GC/MI/FTIR sensitivity was found to be quite similar to the current GC/MS sensitivity (i.e. low nanogram range). Both detection modes were implemented in parallel on some Mattson-Cryolect™ commercial instruments by dividing the effluent from the GC into two parts. This rendered the instrument highly efficient with two compatible and highly complementary identification techniques. Figure 10 shows an example of MI spectra obtained in the original work with ca. 60 ng of dioxin samples.

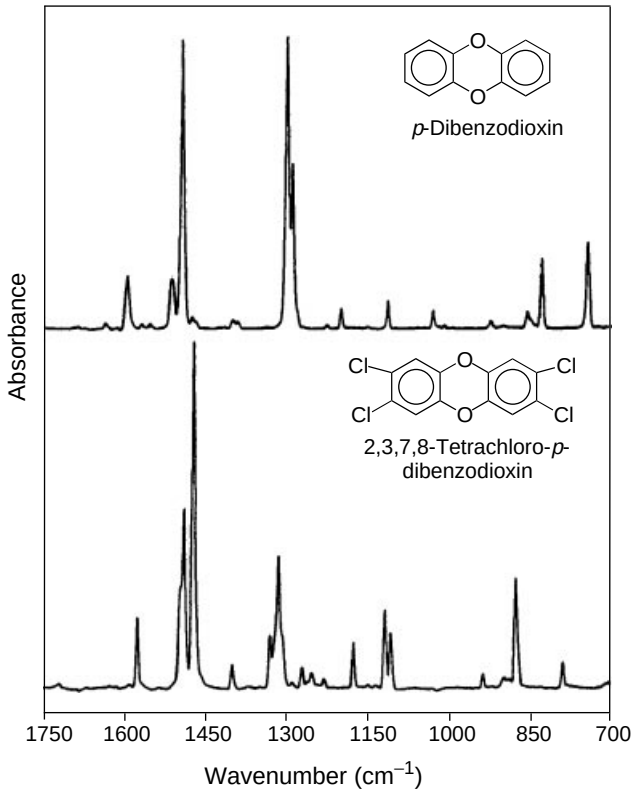

Figure 10 GC/MI/FTIR spectra for ca. 60 ng of *p*-dibenzodioxin and 2,3,7,8-tetrachloro-*p*-dibenzodioxin. (Reproduced by permission of the American Chemical Society from Reedy et al.[27])

Specificity of MI IR spectra, particularly sharp absorption bands, is due to the absence of band broadening following molecular rotation or intermolecular interaction. These spectra can be considered as pure IR vibrational spectra. This specificity allows straightforward discrimination among closely related isomers, discrimination simply impossible to achieve with the LP interface, or even with traditional IR spectra obtained in the condensed phase.[20,27]

The major disadvantages of the GC/MI/FTIR technique are its complexity, the hardware being more sophisticated and requiring more precise construction than the LP interface, and its higher cost. Moreover, the MI interface requires intensive training to maintain and operate, particularly for the optical bench alignment procedures.[28] Another drawback is that argon matrix spectra are different from the conventional spectra obtained in condensed phase and also different from gas-phase spectra. Moreover no commercial library is available with spectra acquired under these conditions. However, spectral specificity with subsequent unparalleled identification power, sensitivity enhancements due to the concentration of the analyte, extensive signal averaging and peak sharpening, and complete compatibility with demanding high-resolution capillary GC are unique features of the MI interface. Unfortunately, for some years the Mattson-Cryolect™ instrument has not been commercially available. A new instrument (vide infra) is available in the US, the MI-TREK™ commercialized by Reedy Scientific Instruments (Bourbonnais, IL).

A major drawback of MI technology is that it operates at the very low temperature of 12 K. Efforts have been made in order to increase the working temperature to the liquid nitrogen temperature using matrixless deposition of eluates at 77 K or MI in xenon at 65 K.[29] Subambient temperatures (−10 to −45 °C) have also been investigated using eluate trapping on an IR-transparent window moved to the focal plane of an FTIR microscope which allowed IR data collection shortly after the eluate was deposited.[30,31] By placing the IR-transparent window in a vacuum chamber and adapting the optics, it was possible to obtain IR transmission spectra of eluates trapped at the liquid nitrogen temperature (77 K) in ''pseudo real time'' with subnanogram sensitivity.[32] This work was the basis for the development of a commercial real-time DD interface for GC/FTIRS.

3.3 The Direct Deposition Interface

Performance characteristics of the DD interface commercialized by Bio-Rad Laboratories as the Tracer™ GC/IR interface were described by Bourne et al.[33] in 1990. A schematic of the optics, sample collection device and

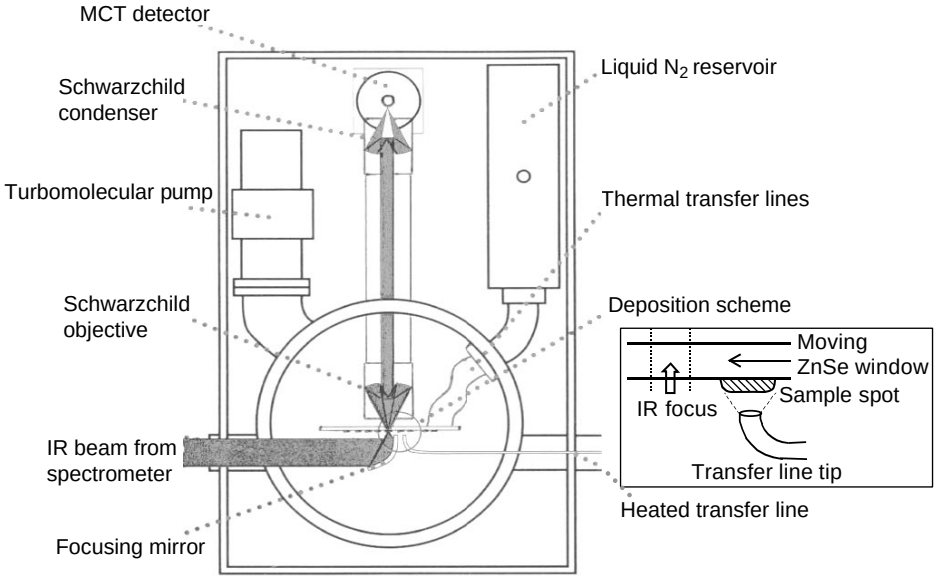

Figure 11 Layout of the commercial DD GC/FTIR interface (Tracer™) with eluate deposition scheme. (Reproduced by permission of Bio-Rad SA.)

the layout of the commercial instrument is presented in Figure 11. The sample slide, mounted on a copper cold-block, is a rectangular ZnSe window transparent to the IR. This assembly is mounted on a motorized XY stage inside a vacuum chamber maintained at 10^{-5} Torr by a turbomolecular pump. The ZnSe window is cooled with flexible thermal transfer lines placed between the moving cold-block and a liquid nitrogen reservoir (Figure 11). With this arrangement the temperature of the window never exceeds 97 K.[33] The heated transfer line from the GC incorporates an orifice of 50 μm internal diameter positioned 50 μm from the surface of the window, close to the focus point of the IR beam. The analytes in the effluent are deposited on the window as small frozen spots of approximately 100 μm width. The window is moved continuously step by step, and each spot passes through the IR beam just after a short time delay, making the technique a "quasi" real-time one (Figure 11). The IR beam is transmitted through the sample and window, and collected for transmission to an MCT detector by a Schwartzchild microscope attachment. As the position and movement of the ZnSe window are under computer

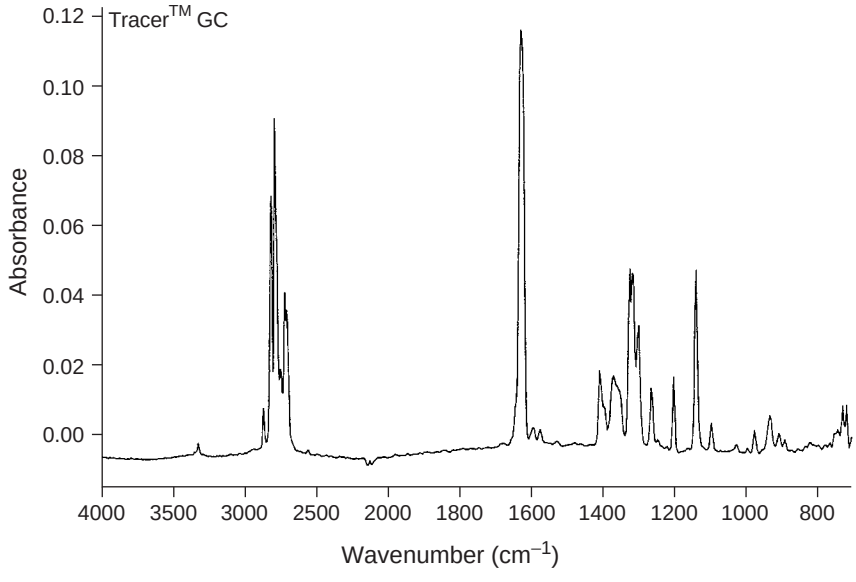

Figure 12 GC/DD/FTIR real-time spectrum of ca. 50 ng of 2-hexanone present in a complex mixture of volatile compounds.

control, the window may be repositioned for post-run IR measurements with extended signal-averaging for improved S/N ratios, as they are with the MI system. The IR spectra are obtained in the condensed (solid) phase, in the traditional transmittance mode. An example is presented in Figure 12. It has been demonstrated that DD spectra may be directly searched against conventional libraries of IR spectra of reference materials prepared as KBr disks.[33,34] Low subnanogram sensitivity down to a few tens of picograms has been demonstrated with this GC/DD/FTIR interface for a variety of chemicals with real-time measurement.[33,34]

3.4 Gas Chromatography/Fourier Transform Infrared Spectra and Sensitivity

The three different GC/IR interfaces based on three different principles lead to different IR spectra. The heated LP interface, the MI cryogenic system and the DD device give gas-phase, matrix-isolated and solid-phase spectra, respectively. This implies that the different spectra acquired in the different modes are significantly different in terms of absorption band intensity and width, and in terms of band position. A shift to higher frequencies

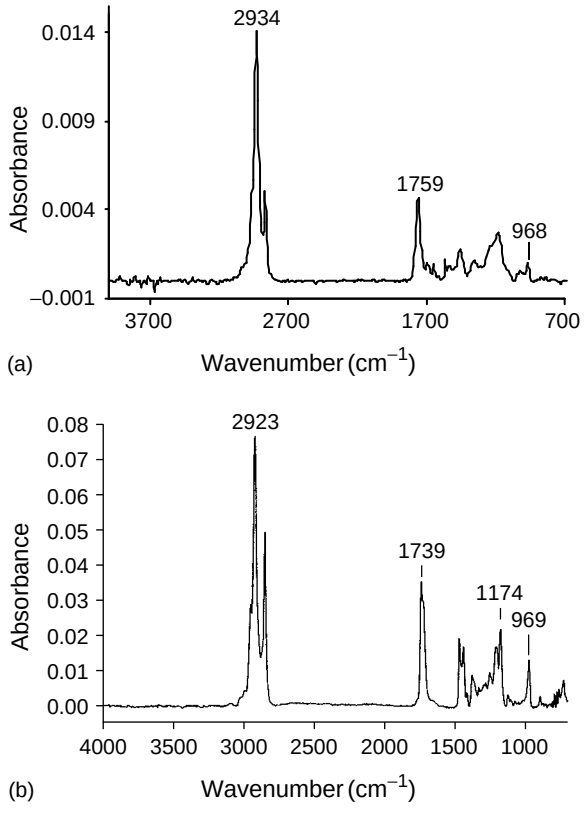

Figure 13 GC/FTIR spectra of methyl elaidate obtained in the vapor phase with an LP interface (a) and in the solid with a DD interface (b).

is noticed on changing from solid to liquid to gaseous state. This change is accompanied by a general narrowing and an intensity decrease for the bands involved in hydrogen bonding in the condensed phase. Gas-phase spectra give broad absorption bands due to enhanced rotational motion in the heated LP. For important characteristic bands that could be weak, this is particularly a handicap. On the other hand, solid-phase spectra obtained with the DD interface give rise to sharper and more intense bands. For comparison purposes, Figure 13 shows the IR spectra of a $C_{18:1}$ unsaturated fatty acid methyl ester possessing a *trans* double bond, methyl elaidate. The position of the main absorption bands (C−H stretching and carbonyl stretching) are different, shifted towards lower frequencies by ca. 10 cm^{-1} for C−H stretching from 2933 in the gas phase to 2923 cm^{-1} in the solid and 20 cm^{-1} for the C=O stretching from 1759 in the vapor phase to 1739 cm^{-1} in the solid state. The respective ratios between these bands are also clearly different. On the contrary, the C−H out-of-plane deformation band, characteristic of the trans double bond, is found at the same wavenumber (968–969 cm^{-1}) for both modes but with a clear advantage for the solid-state spectrum from the intensity standpoint. (This characteristic band which appears invariant is found at the same position around 970 cm^{-1} in MI spectra.[35])

MI spectra are different from both gas-phase and solid-state spectra. As already outlined, they are pure vibrational spectra, resulting with sharp absorption bands. Practically, as it is shown in Figure 14, they can be considered as intermediates between gas-phase spectra (in terms of band position) and solid-phase spectra (in terms of band intensity). For instance, the matrix isolated spectrum of a conjugated *trans,trans*-$C_{18:2}$ diunsaturated fatty acid methyl ester displays C−H and carbonyl stretchings at 2935 and 1754 cm^{-1}, respectively, close to the corresponding gas-phase features, but the respective intensities are those of the solid-state spectrum (Figure 14b) in which however the bands are found at lower frequencies (2923 and 1739 cm^{-1} respectively). On the other hand, the C−H out-of-plane deformation band, characteristic of the conjugated *trans,trans* configuration, is found at exactly the same frequency (990 cm^{-1}) with comparable relative intensities.

Another structural feature worth mentioning here is the hydroxyl stretching vibration of alcohols. As in the solid phase obtained with the DD interface intermolecular forces such as hydrogen bonding are present, the OH stretching bands are significantly broadened, as they are with KBr disks or neat films. Therefore the OH stretching band appears as a very broad band between 3500 and 3000 cm^{-1} (Figure 15). MI spectra of alcohols display intense and sharp absorption bands for the OH

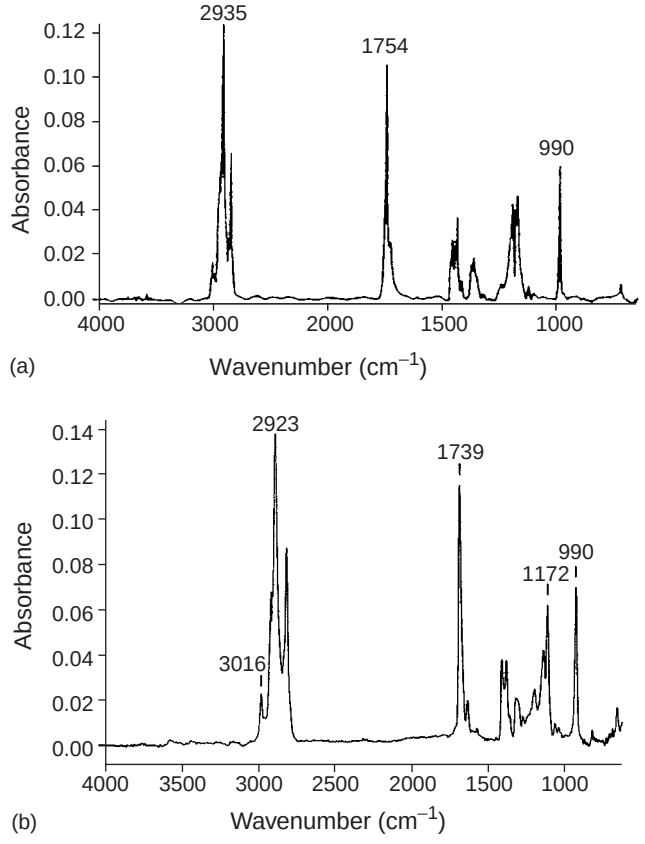

Figure 14 GC/FTIR spectra of a conjugated *E*,*E*-C18 : 2 methyl ester obtained in an argon matrix with an MI interface (a) and in the solid state with a DD interface (b). (Spectrum (a) reproduced by permission of Preston Publications from Mossoba et al.[82])

stretching, resulting from the collection of the separated compounds under MI conditions, that are highly structurally informative.[24] Vapor-phase spectra of alcohols display weak bands for the OH stretching (Figure 7 and Figure 15) revealing the lack of intermolecular interaction, but their positions are also highly structurally informative (i.e. it is easily possible to distinguish between primary, secondary and tertiary alcohols[36]). Moreover, as rotational motion is allowed in the heated interface, intramolecular interactions such as intramolecular hydrogen bonding are allowed, resulting in a significant decrease in the OH absorption frequency, a decrease of high structural significance. Thus, the OH stretching vibration of the compound presented in Figure 15 found at 3611 cm^{-1} could be attributed to a hydroxyl function of a secondary alcohol involved in intramolecular hydrogen bonding (with a carbonyl function revealed by the carbonyl band found at 1738 cm^{-1}) and this observation was of prime importance for the structural characterization of this compound found in a pheromonal secretion of a cockroach and identified as 3-hydroxyhexanal.[37] Another example of the importance of such an observation, a unique feature of gas-phase IR spectra of alcohols, will be presented below (see section 4.3).

As already outlined, only the cryogenic interfaces such as the MI or DD devices possess intrinsic sensitivity fully compatible with demanding GC analyses that can be easily handled by GC/MS. This has been fully documented (see Le Quéré and Sémon[22] for example). Very low detection limits (to the picogram range) have been published for both cryogenic interfaces,[28,33] but these limits have been obtained with extensive signal-averaging in post-GC runs, and not on a regular routine basis. Nevertheless the cryogenic interfaces have a detection limit of one or two orders of magnitude lower than that of the LP interface. For the latter, it seems that no significant improvement should be obtained, due to the technique itself.

One of the claimed advantages of the DD interface is that the spectra obtained in the solid state may be searched in conventional IR spectral libraries, where spectra are generally obtained as KBr disks.[33] Spectra obtained by this interface are characteristic of a solid-phase sample in an amorphous state. However some typical crystalline features such as band splitting and fringing effects can be noticed. Significant differences between the two types of spectra could be observed, and some bands, like the carbonyl ones, are shifted at lower frequencies by about 10 cm^{-1} for solid-state spectra obtained at cryogenic temperature.[38] Moreover, water introduced into the system either by the sample or as a result of poor conditioning gives solid ice on the window, which results in band broadening. It appears necessary to dry the chamber under vacuum for 1 day after opening it to the atmosphere for routine maintenance.[38] In their original work, Bourne et al.[33] demonstrated the possibility of searching a vapor-phase library for nonpolar compounds such as chlorinated pesticides when only the fingerprint region is included for the search. However when the analytes contain some moieties that may be involved in intermolecular interactions such as hydrogen bonding, it is advisable to use a database for spectral searching where the reference materials are in a similar state to the unknown analytes.[33]

Vapor-phase reference spectra are available in various libraries such as the Environmental Protection Agency library of vapor-phase IR spectra, the Sadtler–Heyden spectra,[36] the Aldrich library of FTIR spectra[39] and the Sadtler vapor-phase IR database (Bio-Rad Laboratories, Sadtler Division). To our knowledge, no MI reference spectral library is commercially available. Table 1 summarizes a comparison of the characteristics of the three techniques discussed in this paragraph.

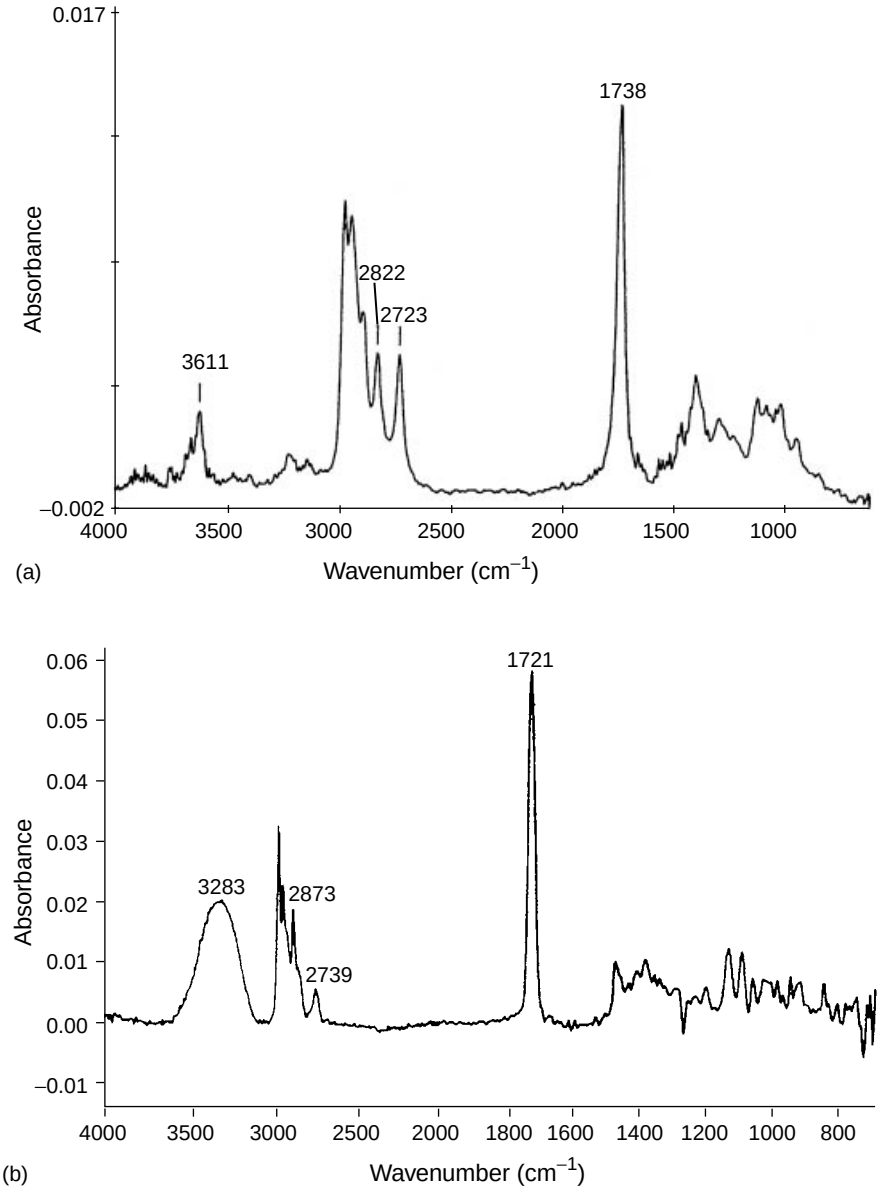

Figure 15 GC/FTIR spectra of 3-hydroxyhexanal obtained in the gas phase (a) with an LP and in the solid state (b) with the DD interface. (a) was obtained with ca. 1 µg injected on column. (b) was obtained with 50 ng injected splitless.

Table 1 Characteristics comparison of GC/FTIR interfaces

Characteristic	MI	DD	LP
Limit of detection	pg level	pg level	ng level
Limit of identification	<1 ng	<1 ng	10 to 1000 ng, depending on volatility and polarity
Cost	+++	++	+
Sample processing	Two-step process Post-run signal-averaging	Pseudo real-time technique Post-run signal-averaging	"On-the-fly", true real time No post-run signal-averaging
Specificity	Matrix-isolated spectra Sharp absorption bands	Solid-state spectra	Vapor-phase spectra Broad absorption bands
Spectral search	No commercial libraries	Universal condensed-phase spectra libraries available	Vapor-phase libraries available

4 SOME APPLICATIONS OF GAS CHROMATOGRAPHY/FOURIER TRANSFORM INFRARED SPECTROMETRY

It is beyond the scope of this article to give an exhaustive overview of the applications of GC/FTIR in analytical chemistry. In fact, all the problems amenable to GC may use FTIR as a powerful detector. However in order to illustrate the capabilities of the technique, some selected examples from selected domains will be presented. An early comprehensive review may be found in a specialized treatise[5] and dedicated contributions appear periodically in specialty journals.

4.1 Environmental

Substantial efforts have been made to apply GC/FTIRS for qualitative and quantitative environmental analysis of typical environmental contaminants.[40,41] The claimed advantages were specificity and selectivity, IR being able to distinguish between isomers, and the complementary nature of GC/FTIRS and GC/MS was soon acknowledged.[41] Efforts made to improve the performance of GC/FTIR interfaces, particularly the LP interface, were very often motivated by environmental issues (see Gurka et al.[42] for example). This system finally exhibited minimum identifiable quantities in the 10–50 ng range for standards, soil samples and a herbicide still-bottom.[43] Quantitative capabilities of GC/FTIRS were also clearly demonstrated for priority pollutants.[44]

Cryogenic GC/IR interfaces showed improved sensitivity: minimum identifiable quantities ranged from 0.5 to 2 ng in one study on polynuclear aromatic hydrocarbons and pesticides using the DD interface.[45] As an example Figure 16 shows the GC/FTIR spectrum of 2 ng DDT (dichlorodiphenyltrichloroethane) obtained on-the-fly with this interface. A sensitivity down to 100 pg has been demonstrated with this interface for on-the-fly detection of hexachlorobenzene.[46] A picogram level quantitation of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in fish extracts has also been reported for the MI interface.[28]

4.2 Flavor and Fragrances

Identification of flavor and fragrance molecules is a major field of application for GC/IR. The complementary nature of GC/IR and GC/MS for the structural analysis of volatile flavor compounds was recognized very early on, with particular emphasis on the unique ability of IR spectroscopy to distinguish individual isomers.[7,21,47,48] The early literature through to the end of 1985 has been covered,[5] with applications to the aroma composition of tropical fruits,[49–51] alcoholic beverages,[52] and the volatiles emanating from intact fruits of *Coffea arabica*.[53] The technique has been demonstrated to be very powerful for the analyses of essential oils, and especially for the structural investigation of terpenes and sesquiterpenes where unequivocal identification by MS alone is very often difficult and when FTIR delivers spectra with clear fingerprints.[11,54–58] Recent investigations used the DD interface for increased specificity and sensitivity, where IR spectra matched the documented literature spectra obtained as liquid or KBr pellets.[59] The usefulness of the MI technique for the analyses of flavor extracts has also been demonstrated.[60]

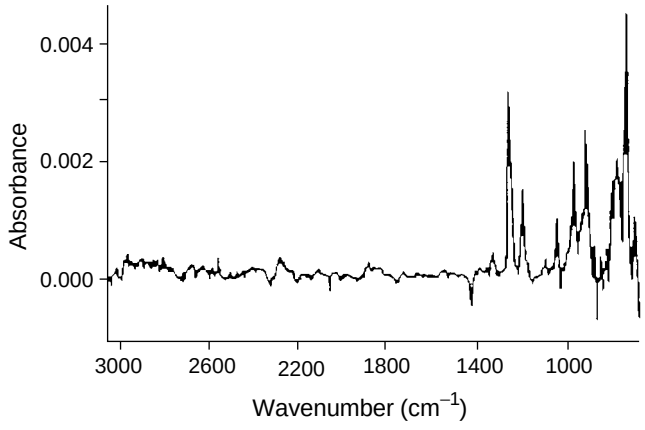

Figure 16 GC/DD/FTIR spectrum of 2 ng DDT. (Reproduced by permission of the American Chemical Society from Gurka et al.[45])

In flavor research coupled GC/FTIRS in conjunction with GC/MS is now a routine tool for characterizing the structure of the volatile components.[22,61,62] The LP system is very often chosen for its ruggedness and low cost, even though the cryofocusing devices present an intrinsic sensitivity one or two orders of magnitude lower.

4.3 Pheromones

Insect pheromones or plant-emitted substances involved in insect–plant interaction are other volatile components that have benefited from the complementary nature of GC/MS and GC/FTIRS. Recent work in the domain uses the sensitivity advantage of the DD interface (see Ferary et al.[63] for example) and coupling with solid-phase microextraction (SPME) adds some advantages.[38] However, a lot of results have been obtained in the vapor phase with the LP interface.[37,64–67] Figure 17 shows the IR spectrum of a female-specific compound isolated from the tergal gland secretion of a primitive cockroach *Cryptocercus punctulatus*. This primitive woodroach lives in galleries in rotten wood and it is the best living representative of the ancestral blattarian-isopteran stock. Besides, several aspects of its life and morphology

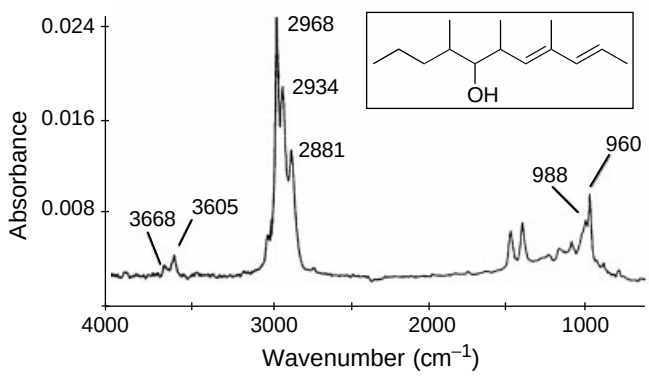

Figure 17 GC/LP/FTIR spectrum of 4,6,8-trimethyl-7,9-undecadien-5-ol from the tergal gland of the primitive cockroach *Cryptocercus punctulatus*. (Reproduced by permission of Plenum Press from Le Quéré et al.[64])

strongly resemble the termites[64] and a better knowledge of its biology is of fundamental interest.

The IR spectrum confirmed the presence of a hydroxyl group in the pheromonal compound, the OH stretching frequency (3668 cm^{-1}) falling in the range of secondary alcohols.[36] Another stretching frequency (3605 cm^{-1}) revealed intramolecular hydrogen bonding, probably resulting from the interaction of one of the carbon–carbon unsaturations present in the molecule, via a $(5 + \pi)$ intramolecular hydrogen-bonded system.[36] This feature, unique to vapor-phase spectra as already outlined, fixes the position of the hydroxyl group relative to the carbon–carbon double bond, and is therefore highly structurally informative. The medium absorption at 960 cm^{-1} was attributed to a CH out-of-plane deformation of a trans disubstituted ethylenic bond. These data greatly helped to propose the structure 4,6,8-trimethyl-7,9(*E*)-undecadien-5-ol for the pheromonal compound.[64] This hypothesis was confirmed later with the synthetic compound.

4.4 Fatty Acids

IR spectroscopy has long been used as a dedicated method to distinguish between cis and trans isomers and to determine the total amount of trans fatty acids in fats and oils. The IR analysis of individual isomers in complex mixtures of unsaturated fatty acids was made possible through the combined GC/FTIR technique. The first example published in the literature described the analysis of unsaturated cyclic fatty acid mixtures isolated from heated oils using an LP interface.[68] This interface type was extensively used to characterize geometrical isomers of polyunsaturated fatty acids (PUFAs) in various applications.[69,70] However, it was soon recognized that greater instrumental sensitivity compatible with HRGC was required to determine the minor components found in mixtures with extremely unbalanced dynamic range of concentration (i.e. levels of minor components such as trans PUFAs less than 1%). In this respect, the MI interface was successfully used by Mossoba et al.[35,71] Examples of characteristic spectra have already been presented above. More recently, it has been demonstrated that the DD interface is also particularly well suited to the study of complex fatty acid mixtures.[72]

5 PERSPECTIVE

GC/FTIRS is a valuable tool for the analytical chemist whose main concern is the structural determination of volatile compounds or any component that can be separated by GC. LP technology has reached a standard of achievement limited in terms of sensitivity and specificity, but still very useful and easy to use and maintain. The cryogenic effluent-trapping techniques, i.e. MI and DD, despite their intrinsic sensitivity are not common instruments in analytical laboratories, probably due to their cost and to their inherent sophistication. The MI Cryolect™-Mattson instrument has not been available for some years, and, to our knowledge, the DD Tracer™ Bio-Rad instrument is no longer commercially available. A new MI system appeared recently, the MI-TREK™ (Reedy Scientific Instruments, Bourbonnais, IL), for which a comprehensive review has been published.[71] Based on the original design (vide supra), the MI-TREK™ has however distinct advantages that will be depicted in the following text.

The interface between the IR spectrometer and the GC is a compact vacuum chamber containing the sample collector and the optics necessary to focus the IR beam on the surface of the collector (Figure 18). The IR beam coming from the interferometer is deflected by a plane mirror to enter the vacuum chamber. Within this vacuum chamber it reflects from a plane mirror towards the first of twin paraboloidal mirrors attached to a single block of aluminum. The two paraboloids have a common focus on the mirrored surface of the collector where the samples are matrix-isolated. The IR beam reflects to the second twin mirror, and then the beam is collimated to the paraboloid mirror at the detector. Contrarily to the original design, the collector disk is in a vertical position (Figure 19). It rotates counter-clockwise during sample collection. This arrangement allows the collection surface where the samples are trapped in a layer of argon matrix to move from the ''sample delivery point'' to the ''IR scan point'' within a short delay, making the technique a ''pseudo real-time'' one like the DD interface. Cooling for the cryogenic surface is provided by a closed-cycle helium refrigeration system which requires no liquid helium.

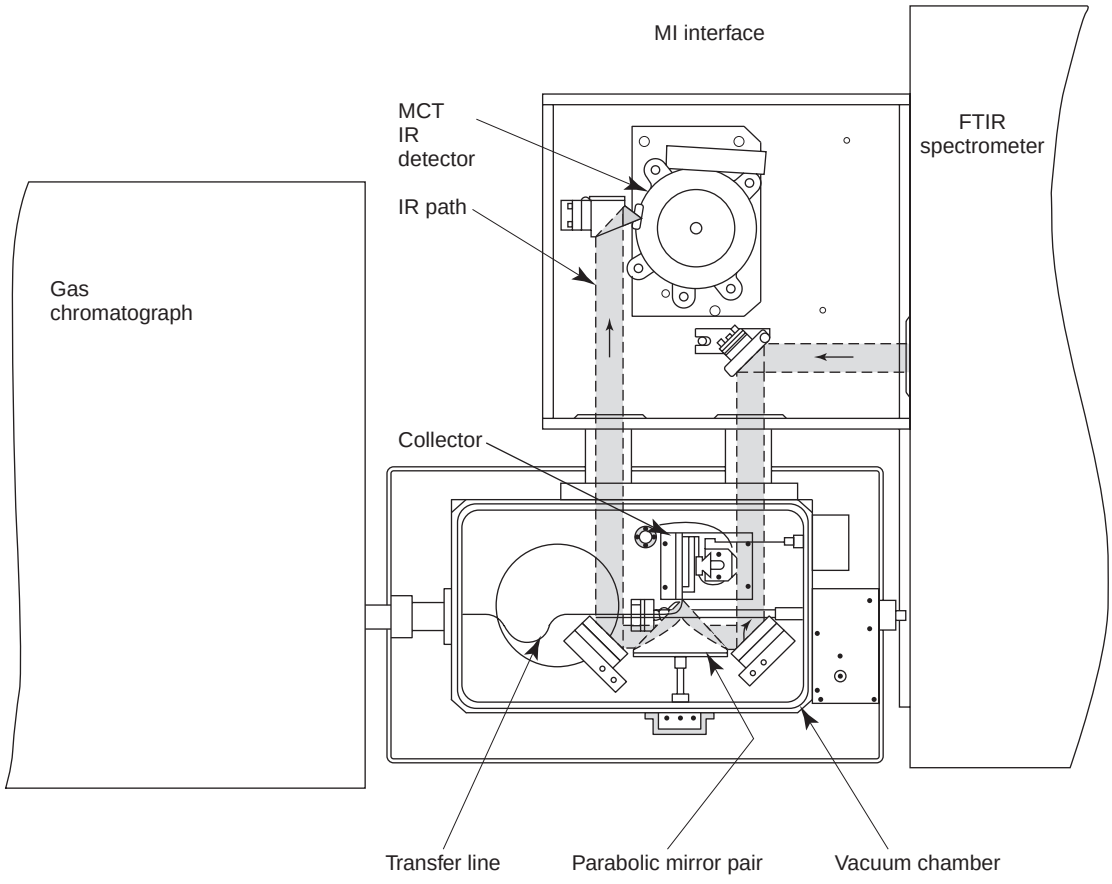

Figure 18 Layout of the MI MI-TREK™ GC/FTIR system. (Reproduced by permission of Marcel Dekker Inc. from Reedy et al.[71])

The refrigerator provides two stages within the vacuum chamber. The first stage operates at 50 K and provides cooling for a reflector shield (Figure 19) which reflects environmental radiant heat and collects background gases such as H_2O and CO_2. The sample collector itself is supported and cooled by the second stage of the refrigerator operating at 10 K.[71] The MI-TREK™ system has been designed to incorporate parallel MS detection by dividing the GC effluent in two parts, with ca. 25% directed to the MS detector. The intrinsic sensitivity of the cryogenic GC/IR interface makes this double-hyphenated technique useful. In fact, all the cryogenic interfaces operating under vacuum are fully compatible with MS, which also operates under vacuum. However simultaneous detection is only possible by dividing the GC effluent in two parts. Contrarily, as the LP system allows on-the-fly IR detection, it is possible to attach an MS detector at the exit of the LP. But as the MS detector is far more sensitive than the IR one, it is also necessary to split the GC effluent (ca. 1/14) to the MS.[44] Many efforts have been made to develop GC/FTIR/MS systems,[73,74] but this double-hyphenated technique is not widespread despite its potential.

Among the technical developments on the combination of chromatography and FTIR spectrometry, the coupling of high-performance liquid chromatography (HPLC) and supercritical fluid chromatography (SFC) to a FTIR spectrometer has received considerable attention. Specific problems related to mobile phases used in HPLC arose with HPLC/FTIRS: no solvent used within the mobile phases is transparent to IR radiation. The use of a flow cell interface necessitates a limitation of the pathlength, resulting in a significant limitation on the sensitivity. Several approaches to eliminate the solvent prior to the IR measurement have been described[31] and subnanogram detection limits have been reported even for reversed-phase HPLC/FTIR using a DD interface.[75] Like GC/FTIRS and HPLC/FTIRS, both flow cell and mobile-phase elimination devices have been described for SFC/FTIRS. Despite the fact that flow cell must accommodate very high pressures,[76] many results have been obtained with this technology, essentially because the transmittance of supercritical CO_2, the most popular fluid used for SFC, is significantly greater than that of most HPLC mobile phases.[31] The method was successfully applied in various fields such

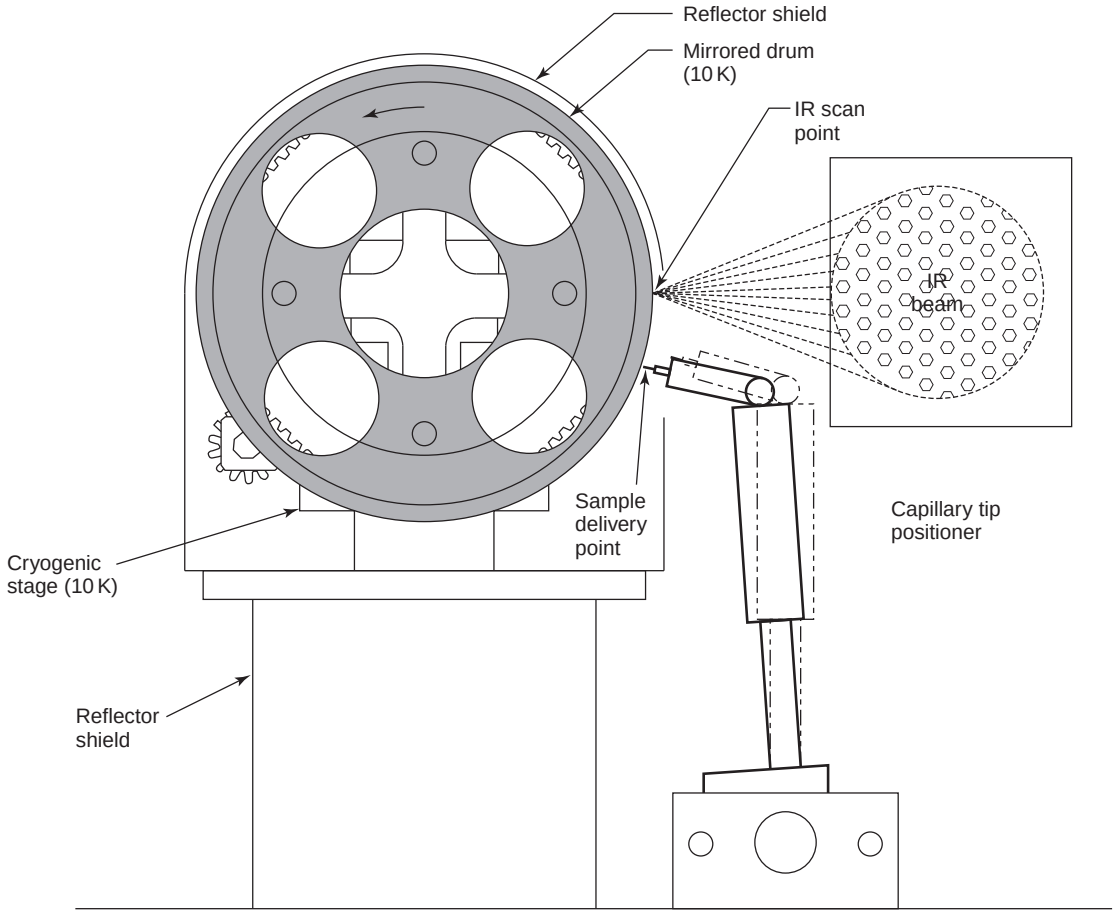

Figure 19 Enlarged view of the sample deposition device of the MI-TREK™ GC/MI/FTIR instrument, showing the sample delivery point and IR scan point. (Reproduced by permission of Marcel Dekker Inc. from Reedy et al.[71])

as flavor research,[77,78] lipids[79] with the benefit of analyzing free fatty acids and triglycerides using the same chromatographic column,[80] monosaccharides, mycotoxins and food additives.[71] Despite significant sensitivity improvements that have been obtained with flow cells (see Wieboldt et al.[81] for example), subnanogram detection limits have been only achieved with a DD interface that eliminates the supercritical mobile phase.[75] On-line SFC/DD/FTIR spectra have been obtained by slightly modifying the DD device used for GC/FTIR, using a window temperature close to room temperature.[75] Excellent results have been obtained with this technique in the analysis of poly(ethylene glycol)s of various average molecular weights.[45] In this mode, as with the GC/DD/FTIR interface, the spectra are obtained in the condensed state, and they may be searched against a conventional IR spectral database.

In conclusion, after this short overview of technique developments, it is clear that the use of FTIR provides sensitive detection of species eluting from a gas chromatograph, and delivers positive identification through the observation of characteristic spectral information. In particular, GC/FTIRS is able to differentiate closely related isomers, a unique feature of IR spectroscopy. Therefore, in analyzing volatile or semi-volatile compounds, GC/FTIR appears to be an essential complement to GC/MS.

ACKNOWLEDGMENTS

The Institut National de la Recherche Agronomique (INRA) and the Conseil Régional de Bourgogne are thanked for financial support. The author is indebted to Christian Giniès and Etienne Sémon for skillful operation of the GC/FTIR instruments at INRA, Laboratoire de Recherches sur les Arômes, and for interesting discussion.

ABBREVIATIONS AND ACRONYMS

DD	Direct Deposition
DDT	Dichlorodiphenyltrichloroethane

DTGS	Deuterated Triglycine Sulfate
FFT	Fast Fourier Transform
FID	Flame Ionization Detector
FTIR	Fourier Transform Infrared
FTIRS	Fourier Transform Infrared Spectrometry
GC	Gas Chromatography
GC/FTIRS	Gas Chromatography/Fourier Transform Infrared Spectrometry
GC/IR	Gas Chromatography/Infrared
GC/MS	Gas Chromatography/Mass Spectrometry
GS	Gram–Schmidt
HPLC	High-performance Liquid Chromatography
HRGC	High-resolution Gas Chromatography
IR	Infrared
LP	Light Pipe
MCT	Mercury Cadmium Telluride
MI	Matrix Isolation
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PUFA	Polyunsaturated Fatty Acid
SFC	Supercritical Fluid Chromatography
SIM	Single-ion Monitoring
S/N	Signal-to-noise
SPME	Solid-phase Microextraction
TIC	Total Ion Current
WCOT	Wall-coated Open Tubular

RELATED ARTICLES

Environment: Trace Gas Monitoring **(Volume 3)**
Matrix Isolation Spectroscopy in Atmospheric Chemistry

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Selective Detectors for Amines • Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Selective Detectors

Food **(Volume 5)**
Flavor Analysis in Food • Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Pesticides **(Volume 7)**
Gas Chromatography and Supercritical Fluid Chromatography with Selective Detection in Pesticide Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Hyphenated Gas Chromatography

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Liquid Chromatography/Infrared Spectroscopy • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry

REFERENCES

1. S.K. Freeman, 'Gas Chromatography and Infrared and Raman Spectrometry', in *Ancillary Techniques of Gas Chromatography*, eds. L.S. Ettre, W.H. McFadden, Wiley-Interscience, New York, 227–267, 1969.
2. R.H. Shaps, A. Varano, 'Stop-flow GC/IR Analysis', *Industr. Res.*, **2**, 86–89 (1977).
3. M.J.D. Low, 'Infrared Fourier Transform Spectroscopy in Flavor Analysis. IV. Spectra of Gas Chromatography Fractions', *J. Agric. Food Chem.*, **19**(6), 1124–1127 (1971).
4. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Interscience, New York, 1986.
5. W. Herres, *HRGC/FTIR: Capillary Gas Chromatography-Fourier Transform Infrared Spectroscopy. Theory and Applications*, Hüthig, Heidelberg, 1987.
6. P.R. Griffiths, J.A. de Haseth, L.V. Azarraga, 'Capillary GC/FTIR', *Anal. Chem.*, **55**(13), 1361A–1387A (1983).
7. W. Herres, 'Capillary GC/FTIR Analysis of Volatiles: HRGC/FTIR', in *Analysis of Volatiles*, ed. P. Schreier, Walter de Gruyter, Berlin, 183–217, 1984.
8. J.A. de Haseth, T.L. Isenhour, 'Reconstruction of Gas Chromatograms from Interferometric Gas Chromatography/Infrared Spectrometry Data', *Anal. Chem.*, **49**(13), 1977–1981 (1977).
9. D.A. Hanna, G. Hangac, B.A. Hohne, G.W. Small, R.C. Wieboldt, T.L. Isenhour, 'A Comparison of Methods Used for the Reconstruction of GC/FTIR Chromatograms', *J. Chromatogr. Sci.*, **17**(8), 423–427 (1979).
10. P. Coffey, D.R. Mattson, J.C. Wright, 'A Programmable GC/FTIR System', *Am. Lab.*, **10**(5), 126–132 (1978).
11. W. Herres, W. Schultze, K.H. Kubeczka, 'Analysis of Essential Oils by HRGC/FTIR: Measurement of Unresolved GC Peaks', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **9**(8), 466–468 (1986).
12. J.M. Bjerga, G.W. Small, 'Reconstruction of Gas Chromatograms from Digitally Filtered Fourier Transform

Infrared Interferograms', *Anal. Chem.*, **61**(10), 1073–1079 (1989).
13. M.J.D. Low, S.K. Freeman, 'Measurement of Infrared Spectra of Gas–Liquid Chromatography Fractions Using Multiple-scan Interference Spectrometry', *Anal. Chem.*, **39**(2), 194–198 (1967).
14. L.V. Azarraga, 'Gold Coating of Glass Tubes for Gas Chromatography/Fourier Transform Infrared Spectroscopy ''Light-Pipe'' Gas Cells', *Appl. Spectrosc.*, **34**(2), 224–225 (1980).
15. K.H. Shafer, A. Bjorseth, J. Tabor, R.J. Jakobsen, 'Advancing the Chromatography of GC/FTIR to WCOT Capillary Columns', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **3**(2), 87–88 (1980).
16. R.S. Brown, J.R. Cooper, C.L. Wilkins, 'Lightpipe Temperature and Other Factors Affecting Signal in Gas Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chem.*, **57**(12), 2275–2279 (1985).
17. D.E. Henry, A. Giorgetti, A.M. Haefner, P.R. Griffiths, D.F. Gurka, 'Optimizing the Optical Configuration for Light-pipe Gas Chromatography/Fourier Transform Infrared Spectrometry Interfaces', *Anal. Chem.*, **59**(19), 2356–2361 (1987).
18. D. Kuehl, G.J. Kemeny, P.R. Griffiths, 'Identification of Peaks in Capillary Column Gas Chromatograms at the Nanogram Level by Dual-beam Fourier Transform Infrared Spectrometry', *Appl. Spectrosc.*, **34**(2), 222–224 (1980).
19. J.R. Cooper, L.T. Taylor, 'Establishing Identification Limits of Model Compounds in Capillary Gas Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chem.*, **56**(11), 1989–1993 (1984).
20. J.F. Schneider, J.C. Demirgian, J.C. Stickler, 'A Comparison of GC/IR Interfaces: The Light Pipe Versus Matrix Isolation', *J. Chromatogr. Sci.*, **24**(8), 330–335 (1986).
21. J.L. Le Quéré, 'Le Couplage CPG/IRTF dans l'Analyse des Arômes', *Analusis*, **20**(7), M46–M49 (1992).
22. J.L. Le Quéré, E. Sémon, 'Les Couplages CPG/SM et CPG/IRTF, Deux Techniques Complémentaires dans l'Analyse Structurale des Substances Volatiles', *Analusis*, **24**(4), M20–M22 (1996).
23. L. Moio, E. Sémon, J.L. Le Quéré, '3-Hydroxy-5-methyl-2-hexanone, a New Compound Characterized by a Melted Cheese Flavor in Dairy Products', *Ital. J. Food Sci.*, **6**(4), 441–447 (1994).
24. G.T. Reedy, S. Bourne, P.T. Cunningham, 'Gas Chromatography/Infrared Matrix Isolation Spectrometry', *Anal. Chem.*, **51**(9), 1535–1540 (1979).
25. S. Bourne, G.T. Reedy, P.T. Cunningham, 'Gas Chromatography/Matrix Isolation/Infrared Spectroscopy: An Evaluation of the Performance Potential', *J. Chromatogr. Sci.*, **17**(8), 460–463 (1979).
26. D.M. Hembree, A.A. Garrison, R.A. Crocombe, R.A. Yokley, E.L. Wehry, G. Mamantov, 'Matrix Isolation Fourier Transform Infrared Spectrometric Detection in the Open Tubular Column Gas Chromatography of Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **53**(12), 1783–1788 (1981).
27. G.T. Reedy, D.G. Ettinger, J.F. Schneider, S. Bourne, 'High-resolution Gas Chromatography/Matrix Isolation Infrared Spectrometry', *Anal. Chem.*, **57**(8), 1602–1609 (1985).
28. M.M. Mossoba, R.A. Niemann, J.Y.T. Chen, 'Picogram Level Quantitation of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin in Fish Extracts by Capillary Gas Chromatography/Matrix Isolation/Fourier Transform Infrared Spectrometry', *Anal. Chem.*, **61**(15), 1678–1685 (1989).
29. R.S. Brown, C.L. Wilkins, 'Cryogenically Cooled Interface for Gas Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chem.*, **60**(14), 1483–1488 (1988).
30. R. Fuoco, K.H. Shafer, P.R. Griffiths, 'Capillary Gas Chromatography/Fourier Transform Infrared Microspectrometry at Subambient Temperature', *Anal. Chem.*, **58**(14), 3249–3254 (1986).
31. P.R. Griffiths, S.L. Pentoney, A. Giorgetti, K.H. Shafer, 'The Hyphenation of Chromatography and FTIR Spectrometry', *Anal. Chem.*, **58**(13), 1349A–1366A (1986).
32. A.M. Haefner, K.L. Norton, P.R. Griffiths, S. Bourne, R. Curbelo, 'Interfaced Gas Chromatography and Fourier Transform Infrared Transmission Spectrometry by Eluate Trapping at 77 K', *Anal. Chem.*, **60**(21), 2441–2444 (1988).
33. S. Bourne, A.M. Haefner, K.L. Norton, P.R. Griffiths, 'Performance Characteristics of a Real-time Direct Deposition Gas Chromatography/Fourier Transform Infrared Spectrometry System', *Anal. Chem.*, **62**(22), 2448–2452 (1990).
34. S. Bourne, S.L. Hill, J.R. Powell, D.R. Prasad, K. Krishnan, 'High Sensitivity Fourier Transform Infrared Spectra of Gas Chromatographic Fractions Using a Liquid-nitrogen Cold-trapping Technique', *Analusis*, **18**(9), 24–29 (1990).
35. M.M. Mossoba, R.E. McDonald, J.Y.T. Chen, D.J. Armstrong, S.W. Page, 'Identification and Quantitation of *trans*-9,*trans*-12-Octadecadienoic Acid Methyl Ester and Related Compounds in Hydrogenated Soybean Oil and Margarines by Capillary Gas Chromatography/Matrix Isolation/Fourier Transform Infrared Spectroscopy', *J. Agric. Food Chem.*, **38**(1), 86–92 (1990).
36. R.A. Nyquist, *The Interpretation of Vapor-phase Infrared Spectra*, Sadtler–Heyden, Philadelphia, 1984.
37. J.P. Farine, C. Everaerts, J.L. Le Quéré, E. Sémon, R. Henry, R. Brossut, 'The Defensive Secretion of *Eurycotis floridana* (Dictyoptera, Blattidae, Polyzosterinae): Chemical Identification and Evidence of an Alarm Function', *Insect Biochem. Mol. Biol.*, **27**(6), 577–586 (1997).
38. J. Auger, S. Rousset, E. Thibout, B. Jaillais, 'Solid-phase Microextraction–Gas Chromatography–Direct Deposition Infrared Spectrometry as a Convenient Method for

the Determination of Volatile Compounds from Living Organisms', *J. Chromatogr. A*, **819**, 45–50 (1998).
39. C.J. Pouchert, *The Aldrich Library of FT-IR Spectra, Vol. 3, Vapor Phase*, Aldrich Chemical Co., Milwaukee, 1989.
40. L.V. Azarraga, C.A. Potter, 'An Integrated GC/FT-IR System for the Analysis of Environmental Pollutants', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **4**(2), 60–69 (1981).
41. K.H. Shafer, M. Cooke, F. DeRoos, R.J. Jakobsen, O. Rosario, J.D. Mulik, 'WCOT Capillary Column GC/FT-IR and GC/MS for Identifying Toxic Organic Pollutants', *Appl. Spectrosc.*, **35**(5), 469–472 (1981).
42. D.F. Gurka, R. Titus, P.R. Griffiths, D. Henry, A. Giorgetti, 'Evaluation of an Improved Single-beam Gas Chromatography/Fourier Transform Infrared Interface for Environmental Analysis', *Anal. Chem.*, **59**(19), 2362–2369 (1987).
43. D.F. Gurka, S.M. Pyle, 'Qualitative and Quantitative Environmental Analysis by Capillary Column Gas Chromatography/Lightpipe Fourier Transform Infrared Spectrometry', *Environ. Sci. Technol.*, **22**(8), 963–967 (1988).
44. D.F. Gurka, I. Farnham, B.B. Potter, S. Pyle, R. Titus, W. Duncan, 'Quantitation Capability of a Directly Linked Gas Chromatography/Fourier Transform Infrared/Mass Spectrometry System', *Anal. Chem.*, **61**(14), 1584–1589 (1989).
45. D.F. Gurka, S. Pyle, R. Titus, E. Shafter, 'Direct-deposition Infrared Spectrometry with Gas and Supercritical Fluid Chromatography', *Anal. Chem.*, **66**(15), 2521–2528 (1994).
46. T. Visser, M.J. Vredenbregt, 'Cryotrapping Gas Chromatography–Fourier Transform Infrared Spectrometry in Environmental Analysis: A Pilot Study', *Vibr. Spectrosc.*, **1**, 205–210 (1990).
47. P. Schreier, H. Idstein, 'High-resolution Gas Chromatography–Fourier Transform Infrared Spectroscopy in Flavor Analysis: Limits and Perspectives', *Z. Lebensm. Unters. Forsch.*, **181**, 183–188 (1985).
48. J.L. Le Quéré, E. Sémon, A. Latrasse, P. Etiévant, 'Gas Chromatography–Fourier Transform Infrared Spectrometry. Applications in Flavor Analysis', *Sci. Aliments*, **7**(1), 93–109 (1987).
49. W. Herres, H. Idstein, P. Schreier, 'Flavor Analysis by HRGC/FT-IR: Cherimoya (*Annona cherimolia*, Mill.) Fruit Volatiles', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **6**(11), 590–594 (1983).
50. H. Idstein, W. Herres, P. Schreier, 'High-resolution Gas Chromatography–Mass Spectrometry and –Fourier Transform Infrared Analysis of Cherimoya (*Annona cherimolia*, Mill.) Volatiles', *J. Agric. Food Chem.*, **32**(2), 383–389 (1984).
51. P. Schreier, H. Idstein, 'The Use of HRGC/FTIR in Tropical Fruit Flavor Analysis', in *Analysis of Volatiles*, ed. P. Schreier, Walter de Gruyter, Berlin, 293–306, 1984.
52. L. Nykänen, P. Savolahti, I. Nykänen, 'First Experiences with HR-GC/FT-IR Analysis of Flavor Compounds in Distilled Alcoholic Beverages', in *Topics in Flavor Research*, eds. R.G. Berger, S. Nitz, P. Schreier, E. Eichhorn, Marzling-Hangenham, 109–123, 1985.
53. J.M. Purcell, P. Magidman, 'Analysis of the Aroma of the Intact Fruit of *Coffea arabica* by GC–FT-IR', *Appl. Spectrosc.*, **38**(2), 181–184 (1984).
54. V.F. Kalasinsky, J.T. McDonald, 'GC/FTIR Spectra of Terpenes', *J. Chromatogr. Sci.*, **21**(5), 193–200 (1983).
55. K.H. Kubeczka, W. Herres, 'Kapillar-Gaschromatographie/FTIR-Spektroskopie von ätherischen Ölen', *GIT Suppl. Chromatogr.*, **4**(5), 33–38 (1984).
56. W. Herres, K.H. Kubeczka, W. Schultze, 'HRGC/FTIR Investigations on Volatile Terpenes', in *Progress in Essential Oil Research*, ed. E.J. Brunke, Walter de Gruyter, Berlin, 507–528, 1986.
57. C. Bicchi, C. Frattini, V. Raverdino, 'Considerations and Remarks on the Analysis of *Anthemis nobilis* L. Essential Oil by Capillary Gas Chromatography and ''Hyphenated'' Techniques', *J. Chromatogr.*, **411**, 237–249 (1987).
58. K.H. Kubeczka, W. Schultze, V. Formacek, W. Herres, 'New Developments in Essential Oils Analysis by Fourier-transform Spectroscopy', in *Flavors and Fragrances: A World Perspective*, eds. B.M. Lawrence, B.D. Mookherjee, B.J. Willis, Elsevier, Amsterdam, 931–950, 1988.
59. T.S. Chamblee, R.L. Karelitz, T. Radford, B.J. Clark, 'Identification of Sesquiterpenes in Citrus Essential Oils by Cryofocusing GC/FT-IR', *J. Essent. Oil Res.*, **9**(2), 127–132 (1997).
60. A.A. Williams, O.G. Tucknott, M.J. Lewis, H.V. May, L. Wachter, 'Examples of Cryogenic Matrix Isolation GC/IR in the Analysis of Flavour Extracts', in *Flavour Science and Technology*, eds. M. Martens, G.A. Dalen, H. Russwurm, John Wiley & Sons, Chichester, 259–270, 1987.
61. F. Paul, 'Spectrométrie Infrarouge et Couplage avec la Chromatographie en Phase Gazeuse', in *Les Arômes Alimentaires*, eds. H. Richard, J.L. Multon, APRIA, Tec & Doc – Lavoisier, Paris, 378–401, 1992.
62. C.J. Mussinan, 'Instrumental Analysis in the Flavor Industry', in *Flavor Science. Sensible Principles and Techniques*, eds. T.E. Acree, R. Teranishi, American Chemical Society, Washington, DC, 169–224, 1993.
63. S. Ferary, J. Auger, A. Touché, 'Trace Identification of Plant Substances by Combining Gas Chromatography–Mass Spectrometry and Direct Deposition Gas Chromatography–Fourier Transform Infrared Spectrometry', *Talanta*, **43**, 349–357 (1996).
64. J.L. Le Quéré, R. Brossut, C.A. Nalepa, O. Bonnard, 'Isolation and Identification of 4,6,8-Trimethyl-7,9-undecadien-5-ol, a Female-specific Compound, in Tergal Gland Secretion of *Cryptocercus punctulatus* Scudder (Dictyoptera: Cryptocercidae)', *J. Chem. Ecol.*, **17**(5), 811–821 (1991).

65. C. Bordereau, A. Robert, O. Bonnard, J.L. Le Quéré, '(3*Z*,6*Z*,8*E*)-3,6,8-dodecatrien-1-ol: Sex Pheromone in a Higher Fungus-growing Termite, *Psedacanthotermes spiniger* (Isoptera, Macrotermitinae)', *J. Chem. Ecol.*, **17**(11), 2177–2191 (1991).
66. J.P. Farine, J.L. Le Quéré, J. Duffy, E. Sémon, R. Brossut, '4-Hydroxy-5-methyl-3(2*H*)-furanone and 4-hydroxy-2,5-dimethyl-3(2*H*)-furanone, Two Components of the Male Sex Pheromone of *Eurycotis floridana* (Walker) (Insecta, Blattidae, Polyzosteriinae)', *Biosci. Biotechnol. Biochem.*, **57**(12), 2026–2030 (1993).
67. A.B. Attygalle, A. Svatos, C. Wilcox, S. Voerman, 'Gas-phase Infrared Spectroscopy for Determination of Double-bond Configuration of some Polyunsaturated Pheromones and Related Compounds', *Anal. Chem.*, **67**(9), 1558–1567 (1995).
68. J.L. Sébédio, J.L. Le Quéré, E. Sémon, O. Morin, J. Prévost, A. Grandgirard, 'Heat Treatment of Vegetable Oils. II. GC/MS and GC/FTIR Spectra of some Isolated Fatty Acid Monomers', *J. Am. Oil Chem. Soc.*, **64**(9), 1324–1333 (1987).
69. J.L. Le Quéré, 'Gas Chromatography–Fourier Transform Infrared Spectrometry in the Analysis of Fatty Acids', in *New Trends in Lipids and Lipoprotein Analyses*, eds. E.G. Perkins, J.L. Sébédio, AOCS Press, Champaign, IL, 232–241, 1995.
70. J.L. Le Quéré, E. Sémon, 'Le Couplage Chromatographie en Phase Gazeuse–spectrométrie Infrarouge à Transformée de Fourier (CPG/IRTF) dans l'Analyse des Acides Gras d'Origine Biologique', *Analusis*, **26**(3), M40–M44 (1998).
71. G.T. Reedy, M.M. Mossoba, 'FTIR and Hyphenated FTIR Techniques for the Analysis of Foods', in *Spectral Methods in Food Analysis*, ed. M.M. Mossoba, Marcel Dekker Inc., New York, 325–396, 1999.
72. E. Sémon, S. Ferary, J. Auger, J.L. Le Quéré, 'Gas Chromatography–Fourier Transform Infrared Spectrometry of Fatty Acids: New Applications with a Direct Deposition Interface', *J. Am. Oil Chem. Soc.*, **75**(2), 101–105 (1998).
73. C.L. Wilkins, G.N. Giss, G.M. Brissey, S. Steiner, 'Direct-linked Gas Chromatography–Fourier Transform Infrared–Mass Spectrometer System', *Anal. Chem.*, **53**(1), 113–117 (1981).
74. C.L. Wilkins, 'Linked Gas Chromatography–Infrared–Mass Spectrometry', *Anal. Chem.*, **59**(8), 571A–581A (1987).
75. K.L. Norton, A.J. Lange, P.R. Griffiths, 'A Unified Approach to the Chromatography–FTIR Interface: GC/FTIR, SFC/FTIR, and HPLC/FTIR with Sub-nanogram Detection Limits', *J. High Resol. Chromatogr.*, **14**(4), 225–229 (1991).
76. C.C. Johnson, J.W. Jordan, L.T. Taylor, D.W. Vidrine, 'On-line Supercritical Fluid Chromatography with Fourier Transform Infrared Spectrometric Detection Employing Packed Columns and a High Pressure Lightpipe Flow Cell', *Chromatographia*, **20**(12), 717–723 (1985).
77. P. Morin, M. Caude, H. Richard, R. Rosset, 'Carbon Dioxide Supercritical Fluid Chromatography–Fourier Transform Infrared Spectrometry', *Chromatographia*, **21**(9), 523–530 (1986).
78. P. Morin, M. Caude, H. Richard, R. Rosset, 'On-line Carbon Dioxide SFC/FTIR in Aroma Research. Perspectives and Limits', in *Flavor Science and Technology*, eds. M. Martens, G.A. Dalen, H. Russwurm, John Wiley & Sons, Chichester, 165–174, 1987.
79. J.W. Hellgeth, J.W. Jordan, L.T. Taylor, M. Ashraf Khorassani, 'Supercritical Fluid Chromatography of Free Fatty Acids with On-line FTIR Detection', *J. Chromatogr. Sci.*, **24**(5), 183–188 (1986).
80. E.M. Calvey, R.E. McDonald, S.W. Page, M.M. Mossoba, L.T. Taylor, 'Evaluation of SFC/FTIR for Examination of Hydrogenated Soybean Oil', *J. Agric. Food Chem.*, **39**(3), 542–548 (1991).
81. R.C. Wieboldt, G.E. Adams, D.W. Later, 'Sensitivity Improvement in Infrared Detection of Supercritical Fluid Chromatography', *Anal. Chem.*, **60**(21), 2422–2427 (1988).
82. M.M. Mossoba, R.E. McDonald, D.J. Armstrong, S.W. Page, 'Identification of Minor C_{18} Triene and Conjugated Diene Isomers in Hydrogenated Soybean Oil and Margarine by GC/MI/FTIR Spectroscopy', *J. Chromatogr. Sci.*, **29**(8), 324–330 (1991).

Infrared Reflection–Absorption Spectroscopy

Peter Hollins
University of Reading, Reading, UK

1 Introduction and Historical Overview 1
2 Principles 2
3 Experimental Techniques 6
3.1 General Considerations 6
3.2 Infrared Sources 9
3.3 Spectrometer Design 11
3.4 Comparison with other Surface Spectroscopic Techniques 14
Acknowledgments 16
Abbreviations and Acronyms 16
Related Articles 16
References 16

Infrared reflection–absorption spectroscopy (IRAS, IRRAS) is an optical technique used to study thin (often submonolayer) films adsorbed on reflective substrates such as metals. Experimentally, it involves measuring the change in the reflectance spectrum of the substrate that accompanies adsorption. In order to maximize sensitivity, the reflection is usually performed at grazing incidence. On metal surfaces, the process is subject to an overriding selection rule which states that only those vibrational modes that have a component of their dipole change perpendicular to the surface can be detected: application of this metal surface selection rule (MSSR) often yields important information about adsorption geometry. Other advantages of the technique include its high sensitivity ($<10^{-4}$ monolayers in favorable instances), its ability to operate under high ambient pressures where other primarily electron-based surface spectroscopies cannot be utilized, and the ease with which its results can be correlated with those from other vibrational spectroscopies, both surface- and bulk-sensitive. A significant current disadvantage is the difficulty in studying low-wavenumber vibrations, but this is being overcome by the development of more intense far-infrared sources. Various acronyms for the technique have been employed – IRAS, IRRAS and Reflection–Absorption Infrared Spectroscopy (RAIRS) have all enjoyed wide currency. The last of these is perhaps the most widely used and is the form advocated by the International Union of Pure and Applied Chemistry (IUPAC).[1,2] Accordingly, it will be adopted in this article.

1 INTRODUCTION AND HISTORICAL OVERVIEW

The first applications of infrared spectroscopy to the study of adsorbed species date back to the late 1930s and its use to characterize supported catalysts had become well established by 1960. However, these early studies employed powdered or porous substrates with high specific areas, typically between 1 and 1000 $m^2\,g^{-1}$, which allowed milligram quantities of the adsorbed species to be introduced into the infrared beam. The 1960s witnessed the rapid development of a range of surface science probes capable of determining adsorbate structures and electronic energy levels on single-crystal and other well-defined substrates, thereby emphasizing the need for complementary techniques of vibrational spectroscopy in this area. The rather small surface area of the samples employed in surface science ($\sim 10^{-4}\,m^2$) corresponds to only nanogram masses of adsorbate in a monolayer, so sensitivity poses a much more severe problem than in the study of high area materials. Despite this, infrared sensitivity to monolayer quantities was quickly attained, initially by Pickering and Eckstrom,[3] who studied the adsorption of carbon monoxide and hydrogen on nickel and rhodium films using a multiple reflection procedure in which the incoming beam was at near-normal incidence to the surface. That such a geometry was far from optimum was recognized by Greenler,[4,5] who developed the underlying theory of RAIRS which will be described in the next section, demonstrating that sensitivity was optimized at high angles of incidence. Subsequent studies have employed glancing incidence beams, almost always using a single-reflection configuration. The first characterization of an adsorbate on a single-crystal substrate was achieved in Pritchard's laboratory at the end of the 1960s.[6] This group also pioneered the integration of RAIRS with other techniques of surface science such as low-energy electron diffraction, surface potential measurements, and various electron spectroscopies. This trend has continued and it is now normal for RAIRS instrumentation to be integrated into a system that also contains other probes to establish surface composition, structure, and topography.

Until the mid-1980s, specialized dispersive spectrometers were almost invariably used for RAIRS. Excellent signal-to-noise (S/N) ratios could be achieved over a limited wavenumber range with such instruments (corresponding to a sensitivity of less than a thousandth of a monolayer for a strongly absorbing species such as carbon monoxide). However, dispersive spectrometers are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

poorly suited to rapid scanning of broad ranges of the infrared spectrum, while the specialized nature of the instruments employed in RAIRS discouraged its wider application. The use of commercial Fourier transform infrared (FTIR) spectrometers in RAIRS studies was pioneered by Chesters, who clearly demonstrated their advantages over dispersive spectrometers,[7] and FTIR instrumentation is now almost universally employed. Recent developments in RAIRS have concentrated on extending the range of application of the technique: important instances of this include extension of the wavenumber range (especially into the far-infrared, to observe vibrations of the primary chemisorptive bond between surface and adsorbate), the use of higher gas pressures to mimic industrial catalytic processes, and even studies in aqueous environments, particularly in the study of electrode reactions.

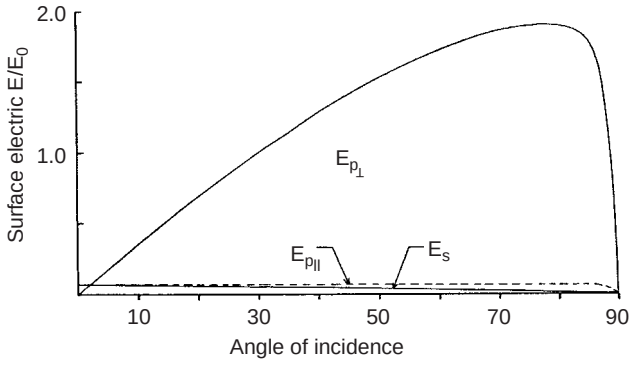

Figure 2 Resultant electric fields relative to the incident field E_0 for infrared radiation incident on a metal with optical constants appropriate to copper at 2000 cm^{-1} ($n = 3$; $k = 30$). $E_{p\|}$ and $E_{p\perp}$ are the components parallel and perpendicular to the surface which arise from initially *p*-polarized radiation. (Reproduced with permission of Elsevier Science from Hollins and Pritchard.[8])

2 PRINCIPLES

The process of reflection at a surface is most easily analyzed by considering the behavior of two polarization components, the *s* component, whose electric field is polarized parallel to the surface and perpendicular to the direction of propagation, and the *p* component, perpendicular to both the *s* component and the line of propagation (Figure 1). Radiation of arbitrary polarization can, of course, be resolved into a combination of these two components.

The process of reflection changes both the magnitude and the direction of the electric field associated with each of these components. A molecule in the vicinity of the surface thus experiences a combination (i.e. the vector sum) of the fields due to the incident and reflected beams. With metals, the direction of the *s* component is almost exactly reversed on reflection, but, since metals have high reflectivities in the infrared, its magnitude is little changed. Electric vectors $\boldsymbol{E}_s$ and $\boldsymbol{E}'_s$ are hence almost exactly equal and opposite, and therefore nearly cancel. The behavior of the *p* component depends somewhat on the angle of incidence, θ, but when θ approaches 90° (i.e. at glancing incidence), the fields combine to give a large component perpendicular to the surface and a much smaller one parallel to it. The θ-dependence of the resultant fields is shown in Figure 2. Bearing in mind that absorption intensity is proportional to the square of the amplitude of the electric field vector, it will be appreciated that at no angle of incidence is there a significant field parallel to the surface, but that an enhanced field perpendicular to the surface arises from the *p* component at angles close to grazing.

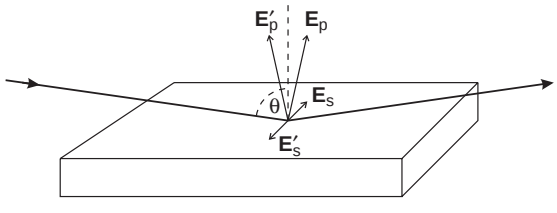

Figure 1 Electric vectors of *s*- and *p*-polarized radiation incident on a metal surface at an angle θ to the normal. Primed and unprimed vectors refer to the reflected and incident beams respectively.

There are two immediate and important consequences of this. The first is that the sensitivity is maximized by employing high angles of incidence. In this regard, the field enhancement already noted is augmented by a simple geometric factor: at glancing angles the beam covers a larger area of surface and hence more molecules. The combined effect of the two factors in a typical case is shown in Figure 3, where it can be seen that the relative sensitivity is sharply peaked at an angle close to grazing. In practice, it is usually not possible to exploit this enhancement fully, partly because the incident beam subtends a range of angles at the surface and partly because at very high angles the projected width of the beam is often larger than the sample width. Bradshaw and Schweizer[9] have shown how the enhancement depicted in Figure 3 is modified by the adoption of a more realistic experimental geometry, concluding that optimum performance should be achieved at an angle a few degrees closer to the normal than the figure would indicate. Nevertheless, the basic principle still applies and experiments are almost always performed using an angle of incidence that is fairly close to grazing.

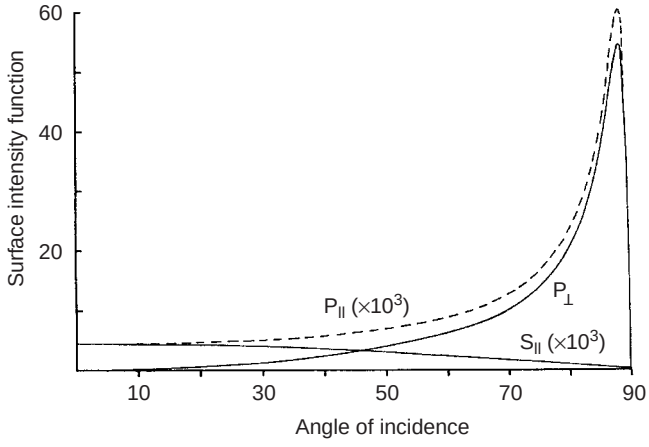

Figure 3 Surface intensity function, $(E/E_0)^2 \sec\theta$, for the field components shown in Figure 2. (Reproduced with permission of Elsevier Science from Hollins and Pritchard.[8])

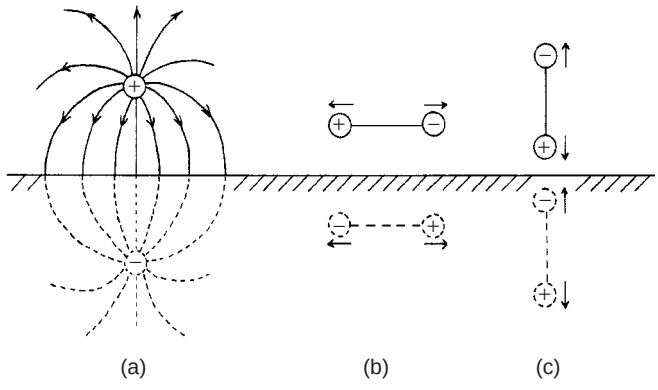

Figure 4 Real and image dipoles at a metal surface. (a) Image effect for a single charge, (b) image of a real dipole parallel to the surface opposes original, (c) enhancement of original when dipole is perpendicular to surface. (Reproduced with permission of Elsevier Science from Pearce and Sheppard.[10])

The second consequence of the optical properties of metal surfaces arises from the fact that it is not possible to establish a large oscillating electric field parallel to the surface. Any dipole lying parallel to the surface will hence be subject to only very small variations in the electric field along its length and will therefore scarcely interact with the incident radiation. We can hence establish the MSSR as developed by Pearce and Sheppard.[10] This states that "only those modes which have vibrational dipole moments with components perpendicular to the metal surface will have measurable intensities". The original discussion of this was couched in terms of image dipole effects which are formally equivalent to the argument already given, but are perhaps easier to grasp intuitively. A famous result in electrostatics[11] states that the field due to a charge situated a distance d above a perfectly conducting plane is identical to that in which the conductor is replaced by a second charge, equal to the first but opposite in sign, placed a distance d below the plane. When this formalism is applied to the two charges that constitute a dipole above a (planar, conducting) metal surface, a clear distinction between parallel and perpendicular dipoles is evident (Figure 4). When the dipole is parallel to the surface, it is opposed by its image. Since the separation between the real and image dipoles is of molecular dimensions and accordingly much smaller than infrared wavelengths, their interactions with infrared radiation cancel essentially perfectly and no absorption can occur. Dipoles perpendicular to the surface, however, reinforce each other, and enhanced absorption is accordingly observed.

The MSSR is often of great help in interpreting RAIRS spectra. For example, at low coverages the spectrum of benzene adsorbed on the silver (100) surface exhibits a single band, whose wavenumber corresponds to that

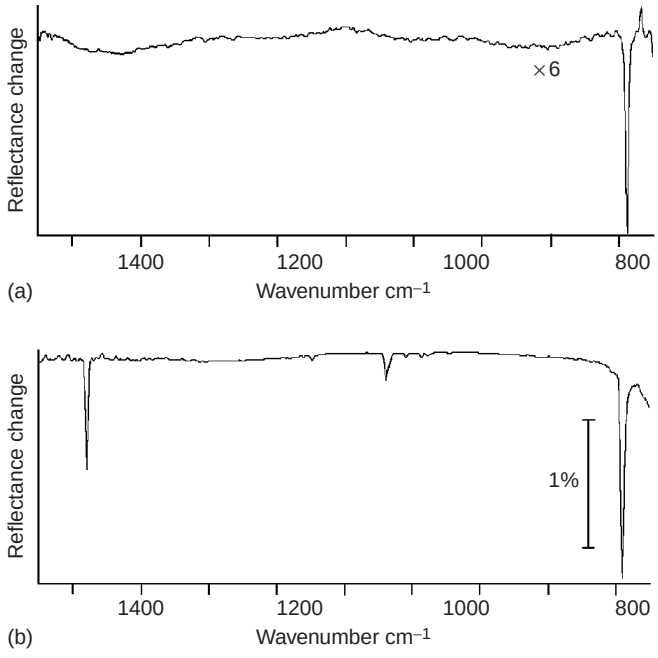

Figure 5 RAIRS spectra of benzene adsorbed on a silver (100) surface. (a) At submonolayer coverages, showing only a band due to the out-of-plane deformation mode, (b) at multilayer coverages, with additional in-plane modes active. (D.A. Slater and P.C. Welch, unpublished work.)

of the out-of-plane deformation mode (Figure 5). This band persists as the coverage is raised to multilayers, but additional bands with wavenumbers corresponding to various in-plane stretches and deformations also appear. The interpretation of these results in terms of the MSSR is that at low coverages the plane of the molecule lies strictly parallel to the surface, so that only the out-of-plane mode is infrared active. At multilayer coverages, the appearance

of extra bands shows that the molecules are no longer all parallel to the surface. This could arise either through tilting of the molecular plane, or by the development of a structure in which some molecules remain parallel to the surface, but others are perpendicular to these: the latter case corresponds to the structure of bulk benzene and is hence the preferred explanation.

Although the MSSR is of great utility, some care must be exercised in its application. One example of this arises in the analysis of spectra from strongly polarizing species such as dioxygen. On a number of substrates oxygen has been found to adsorb molecularly in the peroxide form where the O–O bond is believed to lie strictly parallel to the surface, but where the O–O stretching vibration can nonetheless be observed. The explanation for this is that the metal–oxygen bond is strongly polar and that the charge separation in this depends on the O–O separation. When the O–O stretching vibration is excited, the changes in the O–O bond that occur during the course of the vibration engender concomitant changes in the metal–oxygen bonds, so that an oscillating dipole is established perpendicular to the surface. Since this oscillates at the same frequency as the primary O–O stretching vibration, an absorption band having the same wavenumber as the O–O stretch is observed in the RAIRS spectrum.

It has been shown that modes involving dipole changes which are strictly parallel to a metal surface can sometimes give rise to antiabsorption bands in RAIRS spectra (that is they produce increases in surface reflectivity relative to neighboring regions of the spectrum). This phenomenon, which has chiefly been observed with low-frequency frustrated translations and rotations of the adsorbed species, has been explained in terms of a model that invokes the surface resistivity which the metal conduction electrons experience.[12] Scattering of these electrons by the adsorbate molecules leads to a very broad background absorption in the infrared. However, when the frequency of the incident radiation equals the resonance frequency of a parallel mode, the molecules and the conduction electrons move together, the additional surface resistivity is eliminated, and the background absorption is reduced, leading in turn to the apparent antiabsorption feature. It has been argued that these antiabsorption resonances constitute exceptions to the MSSR, but it is perhaps more useful to envisage them simply as a consequence of an excitation mechanism different from that underlying conventional infrared absorption. Viewed in this context, they are no more exceptions to the MSSR than are the impact excitations of parallel modes observed in vibrational electron energy loss spectroscopy (VEELS) (section 3.4). In any case, since the appearance of these bands is almost precisely opposite to that of an absorption band, confusion between the two is unlikely to arise.

The chief constraint on application of the MSSR, however, is that it is valid only for highly conducting materials, as epitomized by metals. This is true for the whole of the foregoing discussion in this section. Semiconductors and insulating materials lack the high reflectivity of metals in the infrared and accordingly give rise to much weaker absorption bands. RAIRS is therefore a less useful technique for the study of adsorption on such surfaces, which are usually investigated by attenuated total reflectance or by transmission infrared spectroscopy as discussed in a review by Horn.[13] One method of applying RAIRS to such surfaces, however, is to generate (by evaporation or chemical reaction) a thin layer of nonmetal on a metal substrate. Since the wavelength of infrared radiation is long compared with the distances over which chemical interactions take place, it is possible by this means to create a surface whose chemical properties are dominated by those of the overlayer, but which behaves optically as a metal. In the case of silicon overlayers, the real part of the refractive index is close to that of many metals and layers several tens of nanometers in thickness deposited on a metal substrate exhibit optical properties that are essentially metallic in character, as Figure 6 demonstrates. Horn[13] has compared the RAIRS spectrum of a hydrided silicon surface with that of a surface prepared by depositing a silicon layer four atoms thick on a ruthenium substrate and then forming the hydride, finding that the composite substrate yields Si–H stretching bands almost an order of magnitude more intense than the pure silicon, and with a corresponding improvement in S/N ratio.

A similar procedure has been widely used to investigate adsorption on ice surfaces in order to model heterogeneous processes occurring on atmospheric ice particles that are known to be important in stratospheric chemistry. Instead of attempting to carry out RAIRS using a solid ice substrate, a thin (ca. 10 nm) layer of ice is prepared on an inert metal substrate by condensation from the vapor phase. The optical properties of the resulting compound substrate are dependent on wavelength, since ice itself shows strong absorption bands in the mid-infrared, and the design of the RAIRS experiment has to be modified to take this into account.

Returning now to metal surfaces, it might appear that the sensitivity to weakly absorbing species could be enhanced by the use of multiple reflection techniques to sample the adlayer repeatedly. As noted in the previous section, Pickering and Eckstrom[3] employed this method in the earliest RAIRS study, using a multiple reflection White's optics cell in which the mirrors served as the substrate for adsorption. However, Greenler[5] showed that this seemingly attractive possibility for improving sensitivity is limited by the imperfect reflectivity of the substrate, which results in a loss of energy at each

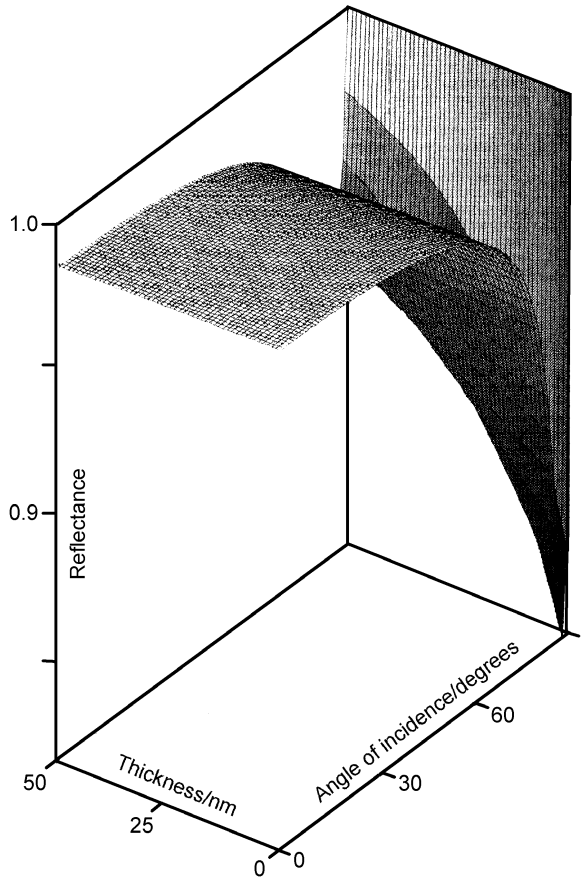

Figure 6 Calculated reflectance at $2000\,cm^{-1}$ of a silicon-on-copper composite interface as a function of silicon thickness and angle of incidence. Metallic behavior persists to substantial thicknesses. Calculated using a program developed by A.B. Horn.

reflection even in regions of the spectrum where the adsorbate does not absorb. In consequence, although the relative absorption intensity increases at each reflection, the absolute intensity attains a maximum after a rather modest number of reflections. If further reflections are attempted, the magnitude of the signal decreases, and, provided that detector noise is the limiting factor (as it usually is), the S/N ratio deteriorates. The number of reflections required to give the maximum signal is given by the integer closest in value to the expression in Equation (1)

$$N_{max} = \frac{\ln(\ln R_0/\ln R)}{\ln(R/R_0)} \qquad (1)$$

where R and R_0 are the reflectivities at the center of the absorption band and away from this band, respectively. Greenler showed that with most metals a single reflection will produce a large fraction (>60%) of the maximum signal theoretically attainable using multiple reflections. The experimental difficulties which attend the implementation of multiple reflection spectroscopy have accordingly been thought to outweigh the small advantages which the technique could in principle provide, at least for single crystal substrates. However, multiple reflections are sometimes employed in the study of polycrystalline materials.

The discussion so far has concentrated on the substrate and no explicit consideration has yet been given to the influence that the optical properties of the adsorbate have on the RAIRS spectrum. The pioneering paper in this area was that of McIntyre and Aspnes,[14] who obtained analytic solutions for the change of reflectance ($\Delta R/R$) for the situation in which an adlayer of thickness d and complex refractive index $n - ik$ overlies a metal substrate. Making approximations valid for submonolayer coverages of an isotropic adsorbate which is fairly weakly absorbing and for an angle of incidence θ which is not too high, these expressions yield the result in Equation (2) for the reflectance change for p-polarized radiation of wavelength λ

$$\frac{\Delta R}{R} = -\left(\frac{4\pi kd}{\lambda}\right)\left(\frac{4\sin\theta\tan\theta}{n^3}\right) \qquad (2)$$

The first factor in this equation corresponds to the familiar attenuation factor εcl in the Beer–Lambert law, while the second corresponds to the field and geometrical enhancements alluded to earlier. This part of the expression takes the same form as that derived from the situation in which the adsorbate is neglected, apart from the inclusion of the factor n^3 in the denominator. Accordingly, to take proper account of the adsorbate, the absorption intensities in Figure 3 should be reduced by this factor, but the conclusions are otherwise unchanged at this level of approximation.

Subsequent work by a number of authors has extended the calculations of McIntyre and Aspnes. Specific additions to the theory include treatment of the adsorbate as a uniaxial or biaxial medium rather than as an isotropic one, extensions to thicker or multiple films and to more strongly absorbing adlayers, consideration of a wider range of substrate optical parameters to include semiconductors and dielectrics as well as metals, and inclusion of the effects of surface roughness. The essential physics behind these calculations is simply an application of Maxwell's equations to the case of an infinite plane electromagnetic wave interacting with a layered structure, and is accordingly well understood in principle. However, as might be expected, increasing complexity of the adlayer system leads to significant computational difficulties, especially when thick multilayers of strong absorbers are being considered.

Early methods of calculation concentrated on finding analytic solutions at appropriate levels of approximation, but more recent work has favored the use of numerical simulation. A particularly comprehensive approach along

these lines has been described by Parikh and Allara,[15] who discuss procedures for the quantitative simulation of both reflection and transmission spectra from multilayer anisotropic systems. As part of its input, such a simulation requires optical parameters for each substrate and adlayer material in the form of a frequency-dependent complex refractive index or permittivity, which in the most general form for an anisotropic medium is described by a tensor quantity. One difficulty arising from this is that for many organic materials this tensor may not be known, either because the large single crystals needed to make the appropriate optical measurements are not available, or because these measurements are themselves too demanding. To circumvent this problem, Parikh and Allara also describe a means for estimating the tensor from experimental measurements made on bulk isotropic material, such as a random dispersion of small crystallites in a matrix of potassium bromide or some similar inert material.

These advanced optical methods allow the simulation of spectra from complex layered systems, such as LangmuirBlodgett and self-assembled films formed from large organic molecules. It should be noted, however, that in practice most experimental results from simpler systems can be dealt with adequately either by using the simplified theory of McIntyre and Aspnes discussed earlier, or even by giving no explicit regard to the optical properties of the adlayer.

In comparing RAIRS measurements with those from other nonsurface infrared spectroscopies, it is often convenient to express the observed intensities in terms of the magnitude of the dynamic dipole involved in the oscillation. The experimental measure of intensity for an absorption band in RAIRS is the integrated absorbance A defined by the relationship in Equation (3)

$$A = \int_{\tilde{\nu}} \ln\left(\frac{R_1}{R_2}\right) d\tilde{\nu} \tag{3}$$

where R_1 and R_2 are respectively the surface reflectivities in the absence and presence of the adsorbate, and $\tilde{\nu}$ represents wavenumber. In the general case, when the surface film has significant thickness, the relationship between this quantity and the magnitude of the dynamic dipole is complicated. However, for a submonolayer film with the dipoles oriented perpendicular to the surface, it reduces to[16] Equation (4)

$$A = \frac{1}{4\pi\varepsilon_0} \frac{4\pi^2 n_s \tilde{\nu}_p \mu^2 \sin\theta \tan\theta}{c\hbar} \tag{4}$$

Here, n_s is the adlayer density (molecules per unit area), $\tilde{\nu}_p$ the wavenumber of the peak absorbance and μ the dynamic dipole. After evaluation of the various constants, this yields Equation (5)

$$\mu = 2.985 \times 10^{-19}\text{C} \sqrt{\frac{A}{\tilde{\nu}_p n_s \sin\theta \tan\theta}} \tag{5}$$

In comparing different oscillators, it is convenient to express the dynamic dipole in terms of the effective charge which moves during the oscillation (e^*). This is given by Equation (6)

$$\mu = e^* \sqrt{\frac{\hbar}{4\pi c m_r \tilde{\nu}_p}} \tag{6}$$

where m_r is the reduced mass of the oscillator. In terms of the molar reduced mass M_r and electronic charge e, this can be expressed as Equation (7)

$$\frac{e^*}{e} = 4.538 \times 10^8 \sqrt{\frac{(A/\text{m}^{-1})(M_r/\text{g mol}^{-1})}{(n_s/\text{m}^{-2})\sin\theta \tan\theta}} \tag{7}$$

For strongly absorbing species, the value of e^* is a substantial fraction of e.

The above equations assume that the real dipole is enhanced by its image (Figure 4): values of μ and e^* twice those given by the above expressions are appropriate if this assumption is not made.

3 EXPERIMENTAL TECHNIQUES

3.1 General Considerations

The requirements for a RAIRS experiment have essentially been defined in the previous section: it is necessary to record the reflectance spectrum of the substrate with a high (S/N) ratio both before and after adsorption, and then to compare the resulting spectra. This final stage is now almost invariably achieved by digital ratioing. As Figure 3 illustrates, the angular range of the incident beam that can be effectively utilized is rather limited. This, combined with the small size of the sample, ensures that the étendue (the conserved product of image size and the solid angle subtended) is much smaller than in most other applications of infrared spectroscopy: values between 10^{-1} and 10^{-2} mm^2 sr are typical. This has implications for the design of the experiment, particularly in terms of the types of source and detector that are appropriate.

The sensitivity of RAIRS is obviously dependent on the attainable noise levels. These are typically in the region of a few parts in 10^5 throughout most of the infrared spectrum, although in the region below about 600 cm^{-1} special sources are needed to achieve this, as will be discussed in section 3.2. This noise level corresponds to

a sensitivity of about 10^{-4} monolayers in the case of adsorbates with particularly strong infrared absorption bands (of which carbon monoxide is the archetype). In less favorable cases, such as the rather weakly absorbing C–H stretching vibrations of hydrocarbons, sensitivity is reduced to a value typically in the range 10^{-2} to 10^{-1} monolayers.

In comparison with many other techniques of surface science, RAIRS possesses the considerable advantage that the probe photon can propagate through nonvacuum environments. This enables the spectrometer to be set up outside the ultra-high vacuum (UHV) chamber, which considerably simplifies operation: not only can materials incompatible with UHV environments be used in construction, but repair and maintenance are facilitated by the absence of any need to break the vacuum in order to perform them. The only modification that is required of the UHV chamber itself is the addition of suitable infrared-transmitting windows. Sapphire or calcium fluoride windows were frequently used during the period when RAIRS was being developed, since they are robust, nonhygroscopic, and available commercially already mounted in UHV-compatible flanges. More recently, alkali halide windows (sodium chloride, potassium bromide, or cesium iodide) have been preferred because of their wider transmission range and, in the first two cases, lower cost. The thermal expansion coefficients of these materials are much higher than that of stainless steel and compensation must be made for this to avoid failure of the window when the UHV system is baked. The usual procedure is to mount the window between O-rings made from a high-temperature elastomer such as Viton, with UHV-compatibility being achieved through differential pumping between these and another O-ring sealing against atmosphere.[17] The alkali halides are also hygroscopic and must be protected against excessive exposure to moist air. Potassium bromide windows permit operation down to 400 cm^{-1}, while the cut-off for cesium iodide lies below 200 cm^{-1}. Cesium iodide is much softer than potassium bromide, but this may prove advantageous in some circumstances, since mechanical or thermal shock is likely to be accommodated by the material deforming, rather than suffering a potentially much more serious fracture.

For many applications, the ideal infrared window material is diamond, which combines great strength with chemical inertness and excellent thermal properties. To obtain good transmission in the infrared, clear diamonds of low nitrogen content (designated as Type IIA diamond) must be used. Until recently, these have been prohibitively expensive in the sizes required for RAIRS applications, but the increasing availability and rapidly diminishing cost of large synthetic diamonds prepared by chemical vapor deposition (CVD) is changing this situation. Natural Type IIA diamonds and CVD diamonds exhibit essentially identical infrared spectra:[18] both are highly transparent throughout the whole of the infrared region (and, indeed, through much of the ultraviolet (UV) and X-ray region), but they do show some absorption around 2000 cm^{-1}. However, this absorption is relatively weak and can be minimized by the use of very thin windows which, because of diamond's great strength, can still support a differential pressure of an atmosphere. The high refractive index of diamond (2.4) necessitates the use of windows mounted at the Brewster angle if excessive losses through reflection are to be avoided.

Although infrared radiation is much more tolerant of gas-phase species than other surface probes such as electrons, it is obviously advantageous to eliminate any such species that exhibit very strong absorption bands. In studies in the mid-infrared region of the spectrum, it is usual to remove water vapor and carbon dioxide from the spectrometer by purging it with air that has been passed through appropriate molecular sieve filters to remove these gases. In the far-infrared, water vapor is particularly troublesome and the use of vacuum spectrometers (operating at about 10^{-2} mbar) is generally advocated.

Until the mid-1980s, almost all RAIRS experiments were performed using dispersive spectrometers, often extensively modified in order to optimize performance. Various techniques were used to improve these spectrometers, including wavelength modulation, polarization modulation, double-beam spectroscopy, and infrared ellipsometry. These methods have been reviewed in some detail elsewhere.[8] Polarization modulation and infrared ellipsometry both achieved an improvement in performance by utilizing the difference in behavior of *s*- and *p*-polarized radiation to obtain an effective double-beam effect, while wavelength modulation produced a signal proportional to the derivative of the conventional spectrum, thereby emphasizing sharp spectral features such as adsorbate absorption bands. More recently, however, these specialized dispersive instruments have been almost entirely displaced by FTIR spectrometers, which are often applied to RAIRS studies with little or no modification of the standard commercial instrument. Nevertheless, there are indications that readoption of now rarely used techniques may prove valuable for some applications, particularly for studies at relatively high gas pressures, where infrared absorption by the phase surrounding the sample is significant. This topic will be considered again in section 3.3.

An extreme case of absorption of infrared radiation by the medium surrounding the surface under study occurs in aqueous solutions, where variants of RAIRS have been widely employed to study adsorption and reaction at

metal electrode surfaces.[19] Two different techniques are usually combined in order to record spectra from surfaces immersed in water, which exhibits very strong infrared absorption bands. First, the path length through the water is minimized by placing the surface very close to the infrared-transmitting window. Separations of the order of micrometers are often used, and prismatic windows are employed in order to achieve the optimum high angle of incidence at the sample. The short path length though the water permits a reasonable proportion of the infrared radiation to be transmitted, but the absorption bands are still strong and would normally swamp any bands due to submonolayer quantities of adsorbate, so an additional means of discriminating against liquid-phase absorption is required. This is achieved by varying the potential applied to the electrode in order to disrupt and reform the adlayer. If, as is usually the case, the infrared spectrum of the liquid phase is unchanged by this change in potential, then the difference between spectra recorded at the two potential differences corresponds to the difference between the surface spectra at these potentials. Spectra from electrodes were first recorded at high sensitivity using a method known as EMIRS (electrochemically modulated infrared spectroscopy) which employed a dispersive spectrometer operating at a slow scan rate in conjunction with a relatively rapid (several hertz) modulation of the potential. As FTIR spectrometers began to supplant the dispersive variety, this was largely replaced by SNIFTIRS (subtractively normalized interfacial Fourier transform infrared spectroscopy), in which a series of scans are taken to obtain a spectrum at a particular potential, the potential is changed, and a second spectrum then recorded. The rapid modulation used in EMIRS is particularly effective in eliminating the effects of drift, and very high S/N ratios are attainable, allowing reflectance changes to be detected at the part per million level. The ultimate sensitivity of SNIFTIRS is somewhat poorer than this, but this disadvantage is usually outweighed by the greater range of potentials that can be employed: while EMIRS is restricted to the study of systems oscillating between two stable configurations, SNIFTIRS can be used to monitor the changes accompanying irreversible voltammetric reactions.

In conventional infrared spectroscopy, FTIR spectrometers exhibit three well-known advantages over scanning dispersive instruments.[20] The first of these, the Fellgett or multiplex advantage, derives from the fact that the spectrometer samples the whole spectral range simultaneously, while a scanning instrument can sample only one resolution element at a time. The measurement time for a spectrum at a given S/N ratio is accordingly reduced by a factor equal to the number of resolution elements, N (numerically equivalent to the width of the spectral range being scanned divided by the resolution). Alternatively, given a fixed measurement time, the S/N ratio is improved by a factor equal to the square root of N. Since N is typically of the order of 10^3, the advantage is generally substantial. In RAIRS, where one seeks to measure a very small change in the signal, the multiplex advantage applies, although it imposes considerable demands on the dynamic range of the spectrometer's detector and analog-to-digital converter (ADC) which will be discussed further in section 3.3.

The second advantage of FTIR spectroscopy, the Jacquinot or throughput advantage, arises because slits are not required in the spectrometer to define a resolution element. Hence, the throughput, or radiation flux, through an FTIR instrument is higher than through a dispersive one, especially at high resolution. In RAIRS, where the small sample size defines a low étendue, it might be expected that the throughput would always be limited by the sample, even with a dispersive spectrometer, and that the Jacquinot advantage would therefore not apply. However, Bradshaw and Schweizer[9] have demonstrated that while this is likely to be true for measurements in the far-infrared, it is often not so in the mid-infrared. They calculate that when using a standard dispersive instrument incorporating a monochromator of 300 mm focal length, the slits rather than the sample become throughput-limiting when the resolution reaches $4\,cm^{-1}$, and that the Jacquinot advantage will accordingly apply at higher resolutions. Allied to the Jacquinot advantage is the observation that when using an FTIR spectrometer, the resolution can be improved simply by increasing the distance through which the moving mirror of the interferometer travels, without otherwise changing the optical arrangement. With a dispersive instrument, the equivalent procedure of reducing the slit width is inevitably more complicated and frequently requires subsequent realignment of the optical path in order to optimize the signal.

The third well-known advantage of FTIR spectrometers, the Connes advantage, relates to the internal calibration of the wavenumber scale provided by the spectrometer's helium–neon laser, and is obviously relevant to RAIRS operation.

As well as these advantages which apply to all areas of infrared spectroscopy, FTIR spectrometers have conferred at least two benefits specifically to RAIRS. The first is the ability to scan an extended wavelength range in a single spectrum. Although the best S/N performance achieved using dispersive spectrometers was probably as good as that subsequently attained by FTIR, it was necessary to optimize the spectrometer for a limited wavenumber range, typically less than $200\,cm^{-1}$ in width. Operation outside this range required grating and filter changes which were time-consuming and, in practice,

rarely performed. Only after the introduction of FTIR technology to RAIRS did the scanning of a wide spectral range become in any way routine. The second benefit derives from the fact that FTIR instruments sample all spectral elements simultaneously. With dispersive spectrometers, the limiting S/N performance was often imposed not by the short-term performance of the detector and its associated electronics, but by much longer-term drifts in detector sensitivity, source temperature, optical alignment, and similar factors. The techniques applied to improve the performance of dispersive spectrometers, particularly wavelength and polarization modulation, achieved their effect mainly by minimizing the influence of these drifts. With FTIR spectrometers, however, such drifts affect all parts of the spectrum almost equally and special techniques are not required to reduce them.

3.2 Infrared Sources

In the vast majority of RAIRS investigations, conventional thermal sources have been used. Such sources are typified by the globar, a silicon carbide rod which is heated resistively, usually to between 1100 and 1500 K, when it behaves as a near-perfect gray body (i.e. constant emissivity) source. Operated in this way, the globar maintains high intensity levels throughout the mid-infrared region between 1000 and 4000 cm^{-1}. It also has sufficient thermal inertia to ensure that it is not overly sensitive to short-term fluctuations in its cooling, and subject to the obvious condition that its power supply must be adequately stabilised, it generally gives good performance in this spectral region, although higher-temperature sources (e.g. a tungsten ribbon at approximately 3000 K) are sometimes preferred at the high wavenumber end of the range.

As one moves toward lower wavenumbers, however, thermal sources become increasingly inadequate, to the extent that few RAIRS studies have investigated the range below 600 cm^{-1}. The reason for this lies in the form of the blackbody radiation curve, as Figure 7 illustrates. It is clear that while the total output integrated over all wavelengths rises very rapidly with temperature (as T^4, following the Stefan–Boltzmann law), the bulk of this additional power is concentrated in the high wavenumber region and intensity at low wavenumbers never approaches that in the mid-infrared. Indeed, in the longwave Rayleigh–Jeans limit the intensity is proportional only to T. Since the output of a 3000 K source is approximately 30 times greater at 2000 cm^{-1} than at 300 cm^{-1}, reasonable power at the lower wavenumber could be obtained from a thermal source only at temperatures in the vicinity of 10^5 K which are obviously unattainable in practice. In many applications of infrared spectroscopy, this loss of intensity at longer wavelengths can be offset by increasing the source size (i.e. étendue), but as already noted, the narrow angular spread of the usable beam in RAIRS reduces the available étendue to a low value, typically in the range 10^{-2}–10^{-1} mm^2 sr. Since increasing the source size is not an effective strategy, a means must be found to increase the absolute brilliance (i.e. power per unit étendue) of the source.

Several groups have now attained this brilliance by the use of synchrotron radiation from a storage ring. Although synchrotron radiation is normally considered as a source of short wavelength radiation (X-rays and vacuum UV), it does contain a significant component in the far-infrared. For example, the synchrotron radiation source at the Daresbury Laboratory, England, operating

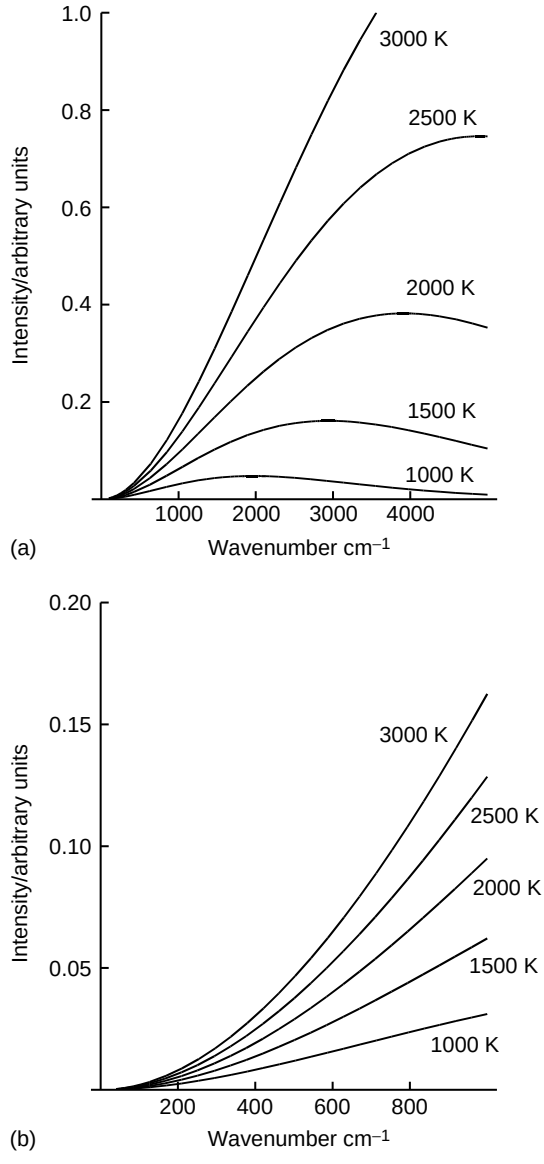

Figure 7 Intensity per unit wavenumber interval for a blackbody source at the temperatures indicated. (a) In the mid-infrared, (b) in the far-infrared.

Table 1 Relative intensity of the Daresbury SRS in comparison with a 1500 K blackbody as a function of wavenumber and the experimental étendue[a]

Étendue (mm^2 sr)	Wavenumber (cm^{-1})			
	3300	1000	330	100
1	0.02	0.08	0.45	4.5
10^{-1}	**0.27**	**0.75**	**4.5**	**45**
10^{-2}	**2.7**	**7.5**	**45**	**450**
1.6×10^{-3}	18	47	270	450
3.5×10^{-4}	81	210	450	450
10^{-4}	270	450	450	450

[a]Entries in bold are typical for a RAIRS experiment.

with an electron current of 300 mA at 2 GeV, has an output in the far-infrared equivalent to that of a blackbody source at a temperature in excess of 650 000 K, about 450 times that of a globar.[21] However, the effective size of the source varies with wavelength, and in the near-infrared the available étendue is small even by the standards of RAIRS, so that thermal sources can actually deliver more power. As one moves toward longer wavelengths, the source étendue increases, and the storage ring starts to deliver more power than a thermal source. Table 1 shows the relative performance of the Daresbury source against a 1500 K blackbody as a function of wavenumber for a range of experimental étendues. Recalling that the relevant range for RAIRS is approximately 10^{-2} to 10^{-1} mm^2 sr, it is seen that the advantage over the thermal source first becomes evident at about 1000 cm^{-1}, that this advantage has increased to approximately one order of magnitude at 330 cm^{-1}, and reaches a limiting value close to 450 at the longest wavelengths, in accordance with the effective blackbody temperature of the synchrotron source.

Infrared beamlines for RAIRS applications are now in operation at several storage rings, with many others under construction.[22] The possibility of extending the technique to longer wavelengths has a number of potential applications, but the one that has been exploited most to date is that of studying the vibrations of the primary chemisorptive bond between substrate and adsorbate. Except in the special case of the lightest atom, chemisorbed hydrogen, these vibrational features usually lie below 700 cm^{-1} and are hence difficult to observe using conventional sources. They are of course especially important in the case of atomic adsorbates, where the atom–substrate bond is necessarily the only one. The chemistry of atomic adsorbates is of major significance in catalysis, since heavy adatoms in the third and subsequent rows of the periodic table – such as sulfur, chlorine, and potassium – are widely recognized to act as poisons or promoters in a great variety of catalytic reactions. Figure 8 shows the first recorded infrared spectrum of such a heavy atomic adsorbate, chlorine on the silver (100) surface, recorded using synchrotron radiation. Other potential topics of interest in the far-infrared include the stretching modes of molecules containing heavy atoms (such as halides and the organometallics employed in semiconductor device fabrication and CVD processes), surface phonons (and their coupling with adsorbed species), and deformation modes, including frustrated rotations and translations.

Certain technical problems attend the use of synchrotron sources for RAIRS, especially in high-energy storage rings, where the electrons have kinetic energies in excess of 1 GeV. Since such kinetic energies are equivalent to several thousand times the rest mass of the electron, major relativistic effects arise. One consequence of these is that length contraction in the direction of motion results in a strongly directional beam throughout most of the electromagnetic spectrum: beam divergence in the X-ray region is less than one milliradian. In the infrared region, however, the source size is comparable with the wavelength, so that diffractive spreading increases the beam divergence to several tens of milliradians. In order to capture this beam using optical components of reasonable size, a mirror must be placed close to the tangent point where the electrons emit the synchrotron radiation: in the Daresbury beamline this mirror is 1.08 m from the tangent point. However, at this distance very little spreading of the

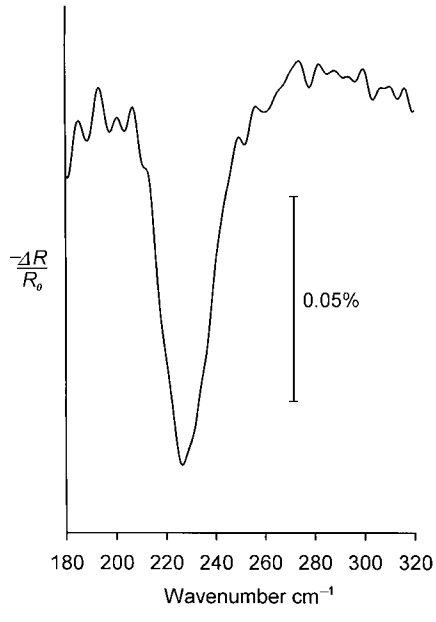

Figure 8 RAIRS spectrum of the c(2 × 2) chlorine adlayer on a silver (100) surface, showing the band arising from the Ag–Cl stretching vibration at 228 cm^{-1}. (Reproduced with permission of the Royal Society of Chemistry from Hollins et al.[16])

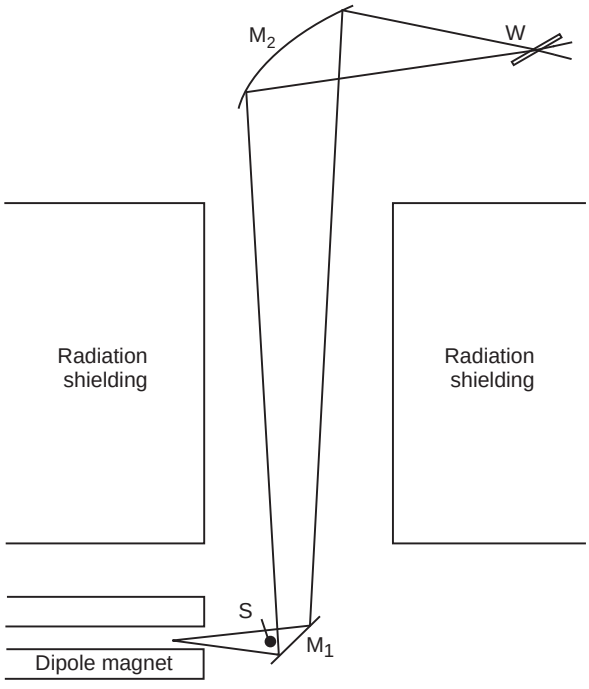

Figure 9 Optical arrangement used to extract an infrared beam from a synchrotron radiation source. S, cooled beam-stop; M_1, cooled plane mirror; M_2, ellipsoidal focusing mirror; W, Brewster-angle diamond window.

shorter wavelength X-ray and UV radiation has occurred, so the central region of the mirror is exposed to a very high flux of these components, leading to massive thermal loads. At Daresbury, these are overcome partly by integral water cooling of the mirror and partly by inserting a cooled stop of copper piping into the beam path immediately before the mirror (Figure 9). (Other solutions are possible and a particularly attractive one which is being explored is the possibility of cutting a central hole in the mirror through which the short-wavelength radiation passes to be utilized elsewhere: hence a single beamline could serve as a combined X-ray/infrared source.) After deflection by this primary mirror, the beam is focused by an ellipsoidal mirror onto a Brewster-angle diamond window which separates the UHV of the storage ring from the RAIRS system.

As well as high intensity in the far-infrared, synchrotron sources possess other potential advantages. The radiation is strongly polarized in the plane of the ring and so can be delivered to the sample as almost pure *p*-polarized radiation. The electrons are also distributed around the ring in evenly spaced ''bunches'' with gaps between them. This results in the emitted radiation having a temporal structure on the nanosecond scale that has potential applications for time-resolved studies, although this has yet to be exploited.

High intensity, polarization, and temporal structure are, of course, major advantages of lasers, which also emit a highly collimated beam that would allow the sharply peaked intensity curve of Figure 3 to be exploited to the full. It is perhaps surprising, therefore, that lasers have not been more widely employed in RAIRS. Attempts to use them have in fact been made by several groups, but the results have generally been inferior to those obtained with thermal sources and have been achieved only with a marked increase in the complexity of the experiment. Although lasers can now offer adequate tunability throughout the infrared region, together with much higher powers than either thermal or synchrotron sources, attaining sufficient stability for RAIRS studies appears to pose major difficulties. However, the potential advantages of lasers are so great that it seems very likely that in the long term these difficulties will be overcome.

3.3 Spectrometer Design

As already noted, most RAIRS work is now carried out using standard commercial FTIR spectrometers with minimal modifications made to accommodate the reflection experiment, which is most commonly carried out with the sample housed in a (relatively large) UHV chamber equipped with facilities for a range of other surface science techniques. Typical chamber and infrared window dimensions constrain the experimental optics to an aperture of f/5 or less, corresponding to an acceptance angle of no more than 12°, but this angle is sufficient to encompass the peak in the surface intensity function (Figure 3). Consideration must also be given to the optical arrangement used to extract the infrared beam from the spectrometer, focus it into the UHV chamber, reflect it from the surface under study, and then refocus it onto the detector. The most usual configuration, shown in Figure 10(a), involves deflecting the beam, after passage through the interferometer, out of the external port with which most commercial FTIR spectrometers are fitted. A (usually parallel) beam of radiation exits the spectrometer from this port and is focused onto the sample using either an alkali halide lens or an arrangement of mirrors. (Use of a lens affords a cheap solution which also allows a particularly simple optical path to be employed, but it suffers from chromatic aberration together with transmission and reflection losses.) After reflection from the surface, the beam is refocused onto an external detector whose signal is fed back into the processing circuitry of the spectrometer. One disadvantage of this configuration is that visual alignment of the optical path external to the spectrometer may be difficult, since the beam splitters used in mid-infrared FTIR instruments often have relatively poor transmission of visible light, while far-infrared beam splitters are frequently totally

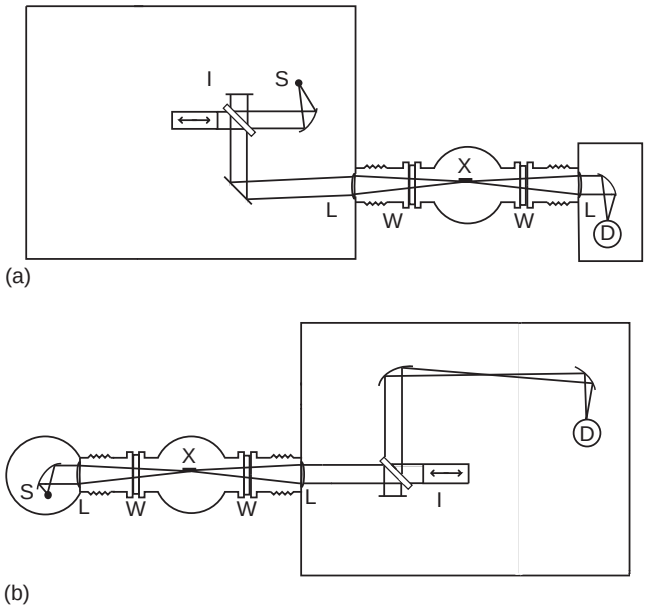

Figure 10 Alternative optical arrangements for performing RAIRS. (a) Source-interferometer-sample-detector, (b) source-sample-interferometer-detector. S, source; I, Michelson interferometer; L, alkali halide lens (or equivalent mirror assembly); W, window to UHV chamber; X, single-crystal sample; D, infrared detector.

opaque in the visible. A common solution is to align the beam visually using a near-infrared quartz beam splitter, which transmits well in the visible, and then to replace this with the appropriate mid- or far-infrared beam splitter for spectroscopic measurement. Alternatively, a sensor card which emits visible light when struck by infrared radiation can be used, but the UHV environment near the sample usually ensures that the sensor card cannot be used in just that section of the optical path where alignment is most critical.

An alternative arrangement that avoids this difficulty is to place the source immediately before the sample and to focus the beam reflected from the sample directly into the interferometer/detector section of the spectrometer, as illustrated in Figure 10(b). The beam at the sample position is then much brighter visually. The chief disadvantage of this design compared with that in Figure 10(a) is that the (cooled) detector receives more thermal infrared emission from the (room temperature) beam splitter. Particularly when working in the far-infrared, the magnitude of this spurious beam splitter emission may be comparable with the true signal and a degradation in performance results. The problem is greatly exacerbated by the fact that commercial FTIR spectrometers are designed for high optical throughput, with an experimental étendue at least an order of magnitude greater than that of a typical RAIRS experiment. Hence, the acceptance angle of the detection optics is large and more radiation is accepted from the beam splitter than is necessary. Reduction of the detector acceptance angle to optimize it for RAIRS experiments would virtually eliminate the problem of beam splitter emission, but this has rarely been attempted.

Chesters[7] and Bradshaw and Schweizer[9] have discussed the constraints placed on a RAIRS spectrometer by the requirement for a high dynamic range. RAIRS differs from many applications of infrared spectroscopy in that it involves detection of a very small change in signal level: the reflectance change following adsorption is usually less than 1% and in some applications may be less than 0.01%. With dispersive instruments, each spectral element is collected separately, so that detection of a 0.01% band requires an S/N ratio of 10^4. With FTIR instruments, a large number, N, of spectral elements are monitored simultaneously and the change in signal is spread over all of these elements. In the absence of the multiplex advantage, this would require an increase in S/N by a factor of N to attain the same sensitivity, but since the whole interferogram contributes to the signal at any point, this factor is reduced to the square root of N. Even so, the required S/N is much higher than with a dispersive instrument: scanning a range of $3200\,\mathrm{cm}^{-1}$ at a resolution of $2\,\mathrm{cm}^{-1}$ generates 1600 spectral elements, so the S/N ratio has to be improved by a factor of 40, to 4×10^5 in the case just discussed.

Such high ratios are achieved by co-adding a large number of individual scans, each with a much poorer S/N value. The FTIR spectrometers used are accordingly rapid-scan instruments with which a single interferogram can be recorded in a fraction of a second. Attention must also be paid to the resolution of the spectrometer's ADC. In the earliest applications of FTIR to RAIRS, 15- or 16-bit ADCs were typically used. Since in normal operation of an FTIR instrument the last two bits (at least) of the ADC contain only noise,[20] the attainable S/N in a single scan from such an instrument is limited by the performance of the ADC alone to a value of about $(2^{13}/\sqrt{1600}) \approx 200$. Improvement of this to the desired value of 10^4 would then require the co-addition of approximately $(10^4/200)^2 = 2500$ scans. This number may be reduced by using ADCs of 20-bit or higher resolution which are found in more modern research instruments.

The infrared detector, too, must be able to support this high dynamic range. Most RAIRS investigations in the mid-infrared range currently use liquid nitrogen cooled mercury cadmium telluride (MCT) detectors, which have high saturation levels so that their dynamic range does not often prove limiting. MCT detectors give good performance over a wide wavenumber range, their performance at long wavelengths having improved dramatically in recent years, but for studies in the far-infrared (including especially the synchrotron work described in the previous

section), they are usually replaced either by copper-doped germanium photoconductive detectors, or by bolometric detectors fabricated from heavily doped silicon or germanium. All these far-infrared detectors require cooling to liquid helium temperatures. If the investigation is restricted to the relatively near infrared (above about $1900\,cm^{-1}$), then the liquid nitrogen cooled indium antimonide detector has a small performance advantage over the MCT.

Interaction with the underlying surface broadens the linewidths of bands in an adsorbate spectrum, typically to several reciprocal centimeters or more. There is thus no requirement for particularly high spectroscopic resolution in RAIRS work. Most experiments are carried out at $2\,cm^{-1}$ or poorer resolution, although in studies of lineshape analysis it may be necessary to reduce this to $1\,cm^{-1}$ or $0.5\,cm^{-1}$.

Although one of the strengths of RAIRS is the ease with which it can be added to an existing UHV apparatus without the need for major modification of either the vacuum system or the spectrometer, there are occasions when changes to one or both of these are required in order to exploit a particular advantage of the technique. One such advantage of RAIRS over many other surface spectroscopies – especially those which use electrons or ions as the surface probe – is that the interaction of the infrared photon with gas-phase species is relatively weak. RAIRS can therefore be used to study surfaces in the presence of relatively high ambient pressures of gases, and even serve as an in situ probe under reaction conditions. It thus possesses the potential to bridge the ''pressure gap'' between the UHV conditions required for the preparation of clean surfaces and the superatmospheric pressures employed in many commercial processes such as catalysis. To exploit this possibility, several researchers have modified a UHV system by the addition of a high pressure RAIRS cell, so that a sample can be prepared and characterized under UHV conditions and then transferred, still under UHV, to the cell, where RAIRS experiments can be performed over a wide pressure range. For example, McDougall and Yates[23] have described a design of cell in which RAIRS experiments can be carried out at pressures of up to two atmospheres (1 atmosphere = 101.325 kPa), far beyond the range of most other surface-sensitive techniques, and they have used this to investigate the reaction of a surface ethylidyne species ($\equiv C-CH_3$) with high pressures of hydrogen. The complete inactivity of gas-phase hydrogen toward infrared absorption, of course, makes this a particularly favorable choice for high-pressure experiments, but such studies are possible even with infrared-active gases. One way of achieving this is to monitor the absorption of both *s*- and *p*-polarized radiation separately: the two components interact equally with any gas-phase species present, but the *s* component is virtually inactive toward species adsorbed on a metal surface (Figure 3), so that the difference in absorption between the two components yields the surface spectrum. By using this approach Judd et al.[24], who also describe an alternative design of high-pressure RAIRS cell, were able to measure the spectrum of submonolayer coverages of carbon monoxide on a nickel surface under carbon monoxide pressures so high that the conventional RAIRS spectrum was swamped by absorption from gas-phase carbon monoxide and from nickel carbonyl generated by the reaction between this and the metal. The spectrometer used in this study was a comparatively primitive dispersive one which only allowed the different polarization components to be measured sequentially, rather than simultaneously, but several groups are working on more sophisticated designs. The techniques of polarization modulation and infrared ellipsometry, which were employed to improve the performance of early dispersive spectrometers but fell into disuse after the widespread adoption of FTIR technology, can measure spectra from *s*- and *p*-polarized radiation simultaneously. There is a strong likelihood that attempts to model catalytic systems at high pressure will lead to the design of FTIR instruments modified to allow some form of polarization-dependent spectroscopy. Whatever technique is adopted, the potential applications of such studies are so diverse and important that they are sure to play a major part in the future development of RAIRS.

Data acquisition times for modern FTIR spectrometers are sufficiently short to allow measurement of the kinetics of some surface processes.[7] This was exploited by Reutt-Robey et al. in an extremely elegant molecular beam experiment which monitored the diffusion of carbon monoxide molecules on stepped platinum surfaces.[25] The spectrometer was configured to scan and store individual interferograms rapidly (at up to 46.5 Hz, limited by the mechanical turn-round time of the moving mirror). Just before the start of one of these scans, the spectrometer initiated, via an RS-232 interface, a rapid (600 μs) pulse of carbon monoxide from the solenoid-actuated supersonic molecular beam doser. This pulse was adsorbed by the stepped platinum crystal surface to give a low surface coverage (~1% of a monolayer) which was initially randomly distributed between the available adsorption sites, but which diffused over time to occupy the thermodynamically favored step sites preferentially. Since molecules adsorbed on the different types of adsorption site exhibit different C–O stretching frequencies, the infrared spectrum can be used to monitor this diffusion process. By recording the proportions of the two species over timescales from a few milliseconds to an hour, and repeating these measurements over a range of substrate temperatures, the activation energy and pre-exponential factor for the surface diffusion process were determined.

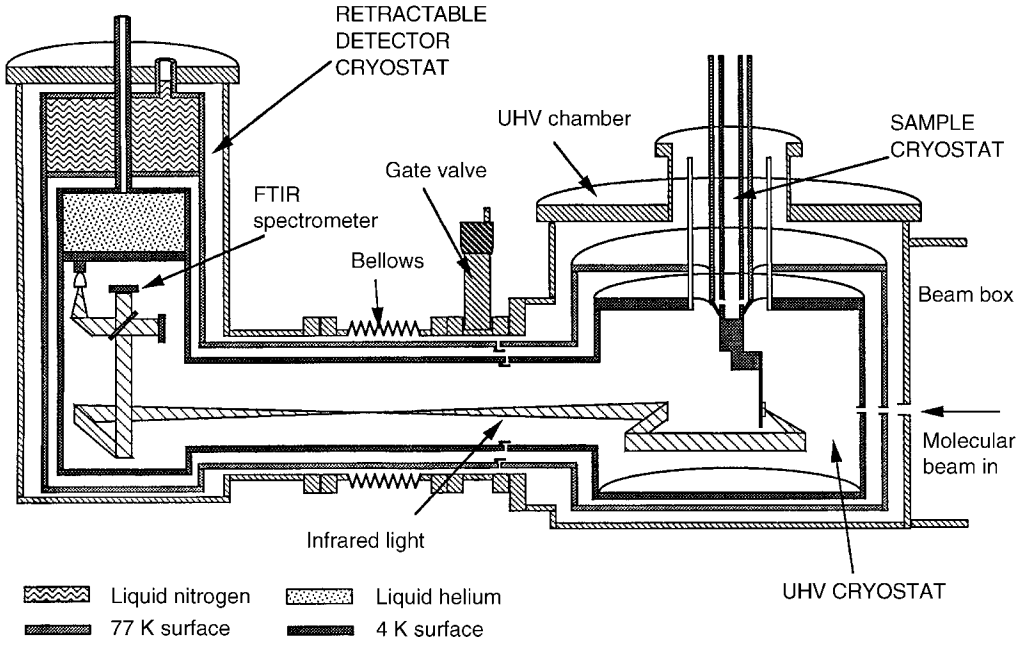

Figure 11 Liquid-helium cooled spectrometer designed for RAIRS and infrared emission measurements. (Reproduced with permission of Elsevier Science from Brown et al.[26])

The most radical redesign of a UHV system and FTIR spectrometer for RAIRS applications must surely be that described by Brown et al.[26] Although primarily designed for measurements of infrared emission and chemiluminescence, this system also served, after the addition of potassium bromide windows, for more conventional RAIRS studies. The design was predicated on the need to reduce spurious thermal signals (alluded to earlier in this section in the context of beam splitter emission) to negligibly low values. To this end, the entire optical path between sample and detector (Figure 11) is shrouded in a light-tight enclosure cooled by liquid helium. The optics of the FTIR spectrometer are also cooled to liquid helium temperatures, and since operation of a laser within the cryogenic enclosure would lead to unacceptable heat dissipation, the position of the moving mirror is sensed capacitatively. The detector for the most sensitive experiments planned is a cadmium-doped germanium bolometer operating at a temperature of 0.3 K and exhibiting noise levels two orders of magnitude lower than those of a typical MCT detector. However, the saturation level of such an ultrasensitive detector is very low, so a more conventional bolometer, with a working temperature of 4 K, is used to measure higher infrared fluxes.

3.4 Comparison with other Surface Spectroscopic Techniques

A range of other techniques are capable of providing vibrational spectroscopic information from the same types of surfaces as RAIRS. The relative advantages and disadvantages of the most important of these are discussed below.

Perhaps the most directly comparable technique is VEELS,[27] which, like RAIRS, has been widely used for fundamental studies of single crystal surfaces. Energy loss spectra are recorded by reflecting a beam of monoenergetic electrons from a surface and then analyzing their energies. Most electrons are scattered elastically, but some (typically, a few per thousand) lose energy to vibrational transitions in the adlayer; a plot of the number of electrons scattered versus their energy loss constitutes the VEELS spectrum, which contains essentially the same information as an infrared spectrum. VEELS rose to prominence in the late 1970s, largely as a result of developments made by Ibach, and at that time it was the only technique capable of recording the entire vibrational spectrum of an adlayer in a single scan. The use of FTIR instruments has since then allowed RAIRS to achieve this also, so the relative merits of the two techniques now lie in other areas.

The relative sensitivities of VEELS and RAIRS are very similar (assuming equal measurement times) in the mid-infrared region, but VEELS has a definite advantage in the far-infrared. This stems not only from the poor intensity of thermal infrared sources in this region (section 3.2), but also from the fact that the performance of most VEELS spectrometers improves at low loss energies. The angular spread of the scattered electrons is smaller at these energies, so the analyzer accepts a

larger number of them, provided that this angular spread remains greater than the acceptance angle, as is usually the case. This effect causes the collection efficiency to scale as $\tilde{\nu}^{-2}$, where $\tilde{\nu}$ is the wavenumber of the loss peak.

For many years, VEELS could only achieve a much poorer spectroscopic resolution than RAIRS, being limited to several tens of reciprocal centimeters. However, Ibach then showed that reduction of space–charge effects in the spectrometer could improve this markedly[28] and spectrometers capable of achieving resolutions of a few reciprocal centimeters are now available commercially. On this basis, RAIRS still maintains a slight advantage, but not nearly to the same extent as formerly. Achievement of high resolution in VEELS, however, is critically dependent on having a well-ordered surface. If the surface becomes poorly ordered (which can happen even with a single-crystal substrate if the adsorbate does not form a well-order overlayer), then scattering becomes more diffuse and it may be necessary to degrade the resolution markedly in order to obtain an acceptable intensity.

The energy-loss process central to VEELS can occur via several mechanisms. The most important of these, dipolar scattering, is formally equivalent to the absorption of an infrared photon, and so obeys the same selection rules, including specifically the MSSR. Other significant mechanisms include impact scattering, which is a shorter-range process than dipole scattering, and negative-ion resonance excitation, which involves the capture of the incident electron by a surface or adsorbate atom, followed by the decay of the negative ion formed in this process. Impact scattering and negative ion resonance are bound by different selection rules from those applicable to dipolar scattering, and the angular distribution of the scattered electrons is also different. VEELS can therefore detect vibrational modes that are inactive in RAIRS and, moreover, it is possible by varying the experimental geometry to establish which of these modes are dipole-allowed and which are not. In some circumstances this additional information can be of considerable use in interpreting the spectrum.

One final area of comparison, where RAIRS certainly maintains an advantage, is in the pressure range over which the techniques can be applied. VEELS, relying as it does on the electrons having sufficiently long mean free paths to travel between the spectrometer and the sample, cannot be employed at pressures much above 10^{-6} mbar (even when the surrounding gas is infrared-inactive), whereas RAIRS has been used at pressures more than nine orders of magnitude higher.

Among photon-based spectroscopies, the most obvious alternative to infrared is probably Raman spectroscopy. Since Raman scattering cross-sections are typically very much smaller than infrared absorption cross-sections, the technique would seem to have a clear disadvantage in sensitivity, but submonolayer sensitivity was achieved using then state-of-the-art technology in the mid-1980s.[29] Prior to that, the observation of surface enhanced Raman scattering (SERS) caused great excitement when it was found that adsorbates on certain surfaces exhibited Raman scattering cross-sections which were typically four to six orders of magnitude greater than expected from the gas-phase cross-sections. Although it has now been established that SERS effects of this magnitude occur only on a limited range of substrates (mainly roughened surfaces of the coinage metals) and there is still some doubt as to whether the molecules enhanced by SERS are representative of the adlayer as a whole, research in the area remains active. In studies of electrode surfaces, where the beam has to be transmitted through aqueous solution, the low Raman scattering cross-section of water offers a potentially large advantage over RAIRS.

Infrared emission spectroscopy, in which the thermal radiation emitted by the adsorbate is measured directly, was used in the mid-1980s by Richards's group at Berkeley to obtain an infrared spectrum of the metal–CO stretching vibration from an extended substrate, before RAIRS had attained sufficient sensitivity to observe such vibrations in the far-infrared.[30] The apparatus used in these studies was remarkably simple, consisting of a dispersive spectrometer recording the spectra in direct current (i.e. unmodulated) mode. As with the Cambridge instrument described earlier,[26] the optics were maintained at liquid helium temperatures and the region between the (room temperature) sample and the spectrometer was screened with cold baffles. Since the blackbody radiation curve for samples at moderate (close to room) temperature peaks at a few hundred reciprocal centimeters, emission spectroscopy is most sensitive to vibrations in this range. As already noted, attempts are currently in progress to develop the technique using more modern instrumentation.

Another vibrational technique that is experiencing rapid development is sum frequency generation (SFG), in which the surface is simultaneously illuminated with a tunable infrared laser and a fixed frequency visible laser.[31] When the frequency of the infrared laser matches that of an appropriate vibration of the adsorbate, coherent anti-Stokes Raman scattering leads to resonant scattering at a frequency equal to the sum of the two laser frequencies. Although experimentally complicated, SFG has two substantial advantages as a surface probe. First, by using a pulsed laser, it offers the prospect of time-resolved measurements on a much more rapid scale than can be achieved by RAIRS. Second, the SFG process does not occur in a centrosymmetric medium, so the technique is inherently specific to the interface between two such media. It therefore offers particular scope for

the investigation of interfaces surrounded by liquids or high pressure gases and, as with Raman spectroscopy, has a potentially large advantage over RAIRS for studies of surfaces in aqueous solution.

ACKNOWLEDGMENTS

I am grateful to Dr David Slater and Dr Philip Welch for recording the spectra shown in Figure 5, and to Dr Andrew Horn for providing me with the program used to generate Figure 6.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-Digital Converter
CVD	Chemical Vapor Deposition
EMIRS	Electrochemically Modulated Infrared Spectroscopy
FTIR	Fourier Transform Infrared
IRAS, IRRAS	Infrared Reflection–Absorption Spectroscopy
IUPAC	International Union of Pure and Applied Chemistry
MCT	Mercury Cadmium Telluride
MSSR	Metal Surface Selection Rule
RAIRS	Reflection–Absorption Infrared Spectroscopy
SERS	Surface Enhanced Raman Scattering
SFG	Sum Frequency Generation
S/N	Signal-to-Noise
SNIFTIRS	Subtractively Normalized Interfacial Fourier Transform Infrared Spectroscopy
UHV	Ultra-high Vacuum
UV	Ultraviolet
VEELS	Vibrational Electron Energy Loss Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Infrared Spectroelectrochemistry

Infrared Spectroscopy **(Volume 12)**
Theory of Infrared Spectroscopy

REFERENCES

1. N. Sheppard, 'English-derived Abbreviations for Experimental Techniques in Surface Science and Chemical Spectroscopy', *Pure Appl. Chem.*, **63**, 887–893 (1991).
2. I. Mills, T. Cvitaš, T. Homann, N. Kallay, K. Kuchitsu, *Quantities, Units and Symbols in Physical Chemistry*, 2nd edition, Blackwell, Oxford, 1993.
3. H.L. Pickering, H.C. Eckstrom, 'Heterogeneous Reaction Studies by Infrared Absorption', *J. Phys. Chem.*, **63**, 512–517 (1959).
4. R.G. Greenler, 'Infrared Study of Adsorbed Molecules on Metal Surfaces by Reflection Techniques', *J. Chem. Phys.*, **44**, 310–315 (1966).
5. R.G. Greenler, 'Reflection Method for Obtaining the Infrared Spectrum of a Thin Layer on a Metal Surface', *J. Chem. Phys.*, **50**, 1963–1968 (1969).
6. M.A. Chesters, J. Pritchard, M.L. Sims, 'Infrared Spectrum and Surface Potential of Carbon Monoxide Chemisorbed on a Copper (100) Single-crystal Surface', *J. Chem. Soc., Chem. Commun.*, 1454–1455 (1970).
7. M.A. Chesters, 'Infrared Spectroscopy of Molecules on Metal Single-Crystal Surfaces', *J. Electron Spectrosc. Relat. Phenom.*, **38**, 123–140 (1986).
8. P. Hollins, J. Pritchard, 'Infrared Studies of Chemisorbed Layers on Single Crystals', *Prog. Surf. Sci.*, **19**, 275–351 (1985).
9. A.M. Bradshaw, E. Schweizer, 'Infrared Reflection-Absorption Spectroscopy of Adsorbed Molecules', in *Spectroscopy of Surfaces*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, 413–483, 1988.
10. H.A. Pearce, N. Sheppard, 'Possible Importance of a ''Metal Surface Selection Rule'' in the Interpretation of the Infrared Spectra of Molecules Adsorbed on Particulate Metals; Infrared Spectra from Ethylene Adsorbed on Silica-supported Metal Catalysts', *Surf. Sci.*, **59**, 205–217 (1976).
11. R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, Addison-Wesley, Reading, MA 1964, lecture II, 6.
12. C.J. Hirschmugl, G.P. Williams, B.N.J. Persson, A.I. Volokitin, 'Adsorbate Vibrational Dynamics in the Anomalous Skin Effect Frequency Region', *Surf. Sci.*, **317**, L1141–L1146 (1994).
13. A. Horn, 'Infrared Spectroscopic Techniques for the Study of Thin Interfacial Films', in *Spectroscopy for Surface Science*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, 273–339, 1998.
14. J.D.E. McIntyre, D.E. Aspnes, 'Differential Reflection Spectroscopy of Very Thin Surface Films', *Surf. Sci.*, **24**, 417–434 (1971).

15. A.N. Parikh, D.L. Allara, 'Quantitative Determination of Molecular Structure in Multilayered Thin Films of Biaxial and Lower Symmetry. Reflection Infrared Vibrational Spectroscopy. I. Reflection Infrared Vibrational Spectroscopy', *J. Chem. Phys.*, **96**, 927–945 (1992).
16. P. Hollins, A.A. Davis, D.A. Slater, M.A. Chesters, E.C. Hargreaves, P.M. Parlett, J.C. Wenger, M. Surman, 'Reflection-absorption IR Spectrum of Chlorine Adatoms on the Silver (100) Surface', *J. Chem. Soc., Faraday Trans.*, **92**, 879–880 (1996).
17. P. Hollins, J. Pritchard, 'Versatile Infrared Window Mounting for Ultrahigh Vacuum Application', *J. Vac. Sci. Technol.*, **17**, 665 (1980).
18. R.S. Sussmann, J.R. Brandon, G.A. Scarsbrook, C.G. Sweeney, T.J. Valentine, A.J. Whitehead, C.J.H. Wort, 'Properties of Bulk Polycrystalline CVD Diamond', *Diamond Relat. Mater.*, **3**, 303–312 (1994).
19. M.J. Weaver, S. Zou, 'Vibrational Spectroscopy of Electrochemical Interfaces: Some Walls and Bridges to Surface Science Understanding', in *Spectroscopy for Surface Science*, eds. R.J.H. Clark, R.E. Hester, John Wiley & Sons, Chichester, 219–272, 1998.
20. P.R. Griffiths, J.A. de Heseth, *Fourier Transform Infrared Spectroscopy*, John Wiley & Sons, New York, 1986.
21. M.A. Chesters, E.C. Hargreaves, M. Pearson, P. Hollins, D.A. Slater, J.M. Chalmers, B. Ruzicka, M. Surman, M.J. Tobin, 'Infrared Spectroscopy and Microscopy at the Daresbury Synchrotron Light Source', *Nuovo Cimento Soc. Ital. Fis. D*, **20**, 439–448 (1998).
22. P. Calvani, P. Roy (eds.), 'Infrared Synchrotron Radiation', Società Italiana di Fisica, Bologna, 1998.
23. G. McDougall, H. Yates, 'Thermal Evolution and Hydrogenation of Ethylidyne on Pt (111) Studied by RAIRS', in *Catalysis and Surface Characterisation*, eds. T.J. Dines, C.H. Rochester, J. Thomson, Royal Society of Chemistry, London, 109–117, 1992.
24. R.W. Judd, H.J. Allen, P. Hollins, J. Pritchard, 'RAIRS at higher pressures: Ni(110)–CO and Cu(111)–CO', *Spectrochim. Acta*, Part A, **43**, 1607–1610 (1987).
25. J.E. Reutt-Robey, D.J. Doren, Y.J. Chabal, S.B. Christman, 'CO Diffusion on Pt (111) with Time-Resolved Infrared-Pulsed Molecular Beam Methods: Critical Tests and Analysis', *J. Chem. Phys.*, **93**, 9113–9129 (1990).
26. W.A. Brown, R.K. Sharma, P. Gardner, D.A. King, D.H. Martin, 'The Design of a Super-Sensitive Infrared Emission Spectrometer for Studies of Adsorbate Dynamics', *Surf. Sci.*, **331–333**, 1323–1328 (1995).
27. H. Ibach, D.L. Mills, *Electron Energy Loss Spectroscopy and Surface Vibrations*, Academic Press, New York, 1982.
28. H. Ibach, *Electron Energy Loss Spectrometers: the Technology of High Performance*, Springer, Berlin, 1991.
29. A. Campion, 'Raman Spectroscopy', in *Vibrational Spectroscopy of Molecules on Surfaces*, eds. J.T. Yates, T.E. Madey, Plenum Press, New York, 345–415, 1987.
30. P.L. Richards, R.G. Tobin, 'Infrared Spectroscopy of Adsorbates on Metals: Direct Absorption and Emission', in *Vibrational Spectroscopy of Molecules on Surfaces*, eds. J.T. Yates, T.E. Madey, Plenum Press, New York, 417–463, 1987.
31. Y.R. Shen, *The Principles of Non-Linear Optics*, John Wiley & Sons, New York, 1987.

Interpretation of Infrared Spectra, A Practical Approach

John Coates
Coates Consulting, Newtown, USA

1 Introduction **1**
2 The Origins of the Infrared Spectrum **2**
3 Spectral Interpretation by Application of Vibrational Group Frequencies **6**
3.1 The Hydrocarbon Species and Molecular Backbone 6
3.2 Simple Functional Groups 9
3.3 The Carbonyl Group 12
3.4 Other Functional Groups Associated with Heteroatoms 13
3.5 Simple Inorganics 14
4 The Practical Situation – Obtaining the Spectrum and Interpreting the Results **15**
4.1 Sample History 16
4.2 Physical Characteristics of the Sample 17
4.3 The Chemistry of the Sample 17
4.4 The Infrared Sampling Method 18
5 An Overview to Infrared Spectral Interpretation – Some Simple Rules and Guidelines **19**
5.1 A Quick Diagnostic Assessment of an Infrared Spectrum 20
Abbreviations and Acronyms **22**
Related Articles **22**
References **23**

The vibrational spectrum of a molecule is considered to be a unique physical property and is characteristic of the molecule. As such, the infrared spectrum can be used as a fingerprint for identification by the comparison of the spectrum from an ''unknown'' with previously recorded reference spectra. This is the basis of computer-based spectral searching. In the absence of a suitable reference database, it is possible to effect a basic interpretation of the spectrum from first principles, leading to characterization, and possibly even identification of an unknown sample. This first principles approach is based on the fact that structural features of the molecule, whether they are the backbone of the molecule or the functional groups attached to the molecule, produce characteristic and reproducible absorptions in the spectrum. This information can indicate whether there is backbone to the structure and, if so, whether the backbone consists of linear or branched chains. Next it is possible to determine if there is unsaturation and/or aromatic rings in the structure. Finally, it is possible to deduce whether specific functional groups are present. If detected, one is also able to determine local orientation of the group and its local environment and/or location in the structure. The origins of the sample, its prehistory, and the manner in which the sample is handled all have impact on the final result. Basic rules of interpretation exist and, if followed, a simple, first-pass interpretation leading to material characterization is possible. This article addresses these issues in a simple, logical fashion. Practical examples are included to help guide the reader through the basic concepts of infrared spectral interpretation.

1 INTRODUCTION

The qualitative aspects of infrared spectroscopy are one of the most powerful attributes of this diverse and versatile analytical technique. Over the years, much has been published in terms of the fundamental absorption frequencies (also known as group frequencies) which are the key to unlocking the structure–spectral relationships of the associated molecular vibrations. Applying this knowledge at the practical routine level tends to be a mixture of art and science. While many purists will argue against this statement, this author believes that it is not possible to teach a person to become proficient as an interpretive spectroscopist by merely presenting the known relationships between structure and the observed spectra. Instead, the practical approach, which has been adopted in this text, is to help the reader appreciate the visual aspects of the spectroscopy and how to interpret these relative to the structure and chemistry of the sample. This is achieved by recognizing characteristic shapes and patterns within the spectrum, and by applying the information obtained from published group frequency data, along with other chemical and physical data from the sample.

Included in the text is a discussion of the interrelationships that exist between the practical side of acquiring the spectrum, the chemistry and physics of the sample under study, the physical interactions of the sample with its environment, and the impact of the structure on the spectrum. In essence, the interpretation of infrared spectra is much more than simply assigning group frequencies. The spectrum is rich in information, and this article is intended to help the reader to extract the maximum using the knowledge available for the sample and the acquired spectral data. One important factor to bear in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

mind is that a successful interpretation is based not only on the *presence* of particular bands within the spectrum, but also the *absence* of other important bands. Complete classes of compounds can be rapidly excluded during the interpretation by the use of *no-band* information.

It must be understood that this article addresses the issue of infrared spectral interpretation from the perspective of the average operator of an infrared instrument. It is not a detailed treatise on the theory of infrared spectroscopy where the modes of vibration are discussed in terms of group theory, and where mathematical models are used to compare theoretical and observed values for the fundamental vibrations of a molecule. There are many excellent texts that cover this subject.[1–4] Instead, this article focuses on the day-to-day problems associated with characterizing a material or attempting to perform some form of identification. One of the main challenges in presenting a text on spectral interpretation is to form a balance between the theory that is needed to appreciate the links between molecular structure and the observed spectrum and the practice. For this reason, a minimum amount of relevant theory is included in the next section, which provides a basic understanding of why the spectrum exists, how it is formed, and what factors contribute to the complexity of observed spectra. It has been assumed that the reader has a fundamental knowledge of molecular theory and bonding, and that there is an understanding of basic structures, in particular for organic compounds.

Infrared spectral interpretation may be applied to both organic and inorganic compounds, and there are many specialized texts dealing with these compounds, in combination and as individual specialized texts. There are too many to reference comprehensively, and the reader is directed to a publication that provides a bibliography of the most important reference texts.[5] However, the most informative general reference texts are included,[6–14] with books by Socrates[10] and Lin-Vien[11] being recommended for general organics, and by Nakamoto[13] and Nyquist et al.[14] for inorganics (salts and coordination compounds). There are numerous specialized texts dealing with specific classes of materials, and undoubtedly polymers and plastics form the largest individual class.[15–17] In this particular case, texts by Hummel and Scholl[16] and Koenig[17] provide a good basic understanding.

The following comments are made relative to the conventions used within this article. The term frequency is used for band/peak position throughout, and this is expressed in the commonly used units of wavenumber (cm^{-1}). The average modern infrared instrument records spectra from an upper limit of around $4000\,cm^{-1}$ (by convention) down to $400\,cm^{-1}$ as defined by the optics of the instrument (commonly based on potassium bromide, KBr). For this reason, when a spectral region is quoted in the text, the higher value will be quoted first, consistent with the normal left-to-right (high to low cm^{-1}) representation of spectra. Also, the terms infrared band, peak and absorption will be used interchangeably within the text to refer to a characteristic spectral feature.

The spectral group frequencies provided in this text were obtained from various literature sources published over the past 30 years, and most of these are included in the cited literature. Every attempt to ensure accuracy has been taken; however, there will be instances when individual functional groups may fall outside the quoted ranges. This is to be expected for several reasons: the influences of other functional groups within a molecule, the impact of preferred spatial orientations, and environmental effects (chemical and physical interactions) on the molecule.

The preferred format for presenting spectral data for qualitative analysis is in the percentage transmittance format, which has a logarithmic relationship ($-\log_{10}$) with respect to the linear concentration format (absorbance). This format, which is the natural output of most instruments (after background ratio), provides the best dynamic range for both weak and intense bands. In this case, the peak maximum is actually represented as a minimum, and is the point of lowest transmittance for a particular band.

2 THE ORIGINS OF THE INFRARED SPECTRUM

In the most basic terms, the infrared spectrum is formed as a consequence of the absorption of electromagnetic radiation at frequencies that correlate to the vibration of specific sets of chemical bonds from within a molecule. First, it is important to reflect on the distribution of energy possessed by a molecule at any given moment, defined as the sum of the contributing energy terms (Equation 1):

$$E_{\text{total}} = E_{\text{electronic}} + E_{\text{vibrational}} + E_{\text{rotational}} + E_{\text{translational}} \quad (1)$$

The translational energy relates to the displacement of molecules in space as a function of the normal thermal motions of matter. Rotational energy, which gives rise to its own form of spectroscopy, is observed as the tumbling motion of a molecule, which is the result of the absorption of energy within the microwave region. The vibrational energy component is a higher energy term and corresponds to the absorption of energy by a molecule as the component atoms vibrate about the mean center of their chemical bonds. The electronic component is linked to the energy transitions of electrons as they

are distributed throughout the molecule, either localized within specific bonds, or delocalized over structures, such as an aromatic ring. In order to observe such electronic transitions, it is necessary to apply energy in the form of visible and ultraviolet radiation (Equation 2):

$$E = h\nu \qquad \text{frequency/energy} \tag{2}$$

The fundamental requirement for infrared activity, leading to absorption of infrared radiation, is that there must be a net change in dipole moment during the vibration for the molecule or the functional group under study. Another important form of vibrational spectroscopy is Raman spectroscopy, which is complementary to infrared spectroscopy. The selection rules for Raman spectroscopy are different to those for infrared spectroscopy, and in this case a net change in bond polarizability must be observed for a transition to be Raman active. The remaining theoretical discussion in this article will be limited to a very simple model for the infrared spectrum. The reader is encouraged to refer to more complete texts[2–4] for detailed discussion of the fundamentals.

While it was stated that the fundamental infrared absorption frequencies are not the only component to be evaluated in a spectral interpretation, they are the essence and foundation of the art. For the most part, the basic model of the simple harmonic oscillator and its modification to account for anharmonicity suffice to explain the origin of many of the characteristic frequencies that can be assigned to particular combinations of atoms within a molecule. From a simple statement of Hooke's law we can express the fundamental vibrational frequency of a molecular ensemble according to Equation (3):

$$\nu = \frac{1}{2\pi c}\sqrt{\frac{\kappa}{\mu}} \tag{3}$$

where ν = fundamental vibration frequency, κ = force constant, and μ = reduced mass. The reduced mass, $\mu = m_1 m_2/(m_1 + m_2)$, where m_1 and m_2 are the component masses for the chemical bond under consideration.

This simple equation provides a link between the strength (or springiness) of the covalent bond between two atoms (or molecular fragments), the mass of the interacting atoms (molecular fragments) and the frequency of vibration. Although simple in concept, there is a reasonably good fit between the bond stretching vibrations predicted and the values observed for the fundamentals.

This simple model does not account for repulsion and attraction of the electron cloud at the extremes of the vibration, and does not accommodate the concept of bond dissociation at high levels of absorbed energy. A model incorporating anharmonicity terms is commonly used to interpret the deviations from ideality and the overall energy–spatial relationship during the vibration of a bond between two atomic centers. The fundamental, which involves an energy transition between the ground state and the first vibrational quantum level, is essentially unaffected by the anharmonicity terms. However, transitions that extend beyond the first quantum level (to the second, third, fourth, etc.), which give rise to weaker absorptions, known as overtones, are influenced by anharmonicity, which must be taken into account when assessing the frequency of these higher frequency vibrations.

Having defined the basis for the simple vibration of an atomic bond, it is necessary to look at the molecule as a whole. It is very easy to imagine that there is an infinite number of vibrations, which in reality would lead to a totally disorganized model for interpretation. Instead, we describe the model in terms of a minimum set of fundamental vibrations, based on a threefold set of coordinate axes, which are known as the normal modes of vibration. All the possible variants of the vibrational motions of the molecule can be reduced to this minimum set by projection on to the threefold axes. It can be shown that the number of normal modes of vibration for a given molecule can be determined from Equations (4) and (5):

$$\text{number of normal modes} = 3N - 6 \text{ (nonlinear)} \tag{4}$$

$$= 3N - 5 \text{ (linear)} \tag{5}$$

where N is the number of component atoms in the molecule.

In practice, apart from the simplest of compounds, most molecules have nonlinear structures, except where a specific functional group or groups generate a predominant linear arrangement to the component atoms. If we calculate the number of modes for a simple hydrocarbon, such as methane (nonlinear, tetrahedral structure), a value of nine is obtained. This would imply that nine sets of absorption frequencies would be observed in the spectrum of methane gas. In reality, the number observed is far less, corresponding to the asymmetric and symmetric stretching and bending of the C–H bonds about the central carbon atom. The reason for the smaller than expected number is that several of the vibrations are redundant or degenerate, that is, the same amount of energy is required for these vibrations. Note that although a small number of vibrational modes is predicted, and in fact observed, the appearance of the methane spectrum at first glance is far more complex than expected, especially at higher spectral resolutions ($<1\,\text{cm}^{-1}$). At relatively high resolutions, a fine structure is superimposed, originating from rotational bands, which involve significantly lower energy transitions. Each of the sets of vibrational–rotational absorptions manifest this superimposed fine structure for low-molecular-weight gaseous compounds, methane being a good example. Several medium-molecular-weight

compounds may also show evidence of some fine structure when studied in the vapor state. For example, it is common to observe the sharp feature (or spike) assigned to the Q-branch of the vibrational–rotational spectrum, as indicated by the vapor spectrum of acetone (Figure 1).

If we proceed up the homologous series from methane (CH_4) to *n*-hexane (C_6H_{14}), there are 20 component atoms, which would imply 54 normal modes. In this case the picture is slightly more complex. Methane is a unique molecule, and only contains one type of C–H group – no other types of bond exist in this molecule. In hexane there are several types of bond and functionality. For reference, a simple two-dimensional representation of the structure is provided in Figure 2(a).

As we can see, there are two terminal methyl groups (CH_3) and four connecting methylene groups (CH_2). Each of these groups has its corresponding C–H stretching and bending vibrations (see later text for the actual absorption frequencies). Also, the methyl groups are linked to a neighboring methylene group, which is in turn linked to neighboring methylene groups, and so on. These linkages feature carbon–carbon bonds. For interpretation, we view the C–H groups as functional groups, giving rise to the common group frequencies, and the C–C linkages as the backbone, producing the skeletal vibrations. As a rule, a group frequency may be applied generally to most compounds featuring the corresponding functional group. In contrast, the skeletal vibrations are unique to a specific molecule. The group frequencies help to characterize a compound, and the

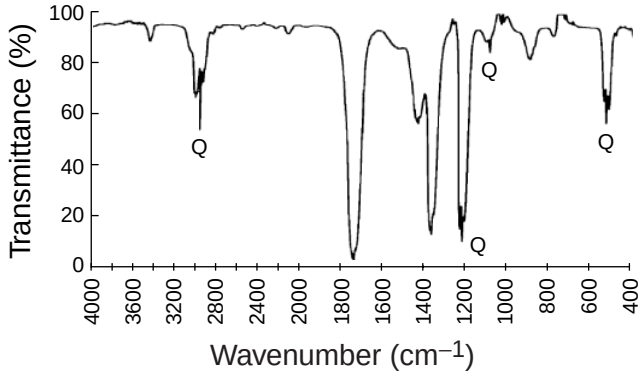

Figure 1 Vapor spectrum of acetone with characteristic Q-branch slitting, denoted by Q. Copyright Coates Consulting.

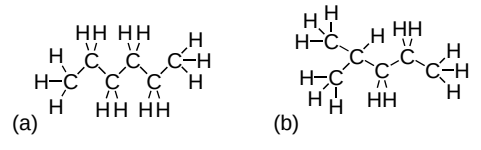

Figure 2 Structures for hexane isomers: (a) *n*-hexane and (b) isohexane (2-methylpentane). Copyright Coates Consulting.

combination of the bands associated with these group frequencies and the skeletal frequencies are used to identify a specific compound. The latter forms the basis of the use of reference spectra for spectral matching by visual comparison or by computer-based searching, for the identification of an unknown from its infrared spectrum.

The group frequencies may be viewed quantitatively, as well as qualitatively. A given absorption band assigned to a functional group increases proportionately with the number times that functional group occurs within the molecule. For example, in the spectrum of *n*-hexane, the intensities measured for the group frequency absorptions assigned to methyl and methylene correspond to four methylene groups and two methyl groups on a relative basis, when compared with other hydrocarbon compounds within a homologous series. For example, if we examine the C–H stretching (or bending) band intensities for CH_3 and CH_2, we will observe that the relative intensities of CH_3 to CH_2 decrease with increase in chain length. Restated, there is less methyl contribution and more methylene contribution with increase in chain length/molecular weight. The reverse holds true if we examine the spectra of linear hydrocarbons with chain lengths shorter than that of hexane.

If we apply these ideas to a different hexane isomer, such as isohexane (2-methylpentane), we would see significant differences in the spectrum. These can be explained by evaluating the structure (Figure 2b), which contains three methyl groups, two methylene groups, and a group that contains a single hydrogen attached to carbon (the methyne group). This adds a new complexity to the spectrum: the main absorptions show differences in appearance, caused by the changes in relative band intensities, splittings of absorptions occur (originating from spatial/mechanical interaction of adjacent methyl groups), and changes are observed in the distributions of the C–C skeletal vibrations, in part due to the splitting by the methyl side chain. Further discussions concerning the impact of chain branching are covered later in this article. Comparison of Figures 3 and 4 provides a graphical representation of the aspects discussed for the hexanes of structurally similar compounds, i.e. *n*-heptane and isooctane.

From a first-order perspective, the idea of the quantitative aspects of the group frequencies carries through for most functional groups, and the overall spectrum is essentially a composite of the group frequencies, with band intensities in part related to the contribution of each functional group in the molecule. This assumes that the functional group does give rise to infrared absorption frequencies (most do), and it is understood that each group has its own unique contribution based on its extinction coefficient (or infrared absorption cross-section).

Returning to the fundamental model, we should now look at the larger picture. In reality, we assign the

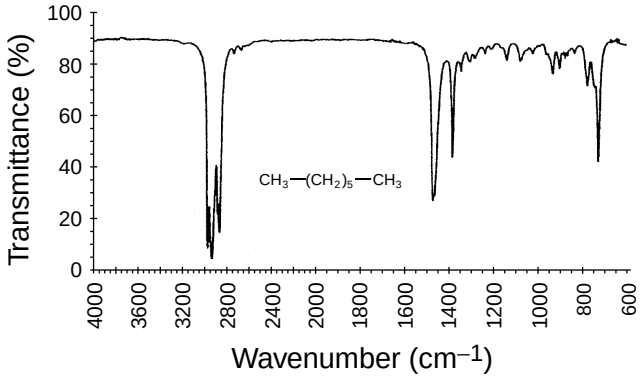

Figure 3 Attenuated total reflectance (ATR) spectrum of *n*-heptane. Copyright Coates Consulting.

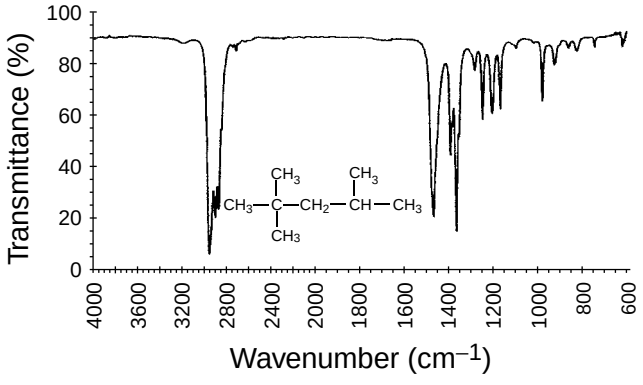

Figure 4 ATR spectrum of 2,2,4-trimethylpentane (isooctane). Copyright Coates Consulting.

observed absorption frequencies in the infrared spectrum to much more that just simple harmonic (or anharmonic) stretching vibrations. In practice, we find that various other deformation motions (angular changes), such as bending and twisting about certain centers within a molecule, also have impact, and contribute to the overall absorption spectrum. By rationalizing the effort needed to move the atoms relative to each other, one can appreciate that it takes less energy to bend a bond than to stretch it. Consequently, we can readily accept the notion that the stretching absorptions of a vibrating chemical bond occur at higher frequencies (wavenumbers) than the corresponding bending or bond deformation vibrations, with the understanding, of course, that energy and frequency are proportionally related. A good example is the C–H set of vibrations, observed in the hydrocarbon spectra, and in virtually all organic compounds. Here, the simple C–H stretching vibrations for saturated aliphatic species occur between 3000 and 2800 cm^{-1}, and the corresponding simple bending vibrations nominally occur between 1500 and 1300 cm^{-1}.

Next in our understanding is that it can take slightly more energy to excite a molecule to an asymmetric than a symmetric vibration. While this might be less intuitive, it is still a rational concept, and therefore easy to understand and accept. Again, we see a good example with the C–H stretch of an aliphatic compound (or fragment), where we observe the asymmetric C–H stretch of the methyl and methylene groups (2960 and 2930 cm^{-1}, respectively) occurring at slightly higher frequency than symmetric vibrations (2875 and 2855 cm^{-1}, respectively for methyl and methylene). For the most part, this simple rule holds true for most common sets of vibrations. Naturally there are always exceptions, and a breakdown of the rationale may occur when other effects come into play, such as induced electronic, spatial or entropy-related effects.

There are many other spatially related scenarios that tend to follow well-orchestrated patterns, examples being in-plane and out-of-plane vibrations, the differences between *cis* and *trans* spatial relationships, and a variety of multicentered vibrations that are defined as twisting or rocking modes. Many of these are exhibited with the C–H vibrations that occur in saturated, unsaturated and aromatic compounds. Molecular symmetry of the static or the dynamic (during vibration) molecule has a large impact on the spectrum, in addition to factors such as relative electronegativity, bond order and relative mass of the participating atoms.

Finally, while discussing the vibrational origins of infrared spectra, it is worth commenting that further complexity may be noted in the spectrum, beyond what is expected based on the fundamentals. As noted, transitions to higher energy levels, although theoretically not allowed, can occur and these give rise to overtone bands, which in the mid-infrared region occur at approximately twice the fundamental frequency for the first overtone. Higher overtones exist, typically the second (3 × fundamental) and third (4 × fundamental), and sometimes higher, and these are observed, with extremely low intensity, relative to the fundamental in the near-infrared spectral regions, between 800 and 2500 nm (12 500 and 4000 cm^{-1}). Other types of bands that can add complexity to a spectrum are combination bands (sum and difference), bands due to transitions from energy states higher than the ground state or ''hot bands'', and bands due to interactions between a weaker overtone or combination band and a fundamental of the same or similar frequency, known as Fermi resonance bands. In the latter case, two relatively strong absorptions are observed, where normally only a single absorption is expected for the fundamental.[3]

As additional functional groups are added to a basic backbone structure, forming a more complex molecule, additional bands are observed, either directly associated with the fundamental vibrations of the functional groups, or indirectly related to interactions between component

functional groups or the basic substructure. Such interactions can be severe, and result in overwhelming distortions in the appearance of the spectrum, a good example being hydrogen bonding. This will be dealt with in depth later.

3 SPECTRAL INTERPRETATION BY APPLICATION OF VIBRATIONAL GROUP FREQUENCIES

This section includes tabulated data relative to the most significant group frequencies for the most common functional groups and structural components found in organic compounds. Brief reference is also made to simple inorganic compounds, in the form of simple ionic species. More detailed listings can be found in published literature, and the reader is encouraged to acquire one or more of these reference texts.[9–13] As already indicated, the use of tabulated data is only a part of the interpretation process, and other facets of the spectrum must be taken into account.

To help gain an understanding of infrared spectral interpretation, it is instructive to start at the root of most organic compounds, namely the fundamental backbone or the parent hydrocarbon structure. We shall start with the simple, aliphatic hydrocarbon, which is at the root of most aliphatic compounds. Aliphatic hydrocarbons exist in simple linear chains, branched chains and in cyclic structures – examples of the linear and branched chain scenarios were provided earlier for hexane isomers. Any one molecule may exist with one or more of these component structures. The infrared spectrum can provide information on the existence of most of these structures, either directly or by inference.

The introduction of unsaturation in the form of a double or triple bond has a profound impact on the chemistry of the molecule, and likewise it has a significant influence on the infrared spectrum. Similarly, the same is observed when an aromatic structure is present within a molecule. Infrared spectroscopy is a powerful tool for identifying the presence of these functionalities. It provides information specific to the group itself, and also on the interaction of the group with other parts of the molecule and on the spatial properties of the group. Examples of these include conjugation between a double bond and another unsaturated center, an aromatic ring or a group, such as a carbonyl (C=O), and the orientation or location of the double bond within the molecule, such as *cis* or *trans* and medial or terminal. It should be noted that *cis*/*trans* relationships are not specific to unsaturated hydrocarbons, and the terminology is referenced elsewhere, such as with secondary amide structures. Again, the associated changes in the spatial arrangement of the groups involved is reflected in the infrared spectrum as additional bands and added complexity.

As we move on to simple organic compounds, where one or more functional groups or heteroatoms are added to the molecule, we see many changes occurring in the spectrum. These result from the bonding associated with the functional group, and also local disturbances to the basic backbone spectrum that relate again to spatial changes and also to local and neighboring electronic effects. Examples of such functionalities are halogens, simple oxygen species, such as hydroxy and ether groups, and amino compounds. Carbonyl compounds, where the added functional group includes the C=O bond, also provide very profound contributions to the spectrum, and because of the wide diversity of these compounds they are best dealt with as a separate class.

A very characteristic group of compounds, from a spectral point of view, are the multiple-bonded nitrogen compounds, such as cyanides and cyanates. These typically have very characteristic absorptions, which are easy to assign, and are free from spectral interferences. The same can be said for some of the hydrides of heteroatoms, such as sulfides (thiols), silanes, and phosphines. Finally, there are other, oxygen-containing functional groups, as encountered in the nitrogen-oxy (NO_x), phosphorus-oxy (PO_x), silicon-oxy (SiO_x), and sulfur-oxy (SO_x) compounds. These are sometimes more difficult to identify from first principles, and a knowledge of the presence of the heteroatom is helpful. The spectra are characteristic, but many of the oxy absorptions occur within a crowded and highly overlapped region of the spectrum, mainly between 1350 and 950 cm^{-1}. Also, many of these compounds feature C–O bonding, which is common in other frequently encountered functionalities such as ethers and esters.

3.1 The Hydrocarbon Species and Molecular Backbone

In this section we include the characteristic absorption frequencies encountered for the parent hydrocarbon species and the associated backbone or substituent group. This includes aliphatic and aromatic structures. The spectral contributions are characterized, as previously noted, as C–H stretching and bending vibrations and C–C vibrations (stretching and bending), which for the most part are unique for each molecule, and are generally described as skeletal vibrations. In the case of aromatic compounds, ring C=C–C stretching and bending vibrations are highly characteristic, and are diagnostic. Likewise, the same can be said for the unsaturated carbon–carbon multiple bonding in alkene and alkyne structures.

Table 1 Saturated aliphatic (alkane/alkyl) group frequencies

Group frequency (cm^{-1})	Functional group/assignment
	Methyl ($-CH_3$)
2970–2950/2880–2860	Methyl C–H asym./sym. stretch
1470–1430/1380–1370	Methyl C–H asym./sym. bend
1385–1380/1370–1365	*gem*-Dimethyl or "iso"- (doublet)
1395–1385/1365	Trimethyl or "*tert*-butyl" (multiplet)
	Methylene ($>CH_2$)
2935–2915/2865–2845	Methylene C–H asym./sym. stretch
1485–1445	Methylene C–H bend
750–720	Methylene $-(CH_2)_n-$ rocking ($n \geq 3$)
1055–1000/1005–925	Cyclohexane ring vibrations
	Methyne ($>CH-$)
2900–2880	Methyne C–H stretch
1350–1330	Methyne C–H bend
1300–700	Skeletal C–C vibrations
	Special methyl ($-CH_3$) frequencies
2850–2815	Methoxy, methyl ether O–CH_3, C–H stretch
2820–2780	Methylamino, N–CH_3, C–H stretch

3.1.1 Saturated Aliphatic and Alicyclic Compounds

See Table 1. The C–H stretch vibrations for methyl and methylene are the most characteristic in terms of recognizing the compound as an organic compound containing at least one aliphatic fragment or center. The bending vibrations help to tell more about the basic structure. For example, a strong methylene/methyl band ($1470\,cm^{-1}$) and a weak methyl band ($1380\,cm^{-1}$), plus a band at $725–720\,cm^{-1}$ (methylene rocking vibration) is indicative of a long-chain linear aliphatic structure (note that splitting may be observed for the 1470 and $720\,cm^{-1}$ bands, which is indicative of a long-chain compound, and is attributed to a crystallinity and a high degree of regularity for the linear backbone structure). In contrast, strong methyl bands, showing significant splitting, and a comparatively weaker methylene/methyl band indicate chain branching, and the possibility of isopropyl or *tert*-butyl substituents (depending on the amount of splitting, and the relative band intensities). A comparison between linear and branched chain hydrocarbons can be seen in Figures 3 and 4, where in the case of isooctane, both isopropyl and *tert*-butyl groups are present.

3.1.2 Unsaturated Compounds

See Table 2. As already commented upon, the saturated hydrocarbon C–H stretching absorptions all occur below $3000\,cm^{-1}$. Any band structures observed between 3150 and $3000\,cm^{-1}$ are almost exclusively indicative of unsaturation (C=C–H) and/or aromatic rings. The

Table 2 Olefinic (alkene) group frequencies

Origin	Group frequency, wavenumber (cm^{-1})	Assignment
C=C	1680–1620	Alkenyl C=C stretch
	1625	Aryl-substituted C=C
	1600	Conjugated C=C
C–H	3095–3075 +3040–3010	Terminal (vinyl) C–H stretch
	3095–3075	Pendant (vinylidene) C–H stretch
	3040–3010	Medial, *cis*- or *trans*-C–H stretch
C–H	1420–1410	Vinyl C–H in-plane bend
	1310–1290	Vinylidene C–H in-plane bend
C–H	995–985 + 915–890	Vinyl C–H out-of-plane bend
	895–885	Vinylidene C–H out-of-plane bend
C–H	970–960	*trans*-C–H out-of-plane bend
	700 (broad)	*cis*-C–H out-of-plane bend

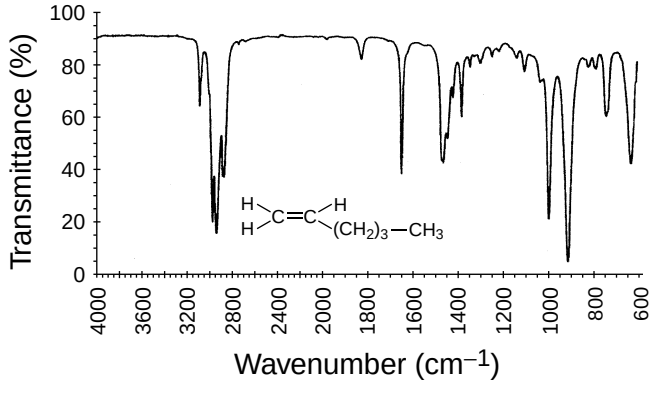

Figure 5 ATR spectrum of 1-hexene. Copyright Coates Consulting.

unsaturated hydrocarbons featuring C=C, with attached hydrogens, usually occur as either a single or a pair of absorptions, in the ranges indicated in Table 2. As noted, the number of bands and their positions are indicative of the double bond location and the spatial arrangement around the double bond. The position of the C=C stretching frequency does vary slightly as a function of orientation around the double bond, but it is less informative than the C–H information. The C–H out-of-plane bending is typically the most informative relative to the location and spatial geometry of the double bond, where terminal and medial double bonds may be clearly differentiated. Figure 5 provides a good example with the spectrum of 1-hexene, which contains the terminal vinyl group. Note that a fully substituted, medial double bond has only the C=C as the sole indicator of the presence of the double bond, unless the bond is conjugated with a second unsaturated site.

Table 3 Aromatic ring (aryl) group frequencies

Origin	Group frequency, wavenumber (cm^{-1})	Assignment
C=C–C[a]	1615–1580	Aromatic ring stretch
C=C–C[a]	1510–1450	Aromatic ring stretch
C–H	3130–3070	Aromatic C–H stretch
C–H	1225–950 (several)	Aromatic C–H in-plane bend
C–H	900–670 (several)	Aromatic C–H out-of-plane bend
	770–730 + 710–690	Monosubstitution (phenyl)
	770–735	1,2-Disubstitution (ortho)
	810–750 + 900–860	1,3-Disubstitution (meta)
	860–800	1,4-Disubstitution (para)
"Combi"[b]	2000–1660 (several)	Aromatic combination bands

[a] C=C–C used as an approximation of the unique aromatic ring bonding.
[b] "Combi" denotes assignment to combination bands.

3.1.3 Aromatic Compounds

See Table 3. The existence of one or more aromatic rings in a structure is normally readily determined from the C–H and C=C–C ring-related vibrations. The C–H stretching occurs above 3000 cm^{-1} and is typically exhibited as a multiplicity of weak-to-moderate bands, compared with the aliphatic C–H stretch. The structure of the bands is defined by the number and positions of the C–H bonds around the ring, which in turn are related to the nature and number of other substituents on the ring. Note that the same applies to the C–H out-of-plane bending vibrations, which are frequently used to determine the degree and nature of substitution on the ring – examples are provided in Figure 6(a–c), with the comparison of the three xylene isomers. This picture often becomes more complex if multiple- or fused-ring structures exist in a compound. The other most important set of bands are the aromatic ring vibrations centered around 1600 and 1500 cm^{-1}, which usually appear as a pair of band structures, often with some splitting. The appearance and ratio of these band structures is strongly dependent on the position and nature of substituents on the ring.

3.1.4 Acetylenic Compounds

See Table 4. Although acetylenic compounds are not very common, the spectrum associated with the C≡C structure can be characteristic. It is instructive to note the impact on the carbon–carbon bond stretching as a function of increase in bond order for the series of single-, double-, and triple-bonded carbon:

C–C stretch: ~1350–1000 cm^{-1} (skeletal vibrations)
C=C stretch: 1680–1620 cm^{-1}
C≡C stretch: 2260–2100 cm^{-1}

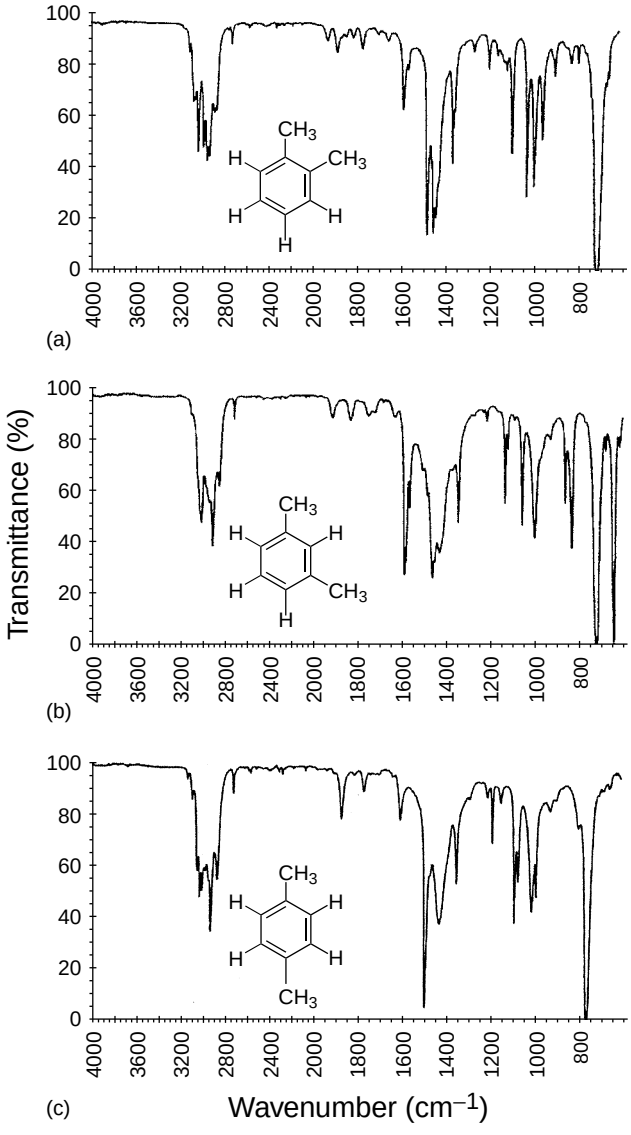

Figure 6 ATR spectra of xylene isomers: (a) *o*-xylene, 1,2-dimethylbenzene; (b) *m*-xylene, 1,3-dimethylbenzene; (c) *p*-xylene, 1,4-dimethylbenzene. Copyright Coates Consulting.

Table 4 Acetylenic (alkyne) group frequencies

Origin	Group frequency, wavenumber (cm^{-1})	Assignment
C≡C	2140–2100	Terminal alkyne (monosubstituted)
C≡C	2260–2190	Medial alkyne (disubstituted)
C–H	3320–3310	Alkyne C–H stretch
C–H	680–610	Alkyne C–H bend
C–H	630 (typical)	Alkyne C–H bend

This increase in bond order produces a corresponding increase in bond strength, which in turn increases the force

constant, κ (see Equation 3), supporting the Hooke's law model described earlier.

As noted in Table 4, the position of the C≡C bond is influenced by whether the group is terminal or medial. The single hydrogen of the terminal acetylene itself is very characteristic, reflecting the labile nature of the acetylenic C–H.

3.2 Simple Functional Groups

Obviously, there is a potentially broad number of molecular fragments that can be considered to be functional groups attached to an organic structure or backbone. This section features the most simple and most common of the functional groups, –C–X, i.e. the halogens (X = F, Cl, Br and I), hydroxy (X = OH), oxy or ether (X = OR, where R = alkyl), and amino ($X = NH_2$, =NH or ≡N). With the exception of the carbonyl functionality, these three basic functional groups cover most of the common occurrences in simple organic compounds. Note that for the oxy/hydroxy and amino functionalities, these are molecular fragments, and they contribute their own set of characteristic absorptions to the spectrum of the compound. In fact, the bonding between the functional group and the backbone is only one part of the overall picture used for the spectral interpretation.

3.2.1 Halogenated Compounds

See Table 5. In principle, the interpretation of the spectra of molecules containing one or more halogens would seem to be straightforward. The functionality is simple, with just a single atom linked to carbon to form the group. With the polar nature of this group, one would expect the spectral contribution to be distinctive. In reality, this is not always the case.

In aliphatic compounds, the C–X bond typically possesses a unique group frequency, which may be assigned to the halogen–carbon stretching. When a single halogen is present, the determination of this group is straightforward. However, if more than one halogen is present, the interpretation is usually more complex. In such cases, the result varies depending on whether the halogens are on the same or different carbon atoms, and, if on different atoms, whether the atoms are close neighbors. This is particularly the case with small molecules, and the resultant spectral complexity arises from the fact that there is restricted rotation about the carbon–carbon bond.

Single bonds usually exhibit free rotation, which would normally mean that there are no preferred spatial orientations for the molecules. However, owing to the size of the halogen atom, relative to the carbon and hydrogen that form the backbone, the molecules tend to exhibit certain specific conformations, where the spatial interaction between neighboring halogen atoms is minimized, and each conformation provides its own contribution to the overall spectrum. It is important to appreciate that this issue of spatial orientation has an impact even on high-molecular-weight compounds, such as the polyhalogenated polymers, e.g. poly(vinyl chloride). Here preferred orientations have an impact on the crystallinity of the polymer, and this in turn has a significant impact on both the spectrum and the physical properties of the material.

Another important issue to consider with halogen substituents is the high electronegativity of the halogen atom. This can have a noticeable impact on the spectrum of neighboring group frequencies, including adjacent hydrogen atoms. In such cases, significant shifting of the C–H frequencies can occur – the direction of the shift being dependent on the location of the C–H, and whether the halogen adds or extracts electron density from the C–H bond – adding strengthens (higher frequency) and extracting weakens (lower frequency). The same influences can be observed with halogen-substituted carbonyl compounds, such as acyl halides and α-substituted acids, where the bond strength of the carbonyl group is increased (see section 3.3). In most cases, both a shift to higher frequency and an increase in absorption strength for the band are observed.

Table 5 only presents the group frequencies for the aliphatic compounds, because no well-defined C–X absorptions are observed for halogen-substituted aromatic compounds. The presence of a halogen on an aromatic ring can be detected indirectly from its electronic impact on the in-plane C–H bending vibrations. Normally, we do not consider the in-plane bending bands to be of use because, as pointed out earlier, these occur in a spectral region that is crowded by other important group frequencies. However, in the case of a halogen-substituted ring, the intensity of these vibrations is enhanced relative

Table 5 Aliphatic organohalogen compound group frequencies

Origin	Group frequency, wavenumber (cm^{-1})[a]	Assignment
C–F	1150–1000	Aliphatic fluoro compounds, C–F stretch
C–Cl	800–700	Aliphatic chloro compounds, C–Cl stretch
C–Br	700–600	Aliphatic bromo compounds, C–Br stretch
C–I	600–500	Aliphatic iodo compounds, C–I stretch

[a] Note that the ranges quoted serve as a guide only; the actual ranges are influenced by carbon chain length, the actual number of halogen substituents, and the molecular conformations present.

to other absorptions by as much as three to four times. For reference, it is informative to compare the intensities for these bands, between 1150 and 1000 cm^{-1}, for the spectra of toluene and chlorobenzene (Figure 7a and b).

3.2.2 Hydroxy and Ether Compounds

See Table 6 for alcohols and hydroxy compounds. The hydroxy function is probably one of the most dominant and characteristic of all of the infrared group frequencies. In most chemical environments, the hydroxy group does not exist in isolation, and a high degree of association is experienced as a result of extensive hydrogen bonding with other hydroxy groups. These hydroxy groups may be within the same molecule (intramolecular hydrogen bonding) or they most likely exist between neighboring molecules (intermolecular hydrogen bonding). The impact of hydrogen bonding is to produce significant band broadening and to lower the mean absorption frequency. The lowering of the frequency tends to be a function of the degree and strength of the hydrogen bonding. In compounds such as carboxylic acids, which exhibit extremely strong hydrogen bonding, forming a stable dimeric structure, a highly characteristic, large shift to lower frequencies is observed.

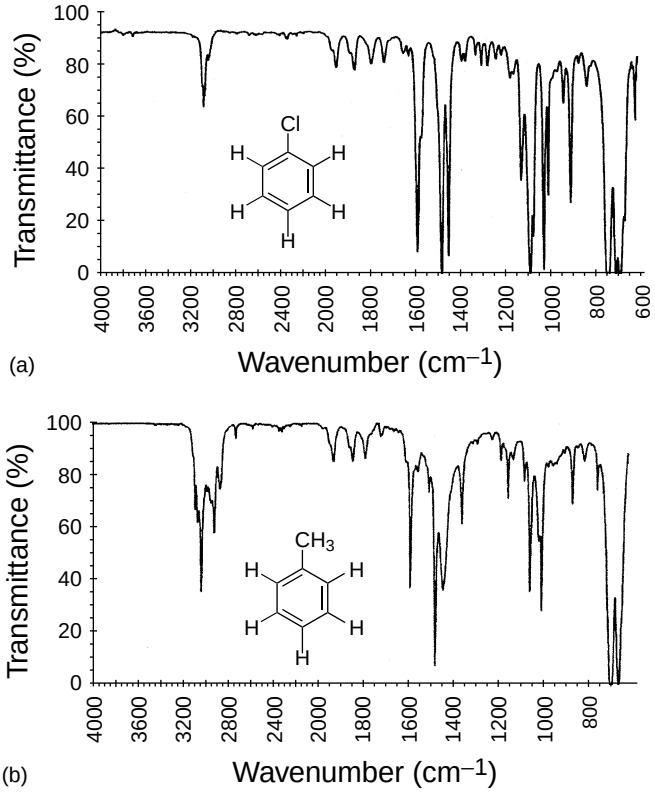

Figure 7 Comparison of ATR spectra of (a) chlorobenzene and (b) toluene. Copyright Coates Consulting.

Table 6 Alcohol and hydroxy compound group frequencies

Origin	Group frequency, wavenumber (cm^{-1})	Assignment
O–H	3570–3200 (broad)	Hydroxy group, H-bonded OH stretch
	3400–3200	Normal "polymeric" OH stretch
	3550–3450	Dimeric OH stretch
	3570–3540	Internally bonded OH stretch
O–H	3645–3600 (narrow)	Nonbonded hydroxy group, OH stretch
	3645–3630	Primary alcohol, OH stretch
	3635–3620	Secondary alcohol, OH stretch
	3620–3540	Tertiary alcohol, OH stretch
	3640–3530[a]	Phenols, OH stretch
O–H	1350–1260	Primary or secondary, OH in-plane bend
	1410–1310	Phenol or tertiary alcohol, OH bend
	720–590	Alcohol, OH out-of-plane bend
C–O	~1050[b]	Primary alcohol, C–O stretch
	~1100[b]	Secondary alcohol, C–O stretch
	~1150[b]	Tertiary alcohol, C–O stretch
	~1200[b]	Phenol, C–O stretch

[a] Frequency influenced by nature and position of other ring substituents.
[b] Approximate center of range for the group frequency.

In special circumstances, where the hydroxy group is isolated – either because of steric hindrance effects or because the sample is in the vapor state or in a dilute solution of a nonpolar solvent – the band is characteristically narrow, and is observed at the natural, higher frequency. This absorption is important for the characterization of certain hindered phenol antioxidants, a commercially important class of compounds in the food, polymer, and formulated oil industries.

It must be appreciated that while the hydroxy absorption is singly one of the most important bands in the infrared spectrum, other vibrations are also important for the actual characterization of the compound. Alcohols exist as three distinct classes – primary, secondary and tertiary – distinguished by the degree of carbon substitution on the central hydroxy-substituted carbon, a single substitution being primary, double substitution being secondary, and triple substitution being tertiary. This is an important fact, because the chemistry and oxidation stability of the alcohol are strongly influenced by the degree of substitution. Whether an alcohol is primary (1°), secondary (2°) or tertiary (3°), may be reflected in the position of the OH stretch absorption, but typically this is determined by the other absorptions, in particular the C–O– stretching frequency. Another absorption of lower importance, but often characteristic, is assigned

to another form of bending vibration, the out-of-plane bend or wagging vibration of the O–H. The OH bending vibrations are broadened by hydrogen bonding as is the stretching absorption, but often to a lesser extent. The differences between primary and secondary alcohols can be appreciated from Figure 8(a) and (b), where the spectra of 1- and 2-octanol are presented.

See Table 7 for ethers and oxy compounds. In some respects, ethers are related to alcohol and hydroxy compounds, where the hydrogen of the hydroxy group is replaced by an aliphatic (alkyl) or aromatic (aryl) molecular fragment. Having stated that, the overall appearance of an ether spectrum is drastically different from that of a related alcohol. This is due to the overwhelming effect of hydrogen bonding on the hydroxy group. However, many of the relationships that exist for the C–O component of the alcohol carry over to the corresponding ether. The relationships that pertain to primary, secondary, and tertiary structures remain intact. The main difference is that one now considers the bonding on both sides of the oxygen, because if carbon is on both sides, then two ether bonds exist. Ethers can exist as simple ethers (same group both sides) and mixed ethers (different groups both sides). Infrared spectroscopy is fairly sensitive for differentiating these ether functions, especially when the structures are mixed aliphatic or aliphatic/aromatic.

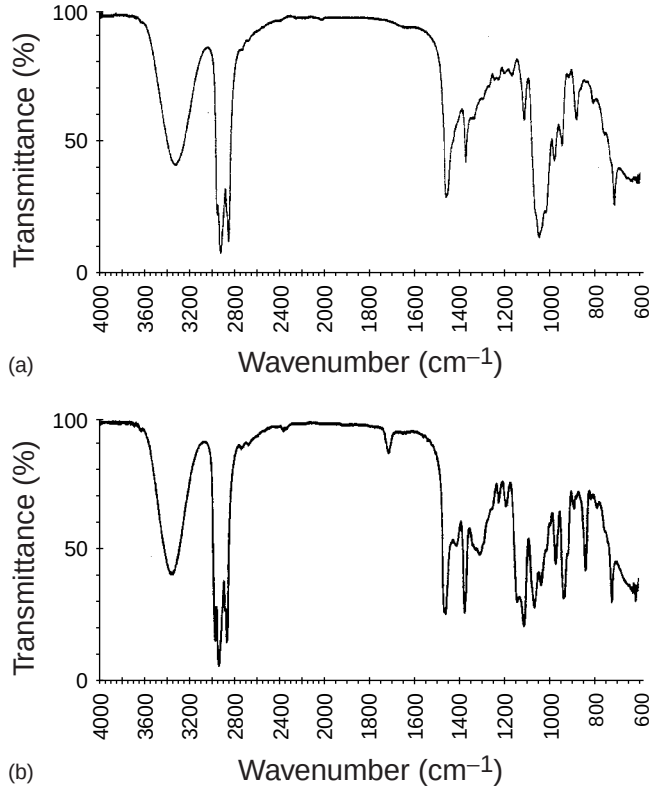

Figure 8 ATR spectra of (a) primary (1-octanol) and (b) secondary (2-octanol) alcohols. Copyright Coates Consulting.

Table 7 Ether and oxy compound group frequencies

Origin	Group frequency, wavenumber (cm^{-1})	Assignment
C–H	2820–2810	Methoxy, C–H stretch (CH_3–O–)
C–O–C	1150–1050	Alkyl-substituted ether, C–O stretch
C–O–C	1140–1070	Cyclic ethers, large rings, C–O stretch
ϕ–O–H	1270–1230	Aromatic ethers, aryl–O stretch
C–O–	~1250 + 890–800[a]	Epoxy and oxirane rings
C–O–O–C	890–820[a]	Peroxides, C–O–O– stretch

[a] Typically very weak, and not very characteristic in the infrared. Tend to be more characteristic in the Raman spectrum.

3.2.3 Amino Compounds

See Table 8. In some respects, the infrared spectra and the characteristic group frequencies of amines tend to

Table 8 Amine and amino compound group frequencies

Origin	Group frequency, wavenumber (cm^{-1})	Assignment
		Primary amino
N–H	3400–3380 +3345–3325	Aliphatic primary amine, NH stretch
N–H	3510–3460 +3415–3380	Aromatic primary amine, NH stretch
N–H	1650–1590	Primary amine, NH bend
C–N	1090–1020	Primary amine, CN stretch
		Secondary amino
>N–H	3360–3310	Aliphatic secondary amine, NH stretch
>N–H	~3450	Aromatic secondary amine, NH stretch
>N–H	3490–3430	Heterocyclic amine, NH stretch
=N–H	3350–3320	Imino compounds, NH stretch
>N–H	1650–1550	Secondary amine, NH bend
C–N	1190–1130	Secondary amine, CN stretch
		Tertiary amino
C–N	1210–1150	Tertiary amine, CN stretch
		Aromatic amino
C–N	1340–1250	Aromatic primary amine, CN stretch
C–N	1350–1280	Aromatic secondary amine, CN stretch
C–N	1360–1310	Aromatic tertiary amine, CN stretch

parallel those of alcohols and ethers. Before proceeding, there is an important distinction to be made between the nomenclature used to describe alcohol structures, compared with that used for amines. The terms primary, secondary, and tertiary are used to describe amines, but the substitution relates to the nitrogen, not the adjoining carbon (as with alcohols). The terminology used is $-NH_2$ primary, >NH secondary and >N– tertiary. As before, these structural differences are important, and they strongly influence the chemistry and the reactivity of the nitrogen and the N–H group (primary and secondary). Again, the infrared spectrum is very diagnostic, and it is possible to differentiate readily the functional group structures, especially for the primary and secondary amino compounds (see Figure 9a–c for an example comparison of primary, secondary, and tertiary aromatic amines).

As with hydroxy compounds, hydrogen bonding is important, but the effect tends to be weaker than for the hydroxy group, and the overall effect on the spectrum is slightly less pronounced. This situation alters in the related ammonium and amino salts, where strong hydrogen bonding is experienced, and a corresponding broadening of the associated NH absorptions is observed.

Note that only the primary and secondary amines exhibit the most characteristic group frequencies, which are associated with the N–H bond. Tertiary amines are very comparable to ethers, and the main diagnostic information is gathered from the C–N vibrations only. An important exception is the methylamino (and dimethylamino) group, which like its counterpart (methoxy in ethers) has a characteristic C–H stretching vibration at lower than the normal C–H stretching frequency for methyl.

3.3 The Carbonyl Group

Carbonyl compounds are not only chemically important, but are also important in the interpretation of infrared spectra. The C=O absorption is almost always one of the most characteristic in the entire spectrum, and it is also most likely to be the most intense spectral feature. Table 9 provides an example listing of some of the common carbonyl frequencies as a function of the particular type of carbonyl group. In essence, a ketone is considered the root compound, with an aldehyde being a special case, where the carbonyl group is terminal, and only has one substituent, the other being a single hydrogen atom. All of the other carbonyl compounds, in a way, can be considered to be derived from the base ketone structure, where one or both alkyl (or aryl) substituents are replaced by another functionality, e.g. from a single hydroxy group, in the case of carboxylic acids, to two ether groups, as in the case of an organic carbonate.

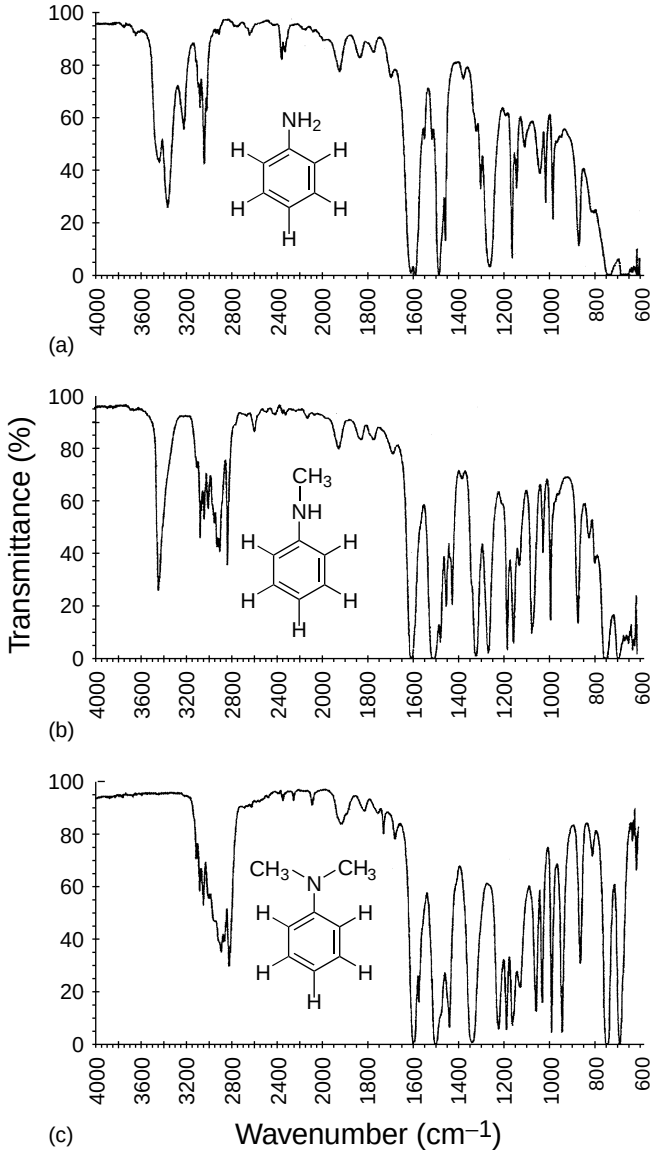

Figure 9 ATR spectra of (a) primary (aniline), (b) secondary (*N*-methylaniline) and (c) tertiary (*N*,*N*-dimethylaniline) aromatic amines. Copyright Coates Consulting.

The actual diagnostic carbonyl absorption frequency is dependent on the electronic characteristics of the substituent group, which in turn define very much the chemical characteristics and reactivity of the specific carbonyl compound. Spatial and structural factors can be important, in particular ring stress, as in the case of lactones (cyclic esters) and lactams (cyclic amides). In this case, the greater the ring stress, the higher is the carbonyl absorption frequency. In common with previous observations for a double-bonded functionality, conjugation plays an important role in the observed carbonyl frequency. This includes connection to an aromatic ring or conjugation to a C=C or another C=O.

Table 9 Example carbonyl compound group frequencies

Group frequency (cm^{-1})	Functional group
1610–1550/1420–1300	Carboxylate (carboxylic acid salt)
1680–1630	Amide
1690–1675/(1650–1600)[a]	Quinone or conjugated ketone
1725–1700	Carboxylic acid
1725–1705	Ketone
1740–1725/(2800–2700)[b]	Aldehyde
1750–1725	Ester
1735	Six-membered ring lactone
1760–1740	Alkyl carbonate
1815–1770	Acid (acyl) halide
1820–1775	Aryl carbonate
1850–1800/1790–1740	Open-chain acid anhydride
1870–1820/1800–1775	Five-membered ring anhydride
2100–1800	Transition metal carbonyls

[a] Lower frequency band is from the conjugated double bond.
[b] Higher frequency band characteristic of aldehydes, associated with the terminal aldehydic C–H stretch.

As previously discussed, a lowering of the parent group frequency is observed. This effect is important for the differentiation of certain types of carbonyl compound, in the determination of whether the carbonyl group is directly or indirectly attached an aromatic ring, e.g. the ability to differentiate aryl acetates from alkyl benzoates. In the case of the acetate, the ring is joined to the ''ether oxygen'' of the ester group, and is not conjugated with the carbonyl, whereas with the benzoate, the ring is directly conjugated with the group, and the carbonyl absorption frequency is correspondingly lowered.

Often, the frequency ranges for the different classes of carbonyl compound overlap, and the carbonyl frequency alone is not sufficient to characterize the functional group. In most cases, spectral information from the other component of the functional group is used for the characterization. Carboxylic acids are a good example, where the C–O, C–O–H and O–H vibrations are highly characteristic. Similarly, esters (C–O–C) and amides (C–N and N–H) are other frequently encountered examples. The frequencies provided earlier for these additional functionalities (C–O, C–N, and N–H) in general apply, although the actual observed frequencies may differ slightly, being modified by the carbonyl group. It is worthwhile returning momentarily to carboxylic acids, because they are unique, in so far as the hydroxy group has direct interaction with the carbonyl group, by the formation of a stable dimeric hydrogen-bonded structure in the condensed phase (solid and liquid). Note that this structure disappears in the vapor state. A characteristic broad feature in the range 3300–2500 cm^{-1}, that overlaps the C–H stretching region, and with a secondary absorption close to 2600 cm^{-1}, is observed for the hydrogen-bonded O–H of most carboxylic acids. The impact of this overlap can be appreciated by reference to the spectrum of butyric acid, provided in Figure 10. Other bands that are associated with the C–O and O–H components tend to be less pronounced, and sometimes may be overlapped with other fingerprint absorptions of the molecule. These are located in the ranges 1320–1210 cm^{-1} (C–O stretch) and 960–850 cm^{-1} (hydrogen-bonded O–H out-of-plane bending).

3.4 Other Functional Groups Associated with Heteroatoms

Potentially there are very large numbers of different organic-based compounds that are associated with one or more heteroatoms. These are in addition to the simple halogen- and amino-based compounds that have already been covered. A detailed discussion of such compounds is beyond the scope of this article. A few illustrative examples are included here in order to provide the reader with a feel for the spectral contributions of commonly encountered compounds that feature heteroatoms. These include triple-bonded and cumulated double-bonded nitrogen compounds, such as cyano and cyanato compounds, hetero-oxy compounds, such as nitro, sulfoxy, phosphoroxy and silicoxy compounds, and thiols and thio-substituted compounds. Many of these belong to important compound classes, and complete texts exist that focus on the spectral attributes of such compounds.

3.4.1 Multiple-bonded and Cumulated Double-bonded Nitrogen Compounds

Nitrogen compounds featuring triple or cumulated double bonds, such as cyanides or nitriles (–C≡N) and cyanates (–O–C≡N), all provide a unique spectrum, typically with a single, normally intense absorption at 2280–2200 cm^{-1}

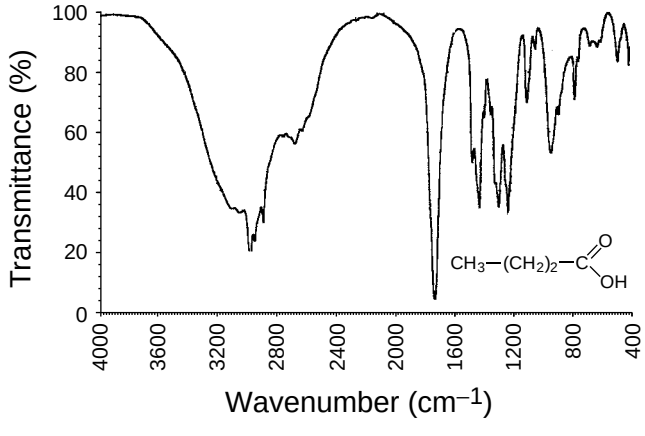

Figure 10 Transmission spectrum of *n*-butyric acid. Copyright Coates Consulting.

(for cyano compounds) and 2285–1990 cm^{-1} (cyanates, isocyanates, thiocyanates, etc.). The band is assigned to the stretching vibration, which is the asymmetric stretch in the case of the cumulated double-bonded compounds (the symmetric stretch is typically weak, and is not diagnostic of the group).

Double-bonded nitrogen groups, such as imino groups (>C=N–) and azo groups (–N=N–), exhibit absorptions close to the carbonyl (C=O) and alkene (C=C) double bond stretching region. While they are characteristic for the functional group, they are sometimes difficult to assign from first principles because of the overlap with other common functional groups in the region. For example, dependent on substitution or location, the >C=N– group can occur in the same spectral region as the C=O of an amide. Examples of the group frequencies for a few common multiple-bonded and cumulated double-bonded compounds are provided in Table 10.

3.4.2 Hetero-oxy Compounds

The group of compounds covered here feature X–O (where X = nitrogen, sulfur, phosphorus, and silicon) and X=O vibrations. In general, the characteristic absorptions span a similar range to that covered by the corresponding carbon–oxygen compounds, with most of the stretching frequencies observed within the main fingerprint spectral region (1500–400 cm^{-1}). Many of the compounds can be considered to be analogs of ethers, especially when an alkoxy group is present, featuring the X–O–C linkage. Interestingly, the nitro group ($-NO_2$) is isoelectronic with the carboxylate ion group ($-CO_2^-$), and both provide very similar spectra for the main functional group. See Table 11 for example group frequencies for hetero-oxy compounds.

3.4.3 Thiols and Thio-substituted Compounds

Thiols and thio-substituted compounds, by definition, can be considered to be the direct analogs of the equivalent oxygenated compounds such as alcohols and ethers. Unlike the oxygen-containing analogs, the equivalent C–S and C–S–H stretching vibrations tend to give rise to very weak absorptions in the infrared spectrum. The C–S and S–H bonds are highly polarizable, and hence produce stronger spectral activity in the Raman spectrum than the infrared spectrum. The higher mass of sulfur, compared with oxygen, results in the characteristic group frequencies occurring at noticeably lower frequencies than the oxygen-containing analogs, as noted in Table 12 for compounds containing S–H and C–S bonds. The thiol S–H is probably the only infrared absorption that can be considered to be of use for the general characterization of these compounds, and most of the others are obtained by inference, or from knowledge of the compound type. Catenation is a unique chemical characteristic of sulfur, where the formation of S–S bonds in extended chains is common, hence the inclusion here of the S–S stretching frequencies.

Table 10 Examples of nitrogen multiple and cumulated double bond compound group frequencies

Group frequency (cm^{-1})	Functional group/assignment
2280–2240	Aliphatic cyanide/nitrile
2240–2220	Aromatic cyanide/nitrile
2260–2240/1190–1080	Cyanate (–OCN and C–OCN stretch)
2276–2240	Isocyanate (–N=C=O asym. stretch)
2175–2140	Thiocyanate (–SCN)
2150–1990	Isothiocyanate (–NCS)
1690–1590	Open-chain imino (–C=N–)
1630–1575	Open-chain azo (–N=N–)

Table 11 Example group frequencies for simple hetero-oxy compounds

Group frequency (cm^{-1})	Functional group/assignment
	Nitrogen-oxy compounds
1560–1540/1380–1350[a]	Aliphatic nitro compounds
1555–1485/1355–1320[a]	Aromatic nitro compounds
1640–1620/1285–1270[a]	Organic nitrates
	Phosphorus-oxy compounds
1350–1250	Organic phosphates (P=O stretch)
1050–990	Aliphatic phosphates (P–O–C stretch)
1240–1190/995–850	Aromatic phosphates (P–O–C stretch)
	Sulfur-oxy compounds
1335–1300/1170–1135[a]	Dialkyl/aryl sulfones
1420–1370/1200–1180[a]	Organic sulfates
1365–1340/1200–1100[a]	Sulfonates
	Silicon-oxy compounds
1095–1075/1055–1020	Organic siloxane or silicone (Si–O–Si)
1110–1080	Organic siloxane or silicone (Si–O–C)

[a] Asymmetric/symmetric XO_2 stretch (NO_2 and SO_2).

3.5 Simple Inorganics

Characterization of compounds via infrared spectroscopy is not limited to organic compounds. Any inorganic compound that forms bonds of a covalent nature within a molecular ion fragment, cation or anion, will produce a characteristic absorption spectrum, with associated group frequencies. In a manner, certain aspects

Table 12 Common group frequencies for thiols and thio-substituted compounds

Group frequency (cm^{-1})	Functional group/assignment
2600–2550	Thiols (S–H stretch)
710–685	Thiol or thioether, CH_2–S– (C–S stretch)
660–630	Thioethers, CH_3–S– (C–S stretch)
715–670	Aryl thioethers, ϕ–S (C–S stretch)
705–570	Disulfides (C–S stretch)
620–600	Disulfides (S–S stretch)
500–430	Aryl disulfides (S–S stretch)
500–470	Polysulfides (S–S stretch)

Table 13 Example group frequencies for common inorganic ions

Group frequency (cm^{-1})	Functional group/assignment
1490–1410/880–860[a]	Carbonate ion
1130–1080/680–610[a]	Sulfate ion
1380–1350/840–815[a]	Nitrate ion
1100–1000	Phosphate ion
1100–900	Silicate ion
3300–3030/1430–1390[a]	Ammonium ion
2200–2000	Cyanide ion, thiocyanate ion, and related ions

[a] Typically, the first absorption is intense and broad, and the second has weak to medium intensity and is narrow. Both often exist as multiple band structures, and this may be used to characterize individual compounds.

have already been covered for inorganic compounds in the form of salts of carboxylic acids and amino and ammonium compounds, which can be extended to metal complexes, chemical fragments associated with hetero-oxy groups (nitrates, sulfates, phosphates, silicates, etc.), and transition metal carbonyl compounds. All complex ionic compounds (containing more than one atom) and coordination compounds produce characteristic spectra. Many of the associated group frequencies can be used diagnostically for characterization. The structure and orientation of the ion or complex, both as an isolated entity or within a crystal lattice, are important factors that affect the appearance and nature of the infrared spectrum. Hydration of compounds (water of crystallization) also has a large effect on the spectrum, and often adds a lot of complexity, in the form of additional absorption bands and structure to existing bands. The subject is far too broad to be covered in this article, and only a few example group frequencies are included here (Table 13). The reader is directed to the recommended standard texts for more detailed information.[13,14]

4 THE PRACTICAL SITUATION – OBTAINING THE SPECTRUM AND INTERPRETING THE RESULTS

Up to this point, the fundamentals of interpretation have been discussed from the most basic concepts of infrared absorption by a molecular species and the impact of chemical functionality on the resultant spectrum. In many ways, this discussion has treated the molecule as a more or less isolated species, with no consideration of the physical state of the actual sample or the environment in which the molecule exists. Physical state and the molecular, chemical and physical environments have a profound effect on the infrared spectrum. As a result, it is just as important to understand and interpret these effects as it is to perform the fundamental interpretation of the functional groups from first principles.

This particular section may be one of the most important for many of the readers, because it reflects the real world, that is taking a sample, preparing it for analysis, and making use of all of the available information on the sample and the spectrum or spectra obtained. Note that sometimes there are several options in the way that a sample may be handled for infrared analysis. It is often beneficial to consider the different options. Before we start to examine the situation it is important to understand the importance of the interpretation and to determine the real requirements. Here are some example scenarios:

1. The sample (or spectrum) is a ''total unknown'' and an identification is required – examples include forensic samples, environmental waste samples, or new discovery samples, where a new material has been synthesized or discovered.
2. The sample (or spectrum) is an unknown and it needs to be characterized or classified – examples include commercial applications where new additives or components are included in a material to provide a specific property; in such cases this could be considered the basis of competitive product analysis.
3. The sample generally is known but the existence of a specific chemical class needs to be determined – examples include contaminant analysis, analysis for toxicology or environmental reasons, material additives, etc.
4. The sample is a complete known and the interpretation is required to confirm the material composition and/or quality – examples include product quality control and the confirmation of a structure or functionality of a newly synthesized material.

There are clearly many other scenarios that can be considered, and the examples presented may cross over into more than one of these scenarios. However, the idea here is to help define a strategy for the interpretation. For the most part, this text will focus on scenarios 1 and 2, although a passing reference may be made to the others. Note also that there are two ways to address the interpretation:

- one is to attempt to assign as many bands as possible using group frequency tables;
- the other is to attempt to build up a picture using information from key spectral bands (with a knowledge of the group frequencies) and rationalizing the spectrum relative to the sample as known, and any known history, including the sample preparation and presentation technique.

The latter approach is the best way to go, and is typically far more rewarding. Also, it is seldom that all bands in the infrared spectrum of a sample can be fully or accurately assigned. There are too many ambiguities and coincidences for a definitive interpretation to be obtained from the spectra of most materials – remembering, of course, that the sample may be a mixture.

To begin, let us start from either the point where a spectrum has been generated or where a sample is presented. In practice, these are very different starting points. The latter is always preferred because with the sample in hand it is possible to gain a lot of first-hand information about nature which will ultimately help in the final interpretation. The term ''unknown'' is used because this is how people often view the interpretation process; however, in reality there are seldom true unknowns – in other words, the reason for the interpretation itself often provides implied information about the sample. Even if a person presents a prerecorded spectrum, it is usually accompanied by some additional information. This at minimum should include a knowledge of how the sample was prepared, whether it is a solid, liquid, or gas, or some mixed state, and maybe some physical characteristics (color, appearance, melting point/boiling point, etc.). For the benefit of presenting a complete picture, we shall assume that the sample is available and that we have the opportunity to perform additional tests and measurements on it. Further, more detailed discussions of this practical approach can be found in the literature.[18–20]

4.1 Sample History

As noted, we are seldom dealing with a true unknown, and typical situations include the following:

- a sample of fiber taken or extracted from a particular environment;
- a contaminant removed from a material;
- a suspicious liquid found leaking out of some drums;
- a residue extracted from a surface;
- a residual liquid or solid remaining after storage or treatment of a product;
- extracted additives or components from a commercial product;
- a mass that was produced after a specific chemical reaction;
- or simply a manufactured product.

Most people can equate with these different sets of circumstances. Each one brings information about the sample.

In the case of a fiber, if it is organic, then it typically fits within certain classes of polymeric materials. It may have orientation properties, which will influence the appearance of the infrared spectrum, depending on how it is sampled, and if there is more than one fiber, then the original material may have been a blend. The environment from where the sample was taken will be important because there may be surface contaminants that may produce spurious bands in the spectrum, which may be removed by some sample pretreatment, such as washing with a solvent. Alternatively, the fiber may be coated, or degraded, thermally or by oxidation.

The characterization of contaminants is a common application for infrared spectroscopy. For this type of exercise it is necessary to know the importance of the contaminant and why it is believed to be a contaminant. Was it extracted from a liquid or solid matrix? if so, what type of matrix, and what was the composition of the matrix? Is it possible that the contaminant originated from the matrix itself, as a breakdown product? – if so, then there should be some chemical similarities or common heritage. Was it embedded within a matrix, such as a plastic/polymer film or a sheet of paper? – if so, is the contaminant itself potentially contaminated by the matrix? This is especially important for interpretation, because the matrix may have a very intense infrared spectrum of its own.

For extracted materials, such as additives removed from a polymer, it is important to know something about the polymer matrix – few extractions are 100% efficient, and it is common to have carry-over. In many such cases, the spectrum obtained reflects that of a mixture rather than a pure compound. Things to bear in mind here are what the original base polymers were, what other additives could be present, and what material, usually a solvent, was used for the extraction.

Looking at residues and residual products of a reaction is similar to the concept of characterizing a contaminant. Residues formed within products are often the result of a breakdown reaction or an unexpected interaction between components. If derived from one or more of the components, then a knowledge of all of the ingredients is essential – again we would be looking for chemically related species in the residue, and reference spectra of each of the ingredients would be important in attempting to characterize the material. In the case of a reaction product, spectra for the reactants and any solvents or substrates (catalysts, etc.) used would be helpful.

Finally, the identification of an end product from its infrared spectrum is common practice. By definition, unless the final product is a pure chemical, most manufactured products are mixtures, often featuring blended solid and/or liquid components from percentage range concentrations to trace amounts. Often at issue is how much can be identified from the interpretation of a single spectrum. It is often necessary to perform some form of preconcentration on one or more ingredients, or to perform a chemical modification to change the spectral response of one or more of the major ingredients. This latter operation may help to remove the spectral interferences from a major component for determining the presence of minor components.

4.2 Physical Characteristics of the Sample

From considerations based on the discussions in the section above, a good working knowledge of the sample allows one to determine basic information about the sample. This information can help in the determination of the best method of sampling. It can also help to indicate if any special treatments to the sample are necessary before proceeding, such as the removal of residual solvents, or the need for selective extraction, in the event that the sample has more than one phase, or is known to be a mixture. Beyond this there are other important physical characteristics of the sample that can assist the interpretation process. As indicated, physical state can be very important. For example, it has been noted that nitro compounds are isoelectronic with organic acid salts. Many simple nitro compounds are liquids, whereas simple carboxylates are usually solids.

There are, however, many other physical characteristics and attributes that can be linked to the sample and its chemistry. This in turn can be linked to important spectral features in the infrared spectrum. Two important characteristics are color and odor. Nitrogen compounds often have color associated with them, e.g. nitro-substituted compounds often have a characteristic yellow or orange color, especially substituted phenols and amines. Similarly, unsaturated nitrogen compounds, such as azo compounds, and highly conjugated olefinic compounds, such as carotenes, are similarly colored. This can be used as a guide to look for the specific functionality in the spectrum.

Likewise, odor can be a useful guide. This author is not advocating that a person should inhale the vapor of a sample directly, for obvious reasons. However, if under cautious sampling conditions a characteristic odor is detected, then this can provide a valuable insight regarding the presence of particular functional groups. Common examples are alcohols, esters, and ketones, which often possess relatively pleasant fragrances, especially when associated with some degree of unsaturation. An almond-like odor might indicate a nitro compound, benzaldehyde, or a cyano compound (caution!), a fish-like odor is often associated with amino or amido compounds, and a putrid or ''bad-cheese'' odor is associated with certain carboxylic acids. With time, one can build up a mental library of common odors and fragrances, and these can be very valuable for initial material screening, by helping one look for the presence of specific functional groups.

Volatility and flammability can be useful guides. Most volatile compounds tend to have a relatively low molecular weight, and they usually are unassociated. This latter point helps to rule out certain hydrogen-bonded species, such as carboxylic acids, amines, and multiple hydroxy-substituted compounds (such as glycols). If a sample is submitted for some form of combustion analysis, it is worthwhile monitoring how it burns. If it burns with a clean, bluish flame, that might indicate that the sample is oxygenated (alcohol or ether). If it has a yellow and/or smoky flame, it indicates an excess of carbon, possible unsaturation or an aromatic compound. The latter can produce a very smoky flame.

4.3 The Chemistry of the Sample

First, there are issues relating to solubility. The old adage, like dissolves like, is very appropriate. Nonpolar materials such as hydrocarbons tend to be more soluble in nonpolar solvents, such as hydrocarbons or chlorinated solvents, whereas polar materials favor solvents such as methanol and acetone, and in extreme cases their solubility may be limited to the most polar solvents, such as water. If a material is potentially identified from its infrared spectrum as a carboxylic acid or an amino compound, then it is worthwhile to check the pH. Also, it is worthwhile to check its ability to form a salt – carboxylic acids react with sodium carbonate or sodium hydroxide to form a carboxylate salt, and amines react with dilute mineral acids, such as hydrochloric acid, to form amine salts (amine hydrochlorides with hydrochloric acid).

Certain compounds react with water; examples are acid halides and acid anhydrides. Alternatively, these

compounds will react with alcohols and amines to form esters and amides. The formation of such derivatives is common practice in the identification of organic compounds, and can be extremely informative in a definitive identification by infrared spectroscopy. Other chemical modifications that can be equally informative are for the identification of olefinic or unsaturated compounds. A simple reaction with bromine (caution – use in solution, or prepare in situ) will often confirm the presence of a double bond. The reaction product can be helpful in characterizing the nature of the unsaturation, especially in the case of a *cis* configuration, which is often spectrally masked by other absorptions in the region.

One similar modification, which is strictly not chemical in nature, is deuterium exchange. In this case, a compound with either a labile hydroxy group or an amino group is placed in equilibrium with a deuterated solvent, such as deuterium oxide (heavy water). The replacement of the O–H in the molecule by O–D, in the case of hydroxy compounds, can help in the confirmation of the hydroxy group, and it can shift an overlapping hydroxy function from the C–H stretching and bending regions, thereby assisting in the determination of the backbone structure. Note that the shift induced by the replacement of O–H by O–D is nearly as much as $1000\,cm^{-1}$.

Finally, an important chemical characteristic to be understood is hydrogen bonding. The infrared spectrum is a wonderful tool for helping to monitor the presence of this phenomenon. However, one must be aware of the fact that hydrogen bonding is usually concentration dependent. If a hydroxy compound is examined in a nonpolar solvent, one should be aware that the hydroxy group spectrum, in particular bands assigned to the stretching vibration, will change shape, position, and intensity as a function of concentration. Also, diagnostically, if the hydroxy group bands do not change significantly with concentration, then this is a strong indication of intramolecular (internal) hydrogen bonding. This latter information is important in helping to determine the spatial relationships of hydroxy groups within a molecule.

The last two sections might seem to be obvious basic knowledge in the chemistry and physical properties of a material. However, just simple observations of the type described can be invaluable in determining the presence of specific functional groups when used in combination with the infrared spectrum. Likewise, the use of negative spectral information, viz. the absence of a characteristic group frequency, is equally important. For example, if a sample has an almond-like odor, but does not have a strong band between 2200 and $2000\,cm^{-1}$, then the compound is more likely to be a nitro compound or benzaldehyde than a cyano compound. If an aldehydic carbonyl group is absent, then the compound probably contains a nitro group.

4.4 The Infrared Sampling Method

The final appearance of the infrared spectrum is very important in the interpretation process. Where possible, it always helps to be able to study the sample as it occurs naturally, without any form of physical modification. This eliminates the possibility of any interactions, or even chemical modifications. A common method for handling solid samples is known as the compressed alkali metal halide pellet method (KBr pellet or disk method). This method of sample preparation is prone to difficulties, and it often requires a skilled operator to produce good-quality spectra on a routine basis. A poor sampling technique with this method can lead to spectral artifacts and distortions, all of which can cause difficulties in the interpretation. Furthermore, some compounds may react with KBr, causing either a liberation of the halogen (oxidants) or halogen exchange (halide salts). The act of applying pressure can also change the appearance of the spectrum, especially when a material exists in more than one polymorphic form. The pressure may cause one polymorph to dominate, thereby changing the entire absorption band distribution of the spectrum. This is especially a problem with certain pharmaceutical products.

Other pretreatment methods, such as melts and cast films for solid samples, can cause similar problems. The cooling process (from melts) or the casting process can induce preferential crystallization of a particular form of the material, and sometimes preferred orientations can result. All of these effects can give rise to nonrepresentative spectra, which can cause interpretation problems.

With liquids, it is always beneficial to examine the sample as it is received. Liquid transmission cells have been used successfully in the past, but care and attention must be paid to the nature of the window material and to the cell pathlength. Low-molecular-weight materials usually require very short pathlengths in order to obtain a good-quality spectrum with most of the absorption bands on-scale. If too many bands bottom out, it is difficult to ascertain the relative intensity of certain key functional groups, and sometimes key factors, such as band splitting, may be missed. Also, if the sample is of low molecular weight, and volatile, it is important to ensure that the cell seals adequately.

In recent years, focus has been placed on reflectance methods of measurement, primarily from the surface of a sample. The most common methods are diffuse reflectance, specular reflectance and/or transflectance, and ATR. Generally, these are nondestructive methods, and require only a minimum amount of sample

preparation. The specular and transflectance methods have found greatest use with microscope-based methods of spectral measurement. Diffuse reflectance and ATR are important laboratory methods, and both lend themselves to rapid and reproducible sampling. However, both methods have an impact on the spectrum, and neither method necessarily provides a true spectrum of the sample free from some form of distortion or aberration. ATR is the simplest to handle, and its main limitation is that it provides a spectrum where the spectrum intensity across the spectrum increases as a function of wavelength. Both diffuse reflectance and specular reflectance can be the most difficult to handle, because both methods can cause severe band distortions, linked to surface and refractive index phenomena. In all three cases, the nature and the origin of the distortions are understood, and can be assigned to known physical properties. If necessary, software algorithms can be applied to correct for most of the distortions – a wavelength correction function in the case of ATR (for qualitative use), and a spectrum transformation, known as the Kramers–Kronig transform, in the case of surface reflectance distortions.

One other technique that is not so common but is important for certain ''difficult-to-handle'' samples is photoacoustic spectroscopy. The accessory, which also functions as the detector, provides a true absorption spectrum from a sample directly, with no sample modification. It is particularly good for samples that naturally have high absorption, such as carbon-filled materials, and for materials that are a good thermal conductor. As with ATR spectra, there is a wavelength dependence of the peak intensity across the spectrum.

The entire subject of sample handling is very broad, and generally falls outside the scope of this article. However, its impact on the final infrared spectrum and its resultant impact on the interpretation process can be very profound. Therefore, the reader is encouraged to read specific texts that address the subject in greater detail.[21–23]

5 AN OVERVIEW TO INFRARED SPECTRAL INTERPRETATION – SOME SIMPLE RULES AND GUIDELINES

This final section is a review of the ideas presented so far as to how to go about the interpretation of an infrared spectrum. With it, there is the implied understanding that the infrared spectrum is only one source of information, and additional information is most likely required for a satisfactory analysis of the spectrum. The following are a set of simple guidelines to help a person through the early stages of rationalizing an infrared spectrum and to get to the first stages of an interpretation. However, note that this is not a definitive set of rules.

First, it is important to appreciate that there are often recurring patterns in the spectrum that are characteristic of certain classes or compounds and/or functional groups. Such patterns can be far more diagnostic in the interpretation process than merely isolated group frequencies. Examples include the overall appearance of the hydrogen-bonded hydroxy functionality, the patterns associated with unsaturation and aromaticity, and the combinations of bands used to characterize different carbonyl compounds. To start, it is worthwhile to determine if the compound is organic (evidence of C–H functionality) and if there is a characteristic chain or ring associated with the compound. Having determined this, it is worthwhile to reflect on the complexity of the spectrum, and to see if one or two characteristic functionalities, associated with simple functional groups (–OH, –NH, C=O, etc.) are present. Often, the like-to-like rule applies – that is, a simple spectrum is usually indicative of a simple molecular structure, and vice versa.

If the spectrum clearly indicates multiple functionality, then be aware that mutual interaction may be occurring, and this may be causing some deviations from ideality, relative to the common group frequency assignments. Also, remember that multiple functionality might reflect a complex structure, but it might also reflect that the sample is a mixture. In both cases, an appreciation of the existence of spectral coincidences is very important. For example, a strong absorption between 1250 and 1150 cm^{-1} can be assigned to numerous types of chemical bonds, from the C–F group to the Si–O group. A careful assessment of the other spectral features, and most likely an understanding of the nature of the sample, will usually help in resolving some of these spectral coincidence situations. The use of known relative band intensities for certain functional groups can also be helpful in determining whether band overlap is occurring, and in particular whether the bands from a group are being obscured by other more intense bands. A good example is the overlap of the *cis* C–H out-of-plane bending vibration, which is typically overlapped by the $-(CH_2)_n-$ rocking vibration of long-chain aliphatic compounds.

Also, if multiple functionality exists, some form of separation scheme may be in order to help determine if the sample is a mixture and, if so, to help separate some of the components. While this may not always be necessary – there are many commercial mixtures identifiable from the original spectrum – it is often required for a more definitive evaluation of a sample from its infrared spectrum. An example of such a mixture is provided in Figure 11. The spectrum of this four-component mixture can be interpreted from a knowledge of the components – acetone, methanol, dichloromethane

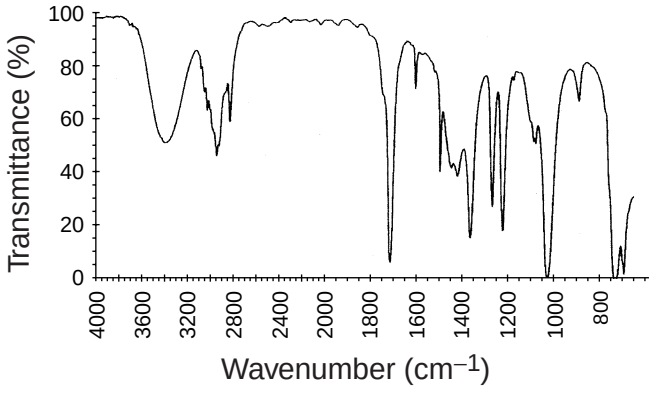

Figure 11 ATR spectrum of a commercial solvent mixture containing methanol, dichloromethane, toluene and acetone. Copyright Coates Consulting.

and toluene. However, the use of separation techniques, such as water-based solvent extraction, and partial distillation can help in determining the identity of the components.

Finally, in terms of using the first-assessment rules outlined in the remainder of this section, it is important to incorporate all known chemical and physical information, including pertinent test data, such as the presence of unsaturation, which may be used to corroborate the spectral data for a suspected double bond. Remember that simple information, such as color and odor, can be equally important – highly conjugated compounds and nitro compounds often give rise to significant color, and numerous functional groups, such as esters, ketones, unsaturation, and nitro groups, have highly characteristic fragrances or odors, even in the solid state.

The following first assessment guide is based on the examination of critical spectral regions for evidence of highly characteristic molecular backbones or fragments and functional groups. An important point to remember here is that negative information, in the form of no-band regions, is extremely important, and helps to eliminate major classes of chemical compounds, thereby reducing the number of possibilities in the final interpretation.

5.1 A Quick Diagnostic Assessment of an Infrared Spectrum

5.1.1 Step 1: Overall Spectrum Appearance

Is the spectrum simple? Are there only a few main characteristic absorption bands, say less than five?

If the spectrum is simple, as defined, then the compound may be a low-molecular-weight organic or inorganic compound, such as a simple salt of a common molecular ion (carbonate, sulfate, nitrate, ammonium, etc.) or a covalent species (chloroform, dichloromethane, methanol, water, etc.). In this case, the physical properties are important – if the material is a crystalline solid, then it is most likely a salt; if it is a mobile, possibly volatile, liquid, then the latter probably applies. Note that there are always exceptions, e.g. bromoform and iodoform, simple covalent compounds, are crystalline solids (both melt at low temperatures, whereas salts do not). One other class of compounds worth mentioning here is simple polymers, such as polyethylene and polytetrafluoroethylene, both of which exhibit very simple spectra when obtained from thin films.

Are some of the bands broad? Hydrogen bonding possibly exists and can be exhibited in hydrates, water solutions, alcohols, ammonium compounds, amino compounds, etc. Note that some inorganics may exhibit a combination of both broad and very narrow bands, even for certain hydrated species. These are often related to the crystalline structure of the compound and the symmetry of certain aspects of the molecular/ionic structure.

5.1.2 Diagnostic Step 2: Testing for Organics and Hydrocarbons – Absorptions in the Region 3200–2700 cm^{-1}

Well-defined and characteristic absorptions in this region are normally characteristic of carbon- and hydrogen-containing species, and are assigned to various forms of C–H stretching.

Are there absorptions above 3000 cm^{-1}? If yes, then the compound is likely to be unsaturated (contains C=C) or aromatic. If isolated absorptions occur at 3010 and/or 3040 cm^{-1}, then the absorbing species is mostly simple olefinic unsaturation.

Is the main absorption below 3000 cm^{-1}? If so, the compound is probably aliphatic. If the main absorptions are approximately 2935 and 2860 cm^{-1}, and there are also absorptions at 1470 and 720 cm^{-1}, then the compound probably contains a long linear aliphatic chain.

5.1.3 Diagnostic Step 3: Testing for Hydroxy or Amino Groups – Absorptions in the Region 3650–3250 cm^{-1}

Hydroxy or amino groups mainly dominate this region, both giving rise to very characteristic band profiles. The presence or absence of hydrogen bonding is well delineated in this region. One other characteristic vibration that is observed in this region, which is unlikely to be confused with any other functional group, is the C–H stretch of a terminal alkyne (acetylenic compound). This exhibits a relatively narrow absorption at

3300 cm^{-1}, and is associated with a second (C≡C) band at 2200 cm^{-1}.

Is the main absorption band in the area broad? If so, the compound is probably a hydroxy or amino group. Knowledge of the presence of nitrogen or a nitrogenous base (pH or titration) helps to support an amino compound.

Are there additional moderate to intense bands in the ranges 1600–1300, 1200–1000 and 800–600 cm^{-1}? If so, the compound is likely to be a simple hydroxy compound. Remember that the simple hydrogen-bonded OH absorption of a hydroxy (alcohol) function has a very characteristic shape.

If the feature is relatively sharp and occurs between 3670 and 3550 cm^{-1}, the compound probably contains a non-hydrogen-bonded hydroxy group, often an alcohol or phenol with a sterically hindered OH group. Note also that this spectral feature is also exhibited by certain inorganics and minerals, and is indicative of a ''free'' OH group, either on the surface, or embedded within a crystal lattice, and free from interactions with other ions or groups.

5.1.4 Diagnostic Step 4: Testing for Carbonyl Compounds – Absorptions in the Region 1850–1650 cm^{-1}

A major band in this region usually indicates the presence of a C=O group (carbonyl compound). Note that certain C=N (imino) and N=N (azo) groups may interfere, example compounds being purines.

Is the absorption at the high end of the range, e.g. 1775 cm^{-1} or above? If yes, the compound probably contains a reactive carbonyl, such as an anhydride, acid halide (acyl halide) or α-halogenated carbonyl, or a strained-ring carbonyl, such as a lactone, or an organic carbonate.

Is the absorption in the middle of the range, 1750–1700 cm^{-1}? If yes, the compound is probably a simple carbonyl compound, such as a ketone, an aldehyde, an ester, or a carboxylic acid.

Is the absorption at the low end of the range, e.g. below 1700 cm^{-1}? If yes, the compound is probably an amide or a carboxylate (carboxylic acid salt). Note that conjugation with another carbonyl group, or a double bond or aromatic ring, will lower the carbonyl absorption by 30–50 cm^{-1}. Therefore, conjugated aldehydes, ketones, esters, and carboxylic acids may fall into this lower end category. If conjugation with a double bond is present, then a second strong absorption should be observed nearby, between 1650 and 1600 cm^{-1}. In the case of the involvement of an aromatic ring, it is important to look for evidence of the characteristic aromatic absorptions (see Diagnostic Step 6).

5.1.5 Diagnostic Step 5: Testing for Unsaturation – Weak to Moderate Absorption in the Region 1670–1620 cm^{-1}

Note that this is not a particularly definitive diagnostic without the support of other corroborating absorptions.

A relatively narrow, weak-to-moderate absorption, normally centered around 1650 cm^{-1}, is indicative of olefinic unsaturation (C=C). A lowering of this frequency, accompanied by intensification of the band, is characteristic of conjugation with another double bond structure, such as C=C, C=O or an aromatic ring (see comments in Diagnostic Step 4).

The test for unsaturation normally involves a check for C–H peaks above 3000 cm^{-1}, and typically as a single, or possibly a pair, of absorptions (dependent on configuration of the double bond) at ~3085 cm^{-1} and/or ~3025 cm^{-1} – not to be confused with the multiple band situation for aromatics (see Diagnostic Step 6).

Are there single or multiple bands between 1000 and 880 cm^{-1} (C–H out-of-plane bending)? If so, this, in combination with the 1650 and 3010 and/or 3040 cm^{-1} bands, can be used as an assignment of unsaturation. This is either terminal (or vinyl $-CH=CH_2$), with a pair of bands, normally located about 990 cm^{-1} and close to 900 cm^{-1}, or trans unsaturation (CH=CH), with a band centered around 965–960 cm^{-1}. A single band around 890 cm^{-1}, in conjunction with the other unsaturation absorptions, can indicate a vinylidene (or pendant, $C=CH_2$) type of olefinic double bond.

Note that a *cis* configuration (CH=CH) is sometimes hard to characterize, because the normally characteristic C–H out-of-plane bending is broad, with only a moderate absorption, and occurs in a region (around 700 cm^{-1}) where it can overlap with absorptions from other structural components, such as long-chain aliphatics, aromatics, and halogen compounds.

5.1.6 Diagnostic Step 6: Testing for Aromatics – Well-defined Absorptions in the Region 1615–1495 cm^{-1}

One, but typically two, sets of bands in the region, with one set around 1600 cm^{-1} and the other around 1500 cm^{-1}, are consistent with an aromatic compound. Sometimes the absorptions occur as single bands, at other times one or more of the band sets may appear split, as an asymmetric doublet. This assignment is usually confirmed by the presence of other bands in the spectrum.

Are there a series of weak-to-moderate absorptions in the region 3150–3000 cm^{-1}, which can be assigned to aromatic C–H stretching? These usually support the presence of the 1600/1500 cm^{-1} aromatic ring bands (see Diagnostic Step 2).

For simple aromatic compounds, a series of weak absorptions may be observed between 2000 and 1700 cm^{-1}. These are multiple combination bands that are noticeably weak, and are typically not to be confused with an isolated weak carbonyl absorption.

The C–H bending vibrations can also be used to support the presence of an aromatic structure. If there are medium-to-strong absorptions, sometimes more than one, between 850 and 670 cm^{-1}, these can be assigned to C–H out-of-plane bending on an aromatic ring. These support the assignment of the 1600/1500 cm^{-1} bands, and the locations of the bands are often indicative of the nature of the substitution of the aromatic ring. For simple structures, it is possible to differentiate mono-, di- (ortho, meta and para), and certain poly-substitutions. Note that a single, strong band may support simple ortho (around 750 cm^{-1}) or para (around 830 cm^{-1}) substitution.

Many organic compounds have multiple band structures in the region 1150–950 cm^{-1}. The in-plane C–H bending vibrations of aromatic compounds typically occur in this region, and can exist as complex band structures (multiple, well-defined bands). They tend not to be diagnostic for many compounds because of conflict and overlap with other functional group absorptions, including some skeletal (backbone) vibrations. However, they are clearly observable in simple hydrocarbons as weak-to-moderate absorptions, and in halogen-substituted aromatic compounds (such as chlorobenzenes) as characteristic strong absorptions.

5.1.7 Diagnostic Step 7: Testing for Multiple Bonding (Often with a Bond Order of 2 or Higher) – Absorption in the Region 2300–1990 cm^{-1}

Multiply bonded nitrogen compounds and cumulated multiple bond compounds, such as cyanides (nitriles), cyanates, isocyanates, thiocyanates, and diazo compounds, exhibit characteristic absorptions in the region 2300–1990 cm^{-1}. Cyanide absorptions vary from weak to moderate to strong, depending on the nature of other substituents in the molecule. Conjugation, including substitution on an aromatic ring, modifies the intensity of this absorption.

If the absorption band is extremely strong, the group is more likely to be a three-center functionality, such as a cyanate, an isocyanate, or a thiocyanate.

Acetylenes and alkyne substituents also exhibit a weak-to-moderate absorption within the same spectral region. If the acetylenic function is internal (medial), then the absorption can be extremely weak. However, a terminal alkyne group (C≡C) band may have moderate intensity, and this is supported by a characteristic C–H stretching absorption around 3310 cm^{-1}. A relatively low-frequency C–H bending vibration around 630 cm^{-1} also confirms the terminal alkyne vibration.

Metal carbonyls have unique bonding, where the bond order between the carbon and oxygen is between two and three, and where the additional electrons are provided by dπ back-bonding from the accompanying transition metal atom. The multiply bonded CO group provides an extremely intense absorption band close to 2000 cm^{-1} (typically between 2100 and 1900 cm^{-1}), the actual position of the band(s) and the complexity of the bands being dependent on the structure of the compound.

Note that an isolated weak-to-moderate intensity absorption above the specified region, and typically between 2700 and 2400 cm^{-1}, is not normally associated with a multiply bonded compound. These absorptions are usually from hydride vibrations, such as from silanes (Si–H), thiols and sulfides (S–H), phosphines (P–H), arsines (As–H), boranes (B–H), etc. Knowledge of the existence of the accompanying heteroatom helps in the assignment of the X–H bonds in these compounds.

This concludes the summary data for a first-pass assessment of an infrared spectrum, based on the most common and characteristic group frequencies. The technique is not foolproof, but for many simple compounds, if one systematically applies the diagnostic tests on the spectrum from a sample, as presented above, it is possible to gain some understanding of the chemical functionality. It is always hoped that more information can be gained from the sample, either by first-hand knowledge or by asking questions, or by performing additional tests. Once a basic interpretation is accomplished, and the sample is broadly characterized, it is recommended that reference spectra are used to try to obtain a more exact match to the sample. Always bear in mind that the sample could be a mixture, such as a commercially formulated/compounded material. In such cases, an exact match may not be possible, unless the reference spectrum exists for the formulation.

ABBREVIATIONS AND ACRONYMS

ATR Attenuated Total Reflectance

RELATED ARTICLES

Infrared Spectroscopy **(Volume 12)**
Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. H.A. Szymanski, R.E. Erickson, *Infrared Band Handbook*, Plenum Press, New York, 2 Volumes, 1970.
2. P. Gans, *Vibrating Molecules: an Introduction to the Interpretation of Infrared and Raman Spectra*, Chapman & Hall, London, 1975.
3. N.B. Colthrup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, San Diego, CA, 1–73, 1990.
4. J.M. Hollas, *Modern Spectroscopy*, 3rd edition, John Wiley & Sons, New York, 1996.
5. J.P. Coates, 'The Interpretation of Infrared Spectra: Published Reference Sources', *Appl. Spectrosc. Rev.*, **31**(1–2), 179–192 (1996).
6. L.J. Bellamy, *Infrared Spectra of Complex Molecules*, Chapman & Hall, New York, Vol. 1, 1975.
7. L.J. Bellamy, *Advances in Infrared Group Frequencies, Infrared Spectra of Complex Molecules*, Chapman & Hall, New York, Vol. 2, 1980.
8. D. Steele, *Interpretation of Vibrational Spectra*, Chapman & Hall, London, 1971.
9. N.B. Colthrup, L.H. Daly, S.E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, San Diego, CA, 1990.
10. G. Socrates, *Infrared Characteristic Group Frequencies*, John Wiley & Sons, New York, 1994.
11. D. Lin-Vien, N.B. Colthup, W.G. Fateley, J.G. Grasselli, *Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press, San Diego, CA, 1991.
12. B. Smith, *Infrared Spectral Interpretation, a Systematic Approach*, CRC Press, Boca Raton, FL, 1999.
13. K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds, Parts A and B*, John Wiley & Sons, New York, 1997.
14. R.A. Nyquist, C.L. Putzig, M.A. Leugers, *Handbook of Infrared and Raman Spectra of Inorganic Compounds and Organic Salts*, Academic Press, San Diego, CA, 1997.
15. J. Haslam, H.A. Willis, D.C. Squirrel, *Identification and Analysis of Plastics*, John Wiley & Sons, New York, 1979; originally published by Iliffe, London, 1972.
16. D.O. Hummel, F.K. Scholl, *Atlas of Polymer and Plastics Analysis*, Verlag Chemie, Weinheim, Vol. 1–3, 1980–81.
17. J.L. Koenig, *Spectroscopy of Polymers*, American Chemical Society, Washington, DC, 1992.
18. J.P. Coates, 'Questions and Answers: What to Do When the Sample Arrives, Part II', *Spectroscopy*, **11**(5), 24–32 (1996).
19. J.P. Coates, 'Into the Unknown: the First in the Series', *Spectroscopy*, **11**(5), 32–39 (1996).
20. J.P. Coates, 'Once More into the Unknown', *Spectroscopy*, **11**(6), 14–23 (1996).
21. J.P. Coates, 'A Review of Sampling Methods for Infrared Spectroscopy', in *Applied Spectroscopy: a Compact Reference for Practitioners*, eds. J. Workman, A.W. Sprinsteen, Academic Press, New York, 49–91, 1998.
22. P.B. Coleman (ed.), *Practical Sampling Techniques for Infrared Analysis*, CRC Press, Boca Raton, FL, 1993.
23. G. Dent, J.M. Chalmers, *Industrial Analysis with Vibrational Spectroscopy*, Royal Society of Chemistry, Cambridge, 1997.

Liquid Chromatography/ Infrared Spectroscopy

Govert W. Somsen
University of Groningen, Groningen, The Netherlands

Tom Visser
National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands

1 Introduction **1**
2 Coupling Modes and Operating Principles **2**
2.1 Flow-cell Approach 2
2.2 Solvent-elimination Approach 3
3 Flow-cell Techniques **4**
3.1 Cell Types and Infrared Detection Modes 4
3.2 Use in Liquid Chromatography 6
3.3 Use in Flow-injection Analysis 7
4 Solvent-elimination Techniques **8**
4.1 Deposition Substrates and Infrared Detection Modes 8
4.2 Early Interfaces 9
4.3 Spray-type Interfaces 10
5 State of the Art **16**
6 Perspective and Future Developments **17**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Analytical techniques that combine liquid chromatography (LC) and infrared (IR) spectroscopy have been developed primarily to permit specific detection and/or identification of sample constituents. LC is an important and extensively used method for the separation of mixtures into their individual components. IR spectroscopy is very useful for the characterization of functional groups and has strong compound-identification capabilities which are especially suited for the differentiation of structural isomers. Over the past years the coupling of LC and IR spectroscopy (LC/IR) has been accomplished by two different approaches. The first and simplest approach is to use a flow cell through which the effluent from the LC column is passed while the IR spectra are continuously measured. The merits of flow-cell IR detection include ease of operation, real-time detection and low maintenance, but its main disadvantage lies in the significant IR-absorption of the solvents commonly used in LC. These absorptions seriously limit both the detection sensitivity and the obtainable spectral information. The second approach involves elimination of the LC solvent prior to IR detection. In this approach an interface is used to evaporate the eluent and deposit the separated compounds onto a substrate suitable for IR detection. The primary advantages of solvent-elimination LC/IR are the possibility to obtain full spectra of the analytes and the considerably enhanced sensitivity when compared to flow-cell detection. Unfortunately, common LC solvents, and particularly aqueous eluents, are not easily removed and therefore the evaporation interfaces are often rather complex. This article reviews the developments, practical aspects, applications and current status of LC/IR, covering both coupling approaches. It follows that despite the unfavorable detection limits, flow-cell LC/IR can be useful for the specific and quantitative detection of major components of mixtures. However, solvent-elimination-based IR-detection should be used when small amounts of sample constituents have to be characterized with a high level of confidence. In general, the practical use of IR detection in LC is still limited, but the advent of various (commercial) flow-cell and interface designs shows that LC/IR is more and more being recognized as a feasible and rewarding technique.

1 INTRODUCTION

IR spectroscopy has a high potential for the elucidation of molecular structures. The IR spectrum of a polyatomic molecule is based on molecular vibrations, each specifically dependent on atomic masses, bond strengths and intra- and intermolecular interactions. As a consequence, the entire IR spectrum of an organic compound provides a unique fingerprint, which can be readily distinguished from the IR-absorption patterns of other compounds including isomers. In other words, when reference spectra are available, most compounds can be unambiguously identified on the basis of their IR spectra. Moreover, characteristic absorption bands can be used for compound-specific detection.

LC is a powerful and versatile separation technique, which can handle a wide range of sample types and compound classes. Because of the widespread use of LC and the (growing) need for analytical procedures that provide confirmation and/or identification of sample constituents, much effort has been – and still is – devoted to the coupling of LC with spectrometric techniques such as mass spectrometry (MS), IR and nuclear magnetic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

resonance (NMR) spectroscopy. Today, with modern Fourier transform infrared (FTIR) instrumentation routinely available, spectra can be recorded from nanogram, or even picogram, amounts of pure substance so that IR detection, in principle, is suited for molecular recognition at analyte levels frequently met in LC. Unfortunately, because of the (spectral) characteristics of the mobile phase, the coupling of LC and IR spectroscopy (LC/IR) is not straightforward and often requires the construction of special flow cells or the development of rather complex interfaces. Therefore, compared with other LC detection modes such as ultraviolet/visible (UV/VIS) absorption spectroscopy or MS, the use of IR detection in LC is still rather limited. Nevertheless, progress in interfacing techniques during the last decade has brought LC/IR to a stage of analytical utility which suggests that LC/IR may well become a commonly available and applied technique.

2 COUPLING MODES AND OPERATING PRINCIPLES

In the first LC/IR systems[1,2] flow cells were used in a fashion analogous to LC with on-line UV/VIS absorption detection. In order to circumvent interfacing difficulties related to the IR absorptions of the mobile phase, in 1979 Kuehl and Griffiths[3] developed the first solvent-elimination based LC/IR set-up in which the eluent is evaporated prior to IR detection. Since then two approaches can be discerned in LC/IR, namely, the flow-cell approach and the solvent-elimination approach. In the contemporary practice of LC/IR both approaches are applied, although the detection limits and spectral information obtained with either approach may differ considerably. The principles, applications, merits and limitations of flow-cell and solvent-elimination LC/IR have been reviewed in a number of books and papers.[4–8]

2.1 Flow-cell Approach

The simplest way to couple LC and IR is to let the column effluent pass directly through a flow cell suited for IR measurements. The IR absorption of the LC effluent is continuously monitored and spectral data are collected on-the-fly and stored throughout the chromatographic run. During or after the run, the spectra and/or IR chromatograms are computed and absorption due to the mobile phase is subtracted. In a flow-cell design, band broadening caused by detection is easily minimized.

Unfortunately, the absolute sensitivity of IR spectroscopy is relatively poor compared to spectrometric techniques like MS, UV/VIS and fluorescence spectroscopy. Moreover, solvents suited for LC generally have many absorption bands in the IR region, which leads to serious limitations of flow cell LC/IR interfacing. Firstly, absorption bands of analytes may be obscured by eluent absorptions. In other words, in flow-cell LC/IR, the spectral information that can be obtained is limited and depends on the window provided by the mobile phase used. Moreover, ill-considered subtraction of solvent absorption bands may lead to the erroneous conclusion that there is no absorption of the analyte in the corresponding spectral regions. Secondly, gradient elution can hardly be applied as the changing composition of the mobile phase frustrates proper subtraction of the background. To some extent, the second derivative of a spectrum can be used to compensate for this problem, but accurate spectral correction is virtually impossible. Thirdly, the pathlength of the flow cell has to be limited to ensure a certain spectral window and sufficient energy reaching the IR detector. For organic solvents commonly used in LC/IR, the effective pathlength rarely exceeds 1 mm which, bearing in mind Beer's law, seriously reduces the analyte detectability. For aqueous eluents, the largest tolerable pathlength is even shorter. As can be seen from the spectrum of water in Figure 1, most of the mid-IR region between 4000 and 1000 cm^{-1} is opaque, even at a very short pathlength of 10 μm. It implies that a practical combination of reversed-phase liquid chromatography (RPLC) and IR via a transmission flow cell is restricted to specific applications. Finally, the use of signal averaging to improve the signal-to-noise ratio (S/N) of the spectra is limited owing to the short time that is available for analysis under dynamic conditions. Occasionally, the principle of stop-flow can be used to enhance the S/N without too much loss in chromatographic performance.

The choice for a certain mobile phase and flow-cell properties is primarily determined by chromatographic considerations and the desired spectral information. In principle, the spectral window of the solvent can be very small for quantitative analysis as the measurement of a single wavenumber, e.g. the band maximum, is

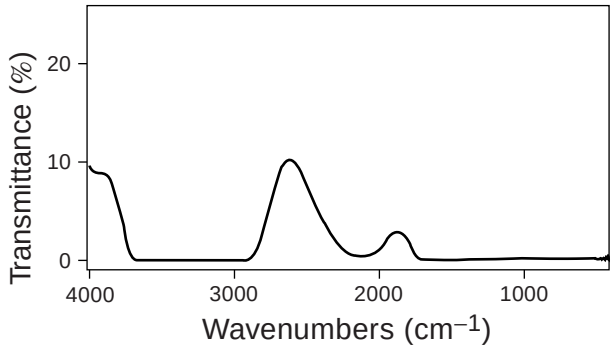

Figure 1 IR transmission spectrum of water recorded using a flow cell with silver bromide windows and an optical pathlength of 10 μm.

sufficient. Contrary, qualitative information desires IR transparency over a much wider spectral region in order to determine the presence or absence of functional groups, or to identify a compound by its IR fingerprint region (1300–600 cm^{-1}). Obviously, the spectral window of the mobile phase is inversely proportional to the cell thickness and thus, an excess of sensitivity can be traded for more spectral information and vice versa. The experimental parameters in LC/IR are, therefore, a compromise between chromatographic and spectroscopic considerations giving maximum cell thickness and solvent transparency.[5–8]

In order to minimize the problems associated with eluent absorption, the choice of the solvent in flow-cell LC/IR is generally restricted to chlorinated alkanes and deuterated solvents. These solvents leave relatively wide windows in the spectrum, although even these eluents inevitably obscure parts of the spectral functional group region (4000–1300 cm^{-1}) and the fingerprint region. The use of a small percentage of a more polar modifier in the eluent, as is quite common in normal-phase liquid chromatography (NPLC), may already prohibit effective detection. Owing to the short optical pathlength, the absolute limit of detection in on-line LC/IR often is in the order of a few micrograms, which implies that analyte concentrations of 1–10 g L^{-1} have to be injected to obtain identifiable spectra.

Tran et al.[9] proposed a completely different flow-cell LC/IR technique with, potentially, higher sensitivity. Different from the conventional way of measuring the IR absorption over a certain wavenumber region, this technique, called IR thermal lens spectrometry, is based on the measurement of the temperature rise that is produced in an illuminated sample by non-radiative relaxation of the energy absorbed by molecules upon IR laser excitation. Detection is carried out either directly, i.e. at a specific wavelength where only the analyte of interest absorbs, or indirectly, i.e. at a wavelength where only the mobile phase absorbs. Obviously, no spectral information is obtained with this quantitative technique, and the wavenumber region that can be covered with IR-tunable lasers is limited.

In conclusion, flow-cell IR detection is not a widely applied technique in LC as result of the interference of absorption bands of the mobile phase and the limited sensitivity of IR spectroscopy. Applications largely focus on dynamic systems where low detection limits are not crucial and solvents appropriate for IR detection can be selected more or less freely, that is, without having unacceptable detrimental effects on the analytical performance. Examples of such techniques are gel permeation chromatography (GPC), also referred to as size-exclusion chromatography (SEC), and flow-injection analysis (FIA). An important feature of the flow-cell LC/IR technique is the ability to monitor the elution of compounds with specific structural characteristics, i.e. functional groups. Besides, apart from regions of solvent opacity, spectra can be obtained instantaneously from any point in a chromatographic peak.

2.2 Solvent-elimination Approach

The compatibility and time-domain difficulties connected to flow-cell IR detection can be circumvented by coupling LC and IR spectrometry via a substrate suitable for IR detection. In this indirect approach, the eluent is eliminated and the chromatographically separated compounds are immobilized on the substrate prior to the collection of IR spectral data. The immobilization of the chromatogram is accomplished by using an interface which evaporates the eluent and continuously deposits the column effluent onto the moving substrate. In this way, interference-free IR spectra of the deposited compounds can be recorded independently from the LC conditions and the sensitivity of the FTIR spectrometer can be fully exploited. A schematic of a solvent-elimination LC/IR set-up is depicted in Figure 2.

Besides the possibility to acquire full spectra, there are some additional advantages of the solvent-elimination approach with respect to flow-cell IR detection. By careful control of the interface performance and the speed of the substrate, concentrated deposits may be obtained which will enhance analyte detectability. Spectral analysis can be performed without any time constraints since the chromatogram is stored on the substrate. This implies that signal averaging can be used and analyte spots can be analyzed repeatedly. For instance, after rapid screening of the deposited trace applying low spectral resolution and only one scan per spectrum, high-resolution spectra with a high S/N can be recorded for a few interesting parts of the chromatogram. Moreover, spectrometric analysis of the stored chromatogram in principle can be carried out at any convenient time or place, which may be helpful when suitable IR facilities are limited and spectrometers have to be shared.

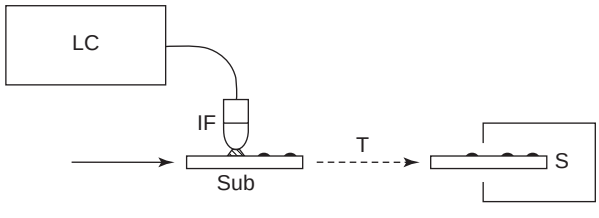

Figure 2 Schematic representation of a set-up for solvent-elimination LC/IR; LC, liquid chromatograph; IF, interface; Sub, deposition substrate; T, transfer; S, spectrometer.

Evidently, the solvent-elimination approach to LC/IR is (technically) more complicated than on-line IR detection. It requires an interface which should adequately effect the evaporation of the eluent and, at the same time, maintain the chromatographic resolution during the deposition process. In this respect, the LC flow rate, the composition of the eluent, the nature of the analytes and the substrate material are important factors. For example, small volumetric flows of a volatile solvent may be readily evaporated, while rapid elimination of aqueous eluents will require an interface with enhanced solvent-elimination power. Elimination of the eluent may also be hampered by the presence of nonvolatile additives such as buffer salts. Next to eluent evaporation, ideally the interface also should provide compound deposition into narrow spots in order to minimize band broadening and achieve optimum IR sensitivity. However, complete eluent evaporation and compact analyte deposition may well be irreconcilable goals. Therefore, in solvent-elimination LC/IR reduced flow rates, nonbuffered eluents and/or eluents with a low (or even zero) water percentage are frequently used.

The compounds analyzed by solvent-elimination LC/IR, of course, should be considerably less volatile than the eluent to accomplish their deposition. Since LC is used for nonvolatiles in particular, this condition is generally met. The quality of the used substrate should not be affected by either the eluent or the deposited compounds. The substrate also must be compatible with the selected IR mode without introducing additional interferences. Furthermore, the physico-chemical properties of the substrate may influence the efficiency of analyte immobilization. For instance, residual eluent easily spreads over a substrate with a hard and smooth surface, while it may be effectively sorbed by a powder.

During the last decades, the combination of LC and IR via solvent elimination has been pursued by several research groups, which designed quite a number of different interface concepts with varying success. The common goal of all these LC/IR systems is to sensitively acquire IR spectra of mixture constituents that are free from spectral interferences and can be used for identification purposes. Through the years, the IR detection limits obtained with solvent-elimination LC/IR have improved gradually from the microgram to the low- or sub-nanogram range. Despite the progress, today there is no single ''perfect'' solvent-elimination interface available: every described system has its specific limitations with respect to, for example, flow rate, composition of the eluent and/or achievable sensitivity. The research activities in the field, which until now have been dominated by the problem of simultaneous eluent-evaporation and compound-deposition, are still on-going. Nevertheless, during the last years, several new commercial interfaces based on the solvent-elimination approach have been introduced.[10,11]

3 FLOW-CELL TECHNIQUES

3.1 Cell Types and Infrared Detection Modes

A variety of flow cells, differing in optical material, pathlength and cell volume, is available for LC/IR detection purposes. The cells, including corresponding beam condensing optics, generally fit in the standard optical bench of the IR spectrometer. The distance between the flow cell, i.e. the IR spectrometer, and the LC system is kept as short as possible in order to minimize deterioration of the chromatography. As noted in section 2.1, the selection of a certain cell type and, particularly, the optical pathlength, depends on the required information.

The choice for a specific window material is mainly determined by the properties of the LC eluent and the spectral region that has to be monitored. A fully IR-transparent material such as potassium bromide (KBr), for instance, cannot be used with RPLC. Instead, more expensive water-insoluble materials such as calcium fluoride and zinc selenide (ZnSe) have to be chosen. The optical and physical properties of some commonly used IR-window materials are presented in Table 1. In all cases, in order to obtain an identifiable IR spectrum a minimum amount of analyte has to be present in the

Table 1 Optical materials for use in IR flow cells

Material	Spectral window (cm^{-1})	Solubility in water	Sensitive to
Silver chloride	4000–350	slightly soluble	complexing agents
Calcium fluoride	4000–1100	insoluble	ammonium salts; acids
Quartz	4000–2400	insoluble	hydrofluoric acid; hot sulfuric acid
Potassium bromide	4000–400	very soluble	water; methanol
Zinc selenide	4000–450	insoluble	acids; strong alkalis
KRS-5	4000–250	slightly soluble	complexing agents
Polyethylene	630–30	insoluble	organic solvents

detection cell during the time of measurement. According to Beer's law, the minimum identifiable concentration decreases when the pathlength of the cell is increased. However, extending the pathlength also results in an increase of the eluent absorbance, thus limiting the spectral window. Commonly, the volume of the flow cell has to be minimized to such an extent that only a small part (1% or less) of an LC-peak volume will fill the cell. Obviously, this places a significant limitation on the obtainable sensitivity. Through the use of microbore LC columns, the volume of the flow cells and LC peaks can be made more compatible and, for that reason, micro-LC/IR is often preferred. Besides a reduced solvent consumption, microbore LC may also offer higher peak concentrations in the cell. At the same time, however, it should be noted that the sample capacity (both in mass and volume) of micro-LC columns is rather limited.[12]

Two types of commercially available flow cells can be distinguished for LC/IR: transmission cells and attenuated total reflection (ATR) cells. The principle of a transmission flow cell is depicted in Figure 3. The basic part of the cell consists of an IR-transparent cavity or two IR-transparent windows separated by a metal spacer. The optical set is mounted into a metal body between flexible rings to prevent breakage. The LC effluent enters and exits the cell via capillary tubes connected to an assembly of universal fittings and the IR beam passes perpendicularly through the LC flow. Additional equipment can be purchased for operation at elevated temperature and pressure. Zero-dead-volume (ZDV) flow cells for IR detection have been developed to minimize the volumes needed to connect the cell to the column, which is very important when microbore LC is used. In this type of cell the LC effluent is directly led into the cavity of the optical element (Figure 4). Flow cells, micro-flow cells and ZDV-cells can be purchased in many variations, rigid and demountable, differing in window material, pathlength (0.001–0.5 mm) and internal volume (0.1–10 μL). Micro-cells are generally used in combination with a beam condenser to obtain a sufficiently high energy throughput and thus enhance the S/N of the recorded signal and/or spectra.

The second type of flow cell is based on the phenomenon of ATR and is called cylindrical internal reflectance cell or CIRCLE® cell.[13] The principle of the CIRCLE® cell is depicted schematically in Figure 5. The cell consists of a cylindrically shaped IR-transparent rod crystal with cone shaped ends that is incorporated in a boat-type cell body made of stainless steel (SS) or glass. The effluent of the LC column flows around the optical crystal while the interrogating IR beam enters the crystal at one end, reflects off the internal surfaces of the crystal and then exits at the other end. The cell body fits into a small optical bench with special input and output optics. Several crystal materials can be used but ZnSe is commonly preferred because of its high IR transparency, high

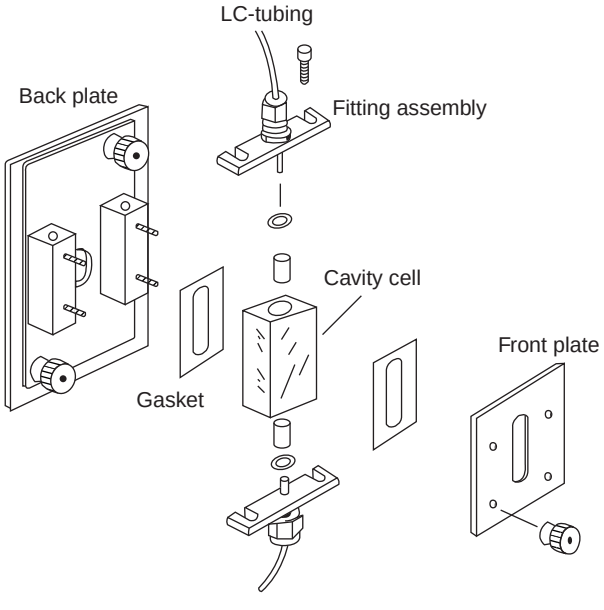

Figure 3 Cavity flow cell for IR transmission detection.

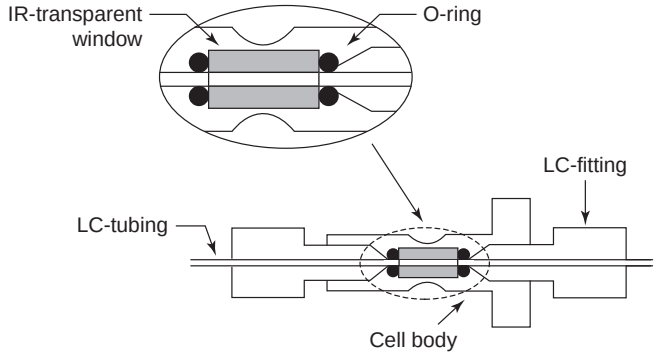

Figure 4 Schematic representation of a ZDV flow cell for IR transmission detection.

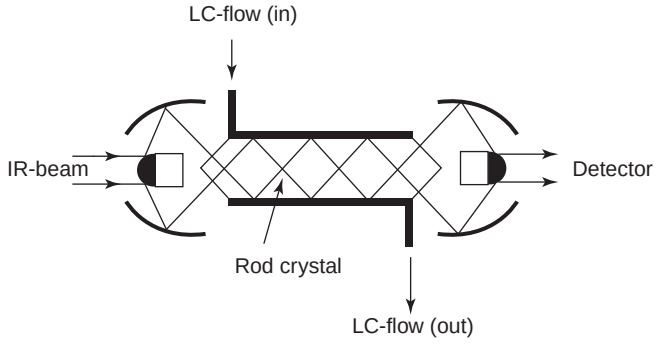

Figure 5 Schematic representation of an IR flow cell for cylindrical internal reflectance detection (CIRCLE® cell).

refractive properties and insolubility in water. CIRCLE® cells can be equipped with heating or cooling jackets too. The cell design tends to involve a relatively large sample volume and efforts have been made to reduce the internal volume of CIRCLE® cells to 1–25 μL in order to allow their efficient application in LC. The effective pathlength of a CIRCLE® cell is defined by the number of reflections in the optical element and, therefore, sensitivity can be enhanced by using longer optical elements. These, however, also imply an increased cell volume and, thus, extra broadening of the LC peaks. CIRCLE® cells cannot be used for quantification in a straightforward manner since the penetration depth of the IR radiation in the LC eluent is limited (typical 1–5 μm) and wavelength dependent. Special algorithms are used to compensate for this.

3.2 Use in Liquid Chromatography

As outlined in the previous sections, flow-cell IR detection is not the principal choice in LC. Yet, it is a valuable alternative and additional method to obtain specific quantitative and structural information on analytes. In view of the relatively poor sensitivity, IR detection is restricted to applications in which low detection limits are not required. In GPC, for example, column capacities and sample concentrations are usually high. Gel permeation chromatography/infrared (GPC/IR) is well suited for the characterization and quantification of compositional changes throughout a (bio)polymer mass distribution. Besides, the structural differences between the polymer components (usually homologues) often are small which implies that one or two spectral windows will suffice for specific detection.

Applications of on-line LC/IR, including automatic subtraction of the solvent background, had already been described in the mid-1970s.[1,2] Injected amounts at submicrogram level were found to be feasible for detection in both NPLC and RPLC separations. In subsequent studies, Taylor et al.[13,14] demonstrated that microbore LC (column i.d., less than 1 mm) offers improved sensitivity compared to conventional LC (column i.d., 4.6 mm) because of the higher sample concentration in the detector cell. Furthermore, use of halogenated hydrocarbons as eluent was shown to offer better spectral specificity owing to the higher IR-transparency of these solvents. A reasonable number of applications of flow cell LC/IR in a variety of disciplines has been developed since. Saunders and Taylor,[15] for example, applied on-line GPC/IR with tetrahydrofuran (THF) as mobile phase for the determination of the nitrogen content of cellulose nitrates. A cylindrical micro flow cell with a pathlength of 1 mm and an internal volume of 4 μL was employed in combination with a beam condenser. The extinction of the O–N–O asymmetric stretching band was used for the quantification of primary and secondary carbon nitration. An improved resolution and quantification was achieved by spectral derivatization. On-line GPC/IR is also a viable analysis method in the polymer field. One method employs high-temperature GPC with flow-cell IR detection to characterize the molecular weight distribution of high-density polyethylenes.[16] Tri- and dichlorobenzene were used as mobile phase as these solvents do not exhibit interfering absorption in the CH_2-stretching band region (3000–2700 cm^{-1}). Furthermore, a quartz flow cell with a relative long pathlength of 1 cm could be used in order to enhance the sensitivity. The method compares favorably with gradient elution fractionation combined with ^{13}C-NMR spectroscopic characterization.

The strong IR absorption of water, methanol and acetonitrile, limits the application of on-line reversed-phase liquid chromatography/infrared (RPLC/IR) to the analysis of samples with relatively high analyte concentrations, such as wines, beverages and cellulose solutions. Various examples of this type of analysis have been described in the literature. Recent applications are the identification and quantification of sucrose, glucose and fructose in soft drinks,[17] and the specific analysis of monosaccharides, alcohols and organic acids in wine.[18] In both studies, separations were achieved with ion-exchange columns using an aqueous eluent. IR detection was performed with a 25-μm pathlength flow cell and minimal identifiable concentrations of typically 1 g L^{-1} were obtained.

When a high sensitivity or selectivity is required, conventional RPLC solvents cannot be used with IR flow cells. Therefore, alternative methods have been developed to circumvent the IR-opacity problems imposed by the aqueous mobile phase, while maintaining the advantages of on-line detection. One approach is the extraction of the analytes into a more IR-transparent solvent. Another approach is the use of deuterated solvents which virtually have the same chromatographic properties but do not absorb in the spectral region of interest. An example of the first method is the dynamic extraction of the analytes from the aqueous effluent into chloroform or carbon-tetrachloride.[19] The on-line liquid–liquid extraction (LLE) is carried out in an extraction coil followed by continuous separation of the aqueous and organic phases by a hydrophobic membrane. Subsequently, the organic phase, carrying the extracted analytes, is monitored in a common IR flow cell. A basically different post-column extraction method has been developed by Messerschmidt.[20] In this method, the RPLC effluent is diluted with water, and the analytes of interest are trapped on several small solid-phase extraction (SPE)

columns. These columns are dried with a flush of nitrogen, and sequentially eluted with a small volume of tetrachloromethane into an IR flow cell allowing the individual spectra of the analytes to be recorded. An additional advantage of this method is the improved minimal identifiable concentration as a result of analyte concentration on the SPE column. This technique has been successfully applied by DiNunzio[21] for the separation and identification of active compounds and degradation products in pharmaceutical samples.

In on-line LC/IR, deuterated solvents can be attractive substitutes for conventional (hydrogenated) solvents. The absolute absorbance of deuterated solvents is usually smaller and, more importantly, their IR absorption bands are shifted to different spectral regions. Additionally, deuterated solvents and their hydrogenated counterparts show very similar elution properties. With respect to on-line extraction procedures, the use of deuterated eluents has advantages in terms of simplicity, speed and maintenance of the chromatographic resolution. A major drawback is their high price. A detailed study on the utility of deuterated eluents in micro-column RPLC, NPLC and GPC with flow-cell IR detection was carried out by Fujimoto et al.[22] It follows that the problem of interfering solvent absorption bands can be (partially) solved by a deliberate choice of a deuterated eluent. Chen and Kou,[23] for instance, used deuterated methanol and water instead of the nondeuterated solvents to overcome the strong interfering absorptions that hinder the effective detection and quantification of lipid fractions by on-line LC/IR. Remsen and Freeman analyzed proteins using on-line gel permeation chromatography/Fourier transform infrared (GPC/FTIR) with a deuterium oxide eluent to remove water and other detection-interfering low-molecular-weight compounds and to achieve a rapid hydrogen exchange.[24] Amounts of 50 μg per protein could be detected in the amide-I region between 1600 and 1300 cm^{-1}, and important information on the protein conformation could be obtained.

3.3 Use in Flow-injection Analysis

FIA is a well-established analytical technique, based on the automated injection of a series of samples in a continuous carrier stream. In FIA various detectors are used, mainly for quantitative purposes. Amongst these detection modes, flow-cell IR is not a frequent choice because of the interfering solvent absorptions and the relative poor sensitivity compared to UV/VIS and fluorescence detection.[25] In several cases, however, on-line flow-injection analysis/infrared (FIA/IR) is an appropriate alternative for the rapid determination and quantification of a specific analyte in a simple mixture. In FIA, there is no need for an eluting solvent and, therefore, in contrast with LC, a carrier solvent suitable for IR detection can be selected quite easily. In this respect, FIA/IR has a wider application range than LC/IR. Still, the flow cells used in FIA/IR are not different from the ones used in LC/IR and the resemblance in detection approach is evident. Commonly, FIA/IR is applied for single-component analysis but since IR spectra comprise a range of absorption frequencies, multi-component analysis can in principle be carried out as well. Obviously, the IR absorption bands of the carrier should not spectrally interfere with the marker band of each analyte.

Guzman et al.[26] compared the characteristics of transmission and CIRCLE® cell detection in FIA/IR for multicomponent analysis of ternary solvent mixtures. With tetrachloromethane as carrier, the transmission cell provided better sensitivity in the continuous-flow mode. The shorter effective pathlength of the CIRCLE® cell was compensated for by applying the stopped-flow technique which allowed data averaging and, thus, enhancement of the S/N of the ATR spectra. Analyte concentrations of 1–2% (v/v) could still be measured.

The principles and applications of flow-cell FIA/IR have been studied extensively by the group of de la Guardia.[27–33] The effects of flow rate, injection volume, cell volume and optical pathlength on the performance of the FIA/IR analysis of *o*-xylene in solutions of hexane, were investigated.[29] Further optimization was achieved in the development of a method for the automated determination of benzene in gasoline.[30] In a further study, incorporation of an analyte-enrichment step was proposed to improve the detection limits of on-line FIA/IR.[31,32] Preconcentration was carried out by trapping the analytes on an SPE cartridge. After drying of the cartridge with nitrogen gas, the analytes were eluted with dichloromethane and detected by IR transmission spectroscopy using a micro flow cell. For aqueous solutions of the pesticide carbaryl and its major metabolite 1-naphthol, detection limits of 0.36 mg L^{-1} and 1.6 mg L^{-1}, respectively, were obtained after preconcentration of 100-mL samples. Quantification was achieved by specific detection of the bands at 1744 cm^{-1} (carbaryl C=O) and 1276 cm^{-1} (1-naphthol C–O) (Figure 6). FIA/IR is also a useful alternative for the quantitative analysis of mineral oil and grease.[33] Analogous to the International Organization for Standardization (ISO) procedure for the quantitative analysis of mineral oil in water and soil, this method is based on measurement of the aromatic, olefinic and aliphatic C–H stretching bands in the 3200–2700 cm^{-1} region. The samples were extracted with tetrachloromethane by LLE or microwave-assisted extraction. Next, the extracts were analyzed by FIA/IR at a sampling frequency of 60 h^{-1} using a quartz flow

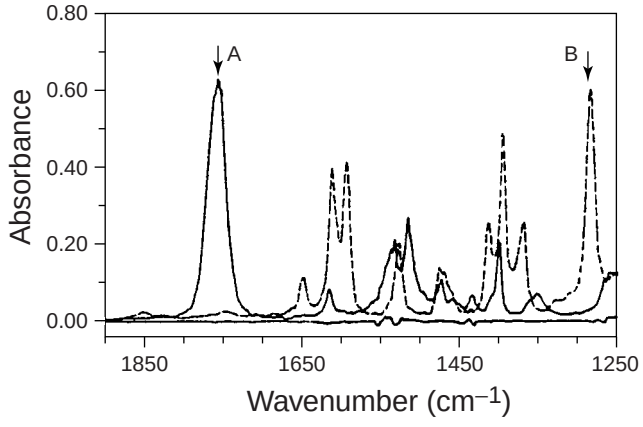

Figure 6 Absorbance spectra of carbaryl (—) and 1-naphthol (- - -) dissolved in $CHCl_3$ showing peaks at (A) 1741 cm^{-1} and (B) 1276 cm^{-1}. (Reproduced by permission of Elsevier Science from Y. Dagbouche, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta*, **314**, 203–212, © 1995.)

cell with a 10-mm pathlength. The detection limit was 1 mg L^{-1} when 300 µL of the extract was injected.

Miller et al.[34] used FIA/IR for the simultaneous detection of succinylcholine chloride and bethanechol chloride, two pharmaceutically important compounds. The compounds were specifically monitored at 1075 cm^{-1} and 953 cm^{-1}, respectively, using a CIRCLE® cell. The detection limit was about 0.02% (w/v) at a throughput of 60 samples per hour. FIA/IR was also applied to the quantification of acetylsalicylic acid, caffeine and paracetamol in pharmaceutical formulations.[27,28] Using a chlorinated solvent the extinction of a characteristic absorption band could be measured for each of the analytes in the 1800–1500 cm^{-1} region. Commercially available flow cells with an optical pathlength of 0.117 mm (5 µL) or 0.17 mm (7.2 µL) were used. Two FIA/IR methods for automated analysis in clinical and process chemistry have been developed by Kellner et al.[35,36] The first method aimed for the determination of glucose and urea after on-line enzymatic digestion to gluconic acid and ammonium carbonate, respectively. A 25-µm pathlength transmission cell was used to monitor the aqueous flow from the enzyme detector. The performance of the system was satisfactory for relative clean samples such as standard solutions, fruit juices and soft drinks. However, for routine application, for example in blood analysis, further improvement of the reproducibility is still required. The second FIA/IR method was used to determine the amyloglucosidase activity during starch hydrolysis processes. The IR flow cell (pathlength, 49 µm) was constructed of two different window materials in order to compromise between high transparency (ZnSe) and low reflectivity (calcium fluoride). Because of the high viscosity of the starch solutions, a stopped-flow method was used to obtain maximum reproducibility and sensitivity. The changes in the IR-spectra of the process mixture appeared to be directly correlated with the enzyme activity.

4 SOLVENT-ELIMINATION TECHNIQUES

4.1 Deposition Substrates and Infrared Detection Modes

The solvent-elimination approach in LC/IR involves the use of an eluent-evaporation interface that deposits the LC-separated compounds onto an IR-compatible substrate. With most described set-ups, after immobilization of one or more chromatograms, the substrate is transferred to the IR spectrometer where spectra from the deposited spots are recorded. The deposited traces on the substrate may be moved (stepwise) through the interrogating IR beam so that, when scanning is complete, continuous IR chromatograms can be reconstructed by computer software. In some designs, IR detection is executed during the immobilization process, within 5–20 s after deposition, allowing spectra to be obtained in real time. Obviously, such a design requires a dedicated detector set-up. Dependent on the type of substrate used (see below) and/or size of the deposited spots, often special optics, such as a (diffuse) reflectance unit, a beam condensor or an IR microscope, are used to scan the deposited compounds.

Basically three types of substrates and corresponding IR modes are used in solvent-elimination LC/IR: powder substrates for diffuse reflectance Fourier transform infrared (DRIFT) detection; metallic mirrors for reflection/absorption (R/A) spectrometry; and IR-transparent windows for transmission measurements. In early solvent-elimination LC/IR designs DRIFT detection of analytes on potassium chloride (KCl) powder was used, but today other, more convenient, detection modes are preferred. DRIFT as such is one of the most sensitive IR modes and sub-microgram identification limits can be achieved when the residual solvent is evaporated quickly from the powder. If the eluent is not highly volatile, it can draw analyte away from the KCl powder surface into the substrate. This will limit the sensitivity, because the effective penetration depth of a DRIFT measurement is not more than 100 µm. To overcome this problem, diffuse transmittance spectrometry has been applied instead of DRIFT, using a layer of KCl powder on an IR-transparent substrate.[37] The main limitations of DRIFT detection in LC/IR, show up during application. Reproducibility is hard to control since factors such as sample homogeneity, sample load and compactness of the powder layer significantly influence the DRIFT analysis. Reorientation of the DRIFT

matrix as a result of sample deposition may lead to a poor background compensation. Careful filling of cups or trays with the powder substrate is very time-consuming and has to be repeated for every analysis. Finally, common substrates such as KCl powder cannot be used in combination with aqueous eluents. In view of the overriding importance of RPLC, this is a very serious restriction. Some authors have used diamond powder as a water-resistant DRIFT substrate, but it is expensive (and thus not disposable) and not easy to clean.

Front-surface aluminum mirrors, which are suitable for R/A detection, are compatible with aqueous eluents and are relatively easy to handle. Compound deposition on this type of substrate requires efficient solvent-elimination interfaces because residual solvent will easily spread over the hard and smooth reflective surface. The band intensities in the R/A spectroscopy are largely governed by a double-pass transmittance mechanism, so that spectral data analogous to transmission data are obtained. Some useful results of solvent-elimination LC/IR using mirrors have been reported, although several authors[38–40] have reported evidence of band asymmetry and spectral distortions. Aspects such as specular and diffuse reflection from the analyte, thickness and microcrystallinity of the spot, and optical characteristics of the substrate affect the shape and intensity of R/A spectra obtained from analytes on aluminum mirrors. The use of an IR-transparent germanium disc with a reflective backing has been proposed in order to reduce spectral distortions. This type of disc is used in the commercially available LC-Transform™ LC/IR interfaces.[10] The cleaning of aluminum mirrors in between analyses is quite delicate: the thin metal layer is fragile and can be damaged easily by rubbing.

Most favorable results in solvent-elimination LC/IR are obtained with IR-transparent deposition substrates that allow straightforward transmission measurements. So far mainly KBr and ZnSe windows have been applied in experimental LC/IR set-ups. These substrates have a hard and smooth surface and, therefore, eluent elimination has to be fast to achieve proper depositions. ZnSe is a water-resistant, inert material, while KBr usually cannot be used in combination with RPLC. Deposited compounds on ZnSe can be removed simply, with water or alcohol, so that one window can be used repeatedly. With ZnSe, good-quality IR spectra with symmetrical band profiles can be recorded from deposited spots. When the size of the sample spots is small and microscopic optics are used for measurement, the sensitivity of ZnSe transmission measurements is higher than the sensitivity of DRIFT measurements.[40] ZnSe windows also cause fewer spectral artifacts than mirror substrates for R/A detection.[39] Many LC/IR studies demonstrated that spectra obtained using ZnSe, closely resemble conventional KBr-disk transmission spectra. Consequently, existing spectral libraries and search programs can be used for identification purposes, which is very important for the acceptance of IR as a valuable detection technique.

The quality and appearance of spectra obtained with solvent-elimination LC/IR will be influenced by the morphology of the deposited analytes. The morphology will depend primarily on parameters such as eluent composition, evaporation rate, temperature and nature of the substrate and the analytes. During solvent elimination some compounds will form nice crystals while others will deposit as an amorphous layer. Also, some analytes will deposit as a smooth film, whereas others may form irregular clusters. When the spot thickness exceeds a certain level, the effect of light scattering may become apparent. A compound may also exhibit polymorphism so that mutually (slightly) different spectra can be obtained for the same compound. In general, IR detection of deposited compounds on IR-transparent substrates does not pose serious problems. However, analyte morphology should always be taken into consideration during spectral interpretation.

In solvent-elimination LC/IR the identification limits usually improve when the width of the analyte spots is decreased. A prerequisite for this gain is the use of the appropriate detector and sampling optics. Optimum solvent-elimination interfaces can produce analyte spots with a width as small as 100–300 μm. For these deposits, the focus of a conventional beam condensor is too large and the use of an IR microscope is indicated. Frequently, the sensitivity enhancement is rationalized by considering the relatively increased spot thickness only (Beer's law!), but this approach is too simple. From more complete S/N considerations it follows that the good sensitivity of IR microscopic detection essentially results from the low noise level of the IR detectors in IR microscopes.[4] In other words, to achieve the most sensitive IR detection in LC, the width of the analyte deposits should have the same dimensions as the microscope detector area (typically, 100–200 μm). Of course, as with any IR experiment, the S/N ratio also can be improved by increasing the measurement time (signal averaging). As outlined in section 2.2, this advantage can be exploited to its full extent in solvent-elimination LC/IR, although at the cost of an increased time of analysis.

4.2 Early Interfaces

The solvent-elimination systems that were developed in an early stage, generally used KCl-powder substrates for DRIFT detection or flat KBr windows for transmission detection. The first working interface for the coupling of NPLC and IR was designed by Kuehl and Griffiths.[3,41]

The organic eluent was led through a heated concentrator tube and dropped into a series of cups filled with KCl powder suited for DRIFT analysis. A carousel rotated the cups into the IR spectrometer where identifiable spectra could be recorded for sub-microgram amounts of analyte. In order to allow for RPLC separations, the aqueous effluent was first on-line extracted with dichloromethane which, after continuous phase separation, was directed through the concentrator to the KCl cups.[42] For extracted compounds good-quality spectra were obtained. The carousel–DRIFT method was also adopted for narrow-bore NPLC (use of 1-mm i.d. column) by reducing the size of the KCl cups and by omitting the concentrator tube.[43] Using a similar set-up, Kalasinsky et al.[44,45] coupled both narrow-bore NPLC and RPLC with DRIFT. The KCl-powder substrate was held either in a ''train'' of compartments or in a continuous trough. Aqueous eluents could be used by on-line conversion of the water into methanol and acetone via a post-column reaction with 2,2′-dimethoxypropane (DMP). The identification limits of these systems were 1–3 μg, typically.

The early DRIFT-based systems for the first time demonstrated that solvent-elimination LC/IR can provide (much) better sensitivity and spectral quality than flow-cell based LC/IR. However, the systems were mechanically complex and tedious to work with, and DRIFT detection appeared to be strongly affected by small disturbances of the KCl-powder surface and by the presence of (residual) water.

Jinno et al.[46–47] proposed the use of micro-LC columns (i.d., 0.3 mm) in solvent-elimination LC/IR in order to alleviate the problem of the evaporation of large eluent volumes. The effluent ($5\,\mu L\,min^{-1}$) from either a GPC or an NPLC was led directly to a moving KBr plate which was covered by a stream of heated nitrogen. Subsequently, the plate with the deposited track was scanned by IR transmission spectroscopy using a 3 × beam condensor. The potential of the approach (termed ''buffer-memory'' technique) was illustrated by the analysis of a mixture of dithiocarbamate metal complexes by three spectroscopic techniques.[47] In a modified set-up the linearly moving KBr plate was replaced by a rotating KBr disk which, after the chromatographic run was finished, was transferred to a special rotation module in the IR spectrometer.[48] In order to permit the use of RPLC, a SS wire net (WN) was used instead of an KBr window.[49] IR transmission measurements were possible because after deposition and drying the analytes were partly suspended in the metal meshes.

The ''buffer-memory'' technique demonstrated the usefulness of the storage of a continuous chromatogram on a flat substrate. Besides, it was considerably simpler than the DRIFT methods. However, at least several micrograms of analyte were needed for a positive IR identification. These amounts often exceeded the sample capacity of the micro-columns and required unrealistically high concentrations to be injected.

4.3 Spray-type Interfaces

When using the ''buffer-memory'' technique for compound deposition on flat substrates, it is not possible to eliminate organic or aqueous eluents at flow rates higher than about $5\,\mu L\,min^{-1}$ without spreading the compounds over a large area of the substrate surface. To achieve a more viable coupling of LC and IR, the use of interfaces with enhanced evaporation power is essential. The solvent-elimination interfaces developed in the last decade all use some kind of spraying to induce rapid eluent evaporation. Heat, gas, electric potential and/or ultrasonic vibrations are used to break up the LC eluent stream into small droplets. Some designs incorporate existing (commercial) equipment, while others have been built from scratch.

In the thermospray (TSP) interface, originally developed for liquid chromatography/mass spectrometry (LC/MS), a directly heated tube evaporates part of the column effluent to an expanding vapor causing nebulization of the remaining effluent. As a result, a mist of desolvating droplets emerges from the tube. In the TSP-based LC/IR systems, nebulization is performed at atmospheric pressure and the spray is directed towards a deposition substrate. Eluent flow rates of up to $1\,mL\,min^{-1}$ can be handled. Griffiths and Conroy[50] reported preliminary results on the use of TSP for LC/IR, but the first really working interface was introduced by Jansen.[51] With a home-made TSP, the LC effluent was sprayed on a SS IR-reflective tape which moved through an optical accessory in the IR spectrometer (Figure 7). Most of the eluent was eliminated directly by the TSP and residual solvent was evaporated off the tape by heating. The immobilized chromatogram was monitored

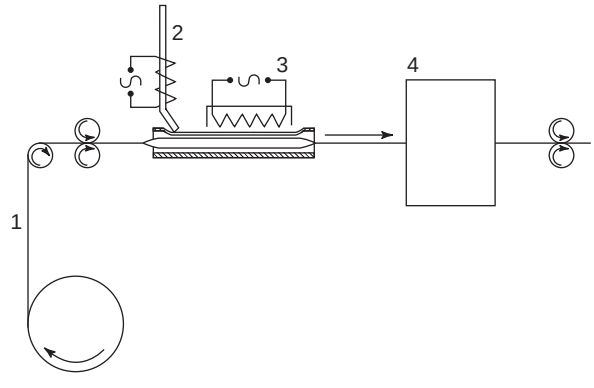

Figure 7 Schematic of TSP–moving belt interface for LC/IR; 1, moving SS tape; 2, TSP; 3, heater; 4, IR reflectance cell mounted in the IR spectrometer.

continuously and full IR spectra were recorded. Several simple polymer samples (20–80 μg injected) were analyzed by GPC/IR using an organic eluent at a flow rate of 0.5–1 mL min^{-1}, but some low-molecular-weight monomers were too volatile to be deposited. The characterization of two Irganox-type polymer additives (100 μg each), which were separated by RPLC using an eluent (0.5 mL min^{-1}) with 30 vol% water, was also shown. Robertson et al.[52] further optimized the TSP–moving belt interface by studying the effect of the TSP temperature and distance to the substrate. The system was used for the analysis of amino acids, saccharides, carboxylic acids, antioxidants and polymers,[53,54] and analyte identification could be achieved down to concentrations of 50 μg mL^{-1} or about 2.5 μg injected.

The main advantage of the TSP-based systems is that relatively high flow rates (0.5–1 mL min^{-1}) of both organic and aqueous eluents can be handled and conventional-size LC can thus be used. Furthermore, spectral data are acquired during the run, which gives the IR detector an essentially on-line character. On the other hand, the high temperature of the TSP may induce analyte losses by evaporation or thermal degradation and the analyte spots on the moving tape are still quite large which results in a moderate IR sensitivity.

The particle beam (PB) interface was modified for LC/IR by de Haseth et al.[55,56] and Wood.[57] In this interface the LC eluent is nebulized by helium and directed into a desolvation chamber where most of the liquid is vaporized. The mixture of gas, vapor and condensed analyte molecules (i.e. particles) is accelerated into the momentum separator where the analytes travel straight through the skimmer cone, while the gas and vapour are pumped away. For IR detection an IR-transparent substrate is placed in the particle-beam path to collect the analytes of interest (Figure 8) and after deposition, the substrate is transferred to the IR spectrometer for analysis. Until now stationary substrates have been used in liquid chromatography/particle beam/infrared (LC/PB/IR), that is, no complete chromatograms, only fractions, were analyzed. The particle beam/infrared (PB/IR) interface can effect the elimination of aqueous eluents at flow rates of up to 0.3 mL min^{-1} as was demonstrated by the analysis of erythrosin B and *p*-nitroaniline (50 μg each) by RPLC/IR.[56] Since the PB interface has strong eluent-elimination capacities, it was believed that interference caused by buffers might be small.[58] The interface indeed appeared to be able to process a 0.3 mL min^{-1} flow of buffered eluent, but the buffer salts were never completely eliminated. Best results were obtained with eluents buffered with ammonium acetate, although buffer bands were clearly present in the IR spectra recorded

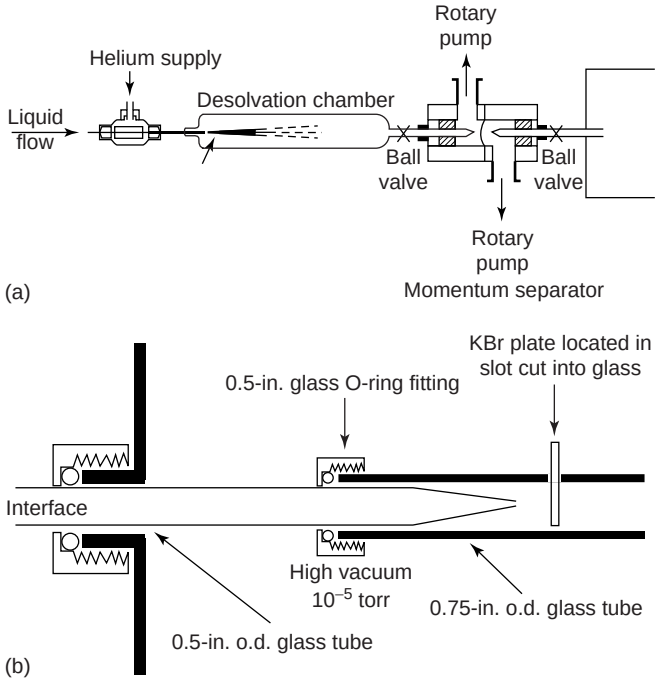

Figure 8 Schematic of PB interface for LC/IR.

from microgram amounts of analyte. When phthalate or phosphate buffers were used, the analyte spectra were completely dominated by absorption bands of the buffer salts. Spectral subtraction procedures could be used to recover spectra from 130-μg depositions but were unsuccessful at the 13-μg level. PB/IR has been used as a tool for the determination of protein structures.[59–61] For β-lactoglobulin and lysozyme it was shown that their structural integrity is maintained during the PB desolvation process and the subsequent deposition on the substrate. In addition, lysozyme appeared to retain its biological activity. The sample loads in these experiments generally were quite high (5–500 μg).

The PB interface can effectively remove both organic and aqueous solvents. However, relevant applications in LC/IR would still require the construction of a device that allows the continuous deposition of a complete chromatogram on a moving substrate. The PB/IR analysis of compounds at the nanogram level has been indicated,[55] but the reported sample quantities mainly are in the (high) microgram range. The modest analyte detectability no doubt is related to the fact that the efficiency of analyte transfer in the PB interface probably is 5–10% only.

The potential of electrospray (ESP) nebulization for micro-LC/IR was studied by Raynor and co-workers.[62] A high electrical potential is used to form a spray of charged droplets at the end of a capillary filled with flowing liquid. As a result of charge density, the initial droplets break up into smaller droplets which

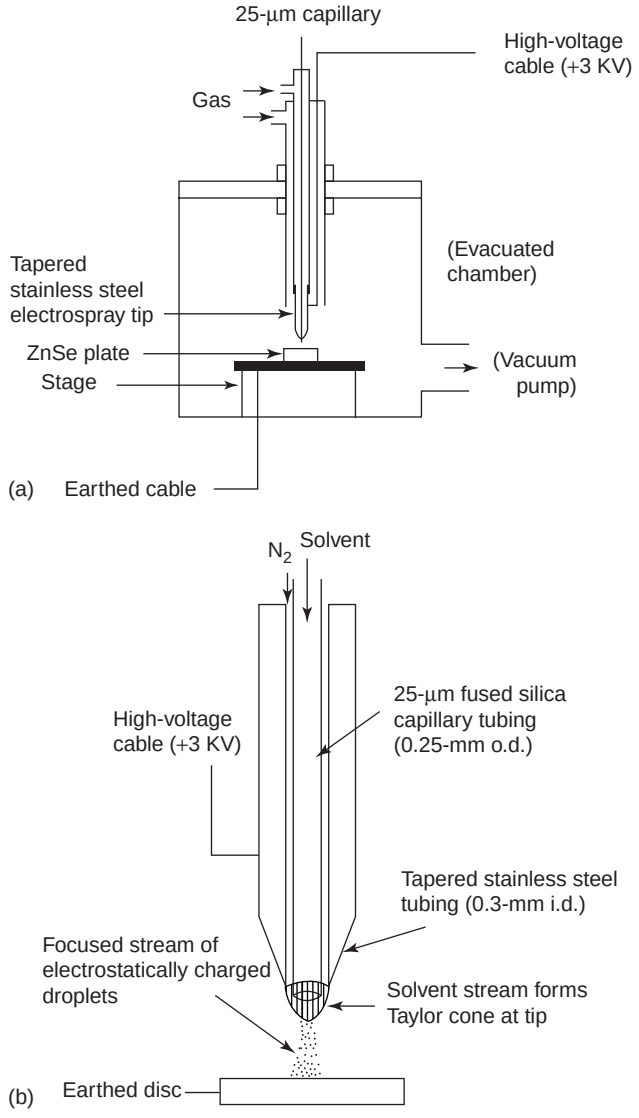

Figure 9 Schematic of ESP interface for (a) micro-LC/IR (b) with the ESP tip in detail. (Reproduced by permission of Hüthig Publishing from M.W. Raynor, K.D. Bartle, B.W. Cook, *J. High Resolut. Chromatogr.*, **15**, 361–366, © 1992.)

facilitate solvent evaporation. Use of low flow rates (typically 1–20 μL min^{-1}) is indicated in order to obtain a stable ESP. The ESP is formed under atmospheric conditions and a sheath flow of nitrogen gas is applied to enhance eluent evaporation (Figure 9). The ESP interface was used to deposit the effluent from a micro-RPLC column onto a ZnSe plate.[62] After LC separation, 20-ng amounts of caffeine and barbital could be analyzed successfully using an IR microscope. Identification of 2 ng caffeine appeared to be possible, although subtraction of the interfering bands from a siliceous impurity was required. Stable ESP conditions were achieved with hexane, dichloromethane, acetonitrile, methanol and several aqueous solvents, but problems were reported for

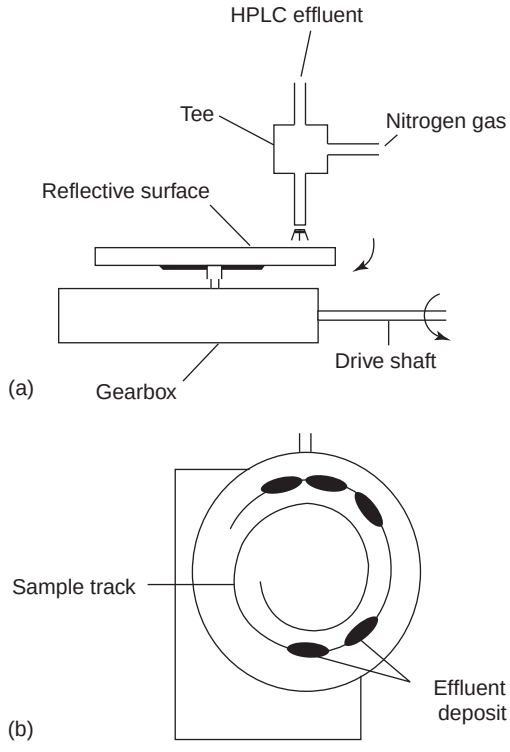

Figure 10 Schematic of a narrow-bore NPLC and IR system with pneumatic nebulization; (a) side view during deposition; (b) top view of collection mirror. (Reproduced by permission from J.J. Gagel, K. Biemann, *Anal. Chem.*, **58**, 2184–2189, © 1986, American Chemical Society.)

pure water. Until now there have been no further studies on electrospray/infrared (ESP/IR) detection in LC.

During pneumatic nebulization a high-speed gas flow is used to disrupt the liquid surface and to form small, fast-moving droplets. Organic solvents can be rapidly evaporated by pneumatic nebulization, while removal of aqueous solvents is possible when the nebulizer gas is heated. Pneumatic nebulizers have been used in several solvent-elimination LC/IR designs, among which are the most successful so far. One of the first nebulizer-based LC/IR methods involved the continuous spraying of the effluent from a narrow-bore NPLC column on a rotating IR-reflective disk (Figure 10).[63] The effluent was mixed with nitrogen gas to form a fine spray and the immobilized chromatogram was analyzed by rotating the disk through a 3 × -condensed IR beam while recording R/A spectra. The system was tested with polycyclic aromatic compounds (200–800 ng each) which were separated using hexane–dichloromethane as eluent (30 μL min^{-1}). Good-intensity spectra were obtained, although some spectral deviations were observed. This pneumatic nebulizer design was improved to accomplish elimination of aqueous solvents.[38] A heated nitrogen flow served as an evaporation-enhancing and spray-focusing sheath gas.

Eluents containing up to 55% water could be handled at $30\,\mu L\,min^{-1}$ and a number of isomeric naphthalenediols (500 ng each) were separated and identified. Again the recorded R/A spectra showed anomalies. These spectral problems could be partially solved by using an IR-transparent germanium disk with a rear surface of aluminum as substrate.[64] This pneumatic nebulizer LC/IR design is commercialized by Lab Connections.[10] The instrument consists of a sample-collection module and an optics module for R/A analysis. So far the commercial interface has been applied mainly in the field of GPC/IR.[65,66] It was also used for the identification of triclosan, an antibacterial agent, in toothpaste.[67]

A simple but effective concentric flow nebulizer (CFN) for the coupling of narrow-bore LC and IR spectrometry was constructed by Lange et al.[68] The interface consists of two concentric fused-silica capillaries. The LC column effluent is led through the inner capillary and heated helium gas through the outer capillary (Figure 11a). The hot gas facilitates the evaporation of the solvent and the focusing of the spray emerging from the inner tube. To enhance the elimination of aqueous eluents, the CFN and the ZnSe substrate were placed in a vacuum chamber (Figure 11b). IR microscopy was used for optimum detection. The CFN can handle eluents with up to 100% water at a flow-rate of $50\,\mu L\,min^{-1}$ and identifiable spectra of analytes can be recorded down to the low-nanogram range. The CFN was also installed in an evacuated compartment which included the IR-microscopic optics and a motor to translate the ZnSe window.[69] With this system, an RPLC effluent ($50\,\mu L\,min^{-1}$) could be continuously deposited on the moving substrate. After immobilization of the chromatogram, spectral data could be collected without the need to transport the substrate from the Chromatograph™ to the IR spectrometer. To further improve the on-line character of the system, a modified CFN was installed on the optical bench of a Tracer® (Biorad, Düsseldorf, Germany) gas chromatography–IR interface which allows spectral acquisition in real time.[70] So far, only some preliminary results with this on-line LC/IR system have been reported. Unfortunately, proper analyte spectra cannot be obtained with the CFN/IR system when using nonvolatile buffers, because of strong co-deposition of buffer salts.[71] However, if sufficient vacuum pump capacity is applied, a 1 mM ammonium acetate buffer can be completely eliminated, although higher ammonium acetate concentrations cause interference.

Somsen et al.[39] proposed a spray-jet interface for the coupling of narrow-bore RPLC and IR spectrometry. In this interface a heated nitrogen flow provides pneumatic nebulization of the column effluent ($20\,\mu L\,min^{-1}$) which is led through a SS needle that

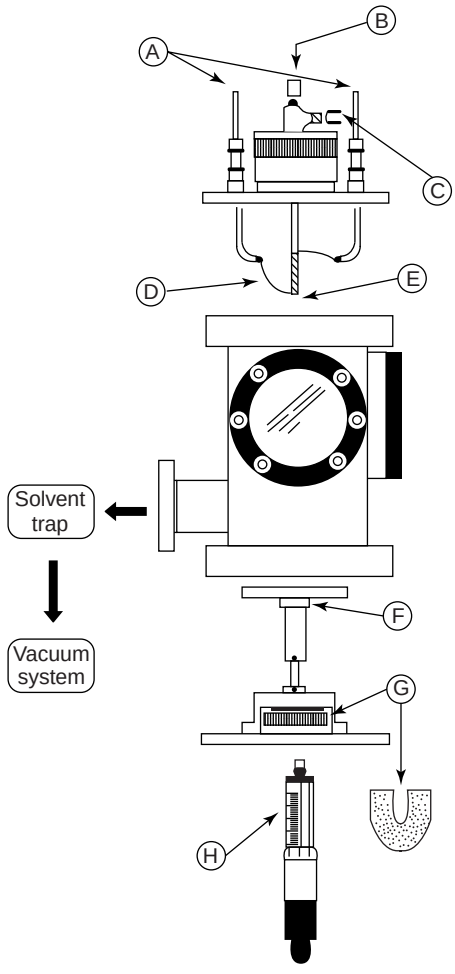

Figure 11 CFN for (a) narrow-bore LC/IR and (b) interface chamber; A, electric connections; B, LC effluent inlet; C, helium gas inlet; D, heating wire; E, concentric fused-silica tubes; F, stage; G, magnet for stage rotation; H, vertical positioner. (Reproduced by permission from A.J. Lange, P.R. Griffiths, D.J.J. Fraser, *Anal. Chem.*, **63**, 782–787, © 1991, American Chemical Society.)

protrudes through a spray nozzle (Figure 12). With ZnSe as substrate and IR microscopy for detection, identification limits in the 10–20-ng range were achieved for quinones and polycyclic hydrocarbons. The narrow-bore RPLC/IR system was used for the impurity profiling of a steroids,[72] for the isomer-specific characterization of chlorinated pyrenes,[73] and for the identification of additives in polymer samples.[74] Furthermore, the suitability of the interface for GPC/IR was demonstrated by analyzing polystyrene oligomers.[74] When RPLC is applied, the spray-jet LC/IR system is limited with regard to the LC flow rate, the water percentage of the eluent and the handling of buffered eluents. In order to take away these limitations, a post-column LLE module (phase segmentor, extraction coil and phase separator), was inserted on-line and

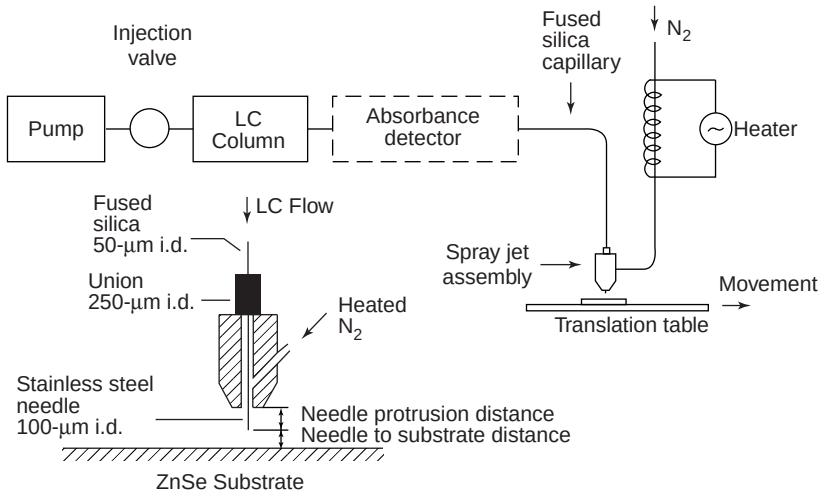

Figure 12 Schematic of the spray-jet interface for narrow-bore LC/IR.

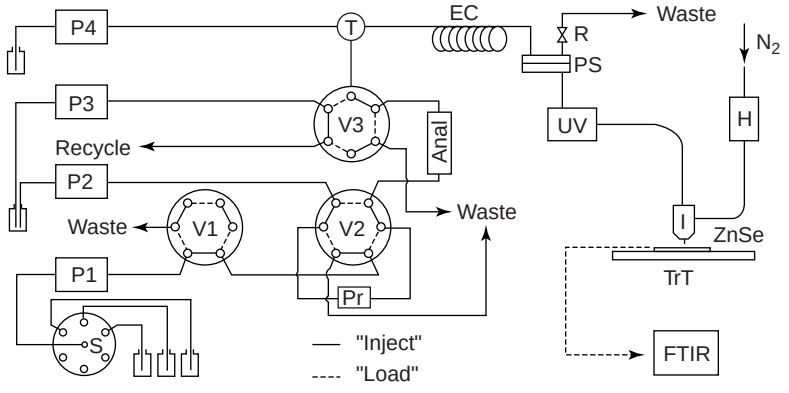

Figure 13 Set-up for LC/IR with on-line SPE and post-column LLE; P1, pump for sample loading and SPE-column rinsing; P2 and P3, eluent pumps; P4, organic phase pump; S, solvent selection valve; V1-V3, six-port injection valves; Pr, SPE column; Anal, analytical column; T, T-piece; EC, extraction coil; PS, phase separator; R, restrictor; UV, UV absorbance detector; I, spray-jet interface; H, gas heater; ZnSe, ZnSe deposition substrate; TrT, translation table; FTIR, FTIR microscope.

the organic phase, carrying the extracted analytes, was sent to the evaporation interface.[75] The resulting liquid chromatography/liquid–liquid extraction/infrared (LC/LLE/IR) system can handle eluents with high water percentages (20–100 vol%) at flow rates up to $0.2\,mL\,min^{-1}$ and provides identification of compounds at the sub-milligram per liter level. Since the salts are not extracted, nonvolatile buffers can be used without causing interference. The detectability of the analytes was further improved by incorporation of on-line SPE for analyte enrichment (Figure 13). With such a system, triazine herbicides, including several isomers, could be identified at the low-microgram per liter level in river water (Figure 14).[76]

In an alternative approach to improve the compatibility of the spray-jet interface with RPLC, the eluent flow rate was reduced to $2\,\mu L\,min^{-1}$.[77] To obtain a useful spray, a make-up liquid ($20\,\mu L\,min^{-1}$ of methanol) was added to the micro-LC effluent. As a consequence, the performance of the interface became independent of the water content of the eluent, so that gradient elution was possible. A micro-precolumn for on-line trace enrichment was applied to improve detection limits. With a 40-μL sample volume, good-quality IR chromatograms and analyte spectra were recorded at the low-milligram per liter level.

In an ultrasonic nebulizer a spray is formed by depositing the LC effluent on a transducer that is vibrating at ultrasonic frequencies. For LC/IR purposes, the spray is directed towards a substrate by a carrier gas. Castles et al.[78] used ultrasonic nebulization for the deposition of compounds separated by narrow-bore RPLC on a diamond-powder substrate suitable for DRIFT detection. Spectra of satisfactory quality were obtained

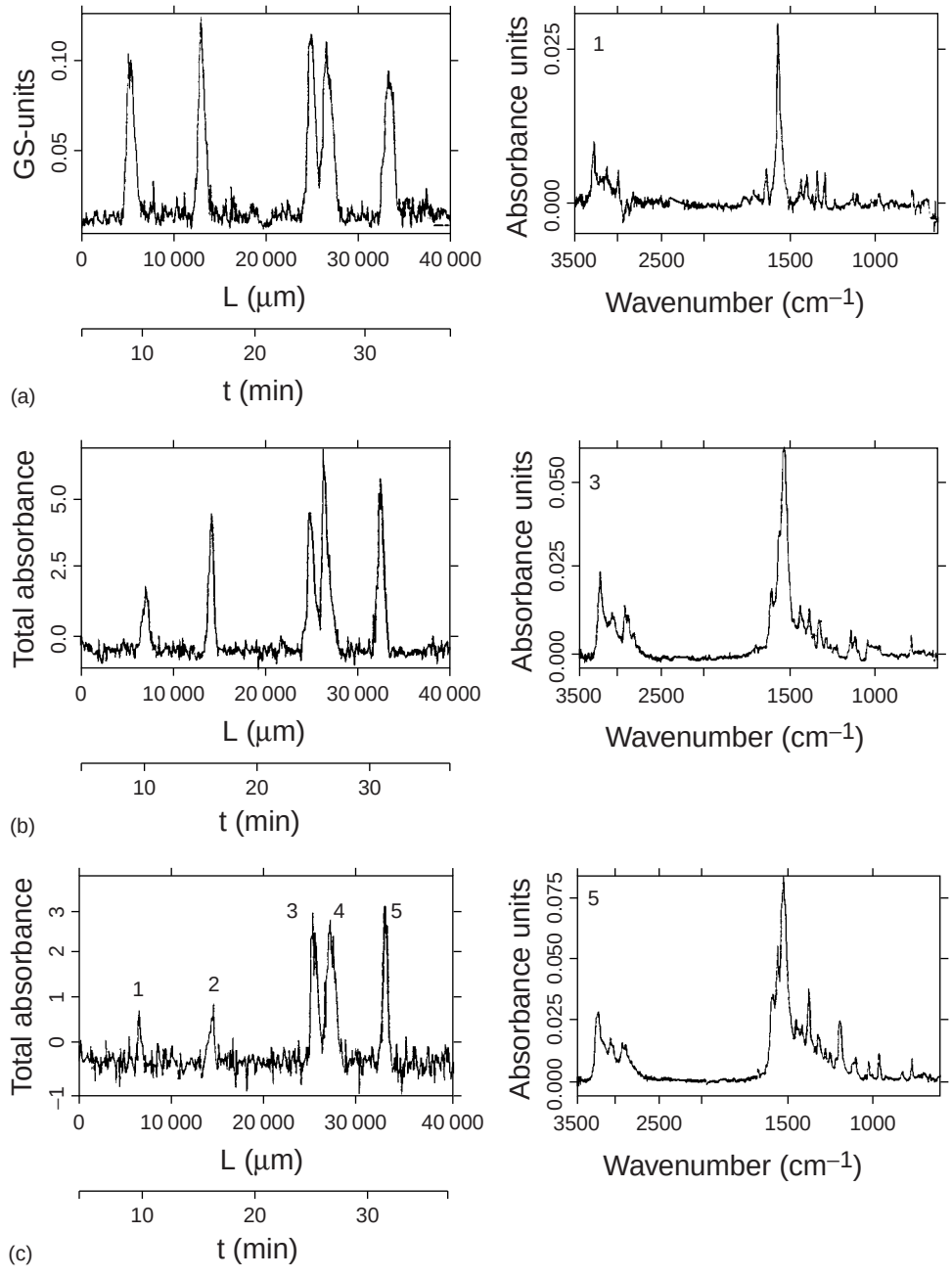

Figure 14 SPE and LC/LLE/IR chromatograms of river Meuse water samples spiked with five triazines: (a) 20 mL (30 μg L^{-1}), (b) 50 mL (6 μg L^{-1}) and (c) 100 mL (2 μg L^{-1}). IR spectra of peaks 1, 3 and 5. Chromatogram representation, (a) Gram–Schmidt or (b and c) spectral window (1650–1500 cm^{-1}). Peaks: 1, simazine; 2, atrazine; 3, sebutylazine; 4, propazine and 5, terbutylazine.

for injections of 3 μg of analyte. In some instances, the complete and direct evaporation of the eluent by the ultrasonic nebulizer was not achieved because the vibrating surface was not uniformly effective and occasionally large droplets were formed which wetted the diamond powder. Dekmezian and Morioka developed an interface for high-temperature GPC/IR which involved an ultrasonic nebulizer.[79] The nebulizer was placed in a vacuum chamber and sprayed the column effluent on a set of heated KBr discs, which were subsequently analyzed by IR transmission spectrometry. The system was applied to the determination of compositional changes of ethylene–propylene rubbers. An interface comprising an ultrasonic nebulizer in a vacuum chamber is used by Lab Connections in their LC-Transform™ 300 Series.[10] This commercial device sprays

the chromatographic effluent on a rotating germanium collection disk suited for R/A analysis (see above). The system was used for the quantitative analysis of copolymers by GPC/IR[80] and for steroid analysis by RPLC/IR.[81]

5 STATE OF THE ART

In the contemporary practice of LC/IR both flow-cell and solvent-elimination approaches are applied. Since the flow-cell procedure cannot get around the detection limitations imposed by the LC eluent, it has developed into a special-purpose method with restricted applicability. The IR absorptions of any solvent invariably take up parts of the mid-IR spectral region and, therefore, the main power of IR spectroscopy, i.e. the reliable identification of compounds, cannot be fully exploited in flow-cell IR detection. Nevertheless, making use of the spectral windows of the eluent, flow-cell IR spectroscopy can serve as a moderately sensitive, compound-specific detection technique. Occasionally, it is used as a universal, fast and low-cost method to obtain quantitative and structural information on major constituents of samples. Various types of flow cells are commercially available, and the experimental set-up and practice of flow-cell LC/IR is relatively simple and well-suited for routine applications.

When the objective of IR detection in LC is the unambiguous identification of (low-level) constituents of mixtures, coupling via solvent elimination should be the approach of choice. Solvent elimination procedures offer the possibility

- to record spectra over the entire mid-IR region without interference from the eluent;
- to perform signal averaging in order to improve the S/N of the spectra; and
- to contain a relatively large part of the chromatographic peak within the IR beam.

As a result, the solvent-elimination approach provides a set-up which features increased sensitivity and enhanced spectral quality, two important conditions for effective analyte identification. The most recent commercial LC/IR systems which are presently available,[10,11] are solvent-elimination devices. Interestingly, also in gas chromatography and supercritical-fluid chromatography analyte-deposition-based IR detection has proven to be more sensitive and versatile than flow-cell-based techniques. On the other hand, one should realize that the vibrational information obtained after solvent elimination is different from the vibrational information obtained with flow-cell detection (condensed-phase spectra against solution-phase spectra). Recently, this difference was used to determine subtle molecular features of drug metabolites which were analyzed by both flow-cell and solvent-elimination LC/IR.[82]

Today, the vast majority of LC separations is carried out by means of RPLC and, not surprisingly, research in the field of LC/IR has concentrated on the development of interfaces that are suitable for the elimination of aqueous eluents. Table 2 summarizes the characteristics of the various solvent-elimination reversed-phase liquid chromatography/Fourier transform infrared (RPLC/FTIR) systems which have been developed during the last years. The systems based on TSP, PB and ultrasonic nebulization

Table 2 Characteristics of solvent-elimination RPLC/IR systems[a]

Type of interfacing	IR mode	Substrate	Eluent flow rate[b] ($\mu L\,min^{-1}$)	Limit of identification[b]		Refs.
				mass (ng)	conc ($mg\,L^{-1}$)	
Concentrator after LLE	DRIFT	KCl powder	800	100	10	42
Deposition after DMP	DRIFT	KCl powder	50	1000	1000	44
Buffer memory	trans	SS/WN	4	1000	10 000	49
TSP	R/A	SS tape	500	10 000	–	51
	R/A	SS tape	1000	1000	25	52
PB	trans	KBr window	300	1000	200	56, 58
ESP	trans-micr	ZnSe	4	1	10	62
Pneumatic nebulizer	R/A	Al mirror	30	30	30	38
	trans-micr	ZnSe	50	1	17	69
	trans-micr	ZnSe	20	5	3	39, 73
...after LLE	trans-micr	ZnSe	200	30	0.2	75
...after SPE + LLE	trans-micr	ZnSe	200	50	0.001	76
...after SPE + make-up	trans-micr	ZnSe	2	20	0.02	77
Ultrasonic nebulizer	DRIFT	Diamond powder	40	1000	1000	78
	R/A	Ge disk	500	100	20	81

[a] DMP, on-line reaction with DMP; make-up, addition of excess methanol; trans, transmission; trans-micr, transmission with IR microscope; –, concentration and injection volume not stated.
[b] Typical values.

can handle relatively high flows of aqueous eluents and allow the use of conventional-size LC, which evidently is an advantage. However, these systems often exhibit moderate, or even unfavorable identification limits and, therefore, their analytical applicability is limited. Until now, the most favorable results have been obtained with pneumatic nebulizers which essentially represent the state of the art in solvent-elimination LC/IR.

The pneumatic interfaces combine rapid solvent elimination with a relatively narrow spray. This implies that analytes can be deposited, for example on ZnSe, in a narrow track of spots and IR microscopy can be applied effectively to achieve mass sensitivities in the low-nanogram range. Bearing in mind that often only part of the injected amount of analyte is actually analyzed, this means that the mass detectability of these LC/IR systems approaches a level close to the minimum that can be identified by IR spectroscopy. The systems based on pneumatic nebulization are limited with regard to flow rate ($2-50\,\mu L\,min^{-1}$) and water percentage of the eluent. The tolerable water content of the eluent depends on the flow rate. Flow rates of $2-5\,\mu L\,min^{-1}$ of even pure water can be eliminated, but for $20-50\,\mu L\,min^{-1}$ flows of aqueous eluents, enhancement of the solvent evaporation efficiency is required, for example by mixing the effluent with nitrogen gas[38] or by placing both the nebulizer and the deposition substrate inside a vacuum chamber.[68] The tedious evaporation of water can also be circumvented by applying on-line LLE of the aqueous effluent with an organic solvent. Such a system allows much higher flow rates ($0.2\,mL\,min^{-1}$) and percentages of water (up to 100 vol%).[75,76] Of course, the required LLE module adds to the complexity of the system and the analytes must have a sufficiently high extraction efficiency. Solvent-elimination RPLC/IR with gradient elution poses the problem of the efficient evaporation of an eluent with a changing water content. One solution involves the increase of the temperature of the nebulization gas during the gradient run,[38] another the addition of excess methanol to a micro-LC effluent in order to mask the changes in its water content.[77]

The injection volumes that can be handled with micro- and narrow-bore LC columns, are at most $1-2\,\mu L$. The identification limits in terms of concentration units of the pneumatic nebulizer-based systems therefore are in the low-milligram per liter range, which is sufficient for a number of analytical applications.[72–74] By using the LLE-pneumatic nebulizer combination, the detectability can be improved to sub-milligram per liter levels because larger LC columns – and, thus, increased injection volumes – can be used. With an adequate trace-enrichment procedure such as on-line SPE, the LC/IR detection limits can be further improved down to the low-microgram per liter level.[76] The use of buffered eluents is generally avoided in solvent-elimination LC/IR, since buffer salts may seriously affect the deposition and detection of the analytes. With pneumatic nebulizers even volatile buffer salts are rarely completely eliminated. In fact, until now the LLE-pneumatic nebulizer combination is the only described LC/IR system which allows the use of nonvolatile buffer salts without introducing interfacing disturbances and/or spectral interference.[75,76]

Effective solvent elimination by the pneumatic nebulizers allows the use of deposition substrates with a hard and smooth surface such as ZnSe windows. With these substrates interference- and distortion-free transmission spectra are obtained which can be readily compared with conventional KBr-disk IR spectra. This implies that libraries of condensed-phase reference spectra can be used for spectral recognition and identification. Compounds of various nature such as quinones, steroids, drugs, polymer additives and herbicides have been analyzed successfully by pneumatic-nebulization-based LC/IR. The interfaces can handle most types of analytes, although too volatile compounds will be evaporated by the nebulizer gas and therefore will not be deposited on the substrate. Thermal degradation of analytes is commonly not observed during pneumatic nebulization, despite the fact that the nebulizer gas is heated to rather high temperatures (70–180 °C). Probably, due to the rapid evaporation of the solvent, the spray droplets are cooled considerably.

6 PERSPECTIVE AND FUTURE DEVELOPMENTS

In the last fifteen years, LC/IR has emerged as a potentially powerful tool for the specific detection of major components (flow-cell approach) or for the identification of (minor) constituents of complex mixtures (solvent-elimination approach). With respect to common LC detection techniques such as UV/VIS absorption detection, the sensitivity of flow-cell IR detection is rather poor and its merits therefore mainly lie in the ability to quantitatively monitor absorptions that are specific for the analyte or for a certain functional group. Since the limitations of flow-cell LC/IR are inherent in the technique, no vast improvements can be anticipated in the future. Some gain in performance may be achieved by optimization of the flow-cell design and use of advanced FTIR spectrometers, but these improvements will be modest and not essentially change the applicability of flow-cell IR detection. Employment of chemometrical techniques, however, may be useful particularly when the signal or spectrum is the result of the absorbance of two or more substances.

The usefulness of solvent-elimination LC/IR to provide structural information and/or identification of unknowns

has been demonstrated convincingly since 1990. Unfortunately, the development of coupling techniques proceeded, and still proceeds, quite slowly and until now most interfaces have been used only by their designers. Nevertheless, the difficulty of solvent-elimination LC/IR, i.e. simultaneous eluent evaporation and analyte deposition, seems to be a technical rather than a fundamental problem. In other words, the development of an overall effective and routinely applicable interface probably is a matter of time, effort and technological innovations. Of course, solvent-elimination LC/IR has to compete with other identification techniques of which today on-line LC/MS undoubtedly is one of the most important. Quite a number of LC/MS interfaces have been developed and commercialized, but still each interface has its specific limitations. Furthermore, MS techniques cannot discriminate between isomers. Hence, even with adequate LC/MS techniques available, there is a need for alternative and complementary detection techniques which independently confirm MS-based identifications and differentiate between structurally highly similar compounds. A recent study demonstrated that LC/IR can make a viable contribution to identification analysis in a research setting that includes MS and NMR.[82] Prerequisite for the successful implementation of IR detection was a good understanding of the relative strengths and weaknesses of each technique, and the integration of analysis in the total research program.

In order to enhance the acceptance of solvent-elimination LC/IR, several items of interest should be considered. The practicality of the technique for real-life samples should be demonstrated more extensively. The applications described so far[72–74,76] indicate that LC/IR can indeed be used for the characterization and unambiguous identification of target and unknown compounds. LC/IR is particularly useful for the distinction of isomeric compounds[77–79] which cannot be distinguished by LC/MS. Another item of attention is the development and use of appropriate on-line sample-treatment procedures to improve analyte detectability. Despite the low-nanogram identification limits, the detectability in concentration units of even the best LC/IR systems will not be sufficient to meet current demands in bio- and environmental analysis. On-line SPE can improve the concentration detectability by one or two orders of magnitude. Obviously, such an improvement is unlikely to be obtained by optimization of interfacing and/or IR detection only. The first studies indicating this advantage in both flow-cell and solvent-elimination IR-detection have already been reported.[31,32,76,77]

Concerning the viability of solvent-elimination LC/IR, the availability and use of commercial interfaces also is essential. The LC-Transform™ interface (Lab Connections) has been available now for several years, but unfortunately few applications have been reported. Because this solvent-elimination system uses a mirror substrate and standard IR equipment, both the IR sensitivity and spectral quality are limited. In a more viable approach an IR-transparent substrate should be used together with microscopic IR detection. Such a configuration is used by the Infrared Chromatograph™ interface (Bourne Scientific). In this commercial and automated design the LC column effluent is deposited on a moving ZnSe window which instantaneously passes through the focused beam of the IR spectrometer allowing spectra and IR chromatograms to be recorded in real time. The placement of the chromatograms on the substrate is controlled by computer which also keeps a record of the position of deposited compounds. The IR Chromatograph™ seems promising but, as it has been introduced only recently, it is still too early to assess its merits. The handling of the obtained spectral data also is a matter of concern in solvent-elimination LC/IR. The identification of analytes on the basis of their IR spectra often is a difficult operation. Therefore, the automated retrieval of spectra in reference collections, and the computerization of spectral interpretation are important. Several such procedures have already been introduced in the vibrational-spectroscopic field and high priority should be given to their implementation in the separation field.

Finally, it should be noted that the solvent-elimination approach in LC is not restricted to IR detection but can, in principle, be applied to any spectrometric technique which requires the compounds of interest to be present as deposits.[83] An example of such an analyte-deposition-based detection technique is matrix-assisted laser desorption/ionization (MALDI) MS. From a technical viewpoint the collector systems developed for the coupling of LC or capillary electrophoresis with MALDI/MS show a strong similarity with solvent-elimination LC/IR systems.

ABBREVIATIONS AND ACRONYMS

ATR	Attenuated Total Reflection
CFN	Concentric Flow Nebulizer
DMP	Dimethoxypropane
DRIFT	Diffuse Reflectance Fourier Transform Infrared
ESP	Electrospray
ESP/IR	Electrospray/Infrared
FIA	Flow-injection Analysis
FIA/IR	Flow-injection Analysis/Infrared
FTIR	Fourier Transform Infrared
GPC	Gel Permeation Chromatography

GPC/FTIR	Gel Permeation Chromatography/Fourier Transform Infrared
GPC/IR	Gel Permeation Chromatography/Infrared
IR	Infrared
ISO	International Organization for Standardization
LC	Liquid Chromatography
LC/IR	Liquid Chromatography/Infrared Spectroscopy
LC/LLE/IR	Liquid Chromatography/Liquid–liquid Extraction/Infrared
LC/MS	Liquid Chromatography/Mass Spectrometry
LC/PB/IR	Liquid Chromatography/Particle Beam/Infrared
LLE	Liquid–liquid Extraction
MALDI	Matrix-assisted Laser Desorption/Ionization
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NPLC	Normal-phase Liquid Chromatography
PB	Particle Beam
PB/IR	Particle Beam/Infrared
R/A	Reflection/Absorption
RPLC	Reversed-phase Liquid Chromatography
RPLC/FTIR	Reversed-phase Liquid Chromatography/Fourier Transform Infrared
RPLC/IR	Reversed-phase Liquid Chromatography/Infrared
SEC	Size-exclusion Chromatography
S/N	Signal-to-noise ratio
SPE	Solid-phase Extraction
SS	Stainless Steel
THF	Tetrahydrofuran
TSP	Thermospray
UV/VIS	Ultraviolet/Visible
WN	Wire Net
ZDV	Zero-dead-volume

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Clinical and Diagnostic Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Infrared Spectroscopy **(Volume 12)**
Gas Chromatography/Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. K.L. Kizer, A.W. Mantz, L.C. Bonar, 'On-line LC/FTIR', *Am. Lab.*, **7**(5), 85–90 (1975).
2. D.W. Vidrine, D.R. Mattson, 'A Practical Real-time Fourier Transform Infrared Detector for Liquid Chromatography', *Appl. Spectrosc.*, **32**, 502–506 (1978).
3. D. Kuehl, P.R. Griffiths, 'Novel Approaches to Interfacing a High Performance Liquid Chromatograph with a Fourier Transform Infrared Spectrometer', *J. Chromatogr. Sci.*, **17**, 471–478 (1979).
4. G.W. Somsen, C. Gooijer, N.H. Velthorst, U.A.Th. Brinkman, 'Coupling of Column Liquid Chromatography and Fourier-Transform Infrared Spectrometry', *J. Chromatogr. A*, **811**, 1–34 (1998).
5. K. Jinno, C. Fujimoto, 'Microcolumn Liquid Chromatography in Hyphenated Techniques (Micro-Column LC/FT-IR)', *Methods Chromatogr.*, **1**, 199–211 (1996).
6. G. Patonay, *HPLC Detection – Newer Methods*, VCH, New York, 1992.
7. R. White, *Chromatography/Fourier Transform Infrared Spectroscopy and its Applications*, Marcel Dekker, New York, 1990.
8. E.S. Yeung, *Detectors for Liquid Chromatography*, John Wiley & Sons, New York, 1986.
9. C.D. Tran, G. Huang, V.I. Grishko, 'Direct and Indirect Detection in Liquid Chromatography by Infrared Thermal Lens Spectrometry', *Anal. Chim. Acta.*, **299**, 361–369 (1995).
10. LC-Transform™ Series 100, 300 and 400, Lab Connections, Marlborough, MA, USA.
11. Infrared Chromatograph™, Bourne Scientific, Newtonville, MA, USA.
12. H. Poppe, 'Spectrometers and High Performance Liquid Chromatography', *Chromatographia*, **24**, 25–32 (1987).
13. R.S. Brown, L.T. Taylor, 'Microbore Liquid Chromatography with Flow Cell Fourier Transform Infrared Spectrometric Detection', *Anal. Chem.*, **55**, 1492–1497 (1983).
14. C.C. Johnson, L.T. Taylor, 'Normal-phase Liquid Chromatography/Fourier Transform Infrared Spectrometry for Analysis of Nonpolar Material with Semipreparative, Analytical and Microbore Columns', *Anal. Chem.*, **55**, 436–441 (1983).
15. C.W. Saunders, L.T. Taylor, 'Determination of the Degree of Oxidation of Cellulose Nitrates via GPC/FTIR using an On-line Flow Cell', *Appl. Spectrosc.*, **45**, 900–905 (1991).

16. K. Nishikida, T. Housaki, M. Morimoto, T. Kinoshita, 'Gel Permeation Chromatography/Fourier Transform Infrared Study of Some Synthetic Polymers II. Instrumentation for the Characterization of Polyethylene', *J. Chromatogr.*, **517**, 209–217 (1990).
17. R. Vonach, B. Lendl, R. Kellner, 'Hyphenation of Ion Exchange High Performance Liquid Chromatography with Fourier Transform Infrared Detection for the Determination of Sugars in Nonalcoholic Beverages', *Anal. Chem.*, **69**, 4286–4290 (1997).
18. R. Vonach, B. Lendl, R. Kellner, 'Fourier Transform Infrared (FTIR) Spectrometry as a Detection System for High Performance Liquid Chromatography (HPLC) for the Determination of Carbohydrates, Alcohols and Organic Acids', Poster presented at 11th ICOFTS, USA, 1997.
19. C.C. Johnson, J.W. Hellgeth, L.T. Taylor, 'Reversed-phase Liquid Chromatography with Fourier Transform Infrared Spectrometric Detection using a Flow Cell Interface', *Anal. Chem.*, **57**, 610–615 (1985).
20. R.G. Messerschmidt, 'Recent Developments in HPLC/FTIR', *Proc. SPIE-Int. Soc. Opt. Eng.*, **553**, 342–343 (1985).
21. J.E. DiNunzio, 'Pharmaceutical Applications of High Performance Liquid Chromatography Interfaced with Fourier Transform Infrared Spectroscopy', *J. Chromatogr.*, **626**, 97–107 (1992).
22. C. Fujimoto, G. Uematsu, K. Jinno, 'The Use of Deuterated Solvents in High-performance Liquid Chromatography/Fourier Transform Infrared Spectrometry', *Chromatographia*, **20**, 112–116 (1985).
23. S.S.-H. Chen, A.Y. Kou, 'High Performance Liquid Chromatography of Phospholipids using Deuterated Solvents for Infrared Detection', *J. Chromatogr*, **307**, 261–269 (1984).
24. E.E. Remsen, J.J. Freeman, 'A Size Exclusion Chromatography/FTIR (SEC/FTIR) Technique for Improved FTIR Spectroscopy of Proteins in D_2O solution', *Appl. Spectrosc.*, **45**, 868–873 (1991).
25. J. Ruzicka, E.H. Hansen. *Flow Injection Analysis*, 2nd edition, John Wiley & Sons, New York, 1988.
26. M. Guzman, J. Ruzicka, G.D. Christian, 'Enhancement of Fourier Transform Infrared Spectrometry by the Flow-injection Technique: Transmittance and Internal Reflectance Cell in a Single-line System', *Vibr. Spectrosc.*, **2**, 1–14 (1991).
27. S. Garrigues, M. Gallignani, M. de la Guardia, 'Simultaneous Determination of Acetylsalicylic Acid and Caffeine in Pharmaceuticals by Flow-injection Fourier Transform Infrared Detection', *Talanta*, **40**, 1799–1807 (1993).
28. Z. Bousain, S. Garrigues, M. de la Guardia, 'Flow-injection Fourier Transform Infrared Spectrometric Determination of Paracetamol in Pharmaceuticals', *Analyst*, **121**, 635–639 (1996).
29. M. de la Guardia, S. Garrigues, M. Gallignani, J.L. Burguera, M. Burguera, 'Flow-injection Fourier Transform Infrared Spectrometric Analysis', *Anal. Chim. Acta* , **261**, 53–57 (1992).
30. M. Gallignani, S. Garrigues, M. de la Guardia, 'Direct Determination of Benzene in Gasoline by Flow-injection Fourier Transform Infrared Spectrometry', *Anal. Chim. Acta*, **274**, 267–274 (1993).
31. S. Garrigues, T. Vidal, M. Gallignani, M. de la Guardia, 'On-line Preconcentration and Flow Analysis-Fourier Transform Infrared Determination of Carbaryl', *Analyst*, **119**, 659–664 (1994).
32. Y. Dagbouche, S. Garrigues, M. de la Guardia, 'Solid phase Preconcentration/Fourier Transform Infrared Spectrometric determination of Carbaryl and 1-Naphthol', *Anal. Chim. Acta* , **314**, 203–212 (1995).
33. Y. Dagbouche, S. Garrigues, M. de la Guardia, 'Flow-injection Fourier Transform Infrared Spectrometric Determination of Oil and Greases: Preliminary Microwave-assisted Extraction Studies', *Analyst*, **121**, 1031–1036 (1996).
34. B.E. Miller, N.D. Danielson, J.E. Katon, 'Aqueous Infrared Pharmaceutical Analysis of Two Choline Compounds by Flow Injection Analysis using the CIRCLE Cell', *Appl. Spectrosc.*, **42**, 401–405 (1988).
35. E. Rosenberg, R. Kellner, 'Flow Injection Analysis with Fourier Transform Infrared Detection for Clinical and Process Analysis', *Fresenius' J. Anal. Chem.*, **348**, 530–532 (1994).
36. R. Schindler, B. Lendl, R. Kellner, 'Determination of Amyloglucosidase Activity using Flow Injection Analysis with Fourier Transform Infrared Spectrometric Detection', *Analyst*, **122**, 531–534 (1997).
37. R.G. Messerschmidt, M.A. Harthcock, 'Infrared Microspectroscopy: Theory and Applications', Marcel Dekker, New York, Chap. 14, 1988.
38. J.J. Gagel, K. Biemann, 'Continuous Infrared Spectroscopic Analysis of Isocratic and Gradient Elution Reversed-phase Liquid Chromatographic Separations', *Anal. Chem.*, **59**, 1266–1272 (1987).
39. G.W. Somsen, R.J. van de Nesse, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, T. Visser, P.R. Kootstra, A.P. J.M. de Jong, 'Spray Jet Assembly Interface for the Coupling of Reversed-phase Narrow-bore Liquid Chromatography and Fourier Transform Infrared Spectrometry', *J. Chromatogr.*, **552**, 635–647 (1991).
40. R. Fuoco, S.L. Pentoney Jr, P.R. Griffiths, 'Comparison of Sampling Techniques for Combined Supercritical Fluid Chromatography and Fourier Transform Infrared Spectrometry with Mobile Phase Elimination', *Anal. Chem.*, **61**, 2212–2218 (1989).
41. D.T. Kuehl, P.R. Griffiths, 'Microcomputer-controlled Interface between a High Performance Liquid Chromatograph and a Diffuse Reflectance Fourier Transform Infrared Spectrometer', *Anal. Chem.*, **52**, 1394–1399 (1980).
42. C.M. Conroy, P.R. Griffiths, P.J. Duff, L.V. Azaragga, 'Interface of a Reversed-phase Liquid Chromatograph

with a Diffuse Reflectance Fourier Transform Infrared Spectrometer', *Anal. Chem.*, **56**, 2636–2642 (1984).
43. C.M. Conroy, P.R. Griffiths, K. Jinno, 'Interface of a Microbore High Performance Liquid Chromatograph with a Diffuse Reflectance Fourier Transform Infrared Spectrometer', *Anal. Chem.*, **57**, 822–825 (1985).
44. K.S. Kalasinsky, J.A.S. Smith, V.F. Kalasinsky, 'Microbore High Performance Liquid Chromatography/Fourier Transform Infrared Spectrometry for Normal-phase and Reversed-phase Liquid Chromatography', *Anal. Chem.*, **57**, 1969–1974 (1985).
45. V.F. Kalasinsky, K.G. Whitehead, R.C. Kenton, J.A.S. Smith, K.S. Kalasinsky, 'HPLC/FTIR Interface for Normal- and Reversed-phase Analytical Columns', *J. Chromatogr Sci.*, **25**, 273–280 (1987).
46. K. Jinno, C. Fujimoto, D. Ishii, ' ''Buffer Memory'' Technique for the Combination of Micro High Performance Liquid Chromatography and Infrared Spectrometry', *J. Chromatogr.*, **239**, 625–632 (1982).
47. C. Fujimoto, K. Jinno, Y. Hirata, 'Liquid Chromatography/Spectrometry with the Buffer Memory Technique', *J. Chromatogr*, **258**, 81–92 (1983).
48. C. Fujimoto, T. Morita, K. Jinno, S. Ochiai, 'Microcolumn Liquid Chromatography/Fourier Transform Infrared Spectrometry Interface using a Potassium Bromide Disk', *Chromatographia*, **23**, 512–516 (1987).
49. C. Fujimoto, T. Oosuka, K. Jinno, 'A New Sampling Technique for Reversed-phase Liquid Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chim. Acta*, **178**, 159–167 (1985).
50. P.R. Griffiths, C.M. Conroy, 'Solvent Elimination Techniques for HPLC/FT/IR', *Adv. Chromatogr.*, **25**, 105–138 (1986).
51. J.A.J. Jansen, 'On-line Liquid Chromatography/Fourier Transform Infrared Spectrometry for the Analysis of Polymers and Additives', *Fresenius' J. Anal. Chem.*, **337**, 398–402 (1990).
52. A.M. Robertson, D. Littlejohn, M. Brown, C.J. Dowle, 'Assessment of a Thermospray Interface for the Coupling of High Performance Liquid Chromatography and Fourier Transform Infrared Spectrometry', *J. Chromatogr*, **588**, 15–24 (1991).
53. A.M. Robertson, D. Farnan, D. Littlejohn, M. Brown, C.J. Dowle, 'Analysis of an (Isodecyl End-capped) Propylenediol Adipate Polyester using Coupled High Performance Liquid Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Proc.*, **30**, 268–271 (1993).
54. M.M. Mottaleb, B.G. Cooksey, D. Littlejohn, 'Evaluation of Gel Permeation Chromatography to Fourier Transform Infrared Spectrometric Detection Using a Modified Thermospray for Analysis of Polystyrene Samples', *Fresenius' J. Anal. Chem.*, **358**, 536–537 (1997).
55. R.M. Robertson, J.A. de Haseth, 'MAGIC-LC/FT-IR Spectrometry', R.F. Browner, *Mikrochim. Acta, II*, 199–202 (1988).
56. R.M. Robertson, J.A. de Haseth, J.D. Kirk, R.F. Browner, 'MAGIC-LC/FT-IR Spectrometry: Preliminary Studies', *Appl. Spectrosc.*, **42**, 1365–1368 (1988).
57. D.J. Wood, 'High Performance Liquid Chromatography/Mass Spectrometry and High Performance Liquid Chromatography/Fourier Transform Infrared Spectroscopy using a Particle-beam Interface', *Spectr. Int.*, **2**(3), 36–40 (1990).
58. R.M. Robertson, J.A. de Haseth, R.F. Browner, 'MAGIC-LC/FT/IR Spectrometry with buffered solvent systems', *Appl. Spectrosc.*, **44**, 8–13 (1990).
59. V.E. Turula, J.A. de Haseth, 'Evaluation of Particle Beam Fourier Transform Infrared Spectrometry for the Analysis of Globular Proteins: Conformation of β-Lactoglobulin and Lysozyme', *Appl. Spectrosc.*, **48**, 1255–1264 (1994).
60. V.E. Turula, J.A. de Haseth, 'Particle Beam LC/FT/IR Spectrometry Studies of Biopolymer Conformations in Reversed-phase HPLC Separations: Native Globular Proteins', *Anal. Chem.*, **68**, 629–638 (1996).
61. V.E. Turula, R.T. Bishop, R.D. Ricker, J.A. de Haseth, 'Complete Structure Elucidation of a Globular Protein by Particle Beam Liquid Chromatography/Fourier Transform Infrared Spectrometry and Electrospray Liquid Chromatography/Mass Spectrometry. Sequence and Conformation of β-Lactoglobulin', *J. Chromatogr. A*, **763**, 91–104 (1997).
62. M.W. Raynor, K.D. Bartle, B.W. Cook, 'Electrospray Micro Liquid Chromatography/Fourier Transform Infrared Micro Spectrometry', *J. High Resolut. Chromatogr.*, **15**, 361–366 (1992).
63. J.J. Gagel, K. Biemann, 'Continuous Recording of Reflection-absorbance Fourier Transform Infrared Spectra of the Effluent of a Microbore Liquid Chromatograph', *Anal. Chem.*, **58**, 2184–2189 (1986).
64. J.J. Gagel, K. Biemann, 'The Continuous Infrared Spectroscopic Analysis of Reversed-phase Liquid Chromatographic Separations', *Mikrochim. Acta, II*, 185–187 (1988).
65. L.M. Wheeler, J.N. Willis, 'Gel Permeation Chromatography/Fourier Transform Infrared Spectrometry Interface for Polymer Analysis', *Appl. Spectrosc.*, **47**, 1128–1130 (1993).
66. T.C. Schunk, S.T. Balke, P. Cheung, 'Quantitative Polymer Composition Characterization with a Liquid Chromatography/Fourier Transform Infrared Spectrometry Solvent Evaporation Interface', *J. Chromatogr. A*, **661**, 227–238 (1994).
67. S.L. Jordan, L.T. Taylor, B. McPherson, H.T. Rasmussen, 'Identification of an Anti-bacterial Agent in Toothpaste via Liquid Chromatography/Fourier Transform Infrared Spectrometry Mobile Phase Elimination', *J. Chromatogr. A*, **755**, 211–218 (1996).
68. A.J. Lange, P.R. Griffiths, D.J.J. Fraser, 'Reversed-phase Liquid Chromatography/Fourier Transform Infrared Spectrometry using Concentric Flow Nebulization', *Anal. Chem.*, **63**, 782–787 (1991).

69. P.R. Griffiths, A.J. Lange, 'On-line Use of the Concentric Flow Nebulizer for Direct Deposition Liquid Chromatography/Fourier Transform Infrared Spectrometry', *J. Chromatogr. Sci.*, **30**, 93–97 (1992).
70. J. Yang, P.R. Griffiths, 'Fabrication of an On-line Interface for Coupling HPLC with FT/IR Spectrometry', *SPIE*, **2089**, 336–337 (1993).
71. A.J. Lange, P.R. Griffiths, 'Use of Buffered Solvent Systems with Concentric Flow Nebulization Liquid Chromatography/Fourier Transform Infrared Spectrometry', *Appl. Spectrosc.*, **47**, 403–410 (1993).
72. G.W. Somsen, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, T. Visser, 'Coupling of LC and FT/IR: Impurity Profiling of Testosterone Undecanoate', *Appl. Spectrosc.*, **46**, 1514–1519 (1992).
73. G.W. Somsen, L.P.P. van Stee, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, T. Visser, 'Isomer and Congener Identification of Chlorinated Pyrenes by Narrow-bore Liquid Chromatography/Fourier Transform Infrared Spectrometry', *Anal. Chim. Acta*, **290**, 269–276 (1994).
74. G.W. Somsen, E.J.E. Rozendom, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, 'Polymer Analysis by Column Liquid Chromatography Coupled Semi-on-line With Fourier Transform Infrared Spectrometry', *Analyst*, **121**, 1069–1074 (1996).
75. G.W. Somsen, E.W.J. Hooijschuur, C. Gooijer, U.A.Th. Brinkman, N.H. Velthorst, T. Visser, 'Coupling of Reversed-phase Liquid Column Chromatography and Fourier Transform Infrared Spectrometry using Postcolumn On-line Extraction and Solvent Elimination', *Anal. Chem.*, **68**, 746–752 (1996).
76. G.W. Somsen, I. Jagt, C. Gooijer, N.H. Velthorst, U.A. Th. Brinkman, T. Visser, 'Identification of Herbicides in River Water using On-line Trace Enrichment Combined with Column Liquid Chromatography/Fourier Transform Infrared Spectrometry', *J. Chromatogr. A*, **756**, 145–157 (1996).
77. T. Visser, M.J. Vredenbregt, G.J. ten Hove, A.P.J.M. de Jong, G.W. Somsen, 'Gradient Elution Liquid Chromatography/Infrared Spectrometry at $\mu g\,L^{-1}$ level using Capillary Column Switching and Addition of a Make-up Liquid. A Preliminary Study', *Anal. Chim. Acta*, **342**, 151–158 (1997).
78. M.A. Castles, L.V. Azarraga, L.A. Carreira, 'Continuous On-line Interface for Reversed-phase Microbore High Performance Liquid Chromatography/Diffuse Reflectance Infrared Fourier Transform Analysis', *Appl. Spectrosc.*, **40**, 673–680 (1986).
79. A.H. Dekmezian, T. Morioka, C.E. Camp, 'Gel Permeation Chromatograph to Collect Solvent-free Polymer Fractions for Composition Drift Analysis', *J. Polym. Sci. B*, **28**, 1903–1915 (1990).
80. M.X. Liu, J.L. Dwyer, 'Quantitative Analysis of Copolymers by GPC/FTIR', *Appl. Spectrosc.*, **50**, 348–356 (1996).
81. J.L. Dwyer, A.E. Chapman, X. Liu, 'Analysis of Steroids by Combined Chromatography/Infrared Spectroscopy', *LC-GC Int.*, **8**, 704–710 (1995).
82. D.E. Pivonka, K.M. Kirkland, 'Research Strategy for the HPLC/FT/IR Analysis of Drug Metabolites', *Appl. Spectrosc.*, **51**, 866–873 (1997).
83. G.W. Somsen, C. Gooijer, U.A.Th. Brinkman, 'Analyte-deposition-based Detection in Column Liquid Chromatography: Concept and Examples', *Trends Anal. Chem.*, **17**, 129–140 (1998).

Microspectroscopy

Patricia L. Lang
Ball State University, Muncie, USA

1 Introduction **1**
1.1 History 1
1.2 Rationale 2
2 Theoretical Considerations in Instrument Design **2**
2.1 Diffraction 2
2.2 Numerical Aperture 4
2.3 Signal-to-noise Ratio 4
3 The Infrared Microscope **5**
3.1 Objectives and Condensers 5
3.2 Apertures 6
3.3 Stages 7
3.4 Visible Illumination and Viewing Systems 7
3.5 Infrared Sources and Detectors 8
3.6 Purge 8
3.7 Automation 8
4 Basic Transmission Techniques and Principles **8**
4.1 Sampling Considerations 8
4.2 Detection Limit 10
5 Basic Reflection Techniques and Principles **11**
5.1 Specular Reflection 11
5.2 Diffuse Reflection 12
5.3 Absorption/Reflection 13
5.4 Attenuated Total Reflection 14
6 Infrared Imaging Systems **15**
6.1 Traditional Mapping 15
6.2 Focal Plane Array Detectors 16
7 Perspective and Future Developments **18**
7.1 Spatial Resolution Below the Diffraction Limit 18
7.2 The Eighth Age of Infrared Spectroscopy 18
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

The essential features of an infrared (IR) microscope accessory include

1. *reflective optics for focusing, collecting, and imaging transmitted or reflected IR radiation from the sample onto a detector,*
2. *one or more variable apertures located in an image plane of the sample which serve to define the area to be recorded, and*
3. *a visible light path which is parfocal and collinear with the IR light path so that the sample can be viewed and positioned.*

The accessory is coupled to a Fourier transform IR (FTIR) spectrometer by a set of transfer optics from the interferometer into the microscope.

IR microspectroscopy provides the analyst with the capability to perform IR transmission, absorption-reflection, specular reflection, and diffuse reflection experiments on microscopic samples. Using special microscope objectives or accessories, attenuated total reflection (ATR) and grazing angle reflection experiments can be performed as well. IR microspectroscopy can also be used to map a sample's heterogeneity using a motorized stage in conjunction with a standard mercury cadmium telluride (MCT) detector. Alternatively, spatial and spectral information can be collected simultaneously using a step-scan interferometer coupled with an IR microscope and a focal plane array (FPA) detector. In either case, the superior contrast attained by functional group imaging and the selectivity found in the mid-IR region has been shown to be a powerful analytical tool.

However, diffraction ultimately limits the quality of the spectral information obtained using the IR microscope, and its effects may include poor spatial resolution, a reduction in photometric accuracy, and unreliable band intensities. Furthermore, unwanted optical effects from microscopic samples can also affect the quality of the spectrum, and sample preparation techniques must often take into account a sample's thickness, diameter, refractive index, and shape.

1 INTRODUCTION

1.1 History

IR microspectroscopy is the coupling of a microscope with an IR spectrometer for the purpose of obtaining spectra on microscopic samples. A mid-IR reflective microscope was first reported in 1949 at Oxford, England in H.W. Thompson's laboratory, and, soon after, the first commercially available IR microscope accessory was manufactured by the Perkin-Elmer Corporation.[1,2] Early investigators were interested in medical and biological applications and recorded spectra of single fibers, tissue sections, and blood smears.[3] But high-quality spectra were difficult to obtain on dispersive IR spectrometers

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

owing to the high energy losses of the microscope. Consequently, the technique remained little used until the development of FTIR spectrometers with their inherent throughput, multiplex, and frequency precision advantages. IR microspectroscopy consequently regained popularity, and a mid-IR microscope designed for the FTIR spectrometer was introduced at the 1983 Pittsburgh Conference and Exposition by Digilab Corporation.[4,5] Two years later the first symposium on the ''Design, Sample Handling, and Applications of Infrared Microscopes'' was held at a meeting sponsored by the American Society for Testing and Materials (ASTM) Committee on Molecular Spectroscopy and the Federation of Analytical Chemistry and Spectroscopy Societies; nine papers presented there were written, peer-reviewed, and published.[6]

1.2 Rationale

Not only does the IR microscope accessory provide the analyst with the capability to obtain IR spectra on extremely small samples, it is a versatile accessory which can used for transmission, absorption-reflection, specular reflection, and diffuse reflection experiments.[7,8,9] With special microscope objectives or accessories, attenuated total internal reflection and grazing angle reflection experiments can also be performed. Consequently, the technique allows for a wide variety of samples to be studied, and the reflection experiments, in general, offer ways to sample both microscopic and macroscopic samples with little or no preparation. Another important use of IR microspectroscopy is that the molecular heterogeneity of samples can be imaged and mapped.[10,11]

However, IR microspectroscopy offers a number of challenges in regard to instrumental design and sampling.[12,13] Diffraction ultimately limits the spatial resolution which can be obtained using the IR microscope, and unwanted optical effects from microscopic samples can affect the quality of the spectrum.[14]

Finally, owing to the selectivity of molecular information which can be attained, the vast number of applications of IR microspectroscopy found in the literature have been in the mid-IR region. The review will therefore focus primarily on the instrumental methods of mid-IR microspectroscopy in the wavelength range from 2 to 20 μm, including how one minimizes the effects of diffraction, refraction, reflection, and absorption by sampling techniques or instrumental parameters. However, near-infrared (NIR) microspectroscopy is an emerging field with distinct advantages in regard to some sampling factors and instrumental design considerations, and these aspects will be addressed as well. For all the apparent advantages of NIR microspectroscopy which will be mentioned, the facts remain that both the inherent functional group contrast needed for NIR imaging and the selectivity which allows for material identification are highly limited.[15] IR microspectroscopy in the mid-IR range is generally far superior and more widely applicable.

2 THEORETICAL CONSIDERATIONS IN INSTRUMENT DESIGN

2.1 Diffraction

Diffraction plays a major role in microspectroscopy when the sample size is on the order of the wavelength of light being used according to Rayleigh's relationship described in Equation (1):

$$D = \frac{1.22\lambda}{\mathrm{NA}} \tag{1}$$

where D is the diameter of the object and λ is the wavelength of light. NA is the numerical aperture which is a measure of the collection ability of an optical system.[7]

The effects of diffraction are graphically illustrated in Figure 1 which shows the diffraction pattern, or Airy disk, which results when a 17 μm IR point source is imaged by an optical system with a NA of 0.58, a typical NA for an IR microscope objective. (17 μm is the longest wavelength in the mid-IR range detected by a narrow-band MCT detector.) Notice that the energy is not imaged to a point but appears as a succession of bright rings which contain some of the energy from the original source. For the experimental conditions described, there will be 85% of

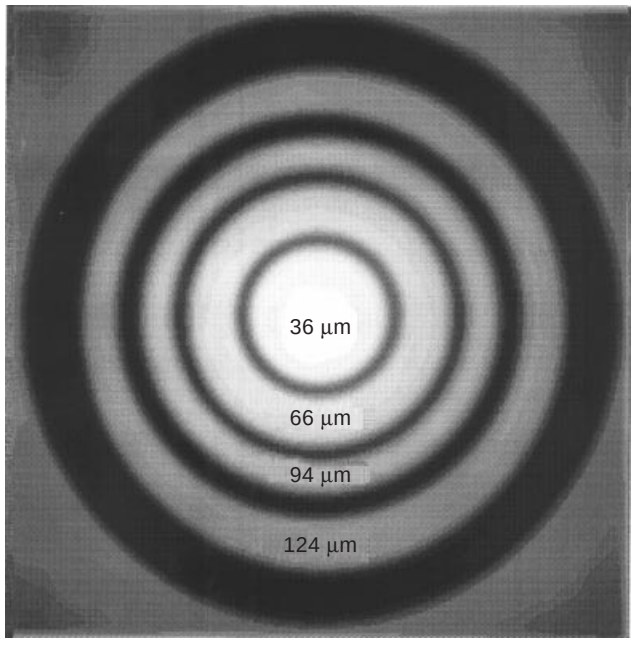

Figure 1 Graphical representation of an Airy disk for a 17 μm point source focused by an optical system with a 0.58 NA. The dimensions given are the outer diameters of the bright rings.

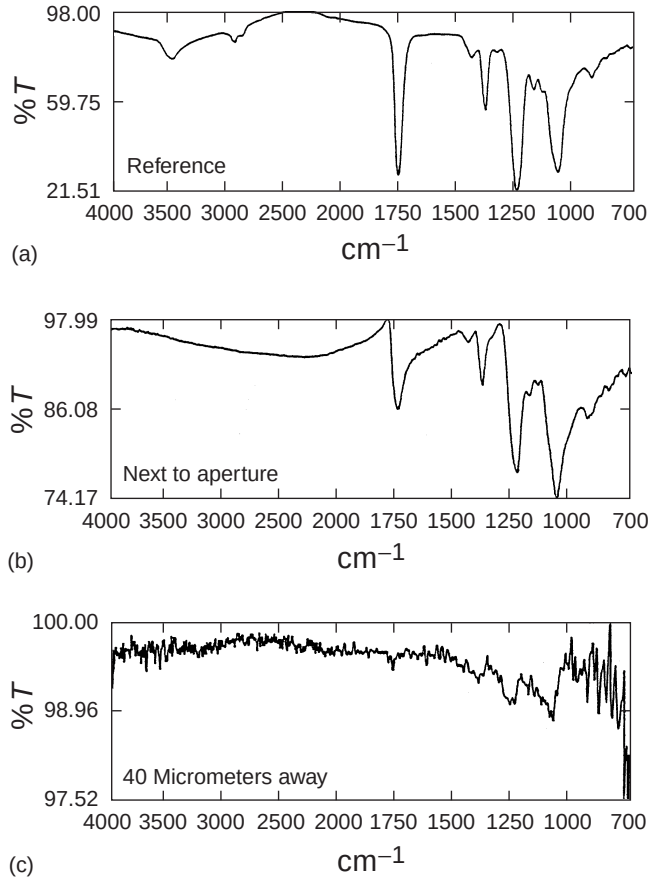

Figure 2 (a) Spectra of a cellulose acetate film recorded within a 8 mm × 240 mm diameter aperture, (b) recorded with the film's edge aligned with the edge of aperture image, and (c) recorded with the film's edge 40 μm away from the aperture's edge. (Reproduced by permission of the Society for Applied Spectroscopy from A.J. Sommer, J.E. Katon, *Appl. Spectrosc.*, **45**(10), 1633 (1991).)

the total original energy in the central spot, which is 36 μm in diameter as calculated by Equation (1). The rest is distributed among the remaining rings, and 5% of the energy still remains beyond the outer diameter of the fourth ring at 124 μm. Likewise, the IR radiation collected from a 17 μm diameter sample will result in a "blurring" of spectral information.[13] The energy which reaches the detector will primarily contain information about the sample, but like the fringes of the Airy disk, radiation which is present outside areas defined by the aperture, called spurious energy, can also reach the detector. Studies show that using a 8 μm effective diameter aperture results in an actual sampling diameter of 48 μm, for example, while using a 24 μm aperture results in a sampling diameter of 56 μm.[14] Diffraction effects can occur at either the high-contrast apertures located in the image plane(s) of the sample which are used to delineate the area or at the sample itself. Some of these effects can be addressed through sampling considerations and will be discussed in section 4.1.

First consider diffraction at the aperture. If an absorbing sample is nearby or adjacent to the sample under study, the spurious energy may result in absorption bands from the neighboring object "contaminating" the IR spectrum of the sample. The spatial resolution of the IR microscope is effectively decreased, and the two objects are not resolved. This is illustrated in Figure 2 which shows spectra of a cellulose acetate film obtained using a 8 mm × 240 mm aperture when the film completely fills the entire aperture (Figure 2a), when the edge of the film, out of view, is lined up with the edge of the aperture (Figure 2b), and after the edge of the film has been moved 40 μm away from the edge of aperture image (Figure 2c). Although there was no sample in the field of view defined by the aperture in the middle and bottom spectrum, spectral contamination is clearly present from the cellulose acetate even from 40 μm away, a situation with important implications for the analysis of multilayered polymer cross-sections.[7,14]

If there is only a nearby hole, that is if the free-standing particle is well removed from any other sample, the stray light reaching the detector may result in an offset of zero percent transmission. The photometric accuracy is decreased, and quantitative measurements are affected.

One way to minimize the above diffraction effects is to use dual remote (confocal) aperturing where an aperture is placed in an image plane located both before and after the sample. For example, Figure 3 shows the IR

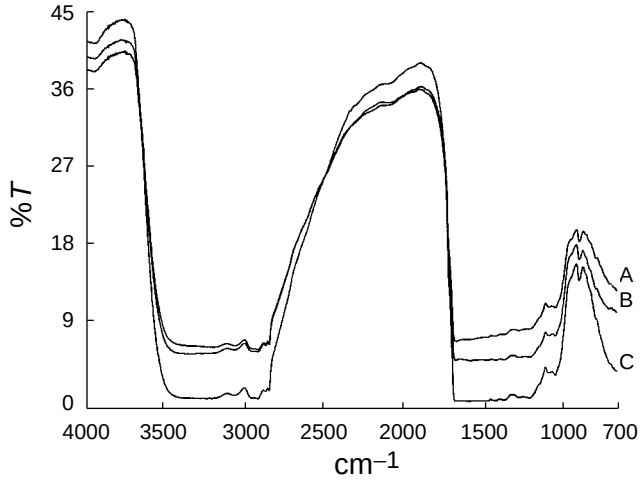

Figure 3 IR spectra of a hair fiber collected with a single aperture placed before the sample, A, a single aperture placed after the sample, B, and with both apertures in place, C. (Reproduced by permission of Marcel Dekker, Inc. from J.E. Katon, A.J. Sommer, P.L. Lang, *Appl. Spectrosc. Rev.*, **25**, 173–211 (1989–1990).)

transmission spectrum obtained on a hair fiber collected with a single aperture (A and B) and with confocal apertures (C). Although the thickness of the fiber is such that the N–H stretching (3300 cm^{-1}) and the amide I and II (1650, 1550 cm^{-1}) absorptions are totally absorbing, a marked offset from zero transmission is observed for single aperture systems, while an improvement is noted for the confocal configuration.[7]

It is important to note that while the theoretical spatial resolution using standard mid-IR microscope accessories is estimated at 34 μm using the previous criteria ($17\,\mu m = \lambda$; $0.58 = NA$), microspectroscopy in the *near*-infrared region, from 1 μm to 4 μm, offers the potential for higher spatial resolution in accordance with Rayleigh's criterion.

2.2 Numerical Aperture

Another variable in IR microspectroscopy which may be altered to our benefit is the NA. NA, is defined by Equation (2):

$$NA = n_1 \sin\theta \tag{2}$$

where n_1 is the refractive index of the medium and θ, is the half angle of collection.[13] Figure 4 illustrates the how NA increases with collection angle.

Increasing the NA of the focusing lens and collecting objective will decrease the beam diameter as well as the distance needed to spatially resolve two closely spaced samples as Equation (1) predicts. It also increases the throughput, Θ, which is the product of the solid angle subtended by the objective and the area of the focused beam, A. According to Messerschmidt[13] the throughput is related to NA by Equation (3):

$$\Theta = 2\pi[1 - (1 - (NA)^2)^{1/2}]A \tag{3}$$

Therefore one could design a microscope of high NA which would allow the use of dual aperturing to reject stray light while at the same time keep an adequate energy signal. Equation (3) predicts that for a system of 0.50 NA there is about five times more signal than for a 0.25 NA system.[13] In addition, Messerschmidt has shown that the 95% confidence level for the spurious energy rejection occurs at a two-wavelength-wide slit width for a dual aperture system with a 0.50 NA versus at a four-wavelength-wide slit for a system with a 0.25 NA. Therefore, a dual-masked system of 0.50 NA can analyze samples two times smaller than a single-masked system of 0.25 NA with an equivalent level of spurious energy rejection and with an increase in signal by a factor of five.

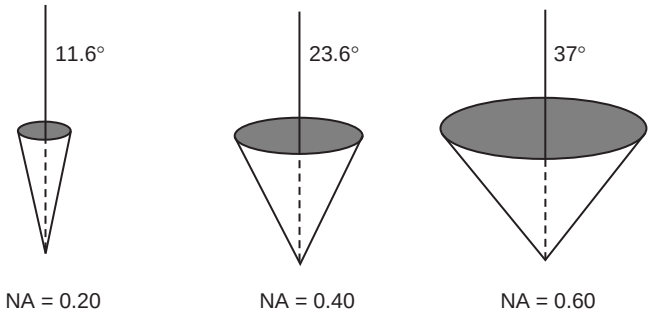

Figure 4 The relationship of NA to collection angle.

As the NA of the lens increases, on the other hand, the working distance decreases.[13] An adequate working distance is necessary for mounting cells and special stages and is therefore an important consideration in transmission experiments. For example, a 0.65 NA lens with 32 × magnification has an 8 mm working distance, while a 0.58 NA lens with 15 × magnification has a 15 mm working distance.[16] In addition, as NA increases, so do spherical aberrations which effectively limit the flatness of fields for wide-area imaging.[15] In other words, rays further from the optical axis are brought to a focus at a location nearer to the lens than those rays closer to the optical axis.

In regard to NIR microspectroscopy, a significant NA advantage can be achieved since refracting (glass) objectives can be employed. Refracting objectives, in comparison to reflective objectives which must be used in the mid-IR, can have higher NA. The use of higher NA objectives can result in an increase in both spatial resolution and throughput as described in Equations (1) and (3) respectively.[15]

2.3 Signal-to-noise Ratio

Griffiths and de Haseth state that the largest source of noise in IR spectrometry is detector noise.[5] Furthermore, the optimum performance is found when the optics are throughput-matched, a condition which exists when the detector is exactly filled by the image of the source.[5] However, in regard to IR microspectroscopy, since the sample is brought into focus and defined by a variable aperture(s), the size of the image at the detector is dependent upon the sample size. Therefore microscope accessories are not designed for the system to be throughput-matched, for in doing so it would be ideal for one specimen size only.

It is customary for microscopes to be designed, aligned, and tested with a 100 μm diameter circular pinhole as the sample. If the optics provide 1 : 1 imaging from the sample to the detector, then a 100 μm diameter image will underfill a 250 μm by 250 μm area detector. Under these conditions, a 100% transmittance line at 2200 to 2100 cm^{-1} typically produces a signal-to-noise ratio (S/N) of 4000 : 1 or greater for a 2 minute collection time using

$4\,cm^{-1}$ spectral resolution, measuring peak-to-peak noise, and with a narrow or medium band MCT detector.

3 THE INFRARED MICROSCOPE

Figure 5 shows the essential features of the standard IR microscope. They include one or more variable apertures located in the image plane of the sample which serve to define the area to be recorded and reflective optics for focusing, collecting, and imaging transmitted or reflected IR radiation from the sample onto a detector.[7] The microscope must also have a visible light path which is parfocal and collinear with the IR light path so that the sample can be easily viewed and positioned. The accessory can be coupled to an FTIR spectrometer by sending an external IR beam via a set of transfer optics from the interferometer into the microscope which is located to one side of a standard IR bench. Another type of instrumentation houses the interferometer, optics, and microscope all in one enclosure.

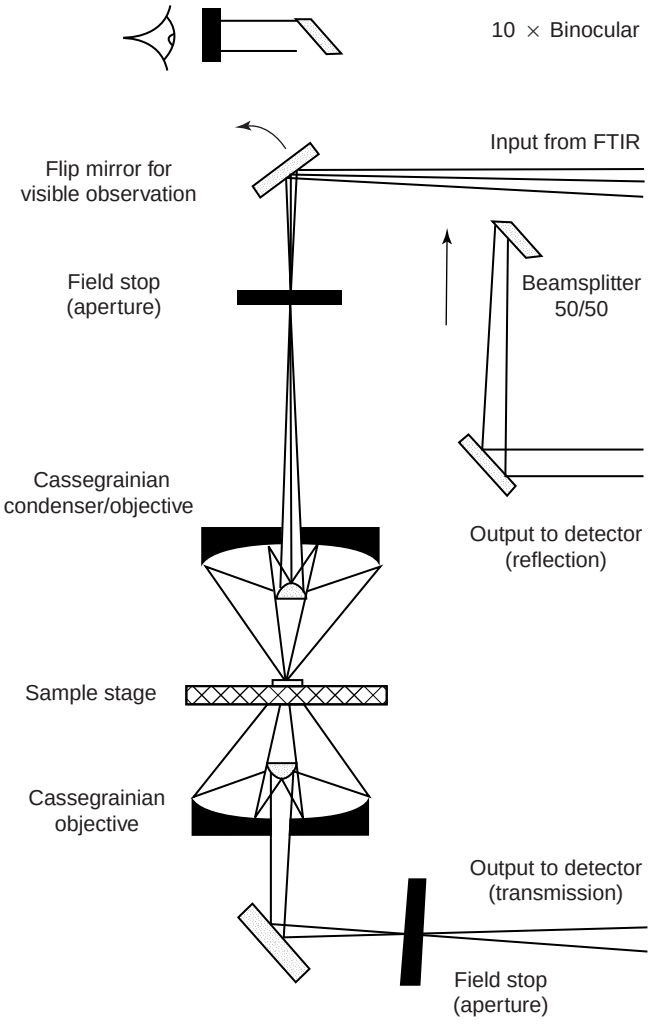

Figure 5 Optical diagram of a standard IR microscope accessory. (Reproduced by permission of Marcel Dekker, Inc. from J.E. Katon, A.J. Sommer, P.L. Lang, *Appl. Spectrosc. Rev.*, **25**, 173–211 (1989–1990).)

3.1 Objectives and Condensers

The objective and the condenser are the main lenses in the IR microscope accessory. Reflecting optics are typically used for these lenses in the mid-IR since they do not absorb IR light, possess chromatic aberrations, or introduce strain birefringents. Although two types can be found in IR microscopes, the on-axis Schwarzschild-type (commonly referred to as the Cassegrain) and the off-axis parabolic lens, all current models use only the Schwarzschild-type reflecting objectives. Figure 6 shows a cross-section of a typical Schwarzschild lens.

The Schwarzschild lens is used as a focusing lens, or objective, because of its high image quality and efficiency. These objectives have NAs which range from 0.28 to 0.60 and magnifications up to 36×. Although there is no dependence of magnification on signal, higher

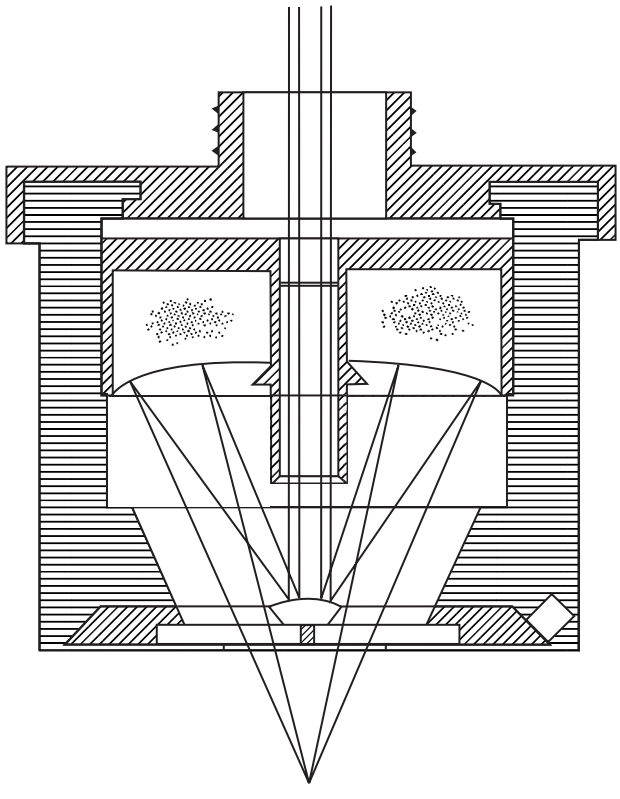

Figure 6 The cross-section of a Cassegrainian lens. (Reproduced by permission of Marcel Dekker, Inc. from J.E. Katon, A.J. Sommer, P.L. Lang, *Appl. Spectrosc. Rev.*, **25**, 173–211 (1989–1990).)

magnification objectives can allow the use of wider apertures to define the sample and therefore reduce diffraction effects at the aperture. However, as noted above, there is also a decrease in working distance. A standard Schwarzschild lens focuses IR radiation onto the sample with light rays ranging from 12° to 35°.

Unique to the reflecting objective is the characteristic known as central obscuration which is defined as the light intensity obstructed by the secondary mirror to that which would have been received by the primary mirror in its absence. Central obscuration is typically minimized to values below 25%.[7] The obscuration has the unwanted effect of reducing image brightness by spreading the diffraction pattern. In comparison with refracting objectives, centrally obscured objectives are limited to lower magnifications and NAs.

In older microscope accessories, one might find an off-axis paraboloid serving as a condenser. But the parabolic lenses, in general, provide a poorer definition of sampling area, produce more glare, and provide a wider focal beam at the sample than the Schwarzschild condensers.[16] The consequence of a larger focal diameter is that for samples smaller than the beam, more spurious energy may be introduced.

Refraction by the sample support is an important consideration in the design of the objectives and condensers. For example, when light passes from a lower refractive index medium to a higher one, the rays are bent toward the normal. The shift, which is also dependent on thickness, t, is described by Equation (4):

$$(n - 1.0)\frac{t}{n} \tag{4}$$

where n is the refractive index of the specimen, and the refractive index of air is assumed to be 1.0.[13] In a visible light microscope the objective is designed with a correction for a standard cover glass, and the condenser, for a standard glass slide. In IR microspectroscopy, the sample might lie on a variety of supports ranging from potassium chloride or zinc sulfide, to diamond or silicon. Therefore supports which have a high refractive index or which are very thick can give rise to large focal shifts. Figure 7 demonstrates that a focal shift of about 400 μm is observed when the support is changed from potassium chloride to zinc sulfide.[7] This shift is significant considering that the typical sample thickness is usually less than 10 μm. Since the beam is not in focus at the sample, it cannot be focused adequately onto the detector, and the result is a decrease in the S/N and photometric inaccuracy. This focal shift can be compensated for by refocusing the condenser. The fixed off-axis system cannot be easily refocused and, thus, puts limitations on the type of sample supports and/or sampling cells one can use.

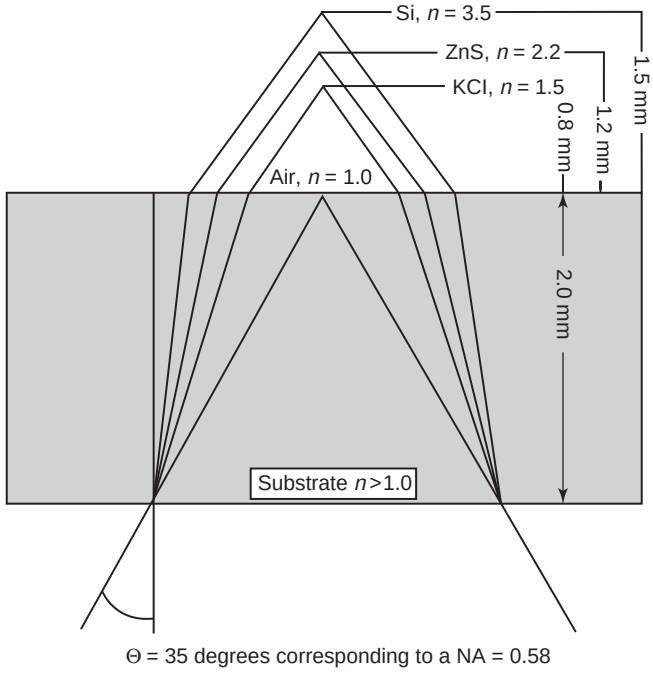

Figure 7 Focal shifts associated with refractive index changes. (Reproduced by permission of Marcel Dekker, Inc. from J.E. Katon, A.J. Sommer, P.L. Lang, *Appl. Spectrosc. Rev.*, **25**, 173–211 (1989–1990).)

A related phenomenon is spherical aberration, due to focal shifts which arise when an IR window is placed over the sample. Some Schwarzschild objectives contain a spherical aberration correction so that the analyst may compensate for these focal shifts.[16] These objectives can compensate for windows ranging from 1 mm to 3 mm in thickness.

3.2 Apertures

Sample definition can be obtained in two ways. One alternative is to place the sample in a pinhole.[17] However, this process is a tedious and time-consuming task. The other alternative is to place and adjust remote apertures in the sample's image planes to define the sample area as previously described and illustrated. Although the placement of an aperture at the image plane is traditionally called a field stop by microscopists, the term aperture has been employed by manufacturers of IR microscopes and will be used here. In all IR microscopes an aperture is found between the sample and the detector. Yet as stated in section 2.1, spurious energy can reach the detector owing to IR light bending around the aperture's edge. To minimize the effects from diffraction, some microscopes have an additional aperture which is positioned in an image plane located between the source and the sample. It has been found

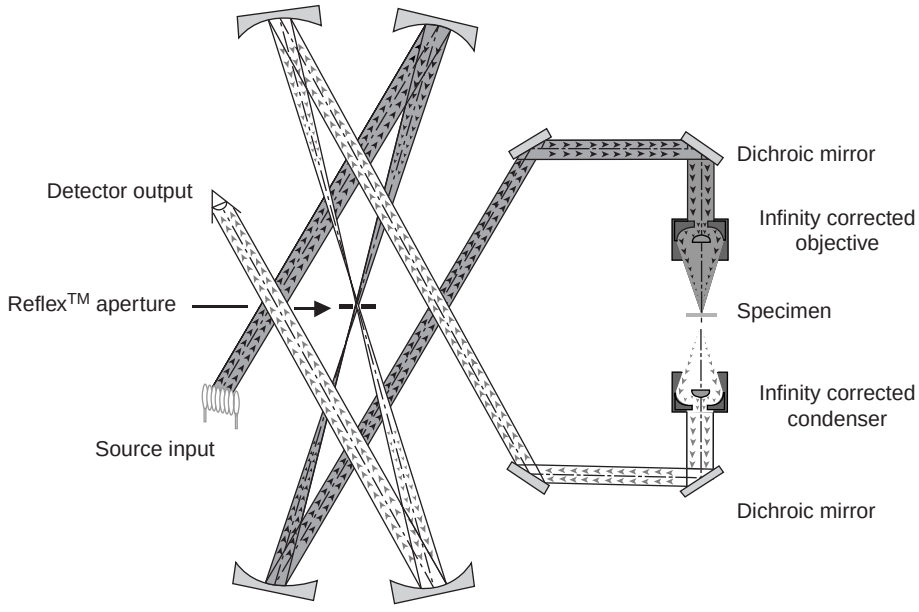

Figure 8 The optical diagram of the Continuμm™ microscope in the transmission mode. (Reproduced by permission of Spectra-Tech, Inc.)

that the latter aperture is more important especially if the specimen plane is flooded with energy or if the specimen is highly scattering.[12,13] The latest Spectra-Tech, Inc. design uses an optical path which results in a single aperture to effectively achieve confocal aperturing as Figure 8 illustrates.

Other microscopes have apertures made of low-E glass which are transparent to visible light, but not IR radiation. Additionally, the Perkin-Elmer AutoIMAGE microscope uses ''whole field of view'' apertures which keep the sample in view when masking an area of interest. The sample area is defined on its visual image using a computer ''growbox''. But during spectral acquisition, motor-driven aperture blades automatically adjust to the dimensions defined by the user.

3.3 Stages

The IR microscope stage needs to accommodate samples, sample supports, and sampling cells of various sizes and shapes. Stages are available to facilitate sample positioning through different amounts of X, Y, and Z travel and can be operated manually or can be motorized for precise, accurate, and easy positioning of samples. Motorized stages, which can be interfaced with the IR spectrometer's data acquisition software via RS-232 communications, are used routinely for mapping the sample. Samples can be stepped as small as 1 μm increments and as precisely as $+/-0.3$ μm. For polarization studies a stage which can provide rotational travel or angular movement is often used. Finally, for the study of samples in controlled temperature environments as well as the study of phase transitions and liquid crystals, a heated/cooled stage is required. These stages incorporate a temperature-, moisture-, and light-resistant window, like diamond, for transmission studies and may provide a range of temperatures from 200 °C to 600 °C.

3.4 Visible Illumination and Viewing Systems

It is a requirement that the IR microscope accessory also function as a visible microscope so that the sample can be found, positioned, and focused. Since the visible microscope is especially important for examining the heterogeneity of a sample, IR microscopes need to allow for the observation of transparent or opaque samples. Consequently, the visible illumination systems must provide a light source below the sample for transmission viewing and a light source above the sample for viewing in reflection. Uniform illumination is critical for maintaining image quality and so Koehler illumination is standard on many microscope accessories. Koehler illumination results when the filament of the light source, a coiled tungsten wire, is imaged into the back focal plane of the objective.[16] The eyepieces for viewing can range from 5× to 20× in magnification and often contain a reticle for the measurement of sample dimensions. Trinocular or quadnocular viewers are available which allow for simultaneous photomicrography or photomicrography and video viewing, respectively.

A wide variety of refracting, transmitting glass objectives are also used for visible viewing or photomicrography, ranging from 4× to 40× in magnification, as well as phase contrast objectives which improve the visibility of colorless samples like certain polymers or unstained biological samples. Microscopes also allow for the positioning of a visible polarizer and analyzer so that various optical properties like refractive indices, angle of extinction, birefringence, and sign of elongation can be measured. Condensers are optically matched to the visible condensers and the type can vary depending on application, such as phase contrast, polarized light, or dark-field microscopy.

3.5 Infrared Sources and Detectors

The IR source used in microspectroscopy is the same as the source on all IR spectrometers, a nichrome wire or a silicon carbide material. Both emit blackbody radiation and operate in the range from 1273 K to 1573 K. However, the standard triglycine sulfate detectors are not used because their sensitivity is too low for microscopic applications. Instead, liquid-nitrogen-cooled semi-conductor detectors are used, the most common being the MCT detector. MCT detectors have different sensitivities, and the greater the sensitivity, the narrower the detection range. For example, a wide-band MCT detector has a range down to $450\,cm^{-1}$ while a narrow-band detector has a range to $580\,cm^{-1}$; however, a wide-band detector has only about 25% of the detectivity. Although MCT detectors are available with sizes ranging from $100\,\mu m \times 100\,\mu m$ up to $2\,mm \times 2\,mm$, the $250\,\mu m \times 250\,\mu m$ is standard equipment for a microscope accessory. Most detectors are housed in the microscope assembly itself.

FPA detectors are discussed in section 6.2.

3.6 Purge

Atmospheric carbon dioxide ($2350\,cm^{-1}$ and $670\,cm^{-1}$) and water ($3400\,cm^{-1}$ and $1640\,cm^{-1}$) absorptions are often present in spectra obtained through an IR microscope. Any change in the amount of these atmospheric gases from the time the background is obtained to the time the sample scan is obtained will result in positive or negative absorptions in the spectrum. An adequate purge must be provided for the IR microscope accessory if a spectrum free from atmospheric interference is to be obtained. A purge can be provided for by enclosing the microscope in plexiglass or by a series of devices: purge enclosures for seal areas below the stage, a bellows to connect the condenser to the aperture location, purge skirt for the objective, and a flexible baffle for the top of the condenser. Short optical paths are also helpful in maintaining a good purge.

However, traditional purge problems may become negligible with the availability of software which easily and accurately compensates for various amounts of unratioed atmospheric absorptions in the spectrum. Perkin-Elmer's Atmospheric Correction software, for example, models the absorption bands from water vapor and carbon dioxide and removes them via a mathematical treatment which not only takes into account the amount of water or carbon dioxide, but subtle band shifts which may arise.

3.7 Automation

At present, most "standard" IR microscopes are fully automated. For example, aperturing is performed interactively with motor-driven aperture blades and focusing achieved with the one touch of a computer button. There is automatic adjustment of microscope configuration between visible viewing and the IR collection mode, as well as from the IR reflection mode to the IR transmission mode. The degree of visible illumination, the position of the Cassegrainian condenser, and the visible magnification can be adjusted through computer control rather than performed manually at the microscope. Furthermore, the motorized stage has become standard equipment and is often used as a multi-sampling device in addition to mapping. Several free-standing samples are placed on the stage, their coordinates programmed into the software, and the microscope is left to find, focus, and record spectra on the individual samples.

4 BASIC TRANSMISSION TECHNIQUES AND PRINCIPLES

A common experiment performed in IR spectroscopy is the transmission experiment, the measurement of the IR radiation which has been transmitted through a sample. This experiment is performed using the IR microscope as the ray traces in Figure 1 and Figure 8 illustrate. The success of the transmission experiment, that is, the quality of the resulting spectrum, is very dependent upon sample preparation. This is because, by virtue of the microscopic sample's size, it is essentially an optical element of the microscope.

4.1 Sampling Considerations

The ideal sample for transmission microspectroscopy is one that does little to disrupt the optical path of the IR radiation. The requirement for such a sample would be an appropriate path length so as not to produce totally absorbing bands in the spectrum, a large enough diameter so that diffraction effects are minimized, and a smooth

contour so that no scattering takes place. Consequently, any preparation of the sample serves to address one or more of this issues.[9]

In macroscopic sampling, we rarely run a transmission spectrum on an undiluted sample with the exception of a thin film. The same consideration must be given to most microscopic samples; they are often too thick to obtain high-quality transmission IR spectra without reducing their path length. Using macroscopic methods, for example, any bands which possess total absorption that result from a thick sample would be readily apparent in the spectra since their transmittance is 0%. When using an IR microscope, however, total absorption may be offset by stray light effects as the spectrum of the hair fiber in Figure 3 shows. On other samples, the effects may be more subtle, and the broad, undefined bands which result may interfere with sample characterization. If one considers that, typically, a transmittance spectrum obtained on a 10-μm thick organic sample needs little or no ordinate expansion, then samples thicker than this will require dilution or path length reduction.[5]

Reducing the path length of a microscopic solid sample can be performed by flattening the sample with a probe tip or stainless steel bearing mounted on an axle handle (called a roller) or by placing the sample in a specially designed pressure cell with IR transparent windows. A common cell used for this is the low-pressure diamond anvil cell whose diamond absorptions occur at 1900–2100 cm^{-1}, a region where few other materials absorb.[18] In fact, IR microscopists have found it convenient to use for flattening most samples because the cell contains the sample while flattening. Often several samples are placed on the diamond cell, flattened, and then the top diamond is removed after which the spectrum is recorded on each subsequent sample. If used in this fashion, the condenser of the microscope must be adjusted to compensate for the high refractive index of the diamond as discussed in section 3.1. If the sample is fairly elastic, pressure must be maintained to keep an adequate path length, so the spectrum must be recorded while the sample is held in place in the pressure cell. In this case, both the condenser and the objective must be adjusted so that signal is not lost owing to refraction effects. When crystalline orientation or morphology need to remain unaltered, slicing the sample to achieve path length reduction may be a better alternative. Although this can be performed by hand using a razor blade, diamond knife, or other sharp device, a microtome can be used to obtain sections of reproducible thickness.[16]

Path-length concerns in the NIR region, however, are less of a problem. Since most NIR bands are a result of overtones or combinations, they are typically much weaker than their corresponding fundamental transitions, and NIR spectra can be recorded on undiluted or unaltered samples as thick as about 0.5 mm.[19,20]

Fortunately, the sampling technique of flattening to reduce path length will also help to minimize diffraction occurring at the sample's edge as well as at the apertures which are used to define the sample. Figure 9 which compares the spectrum of the cellulose acetate *film* (Figure 9b) with the spectrum of a 15 μm × 15 μm *slice* of cellulose acetate (Figure 9a) illustrates the effect of sample diffraction.[21] Both samples are the same thickness, and their spectra were obtained with an identical single remote aperture as previously discussed in section 2.1. Since the film is much larger in area than the area defined by the aperture, the effects from diffraction around the film's edges are not observed. When the spectrum of the film is compared to the spectrum of the slice, one observes that the carbonyl stretch at about 1750 cm^{-1} has a peak intensity of 28% T versus 36% T, and the C–O stretch at about 1225 cm^{-1} has an intensity of 18% T versus 38% T, respectively. IR radiation which has bent around the slice of cellulose acetate has reached the detector, and this spurious radiation is worse at longer wavelengths in accordance with the diffraction theory described in Equation (1). Varying changes in the *relative* intensities are also a result of diffraction. Compare the relative intensity of the methyl bend at 1375 cm^{-1} with the C–O stretch at 1225 cm^{-1}. The relative intensity is about the same in the spectrum of the cellulose acetate slice, while the 1225 cm^{-1} band is much more intense in the spectrum of the film. The longer wavelength band (1225 cm^{-1}) is bent around the sample more in the spectrum of the slice owing to diffraction, and the result is an increase in % T. Other studies clearly show that band intensity ratios change significantly as sample size gets smaller. However, distortions in band *shapes* can result from effects other than diffraction. The asymmetric bands observed in the middle and bottom spectra of Figure 2, for instance, appear to be the result of specular reflections off the side of the microscopic sample.[14]

As discussed in section 2.1, diffraction can occur at the aperture edges or the sample edges. One effect of diffraction around the aperture is a baseline distortion with a loss of transmission at the longer wavelengths.[22] Figure 10 shows that as the aperture becomes smaller, more energy is lost.[23] However, this baseline slope can be ratioed out of the spectrum by using the same aperture settings in the acquisition of the background spectrum as used in the sample spectrum. This is achieved by closing the aperture(s) down around the image of the sample particulate or area of interest, acquiring a single-beam spectrum of the sample, I, then moving the stage to a vacant area of the IR window without resetting the apertures. The single-beam spectrum of the window is acquired, I_0, and the two spectra are ratioed, I/I_0.

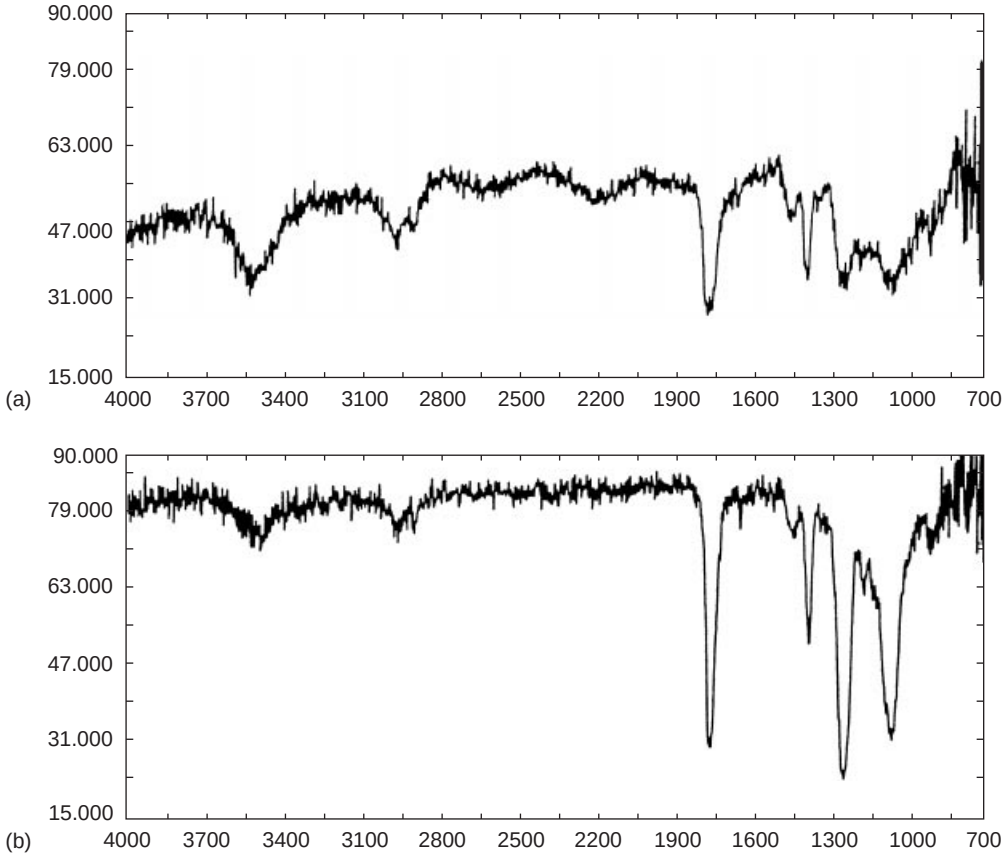

Figure 9 (a) IR spectrum of a 15 μm by 15 μm cellulose acetate slice, and (b) a cellulose acetate film. (Reproduced by permission of Plenum Press from P. Lang, A.J. Sommer, J.E. Katon, 'Infrared and Raman Microspectroscopy: An Overview of Their Use in the Identification of Microscopic Particulates', in: *Particles on Surfaces 2, Detection, Adhesion and Removal*, ed. K.L. Mittal, Plenum Press, New York, 143–161, 1988.)

Another sampling consideration familiar to IR spectroscopists is that caused by the scattering of IR radiation. An opaque, finely powdered sample or one which is rough on a microscopic scale will scatter the light, and the scattering is proportional to the square of the difference in the refractive index of the sample and the surrounding medium.[24] Since the scattering is more marked at shorter wavelengths, the result is that the spectrum of the sample exhibits a sloping baseline. The common correction for macro samples is to suspend them in a medium which has a similar refractive index and coalescing the two into a pellet. KBr, KI, CsBr, CsI, for example, are commonly used media. In microspectroscopy such a ''pellet'' may be formed by placing one or two salt crystals in the sample and pressing down with a pointed probe while rolling it and applying pressure. If the probe has an area of 25 μm × 25 μm, 0.1 pounds of force on the handle becomes translated into about 100 000 lbs in^{-2} at the probe tip, well enough for an adequate pellet.[25] In addition, the sample area is increased along with the concomitant increase in aperture size, thereby minimizing diffraction effects; the signal is increased with a resulting increase in S/N; and the sample is diluted, and thus total absorptions are less likely. Other investigators have found that scattering by opaque particles can be reduced by pressing them directly into KBr windows or suspending them in Nujol.[26,16]

However, some workers still think the use of the diamond anvil superior. The sample preparation is simpler, and problems related to baseline slope from inadequately ground KBr are omitted, in addition to water adsorption on KBr which might prevent accurate hydroxyl group assignments.[27]

4.2 Detection Limit

The eye can detect objects which are below the diffraction limit of visible light, although it may not resolve two objects which are close together.[7] Therefore, diffraction does not necessarily limit the detection of a compound in IR microspectroscopy. However, as sample size

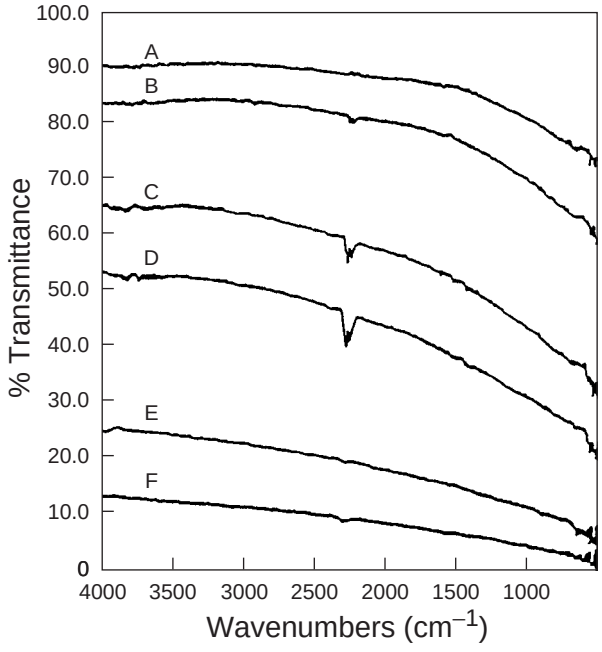

Figure 10 The effect of diffraction caused by small apertures. Spectra A to F correspond to spectra obtained using 29, 25, 21, 17, 12.5, and 8.3 mm (actual) diameter apertures, respectively. (Reproduced by permission of Royal Society of Chemistry from P.H. Turner, *Anal. Proc.*, **23**, 268, 1986.)

approaches the diffraction limit, the quality of the spectrum decreases owing to the loss of energy, and it may be so severe that the absorptions cannot be distinguished from the baseline.

In terms of the detection of small amounts, it depends on the sample's shape, diameter, thickness, and absorption coefficient. Katon has obtained a high-quality transmittance spectrum on an approximate 6-ng particle of isotactic polypropylene, yet the sample was relatively smooth, approximately 10 μm thick and 25 μm wide. The thickness was optimum for a transmission experiment without any ordinate expansion, and the sample was large enough (and consequently the apertures) to not cause any significant diffraction effects.[25]

More impressive is the report of the absorbance IR spectrum of a single, 7-μm diameter red blood cell by Dong and co-workers.[28] Assuming the red cell has a cylindrical shape with a height of 5 μm (probably an overestimation, since red cells are characteristically concave) and an average density of 1.054 g mL^{-1},[29] the experiment indicates a detection limit (DL) of 0.20 ng of material.

Other work has shown that transmission IR microspectroscopy is suitable for the detection of 3.6-nm to 36-nm *thin* organic layers on chalcogenide fibers.[30,31] However, the authors have taken advantage of a sensitivity enhancement which results from the geometrical effects of the cylindrical fiber which serve to concentrate the transmitted radiation from the organic layer and thereby improve the S/N.

5 BASIC REFLECTION TECHNIQUES AND PRINCIPLES

In an IR microscope's reflection collection mode, a split mirror (beamsplitter) is placed in the IR beam to allow 50% of the incident light to be delivered to the objective and focused onto the sample. The light reflected from the sample is collected by the same objective and reflected by the split mirror onto the detector. This is illustrated in Figure 5 and more explicitly illustrated in Figure 11. Depending on the nature of the sample, the analyst can perform three basic experiments when the microscope is in the reflection collection mode: specular reflection, diffuse reflection, and absorption-reflection. Specialized objectives can be installed to perform other reflection experiments, such as, ATR and grazing angle reflection.

5.1 Specular Reflection

Specular reflection occurs when the IR radiation is reflected from the front surface of an optically thick dielectric sample, and the angle of the reflected light is equivalent to the angle of incident light as Figure 12 shows. Although radiation could also be refracted and transmitted, this component is not collected. Back surface reflection could also take place, but if the sample is of the proper thickness, this component will not contribute significantly. In specular reflectance Fresnel's relations apply to the process as Equations (5) and (6) describe:

$$r = \left| \frac{\hat{\eta} - 1}{\hat{\eta} + 1} \right|^2 \tag{5}$$

$$\hat{\eta} = n + ik \tag{6}$$

where r is the reflection coefficient, n is the refractive index, $\hat{\eta}$ is the complex refractive index and k is the extinction coefficient.[32] Since the specular reflectance spectrum is the plot of r versus frequency, and since the reflection coefficient follows the complex refractive index, the specular reflectance spectrum will feature discontinuity in the region of the absorption bands. These first-derivative-like spectral features, or restrahlen bands, can be observed in the top spectrum in Figure 13, a spectrum of a Nylon tie (used to fasten tubing to laboratory glassware).

A mathematical relationship known as the Kramers–Kronig integration can transform specular reflectance spectra into absorbance spectra by calculating k, n, and the phase angle shift, θ (ν).[33,34] The conventional IR absorption spectrum is then the dispersion of k, a plot

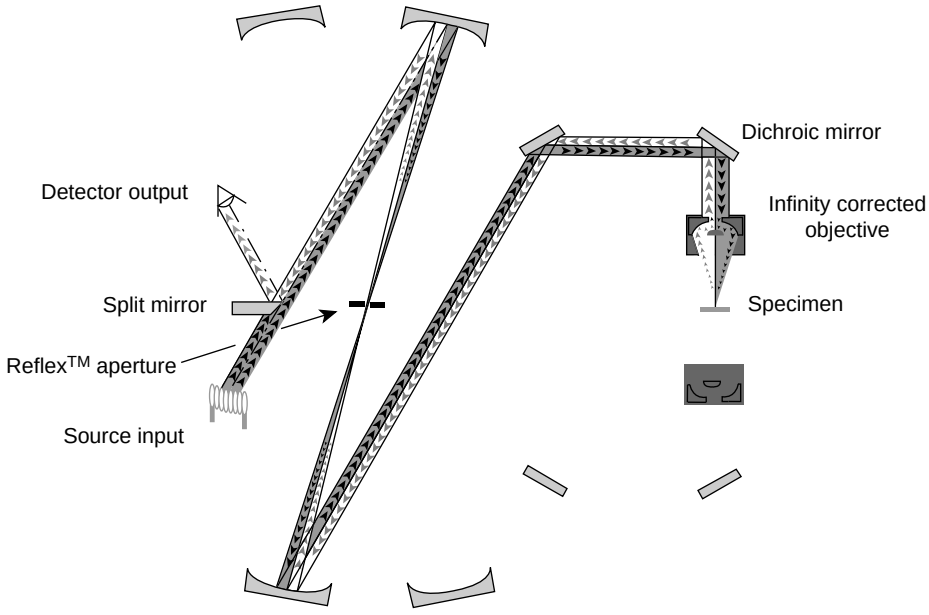

Figure 11 The optical diagram of the Continuμm™ microscope in the reflection mode. (Reproduced by permission of Spectra-Tech, Inc.)

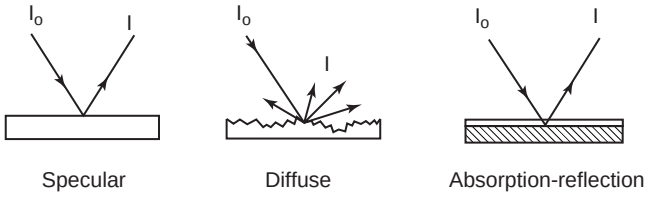

Figure 12 Illustration of specular reflection, diffuse reflection, and absorption-reflection.

of k versus frequency. The bottom spectrum in Figure 13 shows the transformed spectrum of the Nylon tie. The large negative feature at about 3600 cm^{-1} is an anomaly and indicates that the sample was not a perfect specular reflector. In fact, although the sample was optically thick, it was not completely smooth, and one could observe a visible grain in the sample which effected diffuse, random reflections. Although not in this spectrum, another common anomaly that might occur is when unratioed carbon dioxide or water vapor absorptions remain in the specular reflectance spectrum. Those absorption bands will be distorted after the KK transformation since the mathematical treatment presumes all bands result from specular reflection.

Sample preparation for this type of experiment is minimal. A sample is usually placed on top of a microscope slide and placed on the stage of the microscope accessory. The sample area is then defined by the use of one or more remote apertures, and the spectrum acquired. The background spectrum is obtained on a smooth, highly reflective, nonabsorbing surface such as a front-coated gold mirror using the same aperture setting. (This was how the spectra shown in Figure 13 were acquired.) Another sampling method suitable for the study of minerals is to embed the sample in an epoxy resin and polish the surface.[35] Regardless, specular reflection is a low-efficiency process, and therefore the largest sample area possible should be used so that an adequate S/N is obtained.

5.2 Diffuse Reflection

When an absorbing sample is dispersed in a nonabsorbing media, the incident radiation is transmitted through sample particulates and is diffusely reflected off the nonabsorbing particulates, such as ground KBr, as illustrated in Figure 12. This is the diffuse reflection experiment, and it can also be performed with a standard Schartzschild lens in the reflection collection mode. The diffuse reflectance spectrum, R'_∞, is defined as the ratio of the single-beam diffuse reflectance spectrum of the sample to that of a nonabsorbing standard. The Kubelka–Munk theory indicates a linear relationship between the diffuse reflection function, $f(R'_\infty)$ and the sample concentration, c, according to Equations (7) and (8):

$$R'_\infty = \frac{R'_\infty(\text{sample})}{R'_\infty(\text{std})} \tag{7}$$

$$f(R'_\infty) = \frac{(1 - R'_\infty)^2}{2R'_\infty} = \frac{c}{k} \tag{8}$$

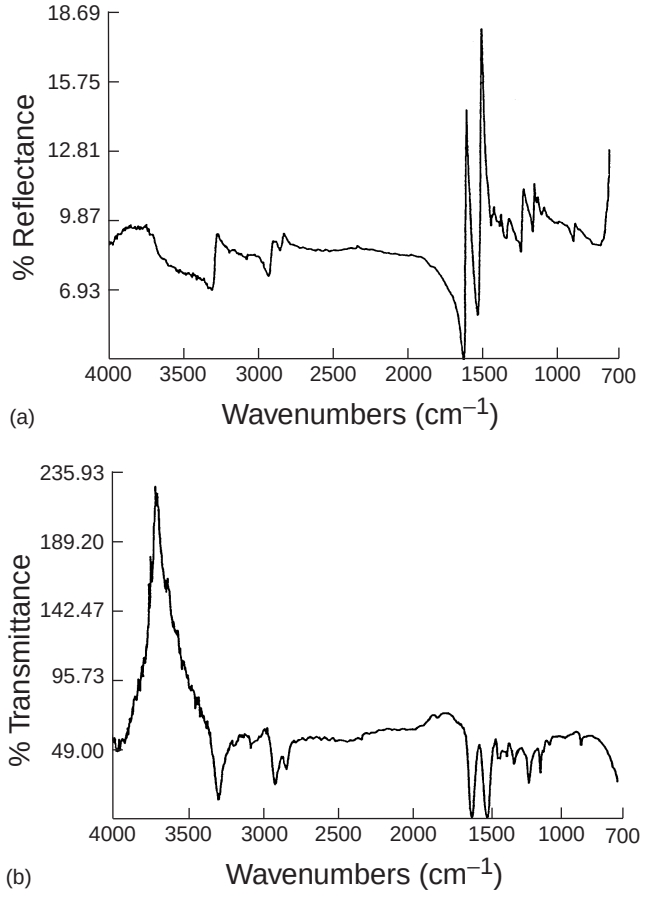

Figure 13 (a) The specular reflection spectrum of a nylon tie; (b) the same spectrum after Kramers–Konig transform. (Reproduced with permission from P.L. Lang, L.J. Richwine, 'The Versatile Sampling Methods of Infrared Microspectroscopy', in: *Practical Sampling Techniques for Infrared Analysis*, ed. P.B. Coleman, 146–162, CRC Press, Boca Raton, FL, © 1993.)

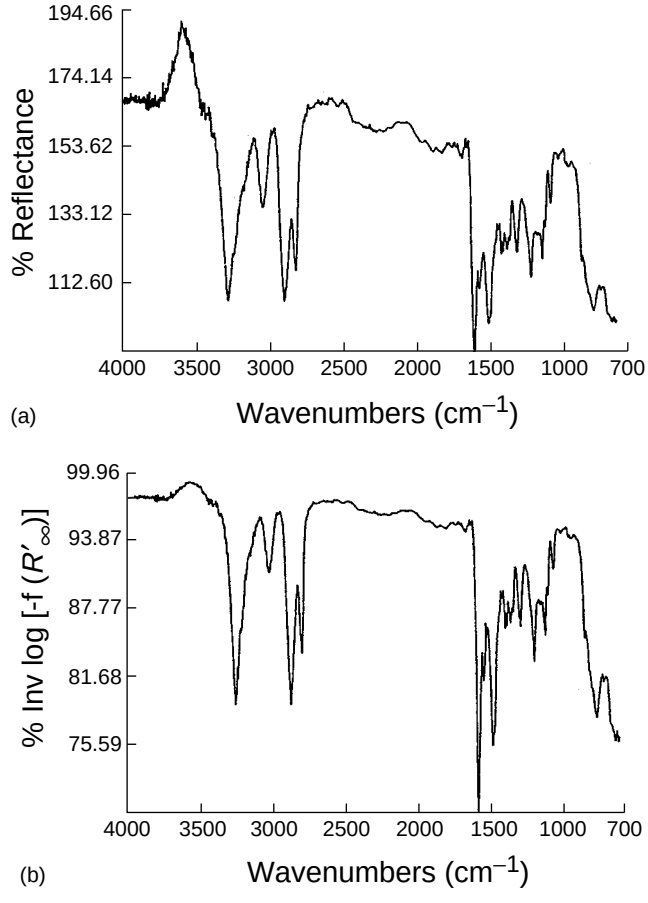

Figure 14 (a) The diffuse reflection spectrum of a nylon tie; (b) the Kubelka–Munk spectrum. (Reproduced with permission from P.L. Lang, L.J. Richwine, 'The Versatile Sampling Methods of Infrared Microspectroscopy', in: *Practical Sampling Techniques for Infrared Analysis*, ed. P.B. Coleman, 146–162, CRC Press, Boca Raton, FL, © 1993.)

where $k = s/(2.30\varepsilon)$, s is the scattering coefficient, and ε is the molar absorptivity. Therefore, one would expect the spectrum, $f(R'_\infty)$ to be similar to an absorbance spectrum as Figure 14(a) shows. After the calculation of the Kubelka–Munk spectrum shown in Figure 14(b), however, the corrected relative intensities are more similar to that of an absorbance spectrum.

Diffuse reflection can vary a great deal with particle size. To minimize this effect, uniform KCl particles which range between 75 μm and 90 μm are recommended because they show little variation in reflectance over the mid-IR region from 2 μm to 20 μm. Another caveat is that if the sample is too concentrated or undiluted, specular reflections may greatly contribute to the spectrum, resulting in the restrahlen features discussed in the previous section. Finally, since microscope optics are not maximized for diffuse reflectance experiments, one needs to obtain the spectrum using the widest apertures possible, in order to maximize energy and consequently S/N.

A typical sample preparation for the microscope accessory is to roughen the surface of the sample with a fine grit emery paper or diamond abrasive paper with the appropriate particle size. The spectrum of the ground sample left on the paper is obtained and ratioed against a spectrum of unused emery paper at the same aperture settings. (This was how the spectra shown in Figure 14 were obtained.) Alternatively, the roughened surface of the actual sample could be analyzed. Diffuse reflectance spectra of dyes separated by thin-layer chromatography have been obtained using the IR microscope.[36]

5.3 Absorption/Reflection

If a sample of the proper thickness is in intimate contact with a reflective surface, a absorption-reflection experiment can be performed when the microscope is in

the reflection collection mode. As illustrated in Figure 12, the incident light passes through the sample, is reflected off the surface, and passes back through the sample again before collection. Since absorption-reflection is approximately a double-pass transmission experiment, the maximum sample thickness is estimated at about half that for the sample measured in transmission. Given that the Schwarzschild objective delivers light at an average incident angle of 25° from the normal, the sample thickness which yields a spectrum with little or no ordinate expansion is estimated at 4.5 μm according to Equation (9):

$$2d = 10 \cos \alpha \tag{9}$$

where α is the angle of incidence and d is the thickness of the film.[5] Although light also reflects at the front surface of the sample, this contribution to the overall reflection will vary depending upon the sample's properties. If the contribution is large, restrahlen bands will be observed. Ideal samples for absorption-reflection experiments are also relatively smooth and flat, and typically tend to be coatings, films, or film-like deposits on metal.

Special objectives can be added to the IR microscope which will deliver the incident light at grazing angles, typically at a range from 65° to 85° as Figure 15 shows. According to Equation (9), measurement at a grazing angle can allow for the analysis of extremely thin films on reflective surfaces of 1 μm or less. The microscope objective offers the capability to obtain spectra on samples microscopic in dimensions in which the nominal area is approximately 50 μm × 50 μm. The Spectra-Tech grazing angle objective, for example, has 30× magnification, a NA which ranges from 0.996 at 85° to 0.51 at 33°, a working distance of 1 mm, a viewing mode, and an option for an IR polarizer. It is important to note that Pepper has found that the grazing angle objective is suitable for quantitative measurements; that is, there is effectively no angular dependence on the incident light intensity.[37]

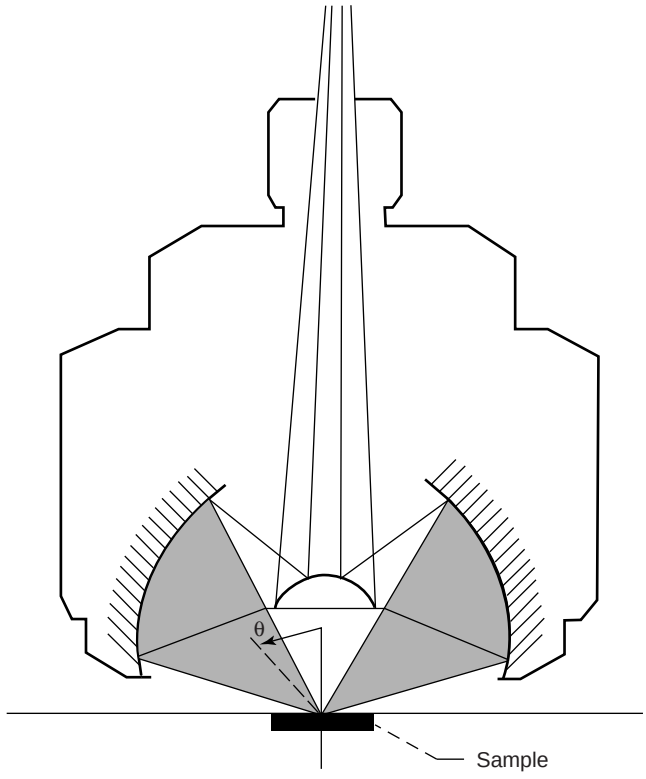

Figure 15 A cross-section of a grazing angle objective. (Reproduced by permission of the Society for Applied Spectroscopy from S.V. Pepper, 'Characterization and Application of a Grazing Angle Objective for Quantitative Infrared Reflection Microspectroscopy', *Appl. Spectrosc.*, **49**(3), 354–360 (1995).)

5.4 Attenuated Total Reflection

A more widely applicable technique for IR microspectroscopy is ATR. To perform this experiment, a special objective can be added to the IR microscope which contains an internal reflection element of relatively high refractive index as Figure 16 shows.

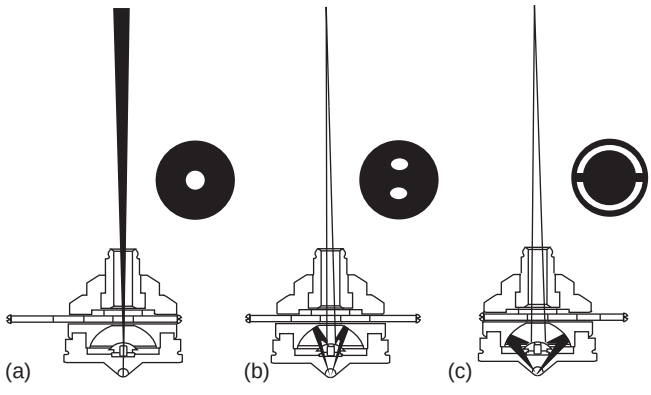

Figure 16 Cross-section of an ATR objective in the viewing, contact, and IR collection mode. (Reproduced by permission of Spectra-Tech, Inc.)

When the IR radiation is introduced into the element, it can be total internally reflected if the incident angle is greater than the critical angle θ_c according to Snell's law described in Equation (10):

$$\theta_c = \sin^{-1} \left(\frac{n_1}{n_2} \right. \tag{10}$$

where n_2 is the refractive index of the element, and n_1 is that of medium with which it is in contact. Although complete internal reflection occurs at the interface between the two media, radiation does in fact penetrate into the rarer medium, and a sample placed in contact with the element may absorb this radiation at the point(s)

of reflection. In contrast to specular reflection, ATR is a very efficient technique with virtually no loss of energy except from absorption by the medium. The standing wave established at the point of reflection is called the evanescent wave and is described by Equation (11):

$$d_p = \frac{\lambda}{2\pi n_e(\sin^2\theta - n_{se}^2)^{1/2}} \quad (11)$$

d_p is the depth of penetration, λ is the wavelength, θ is the angle of incidence, n_e is the refractive index of the element, and n_{se} is the ratio of the refractive index of the sample to that of the element. Thus greater penetration depths are achieved for a particular angle of incidence by altering the internal reflection element. Internal reflection elements such as zinc selenide ($n_e = 2.4$), offer a greater penetration depth than germanium ($n_e = 4.0$), for example. At incident angles of approximately 45° and assuming 1.5 for the sample's refractive index, this corresponds to a depth of 0.2 λ and 0.066 λ, respectively, and offers the possibility to obtain spectra at the surface of a sample. However, if the sample is multilayered or lies on an absorbing substrate, the sample needs to be at least as thick as the depth of penetration, otherwise the medium under the sample will be detected. If a sample is completely transferred to the element and is thinner than the penetration depth, the spectrum of the entire sample will be obtained, not just the surface.

The ATR objective with magnification of 25× allows for sampling the surface of a microscopic sample with a nominal area of 50 μm by 50 μm. The small area of the microscope ATR element in combination with the single bounce reflection has the advantage that complete contact of the sample with the element is more easily achieved. Consequently, baseline distortions associated with the uneven, lower pressures which result from using macroscopic ATR elements are eliminated.[38] In addition, Reffner and Martoglio report another advantage.[16] In the Spectra-Tech ATR objective, which has a hemispherical shaped element, the cross-sectional area of the radiation passing through it is magnified by the refractive index of the element. Using a germanium element, for instance, a 2.50 mm diameter circular aperture in the image plane corresponds to an apparent sample diameter of 100 μm but is an actual sample diameter of 25 μm. Thus, for a germanium ATR microelement, an aperture four times larger than what is usually required for a sample can be used, reducing the diffraction problem at the aperture considerably. This magnification factor associated with micro ATR elements has been verified for samples 60 μm in diameter and larger.[38]

Another theoretical consideration is that since the sample is immersed in the internal reflection element, the spatial resolution should be increased according to Rayleigh's criterion given in Equation (1). This is because the refractive index, n_1, of the medium is increased, and therefore NA is increased according to Equation (2). However, since only half of the beam is introduced into the objective and collected by the objective when the microscope is in the reflection collection mode, the spatial resolution using an ATR element will theoretically only increase by $n_1/2$.[38]

Sample preparation is typically nonexistent using the ATR objective. The sample is brought into view and positioned using a refracting lens which is built into the middle of the ATR objective as Figure 16(a) shows. Then with the slide selector in the intermediate position, Figure 16(b), the stage is adjusted to bring the sample in contact with the internal reflection element by either visual observation or by using an automated pressure sensor. After defining the sample area to be analyzed using a remote aperture, IR radiation is then directed into the internal reflection element as Figure 16(c) illustrates, and a sample spectrum is obtained. The sample spectrum is ratioed to a background spectrum obtained while the objective is not in contact with the sample. Since spectra can be obtained so easily using a micro ATR objective, and because the ATR spectra have equivalent S/Ns as those obtained in transmission on samples of the same size, the use of ATR microspectroscopy has become more widespread.[38]

6 INFRARED IMAGING SYSTEMS

6.1 Traditional Mapping

Using a motorized stage, IR imaging can be performed with a standard MCT detector and microscope accessory in either the reflection or transmission collection mode. IR spectra can be obtained for a contiguous sample area by stepping the stage (a computer-controlled operation) in small increments which are the same as or less than the effective aperture size and recording the spectrum at each step. The corresponding mapping software then plots the intensity of an IR absorption band or absorption band ratio as a function of x-y coordinates, displaying the images as wire frame or contour plots as Figure 17 illustrates.[11] In this fashion, a particular functional group's concentration is mapped across a heterogeneous sample. One can also opt to collect spectra in a straight line across a sample as Figure 18 illustrates.[39]

The main problem associated with mapping is the long collection time. If a high-quality spectrum takes 10 s to obtain at each stage increment, then a map formed from a 64 μm × 64 μm spectral grid (4096 spectra) will take approximately 11 h to obtain. This could correspond to a 225 μm × 225 μm sample stepped across in increments of 15 μm, for example. Another problem, mentioned in section 2.1, is that owing to diffraction at the aperture(s)

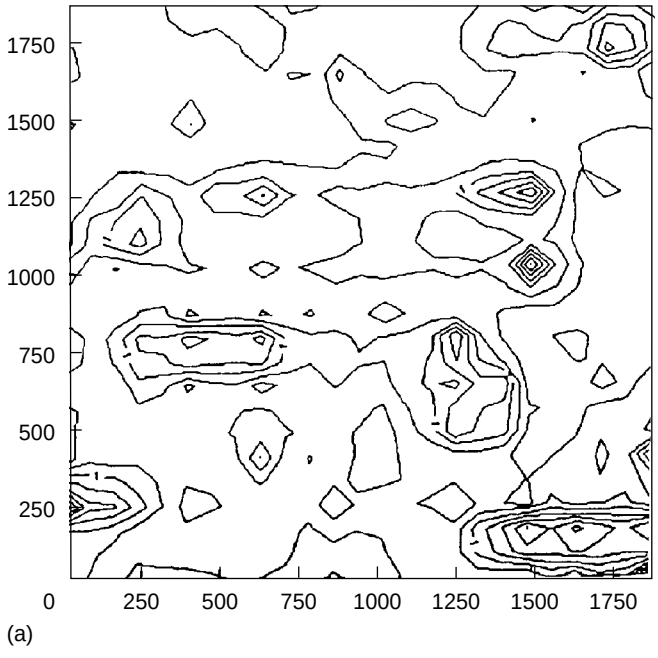

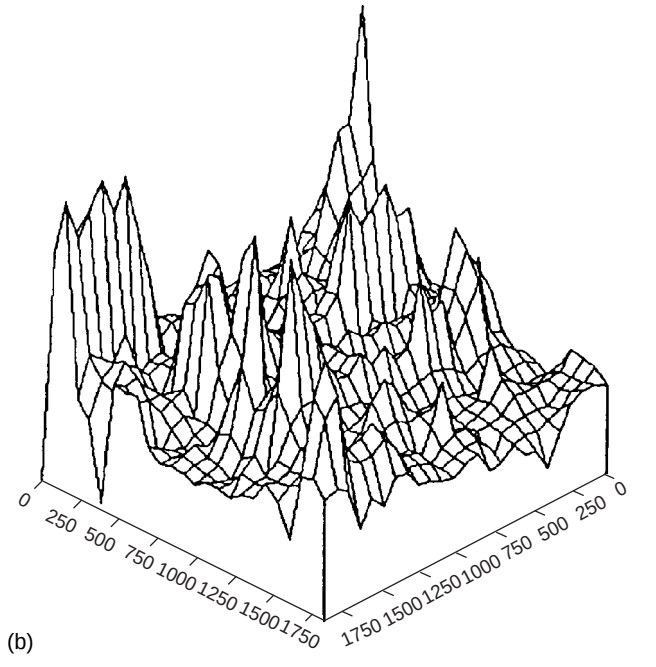

Figure 17 Functional group images of a 2 mm × 2 mm serpentine sample. (a) Contour plot constructed from a Si–O stretching absorption and (b) wire-frame plot constructed from a C–O stretching absorption. (Reproduced by permission of Marcel Dekker, Inc. from K. Krishnan, J.R. Powell, S.L. Hill, 'Infrared Microimaging', in: *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, Inc., New York, 85–110, 1995.)

edge, there will spectral contamination from surrounding areas in each spectrum. However, it has been suggested that if the step size is significantly smaller than the

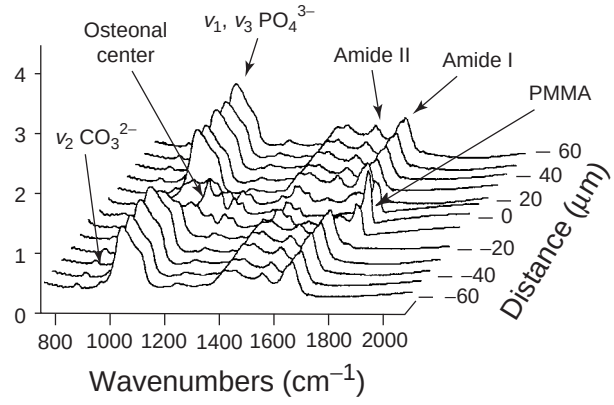

Figure 18 A line map of spectra obtained at 10 μm steps across a human osteonal bone sample. The spectra show that the mineral density increased as a function of increasing distance from the osteonal center ($z = 0$ μm). The spectrum was obtained by transmission on a 5-μm cross-section of a bone sample which had been embedded in polymethyl methacrylate. (Reproduced by permission of Springer-Verlag from E.P. Paschalis, E. DiCarlo, F. Betts, P. Sherman, R. Mendelsohn, A.L. Boskey, *Calcified Tissue Intl.*, **59**, 480–487 (1996).)

effective aperture size (for example, 3 μm versus 15 μm, respectively) that the effective spatial resolution can be increased if chemometric methods are employed in conjunction with mapping. Harthcock has shown that factor analysis of the data can reveal the number of spectrally different layers of a polymer laminate, even when the width of one or more polymer layers is below the diffraction limit, providing that linear additivity holds and experimental noise is within reasonable limits.[40] However, sampling will take considerably longer under these conditions.

6.2 Focal Plane Array Detectors

Spatial and spectral information can now be collected simultaneously using a step-scan interferometer coupled with an IR microscope and a FPA detector.[41] The FPA detectors consist of a grid of very small IR detectors each of which can detect radiation from different locations of the sample. Thus, at each detector element a NIR or mid-IR spectrum is obtained for each region of the sample. One can either observe the spectrum at a particular pixel which corresponds to a particular sampling region or observe the intensity of a particular absorption band for each pixel as Figure 19 shows.[42]

FPA detectors can be fabricated from a small selection of materials including platinum silicide (1.0–5.7 μm), palladium silicide (1.0–3.2 μm), iridium silicide (1.0–8.2 μm), indium antimonide (1.0–5.5 μm), indium gallium arsenide/indium phosphorus (1.0–1.7 μm), germanium (1.0–1.6 μm) and MCT (2.5–12 μm).[15] Digilab's Stingray FPA detectors, for example, are constructed of either

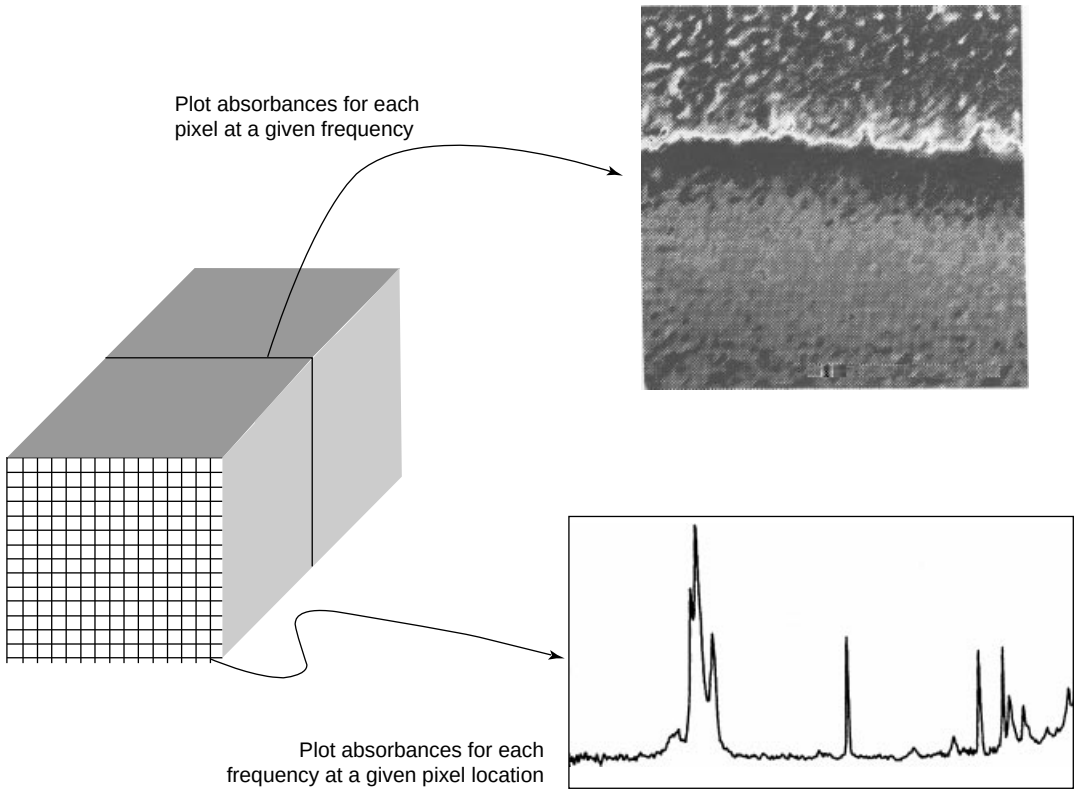

Figure 19 Illustration of a ''stack'' of FPA data. (Reproduced by permission of Spectroscopy from J.L. Koenig, C.M. Snively, 'Fast FTIR Imaging: Theory and Applications', *Spectrosc.*, **13**(11), 22–28 (1998).)

a 128×128 InSb element grid or a 64×64 MCT element grid for work in the NIR region or mid-IR region, respectively. Past instrumentation has been designed using an FPA detector configured with an acousto-optical tunable filter (AOTF)[43] or a tunable bandpass dielectric filter[44] which allows for selection of a wavelength which corresponds to a specific absorption band by changing the spectral bandpass. The imaging process occurs by selecting a wavelength, collecting a spectrum, and then repeating the process. The AOTF and dielectric filters are limited to the very near IR region ($10\,000–5000\,cm^{-1}$) and the NIR region ($4000–2300\,cm^{-1}$) respectively.

More recently, FPA detectors have been used with a step-scan interferometer since as many as 16 000 outputs must be read simultaneously at each retardation interval of the interferometer.[45,46,47] Standard rapid-scanning interferometers which measure the output from a single detector, scan at speeds ranging from millimeters per second to centimeters per second. Since this is too fast to obtain acceptable S/N from the less-sensitive FPA detectors, radiation must be collected over a longer time period. Consequently, at each step of the interferometer, several image frames are obtained and co-added as needed to obtain adequate S/N. Although the time required to Fourier transform and ratio to background is several minutes of processor time, the time required for the collection of 4096 spectra ($16\,cm^{-1}$ resolution, 150 co-added image frames during each 1 second interferometer step, 255 steps, 1 scan) is less than five minutes.[46]

Other than a marked decrease in the collection time of the image, the use of FPA detectors can, in theory, increase the spatial-resolution capabilities of the IR microscope. Since IR spectra are collected across an unapertured sample, there is no diffraction by the aperture. The size and number of individual detector elements in conjunction with the optical configuration define the spatial resolution.[42] For example, if a sample of $500\,\mu m \times 500\,\mu m$ dimensions is focused onto a 64×64 detector, each pixel collects information from a $7.8\,\mu m$ square region of the sample.

Although the image quality and speed are superior using FPA detection, the lower sensitivity of the FPA detectors, often results in individual spectra with lower S/N ratios than spectra obtained mapping with a single-point MCT detector. Furthermore, although it is possible to obtain spectra at spectral resolutions of $4\,cm^{-1}$ or higher using FPA detection, it is currently not practical owing to the difficulty of handling large data sets generated at higher spectral resolutions.[41]

7 PERSPECTIVE AND FUTURE DEVELOPMENTS

7.1 Spatial Resolution Below the Diffraction Limit

The relatively recent coupling of an IR microspectrometer with the IR radiation emission from the National Synchrotron Light Source has shown a significant increase in spatial resolution and S/N performance.[48,49] This result is due to the source being highly collimated and small, 200 μm by 400 μm; it is 100 to 1000 times brighter than from a 1500 K thermal emission source. Experiments have shown that using 6 μm × 6 μm dual remote apertures, the synchrotron system clearly resolved 4 μm layers in a cross-section of a multilayered laminate and provided a S/N greater than 200 : 1.[50]

Spatial resolution can also be improved through the use of pseudo near-field techniques where a small aperture is placed in contact, or nearly in contact, with the sample.[51,52] IR mapping can performed if an aperture turret is constructed to allow for x-y motorized stage movement.[52] In both cases, the aperture effectively prevents the light from spreading into a diffraction pattern. Although the technique substantially enhances the spatial resolution, it generally requires more alignment and increased data-collection times. A related, more elaborate technique is near-field optical microscopy (NFOM) where the evanescent field created by total internal reflection, the near-field, is modulated by the sub-wavelength details of a sample's surface. The modulated field is then detected by the tip of a tapered glass fiber, and an image is constructed when the tip scans the surface by a stepping device. The image is dependent upon both the topography and the sample's absorption properties. The NFOM is an important spectroscopic technique, since theoretically the size of the fiber's tip can control the lateral spatial resolution. Piednoir and Creuzet describe a photon scanning tunneling microscope adapted to perform NFOM in the IR range.[53] Their experiments suggest that with improvements in the fabrication of the probe, an increase in sensitivity can be achieved along with the concomitant improvement in spatial resolution.

7.2 The Eighth Age of Infrared Spectroscopy

In the 1973 editorial, 'The Seven Ages of an Analytical Method,' Herbert Laitinen, Editor of *Analytical Chemistry*, wrote that IR spectroscopy had reached its seventh age.[54] At this stage of analytical development, according to Laitinen, a technique was at the end of its useful life and was gradually being replaced by better analytical methods. But much to the delight of IR spectroscopists, nothing could have been further from the truth. In fact, shortly after the publication of the editorial, IR spectroscopy began to experience an age of explosive growth as a result of the development of the FTIR spectrometer. Noted vibrational spectroscopist, Dr. Bruce Chase, has introduced the idea that vibrational spectroscopy, in general, has undergone an ''Eighth Age'' of analytical development.[55] In regard to IR spectroscopy, the Eighth Age manifested itself in the development of many sophisticated sampling methods including, but not limited to, reflectance spectroscopy(s), photoacoustic spectroscopy, IR emission spectroscopy, gas chromatography linked to FTIR spectrometry, and IR microspectroscopy. Each IR sampling method has its own development that can be described by Laitinen's Ages: experimental measurements in the research lab establish validity of the method, the instrument is introduced into the general laboratory where nonspecialists perform the measurements, the method is accepted and in competition with existing methods, applications appear in a wide variety of scientific areas, and applications become standard methods. Furthermore, because IR microspectroscopy is IR spectroscopy on the microscopic level, we are not surprised that it can be further broken down into its many sampling methods (sections 4, 5, and 6) each of which has gone through a ''mini-life'' of ages. These factors, among others, account for the explosion which Chase refers to as the Eighth Age.

Real-time IR imaging is the newest IR microspectroscopic sampling technique. The validity of the method is being established, it is beginning to be introduced into the general lab, and is offering some competition to traditional mapping techniques. Future developments in *uncooled* FPA detectors for both the near- and mid-IR regions will allow for higher pixel densities, ease of accessibility, and reduced cost.[56] Thus real-time imaging using the IR microspectrometer may well become routine for many applications, including medical diagnoses and NIR quality control.

ACKNOWLEDGMENTS

The author gratefully acknowledges Dr Andre Sommer, Director of the Molecular Microspectroscopy Laboratory at Miami University, for his helpful discussions in preparation of this manuscript.

ABBREVIATIONS AND ACRONYMS

AOTF	Acousto-optical Tunable Filter
ASTM	American Society for Testing and Materials
ATR	Attenuated Total Reflection
DL	Detection Limit
FPA	Focal Plane Array
FTIR	Fourier Transform Infrared

IR	Infrared
MCT	Mercury Cadmium Telluride
NA	Numerical Aperture
NFOM	Near-field Optical Microscopy
NIR	Near-infrared
S/N	Signal-to-noise Ratio

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Microbiology

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Forensic Science **(Volume 5)**
Microspectrophotometry in Forensic Science

Pharmaceuticals and Drugs **(Volume 8)**
Vibrational Spectroscopy in Drug Discovery, Development and Production

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Theory of Infrared Spectroscopy

REFERENCES

1. R. Barer, A.R.H. Cole, H.W. Thompson, 'Infrared Spectroscopy with the Reflecting Microscope in Physics, Chemistry and Biology', *Nature*, **163**, 198–201 (1949).
2. V.J. Coates, A. Offner, E.H. Siegler, Jr, 'Design and Performance of an Infrared Microscope Attachment', *J. Opt. Soc. Amer.*, **43**, 984–989 (1953).
3. E.R. Blout, G.R. Bird, 'Infrared Microspectroscopy', *J. Opt. Soc. Amer.*, **40**, 304 (1950).
4. D.W. Schiering, 'Infrared Microspectroscopic Solutions to Manufacturing and Quality Control Problems', In: *Infrared Microspectroscopy: Theory and applications*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, New York, 229, 1988.
5. P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley and Sons, New York, 430, 1986.
6. P.B. Roush (ed.), *The Design, Sample Handling, and Applications of Infrared Microscopes*, ASTM, Philadelphia, PA, 1985.
7. J.E. Katon, A.J. Sommer, P.L. Lang, 'Infrared Microspectroscopy', *Appl. Spectrosc. Rev.*, **25**, 173–211 (1989–1990).
8. F.J. Bergin, 'Some Novel Applications of an Infrared Microscope', *Appl. Spectrosc.*, **43**(3), 511–515 (1988).
9. P.L. Lang, L.J. Richwine, 'The Versatile Sampling Methods of Infrared Microspectroscopy', In: *Practical Sampling Techniques for Infrared Analysis*, ed. P.B. Coleman, CRC Press, Boca Raton, FL, 146–162, 1993.
10. M.A. Harthcock, S.C. Atkin, 'Infrared Microspectroscopy: Development and Applications of Imaging Capabilities', In: *Infrared Microspectroscopy: Theory and Applications*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, New York, 21–39, 1988.
11. K. Krishnan, J.R. Powell, S.L. Hill, 'Infrared Microimaging', In: *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, New York, 85–110, 1995.
12. R.G. Messerschmidt, 'Minimizing Optical Nonlinearities in Infrared Microspectroscopy', In: *Infrared Microspectroscopy, Theory and Applications*, eds. R.G. Messerschmidt, M.A. Harthcock, Marcel Dekker, New York, 1–19, 1988.
13. R.G. Messerschmidt, 'Minimizing Optical Nonlinearities in Infrared Microspectroscopy', In: *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, New York, 3–39, 1995.
14. S.A. Stewart, A.J. Sommer, 'Optical Non-linearities in Infrared Microspectroscopy: A Preliminary Study into the Effects of Sample Size and Shape on Photometric Accuracy', *Microscopy and Microanalysis*, **3**(Suppl. 2), 837–839 (1997).
15. P.J. Treado, M.D. Morris, 'Infrared and Raman Spectroscopic Imaging', *Appl. Spectrosc. Rev.*, **29**(1), 1–38 (1994).
16. J.A. Reffner, P.A. Martoglio, 'United Microscopy and Spectroscopy' In: *Practical Guide to Infrared Microspectroscopy*, ed. H.J. Humecki, Marcel Dekker, New York, 41–84, 1995.
17. J.C. Shearer, E.C. Peters, 'Fourier Transform Infrared Microspectrophotometry as a Failure Analysis Tool', In: *The Design, Sample Handling, and Applications of*

Infrared Microscopes, ed. P.B. Roush, ASTM, Philadelphia, PA, 27, 1985.
18. J.E. Katon, P.L. Lang, D.W. Schiering, J.F. O'Keefe, 'Instrumental and Sampling Factors in Infrared Microspectroscopy', In: *The Design, Sample Handling, and Applications of Infrared Microscopes*, ed. P.B. Roush, ASTM, Philadelphia, PA, 49–63, 1987.
19. M.J. Smith, R.T. Carl, 'Applications of Microspectroscopy in the Near-Infrared Region', *Appl. Spectrosc.*, **43**(5), 865–872 (1989).
20. B. Chabert, J.L. Gardette, G. Lachenal, I. Stevenson, 'Applications of FT-Raman and FT-IR Microspectroscopy in the Near Infrared to Polymer Analysis', *Mikrochim. Acta*, (Suppl.), **14**, 399–402 (1997).
21. P. Lang, A.J. Sommer, J.E. Katon, 'Infrared and Raman Microspectroscopy: An Overview of Their Use in the Identification of Microscopic Particulates', In: *Particles on Surfaces 2, Detection, Adhesion and Removal*, ed. K.L. Mittal, Plenum Press, New York, 143–161, 1988.
22. D.B. Chase, 'Infrared Microscopy: A Single-Fiber Technique', In: *The Design, Sample Handling, and Applications of Infrared Microscopes*, ed. P.B. Roush, ASTM, Philadelphia, PA, 4–11, 1987.
23. P.H. Turner, 'The Spectroscopy of Small Samples', *Anal. Proc.*, **23**, 268 (1986).
24. A.L. Smith, *Applied Infrared Spectroscopy, Fundamentals, Techniques and Analytical Problem Solving*, John Wiley and Sons, New York, 78, 1979.
25. J.E. Katon, 'Infrared Microspectroscopy. A Review of Fundamentals and Applications', *Micron*, 303–314 (1996).
26. B. Wopenka, P.D. Swan, 'Micro-FTIR Measurements on Individual Interplanetary Dust Particles', *Mikrochim. Acta*, **1**, 183 (1987).
27. O. Ruau, P. Landais, J.L. Gardette, 'Quantitative Analysis of Powdered Organic Matter by Transmission Infrared Microspectroscopy Using a Diamond Window Compression Cell', *Fuel*, **76**(7), 645–653 (1997).
28. A.C. Dong, R.G. Messerschmidt, J.A. Reffner, W. Caughey, 'IR Spectroscopy of a Single Cell – The Human Erythrocyte', *Biochem. Biophys. Res. Commun.*, **156**(2), 752–756 (1988).
29. I. Davidsohn, J.B. Henry, eds., *Todd-Sanford Clinical Diagnosis*, 14th edition, W.B. Saunders, Philadelphia, PA, 137, 1969.
30. R. Kellner, B. Mizaikoff, K. Taga, W. Theib, P. Grosse, 'FTIR-Microspectroscopic Detection of Ultra-thin Organic Films on Chalcogenide Fibers', *Fresenius' J. Anal. Chem.*, **346**, 612–614 (1993).
31. B. Mizaikoff, K. Taga, R. Kellner, 'Toward Theoretical Limits of FT-IR Microspectroscopy for Ultra-thin Organic Layers', *Appl. Spectrosc.*, **47**(9), 1476–1483 (1993).
32. K. Krishnan, 'Applications of the Kramers–Kronig Dispersion Relations to the Analysis of FT-IR Specular Reflectance Spectra', FTS/IR Notes No. 51, Bio-Rad, Digilab Division, Cambridge, MA, 1987.
33. T.S. Robinson, *Proc. Phys. Soc. B.*, **65**, 910 (1952).
34. D.M. Roessler, *Brit. J. Appl. Phys.*, **17**, 1313 (1966).
35. R. Lin, G.P. Ritz, 'Reflectance FT-IR Microspectroscopy of Fossil Algae Contained in Organic-rich Shales', *Appl. Spectrosc.*, **47**(3), 265–271 (1993).
36. S.P. Bouffard, A.J. Sommer, J.E. Katon, S. Godber, 'Use of Molecular Microspectroscopy to Characterize Pigment-Loaded Polypropylene Single Fibers', *Appl. Spectrosc.*, **48**, 1387–1393 (1994).
37. S.V. Pepper, 'Characterization and Application of a Grazing Angle Objective for Quantitative Infrared Reflection Microspectroscopy', *Appl. Spectrosc.*, **49**(3), 354–360 (1995).
38. L. Lewis, A.J. Sommer, 'Attenuated Total Internal Reflection Microspectroscopy of Isolated Particles, An Alternative Approach to Current Methods', *Appl. Spectrosc.*, in press.
39. E.P. Paschalis, E. DiCarlo, F. Betts, P. Sherman, R. Mendelsohn, A.L. Boskey, 'FTIR Microspectroscopic Analysis of Human Osteonal Bone', *Calcified Tissue Intl.*, **59**, 480–487 (1996).
40. R.J. Pell, M.L. McKelvy, M.A. Harthcock, 'Effective Resolution Enhancement of Infrared Microspectroscopic Data by Multiresponse Nonlinear Optimization', *Appl. Spectrosc.*, **57**(5), 634–642 (1993).
41. P. Colarusso, L.H. Kidder, I.W. Levin, J.C. Fraser, J.F. Arens, E.N. Lewis, 'Infrared Spectroscopic Imaging: From Planetary to Cellular Systems', *Appl. Spectrosc.*, **52**(3), 106A–120A (1998).
42. J.L. Koenig and C.M. Snively, 'Fast FTIR Imaging: Theory and Applications', *Spectrosc.*, **13**(11), 22–28 (1998).
43. P.J. Treado, I.W. Levin, E.N. Lewis, 'Near-Infrared Acousto-optic Filtered Spectroscopic Microscopy: A Solid-state Approach to Chemical Imaging', *Appl. Spectrosc.*, **46**(4), 553–559 (1992).
44. E.N. Lewis, I.W. Levin, 'Real-time, Mid-infrared Spectroscopic Imaging Microscopy Using Indium Antimonide Focal-plane Array Detection', *Appl. Spectrosc.*, **49**(5), 672–678 (1995).
45. E.N.Lewis, A.M. Gorbach,C. Marcott, I.W. Levin, 'High-fidelity Fourier Transform Infrared Spectroscopic Imaging of Primate Brain Tissue', *Appl. Spectrosc.*, **50**(2), 263–269 (1996).
46. C. Marcott, R.C. Reeder, E.P. Paschalis, D.N. Tatakis, A.L. Boskey, R. Mendelsohn, 'Infrared Microspectroscopic Imaging of Biomineralized Tissues Using A Mercury-Cadmium-Telluride Focal-plane Array Detector', *Cellular and Molecular Biology*, **44**(1), 109–115 (1998).
47. D.S. Lester, L.H. Kidder, I.W. Levin, E.N. Lewis, 'Infrared Microspectroscopic Imaging of the Cerebellum of Normal and Cytarabine Treated Rats', *Cellular and Molecular Biology*, **44**(1), 29–38 (1998).
48. G.L. Carr, A. Hanfland, G.P. Williams, 'Midinfrared Beamline at the National Synchrotron Light Source Port U2B', *Rev. Sci. Instrum.*, **66**(2), 1643–1645 (1995).

49. G.L. Carr, A. Hanfland, G.P. Williams, 'Performance of an Infrared Microspectrometer at the NSLS', *Rev. Sci. Instrum.*, **66**, 1490–1492 (1995).
50. J.A. Reffner, G.L. Carr, G.P. Williams, 'Infrared Microspectroscopy with Synchrotron Radiation', *Mikrochim. Acta* (Suppl.), **14**, 339–341 (1997).
51. R.G. Messerschmidt, 'Photometric Considerations in the Design and Use of Infrared Microscope Accessories', In: *The Design, Sample Handling, and Applications of Infrared Microscopes*, ed. P.B. Roush, ASTM, Philadelphia, PA, 12–26, 1987.
52. J.J. Sahlin, N.A. Peppas, 'Near-field FTIR Imaging: A Technique for Enhancing Spatial Resolution in FTIR Microscopy', *J. Appl. Polymer Sci.*, **63**, 103–110 (1997).
53. A. Piednoir, F. Creuzet, 'Near-field Optical Microscopy in the Infrared Range', *Micron*, **27**(5), 335–339 (1996).
54. H.A. Laitinen, 'The Seven Ages of an Analytical Method', *Anal. Chem.*, **45**(14), 2305 (1973).
55. B. Chase, 'Vibrational Spectroscopy 1975–1998: The Eighth Age', Paper #481, The 25th Annual Conference of the Federation of Analytical Chemistry and Spectroscopy Societies, Austin, TX, October 11–15, 1998.
56. N. Lewis, L.H. Kidder, P. Colarusso, I.W. Levin, D.J. Walls, 'Uncooled Arrays and More: Directions In Infrared Spectroscope Imaging' Paper #275, The 25th Annual Conference of the Federation of Analytical Chemistry and Spectroscopy Societies, Austin, TX, October 11–15, 1998.

Quantitative Analysis, Infrared

Frédéric Cadet
Université de la Réunion, 97715 Saint Denis messag cedex 9, France-Dom

Miguel de la Guardia
University of Valencia, 46100 Burjassot, Spain

1 Introduction 1
2 Quantification Methods 3
2.1 Selection of the Analysis Region 3
2.2 Classical Quantification Methods 3
2.3 Eigenvector Quantification Methods 5
3 Measured Parameters 7
3.1 Chemical Parameters 8
3.2 Physical Parameters 8
4 Applications of Direct Quantitative Infrared Spectroscopy 8
4.1 Infrared Spectroscopy in Food Analysis 8
4.2 The Use of Infrared Spectroscopy in Enzymatic Analysis 13
5 Applications of Fourier Transform Infrared Analysis Coupled with Other Analytical Methods 14
5.1 Fourier Transform Infrared Determination in Pharmaceuticals 15
5.2 Fourier Transform Infrared Determination in Petroleum Products 17
5.3 Fourier Transform Infrared Determination of Industrial Products 18
5.4 Fourier Transform Infrared Determination of Products of the Paint Industries 19
5.5 Fourier Transform Infrared Determination of Products in the Field of Health 20
5.6 Fourier Transform Infrared Determination of Compounds in the Natural Environment 21
5.7 Quantitative Fourier Transform Infrared Analysis of Compounds at Trace Levels 22
5.8 Other Fourier Transform Infrared Coupled Methods 23
6 Limitations and Perspectives of Infrared Spectroscopy for Quantitative Analysis 24
Acknowledgments 25
Abbreviations and Acronyms 25
Related Articles 25
References 26

In this article, after reviewing the different mathematical methods used in quantification we will describe and discuss their respective advantages and inconveniences. First of all, different examples of applications of infrared (IR) spectroscopy, used with or without other analytical chemistry methods, will be presented. Some examples of direct quantitative analysis in food analysis will be reviewed: lipids (measurement of unsaturation degrees, lipids determination), carbohydrates, proteins (secondary structures, quantitative analysis). We will also focus on original examples of the use of IR spectroscopy combined with enzymes.

Secondly, some examples of Fourier transform infrared (FTIR) analysis coupled with other analytical methods are reviewed in the following fields: FTIR quantification in pharmaceuticals, petroleum, paints and other industrial products and in the fields of health, environment, and trace compounds analysis.

Finally, we will discuss the limitations and perspectives of IR spectroscopy for quantitative analysis. It appears that the most significant developments in the field of quantitative analysis will most probably come through progress in chemometrics and flow analysis (FA) (automation).

1 INTRODUCTION

Absorption spectrometry investigates the interactions between radiation energy and samples. The frequencies, or wavelengths, at which samples absorb radiation and the corresponding absorption intensities, are hereby recorded into a spectrum. These intensities are measured in terms of absorbance, whose values reflect the amount of radiation absorbed by the sample while the amount of radiation which travels through a sample is termed transmittance.

Since functional groups are known to absorb light at specific wavelengths, investigating absorption frequencies provides a means for elucidating chemical compositions. The constitution of an unknown sample can be investigated through the study of its wavelength/absorption spectrum. Investigating the chemical composition of unknown samples in spectrometry is commonly termed qualitative analysis.

It is sometimes useful to go beyond qualitative analysis when one is concerned with the amount of a constituent in a sample. This is commonly known as quantitative analysis.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The intensity of an absorption band is linked to the concentration of its associated constituents in a sample. A calibration model links changes in the concentration of the constituent to variations in absorbance. As a matter of fact, calibration models can be successfully used for the quantitative analysis of constituents in samples.

Near infrared (NIR) extends immediately after visible red light, from 700 nanometers (nm) to 2500 nm or, when expressed in cm^{-1}, from 14 285 to 4000 cm^{-1}. NIR spectroscopy is the most commonly used method for quantitative analysis of major biochemical constituents in food industry. The diversity of products that are analyzed by NIR is increasing.[1–4]

To date, fewer studies are concerned with quantitative analysis in the mid-infrared (MIR) range.[5–14] However, more and more papers are now reporting the use of MIR. Several reasons can be put forward to explain this recent interest in this IR domain for quantitative analysis:

- MIR spectra are complex and each functional group contributes to the spectra, which results in numerous bands overlapping.
- Water is a major constituent in biological products and strongly absorbs MIR radiation.
- IR radiation weakly penetrates solid samples. As a consequence, this method only investigates samples' surfaces.
- With classical systems such as pellets or liquid cells employed to obtain MIR spectra, problems are encountered and it is difficult to establish the band-pass correctly or to clean and mount these systems.

However, the advent of FTIR spectroscopy, coupled with the use of powerful computers and the advent of new techniques such as FA and attenuated total reflectance (ATR), have considerably promoted the use of MIR spectroscopy.

Figure 1 shows the evolution of the scientific literature on quantitative analysis by IR spectrometry and, as can be seen, we are nowadays at the beginning of a new era of quantitative IR.

The use of computers has revolutionized studies in many aspects of biological sciences. Data analysis is a relatively new technique which only became popular from the 1960s onwards. The need for data analysis emerged from scientific disciplines, like physics or analytical physical chemistry, where workers had to manipulate considerable amounts of data. At the same time the computer appeared. It was thereby possible to process a considerable amount of data with great speed.

Data analysis is useful for investigations in which numerous variables occur; the terms multivariate analysis or multi-criterion analysis are used. These call upon mathematical spaces containing *n* dimensions. However, for a long time computers have remained for a long time the privilege of specialists. The advent of more and more powerful microcomputers has allowed the introduction of these multidimensional analyses into research laboratories.

Statistical data analysis has been applied to IR spectral data only recently with the advent of Fourier transformed spectrometry in the late 1970s. It follows that this IR technology particularly generated the ideal data required by these statistical methods.

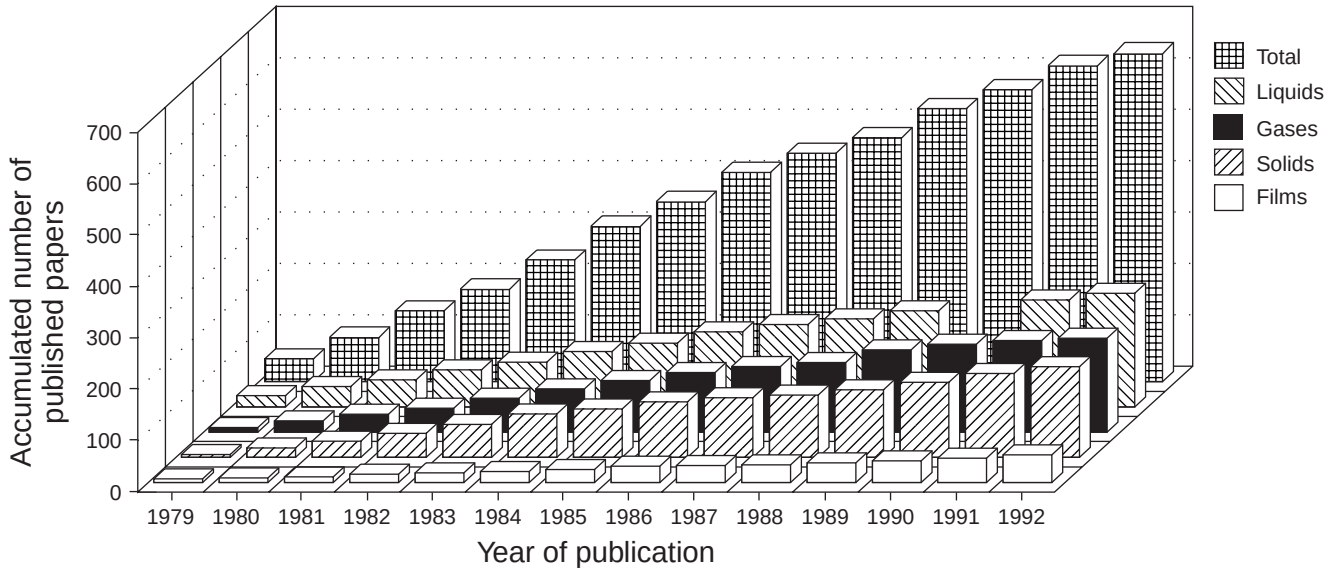

Figure 1 Evolution of the scientific literature on quantitative analysis by IR spectroscopy.

Several mathematical methods can be used for processing IR spectral data in the context of qualitative or quantitative analysis.

In this article, we will first review the different mathematical methods used in quantification and discuss their advantages and inconveniences. We will then describe different example applications of IR spectroscopy used with or without other analytical chemistry methods for the analysis of foods, pharmaceuticals, petroleum products, paints, etc. Original examples of the use of IR spectroscopy combined with the use of enzymes will be also presented. The future prospects of this technique will be discussed.

2 QUANTIFICATION METHODS

In this section, we will describe the most commonly used mathematical methods in IR-based quantitative analysis. Some methods are univariate (solve only one equation based on one measured value and have one calibration value per sample), whereas others are multivariate (solve a series of equations using many measurements per sample for one calibration value). Duckworth[15] has extensively reviewed these methods.

2.1 Selection of the Analysis Region

The choice of the spectral region which correlates best to the concentrations of the constituents to be predicted is one of the most delicate steps in the elaboration of the method of quantitative analysis. Generally, candidate wavelengths or spectral regions are initially chosen among those that are attributed to a constituent. This selection is usually subsequently refined through an iterative process.

The mathematical model that is then used for calculating the concentrations of constituents in unknown samples in quantitative analysis methods is the one determined by the calibration.

2.2 Classical Quantification Methods

2.2.1 Application of the Beer–Lambert Law

The principle of quantitative analysis is that intensities of absorption peaks (absorbance units) are correlated to the amount of constituents. The Beer–Lambert relationship shows that not only are the peak intensities linked to the samples concentrations but that there is also a linear relationship between them, as indicated by Equation (1):

$$A = \varepsilon cl \tag{1}$$

where A is the absorbance, ε the molecular extinction coefficient ($\text{mole}^{-1}\,\text{L}\,\text{cm}^{-1}$), c the concentration ($\text{mole}\,\text{L}^{-1}$) and l the width of the sample (cm) or the bandpass of the cell.

The intensity of an nonoverlapped characteristic absorption band of a compound is hence a function of the concentration of this compound. A calibration model links changes in concentrations to changes in absorbance through the existence of a characteristic constant (the molecular extinction coefficient) and taking into account the bandpass (l).

In cases of substances that do not interact, only one calibration standard is necessary to calculate with precision their concentrations in an unknown sample.

For samples with broad concentration ranges, the calibration is performed on several standards with varying concentrations of the quantified constituent. Spectra of standards are measured so as to determine the absorbance values at each characteristic peak.

In the simplest case, a linear calibration model is obtained: this is the calibration curve. The concentration of an unknown sample can then be easily derived from its absorbance values.

The most commonly used method for calculating a calibration equation is multilinear regression. To calculate calibration equations, several other methods have been developed: least squares regression model, classical least squares model (CLS) and inverse least squares (ILS) model, both using simple or weighed factors.

2.2.2 Multiple Linear Regression

The use of multiple linear regression for determining calibration equations is now well established, as described by several authors. When the data to be predicted are quantitative, the prediction is performed using multilinear regression equations that link calibration data to spectral data. Honigs[16] and Osborne and Fearn[2] have extensively reviewed the use of this method in IR spectroscopy.

The basic principles are summarized below:

The simplest equation that can be established is of the type of Equation (2), where C is the concentration to be predicted, A_l is the absorbance at wavelength l, and a_0 and a_1 are the regression coefficients:

$$C = a_0 + a_1 A_l \tag{2}$$

The wavelength l should be chosen in order to best predict the concentration C.

However, equations of the type of Equation (2) are only applicable for predicting the concentration of one constituent per sample. They would not be convenient for predicting concentrations of other constituents in a mixture where interactions between constituents that absorb at similar wavelength were not taken into account.

In order to take into account these interactions, multilinear regression has been introduced, which is useful for the determination of the concentration of several compounds.

The equation to be used is then as Equation (3):

$$C_j = a_{0j} + a_{1j}A_1 + a_{2j}A_2 + \cdots + a_{nj}A_n \qquad (3)$$

where C_j is the concentration of each one of the compounds to be determined, A_1 are the absorbance values measured at wavelengths from 1 to n, and a_{ij} are the regression coefficients corresponding to each compound j, from 0 to n.

This equation is known as a calibration equation or prediction equation. The absorbances A_1 are the predictive variables and C_j is the predicted variable. The coefficients a_i can be positive or negative terms and can account for interactions between constituents.

Wavelength selection is the main problem in spectroscopic analysis, and wavelength values must be selected in order to be the ones most characteristic of the compounds to be determined and to provide noninterfered data for the compounds. Additionally, in the case of IR analysis, it is necessary to establish an appropriate baseline for the IR spectrum, which could minimize the band overlapping.

The establishment of prediction equations is generally a step-by-step process. The wavelengths are selected as a function of their predictive ability. There are large absorption bands in IR spectra and the data are strongly correlated between them. Wavelength selection is actually not an easy task: replacing a spectral data point by an adjacent one may have no effect on the prediction equation.

To reduce this inconvenience, several prediction equations are retained and tested. The most predictive equation is experimentally determined by testing validation samples.

Prediction equations can also be derived from the first-order and second-order derivative of the absorption zero-order spectra. The first derivative eliminates the general intensity effect of the spectrum and simplifies the baseline selection and the second derivative removes the slope effect.

However, the use of derivative FTIR strongly affects the sensitivity of spectrometric measurement, and thus it must be avoided when constituents are present in the sample at very low concentration values.

Derivation also improves the identification of absorption bands. The selection of the wavelength is thus simplified and the prediction equations obtained give satisfactory results.[17] The quotients of first or second derivatives have also been tested in a simple regression for predicting quantitative values. The inconveniences of these derivation methods are the choice of the calculation steps and, for the quotients of the derivatives, the choice of the terms.

Multilinear regression applied to diffused reflected MIR has only been developed in a few cases. These examples show that it is possible to measure concentrations with satisfactory correlation coefficients.

The use of multilinear regression may lead to prediction equations that are unstable due to the nature of the spectral and quantitative data. The calibration of an instrument, step by step, by multilinear regression does not allow the control of the validation of the prediction equation. The interpretation of the selected wavelengths is not always easy and provides little information on the nature of the samples, owing to possible correlations between quantitative chemical data and spectral deformations.

Prediction equations that have been established from absorbance values do not completely take into account spectral deformations. This problem can be partly solved by using second derivatives.

Another inconvenience of step-by-step multilinear regression, applied to collections of continuous spectra, is that the number of data to be processed is considerable. McClure et al.[18] proposed that the corresponding Fourier coefficients should be calculated. In fact, it was shown that less than 100 coefficients are necessary to describe a spectrum. The selection of the variables of the prediction equation is then performed on these Fourier coefficients. The results of McClure are of similar precision to those obtained by traditional methods. The search for the regression terms is hereby simplified due to the diminution of the number of variables analyzed.

2.2.3 Least Squares Method

The adjustment of the curve following the least squares criterion is based on a linear model. A spectrum S can be expressed as a linear combination of spectra of its constituents, Equation (4):

$$S = S_{i=1}^{n} a_i c_j \qquad (4)$$

where j is the number of constituents, c_j the concentration of the jth constituent and a_i the coefficients obtained for each constituent.

The adjustment of the curves following the least squares criterion is performed as follows: the spectra of pure constituents are measured and the coefficients a_i of the linear combination are calculated by multiple linear regression. The constituents' concentrations are thereby calculated using the coefficients a_i instead of the absorbances as in the stepwise multilinear regression. The advantages and inconveniences of the method are summarized in Table 1.

2.2.4 Classical Least Squares Model

This method is based on the use of Beer's Law to extend the calculation of the absorption coefficients across a

Table 1 Advantages and disadvantages of least squares regression

Advantages	Disadvantages
Easy to understand and calculate	Requires isolated spectral bands that are solely related to the constituent(s) of interest
Calculations are very fast	Cannot be used for complex mixture samples in which the individual constituents have overlapping spectral bands
Used primarily for simple samples: pure compounds, binary mixtures, etc.	Band selection can be difficult or impossible if spectrum of property of interest is not known
	Selecting correct polynomial can be difficult; most spectroscopic systems can be solved with a simple straight line, but this is not 100% true
	Large prediction errors will result from constituents with bands in the same region of the spectrum as the calibration band(s)

Reproduced with permission from J.H. Duckworth, 'Spectroscopic Qualitative Analysis', in *Applied Spectroscopy: A Compact Reference for Practitioners*, eds. J.W. Workman, Jr, A. Springsteen, Academic Press, London, 1998. © 1998 by Galactic Industries Corporation.

much larger portion of the spectrum than the least squares regression model.

Beer's law is rewritten as Equation (5):

$$A_l = K_l C \tag{5}$$

where the absorption coefficient and path length are combined into a single constant K. Measuring the absorbance of a single sample of known concentration and using these values to solve for K can easily solve this equation. Predicting the concentration of an unknown sample is as simple as measuring the absorbance at the same wavelength and rearranging Equation (5):

$$C = \frac{A_l}{K_l}$$

Typically, the CLS calibration method is most useful when applied to samples that have minimal or no interconstituent interactions. The primary application for CLS has been in the analysis of gas-phase samples.

Table 2 below summarizes the advantages and disadvantages of this method.

2.2.5 Inverse Least Squares Model

If the concentrations of all the constituents in the mixtures are not known, there may be a significant error when solving for the absorption coefficients of the known constituents by the CLS approach. One solution to this problem is by rearranging Beer's law and expressing it as Equation (6):

Table 2 Advantages and disadvantages of CLS method

Advantages	Disadvantages
Based on Beer's law	Requires knowledge of the complete composition (concentration of every constituent) of the calibration mixtures
Calculations are relatively fast	Not useful for mixtures with constituents that interact
Can be used for moderately complex mixtures	Very susceptible to baseline; the response at a particular wavelength is due entirely to the calibration constituents
Calibrations do not necessarily require wavelength selection. As long as the number of wavelengths exceeds the number of constituents any number (up to the entire spectrum) can be used	
Using a large number of wavelengths tends to give an averaging effect to the solution, making it less susceptible to noise in the spectra	

Reproduced with permission from J.H. Duckworth, 'Spectroscopic Quantitative Analysis', in *Applied Spectroscopy: A Compact Reference for Practitioners*, eds. J.W. Workman, Jr, A. Springsteen, Academic Press, London, 1998. © 1998 by Galactic Industries Corporation.

$$C = \frac{A_l}{\varepsilon_l l} \tag{6}$$

or, by combining the absorption coefficient (ε_l) and the path length (l) into a single constant as with CLS, express it as Equation (7):

$$C = PA_l + E \tag{7}$$

where, as before, C is the constituent concentration, A_l is the absorbance at the wavelength l, and E is a matrix of concentration (not absorbance) errors.

Because it is not necessary to know the composition of the training mixture beyond the constituents of interest, the ILS method is better suited to more complex types of analysis not handled by the CLS approach. This method has been used for samples ranging from natural products (such as wheat, wood, cotton and gasoline) to manufactured products.

The major features of this method are summed up in Table 3 below.

2.3 Eigenvector Quantification Methods

2.3.1 Principal Component Regression

Multidimensional statistical analyses, such as principal component analyses (PCA), describe variation in multidimensional data by few synthetic variables. These synthetic

Table 3 Advantages and disadvantages of ILS

Advantages	Disadvantages
Based on Beer's law	Wavelength selection can be difficult and time-consuming; must avoid collinear wavelengths
Calculations are relatively fast	Number of wavelengths used in the model is limited by the number of calibration samples
Multivariate model allows calibration of very complex mixtures because only knowledge of constituents of interest is required	Generally, a large number of samples are required for accurate calibration
	Collecting calibration samples and measuring via a primary calibration can be difficult and tedious

Reproduced with permission from J.H. Duckworth, 'Spectroscopic Quantitative Analysis', in *Applied Spectroscopy: A Compact Reference for Practitioners*, eds. J.W. Workman, Jr, A. Springsteen, Academic Press, London, 1998. © 1998 by Galactic Industries Corporation.

variables are linear combinations of all the original variables and have the advantage of having no correlation with each other. Simpler descriptions of data sets are thus obtained with minimal loss of information. These treatments were used for morphological analysis of spectra[19] and for graphical representation of spectra similarity.[20]

PCA is applied to the full spectra that are centered prior (to PCA) according to Equation (8):

$$X_{ij} = A_{ij} - A_j - A_i + A \tag{8}$$

where X_{ij} = centered data; A_{ij} = spectral data ($\log 1/R$) of spectrum i and wavelength j; A_j = mean value of spectral data at wavelength j for every spectrum; A_i = mean value of spectral data of spectrum i for every wavelength; and A = average mean of all spectral data in the collection.

Principal component regression (PCR) was used to establish a prediction equation. PCR is basically a multilinear regression applied to scores and loadings assessed by PCA,[21,22] being the coordinates of data in the new orthogonal axis and the loadings of the directional vectors of the new axis related to the preceding ones. Interest in the introduction of scores according to their predictive ability has already been shown.[23,24]

Concentrations are predicted according to Equation (9):

$$\mathbf{C}_{nl} = \mathbf{X}_{\mathrm{nk}}\mathbf{V}_{\mathrm{kp}}\mathbf{R}_{\mathrm{p}l} \tag{9}$$

where **C** is the column vector of predicted concentrations, **X** is the centered matrix of spectral data, **V** is the matrix of latent vectors of PCA, and **R** is the column vector of the regression coefficients of the prediction equations. n, k, p are respectively the number of samples, the number of wavelengths and the number of significant principal components. The product $\mathbf{V} \cdot \mathbf{R}$ is a vector, the components of which may be interpreted in terms of absorption bands. Plotting the components against the corresponding wavelengths gives a spectral pattern. Peaks correspond to absorption bands which are characteristic of the measured chemical constituents. Troughs indicate that when the concentration increases, the corresponding absorption bands will decrease.[25]

The advantages and disadvantages of this method are indicated in Table 4.

2.3.2 Partial Least Squares

The partial least squares (PLS) method calculates from spectral and concentration or chemical data new variables termed ''factorial coordinates'' which are meaningful with regard to the chemical variables that have to be predicted.[26] The first coordinate is calculated so as to predict as much as possible the chemical variable and the spectral data. The nonpredicted part of the chemical variables and of the spectral data, that is the residuals, are calculated. The second coordinate is then determined so as to best predict the residuals of the chemical variables and of the spectral data. The new residuals are then calculated and the procedure is repeated until a

Table 4 Advantages and disadvantages of PCR analysis

Advantages	Disadvantages
Does not require wavelength selection; any number can be used, usually the whole spectrum or large regions	Calculations are slower than those of most classical methods
Larger number of wavelengths gives an averaging effect, making the model less susceptible to spectral noise	Optimization requires some knowledge of PCA; models are more complex to understand and interpret
PCA data compression allows use of inverse regression to calculate model coefficients; can calibrate only for constituents of interest	No guarantee that PCA vectors directly correspond to constituents of interest
Can be used for very complex mixtures because only knowledge of constituents of interest is required	Generally, a large number of samples are required for accurate calibration
Can sometimes be used to predict samples with constituents (contaminants) not present in the original calibration mixtures	Collecting calibration samples can be difficult; must avoid collinear constituent concentrations

Reproduced with permission from J.H. Duckworth, 'Spectroscopic Quantitative Analysis', in *Applied Spectroscopy: A Compact Reference for Practitioners*, eds. J.W. Workman, Jr, A. Springsteen, Academic Press, London, 1998. © 1998 by Galactic Industries Corporation.

satisfactory precision for the estimation of the chemical variable is obtained.

The new spectrum S is rewritten as Equation (10):

$$S = S + Sa_i c_i + r \tag{10}$$

where S is the mean spectrum of the collection, a_i the ith factorial coordinate, c_i the ith factor and r the residual.

The factors are estimated while calculating the chemical data coordinates and can be written as Equation (11):

$$p = p + Sa_i d_i + rp \tag{11}$$

where p is the mean value of the chemical data, a_i is the ith factorial coordinate calculated by the equation, d_i is the ith factor associated with the chemical data, and rp is the residual.

The d_i factors are estimated during the calculation of the coordinates similarly to the c_i factors defined above. The chemical values for an unknown sample are estimated by calculating a_i, the factorial coordinates, with Equation (10) and then by applying Equation (11) so as to determine the values of p.

There are actually two versions of the PLS algorithm: PLS-1 and PLS-2. The differences between these two methods are subtle but have very important effects on the results. Like the PCR method, PLS-2 calibrates for all constituents simultaneously: one set of scores and one set of eigenvectors or loadings are obtained for calibration during the spectral decomposition. Therefore, the calculated vectors are not optimized for each individual constituent. In PLS-1, a separate set of scores and loading vectors is calculated for each constituent of interest. The method should give more accurate predictions than PCR or PLS-2.

Table 5 summarizes the advantages and disadvantages of the PLS methods.

Table 5 Advantages and disadvantages of PLS method

Advantages	Disadvantages
Combines the full spectral coverage of CLS with partial composition regression of ILS	Calculations are slower than most classical methods, especially PLS-1. Models are more abstract, thus more difficult to understand and interpret
Single-step decomposition and regression; eigenvectors are directly related to constituents of interest rather than largest common spectral variations	Generally, a large number of samples are required for accurate calibration
Calibrations are generally more robust provided that calibration set accurately reflects range of variability expected in unknown samples	Collecting calibration samples can be difficult; must avoid collinear constituent concentrations
Can be used for very complex mixtures because only knowledge of constituents of interest is required	
Can sometimes be used to predict samples with constituents (contaminants) not present in the original calibration mixtures	

Reproduced with permission from J.H. Duckworth, 'Spectroscopic Quantitative Analysis', in *Applied Spectroscopy: A Compact Reference for Practitioners*, eds. J.W. Workman, Jr, A. Springsteen, Academic Press, London, 1998. © 1998 by Galactic Industries Corporation.

2.3.3 *Comparison of Partial Least Squares and Principal Component Regression Methods*

It should be noticed that a spectrum obtained by least squares curve adjustments is in fact a product of spectra similar to the additive model in PCA. The resulting spectra are similar to eigenvectors and the variables with which they are associated are calculated. However the variables and the spectra are not orthogonal. The major advantages of PCA over least squares methods are the orthogonality of the factorial coordinates, but, most importantly, it calculates typical spectra and hence extracts the major characteristic wavelengths.

In the method of Hruscka and Norris,[27] the calculation basis from the typical spectra must be chosen. Even if they are carefully selected so as to be as different as possible, they are nevertheless less dissimilar than eigenvectors are and thus do not detect the most characteristic frequencies of constituents. As a matter of fact, PCA is a more informative method than least squares curve adjustment methods.

The PLS method is interestingly similar to PCA, since factors and coordinates are calculated. The difference lies in the mode of calculation. Data are estimated as a function of the characteristic parameters to be measured. As a consequence, the first calculated coordinates are more predictive. With regard to this, PLS is more powerful than PCA since the information on the first axes is more predictive. However, Martens and Jensen[26] showed that the total number of axes introduced into the regression is identical and that the precision is similar. The two methods are almost identical with regard to the results obtained. The major advantage of PCA is that the underlying theory is simpler than PLS. Hence modeled spectra are easier to interpret.

3 MEASURED PARAMETERS

In an industrial process it is often necessary to monitor and control the successive steps. As is mostly the case in

chemical processes, the methods used for these controls are rather long. It becomes more advantageous to develop alternative methods such as spectroscopic methods, which are more rapid than classical ones, to measure both physical and chemical parameters.

Spectroscopic methods are very interesting when real-time-based measurements are necessary. IR can provide information on raw materials and on the corresponding final products, but additionally it provides an excellent way for process control.

These measurements are advantageous only when it is possible thereafter to control the measured parameter with the aim of improving the efficiency of the process. These parameters are generally physical or chemical.

3.1 Chemical Parameters

It is evident that foods and pharmaceutical products should be carefully analyzed for security reasons and, for these products, the concentration of sample components is extremely important for health and toxicological reasons.

Several parameters can be measured: (i) IR analysis can in many cases be used to analyze the degree of contamination of a sample, hence its purity. In the food industry these aspects are not only important but are moreover imposed by regulations and by the existing quality standards requirements which are very constraining in several countries; (ii) chemical characteristics, the identification of functional groups (OH, NH_2, COOH, etc.) or identification of constituents (sugars, alcohols, fatty acids, proteins, etc.).

The type of calibration is defined as a function of the products to be analyzed. IR spectrophotometers will need to have the species ''introduced'' in all possible matrixes that might occur in a particular reaction/process. It is however possible to select a wavelength and to use the Beer–Lambert law. When several absorption bands recover, it is still possible to select a specific wavelength by using powerful multidimensional data analysis.

In complex mixtures, powerful mathematical tools, such as PLS or PCR, are used for the quantification of a particular constituent. These mathematical treatments can be applied to a part of the spectrum or to the whole spectrum.

It should not be forgotten that the accuracy of the prediction is dependent on the accuracy of the reference method. It should also be remembered that spectroscopic calibration methods demand as much time and resources as a chemical method would demand in order to guarantee their traceability. The advantages of using such methods are their relatively low costs, their accuracy and their speed.

3.2 Physical Parameters

In addition to the concentration values of sample constituents, physicochemical parameters in fluids can be monitored by IR or NIR. These vibrational spectroscopic techniques are very sensitive to matrix effects such as those existing in the presence of hydrogen bonds. Among the physical parameters most currently measured are:

- Viscosity, which is influenced by the number of hydrogen bonds in the samples. It is therefore possible to measure viscosity by IR spectroscopy.
- Density, since it directly influences the transmission of incident light.
- Color, which can be obtained by an ultraviolet/visible (UV/VIS) spectrophotometer; the advantage of FTIR lies in the possibility of verifying that the color is really given by the supposed constituents of the samples.
- Distillation temperature, or practical parameters such as the octane number in gasoline, are closely related to the sample composition and can be easily derived from the IR spectra using appropriate mathematical models.

4 APPLICATIONS OF DIRECT QUANTITATIVE INFRARED SPECTROSCOPY

4.1 Infrared Spectroscopy in Food Analysis

To date, the ease of utilization of spectroscopic methods such as NIR, MIR, UV/VIS and also Raman or nuclear magnetic resonance (NMR) is such that they can be used in routine analysis in quality management and control of food industry production.

NIR is the most commonly used spectroscopic method in the food industry for the quantification of major biochemical constituents. Its success is primarily due to its simplicity and its rapidity. Analytical results can be obtained in few seconds, even if the samples are opaque to IR radiation, such as solid substrates.

The first trials were performed in 1960 under the direction of the United States Department of Agriculture (USDA). A series of investigations were undertaken in order to develop physicochemical methods for rapid characterization of foodstuffs. Norris[28] et al. rapidly assessed the advantage of using the NIR range: NIR is the only region suitable for the diffused reflection technique, which does not require dilution of samples in an optically inactive substance like potassium bromide. In 1962, he developed the first apparatus for the quantification of water by NIR spectroscopy.[28] These trials showed that

constituents such as proteins, lipids and starch hindered the measurement of water content. This problem was alleviated with the application of multilinear regression to spectral data. This not only improved the precision of the quantification of water but also allowed the quantification of other constituents.

From 1970, numerous types of apparatus were specifically devised for specific analytical applications and were commercialized. These devices were spectrophotometers with interferential optical filters and were coupled to microcomputers. The filter-based NIR instruments were able to perform few measurements at some specific wavelengths only. Nowadays, the most sophisticated spectrophotometers are built with interferometers and make use of Fourier transform signal processing.

The field of application of NIR spectroscopy is very wide: it covers the quantification of major constituents such as water, proteins, lipids and sugars in almost every foodstuff.[2,3]

The incorporation of the Fourier transform technique into MIR spectrometers[29] and the advent of new and easy to use techniques[30] such as ATR, photoacoustic spectrometry (PAS), diffused reflectance infrared Fourier transform (DRIFT) has largely contributed to the progress of IR-based analytical methods. In 1991, several workers showed that ATR was potentially a powerful tool for the study of food products.[31–33] With the application of Fourier transforms, it was possible to obtain MIR spectra by diffused reflection. Fuller and Griffiths[34] and Depecker et al.[31] demonstrated the advantages of the diffused reflection technique for the study of biological samples.

The use of MIR spectroscopy for the analysis of foodstuffs is still restricted. Few example applications have been described; Renard et al.[35] used MIR for the analysis of cereals and related products. Other applications for the analysis of milk and dairy products have been proposed.[36]

We will now describe the use of spectroscopy for direct spectroscopic quantitative analysis in the study of lipids, carbohydrates and proteins. We will then show the possible applications of IR spectroscopy to the study of enzymatic reactions. Finally, examples of quantification through FTIR spectrometry coupled to other analytical techniques such as FA, vapor generation, gas chromatography (GC), microwave-assisted vapor-generation, and sequential injection analysis (SIA) will be reviewed (Figures 2–4).

4.1.1 Lipids

The rapid control of the quality of lipids is a major preoccupation in the food industry. NIR has long been used for this purpose. Recent developments in MIR

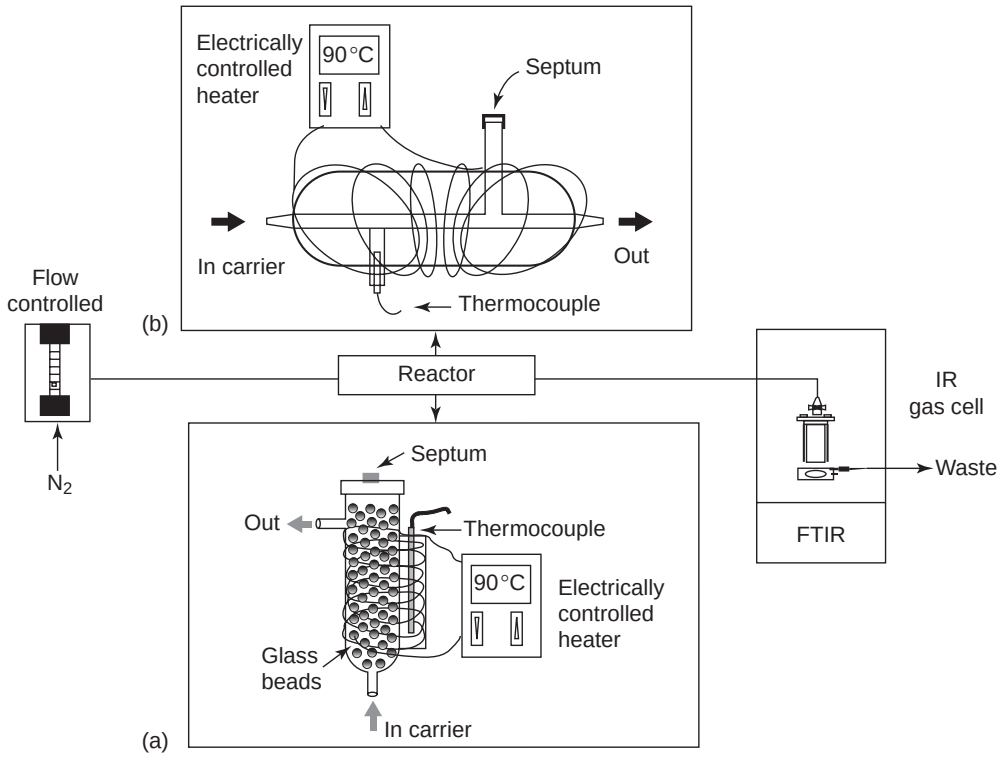

Figure 2 Manifold employed for vapor generation FTIR spectrometry. Inset (a) glass-bead packed reactor, and (b) open reactor.

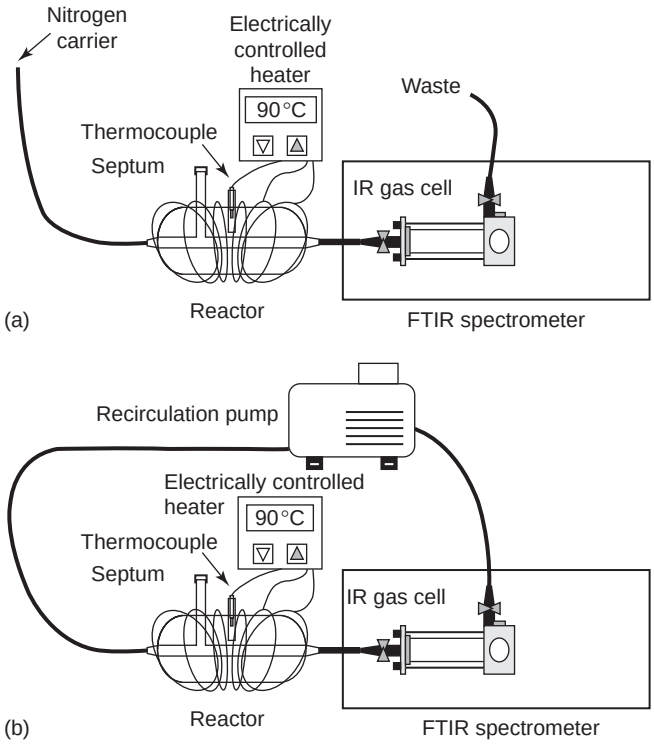

Figure 3 Manifolds employed for vapor generation FTIR analysis. (a) Single-channel flow manifold with a nitrogen carrier flow. (b) Closed FIA manifold.

spectroscopy, such as the advent of the ATR method, which has solved the water absorption problem, have made its use in food analysis easier.

4.1.1.1 Measurement of the Unsaturation Degree of Lipids Holman and Edmondson[37] have shown that the increase or decrease in the intensities of bands in the 2190–2150 nm region is related to the unsaturation degree of lipids. They attributed the bands at 1180, 1680, 2150 and 2190 nm to the group =C−H "*cis*". The bands at 1680 and 1180 nm are the first and second harmonics respectively. They also noticed that NIR did not allow the identification of =C−H "*trans*" in lipids. In 1959, Holman estimated the unsaturation degree of lipids by NIR, MIR and by the iodine index method. Results obtained by MIR and by the classical chemical method were similar. However, since *trans*-unsaturation could not be detected by NIR, the unsaturation degree was systematically underestimated with NIR. Fenton and Crisler[38] developed a NIR-based analytical method in order to determine the *cis*-unsaturation degree in free fatty acids from refined oils. Murray[39] also showed that the absorption bands at 2150 nm and 2190 nm indicated the existence of a *cis*-double bond in free fatty acids.

Wheeler[40] has intensively investigated the structural analysis of lipids through their MIR absorption spectra. MIR is now used routinely for the determination of *cis*- and *trans*-fatty acids in oils and fats and has been adopted as an official routine method by the American Oil Society (AOCS) and the International Union of Pure and Applied Chemistry (IUPAC).

Ahlers et al.[41] have studied the unsaturation degree of pure and sterified free fatty acids. Their investigation was based on the measurement of the intensities of absorption bands at 3015 and 967 cm^{-1}, which were attributed to elongation of *cis* =C−H groups and −CH=CH− *trans*-deformation respectively. The intensity of these two bands increases with the unsaturation degree. The band at 3010 cm^{-1} also often determined the unsaturation

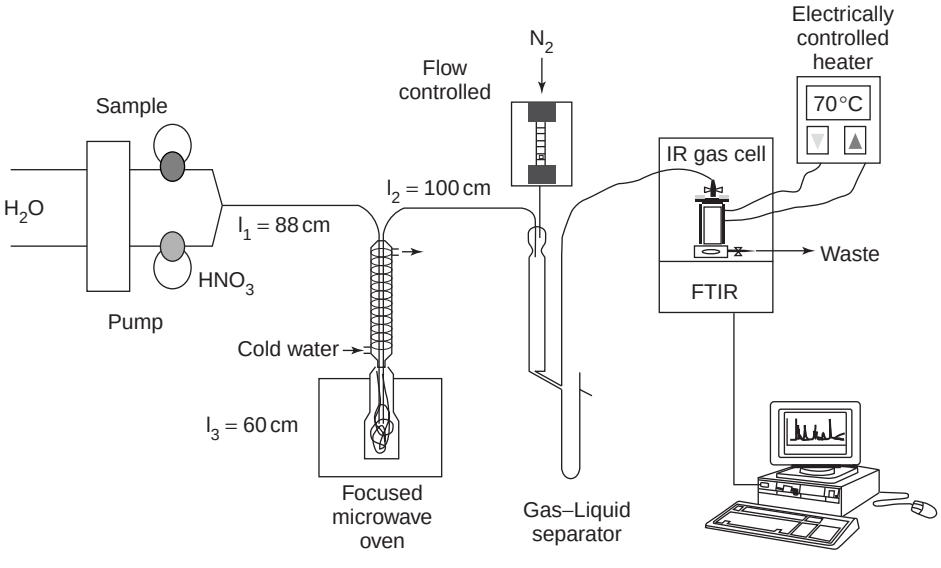

Figure 4 Manifold employed for microwave-assisted vapor-generation FTIR spectrometry: $I_1 = 88$ cm, $I_2 = 100$ cm, $I_3 = 60$ cm.

degree and was attributed to valence vibration of the =CH *cis*-group. Kochhar and Matsui[42] have performed the determination of free *trans*-fatty acid content by a transmission method; the band characteristic of C−H in a *trans* double bond is located at 965 cm^{-1}.

O'Connor[43] has conducted detailed studies in order to differentiate between the *cis*- and *trans*-isomers that were associated with unsaturation.

Arnold and Hartung[44] have determined the unsaturation degree of 25 different oils in the 3700–3200 cm^{-1} region. The unsaturation degree is proportional to the quotient of the intensity of bands at 3015 cm^{-1} and 2860 cm^{-1}. These bands are attributed to the symmetric valence vibration of the $-CH_2-$ group.

Anderson et al.[45] have developed a spectroscopic method for the precise determination of the unsaturation degree in butters and oils. They validated the results found from their method by comparing them with those obtained by the iodine index method. The correlation between IR absorbance and iodine index was verified for butter and mono-hydrogenated oils. Significant correlation coefficients ($r = 0.97$) were obtained (Ahmed and Helal[46]) between absorbance at 1015 cm^{-1} (due to olefin *cis*–*trans* isomerism during hydrogenation) and the iodine index method.

With the use of ATR, Belton et al.[47] have quantified *trans*-unsaturated fatty acids in oils and lipids. The characteristic *trans* =C−H band determined by Belton was located at 976 cm^{-1}.

4.1.1.2 Lipids Determination NIR has long been used as a standard routine method for quantification of lipids in foods and in particular in milk and related compounds.[48–50]

Sato and Kawano[51] showed that characteristic bands of lipids in milk are featured at 1164, 1660, 2144 and 2176 nm. These bands are attributed to *cis*-unsaturations in the fatty acids existing in milk.

Lipids have been quantified by NIR in a great variety of substrates: meat products,[52,53] cereals[54] and oleaginous seeds.[55–57]

In 1986, Murray[39] showed, from a collection of NIR spectra of free fatty acids, that the only difference that can be noticed was in the absorption bands related to the number of CH_2 groups.

MIR absorption bands have been used for the quantification of lipids in milk.[58,59]

With the use of devices equipped with interferential filters, the quantification of lipids by measuring the absorption at 1754 cm^{-1} (characteristic of ester groups) is possible. Since there are on average three ester links in most molecules, this technique is in fact a mean for counting molecules. If the mean molecular weight is constant, the absorption measurements at 1754 cm^{-1} allow the precise assessment of the amount of fats. However, milks show great variations in their fatty acid composition according to their origin and the period of sample collection. In order to overcome this problem, a second filter can be used. This filter permits measurements between 2940 and 2860 cm^{-1} and allows the determination of the average size of fatty acid molecules by measuring the number of C−H bonds.

4.1.2 Quantitative Analysis of Carbohydrates

Both NIR and MIR have been employed for carbohydrate determination in foods.

4.1.2.1 Near Infrared Spectroscopy of Carbohydrates Numerous workers have used NIR for the analysis of sugar contents in foodstuffs. McClure et al.[60] have directly measured total reducing sugars in tobacco leaves. Giangiacomo et al.[61] predicted, with very good precision, the fructose, glucose and sucrose contents of apple juices or of reconstituted[62] mixtures of carbohydrates. Lanza and Li[63] were able to predict total sugar content in fruit juices; Osborne et al.[64] measured sugars in biscuits; Kaffka and Jesenszky[65] determined sugars in wines; Baker and Norris[66] determined the sugar content in cereals and other foodstuffs;[67] Robert et al.[68] analyzed apples, and Ootaka[69] studied tomatoes. Rambla et al.[70] determined glucose, fructose, sucrose and total sugar contents in fruit juices by means of the PLS data treatment of NIR measurements.

Baker[67] quantified starch content and total soluble sugars in fast foods. Bertrand et al.[25] measured total parietal constituents contents in wheat products, including various types of semolina.

NIR can be used to determine the state of starch degradation but this characterization is not supported by a molecular approach. Osborne and Douglas[71] have applied NIR spectroscopy to the prediction of this parameter. In their investigations, they used commercial flour with different degrees of starch degradation. The procedure is useful for the classification of flour according to the extent of starch degradation.

The potential application of NIR spectroscopy in the sugar industry has been assessed. These studies were undertaken on sugarcane juices[72,73] and on beetroot juices.[74,75] Studies performed using an InfraAlyser 450 equipped with 20 filters have shown that it is possible to predict with good precision the sucrose content, as well as the Brix index and humidity, from diluted raw juices. However, these studies were performed on a limited number of samples. In 1989, Ames et al.[76] collected and analyzed up to about 600 NIR spectra of raw sugarcane juices that were collected from several centers in Australia in 1987. The InfraAlyser 500 allowed investigation in the

whole range covering 1100–2500 nm. The results obtained appear to be satisfactory according to the standards of routine analysis in industrial processes. However, a systematic deviation between the polarimetric values and sugar contents predicted by NIR spectroscopy during particular periods in the sugarcane collection has been noticed.

Altenburg and Chou[77] proposed working at 882.6 nm in NIR to determine sugar contents in raw sugarcane juices with a NIR-polarimeter. This work on IR spectroscopy were encouraged by the International Commission for the Uniform Methods of Sugar Analysis (ICUMSA).[78]

4.1.2.2 Mid-infrared Spectroscopy of Carbohydrates

MIR spectroscopy has been used since 1950 for the study of carbohydrates. Structural analysis of carbohydrates from their MIR spectra has been thoroughly investigated.[79–82] Apart from starch, the sugars that are of most interest to the food industry are sucrose, glucose and fructose.

Barker et al.[80] have systematically studied aqueous solutions of carbohydrates by MIR in an attempt to identify bands that are characteristic of sugars in aqueous solutions.

In nonaqueous solvents, there exist IR spectroscopy studies that were carried out on model molecules that featured patterns characteristic of carbohydrates.[82] These investigations essentially focused on the O–H absorption band in the 3700–3300 cm^{-1} region, with the aim of determining sucrose chelation in nonaqueous solutions.

Only a few works are concerned with MIR spectroscopy for direct sugar quantitative determinations. Dupuy et al.[83] have quantified sugars in fruit juices and syrups but they suggested that the results obtained with diffused reflectance did not meet the standards required for a possible use in quality control, particularly for quantitative analysis.[84]

Cadet et al.[33] were the only team to investigate the use of MIR spectroscopy for the study of sugarcane juices. Their work was motivated by the fact that the price of sugarcane is estimated by its sucrose content. Not only is the method used to determine sucrose content in sugarcane fundamental, but also the delay and costs of the analyses are important. Polarimetry and high-performance liquid chromatography (HPLC) are the most commonly used techniques to measure sucrose content in sugarcane. However, it has been shown that polarimetric measurements underestimate sugar content. Measurements by HPLC are not precise enough and the results are not easily reproducible for routine analysis and need a preliminary sample clarification before the analysis is carried out.

MIR ATR can be used to study samples that are difficult to analyze. This is the case for raw sugarcane juice, which is completely opaque.

PCA and PCR were subsequently used to assess and describe MIR spectral data of crude sugarcane juices. The method was validated on a panel of 1267 samples representative of a sugarcane collection.

A peak at 997 cm^{-1} was the most important for the quantification of sucrose. After PCR, the first introduced component gave a correlation coefficient value towards sucrose equal to 0.878 and, with the second introduced component, this value reached 0.949. The bias and the standard deviation of the differences between the reference value and the predicted value are 0.041 g/100 mL and 0.289 g/100 mL respectively. When compared to values obtained between the reference value and direct polarimetric values (−0.163 and 0.087 g mL^{-1}) these values show that the proposed method yields values that are very close to real ones. This method, which uses MIR ATR, allows direct measurements of sucrose content in raw juices and hence is more advantageous for routine analysis and industrial processes than the polarimetric method. Furthermore, the proposed method eliminates costs linked to consumption of the polluting agent lead acetate, and eliminates the lengthy filtration phase. The whole procedure, from raw sugar analysis to prediction of sucrose content, needs only about 25 s for each sample.

Vonach et al.[85] used MIR transmission spectrometry for the direct determination of glucose in human blood.

4.1.3 Quantitative Analysis of Proteins

Proteins are macromolecular structures constituted of polypeptide chains of amino acids that are linked by peptide bonds. It is common to distinguish the different structure and organization of proteins following the nature and the repetitive character of the interactions that exist within the macromolecules.

The primary structure of a protein is defined as its sequence of amino acids.

The secondary structure of a protein is defined by the spatial arrangement of the amino acids in the polypeptide chain. A limited number of particularly favorable spatial configurations are adopted by atoms of each amino acid in a polypeptide chain, which folds to give rise to the well-known secondary structures: the alpha helix and beta strands, as well as less regularly defined structures like turns, loops and coils.

The tertiary structure of a protein is the spatial arrangement of the whole polypeptide chain and arises from the packing of different folding regions owing to interactions between the side chains of the constituent amino acids.

The quaternary structure of proteins describes the arrangement of several polypeptide chains that are associated to form a supramolecular structure.

MIR spectroscopy has been widely used for the determination of polypeptides and protein secondary structures.[86] Miyazawa and Blout[87] carried out normal coordinate calculations on model compounds, which showed three strong IR bands of proteins, called amide I, II and III. The most representative spectral region is that between 1600 and 1700 cm^{-1}, which is characteristic of amide I.

4.1.3.1 Predicting Protein Secondary Structures The secondary structure of a protein is difficult to investigate by physical analytical methods. MIR spectroscopy provides a gateway for studying secondary structures, in particular for quantifying the relative proportioning of alpha helix, beta sheets, turns, etc.

The position of the amide I and II absorption bands depends on the dihedral angles in the peptide units, and hence on the type of periodical structures existing in the investigated protein molecule.[88] Several methods for predicting secondary structures in proteins from IR spectra have been subsequently developed.

Two major methods have been described:

- Methods based on direct analysis of spectra, either by deconvolution,[89] or by calculation of the second derivative.[90]
- Methods that use calibration from a collection of spectra of proteins of known structures.[91,92]

In 1990, Dousseau and Pézolet[93] showed that vibrational spectroscopy in the MIR region can be used to determine with a good precision the secondary structure of proteins. The results obtained can be favorably compared to those obtained by circular dichroism or by Raman spectroscopy.

The proposed method consists of the following steps: subtraction of the spectrum of water, baseline correction and normalization of the spectra, and application of PLS to the corrected spectra – for the prediction of secondary structures.

FTIR spectroscopic studies on membrane proteins were at first mainly of a qualitative nature. Recent developments provide quantitative information as well. For example, PLS FTIR predictions of the secondary structure of the following membrane proteins have been performed: Ca-ATPase, cytochrome oxidase, glucose transporter, rhodopsin, porin, etc. Methods used for quantitative analysis are either based on curve-fitting techniques or pattern recognition.[94]

4.1.3.2 Quantitative Analysis of Proteins The quantitative analysis of proteins by NIR spectroscopy is an alternative to the classical Kjeldahl chemical method and is, as a matter of fact, of great economic importance.

The quantification of proteins in cereals is by far the most important application of NIR spectroscopy. Proteins present several characteristic absorption bands in NIR. The most intense band is located in the 2170 and 2180 nm region.[95] This band is attributed to a combination of the amide I band with that of the amide III band.[96] Quantification of proteins by NIR is as precise as quantification with the Kjeldahl method.

In MIR, a method for the quantitative measurement of α-amino groups from amino acids, peptides and proteins has been proposed by Cadet.[97] The spectral region where the characteristic absorption bands of such groups are located ranges between 1200 cm^{-1} and 1900 cm^{-1}. Water shows a major absorption band in this region (1500–1700 cm^{-1}). This superimposes on the amide I and II bands.

The standard deviation for each and every wavelength, calculated for all the spectra of the calibration set, shows the existence of two absorption bands in the 1500–1900 cm^{-1} region. This means that the observed variations in this zone are not only due to water but are also due to two peaks centered at 1650 cm^{-1} and 1540 cm^{-1} (with a trough at 1600 cm^{-1}) that are characteristic of proteins.

Good predictions were obtained; the mean and standard deviation associated with the predicted concentrations of α-NH_2 (protein) content were 0 g mL^{-1} and 0.12 g mL^{-1} respectively. Hence, it has been clearly established that it is possible to determinine the α-NH_2 content of foods from their MIR spectra.

4.2 The Use of Infrared Spectroscopy in Enzymatic Analysis

Enzymatic studies are often carried out in two different ways: (i) by the determination of the amount of a product or a reagent after reaction through a generally lengthy chemical method or, (ii) by using coupled enzymatic assays, so that a chromophore, like NAD(P), was oxidized or reduced and was finally titrated. This second method is more often used for quantitative analysis through enzymatic reactions that generally do not give chromophoric or fluorescent products directly. However, the use of other enzymes in coupled assays implies a good understanding of their kinetics and can make some studies difficult. As for example is the variation of the activity of an enzyme according to pH. So, for the assay of enzymes with products and reagents that have no absorption in the UV/VIS region, it may be interesting to avoid coupled enzymatic assays and to proceed in a single step.

Recent developments in analytical IR spectroscopy, and particularly the possibility of connecting computers

to spectrophotometers, make this technique easier and give it new scope. Cadet et al.[98] first investigated the potential of MIR spectroscopy combined with multidimensional statistical analysis, using an ATR cell to allow a one-step enzymatic assay. In this way, they proposed an alternate method for enzyme study. Hydrolytic activity of β-fructosidase, with sucrose as substrate, was monitored. Familiar progress curves can be obtained by quantification of the remaining substrate, and enzymatic studies of the kinetic properties of the enzyme become possible. The proposed method offers an alternate analytical procedure for enzyme studies. It also allows the identification of structural changes in a compound during a reaction. Furthermore, the study of the enzyme can be carried out directly without resorting to coupled assays. The authors suggested that the method could be extended to any enzyme, with appropriate calibration.

Schindler et al.[99] proposed an automated method for the determination of amyloglucosidase (glucan 1,4-α-glucosidase) activity in aqueous solutions. Amyloglucosidase is produced from *Aspergillus niger* by submerged fermentation and is frequently used in industry, e.g. for starch liquefaction in food and pharmaceutical industries.[100] The method is based on the direct FTIR spectrometric monitoring of starch hydrolysis catalyzed by amyloglucosidase. It is shown that spectral changes caused by the enzyme activity can be followed by the study of the MIR spectra, hence eliminating the need for additional consecutive reaction steps which are normally required to derive a reaction product detectable by conventional techniques such as UV/VIS spectroscopy or electrochemistry. Unspecific background absorption was eliminated by calculation of a difference spectrum between nonreacted starch and the reaction mixture. To guarantee optimum handling of the viscous starch solution and to obtain high sensitivity, a flow injection analysis (FIA) system using the merging zone approach and the stopped-flow technique was developed. Using a stopped-flow time of 5 min and a 55 g L^{-1} starch substrate solution, linear calibration curves from 50 to 2000 U^{-1} (830–33 300 nkat L^{-1}) were obtained. The coupling of a FTIR spectrometer with a flow system capable of performing chemical reactions in an automated, repeatable and reproducible way was proved to enhance the detection power of the FTIR instrument. By using this approach, a parameter, which is not accessible by a single FTIR measurement, could be successfully determined. Furthermore, this coupling has demonstrated new possibilities for the application of MIR FTIR spectrometry for remote process monitoring and process control and has been applied for the study of other enzymes.[75,101]

5 APPLICATIONS OF FOURIER TRANSFORM INFRARED ANALYSIS COUPLED WITH OTHER ANALYTICAL METHODS

FTIR spectroscopy has been traditionally used to obtain qualitative information of organic sample. But recently it has been demonstrated that FTIR spectroscopy is also an excellent tool to obtain accurate quantitative results from spectra obtained in solid, liquid or gaseous states, in spite of the general problems encountered when using traditional systems of IR measurements.[102–104]

Table 6 provides a picture of the main drawbacks of traditional systems of IR spectra acquisition and the different strategies that can be adapted to solve all these problems.

Among the problems which remain for general applicability of FTIR spectroscopy for quantitative analysis are the solvent and window material transparencies.[105] These are still the main reasons for a limited development in this field. For the analysis of aqueous samples by FTIR, it is necessary to use adequate transmittance or reflectance cells with water-resistant windows. The use of transmittance measurements requires very small path lengths, of the order of 25 µm, due to the strong water absorption, and the use of ATR cells involves poor sensitivity.

The vapor or gaseous state is very useful for carrying out determinations by FTIR. This is due to the high transparency of gases, the low background obtained for this kind of systems and the possibility of using multiple pass gas cells which can provide good sensitivity.[106,107] For these reasons, FTIR spectrometry has been extensively used for detection in GC[108–110] and also in combination with sample pyrolysis.[111–113] More recently, a new strategy for FTIR analysis based on vapor generation from small injected volumes of liquid samples has been

Table 6 Strategies for solving problems of IR spectra acquisition

Sample preparation	Problems	Solutions
Thin films Alkali halide disks	Difficulties in establishing the sample bandpass Bad repeatability	The use of internal standard The use of bands quotient
Gas cells	Leaks Problems in sample preparation Difficult standarization	Derivatization and pyrolysis IR Vapor phase generation
Liquid cells	Compatibility of the cell windows and solvents Difficulties in mounting and cleaning the cells	Use of adivent mixtures Vapor phase generation FIA/FTIR

developed.[6] This technique was applied to the quality control of solvent additives,[110] paint analysis,[7] gasoline component,[114] and to the determination of alcohol in aqueous samples like alcoholic beverages[115] and in blood.[116]

Many examples of on-line determination exist. Most of them deal with the chemistry of synthesis and with the food and pharmaceutical fields.

In recent years, FIA methodology has been incorporated into most branches of analytical chemistry. The second edition of Ruzicka and Hansen's book[117] included 588 references for the use of FIA in UV/VIS determinations, 123 in atomic absorption spectroscopy, 72 in spectrofluorimetry and only 4 in FIA/IR. These values grew up during the past decade. Below, some examples of applications are given.

FTIR has a better selectivity than UV/VIS spectrometry and provides an excellent tool for the determination of mixtures of compounds. On the other hand, in the last few years, FIA/FTIR has been demonstrated to be a synergistic combination highly appropriate for the quality control of drugs.[4,103,118–123]

In general, the analytical procedures developed for IR determination require the use of organic solvents, and frequently chlorinated solvents, such as CCl_4, $CHCl_3$ and CH_2Cl_2, are those that provide the best transparency in the IR range and the lowest limits of detection. However, these reagents are highly toxic and efforts must be made to replace these reagents or to reduce drastically the amount of solvents employed through the automation of measurements and the development of on-line recycling techniques. The FIA/FTIR methodology offers a fast and simple way for automating the sample and measurement operations and provides a clean analytical method.[13,124]

FIA offers an inexpensive and versatile means for the automation of analytical procedures, and hence it has been incorporated in many different techniques.[117,125–127] FTIR spectroscopy is a powerful technique for qualitative and quantitative organic analysis[128,129] and this has led to the development of a flow injection procedure for use with FTIR analysis.

The use of flow injection in conjunction with IR analysis provides a simple and reproducible way to fill and clean the liquid cells. It also allows (i) the rapid automation of the determination and (ii) a strong reduction in the consumption of reagents and in the production of wastes.[130] The use of FTIR permits[121,131–134] an increase in the sensitivity, by accumulating a series of scans in the stopped-flow mode, the continuous monitoring of the baseline of the spectra and the accurate determination of the maximum of absorbance bands.

The recent development of FA procedures in the field of IR spectrometry has solved some of the many drawbacks of quantitative determination using this technique:[121,130–132,135] (i) repeatability and accuracy of the determinations have been improved;[5] (ii) the possibility of direct analysis of real samples has been enhanced;[136] (iii) easy and fast methods for the quantitative quality control of several compounds in the same sample have been provided[137] and (iv) the sensitivity of FTIR determination has been increased by means of on-line coupling with concentration techniques.[138]

The use of FA FTIR derivative spectrometry permits additionally the direct determination of compounds in real samples, without requiring previous separation or clean-up procedures and without problems in establishing the baseline[114] (Figures 5–7).

5.1 Fourier Transform Infrared Determination in Pharmaceuticals

IR spectrometry provides a useful way to identify drugs.[118,137–140] However, traditional techniques employed to obtain the IR spectra, such as alkali halide disks, mulls and thin films, are not adequate either for quantitative analysis, or for the use of transmittance spectra. For this reason, UV spectroscopy was for a long time, and is still currently, employed in the analysis of pharmaceuticals rather than IR. The use of flow cells for handling appropriate solvents and with well-selected spacers and windows opens the possibility of carrying out quantitative IR determinations of drugs, especially for the development of automated procedures based on FIA. Ibuprofen is a nonsteroidal anti-inflammatory analgesic that can be quantified by the combination of FIA/FTIR[136] (Figure 8).

FIA has been successfully used for the determination of acetylsalicylic acid, caffeine,[133] dimenhydrinates and paracetamol.[103]

Propyphenazone and caffeine are important analgesic and stimulant drugs respectively, which are commonly used and combined in pharmaceutical formulations. A procedure has been proposed for the simultaneous FTIR determination of propyphenazone and caffeine in commercial pharmaceutical formulations.[13] The method involves the dissolution of the active principles in $CHCl_3$, followed by filtration of the sample solution, to remove the other molecules than "active principles". Propyphenazone is determined by absorbance measurements at 1595 cm^{-1}, using a baseline established between 2000 and 890 cm^{-1}, and caffeine by using the first derivative values at 1712 cm^{-1}, using propyphenazone and caffeine standards for external calibration. The method was applied in both stopped-flow and flow injection modes, providing precise and accurate results for the quantitative analysis of samples (Figure 9).

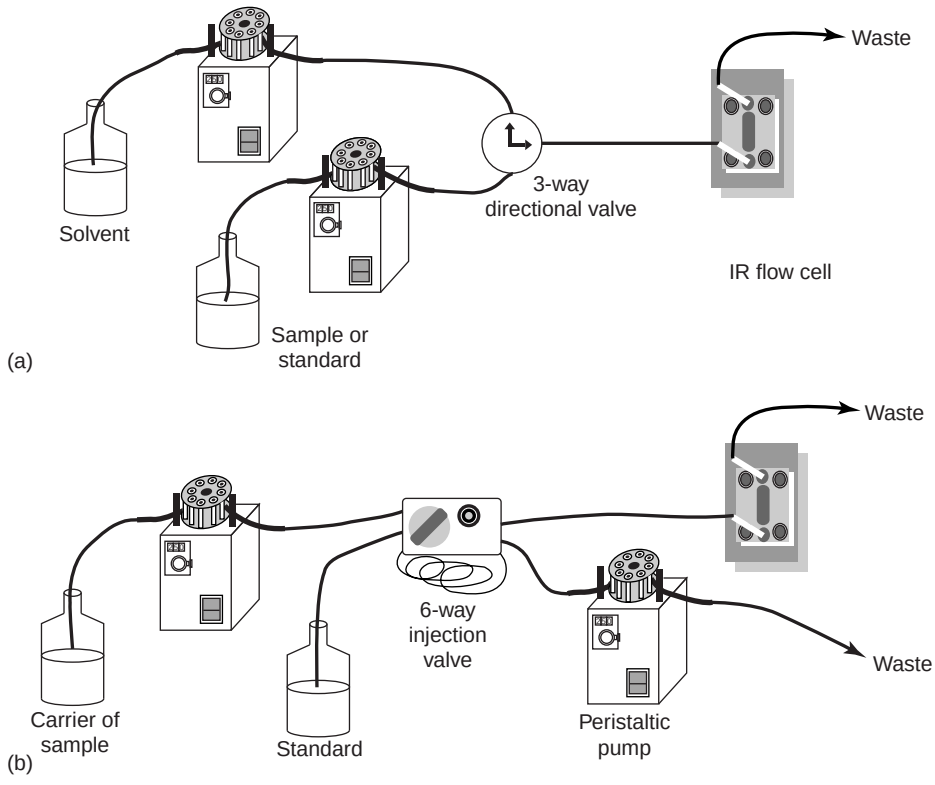

Figure 5 Manifolds employed for: (a) stopped-flow FTIR measurement, and (b) reversed FA.

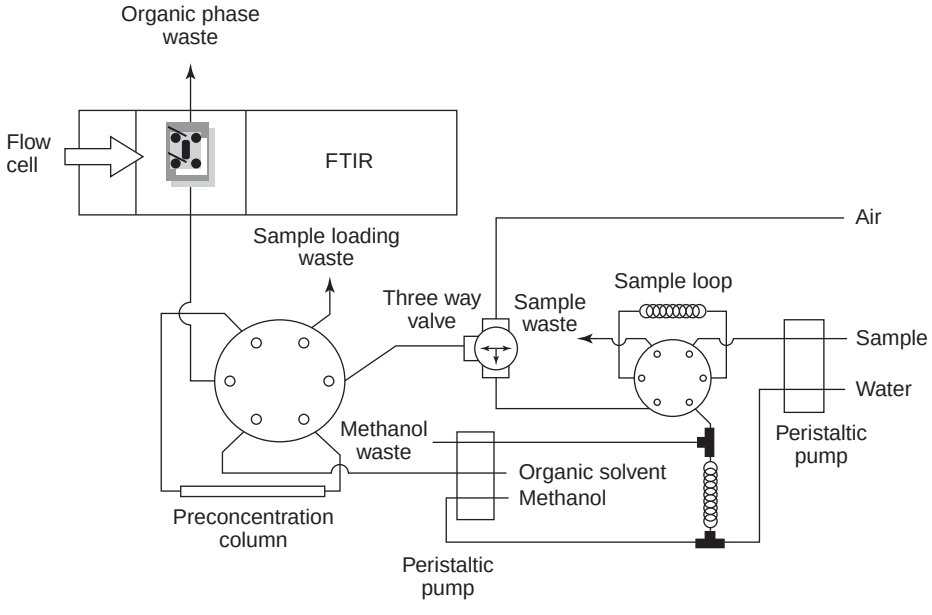

Figure 6 Manifold employed for on-line solid phase extraction and on-line elution of caffeine.

Bouhsain et al.[120] proposed the simultaneous determination of paracetamol, acetylsalicylic acid and caffeine. The simultaneous determination of these three active principles can be performed by using multivariate calibration. Various numerical methods were proposed in combination with molecular spectroscopy, such as Kalman filtering and UV data,[141] multiple linear regression of the NIR reflectance data[142] and multivariate programs applied to UV measurement such as multicomponent simplex analysis and multic,[143] iterative

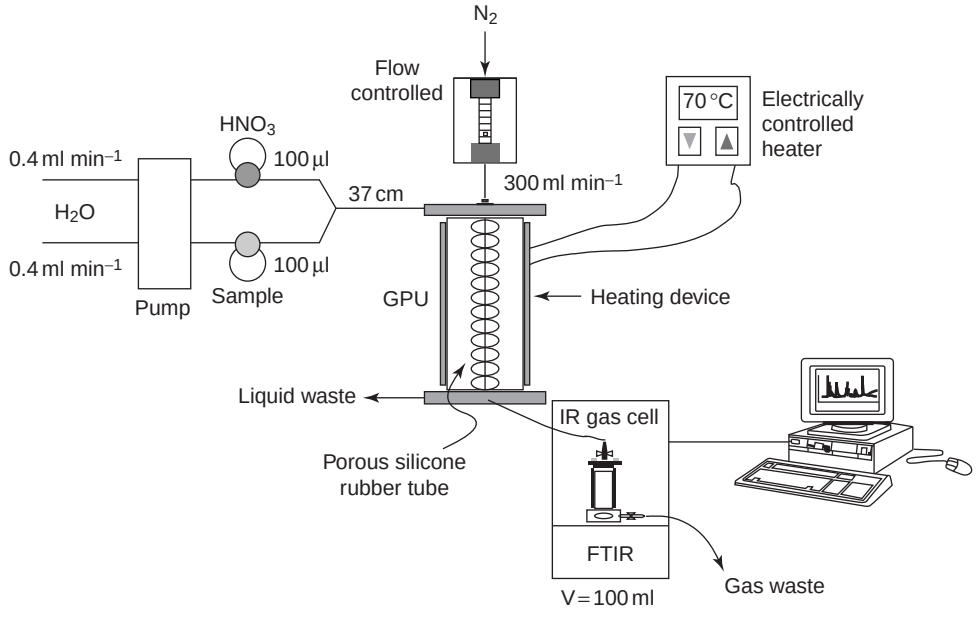

Figure 7 Manifold employed for vapor generation FTIR determination of carbonate in waters. Volumes and flow values are those corresponding to the best conditions of analysis. The porous silicone rubber tube has a 0.8 mm internal diameter, a 1.6 mm external diameter and a length of 60 cm.

target transformation factor,[143,144] step regression[145] and PLS regression.[120] The latter also developed a PLS/FTIR procedure for the fast and accurate simultaneous determination of paracetamol, acetylsalicylic acid and caffeine in commercial pharmaceutical formulations. This reduced preliminary sample treatment to the leaching of compounds to be determined from the pharmaceutical formulations with a 10% v/v solution of ethanol in dichloromethane in an ultrasonic water bath for 20 min, and providing the direct FTIR determination by using a simple matrix calibration including only eight standards.

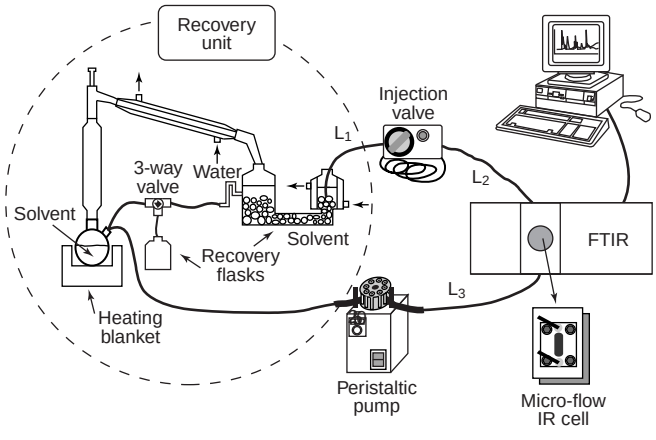

Figure 8 Manifold employed for the FIA/FTIR determination of ketoprofen and on-line solvent recycling. $L_1 = 55$ cm, $L_2 = 78$ cm and $L_3 = 43$ cm.

Garrigues et al.[146] have developed a method for the determination of ibuprofen in pharmaceuticals by FTIR, using the carbonyl band, which the compound presents at 1710 cm^{-1} in carbon tetrachloride solution.

5.2 Fourier Transform Infrared Determination in Petroleum Products

The determination of aromatic hydrocarbons in naphthas and motor gasoline is necessary in several areas of the petroleum, petrochemical and related industries.[147] Detailed information on the aromatic composition of fuel materials, intermediates and commercial products is required for process development and quality control programmes. Legislation in several countries limits the upper concentration of some hydrocarbons in commercial gasoline.

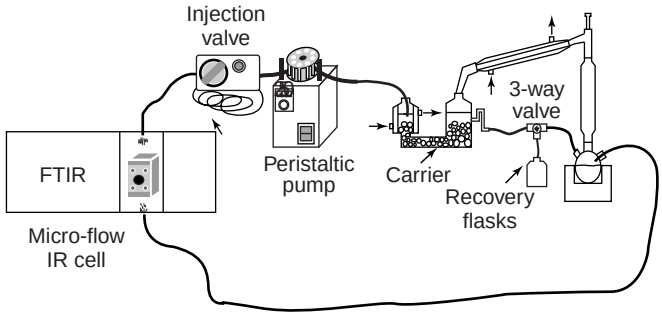

Figure 9 Manifold employed for the clean simultaneous FTIR determination of propyphenazone and caffeine.

FTIR spectrometric techniques are known to provide a rapid means for the direct determination of aromatic compounds.[148] Gallignani et al.[4] proposed a FTIR procedure for the automated determination of benzene in gasoline, based on the use of FIA. The method permits the direct determination of benzene without any pretreatment of samples, with a limit of detection of 0.002% (v/v). Results found by direct analysis agree well with those obtained by off-line and on-line standard addition methods. A rapid quality control procedure was developed, based on the on-line injection of 1 : 10 diluted gasoline samples into a carrier stream of a 0.5% (v/v) solution of benzene in hexane; the baseline corresponds to the upper limit tolerated by the European Economic Community's law for gasoline. On the other hand, the American Society for Testing and Materials (ASTM) has proposed the IR batch determination of benzene in motor and aviation gasoline.[149]

Methyl *tert*-butyl ether is an antiknock agent commonly employed in unleaded gasoline.[114] The additive is frequently determined by GC in gasoline, in synthetics products and in estuarine waters and sediments. A NIR method has been applied to the direct measurement of methyl *tert*-butyl ether in gasoline with an associated standard error of 0.1% (v/v) by using the second overtone NIR region.[150]

Daton et al.[151] used a batch method for the determination of this compound in gasoline based on direct absorbance measurement at 851 cm^{-1} by using ATR. They obtained a relative standard deviation that ranged between 1 and 1.6% for the analysis of samples containing 7% (v/v) of methyl *tert*-butyl ether.

De la Guardia et al.[114] proposed a method for the flow injection FTIR spectrometric determination of methyl *tert*-butyl ether in unleaded gasoline. The method is based on the use of first-order derivative measurements in a flow system. The recovery of methyl *tert*-butyl ether in spiked natural samples varies from 98.2 to 102% and the relative standard deviation was 0.6% for samples containing 0.8% (v/v) methyl *tert*-butyl ether. Results obtained in the analysis of commercial gasoline samples compare well with those found by GC.

The determination of organic compounds in complex matrixes suffers from serious limitations owing to the overlapping of absorbance bands, which prevents direct determinations of a series of compounds and causes strong matrix interference.

The use of derivatives techniques in spectrometry has opened up new possibilities for the resolution of mixtures and compensation for matrix interference.[152–154,131]

Petrol is a complex mixture of volatile flammable liquid hydrocarbons containing hundreds of different natural components and some additives.

Gallignani et al.[9] proposed the simultaneous FA FTIR determination of benzene, toluene and methyl *tert*-butyl ether.

Kerosene is an important fuel for practical purposes. Under different names or with slightly different characteristics, refined kerosene is the selected fuel for civil and military use in jet-fuel turbine aircraft. Therefore it is an important task to assure the kerosene quality. The main kerosene physicochemical characteristics are covered the British Ministry of Defence specification[155] and/or the ASTM guides.[156] Both of them comprise a minimum of 26 tests to be carried out by refinery laboratories and/or external surveillance companies.

Several aircraft fuel quality properties like density, freezing point, flash point, aromatic content, initial boiling point, final boiling point and viscosity have been predicted from the FTIR spectra in the range 4000 to 600 cm^{-1}, using multivariate techniques. Multiple linear regression, PCR and PLS models have been tested. FTIR/PLS repeatability and reproducibility values fall well within the ASTM ranges and the standard error of prediction values is good. Sample manipulation was improved by using a stopped-flow system.[5]

5.3 Fourier Transform Infrared Determination of Industrial Products

Ethanol and also amylene are added to analytical reagent-grade chloroform as stabilizers in order to avoid the decomposition of this chlorinated solvent and the formation of phosgene.[157] The determination of ethanol in chloroform can be carried out by chromatography. However, IR spectrometry, in the near range, offers exciting possibilities for this type of analysis due to the transparency of chloroform and the presence of characteristic bands of ethanol in this region of the spectrum. In the last ten years, a big effort has been made in the development of on-line procedures for process monitoring, based on NIR measurements,[158–160] generally using a combination of chemometrics and fiber-optic probes. Working at a fixed wavelength[4] or using the stopped-flow approach, it is possible to develop the FIA/NIR procedure for a variety of aqueous[9] and nonaqueous[4] samples. López-Anreus et al.[160] propose a simple and direct FIA procedure for the determination of the stabilizing agent ethanol in chloroform samples. The procedure is based on the use of an absorbance band of ethanol in the NIR region at 2272 nm, measured in front of a reference sample of chloroform stabilized with amylene. The method developed provides a limit of detection of 0.0045% (v/v) and a dynamic range up to 10% (v/v). The variation coefficient of six independent analyses of a real sample containing approximately 1% (v/v) of ethanol was as low as 0.4%, while the injection frequency can reach 78 samples per hour.

Polymers are present in a large number of manufactured products and the quality of a polymer is established from different parameters: viscosity, end-group analyses, branching percentage, cross-linking degree, degree of crystallinity, density and molecular weight, which are characteristic of each lot.

Polymers have been investigated by MIR using transmission data[161] and also using NIR.[28] De la Guardia et al.[162] developed a simple procedure for the determination of some average properties of nonylphenol-ethylene oxide condensates, such as average molecular mass, average degree of condensation, the hydrophilic/lipophilic balance (HLB) and the percentage of ethylene oxide, which are characteristic of this kind of surfactant condensate. The method, based on the existence of a regression between the logarithm of the surfactant properties and the logarithm of the ratio of the peak heights of the bands at 960 and 840 cm^{-1}, is precise and accurate. Variation coefficients as low as 0.9% for HLB, 1% for ethylene oxide percentage and 1.6% for molecular mass were obtained.

For the determination of polymers and composite materials by IR, some of the limitations come from the use of extreme temperature and pressure conditions. However, recent developments, such as that proposed by Hanssen and Snail,[163] who developed a diffuse gold integrating sphere which can support up to 250 °C, and the off-the-shelf bidirectional reflectance accessories with controlled atmosphere, which can be used at 500 °C, offer new possibilities for operating in extreme experimental conditions.

5.4 Fourier Transform Infrared Determination of Products of the Paint Industries

Xylol, a solvent commonly used in the paint industry,[48,164,165] is a mixture of *o*-, *m*- and *p*-xylene. Garrigues and de la Guardia,[102] proposed a series of mathematical models for the FTIR determination of *o*-, *m*- and *p*-xylene in xylol without the need to use absorption cells with a known optical path length. The relative errors found in the analysis of binary mixtures are of the order of 1.4%, and those found for ternary mixtures are lower than 3% in most instances, for samples containing from 10 to 75% m/m of each component. In 1992, the same team proposed a FIA/FTIR method for the determination of xylene. The best limit of detection obtained was 0.01% (v/v). The relative standard deviation was 1% for a sample containing 0.4% of *o*-xylene in hexane.

Solvents are necessary in paint formulations in order to improve the solubility of resins and polymers and to improve the fluidity of the final product. Organic solvents traditionally employed in paint manufacture include carbonyl and aromatic compounds. However, the use of water as a paint solvent has always been attractive for reasons of low cost and absence of toxicity risks.[121] Water-borne paints need a minimum application temperature of 5 °C adequate film formation and, because of that, atmospheric conditions can delay its drying time. For this reason, these paints need the addition of a co-solvent at low concentrations; the butylglycol being one of the most commonly employed.[166]

The analysis of solvents employed in paint industry is usually carried out by GC.[167,168] However, this procedure is slow and time-consuming and not very appropriate for the direct analysis of aqueous samples. The use of FTIR allows the direct determination of butylglycol in water-borne paint solvents.[6] The method is based on the injection of 2 μL of untreated paint solvent into an electrically heated Pyrex glass reactor in which the sample is vaporized at 150 °C. The vapor phase generated is transported by means of a nitrogen carrier flow of 400 ml min^{-1} into an IR multiple pass gas cell and the spectrum in the MIR is recorded between 2500 and 600 cm^{-1} as a function of time. Data found for samples are interpolated in two calibration curves obtained by injecting different volumes of pure butylglycol and pure water. The measurement of the area of the flow injection recording obtained from the absorbance of the transient signals in the range 1137–1131 or 1071–1065 cm^{-1} for butylglycol and 1660–1640 cm^{-1} for water allows the direct determination of these two compounds in the same sample. The method provided relative standard deviation values around 1%, a limit of detection of the order of 2.7–4.5 ng, recovery percentage values from 96 to 104% and a sampling frequency of 65 h^{-1}.

Paint solvents generally contain aromatic and carbonyl compounds with a high volatility. In this sense, acetone is nowadays one of the most important solvents in the paint industry, owing to its high solvent power for a wide range of resins and its economic price. The combination between alcohols and acetone, in different proportions, improves its solvent power and, because of that, acetone and isopropanol mixtures are commonly used in nitrocellulose paint formulations.[165] Vapor generation FTIR provides a way for direct determination of acetone and isopropanol in nitrocellulose paints. Spectra were recorded between 2000 and 500 cm^{-1}. Data found for samples were interpolated in two calibration curves obtained by injecting different volumes of pure acetone and isopropanol. The measurement of the height and area of the flow injection recording obtained from the absorbance of the transient signals in the range of 1743–1737 cm^{-1} and 956–950 cm^{-1} allowed the direct quantification of the solvents. The limit of detection of the method was of the order of 0.5–3.9 ng, with a relative standard deviation from 0.09 to 2.8%. The solvent composition in a nitrocellulose paint can be determined by the aforementioned procedure after a

simple treatment consisting of sample dilution with CCl_4 followed by centrifugation at 4000 rpm for five minutes.[7]

A method has been developed for the direct simultaneous determination of *o*-, *m*- and *p*-xylene in xylol by FTIR.[7] The method is based on the injection of samples, diluted in hexane, into a carrier solution of hexane. The absorbance measurement, were carried out at 743, 770 and 796 cm^{-1}. The limit of detection corresponds to 0.03% v/v for *o*- and *m*-xylene and to 0.02% v/v for *p*-xylene, using a 0.13 mm path length micro flow cell.

5.5 Fourier Transform Infrared Determination of Products in the Field of Health

The determination of ethanol in blood is very important, especially in establishing responsibility in traffic accidents. The upper permissible concentration level of ethanol in blood for drivers depends on different national laws: for example, in France, the upper limit is 0.5 g of ethanol per litre of blood. The determination of ethanol in blood is usually carried out, after previous treatment, by headspace GC using a mass-spectrometric detector or a flame ionization detector, or by using a specific instrumentation. Recently, Pérez-Ponce et al.[115] show that this kind of determination can be carried out by FTIR through direct injection of samples in a vaporizer and measurement in the vapor phase, in order to avoid the interference of the sample matrix components. The measurements were made at 1050 cm^{-1} to increase the selectivity. This band corresponds to the C–O stretching vibration in ethanol and has been demonstrated to be useful for breath analysis.[169] The method is based on the injection of a discrete sample volume of 10 μL into an electrically heated Pyrex glass reactor. The ethanol is vaporized in the reactor at a temperature of 90 °C and introduced by means of a N_2 carrier flow inside a long-path IR gas cell. The corresponding FA is recorded as a function of time. With the transient signal obtained in the range 1150–950 cm^{-1}, it is possible to quantify ethanol up to 2 g L^{-1}, with a limit of detection of 0.02 g L^{-1}, and an analysis frequency of 40 samples per hour. This method can be applied to a single drop of finger blood.

IR spectrometry is very useful as a detector in liquid chromatography (LC) to solve complex analyses requiring simultaneous chromatographic and spectrometric discrimination, and provides additional qualitative information not available from conventional LC detector.[170–172]

FTIR spectrometric detection in LC can be accomplished in two ways: (i) on-line detection via the use of flow cells[173,174,228] and (ii) off-line detection after solvent removal.[175,176] The first approach is simpler, but presents limitations regarding excessive absorption caused by the carrier solvents and thus, in general, is less employed. The second approach eliminates the solvent problems but involves the use of expensive interfaces and, because of that, UV/VIS absorbance spectrometry or molecular fluorescence remains one of the most widely used spectrometric detection techniques in LC.[177]

LC has been employed for the analysis of oils and greases, providing a good way of determining fatty acids and triglycerides.[178–180] In some studies, FTIR spectrometry has been employed for the determination of triglycerides, in both normal and reversed-phase,[181] estolide triglycerides,[182] free acids and partial glycerides.[183]

Sterols are the main constituents of the unsaponifiable matter and represent an important group of minor components of oils and greases. Stearic acid, cholesterol and cholesteryl palmitate have been separated in less than 14 min by LC coupled with FTIR spectrometry.[184] General detection was carried out at 2868 or 2855 cm^{-1} and selective detection at 3285 cm^{-1} for stearic acid and 3526 cm^{-1} for cholesterol. In all cases, a baseline was established at 3900 cm^{-1}. The method was validated by application to the determination of cholesterol in animal greases and fish oils. Samples that have previously undergone saponification with KOH were extracted with petroleum ether, redissolved in chloroform and the final solution injected into the chromatographic system. The recoveries were 95–99% for samples spiked with cholesterol and 100–102% for samples spiked with cholesteryl palmitate.

Methanol, originally obtained by the destructive distillation of wood, is nowadays usually manufactured from hydrogen and carbon monoxide or carbon dioxide and also by the oxidation of hydrocarbons. These methods of manufacture and the low taxes imposed on this alcohol, as compared with those on ethanol, make methanol a cheap and dangerous substitute for ethanol. The ingestion, inhalation or skin absorption of methanol can lead to anything from headaches to blindness and even death (the latter from ingestion of 100–200 mL). Therefore, (the latter) analytical control in both beverages and cosmetics is necessary. Colorimetry, GC and immersion refractometry are the three methods recommended by the Official Methods of Analysis of the Association of Official Analytical Chemists (AOAC)[185] for methanol determination in beverages. However, all these methods need the previous treatment of samples, which is, in general, tedious and time-consuming. IR spectrometry has been employed in the identification or quantification of methanol in a series of different samples. Open path FTIR can be employed for qualitative monitoring of methanol vapor in the atmosphere.[186] Quantitative methods have also been proposed for the NIR second derivative determination of methanol in synthetic mixtures of methanol, butanol, acetone, methylene chloride and dichloropropane.[187]

Similarly, methods were proposed for FTIR determination of mixtures of 2,2-dimethoxypropane, acetone and methanol,[188] which is especially useful in the simplex optimization for FTIR remote sensing measurement of SF_6, methanol, acetone and 2-butanone in synthetic mixtures.[189]

For the analysis of real samples, optical fiber NIR spectrometry was used for methanol determination in gasoline.[190] IR gas-filter correlation spectrometry was employed to measure methanol in vehicle exhaust.[191] FTIR was employed for the remote analysis of CO, CO_2 and methanol in motorboat exhausts.[192]

The determination of methanol in the presence of ethanol can be carried out by GC[193–195] and also by GC/FTIR and mass spectrometry (MS).[196] The IR emission from a flame has been also proposed as a means of detection for LC to analyze mixtures of methanol, ethanol and propanol.[197]

The filtering correlation function "*compare*" from Perkin Elmer has been proposed for the discrimination of spectra of mixtures of methanol and ethanol,[198] and thermal lens spectrometry has been employed for the partial quantitative determination of methanol and ethanol in gaseous samples.[199]

Kuehl and Crocombe[200] have established a model for the quantitative analysis of methanol, ethanol and acetone in fermentation broth, showing that by using an ATR cell, synthetic mixtures of the three compounds under study can be analyzed with relative errors below 2%.

Garrigues et al.[201] proposed a FTIR procedure for the direct determination of methanol and ethanol in liquid samples such as alcoholic beverages and eau-de-Cologne. It is based on vapor generation from small injected volumes of untreated samples into a heated Pyrex glass reactor in which ethanol and methanol are volatilized, at a temperature of 80 °C, and introduced into a long-path IR gas cell by means of a N_2 carrier flow. The spectra obtained present two characteristic ethanol bands (1050 and 880 cm^{-1}) and a single characteristic methanol band (1030 cm^{-1}). The measurement of the area of the transient recording obtained for the wavenumber ranges 1025–950 and 950–820 cm^{-1} allows the determination of ethanol and methanol in the same sample by using a simple proportional equation approach. The limit of detection values found for an injection volume of 1 μL were 0.21% v/v for ethanol and 0.04% for methanol, and the relative standard deviation values were between 0.5 and 1.4% for five analyses of samples containing between 30–80% (v/v) ethanol and 0.5–3.5% (v/v) methanol. Acetone, ethyl acetate, 1-propanol and 2-propanol do not interfere with this determination when they are present at concentration levels of the order of those of methanol.

5.6 Fourier Transform Infrared Determination of Compounds in the Natural Environment

The standard protocol for oil and grease determination in polluted waters is a gravimetric procedure based on a liquid–liquid extraction (LLE) with Freon 113.[202] The pioneering studies on IR spectrometric determination of oil and grease in waters were based on CCl_4 extraction.[203] However, the quantitative extraction of all types of oils like hydrocarbons, vegetable oils, animals fats, waxes, soaps, greases and related matter is difficult. Serious problems have been reported for the total extraction of several materials from waters, including gasoline, for which extraction yields of about 50% were found.[204] The use of chlorinated hydrocarbons, like freon and CCl_4, is undesirable because of their ozone-layer depleting effect.[205] Thus it is an urgent task to search for alternative procedures for oil and grease extraction from waters.[206,207]

The use of microwave ovens offers a fast way of improving analytical procedures that include a thermal step. Microwave-assisted procedures have being developed for sample digestion[208] and derivatization.[209] More recently, the use of a microwave-assisted technology has been proposed for the extraction of organic pollutants from sediments for their subsequent analysis by chromatography.[210,211]

In 1997, Daghbouche et al.[184] evaluated LLE, solid phase extraction and microwave-assisted extraction for FTIR spectrometric determination of oil and grease in polluted waters. A fast microwave-assisted quantitative procedure was proposed for the total extraction of oil and grease into CCl_4 from water samples, with an irradiation time of 1 min, at an exit power of 520 W, using closed polytetrafluoroethylene (PTFE) vessels. The FTIR quantification of oils was carried out by the measurement of the area under the absorbance spectra in the range between 3058 and 2780 cm^{-1}, using a baseline established between 3200 and 2700 cm^{-1}. The analytical features of the method and those of LLE/FTIR spectrometry were evaluated and samples spiked with different types of oils examined.

The environmental impact of laboratory practices is not insignificant. Large amounts of toxic compounds are produced every day in research and application laboratories. The storage and destruction of dangerous pollutants contained in laboratory wastes needs adequate decontamination plants, and thus the elimination of waste is expensive and involves strict safety requirements. So it is necessary to substitute traditional methods by clean analytical technologies.[212] FIA provides an interesting way to achieve reduction of reagent consumption and wastes. Using FIA/FTIR it is possible to offer an alternative to the traditional methods, providing a

drastic reduction of pollutant side effects. A semi-micro distillation unit was therefore incorporated into a simple FIA/FTIR manifold for the on-line recovery of the solvent used as carrier. This system was applied to the direct, simple and fast determination of ketoprofen,[124] a nonsteroid anti-inflammatory drug with analgesic and antipyretic properties.[213] The system permits a drastic reduction of reagent consumption and easy and fast sampling and cleaning of the measurement cell, as compared with the official method in the British Pharmacopoeia for ketoprofen determination, which is based on the measurement of the absorbance at 250 nm using methanol as solvent.[214] For the quantification of ketoprofen in pharmaceuticals, the carbonyl bands at $1712\,cm^{-1}$ and $1666\,cm^{-1}$ were employed and the developed method provides a limit of detection of $0.04\,mg\,ml^{-1}$ (3s) and a dynamic range which extends up to $10\,mg\,ml^{-1}$. Typical coefficient of variation values ranged between 1.6 and 3%, with a sample injection frequency greater that 42 per hour. Accurate and precise results were obtained in the analysis of real pharmaceutical samples.

Another example of an important reduction in the side effects of analytical procedures is given by the quantification method of raw sugarcane sugars proposed by Cadet et al.,[98] which avoids the previous clarification of samples required by usual methods of sugar determination (HPLC, polarimetry, GC). This step of clarification needs a large quantity of lead acetate, a strong pollutant which, when released into the environment can pollute surface waters.

There are several methods proposed for the determination of carbonate and bicarbonate in waters, based on both instrumental and volumetric procedures. The official method is based on a simple titrimetric procedure, but a series of problems have been reported with the end-point determination and evaluation of titration data.[185]

There are only few precedents for the IR determination of inorganic carbon in water by batch and FIA. One of these is that proposed for seawater analysis, based on batch acidification with H_3PO_4 and colorimetric or IR determination of CO_2 stripped with N_2.[215] Another is based on the use of a patent cylindrical thin-film reactor with an actively rotating inner cylinder and a central low-pressure mercury lamp. This allows the decomposition of the organic molecules through UV radiation and that of inorganic carbonates by acidification. The release of the CO_2 is carried out by a N_2 flow and its determination is done by IR.[216] More recently, Pérez-Ponce et al.[217] proposed the use of an electrically heated gas-permeation unit (GPU) for the vapor-generation FTIR of total carbonate in waters. The developed method is based on the simultaneous injection of 100 μL of sample and 100 μL of nitric acid in a two-channel manifold with a merging zone. This zone provides the generation of CO_2 through the GPU, heated at 90 °C, from which a stream of N_2 sweeps the CO_2 to a long-path IR gas cell. Absorbance measurements were made in the range from 2500 to $2150\,cm^{-1}$ and the corresponding flow injection recordings were obtained as a function of time. The method provides a limit of detection of 4.6 ppm (expressed as mg of $HCO_3^-\,L^{-1}$) and a sampling frequency of $30\,h^{-1}$. Results obtained for a series of natural waters compare well with those obtained by a titrimetric reference method.

More recently, these authors[14] proposed a microwave-assisted vapor-generation FTIR spectrometric procedure for the determination of total carbonate in waters. The method is based on the simultaneous injection of 100 μL of sample and 300 μL of 0.1 M nitric acid into a two-channel manifold with a merging zone; a coil located inside a microwave oven favors the removal of CO_2. The carbon dioxide is separated from the distilled water employed as a carrier, by means of gas–liquid separator, and is introduced by a nitrogen carrier flow inside a long-path IR absorption gas cell. This provides direct quantification of total carbonate concentration in the range up to 1000 mg $HCO_3^-\,L^{-1}$ with a sampling frequency of 15 samples per hour. The procedure has a limit of detection of 15 mg $HCO_3^-\,L^{-1}$ and a variation coefficient of 1.3% for five independent analyses.

5.7 Quantitative Fourier Transform Infrared Analysis of Compounds at Trace Levels

The use of FTIR for trace analysis is problematic, due to the lack of sensitivity of this technique and the use of small path length values. Substantial efforts were undertaken in order to improve the limit of detection of FTIR, e.g. by using alternative approaches such as the stopped-flow technique.[11,133] Nevertheless they do not provide adequate solutions. Therefore other strategies have been investigated to increase the sensitivity of IR determinations.

The excellent results obtained by on-line preparative concentration of metallic elements by FA and atomic spectrometric techniques[218] offers a model strategy for the improvement of the sensitivity of FTIR determinations. As an example for the determination of pesticides in waters, solid-phase extraction offers an excellent way for improving the limit of detection of FTIR. This strategy was selected because it allows easy on-line preparative concentration and does not require high volumes of organic solvents or the use of complicated phase-separation manifolds.[219–221] Garrigues et al.[138] proposed a new method for the FTIR determination of carbaryl, a pesticide found in natural waters which was

determined at 1746 cm^{-1}. Their procedure permits (i) in-field sampling, (ii) stabilization of pesticides and (iii) fast determination of carbaryl. The sensitivity obtained with this procedure corresponds to 0.044 area units $\mu g^{-1}\,cm^{-1}$ and with a limit of detection 15 $\mu g\,L^{-1}$ of carbaryl, using C18 cartridges for preparative concentration and dichloromethane for on-line elution. The method could be applied to other types of components, providing new possibilities for the application of FTIR to trace analysis.

5.8 Other Fourier Transform Infrared Coupled Methods

SIA was introduced in 1990 by Ruzicka and Marshall[222] and has improved classical FIA in many ways, especially concerning the drastic reduction of reagent consumed and waste generated and the robustness of the resulting analysis system. In classical FIA, the generation of different injection sequences would require a physical rearrangement of the FIA manifold for each sequence. This situation is different in the case of SIA, where only changes in the controlling software are required to create different interpenetrating dispersion profiles between samples and reagents. Schindler et al.[171] employed for the simultaneous determination of glucose, fructose and sucrose in aqueous standards as well as in soft drinks using FTIR detection and multivariate PLS data analysis. Obtained results were compared with those found by an external reference method. Typical deviations from the results were of the order of 2.5%, 4.1% and 2.1% for glucose, fructose and sucrose respectively (Figure 10).

Vonach et al.[135] proposed the modulation of the pH as a novel approach for the determination of phosphate with FIA and FTIR detection.

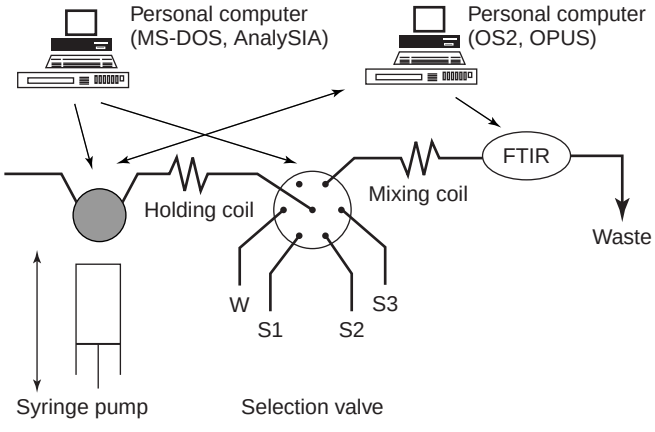

Figure 10 SIA/FTIR manifold: holding coil, 0.5 mm i.d., 250 cm; S1-3, standard solutions; W, water; mixing coil, 0.3 mm i.d., 30 cm. (Reproduced with permission from Lendl et al., *Analytical Chemistry*, **69**(15), 2877–2881, © 1997, American Chemical Society.[226])

The use of FTIR as a detection system in HPLC is of particular interest in analytical chemistry. In particular, the fingerprint region around 1400–900 cm^{-1} provides valuable data for the confirmation of molecule structures, since most organic (and many inorganic) compounds exhibit strong and narrow absorption bands, providing a high degree of qualitative and structural information.

As has been indicated in previous sections, the coupling of HPLC and FTIR can be performed by using a standard flow cell, without mobile phase separation and by introducing an interface that evaporates the eluent and deposits the analytes into a medium compatible with FTIR detection. Another approach for solvent elimination, which is particularly interesting for aqueous solutions, is based on post-column on-line extraction. After extraction with chlorinated solvents (CH_2Cl_2, $CHCl_3$, CCl_4), various interfaces can be applied; regular flow cells;[223,224] loading of the extracts on KBr, with subsequent DRIFT detection;[225] or, more recently, deposition on a zinc selenide window by a spray jet assembly, followed by the analysis of the residues with a FTIR-microscope.[176]

Vonach et al.[135] used FTIR as a molecular-specific detection system for HPLC in an aqueous phase, and developed a method for sugar determination in beverages. The separation was achieved with an isocratic HPLC set-up using an ion exchange column. The FTIR detection of the C–O bands in MIR, between 1000 and 1200 cm^{-1}, was performed in real time with a 25 ml flow cell without elimination of solvent. Characteristic FTIR spectra of sucrose, glucose and fructose were recorded during the separation. The calibration of these sugars in the 5–100 $mg\,mL^{-1}$ range resulted in a linear correlation with a standard deviation of the method of 0.11, 0.07 and 0.11 for sucrose, glucose and fructose respectively. The method was found also to be suitable for the identification and the quantification of minor components in beverages, such as taurine (4 $mg\,mL^{-1}$) and ethanol (0.4 $mg\,mL^{-1}$).

Lendl et al. in 1997[226] developed a flow system containing a micromachined lamella-type porous silicon reaction and a novel MIR fiber-optic flow cell, which was used for the enzymatic determination of sucrose in aqueous solutions. The method relies on the enzymatic hydrolysis of sucrose catalysed by β-fructosidase and on the acquisition of FTIR spectra before and after complete reaction. The enzyme was covalently bound to the porous silicon reactor in a HF/ethanol mixture. For the measurement of small amounts of aqueous solution, a miniaturized flow cell was developed which consisted of two $AgCl_xBr_{1-x}$ fiber tips (0.75 mm diameter) coaxially mounted in a PTFE block at a distance of 23 mm. The flowing stream was directed through the gap between the two fiber tips, which served to define the optical path length and to bring the focused MIR radiation to

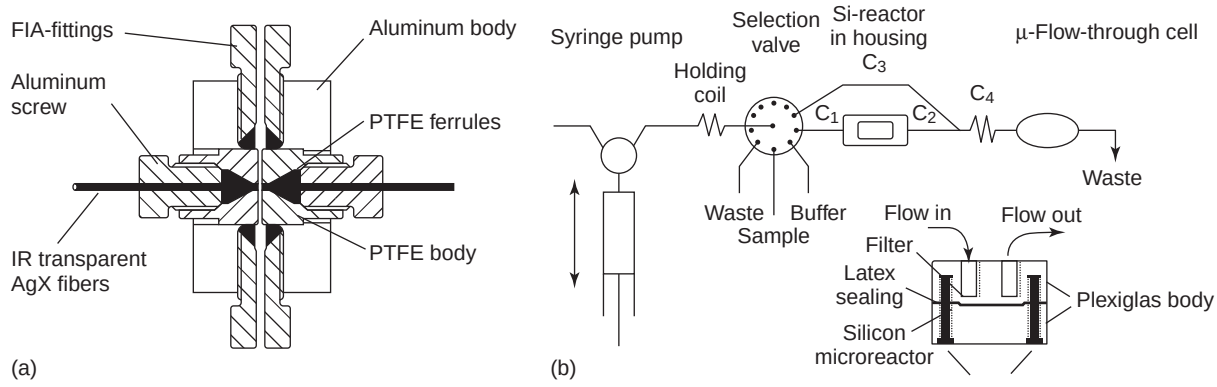

Figure 11 (a) Manifold used for the FTIR detection in miniaturized total analysis systems for sucrose analysis. (b) Sketch of the developed miniaturized fiber-optic flow cell allowing for the acquisition of a FTIR spectrum of 10 nL. The diameter of the AgC_xIBr_{1-x} was 750 mm. (Reproduced with permission from Lendl et al., *Analytical Chemistry*, **69**(15), 2877–2881, © 1997, American Chemical Society.[226])

the place of measurement. Using this set-up, a probed volume of about 10 nL was obtained. The calibration curve was linear up to 10 $mmol^{-1}$ sucrose. Furthermore, the potential of this method was demonstrated by the analysis of binary sucrose/glucose mixtures showing no interference from glucose, and by the successful determination of sucrose in real samples (Figure 11).

The introduction of FTIR as a detector in miniaturized total analysis systems represents a significant step towards the development of techniques capable of obtaining molecular-specific information in miniaturized flowing streams. The huge amount of chemical information contained in FTIR spectra was exploited to reduce the number of enzymatic reactions needed for the determination of sucrose in aqueous standards as well as in real samples such as soft drinks. Both the development of a new molecular specific detector based on a miniaturized IR fiber optic flow cell and the application of a silicon microreactor with in-situ fabricated porous silicon layers as a surface-enlarging matrix for immobilized enzymes, can be considered as some of the most interesting uses of FTIR.

6 LIMITATIONS AND PERSPECTIVES OF INFRARED SPECTROSCOPY FOR QUANTITATIVE ANALYSIS

IR is probably the most significant range of the spectrum in order to provide an appropriate identification of a molecule and, because of that, IR spectrometry has been widely employed to confirm the nature of compounds and particularly to evaluate pure products.

However, the high absorption of IR radiation by water and the presence of a large number of characteristic bands for each functional group inside a molecule decreases sensitivity (notably with the use of small band-pass cells) and selectivity. These problems must be solved in order to obtain accurate quantitative data.

In addition to the pioneering studies carried out during the 1950s, the new age of quantitative analysis through IR spectrometry is based on two important developments in the field of analytical chemistry: chemometrics and FA.

Chemometrics has provided the mathematical tools to solve overlapping problems, which can be found in the interpretation of spectra in NIR and MIR, and to solve the serious problems presented by the influence of the sample environment in both the size and shape of the characteristic bands of analytes. As has been discussed in section 2, the use of multivariate calibration strategies, and specially PCR and PLS, provides the best way to solve these kinds of problems. It offers exciting possibilities for multicomponent determinations, through the use of a reduced number of previously analyzed samples as a calibration set and the prediction of sample concentrations. Through PCR and PLS data treatment it is possible to obtain accurate data for the concentration of several analytes in the same sample and also an estimation of physical properties.

The second milestone in the recent development of quantitative analysis by IR is the use of FA strategies, which improves sample introduction and method automation, and also reduces drastically the side effects of the use of IR-transparent solvents, which are frequently toxic. However, the main contribution of FA to the improvement of quantitative determinations by IR are the possibilities offered by the on-line coupling of derivatization procedures and especially for the on-line preparative concentration and elution of analytes. This increases drastically the limit of detection of the analytical procedures and offers an easy way to solve the problem of some matrix or intermolecular interference.

In the next few years there is expected to be a significant growth of the scientific literature on quantitative methods of determination by IR, both for trace analysis and for the screening of the main components of foods, biological and environmental samples. In this sense the new generation of low-cost FTIR instruments configured for NIR, MIR and FIR, and the development of new methodologies for direct sample analysis, will have a synergistic effect on the use of IR as a tool for routine analysis.

New developments in the remote sensing of air pollutants and recent advances in new strategies for enzymatic analysis and vapor generation-FTIR will be generalized in the following years. This will be achieved through the (i) introduction of fast and robust standardization methodologies, (ii) the use of enzymes in nonaqueous systems and (iii) the on-line treatment of solid samples for a fast screening of volatile organic compounds.

In short, we are at the beginning of a new area of vibrational spectrometry, which will clearly evidence and explore the possibilities of IR for quantitative determination of major, minor and trace components by incorporating the advances in chemometrics and automation.

ACKNOWLEDGMENTS

The authors are grateful to Bernard Offmann, Philippe Rondeau, Sandrine Sers, Christine Robert, Claude Rouch, Marie Renée Thiaw-wing and Salvador Garrigues, who contributed valuable suggestions on this review.

ABBREVIATIONS AND ACRONYMS

AOAC	Association of Official Analytical Chemists
AOCS	American Oil Society
ASTM	American Society for Testing and Materials
ATR	Attenuated Total Reflectance
CLS	Classical Least Squares Model
DRIFT	Diffused Reflectance Infrared Fourier Transform
FA	Flow Analysis
FIA	Flow Injection Analysis
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GPU	Gas-permeation Unit
HLB	Hydrophilic/Lipophilic Balance
HPLC	High-performance Liquid Chromatography
ICUMSA	International Commission for the Uniform Methods of Sugar Analysis
ILS	Inverse Least Squares
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LLE	Liquid–Liquid Extraction
MIR	Mid-infrared
MS	Mass Spectrometry
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
PAS	Photoacoustic Spectrometry
PCA	Principal Component Analyses
PCR	Principal Component Regression
PLS	Partial Least Squares
PTFE	Polytetrafluoroethylene
SIA	Sequential Injection Analysis
USDA	United States Department of Agriculture
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis • Glycolipid Analysis

Chemical Weapons Chemicals Analysis **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Near-infrared Spectroscopy in Food Analysis

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Near-infrared Spectroscopy in Analysis of Crudes and Transportation Fuels

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Pulp and Paper **(Volume 9)**
Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Chemometrics **(Volume 11)**
Chemometrics • Multivariate Calibration of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

General Articles **(Volume 15)**
Quantitative Spectroscopic Calibration

REFERENCES

1. B.G. Osborne, 'Principles and Practice of Near Infrared (NIR) Reflectance Analysis', *J. Food Technol.*, **16**, 13–19 (1981).
2. B.G. Osborne, T. Fearn, *Near Infrared Spectroscopy in Food Analysis*, Longman Scientific and Technical, UK; John Wiley and Sons, New York, 1986.
3. P.C. Williams, K.H. Norris, *Near-infrared Technology in the Agricultural and Food Industry*, American Association of Cereal Chemists, Saint Paul, MN, USA, 1987.
4. M. Gallignani, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta*, **274**, 267–274 (1993).
5. S. Garrigues, J.M. Andrade, M. de la Guardia, D. Prada, *Anal. Chim. Acta*, **317**, 95–105 (1995).
6. E. López-Anreus, S. Garrigues, M. de la Guardia, 'Vapour Generation FTIR Determination of Butylglycol in Water-borne Paint Solvents', *J. Flow Injection Anal.*, **14**, 165 (1997).
7. E. López-Anreus, S. Garrigues, M. de la Guardia, 'Vapor Generation FTIR Direct Determination of Acetone and Isopropanol in Nitrocellulose Paints', *Quimica Analitica*, **17**, 21–28 (1998).
8. F. Cadet, B. Offman, *J. Agric. Food Chem.*, **45**(1), 166–171 (1997).
9. M. Gallignani, S. Garrigues, M. de la Guardia, *Analyst*, **119**, 653 (1994).
10. E. López-Anreus, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta*, **308**, 28–35 (1995).
11. M. Gallignani, S. Garrigues, A. Martinez-Vado, M. de la Guardia, *Analyst*, **118**, 1043 (1993).
12. Z. Bouhsain, S. Garrigues, M. de la Guardia, *Fresenius' J. Anal. Chem.*, **357**, 973 (1997).
13. Z. Bouhsain, S. Garrigues, M. de la Guardia, *Analyst*, **122**, 441 (1997).
14. A. Pérez-Ponce, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta*, **358**, 235–243 (1998).
15. J.H. Duckworth, 'Spectroscopic Quantitative Analysis', in *Applied Spectroscopy: A Compact Reference for Practitioners*, eds. J.W. Workman, Jr, A. Springsteen, Academic Press, London, 1998.
16. D.E. Honigs, 'Near Infrared Analysis', *Analytical Instrumentation*, **14**(1), 1–62 (1985).
17. K.H. Norris, K.J. Kaffa, Predicting Chemical Constituents in Different Materials by Derivative Techniques of Near Infrared Reflectance Spectroscopy, IMEKO, 9th World Congress, Preprint, Vol. 5, 241, 1982.
18. W.F. McClure, Abdul Hamid, F.G. Giesbrecht, W.W. Weeks, 'Fourier Analysis Enhances NIR Diffuse Reflectance Spectroscopy', *Appl. Spectrosc.*, **38**(3), 322–329 (1984).
19. J. Le Nouvel, Etude d'une Famille de Courbes par des Méthodes d'Analyse de Données: Applications à l'Analyse Morphologique de Courbes Provenant de Données Médicales. Thèse Présentée à l'Université de Rennes I, UER Mathématiques et Informatique, 1981.
20. M.F. Devaux, D. Bertrand, P. Robert, M. Qannari, 'Application of Multidimensional Analyses to the Extraction of Discriminant Spectral Patterns from NIR Spectra', *Appl. Spectrosc.*, **42**(6), 1015–1019 (1988).
21. J. Lefebvre, *Introduction aux Analyses Statistiques Multidimensionnelles*, Masson, Paris, 1983.
22. L. Lebart, A. Morineau, N. Tabard, *Techniques de la Description Statistique, Méthodes et Logiciels pour l'Analyse des Grands Tableaux*, Dunod, Paris, 1977.
23. P. Dagnelie, *Analyse Statistique à Plusieurs Variables*, Les Presses Agronomiques de Gembloux, ASBL - Belgique, 1975.
24. D. Bertrand, M. Lila, V. Furtoss, P. Robert, G. Downey, *J. Sci. Food Agri.*, **41**, 299–307 (1987).
25. D. Bertrand, P. Robert, M.F. Devaux, J. Abecassis, 'Assignment of Near Infrared Absorption Bands by Multidimensional Analysis of Spectral Data: Application to the Estimation of Purity in Durum Wheat Products', *Analytical Applications of Spectroscopy*, Royal Society of Chemistry, 450–461, 1988.
26. H. Martens, S.A. Jensen, 'Partial Least Squares Regression: A New Two Stage NIR Calibration Method', in *Progress in Cereal Chemistry and Technology (5a)*, Elsevier, Amsterdam, 607–647, 1983.
27. W.R. Hruschka, K.H. Norris, *Appl. Spectrosc.*, **36**(3), 261–265 (1982).
28. K. Norris, 'Instrumentation of Infrared Radiation', *Trans. Am. Soc. Agric. Eng.*, **5**, 17–20 (1962).
29. T. Hirschfeld, *Eur. Spectrosc. News*, **51**, 13–18 (1983).
30. J.P. Coates, J.M. D'Agostino, C.R. Friedman, *Int. Lab.*, June, 58–65 (1987).
31. C. Depecker, P. Legrand, J.C. Merlin, B. Sombret, 'Contribution de la Réflexion Diffuse Infrarouge à l'Étude

des Composés Biologiques', in *Spectroscopy of Biological Molecules*, eds. A.J.P. Alie, L. Bernard, M. Manfart, John Wiley & Sons, Chichester, 69–75, 1985.
32. F.R. Van de Voort, A.A. Ismail, 'A Rapid FTIR Quality Control Method for Fat and Moisture Determination in Butter', *Trends Food Sci. Technol.*, 13–17 (1991).
33. F. Cadet, D. Bertrand, P. Robert, J. Maillot, J. Dieudonné, C. Rouch, 'Quantitative Determination of Sugarcane Sucrose by Multidimensional Statistical Analysis of their Mid-infrared Attenuated Total Reflectance Spectra', *Appl. Spectrosc.*, **45**(2), 166–172 (1991).
34. M.P. Fuller, P.R. Griffiths, *Anal. Chem.*, **50**(13), 1906–1910 (1978).
35. C. Renard, P. Robert, D. Bertrand, M.F. Devaux, J. Abecassis, *Cereal Chem.*, **64**(3), 177–181 (1987).
36. R.A. Crocombe, N.L. Olson, S.L. Hills, *Quantitative Fourier Transform Infrared Methods for Real Complex Samples*, American Society for Testing and Materials, 95–130, 1987.
37. R.T. Holman, P.R. Edmondson, *Anal. Chem.*, **28**, 1533–1538 (1956).
38. A.J. Fenton, Jr, R.O. Crisler, *J. Am. Oil Chem. Soc.*, **36**, 620–633 (1959).
39. I. Murray, *NIR Spectra of Homologous Series of Organic Compounds*, Int. NIR Symp. Budapest, Hungary, May, 1986.
40. D.H. Wheeler, in *Progress in Chemistry of Fats and Other Lipids*, eds. R.T. Holman, W.O. Berg, T. Malkin, Pergamon Press, London, 268–291, Vol. 2, 1954.
41. S.K.D. Ahlers, R.A. Brett, N.G. McTaggart, *J. Appl. Chem.*, **3**, 433–444 (1953).
42. S.P. Kochhar, T. Matsui, *Food Chem.*, **13**, 85–101 (1984).
43. R.T. O'Connor, *J. Am. Oil Chem. Soc.*, **33**, 1–15 (1956).
44. R.G. Arnold, T.E. Hartung, *J. Food Sci.*, **36**, 166–168 (1971).
45. B.A. Anderson, R. Miller, M.J. Pallansch, *J. Dairy Sci.*, **75**, 156–159 (1974).
46. N.S. Ahmed, F.R. Helal, *Milchwissenschaft*, **32**, 272–273 (1977).
47. P.S. Belton, R.H. Wilson, H. Sadeghi, J. Orabchi, K.E. Peers, 'A Rapid Method of the Estimation of Isolated *Trans* Double Bonds in Oils and Fats Using Fourier Transform Infrared Spectroscopy Combined with Attenuated Total Reflectance', *Lebensm. Wiss. Technol.*, **21**, 153–157 (1988).
48. R. Grapin, R. Jeunet, *Comparaison des Méthodes Actuelles de Dosage de la Matiére Grasse du Lait*, 20 ème Congrès International de Laiterie, Paris, 205–206, 1978.
49. I. Ben-Gera, K.H. Norris, *Cereal J. Agric. Res.*, **18**(3), 117–123 (1968).
50. R.T. Carl, *Quantification of the Content of Milk Using a Partial-least Squares Method of Data Analysis in the Near Infrared*, Nicolet Spectroscopy Research Center, 15–16, 1989.
51. T. Sato, S. Kawano, *J. Dairy Sci.*, **73**, 3408–3413 (1990).
52. I. Ben-Gera, K.H. Norris, *J. Food Sci.*, **33**(1), 64–67 (1968).
53. W.G. Kruggel, R.A. Field, M.L. Riley, H.D. Radloff, K.M. Horton, *J. AOAC*, **64**(3), 692–696 (1981).
54. T. Hymowitz, J.W. Dudley, F.I. Collins, C.M. Brown, *Crop Sci.*, **14**, 713–715 (1974).
55. R. Tkachuk, *J. Am. Oil Chem. Soc.*, **58**, 819–822 (1981).
56. D. Ribaillier, M.F. Maviel, *Revue Française des Corps Gras*, **4–5**, 181–189 (1984).
57. J.A. Panford, P.C. Williams, J.M. De Man, *J. Am. Oil Chem. Soc.*, **65**(10), 1627–1634 (1988).
58. D.A. Biggs, *J. AOAC*, **62**(6), 1202–1210 (1979).
59. D.A. Biggs, *J. AOAC*, **62**(6), 1211–1214 (1979).
60. W.F. McClure, K.H. Norris, W.W. Weeks, *Beitr. Tabakforsch.*, **9**, 13 (1977).
61. R. Giangiacomo, J.B. Magee, G.S. Birth, G.G. Dull, 'Predicting Concentrations of Individual Sugars in Dry Mixtures by Near-infrared Reflectance Spectroscopy', *J. Food Sci.*, **46**, 531–534 (1981).
62. G.G. Dull, R. Giangiacorno, *J. Food Sci.*, **49**, 1601–1603 (1984).
63. E. Lanza, B.W. Li, 'Application of Near Infrared Spectroscopy for Predicting the Sugar Content of Fruit Juices', *J. Food. Sci.*, **49**, 995–998 (1984).
64. B.G. Osborne, T. Fearn, A.R. Miller, S. Douglas, 'Application of Near Infrared Reflectance to the Compositional Analysis of Biscuits and Biscuit Doughs', *J. Sci. Food Agric.*, **35**, 99 (1984).
65. K.J. Kaffka, Jesenszky, 'Application of NIR Techniques in the Determination of Alcohol Extract, Sugar and Titreable Acid Contents in Wines', *Hung. Sci. Instruments*, **58**, 69–74 (1984).
66. D. Baker, K.H. Norris, 'Near Infrared Reflectance Measurement of Total Sugar Content of Breakfast Cereals', *Appl. Spectrosc.*, **39**(4), 618–626 (1985).
67. D. Baker, 'The Determination of Fiber, Starch and Total Carbohydrates in Snack Foods by Near Reflectance Spectroscopy', *Cereal Foods World*, **30**(6), 389–396 (1985).
68. P. Robert, D.D. Bertrand, M. Crochon, J. Sabino, 'A New Mathematical Procedure for NIR Analysis: The Lattice Technique. Application to the Prediction of Sugar Content of Apples', *Appl. Spectrosc.*, **43**(6), 1045–1049 (1989).
69. Y. Ootaka, Determination of Individual Sugar and Acid Contents in Tomatoes in the Fluid State by NIRS, and an Improvement of Calibration Equations Using Manual Selection of Wavelengths', In: ed. K.I Hildrum, *Near Infrared Spectroscopy: Bridging the Gap Between Data Analysis and NIR Applications*, 5th International Conference, Haugesund, Norway, Isakassant, Ellis Horwood, Hemel Hempstead, Herts, 371–376, 1992.
70. F.J. Rambla, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta*, **344**, 41 (1997).

71. B.G. Osborne, S. Douglas, *J. Sci. Food Agric.*, **32**, 328–337 (1981).
72. J. Curtin, *An Evaluation of Near Infrared Spectrophotometry for Sugar Analysis*, CSR R.D. Rapport No. 91, 1986.
73. J.H. Meyer, R.A. Wood, *Rapid Analysis of Cane Juice by Near Infrared Reflectance*, Proc. South African Sugar Technologists' Association, June, 203–207, 1988.
74. G. Vaccari, G. Mantovani, G. Sgualdino, P. Goberti, 'Near Infrared Spectroscopy Utilisation for Sugar Products Analytical Control', *Zukerind*, **112**(9), 800–807 (1987).
75. E.D. Dumoulin, B.P. Azais, J. Guerain, 'Determination of Sugar and Ethanol Content in Aqueous Products of Molasses Distilleries by Near Infrared Spectrophotometry', *J. Food Chem.*, **52**(3), 626–630 (1987).
76. R.V. Ames, S.W. Norton, H.H. Nguyen, 'Control of Sugar Polarization and Moisture Using Near Infrared Reflectance Analysis', Proc. Australian Soc. Sugarcane Technologists, 246–252, 1989.
77. W. Altenburg, C.C. Chou, 'An Alternative Method of Raw Sugar Polarization', *Zuckerind*, **116**, 1041–1046 (1991).
78. A. Emmerich, 'Some Remarks about NIR Polarimetry', *Sugar J.*, 12–14 (1992).
79. C. Randall, R.G. Fowler, N. Fuson, J.R. Dangl, *Infrared Determination of Organic Structures*, Van Nostrand, New York, 1949.
80. S.A. Barker, E.J. Bourne, D.H. Whiffen, *Methods Biochem. Anal.*, **3**, 213–245 (1956).
81. J.E. Katon, J.T. Miller, F.F. Benteley, 'Infrared Spectra of Carbohydrates at Subsequent Temperature', *Carbohydr. Res.*, **10**, 505–506 (1969).
82. J.M. Anthony, 'Second Derivative FT-IR Spectra of Cellulose I and II and Related Mono- and Oligosaccharides', *Carbohydr. Res*, **173**, 185–195 (1988).
83. N. Dupuy, M. Meurens, B. Sombret, P. Legrand, J.P. Huvenne, 'Determination of Sugars and Organic Acids in Fruit Juices by FT–Mid-IR Investigation of Dry Extract', *Appl. Spectrosc.*, **45**, 860–863 (1992).
84. N. Dupuy, J.P. Huvenne, P. Legrand, 'Etude Critique de Méthodes d'Échantillonnage pour l'Analyse Quantitative par Spectrométrie Moyen Infrarouge dans le Domaine de l'Agro-alimentaire', *Industries Agro-Alimentaires*, **110**, 5–10 (1993).
85. R. Vonach, J. Buschmann, R. Falkowski, R. Schindler, B. Lendl, R. Kellner, *Appl. Spectros.*, **52**, 820–822 (1998).
86. A. Elliott, E.J. Ambrose, 'Structure of Synthetic Polypeptides', *Nature*, **4206**, 921–922 (1950).
87. T. Miyazawa, E.R. Blout, 'The Infrared Spectra of Polypeptides in Various Conformations Amide I and II Bands', *J. Am. Chem. Soc.*, **83**, 712–719 (1960).
88. M. Jackson, H.H. Mantsch, *Can. J. Chem.*, **69**, 1639–1642 (1991).
89. W.J. Yang, P.R. Griffiths, D.M. Byler, H. Susi, *Appl. Spectrosc.*, **39**(2), 282–287 (1985).
90. D.M. Byler, H. Susi, *Biopolym.*, **25**, 469–487 (1986).
91. J. Bandekar, *Biochem. Biophys. Acta*, **1120**, 123–143 (1992).
92. W.K. Surewicz, H.H. Mantsch, D. Chapman, *Biochem.*, **32**, 389–394 (1993).
93. F. Dousseau, M. Pézolet, *Biochem.*, **29**, 8771–8779 (1990).
94. P.I. Harris, D. Chapman, 'Fourier Transform Infrared Spectroscopic Studies of Biomembrane Systems', in *Infrared Spectroscopy of Biomolecules*, eds. H.H. Mantsch, D. Chapman, Wiley-Liss, New York, 1996.
95. H. Kamishikiryo, K. Hasegawa, T. Matoba, *Nippon Shokuhin Kogyo Gakkaishi*, **38**(9), 850–857 (1991).
96. K.T. Hecht, D.Z. Wood, *Proc. Roy. Soci.*, **235**, 178–188 (1956).
97. F. Cadet, 'a-NH_2 Determination from Biological Infrared Spectra', *Spectrosc. Lett.*, **29**(5), 919–936 (1996).
98. F. Cadet, F. Wong Pin, C. Rouch, C. Robert, P. Baret, *Biochem. Biophys. Acta*, **1246**(2), 59 (1995).
99. R. Schindler, B. Lendl, R. Kellner, 'Determination of Amyloglucosidase Activity Using Flow Injection Analysis with Fourier Transform Infrared Spectrometric Detection', *Analyst*, **122**, 531–534 (1997).
100. G. Reed, T.W. Nagodowithana, *Enzymes, Biomass, Food and Feed*, 3rd edition, Vol. IV, 1984.
101. B. Lendl, R. Schindler, *Vib. Spectrosc.*, **19**, 1–10 (1999).
102. S. Garrigues, M. de la Guardia, *Analyst*, **116**, 1159 (1991).
103. Z. Bouhsain, B. Hassan, S. Garrigues, M. de la Guardia, *Quim. Anal.*, **14**, 96 (1995).
104. A. Pérez-Ponce, J.M. Garrigues, S. Garrigues, M. de Guardia, *Quim. Anal.*, **14**, 96–101 (1995).
105. R.S. McDonald, *Anal. Chem.*, **56**, 349R (1984).
106. T. Janatuinen, E. Byckling, *Int. Lab.*, **15**(7), 12 (1985).
107. P.L. Hanst, *Fresenius' J. Anal. Chem.*, **324**, 579 (1986).
108. G.A. Eiceman, R.E. Clement, H.H. Hiel, *Anal. Chem.*, **64**, 170R (1992).
109. K.H. Krock, C.L. Wilkins, *Anal. Chim. Acta*, **277**(2), 381 (1993).
110. P. Jackson, G. Dent, D. Carter, D.J. Shofield, J.M. Cholmers, T. Visser, M. Vredenbregt, *J. High. Resolut. Chromatogr.*, **1619**, 515 (1993).
111. M. Bacanu, P. Dumitrescu, C. Simu, *Rev. Chim.*, **40**, 804 (1989).
112. D. Price, A.R. Horrocks, M. Akalin, *Anal. Proc.*, **27**, 148 (1990).
113. J. Laemmel, J.W. Washall, *Labor Praxis*, **17**(3), 66 (1993).
114. M. de la Guardia, M. Gallignani, S. Garrigues, 'Flow-injection Derivative Fourier Transform Infrared Determination of Methyl Tert-butyl Ether in Gasolines', *Anal. Chim. Acta*, **282**, 543–550 (1993).
115. A. Pérez-Ponce, S. Garrigues, M. de la Guardia, *Analyst*, **121**, 923 (1996).

116. A. Pérez-Ponce, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta*, **336**, 123 (1996).
117. J. Ruzicka, E.H. Hansen, *Flow Injection Analysis*, Wiley, New York, 2nd edition, 1988.
118. A.D. Moffat (ed.), *Clarke's Isolation and Identification of Drugs*, 2nd edition, The Pharmaceutical Society of Great Britain, London, 1986.
119. M. Otto, K. Sporreiter, *Mikrochim. Acta*, **III**, 167 (1985).
120. Z. Bouhsain, S. Garrigues, M. de la Guardia, *Analyst*, **121**, 635 (1996).
121. D.K. Morgan, N.A. Danielson, J.E. Katon, *Anal. Lett.*, **8**, 1979 (1988).
122. M.L. Ramos, J.F. Tyson, D.J. Curran, *Anal. Proc.*, **32**, 175 (1995).
123. K.I. Kittrick, N.D. Danielson, J.E. Katon, *Microchem. J.*, **44**, 105 (1991).
124. M.J. Sanchez-Dasi, S. Garrigues, M.L. Cervera, M. de la Guardia, *Anal. Chim. Acta*, **361**, 253–260 (1998).
125. F. Sun, T. Korenaga, *Appl. Spectrosc.*, **50**, 1145 (1996).
126. M. Varcarcel, M.D. Luque de Castro, *Flow Injection Analysis, Principles and Applications*, Ellis Horwood, Chichester, 1987.
127. J.B. Karlberg, C.E. Pacey, *Flow Injection Analysis. A Practical Guide*, Elsevier, Amsterdam, 1989.
128. J.R. Ferraro, J.R. Barrle (eds.), *Fourier Transform Infrared Spectroscopy*, Academic Press, New York, 1982.
129. P.R. Griffiths, J.A. de Haselh, *Fourier Transform Infrared Analysis*, Wiley, New York, 1986.
130. D.J. Curran, W.G. Collier, *Anal. Chim. Acta*, **177**, 259 (1985).
131. B.E. Miller, N.D. Danielson, J.E. Katon, *Appl. Spectrosc.*, **42**, 401 (1988).
132. S.V. Olesick, S.B. French, M. Novotny, *Anal. Chem.*, **58**, 2256 (1986).
133. M. Guzman, J. Ruzicka, G.D. Christian, P. Shelley, *Vibr. Spectrosc.*, **2**(1), (1991).
134. M. Gallignani, S. Garrigues, M. de la Guardia, J.L. Burguera, M. Burguera, *Anal. Chim. Acta*, **261**, 53 (1992).
135. R. Vonach, B. Lendl, R. Kellner, *J. Chromatogr. A.*, **824**, 159–167 (1998).
136. S. Garrigues, M. Gallignani, M. de la Guardia, *Talanta*, **40**, 89 (1993).
137. S. Garrigues, M. Gallignani, M. de la Guardia, *Talanta*, **40**, 1799 (1993).
138. S. Garrigues, M.T. Vidal, M. Gallignani, M. de la Guardia, *Analyst*, **119**, 659 (1994).
139. M.J. de Faubert Maunder, *Practical Hints on Infrared Spectrometry with Particular Reference to Forensic Analysis*, Adam Hilger, London, 1971.
140. Z. Bouhsain, S. Garrigues, M. de la Guardia, *Analyst*, **121**, 1935–1958 (1996).
141. A. Van Loosbroek, H.J.G. Debets, P.M.J. Coenegracht, *Anal. Lett.*, **17**, 779 (1984).
142. K. Molt, M. Egelkraut, *GIT Fachz. Lab.*, **32**, 1311 (1988).
143. G. Sala, S. Maspoch, H. Iturriaga, M. Blanco, V. Cerda, *J. Pharm. Biomed. Anal.*, **6**, 765 (1988).
144. H. Shen, S. Cai, *Gaodeng Xuexiao Huaxue Xuebao*, **14**, 918 (1993).
145. H. Shen, S. Cai, *Fenxi Huaxue*, **22**, 716 (1994).
146. S. Garrigues, M. Gallignani, M. de la Guardia, *Analyst*, **117**, 1849 (1992).
147. N. Kosal, A. Bhari, M.A. Ali, *Fuel*, **69**, 1012 (1990).
148. V. Ramaswamy, S.S. Sarowha, J.D. Singh, *J. Chem. Technol. Biotechnol.*, **36**, 378 (1986).
149. V. Ramaswamy, V. Kothiyal, J.D. Singh, R. Kishna, *J. Chem. Technol. Biotechnol.*, **41**, 255 (1988).
150. B. de Simas, L. Lane, *Adv. Instrum. Control.*, **45**, 451 (1990).
151. E.F. Daton, F.P. Fuller, F.T. White, R.D. Battiste, *Anal. Chem.*, **55**, 408 (1983).
152. R.N. Hager, *Anal. Chem.*, **45**, 1131A (1973).
153. A.F. Fell, *Anal. Proc.*, **15**, 260 (1978).
154. F. Sanchez Rojas, C. Bosch Ojeda, J.M. Cano Pavon, *Talanta*, **35**, 753 (1989).
155. Ministry of Defence, *Defence Standard 1*, London, DERD 2494, September, 1994.
156. *Annual Book of ASTM Standards*, Vol. 05.02. American Society for Testing and Materials, ASTM, Philadelphia, PA, D-3699, 1994.
157. O. Kirk, In *Encyclopedia of Chemical Technology*, ed. M. Grayson, 3rd edition, Wiley, New York, 1979.
158. J. Coates, S. Reber, *Am. Lab.*, **24**, C20 (1992).
159. Y. Verbies, B. Fevre, A. Andre, *Analusis*, **21**, M29–M30 (1993).
160. E. López-Anreus, S. Garrigues, M. de la Guardia, 'Flow Injection Near-infrared Determination of Ethanol in Chloroform', *Fresenius' J. Anal. Chem.*, **351**, 724–728 (1995).
161. M.K. Antoon, D'Esposito, J.L. Koenig, *Appl. Spectrosc.*, **31**(6), 518–524 (1977).
162. M. de la Guardia, J.L. Carrión, J. Medina, *Analysta*, **109**, 457 (1984).
163. L.M. Hanssen, K.A. Snail, Proc. SPIE, The International Society for Optical Engineering, *Passive Infrared Systems Technology*, 148–159, Vol. 807, 1987.
164. J.J. Lagowski, *The Chemistry of Non-aqueous Solvents*, Academic Press, New York, 1978.
165. W.H. Morgans, in *Outlines in Paint Technology*, ed. Griffin, London.
166. E. Castro, J. Garcia, W. Garcia, E. Laloumet, *Manuel de la Pintura en la Construccion*, ANSPI, Barcelona, 1991.
167. F.D. Snell, L.S. Ettre, *Encyclopedia of Industrial Chemical Analysis*, Interscience, New York, 1969.
168. S. Garrigues, 'Paint. Organic Solvent-based Products', in *Encyclopedia of Analytical Science*, Academic Press, London, 1995.
169. C.M. Bell, S.J. Gutowski, S. Young, D. Wells, *J. Anal. Toxicol.*, **16**, 166–168 (1992).
170. Cecon Group, *Pract. Spectrosc.*, **10**, 95 (1990).
171. R. Schindler, M. Watkins, R. Vonach, B. Lendl, R. Kellner, R. Sara, 'Automated Multivariate Calibration in

Sequential Injection Fourier Transform Infrared Spectroscopy for Sugar Analysis', *Anal. Chem.*, **70**, 226–231 (1998).
172. C.A. Bruckner, M.D. Foster, L.R. Lima, R.E. Synovec, R.J. Berman, C.N. Renn, E.L. Johnson, *Anal. Chem.*, **66**, IR (1994).
173. J.W. Hellgeth, L.T. Taylor, *J. Chromatogr. Sci.*, **24**, 519 (1986).
174. K. Jinno, C. Fujimoto, *J. Chromatogr.*, **506**, 443 (1990).
175. K. Jinno, *Chem. Anal.*, **89**, 64 (1986).
176. G.W. Somsen, E.W.J. Hooijschuur, C. Gooijer, U.A. Th. Brinkman, N.H. Velthorst, T. Visser, *Anal. Chem.*, **68**(5), 746 (1996).
177. E.S. Yeung, R.E. Synovec, *Anal Chem.*, **58**(12), 1237A (1986).
178. K.J. Aitzetmüller, *Chromatogr.*, **139**, 61 (1977).
179. W.S. Letter, *J. Liq. Chromatogr.*, **16**(1), 225 (1993).
180. R. Bauza, F. Ysambertt, N. Marquez, G.O. De-Rodriguez, J.L. Lopez, *Sep. Sci. Technol.*, **27**(5), 645 (1992).
181. R.J. Hamilton, S.F. Mitchell, P.A. Sewell, *J. Chromatogr.*, **395**, 33 (1987).
182. K. Payne-Wahl, R. Kleiman, *J. Am. Oil Chem. Soc.*, **60**(5), 1011 (1983).
183. K. Payne-Wahl, G.F. Spencer, R.D. Plattner, R.O. Butterfield, *J. Chromatogr.*, **209**, (1981).
184. Y. Daghbouche, S. Garrigues, A. Morales-Rubio, M. de la Guardia, *Anal. Chim. Acta*, **354**, 161–171 (1997).
185. *Official Methods of Analysis of the Association of Official Analytical Chemists*, AOAC, Washington, DC, Supplement, Chapter 11, 13, March, 1995.
186. A.S. Bangalore, G.W. Small, R.J. Combs, R.B. Knapp, R.T. Kroutil, *Anal. Chim. Acta*, **297**(3), 387–403 (1994).
187. W. Windig, D.A. Stephenson, *Anal. Chem.*, **64**(92), 2735–3742 (1992).
188. I. Panovici, C. Manoliv, L. Costache, *Chim (Bucharest)*, **37**(2), 148–150 (1986).
189. R.E. Shaffer, G.W. Small, *Chemom. Intell. Lab. Syst.*, **32**(1), 95–109 (1996).
190. B.R. Buchanan, D.E. Honigs, *Appl. Spectrosc.*, **41**(8), 1388–1392 (1987).
191. S. Cha, P.A. Gabele, *Opt. Eng.*, **25**(1), 11–12 (1986).
192. J.D. Wang, Y.H. Luo, Z.R. Chen, Y.R. Lu, S. Pan, X.M. Wang, *Anal. Chim. Acta*, **277**(1), 153–155 (1993).
193. J. Wesenberg, M. Knebusch, W. Seyfarth, Z. Wiss, *Humboldt Univ. Berlin, Math. Naturwiss. Reihe*, **31**(6), 603–616 (1982).
194. A. Profumo, C. Riolo, M. Pesavento, A. Francoli, *Ann. Chim.*, **77**(5–6), 509–514 (1987).
195. X.Y. Cai, J.J. Yin, G.D. Hu, *Hewlett Packard Peak*, **3**, 10–11 (1995).
196. H. Laber, J. Schulz, W.R. Sponholz, W. Bremser, *Fresenius' J. Anal. Chem.*, **351**, 530–535 (1995).
197. M.K. Hunson, K.W. Busch, *Anal. Chem.*, **59**, 2603–2609 (1995).
198. B.J. McGrattan, D.W. Schiering, R.A. Hoult, *Spectroscopy (Eugene, Oreg.)*, **4**(9), 39–41 (1989).
199. T. Higashi, T. Imasaka, N. Ishibashi, *Anal. Chem.*, **56**, 2010–2013 (1984).
200. D. Kuehl, R. Crocombe, *Appl. Spectrosc.*, **38**(6), 907–909 (1984).
201. S. Garrigues, A. Pérez-Ponce, S. Garrigues, M. de la Guardia, *Vib. Spectrosc.*, **15**, 219–228 (1997).
202. American Water Works Association and Water Environment Federation, Standard Methods for the Examination of Water and Wastewater, American Public Health Association, Washington, DC, 18th edition, 1992.
203. R.G. Somard, I. Hasegawa, W. Bandaruk, C.E. Headington, *Anal. Chem.*, **23**, 1384 (1951).
204. US Environmental Protection Agency, *Oil and Grease, Total Recoverable*, EPA 413.2, Environmental Protection Agency, Washington, DC, 1978.
205. D. Noble, *Anal. Chem.*, **65**, 695A (1993).
206. J.C. Raia, D.K.J. Lee, P.J. Staley, Proc. Fifteenth Annual EPA Conference on Analysis of Pollutants in the Environment, EPA 821-R-92-007, Environmental Protection Agency, Washington, DC, 299–340, 1992.
207. W.A. Telliard, Proc. Fifteenth Annual EPA Conference on Analysis of Pollutants in the Environment, EPA 821-R-92-007, Environmental Protection Agency, Washington, DC, 291–307, 1992.
208. H.M. Kingston, L.B. Jassie (eds.), *Introduction to Microwave Sample Preparation. Theory and Practice*, American Chemical Society, Washington, DC, 1988.
209. M. de la Guardia (ed.), *Empleo de los Hornos de Microondas en Quimica*, Universidad de Valencia, Valencia, 1990.
210. V. López-Avila, R. Young, W.F. Beckert, *Anal. Chem.*, **66**, 1097 (1994).
211. A. Pastor, E. Vasquez, R. Ciscar, M. de la Guardia, *Anal. Chim. Acta*, **344**, 241–249 (1997).
212. M. de la Guardia, J. Ruzicka, *Analyst*, **120**, 17N (1995).
213. G.G. Liversidge, *Analytical Profiles of Drug Substances*, Academic Press, London, 1981.
214. *British Pharmacopoeia*, HM Stationery Office, London, 1993.
215. C. Goyet, A.K. Snover, *Mar. Chem.*, **44**, 235 (1993).
216. S.A. Huber, F.M. Frimmel, *Anal. Chem.*, **63**, 2122 (1991).
217. A. Pérez-Ponce, S. Garrigues, M. de la Guardia, *Analyst*, **121**, 923–928 (1996).
218. C.W. McLeod, Plenary Lecture Presented at the XIII Reunion National de Espectroscopia, Gandia, Spain, October, 1992.
219. R.J. Bushway, *J. Chromatogr.*, **211**, 135 (1981).
220. J.S. Andrews, T.J. Good, *Am. Lab.*, **14**, 70 (1982).
221. G.A. Junk, J.J. Richard, *Anal. Chem.*, **60**, 451 (1988).
222. J. Ruzicka, G.D. Marshall, *Anal. Chim. Acta*, **237**, 329 (1990).
223. C.C. Johnson, J.W. Hellgeth, L.T. Taylor, *Anal. Chem.*, **57**, 610–615 (1985).

224. J.W. Hellgeth, L.T. Taylor, *Anal. Chem.*, **59**, 295–300 (1987).
225. C.M. Conroy, P.R. Griffiths, P.J. Duff, L.V. Azarraga, *Anal. Chem.*, **56**, 2636–2642 (1984).
226. B. Lendl, R. Schindler, J. Frank, R. Keller, J. Drott, T. Laurell, *Anal. Chem.*, **69**(15), 2877–2881 (1997).

Spectral Data, Modern Classification Methods for

Yvette L. Mallet and Danny H. Coomans
James Cook University, Townsville, Australia

Olivier de Vel
DSTO, Salisbury, Australia

1 Introduction **1**
2 Classification Methods **2**
2.1 Bayesian Quadratic Discriminant Analysis 2
2.2 Bayesian Linear Discriminant Analysis 3
2.3 Regularized Discriminant Analysis 3
2.4 Fisher's Linear Discriminant Analysis 3
2.5 Flexible Discriminant Analysis 4
2.6 Penalized Discriminant Analysis 4
3 Feature Extraction **4**
3.1 Feature Selection 4
3.2 Feature Transformation 5
4 Wavelet Transforms of Discrete Data **5**
4.1 Discrete Wavelet Transform 5
4.2 Wavelet Packet Transform 6
4.3 Local Discriminant Basis Algorithm 8
4.4 Adaptive Wavelet Algorithm 9
5 Applications **10**
5.1 Data Sets 10
5.2 Methodology 10
6 Summary of Results **16**
7 Conclusion **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

A major concern which arises from the classification of spectral data, such as near-infrared (NIR) spectra, is that the number of variables often exceeds the number of cases. In this instance each spectrum represents a case or observation, and each variable corresponds to the wavelengths at which a response (e.g. reflectance) is measured. Classification refers to the process of assigning a spectrum whose class identity is unknown into one of several predefined classes. There are two main strategies for combating the high-dimensional (i.e. more variables than cases) scenario. One strategy is to use a high-dimensional classifier, that is, one which is suited to the ill-posed nature of the data. The second strategy is to reduce the number of variables by a feature extraction method and supply the variables to a low-dimensional classifier, one which is suited to high observation-to-variable ratios. This article presents two high-dimensional classification methods – regularized discriminant analysis (RDA) and penalized discriminant analysis (PDA). The low-dimensional classifiers used include Bayesian linear discriminant analysis (BLDA), Bayesian quadratic discriminant analysis (BQDA), RDA and flexible discriminant analysis (FDA). Common stepwise selection routines are employed to investigate the effectiveness of the wavelet transform as a feature extraction technique. This includes implementation of the local discriminant bases (LDBs) algorithm and the adaptive wavelet algorithm (AWA).

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

In this paper we investigate several strategies for classifying spectral data where the dimensionality or number of variables far exceeds the number of samples or objects. Such poorly posed data lead to a deterioration in performance of many standard classification techniques.[1] The presence of correlated variables can lead to a further deterioration in classification performance. Correlated data often result if the same response is measured at regular intervals in time or space. Although the techniques discussed in this paper can be applied to most highly correlated and highly dimensional data, such as time-varying signals, spectra and images, we focus our applications on the classification of NIR spectral data.

The discriminant problem when applied to NIR spectral data involves assigning an NIR spectrum whose class identity is unknown (Figure 1d) into one of the several predefined categories or classes. Figures 1(a), (b) and (c) show five sampled spectra from three different species of seagrass – *Halophila ovalis, Halodule uninervis/pinifolia* and *Halophila spinulosa*. The different species, which are referred to as species 1, 2 and 3 respectively, form the different group categories. A simple approach is to look for similarities between the unidentified spectrum and the spectra whose class identities are known. In this particular case this task is not straightforward, since the spectra from the different classes share similar characteristics. This problem highlights the relevance of automated discriminant methods for analyzing spectral data.

The general discriminant problem[2,3] can be expressed as assigning a spectra, represented by the p-dimensional vector $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})^T$, into one of $r \in \{1, \ldots, R\}$ classes. The assignment is performed by building a

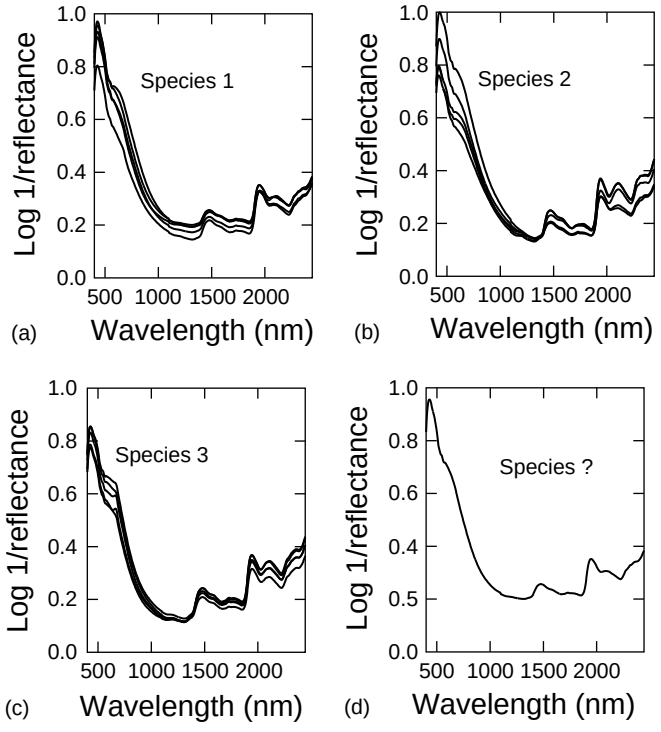

Figure 1 Sample spectra from three species of seagrass. (a) Species 1 (*Halophila ovalis*); (b) species 2 (*Halodule uninervis/pinifolia*); (c) species 3 (*Halophila spinulosa*); (d) spectrum to be identified as species 1, 2 or 3.

discriminant model, using n training spectra where there are n_r spectra $\{\mathbf{x}_{i(r)}\}_{i=1}^{n_r}$ from each class r. The total, $n = \sum_{r=1}^{R} n_r$ training spectra are stored as columns in the $p \times n$ training data matrix $\mathbf{X} = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$. Usually, $p \gg n$ for spectral data.

Several different discriminant or classification models exist. These are divided into two main groups – low- and high-dimensional classifiers. High-dimensional classifiers can be applied to data sets with low observation-to-variable (n/p) ratios. Conversely, low-dimensional classifiers are more suited to data containing high observation-to-variable ratios. The high-dimensional classifiers applied in this paper are PDA and, RDA. The low-dimensional classifiers applied are BLDA and BQDA, and FDA. These methods will be discussed briefly in section 2. For a more detailed account of these techniques the reader is referred to Mallet et al.[4]

If a low-dimensional classifier is to be used in a situation where there are many more variables than objects, then it is recommended that some form of data reduction or feature extraction be implemented. Feature extraction involves selecting a smaller subset of variables. Selection can be from the original set of variables or from a set of transformed variables. Principal components are common examples of transformed variables. Another set of transformed variables that remains relatively unexplored is wavelet coefficients.[5] In this paper we investigate the potential of wavelet coefficients as features for discriminant analysis. Wavelet-based feature extraction methods involve two main procedures. The first procedure is the application of the discrete wavelet transform (DWT) to obtain the wavelet coefficients. The second procedure is the selection of a subset of wavelet coefficients to be supplied to the classification model. Wavelet coefficients are considered useful because they can analyze different segments of data at different resolutions, thus enabling localized features in the spectra to be detected. These features can be extremely useful for discriminating purposes. Previous applications of wavelet-based feature extraction methods for classifying data are available.[4,6–12]

The goal of this paper is not necessarily to find the best discriminant strategy, but to review some traditional methodologies and introduce the reader to some modern classification techniques, which have been divided into two groups:

1. classification of all the original data using a high-dimensional classifier
2. classification using a low-dimensional classifier where features are extracted from either the original data, or the DWT.

The paper proceeds by introducing the low- and high-dimensional classifiers in section 2. Section 3 discusses some feature extraction procedures and section 4 describes the DWT. Section 5 applies the different classification applications to four spectral data sets which exhibit different characteristics. Section 6 summarizes the classification results. Section 7 presents some guidelines for choosing classification techniques and concludes with some final remarks.

2 CLASSIFICATION METHODS

This section briefly describes the discriminant methods that were mentioned in section 1. The classifiers based on Bayes's method are introduced first, followed by a description of the classifiers based on Fisher's linear discriminant analysis (FLDA). Although FLDA is not applied in this article, the method is briefly reviewed as it is the foundation for FDA and PDA.

2.1 Bayesian Quadratic Discriminant Analysis

Bayes's rule assigns a spectrum $\mathbf{x}$ to the class $r \in \{1, \ldots, R\}$, which maximizes the posterior probability,

as shown in Equation (1).

$$P(r|\mathbf{x}) = \frac{p(\mathbf{x}|r)P(r)}{\sum_{r=1}^{R} p(\mathbf{x}|r)P(r)} \quad \text{for } r = \{1, \ldots, R\} \tag{1}$$

Here, $P(r)$ is the prior probability of belonging to class r, and $p(\mathbf{x}|r)$ are the class probability densities, i.e. the probability of spectrum $\mathbf{x}$ arising from group r. BQDA[2,3] assumes that the class probability densities follow a multivariate normal distribution as shown in Equation (2),

$$p(\mathbf{x}|r) = (2\pi)^{-p/2}|\mathbf{S}_r|^{-1/2}\exp[-0.5(\mathbf{x} - \bar{\mathbf{x}}_r)^T\mathbf{S}_r^{-1}(\mathbf{x} - \bar{\mathbf{x}}_r)] \tag{2}$$

where $\mathbf{S}_r = \sum_{i=1}^{n_r}(\mathbf{x}_{i(r)} - \bar{\mathbf{x}}_r)(\mathbf{x}_{i(r)} - \bar{\mathbf{x}}_r)^T/n_r$ is the covariance matrix of class r, and $\bar{\mathbf{x}}_r = \sum_{i=1}^{n_r}\mathbf{x}_{i(r)}/n_r$ is the mean vector of class r. Both $\mathbf{S}_r$ and $\bar{\mathbf{x}}_r$ are sample estimates of the population class covariance matrices and mean vectors, respectively.

2.2 Bayesian Linear Discriminant Analysis

As with BQDA, BLDA[2,3] also assumes that the class probability densities follow a multivariate normal distribution. BLDA makes the added assumption that the class covariance matrices in Equation (2) are equal, and are replaced with a pooled covariance matrix $\mathbf{S}_{\text{pooled}} = \sum_{r=1}^{R} n_r\mathbf{S}_r/n$ to give Equation (3).

$$p(\mathbf{x}|r) = (2\pi)^{-p/2}|\mathbf{S}_{\text{pooled}}|^{-1/2} \times \exp[-0.5(\mathbf{x} - \bar{\mathbf{x}}_r)^T\mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_r)] \tag{3}$$

2.3 Regularized Discriminant Analysis

When the observation-to-variable ratio is small (e.g. less than three) the estimates of the class covariance matrices $\mathbf{S}_r$ become highly variable.[6]

This can cause serious problems in the performance of the classification procedure. To overcome this phenomenon, Friedman[6] has proposed one method of regularizing the class covariance matrix. RDA uses a similar classification rule to that shown in Equation (2) for BQDA. Instead of using the unbiased estimate of the class covariance matrix, RDA replaces $\mathbf{S}_r$ in Equation (2) with the regularized covariance matrix $\mathbf{S}_r(\lambda, \gamma)$ which undergoes a two-stage regularization procedure. Despite the possibility of slightly increasing bias, regularization can produce dramatic improvements in classification.[1,6]

The first regularization procedure replaces the class covariance matrix with a linear combination of the class and pooled covariance matrices, as shown in Equation (4).

$$\mathbf{S}_r(\lambda) = (1 - \lambda)\mathbf{S}_r + \lambda\mathbf{S}_{\text{pooled}} \tag{4}$$

The parameter $\lambda \in [0, 1]$ controls the degree to which the pooled covariance matrix should be used. This form of regularization is similar to that which occurs when reverting from BQDA to BLDA. Replacing the class covariance matrix with the pooled covariance matrix reduces the number of parameters that have to be estimated. In fact, if $\lambda = 1$ then $\mathbf{S}_r(1) = \mathbf{S}_{\text{pooled}}$.

The second regularization procedure is shown in Equation (5).

$$\mathbf{S}_r(\lambda, \gamma) = (1 - \gamma)\mathbf{S}_r(\lambda) + \frac{\gamma}{d}\text{trace}[\mathbf{S}_r(\lambda)]\mathbf{I} \tag{5}$$

$\mathbf{S}_r(\lambda)$ is adjusted by scaling the covariance matrices towards a multiplier of the identity matrix $\mathbf{I}$. The value of $\gamma \in [0, 1]$ determines the degree to which $\mathbf{S}_{(r)}(\lambda)$ is shrunk towards a multiplier of the identity matrix. Different λ and γ values ranging between 0 and 1 are trialled. The pair of values which produces the minimal risk of misclassification is used.

2.4 Fisher's Linear Discriminant Analysis

Many pattern recognition methods are Bayesian- or Fisher-based techniques. The classifiers BLDA, BQDA and RDA are examples of Bayesian-based classifiers. Fisher-based classifiers stem from the time-honoured technique called FLDA. FDA and PDA, which are discussed in sections 2.5 and 2.6, respectively, are largely derived from FLDA.

As shown in Equation (6) FLDA assigns an object $\mathbf{x}$, to the class $r \in \{1, \ldots, R\}$ which minimizes

$$\|\mathbf{x}^T\mathbf{V} - \bar{\mathbf{x}}_r^T\mathbf{V}\|^2 \tag{6}$$

where $\mathbf{V}$ is a $p \times s$ matrix whose ith column is the eigenvector $\mathbf{v}_i$ which corresponds to the ith largest eigenvalue of $\mathbf{S}_W^{-1}\mathbf{S}_B$, and $s \leq \min(R - 1, p)$. Here, $\mathbf{S}_B$ and $\mathbf{S}_W$ are the between- and within-covariance matrices of the data matrix $\mathbf{X}$, respectively. These are calculated from the training data using the definitions in Equations (7) and (8).

$$\mathbf{S}_B = \frac{1}{n}\sum_{r=1}^{R} n_r(\bar{\mathbf{x}}_r - \bar{\mathbf{x}})(\bar{\mathbf{x}}_r - \bar{\mathbf{x}})^T \tag{7}$$

$$\mathbf{S}_W = \frac{1}{n}\sum_{r=1}^{R}\sum_{i=1}^{n_r}(\mathbf{x}_{i(r)} - \bar{\mathbf{x}}_r)(\mathbf{x}_{i(r)} - \bar{\mathbf{x}}_r)^T \tag{8}$$

The solution for $\mathbf{V}$ arises from the formal discriminant problem as shown in Equation (9)

$$\max \mathbf{v}^T\mathbf{S}_B\mathbf{v} \tag{9}$$

subject to the constraint shown in Equation (10).

$$\mathbf{v}^T\mathbf{S}_W\mathbf{v} = 1 \tag{10}$$

2.5 Flexible Discriminant Analysis

FDA[7] is a nonlinear extension of FLDA which incorporates nonparametric regression methods to obtain nonlinear decision boundaries. The method of FDA is quite involved. For this reason we provide a very descriptive review of the method and refer the reader to Hastie et al.[7] and Mallet[8] for more refined details.

FLDA produces linear decision boundaries for separating objects from different classes. One way of producing nonlinear decision boundaries would be to expand the predictor matrix so that it includes quadratic and bilinear terms.[7] This would have the effect of producing discriminant boundaries which exhibit quadratic characteristics. A much more sophisticated approach is used by FDA to obtain much more flexible decision boundaries than quadratic forms.

The nonlinear decision boundaries in FDA are obtained by adaptively computing an expanded set of predictor variables by using nonparametric regression methods such as MARS (multivariate adaptive regression splines)[9] and bruto.[10] The new independent variables are computed with the aim of predicting the columns of a class indicator matrix. Hastie et al.[7] show that the regression coefficients are, by means of a linear transformation, equivalent to the matrix of eigenvectors ($\mathbf{V}$) from FLDA. This relationship is neatly established by using optimal scoring as the link between regression analysis and FLDA (see also Breiman and Ihaka[11]).

2.6 Penalized Discriminant Analysis

PDA[12] follows a similar methodology to that used for FDA. That is, optimal scoring provides the link between regression and classification. The difference between FDA and PDA is in the regression method used. For FDA, the goal is to cater for situations where there is a high observation-to-variable ratio, and to produce flexible decision boundaries for classification. This is achieved by using adaptive regression methods such as MARS and bruto. The goal of PDA is to cater for situations where there is a low observation-to-variable ratio, and also for data containing large amounts of highly correlated variables such as spectra, signals and images. It becomes obvious that the regression methods used by PDA will differ from the regression methods used by FDA. The regression methods for PDA need to be suited to extreme dimensionalities and be somewhat resistant to multicollinearities. Consequently, PDA uses penalized least squares regression (PLSR) methods. Ridge regression[13] can be seen as a trivial PLSR method.

The use of penalized regression methods can be made equivalent to a penalized discriminant method by solving Equation (9) subject to the constraint

$$\mathbf{v}^T(\mathbf{S}_W + \mathbf{\Omega})\mathbf{v} = 1$$

where $\mathbf{\Omega}$ is a penalty matrix which helps prevent overfitting and other computational instabilities.

3 FEATURE EXTRACTION

Feature extraction[2,3] is a dimensionality reduction procedure that selects a subset of p_* variables from a much larger set of p variables. Typically, the dimensionality p_* of the subset of feature variables is less than p, and usually very much less than the number of observations, n. The main goal of feature extraction relative to this article is to extract as much relevant discriminatory information from the data as possible, while simultaneously eliminating redundant features which do not contribute to or have an adverse effect on the classification procedure. It is recommended that some form of feature extraction be implemented prior to the application of low-dimensional pattern classification methods.

The feature extraction procedure may involve selecting features from either the original (untransformed) data, or from data which has been transformed. Hence, feature extraction can consist of two main components – feature selection and feature transformation. These topics are explained in sections 3.1 and 3.2 respectively.

3.1 Feature Selection

Feature selection procedures select a subset of features or variables from the data. Features can be selected from the original data, also referred to as a variable selection, or from data which have been preprocessed or transformed in some other way. When feature selection is from the original data, the feature transformation is via the identity matrix.

One commonly used feature selection criterion for classication is the ratio of the between-to-within variability.[14] That is, for each variable, the between-variability divided by the within-variability is calculated and the variables that produce large values of this quotient are selected, and used as input to the classifier. This measure of between-to-within-variability is referred to as Fisher's criterion.[15] Other feature selection criteria commonly used include the Mahalanobis distance and Wilk's lambda.[16]

One criterion which is applied later in this article is the quadratic probability measure (QPM).[17] Since the QPM is a probabilistic measure, $0 \leq \text{QPM} \leq 1$. Ideally, larger values of the QPM are preferred, since this suggests that the classes can be differentiated with a higher

degree of certainty. The QPM is formulated as shown in Equation (11)

$$\text{QPM} = \frac{1}{n}\sum_{i=1}^{n} a_Q(i) \tag{11}$$

where $a_Q(i)$ is defined as shown in Equation (12).

$$a_Q(i) = \frac{1}{2} + P(r|\mathbf{x}_{i(r)}) - \frac{1}{2}\sum_{r=1}^{R} P(r|\mathbf{x}_i)^2 \tag{12}$$

In section 5 we use a leave-one-out cross-validated quadratic probability measure (CVQPM). Here, the parameters in the appreciation score $a_Q(i)$ are computed in the absence of $\mathbf{x}_i$.

It is worthwhile to give some consideration to procedures that can be used for selecting combinations of variables. For example, two features chosen separately may produce less favourable results than two features chosen in combination. With high-dimensional data it is not realistic to perform an exhaustive search of every possible combination of variables. In this case, stepwise methods can be useful. A forward stepwise search sequentially incorporates variables into the discriminant procedure that contribute to the discrimination power of the model. Generally, variables cease to enter the model when little or no change to the performance of the discriminant model is registered. Conversely, a backward stepwise search removes variables from the model until the performance of the model begins to deteriorate. In the presence of high-dimensional data, it is not advisable to use a backward stepwise search, since initially all the variables will be fed to the classifier, which will lead to ill- and poorly-posed situations, not to mention the computational expense and the numerical instabilities which are likely to arise. If desired, a backward selection scheme could be used, following the implementation of a forward, stepwise selection method.

3.2 Feature Transformation

The procedure for calculating a new set of features is called feature transformation. Examples of spectral features that may be informative include the heights, positions or shapes of peaks. Other features commonly used include principal components and Fourier coefficients, while more recently wavelet coefficients[4,15,18–25] have been investigated. These features are obtained by a transformation called the DWT.

The advantage of using features from the wavelet transform, as opposed to principal components, is that the DWT takes into consideration the ordering or incrementation of the variables. This order information is not contained in the principal components. An alternative procedure is Fourier transform (FT) which does take this ordering into consideration, but offers a broader, more global description of a spectrum or signal. The DWT is able to break the spectra into pieces at different scales and analyze the frequency content of each piece. This allows the DWT to convey localized frequency information about a spectrum, and such information can be quite useful for classification purposes.

4 WAVELET TRANSFORMS OF DISCRETE DATA

4.1 Discrete Wavelet Transform

Consider the finite length, discretely sampled spectrum $\mathbf{x} = (x_1, x_2, \ldots, x_{p=2^J})$ of dimensionality $p = 2^J$ for some $J \in \mathcal{Z}^+$. The DWT for such a spectrum can be likened to a series of filtering operations. There are various procedures such as zero padding which can be implemented if $p \neq 2^J$ (see Bruce and Gao[24] and Misiti[25]). The raw spectrum $\mathbf{x}$, which we also call $\mathbf{c}_J$ for ease of notation, passes through a lowpass $\mathbf{L}$ and a highpass $\mathbf{H}$ filter to produce the decimated sequences $\mathbf{c}_{J-1}$ and $\mathbf{d}_{J-1}$, each of dimensionality 2^{J-1}. The scaling coefficients $\mathbf{c}$ are the outputs from the lowpass filter, and the wavelet coefficients $\mathbf{d}$ are the outputs from the highpass filter. The subscript $j \in \{0, 1, \ldots, J\}$ indicates which level the coefficients are from in the DWT, with J initially being the highest level. The general procedure for the DWT continues by passing the scaling coefficients $\mathbf{c}_j$ at level j through a lowpass and a highpass filter to produce the scaling $\mathbf{c}_{j-1}$ and wavelet coefficients $\mathbf{d}_{j-1}$ at the next lower level $j-1$. The scaling and wavelet coefficients at the lowest level j_o in the DWT, as well as the wavelet coefficients at higher levels, comprise the DWT. That is, the DWT contains the coefficients $\mathbf{c}_{j_o}, \mathbf{d}_{j_o}, \mathbf{d}_{j_o+1}, \ldots, \mathbf{d}_{J-1}$. Because of the decimation procedure which forms part of the filtering operations, the number of scaling and wavelet coefficients halve when moving from one level to the next lower level of the DWT, so that the dimensionality of the scaling or wavelet coefficients at some level j is 2^j.

The filtering operations from any level j to $j-1$ can be defined by the matrix operations

$$\mathbf{c}_{j-1} = \mathbf{L}_j\mathbf{c}_j$$
$$\mathbf{d}_{j-1} = \mathbf{H}_j\mathbf{c}_j$$

In summation notation this can be written as shown in Equations (13) and (14)

$$c_{j-1,i} = \sum_{k=0}^{N_f-1} \ell_k c_{j,2i+k} \tag{13}$$

$$d_{j-1,i} = \sum_{k=0}^{N_f-1} h_k c_{j,2i+k} \tag{14}$$

where there are N_f lowpass (ℓ_k) and highpass (h_k) filter coefficients, respectively. In this discussion periodic boundary conditions were assumed.

$$c_{j,k} = c_{j,2^{j+k}}$$
$$d_{j,k} = d_{j,2^{j+k}}$$

Figure 2 indicates how the wavelet coefficients and scaling coefficients are calculated for some discrete spectrum $\mathbf{x} = (x_0, \ldots, x_7)^T = (c_{3,0}, \ldots, c_{3,7})^T = \mathbf{c}_3$ of length $p = 8$. Obviously, in practice the dimensionality of spectra will usually be much greater than $p = 8$. This example is to simply present the underlying principle of the DWT. Initially, all the data $\mathbf{c}_3$, are passed through the low- and highpass filters to give the scaling $\mathbf{c}_2 = (c_{2,0}, \ldots, c_{2,3})^T$ and wavelet $\mathbf{d}_2 = (d_{2,0}, \ldots, d_{2,3})^T$ coefficients at the next lower level. As one progresses down the tree, the number of elements in each of the bands is reduced by half.

If for example $N_f = 2$, then the lowpass filter coefficient matrices would be defined as in Equations (15–17).

$$\mathbf{L}_3 = \begin{matrix} \ell_0 & \ell_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ell_1 & \ell_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ell_0 & \ell_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ell_0 & \ell_1 \end{matrix} \tag{15}$$

$$\mathbf{L}_2 = \begin{pmatrix} \ell_0 & \ell_1 & 0 & 0 \\ 0 & 0 & \ell_0 & \ell_1 \end{pmatrix} \tag{16}$$

$$\mathbf{L}_1 = \begin{pmatrix} \ell_0 & \ell_1 \end{pmatrix} \tag{17}$$

Likewise, the highpass filter coefficient matrices would have the same structure as the $\mathbf{L}_3$, $\mathbf{L}_2$ and $\mathbf{L}_1$ matrices, the main difference being ℓ_i would be substituted with h_i.

For future reference, band(j, τ) will indicate the τth band at the jth level of the DWT. The band at the top of the tree is then band(3,0). The bands at the next lower level, are band(2,0) and band(2,1). The scaling coefficients at some level j will be contained in band(j,0), and the wavelet coefficients at the same level will be stored in band(j,1).

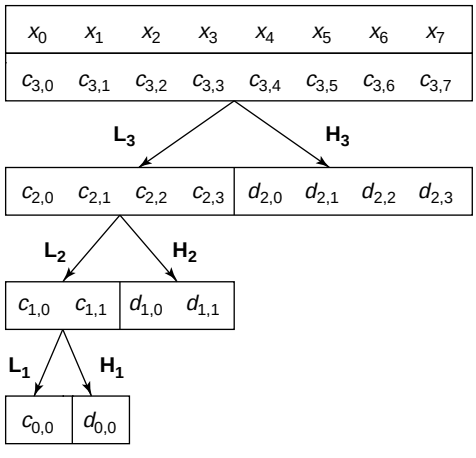

Figure 2 Two-band DWT of a spectrum of dimensionality 8.

Figure 3 shows the effect of performing the DWT on an artificially generated spectrum. The spectrum contains 256 points and consists of a sine curve with a period of 2, sampled over $-\pi$ to π, a block pulse, and another sine curve which has a period of 5 over the same interval $-\pi$ to π. For display purposes only, the DWT is shown for the first six levels $j = 8,7,6,5,4,3$. The original spectrum in band(8,0) undergoes a lowpass filtering process to give the scaling coefficients in band(7,0). The original spectrum also undergoes the highpass filtering process to give the wavelet coefficients which lie in band(7,1).

Next, the scaling coefficients in band(7,0) are passed through the lowpass and highpass filters to give the scaling coefficients in band(6,0) and the wavelet coefficients in band(6,1). The same procedure continues with the scaling coefficients from band(6,0) being filtered to give the scaling coefficients in band(5,0) and the wavelet coefficients in band(5,1). This process could continue for $\log_2 256 = 8$ levels, in which case there would be one scaling coefficient and one wavelet coefficient.

As one moves down the tree, the filtered signals in the scaling bands become smoother and smoother. The lowpass filtering process can be likened to a smoothing procedure followed by decimation (i.e. subsampling). The wavelet bands highlight the information which has not been captured by the scaling bands.

The components with the highest frequency are the first to be removed from the scaling coefficients and captured by the wavelet coefficients. Consider for example the scaling coefficients in band(5,0) – the remains of the $\sin(5t)$ are almost undetected in the scaling coefficients, whereas the coefficients of the $\sin(2t)$ curve are more perceptible. The most noticeable feature in the lower scaling levels of the tree is due to the low-frequency event of the block pulse.

4.2 Wavelet Packet Transform

So far we have only considered filtering the scaling coefficients, but it seems perfectly plausible to filter the wavelet coefficients. The wavelet packet transform (WPT)[5] is obtained by filtering both the scaling and wavelet coefficients.

The WPT has a tree-like structure, where each band in the transform produces two new children bands at the next lower level. The tree-like structure occurs because the detailed (or wavelet) coefficients are filtered through a lowpass and a highpass filter to obtain the next lower level of the WPT. This is done in the same way that the smoothed (or scaling coefficients) are filtered. It is convenient if we introduce the notation $\mathbf{x}^{[j]}(\tau)$

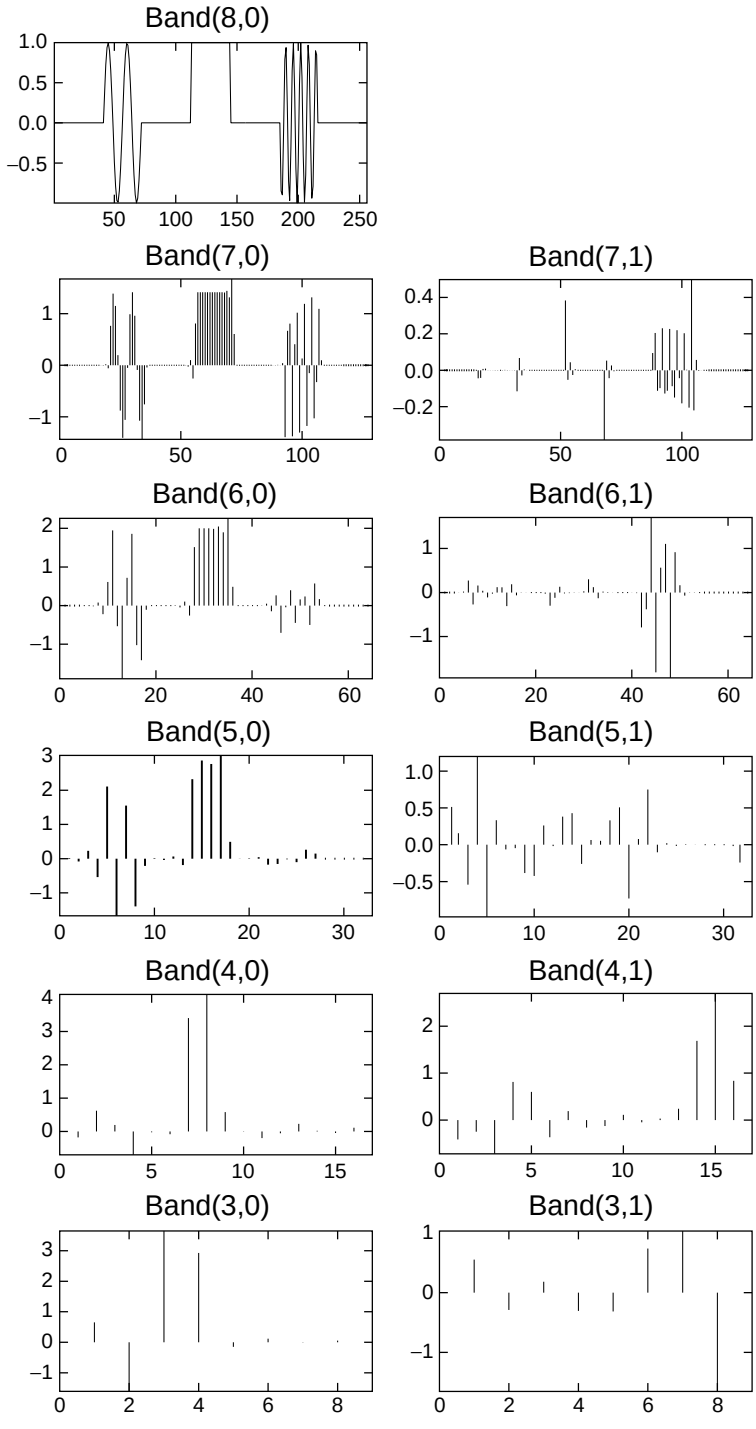

Figure 3 Two-band DWT performed on a generated spectrum to level 3.

to represent the wavelet packet coefficients (WPCs) which occur at the jth level in the τth band of the decomposition.

Figure 4 presents the structure of a WPT for some discretely sampled signal $\mathbf{x} = (x_0, x_1, \ldots, x_{2^J-1})^T = \mathbf{x}^{[J]}(0)$. The $(J-1)$ level of the WPT would be obtained as for the DWT, that is the data are passed through a lowpass and a highpass filter so that

$$\mathbf{x}^{[J-1]}(0) = \mathbf{L}_J\, \mathbf{x}^{[J]}(0)$$
$$\mathbf{x}^{[J-1]}(1) = \mathbf{H}_J\, \mathbf{x}^{[J]}(0)$$

For the WPT, the number of bands doubles from one level to the next (lower) level, since each of the bands in the previous level is passed through a lowpass and a highpass filter. At the next level, there will be four bands of WPCs which are obtained by

$$\mathbf{x}^{[J-2]}(0) = \mathbf{L}_{J-1}\,\mathbf{x}^{[J-1]}(0)$$
$$\mathbf{x}^{[J-2]}(1) = \mathbf{H}_{J-1}\,\mathbf{x}^{[J-1]}(0)$$
$$\mathbf{x}^{[J-2]}(2) = \mathbf{L}_{J-1}\,\mathbf{x}^{[J-1]}(1)$$
$$\mathbf{x}^{[J-2]}(3) = \mathbf{H}_{J-1}\,\mathbf{x}^{[J-1]}(1)$$

The same procedure may continue until there is one WPC in each of the bands. In general, $\mathbf{x}^{[j]}(0) = \mathbf{c}_j$ and $\mathbf{x}^{[j]}(1) = \mathbf{d}_j$.

4.3 Local Discriminant Basis Algorithm

The local discriminant basis (LDB) algorithm of Saito and Coifman[15] selects coefficients from the WPT and then uses these coefficients for classification. The selection of the coefficients involves several steps. The first step is to perform the WPT for each object $\mathbf{x}_{i(r)}$ for $i = 1, \ldots, n, \quad r = 1, \ldots, R$. Saito and Coifman[15] then calculate what they refer to as an ''energy map''. This is done for each class $r = 1, \ldots, R$. The energy maps have the same structure as the WPT, hence the same indices are used to locate items within the energy map (or tree). If $\mathbf{e}^{[j]}_{(r)}(\tau)$ denotes the energy coefficients in band(j, τ) of the energy map for class r, then Equation (18) represents the sum of squares of the coefficients that occur in the same position of the wavelet packet tree divided by a normalization constant.

$$\mathbf{e}^{[j]}_{(r)}(\tau) = \frac{\operatorname{diag}\left(\sum_{i=1}^{n_r} \mathbf{x}^{[j]}_{i(r)}(\tau) \quad \mathbf{x}^{[j]}_{i(r)}(\tau)\right)^T}{\text{const}} \tag{18}$$

The energy maps are obtained from the data objects that belong to class r. The notation $\mathbf{x}^{[j]}_{i(r)}(\tau)$ refers to the WPCs in band(j, τ) of the WPT produced from the object vector $\mathbf{x}_{i(r)}$.

Once the energy maps have been constructed for each class, Coifman[15] apply the best basis algorithm[26] to find an *optimal tree*. This involves selecting bands of coefficients which run horizontally across the WPT without overlapping vertically, i.e. no band can lie directly beneath or above another band. In the case of the LDB algorithm, this set of bands should collectively maximize some discriminant criterion $\mho$.

To find the bands of the optimal tree, one starts at the lowest level j_o in the WPT and works up. Initially, the criterion measures for each of the bands of coefficients at level $j_o + 1$ are compared with the criterion measures for the bands of the coefficients in the descendants at level j_o. If the criterion measure of the parent node is superior to that of the descendant nodes then the descendant nodes are deleted. If the descendant nodes produce a superior criterion measure then the descendant nodes are kept and the parent node is deleted. This procedure continues all the way to the top of the tree and the optimal tree consists of the bands that were not deleted in the elimination process.

A symmetric entropy discriminant criterion[10] is used in section 5 to measure the distinctness of the bands, as shown in Equation (19)

$$\begin{aligned}\mho(j, \tau) &= E_{\text{sym}} \quad \mathbf{e}^{[j]}_{(1)}(\tau), \ldots, \mathbf{e}^{[j]}_{(R)}(\tau) \\ &= \sum_{l} \sum_{r:\, r \neq l} E \quad \mathbf{e}^{[j]}_{(l)}(\tau), \mathbf{e}^{[j]}_{(r)}(\tau) \quad + E \quad \mathbf{e}^{[j]}_{(r)}(\tau), \mathbf{e}^{[l]}_{(r)}(\tau)\end{aligned} \tag{19}$$

where $E(\mathbf{w}, \mathbf{g})$ is calculated as shown in Equation (20)

$$E(\mathbf{w}, \mathbf{g}) = \sum_{i=1}^{p} w_i \log \frac{w_i}{g_i} \tag{20}$$

and $p = \text{length}(\mathbf{w}) = \text{length}(\mathbf{g})$. Here, $\mathbf{w}$ and $\mathbf{g}$ are two vectors whose similarity (or dissimilarity) is to be quantified. The entropy measure requires $w_i = g_i = 1$.

Once the optimal tree has been found, it is necessary to choose a subset of the WPCs, since the number of WPCs corresponding to the best basis is still equal to the dimensionality of the original data vector. Saito and Coifman[15] mention that one way of selecting a subset

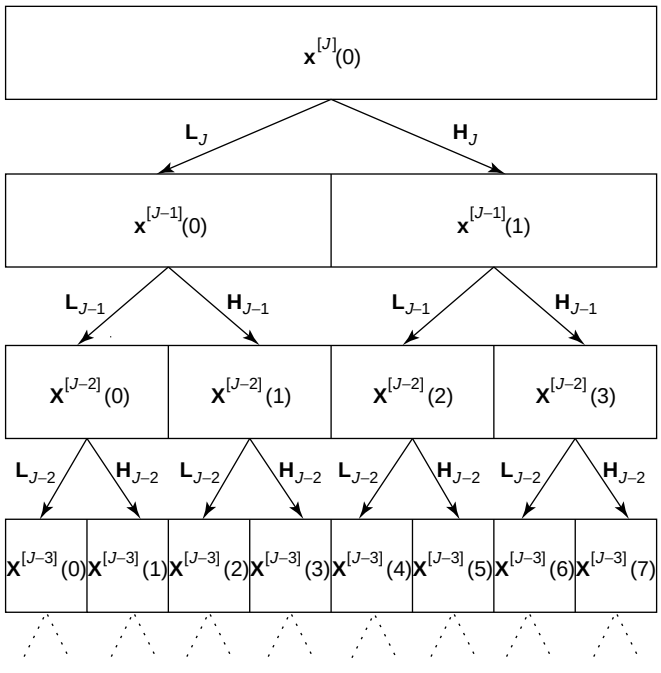

Figure 4 WPT.

is to choose the WPCs from the optimal tree that have the largest ratio of between-groups variance to within-groups variance, as described in section 3. Alternatively, one could select the WPCs based on another criterion, such as the entropy criterion.

4.4 Adaptive Wavelet Algorithm

There are several "off-the-shelf" wavelets that are defined by their respective filter coefficients. These include, for example, the Daubechies wavelets, coiflets, symlets and the Meyer and Haar wavelets (see for example Daubechies[27]). An initial problem is to decide which set (or family) of filter coefficients will produce the best results for a particular application. In practice, several families of filter coefficients may be trialled, and the family that produces the most desirable results is used. It can be advantageous however, to design your own task-specific filter coefficients rather than using a predefined set. The AWA of Mallet et al.[21] avoids the need to preselect a family of filter coefficients. In Mallet et al.,[21] the filter coefficients are designed with the aim of assigning objects to their respective classes with the greatest degree of accuracy.

4.4.1 m-Band Discrete Wavelet Transform

The AWA is presently based on the DWT rather than the WPT, but uses the more general m-band DWT ($m \geq 2, \quad m \in \mathcal{Z}^+$) where there is one lowpass filter, and $m-1$ highpass filters. We prefer to have the dimensionality $p = m^{(J-j_o)}$ and set $J = \lceil \log p / \log m \rceil$.

A three-band DWT for the spectrum $\mathbf{x} = (x_0, x_1, \ldots, x_8)$ is shown in Figure 5. There is one lowpass filter and two highpass filters, producing one set of scaling (or smoothed) coefficients and two sets of wavelet (or detailed) coefficients, respectively. As before, to traverse from one level to the next, only the scaling coefficients are filtered and the number of coefficients in each band is reduced by a factor of $m = 3$.

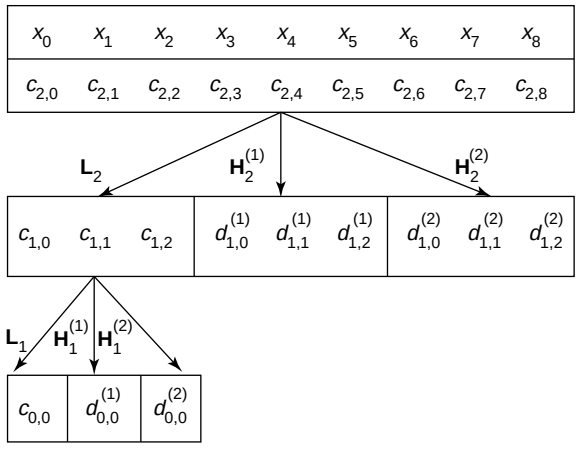

Figure 5 A three-band DWT.

The filtering operations for the general m-band DWT from some level j to level $j-1$ can be described by the matrix operations

$$\mathbf{c}_{j-1} = \mathbf{L}_j \mathbf{c}_j$$
$$\mathbf{d}_{j-1}^{(z)} = \mathbf{H}_j^{(z)} \mathbf{c}_j \qquad z = 1, \ldots, m-1$$

In summation notation one has Equations (21) and (22)

$$c_{j-1,i} = \sum_{k=0}^{N_f-1} \ell_k c_{j,mi+k} \tag{21}$$

$$d_{j-1,i}^{(z)} = \sum_{k=0}^{N_f-1} h_k^{(z)} c_{j,mi+k} \qquad z = 1, \ldots, m-1 \tag{22}$$

where it is assumed

$$c_{j,k} = c_{j,m^j+k}$$
$$d_{j,k}^{(z)} = d_{j,m^j+k}^{(z)}$$

Following the same notation as introduced earlier, band(j, τ) will be referred to as the τth band at the jth level of the DWT. The band at the top of the tree in Figure 5 is band(2,0). At the next level the bands from left to right are referred to as band(1,0), band(1,1) and band(1,2). Similarly, the bands in the last level of the DWT are band(0,0), band(0,1) and band(0,2).

The AWA provides one way of generating suitable filter coefficients for performing the m-band DWT ($m \geq 2$). This is done by placing constraints on the filter coefficients and then searching for suitable filter coefficients that satisfy these constraints. More details on the m-band DWT are available in the literature.[5,28,29]

4.4.2 Filter Coefficient Conditions

There are restrictions imposed on the filter coefficients so that a two-band (or higher) DWT or WPT can be performed. The adaptive wavelet should also adhere to these conditions. Let $\boldsymbol{A}$ denote the $m \times N_f$ matrix, with the first row containing the lowpass filter coefficients and the remaining $m-1$ rows containing the sets of highpass filter coefficients, as shown in Equation (23).

$$\boldsymbol{A} = \begin{matrix} \ell_0 & \ell_1 & \ldots & \ell_{N_f-1} \\ h_0^{(1)} & h_1^{(1)} & \ldots & h_{N_f-1}^{(1)} \\ \vdots & \vdots & & \vdots \\ h_0^{(m-1)} & h_1^{(m-1)} & \ldots & h_{N_f-1}^{(m-1)} \end{matrix} \tag{23}$$

$\boldsymbol{A}$ can be partitioned into $q+1$, $m \times m$ sub-matrices as follows: $\boldsymbol{A} = (\boldsymbol{A}_0 \, \boldsymbol{A}_1 \, \ldots \, \boldsymbol{A}_q)$. The conditions imposed on

the filter coefficients so that an orthogonal wavelet basis exists can be summarized as follows.[30,31]

1. Orthogonality: $\sum_k \boldsymbol{A}_k \boldsymbol{A}_{k+i}^T = \delta_{0i}\boldsymbol{I}$, where $\delta_{0i} = 1$ if $i = 0$, and zero otherwise, and $\boldsymbol{I}$ is the identity matrix.
2. Regularity: $\sum_k \ell_k = \sqrt{m}$.
3. The Lawton matrix condition, $M_{ij} = \sum_k \ell_k \ell_{k+j-mi}$, must have 1 as a simple eigenvalue.

(More sophisticated wavelet and scaling functions will satisfy additional conditions.)

The AWA attempts to find the elements of $\boldsymbol{A}$ which optimize the discriminant problem at hand. Instead of optimizing over each filter coefficient in $\boldsymbol{A}$, the AWA makes use of the factorized form[21,30,31] of a wavelet matrix and the conditions placed therein to reduce the number of parameters to be optimized. It can be shown[31] that $\boldsymbol{A}$ can be constructed from some set of normalized vectors denoted by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_q$ and $\boldsymbol{v}$, where the total number of parameters to be optimized is $N_f - 1$.

Table 1 Description of the spectral data sets used for classification

Data		Class 1	Class 2	Class 3	Class 4	Class 5	Total
Seagrass	Train	55	55	55	–	–	165
	Test	34	34	34	–	–	102
Mineral	Train	20	20	20	20	20	100
	Test	20	20	20	20	20	100
p-Xylene	Train	25	25	25	–	–	75
	Test	25	25	25	–	–	75
Butanol	Train	21	27	–	–	–	48
	Test	21	26	–	–	–	47

5 APPLICATIONS

5.1 Data Sets

Four spectral data sets were used to investigate the various classification procedures. Each data set initially contained 512 variables (i.e. $p = 512$). The data sets are referred to as the seagrass, mineral, *p*-xylene and butanol data sets. The number of training and testing spectra in the group categories is listed in Table 1 for each set of data. Sample spectra from each of these data sets are presented in Figures 6–9.

5.2 Methodology

There are many different classification and feature extraction strategies that can be applied to spectral data.

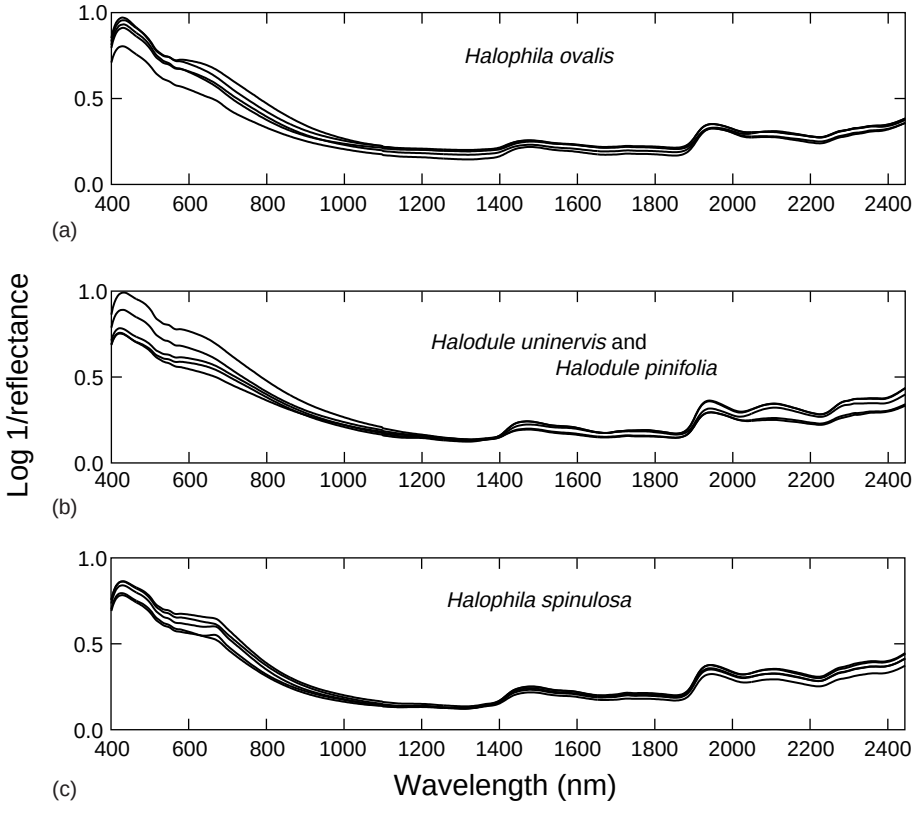

Figure 6 Five sample spectra from the seagrass data.

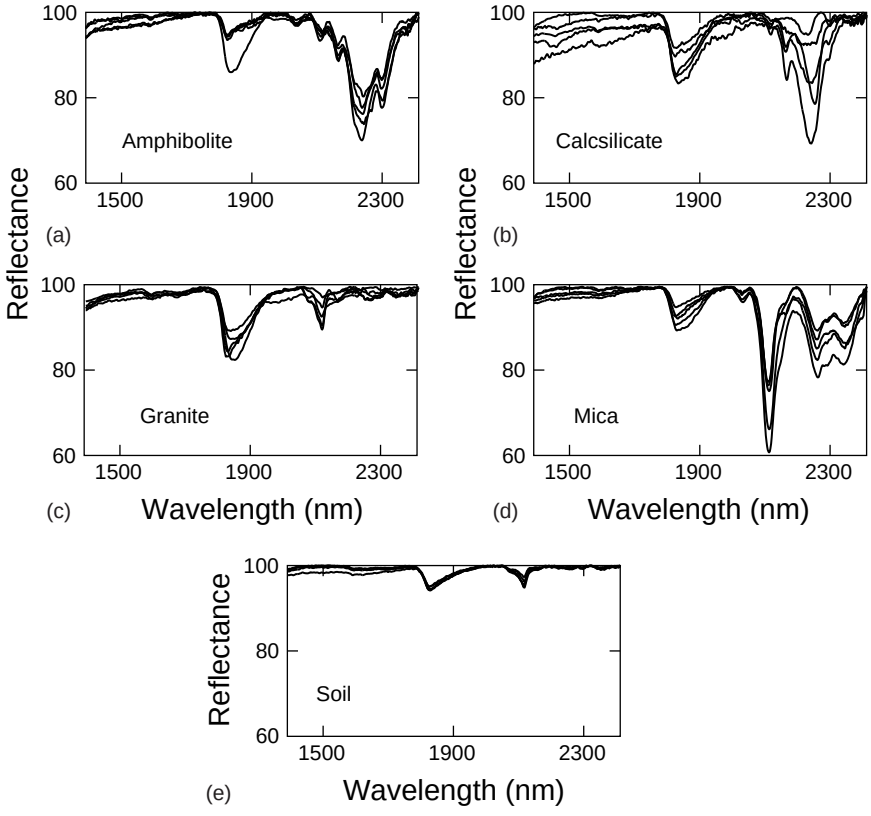

Figure 7 Five sample spectra from the mineral data.

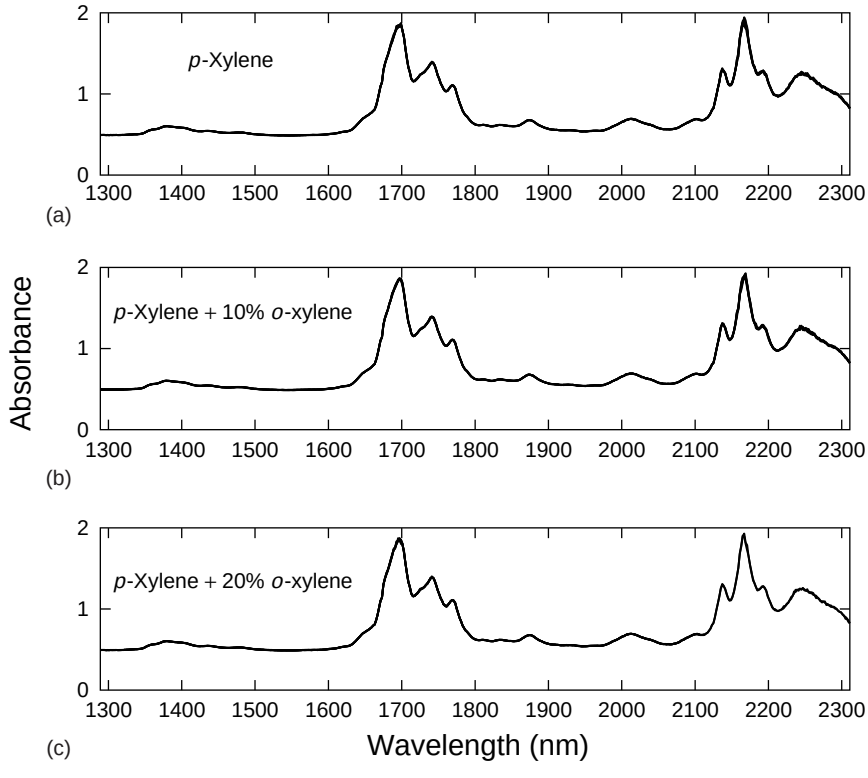

Figure 8 Five sample spectra from the *p*-xylene data.

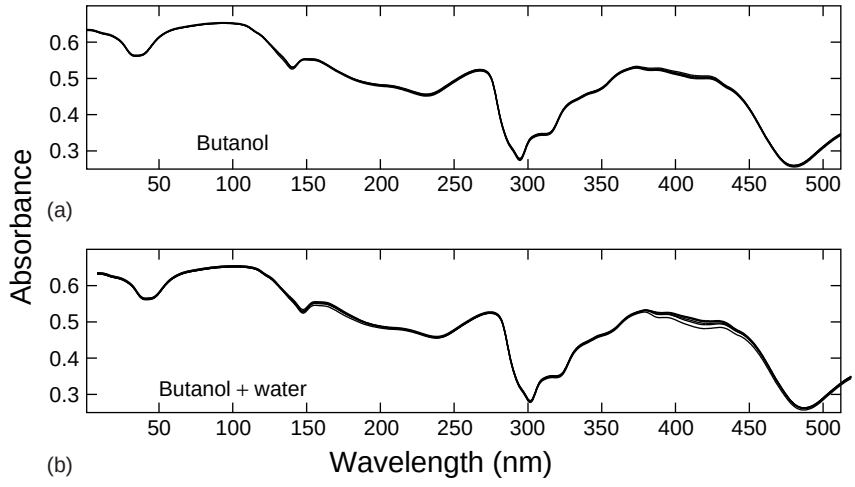

Figure 9 Five sample spectra from the butanol data.

Collectively, the term strategy refers to the particular classifier utilized, such as RDA, or a feature extraction technique. A strategy may even refer to the combination of feature extraction methods with a particular classifier. The classification strategies are described in the following sections.

5.2.1 Stepwise Bayesian Linear and Quadratic Discriminant Analysis

A stepwise feature selection method is used to select variables from the original data set. These selected variables then undergo BLDA and BQDA. If the stepwise feature selection procedure is used in combination with BLDA, the classification strategy is referred to as SBLDA (stepwise Bayesian linear discriminant analysis). Likewise, if BQDA is the classifier then the procedure is called SBQDA (stepwise Bayesian quadratic discriminant analysis).

The stepwise procedure considered here starts with an empty subset, and at each step adds the variable (or wavelength) that produces the largest increase in the CVQPM. The procedure ceases to enter variables into the model when one of the following stopping criteria is reached.

- The change in the correct classification rate (CCR) is less than $1/n$, where n is the number of samples in the data set. That is, from one iteration to the next, the inclusion of another variable does not improve the CCR by more than $1/n$.
- The CCR reaches 100%.

The results of the stepwise strategies for both SBLDA and SBQDA are summarized in Table 2. SBLDA produced a higher test CCR for the seagrass and *p*-xylene data, while SBQDA produced a higher test CCR for the mineral and butanol data. The butanol data set posed the greatest challenge for the stepwise procedures.

Table 2 Percentage of correctly classified objects for SBLDA and SBQDA

Data		SBLDA	SBQDA
Seagrass	Train	99.39	100
	Test	**100**	97.06
Mineral	Train	100	100
	Test	88	**93**
p-Xylene	Train	100	100
	Test	**87.33**	78.67
Butanol	Train	87.50	89.58
	Test	72.39	**76.60**

5.2.2 Flexible and Penalized Discriminant Analysis

The regression method used for FDA was bruto. Bruto is an additive regression model which computes the terms in the model using smoothing splines. The bruto procedures also accommodate a feature selection routine. The regression method used for PDA was ridge regression. Both FDA and PDA were applied to the original data sets without prior feature selection. FDA, when used in conjunction with bruto, can be seen as a monothetic method since, while it is initially supplied with all of the variables, it eventually selects a much smaller subset. PDA (and RDA) on the other hand can be seen as polythetic methods, since all of the variables are entered into the final model. The classification results for FDA and PDA are displayed in Table 3. For all data sets except for the mineral data, FDA outperformed PDA in terms of the misclassification rate of the test data. Again the

Table 3 Percentage of correctly classified objects for FDA and PDA

Data		FDA	PDA
Seagrass	Train	98.18	96.97
	Test	**99.02**	95.10
Mineral	Train	100	100
	Test	95	**100**
p-Xylene	Train	100	86.67
	Test	**86.67**	81.33
Butanol	Train	75	43.68
	Test	**70.21**	43.75

Table 4 Percentage of correctly classified objects for RDA and the regularization parameters

Data		RDA	(λ, γ)
Seagrass	Train	99.39	(1.00, 0.25)
	Test	99.02	
Mineral	Train	100	(1.00, 0.25)
	Test	95	
p-Xylene	Train	100	(0.25, 0.25)
	Test	100	
Butanol	Train	87.50	(1.00, 0.25)
	Test	87.23	

butanol data set was the more challenging data set to classify, especially for PDA.

5.2.3 *Regularized Discriminant Analysis*

RDA was applied to each data set with 25 different (λ, γ) combinations. The combinations resulted from each pairwise combination of the settings 0.00, 0.25, 0.50, 0.75 and 1.00. Table 4 presents the training and test classification rates for each data set as well as the corresponding (λ, γ) grid settings that were used to produce the respective classification rates.

Again, reasonable classification rates were experienced for the seagrass and mineral data. Of interest however is the fact that all of the spectra from the training and testing *p*-xylene data were correctly classified using RDA. From previous applications, the *p*-xylene data (along with the butanol data) posed a challenge for many of the classification strategies.

The $(\lambda, \gamma) = (1.00, 0.25)$ setting was found to be optimal for the seagrass, mineral and butanol data sets. This indicates that for these data sets a pooled covariance matrix is preferred to one that is not pooled. The combination $(\lambda, \gamma) = (0.25, 0.25)$ was used for the *p*-xylene data set, as it weights the individual class covariance matrices more heavily than the pooled class covariance matrix.

5.2.4 *Banded Bayesian Linear and Quadratic Discriminant Analysis*

Perhaps the simplest strategy to incorporate wavelet-based methods with classification strategies is to choose a family of filter coefficients, perform the DWT, and then supply all the coefficients from some band(j, τ) to a classifier. The term ''coefficients'' is used loosely to refer to either wavelet or scaling coefficients. We used the $m = 2$ band DWT with $N_f = 16$ filter coefficients from the Daubechies family. The classifiers used were BLDA and BQDA. Since these are low-dimensional methods, the number of coefficients in band(j, τ) should be small when compared with the sample size, so that an ill- or poorly-posed situation is avoided. The banded approach is a very simple procedure for feature selection of the wavelet coefficients. Previously, Bos and Vrielink[18] used a similar approach, except that the bands of coefficients were supplied to neural networks. The coefficients from band(3,0), band(3,1), band(4,0) and band(4,1) were supplied to the classifiers.

The terms BBLDA (banded Bayesian linear discriminant analysis) and BBQDA (banded Bayesian quadratic discriminant analysis) are adapted if the bands of coefficients are supplied to BLDA or BQDA, respectively. The classification results for BBLDA and BBQDA are summarized in Tables 5 and 6 respectively. The wavelet bands tended to dominate for BBLDA for the mineral, *p*-xylene and butanol data sets. For BBQDA, numerical

Table 5 Classification results for BBLDA

Data		$\mathbf{X}^{[3]}(0)$	$\mathbf{X}^{[3]}(1)$	$\mathbf{X}^{[4]}(0)$	$\mathbf{X}^{[4]}(1)$
Seagrass	Train	98.79	99.39	100	100
	Test	**100**	98.04	**100**	99.02
Mineral	Train	97	95	97	98
	Test	87	90	94	**98**
p-Xylene	Train	62.67	68.00	81.33	80.00
	Test	50.67	58.67	56.00	**61.33**
Butanol	Train	85.42	87.50	93.75	87.50
	Test	82.98	82.98	76.60	**87.23**

The numbers in the square brackets are the level of the transform.

Table 6 Classification results for BBQDA

Data		$\mathbf{X}^{[3]}(0)$	$\mathbf{X}^{[3]}(1)$	$\mathbf{X}^{[4]}(0)$	$\mathbf{X}^{[4]}(1)$
Seagrass	Train	100	100	100	100
	Test	**100**	99.02	**100**	**100**
Mineral	Train	100	100	–	–
	Test	93	**96**	–	–
p-Xylene	Train	88	86.67	100	100
	Test	66.67	72.00	**81.33**	76.00
Butanol	Train	89.58	87.50	100	100
	Test	**74.47**	72.34	63.83	57.45

The numbers in the square brackets are the level of the transform.

instabilities arose for the mineral data when 16 coefficients were supplied to the classifier. This can be attributed to the fact that, for BQDA, the class sample size should be large compared with the dimensionality. For the mineral data there are 20 objects per class, which is only marginally larger than the number of coefficients (16), hence it was not possible to produce accurate results for this setting.

5.2.5 *Stepwise Feature Extraction from the Discrete Wavelet Transform*

A more sophisticated procedure for incorporating wavelet-based methodologies with classification strategies is to use a stepwise selection procedure to select the coefficients which are to be supplied to the respective classifier. As for the banded procedures described in the previous section, the two-band DWT is performed to level 3 using a Daubechies wavelet defined by 16 filter coefficients. The bands shaded in grey in Figure 10 indicate the total set of bands from which wavelet or scaling coefficients may be selected from the stepwise procedure. The same stepwise routine used for SBLDA and SBQDA is also applied here to the coefficients $\mathbf{c}_3, \mathbf{d}_3, \mathbf{d}_4, \ldots, \mathbf{d}_8$. If the coefficients are supplied to BLDA we refer to this classification procedure as SWBLDA, and when the classifier is BQDA the procedure is referred to as SWBQDA.

Also indicated in Figure 10 are the coefficients selected by SWBLDA (asterisk) and SWBQDA (circle). It should be noted that for the challenging *p*-xylene and butanol data sets there are more coefficients selected from higher levels in the DWT, indicating that the classifiers are discriminating using information relating to high-frequency events. For the seagrass and mineral data, however, it should be noted that the classifiers select coefficients from the lower levels, and in fact some scaling coefficients are selected, indicating that information relating to the basic shape of the mineral and seagrass data has been utilized.

The classification results for SWBLDA and SWBQDA are shown in Table 7. The *p*-xylene data set is the only data set which gives a higher test CCR using SWBQDA as opposed to SWBLDA. Whilst SWBLDA does outperform SWBQDA for three of the four data sets in terms of the test misclassification rate, SWBLDA is often outperformed by other classification strategies. SWBLDA does however compare favorably with other classification strategies when applied to the butanol data.

5.2.6 *Local Discriminant Basis*

LDB adds even greater sophistication than the stepwise and banded procedures for combining wavelet features and classification methods. In this section the WPT (with $m = 2$) was calculated to level 3 using the Daubechies

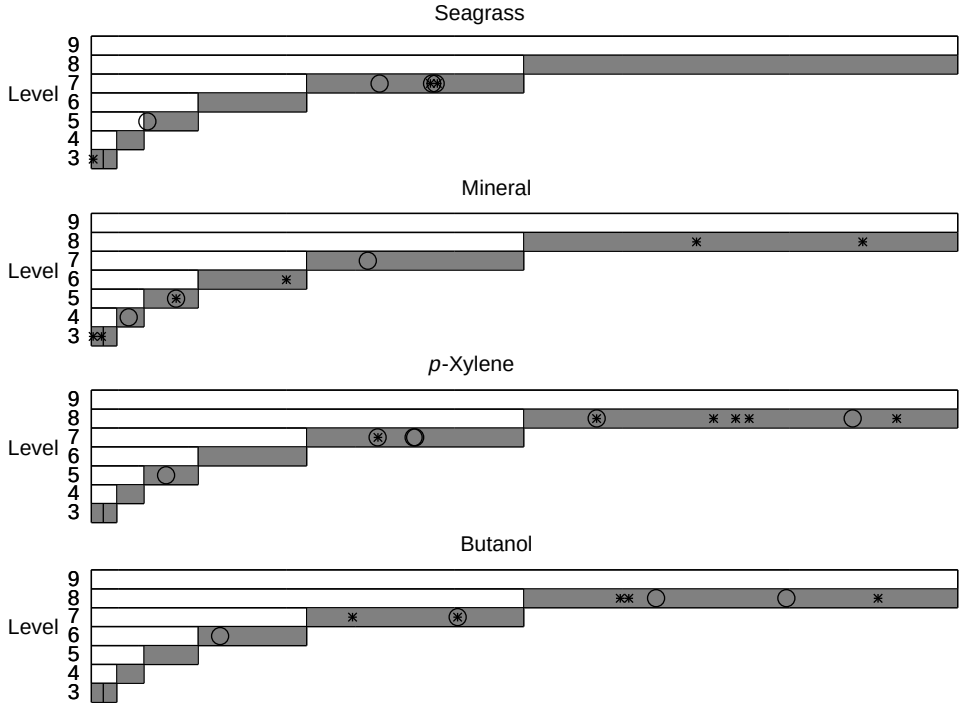

Figure 10 Coefficients selected from the DWT by stepwise wavelet feature extraction using Bayesian linear discriminant analysis (SWBLDA) (asterisks) and stepwise wavelet feature extraction using Bayesian quadratic discriminant analysis (SWBQDA) (circles).

Table 7 Percentage of correctly classified objects for SWBLDA and SWBQDA

Data		SWBLDA	SWBQDA
Seagrass	Train	99.39	100
	Test	**98.04**	97.06
Mineral	Train	100	100
	Test	**93**	90
p-Xylene	Train	98.67	97.33
	Test	81.33	**82.67**
Butanol	Train	85.42	91.67
	Test	**85.11**	74.47

filter coefficients with $N_f = 16$. Once the wavelet packet decomposition had been formed, the best basis (optimal tree) was determined from the energy maps. The criterion used to form the best basis was the symmetric entropy criterion (Equation 19). Figure 11 marks the best basis selected by the LDB algorithm for each of the sets of data. For the *p*-xylene data, the selected best basis was the original data.

Once the best basis had been found, a subset of WPCs was selected from the best basis. It was decided to select the 16 WPCs based on the same discriminant measure which produced the best basis, i.e. symmetric entropy.

The asterisks in Figure 11 show the positions of the 16 WPCs which were selected from the best basis and supplied to the classifier BLDA. The WPCs are quite clustered, which is a likely consequence of selecting the coefficients by a univariate strategy, that is without consideration being given to previously selected features. There is also a tendency for the WPCs to be largely selected from band(4,0) and band(4,1) of the WPT. These bands contain the same coefficients as those in band(4,0) and band(4,1) from the DWT. It should be noted, however, that in Figure 10 there is only one coefficient selected from either of these bands (see mineral data). The conflict in feature selection is a likely consequence of using different selection strategies for selecting features.

The sixteen WPCs with the largest symmetric entropy measures were supplied to the BLDA classifier in a top-down approach using the first $1, 2, \ldots, 16$ coefficients for classification. That is, initially a single WPC (with the largest discriminant measure) was supplied to the classifier. Then, the two WPCs with the largest discriminant measures were supplied to the classifier. This procedure continued until all 16 coefficients had formed part of the BLDA model.

Table 8 gives the classification rates for the training and testing data for each of the discriminant data sets where the first $1, \ldots, 16$ WPCs were selected from the best basis, and supplied to BLDA. The numbers in bold type had the largest (test) CCR and the fewest terms in the discriminant model. This application of the LDB approach closely followed that outlined by Saito and

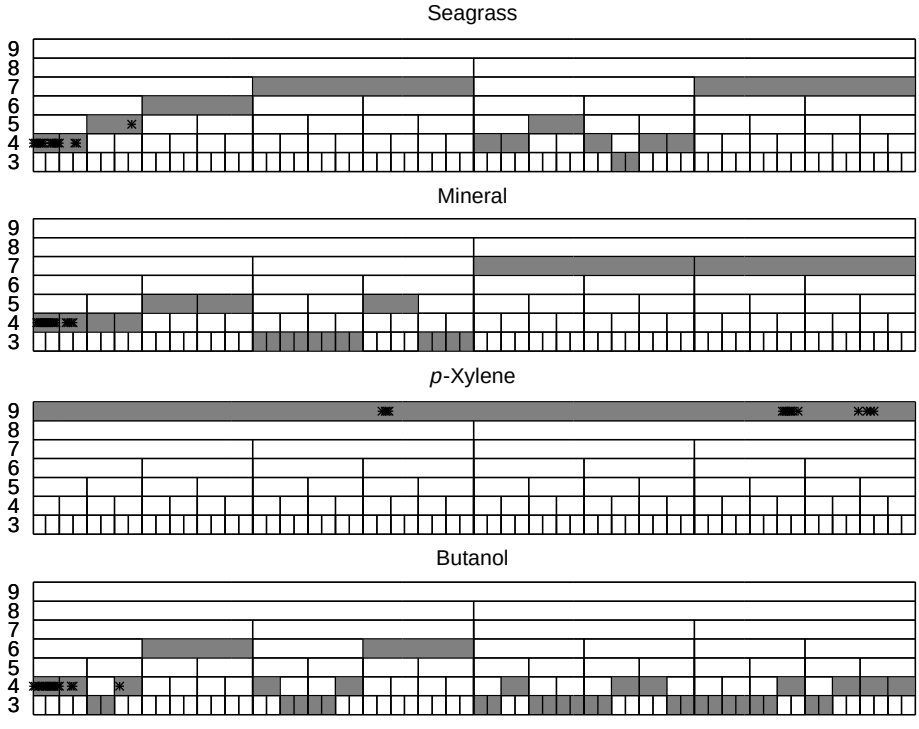

Figure 11 Selected wavelet coefficients (asterisks) from the best bases.

Table 8 Classification performance of the LDB algorithm

No. of WPCs	Seagrass		Mineral		*p*-Xylene		Butanol	
	Train	Test	Train	Test	Train	Test	Train	Test
1	74.55	76.47	67	72	42.67	41.33	72.92	74.47
2	73.33	69.61	87	89	48.00	65.33	75.00	76.60
3	86.67	86.27	88	90	54.67	53.33	79.17	80.85
4	93.94	94.12	93	87	68.00	62.67	81.25	80.85
5	90.91	89.22	93	87	69.33	57.33	83.33	**87.23**
6	93.33	91.18	94	85	72.00	57.33	83.33	80.85
7	98.18	97.06	94	88	78.67	77.33	83.33	80.85
8	100.00	**100.00**	96	88	78.67	**78.67**	83.33	85.11
9	100.00	100.00	96	88	78.67	77.33	85.42	85.11
10	100.00	100.00	96	88	84.00	73.33	83.33	82.98
11	99.39	99.02	97	**93**	84.00	72.00	83.33	80.85
12	99.39	99.02	97	93	84.00	77.33	85.42	82.98
13	100.00	100.00	98	91	88.00	76.00	85.42	82.98
14	100.00	100.00	98	92	89.33	77.33	83.33	78.72
15	100.00	100.00	98	92	89.33	77.33	87.50	76.60
16	100.00	100.00	99	93	92.00	73.33	87.50	76.60

Coifman,[15] the main difference being that BLDA was used, as opposed to FLDA.

5.2.7 *Adaptive Wavelet Algorithm*

Previous feature extraction methods involving the DWT were performed using filter coefficients from the Daubechies family. Many filter coefficients could have been chosen, but the Daubechies filter coefficients were chosen as they tend to be popular. There is no reason, however, that another set of coefficients could not have been chosen. Of course, the problem lies in deciding whether or not better results might be obtained using another set or family of filter coefficients. In this section, task-specific filter coefficients are designed using the AWA described in section 4.3. The underlying principle of the AWA is to avoid deciding which set of filter coefficients to use, and instead to design the filter coefficients to suit the task in hand, which in this case is discriminant analysis.

The AWA is applied using several settings of the m, q and j_o parameters, where j_o is the lowest level of the DWT and $N_f = m(q+1)$. The particular (m, q, j_o) triplets used were (4,3,2), (4,2,2), (8,1,1), (2,5,3), (2,5,4), (2,7,3), and (2,7,4). These settings were chosen because (1) they provide suitable ratios of the dimensionality of the wavelet bands to the sample size and, (2) the number of filter coefficients is of reasonable size. We considered $N_f = 12$ and $N_f = 16$. Mallet et al.[21] describe some heuristics for choosing these parameters.

The discriminant criterion function implemented by the AWA is the CVQPM criterion function. A form of banded selection was performed, whereby the CVQPM was calculated from a band of coefficients. The same coefficients were later supplied to the classifier. The value τ was chosen as the band which gave the highest CVQPM value at initialization for a particular (m, q, j_o) triplet. For more details on this banded selection process see Mallet et al.[21]

Table 9 Percentage of correctly classified objects for AWA

	m	q	j_o	N_f	N_{coef}	τ	Train	Test
Seagrass	4	3	2	16	8	1	100	100
Mineral	2	7	4	16	16	1	99	99
p-Xylene	2	5	4	12	16	1	94.67	86.67
Butanol	2	5	3	12	8	1	93.75	87.23

The settings which produced the highest test CCR are displayed in Table 9 for each of the data sets. The AWA performed consistently well for each data set. It is interesting to note that for three of the four data sets the more common $m = 2$ band DWT was used.

6 SUMMARY OF RESULTS

This section summarizes the results of the various classification procedures which were accumulated in section 5. Figure 12 displays the CCRs for each of the classification strategies. Only the CCRs for the testing data are displayed. To enable easier interpretation of Figure 12, crosses have been used to indicate the results of the high-dimensional methods, circles have been used to indicate the low-dimensional classification performed on the original data, and the asterisks indicate the low-dimensional classification methods based on the coefficients from the DWT. To

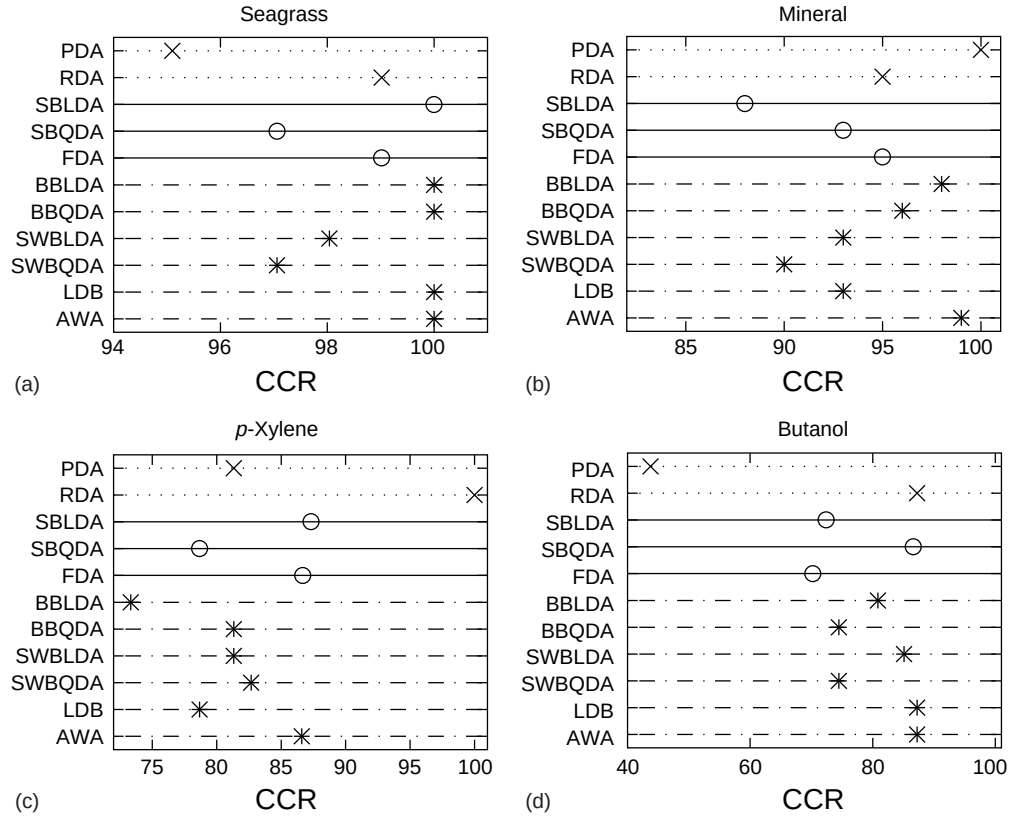

Figure 12 CCRs for each of the discriminant strategies.

Table 10 CCRs for the LDB algorithm

Method	Seagrass		Mineral		*p*-Xylene		Butanol	
	Train	Test	Train	Test	Train	Test	Train	Test
PDA	96.97	95.10	100	**100**	86.67	81.33	43.68	43.75
RDA	99.39	99.00	100	95	100	**100**	87.50	**87.23**
SBLDA	99.39	**100**	100	88	100	87.33	87.50	72.39
SBQDA	100	97.10	100	93	100	78.67	89.50	76.60
FDA	98.18	99.00	100	95	100	86.67	75.00	70.21
BBLDA	100	**100**	98	98	80.00	61.33	87.50	**87.23**
BBQDA	100	**100**	100	96	100	81.33	89.58	74.47
SWBLDA	99.39	98.00	100	93	98.67	81.33	85.42	85.11
SWBQDA	100	97.10	100	90	97.33	82.67	91.67	74.47
LDB	100	**100**	97	93	78.67	78.67	83.33	**87.23**
AWA	100	**100**	99	99	94.67	86.67	93.75	**87.23**

further enhance interpretation of Figure 12, the line types also change with the marker indicators. Table 10 shows the actual CCRs for the training and testing data sets.

The most salient issues that have arisen from the comparative analysis are as follows.

- RDA tends to outperform the other high-dimensional classifier PDA, with the exception being for the mineral data where PDA outperforms RDA.
- If RDA is compared with the low-dimensional methods which entail some form of feature extraction (or selection), then it can clearly be seen that RDA also consistently produces high classification results across each of the sets of data. For the mineral data, however, RDA is outperformed by three of the wavelet-based approaches (and PDA).
- For the classification methods SBLDA, SBQDA and FDA there does not appear to be any one approach that consistently performs well.

- The AWA consistently produces high test classification rates and compares favourably against the high-dimensional classifiers and the classification strategies involving a low-dimensional classifier.
- There appears to be some variability in the performance of the wavelet-based techniques in that, with the exception of the AWA, no wavelet method performs consistently well or consistently outperforms the other strategies.

7 CONCLUSION

This paper has presented several different strategies that can be used for classifying spectral and similar forms of data. We considered two main approaches – using a high-dimensional classifier for all of the original data, and using a low-dimensional classifier on selected features. The features considered were the original variables, and coefficients from the DWT were also explored. The objective of this paper was to compare the performances of each of these strategies. Based on the CCR it was noted that the high-dimensional classifier RDA performed consistently well. It was difficult to observe if any low-dimensional classifier consistently produced outstanding performances. This subjectivity arose due to the different feature selection strategies which were implemented in conjunction with the classifiers BLDA, BQDA and FDA. However, the AWA did consistently produce high CCRs and compared favourably with RDA. This demonstrates that it can be advantageous to design your own task-specific filter coefficients.

Throughout this paper, much attention has been focused on the CCR for comparing techniques. It should be mentioned that there are many other attributes which should be addressed when comparing techniques. Computational complexity is one item which should be considered. Although a comprehensive analysis of computational complexities was not undertaken, in our experience RDA and the AWA tend to be much more computationally expensive than the remaining methods. This is obviously due to the optimization routines which form part of these techniques.

Another attribute is the qualitative information and interpretability that can be obtained from the models. For example, whether the model can discover which variables are more important, or whether the model has graphical capabilities for exploring group separation, for instance. The Bayesian classification methods BLDA, BQDA and RDA are difficult to interpret qualitatively. However, when BLDA and BQDA are combined with stepwise methods then some insight into discriminatory regions of the data is obtained by noting which variables have been selected. This is the case whether selection is from the original data or from wavelet coefficients. The wavelet procedures which involve a banded selection of coefficients do not allow for easy interpretation, although if FLDA was applied to these features then one could examine the size of the discriminant coefficients to indicate which coefficients are more important. A useful feature of both PDA and FDA is that these methods can generate discriminant plots.

To provide more detailed comparisons, further work on more data sets, including simulated data, needs to be performed. Here we have simply introduced some traditional and modern methods for classifying spectral data and quantitatively examined their performances using the CCR, to help provide the reader with more insight into these techniques.

ABBREVIATIONS AND ACRONYMS

AWA	Adaptive Wavelet Algorithm
BBLDA	Banded Bayesian Linear Discriminant Analysis
BBQDA	Banded Bayesian Quadratic Discriminant Analysis
BLDA	Bayesian Linear Discriminant Analysis
BQDA	Bayesian Quadratic Discriminant Analysis
CCR	Correct Classification Rate
CVQPM	Cross-validated Quadratic Probability Measure
DWT	Discrete Wavelet Transform
FDA	Flexible Discriminant Analysis
FLDA	Fisher's Linear Discriminant Analysis
FT	Fourier Transform
LDB	Local Discriminant Basis
MARS	Multivariate Adaptive Regression Splines
NIR	Near-infrared
PDA	Penalized Discriminant Analysis
PLSR	Penalized Least Squares Regression
QPM	Quadratic Probability Measure
RDA	Regularized Discriminant Analysis
SBLDA	Stepwise Bayesian Linear Discriminant Analysis
SBQDA	Stepwise Bayesian Quadratic Discriminant Analysis
SWBLDA	Stepwise Wavelet Feature Extraction Using Bayesian Linear Discriminant Analysis
SWBQDA	Stepwise Wavelet Feature Extraction Using Bayesian Quadratic Discriminant Analysis

WPC	Wavelet Packet Coefficient
WPT	Wavelet Packet Transform

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Statistical Quality Control in Clinical Laboratories

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis • Near-infrared Spectroscopy in Food Analysis

Process Instrumental Methods **(Volume 9)**
Near-infrared Spectroscopy in Process Analysis

Pulp and Paper **(Volume 9)**
Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Interpretation of Infrared Spectra, A Practical Approach • Quantitative Analysis, Infrared • Theory of Infrared Spectroscopy

REFERENCES

1. S. Aeberhard, D. Coomans, O. de Vel, 'Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings', *Pattern Recognit.*, **27**, 1065–1077 (1994).
2. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Boston, 1990.
3. G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 1992.
4. Y. Mallet, D. Coomans, O. de Vel, 'Recent Developments in Discriminant Analysis on High Dimensional Spectral Data', *Chemometr. Intell. Lab. Syst.*, **35**, 157–173 (1996).
5. G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, 1996.
6. J. Friedman, 'Regularized Discriminant Analysis', *J. Am. Stat. Assoc.*, **84**, 165–175 (1989).
7. T. Hastie, R. Tibshirani, A. Buja, 'Flexible Discriminant Analysis', *J. Am. Stat. Assoc.*, **89**, 1255–1270 (1994).
8. Y. Mallet, 'Wavelet Based Feature Extraction Methods for the Classification and Regression of Spectral Data', PhD, James Cook University, Townsville, 1998.
9. J. Friedman, 'Multivariate Adaptive Regression Splines (with discussion)', *Ann. Stat.*, **19**, 1–141 (1991).
10. T. Hastie, R. Tibshirani, *Generalized Additive Models*, Chapman & Hall, London, 1990.
11. L. Breiman, R. Ihaka, 'Nonlinear Discriminant Analysis Via Scaling and Ace', Technical Report, University of California, Berkeley, 1984.
12. T. Hastie, A. Buja, R. Tibshirani, 'Penalized Discriminant Analysis', *Ann. Stat.*, **2**, 73–102 (1995).
13. A. Hoerl, R. Kennard, 'Ridge Regression: Biased Estimation for Nonorthogonal Problems', *Technometrics*, **8**, 27–51 (1970).
14. W. Wu, B. Walczak, D. Massart, K. Prebble, I. Last, 'Spectral Transformation and Wavelength Selection in Near-infrared Spectra Classification', *Anal. Chim. Acta*, **315**, 243–255 (1995).
15. N. Saito, R. Coifman, 'Local Discriminant Bases', in *Mathematical Imaging: Wavelet Applications in Signal and Image Processing II*, eds. A.F. Laine, M.A. Unser, SPIE Proceedings Vol. 2303, Bellingham, WA, SPIE, 1994.
16. SAS Institute Inc., 'SAS/STAT User's Guide', Release 6.03 Edition, SAS Institute Inc., Cary, 1988.
17. D. Coomans, I. Broeckaert, *Potential Pattern Recognition in Chemical and Medical Decision Making Research Studies*, John Wiley & Sons, Letchworth, UK, 1965.
18. M. Bos, J. Vrielink, 'The Wavelet Transform for Preprocessing IR Spectra in the Identification of Mono- and Di-substituted Benzenes', *Chemometr. Intell. Lab. Syst.*, **23**, 115–122 (1994).
19. T. Bailey, T. Sapatinas, K. Powell, W. Krzanowski, 'Signal Detection in Underwater Sound Using Wavelets', *J. Am. Stat. Assoc.*, **93**(441), 73–83 (1998).
20. R. Learned, A. Wilsky, 'A Wavelet Packet Approach to Transient Signal Classification', *Appl. Comp. Harm. Anal.*, **2**, 265–278 (1995).
21. Y. Mallet, D. Coomans, J. Kautsky, O. de Vel, 'Classification Using Adaptive Wavelets for Feature Extraction', *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**, 1058–1066 (1997).
22. R. Tate, D. Watson, S. Eglen, 'Using Wavelets for Classifying Human In Vivo Magnetic Resonance Spectra', in *Wavelets and Statistics*, eds. A. Antoniadis, G. Oppenheim, Springer Verlag, New York, 373, 1995.
23. B. Walczak, B. van den Bogaert, D. Massart, 'Application of Wavelet Packet Transform in Pattern Recognition of Near-IR Data', *Anal. Chem.*, **68**, 1742–1747 (1996).
24. A. Bruce, H. Gao, 'S + Wavelets User's Manual Version 1.0', StatSci (a division of MathSoft, Inc.), Seattle, 1994.
25. M. Misiti, 'Wavelet Toolbox User's Guide', Math Works, Natick, MA, 1996.
26. R. Coifman, M. Wickerhauser, 'Entropy-based Algorithms for Best Basis Selection', *IEEE Trans. Inf. Theory*, **38**, 713–718 (1992).

27. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
28. P. Steffen, P.N. Heller, R.A. Gopinath, C.S. Burrus, 'Theory of Regular *m*-Band Wavelet Bases', *IEEE Trans. Signal Proc.*, **41**(2), 3497–3511 (1993).
29. P. Steffen, P. Heller, R. Gopinath, C. Burrus, 'Theory of *m*-Band Wavelet Bases', *IEEE Trans. Signal Proc.*, **41**, 3497–3511 (1993).
30. J. Kautsky, R. Turcajová, 'Adaptive Wavelets for Signal Analysis', in *Proceedings 6th International Conference on Computer Analysis of Images and Patterns*, Springer Verlag, Prague, 906–911, 1995.
31. R. Turcajová, J. Kautsky, 'Shift Products and Factorizations of Wavelet Matrices', *Numer. Algorithms*, **8**, 27–54, 1994.

Spectral Databases, Infrared

Barbara J. Dębska and Barbara Guzowska-Świder
Rzeszów University of Technology, Rzeszów, Poland

1 Introduction **2**
2 History: More Important Infrared Databases **2**
2.1 Printed Collections of Infrared Spectra 2
2.2 Computer Databases of Infrared Spectra 3
3 Standards in Infrared Databases **6**
4 Structure of Infrared Databases **7**
4.1 Information Stored in Spectral Databases 7
4.2 Spectrum Representation and Spectral Database Organization 8
5 Recognition of Chemical Compound Structure **12**
5.1 Library Search Algorithm 13
5.2 Discovering Spectrum–Structure Correlations from Infrared Databases 14
5.3 Neural Networks as Tools for Structure Recognition 16
6 Multimethod Spectral Databases **18**
Acknowledgments **21**
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Infrared (IR) spectroscopic analysis is one of the most important means of structure determination of organic compounds because it can provide much information about the molecular structures of compounds. Innumerable IR spectra have been measured since the Coblentz collection was published in 1905 and many catalogs (printed collections and, further, computer IR databases) have been developed. These experimental data were very useful in the structure elucidation process, and many studies have been conducted on the relationships between various organic functional groups and their absorption bands. Usually, chemists make structural analyses using two methods. One method is to search a library of standard spectra and to find the closest match to the unknown spectrum [library search (LS) Method]. The other method is to find characteristic spectral features (connected with some parts of a chemical molecule), which is based on the empirical examination of a large number of spectra of known compounds. The conventional process of structure elucidation can be time-consuming, especially in the case of complex, multifunctional compounds. Computers offer the promise of enhanced human productivity in this field. A number of systems for searching a collection of spectral data in order to find reference spectra, identical with or similar to a spectrum of an unknown compound, have been developed. The systems may also retrieve other information, such as molecular formulas, chemical names, Chemical Abstracts Service (CAS) registry numbers, molecular fragments or complete structures of compounds. Most of the search systems are designed for a single type of spectra (e.g. IR spectra), but there are also some multimethod systems that apply various spectral techniques [IR, nuclear magnetic resonance (NMR), mass spectrometry (MS), Raman or ultraviolet/visible (UV/VIS)] to improve the results of a structure recognition process. The weakness of the LS approach is related to the necessity for a powerful computing system and a large database. Also, even a large database may not contain identical or highly similar reference spectra with respect to the sample spectrum. Thus the LS method can solve the problem of structure identification when the database contains a spectrum identical with (or highly similar to) the sample spectrum. Otherwise, to elucidate the structure of an unknown compound, it is necessary to use computer methods that permit the recognition of structural fragments present in the molecule of the analyzed substance. These programs use correlation tables (a list of structural fragments and their spectral characteristic parameters) that can form a knowledge database or be an integral part of the program. Usually, these parameters are taken from the literature, but they can also be generated from a computer library of IR spectra by computer-simulated neural networks, application of a statistical algorithm or other methods. There are systems for structure recognition which test a large number of spectral and structural features in order to calculate decision functions between the classes of compounds. When applied to a spectrum of an unknown structure, these functions indicate the presence or absence of the respective molecular fragments. Methods of this type include various cluster analyses, pattern recognition methods and computer-simulated neural networks.

The first collections of IR spectra (printed and computer IR databases) were often of poor quality, owing to the low technological level of the spectrometers used. Contemporary IR databases contain spectra of a considerably higher quality owing to the advent of high-resolution spectrometers and proper sample preparation (according to the standards required). Apart from large spectral

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

databases (SDBs) containing data on different chemical compounds, a large number of collections dedicated to specialized groups of compounds are offered. Recently, a new type of printed spectral catalogue has appeared on the market: the traditional form of a book is supplemented by a diskette containing peak tables in a digital format, allowing the reader to search for unknowns. As far as the computer databases are concerned, many efforts are aimed at the unification of spectral and chemical structure codes, which will assist an easy exchange of information between various scientific centers. The International Union of Pure and Applied Chemistry (IUPAC) has additionally published a very positive opinion about the Joint Committee on Atomic and Molecular Physical Data (JCAMP) format which makes it very probable that this format will be broadly used in future.

All the above-mentioned issues are discussed in this article, with detailed focus on (1) history of printed and computer IR collections, (2) more important IR databases, (3) standards in IR databases, (4) structure of IR databases, (5) monomethod identification of chemical compounds and (6) application of multimethod SDBs to structure elucidation.

1 INTRODUCTION

Recognition of a chemical structure, realized nowadays by spectroscopic methods, is very important in many research fields, e.g. in chemical analysis, organic synthesis (identification of main reaction products, determination of by-product chemical structures), pharmacology (recognition of the structure of pharmacologically active compounds, investigation of drugs and pesticides metabolism) and others. Irrespective of the applied spectral methods, the basic step in structure elucidation is the interpretation of the spectrum. This process is intellectually difficult, demands a professional education, experience and very often also intuition. For the interpretation of spectra, the chemist uses mainly semiempirical methods which are based on large amounts of previously accumulated reference data (spectra libraries, tables of spectrum–structure correlations). As searching through large collections of data and selecting and applying various empirical rules by conventional means is generally very time-consuming and often boring, many approaches to the computer-aided (semiempirical) interpretation of spectral data have been proposed. Three basically different techniques may be distinguished.

In the most commonly used method, called the LS method, the computer scans a reference spectra collection, comparing the spectrum of an unknown compound with each spectrum of a reference set. The structures of those reference compounds having spectral data most similar to the unknown are considered to be similar to the unknown (analyzed) structure. In the second method, the computer tests (learning phase) a large number of spectral and structural features in order to calculate a decision function between the classes of compounds. When applied to a spectrum of an unknown structure, this function indicates the presence or absence of the respective structural features (substructures, molecular fragments). Methods of this type include various cluster-analysis and pattern-recognition methods. Computer-simulated neural networks can also be acknowledged as a method of this type, although the classification function in the networks is unknown.

The third method is usually employed when the LS fails, i.e. there is a lack of spectra similar to the analyzed spectrum in the spectral library. In this method, spectrum–structure correlations generalized and condensed into a set of interpretation rules are applied to the interpretation of an unknown compound spectrum in order to predict whether the respective substructures are present or absent in the unknown molecule.

The above remarks refer to all kinds of spectra, especially to IR spectrometry. Computer IR SDBs were created before other computer SDBs. The experience gathered during the development of IR databases helped in creating SDBs for other spectral techniques. IR spectrometry was earlier the basic method for chemical structure identification, but later it became only a supporting method for NMR and MS. Nevertheless, most data banks contain IR spectral libraries.

2 HISTORY: MORE IMPORTANT INFRARED DATABASES

2.1 Printed Collections of Infrared Spectra

Before computer IR databases were developed, collections of IR spectra of pure chemical substances had been created in the form of printed catalogs. The first collection of IR spectra was published by Abney and Festing in 1881.[1] This collection contained only 48 spectra. Then, in 1905–1908, Coblentz[2] published his collection of IR spectra containing 135 organic compounds. The development of a quantum theory, wave mechanics and also the possibility of measuring substances at various temperatures and pressures and in different states of concentration indicated the great importance of IR spectroscopy in the investigation of molecular structures. Consequently, the number of IR spectra registered by different laboratories increased dramatically,[3–9] so that various catalogs soon contained even several thousands of spectra. Until the early 1950s, IR collections were printed as books, unfixed sheets (card indexes), slides, microfiches

or microfilms. As the collections of IR spectra increased in content, searching for information became more and more laborious and arduous. Later, a noticeable improvement was obtained by application of cards perforated for spectra presentation and sorting machines to review so-prepared card indexes. The first collections of this kind were elaborated by the American Wyandotte Chemical Corporation in cooperation with the American Society for Testing and Materials (ASTM),[10,11] and the catalog created (comprising 150 000 IR absorption spectra) was constructed using standard Hollerith cards. Various spectra encoded on the cards contained information about band locations, chemical classification of the compound, its chemical formula and a classification number in the original (source) catalog. One year later, the Institute of Applied Spectroscopy in Dortmund and the Institute of Molecular Spectroscopy in London created and disseminated a collection of IR spectra known as the Documentation of Molecular Spectroscopy (DMS) catalog.[3] The system of checking DMS cards was to give the answer to the question, ''Is a spectrum of an unknown compound present in the card index or not?'', and was based on the possibility of defining some classification features, i.e. spectral parameters (the location and intensity of absorption bands at specified intervals), structural parameters (the presence of characteristic structural fragments, e.g. ring systems, substituents, functional groups) and general parameters (e.g. qualitative composition, molecular weight, melting point, boiling point). The classification features were perforated on the card borders (margins) of each collection element (for each chemical compound). As the system was based on punched cards, the searching process could be speeded up by application of electromechanical shakers. The DMS system had a symptomatic feature which was later copied in computer searching systems, namely that it was possible to join different parameters in a logical conjunction within one operation of card-index searching, i.e. to input into a searching matrix of a shaker a set of holes, corresponding to some selected features, to obtain transparency through the cards (i.e. the holes fit together). An example task is to search for ''all cards containing an absorption band located in the 1800–1700 cm^{-1} range, iodide atom and molecular weight below 420''. The system also allowed negative searching, i.e. taking into consideration the lack of a given feature.

The automation of searching through a card index considerably improved the use of catalogs. A user could obtain a reduced set of spectra according to declared earlier parameters, and subsequently look through the remaining spectra manually. In spite of the inconveniences of searching through a large set of spectra, printed catalogs are still acknowledged by chemists, and today handbooks containing IR spectra of chemical substances are still published. For example, books containing IR spectra are as follows: *Raman/IR Atlas of Organic Compounds*[12] (~1000 IR spectra), *The Aldrich Library of IR Spectra*[13] (~12 000 IR spectra), *The Aldrich Library of FTIR (Fourier transform infrared) Spectra*[14] (~18 000 IR spectra), *The Aldrich Library of FTIR Spectra: Vapor Phase*[15] (~6550 IR spectra), *The Aldrich Library of FTIR/Raman Spectra*[16] (~14 000 IR spectra), *Coblentz Society Reference Spectra Collection*[17] (Volume I, 860 IR spectra); some contain only specified types of compounds, e.g. *Coblentz Society Reference Spectra Collection*,[17] Volume II, 282 IR spectra of halogenated hydrocarbons; Volume III, 284 IR spectra of plasticizers and other additives, *Atlas of Polymer and Plastics Analysis*[18] (~6700 IR spectra), *Handbook of Fourier Transform Raman and IR Spectra of Polymers*[19] (500 IR spectra) and *An IR Spectroscopy Atlas for the Coatings Industry*[20] (740 IR spectra).

Some IR collections are published in the form of a WebBook (http://webbook.nist.gov/chemistry), e.g. National Institute of Standards and Technology (NIST) Chemistry WebBook (5000 gas-phase IR spectra). Constantly updated information about books on IR spectra is available on an Internet website, e.g. http://wally.rit.edu/pubs/guides/spectra.html.

2.2 Computer Databases of Infrared Spectra

Developments in hardware technology afforded further progress in the field of SDB creation and the elaboration of algorithms used for searching and interpretation of stored data. The first databases containing IR spectra were coded on magnetic tapes or disks. The type of stored information was changing. It was possible to store not only data about absorption band parameters of the spectrum (see Table 1), but also to record the whole spectrum in a discrete form, which permitted the exact reproduction of a spectral curve (see Table 2). Some more important computer IR databases, developed recently, are presented in Table 3. From the data in these tables, it can easily be concluded that some scientific centers elaborated not only one, but several computer IR databases. The reasons why various IR databases were developed in the same research center were, among others, the application of different sources of information to develop a spectral collection (e.g. SPECTRUM-I was created on the basis of a DMS card index, whereas SPECTRUM-II is based on the Sadtler catalog, Table 1), the use of various algorithms for spectral coding (e.g. ZAPAH and KISIK in Table 1, and three of the Sadtler databases cited in Table 2) and also the separation of certain spectra collections (e.g. a set of polymers in the POLYMER database, see Table 2).

Nowadays, there are two trends of computer database development, i.e. creating larger and larger collections

Table 1 List of some important computer IR databases containing spectra stored as selected parameters of absorption bands

No.	Name of database	Institute which developed the database	Number of IR spectra
1	FIRST-1[21]	University Computing Company, Dallas, TX, USA	143 000
2	IRIS[22]	University Computing Company, Dallas, TX, USA	110 000
3	MIRET[23]	Pennsylvania State University, University Park, PA, USA	92 000
4	IRGO[24]	Chemical Laboratories, West Kirkham, Glendale, MO, USA	150 000
5	SPIR[25]	Canadian Institute for Scientific and Technical Information, Ottawa, Canada	143 000
6	ZAPAH[26,27]	Boris Kidrič Chemical Institute, Ljubljana, Slovenia	40 000
7	KISIK[28]	Boris Kidrič Chemical Institute, Ljubljana, Slovenia	1000
8	IRSPAN[29]	Japan Information Processing Service, Toyohashi, Japan	110 000
9	IRSEARCH[30]	Rzeszów University of Technology, Rzeszów, Poland	1500
10	SPECTRUM-I[31]	Academy of Sciences of the USSR, Siberian Division, Novosibirsk, USSR	18 000
11	SPECTRUM-II[32]	Academy of Sciences of the USSR, Siberian Division, Novosibirsk, USSR	21 000
12	SPEKSU[33]	Technical University, Dresden, Germany	3800
13	IRNET[34]	Rzeszów University of Technology, Rzeszów, Poland	1500
14	IRDC[35]	National Chemical Laboratory for Industry, Tokyo, Japan	19 200

Table 2 List of some of the first computer IR databases with digitized spectral curves

No.	Name of database	Institute which developed the database	Number of IR spectra
1	IR base[36]	Academy of Science of the USSR, Siberian Division, Novosibirsk, USSR	1000
2	COSMOSS[28]	Boris Kidrič Chemical Institute, Ljubljana, Slovenia	102 000
3	POLYMER[37]	Boris Kidrič Chemical Institute, Ljubljana, Slovenia	740
4	Sadtler Digital-4IR[38]	Sadtler Research Laboratories, Philadelphia, PA, USA	100 000
5	Sadtler VARIMAT 8*16[38]	Sadtler Research Laboratories, Philadelphia, PA, USA	100 000
6	Sadtler De Res 32IR[38]	Sadtler Research Laboratories, Philadelphia, PA, USA	100 000

containing spectra of various substances or specialized collections of chemical compounds of a specified type or purpose (polymers and copolymers, plant protection remedies, medicaments, etc.). The application of computer software enables one to generate from a large database some specialized collections, even for individual users (by network mechanisms). This solution was adapted to a database disseminated by Hoechst AG[41] to a network encompassing 20 different laboratories, using FTIR spectrometers from Bruker, Perkin-Elmer, Nicolet and Digilab. The users of this network had access to the central data bank containing selected fragments of Sadtler, Bruker and Hummel collections (6300 standard Sadtler spectra, 9300 Sadtler spectra in the vapor phase, 3200 spectra of Bruker–Merck firm and 2900 spectra from the Hummel–Scholl atlas).

Additionally, access to a small local Hoechst database was included (550 standard spectra, 500 spectra of substances supplemental for polymers, and 300 spectra in the vapor phase). The libraries of the central data bank were accessible only in a read-only procedure, whereas the resources of the second database were continually supplemented (about 25 spectra every week) by analytical laboratories connected to the network (Analytical Lab, Frankfurt; R&D Pigments, Frankfurt; Analytical Lab Kalle, Wiesbaden; Analytical Lab Hoechst, Gersthofen; and others), which produced high-quality spectra.

The analysis of the organization methods of IR databases permits their evaluation. For example, taking into consideration the ASTM catalog created over 30 years (1950–1980), containing at the time about 90 000 standards,[9] it can be stated that limitations imposed on the database at the time of its development (on perforated cards, with application of Hollerith cards with 80 columns) allowed archiving of only some selected bands (coded in micrometers), as opposed to collections containing full discrete spectra, which are nowadays coded in wavenumbers. An additional artifact of the ASTM catalog is the lack of the evaluation of spectrum quality and unclear terminology, e.g. using also some customary chemical names inconsistent with semantics, which led to a certain number of repeated spectra in the database (i.e. spectra of the same substance). The inconsistent terminology also excluded the possibility of cooperation between this database and other collections [e.g. carbon-13 nuclear magnetic resonance (^{13}C-NMR), proton magnetic resonance (^{1}H-NMR), MS, ultraviolet (UV)] which were created simultaneously with IR databases, and they stayed as monomethod manual

Table 3 List of some available computer IR database with digitized spectral curves

No.	Name of database	Number of spectra
1	Coblentz Collection of General Chemicals[39]	1111
2	Sadtler Condensed Phase IR Standards[a]	75 620
3	Sadtler Starter Database[a]	11 800
4	Merck/Sadtler Select FTIR Database[a]	2940
5	Vapor Phase IR Standards[a]	9200
6	Controlled Pyrolyzates of Polymers[b]	2980
7	Electrical Power Plant Materials[b]	1075
8	Hummel/Sadtler Polymers[b]	1920
9	Monomers and Polymers[b]	11 300
10	Surfactants[b]	9240
11	Canadian Forensic Database[b]	3490
12	Dyes, Pigments and Stains[b]	2390
13	Sadtler Environmental Protection Agency (EPA) Vapor Phase[b]	3240
14	Pesticides and Agricultural Chemicals[b]	720
15	Inorganics[b]	1000
16	Organometallics[b]	350
17	Nicolet/Aldrich Condensed Phase Library (FTIR)[40]	10 067
18	Nicolet/Aldrich Condensed Phase Supplement Library (FTIR)[40]	5075
19	Aldrich Vapor Phase FTIR Library[40]	5000
20	Nicolet/Sigma Biochemical Condensed Phase Library[40]	10 411
21	Nicolet Sigma Steroids Condensed Phase Library[40]	3011
22	Nicolet/Aldrich Flavors and Fragrances Vapor Phase Library[40]	677
23	Nicolet Expanded Hummel Polymer Condensed Phase Library[40]	2011
24	Nicolet Polymers, Polymer Additives and Plasticizers Condensed Phase Library[40]	1750
25	Nicolet Food Additives Condensed Phase Library[40]	519

[a] http://www.bio-rad.com/629418.html.
[b] http://www.bio-rad.com/620859.html.

catalogs and related to computer libraries of molecular spectra.

Currently the world's largest database of IR reference spectra of organic compounds is the Sadtler Condensed Phase IR Standards Database. This database contains data on 75 620 different chemical compounds (http://www.bio-rad.com/626559.html). The IR spectra stored in the Sadtler collection are of very high quality. Each compound, before its spectrum is placed in the Sadtler collections, is prepared under standard conditions at Bio-Rad Laboratories, Sadtler Division, using equipment manufactured by leading instrument makers. Samples are checked to ensure that nonlinearity is avoided to maximize reproducibility and provide for the best subtraction of reference spectra. Wherever possible, Sadtler chemists verify every IR spectrum by checking the corresponding NMR and UV spectra. Any spectra which show evidence of composition impurities or reaction with sampling apparatus are rejected. Each spectrum is then reviewed by an independent consultant before inclusion in the database. These quality control procedures provide the best sample purity and structural formula confirmation available for the broad range of compounds in the collection. This database of spectra affords access to the spectra of compounds having desired functional groups and which can be used to establish an identity through empirical comparison. The Sadtler database includes spectra of most simple aliphatic, aromatic, alicyclic and heterocyclic compounds, and also numerous complex materials. Numerous series of homologous compounds, ranging from the very simple to the very complex, which allow spectroscopic studies of trends involving homologs, are also included. The Standards Database is also very useful for the establishment of complex chemical classes of undocumented commercial compounds with similar structures. The molecular structure symbol signifies the molecular structures associated with the database. The structures are available for viewing and substructure searching by special search software.

Apart from the Sadtler database mentioned above, other IR databases for pure organic compounds are also offered (see Table 3). Additionally, a number of dedicated computer collections of IR spectra have been developed, e.g. IR databases for polymers and related compounds, industrial compounds, forensic science, environmental applications and inorganic and organometallic compounds (see Table 3, http://www.bio-rad.com/620859.html).

Apart from the above-mentioned computer IR libraries, there are also IR databases provided by producers of

Table 4 Multimethod spectral computer databases containing IR spectra (a few examples)

Name of database/country	Number of spectra in spectral library					
	IR	MS	^{1}H-NMR	^{13}C-NMR	UV/VIS	Raman
CSS,[35] China	519	489	–	496	–	457
CISOC,[35] China	45 000	73 000	–	2000	–	–
SDBS,[35] Japan	30 000	9000	5800	5500	–	1500
NCLI,[35] Japan	22 500	10 000	6000	5700	–	–
CIS,[45] USA	3000[a]	238 000	–	11 700	–	–
SCANNET,[46] Poland	900	3000	700	500	300	200
Sadtler PC,[47] USA	150 000	–	Several thousand	28 000	–	–
SpecInfo[48] Germany:						
SpecInfo on-line[b]	>18 000	60 000	100 000 NMR spectra			
SpecInfo 3.2[c]	20 000		105 000 ^{13}C-NMR and heteroatom NMR spectra, 2500 ^{1}H-NMR spectra			

PC, personal computer.
[a] IR spectra from IR Search System (Boris Kidrič Institute, Lubljana, Slovenia).
[b] SpecInfo on-line (www.wiley-vch.de/cc/si_onl.html).
[c] SpecInfo 3.2 (www.dl.ac.uk:800/CDS/spectro.html).

IR spectrometers, e.g. a digitized version of *The Aldrich Library of FTIR Spectra*[42] (over 10 600 spectra), and *The Sigma Library of FTIR Spectra*[43] (over 10 400 spectra of biochemicals) as available exclusively from Nicolet Instruments GmbH, or *Bruker FTIR Libraries* (over 13 400 FTIR spectra) offered by Bruker Analytische Messtechnik GmbH.[44] The next step in the development of SDBs was to create multispectral databases. Most of them include libraries of IR spectra (see Table 4). The experience gathered during IR bases development helped in creating spectral bases for other spectral techniques. The content of a given multimethod SDB is usually presented as a list of individual catalogs (IR, MS, UV, etc.) and numbers of spectra related to them. The information about the numeric size of the collections in Table 4 is approximate, and it can differ from current numbers, as the databases are being continuously supplemented and transmitted between various scientific centers which include the new databases in their own collections.

3 STANDARDS IN INFRARED DATABASES

A very important matter for producers of IR databases (printed and computer collections) is to ensure high quality of the IR spectra. It is also important to define one unified data format for computer collections in order to assure accurate transportation of scientific data between different computer systems and software packages. Accurate transportation of data is an essential element of modern integrated laboratory work.[49]

The problem in the quality control of IR SDBs is connected with four main areas:[50] sample preparation, data handling, spectrometer performance and data presentation. With regard to the spectra in an IR collection, attention should be paid to the manner in which a given compound is prepared for an examination. The sample preparation procedures should result in obtaining a reproducible spectrum. The data handling of an IR spectrum downloaded by a computer needs to be controlled by an experienced person in order to avoid any loss of spectral quality and the introduction of any artifacts into the spectrum. Spectrometers used for the registration of reference spectra should be calibrated by standard samples. These three above-mentioned areas determine the quality of reference spectra. All guidelines for the collection of reference IR spectra are formulated by spectroscopic centers. For example, the Coblentz Specifications for IR Reference Spectra of Materials in the Vapor Phase above Ambient Temperature defines the best realistically obtainable spectra as ''reference spectra obtained on high-quality commercial instruments operated at maximum efficiency under conditions consistent with acceptable laboratory practice''. Their precise criteria for sample preparation, spectrometer working parameters and spectrum hard copy (e.g. range of spectrum registration, resolution, noise, baseline, recording continuous with no gaps in wavenumber, purged or evacuated spectrometer, linear abscissa in wavenumber, linear ordinate in absorbance, accuracy with sharp bands readable to 3 cm^{-1}).

In addition to sample and spectra registration conditions, the guidelines for collections of spectra elaborated by expert committees for the submission of spectra to the Spectrendatenbanken-Verbundsystem (Germany)[50] also accept the Joint Committee on Atomic and Molecular

Physical Data Exchange (JCAMP-DX) transfer format of spectral data presentation. McDonald and Wilks are the authors of the first JCAMP-DX standard for IR spectroscopy.[51] The JCAMP-DX is a standard format for IR, UV/VIS, NMR and mass spectra, whereas the Joint Committee on Atomic and Molecular Physical Chemical Structure (JCAMP-CS) format has been proposed for chemical structure coding.[52] The JCAMP-CS format provides definition for exchanging information on compounds and sterochemistry, and also on the two-dimensional (2D) and three-dimensional (3D) atom coordinates of chemical structures. Where possible, care was taken to make the conversion to other formats in use for representing chemical structures as simple as possible.

The JCAMP-DX format is an answer for those chemists who need a standard file format for the exchange of IR and other kinds of spectra between numerous vendor data systems which nowadays use various proprietary file formats. Unification will enable spectra to be transferred between different spectrometers from their own and other laboratories. The usefulness of the JCAMP data exchange formats has been recognized and accepted by such organizations as the International Union of Pure and Applied Chemistry Committee on Printed and Electronic Publications (IUPAC-CPEP) and ASTM (http://lolita.colorado.edu/faq). This leading standard is also used in SpecInfo SDBs. The implementation of the JCAMP standard as a transport format for IR spectra has been a great step forward in the direction of interlaboratory communication. However, the international organization of structural and spectral data storage and archiving has been a serious problem. Nowadays, software utility programs permit the conversion of different formats to the one which would be accepted by a given system. For example, the following formats (in addition to the JCAMP) are currently used in each of the software listed:

- Sadtler software: Galactic SpectraCalc/GRAMS, Digilab 3200/WIN-IR, Nicolet SX/DX/PCIR/OMNIC, Perkin-Elmer IRDM/1600/7000 and others (http://www.bio-rad.com/610483.html, http://www.sadtlersuite.com/spec.html);
- ACD/IR Manager: GRAMS, ASCII (http://www.acdlabs.co.uk/products/ir/ir_mgrhtml).

The open and flexible nature of the JCAMP standard and its easy electronic transmission not only on a PC but also in the Scientific and Technical Information Network (STN) International Network are the reasons why this standard will most probably become the most commonly used standard also by SDBs producers and instrument manufacturers.

According to various literature sources, parallel development of printed and computer IR databases can be observed. Owing to the low technological level of spectrometers, the first collections of IR spectra were often of poor quality. Contemporary IR SDBs contain spectra of a considerably higher quality owing to the advent of high-resolution spectrometers and proper sample preparation (according to the standards required).

In addition to large SDBs containing data on different chemical compounds, many collections dedicated to specialized groups of compounds are offered.

Recently, a new type of printed catalog of spectra has appeared on the market. The traditional form of a book is supplemented by a diskette containing peak tables in a digital format, allowing the reader to search for unknowns[19] (http://www.elsevier.nl/locate/isbn/0444826203).

As far as computer databases are concerned, many efforts are aimed at the unification of spectral and chemical structure codes, which will assist in the easy exchange of information between various scientific centers. IUPAC has additionally published a very positive opinion about the JCAMP format which makes it very probable that this format will be broadly used in the future.

4 STRUCTURE OF INFRARED DATABASES

In this section, some examples of the computer IR database structure is discussed. Since the late 1960s, various computer programs to automate the collection and processing of IR spectra have resulted in different forms of digital representation of spectral data, from the first binary code form[53] to the latest fully discrete form, e.g. OPUS software system (Bruker),[54] SpecInfo 3.2–Input Editor (http://cds3.dl.ac.uk/specinfo/ug_input.html) and Sadtler–IR SearchMaster (http://www.sadtlersuite.com/spec.html).

4.1 Information Stored in Spectral Databases

IR databases usually consist of the following data:

- *Compound identification database (CDB)*. Each record of the file contains the structural and alphanumeric information about a compound, such as systematic and trivial name, chemical formula, molecular weight, CAS registry number, bibliographic data and selected physicochemical parameters (density, refractive index, dipole moment, melting and boiling point, dielectric constant, etc.). Each of these data and their combinations can be used in searching a certain database in order to find all the information about a compound. Some of the input data (e.g. CAS registry

number, molecular structure) enable one to select the compound unequivocally; others (e.g. molecular formula, molecular weight) provide a list of substances fulfilling the introduced search parameters.

- *SDB*. In this database, separate information about the conditions of a spectrum preparation (instrument, temperature, sample preparation, solvent, frequency region, concentration, cell thickness, optical materials of cell, purity, vertical axis, horizontal axis, etc.) is stored in addition to details about the spectrum itself. In principle, each record contains spectral band parameters, a spectrum in a discrete form and usually also a quality index (QI). The value of QI[55] depends on compound purity, the method of sample preparation, the quality of the spectrometer used, etc.

The above data may be grouped in two separate files (CDB and SDB), but usually the structure of a database is much more complex, i.e. the database contains several files that are related to each other by cross-references.

The information about a structure may be input into a database either in an alphanumerical form (such as list of atoms and bonds and connectivity tables) or in a graphic form. An example of an alphanumerical form is the Wiswesser Line Notation (WLN) used in some Sadtler databases (http://www.bio-rad.com/629418.html) or the JCAMP-CS format accepted for chemical structure coding in the SpecInfo system (http://lolita.colorado.edu/faq).

The graphic program for a structure presentation offers easy molecule construction and provides a full range of structure input methods to be used by a chemist. Chemists can easily draw any structure using a mouse and the program's interactive sketch function (see Figure 1). Usually when drawing carbon, hydrogen atoms are automatically labeled. The introduced structure is coded

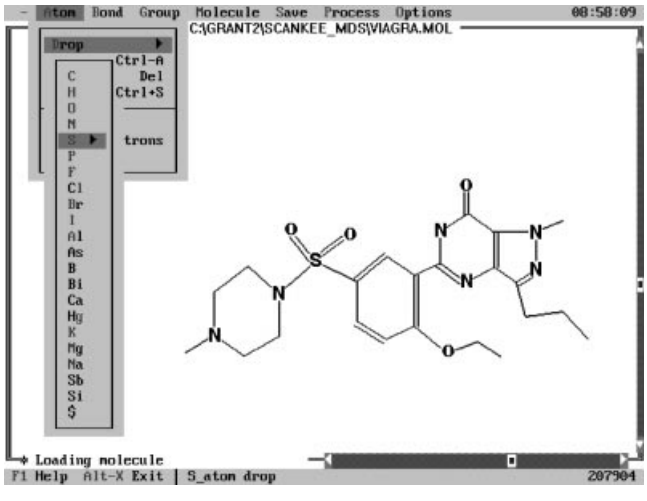

Figure 1 Example of structure edition in computer system.

and changed to a canonical form. This step is necessary because several different pictures of the same structure exist; the canonical form assigns them all the same code, in which the structure is stored in an SDB. Structural information is the most important of the CDB data as it starts the creation of a record for a given compound. Other CDB information can be input optionally.

4.2 Spectrum Representation and Spectral Database Organization

Many systems for archiving collections of spectra on a hard disk (documentary function) have been developed. They give access to information about the structure, spectra and physicochemical parameters of organic compounds, stored in an SDB (information function).

The basis for many computer collections was the ASTM catalog,[11] in which every IR spectrum is presented in a binary (0,1) code generated in the following way: the whole range of a spectrum registration was divided into 0.1 μm intervals; if there was an absorption band in an interval, the interval was given the value 1 and the lack of a band was coded with 0. The virtues of the binary coding form are that it can be stored in a small section of a computer memory (each figure of the code is stored in one bit of memory), and it gives the possibility of using fast algorithms of spectral comparison. The disadvantage of this method is the lack of information about real absorption band parameters (location, intensity and shape).

The development of various computer systems initiated remarkable progress in the creation of molecular spectral computer databases. It was possible to store more information about a spectrum in the databases, which led to changes in the type of data to be stored. The following types of data would be stored:

1. data on the number and location of bands in each spectral coding interval;
2. data on the location of all the bands in a spectrum;
3. data on the location and intensity of chosen absorption bands;
4. data on the location, intensity and factors defining the shape of bands in a spectrum.

None of the mentioned spectrum coding methods would permit the reproduction of a graphic picture of a spectrum. Completion of the already existing spectral data with factors defining shapes of spectra enables one to calculate their curves and finally to display a given spectrum on a monitor screen. Hence the user of the system can compare the analyte spectrum and a spectrum found in the database. The shape of an absorption band can be reproduced by means of a factor called the half bandwidth and an approximating function (in the most general

case it is a sum/product of Gaussian and Lorentzian functions[56]). Other additional parameters can be directional factors of lines approximating absorption band shapes.[57]

Nowadays, the most commonly used method of spectral coding is via the intensity values of spectral curve points with a constant step of discretization. This method demands a substantial memory for the database. A discrete spectrum code is a set of intensity values coded for a constant step of changes for the horizontal axis. In the database, only the information about the values of the intensity of points will be stored.

The possibility of coding whole spectra in a discrete form in a database appeared first in the late 1970s, together with disseminating magnetic carriers of a large capacity and with the possibility of fast access to the stored data (magnetic disks). Additionally, providing spectrophotometers with computers permitted spectrum archiving in real time (on-line), simultaneously with the analysis being conducted (see Figure 2). Usually, each spectrometer is connected only with its own PC computer, but in some laboratories spectrometers are linked together in the LIMS network,[58] which facilitates the creation of a multispectral data bank.

The generation of a molecular spectral library is carried out as shown in Figure 3. Nowadays, the source data are usually transmitted to a computer directly from an analytical instrument (the on-line method). In exceptional cases, the computer databases must be updated by spectra recorded only in an analog form[59] (such spectra were registered by old types of spectrometers, or the data were taken from the literature), and then a digitizer must be used (the off-line method). The spectra stored in a database are supplemented with additional information about their discretization process. They are also transformed according to the format of given library data. All these spectral data (Figure 3) are compatible with any database management program. Most computer systems demand input data prepared in the JCAMP-DX format to be able to build any IR database. A description of the JCAMP file structure can be found on the Internet under http://members.aol.com/rmcdjcamp/xydatst.txt.

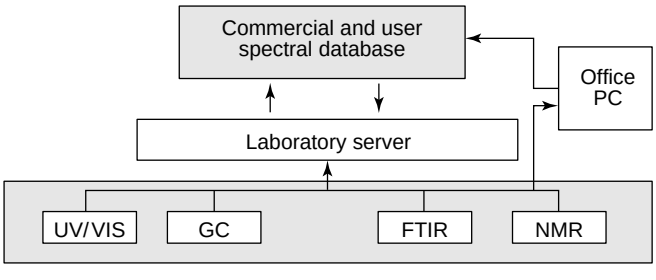

Figure 2 Scheme of Laboratory Information Management System (LIMS).

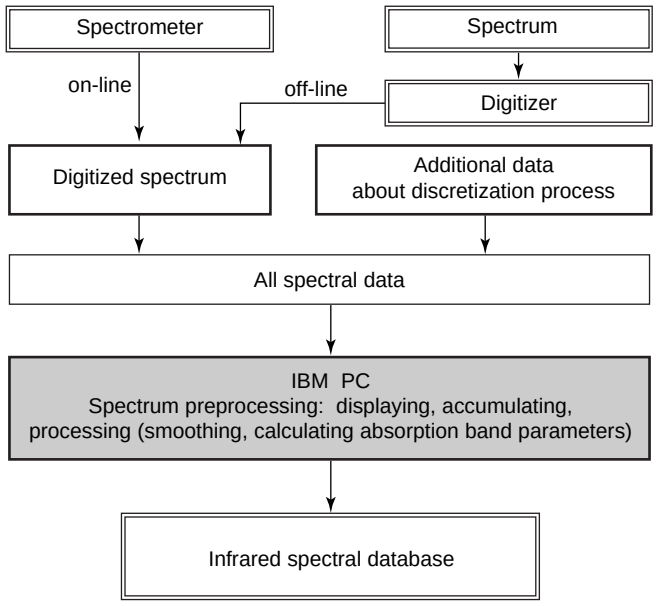

Figure 3 Numerical preprocessing scheme of IR spectrum.

An example of the input IR spectral data is presented in Table 5. The format of the data coded in each line is specifically defined. The JCAMP-DX format accepts three values for step discretization of an IR spectrum (e.g. ##DELTAX = 1.0, 0.5 or 2.0 cm^{-1}). The step of spectrum discretization is a very important parameter in spectrum representation. To choose the optimal step of discretization, the authors of databases take the purpose of the creation of the set of spectra into consideration. If a database is to be a set of references and spectra are to be processed, this step must be smaller (0.5 or 1.0 cm^{-1}); it can be bigger when a set is created only to display an image of a spectrum on the screen, or when only information on whether the database contains a given spectrum is needed. Although the discretization step is a very important parameter in evaluating the quality of spectra stored in a computer database, only few authors publish any information about the size of the step, i.e. the resolution for data exchange files. Such information can be found, for example, in the manuals on the Sadtler databases or on the SpecInfo system.

The use of a spectrum in a fully discrete form is often difficult for LS algorithms, e.g. for the structure identification of an unknown compound. Therefore, in some databases[37,38,60,61] for fast searching, apart from spectra in a discrete form, tables of absorption band parameters calculated from spectra (location, intensity, etc.) are also stored. If a discrete spectrum form is also stored in a computer mass memory, it enables one to display a real spectral curve on a monitor screen. In this case, the optimal number of discretization points depends on the resolution of the monitor (on the number of pixels).

Table 5 Example of JCAMP-DX format for IR spectra

Line format	Remarks
##Resolution = 2	Spectrum resolution
##Deltax = 1.0	Discretization step = 1 cm^{-1}
##Xunits = 1/cm	Wavenumbers
##Yunits Transmittance	Intensity units
##Xfactor = 1.0	Representation of the actual decimal value *x*
##Yfactor = 0.0001	Representation of the actual decimal value *y*
##FirstX = 400	First *x* value
##LastX = 4000	Last *x* value
##NPoints = 3601	Number of discretization points
##FirstY = 0.4480	First *y* value
##XYdata = (X++(Y..Y))	X++ indicates symbolically that each line of data starts with an abscissa value which is incremented between ordinate values. Y..Y indicates that the remainder of each line is filed with a series of ordinate values. The actual data ##XYdata = starts on the next line (400 4480 4618, ..) is a table of ordinates at equal X-intervals
400 4480 4618 4716 4788 4838 4850 4806 4680 4446 4126 3748	
411 3324 2852 2342 1827 1354 953 625 387 251 242 381	
...	

Calculating the values of absorption band parameters is one of the most difficult issues in numerical analysis owing to the complicated character of molecular spectra, which consist of many overlapping absorption bands.

In the 1970s and early 1980s, the problems of the mathematical elaboration of spectral curves were widely discussed in the literature,[56,62–67] connected with problems of structural identification. Many different solutions and detailed numerical algorithms were presented, the most important of which seems to be the algorithm created by Jones et al.[56,64,65] Their numerical algorithm (for calculating the values of absorption band parameters) considers eight basic operations of mathematical spectra elaboration, namely:

- numerical spectrum transformation;
- scale calibration;
- scale correction and transformation;
- numerical filtration;
- deconvolution;
- recognition of overlapping bands;
- localization of isolated and overlapped bands;
- calculation of real parameters of spectral absorption bands.

Commercial software [e.g. PeakFit (http://www.spss.com/software/science/peakfit), STATISTICA (http://www.statsoftinc.com), Galactic GRAMS/386,[40] OMNIC FTIR[40]] works with the above numerical operations for spectrum curve processing.

The software managing commercial SDBs provides access to information included in the databases in order to enable one to use it, but any changes to the information are impossible, as the databases are protected against any interference. Nevertheless, the user can additionally create his or her own database, e.g. by on-line methods, when the spectrometer is directly connected to the computer. Thus, the user has two SDBs available: the commercial one (only for searching) and a personal library. The user has direct access to the user-defined database(s), and can take one of the following actions: add new data, delete data, correct data and delete the whole database. Hence the spectroscopist has a tool for the documentation of his or her work and for the identification of compounds which are present in his or her personal SDB. The commercial databases are enlarged by successive upgrades on a PC or by access to a network data bank, disseminated for users.

As the IR curve depends strongly on the method of sample preparation (e.g. KBr pellet, liquid film, solution in a given solvent, vapor), the storage of several IR spectra for one compound is useful when an SDB is applied to studies concerning the influence of spectrum registration conditions (temperature, pressure, concentration, etc.) on band location and band intensity. In this case, a single structure of a compound is stored and is connected by the cross-references with all its spectra. Special procedures enable one to compare various spectra in order to observe any differences between them. Such a solution is applied in the SCANNET system, with which one can gather five different IR spectra (of one compound) obtained under different conditions. The system also enables observation of the phenomena[61] that cause changes in the spectral band shape and also the disappearance and/or appearance of the bands. The investigation of the phase transition of dichloromethyl *tert*-butyl ketone is shown in Figures 4–6.

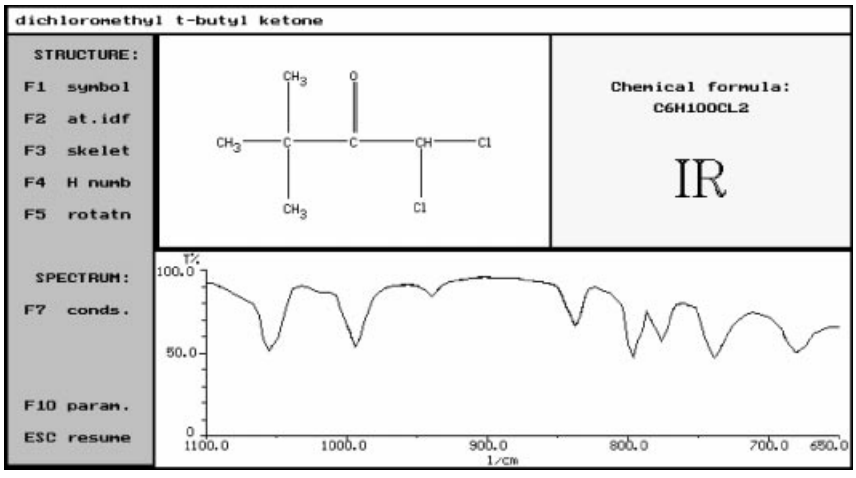

Figure 4 IR spectrum of the crystalline form of dichloromethyl *tert*-butyl ketone.

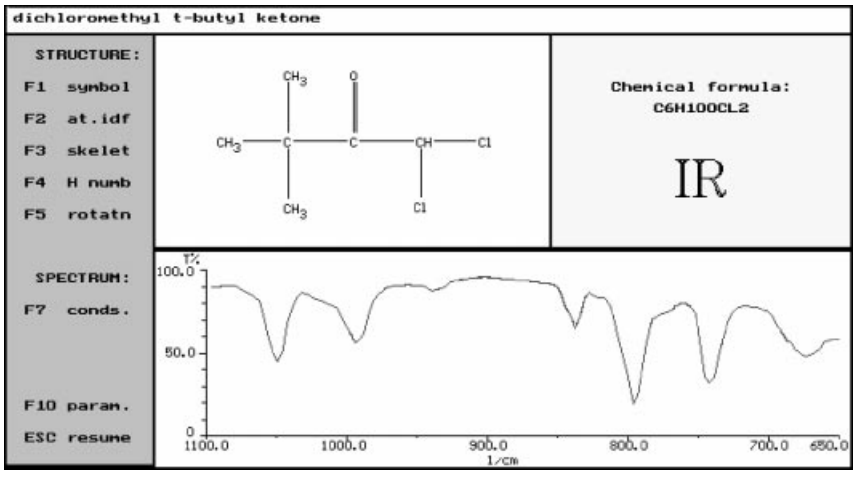

Figure 5 IR spectrum of the liquid form of dichloromethyl *tert*-butyl ketone.

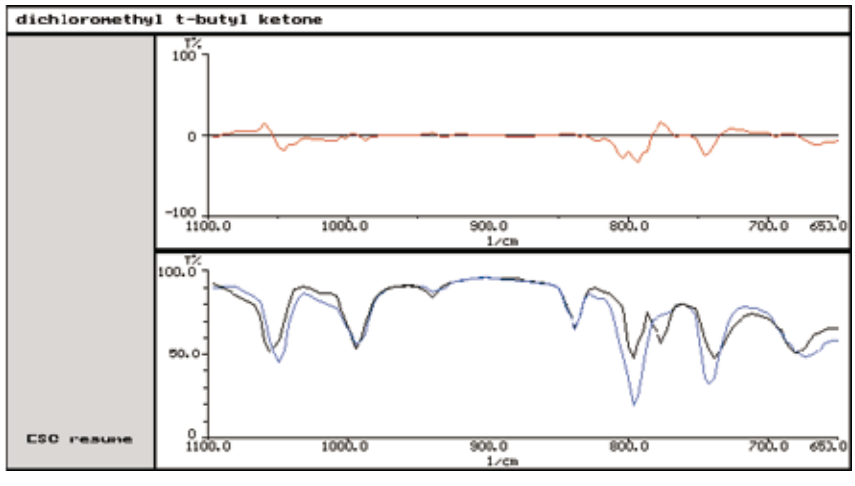

Figure 6 Overlapping and differential IR spectra of dichloromethyl *tert*-butyl ketone.

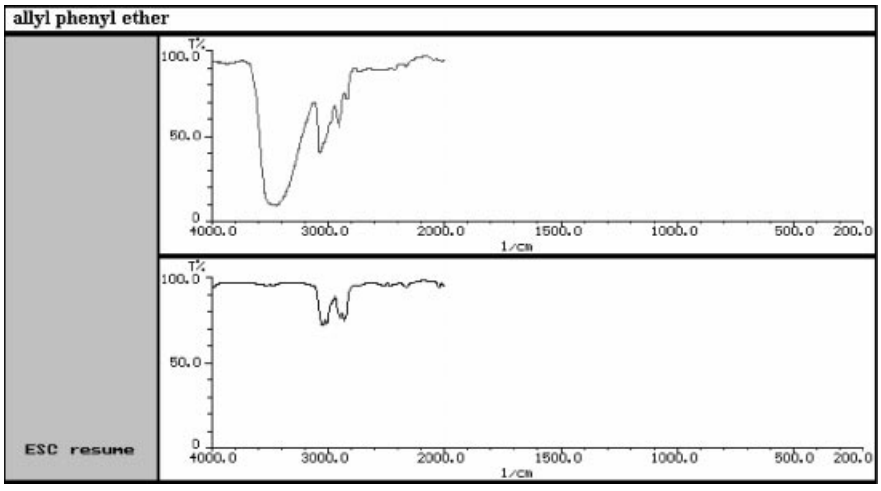

Figure 7 IR spectra of allyl phenyl ether (bottom) and *o*-allylphenol (top).

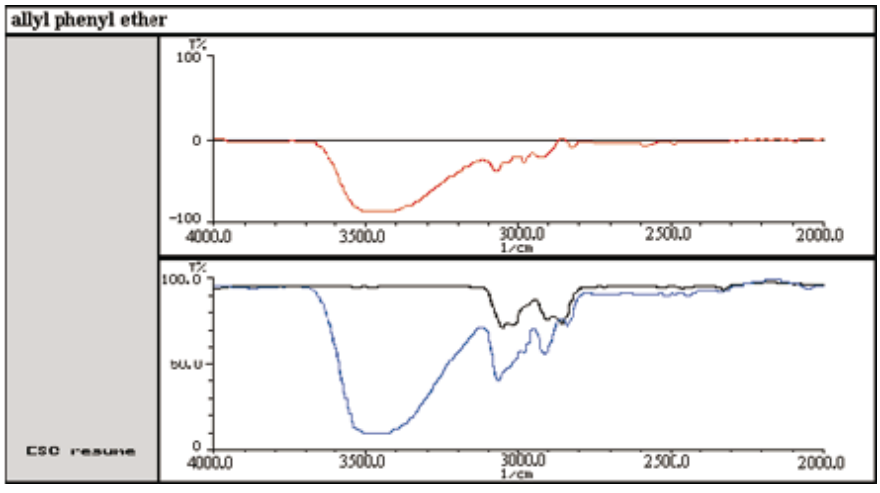

Figure 8 Overlapping and differential IR spectra of allyl phenyl ether and *o*-allylphenol.

The IR spectrum of the crystalline form of the substance is presented in Figure 4 and Figure 5 presents part of the IR spectrum of the liquid form. It can be seen from the overlapping spectra and from the differential curve (see Figure 6) that the spectrum of the substance in the liquid phase contains only one band at 797 cm^{-1} instead of two bands at 799 and 778 cm^{-1}, characteristic of the crystal phase.

The storage of different spectra for one compound can be used for studies of the Claisen rearrangement of allyl phenyl ether to *o*-allylphenol is shown in Figures 7 and 8. On these figures one can observe a strong band at 3430 cm^{-1} (due to the stretching of the O–H bond of the hydroxyl group) in the product spectrum, which is absent in the substrate spectrum.

5 RECOGNITION OF CHEMICAL COMPOUND STRUCTURE

Many attempts have been made in recent years to make the structure elucidation process faster and more reliable by using computers. All the methods presented below (LS algorithms, correlation tables method and neural networks) use IR databases.

The structure of these databases differs for each of these methods. The LS method works on large (tens of thousands) collections of spectra and enables one to find in a database a set of compounds identical with or similar to the analyte compound.[68] In the case of the other two methods, a database is used on a level of generating the identification system. When the correlation tables method is applied, the IR spectral library

(containing from several hundred to several thousand spectra) is used to generate a knowledge base consisting of spectra–structure correlations (presented as tables, interpretation rules or decision trees). The generated knowledge base, together with the inference engine, forms an expert system for structural fragment recognition. The last method uses one specially transformed IR database to learn (to train) how to recognize chosen structural fragments. Once a network has been trained, it is able to identify the fragments without using the IR spectral library.

5.1 Library Search Algorithm

The identification of a chemical compound structure based on the application of an IR library consists[68–71] in looking for the identity or similarity between the spectrum of an unknown substance and one or more spectra from the standard collection. The reasoning is based on the logic Clerc cycle[72] (see Figure 9) assuming that the identity or similarity of the examined and reference substances spectra proves the identity or similarity of their structure. Obviously, some assumptions which would enable one to realize an inference process must be fulfilled (the same method of sample preparation, the same coordinates, similar spectrometer settings, etc.). The control of the data stored in a database is particularly important, as the shape of an IR band is strongly dependent on the mentioned parameters. Therefore, the algorithm comparing the spectra should be preceded by the control of spectrum registration conditions. This algorithm LS algorithm is based on a known paradigm of artificial intelligence – ''describe-and-compare''.[73] The LS algorithm imitates the reasoning of a chemist, who elucidates the structure of a chemical substance by comparing its IR spectrum with successive spectra of known compounds stored in a database. In the case of an automatic (computer) realization of the LS algorithm, not the whole spectral curves are being compared, but only some selected spectral features, so the algorithm for calculating the measure of the identity/similarity of the features must be defined.

It is assumed that for spectral features being compared, one of the following relations exists:[69]

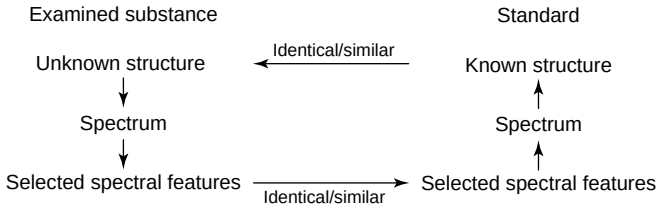

Figure 9 The logical cycle of Clerc.

selected spectral feature	do not inconsiderably considerably	changes for
identical similar different	structures	

These relations are realized according to detailed solutions accepted by a given author. Generally, the LS algorithm is realized in the following steps:[68]

1. the conversion of an unknown compound spectrum into a digital form, exactly the same as that used for reference spectra in a database;
2. the sequential comparison of the numerical representation of an unknown spectrum (of an analyzed compound) with successive spectra in a database;
3. the output of information about the results of searching the database, i.e. a list of spectra identical with or similar to the analyte spectrum, on the basis of a predefined criterion for spectral similarity (strictly, similarity of numerical representations).

The list of these spectra is usually sorted according to decreasing values of their fitting factor, which is a measure of spectral similarity. A value of the fitting factor of 1 refers to complete identity of the spectra being compared, whereas a value of 0 implies a lack of any similarity between them. A chemist is the one to take the final decision about which of the structures proposed by the program is that sought. The LS method of the identification of unknown compounds from their IR spectra has been widely applied for practical purposes. Now, every FTIR instrument has functions for database searches on its minicomputer. However, the success of the database search method depends on the size of the database. That is, in order to identify an unknown spectrum, its precise (or highly similar) reference must be included in the database. However, when the size of the database becomes larger, the search will take longer, which is especially critical in the case of systems on PCs. For this reason, with neural network application a new technique for faster searching of IR databases was elaborated.[74]

An example of a real procedure for the elucidation of an unknown compound structure aided by a LS method is shown in Table 6 and Figures 10 and 11.[75] The unknown compound was identified by means of the IR spectrum. The IR spectrum in a discrete form was the input data; the spectral bands selected for searching are presented in Table 5. After searching the IR database, a list of substances having IR spectra identical with or similar to the input spectrum was obtained. The list contains one compound with a fitting factor of 1.0 (1-pentanol) and

Table 6 Example of IR database searching

Input band parameters	List of substances identical with similar to unknown compound		
	No.	Name of compound	SI
3382–3342 (cm^{-1})	1	1-Pentanol	1.00
1.037–2.387 (Å)	2	1-Hexanol	0.94
	3	1-Butanol	0.94
	4	2-Octanol	0.82
2985–2945 (cm^{-1})	5	2-Methyl-1-pentanol	0.82
2.787–2.387 (Å)	6	1-Heptanol	0.82
2893–2853 (cm^{-1})			
1.513–1.113 (Å)			
1477–1457 (cm^{-1})			
0.536–0.336 (Å)			
1470–1450 (cm^{-1})			
0.480–0.320 (Å)			

SI, similarity index; 1 Å = 10^{-10} m.

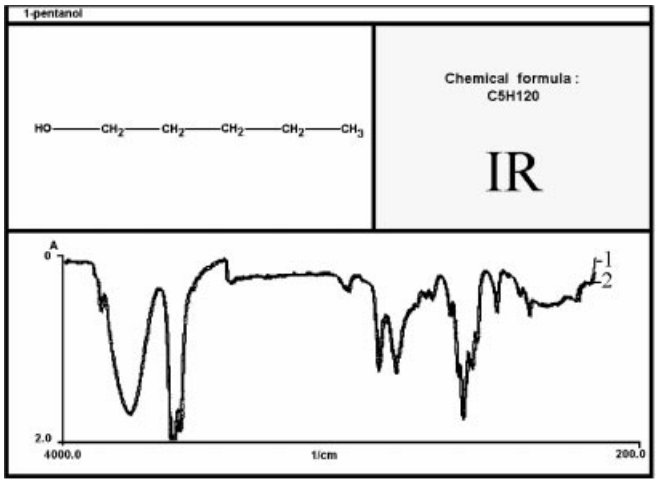

Figure 10 Overlapping IR spectra: (1) unknown spectrum; (2) 1-pentanol.

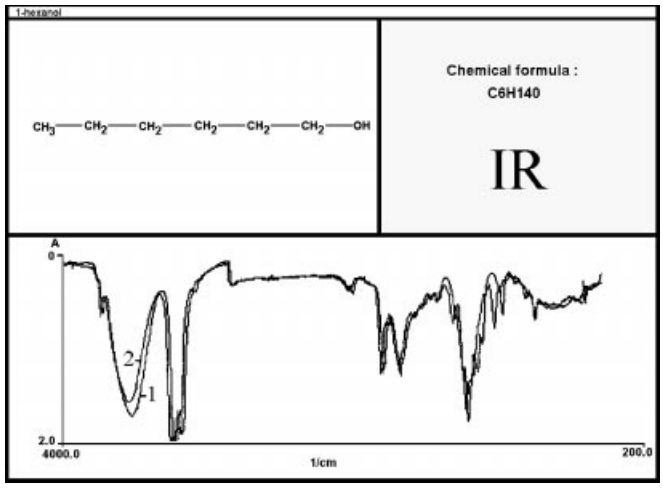

Figure 11 Overlapping IR spectra: (1) unknown spectrum; (2) 1-hexanol.

others with fitting factors <1.0. Overlapping IR spectra are presented in Figures 10 and 11. Here, the unknown compound and 1-pentanol are identical, whereas the spectrum of 1-hexanol differs slightly from the unknown spectrum (fitting factor 0.94). The analyte compound therefore has the structure of 1-pentanol.

The weakness of the LS approach is related to the necessity for a powerful computing system and a large database. Also, even a large database may not contain identical or very similar reference spectra with respect to the sample spectrum. The LS method can thus solve the problem of structure identification when the database contains a spectrum identical with (or very similar to) the sample spectrum. Otherwise, to elucidate the structure of an unknown compound it is necessary to use methods that permit the recognition of structural fragments present in the molecule of the analyte substance. The IR library can also be applied to discover spectrum–structure correlations.

5.2 Discovering Spectrum–Structure Correlations from Infrared Databases

A different approach in spectral interpretation is based on the determination of functional groups which may be present in an unknown compound. Several computer programs have been written in order to mimic the human process of interpreting an IR spectrum of an unknown compound. These programs[76,77] use correlation tables (a list of structural fragments and their spectral characteristic parameters) which can form a knowledge database or be an integral part of the program. Usually, these parameters are taken from literature,[73–88] but they

can also be generated from a computer library of IR spectra.[77,86,89,90] The first approach of using information from the literature was applied in the X-PERT system,[88] which contains a set of molecular fragment libraries and the spectral features of fragments (for IR spectroscopy, the ranges of characteristic frequencies, intensities and band half-widths). For example, the CH_2=CH– group is described by four diagnostic intervals: 880–900, 1395–1440, 1640–1655 and 3010–3120 cm^{-1}. The second approach[86,87] consists in generating spectra–structure correlations from a computer library of IR spectra, which is an automated knowledge acquisition from an IR database. The acquisition of spectral information is done in order to create rules allowing the identification of molecular fragments of an unknown compound from its IR spectrum. The process of searching for spectrum–structure correlations can be realized by computer-simulated neural networks,[89,90] by application of a statistical algorithm[77] or by the combination of a statistical algorithm with an algorithm related to the set theory.[86] The statistical algorithm (see Figure 12) splits the database into subsets with and without a given substructure. Next, it divides the IR spectral range into intervals, and counts the number of spectral features (absorption bands) present in the spectral intervals separately for each subset. Then, the normalized histograms (confidence factor (CF) for each interval) for both subsets are calculated. The difference between these histograms (interpreted as being the effect of only some characteristic spectral features) is converted into correlation tables. After their evaluation by a chemist, the correlations are automatically transformed into rules for substructure identification, a set of rules form a knowledge base being a part of the expert system for structure elucidation.

The basic lemma of the algorithm relying on the set theory applied by Ehreintreich et al.[86] is that every compound containing a substructure S must contain at least one absorption band in every characteristic interval described by a subset of the spectroscopic features: band location, intensity or half-width. This lemma is necessary with respect to strong logical conclusions but it is not fully justified with regard to excitation rules in IR spectroscopy (e.g. problem of center of symmetry). The core of the algorithm lies in a search for a large set of intervals as small as possible. The correlation obtained by these two algorithms can be finished manually by comparison with literature data and next transformed into a rule base for interpretation of IR spectra.

The application of this statistical algorithm for the generation of spectral correlations is presented for the amine group (see Figure 13a–c). The histogram has been plotted as a statistical area diagram using location (cm^{-1}) and CF as coordinates. The rectangular

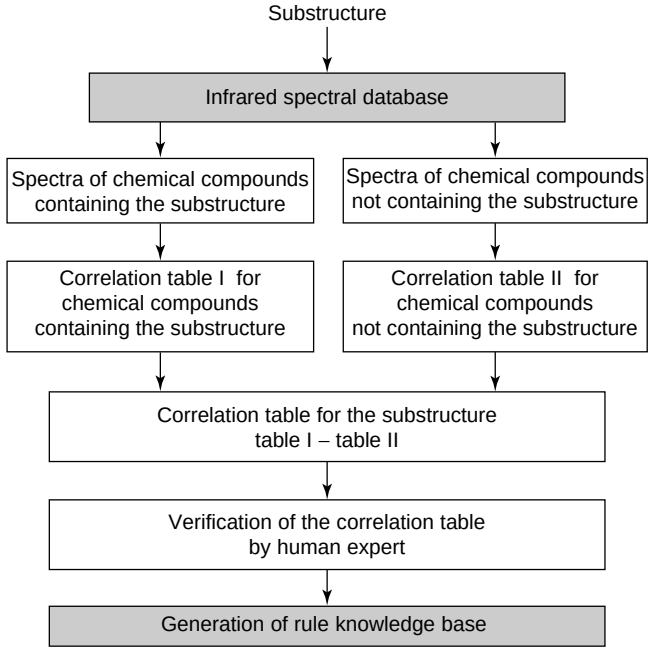

Figure 12 Algorithm corresponding to automatic knowledge acquisition from IR database.

height corresponds to the diagnostic power of the spectral interval. From among all the components of the diagnostic vector (see Figure 13a), the components of the maximum value components (CF < 20%) were chosen (see Figure 13b). Next, they were converted into correlation tables of the investigated substructure (see Figure 13c). One can see (see Figure 13b) that for the $-NH_2$ substructure three diagnostic regions fulfill these conditions. In the first region four intervals (Nos 1, 2, 4 and 5; yellow) were classed, in the second two intervals (Nos 3 and 7; gray) and in the third one interval (No 6; white).

The result of searching for structure correlations depends strongly on the IR database used in the process of searching. The spectral ranges generated from an IR database containing spectra of various types of chemical compounds are wider than the relevant ranges obtained by using dedicated bases (IR collections of carboxylic acids, hydrocarbons, etc.). The interpretation rules are of the general form

IF *conditions* **THEN** *hypothesis*

The *conditions* may be represented by logical expressions:[77,86,88] **A AND B**; **A OR B**, where A and B are the spectral features. In the *hypothesis* represented by logical expressions **C AND D**; **C OR D**, symbols C and D are the substructures identified by A and B spectral features. Thus, the rules can have the form:

IF A THEN C; **IF A AND B THEN C**;

IF A OR B THEN C AND D, etc.

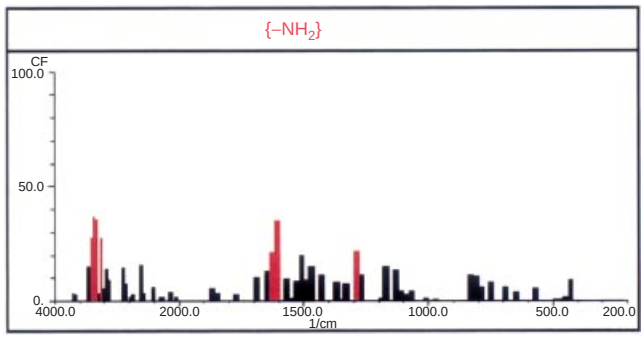

(a)

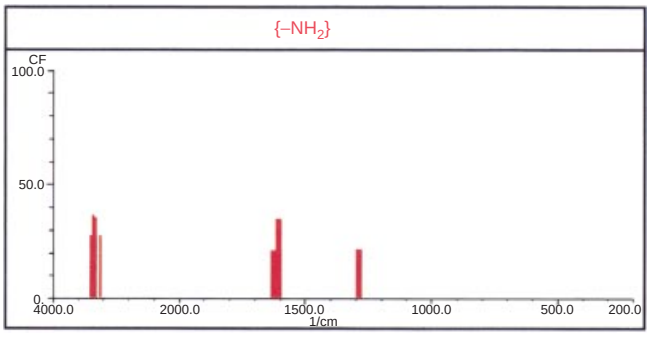

(b)

{–NH2}

Correlation table:

No	Band location [1/cm]	[1/cm]	Band intensity [Tx]	[Tx]	Confidence [x]
1	3320.00	3280.00	32.20	79.60	35.88
2	3360.00	3320.00	33.80	70.90	35.65
3	1620.00	1600.00	2.20	52.50	34.13
4	3400.00	3360.00	10.70	52.97	26.94
5	3240.00	3200.00	33.70	82.90	26.65
6	1300.00	1280.00	21.80	85.90	20.97
7	1610.00	1620.00	13.50	57.80	20.55

(c)

Figure 13 Spectrum–structure correlations generated for $-NH_2$ substructure.

A standard example of the substructure identification rule is as follows:

IF band_location ∈ interval_frequency[1]
AND band_intensity ∈ interval_band_intensity[1]
. . .
AND band_location ∈ interval_frequency[n]
AND band_intensity ∈ interval_band_intensity[n]
THEN substructure **IS** present

where n = number of the bands characteristic of the analyzed substructure.

This type of rule (related to spectral correlations) is then augmented by a set of rules required to control the inference engine, steering the way of reasoning and output of obtained results. The complete knowledge base

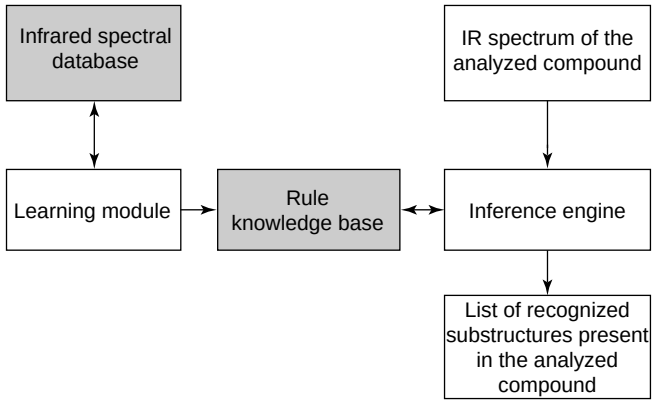

Figure 14 Block diagram of classification algorithm.

generated is in the next step compiled and used by the reasoning section of an expert system. During the consultation, the user enters spectral parameters (band locations and intensities) of the spectrum of an unknown compound. The inference engine accessing the knowledge base generates the list of substructures that would possibly constitute building parts of the investigated structure (see Figure 14).

5.3 Neural Networks as Tools for Structure Recognition

Apart from a LS,[91,92] the use of correlation tables and many different pattern recognition techniques, and also the application of computer-simulated neural networks (abbreviated to neural networks) working on a base of an encoded spectral collection, has been attracting increasing interest in the last few years. The main purpose of this method is to extract structural information from spectral data;[93–105] however, neural networks have also been used for predicting spectral features (i.e. for simulating spectra) from a chemical structure,[90,94,95,106,107] and also for the identification of IR spectra, i.e. to find in the IR spectral library a spectrum identical with the analyte spectrum[74] (the alternative approach to the LS method).

Neural networks[96] are known to be able to learn and then to recognize patterns, so neural networks constitute a means of pattern classification in which an input pattern (a vector the values of which represent object data) is transformed into an output pattern (a vector the values of which represent the desired classification of the object data). In structure elucidation, the input pattern is a representation of an IR spectrum of an unknown compound and the output pattern is a representation of functional groups (substructures) present or absent in the compound. The transformation is achieved by a propagation equation which combines the input vector with coefficients internal to the network. The coefficients are generated by a process called *training* or *learning*.

Starting from random values of the coefficients, output vectors are calculated from a set of input vectors (encoding IR spectra) for which the ''right answers'' (presence or absence of recognized groups) are known, the right answers being encoded in a target vector. After the calculation, a learning rule uses differences between the output values being obtained and target values to modify the coefficients. As this process is repeated, the coefficients gradually converge to values which transform each input pattern to an output pattern closely conforming to the target. Once a neural network has learned to recognize the given substructures, it can be used for the *prediction* of the presence or absence of these substructures in an unknown compound from its encoded IR spectrum.

For substructure recognition, the most commonly used approaches are ''error-back-propagation'' learning methods in multilayer neural networks, and two simple architectures, mono- and multioutputs. Mono-output networks are specialized networks trained to recognize only one particular substructure, whereas multioutput networks are trained to recognize simultaneously several substructures from a single input vector. The hierarchical system[105] of neural networks is half way between one large neural network with many outputs and many mono-output networks. In most work a digitized spectrum was the spectral information used as a network input, so that the number of input elements can vary between approximately 100 and 4000 (for an IR spectrum it is usually about 256 or 512 elements). Another means of spectrum encoding[96] is to divide a spectral range into intervals, assigning each interval to an input unit, and giving for the unit a nonzero value if a band occurs in a frequency interval, or zero otherwise.

An example of the application of neural networks to substructure prediction is the work presented by Novič and Zupan.[90] A training set of about 755 IR spectra was selected from the IR database using the Kohonen layer in a neural network. The spectra in the range from 3500 to 550 cm^{-1} are coded as 512 intensity points, which after Hadamard transformation are reduced to 128 intensity values (1 : 4 reduction). The structures of compounds included in a training set of spectra were coded as 34-dimensional binary vectors. The ones and zeros correspond to the presence or absence, respectively, of a particular fragment (34 structural fragments were defined and coded; see Table 7). These coded IR spectra and structure vectors corresponding to them were the input data to the neural network. It was assumed that the network after a training process will be able to recognize the presence or absence of these structural fragments in a molecule of an unknown compound from the compound's IR spectrum. The counterpropagation neural network, i.e. a two-layer network consisting of a Kohonen and an output layer, was used in the training process. The learned (trained) network was tested on a database of 2529 IR spectra (encoded in the same way as the training spectra set) of known substances. The prediction correctness was calculated and evaluated separately for two sets, a set of compounds containing the tested substructure and a set of compounds without this substructure. The evaluation of the neural network's predictive features was dependent on the number of substructures identified correctly. ''Correct identifying'' means that a given substructure was recognized in a structure of which it was really a part, and it was not recognized where it did not exist. The highest prediction correctness was for the groups carbonyl (96%), carboxyl (94%) and hydroxyl in alcohol (93%). Poor prediction results of less that 35% were obtained for furan (31%), CBr (30%), thiophene (19%) and CI (0%). It can be concluded that the vibrations caused by these and similar fragments are not significant in the IR region, or their influences on IR spectrum are lost among other (more dominant) spectral features.

Very interesting work concerning the correlation of IR spectra with a molecular structure were presented by Gasteiger's group.[106–110] The most important achievement was the successful application to spectrum–structure correlation problems of a novel and original way of coding 3D molecule structures. It is based on equations used in the analysis of the intensity distribution, which can be obtained in electron diffraction experiments. Structure coding and inductive learning algorithms allow one to store the relationship between the 3D structure of a molecule and its IR spectrum in a neural network. The relationship is stored in the weights of a counter-propagation neural network. Such a network is able to simulate an IR spectrum by inputting the coded structure of a molecule. On the other hand, a structural code of a molecule can be predicted by inputting an IR spectrum. As the second step, this structural code can be decoded to obtain Cartesian coordinates of the atoms in the molecule.

A number of authors proved that computer-simulated neural networks provide a practical method for detecting the presence and absence of functional groups on the basis of IR spectral data. Thus, the neural network can be used as a program for structure determination of an unknown chemical compound, and can also cooperate with an expert system for structure elucidation.[89] The combination of a neural network and an expert system can be used in two distinctly different manners. In the first, the network is used in cooperation with an expert system to analyze an unknown compound; in the second, the network is used to help in the creation of rules which the expert system will use. In the first case, the neural network is used directly and simultaneously in connection with the expert system, and the conclusions about the presence

Table 7 Results of prediction for set of substructures identified by neural network[90]

i	Structural fragment	Number of compounds in database	$^{+}p_i^{a}$	$^{-}p_i^{b}$	R_i^{c}
1	OH	630	0.91	0.04	0.93
2	Alcohol	462	0.93	0.03	0.95
3	Prim. alcohol	288	0.60	0.03	0.73
4	Sec. alcohol	171	0.76	0.05	0.83
5	Tert. alcohol	30	0.63	0.00	0.77
6	1,2-Glycol	29	0.43	0.01	0.59
7	Phenol	68	0.77	0.01	0.87
8	Aryl-CH_2OH	31	0.38	0.01	0.54
9	NH	412	0.85	0.04	0.90
10	Prim. amine	250	0.83	0.01	0.90
11	Sec. amine	146	0.71	0.04	0.80
12	CN	842	0.73	0.10	0.77
13	Tert. amine	170	0.60	0.02	0.73
14	C=O	1006	0.96	0.97	0.97
15	COOH	87	0.94	0.00	0.97
16	COO−	415	0.89	0.02	0.91
17	Ester	409	0.89	0.02	0.91
18	Aldehyde	120	0.46	0.02	0.59
19	Ketone	284	0.75	0.04	0.83
20	Amide	65	0.51	0.01	0.67
21	Benzene	1055	0.78	0.09	0.82
22	Naphthalene	28	0.56	0.01	0.71
23	Furan	25	0.31	0.00	0.47
24	Thiophene	29	0.19	0.01	0.31
25	Pyridine	81	0.49	0.02	0.64
26	NO_2	83	0.79	0.01	0.88
27	Aryl-NO_2	60	0.66	0.01	0.79
28	C−O	1456	0.88	0.10	0.88
29	Ether	471	0.62	0.06	0.72
30	C−X	841	0.66	0.13	0.69
31	C−F	218	0.60	0.04	0.72
32	C−Cl	438	0.52	0.11	0.58
33	C−Br	214	0.30	0.06	0.39
34	C−I	42	0.00	0.06	−0.13

[a] $^{+}p_i$ = fraction of correct prediction for compounds containing substructure number i (fraction of recognized substructure).
[b] $^{-}p_i$ = fraction of false prediction for compounds containing substructure number i (fraction of unrecognized substructure).
[c] R_i = reliability of substructure recognition, $R_i = 2(^{+}p_i - {}^{-}p_i)/(1 + |^{+}p_i - {}^{-}p_i|)$.

(or absence) of a substructure are entered into a database of facts which are then used by the expert system. In the second case, the neural network can be used indirectly, providing information (extracted from the coefficients of the network) which enable one to fix the correlations between the substructures and their spectral features. The correlations can next be taken into consideration when writing the rules for the expert system.

6 MULTIMETHOD SPECTRAL DATABASES

The computerized processing of spectral information has become a routine procedure for many laboratories in the last few years. The use of artificial intelligence to combine spectral information obtained by various spectral techniques is essential for the determination of a chemical structure. It is a more efficient solution than the application of only one spectral method because the results obtained by using different spectral techniques will be more accurate than when only one technique is used.

Multimethod SDBs (also called spectral data banks) contain some general information on chemical compounds (see section 4.1, definition of CDB) and/or on bibliographic references, as also monomethod databases obtained by use of various spectral techniques (see Table 4).

In this section, only spectral data banks containing also IR databases will be discussed. Each of the databases uses

a number of special programs, generally specific to a single spectroscopic method, for spectral data analysis and interpretation. The identification of a compound structure consists in the analysis of the results acquired from searching within separate databases, and the results can be combined to increase both scope and accuracy.[104] This function can be fulfilled by the main managing program summing the results of separate database searches. Usually, these results are shown as a list of compounds having bands identical with or similar to those for the examined compound. Finally, a global list including compounds present in all individual lists is created. Multimethod analysis provides more reliable results (higher SI) than the application of only one spectroscopic method. For all the compounds in a global list, all the spectra must comply with the requirements of similarity, i.e.

$$\text{global list} = (\text{IR list}) \text{ and } (^1\text{H-NMR list}) \text{ and } (\text{MS list})$$

Another procedure enables the user to select a different way of creating the global list, e.g.

$$\text{global list} = (\text{IR list}) \text{ and } [(\text{Raman list}) \text{ or } (\text{UV/VIS list})]$$

The chemist can choose ''and/or'' logical symbols and modify the search method according to his or her knowledge of the problem. If it is not possible to create a common results list (structures of a high SI), a list of substructures apparently occurring in the examined compound is built. Next, the probable structures and their spectra are generated. The calculated and experimental spectra are then compared with each other to choose the right structure from the candidate structures set.

Multimethod SDBs can be built by linking some spectral libraries with the already existing databases. The libraries can be obtained from industrial or university laboratories and also from various publications. Before being introduced into the bank, these data are carefully checked and transformed into a standard form. A multimethod system built in such a way includes databases which are not directly linked with each other. As a result of such system searching, separate structure lists are obtained. For example, Bio-Rad offers two Sadtler PC software products [i.e. IR Search Master version 3.0 and NMR Search Master version 3.1 (www.bio-rad.com/610441.html)] for searching various collections of IR and ^{13}C-NMR spectra.

One of the most important multimethod integrated SDBs is SpecInfo, which was developed by BASF for in-house use.[48] Actually, the SpecInfo system is available both as an in-house system through Chemical Concepts (www.dl.ac.uk:800/CDS/spectro.html) and for on-line use through the STN International (www.wiley-vch.de/cc/si_onl.html). The SpecInfo databases are updated thanks to the cooperation between the two experienced partner institutions [the Institute for Spectrochemistry and Applied Spectroscopy in Dortmund (IR and NMR spectra) and the Max-Planck-Institut für Kohlenforschung in Mülheim (mass spectra)[60]]. The SpecInfo on-line version provides information on chemical structures together with spectral information. It permits more than simple similarity searches and provides several run packages to assist users in important structure elucidation matters. For IR spectra, SpecInfo offers various search modes which may be performed using either the complete spectrum or just some of its defined ranges. Recent enhancements include cross-links to the CAS registry and the Beilstein Database.

Another example of multimethod SDBs, with entries connected through the structures of chemical compounds, which allows simultaneous access to all kinds of spectra stored for a given compound, is the Sadtler system. A user-friendly option gives the possibility of observing all the available spectra on a common screen. This option was adapted from, among others, the Sadtler User Library, the Main Manager program of which makes possible full viewing, plotting and coordination of IR and UV spectra, chemical structures, physical properties and bar charts for NMR and mass spectra, all in a sample spreadsheet format.[47] Another example of this kind of system is SCANNET, which assumes that within the identification methodology the results of a given randomly selected spectral interpretation (IR, MS, etc.) should influence the process of subsequent spectral interpretation through the cascade of facts logic.[71] This notion stands for the integrated interpretation of various molecular spectra of a given compound by the logical conjunction of intermediate analyses. The successive steps of the algorithm (see Figure 15) are as follows:

1. selection of a spectral method (i.e. SDB);
2. input of selected absorption band parameters and (if there is access) input of a spectrum in a discrete form;
3. searching the database for input absorption bands; the structure recognition of an unknown substance is made by comparing the input spectrum with

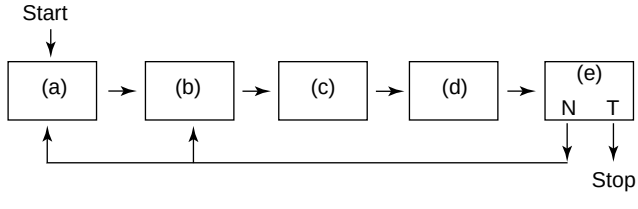

Figure 15 Block diagram of ''cascade of facts logic'' algorithm.

Table 8 Identification of an unknown compound by multispectral database

Cycle no.	Selected spectrum	Input band parameters	Number of compounds found in database	Global list		
				No.	Name of compound	SI
	(a)	(b)	(c)		(d)	
1	IR	3100–2800 (cm^{-1})	25	1	Phenethyl chloride	1.00
		55.0–5.0 (T%)		2	Ethylbenzene	1.00
				3	Isobutylbenzene	1.00
		1640–1580 (cm^{-1})		4	*tert*-Butylbenzene	0.92
		50.0–30.0 (T%)		5	Isopropylbenzene	0.92
				6	*sec*-Butylbenzene	0.92
		1520–1450 (cm^{-1})		7	*n*-Butylbenzene	0.92
		30.0–0.0 (T%)		8	*p*-Fluoronitrobenzene	0.84
				9	Bromobenzene	0.84
		770–670 (cm^{-1})		10	1-3-Dimethylbenzene	0.84
		20.0–0.0 (T%)		11	1-Methyl-3-chlorobenzene	0.84
				12	1-Methyl-2-ethylbenzene	0.84
		3650–3200 (cm^{-1})		13	*p*-Nitrotoluene	0.84
		100.0–0.0 (T%)		14	*o*-Nitrotoluene	0.84
				15	*p*-Tolunitrile	0.84
				16	Ethyl phenyl ether	0.84
				17	Diphenyl	0.84
				18	Nitrobenzene	0.84
				19	Phenyl ethyl ketone	0.84
				20	Ethyl benzoate	0.84
				21	1,2,3-Triethylbenzene	0.80
				22	*p*-Methoxytoluene	0.80
				23	*m*-Hydroxybenzaldehyde	0.80
				24	Chlorobenzene	0.80
				25	Phenethyl alcohol	0.80
2	IR	as above	25	1	Ethylbenzene	1.00
	MS	91.0 (m/z)	20	2	*n*-Butylbenzene	0.90
		100–80.0 (%)		3	Isopropylbenzene	0.90
				4	*sec*-Butylbenzene	0.90
		106.0 (m/z)		5	1,2,3-Triethylbenzene	0.89
		35.0–5.0 (%)		6	1-Methyl-2-ethylbenzene	0.89
				7	1-3-Dimethylbenzene	0.87
		51.0 (m/z)				
		20.0–0.0 (%)				
3	IR	as above	25	1	*n*-Butylbenzene	0.93
	MS	as above	20	2	Isopropylbenzene	0.87
	Raman	1050–950 (cm^{-1})	53	3	*sec*-Butylbenzene	0.87
		100.0–80.0 (–)				

T = transmittance; – = scale of intensity value is relative.

successive reference spectra of standard compounds (chemical compounds of a known chemical structure) stored in a SDB;

4. creation of the global list and analyzing the final list of chemical substances for which the spectra are identical with and/or similar to those of the compound being searched for;
5. the end of a structure elucidation (see Figure 15, STOP command), if the output list satisfies the chemist; he or she can reinterpret the results by presenting the overlapping spectra on one screen and deciding to repeat the search process, choosing (a) new spectral methods or (b) other spectral band parameters.

An example of the use of the cascade of facts logic[46] is a structure recognition of an unknown compound led only on a basis of IR, mass and Raman spectra. The identification is realized in three subsequent cycles. When the investigated structure does not exist in a database, the chosen spectra have a SI smaller than 1.0 (see Table 8). If the process of searching an SDB gives a list of similar spectra, the result of the investigation is a hypothesis about a possible structural similarity between

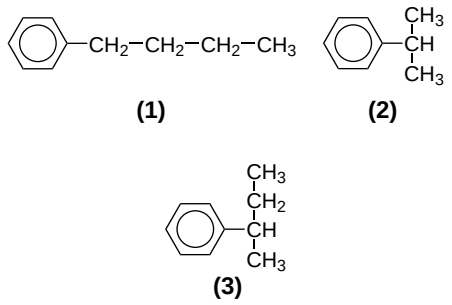

Figure 16 Structures of compounds with the largest degree of similarity of spectra.

the structure of the unknown compound and one or several structures from the obtained list of similar spectra, depending on the SI. The result of the process of IR, mass and Raman spectra interpretation, three proposed structures, is presented in Table 8 and Figure 16. In the first cycle of recognition, the identification was carried out using only the IR spectrum, as a result of which 25 compounds (column c, global list) were found similar to the examined one. In the second cycle, the MS database was searched and 20 compounds (column c) were found similar to the analyzed substance. The comparison of the two lists (IR and MS) reduced the number of candidate compounds to seven substances (column d, global list). In the third cycle, similar analyses were performed for Raman Spectroscopy. From the Raman database 53 spectra were selected as similar to the Raman spectra of the unknown compound. Comparison of the Raman list with the IR and MS global lists reduced the number of candidate compounds to three. All spectra of the selected compounds (Figure 16) are similar to the analyzed compound's spectra, with *n*-butylbenzene being identified as the most similar one (SI = 0.93). As shown, the compound was univocally identified. All spectra of the compound stored in the system database can be seen on one screen (see Figure 17).

ACKNOWLEDGMENTS

The authors are especially grateful to Professor D. Cabrol-Bass (Nice–Sophia Antipolis University, Nice, France), Professor M.E. Elyashberg (All-Russia Research Institute of Organic Synthesis, Moscow, Russia), Professor L.A. Gribov (V.I. Vernadsky Institute of Geochemistry and Analytical Chemistry, Russian Academy of Sciences, Moscow, Russia), Professor Z.H. Hippe (Rzeszów University of Technology, Rzeszów, Poland) and Professor J. Zupan and Dr M. Novič (Institute of Chemistry, Ljubljana, Slovenia) for the reprints of their papers, which were a great help in our work on this article.

We express special acknowledgment to Professor J.T. Clerc (University of Berne, Swiss Federal Institute of Technology, Zurich, Switzerland) for his great interest, encouragement and help in our work.

We appreciate the support of the State Committee for Scientific Research (Warsaw) with Grants 8 8308 92 03 and 8 T11C 004 09 to assist our research in the field of IR SDBs.

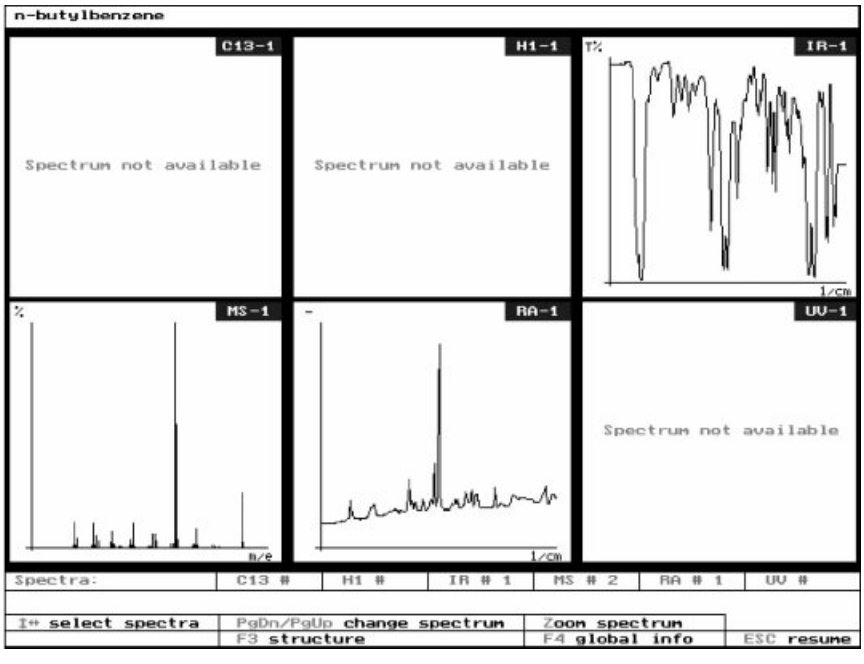

Figure 17 Spectra of *n*-butylbenzene stored in the database.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
CAS	Chemical Abstracts Service
CDB	Compound Identification Database
CF	Confidence Factor
^{13}C-NMR	Carbon-13 Nuclear Magnetic Resonance
DMS	Documentation of Molecular Spectroscopy
FTIR	Fourier Transform Infrared
^{1}H-NMR	Proton Magnetic Resonance
IR	Infrared
IUPAC	International Union of Pure and Applied Chemistry
IUPAC-CPEP	International Union of Pure and Applied Chemistry Committee on Printed and Electronic Publications
JCAMP	Joint Committee on Atomic and Molecular Physical Data
JCAMP-CS	Joint Committee on Atomic and Molecular Physical Chemical Structure
JCAMP-DX	Joint Committee on Atomic and Molecular Physical Data Exchange
LIMS	Laboratory Information Management System
LS	Library Search
MS	Mass Spectrometry
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
PC	Personal Computer
QI	Quality Index
SDB	Spectral Database
SI	Similarity Index
STN	Scientific and Technical Information Network
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
WLN	Wiswesser Line Notation
2D	Two-dimensional
3D	Three-dimensional

RELATED ARTICLES

Infrared Spectroscopy **(Volume 12)**
Interpretation of Infrared Spectra, A Practical Approach
• Theory of Infrared Spectroscopy

Mass Spectrometry **(Volume 13)**
Artificial Intelligence and Expert Systems in Mass Spectrometry

REFERENCES

1. W. Abney, R.E. Festing, *Philos. Trans. R. Soc. London*, **172**, 1530 (1881).
2. W.W. Coblentz, *Investigation of Infrared Spectra, Part I: Infrared Absorption Spectra*, Bull. No. 35, Carnegie Institution Publications, Washington, DC, 1905.
3. *Documentation of Molecular Spectroscopy*, Butterworth, London and Verlag Chemie, Weinheim, 1967–1970.
4. *The Sadtler Handbook of Infrared Spectra*, Heyden, London, 1978.
5. D.O. Hummel, F. Scholl, *Atlas der Kunststoffanalyze*, Carl Hauser Verlag, Munich and Verlag Chemie, Weinheim, 1968.
6. C.D. Craver (ed.), *Coblentz Society Desk Book of Infrared Spectra*, 2nd edition, Coblentz Society, Kirkewood, MO, 1982.
7. C.D. Craver (ed.), *Evaluated Infrared Reference Spectra/Coblentz Society: part through National Standard Reference Data System, with Cooperation from the Joint Committee on Atomic and Molecular Physical Data*, Coblentz Society, Glendale, MO, 1979.
8. *Selected Infrared Spectral Data. Thermodynamics Research Center Data Project*, Thermodynamic Researches Center, Texas A&M University, College Station, TX, 1975.
9. *Sadtler Standard Infrared Grating Spectra (on Microfiche)*, Sadtler Research Laboratories, Philadelphia, PA, 1980.
10. L.E. Kuentzel, 'New Codes for Hollerith-type Punched Cards', *Anal. Chem.*, **23**, 1413–1418 (1951).
11. R.A. Sparks, *Storage and Retrieval of Wyandotte–ASTM Infrared Spectral Data Using an IBM 1401 Computer*, ASTM, Philadelphia, PA, 1964.
12. B. Schrader, *Raman/Infrared Atlas of Organic Compounds*, Verlag Chemie, Weinheim, 1994.
13. Ch.J. Pouchert, *The Aldrich Library of Infrared Spectra*, 3rd edition, Aldrich Chemical, Milwaukee, WI, 1981 (cited in *Aldrich Spectroscopy Catalog*, 1997).
14. *The Aldrich Library of FTIR Spectra*, 2nd edition, Aldrich Chemical, Milwaukee, WI, 1997 (cited in *Aldrich Spectroscopy Catalog*, 1997).
15. Ch.J. Pouchert, *The Aldrich Library of FTIR Spectra: Vapor Phase*, Aldrich Chemical, Milwaukee, WI, 1989 (cited in *Aldrich Spectroscopy Catalog*, 1997).
16. *The Aldrich Library of FTIR/Raman Spectra*, Aldrich Chemical, Milwaukee, WI, 1997 (cited in *Aldrich Spectroscopy Catalog*, 1997).
17. C.D. Craver (ed.), *Coblentz Society Reference Spectra Collection*, Coblentz Society, 1992.

18. D.O. Hummel, *Atlas of Polymer and Plastics Analysis*, Verlag Chemie, Weinheim, 1985.
19. A.H. Kuptsov, G.N. Zhizhin, *Handbook of Fourier Transform Raman and Infrared Spectra of Polymers*, Physical Sciences Data, Vol. 45, Elsevier, Amsterdam, 1998.
20. *An Infrared Spectroscopy Atlas for the Coatings Industry*, Infrared Spectroscopy Committee of the Chicago Society for Coatings Technology, Federation of Societies for Coatings Technology, Philadelphia, PA, 1980.
21. D.S. Erley, 'A Quantitative Evaluation of Several Infrared Searching Systems', *Appl. Spectrosc.*, **25**(2), 200–202 (1971).
22. IRIS, Infrared Information System. User's Guide, Sadtler Research Laboratories, Philadelphia, PA, and University Computing Company, Dallas, TX, 1971.
23. R.W. Sebesta, G.G. Johnson, Jr, 'New Computerized Infrared Substance Identification System', *Anal. Chem.*, **44**(2), 260–265 (1972).
24. C.D. Craver, *Infrared Spectral Search by Computer Time Sharing*, Singer Technical Services, New York, 1979.
25. E.M. Kirby, R.N. Jones, D.G. Cameron, 'SPIR – a Search Program for Infrared Spectra', *CODATA Bull.*, **21**, 18–25 (1976).
26. J. Zupan, D. Hadzi, M. Penca, 'ZAPAH Search System for Infrared Spectra', *Kem. Ind.*, **23**(5), 275–277 (1974).
27. J. Zupan, D. Hadzi, M. Penca, 'A New Retrieval System for Infrared Spectra', *Comput. Chem.*, **1**, 71 (1976).
28. J. Zupan, D. Hadzi, M. Penca, J. Marsel, 'Combined Retrieval System for Infrared, Mass and Carbon-13 Nuclear Magnetic Resonance Spectra', *Anal. Chem.*, **49**(14), 2141–2146 (1977).
29. K. Maeda, Y. Koyama, K. Sato, S. Sasaki, 'Combination of Analytical Spectrometers and Spectroscopic Databases', *Anal. Chim. Acta*, **133**(4), 561–574 (1981).
30. B. Dębska, Z.S. Hippe, A. Kerste, G. Kruczek, *Computer Systems for Polymers and Macromolecular Compound Identification*, Rzeszów University of Technology, Rzeszów, 1976.
31. J.P. Drobyschev, R.S. Nigmatulin, W.I. Lobanov, I.K. Korobeinicheva, W.S. Bocharev, W.A. Koptyug, 'Application of Computers for Recognition of Chemical Compounds for their Spectral Characteristics (a Retrieval System for IR Spectra)', *Vestn. Akad. Nauk SSSR*, **8**, 75–83 (1970).
32. J.P. Drobyschev, R.S. Nigmatulin, W.I. Lobanov, I.K. Korobeinicheva, W.S. Bocharev, W.A. Koptyug, 'Application of Computer for Recognition of Chemical Compounds for their Spectral Characteristics. II (A Retrieval System for IR Spectra)', *Izv. Sib. Otd. Akad. Nauk SSSR*, **2**, 1 (1972).
33. K. Schaarschmidt, R. Riemer, E. Steger, 'Machine Oriented Searching System for Infrared Spectra', *J. Chem.*, **14**(9), 374–375 (1974).
34. A. Kaczmarek, 'Computer System for Structure Recognition and Spectrum–Structure Correlations Investigation', PhD Thesis, University of Mining and Metallurgy, Kracow, 1985.
35. Y. Hu, E.F. Westrum, Jr (eds.), 'Data Activities and Database Developments in China, Japan, and Korea', CODATA Special Report, University of Michigan, 12, 1989.
36. V.V. Komolov, L.N. Drozdov-Tikhomirov, 'Organization of the Database and Exchange Operation in an Automated Information Retrieval System for Infrared Spectroscopy', *Nauchn.-Tekh. Inf. Ser.*, **2**(9), 18–24 (1976).
37. M. Razinger, M. Penca, J. Zupan, M. Janezic, 'Computerized Search System Based on Digitized Infrared Spectra', *Fresenius' J. Anal. Chem.*, **313**(6), 496–499 (1982).
38. *Sadtler Spectra/Digilab FTS-IR*, Sadtler Research Laboratories, Division of Bio-Rad, Philadelphia, PA, 1975.
39. *Infrared Spectroscopy. Accessories and Supplies*, Promotion Pack, Perkin-Elmer, Norwalk, CT, 1993.
40. *Aldrich Spectroscopy Catalog*, Aldrich Chemical, Milwaukee, WI, 1997.
41. E. Zeiberger, T. Fröhlich, 'The Hoechst Infrared and Raman Spectroscopic Database', *Spectrosc. Eur.*, **5**, 16 (1993).
42. Ch.J. Pouchert, *The Aldrich Library of FTIR Spectra*, digitized version of the printed catalog, Aldrich Chemical, Milwaukee, WI, 1985 (cited in *Catalog Handbook of Fine Chemicals, Aldrich 1986–1987*, Aldrich Chemical, Milwaukee, WI, 1988).
43. *The Sigma Library of FT-IR Spectra*, (cited in *Catalog Handbook of Fine Chemicals, Aldrich 1986–1987*, Aldrich Chemical, Milwaukee, WI, 1988).
44. *Bruker FTIR Libraries*, Promotion Pack, Bruker Analytische Messtechnik, Karlsruhe, 1990.
45. W.A. Warr, 'Spectral Databases', *Chemom. Intell. Lab. Syst.*, **10**, 297 (1991).
46. B. Dębska, 'SCANNET – A Spectroscopic Multimethod System for Computer-aided Structural Identification', *Comput. Chem.*, **19**(3), 269–275 (1995).
47. *IR Search Master*, Promotion Pack, Sadtler Division, Bio-Rad Laboratories, Philadelphia, PA, 1994.
48. A. Barth, 'SpecInfo: An Integrated Spectroscopic Information System', *J. Chem. Inf. Comput. Sci.*, **33**, 52–58 (1993).
49. A.N. Davies, R. Lancashire, 'Introducing the Working Party on Spectroscopic Data Standards (JCAMP-DX)', *Chem. Int.*, **19**(2), 39–41 (1997).
50. A.N. Davies, 'Spectroscopy Data Handling', *Spectroscopy Int.*, **4**(1), 18, 20, 21 (1992).
51. R.S. McDonald, P.A. Wilks, 'JCAMP-DX: A Standard Form Exchange of Infrared Spectra in Computer Readable Form', *Appl. Spectrosc.*, **42**, 151–162 (1988).

52. J. Gasteiger, B.M.P. Hendirks, P. Hoever, C. Jochum, H. Somberg, 'JCAMP-CS: A Standard Exchange Format for Chemical Structure Information in Computer-readable Form', *Appl. Spectrosc.*, **45**(1), 4–11 (1991).
53. *Nicolet FTIR Spectral Data Base/Automated Identification System*, Nicolet Analytical Division, Madison, WI, 1971.
54. *OPUS – The Software System for Infrared and Optical Spectroscopy*, Bruker Spectrospin, Karlsruhe, 1992.
55. P.R. Griffiths, C.L. Wilkins, 'Quality Criteria for Digital Infrared Reference Spectra', *Appl. Spectrosc.*, **42**, 538 (1988).
56. K.S. Seshadri, R.N. Jones, 'The Shapes and Intensities of Infrared Absorption Bands – A Review', *Spectrochim. Acta*, **19**, 1013–1085 (1963).
57. V.A. Koptyug, 'Importance of Spectral Collections and Computers in Chemical Structure Studies. Scientific Information Center for Molecular Spectroscopy at the Siberian Division of the Academy of Sciences of the USSR', *J. Chem.*, **15**, 41–47 (1975).
58. E.J. Karjalainen (ed.), *Data Handling in Science and Technology*, Vol. 6, Proceedings of Conference on Scientific Computing and Automation (Europe), 1990, Elsevier, Amsterdam, 1990.
59. H. Sadeghi-Jordadchi, S. Butler, B. Patterson, D. Croft, 'Everything You Wanted to Know about Digitizing Spectra but Were Afraid to Ask', *Spectroscopy*, **8**(3), 8, 10, 12 (1996).
60. *SpecInfo*, Promotion Pack, Chemical Concepts, Weinheim, 1991.
61. B. Dębska, 'The Methodology and Spectroscopic Multimethod System for Computer-assisted Structural Elucidation of Organic Compounds', PhD Thesis, University of Technology, Lódź, 1995.
62. A. Savitzky, S. Geary, A. Muir, R.W. Hannah, presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, 1979.
63. J.P. Hawranek, 'On the Numerical Description of Asymmetric Absorption Bands', *Acta Phys. Pol.*, **A40**, 811–814 (1971).
64. R.N. Jones, 'Computer Programs for Absorption Spectrophotometry', *Appl. Opt.*, **8**(3), 597–601 (1969).
65. R.N. Jones, K.S. Seshadri, N.B.W. Jonathan, 'A Statistical Approach to the Analysis of Infrared Band Profiles', *Can. J. Chem.*, **41**, 750–762 (1963).
66. A. Savitzky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Anal. Chem.*, **36**(8), 1627–1639 (1964).
67. *PeakFit – Peak Analysis Software*, Jandel Scientific, Erkrath, 1994.
68. Z.S. Hippe, *Artificial Intelligence in Chemistry*, Elsevier, Amsterdam, 1991.
69. J.T. Clerc, F. Erni, C. Jost, J. Meili, P. Nëgli, R. Schwarzenbach, 'Computerunterstütze Spektreninterpretation zur Strukturaufklärung organischer Verbindungen', *Fresenius' J. Anal. Chem.*, **264**, 192–195 (1973).
70. J.T. Clerc, R. Knutti, H. Koenitzer, J. Zupan, 'Conversion of a Conventional Infrared Library into Computer-readable Form', *Fresenius' J. Anal. Chem.*, **283**, 177–181 (1977).
71. B. Dębska, 'SCANNET – Multispectral Knowledgebase', *J. Mol. Struct.*, **348**, 131–134 (1995).
72. P.R. Naegeli, J.T. Clerc, 'Computer System for Structural Identification of Organic Compounds from Spectroscopic Data', *Anal. Chem.*, **46**(8), 739A–744A (1977).
73. P.H. Wilson, *Artificial Intelligence*, 3rd edition, Addison-Wesley, Reading, MA, 1992.
74. K. Tanabe, T. Tamura, H. Uesaka, 'Neural Network System for the Identification of Infrared Spectra', *Appl. Spectrosc.*, **46**(5), 807–810 (1992).
75. B. Guzowska-Świder, Z.S. Hippe, 'Structure Elucidation of Organic Compounds Aided by Computer Program System SCANNET', *J. Mol. Struct.*, **275**, 225–234 (1992).
76. H.B. Woodruff, M.E. Munk, 'A Computerized Infrared Spectral Interpreter as a Tool in Structure Elucidation of Natural Products', *J. Org. Chem.*, **42**(10), 1761–1767 (1977).
77. B. Dębska, B. Guzowska-Świder, 'Knowledge Discovery in an Infrared Database', *Comput. Chem.*, **21**(1), 51–59 (1997).
78. S. Sasaki, Y. Abe, Y. Hirota, Y. Ishida, Y. Kudo, S. Ochiai, K. Saito, T. Yamasaki, 'CHEMICS-F: Computer Program System for Structure Elucidation of Organic Compounds', *J. Chem. Inf. Comput. Sci.*, **18**, 211 (1978).
79. S. Sasaki, I. Fujiwara, H. Abe, T. Yamasaki, 'A Computer Program System – New Chemics – for Structure Elucidation of Organic Compounds by Spectral and Other Structural Information', *Anal. Chim. Acta*, **122**, 87 (1980).
80. H.B. Woodruff, G.M. Smith, 'Computer Program for the Analysis of Infrared Spectra', *J. Am. Chem. Soc.*, **52**, 2321–2327 (1980).
81. M.O. Turlson, M.E. Munk, 'Table-driven Procedure for Infrared Spectrum Interpretation', *J. Am. Chem. Soc.*, **55**, 2137–2142 (1983).
82. N.A.B. Gray, *Computer Assisted Structure Elucidation*, John Wiley & Sons, New York, 1986.
83. S. Moldoveanu, C.A. Rapson, 'Spectral Interpretation for Organic Analysis Using an Expert System', *Anal. Chem.*, **59**, 1207–1212 (1987).
84. H. Huixiao, X. Xinquan, 'ESSESA: An Expert System for Elucidation of Structures from Spectra. 1. Knowledgebase of Infrared Spectra and Analysis and Interpretation Programs', *J. Chem. Inf. Comput. Sci.*, **30**, 203–210 (1990).
85. L.A. Gribov, V.A. Dementyev, M.E. Elyashberg, E.Z. Yakapov, 'Automation of Spectrochemical Investigations', *J. Mol. Struct.*, **22**, 161 (1974).

86. F. Ehrentreich, U. Dietze, U. Meyer, H. Schulz, H.M. Klötzer, S. Abbas, M. Otto, '"IR-Spectroscopy" – Computer Program for the Interpretation of Infrared spectra', *Fresenius' J. Anal. Chem.*, **354**, 829–832 (1996).
87. F. Ehrentreich, 'Derivation of Substructures from Infrared Band Shapes by Fuzzy Logic and Partial Cross Correlation Functions', *Fresenius' J. Anal. Chem.*, **359**, 56–60 (1997).
88. M.E. Elyashberg, E.R. Martirosian, Yu.Z. Karasev, H. Thiele, H. Somberg, 'X-PERT: A User-friendly Expert System for Molecular Structure Elucidation by Spectral Methods', *Anal. Chim. Acta*, **337**, 265–286 (1997).
89. D. Ricard, C. Cachet, D. Cabrol-Bass, 'Neural Network Approach to Structural Feature Recognition from Infrared Spectra', *J. Chem. Inf. Comput. Sci.*, **33**, 202–210 (1993).
90. M. Novič, J. Zupan, 'Investigation of Infrared Spectra–Structure Correlation Using Kohonen and Counterpropagation Neural Network', *J. Chem. Inf. Comput. Sci.*, **35**, 454–466 (1995).
91. J. Zupan, *Computer Supported Spectroscopic Databases*, Ellis Horwood, Chichester, 1986.
92. P.C. Jurs, *Computer Software Applications in Chemistry*, John Wiley & Sons, New York, 1986.
93. R.J. Fessenden, L. Györgyi, 'Identifying Functional Groups in IR Spectra Using an Artificial Neural Network', *J. Chem. Soc., Perkin Trans.*, **2**, 1755–1762 (1991).
94. U.M. Weigel, R. Herges, 'Automatic Interpretation of Infrared Spectra: Recognition of Aromatic Substitution Patterns Using Neural Networks', *J. Chem. Inf. Comput. Sci.*, **32**, 723–731 (1992).
95. M. Meyer, T. Weigelt, 'Interpretation of Infrared Spectra by Artificial Neural Networks', *Anal. Chim. Acta*, **265**, 183–190 (1992).
96. M.E. Munk, M.S. Madison, E.W. Robb, 'Neural Network Models for Infrared Spectrum Interpretation', *Microchim. Acta*, **2**(1–6), 505–514 (1991).
97. M. Novič, J. Zupan, '2-D Mapping of Infrared Spectra Using Kohonen Neural Network', *Vestn. Slov. Kem. Drus.*, **39**(2), 195–212 (1992).
98. M. Meyer, K. Meyer, H. Hobert, 'Neural Networks for Interpretation of Infrared Spectra Using Extremely Reduced Spectral Data', *Anal. Chim. Acta*, **282**, 407–415 (1993).
99. T. Visser, H.J. Luinge, J.H. van der Maas, 'Recognition of Visual Characteristics of Infrared Spectra by Artificial Neural Networks and Partial Least Squares Regression', *Anal. Chim. Acta*, **296**, 141–154 (1994).
100. N. Sbirrazzuoli, C. Cachet, D. Cabrol-Bass, T.P. Forrest, 'Indices for Evaluation of Neural Network Performance as Classifier: Application to Structural Elucidation in Infrared Spectroscopy', *Neural Comput. Appl.*, 229–239 (1993).
101. Ch. Klawun, Ch.L. Wilkins, 'Optimization of Functional Group Prediction from Infrared Spectra Using Neural Networks', *J. Chem. Inf. Comput. Sci.*, **36**, 69–81 (1996).
102. Ch. Klawun, Ch.L. Wilkins, 'Joint Neural Network Interpretation of Infrared and Mass Spectra', *J. Chem. Inf. Comput. Sci.*, **36**, 249–257 (1996).
103. M.E. Munk, M.S. Madison, E.W. Robb, 'The Neural Network as a Tool for Multispectral Interpretation', *J. Chem. Inf. Comput. Sci.*, **36**, 231–238 (1996).
104. M.E. Munk, 'Computer-based Structure Determination: Then and Now', *J. Chem. Inf. Comput. Sci.*, **38**, 997–1009 (1998).
105. Ch. Cleva, C. Cachet, D. Cabrol-Bass, T.P. Forrest, 'Advantages of a Hierarchical System of Neural Networks for the Interpretation of Infrared Spectra in Structure Determination', *Anal. Chim. Acta*, **348**, 255–265 (1997).
106. J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, 'Chemical Information in 3D Space', *J. Chem. Inf. Comput. Sci.*, **36**(5), 1030–1037 (1996).
107. C. Affolter, C. Clerc, 'Prediction of Infrared Spectra from Chemical Structures of Organic Compounds Using Neural Networks', *Chemom. Intell. Lab. Syst.*, **18**, 27–39 (1993).
108. J. Gasteiger, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, 'Finding the 3D Structure of a Molecule in Its IR Spectrum', *Fresenius' J. Anal. Chem.*, **359**, 50–55 (1997).
109. J.H. Schuur, P. Selzer, J. Gasteiger, 'The Coding of the Three-dimensional Structure of Molecular Transforms and Its Application to Structure–Spectra Correlations and Studies of Biological Activity', *J. Chem. Inf. Comput. Sci.*, **36**(5), 334–344 (1996).
110. J. Schuur, J. Gasteiger, 'Infrared Spectra Simulation of Substituted Benzene Derivatives on the Basis of a 3D Structure Representation', *Anal. Chem.*, **69**, 2398–2405 (1997).

Sum Frequency Generation Spectroscopy

Arthur McClelland and Zhan Chen
University of Michigan, Ann Arbor, MI, USA

1 Introduction **1**
2 Sum Frequency Generation Theory and Instrumentation **2**
3 Sum Frequency Generation Applications **4**
3.1 Small Molecules at Interfaces Involving Liquids 4
3.2 Polymer Surfaces and Interfaces 7
3.3 Sum Frequency Generation Studies on Proteins and Peptides at Interfaces 17
4 Conclusion **26**
Acknowledgments **26**
References **26**

Sum frequency generation (SFG) vibrational spectroscopy is a nonlinear optical vibrational spectroscopy, which has been shown to have submonolayer surface sensitivity. It has been developed into a powerful technique to investigate molecular structures of surfaces and interfaces. This paper introduces the technique of SFG spectroscopy and discusses several important SFG applications. One of the most important advantages of SFG over many traditional surface-sensitive techniques is that SFG does not require high vacuum to operate. Therefore, it can be applied to investigate liquid surfaces and interfaces, as well as interfacial biological molecules in biologically relevant environments. The SFG applications discussed in this article are (i) simple liquid surfaces and interfaces, (ii) polymer surfaces in water and polymer/silane interfaces, and (iii) interfacial structures of peptides and proteins.

1 INTRODUCTION

The chemistry at surfaces and interfaces is important for everything from high-tech industry to everyday life. The usefulness of many common items is determined by surface properties, for example contact lenses must remain wetted, while raincoats are designed to be nonwetting. The chemistry of surfaces has been tailored for many applications including antibiofouling surfaces for marine vessels, high-temperature-resistant surfaces for space shuttles, and heterogeneous catalysts.

Surface chemistry is also important in biology. The human body contains many different surfaces that are vital to our well-being. Surface chemistry is responsible for protein interactions with cell surfaces, hormone-receptor interactions, and lung function for example. To understand functions of biological molecules at interfaces inside our bodies and to prepare surfaces of advanced materials with desired properties, it is essential to understand the molecular-level structures of these surfaces in situ.

Excellent research on many surfaces and interfaces has been performed using a wide variety of analytical techniques.[1,2] However, many challenges still exist in the field of surface analytical chemistry. Many of the standard surface analytical chemistry techniques such as X-ray photo electron spectroscopy (XPS), secondary ion mass spectrometry (SIMS), scanning electron microscopy (SEM), and scanning tunneling microscopy (STM) require high vacuum to operate. These techniques are powerful surface-sensitive analytical tools, but it is difficult to apply them to the study of liquid surfaces/interfaces or buried solid interfaces.

Various techniques have been researched to study surface structures in liquid environments. For example, freeze-drying XPS has been developed to study polymer surface restructuring in water,[3] under the assumption that the freeze-drying process would not alter the polymer surface structure and thus the polymer surface will have the same structure as that in water. Such a method only provides surface elemental analysis rather than surface functional groups.

Contact angle measurements are easier to perform and can be used to study surface hydrophobicity.[4,5] Many surfaces (e.g. polymer surfaces) restructure when they are exposed to water, but contact angle measurements cannot provide a molecular-level picture of how the polymer surfaces change. These changes should have a molecular origin; that is, they are thought to be caused by the migration or reorientation of groups, segments, or side chains relative to the aqueous phase to minimize the interfacial free energy. Atomic force microscopy (AFM) has been used to study surface morphologies and surface properties (e.g. friction or adhesion) of many materials in a variety of chemical environments including aqueous solutions with excellent spatial resolution, but AFM cannot probe surface chemical structures.[6,7] Different theoretical models have been developed to simulate the details of the surface restructuring in water for decades, but the models need to be verified by molecular-level experiments.

Those techniques that can provide molecular-level structural information and do not require high vacuum, such as attenuated total reflection Fourier transform

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

infrared (ATR-FTIR) spectroscopy and surface enhanced Raman scattering (SERS), are less surface specific than SFG. ATR-FTIR detects signals from many layers, rather than the topmost layer on the surface. In addition, SERS requires special substrates to enhance the signal, which may not be ideal to study different surfaces and interfaces like biosurfaces.

SFG vibrational spectroscopy has been developed into a powerful analytical technique to examine molecular structures of surfaces and buried interfaces.[8–15] SFG is a second-order nonlinear optical spectroscopy, which can provide vibrational spectra of a surface or an interface with submonolayer sensitivity.[10–15] It does not require a high vacuum to operate; therefore it can be used to probe various surfaces/interfaces including liquid surfaces under ambient conditions. It can follow molecular-level chemical structural changes of surfaces and buried interfaces in situ and in real time. SFG spectrometers are more complicated compared to traditional Fourier transform infrared spectroscopy (FTIR) or Raman systems. Owing to the fast development of advanced lasers and nonlinear optical research, SFG equipment is becoming more mature, stable, and user-friendly. SFG research has grown substantially in recent years and numerous academic and industrial labs have been equipped with SFG systems (*see* **Theory of Infrared Spectroscopy**; **Infrared Spectroscopy: Introduction**; **Raman Spectroscopy: Introduction**; **Fourier Transform Raman Instrumentation**; **Raman Microscopy and Imaging**; **Raman Scattering, Fundamentals**; **Ultrafast Laser Technology and Spectroscopy**).

The aim of this article is to introduce the SFG technique and its applications to scientists in other research fields and to beginners in SFG studies who are not very familiar with this technique. Thus, only a brief introduction instead of an in-depth discussion of the SFG theory and technique is presented here. Many quantitative SFG data analysis methods and SFG technical details are omitted. Instead of thoroughly covering all the excellent SFG results from the many SFG research labs on a wide range of different research subjects, this article presents only applications of SFG with respect to several particular research directions to deliver a general idea to the readers about what SFG can examine and why the results obtained from such SFG studies are important and unique. More details can be found in the numerous excellent SFG review papers and original research articles.

The article is organized as follows: after this brief introduction, SFG theory and instrumentation are briefly discussed. Then SFG applications in studying liquid surface and interfaces of small molecules, polymer surfaces and interfaces, and interfacial biological molecules such as proteins and peptides, respectively, are presented. As stated, the purpose of this article is to introduce the SFG technique and its applications to the readers who do not know this technique well. It is hoped that this article will promote their interests in exploring the possibility of applying SFG spectroscopy to solve some of the questions in their research areas.

2 SUM FREQUENCY GENERATION THEORY AND INSTRUMENTATION

SFG spectroscopy provides vibrational spectra of molecules. SFG can elucidate surface/interfacial structures at the molecular level for a centrosymmetric material. FTIR and Raman scattering are the most widely used vibrational spectroscopic techniques. An SFG spectrum is similar to an FTIR or Raman spectrum, which plots the signal intensity as a function of infrared (IR) wavenumber, but an SFG spectrum is specifically from the surface functional groups in a centrosymmetric medium.

FTIR and Raman are linear spectroscopic techniques. FTIR is a one-photon process where an IR photon interacts with a molecule, and if its frequency is in resonance with a vibrational mode of the molecule and if the dipole moment changes during the vibration, the photon can be absorbed. A normal Raman process is a scattering process that involves two photons. The incoming photon can gain or lose energy when it interacts with the molecule. The outgoing photon will contain information about how much energy was gained or lost to the molecule, which is dependent on the vibrational energy levels of the molecule. The energy difference between the two photons matches the energy difference between vibrational states of the molecule. A vibrational mode is Raman active if the polarizability changes during the vibration.

As opposed to FTIR and Raman, SFG is a nonlinear process that involves three photons: two incoming photons and one outgoing photon. The outgoing photon is at the sum frequency of the two input photons. Figure 1 shows the energy-level diagrams of the IR, Raman, and SFG processes. Clearly an SFG process is a combination of an IR process and an anti-Stokes Raman scattering. In a typical SFG experiment using a frequency-scanning

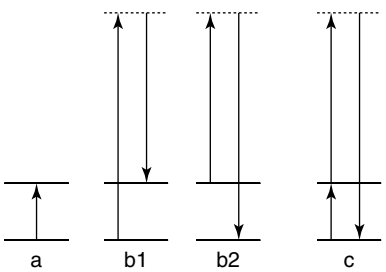

Figure 1 Energy-level diagram of FTIR (a), Stokes and anti-Stokes Raman scattering (b1 and b2), and SFG (c).

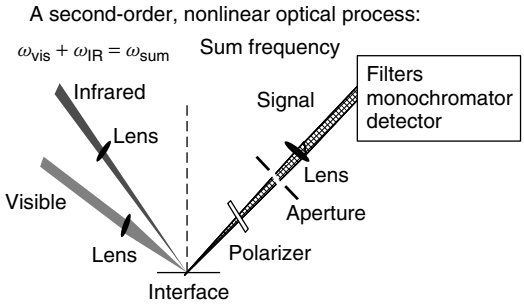

Figure 2 SFG experimental geometry.

system, two input beams, a frequency-fixed visible light and a frequency-tunable infrared light, shine on a surface/interface (Figure 2) and the sum frequency signal beam is detected. The intensity of the sum frequency signal beam is plotted as a function of the input IR wavenumber. When the frequency of the input infrared beam matches a vibrational resonance of molecules on the surface, the SFG signal is enhanced. Only if a vibrational mode is both IR and Raman active can it be observed in SFG experiments.

Under the electric dipole approximation, SFG signal intensity is proportional to the square of second-order nonlinear susceptibility $\chi^{(2)}$ of the material under study. This $\chi^{(2)}$ is a polar tensor, so it will change sign under the inversion operation: $\chi^{(2)}(r) = -\chi^{(2)}(-r)$. However, for a material with inversion symmetry, nothing will be changed under the inversion operation: $\chi^{(2)}(r) = \chi^{(2)}(-r)$. Clearly the only possible solution for the above two equations is $\chi^{(2)} = 0$. Therefore, for materials with inversion symmetry the SFG process is forbidden. Many bulk materials do have inversion symmetry and thus they do not generate SFG signals. However, inversion symmetry is inherently broken at an interface. Here, $\chi^{(2)}$ does not need to be zero, and SFG signals can be detected. Owing to this selection rule, SFG is intrinsically surface specific and has been demonstrated to have a submonolayer surface sensitivity.[10–15]

As mentioned, ATR-FTIR and SERS have been widely applied to study surfaces. However, the surface sensitivity of ATR-FTIR is limited by the penetration depth of the IR light, which is on the order of hundreds of nanometers to a few micrometers. Thus, its surface sensitivity is not comparable to SFG. SERS requires a metal substrate and thus is not as flexible as SFG, which can be used to study many different kinds of surfaces and interfaces. In addition, SERS does not have the submonolayer sensitivity and surface specificity of SFG.

Two different kinds of SFG systems, either a broadband system or a frequency-scanning system, are typically used to collect SFG spectra.[16] A broadband system involves a femtosecond laser, which has a broad frequency spectrum. Therefore, many vibrational modes are excited simultaneously. The signal is typically spectrally spread over a charge-coupled device (CCD) camera and the signals from different vibrational modes are collected simultaneously. Further details have been reported in previous publications and will not be repeated here.[16]

The authors' research group has two frequency-scanning systems.[17] Each system is composed of the following components: a picosecond Nd:YAG laser, a harmonic unit with two KD*P crystals, an optical parametric generation (OPG)/optical parametric amplification (OPA) and difference frequency generation (DFG) system based on Lithium Triborate LiB_3O_5 (LBO) and $AgGaS_2$ crystals, and a detection system. The green visible input beam (532 nm) for an SFG experiment is generated by frequency doubling the fundamental output pulses of 20 ps pulse from the Nd:YAG laser. The IR beam can be tuned from ~1000 to 4300 cm^{-1}, generated from the OPG/OPA and DFG system. By replacing the $AgGaS_2$ crystal with a GaSe crystal, this range may be extended down to 650 cm^{-1}. The diameters of both the visible and IR beams at the surface are about 500 μm. The SFG signal from the surface is collected by a gated photomultiplier tube. The pulse energies of the IR and visible beams are around 100 and 200 μJ, respectively, depending on the samples used in the experiment. SFG spectra with different polarization combinations such as ssp (s-polarized output, s-polarized visible input, and p-polarized IR input), ppp, and sps can be used to deduce orientation information of surface functional groups.

SFG spectra are usually collected from optically flat surfaces, as shown in Figure 2. In the past, most SFG studies have focused on the C–H, N–H, or O–H stretching frequency regions because of experimental equipment limitations. Thus, the data analysis for these functional groups has been extensively studied. Recently, SFG signals in lower frequency regions have been collected and analyzed. In this article, structural information of proteins and peptides at interfaces according to their SFG amide I signals will be extensively discussed.

In addition to identification of various functional groups' presence on the surface from SFG signals (e.g. from characteristic peak centers), orientation information of such surface functional groups can be deduced from SFG spectra collected in different polarization combinations. Details regarding the orientation analysis of such surface functional groups using SFG have been extensively discussed in many previous publications[18–24] and will not be repeated here. The general idea is similar to a polarized FTIR or ATR-FTIR experiment. In a polarized FTIR experiment, when a different s- or p-polarized beam is used, the transition dipole of a vibrational mode of the molecule is projected onto different axes and such projections measured from which the orientation can be deduced. For SFG, the molecular

hyperpolarizability is projected onto the lab coordinate system when a certain polarization combination of input and output laser is used in the experiment. For example, according to the intensity ratio of different vibrational modes of a functional group in the same spectrum, or more frequently intensity ratio of the same peak in SFG spectra collected using different polarization combinations, orientational information can be determined. In addition, more details regarding such an orientation (e.g. orientation distribution) can also be deduced if the SFG absolute intensity can be calibrated.[24]

SFG results (e.g. the presence and orientation of various functional groups at an interface) can lead to a detailed understanding of the interfacial structures and molecular interactions at the interface. For example, if the orientation of certain domains in a protein at an interface using SFG can be determined, it might be possible to deduce the orientation of the entire protein molecule at the interface. Such information is important to understand how the protein interacts with a biomedical implant, how a membrane protein functions, or how an enzyme immobilized on a surface could be used as a biosensor. In the following, several examples to demonstrate various SFG applications will be introduced.

3 SUM FREQUENCY GENERATION APPLICATIONS

After the demonstration of the feasibility of collecting SFG vibrational spectra from interfacial molecules in 1987 by Shen et al.,[25] SFG has been developed into a powerful surface-sensitive analytical technique. It has been widely applied to study many important surfaces and interfaces, including many liquid surfaces and interfaces, which will be discussed in detail in the next section. Such research is still continuing and many papers have been published; thus, it is impossible to even cite all the representative and important papers. It is suggested that interested readers refer to the excellent reviews on this research (e.g. in the works of Eisenthal, Chen et al., Bain, Richmond, Shen, and Gopalakrishnan et al.[10–15]) and those original publications cited in these reviews.

After discussing the SFG studies on liquid surfaces and interfaces in the next section, relatively recent SFG results on polymers and biological molecules at interfaces will be discussed. Similarly, extensive excellent research has been performed in SFG studies on polymers and biological molecules and it is impossible to cover the entire research field. Instead, this paper will mostly focus on the research results from the Chen lab, with which the authors are most familiar.

3.1 Small Molecules at Interfaces Involving Liquids

3.1.1 Water/Air Interfaces

Compared to the surface analytical techniques that require high vacuum to operate, one of the most important advantages of SFG is that it can be used to study liquid surfaces and interfaces. Water is often considered to be one of the most important liquids, and it is important to understand water structures at surfaces and interfaces.

In 1993, Shen and his colleagues published their SFG studies on water surfaces in air.[9] They assigned three major peaks in the SFG spectrum collected in the O–H stretching frequency range (Figure 3). The $\sim$3700 cm^{-1} peak is contributed by free O–H groups (or dangling O–H bonds). Such O–H groups stick out from water into the air; otherwise, they would form hydrogen bonds with other water molecules. The two other peaks centered at $\sim$3200 and $\sim$3400 cm^{-1}, respectively, are hydrogen-bonded O–H stretching signals. Shen et al. referred to these two peaks as ''ice like'' (for $\sim$3200 cm^{-1}) and ''liquid like'' (for $\sim$3400 cm^{-1}) peaks, because they are close in positions with the stretch modes of the bonded O–H in bulk ice and water. However, the assignment for the 3200 cm^{-1} peak is still under debate.[15] From the spectral fitting, it was found that the $\sim$3700 cm^{-1} peak has a different phase compared to the hydrogen-bonded O–H signals ($\sim$3200 and $\sim$3400 cm^{-1}), showing that the hydrogen-bonded O–H groups are facing toward the water bulk. From the calibrated absolute signal intensity, it was deduced that more than 20% of the water molecules at the water/air interface have dangling O–H bonds.

It was found that the dangling O–H signal gradually decreased as methanol was mixed into the bulk water, and completely disappeared as the concentration reached 11% by volume.[26] It was determined that there is a 0.25 monolayer of O–H dangling bonds at the water/air

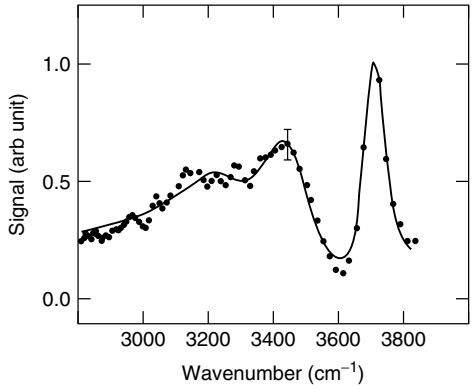

Figure 3 SFG spectrum collected from the pure water/vapor interface with ssp polarization combinations. Solid line is fitting result. (Reprinted figure with permission from Ref. 9. © 1993 by the American Physical Society.)

interface. When the dangling O–H bonds disappeared, the 3200 cm^{-1} peak dominates the spectrum, showing that the interfacial water molecules at the interface after introducing methanol are much more ordered compared to those at the water/air interface. SFG spectra were again collected from the water/air interface recently using different polarization combinations.[14] A new peak at ~3600 cm^{-1} was observed in the sps spectrum, which is assigned to the peak with bonded O–H stretch mode of surface water molecules with one bonded O–H and one dangling O–H. Temperature-dependent SFG signals of the water/vapor interfaces were collected and no strong temperate dependence was observed. The most noticeable change was an increase of the peak around 3400 cm^{-1} (disordered liquid-like peak).[14] This is different from some simulations that predict a strong temperature dependence for the surface structure of water.

3.1.2 Water/Hydrophobic Media Interfaces

Shen et al. also studied water structures at an interface between other hydrophobic surfaces and water, for example, the interface between a layer of octadecyltrichlorosilane (OTS) and water, and hexane and water. At the OTS/water interface,[26] free O–H signal has also been detected. The density of the O–H dangling bonds was measured and it was found that there is one dangling O–H bond in every four water molecules at the interface. This is similar to the water/air interface previously reported. Interestingly, comparing the SFG spectra collected from the OTS/water interface to the water/air interface, the 3200 cm^{-1} peak intensity from the OTS/water interface is much stronger. At a hydrophobic solid/water interface, the surface water molecules form a more ordered ice like structure because of the physical restriction involved in packing molecules against a solid. While at the hexane/water interface, the water spectrum was more similar to that collected from the water/air interface. In addition to the hydrophobic surfaces, water structure at the interface between a partially wettable surface and water was studied using SFG. In this case, the dangling water signal disappeared.[26]

To date, extensive research has been done on water surfaces and interfaces using SFG. Richmond et al. showed that water has a different structure at the CCl_4/water interface than at the water/air interface.[13] At the CCl_4/water interface, the hydrogen-bonded O–H species shift to higher frequencies, indicating significantly weaker hydrogen bonding at the CCl_4/water interface. These hydrogen-bonded water molecules at the CCl_4/water interface have weaker interactions with other water and CCl_4 molecules, leading to a frequency shift. However, the free O–H bonds peak at the CCl_4/water interface shift to 3660 cm^{-1} from the 3700 cm^{-1} at the water/air interface. Richmond's later research showed that this shift is the result of an attractive interaction between the dangling O–H bond and the surrounding CCl_4 molecules. Richmond also extensively investigated the effect of surfactant adsorption on water structure at interfaces. It was shown that the hydrogen bonding of water is highly sensitive to the presence of charged surfactants, with changes occurring in the water spectrum at trace concentrations in the aqueous solutions.[13]

3.1.3 Solution/Air Interfaces

Shultz et al. studied water surface structure at the sulfuric acid as well as sulfate solution surfaces. Sulfuric acid and sulfate aerosols are known to play a central role in atmospheric chemistry. Stratospheric sulfate aerosols are the dominant staging surfaces for reactions involved in ozone depletion. The determination of the distribution and orientation of water on the surface for 40–80 wt% ($0.1x$–$0.4x$) sulfuric acid would greatly help assess plausible mechanisms for stratospheric processes.[15] SFG spectra collected from the sulfuric acid solution in air indicated that the dangling water SFG signal was observed when a small amount of sulfuric acid was added to the water. However, when the solution contained $0.1x$ sulfuric acid, the free O–H signal was greatly reduced. No free OH signal was observed from the surface of solutions that contained a higher sulfuric concentration. The absence of this peak indicated a lack of dangling hydrogen atoms at the surface for stratospherically relevant concentrations. To determine whether the lower temperatures of the stratosphere result in greater partitioning of water to the surface, the SFG signal was examined as a function of temperature for a $0.2x$ solution.[27] Neither hydrogen-bonded water (discussed in the next section) nor free OH water was significantly altered for temperatures as low as 216 K. This showed that the surface structure was not altered at a lower temperature. Shultz et al. also studied surface structures of water at the water-sulfate solution surface. SFG spectra collected from the water/air interface were compared to those from the aqueous solutions of H_2SO_4, $LiHSO_4$, $CsHSO_4$, and $KHSO_4$ (keeping the anion concentration constant at $0.1x$). For H_2SO_4, the SFG intensity increased, especially for the 3150 cm^{-1} peak.

SFG signals from the sulfate solution surfaces showed a less dramatic but similar intensity increase. Schultz et al. believed that ''this intensity increase is attributed to an electric double layer that arises in the solution due to the differential distribution of anions and cations near the surface. Here the anion is larger and more polarizable than the cations. The hydration shell of the cations is therefore more strongly associated than that of the anion, limiting the approach of the cation toward the surface.

Similarly, the greater polarizability of the anions results in a lower enthalpic drive of anions away from the interface. As a result, the tail of the anion distribution extends further toward the interface than that of the cation. This differential distribution gives rise to a double layer just below the surface, and the negative pole of the double layer is directed at the surface.'' For smaller cations, for example H^+ and Li^+ studied here, along with HSO_4 create a stronger field, and therefore a greater enhancement of the 3150 cm^{-1} peak. For the larger cations, K^+ and Cs^+ at $0.01x$, it is easier for them to form ion pair with HSO_4^- and penetrate to the surface. Because of the presence of the ion pairs on the surface, the O–H dangling bonds are reduced, resulting in weaker free O–H signals.

Shen et al. also studied water structure at the sulfuric acid/water interface; details can be found in Raduge et al.[28] In addition to sulfate and bisulfate ions, water structure of halide ion solutions was investigated and compared to the theoretical simulations results. More details of the related research can be found in Gopalakrishnan et al.[15]

Cremer et al. studied how the presence of different Hofmeister anions in the solution altered the water structure next to a polymer.[29] Specifically, SFG spectra were collected from a solution surface with poly (*N*-isopropylacrylamide) (PNIPAM) on the surface (Figure 4). When different ions were added to the solution, the SFG signal from surface PNIPAM exhibited no change but the water signal changed substantially. The SFG signal contributed from O–H stretching was stronger for chaotropic anion solutions. The surface potential changes more for the chaotropic anion solution with PNIPAM on the surface – such results match SFG data quite well. Cremer et al. suggest that chaotropic anions segregate more easily to the solution surface to orient water to generate stronger SFG signals and induce bigger changes in surface potential. This is because ''adsorption behavior can be viewed as a partitioning of the anions between the interface and the bulk solution. The equilibrium constant and the bulk concentration dictate the distribution between these two states. For an anion to be adsorbed at the interface, it needs to undergo partial desolvation. The chaotropic anions have relatively low solvation free energies mostly because of their generally large size and high polarizability.[29] Therefore, a smaller desolvation penalty is paid compared with the kosmotropic anions.''

Ammonium is the only basic gas molecule that is important in environmental chemistry.[27] SFG has been applied to examine surface structures of ammonium aqueous solution. It was interesting to observe that the free O–H peak at ~3700 cm^{-1} actually contained two components at the ammonium solution/air interface (Figure 5). The peak at 3700 cm^{-1} became narrower when

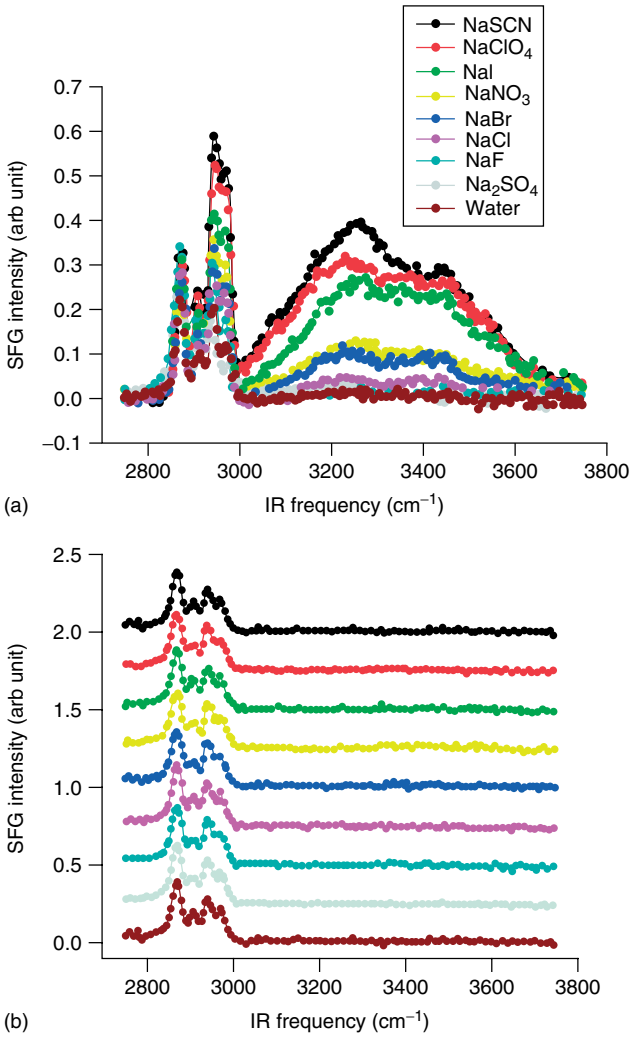

Figure 4 Vibrational Sum Frequency Spectroscopy (VSFS) spectra show specific anion effects on PNIPAM adsorbed at the air/water interface. Each subphase contained 1 M of a given salt as indicated in the legend except for NaF and Na_2SO_4, which are measured with saturated solutions (~0.8 M for both salts). (b) The same experiments repeated with D_2O. The spectra are offset for clarity. (Reprinted with permission from *Journal of American Chemical Society*. © 2007 American Chemical Society.)

as little as $0.005x$ NH_3 was added and disappeared with a concentration of $0.04x$. A broader peak extending about 100 cm^{-1} beginning at 3680 cm^{-1} was observed until $0.33x$ NH_3 was added. When looking at the N–H stretching at ~3310 cm^{-1}, below $0.10x$, the ammonium resonance increased as the free O–H resonance decreases. The observation indicated that at low concentration ammonium titrated the free O–H bonds, simply docking to the dangling hydrogen through the ammonium long pair. When half of the free OH groups reacted with the ammonium, the decrease of the OH peak stopped until the ammonium concentration was doubled ($>0.20x$).

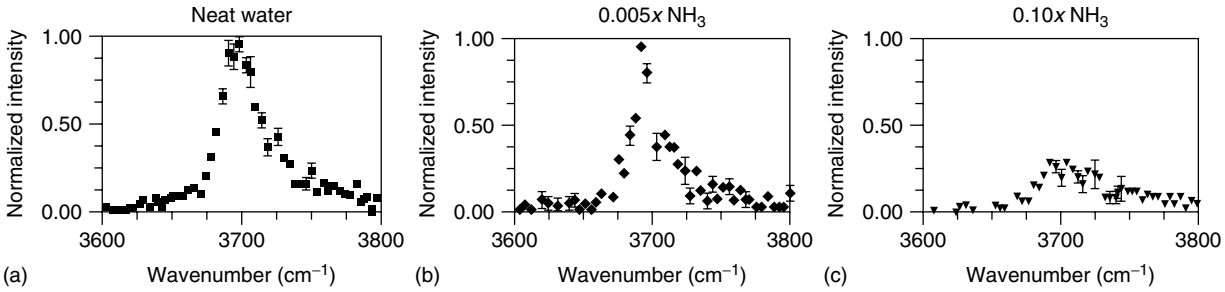

Figure 5 SFG spectra collected from ammonium solution surface: Free OH resonance of water (a) narrows and decreases with as little as $0.005x$ NH_3 (b) The broader resonance, shown for $0.10x$ NH_3 (c) persists to high concentration. (Reprinted with permission from *Journal of Physical Chemistry B*. © 2002 American Chemical Society.)

3.1.4 Water/Hydrophilic Solid Interfaces

Shen et al. studied the water structure at hydrophilic solid/water interfaces. For example, fused silica/water and crystalline quartz/water interfaces were investigated while varying the pH in water. Generally, the SFG spectra that were collected from the silica/water interface or quartz/water interface[26] were similar to those collected from the water surface in air, except there was no O–H dangling bonds signal. Both ice like and liquid like hydrogen-bonded OH signals were detected, indicating that the interfacial water molecules again form a partially ordered hydrogen-bonded network. Both the ~3200 and 3400 cm^{-1} peaks increase with pH, but the ice-like peak grew appreciably at high pH, suggesting a better polar-ordered network at high pH. To determine the absolute orientation of O–H groups at the quartz/water interface at different bulk pH of water, Shen et al. developed a phase-sensitive method from which it was feasible to deduce the O–H orientations.[14] The idea was to use the coherent property of the SFG technique to observe the interference between a bulk nonresonant SFG signal generated from crystalline quartz and the resonant SFG water signal. From such observed signal interference, it is feasible to deduce the phase information of SFG water signal; therefore, both real and imaginary components of the SFG signal from water can be deduced (Figure 6). It was found that the liquid like water maintained the same orientation when pH was varied, but the overall orientation of the ice like water flipped when the pH was increased. This technique of phase interference between a resonant signal of the analyte and a nonresonant background signal is a general method to measure the absolute orientation of interfacial molecules with SFG spectroscopy.

3.1.5 Charge Effect on Interfacial Water Structure

Richmond et al. studied the liquid/air interface of an aqueous solution with the surfactants sodium dodecyl sulfate (SDS), dodecylammonium chloride (dimethylammonium chloride, DAC), or a mixture of SDS and DAC (Figure 7).[13] SDS and DAC carry negative and positive charges, respectively. When the surfactants were added to water, they segregated to the surface, forming a negatively or positively charged surface, inducing strong water ordering near the surface. The SFG water signal became very strong owing to the strong ordering. When the neutral SDS and DAC mixture was added, no SFG water signal was observed.

Chen et al. collected SFG spectra from a protein solution surface in air (Figure 8). When the pH was varied, the SFG water signal had different intensities. When the pH was substantially higher or lower than the isoelectric point of the protein, the protein adsorbed at the solution/air interface carries negative or positive charges, which can orient the water molecules near the interface. This ordering generates much stronger SFG water signal than when the solution pH is close to the protein isoelectric point.[30] Clearly, interfacial charges greatly affect interfacial water ordering, and these changes can be detected using SFG.

3.2 Polymer Surfaces and Interfaces

The performance of many polymers is dominated by their surface properties, such as wettability, friction, lubricity, wearability, chemical reactivity, biological compatibility, biological activity, and permeability. These properties are widely used in many modern technological fields including biomaterials, microelectronics, display technology, coatings, and membranes for separations. Surface properties depend upon the chemical and physical details of the molecular structure at the polymer surface. In order to control surface properties by manipulating surface structure, we must have an extensive database of detailed correlations between properties and structures of the polymer surfaces.

SFG has been extensively applied to study molecular structures of surfaces and interfaces of a variety of polymer materials. In the following, after discussing

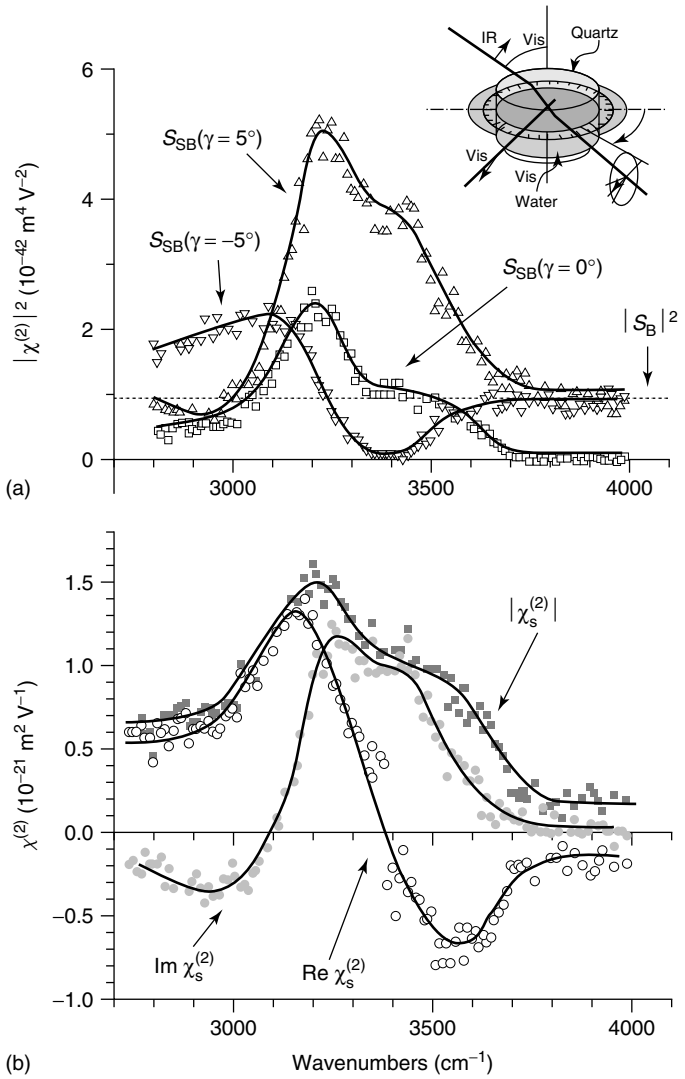

Figure 6 Real and imaginary SFG signals from water at quartz/water interface: An example of phase-sensitive SFG spectroscopic measurement for a quartz/water interface at pH of 6.5. (Reprinted figure with permission from Ostroverkhov, V., Waychunas, G. A., Shen, Y. R. *Phys. Rev. Lett.*, 2005, **94**, No. 046102. Copyright 2005, American Physical Society.)

how to quantify polymer surface structures using SFG, two particular SFG applications will be introduced: (i) polymer surfaces in water and (ii) structures of buried polymer interfaces related to polymer adhesion.

3.2.1 Quantitative Understanding of Polymer Surfaces

Polymers are large molecules and therefore polymer surface structures can be quite complicated. Before discussing how to quantify surface structures of polymer materials, a simpler surface, a surface of a self-assembled monolayer (SAM), for example, a methyl-terminated SAM on an Au surface will be discussed. Assuming the SAM surface has no defects, the surface is completely covered by terminal methyl groups. For an ideal sample, these methyl-terminated groups adopt the same orientation. Therefore, if the distribution of orientation angles of the methyl groups can be measured, it should be a delta distribution. If a surface is known to be completely covered by methyl groups, and if the orientation of all these methyl groups is deduced, the surface structure could be said to be understood quantitatively. Such quantitative structural information of a SAM surface can be obtained using SFG. SFG spectra collected from such a surface will show specific vibrational peaks for methyl groups, for example, 2875 cm^{-1} for symmetric C–H stretching, 2960 cm^{-1} for asymmetric stretching, and 2945 cm^{-1} for Fermi resonance. According to the intensity ratio of the signals detected in SFG spectra using different polarization combinations, the orientation of the methyl groups can be deduced. It is difficult to prepare a ''perfect'' SAM sample. The defects in SAM can change orientation

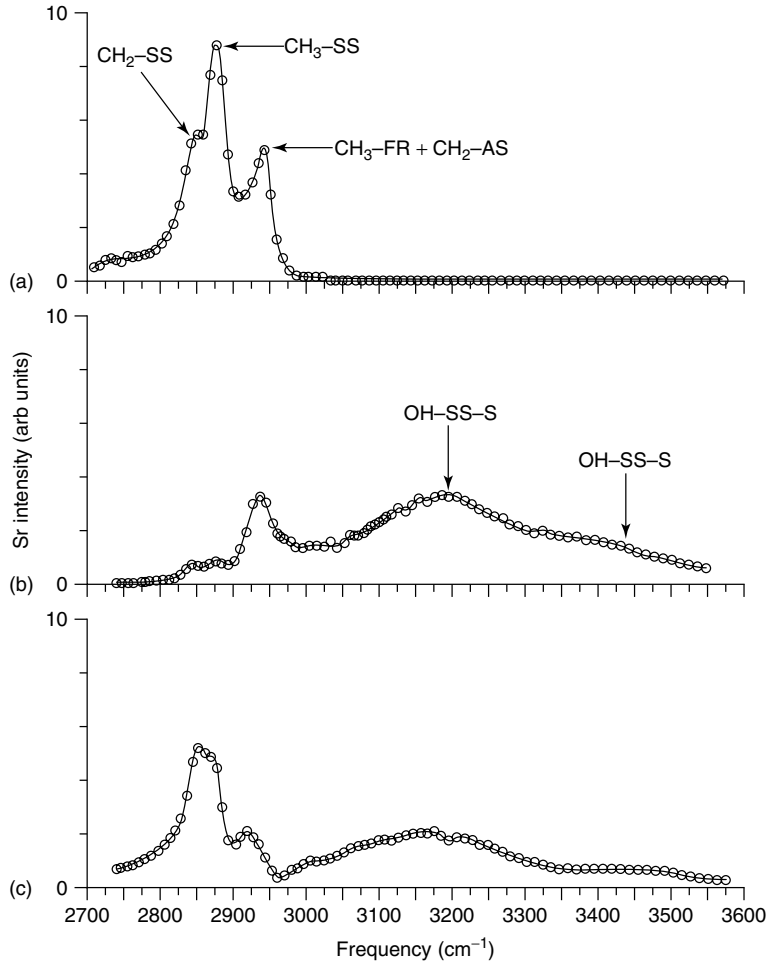

Figure 7 SFG spectra under ssp polarization conditions from the air–water interface of an aqueous solution of (a) a mixture of 0.02 mM dimethylammonium chloride (DAC) and 0.02 mM sodium dodecyl sulfate (SDS), (b) 0.05 mM SDS, and (c) 0.05 mM DAC. Solid curves are a guide to the eye. Symmetric Stretch (SS), Asymmetric Stretch (AS), Fermi Resonance (FR). (Reprinted with permission from *Journal of Physical Chemistry B*. © 1998 American Chemical Society.)

of some of the terminated methyl groups. The orientation distribution of the methyl groups cannot be described by a delta function then, but rather with a distribution with a certain width. If such an orientation distribution can be determined, then the surface with defects can be understood quantitatively.

A polymer surface may be more complicated and can have many different functional groups. The surface can also be quite disordered, meaning that the orientation distribution for each functional group can be very broad. A substantial number of different polymer surfaces have been investigated using SFG. It has been found that most polymer surfaces are dominated by one or two types of functional groups. If the surface-dominating functional groups can be identified, and if orientation distributions of these groups can be quantified, then the surface structure of these polymer materials can be quantified. Therefore, the quantification of polymer surface or interfacial structure can be realized by (i) identifying surface-dominating functional groups and (ii) deducing orientation distributions of such groups.

3.2.2 Sum Frequency Generation Studies on Polymer Surfaces in Water and Other Liquids

SFG has been applied to study surface structures of a variety of polymers in air. Examples of such polymers include polyethylene, polypropylene, various polymethacrylates, polystyrene (PS), poly (dimethyl siloxane) (PDMS), poly (ethylene glycol), poly (propylene glycol), polyimides, nylon, poly (ethylene terephthalate) (PET), and polyurethane.[11] SFG has also been used to study the surface structures of different polymer blends and copolymers. In addition, SFG has been applied to investigate structures of polymer surfaces in liquids and

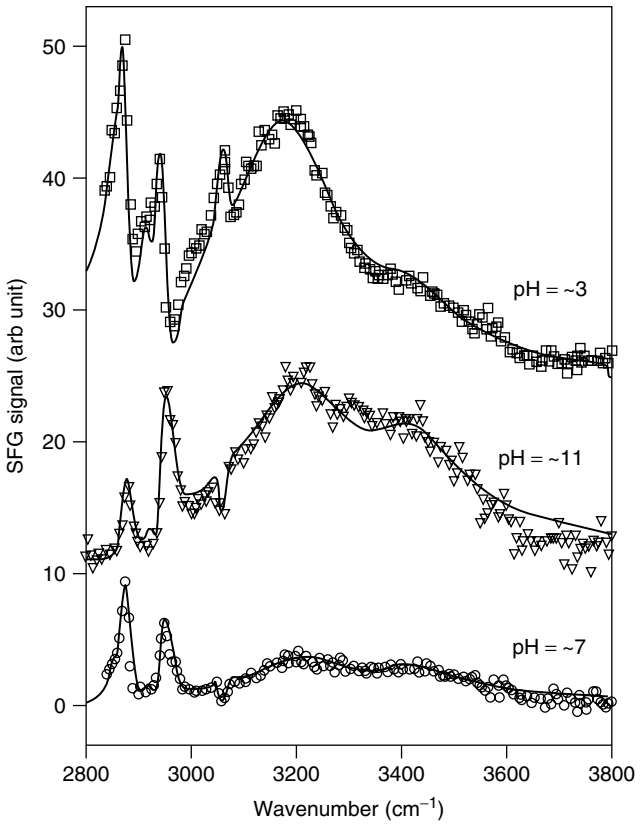

Figure 8 SFG spectra of bovine serum albumin (BSA) solution/air interfaces with different solution pH values. (Reproduced with permission from Ref. 30. Copyright 2002, American Chemical Society.)

buried polymer interfaces, which will be the focus of the discussion here.

Polymer materials are widely used in various liquid environments. For example, polymeric biomaterials are routinely implanted into human bodies. Antibiofouling polymer coatings for marine vessels are used in ocean. Surfaces of polymer membranes for biological molecule separation are exposed to biological molecule solutions when in use. To ensure the desired performance of these polymers, it is important to characterize the polymer surface structures in aqueous environments.

SFG has been developed into a powerful technique to characterize surface-restructuring behavior in liquid environments. SFG spectra can be directly collected from the polymer/liquid interface, probing polymer/liquid interface in situ in real time with a submonolayer sensitivity and surface specificity.

3.2.2.1 Polymer Side Chain Changes in Water – Polymethacrylates SFG spectra from poly (*n*-butyl methacrylate) (PBMA) surface in air, in water, and after removing the sample from water were collected using different polarization combinations (Figure 9).[24] The SFG spectra

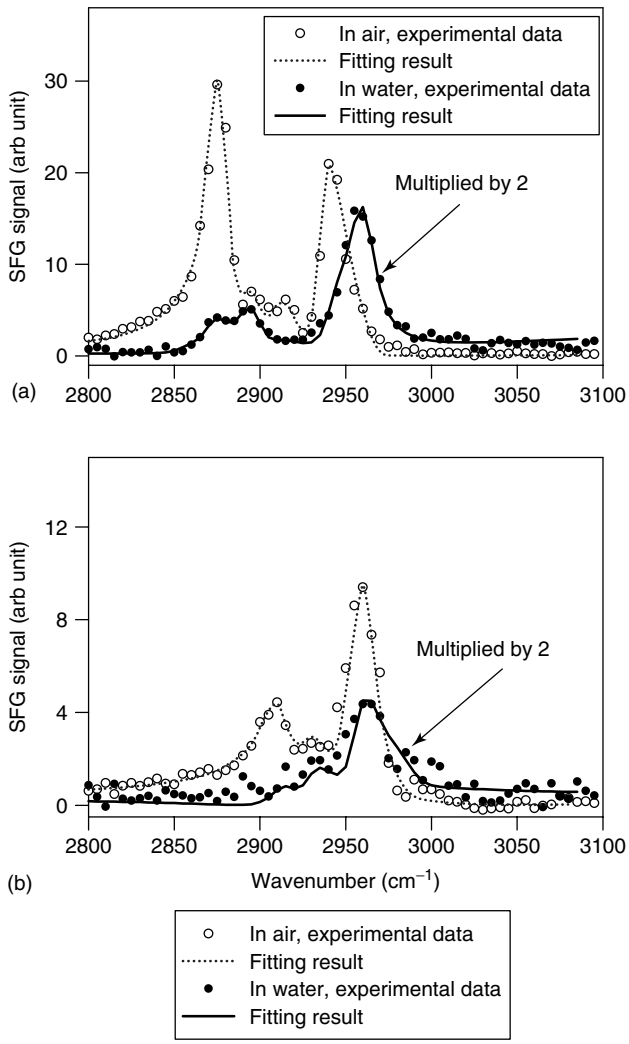

Figure 9 SFG spectra of PBMA in air and in water for (a) ssp and (b) sps polarization combinations. (Reproduced with permission from Ref. 24. Copyright 2002, American Chemical Society.)

collected from PBMA in air show different spectral features compared to those collected from PBMA in water, indicating that PBMA surface exhibits surface restructuring in water. The SFG spectra collected from the PBMA surface after removing it from the water and exposing to air again are similar to those collected before exposure to water, suggesting that the surface restructuring in water is reversible. The SFG spectra collected from PBMA in air and in water were both dominated by the signals contributed from methyl end groups of the ester chains. Comparing the symmetric and asymmetric methyl-stretching peak intensities in the spectra, it was determined that the methyl groups stand up more on the PBMA surface in air, and lie down more in water.

The SFG spectra collected using different polarization combinations can be used to deduce orientation

information of surface functional groups. The SFG polarized spectra measure components of second-order nonlinear optical susceptibility, which are related to different components of hyperpolarizability of the vibrational mode through $\langle \cos \theta \rangle$ and $\langle \cos^3 \theta \rangle$, where "$\langle\rangle$" means the average, and for methyl, θ is the angle between the principal axis of the methyl group and the surface normal. By measuring SFG spectra intensity in different polarization combinations, the ratio between $\langle \cos \theta \rangle$ and $\langle \cos^3 \theta \rangle$ can be determined. Assuming that every methyl has the same orientation – a delta orientation distribution – the orientation angle θ can be deduced. However, it is impossible to deduce $\langle \cos \theta \rangle$ and $\langle \cos^3 \theta \rangle$ individually with this ratio. In addition to collecting the SFG spectra using different polarization combinations, the absolute intensity can also be calibrated. With the help of absolute intensity, it is possible to deduce both $\langle \cos \theta \rangle$ and $\langle \cos^3 \theta \rangle$.

By assuming a Gaussian distribution for the orientation distribution of methyl groups, the orientation distributions of methyl groups on the PBMA surface in air and in water (Figure 10) can be determined. Figure 10 shows that methyl groups have a broader orientation distribution but a smaller average orientation angle in air. Methyl groups stand up more on the surface in air, and the surface is quite disordered (because methyl groups can have a broader range of orientation). In water, the orientation distribution is narrower, showing that the PBMA surface in water is more ordered. The average orientation angle of methyl on the PBMA surface in water is smaller, indicating that they lie down more toward the surface – perhaps due to the unfavorable interaction with water molecules.

To deduce the orientation distribution, it was assumed that the orientation distribution was a Gaussian function. To generalize this methodology for orientation distribution determination, a different trial function called the *maximum entropy function* was used.[24] Mathematically it has been proved that such a maximum entropy function leads to minimum bias under a limited amount of measurements. The orientation distribution deduced by using the maximum entropy function is similar to that of Gaussian function, showing that Gaussian function was a good approximation in this study.

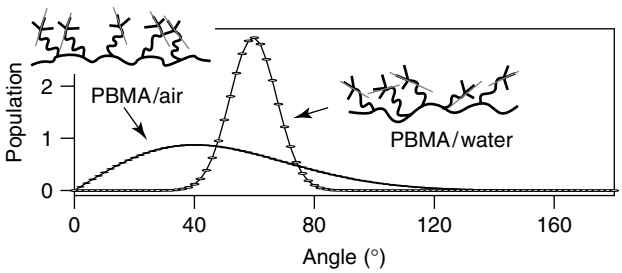

Figure 10 Orientation distribution of methyl groups on the PBMA surface in air and in water. The angle in the x axis is the angle between the methyl principal axis and the surface normal. In air, the methyl groups have a small average angle, showing that they like to stand up on the surface with a broader distribution. In water, they tend to lie down with a narrower distribution (Reproduced with permission from Ref. 24. Copyright 2002, American Chemical Society.)

In addition to the PBMA, other polymethacrylates with different side chain lengths in water were studied. Some of them exhibit similar surface-restructuring behaviors as PBMA, including poly (ethyl methacrylate) (PEMA), poly (propyl methacrylate) (PPMA), and poly (hexal methacrylate) (PHMA).[31] Similar to PBMA, the orientation distributions of surface-dominating end methyl groups on PPMA, PHMA, and poly (octadecyl methacrylate) (PODMA) surfaces were deduced. The methyl end groups of the side chain tend to stand up with a broader distribution in air and lie down more with a narrower orientation distribution in water. They have very similar average orientation angle or orientation distribution in air, but small difference of the methyl group orientation on the PPMA, PBMA, and PHMA surfaces in water was observed (Figure 11). With the increase of the side chain length, the orientation angle increased with a narrower orientation distribution.[31]

The surface structural changes in water for PODMA were different. The average orientation angle and orientation distribution of surface-dominating methyl groups vary only slightly. It was found that PODMA was semicrystalline and therefore the surface was quite rigid and only small orientation changes of surface methyl groups were detected. The research in this section demonstrates that it was feasible to study surface-restructuring behaviors of polymeric materials in water quantitatively.[24,31]

3.2.2.2 Polymer Side Chain Changes in Water – Poly (dimethyl siloxane) PDMS surfaces also exhibit side chain orientation changes in water. PDMS, a good candidate for marine antibiofouling coating, has been extensively studied since the early 1970s.[32–34] However, understanding of its molecular surface structure in the aqueous environment in situ has not been achieved. Different types of PDMS materials can have different antibiofouling properties, because of their different surface structures. To elucidate the biofouling control and release properties of PDMS and their relation to molecular surface structures, SFG was used to examine molecular surface structures in air and surface-restructuring behavior in water of three different types of PDMS materials.[35] They were tetraethoxysilane-cured hydroxy-terminated poly(dimethyl siloxane) (TEOS-PDMS), platinum-cured vinyl-terminated poly(dimethyl

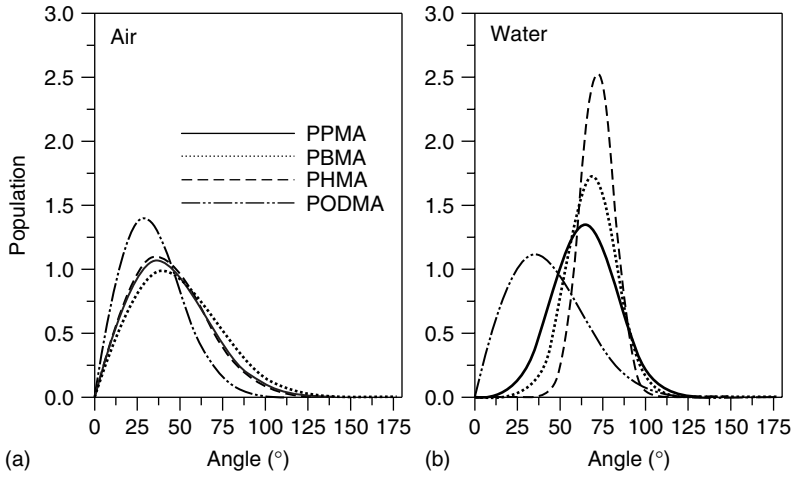

Figure 11 Gaussian orientation distributions for the side chain CH_3 groups on the surfaces of PPMA, PBMA, PHMA, and PODMA in (a) air and (b) in water (Reproduced with permission from Ref. 31. Copyright 2006, American Chemical Society.)

siloxane) (Pt-PDMS), and poly(dimethyl siloxane)-co-polystyrene (PDMS-*co*-PS) copolymer.

Using different polarization combinations of input and output light, SFG spectra were collected from the three PDMS surfaces in air as shown in Figure 12. All the SFG spectra were dominated by SFG signals contributed by the methyl groups on the PDMS side chains, showing that all surfaces are dominated by methyl groups. For Pt-PDMS, there are weak SFG signals due to the presence of $–Si–CH_2–CH_2$ moieties at cross-linking points on the surface. In the spectra collected from the PDMS-*co*-PS surface, there are several peaks above 3000 cm^{-1} that correspond to the aromatic C–H stretches from the phenyl group, indicating that on this surface, in addition to the methyl groups, phenyl groups also have some surface coverage. This research shows that small differences in surface structures of PDMS materials can be seen with SFG spectroscopy.

Surface-restructuring behaviors of these three PDMS materials in water were studied (Figure 13). For comparison, their surface changes in a hydrophobic fluorinated solvent, FC-75, was also studied. The SFG spectra collected from the three PDMS/water interfaces were very different from those of the PDMS/air interfaces. This would indicate that surface restructuring occurred in water. SFG data analysis indicated that when the hydrophobic PDMS surface contacted water the surface groups tended to reorganize by tilting more toward the surface to minimize interfacial free energy. At the PDMS-*co*-PS/water interface, the SFG signals due to methyl group stretches disappeared (Figure 13c) and only the signals due to the phenyl group stretches were visible. The disappearance of the methyl group stretches demonstrated that the methyl groups may reorganize differently at the PDMS-*co*-PS/water interface than at the other two

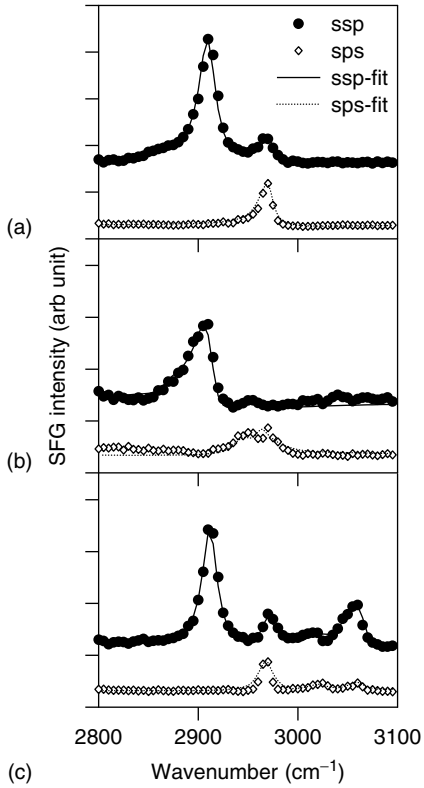

Figure 12 Measured and fitted SFG spectra (ssp and sps) collected from three PDMS surfaces in air: (a) TEOS-PDMS, (b) Pt-PDMS, and (c) PDMS-*co*-PS (Reproduced with permission from Ref. 35. Copyright 2004, American Chemical Society.)

PDMS/water interfaces. The similarity in the SFG signals due to the phenyl group stretches at the PDMS-*co*-PS/air, water, and FC-75 interfaces indicated that the phenyl groups do not experience restructuring when the film contacted different solvents.

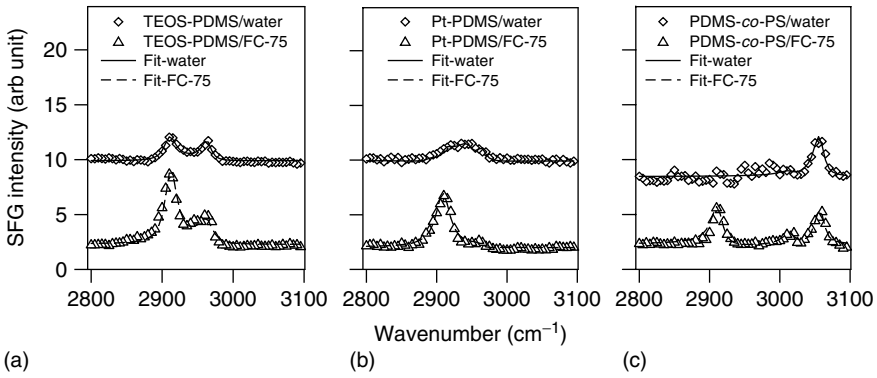

Figure 13 Measured and fitted SFG spectra (ssp) collected from three PDMS/water and PDMS/FC-75 interfaces: (a) TEOS-PDMS, (b) Pt-PDMS, and (c) PDMS-*co*-PS (Reproduced with permission from Ref. 35. Copyright 2004, American Chemical Society.)

Quantitatively, SFG results showed that the methyl groups at the Pt-PDMS/water interface tilt more toward the surface than those at the TEOS-PDMS/water interface. Unlike TEOS-PDMS or Pt-PDMS, the PDMS chains on the PDMS-*co*-PS surface are more mobile owing to the absence of cross-linked networks. Therefore, the molecular reorganization on the PDMS-*co*-PS/water interface became more complete and the interface was totally covered by the Si–O backbone, causing no SFG signals to be detected from the methyl group vibrational modes.

The SFG spectra collected from PDMS/FC-75 interfaces were more similar to the spectra from PDMS surfaces in air than those from PDMS/water interfaces. Detailed analysis of such SFG spectra showed clearly that in FC-75 methyl groups orient more toward the surface normal than those in air.

Through this study, the molecular-level chemical structures of several different PDMS surfaces in air and in liquid environments were studied using SFG. SFG can detect small differences in the surface structures of different PDMS surfaces, which cannot be differentiated by other techniques. Such molecular-level understanding of PDMS surface structures in various environments can provide detailed understanding of differences in antifouling properties of these PDMS materials. Previous research indicates that Pt-cured and alkoxysilane-cured PDMS have different stability in water and different antibiofouling performance.[36] Such differences should be caused by the different surface structures revealed by SFG. This research demonstrates that SFG is a powerful in situ technique to characterize surface structures of biofouling control/release polymer.

3.2.2.3 No Detectable Surface Restructuring in Water Other polymer surface-restructuring behaviors were also detected. The SFG spectra collected from the polymethylmethacrylate (PMMA) surface in water using different polarization combinations have similar spectral features to those collected in air, but weaker intensities.[37] Calculations indicated that the weaker intensities were due to the different refractive indices of water and air, rather than changes in the surface structures. PMMA surfaces were dominated by the ester methyl groups, and surface-dominating ester methyl groups did not exhibit detectable structural changes in water. This may be due to two reasons: (i) PMMA has a high glass transition temperature and therefore PMMA molecules including those on the surface are not very mobile and (ii) the ester methyl groups were more hydrophilic compared to the normal methyl groups. The ester methyl groups have a more favorable interaction with water than normal methyl groups.

3.2.2.4 Irreversible Polymer Surface Restructuring For methacrylates with longer side chains, such as poly (*n*-octyl methacrylate) (POMA) and poly (lauryl methacrylate) (PLMA), the SFG spectra collected in air indicated that their surfaces in air were still dominated by side chains.[31,37] However, substantial methylene signals can also be detected. Longer side chains have more gauche defects, generating stronger methylene signals. SFG signal was not detected from the polymer/water interface for POMA and PLMA, because the interfaces were totally disordered. The results showed that POMA and PLMA surfaces exhibit significant restructurings. The SFG spectra collected from the sample surface after removing the sample from water and exposing to air again did not completely recover. Both POMA and PLMA have very low glass transition temperatures (-20 and $-65\,^{\circ}C$ respectively), meaning the molecules can be very mobile and thus their surfaces exhibit substantial irreversible changes in water.

The SFG study of poly(ethyl acrylate) (PEA) showed that PEA had a similar surface-restructuring behavior as POMA and PLMA.[38] Like POMA and PLMA, PEA also has a very low glass transition temperature of $-21\,^{\circ}C$.

The SFG spectra were also collected from the surface of a PDMS sample that was not fully cured.[35] In air, the SFG spectra were similar to those collected from fully cured PDMS, indicating that the surface was dominated by $SiCH_3$ groups. No SFG signal was detected from the surface in water. The SFG spectra collected from the sample surface after removing the sample from water and exposing to air again did not completely recover. These observations differ from those of the fully cured PDMS materials, but were similar to those of POMA, PLMA, and PEA. It was suggested that this is due to the surface molecules of non fully cured PDMS being quite mobile. From the studies presented in this section, it is suggested that highly mobile molecules on the polymer surfaces in air can become totally disordered in water. After removing the polymer surface from water, the surface structure cannot recover.

3.2.2.5 Slow Surface Restructuring in Water For polymers discussed above, except for PMMA, which does not exhibit a detectable surface change in water, all other polymer surfaces exhibit restructuring behaviors immediately after contacting water. SFG research has shown that for some polymers surface restructuring can be a slow process. Somorjai et al. studied surface structural changes of a polyurethane with PDMS end groups in water.[39] This type of polymer has been used as a biomedical material. In air, because the PDMS silicone end groups are more hydrophobic, they cover the surface. When exposed to water, the more hydrophilic polyurethane backbone gradually segregates to the surface. SFG studies showed that after about 25 h the surface was dominated by the polyurethane backbone.

Chen et al. studied surface structural changes of a commercial polymer Sylgard 184 (unpublished results). As model PDMS surfaces that were discussed above, the Sylgard surface was also dominated by the $SiCH_3$ groups in air. Surface restructuring was detected after Sylgard surface was contacted with water. When the surface was removed from water and exposed to air again, the SFG spectra gradually recovered. The recovery process took tens of minutes.

3.2.2.6 Summary of Sum Frequency Generation Studies on Polymer Surfaces in Water Many important applications of polymers involve placing them in aqueous environments. Examples include biomedical polymer implants, marine antibiofouling coatings, polymer membrane for separation, and polymer coating for biosensors. Owing to the lack of applicable analytical techniques, some previous research studied polymer surface structures in air or in vacuum, and assumed that such surface structures are the same in water. It has been shown recently that SFG can probe molecular-level surface structures of polymers in situ in real time. SFG results indicated clearly that different polymer surfaces can exhibit varied restructuring behaviors. Some polymer surfaces showed reversible side chain orientation changes, others showed no surface change, and still others exhibited irreversible backbone changes. SFG results demonstrated that the assumption that polymer surface structures in air and in water are the same is not always a valid one.

Detailed surface-restructuring behaviors of various polymer surfaces in water revealed by SFG provide excellent experimental evidence for theoretical simulations. Chen et al. showed that molecular dynamics simulations of PS (using ethyl benzene as a PS model) surface in water (unpublished results) match SFG results quite well (unpublished results).

3.2.3 Sum Frequency Generation Studies on Buried Interfaces Involving Polymers to Understand Polymer Adhesion

3.2.3.1 Introduction Adhesion to polymer materials is important in many areas such as the automotive industry, aviation and aerospace, construction, electronics, industrial assembly and maintenance, paints and inks, and beauty and personal care.[40–43] For example, adhesives for polymers ensure the reliability and durability of coatings. Also, adhesives reduce the weight of a spacecraft by eliminating mechanical fasteners such as rivets. Similarly, from aircraft to automobiles to marine vessels, replacing rivets and bolts with adhesives will greatly reduce the vehicle weight and thus the amount of operation fuel, which is vital in energy conservation, environmental protection, and cost reduction. Adhesives also help distribute stresses over a wide area. Some other examples of the wide application of adhesion of polymer materials are in drug delivery, scar therapy, specialty cosmetic and hair replacement, and wound care products.[44] However, up to now the molecular-level understanding of adhesion mechanisms at polymer surfaces/interfaces has not been fully developed, which obstructs the systematic design of adhesives with the required performance.

For polymer adhesion applications, there is an increasing demand for silicone-based materials that can adhere to plastics, without surface priming or pretreatment.[40–42] Silicone elastomers are widely used because of their unique rheological properties, purity, and unsurpassed thermal stability and flexibility at a wide range of temperatures.[41] In particular, hydrosilylation- or addition-cured silicones offer advantages because of their clean cure chemistry and controllable and rapid cure kinetics. However, unlike condensation cure reactions, hydrosilylation-cured siloxanes are not naturally enriched

with polar, moisture-reactive groups and generally require the addition of adhesion-promoting additives.

Silane adhesion promoters are well suited for metals and other inorganic substrates that have reactive hydroxyl functional groups at their surfaces. Research also showed that alkoxysilane-based adhesion promoters could be quite effective in addition-curing silicones for imparting adhesion to some engineering thermoplastics, but the molecular mechanisms for such adhesion are not known.[45] Chen et al. studied the molecular structures of the silane molecules at buried polymer interfaces to understand the mechanisms of adhesion.

3.2.3.2 Molecular Structures of Silanes at Polymer/Silane Interface Chen et al. studied interactions at the interface between various polymers and different silanes. SFG spectra were directly collected from polymer/silane interfaces in situ.[46] The initially examined polymers include PMMA and PS. To avoid spectral confusion, deuterated PMMA (d-PMMA) and d-PS were used in the experiment. Chen et al. observed different structures of various silanes contacting PMMA or PS. The surfaces of both PMMA and PS are dominated by the ester methyl and phenyl groups, respectively.[17,47,48] The silanes that Chen et al. studied are *n*-octadecyltrichlorosilane (OTC, $CH_3(CH_2)_{17}SiCl_3$), and *n*-octadecyltrimethoxysilane (OTM, $CH_3(CH_2)_{17}Si(OCH_3)_3$).

The polymer/OTC and polymer/OTM interfacial SFG spectra did not change with time, indicating that these interfaces were stable. Different interfacial structures of silanes could be deduced by spectral analyses (Figure 14, see Chen et al.[46] for more details), and Chen et al. suggested that such structures were determined by the molecular interactions of the functional groups at these interfaces.[46] For example, the d-PMMA/OTC interface was dominated by the OTC methyl end groups as there were favorable interactions between these groups and the surface-dominating ester methyl groups of PMMA. At the PMMA/OTM interface, methoxy groups segregate to the interface that interacted more favorably with the PMMA ester methyl groups. At the PS/OTC interface, the OTC methylene backbones dominate the spectra. At the PS/OTM interface, both methoxy and methylene groups are detected. The restructuring of the ester methyl groups of PMMA and phenyl groups of PS at these different interfaces was also examined. This research clearly demonstrated that silane molecules adopt varied conformations at polymer/silane interfaces owing to different interfacial molecular interactions.[46]

3.2.3.3 Diffusion of Silanes into Polymers SFG was also applied to investigate structures of amino silane at the polymer/silane interface. Two amino silanes,

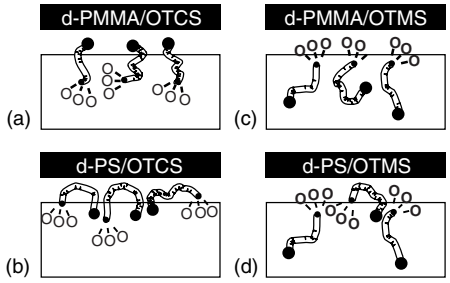

Figure 14 Schematic representations of different silane molecules oriented at the polymer/silane interfaces. The molecular segments (headgroup, backbone, and end group) of the silane molecules are represented as described by the legend. The boxes represent the bulk silane liquid. The SFG signals detected are primarily contributed from the molecular segments at the polymer/silane interfaces, that is, from those segments between the polymer substrate and the bulk. ●$-CH_3$, 〰$-CH_2-$backbone, • Si, ○ CL, o$-OCH_3$ (Reproduced with permission from Ref. 46. Copyright 2003, American Chemical Society.)

3-aminopropyltrimethoxysilane (ATM, $NH_2(CH_2)_3Si(OCH_3)_3$) and *N*-(2-aminoethyl)-3-aminopropyltrimethoxysilane (AATM, $NH_2(CH_2)_2NH(CH_2)_3Si(OCH_3)_3$) were studied.[46,49] The SFG spectra collected from the polymer/silane interfaces when either ATM or AATM were contacted at the PMMA or PS surfaces changed with time and finally decayed to zero. The signal disappearance at the PS interface is because ATM and AATM can dissolve PS. For PMMA, Chen et al. suggested that this phenomenon was due to strong interactions between the silane amino groups and the PMMA surfaces, resulting in the diffusion of the ATM and AATM molecules into the polymer. The diffusion of silane molecules was confirmed by a polymer bilayer experiment.[49] More specifically, the diffusion of AATM through the d-PMMA film was confirmed by SFG studies on the interface between AATM and a d-PMMA/PS two-polymer layer system. Initially the SFG signal from the PS layer was detected. However, after AATM diffused through the d-PMMA film, the PS film was dissolved by the silane, and thus the SFG signal from PS was lost. Chen et al. demonstrated that interdiffusion occurs much faster at the PMMA/ATM interface than at the PMMA/AATM interface. Using d-PMMA films of different thickness, Chen et al. showed that SFG could be used to monitor the chemical structures of moving interfaces during diffusion and to quantitatively deduce the diffusion coefficient (Figure 15). For AATM diffusion into PMMA, the diffusion coefficient was determined to be $(3.80 \pm 0.39) \times 10^{-13}\ cm^2\ s^{-1}$.[49]

SFG results also showed that (3-glycidoxypropyl)trimethoxysilane (γ-GPS, Figure 16) can diffuse through the d-PMMA film.[50] The diffusion of γ-GPS through the d-PMMA film was also confirmed by the polymer bilayer experiment. Similar experiments were carried out at the

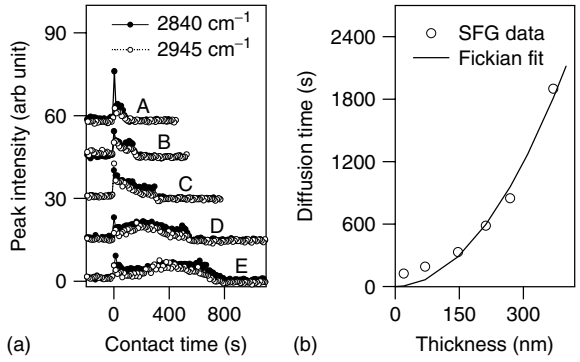

Figure 15 (a) Dependence of the SFG signal intensity at 2840 and 2945 cm^{-1} collected from the d-PMMA/AATM interface as a function of the d-PMMA film thickness. The time taken by the AATM molecules to diffuse through the entire film increases with film thickness. The SFG signal can be detected during the diffusion process, showing that the moving interface is ordered. (b) Fitting of the diffusion time versus the d-PMMA thickness curve via a Fickian Model allows the diffusion coefficient to be deduced (Reproduced with permission from Ref. 49. Copyright 2004, American Chemical Society.)

interface between a silane adhesion–promoting mixture (SAPM) of γ-GPS and a methylvinylsiloxnanol (MVS, $n = 9$) (Figure 16) and the d-PMMA/PS two-polymer layer system. It was found that the diffusion time of the γ-GPS in the SAPM through the d-PMMA film was significantly longer.

3.2.3.4 Hydrogen Bond Formation at the Interface AATM and ATM did not diffuse into all polymer films. For example, stable SFG spectra were detected from PET/ATM interfaces (Figure 17).[51] The shift of the C=O stretching frequency at the PET/ATM interface compared to the PET/air interface indicated the formation of hydrogen bonds between the ester carbonyl groups on the polymer surface and the amino groups of ATM. Detailed analyses of the SFG signals in the C–H stretching region collected from the PET/ATM and PET/BTM interfaces revealed that the conformation of ATM molecules at the interface is different from that of n-butyltrimethoxysilane (BTM). Even though SFG signals of methoxy headgroups were observed from both PET/ATM and PET/BTM interfaces, using the phase analysis previously mentioned in this paper it was shown that the methoxy adopt different orientations at the PET/ATM and PET/BTM interfaces (Figure 18).[51]

3.2.3.5 Molecular Structures of Silane Mixture at Polymer/Silane Interface Chen et al. applied SFG to study interfaces between polymers such as PET and PS, and SAPM.[52] Adhesion tests at Dow Corning Corporation showed that a PDMS elastomer containing the mixture could effectively adhere to PET after curing at 150 °C. Incorporating the silane or the siloxane alone into the elastomer did not achieve good adhesion under the same cure conditions. At the PET/γ-GPS interface, weak methoxy signals were detected using SFG. It was found that the initial SFG signal collected from the PET/mixture interface was also weak; however, it grew with time indicating that the silane methoxy headgroups adopted a much different final orientation or orientational order at this interface than at the polymer/γ-GPS interface. This observation may be used to interpret the different adhesion behavior between the mixture and the single component. Also the SFG signals arising from the mixture were not the same at different polymer interfaces.[52]

3.2.3.6 Preliminary Results of Silanes at Silicone/Plastic Interface SFG spectra of the adhesion-promoting system at the buried silicone/PET interfaces were successfully detected.[53] When adhesion-promoting molecules such as γ-GPS or SAPM were incorporated into the silicone, the segregation of the silane methoxy headgroups were detected at the interface even when the bulk concentration of silane in the silicone is 1.5 wt%. Comparing SFG spectra from samples that were cured for different lengths of time suggests that there was interaction between the γ-GPS with the d_4-PET surface at the polymer/silicone elastomer interface. Diffusion of the silane into the polymer was discounted and the data suggest that both interfacial ordering and reactions may play roles in adhesion in this particular system. When other silanes were incorporated into the silicone, the methoxy headgroups were either not present or not ordered at the d_4-PET/elastomer interface.

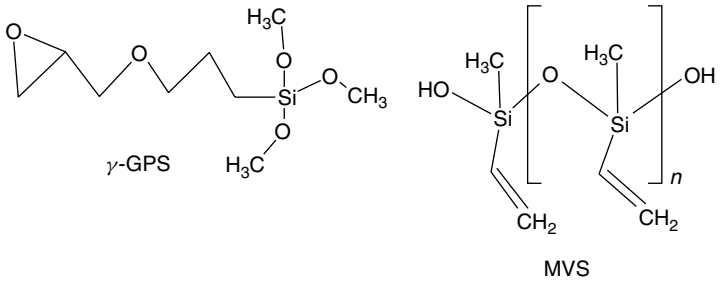

Figure 16 Molecular formulas for (3-glycidoxypropyl)trimethoxysilane (γ-GPS) and methylvinylsiloxnanol (MVS, $n = 9$).

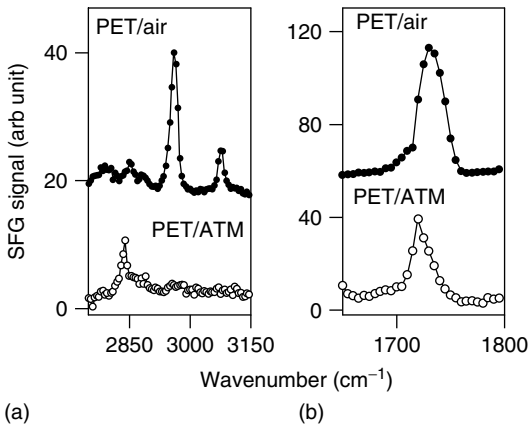

Figure 17 SFG spectra collected from the PET/air and PET/ATM interfaces in the C–H (a) and C=O (b) stretching regions. The shift of the C=O peak at the PET/ATM interface indicates the formation of interfacial hydrogen bonds (Reproduced with permission from Ref. 51. Copyright 2004, American Chemical Society.)

3.2.3.7 Summary From these SFG investigations on interfacial silanes, Chen et al. demonstrated that SFG is a powerful technique to elucidate polymer surface structures and polymer–silane interactions. It was observed that different silane structures at various interfaces were determined by the molecular interactions between the polymer surface functional groups and the functional groups that the silane were composed of, and certain silane molecules can diffuse into polymers. Chen et al. showed that hydrogen bonds formed between plastics and silanes, and mixing siloxane oligomers with silanes changed the silane structure at the plastic/silane interfaces.

3.3 Sum Frequency Generation Studies on Proteins and Peptides at Interfaces

3.3.1 Introduction

ATR-FTIR has been extensively applied to study the interfacial orientation of peptides and proteins through investigations on amide I signals. Amide I signals contain abundant information about protein backbone secondary structures. It has been shown that the amide I signal shifts frequency depending on secondary structure (*see* **Raman Spectroscopy in Analysis of Biomolecules**; **Vibrational Optical Activity of Pharmaceuticals and Biomolecules**; **Biomolecules Analysis: Introduction**; **Circular Dichroism in Analysis of Biomolecules**; **Two-dimensional Vibrational Correlation Spectroscopy in Biomedical Sciences**; **Infrared Spectroscopy of Biological Applications**).

Initially, most SFG studies on interfacial proteins were based on the C–H and N–H stretching ranges, which are mainly contributed by protein side chains, due to instrument limitations. More recently, SFG has been used to study interfacial protein backbone structure/conformation using amide I signals.

SFG has several distinct advantages over conventional ATR-FTIR. The intrinsic surface sensitivity of SFG ensures that SFG only probes proteins/peptides at the interface with no interference from signals contributed from proteins/peptides in the bulk environments. Unlike in FTIR, water-bending modes contribute almost no SFG signal; therefore, SFG amide I signals exhibit no water signal interference. It is necessary to subtract a very large background to obtain the FTIR spectra of interfacial proteins/peptides and this may introduce substantial errors. By collecting SFG spectra using different polarization combinations of these beams, much more structural information can be acquired. SFG is a coherent spectroscopy, and various vibrational modes can interact with each other. Thus, unlike FTIR, SFG spectral-fitting results can be tested experimentally.[54] Site-selective isotope labeling not only shifts the SFG peak of the site but also breaks the local symmetry to make the signal more prominent.

3.3.2 Advantages of Combining Multiple Vibrational Techniques in the Study

Different vibrational spectroscopic techniques measure different structural parameters.[55] For example, if the tilt angle of a structural element (e.g. an α-helix) is θ versus the surface normal, ATR-FTIR can determine the angle through measuring $\langle \cos^2 \theta \rangle$. For ATR-FTIR,

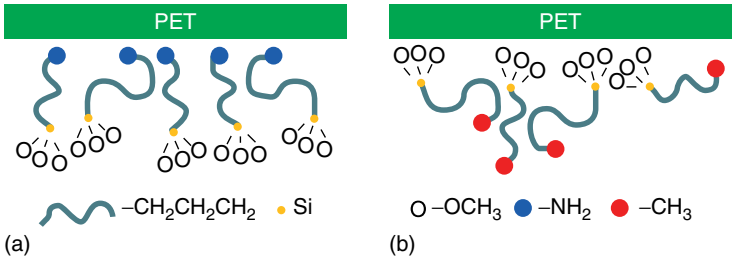

Figure 18 Schematics of (a) PET/ATM and (b) PET/BTM interfaces (Reproduced with permission from Ref. 51. Copyright 2004, American Chemical Society.)

this is the only measurement obtained and can only be used to determine one structural parameter. Therefore, to understand the orientation of the functional group, it must be assumed that all the groups at the interface adopt the same orientation or a δ-orientation distribution. In contrast, SFG can measure both $\langle \cos \theta \rangle$ and $\langle \cos^3 \theta \rangle$.[17,24] Thus, SFG can be used to deduce two different structural parameters, for example, orientation and distribution (assuming a Gaussian distribution function). However, if a chemical functional group adopts three different orientations (e.g. one protein has three α-helices, each with a different orientation in the membrane), then the two SFG measurements are not enough. By combining SFG and ATR-FTIR measurements, however, it should be possible to deduce such distributions. In the future, four-wave mixing (FWM) spectroscopy, which measures $\langle \cos^2 \theta \rangle$ and $\langle \cos^4 \theta \rangle$, could also be included in the study. FWM has the same selection rules as traditional Raman scattering, which are different from those of ATR-FTIR. Thus, FWM may measure $\langle \cos^2 \theta \rangle$ through different vibrational modes from ATR-FTIR. FWM may be used to confirm ATR-FTIR results. In addition, FWM can measure $\langle \cos^4 \theta \rangle$. More measurements can be used to deduce the orientation distribution more accurately, providing a more complete picture.

Figure 19 shows three cases for two α-helices in a protein: they have the same tilt angle, 30° versus the membrane normal, but different absolute orientations. ATR-FTIR can measure their tilt angle but cannot differentiate the three cases. SFG can differentiate these three cases: the top and bottom cases can generate strong SFG signals, while for the middle case no SFG signal can be observed owing to the inversion symmetry of the two helices. The top and bottom cases can be differentiated by collecting SFG spectra interfering with a nonresonant or resonant signal of a known phase, from which the absolute "up" and "down" orientations of the two helices can be deduced. From SFG alone, the orientation of the helices in the middle panel cannot be determined; the only information is that two helices have opposite orientation or they both lie down on the surface. If SFG measurements are combined with ATR-FTIR data, the orientation of the two helices in the middle panel can be determined: one tilts 30° up and the other 30° down. Therefore, the combined vibrational spectroscopic studies can lead to the determination of more structural information.

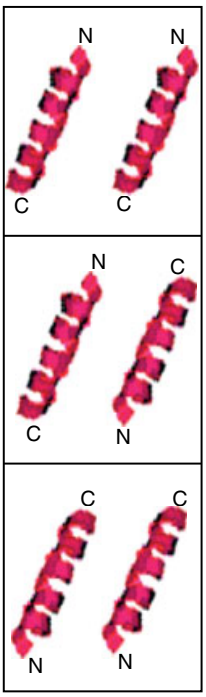

Figure 19 Combined SFG and ATR-FTIR studies can provide more detailed orientation.

When the protein crystal structure is known, it might be possible to deduce the membrane orientation of the entire protein from the orientation of the α-helical secondary structure, allowing a greater understanding of how membrane proteins function. When the crystal structure is not known, measuring orientations of secondary structure elements may help understand the overall structure. Assuming a protein contains several secondary structural motifs, if the orientation of each secondary structure can be measured, then the "general" or "rough" structure/conformation of this protein may be determined (Figure 20). More detailed structure can be achieved if structural information of connections (e.g. coils) between different secondary structural motifs can be further measured (e.g. by isotope labeling). As presented above, when a protein has multiple α-helices, it is difficult

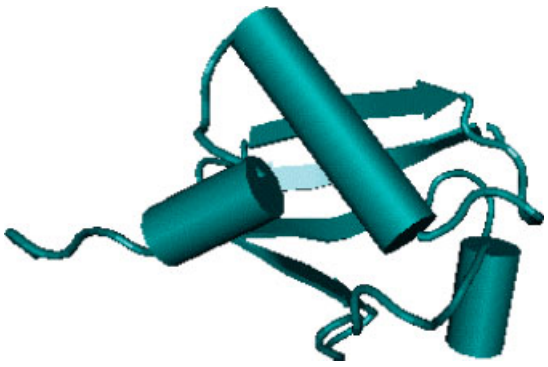

Figure 20 The model protein has three α-helices and several β-sheets. If the orientations of such secondary structures can be deduced, the rough structure or conformation of the protein can be known. A more detailed structure can be understood if the structural information of connections between different secondary structures is achieved.

to measure their orientations using ATR-FTIR alone. The combined SFG, ATR-FTIR, and FWM approach would generate more experimental measurements to deduce more detailed orientation distributions and more complicated structural information.

3.3.3 *Maximum Entropy and Quantitative Analysis of Complicated Orientation Function*

From a mathematical point of view, the task of reconstructing a distribution from its measured orientational parameters (e.g. $\langle \cos^n \theta \rangle$) is known as an *inverse problem*. It has been shown that the maximum entropy function, $G_{\text{total}}(\theta) = \exp\left(\sum_{n=0}^{N} a_n \cos^n \theta\right)$, is the best choice for a limited number of measurements, in the sense that it uses the fewest assumptions about the system.

To demonstrate the power of using multiple measurements, a few examples using artificial distributions will be discussed next. The first example is to assume that $G_{\text{total}}(\theta)$ is a Gaussian distribution with $\theta_0 = 85^\circ$ and $\sigma = 30^\circ$. This could be the distribution of α-helical peptides at an interface with varied orientation of the α-helix. The second function is a sum of two Gaussian distributions with $\theta_1 = 35^\circ$, $\sigma_1 = 10^\circ$, $\theta_2 = 135^\circ$, and $\sigma_2 = 10^\circ$. Such a function could occur from the two coiled-coils in an interfacial fibrinogen molecule. The third example assumes $G_{\text{total}}(\theta)$ to be a combination of four δ functions. A possible physical meaning for the third function could be a monolayer of ordered molecules on a surface. In each molecule, there are four identical chemical groups (e.g. α-helices), which cannot be separated in the frequency domain, but they have different orientations versus the surface normal. The functions $G_{\text{total}}(\theta)$ assumed in the above three examples have the same average angle (85°) but have totally different orientation distributions. With one measurement (e.g. ATR-FTIR alone) only the average orientation angle can be measured, and the differences among these three cases cannot be observed. One measurement is far from adequate to differentiate the functions in these three examples. As mentioned above, using various vibrational techniques, it is possible to measure the following parameters: $\langle \cos \theta \rangle$, $\langle \cos^2 \theta \rangle$, $\langle \cos^3 \theta \rangle$, and $\langle \cos^4 \theta \rangle$.

The deduced functions for the above three hypothetical examples using different numbers of measurements are shown in Figure 21.[55] From this figure, the deduced functions from one measurement (e.g. ATR-FTIR measurement) do not match any of the three real distribution functions. However, with more independent parameters deduced from other experiments, the deduced function can be much closer to the real function. With more measurements, even the case of a complicated distribution with four peaks can be resolved. Certainly, if there is preliminary information regarding these four peaks, for example, if it is known that there are four peaks and each one is a delta function, then only four independent measurements are needed to deduce such a four-peak distribution. This can be the case for membrane proteins: they may more or less have the same orientation. Thus, it is reasonable to assume the distribution of each type of α-helical structure to be a δ distribution. In addition to the application of combined vibrational spectroscopy, isotope labeling and controlled perturbation of interfacial peptides/proteins can generate additional independent measurements, providing more orientational parameters to deduce complicated orientation distribution.

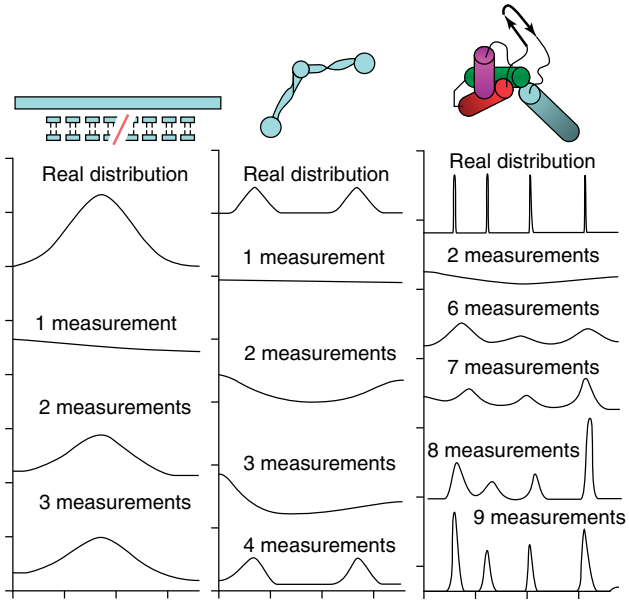

Figure 21 Assumed orientation distributions and deduced distributions with various numbers of hypothetical measurements. The y axis is the normalized population and x-axis is the orientation angle. (a) At least two measurements are needed to closely resemble the assumed Gaussian distribution; (b) at least four measurements are needed to closely resemble the assumed distribution of the two-Gaussian function; and (c) at least eight measurements are needed to closely resemble the assumed distribution of a four-δ function.

3.3.4 *Sum Frequency Generation Applications*

3.3.4.1 Feasibility of Detecting Sum Frequency Generation Amide I Signals Chen et al. demonstrated that it was feasible to collect SFG amide signals from protein molecules at interfaces in situ, allowing for the determination of backbone conformations of interfacial proteins using SFG.[56] To ensure that SFG can detect amide I signals from interfacial proteins, a near-total-reflection sample geometry was adopted (Figure 22). It was clear that the SFG amide I bands from the various proteins were quite different, indicating that secondary structures of various interfacial proteins were different.[56]

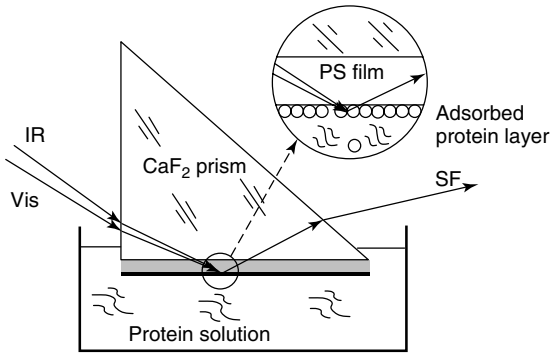

Figure 22 Schematic of the near-total-reflection SFG experimental geometry (Reproduced with permission from Ref. 56. Copyright 2003, American Chemical Society.)

3.3.4.2 Feasibility of Distinguishing Protein Secondary Structures Using Sum Frequency Generation Amide I Signals After demonstrating the feasibility of detecting SFG amide I signals, Chen et al. further showed that these signals can be used to distinguish various protein secondary structures[57] (*see* **Biomedical Spectroscopy: Introduction**).

α-helical Structure MSI-594 (Figure 23a and b) is an analog of the magainin 2 peptide. Figure 23(c) shows the SFG amide I signal collected from MSI-594 at the PS/solution interface. The main amide signal is centered at about 1650 cm^{-1}, showing that the α-helical structure of MSI-594 dominates.[57] The SFG spectrum can be fitted quite well using a single vibrational peak and a nonresonant background. For the fitted result, the peak center is at 1650 cm^{-1}. The SFG amide I signal collected from magainin 2 at the PS/magainin 2 solution interface is also centered at ~1650 cm^{-1}, indicating an α-helical structure (not shown). SFG amide signals from fibrinogen at various polymer/fibrinogen solution interfaces were detected.[58,59] The signals were also dominated by a peak centered at 1650 cm^{-1} owing to the contribution of two α-helical coiled-coils. In addition to the peak center, it was confirmed that this peak was dominated by α-helical contributions using SFG polarization analysis.[58,60]

β-sheet Structure Chen et al. demonstrated experimentally that SFG amide I signals could be detected from β-sheets.[57,61] Figure 24(b) shows the SFG amide I spectrum collected from tachyplesin I at the PS/solution interface and the fitting results. Tachyplesin I is known to have a β-sheet structure in aqueous environments and in membranes. The β-sheet structure of tachyplesin I is robust because such a structure is held by two disulfide bonds (Figure 24a). Therefore, Chen et al. believed that it should maintain a β-sheet structure at the PS/solution interface. The fitting results of Figure 24(b) showed three major peaks at 1645, 1664, and 1688 cm^{-1}, with several minor peaks at lower wavenumbers. Figure 24(c) displays the SFG spectrum collected from the PS/solution interface after the addition of dithiothreitol (DTT) to the solution. DTT reduced the two intramolecular disulfide bonds,[62] which are essential for tachyplesin I to maintain its β-sheet structure (Figure 24d). After the addition of DTT the 1688 cm^{-1} peak disappeared. As discussed in many published papers, amide I signals at 1688 and 1633 cm^{-1} can be ascribed to the B_1/B_3 and B_2 modes of antiparallel β-sheets, respectively.[63–65] The disappearance of the 1688 cm^{-1} peak upon the addition of DTT confirmed that this peak was generated by the β-sheet structure of tachyplesin I at the interface. Chen et al. further showed that SFG chiral spectra could be collected from interfacial tachyplesin, and such chiral

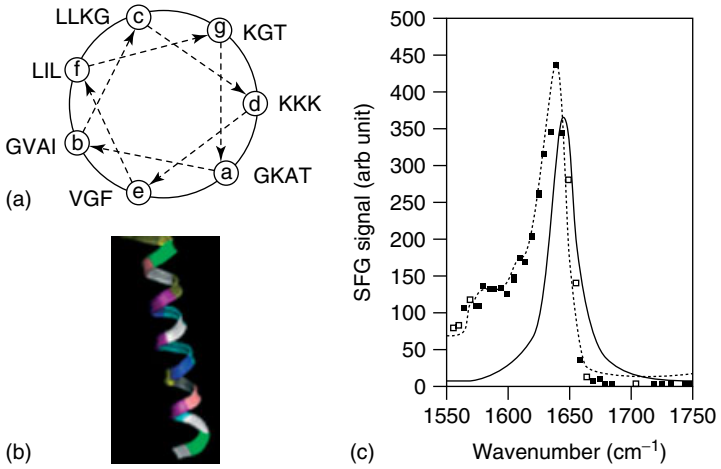

Figure 23 Diagram (a) and ribbon structure (b) of MSI594; (c) ssp SFG spectrum of MSI594 collected from the PS/MSI594 solution interface. Dots: experimental data; line: fitted spectrum. A nonresonant background is used in the fitting. The distorted spectral feature of the experimental data is caused by the nonresonant background.

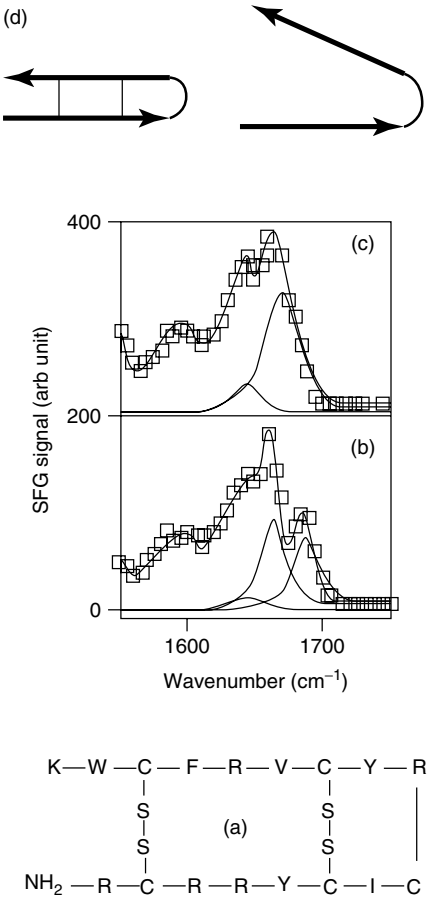

Figure 24 (a) Diagram of tachyplesin I with s–s representing the intrastrand disulfide bonds and (b, c) SFG spectra and fitting results for 0.1mg/mL peptide adsorbed at the PS/solution interface with squares representing the actual data, the line through the data is the fitted spectrum and the curves under the data are the component peaks used to fit the spectra. (b, c) SFG spectra of tachyplesin I before and after contacting 10 mM DTT. (d) Schematic showing the loss of β-sheet structure following addition of DTT.

SFG signals were dominated by contributions from β-sheet structure.[61]

β-turns and Random Structure After the DTT treatment for tachyplesin, the SFG bands at 1645 and 1664 cm^{-1} remained. Following extensive study, Chen et al. suggested that these two peaks are due to turns, random structures, or a combination thereof.[63,64] SFG polarization analysis showed that their behaviors were very different from those of α-helical structures.

3.3.4.3 Determination of α-helical structure orientation SFG has been widely used to determine orientation of functional groups on surfaces and at interfaces using polarization analysis, which has mostly been focused on C–H groups such as methyl and phenyl groups.[18–20] Chen et al. recently extended this analysis to analyze SFG amide I signals using a group theory projection operator method to correlate SFG polarization measurements to α-helical structure orientation.[23] There are two SFG-active amide I bands for α-helical structure, A and E_1 modes. Their SFG measurables χ_{xyz} are related to their hyperpolarizability tensor elements β_{abc} through orientation as follows:

A mode :

$$\chi_{A,xxz} = \chi_{A,yyz} = 1/2N_s[(1+r)\langle\cos\theta\rangle - (1-r)\langle\cos^3\theta\rangle]\beta_{ccc}$$

$$\chi_{A,xzx} = \chi_{A,yzy} = \chi_{A,zxx} = \chi_{A,zyy} = 1/2N_s[(1-r)(\langle\cos\theta\rangle - \langle\cos^3\theta\rangle)]\beta_{ccc}$$

$$\chi_{A,zzz} = N_s[r\langle\cos\theta\rangle + (1-r)\langle\cos^3\theta\rangle]\beta_{ccc}$$

where $r = \beta_{aac}/\beta_{ccc}$

E_1 mode :

$$\chi_{E,xxz} = \chi_{E,yyz} = -N_s[(\langle\cos\theta\rangle - \langle\cos^3\theta\rangle)]\beta_{aca}$$

$$\chi_{E,xzx} = \chi_{E,yzy} = \chi_{E,zxx} = \chi_{E,zyy} = N_s(\langle\cos^3\theta\rangle)\beta_{aca}$$

$$\chi_{E,zzz} = 2N_s(\langle\cos\theta\rangle - \langle\cos^3\theta\rangle)\beta_{aca}$$

Here N_s is the surface density of α-helical repeat units. Since the SFG hyperpolarizability can be deduced from Raman and IR properties of the α-helical molecule, Chen et al. deduced the relations among different hyperpolarizability tensor elements to be $r = 0.54$ and $\beta_{aca} = 0.32\beta_{ccc}$ from experimental measurements of polarized Raman and IR.[66–69] The detailed analysis indicated that only two SFG measurables regarding orientation angle θ were independent: $\langle\cos\theta\rangle$ and $\langle\cos^3\theta\rangle$. Using different polarization combinations of the input and output laser beams, $\langle\cos\theta\rangle$ and $\langle\cos^3\theta\rangle$ could be deduced. If it is assumed that all the α-helical structures on the surface/interface adopt the same orientation, then $\langle\cos\theta\rangle$ and $\langle\cos^3\theta\rangle$ can be replaced by $\cos\theta$ and $\cos^3\theta$. Subsequently, the relationship between an SFG measurable and the orientation angle of the α-helix can be depicted.

3.3.4.4 Orientation Distribution of α-helical Coiled-coils in Fibrinogen at Interfaces: Combined Sum Frequency Generation and Attenuated Total Reflection Fourier Transform Infrared Spectroscopy Studies It might not be accurate to assume all the α-helical structures on a surface/interface to have the same orientation. In these cases, a distribution function should be employed. As discussed above, lacking preliminary information regarding such an orientation distribution, the maximum

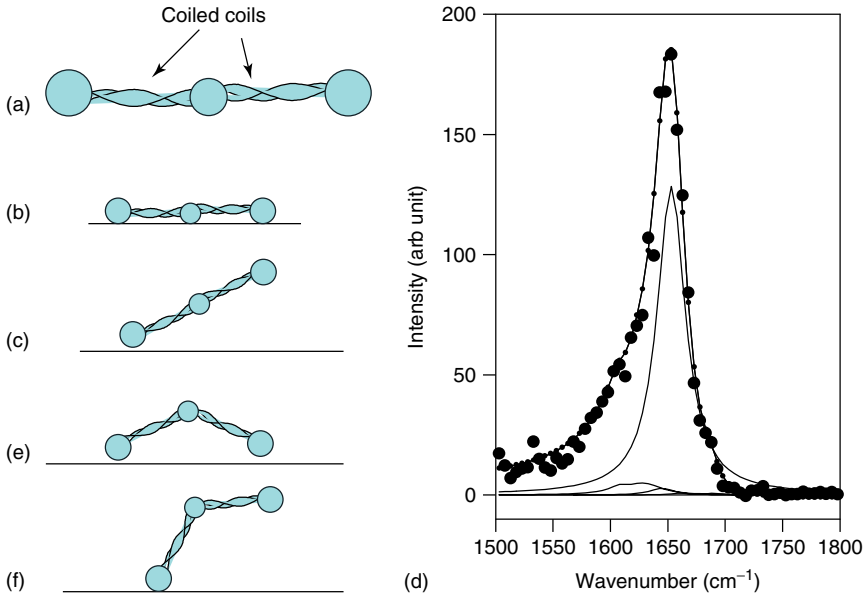

Figure 25 (a) Schematic of the fibrinogen native structure; (b, c, e, and f) possible conformation of fibrinogen after adsorption to PS; (d) ssp SFG spectrum collected from fibrinogen molecules at the PS/solution interface (Reproduced with permission from Ref. 58. Copyright 2006, American Chemical Society.)

entropy function is the best trial function. Chen et al. studied orientation distribution of two α-helical coiled-coils of fibrinogen at the PS/fibrinogen solution interface.

Figure 25(a) shows a schematic of the native fibrinogen structure, which (roughly) possesses inversion symmetry. Therefore, if fibrinogen adopted the native structure after adsorption to a surface (Figure 25b, c), little or no SFG signal would be detected. Experimentally, very strong SFG signal was detected from the PS/fibrinogen solution interface (Figure 25d), showing that adsorbed fibrinogen molecules do not adopt the native structure. The SFG signal was dominated by a peak centered at 1650 cm^{-1}, which was mainly contributed by the α-helical coiled-coils. SFG polarization analysis confirmed that this 1650 cm^{-1} peak was dominated by contributions from α-helical structures. Chen et al. also used the NLOPredict software package from the Simpson group (Purdue University) to determine that fibrinogen domains other than coiled-coils did not contribute noticeable SFG α-helix signals. The observed strong coiled-coil SFG signal indicated that fibrinogen adopted a bent structure after adsorption (Figure 25e, f).

Figure 26(a) shows the bent fibrinogen structure and orientation distribution of coiled-coils deduced by assuming that all the fibrinogen molecules have the same orientation (a δ distribution), using the SFG intensity ratio measured with different polarization combinations of the input/output laser beams. Certainly such a distribution is unlikely because the polymer surface can be very complicated and thus induce different orientations or conformations of the fibrinogen molecules adsorbed. Figure 26(b,

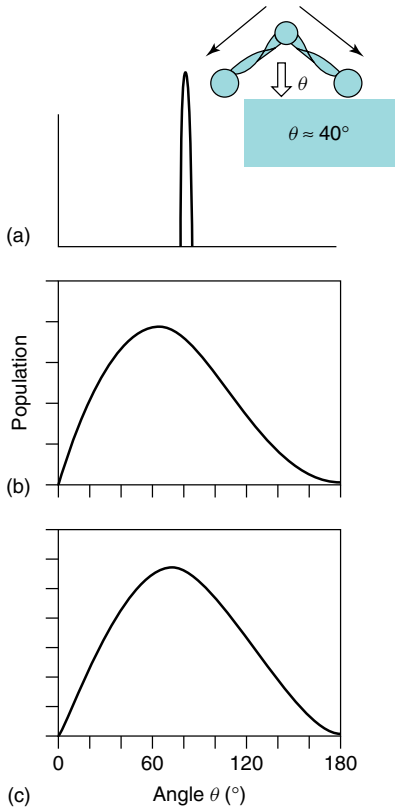

Figure 26 (a) Orientation distribution of α-helical coiled-coils in fibrinogen after adsorption deduced by one SFG measurement and schematic of the fibrinogen bent structure; (b, c) orientation distributions of coiled-coils deduced by two (b) and three (c) measurements. (Reproduced with permission from Ref. 55. Copyright 2007, American Chemical Society.)

c) shows the orientation distributions of α-helical coiled-coils of fibrinogen adsorbed on PS deduced using two SFG measurements (assuming a Gaussian distribution) and using combined SFG and ATR-FTIR (maximum entropy distribution) method, respectively.[55] The two distributions are similar, showing that the Gaussian distribution was a good assumption. However, the two distributions are not identical, showing that a combination of vibrational spectroscopic techniques should deduce the real distribution more accurately. The calculated distribution was quite broad, indicating that adsorbed fibrinogen molecules can adopt a variety of orientations and conformations. Using only ATR-FTIR to study such orientation, as is currently common, would reveal only one measured orientation angle, 52°, which is far from the real picture where coiled-coils adopted a very broad orientation distribution. This research demonstrated the importance of using multiple independent measurements to study orientation of α-helices.

3.3.4.5 Studies of α-helical Structures in Lipid Bilayers Chen et al. also studied peptides in supported single lipid bilayers using SFG. It has been extensively shown that solid-supported lipid bilayers are valid models for cell membranes.[70] SFG was applied to investigate structures of such lipid bilayers.[71–73] For most of the research on supported lipid bilayers, glass, mica, and fused silica were used as solid substrates. Such surfaces can be negatively charged, which may induce interactions with peptides in the lipid bilayer, causing artifacts in the determination of peptide orientation in the bilayer. Chen et al. successfully constructed lipid bilayers on a variety of solid substrates such as SiO_2, CaF_2, and various polymers. It was found that the structure and stability of lipid bilayers on these substrates are identical to those on glass, mica, or fused silica.

Melittin in Membrane It has been shown that melittin adopts α-helical structure in the lipid bilayer by a variety of analytical methods, but the orientation of melittin in a lipid bilayer is still controversial.[70,74,75] Chen et al. first collected polarized ATR-FTIR spectra from melittin in a substrate-supported single lipid bilayer. This generates only one measured parameter: the intensity ratio of the s- and p-polarized ATR-FTIR absorbance. According to the ATR-FTIR results, the average orientation angle is 53.8° versus the surface normal.[76] The SFG spectra from melittin in a lipid bilayer (Figure 27) were then collected. Both spectra shown in Figure 27 can be exclusively fitted by a single peak centered at 1650 cm^{-1}, indicating α-helical structure. Using the SFG intensity ratio-orientation results, Chen et al. found that the SFG experimental data cannot be fitted using a δ-distribution or a Gaussian distribution (Figure 28).

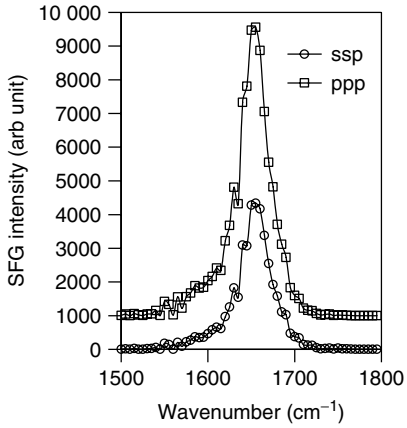

Figure 27 SFG amide I spectra collected from melittin adsorbed onto a 1,2-Dipalmitoyl-*sn*-Glycero-3-[Phospho-*rac*-(1-glycerol)] (DPPG)-supported bilayer at ssp and ppp polarization combinations. A solution concentration of 0.78 mM is used (arbitrary unit). (Reproduced with permission from Ref. 76. Copyright 2007, American Chemical Society.)

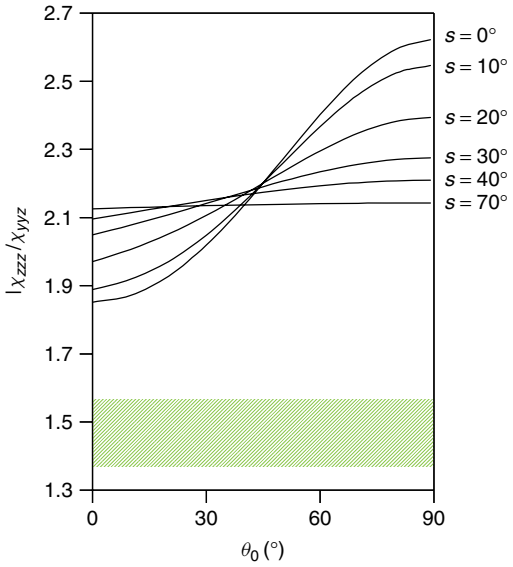

Figure 28 $|\chi_{zzz}/\chi_{yyz}|$ ratio as a function of orientation angle, assuming a δ-distribution or Gaussian distribution function. Shaded area represents the actual experimental result. (Reproduced with permission from Ref. 76. Copyright 2007, American Chemical Society.)

If it is assumed that melittin can adopt two orientations, then combined SFG and ATR-FTIR results showed that the two orientation angles were $\theta_1 = 13°$ and $\theta_2 = 100°$ and coverages were $N_1 = 0.274$ and $N_2 = 0.726$, respectively (Figure 29a). These results were quite different from previous ATR-FTIR measurements, which suggested that these peptides were either perpendicular or parallel to the membrane surface. Chen et al. suggest that both orientations coexist.[76] Such a result should be

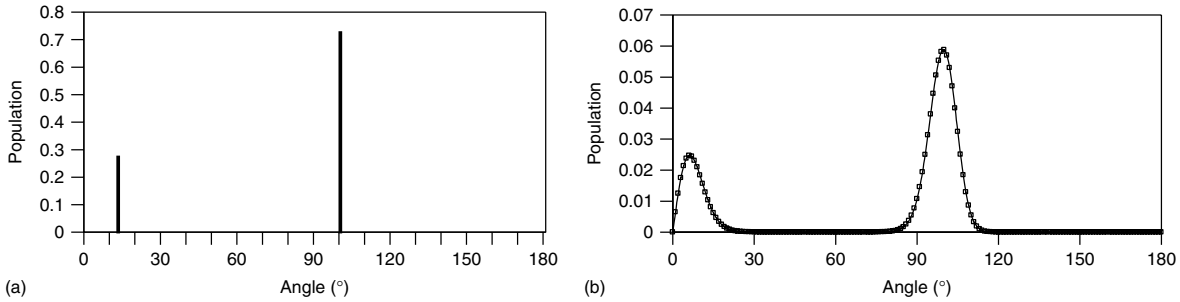

Figure 29 (a) Orientation distribution functions derived on the basis of a dual δ-distribution and (b) orientation distribution function derived on the basis of the maximum entropy theory (Reproduced with permission from Ref. 76. Copyright 2007, American Chemical Society.)

more reliable, because more measurements were used to derive the result.

Without assuming a two-δ distribution function as above, a similar result can be obtained by using a maximum entropy function as the trial orientation distribution. As mentioned, mathematically it can be shown that with limited measurements the maximum entropy function introduces the least bias. The deduced result using the maximum entropy trial function is shown in Figure 29(b). The distribution function in Figure 29(b) is similar to that shown in Figure 29(a) with two preferred orientations (Figure 30), but without initial assumptions.[76] This also suggests that for a protein with two (or even three) α-helical segments, the above method can be used to deduce their orientations.

G Protein Associated with Membrane Heterotrimeric guanine nucleotide-binding proteins (G proteins) are involved in numerous signal transduction pathways,[77,78]

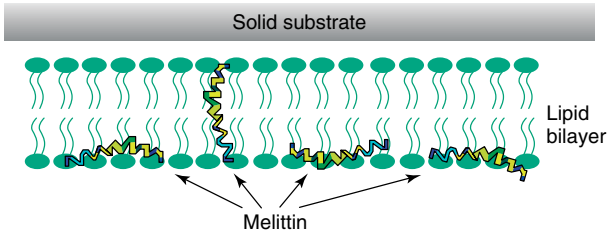

Figure 30 Schematic of two orientations of melittin in the lipid bilayer (Reproduced with permission from Ref. 76. Copyright 2007, American Chemical Society.)

and relay the extracellular signals sensed by G protein–coupled receptors (GPCRs) to downstream effectors. A heterotrimeric G protein comprises three subunits, Gα, Gβ, and Gγ, with Gβ and Gγ forming a tightly associated dimer. Chen et al. studied the orientation of G$\beta\gamma$ in the lipid bilayer using SFG.[79] Two types of G$\beta\gamma$ unit were used in this research: one with

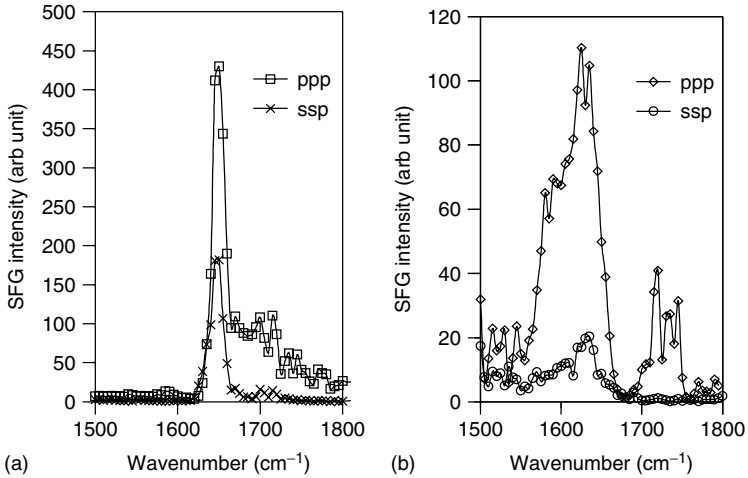

Figure 31 SFG ssp (weaker) and ppp (stronger) amide I spectra of interfacial G$\beta\gamma$ adsorbed onto a 1-Palmitoyl-2-Oleoyl-sn-Glycero-3-[Phospho-rac-(1-glycerol)] (POPG/POPG) bilayer. (a) 5 μL 10 mg mL^{-1} geranylgeranylated G$\beta\gamma$ was injected into the subphase and stronger SFG signals indicative of α-helix structure were observed. (b) 25 μL 10 mg mL^{-1} soluble form of G$\beta\gamma$ was brought into contact with the bilayer. Only relatively weak signals indicative of β-sheet were observed (Reproduced with permission from Ref. 79. Copyright 2007, American Chemical Society.)

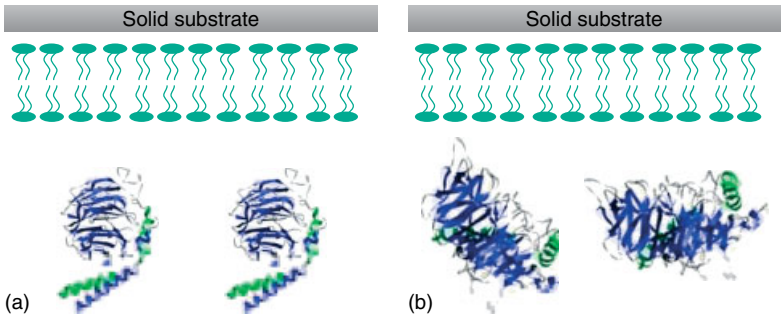

Figure 32 Schematic of G$\beta\gamma$ adsorbed onto a POPG/POPG bilayer deduced from the SFG spectra shown in Figure 31: (a) geranylgeranylated G$\beta\gamma$ and (b) G$\beta\gamma$ with no geranylgeranyl group. (Reproduced with permission from Ref. 79. Copyright 2007, American Chemical Society.)

geranylgeranyl anchor group and the other without. The SFG spectra collected from these two types of G$\beta\gamma$ subunits were very different (Figure 31). The SFG spectra collected from the G$\beta\gamma$ unit with the geranylgeranyl group associated with the lipid bilayer had much stronger signal intensities, even when a much lower protein concentration was used. The SFG signal was dominated by a peak at around 1650 cm^{-1}, characteristic of contributions from α-helical components. Chen et al. suggested that the geranylgeranyl anchor group was inserted into the lipid bilayer and thus G$\beta\gamma$ units associated with the lipid bilayer adopted the same orientation, as shown in Figure 32(a). The β-sheet segments in this orientation did not contribute to SFG signals substantially because they have centrosymmetry, but the α-helices generated strong SFG signals centered at 1650 cm^{-1} because of the ordered orientation. The SFG intensities in the spectra collected from the G$\beta\gamma$ unit without the geranylgeranyl group associated with lipid bilayer were much weaker even though a much higher protein solution concentration was used. Chen et al. suggested that there were no specific interactions between the G$\beta\gamma$ and the bilayer without the presence of the geranylgeranyl anchor. In this case, the SFG signal is dominated by the contributions from the β-sheet structure (Figure 32b), centered around 1630 cm^{-1}. This indicates that it is necessary to have the anchor group on G$\beta\gamma$ to ensure the specific orientation associated with the membrane.

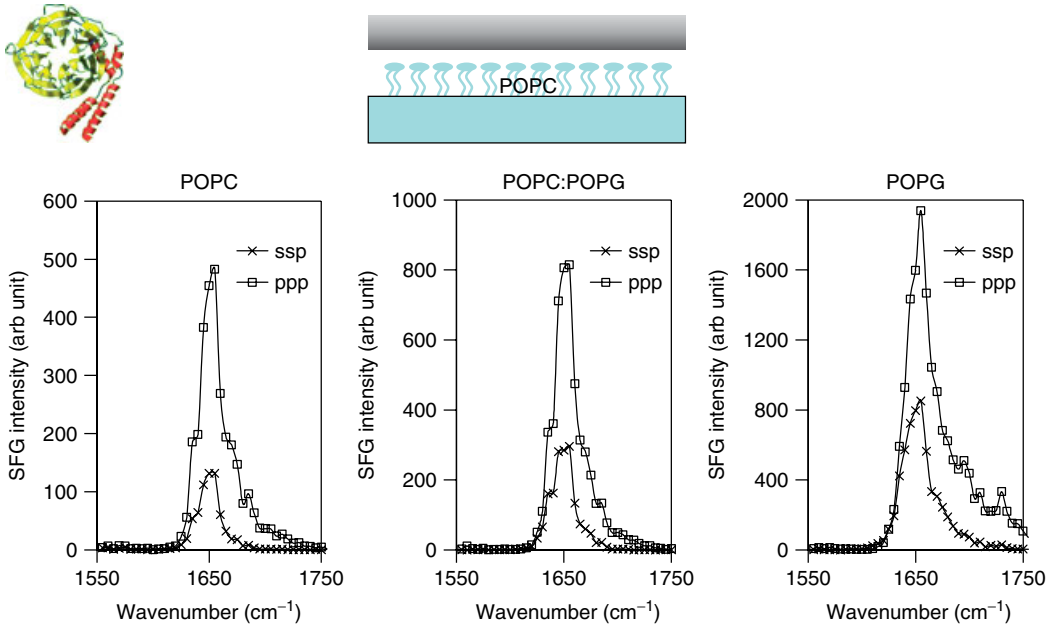

Figure 33 SFG spectra collected from G$\beta\gamma$ associated with different lipid bilayers: POPC:POPC; POPC:(POPC:POPG mixture); and POPC:POPG. The weaker signal in each panel is collected using ssp polarization combination and stronger one using ppp. Intensity ratios of ssp and ppp SFG signals are varied in different lipid bilayers, indicating that G$\beta\gamma$ adopts different orientations. The orientation of G$\beta\gamma$ in the POPC:POPC bilayer is shown in the upper left corner (Reproduced with permission from Ref. 79. Copyright 2007, American Chemical Society.)

The SFG spectra were also collected from $G\beta\gamma$ in lipid bilayers with different compositions (Figure 33): All three bilayers had an inner leaflet of 1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphocholine (POPC), while the outer leaflets that contacted $G\beta\gamma$ were varied as follows: POPG, POPG/POPC 1:1 mixture, and POPC. The SFG ppp/ssp spectral intensity ratios of $G\beta\gamma$ associated with various lipid bilayers were different, showing that $G\beta\gamma$ adopted different orientations owing to different lipid–protein interactions. The $G\beta\gamma$ orientation associated with the POPC:POPC bilayer was deduced and is shown in the upper left corner of Figure 33.

4 CONCLUSION

In conclusion, SFG vibrational spectroscopy is a powerful submonolayer-sensitive surface-specific analytical technique that provides a vibrational spectrum of surface functional groups on a centrosymmetric medium. SFG can provide both chemical specificity and orientation information. It has been applied to the study of a wide variety of systems, including complicated molecules such as polymers, proteins, peptides, and lipids at interfaces.

ACKNOWLEDGMENTS

The authors sincerely acknowledge the support from the Office of Naval Research (N00014-02-1-0832), National Science Foundation (CHE-0315857 and CHE-0449469), the Beckman Foundation, Dow Corning Corporation, and the University of Michigan for their support.

REFERENCES

1. D. Woodruff, T. Delchar, *Modern Techniques of Surface Science*, Cambridge University Press, Cambridge, 1986.
2. G. Somorjai, *Introduction to Surface Chemistry and Catalysis*, Wiley, New York, 1994.
3. K.B. Lewis, B.D. Ratner, *J. Colloid Interface Sci.*, **159**, 77–85 (1993).
4. E. Ruckenstein, S.V. Gourisankar, *J. Colloid Interface Sci.*, **109**, 557–566 (1986).
5. H. Yasuda, E.J. Charlson, E.M. Charlson, T. Yasuda, M. Miyama, T. Okuno, *Langmuir*, **7**, 2394–2400 (1991).
6. S.H. Cohen, M.L. Lightbody, *Atomic Force Microscopy/ Scanning Tunneling Microscopy 3*, Kluwer Academic/ Plenum Publishers, New York, 1999.
7. J. Drelich, K. Mittal, *Atomic Force Microscopy in Adhesion Studies*, VSP, Leiden, Boston, 2005.
8. Y. Shen, *The Principle of Nonlinear Optics*, Wiley, New York, 1984.
9. Q. Du, R. Superfine, E. Freysz, Y.R. Shen, *Phys. Rev. Lett.*, **70**, 2313–2316 (1993).
10. K.B. Eisenthal, *Chem. Rev.*, **96**, 1343–1360 (1996).
11. Z. Chen, Y.R. Shen, G.A. Somorjai, *Annu. Rev. Phys. Chem.*, **53**, 437–465 (2002).
12. C.D. Bain, *J. Chem. Soc., Faraday Trans.*, **91**, 1281–1296 (1995).
13. G.L. Richmond, *Annu. Rev. Phys. Chem.*, **52**, 357–389 (2001).
14. Y.R. Shen, V. Ostroverkhov, *Chem. Rev.*, **106**, 1140–1154 (2006).
15. S. Gopalakrishnan, D.F. Liu, H.C. Allen, M. Kuo, M.J. Shultz, *Chem. Rev.*, **106**, 1155–1175 (2006).
16. J.P. Smith, V. Hinson-Smith, *Anal. Chem.*, **76**, 287A–290A (2004).
17. J. Wang, C.Y. Chen, S.M. Buck, Z. Chen, *J. Phys. Chem. B*, **105**, 12118–12125 (2001).
18. C. Hirose, H. Yamamoto, N. Akamatsu, K. Domen, *J. Phys. Chem.*, **97**, 10064–10069 (1993).
19. C. Hirose, N. Akamatsu, K. Domen, *J. Chem. Phys.*, **96**, 997–1004 (1992).
20. C. Hirose, N. Akamatsu, K. Domen, *Appl. Spectrosc.*, **46**, 1051–1072 (1992).
21. X. Zhuang, P.B. Miranda, D. Kim, Y.R. Shen, *Phys. Rev. B*, **59**, 12632–12640 (1999).
22. A.J. Moad, G.J. Simpson, *J. Phys. Chem. B*, **108**, 3548–3562 (2004).
23. S.H. Lee, J. Wang, S. Krimm, Z. Chen, *J. Phys. Chem. A*, **110**, 7035–7044 (2006).
24. J. Wang, Z. Paszti, M.A. Even, Z. Chen, *J. Am. Chem. Soc.*, **124**, 7016–7023 (2002).
25. J.H. Hunt, P. Guyotsionnest, Y.R. Shen, *Chem. Phys. Lett.*, **133**, 189–192 (1987).
26. Q. Du, E. Freysz, Y.R. Shen, *Science*, **264**, 826–828 (1994).
27. M.J. Shultz, S. Baldelli, C. Schnitzer, D. Simonelli, *J. Phys. Chem. B*, **106**, 5313–5324 (2002).
28. C. Raduge, V. Pflumio, Y.R. Shen, *Chem. Phys. Lett.*, **274**, 140–144 (1997).
29. X. Chen, T. Yang, S. Kataoka, P.S. Cremer, *J. Am. Chem. Soc.*, **129**, 12272–12279 (2007).
30. J. Wang, S.M. Buck, Z. Chen, *J. Phys. Chem. B*, **106**, 11666–11672 (2002).
31. M.L. Clarke, C.Y. Chen, J. Wang, Z. Chen, *Langmuir*, **22**, 8800–8806 (2006).
32. M.E. Callow, A.R. Jennings, A.B. Brennan, C.E. Seegert, A. Gibson, L. Wilson, A. Feinberg, R. Baney, J.A. Callow, *Biofouling*, **18**, 237–245 (2002).

33. J.D. Adkins, A.E. Mera, M.A. RoeShort, G.T. Pawlikowski, R.F. Brady, *Prog. Org. Coat.*, **29**, 1–5 (1996).
34. Y.J. Sun, S.L. Guo, G.C. Walker, C.J. Kavanagh, G.W. Swain, *Biofouling*, **20**, 279–289 (2004).
35. C.Y. Chen, J. Wang, Z. Chen, *Langmuir*, **20**, 10186–10193 (2004).
36. A.E. Mera, R.B. Fox, S. Bullock, G.W. Swain, M.P. Schultz, P. Gatenholm, K.J. Wynne, *Nav. Res. Rev.*, **49**, 4–8 (1997).
37. J. Wang, S.E. Woodcock, S.M. Buck, C.Y. Chen, Z. Chen, *J. Am. Chem. Soc.*, **123**, 9470–9471 (2001).
38. C.Y. Chen, M.L. Clarke, J. Wang, Z. Chen, *Phys. Chem. Chem. Phys.*, **7**, 2357–2363 (2005).
39. D. Zhang, R.S. Ward, Y.R. Shen, G.A. Somorjai, *J. Phys. Chem. B*, **101**, 9060–9064 (1997).
40. A. Kinloch, *Adhesion and Adhesives*, Chapman and Hall, London, 1987.
41. H.K.J. Charles, *Engineered Materials Handbook, Vol. 3, Adhesives and Sealants*, ASM International, 1990.
42. G. Kookootsedes, *Polymeric Materials for Electronics Packaging and Interconnection*, American Chemical Society, Washington, DC, 1989.
43. E.P. Plueddemann, *Silane Coupling Agents*, 2nd edition, Plenum Press, New York, 1991.
44. http://www.dowcorning.com/applications/search/products/default.aspx?WT.svl=2& bhcp=1. (2008).
45. K. Mine, M. Nishio, S. Sumimura, Heat curable organopolysiloxane compositions containing adhesion additives, USA Patent, No. 4,033,924, 1977.
46. C.Y. Chen, C.L. Loch, J. Wang, Z. Chen, *J. Phys. Chem. B*, **107**, 10440–10445 (2003).
47. K.A. Briggman, J.C. Stephenson, W.E. Wallace, L.J. Richter, *J. Phys. Chem. B*, **105**, 2785–2791 (2001).
48. K.S. Gautam, A.D. Schwab, A. Dhinojwala, D. Zhang, S.M. Dougal, M.S. Yeganeh, *Phys. Rev. Lett.*, **85**, 3854–3857 (2000).
49. C.Y. Chen, J. Wang, C.L. Loch, D. Ahn, Z. Chen, *J. Am. Chem. Soc.*, **126**, 1174–1179 (2004).
50. C.L. Loch, D.C. Ahn, A.V. Vazquez, Z. Chen, *J. Colloid Interface Sci.*, **308**, 170–175 (2007).
51. C.L. Loch, D. Ahn, C.Y. Chen, J. Wang, Z. Chen, *Langmuir*, **20**, 5467–5473 (2004).
52. C.L. Loch, D. Ahn, Z. Chen, *J. Phys. Chem. B*, **110**, 914–918 (2006).
53. C.L. Loch, PhD Dissertation, Developing a Molecular-Level Understanding of Adhesion Using Sum Frequency Generation Vibrational Spectroscopy, University of Michigan, Ann Arbor, 2006.
54. J. Wang, M.L. Clarke, Z. Chen, *Anal. Chem.*, **76**, 2159–2167 (2004).
55. J. Wang, Z. Paszti, M.L. Clarke, X.Y. Chen, Z. Chen, *J. Phys. Chem. B*, **111**, 6088–6095 (2007).
56. J. Wang, M.A. Even, X.Y. Chen, A.H. Schmaier, J.H. Waite, Z. Chen, *J. Am. Chem. Soc.*, **125**, 9914–9915 (2003).
57. X.Y. Chen, J. Wang, J.J. Sniadecki, M.A. Even, Z. Chen, *Langmuir*, **21**, 2662–2664 (2005).
58. J. Wang, X.Y. Chen, M.L. Clarke, Z. Chen, *J. Phys. Chem. B*, **110**, 5017–5024 (2006).
59. M.L. Clarke, J. Wang, Z. Chen, *J. Phys. Chem. B*, **109**, 22027–22035 (2005).
60. J. Wang, S.H. Lee, Z. Chen, *J. Phys. Chem. B*, **112**, 2281–2290 (2008).
61. J. Wang, X.Y. Chen, M.L. Clarke, Z. Chen, *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 4978–4983 (2005).
62. Y.J. Li, D.M. Rothwarf, H.A. Scheraga, *J. Am. Chem. Soc.*, **120**, 2668–2669 (1998).
63. S. Krimm, J. Bandekar, *Adv. Protein Chem.*, **38**, 181–364 (1986).
64. A. Barth, C. Zscherp, *Q. Rev. Biophys.*, **35**, 369–430 (2002).
65. E. Vass, M. Hollosi, F. Besson, R. Buchet, *Chem. Rev.*, **103**, 1917–1954 (2003).
66. S.H. Lee, S. Krimm, *Biopolymers*, **46**, 283–317 (1998).
67. S.H. Lee, S. Krimm, *J. Raman Spectrosc.*, **29**, 73–80 (1998).
68. D. Marsh, M. Muller, F.J. Schmitt, *Biophys. J.*, **78**, 2499–2510 (2000).
69. L. Rintoul, E.A. Carter, S.D. Stewart, P.M. Fredericks, *Biopolymers*, **57**, 19–28 (2000).
70. S. Frey, L.K. Tamm, *Biophys. J.*, **60**, 922–930 (1991).
71. J. Liu, J.C. Conboy, *J. Am. Chem. Soc.*, **126**, 8376–8377 (2004).
72. A.W. Doyle, J. Fick, M. Himmelhaus, W. Eck, I. Graziani, I. Prudovsky, M. Grunze, T. Maciag, D.J. Neivandt, *Langmuir*, **20**, 8961–8965 (2004).
73. X. Chen, M.L. Clarke, J. Wang, Z. Chen, *Int. J. Mod. Phys. B*, **19**, 691–713 (2005).
74. I. Constantinescu, M. Lafleur, *Biochim. Biophys. Acta*, **1667**, 26–37 (2004).
75. C. Steinem, H.J. Galla, A. Janshoff, *Phys. Chem. Chem. Phys.*, **2**, 4580–4585 (2000).
76. X.Y. Chen, J. Wang, A.P. Boughton, C.B. Kristalyn, Z. Chen, *J. Am. Chem. Soc.*, **129**, 1420–1427 (2007).
77. W.M. Oldham, H.E. Hamm, *Q. Rev. Biophys.*, **39**, 117–166 (2006).
78. S.R. Neves, P.T. Ram, R. Iyengar, *Science*, **296**, 1636–1639 (2002).
79. X. Chen, A.P. Boughton, J.J.G. Tesmer, Z. Chen, *J. Am. Chem. Soc.*, **129**, 12658–12659 (2007).

Terahertz Spectroscopy

Ingrid Wilke
Rensselaer Polytechnic Institute, Troy, NY, USA

1 Introduction 1
2 Specifications and Performance of Different Methods of THz Spectroscopy 2
2.1 Performance 2
2.2 Merits 3
3 Technical Descriptions 7
3.1 Far-infrared Fourier Transform Spectrometer 7
3.2 Backward-wave Oscillator Terahertz Spectrometer 7
3.3 Time-domain Terahertz Spectrometer 8
4 Overview of Applications of THz Spectroscopy 11
Acknowledgments 11
Related Articles 11
References 12

Terahertz spectroscopy has attracted great interest in recent years due to the development of improved terahertz radiation sources and terahertz radiation detectors as well as novel terahertz spectroscopy systems. This article discusses Fourier transform far-infrared spectroscopy, backward-wave oscillator terahertz spectroscopy, and time-domain terahertz spectroscopy. Terahertz spectroscopy is spectroscopy with electromagnetic radiation in the terahertz (THz) frequency range. THz radiation is invisible. The THz frequency range is located between microwaves and the infrared. THz spectroscopy measures the dielectric properties of liquid and solid as well as inorganic and organic materials in the THz frequency range. Furthermore, THz spectroscopy is a method for vibrational and rotational spectroscopy of gases.

Specifications and merits of Far-infrared Fourier transform spectroscopy, backward-wave oscillator terahertz spectroscopy, and time-domain terahertz spectroscopy for applications in analytical chemistry are compared. The frequency range, frequency resolution, and dynamic range of each method are discussed. Various measurement modalities such as transmission and reflection spectroscopy, phase-sensitive spectroscopy, and time-resolved THz spectroscopy are explained. The physical measurement principle of each method is discussed. The relationship between the measurement observed and the dielectric function of a material are presented for each method. Technical descriptions of the methods are provided. An overview on specific applications of THz spectroscopy is presented.

1 INTRODUCTION

Terahertz spectroscopy describes spectroscopy with electromagnetic radiation in the THz frequency range.[1] The THz frequency range is located between the microwave and the infrared frequency range (Figure 1).[2,3] Previously, this part of the electromagnetic wave spectrum was also referred to as the *submillimeter* or the *far-infrared* (FIR) frequency range. Correspondingly, submillimeter wave spectroscopy and FIR spectroscopy are synonyms of THz spectroscopy.[4–8] A frequency of 1 THz corresponds to a wavelength of 300 μm, a wavenumber of 33 cm^{-1}, or energy of 4.1 meV (Figure 2).

Terahertz spectroscopy measures the dielectric properties of materials. The method provides information on characteristic dynamical properties of materials in the gaseous, liquid, and solid state that occur on picoseconds to femtoseconds timescales. Terahertz spectroscopy is used to identify molecules in the gaseous state via characteristic rotational and vibrational absorption lines. Materials that form periodic crystals can be identified by THz spectroscopy through their phonon-absorption bands in the THz frequency range. Liquids and disordered solid materials generally show broad absorption bands at THz frequencies. However, in this case, it is possible to distinguish between different materials using THz spectroscopy based on dielectric contrast.

Terahertz spectroscopy has a long and successful tradition.[9] The first observation of the rotation spectrum of HCL was published in 1925. Since then, the development of THz spectroscopy and its applications in analytical chemistry has been driven by improvements in THz radiation sources, THz radiation detectors, and THz spectrometers. Accordingly, different approaches to THz spectroscopy have been developed. Earlier, THz spectroscopy was performed with grating spectrometers.[10] This approach was followed by two-beam interferometric spectrometers.[11–15] These types of THz spectrometers rely on thermal radiation sources and THz detectors such as Golay cells, bolometers, photoconductors, or pyroelectric devices.[16] The invention of coherent THz radiation sources such as backward-wave oscillators (BWOs), FIR molecular gas lasers, and THz quantum cascade lasers resulted in the development of scanning frequency THz spectrometers.[17–20] A recent development in THz spectroscopy is time-domain (TD) THz spectroscopy.[1] This method employs very short

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

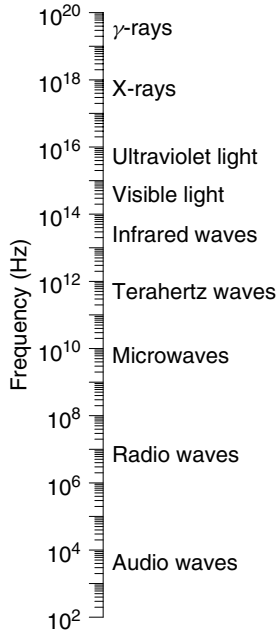

Figure 1 Electromagnetic frequency spectrum.

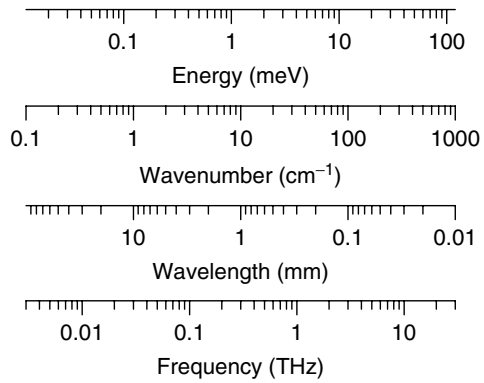

Figure 2 Terahertz frequency range and corresponding wavelength, wavenumber, and energy ranges.

pulses ($\leq 10^{-12}$ s) for broadband spectroscopy in the 0.1–3 THz frequency range. The THz radiation pulses are generated using femtosecond near-infrared (NIR) laser pulses incident on semiconductors or nonlinear optical crystals.[21,22] The THz radiation pulses are detected using photoconducting antennas or electrooptic sampling.[23–26] Time-domain THz spectroscopy measures the complex dielectric function of a material with high dynamic range.[27]

All methods of THz spectroscopy discussed so far have in common that the material to be analyzed is illuminated with THz radiation from a THz radiation source. Another approach to THz spectroscopy is the detection and spectral analysis of THz radiation naturally emitted by a material. This type of THz spectroscopy is used as an experimental method in astronomy. It is performed with THz heterodyne spectrometers.[28]

Today, commercially available THz spectrometers for analytical chemistry are far-infrared Fourier transform spectrometers (FIR-FTS), BWO THz spectrometers, and TD THz spectrometers. The specifications and performance of these three methods are discussed in the subsequent text.

2 SPECIFICATIONS AND PERFORMANCE OF DIFFERENT METHODS OF THz SPECTROSCOPY

2.1 Performance

In analytical chemistry, the important specifications for THz spectroscopy are the frequency range, the frequency resolution, and the dynamic range of measurement. The dynamic range is defined as the ratio of the largest to the smallest change in transmission (or reflection) that can be detected by a measurement, and is generally a function of frequency. The values of the dynamic range quoted in this article refer to maximum performance. Also, this article specifies the performance of commercially available THz spectrometers.

The performances of FIR-FTS, BWO THz spectrometers, and TD THz spectrometers differ by frequency range, frequency resolution, dynamic range, and absolute THz radiation power (Table 1). The frequency ranges of the different methods are also illustrated in Figure 3. A BWO THz spectrometer covers the 0.035–1.4 THz frequency range, whereas that of a TD THz spectrometer extends from 0.030 to 3 THz. The frequency range of operation of a BWO THz spectrometer strongly overlaps with that of a TD THz spectrometer The FIR-FTS spectrometer starts operating at 0.3 THz. The frequency range of this type of instrumentation extends into the infrared part of the electromagnetic spectrum.

The difference in performance between BWO THz spectroscopy and TD THz spectroscopy are frequency

Table 1 Performance of commercially available THz spectrometers

Spectrometer	Frequency range (THz)	Frequency resolution (MHz)	Dynamic range of THz power measurements	Absolute THz power (W)
BWO THz	0.035–1.4	1–10	$10^5 : 1$	10^{-3}
TD THz	0.02–3	1000–10000	$10^8 : 1$	10^{-6}
FIR-FTS	>0.3	1000–10000	300 : 1	10^{-9}

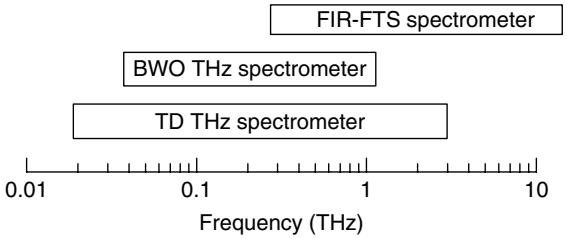

Figure 3 Frequency range of operation of time-domain (TD) spectroscopy, backward-wave oscillator (BWO) spectroscopy, and far-infrared Fourier transform (FIR-FTS) spectroscopy.

resolution and dynamic range. BWO THz spectroscopy allows measurements with 1–10 MHz resolution, whereas the frequency resolution of TD THz spectroscopy is restricted to 1–10 GHz. The dynamic range of TD THz spectroscopy is $10^8 : 1$, whereas that of a BWO spectrometer is $10^5 : 1$.

The frequency range of FIR-FTS starts at 0.3 THz and extends into the midinfrared and NIR (30–300 THz) range. The frequency resolution of FIR-FTS is 1–10 GHz. It is similar to the frequency resolution of TD THz spectroscopy but is lower than that of BWO THz spectroscopy. The dynamic range of FIR-FTS is much lower than that of BWO THz spectroscopy and TD THz spectroscopy.[29]

In summary, the frequency range of BWO THz spectroscopy and TD THz spectroscopy is very similar. BWO THz spectroscopy is the method of choice for analytical chemistry if spectroscopy with megahertz frequency resolution is desired. TD THz spectroscopy is selected for a spectroscopic measurement if it is necessary to measure very small changes in transmission (or reflection) but megahertz frequency resolution is not needed. The dynamic range of a FIR-FTS below a few THz is low compared with that of BWO THz spectroscopy and TD THz spectroscopy. The strength of an FIR-FTS is that it permits broadband spectroscopy from the THz frequency range into the infrared.

2.2 Merits

All three approaches to THz spectroscopy discussed in this article (Table 2) permit measurements in transmission and reflection geometry. BWO THz spectroscopy and TD THz spectroscopy are phase-sensitive methods of spectroscopy. FIR-FTS is a non-phase sensitive measurement method. Phase-sensitive spectroscopy allows the measurement of the real and imaginary part of the dielectric function of a material. Non-phase sensitive spectroscopy measures only the absorption coefficient of a material. TD THz spectroscopy also allows time-resolved spectroscopy with picoseconds to femtoseconds resolution.

In THz transmission spectroscopy, the THz radiation is incident perpendicular onto the surface of the sample (Figure 4a). Part of the THz radiation beam is reflected from the sample surface and part of it is transmitted through the sample. In THz transmission spectroscopy, the THz radiation transmitted through the sample is recorded by the detector. In THz reflection spectroscopy, the angle of incidence onto the sample is off from the sample surface normal. The THz radiation beam

Table 2 Merits of different methods of THz spectroscopy

Measurement	BWO THz	TD THz	FIR-FTS
Modes of operation	Transmission/reflection Phase sensitive Frequency-domain measurement	Transmission/reflection Phase sensitive Time-domain measurement	Transmission/reflection Non-phase-sensitive Space-domain measurement
Physical observables	Transmission Reflection	Transmissivity Reflectivity	Transmissivity Reflectivity
Principles	Frequency tuning Multiple-beam interference of continuous-wave monochromatic THz radiation	Frequency multiplexing Single beam of broadband THz radiation pulse Demultiplexing by Fourier transformation from time-to-frequency domain Time-resolved spectroscopy with femto-to-picosecond resolution (THz- emission spectroscopy, Optical pump/THz probe spectroscopy)	Frequency multiplexing Two-beam interference of continuous-wave broadband THz radiation Demultiplexing by Fourier transformation from space-to-frequency domain

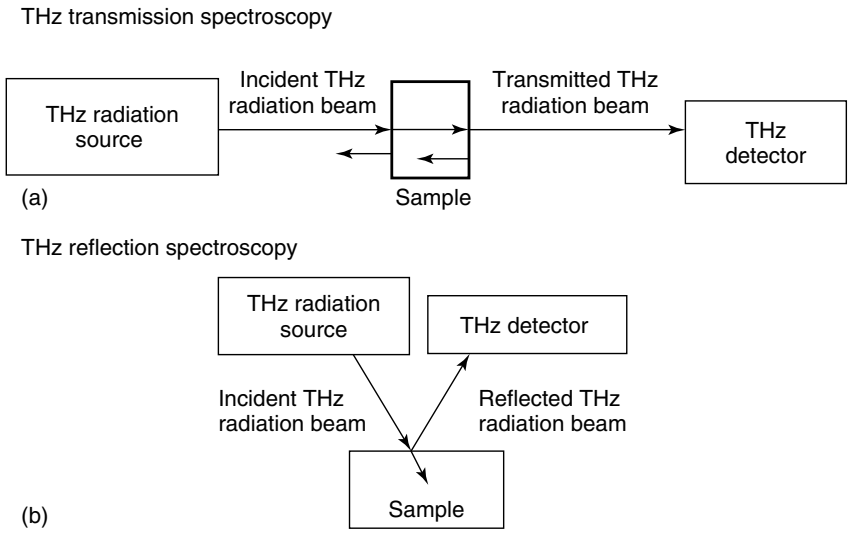

Figure 4 Terahertz spectroscopy measurement modalities. (a) THz spectroscopy in transmission configuration. (b) THz spectroscopy in reflection configuration.

reflected off the sample surface is recorded by the detector (Figure 4b).

The transmissivity $T = I_t/I_0$ or reflectivity $R = I_r/I_0$ of the sample are defined as the ratio of the transmitted I_t or reflected intensity I_r and the intensity of the incident THz radiation I_0. The transmission or reflection coefficient $t = E_t/E_0$ and $r = E_r/E_0$ are the ratio of the transmitted electric field E_t or reflected electric field E_r to the incident THz electric field E_0.

TD THz spectroscopy measures the transmission t (or reflection r) of a sample. The transmission t (or reflection r) of the sample is determined by the complex index of refraction $n = n_0 + ik_0$ of the sample material. Here, n_0 is the index of refraction and k_0 is the attenuation coefficient. The relationships between transmission t (or reflection r) and the complex index of refraction are described by Fresnel's equations.[30]

$$t(\omega) = \frac{4n \exp\left[-\mathrm{i}\frac{\omega}{c}nd\right]}{(n+1)^2\left(1 - \left(\frac{n-1}{n+1}\right)^2 \exp\left[-2\mathrm{i}\frac{\omega}{c}nd\right]\right)} \quad (1)$$

$$r_p(\Theta, n(\omega)) = -\frac{n^2(\omega)\cos\Theta - \sqrt{n^2(\omega) - \sin^2\Theta}}{n^2(\omega)\cos\Theta + \sqrt{n^2(\omega) - \sin^2\Theta}} \quad (2)$$

$$r_s(\Theta, n(\omega)) = -\frac{\cos\Theta - \sqrt{n^2(\omega) - \sin^2\Theta}}{\cos\Theta + \sqrt{n^2(\omega) - \sin^2\Theta}} \quad (3)$$

In Equations (1)–(3), d is the sample thickness, $c = 3 \times 10^8\,\mathrm{m\,s^{-1}}$ is the velocity of light, $\omega = 2\pi f$ is the angular frequency with f being the frequency, and θ is the angle of incidence. Furthermore, r_p and r_s are the reflections in the case of the incident THz electric field being polarized parallel (p) or perpendicular (s) to the plane of incidence. These equations apply for a homogenous sample with parallel front and back surfaces, which are embedded in air. Similar equations have been developed for samples consisting of multiple layers of different materials and a thin conducting layer on a transparent substrate.[30] The complex index of refraction n of the sample is obtained from the measured transmission t and reflection r by numerically solving Equations (1)–(3). Computational routines have been developed for this purpose.[31] The complex index of refraction is related to the complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ and the complex conductivity $\sigma(\omega) = \sigma_1(\omega) + i\sigma_2(\omega)$ by the following relationships $\varepsilon = n^2$, $\varepsilon_1 = {n_0}^2 - {k_0}^2$, $\varepsilon_2 = 2n_0k_0$, and $\sigma = i\omega\varepsilon\varepsilon_0$. The power absorption coefficient α of a material as defined by the Bouguer–Lambert law of transmission $I_t \propto I_0 \exp(-\alpha d)$ is calculated using k_0 and $\alpha = 4\pi k_0/\lambda$, where $\lambda = c/f$ is the wavelength of the THz radiation. TD THz spectroscopy is phase sensitive because a single transmission or reflection measurement provides both the real and imaginary part of the complex index of refraction.

It is possible to perform time-resolved THz spectroscopy with femtoseconds to picoseconds temporal resolution using TD THz spectroscopy.[27] The THz radiation pulses in a TD THz spectrometer are generated using a femtosecond NIR laser pulse. The femtosecond NIR laser beam is divided by a beam splitter on a pump beam, which excites the THz radiation pulse, and a probe beam that gates the THz detector. The temporal profile of the THz radiation pulse transmitted through

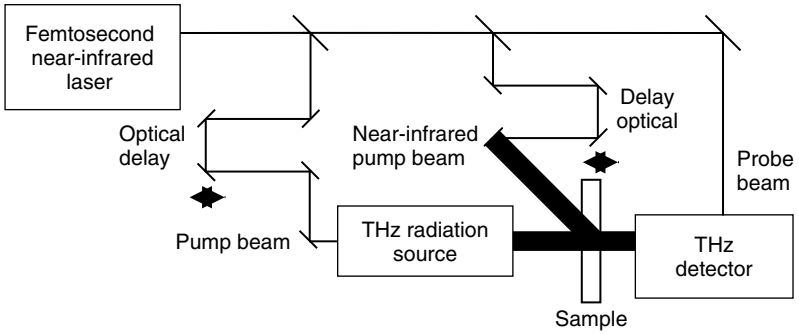

Figure 5 Schematic of experimental arrangements of time-domain THz spectroscopy and optical pump/THz probe spectroscopy.

(or reflected) from the sample is recorded by delaying the probe beam relative to the pump beam by an optical delay line. In optical pump/THz probe spectroscopy, the femtosecond NIR laser beam is divided into three beams: the pump beam for THz pulse generation, the probe beam for gating of the THz detector, and the optical pump beam. The optical pump beam is also incident on the sample surface and is aligned such that the optical beam focus and THz beam focus overlap. Absorption of the femtosecond NIR laser pulse by the sample drives the sample out of equilibrium. The return of the sample to equilibrium is probed by the THz radiation beam by delaying it relative to the femtosecond NIR pulse. This results in time-resolved THz spectroscopy with femtosecond-to-picosecond resolution and provides insight into the dynamic dielectric properties of a material (Figure 5).

BWO THz spectroscopy measures the transmissivity T or reflectivity R of the sample under investigation.[17,32] The principle of BWO THz spectroscopy is multiple-beam interference. This is achieved because of multiple reflections of the continuous-wave THz radiation beam within the plane–parallel front and back surfaces of the sample. For transparent samples, the multiple-beam interference results in transmissivity and reflectivity of the sample changing periodically as a function of the THz frequency. The period of the oscillations is directly related to the index of refraction n_0 of the material by $m\lambda/2 = n_0 d$, where λ is the THz radiation wavelength, m is the number of interference maxima, and d is the sample thickness (Figure 6). The absorption of the material determines the minima in the interference pattern. Low absorption is indicated by low minima. The general relationship between transmissivity T or reflectivity R of the sample and the index of refraction n_0 and attenuation coefficient k_0 is described by Equations (4), (5), and (6).[17]

$$T = \exp\left(-\frac{4\pi k_0 d}{\lambda}\right) \frac{(1-R)^2 + 4R\sin^2\psi}{\left(1 - R\exp\left(-\frac{4\pi k_0 d}{\lambda}\right)\right)^2 + 4R\exp\left(-\frac{4\pi k_0 d}{\lambda}\right)\sin^2\left(\frac{2\pi n_0 d}{\lambda} + \psi\right)} \tag{4}$$

$$R = \frac{(n_0 - 1)^2 + {k_0}^2}{(n_0 + 1)^2 + {k_0}^2} \tag{5}$$

$$\psi = \arctan\left(\frac{2k_0}{n_0^2 + {k_0}^2 - 1}\right) \tag{6}$$

The index of refraction n_0 and attenuation coefficient k_0 of the sample are obtained by fitting Equations (4)–(6) to the measured transmissivity.

In the case of strong absorption of the sample, the interference maxima and minima are less pronounced (Figure 7). In this situation, it is necessary to measure, in addition to the transmissivity T, the phase delay Φ of the THz radiation transmitted through the sample.

$$\Phi = \frac{2\pi nd}{\lambda} - \arctan\frac{k_0(n_0^2 + {k_0}^2 - 1)}{({k_0}^2 + n_0^2)(2 + n_0)n_0} + \arctan\frac{R\exp\left(-\frac{4\pi k_0 d}{\lambda}\right)\sin\left(2\left(\frac{2\pi nd}{\lambda} + \psi\right)\right)}{1 - R\exp\left(-\frac{4\pi k_0 d}{\lambda}\right)\cos\left(2\left(\frac{2\pi nd}{\lambda} + \psi\right)\right)} \tag{7}$$

The index of refraction n_0 and attenuation coefficient k_0 are obtained by fitting Equations (4)–(7) to the measurements of T and Φ.

BWO THz spectroscopy and TD THz spectroscopy are both phase-sensitive methods of THz spectroscopy. They measure the real and imaginary part of the dielectric function of the sample. Because TD THz spectroscopy measures transmission (or reflection), the complex index of refraction of a material is obtained

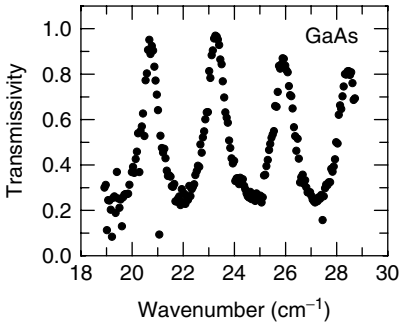

Figure 6 Transmissivity of a plane–parallel gallium arsenide (GaAs) crystal measured with a BWO THz spectrometer.

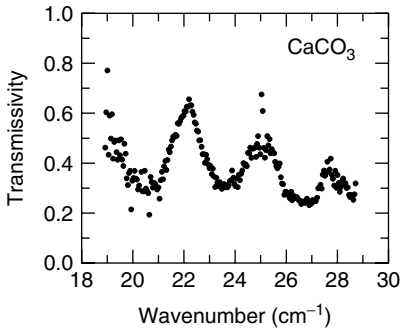

Figure 7 Transmissivity of a plane–parallel calcite ($CaCO_3$) crystal measured with a BWO THz spectrometer.

from two measurements for transparent as well as strongly absorbing materials. In contrast, BWO THz spectroscopy, which measures transmissivity (or reflectivity), requires four measurements to obtain the complex index of refraction for strongly absorbing materials. Moreover, BWO THz spectroscopy measures only the static dielectric properties of a material. An advantage of a continuous-wave BWO THz spectrometer is that this instrument provides high frequency resolution and high dynamic range simultaneously. For a TD THz spectrometer, the dynamic range of the measurement decreases drastically with subgigahertz frequency resolution. The dynamic range and the frequency resolution are linked in TD THz spectroscopy due to the pulsed nature of the radiation and the TD measurement approach.

FIR-FTS also measures the transmissivity (or reflectivity) of a sample under investigation.[11,12,14,15] The principle of FIR-FTS is two-beam interference of broadband THz radiation (Figure 8). Generally, FIR-FTS is performed with a Michelson interferometer and relies on Fourier transformation between the space and frequency domain. It measures the Fourier transform $I(s)$ of the transmission (or reflection) spectrum of a sample as a function of the spatial delay s between two interfering beams, as described in Equation (8). In this equation,

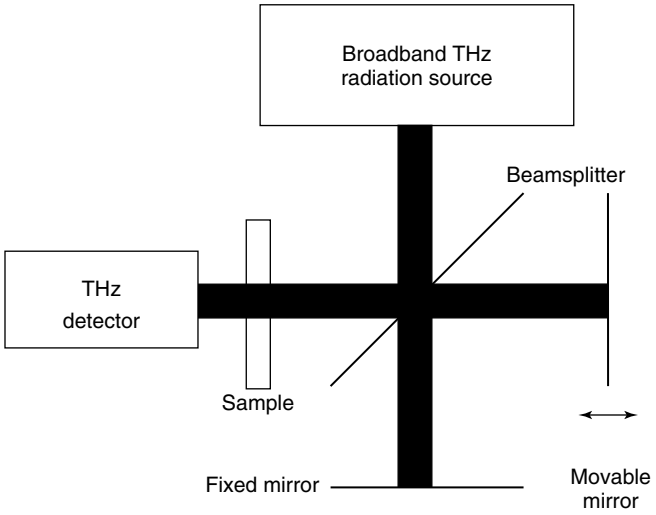

Figure 8 Schematic of a far-infrared Fourier transform spectrometer.

$k = 2\pi/\lambda$ is the wavenumber, I is the total intensity of the broadband THz radiation source, and $I(k)$ is the intensity of the THz radiation in the wavenumber interval $\mathrm{d}k$, as described in Equation (9). The actual transmission (or reflection) spectrum of the sample is obtained by numerically calculating the inverse Fourier transform $I(k)$ of the measurement $I(s)$. The relationship between spectrum $I(k)$ and measured Fourier transform $I(s)$ is described in Equation (10).[33]

$$I(s) = \frac{1}{2}I + \frac{1}{4}\int_{-\infty}^{\infty} I(k)\exp(\mathrm{i}ks)\,\mathrm{d}k \tag{8}$$

$$I = \frac{1}{2}\int_{-\infty}^{\infty} I(k)\,\mathrm{d}k \tag{9}$$

$$I(k) = \frac{2}{\pi}\int_{-\infty}^{\infty}\left\{I(s) - \frac{1}{2}I\right\}\exp(-\mathrm{i}ks) \tag{10}$$

The transmission (or reflection) spectrum measured by FIR-FTS provides information on the absorption frequencies, the relative strength of an absorption line, and the linewidth of the absorption. The absorption coefficient α of the material is obtained by using the Bouguer–Lambert law $I_t \propto I_0\exp(-\alpha d)$ and measuring the transmission spectrum of two samples with different thickness d. FIR-FTS measures the static absorption coefficient of a material. Since FIR-FTS is non-phase sensitive, it does not provide information on the index of refraction of the material. In order to obtain the index of refraction of a material by FIR-FTS, it is necessary to measure the transmission and reflection spectrum of a sample. Another approach for obtaining the index of refraction is performing a Kramers–Kronig analysis of the measured absorption coefficient. However, the accuracy of this computational method requires measurements

of the absorption coefficient over a wide range of frequencies.

The advantage of TD THz and BWO THz spectroscopy over FIR-FTS is the capability of phase-sensitive measurements that allows the direct determination of the index of refraction and attenuation coefficient of materials in the 0.02–3 THz frequency range. The advantage of FIR-FTS is the rapid determination of absorption lines, their relative strength, and linewidth over a wide frequency range of 0.3–300 THz.

TD THz spectroscopy is performed in the time-domain and FIR-FTS spectroscopy is performed in the space domain. Both methods rely on frequency multiplexing, which describes that all frequencies represented in the broadband THz radiation source are incident on the sample simultaneously. For both methods, frequency demultiplexing is performed by Fourier transformation of the measurement to the frequency domain. BWO THz spectroscopy is performed in the frequency domain. The spectrum is recorded by continuously changing the emission frequency of the BWO THz radiation source. The advantage of a BWO-tunable THz radiation source in comparison to a FIR-FTS is that frequencies less than 1 THz are beyond the useful measurement range of most FIR-FTS. The much higher brightness of the BWO THz radiation source results in a much larger signal-to-noise ratio at frequencies below 1.4 THz.

3 TECHNICAL DESCRIPTIONS

3.1 Far-infrared Fourier Transform Spectrometer

Fourier transform spectroscopy is a method used for spectroscopies in the FIR (THz) and ultraviolet frequency range.[3] The basic components of a Fourier transform spectrometer are a broadband THz radiation source, a two-beam (Michelson) interferometer, and a broadband detector (Figure 8). The components are selected according to the frequency range of operation. In the FIR (0.3–20 THz), the radiation source of choice is a globar. Sensitive detectors for FIR radiation are silicon bolometers cooled to temperatures between 1.4 and 4 K. The beam splitters for the interferometer are made from Mylar. Optics used to focus the FIR radiation onto the sample is made from polyethylene or polypropylene.

The globar emits a broad spectrum of FIR radiation. A beam splitter divides the FIR radiation beam into two. One of the beams is reflected from a fixed mirror, whereas the other is reflected from a movable mirror. The path difference between the two beams is changed by moving the mirror. The beam splitter combines the two beams again and steers the FIR radiation to the detector. FIR-FTS measures the interferogram $I(s)$ of the two FIR radiation beams as a function of the path difference between the two beams. The spectrum $I(k)$ is obtained through calculation of the Fourier transform of the interferogram $I(s)$. For spectroscopy, the FIR radiation beam is focused onto a sample and the transmitted (or reflected) FIR radiation is detected. The illustration in Figure 8 is only a schematic of FIR-FTS. Various versions of two-beam interferometers for specific applications of FIR-FTS have been developed and are described in the literature.[14]

3.2 Backward-wave Oscillator Terahertz Spectrometer

The THz radiation source in a BWO THz spectrometer is a BWO,[17] which is an electron beam device. A cathode produces electrons via thermal emission. Subsequently, electrons are accelerated by applying high voltage toward an anode. An applied magnetic field focuses the electrons to a narrow beam. The electron beam interacts with a periodic metallic structure while traveling to the anode. Electromagnetic interactions between the electron beam and the metallic periodic structure result in the generation of THz radiation. The THz radiation is collected by a waveguide and coupled to free space by a horn antenna. The frequency of the THz radiation is determined by the electron velocity, which in turn depends on the applied accelerating voltage. The frequency of THz radiation emitted by a BWO is tunable. Continuous tuning of the THz radiation frequency is achieved by changing the applied voltage. The frequencies of typical BWOs are listed in Table 3. The THz radiation power of a BWO can be as high as 50 mW. The THz radiation power is characterized by high frequency stability $\Delta\nu/\nu = 10^{-5}$,

Table 3 Frequency range of operation and power of backward-wave oscillators

BWO	1	2	3	4	5	6	7
Frequency (THz)	0.035–0.055	0.05–0.08	0.08–0.12	0.12–0.18	0.18–0.26	0.26–0.37	0.37–0.53
THz power (mW)	40	30	30	50	10	10	5
BWO	7	8	9	10	11	12	13
Frequency (THz)	0.37–0.53	0.5–0.7	0.7–0.9	0.8–1.0	0.9–1.1	1.0–1.25	1.2–1.4
THz power (mW)	5	5	5	3	3	2	1

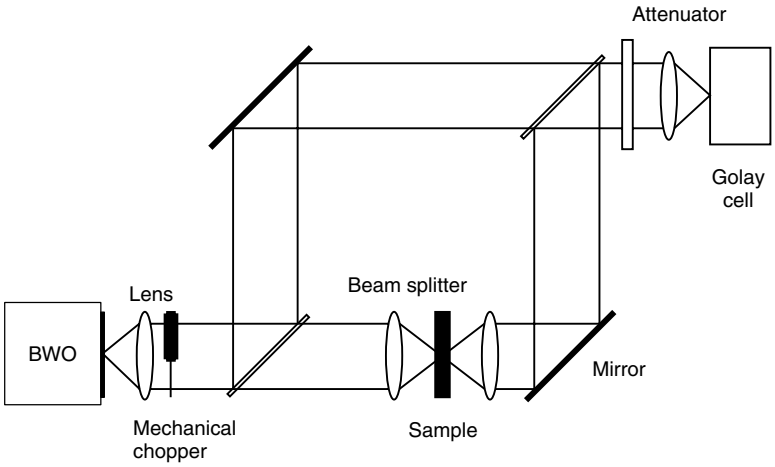

Figure 9 Schematics of a phase-sensitive BWO THz spectrometer.

high long-term power stability $\Delta P/P = 10^{-3}$, and high degree of linear polarization (100 : 1).

Phase-sensitive THz transmission and reflection spectroscopy with BWOs as the THz radiation source is illustrated in Figure 9. In transmission geometry, the THz radiation emitted by the BWO is collected by a lens and split into two beams by a wire-grid beam splitter. One beam is transmitted through the material to be analyzed and the second beam serves as a reference to measure the phase shift induced on the transmitted THz radiation wave by the material. The transmitted beam and the reference beam are combined by a second wire grid and focused on the detector. Depending on whether transmission or reflection measurements are performed, the measured quantities are the transmission coefficient T of the sample and the phase shift Φ_t of the transmitted wave or reflection coefficient R and phase shift Φ_r of the reflected beam. These measured quantities are directly related to the complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ and complex magnetic permeability $\mu(\omega) = \mu_1(\omega) + i\mu_2(\omega)$ of the material under investigation. For nonmagnetic materials, the two unknown quantities ε_1 and ε_2 are obtained from either a transmission or a reflection measurement. For magnetic materials, four unknown quantities have to be determined, ε_1, ε_2, μ_1, and μ_2, using the measured transmission and reflection coefficient and corresponding phase shifts.[32]

3.3 Time-domain Terahertz Spectrometer

A TD THz transmission spectrometer is powered by a femtosecond NIR laser and operates according to the pump–probe principle.[23,24,26] A schematic of the experimental arrangements is displayed in Figure 5. The pump–probe principle is briefly described as follows: The femtosecond NIR laser beam is split in two beams. One of the beams, the pump beam, hit the THz radiation source, which in response to the femtosecond laser pulse releases a pulse of THz radiation. The probe beam gates the THz detector whose response is proportional to the amplitude and the sign of the electric field of the THz pulse. The time profile of the THz pulse is traced by varying the time delay between the pump and probe beam. An example of a measured THz radiation pulse is illustrated in Figure 10. In this experiment, the THz radiation pulse traveled approximately 0.3 m through air with a residual water vapor concentration of $1.7 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$. The Fourier transform of the THz radiation pulse is illustrated in Figure 11. The frequency spectrum extends from 0.45 to 2.5 THz. The Fourier transform shows that the dynamic range of the TD THz spectrometer is a function of frequency. The maximum of the dynamic range for measurements of the THz electric field is 800 : 1 at 0.3 THz. The spectrum in Figure 11 is the THz spectrum of air with a residual water vapor concentration of $1.7 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$. In comparison, Figure 12 displays the spectrum of a THz radiation pulse, which traveled the same distance but through air with a residual water vapor

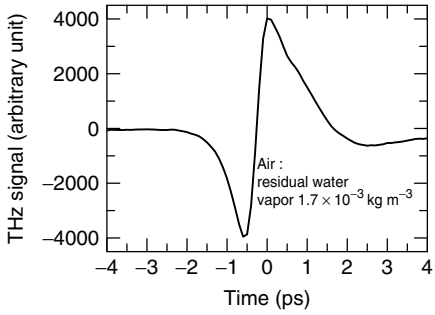

Figure 10 Time-domain measurement of a THz radiation pulse with a TD THz spectrometer. The THz signal is proportional to the amplitude of the THz electric field.

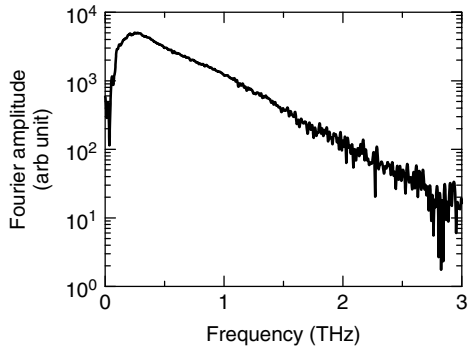

Figure 11 Spectrum of the THz radiation pulse. It is obtained by Fourier transformation of the time-domain measurement of the THz radiation pulse to the frequency domain. In this measurement, the THz radiation pulse propagated through 0.3 m of nitrogen (N_2) with a residual water vapor concentration of $1.7 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$ at room temperature and atmospheric pressure. The dynamic range of the TD THz spectrometer is frequency dependent. The maximum of the dynamic range is at ≅0.3 THz.

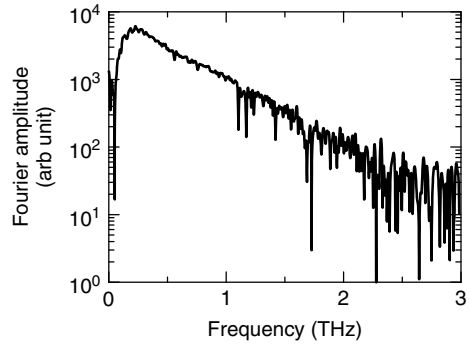

Figure 12 Spectrum of another THz radiation pulse. Again, it is obtained by Fourier transformation of the time-domain measurement of the THz radiation pulse. In this measurement, the THz radiation pulse propagated through 0.3 m of nitrogen (N_2) with a residual water vapor concentration of $6.9 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$ at room temperature and atmospheric pressure. The sharp dips in the spectrum correspond to water vapor absorptions lines in the THz frequency range.

concentration of $6.9 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$. The narrow dips in transmission marked by an arrow correspond to absorption lines of water vapor in the THz frequency range below 1.5 THz. The transmission t of a sample placed in the focus of the THz beam is the ratio of the Fourier transforms of a THz pulse transmitted through the sample and the Fourier transform of a reference, e.g. a THz pulse freely propagating through air.

3.3.1 THz Radiation Emitters for Time-domain THz Spectrometers

THz radiation sources used in TD THz spectrometers are either semiconductors or nonlinear optical crystals. Terahertz radiation emission from semiconductors exposed to femtosecond NIR laser pulses originates either from a nonlinear optical process or from ultrafast photocurrents. Nonlinear optical processes resulting in THz radiation emission are bulk- or surface-field-induced optical rectification of the incident femtosecond NIR laser pulses. Generation of THz radiation pulses due to optical rectification of femtosecond laser pulses either in a semiconductor or in a nonlinear optical crystal is based on difference in frequency mixing of all frequencies within the bandwidth Δf of a femtosecond NIR laser pulse.[26] Terahertz radiation emission due to ultrafast photocurrents can be achieved through acceleration of photocarriers by intrinsic or extrinsic electric fields.[23,24] Intrinsic electric fields occurring at a semiconductor surface are surface depletion/accumulation fields or a photo-Dember field. Extrinsic electric fields are generated by applying laterally a voltage across a gap between two metal electrodes deposited on the semiconductor surface (photoconductive switch).

3.3.2 THz Radiation Detectors for Time-domain THz Spectrometers

The THz radiation detector in a TD THz spectrometer is either a photoconducting antenna or an electrooptic THz detector. A photoconducting antenna for the detection of the temporal waveform of THz radiation pulses consists of a thin-film antenna deposited on a thin semiconductor layer supported by an electrically insulating substrate, which is transparent at THz frequencies.[23,24] The semiconductor layer is manipulated such that it exhibits a high-dark resistivity and femtosecond carrier lifetime. Most commonly, ion-implanted silicon-on-sapphire (SOS) or low-temperature-grown GaAs (LT-GaAs) on semi-insulating GaAs is used for detector fabrication. The thin-film antenna is typically a center-fed dipole, which is approximately 50 μm long and 10 μm wide or a bipolar dipole antenna with 30-μm separation. Both types of antenna have a 5-μm-wide photoconducting gap in the center. A schematic of the experimental arrangements for photoconducting detection is displayed in Figure 13. For the detection, the femtosecond laser probe beam and the THz beam incident perpendicular onto the antenna from opposite sides. The probe beam is focused onto the photoconducting gap of the antenna. The THz radiation is incident through the substrate onto the antenna. The THz radiation is focused by a parabolic mirror and the coupling of the THz radiation to the substrate is enhanced by a small high-resistivity silicon lens that is attached to the substrate. The electrical field of the incident THz radiation couples to the antenna and induces a voltage across the photoconducting gap. A photocurrent $j(t)$ flows under the action of the induced voltage when the probe

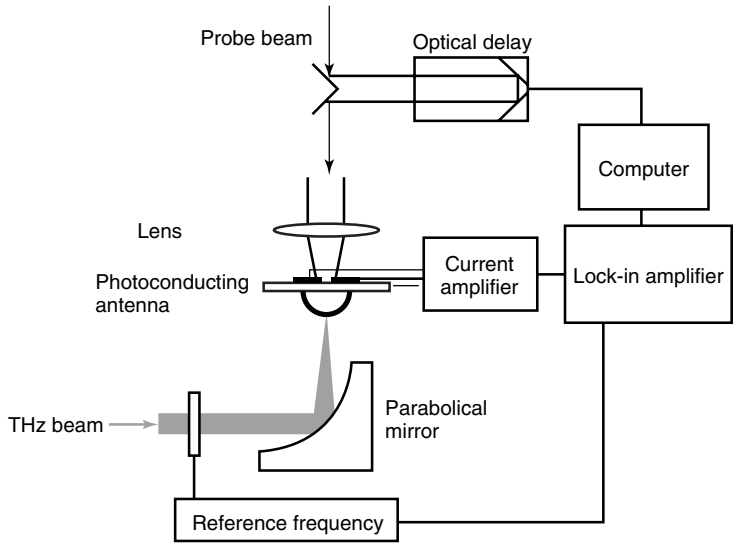

Figure 13 Detection of THz radiation with photoconducting antennas in a TD THz spectrometer.

laser beam excites photocarriers in the photoconducting gap. This photocurrent is the convolution of the transient conductivity $\sigma(t)$ of the semiconductor and the electric field $E(t)$ across the photoconducting gap, as described in Equation (11). The magnitude and sign of the current $j(t)$ are proportional to the amplitude and phase of the electric THz field $E(t)$. The current is measured by a current amplifier.

$$j(t) = \int \sigma(t - t')E(t')\,\mathrm{d}t' \tag{11}$$

The temporal profile of the THz radiation pulse is measured by scanning the femtosecond laser pulse through the THz pulse by an optical delay line.

An alternative to photoconducting antennas for the measurement of the temporal waveform of a THz radiation pulse is electrooptic detection.[25,26] Electrooptic detection of THz radiation pulses relies on the fact that the electric field of a THz radiation beam, which passes through a nonlinear optical crystal such as a ZnTe crystal, induces birefringence in the crystal via the Pockel's effect.[34] This birefringence can be probed by a linearly polarized femtosecond laser pulse, which propagates collinearly with the THz beam through the ZnTe crystal. The phase difference $\Delta\Phi$ between the ordinary and extraordinary ray of the probe laser pulse is directly proportional to the electric THz field. In Equation (12), ω_{opt} is the angular frequency of the probe laser beam, n_0 is the index of refraction of the ZnTe crystal, r_{41} is the electrooptic coefficient of ZnTe, and Δx is the thickness of the ZnTe crystal.

$$\Delta\Phi = \frac{\omega_{\mathrm{opt}}}{c} {n_0}^3 r_{41} E_{\mathrm{THz}} \Delta x \tag{12}$$

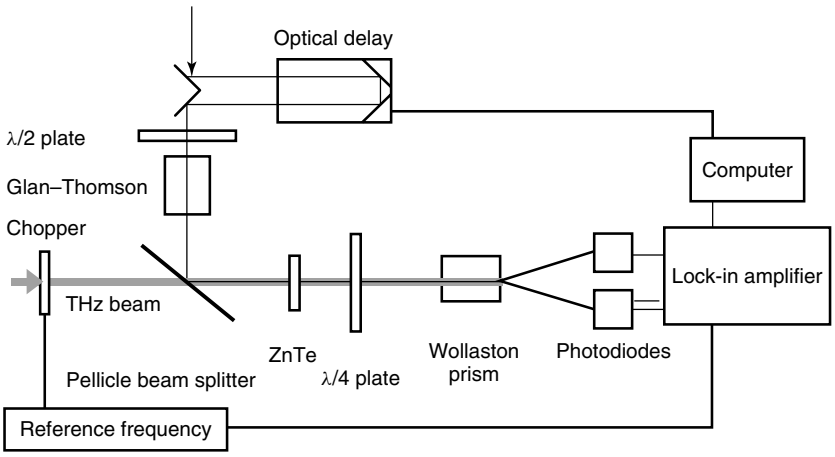

Figure 14 Detection of THz radiation in a TD THz spectrometer by electrooptic sampling.

The experimental arrangements for electrooptic detection are displayed in Figure 14. The setup consists of a λ/2 plate, a Glan–Thomson polarizer, a pellicle beam splitter, a ZnTe crystal, a λ/4 plate, a Wollaston prism, and two photodiodes. Collinear propagation of the probe beam and of the THz beam in the ZnTe crystal is achieved through the pellicle beam splitter. The pellicle beam splitter is transparent for THz radiation but reflects optical light. The λ/2 plate allows a variation in the plane of polarization of the linearly polarized optical laser beam. The Glan–Thomson prism enhances the degree of linear polarization from 100 : 1 at the output of the Ti-sapphire laser to 10^6 : 1.

The induced birefringence of the ZnTe crystal changes the linear polarization of the optical probe pulse to elliptical polarization. Subsequently, the elliptical polarization is changed into a circular polarization by the λ/4 wave plate. The Wollaston prism spatially separates the two orthogonal components of the circularly polarized laser light. Finally, the intensity of each orthogonal component is individually detected by a photodiode. The difference in the photodiode output is measured by a lock-in amplifier. The THz electric field is proportional to the difference in the output of the two photodiodes. The temporal profile of the THz radiation pulse is measured again by scanning the femtosecond laser pulse through the THz radiation pulse.

3.3.3 Optics for Time-domain THz Spectrometers

The THz radiation is collected, steered, and focused with gold-plated off-axis parabolic mirrors. The mirrors have a typical diameter of 25–50 mm. The focal length of the mirrors is typically 25–300 mm. In order to achieve proper THz beam steering, it is important that the THz emitter, sample, and THz detector are accurately positioned at the focal points of the mirrors to avoid spatial transformations of the THz beams. As a result of spatial transformations, phase changes and frequency filtering of the pulse time profile may occur.[35]

4 OVERVIEW OF APPLICATIONS OF THz SPECTROSCOPY

Traditional areas of application of THz spectroscopy have been chemistry, materials science, and physics. A large body of work on THz spectroscopy of gases, liquids, and solids has been reported.[3,4,7,8,36–39] It is beyond the scope of this article to survey and summarize all this work.

As explained previously, FIR spectroscopy is a synonym of THz spectroscopy. Therefore, THz spectroscopy is infrared spectroscopy in the very low frequency range of the infrared (Figure 1). THz spectroscopy measures the dielectric function of materials in the THz frequency range. The dielectric function provides information on the dynamic properties of a material. All selection rules for infrared spectroscopy also apply to THz spectroscopy.

A very fruitful area of research has been THz spectroscopy of gases. Small polar molecules in the gas phase exhibit characteristic absorption lines in the microwave frequency range (Figure 1) and THz frequency range. The absorption frequencies of diatomic molecules with a permanent electric dipole moment are $f = 2h/(8\pi^2 I)(j+1)$, where $j = 0, 1, 2, 3, \ldots$, $h = 6.6 \times 10^{-34}$ Js, the moment of inertia is $I = \mu r^2$, the internuclear distance is r, and $\mu = m_1 m_2/(m_1 + m_2)$ the reduced mass of the molecule. Absorption due to excitation of rotations occurs for molecules with a small reduced mass μ. Examples are HF, HCl, H_2O, and NH_3. The vibration frequency of diatomic molecules is $\omega = 2\pi f = (k/\mu)^{1/2}$, where k is the force constant of the bond and μ is the reduced mass of the molecules. Vibrational frequencies of molecules are found in the THz frequency range for a large reduced mass (at least one heavy atom) or a weak force constant. Examples of weak force constants are atoms bound by hydrogen or van der Waals bonds. Examples of molecules with a large reduced mass are heavy-metal-to-metal or metal-to-ligand bonds.[3]

The understanding of rotational and vibrational transitions of molecules in the THz frequency range is important for the understanding of the transmission of the atmosphere in this frequency range.[38,39]

THz spectroscopy of biomolecules and pharmaceutical applications of THz spectroscopy are emerging as new areas of application.[40–42] A combination of THz spectroscopy and THz imaging is suggested as a novel medical imaging modality.[43]

ACKNOWLEDGMENTS

The TD THz spectroscopy measurements and BWO THz spectroscopy measurements presented in this article were performed by Hanns Selig and Suranja Sengupta. Part of this material is based upon work supported by the National Science Foundation under Grant No. 0619499.

RELATED ARTICLES

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Polymers and Rubbers

Infrared Spectroscopy **(Volume 12)**
Emission Spectroscopy, Infrared

General Articles **(Volume 15)**
Microwave Techniques • Ultrafast Laser Technology and Spectroscopy

REFERENCES

1. S.L. Dexheimer, *Terahertz Spectroscopy – Principles and Applications*, Taylor & Francis, LLC, Boca Raton, 2008.
2. G.W. Chantry, *Modern Aspects of Microwave Spectroscopy*, Academic Press, London, 1979.
3. P.R. Griffith, *Chemical Infrared Fourier Transform Spectroscopy*, John Wiley & Sons, New York, 1975.
4. G.W. Chantry, *Submillimetre Spectroscopy*, Academic Press, London and New York, 1971.
5. G. Gruener, *Millimeter and Submillimeter Wave Spectroscopy of Solids*, Springer Verlag, Berlin, 1998.
6. D.H. Martin, *Spectroscopic Techniques for Far Infrar-Red, Submillimetre and Millimetre Waves*, North-Holland Publishing Company, Amsterdam, 1967.
7. A. Finch, P.N. Gates, K. Radcliffe, F.N. Dickson, F.F. Bentley, *Chemical Applications of Far Infrared Spectroscopy*, Academic Press, London and New York, 1970.
8. L.C. Robinson, *Methods of Experimental Physics Volume 10 – Physical Principles of Far-Infrared Radiation*, Academic Press, New York and London, 1973.
9. J.C. Wiltse, 'History of Millimeter and Submillimeter Waves', *IEEE Trans. Microw. Theory Tech.*, **32**, 1118 (1984).
10. G.R. Wilkinson, D.H. Martin, 'Grating Spectroscopy', in *Spectroscopic Techniques for Far Infrar-Red, Submillimetre and Millimetre Waves*, ed. D.H. Martin, North-Holland Publishing Company, Amsterdam, 69–112, 1967.
11. P.L. Richards, 'Fourier Transform Spectroscopy', in *Spectroscopic Techniques for Far Infrar-Red, Submillimetre and Millimetre Waves*, ed. D.H. Martin, North-Holland Publishing Company, Amsterdam, 33–68, 1967.
12. J.R. Birch, T.J. Parker, 'Dispersive Fourier Transform Spectroscopy', in *Infrared and Millimeter Waves Volume 2 Instrumentation*, ed. K.J. Button, Academic Press, New York, 138–272, 1979.
13. D.H. Martin, 'Polarizing (Martin-Puplett) Interferometric Spectrometers for the Near- and Submillimeter Spectra', in *Infrared and Millimeter Waves Volume 6 Systems and Components*, ed. K.J. Button, Academic Press, New York, 66–149, 1982.
14. L. Genzel, 'Far-Infrared Fourier Transform Spectroscopy', in *Millimeter and Submillimeter Wave Spectroscopy of Solids*, ed. G. Gruener, Springer Verlag, Berlin, 169–220, 1998.
15. J.R. Ferraro, A.J. Rein, 'Applications of Diffuse Reflectance Spectroscopy in the Far-Infrared Region', in *Fourier Transform Infrared Spectroscopy – Applications to Chemical Systems Volume 4*, eds. J.R. Ferraro, L.J. Basile, Academic Press, Orlando, 244–282, 1985.
16. M.F. Kimmitt, *Far-Infrared Techniques*, Pion Limited, London, 1970.
17. G. Kozlov, A. Volkov, 'Coherent Source Submillimeter Wave Spectroscopy', in *Millimeter and Submillimeter Wave Spectroscopy of Solids*, ed. G. Gruener, Springer Verlag, Berlin, 51–110, 1998.
18. B.L. Bean, S. Perkowitz, 'Far Infrared Submillimeter Spectroscopy with an Optically Pumped Laser', in *Infrared and Millimeter Waves Volume 2 Instrumentation*, ed. K.J. Button, Academic Press, New York, 273–298, 1979.
19. G.A. Blake, K.B. Laughlin, R.C. Cohen, K.L. Busarow, D.-H. Gwo, C.A. Schmuttenmaer, D.W. Steyert, R.J. Saykally, 'Tunable Far Infrared Laser Spectrometers', *Rev. Sci. Instrum.*, **62**, 1693 (1991).
20. G. Gagliardi, G. Viciani, M. Inguscio, 'Generation of tunable far-infrared radiation with a quantum cascade laser', *Opt. Lett.*, **7**, 521 (1999).
21. C. Rulliere, *Femtosecond Laser Pulses*, Springer Verlag, Berlin, 1998.
22. A. Bonvalet, M. Joffre, 'Terahertz Femtosecond Pulses', in *Femtosecond Laser Pulses*, ed. C. Rulliere, Springer Verlag, Berlin, 285–306, 1998.
23. M.C. Nuss, J. Orenstein, 'Terahertz Time-Domain Spectroscopy', in *Millimeter and Submillimeter Wave Spectroscopy of Solids*, ed. G. Gruener, Springer Verlag, Berlin, 7–50, 1998.
24. R.A. Cheville, 'Terahertz Time-Domain Spectroscopy with Photoconductive Antennas', in *Terahertz Spectroscopy – Principles and Applications*, ed. S.L. Dexheimer, Taylor & Francis, LLC, Boca Raton, 1–40, 2008.
25. Z. Jiang, X.-C. Zhang, 'Free-Space Electro-Optic Techniques', in *Sensing with Terahertz Radiation*, ed. D. Mittleman, Springer Verlag, Berlin, 155–192, 2003.
26. I. Wilke, S. Sengupta, 'Nonlinear Optical Techniques for Terahertz Pulse Generation and Detection – Optical Rectification and Electrooptic Sampling', in *Terahertz Spectroscopy – Principles and Applications*, ed. S.L. Dexheimer, Taylor & Francis, LLC, Boca Raton, 41–72, 2008.
27. J.S. Baxter, C.A. Schmuttenmaer, 'Time-Resolved Terahertz Spectroscopy and Terahertz Emission Spectroscopy', in *Terahertz Spectroscopy – Principles and Applications*,

ed. S.L. Dexheimer, Taylor & Francis, LLC, Boca Raton, 73–118, 2008.

28. G.W. Schwaab, 'Heterodyne Spectrometers', *Infrared Phys. Techn.*, **40**, 207 (1999).
29. P.Y. Han, M. Tani, M. Usami, S. Kono, R. Kersting, X.-C. Zhang, 'A Direct Comparison between Terahertz Time-Domain Spectroscopy and Far-Infrared Fourier Transform Spectroscopy', *J. Appl. Phys.*, **89**, 2357 (2001).
30. M. Born, E. Wolf, *Principles of Optics: Electromagnetic theory of propagation, interference and diffraction of light*, 7th edition, Cambridge University Press, Cambridge, 1999.
31. L. Duvillaret, F. Garet, J.-L. Coutaz, 'Highly Precise Determination of Optical Constants and Sample Thickness in Terahertz Time-domain Spectroscopy', *Appl. Optics*, **38**, 409 (1999).
32. B. Gorshunov, A. Volkov, I. Spektor, A. Prokhorov, A. Mukhin, M. Dressel, S. Uchida, A. Loidl, 'THz BWO Spectroscopy', *Int. J. Infrared Milli.*, **26**, 1217 (2005).
33. F.K. Kneubuehl, *Repetitorium der Physik*, Teubner, 1994.
34. A. Yariv, *Quantum Electronics*, John Wiley & Sons, Inc., New York, 1989.
35. P. Kuzel, M.A. Khazan, J. Kroupa, 'Spatiotemporal transformations of ultrashort terahertz pulses', *J. Opt. Soc. Am. B*, **16**, 1785 (1999).
36. D.H. Martin, 'Spectrometry between $3\,cm^{-1}$ and $200\,cm^{-1}$', in *Spectroscopic Techniques for Far Infrar-Red, Submillimetre and Millimetre Waves*, ed. D.H. Martin, North-Holland Publishing Company, Amsterdam, 3–32, 1967.
37. J.R. Durig, A.W. Cox Jr., 'Current and Past Research in Far-Infrared Interferometry', in *Fourier Transform Infrared Spectroscopy – Applications to Chemical Systems Volume 1*, eds. J.R. Ferraro, L.J. Basile, Academic Press, New York, 244–282, 1978.
38. J.E. Harris, 'Infrared and Submillimeter Spectroscopy of the Atmosphere', in *Infrared and Millimeter Waves Volume 6 Systems and Components*, ed. K.J. Button, Academic Press, New York, 2–65, 1982.
39. F.A. De Lucia, 'Spectroscopy in the Terahertz Spectral Region', in *Sensing with Terahertz Radiation*, ed. D. Mittleman, Springer Verlag, Berlin, 39–116, 2003.
40. E.J. Heilweil, D.F. Plusquellic, 'Terahertz Spectroscopy of Biomolecules', in *Terahertz Spectroscopy – Principles and Applications*, ed. S.L. Dexheimer, Taylor & Francis, LLC, Boca Raton, 269–298, 2008.
41. M. Claybourn, H. Yang, L. Gradinarsky, J. Johansson, S. Folestad, 'Terahertz Spectroscopy for Pharmaceutical Applications', in *Applications of Vibrational Spectroscopy in Pharmaceutical Research and Development*, eds. D.E. Pivonka, J.M. Chalmers, P.R. Griffith, John Wiley & Sons, Chichester, 337–352, 2007.
42. J.A. Zeitler, T. Rades, P.F. Taday, 'Pharmaceutical and Security Applications of Terahertz Spectroscopy', in *Terahertz Spectroscopy – Principles and Applications*, ed. S.L. Dexheimer, Taylor & Francis, LLC, Boca Raton, 269–298, 2008.
43. M. Koch, 'Bio-Medical Applications of THz Imaging', in *Sensing with Terahertz Radiation*, ed. D. Mittleman, Springer Verlag, Berlin, 295–316, 2003.

Theory of Infrared Spectroscopy

George Turrell
Université des Sciences et Technologies de Lille, Lille, France

1 Introduction **1**
2 The Diatomic Molecule **2**
2.1 Kinetic Energy 2
2.2 Potential Energy 3
2.3 Rigid Rotating Harmonic Oscillator 3
2.4 Vibrating Rotator 4
3 Polyatomic Molecules **4**
3.1 Vibrational Kinetic Energy 4
3.2 Internal Coordinates 5
3.3 Potential Energy 5
3.4 Normal Coordinates 5
3.5 Secular Determinant 6
3.6 The Water Molecule 6
3.7 Molecular Rotation 7
3.8 Classification of Rotators 7
4 Molecular Symmetry **8**
4.1 Symmetry Elements 8
4.2 Group Theory 9
4.3 Coordinate Transformations 10
4.4 Irreducible Representations 10
4.5 Character Tables 11
4.6 Reduction of a Representation: The ''Magic Formula'' 12
4.7 Examples 12
4.8 Symmetry Coordinates 13
4.9 Characteristic Group Frequencies 15
4.10 Spectroscopic Selection Rules 16
4.11 Symmetry and Selection Rules 18
4.12 Absorption Intensities 19
5 Spectra of Gases and Vapors **19**
5.1 Diatomic and Linear Molecules 19
5.2 Spherical Rotators 21
5.3 Symmetric Rotators 21
5.4 Asymmetric Rotators 23
6 Spectra of Neat Liquids and Solutions **23**
6.1 Solvent Effects 23
6.2 Absorption Bandshapes 24
6.3 Band Broadening 24
6.4 The Hydrogen Bond 25
7 Spectra of Solids **26**
7.1 The Linear Diatomic Chain 26
7.2 Three Dimensions: Crystalline Solids 27
7.3 The Factor-group Analysis of Crystals 28
7.4 Polymers 29
7.5 Disordered Solids 30
8 Conclusion **31**
Abbreviations and Acronyms **31**
Related Articles **31**
References **31**

This article is devoted to the fundamentals of infrared spectroscopy. While it is well known that infrared spectroscopy is an extremely important analytical technique, its theoretical basis is not easily assimilated by most analytical chemists. It must be admitted that a certain understanding of abstract concepts, such as those encountered in quantum mechanics, is required. However, with all apologies to the reader, it is the objective here to present the essentials of these ideas, as needed in analytical applications of infrared spectroscopy. It should be noted that most of these principles apply equally to Raman spectroscopy, a subject which is described in other articles.

It is the Born–Oppenheimer approximation that provides the theoretical basis for the analysis of infrared spectra. The rotational and vibrational degrees of freedom of a molecule are specified to allow the spectra of molecules in the gas phase to be interpreted. It is shown that the logical extension of these arguments can be employed in the analysis of the spectra of liquids and crystalline solids.

In later sections, an effort has been made to explain how the basic ideas developed earlier can be applied to the interpretation of the infrared spectra of hydrogen-bonded molecules, polymers and amorphous systems. A brief description of the notion of characteristic group frequencies, as routinely employed in organic chemistry, has been included to allow the practicing chemist to appreciate the possibilities – as well as the limitations – of this most important approach to the subject. To advance his appreciation of the advantages of infrared spectroscopy in analytical chemistry, the reader should refer to the various applications, as listed at the end of this article.

1 INTRODUCTION

The theoretical basis of the infrared spectroscopy of molecular systems is established by the analysis of the dynamics of the nuclear displacements. The Born–Oppenheimer approximation is of course the fundamental argument in the treatment of this problem. As the masses of the electrons and the nuclei in a molecule are significantly different, the much lighter electrons

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

move rapidly to create the so-called electron cloud which ''sticks'' the nuclei into relatively fixed equilibrium positions. The resulting geometry of the nuclear configuration is usually referred to as the molecular structure. It will be shown in this article that this molecular geometry plays an essential role in the determination of the vibrational and rotational spectra, both infrared and Raman, of a given molecule.

The analytical application of infrared spectroscopy can without doubt be founded on the work of Coblentz. His early publication, which has been reprinted, represents the most important contribution to the development of analytical applications of infrared spectroscopy.[1] However, the more practical or ''routine'' use of infrared spectroscopy as an analytical tool had to await the evolution of the necessary instrumentation, for a large part the result of military requirements during World War II. Thus, after a half century the early observations by Coblentz were confirmed and applied to what is now the field of analytical infrared spectroscopy. Initial applications were for the most part limited to the mid-infrared region, from 4000 to approximately $500\,\mathrm{cm}^{-1}$, by the instrumentation available. However, it is just this region that yields spectral features that can be identified, at least approximately, with organic functional groups. Furthermore, it is this region that is important in the analysis of molecular vibrations, as it is here that most vibrational fundamentals are observed. With the introduction of matrix methods, followed by the development of digital computers, the calculation of molecular force fields from observed vibrational spectra became possible. These results have provided invaluable information concerning the nature of the chemical bond.

The objective of this article is to summarize the theoretical basis of this important area of analytical chemistry. It must be admitted that the mathematical treatment of certain aspects can become rather demanding. An effort will be made to reduce these difficulties to a minimum, to allow the practicing analytical spectrocopist to appreciate the bases of the theoretical problem, without being obliged to delve into the mathematical details.

2 THE DIATOMIC MOLECULE

2.1 Kinetic Energy

The vibrational and rotational motions of a free diatomic molecule, as in an ideal gas, can be simply represented by the displacement of two point masses m_1 and m_2 in free space. The kinetic energy of the two-particle system is then given in Cartesian coordinates by Equation (1):

$$T = \tfrac{1}{2}[m_1(\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2) + m_2(\dot{x}_2^2 + \dot{y}_2^2 + \dot{z}_2^2)] \qquad (1)$$

The introduction of the reduced mass μ, defined by Equation (2):

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \qquad (2)$$

allows the kinetic energy to be expressed as a function of the relative positions of the two particles. Furthermore, if the spherical coordinates r, θ, φ with respect to the center of mass are employed to describe the position of a hypothetical particle of mass μ, the Schrödinger equation for this system becomes (Equation 3)

$$-\frac{h^2}{8\pi^2\mu}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right]\psi + V(r)\psi = E\psi \qquad (3)$$

where h is Planck's constant. The usual procedure for the separation of independent variables can then be made by the introduction of Equation (4):

$$\psi(r,\theta,\varphi) = R(r)Y(\theta,\varphi) \qquad (4)$$

The functions $Y(\theta,\varphi)$ are the well-known spherical harmonics. The resulting expression for the r-dependent function is then of the following form (Equation 5):

$$-\frac{h^2}{8\pi^2\mu r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}}{\mathrm{d}r}\right)R(r) + \frac{h^2J(J+1)}{8\mu r^2}R(r) + V(r)R(r) = ER(r) \qquad (5)$$

It is important to note that the second term in brackets contains the factor $J(J+1)$, which arises from the boundary conditions imposed on the spherical harmonics. The requirement that $J = 0, 1, 2, 3, \ldots$, has the effect of quantizing the angular momentum associated with the rotation of the diatomic molecule. Clearly, the solutions to Equation (5) depend on the form of the potential function which, in this case, governs the interaction between the two nuclei. The rotation of the system, which produces the centrifugal force, contributes the energy given by the second term in Equation (5). Thus, the square of the angular momentum is quantized according to Equation (6):

$$\mathcal{M}^2 = \frac{h^2}{4\pi^2}J(J+1) \qquad (6)$$

(see section 3.8).

To take the simplest case, suppose that the distance between the two masses is held constant, $r = r_e$. Then, as $V(r) = 0$, Equation (5) reduces to Equation (7):

$$E_{\mathrm{rot}} = \frac{h^2J(J+1)}{8\pi^2\mu r_e^2} \qquad (7)$$

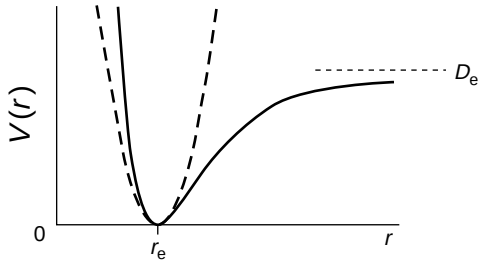

Figure 1 The potential function for a diatomic molecule (solid curve) and the harmonic approximation (dashed curve).

where E_{rot} is the energy of the rigid rotor. This approximation is usually the first step in the interpretation of the pure rotational spectra of diatomic molecules, as observed in the microwave and far-infrared regions.

2.2 Potential Energy

According to the Born–Oppenheimer approximation, the effect of the electron cloud is to produce an interaction potential of the form shown by the solid curve in Figure 1. Although a number of empirical functions exist to represent this function, as a first approach, the harmonic approximation is employed. Here, the development of the potential function $V(r)$ is carried out in a Taylor series in the following form (Equation 8):

$$V(r) = V_0 + \left(\frac{dV}{dr}\right)_{r_e}(r - r_e) + \frac{1}{2}\left(\frac{d^2V}{dr^2}\right)_{r_e}(r - r_e)^2 + \cdots \tag{8}$$

As the origin of the potential energy scale is arbitrary, V_0 can be chosen equal to zero. The second term in Equation (8) vanishes, because the slope of the curve at the minimum, $r = r_e$, is of course zero. The first nonvanishing term is thus the quadratic term in $x = r - r_e$. In the harmonic approximation higher terms in this development are neglected and the potential function becomes (Equation 9)

$$V(r) = \tfrac{1}{2}f(r - r_e)^2 = \tfrac{1}{2}fx^2 \tag{9}$$

where the force constant f can be identified with the curvature of the potential function at the minimum.

The solution to the one-dimensional Schrödinger equation with the potential function given by Equation (9) leads to wave functions expressed as Hermite polynomials and the energy in the following form (Equation 10):

$$E_{\text{vib}} = h\nu_e(v + \tfrac{1}{2}) \tag{10}$$

where $v = 0, 1, 2, \ldots$ is the vibrational quantum number. The quantity $\nu_e = (1/2\pi)\sqrt{f/\mu}$ is equal to the vibrational frequency of the corresponding classical oscillator.[2]

2.3 Rigid Rotating Harmonic Oscillator

With the rotational energy given by Equation (7) and the vibrational energy given by Equation (10), the total energy takes the simple form of Equation (11):

$$E_{J,v} = E_{\text{rot}} + E_{\text{vib}} = \frac{h^2J(J+1)}{8\pi^2\mu r_e^2} + h\nu_e\left(v + \frac{1}{2}\right. \tag{11}$$

The practicing spectroscopist usually writes this expression as Equation (12):

$$\frac{E_{J,v}}{hc} = \bar{\nu}_e\left(v + \frac{1}{2}\right. + \overline{B}_eJ(J+1) \tag{12}$$

where $\overline{B}_e$ is the rotational constant expressed in cm^{-1} and $\bar{\nu}_e$ is the vibrational wavenumber, often referred to (incorrectly) as the vibrational "frequency". Equation (11) is represented by the energy-level diagram presented in Figure 2. The figure is not to scale, as the relative separation between the fundamental vibrational levels $v = 0$ and those of the excited state $v = 1$ is usually much greater than indicated.

A first approximation to the calculation of the rotation–vibration spectrum of a diatomic molecule can be made on the basis of Figure 2. It is sufficient to impose the selection rules $\Delta v = 1$ for vibrational absorption

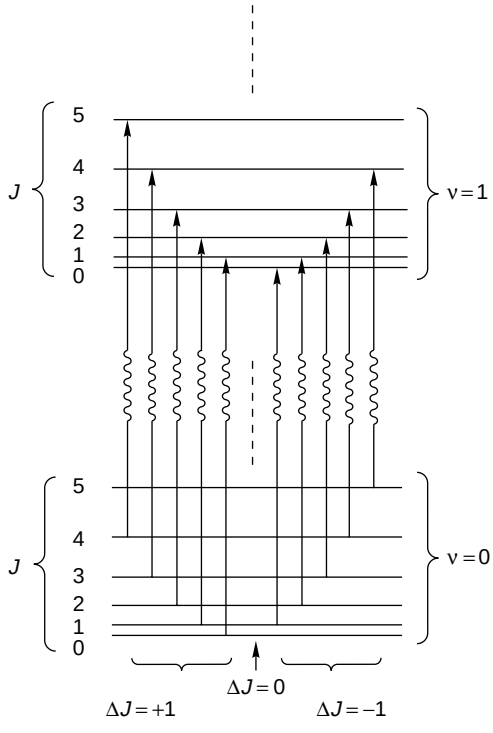

Figure 2 The energy levels for a diatomic rigid rotator.

and $\Delta J = \pm 1$ for the accompanying rotational transitions, as shown. It is important to note that there is no transition that corresponds to the purely vibrational frequency, as the rotational transition $\Delta J = 0$ is forbidden by the rotational selection rules. This analysis indicates that the fundamental transition $v = 0 \Rightarrow v = 1$ should yield a series of equally spaced lines, with a missing element at the center, where $\Delta J = 0$ (see section 5.1).

2.4 Vibrating Rotator

The simple description presented above of the rotation–vibration spectrum of a diatomic molecule was based on the rigid rotator–harmonic oscillator approximation. However, if Equation (12) is generalized to include the anharmonicity of the vibration, in addition to the effects of centrifugal distortion and rotation–vibration coupling, it becomes (Equation 13)

$$\frac{E_{J,v}}{hc} = \bar{\nu}_e \left(v + \frac{1}{2} \right. + \overline{B}_e J(J+1) - \bar{\nu}_e x_e \left(v + \frac{1}{2} \right)^2 - \overline{D} J^2 (J+1)^2 - \bar{\alpha}_e \left(v + \frac{1}{2} \right. J(J+1) \qquad (13)$$

The various parameters appearing in Equation (13) can be evaluated from the rotation–vibration spectra of diatomic molecules or with the use of an empirical function to represent the curve shown in Figure 1. In the latter case, the Morse function (Equation 14)

$$V(r) = D_e[1 - e^{-a(r-r_e)}]^2 \qquad (14)$$

which is one of the simplest, usually yields satisfactory results.[2] In Equation (14) D_e is the energy of dissociation and the parameter a is $\sqrt{f/2D_e}$.

3 POLYATOMIC MOLECULES

Within the framework of the Born–Oppenheimer approximation the electronic energy of a molecule is considered to be a function of its instantaneous nuclear configuration. What is often referred to as the ''electronic problem'' is then the solution of the Schrödinger equation for the movement of the electrons for each fixed configuration of the N atoms in the molecule. The values of the energy as functions of the molecular configuration represent the effective potential function that governs the movement of the nuclei. The solutions to this problem yield the various possible values of the energy E_{nucl} for each electronic state of the molecule.

This energy, associated with the movement of the nuclei, is usually decomposed, at least in a first approximation, in the following form (Equation 15):

$$E_{nucl} = E_{trans} + E_{rot} + E_{vib} + \text{interactions} \qquad (15)$$

The interaction terms are not always negligible, although in analytical applications of vibrational spectroscopy it is usually an acceptable approximation. Furthermore, the translational contribution to the energy is in most cases classical. In other words, it does not have spectroscopic significance. In the case of solids, however – particularly at low frequencies – its effects should not be overlooked.

The conclusions of the above arguments are that for practical purposes the energy of an isolated molecule can be expressed as the sum of two terms, corresponding to its rotational and vibrational energies. This approximation is the initial step in the interpretation of the spectra of gases, as summarized in section 5. The rotation of molecules in condensed phases is restricted by the intermolecular forces. Thus, as a first approximation, the infrared and Raman spectra of liquids and solids can be analyzed with the consideration of only the vibrational energy of the system. However, the intermolecular forces often have important consequences, as they determine to a large extent the shapes of vibrational bands and are responsible for ''solvent effects'' in the spectra of solutions. These questions are considered later.

3.1 Vibrational Kinetic Energy

The vibrational energy of an isolated molecule will now be written as the sum of the kinetic and potential energies (Equation 16):

$$E_{vib} = T + V \qquad (16)$$

A set of $3N$ Cartesian displacement coordinates in the following form (Equation 17):

$$\xi = \begin{matrix} \Delta x_1 \\ \Delta y_1 \\ \Delta z_1 \\ \text{----} \\ \Delta x_2 \\ \Delta y_2 \\ \Delta z_2 \\ \text{----} \\ \cdot \\ \cdot \\ \cdot \\ \text{----} \\ \Delta x_N \\ \Delta y_N \\ \Delta z_N \end{matrix} \qquad (17)$$

can be employed to express the kinetic energy. If a diagonal matrix of the atomic masses is defined by Equation (18):

$$\boldsymbol{M} = \begin{matrix} m_1 & & & & & & & & \\ & m_1 & & & & & & & \\ & & m_1 & & & 0 & & & \\ & & & m_2 & & & & & \\ & & & & m_2 & & & & \\ & & & & & m_2 & & & \\ & & & & & & \cdot & & \\ & & & 0 & & & & m_N & \\ & & & & & & & & m_N \\ & & & & & & & & & m_N \end{matrix} \qquad (18)$$

the kinetic energy in terms of these matrices takes the compact form (Equation 19)

$$T = \tfrac{1}{2}\dot{\tilde{\boldsymbol{\xi}}}\boldsymbol{M}\dot{\boldsymbol{\xi}} \qquad (19)$$

where the dot over the vector defined by Equation (17) indicates that the time derivative has been taken and the tilde represents the transpose operation.

3.2 Internal Coordinates

It is usually convenient to work with a set of internal displacement coordinates, $\boldsymbol{S}$, as they have chemical significance. In the limit of small amplitudes of atomic displacements, the two sets of coordinates are linearly related. Thus (Equation 20),

$$\boldsymbol{S} = \boldsymbol{B}\boldsymbol{\xi} \qquad (20)$$

where the matrix $\boldsymbol{B}$ is determined by the equilibrium geometry of the molecule. Equation (19) allows the kinetic energy to be expressed in internal coordinates as Equation (21):

$$T = \tfrac{1}{2}\dot{\boldsymbol{S}}\boldsymbol{G}^{-1}\dot{\boldsymbol{S}} \qquad (21)$$

where the matrix $\boldsymbol{G}$ is defined by $\boldsymbol{G} \equiv \boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{B}$. Although, in principle, the $\boldsymbol{G}$ matrix can be constructed with a knowledge of the atomic masses and the molecular geometry, tables of its elements are available that were developed many years ago.[3] In present-day computer programs, it is constructed and employed directly to calculate the vibrational frequencies and the specific forms of the normal modes of vibration, as summarized in the following sections.

3.3 Potential Energy

Although the equilibrium configuration of a molecule can usually be specified, at ordinary temperatures all of the atoms undergo oscillatory motions. The forces between the atoms in the molecule are described by a Taylor series of the intramolecular potential function in the internal coordinates. This function can then be written in the following form (Equation 22):

$$V = \frac{1}{2}\sum_{t,t'}\left(\frac{\partial^2 V}{\partial S_t \partial S_{t'}}\right)_0 S_t S_{t'} = \frac{1}{2}\sum_{t,t'} f_{t,t'} S_t S_{t'} \qquad (22)$$

an expression that includes the first nonzero terms in the expansion of the potential function with respect to the equilibrium configuration. Clearly, Equation (22) is a generalization of the harmonic potential for the diatomic molecule (Equation 9). Although higher terms in this series can be added to take into account the anharmonicity of the molecular vibrations, the harmonic approximation is employed in most practical calculations of the vibrational frequencies of a polyatomic molecule. Equation (22) can be written in matrix form as Equation (23):

$$V = \tfrac{1}{2}\boldsymbol{S}\boldsymbol{F}\boldsymbol{S} \qquad (23)$$

where $\boldsymbol{F}$ is a symmetric matrix of the force constants. Although the force constants can in principle be calculated from the molecular electronic wave functions, the results are not yet quantitatively reliable. These quantities are usually treated as variable parameters in the vibrational problem. Their evaluation from experimental spectroscopic data is of importance, both as a means of testing the results of ab initio electronic calculations and for their chemical significance. For example, the knowledge of the value of a force constant associated with the stretching of a chemical bond provides some insight into the nature of the bond, including semiquantitative measures of both its length and its dissociation energy.

3.4 Normal Coordinates

It is now fundamental to define the normal coordinates of this vibrational system – that is to say, the nuclear displacements in a polyatomic molecule. Again in the limit of small amplitudes of vibration, the normal coordinates in the form of the vector $\boldsymbol{Q}$ are related to the internal coordinates by a linear transformation, viz. (Equation 24)

$$\boldsymbol{S} = \boldsymbol{L}\boldsymbol{Q} \qquad (24)$$

The kinetic and potential energies given by Equations (21) and (23) become (Equation 25)

$$T = \tfrac{1}{2}\dot{\boldsymbol{S}}\boldsymbol{G}^{-1}\dot{\boldsymbol{S}} = \tfrac{1}{2}\dot{\boldsymbol{Q}}\boldsymbol{L}\boldsymbol{G}^{-1}\boldsymbol{L}\dot{\boldsymbol{Q}} = \tfrac{1}{2}\dot{\boldsymbol{Q}}\boldsymbol{E}\dot{\boldsymbol{Q}} \qquad (25)$$

and (Equation 26)

$$V = \tfrac{1}{2}\boldsymbol{SFS} = \tfrac{1}{2}\boldsymbol{QLFLQ} = \tfrac{1}{2}\boldsymbol{Q\Lambda Q} \tag{26}$$

respectively. The matrices $\boldsymbol{E}$ and $\boldsymbol{\Lambda}$ are both diagonal. The former is the unit matrix, while the latter is composed of elements $\boldsymbol{\lambda}_k = 4\pi^2 v_k^2$. The vibrational frequencies are given by v_k, where the subscript k identifies the normal mode of vibration. An inspection of Equations (25) and (26) identifies the role of the normal coordinates, namely, to eliminate simultaneously all cross-terms in the expressions for the kinetic and potential energies. This condition can be taken as the definition of the normal modes of vibration.

3.5 Secular Determinant

The last step in the calculation of the frequencies of molecular vibrations, as observed in the infrared spectra, is carried out by combining Equations (25) and (26). The vibrational energy of a polyatomic molecule is then given in this, the harmonic approximation, by Equation (27):

$$E_{\text{vib}} = T + V = \tfrac{1}{2}\dot{\boldsymbol{Q}}\boldsymbol{E}\dot{\boldsymbol{Q}} + \tfrac{1}{2}\boldsymbol{Q\Lambda Q} \tag{27}$$

From Equations (25) and (26), it is apparent that $\boldsymbol{LG^{-1}L = E}$ and $\boldsymbol{LFL = \Lambda}$. With the definition of the **G**-matrix, $\boldsymbol{L = L^{-1}G}$. Therefore (Equation 28),

$$\boldsymbol{L^{-1}GFL = \Lambda} \tag{28}$$

Hence, as is often stated, the determination of the normal coordinates is equivalent to the successful search for a matrix $\boldsymbol{L}$ that diagonalizes the product $\boldsymbol{GF}$ via a similarity transformation. This system of linear, simultaneous, homogeneous equations can be written in the following form (Equation 29):

$$\sum_{t'}[(\boldsymbol{GF})_{tt'} - \delta_{tt'}\lambda_k]L_{t'k} = 0 \tag{29}$$

where the Kronecker delta is defined by Equation (30):

$$\delta_{tt'} = \begin{cases} 1, & \text{if } t = t' \\ 0, & \text{if } t \neq t' \end{cases} \tag{30}$$

This set of equations for the elements of $\boldsymbol{L}$ can be resolved by application of Cramer's rule. Then, a nontrivial solution exists only if the determinant of the coefficients vanishes, or (Equation 31)

$$|\boldsymbol{GF} - \boldsymbol{E}\lambda_k| = 0 \tag{31}$$

This condition on the so-called secular determinant is the basis of the vibrational problem. The roots of Equation (31), λ_k, are the eigenvalues of the matrix product $\boldsymbol{GF}$, while the columns of $\boldsymbol{L}$, the eigenvectors, determine the forms of the normal modes of vibration. These relatively abstract relations become more evident with the consideration of an example.

3.6 The Water Molecule

As a simple illustration of the development of the secular determinant, consider the water molecule. A reasonable set of internal coordinates consists of the changes in lengths of the two bonds and the variation in the bond angle. Thus, from Equation (20) and Figure 3 (Equation 32)

$$\boldsymbol{S} = \begin{pmatrix} \Delta r_1 \\ \Delta r_2 \\ \Delta\alpha \end{pmatrix} = \boldsymbol{B}\boldsymbol{\xi}$$

$$= \left(\begin{array}{ccc:ccc:ccc} 0 & -s & -c & 0 & 0 & 0 & 0 & s & c \\ 0 & 0 & 0 & 0 & s & -c & 0 & -s & c \\ 0 & -c/r & s/r & 0 & c/r & s/r & 0 & 0 & -2s/r \end{array}\right) \begin{matrix} \Delta x_1 \\ \Delta y_1 \\ \Delta z_1 \\ \hdashline \Delta x_2 \\ \Delta y_2 \\ \Delta z_2 \\ \hdashline \Delta x_3 \\ \Delta y_3 \\ \Delta z_3 \end{matrix} \tag{32}$$

where $s = \sin\alpha/2$, $c = \cos\alpha/2$ (α being the equilibrium bond angle) and r is the equilibrium O–H bond length, as shown in Figure 3. The masses of the hydrogen and oxygen atoms are arranged according to Equation (18) in the order established by the vector $\boldsymbol{\xi}$, namely (Equation 33),

$$\boldsymbol{M} = \begin{pmatrix} m_{\text{H}} & & & & & & & & \\ & m_{\text{H}} & & & & & & & \\ & & m_{\text{H}} & & & 0 & & & \\ & & & m_{\text{H}} & & & & & \\ & & & & m_{\text{H}} & & & & \\ & & & & & m_{\text{H}} & & & \\ & & 0 & & & & m_{\text{O}} & & \\ & & & & & & & m_{\text{O}} & \\ & & & & & & & & m_{\text{O}} \end{pmatrix} \tag{33}$$

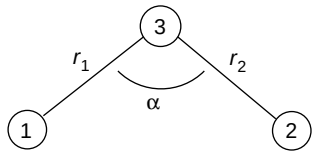

Figure 3 Internal coordinates for the water molecule.

The $\boldsymbol{G}$ matrix, as calculated from its definition, becomes (Equation 34)

$$\boldsymbol{G} = \boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{B}$$

$$= \begin{matrix} \mu_H + \mu_O & \mu_O \cos\alpha & -\dfrac{\mu_O \sin\alpha}{r} \\ \mu_O \cos\alpha & \mu_H + \mu_O & -\dfrac{\mu_O \sin\alpha}{r} \\ -\dfrac{\mu_O \sin\alpha}{r} & -\dfrac{\mu_O \sin\alpha}{r} & \dfrac{2\mu_H + 2\mu_O(1-\cos\alpha)}{r^2} \end{matrix} \tag{34}$$

where $\mu_H = 1/m_H$ and $\mu_O = 1/m_O$.

The force-constant matrix based on internal coordinates is of the following form (Equation 35):

$$\boldsymbol{F} = \begin{matrix} f_r & f_{rr} & f_{r\alpha} \\ f_{rr} & f_r & f_{r\alpha} \\ f_{r\alpha} & f_{r\alpha} & f_{\alpha\alpha} \end{matrix} \tag{35}$$

where f_r is the coefficient of $(\Delta r_1)^2$ and $(\Delta r_2)^2$, and f_α is the coefficient of $(\Delta\alpha)^2$ in the potential energy expression. These constants are referred to as the principal or valence force constants. The off-diagonal constant f_{rr} represents an interaction constant which is the coefficient of $(\Delta r_1 \Delta r_2)$ and $(\Delta r_2 \Delta r_1)$ in the potential energy expression, while the interaction constant $f_{r\alpha}$ is the coefficient of all four terms of the type $(\Delta r_1 \Delta\alpha)$. By taking the product of the above matrices $\boldsymbol{G}$ and $\boldsymbol{F}$, the secular equation in the form of Equation (31) can be readily found.

In the construction of the matrix $\boldsymbol{F}$ of Equation (35), the symmetrical equivalence of the two O–H bonds was taken into account. Nevertheless, it contains four independent force constants. As the water molecule has only three fundamental vibrational frequencies, at least one interaction constant must be neglected or some other constraint introduced. If all of the off-diagonal elements of $\boldsymbol{F}$ are neglected, the two principal constants, f_r and f_α, constitute the valence force field (VFF) for this molecule. However, to reproduce the three observed vibrational frequencies this force field must be modified to include the interaction constant f_{rr}. This example will be reconsidered in section 4.7, after the question of molecule symmetry has been treated.

3.7 Molecular Rotation

The rotational energy of a molecule can be written in the following form (Equation 36):

$$E_{\text{rot}} = \tfrac{1}{2}\boldsymbol{\omega}\boldsymbol{I}\boldsymbol{\omega} \tag{36}$$

where it has been assumed that the molecule acts as a rigid rotator. In effect, in the limit of infinitesimal vibrational amplitudes, the moment of inertia tensor, $\boldsymbol{I}$, which becomes equal to that of a rigid rotator, is of the following form (Equation 37):

$$\boldsymbol{I} = \begin{matrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{matrix} \tag{37}$$

The Cartesian coordinates employed in Equation (37) have been assumed to be attached to the rigid, rotating molecule. It should be noted that the moment of inertia tensor is symmetric with respect to the principal diagonal; thus, $I_{xy} = I_{yx}$, etc. The angular velocity associated with the rotation of the molecule is represented here by the vector $\boldsymbol{\omega}$. Equation (36) can then be written out in the following form (Equation 38):

$$E_{\text{rot}} = \tfrac{1}{2}[I_{xx}w_x^2 + I_{yy}w_y^2 + I_{zz}w_z^2 + 2I_{xy}w_x w_y + 2I_{yz}w_y w_z + 2I_{zx}w_z w_x] \tag{38}$$

It is in general possible to find a Cartesian coordinate system attached to the molecule such that the tensor I takes a diagonal form. In terms of these so-called principal axes Equation (38) is simplified in that all cross-terms are eliminated (Equation 39):

$$E_{\text{rot}} = \tfrac{1}{2}(I_a w_a^2 + I_b w_b^2 + I_c w_c^2) \tag{39}$$

The definition of the principal axes labeled a, b, c is usually dictated by the symmetry properties of the molecule (see section 4.5).

3.8 Classification of Rotators

With the use of principal axes it becomes possible to classify the types of rotator as follows:

1. linear: $I_a = I_b \neq I_c = 0$;
2. spherical: $I_a = I_b = I_c \neq 0$;
3. symmetric: $I_a = I_b \neq I_c \neq 0$;
4. asymmetric: $I_a \neq I_b \neq I_c$.

Within the framework of the rigid-rotor approximation, it then becomes possible to express in closed form the energy of each of the first three types of rotator. From Equation (36) with the angular velocity given by $\boldsymbol{\omega} = \begin{pmatrix} w_a \\ w_b \\ w_c \end{pmatrix}$, the rotational energy can be written in the following form (Equation 40):

$$E_{\text{rot}} = \frac{1}{2}\boldsymbol{\omega}\boldsymbol{I}\boldsymbol{\omega} = \frac{1}{2}\left(\frac{{}^{2}_{a}}{I_a} + \frac{{}^{2}_{b}}{I_b} + \frac{{}^{2}_{c}}{I_c}\right. \tag{40}$$

where ${}_a$, ${}_b$ and ${}_c$ are the components of the angular momentum with respect to the principal axes a, b and c, respectively.[3,4]

In the quantum-mechanical treatment of the rotational problem, it is shown that the square of the total angular momentum is quantized,[3,4] namely (Equation 41)

$$\mathcal{M}^2 = \quad {}^2_a + \quad {}^2_b + \quad {}^2_c = \frac{h^2}{4\pi^2} J(J+1) \qquad (41)$$

where the rotational quantum number is defined by $J = 0, 1, 2, 3, \ldots$ (see section 2.1).

3.8.1 Linear Rotators

If c is chosen as the unique axis of a linear molecule, $\quad_c = 0$ and the rotational energy is given by Equation (42):

$$E_{\text{rot}} = \frac{h^2}{8\pi^2 I} J(J+1) \qquad (42)$$

where $I = I_a = I_b$ is the moment of inertia of the molecule. It should be noted that there is a degeneracy equal to $2J + 1$, as the quantum number $M = 0, \pm 1, \pm 2, \ldots, \pm J$.

3.8.2 Spherical Rotators

If the principal moments of inertia are equal, viz. $I = I_a = I_b = I_c$, Equation (36) leads to the same expression for the energy as given by Equation (42), i.e. (Equation 43)

$$E_{\text{rot}} = \frac{h^2}{8\pi^2 I} J(J+1) \qquad (43)$$

where, as above, I is the moment of inertia of the molecule at equilibrium. Then, as $\quad_c = (h/2\pi)K$, where the additional quantum number $K = 0, \pm 1, \pm 2, \ldots, \pm J$, the resulting degeneracy is equal to $(2J + 1)^2$.

3.8.3 Symmetric Rotators

In this case the principal moments are related by $I = I_a = I_b \neq I_c \neq 0$, where again the c-axis has been chosen as the unique direction. The classical rotational energy can then be given as Equation (44):

$$E_{\text{rot}} = \frac{{}^2_a + \quad {}^2_b}{2I_a} + \frac{{}^2_c}{2I_c} = \frac{\mathcal{M}^2}{2I_a} + \frac{{}^2_a}{2}\left(\frac{1}{I_c} - \frac{1}{I_a}\right. \qquad (44)$$

With the use of the relations given above, the quantum-mechanical expression becomes (Equation 45)

$$E_{\text{rot}} = \frac{h^2}{8\pi^2 I_a} J(J+1) + \frac{h^2 K^2}{8\pi^2}\left(\frac{1}{I_c} - \frac{1}{I_a}\right. \qquad (45)$$

Equation (45) allows two types of symmetric rotator to be defined. If $I_c < I_a$, the rotator is elongated in the direction of the c-axis. It is said to be a prolate top, as, say, an American football. The second term on the right-hand side of Equation (45) is then positive. On the other hand, if $I_c > I_a$, the rotator is flattened with respect to the c direction, as the earth – or in the limit, a pancake. In this case the second term on the right-hand side of Equation (45) is negative and the rotator is an oblate top. These two types of rotator can usually be distinguished in their rotation–vibration spectra by the contribution of the second term. The degeneracy for this class of rotator is rather complicated, viz. $2J + 1$ if $K = 0$ and $2(2J + 1)$ if $|K| > 0$ (see section 5.3).

3.8.4 Asymmetric Rotators

In the most general case, all three principal moments of inertia are different: $I_a \neq I_b \neq I_c$. Unfortunately, most molecules fall into this class. No analytical expression can be written for the rotational energy of an asymmetric rotator. However, extensive tables for the energy levels were calculated many years ago. More recently, computer programs have been developed to carry out their numerical evaluation for a given molecular structure.

4 MOLECULAR SYMMETRY

Although all molecules are in constant thermal motion, when all of their atoms are in their equilibrium positions, a specific geometrical structure can usually be defined. The normal modes of vibration of such a structure, and hence an interpretation of its vibrational spectrum, can then be calculated – at least in the harmonic approximation.

4.1 Symmetry Elements

The mathematical description of the structure of a molecule is based on an application of group theory. It is necessary, then, to define certain symmetry operations. Although there are various ways of specifying these operations, geometrical descriptions are certainly the easiest to understand. Each symmetry operation (aside from the trivial one, E, the identity) is associated with an element of symmetry. Thus, for example, certain molecules are said to be planar. Well-known examples are water, boron trifluoride and benzene, whose structures can be drawn in the forms shown in Figure 4(a–c).

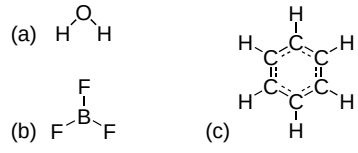

Figure 4 Planar molecules: (a) water; (b) boron trifluoride; (c) benzene.

The symmetry operations and their descriptions, and the notation established by Schönflies, are given in Table 1. To identify the symmetry group it is necessary to visualize all of the operations which characterize the structure of the molecule. As there is always at least one point in space that is not affected by any of these operations, the group is referred to as a point group. Usually with experience, and the aid of molecular models, the symmetry operations are not too difficult to find.

The molecules shown in Figure 4(a–c) are planar; thus, the paper on which they are drawn is an element of symmetry and the reflection of all points through the plane yields an equivalent (superposable) structure. The process of carrying out the reflection is referred to as the symmetry operation σ. However, as the atoms of these molecules are essentially point masses, the reflection operations result simply in the inversion of the direction (the coordinate) perpendicular to the plane.

In general, a molecule may have a number of elements of symmetry. For example, the water molecule has, in addition to the plane of symmetry indicated above, a second perpendicular to the first. The reflection operation with respect to this plane inverts the y direction, as well as interchanging the identification of the hydrogen atoms (see Figure 5).

The presence of two perpendicular planes of symmetry implies the existence of another element of symmetry,

Table 1 Operations used to describe molecular symmetry

Symmetry operation	Symbol	Symmetry element
Identity	E	–
Reflection in a plane	σ	Plane of symmetry
Rotation k times by $\theta = 2\pi/n$	C_n^k	Proper rotation axis of order n
Rotation by $\theta = 2\pi/n$ followed by a reflection in the plane $\perp$ to the axis	S_n^k	Improper rotation axis of order n
Inversion with respect to a center	$i \equiv S_{2k}^k$ (k odd)	Center of symmetry

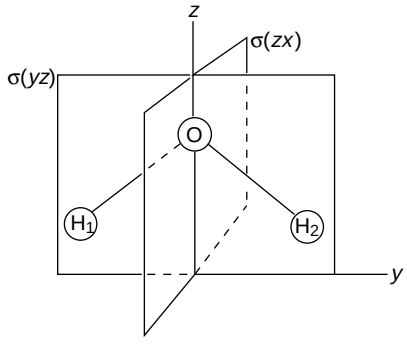

Figure 5 The planes of symmetry in the water molecule.

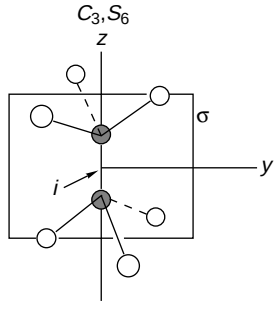

Figure 6 The symmetry elements of the ethane molecule.

the twofold axis. It is an example of a binary rotation axis, denoted by the symbol C_2, which is in this case collinear with the z-axis in Figure 5.

Examples of the other symmetry elements shown in Table 1 are illustrated in Figure 6. For the ethane molecule in the staggered configuration, the vertical (z) axis is collinear with the ternary axis C_3 and a rotation–reflection axis, S_6. The symmetry of this molecule is also characterized by a center of inversion, i, as well as the plane of symmetry σ, as shown (Figure 6), and two similar ones designated σ' and σ''.

It should be apparent that more than one operation can be associated with each symmetry element. In the case of the ethane molecule considered above, the rotation about the C_3 axis by 120° can be carried out once or twice, as C_3^1 or C_3^2, respectively. However, the operation C_3^3 need not be considered as it corresponds to the identity, E, defined below. Similarly, after some reflection it is found that $S_6^2 \equiv C_3^1$, $S_6^3 \equiv i$ and $S_6^4 \equiv C_3^2$.

4.2 Group Theory

Without entering into the mathematical details, it is important to point out that the ensemble of symmetry operations that characterizes the structure of a molecule forms a group.[5,6] That is,

1. The combination of any two members of the ensemble is an element of the ensemble.
2. There is one element of the ensemble which permutes with all of the others and leaves them unchanged. It is known as the identity, E (German: Einheit).
3. The associative law of combination is valid.
4. Each element of the ensemble has an inverse that is also a member of the ensemble.

In (1), the combination of two members of the ensemble, that is, two symmetry operations, implies their successive application. It should be noted that in general these operations do not commute. The number of independent symmetry operations in a group is called the order of the group.

4.3 Coordinate Transformations

In the foregoing discussion, the symmetry operations have been described as ''rotations'', ''reflections'', etc. that is, as dynamic operations. However, strictly, symmetry operations are really coordinate transformations that modify the mathematical statement of the atomic positions, rather than the positions themselves. Furthermore, these transformations are subject to the condition that they cannot modify the potential energy of the molecule. This property follows from the fact that any coordinate system is merely a bookkeeping scheme without physical significance.

Returning to the water molecule as an example, the Cartesian displacement coordinates as given in Equation (32), are of the following form (Equation 46):

$$\boldsymbol{\xi} = \begin{matrix} \Delta x_1 \\ \Delta y_1 \\ \Delta z_1 \\ \text{----} \\ \Delta x_2 \\ \Delta y_2 \\ \Delta z_2 \\ \text{----} \\ \Delta x_3 \\ \Delta y_3 \\ \Delta z_3 \end{matrix} \tag{46}$$

for this triatomic system. A symmetry operation $\boldsymbol{R}$ results in a coordinate transformation (Equation 47):

$$\boldsymbol{\xi}' = \boldsymbol{R}\boldsymbol{\xi} \tag{47}$$

to form the vector $\boldsymbol{\xi}'$. Specifically, for the reflection $\boldsymbol{R} = \sigma(zx)$ the matrix representation is given by Equation (48):

$$\boldsymbol{R} = \left[\begin{array}{ccc:ccc:ccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hdashline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{48}$$

The nine-by-nine matrix that represents a given operation $\boldsymbol{R}$ is composed of three-by-three submatrices, as shown. Those submatrices that lie along the principal diagonal contain nonzero elements if the coordinates of the atom in question are not permuted by the operation. Thus, in this example it is apparent that in Equation (48) the submatrix associated with the oxygen atom remains along the diagonal, while those for the two equivalent hydrogen atoms are displaced.

It is perhaps not surprising that the specific form of a matrix representation, such as that given in Equation (48), depends on the chosen basis coordinates – in this case the elements of the vector $\boldsymbol{\xi}$. However, there is always one quantity that is invariant, namely the trace of the representation matrix, that is, the sum of its diagonal elements. In group theory it is known as the character, as it characterizes the representation.

Each submatrix $\boldsymbol{R}_i$ appearing in a representation such as $\boldsymbol{R}$ in the above example can be written in the following general form (Equation 49):

$$\boldsymbol{R}_i = \begin{bmatrix} \cos\theta_j & -\sin\theta_j & 0 \\ \sin\theta_j & \cos\theta_j & 0 \\ 0 & 0 & \pm 1 \end{bmatrix} \tag{49}$$

where $\theta_j = 2\pi/n$ is the angle of (counterclockwise) rotation for a given symmetry operation (or class of symmetry operation), as given in Table 1. Then, for each atom i the contribution to the character of the representation is given by Equation (50):

$$\text{Trace } \boldsymbol{R}_i = (2\cos\theta_j \pm 1) \tag{50}$$

where in Equations (49) and (50) the positive sign is employed for the so-called proper operations (E and C_n) and the negative sign for the improper ones (σ, S_n and i). Clearly, for the example of the operation σ, Equation (50) yields $2\cos 0 - 1 = 1$ for the trace of the submatrix for the oxygen atom (see Equation 48). This result can be generalized to obtain the following expression for the character of the representation of any symmetry operation (Equation 51):

$$\chi_j = m_j(2\cos\theta_j \pm 1) \tag{51}$$

where m_j is the number of atoms in the molecule that are not permuted by the symmetry operations of class j. This relation, which is independent of any choice of basis coordinates, will be applied repeatedly in the following applications.

4.4 Irreducible Representations

As indicated above, there may be many – even an infinity – of matrix representations of a given symmetry operation, depending on the choice of basis coordinates. However, there exists a particular set of basis coordinates in terms of which the representation matrix is reduced to block-diagonal form. This result is shown symbolically in Figure 7. Each submatrix $\boldsymbol{\Gamma}^{(k)}$ is called an irreducible representation (IR), as there exist no

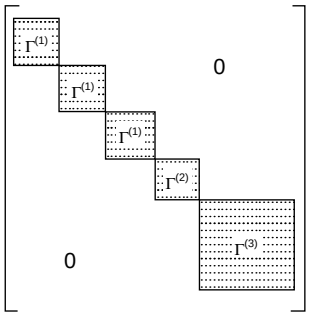

Figure 7 The form of a reduced representation.

basis coordinates that can result in a further reduction. Here, for example, the arbitrarily chosen first IR occurs three times, the second once and the third once again. In this figure the smaller squares ($k = 1, 2$) indicate one-dimensional matrices, while $\mathbf{\Gamma}^{(3)}$ is an IR of order two.

The result shown schematically in Figure 7 can be expressed mathematically in the following form (Equation 52):

$$\mathbf{\Gamma} = \sum_{\gamma} n^{(\gamma)} \mathbf{\Gamma}^{(\gamma)} \tag{52}$$

if it is understood that the summation is special. It indicates that the direct sum is to be taken, that is, the various IRs are arranged along the diagonal of the reduced representation. Each IR, identified by the superscript (γ), may appear more than once, or not at all. Thus, $n^{(\gamma)}$ is equal to the number of times that the IR $\mathbf{\Gamma}^{(\gamma)}$ appears in the reduced representation.

While it is not, in general, necessary to find the explicit forms of the IR matrices, their characters are readily available from standard tables. This information is sufficient for the resolution of most problems encountered in molecular spectroscopy.

4.5 Character Tables

The tables of the characters of the IRs have the general form shown in Table 2. For a point group the columns are headed by the symbols $_j$, the classes of symmetry operation. Each class may be composed of one or more symmetry operations, although by tradition the arbitrarily chosen first class consists uniquely of the identity operation, E. The rows are labeled by the identification of the IRs. In a given IR the operation E corresponds to a unit matrix whose order is equal to the dimension of the representation. Hence the resulting character is equal to the dimension of the representation and the dimension of each IR can be easily determined by inspection of the first column of characters in the table. It should be noted that the matrix of characters is square, that is, the number of IRs is equal to the number of classes of symmetry operation.

A further property of the character tables arises from the fact that every symmetry group has an IR that is invariant under all of the group operations. This IR is a one-by-one matrix composed of the number one for every class of operation. Obviously, the characters, are all then equal to one. As this IR is by convention taken to be $\mathbf{\Gamma}^{(1)}$, the first row of each character table consists solely of ones. The significance of the character tables will become more apparent by consideration of an example.

The water molecule in its equilibrium configuration is represented in Figure 5. The symmetry elements $\sigma(zx)$ and $\sigma(yz)$ are shown. The intersection of these two planes is the z-axis, which is collinear with the binary axis of symmetry, C_2. The symmetry operations with respect to these elements, plus of course the identity E, define the group $_{2v}$ in the notation of Schönflies. It is then said that the water molecule "belongs" to the symmetry group $_{2v}$.

The water molecule has $3N = 9$ degrees of freedom. Of these, three describe the position in space of the center of mass of the molecule, as specified by a vector with components T_x, T_y and T_z. Similarly, the angular momentum vector associated with the rotation of the molecule, is identified by its components R_x, R_y and R_z. These coordinates, which represent the six external degrees of freedom, are classified for molecules of $_{2v}$ symmetry in Table 3.

The classification of rotators made in section 3.8 can now be reconsidered in terms of the symmetry elements defined in Table 1. If a molecule has one symmetry

Table 2 The general form of a character table

	$_1 = E$	$_2$	$_3$	$\cdots$	$_k$
$\mathbf{\Gamma}^{(1)}$	$\chi_1^{(1)}$	$\chi_2^{(1)}$	$\chi_3^{(1)}$	$\cdots$	$\chi_k^{(1)}$
$\mathbf{\Gamma}^{(2)}$	$\chi_1^{(2)}$				
$\mathbf{\Gamma}^{(3)}$	$\chi_1^{(3)}$				
$\vdots$	$\vdots$				
$\mathbf{\Gamma}^{(k)}$	$\chi_1^{(k)}$	$\cdots$	$\cdots$	$\cdots$	$\chi_k^{(k)}$

Table 3 The character table for the point group $_{2v}$ and the analysis of the water molecule

$_{2v}$	E	C_2	$\sigma\,(zx)$	$\sigma\,(yz)$		$n^{(\gamma)}$ (tot)	$n^{(\gamma)}$ (vib)
$\mathbf{\Gamma}^{(1)} = A_1$	1	1	1	1	T_z	3	2
$\mathbf{\Gamma}^{(2)} = A_2$	1	1	−1	−1	R_z	1	0
$\mathbf{\Gamma}^{(3)} = B_1$	1	−1	1	−1	T_x, R_y	2	0
$\mathbf{\Gamma}^{(4)} = B_2$	1	−1	−1	1	T_y, R_x	3	1
χ_j	9	−1	1	3			

axis of threefold or higher symmetry, it is a symmetric rotator. Two angular momentum components will appear together, e.g. (R_x, R_y) in a doubly degenerate IR. If the molecule has two or more axes of greater than twofold symmetry, it is a spherical rotator, and all three components will be grouped in a three-dimensional IR. Clearly, the water molecule is an example of an asymmetric rotator.

4.6 Reduction of a Representation: The "Magic Formula"

The relation between a reducible representation and the IRs is given by Equation (52). In practical applications, it is essential to determine the values of $n^{(\gamma)}$, the number of times that each IR enters the direct sum shown in Figure 7. As χ_j is merely the sum of the diagonal elements of Γ for each class of operation j, it is also equal to the sum of the traces of the individual submatrices $\Gamma^{(\gamma)}$, each multiplied by $n^{(\gamma)}$. It follows that (Equation 53)

$$\chi_j = \sum_{\gamma=1}^{k} n^{(\gamma)} \chi_j^{(\gamma)} \quad j = 1, 2, \ldots, k \tag{53}$$

where $\chi_j^{(\gamma)}$ is the character of the IR $\Gamma^{(\gamma)}$ of an operation of the class $_j$. This set of simultaneous equations can be solved with aid of the orthogonality properties of the characters. The result is given by Equation (54):

$$n^{(\gamma)} = \frac{1}{g} \sum_{j=1}^{k} g_j \chi_j^{(\gamma)^*} \chi_j \tag{54}$$

This expression is of such widespread applicability that it is often referred to as the "magic formula". In it, g is the order of the group, that is, the total number of independent symmetry operations. The characters for each class of operation $\chi_j^{(\gamma)}$, are given in the character table for the symmetry group of the molecule, while χ_j is the corresponding character of the reducible representation. The latter can be calculated, for example, from Equation (51). Each product in the summation appearing in Equation (54) is weighted by the quantity g_j, the number of independent symmetry operations in each class.[3,6]

4.7 Examples

In spite of the apparent complexity of the magic formula, its application is fairly simple, as shown in the following examples. The water molecule was described above, where it was indicated that it is of symmetry $_{2v}$. The character table is presented in Table 3, where it is seen that the order of the group is equal to four and that each class is composed of only a single symmetry operation. The characters for the reducible representation, as calculated with the use of Equation (51), are shown in the last line of Table 3. Thus, Equation (54) yields the values of $n^{(\gamma)}$(tot) given in the first column to the right of the character table. This result indicates that the symmetry of the $3N$ degrees of freedom of the water molecule (or any triatomic molecule of symmetry $_{2v}$) can be described by the direct sum ($\oplus$) of IRs (Equation 55):

$$\Gamma = 3A_1 \oplus A_2 \oplus 2B_1 \oplus 3B_3 \tag{55}$$

As indicated above, for the $3N - 6$ normal vibrations of a nonlinear molecule in free space, the coordinates of translation and rotation can be removed. The symmetry of these external coordinates, T_x, T_y, T_z and R_x, R_y, R_z, is given in each character table. The difference for this example, $2A_1 \oplus B_2$, describes the symmetry of the three normal modes of vibration of the molecule. Two of them are symmetric with respect to all four of the symmetry operations of the point group. However, the mode of species B_2 is antisymmetric under the operations C_2 and $\sigma(zx)$, as shown in Table 3.

As a second example, the symmetry elements of the ethane molecule in its staggered conformation are illustrated in Figure 6. The ensemble of 12 symmetry operations forms the group $_{3d}$, whose character table is presented in Table 4. Note that in this case the IRs E_g and E_u are doubly degenerate, as indicated by the number two in the column headed E, the identity. The characters of the reducible representations are given in the last line. As in the previous example, the structure of the reduced representation is determined by the values of $n^{(\gamma)}$(tot) determined by application of the magic formula. The removal of the external coordinates leads to the values of $n^{(\gamma)}$(vib) shown in the last column. The values of $n^{(\gamma)}$(vib) given for the doubly degenerate species are the number of pairs of normal modes. Thus, for example, there are three pairs of normal modes of symmetry E_u, corresponding to three distinct vibrational frequencies. Note that the subscripts g and u (German: gerade, ungerade) specify the symmetry with respect to the inversion operation, as indicated in Table 4.

Table 4 Character table for the group $_{3d}$ and the analysis of the normal modes of vibration of ethane

$_{3d}$	E	$2C_3$	$3C_2$	i	$2S_6$	$3\sigma_d$		$n^{(\gamma)}$ (tot)	$n^{(\gamma)}$ (vib)
A_{1g}	1	1	1	1	1	1		3	3
A_{2g}	1	1	−1	1	1	−1	R_z	1	0
E_g	2	−1	0	2	−1	0	(R_x,R_y)	4	3
A_{1u}	1	1	1	−1	−1	−1		1	1
A_{2u}	1	1	−1	−1	−1	1	T_z	3	2
E_u	2	−1	0	−2	1	0	(T_x,T_y)	4	3
χ_j	24	0	0	0	0	4			

4.8 Symmetry Coordinates

The characters χ_j for the examples in the previous section were calculated with the use of Equation (51), that is, on the basis of Cartesian displacement coordinates. Alternatively, it is often desirable to employ a set of internal coordinates as the basis. However, they must be well chosen so that they are sufficient to describe the vibrational degrees of freedom of the molecule and that they are linearly independent. The latter condition is necessary to avoid the problem of redundancy. Even when properly chosen, the internal coordinates still do not usually transform following the symmetry of the molecule. Once again, the water molecule provides a very simple example of this problem.

The internal coordinates for the water molecule are defined in Figure 3. The effect of each symmetry operation of the symmetry group $_{2v}$ on these internal coordinates is specified in Table 5. Clearly, the internal coordinate $\Delta\alpha$ is totally symmetric, as the characters $\chi_j(\Delta\alpha)$ correspond to those given for the IR A_1. On the other hand, the characters $\chi_j(\Delta r)$, as shown, cannot be identified with a specific IR. By inspection of Table 5, however, it is apparent that the direct sum $A_1 \oplus B_2$ corresponds to the correct symmetry of these coordinates. In more complicated cases the magic formula can always be employed to achieve the correct reduction of the representation in question.

From the above analysis, as shown in the last column of Table 5, it should be obvious that two linear combinations of the internal coordinates Δr can be formed whose symmetry corresponds to the IRs A_1 and B_2. They are $\Delta r_1 + \Delta r_2$ (A_1) and $\Delta r_1 - \Delta r_2$ (B_2), as can be verified by inspection of the character table.

Coordinates such as these, which have the symmetry properties of the point group, are known as symmetry coordinates. As they transform in the same manner as the IRs when used as basis coordinates, they factor the secular determinant into block-diagonal form.[3] Thus, while normal coordinates must be found to diagonalize the secular determinant, the factorization resulting from the use of symmetry coordinates often provides considerable simplification of the vibrational problem. Furthermore, symmetry coordinates can be chosen a priori by a simple analysis of the molecular structure.

Table 5 Symmetry of the internal coordinates for the water molecule

$_{2v}$	E	C_2	$\sigma(zx)$	$\sigma(yz)$	$n^{(\gamma)}(\Delta\alpha)$	$n^{(\gamma)}(\Delta r)$
A_1	1	1	1	1	1	1
A_2	1	1	−1	−1	0	0
B_1	1	−1	1	−1	0	0
B_2	1	−1	−1	1	0	1
$\chi_j(\Delta\alpha)$	1	1	1	1		
$\chi_j(\Delta r)$	2	0	0	2		

In the example considered above, $\Delta r_1 - \Delta r_2$ is the only symmetry coordinate of species B_2. Thus, it results in a factor of degree one in the completely reduced secular determinant. It is therefore a normal coordinate. On the other hand, the two normal coordinates of species A_1 are linear combinations of the symmetry coordinates $\Delta\alpha$ and $\Delta r_1 + \Delta r_2$. They can only be found by solution of the secular equations.

It is usually convenient to normalize the symmetry coordinates. Hence, for the example considered here, the three symmetry coordinates take the form (Equation 56)

$$\text{species } A_1: \quad \begin{aligned} \mathsf{S}_1 &= \Delta\alpha \\ \mathsf{S}_2 &= \tfrac{1}{\sqrt{2}}(\Delta r_1 + \Delta r_2) \end{aligned} \tag{56}$$

and (Equation 57)

$$\text{species } B_2: \mathsf{S}_3 = \tfrac{1}{\sqrt{2}}(\Delta r_1 - \Delta r_2) \tag{57}$$

where the normalizing factor $_p$ is determined by the condition (Equation 58)

$$_p{}^2 \sum_t \mathsf{S}_t^2 = 1 \tag{58}$$

for each symmetry coordinate S_p. In general form, the relations given by Equations (56) and (57) are expressed by the matrix relation (Equation 59)

$$\mathsf{S} = \boldsymbol{U}\boldsymbol{S} \tag{59}$$

which for the above example becomes (Equation 60)

$$\begin{pmatrix} \mathsf{S}_1 \\ \mathsf{S}_2 \\ \mathsf{S}_3 \end{pmatrix} = \begin{matrix} 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{matrix} \begin{pmatrix} \Delta r_1 \\ \Delta r_2 \\ \Delta\alpha \end{pmatrix} \tag{60}$$

It should be noted that when the symmetry coordinates have been normalized, $\boldsymbol{U}$ is orthogonal, that is, $\boldsymbol{U} = \boldsymbol{U}^{-1}$. The matrices $\boldsymbol{U}$ can be constructed with the use of projection operators. For lack of space, this method will not be presented here.

To show how use is made of symmetry coordinates as the bases of the vibrational problem, reconsider the kinetic and potential energies as given earlier, e.g. (Equation 61)

$$2T = \dot{\boldsymbol{S}}\boldsymbol{G}^{-1}\boldsymbol{S} = \dot{\boldsymbol{Q}}\boldsymbol{E}\dot{\boldsymbol{Q}} \tag{61}$$

and (Equation 62)

$$2V = \boldsymbol{S}\boldsymbol{F}\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q} \tag{62}$$

From Equation (59), we have (Equation 63)

$$\boldsymbol{S} = \boldsymbol{U}^{-1}\mathsf{S} = \boldsymbol{U}\mathsf{S} \tag{63}$$

hence (Equation 64)

$$2T = \dot{\mathsf{S}}\mathsf{G}^{-1}\dot{\mathsf{S}} = \dot{\boldsymbol{Q}}\boldsymbol{E}\dot{\boldsymbol{Q}} \tag{64}$$

and (Equation 65)

$$2V = \mathsf{SFS} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q} \tag{65}$$

In Equations (64) and (65) the definitions $\mathsf{G} \equiv \boldsymbol{UGU}$ and $\mathsf{F} \equiv \boldsymbol{UFU}$ have been introduced. These equations lead to the secular determinant as before (Equation 66):

$$|\mathsf{GF} - \boldsymbol{E}\lambda_k| = 0 \tag{66}$$

The eigenvectors L with this basis can be found from $\mathsf{L} = \boldsymbol{UL}$.

Returning now to the example of the water molecule, with $\boldsymbol{F}$ given by Equation (35) and $\boldsymbol{U}$ in Equation (60), the transformation $\boldsymbol{UFU}$ gives Equation (67):

$$\mathsf{F} = \left[\begin{array}{cc:c} f_\alpha & \sqrt{2}f_{r\alpha} & 0 \\ \sqrt{2}f_{r\alpha} & f_r + f_{rr} & 0 \\ \hdashline 0 & 0 & f_r - f_{rr} \end{array}\right] \tag{67}$$

Similarly, $\boldsymbol{UGU}$ leads to Equation (68):

$$\mathsf{G} = \left[\begin{array}{cc:c} 2\mu_H/r^2 + 2\mu_O \times\dfrac{(1-\cos\alpha)}{r^2} & -\dfrac{\sqrt{2}\mu_O \sin\alpha}{r} & 0 \\ -\dfrac{\sqrt{2}\mu_O \sin\alpha}{r} & \mu_H + \mu_O \times(1+\cos\alpha) & 0 \\ \hdashline 0 & 0 & \mu_H + \mu_O \times(1-\cos\alpha) \end{array}\right] \tag{68}$$

As both F and G are partitioned by the use of symmetry coordinates, the secular determinant is factored accordingly. The problem of calculating the vibrational frequencies is thus divided into two parts: solution of a linear equation for the single frequency of species B_2 and of a quadratic equation for the pair of frequencies of species A_1.

On the other hand, it is often of interest to calculate force constants from observed vibrational frequencies. However, it is not generally possible to derive analytical expressions for the force constants as functions of the frequencies and the molecular geometry. The calculation is necessarily an iterative one. Starting with a set of assumed force constants – usually obtained by analogy with similar bonds in other molecules – the values are refined until a suitable set is found. The set that yields the best agreement between calculated and observed frequencies constitutes the accepted force field for the molecule.

From the above example, it is apparent that there may be (and usually are) more unknown force constants than observed frequencies. If no additional sources of data are available, it is necessary to make some assumptions to simplify the force field. Often some or all of the off-diagonal elements in the $\boldsymbol{F}$ matrix (Equation 23) are neglected, leading to the so-called VFF or modified valence force field (MVFF), respectively.

Assuming that a reasonable force field is known, the solution of the above equations to obtain the vibrational frequencies of water is not difficult. However, in more complicated molecules it very rapidly becomes a formidable one. If there are N atoms in the molecule, there are $3N$ total degrees of freedom and $3N - 6$ for the vibrational frequencies. The molecular symmetry can often aid in simplifying the calculations, although in large molecules there may be no true symmetry. In some cases the notion of local symmetry can be introduced to simplify the calculation of vibrational frequencies and the corresponding forms of the normal modes of vibration.

Regardless of the force field chosen, the calculation of vibrational frequencies by the method outlined above is based on the harmonic approximation. Tabulated values of force constants can be used to calculate vibrational frequencies, for example, of molecules whose vibrational spectra have not been observed. However, as anharmonicities have been neglected in the above analysis, the resulting frequency values are often no better than ±5% with respect to those observed.

For the example of the water molecule, it is of interest to calculate the forms of the vibrational modes, as obtained from the evaluation of the matrix $\mathsf{L} = \boldsymbol{UL}$. The results can be presented most simply in the form shown in Figure 8. The calculation of the specific form of the normal modes is relatively complicated, although with the aid of current computer programs it becomes routine – at least for relatively simple molecules.

It is apparent from Figure 8 that the normal modes of vibration of the water molecule, as calculated from the eigenvectors, can be described approximately as a symmetrical stretching vibration (ν_1) and a symmetrical bending vibration (ν_2). It should be emphasized that the two normal modes of species A_1 are linear combinations of the two internal symmetry coordinates given by Equations (56) and (57). On the other hand, as there is only one symmetry coordinate of species B_2 in this

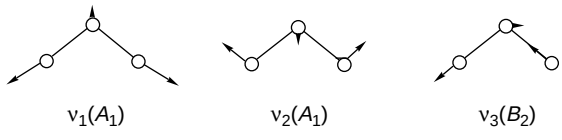

Figure 8 The normal modes of vibration of the water molecule.

case, it is identical with the normal coordinate shown for ν_3 and can be correctly described as an antisymmetric stretching vibration.

It has been shown that the potential energy distribution provides an approximate method to evaluate the relative contribution of each symmetry coordinate to a given normal mode of vibration. From the definition of the symmetry coordinates (Equation 59), the relation (Equation 69)

$$\mathsf{LFL} = \Lambda \tag{69}$$

or (Equation 70)

$$\sum_{pp'} \mathsf{L}_{pk}\mathsf{L}_{p'k}\mathsf{F}_{pp'} = \lambda_k \quad k = 1, 2, \ldots, 3N-6 \tag{70}$$

can be derived. As the diagonal elements of F are usually at least an order of magnitude greater than the off-diagonal elements, the relation (Equation 71)

$$\frac{1}{\lambda_k}\sum_{p} \mathsf{L}_{pk}^2\mathsf{F}_{pp} \approx 1 \quad k = 1, 2, \ldots, 3N-6 \tag{71}$$

is approximately equivalent to Equation (70). Each term on the left-hand side of Equation (71) represents the approximate contribution of each symmetry coordinate S_p to the normal coordinate Q_k. The evaluation of this potential-energy distribution, and also the form of the normal vibrations, is usually carried out by current computer programs for the calculation of vibrational frequencies.

4.9 Characteristic Group Frequencies

Organic chemists are familiar with the use of infrared spectroscopy to identify functional groups in an unknown compound.[7] This method of qualitative analysis is possible because the infrared absorption bands of many common functional groups vary but little from one molecule to another. For example, the band assigned to the carbonyl group, C=O, in ketones is usually found in the range $1710 - 1720\,\mathrm{cm}^{-1}$. Similarly, in aldehydes it is between 1715 and $1735\,\mathrm{cm}^{-1}$ and in organic acids from 1750 to $1770\,\mathrm{cm}^{-1}$. The assignment of these bands to the carbonyl group implies that the normal modes of vibration corresponding to the observed frequencies are localized, that is, they involve only (or at least primarily) the internal coordinate that is the variation in the length of the C=O bond. Such assignments, which have been established empirically, are very useful, although they should be applied with caution. In this particular case the vibrational frequency associated with the carbonyl group is characteristic because the force constant of the carbon–oxygen double bond has approximately twice the value of the nearby single bonds. Substitution on an adjacent carbon atom will, however, produce frequency shifts, as indicated above.

From a theoretical point of view, if the matrices $\boldsymbol{G}$ and $\boldsymbol{F}$ are expressed in internal coordinates, the vibrational frequencies are determined from the eigenvalues of their product (Equation 31). In the case of a carbonyl compound, the change in the C=O distance could be chosen as the internal coordinate associated with the element $(\boldsymbol{GF})_{11}$. In the limit as either off-diagonal element $(\boldsymbol{GF})_{12}$ or $(\boldsymbol{GF})_{21}$ is much smaller than the diagonal element $(\boldsymbol{GF})_{11}$, the chosen internal coordinate becomes a normal coordinate, that is, it is ''uncoupled'' from the rest of the molecule. The normal coordinate is then localized and the observed frequency can be correctly described as a carbonyl vibration.

A more complicated example of characteristic group frequencies can be illustrated by consideration of the methyl group, CH_3, as shown in Figure 9. It is assumed in this model that the methyl group is attached to the molecular framework, represented here by an infinite mass. If the symmetry of the methyl group is not modified by the presence of the molecular framework, the point group is $_{3v}$ and the analysis can be carried out as illustrated in section 4.8. The results are summarized in Table 6, where the figures given in the last line were calculated with the use of Equation (51). Application of the magic formula yields the reduced representation for the $3N = 15$ degrees of freedom of the methyl group attached to the mass X. Removal of the external coordinates leads to the result shown in the last column for the vibrational degrees of freedom, $3A_1 \oplus 3E$.

The 10 internal coordinates for the attached methyl group can be combined to form the symmetry coordinates

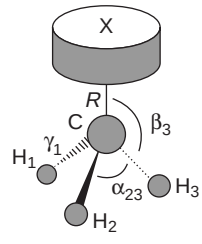

Figure 9 The methyl group attached to an infinite mass X.

Table 6 The character table for the point group $_{3v}$ and the analysis of the methyl group

$_{3v}$	E	$2C_3$	$3\sigma_v$		$n^{(\gamma)}$ (tot)	$n^{(\gamma)}$ (vib)
A_1	1	1	1	T_z	4	3
A_2	1	1	−1	R_z	1	0
E	2	−1	0	(T_x, T_y) (R_x, R_y)	5	3
χ_j	15	0	3			

given by Equation (72):

$$\text{species } A_1: \quad \begin{aligned} \mathsf{S}_1 &= \tfrac{1}{\sqrt{3}}(\Delta r_1 + \Delta r_2 + \Delta r_3) \\ \mathsf{S}_2 &= \tfrac{1}{\sqrt{6}}(\Delta\alpha_{23} + \Delta\alpha_{31} + \Delta\alpha_{12} \\ &\quad - \Delta\beta_1 - \Delta\beta_2 - \Delta\beta_3) \\ \mathsf{S}_3 &= \Delta R \end{aligned} \tag{72}$$

and Equation (73):

$$\text{species } E: \quad \begin{aligned} \mathsf{S}_{4a} &= \tfrac{1}{\sqrt{6}}(2\Delta r_1 - \Delta r_2 - \Delta r_3) \\ \mathsf{S}_{5a} &= \tfrac{1}{\sqrt{6}}(2\Delta\alpha_{23} - \Delta\alpha_{31} - \Delta\alpha_{12}) \\ \mathsf{S}_{6a} &= \tfrac{1}{\sqrt{6}}(2\Delta\beta_1 - \Delta\beta_2 - \Delta\beta_3) \\ \mathsf{S}_{4b} &= \tfrac{1}{\sqrt{2}}(\Delta r_2 - \Delta r_3) \\ \mathsf{S}_{5b} &= \tfrac{1}{\sqrt{2}}(\Delta\alpha_{31} - \Delta\alpha_{12}) \\ \mathsf{S}_{6b} &= \tfrac{1}{\sqrt{2}}(\Delta\beta_2 - \Delta\beta_3) \end{aligned} \tag{73}$$

It is then apparent that the three normal modes of vibration of species A_1 are linear combinations of the three symmetry coordinates given by Equation (72). The three frequencies corresponding to these three normal modes are by convention designated by ν_1, ν_2 and ν_3. The symmetry coordinates of species E given by Equation (73) consist of three pairs, identified by the subscripts a and b. The coordinates labeled a are symmetric with respect to the plane defined by H_1, C and X, while the three labeled b are antisymmetric. Appropriate linear combinations of the three members of each pair are the normal coordinates of this species. Either set of three symmetry coordinates, when substituted into the secular determinant, yields the corresponding vibrational wavenumbers of the modes ν_4, ν_5 and ν_6. Representative results are summarized in Table 7.

The fourth column in Table 7 presents an evaluation of the contribution of each symmetry coordinate to a given normal mode of vibration of the methyl group. It is obvious that the often-used descriptions, as given in the last column of Table 7, are only approximate. It should be noted that the quantitative evaluation of the contribution of each symmetry coordinate to a given normal mode can be estimated by the calculation of the potential energy distribution (see Equation 71).

Table 7 The normal modes of vibration of an isolated methyl group

Mode	Species	Wavenumber (cm^{-1})	Symmetry coordinates	Approx. description
ν_1	A_1	2880	S_1 (100%)	Sym. CH stretch
ν_2	A_1	1250–1400[a]	S_2 (~80%)	Sym. CH bend
ν_3	A_1	~530[a]	S_3 (60–80%), S_2 (~30%)	CX stretch
ν_4	E	2960	S_4 (100)%	Antisym. CH stretch
ν_5	E	1450	S_5 (~80%), S_6 (~20%)	Antisym. CH bend
ν_6	E	960	S_6 (80%), S_5 (~20%)	CH_3 rocking

[a] The wavenumber of these modes, and also the relative contributions of the different symmetry coordinates, are very sensitive to the value of the force constant for the CX bond.

4.10 Spectroscopic Selection Rules

The relation between classical and quantum mechanics was pointed out in section 2.2. There it was shown that the difference between successive energy levels for the harmonic oscillator in quantum mechanics, when divided by Planck's constant, yields the value of the vibrational frequency of the equivalent classical oscillator. This result is an example of the correspondence principle. In the preceding section the vibration of polyatomic molecules is treated in a purely classical manner. With the use of normal modes the molecular vibrations become separated, each possessing a characteristic vibrational frequency. It then becomes possible to consider each normal mode as a quantum-mechanical oscillator, thanks to the correspondence principle. This fact is of extreme importance in vibrational spectroscopy, as it permits the calculation of the harmonic frequencies of vibration of the molecules with the use of the methods of classical mechanics (see section 3.5). The extension of the model to include the anharmonicity provides a basis for the qualitative interpretation of observed overtones and combinations of the fundamental vibrational frequencies.

In absorption spectroscopy the electric vector $\boldsymbol{\mathcal{E}}$ of the incident radiation couples with the electric dipole moment $\mathbf{m}$ of the molecular system. The resulting time-dependent perturbation of the Hamiltonian (Equation 74),

$$\boldsymbol{H}' = -\mathbf{m} \cdot \boldsymbol{\mathcal{E}} \tag{74}$$

induces a transition in the molecule and a corresponding absorption of energy. The probability per second of the transition from the state n to the state m is given by the Einstein coefficient for absorption, which can be written in the following form (Equation 75):

$$B_{n\to m} = \frac{8\pi^3}{3h^2} \; \langle m|\mathrm{m}_X|n\rangle^2 + \langle m|\mathrm{m}_Y|n\rangle^2 + \langle m|\mathrm{m}_Z|n\rangle^2 \tag{75}$$

This result is obtained by application of time-dependent perturbation theory.[6] It also applies to the processes of spontaneous and induced emission. It should be noted that the factor before the brackets on the right-hand side of Equation (75) depends on the units employed and is unimportant for the purposes of this article.

In Equation (75) the notation of Dirac has been used to represent the appropriate integrals. Thus, for example (Equation 76),

$$\langle m|\mathrm{m}_X|n\rangle = \int \psi_m^* \mathrm{m}_X \psi_n \, \mathrm{d}\tau \tag{76}$$

where ψ_n and ψ_m are the wave functions for the initial and final states, respectively, and $\mathrm{d}\tau$ represents the volume element in the space in question. For readers familiar with quantum mechanics it is apparent that when $n = m$, Equation (76) expresses the average value of the x-component of the dipole moment of the molecule in that state. However, when $n \neq m$ the integral in Equation (76) is known as the transition moment for this component. This, and the two analogous terms in Equation (75), determine the selection rules for this transition. Clearly, if all three transition moments are equal to zero, the transition cannot take place – it is said to be forbidden. The vanishing (or nonvanishing) of an integral of this type can be established by consideration of the symmetry of the integrand. It is here again that group theory, as outlined in section 4.2, plays an important role.[5]

In the rigid-rotator approximation, the energy was written as the sum given by Equation (11). To the same degree of approximation the wave function for a free molecule is simply the product in Equation (77):

$$\psi \approx \psi(\mathrm{rot})\psi(\mathrm{vib}) \tag{77}$$

The dipole moment in Equation (74) has been expressed in space-fixed coordinates X, Y, Z in the equations that follow. However, it is necessary to establish a Cartesian coordinate system x, y, z attached to the molecule and related to the space-fixed system by Equation (78):

$$\mathbf{m}(X, Y, Z) = \mathbf{\Phi}\mathbf{m}(x, y, z) \tag{78}$$

where $\mathbf{\Phi}$ is a three-by-three matrix whose elements are direction cosines connecting the various pairs of axes.[4] With the use of Equations (77) and (78) the transition moment for the component m_X becomes (Equation 79)

$$\begin{aligned}
\langle m|\mathrm{m}_X|n\rangle = {} & \int \psi_m^*(\mathrm{rot})\Phi_{Xx}\psi_n(\mathrm{rot})\,\mathrm{d}\tau_{\mathrm{rot}} \\
& \times \int \psi_m^*(\mathrm{vib})\mathrm{m}_x\psi_n(\mathrm{vib})\,\mathrm{d}\tau_{\mathrm{vib}} \\
& + \int \psi_m^*(\mathrm{rot})\Phi_{Xy}\psi_n(\mathrm{rot})\,\mathrm{d}\tau_{\mathrm{rot}} \\
& \times \int \psi_m^*(\mathrm{vib})\mathrm{m}_y\psi_n(\mathrm{vib})\,\mathrm{d}\tau_{\mathrm{vib}} \\
& + \int \psi_m^*(\mathrm{rot})\Phi_{Xz}\psi_n(\mathrm{rot})\,\mathrm{d}\tau_{\mathrm{rot}} \\
& \times \int \psi_m^*(\mathrm{vib})\mathrm{m}_z\psi_n(\mathrm{vib})\,\mathrm{d}\tau_{\mathrm{vib}}
\end{aligned} \tag{79}$$

To separate completely the integrals over rotational and vibrational coordinates, each component of the dipole moment is expanded about the equilibrium configuration in a Taylor's series in normal coordinates. That is (Equation 80),

$$\mathrm{m}_x = \mathrm{m}_x^0 + \sum_{k=1}^{3N-6} \left(\frac{\partial \mathrm{m}_x}{\partial Q_k}\right)_0 Q_k + \cdots \tag{80}$$

and similarly for the other components. The higher terms in Equation (80), which account for the so-called electrical anharmonicity, are almost invariably neglected. Expressions of the form of Equation (80) for the various components are substituted into the three equations for the matrix elements, Equation (79) being one of them. Then, for example, the first term of Equation (79) becomes (Equation 81)

$$\begin{aligned}
\langle m|\mathrm{m}_{Xx}|n\rangle = {} & \mathrm{m}_x^0 \int \psi_m^*(\mathrm{rot})\Phi_{Xx}\psi_n(\mathrm{rot})\,\mathrm{d}\tau_{\mathrm{rot}} \\
& + \sum_{k=1}^{3N-6} \left(\frac{\partial \mathrm{m}_x}{\partial Q_k}\right)_0 \int \psi_m^*(\mathrm{vib})Q_k\psi_n(\mathrm{vib})\,\mathrm{d}\tau_{\mathrm{vib}} \\
& \times \int \psi_m^*(\mathrm{rot})\Phi_{Xx}\psi_n(\mathrm{rot})\,\mathrm{d}\tau_{\mathrm{rot}}
\end{aligned} \tag{81}$$

where the integral over the vibrational coordinates normalizes to unity in the first term. Evaluation of integrals such as the first of Equation (81) results in the rotational selection rules that govern pure-rotational transitions in the microwave and far-infrared spectra of gases. These transitions involve no change in the vibrational state of the molecule.

The second term in Equation (81) contains the product of two integrals, one over rotational coordinates and the other over vibrational coordinates. Therefore, the nonvanishing of this term places simultaneous restrictions on changes of rotational and vibrational states. It is the origin of the selection rules for rotation–vibration absorption bands in the infrared spectral region. The evaluation of the integrals over rotational coordinates depends on the type of rotator involved, being relatively straightforward for linear or symmetric tops (see section 3.8).

The integrals over the vibrational coordinates in Equation (81) can be easily calculated in the harmonic approximation. In this case the properties of the Hermite polynomials yield directly the vibrational selection rule $\Delta v_k = \pm 1$, where v_k is the vibrational quantum number for the kth normal mode. As most molecules are in their vibrational ground state ($v_k = 0$ for all values of k) at ordinary temperatures, the most important features in the absorption spectrum of a molecule are those corresponding to the transitions $v_k = 0 \Rightarrow v_k = 1$ for a given k. This process is the absorption of but a single

quantum of vibrational energy. The absorption band observed in the infrared region is then referred to as a vibrational fundamental.

The selection rule given above, $\Delta v_k = \pm 1$, applies in the harmonic approximation; it will break down when anharmonicity is important. The effect of anharmonicity is (i) to modify the regular spacing of energy levels and (ii) to relax the harmonic selection rule. Thus, in the spectra of real molecules, absorption features are often observed at frequencies of approximately $2\nu_k$, $3\nu_k$, etc. the overtones, and at approximately $\ell\nu_k \pm \ell'\nu_k(\ell, \ell' = 1, 2, \ldots)$, the combinations. The quantitative analysis of these effects requires the addition of anharmonic terms in the potential energy. They have been neglected in treatments of all but the simplest (triatomic) polyatomic molecules.

4.11 Symmetry and Selection Rules

The vanishing (or nonvanishing) of an integral such as those given in Equation (79) can be determined by consideration of the symmetry properties of the integrand. It is sufficient to recall that in elementary calculus the symmetry (odd or even) of a function determines the value of its integral between symmetric limits. The same principle can be generalized and applied to the present problem. Thus, the symmetry properties of an integral such as Equation (82)

$$\psi_m^*(\text{vib})\text{m}_x\psi_n(\text{vib})\,\text{d}\tau_{\text{vib}} \tag{82}$$

can be established by consideration of the symmetry of each factor in the integrand. This approach to the problem is general, although in certain cases, particularly when degenerate wavefunctions are involved, it can be complicated. The interested reader is referred to the book by Wilson, Decius and Cross for the analysis of this problem.[3]

In its simplest form, as the IR of each factor in the integrand of Equation (82) can be established with reference to the symmetry group of the molecule, the overall symmetry of the integrand can be determined. If the IR for each factor has been found, their direct product describes the resulting symmetry. This representation, which can then be reduced, viz. (Equation 83)

$$\mathbf{\Gamma}^{(\gamma'')} \otimes \mathbf{\Gamma}^{(\gamma')} \otimes \mathbf{\Gamma}^{(\gamma)} = \sum_{\ell} n^{(\ell)}\mathbf{\Gamma}^{(\ell)} \tag{83}$$

is used to determine the selection rule for any transition $n \Rightarrow m$. In Equation (83), $\mathbf{\Gamma}^{(\gamma)}$ and $\mathbf{\Gamma}^{(\gamma'')}$ are the IRs of the initial and final states, respectively, while in this example the x-component of the dipole moment belongs to the IR $\mathbf{\Gamma}^{(\gamma)}$. The general rule follows. If the reduced representation, as given by the right-hand side of Equation (83) does not contain the totally symmetric IR, $\Gamma^{(1)}$, the integral given by Equation (82), is equal to zero. The transition $n \Rightarrow m$ involving the x-component of the dipole moment is then forbidden. It should be noted that the initial state n is not necessarily the ground state of the molecule.

As an illustration of the analysis outlined in the preceding paragraph, consider a fundamental absorption band of the ethane molecule. The symmetry of this molecule, which was described in section 4.7, leads to the character table given as Table 4. The structure of the reduced representation is then established with the use of the magic formula. These results are shown in the last two columns of Table 4. For fundamental transitions, those which originate from the (always) totally symmetric ground state of the molecule, their symmetry is thus described by $\mathbf{\Gamma}^{(\gamma)} = \mathbf{\Gamma}^{(1)}$. From Table 4 it is apparent that there are two normal modes of symmetry A_{2u}. The symmetry of the dipole moment of the molecule is characterized by the components T_z, of symmetry A_{2u}, and the pair of components (T_x, T_y), of symmetry E_u.

If, for example, a fundamental transition such as $A_{1g} \Rightarrow A_{2u}$ is considered, Equation (81) becomes Equation (84):

$$A_{1g} \otimes A_{2u} \otimes A_{2u} = A_{1g} \tag{84}$$

as can be easily established by multiplying the characters given in Table 4. Here it has been assumed that it is the z-component of the dipole moment that plays an active role in the transition. For this example, then, these fundamental transitions are allowed, as the right-hand side of Equation (84) contains the totally symmetric IR, A_{1g}.

For the three vibrational fundamentals of species E_u, only the x, y coordinates of the dipole moment can provide infrared activity. This result is apparent from Equation (85):

$$A_{1g} \otimes E_u \otimes E_u = A_{1g} \oplus A_{2g} \oplus E_g \tag{85}$$

which is established by taking the appropriate products of the characters given in Table 4 and making the reduction of the representation with the use of the magic formula. As Equation (85) contains the totally symmetric representation A_{1g}, these vibrational fundamentals of ethane are active in its infrared spectrum. The other normal modes of this molecule, as summarized in the right-hand column of Table 4, cannot be infrared active.

The symmetry arguments presented in the preceding two paragraphs are general. They apply not only to the vibrational fundamentals, but also to the overtones and combinations. However, for the vibrational fundamentals they lead to a very simple rule, namely that for a normal mode to produce a fundamental transition it must belong to an IR which also contains a component of the

dipole moment. The more general method of analysis, as described above, is presented in more detail in a number of advanced texts.[3,6,8]

4.12 Absorption Intensities

From the above discussion, it is apparent that the possibility of absorption of radiation by a molecule depends on its dipole moment in the case of pure-rotational spectra and on its dipole moment derivatives for vibrational spectra. Hence it is not surprising that the observed intensities of absorption lines or bands can be used as quantitative measures of the molecular dipole moment and its derivatives.

If a beam of light of intensity falls on a layer of sample material of infinitesimal thickness $d\ell$, the decrease in intensity of the beam is given by Lambert's law (Equation 86):

$$-d \quad = a \quad d\ell \tag{86}$$

where a is the absorption coefficient. This expression can be integrated to yield

$$\ln \frac{}{0} = -a\ell \tag{87}$$

where ℓ is the thickness of the sample and $_0$ is the intensity of the radiation incident on it. As the absorption coefficient, a, is proportional to $N_n B_{n\Rightarrow m}$, light absorption depends on the number of molecules present in the initial state of the system, in addition to the nonvanishing of the Einstein coefficient. The former condition is, of course, the origin of Beer's law, the basis of the application of absorption spectroscopy in quantitative analysis. It is important to note that from a theoretical point of view, the absorption coefficient is proportional to the area of a given absorption band and not necessarily to its peak height. This question will be treated in section 6 in connection with the analysis of band shapes.

5 SPECTRA OF GASES AND VAPORS

The rotational transitions which accompany a given vibrational transition in a polyatomic molecule produce a rotational structure on the infrared absorption band. For molecules with relatively large moments of inertia, this structure is often not observed, particularly at relatively low spectral resolution. However, small molecules in the gaseous phase yield spectra with easily resolved fine structure that can provide considerable information.[9]

5.1 Diatomic and Linear Molecules

The energy of a rotating, vibrating diatomic molecule was considered in section 2.3. There it was shown that its energy is given by Equation (12), at least in the rigid-rotor approximation. This model was then generalized to take into account vibrational anharmonicity, rotation–vibration coupling and centrifugal distortion (Equation 13).

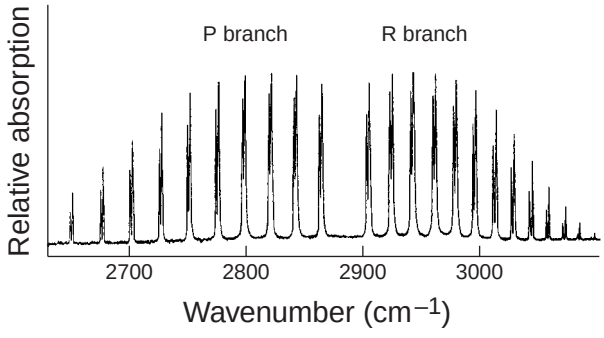

Figure 10 The fundamental rotation–vibration spectrum of gaseous HCl.

The fundamental rotation–vibration band of HCl is presented in Figure 10 as an example. The series of lines on the low-frequency side constitute the P branch, which arises from the selection rule $\Delta J = -1$. The analogous features on the high-frequency side form the R branch ($\Delta J = +1$). This spectrum was obtained under medium resolution. However, it is sufficient to show that each "line" in the rotational fine structure is in fact a doublet. This splitting of each line is due to the isotope effect. Specifically, the observed spectrum can be considered to be the superposition of two spectra, those of $H^{35}Cl$ and $H^{37}Cl$. Those rotational lines that are at lower frequencies are due to the isotopic species with the higher moment of inertia. The relative intensities of the two series of lines is a direct measure of the isotopic ratio for the chlorine atom in natural abundance. The gap in the middle of the spectrum is a manifestation of the selection rules, for which $\Delta J = 0$ is forbidden.

Those polyatomic molecules for which, in the equilibrium configuration, all of the atoms lie along a straight line, can be analyzed similarly. In this case there are $3N - 5$ normal modes of vibration, each with a wavenumber $\bar{\nu}_k$. The vibrational energy is given by Equation (88):

$$\frac{E_{\text{vib}}}{hc} = \sum_{k=1}^{3N-5} \bar{\nu}_k \left(v_k + \frac{1}{2} \right. \tag{88}$$

in the harmonic approximation, where $v_k = 0, 1, \ldots, 3N - 5$. However, in the case of degeneracy, in which two modes have the same frequency, two identical terms in Equation (88) will be obtained. Thus, it is convenient to rewrite this equation in the following form (Equation 89):

$$\frac{E_{\text{vib}}}{hc} = \sum_{i} \bar{\nu}_i \left(v_i + \frac{d_i}{2} \right. \tag{89}$$

where the summation is now over all wavenumbers and the degree of degeneracy of each mode is equal to d_i.

The addition of the rotational contribution to the energy, as given by the second term on the right-hand side of Equation (12), to Equation (89) leads to Equation (90):

$$\frac{E_{\text{vib}} + E_{\text{rot}}}{hc} = \sum_i \overline{\nu}_i \left(v_i + \frac{d_i}{2} \right) + \overline{B}_e J(J+1) \quad (90)$$

in the rigid-rotor, harmonic oscillator approximation. Once again, this result can be generalized to include the effects of anharmonicity, centrifugal distortion and rotation–vibration coupling. It should be noted that consideration of the anharmonicity will introduce cross-terms between the various modes, which, strictly, are no longer normal modes of vibration. Furthermore, a relatively more important contribution to rotation–vibration coupling arises in this case from Coriolis forces for which a term must be added to take into account the angular momentum around the molecular axis. This term, which enters for modes involving bending of the molecule, requires the definition of an additional quantum number, $\ell = 1, 2, \ldots$. In the case of bending modes the condition imposed on this quantum number, $J = \ell, \ell + 1, \ell + 2 \ldots$, results in certain missing rotational lines and in splitting of others (ℓ doubling) due to the removal of the twofold degeneracy. While these effects are essential in the quantitative analysis of the rotation–vibration spectra of small molecules, they become less important as the moment of inertia increases (and hence the value of $\overline{B}_e$ decreases).

As a first approximation, the rotation–vibration spectrum of a linear molecule can be interpreted on the basis of Equation (90). However, the appropriate selection rules must be applied. In the case of a diatomic molecule the selection rule $\Delta J = \pm 1$ was used because the change in dipole moment is of course along the molecular axis. The same rule applies for linear polyatomic molecules when the dipole moment derivative is similarly directed. However, for vibrational fundamentals arising from bending vibrations of linear molecules, the dipole moment variation is oriented perpendicular to the molecular axis. The selection rule in this case is $\Delta J = 0, \pm 1$. Reference to Figure 2 then shows that the transitions of the type $\Delta J = 0$ will result in a series of lines at approximately the same wavenumber. These features constitute the Q branch in the rotation–vibration spectrum.

It is of interest to regard the infrared spectra of a simple linear molecule on the basis of the general theory presented above. As an example, consider the carbon dioxide molecule, whose equilibrium configuration is described by the point group $_{\infty}$h. The three fundamental vibrations of this molecule are represented in Figure 11. It should be noted that ν_2 is doubly degenerate, leading to

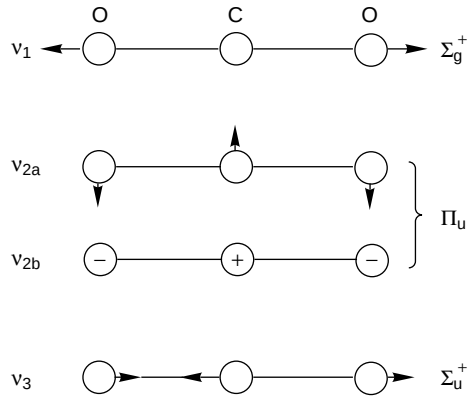

Figure 11 The normal modes of vibration of the carbon dioxide molecule. The IRs in point group $_{\infty}$h are shown on the right.

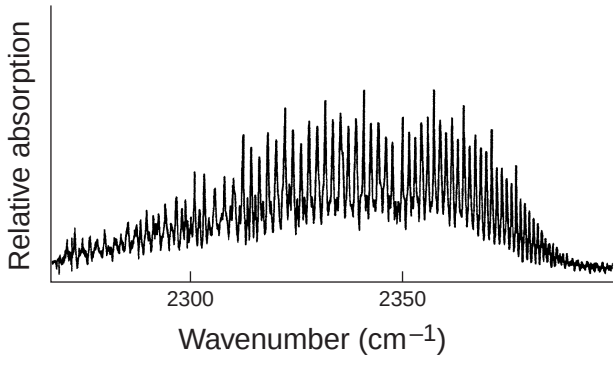

Figure 12 Portions of the infrared spectrum of carbon dioxide gas: the ν_3 fundamental.

a nonzero angular momentum about the molecular axis and the symbol Π for the IR. Only the two vibrational fundamentals ν_3 and ν_2 are infrared-active.

The infrared spectrum of carbon dioxide in the region of its vibrational fundamental ν_3 is shown in Figure 12. This band, which arises from the asymmetric stretching vibration, displays the characteristic P–R branches, analogous to those in the spectrum of a diatomic molecule. Furthermore, the intensity distribution of these branches is similar as illustrated in Figure 10 for hydrogen chloride. A weak, alternating variation in the intensities of the individual lines can be detected in Figure 12. This intensity pattern is due to the effect of the spins of the nuclei that are interchanged under the inversion operation.[9]

The ν_2 fundamental absorption band of carbon dioxide is presented in Figure 13. It is characterized by an intense Q branch, which arises from the series of lines due to the $\Delta J = 0$ transitions. These lines are not superimposed because the moments of inertia of the molecule are not exactly the same in the fundamental and first excited vibrational states. Under high resolution these lines can be clearly resolved in this case. The P and R branches

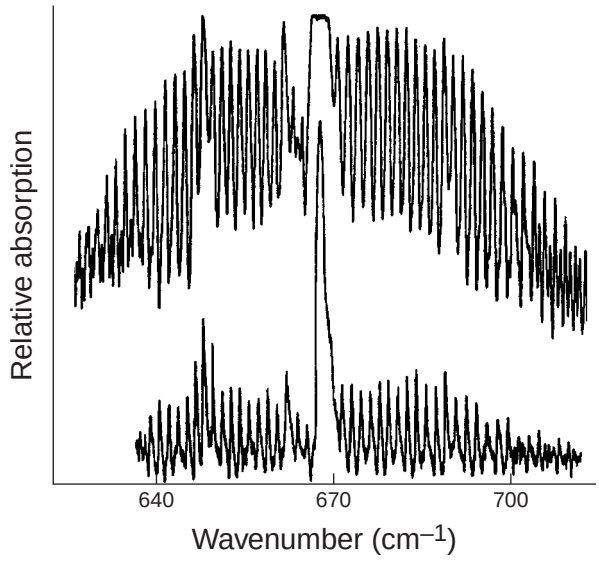

Figure 13 Portions of the infrared spectrum of carbon dioxide gas: the ν_2 fundamental.

in Figure 13 also display the characteristic intensity distribution. The detailed analysis of a perpendicular band such as this one requires inclusion of the effect of Coriolis forces, as the angular momentum about the molecular axis is not equal to zero ($\ell = 1$).

From this brief summary of the infrared spectra of linear molecules, it is seen that the vibrations can be classified as stretching or bending modes. The former give rise to parallel bands, while the latter are characterized by perpendicular bands. These descriptions refer to the direction of the dipole moment variation with respect to the molecular axis. The most obvious difference between parallel and perpendicular bands is in the absence or presence, respectively, of the Q branch. Although for larger molecules in the gas phase the rotational structure of absorption bands may not be resolvable, the observed band envelopes often allow parallel and perpendicular bands to be distinguished.

5.2 Spherical Rotators

The term ''spherical top'' is employed to describe a molecule for which, in its equilibrium configuration, all three moments of inertia are equal (see section 3.8). Then, as for linear molecules, the energy in the rigid-rotator approximation is given by Equation (43). However, it should be noted that the degeneracy of the rotational levels becomes equal to $(2J + 1)^2$. This result has an effect on the intensity distribution of a given branch, or its envelope, of an absorption band.

Most molecules in this class whose rotation–vibration spectra have been studied have tetrahedral symmetry. Thus, from the character table for point group T_d it is seen that infrared activity is limited to triply degenerate vibrations of species F_2. As the selection rules appropriate to the F_2 fundamentals are $\Delta J = 0, \pm 1$, rotation–vibration bands with a simple P–Q–R branch structure are predicted. However, as the vibrational levels are degenerate in this case, the effect of Coriolis interaction can become important. This contribution to rotation–vibration interaction results in splitting of the triply degenerate levels into three distinct energies.

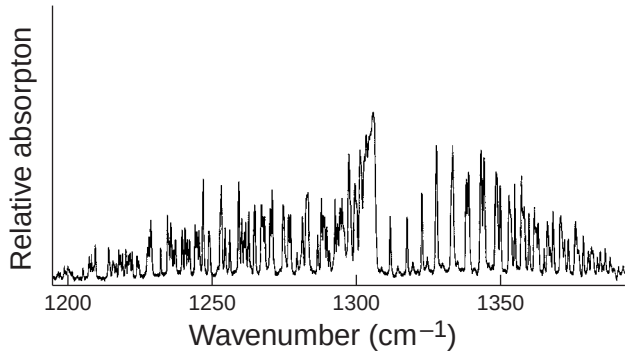

Figure 14 The ν_4 fundamental in the infrared spectrum of methane.

The magnitude of the Coriolis splitting varies considerably from one band to another. This fact is exemplified in Figure 14, where the spectrum of the ν_4 fundamental of methane is presented. Although the ν_3 band exhibits the expected simple structure, the ν_4 band is seriously perturbed as a result of interaction with the nearby infrared-inactive vibration ν_2.

From the above example of the infrared spectrum of a spherical rotator, it should be obvious that a complete, quantitative analysis of the rotation–vibration spectrum is not straightforward. Although the broad features of a vibrational fundamental of a molecule in this class are similar to those of a linear molecule, Coriolis interaction often complicates the interpretation.

5.3 Symmetric Rotators

In Equation (45) for the rotational energy of a rigid symmetric top, (Equation 91)

$$\frac{E_{\text{rot}}}{hc} = \overline{B}J(J+1) + (\overline{C} - \overline{B})K^2 \qquad (91)$$

where the rotational constants expressed in wavenumbers are defined with respect to the principal axes. Note that in this case the c-axis has been chosen as the axis of symmetry; thus, $\overline{A} = \overline{B} \neq \overline{C}$. In Equation (91), the rotational quantum numbers are defined as before, namely, $J = 0, 1, 2, \ldots$ and $K = 0, \pm 1, \pm 2, \ldots, \pm J$. If the rotator is an oblate spheroid, $\overline{C} < \overline{B}$ and the second term in Equation (91) is negative. For the prolate top this

term is positive (if the a-axis is chosen to be the unique one, the axis of symmetry, $\overline{B} = \overline{C}$ and the rotational constant $\overline{C}$ is replaced by $\overline{A}$ in Equation 91). As in this approximation the energy depends on K^2, each level for $|K| > 0$ is doubly degenerate. The sign, for a given value of K, represents classically the direction of rotation of the angular momentum about the symmetry axis. Finally, it is apparent that if $\overline{B} = \overline{C}$, Equation (91) reduces to the spherical-top expression.

As a first approximation the energy of a symmetric-top molecule can be written as the sum of those of a rigid rotor and $3N - 6$ harmonic oscillators. However, in a higher approximation it is often necessary to take into consideration the effect of molecular vibrations on both moments of inertia. For nondegenerate vibrations it is usually sufficient to add terms that represent simple rotation–vibration coupling. For degenerate vibrations, however, the effect of Coriolis forces is much more important. As in the case of spherical rotators, they can produce a coupling of two vibrations of different species that increases with the rotational energy and a splitting of the degenerate vibrational levels. This effect is more important if the frequencies of the two vibrations are not widely separated. For practical purposes, the selection rules outlined in the following paragraph are not modified by these interactions.

For symmetrical rotators the selection rules for rotational transitions that accompany a vibrational one depend on the direction of the transition moment with respect to the top axis. Thus, as in the case of linear rotators, parallel and perpendicular bands have different characteristic structures. For parallel bands, the selection rules are given by Equation (92):

$$\Delta K = 0, \quad \Delta J = 0, \pm 1 \tag{92}$$

with the condition that $\Delta J = 0$ transitions do not take place when $K = 0$. The corresponding selection rules that determine the structure of perpendicular bands are (Equation 93)

$$\Delta K = \pm 1, \quad \Delta J = 0, \pm 1 \tag{93}$$

The possible transitions are summarized in Figure 15(a) and (b).

If the interaction of rotation and vibration is weak, as is often the case, the above analysis indicates that the parallel bands of a symmetric rotator should exhibit the characteristic P–Q–R branch structure. However, the corresponding perpendicular bands are usually much more complicated because of the Coriolis forces. Furthermore, all nonplanar molecules in this class have the possibility of inversion. The most important case is certainly that of the ammonia molecule. It has a pyramidal structure whose equilateral triangular base

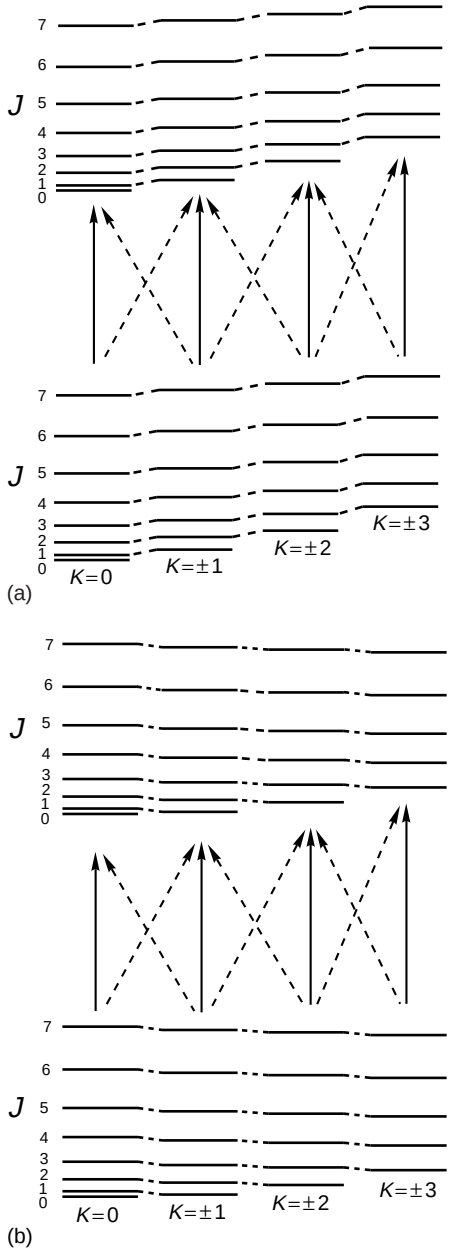

Figure 15 The energy levels (schematic) of symmetric rotators: (a) prolate; (b) oblate. Transitions indicated by solid arrows are limited to $\Delta K = 0$ (parallel bands) and those by dashed arrows by $\Delta K = \pm 1$ (perpendicular bands).

is determined by the equilibrium positions of the three hydrogen atoms. Its symmetry group is thus $_{3v}$. This structure can be inverted in the sense that the three protons can pass to equivalent positions on the other side of the nitrogen atom. The relatively low barrier to this process (approximately 5.8 kcal mol^{-1}) results in splitting of the vibrational energy levels. The separation of the two components corresponds to a frequency of 23.87 GHz, or approximately 0.79 cm^{-1}. Thus, an absorption line is

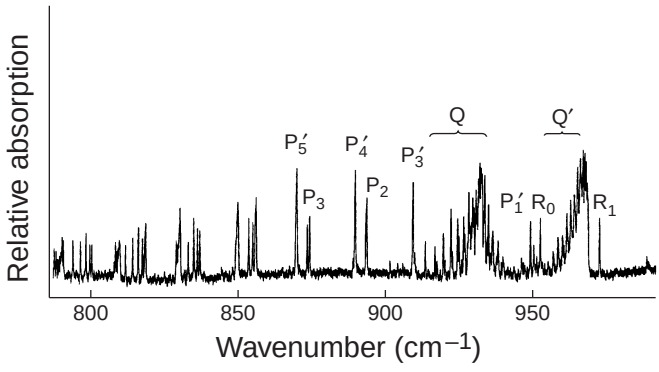

Figure 16 A portion of the ν_2 fundamental of ammonia.

readily observable at this frequency in the microwave region. The effect of this inversion process results in a doubling of the rotation–vibration spectrum, as shown in Figure 16. The two Q branches are clearly shown, each with its associated P and R branches.

5.4 Asymmetric Rotators

Very little can be said here concerning the rotation–vibration spectra of asymmetric rotators, except that they are in general very complicated. As indicated earlier, no analytical expression can be derived for the rotational energies of molecules in this class. However, computer programs are now available that provide a means of fitting observed rotational fine structure, as observed on vibrational bands in the infrared region.

6 SPECTRA OF NEAT LIQUIDS AND SOLUTIONS

The observed vibrational frequencies of molecules in neat liquids or in solution are often employed as a means of identification. However, it is well known that the frequency of a given mode of vibration of a molecule in the liquid state depends on the nature of the surrounding molecules. The vibrational frequency shifts displayed by a molecule as it passes from the free state (as in ideal gas) to its role as a molecule in a liquid environment is due to intermolecular forces. It may undergo interactions with identical molecules in a neat liquid or it may become a solute in a liquid solvent. The interaction of this molecule with its neighbors, which depends on physical quantities such as the polarity of the nearby molecules and their polarizability, have been employed in simple models of intermolecular interactions. Probably the earliest treatment of these so-called solvent effects was due to Kirkwood and to Bauer and Magat.[6] Their model, which has been modified by many authors, provides the initial basis for the interpretation of the effect of the environment on the vibrational spectra of molecules in the liquid state.

6.1 Solvent Effects

The above approach to the evaluation of the environmentally induced frequency shifts provided a semiempirical basis of the analysis of solvent effects. It was useful in the days when infrared instrumentation was relatively primitive and spectroscopists were content to measure frequencies of vibration, or perhaps the widths of absorption bands. Such quantities yielded extremely valuable analytical information concerning molecular interactions in liquids. However, with the refinement of instrumentation (in particular the development of Fourier transform techniques), as well as an important evolution of the theory of molecular dynamics in condensed phases, it has become of interest to measure and interpret the exact profile of an absorption band of a molecule in the liquid state. This approach to the analysis of the profiles of absorption bands of liquid samples will be developed briefly in section 6.3.

Some infrared spectra of acetonitrile in the liquid state are presented in Figure 17. The spectrum of the neat liquid exhibits a strong absorption band at 2253 cm^{-1}, which is assigned to the C≡N stretching vibration. The weaker feature at 2293 cm^{-1} is probably the first overtone of a mode approximately described as a C−C bond stretching.

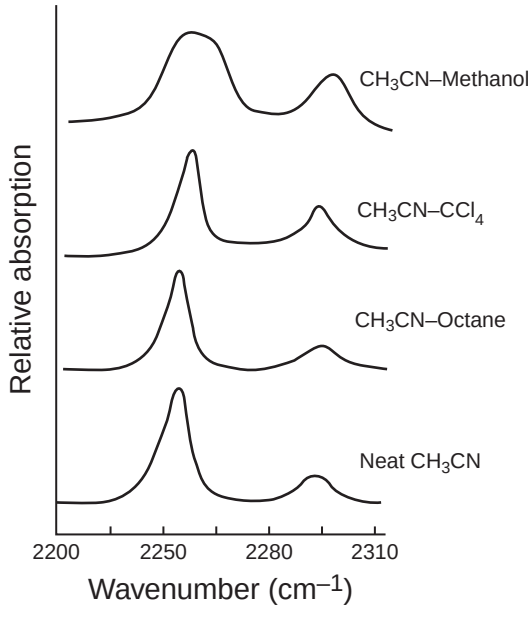

Figure 17 Infrared spectra of acetonitrile. (Reproduced by permission of Wiley-VCH from G. Turrell, 'Bandshapes in the Vibrational Spectroscopy of Liquids I. Infrared Absorption', *Spectrosc. Eur.*, **9**(4), 8–15 (1997).)

Spectra of the same molecule in several commonly used solvents are also shown in Figure 17. It is apparent that the position of the minimum transmission of the C≡N stretching fundamental is a function of the molecular environment. On the other hand, the C−C overtone is relatively insensitive to the intermolecular forces. A careful inspection of these spectra reveals certain changes in the profiles of the observed bands. Both the widths of the C≡N stretching bands and their transmission in the wings vary with the solvent used. Furthermore, slight band asymmetries can be detected when the abscissae are developed. In particular, the broadening of the C≡N stretching band of acetonitrile provides evidence of hydrogen-bond formation with the methanol solvent. This particular interaction will be discussed in section 6.4.

6.2 Absorption Bandshapes

The shape of the wings of an absorption band of a small molecule in solution is primarily due to its hindered rotational (or orientation) motion. The effect becomes more clearly evident in the case of a diatomic solute molecule. Some early spectra of HCl in various solutions are shown in Figure 18, where they are compared with a schematic representation of the spectrum of the gas. The spectra of the dissolved molecule are characterized by the appearance of a central absorption peak analogous to the (forbidden) Q branch in the spectrum of the gas at ν_0 and its shift to lower frequencies with increasing solvent perturbation. Furthermore, the wings of the bands become less prominent as the central feature increases in intensity. It should be noted that CCl_4, a commonly used solvent in organic analytical spectroscopy, is not as ''inert'' as is often supposed. Thus, the vibrational frequencies of a molecule (especially those involving H atoms) in this solvent do not correspond to those of the isolated molecule. Furthermore, the frequency shifts become greater in polar solvents, as exemplified by the spectrum of HCl in CH_2Cl_2.

6.3 Band Broadening

When a hypothetical isolated molecule interacts with electromagnetic radiation, the processes of absorption, induced emission and/or light scattering may result. The spectra obtained are characterized by the bands due to transitions between the various quantum states of the molecules. The fact that a given transition produces a broadened feature is fundamentally due to the uncertainty principle and the Doppler effect. These origins of spectral broadening are only important in the study of gaseous samples under extreme conditions (e.g. very low pressures and temperatures – as in the stratosphere). For the routine applications considered here, where a sample

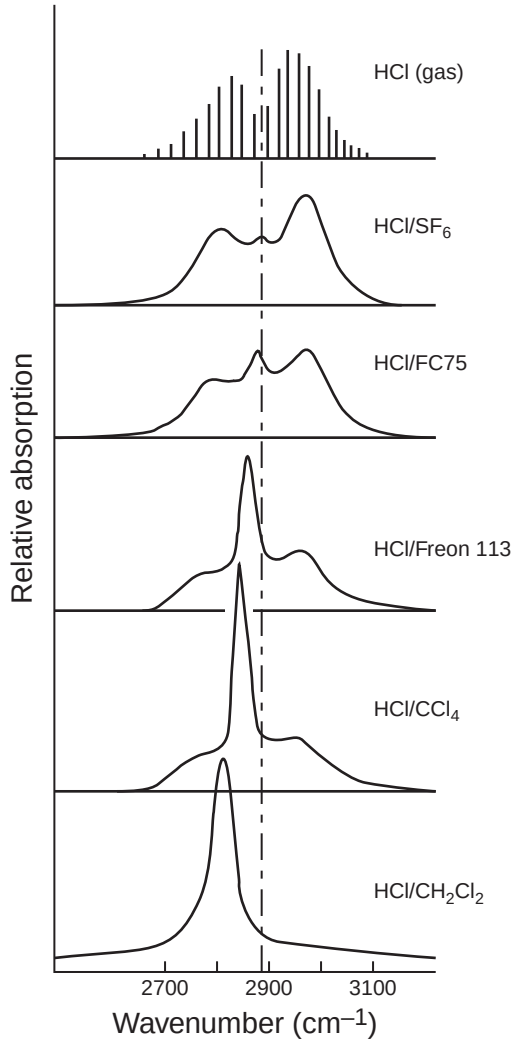

Figure 18 Infrared spectra of HCl in various solvents compared with that of the gas. (Reproduced by permission of Wiley-VCH from G. Turrell, 'Bandshapes in the Vibrational Spectroscopy of Liquids I. Infrared Absorption', *Spectrosc. Eur.*, **9**(4), 8–15 (1997).)

consists of a very large number of molecules in a liquid, the broadening and frequency shift of a given band is produced by the intermolecular interactions. Furthermore, the time dependence of these interactions results in characteristic band shapes or ''profiles''.[10]

In general, the shape of a spectral band can be described by a series of moments defined by Equation (94):

$$\mathrm{M(n)} \equiv \int_{-\infty}^{+\infty} (w - w_0)^n \hat{} (w)\, \mathrm{d}w \qquad (94)$$

where $n = 0, 1, 2 \ldots$ and $w \equiv 2\pi\nu$. The quantity w_0, which is known as the ''shifted band origin'', corresponds to the peak position of a symmetric band. Clearly, when $n = 0$, the moment M(0) is just the area of the band, the absolute intensity referred to in section 4.12. The first

moment ($n = 1$) is the shift of the position of the peak and the second moment is essentially a measure of the bandwidth. Higher moments provide specific information regarding solvent–solute interaction.

From a theoretical point of view, it is the Fourier transform of the spectral profile that leads to a description of the temporal behavior of a molecule in solution. The result is the autocorrelation function that at each instant specifies the orientation of the molecule. While this information is, at least at the present time, of limited analytical importance, it provides quantitative support for the molecular-dynamic description of the liquid state.

6.4 The Hydrogen Bond

This summary of the effects of intermolecular interactions in neat liquids, and thus those due to solute–solvent interaction, has been limited to the case of weak intermolecular forces. If a given molecule is surrounded by polar molecules, these forces become much stronger. Furthermore, in certain cases specific interactions may occur, leading, for example to chemical reactions. An intermediate situation may involve the formation of hydrogen bonds. Infrared spectroscopy provides the most important tool for the study of these important interactions.[11]

The vibrational spectrum of a simple hydrogen bond can be characterized by considering the degrees of freedom of the structure X–H···Y. In most cases the atoms X and Y are at least an order of magnitude heavier than the proton and the vibrational modes of this system can be described approximately by:

1. An X–H stretching mode of wavenumber $\bar{\nu}_s$, $X-\overrightarrow{H}\cdots Y$, where $3500 \geq \bar{\nu}_s \geq 1700\,cm^{-1}$.
2. An X–H···Y bending mode, $X-\underset{\uparrow}{H}\cdots Y$, of wavenumber in the region $1700 \geq \bar{\nu}_b \geq 800\,cm^{-1}$. It is doubly degenerate if the symmetry of rotation about the XHY axis is threefold or higher. If the XHY bond is bent, or if the groups X and Y lower the symmetry of the system, $\bar{\nu}_b$ will in principle be split into a doublet.
3. A stretching mode of wavenumber $\bar{\nu}_{XY}$, $\overleftarrow{X}-H\cdots\overrightarrow{Y}$, where $600 \geq \bar{\nu}_{XY} \geq 50\,cm^{-1}$.
4. A torsional mode $\bar{\nu}_t \geq 500\,cm^{-1}$, which occurs when X and Y are nonlinear polyatomic groups.

Although the appearance of (3) and (4) could be employed as a criterion for hydrogen-bond formation, it is usually more convenient and reliable to use the observed shifts of $\bar{\nu}_s$ and $\bar{\nu}_b$ that accompany it. Studies of a large number of hydrogen-bonded systems have shown that as $\bar{\nu}_s$ decreases, $\bar{\nu}_b$ shifts to higher wavenumbers, and the intensity and width of the $\bar{\nu}_s$ absorption band increases with increasing strength of the hydrogen bond.

A characteristic feature of the hydrogen bond, as observed in its infrared spectrum, is the extreme broadening of the $\bar{\nu}_s$ band. This behavior, which is unique to hydrogen-bonded systems, has been the subject of much discussion. Regardless of its origin, it is of obvious diagnostic value, as it is the most important criterion for hydrogen-bond formation.

An early infrared spectroscopic investigation of hydrogen bonding in methanol illustrates the specific information that can be obtained (Figure 19). From the results of this work it was possible to distinguish the absorption band associated with the O–H stretching vibration of the monomer in a relatively inert solvent from those resulting from dimers and higher polymers. The spectra were obtained from solutions of methanol in toluene. Spectrum A exhibits a relatively sharp band at $3607\,cm^{-1}$ which is assigned to the O–H stretching vibration of the monomer. This feature is clearly dominant in dilute solution. The broad band at $3502\,cm^{-1}$ arises from the formation of dimers as a result of hydrogen-bond formation. It can be considered to be characteristic of this type of interaction. The subtraction of the spectrum of the more dilute solution effectively eliminates the monomer band and reveals the weaker band at $3582\,cm^{-1}$. This band can thus be attributed to the stretching of a "free" OH group in the dimer. This result leads to the conclusion that the dimer has an open structure, rather than a cyclic one.

The example chosen here to illustrate the spectroscopic consequences of hydrogen-bond formation is from studies of the liquid state. It is of particular interest in demonstrating the evolution of this fundamental aspect of infrared

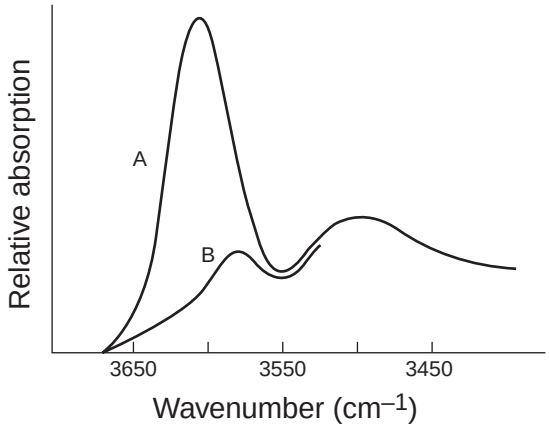

Figure 19 Infrared absorption spectra of methanol in toluene. (A) Concentration 0.25 M. (B) Difference spectrum obtained by subtracting that of a 0.1 M solution. (Redrawn by permission of Elsevier from Figure 1 in L.J. Bellamy, R.J. Pace, 'Hydrogen Bonding by Alcohols and Phenols – I. The Nature of the Hydrogen Bond in Alcohol Dimers and Polymers', *Spectrochim. Acta*, **22**, 525–533 (1966).)

spectroscopy. More recent applications to crystals, glasses and, in particular, molecules of biological interest are described elsewhere in this volume.

7 SPECTRA OF SOLIDS

7.1 The Linear Diatomic Chain

The interpretation of the vibrations of crystalline solids is based on the one-dimensional model developed by Born.[6,12] This simple system was originally employed to estimate the specific heats of crystals, in particular, those of the sodium chloride structure. It has subsequently served as the basis for the interpretation of the vibrational spectra of crystals. The analysis of this model, the diatomic chain, will now be briefly summarized.

Consider a linear chain consisting of two types of atom arranged alternately, as shown in Figure 20. As indicated above, this one-dimensional model can be used to represent a simple ionic crystal. The distance d is then the dimension of the unit cell which contains two particles. The forces on atoms of type 1 and 2 are given by, respectively, Equation (95):

$$F_{2n} = -f'(\rho_{2n} - \rho_{2n-1}) + f(\rho_{2n+1} - \rho_{2n}) = m_1\ddot{\rho}_{2n} \quad (95)$$

and Equation (96):

$$F_{2n+1} = -f(\rho_{2n+1} - \rho_{2n}) + f'(\rho_{2n+2} - \rho_{2n+1}) = m_2\ddot{\rho}_{2n+1} \quad (96)$$

where ρ_n is the displacement of atom n along the direction of the chain. The force constants f and f' are assumed to be different, at least in this initial formulation of the problem.

Periodic solutions of Equations (95) and (96) can be proposed in the form of Equation (97):

$$\rho_{2n} = A_1 e^{2\pi i(\nu t - nkd)} \quad (97)$$

and Equation (98):

$$\rho_{2n+1} = A_2 e^{2\pi i\ \nu t - \left(n+\frac{1}{2}\right)kd} \quad (98)$$

Substitution of these expressions in Equations (95) and (96) leads to the pair of linear, homogeneous

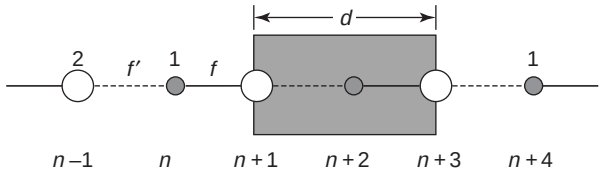

Figure 20 The linear diatomic chain.

Equations (99) and (100) for the amplitudes A_1 and A_2:

$$[4\pi^2\nu^2 m_1 - (f + f')]A_1 + (fe^{\pi ikd} + f'e^{-\pi ikd})A_2 = 0 \quad (99)$$

and

$$(f'e^{\pi ikd} + fe^{-\pi ikd})A_1 + [4\pi^2\nu^2 m_2 - (f' + f)]A_2 = 0 \quad (100)$$

A nontrivial solution of these equations exists only if the determinant of the coefficients vanishes. Thus, the secular determinant becomes Equation (101):

$$\begin{vmatrix} f + f' - 4\pi^2\nu^2 m_1 & -(fe^{\pi ikd} + f'e^{-\pi ikd}) \\ -(f'e^{\pi ikd} + fe^{-\pi ikd}) & f' + f - 4\pi^2\nu^2 m_2 \end{vmatrix} = 0 \quad (101)$$

whose roots are given by Equation (102):

$$\nu^2 = \frac{f + f'}{8\pi^2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) \pm \frac{1}{4\pi^2}\sqrt{\left(\frac{f + f'}{2}\right)^2\left(\frac{1}{m_1} + \frac{1}{m_2}\right)^2 - \frac{4ff'\sin^2(\pi kd)}{m_1 m_2}} \quad (102)$$

For the special case in which all bonds are identical, $f = f'$, which implies equal spacing of the atoms, and Equation (102) reduces to Equation (103):

$$\nu^2 = \frac{f}{4\pi^2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) \pm \sqrt{\left(\frac{1}{m_1} + \frac{1}{m_2}\right)^2 - \frac{4\sin^2(\pi kd)}{m_1 m_2}} \quad (103)$$

The vibrational frequencies calculated from Equation (103) are plotted as functions of the wavenumber k in Figure 21, where it has been assumed that $m_1 < m_2$. Clearly, it is the positive roots of Equation (103) that are of physical significance. The curve that passes to the origin is called the acoustical branch, because the frequencies fall in the region of sonic or ultrasonic waves. The upper curve, referred to as the optical branch, represents frequencies in the optical (infrared) spectral region. The latter are of prime importance for the subject of this article.

The forms of the normal vibrations for each branch are given schematically in Figure 21 for the limiting values $k = 0$ and $k = 1/2d$. It should be noted that at $k = 0$ the optical branch represents a simple stretching of the bond between the two atoms for which their center of mass remains fixed. As the two atoms are assumed to be different, the oscillating dipole moment which results can interact with incident radiation, producing absorption in the infrared region. The acoustic vibration, on the other hand, does not result in a change in dipole moment at $k = 0$.

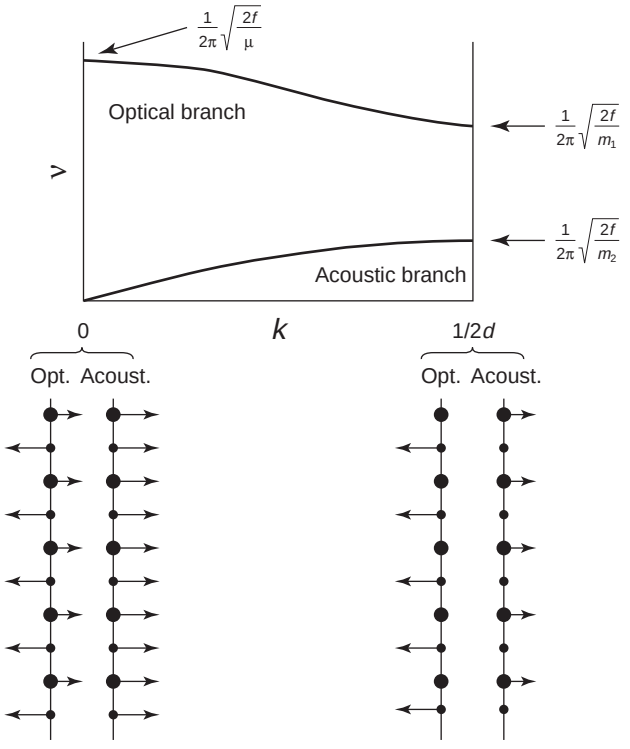

Figure 21 Dispersion curves and the forms of the normal modes for the longitudinal vibrations of the diatomic chain. The arrows represent atomic displacements. (Reproduced by permission of Academic Press from G. Turrell, *Infrared and Raman Spectra of Crystals*, Academic Press, London, 70, 1972.)

From Figure 21 it is seen that wave solutions of the type of Equations (97) and (98) do not exist for frequencies between $(1/2\pi)\sqrt{2f/m_1}$ and $(1/2\pi)\sqrt{2f/m_2}$. In this gap the wavenumber k is imaginary; hence waves having frequencies in this region are damped.

The diatomic chain illustrated in Figure 20 can also be used as a model of a simple one-dimensional molecular crystal, if, for example, f is associated with a covalent bond and f' is used to represent the much weaker intermolecular forces. If $f \gg f'$, the roots of Equation (103) are given approximately by Equation (104):

$$\nu_{\text{int}} \approx \frac{1}{2\pi} \sqrt{\frac{f}{\mu}} \tag{104}$$

and Equation (105):

$$\nu_{\text{ext}} \approx \frac{1}{\pi} \sqrt{\frac{f'}{m_1 + m_2}} \, |\sin(\pi k d)| \tag{105}$$

The first solution (Equation 104), which is independent of k, corresponds to the vibrational frequency of an isolated diatomic molecule, as in the ideal-gas phase. For all values of k the center of mass of each molecule remains fixed in its unit cell. The second solution arises from translational motion of the rigid diatomic molecules. The approximate solutions given by Equations (104) and (105) are appropriate to the limiting case, $f \gg f'$. In general, ν_{int} exhibits some dependence on k. Nevertheless, the above approximation is the basis of the separation which is often made of internal and external motion in crystals containing molecules or polyatomic ions.

From the preceding analysis it should be apparent that the vibrational frequencies of this one-dimensional system spread over two distinct regions. For the acoustic branch the frequencies lie between zero and $(1/2\pi)\sqrt{2f/m_1}$, whereas for the optical branch they are between $(1/2\pi)\sqrt{f/\mu}$ and $(1/2\pi)\sqrt{2f/m_1}$. In each case the former values correspond to the origin of the wavenumber scale, $k = 0$, and the latter are the vibrational frequencies at the limit of the first Brillouin zone. This limit is equal to $k = 1/2d$ in this, the one-dimensional case. The distribution of frequencies within the two limits specified above is usually referred to as the density of vibrational states. Not all of these frequencies can result in spectroscopic features, as the question of selection rules must be considered.

7.2 Three Dimensions: Crystalline Solids

The static structure of a perfect crystal can be described by successive translations of a unit cell in space. This repetition constitutes a symmetry operation in the sense that all such unit cells are considered to be identical. Thus, a symmetry group of these operations can be defined. The application of this principle leads to the definition of the 230 space groups of crystallography. The justification for this model supposes that the crystal is infinitely large in three-dimensional space, or at least large with respect to the dimensions of the unit cell. A mathematical artifice, the so-called cyclic boundary conditions, is sometimes introduced to achieve the equivalent result.

In the application of interest in this article, it is the primitive unit cell that is employed to generate the three-dimensional space of the crystal. This cell, which is the smallest geometrical motif that can be defined, is in some cases identical with the crystallographic unit cell, although it may otherwise have half or one-quarter its volume. This question, which has been addressed in several references, is basic in the spectroscopic applications of crystallography, as the vibrational spectroscopy of crystals involves their dynamic, rather than their static, description.[8,9] A starting point in the analysis of their dynamics is provided by the model of the diatomic chain, as outlined above.

The complexity of the three-dimensional vibrational problem increases very rapidly with the number of atoms in the unit cell. However, by analogy with the

one-dimensional model, solutions to the equations of motion can be written in the harmonic approximation as Equation (106):

$$\rho_n = \boldsymbol{A}_n \exp[2\pi i(\nu t - \boldsymbol{k}\boldsymbol{x}_\ell)] \qquad (106)$$

where ρ_n is a vector that specifies the displacement of atom n at a given instant. The amplitude and frequency of the descriptive wave in the crystal are represented by $\boldsymbol{A}_n$ and ν, respectively. The equilibrium position of each atom is located with respect to a suitable space-fixed origin by the vector $\boldsymbol{x}_\ell$, while $\boldsymbol{k}$ is the wave vector, a quantity that has the magnitude of reciprocal wavelength and is in the direction of propagation of the descriptive wave. As in the one-dimensional case, the dispersion relations determine the density of vibrational states in the crystal.

There are three dispersion curves for each particle in the primitive unit cell, as each has three degrees of freedom. Thus, for the simple example of a molecular crystal whose unit cell contains σ molecules, each of which is composed of N atoms, as many as $3N\sigma$ separate curves (branches) may exist. Of course, the actual number of branches may be reduced owing to degeneracies in the system. Three of these degrees of freedom correspond to acoustic modes; hence there are up to $3N\sigma - 3$ optical branches. The general form of the dispersion relations for such a system is sketched in Figure 22 for a given direction of the wave vector. The various branches have, in general, different shapes depending on the direction of $\boldsymbol{k}$.

As pointed out in the previous section in connection with the vibration of the diatomic chain, intramolecular forces are relatively strong in a molecular crystal. Hence the frequencies of the $3N - 6$ vibrational fundamentals of a molecule in the gas phase are not severely modified when the molecule is in a crystalline environment. If there are σ molecules per unit cell, there are $\sigma(3N - 6)$ possible ''internal'' frequencies – although degeneracies are usually present to reduce this number. As in the

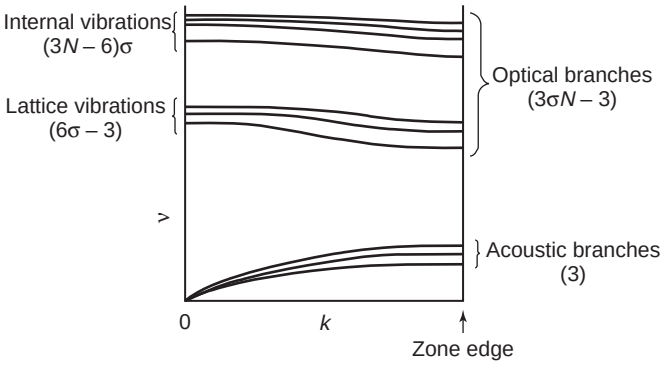

Figure 22 Dispersion curves for a simple crystal. (Reproduced by permission of Academic Press, from G. Turrell, *Infrared and Raman Spectra of Crystals*, Academic Press, London, 76, 1972.)

corresponding limiting case of the diatomic chain, these internal frequencies are usually only slightly dependent on the wave vector, $\boldsymbol{k}$.

The remaining degrees of freedom are derived from the 3σ translational and 3σ rotational motions of the (nonlinear) molecules. However, three of these degrees of freedom are responsible for the acoustic modes of the crystal, leaving $6\sigma - 3$ as the maximum number of optical external modes. These latter, which are referred to by spectroscopists as the ''lattice vibrations'', usually fall in the far-infrared region of the spectrum of a crystal.

7.3 The Factor-group Analysis of Crystals

The analysis of the infrared and Raman spectra of crystals is based on the notion of the factor group. In effect, if a perfect crystal can be described by a space group, a symmetry group that characterizes the unit cell can also be defined. Although it is not, strictly, a point group, it has the same form as a point group, viz. it is isomorphic. As only symmetry axes of order 1, 2, 3, 4 or 6 are possible for a unit cell that can fill three-dimensional space, only 32 such groups exist. They identify the crystal classes. On the basis of this argument, the space group of a given crystal can be factored in the following form (Equation 107):

$$= \quad \otimes \qquad (107)$$

where is the group of translations of the unit cell.[6] The symbol $\otimes$ represents here the (semi)direct product. It can be shown that the point group that is isomorphic with the unit cell group can be used to determine the normal modes of vibration of the molecules in the unit cell and the selection rules for the vibrational fundamentals. Once the appropriate unit-cell group has been identified, the vibrational analysis is similar to that of an isolated molecule.

As an illustration of the factor-group method, consider crystalline methyl chloride. The symmetry of the molecule is described by the point group $_{3v}$ and the discussion presented in section 4.9 is appropriate. With the use of the character table (Table 6) the vibrational analysis of the molecule can be made as before. The results are summarized in the first column of Figure 23. The internal vibrations are numbered in decreasing order of frequency, starting with the most symmetrical IR (A_1 in this case) following Herzberg.[7] The three doubly degenerate frequencies of E species are numbered similarly. The representations of the six external degrees of freedom are also identified in Figure 23.

From the results of X-ray analysis, it is found that methyl chloride crystallizes in the side-centered orthorhombic system. Its space group is $^{12}_{2v} \equiv Bm2_1b$, with four molecules in the crystallographic unit cell. However, the primitive-symmetric or Wigner–Seitz cell contains only two molecules. They are located at sites

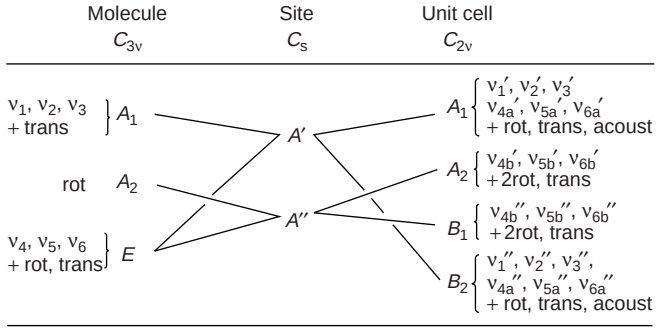

Figure 23 The correlation diagram for methyl chloride.

(special positions) of symmetry $_s$. One of the three planes of symmetry of the isolated molecule corresponds to the plane at the site. Of course, it contains the chlorine and carbon atoms, in addition to one of the three hydrogens. The other two hydrogen atoms are exchanged by the reflection operation with respect to this plane. The site group is a subgroup of the symmetry group of the molecule. Thus, the correlation shown on the left of Figure 23 is easily established.

The required unit-cell group is $_{2v}$, as found by simply suppressing the superscript in the Schönfiles notation for the space group. In this case the site group is also a subgroup of the unit-cell group. The correlation shown on the right of Figure 23 is established by inspection of the character tables (see Table 5).

The interpretation of the results obtained above by the analysis of the methyl chloride example can be summarized as follows. The three nondegenerate internal vibrations are not modified by the site symmetry in the crystal. The three degenerate modes of species E are, however, split by the symmetry of the local crystalline field. These three internal vibrations will then be expected to be observed as doublets in the infrared spectrum because of this site splitting.

As there are two molecules per primitive unit cell, coupling of the molecular modes of vibration becomes possible. In the present example, the three frequencies of species A_1 in the molecule are split by this correlation (or dynamic) coupling. According to the character table for the group $_{2v}$, both components of each doublet will be infrared-active. As for the three vibrations of species E in the molecule, each has been subjected to site splitting. Furthermore, each component will be further split due to coupling with its partner in the unit cell. However, as shown in Figure 23, the doublets produced by the splitting of the A'' species of the molecule in its site will produce one component in the infrared-inactive species A_2. The other component passes to species B_1 and is therefore active. The overall result for the E species of the molecule is to produce an infrared-active triplet for each of the frequencies ν_4, ν_5 and ν_6.

The magnitudes of the static and dynamic splittings described above are in general of the same order of magnitude, viz. a few wavenumbers. They cannot, therefore, be distinguished unless systematic investigations of oriented single crystals can be made. Such studies of infrared dichroism with the use of polarized infrared radiation are of extreme value, not only for structural investigations of single crystals, but also in the determination of the optical properties of films. Examples include the problem of molecular orientation of films deposited on solid substrates and that of stretched polymers.

7.4 Polymers

The linear-chain model developed in section 7.1 served as the basis for the dynamic description of a crystal. It can also be employed in the analysis of the vibrations of polymers.[6,13] It is often the case that the forces between successive motifs in a polymer chain are much stronger than the forces between adjacent chains. A given chain can then be developed by successive translation of the particular motif of which it is composed. The group of translational operations constitute the line group, while the particular motif represents an effective unit cell in this model.

Consider polyethylene as a polymer chain that can be analyzed by application of the line-group method. The repetitive motif is represented in Figure 24. The carbon skeleton of the extended chain is assumed to have a planar $[\sigma(yz)]$ zigzag structure along the y-direction and a plane of symmetry $[\sigma(zx)]$ perpendicular to it. There is a binary axis of symmetry $[C_2(x)]$ perpendicular to the plane of the carbon chain, passing through the center of inversion (i) shown in Figure 24, as well as another $[C_2(z)]$ that passes through each carbon atom. The other symmetry elements are a screw axis and a glide plane, analogous to those encountered in crystallography. They are denoted here by $C_2^s(y)$ and $\sigma^g(xy)$, respectively. The ensemble of operations associated with these elements, with the addition of the identity, form the factor group. It is isomorphic with the point group $_{2h}$. The 18 degrees of freedom of the six-atom repeat unit can be classified as shown in Table 8 with the use of the method outlined in section 7.3.

As the polymer chain has been assumed to be isolated, the external motions of translation in the x- and y-directions, as well as the rotation about the z-axis have

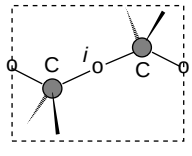

Figure 24 The C_2H_4 motif of the polyethylene chain.

Table 8 Line-group analysis of the extended polyethylene chain

$_{2h}$	E	$C_2(z)$	$C_2(y)$	$C_2(x)$	i	$\sigma(xy)$	$\sigma(zx)$	$\sigma(yz)$		$n^{(\gamma)}$ (tot)	$n^{(\gamma)}$ (vib)
A_g	1	1	1	1	1	1	1	1		3	3
B_{1g}	1	1	−1	−1	1	1	−1	−1	R_z	1	0
B_{2g}	1	−1	1	−1	1	−1	1	−1	R_y	3	2
B_{3g}	1	−1	−1	1	1	−1	−1	1	R_x	2	1
A_u	1	1	1	1	−1	−1	−1	−1		1	1
B_{1u}	1	1	−1	−1	−1	−1	1	1	T_z	3	2
B_{2u}	1	−1	1	−1	−1	1	−1	1	T_y	2	1
B_{3u}	1	−1	−1	1	−1	1	1	−1	T_x	3	2
χ_j	18	−2	0	0	0	0	6	2			

zero frequencies and can be removed, as for a gaseous molecule. These degrees of freedom are sometimes referred to as ''nongenuine'' vibrations. Translation in the z-direction is analogous to the acoustic vibrations of a crystal lattice, as described in the previous section. Thus, for in-phase motions of all repeat units this frequency is also equal to zero.

The number of vibrational modes in each IR is given in the last column of Table 8. Infrared activity is limited to the three species B_{1u}, B_{2u} and B_{3u}. The two B_{1u} modes are combinations of the symmetric stretching (ν_s) and bending (δ) coordinates illustrated in Figure 25. The B_{3u} modes are combinations of the antisymmetric stretching mode (ν_a) and the rocking coordinate (ρ), while the wagging coordinate (w) shown for species B_{2u} can be identified with the corresponding normal coordinate.

The analysis presented above for the vibrations of extended, normal polymer chains can be generalized to include, for example, helical chains. This problem has been described in a number of references.[6,12] Furthermore, the question of the interaction of polymer chains in a solid has also been treated. It should be emphasized that in the development of these models it has been assumed that the chains are of infinite length, or at least that their symmetry under translation is preserved. In the case of chains of finite length, the problem becomes much more complicated, as the appropriate boundary conditions must be imposed at each end of the chain. The reader is referred to more advanced presentations for the effects of these conditions on the vibrational spectra of polymer chains.[6,12]

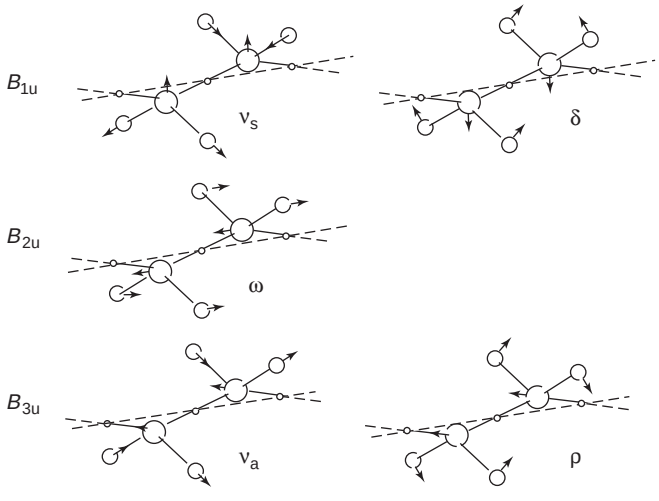

Figure 25 The infrared-active representations and symmetry coordinates for the extended polyethylene chain. (Reproduced by permission of Academic Press, from G. Turrell, *Infrared and Raman Spectra of Crystals*, Academic Press, London, 237, 1972.)

7.5 Disordered Solids

A complete theoretical analysis of disordered solids has not as yet been developed. However, a qualitative description of these systems can be obtained on the basis of the preceding analysis of the vibrations of crystals. In section 7.3 the perfect crystal was taken as a model. However, if its translational symmetry is perturbed, the factor-group method can no longer be applied. Thus, the concept of a repetitive unit cell in space is no longer valid, as a given motif is not exactly reproduced by simple translations. Furthermore, the force field involved is not repetitive. Thus, there are three perturbations to spectroscopic analysis that are produced by the disorder of the system. The first of them has an immediate effect on the selection rules that govern vibrational transitions. All of the frequencies that contribute to the density of vibrational states in a given crystal become spectroscopically active. The result is a broadening of the infrared absorption bands, as can be seen from the simple example shown in Figure 26. The structural changes involved in the transitions from the pure crystal to a disordered system result in the other two effects, namely, modifications of the force field and kinetic effects that contribute to the effective $\boldsymbol{G}$ matrix. Obviously, the quantitative treatment of this problem is difficult.

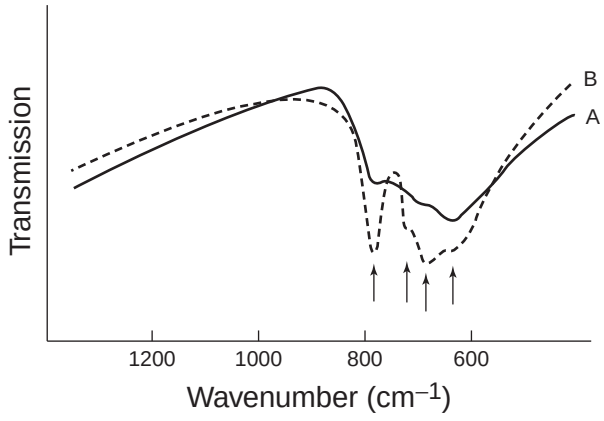

Figure 26 Spectra of TeO_2: (A) glass; (B) crystal. (Redrawn by permission of Kluwer Academic Publishers, from Y. Dimitrev, V. Dimitrov, M. Arnaudov, 'IR Spectra and Structures of Tellurite Glasses', *J. Mater. Sci.*, **18**, 1358–1358 (1983).)

8 CONCLUSION

From the contents of this article, it is apparent that the basic theory of infrared spectroscopy was developed many years ago. The more recent evolution of this subject has been primarily in the areas of computational methods and instrumentation. The impact of computer science has made possible the numerical treatment of spectroscopic data and, in particular, the application of the Fourier transform technique. The future will certainly see advances in the development of rapid time-resolved infrared spectroscopy, which should soon become an invaluable tool in the study of chemical reactions.

On the purely theoretical side, there remains the fundamental problem of disordered systems. These include glasses and other amorphous materials, in addition to the liquid state. The latter can be considered to be a system that is disordered in both space and time. Although it is the state of matter most generally employed in analytical spectroscopy, it still presents the most difficult challenge to molecular theorists.

ABBREVIATIONS AND ACRONYMS

IR	Irreducible Representation
MVFF	Modified Valence Force Field
VFF	Valence Force Field

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Microbiology

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Infrared Spectroscopy **(Volume 12)**
Emission Spectroscopy, Infrared • Interpretation of Infrared Spectra, A Practical Approach • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Databases, Infrared • Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

REFERENCES

1. W.W. Coblentz, *Investigations of Infrared Spectra*, Carnegie Institution of Washington, Washington, DC, 1904; republished by the Coblentz Society, French Village, MO, 1962.
2. L. Pauling, E. Bright Wilson, Jr, *Introduction to Quantum Mechanics*, McGraw-Hill, New York, 1935; republished by Dover Publications, Mineola, NY, 1985.
3. E. Bright Wilson, Jr, J.C. Decius, P.C. Cross, *Molecular Vibrations*, McGraw-Hill, New York, 1955.
4. I.N. Levine, *Molecular Spectroscopy*, John Wiley & Sons, New York, 1975.
5. F.A. Cotton, *Chemical Applications of Group Theory*, 2nd edition, John Wiley & Sons, New York, 1971.
6. G. Turrell, *Infrared and Raman Spectra of Crystals*, Academic Press, London, 1972.
7. L.J. Bellamy, *The Infrared Spectra of Complex Molecules*, Vol. I, John Wiley & Sons, New York, 1954; Vol. II, 3rd edition, Chapman & Hall, New York, 1980.
8. S. Califano, *Vibrational States*, John Wiley & Sons, Chichester, 1976.

9. G. Herzberg, *Molecular Spectra and Molecular Structure II. Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand, Princeton, NJ, 1945.
10. W.G. Rothschild, *Dynamics of Molecular Liquids*, John Wiley & Sons, New York, 1984.
11. G.C. Pimentel, A.L. McClellan, *The Hydrogen Bond*, Freeman, San Francisco, 1960.
12. J.C. Decius, R.M. Hexter, *Molecular Vibrations in Crystals*, McGraw-Hill, New York, 1977.
13. R. Zbinden, *Infrared Spectroscopy of High Polymers*, Academic Press, New York, 1964.

Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

Guido Busca and Carlo Resini
University of Genova, Italy

1 Introduction **1**
2 Vibrational Modes of Inorganic Solids **2**
2.1 Molecular Species 2
2.2 Crystalline Solids 2
2.3 Amorphous Solids 3
3 Infrared Spectroscopy of Inorganic Solids **3**
3.1 Transmission–Absorption Technique 4
3.2 Reflection Techniques 5
3.3 Diffuse Reflectance 7
3.4 Emission Technique 8
3.5 Photoacoustic and Photothermal Techniques 8
4 Raman Spectroscopy of Inorganic Solids **9**
5 Inelastic Neutron Scattering Spectroscopy **10**
6 Applications of Vibrational Spectroscopies to the Characterization of Inorganic Solids **11**
6.1 Simple Anhydrous Oxides 12
6.2 Complex Oxides 17
6.3 Vibrational Features of Complex Oxides of Practical Interest 18
6.4 Metal Hydroxides 18
6.5 Hydrated Compounds 20
6.6 Oxo Salts 20
6.7 Other Inorganic Materials 25
6.8 Unusual Applications 28
Abbreviations and Acronyms **28**
Related Articles **28**
References **29**

The vibrational spectroscopies are those producing the entire (or part) vibrational spectrum of a sample, i.e. the collection of transition energies between its vibrational states. The simplest way to cause vibrational excitation is to allow the chemical species in its vibrational ground state to absorb an energy quantum from electromagnetic radiation of an appropriate energy, i.e. in the medium infrared (IR) or in the far infrared (FIR) regions. Transmission–absorption IR spectroscopy comprises the analysis of the quanta that are actually absorbed by a polyatomic chemical species and those that are not absorbed (so are transmitted). Other IR techniques relate to the analysis of the radiation that is emitted, diffused, or reflected by the sample. Another way to obtain part or all of the vibrational spectrum of a chemical species is to look at the inelastic scattering of beams other than IR radiation. Raman spectrometry is the analysis of the ultraviolet (UV), visible or near infrared (NIR) radiation that is scattered inelastically by the sample. Inelastic neutron scattering (INS) is the analysis of the energy of formerly monochromatic neutron beams after inelastic scattering by the sample. Analyses of the number and energies of the vibrational transitions of the sample, detected by each technique, yield information on the geometric (molecular and crystallographic) structure of the sample and on the strength and nature of the chemical bonds it contains.

1 INTRODUCTION

Any chemical species containing more than one atom – polyatomic molecules, molecular ions, molecular radicals, and any crystal – can be described in terms of atomic positions, bond lengths, and bond angles. However, any atom vibrates around its equilibrium position, giving rise to vibrational modes. At absolute zero temperature all polyatomic chemical species lie at the vibrational ground state. However, they can undergo vibrational excitation, thereby occupying higher-energy vibrational states. The energy differences between nearest vibrational states lie in the medium and FIR ranges, generally defined by the wavenumber limits: $4000–400\,cm^{-1}$, i.e. $0.496 > E > 0.0496\,eV$ for medium IR; and $400–10\,cm^{-1}$, i.e. $0.0496 > E > 1.24 \times 10^{-3}\,eV$ for FIR.

The simplest way to cause vibrational excitation is to allow the chemical species to absorb an energy quantum from electromagnetic radiation of an appropriate energy, so giving rise to a vibrational transition from the ground state to an excited vibrational state. Analysis of the quanta that are actually absorbed by a polyatomic chemical species, and those that are not absorbed (so are transmitted), gives information on the vibrational structure of these species and, consequently, on its chemical and geometric structure. These are the fundamentals of transmission/absorption IR spectroscopy, which has been under continuous development since the late 19th century and has became one of the most widely used techniques for chemical analysis.[1]

Over time, it has been recognized that other IR techniques (reflection, emission, and photothermal techniques) can also be applied, and that IR spectroscopy only gives partial information on the vibrational structure

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of any polyatomic species. In fact, selection rules apply to IR light absorption phenomena, so that only those vibrational modes associated with changes in the molecular dipolar moment can be directly excited.

Another way to observe the vibrational spectrum of a chemical species is to look at the inelastic scattering of beams other than IR radiation. When such beams are scattered by a polyatomic species, some photons lose or gain energy as a result of interaction with the vibrational states of the molecule. The vibrational spectrum is obtained by measuring the difference between the energies of the incident beam and the small fraction that is inelastically scattered. If the incident photons are of relatively high energy – UV, visible, or NIR – we are dealing with Raman spectroscopy.[2] Raman observed that if a monochromatic high-energy radiation interacts with a chemical species, the scattered radiation is no longer monochromatic, now containing photons whose energy is produced by the sum or the difference of the original radiation energy with the energies of some vibrational transitions of the chemical species. As the selection rules for Raman-active modes do not correspond to IR-active modes in centrosymmetric chemical species, and only partly correspond (generally) for noncentrosymmetric species, Raman spectroscopy is in most cases complementary to IR spectroscopy.

The vibrational spectra of polyatomic chemical species also can be probed by the inelastic scattering of particles other than photons. The use of electrons gives rise to the techniques of electron energy loss spectroscopy (EELS) and inelastic electron tunneling spectroscopy (IETS). The incident particles can be neutrons, giving rise to INS spectroscopy, or atoms, giving rise to atom scattering spectroscopy (ASS). These techniques, developed in the 1960s by physicists, have found application in surface chemistry and physics. However, some may also be applied to bulk solids, such as geological and inorganic materials.

2 VIBRATIONAL MODES OF INORGANIC SOLIDS

2.1 Molecular Species

The total number of degrees of freedom associated with a chemical species containing N atoms is $3N$. If this chemical species is a molecule in the gaseous state, three of these degrees of freedom are associated with its translations and another three to its rotations, so that in the most general case six modes are associated with external motion (rotations and translations). The remaining $3N - 6$ degrees of freedom are associated with internal, i.e. vibrational, modes. However, if the molecule is linear, only two degrees of freedom are associated with rotation, because no rotational freedom exists around the molecular axis. Thus, in this particular case, the number of degrees of freedom associated with vibrations is $3N - 5$. A complete treatment of the principles of vibrational spectroscopy is beyond the aim of this article. The methods for determination of the number and the optical activity of the vibrational modes of molecular species have been given elsewhere.[3–8]

The $3N - 6$ (or $3N - 5$) degrees of vibrational freedom give rise to vibrational modes that differ in their symmetry with respect to the symmetry elements of the molecular point group and in their multiplicity. The nondegenerate modes are denoted as A or B in relation to their symmetry or antisymmetry with respect to the rotation about the principal symmetry axis. Doubly degenerate modes and triply degenerate modes are denoted E and F, respectively. The subscripts g and u denote modes that are symmetric or antisymmetric, respectively, with respect to the center of symmetry (if any); the superscript symbols $'$ and $''$ distinguish modes that are symmetric or antisymmetric, respectively, with respect to symmetry planes not containing the principal axis ($\sigma \neq \sigma_v$). The subscripts 1, 2, 3 are related to the symmetry with respect to other symmetry axes.

Character tables, which can be found in several vibrational spectroscopy books, allow the determination, for any molecular point group, of the species (or irreducible representations) in relation to the symmetry elements typical of that group. As discussed below, the classification in terms of a particular symmetry species determines the activity – IR activity, Raman activity, both IR and Raman activity, or inactivity – of any mode.

2.2 Crystalline Solids

When a crystalline solid is considered, the N atoms present in the smallest (primitive) Bravais cell must be taken into consideration to count the fundamental vibrational modes. They give rise to $3N$ total degrees of freedom, three of which give rise to translational modes of the cell as a whole, which are the acoustic modes. Thus, the vibrational degrees of freedom (optical modes) are, in this case, $3N - 3$.

If the solid is molecular, the molecules (supposed to be formed by M atoms, where $M = N/r$ with r being the number of molecules in the smallest Bravais cell) can be treated as for the gas phase, thus giving rise to $3M - 6$ (or $3M - 5$ if linear) vibrations for each molecule. The degrees of freedom associated with the external modes of every molecular unit (which are $6r$ for nonlinear molecules and $5r$ for linear molecules) give rise to lattice vibrations (frustrated translations and rotations) and to three acoustic modes. However, the internal vibrations

of each molecules should in principle give rise to *r*-fold splitting, as a result of the coupling of the vibrations within its primitive unit cell as a whole.

The analysis of the symmetry of the chemical species under study (i.e. the point group for a free molecule, or the space and factor groups for a crystal[3]), according to the site symmetry of every atom, allows the determination of the irreducible representation of the total modes. After subtraction of the translational and rotational modes (the acoustic modes for the crystals), the irreducible representation of the vibrational (or optical) modes can be obtained. This means that the number of vibrational modes belonging to the symmetry species associated with the molecular or crystal symmetry can be counted. Consequently, the number of active modes can be counted, according to the symmetry selection rules of the different techniques (in particular IR and Raman).

In principle, every vibrational mode is due to motions of the chemical species (the molecule or the primitive unit cell) as a whole. Thus, molecular and crystal dynamics calculations should rigorously define the motion of every atom of the chemical species in terms of vibrational modes. For large and complex chemical species, this approach gives rise to a very complex picture so that the results are sometimes not easily interpreted and comparison between the vibrational behavior of similar species is frequently difficult.

However, a group approximation can be frequently used, in which the vibrational modes can be attributed approximately to the movements of small groups of atoms. This is especially the case if very different bonds are present in the molecule, so that the coupling of their movements is negligible. In this way, it is possible to dissect the chemical species under study and consider separately the vibrations of groups of a few atoms (such as the functional groups in organic compounds). This makes possible a satisfactory, although approximate, interpretation of the vibrational spectra of polyatomic molecules and of some crystals, and allows easy comparison between the vibrational features of related chemical species.

In the case of crystalline solids, more than one equivalent structural unit may be present in the primitive cell. This results in splittings of the fundamental vibrational modes of these units. In the case of many crystalline solid materials covalent units (e.g. oxo anions for oxo salts) are present, together with other groups bonded by ionic bonds (e.g. the cations in the oxo salts). According to the group approximation, the internal vibrations of the covalent units can be considered separately from their external vibrations (hindered rotations and translations of the group that finally contribute to the lattice vibrations and to the acoustic modes of the unit cell) and those of the other units. If a number of covalent structural units are present in the primitive cell, their internal modes become split.

In the spectra of solids, the polar phonons (those that are IR active) split into two components, the transverse optical (TO) mode and the longitudinal optical (LO) mode. This is because the electric field associated with the transverse wave is zero, whereas that associated with the longitudinal wave is nonzero; when coupling of these modes occurs with the electric fields associated with the vibration, this gives rise to the separation $\nu_{LO} > \nu_{TO}$. This factor is relevant to the shape and interpretation of the IR spectra of solid materials, as discussed below.

2.3 Amorphous Solids

Amorphous solids differ from crystalline solids because no long-range order occurs. So, variable coupling exists between the vibrational modes of similar or equivalent structural units. Consequently, amorphous solids can be treated like liquids and gases. The vibrational spectra of amorphous materials can present a smaller number of broader features than those of corresponding crystalline materials, whereas crystal coupling effects can produce multiple sharp features. An example of this is treated below in some detail, concerning the forms of silica.

3 INFRARED SPECTROSCOPY OF INORGANIC SOLIDS

As already mentioned, a variety of IR techniques are in use to give information on the structure of different solids. When electromagnetic radiation is incident on a sample, the light may (essentially) be absorbed, reflected, or transmitted. Equation (1) follows from the conservation of energy:

$$I_0 = I_R + I_T + I_A \tag{1}$$

where I_0, I_R, I_T, and I_A are the intensities of the incident, reflected, transmitted, and absorbed radiation, respectively. Dividing Equation (1) by I_0 gives Equation (2):

$$R + T + A = 1 \tag{2}$$

where $R = I_R/I_0$, $T = I_T/I_0$, and $A = I_A/I_0$ are the apparent reflectance, transmittance, and absorptance, respectively. In any IR technique, two selection rules are valid. They are simplified as follows:

$$\Delta v = \pm 1 \tag{3}$$

$$\left(\frac{\partial \mu}{\partial Q}\right)_0 \neq 0 \tag{4}$$

Equation (3), where v is the vibrational quantum number, means that only transitions between nearest

vibrational states can directly occur in the case of the harmonic oscillator. However, this condition is relaxed in the case of anharmonic oscillators, so that not only fundamentals but also overtones and combination modes can sometimes be observed, although they are usually weak. Equation (4) indicates that only vibrational modes associated with a change in the dipole moment μ of the molecule between the extremes of the atomic displacements (Q is the normal coordinate) different from zero can be directly excited. According to the Lambert–Beer law (Equation 5):

$$A = -\log_{10}\frac{I_T}{I_0} = \varepsilon c l \tag{5}$$

the absorbance A linearly depends on the molar concentration c of the absorbing species, through the molar absorption coefficient ε and the sample path length l. This relationship (which is rigorously valid only in non-scattering media) is the basis for quantitative analysis performed through IR spectroscopies. Recent work has shown how to extend the range of Beer's law in FTIR (Fourier transform infrared) spectrometry.[9]

However, the absorbance at a particular wavelength also depends upon the population of the corresponding vibrational states. According to Boltzman's law, Equation (6),

$$\frac{N_i}{N_0} = \frac{g_i}{g_0}e^{-\Delta E/kT} \tag{6}$$

where N_i are the state populations and g_i are the state multiplicities, ΔE is the energy difference between the states 0 and i and k is the Boltzmann constant. The ΔE values for fundamental vibrational transitions fall in the IR region, so that the population of the first excited states is very small at room temperature and lower than the population of the ground state even at 1000 K. This means that only transitions originating from the ground state can be ordinarily excited at room or relatively low temperature.

Commercial FTIR spectrometers, available at moderate costs, offer short recording times (one to a few minutes at most), high resolution (0.5–4 cm^{-1}) and are capable of obtaining desirable signal-to-noise ratios for most types of sample. These performances fulfill the requirements for the study of inorganic materials, in relation to the intrinsic line width of their modes (generally some tens of wavenumbers) and the optical properties of the samples. The different experimental set-ups are described below.

3.1 Transmission–Absorption Technique

This technique was originally applied by the pioneers of IR and is probably still the most widely used, even though its application to solids involves problems of sample form and preparation. Working with monocrystals or large particles is difficult, due to the high amount of absorbing species necessarily placed in the IR beam. However, this technique can be successfully applied to powdered materials, even though a problem arises from the radiation scattering.

When a radiation is incident on a nonabsorbing layer consisting of N particles per unit volume, its intensity decreases exponentially with the sample thickness x, following Equation (7):

$$I_T = I_0 e^{-\varphi x} = I_0 e^{-N\sigma x} \tag{7}$$

Thus an apparent absorbance A' can be defined (Equation 8) for an absorbing and scattering medium:

$$\begin{aligned} A' &= -\log_{10}\frac{I_T}{I_0} = 0.43N\sigma x = 0.43N(\varepsilon' + s)x \\ &= A + 0.43Nsx = A + S' \end{aligned} \tag{8}$$

where φ is the linear attenuation coefficient and σ is the linear attenuation coefficient of one particle. The coefficient σ contains a component due to absorption, ε', and a component due to scattering, s. So the apparent absorbance is composed of the true absorbance A superimposed upon the component due to the light scattering S'. This simplified approach is similar to that given by Henry[10] who proposed Equation (9) for the component dependent on scattering:

$$S' = kx\,d\frac{(n-1)^2}{\lambda^2} \tag{9}$$

where k is a constant independent of λ but dependent on the particle arrangement, d is the particle diameter and n is the refraction index of the material. However, according to the Rayleigh theory, S' depends on the third power of the particle diameter (and so linearly on the particle volume) and on the fourth power of the radiation frequency ν. Studies of effect of scattering on the IR transmission spectra of powder samples[11,12] have shown that if the particle diameter is sufficiently small (considerably smaller than the IR wavelength), the scattering of the IR radiation is very small. Thus, finally, the apparent absorbance of a layer of a powder depends on its own absorptivity, on the disk thickness and density, on the particle size, and on the wavelength.

To obtain a good quality powder sample for the transmission–absorption IR technique, it is necessary to prepare a layer appropriately diluted and sufficiently thin. The most used technique uses KBr pressed disks. KBr is an easily available powdered material which is not absorbing in the medium IR region (down to near

Table 1 Materials used for discs and windows that transmit medium IR or FIR radiation

Name	Formula	Frequency range (cm^{-1})	Comments
Sodium chloride	NaCl	$\nu > 590$	Hygroscopic, water soluble
Potassium bromide	KBr	$\nu > 340$	Hygroscopic, water soluble
Cesium iodide	CsI	$\nu > 200$	Very hygroscopic
Calcium fluoride	CaF_2	$\nu > 1140$	Not hygroscopic, allows high temperature
Barium fluoride	BaF_2	$\nu > 840$	Not hygroscopic, allows high temperature
Silver chloride	AgCl	$\nu > 450$	Soft, not hygroscopic
Silver bromide	AgBr	$\nu > 300$	Soft, not hygroscopic
Polyethylene	PE	$625 > \nu > 33$	Low temperature, no organic solvents
Quartz	SiO_2	$\nu > 2500$	Only high ν
Sapphire	α-Al_2O_3	$\nu > 1780$	Only high ν
Zinc sulfide	ZnS	$\nu > 840$	Not hygroscopic
Germanium	Ge	$\nu > 600$	Hard, brittle, high reflection loss
Zinc selenide	ZnSe	$\nu > 455$	Low absorption
Cadmium telluride	CdTe	$\nu > 320$	Quite soft
Thallium bromide, iodide	Tl(Br,I)	$\nu > 250$	Toxic
Silicon	Si	$\nu > 660-400 > \nu$	Hard, brittle, high reflection loss

400 cm^{-1}, i.e. it cuts out the FIR). It can be easily mixed homogeneously with the powder to be investigated, and pressed, thus obtaining diluted self-supporting disks very useful for IR transmission. The spectra so obtained give rise to a baseline due to scattering that gives very high transmission in the low-frequency region (down to the cut-off near 400 cm^{-1} due to absorption by KBr itself). Other materials (see Table 1) give similar pressed disks, with cut-off limits at even lower frequencies, but are more expensive. Alternatively, the powders can be simply deposited in the form of a thin layer on a disk of a transmitting medium.

As KBr and the other alkali halides absorb in the FIR region, other binding materials must be used. The preparation of disks with polyethylene (which is transparent in most of the FIR region) is the most used technique. However, in this range the scattering is very low because of the long wavelengths involved, so that it is sometimes possible to simply deposit the powder on a disk of a transmitting material (see Table 1) using a solvent that later evaporates.

Another technique, used mostly in the past, is to mull the powder with organic liquids, such as liquid paraffin (nujol) and perhalogenated organics. The powder is mixed to make a paste from which a layer is formed. Absorptions from the mulling liquid (i.e. in the C–H stretching and deformation regions for paraffin oils, or C–X modes for halocarbons) also occur and obscure some spectral regions.

3.2 Reflection Techniques

When incident radiation interacts with a flat surface, the reflectances of the parallel and perpendicular polarized light with respect to the incidence plane, $R_\parallel$ and $R_\perp$, are given by Equations (10) and (11):

$$R_\parallel = \frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}}^2 \tag{10}$$

$$R_\perp = \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon + \sin^2\theta}}^2 \tag{11}$$

where ε is the complex dielectric constant of the material (assumed to be isotropic) and θ is the incidence angle. From these relationships it follows that the reflected light is mainly polarized perpendicular ($R_\perp$ is large, whereas $R_\parallel$ has a minimum for $0 < \theta < 90$). When $\theta = 0$, substituting $\varepsilon = (n - ik)^2$ gives Equation (12):

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \tag{12}$$

where n and k are the real and imaginary parts of the absolute index of refraction. For conducting materials, such as metals, the refraction index is totally imaginary, which means that $n = 0$ so that $R = 1$ at any frequency, and the radiation is totally reflected. This is the case of total (or specular) reflection.

For insulating or weakly semiconducting materials, such as most inorganic compounds, $k = 0$ and n is nearly constant (i.e. the index of refraction is real) in the IR spectrum except near the absorption frequencies, i.e. for $\nu > \nu_{LO}$ and $\nu < \nu_{TO}$. This means that outside of the latter skeletal region, monocrystal surfaces of inorganic compounds only weakly reflect the radiation, R being smaller with greater n. This means that all light is essentially transmitted or, better, refracted. However, for $\nu_{LO} > \nu > \nu_{TO}$, the index of refraction is imaginary and $k > 0$ goes through a maximum. This means that near the skeletal absorption region, R again becomes

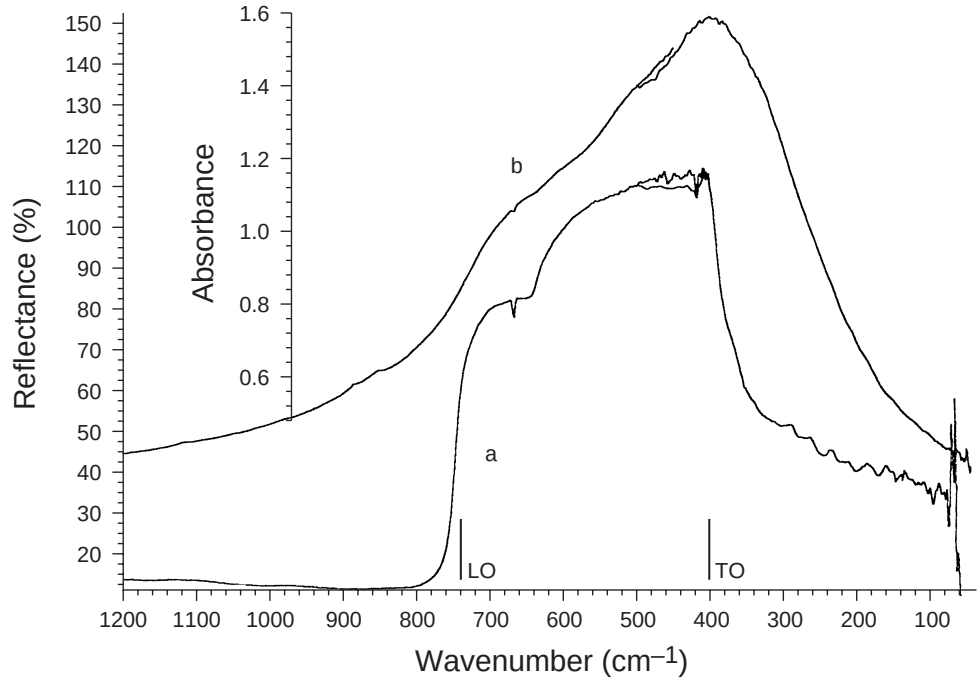

Figure 1 (a) FTIR reflectance spectrum of an MgO monocrystal (incidence angle 26.5°, face [001]). (b) FTIR and FT/FIR absorption–transmission spectra of MgO powder. (Reproduced with permission from G. Busca, *Catal. Today*, **27**, 323–352 (1996).)

near to 1, i.e. the skeletal vibrations of inorganic compounds give rise not only to absorbed radiation in transmission experiments but also to reflected radiation in the reflection experiments (the reststrahlen effect). Thus, the specular reflectance for insulating materials, both in the form of monocrystals[5] and in the form of sintered pellets,[13] is frequently the basis for the best determination of the skeletal spectrum, as far as the IR-active modes are concerned.

Figure 1 shows the reflection spectrum of an MgO monocrystal compared with the transmission–absorption spectrum of MgO powder. MgO (periclase) has the rock-salt structure (space group $Fm3m \equiv O_h^5$, with $Z = 4$, and with one formula unit only in the smallest Bravais cell). The six degrees of freedom associated with the two atoms in the primitive cell give rise to a triply degenerate acoustic mode and to a triply degenerate optical mode. The irreducible representation for the optical modes is given by Equation (13):

$$\Gamma_{\text{opt}} = F_{1u}(\text{IR}) \tag{13}$$

Thus, rock-salt-type compounds are Raman-silent (as far as the fundamentals are concerned, see below) and only show one IR-active fundamental mode. The TO quoted for MgO at $401\,\text{cm}^{-1}$ (Figure 1, curve a) corresponds to the lower frequency limit of the reflectance band, whereas the LO corresponds to the higher-energy limit of the reflectance, and is quoted at $718\,\text{cm}^{-1}$. For $\nu_{\text{LO}} > \nu > \nu_{\text{TO}}$ the reflectance is strong and almost constant, but outside these limits the reflectance is very small. The transmission–absorption IR spectrum of the MgO powder shows the maximum slightly above ν_{TO} ($407\,\text{cm}^{-1}$ in the spectrum of a polyethylene disk) and a shoulder near ν_{LO}. Other components arise from particles with different morphologies[14] although it has been shown that the microcrystal powder spectrum of MgO is also affected by surface relaxation.[15]

The total reflection of IR incident light at any wavelength from a surface can occur in two cases: (a) when the beam arises from a transmitting medium and is incident on the surface of a conducting material; and (b) when the beam arises and is incident within a transmitting medium, but the medium from which the beam arises has a higher refraction index than the other, and the incidence angle is greater than the limit for refraction.

It has been recognized in the first case that if molecules are adsorbed on the surface of a metal, only part of the grazing incident radiation is reflected, with part being absorbed by the adsorbed species. This absorption is greatly increased if the incident radiation is polarized perpendicular to the metal surface. This is infrared reflection absorption spectroscopy (IRRAS) which, mainly after the work of Greenler[16] and Pritchard,[17] is applied widely to surface studies on metal surfaces. Under these conditions, the additional selection rule of Equation (14)

applies:

$$\left(\frac{\partial \mu_{\perp}}{\partial Q}\right)_0 \neq 0 \tag{14}$$

which means that only the adsorbate vibrations associated with a change in the dipole moment perpendicular to the crystal surface are detectable. The total reflection technique can be applied to detect the growth of oxidic layers over metal surfaces (i.e. the skeletal bands of the oxide layers) and this is a widely applied technique in corrosion and electrochemical research,[18,19] colloid chemistry,[20] coating technology,[21] and heterogeneous catalysis.[16,17]

The second case is internal reflection spectroscopy, utilized in attenuated total reflection (ATR)[22] where an absorbing layer is deposited on one or two external surfaces of a prism. The light penetrates the prism from a free surface and is multiply reflected by the other faces. During the reflection, part of the light penetrates the external layer and is absorbed by it. Thus, in the case of the internal reflection techniques the reflectance is given by Equation (15):

$$R = 1 - A = 1 - \varepsilon c d \tag{15}$$

where d is the thickness of the layer penetrated by the radiation, c is the molar concentration of the absorbing species in that layer, and ε is the molar absorption coefficient of that species. This technique has found application in the examination of several bulk solids (e.g. polymers[23,24]), electronic devices (e.g. those constituted by silica layers on silicon[25,26]), and glasses.[27] This technique can also be applied to metal layers at the interface with a solution and in this case the signal can be surface enhanced.[28] This technique is of wide interest in the fields of electrochemistry[29] and colloid chemistry.[30] An interesting variation is cylindrical internal reflection (CIR).[31–33] As an example of use, a hydrogel slurry can be analyzed using cylindrical crystals immersed in the medium so that IR radiation is multiply transmitted and reflected at the internal surface. This technique could be successfully applied to heterogeneous catalytic systems at the water/solid interface, or in stages of catalyst preparation (e.g. zeolite syntheses).

Another interesting variant of the total reflection technique is surface electromagnetic wave spectroscopy (SEWS), which consists in the generation of a surface plasmon on a substrate by frustrated total internal reflection in a prism located a few micrometers from the surface. This plasmon is decoupled by a second prism. Some interesting results have been reported with this technique concerning surface species on alumina.[34]

3.3 Diffuse Reflectance

The IR spectroscopy of powders in the diffuse reflectance mode has grown strongly, aided by the commercialization of appropriate attachments and cells. This technique is particularly attractive because it does not require much sample preparation (the powder is simply deposited in a sample holder). As this technique takes advantage of light scattering, it is very useful for studies in surface chemistry.

This technique requires collection of the radiation scattered by the sample with appropriate collecting mirrors, such as an integrating sphere. Most photons are essentially simply scattered, but those corresponding to the energies of vibrational transitions are potentially absorbed. The interpretation of diffuse reflectance spectra is based on the phenomenological theory of Kubelka and Munk.[35–37] The Kubelka–Munk (K–M) function is given by Equation (16):

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} \tag{16}$$

where R_∞ is the reflectance of an infinitely thick layer and may in practice be substituted by $R_\infty - R'_\infty$ (i.e. by the reflectance spectrum from which the reflectance of a transparent reference material such as KBr has been subtracted). The K–M function depends linearly on the absorption coefficient k (such that $k = 2303\varepsilon c$, where ε is the molar absorption coefficient and c is the molar concentration) and inversely on the scattering factor s. Obviously, the greater the absorption coefficient and the smaller the scattering, the higher the K–M function. An approximation of the K–M function is that k is treated as a variable but s is assumed constant. This is not true in the vicinity of strong absorptions.

More realistic and complex theories have subsequently been developed,[38–41] according to the quite complex nature of the phenomena involved in the diffuse reflectance of light. Experimental studies[42–45] showed that this technique is affected by particle size, granulometric distribution, and the index of refraction of the particles; the latter has an important role when the particle size is near the wavelength of the IR radiation. Diffuse reflectance infrared Fourier transform (DRIFT) studies in the FIR region allow the detection of the skeletal spectra of materials, such as mixed oxide catalysts, pigments, metal halides.[46]

One important practical problem in diffuse reflectance measurement is the need to cancel the specular reflectance by the front surface of the sample, which generates negative bands in the DRIFT spectrum, so causing apparent shifts in the true absorption bands. This is achieved by using appropriate cell designs in commercial DRIFT attachments (see Figure 2).[47,48]

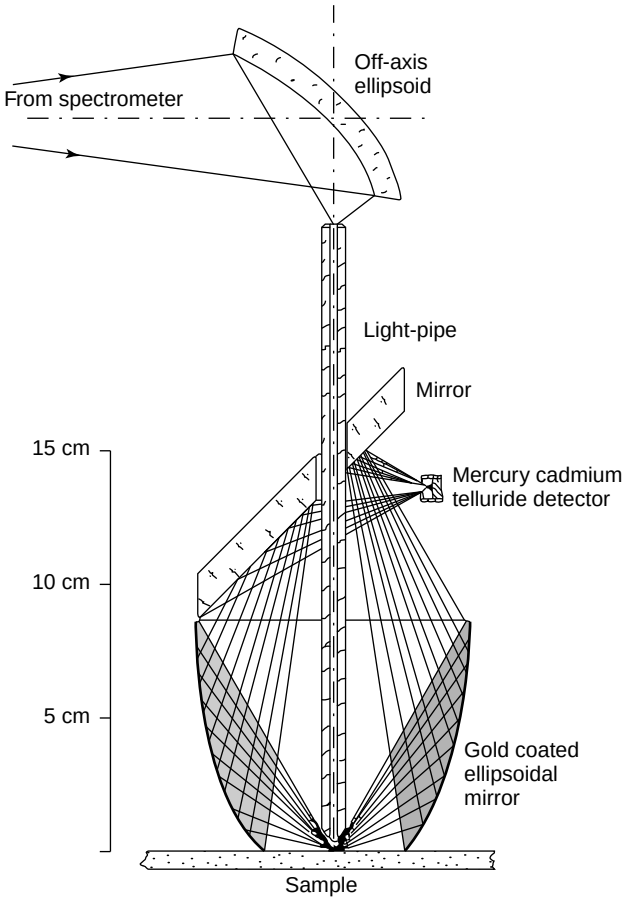

Figure 2 Diagram of a DRIFT accessory designed to avoid the collection of totally reflected light. (Reproduced with permission from A.N. Davies, *Spectrosc. Eur.*, **6**(4), 25–28 (1994).)

3.4 Emission Technique

According to Kirchhoff's law (Equation 17),

$$\varepsilon = \frac{I_{em}}{I_b} = \alpha \tag{17}$$

the emittance ε (i.e. the ratio of the light emitted by the sample I_{em} with respect to that emitted by a blackbody at the same temperature, I_b) is equal to the absorptance, α, of the sample. The emissivity is proportional to the fourth power of the temperature difference between the emitting sample and the detector (Stefan's law). This implies that emissivity is sufficiently strong at relatively high temperatures to give a good signal-to-noise ratio in a large part of the IR spectrum. In fact, the blackbody emission below 1000 K shows its maximum in the medium IR region. According to Koenig[23,24] the emission spectra of polymers can be taken at above 70 °C with sufficiently good signal-to-noise ratios in the range 1800–450 cm^{-1} (outside this range the signal-to-noise ratio is very low). As pointed out by Griffiths et al.,[49] temperatures well above 500 °C are needed to have a sufficiently good signal-to-noise ratio in the CH stretching region near 3000 cm^{-1}.

A number of problems arise in connection with the use of infrared emission spectroscopy (IRES). One of them arises from the existence of temperature gradients that can cause self-absorption of the emitted radiation by the colder outer parts of the sample itself; another is concerned with selective reflection that occurs in the vicinity of strong absorption bands. This reduces the absorptance and hence the emittance. Moreover, perturbations can be created by reflections and emission by the cell elements. However, these problems can in part be overcome so that IRES can be successfully used in the fields of polymers[23,24] and corrosion science.[50] The IRES of metal oxides in the field of surface chemistry and catalysis has been reviewed recently by Sullivan et al.[51] These authors reported several examples of emission spectra of oxide catalysts and of adsorbates on supported metals, and cite at least 11 papers concerning metal oxide surfaces. Some of the advantages of this technique are the very easy sample preparation and its easy applicability at high temperatures (150–400 °C). It can be applied to investigate the temperature dependence of the radiative properties of materials, such as glasses.[52] Figure 3 is a diagram of an IRES attachment.

3.5 Photoacoustic and Photothermal Techniques

When an IR beam is incident on a solid surface it can be absorbed in part, and this leads to its conversion into heat. If the beam is modulated (as in any interferometers such as those of FTIR instruments) and the solid is in contact with a gas (air, He, Ar, ...), its conversion

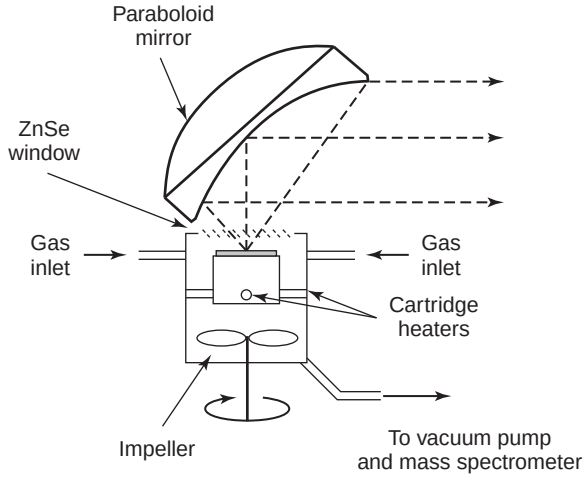

Figure 3 Experimental arrangements for IRES. (Reproduced with permission from D.H. Sullivan, W. Curtis Conner, M.P. Harold, *Appl. Spectrosc.*, **46**, 811–818 (1992).)

to heat gives rise to an acoustic signal. In fact, the periodic temperature rise so obtained causes a periodic modulation of a gas pressure in the cell that can be detected by a sensitive microphone (Figure 4). This acoustic signal will be the more intense the stronger is the absorption at a particular wavelength. Reflected and scattered light are not absorbed and hence do not cause a photoacoustic signal. However, light absorbed by the gas over the sample causes signals. This makes necessary the use of monoatomic nonabsorbing gases (He, Ar, . . .). The photoacoustic effect, discovered as early as in 1880 by Bell,[53] could be successfully applied essentially only after the work of Rosencwaig and Gersho[54,55] in 1976. The main limitations of this technique are: (a) it requires a gaseous atmosphere; (b) the cell needs a microphone close to the sample (see Figure 7) so the sample cannot be heated and otherwise activated conveniently; and (c) the technique has an intrinsically low signal-to-noise ratio. However, the photothermal effect is much more efficient for species in the vapor phase than for bulk and surface species. Photoacoustic spectrometry (PAS) is not exactly a surface spectroscopy, because the penetration of the thermal effect is always significant. The extent of this depends on the modulation frequency, which varies with wavenumber in a Fourier transform (FT) spectrum. The PAS technique has found successful applications in polymer science,[23,24] surface carbon studies and coal characterization,[56,57] and in heterogeneous catalysis.[49,58–62]

Another technique is photothermal beam deflection spectroscopy (PBDS), based on the mirage effect first reported by Boccarra et al.,[63] and developed in the IR region by Low et al.[64] In this case, the periodic temperature rise caused by the absorption of the modulated IR radiation (i.e. the photothermal effect) is detected optically because it causes periodic deflections of a laser beam passing close to the surface of the

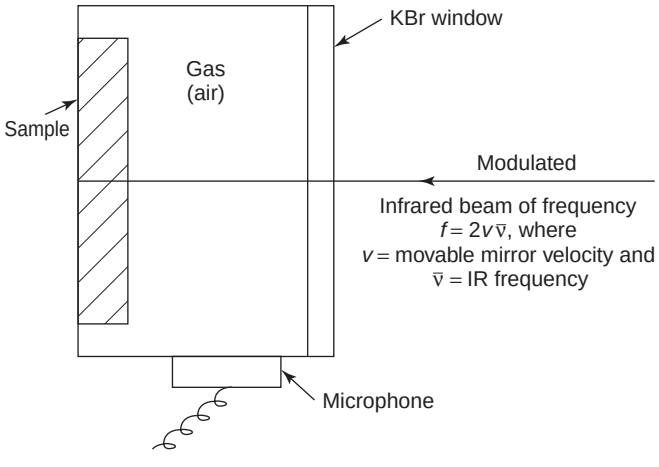

Figure 4 Diagram of a photoacoustic cell.

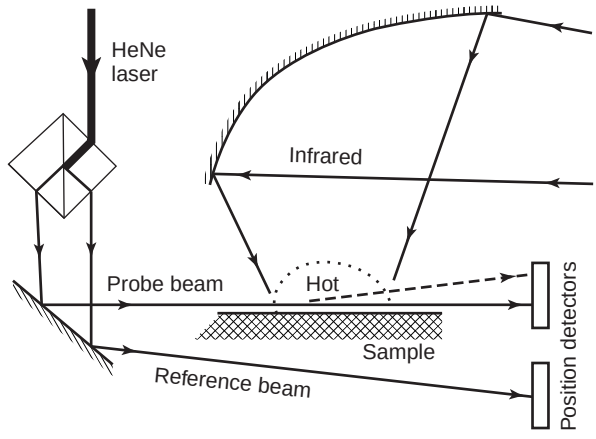

Figure 5 Diagram of FTIR PBDS set-up. (Reproduced with permission from C. Morterra, M.J.D. Low, *Langmuir*, **1**, 320–326 (1985).)

solid sample (Figure 5). The PBDS technique has some advantages over PAS because of its lower limits of sample dimensions, but has disadvantages because of the critical geometric set-up. As for PAS, PBDS can have advantages with respect to traditional IR for the detection of surface vibrations in very opaque materials. This has resulted in the application of PBDS to carbons and coals.[56,57]

4 RAMAN SPECTROSCOPY OF INORGANIC SOLIDS

The Raman effect is based on the inelastic scattering of electromagnetic radiation by molecular vibrations. When the radiation is incident on a sample, some photons are scattered due to the dipole moments induced through the static polarizability of the electrons. This is Rayleigh or elastic scattering. However, a small number of photons are scattered through modulation of the polarizability by electronic, vibrational, or rotational motion, leading to Raman or inelastic scattering.

Using a simple classical model it is easy to demonstrate[65] that when a chemical species is irradiated at the frequency ν_0 and its polarizability is modulated at the vibrational frequency ν_1, the time dependence of the induced dipole moment μ is given by Equation (18),

$$\mu(t) = \alpha_0 E_0 \cos 2\pi\nu_0 t + \frac{1}{2}\left(\frac{\partial\alpha}{\partial Q}\right)_0 Q_0 E_0 \cos 2\pi(\nu_0 + \nu_1)t + \frac{1}{2}\left(\frac{\partial\alpha}{\partial Q}\right)_0 Q_0 E_0 \cos 2\pi(\nu_0 - \nu_1)t \quad (18)$$

where α is the polarizability, which is a function of the normal coordinate Q in a Taylor's series

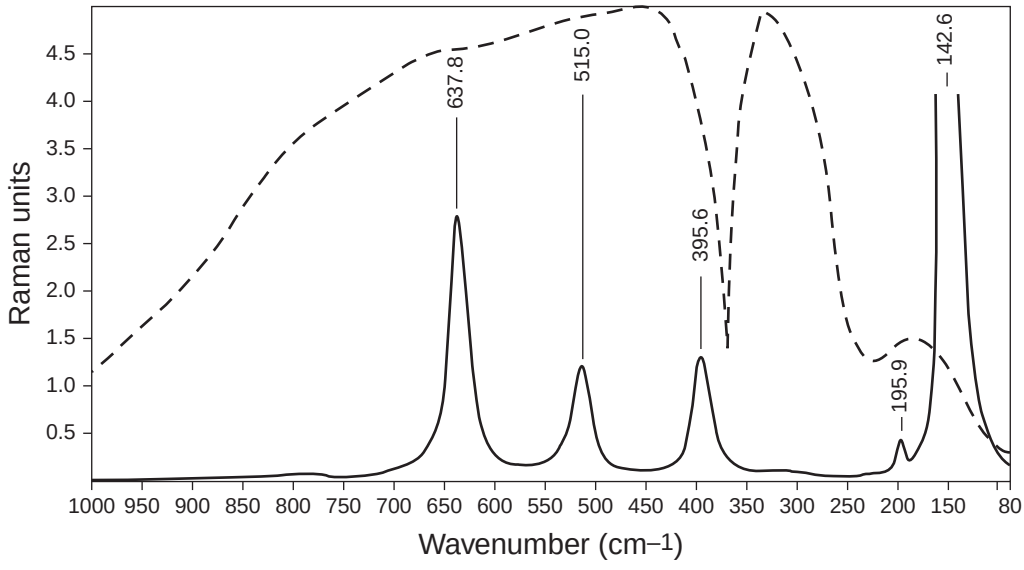

Figure 6 FTIR transmission–absorption spectrum (KBr and polyethylene disks, broken line) and FT/Raman spectrum (full line) of a commercial TiO_2 anatase powder. (Reproduced with permission from G. Busca, *Catal. Today*, **27**, 323–352 (1996).)

(Equation 19):

$$\alpha = \alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q + \frac{1}{2}\left(\frac{\partial^2 \alpha}{\partial Q^2}\right)_0 Q^2 + \cdots \qquad (19)$$

where α_0 is the static polarizability and E_0 is the amplitude of the radiation. The three terms in Equation (18) correspond to Rayleigh scattering (with $\nu = \nu_0$), to anti-Stokes–Raman scattering (with $\nu = \nu_0 + \nu_1$) and to Stokes–Raman scattering (with $\nu = \nu_0 - \nu_1$), respectively. From a quantum-mechanical treatment it is possible to rationalize why anti-Stokes–Raman scattering is far weaker than Stokes–Raman scattering, in particular for high frequency modes.

It is possible to conclude that Raman scattering occurs only if Equations (20) and (21) hold:

$$\Delta v = \pm 1 \qquad (20)$$

$$\left(\frac{\partial \alpha}{\partial Q}\right)_0 \neq 0 \qquad (21)$$

This also means that in Raman spectroscopy only transitions between nearest levels and those associated with changes in the polarizability α upon motion are allowed for a harmonic oscillator. The first selection rule is relaxed for anharmonic oscillators. Group theory shows that for centrosymmetric chemical species Raman-active modes are IR inactive, and vice versa. This is the mutual-exclusion rule. Other modes can be both IR and Raman inactive. For noncentrosymmetric molecular species, modes also occur that are both IR and Raman active.

The Raman scattering process is inherently weak (only a small fraction of photons are scattered, about 10^{-6}) and this limit is expected to be even more important for surface species. The intensity of the Raman signal of adsorbed species can be enhanced by surface-enhanced Raman scattering (SERS) that is, however, limited to rough surfaces of a few highly reflective metals.[66,67]

Raman spectroscopy of inorganic materials offers advantages over IR spectroscopy because of the weak Raman scattering intensity from the skeletal vibrations of ionic oxides and (generally) the absence of TO/LO splitting, offering rather high resolution, and high spatial resolution allowing microspectroscopy. As an example, the FTIR and FT/Raman skeletal spectra of a TiO_2 anatase powder sample are compared in Figure 6.

One of the problems with Raman spectroscopy is the superimposition of the Raman effect by fluorescence phenomena. To avoid this, FT/Raman spectrometers working with NIR lasers have been developed. Alternatively, anti-Stokes–Raman spectrometry may be used in spite of the weakness of the anti-Stokes–Raman signal.[68]

5 INELASTIC NEUTRON SCATTERING SPECTROSCOPY

Among the techniques based on the inelastic scattering of particles, the INS techniques allow important data to be obtained for the vibrational characterization of solids. The INS technique[69,70] consists of the analysis of the energy distribution among neutrons inelastically scattered by the chemical species under study, generally presented in a dispersed form such as in the form of powders. If the instruments operate with a polychromatic continuous

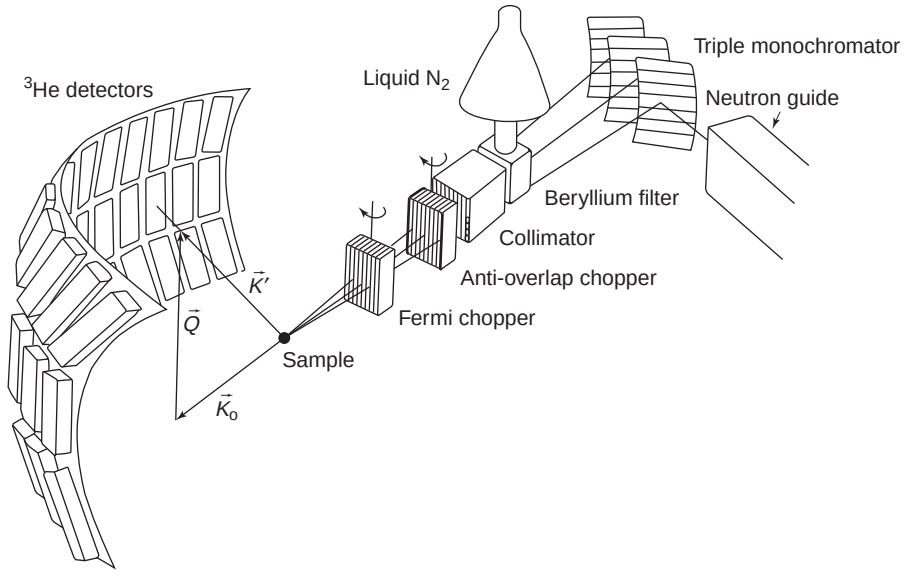

Figure 7 Diagram of a time-of-flight INS spectrometer. (Reproduced with permission from H. Jobic, in *Catalyst Characterization, Physical Techniques for Solid Materials*, eds. B. Imelik, J.C. Vedrine, Plenum Press, New York, 347–375, 1994.)

neutron source (from a fission nuclear reactor), the dispersion of the scattered beam is obtained before the sample by Bragg reflection from a single crystal operating as a diffraction monochromator.

However, if the instrument works with a pulsed neutron beam (spallation source) the energy analysis is obtained by using the time-of-flight technique, i.e. by measuring the time of arrival at the detector of the scattered neutrons after the pulse, through the source–sample–detector route. This arrangement is shown in Figure 7.

The INS technique has many disadvantages with respect to the much easier IR technique, essentially associated with the limited availability of a neutron beam source, the complexity of the neutron energy dispersion, and the neutron detection complexities. Moreover, the resolution is much smaller (near 2% of the energy transfer, which means 20 cm^{-1} at 1000 cm^{-1}); to increase the signal-to-noise ratio the spectra are usually recorded at very low temperatures.

In principle, there is a great advantage of INS with respect to both IR and Raman spectroscopies, because there are no selection rules and so all vibrational transitions are in principle INS active. However, the intensity of the INS peaks depends on the INS cross-section of all the atoms involved in the corresponding vibrational mode. This cross-section is characteristic of each element and does not depend on its chemical environment; it is at least 40 times greater for hydrogen than for any other element. Thus, the peaks associated with modes that involve hydrogen motions dominate the spectra, whereas those that do not involve hydrogen motions can be so weak as to be undetectable.

It is evident that INS spectra are particularly useful for the detection of vibrational transitions involving hydrogen-containing species and hydroxy groups. This is particularly true for the vibrational transitions associated with adsorbed hydrocarbons, hydroxyls and water on zeolites[71,72] and of hydrogen and absorbed species adsorbed on metallic catalysts (the latter, when unsupported, cannot be studied by IR and Raman), as well as hydride species on oxide and sulfide catalysts.[73–76] In the latter cases, these absorptions can fall into the region below 1000 cm^{-1}, where metal–oxygen and metal–sulfur skeletal vibrations occur and dominate IR and Raman spectra. As the skeletal vibrations of oxides and sulfides are very weak in INS spectra, those involving hydrogen are detected easily. This technique has been applied successfully to the analysis of hydrogen bonds in solids, such as the analysis of hydrogen bonds in metal hydroxides including iron oxyhydroxides,[77] the free proton dynamics in γ-MnO_2,[78] and hydrogen bonding in ices.[79] Low-energy vibrational modes in oxides have also been studied.[80]

6 APPLICATIONS OF VIBRATIONAL SPECTROSCOPIES TO THE CHARACTERIZATION OF INORGANIC SOLIDS

Several collections of IR, Raman, or both IR and Raman spectra of inorganic materials and minerals have been published[81–84] and are available electronically.[85] The following discussion briefly reviews some of the

applications of vibrational spectroscopies in the fields of characterization of such materials. A database of standard Raman spectra of minerals and related inorganic crystals has been published.[86]

6.1 Simple Anhydrous Oxides

Within basic inorganic chemistry[87–89] the oxides of nonmetals are defined as acidic oxides or anhydrides, whereas the oxides of metals are denoted as basic oxides. Some oxides have both acidic and basic character and are consequently defined as amphoteric. The bulk properties of these compounds are related to their crystal chemical features and are also reflected in their bulk vibrational properties.

6.1.1 Ionic/Covalent Nature of Solid Oxides and Their Vibrational Features

The oxides of low-valency metals (i.e. with cations of oxidation number $\leq +4$) are typically ionic compounds. They are most frequently easily obtained in crystalline forms.

For sufficiently small cations ($r < 0.8$ Å) the oxide structures are determined by cubic or hexagonal close packing of oxide ions with the cations in tetrahedral or octahedral interstices.[90–93] Bigger cations do not enter such interstices, so that structures with less densely packed oxide ions (e.g. simple cubic) and higher coordinations for the cations must occur (seven or eight). In ionic metal oxides coordination of the cations (four to eight) is generally higher than the valency (one to four) and this also occurs for the coordination of O^{2-} oxide ions (three to six).

The oxides of nonmetals are characterized by the considerable covalent nature of the bond between the nonmetal and oxygen, in agreement with the high electronegativity of the nonmetal elements. This frequently results in molecular structures for the nonmetal oxides, some of which are gaseous or liquid in normal conditions. Some form cyclic polymeric molecules and are solid in normal conditions (e.g. Se_4O_{12}, P_4O_{10}). By decreasing the element electronegativity, covalent network structures are formed for semimetal oxides (e.g. B_2O_3 or SiO_2 in their room-pressure forms). Although they can form crystalline phases, the solid oxides of semimetals are also characterized by the existence of very stable amorphous states and by their glass-forming properties.[94] These metastable amorphous or vitreous structures find many applications (e.g. for silica) and can be converted into stable crystalline phases at very high temperatures.

The covalent nature of the solid nonmetal oxides (molecular or network) is shown by the coordination around the element that either corresponds to its valency or is even lower. In such structures, the oxygen is either bridging between two metal atoms or bonded to one atom only with a former double bond.

The bulk acidity of these nonmetal solid oxides is associated with the acidity of the product of their reaction with water, giving an oxo acid. The acidity of oxo acids is associated with the strong covalency of the nonmetal–oxygen bond, which subtracts electronic charge from the oxygen atom thus allowing the O–H bond to be highly polarized and stabilizing the negative charge of the anion produced by dissociation. However, the acid strength of oxo acids is primarily enhanced by the presence of doubly bonded oxygen atoms around the nonmetal atom, which allows delocalization of the negative charge produced by the acid dissociation. For this reason, the higher the oxidation state of the nonmetal, the more the oxygen atoms are doubly bonded, and the stronger is the acidity of the oxo acid.

The oxides of the metal elements in very high oxidation states (e.g. Mn^{VII}, Cr^{VI}, Mo^{VI}, W^{VI}, V^{V}, Nb^{V}, Ta^{V}) are also frequently denoted as anhydrides, and give rise, by reaction with water, to acid species (sometimes polyoxoacids) that can be strongly acidic. Due to the high oxidation states of the metal elements in these compounds, their electronegativity is very high,[95] so the oxides and the derived acids have nearly the same properties as those from nonmetal oxides and acids.

The structural properties of the compounds of metals in very high oxidation states ($\geq +5$) differ from those of nonmetals, in that the coordination around the metal atom is elastic and can be very distorted. For example, the overall coordination of pentavalent vanadium[96] and hexavalent molybdenum in their oxides is frequently quite arbitrarily defined, and can vary from four to six depending on the environment. Sixfold coordination for these ions is frequently very distorted, with one or two very short metal–oxygen bonds and one or two very long so that, excluding the longest bonds, sixfold coordination frequently approaches fourfold coordination.[97]

In network oxide structures the metal–oxygen bonds are all almost equivalent, so that their movements couple extensively. Also, the stretching and deformation vibrations mix. As a result, assignments of the different vibration modes can be very approximate. However, the higher-frequency vibrational features are generally associated with stretching modes. The position of these modes can be assumed to be primarily dependent on the coordination of the involved metal ions and on the extent of the condensation of the resulting polyhedra (Table 2). This approach has been proposed by Tarte[98,99] and is a useful approximation for the interpretation of the IR and Raman skeletal spectra of oxides of unknown structure. However, and perhaps more correctly, analysis of the spectra suggests that the higher-frequency modes are

Table 2 Absorption range (cm^{-1}) of some elements and coordinated compounds

Atom X	Tetrahedral XO_4 (isolated)	Octahedral XO_6	
		Pure oxide	Isolated complex
Ge^{IV}	700–800	700–800	~500
Ti^{IV}	650–800	–	500–400
Al^{III}	700–850	–	500–350
Ga^{III}	600–750	450–350	–
Cr^{III}	–	450–300	–
Fe^{III}	550–650	–	400–300
Mg^{II}	550–700	450–300	–
Zn^{II}	400–650	<300	–
Fe^{II}	–	<350	–

mainly to be assigned to motions of the lighter oxygen atoms, so that they can be assigned to stretchings of oxygen in different coordination states. This alternative approach, first proposed by Beattie and Gilson,[100] is shown in Table 3.

Most simple oxide structures have been the object of extensive studies and complete vibrational analysis. Some examples are summarized in Table 4.

6.1.2 Silica

The different polymorphs of silica – α- and β-quartz, α- and β-cristobalite, tridymites, coesite, morganite, and silicalite – all have tetrahedrally coordinated silicon. Their vibrational structures, as well as that of amorphous or vitreous silica,[101,102] where Si is also tetrahedral, can therefore be discussed in parallel.

The IR spectra of all tetrahedral silica polymorphs have bands in three regions: 1300–950 cm^{-1}, very strong, hereinafter denoted ν_1; 850–600 cm^{-1}, medium strength, ν_2; and near 450 cm^{-1}, very strong, ν_3. Raman spectra show very weak peaks in the 1300–950 cm^{-1} and 850–700 cm^{-1} regions, and one or more very strong peaks below 550 cm^{-1}, in a position very sensitive to the overall crystal structure of the solid (ν_4). These features, also present in the spectra of amorphous or vitreous silica, represent vibrations of the basic structural units of both crystalline and amorphous silica, namely SiO_4 tetrahedra and bridging oxygens. Additional sharp peaks can be found in the IR and Raman spectra of crystalline polymorphs in the region below 700 cm^{-1}, which are associated with splittings of the above modes, due to crystal structure effects, and to torsional lattice modes.

The interpretation of the spectra of silica polymorphs can be attempted on the basis of a model assuming the bridging Si_2O, or T–O–T monomeric unit, as done by Bell and Dean[103] and by Galeener[104] for amorphous silica. This approach gives rise to a picture agreeing with that arising from the lattice dynamics calculations of van Santen et al.[105,106]

An oxygen atom in any tetrahedral-based SiO_2 polymorph or zeolite is bonded symmetrically with two tetrahedral cations (T) giving rise to a bent T–O–T group. Assuming that this group is symmetrical and isolated, the site symmetry of oxygen is C_{2v}. So, the three degrees of freedom of any oxygen atom consist of two in-plane movements (A_1 and B_1) and one out-of-plane movement (B_2). The B_2 movement is associated with the out-of-plane deformation of the T–O–T group, the B_1 with the asymmetric stretching, whereas the A_1 mode has the character of both a symmetric stretching and an in-plane bending mode (scissoring). According to this, the three main IR bands observed in all silica polymorphs are classically assigned to the asymmetric stretching (arising from the A_1 mode, ν_1), to the in-plane bending (arising

Table 3 Typical oxygen-atom vibrational modes for metal oxide structures

Situation	Number of oxygen atoms	Number of vibrational modes	Stretchings	Frequency region (cm^{-1})	Deformations
Terminal					
M=O	1	3	1	1500–800	1 Degenerate
MO_2	2	6	Asym + sym	1300–700	Bending, twisting, wagging, rocking
MO_3	3	9	Asym(2) + sym	1300–700	2 Asym bendings, sym bending, 2 rockings, torsion
Bridging					
M–O–M	1	3			
Bent			Asym + sym	1200–600	Rocking
Linear			Asym		Bending degenerate
Triply bridging					
M_3O	1	3			
Pyramidal			Asym(2) + sym	<600	
Planar			Asym(2)		Out-of-plane

Table 4 Crystal structures and ion coordinations in some solid binary oxide structures, and the irreducible representations of the optical modes

Cation valence	Structure type	International symbols	Schönflies notations	Z	Irreducible representations[a]
2	Rock salt	$Fm3m$	O_h^5	4	F_{1u} (IR)
	Zinc blende	$F\bar{4}3m$	T_d^2	4	F_2(IR,R)
	Wurtzite	$P6_3mc$	C_{6v}^4	2	A_1(IR,R) + $2B_1$(in) + E_1(IR,R) + $2E_2$(R)
	Tenorite	$C2/c$	C_{2h}^6	4	A_g(R) + $2B_g$(R) + $3A_u$(IR) + $3B_u$(IR)
3	Corundum	$R\bar{3}c$	D_{3d}^6	6	$2A_{1g}$(R) + $2A_{1u}$(in) + $3A_{2g}$(in) + $3A_{2u}$(IR) + $5E_g$(R) + $5E_u$(IR)
	"A"	$P\bar{3}m1$	D_{3d}^3	1	$2A_{1g}$(R) + $2E_g$(R) + $2A_{2u}$(IR) + $2E_u$(IR)
	"C" Bixbyite	$Ia\bar{3}$	T_h^7	16	$4A_g$(R) + $4E_g$(R) + $14F_g$(R) + $5A_u$(in) + $5E_u$(in) + $16F_u$(IR)
2, 3	Spinel	$Fd3m$	O_h^7	8	A_{1g}(R) + E_g(R) + F_{1g}(in) + $3F_{2g}$(R) + $2A_{2u}$(in) + $2E_u$(in) + $4F_{1u}$(IR) + $2F_{2u}$(in)
4	α-Quartz	$P3_221$	D_3^6	3	4A1(R) + 4A2(IR) + 8 E (R,IR)
	Rutile	$P4_2/mnm$	D_{4h}^{14}	2	A_{1g}(R) + A_{2g}(in) + B_{1g}(R) + B_{2g}(R) + E_g(R) + A_{2u}(IR) + $2B_{1u}$(in) + $3E_u$(IR)
	Anatase	$I4_1/amd$	D_{4h}^{19}	4	A_{1g}(R) + A_{2u}(IR) + $2B_{1g}$(R) + B_{2u}(in) + $3E_g$(R) + $2E_u$(IR)
	Fluorite	$Fm3m$	O_h^5	4	F_{1u} (IR) + F_{2g}(R)
	Baddeleyite	$P2_1/c$	C_{2h}^5	4	$9A_g$(R) + $9B_g$(R) + $8A_u$(IR) + $7B_u$(IR)
5	V_2O_5	$Pmmn$	D_{2h}^{13}	2	$7A_g$(R) + $7B_{1g}$(R) + $3B_{2g}$(R) + $4B_{3g}$(R) + $3A_u$(in) + $3B_{1u}$(IR) + $6B_{2u}$(IR) + $6B_{3u}$(IR)
6	ReO_3	$Pm3m$	O_{h1}	1	$2F_{1u}$(IR) + F_{2u}(in)

[a] R = Raman; in = inactive.

from the B_1 mode, ν_2) and to the out-of-plane rocking (arising from the B_2 mode, ν_3) of the T−O−T groups. However, in this manner the movements of Si (or T) atoms are neglected, and the interpretation of the Raman spectra is difficult. However, if account is taken of an isolated symmetric Si_2O molecule (C_{2v} point group) where Si atoms also are free to move, its total of nine degrees of freedom (six of which arise from silicon and three from oxygen) constitute three translations, three rotations, and three vibrations. The three vibrations are the asymmetric stretching (B_1), the symmetric stretching (A_1) and the bending mode (A_1), whereas the B_2 mode becomes a rotation. So, if Si atoms are free to move, the in-plane motion of oxygen (A_1 symmetry) splits into two components, a bending mode (ν_2) and a symmetric stretching mode (ν_4). By analogy, and by developing the central force network model of Sen and Thorpe,[107] Galeener[104] showed that four main vibrational modes are expected for a tetrahedral silica network, i.e. a symmetric stretching, an antisymmetric stretching, a bending, and a rocking mode. In the C_{2v} symmetry of the T−O−T bent units, the modes with A_1, B_1, and B_2 symmetry are both IR and Raman active. The strongest Raman mode, ν_4, is assumed by Galeener[104] to be a symmetric stretching mode mainly because the Raman peak intensity is generally greater for stretching than for bending modes. However, the coupling of this mode with the scissoring mode provides an explanation of the low frequency of the strongest Raman peak (550–350 cm^{-1}) in all silica polymorphs. It is also clear that in the solid state, in order for all T−O bonds to expand in phase, most T−O−T angles must contract simultaneously. This mode is consequently highly sensitive to the structure of the polymorph, as is indeed observed. The position of the strongest Raman peak, ν_4, in fact, strongly depends on the type of rings present in silicas and silicates, as shown by Sharma et al.,[108] and on the density of the unit cell, as shown by Kingma and Hemley for silicas.[102] This is shown in Table 5.

As for the IR spectrum, the strongest complex band is ν_1, observed in the 1300–950 cm^{-1} region, and is associated with the asymmetric stretching of the T−O−T bridges. However, as recently pointed out by Kamitsos et al.,[109] this mode is split into two components even in the case of vitreous silica, due to either the in-phase ($\nu_{1'}$) or the out-of-phase coupling ($\nu_{1''}$) of the asymmetric stretching modes of the nearest T−O−T groups. In other words, this mode couples with the symmetric and asymmetric stretching of the four Si−O bonds of the SiO_4 octahedra. As shown in Table 5, it seems clear that the position of both $\nu_{1'}$ and $\nu_{1''}$ components is also somewhat

Table 5 Densities (d, g cm^{-3}) and positions of the most intense (underlined) vibrational peaks (cm^{-1}) in the IR and Raman (R) spectra of different forms of silica

Mode	Silicalite ($d = 1.76$)		α-Tridymite ($d = 2.28$)		α-Cristobalite ($d = 2.33$)		α-Quartz ($d = 2.65$)		Coesite ($d = 2.91$)		Amorphous SiO_2 –	
	IR	R	IR	R	IR	R	IR	R	IR	R	IR	R
$\nu_{1''}$	1232	1230	1219		1197	1193		1235	1220		1208	1200
	1170		1145		1160		1166	1162	1160	1164		
										1144		
$\nu_{1'}$	1104	1095	1108		1096	1076	1086	1085	1085	1065	1100	1060
			1073					1069	1040	1036		
SiOH	960	975									950	970
		835						808	814	815		
	810	820	820					808	814	815		
ν_2	807	803	795		798	796	798	796	795	785	805	800
	790	790				785	778	778				
							694	696	683			
	600								600			608[a]
	562		585		624				560			
	552		541		525		511	511			560	
	545											490[a]
	495	473						495	490		466	
ν_3	447	450	485		493		461	(464)	435	425	465	
	420	435						450				
	375											
	350											
				457								
				449								
				422		421		465		521		430
ν_4	380		403				401		466			
		365			385		397	394	390	427		
			338				373	365	340	355		
			309		298			355		326		
			279				263	265		269		
						230		206		204		
Lattice										176		
										151		
					146	110		128		116		

[a] Defect bands.

sensitive to the size of the rings and/or to the density of the different silica polymorphs. This is particularly true for $\nu_{1''}$, whose shift is stronger, so that it is only sometimes resolved from $\nu_{1'}$.

The ν_1 modes are detectable in nearly the same position as in the Raman spectra, but are very weak there because of the small change in polarizability of T–O–T bonds during the asymmetric stretching if the two T atoms are equal.

As discussed above, the ν_2 mode is essentially a bending mode of the T–O–T bridges, although mixed with a symmetric stretching mode, as assumed by Flanigen et al.[110] Its assignment predominantly to a bending mode is mainly due to its weakness in the Raman spectra, although its position is more typical of a symmetric stretching mode. This mode appears to be the least sensitive to the structure, although in some cases it splits definitely into more components, as in the case of α-quartz where three well-resolved sharp bands are found.

The lowest-frequency IR mode ν_3 is associated with the out-of-plane deformation of the T–O–T bridges, i.e. a rocking mode. This explains why this mode is essentially silent in the Raman spectrum, while being very strong in the IR region 430–490 cm^{-1} in all silica forms, as well as in zeolites.

In addition to the four vibrational modes discussed here that are common to all silicas, network silicates, and aluminosilicates, other structure-sensitive features are

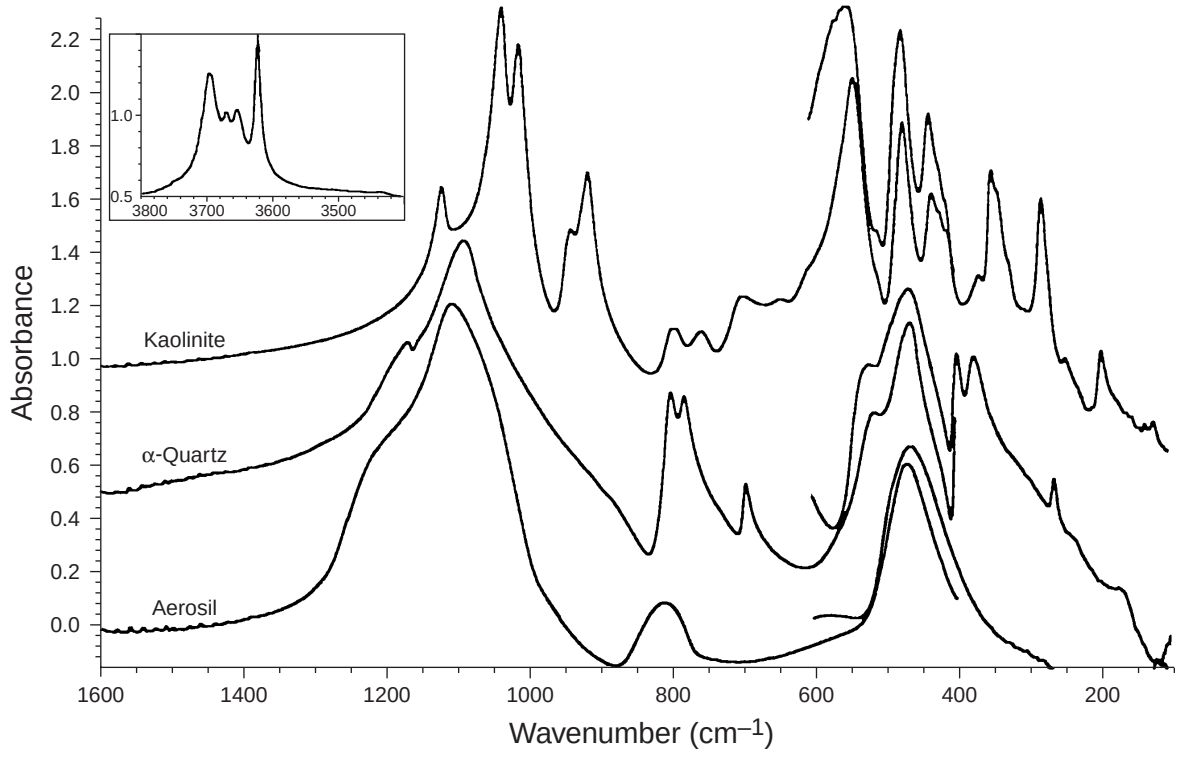

Figure 8 FTIR and FT/FIR transmission spectra of amorphous silica (aerosil), α-quartz and kaolinite powders (KBr and polyethylene disks).

frequently present. They are associated with the splitting of these vibrational modes (bands above 350 cm^{-1}) or to torsional lattice modes (below 400 cm^{-1}).

Figure 8 compares the FTIR and FT/FIR transmission–absorption spectra of an amorphous silica sample (aerosil) and of a crystalline powder (α-quartz). The ν_1, ν_2, and ν_3 absorptions (broad) are evident for amorphous silica. In the case of quartz, more, sharper components are observable, as expected: 10 maxima can be easily observed at 1165, 1086, 798, 779, 694, 519, 463, 397, 373, and 263 cm^{-1}. The irreducible representation for α-quartz (taking the right-handed enantiomorph with space group $P3_221 = D_3^6$, with $Z = 3$) is $4A_1(R) + 4A_2(IR) + 8E$ (R,IR) so that 12 IR-active fundamental modes are expected, each one undergoing TO/LO splitting. In the ν_1 region (asymmetric Si–O–Si stretching) three fundamental modes are expected, so that in the maximum at 1086 cm^{-1}, which shows shoulders on both sides, more components are superimposed. Three resolved components are found as expected in the ν_2 region. One fundamental mode is expected to be very weak near 130 cm^{-1},[111] but could not be detected.

The difference between the network silicate structure of the silica and the layer silicate structure of kaolinite is mainly shown by the presence, in the case of kaolinite, of a strong band with maxima at 940 and 913 cm^{-1}, typically due to Si–(OH) terminal stretchings. In this region silicas and network silicates do not absorb.

Another form of silica, called silicalite, belongs to the MFI IUPAC zeolite code, which belongs to the $P2_1/n = C_{2h}^5$ monoclinic space group (number 14) with $Z = 96$ (1,18). It transforms into an orthorhombic structure, belonging to the $Pnma = D_{2h}^{16}$ space group (number 62) between 350 and 363K. Substituted silicalites like ZSM5 zeolite (see below) and Ti-silicalite take the orthorhombic structure even at room temperature, the transition temperature being strongly shifted to lower temperatures.

The factor group analysis for monoclinic silicalite[112] gives the following irreducible representation for the optical modes:

$$\Gamma_{opt} = 216\ A_g\ (R) + 216\ B_g\ (R) + 215\ A_u\ (IR) + 214\ B_u\ (IR)$$

For the orthorhombic structure the following irreducible representation is obtained:

$$\Gamma_{opt} = 110\ A_g\ (R) + 106\ B_{1g}\ (R) + 110\ B_{2g}\ (R) + 106\ B_{3g}\ (R) + 106\ A_u\ (\text{inactive}) + 109\ B_{1u}\ (IR) + 105\ B_{2u}\ (IR) + 109\ B_{3u}\ (IR)$$

Accordingly, 429 IR-active modes and 432 Raman-active modes are expected for the monoclinic structure and 323 IR-active modes and 432-Raman active modes are expected for the orthorhombic form. The maxima really observable for silicalite are reported, together with those of other silica polymorphs, in Table 5. We observed only 16 components in the IR spectrum and 11 components in the Raman spectrum, because of the superimposition of many of the expected fundamentals. However, careful observation of the spectra, with the help of analysis of the perturbations arising from isomorphic substitution and with the aid of derivative spectra, showed the presence of a great number of very weak components (shoulders) in the spectra.

6.2 Complex Oxides

There are many ways to combine oxide components, producing a so-called mixed oxide. They depend not only on the preparation procedure adopted, but also on the nature of the component oxides. We will summarize them here below.

6.2.1 Mixing of Ionic Oxides

Several pairs of ionic oxides with the same crystal structure present extended or total reciprocal solubility. In other cases only limited solubility occurs, mainly due to their too-different cationic size. Limited reciprocal solubilities are also found for oxides having different structures but the same stoichiometry. Very limited solubilities are frequently found for ionic oxides having different stoichiometries, and these form mainly metastable phases. In any case, the resulting solid solutions retain the ionic character of the metal–oxygen bond. The characteristics of the vibrational behavior of such solid solutions are similar to those of the component oxides, although small band shifts are generally found between the positions typical for the spectra of the two component oxides.

Alternatively, oxides with different stoichiometry can mix, giving rise to ternary phases that are characterized by structures very different from those of the pure compounds. If the two cations are both sufficiently small, mixed oxide phases with oxide close packing structures (either hcp or ccp) are formed, with cations in tetrahedral or octahedral interstices. Spinels (formed by a bivalent and a trivalent metal in a 2 : 1 atomic ratio) and ilmenites (a bivalent and a tetravalent element in a 1 : 1 atomic ratio) are examples. Alternatively, one of the two cations can enter into oxide close packing, but the other can be too large, needing higher coordination numbers than six. This generates other structures such as the β-aluminas (a monovalent or a bivalent element with a trivalent element in ratios 1 : 17–1 : 19) and perovskites (two trivalent elements or one bivalent and one tetravalent in 1 : 1 atomic ratios).

From the point of view of vibrational spectroscopies, the treatment of such complex oxides does not differ significantly from those of the pure ionic oxides. Data on ternary oxide phases with the ABO_3 stoichiometry are given in Table 6.

6.2.2 Mixing of Covalent Oxides

Covalent oxides can also give rise to more or less extended solid solutions (such as those of MoO_3 and WO_3) and compounds, which can be treated similarly to the pure covalent oxide structures; sometimes they are treated as the salts (see below).

Table 6 Crystal structures of some solid ternary-oxide structures with the stoichiometry ABO_3, and the irreducible representations of the optical modes

Structure type	International symbols	Schönflies notations	Z	Irreducible representations[a]
Ilmenite	$R\bar{3}$	C_{3i}^2	6	$5A_g\,(R) + 5E_g(R) + 4A_u(IR) + 4E_u(IR)$
$LiNbO_3$	$R3c$	C_{3v}^6	6	$4A_1(IR) + 5A_2(R) + 9E\,(IR,R)$
Cubic perovskite	$Pm3m$	O_h^1	1	$3F_{1u}(IR) + F_{2u}(in)$
Tetragonal perovskite	$P4mm$	C_{4v}^1	1	$3A_1(IR,R) + 4E(IR,R) + B_1(R)$
Orthorhombic perovskite	$Pnma$	D_{2h}^{16}	4	$7A_g(R) + 7B_{1g}(R) + 5B_{2g}(R) + 5B_{3g}(R) + 8A_u(in) + 7B_{1u}(IR) + 9B_{2u}(IR) + 9B_{3u}(IR)$
Rhombohedral perovskite	$R\bar{3}c$	D_{3d}^6	2	$A_{1g}(R) + 3A_{2g}(in) + 4E_g(R) + 2A_{1u}(in) + 2A_{2u}(IR) + 4E_u(IR)$
Calcite	$R\bar{3}c$	D_{3d}^6	2	$A_{1g}(R) + 3A_{2g}(in) + 4E_g(R) + 2A_{1u}(in) + 2A_{2u}(IR) + 4E_u(IR)$
Aragonite	$Pmcn$	D_{2h}^{16}	4	$9A_g(R) + 6B_{1g}(R) + 6B_{2g}(R) + 9B_{3g}(R) + 6A_u(in) + 8B_{1u}(IR) + 8B_{2u}(IR) + 5B_{3u}(IR)$

[a] R = Raman; in = in active.

6.2.3 Mixing of Covalent and Ionic Oxides: the Oxo Salts

Ionic and covalent oxides mostly give rise to ternary phases, known as oxo salts. Examples are metal silicates, phosphates, molybdates, tungstates, chromates, and vanadates. Their structures have little in common with those of the pure oxides. Frequently, different salts can be produced starting from the same oxide components, in relation to the different compositions, as for *ortho-*, *pyro-*, and *meta*-vanadates, silicates, and phosphates.

The oxo salts contain covalent structures (the oxo anions) ionically bonded to cations. This allows easy distinction of some vibrational modes involving the nonmetal atom (or the higher oxidation state metal) from those involving the metal atom.

6.3 Vibrational Features of Complex Oxides of Practical Interest

6.3.1 Cuprate Superconductors

Vibrational spectroscopies have largely been used to characterize these superconductor ceramics.[113] However, most studies involve Raman spectroscopy and IR reflection measurements of monocrystals. Additionally, the interpretation of such measurements is mostly performed using lattice dynamics calculations. This approach, although very complete in principle, does not always give rise to a good fit between calculated and theoretical frequencies and does not allow easy comparison between the vibrational features of different structures. The interpretation of IR spectra of superconducting powders based on the group approximation has also been attempted.[114,115] From these studies, some conclusions can be drawn allowing the comparison of the spectra of different superconducting cuprates:

- The highest frequency IR and Raman modes are due to stretchings of the basal Cu–O bonds.
- The position of the higher-frequency Cu–O asymmetric stretching mode easily correlates to the length of the basal Cu–O bonds.
- Loss of resolution for the individual components is typical for samples approaching the composition that allows superconductivity to occur at low temperature.[114,116]
- Modes parallel to the Cu–O square planes are screened out when the structures become able to undergo the superconducting transition due to doping effects, such as for YBCO,[117] $Nd_{2-x}Ce_xCuO_4$,[114,118] $La_{2-x}Ba_xCuO_4$, and $La_{2-x}Sr_xCuO_4$.[115]

6.3.2 Ferrite Magnetic Materials

Many ferrimagnetic materials, used as magnetic pigments or as permanent magnets, belong to two classes of ferrites. The soft ferrites are based on inverted spinel structures (Fe_3O_4, $MnFe_2O_4$, mixed $(Mn,Zn)Fe_2O_4$, $MgFe_2O_4$, $CoFe_2O_4$). The cubic spinel structure, irrespective of the cation distribution of normal, inverted or random spinels, belongs to the $O_h^7 = Fd3m$ space group with $Z = 8$. Five Raman-active modes and four IR-active modes occur; both IR[119–121] and Raman spectroscopies[122,123] have been used to characterize such structures. Also the magnetic ferric oxide maghemite (γ-Fe_2O_3) can be assumed to belong to the same family although its distorted defective spinel-like structure is far more complex, and possibly still not defined with certainty. The IR[124] and Raman[123] spectra of γ-Fe_2O_3 are also definitely more complex than that of spinels, showing that the maghemite structure is actually a superstructure. The other class of ferrite materials is the hard ferrites. The best known example is barium ferrite, $BaFe_{12}O_{19}$, which takes the magnetoplumbite structure (space group $P63/mmc = D_{6h}^4$, with $Z = 2$) of $PbFe_{12}O_{19}$. IR studies have been devoted to barium ferrite magnetic materials, giving further information on the structure and the likely dynamic disorder involving one ferric atom between two sites, one located in a mirror plane and the other out of it.[124,125]

6.3.3 β-Alumina-type Ionic Conductors

One of the most relevant families of ionic conductor solids are the so-called β-aluminas. The parent compound is sodium β-alumina whose composition (when stoichiometric) is $NaAl_{11}O_{17}$. Spectroscopic investigations have been devoted to this compound[126,127] as well as to other compounds of the same family.[124]

These phases are characterized by the existence of layered blocks of a trivalent element (most frequently Al^{3+}, but also Fe^{3+} and Ge^{3+}) with the same spinel structure separated by mirror planes. The big monovalent (Na or K), bivalent (Ba and sometimes Ca or Sr) and trivalent (La) cations, which do not enter the cubic close packed array of oxide ions of the spinel blocks, are located together with oxide anions bridging between the blocks. In these planes the cations are able to diffuse. The metal–oxygen stretchings involving such ions located in the mirror planes fall in the low-frequency (FIR) region. For this reason FIR and Raman spectroscopy have been devoted to studies of the coordination and mobility of such cations.[124,128,129]

6.4 Metal Hydroxides

The solid crystalline hydroxides are generally formed from condensed polyhedra containing the metal element surrounded by oxygen atoms. The hydrogen atoms are located externally to the polymeric structures and are involved in hydrogen bonding. Thus the skeletal spectra of hydroxides show features typically arising from the OH

groups, and features to be assigned to the vibrations of MO_x polyhedra. For the latter features, the considerations already made for oxides are valid.

The vibrations of hydroxy groups are typically composed of O–H stretchings (3800–2000 cm^{-1}, depending on the extent of hydrogen bonding), in-plane bending (1200–800 cm^{-1}), and out-of-plane deformation (generally below 1000 cm^{-1}). The multiplicity of such modes depends upon the number of hydroxy groups present in the smallest Bravais cell and in the coupling of their vibrations, and is also dependent on the hydrogen-bonding patterns.

The vibrational spectra of alkali metal hydroxides have been reviewed by Ryskin and Farmer.[130] The centrosymmetric structures of LiOH and NaOH, both containing two formula units per unit cell, give rise to two OH stretching modes, one of which is IR active and the other Raman active. In the case of KOH, whose structure is not centrosymmetric and again contains two formula units per unit cell, a split both of IR- and Raman-active peaks is detectable. The positions of these peaks (in the 3680–3600 cm^{-1} region) show that no hydrogen bonding occurs in these structures.

The structure of brucite $Mg(OH)_2$ is assumed by most bivalent hydroxides. Every layer of $Mg(OH)_2$ contains MgO_6 octahedra with three triply-bridging hydroxy groups pointing up and three pointing down, alternately. The unit cell of the brucite structure contains only one $Mg(OH)_2$ unit, so 12 optical modes should exist. Two OH stretching modes exist, one Raman active (symmetric stretching, A_{1g}, 3655 cm^{-1}) and one IR active (asymmetric OH stretching, A_{2u}, 3700 cm^{-1}). The positions of these modes show that no hydrogen bonds occur in this structure. Additional two doubly-degenerate deformation modes occur, again one Raman active (E_g) and one IR active (E_u). Finally, two IR-active modes ($A_{2u} + E_u$) and two Raman-active modes ($A_{1g} + E_g$) are associated with vibrations of the MgO_6 octahedra. Lutz et al. compared the vibrational structures (IR, Raman, and INS) of several brucite-type hydroxides,[131] as well as oxyhalides of bivalent metals in the fundamental[132] and overtone regions.[133]

The goethite or diaspore structures of α-AlOOH, α-FeOOH, α-CrOOH and α-GaOOH have orthorhombic unit cells, belonging to $Pbnm = D_{2h}^{16} = 62$ space group with $Z = 4$. The overall unit cell and the smallest Bravais cell coincide and contain 16 atoms; thus 45 vibrational modes are expected. All atoms occupy the 4*c* Wyckoff position in the cell on a reflection plane (C_s site symmetry). According to literature data,[134,135] and separating the modes due to the hydrogen-bonded OH groups from those of the Me–O skeleton, the irreducible representation for the IR optical modes of α-MeOOH is as follows:

$$\Gamma_{Me-O} = 5B_{1u}\,(IR) + 2B_{2u}\,(IR) + 5B_{3u}\,(IR)$$

$$\Gamma_{O-H} = 2B_{1u}\,(IR) + 1B_{2u}\,(IR) + 2B_{3u}\,(IR)$$

Consequently, the IR spectrum is expected to contain 17 fundamental modes. Clearly, not all will be resolved. A typical feature of the spectra of these compounds is the presence of two IR-active OH stretchings, giving rise to a strong band with a shoulder at the higher frequency in the IR spectra (3126 and 3050 cm^{-1} for goethite, 3000 and 2920 cm^{-1} for diaspore, and 2913 and 2840 cm^{-1} for α-GaOOH) as well as two strong bending modes (891 and 796 cm^{-1} for goethite, 1080 and 970 cm^{-1} for diaspore, and 1019 and 952 cm^{-1} for α-GaOOH), which have been discussed in terms of the geometry of the hydrogen-bonding situations. The materials α-(Cr,Fe)OOH[136] and α-(Fe,Ga)OOH,[137] forming goethite-type solid solutions, have also been characterized by IR. The INS spectra of goethite have also been reported.[77]

X-ray diffraction (XRD) analysis indicates that boehmite γ-AlOOH and lepidocrocite γ-FeOOH crystallize in the space group $Cmcm = D_{2h}^{17}$, i.e. space group 63 with $Z = 4$. However, XRD analysis does not reveal the position of the hydrogen atoms. To obtain information on the position of the hydrogen atoms, vibrational spectroscopies can be used. The IR spectra are dominated by two well-split OH stretchings at 3295 and 3090 cm^{-1} for boehmite and at 3390 and 3125 cm^{-1} for lepidocrocite. Some authors have concluded that the IR spectra of boehmite, in particular the modes associated with the vibrations of the hydroxy groups, are consistent with the space group *Cmcm*,[134,138] whereas others have concluded that a lower symmetry space group applies, namely $Cmc2_1 = C_{2v}^{12}$ of space group 36.[139] Publication of the Raman spectrum by Kiss et al.[140] confirmed that the mutual exclusion rule applies, so the noncentrosymmetric space group $Cmc2_1$ must be ruled out, the true space group for boehmite being *Cmcm*.

More recently, boehmite spectra have been compared with those of samples containing small amounts of silica.[141] Three different behaviors for the IR bands of boehmite upon Si addition can be noticed: (a) some bands are essentially not perturbed or are only slightly shifted by silicon addition; (b) other bands progressively decrease in intensity; and (c) others are eliminated even with a very small amount of silica addition. In particular, the bands lost through silicon addition are those at 325, 398, 412, 1161, 3088 and 3305 cm^{-1}, i.e. bands associated with the vibrations of the Al–OH groups. The bands near 1070 and 735 cm^{-1} decrease progressively in intensity by Si addition. However, the three bands at 608, 483 and

Table 7 Assignment of the observed IR and Raman peaks of boehmite and changes caused by Si addition

IR bands (cm^{-1})	Assignments	Behavior on Si addition
3302	ν O–H	Removed
3088	ν O–H	Removed
2097	Combination	
1972	Combination	
1161	δ O–H	Removed
1072	δ O–H	$\rightarrow$1066, slowly decreasing in intensity
sh 870	Combination	Superimposed by ν Si–O
sh 775	Combination	Superimposed by ν Si–O
735	γ O–H	Slowly decreasing in intensity
608	ν Al–O	$\rightarrow$619
483	ν Al–O	$\rightarrow$470, $\rightarrow$478
412	ν Al–OH	Removed
398	ν Al–OH	Removed
368	ν Al–O	$\rightarrow$357
325	ν Al–OH	Removed

sh = shoulder; $\rightarrow$ = shift to.

368 cm^{-1} shift only slightly, but retain their strengths and shapes.

This behavior has allowed slight modification of the assignments of the low-frequency Al–O stretching bands, as shown in Table 7, with respect to those reported in the literature. Thus, as described above, six IR-active modes are expected for the Al–O stretchings of boehmite. Three of them, which are not perturbed by Si addition, can be assigned to Al–O stretchings (the bands at 608, 483, and 368 cm^{-1} in pure boehmite). The other three, which are eliminated by silicon addition, could be due to Al–(OH) stretching modes (at 412, 398, and 325 cm^{-1}) in pure boehmite.

The addition of Si causes new bands to occur at 3450, 880, 800, and near 405 cm^{-1}. The bands at 880, 800 and near 405 cm^{-1} are probably associated with Si–O modes. They can be associated with orthosilicate species $(SiO_4)^{4-}$, the first two bands being due to the symmetric and asymmetric stretchings and the last to a deformation mode (see below). The evident perturbation of the modes associated with the Al–OH groups suggests that most silicate species go to the interlayer region between the boehmite layers.

6.5 Hydrated Compounds

Gas-phase or isolated water molecules belong to the C_{2v} point group and give rise to three optical modes, $2A_1 + B_1$, all of which are IR and Raman active. The two stretchings are at 3756 cm^{-1} (B_1, asymmetric stretching) and at 3657 cm^{-1} (A_1, symmetric stretching), whereas the bending mode (A_1) is at 1594 cm^{-1}. When water molecules are incorporated into a crystal, their six external modes (arising from frustrated rotations and translations) contribute to lattice modes called librations of water molecules (wagging, rocking and twisting modes) and to the acoustic modes. When water is coordinated through oxygen lone pairs to cations (aquo complexes) the metal–oxygen stretching mode and the corresponding deformation modes appear and couple with the other metal–ligand vibrations. The metal–oxygen stretchings are usually located below 600 cm^{-1}. At slightly higher frequencies (900–500 cm^{-1}) the wagging and twisting modes of coordinated water molecules become internal vibrations of the aquo complex. Coordination per se causes relatively weak perturbations of the stretching and bending modes of water. However, coordination is frequently associated with additional hydrogen-bonding of water, with other water molecules, or with anions present in the crystal structure. This can result in strong downward shifts (to 2000 cm^{-1}) and broadening of the stretching absorptions, whereas the bending modes are not very sensitive and always fall in the region 1700–1550 cm^{-1}. Asymmetric hydrogen bonding (bonding to only one of the two protons of water) gives rise to spectra where sharp high-frequency peaks occur together with broad low-frequency OH stretching modes.[142]

Spectra of molecular water molecules in crystals and complexes have been collected by Ryskin[130] and by Nakamoto.[8] Protonated forms of water such as the oxonium ion H_3O^+[143] have also been investigated.

6.6 Oxo Salts

The oxo salts involve oxo anions, which are internally held by covalent bonds between the nonmetal atom (or the metal in a very high oxidation state) and oxygen, giving rise to M–O bond orders ranging between 1 and 2. The oxo anions are further bonded through ionic interactions with the metal cations. The vibrational spectra of oxo salts are therefore dominated, in their high-frequency ranges, by the internal vibrations of the oxo anion. The following sections briefly review the vibrational structure of such anions, in relation to the point group symmetry they can take. Obviously, the splittings of these internal vibrations occur in principle as a result of couplings between the modes of more than one oxo-anionic unit simultaneously present in the smallest Bravais cell.

However, when the unit cell is centrosymmetric the mutual exclusion rule is valid, so that Raman-active modes are IR inactive and vice versa. In practice, for centrosymmetric cells containing N oxo anions, every internal vibrational mode of the oxo anion gives rise to $N/2$ IR-active modes and $N/2$ Raman-active modes.

The cation–anion vibrations (lattice vibrations) are mainly located in the FIR region and their assignments,

based on similar considerations to those reported for ionic oxides, are frequently difficult.

6.6.1 *Salts with Trigonal-planar and Pyramidal MO_3^{n-} Oxo Anions*

The movements of the four atoms present in MO_3^{n-} oxo anions give rise to a total of 12 modes, six of which are associated with the movements of the anion as a whole (rotations and translations, which contribute to the lattice vibrations and to the acoustic modes of the solid). So, six ''internal'' vibrations occur. The highest symmetry point group for this ion is D_{3h}, occurring when it is planar trigonal and isolated. This is typical for isolated carbonate, nitrate and orthoborate anions. In this case, as shown in Table 8, the symmetric stretching gives rise to a nondegenerate Raman-active mode, whereas the asymmetric stretching is doubly degenerate and both IR and Raman active. Two deformation modes occur, the out-of-plane deformation (δ_{oop}) which is IR active, and in-plane deformation (δ_{ip}) which is degenerate and both IR and Raman active. By lowering the symmetry, e.g. to the C_{2v} point group (as occurs when only two of the three oxygen atoms are equivalent), the doubly-degenerate modes split and all six modes become both IR and Raman active.

The highest symmetry point group for nonplanar pyramidal MO_3^{n-} oxo anions (such as sulfite, selenite, and tellurite anions) is C_{3v}. In this case as well, all modes are both IR and Raman active. The symmetric stretching and deformations are nondegenerate whereas the asymmetric modes are doubly degenerate. Such a degeneracy is broken when symmetry is further lowered. Typical positions for the corresponding modes are reported in Table 9.

The commonest crystal structures for AMO_3 salts that have been the subject of vibrational studies are calcite ($CaCO_3$, $AlBO_3$, $FeBO_3$, $LuBO_3$, $NaNO_3$),[144] dolomite ($CaMg(CO_3)_2$), and aragonite ($CaCO_3$, $LaBO_3$[145]), including aragonite–strontianite solid solutions[146] and nitrates of bivalent metals.[147]

Table 8 Normal modes of vibrations and their activity for isolated MO_3^{n-} oxo anions

Planar	ν_{sym}	δ_{oop}	ν_{asym}	δ_{ip}
D_{3h}	A_1'(R)	A_2''(IR)	E′(IR, R)	E′(IR, R)
C_{2v}	A_1(IR,R)	B_2(IR,R)	A_1(IR,R)	A_1(IR,R)
			B_1(IR,R)	B_1(IR,R)
Pyramidal	ν_{sym}	δ_{sym}	ν_{asym}	δ_{asym}
C_{3v}	A_1(IR,R)	A_1(IR,R)	E(IR,R)	E(IR,R)
C_s	A′(IR,R)	A′(IR,R)	A′(IR,R)	A′(IR,R)
			A″(IR,R)	A″(IR,R)

R = Raman.

Table 9 Typical band positions (cm^{-1}) for MO_3^{n-} oxo anions

Planar	ν_{sym}	δ_{oop}	ν_{asym}	δ_{ip}
BO_3^{3-}	1100–900	750–650	1500–1250	600–550
CO_3^{2-}	1100–1050	900–850	1480–1380	750–670
NO_3^-	1080–1030	840–800	1490–1320	750–690
Pyramidal	ν_{sym}	δ_{sym}	ν_{asym}	δ_{asym}
SO_3^{2-}	1000–960	670–610	1000–910	520–450
SeO_3^{2-}	850–780	500–400	750–700	400–350
TeO_3^{2-}	800–750	430–330	730–680	350–300

Table 10 Normal modes of vibration and their activities for tetrahedral oxo anions, and typical vibrational frequencies

Environment	ν_{sym}	δ_{sym}	ν_{asym}	δ_{asym}
Normal vibrational modes and their activities				
T_d	A_1(R)	E (R)	F_2(IR,R)	F_2(IR,R)
C_{3v}	A_1(IR,R)	E(IR,R)	A_1(IR,R)	A_1(IR,R)
			E(IR,R)	E(IR,R)
C_{2v}	A_1(IR,R)	A_1(IR,R)	A_1(IR,R)	A_1(IR,R)
		A_2(R)	B_1(IR,R)	B_1(IR,R)
			B_2(IR,R)	B_2(IR,R)
C_1	A(IR,R)	2A(IR,R)	3A(IR,R)	3A(IR,R)
Typical band positions (cm^{-1})				
SiO_4^{4-}	850–800	450–300	1000–850	650–450
GeO_4^{4-}	800–600	350–250	850–650	550–400
PO_4^{3-}	1000–920	420–350	1080–950	600–530
AsO_4^{3-}	900–830	400–320	880–750	500–410
SbO_4^{3-}	800–500	400–300	800–500	400–300
VO_4^{3-}	915–800	500–350	900–730	500–350
NbO_4^{3-}	800–500	400–300	800–500	400–300
SO_4^{2-}	1070–950	520–410	1270–1030	670–570
SeO_4^{2-}	860–810	370–300	940–840	460–390
CrO_4^{2-}	900–830	360–330	960–860	410–330
MoO_4^{2-}	940–860	360–310	910–740	(360–310)
WO_4^{2-}	960–900	355–320	870–830	(355–320)
MnO_4^-	850–820	420–380	940–880	(420–380)
ReO_4^-	1000–950	350–300	950–900	350–300

R = Raman.

6.6.2 *Salts with Tetrahedral MO_4^{n-} Oxo Anions*

Most of the MO_4^{n-} oxo anions take a tetrahedral-like coordination. In this case there are nine internal vibrations plus six external vibrations contributing to lattice vibrations and to acoustic modes. When the ion takes its highest symmetry (T_d point group) the symmetric stretching and the symmetric deformation are Raman active, the last being doubly degenerate. The asymmetric stretching and deformation are both IR and Raman active and are triply degenerate. By lowering of the symmetry, the degeneracies are broken and all modes can become IR

and Raman active. This is shown in Table 10, where the typical positions of such modes are also reported for different oxo anions.

Among the different structures of salts containing tetrahedral-like oxo anions AMO_4, scheelite structures (molybdates of Ca, Sr, Ba, Pb, and Cd, tungstates of Ca, Sr, Ba, and Pb) have been the subject of many studies.[148–153] The structure of α-$MnMoO_4$ (isostructural with β-$MgMoO_4$ and β-$CoMoO_4$)[154] and the structures taken by divalent chromates[155] and sulfates[156] have also been investigated. As for compounds with the formula A_2MoO_4, studies are reported for spinel structures (Na and Ag molybdate, Na tungstate, Ni silicate, and several germanates),[157,158] olivine and monticellite structures such as Mg_2SiO_4 and some germanates,[159,160] beryllium silicate structures (Be silicate, Li molybdate and tungstate[149]), and isomorphous alkali chromates and sulfates.[161,162] Among compounds with the formula $A_3(MO_4)_2$ studies have been reported on Mg, Ca, Sr and Ba vanadates[163,164] as well as Ba and Sr orthophosphates, arsenates, chromate and manganate(VI).[165] The IR and Raman spectra of the perrhenate compound Hg_2ReO_5[166] and of Li_4SiO_4 and Li_4GeO_4[167] have been reported.

The structure of $Mg_3(VO_4)_2$ and the coordination of the vanadate ion in this structure are shown in Figure 9. This compound crystallizes in the $Cmca = D_{2h}^{18}$ space group with $Z = 4$. The crystallographic unit cell contains two Bravais cells. Figure 10 shows the IR and Raman spectra. In Table 11 the distribution and assignment of the optical modes are summarized, whereas Table 12 offers a tentative detailed assignment for the V–O stretching mode, showing the effect of decreasing of the symmetry of the anion in the crystal and of coupling of the four units.

The tetrahedral oxo anions can retain one or two protons in the crystal structure, giving rise to monohydrogen or dihydrogen ortho anions. In this case, one or two of the oxygens are certainly unequivalent to the others, and the vibrations of the O–H group also appear. Several studies have appeared concerning metal monohydrogen orthophosphates and dihydrogen orthophosphates.[168–170]

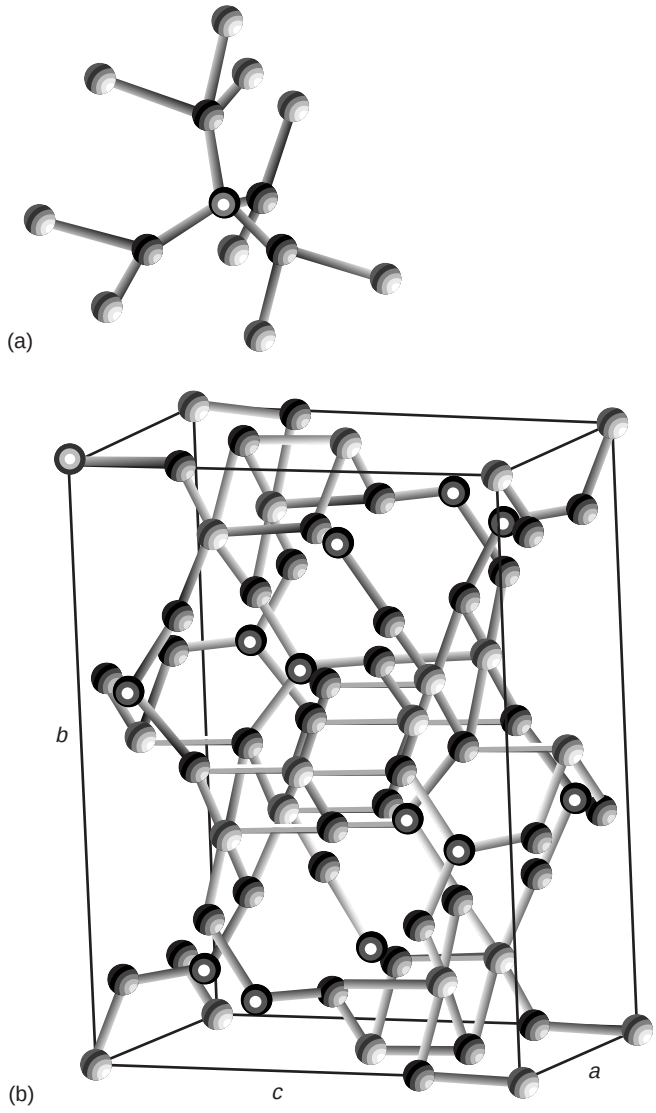

Figure 9 Structures of (a) the anion of $Mg_3(VO_4)_2$ and (b) the unit cell. (Reproduced with permission from G. Busca et al., *J. Chem. Soc. Faraday Trans.*, **90**, 1161–1170 (1994).)

6.6.3 Salts with Condensed Tetrahedral Oxo Anions

6.6.3.1 Dimeric Condensed Tetrahedral Anions $[X_2O_7]^{n-}$
Salts where the tetrahedral oxo anions are condensed through common oxygen atoms are denoted by the prefix "pyro" (for dimeric species) or "meta" (for polymeric species). The oxo anions involved have terminal M–O bonds (six in total for dimeric pyro species, two of each atom for polymeric meta anions), and additionally contain M–O–M bridges. Such bridges are generally bent, but they can sometimes be linear, such as for some pyrosilicates, pyrogermanates, pyrophosphates, and pyroarsenates. When the bridge is bent it gives rise to an asymmetric stretching, bending/symmetric stretching modes (both are in-plane modes) and an out-of-plane rocking mode (like the Si–O–Si bridges of silica, discussed above). For silicates, the ν_{as} Si–O–Si and ν_{sym} Si–O–SiO occur in the regions 1050–900 cm^{-1} and near 650 cm^{-1}, respectively.[171,172] According to Lazarev[173] a linear relationship exists between the bond angle and the parameter $\Delta = (\nu_{as} - \nu_{sym})/(\nu_{as} + \nu_{sym})100$. When the bridge is linear (as in thortveitite, $Sc_2Si_2O_7$), the symmetric stretching could become IR inactive (or very weak if the environment is symmetric). The difference $\nu_{as} - \nu_{sym}$ is apparently a little higher, with the

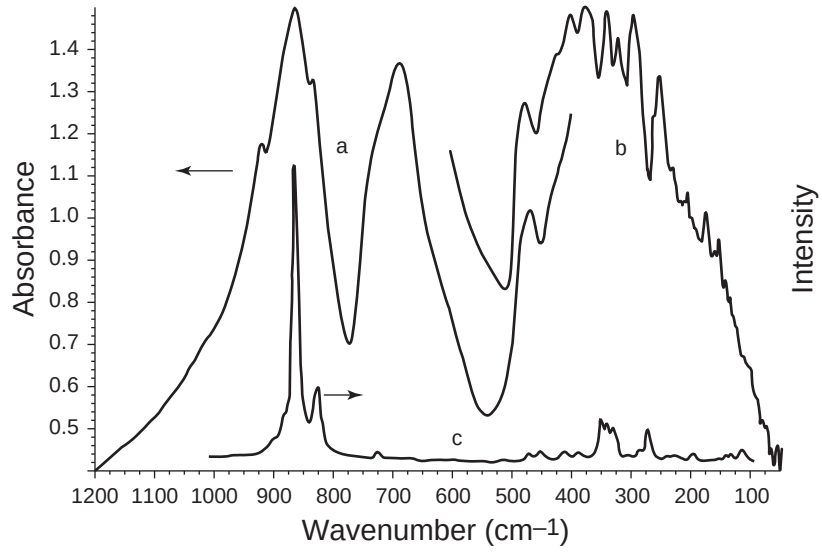

Figure 10 (a) FTIR, (b) FT/FIR, and (c) laser Raman spectra of $Mg_3(VO_4)_2$ powder. (Reproduced with permission from G. Busca et al., *J. Chem. Soc. Faraday Trans.*, **90**, 1161–1170 (1994).)

Table 11 Distribution and assignments of the fundamental modes of orthorhombic $Mg_3(VO_4)_2$

Symmetry	Activity[a]	Total	Acoustic	Optical	Lattice	VO_4		Origin			
						Librational	Internal	ν_1	ν_2	ν_3	ν_4
A_g	R	10	0	10	3	1	6	1	1	2	2
B_{1g}	R	8	0	8	3	2	3	0	1	1	1
B_{2g}	R	7	0	7	2	2	3	0	1	1	1
B_{3g}	R	11	0	11	4	1	6	1	1	2	2
A_u	–	8	0	8	3	2	3	0	1	1	1
B_{1u}	IR	13	1	12	5	1	6	1	1	2	2
B_{2u}	IR	12	1	11	4	1	6	1	1	2	2
B_{3u}	IR	9	1	8	3	2	3	0	1	1	1
Total		78	3	75	27	12	36	4	8	12	12

[a] R = Raman.

symmetric stretching mode shifted somewhat down to 660–630 cm^{-1}.[174,175]

Several studies have been devoted to compounds containing condensed tetrahedral anions $[X_2O_7]^{n-}$[176] such as pyrovanadates,[165] dichromates, pyrosulfates, pyroselenates, pyrophosphates, pyroarsenates,[177] and pyroniobates.[178] Typical of these compounds are the vibrations of the bridging oxygens. Additionally, the spectra show the vibrations of the terminal MO_3. Each terminal MO_3 moiety gives rise to one symmetric and two asymmetric stretchings, one symmetric and two asymmetric bendings, as well as to one twisting and two rocking modes.

6.6.3.2 Metasilicates, Layer Silicates, and Framework Silicates The vibrational spectra of the different kinds of chain and ring silicates are dominated by the vibrations of the Si–O–Si bridges (as previously discussed for amorphous silicas and for the pyrosilicates) and of Si–O– terminal bonds.[179] Layer silicates[180,181] also present absorptions for Si–O–Si bridges and terminal silanol groups, Si–OH. Both terminal silicate bonds $Si-O^- M^+$ and silanols give rise to strong IR bands in the region 950–1000 cm^{-1}, which frequently occur also in the Raman spectra. Cyclic structures and particular conformations of the chains can give rise to characteristic features, particularly in the lower frequency regions (Si–O–Si rockings and lattice modes). Terminal silanols, such as those present in the case of layer silicates, also give rise to OH stretching modes strong in IR, such as those of kaolinite (see Figure 8), and these have been characterized by Raman spectrometry.[182] Systematic treatments of the vibrational structures of silicates have been reported.[81,82,173]

Zeolites are microporous framework aluminosilicate materials of great interest in heterogeneous catalysis and

Table 12 Assignment of the fundamental stretching bands of orthorhombic $Mg_3(VO_4)_2$ (cm^{-1}) and their IR-active combinations

Mode T_d	Site symmetry factor group				Activity[a]
	C_{3v}	C_s	D_{2h}	Symmetry	
			915 + 881 1790	B_{1u} or B_{2u}	IR
			862 + 861 1720	B_{1u} or B_{2u}	IR
			861 + 827 1672	B_{1u} or B_{2u}	IR
			862 + 485 1347		IR
			915	B_{1u} or B_{2u}	IR
			897	A_g	R
	888 $A_1 \rightarrow A'$				
			881	B_{3g}	R
			862	A_g	R
			861	B_{2u} or B_{1u}	IR
ν_1 A_1 827	837 $A_1 \rightarrow A'$				
			833	B_{1u}, B_{2u}	IR
			827	B_{3g}	R
ν_3 F_2 780					
			740		IR
	700 $E \rightarrow A'' + A'$				
			724		R
			687		IR
			(690)		R

[a] R = Raman.

adsorption technology. Some are natural materials, others are synthetic. The electrostatic imbalance arising from the substitution of Al for Si in the silica-like framework is balanced by protons or extraframework cations lying in the cavities. The protonic forms are used as solid acids, whereas the cation-exchanged forms are very useful for selective adsorption. Figure 11 shows the IR and Raman spectra of a typical protonic zeolite, H-ZSM5. This is the Al-containing form of orthorhombic silicalite, discussed above. For this structure there are 323 IR-active modes and 432 Raman-active modes. These spectra show the most typical features of network silica, as discussed above, with few splittings because of the superimposition of many vibrational modes arising from the same vibrations of the basic Si–O–Si units. Figure 12 shows the IR spectrum of the zeolite NaX, a typical cation-containing zeolite whose composition is $Na_{12}Al_{12}Si_{12}O_{48}$. The structure was first assumed to be cubic with a disordered distribution of Si and Al (*Pm3m* space group), but it became clear that Si and Al alternate in an orderly fashion. This makes it necessary to multiply the unit cell by eight (*Fm3c* space group). An enormous number of fundamentals are expected in both IR and Raman spectra. The strong vibrations expected for the zeolite framework are those above 400 cm^{-1}, whereas the bands in the FIR are essentially due to the vibrations of the Na cations in the cavities.[183] IR spectroscopy has been used extensively for zeolite characterization,[110] and Raman spectroscopy is becoming increasing popular.[184]

6.6.3.3 Glasses Raman and IR spectroscopies are unique techniques for the investigation of the structure of glasses. Many studies have been reported concerning silicate-based glasses,[185] in their anhydrous or hydrated states.[186] Both IR and Raman spectra give information on the nature of Si–OH and $Si-O^- M^+$ terminal bonds with respect to Si–O–Si bridges and of any cyclic structures. The FIR spectra can give information on the coordination state of the cations. Many different compositions have been investigated, such as La aluminosilicates[187] and glasses belonging to the $Li_2O-ZrO_2-SiO_2$ system.[188] Comparisons between the IR spectra of amorphous and crystalline materials belonging to the $MgO-Al_2O_3-SiO_2$ and $Li_2O-Al_2O_3-SiO_2$ systems, including cordierite and β-eucriptite, have also been reported.[189]

6.6.4 Salts with Condensed Octahedral-like Oxo Anions

Chains of oxo anions can condense, giving rise to octahedral-like chains or multiple chains. This is the case for several metavanadates forming double chains $[(V_2O_6)^{2-}]_n$, such as Mg metavanadate. The structure of this compound is shown in Figure 13. In this case vibrational features appear due to triply-bridging O atoms and to pairs of short divanadyl VO_2 bonds.[164] Similarly orthotungstate ions, such as those of metal tungstates in the wolframite structure, condense giving rise to single chains of corner-sharing octahedra (Figure 14). Correspondingly, the spectra can be interpreted in terms

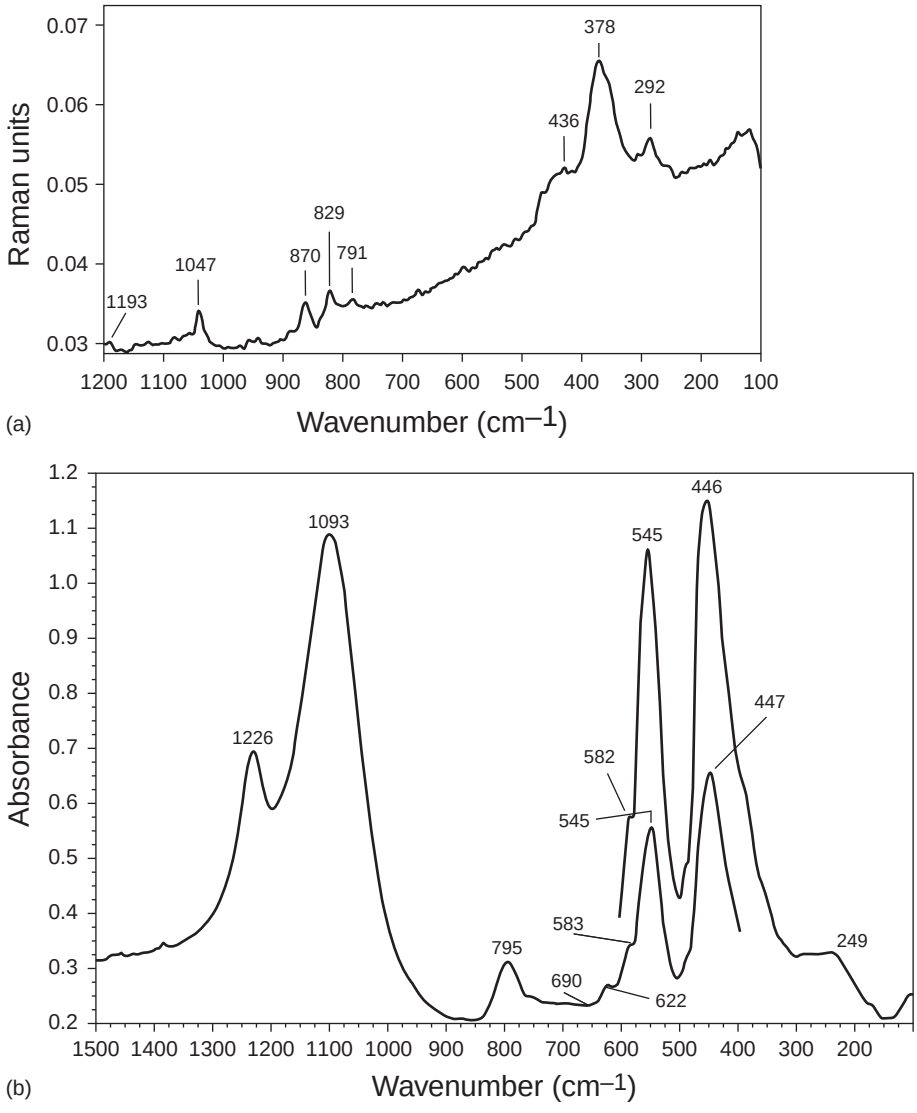

Figure 11 (a) FT/Raman and (b) FTIR and FT/FIR spectra of HZSM5 zeolite.

of bridging W−O−W oxygens and of pairs of diwolframyl WO_2 terminal groups.[153]

6.7 Other Inorganic Materials

6.7.1 Elementary Network Solids

Diamond, silicon, and germanium are elementary solids. They crystallize in the diamond cubic structure, space group O_h^7, $Fd3m$, with $Z = 2$. Consequently, only one triply-degenerate mode occurs which is Raman active. No IR-active fundamentals occur. However, overtones or impurity features can be detectable. Moreover, the Raman-active fundamental appears sometimes to gain some IR activity. For diamond, the fundamental Raman mode was determined as early as in 1930 to be at $1332.5\,cm^{-1}$,[190] but tends to shift down to $1320\,cm^{-1}$ in powders of small particle size.[191] Photoacoustic spectra of diamond allow the detection of adsorbed hydrogen and oxygen.[192] In the case of silicon monocrystals the Raman peak is at $520\,cm^{-1}$[193] and is strongly broadened for amorphous silicon powder. IR spectroscopy is applied to detect and quantify oxygen,[194] hydrogen,[195] and carbon[196] impurities in or on silicon, by measuring the intensities of the Si−O, Si−H, and Si−C IR-active stretching mode. Fundamental TO modes of Ge and cubic Sn fall near $300\,cm^{-1}$ and near $200\,cm^{-1}$, respectively. Raman spectroscopy has also enabled the characterization of tetragonal Ge nanoparticles.[197]

Pure elementary carbon can crystallize in the form of graphite. The vibrational structure has been determined by Fateley et al.[3] in terms of the space group C_{6v}^4, $P6_3mc$

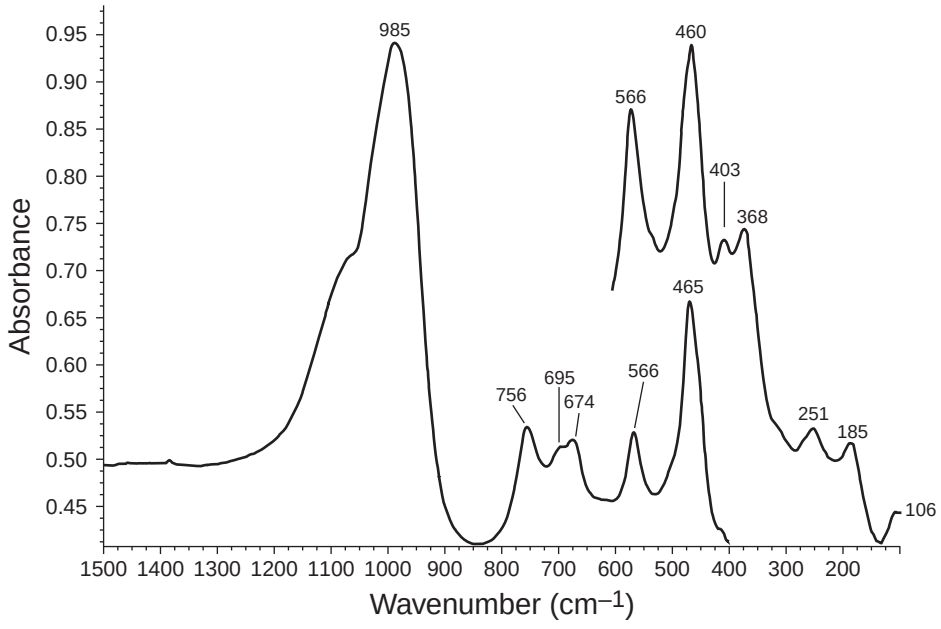

Figure 12 FTIR and FT/FIR spectrum of NaX zeolite.

with $Z = 4$. Due to its conducting behavior, it might not be expected to give either Raman or IR spectra. However, it is reported to give a strong Raman peak at 1580 cm^{-1} and no IR bands.[198] This is consistent with the vibrational analysis performed by Fateley et al.[3] assuming that the interaction between the layers is extremely weak and that the vibrational modes of different layers do not couple.

Vibrational and, in particular, Raman spectroscopy have also been successfully applied to glassy carbon,[199] diamite and carbon nanotubes,[200,201] fullerenes,[202] and metal fullerides.[203] In all cases the strongest Raman peaks fall in the region 1600–1300 cm^{-1}, characterizing a C–C stretching with a bond order intermediate between 1 and 2.

6.7.2 Binary Semiconductors

Most of group III–V semiconductors (cubic BN, and B, Al, Ga, and In phosphides, arsenides and antimonides) crystallize in the cubic zinc blende (sphalerite) structure ZnS (T_d^2, $F\bar{4}3m$) with $Z = 1$. This structure is also taken by β-SiC. This is a superstructure of the diamond structure. In this case there is only one triply-degenerate optical mode that is both IR and Raman active. Accordingly, this mode is polar and undergoes LO/TO splitting (see above) so that two peaks occur in the IR and Raman spectra. The position of the TO mode is sensitive to the weight of the atoms involved. It is found at 795 cm^{-1} for SiC and at 830 cm^{-1} for BP. For heavier atoms down it appears in the FIR region at 400–100 cm^{-1}.[204,205] These materials are therefore useful for windows for the medium IR. The IR reflection spectra allow determination of the composition of semiconducting solid solutions, such as $Al_xGa_{1-x}As$ and $In_xGa_{1-x}As$.[206]

Most group II–VI semiconductors take the hexagonal wurtzite structure, C_{6v}^4, $P6_3c/m$, as does one of the polymorphs of ZnS (wurtzite), ZnO, ZnSe, ZnTe, CdS, and CdSe. This structure is also taken by the nitrides of Al, Ga, In, and Tl, and by α-SiC. The vibrational structure is more complex here, with two Raman-active modes (E_2), two modes both IR and Raman active ($A_1 + E_1$) and one inactive mode (B_1) being expected.

6.7.3 Nonoxide Ceramics

Raman and IR spectroscopies are widely used to investigate nonconducting nonoxide ceramic materials. Aluminum nitride takes the wurtzite-type structure. The active modes have been found by Raman spectrometry to lie at 252 (E_2), 614 (A_1,TO), 660 (E_2), 673 (E_1,TO), 893 (A_1,LO), and 916 cm^{-1} (E_1,LO).[207] IR spectroscopy enables distinction between the polymorphs of silicon nitride, namely α-Si_3N_4, β-Si_3N_4,[208] and amorphous phases.[209]

The materials β-Si_3N_4 and β-Ge_3N_4 crystallize in the hexagonal C_{6h}^2 ($P6_3/m$) space group with $Z = 2$; their IR spectra show that Si–N–Si asymmetric stretchings fall in the region 1100–900 cm^{-1}, whereas Ge–N–Ge asymmetric stretchings fall in the region 920–800 cm^{-1}.[210] The substitutional solid solutions of Al and oxygen into β-Si_3N_4 give rise to "β'-sialons" (Si-Al oxynitrides). This substitution gives rise to broadening and shifts of the

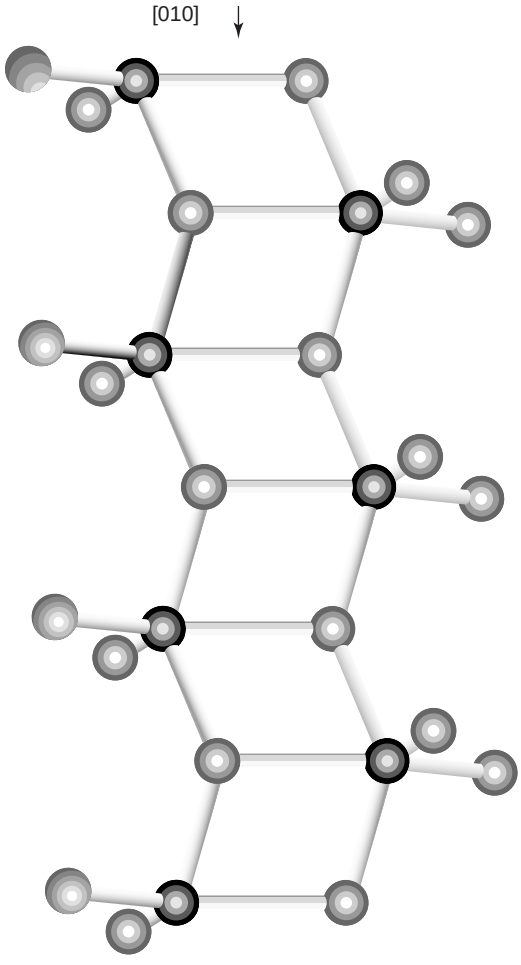

Figure 13 Structure of the anion of Mg metavanadate. (Reproduced with permission from G. Busca et al., *J. Chem. Soc. Faraday Trans.*, **90**, 1161–1170 (1994).)

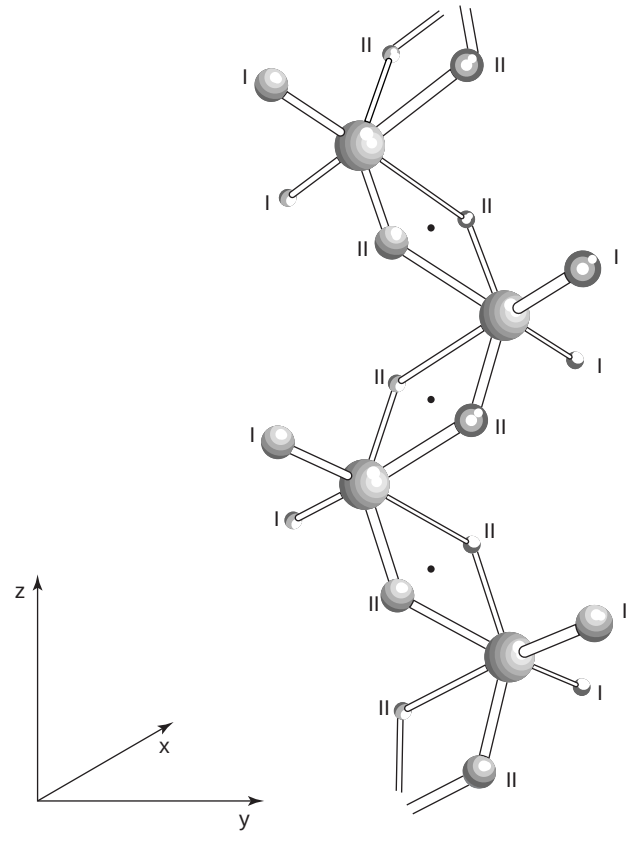

Figure 14 Structure of the anion of Cd tungstate. (Reproduced with permission from M. Daturi et al., *J. Phys. Chem.*, **101**, 4358–4369 (1997).)

IR absorption bands, allowing the determination of their composition.[211] The vibrational spectra of silicon oxynitride Si_2ON_2, where an oxygen bridge forms between SiN_3 units, has been the subject of skeletal vibrational studies[212] and lattice dynamics calculations.[213]

Boron nitride takes the zinc-blende diamond-like structure with tetrahedrally coordinated B and N atoms or, alternatively, a hexagonal graphite-like structure (h-BN, D_{6h}^4 space group). In this case both types of atoms have trigonal coordination. Two B–N stretching modes are detectable, one IR active and the other Raman active (1366 and 1373 cm^{-1}, respectively), characterizing the vibrations of BN_3 planar units.[214] A turbostratic sample shows vibrational features close to those of h-BN.[215] Raman and IR spectroscopies have recently been applied to the study of stresses undergone by ceramic materials such as Al and Si nitrides.[216,217] The antibixbyite compounds α-Ca_3N_2 and Mg_3N_2 have been the subject of a recent study.[218]

The material β-SiC takes the cubic zinc-blende structure. The splitting of the sole optical mode gives the transverse mode at 796 cm^{-1} and the longitudinal mode at 972 cm^{-1}, detectable both by IR and Raman.[219] The wurtzite structure is adopted by α-SiC, and other polymorphs present highest frequency fundamentals in similar frequency ranges.[220] IR spectra enable the detection of hydrogen and oxygen impurities in these materials.[221] The vibrational spectra of boron carbide $B_{13}C_2$ have also been investigated.[222]

6.7.4 Halides and Sulfides

The halides of most alkali metals take the rock-salt structure, which is Raman silent with only one triply-degenerate IR-active mode (see Table 1). The maximum of the transverse mode is located below 400 cm^{-1} in all cases[204] and shifts to lower frequencies on increasing the weight of either anion or cation. The CsCl structure taken by Cs and Tl^I halides (space group $Pm3m = O_h^1$ space group) also gives rise to only one triply-degenerate IR mode which is located (according to the relevant weight of the cations involved) in the FIR. This makes

both NaCl-type and CsCl-type halides useful materials for windows for the medium IR region. Several halides of divalent metals take the rutile structure (some slightly distorted), or the CdI_2 structure which is strongly related to the brucite structure of $MgCl_2$. Fluorite CaF_2 ($Fm3m = O_h^5$ space group, $Z = 4$) is largely used as a window material for IR due to its nonsolubility in water. The halides of Ba and Pb(II) crystallize in the $Pbnm = D_{2h}^{16}$ space group with $Z = 4$. Their IR spectra are quite complex.[223] However, alkaline earth halides absorb IR at higher frequencies than the corresponding alkali metal halides (the highest frequency IR absorption maximum is at 410 cm^{-1} for CaF_2 and at 465 cm^{-1} for rutile-type MgF_2).

Sulfur is heavier than oxygen, hence vibrational modes involving metal–sulfur bonds fall at lower frequencies than for the corresponding metal–oxygen bonds. Bivalent metal chalcogenides constitute the group II–VI class of semiconductors, taking either the zinc-blende or the wurtzite structures (see above). The position of the fundamental modes of zinc chalcogenides decreases by increasing the chalcogen weight.[204] The fundamental modes of Ca, Sr, and Ba sulfides[224] and Zn and Cd chalcogenides[225,226] fall in the FIR region (below 400 cm^{-1}). Rare earth sesquisulfides have been the object of recent Raman investigations.[227,228] The sulfides of tetravalent Mo and W have a well-known layered structure, but they can also be prepared in fullerene-like and nanotube-like forms. These MoS_2 and WS_2 polymorphic forms have been characterized by Raman spectroscopy.[229] The spectra of several sulfide and chalcogenide structures have been reported and discussed.[81–85]

6.8 Unusual Applications

The application of vibrational spectroscopies to different fields of inorganic materials research is expanding. Among the many interesting and unusual applications are Raman spectroscopic sensors for in situ mineral characterization on planetary surfaces,[230] the study of decorated stained glass[231] to characterize historic pigments[232] such as those used in native American Indian rock art[233] and by Romans,[234] and determination of the composition of medieval pottery fragments from the south of Italy.[235]

Vibrational techniques are used in the characterization of minerals and hydrated products[236] of construction cement, ferroelectric and ferroelastic materials,[237–239] nonlinear optical materials,[240] in petrography,[241] and the analysis of soils.[242] Raman techniques have been applied in ceramics science for the analysis of polymorphs of silicate materials[243] and their thermal expansion properties.[244] FIR is used to characterize dielectric ceramics for microwave applications.[245] Vibrational spectroscopies are used to characterize solid catalysts,[246] and to investigate inorganic materials of biological and medical interest, such as urinary stones[247] and hydroxyapatite.[248]

ABBREVIATIONS AND ACRONYMS

ASS	Atom Scattering Spectroscopy
ATR	Attenuated Total Reflection
CIR	Cylindrical Internal Reflection
DRIFT	Diffuse Reflectance Infrared Fourier Transform
EELS	Electron Energy Loss Spectroscopy
FIR	Far Infrared
FT	Fourier Transform
FTIR	Fourier Transform Infrared
IETS	Inelastic Electron Tunneling Spectroscopy
INS	Inelastic Neutron Scattering
IR	Infrared
IRES	Infrared Emission Spectroscopy
IRRAS	Infrared Reflection Absorption Spectroscopy
K–M	Kubelka–Munk
LO	Longitudinal Optical
NIR	Near Infrared
PAS	Photoacoustic Spectrometry
PBDS	Photothermal Beam Deflection Spectroscopy
SERS	Surface-enhanced Raman Scattering
SEWS	Surface Electromagnetic Wave Spectroscopy
TO	Transverse Optical
UV	Ultraviolet
XRD	X-ray Diffraction

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Clinical and Diagnostic Analysis • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications • Vibrational Optical Activity of Pharmaceuticals and Biomolecules

Chemical Weapons Chemicals Analysis **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Soil Instrumental Methods

Food **(Volume 5)**
Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis

Peptides and Proteins **(Volume 7)**
Fourier Transform Infrared Spectroscopy in Peptide and Protein Analysis

Polymers and Rubbers **(Volume 9)**
Infrared Spectroscopy in Analysis of Plastics Recycling • Infrared Spectroscopy in Analysis of Polymer Crystallinity • Infrared Spectroscopy in Analysis of Polymer Degradation • Infrared Spectroscopy in Analysis of Polymer Structure–Property Relationships • Infrared Spectroscopy in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Infrared Spectroscopy in Process Analysis • Near-infrared Spectroscopy in Process Analysis • Raman Spectroscopy in Process Analysis

Pulp and Paper **(Volume 9)**
Fourier Transform Infrared Spectroscopy in the Pulp and Paper Industry

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Infrared Spectroelectrochemistry

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Emission Spectroscopy, Infrared • Gas Chromatography/Infrared Spectroscopy • Infrared Reflection–Absorption Spectroscopy • Interpretation of Infrared Spectra, A Practical Approach • Liquid Chromatography/Infrared Spectroscopy • Microspectroscopy • Quantitative Analysis, Infrared • Spectral Data, Modern Classification Methods for • Spectral Databases, Infrared • Theory of Infrared Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation • Raman Microscopy and Imaging • Raman Scattering, Fundamentals

REFERENCES

1. R.N. Jones, 'Analytical Applications of Vibrational Spectroscopy. An Historical Review', in *Chemical, Biological and Industrial Applications of Infrared Spectroscopy*, ed. J.R. Durig, Wiley, New York, 1–50, 1985.
2. C.V. Raman, 'A New Radiation', *Ind. J. Phys.*, **2**, 387–398 (1928).
3. W.G. Fateley, F.R. Dollish, N.T. McDevitt, F.F. Bentley, *Infarared and Raman Selection Rules for Molecular and Lattice Vibrations*, Wiley, New York, 1972.
4. T.R. Gilson, P.J. Hendra, *Laser Raman Spectroscopy*, Wiley, Chichester, 1970.
5. G. Turrell, *Infrared and Raman Spectra of Crystals*, Academic Press, London, 1972.
6. J.C. Decius, R.M. Hexter, *Molecular Vibrations in Crystals*, McGraw-Hill, New York, 1977.
7. K. Nakamoto, *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, Wiley, New York, 1986.
8. J.R. Ferraro, K. Nakamoto, *Introductory Raman Spectroscopy*, Academic Press, San Diego, CA, 1994.
9. C. Zhu, P.R. Griffiths, 'Extending Range of Beer's Law in FT/IR Spectrometry. Part I Theoretical Study of Norton–Beer Apodization Functions', *Appl. Spectrosc.*, **52**, 1403–1408 and 1409–1413 (1998).
10. R.L. Henry, 'The Transmission of Powder Films in the Infrared', *J. Opt. Soc. Am.*, **38**, 775–789 (1948).
11. R. Lejeune, G. Duykaerts, 'Note Concernant les Spectres Infrarouge de Poudres', *Spectrochim. Acta*, **6**, 194–197 (1954).
12. G. Duykaerts, *Analyst*, **84**, 201 (1959).
13. M. Bliss, B.L. Walden, W.B. White, 'Measurement of Infrared Spectra of Dense Ceramics by Specular Reflectance Spectroscopy', *J. Am. Ceram. Soc.*, **73**, 1078–1083 (1990).
14. S. Hayashi, N. Nakamori, J. Hirono, H. Kanamori, 'Infrared Study of Surface Vibration Modes in MgO Small Cubes', *J. Phys. Soc. Jpn.*, **43**, 2006–2012 (1977).
15. F. Boccuzzi, S. Coluccia, G. Ghiotti, A. Zecchina, 'Influence of Surface Relaxation on the IR Spectrum of

Microcrystalline MgO', *Chem. Phys. Lett.*, **78**, 388–391 (1981).
16. R.G. Greenler, 'Infrared Study of Adsorbed Molecules on Metal Surfaces by Reflection Techniques', *J. Chem. Phys.*, **44**, 310–315 (1966).
17. J. Pritchard, 'Reflection–Absorption Infrared Spectroscopy', in *Chemical Physics of Solids and their Surfaces*, RSC, London, 157–179, Vol. 7, 1978.
18. D.K. Ottesen, A.S. Nagelberg, *Thin Solid Films*, **73**, 347 (1980).
19. W. Plieth, W. Kozlowski, T. Twomey, 'Reflectance Spectroscopy and Ellipsometry of Organic Monolayers', in *Adsorption of Molecules and Metal Electrodes*, eds. J. Lipkowski, P.N. Ross, VCH, Weinheim, 239–249, 1992.
20. A. Ulman, 'Fourier Transform Infrared Spectroscopy of Langmuir–Blodgett and Self-assembled Films', in *Fourier Transform Infrared Spectroscopy in Colloid and Interface Science*, ed. D.R. Scheuing, ACS Symp. Ser. 447, ACS, Washington, 144–159, 1990.
21. F.J. Boerio, J.P. Boerio, R.C. Bozian, 'Infrared Spectroscopy of Thin Organic Films on Metal Surfaces', *Appl. Surf. Sci.*, **31**, 42–58 (1988).
22. N.J. Harrick, *Internal Reflection Spectroscopy*, Interscience, New York, 1967.
23. J.L. Koenig, 'Fourier Transform Infrared Spectroscopy of Polymers', in *Spectroscopy: NMR, Fluorescence, FT/IR*, Advances in Polymer Science, eds. H.J. Cantow, G. Dall'Asta, K. Dusek, J.D. Ferry, H. Fujita, M. Gordon, G. Heinrici-Olivè, H.H. Kausch, J.P. Kennedy, W. Kern, S. Okamura, S. Olivè, C.G. Overberger, T. Saegusa, G.V. Schulz, W.P. Slichter, J.K. Stille, Springer Verlag, Berlin, 89–154, Vol. 54, 1984.
24. J.L. Koenig, *Spectroscopy of Polymers*, ACS, Washington, 1992.
25. K.H. Beckmann, N.J. Harrick, 'Hydrides and Hydroxyls in Thin Silicon Dioxide Films', *J. Electrochem. Soc.*, **118**, 614–619 (1971).
26. R. Koba, J.R. Monkowski, R.E. Tressler, 'The Hygroscopic Behavior of SiO_2 and $Si_xO_yCl_z$ Thin Films on Si Using ATR FTIR Spectra', in *Advances in Materials Characterization II*, eds. R.L. Snyder, R.A. Condrate, P.F. Johnson, Plenum Press, New York, 301–313, 1985.
27. T. Uchino, T. Sakka, K. Hotta, M. Iwasaki, 'Attenuated Total Reflectance Fourier-transform Infrared Spectra of a Hydrated Sodium Silicate Glass', *J. Am. Ceram. Soc.*, **72**, 2173–2175 (1989).
28. M. Osawa, K. Ataka, K. Yoshii, T. Yatsuyanagi, 'Surface-enhanced Infrared ATR Spectroscopy for in situ Studies of Electrode/Electrolyte Interfaces', in *Vibration at Surfaces 1993*, eds. M. Rocca, G. Bracco, U. Valbusa, Elsevier, Amsterdam, 371–379, 1993.
29. B.W. Johnson, J. Banhofer, K. Doblhofer, B. Pettinger, 'Practical Considerations of the ir Attenuated-total-reflection (ir-ATR) Technique for Electrochemical Investigations', *Electrochim. Acta*, **37**, 2321–2329 (1992).
30. D.R. Scheuing (ed.), *Fourier Transform Infrared Spectroscopy in Colloid and Interface Science*, ACS Symp. Ser. 447, ACS, Washington, 1990.
31. P. Wilks, *Eur. Spectrosc. News*, **55**, 26 (1984).
32. W.R. Moser, J.C. Chaing, J.E. Cnossen, 'A New FT/IR Technique for the in situ Study of the Mechanism of Solid State Materials Preparations', in *Advances in Materials Characterization II*, eds. R.L. Snyder, R.A. Condrate, P.F. Johnson, Plenum Press, New York, 315–327, 1985.
33. E.H. Braue, M.J. Pannella, 'Consistency in Circle Cell FT/IR Analysis of Aqueous Solutions', *Appl. Spectrosc.*, **41**, 1057–1067 (1987).
34. P.A. Safranovsky, L.A. Kuzik, G.H. Zhinzhin, 'Low Energy Non Linear Surface Effects and Hole Burning in IR Absorption Spectra of Natural Alumina Films', *J. Electron. Spectrosc. Relat. Phenom.*, **54/55**, 457–468 (1990).
35. P. Kubelka, F. Munk, *Z. Tech. Phys.*, **12**, 593 (1931).
36. G. Kortum, *Reflectance Spectroscopy*, Springer, New York, 1969.
37. M.P. Fuller, P.R. Griffiths, 'Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry', *Anal. Chem.*, **50**, 1906–1910 (1978).
38. Z. Bodo, *Acta Phys. Hung.*, **14**, 311 (1962).
39. N.T. Melamed, 'Optical Properties of Powders. Part I. Optical Absorption Coefficients and the Absolute Value of the Diffuse Reflectance. Part II. Properties of Luminescent Powders', *J. Appl. Phys.*, **34**, 560–570 (1963).
40. E.L. Simmons, 'Modification of the Particle-model Theory of Diffuse Reflectance Properties of Powdered Samples', *J. Appl. Phys.*, **46**, 344–349 (1975).
41. E.L. Simmons, 'Diffuse Reflectance Spectroscopy: a Comparison of the Theory', *Appl. Opt.*, **14**, 1380–1386 (1975).
42. P.J. Brimmer, P.R. Griffiths, 'Angular Dependence of Diffuse Reflectance Infrared Spectra. Part II: Effect of Polarization', *Appl. Spectrosc.*, **41**, 791–797 (1987).
43. P.J. Brimmer, P.R. Griffiths, 'Angular Dependence of Diffuse Reflectance Infrared Spectra. Part III: Linearity of Kubelka–Munk Plots', *Appl. Spectrosc.*, **42**, 242–247 (1988).
44. D.J.J. Fraser, P.R. Griffiths, 'Effect of Scattering Coefficient on Diffuse Reflectance Infrared Spectra', *Appl. Spectrosc.*, **44**, 193–199 (1990).
45. K. Moradi, C. Depecker, J. Corset, 'Diffuse Reflectance Infrared Spectroscopy: Experimental Study of Nonabsorbing Materials and Comparison with Theories', *Appl. Spectrosc.*, **48**, 1491–1496 (1994).
46. J.R. Ferraro, A.J. Rein, 'Applications of Diffuse Reflectance Spectroscopy in the Far-infrared Region', in *Fourier Transform Infrared Spectroscopy, Applications to Chemical Systems*, eds. J.R. Ferraro, L.J. Basile, Academic Press, New York, 243–282, Vol. 4, 1985.
47. A.N. Davies, 'Reflections, or Are Your Spectra Only Skin Deep?', *Spectrosc. Eur.*, **6**(4), 25–28 (1994).

48. B. Li, R.D. Gonzalez, 'Design and Construction of a DRIFTS Accessory and an in situ Heatable Sample Cell', *Appl. Spectrosc.*, **52**, 1488–1491 (1998).
49. P.R. Griffiths, K.W. Van Every, N.A. Wright, I.M. Hamadeh, in *Chemical, Biological and Industrial Applications of Infrared Spectroscopy*, ed. J.R. Durig, Wiley, New York, 253–276, 1985.
50. L.M. Gratton, S. Paglia, F. Scattaglia, M. Cavallini, 'Infrared Emission Spectroscopy Applied to the Oxidation of the Molybdenum', *Appl. Spectrosc.*, **32**, 310–316 (1978).
51. D.H. Sullivan, W.C. Conner, M.P. Harold, 'Surface Analysis with FT/IR Emission Spectroscopy', *Appl. Spectrosc.*, **46**, 811–818 (1992).
52. J.R. Markham, J.E. Cosgrove, M.D. Carangelo, P.R. Solomon, P.E. Best, 'Temperature Profiling of Glass by FT/IR Emission Spectroscopy', *J. Am. Ceram. Soc.*, **78**, 897–902 (1995).
53. A.G. Bell, *Am. J. Sci.*, **29**, 305 (1980).
54. C.A. Rosencwaig, A. Gersho, *J. Appl. Phys.*, **47**, 64 (1976).
55. C.A. Rosencwaig, *Photoacoustics and Photoacoustic Spectroscopy*, Wiley, New York, 1980.
56. K. Kinoshita, *Carbon, Electrochemical and Physicochemical Properties*, Wiley, New York, 107–173, 1988.
57. B.M. Lynch, J.A. MacPhee, 'Photoacoustic FTIR Spectroscopy in Detection of Coal Oxidation', in *Chemistry of Coal Weathering*, ed. C.R. Nelson, Elsevier, Amsterdam, 83–106, 1989.
58. J.F. McClelland, 'Photoacoustic Spectroscopy', *Anal. Chem.*, **55**, 89A–105A (1983).
59. E.M. Eyring, S.M. Riseman, F.E. Massoth, 'Photoacoustic Spectroscopy of Catalyst Surface', in *Catalytic Materials, Relationship Between Structure and Reactivity*, eds. T.E. Whyte, Jr, R.A. Dalla Betta, E.G. Derouane, R.T.K. Baker, ACS, Washington, 399–410, 1984.
60. J.G. Highfield, J.B. Moffatt, 'Characterization of 12-Tungstophosphoric Acid and Related Salts Using Photoacoustic Spectroscopy in the Infrared Region. I. Thermal Stability and Interactions with Ammonia', *J. Catal.*, **88**, 177–187 (1984).
61. J.G. Highfield, J.B. Moffatt, 'Characterization of 12-Tungstophosphoric Acid and Related Salts Using Photoacoustic Spectroscopy in the Infrared Region. II. Interactions with Pyridine', *J. Catal.*, **89**, 185–195 (1984).
62. J.G. Highfield, J.B. Moffatt, 'Elucidation of Mechanism of Dehydration of Methanol Over 12-Tungstophosphoric Acid Using Infrared Photoacoustic Spectroscopy', *J. Catal.*, **95**, 108–119 (1985).
63. A.C. Boccara, D. Fournier, J. Badoz, 'Thermooptical Spectroscopy: Detection by the ''Mirage Effect'' ', *Appl. Phys. Lett.*, **36**, 130–132 (1980).
64. M.J.D. Low, C. Morterra, A.G. Severdia, J.M.D. Tascon, in *Catalytic Materials, Relationship Between Structure and Reactivity*, ACS, Washington, 411, 1984.
65. P. Bruesch, *Phonons: Theory and Experiments*, Parts I, II and III, Springer Verlag, Berlin, 1986.
66. R.K. Chang, T.E. Furtak (eds.), *Surface Enhanced Raman Scattering*, Plenum, New York, 1982.
67. J.M. Stencel, *Raman Spectroscopy for Catalysis*, Van Nostrand, New York, 1990.
68. Y. Iida, M. Furukawa, K. Kato, H. Morikawa, 'In situ Anti-stokes Raman Monitoring of Gel-to-Anatase Phase Transformation of Titania', *Appl. Spectrosc.*, **51**, 673–677 (1997).
69. H. Jobic, 'Application of Neutron Scattering to Catalysis', in *Catalyst Characterization, Physical Techniques for Solid Materials*, eds. B. Imelik, J. Vedrine, Plenum Press, New York, 347–375, 1994.
70. S.F. Parker, 'Vibrational Spectroscopy with Neutrons', *Spectrosc. Eur.*, **6**, 14–20 (1994).
71. H. Jobic, 'Molecular Motions in Zeolites', *Spectrochim. Acta*, **48A**, 293–312 (1992).
72. H. Jobic, A. Tuel, M. Krossner, J. Sauer, 'Water in Interaction with Acid Sites in HZSM5 does not Form Hydroxonium Ions. A Comparison Between Neutron Scattering Results and ab initio Calculations', *J. Phys. Chem.*, **100**, 19 545–19 550 (1996).
73. J.M. Nicol, 'Chemisorbed Hydrogen and Hydrogenous Molecules', *Spectrochim. Acta*, **48A**, 313–327 (1992).
74. E. Fillaux, C. Cachet, H. Ouboumour, J. Tomkinson, C. Levy-Clement, L.T. Yu, 'Inelastic Neutron Scattering Study of γ-MnO_2: The Dynamics of $(H^+)_4$ Entities in Mn^{4+} Vacancies', *Chem. Phys.*, **164**, 311–319 (1992).
75. T.J. Udovic, R.D. Kelley, in *Hydrogen Effects in Catalysis*, eds. Z. Paal, P.G. Menon, Dekker, New York, 167, 1988.
76. H. Jobic, G. Glugnet, M. Lacroix, S. Yuan, C. Mirodatos, M. Breysse, 'Identification of New Hydrogen Species Adsorbed on Ruthenium Sulfide by Neutron Spectroscopy', *J. Am. Chem. Soc.*, **115**, 3654–3657 (1993).
77. P.G. Hall, N.S. Clarke, S.P.C. Maynard, 'Inelastic Neutron Scattering (TFXA) Study of Hydrogen Modes in α-FeOOH (Goethite) and γ-FeOOH (Lepidocrocite)', *J. Phys. Chem.*, **99**, 5666–5673 (1995).
78. Y. Fillaux, S.M. Bennington, J. Tomkinson, L.T. Yu, 'Inelastic Neutron Scattering Study of Free Proton Dynamics in Gamma-MnO_2', *Chem Phys.*, **209**, 111–125 (1996).
79. J. Li, 'INS Studies of Hydrogen Bonding in Ices', *J. Chem. Phys.*, **105**, 6733–6755 (1996).
80. R.S. Klein, G.E. Kugel, B. Hennion, 'INS Measurements on the TA and Lowest Optic Modes in $KTaO_3$ Doped with Li', *J. Phys. Condens. Matter*, **8**, 1109–1121 (1996).
81. V.C. Farmer, *The Infrared Spectra of Minerals*, Mineralogical Society Monograph 4, Mineralogical Society, London, 1974.
82. J.A. Gadsden, *Infrared Spectra of Minerals and Related Inorganic Compounds*, Butterworths, London, 1975.
83. W.P. Griffith, 'Advances in the Raman and Infrared Spectroscopy of Minerals', in *Spectroscopy of Inorganic-Based Materials*, eds. R.J.H. Clark, R.E. Hester, Wiley, New York, 119–186, Vol. 14, 1987.

84. R.A. Nyquist, R.O. Kagel, *Infrared Spectra of Inorganic Compounds*, Academic Press, New York, 1971.
85. R. Nyquist, 'Electronic Handbook of Infrared and Raman Spectra of Inorganic Compounds and Organic Salts', Academic Press, New York, 1999.
86. Wang, J. Han, L. Guo, J. Yu, P. Zeng, 'Database of Standard Raman Spectra of Minerals and Related Inorganic Crystals', *Spectrochim. Acta*, **A48**, 959–968 (1994).
87. D.W. Smith, 'An Acidity Scale for Binary Oxides', *J. Chem. Educ.*, **64**, 480–481 (1987).
88. F.A. Cotton, G. Wilkinson, *Advanced Inorganic Chemistry*, 4th edition, Wiley, New York, 1980.
89. T. Moeller, *Inorganic Chemistry, a Modern Introduction*, Wiley, New York, 1982.
90. A.R. West, *Solid State Chemistry and its Applications*, Wiley, New York, 1984.
91. A.F. Wells, *Structural Inorganic Chemistry*, Clarendon Press, Oxford, 1975.
92. J. Haber, 'Crystallography of Catalyst Types', in *Catalysis, Science and Technology*, eds. J.R. Anderson, M. Boudart, Springer Verlag, Berlin, 13–95, Vol. 2, 1981.
93. B.G. Hyde, S. Andersson, *Inorganic Crystal Structures*, Wiley, New York, 1989.
94. W. Vogel, *Chemistry of Glass*, The American Ceramic Society, Columbus, OH, 1985.
95. R.T. Sanderson, *Chemical Bonds and Bond Energy*, 2nd edition, Academic Press, New York, 1976.
96. J. Haber, A. Kozlowska, R. Kozlowski, 'The Structure and Redox Properties of Vanadium Oxide Surface Compounds', *J. Catal.*, **102**, 52–63 (1986).
97. M. Daturi, G. Busca, M.M. Borel, A. Leclaire, P. Piaggio, 'Vibrational and XRD Study of the System $CdWO_4$-$CdMoO_4$', *J. Phys. Chem.*, **101B**, 4358–4369 (1997).
98. P. Tarte, 'Etude Infra-rouge des Orthosilicates et des Orthogermanates, une Nouvelle Methode d'Interpretation des Spectres', *Spectrochim. Acta*, **18**, 467–483 (1962).
99. P. Tarte, 'Infrared Spectra of Inorganic Aluminates and Characteristic Vibrational Frequencies of AlO_4 Tetrahedra and AlO_6 Octahedra', *Spectrochim. Acta*, **23A**, 2127–2143 (1967).
100. I.R. Beattie, T.R. Gilson, 'Oxide Phonon Spectra', *J. Chem. Soc. A*, 2322–2327 (1969).
101. M. Handke, W. Mozgawa, 'Vibrational Spectroscopy of the Amorphous Silicates', *Vibrat. Spectrosc.*, **5**, 75–84 (1993).
102. K.J. Kingma, R.J. Hemley, 'Raman Spectroscopic Study of Microcrystalline Silica', *Am. Mineral.*, **79**, 269–273 (1994).
103. R.J. Bell, P. Dean, 'Atomic Vibrations in Vitreous Silica', *Discuss. Faraday Soc.*, **50**, 55–61 (1970).
104. F.L. Galeener, 'Bands Limits and the Vibrational Spectra of Tetrahedral Glasses', *Phys. Rev. B*, **19**, 4292–4297 (1979).
105. R.A. Van Santen, D.L. Vogel, 'Lattice Dynamics of Zeolites', *Adv. Solid State Chem.*, **1**, 151–224 (1989).
106. A.J.M. deMan, B.N.H. van Beest, M. Leslie, R.A. van Santen, 'Lattice Dynamics of Zeolitic Silica Polimorphs', *J. Phys. Chem.*, **94**, 2524–2534 (1990).
107. P.N. Sen, M.F. Thorpe, 'Phonon in AX2 Glasses: From Molecular to Band-like Modes', *Phys. Rev. B*, **15**, 4030–4038 (1977).
108. S.K. Sharma, J.F. Mammone, M.F. Nicol, *Nature*, **292**, 140–141 (1981).
109. E.I. Kamitsos, A.P. Patsis, G. Kordas, 'Infrared-reflectance Spectra of Heat Treated, Sol-gel-derived Silica', *Phys. Rev. B*, **48**, 12 499–12 505 (1993).
110. E.M. Flanigen, in *Zeolite Chemistry and Catalysis*, ed. J.A. Rabo, ACS, Washington, 80, 1979.
111. J. Etchepare, M. Merian, L. Smetankine, 'Vibrational Normal Modes of SiO_2. I. α and β Quartz', *J. Chem. Phys.*, **60**, 1873–1876 (1974).
112. E. Astorino, J. Peri, R.J. Willey, G. Busca, 'Spectroscopic Characterization of Silicalite-1 and Titanium Silicalite-1', *J. Catal.*, **157**, 482–500 (1995).
113. R. Feile, 'Lattice Vibrations in High-T_c Superconductors: Optical Spectroscopy and Lattice Dynamics', *Phys. C*, **159**, 1–32 (1989).
114. M. Daturi, G. Busca, E. Magnone, M. Ferretti, 'FT/IR Skeletal Study of $RBa_2Cu_3O_{7-x}$ (R = Ln or Y) and $Nd_{2-x}Ce_xCuO_4$ Cuprate Powders', *J. Solid State Chem.*, **119**, 36–44 (1995).
115. M. Napoletano, J.M. Gallardo Amores, E. Magnone, G. Busca, M. Ferretti, 'Skeletal Infrared Spectra and Structural Properties of $La_{2-x}Sr_xCuO_4$ and $La_{2-x}Ba_xCuO_4$ Cuprate Powders in the $0 \leq x \leq 0.125$ Region', *Phys. C Supercond.*, **319**, 229–237 (1999).
116. M. Daturi, M. Ferretti, E. Magnone, G. Busca, 'Thermal Stability and Structural Aspects of $Y_{1-x}Pr_xBa_2Cu_3O_7$ Solid Solution', *Nuovo Cimento*, **19D**, 1111–1116 (1997).
117. J. Humlicek, A.P. Litvinchuk, W. Kress, B. Lederle, C. Thomsen, M. Cardona, H.U. Habermeier, I.E. Trofimov, W. Konig, 'Lattice Vibrations of $Y_{1-x}Pr_xBa_2Cu_3O_7$: Theory and Experiment', *Phys. C*, **206**, 345–359 (1993).
118. R.P.S.M. Lobo, C. Allancon, F.J. Gotor, J.-M. Bassat, J.-P. Loup, P. Odier, K. Dembinski, F. Gervais, C. Champeaux, P. Marchet, A. Catherinot, 'Analysis of Infrared-visible-near-ultraviolet Reflectivity of Conducting and Superconducting Oxides', *Phys. C*, **235–240**, 1071–1072 (1994).
119. W.B. White, B. De Angelis, 'Interpretation of the Vibrational Spectra of Spinels', *Spectrochim. Acta*, **23A**, 985–995 (1967).
120. J. Preudhomme, P. Tarte, 'Infrared Studies of Spinels-III. The Normal II-III Spinels', *Spectrochim. Acta*, **27A**, 1817–1835 (1971).
121. G.C. Allen, M. Paul, 'Chemical Characterization of Transition Metal Spinel-type Oxides by IR Spectroscopy', *Appl. Spectrosc.*, **49/4**, 451–458 (1995).

122. H.D. Lutz, B. Muller, H.J. Steiner, 'Lattice Vibration Spectra. LIX. Single Crystal Infrared and Raman Studies of Spinel Type Oxides', *J. Solid State Chem.*, **90**, 54–60 (1991).
123. M.I. Baraton, G. Busca, V. Lorenzelli, R.J. Willey, 'Characterization of High-area $Mg_xFe_2O_{3+x}$ $(0 \leq x \leq 2)$ Monophasic Powders by Vibrational Spectroscopies', *J. Mater. Sci. Lett.*, **13**, 275–277 (1994).
124. M. Bellotto, G. Busca, C. Cristiani, G. Groppi, 'FT/IR Skeletal Powder Spectra of Ba-ß-aluminas with Compositions $BaAl_9O_{14.5}$, $BaAl_{12}O_{19}$ and $BaAl_{14}O_{22}$ and of Ba-ferrite $BaFe_{12}O_{19}$', *J. Solid State Chem.*, **117**, 8–15 (1995).
125. J. Kreisel, G. Lucazeau, H. Vincent, 'Raman Spectra and Vibrational Analysis of $BaFe_{32}O_{19}$ Hexagonal Ferrite', *J. Solid State Chem.*, **137**, 127–137 (1998).
126. R. Frech, J.B. Bates, 'Raman, i.r. Reflection and Emission Spectra of Sodium β-Alumina', *Spectrochim. Acta*, **35A**, 685–694 (1979).
127. Ph. Colomban, G. Lucazeau, 'Vibrational Study of and Conduction Mechanism in β-Alumina. I. Stoichiometric β-Alumina', *J. Chem. Phys.*, **72**, 1213–1224 (1980).
128. W. Hayes, F.L. Pratt, 'Far-infrared Absorption of Barium and Lead β-Alumina', *J. Phys. C Solid State Phys.*, **19**, 929–936 (1986).
129. G. Mariotto, G.C. Farrington, 'Raman Spectra of Cadmium-containing Beta Alumina Crystals', *Solid State Ion.*, **18&19**, 619–623 (1986).
130. Ya.I. Ryskin, 'The Vibrations of Protons in Minerals: Hydroxyls, Water and Ammonium', in *The Infrared Spectra of Minerals*, ed. V.C. Farmer, Mineralogical Society Monograph 4, Mineralogical Society, London, 137–181, 1974.
131. H.D. Lutz, H. Moller, M. Schmidt, 'Lattice Vibration Spectra. Part LXXXII. Brucite-type Hydroxide $M(OH)_2$ (M = Ca, Mn, Co, Fe, Cd) – IR and Raman Spectra, Neutron Diffraction of $Fe(OH)_2$', *J. Mol. Struct.*, **328**, 121–132 (1994).
132. H.D. Lutz, K. Beckenkamp, St. Peter, 'Laurionite-type $M(OH)_x$ (M=Ba, Pb; X=Cl, Br, I) and Sr(OH)I. An IR and Raman Spectroscopic Study', *Spectrochim. Acta*, **A51**, 755–767 (1995).
133. B. Weckler, H.D. Lutz, 'Near-infrared Spectra of M(OH)Cl(M=Ca, Cd, Sr), Zn(OH)F, γ-$Cd(OH)_2$, $Sr(OH)_2$ and Brucyte-type Hydroxides $M(OH)_2$ (M=Mg, Ca, Mn, Fe, Co, Ni, Cd)', *Spectrochim. Acta*, **52A**, 1507–1513 (1996).
134. M.C. Stegmann, D. Vivien, C. Mazieres, 'Etude des Modes de Vibration Infrarouge dans les Oxyhydroxydes d'Aluminum Boehmite et Diaspore', *Spectrochim. Acta*, **29A**, 1653–1663 (1972).
135. P.G. Hall, N.S. Clarke, S.C.P. Maynard, 'Inelastic Neutron Scattering (TFXA) Study of Hydrogen Modes in α-FeOOH (goethite) and γ-FeOOH (Lepidocrocite)', *J. Phys. Chem.*, **99**, 5666 (1995).
136. G. Busca, G. Ramis, M.C. Prieto, V. Sanchez Escribano, 'Preparation and Characterization of $Fe_{2-x}Cr_xO_3$ Mixed Oxide Powders', *J. Mater. Chem.*, **3**, 665–673 (1993).
137. J.M. Gallardo-Amores, V. Sanchez Escribano, G. Busca, 'Synthesis and Characterization of Fe-Ga Mixed Hydroxides', *J. Mater. Chem.*, **9**, 1161–1166 (1999).
138. J.J. Fripiat, H. Bosmans, P.G. Rouxhet, 'Proton Mobility in Solids. I. Hydrogenic Vibration Modes and Proton Delocalization in Boehmite', *J. Phys. Chem.*, **71**, 1097–1111 (1967).
139. J.D. Russell, V.C. Farmer, D.G. Lewis, 'Lattice Vibrations of Boehmite (γ-AlOOH): For a C_{2v}^{12} Rather Than a D_{2h}^{17} Space Group', *Spectrochim. Acta*, **34A**, 1151–1153 (1978).
140. A.B. Kiss, G. Keresztury, L. Farkas, 'Raman and i.r. Spectra and Structure of Boehmite (γ-AlOOH). Evidence for the Recently Discorded D_{2h}^{17} Space Group', *Spectrochim. Acta*, **36A**, 653–658 (1980).
141. M. Trombetta, G. Busca, R.J. Willey, 'Characterization of Silica-containing Aluminum Hydroxide and Oxide Aerogels', *J. Colloids Interface Sci.*, **190**, 416–426 (1997).
142. S.E. Horsley, D.V. Nowell, D.T. Stewart, 'The Infrared and Raman Spectra of α-Zirconium Phosphate', *Spectrochim. Acta*, **30A**, 535–541 (1974).
143. T.G. Koch, N.S. Holmes, T.B. Roddis, J.R. Sodeau, 'Low-temperature Reflection/Absorption IR Study of Thin Films of Nitric Acid Hydrates and Ammonium Nitrate Adsorbed on Gold Foil', *J. Chem. Soc. Faraday Trans.*, **92**, 4787–4792 (1996).
144. W.B. White, 'The Carbonate Minerals', in *The Infrared Spectra of Minerals*, ed. V.C. Farmer, Mineralogical Society Monograph 4, Mineralogical Society, London, 227–283, 1974.
145. A. Rulmont, J.P. Flamme, M.J. Potttier, B.M. Wanklyin, 'Birefringence Effect in the Raman Spectrum of a Crystal Which is not Cut Parallel to the Principal Axes – II. Application to a Single Crystal of $LaBO_3$', *Spectrochim. Acta*, **35A**, 635–639 (1979).
146. J.M. Alia, Y. Diaz de Mera, H.G.M. Edwards, P. Gonzalez Martin, S. Lopez Andres, 'FT-Raman and Infrared Spectroscopic Study of Aragonite–Strontianite ($Ca_xSr_{1-x}CO_3$) Solid Solution', *Spectrochim. Acta*, **A53**, 2347–2362 (1997).
147. M.H. Brooker, J.B. Bates, 'On the Strucutre of the Cubic Crystals: $Ca(NO_3)_2$, $Sr(NO_3)2$, $Ba(NO_3)_2$ and $Pb(NO_3)_2$. An Infrared and Raman Study', *Spectrochim. Acta*, **29A**, 439–452 (1973).
148. A.S. Barker, 'Infrared Lattice Vibrations in Calcium Tungstate and Calcium Molybdate', *Phys. Rev. A*, **135**, 742–747 (1964).
149. G.M. Clark, W.P. Doyle, 'Infrared Spectra of Anhydrous Molybdates and Tungstates', *Spectrochim. Acta*, **22**, 1441–1447 (1966).

150. R.K. Khanna, E.R. Lippincott, 'Infrared Spectra of Some Scheelite Structures', *Spectrochim. Acta*, **24A**, 905–908 (1968).
151. P. Tarte, M. Liegeois-Duyckaerts, 'Vibrational Studies of Molybdates, Tungstates and Related Compounds – I. New Infrared Data and Assignments for the Scheelite-type Compounds', *Spectrochim. Acta*, **28A**, 2029–2036 (1972).
152. M. Liegeois-Duyckaerts, P. Tarte, 'Vibrational Studies of Molybdates, Tungstates and Related Compounds – II. New Raman Data and Assignments for the Scheelite-type Compounds', *Spectrochim. Acta*, **28A**, 2037–2051 (1972).
153. M. Daturi, G. Busca, M.M. Borel, A. Leclaire, P. Piaggio, 'A Vibrational and XRD Study of the System $CdWO_4$-$CdMoO_4$', *J. Phys. Chem. B*, **101**, 4358–4369 (1997).
154. I. Kanesaka, H. Hashiba, I. Matssura, 'Polarized Raman Spectrum and Normal Coordinate Analysis of α-Manganese Molybdate ($MnMoO_4$), *J. Raman Spectrosc.*, **19**, 213–218 (1988).
155. P.P. Cord, P. Courtine, G. Pannetier, 'Spectre de Vibration des Chromates et des Molybdates des Metaux Divalents de la Premiere Serie de Transition', *Spectrochim. Acta*, **28A**, 1601–1613 (1972).
156. P.M. Nikolic, P. Mihailovic, D.M. Todorovic, 'For Infrared and Infrared Properties of Single Crystal Anglesite', *Spectrochim. Acta*, **52A**, 131–137 (1996).
157. J. Preudhomme, P. Tarte, 'Infrared Studies of Spinels – IV Normal Spinels with a High-Valency Tetrahedral Cation', *Spectrochim. Acta*, **28A**, 69–79 (1972).
158. P. Tarte, 'Etude Infra-rouge des Orthosilicates et des Orthogermanates – III, Structure du Type Spinelle', *Spectrochim. Acta*, **19**, 49–71 (1963).
159. P. Tarte, 'Etude Infra-rouge des Orthosilicates et des Orthogermanates – II, Structure du Type Olivine et Monticellite', *Spectrochim. Acta*, **19**, 25–47 (1963).
160. M.Th. Paques-Ledent, P. Tarte, 'Vibrational Studies of Olivine-type Compounds – I. The ir and Raman Spectra of the Isotopic Species of Mg_2SiO_4', *Spectrochim. Acta*, **29A**, 1007–1016 (1973).
161. R.L. Carter, C.E. Bricker, 'Laser-Raman Spectra of Crystalline K_2CrO_4, Rb_2CrO_4 and Cs_2CrO_4', *Spectrochim. Acta*, **27A**, 569–580 (1971).
162. F.M. Njagi, U.A. Jayasooriya, S.F.A. Kettle, R. Autchakit, P. Bhanthumnavin, W. Trakarnpruk, 'Latent Site Symmetry Effects in the Vibrational Spectra of Potassium Sulfate, Potassium Chromate and their Mixed Crystals', *Spectrochim. Acta*, **48A**, 1403–1408 (1992).
163. P. Tarte, J. Thelen, 'Spectre Vibrationnel des Composes du Type $Ba_3(XO_4)_2$ et $Sr_3(XO_4)_2$ (X^V = P, As, V, Cr, Mn)', *Spectrochim. Acta*, **28A**, 5–14 (1972).
164. G. Busca, G. Ricchiardi, D. Siew Hew Sam, J.C. Volta, 'Spectroscopic Characterization of Magnesium Vanadate Catalysts. Part 1 – Vibrational Characterization of $Mg_3(VO_4)_2$, $Mg_2V_2O_7$ and MgV_2O_6 Powders', *J. Chem. Soc. Faraday Trans.*, **90**, 1161–1170 (1994).
165. S.D. Ross, 'Phosphates and Other Oxy-anions of Group V', in *The Infrared Spectra of Minerals*, ed. V.C. Farmer, Mineralogical Society Monograph 4, Mineralogical Society, London, 383–421, 1974.
166. E.J. Baran, M.S. Schriewer-Pottgen, W. Jeitschko, 'Vibrational Spectrum of Hg_2ReO_5', *Spectrochim. Acta*, **A52**, 441–444 (1996).
167. V.V. Fomichev, A.V. Proskuryakova, 'Vibrational Spectra and Energy Characteristics of the Superionics Li_4SiO_4 and Li_4GeO_4', *J. Solid State Chem.*, **134**, 232–237 (1997).
168. S.E. Horsley, D.V. Nowell, D.T. Stewart, 'The Infrared and Raman Spectra of α-Zirconium Phosphate', *Spectrochim. Acta*, **30A**, 535–541 (1974).
169. L.W. Schroeder, T.H. Jordan, W.E. Brown, 'Raman Spectrum from a Single Crystal of $SnHPO_4$', *Spectrochim. Acta*, **37A**, 21–29 (1981).
170. V. Ramakrishnan, G. Aruldhas, 'Raman and Infrared Studies of Sodium Phosphate Hydrate ($Na_2HPO_4 \cdot 2H_2O$)', *J. Raman Spectrosc.*, **18**, 145–146 (1987).
171. M. Gabelica-Robert, P. Tarte, 'Vibrational Spectrum of Akermanite-like Silicates and Germanates', *Spectrochim. Acta*, **35A**, 649–654 (1979).
172. M. Gabelica-Robert, P. Tarte, 'Vibrational Spectrum of Fresnoite ($Ba_2TiOSi_2O_7$) and Isostructural Compounds', *Phys. Chem. Miner.*, **7**, 26–30 (1981).
173. A.N. Lazarev, *Vibrational Spectra and Structure of Silicates*, Consultants Bureau, New York, 1972.
174. S.D. Ross, 'The Vibrational Spectrum of Thorveitite', *Spectrochim. Acta*, **27A**, 1837–1843 (1971).
175. P. Tarte, M.J. Pottier, A.M. Procés, 'Vibrational Studies of Silicates and Germanates – V. IR and Raman Spectra of Pyrosilicalites and Pyrogermanates with a Linear Bridge', *Spectrochim. Acta*, **29A**, 1017–1027 (1973).
176. R.G. Brown, S.D. Ross, 'Vibrational Spectra of Some Condensed Tetrahedral Anions $[X_2O_7]^{n-}$', *Spectrochim. Acta*, **28A**, 1263–1274 (1972).
177. R. Hubin, 'Spectre d'Absorption Infra-rouge des Pyrophosphates et Pyroarseniates Cubique et Ortorombique de Thorium', *Spectrochim. Acta*, **27A**, 311–319 (1971).
178. E. Husson, Y. Repelin, N.Q. Dao, H. Brusset, 'Etude par Spectrophotometries d'Absorption Infrarouge et de Diffusion Raman des Niobiates de Structure Columbite', *Spectrochim. Acta*, **33A**, 995–1001 (1977).
179. V.C. Farmer, 'Orthosilicates, Pyrosilicates and Other Finite-chain Silicates', in *The Infrared Spectra of Minerals*, ed. V.C. Farmer, Mineralogical Society Monograph 4, Mineralogical Society, London, 285–303, 1974.
180. V.C. Farmer, J.D. Russel, 'Effects of Particle Size and Structure on the Vibrational Frequencies of Layer Silicates', *Spectrochim. Acta*, **22**, 389–398 (1966).
181. V.C. Farmer, 'The Layer Silicates', in *The Infrared Spectra of Minerals*, ed. V.C. Farmer, Mineralogical

Society Monograph 4, Mineralogical Society, London, 331–363, 1974.
182. U. Johansson, R.L. Frost, W. Forsling, J.T. Kloprogge, 'Raman Spectroscopy of the Kaolinite Hydroxyls at 77 K', *Appl. Spectrosc.*, **52**, 1277–1282 (1998).
183. W.M. Butler, C.L. Angell, W.Mc. Allister, W.M. Risen, 'Far-infrared Study of Cation Motion in Dry and Solvated Mono- and Divalent Cation Containing Zeolites X and Y', *J. Phys. Chem.*, **81**, 2061–2068 (1977).
184. C. Brémard, D. Bougeard, 'Raman Scattering in Zeolites and Molecular Sieves', *Adv. Mater.*, **7**, 10–25 (1995).
185. N. Sahba, T.J. Rockett, 'Infrared Absorption Coefficients of Silica Glasses', *J. Am. Ceram. Soc.*, **75**, 209–212 (1992).
186. T. Uchino, T. Sakka, M. Iwasaki, 'Interpretation of Hydrate States of Sodium Silicate Glasses by Infrared and Raman Analysis', *J. Am. Ceram. Soc.*, **74**, 306–313 (1991).
187. A. Aronne, S. Esposito, P. Pernice, 'FTIR and DTA Study of Lanthanum Aluminosilicate Glasses', *Mater. Chem. Phys.*, **51**, 163–168 (1997).
188. A.P. Novaes de Oliveira, C. Leonelli, T. Manfredini, G.C. Pellacani, G. Ramis, M. Trombetta, G. Busca, 'Properties of Glasses Belonging to the Li_2O-ZrO_2-SiO_2 System', *Phys. Chem. Glasses*, **39**, 213 (1998).
189. D. Mazza, M. Lucco-Borlera, G. Busca, A. Delmastro, 'High-quartz Solid-solution Phases from Xerogels with Composition $2MgO \cdot 2Al_2O_3 \cdot 5SiO_2$ (μ-cordierite) and $LiO_2 \cdot Al_2O_3 \cdot nSiO_2$ ($n = 2$ to 4) (β-eucryptite): Characterization by XRD, FTIR and Surface Measurements', *J. Eur. Ceram. Soc.*, **11**, 299–308 (1993).
190. C. Ramaswamy, 'The Raman Effect in Diamond', *Ind. J. Phys.*, **5**, 97–104 (1930).
191. X.-Z. Zhao, K.A. Cerian, R. Roy, W.B. White, 'Downshift of Raman Peak in Diamond Powders', *J. Mater. Res.*, **13**, 1974–1976 (1998).
192. T. Ando, S. Inoue, M. Ishii, M. Kamo, Y. Sato, O. Yamada, T. Nakano, 'Fourier-Transform Infrared Photoacoustic Studies of Hydrogenated Diamond Surfaces', *J. Chem. Soc. Faraday Trans.*, **89**, 749–751 (1993).
193. S. Veprek, Z. Iqbal, H.R. Oswals, A.P. Webb, 'Properties of Polycrystalline Silicon Prepared by Chemical Transport in Hydrogen Plasma at Temperature Between 80 and 400 °C', *J. Phys. C*, **14**, 295–308 (1981).
194. T. Iizuka, S. Takasu, M. Tajima, T. Arai, T. Nozaki, N. Inoue, M. Watanabe, 'Determination of Conversion Factor for Infrared Measurements of Oxygen in Silicon', *J. Electrochem. Soc.*, **132**, 1707–1713 (1985).
195. T.M. Kramer, W.E. Rhine, H.K. Bowen, 'Surface Characterization of Hydrogenated Silicon Powder Made by the Laser-induced Decomposition of Silane', *Adv. Ceram. Mater.*, **3**, 244–248 (1988).
196. J.L. Regolini, J.P. Stoquert, C. Ganter, P. Siffert, 'Determination of the Conversion Factor for Infrared Measurements of Carbon in Silicon', *J. Electrochem. Soc.*, **133**, 2165–2168 (1986).
197. S. Rath, S. Sato, H. Ono, S. Niozaki, H. Morisaki, 'Evidence of a Tetragonal Structure of Germanium Nanocrystals Prepared by the Cluster-beam Deposition Technique', *Mater. Chem. Phys.*, **54**, 244–246 (1998).
198. F. Tuinstra, J.L. Koenig, 'Raman Spectrum of Graphite', *J. Chem. Phys.*, **53**, 1126–1130 (1970).
199. M.I. Nathan, J.E. Smith, K.N. Tu, 'Raman Spectra of Glassy Carbon', *J. Appl. Phys.*, **45**, 2370 (1974).
200. P.V. Huong, 'The New Carbon Structures and Their Electronic Conduction', *Spectrosc. Eur.*, **0/1**, 24–26 (1997).
201. S.L. Fang, A.M. Rao, P.C. Eklund, P. Nikolaev, A.G. Rinzler, R.E. Smalley, 'Raman Scattering Study of Coalesced Walled Carbon Nanotubes', *J. Mater. Res.*, **13**, 2405–2411 (1998).
202. T.J. Dennis, J.P. Hare, H.W. Kroto, R. Taylor, D.R.M. Walton, P.J. Hendra, 'The Vibrational Raman Spectra of C_{60} and C_{70}', *Spectrochim. Acta*, **47A**, 1289–1292 (1991).
203. R.C. Haddon, A.F. Hebrad, M.J. Rosseinsky, D.W. Murphy, S.H. Glarum, T.T.M. Plastra, A.P. Ramirez, S.J. Duclos, R.M. Fleming, T. Siegrist, R. Tycko, in *Fullerenes, Synthesis Properties and Chemistry of Large Carbon Clusters*, eds. G.S. Hammond, V.J. Kuck, ACS symp. Ser.- ACS, Washington, 78, Vol. 481, 1992.
204. J.N. Plendl, 'Characteristic Energy Absorption Spectra of Solids', in *Far-Infrared Properties of Solids*, eds. S.S. Mitra, S. Nudelman, Plenum Press, London, 387–450, 1970.
205. C. Kittel, *Introduction to Solid State Physics*, 6th edition, Wiley, New York, 1986.
206. J.A.A. Engelbrecht, J.R. Botha, 'Infrared Reflectivity Determination of Alloy Composition in Al*x*Ga1-*x*As and In*x*Ga1-*x*As Structures', *Appl. Spectrosc.*, **51**, 433–437 (1997).
207. L.E. McNeil, M. Grimsdich, R.H. French, 'Vibrational Spectroscopy of Aluminum Nitride', *J. Am. Ceram. Soc.*, **76**, 1132–1136 (1993).
208. J.P. Luomgo, 'Infrared Characterization of α- and β-Crystalline Silicon Nitride', *J. Electrochem. Soc.*, **130**, 1560–1562 (1983).
209. J.P. Luongo, 'IR Study of Amorphous Silicon Nitride Films', *Appl. Spectrosc.*, **38**, 195–199 (1984).
210. M.I. Baraton, R. Marchand, P. Quintard, 'Comparative Infrared Study of Silicon and Germanium Nitrides', *J. Mol. Struct.*, **143**, 9–12 (1986).
211. M.I. Baraton, P. Quintard, 'On the Determination of the Composition of Sialons by Infrared Spectroscopy', *J. Mol. Struct.*, **143**, 239–242 (1986).
212. M.I. Baraton, J.C. Labbe, P. Quintard, G. Roult, 'L'oxynitrure de Silicium: Si_2N_2O. I. Attribution des Absorptions du Spectre Infrarouge aux Vibrations Fondamentales', *Mater. Res. Bull.*, **20**, 1239–1250 (1985).
213. A.P. Mirgoredsky, M.I. Baraton, P. Quintard, 'Lattice Dynamics of Silicon Oxynitride, Si_2N_2O: Vibrational Spectrum, Elastic and Piezoelectric Properties', *J. Phys. Condens. Matter*, **1**, 10 052–10 066 (1989).

214. R. Geick, C.H. Perry, G. Rupprecht, 'Normal Modes in Hexagonal Boron Nitride', *Phys. Rev.*, **146**, 543–547 (1966).
215. M.I. Baraton, L. Boulanger, M. Cauchetier, V. Lorenzelli, M. Luce, T. Merle, P. Quintard, Y.H. Zhou, 'Nanometric Boron Nitride Powders: Laser Synthesis Characterization and FT/IR Surface Study', *J. Eur. Ceram. Soc.*, **13**, 371–378 (1994).
216. N. Muraki, V. Sergo, G. Pezzotti, G. Katagiri, S. Meriani, T. Nishida, 'Raman Piezo-spectroscopic Behavior of Aluminum Nitride', *Appl. Spectrosc.*, **51**, 1761–1765 (1997).
217. K. Honda, S. Yokoyama, S. Tanaka, 'Stress Dependence of Infrared Active Mode of beta-Si_3N_4 Appearing at 1020 cm^{-1} Measured by Micro FT/IR Spectroscopy', *Appl. Spectrosc.*, **52**, 1274–1276 (1998).
218. A.M. Heyns, L.C. Prinsloo, K.J. Range, M. Stassen, 'The Vibrational Spectra and Decomposition of α-calcium Nitride (α-Ca_3N_2) and Magnesium Nitride (Mg_3N_2)', *J. Solid State Chem.*, **137**, 33–41 (1998).
219. P. Krautwasser, G.M. Begun, P. Angelini, 'Raman Spectral Characterization of Silicon Carbide Nuclear Fuel Coatings', *J. Am. Ceram. Soc.*, **66**, 424–434 (1983).
220. L. Patrick, W.J. Choyke, 'Lattice Absorption Bands in SiC', *Phys. Rev.*, **123**, 813–815 (1961).
221. G. Ramis, M. Cauchetier, P. Quintard, G. Busca, V. Lorenzelli, *J. Am. Ceram. Soc.*, **72**, 1692–1697 (1989).
222. K. Shirai, S. Emura, 'Lattice Vibrations of Boron Carbide', *J. Solid State Chem.*, **133**, 93–96 (1997).
223. A. Hadni, G. Morlot, F. Brehat, 'Spectres d'Absorption dans l'Infrarouge Lointain de Sept Halogenures Anhydres et Hydrates: $BaCl_2$, $BaBr_2$, $SrBr_2$, $SrCl_2$, $PbCl_2$, $PbBr_2$ et NaI', *Spectrochim. Acta*, **24A**, 1167–1175 (1968).
224. C. Karr, J.J. Kovach, 'Far-infrared Spectroscopy of Minerals and Inorganics', *Appl. Spectrosc.*, **23**, 219–223 (1969).
225. J. Shamir, S. Larach, 'Concentration Dependence of Raman Scattering from Mixed Zinc Cadium Sulfide Polycrystallites', *Spectrochim. Acta*, **27A**, 2105–2108 (1971).
226. A. Mitsuishi, H. Yoshinaga, S. Fujita, *J. Phys. Soc. Jpn.*, **13**, 1235 (1958).
227. D.S. Knight, W.B. White, 'Raman Spectroscopic Study of the Rare Earth Sesquisulfides', *Spectrochim. Acta*, **46A**, 381–387 (1990).
228. D.S. Knight, W.B. White, 'Raman Line Widths in Rare-earth Sesquisulfide Solid Solutions', *J. Am. Ceram. Soc.*, **79**, 1394–1396 (1996).
229. G.L. Frey, R. Tenne, M.J. Matthews, M.S. Dresselhaus, G. Dresselhaus, 'Optical Properties of MS2 (M=Mo,W) Inorganic Fullerene-like and Nanotube Materials – Optical Absorption and Resonance Raman Measurements', *J. Mater. Res.*, **13**, 2412–2417 (1998).
230. Wang, L.A. Haskin, E. Cortez, 'Prototype Raman Spectroscopic Sensor for in situ Mineral Characterization on Planetary Surfaces', *Appl. Spectrosc.*, **52**, 477–487 (1998).
231. H.G.M. Edwards, J.K.F. Tait, 'FT-Raman Spectroscopic Study of Decorated Stained Glass', *Appl. Spectrosc.*, **52**, 679–682 (1998).
232. I.M. Bell, R.J.H. Clark, P.J. Gibbs, 'Raman Spectroscopic Library of Natural and Synthetic Pigments (pre 1850 A.D.)', *Spectrochim. Acta*, **A53**, 2159–2179 (1997).
233. H.G.M. Edwards, L. Drummond, J. Russ, 'FT-Raman Spectroscopic Study of Pigments in Native American Indian Rock Art: Seminole Canyon', *Spectrochim. Acta*, **A54**, 1849–1856 (1998).
234. P. Mirti, L. Appolonia, A. Casoli, R.P. Ferrari, E. Laurenti, A. Amisano Canesi, G. Chiari, 'Spectrochemical and Structural Studies on a Raman Sample of Egyptian Blue', *Spectrochim. Acta*, **51A**, 437–446 (1995).
235. J.H. Robin Clark, M. Lucia Curri, C. Laganara, 'Raman Microscopy: the Identification of Lapis Lazuli on Medieval Pottery Fragments from the South of Italy', *Spectrochim. Acta*, **A53**, 597–603 (1997).
236. A.H. Delgado, R.M. Paroli, J.J. Beaudoin, 'Comparison of IR Techniques for the Characterization of Construction Cement Minerals and Hydrated Products', *Spectrochim. Acta*, **A50**, 970–976 (1996).
237. G. Godefroy, B. Jannot, in *Diffusionless Phase Transitions and Related structures in Oxides*, ed. C. Boulesteix, Trans Tech Pub., Zurich, 81–132, 1992.
238. G. Suresh, R. Ratheesh, V.U. Nayar, M. Ichikawa, 'Raman Spectroscopic Investigation on the Phase Transitions in $Rb_3H(SeO_4)_2$ and $Rb_3D(SeO_4)_2$ Single Crystals', *Spectrochim. Acta*, **52A**, 465–470 (1996).
239. G. Suresh, R. Ratheesh, V.U. Nayar, M. Ichikawa, G. Kresztury, 'IR and Polarized Raman Spectra of a Single Crystal $Rb_3H(SeO_4)_2$ in the Ferroelastic Phase', *Spectrochim. Acta*, **51A**, 429–435 (1995).
240. V. Vivekanandan, S. Selvasekarapandian, P. Kolandaivel, M.T. Sebastian, S. Suma, 'Raman and FT/IR Spectroscopic Characterization of Flux Grown $KTiOPO_4$ and $KRbTiOPO_4$ Non-linear Optical Crystals', *Mater. Chem. Phys.*, **49**, 204–210 (1997).
241. N. Guilhaumou, P. Dumas, G.L. Carr, G.P. Williams, 'Synchrotron Infrared Microspectrometry Applied to Petrography in Micrometer-scale Range: Fluid Chemical Analysis and Mapping', *Appl. Spectrosc.*, **52**, 1029–1034 (1998).
242. G. Hazel, F. Bucholtz, I.D. Aggarwal, G. Nau, K. Ewing, 'Multivariate Analysis of mid-IR FT/IR Spectra of Hydrocarbon-contaminated Wet Soils', *Appl. Spectrosc.*, **51**, 984–989 (1997).
243. C. Remy, B. Reynard, M. Madon, 'Raman Spectroscopic Investigations of Dicalcium Silicate: Polymorphs and High-temperature Phase Transformations', *J. Am. Ceram. Soc.*, **80**, 413–423 (1997).
244. V. Swamy, L.S. Dubrovisnky, F. Tutti, 'High-temperature Raman Spectra and Thermal Expansion of

Wollastonite', *J. Am. Ceram. Soc.*, **80**, 2237–2247 (1997).
245. K. Fukuda, R. Kitoh, I. Awai, 'Far-infrared Reflection Spectra of Dielectric Ceramics for Microwave Applications', *J. Am. Ceram. Soc.*, **77**, 149–154 (1994).
246. G. Busca, 'Vibrational Spectroscopy of Adsorbed Molecules and Surface Species on Metal Oxides', *Catal. Today*, **27**, 323–541 (1996).
247. C.G. Kontoyannis, N.C. Bouropoulos, P.G. Koutsoukos, 'Urinary Stone Layer Analysis of Mineral Components by Raman Spectroscopy, IR Spectroscopy and X-ray Powder Diffraction: a Comparative Study', *Appl. Spectrosc.*, **51**, 1205–1209 (1997).
248. H. Tsuda, J. Arias, B. Leon, J. Arends, 'Necessary Precautions in the Raman Analysis of Calcium Phosphate Minerals Using 1.06 μm YAG Laser Excitation', *Appl. Spectrosc.*, **52**, 1122–1126 (1998).

Kinetic Determinations: Introduction

Horacio A. Mottola
Oklahoma State University, Stillwater, USA

1 Equilibrium and Kinetic Determinations **1**
1.1 Comparison of Thermodynamic and Dynamic Measurements 1
1.2 Approaches to Dynamic Measurement of Concentration 2
1.3 Equilibrium and Pseudo-equilibrium Measurements to Support Kinetic-based Determinations 4
2 Historical Perspective **4**
2.1 Early Tests and Publications 4
2.2 From Timers and Watches to Semi-automated and Automated Reaction Rate Methodology 6
2.3 From Determination of Inorganic Species to the Determination of Organic and Biochemicals 7
3 Kinetics in Analytical Chemistry: Beyond Determinations **8**
3.1 Chemical Reactions and Physical Processes in Analytical Chemistry 8
3.2 Some Areas of Analytical Chemistry in which Kinetics Play a Direct or Indirect Role 8
Further Reading **9**
Books and Monographs 9
Reviews and Other Useful Publications 9
References **10**

This introduction opens by discussing the distinction between kinetic and thermodynamic determinations. After defining equilibrium, pseudo-equilibrium, and dynamic measurements, a historical perspective is given that covers early tests and publications to the present overall scope of kinetic determinations. In closing, a listing of selected bibliography is presented; this follows a short narrative emphasizing that kinetics goes beyond determinations in analytical chemistry.

1 EQUILIBRIUM AND KINETIC DETERMINATIONS

The two pillars of the physical sciences are their kinetic and equilibrium components.

1.1 Comparison of Thermodynamic and Dynamic Measurements

All physical and chemical processes, once triggered, evolve to equilibrium by a pre-equilibrium evolution defined as the kinetics of the processes. Practically, however, nothing in this world is perfectly static. Physical and chemical processes do not abide; they occur. Physics and chemistry are scientific disciplines focused on changes that take place in time. Changes occur even in systems considered to be at equilibrium. For all practical purposes, however, the chemical composition of a system and its physical properties remain constant once the system has reached equilibrium, but both change with time during the dynamic evolution to equilibrium.

The rate of transient physical processes as well as the rate of chemical changes can both be used for quantitative analytical purposes. Kinetic determinations based on fluorescence or phosphorescence, for instance, involve primarily physical processes. The bulk of analytical applications based on measurements during transient processes, however, exploit chemical changes. Consequently, this section of this *Encyclopedia* focuses, primarily, on chemical kinetics without neglecting physical processes. Experimentally, measurements of the system composition are possible while reaching equilibrium (kinetic measurements) or with the system at equilibrium (thermodynamic measurements). Consequently, the contemporary analytical chemist can choose to make measurements (e.g. determination of the concentration of a particular analyte) during evolution to equilibrium or at equilibrium. Essentially, kinetic determinations are based on measurements while the composition of the system is changing with time and without disturbing the kinetic process. It is possible, however, to freeze the composition of the system at any given time during the evolution to equilibrium (quenching a given reaction), and to make the 'thermodynamic measurement' at this 'forced equilibrium' situation. Then, relate the composition to the known kinetic behavior of the analyte to arrive at its original concentration in the sample. This practice is seldom used today, but was quite common some years ago, and although technically the measurement is a thermodynamic one, the methodology is classified as resulting in a kinetic determination.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

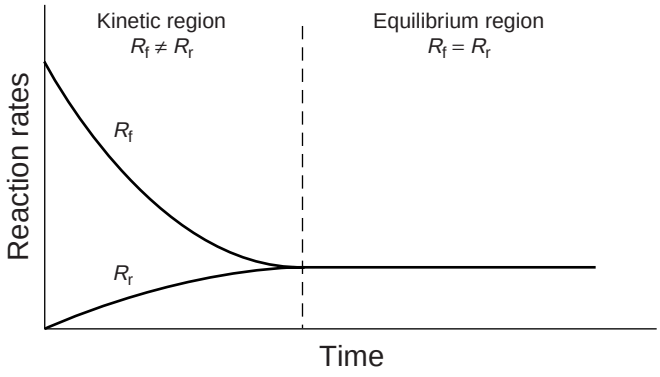

Figure 1 Typical rate–time profile. R_f, rate of forward change; R_r, rate of reversed change. Kinetic (dynamic) measurements are performed in the kinetic region, thermodynamic measurements in the equilibrium region.

The simple formulation, Equation (1):

$$\text{REACTANTS} \longleftrightarrow \text{PRODUCTS} \tag{1}$$

and the corresponding kinetic symbolism, (Equation 2), in which t = time:

$$\frac{dC_{\text{PRODUCTS}}}{dt} = -\frac{dC_{\text{REACTANTS}}}{dt} = \text{rate of reaction} \tag{2}$$

allows the time regions which define kinetic and thermodynamic measurements to be visualized. This is shown in Figure 1.

The rate profile shown in Figure 1 is exhibited by all chemical reactions and physical processes regardless of their complexity. Although the didactic value of this figure is obvious for illustration of the two measurement approaches available for analytical determinations, direct measurement of reaction rates is not practical. What the chemist usually measures directly is some electrical signal, S, related to concentration. In some cases (e.g. conductance or amperometric measurements), S is linearly related to concentration (Equation 3):

$$S = \S C_R \tag{3}$$

in which § is the proportionality constant or transfer function in electrical units per concentration units and C_R represents the concentration of a given reactant at a given time. In some instances, S is not a linear function of C_R (e.g. optical absorption or potentiometric measurements), and the direct instrument response can be written as Equation (4):

$$S = fC_R \tag{4}$$

where f is an arbitrary function. The mathematics is more complicated, generally resulting in nonlinear calibrations.

Table 1 Most common measurement approaches to extract concentration and rate information while the system is evolving to equilibrium

Single point measurements:	Initial rate: derivative, fixed and variable time
	Derivative or $\Delta S/\Delta t$ at $t \neq 0$ (t at which more than 10% reaction has been completed)
Multi point measurements:	Derivative, $\Delta S/\Delta t$, and regression methods
Methods based on kinetic plots:	Method of the tangents

In any event, C_R can always be obtained from S. With a knowledge of concentration and its change with time, reaction rates can then be calculated. Physical changes in the kinetic region are also followed by the change of an electrical signal from which rate information is extracted.

1.2 Approaches to Dynamic Measurement of Concentration

Table 1 summarizes the most important approaches for kinetic determinations based on concentration estimates obtained while the system is evolving to the equilibrium position. An understanding of the operational characteristics of kinetic-based determinations requires a background in the mathematical formulation and manipulation of rate expressions. Most of this background is addressed in detail in several of the articles within this section. In general, however, such a background can be kept simple considering that chemical reactions employed in virtually all kinetic determinations are bimolecular. The analyte species, A, is one of the reactants, and the other reactant, R, is present in large excess rendering pseudo first-order conditions under which:

$$C_{R,\text{ at any time } t} = C_{A,\text{ initial}}e^{-kt} \tag{5}$$

where k = rate proportionality constant (rate coefficient). Equation (5) implies that the initial concentration of analyte can be obtained with a knowledge of S at any given t, k, and § or f, provided the mechanism of the reaction remains constant.

Although the conversion of C_A from S is built into the computer software used with contemporary instrumentation, a knowledge of the value of k often requires independent experimentation. There are, however, experimental/instrumental means that circumvent the need for a knowledge of the value of k. For example, if the value of S is collected only during the initial 1 to 2%, in some cases even 10%,[1] completion of the total process, conditions (e.g. the concentrations of A and R)

remain virtually unchanged, the reversed process can be ignored because $C_{\text{PRODUCT}} \approx 0$, and a pseudo-zero-order process is in operation such as Equation (6):

$$-\left(\frac{dC_{A,\text{ at any time } t}}{dt}\right) \approx k\, C_{A,\text{ initial}} C_{R,\text{ initial}} \approx \text{initial rate} \approx \text{constant} \qquad (6)$$

Equation (6) provides the basis of a very common approach to dynamic measurement of concentration, the initial rate method, since if the initial rate is plotted against $C_{A,\text{ initial}}$ should yield a straight line useful as a calibration plot for the determination of $C_{A,\text{ initial}}$. Three approaches are used to extract initial rate information, they are:

1. the derivative or slope approach, in which electronic differentiation provides dS/dt;
2. the fixed time or constant time approach;
3. the variable time approach.

The fixed time and variable time approaches result from the use of the integral forms of the rate equation, but under conditions such that Equation (6) is approximately valid. The fixed time method involves measurement of a small signal change, ΔS, at a finite but short Δt close to $t = 0$. In the variable time approach, Δt is measured for a finite but small ΔC_R or ΔC_A close to $C_{R,\text{ initial}}$ or $C_{A,\text{ initial}}$. In both cases, the ratio $\Delta S/\Delta t$ gives an approximate estimate of the initial rate.

Initial rate measurements are often recommended because the back reaction does not contribute appreciably to the overall rate, and complications from side reactions are less possible at the beginning of any reaction.

To determine the initial concentration of a desired species, monitoring of the concentration of at least one of the reactants (or products) as a function of time is needed. Continuous measurement of the change in chemical composition is possible by physical, but not by chemical methods, and the observable rates are limited only by the response time of the instruments. With well behaved kinetics and the concentration–time profile obtained by continuous monitoring, two other approaches based on dynamic measurement of concentration can be identified:

1. the derivative approach, and
2. methods based on kinetic plots such as the method of the tangents.

The derivative approach considers that at any time t the value of dS/dt is indicative of the instantaneous rate. If most of the concentration–rate profile is available, relaying on a single derivative (single point measurement) is unwise. This is particularly true now that we enjoy the brute force of the digital computer. The computer makes multipoint-based measurements possible, and this offers two main advantages:

1. less experimental work is needed to seek discriminating variables, and
2. the use of a large number of data points reduces errors.

Table 2 Summary of most common kinetic-based methods

Catalytic methods	• Enzymatic methods employing soluble enzyme preparations • Nonenzymatic catalytic methods
Uncatalyzed reaction-rate methods	• Single-component determinations • Multicomponent determinations (differential reaction rate methods)
Luminescence-based methods	
Determinations in heterogeneous systems	• Kinetic methods based on electrode reactions • Enzymatic methods employing immobilized enzyme preparations

Which approach is used, of those listed in Table 1, primarily depends on available instrumental and/or computational capabilities, and on the type of kinetic method being applied. Table 2 summarizes the main types of methods used in kinetic-based determinations. Except for special cases, all these methods use the measurement approaches listed in Table 1. The measured signal, proportional to concentration, is in all these cases used to derive rate information. The converse approach of estimating the equilibrium concentration (signal) from signals measured under dynamic conditions is less often used. This approach characterizes the so-called predictive methods.[2] Predictive methods are of relatively recent application for kinetic-based determination, but their foundation can be traced back to considerations advanced in 1926 by Christie Smith.[3]

Some approaches to mixing[4] (e.g. centrifugal mixing or continuous-flow mixing) also provide unique means for measurements under dynamic conditions. Unsegmented continuous-flow mixing of the type encountered in the approach commonly known as 'flow injection analysis' provides, in its simplest and most commonly used form, fixed-time measurements under dynamic conditions.[5] Programmed continuous-stopped-continuous-flow procedures in unsegmented systems provide another avenue for signal measurement under dynamic conditions[5] of utility for kinetic-based determinations.

1.3 Equilibrium and Pseudo-equilibrium Measurements to Support Kinetic-based Determinations

The progress to equilibrium of any process can be terminated by quenching the reaction. Quenching is possible by several means, including rapid cooling, precipitation of one reactant, addition of free radical scavengers, or addition of a complexing agent or other chemical species that renders inactive one of the reactants. Quenching by adding an acid or a base can be used, for example, to stop most enzyme-catalyzed reactions. Since the progress to equilibrium has been interrupted, for all practical purposes Equation (7) is valid:

$$-\frac{dC_{REACTANTS}}{dt} = \frac{dC_{PRODUCTS}}{dt} = 0 \qquad (7)$$

and the system, composition-wise, is at 'equilibrium', and therefore the measurement of S is not performed under dynamic conditions. The corresponding values of S, however, can be used to calculate concentrations at given times and apply, for example, the fixed-time or variable-time approach to effect a kinetic-based determination. In such cases one of the times or one of the concentration values may be equal to zero. The speed of today's signal detection/manipulation devices has practically made obsolete quenching for kinetic-based determinations, but quenching is always an approach available to the analytical chemist.

A unique steady-state situation (pseudo-equilibrium condition) is imposed on a system by stat procedures.[6] These procedures are an example of kinetic-based determinations in which a solution containing a chemical species capable of counteracting a chemical change in the system is added to the system. If the rate of delivery matches the rate of change in the system, a steady state situation is approached (stat condition) in which $dS/dt = 0$. The rate for restoring the value of S serves to deduce quantitative information for determinative purposes. This approach has its origins in the determination of enzyme activity with the maintenance of pH at about a constant value (generally the optimum pH for the enzyme) by a pH-stat procedure, but its report in the literature for other type of kinetic-based determinations is scarce.

2 HISTORICAL PERSPECTIVE

Difficult as it is to establish historical priorities, events that left a historical mark – e.g. journal papers and books – can be recognized.

Historically, equilibrium-based determinations predated kinetic-based ones, and they dominated the arena of chemical analysis for a long period of time. During this dominance, kinetics was considered mainly in the context of undesirable effects (e.g. the sluggishness of attainment of some end-points in redox titrimetry). In chemical analysis, the presence of kinetic concepts was very limited in those years, although an outstanding example of positive consideration was present in the recognition that the rate of crystal growth greatly influences the filterability and purity of a precipitate. Precipitation from homogeneous solution,[7] an ingenious application of chemical kinetics, seems to date back to 1858 with the work of Chancel[8] on the separation of iron and aluminum employing thiosulfate.

Difficult as it is to establish historical priorities, a few examples can be cited and are included in the following section.

2.1 Early Tests and Publications

Nonenzymatic catalysis seems to provide the first recorded analytical development of a kinetic-based method. As early as 1876, Guyard (who was also known by the alias Hugo Hamm) discussed the catalytic effect of vanadium on the perchlorate ion oxidation of aniline to yield aniline black and proposed the use of this reaction to detect the presence of vanadium.[9] A decade later, Osmond[10] described a semiquantitative estimation of vanadium by considering that the amount of aniline black formed after a given, fixed, time is proportional to the amount of vanadium present in the sample. The paper was written by Osmond after Georges Witz' death, but was inspired by Witz' studies on certain chemical modifications of cotton, and Witz is listed as coauthor. Oxycellulose threads were dipped into vanadium-containing solutions and subsequently into an aniline–chlorate mixture for a fixed time. The practical minimum detectable quantity can be estimated as 0.000005 mg of vanadium, but the volume of sample used, unfortunately, is not indicated in the publication. This paper is one of the first to report a nonenzymatic quantitative estimation based on a catalytic procedure, and using a fixed time procedure. Osmond also indicates in the same publication that the method was used to recognize the presence or absence of vanadium in samples of different natural mineral waters. In 1881, a few years earlier than the publication of the paper with Witz and Osmond authorship, William Roberts[11] published what can be considered the first application of the variable-time procedure. Roberts was a physician in the Manchester Royal Infirmary and a professor of clinical medicine at Owens College. The procedure proposed by Roberts was inspired by the work of Brown and Heron[12] on the transformation of starch by malt infusions, and the work of Johan Gustaf Kjeldahl on the measurement

of the comparative activity of malt infusions and of saliva at the Carlsberg Institute for research on brewing and the beer industry. Contributions from clinical medicine and industrial enzymology have undoubtedly left an indelible mark in the early history of kinetic-based determinations. Almost half a century passed until catalytic-based determinations were cited again in the records of chemical literature. This time Sandell and Kolthoff[13] introduced the determination of iodide based on its catalytic action on the cerium(IV) oxidation of arsenic(III) by a variable-time procedure. Almost simultaneously, in Hungary, Szebellédy and Ajtai[14] reported a catalytic-based determination of as little as 0.6 ng of vanadium. A large number of determinations based on catalysis have been described since then and many found a place in the first monograph on kinetic methods of determination authored by Yatsimirskii and published in 1966[15] (see Figure 2). A monograph by Arkosi[16] on 'catalytic analysis' preceded by about 10 years the book by Yatsimirskii. Yatsimirskii's book, however, appeared at a time when instrumentation was making measurements under dynamic conditions competitive, and consequently had a significant impact on the development of the topic.

As early as 1921, Clarens[17] applied the laws of chemical kinetics to the dissolved oxygen oxidation of tannins and proposed a fractional determination of tannins in general and in wines in particular. Clarens paper seems to be one of the earliest uses of differential rates for the in situ determination of more than one species in solution. It also points to a pioneering of a kinetic-based approach utilizing uncatalyzed reaction rates. In contrast with catalytic methods, simultaneous determination of more than one chemical species (differential rate methods) are not necessarily aimed at determining low concentrations and most of the reactions employed are uncatalyzed. Relatively more recently, a paper by

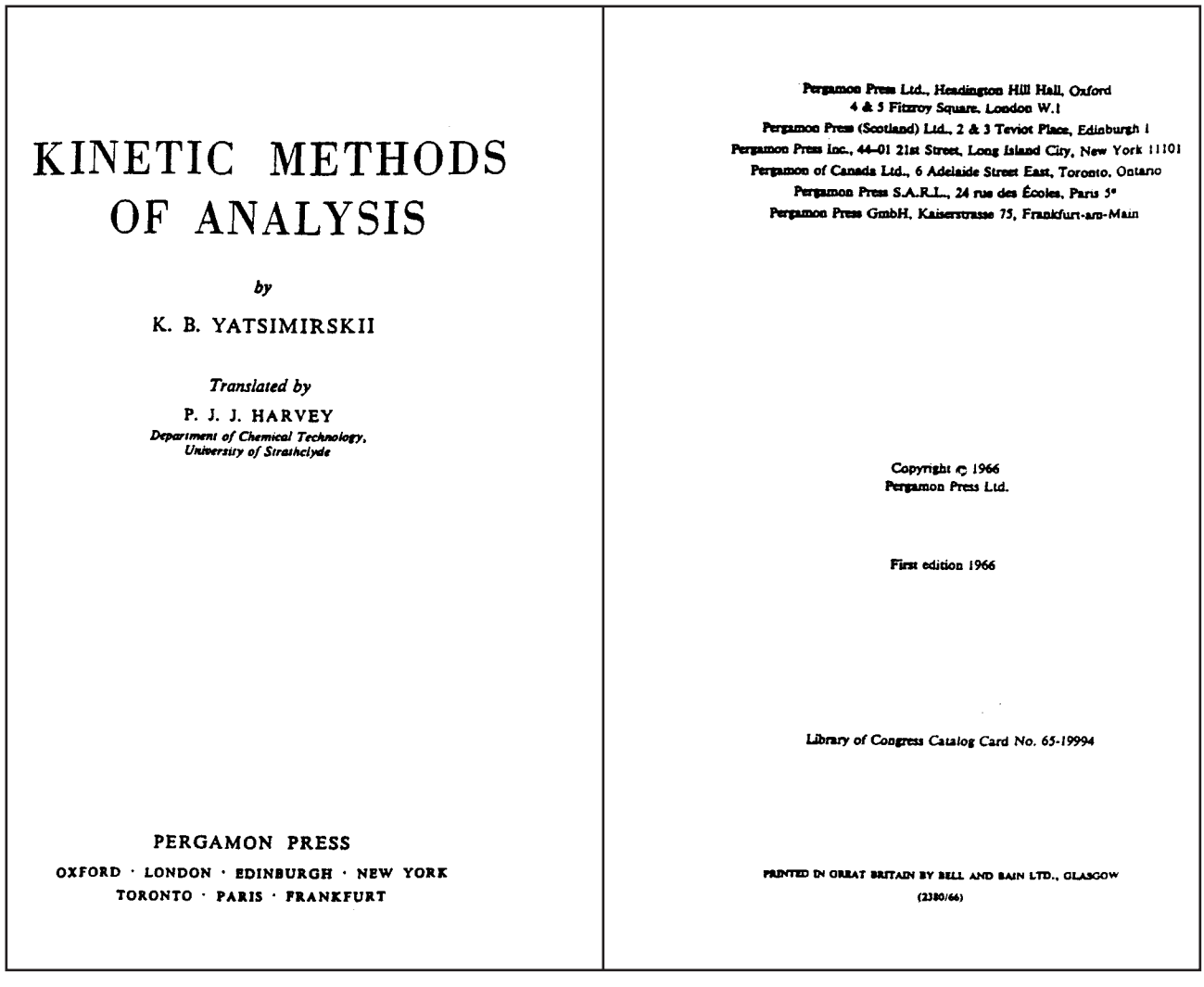

KINETIC METHODS OF ANALYSIS

by

K. B. YATSIMIRSKII

Translated by

P. J. J. HARVEY

Department of Chemical Technology, University of Strathclyde

PERGAMON PRESS

OXFORD · LONDON · EDINBURGH · NEW YORK

TORONTO · PARIS · FRANKFURT

Pergamon Press Ltd., Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1
Pergamon Press (Scotland) Ltd., 2 & 3 Teviot Place, Edinburgh 1
Pergamon Press Inc., 44-01 21st Street, Long Island City, New York 11101
Pergamon of Canada Ltd., 6 Adelaide Street East, Toronto, Ontario
Pergamon Press S.A.R.L., 24 rue des Écoles, Paris 5e
Pergamon Press GmbH, Kaiserstrasse 75, Frankfurt-am-Main

Copyright © 1966
Pergamon Press Ltd.

First edition 1966

Library of Congress Catalog Card No. 65-19994

PRINTED IN GREAT BRITAIN BY BELL AND BAIN LTD., GLASGOW
(2380/66)

Figure 2 Front and reverse of the title page of Yatsimirskii's book 'Kinetic Methods of Analysis'.

Kolthoff and Lee[18] demonstrated that kinetics, via differential reaction rate methods, can give access to medium and major components in a mixture, and thus complementing the accessibility to very low concentrations afforded by catalytic methods. Although the procedure introduced by Kolthoff and Lee, known today as the single-point method, has found infrequent use, it initiated a period extended through the 1950s and 1960s in which differential rate methods drew attention, mainly for the determination of organic species. Lee's chapter in the series 'Organic Analysis'[19] and the book by Fritz and Hammond[20] on quantitative organic analysis emphasize the impact of simultaneous determination of more than one organic chemical in a mixture exploiting rate differences. The chapter by Lee[19] offered the interesting forecast that enzymes would find a place in analytical chemistry as purity of preparations and commercial availability would make them attractive in kinetic-based methodology. Time has proved Lee's forecast as true. Stetter,[21] however, provided one of the earliest documentations of the use of impure enzyme preparations in a variety of analytical procedures.

2.2 From Timers and Watches to Semi-automated and Automated Reaction Rate Methodology

Mixing and timing have been the two experimental operations critical for the development of kinetic methods. Manual mixing (Figure 3) and watches provided the earliest means for implementing such experimental operations, and their crude performance did little to favor the use of kinetics in analytical determinations.

Resistance to accept kinetic-based determinations persisted throughout the 1950s; the main arguments being that time was an elusive variable to measure and/or that concentration was difficult to ascertain at given times if the system was evolving to equilibrium. The 1960s, however, brought accessibility to electronic devices, particularly the use of operational amplifiers, that demonstrated as possible the accurate and precise measurement of signals during evolution to equilibrium: both time and concentration could be reliably known simultaneously. The road to automation of reaction-rate methodology was opened in the 1960s (see Mottola, Kinetic Aspects of Analytical Chemistry, 8–11) and worth mentioning for their pioneering nature are the papers by Malmstadt and Hicks[23] and Blaedel and Hicks.[24] Undoubtedly, the use of strip-chart recorders played an important role in the late 1950s and early 1960s, allowing the recording of signal change with time in relatively slow chemical changes. Figure 4, taken from Yatsimirskii's book illustrates the use of such read-out devices.

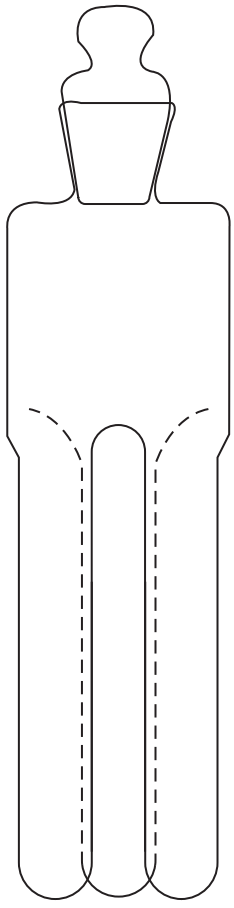

Figure 3 A container designed by L.I. Budarin to aid manual mixing of reactants. The container had three compartments for three different solutions, and could be immersed into a water bath for temperature control. When the desired temperature was reached, the container was turned upside down and the contents energetically mixed. Time measurement was started when the container was turned upside down. This type of container was used by Bontchev et al.,[22] for example, to determine silver in human saliva by a catalytic method. (Reproduced by permission of Pergamon Press from K.B. Yatsimirskii, 'Kinetic Methods of Analysis', Oxford, 30, 1966.)

Malmstadt and Hicks[23] reported on how to assemble, from commercially available units and interface it with a controller, an instrument that greatly facilitated the rate-based enzymatic determination of glucose in clinical samples. Blaedel and Hicks[24] introduced the first unsegmented continuous-flow sample/reagent(s) processing system for the continuous measurement of the rate of an enzyme-catalyzed reaction.

The logical next historical development in kinetic-based determinations was a spin-off from electronics: the exploitation of computers for feedback/feedforward control of instrumentation, signal collection and manipulation, and data treatment (e.g. application of multipoint

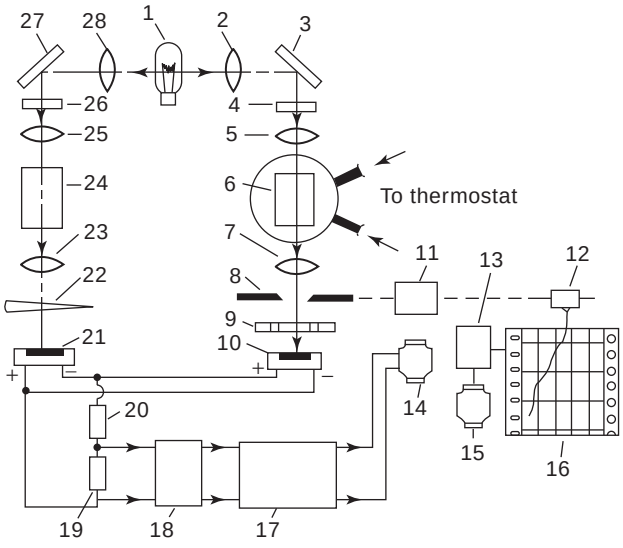

Figure 4 Block diagram of an instrument for automatic recording of absorbance change with time. 1: Lamp. 2, 28: Condensing lenses. 3, 27: Mirrors. 4, 26: Light filters. 5, 7, 23, 25: Lenses. 6, 24: Cells (cell 6, thermostatically controlled). 8: Diaphragm. 9: Stop. 10, 21: Photodetectors. 11, 13: Reductors. 12: Recorder pen. 14: Reversing motor. 15: Synchronous motor. 16: Recorder chart. 17: Amplifier. 18: Vibrotransformer. 19, 20: Resistances. 22: Circular gray stop. (Reproduced by permission of Pergamon Press from K.B. Yatsimirskii, 'Kinetic Methods of Analysis', Oxford, 27, 1966.)

regression, Kalman filtering, predictive approaches, and simulations *to name a few, and which are covered in other sections*). These, however, take us too close to today's developments and away from the historical roots that pertain to this article. The interfacing of a general purpose digital computer for on-line processing of reaction rate data, for example, dates back to the late 1960s.[25]

2.3 From Determination of Inorganic Species to the Determination of Organic and Biochemicals

The developments in the field of kinetic-based determinations have paralleled the evolution experienced by chemical analysis throughout the years. During the years in which thermodynamic-based measurements dominated the analytical subdiscipline of chemistry, the attention was focused on what can be called 'mineral analysis'. In those years then, the focus was on the determination of inorganic species. Catalytic determinations, which opened the road to measurements under dynamic conditions, did not escape such an influence, and most of the early catalytic determinations were of inorganic species. Therefore, is not surprising to find a few catalytic determinations cited as part of 'inorganic microchemistry' in some early reviews.[26] More than half of the citations in these reviews, however, reported the use of organic reagents for equilibrium-based measurements aimed at the determination of inorganic species, a 'hot' topic at the time. In the middle to late 1940s, the need for training inorganic chemistry for analytical chemists became more widely recognized. This is perhaps best summarized in the following words lifted from the Preface of a timely book on functional analysis by Siggia:[27]

> I recognize the great need which exists in our colleges and universities for formal training in the field of quantitative organic analysis, and I know that very few schools offer this training ... chemists are needed who are trained to do analytical work with compounds of this type. It is my belief that quantitative organic analysis should be included in the regular analytical courses that are usually entirely inorganic.

These were the times when elemental analysis data was beginning to be supplemented with information extracted from functional group analysis based on acid–base and redox titrimetry as well as on the measurement of gases consumed or liberated by reactions of analyte and reagents. In parallel, instrumental techniques such as chromatography, electrophoresis, polarimetry, light absorption, and polarography – to name a few making an impact at that time – were helping to blur the inorganic prevalence. Kinetic-based analytical methods backed the trend in the 1950s and 1960s which can be considered to be triggered by Kolthoff and Lee's paper introducing the single point method.[18] Statements such as those by Fritz and Hammond[20] in their book on quantitative organic analysis (page 167):

> It is the opinion of the authors that the analyses described in which the differential kinetic procedure was used could not have been solved by any other method of comparable simplicity.

as well as that of Lee forecasting the impact of enzymes on analytical chemistry,[19] accurately framed the position of kinetic-based determinations paralleling the evolution of analytical chemistry itself (page 242):

> There is no doubt that methods based on catalysis by enzymes will increase in number and practicability as advances are made in the study of these substances and as more enzymes of known activity and purity become commercially available.

Transition from inorganic to organic and to biochemical analysis was paved at that time and kinetic-based measurements played an important role in such transition.

3 KINETICS IN ANALYTICAL CHEMISTRY: BEYOND DETERMINATIONS

Many analytical techniques or approaches, if scrutinized closely, are perceived to be of kinetic nature or to involve key kinetic components.

Chemical reactions in solution provided the early foundation and constitute an important portion of the subdiscipline we know today as analytical chemistry. Equally important, however, is the realization that many analytical methods and techniques are solely based on the measurement of a physical property of the system or chemical species of interest. In either case – whether a reaction-based or a physically-based measurement – measurements under dynamic conditions can not only generate information useful for determination of chemical species, but also for the fundamental understanding of the process(es) involved. In the realm of analytical chemistry, the relevance of kinetics goes beyond determinations.

3.1 Chemical Reactions and Physical Processes in Analytical Chemistry

The bulk of analytical chemistry based on chemical reactions focus on solution chemistry, particularly with water as solvent. Traditionally, four basic types of reactions have been used, and they still continue to be used, in both equilibrium-based and kinetic-based methods. They are:

1. precipitation (ion-exchange reactions),
2. acid–base (proton-exchange reactions),
3. redox (electron-exchange reactions), and
4. complexation (ligand-exchange reactions).

In kinetic-based methods, proton transfer and precipitation reactions have had very limited use. On the other hand, redox reactions dominated applications based on nonenzymatic catalytic determinations, and ligand exchange reactions have been frequently used in the development of differential rate methods.

Although not as easily recognized, rates of change of physical properties of a system play important roles in contemporary analytical chemistry. The rate of mass transfer is a case at hand. Diffusion and convection as time-dependent processes, for example, are particularly critical in electroanalytical chemistry, separations, and continuous-flow procedures. In the field of spectroscopy, excitation–absorption–emission are all rate driven processes critical in atomic as well as molecular spectroscopy.

3.2 Some Areas of Analytical Chemistry in which Kinetics Play a Direct or Indirect Role

Although the theme of this Encyclopedia is focused on determinations and instrumentation, a few words about other roles of kinetics in analytical chemistry are in order. Practically, in all areas of interest to analytical chemistry, kinetics play an important direct or indirect role. This is because time-dependent chemical and physical processes are ingrown into most aspects of analytical chemistry. This is true from very classical methodologies (e.g. gravimetric analysis) to rather recent tools (e.g. glow discharge optical emission spectroscopy and cavity ring-down laser absorption spectroscopy). The theoretical aspects of precipitation are derived from theories on nucleation and crystal growth.[28] Kinetics is of interest in many separation approaches such as solvent extraction (for example in metal ion separation[29]), ion exchange,[30] and chromatography in general.[31] Both the rate of electron transfer at the electrode/solution interface and chemical kinetics in and in the vicinity of the double layer are of paramount importance in electroanalytical chemistry.[32]

Direct injection enthalpimetry is one of the thermal methods in which kinetics is most obvious since it is based on monitoring the change in temperature with time.[33] Other thermal methods reveal kinetic in their theoretical foundations such as it is the case in differential thermal analysis.[34] Continuous-flow sample/reagent(s) processing for the so-called 'wet chemistry' are all approaches plentiful in kinetics.[5,35]

The enumeration of analytical approaches rich in kinetics becomes almost endless, including for example typical spectroscopic (e.g. nuclear magnetic resonance) and non-spectroscopic methods (e.g. mass spectrometry), but before calling to a close this section a mentioning of some well established spectroscopic approaches (e.g. absorbance-based measurements and luminescence-based methods) as well as some not so popularized as yet (e.g. glow discharge emission and cavity ring-down absorptiometric measurements) is pertinent. The kinetic nature of fluorescence and phosphorescence is well established,[36] and even one of the distinctive aspects that distinguish phosphorescence from fluorescence is rate related, i.e. the decay periods are much longer in phosphorescence because relaxation from the triplet to the ground state singlet is forbidden by spin symmetry.[37]

Bioluminescence is a special case of chemiluminescence in which the chemical reaction responsible for photon emission occurs or can occur in nature. Both, however, are observed when an electronically excited product is produced that either luminesces or transfers its excess energy to another molecule that then luminesces. Kinetically, such situations result in series processes

leading to transient signals and which can be symbolized as Equations (8–11):

$$A + B \longrightarrow C^* \quad (8)$$

$$C^* \longrightarrow C + h\nu \quad (9)$$

or

$$C^* + D \longrightarrow D^* + C \quad (10)$$

$$D^* \longrightarrow D + h\nu \quad \text{or} \quad D' + h\nu \quad (11)$$

A free energy change of at least 44 kcal mol^{-1} is required for emission of light in the visible region of the spectrum. This is usually provided by redox reactions involving oxidants such as O_2 and H_2O_2 and cyclic hydrazides (e.g. luminol or acridine derivatives).[38] Temporal resolution of events in the picosecond and even femtosecond range is of interest in analytical detection, particularly in time-resolved spectroscopic measurements.[39]

Kinetics is not only prominent in well established spectroscopic techniques. Rather recent additions to the spectroscopic tool chest are successful because of critical kinetic components, or because the measurement approach itself is kinetic in nature. In glow discharge optical emission spectroscopy,[40] for example, signals are directly related to the supply rate of element into the plasma. In the so-called cavity ring-down spectroscopy,[41] the measurement of photon absorption is based on a time measurement: the time for the detected photomultiplier response (signal) to decay to $1/e$ of its original value.

FURTHER READING

Books and Monographs

K.B. Yatsimirskii, 'Kinetic Methods of Analysis', Translated by P.J.J. Harvey, Vol. 25 of the *International Series of Monographs in Analytical Chemistry*, eds. R. Belcher, L. Gordon, Pergamon Press, Oxford, 1966.

H.B. Mark, Jr, G.A. Rechnitz with the assistance of R.A. Greinke, 'Kinetics in Analytical Chemistry', Vol. 24 of *Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications*, eds. P.J. Elving, I.M. Kolthoff, Interscience, New York, 1968.

H. Müller, M. Otto, G. Werner, *Katalytische Methoden in der Spurenanalyze*, Akademische Verlagsgessellschaft Geets & Portig K.-G., Leipzig, 1980.

D. Pérez-Bendito, M. Silva, *Kinetic Methods in Analytical Chemistry*, Translation eds. J.R. Majer, R.A. Chalmers, Ellis Horwood, Chichester, 1988.

H.A. Mottola, 'Kinetic Aspects of Analytical Chemistry', Vol. 96 of *Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications*, eds. J.D. Winefordner, I.M. Kolthoff, John Wiley & Sons, New York, 1988.

M. Kopanica, V. Stará, 'Kinetic Methods in Chemical Analysis', in Vol. XVIII of *Comprehensive Analytical Chemistry*, ed. G. Svehla, Elsevier, Amsterdam, 1983, 17–250.

G.G. Guilbault, *Enzymatic Methods of Analysis*, Pergamon Press, Oxford, 1970.

P.W. Carr, L.D. Bowers, 'Immobilized Enzymes in Analytical and Clinical Chemistry', Vol. 56 of Chemical Analysis: *A Series of Monographs on Analytical Chemistry and its Applications*, eds. P.J. Elving, J.D. Winefordner, I.M. Kolthoff, John Wiley & Sons, New York, 1980.

G.G. Guilbault, Editor, 'Analytical Uses of Immobilized Enzymes', *Modern Monographs in Analytical Chemistry*, Dekker, New York, Vol. 2, 1984.

Reviews and Other Useful Publications

General

1964–Present: Biennial reviews (Fundamental Reviews) in the journal Analytical Chemistry. This series started in 1964 with a review entitled 'Kinetic Aspects of Analytical Chemistry', authored by G.A. Rechnitz, and since then has appeared every two years with occasional change in authors.

H.L. Pardue, 'Holistic View of Chemical Analysis with Emphasis on Kinetic Methods', *Quim. Anal. (Bellaterra)*, **II**, (Extra), 24–34 (1983).

H.L. Pardue, 'Unified View of Kinetic-based Analytical Methods with Emphasis on Ruggedness', *Analyst*, **121**, 385–390 (1996).

Accounts of Evolution in Kinetic Determinations

H.A. Mottola, 'An Account of Kinetic Determinations and Other Kinetic Aspects of Analytical Chemistry: From Córdoba to Erlangen', *Analyst*, **118**, 675–679 (1993).

H.A. Mottola, 'From Erlangen to Moscow: Minor Redirections', *Analyst*, **121**, 381–383 (1996).

Definitions, Nomenclature, and Related Topics

H.B. Mark, Jr, 'Development and Publication of New Methods in Kinetic Analysis', *Talanta*, **20**, 257–266 (1973). [Recommendations made regarding research for development of new methods and/applications. Presentation of the results for publication also discussed].

H.L. Pardue, 'A Comprehensive Classification of Kinetic Methods of Analysis Used in Clinical Chemistry', *Clin. Chem.*, **23**, 2189–2201 (1977).

'Nomenclature of kinetic methods of analysis (IUPAC Recommendations 1993)', *Pure Appl. Chem.*, **65**, 2292–2298 (1993) [Names and definitions frequently used in kinetic methods].

H. Müller, 'Catalytic Methods of Analysis: Characterization, Classification and Methodology', *Pure Appl. Chem.*, **67**, 601–613 (1995).

K.J. Laidler, 'A Glossary of Terms Used in Chemical Kinetics, Including Reaction Dynamics', *Pure Appl. Chem.*, **68**, 149–192 (1996).

Catalytic Determinations

H.A. Mottola, 'Catalytic and Differential Rate Methods', *CRC Crit. Rev. Anal. Chem.*, **4**, 229–280 (1975).
H. Müller, 'Catalymetric Methods of Analysis', *CRC Crit. Rev. Anal. Chem.*, **13**, 313–372 (1982).
H.A. Mottola, 'Catalytic Titrants and Catalytic Indication of End-Points', *Talanta*, **16**, 1267–1275 (1969).
F.F. Gaál, 'Catalytic Titrations', *Analyst*, **112**, 739–751 (1987).
P.R. Bontchev, 'Catalytic Reactions-I, Mechanisms', *Talanta*, **17**, 499–513 (1970);
P.R. Bontchev, 'Catalytic Reactions-II, Activation', *Talanta*, **19**, 675–685 (1972);
Gy. Bazsa, M.T. Beck, 'Autocatalysis, Autoinhibition, Self-Catalysis, Self-Inhibition: Specific Kinetic Effects of the Reaction Products and the Reactants', *Acta Chim. Acad. Sci. Hung.*, **73**, 425–441 (1972);
H.A. Mottola, 'Some Considerations in Reaction Rate Methods – Analytical Use of Modifying Effects of Ligands in Some Electron-Transfer Reactions', *Anal. Chim. Acta*, **71**, 443–450 (1974) [Inhibition, True Metal-Complex Catalysis, and Promoting Effects].
S. Yamada, A. Murata, M. Tanaka, 'How to Design a Kinetic-Catalytic Method of Analysis Based on Ligand Substitution Reactions', *Mikrochim. Acta*, **III**, 291–297 (1988).
T. Kawashima, S. Nakano, M. Tabata, M. Tanaka, 'Catalytic Reactions as Applied to Trace Analysis', *Trends in Anal. Chem.*, **16**, 132–139 (1997).

Enzymatic Methods

H.A. Mottola, 'Enzymes as Analytical Reagents: Substrate Determinations with Soluble and with Immobilised Enzyme Preparations', *Analyst*, **112**, 719–727 (1987).

Differential Rate Methods

H.A. Mottola, 'Catalytic and Differential Rate Methods', *CRC Crit. Rev. Anal. Chem.*, **4**, 229–280 (1975).
D. Pérez-Bendito, 'Inorganic Differential Kinetic Analysis', *Analyst*, **109**, 891–899 (1984).
D. Pérez-Bendito, 'Approaches to Differential Reaction-Rate Methods', *Analyst*, **115**, 689–698 (1990).
B.M. Quencer, S.R. Crouch, 'Multicomponent Kinetic Methods', *CRC Crit. Rev. Anal. Chem.*, **24**, 243–262 (1993).

Determinations Based on Uncatalyzed Reactions

H.A. Mottola, 'Kinetic Determinations of Reactants Utilizing Uncatalyzed Reactions', *Anal. Chim. Acta*, **280**, 279–287 (1993).

Electrocatalysis

H.B. Mark, Jr, 'Electrocatalysis and its Application to Analysis', *Analyst*, **115**, 667–678 (1990).

Instrumentation

H.V. Malmstadt, C.J. Delaney, E.A. Cordos, 'Reaction Rate Methods of Chemical Analysis', *CRC Crit. Rev. Anal. Chem.*, **2**, 559–619 (1972).
H.V. Malmstadt, C.J. Delaney, E.A. Cordos, 'Instruments for Rate Determinations', *Anal. Chem.*, **44**(12), 79A–89A (1972).
S.R. Crouch, in 'Computers in Chemistry and Instrumentation', Vol. 3, *Spectroscopy and Kinetics*, eds. J.S. Mattson, H.B. Mark, Jr, H.C. MacDonald, Jr, Dekker, New York, 1973, Chapter 3.
M. Silva, 'Recent Strategies in Automated Reaction Rate Based Determinations', *Analyst*, **118**, 681–687 (1993).

Use of Micelles in Kinetic Determinations

M.L. Carreto, S. Rubio, D. Pérez-Bendito, 'Organic Microheterogeneous Systems in Kinetic Analysis. Self-assembled Systems', *Analyst*, **121**, 33R–44R (1996).

Selectivity

G. Werner, 'Selectivity in Catalytic Methods of Determination', *Quim. Anal. (Bellaterra)*, **II** (Extra), 68–84 (1983).
M. Valcárcel, 'Selectivity and Kinetics in Analytical Chemistry', *Analyst*, **112**, 729–737 (1987).

Incidental Kinetics in Analytical Chemistry

H.A. Mottola, 'Some Kinetic Aspects Relevant to Contemporary Analytical Chemistry', *J. Chem. Educ.*, **58**, 399–403 (1981).
H.A. Mottola, 'Some Incidental Kinetic Aspects in Analytical Chemistry', *Analyst*, **115**, 679–683 (1990).

REFERENCES

1. R.D. Allison, D.L. Purich, 'Enzyme Kinetics and Mechanism, Part A, Initial Rate and Inhibitors Methods', in *Methods in Enzymology*, ed. D.L. Purich, Academic Press, New York, 4, Vol. 63, 1979.
2. R.C. Harris, E. Hultman, 'Kinetic Methods That are Independent of the Rate of Reaction', *Clin. Chem.*, **29**, 2079–2081 (1983).
3. R. Christie Smith, 'Some Considerations of the Reaction Constant Equation, and a Simple Method of Determining the End-Point', *Philos. Mag. J. Sci.*, **1**, 496–499 (1926).
4. H.A. Mottola, *Kinetic Aspects of Analytical Chemistry*, John Wiley & Sons, New York, Chapter 9, 1988.
5. (a) M. Valcárcel, M.D. Luque de Castro, *Flow Injection Analysis. Principles and Applications*, Ellis Horwood, Chichester, UK, Chapter 9, 1987; (b) J. Ruzicka, E.H. Hansen, *Flow Injection Analysis*, 2nd edition, John Wiley & Sons, New York, Chapter 2, 1988.

6. C.F. Jacobsen, J. Leonis, K. Linderstrom-Lang, M. Ottesen, *Methods of Biochemical Analysis*, ed. D. Glick, Interscience, New York, 171, Vol. 4, 1957.
7. L. Gordon, M.L. Salutsky, H.H. Willard, *Precipitation from Homogeneous Solution*, John Wiley & Sons, New York, 1959.
8. M.G. Chancel, 'De L'emploi des Hyposulfites dans L'analyze – Application á la Séparation Directe du Fer D'avec L'alumine', *Compt. rend.*, **46**, 987–990 (1858).
9. M.A. Guyard, 'De la Formation du Noir D'aniline, au Moyen se Sels de Vanadium, et de la Théorie de la Formation du Noir D'aniline', *Bull. Soc. Chim. Paris*, **25**, 58–64 (1876).
10. G. Witz, F. Osmond, 'Essais sur l'Application des Propriétés de Poxycellulose au dosage du vanadium', *Bull. Soc. Chim. Paris*, **45**, 309–315 (1885).
11. W. Roberts, 'On the Estimation of the Amylolytic and Proteolytic Activity of Pancreatic Extracts', *Proc. Roy. Soc.*, **32**, 145–161 (1881).
12. H.T. Brown, J. Heron, 'Contributions to the History of Starch and its Transformations', *J. Chem. Soc.*, **35**, 596–654 (1879).
13. (a) E.B. Sandell, I.M. Kolthoff, 'Chronometric Catalytic Method for the Determination of Micro Quantities of Iodine', *J. Am. Chem. Soc.*, **56**, 1426 (1934); (b) E.B. Sandell, I.M. Kolthoff, 'Micro Determination of Iodine by a Catalytic Method', *Mikrochim. Acta*, 9–25 (1937).
14. (a) L. Szebellédy, M. Ajtai, 'Die Aktivierung der Katalytischen Reaktionen des Vanadins mit Kaliumhydrotartrat', *Mikrochemie*, **25**, 258–262 (1938); (b) L. Szebellédy, M. Ajtai, 'Die Quantitative Bestimmung von Vanadim Mittels Aktivierter Katalyse', *Mikrochemie*, **26**, 87–94 (1939).
15. K.B. Yatsimirskii, 'Kinetic Methods of Analysis', in the *International Series of Monographs in Analytical Chemistry*, eds. R. Belcher, L. Gordon, Pergamon Press, Oxford, Vol. 25, 1966.
16. Z. Arkosi, *Ein Beitrag zur Mikrobestimmung einiger hochwertiger Schwermetallionen auf Grund ihrer Katalytischen Wirkung, Katalytische Analyze*, Juris Verlag, Zürich, 1956.
17. M.J. Clarens, 'Application des Lois de la Cinétique Chimique á L'analyze Quantitative: Du Dosage Fractionné des Tannins en Général et des Tannins du Vin, en Particulier', *Bull. Soc. Chim. France (Mémoires)*, **29**, 837–852 (1921).
18. I.M. Kolthoff, T.S. Lee, 'Use of Perbenzoic Acid in Analysis of Unsaturated Compounds. II. Determination of External Double Bonds in Synthetic Rubbers', *J. Polymer Sci.*, **2**, 206–219 (1947).
19. T.S. Lee, in *Organic Analysis*, eds. J. Mitchell, Jr, I.M. Kolthoff, E.S. Proskauer, A. Weissberger, Interscience, New York, 237–252, Vol. 2, 1954.
20. J.S. Fritz, G.S. Hammond, *Quantitative Organic Analysis*, John Wiley & Sons, New York, Chapter 9, 1957.
21. H. Stetter, *Enzymatische Analyze*, Verlag Chemie, Weinheim, 1951.
22. A.A. Alexiev, P.R. Bontchev, I. Todorov, 'A Chemical Catalytic Method for Determination of Silver in Human Saliva', *Archs. Oral Biol.*, **18**, 1461–1467 (1973).
23. H.V. Malmstadt, G.P. Hicks, 'Determination of Glucose in Blood Serum by a New Rapid and Specific Automatic System', *Anal. Chem.*, **32**, 394–398 (1960).
24. W.J. Blaedel, G.P. Hicks, 'Continuous Analysis by Measurement of the Rate of Enzyme Catalyzed Reactions. Glucose Determination', *Anal. Chem.*, **34**, 388–394 (1962).
25. G.E. James, H.L. Pardue, 'Computer Assisted Reaction-Rate Analyzes', *Anal. Chem.*, **41**, 1618–1623 (1969).
26. (a) P.W. West, 'Inorganic Microchemistry', *Anal. Chem.*, **26**, 121–128 (1954); (b) *Anal. Chem.*, **30**, 748–759 (1958); (c) *Anal. Chem.*, **34**, 104R–111R (1962).
27. S. Siggia, *Quantitative Organic Analysis via Functional Groups*, John Wiley & Sons, New York, 1949.
28. A.E. Nielsen, *The Kinetics of Precipitation*, Macmillan, New York, 1964.
29. P.R. Danesi, R. Schiarizia, 'The Kinetics of Metal Solvent Extraction', *CRC Crit. Rev. Anal. Chem.*, **10**, 1–126 (1980).
30. F. Helfferich, *Ion Exchange*, McGraw-Hill, London, 1962.
31. J.C. Giddings, *Dynamics of Chromatography, Part I: Principles and Theories*, Dekker, New York, 1965.
32. (a) P. Delahay, *Double Layer and Electrode Kinetics*, Wiley-Interscience, New York, 1965; (b) K.J. Vetter, *Electrochemical Kinetics (Theoretical and Experimental Aspects)*, Academic Press, New York, 1967; (c) H.R. Thirsk, J.A. Harrison, *A Guide to the Study of Electrode Kinetics*, Academic Press, London, 1972; (d) E. Gileadi, *Electrode Kinetics for Chemists, Chemical Engineers, and Materials Scientists*, VCH Publishers, New York, 1993.
33. J. Barthel, *Thermometric Titrations*, Wiley-Interscience, New York, 66–70, 1975.
34. M.I. Poppe, M.D. Judd, *Differential Thermal Analysis*, Heyden, Philadelphia, 1977.
35. W.H.C. Walker, in *Continuous Flow Analysis: Theory and Practice*, ed. W.B. Furman, Dekker, New York, Chapter 7, 1976.
36. (a) I.M. Warner, 'Molecular Fluorescence and Phosphorescence', in *Instrumental Analysis*, 2nd edition, eds. G.D. Christian, J.E. O'Reilly, Allyn and Bacon, Boston, Chapter 9, 1986; (b) R.J. Hurtubise, 'Solid-matrix Luminescence Analysis: Photophysics, Physicochemical Interactions and Applications', *Anal. Chim. Acta*, **351**, 1–22 (1997).
37. R. Chang, *Basic Principles of Spectroscopy*, McGraw-Hill, New York, Chapter 12, 1971.
38. W.R. Seitz, 'Chemiluminescence and Bioluminescence Analysis: Fundamentals and Biomedical Applications', *CRC Crit. Rev. Anal. Chem.*, **13**, 1–58 (1981).
39. V. Bruckner, K.-H. Feller, U.-W. Grummt, *Application of Time-resolved Optical Spectroscopy*, Elsevier, Amsterdam, 1990.

40. R. Payling, in *Glow Discharge Optical Emission Spectrometry*, eds. R. Payling, D.G. Jones, A. Bengtson, John Wiley & Sons, Chichester, UK, 1997.
41. (a) J.J. Scherer, J.B. Paul, A. O'Keefe, R.J. Saykally, 'CRLAS: A New Analytical Technique for Cluster Science', in *Advances in Metal and Semiconductor Clusters*, ed. M.A. Duncan, JAI Press, Greenwich, CT, 149–180, Vol. 3, 1995; (b) M.D. Wheeler, S.M. Newman, A.J. Orr-Ewing, M.N.R. Ashfold, 'Cavity Ring-Down Spectroscopy', *J. Chem. Soc., Faraday Trans.*, **94**, 337–351 (1998).

Catalytic Kinetic Determinations: Nonenzymatic

Takuji Kawashima
University of Tsukuba, Tsukuba, Japan

Norio Teshima
Aichi Institute of Technology, Toyota, Japan

Shigenori Nakano
Tottori University, Tottori, Japan

1 Introduction 2
2 Theoretical Bases of Kinetic–Catalytic Methods of Analysis 2
3 Indicator Reactions 3
3.1 Redox Reactions 3
3.2 Ligand Substitution and Complexation Reactions 6
4 Activators and Surfactants in Catalyzed Reactions 6
4.1 Activators 6
4.2 Surfactants 7
5 Methods of Determination 7
5.1 Tangent and Initial Rate Methods 7
5.2 Fixed Time Methods 8
5.3 Variable Time Methods 9
5.4 Measurements of Induction Period 9
5.5 Analysis of Mixtures by Differential Rates 9
5.6 Flow-injection Analysis 10
6 Titrimetric Methods with Catalytic Endpoint Indication 10
7 Kinetic–Catalytic Methods for Trace Elements 14
7.1 Catalytic Methods Using Redox Reactions 14
7.2 Catalytic Methods Using Ligand Substitution and Complex Formation Reactions 21
Acknowledgments 24
Abbreviations and Acronyms 24
Popular Names 25
Related Articles 26
References 27

Although a catalyst does not shift the position of equilibrium but it does markedly change the reaction rate, because the catalyst lowers the activation energy of a reaction providing an alternative favorable route for the formation of the products. The reacted catalyst is regenerated and is repeatedly involved in an indicator reaction. When the rate depends on the concentration of the catalyst, it can be determined by measuring the change in the physical properties of the reactants and/or products in the indicator reaction. The change in a physical property such as absorbance, fluorescence, chemiluminescence (CL) or electrode potential of a reaction system is usually monitored with time. This methodology is called catalytic methods or kinetic–catalytic methods of analysis.

Kinetic–catalytic methods of analysis based on catalyzed reactions have progressed recently and have been applied to trace analyses for various catalysts (element) because of their extremely high sensitivity, low limit of detection and selectivity. Several reactions such as redox, ligand substitution and complexation reactions have been utilized as indicator reactions for the development of kinetic–catalytic methods of trace analyses. Most of the catalysts in redox reactions are metal ions or inorganic anions having different oxidation states such as cobalt, copper, chromium, iron, manganese, vanadium, selenium, iodide, bromide and nitrite. Oxidants capable of regenerating the catalyst are hydrogen peroxide, chlorate, bromate, periodate and dissolved oxygen. Reductants include ascorbic acid, hydroxylamine, hypophosphorous acid, alkali sulfides, dithiothreitol and tin(II).

Certain substances significantly increase the rate of catalyzed reactions. These substances are referred to as activators and have been defined as substances that do not catalyze the indicator reaction but enhance the catalytic effect of the catalyst. Activators provide high sensitivity and selectivity in kinetic–catalytic methods. Such improved specificity can be expected to eliminate the preliminary separation in many instances.

As the measurement includes time as an experimental variable, care is needed to ensure that mixing of reagents takes place at regular time intervals to obtain highly accurate results using a manual procedure. Such disadvantages in the manual system can be overcome by using flow-injection analysis (FIA), in which the reaction time can easily be controlled by varying the flow rate of the solutions and the length of the reaction coil.

Kinetic–catalytic methods allow the determination of various elements at nanogram and picogram levels (10^{-11} M) with a simple and reasonable equipment. The detection limits (DLs) of the methods are of the order of picograms per milliliter with good accuracy and precision and are almost equal to those for inductively coupled plasma mass spectrometry (ICPMS).

The application of catalytic effects of metal ions in titrimetry has been recognized for a long time. This methodology is called catalytic titration and involves two

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

consecutive reactions, i.e. titration reaction and indicator reaction. The determinable concentration is almost 10^{-6} M of the metal ions.

1 INTRODUCTION

Methods of chemical analysis for the analyte in a mixture are mainly classified into two classes, *thermodynamic* (or *static*) and *kinetic* (or *dynamic*) classes, according to the nature of measurement employed. A thermodynamic method generally involves changing the equilibrium conditions of the reaction system to proceed essentially to completion, that is, the equilibrium greatly favors the reaction products, and a physical property of the solution, corresponding to the concentration of either the reactants or the products, is measured by using conventional analytical methods such as spectrophotometry, fluorometry, chemiluminescent method, atomic absorption spectrometry, flame spectrometry, an electrochemical method, and so on. A kinetic method involves the measurement of the rate of the chemical reaction under nonequilibrium conditions as a function of time. Furthermore, kinetic methods of analysis are classified into two categories, catalytic and noncatalytic determinations. These methods have been subdivided according to the type of reaction and to the procedure employed for determining single or multiple components in a mixture. Classification of kinetic methods is listed in Table 1.

The change in a physical property such as absorbance, fluorescence, CL, electrode potential of a system is usually monitored. Numerous batch and flow-injection methods have been reported for the catalytic determination of trace elements. Kinetic–catalytic methods are usually much more sensitive than thermodynamic ones based on stoichiometric reactions. The sensitivity of the kinetic–catalytic methods of analysis is estimated to be 10^{-16} M or 10^{-5} pg mL^{-1}, under the most suitable conditions, when a spectrophotometric method is used for determining reaction rate.[1] The use of activators in the kinetic–catalytic methods increases further the sensitivity and selectivity of the methods. Improved specificity can be expected to eliminate the preliminary separation in many instances. The details of the action of activators will be discussed in section 4.

Table 1 Classification of kinetic methods based on the type of reaction employed

Catalytic methods
Nonenzymatic: catalytic effects of chemical species on redox and ligand substitution reactions
Enzymatic: substrate determinations employing soluble and immobilized enzymes
Electrochemical: electrode reaction
Noncatalytic methods
Determination of a single species
Determination of multicomponents (differential reaction rate methods)

A catalytic reaction was first applied to quantitative analysis by Guyard who reported a catalytic determination of vanadium based on its catalytic effect on the oxidation of aniline by chlorate in 1876.[2] Nearly 60 years later, Sandell and Kolthoff reported the iodide determination using its catalysis on the reaction of As(III) with Ce(IV),[3,4] Equation (1)

$$\text{As(III)} + 2\text{Ce(IV)} \longrightarrow \text{As(V)} + 2\text{Ce(III)} \qquad (1)$$

This method is probably the only catalytic method for iodine as iodide that has been extensively used. After that, kinetic–catalytic methods have been recognized as offering a valuable approach to trace analysis, because of their high sensitivity, low limit of detection and good selectivity, and have been the subject of reviews[5–23] and monographs.[24–27] Since 1964, biennial reviews entitled ''Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry'' have been published in *Analytical Chemistry*.[28–34] The contents of the reviews consist of kinetic methods for determination of catalyst, kinetic methods based on activation or inhibition of catalysis, titrimetric methods with catalytic end-point indication, kinetics and mechanisms of some catalyzed reactions of analytical interest, applications of luminescence, kinetic methods based on uncatalyzed reactions, differential methods (simultaneous multicomponent determinations), instrumentation, and so on. According to Mottola, the idea of an international gathering of chemists interested in all kinetic aspects of analytical chemistry was conceived in the late 1970s. Since the First International Symposium on Kinetics in Analytical Chemistry took place in Cordoba, Spain, 1983, a series of symposia has been held every three years.[35–38] The details of many of the contributions presented at these meetings from the second symposium can be found in *The Analyst*.

This article focuses on kinetic–catalytic methods for nonenzymatic reactions in homogeneous systems.

2 THEORETICAL BASES OF KINETIC–CATALYTIC METHODS OF ANALYSIS

A catalyst is a substance that alters the rate of a reaction without being consumed or shifting the position of equilibrium of a reaction system. The catalyst lowers the activation energy of a reaction, providing an alternative favorable route for the formation of the products. Reacted

catalyst is regenerated and repeatedly involved in the reaction, although the catalyst is chemically unchanged at the end of the reaction. The rate of catalyzed reaction is proportional to the concentration of the catalyst under given conditions and can therefore be employed for the catalyst determination by using a suitable procedure.

In the kinetic–catalytic methods, the main reaction in which the species participate is called the indicator reaction. This reaction is always thermodynamically favored to completion, but is kinetically unfavorable. One type of indicator reaction frequently used in the kinetic–catalytic methods is a redox reaction in which the catalyst M, usually a multivalent ion, changes its oxidation state during the reaction as follows, Equations (2–4):

$$S + R \longrightarrow P + Q \quad (2)$$

$$S + M \longrightarrow P + M' \quad (3)$$

$$M' + R \longrightarrow Q + M \quad (4)$$

where S and R are the substrate and reagent, respectively, and P and Q are the reaction products in the indicator reaction, Equation (2). The indicator reaction is usually very slow. However, when an ion M acts as a catalyst, it accelerates the formation of the product P and is then converted into a different oxidation state, M′ by reaction shown in Equation (3). If M′ is regenerated again to M by reaction shown in Equation (4), the formation of P is catalyzed by a trace amount of M. Thus small amounts of M even at subnanogram levels can be determined by measuring the rate of the formation of P. It is also true from Equation (4) that M′ can be determined as well as M.

The rate of reaction (2) catalyzed by M, v, is expressed as Equation (5):

$$v = k_u[S]^a[R]^b + k_c[S]^a[R]^b[M]^c \quad (5)$$

where k_u and k_c are the rate constants of uncatalyzed and catalyzed reactions, respectively, and a, b and c are reaction orders with respect to each species. In most cases, the reaction is first-order in S and R, and also in M. Then rate Equation (5) can be written as:

$$-\frac{d[S]}{dt} = \frac{d[P]}{dt} = k_u[S][R] + k_c[S][R][M] \quad (6)$$

The concentration of the catalyst in a mixture can be determined by monitoring the rate of decrease in concentration of S or the rate of an increase in concentration of one of the products. It is preferable that the rate constant of uncatalyzed reaction, k_u, is smaller than that of catalyzed reaction, $k_c[M]$, in order to obtain high sensitivity and precise results.

3 INDICATOR REACTIONS

A large number of indicator reactions are used for the kinetic–catalytic methods in homogeneous systems. Nonenzymatic indicator reactions can be classified into two broad categories: (1) redox reaction and (2) ligand substitution and complex formation reactions.

3.1 Redox Reactions

Redox reactions have frequently been used as the indicator reactions in kinetic–catalytic methods with spectrophotometric detection. Reactant S used in the redox reaction is an inorganic or organic compound, which is converted into a colored or colorless compound P by an oxidant and/or a reductant R. Oxidation reactions of organic compounds are likely to be used most frequently as the indicator reactions for the determination of metal ions based on their catalysis. The absorbance change of the reactant S and/or the product P is usually monitored. Typical organic compounds used in the indicator reactions are presented in Figure 1. Oxidants used in the indicator reactions are hydrogen peroxide, chlorate, bromate, peroxodisulfate, periodate and dissolved oxygen, and reductants such as ascorbic acid, hydrazine, hydroxylamine, hypophosphorous acid, alkali sulfides, thiosulfate and tin(II), etc. are also used. The role of oxidants is simply to regenerate the catalyst M. Although these oxidants have high oxidation potential for oxidizing S to P, the oxidation rate is very slow, and the regeneration process of M′ to M proceeds very rapidly. The concentration of catalyst M is kept constant during the reaction, as a result of the regeneration of M. Thus a trace amount of M can be determined. Typical examples of catalysts are metal ions such as iron(II, III), copper(I, II), manganese(II, IV, VII), vanadium(IV, V), cobalt(II, III), chromium(III, VI) and nonmetal ions such as iodide and bromide.

CL processes are closely related to redox reactions. Generally, CL is observed in a redox reaction which yields an electronically excited product; it emits luminescence or transfers its energy to another luminescent molecule present.[39,40] As shown in Figure 2, commonly used reagents for the CL reactions are acyl hydrazides (e.g. luminol), acridine derivatives (e.g. lucigenin), imidazoles (e.g. lophine) and peroxyoxalates (e.g. bis(2,4,6-trichlorophenyl) oxalate (TCPO) and bis(2,4-dinitrophenyl) oxalate (DNPO)).

Of these, luminol (5-amino-2,3-dihydro-1,4-phthalazinedione) is widely used in the CL reactions. The CL reaction of luminol was first reported by Albrecht in 1928[41] and involves the oxidation of luminol by oxidants such as oxygen, hydrogen peroxide, hypochlorite, persulfate, and so on in an alkaline solution.[42,43] The CL reaction of luminol with hydrogen peroxide has been

known as one of the most sensitive systems in analytical use and has been applied for the determination of metal ions such as cobalt(II), copper(II), chromium(III), iron(II) and so on, at the nanogram level. In addition, the luminol/H_2O_2/hexacyanoferrate(III) system has been widely used for the determination of hydrogen peroxide and biosubstances.

Lucigenin (10,10′-dimethyl-9,9′-biacridinium dinitrate) is the second most popular CL reagent and its CL properties were first observed by Glue and Petsch in

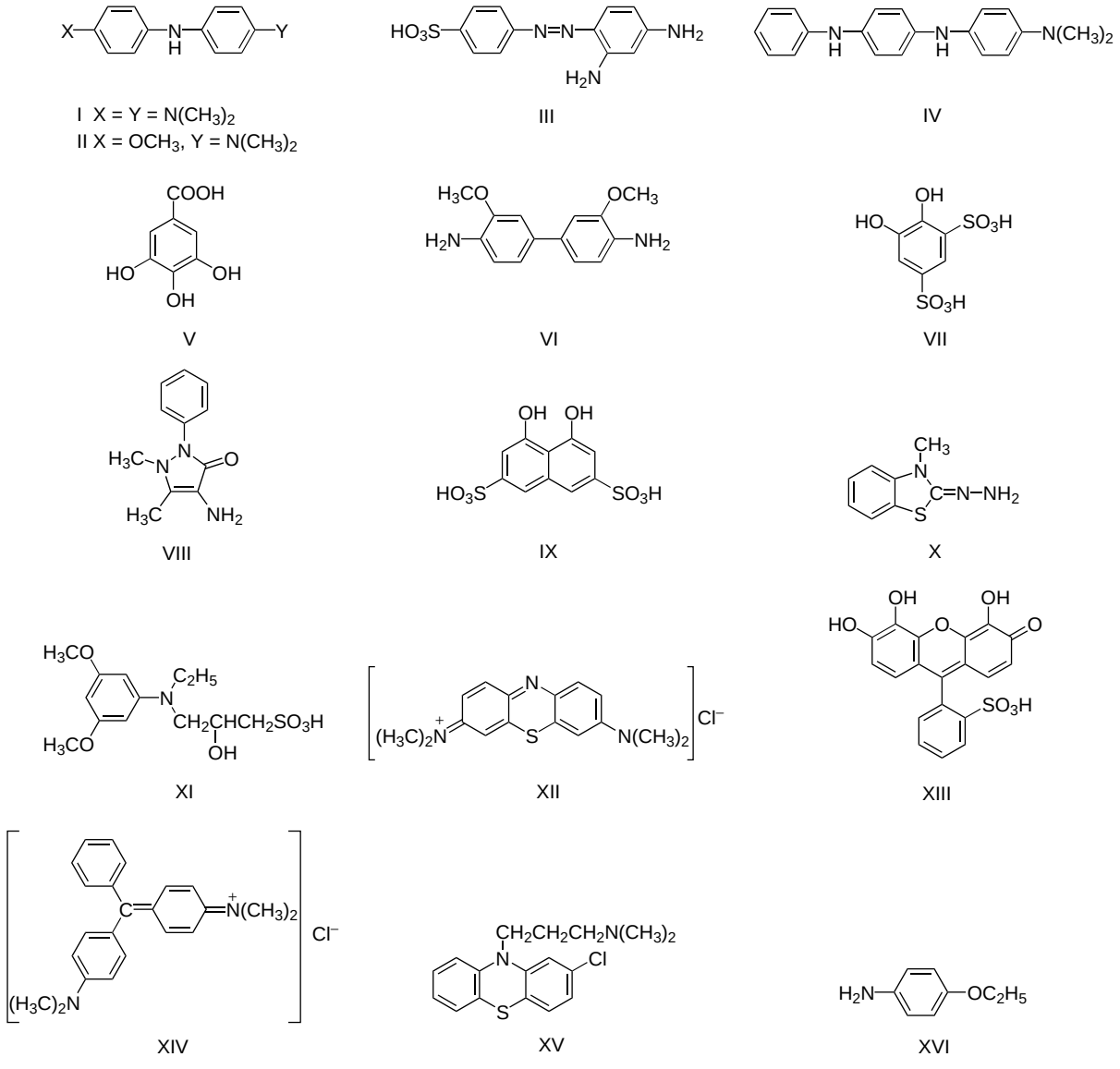

Figure 1 Representative examples of organic compounds used in indicator reactions. I, Bindschedler's green leuco base (4,4-bis(dimethylamino)diphenylamine); II, MDP (Variamine Blue leuco base analog) *N*-(*p*-methoxyphenyl)-*N*′,*N*′-dimethyl-*p*-phenylenediamine; III, azo dye formed by reaction of HBS (*p*-hydrazinobenzenesulfonic acid) with PDA (*m*-phenylenediamine); IV, dye formed by reaction of PPDA (*N*-phenyl-*p*-phenylenediamine) with DMA (*N*,*N*-dimethylaniline); V, gallic acid (3,4,5-trihydroxybenzoic acid); VI, *o*-dianisidine; VII, Tiron (1,2-dihydroxybenzene-3,5-disulfonic acid); VIII, AA, 4-aminoantipyrine (4-amino-1,5-dimethyl-2-phenyl-3-pyrazolone); IX, chromotropic acid (1,8-dihydroxy-3,6-naphthalenedisulfonic acid); X, MBTH (3-methyl-2-benzothiazolinone hydrazone); XI, DAOS (*N*-ethyl-*N*-(2-hydroxy-3-sulfopropyl)-3,5-dimethoxyaniline); XII, methylene blue (3,7-bis(dimethylaniline)-phenazothionium chloride); XIII, Pyrogallol Red (pyrogallol sulfonephthalein); XIV, malachite green (bis(*p*-dimethylaminophenyl)phenylmethylium chloride); XV, chlorpromazine (2-chloro-10-(3-dimethylaminopropyl)phenothiazine; XVI, *p*-phenetidine; XVII, DBPT (4,4′-dihydroxybenzophenone thiosemicarbazone); XVIII, MTT (3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2*H*-tetrazolium bromide); XIX, gallocyanine (1-carboxy-7-(dimethylamino)-3,4-dihydroxyphenoxazine-5-ium chloride); XX, DPC (diphenylcarbazone) (phenylazoformic acid 2-phenylhydrazide); XXI, indigo carmine (5,5′-indigotindisulfonic acid).

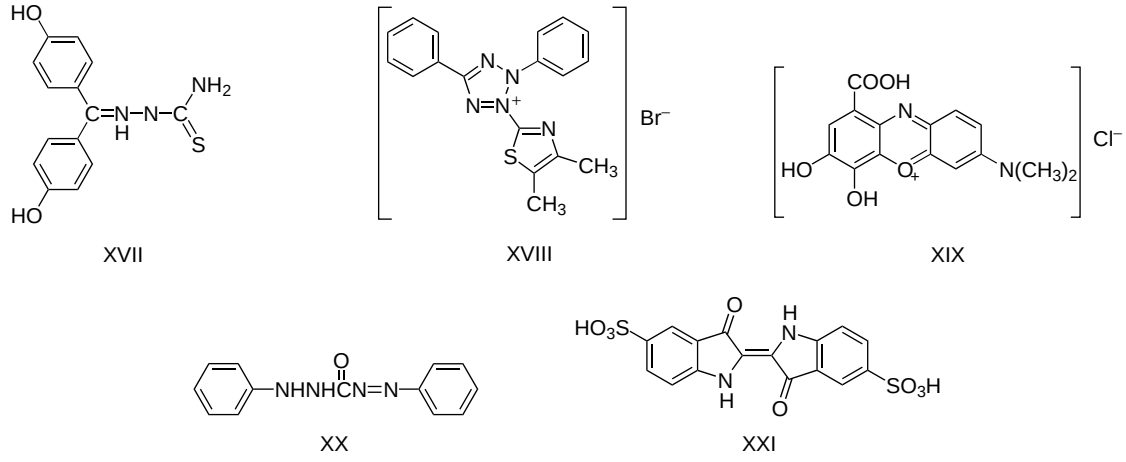

Figure 1 (*Continued*)

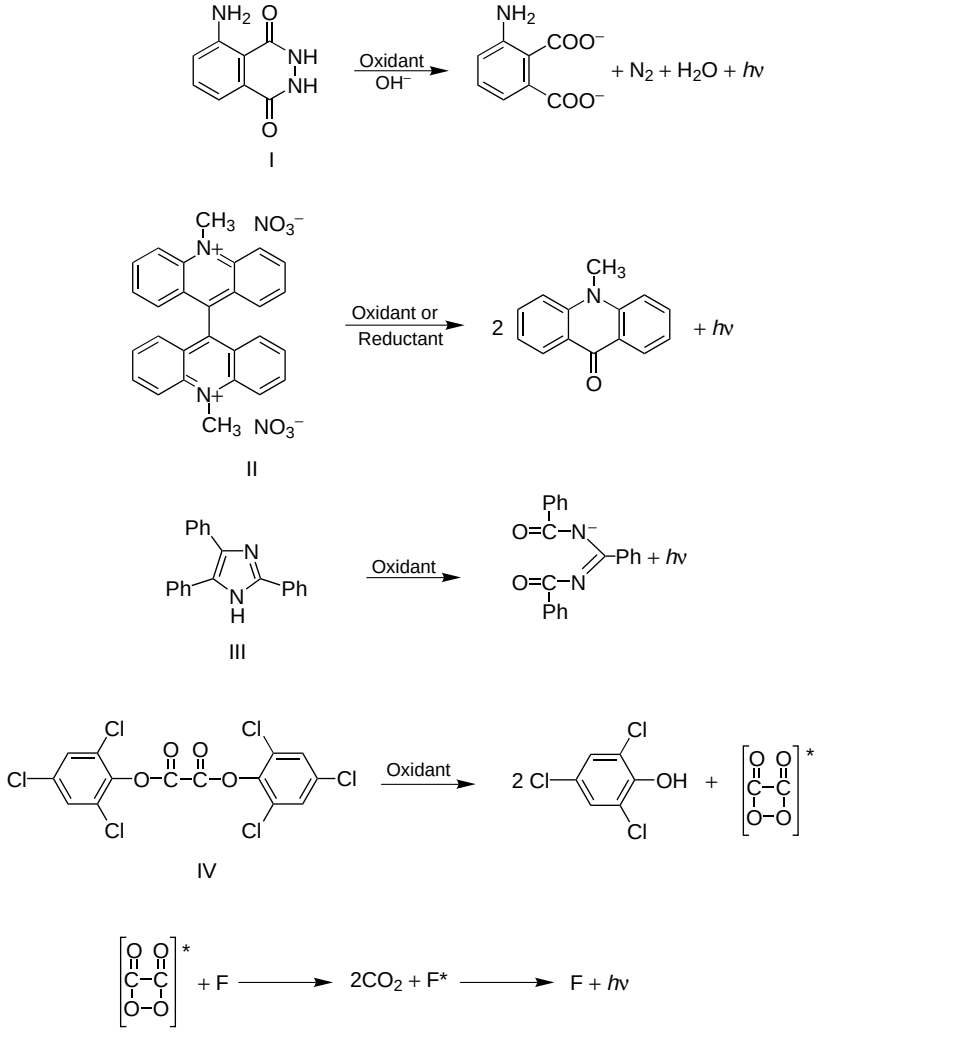

Figure 2 Representative CL compounds and their postulated schemes. I, luminol; II, lucigenin; III, lophine; IV, TCPO (F denotes a fluorophore).

1935.[44] The oxidation of lucigenin occurs with peroxide under alkaline media to produce a blue light emission as well as a green luminescent light when a catalyst, e.g. heavy metal ion, is present. This blue light is spectrally the same as the fluorescent light of the oxidized product of lucigenin, *N*-methylacridone.[43] As *N*-methylacridone is insoluble in water, a micellar system containing a suitable surfactant is used to solubilize the precipitate.[45] There are a large number of analytical applications of the CL reaction of lucigenin for the determination of metal ions.

The CL of lophine (2,4,5-triphenylimidazole) was first discovered by Radzisewski in 1877.[46] The CL reaction of lophine involves the initial oxidation of the lophine anion to its free radical, reaction of the radical with molecular oxygen to give a peroxide, and reaction of the peroxide with base to yield light.[47] The oxidation reaction of lophine is also catalyzed by various transition metal ions in the presence of suitable oxidants.

Peroxyoxalate CL involves its oxidation with hydrogen peroxide to produce an energy-rich intermediate. This metastable intermediate can provide enough chemiexcitation energy for the production of an excited state fluorophore. Through the production of a charge transfer complex between the intermediate and a fluorophore, the actual chemiexcitation step occurs. This charge transfer complex dissociates to produce the chemically excited fluorophore which returns to the ground state via emission of a photon.[48]

3.2 Ligand Substitution and Complexation Reactions

Kinetic–catalytic determinations not involving redox reaction processes have also been developed by many authors. An example of this category is the substitution of ligands and/or metal ions in metal complexes. This reaction is homogeneously catalyzed by minute amounts of relatively soft metal ions such as mercury, cadmium as well as ammonia and some other species.

A substitution reaction of one ligand by another L in a metal complex maximum absorbance (MA) or MAX and a double exchange reaction between M_1L and M_2X are widely used in kinetic–catalytic methods, according to the general scheme shown in Equations (7–9):

$$MA + L \longrightarrow ML + A \tag{7}$$

$$MAX + L \longrightarrow ML + A + X \tag{8}$$

$$M_1L + M_2X \longrightarrow M_1X + M_2L \tag{9}$$

An example of Equation (7) is the reaction of hexacyanoferrate(II) with *p*-nitrosodiphenylamine (*p*-NDA) in aqueous solution, which is catalyzed by mercury(II), Equation (10).[49,50]

$$Fe(CN)_6^{4-} + p\text{-NDA} \longrightarrow Fe(CN)_5 \cdot p\text{-NDA}^{3-} + CN^- \tag{10}$$

The concentration of the green product, $[Fe(CN)_5 \cdot p\text{-NDA}]^{3-}$ ($\lambda_{max} = 640$ nm), is proportional to the amount of mercury(II) present in a solution at a fixed time. The determinable range for mercury(II) is 1×10^{-7}–1×10^{-6} M.

An example of Equation (9) is the exchange reaction between triethylenetetraminenickel(II) (NiT^{2+}) and ethylenediaminetetraacetatocuprate(II) (CuY^{2-}), which is catalyzed by EDTA (ethylenediamine-*N*,*N*,*N*′,*N*′-tetraacetic acid) or Trien (triethylenetetraamine) and is inhibited by any metal ions complexing these ligands,[51,52] Equation (11)

$$NiT^{2+} + CuY^{2-} \longrightarrow CuT^{2+} + NiY^{2-} \tag{11}$$

The exchange rate is monitored by measuring the absorbance of CuT^{2+} as a function of time. Trace amounts of EDTA (10^{-7} M levels) and metal ions (10^{-8} M levels) can be determined.

4 ACTIVATORS AND SURFACTANTS IN CATALYZED REACTIONS

Generally a catalyzed reaction proceeds more quickly than an uncatalyzed reaction with an increase in temperature and reaction time. However, the increase in temperature and time also increases the rate of uncatalyzed reaction. It is possible to increase the sensitivity of kinetic–catalytic methods by the application of activators and/or surfactants.

4.1 Activators

An activator is defined as a substance that does not catalyze the indicator reaction, Equation (2), but increases its rate in the presence of a catalyst. The use of activators also provides greater selectivity in kinetic–catalytic methods. According to Bontchev,[6] activators are classified into three groups depending on the role of the catalyst in the reaction, the nature of the catalyst–activator interaction and the stage of the reaction affected in the presence of the activator:

1. activators affecting the catalyst–substrate interaction
2. activators participating in the regeneration of the catalyst
3. activators acting indirectly in the catalytic process.

Generally, activation effects are limited to metal-catalyzed reactions and connected to the complex-forming tendency of metal ions. Furthermore, the dependence of the rate of the reaction on activator concentration shows a maximum. Such a maximum is a characteristic feature of all types of activator which are concerned

with the formation of ternary complexes of the type activator–metal–substrate.[6] Thus, the coordination sphere of the catalyst is fully occupied by the activator and any complexation with substrate does not occur at higher concentrations of activator. For example, 1,10-phenanthroline (phen) and 2,2′-bipyridine (bpy) act as effective activators for the catalysis of iron(III),[53] manganese(II)[54] and copper(II)[55] in the indicator reactions and the dependence of the concentration of these activators shows a maximum. 5-Sulfosalycilic acid, tartaric acid and 1,2-dihydroxy-3,5-benzenedisulfonic acid (Tiron) act as effective activators for vanadium(V)-catalyzed reactions.[56,57] EDTA acts as an effective activator only for the chromium(III)-catalyzed reactions,[58] although EDTA masks the catalysis of other metal ions.

4.2 Surfactants

Surfactant molecules associate dynamically to form assemblies such as micelles, reversed micelles and bilayer aggregates (lamela or vesicle structure).[22,59,60] At concentrations above the critical micelle concentration (CMC) of surfacants, the following capabilities are expected:

1. solubilize, concentrate and organize reactants according to hydrophobic, electrostatic and specific interactions
2. alter the effective microenvironment, i.e. polarity, viscosity, acidity
3. alter the chemical pathways and reaction rates
4. alter the spectral parameters of solubilized chemical species
5. stabilize the chemical species with optical transparency.

These properties are of interest in analysis and can be useful in kinetic–catalytic methods as well as in equilibrium-based ones. The use of organized surfactant molecular assemblies increases the sensitivity, selectivity and precision of the methods. For example, the copper(II)-catalyzed reaction between *N*,*N*-dimethyl-*p*-phenylenediamine (DPD) and DMA in the presence of hydrogen peroxide is enhanced by sodium dodecylsulfate (SDS) and dodecyltrimethylammonium bromide (DTAB) micelles.[61] The DL and specificity of the copper(II) determination in the presence of SDS or DTAB are improved compared with those in their absence.[62]

5 METHODS OF DETERMINATION

In order to determine the concentration of a catalyst or a reactant in solution by the kinetic-based method, the change in the physical property of the monitored species as a function of time must be measured and kinetic curves are needed.

The procedures for determining the catalyst concentration by measuring the rate of catalyzed reaction are the initial rate method (tangent method), fixed time method and variable time (fixed concentration) method.[1,24–27] The induction period measurements are also applied to some catalytic reactions in which there is a time delay between reactants and the appearance of change in the chemical composition of the system.

Equation (6) can be simplified as follows, Equations (12) and (13):

$$\frac{\mathrm{d[P]}}{\mathrm{d}t} = F_\mathrm{u} + F_\mathrm{c}[\mathrm{M}] \tag{12}$$

$$\frac{\Delta[\mathrm{P}]}{\Delta t} = F_\mathrm{u} + F_\mathrm{c}[\mathrm{M}] \tag{13}$$

where F_u and F_c are the conditional constants of uncatalyzed and catalyzed reactions, involving the concentrations of the reactants, respectively. These equations are the basis of determining the concentration of the catalyst.

5.1 Tangent and Initial Rate Methods

The rate of a reaction, v, corresponds to the concentration of the catalyst as exemplified by Equation (14)

$$v = \frac{\mathrm{d[P]}}{\mathrm{d}t} = F_\mathrm{u} + F_\mathrm{c}[\mathrm{M}] \tag{14}$$

The change in a physical property of a monitored species is observed continuously and recorded versus time (Figure 3a). The slope of the resulting straight line, $\tan\alpha$, is a measure of the catalyst concentration, known as the tangent method. The plot of $\tan\alpha$ versus the concentration of catalyst is employed as a calibration graph (Figure 3b). The initial rate of the reaction is also measured from the slope of kinetic curves; this approach is called the initial rate method. Although the measurement of the initial rate requires the kinetic curve, it has some advantages:

1. The reaction obeys pseudo-zero order kinetics with respect to the reactants, because these concentrations do not change appreciably.
2. The back reaction which contributes to the rate of overall reaction is virtually nonexistent, because the amount of product formed during the initial period is very small.
3. Complications arising from any side reactions are minimized during this period. Thus, the reproducibility and the selectivity of the initial rate procedure

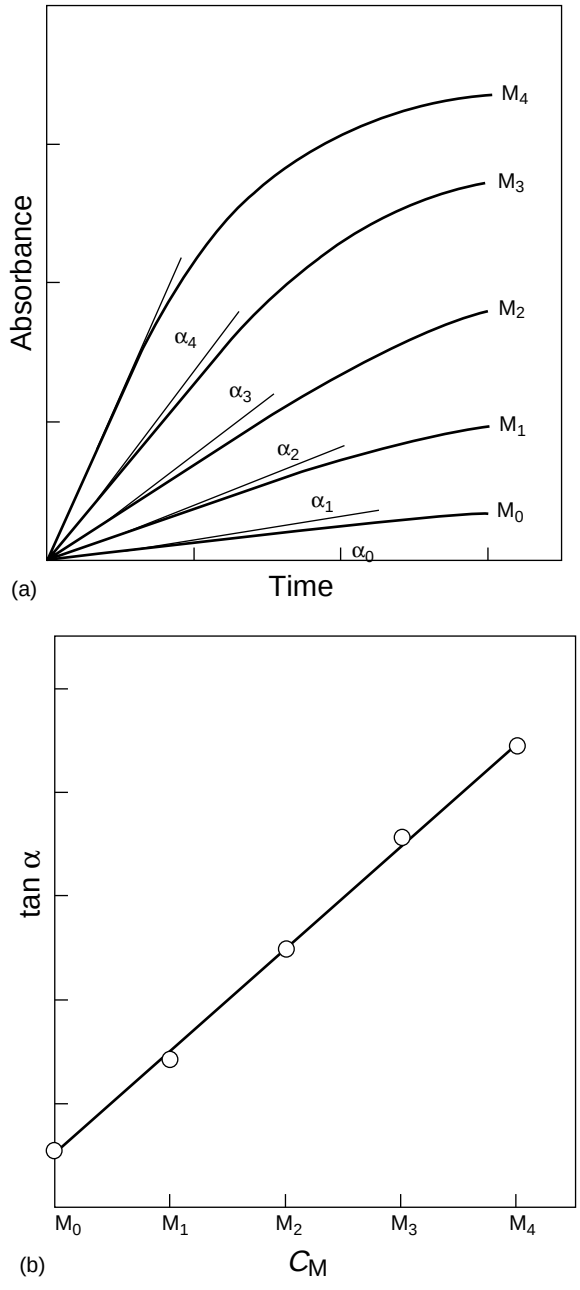

Figure 3 Absorbance – time curves (a) and calibration curve for catalyst determination by the tangent method (b).

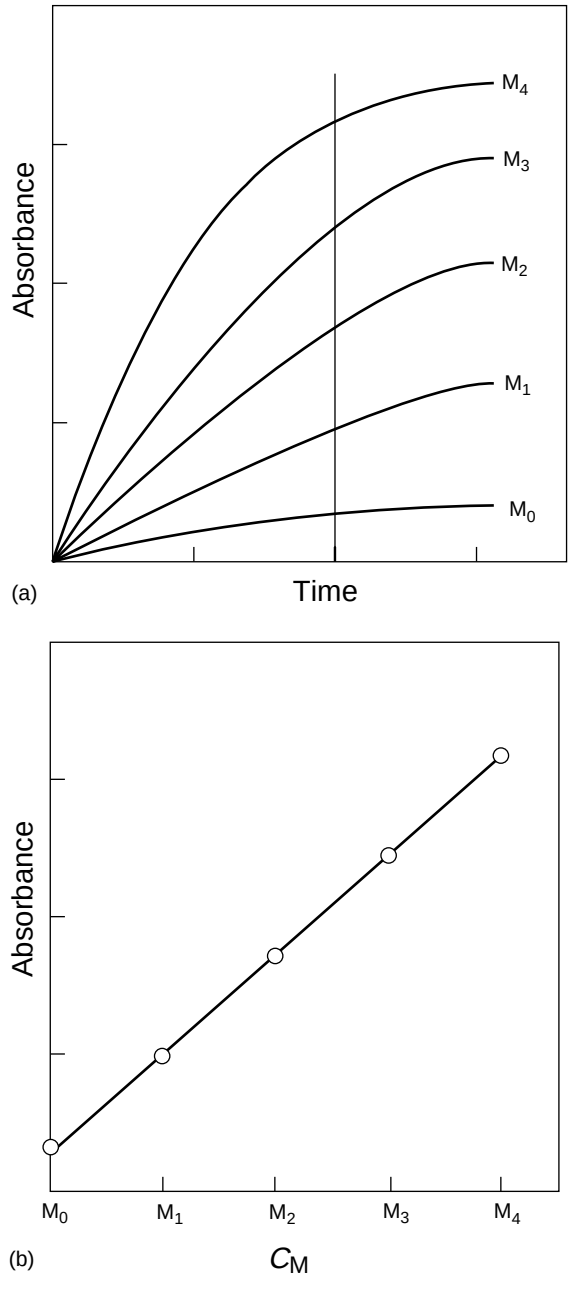

Figure 4 Absorbance – time curves (a) and calibration curve for catalyst determination by the fixed time method (b).

are better than those of a fixed time procedure. In addition, the use of the initial rate as a parameter can be obtained over a wider determinable range, i.e. two or three orders of magnitude.

5.2 Fixed Time Methods

This method involves the determination of a physical property corresponding to the concentration of a reactant or a product at a definite time after the initiation of the reaction. For the fixed-time method, Equation (13) can be rewritten as Equation (15):

$$\Delta[\mathrm{P}] = (F_\mathrm{u} + F_\mathrm{c}[\mathrm{M}])\Delta t \qquad (15)$$

Figure 4(a) shows the kinetic curves in the absence and the presence of the catalyst and Figure 4(b) shows a plot of an absorbance at a fixed time versus the concentration of the catalyst. From the latter curve (calibration curve), the concentration of the catalyst can be determined.

5.3 Variable Time Methods

The time required to attain a constant change in composition may be measured. For the variable time method, Equation (13) can be rewritten as Equation (16):

$$\frac{1}{\Delta t} = \frac{F_u + F_c[M]}{\Delta[P]} \tag{16}$$

From the kinetic curves shown in Figure 5(a), a plot of $1/\Delta t$ as a function of [M] (Figure 5b) is obtained and is used for the determination of the catalyst.

5.4 Measurements of Induction Period

Some reactions used in the catalytic methods are characterized by the occurrence of induction periods. Reactions of this type generally proceed in two or more stages. The product of the first slow stage behaves as reactants for the second and final stages. The concentration of the catalyst is obtained from the following relationships, Equations (17) and (18)

$$[M] = \frac{\text{const}}{t_i} \tag{17}$$

$$[M] = \frac{\text{const}}{t_i^2} \tag{18}$$

where const denotes a conditional constant for a given reaction and t_i is an interval time of an induction period. Figure 6(a) shows typical absorbance versus time curves and Figure 6(b) shows a plot of $1/t_i$ versus catalyst concentrations.

5.5 Analysis of Mixtures by Differential Rates

Differential reaction rate methods, based on the different rate at which two or more species interact with a common reagent, permit analysis of two or more species without prior separation.[63] Although these methods are based on an uncatalyzed reaction, they are also applied to the determination of mixtures acting as catalyst for a given catalyzed reaction. Differential rate methods are classified into two categories depending on the approach used:[26,27]

1. those based on graphical evaluations (e.g. logarithmic extrapolation method), and
2. those based on mathematical complications (e.g. proportional equation method).

An example is the simultaneous determination of Ru and Os based on their catalysis in the reaction of Ce(IV) with As(III).[64] With concentrations of 5×10^{-3} M Ce(IV) and 1×10^{-3} M As(III), conditions prevailing

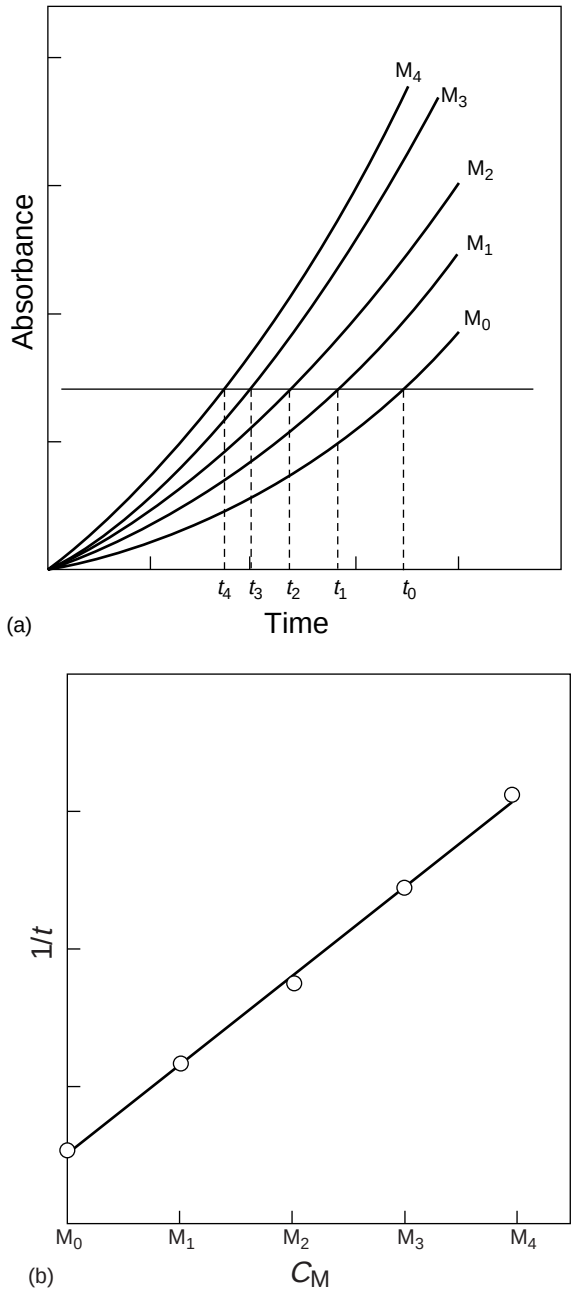

Figure 5 Absorbance – time curves (a) and calibration curve for catalyst determination by the variable time method (b).

for selectivity for ruthenium, the initial reaction rate of the catalyzed reaction in 2 M H_2SO_4 is given by Equation (19)

$$v_1 = 240[RuO_4] + 90.4[OsO_4] \tag{19}$$

With concentrations of 5×10^{-4} M Ce(IV) and 1×10^{-3} M As(III), the latter being added first, and conditions prevailing for selectivity for osmium, the resulting rate

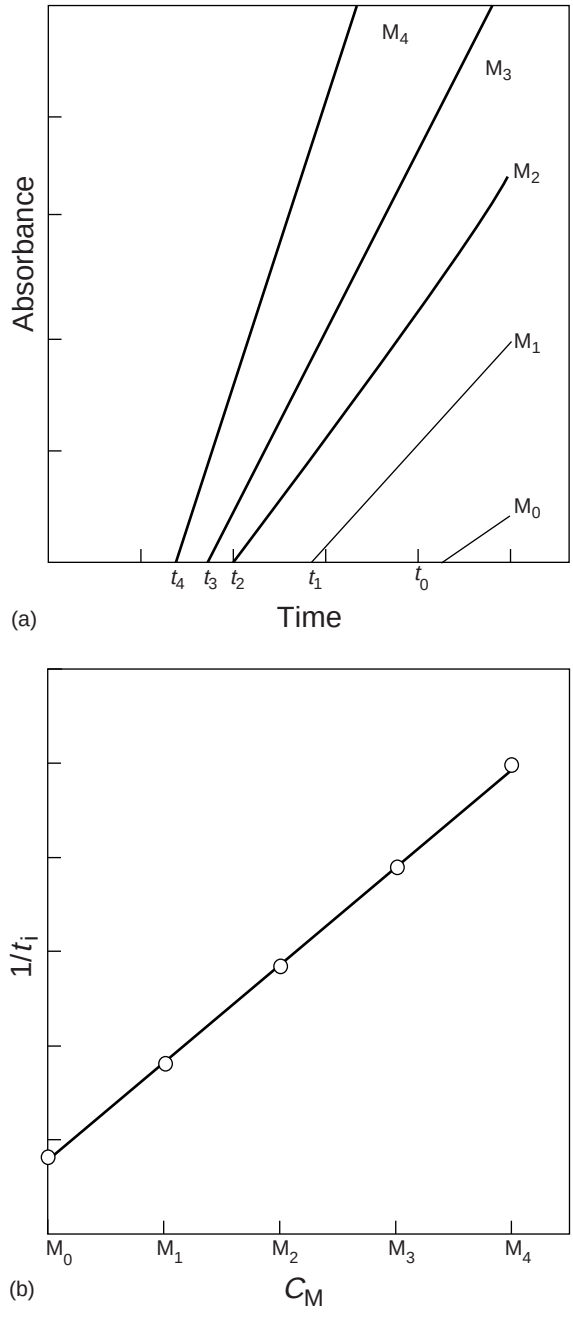

Figure 6 Absorbance – time curves (a) and calibration curve for catalyst determination by the measurement of induction period (b).

equation is expressed as Equation (20):

$$v_2 = 2.69[RuO_4] + 78.0[OsO_4] \qquad (20)$$

By measuring the rates under these two sets of conditions, mixtures of RuO_4 and OsO_4 in the range 2×10^{-10}–1×10^{-7} M can be determined with relative standard deviations ranging 0.51–4.1%.

5.6 Flow-injection Analysis

Since kinetic methods include time as an experimental variable, care is needed in the mixing of reagents at regular time intervals to obtain highly accurate results. Furthermore, batch methods are sometimes time-consuming and laborious. Such disadvantages in the batch method can be overcome by using FIA, which is firmly established as a fast precise accurate efficient and extremely versatile analytical technique. The FIA method is a suitable technique for catalytic methods of analysis because the rate of the catalyzed reaction can easily be controlled by fixing the flow rate of the solutions and the length of the reaction coil, yielding reproducible mixing. The fundamental characteristic of FIA is its intrinsically kinetic nature. The adaptation of FIA to catalytic methods can lead to many advantages, namely rapid and reproducible mixing of sample and reagents solutions can be achieved in the closed flow system without any contamination from the environment and a large number of samples can be treated automatically in a short time.[65] FIA procedures for trace elements using of kinetic–catalytic methods have been reviewed.[21]

6 TITRIMETRIC METHODS WITH CATALYTIC END-POINT INDICATION

In 1960, Erdey and Buzas[66] first reported the titration of EDTA with copper(II) employing copper(II)-catalyzed luminol and lucigenin CL as indicator reactions which were later studied in depth by Mottola.[67,68] Since the appearance of this work, papers on titrimetric methods with catalytic end-point indication (catalytic titration) for the determination of inorganic and organic compounds have increased. Since 1960, about 200 papers on catalytic titrimetric procedures have been published for the determination of catalysts and other species by photometric, potentiometric, biamperometric and thermometric techniques.

The methodology involves two consecutive reactions, i.e. titration and indicator reactions. The titration reaction is expressed as Equation (21):

$$L + M \longrightarrow ML \qquad (21)$$

where L, M and ML are the titrated component (inhibitor), titrant (catalyst) and titration product (catalytically inactive), respectively.

As described in section 2, the indicator reaction is expressed as Equation (2):

$$S + R \longrightarrow P + Q \qquad (2)$$

A small excess of the titrant M added after the equivalence point catalyzes the indicator reaction and this

catalyzed reaction is used to detect the end-point of the titration. As previously noted, rapid regeneration of M by Equation (4) results in the catalytic formation of the product P, and thus trace amounts of M lead to marked color changes in the end-point of the solution titrated. It is thus possible to determine trace amounts of the catalyst M when a small excess of the titrant M is added to the known amount of the titrated component such as EDTA.

Some examples of the titrimetric methods for the determination of metals (catalyst) or inhibitors are compiled in Table 2.[67,69–81] Other interesting approaches are briefly reviewed below.

The oxidation of 1,4-dihydroxyphthalimide dioxime (DHPO) by iodate is catalyzed by nickel(II); the oxidation product has an absorption maximum at 510 nm. The indicator reaction is used for the determination of inhibitory substances such as EDTA and CDTA, and also for the indirect titration of nickel(II), mercury(II) and iron(III).[82] Amounts of EDTA and CDTA in the range 700–6000 μg are determined.

A semiautomatic spectrophotometric method of end-point indication for indirect catalytic titrations of alkaline earth ions has been developed.[83] The manganese(II)-catalyzed autoxidation of 1,4-dihydroxyphthalimide dithiosemicarbazone is used as an indicator reaction. Calcium and magnesium (10–150 μg), strontium (20–250 μg), and barium (20–450 μg) can be determined by adding a known excess of EDTA to the sample solution and back-titrating the unconsumed inhibitor agent with standard manganese(II) solution. This method is applied to the determination of the total concentration of calcium and magnesium (total hardness) in natural waters and milk.

Gaal et al.[84] investigated the applicability of the conductometric detection to catalytic end-point indication of neutralization titrations. Conductometric and high-frequency conductometric methods are used for the titration of sodium acetate, as well as some organic bases (antipyrine, aminopyrine, caffeine, etc.) with perchloric acid by using the acetic anhydride–methanol or acetic anhydride–formic acid system which is catalyzed by excess titrant (perchloric acid).

A simple catalytic end-point indication has been developed for total concentrations of heavy metal ions which form more stable EDTA complexes than manganese(II).[85] The method is based on the substitution reaction of manganese(II)–EDTA and heavy metal ions to produce manganese(II). The end-point is determined by the manganese(II)-catalyzed oxidation of Acid Blue 45 dye (1,5-dihydroxy-4,8-diaminoanthraquinone-2,6-disulfonate) with hydrogen peroxide as a result of decoloration. The method is applicable to the determination of heavy metal ions in river and groundwater samples.

A semiautomatic spotting method is used for end-point detection.[86] In this method metal ion catalysts in the range 3–70 $\mu g\,mL^{-1}$ are titrated with EDTA. The necessary assembly consists of a titration vessel, mixing chamber, automatic syringe burette, PTFE (polytetrafluoroethylene)-coated belt (a moving band) and connection tubes. In an example copper(II) is visually determined by using a copper-catalyzed hydroquinone/hydrogen peroxide reaction in the presence of bpy as activator. The titration vessel is completely filled with a sample solution of copper(II) (avoiding gas bubbles). The reagent solutions for the indicator reaction are delivered to the mixing chamber from two automatic syringe burettes at a constant speed. After a few drops of the reagent, solutions leave the mixing chamber for the belt (which is not moving yet), the EDTA titrant is propelled to the titration vessel at a constant speed by the syringe burette. Simultaneously, the PTFE-coated belt starts to move at a constant speed. Because the titration vessel is completely filled with the sample solution, excess solution containing EDTA and copper(II) is pushed out of the titration vessel into the mixing chamber where the catalyzed reaction occurs. Before the end-point, the drops on the moving belt are red–brown caused by formation of products of the copper-catalyzed reaction; after the end-point, the drops are colorless. The first drop which shows a color change corresponds to the end-point. Calibration graphs are prepared by plotting titration time versus copper concentration.

The catalytic titration of aminopolycarboxylic acids (EDTA, GEDTA, CDTA and DTPA) using manganese(II) as a titrant and the manganese-catalyzed periodate–diethylaniline reaction, which acts as an indicator reaction, has been developed to demonstrate the usefulness of flow injection pseudotitration in catalytic titrimetry.[87] The titrant and each reagent for the indicator reaction are mixed in a 100-cm mixing coil and the mixing chamber, and then allowed to react in a 3-m reaction coil. When an aminopolycarboxylic acid sample solution (200 μL) is injected into the reacting flowstream, the catalysis is inhibited by the formation of the manganese–aminopolycarboxylic acid complex, and the uncatalyzed reaction is monitored, resulting in the transmittance being increased from 3% to 75%. The FIA peak width depends on the aminopolycarboxylic acid concentration in the range 5×10^{-6}–1×10^{-2} M.

Raya-Saro and Perez-Bendito[73,74] reported a method for the simultaneous catalytic titration of binary mixtures of mercury(II) with copper(II) or cadmium at the micromolar level by using the oxidation of DBPT by hydrogen peroxide as an indicator reaction. Copper(II) and mercury(II) ions act as catalyst and inhibitor for the indicator reaction, respectively. The combination of the

Table 2 Titrations with catalytic end-point indication

Indicator reaction	Analyte	Titrant	Titrand	Detection technique	Dynamic range	Application	Ref.
Ascorbic acid + O_2	Cyanide	Cu(II)	Cyanide	Photometry	10^{-7} M level		67
Malachite green + IO_4^-	EDTA	Mn(II)	EDTA	Photometry	0.54×10^{-5}–3.21×10^{-5} M		69
	CDTA	Mn(II)	CDTA	Photometry	0.50×10^{-5}–3.03×10^{-5} M		69
	DTPA	Mn(II)	DTPA	Photometry	1.00×10^{-5}–3.00×10^{-5} M		69
	EDTA-OH	Mn(II)	EDTA-OH	Photometry	1.02×10^{-5}–3.06×10^{-5} M		69
Phenolphthalein + H_2O_2	Cu(II)	Cu(II)	EDTA	Photometry	4.6×10^{-3}–7.9×10^{-3} M		70
	Mn(II)	Mn(II)	EDTA	Photometry	6.9×10^{-3}–8.7×10^{-3} M		70
o-Dianisidine + H_2O_2	Ascorbic acid	Cr(VI)	Ascorbic acid	Photometry	1.3×10^{-4}–1.3×10^{-3} M		72
DBPT + H_2O_2	EDTA	Cu(II)	EDTA	Photometry	370–1100 µg		73
	GEDTA	Cu(II)	GEDTA	Photometry	380–1140 µg		73
	Cu(II)[a]	Cu(II)	EDTA	Photometry	10–120 µg		73
	Ni(II)[a]	Cu(II)	EDTA	Photometry	10–120 µg		73
	Mn(II)[a]	Cu(II)	EDTA	Photometry	10–120 µg		73
	Fe(III)/Cu(II)[a,b]	Cu(II)	EDTA or GEDTA	Photometry	0.2–3.0 µmol	Alloy	77[c]
	Fe(III)/Ni(II)[a,b]	Cu(II)	EDTA or GEDTA	Photometry	0.2–3.0 µmol	Steel and alloy	77[c]
	Fe(III)/Mn(II)[a,b]	Cu(II)	EDTA or GEDTA	Photometry	0.2–3.0 µmol	Alloy	77[c]
	Cystine/cysteine[b]	DBPT	Hg(II), Cu(II)	Photometry	µmol levels	Urine	78[d]
Acetaldehyde	Cu(II)	OH^-	Cu(II)	Thermometry	0.1 mequiv.		75[e]
	Ni	OH^-	Ni	Thermometry	0.1 mequiv.		75[e]
	Cr(III)	OH^-	Cr(III)	Thermometry	0.1 mequiv.		75[e]
	Fe(III)	OH^-	Fe(III)	Thermometry	0.1 mequiv.		75[e]
	Al	OH^-	Al	Thermometry	0.1 mequiv.		75[e]
	Co	OH^-	Co	Thermometry	0.1 mequiv.		75[e]
	Mn(II)	OH^-	Mn(II)	Thermometry	0.1 mequiv.		75[e]
	Cd	OH^-	Cd	Thermometry	0.1 mequiv.		75[e]
	Zn	OH^-	Zn	Thermometry	0.1 mequiv.		75[e]
	Pb(II)	OH^-	Pb(II)	Thermometry	0.1 mequiv.		75[e]

$HBTS + H_2O_2$	EDTA	Fe(III)	EDTA	Fluorimetry	1.25×10^{-5}–3.75×10^{-5} M		76
	Fe(III)[a]	Fe(III)	EDTA	Fluorimetry	20–70 μg		76
	Ca(II)[a]	Fe(III)	EDTA	Fluorimetry	20–100 μg		76
	Mg(II)[a]	Fe(III)	EDTA	Fluorimetry	20–100 μg		76
	Sr(II)[a]	Fe(III)	EDTA	Fluorimetry	20–140 μg		76
	Ba(II)[a]	Fe(III)	EDTA	Fluorimetry	20–100 μg		76
Mn(III) + As(III)	Ag(I)	I^-	Ag(I)	Thermometry	0.5–500 μg		71
	Hg(II)	I^-	Hg(II)	Thermometry	0.5–500 μg		71
	Pd(II)	I^-	Pd(II)	Thermometry	0.2–500 μg		71
Ce(IV) + As(III)	Au(III)	I^-	Au(III)	Potentiometry	30 μg mL^{-1}	Pharmaceutical preparations	79
$I^- + H_2O_2$	F^-	Th(IV)	F^-	Potentiometry	16–32 μg mL^{-1}	Pharmaceutical preparations	80
$S_2O_8^{2-}$ + sulfanilic acid	Br^-	Ag(I)	Br^-	Photometry	10–20 μmol[f]	Pharmaceutical preparations	81
$S_2O_8^{2-}$ decomposition	Br^-	Ag(I)	Br^-	Potentiometry	10–20 μmol[f]	Pharmaceutical preparations	81
H_2Q + acetic anhydride	Br^-	ClO_4^-	Br^-	Thermometry	10–20 μmol[f]	Pharmaceutical preparations	81

CDTA, *trans*-1,2-cyclohexanediamine-*N*,*N*,*N*′,*N*′-tetraacetic acid; DTPA, diethylenetriamine-*N*,*N*,*N*′,*N*″,*N*″-pentaacetic acid; EDTA-OH, *N*-(2-hydroxyethyl)ethylenediamine-*N*,*N*′,*N*′-triacetic acid; GEDTA, *O*,*O*′-bis(2-aminoethyl)ethyleneglycol-*N*,*N*,*N*′,*N*′-tetraacetic acid; HBTS, 2-hydroxybenzaldehyde thiosemicarbazone; H_2Q, hydroquinone.

[a] By the indirect back-titration method.

[b] Binary mixture.

[c] Two sample aliquots are necessary. Triethanolamine or fluoride was used as a masking agent for iron(III) in the determination of copper, nickel or manganese only.

[d] Copper(II) and mercury(II) act as a catalyst and an inhibitor, respectively. Cysteine has a counter-inhibitory effect on mercury. Cystine is reduced to cysteine by hydroxylamine. Two catalytic titration runs, in the absence and presence of hydroxylamine, are needed for each sample.

[e] In this method free mineral acids have been simultaneously determined with turbidimetric end-point indication.

[f] Amounts of the active ingredients (e.g. hyoscine butyl bromide, pyridostigmine bromide and neostigmine bromide).

catalytic effect of copper(II) and the inhibitory effect of mercury(II), along with the blocking of the catalytic cycle by EDTA are used as the basis for titrimetric methods for individual and simultaneous titrations of mercury and copper or cadmium, with catalytic end-point detection.

A gaseous catalyst (hydrogen sulfide) is also used for the end-point indication in titrimetric analysis of metal ions with the hydrogen sulfide-catalyzed iodine–azide reaction as an indicator reaction.[88] Sodium sulfide can be used as a standard solution for the determination of metal ions such as antimony, nickel, iron, mercury plus cadmium, lead plus copper and silver plus copper (precipitation) and permanganate (redox reaction) both in acidic solution. Once the metal ions are fully precipitated or reduced by sodium sulfide, hydrogen sulfide is liberated and transferred by a stream of nitrogen gas into the separate indicator vessel containing the solution of sodium azide, iodine and starch. Thus the color of the solution disappears at the end-point.

7 KINETIC–CATALYTIC METHODS FOR TRACE ELEMENTS

A large number of publications devoted to kinetic–catalytic determination of trace elements that obtain higher sensitivity and better selectivity have been reported, with a marked trend toward its application to environmental, clinical, pharmaceutical and agricultural samples. The determination of a single species in a mixture is the most popular area of the kinetic–catalytic method. Relatively few methods are reported for the simultaneous determination of two components. This section describes the kinetic–catalytic determination for trace elements using redox and ligand substitution reactions.

7.1 Catalytic Methods Using Redox Reactions

Inorganic and organic redox reactions are extensively used for kinetic–catalytic methods. These methods focus on the determination of transition metal ions such as copper, cobalt, manganese, iron and vanadium. Tables 3–9[53–56,58,61,62,89–263] show selected methods

Table 3 Kinetic–catalytic methods for determination of copper

Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application	Ref.
DPD + DMA + H_2O_2	NH_3	Photometry	Fixed time	0.2–2		62
	NH_3	Photometry	Fixed time	0.05–1	Water	89
	DTAB or SDS	Photometry	Fixed time	0.8–6		61
HBS + PDA + H_2O_2	Acetate	Photometry	Fixed time	0.1–2	Water	90
p-Anisidine + DMA + H_2O_2	NH_3	Photometry	Fixed time	2–20	Water	91
PPDA + DMA + H_2O_2	NH_3	Photometry	Fixed time	0.4–4	Water	92
+ PDA + H_2O_2	NH_3/Py/Tween 80	Photometry	FIA	0.2–30	Water	93
DPKH + O_2		Fluorimetry	FIA	8–30		94
MBTH + DMA + H_2O_2	bpy	Photometry	Tangent	0.1–5	Water	95
	NH_3	Photometry	Fixed time	0.1–1.2	Water	55
	bpy	Photometry	FIA	0.05–50		96
+ DAOS + H_2O_2	Py	Photometry	Fixed time	0.002–0.1	Human hair	97
Chromotropic acid + H_2O_2	Py	Photometry	Fixed time	12.7–190.6	Water	98
HBAA + $S_2O_8^{2-}$		Photometry	Initial rate	0.5–18	Water and wine	99
Hydroxylamine + O_2		Thermometry	Tangent	50–500		100
DBPT + H_2O_2		Photometry	Initial rate	3.5–25	Serum	101
L-Ascorbic acid + O_2 + *o*-PDA		Fluorimetry	Fixed time	0.2–8	Water	102
Hydroquinone + H_2O_2	Py	Photometry	Segmented FIA	0.4–1		103
	Py	Photometry	FIA	5–100	Water	104
$S_2O_3^{2-}$ + Fe(III)		Photometry	Closed-loop FIA	250–20 000	Blood serum	105
		Biamperometry	FIA	200–800	Blood plasma	106
$SbCl_6^-$ + NH_2OH		Potentiometry	Batch	5–510	Zn-metal, ores and water	107
FMN + H_2O_2		CL	FIA	0.6–6000	Pepperbush and sediment	108
Fluorescein + NH_2OH		CL	Batch	1–20	Serum	109

DPKH, 2,2′-dipyridylketone hydrazone; HBAA, 3-hydroxybenzaldehyde azine; *o*-PDA, *o*-phenylenediamine; FMN, flavin mononucleotide; Py, pyridine.

Table 4 Kinetic–catalytic methods for determination of cobalt

Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application	Ref.
Diphenylcarbazone + H_2O_2	DMSO	Photometry	Tangent	0.001–0.01		110
Chromotropic acid + H_2O_2		Photometry	FIA	2–100		111
Tiron + H_2O_2		Photometry	FIA	0.005–0.2	Water/cation exchanger	112
		Photometry	FIA	0.03–0.5	Seawater	113
PPDA + H_2O_2	Tiron	Photometry	FIA	0.02–0.2	Seawater	114
DPDA + H_2O_2	Tiron	Photometry	FIA	0.01–0.1	Seawater	115
Pyrogallol Red + H_2O_2		Photometry	Fixed time/ Variable time/Initial rate	0.2–13		116
Protocatechuic acid + H_2O_2		Photometry	FIA	0.005–0.08	Water/HQ column/cation exchanger	117
Fluorescein + H_2O_2	Tiron	Fluorimetry	Fixed time	0–0.24	Vitamin B_{12}	118
Amide black 10B + H_2O_2	Tiron	Photometry	Fixed time	5–30	Ni compound	119
AHHNS + H_2O_2	Tiron	Photometry	Fixed time	0.2–1.2		120
DHBA + H_2O_2		Thermometry	Tangent	1–16	Vitamin B_{12}	121
L-Adrenaline + H_2O_2		Photometry	Initial rate Fixed time	0.47–4.7	Urine	122
Co(III)-Br-PAPS + HSO_5^-		Photometry	Induction period	0.06–74		123
MBTH + DMA + H_2O_2	Tiron/HCO_3^-/ SDS	Photometry	FIA	0.04–0.5	Biological	124
SO_3^{2-} + O_2		Thermometry	Tangent	5–30		125
Gallic acid + H_2O_2		CL	FIA	0.0005–0.012	Water/HQ column	126
Pyrogallol Red + H_2O_2	CTAB	CL	Batch	0.05–0.4	Water	127
Luminol + H_2O_2		CL	FIA	0.001–1000	Water/IC	128
+ IO_4^-		CL	FIA	0.01–12	Water	129
Phen + H_2O_2	CTAB	CL	FIA	1–100		130
THCPF + H_2O_2	CTAB	CL	Batch	0.5–6	Water	131

Tiron, 1,2-dihydroxybenzene-3,5-disulfonic acid; DPDA, *N*,*N*′-diethyl-*p*-phenylenediamine; AHHNS, 1-amino-8-hydroxy-7-(*p*-hydrophenylazo)-3,6-naphthalenedisulfonic acid; DHBA, 3,4-dihydroxybenzoic acid, Br-PAPS, 2-(5-bromo-2-pyridylazo)-5-(*N*-propyl-*N*-sulfopropylamino)phenolate; THCPF, 2,6,7-trihydroxy-9-(4′-chlorophenyl)-3-fluorone; DMSO, dimethyl sulfoxide; CTAB, cetyltrimethylammonium bromide; HQ, 8-hydroxyquinoline; 8-quinolinol; IC, ion chromatography. Other abbreviations as in Table 3.

based on catalyzed-redox reactions. Of these, spectrophotometry occupies a predominant position because of the wide use of colored organic compounds.

When the reactants of the indicator reaction have high absorptivity for monitoring the reaction rate, it is preferable to combine the catalyzed reaction with a second coupling reaction yielding a colored product. An oxidative coupling reaction of *p*-anisidine with DMA in the presence of hydrogen peroxide is used as an indicator reaction for the kinetic–catalytic determination of Fe(III)[53] and Cu(II).[91] That is, *p*-anisidine reacts with DMA to form Variamine Blue B leuco base analog, MDP. MDP is oxidized to a blue compound *N*-(*p*-methoxyphenyl)-*N*′,*N*′-dimethyl-1,4-benzoquinonediiminium ion (MDB) by trace amounts of Fe(III) and Cu(II) in the presence of H_2O_2. By using the reactants separately, the blank values can be kept low, decomposition and oxidation of the product during storage being avoided. Similarly, an oxidative coupling reaction of MBTH with DMA in the presence of H_2O_2 is used as an indicator reaction for the catalytic determination of Cu(II),[55,95,96] Cr(III),[58] Mn(II),[54,154] Co(II)[124] and Fe(II, III)[149] at nanograms per milliliter levels, Equations (22) and (23)

$$\text{MBTH} + \text{DMA} + \text{M}^{(n+1)+} \longrightarrow \text{indamine dye} + \text{M}^{n+} \quad (\lambda_{max} = 590\,\text{nm}) \qquad (22)$$

$$\text{M}^{n+} \xrightarrow{\text{H}_2\text{O}_2} \text{M}^{(n+1)+} \qquad (23)$$

In these reaction systems, the use of activators allows an increase in the sensitivity. Among them, catalytic effects of Cr(III) and Co(II) in the presence of activators are worthy of remark. EDTA acts as an activator for Cr(III), because the Cr(III)–EDTA complex catalyzes the decomposition of H_2O_2. On the other hand, Cr(VI) has hardly any catalytic action in this reaction because of the formation of a peroxochromium(VI) complex and slow production of Cr(III) from this complex.[58] In the Co(II)-catalyzed reaction system, the extremely high activating effect is obtained by using two activators,

Table 5 Kinetic–catalytic methods for determination of iron

Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application	Ref.
p-Anisidine + DMA + H_2O_2		Photometry	Fixed time	2–20	Chemicals	132
	Phen	Photometry	Fixed time	1–14	Water	53
PPDA + DMA + H_2O_2	Acetate	Photometry	Fixed time	0.6–6	Water	133
	Acetate/Tween 80	Photometry	FIA	0.5–6	Water	134
Leuco malachite green + H_2O_2	Phen	Photometry	FIA	100–1000		135
DABA + H_2O_2		Photometry	Tangent	1.2–11	Water	136
Chromotropic acid + H_2O_2		Photometry	Tangent	11–167	Water	137
DPD + H_2O_2		Photometry	Fixed time	0.04–3.2	Metals and water	138
Sulfanilic acid + IO_4^-	Phen	Photometry	Tangent	1–55		139
	Phen	Thermometry	Tangent	50–220	Water	140
	Phen/CPC	Photometry	Initial rate	50–220	Human serum	141
Pyrogallolphthalein + $S_2O_8^{2-}$	Triton X	Photometry	Fixed time	1–60	Water	142
Reduced Rhodamine B + H_2O_2	SCN^-	Photometry	Fixed time	0.04–150	Water	143
Hydroxylamine + O_2	EDTA	Photometry	FIA	3.5–150	Water	144
AA + DMA + H_2O_2	Acetate	Photometry	Fixed time	0.2–4	Water	145
Chlorpromazine + H_2O_2		Photometry	Tangent	5–200	Water	146
Pyrogallol-5-sulfonate + H_2O_2		Photometry	Tangent	2–75	Water	147
p-Phenetidine + IO_4^-	Phen	Photometry	FIA	0.7–500	Water	148
MBTH + DMA + H_2O_2		Photometry	Initial rate	1–1000	Water	149
Brilliant sulfoflavine + H_2O_2		CL	FIA	0.05–50	Water/cation exchanger	150
Luminol + H_2O_2		CL	FIA	0.001–0.1	Water/HQ column	151
Phen + H_2O_2	Zephiramine	CL	FIA	0.1–1		152

DABA, 3,5-diaminobenzoic acid; CPC, cetylpyridinium chloride. Other abbreviations as in Tables 3 and 4.

i.e. Tiron and hydrogen carbonate.[124] In addition, the sensitivity of the method with FIA is improved in the presence of SDS at concentrations above its CMC.[124] The peak height in the presence of SDS is about eight times as high as that without it. By using modified Trinder's reagents instead of DMA as a coupling agent for MBTH, the catalytic effect of Cu(II) is examined.[97] It is found that DAOS is the most sensitive reagent for Cu(II) and pyridine is selected as the most effective activator for Cu(II) in this system. A Cu(II) concentration as low as 10^{-11} M can easily be determined by using the fixed time procedure, and the method has been successfully applied to the copper determination in real samples.

The oxidative coupling of PPDA with DMA or PDA in the presence of an oxidizing agent is utilized for the catalytic determination of Cu(II),[92,93] Fe(II, III)[133,134] and V(V)[174] with manual and FIA procedures. For the determination of Cu(II) and Fe(II, III), hydrogen peroxide is used as an oxidant. Activators for Cu(II) and Fe(II, III) are ammonia and acetate, respectively. In the FIA determination, Tween 80 acting as a nonionic surfactant improves the sensitivity and sampling frequency. Concentrations of Cu(II) and Fe(II, III) at 10^{-9}–10^{-8} M levels can be determined from the increase in the absorbance of the product at a fixed time. For the V(V) determination, bromate and tartrate are used as an oxidant and an activator, respectively.

The oxidative coupling of AA with DMA in the presence of bromate is used for the catalytic determination of V(IV, V) in the manual[178,179] and FIA[57] mode. In this reaction system, SSA and Tiron act as activators. The combination of this method and solvent extraction of vanadium with *N*-cinnamoyl-*N*-(2,3-xylyl)hydroxylamine (CXA) allows the differential determination of V(IV) and V(V) in natural waters.[179]

Fluorimetry is an effective tool for monitoring the rate of reaction. However, these are fewer catalytic fluorimetric methods than spectrophotometric ones. Valcarcel et al.[153] have reported catalytic methods based on fluorimetric reactions; the H_2O_2 oxidation of PPH is adopted for the determination of Mn(II). Applying the tangent, fixed time and fixed intensity procedures, Mn(II) concentration in the range 0.05–8 ng mL^{-1} can be determined.

Several kinetic–catalytic methods for the determination of catalysts have been developed by using thermometric methods. Although the sensitivity of the methods is lower than that of spectrophotometric and fluorimetric detections, thermometric monitoring has other advantages; the presence of colored compounds and/or precipitates in the samples does not affect the thermometric detection. On the basis of Co(II)-catalyzed oxidation of DHBA by H_2O_2, a thermometric determination of Co(II) is developed with the aid of a personal

Table 6 Kinetic–catalytic methods for determination of manganese

Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application	Ref.
PPH + H_2O_2		Fluorimetry	Fixed time /Tangent/Variable time	0.05–8	Milk, cream and cheese	153
MBTH + DMA + O_2 + H_2O_2	Phen, bpy	Photometry	Fixed time	0.1–2	Water	54
		Photometry	FIA	4–30	Biological materials	154
Succinimide dioxime + O_2		Photometry	FIA	0.2–1300	Chemicals and food	155
N,*N*-Diethylaniline + IO_4^-		Photometry	FIA	0.02–1.0	Water	156
		Photometry	FIA	0.01–15	Seawater	157
Malachite green + IO_4^-	NTA	Photometry	Stopped flow	0.1–15		158
		Photometry	FIA	0.01–0.07	Chemicals and Ti-metals	159
		Photometry	Fixed time	0.2–2	Chinese medicine	160
Tiron + H_2O_2	Phen	Photometry	Stopped flow FIA	5.5–22	Water	161
	Phen	Thermometry	Initial rate	1–120	Water, beer and wine	162
Photoactivated oxidation of SO_3^{2-}	Acac/Rose Bengal	Potentiometry	Tangent	0.3 (DL)	Water	163
AHHNS + H_2O_2	Phen or Tiron	Photometry	Fixed time	0.04–1	Water	164
DPD+PDA + H_2O_2	Trien + Tiron	Photometry	FIA	0.05–1	Water	165
Alizarin S + H_2O_2		Thermometry	Initial rate	10–100	Brass	166
Rhodamine 6G + IO_4^-	NTA	Fluorimetry	Fixed time	0.04–1	Hair, urine, fish and water	167
Rhodamine B + IO_4^-	Phen	Photometry	Fixed time	0.1–5	Chinese tea leaves	168
DHN + ethylenediamine + H_2O_2		Fluorimetry	Fixed time	0.010–0.14	Water	169
DHBA + H_2O_2		Photometry	FIA	0.5–15	Solar salts/cation exchanger	170
3,3′,5,5′-Tetramethylbenzidine + IO_4^-		Photometry	Fixed time	0.06–2	Water	171
Decomposition of MnO_4^-		CL	Batch	0.1–10	Bauxite	172
Luminol + IO_4^-		CL	FIA	0.02–9	Water	130

PPH, pyridoxal 2-pyridylhydrazone; DHN, 2,3-dihydroxynaphthalene; NTA, nitrilotriacetic acid; Acac, acetylacetone. Other abbreviations as in Tables 3–5.

computer system.[121] The calibration graph is linear in the range 1–16 ng mL^{-1} Co(II) and the method is applied to the pharmaceutical sample.

Thermal lens spectrometry (TLS) is 100–1000 times more sensitive than conventional spectrophotometry and is used for monitoring the rate of the bromate oxidation of aniline catalyzed by V(V) in the presence of 8-quinolinol or pyrocatechol as an activator.[189] The DL of the TLS method is 100 times lower than that of the spectrophotometric method.

The oxidation of CL reagent with a suitable oxidant has been applied to the kinetic–catalytic determination of transition metal ions by many workers. CL methods have some advantages in simplicity of instrumentation, low limits of detection and a wide dynamic range. Further, the use of a CL reaction with FIA techniques allows rapid and reproducible determination. However, the application of the method to real samples is difficult because of its lack of specificity. Several separation procedures such as ion exchange, solvent extraction and a membrane phase separator are used prior to the determination of a catalyst as well as the use of a masking agent. A luminol CL reaction is utilized for the determination of transition metal ions by FIA with separation processes.[128,151] The CL oxidation of gallic acid with H_2O_2 is adopted for the flow-injection determination of Co(II) with an in-line column of immobilized 8-quinolinol.[126] The method enables the determination of picomolar levels of Co(II) in seawater. The oxidation of brilliant sulfoflavine with H_2O_2 is adapted for the CL–FIA determination of Fe(II).[150] A cation column is used to preconcentrate the iron in seawater. The DL of the method is 0.45×10^{-9} M when 4.4 mL of sample is passed through the column.

There are several methods for the determination of anions containing nonmetals, although most catalytic determinations are intended to determine various metals.[26,27] Iodide ion can be determined by its catalytic effect on the redox reaction between As(III) and Ce(IV)

Table 7 Kinetic–catalytic methods for determination of vanadium

Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application	Refs.
HBS + NA + ClO_3^-	Tartrate	Photometry	Fixed time	10–100		173
+ PDA + ClO_3^-		Photometry	Fixed time	4–40		56
PPDA + DMA + BrO_3^-	Tartrate	Photometry	Fixed time	0.1–1	Water	174
Gallic acid + BrO_3^-		Photometry	Fixed time	0.03–1	Water/Ion exchanger	175
Chlorpromazine + BrO_3^-		Photometry	Tangent	10–400	Petroleum products	176
	Tartaric acid	Photometry	Initial rate	0.2–150	Water	177
AA + DMA + BrO_3^-	SSA	Photometry	Fixed time	0.05–1	Water	178, 179
	Tiron	Photometry	FIA	0.05–2	Water	57
Chromotropic acid + BrO_3^-		Thermometry	Tangent	0.02–2.5	Petroleum derivatives	180
		Photometry	FIA	0.02–2.5	Water/Sephadex column	181
p-Phenetidine + catechol + BrO_3^-	CPC	Photometry	Tangent	0.05–10		182
DABA + BrO_3^-		Photometry	Tangent	3–50	Water and steel	183
DHPO + BrO_3^-		Photometry	Tangent/Fixed time/Variable time	10–400	Particulate matter and serum	184
Pyrogallol Red + BrO_3^-		Thermometry	Initial rate	10–150	Steel	185
Gallocyanine + BrO_3^-		Photometry	Fixed time	0.01–500	Petroleum, food and steel	186
Perphenazine + BrO_3^-	Citric acid	Photometry	Tangent	0.08–6.5	Water	187
o-PDA + BrO_3^-	Gallic acid	Fluorimetry	Fixed time	0.002–0.8	Water	188
Aniline + BrO_3^-	HQ or pyrocatechol	Thermal lens	Tangent	0.05–5		189
Indigo carmine + BrO_3^-		Photometry	Induction period/Tangent	300–3000	Petroleum and milk powder	190
Cinchomeronic hydrazide + H_2O_2		CL	Batch	40–1000	Steel	191
Luminol + $Fe(CN)_6^{4-}$		CL	FIA	10–10000	Geochemical sample and human hair	192

NA, 1-naphthylamine; SSA, 5-sulfosalicylic acid. Other abbreviations as in Tables 3–6.

in sulfuric acid solution.[3,4] This reaction proceeds very slowly, but traces of iodide increase the rate of the reaction. The rate of the reaction is proportional to the iodide concentration: the time required for the yellow color of Ce(IV) to disappear is inversely proportional to the amounts of iodide present in solution. The determination of iodide based on this reaction has been used extensively by many workers.[3,4,264–268] This reaction is also catalyzed by Os,[64,269–272] Ru[64,273] and Cr[272] in the same way as iodide.

The catalytic effect of nitrite on the bromate oxidation of phenosafranine is used for the fluorimetric determination in the range 0.9–1.38 ng mL^{-1} nitrite using the manual procedure.[254] This reaction system is applied to the FIA in the stopped-flow mode; the dynamic range of the FIA method is 46–460 ng mL^{-1}. The oxidative coupling of PPDA with DMA in the presence of bromate is also utilized for the nitrite determination; nitrite at the amounts of 2–100 ng mL^{-1} in river water can be determined using the FIA mode.[261]

Many differential kinetic methods based on uncatalyzed reactions have been proposed for the analysis of mixtures involving closely related compounds without prior separation. By using catalyzed reactions, a few methods for the simultaneous determination of two catalysts are reported; they are listed in Table 10.[274–286] A method for the differential determination of Mo(VI) and W(VI) based on their different catalytic effects on the oxidation of chloropromazine by H_2O_2 has been proposed.[284] The initial rates at two different H_2O_2 concentrations and the corresponding rate equations are obtained; mixtures of Mo(VI) and W(VI) in the 10^{-7} M range can be determined. The method is applied to the

Table 8 Kinetic–catalytic methods for determination of elements

Element	Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range ($ng\,mL^{-1}$)	Application	Ref.
Ag(I)	PCAPH + $S_2O_8^{2-}$	Phen	Fluorimetry	Tangent	4–100	Panchromatic plates	193
	Pyrogallol Red + $S_2O_8^{2-}$	Phen	Photometry	Tangent	0.85–21		194
	Gallocyanine + $S_2O_8^{2-}$	Phen	Photometry	FIA	0.3–4000	Water	195
Au(III)	MDPASA + V(V)		Photometry	Tangent/Fixed time	250–15000	Plating electrolyte	196
Cr(III)	*o*-Dianisidine + H_2O_2		Photometry	Fixed time	5–20		197
	MBTH + DMA + H_2O_2	EDTA	Photometry	Fixed time	0.5–10		58
	Sb(III) + $Fe(CN)_6^{3-}$		Photometry	Initial rate	5.2–1140		198
	Luminol + H_2O_2		CL	FIA	0.01–6	Water and food	199
	Pyrogallol + IO_4^-		CL	FIA	5–100		200
Cr(VI)	*o*-Dianisidine + H_2O_2		Photometry	Tangent	2–40	Salts	201
	Lophine + H_2O_2		CL	FIA	20–3000	Water/Activated alumina	202
Hg(II)	Photo-oxidation of EDTA + I^-		Amperometry	Variable time	0.4–42	Zinc and lens cleaning solutions	203
	Safranine + I^-		Photometry	FIA	500–9000	Human hair	204
Ir(IV)	MDPASA		Photometry	Tangent	0.02 (DL)	Synthetic mixture	205
Mo(VI)	I^- + H_2O_2	CPC	Photometry	Tangent	2–150	Soil	206
			Photometry	FIA	0–100	Plants and food	207
	$S_2O_3^{2-}$ + H_2O_2		Thermometry	Initial rate	5–60	Vegetables	208
	L-Ascorbic acid + H_2O_2 + *o*-PDA		Fluorimetry	Fixed time	0.04–3	Water	209
	Nile Red + hydrazine		Photometry	Fixed time	50–30 000	Plants, steel alloy	210
	Toluidine Blue + Sn(II)	Phen	Photometry	Fixed time	8–770		211
	Thionine + hydrazine		Photometry	Fixed time	50–4000	Steel alloy/Cation exchanger	212
Ni(II)	DPC + H_2O_2	DMSO	Photometry	Tangent	1.8×10^4–3×10^5		210
	PH + $S_2O_8^{2-}$		Fluorimetry	Tangent	50–300		213
	Decomposition of MnO_4^-		Thermometry	Initial rate	100–700	Petroleum derivatives	214
	Acetone + H_2O_2 + ClO^-		CL	Batch	10–1000	Water	215

(*continued overleaf*)

Table 8 (*continued*)

Element	Indicator reaction	Activator or surfactant	Detection technique	Procedure	Dynamic range ($ng\,mL^{-1}$)	Application	Ref.
Os(IV, VIII)	Pyrogallol Red + BrO_3^-		Photometry	Tangent	0.65–1400	MIBK extraction	216
	Gallocyanine + BrO_3^-		Photometry	Fixed time	0.1–1200	Synthetic sample	217
	Salicylfluorone + H_2O_2		Fluorimetry	Fixed time	0.008–0.6	Synthetic sample and ore	218
Pb(II)	Resazurin + S^{2-}		Photometry	Stopped flow FIA	1–100	Water	219
	DEAPH + H_2O_2		CL	Batch	0.05–5		220
Pd(II)	Purpurin + H_2O_2		Photometry	Tangent	50–2500	Synthetic sample	221
Pt(IV)	DPKH + O_2		Fluorimetry	Initial rate	200–600		222
Re(IV)	BABAT + Sn(II)		Photometry	Variable time	10–1000	Copper sandstone	223
Rh(II, IV)	Cu(II) + IO_4^-		Photometry	Tangent	0.5–10	Ore	224
Ru(III, VIII)	MDPASA + IO_4^-		Photometry	Fixed time	0.05–0.5	Synthetic sample	225
	p-Ethoxychrysoidine + IO_4^-		Photometry	Tangent	0.2–3	Synthetic sample	226
	Fe(II)-phen + IO_4^-		Photometry	FIA	10–120		227
Se(IV)	TAN + PO_2^{2-}		Photometry	Fixed time	5–30		228
	HBS + NA + ClO_3^-		Photometry	Fixed time	10–120		229
	HBS + PDA + ClO_3^-		Photometry	Fixed time	8–80		230
	TNBT + dithiothreitol		Photometry	Fixed time	2–1000		231
	MTT + dithiothreitol		Photometry	FIA	0.004–3.8	Lobster/ODS column	232
	Methylene blue + S^{2-}		Photometry	Fixed time	200–1500	Water and bird fodder corrector	233
	Methylene blue + S^{2-}	CTAB	Photometry	Induction period	47–505	Fly ash	234
	Gallocyanine + S^{2-}		Photometry	Fixed time	10–500	Shampoo	235
Tc(VII)	Methylene blue + Sn(II)		Photometry	Initial rate	120–1400	Synthetic sample and seawater	236
Te(IV)	Toluidine blue + S^{2-}		Photometry	FIA	80–3800	Water	237
Ti(IV)	PANH + O_2		Fluorimetry	FIA	4–120		238
U(IV)	Phosphomolybdic acid + I^-		Photometry	Fixed time	20–12000	Chemicals	239
W(VI)	Chlorpromazine + H_2O_2		Photometry	Initial rate	2–10000	Water	240

PCAPH, pyrocatechol-1-aldehyde 2-pyridylhydrazone; MDPASA, *N*-methyldiphenylamine-4-sulfonic acid; PH, pyridoxal hydrazone; DEAPH, 4-diethylaminophthalohydrazide; BABAT, benzeneazobenzeneazothiocyanate; TAN, 1,4,6,11-tetraazanaphthacene; TNBT, tetranitro blue tetrazolium; PANH, picolinealdehyde nicotinoyl hydrazone. Other abbreviations as in Tables 3–7.

Table 9 Kinetic–catalytic methods for determination of anions

Anion	Indicator reaction	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application	Ref.
Br^-	PCV + H_2O_2	Photometry	Fixed time	4–300	Water	241
		Photometry	FIA	10–600	Water	242
CN^-	*p*-Nitrobenzaldehyde + *o*-dinitrobenzene	Photometry	FIA	10–50	Water	243
I^-	Chlorpromazine + BrO_3^- + H_2O_2	Photometry	Tangent	5–70	Salts and rat thyroid	244
		Photometry	Initial rate	0.2–10	Water	245
		Photometry	Tangent/ MA method	0.5–200 0.1–7		246
	Tetrabase + Chloramine T	Photometry	FIA	0.1–10	Water	247
	Promethazine + H_2O_2	Photometry	Tangent	0.1–12	Water	248
	As(III) + Ce(IV)	Photometry	Intersection time	10–100	Table salt	249
	$SCN^- + NO_2^-$	Photometry	FIA	0.4–9	Water	250
	$SbCl_6^- + NH_2OH$	Potentiometry	Tangent	10–400	Iodinated salt	251
NO_2^-	PAPH + BrO_3^-	Photometry	Initial rate	40–4000	Water	252
	Thionine + BrO_3^-	Photometry	Fixed time	0.3–55	Water and sausage	253
	Phenosafranine + BrO_3^-	Fluorimetry	Initial rate Stopped flow FIA	0.9–13.8 62–620	Water and food	254
	Pyrogallol Red + BrO_3^-	Photometry	Fixed time	3–1000	Water	255
	Brilliant Cresyl blue + BrO_3^-	Photometry	Fixed time	0.5–1000	Water, sausage and salt	256
	Chlorpromazine + H_2O_2	Photometry	MA method	3–1500	Water	257
	Chlorophosphonazo-pN + BrO_3^-	Photometry	FIA	50–1000	Water and soil	258
	Prochlorperazine + BrO_3^-	Photometry	Initial rate	0.8–70	Water	259
	Thymol blue + BrO_3^-	Photometry	Fixed time	0.4–1500	Water	260
	PPDA + DMA + BrO_3^-	Photometry	FIA	2–100	Water	261
S^{2-}	Toluidine Blue + Se(IV)	Photometry	Fixed time	280–1700	Water	262
	Azide + I	Photometry	Fixed time	20–500	Water and steel	263

Tetrabase, 4,4′-bis(dimethylamino)diphenylmethane; PAPH, pyridine-2-aldehyde 2-pyridylhydrazone. Other abbreviations as in Tables 3–8.

determination of these ions in hot spring water. The simultaneous determination of these ions in a single kinetic run is of interest.

Principal component regression (PCR) has been employed in data processing of multicomponent analysis. Zhu et al.[285] have developed a simultaneous determination of Os(VIII) and Ru(IV) based on their catalytic action on the reaction between As(III) and Ce(IV), applying PCR to calculate the concentration of each catalyst.

7.2 Catalytic Methods Using Ligand Substitution and Complex Formation Reactions

The catalytic determinations of some metal ions and inorganic anions using ligand substitution and complex formation reactions are shown in Table 11.[50,287–303] The ligand substitution reaction of the Hg(II) complex of CAC with CDTA (Equation 24) is used for the catalytic determination of trace amounts of ammonia.[295]

$$Hg^{II}\text{–CAC} + \text{CDTA} \longrightarrow Hg^{II}\text{–CDTA} + \text{CAC} \quad (24)$$

The rate equation for Equation (24) in the presence of ammonia is expressed as Equations (25) and (26):

$$-\frac{d[Hg^{II}\text{–CAC}]}{dt} = k_0[Hg^{II}\text{–CAC}] \quad (25)$$

and

$$k_0 = \frac{k_1' + k_2'[H^+] + k_3'[H^+][NH_3]}{[\text{CAC}][\text{CDTA}]} \quad (26)$$

where k_1', k_2' and k_3' refer to conditional rate constants. This reaction path, corresponding to the third term on the right-hand side of Equation (26), is expressed as Equations (27) and (28)

$$Hg\text{–CAC} + NH_3 \longrightarrow HgNH_3 + \text{CAC} \quad (27)$$

$$HgNH_3 + \text{CDTA} \longrightarrow Hg\text{–CDTA} + NH_3 \quad (28)$$

This provides the basis of the determination of ammonia. From the absorbance of Hg^{II}–CAC complex at a fixed time, the concentration of ammonia at micromolar levels can be determined. Interferences from cationic and anionic ions are avoided by distillation of ammonia at room temperature.

Table 10 Simultaneous determination of catalysts

Catalysts	Indicator reaction	Detection technique	Procedure	Dynamic range (ng mL^{-1})	Application/ additive	Ref.
Ag(I)–Hg(II)	$[Fe(CN)_6]^{4-}$ + bpy	Photometry	FIA	1–64/Ag(I) 0.5–75/Hg(II)	Water/Ligand substitution	274
Cu(II)–Hg(II)	DPKH/DPDKPH + H_2O_2	Fluorimetry	Stopped flow FIA	2000–5120/Cu(II) 1000–6000/Hg(II)		275
Cr(VI)–W(VI)	$I^- + H_2O_2$	Photometry	Initial rate	15–300/Cr(VI) 5–100/W(VI)		276
Fe(II)–Mn(II)	SAT + H_2O_2	Fluorimetry	FIA	40–500/Fe(II) 40–600/Mn(II)		277
Fe(II)–Sb(III)	Cr(VI) + I^-	Photometry	Stopped flow FIA	0–3500/Fe(II) 0–3700/Sb(III)	Water and zinc	278
Fe(II)–V(V)	Chromotropic acid + BrO_3^-	Photometry	FIA	150–1500/Fe(III) 3–700/V(V)	Food	279
Mn(II)–Pb(II)	Tiron + H_2O_2	Photometry	Tangent	0.5–10/Mn(II) 50–2000/Pb(II)	Activating effect of Pb(II)	280
Mo(VI)–W(VI)	DAP + H_2O_2	Photometry	Tangent	96–480/Mo(VI) 184–919/W(VI)	Steel	281
	$I^- + H_2O_2$ (Starch)	Photometry	Stopped flow FIA	0–1100/Mo(VI) 0–2000/W(VI)	Steel	282
		Photometry	FIA	19–307/Mo(VI) 37–588/ W(VI)	Standard mineral sample	283
	Chlorpromazine + H_2O_2	Photometry	Initial rate	9.6–77/Mo(VI) 18–147/W(VI)	Water	284
Os(VIII)–Ru(IV)	As(III) + Ce(IV)	Photometry	Tangent	2.3–13/Os(VIII) 1–5.8/Ru(IV)		285
Br^-–I^-	PCV + H_2O_2	Photometry	Initial rate	10–100/Br^- 5–50/I^-	Water	286

DPDKPH, dipyridyldiketone phenylhydrazone; SAT, salicylaldehyde thiosemicarbazone; DAP, 2,4-diaminophenol; PCV, pyrocatechol violet. Other abbreviations as in Tables 3–9.

The ligand substitution of Hg^{II}–4-(2-pyridylazo)resorcinol (H_2PAR) complex of CDTA is catalytically accelerated by a trace amount of iodide, Equation (29)

$$Hg(PAR)_2^{2-} + HCDTA^{3-} + H^+ \longrightarrow Hg(CDTA)^{2-} + 2HPAR^- \quad (29)$$

According to detailed studies of the kinetics and mechanism of the reaction,[300] $Hg(PAR)_2^{2-}$ dissociates to Hg(PAR) which then reacts with iodide to form a mixed ligand complex, $Hg(PAR)I^-$, which is more reactive with CDTA than Hg(PAR). The rate-determining step is the dissociation of the PAR molecule from Hg(PAR), $Hg(PAR)OH^-$ and $Hg(PAR)I^-$. Iodide and hydroxide ions enhance the rate of the dissociation of PAR molecule by their electron donation. The rate equation of Equation (29) is given as follows, Equations (30) and (31):

$$-\frac{d[Hg(PAR)_2^{2-}]}{dt} = k_0[Hg(PAR)_2^{2-}][HCDTA^{3-}] \quad (30)$$

and

$$k_0 = (k_1'[H^+] + k_2' + k_3'[H^+][I^-])[HPAR^-]^{-1} \quad (31)$$

From Equation (31), the conditional constant k_0 is a linear function of the iodide concentration when PAR is present in large excess and the pH is constant. By measuring the absorbance change at 500 nm, up to 2×10^{-7} M iodide can be determined. The method is applied to rainwater after separation of interfering ions using a cation exchanger. From the kinetic investigation of the ligand substitution reaction between the Hg^{II}–PAR complex and CDTA together with the catalytic effect of some inorganic ligands on this reaction, the difference

Table 11 Kinetic–catalytic methods for determination of elements using ligand substitution reactions

Catalyst	Indicator reaction	Detection technique	Procedure	Dynamic range or detection limit (ng mL^{-1})	Application/ additive	Ref.
Ag(I)	$Fe(CN)_5(NH_3)^{3-}$ + PDTS	Photometry	Fixed time	20–500		287
	$Fe(CN)_6^{4-}$ + INH	Photometry	Fixed time	30–150		288
	Hg(II)-EDTA + Cu(II)-$DDTC_{org}$	Photometry	SEFIA	22–86		289
Au(III)	$Fe(CN)_5(NH_3)^{3-}$ + PDTS	Photometry	Fixed time	100–3000		287
Cd(II)	Mn(II) + TPPS	Photometry	Fixed time	5.6–56	Pb(II) is removed by coprecipitation	290
Hg(II)	$Fe(CN)_5(NH_3)^{3-}$ + PDTS	Photometry	Fixed time	10–400		287
	Mn(II) + TPPS	Photometry	Fixed time	0.2		291
	$Fe(CN)_6^{4-}$ + *p*-NDA	Photometry	Fixed time	20–200		50
	$Fe(CN)_6^{4-}$ + INH	Photometry	Fixed time	120–480		288
	$Fe(CN)_6^{4-}$ + TARPO	Photometry	Fixed time	10–900	Water	292
	$Fe(CN)_6^{4-}$ + Phen	Photometry	Tangent	10–80	SDS	293
Pb(II)	Mn(II) + TPPS	Photometry	Fixed time	20	Water/Cd(II) is masked by cyanide	294
NH_3	Hg(II)-CAC + CDTA	Photometry	Fixed time	0–66	Water/distillation	295
F^-	Zr(IV) + XO	Photometry	Fixed time	5–100	Water/CTAC Interfering ions are separated	296
	Zr(IV) + XO or MTB	Photometry	FIA	10–2000	Water	297
HCO_3^-/CO_3^{2-}	Cr(III) + Chromazurol S	Photometry	Tangent	500–10 000	Salt	298
	Cr(III) + EDTA	Photometry	FIA	1×10^4–3×10^5	Water	299
I^-	Hg(II)-PAR + CDTA	Photometry	Fixed time	0–25	Water/cation-exchanger	300
PO_4^{3-}	Zr(IV) + XO	Photometry	Fixed time	50–300	Water/CTAC Interfering ions are separated	301
SO_3^{2-}	Hg(II)-PAR + CDTA	Photometry	Fixed time	2.4–120		302
$S_2O_3^{2-}$	Hg(II)-PAR + CDTA	Photometry	Fixed time	2.2–78		302
SO_4^{2-}	Zr(IV) + XO	Photometry	Fixed time	5×10^2–1×10^4	Water/CTAC	303

PDTS, 3-(2-pyridyl)-5,6-diphenyl-1,2,4-triazine-*p*,*p*′-disulfonic acid; INH, isonicotinoyl hydrazine; $DDTC_{org}$, diethyldithiocarbamic acid in organic solvent; TPPS, 5,10,15,20-tetra(*p*-sulfonatophenyl)porphine; TARPO, 4-(2′-thiazolylazo)resacetophenone oxime; CAC, 3,3′-bis[*N*,*N*-bis(carboxymethyl)aminomethyl]*o*-cresolphthalein; XO, xylenol orange; MTB, methylthymol blue; PAR, 4-(2-pyridylazo)resorcinol; CTAC, cetyltrimethylammonium chloride; SEFIA, solvent extraction flow injection analysis. Other abbreviations as in Tables 3–10.

in reactivity between sulfite or thiosulfate and sulfate is established; the differential kinetic determinations of sulfite and thiosulfate are designed from the observations. Sulfite up to 1.5×10^{-6} M and thiosulfate up to 7×10^{-7} M can be determined in the presence of 10^4 and 2.5×10^4-fold molar amounts of sulfate under the optimized conditions.[302]

Catalytic determinations of Ag(I) and Hg(I) are proposed using ligand substitutions of cyano and aminocyano complexes of iron with chromogenic organic ligands such as phen,[293] PDTS[287] and INH.[288]

A metal exchange reaction between ethylenediaminetetraacetatomercury(II) (Hg^{II}–EDTA) in aqueous phase and bis(diethyldithiocarbamato)copper(II) (Cu^{II}–DDTC) in the organic phase is catalyzed by trace amounts of Ag(I), Equation (32).[289]

$$Cu^{II}\text{–}DDTC_{org} + Hg^{II}\text{–}EDTA \longrightarrow Hg^{II}\text{–}DDTC_{org} + Cu^{II}\text{–}EDTA \quad (32)$$

where org denotes the organic phase. By measuring the decrease in the absorbance of the Cu^{II}–DDTC complex, Ag(I) at 10^{-7} M levels can be determined with the flow-injection mode.

Complex formations of metal ions with porphyrine derivatives are attractive reactions for kinetic-based methods. The rate of the incorporation of Mn(II) into TPPS is very slow, so that little spectral change is observed even 60 min after initiation of the reaction at 25 °C.[290] In the presence of Hg(II) or Cd(II), a rapid spectral

change occurs. Mercury(II) also accelerates the complex formation of H_2TPPS with Ni(II), Co(II) and Cu(II). In the metalloporphyrin formation reaction catalyzed by metal ions, the ionic radius (r) of the metal ion acting as a catalyst is an important factor. Large metal ions such as Cd^{2+} ($r = 95$ pm), Cu^{+} ($r = 96$ pm), Pb^{2+} ($r = 73$ pm) and Hg^{2+} ($r = 114$ pm) give rise to an appreciable catalytic effect, while medium-sized ions like Cu^{2+} ($r = 73$ pm) and Zn^{2+} ($r = 75$ pm) exert hardly any catalytic effect.

In the formation of Mn^{II} (TPPS) catalyzed by Cd(II), the following reaction paths are proposed from the mechanical study,[290] Equations (33–35):

$$\mathrm{Mn(II)} + \mathrm{TPPS} \longrightarrow \mathrm{Mn^{II}TPPS\ (slow)} \quad (33)$$

$$\mathrm{Cd(II)} + \mathrm{TPPS} \rightleftharpoons \mathrm{Cd^{II}TPPS} \quad (34)$$

$$\mathrm{Cd^{II}TPPS} + \mathrm{Mn(II)} \longrightarrow \mathrm{Mn^{II}TPPS} + \mathrm{Cd(II)\ (fast)} \quad (35)$$

Cadmium(II), having a large ionic radius, will distort the configuration of the porphine core which favors the attack of porphine nucleus by Mn(II) from behind (Equation 34). After the incorporation of Mn(II) (Equation 35), Cd(II) is liberated and takes part in the formation of Mn^{II}TPPS. On the basis of this consideration, the catalytic determination of Cd(II) has been developed: as little as 10^{-7} M Cd(II) can be determined from the decrease in absorbance of TPPS using a fixed time procedure.[290] Since lead(II) interferes with the determination, it needs to separated from Cd(II) by coprecipitation with manganese(IV) oxide.

Mercury(II) and Pb(II) also act as catalysts for the complex formation of TPPS with Mn(II) (Equation 33) in the same way as Cd(II). The catalytic methods for the determination of Hg(II) and Pb(II) are constructed using this complex formation. After separation of metallic mercury by distillation, concentrations of Hg(II) as low as 10^{-8} M can be determined.[291] To determine Pb(II), interferences from Hg(II) and Cd(II) should be avoided by the addition of hydroxylamine and cyanide.[294] Lead(II) concentrations at 10^{-7} M levels can be determined.

Although the rate of complex formation of Cr(III) with a ligand such as chromazurol S[298] and EDTA[299] is slow, this reaction is catalytically accelerated in the presence of carbonate and/or bicarbonate. By using these systems, the anions can be determined by spectrophotometry. The catalytic effect of fluoride[296,297] and sulfate[303] on the complex formation of Zr(IV) with XO is utilized for the determination of these anions.

The vast development in kinetic methods is a result of improvements in the sensitivity and selectivity brought about by the use of catalyzed reactions. These have facilitated the determination of microconstituents in various types of samples by straightforward and inexpensive instrumentation.

The number of papers devoted to kinetic photometric determination continues to be high, with a marked trend towards the application of the methods development to real sample analyses. Methods focus on the determination of transition metal ions like copper, manganese and iron, and are based on their catalytic effects on redox reactions. Relatively few methods for the determination of nonmetals and organic compounds are reported.

ACKNOWLEDGMENTS

This article was inspired by Professor A. Meyers and the editors of *Encyclopedia of Analytical Chemistry*. During the writing of this manuscript, the constant encouragement and valuable advice of the editors have been given. We also wish to thank Managing Editor, Dr Adam Holyoake, for his encouragement and constructive comments on this manuscript.

ABBREVIATIONS AND ACRONYMS

AA	4-Aminoantipyrine
Acac	Acetylacetone
AHHNS	1-Amino-8-hydroxy-7-(*p*-hydrophenylazo)-3,6-naphthalenedisulfonic Acid
BABAT	Benzeneazobenzeneazothiocyanate
bpy	2,2′-Bipyridine
Br-PAPS	2-(5-Bromo-2-pyridylazo)-5-(*N*-propyl-*N*-sulfopropylamino)phenolate
CAC	3,3′-Bis[*N*,*N*-bis(carboxymethyl)aminomethyl]*o*-cresolphthalein
CDTA	*Trans*-1,2-cyclohexanediamine-*N*,*N*,*N*′,*N*′-tetraacetic Acid
CL	Chemiluminescence
CMC	Critical Micelle Concentration
CPC	Cetylpyridinium Chloride
CTAB	Cetyltrimethylammonium Bromide
CTAC	Cetyltrimethylammonium Chloride
CXA	*N*-Cinnamoyl-*N*-(2,3-xylyl)hydroxylamine
DABA	3,5-Diaminobenzoic Acid
DAOS	*N*-Ethyl-*N*-(2-hydroxy-3-sulfopropyl)-3,5-dimethoxyaniline
DAP	2,4-Diaminophenol
DBPT	4,4′-Dihydroxybenzophenone Thiosemicarbazone
DDTC	Diethyldithiocarbamic Acid

DEAPH	4-Diethylaminophthalohydrazide
DHBA	3,4-Dihydroxybenzoic Acid
DHN	2,3-Dihydroxynaphthalene
DHPO	1,4-Dihydroxyphthalimide Dioxime
DL	Detection Limit
DMA	*N*,*N*-Dimethylaniline
DMSO	Dimethyl Sulfoxide
DNPO	Bis(2,4-dinitrophenyl) Oxalate
DPC	Diphenylcarbazone
DPD	*N*,*N*-Dimethyl-*p*-phenylenediamine
DPDA	*N*,*N*′-Diethyl-*p*-phenylenediamine
DPDKPH	Dipyridyldiketone Phenylhydrazone
DPKH	2,2′-Dipyridylketone Hydrazone
DTAB	Dodecyltrimethylammonium Bromide
DTPA	Diethylenetriamine-*N*,*N*,*N*′,*N*″,*N*″-pentaacetic Acid
EDTA	Ethylenediamine-*N*,*N*,*N*′,*N*′-tetraacetic Acid
EDTA-OH	*N*-(2-Hydroxyethyl)ethylenediamine-*N*,*N*′,*N*′-triacetic Acid
FIA	Flow-injection Analysis
FMN	Flavin Mononucleotide
GEDTA	*O*,*O*′-Bis(2-aminoethyl)ethyleneglycol-*N*,*N*,*N*′,*N*′-tetraacetic Acid
HBAA	3-Hydroxybenzaldehyde Azine
HBS	*p*-Hydrazinobenzenesulfonic Acid
HBTS	2-Hydroxybenzaldehyde Thiosemicarbazone
HQ	8-Hydroxyquinoline; 8-Quinolinol
IC	Ion Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
INH	Isonicotinoyl Hydrazine
MA	Maximum Absorbance
MBTH	3-Methyl-2-benzothiazolinone Hydrazone
MDB	*N*-(*p*-Methoxyphenyl)-*N*′,*N*′-dimethyl-1,4-benzoquinonediiminium Ion
MDP	(Variamine Blue leuco base analog) *N*-(*p*-methoxyphenyl)-*N*′,*N*′-dimethyl-*p*-phenylenediamine
MDPASA	*N*-Methyldiphenylamine-4-sulfonic Acid
MTB	Methylthymol Blue
MTT	3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl-2*H*-tetrazolium Bromide
NA	1-Naphthylamine
p-NDA	*p*-Nitrosodiphenylamine
NTA	Nitrilotriacetic Acid
PANH	Picolinealdehyde Nicotinoyl Hydrazone
PAPH	Pyridine-2-aldehyde 2-Pyridylhydrazone
PAR	4-(2-Pyridylazo)resorcinol
PCAPH	Pyrocatechol-1-aldehyde 2-pyridylhydrazone
PCR	Principal Component Regression
PCV	Pyrocatechol Violet
PDA	*m*-Phenylenediamine
o-PDA	*o*-Phenylenediamine
PDTS	3-(2-Pyridyl)-5,6-diphenyl-1,2,4-triazine-*p*,*p*′-disulfonic Acid
PH	Pyridoxal Hydrazone
phen	1,10-Phenanthroline
PPDA	*N*-Phenyl-*p*-phenylenediamine
PPH	Pyridoxal 2-Pyridylhydrazone
PTFE	Polytetrafluoroethylene
Py	Pyridine
SAT	Salicylaldehyde Thiosemicarbazone
SDS	Sodium Dodecylsulfate
SEFIA	Solvent Extraction Flow Injection Analysis
SSA	5-Sulfosalicylic Acid
TAN	1,4,6,11-Tetraazanaphthacene
TARPO	4-(2′-Thiazolylazo)resacetophenone Oxime
TCPO	Bis(2,4,6-trichlorophenyl) Oxalate
THCPF	2,6,7-Trihydroxy-9-(4′-chlorophenyl)-3-fluorone
TLS	Thermal Lens Spectrometry
TNBT	Tetranitro Blue Tetrazolium
TPPS	5,10,15,20-Tetra(*p*-sulfonatophenyl)-porphine
Trien	Triethylenetetraamine
XO	Xylenol Orange

POPULAR NAMES

Acid blue 45	1,5-Dihydroxy-4,8-diamino-anthraquinone-2,6-disulfonate
Amide black 10B	1-Amino-8-hydroxy-2-(*p*-nitrophenylazo)-7-(phenylazo)-3,6-naphthalene-3,6-disulfonic Acid
4-Aminoantipyrine	4-Amino-1,5-dimethyl-2-phenyl-3-pyrazolone
p-Anisidine	4-Methoxybenzenamine; 4-Methoxyaniline
Arizarin S	3,4-Dihydroxyanthraquinone-2-sulfonic Acid
Brilliant cresyl blue	2-Methyl-3-amino-7-diethylaminophenazonium Chloride
Brilliant sulfoflavine	4-Amino-*N*-(*p*-tolyl)naphthalimide-3-sulfonate
Chloramine T	*N*-Chloro-*p*-toluenesulfonamide Hydrochloride
Chlorophosphonazo-pN	2-(4-Chloro-2-phosphonophenylazo)-7-(4-nitrophenylazo)-1,8-dihydroxynaphthalene-3,6-disulfonic Acid

Chromazurol S	2,6-Dichloro-4′-hydroxy-3′,3″-dimethyl-3-sulfofuchsone-5′,5″-dicarboxylic Acid
Chromotropic Acid	1,8-Dihydroxy-3,6-naphthalenedisulfonic Acid
Chlorpromazine	2-Chloro-10-(3-dimethylaminopropyl)phenothiazine
Cinchomeronic hydrazide	1,4-Dione-2,3-dihydropyrido[3,4-*d*]pyridazine
o-Dianisidine	3,3′-Dimethoxy-4,4′-diaminobiphenyl
Diphenylcarbazone	Phenylazoformic Acid 2-Phenylhydrazide
Dithiothreitol	*Threo*-2,3-Dihydroxy-1,4-dithiolbutane
p-Ethoxychrysoidine	4-Ethoxy-2′,4′-diaminoazobenzene Hydrochloride
Fluorescein	9-(2′-Carboxyphenyl)-6-hydroxy-3-isoxanthenone
Gallic Acid	3,4,5-Trihydroxybenzoic Acid
Gallocyanine	1-Carboxy-7-(dimethylamino)-3,4-dihydroxyphenoxazine-5-ium Chloride
Indigo carmine	5,5′-Indigotindisulfonic Acid
Lophine	2,4,5-Triphenylimidazole
Lucigenin	10,10′-Dimethyl-9,9′-biacridinium Dinitrate
Luminol	5-Amino-2,3-dihydro-1,4-phthalazinedione
Malachite green	Bis(*p*-dimethylaminophenyl)-phenylmethylium Chloride
Methylene blue	3,7-Bis(dimethylaniline)-phenazothionium Chloride
MTB	3,3′-Bis[*N*,*N*-di(carboxymethyl)aminomethyl]-thymolsulfonaphthalein
Nile Red (nile blue A oxazone)	9-Diethylamino-5*H*-benzo[*a*]-phenoxazin-5-one
Perphenazine	2-Chloro-10-{3-[1-(2-hydroxyethyl)-4-piperazinylpropyl]-phenothiazine}
p-Phenetidine	4-Ethoxyaniline
Phenolphthalein	4′,4″-Dihydroxytriphenylmethane-2-carboxylic Acid
Phenosafranine	3,7-Diamino-5-phenylphenazinium Chloride
Promethazine	10-(2-Dimethylamino-2-methylethyl)phenothiazine
Protocatechuic Acid	3,4-Dihydroxybenzoic Acid
Purpurine	1,2,4-Trihydroxyanthraquinone
Prochlorperazine	3-Chloro-10-[3-(4-methyl-1-piperazinyl)-propyl]phenothiazine
PCV	3,3′,4-Trihydroxyfuchsone-2′-sulfonic Acid
Pyrogallolphthalein	4,5-Dihydroxyfluorescein; 9-(2′-Carboxyphenyl)-4,5,6-trihydroxy-3*H*-xanthen-3-one
Pyrogallol Red	Pyrogallol Sulfonephthalein; 9-(2′-Sulfophenyl)-4,5,6-trihydroxy-3*H*-xanthen-3-one
Resazurin	7-Hydroxy-3*H*-phenoxazin-3-one 10-Oxide
Rhodamine B	9-(2-Carboxyphenyl)-3,6-bis-(diethylamino)xanthylium Chloride
Rhodamine 6G	Ethyl-2-(6-ethylamino-3-ethylimino-3*H*-xanthen-9-yl)-benzoate Hydrochloride
Safranine	3,7-Diamino-2,8-dimethyl-5-phenylphenazonium Chloride
Sulfanilic Acid	4-Aminobenzenesulfonic Acid
Tetrabase	4,4′-Bis(dimethylamino)-diphenylmethane
Thionine	3,7-Diaminophenothiazine-5-ium Chloride
Tiron	Disodium Salt of 1,2-Dihydroxybenzene-3,5-disulfonic Acid
Thymol blue	Thymolsulfonephthalein; 4,4′-(3*H*-2,1-benzoxathiol-3-ylidene)bis[5-methyl-2-(1-methylene)phenol] *S*,*S*-Dioxide
Toluidine Blue	2-Methyl-3-amino-7-dimethylaminophenazothionium Chloride
Variamine Blue B leuco base analog	*N*-(*p*-Methoxyphenyl)-*N*′,*N*′-dimethyl-*p*-phenylenediamine
Xylenol Orange	3,3′-Bis[*N*,*N*-di(carboxymethyl)-aminomethyl]-*o*-cresolsulfonenaphthalein
Zephiramine	Benzyldimethyltetradecylammonium Chloride

RELATED ARTICLES

Kinetic Determinations **(Volume 12)**
Differential Rate Determinations

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics

REFERENCES

1. K.B. Yatsimirskii, *Kinetic Methods of Analysis*, Pergamon Press, Oxford, 1966.
2. M.A. Guyard, 'De la Formation du Noir d'aniline, au Moyen des Sels de Vanadium, et de la Theorie du la Formation du Noir d'aniline', *Bull. Soc. Chim., Paris*, **25**, 58–64 (1876).
3. E.B. Sandell, I.M. Kolthoff, 'Chronometric Catalytic Method for the Determination of Micro Quantities of Iodine', *J. Am. Chem. Soc.*, **56**, 1426 (1934).
4. E.B. Sandell, I.M. Kolthoff, 'Micro Determination of Iodine by a Catalytic Method', *Mikrochim. Acta*, **1**, 9–25 (1937).
5. P.R. Bontchev, 'Catalytic Reactions I: Mechanisms', *Talanta*, **17**, 499–513 (1970).
6. P.R. Bontchev, 'Catalytic Reactions II: Activation', *Talanta*, **19**, 675–685 (1972).
7. H.B. Mark, Jr, 'Reaction Rate Methods in Analysis', *Talanta*, (Talanta Review), **19**, 717–746 (1972).
8. T. Kawashima, Y. Yonehara, 'Catalytic Methods of Analysis', *Bunseki Kagaku*, **21**, 825–834 (1972).
9. M. Tanaka, 'Kinetic Method of Analysis', *Bunseki*, 426–432 (1976).
10. H. Weisz, 'Application of Catalyzed Reactions in Analytical Chemistry', *Angew. Chem., Int. Ed. Engl.*, **15**, 150–156 (1976).
11. M. Otto, H. Mueller, G. Werner, 'Selectivity of Catalytic Methods of Determination', *Talanta*, **25**, 123–130 (1978).
12. M. Valcarcel, F. Grases, 'Fluorimetric Reaction-rate Methods of Inorganic Analysis: A Review', *Talanta*, **30**, 139–143 (1983).
13. N.M. Ushakova, I.F. Dolmanova, 'Kinetic Methods of Determination of Nonmetals', *Zh. Anal. Khim.*, **38**, 1515–1531 (1983).
14. G.A. Milovanovic, 'Determination of Organic Substances by Kinetic–Catalytic Methods of Analysis', *Microchem. J.*, **28**, 437–457 (1983).
15. S. Nakano, T. Kawashima, 'Kinetic Methods of Analysis: Catalytic Methods (A Review)', *Bunseki*, 317–324 (1987).
16. K.B. Yatsimirskii, L.P. Tikhonova, 'Catalytic Kinetic Methods for the Determination of Platinum Metals', *Talanta*, **34**, 69–75 (1987).
17. S. Yamada, A. Murata, M. Tanaka, 'How to Design a Kinetic–Catalytic Method of Analysis Based on Ligand Substitution Reactions', *Mikrochim. Acta*, **III**, 291–297 (1988).
18. F.F. Gaal, 'Catalytic Titrations', *Analyst*, **112**, 739–751 (1987).
19. M. Valcarcel, 'Selectivity and Kinetics in Analytical Chemistry', *Analyst*, **112**, 729–737 (1987).
20. R. Lobinski, Z. Marczenko, 'Recent Advances in Ultraviolet–Visible Spectrophotometry', *CRC Crit. Rev. Anal. Chem.*, **23**, 55–111 (1992).
21. T. Kawashima, S. Nakano, 'Flow Injection Analysis of Trace Elements by Use of Catalytic Reactions', *Anal. Chim. Acta*, **261**, 167–182 (1992).
22. D. Perez-Bendito, S. Rubio, 'Micellar Catalysis in Reaction-rate Methods', *Trends Anal. Chem.*, **12**, 9–18 (1993).
23. T. Kawashima, S. Nakano, M. Tabata, M. Tanaka, 'Catalytic Reactions as Applied to Trace Analysis', *Trends Anal. Chem.*, **16**, 132–139 (1997).
24. H.B. Mark, Jr, G.A. Rechnitz, *Kinetics in Analytical Chemistry*, Interscience, New York, 1968.
25. G.G. Guilbault, 'Application of Kinetics and Catalysis to Analysis', in *Treatise on Analytical Chemistry, Part 1, Theory and Practice*, 2nd edition, eds. I.M. Kolthoff, P.J. Elving, John Wiley & Sons, New York, 663–710, Vol. 1, 1978.
26. H.A. Mottola, *Kinetic Aspects of Analytical Chemistry*, John Wiley & Sons, New York, 1988.
27. D. Perez-Bendito, M. Silva, *Kinetic Methods in Analytical Chemistry*, Ellis Horwood, New York, 1988.
28. G.A. Rechnitz, 'Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **36**, 453R–457R (1964); **38**, 513R–517R (1966); **40**, 455R–458R (1968).
29. G.G. Guilbault, 'Use of Enzymes in and Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **42**, 334R–349R (1970).
30. R.A. Greinke, H.B. Mark, Jr, 'Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **44**, 295R–300R (1972); **46**, 413R–420R (1974); **48**, 87R–93R (1976); **50**, 70R–76R (1978).
31. H.A. Mottola, H.B. Mark, Jr, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **52**, 31R–40R (1980); **54**, 62R–83R (1982); **56**, 96R–112R (1984); **58**, 264R–279R (1986).
32. H.A. Mottola, D. Perez-Bendito, H.B. Mark, Jr, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **60**, 181R–200R (1988); **62**, 441R–461R (1990).
33. H.A. Mottola, D. Perez-Bendito, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **64**, 407R–428R (1992); **66**, 131R–162R (1994); **68**, 257R–289R (1996).
34. S.R. Crouch, T.F. Cullen, A. Scheeline, E.S. Kirkor, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **70**, 53R–106R (1998).

35. H.A. Mottola, 'Second International Symposium on Kinetics in Analytical Chemistry – Preveza Beach, Greece 9–12 September, 1986', *Analyst*, **112**, 717 (1987).
36. H.A. Mottola, 'Third International Symposium on Kinetics in Analytical Chemistry – Dubrovnik Cavtat, Yugoslavia, 25–28 September, 1989', *Analyst*, **115**, 665 (1990).
37. U. Nickel, 'Fourth International Symposium on Kinetics in Analytical Chemistry, Erlangen, Germany, September 27–30, 1992', *Analyst*, **118**, 674 (1993).
38. H.A. Mottola, 'Perspective – From Erlangen to Moscow: Minor Redirections', *Analyst*, **121**, 381–383 (1996).
39. L.J. Kricka, G.H.G. Thorpe, 'Chemiluminescent and Bioluminescent Methods in Analytical Chemistry (A Review)', *Analyst*, **108**, 1274–1296 (1983).
40. K. Robards, P.J. Worsfold, 'Analytical Applications of Liquid-phase Chemiluminescence', *Anal. Chim. Acta*, **266**, 147–173 (1992).
41. H.O. Albrecht, 'Uber die Chemiluminescenz des Aminophthalsaurehydrazids', *Z. Phys. Chem.*, **136**, 321–330 (1928).
42. M.M. Rauhut, A.M. Semsel, B.G. Roberts, 'Reaction Rates, Quantum Yields, and Partial Mechanism for the Chemiluminescent Reaction of 3-Aminophthalhydrazide with Aqueous Alkaline Hydrogen Peroxide and Persulfate', *J. Org. Chem.*, **31**, 2431–2436 (1966).
43. L. Erdey, I. Buzas, 'Complexometric and Argentometric Titrations Using Chemiluminescent Indicators', *Anal. Chim. Acta*, **22**, 524–531 (1960).
44. K. Glue, W. Petsch, 'Die Chemiluminescenz der Dimethyldiacridyliumsalze', *Angew. Chem.*, **48**, 57–59 (1935).
45. A. Ingvarsson, C.L. Flurer, T.E. Riehl, K.N. Thimmaiah, J.M. Williams, W.L. Hinze, 'Improvement in 10,10′-Dimethyl-9,9′-biacridinium Dinitrate Analytical Chemiluminescence Measurements by Use of Reactive Hydroxide Counterion Alkyltrimethylammonium Micellar Surfactants', *Anal. Chem.*, **60**, 2047–2055 (1988).
46. B. Radziszewski, 'Untersuchungen uber Hydrobenzamid, Amarin und Lophin', *Ber. Dtsch. Chem. Ges.*, **10**, 70–75 (1877).
47. E.H. White, M.J.C. Harding, 'Chemiluminescence in Liquid Solutions: The Chemiluminescence of Lophine and Its Derivatives', *Photochem. Photobiol.*, **4**, 1129–1155 (1965).
48. K. Imai, R. Weinberger, 'Aryl Oxalate Chemiluminescence Detection for High-performance Liquid Chromatography', *Trends Anal. Chem.*, **4**, 170–175 (1985).
49. M. Phull, P.C. Nigam, 'Kinetics and Mechanism of Mercury(II)-catalyzed Replacement of Cyanide in Hexacyanoferrate(II) by *p*-Nitrosodiphenylamine', *Talanta*, **28**, 591–597 (1981).
50. M. Phull, H.C. Bajaj, P.C. Nigam, 'Trace Element Determination in Environmental Pollution: A Kinetic Method for Mercury Estimation', *Talanta*, **28**, 610–612 (1981).
51. D.C. Olson, D.W. Margerum, 'Coordination Chain Reactions. Exchange Between Ethylenediaminetetraacetatocuprate(II) and Triethylenetetraminenickel(II)', *J. Am. Chem. Soc.*, **85**, 297–303 (1963).
52. D.W. Margerum, R.K. Steinhaus, 'Ultratrace Determination of Metals Using Coordination Chain Reactions', *Anal. Chem.*, **37**, 222–228 (1965).
53. T. Kawashima, N. Hatakeyama, M. Kamada, S. Nakano, 'Spectrophotometric Determination of Ultratrace Amounts of Iron by Its Catalytic Effect', *Nippon Kagaku Kaishi*, 84–90 (1981).
54. S. Nakano, A. Ohta, T. Kawashima, 'Catalytic Determination of Nanogram Amounts of Mn(II) by the Aerial Oxidative Coupling Reaction of 3-Methyl-2-benzothiazolinone Hydrazone with *N*,*N*-Dimethylaniline', *Mikrochim. Acta*, **II**, 273–282 (1985).
55. S. Nakano, H. Ihara, M. Tanaka, T. Kawashima, 'Catalytic Determination of Ultratrace Amounts of Cu(II) by the Oxidative Coupling Reaction of 3-Methyl-2-benzothiazolinone Hydrazone with *N*,*N*-Dimethylaniline', *Mikrochim. Acta*, **I**, 455–465 (1985).
56. T. Kawashima, S. Karasumaru, M. Hashimoto, S. Nakano, 'Catalytic Determination of Trace Amounts of Vanadium(V) by the Oxidative Coupling Reaction of *p*-Hydrazinobenzenesulfonic Acid with *m*-Phenylenediamine', *Nippon Kagaku Kaishi*, 175–178 (1981).
57. S. Nakano, M. Tago, T. Kawashima, 'Catalytic Determination of Nanogram Amounts of Vanadium in Natural Water by Flow Injection Analysis', *Anal. Sci.*, **5**, 69–72 (1989).
58. S. Nakano, S. Hinokuma, T. Kawashima, 'A Kinetic Determination of Ultratrace Amounts of Chromium(III) by Its Catalytic Effect', *Chem. Lett.*, 357–360 (1983).
59. J.H. Fendler, E.J. Fendler, *Catalysis in Micellar and Macromolecular Systems*, Academic Press, New York, 1975.
60. S. Rubio, D. Perez-Bendito, 'Micellar Media in Kinetic Determinations', *Anal. Chim. Acta*, **224**, 185–198 (1989).
61. M.L. Lunar, S. Rubio, D. Perez-Bendito, 'Micellar Catalysis in Kinetic Methods of Analysis: Improvement of Spectrophotometric Catalytic Determination of Copper', *Talanta*, **39**, 1163–1173 (1992).
62. S. Nakano, M. Sakai, M. Tanaka, T. Kawashima, 'Spectrophotometric Determination of Nanogram Level of Copper(II) by Its Catalytic Effect', *Chem. Lett.*, 473–476 (1979).
63. D. Perez-Bendito, 'Inorganic Differential Kinetic Analysis (A Review)', *Analyst*, **109**, 891–899 (1984).
64. J.B. Worthington, H.L. Pardue, 'Kinetics and Analytical Applications of the Ruthenium Catalyzed Reaction Between Cerium(IV) and Arsenic(III) in Sulfuric Acid Medium', *Anal. Chem.*, **42**, 1157–1164 (1970).
65. J. Ruzicka, E.H. Hansen, *Flow Injection Analysis*, 2nd edition, John Wiley & Sons, New York, 1988.
66. L. Erdey, I. Buzas, 'Complexometric and Argentometric Titrations Using Chemiluminescent Indicators', *Anal. Chim. Acta*, **22**, 524–531 (1960).

67. H.A. Mottola, H. Freiser, 'Use of Metal Ion Catalysis in Detection and Determination of Microamounts of Complexing Agents', *Anal. Chem.*, **40**, 1266–1268 (1968).
68. H.A. Mottola, 'Catalytic Titrants and Catalytic Indication of End-points', *Talanta*, **16**, 1267–1275 (1969).
69. H.A. Mottola, 'Titration of Microgram Amounts of Aminopolycarboxylic Acids with Manganese(II) and Catalytic End-point Indication', *Anal. Chem.*, **42**, 630–635 (1970).
70. S. Abe, T. Matsuo, 'The Use of Phenolphthalein as a Catalytic Indicator for the Determination of Copper and Manganese by EDTA Titration', *Bunseki Kagaku*, **20**, 1168–1172 (1971).
71. N. Kiba, M. Furusawa, 'Catalytic Thermometric Titration of Silver(I), Mercury(II) and Palladium(II)', *Anal. Chim. Acta*, **98**, 343–348 (1978).
72. S. Pantel, H. Weisz, 'Catalytic End-point Indication in Redox Titrations', *Anal. Chim. Acta*, **116**, 421–425 (1980).
73. T. Raya-Saro, D. Perez-Bendito, 'Semi-automatic Catalytic Titration of Aminopolycarboxylic Acids and Metal Ions with the 4,4′-Dihydroxybenzophenone Thiosemicarbazone–Hydrogen Peroxide–Copper(II) System', *Analyst*, **108**, 857–863 (1983).
74. T. Raya-Saro, D. Perez-Bendito, 'Simultaneous Semiautomatic Catalytic Titration of Binary Mixtures of Mercury(II) with Copper(II) or Cadmium at the Micromolar Level', *Anal. Chim. Acta*, **182**, 163–175 (1986).
75. E.J. Greenhow, M. Kashanipour, 'Determination of Metal Salts in the Presence of Mineral Acids by a Combination of Turbidimetric and Catalytic Thermometric Titrimetry', *Analyst*, **109**, 931–935 (1984).
76. A. Moreno, M. Silva, D. Perez-Bendito, M. Valcarcel, 'Fluorimetric Semiautomatic Catalytic Titrations', *Analyst*, **109**, 249–253 (1984).
77. T.R. Saro, D. Perez-Bendito, 'Semi-automatic Catalytic Titration of Metal Ion Mixtures and Its Application to Metallurgical Samples', *Anal. Chim. Acta*, **172**, 273–283 (1985).
78. M. Marquez, M. Silva, D. Perez-Bendito, 'Semi-automatic Catalytic Titration Method for the Analysis of Cystine–Cysteine Mixtures and Its Application to Urine Samples', *Analyst*, **113**, 1373–1376 (1988).
79. B.F. Abramovic, M.M. Marinkovic, F.F. Gaal, 'Titrimetric Determination of Gold in Some Pharmaceutical Preparations', *Analyst*, **115**, 79–83 (1990).
80. B.F. Abramovic, F.F. Gaal, S.D. Cvetkovic, 'Titrimetric Determination of Fluoride in Some Pharmaceutical Products Used for Fluoridation', *Talanta*, **39**, 511–515 (1992).
81. B.F. Abramovic, K.S. Horvath, F.F. Gaal, 'Microcomputer-aided Titrimetric Determination of Bromine-containing Active Ingredients in Some Drug Formulations', *Analyst*, **118**, 899–903 (1993).
82. A.G. Hens, M. Ternero, D. Perez-Bendito, M. Valcarcel, 'Semi-automatic Catalytic Titrations Based on the 1,4-Dihydroxyphthalimide Dioxime/Iodate/Nickel System', *Mikrochim. Acta*, **I**, 375–384 (1979).
83. M. Ternero, D. Perez-Bendito, M. Valcarcel, 'Semiautomatic Indirect Titration of Alkaline-earth Ions with Catalytic End-point Indication', *Microchem. J.*, **26**, 61–67 (1981).
84. F.F. Gaal, B.F. Abramovic, R.I. Cservenak, 'Some Observations on Catalytic Conductometric Titrations', *Microchem. J.*, **34**, 295–304 (1986).
85. S. Abe, A. Kanno, K. Hasegawa, M. Endo, 'A Simple Analytical Method for Monitoring the Total Concentration of Heavy Metal Ions in Environmental Waters', *Int. J. Environ. Anal. Chem.*, **42**, 45–52 (1990).
86. H. Weisz, J. Schlipf, 'Titrimetric Analysis with Catalytic End-point Detection and a Spotting Technique', *Anal. Chim. Acta*, **147**, 247–254 (1983).
87. M.A. Koupparis, P. Anagnostopoulou, H.V. Malmstadt, 'Automated Flow-injection Pseudotitration of Strong and Weak Acids, Ascorbic Acid and Calcium, and Catalytic Pseudotitrations of Aminopolycarboxylic Acids by Use of a Microcomputer-controlled Analyzer', *Talanta*, **32**, 411–417 (1985).
88. H. Weisz, J. Schlipf, 'Gaseous Catalysts for End-point Indication in Titrimetric Analysis in the Microgram Range', *Anal. Chim. Acta*, **121**, 257–263 (1980).
89. S. Nakano, H. Hayashi, T. Kawashima, 'Determination of Nanogram Amounts of Copper(II) by the Catalytic–Photometric Method', *Anal. Sci.*, **9**, 695–699 (1993).
90. S. Nakano, H. Enoki, T. Kawashima, 'A New Catalytic Determination of Copper(II) by Means of the Oxidative Coupling Reaction of *p*-Hydrazinobenzenesulfonic Acid with *m*-Phenylenediamine', *Chem. Lett.*, 1173–1176 (1980).
91. S. Nakano, K. Kuramoto, T. Kawashima, 'Determination of Ultratrace Amounts of Copper by the Catalytic Oxidation of Variamine Blue B Leuco Base Analog', *Nippon Kagaku Kaishi*, 91–97 (1981).
92. S. Nakueno, M. Tanaka, M. Fushihara, T. Kawashima, 'Catalytic Determination of Nanogram Amounts of Cu(II) by the Oxidative Coupling Reaction of *N*-Phenyl-*p*-phenylenediamine with *N*,*N*-Dimethylaniline', *Mikrochim. Acta*, **I**, 457–465 (1983).
93. S. Nakano, K. Nakaso, K. Noguchi, T. Kawashima, 'Catalytic Flow-injection Determination of Copper at Nanogram Levels by Using Color Formation of *N*-Phenyl-*p*-phenylenediamine with *m*-Phenylenediamine in the Presence of Pyridine and Ammonia as Activators', *Talanta*, **44**, 765–770 (1997).
94. F.L. Boza, M.D. Luque de Castro, M. Valcarcel, 'Catalytic–Fluorimetric Determination of Copper at the Nanograms per Millilitre Level by Flow Injection Analysis', *Analyst*, **109**, 333–337 (1984).
95. F. Holz, 'Photometric Determination of Ultra Traces of Copper by Measuring Its Catalytic Effect on Oxidative

Coupling of *N*,*N*-Dimethylaniline and 3-Methyl-2-benzothiazolinohydrazone by Hydrogen Peroxide in the Presence of 2,2′-Dipyridyl as an Activator', *Fresenius' Z. Anal. Chem.*, **319**, 29–37 (1984).

96. K. Satoh, N. Iwamura, N. Teshima, S. Nakano, T. Kawashima, 'Catalytic Flow Injection Analysis of Copper(II) and the Activation Effects of Polypyridines', *J. Flow Injection Anal.*, **10**, 245–254 (1993).
97. S. Ohno, N. Teshima, T. Watanabe, H. Itabashi, S. Nakano, T. Kawashima, 'Determination of Ultratrace Amounts of Copper(II) by Its Catalytic Effect on the Oxidative Coupling Reaction of 3-Methyl-2-benzothiazolinone Hydrazone with *N*-Ethyl-*N*-(hydroxy-3-sulfopropyl)-3,5-dimethoxyaniline', *Analyst*, **121**, 1515–1518 (1996).
98. D.G. Themelis, G.S. Vasilikiotis, 'Kinetic Determination of Trace Amounts of Copper(II) Using its Catalytic Effect on the Oxidation of Chromotropic Acid by Hydrogen Peroxide', *Analyst*, **112**, 797–801 (1987).
99. A. Velasco, M. Silva, M. Valcarcel, 'Kinetic Photometric Determination of Copper at the Nanograms per Millilitre Level in Real Samples', *Anal. Chim. Acta*, **299**, 107–114 (1990).
100. E. Gomez, J.M. Estela, V. Cerda, 'Thermometric Determination of Copper(II) Based on Its Catalytic Effect on the Oxidation of Hydroxylamine by Dissolved Oxygen', *Thermochim. Acta*, **165**, 255–260 (1990).
101. M. Marquez, M. Silva, D. Perez-Bendito, 'The Continuous-addition-of-reagent Technique in Catalytic Kinetic-based Determinations', *Anal. Lett.*, **23**, 1357–1369 (1990).
102. S. Kawakubo, H. Kato, M. Iwatsuki, 'Catalytic Spectrofluorimetric Determination of Copper Using Aerial Oxidation of Ascorbic Acid in the Presence of *o*-Phenylenediamine', *Analyst*, **119**, 2119–2122 (1994).
103. T.V. Rodionova, M.K. Beklemishev, E.V. Kletskina, L.V. Dracheva, Y.A. Zolotov, 'Continuous-flow Analysis: Determination of Copper in Process Solutions', *Zh. Anal. Khim.*, **49**, 813–816 (1994).
104. S. Cao, J. Zhong, K. Hasebe, W. Hu, 'On-Line Monitoring of Trace Amounts of Copper(II) in Steam Condensate and Boiler Feed-water by Flow Injection Analysis Based on Its Catalytic Effect on the Oxidation of Hydroquinone by Hydrogen Peroxide', *Anal. Chim. Acta*, **331**, 257–262 (1996).
105. S.M. Ramasamy, H.A. Mottola, 'Flow Injection (Closed-loop Configuration) Catalytic Determination of Copper in Human Blood Serum', *Anal. Chim. Acta*, **127**, 39–46 (1981).
106. J. Michalowski, M. Trojanowicz, 'Catalytic Determination of Copper in Blood Plasma Using Flow-injection Biamperometry', *Anal. Chim. Acta*, **281**, 299–304 (1993).
107. C. Sanchez-Pedreno, J.A. Ortuno, J. Martinez-Rodenas, 'Kinetic–Potentiometric Determination of Copper in Real Samples with a New Antimony(V) Ion-selective Electrode', *Fresenius' J. Anal. Chem*, **344**, 100–103 (1992).
108. M. Yamada, H. Kanai, S. Suzuki, 'Flavin Mononucleotide Chemiluminescence for Determination of Traces of Copper(II) by Continuous Flow and Flow Injection Methods', *Bull. Chem. Soc. Jpn.*, **58**, 1137–1142 (1985).
109. J. Lin, T. Hobo, 'Chemiluminescence Investigation of NH_2OH–Fluorescein–Cu^{2+} System and Its Application to Copper Analysis in Serum', *Talanta*, **42**, 1619–1623 (1995).
110. O.I. Mel'nikova, T.N. Shekhovtsova, I.F. Dolmanova, 'Effect of Organic Solvents on the Oxidation of Diphenylcarbazone with Hydrogen Peroxide, Catalyzed by Nickel and Cobalt. Kinetic Methods of Determining Nickel and Cobalt', *Zh. Anal. Khim.*, **35**, 1960–1964 (1980).
111. T. Yamane, 'Flow Injection Determination of Trace Cobalt(II)', *Nippon Kagaku Kaishi*, 93–97 (1982).
112. C. Maekoya, F. Mizuniwa, K. Usami, K. Osumi, 'Flow Injection Determination of ppt Levels of Cobalt(II) in Water by the Use of Contact Catalytic Reaction', *Nippon Kagaku Kaishi*, 1023–1027 (1983).
113. K. Isshiki, E. Nakayama, 'Determination of Ultratrace Amounts of Cobalt by Catalysis of the Tiron–Hydrogen Peroxide Reaction with an Improved Continuous-flow Analysis System', *Talanta*, **34**, 277–281 (1987).
114. T. Kawashima, T. Minami, M. Ata, M. Kamada, S. Nakano, 'Catalytic Determination of Ultratrace Amounts of Cobalt by Flow Injection Analysis', *J. Flow Injection Anal.*, **2**, 40–49 (1985).
115. A. Malahoff, I.Ya. Kolotyrkina, L.K. Shpigun, 'Shipboard Determination of Dissolved Cobalt in Seawater Using Flow Injection with Catalytic Spectrophotometric Detection', *Analyst*, **121**, 1037–1041 (1996).
116. M. Llobat-Estelles, A. Sevillano-Cabeza, J. Medina-Escriche, 'Kinetic–Catalytic Determination of Cobalt by Oxidation of Pyrogallol Red by Hydrogen Peroxide', *Analyst*, **111**, 193–195 (1986).
117. T. Yamane, K. Watanabe, H.A. Mottola, 'Continuous-flow System for the Determination of Cobalt in Sea and River Water', *Anal. Chim. Acta*, **207**, 331–336 (1988).
118. I. Mori, Y. Fujita, Y. Nakahashi, K. Ikuta, K. Kato, 'Selective Fluorimetric Determination of Trace Amounts of Cobalt(II) with the System of Fluorescein–Hydrogen Peroxide–Tiron', *Bull. Chem. Soc. Jpn.*, **62**, 2536–2540 (1989).
119. K. Watanabe, A. Tsutsumi, N. Koura, 'Determination of Cobalt Based on Its Catalytic Effect on the Decomposition-reaction of Amide Black 10B with Hydrogen Peroxide', *Bunseki Kagaku*, **38**, 691–696 (1989).
120. K. Watanabe, A. Tsutsumi, N. Koura, 'Synthesis of Monoazo Derivatives of 1-Amino-8-naphthol-3,6-disulfonic Acid as New Reagents for Catalytic Analysis of Cobalt and the Effect of Substituted Groups on their Analytical Sensitivity', *Bunseki Kagaku*, **39**, 375–380 (1990).

121. M.T. Oms, R. Forteza, V. Cerda, 'Automatic Kinetic Thermometric Method for the Determination of Ultratraces of Cobalt', *Anal. Chim. Acta*, **258**, 177–182 (1992).
122. M.G. Angelova, A.A. Alexiev, 'Kinetic Spectrophotometric Method for the Determination of Urinary Cobalt Based on Its Catalytic Effect on the Oxidation of L-Adrenaline Hydrochloride by Hydrogen Peroxide', *Anal. Chim. Acta*, **290**, 363–369 (1994).
123. M. Endo, M. Ishihara, T. Yotsuyanagi, 'Autocatalytic Decomposition of Cobalt Complexes as an Indicator System for the Determination of Trace Amounts of Cobalt and Effectors', *Analyst*, **121**, 391–394 (1996).
124. Z. Yu, N. Teshima, S. Nakano, T. Kawashima, 'Flow Injection Photometric Determination of Sub-nanogram Amounts of Cobalt by its Catalysis of the Oxidative Coupling of 3-Methyl-2-benzothiazolinone Hydrazone with *N*,*N*-Dimethylaniline in a Micellar Medium', *Talanta*, **43**, 1519–1525 (1996).
125. F. Grases, J.G. March, R. Forteza, V. Cerda, 'Sensitive Reaction-rate Determination of Co(II) Based on Its Catalytic Action in the Aerial Oxidation of Sulphite by Use of a Simple Monitored Thermometric Technique', *Thermochim. Acta*, **73**, 181–186 (1984).
126. C.M. Sakamoto-Arnold, K.S. Johnson, 'Determination of Picomolar Levels of Cobalt in Seawater by Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chem.*, **59**, 1789–1794 (1987).
127. G.N. Chen, C.S. Huang, 'A Study of the Chemiluminescence of Some Acidic Triphenylmethane Dyes', *Talanta*, **35**, 625–631 (1988).
128. H. Sakai, T. Fujiwara, M. Yamamoto, T. Kumamaru, 'Determination of Ultra-trace Levels of Cobalt by Ion Chromatographic Separation and Chemiluminescence Detection', *Anal. Chim. Acta*, **221**, 249–258 (1989).
129. Q. Lin, A. Guiraum, R. Escobar, F.F. de la Rosa, 'Flow Injection Chemiluminescence Determination of Cobalt(II) and Manganese(II)', *Anal. Chim. Acta*, **283**, 379–385 (1993).
130. K. Watanabe, T. Yamasaki, M. Itagaki, 'Chemiluminescence Determination of Cobalt(II) with 1,10-Phenanthroline in Presence of Surfactants', *Bunseki Kagaku*, **45**, 897–902 (1996).
131. Z. Xie, F. Zhang, Y. Pan, 'Chemiluminescence Detection of Trace Amounts of Cobalt(II) with 2,6,7-Trihydroxy-9-(4′-chlorophenyl)-3-fluorone–Hydrogen Peroxide–Cetyltrimethylammonium Bromide System', *Analyst*, **123**, 273–275 (1998).
132. T. Kawashima, Y. Kozuma, S. Nakano, 'A New Catalytic Determination of Ultratrace Amounts of Iron with *p*-Anisidine and *N*,*N*-Dimethylaniline', *Anal. Chim. Acta*, **106**, 355–360 (1979).
133. S. Nakano, M. Odzu, M. Tanaka, T. Kawashima, 'Catalytic Determination of Nanogram Amounts of Fe(II, III) by the Oxidative Coupling Reaction of *N*-Phenyl-*p*-phenylenediamine with *N*,*N*-Dimethylaniline', *Mikrochim. Acta*, **I**, 403–411 (1983).
134. S. Nakano, K. Tsujii, T. Kawashima, 'Flow Injection Photometric Determination of Nanogram Levels of Iron Based on the Catalysis of Oxidative Coupling of *N*-Phenyl-*p*-phenylenediamine with *m*-Phenylenediamine in a Micellar Medium', *Talanta*, **42**, 1051–1056 (1995).
135. H. Muller, V. Muller, 'Kinetisch–Katalytische Eisenspurebestimmung unter Flow-injection – Bedingungen mittels Leukomalachitgrun – Wasserstoffperoxid – Reaktion', *Z. Chem.*, **26**, 142–143 (1986).
136. A.C. Zotou, C.G. Papadopoulos, 'Indicator Reaction for the Kinetic–Spectrophotometric Determination of Nanogram Amounts of Iron', *Analyst*, **112**, 787–789 (1987).
137. D.G. Themelis, G.S. Vasilikiotis, 'Catalytic Determination of Nanogram Amounts of Iron(III) Using Its Catalytic Effect on the Oxidation of Chromotropic Acid by Hydrogen Peroxide', *Analyst*, **112**, 791–795 (1987).
138. K. Hirayama, N. Unohara, 'Spectrophotometric Catalytic Determination of an Ultratrace Amount of Iron(III) in Water Based on the Oxidation of *N*,*N*-Dimethyl-*p*-phenylenediamine by Hydrogen Peroxide', *Anal. Chem.*, **60**, 2573–2577 (1988).
139. A.A. Alexiev, A.M. Stoyanova, 'A New Catalytic Reaction for the Determination of Nanomolar Concentrations of Iron(III)', *Anal. Lett.*, **21**, 1515–1532 (1988).
140. R. Forteza, J.M. Estela, V. Cerda, 'Sensitive Thermometric Catalytic Method for the Determination of Iron(III) Based on Its Catalytic Action on the Oxidation of Sulphanilic Acid by Sodium Periodate', *Analyst*, **115**, 749–751 (1990).
141. A. Alexiev, S. Rubio, M. Deyanova, A. Stoyanova, D. Sicilia, D. Perez Bendito, 'Improved Catalytic Photometric Determination of Iron(III) in Cetylpyridinium Premicellar Aggregates', *Anal. Chim. Acta*, **295**, 211–219 (1994).
142. I. Mori, Y. Fujita, K. Ikuta, Y. Nakahashi, K. Kato, K. Tamura, M. Ohji, 'Kinetic– and Catalytic–Spectrophotometric Determination of Iron by Pyrogallolphthalein and Potassium Perdisulphate in Non-ionic Surfactant Micellar Medium', *Fresenius' Z. Anal. Chem.*, **334**, 49–52 (1989).
143. D. Guo-Zhong, J. Zhi-Liang, 'A New Catalytic Kinetic Spectrophotometric Method for Determination of Iron', *Talanta*, **36**, 1107–1110 (1989).
144. A. Cladera, E. Gomez, J.M. Estela, V. Cerda, 'Determination of Iron by Flow Injection Based on the Catalytic Effect of the Iron(III)-Ethylenediaminetetraacetic Acid Complex on the Oxidation of Hydroxylamine by Dissolved Oxygen', *Analyst*, **116**, 913–917 (1991).
145. S. Nakano, M. Sakai, M. Kurachi, T. Kawashima, 'Kinetic Determination of Iron Based on Its Catalytic Action on the Oxidative Coupling of 4-Aminoantipyrine with *N*,*N*-Dimethylaniline', *Microchem. J.*, **49**, 298–304 (1994).

146. T. Tomiyasu, H. Sakamoto, N. Yonehara, 'Spectrophotometric Determination of Trace Amounts of Iron by Its Catalytic Effect on the Chlorpromazine–Hydrogen Peroxide Reaction', *Anal. Sci.*, **10**, 761–764 (1994).
147. S.S. Mitic, G.Z. Miletic, M.V. Obradovic, 'Catalytic Determination of Nanogram Amounts from Iron(III) Using Its Catalytic Effect on the Oxidation of Sodium Pyrogallol-5-sulphonate by Hydrogen Peroxide', *Talanta*, **42**, 1273–1278 (1995).
148. I.Ya. Kolotyrkina, L.K. Shpigun, Y.A. Zolotov, A. Malahoff, 'Application of Flow Injection Spectrophotometry to the Determination of Dissolved Iron in Seawater', *Analyst*, **120**, 201–206 (1995).
149. T. Tomiyasu, N. Teshima, S. Nakano, T. Kawashima, 'Kinetic–Catalytic Determination of Traces of Iron by the Oxidative Coupling Reaction of 3-Methyl-2-benzothiazolinone Hydrazone with *N*,*N*-Dimethylaniline', *Talanta*, **47**, 1093–1098 (1998).
150. V.A. Elrod, K.S. Johnson, K.H. Coale, 'Determination of Subnanomolar Levels of Iron(II) and Total Dissolved Iron in Seawater by Flow Injection Analysis with Chemiluminescence Detection', *Anal. Chem.*, **63**, 893–898 (1991).
151. H. Obata, H. Karatani, E. Nakayama, 'Automated Determination of Iron in Seawater by Chelating Resin Concentration and Chemiluminescence Detection', *Anal. Chem.*, **65**, 1524–1528 (1993).
152. K. Watanabe, T. Yamasaki, M. Itagaki, 'Chemiluminescence Determination of Iron(III) with 1,10-Phenanthroline', *Bunseki Kagaku*, **45**, 407–413 (1996).
153. S. Rubio, A. Gomez-Henz, M. Valcarcel, 'Fluorimetric Determination of Manganese at the Nanogram Level by Catalytic Oxidation of Pyridoxal 2-Pyridylhydrazone by Hydrogen Peroxide', *Analyst*, **109**, 717–722 (1984).
154. Y. Miyata, T. Hirano, S. Nakano, T. Kawashima, 'Determination of Manganese by Flow Injection Analysis Based on its Catalytic Effect on the Oxidative Coupling Reaction of 3-Methyl-2-benzothiazolinone Hydrazone with *N*,*N*-Dimethylaniline', *Anal. Sci.*, **7**, 97–101 (1991).
155. S. Maspoch, M. Blanco, V. Cerda, 'Catalytic Determination of Manganese at Ultra-trace Levels by Flow Injection Analysis', *Analyst*, **111**, 69–72 (1986).
156. Y.A. Zolotov, L.K. Shpigun, I.Y. Kolotyrkina, E.A. Novikov, O.V. Bazanova, 'The Trace Determination of Some Heavy Metals in Waters by Flow Injection Spectrophotometry and Potentiometry', *Anal. Chim. Acta*, **200**, 21–33 (1987).
157. I.Ya. Kolotyrkina, L.K. Shpigun, Y.A. Zolotov, G.I. Tsysin, 'Shipboard Flow Injection Method for the Determination of Manganese in Seawater Using In-valve Preconcentration and Catalytic Spectrophotometric Detection', *Analyst*, **116**, 707–710 (1991).
158. M.C. Quintero, M. Silva, D. Perez-Bendito, 'Evaluation of the Analytical Use of the Manganese-catalysed Malachite Green–Periodate Reaction by the Stopped-flow Technique', *Talanta*, **36**, 1091–1094 (1989).
159. C. Zhang, S. Kawakubo, T. Fukasawa, 'Determination of Trace Manganese in High-purity Titanium, Silicon and Mineral Acids by Flow Injection Method Based on a Catalytic Reaction', *Anal. Chim. Acta*, **217**, 23–30 (1989).
160. L. Jianli, W. Budong, Z. Fenyan, 'Determination of Free State Manganese(II) in a Decoction of a Traditional Chinese Medicine Based on a Kinetic Spectrophotometric Method', *Analyst*, **118**, 1213–1215 (1993).
161. J. Yang, C. Ma, S. Zhang, Z. Shen, 'Flow Injection Catalytic Kinetic Determination of Manganese Using Stopped-flow and Gradient Calibration', *Anal. Chim. Acta*, **235**, 323–327 (1990).
162. E. Gomez, J.M. Estela, V. Cerda, 'Kinetic Thermometric Determination of Traces of Manganese by Its Catalytic Effect on the Reaction between Tiron and Hydrogen Peroxide', *J. Thermal. Anal.*, **37**, 195–202 (1991).
163. J. Mattusch, G. Werner, H. Mueller, 'Determination of Manganese(II) by a Photoactivated, Catalytic Method', *Analyst*, **116**, 53–57 (1991).
164. K. Watanabe, F. Seto, A. Tsutsumi, 'Catalytic Analysis of Manganese Using 1-Amino-8-naphthol-3,6-disulfonic Acid Monoazo Compounds', *Bunseki Kagaku*, **41**, 37–42 (1992).
165. S. Nakano, M. Nozawa, M. Yanagawa, T. Kawashima, 'Flow Injection Determination of Subnanogram Amounts of Manganese by Catalysis of the Oxidative Coupling of *N*,*N*-Dimethyl-*p*-phenylenediamine with *m*-Phenylenediamine', *Anal. Chim. Acta*, **261**, 183–188 (1992).
166. J.M. Estela, A. Caro, R. Forteza, V. Cerda, 'Kinetic Determination of Manganese by Its Catalytic Effect on the Oxidation of Alizarin S by Hydrogen Peroxide', *Thermochim. Acta*, **200**, 467–474 (1992).
167. G. Zhang, D.X. Cheng, S. Feng, 'Study on Catalytic Fluorimetric Determination of Trace Manganese', *Talanta*, **40**, 1041–1047 (1993).
168. R. Liu, A. Zhang, D. Liu, S. Wang, 'Determination of Manganese in Chinese Tea Leaves by a Catalytic Kinetic Spectrophotometric Method', *Analyst*, **120**, 1195–1197 (1995).
169. K. Watanabe, K. Rokugawa, M. Itagaki, 'Fluorimetric Determination of Manganese(II) by the Catalytic Oxidation of 2,3-Dihydroxynaphthalene in the Presence of Ethylenediamine and Hydrogen Peroxide', *Bunseki Kagaku*, **44**, 933–938 (1995).
170. T. Yamane, K. Koshino, 'Flow Injection Systems for Determination of Trace Manganese in Various Salts by Catalytic Photometric Detection', *Talanta*, **43**, 963–969 (1996).
171. M.K. Beklemishev, T.A. Stoyan, I.F. Dolmanova, 'Sorption–Catalytic Determination of Manganese Directly on a Paper-based Chelating Sorbent', *Analyst*, **122**, 1161–1165 (1997).
172. Z. Fan, L. Qingxiong, 'A New Chemiluminescence System: MnO_4^-–Na_2CO_3–KOH and Its Application in the

Determination of Manganese', *Talanta*, **40**, 1557–1561 (1993).
173. M. Tanaka, N. Awata, 'Catalytic Determination of Submicrogram Amounts of Vanadium by Means of the Oxidative Coupling Reaction of Phenylhydrazine-*p*-sulfonic acid with α-Naphthylamine', *Anal. Chim. Acta*, **39**, 485–490 (1967).
174. S. Nakano, E. Kasahara, M. Tanaka, T. Kawashima, 'A Kinetic Determination of Nanogram Amounts of Vanadium by Its Catalytic Effect', *Chem. Lett.*, 597–600 (1981).
175. T. Fukasawa, T. Yamane, 'Determination of Trace Vanadium in Natural Waters by a Combined Ion Exchange–Catalytic Photometric Method', *Anal. Chim. Acta*, **88**, 147–153 (1977).
176. M.H. Cordoba, P. Vinas, C. Sanchez-Pedreno, 'Determination of Vanadium in Petroleum Products by a Catalytic Method', *Analyst*, **110**, 1343–1345 (1985).
177. S. Kawakubo, B. Liang, M. Iwatsuki, T. Fukawasa, 'Effect of Tartrate on Vanadium-catalyzed Chlorpromazine–Bromate Redox Reaction and Its Application to the Determination of Vanadium in Natural Waters', *Analyst*, **119**, 1391–1395 (1994).
178. S. Nakano, C. Yamada, M. Sakai, T. Kawashima, 'Catalytic Determination of Nanogram Amounts of Vanadium by the Oxidative Coupling Reaction of 4-Aminoantipyrine with *N*,*N*-Dimethylaniline', *Anal. Sci.*, **2**, 61–65 (1986).
179. S. Nakano, S. Kinoshita, M. Ikuta, T. Kawashima, 'Fractional Determination of Nanogram Amounts of Vanadium(IV) and Vanadium(V) in Natural Water by a Catalytic Method Combined with Solvent Extraction', *Anal. Sci.*, **6**, 435–438 (1990).
180. R.M. Marce, M. Calull, J. Torres, F. Borrull, 'Thermometric Determination of Trace Amounts of Vanadium(V) in Petroleum Products Using the Chromotropic Acid–Bromate Reaction', *Analyst*, **113**, 505–508 (1988).
181. T. Yamane, Y. Osada, M. Suzuki, 'Continuous Flow System for the Determination of Trace Vanadium in Natural Waters Utilizing In-line Preconcentration/Separation Coupled with Catalytic Photometric Detection', *Talanta*, **45**, 583–589 (1998).
182. M.L. Lunar, S. Rubio, D. Perez-Bendito, 'Combination of Micellar and Chemical Catalysis as a Means of Enhancing the Sensitivity of Catalytic Kinetic Determinations', *Anal. Chim. Acta*, **237**, 207–214 (1990).
183. A.C. Zotou, C.G. Papadopoulos, 'Kinetic–Spectrophotometric Determination of Trace Amounts of Vanadium(V)', *Analyst*, **115**, 323–327 (1990).
184. M.L. Camacho, M.T. Rodriguez, M.C. Mochon, A.G. Perez, 'Kinetic–Spectrophotometric Determination of Traces of Vanadium(V) by Its Catalytic Effect on the Oxidation of 1,4-Dihydroxyphthalimide Dioxime with Bromate', *Anal. Chim. Acta*, **244**, 89–97 (1991).
185. R. Forteza, J.M. Estela, V. Cerda, 'Kinetic Enthalpimetric Determination of Vanadium(V) by Its Catalytic Effect on the Oxidation Rate of Pyrogallol Red by Potassium Bromate', *Analyst*, **116**, 1171–1173 (1991).
186. A.A. Ensafi, A. Kazemzadeh, 'Highly Sensitive Spectrophotometric Kinetic Determination of Vanadium by Catalytic Effect on the Gallocyanine–Bromate Reaction', *Anal. Chim. Acta*, **298**, 27–32 (1994).
187. A.A. Mohamed, M. Iwatsuki, T. Fukawasa, M.F. El-Shahat, 'Catalytic Determination of Vanadium Using the Perphenazine–Bromate Redox Reaction and a Citric Acid Activator', *Analyst*, **120**, 2281–2285 (1995).
188. S. Kawakubo, K. Ogihara, M. Iwatsuki, 'Catalytic Spectrofluorimetric Determination of Vanadium Using Oxidation of *o*-Phenylenediamine with Bromate in the Presence of Gallic Acid', *Analyst*, **120**, 2719–2723 (1995).
189. M.A. Proskurnin, N.V. Osipova, V.V. Kuznetsova, E.K. Ivanova, A.G. Abroskin, 'Thermal Lens Determination of Vanadium(V) and Its Activators by Oxidation of Aniline by Bromate Ions', *Analyst*, **121**, 419–423 (1996).
190. A. Massoumi, H. Tavallali, 'Kinetic Spectrophotometric Determination of Vanadium by Catalytic Effect on the Indigo Carmine–Bromate Reaction', *Anal. Lett.*, **31**, 193–206 (1998).
191. A.N. Sierra, J.S.D. Alegria, A.C. Martin, 'Chemiluminescent Determination of Vanadium in Steel', *Fresenius' J. Anal. Chem.*, **345**, 437–440 (1993).
192. W. Qin, Z. Zhang, C. Zhang, 'Chemiluminescence Flow System for Vanadium(V) with Immobilized Reagents', *Analyst*, **122**, 685–688 (1997).
193. A.M. Afonso, J.J. Santana, F.G. Montelongo, 'Kinetic Spectrofluorimetric Determination of Silver, Based on Its Catalytic Effect on the Oxidation of Pyrocatechol-1-aldehyde 2-Pyridylhydrazone by Peroxodisulphate in the Presence of 1,10-Phenanthroline as Activator', *Talanta*, **33**, 779–783 (1986).
194. A. Sevillano-Cabeza, J. Medina-Escriche, M. Llobat-Estelles, M. Martin-Penella, 'Study of the Silver-catalysed Pyrogallol Red–Peroxodisulphate Reaction with 1,10-Phenanthroline as Activator by a Kinetic–Spectrophotometric Method', *Analyst*, **111**, 1417–1422 (1986).
195. A.A. Ensafi, K. Zarei, 'Flow Injection Determination of Silver with Spectrophotometric Detection', *Fresenius' J. Anal. Chem.*, **358**, 475–479 (1997).
196. S.P. Mushtakova, S.S. Khmelev, L.F. Kozhina, N.N. Bylinkina, 'Kinetic Method for Determination of Gold in Gold-plating Electrolytes', *Zh. Anal. Khim.*, **46**, 2411–2415 (1991).
197. I.F. Dolmanova, G.A. Zolotova, T.N. Shekhovtsova, V.M. Peshkova, 'Kinetic Determination of Small Amounts of Chromium(III)', *Zh. Anal. Khim.*, **25**, 2136–2138 (1970).
198. S.T. Nandibewoor, V.A. Morab, 'Chromium(III)-catalyzed Oxidation of Antimony(III) by Alkaline Hexacyanoferrate(III) and Analysis of Chromium(III) in Microamounts by a Kinetic Method', *J. Chem. Soc., Dalton Trans.*, 483–488 (1995).

199. R. Escobar, Q. Lin, A. Guiraum, F.F. de la Rosa, 'Flow Injection Chemiluminescence Method for the Selective Determination of Chromium(III)', *Analyst*, **118**, 643–647 (1993).
200. S. Nakano, M. Fukuda, S. Kageyama, H. Itabashi, T. Kawashima, 'Flow Injection Determination of Chromium(III) by Pyrogallol Chemiluminescence', *Talanta*, **40**, 75–80 (1993).
201. I.F. Dolmanova, G.A. Zolotova, L.V. Tarasova, V.M. Peshkova, 'Determination of Microamounts of Chromium(VI) by Means of the Catalytic Oxidation of *o*-Dianisidine by Hydrogen Peroxide', *Zh. Anal. Khim.*, **24**, 1035–1039 (1969).
202. T. Kamidate, K. Yamaguchi, T. Segawa, H. Watanabe, 'Lophine Chemiluminescence for Determination of Chromium(VI) by Continuous Flow Method', *Anal. Sci.*, **5**, 429–433 (1989).
203. C. Sanchez-Pedreno, M.T. Sierra, M.I. Sierra, A. Sanz, 'Catalytic Effect of Mercury(II) on the Photo-oxidation of EDTA by Iodine: Kinetic Determination of Mercury', *Analyst*, **113**, 145–148 (1988).
204. A. Safavi, M.R. Baezzat, 'Kinetic FIA Determination of Hg(II) Based on Its Catalytic Effect on the Reaction Between Safranin and Iodide', *Anal. Lett.*, **29**, 807–819 (1996).
205. S.P. Mushtakova, A.P. Gumenyuk, S.S. Khmelev, N.A. Batasheva, 'Oxidation of *N*-Methyldiphenylamine-4-sulfonic Acid in the Catalytic Determination of Iridium(IV)', *Zh. Anal. Khim.*, **48**, 1218–1225 (1993).
206. M.L. Lunar, S. Rubio, D. Perez-Bendito, 'Use of the Triiodide-Hexadecylpyridinium Chloride Micellar System for the Kinetic Determination of Molybdenum(VI)', *Analyst*, **118**, 715–718 (1993).
207. E.N.V.M. Carrilho, F.J. Krug, E.A.G. Zagatto, 'Correction of Iron Interface in the Spectrophotometric Flow Injection Catalytic Determination of Molybdenum in Plants', *Talanta*, **42**, 2021–2026 (1995).
208. A. Caro, J.M. Estela, R. Forteza, V. Cerda, 'Kinetic Thermometric Determination of Mo(VI) in Vegetables by Its Catalytic Effect on the Reaction Between Hydrogen Peroxide and Thiosulphate Ion', *Thermochim. Acta*, **224**, 271–279 (1993).
209. S. Kawakubo, H. Suzuki, M. Iwatsuki, 'Catalytic Spectrofluorometric Determination of Ultratrace Molybdenum in Natural Fresh Water', *Anal. Sci.*, **12**, 767–771 (1996).
210. A.A. Ensafi, 'Catalytic–Spectrophotometric Determination of Molybdenum in Plant and Steel with Nile Red–Hydrazine Dihydrochloride Reaction', *Anal. Lett.*, **25**, 2339–2354 (1992).
211. S.B. Jonnalagadda, M. Dumba, 'Studies on the Molybdenum Catalysed Reaction Between Toluidine Blue and Stannous Chloride', *Fresenius' J. Anal. Chem.*, **345**, 673–678 (1993).
212. H. Parham, N. Pourreza, R. Marandi, 'Kinetic–Spectrophotometric Determination of Molybdenum(VI) Based on Its Catalytic Effect on the Thionine–Hydrazine Reaction', *Anal. Lett.*, **31**, 179–191 (1998).
213. M.A. Lopez-Fernandez, A. Gomez-Hens, D. Perez-Bendito, 'Kinetic-fluorimetric Determination of Traces of Nickel by Its Catalytic Effect on the Oxidation of Pyridoxal Hydrazone by Potassium Peroxidisulphate', *Anal. Lett.*, **17**, 507–521 (1984).
214. F. Borrull, J. Torres, V. Cerda, 'Kinetic–Thermometric Determination of Nickel(II) Based on Its Catalytic Action in the Decomposition Reaction of Permanganate', *Analyst*, **112**, 1453–1455 (1987).
215. L. Xiaohu, L. Minggang, Z. Guiwen, 'Determination of Nickel in Water by Chemiluminescence', *Analyst*, **116**, 747–748 (1991).
216. A.A. Ensafi, A. Safavi, 'Sensitive Spectrophotometric Kinetic Determination of Osmium by Catalysis of the Pyrogallol Red–Bromate Reaction', *Anal. Chim. Acta*, **244**, 231–238 (1991).
217. A.A. Ensafi, E. Shamss-E-Sollari, 'Spectrophotometric Reaction Rate Method for the Determination of Osmium by Its Catalytic Effect on the Oxidation of Gallocyanine by Bromate', *Talanta*, **41**, 1651–1655 (1994).
218. Q. Zhu, G. Gong, H. Wang, 'Kinetic–Catalytic Determination of Osmium by Fluorescence Quenching with Salicylfluorone and Hydrogen Peroxide', *Mikrochim. Acta*, **116**, 197–204 (1994).
219. M. Li, G.E. Pacey, 'Lead Determination at the ppb Level Using Stopped-flow FIA', *Talanta*, **42**, 1857–1865 (1995).
220. A.T. Pilipenko, O.A. Zaporozhets, 'Chemiluminescence Determination of Lead with 4-Diethylaminophthalhydrazide', *Zh. Anal. Khim.*, **44**, 73–79 (1989).
221. G.Z. Miletic, S.S. Mitic, 'Catalytic Determination of Palladium by the Oxidation Reaction of Purpurin with Hydrogen Peroxide', *Zh. Anal. Khim.*, **49**, 508–511 (1994).
222. F. Grases, J.M. Estela, F. Garcia-Sanchez, M. Valcarcel, 'A Kinetic Fluorimetric Method for Determination of Microamounts of Pt(IV) with 2,2′-Dipyridylketone Hydrazone', *Anal. Lett.*, **13**, 181–189 (1980).
223. E.I. Plastinina, L.V. Borisova, R.F. Gur'eva, 'Kinetic Determination of Trace Amounts of Rhenium in Alkaline Solutions', *Zh. Anal. Khim.*, **47**, 2052–2056 (1992).
224. L.P. Tikhonova, I.P. Svarovskaya, I.A. Fedotova, G.N. Solov'eva, 'Kinetic Determination of Rhodium in Nickel- and Copper-based Materials', *Zh. Anal. Khim.*, **46**, 1176–1180 (1991).
225. A.P. Gumenyuk, V.V. Aleshina, S.P. Mushtakova, 'A Sensitive Kinetic Method of Determining Ruthenium', *Zh. Anal. Khim.*, **41**, 1400–1404 (1986).
226. P. Vinas, M.H. Cordoba, C. Sanchez-Pedreno, 'Catalytic Method for the Determination of Ruthenium in the Presence of Other Noble Metals', *Analyst*, **112**, 1019–1022 (1987).
227. A.M. Almuaibed, A. Townshend, 'Flow Injection Spectrophotometric Catalytic Method for the Determination of Ruthenium(III)', *Microchem. J.*, **48**, 210–214 (1993).

228. T. Kawashima, M. Tanaka, 'Determination of Submicrogram Amounts of Selenium(IV) by Means of the Catalytic Reduction of 1,4,6,11-Tetraazanaphthacene', *Anal. Chim. Acta*, **40**, 137–143 (1968).
229. T. Kawashima, S. Nakano, M. Tanaka, 'Catalytic Determination of Submicrogram Amounts of Selenium(IV) by Means of the Oxidative Coupling Reaction of Phenylhydrazine-*p*-sulfonic Acid with 1-Naphthylamine', *Anal. Chim. Acta*, **49**, 443–447 (1970).
230. T. Kawashima, S. Kai, S. Takashima, 'Catalytic Determination of Ultratrace Amounts of Selenium(IV)', *Anal. Chim. Acta*, **89**, 65–70 (1977).
231. W.C. Hawkes, 'Spectrophotometric Determination of Traces of Selenium by Catalytic Reduction of Tetranitro Blue Tetrazolium', *Anal. Chim. Acta*, **183**, 197–205 (1986).
232. E. Aoyama, N. Kobayashi, M. Shibata, T. Nakagawa, H. Tanaka, 'Determination of Selenium by Flow Injection Analysis Based on the Selenium(IV)-catalyzed Reduction of 3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl-2H Tetrazolium Bromide', *Anal. Sci.*, **7**, 103–107 (1991).
233. J.L. Bernal, M.J. del Nozal, L. Deban, F.J. Gomez, O. de Uria, J.M. Estela, V. Cerda, 'Modification of the Methylene Blue Method for Spectrophotometric Selenium Determination', *Talanta*, **37**, 931–936 (1990).
234. B. Arikan, M. Tuncay, R. Apak, 'Sensitivity Enhancement of the Methylene Blue Catalytic – Spectrophotometric Method of Selenium(IV) Determination by CTAB', *Anal. Chim. Acta*, **335**, 155–167 (1996).
235. A.A. Ensafi, G.B. Dehaghi, 'Kinetic–Spectrophotometric Determination of Trace Amounts of Selenium with Catalytic Reduction of Gallocyanine by Sulfide', *Anal. Lett.*, **28**, 335–347 (1995).
236. F. Grases, J.G. March, 'Determination of Technetium by Reduction of Methylene Blue with Tin(II)', *Anal. Chem.*, **57**, 1419–1422 (1985).
237. M.F. Mousavi, M.R. Almasian, 'Flow Injection Spectrophotometric Determination of Trace Amounts of Tellurium by Catalytic Reduction of Toluidine Blue', *Anal. Lett.*, **29**, 1851–1860 (1996).
238. F. Lazaro, M.D. Luque de Castro, M. Valcarcel, 'Catalytic–Fluorimetric Determination of Ti(IV) by Flow Injection Analysis', *Anal. Lett.*, **18**, 1209–1220 (1985).
239. W. Jianhua, H. Ronghuan, W. Jianguo, 'Kinetic Determination of Uranium(IV) Based on Its Catalytic Effect on the Phosphomolybdic Acid – Iodide Reaction', *J. Radioanal. Nucl. Chem.*, **162**, 57–62 (1992).
240. T. Tomiyasu, 'Kinetic Spectrophotometric Determination of Traces of Tungsten(VI) by the Catalytic Reaction of Chlorpromazine with Hydrogen Peroxide. A Mechanistic Study', *Anal. Chim. Acta*, **312**, 179–187 (1995).
241. N. Yonehara, A. Kawasaki, H. Sakamoto, M. Kamada, 'Spectrophotometric Determination of Traces of Bromide Based on Its Catalysis of the Pyrocatechol Violet/Hydrogen Peroxide Reaction', *Anal. Chim. Acta*, **206**, 273–280 (1988).
242. N. Yonehara, S. Akaike, H. Sakamoto, M. Kamada, 'Flow Injection – Spectrophotometric Determination of Trace Amounts of Bromide by Its Catalytic Effect on the Hydrogen Peroxide Oxidation of Pyrocatechol Violet', *Anal. Sci.*, **4**, 273–276 (1988).
243. A. Sukuragawa, Y. Kinoshita, T. Okutani, 'Flow Injection Spectrophotometric Determination of Trace Amounts of Cyanide Ion with *p*-Nitrobenzaldehyde and *o*-Dinitrobenzene on the Basis of Catalytic Effect', *Bunseki Kagaku*, **40**, T41–T46 (1991).
244. P. Vinas, M.H. Cordoba, S. Sanchez-Pedreno, 'Kinetic Determination of Iodide, Based on the Chlorpromazine–Bromate Reaction', *Talanta*, **34**, 351–354 (1987).
245. T. Tomiyasu, H. Sakamoto, N. Yonehara, 'Spectrophotometric Determination of Trace Amounts of Iodide by Its Catalytic Effect on the Chlorpromazine–Hydrogen Peroxide Reaction', *Anal. Sci.*, **8**, 293–298 (1992).
246. B. Liang, S. Kawakubo, M. Iwatsuki, T. Fukawasa, 'Catalytic Determination of Iodide by a Maximum Absorbance Method Using the Oxidation Reaction of Chlorpromazine with Hydrogen Peroxide', *Anal. Chim. Acta*, **282**, 87–94 (1993).
247. N. Yonehara, S. Kozono, H. Sakamoto, 'Flow Injection – Spectrophotometric Determination of Trace Amounts of Iodide by its Catalytic Effect on the 4,4′-Bis(dimethylamino)diphenylmethane – Chloramine T Reaction', *Anal. Sci.*, **7**, 229–234 (1991).
248. A.A. Mohamed, M. Iwatsuki, M.F. El-Shahat, T. Fukasawa, 'Catalytic Determination of Iodide Using the Promethazine–Hydrogen Peroxide Redox Reaction', *Analyst*, **120**, 1201–1204 (1995).
249. G. Lopez-Cueto, J.M. Santiago, N. Grane, 'Catalytic Determination of Iodine by Continuous Addition of Catalyst to a Reference Solution with Photometric Probe Monitoring', *Talanta*, **39**, 349–354 (1992).
250. V. Fournier-Bidoz, P. Foster, J.M. Quinault, 'De'termination de l'iode Inorganique dans les Eaux de Pluie', *Analusis*, **20**, 627–631 (1992).
251. C. Sanchez-Pedreno, J.A. Ortuno, A.L. Ros, 'Kinetic–Potentiometric Determination of Iodide with an Antimony(V) Coated-wire Ion-selective Electrode', *Analyst*, **117**, 1619–1621 (1992).
252. R. Montes, J.J. Laserna, 'Spectrophotometric Reaction-rate Method for the Determination of Nitrite in Waters with Pyridine-2-aldehyde 2-Pyridylhydrazone', *Talanta*, **34**, 1021–1026 (1987).
253. M. Jiang, F. Jiang, J. Duan, X. Tang, Z. Zhao, 'Spectrophotometric Determination of Trace Nitrite Based on Catalytic Oxidation of Thionine by Potassium Bromate', *Anal. Chim. Acta*, **234**, 403–407 (1990).
254. T. Perez-Ruiz, C. Martinez-Lozano, V. Tomas, 'Kinetic–Spectrofluorimetric Determination of Nitrite and Nitrate at the Nanograms per Millilitre Level by Manual and Flow Injection Methods', *Anal. Chim. Acta*, **265**, 103–110 (1992).

255. A.A. Ensafi, M. Samimifar, 'Kinetic-spectrophotometric Determination of Low Levels of Nitrite by Catalytic Reaction Between Pyrogallol Red and Bromate', *Talanta*, **40**, 1375–1378 (1993).
256. A.A. Ensafi, B. Rezaii, 'Kinetic–Spectrophotometric Determination of Nitrite by Its Catalytic Effect on the Oxidation of Brilliant Cresyl Blue by Bromate', *Microchem. J.*, **50**, 169–177 (1994).
257. B. Liang, M. Iwatsuki, T. Fukasawa, 'Catalytic Spectrophotometric Determination of Nitrite Using the Chlorpromazine–Hydrogen Peroxide Redox Reaction in Acetic Acid Medium', *Analyst*, **119**, 2113–2117 (1994).
258. C. Xingguo, W. Ketai, H. Zhide, Z. Zhengfeng, 'Kinetic Flow Injection Spectrophotometric Determination of Nitrite by Its Catalytic Effect on the Oxidation of Chlorosphophonazo-pN by Bromate', *Anal. Lett.*, **29**, 2015–2023 (1996).
259. A.A. Mohamed, M.F. El-shahat, T. Fukasawa, M. Iwatsuki, 'Sensitive Determination of Nitrite Using Its Catalytic Effect on the Bromate Oxidation of Prochlorperazine', *Analyst*, **121**, 89–92 (1996).
260. A. Afkhami, F. Jalali, 'Spectrophotometric Determination of Trace Amounts of Nitrite Ion Based on Its Catalytic Effect in the Reaction between Thymol Blue and Bromate', *Microchem. J.*, **57**, 224–230 (1997).
261. R. Kadowaki, S. Nakano, T. Kawashima, 'Sensitive Flow Injection Colorimetry of Nitrite by Catalytic Coupling of *N*-Phenyl-*p*-phenylenediamine with *N*,*N*-Dimethylaniline', *Talanta*, **48**, 103–107 (1999).
262. M.F. Mousavi, M. Shamsipur, 'Spectrophotometric Determination of Trace Amounts of Sulfide Based on Its Catalytic Reduction of Toluidine Blue', *Bull. Chem. Soc. Jpn.*, **65**, 2770–2772 (1992).
263. A.A. Ensafi, M. Samimifar, 'Spectrophotometric Reaction-rate Method for the Determination of Sulfide by Catalytic Action on the Oxidation of Sodium Azide by Iodine', *Anal. Lett.*, **27**, 153–167 (1994).
264. B. Rogina, M. Dubravcic, 'Microdetermination of Iodides by Arresting the Catalytic Reduction of Ceric Ions', *Analyst*, **78**, 594–599 (1953).
265. P.A. Rodriguez, H.L. Pardue, 'Kinetics of the Iodide-catalyzed Reaction between Cerium(IV) and Arsenic-(III) in Sulfuric Acid Medium', *Anal. Chem.*, **41**, 1369–1376 (1969).
266. P.A. Rodriguez, H.L. Pardue, 'Analytical Application of the Iodide and Osmium Catalyzed Reaction between Cerium(IV) and Arsenic(III)', *Anal. Chem.*, **41**, 1376–1380 (1969).
267. G. Knapp, H. Spitzy, 'Untersuchungen zur Optimierung der Reaktionsbedingungen fur die Katalytische Jodwirkung auf das System Ce(IV)-Arsenige Saure (Eine Modifizierte Sandell–Kolthoff-reaktion)', *Talanta*, **16**, 1353–1360 (1969).
268. T. Deguchi, A. Tanaka, I. Sanemasa, H. Nagai, 'Flow Injection Analysis of Trace Amounts of Iodide Ion Using a Catalytic Method', *Bunseki Kagaku*, **32**, 23–28 (1983).
269. R.D. Sauerbrunn, E.B. Sandell, 'Determination of Submicrogram Quantities of Osmium by Catalysis', *Mikrochim. Acta*, **1–2**, 22–33 (1953).
270. H.L. Pardue, R.L. Habig, 'Trace Analysis with Automatic Printout Based on Reaction Rate Measurements: Determination of Osmium', *Anal. Chim. Acta*, **35**, 383–390 (1966).
271. R.L. Habig, H.L. Pardue, J.B. Worthington, 'Kinetics of Osmium-catalyzed Reaction between Cerium(IV) and Arsenic(III) in Sulfuric Acid Medium', *Anal. Chem.*, **39**, 600–605 (1967).
272. D. Sicilia, S. Rubio, D. Perez-Bendito, 'Some Observations on the Use of Micellar Media in the Reaction between Cerium(IV) and Arsenic(III) Catalyzed by Metal Ions', *Analyst*, **115**, 1613–1616 (1990).
273. C. Surasiti, E.B. Sandell, 'Determination of Submicrogram Quantities of Ruthenium by Catalysis of the Cerium(IV)–Arsenic(III) Reaction', *Anal. Chim. Acta*, **22**, 261–269 (1960).
274. J. Wang, R. He, 'Stopped-flow Injection Kinetic Determination of Multicomponent Samples: Simultaneous Determination of Mercury(II) and Silver(I)', *Anal. Chim. Acta*, **294**, 195–199 (1994).
275. F. Lazaro, M.D. Luque de Castro, M. Valcarcel, 'Simultaneous Catalytic–Fluorimetric Determination of Copper and Mercury by Flow Injection Analysis', *Fresenius' Z. Anal. Chem.*, **320**, 128–132 (1985).
276. Z. Zhu, C. Han, Z. Gu, R. Chen, 'Simultaneous Determination of Chromium(VI) and Tungsten(VI) Based on Their Catalytic Effect on the Reaction Between Hydrogen Peroxide and Iodide', *Analyst*, **119**, 2251–2255 (1994).
277. F. Lazaro, M.D. Luque de Castro, M. Valcarcel, 'Sequential and Differential Catalytic–Fluorimetric Determination of Manganese and Iron by Flow Injection Analysis', *Anal. Chim. Acta*, **169**, 141–148 (1985).
278. W. Jianhua, H. Ronghuan, 'The Mutual Catalytic Effect in Kinetic Reactions and Its Application to Simultaneous Determination of Binary Components', *Mikrochim. Acta*, **117**, 23–30 (1994).
279. H. He, Y. Ci, W. Chang, W. Gong, 'Simultaneous Determination of Iron(III) and Vanadium(V) by Use of a Kinetic–Spectrophotometric and Rapid Mixing Flow System', *Talanta*, **38**, 1159–1162 (1991).
280. E. Gomez, J.M. Estela, V. Cerda, 'Simultaneous Thermometric Determination of Manganese and Lead Based on Their Catalytic Effect on the Oxidation of Tiron by Hydrogen Peroxide', *Anal. Chim. Acta*, **251**, 305–311 (1991).
281. C.G. Papadopoulos, A.C. Zotou, 'Kinetic–Spectrophotometric Determination of Molybdenum(VI) and Tungsten(VI) in Mixtures', *Mikrochim. Acta*, **106**, 203–210 (1992).
282. W. Jianhua, H. Ronghuan, 'Synergetic Catalytic Effect of Molybdate and Tungstate on the Hydrogen

Peroxide–Iodide System and Its Analytical Applications', *Anal. Chim. Acta*, **303**, 241–246 (1995).
283. R. Liu, D. Liu, A. Sun, G. Liu, 'Simultaneous Determination of Molybdenum and Tungsten Using a Flow Injection System and without Pre-separation', *Analyst*, **120**, 565–568 (1995).
284. T. Tomiyasu, 'Kinetic–Mechanistic Study of the Chlorpromazine–Hydrogen Peroxide Reaction Catalyzed by Molybdenum(VI) and Tungsten(VI), and Their Differential Determination', *Anal. Chim. Acta*, **349**, 43–52 (1997).
285. Z. Zhu, Z. Gu, R. Chen, C. Han, B. Lu, 'Simultaneous Determination of Catalysts Based on the Differences in the Characteristic Rate Spectra of Catalytic Kinetics', *Anal. Chim. Acta*, **298**, 19–26 (1994).
286. N. Yonehara, T. Yamane, T. Tomiyasu, H. Sakamoto, 'Simultaneous Spectrophotometric Determination of Traces of Bromide and Iodide Based on Their Catalytic Effects on the Pyrocatechol Violet–Hydrogen Peroxide Reaction', *Anal. Sci.*, **5**, 175–179 (1989).
287. M.K. Gadia, M.C. Mehra, 'Analytical Reactions of Substituted Cyanoferrates: 2. Pentacyanoamminoferrate(II) in Catalytic Spectrophotometric Determination of Subparts per Million Amounts of Ag^{+}, Au^{3+} and Hg^{2+} in Solution', *Microchem. J.*, **23**, 278–284 (1978).
288. K.M. Rao, T.S. Reddy, S.B. Rao, 'Determination of Trace Amounts of Mercury(II) and Silver(I): a Catalytic Method', *Analyst*, **113**, 983–985 (1988).
289. H. Itabashi, M. Takahata, H. Kawamoto, H. Akaiwa, 'Catalytic Determination of Silver(I) by Extractive Flow Injection Analysis', *Talanta*, **45**, 549–555 (1998).
290. M. Tabata, M. Tanaka, 'A Kinetic Method for the Determination of Nanogram Amounts of Cadmium(II) by Its Catalytic Effect on the Complex Formation of Manganese(II) with $\alpha, \beta, \gamma, \delta$-Tetra(*p*-sulphonatophenyl)porphine', *Mikrochim. Acta*, **II**, 149–158 (1982).
291. M. Tabata, M. Tanaka, 'A Kinetic Method for the Determination of a Nanogram Amount of Mercury(II) by Its Catalytic Effect on the Complex Formation Reaction of Manganese(II) with $\alpha, \beta, \gamma, \delta$-Tetraphenylporphinesulfonate', *Anal. Lett.*, **13**, 427–438 (1980).
292. A. Ramesh, 'Catalytic Method for the Determination of Trace Amounts of Mercury in Environmental Samples', *Analyst*, **118**, 945–946 (1993).
293. D. Sicilia, S. Rubio, D. Perez-Bendito, 'Kinetic Determination of Hg(II) Based on Its Accelerating Effect on the Reaction Between Hexacyanoferrate(II) and 1,10-Phenanthroline Catalyzed by Micelles', *Talanta*, **38**, 1147–1153 (1991).
294. M. Tabata, 'Kinetic Method for the Determination of Nanogram Amounts of Lead(II) Using Its Catalytic Effect on the Reaction of Manganese(II) with 5,10,15,20-Tetrakis(4-sulphonatophenyl)porphine', *Analyst*, **112**, 141–144 (1987).
295. M. Tabata, S. Funahashi, M. Tanaka, 'A Kinetic Method for the Determination of Ammonia by Its Catalytic Effect on the Ligand Substitution Reaction of the Mercury(II)–*o*-Cresolphthalein Complexone Complex with *trans*-1,2-Diaminocyclohexane-*N*,*N*,*N*′,*N*′-tetraacetic Acid', *Anal. Chim. Acta*, **62**, 289–295 (1972).
296. A. Sakuragawa, M. Tsukada, T. Okutani, S. Utsumi, 'Spectrophotometric Determination of a Minute Amount of Fluoride Ion by Catalysis of the Zirconium Xylenol Orange Reaction', *Bunseki Kagaku*, **31**, 224–229 (1982).
297. K. Toda, I. Sanemasa, T. Deguchi, 'Flow Injection Analysis of Fluoride Ions Using Catalytic Reactions', *Bunseki Kagaku*, **34**, 31–35 (1985).
298. R.P. Pantaler, I.V. Pulyaeva, 'Kinetic Determination of Carbonates in Alkali Metal Halides of Special Purity', *Zh. Anal. Khim.*, **40**, 1854–1858 (1985).
299. N. Maniasso, S. Sato, M.F. Gine, A.O. Jacintho, 'Catalytic Determination of Dissolved Inorganic Carbon in Natural Waters by Flow Injection Spectrophotometry', *Analyst*, **121**, 1617–1619 (1996).
300. S. Funahashi, M. Tabata, M. Tanaka, 'Determination of Submicrogram Amounts of Iodide by Its Catalytic Effect on the Substitution Reaction of Mercury(II)-PAR Complex with CyDTA', *Anal. Chim. Acta*, **57**, 311–315 (1971).
301. A. Sakuragawa, 'Spectrophotometric Determination of a Micro Amount of Phosphate Ion by Its Catalytic Effect of the Zirconium(IV)–Xylenol Orange Reaction', *Bunseki Kagaku*, **35**, 926–930 (1986).
302. S. Yamada, F. Umika, M. Nakamura, S. Nakamura, 'Catalytic Effect of Some Inorganic Ligands on a Ligand Substitution Reaction Involving Mercury(II) and Its Application as a Differential Kinetic Method of Analysis', *Talanta*, **43**, 1715–1720 (1996).
303. A. Sakuragawa, T. Okutani, 'Spectrophotometric Determination of a Micro Amount of Sulfate by Its Catalytic Effect of the Zirconium–Xylenol Orange Reaction', *Bunseki Kagaku*, **32**, 382–386 (1983).

Data Treatment and Error Analysis in Kinetics

Juan J. Baeza-Baeza and Francisco F. Pérez-Pla
University of Valencia, Valencia, Spain

1 Introduction **1**
2 Data Treatment in Chemical Kinetics **1**
2.1 The Scope of Chemical Kinetics 1
2.2 The Reaction Rate 2
2.3 The Rate Law 2
2.4 The Mechanism of Reaction and the Rate Law 3
2.5 Determination of Simple Rate Laws 4
2.6 Treatment of Complex Mechanism 8
3 Error Analysis in Chemical Kinetic Methods **12**
3.1 Factors Influencing the Reaction Rate 12
3.2 Methods Based on First-order Reactions for the Determination of One Species 13
3.3 Multicomponent Methods 18
Abbreviations and Acronyms **24**
Related Articles **24**
References **24**

Methods of data treatment in kinetics are reviewed from a practical viewpoint. The equations to describe the kinetics of chemical reactions and the methodology to obtain the models are presented. Moreover, the integration methods that are needed for the simulation of complex reactions are explained, from the simple first-order reaction to the complex general mechanism.

However, it is essential to know the main sources of error in data treatment in order to reduce their effects on the results. Stress is put on the analysis of the errors due to the simplification of complex models. The different factors that influence the kinetics of the reaction are reviewed. How to take advantage of the kinetic information contained in the experimental data with an analytical aim is described. Finally, the propagation of errors in the analytical results is studied, reviewing both the simple univariate methods, used to determine one species, as well as the complex multivariate methods devoted to the determination of multicomponent mixtures.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Modern kinetic methods have benefited from the development of sophisticated data treatment techniques, the increasing availability of inexpensive computers, and the tremendous advances in instrumentation that is able to provide reliable data. For many years, most kinetic-based methods have been performed under pseudo-first-order conditions, which use simpler mathematical procedures. This approach is simplification of the reaction mechanism, carried out with an excess of reagent concentration, and aims to produce a reduction in the reaction order. However, the effective exploitation of the analytical potential of kinetics determinations requires the processing, without simplifications, of the actual reaction mechanism. Several approaches, such as the fractional time method, can be applied to elucidate the rate equations from which the concentrations of the species involved in the reaction can be calculated as a function of time. The differential nature of rate equations is a particular characteristic of kinetic methods, and it makes necessary the use of numerical methods to integrate accurately the rate equations.

Most kinetic methods are strongly dependent on factors affecting the reaction rate, such as temperature, ionic strength, or pH. Thus, reproducibility in the experimental conditions in each determination is critical, because variations can lead to important errors. Therefore, to assure the quality of the results, an effective data treatment in kinetics determinations requires an error analysis that quantifies the uncertainty inherent to the kinetic measurement. The study of error sources and propagation provides a means to obtain the optimal experimental design, thus improving the final results. An additional control of errors can be obtained by using multivariate methods, aimed at determining simultaneously several species, together with multidimensional data. The Kalman filter (KF) algorithm has been widely used, giving excellent results and providing error compensation capabilities, so that a large variation in rate constants is tolerated with little error.

2 DATA TREATMENT IN CHEMICAL KINETICS

2.1 The Scope of Chemical Kinetics

Chemical kinetics is concerned with the measurement and interpretation of the reaction rates, and, traditionally, it has been divided into molecular kinetics, dealing with the description of how reactions proceed at a molecular level, and formal kinetics, devoted to the empirical study of the reacting systems. More precisely, the objective of formal

kinetics is to induce from experience a chemical model, *the reaction mechanism*, explaining how the reaction takes place, and to develop a suitable mathematical model, *the rate law*, explaining the concentration variation with time of all chemical species of the reacting system.

Analytical chemists are mainly interested in the methodology of formal kinetics, since the mechanism of a reaction and the rate law should be known prior to their usage as an analytical tool.

2.2 The Reaction Rate

Let us suppose a reaction taking place according to the stoichiometric equation shown in Equation (1):

$$\nu_A A + \nu_R R \longrightarrow \nu_P P \tag{1}$$

where A and R are reactants and P the product. The rate of such a process in a closed reactor is defined in Equation (2) as the variation with time in the number of moles generated by the chemical reaction:

$$v = \frac{1}{\nu_i}\frac{1}{V}\frac{dn_i}{dt} \tag{2}$$

where n_i is the number of moles of species i, V the reactor volume, and ν_i the stoichiometric coefficient, which is taken as positive for products and negative for reactants to ensure the positive sign of the reaction rate. The introduction of ν_i and V makes the above definition independent of the chemical species used to follow the reaction, and on the reacting system size, respectively. Equation (2) turns into the usual definition of reaction rate, when volume remains constant during the reaction as shown in Equation (3):

$$v = \frac{1}{\nu_i}\frac{d}{dt}\frac{n_i}{V} = \frac{1}{\nu_i}\frac{dc_i}{dt} \tag{3}$$

where c_i is the concentration of the ith reactant. Strictly speaking, the volume remains constant only for gas reactions taking place in a closed rigid reactor. For reactions in solution, Equation (3) does not stand exactly, since the volume of the reaction mixture depends on composition and, consequently, on time. Nevertheless, the volume can be considered as a constant for dilute solutions reacting isothermally. On the other hand, most of the physical properties used to follow the kinetics of a reaction depend linearly on the molar concentration of species.

2.3 The Rate Law

Reactions can be divided into elementary and complex according to their kinetic mechanism. An *elementary* reaction takes place in a single step through a unique transition state. The transition state is the complex formed by the reactants to achieve the products. Figure 1 illustrates the energy changes during this process. There is an energy barrier to the reaction because reactants must take energy to form the transition state. The difference in energy between the transition state and the reactants is called the activation energy. On the other hand, complex reactions consist of two or more elementary steps.

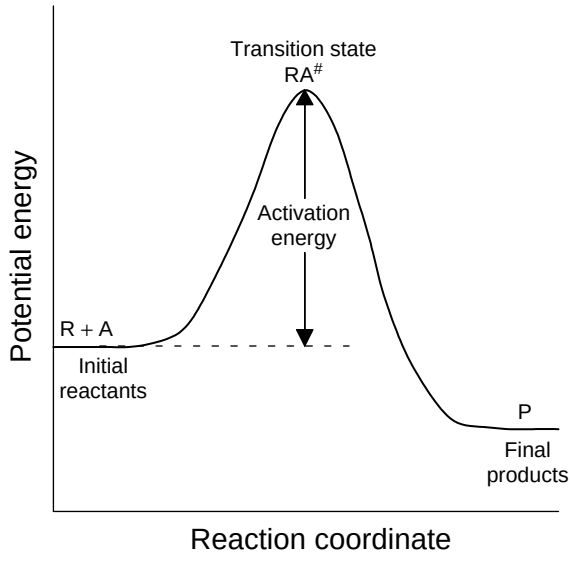

Figure 1 Potential energy–reaction coordinate diagram for a hypothetical elementary reaction. (The final energy is not always lower than the initial energy.)

The rate law relates the reaction rate with system composition (at constant temperature and pressure), and its mathematical form can be established from stoichiometry only for elementary reactions. According to the Law of Mass Action, the general rate equation of an elementary reaction is shown by Equation (4):

$$v = \frac{1}{\nu_i}\frac{dc_i}{dt} = k\prod_{i=1}^{M} c_i^{\nu_i} \tag{4}$$

where k is the rate constant (or specific rate) of the reaction, M the number of reactants, and ν_i is taken as positive. The rate constant depends on pressure, on temperature, and, for reactions taking place between ions, on electrolyte concentration.

For complex reactions, the comprehensive experimental work carried out in chemical kinetics from the beginning of the 20th century allows the rate laws to be classified into simple and complicated.

A simple rate law can be expressed, in a similar manner to elementary reactions, by Equation (5):

$$v = \frac{1}{\nu_i}\frac{dc_i}{dt} = k\prod_{i=1}^{M} c_i^{\gamma_i} \tag{5}$$

where γ_i is the reaction order with regard to the ith reactant. The reaction orders are empirical parameters induced from experience and whose values coincide with the stoichiometric coefficients only for elementary reactions. The sum of the reaction orders is the *overall order* of the reaction. For reactions with simple rate laws, the change of concentrations with time can be described by a single extent of reaction. The study of the kinetics of a single reaction is reduced to calculating k and the reaction orders.

2.4 The Mechanism of Reaction and the Rate Law

As previously stated, elementary reactions take place though a unique transition state and their rate law can be easily established. However, most chemical reactions are not elementary, and must be described by a *reaction mechanism*, which shows how the reaction takes place and must include all the contributing elementary reactions and the intermediate species formed. Thus, the reaction mechanism is the *chemical model* associated with a reaction while the rate law is the associated *mathematical model*. Once the mechanism is postulated, the mathematical model, which consists of an ordinary differential equation (o.d.e.) system, can be easily derived. Each differential equation of the system is the global derivative of the species concentration with regard to time, which is obtained by summation over the reaction rates of the elementary reactions where the species takes part, multiplied by its stoichiometric coefficient.

Let us consider, as an example, a simple enzyme-catalyzed reaction with the stoichiometry shown in Equation (6):

$$\mathrm{E} + \mathrm{S} \longrightarrow \mathrm{E} + \mathrm{P} \tag{6}$$

where E is the enzyme and S the substrate. Equation (7) is the simplest kinetic mechanism that can account for this reaction with three elementary steps:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{7}$$

where ES is an enzyme–substrate complex. From this mechanism, it is possible to build the differential equation system shown in Equations (8–11):

$$\frac{\mathrm{d[E]}}{\mathrm{d}t} = -k_1[\mathrm{E}][\mathrm{S}] + k_{-1}[\mathrm{ES}] + k_2[\mathrm{ES}] \tag{8}$$

$$\frac{\mathrm{d[S]}}{\mathrm{d}t} = -k_1[\mathrm{E}][\mathrm{S}] + k_{-1}[\mathrm{ES}] \tag{9}$$

$$\frac{\mathrm{d[ES]}}{\mathrm{d}t} = k_1[\mathrm{E}][\mathrm{S}] - k_{-1}[\mathrm{ES}] - k_2[\mathrm{ES}] \tag{10}$$

$$\frac{\mathrm{d[P]}}{\mathrm{d}t} = k_2[\mathrm{ES}] \tag{11}$$

The concentration of species is obtained by integration of the o.d.e., but most o.d.e. systems do not have analytical solutions and must be integrated numerically. However, taking into account the relative rate at which the elementary reactions occur, sometimes the mathematical model can be simplified. The most popular simplifications arise from the use of the *steady-state* and the *pre-equilibrium state* hypotheses, together with the assumption of the existence of a *rate-limiting step*.

2.4.1 Steady-state Hypothesis

Usually, reactions proceed through highly reactive species such as free radicals, ions, or reaction sites on the surface of a catalyst or enzymes. Owing to the intrinsic instability of these intermediates, their concentration will tend to remain low and constant compared to those of reactants and products. These facts are the basis of the steady-state hypothesis, in which the accumulation rate of a highly reactive intermediate is considered null and its associated differential equations equal to zero. Thus, its o.d.e. becomes an algebraic equation that allows the steady concentration of the intermediate to be related to the concentration of reactants and products. Since the concentration of the highly reactive intermediate remains low compared to that of reactants and products, it can be neglected in the concentration balances.

Steady-state hypothesis is often used to simplify enzymatic catalysis data treatment because once the steady state is accomplished for the enzyme–substrate complex, ES, the differential Equation (10) turns into the algebraic equation shown in Equation (12):

$$k_1[\mathrm{E}][\mathrm{S}] - k_{-1}[\mathrm{ES}] - k_2[\mathrm{ES}] = 0 \tag{12}$$

This equation allows us to express the ES concentration as a function of the reactant concentrations:

$$[\mathrm{ES}] = \frac{k_1}{k_{-1} + k_2}[\mathrm{E}][\mathrm{S}] \tag{13}$$

The overall reaction rate is determined by the product formation rate and after substituting Equation (13) into Equation (11), it yields Equation (14):

$$v = \frac{\mathrm{d[P]}}{\mathrm{d}t} = k_2[\mathrm{ES}] = \frac{k_2 k_1}{k_{-1} + k_2}[\mathrm{E}][\mathrm{S}] \tag{14}$$

This solution will be valid as long as the steady-state approximation remains valid. At the beginning of the reaction this hypothesis cannot be applied because the concentration of the intermediate is building up from its initial value of zero to the steady-state value.

2.4.2 Rate-limiting and Pre-equilibrium Steps

The reaction rate of the elementary steps of a mechanism can be very different. When there is a step much slower

than the others, it controls the rate of the overall reaction and the formation rate of final products equals the rate of this step, that is called the *rate-limiting step*. On the other hand, some mechanisms exhibit a rate-limiting step preceded by a fast reversible step. Then, it can be considered that the fast reversible step reaches the equilibrium being unperturbed by the following much slower step. Now the concentration of the intermediate can be obtained by solving the equilibrium equation. A system of this type is said to possess a fast *pre-equilibrium step*. Proton transfers constitute a very important class of fast pre-equilibrium step as in acid-catalyzed reactions. Pre-equilibria steps can be found in other catalyzed reactions. One simple mechanism often encountered is shown in Equations (15) and (16):

$$\mathrm{R} + \mathrm{C} \Longleftrightarrow \mathrm{RC} \tag{15}$$

$$\mathrm{RC} + \mathrm{A} \xrightarrow{k_c} \mathrm{P} + \mathrm{C} \tag{16}$$

where the first step is a fast equilibrium and the second the rate-limiting step, k_c being the catalytic rate constant and C the catalyst. The reaction rate is given by Equation (17):

$$v = \frac{\mathrm{d[P]}}{\mathrm{d}t} = -\frac{\mathrm{d[A]}}{\mathrm{d}t} = k_c[\mathrm{RC}][\mathrm{A}] \tag{17}$$

The concentration of the intermediate complex, RC, can be obtained by applying the pre-equilibrium hypothesis as shown in Equation (18):

$$K = \frac{[\mathrm{RC}]}{[\mathrm{R}][\mathrm{C}]} \longrightarrow [\mathrm{RC}] = K[\mathrm{R}][\mathrm{C}] \tag{18}$$

from which the rate law can be given by Equation (19):

$$v = k_c K[\mathrm{R}][\mathrm{C}][\mathrm{A}] \tag{19}$$

Therefore, many complex reactions exhibit simple rate laws when the steady-state or the pre-equilibrium and limiting reaction steps hypothesis can be applied.

2.5 Determination of Simple Rate Laws

Most of the reactions of analytical interest follow the rate law given by Equation (5) after applying the steady-state or rate-limiting step approximations to the o.d.e. derived from their mechanisms. Generally, the study of the kinetics of a reaction involves measurement of a physical property like those summarized in Table 1. For this reason, it is convenient to express the rate law as a function of the measured physical property.

2.5.1 Expression of the Concentrations as a Function of a Physical Property

Let us assume that the kinetics of a reaction is monitored by measuring a physical property $S(t)$ that is a linear function of the composition of the reaction mixture given by Equation (20):

$$S = S_b + \sum_{i=1}^{N} s_i c_i \tag{20}$$

where N is the number of species in the system, S_b is the property measured from the blank and s_i is the proportionality coefficient between the concentration of the ith species and the physical property; the values of these constants can be estimated from equilibrium experiments when possible. The s coefficients are supposed to be independent of time and composition. If the rate law of the reaction is simple, the concentration variations will be described by a single extent of reaction, ξ, according to

Table 1 Physical properties commonly used to study the kinetics of a reaction

Physical property	Commentary
Light absorption (ultraviolet, visible, infrared)	Sensitive. Possibly the most frequently used physical property. Specially useful to follow fast kinetics
Light emission (fluorescence)	More specific and sensitive than light absorption. Allows following fast kinetics at the trace level
Electric conductivity	Little selectivity. The specific conductivity of ions is affected by the ionic environment and changes with the reaction mixture composition
Volume change	Little sensitivity. Only useful to follow slow reactions leading to the gelification or solidification of the reactive mixture
Pressure change	Property used to follow slow reactions in the gas phase
Rotation of the polarization plane of light	Property used to follow kinetics involving optically active species
Heat interchange	Little sensitivity and selectivity. Used to follow polymerization reactions or phase change kinetics
Multinuclear nuclear magnetic resonance absorption	Little sensitivity but selective. Used to follow slow organic reactions or some isomerization processes. Sensitive to the saturation of the signal. The property is not always linear

Equation (21):

$$c_i = c_{0,i} + \nu_i \xi \tag{21}$$

where $c_{0,i}$ is the initial concentration of ith species. The substitution of Equation (21) into (20) yields:

$$S = S_b + \sum_{i=1}^{N} s_i (c_{0,i} + \nu_i \xi)$$
$$= \left(S_b + \sum_{i=1}^{N} s_i c_{0,i} \right. + \left(\sum_{i=1}^{N} s_i \nu_i \right. \quad \xi = S_0 + S^0 \xi \tag{22}$$

where S_0 is the property value at zero time. From Equations (21) and (22), Equation (23) can be obtained:

$$\xi = \frac{S - S_0}{S^0} = \frac{c_i - c_{0,i}}{\nu_i} \tag{23}$$

At the end of the reaction, it becomes Equation (24):

$$\frac{S_\infty - S_0}{S^0} = \frac{c_{\infty,i} - c_{0,i}}{\nu_i} \tag{24}$$

where $c_{\infty,i}$ is the concentration of the ith species when the reaction has finished, and S_∞ the signal value at infinite time. From Equations (23) and (24), we obtain Equation (25):

$$\alpha = \frac{c_i - c_{\infty,i}}{c_{0,i} - c_{\infty,i}} = \frac{S_\infty - S}{S_\infty - S_0} \tag{25}$$

Let us call A the limiting reactant for which the final concentration, a_∞, is zero, then Equation (25) gives:

$$\alpha = \frac{a}{a_0} = \frac{S_\infty - S}{S_\infty - S_0} \tag{26}$$

Equation (26) is the basis of the data treatment of simple rate laws, because they can be expressed as a function uniquely of the α variable. It is important to realize that α is a dimensionless variable ranging from 0 to 1, and that it depends directly on the signal value.

2.5.2 First-order Reactions

For an elementary first-order reaction the stoichiometric equation will be given by Equation (27):

$$\mathrm{A} \xrightarrow{k} \mathrm{P} \tag{27}$$

and the corresponding differential rate equation by Equation (28):

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -ka \tag{28}$$

where a is the concentration of reactant A at time t. After separating variables, this equation can be integrated as in Equation (29):

$$\int_{a_0}^{a} \frac{\mathrm{d}a}{a} = -\int_{0}^{t} k \, \mathrm{d}t \tag{29}$$

where a_0 is the concentration of reactant A at time zero. The integration yields Equation (30):

$$\ln\left(\frac{a}{a_0}\right. = \ln\left(\frac{S_\infty - S}{S_\infty - S_0}\right. = -kt \tag{30}$$

where Equation (26) has been used.

This expression can be rewritten as Equation (31) in terms of the variation of the physical property with time:

$$S = S_\infty - (S_\infty - S_0)\mathrm{e}^{-kt} \tag{31}$$

Finally, first-order reactions are characterized by a semi-reaction time ($t_{1/2}$, time needed to decrease the initial concentration of limiting reactant to half of its initial value) independent of the initial composition of the mixture, as shown in Equation (32):

$$t_{1/2} = \frac{\ln(2)}{k} \tag{32}$$

Many of the reactions of analytical interest can be treated as first-order reactions when experimental conditions are chosen adequately. For instance, a second-order reaction between reactants A and R behaves as a first-order reaction when one of the reactants, R, is present in excess ($[\mathrm{R}] \cong [\mathrm{R}]_0 \gg [\mathrm{A}]_0$) as shown in Equations (33) and (34):

$$\mathrm{A} + \mathrm{R} \xrightarrow{k_2} \mathrm{P} \tag{33}$$

$$\nu = k_2[\mathrm{R}][\mathrm{A}] \xrightarrow{[\mathrm{R}] \cong [\mathrm{R}]_0 \gg [\mathrm{A}]_0} \nu \cong k_2[\mathrm{R}]_0[\mathrm{A}] = k_1[\mathrm{A}] \tag{34}$$

In this case, we say that the second-order reaction has been transformed into a pseudo-first-order reaction. The ability to reduce the reaction order is a very valuable experimental tool, that often permits the simplification of the reaction kinetics.

2.5.2.1 Calculation of the Rate Constant The determination of the rate law for first-order reactions reduces to the calculation of the rate constant. The fitting of experimental data to Equation (35):

$$\ln(|S_\infty - S|) = \ln(|S_\infty - S_0|) - kt \tag{35}$$

gives the rate constant from the slope and, from the intercept, the initial concentration can be obtained. However, this method is not applicable when the final property cannot be accurately measured. This is the case for very slow reactions for which it is not

valuable to wait for completion, or when there are slow secondary processes which make it difficult to know the exact contribution of the main reaction to the value of S_∞. In such cases, it is recommended to use the methods developed by Guggenheim or Kezky–Mangerlsdorf–Swinbourne (KMS).[1]

For both methods the physical property is measured at two time series, $\{t_1, t_2, \ldots\}$ and $\{t_1 + \Delta t, t_2 + \Delta t, \ldots\}$, where Δt is a constant time increment. The *Guggenheim* method takes the differences between the physical properties measured in both series as shown in Equation (36):

$$S(t + \Delta t) - S(t) = (S_\infty - S_0)(1 - e^{-k\Delta t})e^{-kt} \quad (36)$$

This expression can be linearized by taking logarithms on both sides of the equation to give Equation (37):

$$\ln|S(t + \Delta t) - S(t)| = -kt + \ln[|S_\infty - S_0|(1 - e^{-k\Delta t})] \quad (37)$$

Thus, a fit of the logarithm of the difference in the physical property values versus time should give a line, from whose slope the rate constant can be calculated. In the least-squares fitting of Equation (37) it is recommended to use the weight as in Equation (38):

$$w(t) = [S(t + \Delta t) - S(t)]^2 \quad (38)$$

to correct the perturbation due to the logarithm linearization.[2]

The *KMS* method relates directly the data of both series, avoiding the signal difference calculation, as in Equation (39):

$$S(t + \Delta t) = S(t)e^{-k\Delta t} + S_\infty(1 - e^{-k\Delta t}) \quad (39)$$

Thus, a plot of $S(t + \Delta t)$ versus $S(t)$ yields a line with slope equal to $e^{-k\Delta t}$, from which it is possible to calculate the rate constant. In both methods the value of the time increment is an important factor, as illustrated in Figure 2. If Δt is small, the relative error in the signal difference will be large; on this basis, Δt should be as large as possible. However, the larger Δt the smaller the number of data points used; then, an intermediate Δt must be taken. It is recommended to take Δt between $t_{1/2}$ and $2t_{1/2}$.

Finally, data can be fitted to Equation (31) directly using a nonlinear method. In this case k, S_0 and S_∞ are adjustable parameters.

2.5.3 General Data Treatment for Simple Rate Laws

Let us suppose a reaction takes place according to Equation (1), and following the simple rate law given by Equation (40):

$$\frac{1}{\nu_A}\frac{da}{dt} = \frac{1}{\nu_R}\frac{dr}{dt} = -ka^p r^q \quad (40)$$

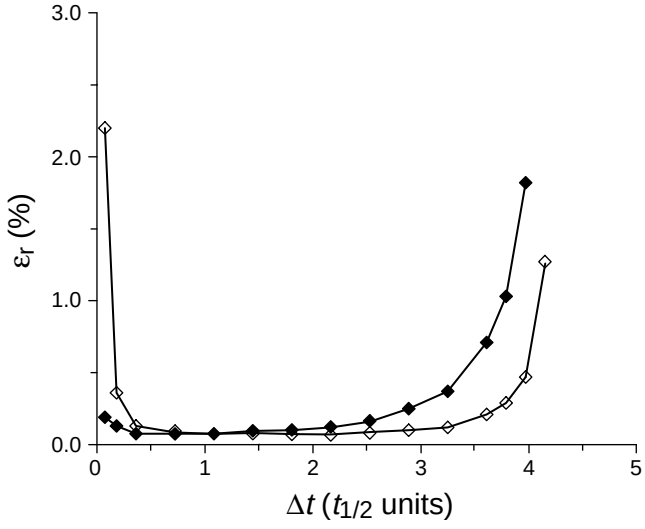

Figure 2 Average relative error (percentage) in the evaluation of a first-order rate constant ($k = 1\,s^{-1}$) using the Guggenheim (◇) and KMS ( ) methods, versus the Δt used.

where the stoichiometric coefficients are taken as positive, and p and q are the reaction orders with regard to the limiting reactant A and the co-reactant R, respectively. According to Equation (21), the concentrations of A and R are related by Equation (41):

$$r = r_0 - \frac{\nu_R}{\nu_A}(a_0 - a) = r_\infty + \frac{\nu_R}{\nu_A}a \quad (41)$$

The substitution of Equation (41) into Equation (40) leads to Equation (42):

$$\frac{da}{dt} = -\nu_A k a^p \left(r_\infty + \frac{\nu_R}{\nu_A}a\right)^q \quad (42)$$

This equation is commonly expressed as a function of the α variable defined in section 2.5.1. Taking into account that $\alpha = a/a_0$ yields Equation (43):

$$\frac{d\alpha}{dt} = -(\nu_A k a_0^{p+q-1})\alpha^p \left(\frac{r_\infty}{a_0} + \frac{\nu_R}{\nu_A}\alpha\right)^q \quad (43)$$

The benefits of this change of variable are evident; since we know how to relate the α value to a physical property, it will be possible to fit directly the measured property expressed in terms of α. In addition, the resulting differential equation has the dimensionless α as a unique variable, and analytical solutions for a large number of reaction orders, p and q, are tabulated.[3] Now, the rate law will be determined when the parameters k, p and q are known. Indeed, the rate law depends on the ratio of the stoichiometric coefficients; thus the study of the stoichiometry of the reaction must be done prior to the kinetic study. The objective of the next section is to give a review of methods allowing the determination of k, p and q.

2.5.4 Determination of the Reaction Orders and the Rate Constant

2.5.4.1 The Fractional Time Method The fractional time method[3] makes it possible to estimate the reaction orders and the rate constant from two kinetic experiments, one with the reactants being present initially at the stoichiometric ratio ($r_\infty = a_\infty = 0$), and the other carried out with excess of one of the reactants ($r_0 \gg a_0$, then $r_\infty/a_0 \gg \alpha$). In these cases, Equation (43) reduces to:

$$-\frac{\mathrm{d}\alpha}{\mathrm{d}t} = k^*\alpha^m \tag{44}$$

where k^* is a constant and $m = p + q$ in the first case, or $m = p$ in the second. The integration of Equation (44), when $m \neq 1$, leads to Equation (45):

$$\frac{1}{1-m}(1-\alpha^{1-m}) = k^*t \tag{45}$$

This method takes a series of pairs (t, α) covering a wide interval of α values [for instance $\alpha \in (0.1, 0.95)$], as shown in Equation (46):

$$\alpha_1 = f\alpha_0, \alpha_2 = f\alpha_1, \ldots, \alpha_j = f\alpha_{j-1}, \ldots, \alpha_n = f\alpha_{n-1} \tag{46}$$

where f is a constant in the interval $0 < f < 1$. Very often, some values will not be available because no measurements have been done, and, consequently, they will have to be interpolated, for instance by using the orthogonal polynomial method.[4]

Equation (47) shows the difference between two consecutive values of α, deduced from Equation (45):

$$\frac{1}{1-m}\left(\alpha_{j-1}^{1-m} - \alpha_j^{1-m}\right) = k^*(t_j - t_{j-1}) \tag{47}$$

which can be rearranged to Equation (48):

$$\alpha_j^{1-m}\frac{1-f^{m-1}}{k^*(1-m)} = (t_j - t_{j-1}) \tag{48}$$

On taking logarithms this leads to Equation (49):

$$\ln(t_j - t_{j-1}) = \ln\left(\frac{1-f^{m-1}}{k^*(1-m)}\right) + (1-m)\ln(\alpha_j) \tag{49}$$

A plot of the logarithm of $t_j - t_{j-1}$ versus the logarithm of α_j should yield a line whose slope and intercept give information on the value of m and the rate constant, respectively. Once the global order $p + q$ and the partial order p are known, the partial order q can be obtained by difference.

2.5.4.2 The Initial Rate Method The differential methods use the derivative of the physical property with respect to time, instead of the signal itself. The differential methods that use all the information (t, S) available are rarely applied because the numerical differentiation of all signal profile could increase remarkably the noise level. Nevertheless, the last affirmation is no longer true for experimental kinetic methods that measure the reaction rate directly as the electrochemical amperometric monitoring of an electrode reaction. An advantage of differential methods is that they do not need to integrate any differential equation since the reaction rates are dependent on the composition of the system through an algebraic equation system.

The initial rate method is a differential method that uses, as a dependent variable, the reaction rate at zero time, at which the composition of the system is known. Let us assume a reaction is taking place according to Equation (1). If the rate law is simple, the initial reaction rate is as shown in Equation (50):

$$v_0 = ka_0^p r_0^q \tag{50}$$

where the subscript "0" refers to zero time. Taking logarithms on both sides of the expression leads to Equation (51):

$$\ln(v_0) = \ln(k) + p\ln(a_0) + q\ln(r_0) \tag{51}$$

In this equation there are three parameters, k, p and q; thus it is necessary to measure at least three reaction rates to determine them by fitting Equation (51). The experimental design should be such that one initial concentration is kept constant whereas the other is changed and vice versa. In this way each set of data can be related independently to parameters p and q.

The correct evaluation of the initial rate is an important question. It can be estimated as the ratio between the variation of the signal and the interval of time elapsed at the beginning of the reaction. Therefore, for a first-order reaction given by Equation (52):

$$v_0^{\mathrm{est}} = \frac{S_2 - S_1}{t_2 - t_1} = (S_\infty - S_0)\frac{\mathrm{e}^{-kt_1} - \mathrm{e}^{-kt_2}}{t_2 - t_1} \tag{52}$$

The error associated can be calculated by means of a Maclaurin series expansion of the exponential term, shown in Equation (53):

$$\mathrm{e}^{-kt} = 1 - kt + \frac{1}{2}(kt)^2 - \cdots = \sum_{i=0}^{\infty}(-1)^i\frac{(kt)^i}{i!} \tag{53}$$

Since kt is close to zero near the origin, only the first three terms can be used, as shown in Equation (54):

$$v_0^{\mathrm{est}} \cong (S_\infty - S_0)k\left(1 + \frac{kt_1 + kt_2}{2}\right. \tag{54}$$

The exact initial rate can be evaluated by differentiating Equation (31) as shown in Equation (55):

$$v_0 = \left(\frac{\mathrm{d}S}{\mathrm{d}t}\right)_{t=0} = (S_\infty - S_0)k \tag{55}$$

Finally, the relative systematic error is given in Equation (56):

$$\varepsilon_r = \frac{v_0^{\mathrm{est}} - v_0}{v_0} = -\frac{kt_1 + kt_2}{2} \cong -\frac{\xi_{r1} + \xi_{r2}}{2} \tag{56}$$

where ξ_r is the relative extent of reaction, given by Equation (57):

$$\xi_r = \frac{a_0 - a}{a_0} = 1 - \mathrm{e}^{-kt} \tag{57}$$

and $\xi_r \cong kt$ for values of time small enough. Thus the relative systematic error will coincide with the average relative extent of reaction of the experimental points used to evaluate the initial rate.

Commonly, the signal values are not reliable close to the origin, due to the mixing effect and the induction period; therefore, a better approach is to fit the initial measured pairs (t, S) to an orthogonal polynomial series[5] from which the smoothed derivative is calculated and extrapolated at zero time.

The initial rate measurements are faster than those based on integral methods since the reaction is not required to reach completion, and can be applied to very slow reactions. Another advantage is that complications from side reactions are avoided. However, the mechanistic conclusions obtained about the reaction orders are only valid for the beginning of the reaction since these parameters can change, especially for reactions with complicated rate laws. Therefore, it is always highly recommended to complete the study with the aid of an integral method.

2.5.4.3 Use of Integrated Equations The use of the integrated rate equation is, maybe, the most popular technique to determine the rate constant. This method is acceptable once the reaction orders are known, because the integrated rate equation depends on the p and q values. When reaction orders are unknown, or if it is suspected that they can take fractional values, the methods given above are preferable, since in this case, the use of the integrated equation should be based on a trial and error method, and it is really difficult to distinguish between two close fractional values. Table 2 shows integrated equations[6] found frequently in the experimental work.

2.5.4.4 Mixed Simple–Complicated Equations Very often, it is found that reactions with complicated rate equations present a simple form for one of the reactants and a complicated form for the other. In this case, the rate law could be written as shown in Equation (58):

$$-\frac{\mathrm{d}a}{\mathrm{d}t} = f(r)a^p \tag{58}$$

where it was supposed that the kinetic behavior of reactant A was simple while reactant R presented a complicated form. In this situation, the kinetic study is carried out by running a reaction series where the reactant R is in excess.[7] Under these experimental conditions, where less than 1–10% of R is consumed, the rate law takes the simple form shown in Equation (59):

$$-\frac{\mathrm{d}a}{\mathrm{d}t} = f(r)a^p \approx f(r_0)a^p = k_{\mathrm{app}}a^p \tag{59}$$

Now, the techniques seen in the above sections enable us to find the reaction order with respect to the limiting reactant, but the dependence of the rate law on r must be deduced from a plot of k_{app} versus the initial concentration r_0.

A typical example is the dependence of most enzymatic reactions on pH; the rate presents a simple dependence on the substrate concentration, but the dependence on the proton concentration frequently exhibits a sigmoidal or bell-shaped curve explained by the acid–base pre-equilibria established among the enzyme, activated complex, and substrate with protons.

There is not a systematic procedure to deduce the mechanism of a reaction from its $k_{\mathrm{app}}(r_0)$ curve. Generally, several mechanisms are proposed based on additional information such as the structure of the products or the reactivity of the functional groups involved. Once the reaction mechanism has been derived, it is simplified on the basis of the steady state or the presence of pre-equilibria, and a rate law is obtained from which it is possible to get a theoretical equation for $k_{\mathrm{app}}(r_0)$, which is compared to the experimental data in order to calculate the more relevant kinetic parameters.

The procedure does not ensure the validity of a mechanism whose curve $k_{\mathrm{app}}(r_0)$ fits correctly the experimental one, but permits us to reject all mechanisms from which it is not possible to deduce a reasonable shape for $k_{\mathrm{app}}(r_0)$.

2.6 Treatment of Complex Mechanism

Complicated rate laws arise from complex reactions that need more than one differential equation and two or more extents of reaction to be described. For instance, the reaction used for the determination of oxygen peroxide with dichromate in acidic media[8] proceeds in two steps, the formation of the blue peroxychromic acid and its decomposition to yield aqueous Cr(III). The reaction mechanism consists of several elementary steps, and even

Table 2 Summary of integrated rate expressions for selected single reaction rate equations

p	q	Integrated rate law
0	0	$1-\alpha = ka_0^{-1}t$
$\frac{1}{2}$	0	$1-\alpha^{1/2} = \frac{k}{2}a_0^{-1/2}t$
1	0	$\ln\left(\frac{1}{\alpha}\right. = kt$
2	0	$\frac{1}{\alpha} - 1 = ka_0t$
0	$\frac{1}{2}$	$(1+c)^{1/2} - (\alpha+c)^{1/2} = \frac{k}{2}a_0^{-1/2}v^{1/2}t$
$\frac{1}{2}$	$\frac{1}{2}$	$\ln\ \frac{(1+c)^{1/2}+1}{(\alpha+c)^{1/2}+\alpha^{1/2}} = \frac{k}{2}v^{1/2}t$
0	1	$\ln\left(\frac{1+c}{\alpha+c}\right. = kvt$
1	1	$c>0, c<0$ $\ln\ \frac{\alpha+c}{\alpha(1+c)} = kca_0vt$
		$c=0$ $\frac{1}{\alpha} - 1 = ka_0vt$
0	2	$\frac{1}{a+c} - \frac{1}{1+c} = ka_0v^2t$
2	1	$c>0, c<0$ $\ln\ \frac{(1+c)\alpha}{\alpha+c} - c\left(1+\frac{1}{\alpha}\right. = c^{-2}ka_0^2vt$
		$c=0$ $\frac{1}{\alpha^2} - 1 = ka_0^2vt$
1	2	$c>0, c<0$ $\ln\ \frac{\alpha+c}{(1+c)\alpha} + c\left(\frac{1}{1+c} - \frac{1}{\alpha+c}\right. = c^{-2}ka_0^2v^2t$
		$c=0$ $\frac{1}{\alpha^2} - 1 = ka_0^2v^2t$

a, Concentration of limiting reactant A; a_0, r_0, initial concentration of species A and R; r_∞, final concentration of species R. $\alpha = a/a_0$, $c = r_\infty v_A/a_0v_R$.

applying the pre-equilibrium and the steady-state simplifications, it needs two reaction extent variables, one for the formation of the peroxychromic acid and the other for its decomposition, to describe all concentration variations.

Methods described in section 2.5 cannot be used for dealing with complex reactions with complicated rate laws, since the α variable relating the physical response and concentrations was deduced on the basis that a unique extent of reaction was needed.

The general data treatment of complex reactions, including those with complicated behavior, is carried out by developing a system of o.d.e.s from the reaction mechanism. The resolution of this system yields the theoretical concentration values, from which the monitored physical property can be estimated as a function of time. The comparison of estimated and experimental signal values by a regression method allows the model validity to be checked.

In section 3.3.2, several regression methods will be studied, but the next sections are devoted to describing methods capable of calculating the concentrations at any time from any reaction mechanism. Complex mechanisms can be conveniently divided into first-order mechanisms (first-order networks) and general mechanisms due to the different mathematical techniques that can be used to solve the o.d.e. system associated with first-order networks.

2.6.1 First-order Networks

A first-order network arises from reaction mechanisms made up entirely of first-order elementary steps. Typical

experiments leading to first-order complex mechanisms are those conducted under pseudo-first-order conditions, or those based on perturbation methods. The general o.d.e. system associated with these models is linear and takes the general form shown in Equation (60):

$$\frac{d\mathbf{c}}{dt} = \mathbf{A}\mathbf{c} \tag{60}$$

where $\mathbf{c}$ is a vector containing the species concentration, t is the reaction time, and $\mathbf{A}$ is a square $(N \times N)$ matrix, the elements of which are shown in Equation (61):

$$a_{ij} = k_{ji}, \quad a_{ii} = -\sum_{j=1}^{N} k_{ij} \tag{61}$$

where N is the number of species in the reaction system, and k_{ij} the rate coefficient for the reaction that converts i species into j species (many k_{ij} can have a zero value).

2.6.1.1 Integration of First-order Networks The first-order mechanisms are classified as reversible or irreversible. For the reversible mechanisms, the analytical solution is given by Equation (62):

$$c_i(t) = \sum_{j=1}^{N} u_{ij} g_j e^{\lambda_j t} \tag{62}$$

where c_i is the ith species concentration at time t, u_{ij} the elements of matrix $\mathbf{U}$ that contain the eigenvectors of matrix $\mathbf{A}$, λ_j being its jth eigenvalue, and g_j is a constant related to the initial conditions of the problem.

The reaction matrix $\mathbf{A}$ derived from a reversible network can be converted into a real symmetrical matrix, $\mathbf{A'}$, by applying the similarity transformation shown in Equation (63):

$$\mathbf{A'} = \mathbf{D}^{-1}\mathbf{A}\mathbf{D} \tag{63}$$

where $\mathbf{D}$ is a diagonal matrix containing the square root of the equilibrium concentrations of the species. The eigenvectors and eigenvalues of a symmetrical real matrix are always real and can be obtained by the simpler Jacobi's diagonalization method.[4] From the matrix $\mathbf{U'}$, containing the eigenvectors of matrix $\mathbf{A'}$, the matrix $\mathbf{U}$ is calculated as shown in Equation (64):

$$\mathbf{U} = \mathbf{D}\mathbf{U'} \tag{64}$$

The $\mathbf{g}$ vector ensures the compatibility of the general solution with the boundary conditions of the problem, and is calculated from the initial species concentration and the eigenvectors of $\mathbf{A}$. At zero time, Equation (62) gives Equation (65):

$$c_i(0) = \sum_{j=1}^{N} u_{ij} g_j \rightarrow \mathbf{c}(0) = \mathbf{U}\mathbf{g} \rightarrow \mathbf{g} = \mathbf{U}^{-1}\mathbf{c}(0) \tag{65}$$

This algorithm is faster than numerical methods and is not affected by the stiff nature of the kinetic system. Another advantage is that the calculation time is not affected severely by the number of experimental points.

The above method cannot be applied to an irreversible first-order mechanism because the reaction matrix $\mathbf{A}$, derived from an irreversible network, cannot be transformed into a real symmetrical matrix and it could have an incomplete set of eigenvalues. This leads to the reduction of exponential terms in Equation (36), and to the inclusion of polynomial terms, as illustrated in Table 3 for the consecutive mechanism when $k_1 = k_2$. Thus, for irreversible networks, the general solution for species concentration is given by Equation (66):

$$c_i(t) = \sum_{j=1}^{N} u_{ij} \frac{t^{m_j - 1}}{(m_j - 1)!} e^{\lambda_j t} \tag{66}$$

where m_j is the multiplicity of the jth eigenvalue of $\mathbf{A}$. Now, the Laplace transform technique[9] can be used to integrate irreversible first-order networks.[10]

This procedure is expected to work faster than numerical methods and is hardly dependent on the number of experimental points. The technique, however, needs more calculus time than the eigenvector method used for reversible networks, since the $\mathbf{U}$ matrix is calculated by solving N linear equation systems. This fact makes the procedure more size-dependent than the eigenvector method.

Table 3 Integrated rate laws for common first-order mechanisms[a]

Equilibrium reactions		
$A \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} B$	$a = a_0 \left\{1 - \frac{k_1}{k_1 + k_{-1}}[1 - e^{-(k_1 + k_{-1})t}]\right\}$	
	$b = a_0 \frac{k_1}{k_1 + k_{-1}}[1 - e^{-(k_1 + k_{-1})t}]$	
Consecutive reactions		
$A \xrightarrow{k_1} B \xrightarrow{k_2} C$	$a = a_0 e^{-k_1 t}$	
	$b = a_0 \frac{k_1}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t})$	$k_1 \neq k_2$
	$b = a_0 k t e^{-kt}$	$k_1 = k_2$
	$c = a_0 - a - b$	
Parallel reactions		
$A \overset{k_1}{\longrightarrow} B$, $A \overset{k_2}{\longrightarrow} C$	$a = a_0 e^{-(k_1 + k_2)t}$	
	$b = a_0 \frac{k_1}{k_1 + k_2}[1 - e^{-(k_1 + k_2)t}]$	
	$c = a_0 \frac{k_2}{k_1 + k_2}[1 - e^{-(k_1 + k_2)t}]$	

[a] Initial concentrations for A, B, C were a_0, 0, 0 respectively.

2.6.2 Data Treatment of a General Kinetic Model

When reactions higher than first-order appear in the reaction mechanism, the o.d.e. shown in Equation (67) describes the general reaction system:

$$\frac{dc_n}{dt} = \sum_{i=1}^{R} \nu_{in} k_i \prod_{j=1}^{N} c_j^{\gamma_{ij}} \tag{67}$$

where N is the number of species in the system, R the number of reactions, k_i is the rate coefficient associated with the ith reaction, ν_{in} the stoichiometric coefficient of the nth species in the ith reaction, and γ_{ij} the reaction order of the jth species in the ith reaction. In order to generalize the problem, the reaction orders, γ_{ij}, can take any real number. However, if the reaction mechanism is used, only elementary steps are involved and the orders will be equal to the stoichiometric coefficients.

Frequently, the number of reactions is lower than the number of species, and the number of differential equations can be reduced by carrying out a change of variable. The new variables will be the extents of reaction, ξ, which are linearly related to the concentrations as shown in Equation (68):

$$c_n = c_{n,0} + \sum_{j=1}^{R} \nu_{nj} \xi_j \tag{68}$$

where $c_{n,0}$ is the initial concentration of the n species. The new o.d.e. is described by Equation (69):

$$\frac{d\xi_i}{dt} = k_i \prod_{n=1}^{N} \left(c_{n,0} + \sum_{j=1}^{R} \nu_{nj} \xi_j \right)^{\gamma_{in}} \tag{69}$$

Now the mathematical model is a system of R differential equations instead of N.

2.6.3 Numerical Integration of Kinetic Systems

The ordinary methods used to solve numerically o.d.e. systems are based on the calculation of a solution at time $t + h$ (where h is the integration step) from a solution known at time t, an integration scheme that is known as *explicit*. The value of h cannot be taken arbitrarily because it controls the accuracy at which the new solution is calculated and, thus, the stability of the integration procedure; small steps waste calculation time and increase the round-off error, and unreasonably large steps result in a diverging integration.

Several explicit integration methods can be applied to kinetic differential equation systems, such as the Runge–Kutta type methods or the predictor–corrector methods.[4] However, step size control is cumbersome, and high accuracy is difficult to obtain. A better approach for kinetic problems is to apply the explicit Taylor method,[11] which allows a direct and complete control of the step size.

The Taylor method is based on developing the solution as a polynomial series. It is clear that the solution at times t and $t + h$ are related by the Taylor formula shown in Equation (70):

$$\mathbf{c}^{(0)}(t+h) = \mathbf{c}^{(0)}(t) + \sum_{k=1}^{Q} \frac{h^k}{k!} \mathbf{c}^{(k)}(t) \tag{70}$$

where $\mathbf{c}$ is a vector containing the concentrations, and the superscripts in parentheses refer to the derivative order. It is obvious that this expression can be applied if the Q first derivatives are known. This task can be done as follows. The first derivative of the concentration of the nth species at time t is given by Equation (71):

$$c_n^{(1)} = \sum_{i=1}^{R} \nu_{in} \quad k_i \prod_{j=1}^{N} c_j^{\gamma_{ij}} \quad = \sum_{i=1}^{R} \nu_{in} \Phi_i^{(1)} \tag{71}$$

Thus, deriving the above expression $(k + 1)$ times yields Equation (72):

$$c_n^{(k+1)} = \sum_{i=1}^{R} \nu_{in} \Phi_i^{(k+1)} \tag{72}$$

where the $(k + 1)$th derivative of Φ_i can be calculated with the algorithm in Equation (73), derived from its own definition after applying the Leibniz rule:[12]

$$\Phi_i^{(k+1)} = \sum_{j=1}^{k} \binom{k-1}{j-1} \quad T_i^{(j)} \Phi_i^{(k+1-j)} \tag{73}$$

where $T_i^{(j)}$ is obtained from Equations (74) and (75):

$$T_i^{(j)} = \sum_{n=1}^{N} \gamma_{in} f_n^{(j)} \tag{74}$$

$$f_n^{(j)} = \frac{c_n^{(j)}}{c_n^{(0)}} - \sum_{i=1}^{j-1} \binom{k-1}{i-1} \quad f_n^{(i)} \frac{c_n^{(j-i)}}{c_n^{(0)}} \tag{75}$$

The recurrent algorithm starts with the values shown in Equation (76):

$$f_n^{(1)} = \frac{c_n^{(1)}}{c_n^{(0)}}, \quad T_i^{(1)} = \sum_{n=1}^{N} \gamma_{in} f_n^{(1)} \tag{76}$$

The Taylor method allows the integration error to be controlled in a straightforward way. Equation (70) indicates that the error committed in the calculation of nth concentration at time $t + h_n$ is given by the $Q + 1$ term in the Taylor series shown in Equation (77):

$$O_n(h_n) = \frac{h_n^{Q+1}}{(Q+1)!} c_n^{(Q+1)} \tag{77}$$

It is desirable that the relative error is below a given value:

$$O_n(h_n) \leq \varepsilon c_n^{(0)} \tag{78}$$

The combination of Equations (77) and (78) affords the desired result:

$$h_n \leq \left\{ \varepsilon (Q+1)! \ \frac{c_n^{(0)}}{c_n^{(Q+1)}} \right\}^{\frac{1}{Q+1}} \tag{79}$$

Finally, as each species requires its own step, convergence is ensured by taking the smallest integration step as given in Equation (80):

$$h = \min(h_n) \tag{80}$$

This method has been shown to be better than other explicit methods for accurately integrating nonstiff and moderately stiff kinetic models,[11] and it can be modified[13] to cope with stiff systems.

The stiffness of a kinetic problem arises from mechanisms where steps coexist with very different rates. This drastically slows down the numerical integration process, because the integration step depends inversely on the rate constants. Thus, the existence of fast processes implies a small step slowing down the whole integration procedure. Stiffness is a serious problem because the nonlinear least-squares (NLS) minimization routines could need the o.d.e. system to be integrated several thousands of times, and during the minimization process there is a high probability of finding a stiff area. If such an event occurs and a nonstiff integration method is being used, the minimization procedure will suddenly slow down, taking much more time to complete a minimization cycle.

For first-order o.d.e. systems, stiffness can be solved by a change of variable that can decouple the o.d.e. systems. Then each o.d.e. can be integrated separately. This is the strategy followed by the methods shown in section 2.6.1.1. However, the situation is rather more complicated for nonlinear o.d.e. systems where this procedure is not operative.

It has been well established that the integration routines able to solve stiff problems should be based on implicit or semi-implicit numerical methods.[14] That means that the knowledge of the solution of the o.d.e. at time $t + h$ is obtained by solving an algebraic equation system, whose unknown variables are, precisely, the o.d.e. solutions. The procedure presents obvious advantages for stiff systems that can be solved in a short space of time, but implicit methods are much more complex and have disadvantages when they are applied to nonstiff problems due to the time employed in solving the algebraic system. In such cases, a waste of calculation time and a relatively poor accuracy in the numerical solution could be observed. There are several types of algorithms available to solve stiff o.d.e. systems,[4] including the *Bader–Deuflhard* algorithm and the *Gear* method. Also an *implicit Taylor* method can be developed to obtain precise solutions for nonlinear stiff systems.

3 ERROR ANALYSIS IN CHEMICAL KINETIC METHODS

The study of sources of error and their effects on results is an essential part of data treatment. Errors can be divided into random and systematic. In this section, random errors are those inherent to the measurement process, affecting the results in an unpredictable way, and they can be averaged away using enough data. Random errors can be quantified by the standard deviation of the results.

However, systematic errors arise mainly from uncontrolled changes in the factors that modify the reaction rate, and from the simplifications made in the mathematical models. This kind of error will not average away using any amount of data. Sources of systematic errors must be carefully controlled because it is difficult to detect this kind of error.

3.1 Factors Influencing the Reaction Rate

The reaction rate, and therefore most of the kinetic methods, are strongly dependent on several factors such as temperature, reactant concentration, ionic strength, and solvent nature. The reaction rate can be adjusted to a practical range controlling these factors. However, variations out of control can lead to important systematic errors.

The Arrhenius equation shown in Equation (81) is an empirical relationship describing the dependence of the rate constant, k, on the absolute temperature, T:

$$k = A\mathrm{e}^{-\frac{E_a}{RT}} \tag{81}$$

where E_a is the activation energy, with dimensions of energy per mole, A is the pre-exponential factor having the units of k, and R is the universal gas constant ($8.31\,\mathrm{J\,K^{-1}\,mol^{-1}}$).

For many homogeneous reactions, the activation energy ranges between 50 and $100\,\mathrm{kJ\,mol^{-1}}$, and thus, the rate constant increases by a factor of 2–4 per 10 °C rise in room temperature. For instance, a temperature variation of 0.1 °C produces a 1% change in the rate constant at room temperature for a reaction having an activation energy of $75\,\mathrm{kJ\,mol^{-1}}$.

The reactant concentration has a proportional effect on the reaction rate. For a pseudo-first-order reaction, the kinetic constant is directly proportional to the reagent concentration as shown in Equation (82):

$$k_1 = k_2[\mathrm{R}] \tag{82}$$

Therefore, the concentration errors are correlated with those of the reaction rate.

The simplest treatment to quantify the influence of ionic strength on the reaction rate is based on the transition-state theory. This theory considers that a fast pre-equilibrium step is established among reactants and the activated complex as shown in Equation (83):

$$\mathrm{A} + \mathrm{R} \Longleftrightarrow \mathrm{AR}^{\#} \longrightarrow \mathrm{P} \tag{83}$$

The reaction rate of this system is given by Equation (84):

$$v = k[\mathrm{AR}^{\#}] = kK^{\#}\frac{\gamma_{\mathrm{A}}\gamma_{\mathrm{R}}}{\gamma_{\mathrm{AR}^{\#}}}[\mathrm{A}][\mathrm{R}] \tag{84}$$

where $K^{\#}$ is the thermodynamic equilibrium constant for the formation of the activated complex, and γ_{A}, γ_{R}, and $\gamma_{\mathrm{AR}^{\#}}$ the activity coefficients of the species. The activity coefficient can be related to the ionic strength, I, by the extended Debye–Hückel equation:

$$\log \gamma_i = -\frac{AZ_i^2\sqrt{I}}{1 + a_iB\sqrt{I}} + b_iI \tag{85}$$

where Z_i is the charge of the ith species, and b_i becomes significant for neutral species or high ionic strengths. Considering that the product a_iB is usually taken as unity, and that $Z_{\mathrm{AR}^{\#}} = Z_{\mathrm{A}} + Z_{\mathrm{R}}$, substitution of the activity coefficients in Equation (84) by the expression in Equation (85) leads to Equation (86):

$$\log k = \log k_0 + \frac{2AZ_{\mathrm{A}}Z_{\mathrm{R}}\sqrt{I}}{1 + \sqrt{I}} + bI \tag{86}$$

where k_0 is the rate constant at zero ionic strength. This expression indicates that changes in the ionic strength scarcely affect the rate of reactions taking place between neutral species, or an ion and a neutral molecule, since the product $Z_{\mathrm{A}}Z_{\mathrm{R}}$ cancels out and b is usually small. In addition, the rate of a second-order reaction between equally charged ions should increase as the ionic strength increases, whereas the rate of a reaction between oppositely charged ions should decrease. For instance, an increase of ionic strength from 0.01 to 0.011 produces a 1% change in the rate constant for a reaction taking place between monocharged ions. The effect is considerably higher for low ionic strengths; thus, a change of 7% is observed when the ionic strength increases from 0 to 0.001.

The rate of a chemical reaction can be modified by the nature of the solvent. In spite of its complexity, Equation (87) shows that a linear variation in $\ln(k)$ with the inverse of the dielectric constant of the solvent, $1/\varepsilon$, is observed with reasonable accuracy:[15]

$$\ln k = \ln k_0 + m\frac{1}{\varepsilon} \tag{87}$$

where k_0 is the rate constant in a medium of infinite ε, and m a constant depending on temperature, charge, ionic radius, and the dipole moments of reactants and activated complex. Equation (87) suggests that an increase in solvent polarity favors those reactions where an activated polar complex is formed, and slows down those reactions occurring through a nonpolar activated complex. Hence, the solvent has a drastic influence on the rate constant. For instance, the reaction rate of triethylamine with ethyl iodide suffers a 4200% increase when changing from benzene to acetonitrile.[1]

Finally, it could be of interest to compare equilibrium and rate measurements. According to the relationship in Equation (88) relating the equilibrium constant to the forward and reverse rate constants:

$$K = \frac{k_{\mathrm{f}}}{k_{\mathrm{r}}} \tag{88}$$

the effect of the experimental conditions might be lower on the equilibrium constant than on the rate constants due to the compensation of errors in Equation (88). Therefore, in order to obtain good results with kinetic-based methods, the experimental conditions must be controlled more carefully than is usual in the equilibrium-based methods.

3.2 Methods Based on First-order Reactions for the Determination of One Species

From a practical analytical viewpoint, most kinetic methods for the determination of a single species are carried out in pseudo-first-order conditions. Table 4 shows the most frequently used experimental conditions in uncatalyzed, catalyzed, and enzymatic methods in order to assure pseudo-first-order kinetics. Depending on the original mechanism, the pseudo-first-order rate constant will be a function of several experimental factors, such as the reagent concentration. The rate law in Equation (89) summarizes the different cases in a general form:

$$v = \frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = -(k_{\mathrm{c}}[\mathrm{C}]_0 + k_{\mathrm{u}}[\mathrm{R}]_0)[\mathrm{A}] = -k[\mathrm{A}] \tag{89}$$

where k_{c} and k_{u} are rate constants for the catalyzed and uncatalyzed reactions, respectively. On integration Equation (89) leads to Equation (90):

$$[\mathrm{A}] = [\mathrm{A}]_0\mathrm{e}^{-kt} = [\mathrm{A}]_0\mathrm{e}^{-(k_{\mathrm{c}}[\mathrm{C}]_0 + k_{\mathrm{u}}[\mathrm{R}]_0)t} \tag{90}$$

Now, after taking into account the linear dependency of the response upon concentration given by Equation (31)

Table 4 Mechanisms frequently used in kinetic analysis and common simplifications

Reaction type	Mechanism	Original o.d.e.	Simplification	Applied o.d.e.
Uncatalyzed	$A + R \xrightarrow{k_2} P$	$\frac{d[P]}{dt} = k_2[R][A]$	$[R] \gg [A]; [R] \approx [R]_0$	$v = k_2[R]_0[A]$
Catalyzed	$S + R \xrightarrow{k_u} P$ $R + C \Longleftrightarrow I$ $I + S \xrightarrow{k_c} P + C$	$\frac{d[P]}{dt} = k_c[I][S] + k_u[R][S]$	$[R] \gg [S]$ $[R] \approx [R]_0$ $[I] \approx [C]_0$	$v = (k_c[C]_0 + k_u[R]_0)[S]$
Enzymatic	$S + E \Longleftrightarrow ES \xrightarrow{k_2} P + E$	$\frac{d[P]}{dt} = \frac{k_2[E]_0[S]}{K_m + [S]}$	$[S] \ll K_m$ $[S] \gg K_m$	$v = \frac{k_2}{K_m}[E]_0[S]$ $v = k_2[E]_0$

for first-order reactions, we can write Equation (91):

$$S = S_\infty - s[A]_0 e^{-kt} = S_0 + s[A]_0(1 - e^{-kt}) \quad (91)$$

where s is a linear combination of the proportionality constants of the different species with signal. For instance, $s = s_P - s_A - s_R$ for the reaction $A + R \longrightarrow P$.

Data treatment can be done by applying several strategies.[16] The different data used by each treatment appear in the kinetic profile shown in Figure 3. In the fixed time method the signal values S_1 and S_2 are measured at constant times t_1 and t_2. On the contrary, in the variable time method the time values are measured, whereas the signal values remain constant. In the initial rate method, initial rate, v_0, is measured as the slope, near the origin, of the kinetic curve. Finally, the predictive kinetic (PK) method[17] is based on the extrapolation of a fitted kinetic profile to the predicted final signal, S_∞.

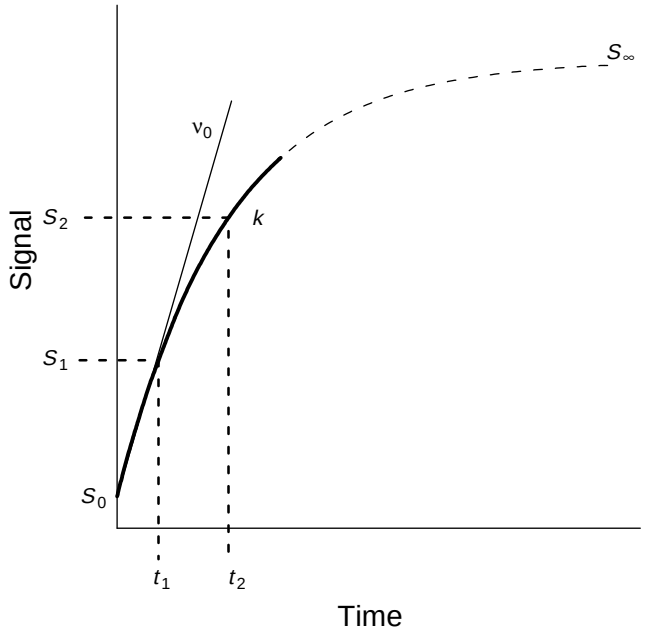

Figure 3 Signal–time profile for a first-order reaction. Some data treatment used in kinetic methods are shown.

3.2.1 *Implicit Systematic Errors of the Data Treatment*

The fixed and variable time methods are based on the change of the signal during a time interval Δt. For a first-order reaction we obtain Equation (92):

$$\frac{\Delta S}{\Delta t} = s[A]_0 \frac{e^{-kt_1} - e^{-kt_2}}{t_2 - t_1} = s[A]_0 e^{-kt_1} \frac{1 - e^{-k\Delta t}}{t_2 - t_1} \quad (92)$$

where $t_2 = t_1 + \Delta t$. The expansion of the exponential function in a Maclaurin series, Equation (53), yields Equation (93):

$$\frac{\Delta S}{\Delta t} = sk[A]_0 e^{-kt_1} \sum_{i=1}^{\infty} (-1)^i \frac{(k\Delta t)^{i-1}}{i!} = k[A]_0\Omega$$
$$= (k_c[C]_0 + k_u[R]_0)[A]_0\Omega \quad (93)$$

where Ω is a calibration factor depending on the products kt_1 and kt_2.

The determination process can be divided into the calibration and the sample measurement steps. The calibration step allows us to calculate the value of Ω from standards of known concentration. This value will be used in the determination of analyte concentration in the sample. The result will fit to the real concentration, when the factor Ω remains constant between the calibration and the sample measurements. Taking t_1 the closest to zero reduces the error and increases the sensitivity. On the other hand, the more similar the calibration and sample concentrations, the closer the Ω values. The data treatment also has an important effect. When the initial concentration of the limiting reagent, A, is determined, the rate constant remains invariable between the calibration procedure and the sample determination. Thus, Ω will keep constant for the fixed time method, but not for the variable time method, and a systematic error will be produced. This situation is also applicable to

the determination of substrate concentration in enzyme-catalyzed reactions which exhibit a similar pseudo-first-order kinetic behavior, as shown in Table 4.

When a catalyst is determined, the rate constant will change between the standards and the sample, and thus Ω will vary for the fixed time method, introducing a systematic error. However, the product kt will remain constant for the variable time method, since this quantity is related to the signal as Equation (94) shows:

$$kt = -\ln\left(\frac{S_\infty - S}{S_\infty - S_0}\right) \tag{94}$$

where S, S_∞, and S_0 are constants, making Ω invariable. The same situation is found for uncatalyzed reactions when the species to be determined is the reagent in excess, R, but now a small systematic error is committed when R affects the signal value.

Another approach, shown in Equation (95), takes the logarithm of the signal ratio:

$$\frac{1}{\Delta t}\ln\frac{S_\infty - S_1}{S_\infty - S_2} = k = k_c[C]_0 + k_u[R]_0 \tag{95}$$

In this case, the fixed time method can be used without systematic error when S_∞ is known. When it is unknown, and the logarithm of the signal ratio is used, Equation (96) is obtained:

$$\ln\frac{S_1}{S_2} = k\Delta t + \ln\frac{S_\infty e^{kt_1} - S_\infty + S_0}{S_\infty e^{kt_2} - S_\infty + S_0} \tag{96}$$

Now, a systematic error is expected, but smaller than that obtained using ΔS, because of the quotient error compensation in Equation (96).

Figure 4 shows the performance of the fixed and variable time methods. According to the above conclusions, the variable time method will give the best results when a catalyst is determined, whereas the fixed time method will be superior for the determination of the limiting reagent for uncatalyzed reactions.

Ingle and Crouch[18] have also shown that the variable time methodology has a wider range of catalyst concentration amenable to determination, in spite of the difficulty involved in developing a general treatment, owing to the large variety of possible mechanisms encountered. Furthermore, the authors concluded, on the basis of a study of the theoretical and experimental factors influencing the accuracy of reaction rate measurements, that the variable time approach is superior in most cases where nonlinear response curves are obtained.

A different approach can be applied by fitting the whole kinetic curve to the integrated equation to obtain k and $[A]_0$. When the rate constant is the analytical parameter, the method of tangents can be used. In this method, the linear fit of Equation (35) yields, as the slope, the rate constant, which is linearly related to the catalyst concentration. When the final signal is unknown, the Guggenheim or KMS methods should be used as illustrated in section 2.5.2.1. Finally, since the method of tangents performs more measurements to obtain the rate constant, its application is expected to result in better precision than the variable or fixed time methods. However, the method of tangents is highly affected by model nonlinearities and fails when the kinetics is complex, and the integration of the rate equations is difficult or impossible.

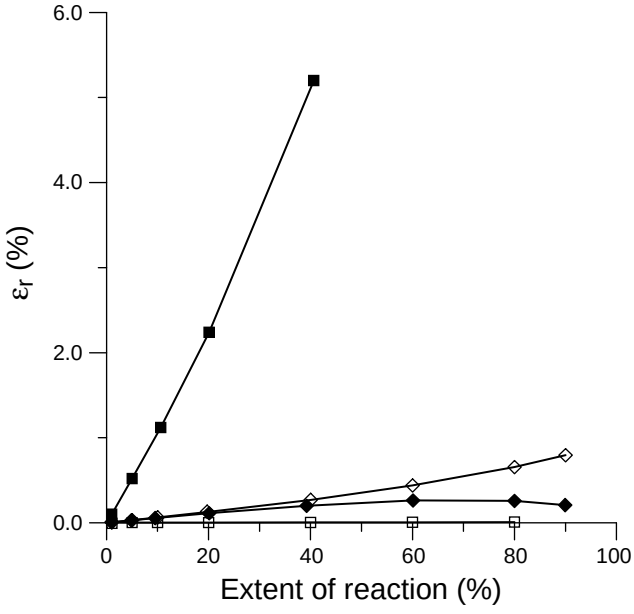

Figure 4 Variation of the mean relative error of the data used in the determination of two samples of 2.5 and 3.5 μM, with increasing extent of reaction. A calibration of five points (1, 2, 3, 4 and 5 μM) and the following conditions were applied: ( ) fixed time and (◇) variable time, for the uncatalyzed determination of a substance using a pseudo-first-order reaction with $[R]_0 = 50\,\mu M$ and $k = 0.1\,s^{-1}\,\mu M^{-1}$; (■) fixed time and ( ), variable time for the determination of a catalyst with $k_c = 1.0$ and $k_u = 0\,s^{-1}\,\mu M^{-1}$.

When the initial concentration of the limiting reagent is to be determined, the PK method can be applied. The method is based on the extrapolation of a fitted kinetic profile to predict the signal at the end of the reaction. The advantage of this approach is the bypassing of errors due to experimental changes in k that go with changes in temperature, reagent concentration, or ionic strength. Two approaches can be used, the consecutive linear determination of k and $[A]_0$, or the simultaneous nonlinear determination of k, $[A]_0$, and S_0, based on Equation (91). Baeza-Baeza et al.[19] evaluated several methods for the determination of an analyte following pseudo-first-order kinetics, and observed that the PK method gave the best results, specially when the average

rate constant of all kinetic runs is used instead of the individual ones. This fact indicates that random errors in the constant have an important effect on the extrapolated final signal; this effect decreases when the extent of the reaction increases, that is, as the extrapolation risk decreases.

The rate methods are differential approaches based on the relation of the reaction rate with concentration as shown in Equation (97):

$$v = \frac{\mathrm{d}S}{\mathrm{d}t} = sk[\mathrm{A}]_0 \mathrm{e}^{-kt} \tag{97}$$

The initial rate method is based on the measurement of the slope of the kinetic curve at initial times, as is shown in Equation (54). As stated by Equation (56), the smallest possible error is obtained when the measurement is made as close to the origin as the mixing time and induction period permit. It is recommended that no more than 2% reactant conversion be allowed to occur. In addition, a reduction in the systematic error can be done if the initial rate is obtained by fitting the initial values of the signal to an orthogonal polynomial series.

On the other hand, an error compensation occurs between calibration and determination in the global process. Therefore, bearing in mind that Equation (54) is a simplified version of Equation (93), similar conclusions to those previously outlined will be obtained when the model follows first-order kinetics near the origin. However, for catalyzed reactions, including enzymatic ones, the initial kinetics could not follow a first-order model owing to the induction period necessary to reach the stationary state. The advantages generally attributed to initial rate measurements include the minimal complications from side reactions, and the increased sensitivity over other rate measurements, since the exponential value in Equation (97) is unity when $t = 0$. Finally, it should be realized that the success of the initial rate measurements depends greatly on the availability of sufficiently sensitive and reproducible monitoring of small signal changes.

A special case is found in the determination of the enzyme activity because, as shown in Table 4, the reaction follows a zero-order law under saturation conditions. In order to study the error of this treatment, an integration of the Michaelis–Menten equation must be done.[20] A good approximation is to suppose that $[\mathrm{S}] \cong [\mathrm{S}]_0 - [\mathrm{P}]$ when $[\mathrm{S}]_0 \gg [\mathrm{E}]_0$; now, considering that the signal is produced by the product of the reaction, $S = s_\mathrm{P}[\mathrm{P}]$, the integrated equation has the form shown in Equation (98):

$$S = s_\mathrm{P} k_2 [\mathrm{E}]_0 t - s_\mathrm{P} K_\mathrm{m} \ln\left(\frac{S_\infty}{S_\infty - S}\right. \tag{98}$$

This expression allows signal and time variations to be related as shown in Equation (99):

$$\frac{\Delta S}{\Delta t} = s_\mathrm{P} k_2 [\mathrm{E}]_0 - s_\mathrm{P} K_\mathrm{m} \frac{\ln[(S_\infty - S_1)/(S_\infty - S_2)]}{\Delta t} \tag{99}$$

The final signal, S_∞, is expected to be large compared to S_1 and S_2 due to the saturation conditions, and, thus, the logarithm can be expanded by means of a Taylor series to Equation (100):

$$\ln\left(\frac{S_\infty - S_1}{S_\infty - S_2}\right. \quad \cong \frac{S_2 - S_1}{S_\infty - S_2} = \frac{\Delta S}{S_\infty - S_2} \tag{100}$$

leading to Equation (101):

$$\frac{\Delta S}{\Delta t} = \frac{s_\mathrm{P} k_2 (S_\infty - S_2)}{S_\infty - S_2 + s_\mathrm{P} K_\mathrm{m}} [\mathrm{E}]_0 = m[\mathrm{E}]_0 \tag{101}$$

Thus, if the variable time method is applied, the slope, m, will remain constant since S_2 has a constant value. On the other hand, S_2 varies in the fixed time method, and m is not constant, resulting in a systematic error. Figure 5 shows the application of both methods to the enzymatic activity determination by using an increasing time interval. The variable time method gives better results when the extent of the reaction is high enough to ensure the validity of the stationary state and the Michaelis–Menten equation. However, this hypothesis is not valid at the beginning of

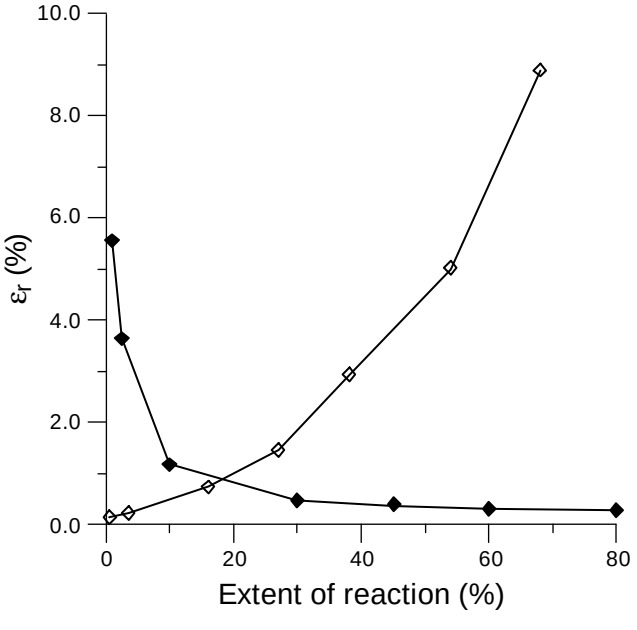

Figure 5 Variation of the mean relative error of the data used with increasing extent of reaction in the determination of two enzyme concentrations (2.5 and 3.5 nM), applying the fixed time (◇) and the variable time ( ) approach. A calibration of five points (1, 2, 3, 4 and 5 nM) was used. The kinetic constants of the Michaelis–Menten mechanism were $k_1 = 50$, $k_{-1} = 20$, $k_2 = 200$ and $[\mathrm{S}]_0 = 1\,\mu\mathrm{M}$.

the reaction. The fixed time method improves itself as the time interval decreases, approaching the data treatment of the initial rate method, whereas the variable time method gives worse and worse results because it evaluates the initial rate at different times. It must be realized that the reaction rate increases almost linearly with time at the beginning of reaction, as shown in Equation (102):

$$v_{t\to 0} = k_2[\mathrm{ES}]_{t\to 0} \cong \frac{k_2 k_1}{k_{-1}+k_2}[1-\mathrm{e}^{-(k_{-1}+k_2)t}][\mathrm{S}]_0[\mathrm{E}]_0 \cong k_2 k_1 t[\mathrm{S}]_0[\mathrm{E}]_0 \tag{102}$$

where $[\mathrm{ES}]_{t\to 0}$ has been obtained by integration of Equation (10) supposing that S and E concentrations remain constant and equal to their initial values. Thus, the fixed time method will yield better results near $t = 0$.

3.2.2 *Propagation of Systematic Errors*

The study of the sources of systematic errors can lead to an optimum experimental design minimizing the error propagation. Systematic errors can arise from changes in the rate constant between calibration and sample determination, owing to variations in the factors that affect the reaction rate.

When the fixed-time method is applied to an uncatalyzed reaction, the linear calibration model will be as shown in Equation (103):

$$\Delta S = m[\mathrm{A}]_0 + n \tag{103}$$

where m and n are the calibration parameters. A systematic error in the rate constant, ε_k, will cause the systematic error given in Equation (104) for the initial concentration of the analyte:

$$\varepsilon_\mathrm{A} = \frac{\partial \Delta S}{\partial k}\frac{\varepsilon_k}{m} = s[\mathrm{A}]_0(t_2\mathrm{e}^{-kt_2} - t_1\mathrm{e}^{-kt_1})\frac{\varepsilon_k}{m} \tag{104}$$

The error will be zero if the times are chosen according to Equation (105):

$$t_1\mathrm{e}^{-kt_1} = t_2\mathrm{e}^{-kt_2} \tag{105}$$

The function $t\mathrm{e}^{-kt}$ has a bell shape with a maximum at $t = 1/k$. As time moves away from this value, the function decreases. As a consequence, the optimal approach consists of choosing the biggest t_2 possible to lower the $t\mathrm{e}^{-kt}$ function, and resolving Equation (105) for t_1.

When a catalyzed reaction is treated by the variable time approach, the analytical function is $1/\Delta t$, and the error is given by Equation (106):

$$\varepsilon_\mathrm{C} = \frac{\partial(1/\Delta t)}{\partial k_\mathrm{c}}\frac{\varepsilon_{k_\mathrm{c}}}{m} = \frac{[\mathrm{C}]_0}{\ln[(S_\infty - S_1)/(S_\infty - S_2)]}\frac{\varepsilon_{k_\mathrm{c}}}{m} \tag{106}$$

In this case, the error decreases when ΔS increases, since the logarithm also increases.

Regarding the initial rate method, taking into account that in this case $m = sk$, its application leads to a constant error as shown in Equation (107):

$$\varepsilon_\mathrm{A} = \frac{\partial(v_0)}{\partial k}\frac{\varepsilon_k}{m} = s[\mathrm{A}]_0\frac{\varepsilon_k}{m} = [\mathrm{A}]_0\frac{\varepsilon_k}{k} \tag{107}$$

Thus, the relative error in k propagates invariable to the relative concentration error. Holler et al.[21] studied the propagation of errors in the instantaneous rate measurement. From Equation (97) is obtained Equation (108):

$$\varepsilon_\mathrm{A} = \frac{\partial(v)}{\partial k}\frac{\varepsilon_k}{m} = s[\mathrm{A}]_0(1-kt)\mathrm{e}^{-kt}\frac{\varepsilon_k}{m} = [\mathrm{A}]_0(1-kt)\frac{\varepsilon_k}{k} \tag{108}$$

This expression shows that the reaction rate values are independent of changes in parameters affecting the value of the rate coefficient at a time equal to $1/k$. Besides, when the initial rate is measured, time equals zero and the propagated error is one of the largest.

The precision in the determination of an analyte by using the PK extrapolation method depends significantly on the extent of the reaction concerned. Let us suppose that an extrapolation is carried out using only one data point as in Equation (109):

$$[\mathrm{A}]_0 = \frac{S - S_0}{s(1-\mathrm{e}^{-kt})} \tag{109}$$

From Equation (109) it is clear that the error in $[\mathrm{A}]_0$ due to an error committed in k is given by Equation (110):

$$\varepsilon_\mathrm{A} = \frac{\partial[\mathrm{A}]_0}{\partial k}\varepsilon_k = -\frac{[\mathrm{A}]_0 t\mathrm{e}^{-kt}}{1-\mathrm{e}^{-kt}}\varepsilon_k \tag{110}$$

Now, taking into account that the extent of reaction is $\xi = 1 - \mathrm{e}^{-kt}$, the relative error in $[\mathrm{A}]_0$ will be as shown in Equation (111):

$$\varepsilon_{r_\mathrm{A}} = \frac{1-\xi}{\xi}[\ln(1-\xi)]\varepsilon_{r_k} \tag{111}$$

Carr[22] studied this effect showing that the relative precision improves monotonically with increasing ξ, and that the relative uncertainty in $[\mathrm{A}]_0$ tends to the relative uncertainty in k at small values of ξ; finally, when the reaction approaches completion ($\xi \to 1$), the uncertainty in the measurement of k causes no appreciable systematic error in the estimated concentration.

Another kind of systematic error arises from the presence of an interferent species, B, in the sample. In this case, the variation of the signal in the sample for an uncatalyzed reaction is given by Equation (112):

$$\Delta S = s_\mathrm{A}[\mathrm{A}]_0(\mathrm{e}^{-k_\mathrm{A}t_1} - \mathrm{e}^{-k_\mathrm{A}t_2}) + s_\mathrm{B}[\mathrm{B}]_0(\mathrm{e}^{-k_\mathrm{B}t_1} - \mathrm{e}^{-k_\mathrm{B}t_2}) \tag{112}$$

The relative error in the determination of A, supposing the calibration standards are free of B, and applying the fixed time method, is given by Equation (113):

$$\varepsilon_{r_A} = \frac{s_B}{s_A}\frac{[B]_0}{[A]_0}\frac{e^{-k_B t_1} - e^{-k_B t_2}}{e^{-k_A t_1} - e^{-k_A t_2}} \tag{113}$$

This expression shows that the error increases when k_B increases and decreases when t_1 and Δt decrease. Obviously, error decreases with s_B and $[B]_0$.

The error of the variable time method for a catalyzed reaction with interference from another catalyst is given by Equation (114):

$$\varepsilon_{r_C} = \frac{[B]_0}{[C]_0}\frac{k_C}{k_B} \tag{114}$$

where the error depends only on the reaction rates and the initial concentration ratio. If the rate method has been used, the error will be as shown in Equation (115):

$$\varepsilon_{r_A} = \frac{s_B}{s_A}\frac{k_B}{k_S}\frac{[B]_0}{[A]_0}\frac{e^{-k_B t}}{e^{-k_A t}} \tag{115}$$

which is a minimum at $t = 0$, that is, for the initial rate method.

3.2.3 Propagation of Random Errors

The error propagation theory must be applied to obtain the parameter uncertainty due to random errors. For small independent errors in the variables, the propagation error rule is given by Equation (116):

$$\sigma_y^2 = \sum \left(\frac{\partial y}{\partial x_i}\right)^2 \sigma_{x_i}^2 \tag{116}$$

When an uncatalyzed reaction is used to determine the concentration of a substance by applying the fixed time method with, for simplicity, a unique calibration point, the result is given by Equation (117):

$$[A]_{sm} = \frac{\Delta S_{sm}}{\Delta S_{cl}}[A]_{cl} \tag{117}$$

where the subscript "sm" refers to the sample and "cl" to the calibration measurements. The mean square relative error of the concentration is given by Equation (118):

$$\frac{\sigma_A^2}{[A]_{sm}^2} = \frac{\sigma_{\Delta S_{sm}}^2}{\Delta S_{sm}^2} + \frac{\sigma_{\Delta S_{cl}}^2}{\Delta S_{cl}^2} \tag{118}$$

This expression indicates that the relative square error in the evaluated concentration is the addition of the relative square error of the measured signal differences. The variance of the measured signal is given by Equation (119):

$$\sigma_{S_i}^2 = \sigma_S^2 + \left(\frac{\partial S_i}{\partial t}\right)^2 \sigma_t^2 \tag{119}$$

The signal is affected by its own uncertainty, σ_S^2, and by the uncertainty introduced for the lack of precision in time.

Now, supposing homocedastic data, Equation (120) is obtained:

$$\frac{\sigma_{\Delta S}^2}{\Delta S^2} = \frac{\sigma_{S_1}^2 + \sigma_{S_2}^2}{\Delta S^2} = 2\frac{\sigma_S^2}{\Delta S^2} + \frac{(S_\infty - S_2)^2 + (S_\infty - S_1)^2}{\Delta S^2}k^2\sigma_t^2 \tag{120}$$

The uncertainty decreases as ΔS increases, reaching a minimum when S_2 tends to S_∞, and S_1 tends to the smallest signal value. On the other hand, the error increases as the reaction rate increases, owing to the higher uncertainty in the signal measurement produced by the time uncertainty.

When the variable time method is used to determine the concentration of catalyst C, the relative square error in the concentration is a function of the time interval measured as shown in Equation (121):

$$\frac{\sigma_C^2}{[C]_{sm}^2} = \frac{\sigma_{\Delta t_{sm}}^2}{\Delta t_{sm}^2} + \frac{\sigma_{\Delta t_{cl}}^2}{\Delta t_{cl}^2} \tag{121}$$

and the relative square error of the time difference is given by Equation (122):

$$\frac{\sigma_{\Delta t}^2}{\Delta t^2} = 2\frac{\sigma_t^2}{\Delta t^2} + \frac{\sigma_S^2}{k^2\Delta t^2}\quad\frac{1}{(S_\infty - S_1)^2} + \frac{1}{(S_\infty - S_2)^2} \tag{122}$$

In this case, the error depends on two opposite factors, the reduction of uncertainty when Δt increases, and the lack of precision in the time measurement when S_2 nears S_∞. A graphical inspection of Equation (122) shows that precision increases as S_∞ increases, S_1 decreases, and S_2 takes a value around 0.7 times S_∞, at which Equation (122) has a minimum depending on σ_t^2, σ_S^2, and k values.

3.3 Multicomponent Methods

Kinetic methods based on the different rates at which the components of a mixture react with a common reagent can be used to determine several analytes simultaneously.[23] However, kinetic-based multicomponent determinations have found fewer applications than those devoted to the determination of only one species. This reflects the fact that conditions for determining more than one species simultaneously, and suitable multicomponent reactions, are difficult to find.

3.3.1 Differential Methods

Several strategies have been employed[24] in the simultaneous determination of two or more species. For two analytes, one of the most popular approaches has been the

logarithmic extrapolation method, because of its simplicity and its greater accuracy than other methods when the ratio of the rate constants is large. This method assumes that the signal is due to the slowest-reacting component, A, after reacting to completion the faster component, B. Equation (123) gives:

$$\ln(S_\infty - S) \cong -k_A t + \ln(s[A]_0) \tag{123}$$

Thus, fitting the final data points, the concentration of the slowest component could be obtained. The faster component concentration can be evaluated from the difference between the final and the initial signal. Better results are obtained if weights, $w_i = (S_\infty - S_i)^2$, are used[2] to compensate the logarithmic transformation. The systematic error of the method depends on the ratio of the rate constants, the extent of the reactions, and the uncertainty in the final signal value. Figure 6 shows that the error decreases as the extent of the faster reaction increases and the extent of the slower reaction decreases. On the other hand, the systematic errors in the final signal can produce dramatic effects on the results. From Equation (123) and supposing that the value of k_A is known, the error yields by the extrapolation from one data point, S, will be given by Equation (124):

$$\frac{\varepsilon_A}{[A]_0} = \frac{\varepsilon_{S_\infty}}{S_\infty - S} \tag{124}$$

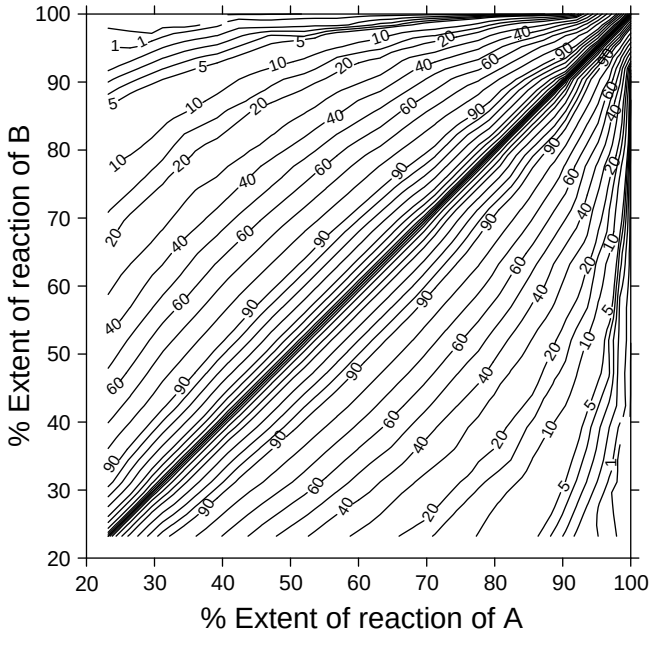

Figure 6 Relative error in the determination of the slower species of a binary mixture using the logarithmic extrapolation method, as a function of the extents of reaction at the first experimental point taken for the extrapolation.

The overall process leads to an error that will be a function of a combination of terms like Equation (124) for all the points used. For this reason, to reduce the effect of the propagation of a systematic error in S_∞, the points must be taken as near the origin as the reaction rates ratio permits.

Another widely used approach is the method of proportional equations, which involves the resolution of a system of simultaneous equations to obtain analyte concentrations. For two substances, Equation (125) relates the signal value to the analyte concentration:

$$\begin{aligned} S - S_0 &= s_A[A]_0(1 - e^{-k_A t}) + s_B[B]_0(1 - e^{-k_B t}) \\ &= \Psi_A[A]_0 + \Psi_B[B]_0 \end{aligned} \tag{125}$$

The above expression indicates that the analyte concentrations can be obtained by solving a system of linear equations when the signal is measured at two different times. This method does not require a prior knowledge of S_∞, but the proportionality factors, Ψ, must be carefully evaluated at each time from pure solutions of A and B separately. Although the knowledge of s and k values should permit theoretical calculations of Ψ, the empirical evaluation is preferred to minimize the effect of experimental variables. The solutions of the system for two species is given by Equation (126):

$$[A]_0 = \frac{\Psi_{2B}S_1 - \Psi_{1B}S_2}{D} \quad [B]_0 = \frac{\Psi_{1A}S_2 - \Psi_{2A}S_1}{D} \tag{126}$$

where the S values have been corrected with regard to the initial signal, and $D = \Psi_{1A}\Psi_{2B} - \Psi_{2A}\Psi_{1B}$ is the determinant of the $\boldsymbol{\Psi}$ matrix.

The study of the effect of experimental errors on the evaluated concentrations will be based on the hypothesis that errors affect the measured signals due to the uncertainties in time, signal measurement, experimental conditions, and so on. The error in calibration will affect the proportionality constants; for instance, the error in Ψ_{1A} is given by Equation (127):

$$\Psi_{1A} = \frac{S^0_{1A} + \varepsilon_{1A}}{[A]_{cl}} = \Psi^0_{1A} + \frac{\varepsilon_{1A}}{[A]_{cl}} \tag{127}$$

where $[A]_{cl}$ is the concentration of A species in the calibration. By introducing all the measurement errors in Equation (126), expressions for the error concentrations are given by Equations (128) and (129):

$$\varepsilon_A \cong \frac{\begin{aligned}&\Psi_{1B}\left(\frac{[B]_{sm}}{[B]_{cl}}\varepsilon_{2B} + \frac{[A]_{sm}}{[A]_{cl}}\varepsilon_{2A} - \varepsilon_2\right) \\ &-\Psi_{2B}\left(\frac{[B]_{sm}}{[B]_{cl}}\varepsilon_{1B} + \frac{[A]_{sm}}{[A]_{cl}}\varepsilon_{1A} - \varepsilon_1\right)\end{aligned}}{D} \tag{128}$$

$$\varepsilon_B \cong \frac{\Psi_{2A}\left(\frac{[B]_{sm}}{[B]_{cl}}\varepsilon_{1B} + \frac{[A]_{sm}}{[A]_{cl}}\varepsilon_{1A} - \varepsilon_1\right) - \Psi_{1A}\left(\frac{[B]_{sm}}{[B]_{cl}}\varepsilon_{2B} + \frac{[A]_{sm}}{[A]_{cl}}\varepsilon_{2A} - \varepsilon_2\right)}{D} \quad (129)$$

These equations indicate that the concentration errors depend on: the value of the uncertainty in the measurements; the value of the proportionality constants, and thus on the experimental design; and the relationship between the concentrations in the calibration and the sample. Besides, the systematic errors can be partially compensated when they are in the same kinetic curve or reinforced when they are at the same time in the calibration curves. On the other hand, the systematic errors affect each substance in an opposite way, that is, when the error of A is positive the error of B will be negative and vice versa. Finally, errors in the first experimental point will have a larger effect because the proportionality constants of the second calibration time, Ψ_{2A} and Ψ_{2B}, are higher than those of the first calibration time, Ψ_{1A} and Ψ_{1B}.

The optimal experimental design will be the one that makes the determinant D the maximum. Equation (130) gives:

$$D = s_A s_B \left(1 - e^{-k_A t_1}\right)\left(1 - e^{-k_B t_2}\right) - \left(1 - e^{-k_A t_2}\right)\left(1 - e^{-k_B t_1}\right) \quad (130)$$

This determinant is maximum when t_2 is the highest possible, and t_1 is given by Equation (131):

$$t_1 = \frac{\ln[(k_B/k_A)(1 - e^{-k_A t_2})/(1 - e^{-k_B t_2})]}{k_B - k_A} \quad (131)$$

A drawback of proportional equations is that although a fairly large number of data points is obtained, only a small fraction, two in a binary mixture, are used in the computation, thereby limiting precision. On the other hand, although theoretically there is no limitation on the number of components that could be determined, a practical limit only allows the resolution of mixtures of up to four components with acceptable errors, as long as their rate constants are sufficiently different. The high propagation of experimental errors, as shown in Equations (128) and (129), imposes the restriction. However, this method provides better precision and accuracy than other differential methods.[25]

3.3.2 Regression Methods

Experimental data can be condensed by fitting them to a model that depends on adjustable parameters. Generally, these parameters are the rate constants in the calibration step and the analyte concentrations during the determination procedure. An objective function must be chosen to measure the agreement between data and model response with regard to a particular set of parameter values. The fitting problem reduces to a minimization process that searches the set of parameter values that minimize the objective function. Finally, statistical tools are needed to measure the uncertainty of the determined parameters.

3.3.2.1 Least-squares Regression The least-squares fitting is a maximum likelihood estimation of the parameters, provided that the measurement errors are independent and normally distributed.[4] For a single response, the least-squares can be written as in Equation (132):

$$\chi = \sum_{i=1}^{N} w_i (Y_i - S_i)^2 = \sum_{i=1}^{N} w_i r_i^2 \quad (132)$$

where w_i is the weight assigned to the ith point, Y_i the experimental signal, S_i the signal value predicted by the model, and r_i the residual or difference between the experimental and predicted values.

Conventionally, the weight of an experimental point, w_i, has been taken as inversely proportional to its variance, σ_i, for uncorrelated measurements, as given by Equation (133):

$$w_i = \frac{1}{\sigma_i^2} \quad (133)$$

In this way, the more uncertain experimental data are considered less. Another kind of weight arises from the use of a transformed variable,[2] Z, from the raw experimental variable, Y. Equation (134) gives:

$$w_i = \frac{(1/\sigma_i^2)}{(dZ_i/dY_i)^2} \quad (134)$$

For instance, when a logarithmic transformation is performed the weights must be $w_i = (S_i/\sigma_i)^2$. It can be shown that the transformation causes a distortion in the informative structure of the data[26] and that weights given in Equation (134) restore the original structure.

The minimization of the least-squares function can be done using several methods.[27] The Newton–Raphson method is based on the Taylor's series expansion of the gradient, $\mathbf{G}$, as shown by Equation (135):

$$\mathbf{G}(\mathbf{P} + \Delta\mathbf{P}) = \mathbf{G}(\mathbf{P}) + \mathbf{H}(\mathbf{P})\Delta\mathbf{P} + \ldots \quad (135)$$

where $\mathbf{P}$ and $\Delta\mathbf{P}$ are the parameter and the parameter increment vectors, and $\mathbf{H}$ the Hessian matrix. By neglecting the higher-order terms and setting the gradient equal to zero, Equation (135) can be easily solved for the value of $\Delta\mathbf{P}$ that leads to the desired minimum as shown in Equation (136):

$$\Delta\mathbf{P} = -\mathbf{H}(\mathbf{P})^{-1}\mathbf{G}(\mathbf{P}) \quad (136)$$

In general, the higher-order terms in Equation (135) are not negligible, and hence an iterative procedure has to be used. In least-squares regression problems, the elements of gradient and Hessian are given by Equations (137) and (138):

$$g_i = \frac{\partial \chi}{\partial P_i} = -2\sum_{k=1}^{N} w_k \frac{\partial S_k}{\partial P_i} r_k \tag{137}$$

$$h_{ij} = \frac{\partial^2 \chi}{\partial P_i \partial P_j} \cong 2\sum_{k=1}^{N} w_k \frac{\partial S_k}{\partial P_i}\frac{\partial S_k}{\partial P_j} \tag{138}$$

where the second-order derivative terms, in the Hessian element, have been dismissed. Using this approximation in Equation (136) leads to the iterative expression shown in Equation (139):

$$\mathbf{P}_{n+1} = \mathbf{P}_n + (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{r} \tag{139}$$

where $\mathbf{W}$ is the diagonal matrix of weights, $\mathbf{r}$ the residuals vector, $\mathbf{X}$ the Jacobian matrix ($x_{ij} = \partial S_i/\partial P_j$), and $\mathbf{X}^{\mathrm{T}}$ denotes the transpose of the matrix $\mathbf{X}$. This procedure is often called the Gauss–Newton method.

The least-squares function of a lineal model is quadratic and the minimum is reached in a single step. Then, taking the origin as the initial parameters vector, Equation (139) leads to Equation (140):

$$\mathbf{P}^* = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Y} \tag{140}$$

where $\mathbf{P}^*$ is the optimum parameter vector.

An important part of the optimization process is to determine the uncertainty of the estimated parameters.[28] From the linearized model shown in Equation (136), the variance–covariance matrix, $\mathbf{C}$, that contains estimates of the errors in the parameters, is given by Equation (141):

$$\mathbf{C} = \frac{\chi^*}{N-n}\mathbf{H}^{-1} \tag{141}$$

where χ^* is the value of the least-squares objective function at the minimum, N the number of experimental points and n the number of fitted parameters. Thus, the estimated standard deviation in the ith parameter is given by Equation (142):

$$\sigma_{P_i} = \sqrt{C_{ii}} \tag{142}$$

It is important to realize that the Hessian matrix $\mathbf{H}$ is a curvature matrix of the least-squares objective function, and low curvatures around the optimum result in high uncertainties in the estimated parameters. The curvature decreases when the correlation of predictor variables (the columns of the $\mathbf{X}$ matrix) increases. In multicomponent analysis using kinetic data, the correlation could be high because the kinetic signals usually have similar profiles, leading to low curvatures and large uncertainties. Obviously, the problem increases when the rate constants become more similar.

3.3.2.2 Error Sensitivity Analysis in Least-squares Regression It is well known that the accuracy of a determination can be ruined by small systematic errors in the experimental conditions such as reagent concentration, time, temperature, or calibration parameters. To evaluate the influence of uncertainties in experimental variables, an error sensitivity analysis must be performed. The error sensitivity of the parameter P with respect to an error in the variable X can be obtained using Equation (143):[29]

$$\delta_{PX} = \frac{\mathrm{d}P^*}{\mathrm{d}X} = -\frac{h_{PX}}{h_{PP}} \tag{143}$$

where P^* is the optimized parameter, and h_{PX} and h_{PP} are the corresponding elements of the Hessian matrix. This equation gives the estimated error in the determination of P by least-squares regression, when the systematic error in the variable X equals unity. It must be noted that X can be any variable having an influence on the value of the experimental points, and an explicit relationship between X and P is not required. In this way, the relative systematic error in P owing to a relative error in X will given by Equation (144):

$$\varepsilon_{rP} = -\frac{h_{PX}}{h_{PP}}\frac{X}{P}\varepsilon_{r_X} \tag{144}$$

When several experimental variables are affected by errors, the propagation of systematic errors must be applied as in Equation (145):

$$\varepsilon_P = \delta_{PX_1}\varepsilon_{X_1} + \delta_{PX_2}\varepsilon_{X_2} \tag{145}$$

Finally, when several parameters are simultaneously fitted, the propagation of errors rule leads to a system of lineal equations. Equations (146) and (147) show that for two parameters and one variable we have:

$$\varepsilon_{P_1} = \delta_{P_1X}\varepsilon_X + \delta_{P_1P_2}\varepsilon_{P_2} \tag{146}$$

$$\varepsilon_{P_2} = \delta_{P_2X}\varepsilon_X + \delta_{P_2P_1}\varepsilon_{P_1} \tag{147}$$

The resolution of this system will yield the propagated errors in the parameters.

Table 5 shows the relative error sensitivities for the determination of the initial concentration and rate constant for a second-order reaction at several reagent concentrations. The rate constant, k, is more affected by systematic errors than the initial concentration. On the other hand, error propagation is, in general, lower when pseudo-first-order experimental conditions are used.

3.3.2.3 Principal Component Regression Least-squares regression techniques can be misused if care is not

Table 5 Relative error sensitivities for the determination of the initial concentration and the rate constant

	$[A]_0$			k		
	k	$[R]_0$	S_b	$[A]_0$	$[R]_0$	S_b
$[R]_0 = 1$	−0.35	−0.35	−0.64	−2.4	−1.0	−1.7
$[R]_0 = 0.1$	−0.34	−0.35	−0.65	−2.4	−1.1	−1.8
$[R]_0 = 0.011$	−0.26	−0.73	−0.88	−2.9	−2.3	−2.8

Experimental conditions were: $[A]_0 = 0.01$, $s = 100$, $k = 1$ and 120 experimental points were used taking 95% extent of reaction. Relative error in blank signal, S_b, was referred to $S_\infty/2$.

applied in ascertaining the suitability of the model for describing the data. On the other hand, the correlation between the kinetic behavior of the different substances limits the number of species that can be simultaneously determined. Principal component regression (PCR), partial least-squares (PLS) regression and KF regression are procedures that have been applied to obtain kinetic parameters from noisy data, with models that are not well fixed, and to reduce the effects of correlation.

A major drawback of multiple least-squares is derived from the increase of collinearity among the predictor variables (the columns of the **X** matrix). Collinearity arises from similar signal responses for some analytes, from badly conditioned experimental designs, or from an overdimensioned number of predictor variables. All these problems increase when a large number of independent variables are involved. In this case, the Hessian becomes nearly singular and the inversion leads to a large variance in the parameters, as shown in Equation (141). This is a common situation when applying kinetic methods, due to the similar behavior of time-dependent signals in multicomponent analysis. PCR copes with this problem by performing a dimensional reduction by a principal component analysis of **X**.

The principal component analysis involves the rotation of coordinate axes to orientate them in the direction that maximizes the empirical variance of the predictor variables.[30] The new axes are called eigenvectors or loadings, **V**, and the new predictor variables, principal components, scores, or factors **T**. These scores are not correlated and, thus, principal component analysis compresses **X** to its most relevant factors **T**. The matrix **X** must be mean column-centered, previously. The transformation is given by Equation (148):

$$\mathbf{X} = \mathbf{T}\mathbf{V}^{\mathrm{T}} + \mathbf{E} \tag{148}$$

where **E** is the matrix of residuals of **X** generated when the number of factors that account for the variability in **X** is less than the number of original predictor variables. Thereby, the common structures in **X** are compressed into a stabilized, more easily interpretable model, leaving out much of the noise as residuals. Eigenvector decomposition can be performed using several algorithms.[4,31] The nonlinear iterative partial least-squares (NIPALS) method[30] is a simple algorithm that extracts one factor at a time, employing an iterative procedure to obtain the loading and score vectors from the residual matrix, obtained after estimation of the previous factors. In a second step, the Y-variables are regressed on **T**. Since the principal components are uncorrelated, the regression can proceed factor by factor, choosing only the factors that involve real variability of the **X** matrix and rejecting those that only account for noise. The PCR estimate of the parameters, using the most relevant factors, can be written as Equation (149):

$$\mathbf{P}_{\mathrm{PCR}} = \mathbf{V}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y} \tag{149}$$

Since the scores in **T** are uncorrelated, Equation (150) gives:

$$\mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathrm{diag}\left(\frac{1}{v_m}\right. \tag{150}$$

where v_m is the eigenvalue of the mth factor. The eigenvalues show how much variability each factor removes from **X**.

The main advantage of PCR is that it allows redundant factors and noise variability to be removed by performing a dimension reduction. However, an important problem is how to define the number of relevant factors. A good calibration practice is to extract more factors than are expected to be needed, and afterwards eliminate those that appear to be irrelevant or unreliable. Finally, it is important to realize that PCR coincides with multivariate least-squares regression when all factors are used.

3.3.2.4 Partial Least-squares Regression This method differs from PCR by including the dependent variable in the data compression and decomposition operations; this makes the method more complex as two loading vectors are required to provide orthogonality of the factors. The factors, called latent variables, are chosen in order to use **X** and **Y** efficiently, and provide a strong fit between them. In this way, the PLS factors can be seen as modified principal components improving the correlation. In the PLS algorithm, two principal component analyses are carried out, one for **X** and another for **Y**, by applying NIPALS to the column-centered matrices **X** and **Y** as shown in Equations (151) and (152):

$$\mathbf{X} = \mathbf{T}\mathbf{V}^{\mathrm{T}} + \mathbf{E} \tag{151}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{B}^{\mathrm{T}} + \mathbf{F} \tag{152}$$

where **V** and **B** are the loading matrices and **T** and **U** the score matrices for **X** and **Y**, respectively. A linear

relationship as shown in Equation (153) between latent variables is set to enhance the correlation between the predictors and the response:

$$\mathbf{U} = \mathbf{TQ} + \mathbf{D} \tag{153}$$

where **Q** is a diagonal matrix that accounts for the correlation between latent variables, and **D** is the residuals matrix. Equation (154) relates **T** to **Y**:

$$\mathbf{Y} = \mathbf{TC}^{\mathrm{T}} + \mathbf{F} \tag{154}$$

Finally, the estimated parameters can be obtained by using the expression in Equation (155):

$$\mathbf{P} = \mathbf{W}(\mathbf{V}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{C}^{\mathrm{T}} \tag{155}$$

where **W** is the matrix of **X**-factor weights that correlate each **u** vector with **X** residuals. Equation (156) gives:

$$\mathbf{w}_j = c\mathbf{E}_j^{\mathrm{T}}\mathbf{u}_j \tag{156}$$

where c is the scaling constant that makes the length of the final $\mathbf{w}_j$ vector equal to 1, and $\mathbf{E}_{\mathrm{j}}$ is obtained from Equation (157):

$$\mathbf{E}_j = \mathbf{X} - \sum_{i=1}^{j-1} \mathbf{t}_i\mathbf{p}_i^{\mathrm{T}} \tag{157}$$

and at the beginning $\mathbf{E}_1 = \mathbf{X}$.

When all factors are used, the PLS predictor is equal to the least-squares predictor. PLS and PCA are the so-called soft-modeling methods in which an empirical model is derived from the data. The model is built up from a number of calibration samples. The use of these methods in kinetic problems[32] is growing due to their efficiency in dealing with problems from incomplete models, imprecise data sets, and interactions among the components.

3.3.2.5 Kalman Filter The KF is a recursive algorithm that introduces a new experimental point at each calculation step, and optimizes the point residual and the mean error of the estimates of the parameters (state variables). The linear KF requires the model to be linear with respect to the state variables, and has been used for pseudo-first-order kinetic determinations.[33] In this algorithm, as in linear least-squares, the rate constants must be well-known and are supposed to be invariant from run to run. This assumption is a serious drawback that affects the results, since the pseudo-first-order rate constants are functions of many experimental factors (temperature, pH, ionic strength, and so on).

The extended KF must be used with nonlinear models. This algorithm is based on a linear term truncated Taylor's series expansion about the current estimate of the parameters. The extended KF is a statistically nonoptimal procedure, and more prone to the influence of the initial guesses of the parameters than the linear form, due to the inherent error caused by truncation of higher-order terms. Rate constants can be included among the parameters to be estimated and adjusted in the process of the filter, as in the NLS method. Both the least-squares and the KF algorithms are model-based, or hard-modeling methods, in which accurate model information is indispensable to obtain good results.

The extended Kalman algorithm for time-invariant parameters can be set out in the following steps:

1. Initialize $\mathbf{P}(0)$ and $\mathbf{x}(0)$, **P** being the error covariance matrix, related to the inverse of the Hessian in least-squares, and **x** the vector of the parameters to be estimated. It is recommended to choose for $\mathbf{P}(0)$ a diagonal matrix with very large diagonal elements, expressing the large uncertainty in the chosen starting values of $\mathbf{x}(0)$.
2. Compute the linearized measurement function vector, $\mathbf{h}(j)$, that coincides with the jth row of the Jacobian matrix as shown in Equation (158):

$$\mathbf{h}(j)_k = \frac{\partial S[\mathbf{x}(j-1)]_j}{\partial x_k} \tag{158}$$

   where S is the model estimated signal.
3. Update the Kalman gain vector, $\mathbf{k}(j)$, and the error covariance matrix using Equations (159) and (160):

$$\mathbf{k}(j) = \frac{\mathbf{P}(j-1)\mathbf{h}(j)}{\mathbf{h}^{\mathrm{T}}(j)\mathbf{P}(j-1)\mathbf{h}(j) + V(j)} \tag{159}$$

$$\mathbf{P}(j) = \mathbf{P}(j-1) - \mathbf{k}(j)[\mathbf{h}^{\mathrm{T}}(j)\mathbf{P}(j-1)] \tag{160}$$

   where $V(j)$ is the variance of the jth experimental point. The Kalman gain vector is chosen to be the gain or blending factor that minimizes the a posteriori error covariance matrix $\mathbf{P}(j)$.[34]
4. Predict the value of the signal at point j, $S(j)$, with the parameters obtained in the $j-1$ cycle, and evaluate the residual or innovation, $I(j)$, using Equation (161):

$$I(j) = Y(j) - S[\mathbf{x}(j-1)]_j \tag{161}$$

5. Update the predicted parameters using Equation (162):

$$\mathbf{x}(j) = \mathbf{x}(j-1) + \mathbf{k}(j)I(j) \tag{162}$$

6. Return to step 2, until all experimental points have been included.

When the parameters to be estimated are not constant but vary with time, a dynamic KF can be used in which a new step must be included between steps 5 and 6.

Table 6 Average absolute relative errors of the estimated concentrations of two binary mixtures using spectral-kinetic data and different algorithms of data treatment[a]

No. of wavelengths:	*m*-ABA/*p*-ABA[b]			*o*-ABA/*m*-ABA		
	1	6	41	1	6	36
NLS[c]	–[d]	56/20	61/7.8	–[d]	–[d]	–[d]
Linear Kalman	2.7/2.9	2.9/5.4	3.4/4.6	23/23	14/13	8.0/9.1
Extended Kalman	8.3/4.6	4.0/3.8	3.5/3.6	21/25	4.1/3.3	2.4/3.3
PLS	2.6/1.8	2.6/1.7	2.8/1.9	13/17	2.8/2.8	2.5/2.7

[a] From Xie et al.[35]
[b] ABA, aminobenzoic acid. Coupling reaction of diazotized sulfanilamide with the *o*-, *m*- and *p*-ABAs. The pseudo-first-order rate constants for the reactions performed at pH = 3.9 were $k_o = 4.67$, $k_m = 9.24$, and $k_p = 1.45$; these values are highly affected by pH variations.
[c] NLS regression using the Powell algorithm.[27]
[d] Extremely large absurd values.

5′. Extrapolate the parameters vector and the error covariance matrix using Equations (163) and (164):

$$\mathbf{x}(j+1|j) = \mathbf{F}(j+1, j)\mathbf{x}(j|j) + \mathbf{w}(j) \quad (163)$$

$$\mathbf{P}(j+1|j) = \mathbf{F}(j+1, j)\mathbf{P}(j|j)\mathbf{F}^{\mathrm{T}}(j+1, j) + \mathbf{Q}(j) \quad (164)$$

where the matrix $\mathbf{F}(j+1, j)$ is the system transition matrix, which describes how the system state changes from time t_j to t_{j+1}. The vector $\mathbf{w}(j)$ consists of the noise contribution to each of the parameters, and $\mathbf{Q}(j)$ is the variance–covariance matrix of the system noise which contains the variance of $\mathbf{w}$.

In line with KF literature, in the above description the common notation has been used in spite of the difference with the notation in other multivariate methods. The results of the KF depend on the initial estimates of the parameters. On the other hand, they rely on the initial guesses for the variance of the corresponding estimated values, and on the variance of the measurement noise.

Table 6 shows the average absolute relative error produced in the simultaneous determination of two substances using spectral-kinetic data, and different regression methods with three different numbers of wavelengths.[35] For the NLS the estimated error was extremely large for some of the mixtures. On the other hand, PLS and extended KF gave satisfactory results if the difference of the rate constants was large enough (*m*-ABA/*p*-ABA mixture). However, for closer rate constants (*o*-ABA/*m*-ABA mixture), a larger number of wavelengths must be used to improve the results. It can be deduced that PLS gave better results than the KFs, and that the extended KF performed better than the linear one for these experimental data. Besides, the more wavelengths used, the less the influence of the errors on the rate constants. Moreover, the influence of the number of wavelengths on the extended KF was not so significant as in the linear KF, which was attributed to the correction of the rate constants.

ABBREVIATIONS AND ACRONYMS

ABA	Aminobenzoic Acid
KF	Kalman Filter
KMS	Kezky–Mangerlsdorf–Swinbourne
NIPALS	Nonlinear Iterative Partial Least-squares
NLS	Nonlinear Least-squares
o.d.e.	Ordinary Differential Equation
PCR	Principal Component Regression
PK	Predictive Kinetic
PLS	Partial Least-squares

RELATED ARTICLES

Chemometrics **(Volume 11)**
Classical and Nonclassical Optimization Methods • Multivariate Calibration of Analytical Data • Soft Modeling of Analytical Data

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Catalytic Kinetic Determinations: Nonenzymatic • Differential Rate Determinations • Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Uncatalyzed Kinetic Determinations

REFERENCES

1. K.A. Connors, *Chemical Kinetics. The Study of Reaction Rates in Solution*, Wiley, New York, 1997.

2. W.E. Deming, *Statistical Adjustment of Data*, Dover, New York, 1984.
3. C.H. Bramford, C.H.F. Tipper, *Comprehensive Chemical Kinetics*, Elselvier, Amsterdam, Vol. 1, 1969.
4. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1992.
5. P. Gans, *Data Fitting in the Chemical Sciences*, John Wiley, Chichester, 1992.
6. M.R. Spiegel, *Mathematical Handbook of Formulas and Tables*, McGraw-Hill, New York, 1998.
7. J.H. Espenson, *Chemical Kinetics and Reaction Mechanisms*, McGraw-Hill, New York, 1995.
8. P. Moore, S.F.A. Kettle, R.G. Wilkins, 'The Kinetics of Formation of Blue Peroxychromic Acid in Aqueous Solution', *Inorg. Chem.*, **5**(3), 466–467 (1966).
9. M.R. Spiegel, *Theory and Problems of Laplace Transforms*, McGraw-Hill, New York, 1983.
10. X.Z. Zhang, A. Strand, H.D. White, 'A General Pre-steady Solution to Complex Reaction Mechanisms', *Anal. Biochem.*, **176**(2), 427–431 (1989).
11. J.J. Baeza-Baeza, F. Pérez-Pla, G. Ramis-Ramos, 'On the Integration of Kinetic Models Using a High-order Taylor Series Method', *J. Chemom.*, **6**, 231–246 (1992).
12. M. Hazewinkel (ed.), *Encyclopaedia of Mathematics*, Kluwer Academic Publishers, Dordrecht, Vol. 5, 1990.
13. J.J. Baeza-Baeza, F. Pérez-Pla, G. Ramis-Ramos, 'Stiffness-adaptative Taylor Method for the Integration of Non-stiff and Stiff Kinetic Models', *J. Comput. Chem.*, **13**(7), 810–820 (1992).
14. C.W. Gear, *Numerical Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
15. K.J. Laidler, *Chemical Kinetics*, Addison-Wesley, Oxford, 1987.
16. H.A. Mottola, *Kinetic Aspects of Analytical Chemistry*, Wiley & Sons, New York, 1988.
17. G.E. Mieling, H.L. Pardue, 'Kinetic Method that is Insensitive to Variables Affecting Rate Constants', *Anal. Chem.*, **50**(12), 1611–1618 (1978).
18. J.D. Ingle, Jr, S.R. Crouch, 'Signal-to-noise Ratio Theory of Fixed-time Spectrophotometric Reaction Rate Measurements', *Anal. Chem.*, **45**(2), 333–338 (1973).
19. J.J. Baeza-Baeza, K. Matsumoto, H.A. Mottola, 'Comparative Studies on Data Collection and Data Treatment in Kinetic-based Determinations with Two Rotating Bioreactor/Amperometric Detection Systems', *Anal. Chim. Acta*, **283**, 785–793 (1993).
20. A. Cornish-Bowden, *Analysis of Enzyme Kinetic Data*, Oxford University Press, Oxford, 1995.
21. F.J. Holler, R.K. Calhoun, S.F. McClanahan, 'Minimization of Errors in Fixed-time Reaction Rate Methods by Optimization of Measurement Time', *Anal. Chem.*, **54**(4), 755–761 (1982).
22. P.W. Carr, 'Kinetic and Equilibrium Assays Based on the Effect of Extent of Reaction and the Uncertainty of Rate Parameters', *Anal. Chem.*, **50**(12), 1602–1607 (1978).
23. B.M. Quencer, S.R. Crouch, 'Multicomponent Kinetic Methods', *Crit. Rev. Anal. Chem.*, **24**, 243–262 (1993).
24. D. Pérez-Bendito, M. Silva, *Kinetic Methods in Analytical Chemistry*, Ellis Horwood, Chichester, 1988.
25. B.G. Willis, W.H. Woodruff, J.M. Frysinger, D.W. Margerum, H.L. Pardue, 'Simultaneous Kinetic Determinations of Mixtures by On-line Regression Analysis', *Anal. Chem.*, **42**(12), 1350–1355 (1970).
26. J.J. Baeza-Baeza, G. Ramis-Ramos, 'Reduction of the Relative Standard Deviation in the Least-squares Fitting of Linearized Equations by Using Pointwise Sensitivity Weights', *Anal. Chim. Acta*, **316**, 173–184 (1995).
27. E.K.P. Chong, *An Introduction to Optimization*, Wiley–Interscience, New York, 1995.
28. S.S. Rao, *Engineering Optimization*, Wiley & Sons, New York, 1996.
29. J.J. Baeza-Baeza, G. Ramis-Ramos, C. Mongay Fernández, 'Comparative Study of Several Programs Used in the Potentiometric Evaluation of Equilibrium Constants Including an Error Sensitivity Analysis', *Anal. Chim. Acta*, **223**, 419–427 (1989).
30. H. Martens, T. Naes, *Multivariate Calibration*, John Wiley, Chichester, 1989.
31. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Elsevier, Amsterdam, Vol. 2, 1998.
32. T.F. Cullen, S.R. Crouch, 'Multicomponent Kinetic Determinations Using Multivariate Calibration Techniques', *Mikrochim. Acta*, **126**(1–2), 1–9 (1997).
33. R. Jiménez-Prieto, A. Velasco, M. Silva, D. Pérez-Bendito, 'Kalman Filtering of Data from First- and Second-order Kinetics', *Talanta*, **40**(11), 1731–1739 (1993).
34. R.G. Brown, P.Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, Wiley & Sons, New York, 1992.
35. Y.L. Xie, J.J. Baeza-Baeza, G. Ramis-Ramos, 'A Comparative Study of Several Chemometric Methods Applied to the Treatment of Two-way Kinetic-spectral Data for Mixture Resolution', *Anal. Chim. Acta*, **321**, 75–95 (1996).

Differential Rate Determinations

M. Dolores Pérez-Bendito
University of Córdoba, Spain

1 Introduction **1**
2 Theoretical and Methodological Background **2**
2.1 General Considerations 2
3 Kinetic Discrimination: Classical Methods **3**
3.1 Logarithmic Extrapolation and Proportional Equations 3
3.2 Synergism 4
4 Mathematical Discrimination: Kinetometric Methods **5**
4.1 Multipoint Curve-fitting Methods 5
4.2 Kalman Filtering 5
4.3 Artificial Neural Networks 6
5 Spectral/Instrumental Discrimination **8**
5.1 Multivariate Methods Using Multiwavelength Spectroscopy 8
5.2 Kinetic Wavelength-pair and H-Point Methods 9
5.3 Miscellaneous Methods 10
6 Physicochemical Discrimination: Micellar Catalysis **11**
7 Practical Implementation: Automated Methods **13**
7.1 Mixing Modes and Instrumental Assemblies 13
7.2 High-pressure Stopped-flow Technique 14
7.3 Low-pressure Continuous Flow Techniques 15
7.4 Low-pressure Continuous Addition of Reagent Technique 16
7.5 Miscellaneous Instrumental Approaches 16
8 Multicomponent Kinetic Determinations as Problem Solvers **18**
Acknowledgment **18**
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Kinetic differential rate determinations are a measurement approach in kinetic methods of analysis where two or more components are determinated based on the different rates at which they interact with a common reagent.

1 INTRODUCTION

Multicomponent analysis is one of the main goals of the modern analytical chemistry. The simultaneous or sequential determination of several analytes in the same sample can be accomplished without or with physical separation in space (e.g. in chromatographic techniques). Multicomponent kinetic determinations, traditionally called ''differential reaction rate methods'', are based on the different rate at which two on more species in solution react with a common reagent.[1] They allow the variation of the analytical signal yielded by each species to be ''separated'' on a different timescale, thereby avoiding the need to separate the components physically. Separation techniques are usually time-consuming and may result in sample loss and contamination. Hence differential rate methods are of great interest in the context of kinetic methods of analysis as they are effective alternatives to multicomponent analysis, especially when only a few components in a mixture are to be determined. The principles and features of differential rate methods have been compiled in two books on kinetic methods[2,3] and also in several reviews on the topic.[4,5]

Classical differential rate methods, developed in the 1960s–1970s, were based on simple mathematical equations and required only a few kinetic data for application. They were only applied to first- or pseudo-first-order reactions in order to resolve binary mixtures with a rate constant ratio above about three. Modern differential rate methods combine rate discrimination with mathematical and/or spectral/instrumental discrimination to provide much better performance and a wider scope of application by virtue of their using much greater amounts of information. The development of new computerized data processing methods has been the key to both the implementation of variably complex mathematical algorithms (kinetometric approaches) and the handling of multidimensional data such as the multiwavelength readings provided by diode array spectrophotometers. These methods have expanded the scope of conventional reaction rate methods to reaction orders higher than first- or pseudo-first order, rate constant ratios near unity and even mixtures exhibiting mutual kinetic effects (synergism). As the efficiency of mathematical discrimination and/or spectral discrimination increases, the need for kinetic discrimination (based on the difference between

reaction rates) is reduced, so much so that a system with completely overlapped kinetics can be effectively resolved provided that some spectral difference exists at the wavelength considered. These methods thus afford the kinetic resolution of mixtures of analytes on the basis of spectral differences rather than kinetic differences; as a result, the designation ''multicomponent kinetic methods'' is more appropriate than differential reaction rate methods. Finally, rate discrimination can also be combined with the physicochemical discrimination provided by micellar systems.

The potential of these modern methods can only be fully exploited for analytical purposes by using automated methods, which provide results of increment quality and enable multicomponent determinations in routine analyses. Various automated approaches (continuous and stopped-flow (SF) included) according to the way sample and reactants are mixed are possible. Their performance in terms of mixing expeditiousness and homogeneity depends on the rates of the reactions involved, especially with fast reactions (viz. with lifetimes of a few milliseconds or shorter).

This article describes currently available methods for kinetic multicomponents analysis that provide high-quality results and encompass a wide range of analytes in mixtures and analyte concentration ratios when used in combination with an automated method suitable for routine analyses.

2 THEORETICAL AND METHODOLOGICAL BACKGROUND

2.1 General Considerations

Differential kinetic methods have traditionally focussed on mixtures of components reacting at slightly or strongly different rates; however, methods for mixtures exhibiting small rate differences account for the majority of the work in this field. The analytical applicability of these methods is dependent on the relative values of the rate constants and on the ratio between the concentrations of the mixture components.

Let A and B be two substances in a mixture that react with a reagent R to yield two products (P and P′) which, though different in nature, are essentially similar with regard to the analytical property to be measured, according to the reactions shown here (Equations 1 and 2):

$$A + R \xrightarrow{k_A} P \quad (1)$$

$$B + R \xrightarrow{k_B} P' \quad (2)$$

where k_A and k_B are the rate constants for reaction of A and B, respectively, which are first-order with respect to each mixture components (i.e. $[R]_0 \gg [A]_0$ and $[R]_0 \gg [B]_0$).

If the mixture to be resolved exhibits a large difference between the reaction rates (i.e. k_A/k_B is sufficiently different from unity), then the analysis is quite simple because each species can be treated separately. In this case, one can assume that, over any time interval, only one species will react since the other either will have already reacted completely or be reacting so slowly that it could hardly interfere with the reaction of interest.

If $k_A > k_B$ but B reacts to an appreciable extent over a given time interval, then the ratio of their rate constants, k_A/k_B, needed to obtain a given error in the concentration of A can be determined from the ratio of the respective first-order rate integrated equations for both components (Equations 3 and 4):

$$\ln\left(\frac{[A]_0}{[A]_t}\right) = k_A t; \quad \ln\left(\frac{[B]_0}{[B]_t}\right) = k_B t \quad (3)$$

i.e.

$$\frac{\ln([A]_0/[A]_t)}{\ln([B]_0/[B]_t)} = \frac{k_A}{k_B} \quad (4)$$

This equation is time-dependent and relates the extent of both reactions to their rate constants. If a maximum error of 1% is considered to be acceptable (i.e. if there is still 1% of A present when less than 1% of B has reacted) then the k_A/k_B ratio in Equation (4) should be at least 463. In practice, a minimum ratio of 500 is commonly adopted.

As the ratio of the two rate constants, k_A/k_B, approaches unity, the errors in the estimated concentrations of the components in the mixture become larger. Hence, traditional methods for small rate differences involve the selective determination of the more rapidly reacting species, either by altering the kinetics of a system via appropriate changes in the reaction conditions (e.g. temperature, solvent, reactant concentrations) or by masking the species with a suitable inhibitor. The concentration of the second species can be obtained by a similar approach or by the difference from a determination of the sum of the two concentrations. However, the most frequently used methods for the simultaneous determination of species with small differences in their rate constants are based on mathematical approaches for systems having unalterable rate constants (such constants often can not be separated to a sufficient degree by either of the above-described types of method). The most popular approaches to differential rate determinations

historically have been logarithmic extrapolation methods and the method of proportional equations (classical methods).[2]

These methods are available for various reaction orders[2] of which the pseudo-first is the most widely considered. They assume the simultaneous contribution of the two-reaction component to the analytical signal according to Equations (1) and (2). If the pseudo-first order reaction rates of two substances are different and independent of each other (i.e. if no mutual kinetic effects are present), then the sum of the contribution of A and B at time t, C_t, is given by Equation (5)

$$C_t = [\mathrm{A}]_t + [\mathrm{B}]_t = [\mathrm{P}]_\infty - [\mathrm{P}]_t = [\mathrm{A}]_0 \exp(-k_\mathrm{A} t) + [\mathrm{B}]_0 \exp(-k_\mathrm{B} t) \qquad (5)$$

where $[\mathrm{A}]_t$ and $[\mathrm{B}]_t$ are obtained, respectively, from the exponential forms of the integrated rate equations (Equation 3). $[\mathrm{P}]_\infty$ is the maximum total product concentration, which is usually measured after allowing the reaction to proceed for a sufficiently long time and is in itself a measure of the total initial concentration of reactants, $[\mathrm{A}]_0 + [\mathrm{B}]_0$, [P] can be replaced with a measurable property proportional to the product concentration such as the absorbance, fluorescence or chemiluminescence intensity, or potential. Equation (5) is the basis for application of classical and some kinetometric methods.

3 KINETIC DISCRIMINATION: CLASSICAL METHODS

3.1 Logarithmic Extrapolation and Proportional Equations

The *logarithmic extrapolation method* is the most commonly used graphical linear method for the simultaneous kinetic determination of species. The method is based on Equation (5) and uses plots of $\log C_t$ or $\log[(\mathrm{P})_\infty - (\mathrm{P})_t]$ as a function of time (Figure 1). The result is a curve unless $k_\mathrm{A} = k_\mathrm{B}$, in which case a straight line is obtained. If species A disappears at a higher rate than B (i.e. if $k_\mathrm{A} > k_\mathrm{B}$), then $[\mathrm{A}]_t \to 0$ and the curve eventually becomes a line (Figure 1), Equation (6)

$$\log C_t = \log([\mathrm{P}]_\infty - [\mathrm{P}]_t) = \log [\mathrm{B}]_0 - \frac{k_\mathrm{B} t}{2.303} \qquad (6)$$

The intercept of the curve provides the initial concentration of B, $[\mathrm{B}]_0$, that of the other component, $[\mathrm{A}]_0$, being obtained by difference once $[\mathrm{A}]_0 + [\mathrm{B}]_0$ has been determined from $[\mathrm{P}]_\infty$. If a photometric detection technique is applied, then Equation (6) can be written as Equation (7):

$$\log(D_\infty - D_t) = \log \varepsilon_\mathrm{B}[\mathrm{B}]_0 - \frac{k_\mathrm{B} t}{2.303} \qquad (7)$$

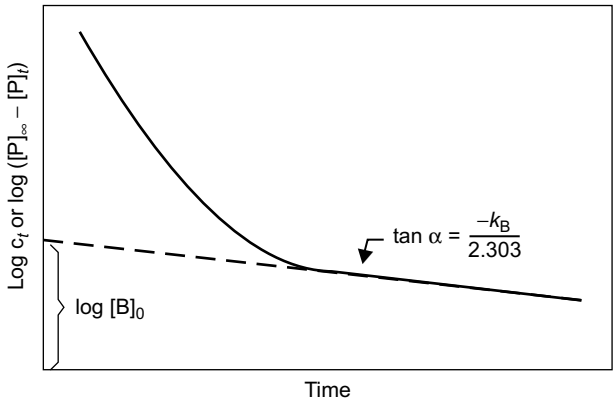

Figure 1 Logarithmic extrapolation method for first-order reactions.

where D_∞ and D_t are the absorbances measured after the reaction has developed to completion and after a time t near the start of the linear region, and ε_B is the molar absorptivity of B (known beforehand).

This method features several advantages, namely: (1) no prior knowledge of the rate constant of A or B is necessary; (2) it is less prone to error as the linear plot is constructed from several points; (3) it can be applied to ternary mixtures provided the rate constants of the three components are sufficiently different (even in the presence of synergistic effects) as it is not limited to pseudo-first-order reactions; (4) rate constant ratios as low as five are acceptable under certain conditions (e.g. with favorable concentration ratios and if somewhat large errors are acceptable). However, at least 99% of the more reactive species should be consumed in the process if valid data are to be obtained and the total initial concentration of all the species involved should be known for the method to be applicable. In addition, the method is somewhat laborious as it entails the continuous recording of the measured property and obtaining the required $[\mathrm{P}]_\infty - [\mathrm{P}]_t$ values from the resulting curve. For two reactions following second-order kinetics in which $[\mathrm{R}]_0 \approx [\mathrm{A}]_0 + [\mathrm{B}]_0$, the reaction rate can be expressed by Equation (8)

$$\frac{-\mathrm{d}[\mathrm{R}]}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t} = k_\mathrm{A}[\mathrm{R}]_t[\mathrm{A}]_t + k_\mathrm{B}[\mathrm{R}]_t[\mathrm{B}]_t \qquad (8)$$

where x is the amount of R consumed at any time t. If A reacts faster than B, then, when the [A] has reacted essentially to completion, Equation (8), after integration, gives Equation (9)

$$x = k_\mathrm{A}[\mathrm{B}]_t([\mathrm{R}]_0 - x)t + [\mathrm{A}]_0 \qquad (9)$$

$[\mathrm{A}]_0$ is then determined from the extrapolated intercept at $t = 0$ of a plot of x versus $[(\mathrm{R})_0 - x]t$.

The *proportional equation method* is a mathematical approach of wide use for the resolution of mixtures of closely related species; it is usually less time-consuming

than the previous method (especially if a microcomputer is used) and in the most cases requires no prior knowledge of the total concentrations of the mixture component. A rate constant ratio of four is usually acceptable and the ease with which the method can be automated make it suitable for fast reactions. However, it cannot be applied to processes subject to synergistic effects because the measured property must be additive in nature. On the other hand, the rate constants must be carefully calculated; if the reaction involves complex kinetics, then the applicability of the method relies on the accurate calculation of the different proportionality constants.

The practical procedure for pseudo-first rate reactions entails measuring C_t at two or more reaction times and formulating two Equations (10) and (11) similar to Equation (5):

$$[P]_\infty - [P]_{t_1} = [A]_0 \exp(-k_A t_1) + [B]_0 \exp(-k_B t_1) \quad (10)$$

$$[P]_\infty - [P]_{t_2} = [A]_0 \exp(-k_A t_2) + [B]_0 \exp(-k_B t_2) \quad (11)$$

the resolution of which provides the initial concentrations of the different components as functions of their respective rate constants (known beforehand).

Not only time, but any other experimental variable such as temperature or one related to the solvent or the physicochemical characteristics of the reactants can be used to formulate the simultaneous equations from which the required concentrations are to be calculated. Thus, the general equations for a binary mixture will be of the form (Equations 12 and 13)

$$f_1 = k_{A,1}[A]_0 + k_{B,1}[B]_0 \quad (12)$$

$$f_2 = k_{A,2}[A]_0 + k_{B,2}[B]_0 \quad (13)$$

where f is any measurable parameter determined under two different sets of conditions, and $k_{A,1}$ and $k_{A,2}$ are empirical or proportionality constants obtained separately for component A under different conditions in the same way as $k_{B,1}$ and $k_{B,2}$ are obtained for component B. Although any experimental variable is usable provided $k_{A,1}k_{B,2} \neq k_{A,2}k_{B,1}$, the optimum results are obtained with $k_{A,1}/k_{B,1} > 1$ for one set of conditions and $k_{A,2}/k_{B,2} < 1$ for the other.

Applications of these classical methods, most of which involve the resolution of binary mixtures of metal ions, have been compiled.[5]

3.2 Synergism

Both classical differential reaction-rate methods and new approaches rely on the assumption that the reaction rates of the species in the mixture are independent of each other. When this is not so, the mixture is said to be subject to a mutual kinetic or synergistic effect. Although synergism is not a frequent phenomenon, it is the chief source of error in differential kinetic analysis and should therefore be duly studied and controlled to ensure the accurate resolution of mixtures. Only the second-order logarithmic extrapolation and the Connors graphical extrapolation methods[2] are applicable in this event; in fact, if the cause of synergism is known, the reaction conditions are usually modified to avoid it.

Several approaches to the simultaneous resolution of mixtures subject to synergism have been reported. One is based on the Connors graphical interpolation method and involves using equations that can be formulated on the basis of any experimental parameter provided it is proportional to the concentration of the species concerned.[6] The method involves matching the concentration–time profile of a simple mixture to that of a reference mixture so that the initial concentrations of sample and reference are identical. Consider a mixture of two components, A and B, which react under pseudo-first order conditions with a reagent R to yield the same product, P. First, the concentration of P is measured as a function of time ($[P]_t^s$). Next, reference mixtures are reacted under the same conditions and $[P]_t^r$ is measured as a function of time. Once the concentration–time profiles for the sample and reference have been fully matched throughout the time range, $[P]_t^s = [P]_t^r$ and hence $[A]_0^r = [A]_0^s$ and $[B]_t^r = [B]_t^s$.

In practice, the calculation of $[A]_0^s$ involves the construction of several calibration graphs from plots of the final absorbance of component B in the presence of different known amounts of A. The intercepts of these plots are dependent on the concentration of component A. The concentration of B is calculated from the plots of the difference in the initial rates (IR) of the concentration–time curves for the sample and several reference solutions, viz. $(IR)^s - (IR)^r$ versus [B]. The method is also useful at low rate-constant ratios, which often pose problems in differential rate analyses. Its performance was evaluated with fast direct complex-formation reactions between cobalt and nickel and pyridoxal thiosemicarbazone (PTSC), which were subject to a synergistic effect, using a customized SF mixing module. The rate constant ratio was $k_{Ni}/k_{Co} = 2.63$. The determination of nickel in real samples entailed the prior determination of their cobalt content by applying the above-described procedure for calculation of $[A]_0^s$. Once this concentration had been evaluated, different known amounts of nickel in the range 1.0–5.0 $\mu g\,mL^{-1}$ were added to a sample aliquot containing a known cobalt concentration between 1.0–2.5 $\mu g\,mL^{-1}$, and reacted with PTSC.

4 MATHEMATICAL DISCRIMINATION: KINETOMETRIC METHODS

As noted in section 3, classical differential kinetic methods are based on simple mathematical equations and require only a few data points and rate constant ratios higher than about four for binary mixtures. The inception of digital computers has facilitated the use of multipoint methods that provide more information by exploiting the whole kinetic curve. This trend has materialized in the application of mathematical algorithms based on specific kinetic models[7] which allow the concentrations of two or more analytes in mixtures to be determined, whether or not the rate constants and reactions order involved are known beforehand, and for k_A/k_B ratios closer to unity.

Kinetometrics is a concept of increasing interest[8] as a result of dynamic systems frequently exhibiting a nonlinear relationship that can readily be modeled with computational aids. Curve fitting and related algorithms have proved useful in simultaneous kinetic determinations. In this section, the usefulness of recent kinetometric approaches such as Kalman filtering and artificial neural networks (ANN) is also discussed.

4.1 Multipoint Curve-fitting Methods

Curve-fitting methods compute values of parameters that represent the best fit of some mathematical model to signal–time data. Although curve fitting methods excel over direct computation methods (classical methods) for processing linear responses (pseudo-zero-order processes), they are also especially suitable for nonlinear responses.[7] Fits to nonlinear responses can be used to predict or extrapolate data obtained in another time region of the process.

In multipoint curve-fitting methods for kinetic analysis of mixtures[9,10] signal–time data pairs collected over a fixed period (7–8 half-lives of the slowest reacting component) are fitted to a mathematical model descriptive of the kinetic behavior of the mixture. The model frequently involves monitoring two or three pseudo-first-order reactions simultaneously through absorbance changes. The absorbance estimates as $D_\infty - D_0$ (D_∞ and D_0 being the absorbances at infinite and zero time, respectively) for each component, which result in the best least-squares fit to the experimental model, are used to compute the concentration of each component. This requires a prior knowledge of the pseudo-first-order rate constant of each component and the absorptivity of its product; also, no mutual kinetic effect should be present. This approach also affords satisfactory results from data processed over two half-lives of the slower reacting component by using nonlinear and multiple linear curve-fitting methods that yield virtually identical results.

Other reported methodologies for the resolution of mixtures of species include a new algorithm developed by Schechter[11] to resolve binary mixtures which requires no prior knowledge of the reaction orders or rate constants of the species involved and was tested on simulated data only. In addition, it allows the concentrations of the mixture components to be determined over broad ranges and the corresponding rate constants and reaction orders to be calculated. The algorithm, however, is somewhat less precise with reaction orders of exactly unity.

4.2 Kalman Filtering

Kalman filtering is an alternative curve-fitting approach that is gaining increasing interest in various fields of analytical chemistry and can be used to compute analyte concentrations or activities in enzymatic reactions. It is one of the most powerful algorithms available today for the simultaneous kinetic determination of species in mixtures (it is only surpassed by the neural network approach). Although it can be applied to single-component systems in order to calculate the initial concentration of the analyte and its first-order rate constant, its main applications are the correction of errors associated with between-run variations and, especially, simultaneous kinetic determinations. Thanks to its recursive nature, determinations can be performed in real time (thereby expediting the process) provided half lives allow it. It has proved useful for fitting linear and nonlinear (extended Kalman filter) kinetic data, depending on whether or not the rate constants of the two reactions are known beforehand (although the results are better when they are).

When the algorithm is used in its nonlinear form, its performance relies heavily on the initial parameter chosen and the number of iterations performed. Figure 2 illustrates how the Kalman filter can be used in multicomponent kinetic analysis. It is based on the rate equation for a binary mixture (Equation 5), which, taking into account that $[P]_t = [A]_0 - [A]_t + [B]_0 - [B]_t$ and that $[P]_t$ is replaced with the total absorbance at time t, D_t, can be expressed by Equation (14)

$$D_t = D_A[1 - \exp(-k_A t)] + D_B[1 - \exp(k_B t)] + C \quad (14)$$

where D_A and D_B and k_A and k_B are the net absorbance change and the pseudo-first-order rate constants for the two components, respectively, and C is the contribution of the background absorbance.

Investigations have shown this algorithm to be superior to classical multicomponent kinetic methods such as those of logarithmic extrapolation, single point and proportional equations in many respects, especially with regard to the affordable rate constant ratios of reactions for a binary mixture, which can be closer to unity than in the above-mentioned classical methods. Such is the case with

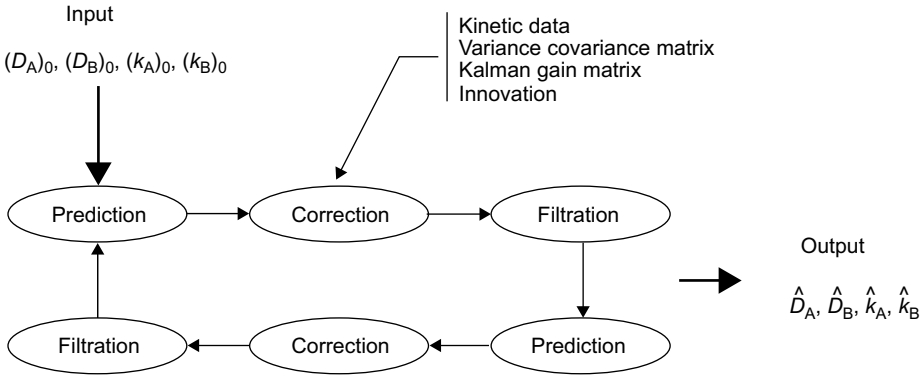

Figure 2 Schematic diagram illustrating the application of the extended Kalman filter in multicomponent kinetic analysis. (Reproduced from Pérez-Bendito, Silva, *Trends Anal. Chem.* Copyright (1996) by permission of Elsevier Science.[8])

Table 1 Analysis of synthetic cortisone–hydrocortisone mixtures by use of the extended and linear Kalman filters and classical differential reaction rate methods[a]

Method	Corticosteroid (μg)		Cortisone		Hydrocortisone	
	Cortisone	Hydrocortisone	Found (μg)	Error (%)	Found (μg)	Error (%)
Extended Kalman filter	200	200	196.4	−1.80	206.7	3.35
	80	200	75.3	−5.87	207.0	3.50
	200	80	216.7	8.36	87.8	9.75
	200	400	205.4	2.70	401.3	0.32
	120	400	111.8	−6.83	430.5	7.62
	160	240	170.0	6.25	246.9	2.87
Linear Kalman filter	200	200	194.2	−2.90	197.2	−1.40
	80	200	76.6	−4.25	193.0	−3.50
	200	80	214.3	7.15	75.2	−6.00
	200	400	194.9	−2.55	411.2	2.80
	120	400	111.8	−6.83	436.5	9.12
	160	240	168.4	5.25	249.7	4.04
Logarithmic extrapolation	200	200	263.3	31.65	108.1	−45.95
	80	200	–[b]	–	–	–
	200	80	207.2	3.60	68.7	−14.12
	200	400	358.8	79.40	206.3	−48.42
	120	400	78.5	−34.58	455.0	13.75
	160	240	287.1	79.44	83.0	−65.41
Proportional equation	200	200	222.7	11.35	166.3	−16.85
	80	200	113.6	42.00	159.3	−20.35
	200	80	–	–	–	–
	200	400	280.0	40.00	313.1	−21.72
	120	400	155.2	29.33	386.8	−3.30
	160	240	228.6	42.87	168.8	−29.67

[a] Reproduced from Xiong et al., *Anal. Chim. Acta* by permission of Elsevier.[12]
[b] Negative concentration values.

the resolution of mixtures of cortisone and hydrocortisone by reaction with Tetrazolium Blue, the rate constant ratio of which is 1.8; the results compared favorably with those of classical methods[12] (Table 1). Other mixtures of interest have also been satisfactorily resolved by the Kalman filtering.[8] The use of this filter is not limited to pseudo first-order reactions; in fact, it also allows the resolution of mixtures of species following different kinetics (pseudo-first order and second order).[13] The filter model used for this purpose was assessed both on simulated data and on the analysis of cysteine–ascorbic acid and gluthatione–ascorbic acid mixtures, which were resolved with good accuracy over wide concentration ranges.

4.3 Artificial Neural Networks

ANN, also called computational neural networks (CNN), are increasingly used in analytical chemistry owing to

their ability to "learn" arbitrary nonlinear relationships between input and output spaces.[8] An ANN can adapt itself autonomously to learn to recognize key features in a data set by repeatedly examining the same or similar data (the training set). ANNs can process information rapidly once the system is trained since the training process may require several hours of dedicated computer time depending on the particular training method. Several training methods have so far been used, the multilayer back-propagation method being the most widely employed so far.

The intrinsic features of ANNs afford the resolution of complex mixtures of species without the aid of spectral discrimination[14] even if the systems involved have almost identical rate constants and exhibit synergism. By contrast, multivariate calibration methods (MCMs) such as nonlinear least-squares regression (NLR) and principal components regression (PCR) (and even the Kalman filter) are effective approaches to multicomponent kinetic determination when they combine kinetic and spectral discrimination (multiwavelength data) as shown in the next section.

The type and number of input data to be used in an ANN must be carefully chosen in order to ensure accurate results in the concentration of the mixture components, with minimal training costs. This has been accomplished by developing a straightforward methodology.[14] For a binary mixture, Equation (14) can be fitted to Equations (15), the exponential function using NLR:

$$\Delta D_t = \Delta D_T[1 - \exp(-\hat{k}t)] + C \qquad (15)$$

where ΔD_T and $\hat{k}$ are related to the true values of ΔD_A, ΔD_B, k_A and k_B. $\hat{k}$ can be considered the estimated rate constant for the mixture, which approaches k_A as the contribution of component A in the mixture increases and vice versa. ΔD_T and $\hat{k}$ values obtained from the kinetic curve for the mixture concerned were used as input to the ANN in order to obtain ΔD_A and ΔD_B or the mixture concentrations as output data by means of a very simple neural network architecture such as 2:4s:2l (which means 2 input data, 4 nodes in the hidden layer and 2 output data) with no prior knowledge of the respective rate constants. A scheme of the operating mode of an ANN corresponding to this architecture is shown in Figure 3.

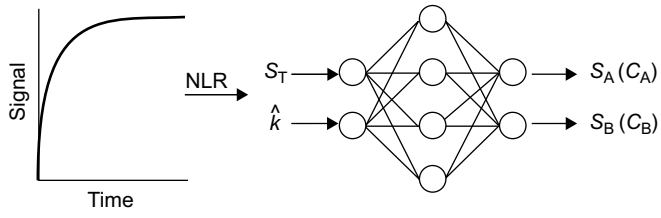

Figure 3 Schematic diagram of the foundation and architecture of the ANN used for multicomponent kinetic determinations. (Reproduced from Ventura et al., *Anal. Chem.*,[14] by permission. Copyright (1995) American Chemical Society.)

This method was evaluated in the resolution of mixtures of 2-chlorophenol (2-ChPh) and 3-chlorophenol (3-ChPh) (rate constant ratio 1.4) with fairly good results: the values of the relative standard error of prediction (%SEP) were 4.23 and 6.20%, respectively, and the analyte concentration ratio range was 1:9–9:1. The performance of the method compared favorably with that of the Kalman filter (Table 2), which yielded relative errors near

Table 2 Comparison of the quality achieved in the resolution of 2-ChPh/3-ChPh mixtures using ANNs and Kalman filter[a]

[2-ChPh]/[3-ChPh]	ANNs error (%)		Kalman filtering error (%)	
	2-ChPh	3-ChPh	2-ChPh	3-ChPh
9:1	2.41	−10.87	−2.50	39.5
5:1	−3.45	5.52		
4:1	−4.06	6.96	−3.25	13.0
3:1[b]	−1.35	4.65	2.00	6.83
2:1	1.91	4.31	0.00	1.25
1:1	−0.95	2.82	−2.00	−4.00
1:2[b]	−0.79	0.81	2.37	−3.50
1:3[b]	−1.91	−2.74	8.96	−9.78
1:4	−1.27	2.96	8.75	−6.06
1:5	2.27	−2.25		
1:6	−4.67	4.32		
1:7	−9.92	2.61		
1:9	−10.23	2.31	21.5	−5.66

[a] Reproduced from Ventura et al., *Anal. Chem.*, by permission. Copyright (1995), American Chemical Society.[14]

[b] The Kalman filter results at these ratios correspond to exactly 2.3:1, 1:1.5 and 1:2.3, respectively.

to or about 10% in the resolution of 2-ChPh–3-ChPh mixtures in ratios from only 4:1 to 1:4.[14] On the other hand, the proposed network architecture could be extrapolated to multicomponent analysis involving other kinetic situations such as species following different kinetics and the presence of synergistic effects. However, realizing the actual potential of ANNs in kinetic analysis requires further research.

5 SPECTRAL/INSTRUMENTAL DISCRIMINATION

The kinetic methods described in this section are also based on mathematical algorithms but rely heavily on the spectral discrimination afforded by multiwavelength data simultaneously provided by special optical spectroscopic devices (instrumental discrimination) such as diode array spectroscopy and the use of T-optic configurations. In general, kinetic discrimination is also present to a greater or lesser extent. However, the use of multiple wavelengths allows systems having rate constants too close for analysis by normal means to be successfully determined provided they exhibit any spectral differences. In some cases, instrumental discrimination enables the kinetic determination of mixture components irrespective of their kinetics provided sufficient spectral differences between them exist.

5.1 Multivariate Methods Using Multiwavelength Spectroscopy

The use of several wavelengths in kinetic experiments allows a greater variety of systems to be successfully analyzed. In fact, rapid-scanning spectrophotometers such as diode array models were the earliest instruments used to perform multiwavelength studies in multicomponent kinetic analysis.[5,10] Full spectra can be recorded within a few seconds and simultaneous absorbance measurements made at several wavelengths instead of the absorbance–time data provided by conventional instrumentation are used as kinetic data. Thus, for a kinetic reaction involving a single component at an initial concentration C_0, absorbance changes with time can be expressed by Equation (16)

$$A_{\lambda,t} = \varepsilon_\lambda l C_0[1 - \exp(-kt)] + B_{\lambda,C} \tag{16}$$

For a mixture of two components 1 and 2 the absorbance is given by Equation (17)

$$A_{\lambda,t} = \varepsilon_{\lambda,1} l C_{0.1}[1 - \exp(-k_1 t)] + \varepsilon_{\lambda,2} l C_{0.2}[1 - \exp(-k_2 t)] + B_{\lambda,C} \tag{17}$$

Finally, for a mixture of n components, the simplified expression in Equation (18) holds:

$$A_{\lambda,t} = \sum_{i=1}^{n} K_{(i,\lambda,t)} C_{0.i} + B_{\lambda,c} \tag{18}$$

where $A_{\lambda,t}$ is the absorbance measured at wavelength λ at time t, $\varepsilon_{\lambda,\,1\ldots n}$ are the molar absorptivities of the components at the working wavelength, ℓ is the pathlength, $C_{0.1\ldots n}$ is the initial concentration of each component, $k_{1\ldots n}$ are the respective rate constants, $B_{\lambda C}$ is the background absorbance at $t = 0$ and $K_{(1,\lambda,t)}$ is a constant which depends on the particular component, its rate constant, wavelength and time.

Equation (18) is in itself a system of nonlinear equations that can be solved using various chemometric procedures such as the Newton–Gauss algorithm.[15] Although the molar absorptivity, rate constant and concentration of each mixture component can be determined simultaneously, the most interesting application is the determination of rate constants and concentrations as a compromise between computational costs and the ability to introduce corrective actions for fluctuations in the rate constants arising from environmental factors (error-compensated method), taking into account that absorptivities can be calculated individually with greater accuracy from equilibrium data.

The resolution of cobalt and nickel in mixtures (including the estimation of the rate constant) using the ligand-exchange reaction with EGTA (ethylene glycol bis(2-aminoethyl ether *N*,*N*,*N*′,*N*′-tetraacetic acid)) and PAR (4-(2-piridylazo)resorcinol) is one of the systems most widely used to validitate this approach.[15]

The application of the MCM to multicomponent kinetic analyses entails making several mixtures of known composition and recording their absorbances at different wavelengths and times. The wealth of information thus obtained must be reduced in order to abstract the analytically relevant information by using principal component analysis (PCA); also, calibration can be performed by using two similar MCMs such as PCR and partial least-squares regression (PLS). PLS breaks down both the measured data and the concentration calibration matrices into the product of two smaller matrices. This involves a data compression step where the measured absorbances are reduced to a small number of values called ''scores'' in a new coordinate system. The new axes are called ''principal components'' or ''factors''. PLS assumes a linear relationship between the concentrations and scores. The regression coefficients from each original variable to each principal component are called ''loadings''. Once the model has been fully developed, it can be used to determine analyte concentrations simply by using the data set for mixtures of unknown concentrations.

The effect of different variables (e.g. the influence of noise on both the analytical signal and the rate constants) on the quality of resolution of binary mixtures by PLS has also been assessed from simulated data, as has the potential of this algorithm for multicomponent kinetic determinations in nonlinear systems (e.g. in the presence of synergistic effects).[16] In contrast to curve-fitting and Kalman filtering approaches, these MCMs require no prior knowledge of rate constants or absorbance coefficients and entail no explicit adoption of a given kinetic model. Regarding the rate constant ratio for mixture components, because the resolution of the mixture relies on both spectral and kinetic discrimination, even mixtures of more than two components with near-unity ratios can be resolved. However, the rate constant ratio has a great influence on the mixture resolution in the presence of synergistic effects (it must be higher than three), as shown by applying PLS to simulated data.[16]

On the other hand, a procedure based on factor analysis was satisfactorily applied to kinetic data simulated from Gaussian and real spectra for Co(II) and Ni(II) using the above-mentioned ligand exchange reaction.[17]

Kalman filtering is a more effective approach when it is combined with spectral discrimination. Thus, the extended form of this algorithm has been fostered by combination with multiwavelength photometric determination, which is particularly effective for the simultaneous kinetic resolution of complex mixtures. Simulations have shown that mixtures can be successfully determined even if the rate constants of the two reactants are identical, as long as some spectral difference existed.[18] Quencer and Crouch[19] established the basis for the simultaneous determination of species from both their spectral and kinetic differences. Their approach allows resolution of more closely related species, thanks to the fast computation capabilities derived from the recursive nature of the Kalman filter. How large the differences need be to ensure accurate results depends on the number of wavelengths used for analysis; however, the larger the number of wavelengths used, the stronger is the overlap that can be tolerated. The accuracy of the determination is very good, even with a high degree of overlap and a few wavelengths; also, systems with completely overlapped kinetics can be resolved provided spectral differences exist at the chosen wavelength, and systems with completely overlapped spectra can be successfully determined as long as sufficient kinetic differences exist.[19] In the latter case, rate constants must differ by a factor of about 2.5–3 for acceptable results to be achieved using the extended Kalman filter.

5.2 Kinetic Wavelength-pair and H-Point Methods

Although the principle of these methods is different, both have been developed for the ultraviolet–visible spectroscopic kinetic analysis of two-component systems involving the selection of one or two wavelength pairs.

The *kinetic wavelength-pair* or *dual wavelength method*[20] involves measuring the difference in the rate of change of absorbance as a function of time at two preset wavelength pairs. The rate of change of absorbance, $u = \Delta A/\Delta t$, is dependent on the wavelength used and should not be confused with the chemical reaction rate, which is the same at any wavelength for a given concentration. The aim of this method is to enhance the accuracy in simultaneous multicomponent resolutions by increasing the differences between the absorptivity coefficient ratios of the components in the mixture at the preset wavelengths. A diode array spectrophotometer and a computer furnished with suitable software are required for application of this method.

For a mixture of two components M and N, it application involves measuring u at two wavelength pairs, (λ_1, λ_2) and (λ_3, λ_4) (for simplicity subscripts 1–4 in the expressions below denote measurements made at λ_1–λ_4 respectively), which are chosen in such a way that the difference between the rates of change of the absorbance for the first wavelength pair is as large as possible for one of the components (e.g. N), and the opposite holds for the second wavelength pair. The data set thus obtained allows determination of the concentrations of M and N in the mixture by using a single aliquot. Hence, if additivity of the rates of absorbance change at each wavelength is fulfilled, Equation (19) follows:

$$u_i = \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{M},i} + \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{N},i} \tag{19}$$

where i is the selected wavelength.

Taking into account the differences between wavelength pairs, the following systems of Equations (20) and (21) can be established:

$$u_2 - u_1 = \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{M},2} + \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{N},2} - \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{M},1} - \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{N},1} \tag{20}$$

$$u_4 - u_3 = \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{M},4} + \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{N},4} - \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{M},3} - \left(\frac{\Delta A}{\Delta t}\right)_{\mathrm{N},3} \tag{21}$$

where $u_2 > u_1$ and $u_4 > u_3$.

Parameter u is related to the chemical reaction rate, v, through Equation (22) and (23):

$$v_M = \frac{\Delta[M]}{\Delta t} = \frac{1}{\varepsilon_{M,i}} \left(\frac{\Delta A}{\Delta t} \right)_{M,i} = k_M[M] \quad (22)$$

$$v_N = \frac{\Delta[N]}{\Delta t} = \frac{1}{\varepsilon_{N,i}} \left(\frac{\Delta A}{\Delta t} \right)_{N,i} = k_N[N] \quad (23)$$

where k_M and k_N are the first-order rate constants and a path length of 1 cm is assumed. A combination of Equations (20), (21), (22) and (23) yields Equations (24) and (25):

$$u_{2-1} = u_2 - u_1 = k_M(\varepsilon_{M,2} - \varepsilon_{M,1})[M] + k_N(\varepsilon_{N,2} - \varepsilon_{N,1})[N] \quad (24)$$

$$u_{4-3} = u_4 - u_3 = k_M(\varepsilon_{M,4} - \varepsilon_{M,3})[M] + k_N(\varepsilon_{N,4} - \varepsilon_{N,3})[N] \quad (25)$$

which allows the concentrations of M and N to be simultaneously determined.

The four wavelengths used are selected and u calculated by computer. The conditions yielding the most accurate and sensitive estimate of M and N concentrations are those at which the differences between $(\varepsilon_{M,2} - \varepsilon_{M,1})/(\varepsilon_{N,2} - \varepsilon_{N,1})$ and $(\varepsilon_{M,4} - \varepsilon_{M,3})/(\varepsilon_{N,4} - \varepsilon_{N,3})$ are maximal and parameters $u_2 - u_1$ and $u_4 - u_3$ are as large as possible.

Considerable simplification is introduced when only one component contributes at each wavelength pair (e.g. $(\Delta A/\Delta t)_{N,2} = (\Delta A/\Delta t)_{N,1}$ and $(\Delta A/\Delta t)_{N,4} = (\Delta A/\Delta t)_{N,3}$). In that case, Equations (24) and (25) become Equations (26) and (27)

$$u_M = u_2 - u_1 = k_M(\varepsilon_{M,2} - \varepsilon_{M,1})[M] \quad (26)$$

$$u_N = u_4 - u_3 = k_N(\varepsilon_{N,4} - \varepsilon_{N,3})[N] \quad (27)$$

Hence u_M is independent of the concentration of N and u_N is independent of that of M.

The kinetic wavelength-pair method considerably enhances accuracy relative to some classical kinetic approaches (e.g. logarithmic extrapolation, single point and proportional equation methods), as shown in the simultaneous determination of formaldehyde and acrolein by reaction with 3-methylbenzodiazolin-2-one hydrazone.[20]

There are some advantages associated with the use of the kinetic wavelength-pair method. Thus, the resolution is based on kinetic and spectral differences. Therefore, the simultaneous kinetic determination of species with rate constant ratios close to unity is possible whenever enough spectral differences exist. On the other hand, application of the method requires no use of complex chemometric procedures and, in principle, there are no restrictions with regard to the number of components that can be determined. Its main drawback is a low sensitivity compared with methods based on peak wavelengths. The method has also been extended to the fixed-time kinetic approach[21] and has been applied to the resolution of binary mixtures of arsenate and phosphate,[21] *o*-cresol and *m*-cresol,[22] as well as to ternary mixtures of phenothiazines.[23]

The *H-point standard addition method* (HPSAM)[24] is identificated for two cases, namely the determination of two species of which only one evolves with time, and that of two species with overlapped time evolutions.

For a mixture of two components C_1 and C_2, the determination of the concentration of C_1 at equilibrium by the HPSAM entails selecting two wavelengths, λ_1 and λ_2, lying on each side of the absorption maximum of C_2 so that the absorbance of the latter component is the same at both wavelengths. Then, known amounts of C_1 are successively added to the mixture and the resulting absorbances at the two wavelengths are measured. The two straight lines thus obtained intersect at the so-called H-point $(-C_H, A_H)$, where $-C_H = (-C_{C_1})$ is the unknown concentration of C_1 and $A_H = (A_{C_2})$ is the analytical signal of C_2. When one of the species evolves with time, the variables to be fixed are two times, t_1 and t_2. Species C_2, which should not evolve with time over this range, should have constant absorbance. This contrasts with the equilibrium HPSAM, where two wavelengths are chosen. A plot is then constructed of added $[C_1]$ on the y-axis and ΔA for the two times on the x-axis. The point where $\Delta A = 0$ is equal to $-[C_1]$.

When both species evolve with time, $[C_1]$ can be calculated by plotting ΔA for the two times against the added concentration of C_1 at two wavelengths λ_1 and λ_2 provided the absorbances of C_2 at these two wavelengths are the same.

5.3 Miscellaneous Methods

The continuous addition of reagent (CAR) technique[25] is also applicable to multicomponent kinetic determinations. This method allows the resolution of mixtures of species basically thanks to the special way in which sample and reagents are mixed, discrimination being mainly of the instrumental type.

This method is based on the continuous addition of a reagent at a constant rate to the species to be determined.[25] The model for this method was developed for a single component but has readily been extended to multiple reacting species. If a solution of a reagent R of concentration $[R]_0$ is added at a constant rate u to a volume V_0 of a solution containing the analyte C, then the overall rate of the process will be given by Equation (28)

$$-\frac{d[C]}{dt} = k \left(\frac{u[R]_0}{V_0} \right)_{t[C]_0} \quad (28)$$

on the assumption that, under short reaction times, dilution of the analytes can be neglected. k is the second-order reaction rate. Equation (28) is used to construct a linear calibration graph by plotting the initial slopes of the kinetic curves as a function of $[C]_0$ over a fixed time interval in all experiments.

Two or more kinetic curves corresponding to as many components are sequentially obtained by recording the changes in the analytical signal as function of time; the respective slopes are proportional to the individual concentration of each component.[26] The reactions of mixtures whose components react very rapidly under pseudo first-order conditions are slowed on switching to pseudo second-order kinetic conditions, which allows the separation of the analytes on different timescales according to their reactivities and reaction rates. These reactions are independent of each other and are related to the characteristic parameters of the CAR technique shown above and to the concentration of each component through Equation (28), established from initial rate measurements. For a mixture of two components A and B these reaction rates are given by Equations (29) and (30)[26]

$$-\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = k_{\mathrm{A}}\left(\frac{u[\mathrm{R}]_0}{V_0} \quad t[\mathrm{A}]_0 \right. \tag{29}$$

$$-\frac{\mathrm{d}[\mathrm{B}]}{\mathrm{d}t} = k_{\mathrm{B}}\left(\frac{u[\mathrm{R}]_0}{V_0} \quad t[\mathrm{B}]_0 \right. \tag{30}$$

If $k_{\mathrm{A}} > k_{\mathrm{B}}$, only component A will react with the reagent, R, added from the autoburette; once component A has almost completely reacted, the addition of a fresh volume of R starts the reaction of component B and the concentrations of A and B are calculated from Equations (29) and (30), respectively. This method involves no complex mathematical treatment and can also be used to resolve multicomponent mixtures by using n equations in unknowns, provided appropriate values of adequate k_n/k_{n-1}, k_{n-1}/k_{n-2} and so on can be obtained in order to minimize errors.

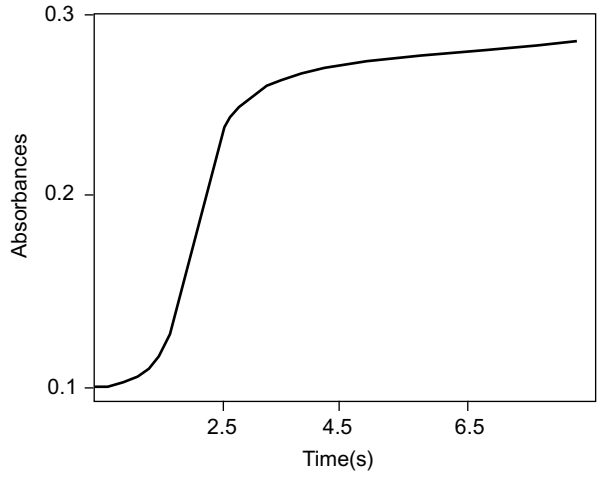

Figure 4 Absorbance–time curve provided by the CAR technique for a mixture containing $1.2\,\mu g\,mL^{-1}$ zineb and maneb. (Reproduced by permission of the Royal Society of Chemistry, from Quintero et al.[27])

Mixtures of metal ions such as copper(II) and iron(III),[26] and pesticides such as zineb and maneb,[27] have been resolved by using complex-formation reactions in both cases. Figure 4 shows the absorbance–time curve obtained for a mixture of zineb and maneb.

6 PHYSICOCHEMICAL DISCRIMINATION: MICELLAR CATALYSIS

Micellar catalysis can be a powerful tool for implementation of differential reaction rate methods. Micelles can alter the apparent rate constant ratio of two or more species that interact with a common reagent by both altering their intrinsic reactivity and, more generally, binding to a different extent to different analytes. According to the pseudophase model,[28] the observed second-order rate constant for a bimolecular reaction between A and B in the presence of micelles, k_{exp}, is given by Equation (31)

$$k_{\mathrm{exp}} = \frac{(k_{\mathrm{M}}P_{\mathrm{A}}P_{\mathrm{B}} + k'_{\mathrm{M}}P_{\mathrm{A}} + k''_{\mathrm{M}}P_{\mathrm{B}})CV + k_{\mathrm{W}}(1 - CV)}{(1 + K_{\mathrm{A}}C)(1 + K_{\mathrm{B}}C)} \tag{31}$$

where the subscripts M, w, A and B denote quantities related to the micellar phase, aqueous phase and reactants, respectively; P_{A} and P_{B} are the partition coefficients of the reactants (i.e. $P_{\mathrm{A}} = [\mathrm{A}]_{\mathrm{M}}/[\mathrm{A}]_{\mathrm{w}}$); C is the surfactant concentration (molarity) minus the critical micelle concentration; V is the molar volume of surfactant; the factors CV and $(1 - CV)$ are the volume fractions of the micellar and aqueous phase, respectively; and K_{A} and K_{B} are the binding constants of the reactants, viz. $K_{\mathrm{A}} = (P_{\mathrm{A}} - 1)V$ and $K_{\mathrm{B}} = (P_{\mathrm{B}} - 1)V$; k_{M} and k_{w} denote the rate constants of the reactions taking place in the micellar and aqueous phase, respectively; k'_{M} is the rate constant of the reaction resulting from collisions between reactant A in the micellar phase and reactant B in the aqueous phase, and k''_{M} is the rate constant of the reaction between reactant B in the micellar phase and reactant A in the aqueous phase.

The efficiency of micellar catalysis, a parameter of interest in kinetic methodology, will be given by the ratio $(k_{\mathrm{exp}}/k_{\mathrm{w}})$, which depends on two factors, Equation (32):

$$\frac{k_{\mathrm{exp}}}{k_{\mathrm{w}}} = \frac{k_{\mathrm{M}}}{k_{\mathrm{w}}}\frac{K_{\mathrm{A}}K_{\mathrm{B}}}{V(\sqrt{K_{\mathrm{A}}} + \sqrt{K_{\mathrm{B}}})^2} \tag{32}$$

The first factor $(k_{\mathrm{M}}/k_{\mathrm{w}})$ takes into account the changes in intrinsic reactivity when the reactants are transferred

from water into the micelle phase and hence the specific effect of the micelles on the transition state of the reaction. The second factor depends on the effective reactant concentration in the micellar pseudophase and is dependent on the binding constants of the reactants to the micelles (K_A, K_B) and the molar volume of the surfactant (V). Analysis of different reactions developed in the presence of micelles in the light of different models showed that the observed increase in the reaction rate was mainly caused by the reactant concentrating in a low volume in the micellar pseudo-phase since the values of k_M and k_w were similar.

For two analytes with different partition coefficients, the ratio between their rate constants will depend on the surfactant concentration. As a result, micellar aggregates might be useful for the resolution of mixtures of species with similar reactivities in aqueous media or in those cases where a given component takes too long to react and thus precludes analytical measurements.

There are many other features of micellar aggregates that can be of great value in multicomponent kinetic analysis, Thus, micelles have been used to alter reaction pathways, modify the mechanisms through which these reactions develop and alter product distribution in reactions where two or more products are generated. Consequently, micelles can alter the physicochemical features of the reactant or product of the reaction concerned and give rise to spectral shifts, changes in molar absorptivity and so on, which make them firm candidates for kinetic multicomponent analysis.

Micellar catalysis has been applied to the determination of binary mixtures of cyanide, sulfide and sulfite based on their reaction with 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB).[29] The anion mixture could not be resolved by conventional differential kinetic analysis in an aqueous medium because of the similar reactivity of sulfide and sulfite ions and the slowness of the reaction between cyanide and DTNB, which took ca. 100 min to complete. The dependences obtained by plotting the observed rate constants, k_{exp}, of the reactions among the three ions and DTNB as a function of the cetyltrimethylammonium bromide (CTAB) concentrations (Figure 5) showed that, by selecting an appropriate concentration of surfactant, adequate differences between the rate constants were obtained and binary mixtures of the anions at micromolar concentrations could be resolved by using a multiple linear regression method.

Micellar control of reaction pathways leading to specific products can be of great use in multicomponent kinetic analysis. One representative example of the potential of

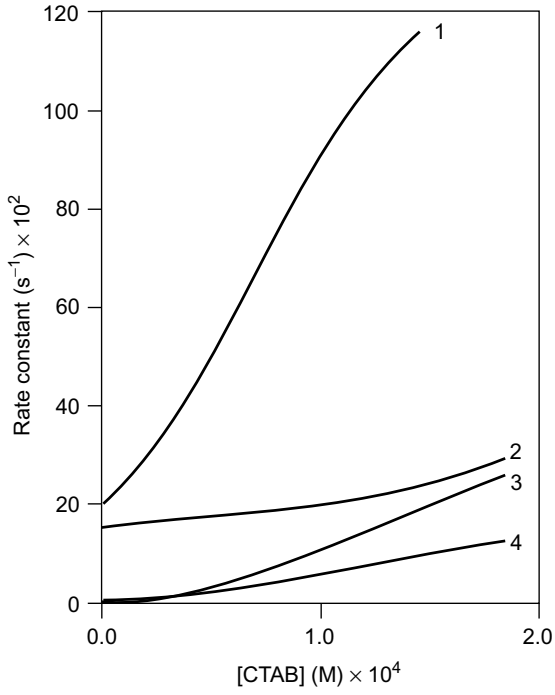

Figure 5 Dependence of the rate constants for the sulfide–DTNB, (1) sulfite–DTNB (2) and cyanide–DTNB (3, 4) systems on the CTAB concentration. Curves 1, 2, 4 obtained at pH 8; curve 3 obtained at pH 10. (Reproduced from González et al., *Anal. Chem.*, [29] by permission. Copyright (1993) American Chemical Society.)

this micellar effect is the simultaneous determination of V(V), Cr(VI) and Ti(IV) based on their reactions with pyrogallol red (PR) in the presence of dodecyltrimethylammonium bromide (DTAB) micelles.[30] In an aqueous medium, PR is gradually decolorized by oxidation with chromium(VI) and vanadium(V), so the simultaneous resolution of these two ions is rendered impossible by their similar reactivity. Titanium(IV) interferes with the individual determinations of these ions by forming a complex with PR. DTAB micelles have the following effects on the process: they catalyze the oxidation of PR by Cr(VI); they allow the oxidation reaction between V(V) and PR to be completely displaced by a complex-formation reaction between the two species, and they also alter the photometric features of the Ti(IV)–PR complex. These changes allow the simultaneous resolution of the three ions from measurements made at three different wavelengths corresponding to as many reaction products. The linear determination ranges thus achieved are between 0.2 and 1 μg mL^{-1} and the relative errors made in the resolution of the ternary mixture vary from 0.7 to 1.4%.

The use of micellar media can be a powerful tool for enhancing kinetic determinations and facilitating kinetic multicomponent analyses.

7 PRACTICAL IMPLEMENTATION: AUTOMATED METHODS

The conversion of differential kinetic methods of purely academic interest into practical methods for routine analysis entails a certain degree of automation. Electronic devices and computers have played a decisive role insofar as they have solved the problems associated with the measurement of time (i.e. especially in fast processes) by permitting automated acquisition of data during development of the process. Differential kinetic methods use instruments of varying degrees of complexity, from conventional elementary systems to sophisticated, completely automated instruments capable of collecting the sample, transporting it to the detector and processing the results obtained. This latter trend has been significantly fostered by the development of automated instrumentation for the computation of reaction rates such as the incorporation of SF mixing chambers into conventional photometric (including diode array) or fluorimetric detectors (for fluorescence or chemiluminescence measurements), the use of a simple automatic titrators and the implementation of continuous-flow techniques on simple manifolds, among others.[2,4,31,32]

Kinetic measurements are free from effects of factors that introduce errors in absolute values (e.g. turbidity, the presence of other absorbing or fluorescent substances provided they do not take part in the reaction of interest or alter the parameter response); as a result they are based on temporal charges in the parameter used to monitor the reaction (absorbance, fluorescence, intensity, potential) rather than on their absolute values. However, strict timing and temperature control (to within 0.01–0.1 °C) are essential for kinetic methods, which thus require modern powerful instrumentation. The choice of a given method for implementing rate measurements of a reaction is dictated by its half life. Thus, the instrumentation required to monitor slow reactions is typically simpler than that needed for fast reactions.

The process involved in measuring the rate of two or more components in a sample by using a common reagent comprises the following steps:

1. preparation, measurement, transport and mixing of reactants
2. signal monitoring and transducing at a constant temperature by measuring changes in a property of the reactants or products
3. timed acquisition of data (viz. collection of signals at different times) which are simultaneously or sequentially processed, usually with the aid of a computer.

Fully automated systems execute all three steps above, whereas partly automated instruments usually perform steps (2) and (3) only, and process data in a sequential rather than simultaneous manner.

In this section are briefly described currently available mixing modes and instrumental measurements such as flow techniques (low-pressure continuous and high-pressure SF techniques), used with single- or multi-wavelength photometric detection and luminescence detection, and the low-pressure CAR technique with photometric or luminescence detection, in addition to other approaches such as the T-optical fluorescence technique.

The increasing popularity of multicomponent kinetic methods is reflected in the variety of real analytical problems which are currently being addressed with their help.[4,31]

7.1 Mixing Modes and Instrumental Assemblies

The first step in the kinetic process, that is the mixing of sample and reagents, is common to other analytical methods but is a key step in kinetic methods as it involves measuring the parameter via which the process is monitored. The way in which reactants are mixed to start the analytical reaction depends on whether the external experimental conditions will remain unchanged (closed systems) or not (open systems). Closed systems are unsuitable for fast reactions while open systems are specially suited to reactions with fast kinetics. The sample must be thoroughly mixed with the reagent(s) at a constant temperature before it is transferred to the detection system for measurement. This can be done manually, which is usually the case with closed systems, or automatically. Manual mixing can be accomplished simply by homogenizing the reactants in the reaction vessel, which can be a spectrophotometric cuvette or an electrode cell. The instant at which the last reactant is added is taken as the reaction start time ($t = 0$) and must therefore be determined with great accuracy.

Automated mixing of sample and reagents calls for special mechanical/electronic devices when fast reactions, which as noted earlier require the use of open systems, are addressed. This increases the reproducibility and throughput of multicomponent kinetic analysis and is the reason why automated open systems are preferable even when slow reactions are involved.

The most frequently used open systems can be classified into continuous or flow systems and discrete systems. The most popular and widely employed techniques in connection with such systems for multicomponent kinetic determinations in routine analyses are flow injection and SF among the former and the CAR technique among discrete systems.

As a rule, the flow rate of the sample and/or reagents in the mixing system can be zero or constant while the kinetic curve is being recorded. The former alternative is the basis for SF systems; on the other hand, in continuous addition analyses, the flow rate of one of the reaction ingredients is nonzero and constant. Throughout the measurement interval, the reactants can be circulated at a given, identical flow rate in flow injection analyses.

SF is highly versatile and often superior to other flow techniques such as continuous-flow analysis (CFA), both in its air-segmented CFA and in its unsegmented mode (flow-injection analysis, FIA). Thus, SF can be applied to the resolution of mixtures following very fast kinetics with higher sample throughput and lower reagent consumption than those afforded by FIA or CFA.

The choice of a given technique will be dictated not only by its technical features, but also by the half lives of the reactions concerned as shown below.

7.2 High-pressure Stopped-flow Technique

This technique is the most commonly used for the resolution of mixtures involving fast reactions (namely those with half lives of a few milliseconds to a few seconds). In conventional SF systems, the sample and reagents are mixed at high pressure by means of two driving syringes that are actuated manually or automatically (e.g. by pneumatic device) which are inserted into a flow cell or mixing chamber that can also act as the observation cell. The flow is then abruptly stopped by using a third syringe and the analytical signal is recorded as a function of time. The system can be controlled by computer which can also be used for data acquisition and processing. The dead time of the instrument, which is determined by the mixing efficiency (viz. the mixing time) and the transport and stop times, must be smaller than the half life of the reaction of interest by about two orders of magnitude. The high cost of SF instruments (which has delayed its use in routine analyses) can be avoided by using commercially available SF mixing modules to fit to conventional and diode array photometric, fluorimetric and chemiluminescence detectors for implementation of the SF techniques. Figure 6 depicts a SF mixing module controlled by a simple computer interfaced on-line to it. The module is accomodated vertically above the detector.[33] This modular SF system is applicable to reactions with half lives in the millisecond range. Analytical results are rapidly obtained by using computer software for differential kinetic methods, this ensures high sample throughput, reproducibility and accuracy. The modular system has proved fairly efficient in the spectrophotometric or spectrofluorimetric determination of binary mixtures of metal ions, biogenic amines, drugs and pesticides.[4] Some of the ensuring methods are based on slow reactions, which also benefit from the simplicity, rapidity, precision and minimum sample and reagent manipulation of the SF technique.[4]

The above-described SF module was also incorporated into a commercial diode array spectrophotometer to resolve multicomponent mixtures by kinetic methodology. The ability to make absorbance measurements as a simultaneous function of time and wavelength is a powerful way of accomplishing kinetic multicomponent analyses, as shown in section 5. Two or more reactions involving the same or different reagents can be simultaneously monitored in this way. When the absorption spectra for the mixture components are sufficiently resolved (i.e. in the simplest case), kinetic data can be readily obtained at each absorption maximum without the need to apply differential kinetic methods based on mathematical approaches, provided the reaction rates of the components are independent of one another. Figure 7 shows the three-dimensional plot for a mixture of hydrazine and phenylhydrazine determined by their corresponding condensation reactions with *p*-dimethylaminobenzaldehyde (DAB), as well as their respective kinetic curves, which were computed simultaneously. These data allowed the individual and/or simultaneous determination of these compounds by applying a computerized initial-rate method.[34]

SF differential rate methods have also been applied to energy-transfer sensitized luminescence reactions based on the energy transfer from an excited organic compound analyte to europium or terbium ion which gives a characteristic line-type emission at 612 nm [Eu(III)] or 545 nm

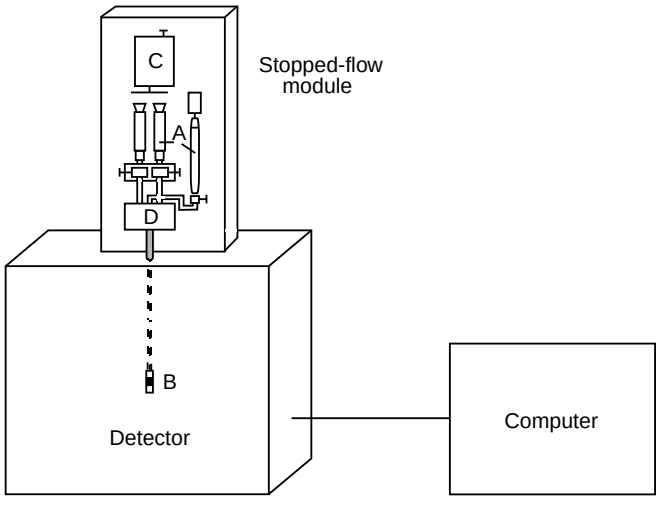

Figure 6 Schematic diagram of a modular SF system. A, Driving and stopping syringes; B, mixing-observation chamber; C, pneumatic propulsion system; D, thermostatted chamber. (Reproduced by permission of the Royal Society of Chemistry, from Pérez-Bendito.[4])

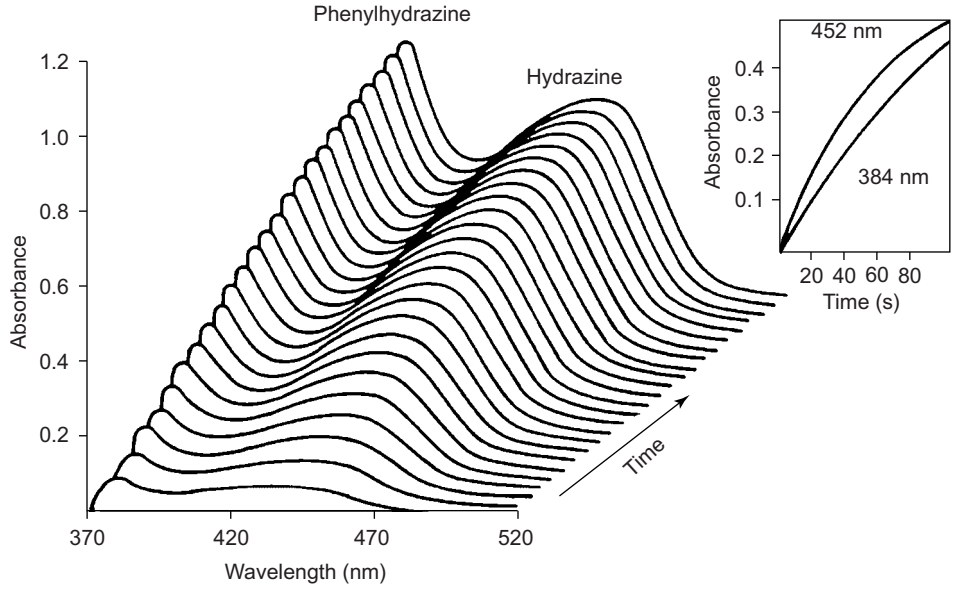

Figure 7 Temporal variation of the absorption spectrum for the phenylhydrazine–hydrazine–DAB system. λ_{max} of phenylhydrazine = 452 nm, λ_{max} of hydrazine = 384 nm, $T = 30$ s. (Reproduced by permission of the Royal Society of Chemistry, from Pérez-Bendito.[41])

[Tb(III)]. These methods rely on the different rates of formation of a lanthanide–analyte–ligand ternary complex in the presence of a surfactant, which is very fast and require the use of the SF technique. Tetracycline–doxycyclines in mixtures were simultaneously determined in this way, in untreated serum samples using the proportional equation method. The kinetic curves were obtained by placing Eu(III) and thenoyltrifluoroacetone (TTA) in one of the syringes and tetracycline plus Triton X-100 in the other.[35] A similar approach was also developed for the simultaneous determination of salicylic acid and diflunisal with terbium(III)–EDTA (*N*,*N*,*N*′,*N*′-ethylenediaminetetraacetic acid) and CTAB in a basic medium.[36]

7.3 Low-pressure Continuous Flow Techniques

The unsegmented continuous-flow technique, so also called ''FIA''[37] is one other automated approach for implementing the differential reaction rate method. Although it is unsuitable for fast reactions (mixing is not efficient enough owing to the low pressure in the merging streams and dispersion decreases the sensitivity) it has been used in multicomponent kinetic determinations by virtue of the very simple instrumental set-up required. Typically, this consists of a peristaltic pump for propelling a carrier, into which the sample is injected by means of an injection valve, and one or more reagent streams that merge at a mixing point and then pass through a reaction coil of appropriate length before they reach a flow-cell placed in the detector. The analytical signal is a sharp peak the height of which is directly related to the analyte concentration. The degree of complexity of this configuration depends on whether it includes one or several detectors, simple or multiple injection, whether or not it involves splitting or stopping the flow and so on.

Flow-injection differential reaction rate methods are essentially based on the formation of two reaction zones reacting in the detector at different times, the use of multiple detection points separated by delay coils or the iterative passage of the reacting plug by the same detection point. Depending on the type of data obtained, there are two basic types of flow-injection differential reaction-rate methods, that is those based on two signals measured at different reaction times and those relying on multipoint signal–time curves. From a practical point of view, the most outstanding differential reaction-rate methods implemented in simple flow manifolds rely on the formation of two reaction zones based on the differential contribution of each analyte to each peak resulting from differences in reaction rate. The mixture is resolved by applying the classical proportional equation method (section 3.1). One example is the analysis of mixtures of cobalt and nickel based on the formation of their complexes with 2-hydroxybenzaldehyde thiosemicarbazone.[38]

The other type of method involves the iterative passage of the reacting plug through a nondestructive detector in order to obtain multipeak recordings that can be used for kinetic purposes since the envelopes of their maxima and minima resemble typical kinetic curves.[39]

7.4 Low-pressure Continuous Addition of Reagent Technique

The special way in which the sample and reagents are mixed in this technique affords the resolution of mixtures by direct kinetic methods. The description and theoretical formulation of this technique are given in section 5.3. The experimental set-up for implementation of the CAR technique is shown in Figure 8. A typical CAR assembly consists of (1) an addition unit composed of an autoburette, a magnetic stirrer and a thermostatted vessel as reaction cell, (2) a photometric/fluorimetric or chemiluminescence detection unit that can includes an immersion probe for photometric measurements and (3) a computer. The reaction vessel can be of variable capacity from 60 mL to about 2 mL and typically holds sample volumes in the microliter range. This makes the technique especially useful when the amount of sample available is limited (e.g. with serum samples).

The CAR technique is a major alternative to the SF technique in tackling direct rate measurements of reactions with half lives of a few milliseconds, each component of the mixture is determined in a sequential manner in a single run of sample. Typical examples are the photometric determination of binary mixtures of carbamate pesticides (see Figure 4) in commercial formulations,[27] and of copper and iron in serum samples by use of a miniaturized reaction chamber[26] and photometric detection.

The CAR technique has also been used for the resolution of mixtures involving chemiluminescent reactions by using peroxyoxalate energy-transfer processes. 2,4-Trichlorophenylperoxyoxalate (TCPO) was added from the autoburette to the sample (a mixture of trimeprazine and methotrimeprazine) in excess hydrogen peroxide in the reaction vessel.[40] A single kinetic curve (sum of the two reactions) was obtained and the kinetometric ANN approach (see section 4.3) was used to resolve the mixture, the components of which exhibited a similar kinetic behavior and synergism. The CAR technique suppresses the background emission associated with this type of chemiluminescence reaction, thereby increasing the signal-to-noise ratio.

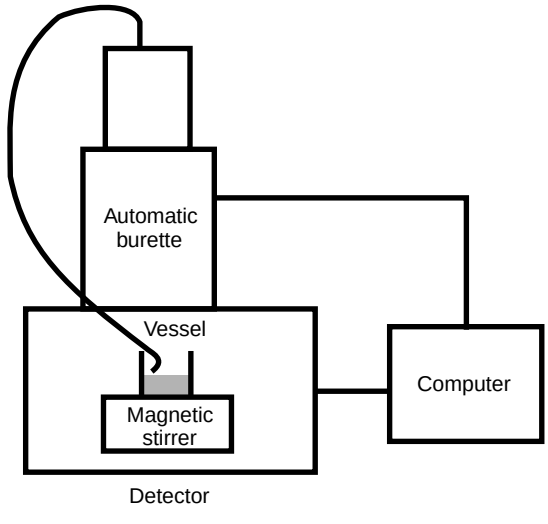

Figure 8 Experimental set-up for implementation of the CAR technique.

7.5 Miscellaneous Instrumental Approaches

Kinetic fluorimetric methods for the simultaneous resolution of mixtures usually involve chemical reactions that must meet several requirements when a conventional L-format spectrofluorimeter is used, namely: (1) similar maximum excitation and emission wavelengths for their reaction products and (2) differential kinetic behavior. In addition, if the reaction rate of any of the systems involved is very fast, the reactants cannot be mixed manually batchwise since the reaction is finished by the time the instrument starts acquiring the kinetic curve. However, the use of a spectrofluorimeter in a T-format configuration (one with two emission paths symmetrically arranged on both sides of the sample compartment), which is commercially available, allows two chemical systems with similar excitation wavalengths but different emission wavelengths to be monitored. The absence of overlap in the emission maxima allows kinetic data to be obtained from each individual system, even if both possess similar rate constants.

When the chemical system involves fast reactions, the combined use of one such spectrofluorimeter with the SF technique allows kinetic data to be obtained and provides a means for automation and rapid handling of reagents, both of which are highly desirable in routine analyses. Figure 9 shows a scheme of this approach, which provides ''instrumental discrimination'' for the resolution of neomycin and promethazine using two different reactions, namely *o*-phthaldialdehyde (OPA) in the presence of *N*-acetylcysteine to yield a fluorescent substituted isoindole by its condensation with neomycin and the oxidation of promethazine to a fluorescent sulfoxide by dissolved oxygen.[41] The sample solution, in a borate buffer medium, was placed in one of the syringes, the other being filled with an aqueous solution containing OPA, *N*-acetylcysteine and borate buffer. The common excitation wavelength used was 335 nm; the oxidation of promethazine was monitored on channel B at 370 nm and the condensation of neomycin on channel A at 460 nm. The respective kinetic data were processed by a microcomputer and the initial rate method was used to determine each component in the mixture.

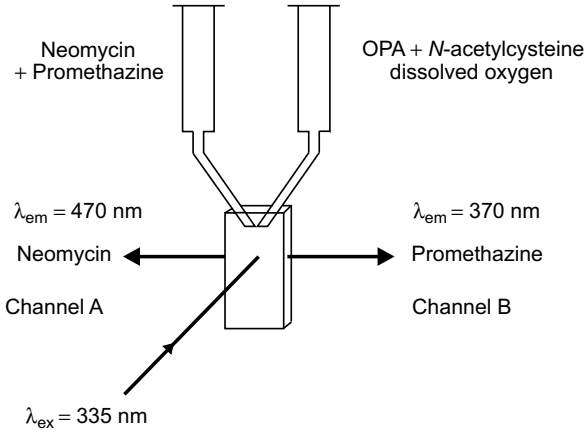

Figure 9 Scheme showing the SF/T-format spectrofluorimetric approach.

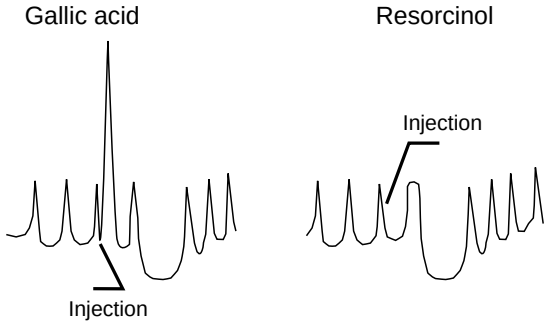

Figure 11 Pt-electrode potential–time dependence for gallic acid and resorcinol perturbation on the oscillating reaction for 0.55 μmol of gallic acid and 6.0 μmol of resorcinol. (Reproduced from Jiménez-Prieto et al., *Anal. Chim. Acta.*[43] Copyright (1996) by permission of Elsevier Science.)

The introduction of the analyte pulse perturbation (APP) technique[42] has opened up new prospects for differential rate determinations based on oscillatory reactions, which are a special type of far-from-equilibrium dynamic system. This technique allows the steady state of an oscillating reaction to be rapidly restored after each perturbation with a microvolume of the analyte added to a continuous-flow stirred tank reactor (CSTR); this affords many determinations on the same oscillating system in a simple rapid manner. Thus, two or more different responses corresponding to as many components are sequentially obtained in the oscillating system; the respective perturbation (i.e. change in period or signal amplitude of the oscillating curves) is proportional to the individual concentration of each component.[43] This approach is very simple since it requires none of the complex mathematical treatments used in kinetometric approaches. Figure 10 shows a schematic depiction of the experimental set-up required for the resolution of gallic acid/resorcinol mixtures[43] based on their perturbation (Figure 11) on the copper(II)-catalyzed oscillating reaction between hydrogen peroxide and sodium thiocyanate, using the

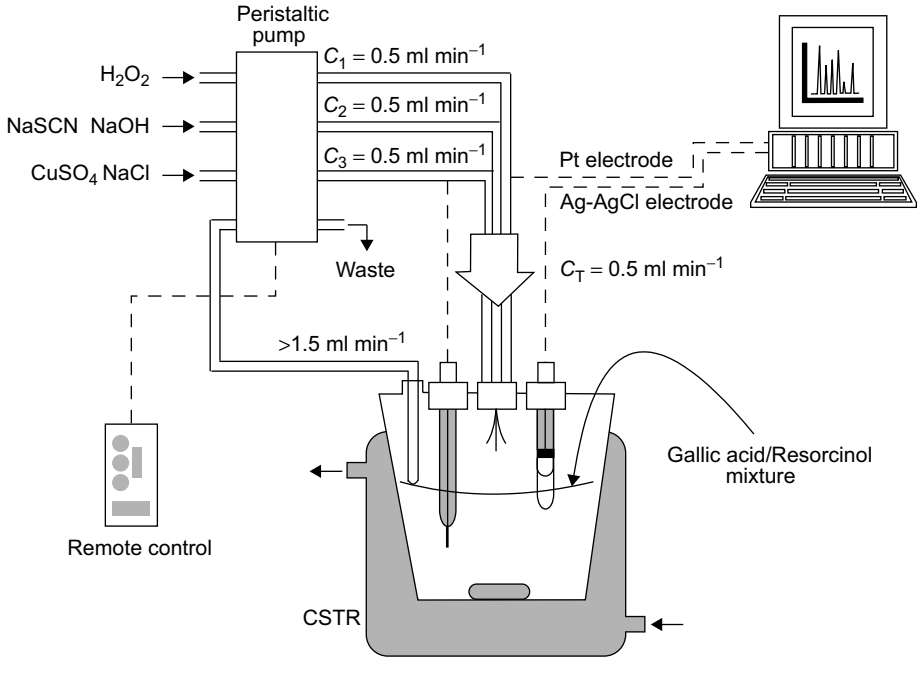

Figure 10 Schematic depiction of the experimental set-up used for the resolution of gallic acid/resorcinol mixtures by the APP technique. (Reproduced from Jiménez-Prieto et al., *Anal. Chim. Acta.*[43] Copyright (1996) by permission of Elsevier Science.)

APP technique as implemented in a thermostatted CSTR, potentiometric detection, a peristaltic pump to dispense the reagent continuously and keep the volume in the CSTR constant, and a computer to acquire and process data.

8 MULTICOMPONENT KINETIC DETERMINATIONS AS PROBLEM SOLVERS

One of the best indicators of the consolidation of differential reaction-rate methods is their increasing use for solving real problems.[2,4]

Thus speciation has also been addressed by using these methods. Distinguishing between the forms (free or bound) or oxidation states in which a given species can occur in a sample is possible by applying differential rate methods. Fe(III) and Fe(II) can be simultaneously determined in iron-rich groundwater through the aerial oxidation of Fe(II) in the presence of Tiron and acetate ion; the Fe(III) formed is subsequently complexed with Tiron and the reaction is monitored spectrophotometrically. The concentrations of Fe(III) and Fe(II) are obtained from the initial and equilibrium absorbance values, respectively, which are calculated by nonlinear least-squares fitting (without the need for any prior rate constant measurement).[44] The same iron ions were determined in wines by the formation of the Fe(III)–thiocyanate complex, using the SF technique to process the two samples needed for analysis.[33] In the speciation of Sb(III) and Sb(V), the differential rate of reaction of the two ions with mandelic acid enabled their discrimination.[45] The different forms in which a metal ion can occur in a given medium such as water or soil (e.g. free, complexed or bound to organic molecules) can be discriminated by adapting conventional differential reaction-rate methods to the analysis of mixtures for which neither the number of components nor their rate constants is known. Such is the case with the kinetic speciation of Ni(II) bound to fulvic acid in a soil.[46]

Differential kinetic analyses have also proved useful in various analytical areas of interest such as environmental and clinical chemistry for the determination of organic and inorganic species in mixtures. Mixtures of carbamate pesticides (e.g. carbaryl + 1-naphthol, carbaryl + 1-naphthol + 2-naphthol and carbofuran + propoxur) in different types of water samples were resolved with satisfactory results by the SF technique.[4] The micellar catalysis approach described in section 6 was also applied to the simultaneous kinetic determination of arsenate and phosphate in water samples.[47] Mixtures of epinephrine and norepinephrine in urine[48] were resolved by the automatic method described in section 7.2. Mixtures of phenothiazines in human serum samples,[23] and *o*-cresol and *m*-cresol in urine samples,[22] were resolved based on their respective oxidative coupling reactions, which exhibit strong kinetic and spectral overlap, using the kinetic wavelength pair-method described in section 5.2. These are but a few examples of the uses of differential rate methods with clinical samples that involve little or no sample treatment.

Pharmaceutical analysis has also benefited from differential rate methods.[49] Thus, psychotropic drugs such as epinephrine and norepinephrine, chlorpromazine and perphenazine, and thiamine and cocarboxylase, among others, have been determined in different pharmaceutical preparations by using (mainly) proportional equation methods. Differential rate methods have also been applied to food analysis. Thus, the simultaneous determination of ampicillin and tetracycline in milk[50] was accomplished by using the SF/T-format spectrofluorimeter described in section 7.5, as was the resolution of mixtures of the antioxidants butylated hydroxyanisole and propyl gallate in commercial food samples (chicken and vegetable group, corn, sunflower, olive oils, and margarine) using SF diode array detection.[51]

Most of these methods feature high sampling rates and have been promoted by the inception of the above-described automatic techniques. The expeditiousness with which samples can be processed and the ease with which the data produced can be treated have helped popularize multicomponent analytes by differential rate methods.

ACKNOWLEDGMENT

The author gratefully acknowledges financial support from Spain's DGICyT (Project PB96-0984).

ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Network
APP	Analyte Pulse Perturbation
CAR	Continuous Addition of Reagent
CFA	Continuous-flow Analysis
CNN	Computational Neural Network
CSTR	Continuous-flow Stirred Tank Reactor
CTAB	Cetyltrimethylammonium bromide
DAB	*p*-Dimethylaminobenzaldehyde
DTAB	Dodecyltrimethylammonium bromide
DTNB	5,5′-Dithiobis(2-nitrobenzoic acid)
EDTA	*N*,*N*,*N*′,*N*′-Ethylenediaminetetraacetic Acid

EGTA	Ethylene Glycol bis(2-aminoethyl ether *N*,*N*,*N*′,*N*′-tetraacetic acid)
FIA	Flow-injection Analysis
HPSAM	H-Point Standard Addition Method
MCM	Multivariate Calibration Method
NLR	Nonlinear Least-squares Regression
OPA	*o*-Phthaldialdehyde
PAR	4-(2-Piridylazo)resorcinol
PCA	Principal Component Analysis
PCR	Principal Components Regression
PLS	Partial Least-squares Regression
PR	Pyrogallol Red
PTSC	Pyridoxal Thiosemicarbazone
SEP	Relative Standard Error of Prediction
SF	Stopped-flow
TCPO	Trichlorophenylperoxyoxalate
TTA	Thenoyltrifluoroacetone
2-ChPh	2-Chlorophenol
3-ChPh	3-Chlorophenol

RELATED ARTICLES

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Catalytic Kinetic Determinations: Nonenzymatic • Data Treatment and Error Analysis in Kinetics • Electrocatalysis-based Kinetics Determinations • Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics • Luminescence-based Kinetic Determinations • Uncatalyzed Kinetic Determinations

REFERENCES

1. IUPAC, 'Nomenclature of Kinetic Methods of Analysis', *Pure Appl. Chem.*, **65**, 2291–2298 (1993).
2. D. Pérez-Bendito, M. Silva, *Kinetic Methods in Analytical Chemistry*, Ellis Horwood, Chichester, 1988.
3. H.A. Mottola, *Some Kinetic Aspects of Analytical Chemistry*, Wiley, New York, 1988.
4. D. Pérez-Bendito, 'Approaches to Differential Reaction-rate Methods', *Analyst*, **115**, 689–698 (1990).
5. B.M. Quencer, S.R. Crouch, 'Multicomponent Kinetic Methods', *Crit. Rev. Anal. Chem.*, **24**(3), 243–262 (1993).
6. A. Loriguillo, M. Silva, D. Pérez-Bendito, 'Interpolation Method for Treatment of Mutual Kinetic Effect', *Anal. Chim. Acta*, **212**, 223–234 (1988).
7. M.D. Love, M.L. Pardue, 'Systematic Comparison of Data-processing Options for Kinetic-based Single-component Determinations of Non-catalysts. Part 1. Review, Systematic Classification, Mathematical Descriptions, Performance, Characteristics and Perspectives', *Anal. Chim. Acta*, **299**, 195–208 (1994).
8. D. Pérez-Bendito, M. Silva, 'Recent Advances in Kinetometrics', *Trends Anal. Chem.*, **15**, 232–240 (1996).
9. H.L. Pardue, 'Kinetic Aspect of Analytical Chemistry', *Anal. Chim. Acta*, **216**, 69–107 (1989).
10. G.M. Ridder, D.W. Margerum, 'Simultaneous Kinetic and Spectral Analysis with a Vidicon Rapid-scanning Stopped-flow Spectrophotometers', *Anal. Chem.*, **49**, 2098–2108 (1977).
11. I. Schechter, 'Simultaneous Determination of Mixtures by Kinetic Analysis of General-order Reactions', *Anal. Chem.*, **64**, 729–737 (1992).
12. R. Xiong, A. Velasco, M. Silva, D. Pérez-Bendito, 'Performance of the Kalman Filter Algorithm in Differential-rate Methods', *Anal. Chim. Acta*, **251**, 313–319 (1991).
13. R. Jiménez-Prieto, A. Velasco, M. Silva, D. Pérez-Bendito, 'Kalman Filtering of Data From First and Second-order Kinetics', *Talanta*, **40**, 1731–1739 (1993).
14. S. Ventura, M. Silva, D. Pérez-Bendito, C. Hervás, 'Multicomponent Kinetic Determinations Using Artificial Neural Networks', *Anal. Chem.*, **67**, 4458–4461 (1995).
15. A. Cladera, E. Gómez, J.M. Estela, V. Cerdá, J.L. Cerdá, 'Computer Method for the Simultaneous Kinetic Determination of Compounds in Mixtures Based on the Use of Diode Array Spectrophotometry', *Anal. Chim, Acta*, **272**, 339–344 (1993).
16. M. Blanco, Y. Coello, H. Iturriaga, S. Maspoch, J. Riba, E. Rovira, 'Kinetic Spectrophotometric Determination of Ga(III)–Al(III) Mixtures by Stopped-flow Injection Analysis Using Principal Component Regression', *Talanta*, **40**, 261–267 (1993).
17. A. Cladera, E. Gómez, J.M. Estela, V. Cerdá, 'Resolution of Simultaneous Kinetic Spectrophotometric Process by Factor Analysis', *Anal. Chem.*, **65**, 707–715 (1993).
18. B.M. Quencer, S.R. Crouch, 'The Extended Kalman Filter for Multiwavelength Multicomponent Kinetic Determinations', *Analyst*, **118**, 695–701 (1993).
19. B.M. Quencer, S.R. Crouch, 'Multicomponent Kinetic Determination of Lanthanides with Stopped-flow, Diode Array Spectrophotometry and the Extended Kalman Filter', *Anal. Chem.*, **66**, 458–463 (1994).
20. J.M. Peña, S. Rubio, D. Pérez-Bendito, 'Kinetic Wavelength-pair Method as an Alternative Approach to Simultaneous Multicomponent Analysis', *Anal. Chim. Acta*, **244**, 81–88 (1991).
21. M. López-Carreto, D. Sicilia, S. Rubio, D. Pérez-Bendito, 'Simultaneous Determination of Arsenate and Phosphate by Use of Kinetic Wavelength pair Method', *Anal. Chim. Acta*, **283**, 481–488 (1993).
22. M. López-Carreto, L. Lunar, S. Rubio, D. Pérez-Bendito, 'Simultaneous Spectrophotometric Determination of *o*-cresol and *m*-cresol in Urine by Use of the Kinetic Wavelength Pair Method', *Analyst*, **121**, 1647–1652 (1996).

23. M. López-Carreto, L. Lunar, S. Rubio, D. Pérez-Bendito, 'Simultaneous Spectrophotometric Determination of Chlorpromazine, Perphenazine and Acetopromazine by Use of the Kinetic Wavelength Pair-method', *Anal. Chim. Acta*, **349**, 33–42 (1997).
24. F. Bosch-Reig, P. Campins-Falcó, A. Sevillano-Cabeza, R. Hernáez-Hernández, C. Molins-Legua, 'Development of the H-point Standard Additions Method for Ultraviolet–Visible Spectroscopic Kinetic Analysis of Two-component Analysis Systems', *Anal. Chem.*, **63**, 2424–2429 (1991).
25. M. Márquez, M. Silva, D. Pérez-Bendito, 'Direct Rate Measurements of Millisecond Reactions by Continuous Addition of Reagent for Routine Analysis', *Analyst*, **113**, 1733–1736 (1988).
26. M. Márquez, M. Silva, D. Pérez-Bendito, 'Continuous Addition of Reagent Technique: A New Approach to Differential Reaction-rate Methods', *Anal. Chim. Acta*, **239**, 221–227 (1990).
27. M.C. Quintero, M. Silva, D. Pérez-Bendito, 'Simultaneous Kinetic Determination of Zineb and Maneb by the Continuous Addition of Reagent Technique', *Analyst*, **115**, 1261–1264 (1990).
28. M. López-Carreto, S. Rubio, D. Pérez-Bendito, 'Organic Microheterogeneous Systems in Kinetic Analysis. Self-assembled Systems. A Review', *Analyst*, **121**, 33R–44R (1996).
29. V. González, B. Moreno-Cordero, D. Sicilia, S. Rubio, D. Pérez-Bendito, 'Micellar Catalysis in Kinetic Multicomponent Analysis: Simultaneous Determination of Binary Mixture of Cyanide, Sulfide and Sulfite Ions', *Anal. Chem.*, **65**, 1897–1902 (1993).
30. D. Sicilia, S. Rubio, D. Pérez-Bendito, 'Micellar Effects on Reaction Kinetics. Part I. Simultaneous Determination Chromium(VI), Vanadium (V) and Titanium (IV)', *Anal. Chim. Acta*, **284**, 149–157 (1993).
31. D. Pérez-Bendito, M. Silva, A. Gómez-Hens, 'Automated Kinetic-based Determinations for Routine Analysis: Recent Developments', *Trends Anal. Chem.*, **8**, 302–308 (1989).
32. M. Silva, 'Recent Strategies in Automated Reaction-rate Based Determinations', *Analyst*, **118**, 681–688 (1993).
33. A. Loriguillo, M. Silva, D. Pérez-Bendito, 'Versatile Automatic Stopped-flow System for Routine Analysis', *Anal. Chim. Acta*, **199**, 29–40 (1987).
34. M.C. Gutiérrez, A. Gómez-Hens, D. Pérez-Bendito, 'Modular Stopped-flow Diode-array Detection System for Simultaneous Kinetic Analysis', *Anal. Chim. Acta*, **225**, 115–122 (1989).
35. P. Izquierdo, A. Gómez-Hens, D. Pérez-Bendito, 'Simultaneous Stopped-flow Determination of Tetracycline and Doxycycline in Serum Based on Lanthanide-sensitized Luminescence', *Anal. Lett.*, **27**, 2303–2316 (1994).
36. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Determination of Salicylic Acid, Diflumisal and Their Mixtures Based on Lanthanide-sensitized Luminescence', *Anal. Chim. Acta*, **329**, 135–141 (1996).
37. M. Valcárcel, M.D. Luque de Castro, *Flow Injection Analysis. Principles and Applications*, Ellis Horwood, Chichester, 1987.
38. A. Fernández, M.D. Luque de Castro, M. Valcárcel, 'Formation of Two Reaction Zones in Flow-injection Systems for Kinetic Determinations of Cobalt and Nickel', *Anal. Chim. Acta*, **193**, 107–118 (1997).
39. A. Ríos, M.D. Luque de Castro, M. Valcárcel, 'Analytical Potential of Flow-reversal Injection Analysis', *Anal. Chem.*, **60**, 1540–1545 (1988).
40. C. Hervás, S. Ventura, M. Silva, D. Pérez-Bendito, 'Computational Neural Network for Resolving Non-linear Multicomponent Systems Based on Chemiluminescence-methods', *J. Chem. Inf. Comput. Sci.*, **38**, 1119–1124 (1998).
41. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Use of a Stopped-flow/T-format Spectrofluorimeter for Simultaneous Kinetic Analysis', *Anal. Chim. Acta*, **310**, 453–459 (1995).
42. R. Jiménez-Prieto, M. Silva, D. Pérez-Bendito, 'Analyte Pulse Perturbation Techniques: A Tool for Analytical Determinations in Far-from-equilibrium Dynamic Systems', *Anal. Chem.*, **67**, 729–734 (1995).
43. R. Jiménez-Prieto, M. Silva, D. Pérez-Bendito, 'Simultaneous Determination of Gallic Acid and Resorcinol Based on an Oscillating Chemical Reaction by the Analyte Pulse Perturbation Technique', *Anal. Chim. Acta*, **334**, 323–330 (1996).
44. S. Abe, T. Saito, M. Suda, 'Simultaneous Determination of Iron(II) and Iron (III) in Aqueous Solution by Kinetic Spectrophotometry with Tiron', *Anal. Chim. Acta*, **181**, 203–209 (1986).
45. S. Sato, 'Differential Determination of Antimony(III) and Antimony(V) by Solvent Extraction–Spectrophotometry with Mandelic and Malachite Green, Based on the Difference in Reaction Rates', *Talanta*, **32**, 341–344 (1985).
46. J.A. Lavigne, C.H. Longford, M.K.S. Mak, 'Kinetic Study of Speciation of Nickel(III) Bound to a Fulvic Acid', *Anal. Chem.*, **59**, 2616–2620 (1987).
47. M. López, S. Lunar, S. Rubio, D. Pérez-Bendito, 'Application of Micellar Catalysis to the Simultaneous Kinetic Determination of Arsenate and Phosphate', *Quim. Anal. (Spain)*, **15**, 283–291 (1996).
48. M.P. Llavero, S. Rubio, A. Gómez-Hens, D. Pérez-Bendito, 'Improved Hydroxyindole Method for the Simultaneous Stopped-flow Spectrofluorimetric Determination of Epinephrine and Norepinephrine in Urine', *Anal. Chim. Acta*, **229**, 27–33 (1990).
49. D. Pérez-Bendito, A. Gómez-Hens, M. Silva, 'Advances in Drug Analysis by Kinetic Method', *J. Pharm. Biomed. Anal.*, **14**, 917–930 (1996).
50. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Simultaneous Determinations of Ampicillin and Tetracycline in

Milk by Using a Stopped-flow/T-format Spectrofluorimeter', *Talanta*, **44**, 1883–1889 (1997).
51. M.P. Aguilar-Caballos, A. Gómez-Hens, D. Pérez-Bendito, 'Simultaneous Kinetic Determination of Butylated Hydroxyanisole and Propyl Gallate by Coupling-flow Mixing Technique and Diode-array Detection', *Anal. Chim. Acta*, **354**, 173–179 (1997).

Electrocatalysis-based Kinetics Determinations

Florinel Gabriel Bănică
Norwegian University of Science and Technology, Trondheim, Norway

Ana Ion
Politehnica University of Bucharest, Bucharest, Romania

1 Introduction **1**
2 Electrocatalysis at Bare Electrodes **2**
2.1 Redox-type Catalytic Processes 2
2.2 Indirect Electrocatalytic Methods 12
3 Electrocatalysis at Chemically Modified Electrodes **15**
3.1 General Aspects 15
3.2 The Modifier as a Thin Layer 16
3.3 Electrocatalysis at Redox Polymer Modified Electrodes 16
3.4 Electrodes Modified by Inorganic Films 17
3.5 Carbon Paste Electrodes and Other Composite Electrodes 17
3.6 Electrocatalysis at Electronically Conducting Polymer Electrodes 17
3.7 Guidelines for Analytical Applications of Electrocatalysis at Chemically Modified Electrodes 18
4 Amperometric Enzymatic Determinations **18**
4.1 Background 18
4.2 Main Classes of Enzymes Used for Amperometric Sensors 19
4.3 Configuration of the Enzymatic Amperometric Sensor 20
4.4 Immobilization Methods 20
4.5 Redox Mediators for Amperometric Enzymatic Sensors 21
4.6 Sensor Response: Modeling Enzyme Electrode Processes 21
4.7 Examples of Applications 21
5 Conclusions **22**
List of Symbols **22**
Abbreviations and Acronyms **23**
Related Articles **23**
References **24**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The term ''electrocatalysis'' denotes the catalysis of an electrochemical reaction. In accord with the general definition of catalysis[1] the effect of an electrocatalyst is to increase the rate of an electrochemical reaction without modifying the overall standard Gibbs energy change in the reaction. Taking the view of Faraday's law, the rate of an electrochemical reaction is expressed by the electrolytic current. Consequently, an increase in the electrolytic current is the main result of electrocatalysis. The benefits of electrocatalysis include making possible the quantitation of the catalyst itself or of some substrate compounds that are not electrochemically active or are not generally amenable to direct electrochemical determination. An appreciable improvement in selectivity may be brought about by electrocatalysis. Additionally, appropriate engineering of catalytic layers on electrode surfaces has led to the preparation of a huge variety of chemical and biochemical sensors. However, close control of the experimental conditions (temperature, electrolyte composition) is often required when dealing with electrocatalytic processes. Usually, such requirements do not give rise to any particular difficulties.

1 INTRODUCTION

In contrast to conventional chemical reactions, the activation energy of an electrochemical reaction depends not only on the molecular/chemical factors (included in the so-called thermal term) but also on the electrode potential. That is why the rate constants for electrochemical reactions are exponential functions of the electrode potential. For the most simple electrochemical reaction (Equation 1)

$$\mathrm{Ox} + \mathrm{e} \rightleftharpoons \mathrm{Red} \tag{1}$$

the current–potential relationship is expressed by the Butler–Volmer relationship (Equation 2):[2]

$$i = i_0 \left\{ \exp\left(\frac{-\alpha F \eta}{RT} \right) - \exp\left(\frac{(1-\alpha) F \eta}{RT} \right) \right\} \tag{2}$$

where i_0 is the exchange current, α is the transfer coefficient ($\alpha \leq 1$), F is the Faraday constant, T is the absolute temperature and R is the gas constant. η denotes the overvoltage, i.e. the deviation of the actual electrode potential E from the equilibrium value E_e (Equation 3):

$$\eta = E - E_e \tag{3}$$

The expression for the exchange current (Equation 4) includes the standard rate constant for the electrode reaction, k_s

$$i_0 = FAk_s c_{\mathrm{Ox}}^{(1-\alpha)} c_{\mathrm{Red}}^{\alpha} \tag{4}$$

where A is electrode area, and c_{Ox}, and c_{Red} stand for the bulk (equilibrium) concentrations. According to Equation (2), the absolute value of the current rises exponentially with the increase in the absolute value of the overvoltage. In practise, a nonelectrochemical step always limits the current increase to a value that is termed the limiting current. Under such circumstances, the rate-determining step can be either the transport of the reactant towards the electrode surface or a chemical reaction producing the electrochemically active species. In the last case, the term ''kinetic current'' is used to denote this special type of limiting current.

The effect of electrocatalysis is an increase in the standard rate constant, k_s, which is followed by a rise in current value at a constant overvoltage. Because this effect can be masked by the intervention of a nonelectrochemical rate-determining step, the most straightforward mark of electrocatalysis is the shift of the electrode process to a lower overvoltage (in absolute value).

An electrochemical reaction, involving the electron transfer between two different phases, is intrinsically a heterogeneous process. However, electrocatalysis can be either heterogeneous or homogeneous in character. In the first case, the catalyst is confined to the electrode surface by adsorption, chemical bonding, adhesion or some other immobilization procedure. In the second case, the catalyst is present in the bulk of the solution and interface processes do not influence the chemical steps that involve it. A much more detailed approach of the general theory of electrocatalysis is available in the literature, with emphasis on technological applications.[2,3] Some monographs on kinetic methods of analysis also include sections on electrocatalysis applications in electroanalytical chemistry.[4,5] This topic is periodically surveyed in reviews published by the Journal *Analytical Chemistry*, jointly with either *Kinetic Methods of Analysis* or *Electroanalytical Chemistry*.

2 ELECTROCATALYSIS AT BARE ELECTRODES

A characteristic feature of bare-electrode processes is the occurrence of both the catalyst and the substrate in the bulk solution. Consequently, the process involves mass transfer processes for both. According to the nature of the catalyst-involving step, three kinds of electrocatalytic processes at bare electrodes can be distinguished: redox-type catalytic processes (mediated electron transfer), processes with proton transfer as a regenerative step (catalytic hydrogen evolution), and processes with a complex-forming regenerative step (catalytic reduction of some metal ions). Most of the research in this field has been done with mercury electrodes using either polarography or voltammetry at a stationary electrode, which is why the following discussion relates mostly to reduction-based processes. However, the conclusions can be easily transposed to anodic reactions.

2.1 Redox-type Catalytic Processes

2.1.1 Reaction Mechanism and Theoretical Approach

A general reaction scheme for redox-type catalytic processes can be formulated as Equations (5–8), shown in Scheme 1.

$$P + e \rightleftharpoons Q \quad (5)$$

$$Q + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B + P \quad (6)$$

$$B \xrightarrow{k} C \quad (7)$$

$$A + e \rightleftharpoons B \quad (8)$$

Scheme 1

The couple P/Q plays here the role of an electron carrier from the electrode to the substrate A via Equations (5) and (6). The term ''mediator'' is frequently used to denote the couple P/Q or one of its members. Equation (6) represents the regeneration step involving the substrate A, leading to the reformation of the direct electron acceptor, P. In principle, the direct electrochemical conversion of the substrate (Equation 8) is also possible. However, a prerequisite for the catalytic process is the occurrence of this reaction at a more negative potential than that of Equation (5). The couple A/B must be a more powerful oxidant than P/Q (i.e. $E^0_{A/B} > E^0_{P/Q}$), but Equation (8) occurs with a high activation energy (i.e. high overvoltage) and is shifted to potentials much more negative than its standard potential. The electrolytic current for the above mechanism is known as the catalytic current. The term ''parallel reaction'' is often used for Equation (6), with Equation (5) being the main step. An electrochemical reaction with a parallel chemical regeneration (EC′) is distinct from an electrochemical reaction with a subsequent chemical reaction (EC; i.e. with no regeneration). The subscripts ''r'' or ''ir'' attached to either E or C indicate the reversible or irreversible character of Equations (5) and (6), respectively. From the kinetic standpoint, the reversible or nernstian behavior of Equation (5) is defined by the condition $k_s \rightarrow \infty$ (where k_s is the standard rate constant for Equation 5) whereas an irreversible character corresponds to a low value of k_s. In thermodynamic terms, the nernstian character of

the reaction means that the Nernst equation relative to interface concentrations of both P and Q is fulfilled. The occurrence of Equation (7) is usually neglected.

A typical example of a redox-type catalyzed process is the reduction of Fe^{3+} in the presence of hydrogen peroxide[6,7] (Figure 1). In this case, the substrate is H_2O_2, whereas the couple Fe^{3+}/Fe^{2+} plays the role of the electron carrier, P/Q, according to Equations (9) and (10):

$$Fe^{3+} + e \rightleftharpoons Fe^{2+} \quad (9)$$

$$2Fe^{2+} + H_2O_2 + 2H^+ \longrightarrow Fe^{3+} + 2H_2O \quad (10)$$

The main analytical application of this kind of electrode process is the analytical determination of P with an enhanced sensitivity due to the regeneration step (Equation 6). Accordingly, an increase in the reduction current of P occurs at the expense of the substrate A.

The theoretical approach aims at establishing the relationship between the limiting current on one side and the rate constant and reactant concentrations on the other side. A time parameter is always included in the kinetic equation, because of the dynamic character of the electrode process.

Most of the available theoretical treatments assume that the substrate is present in a high excess as compared with the mediator (so that Equation 6 is pseudomonomolecular) and Q is not initially present in the solution. Substrate excess is favorable from the standpoint of the analytical determination of P because maximum sensitivity is thus achieved. In addition, it is assumed that Equation (6) occurs by outer-sphere electron transfer and, consequently, does not involve the formation of an intermediate adduct. Adsorption of the compounds involved in the electrode process brings about some complications and is usually neglected. Under these circumstances Equation (6) occurs in a solution layer (the reaction layer) adjacent to the electrode surface. Under such ideal conditions, the regeneration step is a homogeneous reaction, with no effects from the electrode–solution interface. Equation (7) is assumed to be very fast and to have no kinetic relevance. These assumptions should be carefully considered when doing kinetic investigations, but are generally not critical from the standpoint of analytical applications.

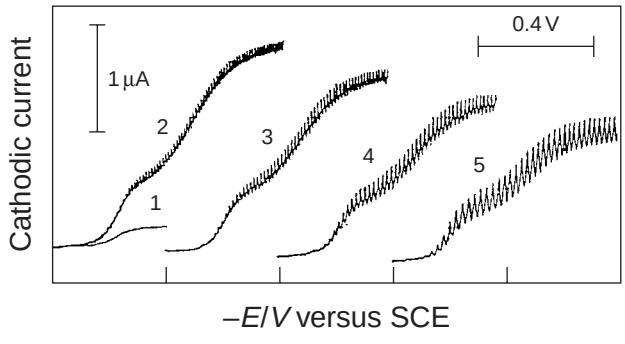

Figure 1 Direct current (DC) polarograms showing the catalytic effect of the Fe^{3+}/Fe^{2+} couple on the electrochemical reduction of hydrogen peroxide. Background electrolyte: 0.1 M KOH, 4×10^{-2} M triethanolamine. Curve 1 shows the reduction wave of Fe^{3+} (1×10^{-6} M) in the absence of H_2O_2 (Equation 9). Curves 2–5 show the electrochemical processes in the presence of both Fe^{3+} (1×10^{-6} M) and H_2O_2 (1.5×10^{-4} M) for the following values of the mercury column height, h (in cm): (1 and 2) 90; (3) 75; (4) 60; (5) 45. All polarograms start at −0.8 V vs SCE (saturated calomel electrode). A comparison of the first wave on curve 2 with the wave on the curve 1 demonstrates the increase in the limiting current due to Equation (10). The limiting current of this wave is independent of h (curves 2–5). This proves that Equation (10) (and not mass transfer processes) is the rate-determining step. The second wave on curves 2–5 is due to the diffusion-controlled reduction of H_2O_2 (Equation 8 in Scheme 1).

Theoretical treatments for the above mechanism have been developed for various electrochemical methods and are amply reviewed.[6–13] The following discussion is confined to electrochemical methods currently used in chemical analysis. However, such methods also provide valuable tools for the kinetic investigation of any chemical reaction that may occur as a regeneration step in a catalytic electrode process. In addition, this kind of investigation can provide useful information for selecting an appropriate mediator when developing chemically modified electrodes (CMEs) with electrocatalytic properties (sections 3 and 4).

Generally, two kinds of theoretical approach may be used when dealing with any class of electrode process coupled with chemical reactions.[6–9] The rigorous approach proceeds by solving the mass-transfer differential equations that include some additional terms corresponding to the formation or elimination of the reactants (e.g. P, Q, and A) by the chemical reactions. For simplicity, it is often assumed that the diffusion coefficients have the same value for all the species involved. Although only numerical solutions for such equations are available, under limiting conditions it is still possible to obtain an analytical expression for the current as a function of the rate constant, reactant concentrations, and a time-parameter characteristic for each particular electrochemical method.

The more simple, but approximate, approach is based on the concept of the reaction layer. This is the solution layer adjacent to the electrode surface, where the equilibrium of Equation (6) is disturbed by the occurrence of Equation (5). If the regeneration rate is very fast, the diffusion of P may be neglected and the single species taking part in the diffusion process is A. Furthermore, it is assumed that the thickness of the reaction layer is very small compared to the thickness of the diffusion layer for A. Consequently the concentration of A inside the reaction layer is assumed to be constant. In addition,

it is supposed that the steady state occurs within the reaction layer, i.e. the formation of P by Equation (6) is completely balanced by its removal via Equation (5). Hence, the reaction layer method is restricted to the limiting case of a very fast parallel chemical reaction.

An alternative to the above methods is the digital simulation of the electrode process.[14] The mass transfer model is discretized into events that occur within small time increments and small intervals of distance from the electrode surface. In this way, the concentration versus distance profiles can be calculated by successive iterations. Initial assumptions about the mechanism and the rate parameters must be made, and this makes unavoidable the use of the diagnostic criteria and relevant relationships derived by the previous methods. However, digital simulation is very useful when checking the validity of some approximations or for investigating complicated processes that cannot be handled by solving the relevant set of differential equations.

2.1.2 Kinetic Equations

2.1.2.1 Direct Current Polarography Most of the pioneering work in the field of catalytic electrode processes was done by the Czech school of polarography and is amply reviewed elsewhere.[6–8,11–13] As a rule, the effect of sphericity is neglected and the dropping mercury electrode is regarded as an expanding plane electrode. This common approximation is valid provided that the thickness of the diffusion layer is much smaller than the electrode radius.

Taking into account the restrictions mentioned in the previous section, the solution to the system of differential equations for the diffusion of P and Q, accompanied by the irreversible pseudo-first-order Equation (6), is expressed by Equation (11):[6]

$$\frac{i_c}{i_d} = \Psi(\chi_1) \tag{11}$$

Here, i_c is the mean limiting catalytic current and i_d is the mean diffusion current of P (which can be measured in the absence of the substrate A). The dimensionless parameter χ_1 is defined in Equation (12):

$$\chi_1 = k_1 c_A^0 t_1 \tag{12}$$

where c_A^0 is the concentration of the substrate and t_1 is the drop time for the dropping mercury electrode. The ratio i_c/i_d represents the current amplification brought about by the regeneration step. A high regeneration rate, due for example to a high excess of substrate ($\chi_1 > 10$), has a beneficial effect on the sensitivity. Under these conditions, the amplification factor is given by the following asymptotic form (Equation 13):

$$\frac{i_c}{i_d} = 0.81(k_1 c_A^0 t_1)^{1/2} \tag{13}$$

The above treatment is applicable only to the limiting-current region, i.e. for potential values negative enough to make the charge transfer reaction (Equation 5) very fast and kinetically irrelevant. An approximate expression for the current–voltage curve, which is valid for any potential value on the kinetic wave, is also available.[15] Accordingly, the catalytic wave is described by the well-known equation of the polarographic wave (Equation 14):

$$i = \frac{i_c}{1 + \exp[\alpha n F(E - E_{1/2})/RT]} \tag{14}$$

where i is the current as a function of electrode potential E, n is the stoichiometric coefficient of the electron in the electrode reaction (Equation 5) and $E_{1/2}$ is the half-wave potential. For the case of a reversible electrode reaction $E_{1/2} = E_{P/Q}^0$ and $\alpha = 1$. For the opposite case of an irreversible electrode reaction, under the limiting conditions of a very fast parallel reaction ($\chi_1 > 10$), $E_{1/2}$ is dependent on both k_s and the rate constant k_1 (Equation 15):

$$E_{1/2} = E^0 + \left(\frac{RT}{\alpha n F}\right) \ln \frac{k_s}{(D_P k_1 c_A^0)^{1/2}} \tag{15}$$

where D_P is the diffusion coefficient of P. The above relationships appear as particular cases of a more general treatment which is not restricted to the steady-state assumption and refers to an electrode expanding with any power law.[16] Because most of the available theoretical treatments deal with the expanding plane electrode, the effect of electrode sphericity has been assessed[17] and it was concluded that i_c is more affected than $E_{1/2}$ by this approximation; however, only slight corrections are required.

If the regeneration step (Equation 6) is a second-order reaction, a subunitary correction factor should be added to Equation (13).[6] This obviously brings about a decrease in the sensitivity of the determination of P.

DC polarography provides some straightforward diagnostic tests for the occurrence of the catalytic electrode process. The most typical case is that corresponding to Equation (13). This implies a very fast regeneration of P and no concentration gradients for A (which is present in a high excess) or P and Q (which are cycled inside the reaction layer). Consequently, the overall electrode process is independent of mass-transport phenomena.[6] Experimentally this is proved by the independence of the limiting catalytic current, i_c, on the mercury pressure above the mercury electrode (expressed by the height of the mercury level above the lower end of the capillary

electrode, h) (Figure 1). This contrasts the behavior of the diffusion current, which is directly proportional to $h^{1/2}$. Under intermediate conditions, the i_c versus $h^{1/2}$ plot gives a straight line with a nonzero intercept. The investigation of the deuterium isotope effect is a valuable test for detecting the participation of hydrogen species in the rate-determining chemical step.[18]

2.1.2.2 Linear Sweep Voltammetry and Cyclic Voltammetry at Stationary Electrodes Linear sweep voltammetry (LSV) is an essentially nonsteady-state method under convection-free conditions, involving a potential scan according to Equation (16):

$$E = E_i + vt \tag{16}$$

where E_i is the initial potential, v is the scan rate (in $V\,s^{-1}$), and t is the time variable. The output is the current–potential curve (known as a voltammogram) which displays asymmetric peaks corresponding to each electrode reaction. Cyclic voltammetry (CV) involves potential cycling with constant speed between two potential values. This makes it possible to detect the product of the initial scan from the relevant peak recorded during the reverse process.

The theory of the catalytic process in LSV, as worked out by Savéant and Vianello[19] and Nicholson and Shain,[20] is amply described elsewhere.[11–13] It is assumed that $c_A^0 \gg c_p^0$ and $k_2 = 0$, so that Equation (6) is pseudomonomolecular and irreversible. The perturbation of the system is expressed by the dimensionless sweep rate $a = nFv/RT$, whereas the mobility of the chemical reaction is shown by the pseudomonomolecular rate constant $k_1 c_A^0$. The treatment assumes a planar electrode (i.e. semi-infinite planar diffusion) but the results are reasonably correct for spherical electrodes as well. This is because the relevant processes occur in the reaction layer, which has a thickness much lower than the electrode radius.

The general solution of the differential equations for the mass transfer accompanied by the chemical reaction gives the current–potential relationship in the form of a series. First, the case of the reversible electrode reaction is considered (i.e. $E = E_{P/Q}^0 + (RT/nF)\ln(c_{P,i}/c_{Q,i})$; the subscript i denotes concentrations at the electrode/solution interface). If the regeneration is very fast ($k_1 c_A^0/a \gg 1$), the current–potential curve assumes the shape of the steady-state voltammetric wave and is described by Equation (14) where $\alpha = 1$ and $E_{1/2}$ is the polarographic half-wave potential for the electron transfer reaction (Equation 5) (i.e. $E_{1/2} = E_{P/Q}^0 - (RT/2nF)\ln(D_P/D_Q)$; usually, $E_{1/2} \approx E_{P/Q}^0$ because $D_P \approx D_Q$). Under these conditions, the value of i_c is dictated by the rate of the parallel reaction, according to Equation (17):

$$i_c = nFAc_P^0(D_P k_1 c_A^0)^{1/2} \tag{17}$$

Accordingly, i_c is not dependent on the scan speed, in contrast with the peak current for the diffusion-controlled process, i_p, which is recorded in the absence of A and is directly proportional to $v^{1/2}$. This is an important diagnostic test for the catalytic process, in addition to the curve shape and the absence of the hysteresis on a CV run (provided that the reversal potential is not more negative than $(E_{1/2} + 35/n)$ mV). Equation (17) enables a straightforward determination of the rate constant k_1.

The current amplification brought about by the regeneration of the electron acceptor is defined in Equation (18):

$$\frac{i_c}{i_d} = 2.242(k_1 c_A^0)^{1/2} a^{-1/2} \tag{18}$$

At a very high scan speed ($k_1 c_A^0/a \ll 1$) the effect of the regeneration becomes less evident and the shape of the voltammetric pattern becomes an asymmetric peak. At the limit, the timescale is so short that Equation (6) is not disturbed any longer and the rate-determining step is the diffusion of P.

If Equation (6) is reversible, both k_1 and k_2 must be taken into account and the limiting current recorded at a low sweep rate ($k_1 c_A^0/a \gg 1$) is given by Equation (19):[11,21]

$$i_c = nFAc_P^0[D_P(k_1 c_A^0 + k_2)]^{1/2} \tag{19}$$

In the case of the irreversible electrode reaction the rate of Equation (5) is expressed by the Butler–Volmer equation and the dimensionless sweep rate is defined as $b = \alpha nFv/RT$. The current–voltage relationship is available as a series.[11,20] However, for the analytically important case of a very fast regeneration ($k_1 c_A^0/b \gg 1$), the current–voltage relationship is in accord with Equation (14), i.e. it has a sigmoid shape with a limiting current i_c as given by Equation (17). The effect of irreversibility is demonstrated by the shape of the curve, which is more spread out than in the case of the reversible reaction (because $\alpha < 1$), and by the dependence of $E_{1/2}$ on the rate constants for Equations (5) and (6).

The LSV technique has limited analytical applications but it is a very valuable tool for investigating catalytic electrode processes of analytical interest, involving either synthetic or biogenic catalysts.

2.1.2.3 Differential Pulse Polarography, Square-wave Voltammetry, and Alternating Current Polarography The above methods are characterized by the superposition of a periodic, small amplitude voltage (10–100 mV)

over the constant polarization potential. The superimposed signal is either square shaped (for differential pulse polarography (DPP) and square-wave voltammetry (SWV)) or is a sine wave (for alternating current (AC) polarography). An outstanding sensitivity is achieved owing to the ability of such methods to distinguish the electrolytic current (which bears the analytical information) from the disturbing capacity current. Additional improvement in the sensitivity and the detection limit is brought about by coupling the electrode reaction (Equation 5) with the chemical regeneration step (Equation 6).

In the case of DPP[11,22] an amplification factor of $(\tau\pi k_1 c_A^0)^{1/2}$ results from the theoretical analysis of the $E_rC'_{ir}$ process for the most favorable case of a very fast regeneration (where τ is the pulse duration). A treatment for the more general case, involving stoichiometric factors other than unity in Equations (5) and (6), is also available.[23]

The most successful version of the square-wave methods is SWV, which involves a fast variation of the polarization potential (typically, $10\,V\,s^{-1}$), thus combining a high sensitivity with a very short duration of the run. The theoretical treatment for the $E_rC'_{ir}$ mechanism shows that in the purely kinetic region ($k_1 c_A^0 p > 2$; p is the square-wave period) the response (known as the net current, i.e. the difference between current values sampled at the extremes of the square wave pulse) is proportional to $(k_1 c_A^0 p)^{1/2}$ and c_P^0.[24] The modulation frequency has no effect on the current, but the lowest possible (e.g. 10 Hz) is recommended because this leads to a very low capacity current. An outstanding detection limit (2.5×10^{-8} M) for Ti^{IV} determination by catalytic SWV in the presence of ClO_3^- as a substrate was reported.[25]

AC polarography has not enjoyed so much attention as the other methods, although the theoretical treatment performed for the $E_rC'_{ir}$ mechanism, assuming planar diffusion at a constant-area electrode, demonstrates the same kind of favorable effect of the parallel reaction (as pointed out before). In the most favorable case of a very high reaction rate ($k_1 c_A^0 \gg w$; w is the angular frequency of the modulating sine wave) an increase in the peak current by a factor of $(k_1 c_A^0 / w)^{1/2}$ is predicted, compared with the diffusion current of P (i.e. the peak current recorded for $k_1 c_A^0 = 0$).[11,26] Under the above assumption, the peak current is not dependent on w, because of the fast regeneration of P, but it is directly proportional to $(k_1 c_A^0 D_P)^{1/2}$ and c_P^0. A comprehensive review of the theory for AC methods including catalytic electrode reactions provides more details.[9]

2.1.2.4 Hydrodynamic Voltammetry The methods included in this class involve electrolysis under controlled and reproducible electrolyte flow conditions. Consequently, the mass-transfer process to the electrode fulfills the conditions for the steady state. Among the systems in this class, the most common are rotating disc, wall-jet, and the flow-through (tubular or channel) electrodes.[27] The latter types are applicable to continuous analytical measurement.

A general solution for the rotating disc electrode (RDE) is available in the case of the $E_rC'_{ir}$ mechanism (Equation 20):[11,27]

$$\frac{i_c}{i_d} = x\,\mathrm{cotanh}(x) \tag{20}$$

where $x = 1.61(\nu/D_P)^{1/6}(k_1 c_A^0/w)^{1/2}$, ν is the cinematic viscosity, and w is the angular velocity of the rotating disc. For a very fast regeneration the current amplification is $i_c/i_d = x$.[11,27] Under the same conditions, i_c is proportional to c_A^0 (via i_d) and is not dependent on w.

The theory for the EC′ mechanism at a channel electrode[28] gives the (i_c/i_d) ratio as a function of a dimensionless kinetic parameter including the rate constant k_1 and the electrolyte flow rate. The solution is available both as a series and as working curves. A number of variations in the EC′ mechanism occurring at a channel electrode were recently examined by numerical computation.[29] For the most simple case of an $E_rC'_{ir}$ mechanism, the current–voltage curves have the typical shape for a steady-state process, in agreement with Equation (14) with $\alpha = 1$. Although the solutions are more complicated than for the RDE, it is still clear that the amplification factor increases with an increase in both k_1 and c_A^0.

2.1.2.5 Adsorption Effects Adsorption at the electrode interface is a common feature in electrochemistry of organic compounds or complexes containing organic ligands. It brings about a series of theoretical complications that originate in the deviation from the equilibrium state of the adsorption process or in the effect of electrode potential on the parameters of the adsorption isotherm. However, the adsorption of one or more participants in the catalytic process may favorably enhance the overall reaction rate as a consequence of the augmented local concentrations. Special advantages were demonstrated in the case of adsorptive stripping voltammetry (ASV) where the adsorption represents the basis for the preconcentration step.

A theoretical treatment for heterogeneous redox catalysis was performed under the assumption that the mediator couple is confined to the electrode surface as a monomolecular layer, either by irreversible adsorption or by another method of electrode modification.

When dealing with the CV of a nernstian system ($E_rC'_{ir}$) the dimensionless kinetic parameter was defined as in

Equation (21):[30]

$$\Lambda_s = k_1 \Gamma_T \left(\frac{RT}{FvD_A} \right)^{1/2} \quad (21)$$

where Γ_T is overall surface coverage ($\Gamma_T = \Gamma_P + \Gamma_Q$). The overall current is the algebraic sum of the current produced by Equation (5) alone (i_s) and the purely kinetic component brought about by the occurrence of Equation (6) in the adsorbed state, i_c. The effects of Equation (6) on the peak current and potential are available in a graphical form, as a function of both Λ_s and the departure from $E^0_{P/Q}$.[30] With the increase in Λ_s, the i_c versus E curve changes in shape, from a sigmoid hysteresis-free curve to an asymmetrical peak. Simultaneously, the overall process turns from a kinetically controlled one (for small Λ_s values) to control by the diffusion of A when $\Lambda_s \to \infty$. This change is accompanied by a strong increase in i_c and the shift of half-peak potential towards more negative values. The limiting value of i_c still cannot overcome the diffusion-controlled current of A at a bare electrode. Such conditions are suitable for the determination of the substrate.

If Equation (5), occurring in the adsorbed state, is followed by the formation of an adsorbed Q–A complex, an intramolecular electron transfer proceeds as Equation (6) and the product decomposes, leading subsequently to the regeneration of P.[31] Assuming that the complex formation is at equilibrium and that Equation (5) is reversible, the relevant kinetic parameter Θ was defined as in Equation (22):

$$\Theta = k_3 K_{QA} \Gamma_T \left(\frac{RT}{FvD_A} \right)^{1/2} \quad (22)$$

where k_3 is the rate constant for the intramolecular electron transfer, K_{QA} is the equilibrium constant for complex formation and $\Gamma_T = \Gamma_P + \Gamma_Q + \Gamma_{QA}$. At a small, fixed value of $K_{QA}c^0_A$ ($K_{QA}c^0_A \leq 0.1$), an increase in Θ from 0.1 to 30 makes the peak current i_c increase towards a limit which corresponds to kinetic control by the diffusion of A. Simultaneously, the shape of the i_c–E curve changes from sigmoid (which is characteristic of the absence of a noticeable concentration gradient of A) to the usual, peak-shaped form of the voltammogram. At the same time, the peak potential, which is dictated by the P : Q concentration ratio in the adsorbed state, shifts towards more positive values. At a constant and small value of Θ (kinetic control), the increase in $K_{QA}c^0_A$ induces a decline in i_c with no marked change in the peak potential. The effect of $K_{QA}c^0_A$ is much less marked for $\Theta > 10$.

The above results also provide an interpretation of electrocatalysis at modified electrodes. With the bare electrode, at least some qualitative insight into heterogeneous catalysis can be drawn. For example, the above conclusion can be used for interpretation of the processes occurring in the quantitative determination of the catalyst by stripping voltammetry with adsorptive preconcentration. Under such circumstances, the presence of the catalyst in the bulk solution (usually, under 10^{-7} M) can be neglected and the catalytic effect is due to the adsorbed species only. In this case, one major deviation from the above treatments may be the reversible character of catalyst adsorption, leading to gradual desorption during the potential scan.

2.1.2.6 Electrocatalysis at Ultramicroelectrodes The use of the UME (ultramicroelectrode), defined as an electrode with the characteristic dimension of 1 μm or less,[32] shows a series of advantages: small interference arising from capacity current and electrolyte resistance, allowing the use of highly resistive media including nonpolar solvents; the possibility of approaching steady-state mass transfer without convection; and suitability to microscale and in vivo analysis. Various UME shapes are disc, ring, disc array, hemisphere, band, cylinder and interdigitated array.

A particular feature of the UME is that the steady-state mass transfer occurs even in the absence of the convection, provided that the characteristic dimension (e.g. sphere radius r_s) is much lower than the thickness of the diffusion layer ($r_s \ll (D\tau)^{1/2}$, where τ is the timescale of the experiment, such as the pulse duration in DPP or duration of potential scan in LSV).[33] Accordingly, a more general definition of the UME is proposed by Amatore[34] who assumes that any electrode/electrolyte system shows the behavior of the UME if $r_s \ll \delta$, where δ is the thickness of the convection-free solution layer at the electrode surface. The factor δ depends not only on D and τ but also on the hydrodynamic conditions and electrolyte viscosity. Under such conditions, radial (convergent) diffusion prevails over planar diffusion. Rather than the dimension itself, the occurrence of radial diffusion is the main characteristic of a UME. In principle, this definition does not impose an upper limitation in the electrode dimension. However, it must be stressed that the advantages emerging from the small-capacity current and very small ohmic voltage drop arise only if the electrode dimension is under 100 μm.

The use of the UME for kinetic investigations has been reviewed by Montenegro.[33] Under steady-state conditions, the diffusion-controlled reduction of the P species (Equation 5 alone) gives a sigmoid i–E curve (in accord with Equation 14) with a plateau corresponding to the limiting diffusion current i_d. The occurrence of the $E_rC'_{ir}$ mechanism (with pseudo-first-order regeneration) induces a higher plateau current (i_c), according to

Equation (23):[35]

$$\frac{i_c}{i_d} = 1 + \Lambda_m^{1/2} \tag{23}$$

with i_d as in Equation (24):

$$i_d = \frac{FADc_P^0}{r_s} \tag{24}$$

The dimensionless kinetic parameter Λ_m is defined in Equation (25):

$$\Lambda_m = \frac{r_s^2 c_A^0 k_1}{D} \tag{25}$$

Here, r_s is the radius for a hemispherical electrode (e.g. a mercury hemisphere electrolytically deposed on a Pt microdisc). The rate constant k_1 can be determined from the slopes of the i_c/i_d ratio versus r_s plot (at constant c_A^0) or i_c/i_d ratio versus $(c_A^0)^{1/2}$ plot (at constant r_s). Alternatively, the above equations are easily rearranged aiming at data analysis by means of the i_c versus $1/r_s$ plot.[36] The pseudo-first-order conditions are fulfilled if $c_A^0 \gg k_1 r_s^2 c_P^2/D$.[37] However, the occurrence of the steady state is favored by a high Λ_m/q ratio, where q is the dimensionless potential sweep rate ($q = (Fvr_s^2/RTD)^{1/2}$).[35] Empirically, the occurrence of the steady state can be demonstrated by the independence of i_c on v or by performing a logarithmic plot of the curve according to Equation (14). A more general treatment for the catalytic mechanism at a hemispherical electrode with application to chronoamperometry, cyclic staircase voltammetry, and cyclic LSV was published recently.[38]

In the case of a second-order regeneration, the solution of the relevant differential equations for a spherical electrode leads to Equation (26) (under the assumption of equal diffusion coefficients for all the species involved):[39]

$$\frac{i_c}{i_d} = 1 - \left(\frac{k_1 r_s^2 c_P^0}{2D} + 0.5\left(\frac{k_1 r_s^2 c_P^0}{D}\right)^2 + (4\Lambda_m)\right)^{1/2} \tag{26}$$

In contrast to the pseudo-first-order reaction, the relationship between the i_c/i_d ratio and r_s is nonlinear in this case. Interestingly, the pseudo-first-order kinetic regime appears as a limiting case of the second-order one if $c_A^0 \gg k_1 r_s^2 (c_P^0)^2/D$. In other words, for small values of $k_1 r_s^2$ the process may assume pseudo-first-order character even if c_A^0 is close to c_P^0. Under these conditions, A is not depleted in the reaction layer and the A/P concentration ratio is very high here. Consequently, Equation (26) assumes the simple form of Equation (23).

The disc electrode (with radius r_d) was approached initially by asserting that it behaves like a hemisphere with an equivalent area and it is, therefore, sufficient to substitute r_s by $\pi r_d/4$ in the relevant equations for the hemispherical electrode. It was shown later that this is a very rough approximation in the case of the catalytic process and rigorous equations were derived accordingly.[40]

The effect of the difference between the diffusion coefficients was investigated by using the reaction layer concept.[41] Kinetic equations were derived under such conditions for either the first- or second-order homogeneous catalytic reaction at an ultramicrodisc electrode. In accord with the results for the spherical UME presented above, the reaction order can be controlled by changing the dimension of the UME. Thus, the second-order reaction can be changed to quasi-first-order by decreasing the dimension of the UME.[42]

It was stated that the microcylinder electrode shows important advantages from the standpoint of both theory and experiment. The theory of the EC'_{ir} mechanism with a second-order parallel reaction was worked out using the concept of the reaction layer and was applied to the Fe^{III}–ethylenediamine tetraacetic acid (EDTA)–H_2O_2 system.[43]

Although the steady state seems to be typical of the UME, the nonsteady state may also be attained. This occurs for a very low timescale ($r_d \gg (D\tau)^{1/2}$) and involves the planar diffusion. Under such conditions, the experimental data can be dealt with in the classical way (sections 2.1.2.2 and 2.1.2.3), with the advantage of a very low τ value, which allows the investigation of very fast electrode reactions.

2.1.2.7 Intramolecular Electron Transfer as a Regeneration Step As shown before, most of the available theoretical treatments of catalytic electrode reactions assume that the parallel chemical reaction proceeds via an outer sphere mechanism. Nevertheless, the occurrence of the opposite mechanism, involving a catalyst–substrate adduct, cannot be ruled out as was mentioned in section 2.1.2.5. The formation of a substrate–catalyst complex may occur, for example, in the case of some processes involving oxo cations as catalysts and hydrogen peroxide as the substrate.[44] This kind of mechanism is easily detected in DC polarography by the direct proportionality relationship between the i_d/i_c ratio and $1/c_A^0$.[45] The catalytic limiting current is directly proportional to substrate concentration under pseudo-first-order conditions and for $c_A^0 \gg K_{QA}$. Such conditions are appropriate for the substrate determination with a sensitivity that is directly proportional to the concentration of the catalyst.

2.1.3 Some Examples of Parallel Electrocatalytic Systems

Many early investigations were performed by DC polarography and a detailed compilation of the results

is available.[44] Data included in this review could provide useful information for developing analytical methods based on more advanced electroanalytical techniques.

Commonly, inorganic oxidants are used as substrates, including either neutral (hydrogen peroxide, hydroxylamine) or anionic species (chlorite, chlorate, perchlorate, bromate, iodate, nitrate, nitrite, persulfate ions). The mediator may be, in principle, any transition metal which shows two stable oxidation states. Metals in groups Ib (Cu), IVb (Ti^{IV}), Vb (V^{V}), VIb (Cr^{III}, Mo^{VI}, W^{VI}), VIII (Fe^{III}, Co^{III}, Ru^{III}, Os^{VIII}) as well as some lanthanides (Y^{III}) and actinides (U^{VI}) can be mentioned in this connection and appropriate methods for the polarographic determination of such metal ions were quoted.[44] An organic mediator (4-hydroxy-2,2,6,6-tetramethyl-piperidinyloxy radical) which catalyzes the anodic oxidation of hydrazine on a carbon electrode was also reported.[46]

Although Scheme 1 may accurately describe many real systems, frequent deviations occur due, for example, to the simultaneous occurrence of different chemical states of the catalyst, double-layer effects on reactions involving charged particles, and adsorption of complexes with organic ligands. A careful empirical optimization of the experimental parameters still allows the use of complicated processes for analytical purposes, even if the rigorous kinetic investigation is not possible.

Because the substrate is very often a covalent compound, Equation (6) may consist of several steps involving free-radical species. For example, in the $Fe^{III}/Fe^{II}-H_2O_2$ system this reaction occurs as shown in Equations (27) and (28):[6,8]

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + OH^- + OH\bullet \qquad (27)$$

$$Fe^{2+} + OH\bullet \longrightarrow Fe^{3+} + OH^- \qquad (28)$$

Analogously, hydroxylamine reduction by either Fe^{2+} or Ti^{III} involves $NH_2\bullet$ or $OH\bullet$ radicals.[6,8]

In the presence of a radical scavenger, such as ethanol or an organic ligand in excess, the stoichiometry of the regeneration step will change. Accordingly, the stoichiometric factor of Equation (5) should be included in the kinetic equation as a multiplier for k_1. This factor indicates how many times Equation (5) occurs for each occurrence of Equation (6), i.e. it is the ratio of the stoichiometric coefficients of Q in Equations (6) and (5). Thus, for the $Fe^{3+}-H_2O_2$ system (Equations 9 and 10), the stoichiometric factor is two but it decreases towards one in the presence of ethanol. Consequently, the presence of radical scavengers will produce a drop in sensitivity. Detailed investigations on this topic have been made with special reference to LSV.[19]

Hydrogen peroxide is by far the most widely applicable substrate and many of the above-mentioned mediators react with it giving catalytic currents for potential analytical applications. Some particular features are the effect of pH on H_2O_2 oxidizing power and the formation of peroxo complexes of various catalyst ions (Ti, V, Mo, W, Nb, Hf, Zr, Ta) followed by a shift in the redox potential of the metal ion to less negative values.

Hydroxylamine also shows a broad area of applications (for Nb, Ru, and W). Chlorate and bromate ions are also strong oxidants with many possible applications, but the negative charge may lead to intricate polarographic shapes, especially when the mediator itself is an anion, such as Mo^{VI} or W^{VI}. Worthy of mention is the catalytic current of selenium (after reduction from SeO_3^{2-} to H_2Se and redox cycling between this state and Se^0) in the presence of BrO_3^-. Perchlorate ion is a milder oxidant and only Mo^{VI} and W^{VI} give catalytic currents in its presence. The catalytic activity of Mo^{VI} and W^{VI} is depressed by complex formation. Iodate and periodate provide catalytic currents for organic sulfur compounds, such as for the cystine/cysteine couple.

2.1.4 Analytical Applications

Redox-type electrocatalytic processes can be used to determine the catalyst, the substrate, an activator or an inhibitor of the redox reaction; the first of these applications is most usual. Substrate determination may be more conveniently done by electrocatalysis at a modified electrode (section 3), although the advantages of bare electrodes (readily available, reproducible response and a broad field of application) can very often decide in favor of the latter kind of electrode.

2.1.4.1 Selection of Experimental Conditions The experimental conditions for either catalyst or substrate determination can be decided using the kinetic zone diagram.[47] This shows the occurrence of various kinetic regimes as a function of two dimensionless parameters that include both intrinsic (rate constants) and operational factors (the timescale and reactant concentrations). The simplest procedure for dealing with the kinetic data corresponds to limiting kinetic regimes and such conditions are also optimal for analytical determination. Figure 2 shows the kinetic zone diagram for the $E_rC'_{ir}$ mechanism under the conditions of LSV. The relevant parameters are $\gamma = c_A^0/c_P^0$ (the excess factor) and $\lambda_1 = (k_1 c_P^0/v)(RT/F)$. The arrows on the right-hand side of the figure represent the direction and the magnitude (in log units) of the shift produced by intrinsic (k_1) and operational (v, c_A^0, c_P^0) factors.

In the purely kinetic zone K, fast substrate consumption leads to a purely kinetic regime involving mutual compensation between the parallel chemical reaction and the diffusion process. A limiting case of this situation that corresponds to a very low λ_1/γ ratio is represented by

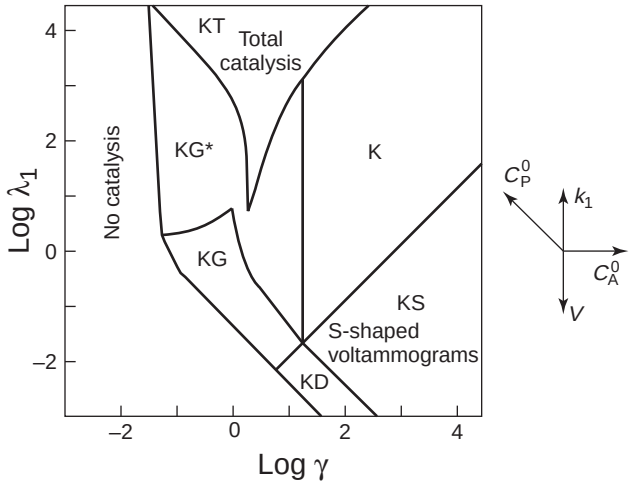

Figure 2 Kinetic zone diagram for the $E_rC'_{ir}$ mechanism in LSV. (Adapted from Savéant and Su.[47])

the KS zone. Here the substrate concentration is constant throughout the solution and the diffusion effect is negligible. Consequently, the voltammogram shows the characteristics of the steady state, that is, it is S-shaped, with the half-wave potential equal to $E^0_{P/Q}$ and the plateau current is in accord with Equation (17), i.e. directly proportional to c^0_P. Such conditions are suitable for catalyst determination with the sensitivity proportional to $(c^0_A)^{1/2}$.

In the opposite case of total catalysis (KT), the substrate consumption is so fast that the rate-determining step is the diffusion of the substrate. This makes this kinetic zone suitable for substrate determination. Under these conditions, the catalytic process produces a peak located well ahead of the standard potential of the P/Q couple.

The KD zone represents the transition from pure diffusion control (no catalysis) to pure kinetic control. In the KG and KG* zones, some complications arise from the tendency of the voltammogram to split into two waves, the potential separation and relative heights of which are functions of both λ_1 and γ.

As is usual in kinetic analysis, the signal is strongly dependent on the chemistry of the system. The mediator-binding ligands (activators) may strongly influence the rate of the regeneration reaction and its proper selection will improve the sensitivity dramatically. The pH is also an important factor, as it may influence various processes such as metal ion hydrolysis, complex formation and substrate protonation. The presence of radical scavengers, with negative effects on sensitivity, has to be avoided. Temperature should be carefully controlled in view of its strong effect on the catalytic current via the rate constant k_1. However, this could be an advantage as the sensitivity strongly increases with temperature.

It is also worth mentioning the potential applications of electrocatalysis in microheterogeneous fluids. Surfactant aggregates (micelles, microemulsion) can enhance the rate of the parallel reaction by bringing the reactants together in local higher concentrations. This results in enhanced sensitivity and makes nonpolar substances amenable to electroanalysis in an aqueous milieu.[48]

2.1.4.2 Catalyst Determination This is by far the most important analytical application of redox catalysis at bare electrodes, enabling an appreciable enhancement in sensitivity. Substrate selection is important for achieving convenient analytical figures of merit. First, the substrate must be selected taking into account its redox potential with respect to that of the mediator (section 2.1.1). Additionally, a high rate constant for the parallel reaction, as well as some affinity of the substrate towards the mediator, may improve the sensitivity.[45] Although dated, the survey by Mark and Rechnitz[49] of the redox reactions from the kinetic viewpoint provides a good theoretical starting point for rationalizing the analytical applications of redox catalytic electrode processes. More recent monographs afford a deeper insight of the molecular basis of redox reaction.[50]

Most of the early applications of redox catalysis at bare electrodes have been reviewed[44] and refer to industrial, geological and biological samples. The most impressive progress in this field results from the combination of ASV with redox catalytic processes. In ASV the preconcentration is made by the adsorption of the analyte (here, a metal ion in complex form) on the electrode surface at an appropriate potential. During the subsequent potential scan, the adsorbed species undergoes an electrode reaction that produces the analytical signal as a current peak.[51] This method is very convenient for the determination of most of the transition metals that cannot be concentrated as an amalgam in the mercury drop electrode. If the electrode reaction is accompanied by a redox reaction, which regenerates the initial form, the sensitivity is greatly enhanced. Also, the selectivity can be ameliorated in this way. The selection of the ligand is crucial – it must possess a high surface activity so as to enable the adsorptive preconcentration and it must not alter the redox properties of the analyte. It is particularly advantageous that the ligand (which is usually added in excess) also does act as a substrate. In this case, the parallel reaction involves two adsorbed partners and the reaction rate may increase accordingly.

Some selected examples of analytical applications of redox catalysis at bare electrodes are included in Table 1.

2.1.4.3 Substrate or Ligand Determination The catalytic determination of the substrate may be advantageous from several standpoints. First, it makes possible

Table 1 Selected examples of voltammetric methods of analysis using catalytic redox processes for signal enhacement

Analyte	Ligand (L)	Substrate	Supporting electrolyte	Method[a]	Detection limit (M)	Ref.
Co	Dimethylglyoxime (0.1 mM)	NO_2^- (0.5 M)	0.1 M Ammonia/ ammonium chloride buffer	LS–ACSV: $E_d = -0.5$ V (Ag|AgCl (3 M KCl))	1×10^{-10}	52
$Cr^{III} + Cr^{VI}$ (in natural water samples)	Diethylenetriamine-pentaacetic acid (DTPA) (0.01 M)	NO_3^- (0.5 M)	0.04 M CH_3COONa, pH 6.2	DP–ACSV: $E_d = -1.0$ V (Ag|AgCl (sat. KCl))	5×10^{-10}	53
Fe^{III} and Fe^{II} (in seawater)	1-Nitroso-2-naphtol (20 μM)	BrO_3^- (40 mM)	0.01 M HEPES buffer (pH 8)	LS–ACSV: $E_d = -0.05$ to -0.04 V (Ag|AgCl (3 M KCl))	8×10^{-11}	54
Ge (in soils and vegetables)	Catechol (1.5 mM)	BrO_3^- (0.1 M)	0.07 M Acetate buffer (pH 4.6)	DC polarography	10^{-9}	55
Mo and W (in seawater)	Benzilic acid (0.3 mM) + 2-methyl-8-quinolinol (0.2 mM)	ClO_3^- (0.5 M)	0.03 M H_2SO_4	DC polarography (after preconcentration on a resin column)	5×10^{-10} (Mo) 5×10^{-11} (W)	56
Ti (in seawater)	Mandelic acid (4 mM)	ClO_3^- (45 mM)	NH_4Cl, pH 3	DP–ACSV: $E_d = -0.1$ V (Ag|AgCl (3 M KCl))	10^{-12}	57
Ti (in pure Al)	Cupferron[b] (0.1 mM)	Cupferron	0.1 M HCl in 10% v/v ethanol	LS–ACSV: $E_d = -0.2$ V (SCE)	6×10^{-11}	58
Ti and Mo (in natural water)	Mandelic acid (4 mM)	ClO_3^- (45 mM)	NH_4Cl, pH 3	DP–ACSV: $E_d = -0.1$ V (Ag|AgCl (3 M KCl))	10^{-10} (Mo) 10^{-9} (Ti)	59
U (in ground and river water)	Cupferron[b] (0.005 mM)	Cupferron	Acetate buffer (0.1 M, pH 4.5)	DP–ACSV ($E_d = -0.2$ V) or on-line DP–ACSV flow analysis	4×10^{-10}	60
V	Cupferron (0.02 mM)	BrO_3^- (6 mM)	5 mM Acetate buffer (pH 4.8)	DP–ACSV, $E_d = +0.1$ V	5×10^{-12}	61

[a] ACSV, adsorptive cathodic stripping voltammetry; DP, differential pulse; LS, linear scan.
[b] The ligand also acts as a substrate/oxidant.

the determination of some compounds that show electrochemical properties not suitable for electroanalytical quantitation. Second, the electrochemical process proper involves the mediator and, by an appropriate mediator selection, the electrochemical response may be shifted beneficially along the potential axis. No special advantages from the standpoint of sensitivity can be expected under the conditions of the steady-state methods (namely DC polarography and RDE voltammetry) because the response cannot overcome the diffusion current of the substrate. However, the sensitivity and detection limit can be substantially improved when using nonsteady-state methods (e.g. DPP, SWV or AC polarography) provided that the mediator undergoes a reversible electrode reaction. This advantage results from the dependence of the response on the rate constant for the electrode reaction, k_s. An important detail related to the form of the calibration graph is the occurrence of a nonzero signal even in the absence of the analyte substrate. This is due to the simple electrode reaction of the mediator.

Although not so frequent as the applications for catalyst determination, the catalytic substrate determination may prove very useful as, for example, in the determination of trace nitrate (with U^{VI} as catalyst),[62] H_2O_2 (on a copper electrode, with anodically generated Cu^I as catalyst),[63] hydrazine,[46] or hydroxylamine[64]

(with 4-hydroxy-2,2,6,6-tetramethyl-piperidin-1-oxyl as catalyst).

Catalytic redox processes may also be used for the determination of a ligand, provided that the complex compound differs from the noncomplexed mediator in both the standard potential and catalytic activity. The distinct wave of the complex is amplified by an appropriate substrate in order to improve the sensitivity. In this respect, the determination of some hydroxyacids and polyphenols using either the $Ti^{IV}-ClO_3^-$ or $Mo^{VI}-ClO_3^-$ systems has been investigated.[45] Analogously, phosphate ion may be determined as molybdophosphoric acid in the presence of ClO_3^-.[65] Nitrite ion may be determined after disproportionation to nitrogen oxide which is thereafter included in the $[Fe(SCN)(NO)]^+$ complex. This gives a catalytic current in the presence of dissolved oxygen.[66]

2.2 Indirect Electrocatalytic Methods

This class of methods includes electrode processes with no electron transfer to or from the catalyst species. Interaction with the substrate occurs so as to decrease the activation energy of the electrode reaction and thus generates a new electrochemical signal with a distinct position on the potential axis.

2.2.1 Ligand-catalyzed Metal Ion Reduction

The catalytic reduction of some metal ions was first detected for Ni^{2+} or Co^{2+} by DC polarography in the presence of some aromatic diamines,[67] pyridine or its derivatives,[67,68] and aminothiols such as cysteine and its derivatives.[69–71]

The catalytic process induces a prewave whose half-wave potential may be hundreds of millivolts positive with respect to that of the main (irreversible) wave due to the hydrated metal ion, M^{n+} (Figure 3). It is obvious that a complex compound is reduced in the potential range of the prewave. The catalytic character of this process is demonstrated by the high i_k/i_d ratio, where i_k is the limiting current of the prewave and i_d denotes the expected diffusion current of the reducible complex compound. That is, the catalytic prewave occurs at a very low ligand/metal ratio, usually less than 0.1. The solution pH is an important parameter that decides the occurrence of the prewave via the formation of the reducible complex. Very often, the prewave is bell shaped (Figure 3).

An elementary description of the catalytic reduction of the metal ion is given by Equations (29) and (30) where L is the ligand catalyst (Scheme 2). The charges for L and the complex species are omitted for simplicity.

Provided that the metal ion is present in high excess and is not significantly depleted, the rate-determining

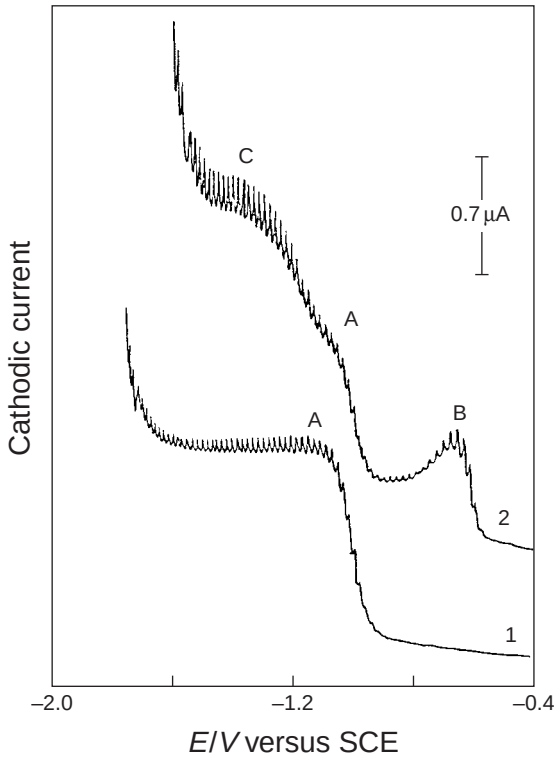

Figure 3 Catalytic nickel prewave (B) and catalytic hydrogen prewave (C) in the presence of cysteine. The background electrolyte is phosphate–acetate buffer, pH 6.5. Ni^{2+} concentration is 2×10^{-4} M (the nickel diffusion wave is indicated as (A)). Cysteine concentrations: (1) 0 M; (2) 6×10^{-5} M.

$$[ML] + ne^- \longrightarrow M(0) + L \quad (29)$$

$$L + M^{n+} \rightleftharpoons [ML] \quad (30)$$

Scheme 2

step is the regeneration of the reducible complex. This characteristic is demonstrated by the absence of any mass transfer effect on the prewave current. Experimentally, the absence of any effect of the mercury column pressure (in DC polarography[72]) or potential scan speed (in LSV on a stationary electrode[73,74]) is an indication of kinetic control by the chemical regeneration step.

This simple reaction scheme is often complicated by additional factors: ligand protonation; occurrence of a series of successive complexes; ligand adsorption (often under nonequilibrium conditions); the effect of electrode potential on the adsorption degree; and double-layer effects on the concentrations of charged reactants in the reaction layer. Attempts to set up a theoretical approach taking into account the above factors yielded only a partial description of the process.[67,68] However, a search for

systems showing simpler behavior enabled the catalytic current theory presented in section 2.1.2 to be used, by substituting [ML] for P, L for Q, and M^{n+} for A. This procedure is suitable in the absence of adsorption and if the formation of higher complexes is prevented. In this way, either DC polarography or LSV has been used for the determination of rate and equilibrium constants for the parallel, complex-forming reaction in the case of nickel reduction catalyzed by nicotinamide[72,73] or 6-mercaptopurine riboside.[74]

When the ligand adsorption has to be accounted for, two main problems arise. First, when using the dropping mercury electrode, the surface concentration is not accurately given by an adsorption isotherm because the equilibrium state cannot be achieved during the short lifetime of each drop. Conversely, if a hanging mercury electrode is used instead, the adsorption equilibrium could be attained by a prolonged contact with the solution at a given potential. However, during the potential scan, the equilibrium is disturbed by the shift in the value of the adsorption coefficient.[75] A favorable situation arises when the diffusion-controlled adsorption of the catalyst is the rate-determining step because, under such conditions, the kinetic equation relating the current to catalyst and reactant concentration can be derived.[71]

As far as the stoichiometry of the reducible complex is concerned, a direct correlation of the equilibrium data with the prewave current in the Ni^{2+}–penicillamine system demonstrated that the metal/ligand ratio is 1 : 1.[76] This value is similar to that determined in an indirect way for other ligands[71,72,74] and seems to be typical for the catalytic reduction of the metal ion.

The catalytic prewave of the metal ion enables sensitive determination of the ligand catalyst, which may be electrochemically inactive and not amenable to electrochemical determination by other methods. The sensitive determination of pyridine,[67] methionine,[77] or the alkaloid pilocarpine[78] in the presence of Ni^{2+} can be mentioned as examples. Noteworthy progress in this field has emerged by using catalytic reduction of the metal ion as a detection process in stripping voltammetry.[79–82] Detection limits down to 10^{-9} M can be achieved in this way for compounds showing catalytic properties in the reduction of Ni^{2+} or Co^{2+}.

Moreover, the dependence of catalytic activity on subtle features of the catalyst structure enables selective determination of the catalyst in the presence of related compounds. For example, *o*-phenylenediamine can thus be determined in the presence of its meta or para isomers by DC polarography.[67] Analogously, cysteinyl dipeptides have been determined in the presence of their constituent amino acids.[71] Cathodic stripping voltammetry allows the simultaneous determination of either cysteine or penicillamine (as a catalyst) and *N*-acetyl cysteine (a noncatalytic compound) in the presence of Ni^{2+}.[79,80]

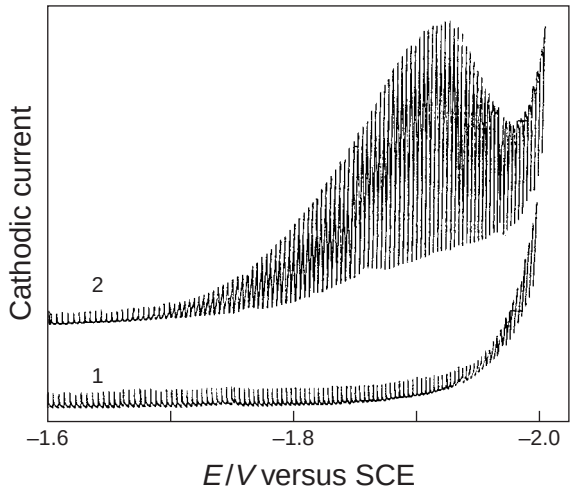

Figure 4 Catalytic hydrogen evolution on the dropping mercury electrode produced by an organic base. Curve 1 – background electrolyte alone (0.05 M borax, pH 9.2). Curve 2 – as for 1, with 3×10^{-6} M quinine as catalyst.

2.2.2 *Hydrogen Evolution on Mercury Electrodes Catalyzed by Organic Bases*

Hydrogen evolution on the mercury electrode occurs with an extremely high overvoltage. Some organic compounds (proteins, various alkaloids, pyridines and their derivatives, other heterocyclic nitrogen compounds) may reduce the hydrogen overvoltage giving a typical polarographic wave (Figure 4). Its current is much higher than any expected diffusion current in the system. This wave was ascribed a catalytic character[83] and is sometimes called a presodium wave, in view of the experimental circumstances that led to its discovery. The organic catalyst must be able to bind a proton, which subsequently enters the electrode reaction. That is why the catalytic wave obviously occurs in buffer solutions at a pH close to the pK_a of the catalyst.

Several interpretations of the catalytic hydrogen evolution by organic nitrogen bases have been proposed.[83] The most elaborate[84] postulates that the electron uptake by the protonated catalyst (BH^+) produces a radical, BH•. Thereafter, the radical dimerizes leading to regeneration of the catalyst in the nonprotonated form ($2BH\bullet \longrightarrow 2B + H_2$). The B• form is subsequently protonated via reaction with the acid component of the buffer. The bell-shaped form of the catalytic wave was explained by two opposite effects of the potential shift towards more negative values – an increase in the rate constant for the electron transfer reaction (according to the Butler–Volmer equation), and a drop in the surface concentration of the catalyst.

This interpretation was recently criticized from two standpoints.[85,86] First, it is stressed that the catalyst is a reduction product of the nitrogen base present in the solution and not the base itself. Second, a key step in the mechanism, i.e. dihydrogen production by radical dimerization, has never been proved directly by the methods of organic chemistry.

Despite the uncertain mechanism, this electrode process has proved useful in the very sensitive determination of various compounds of pharmaceutical relevance.[87] Catalytic hydrogen evolution has been used as a detection process in ASV for the determination of either anticancer alkaloids,[88] or ammonia (after reacting it with formaldehyde to give methyleneimine).[89]

2.2.3 *Platinum Metals Catalysis of Hydrogen Evolution on Mercury Electrodes*

The wave recorded in acidic solutions containing platinum metal ions was ascribed to hydrogen evolution on metal microcrystals that form on the mercury electrode surface by metal ion reduction.[83] The wave height increases in the sequence $Ru^{III} > Rh^{III} > Ir^{IV} > Pt^{IV} > Pd^{IV}$ and enables extremely sensitive determinations (e.g. $5-10^{-10}$ M ruthenium).

More recently it has been shown that some platinum complexes with organic ligands (such as ethylenediamine[90,91] or formazone[92,93]) behave analogously. Moreover, the organic ligand enables adsorptive accumulation of the metal ion in view of the stripping voltammetric determination with detection by means of the catalytic hydrogen current.[88,93] It is thus possible to detect Pt at the ultratrace level in biological, environmental and geological samples. The ligand itself, or a compound that generates the ligand by an appropriate chemical reaction, can also be determined in this way.[94] Analogous applications for other platinum group metals have been reported.

2.2.4 *Hydrogen Evolution on Mercury Electrodes Catalyzed by Transition Metal Complexes*

The most typical behavior of this feature is shown by sulfur-containing amino acids or proteins in the presence of cobalt ions in the ammonia buffer at pH about 9.5. The catalytic wave occurs in the potential range of the diffusion current for Co^{2+} reduction and has the shape of a round maximum in the case of the small molecular mass compounds (Figure 5). Under the same conditions, proteins give a particular double wave (Figure 6).[83,95–98] This kind of polarographic wave is often termed a Brdička wave after the name of the discoverer. Under the same conditions, Ni^{2+} gives lower catalytic currents. Analogous electrode processes may occur in the presence of other ligands containing sulfur, selenium, nitrogen or phosphorus.[97]

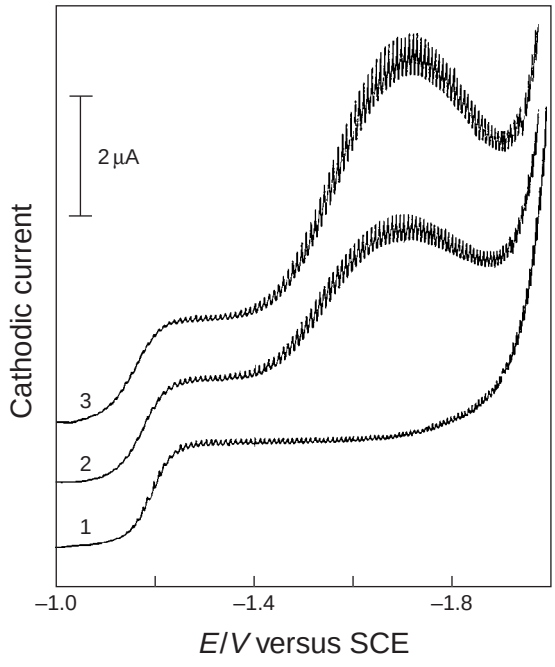

Figure 5 Catalytic hydrogen evolution on the dropping mercury electrode produced by Co^{2+} (5×10^{-4} M) and cysteine. Background electrolyte was NH_3, NH_4Cl (0.1 M each) +0.005% gelatine. Cysteine concentrations: (1) 0 μM; (2) 1 μM; (3) 2 μM.

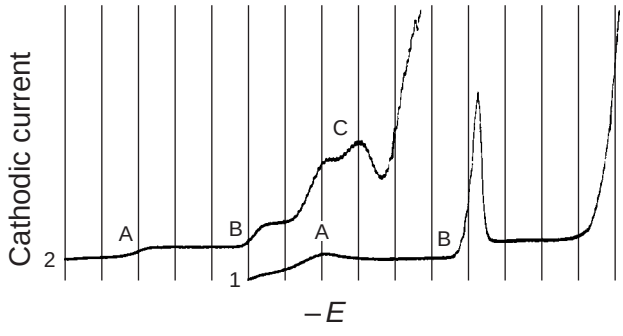

Figure 6 Brdička double wave of protein. (Curve 1) NH_3, NH_4Cl (0.1 M each) and $Co(NH_3)_6Cl_3$, 0.001 M. (Curve 2) 400 times diluted blood serum added. Electrochemical reactions are (A) $Co^{3+} \longrightarrow Co^{2+}$, (B) $Co^{2+} \longrightarrow Co$, and (C) catalytic double wave. Wave B on curve 1 shows a first-order polarographic maximum that is removed by protein adsorption in the case of curve 2. (Adapted from Heyrovský and Kůta.[83])

It is clear that the catalytic wave is due to a metal complex, which must be present at the electrode surface. A common feature is that the ligand is strongly polarizable and may accept electron pairs from the metal ion in an empty π or d orbital. In this way, unusual low-valence states of the metal ion (+1 or 0) may be stabilized. The catalytic complex is usually a chelate species, although this is not a general rule.

Two different assumptions have been made about the nature of the catalytic complex. The first assumes that this complex contains the metal ion in the normal oxidation state (2+), but its reduction potential is shifted to more negative values as an effect of complexation.[97] This complex is reduced in the potential range of the catalytic wave, simultaneously with catalytic hydrogen evolution. Consequently, the gradual decrease in the interface concentration of the catalytic complex produces a drop in the wave current after reaching a maximum (Figure 5).

The alternative interpretation assumes the catalyst to be a zero-charged metal complex, which results from the electroreduction of the normal complex present in the solution.[98] In this instance, the round shape of the catalytic hydrogen wave may be ascribed to the effect of the electrode potential on the surface concentration of the adsorbed organic ligand.[84] It is possible that either of the above mechanisms may be operative, depending on the properties of the ligand. As a common detail, both interpretations point out that the maximum current is not a true limiting current.

The occurrence of a double wave in the case of the proteins may be due to either the occurrence of two catalytic complexes with different properties[97] or to some changes in the adsorption layer due to the shift in the potential value.[84] Despite the complexity of the reaction mechanism, the effect of some experimental parameters on the protein double wave is fairly accounted for by the equations derived under the assumption of the diffusion-controlled catalyst adsorption.[99]

A more advanced decrease in hydrogen overvoltage on the mercury electrode occurs in the presence of Ni^{2+} and cysteine or related compounds, at pH 6–7.[100,101] The catalytic hydrogen prewave thus produced ($E_{1/2}$ about -1.2 V versus SCE; Figure 3) shows many distinct characteristics when compared with the Brdička wave and is much more sensitive to small changes in ligand structure. In this instance, the occurrence of hydride ion as an intermediate that coordinates the metal center is highly probable, in accord with recent representations of hydrogen evolution catalyzed by metal complexes.[102] The relevant electrode process was suggested as a model for hydrogenase-catalyzed hydrogen bioproduction.[101,103]

Catalytic hydrogen evolution enables very sensitive determination of the organic ligand, and various applications in biochemistry and medicine have been reported.[96] Much effort has been devoted to establishing a cancer test based on the Brdička wave.[95,96] Although such a method appears now to be of purely historic interest, the Brdička reaction still proves useful in some applications.[104–106] After being optimized for DPP, it now enables the determination of minute amounts of proteins (0.1–1 mg L^{-1}).[107] In this version, it is one of the standard methods for the determination of metallothionein.[108,109] The Brdička wave also enables the simultaneous determination of hemoglobin and cysteine.[110] Other sulfur-containing organic compounds may be similarly determined.[111,112]

The catalytic hydrogen evolution in the Co^{2+}–SH^{-} system has been employed as a detection process in cathodic stripping voltammetry of sulfide ion on the mercury electrode.[113] Some organic sulfur-containing compounds can be determined in the same way after the carbon–sulfur bond splits during the deposition step at a sufficiently positive potential.[114]

The catalytic wave is also useful for the determination of the metal ion itself, as a component of the catalytic complex. This method brings about a spectacular enhancement in sensitivity. Accordingly, the most important electroanalytical methods for nickel and cobalt determination are based on the catalytic hydrogen evolution in the presence of dimethylglyoxime.[115] The catalytic ligand may also enable adsorptive preconcentration of the metal ion on the hanging mercury drop electrode, in view of the stripping voltammetric determination, with detection by the catalytic evolution of hydrogen.[115,116]

Some metal complexes of practical importance (e.g. fungicides) can be determined by the catalytic wave induced by the constituent sulfur-containing ligand.[117] Some metal ions can also be determined indirectly, after performing the extraction as a chelate with a ligand that is able to produce the catalytic hydrogen wave (dithiocarbamates, xantogenates, 8-mercaptoquinoline).[117]

3 ELECTROCATALYSIS AT CHEMICALLY MODIFIED ELECTRODES

3.1 General Aspects

The properties of the electrode can be deliberately modified by immobilizing a reagent on its surface so that the modified electrode displays the properties of the immobilized compound. Reagent immobilization can be done by chemical (irreversible) adsorption, covalent bonding, or film deposition.[118,119] The composite electrodes, consisting of a mixture of conducting powder (e.g. carbon), a modifier and a hydrophobic binder, may also be included in this class. Depending on the modifier properties, the electrode may achieve various functions, such as preconcentration of the analyte, permselectivity, detection of some electroinactive species, or electrocatalysis.

As a research field, electrocatalysis at a CME is mostly directed towards mediated electron transfer via an immobilized redox couple. An immobilized redox couple may act as a shuttle for electron transfer between

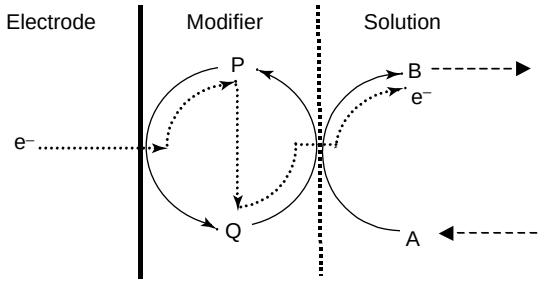

Figure 7 Diagram of mediated electron transfer at a CME for an oxidation process.

the electrode and some substrate, which would otherwise undergo a slow electrochemical reaction. This process is fairly described by Equations (5–7) with the important difference that the mediator is bound to the electrode surface and only the substrate and the products take part in mass transfer processes in the bulk of the solution (Figure 7). The catalyzed reaction often involves the transfer of several electrons, the breaking of some chemical bonds and the formation of new ones. In order to increase the reaction rate for such a complicated process, some intermediate must be stabilized by interaction with the modifier.

The thermodynamic requirement for electrocatalysis to occur is the same as for the homogeneous system (section 2.1.1). The catalyzed reaction proceeds at the formal potential of the mediator couple, provided that no catalyst–substrate adduct forms (Figure 8). The mediator must be able to undergo a very fast electron transfer

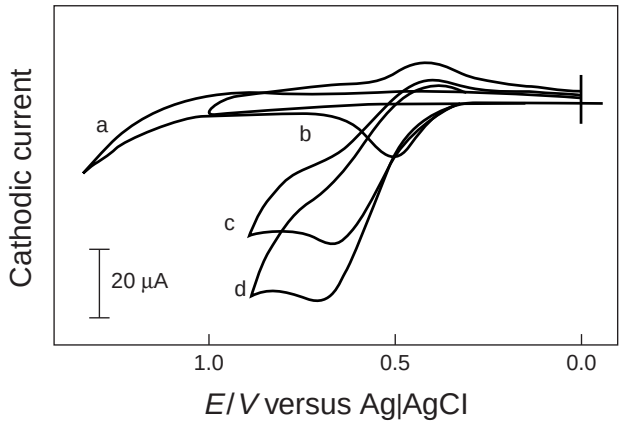

Figure 8 CV data demonstrating the electrocatalysis of hydroxylamine oxidation at a CME (glassy carbon electrode modified with a cobalt(II) hexacyanoferrate film). Background electrolyte, acetate buffer (pH 4.5), 0.1 M $NaNO_3$; scan rate, $0.1\,V\,s^{-1}$. (a) Bare carbon electrode, 9 mM NH_2OH. (b) Modified electrode, no NH_2OH (the CV peaks are due to the electrochemical processes involving the immobilized modifier). (c, d) As (b) with 6 and 9 mM NH_2OH, respectively. (Adapted from Chen.[133])

reaction, i.e. the reorganization of the coordination sphere should involve a small Gibbs energy change. This condition is fulfilled by complexes of transition metals (e.g. ferrocene and cyano complexes of iron and ruthenium) and organic molecules with a large π-bond system (e.g. phenoxazine, tetrathiafulvalinium tetracyanoquinodimethane (TCNQ)). A comprehensive description of various mediators with modified carbon electrodes is available.[120]

Metal powders dispersed in a conductive matrix are efficient catalysts for the electrooxidation of various organic compounds. Specific interaction with the substrate rather than mediated electron transfer may be the source of the catalytic activity in this instance.

3.2 The Modifier as a Thin Layer

Submonolayer or monolayer coverage is achieved by irreversible adsorption from solution or by placing a small, accurately measured volume of the modifier solution on the electrode surface and letting the solvent evaporate. This straightforward method suffers from a poor reproducibility, and leaching of the modifier slowly damages the resulting layer. It is still useful for a preliminary investigation of some modifiers.[118,119]

3.3 Electrocatalysis at Redox Polymer Modified Electrodes

3.3.1 Homogeneous Polymer Films

A redox polymer contains spatially localized redox sites. They can be bound covalently by being built into the chain or as a pendant group, such as osmium and ruthenium poly(pyridyl) polymers.[121] An alternative is electrostatic binding to an ion-exchange polymeric system, such as $Ru(bpy)_3^{3+/2+}$ included in a Nafion™ layer. In contrast to the monolayer, the polymer film forms a tridimensional structure containing up to $5 \times 10^{-6}\,mol\,cm^{-2}$ redox groups, providing a far greater number of reactive sites.

Most polymer modification procedures can be classified as either deposition of preformed polymers or electrochemical polymerization. Charge transport inside the polymer layer occurs by an electron exchange process between neighboring sites (self-exchange reactions, electron hopping, and redox conduction are alternative terms for this process). Simultaneously, transport of the co-ion in the opposite direction occurs to preserve electroneutrality, which is why the nature and the concentration of the background electrolyte may play a very important role in the charge transport process.

Redox catalysis at a CME occurs via a substrate–mediator electron-transfer reaction that may proceed either at the outer surface of the polymer film (surface reaction), or inside the coating with the substrate undergoing diffusion within the modifier layer.[122–124]

Two different theoretical models, leading to similar results, are commonly quoted. In the model of Albery and Hillman[125] two reaction layers are assumed. The first, located close to the film/electrolyte interface, is the zone where the substrate reaction with the mediator occurs. The other (near the electrode/film interface) corresponds to the electron transfer to (or from) the mediator. In the approach of Savéant et al.[126,127] no reaction layers are identified but the process is described in terms of four characteristic currents: substrate diffusion in the solution; substrate diffusion across the film; diffusion-like electron transfer across the film; cross-exchange between the redox center and substrate in the film. Steady-state methods (e.g. RDE[126] or UME[128,129]) provide the best conditions for the kinetic analysis of the catalyzed reactions at CMEs.

3.3.2 Microheterogeneous Polymer Layers

As previously discussed, the modifier film performs two parallel functions simultaneously, namely charge transport to the catalytic site and the catalytic activity. Conversely, a distinct component fulfills each function in the case of a microheterogeneous polymer. This provides more flexibility when designing a CME for a selected purpose. The coating may contain metal or metal oxide microparticles dispersed inside an electronically conducting polymer (polypyrrole, polyaniline), an ionically conducting polymer (Nafion™), or a redox polymer. The kinetic analysis of such a system must take into account two additional factors: substrate diffusion towards the catalyst particle within the matrix, and the charge transfer reaction at the particle surface.[122–124]

3.4 Electrodes Modified by Inorganic Films

Electrodes modified with sparingly soluble inorganic films promise better stability than polymer film CMEs.[130] For example, metallophthalocyanines and porphirines, as components of composite modified electrodes, are efficient catalysts for the oxidation of hydrazine, hydroxylamine, sulfhydryl compounds, α-keto acids and ascorbic acid[130,131] or for the reduction of oxygen, nitrate ion, and nitrogen oxide.[130]

Zeolites and clays show the particular advantage of performing size and shape discrimination. Although electrochemistry at zeolite and clay CMEs is a very dynamic field,[132] the analytical applications of electrocatalysis in such systems are still rather scarce.

Films of Prussian blue or its analogs ($KMe^{II}Fe^{III}(CN)_6$; Me = Fe, Co or Ni) can be formed on the electrode surface by electrolysis in a solution containing $Fe(CN)_6^{3-}$ and Me^{2+} ions. Such films display both intrinsic redox properties and also some size/shape exclusion characteristics. The oxidation of hydrazine, hydroxylamine, nitrite[133] and As^{III},[130] as well as reduction of O_2, H_2O_2 and Fe^{III},[130] are examples of electrochemical reactions catalyzed by such CMEs.

Other inorganic films with promising properties have been reviewed.[130] Lower metal oxides (PtOH, AuOH), nonstoichiometric, mixed-valence metal oxides (tungsten and molybdenum bronzes), and doped metal oxides show catalytic properties towards the oxidation of various organic compounds.

3.5 Carbon Paste Electrodes and Other Composite Electrodes

Graphite powder (5–20 μm particles) is mixed with an inert binder, such as paraffin or mineral oil, to form a paste that is typically 70% carbon by weight. The paste is packed into a shallow well in an inert holder with electrical contact at the back. A solid composite electrode may be prepared by using a filler like Kel-F™, Teflon®, epoxy resin or polyethylene.[134] To prepare a CME, a modifier reagent is either dissolved in the binder or physically mixed with the paste.[134,135] This is one of the most straightforward ways to prepare CMEs with highly reproducible properties.[134,136] Screen-printed electrodes, prepared by adding the catalyst to carbon ink,[137] can be also included in this class. A large diversity of inorganic (metal phthalocyanines, metalloporphyrins, hexacyanoferrates(III), metal powders, metal oxides and various complexes) or organic (phenoxazines, phenothiazines, quinone/hydroquinone) redox mediators are effective modifiers for carbon paste electrodes.

3.6 Electrocatalysis at Electronically Conducting Polymer Electrodes

In contrast to a redox polymer that contains distinct redox sites separated by nonconducting molecular fragments, an electronically conducting polymer (such as polypyrrole, polythiophene and their analogs) possesses extended π orbitals that facilitate continuous electron flow along the macromolecule. In some cases electrocatalytic activity is ascribed to the polymer itself, with no additional components. Thus the electrocatalytic mechanism may not be redox-mediated electron transfer, but the result of some favorable interaction between the substrate or some intermediates and the organic electrode surface. Some promising applications (for the detection of neurotransmitters, ascorbic acid, and other compounds of biological relevance) have been reviewed.[138]

3.7 Guidelines for Analytical Applications of Electrocatalysis at Chemically Modified Electrodes

From the analytical standpoint, electrocatalysis at CMEs can be used in two ways: the determination of the substrate by means of the catalyst-modified electrode; and for the determination of the catalyst by an electrochemical stripping method, after its accumulation on a properly designed CME.[139]

Substrate determination represents by far the most frequent application of CMEs for the analysis of various samples of environmental, biomedical and biotechnological relevance.[120,130,132,134,136–138,140–142]

Optimum conditions for electrocatalytic determination of the substrate at a CME are met when the rate-determining step is the mass transfer of the analyte in solution while charge propagation within the film is very fast.[142] A linear calibration graph results under such conditions and the kinetic complications are absent. Inorganic species appropriate for quantitation by such a method include oxygen, hydrogen peroxide, hydrazine, hydroxylamine, nitrogen and sulfur anions. Many more applications have been reported for organic and biochemical analyses (cysteine and derivatives, thiourea and derivatives, carbohydrates, polyalcohols, α-ketoacids).

Although CMEs have important advantages as sensors for batch analysis, the most promising applications are chemical sensors for field analysis, single-use sensors, and detectors for the analysis in flowing systems (flow injection analysis, liquid chromatography).[143] For detection in liquid chromatography, the problem of selectivity is not critical. Conversely, a low selectivity may be an advantage as it allows the detection of different compounds after performing the chromatographic separation. A serious problem with such detectors is the mechanical abrasion, but a permeable protecting coating over the catalytic layer can prevent this inconvenience.

As a result of their small size, modified UMEs have special advantages for microchemical-scale analysis, in vivo analysis,[144] or as detectors in capillary electrophoresis.[145]

4 AMPEROMETRIC ENZYMATIC DETERMINATIONS

4.1 Background

The outstanding specificity of enzymes as catalysts makes them very attractive for analytical purposes. Hence, enzymatic analysis is a well-established branch of analytical chemistry. Electrochemical methods can be used to monitor the enzymatic reaction occurring in the solution phase, but such a procedure is obviously not convenient owing to the consumption of the enzyme and other expensive reagents. The main interest in this field is therefore directed towards the development of electrochemical enzymatic sensors. The sensor consists of a reagent layer (including the enzyme) in close contact with a transducer that generates the measurable signal. The reagent layer performs the sensing (recognition) function by enzyme-catalyzed conversion of the substrate. Either potentiometric or amperometric transduction methods have been proposed, both showing some particular advantages. Whereas potentiometric transduction involves passive electrochemical devices, the amperometric version frequently relies on the integration of the electrochemical transducer in the electron transfer chain with the substrate at one end and the electrode at the other. At least one step in this dynamic process is an enzyme-catalyzed reaction, which justifies the classification of such procedures as electrocatalysis-based methods. Owing to the electrochemical nature of the transduction process, the most suitable enzymes belong to the oxido-reductase class, with glucose oxidase (GOD; EC 1.1.3.4) being a characteristic example.[146]

A simplified view of the function of an amperometric enzymatic sensor is described by Figure 9, for the particular case of substrate oxidation. Accordingly, the electron is transferred from the substrate S, to the cofactor $(CF)_{Ox}$, which is permanently or temporarily bound to the enzyme backbone. The reduced cofactor is thereafter reoxidized by a suitable electron acceptor, A_1.

Several transduction procedures have been proposed. The first involves the electrode reaction of the natural electron acceptor, A_1, or its product, A_2. Alternatively, an artificial mediator that shuttles the electrons from the prosthetic group of the enzyme to the electrode, through a regenerative process, may substitute the natural electron acceptor. Finally, it was attempted to perform the direct (nonmediated) electron transfer from the prosthetic group of the enzyme to the electrode. In this way, the electrode assumes the role of the cosubstrate A_1. It is sometimes claimed that only this latter case truly belongs to bioelectrocatalysis,[147,148] although such a sharp distinction in terminology overshadows some of

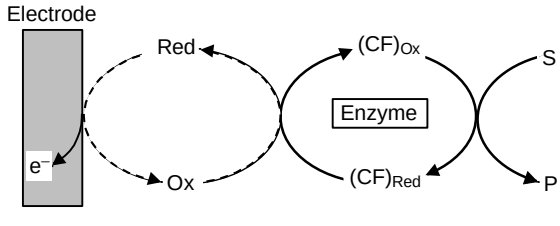

Figure 9 Simplified scheme of an enzymatic amperometric sensor.

the practical advantages. Any of the above three methods can be the basis for the transduction process.

4.2 Main Classes of Enzymes Used for Amperometric Sensors

Oxidases contain a redox cofactor that accepts electrons from the substrate and makes use of molecular oxygen as a reoxidation agent (electron and hydrogen ion acceptor, A_1 in the catalytic cycle). A typical example is the glucose sensor,[149] which is based on glucose oxidation by molecular oxygen catalyzed by GOD (Scheme 3). Here FAD (flavin adenine dinucleotide) represents the prosthetic group of GOD, formally acting as a cofactor.

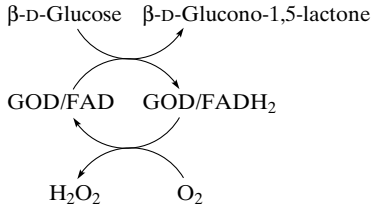

Scheme 3

The transduction may be performed by monitoring either the concentration of O_2 (cathodic reaction) or that of H_2O_2 (anodic reaction).[150] Very often, this kind of sensor is said to be unmediated, because no artificial mediator is involved. However, from a mechanistic standpoint, the term ''unmediated sensor'' is more appropriate for that involving direct electron transfer between enzyme and electrode. The detection of either O_2 or H_2O_2 enjoys outstanding practical applications, but suffers some drawbacks that arise from interference by the electroactive compounds that are usually present in biological samples (ascorbic acid, glutathione, uric acid). In addition, these procedures are dependent on the pH and the oxygen content in the sample, and require additional manipulations (such as sample dilution with an oxygen-saturated buffer). Hence, artificial mediators are alternatively used to shuttle the electrons from the enzyme to an electrode. A typical example is represented by ferrocene (bis(η^5-cyclopentadienyl) iron) or its derivatives, which can be localized on a graphite foil in order to shuttle electrons from GOD to the electrode (Figure 10).[151–153] Ferrocene electrochemical oxidation is a reversible process that occurs at a low electrode potential, thus preventing the disturbing electrode reaction.

Dehydrogenases are enzymes that carry out their catalytic function with the help of an electron acceptor other than O_2. Nicotinamide adenine dinucleotide, (NAD^+/NADH) or nicotinamide adenine dinucleotide phosphate ($NADP^+$/NADPH) are low molecular redox cofactors that can accept (or donate, respectively) two electrons and a proton. A distinct class includes dehydrogenases acting on −CHOH groups (such as lactate dehydrogenase or glucose-6-phosphate dehydrogenase), according to Equation (31):

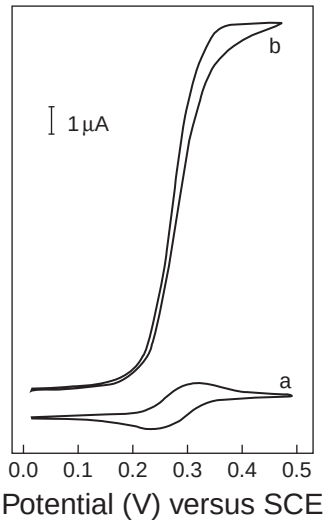

Figure 10 Electrocatalysis of glucose oxidation demonstrated by LSV at a graphite electrode in phosphate buffer (0.1 M, pH 7). (a) D-Glucose, 50 mM; ferrocenemonocarboxylic acid, 0.5 mM. (b) As for (a), but with the addition of 10.9 μM GOD.[151] The shape of the CV curve b is typical for a process with mediator regeneration as the rate-determining step and no substrate depletion (section 2.1.2.2).

$$NAD^+ + RR'CH{-}OH \longrightarrow NADH + RR'C{=}O + H^+ \quad (31)$$

Here, NAD(P) acts as a cosubstrate, i.e. it is temporarily bound to the enzyme molecule jointly with the substrate in order to perform the chemical conversion of the latter, whereas the redox regeneration of the factor occurs in the solution phase, after leaving the enzyme molecule. This characteristic represents a major advantage from the standpoint of coupling the enzymatic reaction with the electrochemical transduction process.

When using such enzymes, the transduction can be performed by the electrochemical oxidation of NAD^+, either by mediated electron transfer[152] or directly on organic salt electrodes.[152,154] Direct oxidation of NAD^+ on metal or carbon electrodes requires a high overvoltage and is disturbed by interference and secondary processes.

Over 250 enzymes belongs to this class, thus offering a wide range of analytical applications based on NAD(P) electrochemistry. The retention of the cofactor in the reaction layer, either by chemical immobilization or by an exclusion principle, is essential because of its solubility in water.

Table 2 Several versions of mono- and bienzyme sensors

Mechanism and designation	Analyte	Sensing process (enzyme reactions)	Transduction process	Response
Monoenzyme (sensor)	S	$S + A_1 \longrightarrow Pr + A_2$	$A_2 \pm ne \longrightarrow A_3$	Current increases with S concentration
			$A_1 \pm ne \longrightarrow A_3$	Current decreases with S concentration
Bienzyme (sequence sensor)	S	$S + A_1 \longrightarrow Z + A_2$ (enzyme 1) $Z + A_3 \longrightarrow Pr + A_4$ (enzyme 2)	$A_4 \pm ne \longrightarrow A_5$	Current increases with S concentration
Bienzyme (competition sensor)	Y	$S + A_1 \longrightarrow Pr + A_2$ (enzyme 1) $S + Y \longrightarrow A_3 + A_4$ (enzyme 2)	$A_2 \pm ne \longrightarrow A_5$	Current decreases with Y concentration
Bienzyme (substrate-recycling sensor)	S	$S + A_1 \longrightarrow Pr + A_2$ (enzyme 1) $Pr + A_3 \longrightarrow S + A_4$ (enzyme 2)	$A_2 \pm ne \longrightarrow A_5$	Current amplification due to substrate regeneration

Peroxidases are redox enzymes that can shuttle electrons to hydrogen peroxide (or small organic peroxides) from electron donors. The most common is horseradish peroxidase (HRP), which can accept electrons from virtually any reducing agent, such as ferrocyanide, phenol, *ortho*- and *para*-phenilenediamines, ascorbate, iodide, or ferrocene. The use of peroxidases in conjunction with hydrogen peroxide-producing oxidases is well established in analytical chemistry, including the field of amperometric biosensors.[141]

4.3 Configuration of the Enzymatic Amperometric Sensor

An enzyme amperometric sensor includes most of the features of a CME (section 3). The main problems when designing an enzyme sensor are concerned with the selection of the enzyme itself, the electrode material, additional reagents such as the mediator, and the procedures for the immobilization of enzyme and reagents. These problems are interrelated and must be dealt with in an integrative way, taking into account the expected figures of merit: sensitivity, the shape of the response function, the extent of the dynamic range, response time, and robustness.

Carbon is often the preferred electrode material, either in the form of carbon paste[155] or as a screen-printed electrode.[120,141,156] It is sometimes convenient to use a CME as a transducer instead of a plain electrode. For example, phthalocyanine-modified carbon electrodes are very convenient for the electrocatalytic detection of hydrogen peroxide produced in oxidase-catalyzed processes (e.g. glucose or uric acid sensors). Conducting organic salts, sometimes dubbed ''organic metals'', are the materials of choice for performing direct electron transfer to or from some biogenic compounds. Their use brings about a substantial simplification in sensor configuration.[154] Conducting organic salts are made by the combination of an electron donor with an acceptor. Such compounds are planar molecules with delocalized π-electron density, both above and below the molecular plane, and show a nonisotropic electrical conductance. A typical example in this class is the donor tetrathiafulvalene (TTF) combined with the acceptor TCNQ.

In addition to the obvious case of the monoenzyme sensor, inclusion of several enzymes is sometimes advantageous for interference removal (using an interferent-specific enzyme), mediated transduction using two mediators coupled with appropriate enzymes, indirect determination by competing enzyme reaction, or signal amplification by substrate recycling (Table 2).[157]

4.4 Immobilization Methods

Immobilization is a crucial step in preparing an enzymatic sensor. The purpose of this step is to produce an enzyme layer that is as thin as possible and to preserve the highest possible specific activity of the enzyme.

Any method of enzyme immobilization may be used to build up the sensing layer of an enzymatic sensor: immobilization behind a permselective membrane, covalent attachment to a membrane, entrapment in polymer gels, covalent cross-linking with bi- or multifunctional reagents, covalent coupling to a polymeric support or directly to the functionalized electrode surface.[158–160] The use of carbon as an electrode material has promoted immobilization into organic phase composite structures (carbon paste, carbon cement, carbon epoxy resins,[134,137,141,155] or screen-printed layers[141]). Either mono- or multilayers can be prepared by enzyme binding to the self-assembled monolayer of a functionalized thiol.[161,162]

Electropolymerized films[163–166] represent a very attractive way of preparing the enzyme layer as the polymers provide a favorable microenvironment for the enzyme. The preparation procedure is straightforward and easily monitored by the electrochemical parameters (sections 3.3 and 3.6). In addition, an electronically conducting polymer also can perform coupling of the electron transfer between the molecular species and the electrode.

Either entrapment of the enzyme during the electrochemical preparation of the polymer film or covalent binding to a functionalized polymer layer may be employed as the immobilization procedure. The second, multistep method is more flexible and offers more possibilities for securing the expected properties to the enzyme-containing modifier film. Alternatively, it is possible to perform electrochemical copolymerization of the plain monomer (e.g. pyrrole) and the enzyme-bound monomer.

Some changes in enzyme properties due to immobilization may occur. Hence, the evaluation of some parameters (such as recovery of enzyme activity, effectiveness factor, enzyme loading, sensor response and extension of the linear range, pH and temperature effects) enables the effectiveness of the immobilization method to be assessed.[160]

4.5 Redox Mediators for Amperometric Enzymatic Sensors

The use of a mediator makes the transduction process virtually independent of the ambient conditions and, simultaneously, shifts the working electrode potential to a convenient range, where the occurrence of a disturbing electrode reaction is less probable. An effective mediator is stable in both the oxidized and reduced forms; it should react rapidly with the enzyme and it should also exhibit a fast and reversible electrode reaction, with a low and pH-independent overpotential. The most frequently used artificial mediators include ferrocenes, hexacyanoferrate, phenoxazines, phenothiazines and phenazines, quinone/hydroquinone, TTF, and TCNQ.[120,146,152,155]

Electropolymerized films offer a very convenient alternative for the integration of redox mediators in the modifier layer. Negatively charged modifiers (ferro/ferricyanide, ferrocene carboxylate, and hydroquinonesulphonate) may be entrapped as counterions in electronically conducting polymers. The poor long-term stability of such systems, due to mediator leakage, has prompted the use of redox polymers (section 3.3) as the immobilization environment for the enzyme.[164,165] In this case, the polymer chain has to exhibit some flexibility so as to allow redox groups to reach an enzyme site.

4.6 Sensor Response: Modeling Enzyme Electrode Processes

A linear response over a broad concentration range is the ideal for any kind of sensor. Nevertheless, at least one step in the reaction sequence, namely the enzyme-catalyzed reaction, obeys a nonlinear kinetic equation and the situation is even more complicated in the case of the mediated electron transfer and multienzyme sensors. Consequently, careful optimization of sensor configuration, using both empirical and theoretical approaches, is essential for good performance. Generally, the electrode reaction, which represents the transduction process, is carried out with a sufficiently high overvoltage to make it very fast and kinetically irrelevant. The linear response occurs as a result of the mass transfer control of the overall reaction kinetics. Diffusion of the substrate or intermediate products through permselective membranes, or within the enzyme layer, could grant the appropriate mass transfer limitation.

The modeling of bioelectrocatalytic sensors is closely connected to the more general problem of mass transfer processes accompanied by reactions catalyzed by immobilized enzymes, a topic belonging to bioengineering.[167] Owing to the variety of sensor configurations, a general theoretical treatment is not possible. However, some general strategies have been outlined.[168–171]

In common with electrochemical kinetics, either mathematical analysis (using some numerical procedures) or digital simulation can be used to model the sensor process (see section 2.1.1). The most simple case is homogeneous mediation, which shows the characteristics of the EC′ mechanism (section 2.1) for the limiting case of very fast enzyme–substrate kinetics. This may be useful in preliminary investigations, but the actual sensor needs enzyme immobilization, and a series of additional factors must be considered: the number of enzyme reactions (Table 2); the number of layers (either reactive or protective); the kinetics of the enzyme reaction; the occurrence of either the steady state or transient behavior; the nature of the electroactive compound involved in the transduction process; and the required overvoltage. It is usually assumed that the diffusion proceeds along a single axis (planar diffusion) and obeys Fick's laws. It is clear that modeling the processes occurring in enzyme electrodes is not an easy task, but it provides substantial rewards – a better understanding of the mechanism and kinetics operating in the system, and a logical basis for the design and optimization of the sensor.

4.7 Examples of Applications

Amperometric enzymatic sensors represent the most successful class of biospecific sensors,[120,137,141,155,172,173] due mostly to the direct coupling of the recognition and transduction processes. They encompass the benefits of the dynamic electrode processes, such as linear response over a wide concentration range and the possibility of tuning the selectivity by the electrode potential.

The technological aspects of sensor production are an important area of research. It must be pointed out that an amperometric sensor is actually an electrochemical cell that needs, in addition to the sensing device, a reference electrode and, sometimes, an auxiliary electrode for operation in a three-electrode configuration. Modern

Table 3 Application of some commercial oxidase-based amperometric sensors using O_2 as the electron acceptor. (Adapted from J.R. Woodward, R.B. Spokane[177])

Analyte	Enzyme	Sample
Glucose	GOD	Blood, serum, plasma, dextrose in vegetables, ice cream, cereals, peanut butter
Sucrose	GOD/mutarotase/ invertase	Vegetables, ice cream, peanut butter, baked goods, cereal, effluent monitoring
Lactate		Blood, serum, plasma, spinal fluid; lunch meats, cooked foods
Lactose	Galactose oxidase	Cheese, whey
Ethanol, methanol	Alcohol oxidase	Alcohol in beers and wine; aspartame (after pretreatment with chymotrypsin)
Glutamate	Glutamate oxidase/ glutaminase	Glutamine in cell cultures
Choline	Choline oxidase	Infant formula, pet foods

microelectronics enables facile fabrication of the sensor, even for single-use application or at the micro scale, for in vivo measurements.[174,175] The use of microorganisms or tissue instead of isolated enzymes may achieve both operational and practical advantages.[155]

The applications of amperometric enzyme sensors encompass health care, the food and fermentation industries, defense, and environmental monitoring.[176] Glucose monitoring in blood is the first and still most widely used application, but other low-molecular-weight substances are currently determined by commercially available enzymatic sensors (Table 3).

5 CONCLUSIONS

Catalytic electrochemical processes may be used for analytical purposes either by adding the appropriate reagents to the sample solution or by including them in the structure of a CME. In the first case, the advantages of electrocatalysis are conveniently exploited using general-purpose devices such as the many versions of mercury electrodes. In the second case, it is possible to develop a specialized chemical sensor for field applications, single-use sensors, or detectors for the analysis in flowing fluids (flow injection analysis, liquid chromatography). Microdevices (UMEs, microchips) enable electrocatalysis-based sensors to perform analyses on very small sample volumes, in vivo analysis, or for detection in capillary electrophoresis. Analytical applications encompass a large variety of substances – from metal ions and simple inorganic species to various organic species, and many essential biological compounds.

LIST OF SYMBOLS

A	electrode area
$a = nFv/RT$	dimensionless sweep rate in LSV for a reversible electrode reaction
$b = \alpha nFv/RT$	dimensionless sweep rate in LSV for an irreversible electrode reaction
c	concentration of the species indicated by the subscript. The superscript "0" refers to bulk (equilibrium) concentration
D	diffusion coefficient
E	electrode potential (on the standard hydrogen electrode scale)
E^0	standard redox potential for the redox couple shown by the subscript
$E_{1/2}$	half-wave potential
E_e	equilibrium electrode potential
E_i	initial potential in LSV
F	Faraday constant
h	the height of the mercury level above the lower end of the mercury capillary electrode
i	electrolytic current
i_c	catalytic current
i_d	diffusion current
i_0	exchange current
i_k	limiting current of the metal ion catalytic prewave
k	rate constant for Equation (7)
k_1, k_2	homogeneous forward and backward rate constants, respectively, for the regeneration step
k_3	rate constant for the intramolecular electron transfer inside the catalyst–substrate adduct
K_a	acid ionization constant
K_{QA}	equilibrium constant for the formation of the catalyst–substrate adduct
k_s	conditional rate constant for the electrode reaction
n	stoichiometric coefficient of the electron in the electrode reaction
p	square-wave period in SWV
q	dimensionless potential sweep rate for LSV at a UME

R	gas constant
r_d	disc UME radius
r_s	spherical UME radius
T	absolute temperature
t	time variable
t_1	drop-time for the dropping mercury electrode
v	potential scan rate
α	transfer coefficient
δ	the thickness of the convection-free solution layer at electrode surface (diffusion layer)
χ_1	dimensionless kinetic parameter in DC polarography
Γ	surface coverage
η	overvoltage
Λ_s	dimensionless kinetic parameter for the case of the heterogeneous regeneration reaction with no substrate–catalyst adduct formation
Λ_m	dimensionless kinetic parameter for redox catalysis at a UME
ν	cinematic viscosity
Θ	dimensionless kinetic parameter for the case of the heterogeneous regeneration reaction with substrate–catalyst formation
τ	pulse duration in DPP; more generally, characteristic time constant for the electrochemical experiment
w	angular frequency of the modulating sine wave in AC polarography or angular velocity for the RDE

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
ASV	Adsorptive Stripping Voltammetry
CME	Chemically Modified Electrode
CV	Cyclic Voltammetry
DC	Direct Current
DPP	Differential Pulse Polarography
EC	Electrochemical Reaction with a Subsequent Chemical Reaction
EC′	Electrochemical Reaction with a Parallel Chemical Regeneration
EDTA	Ethylenediamine Tetraacetic Acid
FAD	Flavin Adenine Dinucleotide
GOD	Glucose Oxidase
HRP	Horseradish Peroxidase
LSV	Linear Sweep Voltammetry
NAD^+/NADH	Nicotinamide Adenine Dinucleotide
$NADP^+$/NADPH	Nicotinamide Adenine Dinucleotide Phosphate
RDE	Rotating Disc Electrode
SCE	Saturated Calomel Electrode
SWV	Square-wave Voltammetry
TCNQ	Tetracyanoquinodimethane
TTF	Tetrathiafulvalene
UME	Ultramicroelectrode

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Glucose, In Vivo Assay of

Biomolecules Analysis **(Volume 1)**
Voltammetry In Vivo for Chemical Analysis of the Living Brain • Voltammetry In Vivo for Chemical Analysis of the Nervous System

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Glucose Measurement • Urinalysis and Other Bodily Fluids

Environment: Water and Waste **(Volume 3)**
Heavy Metals Analysis in Seawater and Brines • Inorganic Environmental Analysis by Electrochemical Methods

Environment: Water and Waste cont'd **(Volume 4)**
Soil Instrumental Methods

Field-portable Instrumentation **(Volume 5)**
Microelectromechanical Systems Technology Applied to the Miniaturization of Field Instrumentation

Pharmaceuticals and Drugs **(Volume 8)**
Alkaloids, Pharmaceutical Analysis of

Steel and Related Materials **(Volume 10)**
Noble Metals, Analytical Chemistry of

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Neurotransmitters, Electrochemical Detection of • Organic Electrochemical Mechanisms • Pulse Voltammetry • Selective Electrode Coatings for Electroanalysis • Self-assembled Monolayers on Electrodes

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Catalytic Kinetic Determinations: Nonenzymatic • Enzymatic Kinetic Determinations

REFERENCES

1. G. Svehla, 'Nomenclature of Kinetic Methods of Analysis', *Pure. Appl. Chem.*, **65**(10), 2291–2298 (1993).
2. J.O'M. Bockris, K.N. Reddy, *Modern Electrochemistry*, Plenum Press, New York, 1141–1173, Vol. 2, 1970.
3. A.J. Appleby, 'Electrocatalysis', in *Comprehensive Treatise of Electrochemistry*, eds. B.E. Conway, J.O'M. Bockris, E. Yeager, S.U.M. Khan, R.E. White, Plenum Press, New York, 173–240, Vol. 7, 1983.
4. H.B. Mark, Jr, G.A. Rechnitz, *Kinetics in Analytical Chemistry*, Interscience, New York, 56–60, 1968.
5. H.A. Mottola, *Kinetic Aspects of Analytical Chemistry*, John Wiley, New York, 89–93, 1988.
6. J. Heyrovský, J. Kůta, *Principles of Polarography*, Academic Press, London, 380–393, 1966.
7. R. Brdička, V. Hanuš, J. Koutecky, 'General Theoretical Treatment of the Polarographic Kinetic Currents', in *Progress in Polarography*, eds. P. Zuman, I.M. Kolthoff, Interscience, New York, 145–199, Vol. 1, 1962.
8. R. Guidelli, 'Chemical Reactions in Polarography', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 149–374, Vol. 5, 1971.
9. M. Sluyters-Rechbach, J.H. Sluyters, 'Alternating Current and Pulse Methods', in *Comprehensive Chemical Kinetics*, eds. C.H. Bamford, R.G. Compton, Elsevier, Amsterdam, 317–354, Vol. 26, 1986.
10. C.P. Andrieux, J.M. Savéant, 'Electrochemical Reactions', in *Investigation of Rates and Mechanisms of Reactions. Part II*, ed. C.F. Bernasconi, John Wiley, New York, 305–390, 1986.
11. Z. Galus, *Fundamentals of Electrochemical Analysis*, Ellis Horwood, New York, 387–409, 1994.
12. A.J. Bard, L.J. Faulkner, *Electrochemical Methods. Fundamentals and Applications*, John Wiley, New York, 429–485, 1980.
13. O. Fischer, E. Fischerova, 'Voltammetric Methods for the Investigation of Chemical Kinetics', in *Experimental Techniques in Bioelectrochemistry*, eds. V. Brabec, D. Walz, G. Milazzo, Birkhäuser Verlag, Basel, 159–249, Vol. 3, 1996.
14. D.K. Gosser, Jr, *Cyclic Voltammetry: Simulation and Analysis of Reaction Mechanisms*, VCH, New York, 1993.
15. C. Nishihara, 'Note on Approximate Equations of DC Polarographic Current–Potential Curves for a Catalytic Wave', *J. Electroanal. Chem.*, **134**, 171–176 (1982).
16. J. Galvez, C. Serna, R. Saura, J. Zapata, 'Current–Potential Curves for a Catalytic Mechanism with Non-Nernstian Behavior', *J. Electroanal. Chem.*, **199**, 27–35 (1986).
17. J. Galvez, A. Molina, C. Serna, R. Saura, 'DC Polarography: Effects of Electrode Sphericity on the Catalytic Currents with Non-Nernstian Behavior', *J. Electroanal. Chem.*, **199**, 37–45 (1986).
18. A. Căluşaru, 'Investigation of Isotope Effects with the aid of Kinetic and Catalytic Polarographic Currents', *Russian Chem. Rev.*, **48**(2), 118–129 (1979).
19. J.M. Savéant, E. Vianello, 'Potential-sweep Chronoamperometry: Kinetic Currents for First-order Chemical Reaction Parallel to Electron-transfer Process (Catalytic Currents)', *Electrochim. Acta*, **10**, 905–920 (1965).
20. R.S. Nicholson, I. Shain, 'Theory of Stationary Electrode Polarography: Single Scan and Cyclic Methods Applied to Reversible, Irreversible, and Kinetic Systems', *Anal. Chem.*, **36**, 706–723 (1964).
21. L. Rampazzo, 'A Note on the Theory of Catalytic Currents in Linear Sweep Polarography. A Reversible Chemical Reaction (Pseudo-first Order) Parallel to the Electron Transfer Process', *J. Electroanal. Chem.*, **14**, 117–118 (1967).
22. M.-H. Kim, R. Birke, 'Differential Pulse Polarography for a First-order Catalytic Process', *Anal. Chem.*, **55**(6), 522–527 (1983).
23. J. Galvez, A. Molina, C. Serna, 'Pulse Polarography. VIII. Theory for Electrode Processes with First-order Regeneration Mechanisms', *J. Electroanal. Chem.*, **121**, 85–92 (1981).
24. J.J. O'Dea, J. Osteryoung, R.A. Osteryoung, 'Theory of Square Wave Voltammetry for Kinetic Currents', *Anal. Chem.*, **53**, 695–701 (1981).
25. J. Zeng, R.A. Osteryoung, 'Square-wave Voltammetry for a Pseudo-first-order Catalytic Process', *Anal. Chem.*, **58**, 2766–2771 (1986).
26. K.R. Bullock, D.E. Smith, 'Fundamental and Second-harmonic Alternating Current Polarography of Electrode Processes with Coupled First-order Catalytic Chemical Reactions: Theory and Experimental Results with the Iron Triethanolamine–Chlorite System', *Anal. Chem.*, **46**, 1567–1576 (1974).
27. C.M.A. Brett, A.M.C.F.O. Brett, 'Hydrodynamic Electrodes', in *Comprehensive Chemical Kinetics*, eds. C.H. Bamford, R.G. Compton, Elsevier, Amsterdam, 355–441, Vol. 26, 1986.
28. K. Aochi, K. Tokuda, H. Matsuda, 'Hydrodynamic Voltammetry at Channel Electrodes. I. Catalytic Currents', *J. Electroanal. Chem.*, **76**, 217–233 (1977).
29. R.G. Compton, A.C. Fischer, R.A. Spackman, 'Homogeneous Catalysis of Electrochemical Reactions. Channel Electrode Voltammetry and EC′ Mechanism', *Electroanalysis*, **4**(2), 167–182 (1992).
30. K. Aochi, K. Tokuda, H. Matsuda, 'Linear Sweep and Cyclic Voltammetry for Electrocatalysis at Modified

Electrodes with Very Thin Films', *J. Electroanal. Chem.*, **199**, 69–79 (1986).
31. Y. Xie, F.C. Anson, 'Analysis of the Cyclic Voltammetric Responses Exhibited by Electrodes Modified with Monolayers of Catalysts in the Absence and Presence of Substrate', *J. Electroanal. Chem.*, **384**, 145–153 (1995).
32. R.M. Wightman, D.O. Wipf, 'Voltammetry at Ultramicroelectrodes', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 267–353, Vol. 15, 1989.
33. M.I. Montenegro, 'Applications of Microelectrodes in Kinetics', in *Research in Chemical Kinetics*, eds. R.G. Compton, G. Hancock, Elsevier, Amsterdam, 1–80, Vol 2, 1994.
34. C. Amatore, 'Electrochemistry at Ultramicroelectrodes', in *Physical Electrochemistry*, ed. I. Rubinstein, Marcel Dekker, New York, 131–208, 1995.
35. G. Diao, Z. Zhang, 'The Theory of Catalytic Electrode Processes at a Hemispherical Ultramicroelectrode and its Applications for the Catalytic Behavior of the Sixth Reduction Wave of Fullerene, C_{60}', *J. Electroanal. Chem.*, **429**, 67–74 (1997).
36. K.B. Oldham, 'Steady-state Microelectrode Voltammetry as a Route to Homogeneous Kinetics', *J. Electroanal. Chem.*, **313**, 3–16 (1991).
37. M. Fleischmann, F. Lasserre, J. Robinson, D. Swan, 'The Application of Microelectrodes to the study of Homogeneous Processes Coupled to Electrode Reactions. I. EC′ and CE Reactions', *J. Electroanal. Chem.*, **177**, 97–114 (1984).
38. A. Molina, C. Serna, J. Gonzalez, 'General Analytical Solution for a Catalytic Mechanism in Potential Step Techniques at Hemispherical Microelectrodes: Applications to Chronoamperometry, Cyclic Staircase Voltammetry and Cyclic Linear Sweep Voltammetry', *J. Electroanal. Chem.*, **454**, 15–31 (1998).
39. G. Denuault, M. Fleischmann, D. Pletcher, D.R. Tutty, 'Development of the Theory for the Interpretation of Steady State Limiting Currents at a Microelectrode. EC′ Process: First and Second Order Reactions', *J. Electroanal. Chem.*, **280**, 243–254 (1990).
40. C.G. Phillips, 'The Steady, Diffusion-limited Current at a Disc Microelectrode with a First-order EC′ Reaction', *J. Electroanal. Chem.*, **296**, 255–258 (1990).
41. G.L. Che, S.J. Dong, 'Application of Ultramicroelectrodes in Studies of Homogeneous Catalytic Reactions. 2. A Theory of Quasi-1st and 2nd-order Homogeneous Catalytic Reactions', *Electrochim. Acta*, **37**(15), 2695–2699 (1992).
42. G.L. Che, S.J. Dong, 'Application of Ultramicroelectrodes in Studies of Homogeneous Catalytic Reactions. 3. The Condition for Quasi-1st and 2nd-order Homogeneous Catalytic Reactions at Ultramicrodisk Electrodes in the Steady-state', *Electrochim. Acta*, **37**(15), 2701–2705 (1992).
43. K. Aoki, M. Ishida, K. Tokuda, 'Voltammetry at Microcylinder Electrode. VI. Second-order Reaction of Fe(EDTA) with H_2O_2', *J. Electroanal. Chem.*, **245**, 39–50 (1988).
44. P.M. Zaitsev, S.I. Zhdanov, T.D. Nikolaeva, 'Catalytic Polarographic Currents of Oxidants', *Russ. Chem. Rev.*, **51**(6), 968–989 (1982).
45. Yu.S. Milyavskii, 'Choice of Conditions for Using Catalytic Polarographic Wave in Analytical Practice', *J. Anal. Chem. USSR*, **34**(9), 1293–1299 (1979).
46. H. Xia, H.L. Li, D.Q. Yang, 'Voltammetric Determination of Hydrazine Based on Catalytic Reaction in the Presence of 4-Hydroxy-2,2,6,6-tetramethylpiperdinyloxy (TEMPOL) Radical', *Electroanalysis*, **9**(18), 1429–1431 (1997).
47. J.M. Savéant, K.B. Su, 'Homogeneous Redox Catalysis of Electrochemical Reactions. VI. Diagram Representation of the Kinetic Regims', *J. Electroanal. Chem.*, **171**, 341–349 (1984).
48. J.F. Russling, 'Electrocatalytic Systems Organized by Micelles', *Trends Anal. Chem.*, **7**(7), 266–269 (1988).
49. H.B. Mark, Jr, G.A. Rechnitz, *Kinetics in Analytical Chemistry*, Interscience, New York, 232–317, 1968.
50. R.G. Wilkins, *Kinetics and Mechanism of Reactions of Transition Metal Complexes*, VCH, Weinheim, 257–297, 1991.
51. M.G. Paneli, A. Voulgaropoulos, 'Applications of Adsorptive Stripping Voltammetry in the Determination of Trace and Ultratrace Metals', *Electroanalysis*, **5**(5–6), 355–373 (1993).
52. A. Bobrowski, A.M. Bond, 'Exploitation of the Nitrite Catalytic Effect to Enhance the Sensitivity and Selectivity of the Adsorptive Stripping Voltammetric Method for the Determination of Cobalt with Dimethylglyoxime', *Electroanalysis*, **4**, 975–979 (1992).
53. J. Golimowski, P. Valenta, H.W. Nürnberg, 'Trace Determination of Chromium in Various Water Types by Adsorption Differential Pulse Voltammetry', *Fresenius Z. Anal. Chem.*, **322**(3), 315–322 (1985).
54. A.P. Aldrich, C.M.G. van den Berg, 'Determination of Iron and its Redox Speciation in Seawater Using Catalytic Cathodic Stripping Voltammetry', *Electroanalysis*, **10**(6), 369–373 (1998).
55. J. Shi, K. Jiao, 'Adsorptive Complex Catalytic Polarographic-determination of Germanium in Soils and Vegetables', *Anal. Chim. Acta*, **309**, 103–109 (1995).
56. Y. Sohrin, K. Isshiki, E. Nakayama, S. Kihara, M. Matsui, 'Simultaneous Determination of Tungsten and Molybdenum in Seawater by Catalytic Current Polarography after Preconcentration on a Resin Column', *Anal. Chim. Acta*, **218**, 25–35 (1989).
57. K. Yokoi, C.M.G. van den Berg, 'Determination of Titanium in Seawater Using Catalytic Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **245**, 167–176 (1991).

58. B. Ye, S. Yang, 'Adsorptive-catalytic Stripping Voltammetry for Determination of Ultratrace Titanium', *Talanta*, **41**(4), 537–540 (1994).
59. K. Yokoi, C.M.G. van den Berg, 'Simultaneous Determination of Titanium and Molybdenum in Natural-waters by Catalytic Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **257**, 293–299 (1992).
60. J. Wang, R. Setiadiji, 'Selective Determination of Trace Uranium by Stripping Voltammetry Following Adsorptive Accumulation of the Uranium–Cupferron Complex', *Anal. Chim. Acta*, **264**, 205–211 (1992).
61. J. Wang, B. Tian, J. Lu, 'Adsorptive Catalytic Stripping Measurements of Ultratrace Vanadium in the Presence of Cupferron and Bromate', *Talanta*, **39**(10), 1273–1276 (1992).
62. H. Hemmi, K. Hasebe, K. Ohzeki, T. Kambara, 'Differential Pulse Polarographic Determination of Nitrate in Environmental Materials', *Talanta*, **31**(5), 319–323 (1984).
63. M. Somasundrum, K. Kirtikara, M. Tanticharoen, 'Amperometric Determination of Hydrogen Peroxide by Direct and Catalytic Reduction at a Copper Electrode', *Anal. Chim. Acta*, **319**, 59–70 (1996).
64. H. Xia, H. Li, 'Catalytic Activity of 4-Hydroxy-2,2,6,6-tetramethyl-piperdin-1-oxyl in Homogeneous Electrooxidation of Hydroxylamine', *Collect. Czech. Chem. Commun.*, **63**, 1994–2000 (1998).
65. I. Helms, A. Möller, F. Scholz, 'Extraction-voltammetric Determination of Phosphorus', *Fresenius J. Anal. Chem.*, **349**, 575–577 (1994).
66. Z. Zhao, X. Cai, 'Determination of Trace Nitrite by Catalytic Polarography in Ferrous Iron Thiocyanate Medium', *J. Electroanal. Chem.*, **252**, 361–370 (1988).
67. H.B. Mark, Jr, 'Electrocatalysis and its Applications to Analysis. Plenary Lecture', *Analyst*, **115**(6), 667–678 (1990).
68. Ya.I. Tur'yan, 'Electric Catalytic Discharge of Metal Ions on Mercury Electrodes under Catalysis by Ligands', *J. Electroanal. Chem.*, **338**, 1–39 (1992).
69. J. Kůta, 'Die Ausnutzung der Adsoption in der Polarographische Analyse', *Z. Anal. Chem.*, **216**(2), 243–254 (1966).
70. A. Căluşaru, V. Voicu, 'Pre-vague Catalytique du Cobalt en Presence de Seleno-cystine', *J. Electroanal. Chem.*, **20**, 463–477 (1969).
71. F.G. Bănică, 'Polarographic Catalytic Nickel Prewave Induced by some Cysteine-containing Dipeptides', *Talanta*, **32**(12), 1145–1150 (1985).
72. F.G. Bănică, L.I. Doicin, A. Fodor, A. Ion, 'Ligand-catalysed Metal Ion Reduction. Kinetics and Analytical Applications of Nickel Reduction Catalysed by Nicotinamide', *Electroanalysis*, **6**, 1051–1056 (1994).
73. F.G. Bănică, 'Voltammetric Study of a Catalytic Electrode Process with Parallel Complexation Reaction: Reduction of Nickel Ion in the Presence of Nicotinamide', *Rev. Roum. Chim.*, **35**(6), 693–699 (1990).
74. A. Ion, F.G. Bănică, C. Luca, 'Ligand-catalysed Metal Ion Reduction. Voltammetric Determination of Rate and Formation Constants for the Nickel Complex with 6-Mercaptopurine-9-D-riboside', *J. Electroanal. Chem.*, **444**, 11–18 (1998).
75. S.G. Mairanovskii, 'Double Layer and Related Effects in Classical Polarography', in *Progress in Polarography*, eds. P. Zuman, L. Meites, Wiley-Interscience, New York, 287–357, Vol. III, 1972.
76. F.G. Bănică, A. Ion, 'Electrochemical Investigations of the Nickel(II)–Penicillamine System. II. Direct Identification of Complex Species Involved in Catalytic Nickel Reduction on the Dropping Mercury Electrode', *Collect. Czech. Chem. Commun.*, **63**, 995–1006 (1998).
77. J.M. Lopez Fonseca, M.C. Arredondo, 'Polarographic Determination of Methionine by Measurement of the Nickel(II) Catalytic Pre-wave', *Analyst*, **107**(8), 903–907 (1982).
78. J.M. Lopez Fonseca, D. Fojon, 'Polarographic Determination of Pilocarpine Using the Catalysed Prewave of Nickel(II)', *Anal. Chim. Acta*, **230**, 97–103 (1990).
79. F.G. Bănică, J.C. Moreira, A.G. Fogg, 'Application of Catalytic Stripping Voltammetry for the Determination of Organic Sulphur Compounds. Behavior of Cysteine, Cystine and N-Acetyl-cysteine in the Presence of Nickel Ion', *Analyst*, **119**(2), 309–318 (1994).
80. F.G. Bănică, A.G. Fogg, A. Ion, J.C. Moreira, 'Cathodic Stripping Voltammetry of Sulphur-containing Amino Acids and Peptides in the Presence of Nickel Ion. Catalytic and Inhibiting Effects', *Anal. Lett.*, **29**(8), 1415–1429 (1996).
81. A.G. Fogg, R. Ismail, R. Ahmad, F.G. Bănică, 'Cathodic Stripping Voltammetric Determination of 2-Mercaptobenzothiazole Using the Catalytic Cobalt Peak', *Talanta*, **44**, 491–495 (1997).
82. A. Ion, F.G. Bănică, C. Luca, 'Catalytic Cathodic Stripping Voltammetry of 6-Mercaptopurine-9-D-riboside in the Presence of Nickel Ion', *Electroanalysis*, **9**(12), 945–949 (1997).
83. J. Heyrovský, J. Kůta, *Principles of Polarography*, Academic Press, London, 407–428, 1966.
84. S.G. Mairanovskii, *Catalytic and Kinetic Waves in Polarography*, Plenum Press, New York, 241–285, 1968.
85. V.N. Leibson, A.P. Churilina, A.S. Mendkovich, V.P. Gultyai, 'New Ideas of the Mechanism of Catalytic Hydrogen Evolution in the Buffer Solution of Organic Compounds', *J. Electroanal. Chem.*, **261**, 165–176 (1989).
86. R.L. Birke, R.-A. Gu, J.-M. Yau, M.-H. Kim, 'Polarographic Catalytic Hydrogen Waves in Aquocobalamin and Methylcobalamin Solutions', *Anal. Chem.*, **56**(9), 1716–1722 (1984).
87. M. Březina, P. Zuman, *Polarography in Medicine, Biochemistry and Pharmacy*, Interscience, New York, 351–380, 1958.

88. J. Wang, J. Zadeii, M.S. Lin, 'Ultrasensitive Voltammetric Measurements Based on Coupling Hydrogen Catalytic Systems with Controlled Interfacial Accumulation of the Catalyst', *J. Electroanal. Chem.*, **237**, 281–287 (1987).
89. A.-M. Harbin, C.M.G. van den Berg, 'Determination of Ammonia in Seawater Using Catalytic Cathodic Stripping Voltammetry', *Anal. Chem.*, **65**(23), 3411–3416 (1993).
90. P.W. Alexander, R. Hoh, L.E. Smythe, 'Trace Determination of Platinum. I. A Catalytic Method Using Pulse Polarography', *Talanta*, **24**, 543–548 (1977).
91. O. Vrana, V. Kleinwächter, V. Brabec, 'Determination of Platinum in Urine by Differential Pulse Polarography', *Talanta*, **30**(4), 288–290 (1983).
92. Z. Zhao, H. Freiser, 'Differential Pulse Polarographic Determination of Trace Levels of Platinum', *Anal. Chem.*, **58**(6), 1498–1501 (1986).
93. C.M.G. Van den Berg, G.S. Jacinto, 'The Determination of Platinum in Seawater by Adsorptive Cathodic Stripping Voltammetry', *Anal. Chim. Acta*, **211**, 129–139 (1988).
94. J. Wang, Z. Taha, 'Catalytic-adsorptive Stripping Voltammetric Measurements of Hydrazines', *Talanta*, **35**(12), 965–968 (1988).
95. R. Brdička, M. Březina, V. Kalous, 'Polarography of Proteins and its Analytical Aspects', *Talanta*, **12**(12), 1149–1162 (1965).
96. M. Březina, P. Zuman, *Polarography in Medicine, Biochemistry and Pharmacy*, Interscience, New York, 585–683, 1958.
97. A. Căluşaru, 'Décharge Catalytique de l'Hydrogène en Présence de Cobalt et de Nickel', *J. Electroanal. Chem.*, **15**, 269–307 (1967).
98. P. Mader, 'Polarographic Investigation of Systems Containing Cobalt (II) and Cysteine-like Compounds with Special Reference to their Ability to Facilitate Hydrogen Evolution from Buffered Aqueous Solutions', *Collect. Czechoslov. Chem. Commun.*, **36**, 1035–1064 (1971).
99. M. Senda, T. Ikeda, K. Kano, I. Tokimitsu, 'Theory of Brdička Current and the Dependence of the Current on Cobalt Ion Concentration', *Bioelectrochem. Bioenerg.*, **9**, 253–263 (1982).
100. A. Căluşaru, V. Voicu, 'Catalytic Hydrogen Prewave in the Presence of Seleno-cystine and Nickel', *J. Electroanal. Chem.*, **43**, 257–266 (1973).
101. F.G. Bănică, 'Catalytic Hydrogen Evolution on the Mercury Electrode in the Presence of Nickel–Aminothiol Complexes. A Tentative Model Process of Hydrogenase-catalysed Hydrogen Bioproduction', *Bull. Soc. Chim. Fr.*, **128**, 697–703 (1991).
102. I. Bhugun, D. Lexa, J.-M. Savéant, 'Homogeneous Catalysis of Electrochemical Hydrogen Evolution by Iron(0) Porphyrins', *J. Am. Chem. Soc.*, **118**(16), 3982–3983 (1996).
103. M.A. Halcrow, G. Christou, 'Biomimetic Chemistry of Nickel', *Chem. Rev.*, **94**(8), 2421–2481 (1994).
104. M. Senda, T. Ikeda, T. Kakutani, K. Kano, 'Polarographic Studies on Proteins', *Bioelectrochem. Bioenerg.*, **8**, 151–165 (1981).
105. V. Brabec, V. Vetterl, O. Vrana, 'Electroanalysis of Biomacromolecules', in *Experimental Techniques in Bioelectrochemistry*, eds. V. Brabec, D. Walz, G. Milazzo, Birkhäuser Verlag, Basel, 287–359, Vol. 3, 1996.
106. J.M. Séquaris, 'Analytical Voltammetry of Biological Molecules', in *Analytical Voltammetry*, eds. M.R. Smith, J.G. Vos, Elsevier, Amsterdam, 115–157; 138–140, 1992.
107. E. Palecek, Z. Pechan, 'Estimation of Nanogram Quantities of Proteins by Pulse-polarographic Technique', *Anal. Biochem.*, **42**, 59–67 (1971).
108. B. Raspor, J. Pavicic, 'Electrochemical Methods for Quantification and Characterization of Metallothioneins Induced in *Mytilus galloprovincialis*', *Fresenius J. Anal. Chem.*, **354**, 529–534 (1996).
109. R. Wagemann, J.G. Dick, J.F. Klaverkamp, 'Metallothionein Estimates in Marine Mammal and Fish Tissues by Three Methods: ^{203}Hg Displacement, Polarography and Metal-summation', *Int. J. Environ. Anal. Chem.*, **54**(2), 147–160 (1994).
110. M. Kuik, K. Krassowski, 'Simultaneous Determination of Hemoglobin and Cysteine Based on Brdička Catalytic Current', *Collect. Czechoslov. Chem. Commun.*, **56**, 157–166 (1991).
111. A. Căluşaru, 'Hydrogen Catalytic Waves for the Polarographic Determination of 2-Mercaptobenzothiazole in the Presence of Cobalt(II)', *Analyst*, **115**(9), 1223–1228 (1990).
112. N. Spătaru, A. Căluşaru, 'Polarographic Determination of Diethyldithiocarbamate Using Kinetic and Catalytic Processes', *Electroanalysis*, **7**(6), 583–586 (1995).
113. F.G. Bănică, N. Spătaru, T. Spătaru, 'Catalytic Hydrogen Evolution in the Presence of Sulfide and Cobalt Ions. A Study by Cathodic Stripping Voltammetry on the Hanging Mercury Drop Electrode', *Electroanalysis*, **9**, 1341–1347 (1997).
114. F.G. Bănică, N. Spătaru, 'Catalytic Hydrogen Evolution in the Presence of Methylthiohydantoin-glycine and Cobalt(II) ion. A Study by Cathodic Stripping Voltammetry at a Hanging Mercury Drop Electrode', *Talanta*, **48**, 491–494 (1999).
115. S.G. Mairanovskii, G.V. Prokhorova, E.A. Osipova, 'On the Nature of Polarographic Waves in Nickel and Cobalt Dimethylglyoxime Solutions', *J. Electroanal. Chem.*, **266**, 205–214 (1989).
116. M.G. Paneli, A. Voulgaropoulos, 'Determination of Ni and Co Using 2-Quinolinethiol by Adsorptive Voltammetry', *Fresenius J. Anal. Chem.*, **341**, 716–719 (1991).
117. V.F. Toropova, H.C. Budnikov, N.A. Ulakhovich, E.P. Medyantseva, 'Catalytic Evolution of Hydrogen on a Mercury Electrode in Cyclic Voltammetry and Its

Analytical Applications', *J. Electroanal. Chem.*, **144**, 1–10 (1983).
118. R.V. Murray, 'Chemically Modified Electrodes', in *Electroanalytical Chemistry*, ed. A.J. Bard, Marcel Dekker, New York, 191–367, Vol. 13, 1984.
119. G.G. Wallace, 'Chemically Modified Electrodes', in *Chemical Sensors*, ed. T.E. Edmonds, Blackie, London, 132–154, 1988.
120. S.A. Wring, J.P. Hart, 'Chemically Modified, Carbon-based Electrodes and their Application as Electrochemical Sensors for the Analysis of Biologically Important Compounds', *Analyst*, **117**(8), 1215–1229 (1992).
121. D.M. Kelly, J.G. Vos, 'Osmium and Ruthenium Poly(pyridyl) Redox Polymers as Electrode Coatings', in *Electroactive Polymer Electrochemistry. Part 2: Methods and Applications*, ed. M.E.G. Lyons, Plenum Press, New York, 173–232, 1996.
122. R.J. Mortimer, 'Dynamic Processes in Polymer Modified Electrodes', in *Research in Chemical Kinetics*, eds. R.G. Compton, G. Hancock, Elsevier, Amsterdam, 261–311, Vol. 2, 1994.
123. M.E.G. Lyons, 'Electrocatalysis Using Electroactive Polymer Films', in *Electroactive Polymer Electrochemistry. Part 1: Fundamentals*, ed. M.E.G. Lyons, Plenum Press, New York, 237–374, 1994.
124. M.E.G. Lyons, 'Electrocatalysis Using Electroactive Polymers, Electroactive Composites and Microheterogeneous Systems', *Analyst*, **119**(5), 805–826 (1994).
125. W.J. Albery, A.R. Hillman, 'Transport and Kinetics in Modified Electrodes', *J. Electroanal. Chem.*, **170**, 27–49 (1984).
126. C.P. Andrieux, J.M. Savéant, 'Kinetics of Electrochemical Reactions Mediated by Redox Polymer Films. Irreversible Cross-exchange Reactions: Formulation in Terms of Characteristic Currents for Stationary Techniques', *J. Electroanal. Chem.*, **134**, 163–166 (1982).
127. C.P. Andrieux, J.-M. Dumas-Bouchiat, J.M. Savéant, 'Kinetics of Electrochemical Reactions Mediated by Redox Polymer Films. New Formulation and Strategies for Analysis and Optimization', *J. Electroanal Chem.*, **169**, 9–21 (1984).
128. G. Che, S. Dong, 'Electrochemical Catalysis at an Ultramicroelectrode Modified with Redox Species', *Electrochim. Acta*, **38**(4), 581–588 (1993).
129. M.E.G. Lyons, T. Bannon, S. Rebouillat, 'Reaction/diffusion at Conducting Polymer Ultramicroelectrodes', *Analyst*, **123**, 1961–1966 (1988).
130. J.A. Cox, R.K. Javorski, P.J. Kulesza, 'Electroanalysis with Electrodes Modified by Inorganic Films', *Electroanalysis*, **3**, 869–877 (1991).
131. J.H. Zagal, 'Metallophthalocyanines as Catalysts in Electrochemical Reactions', *Coord. Chem. Rev.*, **119**, 89–136 (1992).
132. A. Walcarius, 'Zeolite-modified Electrodes: Analytical Applications and Prospects', *Electroanalysis*, **8**(11), 971–986 (1996).
133. S.-M. Chen, 'Characterization and Electrocatalytic Properties of Cobalt Hexacyanoferrate Films', *Electrochim. Acta*, **43**(21–22), 3359–3369 (1998).
134. K. Kalcher, J.-M. Kauffmann, J. Wang, I. Svancara, K. Vytras, C. Neuhold, Z. Yang, 'Sensor Based on Carbon Paste in Electrochemical Analysis: A Review with Particular Emphasis on the Period 1990–1993', *Electroanalysis*, **7**(1), 5–22 (1995).
135. R.L. McCreery, K.K. Cline, 'Carbon Electrodes', in *Laboratory Techniques in Electroanalytical Chemistry*, eds. P.T. Kissinger, W.R. Heineman, Marcel Dekker, New York, 293–332, 1996.
136. K. Kalcher, 'Chemically Modified Carbon Paste Electrodes in Voltammetric Analysis', *Electroanalysis*, **2**, 419–433 (1990).
137. J.P. Hart, S.A. Wring, 'Screen-printed Voltammetric and Amperometric Electrochemical Sensors for Decentralized Testing', *Electroanalysis*, **6**(8), 617–624 (1994).
138. H.B. Mark, N. Atta, Y.L. Ma, K.L. Petticrew, H. Zimmer, Y. Shi, S.K. Lunsford, J.F. Rubinson, A. Galal, 'The Electrochemistry of Neurotransmitters at Conducting Organic Polymer Electrodes – Electrocatalysis and Analytical Applications', *Bioelectrochem. Bioenerg.*, **38**(2), 229–245 (1995).
139. Z.Q. Gao, K.S. Siow, A. Ng, 'Catalytic Voltammetric Determination of Molybdenum at a Chemically Modified Carbon Paste Electrode', *Electroanalysis*, **8**(12), 1183–1187 (1996).
140. D. Leech, 'Analytical Applications of Polymer-modified Electrodes', in *Electroactive Polymer Electrochemistry. Part 2: Methods and Applications*, ed. M.E.G. Lyons, Plenum Press, New York, 269–296, 1996.
141. M.A.T. Gilmartin, J.P. Hart, 'Sensing with Chemically and Biologically Modified Carbon Electrodes. A Review', *Analyst*, **120**(4), 1029–1045 (1995).
142. W. Kutner, J. Wang, M. L'Her, R.P. Buck, 'Analytical Aspects of Chemically Modified Electrodes: Classification, Critical Evaluation and Recommendations', *Pure. Appl. Chem.*, **70**(6), 1301–1318 (1998).
143. R.P. Baldwin, K.N. Thomsen, 'Chemically Modified Electrodes in Liquid Chromatography Detection: A Review', *Talanta*, **38**(1), 1–16 (1991).
144. R.D. Oneill, 'Microvoltammetric Techniques and Sensors for Monitoring Neurochemical Dynamics in vivo – a Review', *Analyst*, **119**(5), 767–779 (1994).
145. T.J. O'Shea, S.M. Lunte, 'Chemically-modified Microelectrodes for Capillary Electrophoresis Electrochemistry', *Anal. Chem.*, **66**(2), 307–311 (1994).
146. J. Raba, H.A. Mottola, 'Glucose Oxidase as an Analytical Reagent', *Crit. Rev. Anal. Chem.*, **25**(1), 1–42 (1995).
147. M.R. Tarasevich, 'Bioelectrocatalysis', in *Comprehensive Treatise of Electrochemistry*, eds. S. Srinivasan, Yu.A. Chizmadzhev, J.O'M. Bockris, B.E. Conway, E. Yeager, Plenum Press, New York, 231–296, Vol. 10, 1985.

148. A.L. Ghindilis, P. Atanasov, E. Wilkins, 'Enzyme-catalyzed Direct Electron Transfer: Fundamentals and Analytical Applications', *Electroanalysis*, **9**(9), 661–674 (1997).
149. L.C. Clark, Jr, C. Lyons, 'Electrode Systems for Continuous Monitoring in Cardiovascular Surgery', *Ann. NY Acad. Sci.*, **102**, 29–45 (1962).
150. G.S. Wilson, D.R. Thévenot, 'Unmediated Amperometric Enzyme Electrodes', in *Biosensors: A Practical Approach*, ed. A.E.G. Cass, Oxford University Press, Oxford, 1–17, 1990.
151. A.E.G. Cass, G. Davis, G.D. Francis, H.A.O. Hill, W.J. Aston, I.J. Higgins, E.V. Plotkin, L.D.L. Scott, A.P.F. Turner, 'Ferrocene-mediated Enzyme Electrode for Amperometric Determination of Glucose', *Anal. Chem.*, **56**(4), 667–671 (1984).
152. M.F. Cardosi, A.P.F. Turner, 'The Realization of Electron Transfer from Biological Molecules to Electrodes', in *Biosensors: Fundamentals and Applications*, eds. A.P.F. Turner, I. Karube, G.S. Wilson, Oxford University Press, Oxford, 257–275, 1987.
153. H.A.O. Hill, G.S. Sanghera, 'Mediated Amperometric Enzyme Electrodes', in *Biosensors: A Practical Approach*, ed. A.E.G. Cass, Oxford University Press, Oxford, 19–46, 1990.
154. P.N. Bartlett, 'Conducting Organic Salt Electrodes', in *Biosensors: A Practical Approach*, ed. A.E.G. Cass, Oxford University Press, Oxford, 47–95, 1990.
155. L. Gorton, 'Carbon Paste Electrodes Modified with Enzymes, Tissues, and Cells', *Electroanalysis*, **7**(1), 23–45 (1995).
156. A. Schmidt, I. Rohm, P. Rüger, W. Weise, U. Bilitewski, 'Application of Screen Printed Electrodes in Biochemical Analysis', *Fresenius J. Anal. Chem.*, **349**, 607–612 (1994).
157. F. Scheller, F. Schubert, *Biosensors*, Elsevier, Amsterdam, 184–211, 1992.
158. L. Goldstein, G. Manecke, 'The Chemistry of Enzyme Immobilization', in *Applied Biochemistry and Bioengineering*, eds. L.B. Wingard, Jr, E. Katchalski-Katzir, L. Goldstein, Academic Press, New York, 23–126, 1976.
159. E.A.H. Hall, *Biosensors*, Open University Press, Milton Keynes, 216–267, 1990.
160. F. Scheller, F. Schubert, *Biosensors*, Elsevier, Amsterdam, 50–65, 1992.
161. Th. Wink, S.J. van Zuilen, A. Bult, W.P. van Bennekom, 'Self-assembled Monolayers for Biosensors', *Analyst*, **122**(4), R43–R50 (1997).
162. I. Willner, E. Katz, B. Willner, 'Electrical Contact of Redox Enzyme Layers Associated with Electrodes: Routes to Amperometric Biosensors', *Electroanalysis*, **9**(13), 965–977 (1997).
163. P.N. Bartlett, J.M. Cooper, 'A Review of the Immobilization of Enzymes in Electropolymerized Films', *J. Electroanal. Chem.*, **362**, 1–12 (1993).
164. W. Schuhmann, 'Conducting Polymer Based Amperometric Enzyme Electrodes', *Microchim. Acta*, **121**, 1–29 (1995).
165. P.N. Bartlett, J.M. Cooper, 'Applications of Electroactive Polymers in Bioelectrochemistry and Bioelectronics', in *Electroactive Polymer Electrochemistry. Part 2: Methods and Applications*, ed. M.E.G. Lyons, Plenum Press, New York, 233–266, 1996.
166. E.I. Iwuoha, M.R. Smith, 'Polymer-based Amperometric Biosensors', in *Electroactive Polymer Electrochemistry. Part 2: Methods and Applications*, ed. M.E.G. Lyons, Plenum Press, New York, 297–328, 1996.
167. J.-M. Engasser, C. Horvath, 'Diffusion and Kinetics with Immobilized Enzymes', in *Applied Biochemistry and Bioengineering*, eds. L.B. Wingard, Jr, E. Katchalski-Katzir, L. Goldstein, Academic Press, New York, 127–220, 1976.
168. W.J. Albery, D.H. Craston, 'Amperometric Enzyme Electrodes: Theory and Experiment', in *Biosensors: Fundamentals and Applications*, eds. A.P.F. Turner, I. Karube, G.S. Wilson, Oxford University Press, Oxford, 180–210, 1987.
169. P.N. Bartlett, P. Tebbutt, R.G. Whitaker, 'Kinetic Aspects of the use of Modified Electrodes and Mediation in Bioelectrochemistry', *Prog. React. Kinet.*, **16**, 55–155 (1991).
170. F. Scheller, F. Schubert, *Biosensors*, Elsevier, Amsterdam, 66–82, 1992.
171. P.N. Bartlett, K.F.E. Pratt, 'Modeling of Processes in Enzyme Electrodes', *Biosensors and Bioelectronics*, **8**(9–10), 451–462 (1993).
172. J. Kas, M. Marek, M. Stastny, R. Volf, 'Biosensors with Electrochemical Transducers', in *Experimental Techniques in Bioelectrochemistry*, eds. V. Brabec, D. Walz, G. Milazzo, Birkhäuser Verlag, Basel, 361–453, Vol. 3, 1996.
173. F. Scheller, F. Schubert, *Biosensors*, Elsevier, Amsterdam, 291–322, 1992.
174. G.A. Urban, G. Jobst, 'Sensor Systems', in *Microsystems Technology in Chemistry and Life Science*, eds. A. Manz, H. Becker, Springer-Verlag, Berlin, 189–213, Vol. 194, 1998.
175. M. Lambrechts, W. Sansen, *Biosensors: Microelectrochemical Devices*, Institute of Physics Publishing, Bristol, 1992.
176. G. Ramsey (ed.), *Commercial Biosensors: Application to Clinical, Bioprocess and Environmental Samples*, John Wiley, New York, 1998.
177. J.R. Woodward, R.B. Spokane, 'Applications of Biosensor-based Instruments to the Bioprocess Industry', in *Commercial Biosensors: Application to Clinical, Bioprocess and Environmental Samples*, ed. G. Ramsey, John Wiley, New York, 227–264, 1998.

Enzymatic Kinetic Determinations

Regina Hüttl
Technische Universität Bergakademie Freiberg, Germany

1 Introduction **1**
2 Features of Enzymatic Reactions **2**
2.1 High Specificity of Enzymes 2
2.2 Temperature 2
2.3 pH Value 2
2.4 Cofactors 2
2.5 Ionic Strength 2
3 Reaction Mechanism and Kinetic Equations **2**
3.1 Michaelis–Menten Equation 3
3.2 Integrated Michaelis–Menten Equation 3
3.3 Michaelis Constant and Enzyme Activity 4
3.4 Determination of Kinetic Parameters 5
4 Modifying the Kinetic by Effectors **6**
4.1 Activators 6
4.2 Inhibitors 6
5 Enzymatic Determinations **7**
5.1 Determination of Substrates 7
5.2 Determination of Enzymes 9
5.3 Determinations of Effectors 11
6 Immobilized Enzymes for Determinations in Flow Systems **13**
7 Immobilized Enzymes in Biosensor Development **15**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Kinetic methods for analytical applications can be divided into noncatalytic and catalytic methods. Enzyme kinetic determinations are a special type of catalytic methods. Enzymatic reactions are used analytically to determine enzyme activities – for instance in the diagnosis of diseases, substrate concentrations in food industry or medicine and concentrations of effectors – for example in trace analysis. Due to the high selectivity of enzymatic analysis the importance of this method in various fields is growing rapidly. The widest use is observed in clinical chemistry and in food chemistry.

This article can only give a general survey of the great potential of enzymatic analysis, with references to more detailed literature.

Enzymatic reactions are characterized by some special features, explained after the introduction. The theoretical background for kinetic determination is the comprehension of the principles on enzymatic kinetics. In section 3 the simplest mechanism, outlined by Michaelis and Menten and several transformations of this equation will be described. In this section two kinetic parameters, enzyme activity and the Michaelis constant, will be introduced. Section 4 deals with substances, so called effectors, influencing the rate of an enzymatic reaction. Simple mechanisms of these effectors will be shown. The different methods for determination of substrates, enzymes and effectors, illustrated by many examples, are represented in section 5. While substrate concentrations can be determined using both ways, equilibrium or kinetic methods, the determination of enzymes and effectors can only be performed by kinetic approaches. In the last two sections the application of immobilized enzymes in flow systems and biosensor development will be shown.

1 INTRODUCTION

Kinetic methods are useful tools in analytical chemistry. The use of enzymes in kinetic-based determinations is widely applied for the analysis of many substances, for instance in medicine or the food industry, for the analysis of enzymes, especially enzyme activities in clinical diagnostics, and for the determination of inhibitors or activators of enzymes, i.e. trace analysis. The enzymatic analysis method combines the exceptional selectivity with high sensitivity. This article can only attempt a general survey of the enormous potential of the enzymatic analysis, with references to more detailed books or reviews.

At the beginning of this article the special features of enzymatic reactions are characterized and the theoretical background for the application of enzymatic catalysis, the comprehension of the principles on enzymatic kinetics, is described. The rate of the enzymatic reaction can be either increased or decreased by the effect of some substances known as activators or inhibitors, respectively. Simple mechanisms of the effectors are explained in section 4. Examples of the analytical application of enzymatic catalysis for the determination of substrates, enzymes and effectors are given. The use of immobilized enzymes in flow systems and biosensors is presented in the last, but not less interesting, two sections.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 FEATURES OF ENZYMATIC REACTIONS

All enzymes are proteins produced by living cells that catalyze specific reactions with a high degree of efficiency. In the following the special features of enzymes will be represented, including some relevant information for their analytical uses.

2.1 High Specificity of Enzymes

Enzymes are very highly specific in their choice of substrate. Think of the 'key-lock' theory by Fischer or the 'induced fit' hypothesis, suggested by Koshland. Therefore, enzymes are unique analytical reagents. The ideal enzyme for analytical purposes would catalyze only one chemical reaction.

2.2 Temperature

Most enzymatic reactions are carried out within the physiological temperature range of 25–37 °C. As is known from other chemical reactions, there is an increase in reaction rate with temperature. In view of the fact that the enzyme is a protein, it becomes denatured above about 50 °C. Figure 1 shows a typical curve for the dependence of enzymatic activity on temperature. The optimum operating temperature differs from one enzyme to another. For enzymatic reactions it is very important to control the temperature during the measurement. Enzymes should be stored at 4 °C or less, since they can lose activity when stored at room temperature. Many producers of enzyme preparations use some substances to stabilize the preparations, but for some analytical purposes (see section 5.3) these substances can disturb the detection, for example the detection of inhibitors.

2.3 pH Value

Most enzymes are very sensitive to changes in pH. The effect of pH is due to an alteration in the conformation

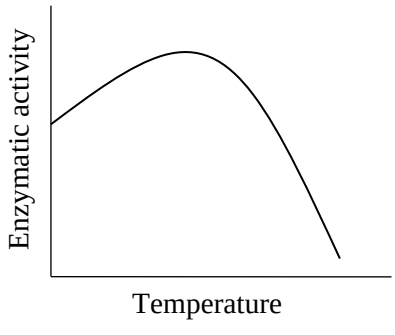

Figure 1 Effect of temperature on the activity of a typical enzyme.

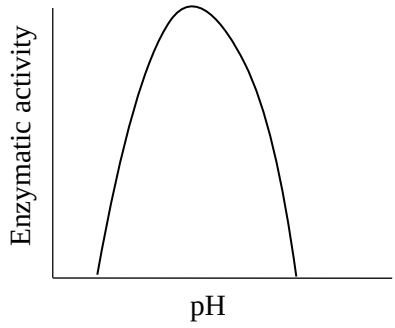

Figure 2 Effect of pH on the activity of a typical enzyme.

of the protein structure or the ionization of the active site of the enzyme or substrate. As seen in Figure 2 there is an optimum in pH value (generally in the pH range of 5–7). Therefore it is very important to keep the pH value constant during the measurement. The use of buffer is advisable. Extreme pH value leads also to denaturation of the protein. The proper choice of pH is especially important in enzyme assays with enzyme-coupled systems, because there can be different pH optima for the coupled enzymes.

2.4 Cofactors

Many enzymes need the presence of a nonprotein component (molecular mass generally between 10^2–10^3 g mol^{-1}), called cofactors, for catalytic activity. In this case, the inactive protein component is termed *apoenzyme* and the active enzyme, including cofactor, the *holoenzyme*. If the cofactor is an organic molecule it is called *coenzyme*. For analytical purposes, e.g. in photometric analysis, the coenzyme nicotinamide adenine dinucleotide (NAD) is widely used (see section 5.1). Other cofactors may be metal ions, e.g. zinc or iron ions. Their analytical importance is given in section 5.3.

2.5 Ionic Strength

Enzymatic reaction rates can be changed by the presence of some salts. The addition of small amounts of neutral salts may increase the solubility of a protein. Salts can cause changes in ionization of amino acid side chains. Therefore the choice of buffer recipe for any given pH is very important for the reaction rate and the mechanism of an enzymatic reaction.

3 REACTION MECHANISM AND KINETIC EQUATIONS

The following kinetic considerations are restricted to the classical approach by Michaelis and Menten. Deviations

from the hyperbolic dependence of the reaction rate on the substrate concentration will not be represented. Furthermore, in most cases only single substrate reactions are significant for the analytical application of enzymes.

3.1 Michaelis–Menten Equation

In the early 1900s two biochemists, Michaelis and Menten worked out a theory for the kinetic analysis of enzyme-catalyzed reactions. The simplified mechanism can be written as Equation (1):

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \{\mathrm{ES}\} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{P} + \mathrm{E} \tag{1}$$

where E is the enzyme, S the substrate, {ES} the intermediate enzyme–substrate complex and P the product(s). The k's are the rate constants. The substrate concentration is much larger than the enzyme concentration so that the formation of {ES} does not alter the substrate concentration. This reaction scheme has been treated both as pre-equilibrium[1] and as a steady-state case.[2] Then, the Michaelis–Menten equation becomes Equation (2):

$$-\frac{\mathrm{d[S]}}{\mathrm{d}t} = \frac{k_2[\mathrm{S}][\mathrm{E}]_0}{K_\mathrm{m} + [\mathrm{S}]} \tag{2}$$

K_m is the Michaelis constant, [S] the substrate concentration and $[\mathrm{E}]_0$ the total concentration of enzyme present. This expression shows the relationship between reaction rate and concentration of both substrate and enzyme. Equation (2) is the base of determination of substrate concentrations and enzyme activities.

Further on, we know that when the substrate concentration is very high, all the enzyme is present as the enzyme–substrate complex and the maximum rate v_max is reached (enzyme saturation). Under these conditions we have Equation (3):

$$v_\mathrm{max} = k_2[\mathrm{E}]_0 \tag{3}$$

We can substitute v_max in Equation (2) and get Equation (4):

$$-\frac{\mathrm{d[S]}}{\mathrm{d}t} = \frac{\mathrm{d[P]}}{\mathrm{d}t} = \frac{v_\mathrm{max}[\mathrm{S}]}{K_\mathrm{m} + [\mathrm{S}]} = \mathrm{v} \tag{4}$$

Equations (2) and (4) are only precisely valid in the case of the initial reaction rate v_0 (Equation 5):

$$v_0 = \frac{v_\mathrm{max}[\mathrm{S}]_0}{K_\mathrm{m} + [\mathrm{S}]_0} \tag{5}$$

For determinations of substrate concentration and enzyme activities we have to discuss two experimental possibilities:

1. $[\mathrm{S}] \gg K_\mathrm{m}$, thus the initial reaction rate v_0 is directly proportional to the enzyme concentration (Equation 6):

$$v_0 = \frac{k_2[\mathrm{S}]_0[\mathrm{E}]_0}{[\mathrm{S}]_0} = k_2[\mathrm{E}]_0 \tag{6}$$

   Since $v_\mathrm{max} = k_2\,[\mathrm{E}]_0$, this attempt is commonly called saturation conditions.
2. $[\mathrm{S}] \ll K_\mathrm{m}$, thus the initial reaction rate v_0 is directly proportional to the substrate concentration, assuming the determination has been carried out under constant enzyme condition ($v_\mathrm{max}/K_\mathrm{m} = \text{constant}$) (Equation 7):

$$v_0 = \frac{k_2[\mathrm{S}]_0[\mathrm{E}]_0}{K_\mathrm{m}} = \frac{v_\mathrm{max}}{K_\mathrm{m}}[\mathrm{S}]_0 \tag{7}$$

When the rate of an enzymatic reaction v is plotted against the substrate concentration [S] a hyperbolic curve is obtained (see Figure 3). Enzyme-catalyzed reactions show a first-order dependence of the reaction rate at low concentrations of substrate and as the substrate concentration rises, the reaction rate approaches a maximum rate v_max, the enzyme saturation. This is the region of zero-order kinetics. At intermediate substrate concentrations, the relationship between the reaction rate and the substrate concentration follows neither first-order nor zero-order kinetics.

3.2 Integrated Michaelis–Menten Equation

Mentioned in section 3.1, the Michaelis–Menten equation is only precisely valid in the case of the initial reaction rate v_0, i.e. when the substrate concentration $[\mathrm{S}] = [\mathrm{S}]_0$, or when less than 5% of [S] is converted to product P.[3] Under many experimental conditions it

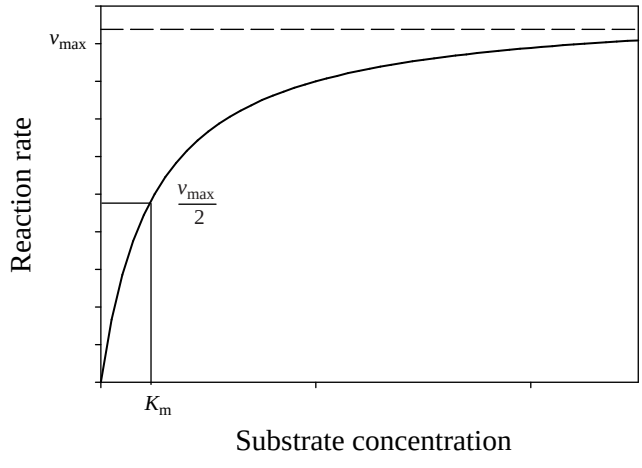

Figure 3 Reaction rate plotted against substrate concentration for an enzyme-catalyzed reaction.

may not be feasible to restrict the reaction extent to 5% or less. In contrast to this, the integrated form of the Michaelis–Menten equation is valid over the entire course of the reaction. This we can write as Equation (8):

$$[S] = [S]_0 - [P] \tag{8}$$

[P] and [S] are the product or substrate concentration at time t.

Insertion of Equation (8) into Equation (4) leads to Equation (9):

$$v = \frac{d[P]}{dt} = \frac{v_{max}\left([S]_0 - [P]\right)}{[S]_0 - [P] + K_m} \tag{9}$$

Integration of Equation (9) between time $t = 0$ and any time t and between the corresponding two product concentrations 0 and [P] leads to Equation (10):

$$t = \frac{K_m}{v_{max}} \ln \frac{[S]_0}{[S]_0 - [P]} + \frac{[P]}{v_{max}} \tag{10}$$

Equation (10) presents the relationship between product concentration [P] and time t. Moreover, Equation (10) can be rearranged to give Equation (11):

$$t = \frac{K_m}{v_{max}} \ln \frac{[S]_0}{[S]} + \frac{[S]_0 - [S]}{v_{max}} \tag{11}$$

The application of Equation (10) for the determination of K_m and v_{max} (or the enzyme activity) is described in section 3.4.

3.3 Michaelis Constant and Enzyme Activity

The Michaelis constant, K_m, is defined as the substrate concentration at half the maximum reaction rate (Equation 12):

$$\text{when} \quad K_m = [S] \quad v = \frac{[S]v_{max}}{2[S]} = \frac{v_{max}}{2} \tag{12}$$

Thus K_m will have the same units as the substrate concentration. Using the two methods for treatment of enzyme-catalyzed reaction Equation (1), one can obtain Equations (14) and (16):

1. Michaelis–Menten assumption of equilibrium between substrate, enzyme and substrate–enzyme complex (pre-equilibrium or rapid equilibrium), $k_2 \ll k_{-1}$, Equation (13):

$$E + S \overset{K_s}{\rightleftharpoons} \{ES\} \xrightarrow{k_2} P + E \tag{13}$$

$$K_m^{I} = K_s = \frac{k_{-1}}{k_1} \tag{14}$$

K_s = dissociation constant for the enzyme–substrate complex

2. For Equation (15):

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \{ES\} \xrightarrow{k_2} P + E \tag{15}$$

Briggs–Haldane steady-state assumption, Equation (16):

$$K_m^{II} = \frac{k_{-1} + k_2}{k_1} \tag{16}$$

The Briggs–Haldane assumption is the more general one, eliminating the requirement for the enzyme–substrate complex to be in equilibrium with free enzyme and free substrate. Thus, K_m is a dynamic or pseudo-equilibrium constant. K_m is a constant for a given substrate under well-defined experimental conditions such as temperature, buffer, pH and ionic strength. It represents the affinity between the substrate and the enzyme. A small value of K_m means a high affinity between substrate and enzyme and the enzyme becomes saturated at small concentrations of substrate. Michaelis constants for enzyme usually range from 10^{-2} to 10^{-5} mol L^{-1}. For enzymatic analysis methods a small value of K_m results in a very limited linear concentration range in calibration graphs based on initial reaction rate as a function of substrate concentration. If we know K_m, we can adjust the experimental conditions for determining substrate concentration or enzyme activity (see section 3.1). The activity of an enzyme preparation varies from one source to another, since the purification of the enzyme is in general not complete. Because of this, in most preparations the actual molar concentration of enzyme is unknown, and therefore the quantity generally used to express the enzyme activity is the *unit*, defined in terms of reaction rate.

One international unit (U) of enzyme is defined as that amount catalyzing the conversion of micromole substrate to product in 1 minute under optimal conditions.

The declaration of the optimal conditions has to include the temperature (usually 25 or 30 °C), pH value and buffer system. For determining the enzyme activity the substrate concentration should be sufficient for saturation. The Nomenclature Commission of the International Union of Biochemistry has recommended the use of the *katal* (kat), which differs from the definition of the unit through the declaration of 1 mole substrate and time of 1 second:

$$1\,\text{kat} = 6 \times 10^7\,\text{U}$$

The *specific activity* of an enzyme preparation is the number of units per mg protein or in the case of solution or suspension one can declare the number of units per

milliliter. If the molecular mass of an enzyme is known it is possible to express the *molecular activity* in units per μmol enzyme.

In Palmer[4] one can also find the terms: *turnover number* k_{cat} and *catalytic efficiency* k_{cat}/K_m. The constant k_{cat} is defined in Equation (17):

$$k_{cat} = \frac{v_{max}}{[E]_0} \tag{17}$$

If the enzyme concentration is known the ratio k_{cat}/K_m is a good measure for catalytic efficiency of an enzyme.

3.4 Determination of Kinetic Parameters

The determination of the kinetic parameters v_{max} and K_m can be carried out by measuring the initial reaction rate as the substrate concentration is varied. If experimentally possible, the substrate concentration should be varied in the range from $K_m/10$ to 10 K_m. In the literature there are some linearization procedures (see Figure 4):

3.4.1 The Lineweaver–Burk Plot

The linearization of the Michaelis–Menten equation Equation (5) by Lineweaver and Burk (Figure 4a) is the most widely and traditionally used method. It is a double reciprocal plot (Equation 18):

$$\frac{1}{v_0} = \frac{K_m}{v_{max}}\frac{1}{[S]_0} + \frac{1}{v_{max}} \tag{18}$$

From the slope (K_m/v_{max}) or the intercept on the ordinate ($1/v_{max}$) or the intercept on the abscissa ($-1/K_m$) one can obtain the two kinetic parameters. But such a double-reciprocal method has a strong distorting effect on the data.

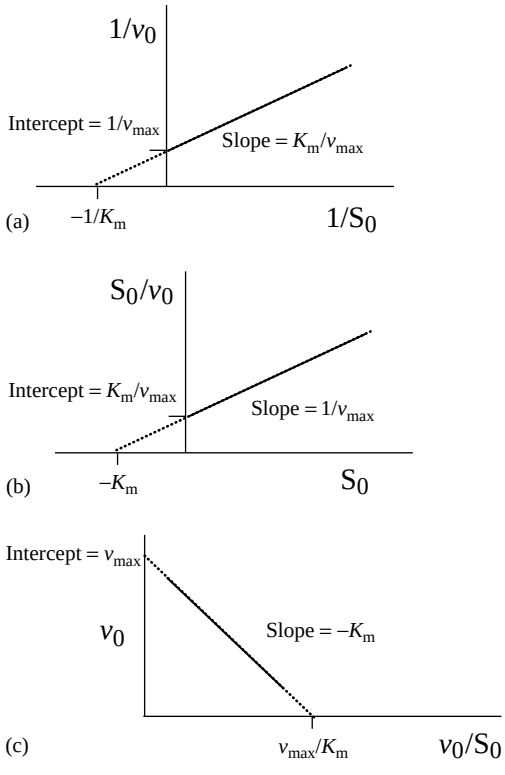

Figure 4 Transformations of the Michaelis–Menten equation to determine v_{max} and K_m. (a) Lineweaver–Burk plot; (b) Hanes plot; (c) Eadie–Hofstee plot.

3.4.2 The Hanes Plot

The plot $[S]_0/v_0$ versus $[S]_0$ due to Hanes and Woolf (Figure 4b) is much better with regard to this effect. The data are more evenly weighted over the whole concentration range. Equation (18) may be rearranged to give Equation (19):

$$\frac{[S]_0}{v_0} = \frac{1}{v_{max}}[S]_0 + \frac{K_m}{v_{max}} \tag{19}$$

From this linear plot the parameters K_m and v_{max} can be calculated.

3.4.3 The Eadie–Hofstee Plot

The Eadie–Hofstee plot (Figure 4c) is the straight line graph v_0 versus $v_0/[S]_0$, Equation (20):

$$v_0 = -K_m\frac{v_0}{[S]_0} + v_{max} \tag{20}$$

Both plots (Eadie–Hofstee plot and Hanes plot) are less frequently used, but they are better suited for kinetic determination than the Lineweaver–Burk plot.

3.4.4 Integrated Michaelis–Menten Equation

All above-mentioned methods require the measurement under initial conditions. If it is not possible to fulfill this requirement – for example, it may be difficult to determine very low concentrations of product – one can use the integrated Michaelis–Menten equation (see section 3.2) for the determination of v_{max} and K_m. The linear transformation of Equation (10) gives Equation (21):

$$\frac{t}{[P]} = \frac{K_m}{v_{max}}\frac{1}{[P]}\ln\frac{[S]_0}{[S]_0 - [P]} + \frac{1}{v_{max}} \tag{21}$$

The Michaelis constant and maximum rate can be determined by measuring the concentration of product several times during the reaction and plotting the respective values as shown in Figure 5.

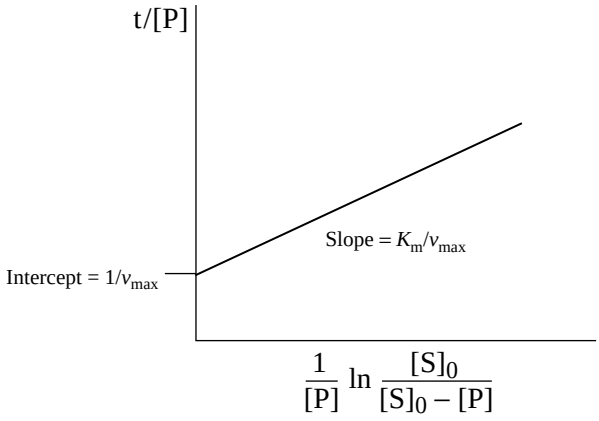

Figure 5 Plot of integrated Michaelis–Menten equation.

4 MODIFYING THE KINETIC BY EFFECTORS

The rate of the enzymatic reaction can be either increased or decreased by effectors. While activators accelerate the enzymatic reaction, inhibitors reduce the rate of the enzymatic reaction. For this reason all enzymatic analyses of effectors involve the determination of reaction rates (different kinetic methods are described in section 5 or in Pérez-Bendito and Silva[5]).

4.1 Activators

An activator increases the catalytic effectiveness or the activator transforms the inactive apoenzyme (see section 2.4) to the active holoenzyme. Various substances can activate the same enzyme. In consequence, there is a lack of specificity for determination of individual ions.

The course of the initial rate in the presence of an activator is similar to that of substrate concentration, discussed in section 3.1. At low activator concentration one can find that the initial rate is directly proportional to the concentration of the activator. But at higher concentrations the enzyme is maximally activated (similar to the enzyme saturation at high substrate concentrations) and the initial rate becomes independent of the concentration of the activator. Consequently the determination of effectors has to be carried out at low levels of activator concentrations. Examples for the analytical application, e.g. the determination of metallic ions, are given in section 5.3.

4.2 Inhibitors

Inhibitors are substances which cause a decrease in the rate of a catalytic reaction. The enzyme inhibition can be either reversible or irreversible. The characteristic of the reversible inhibition is the equilibrium between enzyme and inhibitor. In contrast, the irreversible inhibitor binds covalently to the enzyme. Seen from Figure 6, the enzymatic reaction rate can be influenced by several substances. We distinguish:

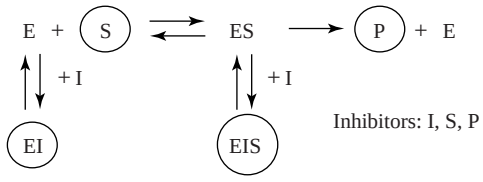

Figure 6 Schematic representation of the various forms of inhibition.

- substrate inhibition
- product inhibition
- competitive inhibition (formation of enzyme–inhibitor complex EI)
- uncompetitive inhibition (formation of enzyme–inhibitor–substrate complex EIS)
- noncompetitive inhibition (equal affinity of formation of EI and EIS).

4.2.1 Competitive Inhibition

The inhibitor competes with substrate (or coenzyme) for the active centre of the enzyme. Since the inhibitor forms the enzyme-inhibitor complex, the proportion of free enzyme will be reduced and the rate of the enzymatic reaction decreases. Hence, the competitive inhibition can be made ineffective by an excess of substrate. In most cases we find that the chemical structure of the substrate is similar to that of the competitive inhibitor. Figure 7(a) shows the Lineweaver–Burk plot of the competitive inhibited enzyme reactions. The curves increase in slope (increase in K_m), while the intercept on the $1/v_0$ axis remains constant. The maximum rate in the presence of a competitive inhibitor is always equal to v_{max} of the uninhibited reaction.

4.2.2 Uncompetitive Inhibition

The inhibitor binds only to the enzyme–substrate complex ES. Figure 7(b) shows the Lineweaver–Burk plot of the uncompetitive inhibited enzyme reaction. The uncompetitive inhibitor decreases v_{max} and K_m. The decrease of K_m means an apparent increase in the affinity of the enzyme for its substrate.

4.2.3 Noncompetitive Inhibition

The inhibitory mechanism, where the inhibitor binds with equal affinity to both the enzyme and the enzyme–substrate complex, is called noncompetitive

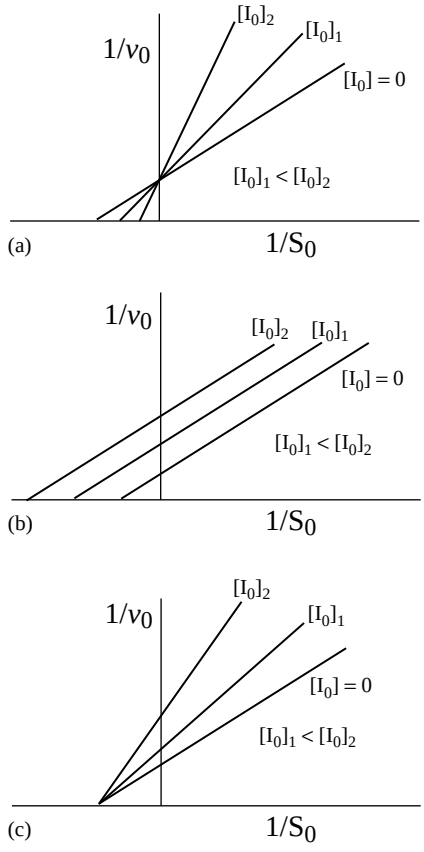

Figure 7 Lineweaver–Burk plots for (a) competitive inhibition; (b) uncompetitive inhibition; (c) noncompetitive inhibition.

inhibition. Noncompetitive inhibitors bind to the enzyme at a site that is distinct from the substrate binding site.[6]

Hence, this type of inhibition cannot be overcome by increasing the substrate concentration. Figure 7(c) illustrates that the noncompetitive inhibition leads to an apparent decreasing of v_{max}, while the Michaelis constant is not altered.

4.2.4 Substrate Inhibition

As represented in section 3.1, the initial rate of an enzyme reaction increases with increasing initial substrate concentration to a maximum value v_{max}. But in some cases we can find a decreased initial rate at high substrate concentrations, seen in Figure 8. There, an excess of substrate can inhibit its own conversion to product.

4.2.5 Product Inhibition

The product of some enzyme reaction can also be an inhibitor. The product may bind to the active site of the enzyme. Consequently the active site is blocked and cannot bind further substrate molecules. Thus, the rate of the enzyme reaction decreases.

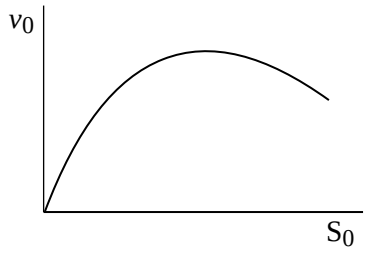

Figure 8 Substrate inhibition.

The represented types of enzyme inhibition are not complete. More detailed information can be found in Segel.[7] Examples for analytical application are given in section 5.3.

5 ENZYMATIC DETERMINATIONS

The methods to be used in monitoring enzyme-catalyzed reactions are dictated by the chemistry and physical chemistry of the reaction. We can distinguish three classes:

- spectrophotometric
- electrochemical
- calorimetric.

In the following only some typical examples of the determination of substrates, enzymes and effectors, using different monitoring techniques, are represented. In the literature one enzyme plays an exceptional role – glucose oxidase (GOD) as analytical reagent is the most widely used enzyme.[8]

5.1 Determination of Substrates

The substrate determination can be carried out by both, equilibrium and kinetic approaches. The *equilibrium method*, also called *end-point method*, means the measurement of the total change, almost all of the substrate being converted to product. The analysis can be performed by measuring the appearance of product or the disappearance of substrate. To ensure rapid progress towards equilibrium, the concentration of the used enzyme should be high. Any other substances necessary for enzyme-catalyzed reaction, such as cofactors (coenzyme or metal ions), should be in excess, so that they do not limit the reaction. Because of the possible side reactions or instability of products or reactants, the enzyme reactions are often not stoichiometric with respect to the substrate concentration. Therefore the calibration curves must be prepared under the selected experimental conditions. There, the measured quantity is related to known concentrations of substrates.

The *kinetic determinations* of substrates should be carried out under conditions where the concentration of substrate is in the range to fulfill the first order circumstances ($[S]_0 \ll K_m$, see Equation (7) section 3.1). In Christian[9] one can find three forms:

- measure the time required for the reaction to convert a certain amount of substrate or form a certain amount of product
- measure the amount of substrate converted or product generated in a given time and
- continuous measurement of substrate or product concentration as a function of time, or initial slope method.

The initial slope method means that the change in some physicochemical parameters (absorption, pH or temperature), related to the change of concentration, are plotted as a function of time. Kinetic methods are generally more rapid than equilibrium methods, but also more sensitive to enzyme effectors.

Analytical applications of enzyme-catalyzed reactions for determining substrates are widely used in many fields. Table 1 gives some examples for applications in food industry, medicine and environmental analysis; further pertinent information can be found in Bergmeyer.[10]

As seen from Table 1, many reactions with dehydrogenases need the coenzyme NAD. The NADH and the oxidized form (NAD^+) differ in their ultraviolet absorption spectra. NAD does not adsorb photons in the 340 nm region of spectra, whereas its reduced form does.

Such differences in absorption spectra of products, substrates or other reactants are widely used in the

Table 1 Typical determinations of substrates using enzymes

Substrate	Enzyme	E.C. number	CAS registry number	Reaction	Examples for application
A Food industry					
ethanol	alcohol dehydrogenase	1.1.1.1	9031-72-5	ethanol + NAD^+ ⟶ acetaldehyde + NADH + H^+	alcoholic drinks
sucrose	invertase	3.2.1.26	9001-57-4	sucrose + H_2O ⟶ glucose + fructose	fruit juices, syrups
glutamic acid	glutamic dehydrogenase	1.4.1.3	9029-12-3	glutamate + NAD^+ + H_2O ⟶ 2-oxoglutarate + NADH + NH_4^+	spices, meat
ascorbic acid	ascorbatoxidase	1.10.3.3	9029-44-1	ascorbate + $\frac{1}{2}O_2$ ⟶ dehydroascorbate + H_2O	fruit juices, drinks, meat
sorbitol	sorbitol dehydrogenase	1.1.1.14	9028-21-1	sorbitol + NAD^+ ⟶ fructose + NADH + H^+	foodstuffs for diet, fruits
B Medicine					
cholesterol	cholesterol oxidase	1.1.3.6	9028-76-6	cholesterol + O_2 ⟶ 4-cholesten-3-one + H_2O_2	blood analysis of risk factors
glucose	GOD	1.1.3.4	9001-37-0	glucose + O_2 + H_2O ⟶ gluconic acid + H_2O_2	diagnosis of diabetes
	glucose dehydrogenase	1.1.1.47	9028-53-9	glucose + NAD^+ + H_2O ⟶ gluconic acid + NADH + H^+	
lactic acid	D-LDH	1.1.1.28	9028-36-8	(R)lactate + NAD^+ ⟶ pyruvate + NADH + H^+	sports medicine
uric acid	uricase	1.7.3.3	9002-12-4	urate + O_2 + $2H_2O$ ⟶ allantoin + CO_2 + H_2O_2	diagnosis of gout
penicillin	penicillinase	3.5.2.6	9001-74-5	penicillins + H_2O ⟶ penicilloic acid	diagnosis of allergy
C Environment					
phenol	tyrosinase	1.14.18.1	9002-10-2	phenol + $\frac{1}{2}O_2$ ⟶ *o*-benzochinone	drinking water, waste water
urea	urease	3.5.1.5	9002-13-5	urea + $2H_2O$ + H^+ ⟶ HCO_3^- + $2NH_4^+$	bathwater
nitrate	nitrate reductase	1.6.6.1	9013-03-0	nitrate + NADH + H^+ ⟶ nitrite + NAD^+ + H_2O	water

NADH, reduced form of nicotinamide adenine dinucleotide; LDH, lactate dehydrogenase.

spectrophotometric method to follow the progress of enzyme reactions.

As an example, the amount of NADH produced according to Equation (22) is a measure of the amount of acetaldehyde present in wine or beer.[11]

$$\text{Acetaldehyde} + NAD^+ + H_2O \xrightarrow{\text{Al-DH}} \text{acetic acid} + NADH + H^+ \quad (22)$$

(Al-DH = aldehyde dehydrogenase, E.C. 1.2.1.3, 9028-86-8)

Moreover, NADH is used as a coupling agent. Two or more reactions occur in sequence. At the end of such a sequence, NAD^+ is generated or consumed. For example, the determination of creatinine or creatine is represented by Equation (23–26):

$$\text{Creatinine} + H_2O \xrightarrow{\text{CAH}} \text{creatine} \quad (23)$$

$$\text{Creatine} + ATP \xrightarrow{\text{CK}} \text{creatine phosphate} + ADP \quad (24)$$

$$ADP + \text{phosphoenolpyruvate} \xrightarrow{\text{PK}} ATP + \text{pyruvate} \quad (25)$$

$$\text{Pyruvate} + NADH + H^+ \xrightarrow{\text{LDH}} \text{L-lactate} + NAD^+ \quad (26)$$

The amount of NADH consumed is stoichiometric with the amount of creatine or creatinine.[12]

(CAH creatinine amidohydrolase, E.C. 3.5.2.10, 9025-13-2
CK creatinine kinase, E.C. 2.7.3.2, 9001-15-4
PK pyruvate kinase, E.C. 2.7.1.40, 9001-59-6
LDH E.C. 1.1.1.27, 9001-60-9)
ADP adenosine diphosphate
ATP adenosine triphosphate

The coupling of enzyme-catalyzed reactions with further reactions is commonly applied in enzymatic analysis, chiefly by using spectrophotometric or electrochemical methods. In contrast, the *calorimetric determination (enthalpimetry)* of substrate is much more general.[13] Each enzyme reaction with a defined product can be used. There is no demand for colored products or electrochemically active reactants. The measured heat is proportional to the substrate concentration.

Furthermore, Table 1 shows that the reaction of oxidases consumes oxygen. That makes them accessible to an *electrochemical method*, the amperometric detection. For the *amperometric determination* one can also use the production of hydrogen peroxide, Equation (27):

$$\text{glucose} + O_2 + H_2O \xrightarrow{\text{GOD}} \text{gluconic acid} + H_2O_2 \quad (27)$$

The hydrogen peroxide is electrochemically oxidized at the platinum cathode, Equation (28):

$$H_2O_2 \longrightarrow O_2 + 2H^+ + 2e^- \quad (28)$$

The current obtained is proportional to the glucose concentration (principle of Clark oxygen electrode).

The substrate concentration of such an oxidase-catalyzed reaction (Equation 27) can also be measured with an *iodide-selective electrode*[14], Equation (29):

$$H_2O_2 + 2I^- + 2H^+ \xrightarrow{\text{Mo(VI)}} I_2 + 2H_2O \quad (29)$$

The decrease in iodide concentration is proportional to the substrate concentration, for instance glucose.

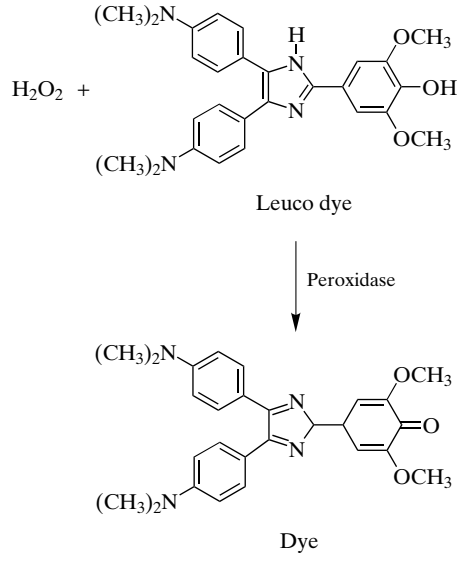

In H_2O_2-generating systems, substrate concentrations can also be determined by using *indicator reactions*.[15] One example is the determination of glucose with indicator reactions based on the peroxidase (E.C. 1.11.1.7, 9003-99-0) catalyzed oxidation of single chromogen (Equation 30).

An interesting way to increase the sensitivity of enzymatic substrate determinations is *amplification by cycling of enzyme-catalyzed reactions*. The principle based on the following scheme:

The substrate to be determined is converted by an initial enzyme to a product. This product is the substrate of a second enzyme, which regenerates the substrate. Therefore it can be converted by the first enzyme again and thus the cycle is reinitiated.

A typical example is the determination of NADP by a cycling of NADPH between two enzyme reactions.[16]

5.2 Determination of Enzymes

Since enzymes are catalysts, affecting the rate of the reaction, not the equilibrium, their activity must be

measured by a *kinetic* method. The kinetic determination of enzymes should be carried out under conditions where the substrate concentration is in excess and all other substances which influence the rate of reaction are present in fixed and non-limiting concentrations. Then the value of K_m can be neglected and the rate of the enzyme reaction depends only on the enzyme concentration (zero order conditions, $[S]_0 \gg K_m$ see Equation (6) section 3.1). The three possible forms of determination are identical to the *kinetic* determination of substrates (see section 5.1). Because of the influence on the enzyme activity (see section 2), the pH value and the temperature have to keep constant during the measurement.

Enzyme determinations are chiefly applied in clinical diagnostics, food industry and monitoring biotechnological processes. Table 2 shows a selection of interesting enzymes; further information is found in Bergmeyer.[10]

As seen in the Table 2, *spectrophotometric monitoring* of NADH consumption as a function of time can be used to determine the enzyme activity of various enzymes.

The determination of amylase activity by measuring the produced reducing sugar as a function of time is an example of the application of the *colorimetric method*. An alkaline solution of 3,5-dinitrosalicylic acid and K-Na-tartrate is used as a chromogenic reagent.[17] Another, even more simple, case of *colorimetric method* is the determination of the AP activity. There the formation of the yellow *p*-nitrophenol under alkaline conditions is measured.

Furthermore the *pH-stat method* is used to determine enzyme activities. Monitoring enzyme reaction by pH-stat method means that the pH is maintained at constant value by frequent addition of alkali. The rate at which the base is added is proportional to the enzyme activity. The determination of lipase activity in serum (Equation 31) is a typical example.[18]

$$\text{Triglyceride} + H_2O \xrightarrow{\text{lipase}} \text{diglyceride} + \text{fatty acid} \quad (31)$$

Table 2 Typical examples for enzyme determinations

Enzyme	E.C. number	CAS registry number	Reaction	Application
glutamic-oxaloacetic transaminase (GOT)	2.6.1.1	9000-97-9	asparate + α-ketoglutarate $\xrightarrow{\text{GOT}}$ glutamate + oxaloacetate oxaloacetate + NADH + H^+ $\xrightarrow{\text{MDH}}$ malate + NAD^+ (MDH malate dehydrogenase, E.C. 1.1.1.37, 9001-64-3)	clinical diagnostics, fresh meat control
glutamic-pyruvic transaminase (GTP)	2.6.1.2	9000-86-6	L-alanin + α-ketoglutarate $\xrightarrow{\text{GPT}}$ glutamate + pyruvate pyruvate + NADH + H^+ $\xrightarrow{\text{LDH}}$ lactate + NAD^+	clinical diagnostics (liver diagnostic, bilious attack)
LDH	1.1.1.27	9001-60-9	pyruvate + NADH + H^+ $\xrightarrow{\text{LDH}}$ lactate + NAD^+	clinical diagnostics (cardiac infarction)
CK	2.7.3.2	9001-15-4	creatinine phosphate + ADP $\xrightarrow{\text{CK}}$ creatine + ATP	clinical diagnostics (cardiac infarction)
alkaline phosphatatse	3.1.3.1	9001-78-9	*p*-nitrophenylphosphate + H_2O $\xrightarrow{\text{AP}}$ *p*-nitrophenol + phosphate	food industry (milk products)
α-amylase	3.2.1.1	9000-90-2	polysaccharide containing alpha-(1-4)-linked glucose units + H_2O $\longrightarrow$ maltooligosaccharides	biotechnological processes
lipase	3.1.1.3	9001-62-1	triglyceride + H_2O $\xrightarrow{\text{lipase}}$ diglyceride + fatty acid	biotechnological processes detergent industry
catalase	1.11.1.6	9001-05-2	$H_2O_2 \longrightarrow H_2O + O_2$	food stuffs
alanine aminopeptidase	3.4.11.2	9054-63-1	aminoacyl-peptide + H_2O $\longrightarrow$ amino acid + oligopeptide	waste water
peroxidase	1.11.1.7	9003-99-0	donor + $H_2O_2 \longrightarrow$ oxidized donor + H_2O	food stuffs

AP, alkaline phosphatase.

The produced fatty acid is neutralized by counter titration with NaOH. Normally olive oil is used as the substrate. More examples for enzyme determinations and automated pH-stat systems can be found in Gul et al.[19]

Among the pH-stat systems, other *electrochemical methods* such as *conductometry* (change in overall ionic mobility is measured as a function of time), *amperometry* (O_2 consumption or H_2O_2 production is measured as a function of time) or *ion-sensitive electrodes* (see section 5.1) are also usable.

The *calorimetric determination* of invertase activity is described.[20] The sucrose hydrolysis shown in Equation (32) was performed in an isoperibolic calorimeter (invertase, E.C. 3.2.1.26, 9001-57-4). The kinetic evaluation of the calorimetric curves qualified to determine the enzymatic activity.

$$\text{sucrose} + H_2O \xrightarrow{\text{invertase}} \text{glucose} + \text{fructose} \quad (32)$$

Finally, *electrophoresis* is an important method for determining the catalytic activity of nucleases.[21]

5.3 Determinations of Effectors

The rate of the enzyme reaction can be influenced by various substances (see section 4). This characteristic feature of enzyme reactions can be used for analytical purposes, the determination of *activators* and the determination of *inhibitors*. These methods are especially important in trace metal determinations, since in most cases only a single oxidation state of metal is determined. The enzymatic analysis based on activator effects is not as frequently used as the determination based on inhibitor effects. Because effectors influence the kinetic of the enzyme reaction, their determination requires the monitoring of enzyme reactions as a function of time, i.e. kinetic determinations.

5.3.1 Activators

Many enzymes require for their catalytic activity *cofactors* (see section 4.1) or the 'true substrate' (SA), where A is generally a metal ion.[22] For example, reactions involving ATP require Mg^{2+} ions to form $MgATP^{2-}$ ($A = Mg^{2+}$).

For instance, the amounts of magnesium can be determined according to Equation (33):

$$\text{luciferin} + O_2 + \text{ATP} \xrightarrow[Mg^{2+}]{\text{luciferase}} \text{oxyluciferin} + \text{ADP} + PO_4^{3-} \quad (33)$$

The formation of oxyluciferin is combined with green *chemiluminescence*. Trace amounts of magnesium (ppb levels) can be determined.[23]

Other inorganic substances combine with the enzyme at a specific activator site. This activation may be essential or nonessential. Table 3 shows some examples of activation by inorganic substances. Pertinent information on the requirement or effectiveness of some substances for the catalytic activity of multitude of enzymes can be found in Schomburg and Salzmann.[24] These details are important for the application of activator effects for analytical purposes.

The experimental procedure of trace element analysis, using *metallo-enzymes* (specific metal ions are located at the active centers of the enzyme) is characterized by the removal of the metal ion from the holoenzyme (see section 4.1) by treatment with chelating agents (for instance EDTA). The resultant *apoenzyme* is catalytically inactive. The activator assay consists of the respective apoenzyme, its natural substrate, and a sample containing the metal ion. Then the restoration of enzyme activity is proportional to the amount of metal ion in the sample. In this way trace amounts of metal ions can be determined. An example is the determination of zinc ions by using AP (Equation 34):

Table 3 Examples for activation of enzymes by inorganic substances

Enzyme	E.C. number	CAS number	Activators	Ref.
arginase	3.5.3.1	9000-96-8	Mn^{2+}	25
citrate lyase	4.1.3.6	9012-83-3	Mg^{2+}, Mn^{2+}, Fe^{2+}, Zn^{2+}	25
enolase	4.2.1.11	9014-08-8	Mg^{2+}	25
cytochrome oxidase	1.9.3.1	9001-16-5	Cu^{2+}	26
carboxypeptidase A	3.4.17.1	11075-17-5	Zn^{2+}	26
creatinine kinase	2.7.3.2	9001-15-4	Mg^{2+}, Mn^{2+}	26
glycerokinase	2.7.1.30	9030-66-4	Mg^{2+}	26
isocitrate dehydrogenase	1.1.1.41	9001-58-5	Mn^{2+}, Zn^{2+}, Co^{2+}	27
phosphofructokinase	2.27.1.11	9001-80-3	K^+	27
AP	3.1.3.1	9001-78-9	Zn^{2+}, Mg^{2+}, Mn^{2+}, Co^{2+}	25 28
collagenase	3.4.24.3	9001-12-1	Ca^{2+}	25
luciferase	1.13.12.7	61970-00-1	Mg^{2+}	27

$$\text{4-nitrophenyl phosphate} + H_2O \xrightarrow{AP} \text{4-nitrophenol} + \text{phosphate} \quad (34)$$

The enzyme activity is determined by measuring the rate of release of 4-nitrophenolate ion (alkaline pH) at 410 nm in a recording spectrometer. The method is so sensitive that 0.5 ng zinc can be determined in small volumes (0.5 µl) of plasma.[29]

Many enzymes require for their catalytic activity a specific *coenzyme* (see Table 4). At low concentrations of enzyme the degree of activation of such an enzyme is proportional to the concentration of coenzyme. That is in contrast with the assay for determining substrate concentrations (see section 5.1), where the coenzyme should be in excess.

For instance, ATP, an important intermediate in energy metabolism, can be determined by the following reaction scheme (Equations 35 and 36):

$$\text{ATP} + \text{glucose} \xrightarrow{HK} \text{glucose-6-phosphate} + \text{ADP} \quad (35)$$

Table 4 Some examples of coenzymes

Coenzyme	Abbreviation	Apoenzyme	Function
adenosine triphosphate	ATP	kinases, transferases	transphosphorylation
coenzyme A	CoA	acyltransferases, thioligases	acyl group transfer
flavine adenine dinucleotide	FAD	oxidases	proton transfer
nicotinamide adenine dinucleotide	NAD	dehydrogenases	proton transfer
pyridoxal phosphate	PAL	transaminases	amino group transfer

$$\text{glucose-6-phosphate} + NADP^+ \xrightarrow{G6PDH} \text{glucono-1,5-lactone 6-phosphate} + NADPH \quad (36)$$

(HK Hexokinase, E.C. 2.7.1.1, 9001-51-8
G6PDH glucose 6-phosphate dehydrogenase, E.C. 1.1.1.49, 9001-40-5)

The reaction can be monitored by the formation of NADPH spectrophotometrically at 340 nm.

5.3.2 *Inhibitors*

In the literature one can find a variety of enzyme inhibitors. Table 5 shows a selection of well-known inorganic and organic inhibitors. The relative effect on enzyme activity of different substances is characterized by the IC_{50} value. The IC_{50} value is the inhibitor concentration giving 50% inhibition.[30] Several analytical methods have been developed based on the inhibition of enzyme reactions.

As seen from Table 5, one enzyme often has various inhibitors. Therefore, investigations of inhibitor effects are not as practicable for single-component analysis as for determination of summary or group specific parameters. An example is the screening of pollutants in waste water. Furthermore, enzyme inhibition tests are suitable to preselect polluted samples.

In the following some examples for analytical application (see also sections 6 and 7) are represented. There is a variety of methods for determination of inhibitors, dictated by the physicochemistry of the enzyme reaction to be monitored.

The urease-catalyzed urea hydrolysis can be inhibited by various metal ions (see Table 5). The inhibitor effect is used for the screening of heavy metals in water. An example is the *microtiter-plate urease inhibition assay*

Table 5 Some examples for inhibitors

Enzyme	Inhibitor	E.C. number	CAS number	Ref.
delta-aminolevalinic dehydratase	Pb^{2+}	4.2.1.24	9036-37-7	31 Vol. IV
urease	Cd^{2+}, Hg^{2+}, Cu^{2+}, Zn^{2+}	3.5.1.5	9002-13-5	32
lipase	Fe^{2+}, Zn^{2+}, Co^{2+}	3.1.1.3	9001-62-1	33
arylsulfatase	phosphate, sulfate	3.1.6.1	9016-17-5	33
catalase	ascorbic acid	1.11.1.6	9001-05-2	32
cholinesterase	pesticides	3.1.1.7	9000-81-1	32
beta-galoctosidase	organic SH-compounds	3.2.1.23	9031-11-2	33
dihydrofolate reductase	methotrexate (MTX)	1.5.1.3	9002-03-3	31 Vol. XII
invertase	Ag^{+}, Hg^{2+}	3.2.1.26	9001-57-4	32
phosphoenolpyruvate carboxylase	pesticide, Hg^{2+}, Co^{2+}, Pb^{2+}, Zn^{2+}, Cu^{2+}	4.1.1.31	9067-77-0	34
GOD	Ag^{+}, Hg^{2+}, Pb^{2+}	1.1.3.4	9001-37-0	32

based on the determination of the released ammonia with an indophenol reaction. Highest sensitivity was found for Hg^{2+} (detection limit of $0.05\,\mu g\,L^{-1}$).[35]

Pesticide pollution in water is a serious problem, especially for drinking water quality. The acetylcholinesterase (ACHE) inhibition test is a well-known method for the determination of organophosphorous acid esters and *N*-methyl carbamates. *Photometric* tests are widely used. The hydrolysis of acetylthiocholine by ACHE (E.C. 3.1.1.7, 9000-81-1) is measured (Equations 37 and 38):

$$\text{acetylthiocholine} + H_2O \xrightarrow{\text{ACHE}} \text{thiocholine} + CH_3COOH \quad (37)$$

$$\text{thiocholine} + \text{dithiobisnitrobenzoate} \longrightarrow \text{thionitrobenzoate (yellow dye)} \quad (38)$$

The entire inhibition effect, in comparison with the uninhibited ACHE, is given in paraoxon equivalent.[36]

This inhibition effect can also be measured by *amperometric* detection (Equations 39–41)

$$\text{acetylcholine} + H_2O \xrightarrow{\text{ACHE}} \text{choline} + CH_3COOH \quad (39)$$

$$\text{choline} + 2O_2 + H_2O \xrightarrow{\text{choline oxidase}} \text{betaine aldehyde} + 2H_2O_2 \quad (40)$$

$$H_2O_2 \xrightarrow{\text{anodic oxidation}} O_2 + 2e^- + 2H^+ \quad (41)$$

(choline oxidase, E.C. 1.1.3.17, 9028-67-5).

The inhibition of ACHE is proportional to pesticide concentration and measured through the decrease of the current.[37]

Inhibition studies are also suitable to distinguish between oxidation states of metals. As an example, *calorimetric investigations* of urease inhibition of heavy metals show that As(III)-ions inhibit the urea hydrolysis by competitive mechanism, while As(V)-ions do not inhibit the enzyme in the comparable concentration range.[38] For the evaluation of the potential of toxicological hazard the differentiation of arsenic content should be important.

6 IMMOBILIZED ENZYMES FOR DETERMINATIONS IN FLOW SYSTEMS

In analytical chemistry the use of immobilized enzymes is rapidly increasing. The term *immobilized enzymes* stands for a preparation in which an enzyme is confined or localized in a relatively well defined region. Enzymes may be immobilized by covalent bindings, by crosslinking, adsorption or entrapment within various matrices, for instance membranes or polymeric gels. Different techniques used for the immobilization of enzymes are described extensively.[39,40]

Enzyme immobilization is an economical requirement, since an immobilized enzyme can easily separate from the reaction solution and becomes therefore reusable. Besides, immobilization frequently leads to an improvement of thermal and/or chemical stability of enzyme preparation in comparison to the soluble or native enzyme. On the other hand, the immobilization often causes a reduction of catalytic activity. Immobilized enzyme systems always differ from the soluble counterparts by the K_m value or the optimal pH condition.

A further advantage of immobilization is the possibility to prepare *multienzyme systems*, where two or more enzymes immobilized on the same carrier. Immobilized multienzyme systems are suitable to catalyze sequential sets of reactions.

The application of immobilized enzymes in flow systems is a very convenient way for rapid and automated enzymatic determinations. The most common configurations in such flow systems are packed bed reactors or open-tube wall reactors.

In packed bed reactors especially covalent enzyme binding on controlled pore glass (CPG) or polymeric carrier is used. In open-tube reactors the enzymes are bounded to the inner surface of tubes. The sample-handling technique used in packed bed reactors and open-tube wall reactors is the flow injection analysis (FIA). FIA means short analysis times, and ease of automation. FIA is based on the injection of the liquid sample into a nonsegmented continuous buffer stream.[41] The injected sample forms a zone in the stream, which is then flowing into the enzyme reactor. There the enzyme reaction takes place and the resultant change in a physicochemical parameter (e.g. absorption, temperature, pH) is registered by the detector. Figure 9 shows a practicable arrangement of the flow system for enzymatic determinations in a simplified representation.

The detector signal during the passage of the sample (substrate zone) has the form of a peak. The analytical information is given by the peak height, the peak area or

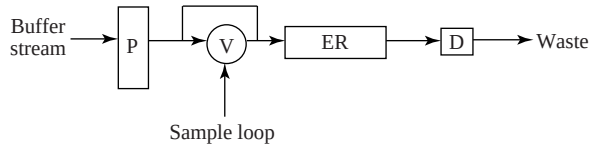

Figure 9 Simplest arrangement of enzymatic flow system. P: pump; V: valve; ER: enzyme reactor; D: detector.

the width of the peak, which is analogous to the exploration in chromatographic measurements. The enzymatic flow systems are calibrated with standard solutions before their application. The operating conditions, as pH, buffer system or buffer flow stream have to be optimized. In FIA systems kinetic controlled reactions take place.

The arrangements of enzymatic flow injection systems vary with the analytical problem to be solved. A review of FIA techniques is given in Ružička and Hansen;[42] the authors introduced this method in the analytical chemistry. Multiline manifolds, combination of several enzyme reactors, enzyme reactors with multienzyme immobilization are described, also.

Enzymatic flow systems are suitable to determine substrates, coenzymes, inhibitors or enzyme activities, and for application such as on-line monitoring, control of biotechnological processes or medicine.

In the following some interesting examples of application of immobilized enzymes in flow systems are represented. The use of immobilized enzymes to construct biosensors, which can also incorporate in flow systems, is explained in section 7.

A FIA system for measuring of acetate, ethanol, glucose, L-lactate and pyruvate is developed by Schürgel and Rhee,[43] and described by Equations (42–46). Alcohol oxidase (AOD, E.C. 1.1.3.13, 9073-63-6), GOD (E.C. 1.1.3.4, 9001-37-0), LDH (E.C. 1.1.1.27, 9001-60-9), lactate monooxygenase (LMO, E.C. 1.13.12.4, 9028-72-2) and sarcosine oxidase (SOD, E.C. 1.5.3.1, 9029-22-5) were immobilized on VA-Epoxy-Biosynth E3 carrier. In that publication the influence of *metabolites* on enzyme based FIA is investigated. Three different detection methods are used. The oxygen consumption (Equations 42–44) is evaluated with an oxygen electrode, the peroxide production (Equation 46) is detected with the redox indicator 2,2′-azino-di(3-ethylbenzthiazoli-sulfonic acid-6) (ABTS), and the NADH reduction (Equation 45) is evaluated spectrophotometrically.

$$\text{ethanol} + O_2 \xrightarrow{\text{AOD}} \text{acetaldehyde} + H_2O_2 \quad (42)$$

$$\text{glucose} + O_2 \xrightarrow{\text{GOD}} \text{gluconic acid} + H_2O_2 \quad (43)$$

$$\text{L-lactate} + O_2 \xrightarrow{\text{LMO}} \text{acetate} + CO_2 + H_2O \quad (44)$$

$$\text{pyruvate} + \text{NADH} + H^+ \xrightarrow{\text{LDH}} \text{L-lactate} + \text{NAD}^+ \quad (45)$$

$$\text{sarcosine} + H_2O + O_2 \xrightarrow{\text{SOD}} \text{glycine} + \text{formaldehyde} + H_2O_2 \quad (46)$$

Since acetate is a competitive inhibitor of the reaction with SOD, the reduction of the reaction rate (Equation 46) is proportional to the concentration of acetate.

A recent trend in analytical chemistry is the miniaturization. As an example *capillary flow injection* (CFI) is combined with an immobilized enzyme reactor (open tubular capillary reactor). The glucose determination in juice and soft drinks has been described.[44]

Chloramine is used as bleaching and disinfecting agent or antiseptic. Its determination, based on an inhibition reaction, is discussed.[45] Chloramine is an *inhibitor* of peroxidase. The flow injection peroxidase reactor is made out of a steel tube in which the enzyme immobilizate (polymer carrier) is filled. A detection limit for chloramide of 26 μmol L^{-1} was obtained.

An automated system for *multichannel* FIA, in which up to six different analytes can be determined sequentially, is described.[46] The enzyme reactors are configured in parallel and only one fluorescence detector is necessary for the sequentially implemented determinations. Furthermore the FIA *stopped flow method* is used to control the analyte conversion and to optimize multistep enzymatic assays.

FIA technique in combination with immobilized enzymes is a powerful tool for on-line monitoring of cultivation processes. The concentrations of penicillin and glucose have been measured online during penicillin production by the filamentous fungus *Penicillium chrysogenum*.[47] In this publication a development of FIA, called *sequential injection analysis* (SIA), has been employed. SIA allows the determination of more than one component in the same manifold.

An example of the use of *coimmobilized enzymes* is given.[48] G6PDH and diaphorase (E.C. 1.6.99.1, 9001-68-7) are employed to enhance the sensitivity of NAD(P) coenzymes as on-line amplifiers on substrate recycling (see section 5.1) in a chemiluminometric flow injection system.

The combination of immobilized enzymes in flow systems with calorimetric detection is used in the *enzyme thermistor device* (ET).[49] The immobilized enzymes arranged in columns and the temperature is generally measured at the top of the column with a thermistor. As a result of the enzyme catalyzed reaction, a temperature peak can be detected. The height of the peak is proportional to the enthalpy change, corresponding to a definite substrate concentration. Normally, the ET device is calibrated with known concentrations of an analyte. With such an arrangement many substances can be determined.[50]

Another recently developed flow instrument is *bypass trapped flow analysis system* (ByT-FAS).[53] The ByT-FAS is applied to evaluate enzyme kinetic parameters by use of coupled enzyme assay (HK and G6PDH) and fluorescence detection.

Flow injection analytical biosystems based on *open-tube and packed bed enzyme reactors* have been compared.[52] Three different enzymes: urease (E.C. 3.5.1.5, 9002-23-5); creatine iminohydrolase (E.C. 3.5.4.21, 37289-15-9); and creatine amidohydrolase (E.C. 3.5.3.3, 37340-58-2) were covalently immobilized on CPG beads as the filling for the packed-bed reactor and immobilized covalently on the inner wall of nylon tube reactors. The authors notice, that packed bed reactors would be ideal for applications requiring a high sensitivity or measurements of low concentration, whereas open-tubular reactors will be best for applications involving samples with a high concentration or requiring saving in the reagents used or shorter analysis times.

7 IMMOBILIZED ENZYMES IN BIOSENSOR DEVELOPMENT

One of the first applications of immobilized enzymes in analytical chemistry was the construction of an enzyme electrode for glucose determination. An oxygen electrode, Clark type, was combined with a thin layer of GOD entrapped in a polyacrylamide gel.[53] Such enzyme electrode form a major category of so called *biosensors*.

Literature lacks a standardized definition of the term biosensor. Generally, the biosensor is a measuring device that contains a biological element.[54] The biological or biorecognition element may be an enzyme, antibody or microorganism system. Besides the biological element, or in sensor development called receptor, the biosensor consists of a transducer, that interprets the interaction of the analyte with the receptor as an electrical signal, related to the concentration (or activity) of a chemical or biochemical analyte (see Figure 10). For instance an enzymatic catalyzed reaction may produce ions, electrons, heat, color or changes in mass, used in the transducer to provide an electrical output. The receptor as a biorecognition element imparts the required high analyte selectivity to the transducer.

Usually, the enzyme is immobilized directly at the surface of the transducer or entrapped in gel or rather in membranes, fixed immediately at the transducer. Pertinent information about the immobilization of biological components for biosensor application is published by Eggins.[55] The optimization of the enzyme immobilization is one of the critical issues of biosensor development. Problems with long term stability of enzyme layers and long response times of biosensors have to be solved. Frequently, membranes used to separate the enzyme layer and the medium. Because of this, additional barriers, not only the enzyme kinetics but also diffusion effects, have to be taken into account. Furthermore, the biosensor signal is also influenced by temperature, inhibitors and pH value. In many practical applications, the highest efficiency can be achieved in combination with FIA systems.

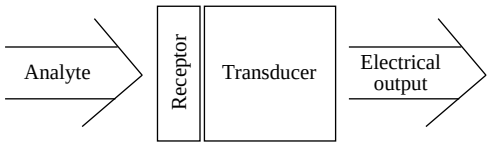

Figure 10 General principle for a biosensor.

Advantages of biosensors are the detection of biological and chemical substances without the need for complex sample processing, simple method for use outside of a laboratory and the great potential of miniaturization by combination with the microsystem technique. Biosensors excel by a high selectivity and sensitivity.

In the following some examples of biosensors from the four principal transducer classes

- amperometric
- potentiometric
- optic
- calorimetric

are represented.

An extensive representation of the different transducer principles is given in Göpel, Hesse and Zemel.[56]

Amperometric biosensors monitor currents in the transducer generated when electrons are exchanged between a biological system and an electrode. Many blood glucose sensors are based on this principle. A prominent representative is the glucose analyzer Yellow Springs Instrument for the measurement of glucose (according to Equations 27 and 28, section 5.1) in whole blood. The applicability of this instrument has been expanded for the determination of for instance lactate, alcohol, uric acid and ascorbic acid. A list of enzyme electrode-based autoanalyzers and their application are given in Kress-Rogers.[51]

Potentiometric devices measure the accumulation of charge density at the surface of an electrode, especially pH-sensitive electrodes are used. Any enzyme reaction, affecting the acid–base equilibrium can be applied. Potentiometric enzyme sensors based on ion-sensitive field effect transistors (ISFETs) are also called enzyme field effect transistors (ENFETs). ENFETs are chiefly used for the determination of glucose, urea and penicillin. Moreover potentiometric enzyme sensors are employed for the determination of inhibitors. Copper(II) in water is determined in the ppm range without preconcentration, based on the urease inhibition and by means of an ISFET.[58] The urease is immobilized onto the pH-sensitive gate area of an ISFET. Authors used a photopolymer system

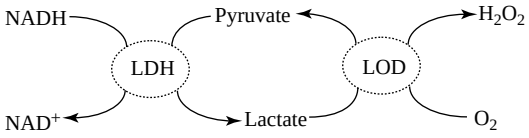

Figure 11 Schematic representation of the operating principle.

with urease to prepare photolithographically patterned enzyme membranes.

An example for the *optical* transducer principle are fiber-optic biosensors. A dual-enzyme fiber-optic biosensor for pyruvate determination is described.[59] This biosensor is based on the fluorometric detection of NADH and immobilized enzyme recycling system (see section 5.1) to amplify the sensitivity. The enzyme layer composed of lactate oxidase (LOD) and LDH is prepared at the tip of a sensing optrode. The described operating principle is illustrated in Figure 11.

The temperature changes during enzymatic reactions, as a consequence of heat exchange, are detected in biosensors with *thermal* transducer principle. A comprehensive representation of such transducer principle is given in Scheper.[60]

An integrated *thermal biosensor* array for multianalyte determination is described.[61] A single microchannel is partitioned into three distinct detection regions. Each of the regions is filled with one immobilized enzyme and contains a pair of thermistors for differential measurement of temperature changes, released by the enzymatic reaction. The pair of thin-film thermistors are placed upstream and downstream of each enzyme matrix. An enzyme-free region is arranged between two adjacent enzyme regions to damp thermal carryover. The authors demonstrate this sensor array with determinations of glucose, urea and penicillin. They applied GOD, urease and ß-lactamase (E.C. 3.5.2.6, 9001-74-5), immobilized on *N*-hydroxysuccinimide (NHS)-activated agarose beads.

Xie et al.[62] applied a *combination of two transducer principles*: a hybrid biosensor for simultaneous *electrochemical* and *thermometric* detection for determination of catechol. This biosensor is based on the tyrosinase (1.14.18.1, 9002-10-2) catalyzed reaction Equation (47):

$$\text{catechol} + \frac{1}{2}O_2 \xrightarrow[\text{response}:\Delta q]{\text{tyrosinase}} \text{1,2 benzochinon} + H_2O \quad (47)$$

with the electrochemical regeneration of catechol by Equation (48):

$$\text{1,2 benzochinon} + 2H^+ + 2e^- \xrightarrow[\text{response}:\Delta I]{} \text{catechol} \quad (48)$$

The hybrid biosensor concept uses the individual advantage of thermal and electrochemical detection principle.

ABBREVIATIONS AND ACRONYMS

ABTS	2,2′-Azino-di(3-ethylbenzthiazoli-sulfonic Acid-6)
ACHE	Acetylcholinesterase
ADP	Adenosine Diphosphate
Al-DH	Aldehyde Dehydrogenase
AOD	Alcohol Oxidase
AP	Alkaline Phosphatase
ATP	Adenosine Triphosphate
ByT-FAS	Bypass Trapped Flow Analysis System
CAH	Creatinine Amidohydrolase
CFI	Capillary Flow Injection
CK	Creatinine Kinase
CPG	Controlled Pore Glass
ENFETs	Enzyme Field Effect Transistors
ET	Enzyme Thermistor
FIA	Flow Injection Analysis
GOD	Glucose Oxidase
G6PDH	Glucose 6-phosphate Dehydrogenase
HK	Hexokinase
ISFET	Ion-sensitive Field Effect Transistor
LDH	Lactate Dehydrogenase
LMO	Lactate Monooxygenase
LOD	Lactate Oxidase
NAD	Nicotinamide Adenine Dinucleotide
NADH	Reduced Form of Nicotinamide Adenine Dinucleotide
NHS	*N*-hydroxysuccinimide
PK	Pyruvate Kinase
SIA	Sequential Injection Analysis
SOD	Sarcosine Oxidase

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Biosensor Design and Fabrication • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis

Food **(Volume 5)**
Enzyme Analysis and Bioassays in Food Analysis

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis

REFERENCES

1. L. Michaelis, M.L. Menten, 'Die Kinetik der Invertinwirkung', *Biochem. Z.*, **49**, 333–369 (1913).

2. G.E. Briggs, J.B.S. Haldane, 'A Note on the Kinetics of Enzyme Action', *Biochem. J.*, **19**, 338–339 (1925).
3. I.H. Segel, *Enzyme Kinetics*, Wiley & Sons, Inc., New York, 54, 1993.
4. T. Palmer, *Understanding Enzymes*, Prentice Hall, New York, 112, 1995.
5. D. Pérez-Bendito, M. Silva, *Kinetic Methods in Analytical Chemistry*, Ellis Horwood, Chichester, UK, 84–102, 1988.
6. R.A. Copeland, *Enzymes*, VCH Publishers, Inc., New York, 191, 1996.
7. I.H. Segel, *Enzyme Kinetics*, Wiley & Sons, Inc., New York, Chapters 3, 4 and 8, 1993.
8. J. Raba, H. Mottola, 'Glucose Oxidase as an Analytical Reagent', *CRC Crit. Rev. Anal. Chem.*, **25**(1), 1–42 (1995).
9. G.D. Christian, *Analytical Chemistry*, Wiley & Sons, Inc., New York, 574, 1994.
10. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, 1983.
11. *Methods of Biochemical Analysis and Food Analysis*, Boehringer Mannheim GmbH, 6–7, 1989.
12. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, 492, Vol. VIII, 1983.
13. J. Lerchner, R. Oehmgen, G. Wolf, P. le Parlour, J.L. Daudon, 'Supermicrocalorimetric Devices for the Investigation of Small Samples', *High Temperature-High Pressures*, **30**, 701–708 (1998).
14. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, 394, Vol. I, 1983.
15. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, 210–215, Vol. I, 1983.
16. H.A. Mottola, 'Kinetic Aspects of Analytical Chemistry', in *Chemical Analysis*, ed. J.D. Winefordner, Wiley & Sons, New York, 68–70, Vol. 96, 1988.
17. H.U. Bergmeyer, *Methoden der Enzymatischen Analyse*, Akademie Verlag, Berlin, 394–395, Vol. I, 1970.
18. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, 391, Vol. I, 1983.
19. S. Gul, S.K. Sreedharan, K. Brocklehurst, 'Enzyme Assays', in *Essential Data*, eds. D. Rickwood, B.D. Hames, John Wiley & Sons, 44–49, 1998.
20. R. Hüttl, K. Oehlschläger, G. Wolf, 'Calorimetric Investigations of the Enzyme Catalysed Sucrose Hydrolysis', *Thermochim. Acta*, **325**, 1–4 (1999).
21. W. Gerhartz, *Enzymes in Industry*, VCH Verlagsgesellschaft mbH, Weinheim, 26–28, 1990.
22. I.H. Segel, *Enzyme Kinetics*, Wiley & Sons, Inc., New York, 242, 1993.
23. G.G. Guilbault, 'Enzymatic Methods of Analysis', in *Analytical Chemistry*, eds. R. Belcher, H. Freiser, Pergamon Press, Elmsford, NY, Vol. 34, 1970.
24. D. Schomburg, M. Salzmann, *Enzyme Handbook*, Springer-Verlag, New York, 1990.
25. K.E. Geckeler, H. Eckstein (eds.), *Bioanalytische und Biochemische Labormethoden*, Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, 74, 1998.
26. R.A. Copeland, *Enzymes*, VCH Publishers, Inc., New York, 64, 1996.
27. G.G. Guilbault, 'Enzymatic Methods of Analysis', in *Analytical Chemistry*, eds. R. Belcher, H. Freiser, Pergamon Press, Elmsford, New York, 183, Vol. 34, 1970.
28. H.-U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, Vol. VII, 1983.
29. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie, Weinheim, 597–599, Vol. VII, 1983.
30. R.A. Copeland, *Enzymes*, VCH Publishers, Inc., New York, 204, 1996.
31. H.U. Bergmeyer, *Methods of Enzymatic Analysis*, 3rd edition, Verlag Chemie GmbH, Weinheim, 1983.
32. G.G. Guilbault, 'Enzymatic Methods of Analysis', in *Analytical Chemistry*, eds. R. Belcher, H. Freiser, Pergamon Press, Elmsford, New York, 200–202, Vol. 34, 1970.
33. K.E. Geckeler, H. Eckstein (eds.), *Bioanalytishce und biochemische Labormethoden*, Friedr. Vieweg & Sohn Verlagsgesellchaft mbH, 74, 1998.
34. K. Stein, G. Ohlenbusch, 'Inhibition of the Enzyme Phosphoenolpyruvate-carboxylase (PEPC) by Different Pollutants', *Talanta*, **44**, 475–481 (1997).
35. E. Wittekind, M. Werner, A. Reinicke, A. Herbert, P. Hansen, 'A Microtiter-plate Urease Inhibition Assay – Sensitive, Rapid and Cost-effective Screening for Heavy Metals in Water', *Environmental Technology*, **17**(6), 597–603 (1996).
36. U. Obst, R. Renneberg, 'Enzymtests', in *Biochemische Methoden zur Schadstofferfassung im Wasser*, eds. Fachgruppe Wasserchemie der GDCH, VCH Verlagsgesellschaft mbH, Weinheim, 21–36, 1993.
37. N. Mionetto, R. Rouillon, J.L. Marty, 'Inhibition of Acetylcholinesterase by Organophosphorus and Carbamates Compounds. Studies on Free and Immobilised Enzymes', *Z. Wasser-Abwasser-Forschung*, **25**, 171–174 (1992).
38. K. Oehlschläger, R. Hüttl, G. Wolf, 'Calorimetric Investigations of Urease Inhibition by Heavy Metal Ions', *Thermochim. Acta*, **310**, 185–189 (1998).
39. B. Atkinson, F. Mavituna, *Biochemical Engineering and Biotechnology Handbook*, Stockton Press, Basingstoke, UK, 568–599, 1991.
40. G.F. Bickerstaff, 'Immobilisation of Enzymes and Cells', in *Methods in Biotechnology*, ed. J.M. Walker, Humana Press, Totowa, NJ, 1997.
41. J. Ružička, 'Flow Injection Analysis', in *Chemical Analysis*, ed. J.D. Winefordner, Wiley & Sons, Inc., New York, 15, Vol. 62, 1988.
42. J. Ružička, 'Flow Injection Analysis', in *Chemical Analysis*, ed. J.D. Winefordner, Wiley & Sons, Inc., New York, Vol. 62, 1988.
43. J.I. Rhee, K. Schügerl, 'The Influence of Metabolites on Enzyme Based Flow Injection Analysis', *Anal. Chim. Acta*, **355**, 55–62 (1997).

44. Y. Shi, S.R. Crouch, 'Micro-scale Determination of Glucose by Capillary Flow Injection with an Immobilised Enzyme Reactor', *Anal. Chim. Acta*, **381**, 165–173 (1999).
45. S. Buck, K. Stein, G. Schwedt, 'Use of a Peroxidase Reactor in Flow Injection Analysis for the Determination of Chloramine and the Inhibition Kinetics', *Anal. Chim. Acta*, **390**, 141–146 (1999).
46. U. Spohn, J. van der Pol, R. Eberhardt, B. Joksch, C. Wandrey, 'An Automated System for Multichannel Flow Injection Analysis', *Anal. Chim. Acta*, **292**, 281–295 (1994).
47. R.W. Min, J. Nielsen, J. Villadsen, 'On-line Monitoring of Glucose and Penicillin by Sequential Injection Analysis', *Anal. Chim. Acta*, **320**, 199–205 (1996).
48. T. Yao, H. Ogawa, T. Nakahara, 'Highly Selective and Sensitive Detection of NADP Coenzymes Using Coimmobilised Glucose-6-phosphate Dehydrogenase/Diaphorase Reactors as On-line Amplifiers Based on Substrate Recycling in a Chemiluminometric Flow Injection System', *Talanta*, **42**, 1297–1303 (1995).
49. B. Danielsson, 'Enzyme Thermistor Devices', in *Bioprocess Technology*, ed. L.J. Blum, Marcel Dekker, New York, 83–105, Vol. 15, 1991.
50. B. Xie, K. Ramanathan, B. Danielsson, 'Principles of Enzyme Thermistor Systems: Application to Biomedical and Other Measurements', in *Advances in Biochemical Engineering/Biotechnology*, ed. T. Scheper, Springer-Verlag, New York, 1–33, Vol. 64, 1999.
51. S.W. Hillard, K.K. Stewart, 'A Bypass Trapped-flow Analysis System Evaluation of Enzyme Kinetic Parameters with a Coupled Enzyme Assay and Fluorescence Detection', *Talanta*, **45**, 507–512 (1998).
52. M. Jurkiewicz, S. Alegret, E. Fábregas, 'Comparison of Flow Injection Analytical Biosystems Based on Open-tube and Packed-bed Enzyme Reactors', *Anal. Chim. Acta*, **370**, 47–58 (1998).
53. S.J. Updicke, G.P. Hicks, 'The Enzyme Electrode', *Nature*, **214**, 986–988 (1967).
54. D.G. Buerk, *Biosensors*, Technomic Publishing Company, Inc., Lancaster, 1–2, 1993.
55. B.R. Eggins, *Biosensors – An Introduction*, John Wiley & Sons Ltd. and B.G. Teubner, Chichester, Chapter 3, 31–50, 1996.
56. W. Göpel, T.A. Jones, M. Kleitz, I. Lundström, 'Chemical and Biochemical Sensors Part I and II', in *Sensors – A Comprehensive Survey*, eds. W. Göpel, J. Hesse, J.N. Zemel, VCH Verlagsgesellschaft mbH, Weinheim, Vol. 2/3, 1992.
57. E. Kress-Rogers, *Handbook of Biosensors and Electronic Noses: Medicine, Food, and the Environment*, CRC Press, 250–251, 1997.
58. A. Zürn, H. Müller, 'Photolithographically Patterned Enzyme Membranes for the Detection of Pesticides and Copper(II) Based on Enzyme Inhibition', *Fresenius J. Anal. Chem.*, **346**, 589–593 (1993).
59. W. Zhang, H. Chang, G.A. Rechnitz, 'Dual-enzyme Fiber Optic Biosensor for Pyruvate', *Anal. Chim. Acta*, **350**, 59–65 (1997).
60. T. Scheper (ed.), 'Thermal Biosensors, Bioactivity, Bioaffinity', in *Advances in Biochemical Engineering/Biotechnology*, Springer-Verlag, New York, Vol. 64, 1999.
61. B. Xie, B. Danielsson, 'An Integrated Thermal Biosensor Array for Multianalyte Determination Demonstrated with Glucose, Urea, and Penicillin', *Anal. Lett.*, **29**(11), 1921–1932 (1996).
62. B. Xie, X. Tang, U. Wollenberger, G. Johansson, L. Gorton, F. Scheller, B. Danielsson, 'Hybrid Biosensor for Simultaneous Electrochemical and Thermometric Detection', *Anal. Lett.*, **30**(12), 2141–2158 (1997).

Instrumentation for Kinetics

Emil A. Cordos
University ''Babes-Bolyai'', Romania

1 Introduction **1**
2 Sample and Reagent Preparation, Reactant Proportioning and Mixing **3**
2.1 Stopped-flow Analyzers 3
2.2 Flow Analyzers 7
2.3 Continuous Reagent Addition Method 10
3 Reaction Monitoring Systems **11**
3.1 Single-channel Detection 11
3.2 Multichannel Detection 14
3.3 Combined Detection 15
4 Data Handling **16**
4.1 Direct Computation Methods 17
4.2 Curve-fitting Methods 18
Abbreviations and Acronyms **19**
Related Articles **19**
References **20**

In kinetic methods or reaction-rate methods of analysis, the rate of a given reaction is measured and related to the concentration of the sought-for species. The instrumentation in kinetics is primarily designed to monitor the evolution of the analytical signal with time. Unlike in the traditional equilibrium-based approaches, where the signal is measured after the reaction has reached completion, in kinetic methods the signal is usually measured in the first few percent of the total reaction time. Therefore, the advantages and limitations of kinetic methods should be considered in their relations with equilibrium methods. Kinetic methods are faster than equilibrium-based methods since the user does not have to wait for a reaction to go to completion. They are relative methods that depend on changes in concentration and not on its absolute value and they could be more selective than equilibrium methods since the difference between reaction constants could be used to discriminate against an interfering species. The limitations of kinetic methods are a lower precision and a strong dependence on experimental conditions. The main factors that cause these limitations are the measured signal, which is small in comparison with that from an equilibrium reaction, and the dependence of rate constants on temperature, pH, ionic strength, etc. Therefore, the measurement systems used in kinetic methods should be sensitive and the experimental conditions should be kept strictly constant. The best results in kinetic methods are obtained with instrumental systems having a high degree of automation.

The main operations that should be carried out by the instrumentation in kinetic determinations are the mixing of the reactants, measurement of the detector signal in time and processing of data to obtain the reaction rate and concentration. The mixing of the reactants should be done in a time shorter than the reaction half-life and it is critical for fast reactions where required mixing times are of the order of milliseconds. In these cases the mixing is accomplished by mechanical systems that rapidly drive the reactants through a mixing chamber into an observation chamber. For slower reaction the best mixing procedures are based on automated flow analyzers that have different manifolds configurations adapted both to sample processing and to reactant mixing. After mixing, the reaction could be monitored in time either by stopping the flow and recording the signal or by passing the reacting plug several times through the detector and reading the signal after each pass. The measurement of the detector signal, after the reagents are mixed, could be done with a large variety of detectors, but the most commonly used detection systems are photometric and electrochemical. Photometric detection provides a short response time, specificity and the possibility of collecting full spectra very rapidly if photodetector arrays are used. Electrochemical detectors are simple and relatively inexpensive. The concentration of the analyte is determined from the slope of the detector signal. In many routine procedures the slope is determined by measuring the signal at one or two fixed points in time or by determining the time required by the signal to reach a predetermined level. These procedures have the advantage that they are rapid and need no supplementary instrumentation for data processing. Another approach in kinetic data processing, based on computer software, computes a curve that fits the experimental data. The slope or concentration is then extracted from the mathematical expression of the fitting curve.

1 INTRODUCTION

Kinetic methods of analysis are based on the determination of the rate of a given reaction in which the sought-for species is involved. The reaction rate is then correlated with the concentration. Consequently, the instrumentation in kinetics is primarily designed to monitor the evolution of the analytical signal with the time.

The use of kinetic methods is always considered in their relation with the well-established equilibrium methods because many analytical determinations could be

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

performed by either of the two types of method.[1–3] The main advantages of kinetic methods concern speed, relativity and selectivity. Kinetic methods have a potential speed advantage over equilibrium-based methods since the user does not have to wait for a reaction to go to completion. They are inherently relative methods and do not depend on absolute concentration but on changes in concentration. Kinetic methods could be more selective than equilibrium methods. The difference between reaction constants could be used to discriminate against an interfering species or to perform multicomponent determinations. Kinetic methods permit the determination of catalyst concentration, which cannot be readily determined by equilibrium-based methods.

The limits of kinetic methods should be carefully considered when using or designing instrumentation for kinetic determinations. The change in analytical signal that is measured is small in comparison with that from an equilibrium reaction. The consequence is a lower precision and the need to have very sensitive measurement systems. The rate of a reaction is dependent, via the rate constants, on experimental conditions: temperature, pH, ionic strength, etc. Therefore, the experimental conditions should be kept strictly constant. This is why automated systems give the best results in kinetic determinations and many of the modern instruments involved in rate methods have a high degree of automation.

The main operations in a kinetic method of analysis are the mixing of the reactants, measurement of the detector signal in time and processing of data to obtain the reaction rate and concentration. For each of these stages a unit in a block diagram of the instrument could be identified. A general block diagram of a complete instrumentation system for reaction rate methods is given in Figure 1.

The mixing of the reactants is a key operation in kinetic methods. In most kinetic methods the use of the initial rate of reaction is recommended, which means that the rate measurement should be made just after the reaction was initiated. Therefore, the time taken for mixing should be shorter than the reaction half-life. In the case of fast reactions, the required mixing times are of the order of milliseconds. For these reactions the mixing is accomplished by mechanical systems that drive the reactants through a mixing chamber, rapidly enough to have turbulent flow and rapid and thorough mixing. The observation is made in a region very close to the point where the reactants were completely mixed.

For slower reactions, the mixing could be done manually or with automated systems. Flow analyzers are the most widespread instrumental solution for the automation of kinetic methods. Once the reactants have been mixed in the analyzer manifold, the reaction could be monitored in time either by stopping the flow and recording the signal or by passing the reacting plug several times through the detector and reading the signal after each pass.

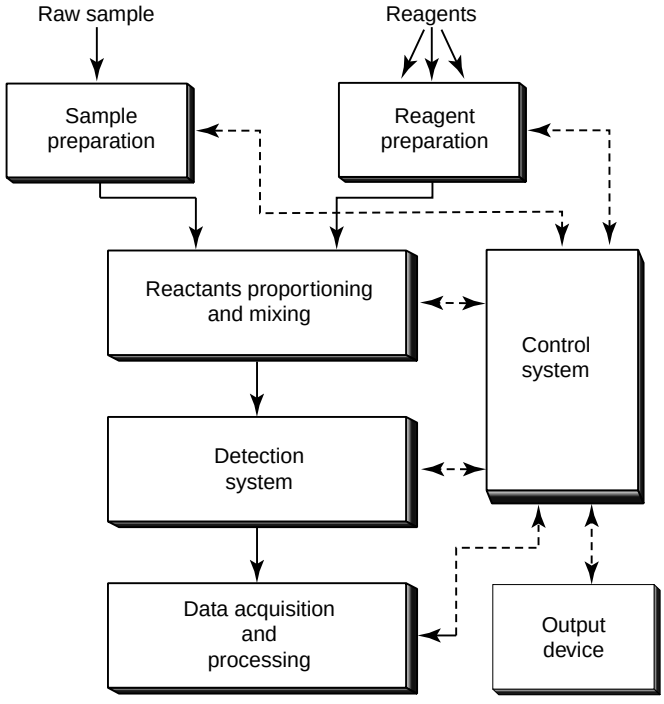

Figure 1 Block diagram of a complete instrumentation system for reaction rate methods. [Reprinted with permission from H.V. Malmstadt, E.A. Cordos, C.J. Delaney, *Anal. Chem.*, **44**(12), 26A–41A. Copyright 1972 American Chemical Society.]

The monitoring of the analytical signal after the reagents are mixed could be done with a large variety of detectors that should be selected considering the type of reaction and its rate. The detectors should be sensitive enough to detect small changes in the monitored parameter and their response time must be much shorter than the reaction half-life. The most commonly used detection systems are electrochemical and photometric. Electrochemical detectors are simple and relatively inexpensive. However, most detection methods used in kinetic determinations are photometric. Photometric detection is attractive because of the very short response time of modern photodetectors and the specificity of the spectrophotometric measurements. A wealth of information could be obtained if photodetector arrays are used to collect full spectra during the initial stage of the reaction.

In the final step of analysis the concentration should be determined from the slope of the detector signal. The slope could be determined explicitly if the signal was recorded in time, as a tangent to the kinetic curve. For routine applications, however, it is more convenient to simplify the slope determination by measuring the signal at one or two fixed points in time or to determine the

time required for the signal to reach a predetermined level (fixed-time and variable-time methods). These procedures predominate in routine analysis and even in many publications since they are rapid and need little or no supplementary instrumentation for data processing.

A more modern approach to kinetic data processing is based on computer software. Instead of the direct determination of the slope and concentration, a curve that fits the experimental data is computed. All the required data are then extracted from the mathematical expression for the fitting curve. The advantages of a curve-fitting procedure are multiple, e.g. a reduced effect of the noise and the possibility of extracting more information about rate constants and about the values of the equilibrium response by using a predictive manner approach. In the same category of advantages are the applications of error compensation methods and multicomponent analysis.

2 SAMPLE AND REAGENT PREPARATION, REACTANT PROPORTIONING AND MIXING

The sample and reagent preparation and the proportioning and mixing of the reactants are the most important, and sometimes the weakest, links in the chain of the instrumentation system. The development and automation of kinetic instrumentation in the last three decades focused on these parts of the diagram in Figure 1.

The sample preparation system must provide treatment operations that include dissolution in the case of solid samples, pH adjustment, extraction and removal of interfering components. The reagent preparation, in addition to simply preparing the reagent, is often used to dispense and mix the reagent in form of a single composite reagent.

The first approach to sample and reagent handling and mixing was that of manually pipeting sample and reagents into a reaction vessel that served both as mixing chamber and observation cell (the pour-and-shake technique). For a slow reaction the reactants could be mixed externally and then transferred to the observation cell. Most spectrophotometers have a pumping system that could aspirate the sample and reagent mixture and transfer it in a cuvette.

Today most kinetic instrumentation is automated to various degrees and adapted to fast or slow reactions. The sample and reagent preparation is often accomplished within the instrumental system, on-line, especially in flow analyzers. There are two main categories of techniques and analyzers in approaching the reactants preparation and mixing: stopped-flow analyzers for fast reaction and flow analyzers for slower reactions. A different novel approach that could be applied for fast reactions is based on continuous reagent addition.

2.1 Stopped-flow Analyzers

The stopped-flow technique consists of rapidly mixing the reactants by forcing the solutions through a mixing chamber and into an observation cell. The flow of solution is abruptly stopped, creating a back-pressure which completes the mixing, and the rate measurement is rapidly made. The more sudden the stopping, the faster the reaction can be observed. The stopped-flow technique is suitable for fast reactions with half-lives of a few milliseconds to few seconds.

The term ''stopped-flow'' is also used in continuous-flow techniques. It indicates a similar principle of operation that consists in halting the flow, recording the detector signal and measuring the slope of the obtained kinetic curve. However, the operating conditions are different from that of the normal stopped-flow technique. The relatively low flow rate of the solutions and the fact that sample and reagents are not mixed under high pressure and turbulent flow conditions does not allow the measurement of fast reactions. Therefore, the term ''stopped-flow mode''[4] will be used for this type of operation in continuous-flow techniques.

The basic components of a stopped-flow system are the driving syringes, mixing chamber, observation cell and the stopping syringe. The detection system used in connection with the observation cell should measure a certain property of the mixed solutions. The detection systems are spectrophotometric, fluorescence or both.[5] A few other types of detection systems for analytical application have also been devised.[6,7] However, many detection systems that are not photometric were intended rather for kinetic studies of reactions.

An example of an automatic stopped-flow system is shown schematically in Figure 2.[1] The driving syringes could be actuated by a motor and cam device or by a pneumatic system. Upon activation of the system, the sample and the reagents are drawn into the syringes. Then, after the three-way stopcock has been set in a proper position by the stopcock motor, the actuating system pushes the pistons and the solutions are rapidly driven out of the syringes. The solutions are forced into the mixing chamber and on to the observation cell. The reactants mixture flows through the observation cell until the flow is abruptly stopped by the stopping syringe. From the instant the syringe hits the mechanical stop the reaction rate curve could be recorded or the data fed to a computer. A family of stopped-flow kinetic curves obtained for the determination of lysozyme with fluorimetric detection is given in Figure 3.[8]

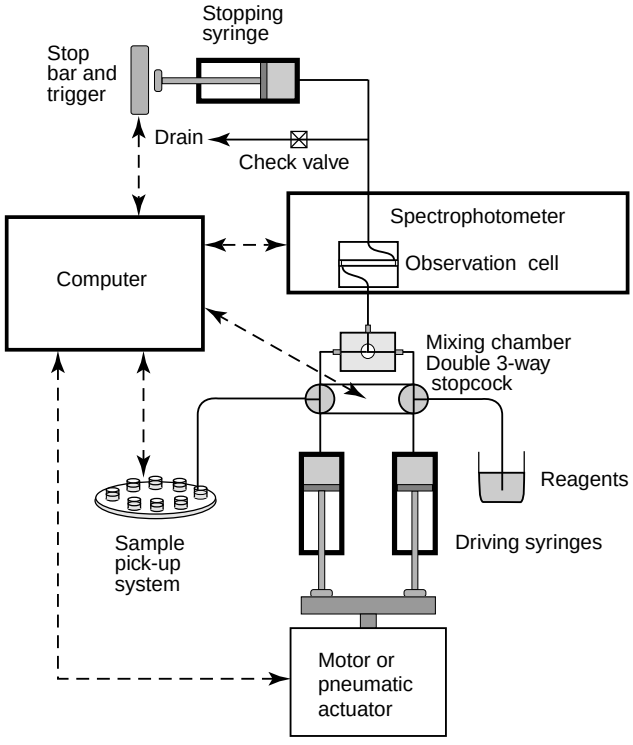

Figure 2 Schematic diagram of an automatic stopped-flow system. [Reproduced with permission from V. Malmstadt, E.A. Cordos, C.J. Delaney, *Anal. Chem.*, **44**(12), 26A–41A. Copyright 1972 American Chemical Society.]

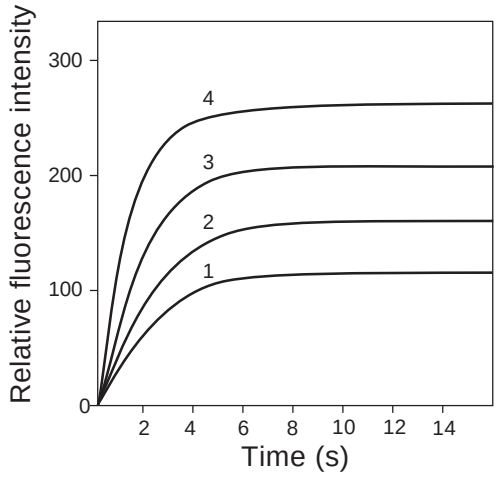

Figure 3 Stopped-flow kinetic curve family. Kinetic curves with fluorimetric detection obtained for cresyl violet–sodium dodecyl sulfate–lysozyme system at different lysozyme concentrations: (1) 10, (2) 20, (3) 30 and (4) 40 mg mL^{-1}. Note that the initial reaction rate could be obtained within only 2 s. [Reprinted from B. Gala, A. Gómez-Hens, D. Pérez-Bendito, *Talanta*, **43**(9), 1413–1419 (1996), Copyright 1996, with permission from Elsevier Science.]

The efficiency of the stopped-flow system is dependent upon the flow velocity, mixer design and the distance between the mixer and the observation cell. It implies that the flow of solution must be turbulent. At a Reynolds number of about 2000 for short tubes with streamline entry, turbulent conditions, and consequently good mixing, are achieved at flow rates usually higher than 1 m s^{-1}. However, a very high flow rate could lead to cavitation effects produced by the unequal hydrostatic pressure of the streaming fluid in different portions of the tube. This will affect the precision of the reaction rate determinations, especially for spectrophotometric detection. Therefore, a compromise is made so that the flow rate is high enough to produce a turbulent flow and low enough to prevent cavitation.[9]

In obtaining a fast flow rate, the quality of the driving and stopping syringes is a key factor. The syringes are usually made from frosted-glass barrels and ceramic-tipped plungers. The problems that may arise in syringe functioning are those resulting from the difference in thermal expansion of the barrel material and plunger and those related to sealing and lubrication. Both of these problems could lead to breakage or ''blow-by'' during rapid acceleration that is required to obtain fast flow rates and to reduce the instrument dead time. Part of these could be avoided by using smooth-bore glass barrels and Teflon™ plunger tips.[10]

Most stopped-flow instruments are provided with thermostating systems for the driving syringes and reagents reservoirs to insure the constant temperature required by reaction rate measurements. Uncontrolled temperature changes may have undesirable effects on the rate of studied reaction and on the precision of analytical measurements. Temperature changes may result from the thermal gradient between different parts of the system or by heating during the compression and mixing of the reagents.[11]

The mixing efficiency, in turn, is dependent of mixing design. High mixing efficiencies are obtained using tangentially offset jets that promote a rotary motion of the fluid.[12,13] An increased number of jets and a higher jet velocity also contribute to higher efficiency. For a given mixer it means more jets with a smaller jet bore diameter. To avoid excess pressure drop in the mixer, the total area of the jet bores should be equal to that of the section of the observation cell. More than one mixer could be used, in cascade, both to increase the efficiency and to mix more than two different solutions.

A measure of the mixing efficiency is the dead time. It is defined as the time interval from the initial contact of the two solutions in the mixing chamber to the observation measurement point in the cell. If d_m is the distance, in meters, from the mixing chamber outlet to an observation point where the solutions are thoroughly mixed and v is the flow velocity in meters per second, the dead time is

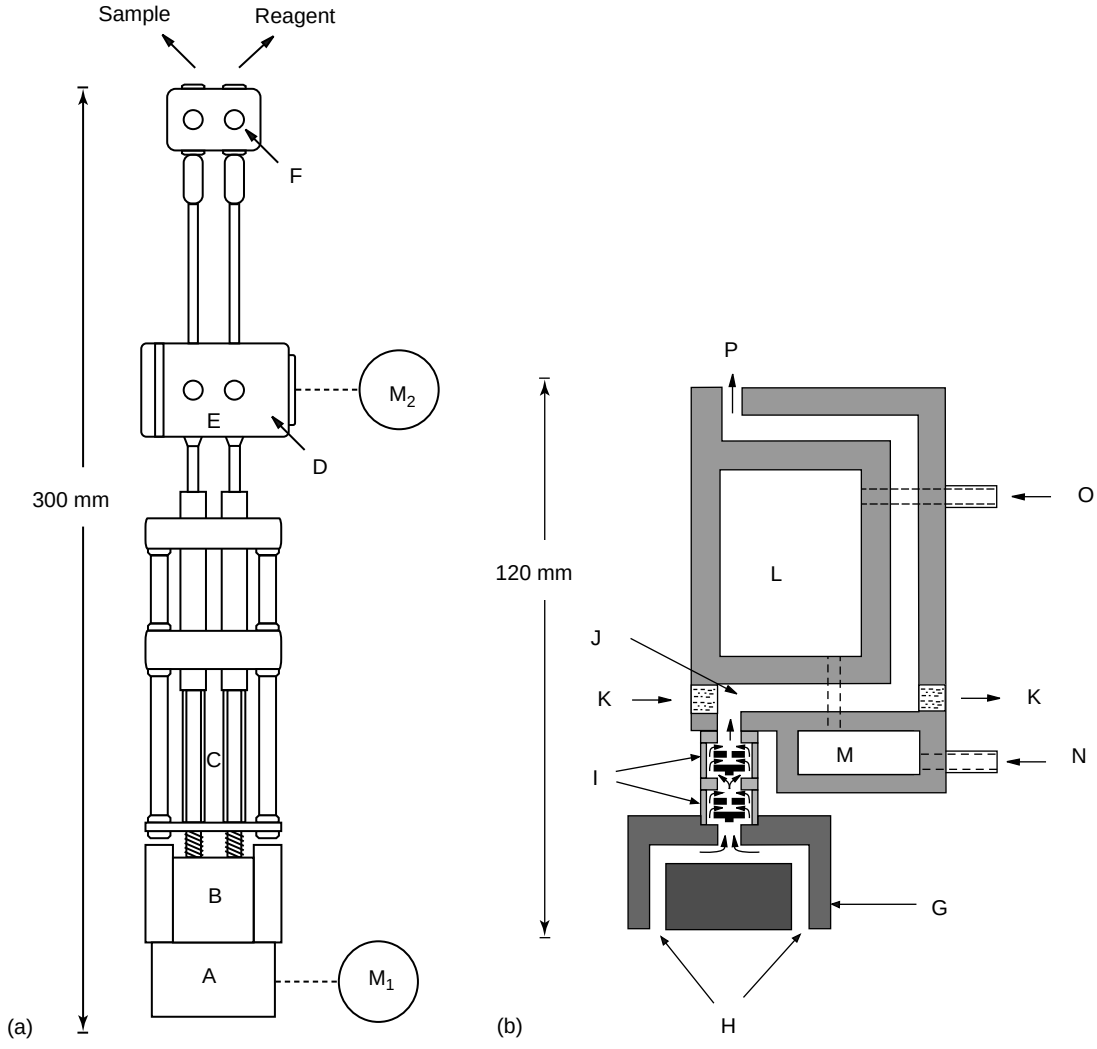

Figure 4 Driving syringe block and mixing system. (a) Reagent delivery unit. A, Teflon™ cam; B, sliding brass bar; C, gas-tight syringes with stainless-steel springs; D, Teflon™ double three-way stop-cock; E, outlet to delivery block; F, three-way Hamilton valves; M_1 and M_2, synchronous motors. (b) Vertical section of delivery block, mixers, and observation chamber. G, delivery block; H, reagent and sample inlets; I, tangential mixers; J, observation cell; K, quartz windows; L and M, thermostating jacket; N, thermostating water outlet; P, reaction mixture outlet. [Reproduced by permission of CRC Press LLC from H.V. Malmstadt, C.J. Delaney, E.A. Cordos, *CRC Crit. Rev. Anal. Chem.*, **2**(1), 559–619 (1972).]

given by Equation (1):

$$t_m = \frac{d_m}{v} \tag{1}$$

It is desirable to provide a dead time as small as possible. This could be accomplished either by improving the mixing design and thus decreasing the distance d_m from the mixer outlet to the point where the solutions are mixed or by increasing the flow velocity, v. Typically the dead time is in the millisecond range.[14,15]

In most instruments the mixer and the observation cell are part of the same unit. A schematic diagram of the driving syringe block and a vertical section of mixing and observation cell with spectrophotometric detection are given in Figure 4(a) and (b).[9] The syringe block could be considered as an aliquoting system that delivers solutions to the mixer. The sample and the composite reagent are mixed in a two-stage mixing chamber, introduced into the observation cell and pushed further through the block outlet to the stopping syringe. To minimize the dead time, the mixing chamber is located as close as possible to the observation cell.

The mixers and the observation cell are built from stainless steel, quartz,[16,17] fused silica[18] or polymers (Kel-F®).[19] In some of the commercial stopped-flow units the mixing chamber and observation cell are designed as a single block that could be easily fitted and removed from the cuvette compartment

of the existing spectrophotometers.[18] The design of stopped-flow syringe block implies that the volumes of solution delivered or their ratio are fixed. Therefore, a composite reagent is often used. To facilitate the adaptation of the method to more complex chemical reactions, mixing devices that allow predilution of the reactants have been devised.[20,21]

2.1.1 *Pulse-accelerated Flow*

Pulse-accelerated flow (PAF) is a procedure that allows the measurement of single-step reactions with half-lives from 0.5 ms to 4 μs.[22] The instrumentation is basically that of a stopped flow with motor-driven syringes (Figure 5). Instead of completely stopping the flow, a deceleration is applied and the flow velocity is reduced from 21 to 2 m s^{-1}. Reactions are observed as mixing takes place. Chemical rate processes are resolved from physical mixing rate process by variation of flow velocities under turbulent flow conditions. An integrating system of observation or position-resolved observation could be used. For position-resolved observation see section 3.2.

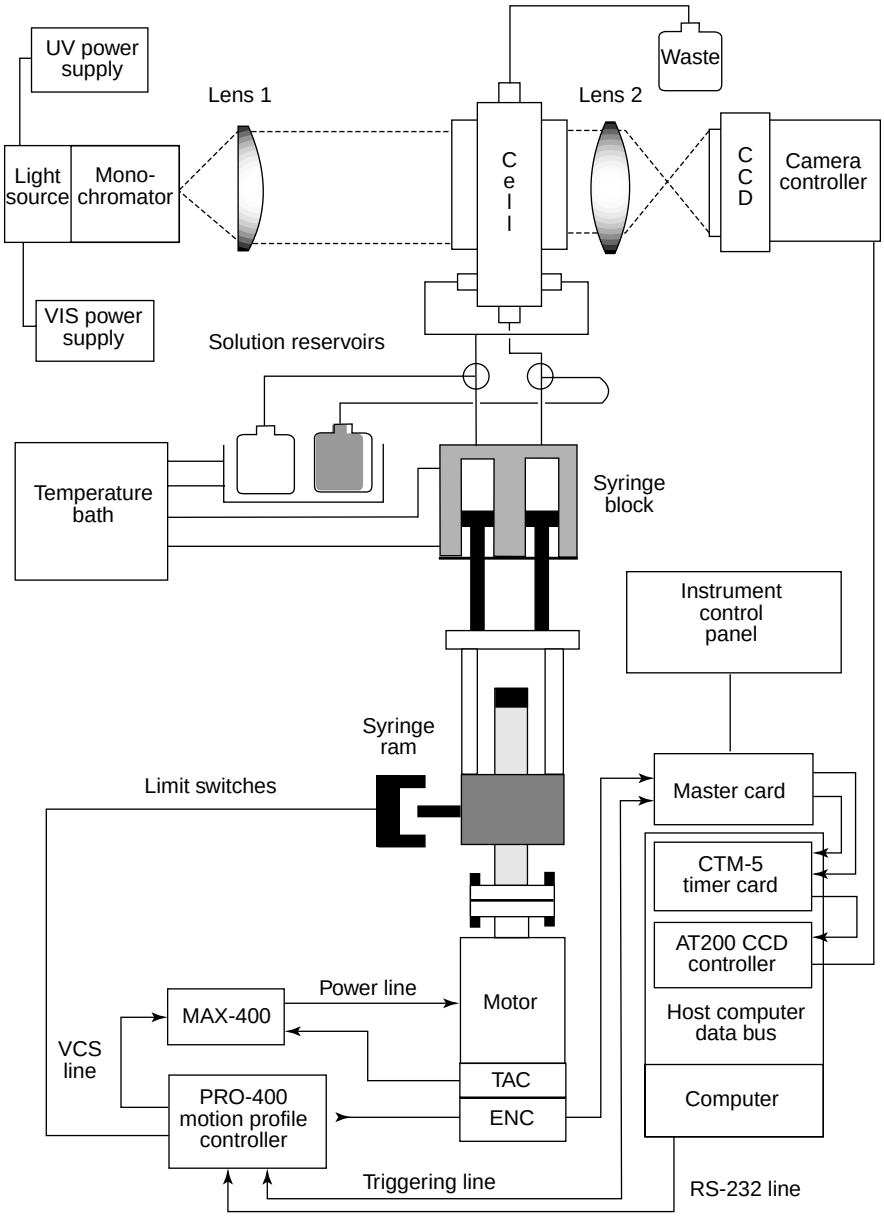

Figure 5 Schematic diagram of PAF spectrometer with position-resolved observation. [Reprinted with permission from M.R. McDonald, T.X. Wang, M. Gazda. W.M. Scheper, S.W. Evetts, D.W. Margerum, *Anal. Chem.*, **69**(17), 3513–3520. Copyright 1997 American Chemical Society.]

2.2 Flow Analyzers

Flow analyzers are the most widespread type of automated instruments that could be adapted both to equilibrium methods and kinetic methods. Flow analyzers are based on the sample treatment as it is transported by a flowing stream through a manifold made of interconnected tubing. There are two main techniques in accomplishing this method. One is based on the segmentation of the reagents flow with air bubbles [air-segmented continuous-flow analysis (ASCFA)], the sample being introduced into the flow by aspiration by the same pumping unit as for the reagents.[23] In the second technique, the reagent flow is continuous, nonsegmented, and the sample is injected into a carrier flow at a certain point of the manifold [flow-injection analysis (FIA)].[24,25] Combinations between these two techniques are also known.[26]

The main components of a flow analyzer are a propelling and aspiration system for reagents and sample, tubing manifold and a detector unit. For the air-segmented technique a debubbling device is included prior to the detection or an electronic bubble gate is used, so that the detection is done only when the solution is in the detector unit.[27] For FIA a sample injection valve is added, usually on the carrier stream.

The most frequently used device for obtaining a constant reagent flow is the peristaltic pump, although a number of other devices have been used: gas-pressure propulsion, syringes, linear and sinusoidal pumps.[28,29] Usually a six-channel peristaltic pump is sufficient for a single-channel analyzer.

The manifold consists of a coiled tube that insures radial mixing of sample and reagent. If more reagents have to be used in succession more streams are confluenced by adding more coils. Coiled tubes provide better mixing than straight tubing.[30] Tubes are made of glass, fused silica, plastics with inner diameters from about 1 mm down to 100 µm for capillary flow injection.[31–35] In the manifold structure could be included miniaturized packed reactors to process the sample further. The packed reactors could contain reagents,[36] immobilized enzymes,[37,38] ion exchangers[39] or polymer-based C_{18} materials.[40] Finally, the stream of reaction products flows through a detection system adapted to the small volumes of solutions.

Using the above-mentioned basic modules of the instrumentation a number of techniques are applied for reaction rate measurements.

2.2.1 Flow Analyzers in Stopped-flow Mode

The instrumentation includes the basic flow analysis set-up and a stop-and-go sequence provided by the computer or the circuitry that controls the instrument. The stream is started and when the mixture of sample and reagent

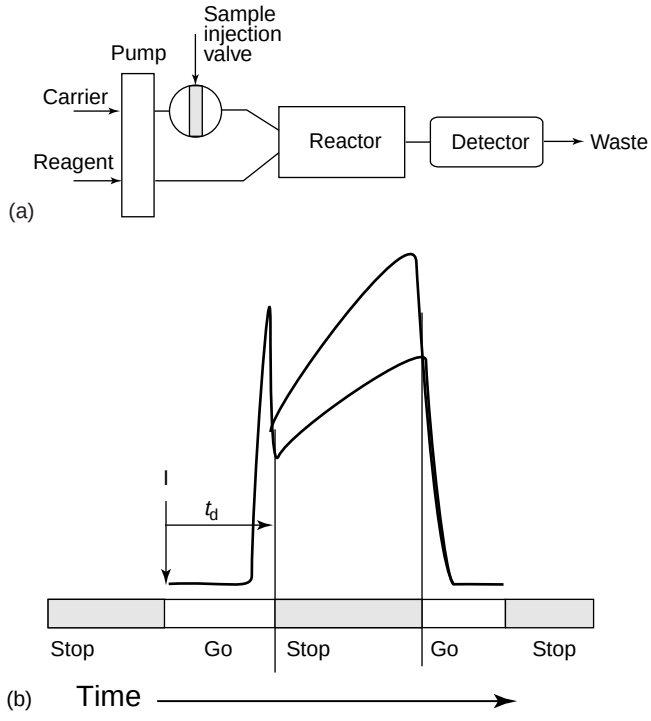

Figure 6 Schematic manifold of a flow injection system used in stopped-flow mode for reaction rate measurement (a) and the rate curves for two analyte concentrations (b). [Reprinted from J. Ruzicka, *Anal. Chim. Acta*, **261**(1–2), 3–10 (1992), Copyright 1992, with permission from Elsevier Science.]

reaches the detector the flow is stopped and the signal changes in time is monitored.[41] It could be accomplished as either an FIA[42] or an ASCFA technique.[43] An example of a stopped-flow injection instrumental set-up and the resulting kinetic curve is given in Figure 6(a) and (b). Here, the reactor is the called the sum of tubing, coils and T-mixers of the manifold where the reaction and/or subsequent reactions take place. By turning the sample valve a given amount of sample is injected into the carrier stream. The first rising portion of the signal indicates the moment the reacting plug has reached the detector. As the stream flows, the signal begins to decrease as the front of the reacting plug passes by. Stopping the flow after a time t_d (delay time) will result in an increase in the signal as reaction proceeds in the observation cell. The slope of the signal during the stop interval (reaction rate) is proportional to the analyte concentration. An example of a kinetic curve family obtained by stopped-flow injection is given in Figure 7 for glucose standards in glucose determination by the Trinder enzymatic reaction.[44]

2.2.2 Sequential Injection Stopped Flow

The techniques of sequential injection (SI) is based on the dispersion and mutual penetration of the sample and

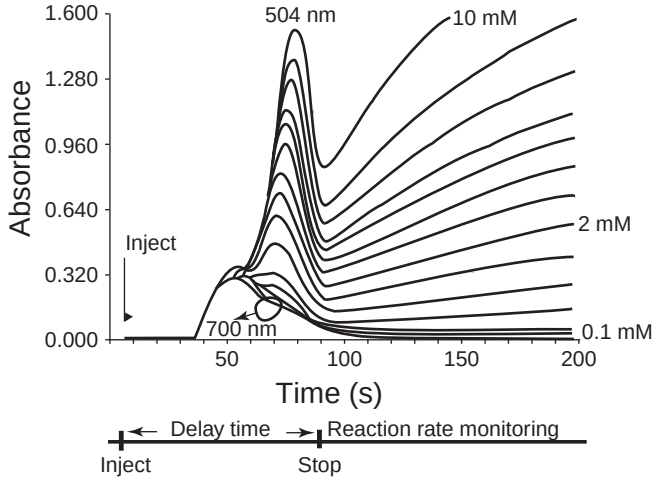

Figure 7 Kinetic curves family obtained by FIA in the stopped-flow mode. The curves represent the absorbance signals at 504 nm in monitoring glucose concentration by use of Trinder reaction. Carrier flow was stopped at 90 s after injection. [Reprinted from G.D. Christian, J. Ruzicka, *Anal. Chim. Acta*, **261**(1–2), 11–21 (1992), Copyright 1992, with permission from Elsevier Science.]

reagents zones as they are stacked in a holding coil, adjacent to each other. The sampling valve has multiple ports so that sample, reagents and the carrier could be alternatively aspirated and propelled through the reactor to the detector (SI).[45] A diagram of the instrumental set-up and its functioning in the stopped-flow mode is given in Figure 8(a–e).

A high-precision piston pump or a peristaltic pump could be used to control the forward and reverse movement of the liquid flow. The first step is the aspiration of precisely measured volumes of carrier, sample and reagent into the holding coil (Figure 8a). Then the pump movement is reversed and the stack is propelled through the selector valve and reactor (Figure 8b) to the detector. The reversal of the flow creates a composite zone in which sample and reagent zones mutually penetrate. When the stack reaches the detector, the pump stops the flow (Figure 8c) and from this moment the kinetic curve could be recorded or the reaction rate could be determined (Figure 8e).

SI requires a simpler mechanics than flow injection. It uses a single pump, a single valve and a single channel. By designing appropriate software the SI technique becomes very versatile in configuring the analytical systems. Injection volumes, reaction times and zone dispersion could be changed via the computer by varying stroke volumes, flow rates and stopped-flow time.

In a recent kinetic application, SI was used for mechanistic studies and kinetic determination of bromazepam.[46]

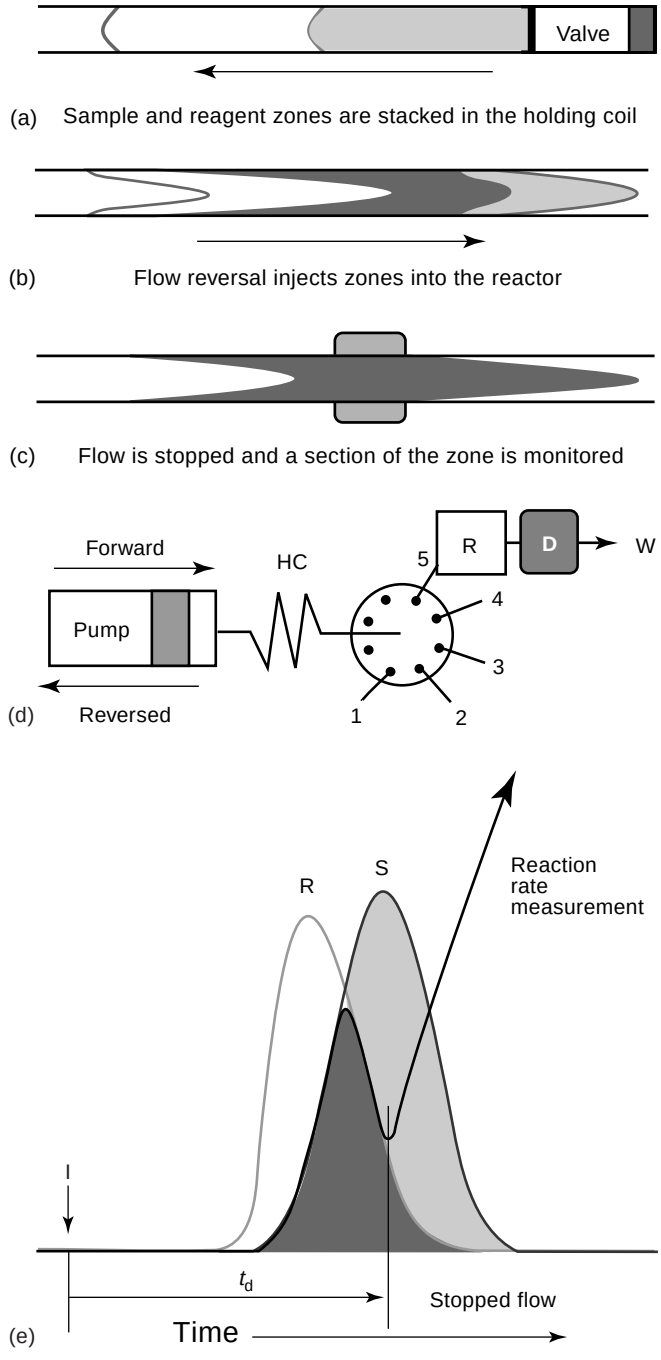

Figure 8 SI combined with stopped-flow method. (a–c) Structure of stacked and reacting zones at different stages of the SI analysis. (d) Schematic diagram of the sequential flow system. HC, holding coil; R, reactor; D, detector; W, waste; 1, port for aspiration of carrier; 2, port for aspiration of sample solution; 3, port for aspiration of reagent solution; 4, port for aspiration of the second reagent, if required. (e) Representation of mutual penetration of reagent and sample zones and the resulting reaction-rate curve as recorded during the stopped-flow period. R, reagent; S, sample; I, point of injection, t_d, delay time. [Reprinted from G.D. Christian, J. Ruzicka, *Anal. Chim. Acta*, **261**(1–2), 11–21 (1992), Copyright 1992, with permission from Elsevier Science.]

2.2.3 Flow Recycle Analyzers

The kinetic information is obtained by passing the reacting plug several times through the detector until it is completely dispersed into the carrier.[47,48] After each pass the signal shows a peak. The number of peaks obtained per sample is a function of its volume, flow rate and the reactor length. The envelope of the maximum (or minimum) of these peaks defines a kinetic curve that allows the application of conventional kinetic methods of determination and calculations of the partial reaction orders and rate constants. The technique could be implemented both with unsegmented flow or air-segmented flow systems. The air-segmented flow system gives lower dispersion and shorter peaks than nonsegmented flow procedure.[49]

The instrumentation involves, in addition to the basic modules of a regular flow analyzer, a switching valve and a unidirectional peristaltic pump. A diagram of a flow recycle system, unsegmented type, and a typical kinetic curve are given in Figure 9(a–c).[50] It works as an open–closed system. The sample is injected and carried until it passes through the detection system. Then the valve is switched and the reacting plug is transported in a closed loop by the unidirectional peristaltic pump. Instrumental versions of this technique in which the recycling is done using only the main peristaltic pump and a pair of switching valves are also known.[51]

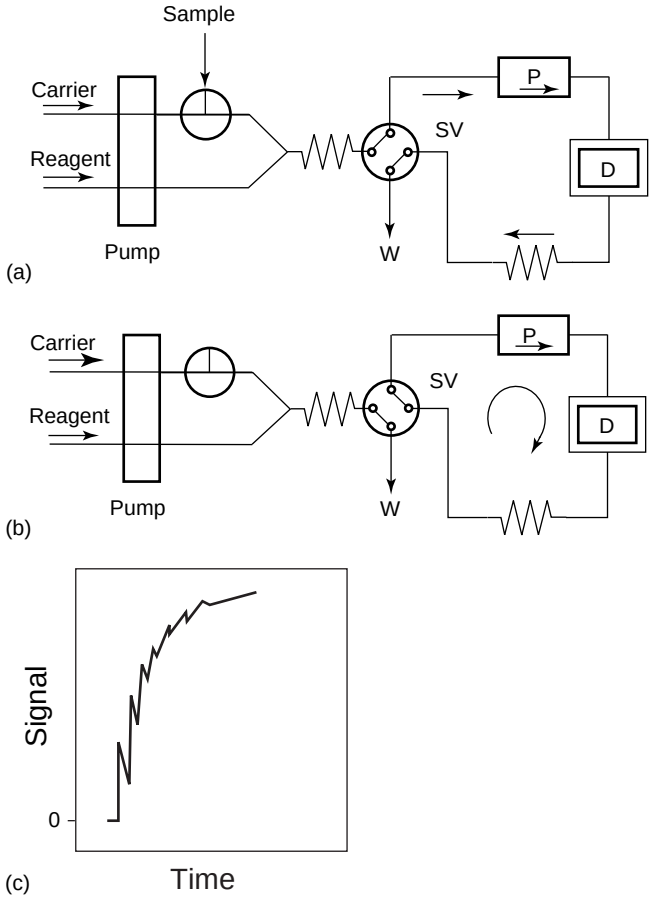

Figure 9 Diagram of a flow recycle system, unsegmented flow technique. SV, switching valve; P, unidirectional pump; D, detector; W, waste. (a) Sample injection and delay period until the reacting plug reaches the detector. (b) Closing of a loop by actuating the switching valve. (c) Typical plot of the detector signal. [Adapted by permission of The Royal Society of Chemistry from M. Silva, *Analyst*, **118**(6), 681–688 (1993).]

2.2.4 Flow Reversal Analyzers

The kinetic response of this type of analyzers is obtained by passing the reacting plug several times through the detector, but alternately in opposite directions.[52–54] The manifold for an air-segmented flow reversal analyzer and the corresponding response curve are given in Figure 10(a–c). The reversal of the flow direction, via the selecting valve, is activated when the detector signal falls

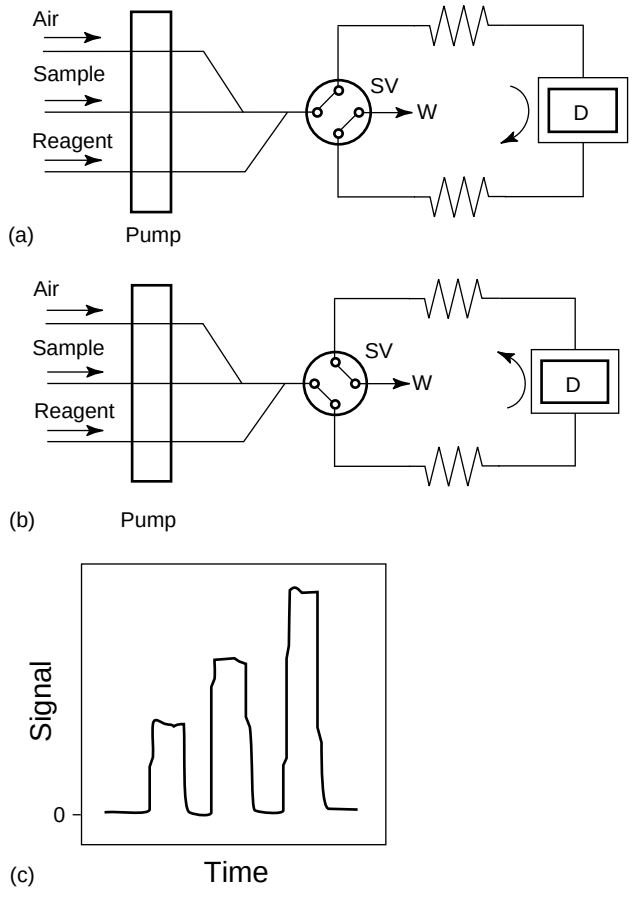

Figure 10 Diagram of an air-segmented flow reversal analyzer. SV, switching valve; D, detector; W, waste. (a) and (b) flow reversal by the switching valve; (c) typical plot of the detector signal. [Adapted by permission of The Royal Society of Chemistry from M. Silva, *Analyst*, **118**(6), 681–688 (1993).]

to zero. Compared with the normal flow, peaks obtained by flow reversal are narrower and more Gaussian, with lower dispersion. This is mainly because of the way the flow is circulated and the fact that the sample does not travel through the whole flow injection system. This type of analyzer is particularly well suited for very slow reactions. Flow reversal analyzers are used both in unsegmented flow and in air-segmented flow modes.[49]

2.3 Continuous Reagent Addition Method

In this method, the reagent with which the analyte reacts is added continuously, at a constant rate, to the solution of the analyte and other reaction ingredients.[55] The instrumentation is fairly simple and consists of a reaction vessel with an adequate stirring and an automatic burette. Using an immersion probe for photometric detection allows for smaller reaction vessels and lower sample consumption (Figure 11a).[50] The analyzers of this type are especially suited to fast reactions and offer a major alternative to the stopped-flow analyzers. The analyte concentration during this addition varies because of both reaction and dilution. The overall reaction is of second order.

A typical kinetic curve is given in Figure 11(b). There are two portions on this curve, which are of analytical interest. The first is the initial concave segment that conforms to Equation (2):

$$\Delta S = 0.5\varepsilon_p u V_0^{-1} k[\mathrm{R}]_0[\mathrm{A}]_0 t^2 \qquad (2)$$

where ΔS is the change in analytical signal, ε_p the molar absorbtivity of the product, V_0 the initial sample volume, k the second-order rate constant, $[\mathrm{R}]_0$ the reagent concentration in the automatic burette, $[\mathrm{A}]_0$ the initial concentration of the analyte and t time. Equation (2) shows that the analytical signal is directly proportional to t^2. The analyte concentration could be determined from this segment of the kinetic curve using the initial reaction method.

The second segment of interest on the kinetic curve is the linear intermediate portion that follows Equation (3):

$$\left(\frac{\mathrm{d}S}{\mathrm{d}t}\right)_m \approx 0.55\varepsilon_p[\mathrm{A}]_0 \left(\frac{k[\mathrm{R}]_0}{V_0}\right)^{1/2} \qquad (3)$$

where the subscript m signifies the maximum reaction.

In comparing the two possible approaches to the kinetic curve, the maximum reaction method yields a smaller error, for instrumental reasons, while the initial reaction method is less affected by side reactions since the measurements are made at the beginning of the reaction. However, the maximum reaction method is preferred since it is less sensitive to between-run variations arising from the rate constants and the measurements are made on a much larger linear portion.[56]

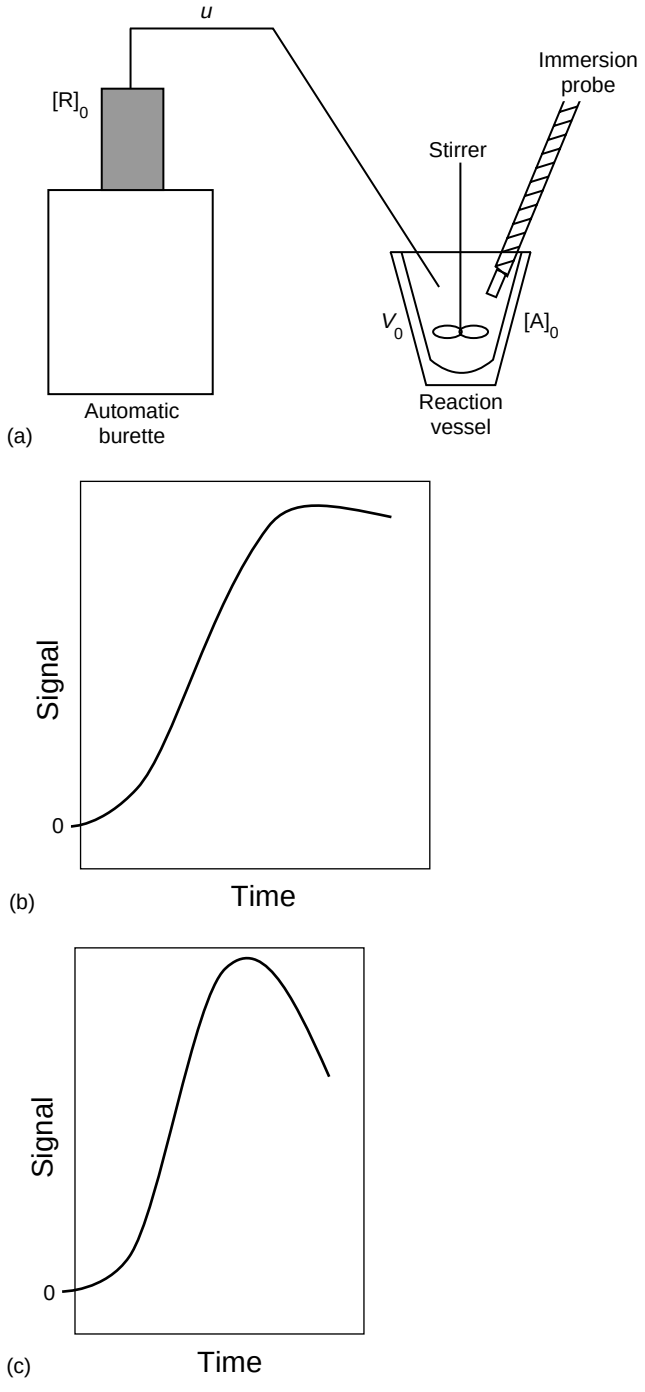

Figure 11 Schematic diagram of the instrumentation and kinetic curve for continuous reagent addition method. (a) Instrumentation used with photometric detection; (b) typical kinetic curve; (c) recording obtained for a chemical system involving an undesirable reaction. [Reproduced by permission of The Royal Society of Chemistry from M. Silva, *Analyst*, **118**(6), 681–688 (1993).]

The continuous reagent addition method was applied for single-element determinations[57–61] and on pharmaceuticals,[62–66] biological[67] and clinical[68] samples.

2.3.1 Continuous-flow Stirred Tank Reactor

The principle of the continuous flow of reagent could be used to apply the analyte pulse perturbation (APP) technique on oscillating chemical reactions and to use these reactions for analytical purposes. The most convenient instrumental set-up for APP is the continuous-flow stirred tank reactor (CSTR). The continuous-flow component of a CSTR is provided by a peristaltic pump, that pumps the required reagents into a thermostatted receiver. The reagent mixture is well stirred and its volume is kept constant by an overflow channel provided by the same peristaltic pump. The reaction is monitored in the receiver, usually by photometric or electrochemical detection. The CSTR could be considered a homogeneous system, where mass and energy are continuously exchanged with the surrounding environment and the reactions occur in far from thermodynamic equilibrium conditions.[69]

The APP technique is based on the change in amplitude or period of an oscillating reaction produced by a microvolume of an analyte solution added to the reactant mixture.[70] The reactions most frequently used are the Belousov–Zhabotinskii reaction, based on the oxidation of malonic acid by bromate ions in concentrated sulfuric acid, and copper oscillators based on the oxidation of KSCN by H_2O_2 in a strongly alkaline medium, which is catalyzed by traces of copper. The APP technique has been used on copper oscillators for the determination of vitamin B_6,[71] gallic acid and paracetamol,[72,73] resorcinol[74] and glutathione.[75]

3 REACTION MONITORING SYSTEMS

Any suitable means of monitoring the concentration of a chemical species can be used in reaction rate methods. Most frequently, spectrochemical and electrochemical methods are used for monitoring but a smaller number of methods based on other chemical or physical parameters have been described. In selecting an appropriate transducer, the nature of the monitored species, the type of reaction and its rate must be considered. The transducer should be sufficiently sensitive to provide detection of very small changes in measured reaction parameter and must have a response time which is short relative to the reaction half-time to prevent distortion of the resultant rate curve.

The kinetic information could be obtained using single- or multichannel detection systems. Single-channel detection systems produce a single response, proportional to the reactant or product concentration versus time. Multichannel or multidimensional detection systems provide simultaneous (or quasi-simultaneous) responses from the same chemical reaction. The multidimensional detection could be achieved by using a single type of detector or by combining more types of detectors.[2]

3.1 Single-channel Detection

3.1.1 Electrochemical Monitoring

Electrochemical-based devices are attractive mainly because of instrumental simplicity and relatively low cost. Most of the instruments are based on potentiometric and amperometric detection but detectors based on conductometry, chronoamperometry and voltammetry are also used. Since most kinetic determinations are accomplished by means of flow analyzers, the detectors for this type of instrumentation will be considered.

Potentiometric detection is based on monitoring the potential of an indicating electrode. The main indicating electrodes in kinetic determinations are metallic electrodes, carbon-based electrodes, ion-selective electrodes (ISEs) and enzyme electrodes. The reference electrode is usually an Ag/AgCl electrode.

Copper-wire electrodes were used for the detection of amines in a flow-injection system.[76] A flow cell containing eight Ag electrodes in the carrier channel and one Ag electrode in the reference channel was used for determination of Ag^+ and iodide ions.[77] The potential of a silver metal electrode was measured to detect the metal cyanide complex formation. Ag, Hg(II), Fe(III), Cu(II), Zn, Co(II) or Ni(II) were determined in a flow-injection analyzer with 0.29 mM cyanide solution as carrier.[78] Aluminum ions were detected by an Al-wire indicator electrode with an Ag-wire reference electrode in a carrier solution of 1 mM fluoride in 0.05 M acetate buffer of pH 5.[79]

ISEs and biosensors are known for many analytical species. The electrode response time varies according to both the type of electrode and the conditions under which is used. For detection in flow analyzers where the electrode volume should be small and its response time short, the ionophore is incorporated in a polymer and applied as a thin layer on a metallic wire (ion-selective coated-wire electrodes).[80–82]

A solid-state Br^- electrode in conjunction with a double-junction Ag/AgCl reference electrode was used for potentiometric monitoring of Br in the kinetic determination of pyridoxine.[83] A Br^--selective electrode is used for monitoring the potential oscillations in the Belousov–Zhabotinskii reaction in kinetic determination of hexacyanoferrates.[84] Fluorine-sensitive electrodes have been applied for a number of kinetic determinations. The determination of carbocysteine in colored pharmaceuticals based on monitoring of its reaction with 1-fluoro-2,4-dinitrobenzene, catalyzed by

hexamethyltrimethylammonium bromide micelles, was performed using a F^--selective electrode.[85] The same type of electrode was used for kinetic determination of ethylenediaminetetraacetic acid (EDTA) and citrate by the displacement of fluoride from aluminum(III) fluoride complexes.[86] A kinetic method based on a fluoride-selective electrode was applied in the determination of alkaline phosphatase in serum.[87] Kinetic determination of primary and secondary amines was accomplished based on their reaction with 1-fluoro-2,4-dinitrobenzene monitored by an F^--selective electrode.[88]

Copper was determined based on its catalytic effect on the reaction between hexachloroantimonate and hydroxylamine. The rate of the reaction was measured potentiometrically with an Sb(V) ISE using 1,2,4,6-tetraphenylpyridinium iodide as electroactive material in a poly(vinyl chloride) (PVC) membrane.[89] A Pb ISE was used for determination of citrate by monitoring the citrate in drinks and pharmaceutical products. The method is based on inhibition, by citrate, of crystalline growth of lead carbonate seed crystals.[90]

Amperometric detection involves monitoring the current between a working electrode and an auxiliary electrode at a given potential difference between the working and a reference electrode. The working electrode could be metal (Pt, Co, W), modified metal (metal hexacyanoferrate, Pt enzyme) or carbon-based (vitreous carbon, graphite, carbon paste, modified carbon paste). The auxiliary electrode is Pt, a carbon-based electrode or sometimes metallic parts of a cell. The reference electrode is Ag/AgCl or a saturated calomel electrode (SCE). The potential of the working electrode is a function of the species to be measured and of the type of electrode. Detection limits are in the range of 1 μM. Lower detection limits could be achieved with chemically modified electrodes. Detection cells could have a volume in the range of microliters.

Amperometric detection has been extensively used in conjunction with rotating bioreactors. The cell comprises a stationary Pt-ring electrode, a stainless-steel auxiliary electrode and an Ag/AgCl reference electrode.[91] It was used for determination of glucose, via the H_2O_2 produced enzymically, kinetic determination of fructose and ascorbate in food,[92] determination of choline esters in pharmaceuticals[93] and in comparative studies of bioreactor systems.[94] An amperometric oxygen sensor with a Pb anode, an Ag cathode and a polyethylene membrane was used for the determination of Co by measuring the rate of its reaction with dissolved oxygen.[95] For kinetic determination of As(III) the limiting diffusion current is measured, in a solution of KBr and KIO_4, with a polarograph with a Pt indicator electrode at 0.4 V and an SCE. The oxidation of KBr by KIO_4 is subject to an induction period, the length of which is directly proportional to the As(III) concentration.[96] Hg could be determined based on its catalytic effect on the photooxidation of EDTA by iodine. The reaction rate is measured amperometrically from the time required for the iodine reduction current to decay to zero at −100 mV vs SCE.[97] The determination of nitrate is based on the same photochemical reaction as above, but with an inhibitory effect. A platinum indicator electrode and an SCE is dipped into the EDTA and iodine solution and the current–time curve is recorded at a constant potential of 100 mV.[98]

Biamperometric detection is based on the measurement of current between two identical electrodes. The detector is simpler than in amperometric detection and consists of two opposing platinum electrodes in contact with the carrier stream. A potential difference of 10–500 mV is applied between the electrodes and the resulting current is measured. The current can flow only when both reduced and oxidized forms of a couple are present in the cell. The detection limits are in the range of 1 μM. Biamperometric detection was used in the kinetic determination of vitamin B_1 during their reaction with coulometrically generated iodine.[99]

For *conductometric detection*, a computer-controlled bipolar pulse conductivity system has been developed for use as a detector in reaction rate determinations. The system could perform discrete conductance measurements in 30 μs and could be used as detector in stopped-flow studies.[6] Kinetic monitoring based on conductometry was applied to the synthesis of *N*-α-benzyloxycarbonylalanylphenylalaninamide in the presence of *Pseudomonas aeruginosa* elastase. After the reaction was initiated the change in conductance was recorded allowing a precise measurement of the initial reaction rate.[100] A conductometer with facilities conductivity compensation was used to monitor the bromate–iodide–ascorbic acid Landolt reaction. The time interval between the initial rise in conductivity on mixing the solutions and the attainment of a minimum conductivity value was taken as the reaction time.[101]

The effect of lanthanide ions on the kinetics of glutamate dehydrogenase with NADH as coenzyme was monitored with a *chronoamperometric method*. The initial reaction rate, and hence the activity of glutamate dehydrogenase, was determined by measuring the decrease in the steady-state limiting current of NADH oxidation at a vitreous C electrode at ±0.8 V vs SCE.[102] *Chronopotentiometry* with a Br^--selective chalcogenide glass electrode was applied in the kinetic determination of vanadium in natural waters and wastewater.[103]

Polarographic-based detection was used for catalyzed reactions with the reaction step and detection step performed separately. The catalyzed reaction is quenched after a fixed time. The concentration of one of the

species involved in the reaction is determined by a form of polarography. This procedure was applied to the determination of ascorbic acid and Fe(III) using single-sweep oscillopolarography[104,105] and nitrate by differential-pulse polarography[106] with a dropping-mercury electrode as the working electrode and a Pt counter electrode.

3.1.2 Photometric Monitoring

Photometric monitoring makes up the bulk of detection systems used in reaction rate determination, as demonstrated in the comprehensive reviews by Mottola et al., dedicated to kinetic determinations, published biannually in *Analytical Chemistry*.[107–109]

Photometric systems for measuring ultraviolet and visible absorbed, fluorescent or chemiluminescence radiation are equipped almost exclusively with photomultiplier tubes or photodiodes (PDs). Both of these transducers have a response time in the nanosecond range, allowing very fast reactions to be followed. The type and model of transducer selected for a given type of chemical system and spectroscopic method should provide the highest signal-to-noise ratio. Photomultiplier tubes exhibit better characteristics than PDs when low light levels are measured (photocathodic current below 10^{-9} A) and are recommended for measurements requiring a narrow absorption bandwidth and for fluorescence, phosphorescence and chemiluminescence spectroscopy.[110] PDs are advantageous for the measurement of higher light intensities as in molecular absorption spectroscopy where the large bandwidth allows the use of filters or monochromators with larger bandwidth.[111]

Most detection systems are commercially available spectrophotometers, which could be used, for slow reaction, with no major modification. For spectrophotometers used as a detection system in flow analysis, a flow-through cell should be adapted to the cell compartment of the spectrophotometer.

A series of flow-through cells were used in connection with colorimeter heads using light-emitting diodes (LEDs) as light sources and PDs as detectors. Such a colorimeter head is very cheap, has a small volume and does not require very stabilized sources. The band gap in LEDs could be altered in the manufacture so that LEDs having source peak emission at different wavelengths in the visible region are currently available. The LED emission band-pass is less than 15 nm at half emission peak height, which is comparable to the band-pass of an interference filter. Such a colorimeter head was built in a double-beam version with an LED of appropriate wavelength and two PDs. One PD is mounted in the light path and the other perpendicular to the path. A beam splitter fabricated from a rod cut at one end at a 45° angle reflects about 50% of the light to the detecting PD while the remainder of the light strikes the reference PD.[112]

3.1.3 Fluorimetric and Chemiluminescence Monitoring

Fluorimetric and chemiluminescence monitoring implies the detection of relatively low light intensities. Most commercial spectrofluorimeters are provided with facilities to measure both fluorimetric and chemiluminiscence signals.

Fluorimetric detection is a multiparameter technique that involves an absorption wavelength and a luminescence wavelength. The bandwidth and the polarization of both light beams could be separately controlled. This, in combination with the control of other parameters, such as beam modulation, provides the fluorimetric methods with a high selectivity.[113]

Fluorimetric monitoring in reaction rate methods has been applied for both catalyzed and noncatalyzed reactions. Kinetic fluorimetric determination has been reported for a number of metallic ions, e.g. iron,[114] cerium,[115] aluminum,[116] tin,[117] nickel[118] and vanadium.[119] The combination of stopped-flow or flow injection in the stopped-flow mode and fluorimetric monitoring of reaction rate has been used for pharmaceuticals, drugs, pesticides and biological samples.[8,120–126] The potential of reaction rate methods in pesticide and drug determinations was demonstrated by using stopped-flow immunoassay. The negative effect of the sample matrix could be avoided by monitoring the initial rate of an antigen–antibody reaction. Using fluorescence polarization as the detection system significantly enhances the selectivity of the method.[127–129] In this case, a polarizer polarizes the excitation beam while a second polarizer is introduced in the luminescence beam. Light intensity measurements could be made parallel or perpendicular to the polarization of the exciting light.

Chemiluminescence detection in reaction rate measurements involves a highly efficient system of mixing and a fast light detector that are motivated by the transient nature of the light emitted by chemiluminescent systems. This is why this type of detection is used in conjunction with stopped-flow techniques or continuous reagent addition coupled with spectrofluorimeters. Chemiluminescence detection was used for the determination of nitrogen oxides in environmental samples, based on $NO-O_3$ chemiluminescence[130] and in studies of the effects of surfactants on the rate of peroxyoxalate chemiluminescence reaction between bis{*N*-[2-(*N*′-methyl-2′-pyridiniumyl)ethyl]-*N*-[(trifluoromethyl)sulfonyl]}oxamide, H_2O_2 and a fluorophore.[131] Also, in the determination of serum creatinine using creatininase, creatinase and sarcosine oxidase, the detection was based on the chemiluminometric rate assay of H_2O_2 produced.[132] The continuous addition reagent technique was tested

for the Cu(II)-catalyzed reaction between luminol and H_2O_2 using a fluorimeter in energy mode at 425 nm for chemiluminescent reaction-rate determinations.[61] A stopped-flow procedure combined with chemiluminescence detection was reported for multicomponent kinetic analysis of chemiluminescence enhancers (protocatechuic and caffeic acids).[133] The chemiluminiscence of the bis(2,4,6-trichlorophenyl) oxalate–H_2O_2 system was used for the kinetic determination of phenothiazine drugs. The method was applied to the determination of acepromazine in horse plasma.[60] Chemiluminescence detection was used for the determination of alkaloids (psychotropic indole derivatives) by the maximum reaction rate method. An application of psilocin determination in mushrooms has been reported.[134] A kinetic method for tertiary aliphatic amines with chemiluminescence detection is based on the reaction between tertiary aliphatic amines and sodium hypochlorite in alkaline medium and in the presence of fluorescein as sensitizer. The chemiluminescence intensity was monitored using a photomultiplier at a rate of one reading every 7.5 ms. Detection limits were 2.7–7.7 mol. The method was applied to determine trimethylamine in fish and to study the degradation of fish during storage.[67]

Photon counting, as a detection technique fitted for very low light intensities, has been used in fluorimetry for drug determinations. A stopped-flow reactor was used in conjunction with a photon-counting spectrofluorimeter, equipped with a xenon-arc source, for the determination of nordazepam in urine and imipramine in blood serum.[127]

3.1.4 Other Types of Detection

Thermometric detection could be used for catalytic reactions. After the start of reaction the initial temperature change is usually proportional to the catalyst concentration if the reaction is first order with respect to the catalyst. The temperature of the reacting mixture is monitored using a thermistor connected to a Wheatstone bridge and having a sensitivity of $0.01\,^{\circ}C\,mV^{-1}$. Thermometric detection requires simple instrumentation and avoids interference from colored compounds and precipitates. Reaction rate determination by thermometry has been applied to the determination of single elements[135–142] and mixtures of elements[143–145] based on their catalytic effect in redox reactions and for the determination of enzymes.[146]

Surface acoustic wave (SAW) sensors are mass detectors that can measure a variety of other physical properties at the sensor surface. Most SAW devices consist of a piezoelectric crystal with two sets of interdigital transducers laid down on the crystal surface. When a time-varying radiofrequency potential is applied to one set of interdigital transducers, an acoustic wave is launched in the surface of the piezoelectric crystal. This wave travels across the surface of the crystal and is received by the second set of transducers, which converts the mechanical energy of the wave back into an electrical signal. The acoustic wave generated is very sensitive to material present at the crystal surface. The acoustic device is converted into a chemical sensor by application of chemically selective layer on the resonant crystal. The resonant frequency of the SAW device would be altered if a chemical component interacts with the selective layer.[147]

The use of SAWs in analysis is increasing because of their selectivity and sensitivity. In recent years a series of applications of SAW as detectors have been published for kinetic determinations of alkaline phosphatase and acid phosphatase,[148,149] trypsin and antitrypsin in human serum,[150,151] pancreatic lipase,[152] urea in human blood,[153] arginase–urease,[154] organophosphorus nerve agents[155] and organophosphorus pesticides.[156]

3.2 Multichannel Detection

Multichannel detection is the basis for multidimensional instrumentation. It offers a great deal more information about time-dependent systems than could be obtained from monitoring one signal versus time. Multichannel detection is accomplished mainly as photometric detection based on photodetector arrays. Such an array consists of a number of small photosensitive detectors, called pixels, which are part of a single integrated circuit. The array could be linear or two-dimensional. Typically, a linear array could have from 64 to 4096 photosensitive elements on a photoactive surface 2.5–50 mm long and a few millimeters wide. Linear arrays with 8196 elements are also known. A two-dimensional array could have from 64×64 to 4096×4096 pixels. The pixel size ranges from 9 to 27 μm. The sensitive element could be a PD, charged-coupled device (CCD) or charge-injection device (CID). Most arrays have a spectral responsivity from 200 to 900 nm but devices that respond from 1 to 1000 nm are also available.[157]

By placing the photodetector array in the focal plane of a spectrometer, each element provides the information for a different wavelength and a full spectrum in absorption, fluorescence or chemiluminescence mode could be obtained in a matter of milliseconds or less. This allows for the monitoring of multiple species, observation of transient intermediates and correction for background absorbances. An example of a time–wavelength recording using diode arrays is given in Figure 12 for the iron-catalyzed reaction between leucomalachite green and peroxodisulfate.[158] The absorbance peak on the left indicates a catalytic effect produced by iron(II) and the higher peak in the middle a catalytic effect of both iron(II) and iron(III).

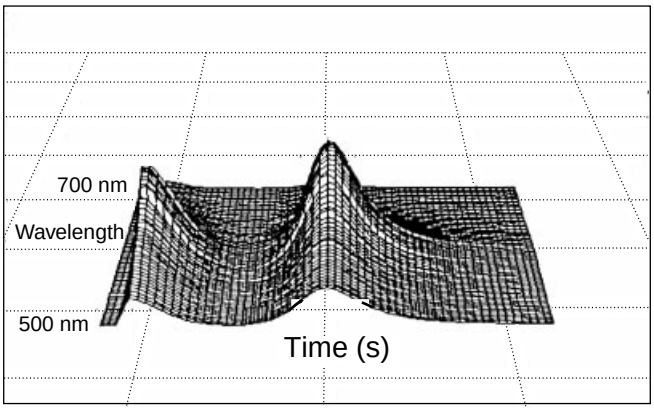

Figure 12 Examples of time–wavelength recording using PD arrays for the iron-catalyzed reaction between leucomalachite green and peroxodisulfate. [Reprinted from H. Muller, V. Muller, E.H. Hansen, *Anal. Chim. Acta*, **230**, 113–123 (1990), Copyright 1990, with permission from Elsevier Science.]

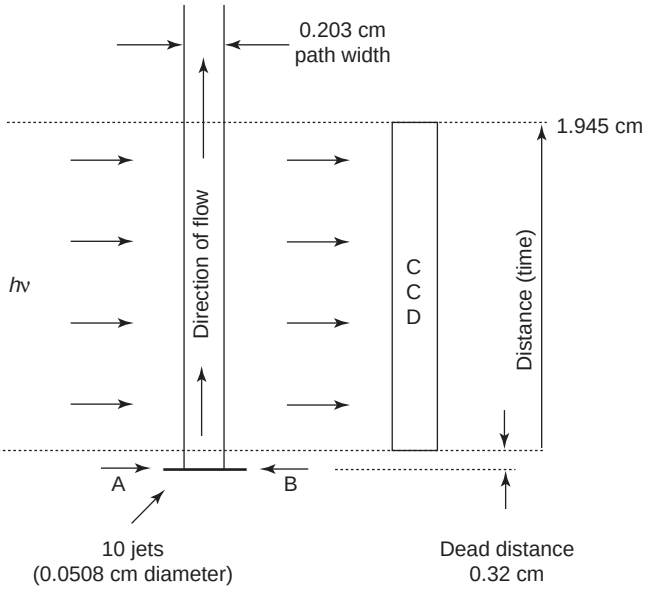

Figure 13 Position-resolved observation with photodetector array. A masked CCD array is placed parallel with the observation tube. The illumination permits observation perpendicular to the direction of solution flow. [Reprinted with permission from M.R. McDonald, T.X. Wang, M. Gazda. W.M. Scheper, S.W. Evetts, D.W. Margerum, *Anal. Chem.*, **69**(17), 3513–3520. Copyright 1997 American Chemical Society.]

Two types of spectra, absorption and fluorescence, could be also obtained simultaneously. A bifurcated fiber-optic light pipe and a spectrograph with an intensified array detector are used. The measurement circuit is operated in the time-sharing mode so that a single array collects both the absorbance and fluorescence signals.[159]

Spectrofluorimeters in a T format configuration could be used for the simultaneous determination of two species. In order to avoid spectral overlap the method requires that the reaction products of the analytes have similar excitation wavelengths and very different emission wavelengths.[160] The fluorescence intensity, provided by the same measurement cell, could be read on two channels perpendicularly on the excitation beam, at two different wavelengths. An appropriate set of filters and monochromators could be used to provide the excitation wavelength and the emission wavelengths. Since in most instruments there is only one emission monochromator, the wavelength of the second emission arm is set by a filter. This experimental set-up was used for simultaneous determination of nordazepam and imipramine in blood serum by fluoroimmunoassay,[127] for neomycin and promethazine in pharmaceutical preparations[160] and for ampicillin and tetracycline in milk, by stopped-flow fluorimetry.[161]

The photodetector arrays could be used for position-resolved observation. The detector is placed in a parallel position along the observation cell of a continuous-flow system. The data collected at different points along the cell by each element of the array correspond to different reaction times. This method of observation is exemplified in Figure 13 for a PAF spectrometer.[22]

The reagent and sample solutions are brought together in a 10-jet radial mixer and enter the observation cell with a flow velocity of $21\,m\,s^{-1}$. A decelerating push is given and the flow velocity decreases to $2\,m\,s^{-1}$. During the push, the reactions are observed perpendicular to the direction of the flow by a masked CCD array detector which provides, at the end of the push, a complete set of three-dimensional data, in 128 points, on position, velocity and transmittance.

3.3 Combined Detection

The rate measurements are more susceptible to variations in experimental conditions than equilibrium-based measurements. Combined detection could be used to minimize the effect of different parameters on the precision of rate measurements. Instead of maintaining the experimental conditions perfectly constant, the parameters that influence the precision could be measured and then a normalized rate calculated. An instrumental set-up for monitoring fluorescence signal, temperature and pH is given in Figure 14.[162] The fluorescence signal is simultaneously monitored with temperature and pH, the main experimental factors that could influence the precision of reaction rate measurements.

By means of dedicated software, the measured rate is normalized to a value that would be obtained at predetermined temperature and pH values. By applying the software correction the precision of rate measurement improves with a factor of four for temperature variations

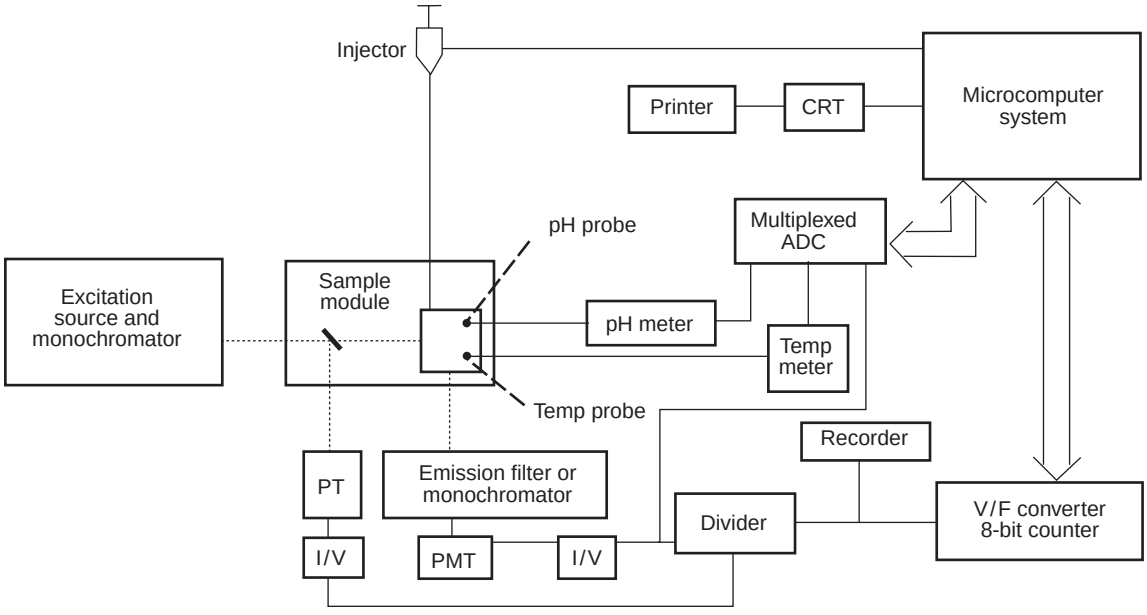

Figure 14 Block diagram of instrumentation for monitoring and compensating for between-run temperature and pH variations in kinetic-based determinations. [Reprinted from G.L. Campi, J.D. Ingle, Jr, *Anal. Chim. Acta*, **224**, 275–287 (1989), Copyright 1989, with permission from Elsevier Science.]

higher than 1 °C and more than one order of magnitude for a pH variation greater than 0.28.

Specialized spectroelectrochemical detectors for flow injection were built so that absorption spectra could be recorded simultaneously with potentiometric measurements. In one of these systems,[163] the cells have a rectangular flow channel of 15 × 5 mm with a reticulated vitreous-carbon working electrode, an Ag/AgCl reference electrode and an open optical window. The dead volume of the cells is 80 µL. Absorption spectra were recorded with a 2-nm slit width and a 2-s response time.

4 DATA HANDLING

The final step of a kinetic method of analysis is the determination of the analyte concentration based on the reaction rate measurement. In turn, the reaction rate is obtained by processing the signal provided by the reaction monitoring system.

In the first stages of the data processing the slope of the signal (or the slope of the tangent to the signal) should be determined. In some analytical systems the rate curve is recorded and the slope is determined, by various means, from this recording. However, for many routine analyses, the recording of a rate curve could be a tedious or an unnecessary operation. Therefore, simpler ways to extract the kinetic information were devised, so that one or a few data from the detector are sufficient for a point on the calibration curve.

Table 1 Classification of kinetic methods of analysis based on data processing procedures. [Reprinted from H.L. Pardue, *Anal. Chim. Acta*, **16**(1–2), 69–107 (1989), Copyright 1989, with permission from Elsevier Science]

I. Direct computation methods	II. Curve-fitting methods
1. Integral methods	1. Integral methods
A. Fixed-time	A. Linear responses
a. One-point	B. Nonlinear responses
b. Two-point	a. Direct computation
c. Multipoint	b. Predictive (extrapolation)
B. Variable-time	2. Rate Methods
a. One-point	
b. Two-point	
2. Rate methods	
A. Initial-rate	
B. Intermediate-rate	
3. Integrated-signal methods	

Much effort has been devoted to developing more efficient procedures for reaction rate measurement, starting with analog systems and continuing with computer-based methods. In an exhaustive review, Pardue[164] classified, in a broadly applicable scheme, the options available in processing the reaction rate data. A simplified version of this classification is presented in Table 1. For single-component methods the classification scheme has two main categories: direct computation methods and curve-fitting methods.

The methods in the first category include those that could be implemented by manual data manipulation and automation with analog electronic devices. The data that represent the analytical signal are used to compute the concentration directly. The second category was developed after the advent and the wide availability of low-cost computers in analytical chemistry and involves curve-fitting procedures.

4.1 Direct Computation Methods

4.1.1 Integral Methods

The integral methods subcategory includes some basic methods that were applied from the early stages of kinetic determination development: fixed-time and variable-time methods.[9,165–167] A schematic representation of the principle of these two methods is given in Figure 15(a) and (b), respectively.

4.1.2 Fixed-time Method

In the fixed-time method, the change in the reaction-monitor signal is measured over a predetermined time interval, Δt. The method could be implemented in making the signal measurement at one point, two points or multiple points. The fixed-time method has two main attractive features. The first is that the measured signal varies in direct proportion to the analyte concentration. If the signal is a linear function of concentration, such as absorbance in spectrophotometric detection, the signal change over Δt varies linearly with concentration. This makes the data much easier to process and the concentration could be computed using just a multiplication factor. The second attractive feature is related to the fixed interval of time used for measurement. It is more convenient to automate the method if a fixed time sequence is used. In manual processing of a large number of samples, a fixed-time method facilitates the sample processing sequence, so that the signal could also be measured in a sequence.

However, in applying the fixed-time method, the relationship between signal and concentration should be known exactly. Anything that affects this relationship should be taken into consideration for the full range of analyte concentration implied in the determination. The method is less attractive in the case of detectors that do not respond linearly with concentration.

In the one-point approach of the fixed-time method, the time interval for measuring the signal is considered from the origin of the rate curve. It is the simplest form of data processing, since a single measurement is made, but it has the disadvantages of not compensating for a fixed background or for curve nonlinearity. In the two-point mode, two signal measurements are made with a Δt time interval.

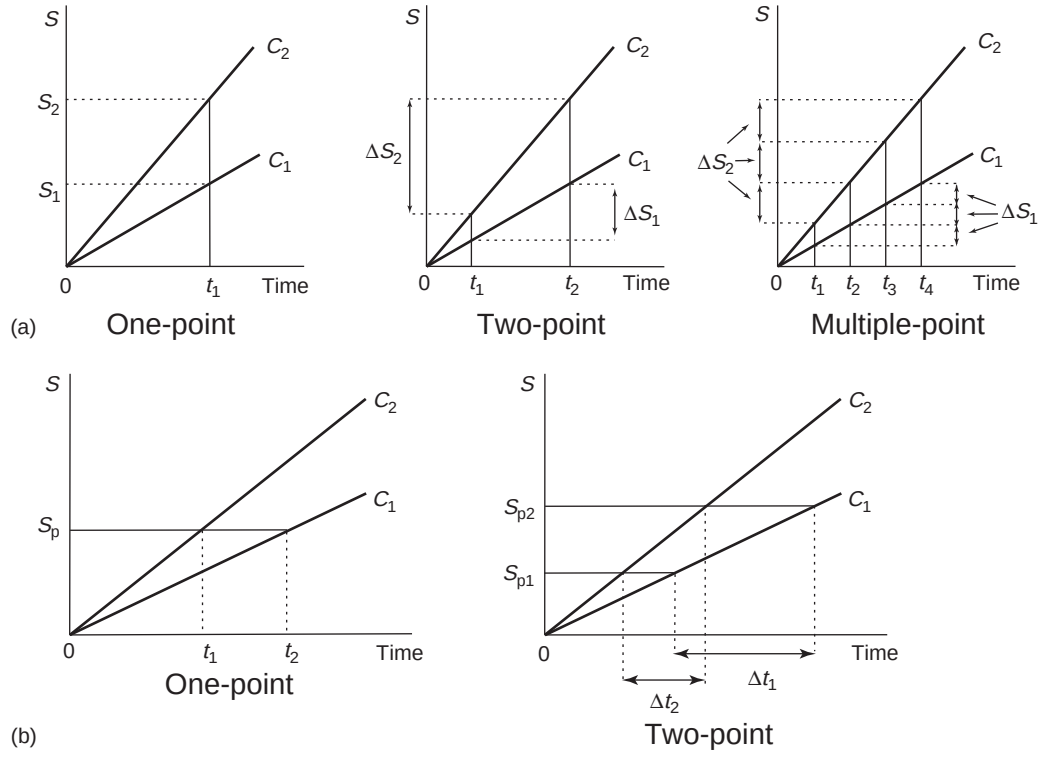

Figure 15 (a) Fixed-time and (b) variable-time methods exemplified on a pseudo-first-order reaction.

The signal difference, ΔS, for the two measurement points is used to compute the concentration. In comparison with the one-point approach, the two-point mode could compensate for the fixed background. However, it cannot detect the curve nonlinearity. The multiple-point approach implies the measurement of the signal difference between multiple points in time. The main merit of this approach is that it can detect nonlinear responses.

4.1.3 Variable-time Method

The variable-time method involves the measurement of the time required by the reaction-monitor signal to change by a predetermined amount. The reciprocal of the measured time interval is proportional to the concentration of the sought-for species. Unlike the fixed-time methods, here is not necessary to know the mathematical relationship between concentration and detector signal and the nonlinearity of the detector response is of little importance. The detection system should be sensitive enough to detect the predetermined reference point(s).

Among the disadvantages of the method is the fact that the measured parameter, the time interval, varies inversely with the analyte concentration following a relationship that is relatively complex except for zero-order reactions. Since the time for a reaction to reach a predetermined signal level is not predictable, it is less convenient to use the variable-time method for manually processing a large number of samples or for automation. In some cases the concentration of the sought-for species could be too small and the detector signal could not reach the predetermined level. This sets a limit for the lower concentration that could be determined in a convenient time span. Also, the systems using the variable-time method are more susceptible to random spikes of the signal that may cause false triggering of the measurement system.

The one- and two-point approaches could be applied for the variable-time method. Their relative merits and shortcomings are the same as for the fixed-time method.

4.1.4 Rate Methods

This category includes the methods in which the rate is determined instantaneously direct from the detector signal.[168] They could be considered as *derivative methods* since the electronic device used for rate determination is a derivative circuit. The circuit is very simple and provides the instantaneous value of the tangent at one point of the rate curve. Paradoxically, this is a limiting factor because at the required sensitivity for rate measurement the rate curve is accompanied by noise. The instantaneous slope of the noise is usually much greater than that of the signal and the final result is an amplification of the noise signal.[9]

4.1.5 Integrated-signal Methods

Integrated-signal methods use the integral of a portion of the rate curve. In the only method from this group the detector signal is integrated on two equal segments of the rate curve.[169] The integrated curve portions could be adjacent or separated by a certain time interval. For zero-order processes the difference between the integrated signals is proportional to reaction rate. The principal advantage of this approach is very low noise susceptibility. This particular method could be considered a form of fixed-time two-point method in which the noise is averaged by integration.

Neither rate methods nor integrated-signal methods have been used extensively in routine applications. One of the reasons is that their implementation requires a specialized device, a ratemeter. Also, integral methods, fixed-time or variable-time, were more convenient for routine analysis since they could be applied by simply reading the level of the detector signal or measuring an interval of time. An integrated-signal method that offers greater noise immunity and reliable results was developed at about the same time when the use of computers was expanding in analytical instrumentation. The principle of the method could be implemented much more easily by software so there was no reason to develop specialized instrumentation.

4.2 Curve-fitting Methods

The methods in this category are based on in-line or off-line data processing by a computer. Unlike direct computation methods, where the reaction rate is obtained by using a closed-form algorithm applied to a few experimental points, here a mathematical model is computed that represents the ''best fit'' of data obtained from the detector. The values of the parameters required by the kinetic determination are obtained from the equation that resulted from the fit. The curve-fitting procedure could be applied for either linear or nonlinear responses.

Linear responses are the commonest applications since they correspond to zero-order and pseudo-zero-order processes that are most frequent in kinetic determinations. A least-squares algorithm or a Kalman filter could be used.[170,171] Several well-known software packages such as ORIGIN® (Microcal) and EXCEL® (Microsoft) include a linear regression fitting option. In addition to obtaining the slope of the rate curve, which is the primary objective of the method, some other statistical parameters could be computed: intercept linearity, correlation coefficients and errors.[172]

The curve-fitting for *nonlinear responses* could be used to compute directly the concentration from the slope. The slope could be determined either at the reaction

beginning, where it approaches zero-order behavior, or at another point on the rate curve. In the former case the data manipulation is reduced to that of a linear response. In the latter case the concentration could be computed from the slope if some of the experimental parameters are kept constant. For a first-order reaction this means that the slope is always measured at the same point in time and the rate constant does not change during the reaction.

The curve-fitting procedure could be used in a predictive manner. The kinetic curve based on data collected in one stage of the reaction could be extrapolated to another time range, without the need to run the reaction to completion. It is possible to predict the value of absorbance corresponding to the steady state in a spectrophotometrically monitored reaction or the steady-state velocity from data collected before the steady state was achieved.[173–175]

Data processing by computer offers the possibility to apply error-compensating methods to overcome the sensitivity of kinetic methods to conditions that influence the rate constant and to implement multicomponent kinetic methods.[176] Details of these features of data processing step of kinetic methods are given in the article **Data Treatment and Error Analysis in Kinetics**.

ABBREVIATIONS AND ACRONYMS

APP	Analyte Pulse Perturbation
ASCFA	Air-segmented Continuous-flow Analysis
CCD	Charged-coupled Device
CID	Charge-injection Device
CSTR	Continuous-flow Stirred Tank Reactor
EDTA	Ethylenediaminetetraacetic Acid
FIA	Flow-injection Analysis
ISE	Ion-selective Electrode
LED	Light-emitting Diode
PAF	Pulse-accelerated flow
PD	Photodiode
PVC	Poly(vinyl chloride)
SAW	Surface Acoustic Wave
SCE	Saturated Calomel Electrode
SI	Sequential Injection

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Biomolecules Analysis **(Volume 1)**
Fluorescence-based Biosensors

Carbohydrate Analysis **(Volume 1)**
Monosaccharides and Sugar Alcohol Analysis

Chemical Weapons Chemicals Analysis **(Volume 1)**
Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Automation in the Clinical Laboratory • Drugs of Abuse, Analysis of • Electroanalysis and Biosensors in Clinical Chemistry • Electroanalytical Chemistry in Clinical Analysis • Electrolytes, Blood Gases, and Blood pH • Glucose Measurement • Immunochemistry • Laboratory Instruments in Clinical Chemistry, Principles of • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Flow-injection Techniques in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Luminescence in Environmental Analysis

Food **(Volume 5)**
Enzyme Analysis and Bioassays in Food Analysis • Fluorescence Spectroscopy in Food Analysis • Pesticides, Mycotoxins and Residues Analysis in Food • Vitamins Analysis in Food

Forensic Science **(Volume 5)**
Immunoassays in Forensic Toxicology

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Pesticides **(Volume 7)**
Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Herbicide Residues in Biota, Analysis of • Immunochemical Assays in Pesticide Analysis • Organochlorine, Pyrethrin and Pyrethroid Insecticides: Single Class, Multiresidue Analysis of • Organophosphorus Pesticides in Water and Food, Analysis of • Pesticides in Water: Sampling, Sample Preparation, Preservation • Pesticides (New Generation) and Related Compounds, Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Alkaloids, Pharmaceutical Analysis of • Antibiotics, Pharmaceutical Analysis of • Vitamins: Fat and Water Soluble, Analysis of

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Signal Processing in Analytical Chemistry

Electroanalytical Methods **(Volume 11)**
Chemiluminescence, Electrogenerated • Ion-selective Electrodes: Fundamentals

Electronic Absorption and Luminescence **(Volume 12)**
Phosphorescence Measurements, Applications of

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Catalytic Kinetic Determinations: Nonenzymatic • Data Treatment and Error Analysis in Kinetics • Differential Rate Determinations • Electrocatalysis-based Kinetics Determinations • Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Luminescence-based Kinetic Determinations • Uncatalyzed Kinetic Determinations

General Articles **(Volume 15)**
Karl Fischer Moisture Determination

REFERENCES

1. H.V. Malmstadt, E.A. Cordos, C.J. Delaney, 'Automated Reaction Rate Methods of Analysis', *Anal. Chem.*, **44**(12), 26A–41A (1972).
2. S.R. Crouch, 'Trends in Kinetic Methods of Analysis', *Anal. Chim. Acta*, **283**, 453–470 (1993).
3. S.R. Crouch, 'Kinetic Methods of Analysis: How do They Rate?', *J. Chin. Chem. Soc.*, **41**(3), 221–229 (1994).
4. A. Gómez-Hens, D. Pérez-Bendito, 'The Stopped-flow Technique in Analytical Chemistry', *Anal. Chim. Acta*, **242**, 147–177 (1991).
5. P. Bugnon, G. Laurenczy, Y. Ducommun, P.Y. Sauvageat, A.E. Merbach, R. Ith, R. Tschanz, M. Doludda, R. Bergbauer, E. Grell, 'High-pressure Stopped-flow Spectrometer for Kinetic Studies of Fast Reactions by Absorbance and Fluorescence Detection', *Anal. Chem.*, **68**(17), 3045–3049 (1996).
6. K.J. Caserta, F.J. Holler, S.R. Crouch, C.G. Enke, 'Computer Controlled Bipolar Pulse Conductivity System for Applications in Chemical Rate Determinations', *Anal. Chem.*, **50**(11), 1534–1541 (1978).
7. J. Wang, H.D. Dewald, 'Rapid Stopped-flow Voltammetry with Potential Scanning', *Anal. Chim. Acta*, **136**, 7–84 (1982).
8. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Long-wavelength Stopped-flow Fluorimetric Determination of Lysozyme', *Talanta*, **43**(9), 1413–1419 (1996).
9. H.V. Malmstadt, C.J. Delaney, E.A. Cordos, 'Reaction-rate Methods of Chemical Analysis', *CRC Crit. Rev. Anal. Chem.*, **2**(1), 559–619 (1972).
10. J.F. Evans, E.E. Bancroft, H.N. Blount, 'Stopped-flow Syringes with Teflon™ [PTFE] Plunger Tips and Smooth-bore Glass Barrels', *Anal. Chem.*, **53**(2), 359–360 (1981).
11. F.J. Holler, C.G. Enke, S.R. Crouch, 'Critical Study of Temperature Effects in Stopped-flow Mixing Systems', *Anal. Chim. Acta*, **117**, 99–113 (1980).
12. Q.H. Gibson, L. Milnes, 'Apparatus for Rapid and Sensitive Spectrophotometry', *Biochem. J.*, **91**(1), 161–171 (1964).
13. S.R. Crouch, F.J. Holler, P.K. Notz, P.M. Beckwith, 'Automated Fast Analysis with Stopped-flow Spectrophotometry', *Appl. Spectrosc. Rev.*, **13**, 165–259 (1977).
14. P.M. Beckwith, S.R. Crouch, 'Automated Stopped-flow Spectrophotometer with Digital Sequencing for Millisecond Analyses', *Anal. Chem.*, **44**(2), 221–227 (1972).
15. M. Anson, P.M. Bayley, 'Measurement of Circular Dichroism at Millisecond Time Resolution: Stopped-flow Circular-dichroism System', *J. Phys. E*, **7**(6), 481–486 (1974).
16. N. Papadakis, R.B. Coolen, J.L. Dye, 'Variable-temperature Computerized Dual-beam, Rapid-scanning Stopped Flow Apparatus for the Study of Air-sensitive Systems and Transients in Enzyme Reactions', *Anal. Chem.*, **47**(9), 1644–1649 (1975).
17. F.J. Holler, W.C. Mateyka, 'Modular Chemically Inert Mixer for Flow and Stopped-flow Experiments', *Anal. Chem.*, **52**(2), 354–355 (1980).
18. S. Lever, S. Crooks, 'Study of Fast Chemical Reactions Using Stopped-flow Photometry and the Preparative Quench Technique', *Int. Lab.*, **15**(8), 24–34 (1985).
19. R.Q. Thompson, S.R. Crouch, 'Stopped-flow Instrument for Reaction-rate Measurements with Immobilized Enzymes', *Anal. Chim. Acta*, **144**, 155–163 (1982).
20. R.A. Harvey, W.O. Borcherdt, 'Fluid-delivery System for a Variable-ratio Stopped-flow Mixing Device', *Anal. Chem.*, **44**(11), 1926–1928 (1972).
21. M. Goez, 'Inexpensive Pre-mixer for Stopped-flow Apparatus', *J. Phys. E*, **21**(5), 440–442 (1988).
22. M.R. McDonald, T.X. Wang, M. Gazda, W.M. Scheper, S.W. Evetts, D.W. Margerum, 'Pulse-accelerated-flow Spectrometer with Position Resolved Observation', *Anal. Chem.*, **69**(17), 3513–3520 (1997).
23. W.B. Furman, *Continuous Flow Analysis: Theory and Practice*, Marcel Dekker, New York, 1976.

24. J. Ruzicka, E.D. Hansen, 'Flow Injection Analysis. Part X. Theory, Techniques and Trends', *Anal. Chim. Acta*, **99**, 37–36 (1978).
25. J. Ruzicka, E.D. Hansen, *Flow Injection Analysis*, 2nd edition, Wiley, New York, 1988.
26. Y. Hsieh, S.R. Crouch, 'Air-segmented Flow Injection: A Hybrid Technique for Automated, Low Dispersion Determinations', *Anal. Chim. Acta*, **303**(2–3), 231–239 (1995).
27. C.J. Patton, M. Rabb, S.R. Crouch, 'Electronic Bubble Gate for Colorimetric, Air-segmented, Continuous-flow Analyzers', *Anal. Chem.*, **54**(7), 1113–1118 (1982).
28. D.P. Arnold, R.M. Peachey, J.D. Petty, D.R. Sweatman, 'Automated Titrations Using a Discontinuous Programmed-flow Analyzer', *Anal. Chem.*, **61**(18), 2109–2116 (1989).
29. J. Ruzicka, G.D. Marshall, G.D. Christian, 'Variable Flow Rates and a Sinusoidal Flow Pump for Flow-injection Analysis', *Anal. Chem.*, **62**(17), 1861–1866 (1990).
30. C.C. Painton, H.A. Mottola, 'Chemical Kinetic Contributions to Practical Dispersion in Flow Injection Analysis', *Anal. Chem.*, **53**(11), 1713–1715 (1981).
31. D.M. Spence, S.R. Crouch, 'An Investigation of Internal Pressures in Capillary Flow-injection Systems', *Anal. Chim. Acta*, **358**(1), 95–101 (1998).
32. D.M. Spence, S.R. Crouch, 'Factors Affecting Zone Variance in a Capillary Flow-injection System', *Anal. Chem.*, **69**(2), 165–169 (1997).
33. D.M. Spence, A.M. Sekelsky, S.R. Crouch, 'Design Considerations for Capillary Flow-injection Systems', *Instrum. Sci. Technol.*, **24**(2), 103–113 (1996).
34. S. Liu, P.K. Dasgupta, 'Flow-injection Analysis in the Capillary Format Using Electro-osmotic Pumping', *Anal. Chim. Acta*, **268**(1), 1–6 (1992).
35. S. Liu, P.K. Dasgupta, 'Electro-osmotically Pumped Capillary Flow-injection Analysis. Valve-based Injection Systems and Sample Throughput', *Anal. Chim. Acta*, **283**(2), 739–745 (1993).
36. A.M. Almuaibed, A. Townshend, 'Indirect Flow-injection–Spectrophotometric Determination of Molybdenum(VI) Using an Online Cadmium Reductor', *Microchem. J.*, **49**(1), 56–60 (1994).
37. K.S. Kurtz, S.R. Crouch, 'Design and Optimization of a Flow-injection System for Enzymic Determination of Galactose', *Anal. Chim. Acta*, **254**(1–2), 201–208 (1991).
38. A.M. Almuaibed, A. Townshend, 'Flow-injection Amperometric and Chemiluminescence Individual and Simultaneous Determination of Lysine and Glucose with Immobilized Lysine Oxidase and Glucose Oxidase', *Anal. Chim. Acta*, **338**(1–2), 149–154 (1997).
39. A.M. Almuaibed, A. Townshend, 'Flow Procedure for Ion-exchange Preconcentration and Online Spectrophotometric – A Determination of Iron(III) as its Thiocyanate Complex', *Fresenius' J. Anal. Chem.*, **352**(6), 601–602 (1995).
40. J. Ruzicka, A. Arndal, 'Sorbent Extraction in Flow-injection Analysis and its Application to Enhancement of Atomic Spectrometry', *Anal. Chim. Acta*, **216**(1–2), 243–255 (1989).
41. J. Ruzicka, T. Guebeli, 'Principles of Stopped-flow Sequential Injection Analysis and its Application to the Kinetic Determination of Traces of a Proteolytic Enzyme', *Anal. Chem.*, **63**(17), 1680–1685 (1991).
42. J. Ruzicka, 'The Second Coming of Flow Injection Analysis', *Anal. Chim. Acta*, **261**(1–2), 3–10 (1992).
43. Y.S. Hsieh, S.R. Crouch, 'Stopped-flow, Air-segmented Continuous Flow: Kinetic Determinations of Glucose and Phosphate in Wine and Serum Samples', *Anal. Chim. Acta*, **284**(1), 159–166 (1993).
44. G.D. Christian, J. Ruzicka, 'Exploiting Stopped-flow Injection Methods for Quantitative Chemical Analysis', *Anal. Chim. Acta*, **261**(1–2), 11–21 (1992).
45. J. Ruzicka, G.D. Marshall, 'Sequential Injection: A New Concept for Chemical Sensors, Process Analysis and Laboratory Assays', *Anal. Chim. Acta*, **237**(2), 329–343 (1990).
46. S.M. Sultan, F.E.O. Suliman, 'Use of a Sequential-injection Technique for Mechanistic Studies and Kinetic Determination of Bromazepam Complexed with Iron(II) in Hydrochloric Acid', *Analyst*, **121**(5), 617–621 (1996).
47. A. Rios, M.D. Luque-de-Castro, M. Valcarcel, 'Multidetection in Unsegmented Flow Systems with a Single Detector', *Anal. Chem.*, **57**(9), 1803–1809 (1985).
48. E.B. Townsend, S.R. Crouch, 'Computer Considerations for a Flow Reversal/Flow Recycle Continuous Flow Analysis System', *Trends Anal. Chem.*, **11**(3), 90–96 (1992).
49. Y.S. Hsieh, S.R. Crouch, 'Flow Reversal and Flow Recycle Air-segmented Flow Injection for Simultaneous Determination of a Binary Mixture', *Anal. Chim. Acta*, **304**(3), 333–337 (1995).
50. M. Silva, 'Recent Strategies in Automated Reaction Rate Based Measurements', *Analyst*, **118**(6), 681–688 (1993).
51. S.R. Crouch, 'Kinetic Methods for Intelligent Automation', *Chemom. Intell. Lab. Syst.*, **8**(3), 259–273 (1990).
52. A. Rios, M.D. Luque-de-Castro, M. Valcarcel, 'Analytical Potential of Flow-reversal Injection Analysis', *Anal. Chem.*, **60**(15), 1540–1545 (1988).
53. M. Romero-Saldana, A. Rios, M. Valcarcel, 'Continuous Sample Monitoring by Flow-reversal Methodology', *Fresenius' J. Anal. Chem.*, **342**(7), 547–551 (1992).
54. D. Betteridge, P.B. Oates, A.P. Wade, 'Control of Dispersion and Variation of Reaction Coil Length in Flow-injection Analyzers by Flow Reversals', *Anal. Chem.*, **59**(8), 1236–1238 (1987).
55. M. Marquez, M. Silva, D. Pérez-Bendito, 'Direct Rate Measurements on Millisecond Reactions by Continuous Addition of Reagent for Routine Analysis', *Analyst*, **113**(11), 1733–1736 (1988).

56. A. Velasco, M. Silva, D. Pérez-Bendito, 'Processing Analytical Data Obtained from Second-order Reactions by Using Continuous Reagent Addition', *Anal. Chem.*, **64**(20), 2359–2365 (1992).
57. M. Marquez, M. Silva, D. Pérez-Bendito, 'Continuous-addition-of-reagent Technique in Catalytic Kinetic-based Determinations', *Anal. Lett.*, **23**(7), 1357–1369 (1990).
58. M. Marquez, M. Silva, D. Pérez-Bendito, 'Continuous Addition of Reagent Technique: A New Approach to Differential Reaction-rate Methods', *Anal. Chim. Acta*, **239**(2), 221–227 (1990).
59. M. Marquez, M. Silva, D. Pérez-Bendito, 'Automated Kinetic Determination of Mercury(II) and its Application to Urine Samples', *Microchem. J.*, **43**, 229–236 (1991).
60. J. Cepas, M. Silva, D. Pérez-Bendito, 'Improved Peroxyoxalate Chemiluminescence-based Determinations by use of Continuous Reagent Addition to Remove Background Emission', *Anal. Chem.*, **66**(22), 4079–4084 (1994).
61. J. Cepas, M. Silva, D. Pérez-Bendito, 'Potential of the Continuous Addition of Reagent Technique for Chemiluminescent Reaction-rate Determinations', *Anal. Chim. Acta*, **285**(3), 301–308 (1994).
62. M. Marquez, M. Silva, D. Pérez-Bendito, 'Kinetic Determination of Vitamin B12 in Pharmaceuticals by the Continuous Addition of Reagent Technique', *J. Pharm. Biomed. Anal.*, **8**(7), 563–567 (1990).
63. M. Marquez, M. Silva, D. Pérez-Bendito, 'Kinetic Determination of Sulfonamides at the Millimolar Level by the Continuous Addition of Reagent Technique', *Anal. Chim. Acta*, **237**(2), 353–359 (1990).
64. D. Pérez-Bendito, A. Gómez-Hens, M. Silva, 'Advances in Drug Analysis by Kinetic Methods', *J. Pharm. Biomed. Anal.*, **14**(8–10), 917–930 (1996).
65. L. Ayllon, M. Silva, D. Pérez-Bendito, 'Improved Automatic Kinetic Method for the Determination of Various Corticosteroids', *J. Pharm. Sci.*, **83**(8), 1135–1141 (1994).
66. M. Carmona, M. Silva, D. Pérez-Bendito, 'Automatic Kinetic Determination of Oxazepam by the Continuous Addition of Reagent Technique', *Talanta*, **39**(9), 1175–1180 (1992).
67. M. Cobo, M. Silva, D. Pérez-Bendito, 'Chemiluminescence Measurement of Tertiary Amines Using the Continuous-addition-of-reagent Technique. Determination of Trimethylamine in Fish Samples', *Anal. Chim. Acta*, **356**(1), 51–59 (1997).
68. M. Marquez, M. Silva, D. Pérez-Bendito, 'Performance of the Continuous Addition of Reagent Technique in Fluorimetric Reaction-rate Methods. Determination of Thiamine at the Nanomolar Level', *Anal. Lett.*, **22**(11–12), 2485–2500 (1989).
69. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Approaching the Use of Oscillating Reactions for Analytical Monitoring', *Analyst*, **123**(2), 1R–8R (1998).
70. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Analyte Pulse Perturbation Technique: A Tool for Analytical Determinations in Far-from-equilibrium Dynamic Systems', *Anal. Chem.*, **67**(4), 729–734 (1995).
71. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Analytical Assessment of the Oscillating Chemical Reactions by Use of Chemiluminescence Detection', *Talanta*, **44**(8), 1463–1472 (1997).
72. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Application of Oscillating-reaction-based Determinations to the Analysis of Real Samples', *Analyst*, **122**(3), 287–292 (1997).
73. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Determination of Gallic Acid by an Oscillating Chemical Reaction Using the Analyte Pulse Perturbation Technique', *Anal. Chim. Acta*, **321**(1), 53–60 (1996).
74. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Simultaneous Determination of Gallic Acid and Resorcinol Based on an Oscillating Chemical Reaction by the Analyte Pulse Perturbation Technique', *Anal. Chim. Acta*, **334**(3), 323–330 (1996).
75. R. Jimenez-Prieto, M. Silva, D. Pérez-Bendito, 'Determination of Trace Amounts of Reduced Glutathione by a Chemical Oscillating Reaction', *Analyst*, **121**(4), 563–566 (1996).
76. Z-L. Chen, P.W. Alexander, 'Potentiometric Detection of Aliphatic Amines by Flow-injection Analysis and Ion-interaction Chromatography with a Metallic Copper Electrode', *J. Chromatogr. A*, **758**(2), 227–233 (1997).
77. P.W. Alexander, T. Dimitrakopoulos, D.B. Hibbert, 'Operational Assessment of a Potentiometric Eight-sensor Flow Cell in a Portable Flow-injection Analyzer', *Field Anal. Chem. Technol.*, **1**(1), 31–37 (1996).
78. D.P. Root, D. Wijesuriya, P.W. Alexander, 'Steady-state and Flow-injection Response of a Metallic Silver Electrode for Potentiometric Detection of Metal Cyanide Complex Formation', *Anal. Chim. Acta*, **225**(1), 105–113 (1989).
79. J.C. Nglia, P.W. Alexander, D.B. Hibbert, 'Determination of Aluminium Ions by Direct Potentiometry in a Flow System', *Electroanalysis*, **6**(11–12), 990–995 (1994).
80. P.W. Alexander, T. Dimitrakopoulos, D.B. Hibbert, 'Photocured Ammonium and Hydrogen Ion Selective Coated-wire Electrodes Used Simultaneously in a Portable Battery-powered Flow Injection Analyzer', *Electroanalysis (NY)*, **9**(17), 1331–1336 (1997).
81. P.W. Alexander, T. Dimitrakopoulos, D.B. Hibbert, 'A Photo-cured Coated-wire Potassium-ion-selective Electrode for Use in Flow-injection Potentiometry', *Electroanalysis (NY)*, **9**(11), 813–817 (1997).
82. P.W. Alexander, T. Dimitrakopoulos, D.B. Hibbert, 'A Photo-cured Coated-wire Calcium Ion-selective Electrode for Use in Flow-injection Potentiometry', *Talanta*, **44**(8), 1397–1405 (1997).

83. S.A. Halvatzis, M.M. Timotheou-Potamia, C.E. Efstathiou, 'Kinetic Study of *N*-Bromosuccinimide Reactions and Kinetic Determination of Pyridoxine Using a Bromide-selective Electrode', *Talanta*, **40**(8), 1213–1220 (1993).
84. M. Jiang, Y. Li, X. Zhou, Z. Zhao, J. Wang, J. Mo, 'Kinetic Determination of Hexacyanoferrates by Their Inhibition of an Oscillating Chemical Reaction', *Anal. Chim. Acta*, **236**(2), 411–416 (1990).
85. J.G. Pentari, C.E. Efstathiou, M.A. Koupparis, 'Kinetic Determination of Carbocysteine in Syrup Based on Its Reaction with 1-Fluoro-2,4-dinitrobenzene Monitored with a Fluoride-selective Electrode', *Int. J. Pharm.*, **77**(1), 41–46 (1991).
86. J.G. Pentari, C.E. Efstathiou, T.P. Hadjiioannou, 'Kinetic Determination of EDTA and Citrate by the Displacement of Fluoride from Aluminium(III)-Fluoride Complexes and Use of a Fluoride Ion-selective Electrode', *Talanta*, **38**(3), 295–301 (1991).
87. W.P. Venetz, C. Mangan, I.W. Siddiqi, 'Kinetic Determination of Alkaline Phosphatase Activity Based on Hydrolytic Cleavage of the P–F Bond in Monofluorophosphate and Fluoride Ion-selective Electrode', *Anal. Biochem.*, **91**(1), 127–132 (1990).
88. E. Athanasiou-Malaki, M.A. Koupparis, T.P. Hadjiioannou, 'Kinetic Determination of Primary and Secondary Amines Using a Fluoride-selective Electrode and Based on Their Reaction with 1-Fluoro-2,4-dinitrobenzene', *Anal. Chem.*, **61**(13), 1358–1363 (1989).
89. C. Sanchez-Pedreno, J.A. Ortuno, J. Martinez-Rodenas, 'Kinetic–Potentiometric Determination of Copper in Real Samples with a New Antimony(V) Ion-selective Electrode', *Fresenius' J. Anal. Chem.*, **344**(3), 100–103 (1992).
90. F. Grases, C. Genestar, 'Potentiometric Technique for Kinetic Determination of Citrate', *Talanta*, **38**(12), 1353–1357 (1991).
91. K. Matsumoto, J.J. Baeza-Baeza, H.A. Mottola, 'Continuous-flow Sensor Strategy Comprising a Rotating Bioreactor and a Stationary Ring Amperometric Detector', *Anal. Chem.*, **65**(5), 636–639 (1993).
92. K. Matsumoto, J.J. Baeza-Baeza, H.A. Mottola, 'Simultaneous Kinetic-based Determination of Fructose and Ascorbate with a Rotating Bioreactor and Amperometric Detection: Application to the Analysis of Food Samples', *Anal. Chem.*, **65**(13), 1658–1661 (1993).
93. B. Lopez-Ruiz, M. Sanchez-Cabezudo, H.A. Mottola, 'Enzymic Determinations with Rotating Bioreactors and Continuous-flow–Stopped-flow Processing. Determination of Choline Esters in Pharmaceuticals', *Analyst*, **121**(11), 1695–1698 (1996).
94. J.J. Baeza-Baeza, K. Matsumoto, H.A. Mottola, 'Comparative Studies on Data Collection and Data Treatment in Kinetic-based Determinations with Two Rotating Bioreactor/Amperometric Detection Systems', *Anal. Chim. Acta*, **283**(2), 785–793 (1993).
95. F. Matei, V. Paraschivescu, G. Ionescu, 'Application of Electrochemical Sensor for Oxygen in Kinetic Determination of Cobalt', *Rev. Chim (Bucharest)*, **42**(8–9), 420–423 (1991).
96. I.I. Alekseeva, L.V. Kurtova, 'Kinetic Determination of Trace Amounts of Arsenic with Amperometric Reaction Rate Detection', *Zh. Anal. Khim.*, **43**(8), 1449–1452 (1988).
97. C. Sanchez-Pedreno, M.T. Sierra, M.I. Sierra, A. Sanz, 'Catalytic Effect of Mercury(II) on the Photo-oxidation of EDTA by Iodine: Kinetic Determination of Mercury', *Analyst*, **113**(1), 145–148 (1988).
98. C. Sanchez-Pedreno, M.T. Sierra, M.I. Sierra, A. Sanz, 'Kinetic Determination of Trace Amounts of Nitrite Based on Its Inhibitory Effect on the Photochemical Reaction Between Iodine and Ethylenediaminetetraacetic acid', *Analyst*, **112**(6), 837–840 (1987).
99. W. Ciesielski, 'Kinetic Determination of Vitamin B1 and Cystine in an Open System', *Chem. Anal. (Warsaw)*, **33**(6), 987–992 (1988).
100. S. Rival, C. Besson, J. Saulnier, J.M. Wallach, 'Use of Conductometry for the Monitoring of Thermodynamically-controlled Enzymic Peptide Synthesis', *Anal. Chim. Acta*, **312**(2), 213–216 (1995).
101. J.A. Amberson, G. Svehla, 'Catalytic Determination of Vanadium with the Bromate Iodide–Ascorbic Acid Landolt Reaction and Conductometric Monitoring', *Anal. Chim. Acta*, **185**(2), 201–208 (1986).
102. W.K. Xin, X.X. Gao, 'Study of the Effect of Lanthanide Ions on the Kinetics of Glutamate Dehydrogenase by a Chronoamperometric Method', *Analyst*, **121**(5), 687–690 (1996).
103. L.N. Moskvin, D.V. Golikov, J.P. Josse, M.F. Grigoreva, I.V. Kruglova, 'Catalytic Determination of Trace Vanadium in Solution with Ionometric Measurement of the Reaction Rate', *Zh. Anal. Khim.*, **47**(7), 1304–1309 (1992).
104. Z.L. Jiang, A.H. Liang, 'Catalytic Kinetic Determination of Trace Amounts of Ascorbic Acid with Single-sweep Oscillopolarography', *Anal. Chim. Acta*, **278**(1), 53–58 (1993).
105. S.H. Wang, L.Y. Du, A.M. Zhang, B. Li, 'Catalytic Kinetic Determination of Trace Amounts of Iron(III) with Polarographic Detection', *Anal. Lett.*, **30**(11), 2099–2107 (1997).
106. A.A. Ensafi, M.S. Kolagar, 'Catalytic Kinetic Determination of Ultratrace Amounts of Nitrite with Detection by Differential Pulse Polarographic Method', *Anal. Lett.*, **28**(7), 1245–1260 (1995).
107. H.A. Mottola, D. Pérez-Bendito, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **68**(12), 257R–289R (1996); **66**(12), 131R–162R (1994); **64**(12), 407R–428R (1992).
108. H.A. Mottola, D. Pérez-Bendito, H.B. Mark, Jr, 'Kinetic Determinations and Some Kinetic Aspects of Analytical

Chemistry', *Anal. Chem.*, **62**(12), 441R–461R (1990); **60**(12), 181R–200R (1988).
109. H.A. Mottola, H.B. Mark, Jr, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **58**(5), 264R–279R (1986); **56**(5), 96R–112R (1984).
110. R.W. Engstrom, *RCA Photomultiplier Handbook*, PMT-62, RCA Corp., Solid State Division, Lancaster, PA, 1980.
111. M. Trojanowicz, P.J. Worsfold, J.R. Clinch, 'Solid-state Photometric Detectors for Flow-injection Analysis', *Trends. Anal. Chem.*, **7**(8), 301–305 (1988).
112. D.J. Hooley, R.E. Dessy, 'Continuous-flow Kinetic Techniques in Flow-injection Analysis', *Anal. Chem.*, **55**(2), 313–320 (1983).
113. I.M. Warner, G. Patonay, M.P. Thomas, 'Multi-dimensional Luminescence Measurements', *Anal. Chem.*, **57**(3), 463A–483A (1985).
114. A. Moreno, M. Silva, D. Pérez-Bendito, M. Valcarcel, 'Catalytic Effect of Iron(III) on Oxidation of 2-Hydroxybenzaldehyde [Salicylaldehyde] Thiosemicarbazone by Hydrogen Peroxide. Kinetic Fluorimetric Determination of Iron', *Anal. Chim. Acta*, **157**(2), 333–342 (1984).
115. F. Grases, C. Genestar, J.G. March, 'Kinetic Fluorimetric Determination of Cerium with 1,5-Diphenyl-3-styryl-2-pyrazoline', *Microchem. J.*, **29**(2), 237–242 (1984).
116. G.L. Campi, J.D. Ingle, Jr, 'Fluorimetric Reaction-rate Method for the Determination of Aluminium with Acid Alizarin Garnet R', *Anal. Chim. Acta*, **224**(2), 363–372 (1989).
117. M.E. Martinez-Izquierdo, J.S. Durand, A. Cabrera-Martin, S. Vicente-Pérez, 'Photokinetic Study of the Rose Bengal B–Tin(II) Reaction. Determination of Tin in Metals', *Anal. Chim. Acta*, **325**(1–2), 81–86 (1996).
118. M.A. Lopez-Fernandez, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic–Fluorimetric Determination of Traces of Nickel by Its Catalytic Effect on the Oxidation of Pyridoxal Hydrazone by Potassium Peroxodisulfate', *Anal. Lett.*, **17**(A6), 507–521 (1984).
119. S. Kawakubo, K. Ogihara, M. Iwatsuki, 'Catalytic Spectrofluorimetric Determination of Vanadium Using Oxidation of *o*-Phenylenediamine with Bromate in the Presence of Gallic Acid', *Analyst*, **120**(11), 2719–2723 (1995).
120. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Stopped-flow Determination of Diphacinone Based on Lanthanoid-sensitized Luminescence', *Anal. Chim. Acta*, **280**(1), 163–168 (1993).
121. L. De-La-Pena, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Determination of Nortriptylin in Pharmaceutical Samples by Use of Photometric and Fluorimetric Detection', *J. Pharm. Biomed. Anal.*, **13**(3), 199–203 (1995).
122. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Individual and Joint Stopped-flow Kinetic Spectrofluorimetric Determination of Neomycin and Tyrothricin', *Anal. Chim. Acta*, **303**(1), 31–37 (1995).
123. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Stopped-flow Kinetic Determination of Nalidixic Acid and Norfloxacin Based on Lanthanide-sensitized Fluorescence', *Anal. Chim. Acta*, **303**(1), 39–45 (1995).
124. H.P. Huang, R.X. Cai, Y.M. Du, Z.X. Lin, Y.E. Zeng, 'Micelle-enhanced Spectrofluorimetric Assay for Laccase Activity by a Flow-injection Stopped-flow Technique', *Anal. Chim. Acta*, **318**(1), 63–69 (1995).
125. J.A. Murillo-Pulgarin, L.F. Garcia-Bermejo, 'Flow-injection Stopped-flow Spectrofluorimetric Kinetic Determination of Paracetamol Based on Its Oxidation Reaction by Hexacyanoferrate(III)', *Anal. Chim. Acta*, **333**(1–2), 59–69 (1996).
126. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Determination of Salicylic Acid, Diflunisal and Their Mixture Based on Lanthanide-sensitized Luminescence', *Anal. Chim. Acta*, **329**(1), 135–141 (1996).
127. A. Gaikwad, A. Gómez-Hens, D. Pérez-Bendito, 'Use of Stopped-flow Fluorescence–Polarization Immunoassay in Drug Determinations', *Anal. Chim. Acta*, **280**(1), 129–135 (1993).
128. D. Pérez-Bendito, A. Gómez-Hens, A. Gaikwad, 'Direct Stopped-flow Fluorescence Polarization Immunoassay of Abused Drugs and Their Metabolites in Urine', *Clin. Chem.*, **40**(8), 1489–1493 (1994).
129. E.G. Matveeva, M.P. Aguilar-Caballos, S.A. Eremin, A. Gómez-Hens, D. Pérez-Bendito, 'Use of Stopped-flow Fluoroimmunoassay in Pesticide Determination', *Analyst*, **122**(8), 863–866 (1997).
130. O.C. Zafiriou, M.B. True, 'Interferences in Environmental Analysis of Nitric Oxide by Nitric Oxide Plus Ozone Detectors: A Rapid Screening Technique', *Environ. Sci. Technol.*, **20**(6), 594–596 (1986).
131. N. Dan, M.L. Lau, M.L. Grayeski, 'Micellar-enhanced Aqueous Peroxyoxalate Chemiluminescence', *Anal. Chem.*, **63**(17), 1766–1771 (1991).
132. M. Tabata, M. Totani, J. Endo, 'Co-immobilized Enzyme Columns in Determining Serum Creatinine Using Creatininase, Creatinase and Sarcosine Oxidase by Flow-injection Analysis and Chemiluminescence Detection', *Anal. Chim. Acta*, **262**(2), 315–321 (1992).
133. A. Navas-Diaz, J.A. Gonzalez-Garcia, 'Non-linear Multi-component Kinetic Analysis for the Simultaneous Stopped-flow Determination of Chemiluminescence Enhancers', *Anal. Chem.*, **66**(7), 988–993 (1994).
134. J. Cepas, M. Silva, D. Pérez-Bendito, 'Sensitive Peroxyoxalate Chemiluminescence Determination of Psychotropic Indole Derivatives', *Analyst*, **121**(1), 49–54 (1996).
135. J. Mateu, R. Forteza, V. Cerda, 'Kinetic–Thermometric Determination of Vanadium in Atmospheric Aerosols Based on Its Catalytic Effect on the Oxidation of Tannic Acid by Bromate Ion', *Thermochim. Acta*, **247**(2), 457–464 (1994).

136. J. Mateu, R. Forteza, V. Cerda, M. Colom-Altes, 'Kinetic Thermometric Determination of Vanadium by its Catalytic Action on the Oxidation of Chromotropic Acid by Bromate. Application to the Analysis of Atmospheric Aerosols', *Analyst*, **119**(5), 1077–1080 (1994).
137. E. Gomez, J.M. Estela, V. Cerda, 'Kinetic Thermometric Determination of Traces of Manganese by its Catalytic Effect on the Reaction Between Tiron and Hydrogen Peroxide', *J. Thermal Anal.*, **37**(1), 195–202 (1991).
138. A. Caro, G. Cortes, V. Cerda, 'Kinetic–Thermometric Determination of Manganese(II) Based on Its Catalytic Action on the Oxidation of Sulfanilic Acid by Periodate', *Analyst*, **115**(6), 753–755 (1990).
139. F. Borrull, P. Ramirez-de-la-Piscina, N. Homs, J. Torres, V. Cerda, 'Simple Kinetic–Thermometric Determination of Sub-microgram Quantities of Ruthenium Based on Its Catalytic Effect on the Cerium(IV)–Arsenic(III) Reaction', *Thermochim. Acta*, **12**, 209–216 (1988).
140. F. Borrull, J. Torres, V. Cerda, 'Kinetic–Thermometric Determination of Nickel(II) Based on Its Catalytic Action in the Decomposition Reaction of Permanganate', *Analyst*, **112**(10), 1453–1455 (1987).
141. F. Borrull, V. Cerda, 'Kinetic–Thermometric Determination of Sub-microgram Quantities of Palladium(II)', *Thermochim. Acta*, **11**, 373–377 (1987).
142. R. Forteza, V. Cerda, 'Specific Thermometric Determination of Trace Vanadium in Steel Based on Its Catalytic Effect on the Oxidation of Gallic Acid by Bromate', *Anal. Chem.*, **58**(2), 453–455 (1986).
143. R. Forteza, A. Canals, A. Caro, J.M. Estela, V. Cerda, 'Simultaneous Kinetic–Thermometric Determination of Traces by the Inhibitory Effect of Tungsten(VI) on the Vanadium(V)-catalyzed Oxidation of Thiosulfate Ion by Hydrogen Peroxide', *Fresenius' J. Anal. Chem.*, **345**(6), 432–436 (1993).
144. E. Gomez, J.M. Estela, V. Cerda, 'Simultaneous Thermometric Determination of Manganese and Lead Based on Their Catalytic Effect on the Oxidation of Tiron by Hydrogen Peroxide', *Anal. Chim. Acta*, **251**(1–2), 305–311 (1991).
145. F. Borrull, V. Cerda, 'Kinetic–Thermometric Determination of Cobalt(II) and Copper(II) Based on Their Catalytic Effect on Permanganate Decomposition in Basic Media', *Thermochim. Acta*, **13**(2), 263–268 (1989).
146. R. Forteza, E. Gomez, J.M. Estela, V. Cerda, 'Sequential Kinetic–Thermometric Determination of the Activity of Peroxidase and Catalase Using Catechol as Substrate and Inhibitor for Their Reaction with Hydrogen Peroxide', *Anal. Chim. Acta*, **284**(2), 453–459 (1993).
147. J.W. Grate, S.J. Martin, R.M. White, 'Acoustic Wave Micro-sensors. Parts I and II', *Anal. Chem.*, **65**(21), 940A–948A (1993); **65**(22), 987A–996A (1993).
148. Q.Y. Cai, W. Wei, R.H. Wang, L.H. Nie, S.Z. Yao, 'Measurement of Serum Alkaline Phosphatase with a Surface Acoustic Wave Impedance Sensor Device', *Anal. Sci.*, **13**(1), 121–125 (1997).
149. Q.Y. Cai, R.H. Wang, L.H. Wu, L.H. Nie, S.H. Yao, 'Surface Acoustic Wave Enzyme Sensor Applied to the Kinetic Assay of Acid Phosphatase', *Analyst*, **120**(12), 2833–2836 (1995).
150. Q.Y. Cai, R.H. Wang, L.Y. Wu, L.H. Nie, S.Z. Yao, 'Surface Acoustic Wave (SAW)–Impedance Sensor for Kinetic Assay of Trypsin', *Microchem. J.*, **55**(3), 367–374 (1997).
151. Q.Y. Cai, W. Wei, R.H. Wang, L.H. Nie, S.Z. Yao, 'Kinetic Assay of Antitrypsin in Human Serum by a Surface Acoustic Wave (SAW) Impedance Sensor', *Fresenius' J. Anal. Chem.*, **356**(1), 96–97 (1996).
152. K. Ge, D.Z. Liu, K. Chen, L.H. Nie, S.Z. Yao, 'Assay of Pancreatic Lipase with the Surface Acoustic Wave Sensor System', *Anal. Biochem.*, **226**(2), 207–211 (1995).
153. D.H. Liu, K. Ge, K. Chen, L.H. Nie, S.H. Yao, 'Clinical Analysis of Urea in Human Blood by Coupling a Surface Acoustic Wave Sensor with Urease Extracted from Pumpkin Seeds', *Anal. Chim. Acta*, **307**(1), 61–69 (1995).
154. D.Z. Liu, L.H. Nie, S.Z. Yao, 'Kinetic Analysis of Arginase–Urease Coupled Reaction with a Surface Acoustic Wave Enzyme Sensor System', *Talanta*, **43**(4), 667–674 (1996).
155. M.S. Nieuwenhuizen, J.L.N. Harteveld, 'Studies on a Surface Acoustic Wave (SAW) Dosimeter Sensor for Organophosphorus Nerve Agents', *Sens. Actuators B*, **40**(2–3), 167–173 (1997).
156. W. Wei, R.H. Wang, L.H. Nie, S.Z. Yao, 'Enzymic Determination of Organophosphorus Pesticides with a Surface Acoustic Wave Impedance Sensor', *Instrum. Sci. Technol.*, **25**(2), 157–167 (1997).
157. J.V. Sweedler, 'Charge Transfer Device Detectors and Their Applications to Chemical Analysis', *Crit. Rev. Anal. Chem.*, **24**(1), 59–98 (1993).
158. H. Muller, V. Muller, E.H. Hansen, 'Simultaneous Differential Rate Determination of Iron(II) and Iron(III) by Flow Injection Analysis', *Anal. Chim. Acta*, **230**, 113–123 (1990).
159. M.A. Victor, S.R. Crouch, 'Absorbance-corrected Synchronous Fluorescence with a Fiber-optic-based Fluorimeter', *Appl. Spectrosc.*, **49**(7), 1041–1047 (1995).
160. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Use of a Stopped-flow/T-format Spectrofluorimeter for Simultaneous Kinetic Analyses', *Anal. Chim. Acta*, **310**(3), 453–459 (1995).
161. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Simultaneous Determination of Ampicillin and Tetracycline in Milk by Using a Stopped-flow/T-format Spectrofluorimeter', *Talanta*, **44**(10), 1883–1889 (1997).
162. G.L. Campi, J.D. Ingle, Jr, 'Monitoring and Compensating for Between-run Temperature and pH Variations in Kinetic-based Determinations', *Anal. Chim. Acta*, **224**, 275–287 (1989).

163. H.D. Dewald, J. Wang, 'Spectroelectrochemical Detector for Flow-injection Systems and Liquid Chromatography', *Anal. Chim. Acta*, **166**, 163–170 (1984).
164. H.L. Pardue, 'Kinetic Aspects of Analytical Chemistry', *Anal. Chim. Acta*, **16**(1–2), 69–107 (1989).
165. W.J. Blaedel, G.P. Hicks, 'Continuous Analysis by Measurement of the Rate of Enzyme Catalyzed Reactions', *Anal. Chem.*, **34**(3), 388–394 (1962).
166. J.S. Ingle, S.R. Crouch, 'Theoretical and Experimental Factors Influencing the Accuracy of Analytical Rate Measurements', *Anal. Chem*, **43**(6), 697–701 (1971).
167. H.L. Pardue, 'Comprehensive Classification of Kinetic Methods of Analysis Used in Clinical Chemistry', *Clin. Chem.*, **23**(12), 2189–1201 (1977).
168. H.L. Pardue, W.E. Dahl, 'All-electronic System for the Automatic Measurement of Slopes of Reaction-rate Curves', *J. Electroanal. Chem.*, **8**(4), 268–276 (1964).
169. E.A. Cordos, S.R. Crouch, H.V. Malmstadt, 'An Automated Digital Readout System for Reaction Rate Methods', *Anal. Chem.*, **40**(10), 1812–1818 (1968).
170. S.D. Brown, 'Kalman Filter in Analytical Chemistry', *Anal. Chim. Acta*, **181**, 1–26 (1986).
171. S.C. Rutan, S.D. Brown, 'Estimation of First-order Kinetic Parameters by Using the Extended Kalman Filter', *Anal. Chim. Acta*, **167**, 3–37 (1985).
172. Y-S. Hsieh, S.R. Crouch, 'Determination of Enzyme Substrates with an Extended Range of Linearity', *Anal. Chim. Acta*, **309**(1–3), 251–258 (1995).
173. I. Laios, D.M. Fast, H.L. Pardue, 'Development and Evaluation of Conditions for the Kinetic Quantitation of Acetoacetate in Body Fluids', *Anal. Chim. Acta*, **180**, 429–443 (1986).
174. R.S. Harner, H.L. Pardue, 'Regression Data-processing Methods for Parallel Zero- and First-order Processes Applied to Lactate Dehydrogenase Sub-unit Determinations', *Anal. Chim. Acta*, **127**, 23–38 (1981).
175. J.D. Lin, H.L. Pardue, 'Multi-point Kinetic Methods Evaluated for Quantitation of Theophylline Prosthetic Group Label Immunoassay', *Clin. Chem*, **28**(10), 2081–2087 (1982).
176. D. Pérez-Bendito, M. Silva, 'Recent Advances in Kinetometrics', *Trends Anal. Chem.*, **15**(6), 232–240 (1996).

Luminescence-based Kinetic Determinations

A. Gómez-Hens
Department of Analytical Chemistry, University of Córdoba, Spain

1 Introduction 1
2 Theoretical Principles 2
3 Photoluminescence 4
3.1 Fluorescence 4
3.2 Phosphorescence 6
3.3 Time-resolved and Phase-resolved Luminescence 7
3.4 Lanthanide-sensitized Luminescence 7
4 Chemiluminescence 8
4.1 Conventional Chemiluminescence 8
4.2 Electrogenerated Luminescence 8
4.3 Bioluminescence 9
5 Immunoassay 9
6 Automation 11
6.1 Flow Injection 11
6.2 Stopped-flow 12
6.3 Continuous Addition of Reagent 13
7 Simultaneous Determinations 14
8 Final Remarks 16
Acknowledgments 16
List of Symbols 16
Abbreviations and Acronyms 16
Related Articles 16
References 17

Luminescence-based kinetic determinations are of great analytical interest and flexibility as they combine the high selectivity of kinetic methodology with the low detection limits afforded by the different luminescence techniques.

1 INTRODUCTION

Luminescence techniques can be included among monitoring techniques for kinetic determinations. However, the typically high selectivity and low detection limits of the ensuing methods make them especially attractive for kinetic analysis, where measured concentration changes are usually quite small. Like other kinetic methods, the growth of luminescence-based kinetic methods has relied heavily on the availability of suitable instrumentation. Thus, many reaction rate methods were initially less precise and more time-consuming and tedious than conventional equilibrium methods because the reaction rate had to be manually extracted from the recorder tracing, which decreased the precision of the results and increased the data treatment time. In addition, the instability of luminescence instruments precluded wider use of dynamic measurements based on using luminescence detection. These problems have been overcome with the inception of electronic devices and computers that permit the automated acquisition of data as the process develops, and afford enhanced data handling. The availability of stable luminescence instruments, automated precise sample and reagent delivery systems, and improved temperature-control equipment has minimized the shortcomings of luminescence-based kinetic determinations, which are now effective alternatives to equilibrium determinations as shown by the abundant literature on their applications.

The use of dynamic luminescence measurements affords very low detection limits (better than those obtained with equilibrium methods in some instances). Thus, when a luminescent system yields a static background signal, which is very usually the case in analyses of real samples, discriminating it from the equilibrium analytical signals obtained at very low analyte concentrations can be difficult because both are static signals. However, when the reaction rate is used as the analytical parameter, its dynamic character makes it less strongly dependent on the background signal provided this does not change with time. Also, analytical measurements made at the beginning of the reaction are known to avoid or minimize the potential inhibitory effect of other constituents of the sample matrix.

The first reported application of the combination of luminescence techniques and kinetic methodology was the fluorimetric monitoring of initial reaction rates for the determination of the rate constants involved in the association and dissociation of flavin mononucleotide (FMN) with its apoenzyme.[1] The study, which was carried out by using a filter fluorimeter, led to the development of the many luminescence-based kinetic methods available today. Early fluorimetric reaction rate methods were reviewed by Ingle and Ryan in 1981,[2] who noted the scant practical use of kinetic fluorimetric methods. However, continual improvements in available equipment have helped to consolidate fluorimetry and other luminescence techniques to be consolidated as alternatives to the analysis of real samples using kinetic methodology.

Broadly speaking, luminescence-based kinetic determinations can be classified into four groups, as follows.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1. Determinations based on the rate of change of chemical composition with time, which is monitored by measuring the variation of the luminescence signal with time. These determinations involve the use of conventional kinetic approaches such as the tangent, fixed time, and fixed signal (or variable time) methods.[3] Determinations based on measurements of the induction period, which arises in Landolt reactions[4] are also included in this group.
2. Determinations based on the use of flow systems with luminescence detection. Measurements are obtained when neither physical (homogenization) nor chemical equilibrium has been attained by the time that the sample zone reaches the continuous detector, so these determinations are doubly kinetic in nature.
3. Determinations based on measuring the lifetime of the excited state of a chemical species, which is obtained as the molecules return from the excited state to the ground state.
4. Determinations based on the use of chemiluminescence (CL) reactions in which the intensity of the emitted light is recorded as a function of the time and the peak intensity or the integrated area under all or part of the intensity–time curve is used as the analytical parameter.

The different luminescence techniques are included in the first and second group as they are used only as detection systems; by contrast, those in the third and fourth group unavoidably involve two luminescence processes: photoluminescence (fluorescence or phosphorescence) and CL. All these determinations involve making analytical measurements while the system is under nonequilibrium conditions, so kinetics play a significant role. Although each group is commented on to some extent here, this chapter deals mainly with traditional kinetic determinations,[5,6] i.e. those in the first group, which also includes CL and flow methods based on kinetic methodology. For a more systematic discussion, they have been classified according to the luminescence process involved. The special features of immunoassay and automation justify coverage in separate sections. The last section describes selected examples of simultaneous determinations.

2 THEORETICAL PRINCIPLES

All luminescence phenomena are processes of an intrinsically kinetic nature as they involve the emission of ultraviolet, visible or near-infrared radiation from a molecule or an atom resulting from the transition of an electronically excited state to a lower energy state (usually the ground state). Thus, in photoluminescence processes, which involve the use of electromagnetic radiation as the exciting source and the transition from a singlet (fluorescence) or a triplet (phosphorescence) electronically excited state, measurements are carried out under conditions of dynamic change controlled by the rates of deactivation processes that follow the excitation. Figure 1 shows the typical rate constants for the different excited state processes.

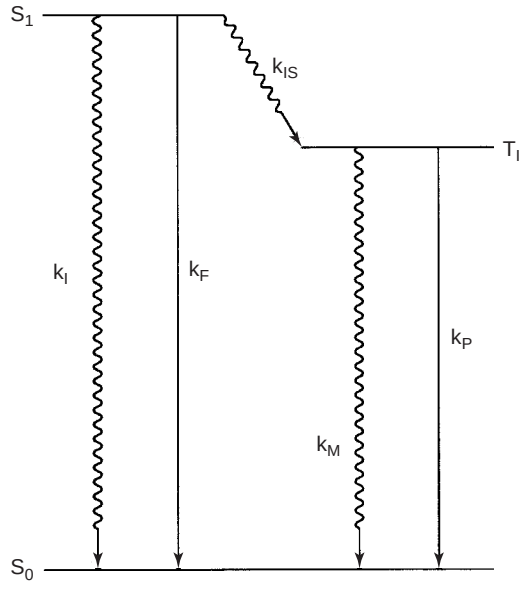

Figure 1 Rate constants involved in the different excited-state processes: (S_0) ground state; (S_1) lowest excited singlet state; (T_1) lowest triplet state; (k_I) internal conversion for lowest excited singlet state; (k_F) fluorescence; (k_M) intersystem crossing between lowest triplet state and ground state; (k_P) phosphorescence; (k_{IS}) intersystem crossing. Wavy lines indicate nonradiative transitions.

The excited states responsible for luminescence phenomena have finite lifetimes that can be measured by applying time-resolved methodologies and used as additional analytical parameters. In principle, time-resolved measurements contain more information than do steady-state measurements, since steady-state values represent the time averages of time-resolved determinations. The temporal information given by the fluorescence or phosphorescence lifetime can be employed, usually in combination with other parameters, to identify compounds. Also, if two or more luminescent compounds have similar absorption and emission spectra but different luminescence decay times, the compounds can be distinguished from one another by differential kinetic analysis.

In general, the excitation of a luminophore with an infinitely short pulse of light provides the best approach to explaining the luminescence decay. The variation of

the luminescence intensity as a function of time after the excitation pulse is applied is assumed to conform to an exponential decay curve, since the luminescence decay process usually obeys first-rate kinetics (Equation 1):

$$I_t = I_0\, e^{-kt} = I_0\, e^{-(t/\tau)} \tag{1}$$

where I_t is the luminescence intensity at time t, I_0 is the maximum luminescence intensity during excitation and t is the time after removal of the exciting radiation. Parameter k is the first-order rate constant for the decay process and equals the reciprocal of the mean lifetime of the excited state, τ. When $t = \tau$, the luminescence intensity is equal to 1/e of its initial value. Thus, the luminescence lifetime is the time required for the luminescence intensity to decrease to 1/e of its initial value.[7] Experimental lifetimes, whether for fluorescence or phosphorescence, indicate the overall rate at which the excited state is deactivated, through both radiative and nonradiative processes.[8,9]

The intrinsic luminescence lifetime, τ_0, can be experimentally obtained, provided luminescence is the only mechanism by which the excited molecule returns to the ground state. For a fluorescent process, the relationship between τ_F (observed fluorescence lifetime) and τ_{F0} (intrinsic luminescence lifetime) is given by Equation (2)

$$\tau_F = \phi_F \tau_{F0} \tag{2}$$

where ϕ_F is the fluorescence quantum yield, which can be defined in terms of rate constants of fluorescence (k_F), internal conversion for lowest excited singlet state (k_I) and intersystem crossing (k_{IS}) as in Equation (3):

$$\phi_F = \frac{k_F}{k_F + k_I + k_{IS}} \tag{3}$$

Also, τ_F can be defined as a function of such rate constants (Equation 4):

$$\tau_F = \frac{1}{k_F + k_I + k_{IS}} \tag{4}$$

If a biomolecular quenching process were competing for fluorescence, then the term $k_Q[Q]$ would be added to the denominator, where k_Q is the rate constant for biomolecular quenching and $[Q]$ the concentration of quencher.

Similar expressions can be obtained for phosphorescence decay, but the relationship between the quantum yield (ϕ_P) and intrinsic (τ_{P0}) and observed (τ_P) lifetimes is different in this instance, as the triplet formation quantum yield (ϕ_T) must be also included, as in Equation (5):

$$\tau_P = \frac{\phi_P \tau_{P0}}{\phi_T} \tag{5}$$

where $\phi_P = [k_P/(k_P + k_M)][k_{IS}/(k_{IS} + k_I + k_F)]$, in which k_M denotes the intersystem crossing between the lowest triplet state and the ground state rate constant, and $\phi_T = k_{IS}/(k_{IS} + k_I + k_F)$. Thus, τ_P is given by Equation (6):

$$\tau_P = \frac{1}{k_P + k_M} \tag{6}$$

Additionally, if a quenching process were present, the term $k_Q[Q]$ would be added to the denominator, as in fluorescence. The situation would also be more complex if other phenomena such as photochemical or energy transfer processes were present.

Phosphorescence triplet states of organic molecules usually have relatively long mean lifetimes (at the level of milliseconds or greater), so measuring τ_P is fairly simple and can be accomplished with conventional commercial phosphorescence spectrometers. However, τ_F is much shorter, at the level of 10^{-8}–10^{-11} s, and its measurement is electronically a much more difficult task. Two general approaches to determining the fluorescence lifetime have been developed that are named time-resolved and phase-resolved methods, and are based on the use of two different types of techniques, namely: pulse fluorimetry, which relates measurements performed in the time domain; and phase and modulation fluorimetry, which relate to the frequency domain.[10,11]

With CL, where the molecule in the electronically excited state is produced by a chemical, biochemical (bioluminescence) BL or electrochemical (electroluminescence) reaction, the process is probably the most markedly kinetic among luminescent phenomena, as the emission intensity, I_L, in photons emitted per seconds, is directly proportional to the rate of the reaction involved, dC/dt, in molecules reacting per second, according to Equation (7):

$$I_L = \phi_L \frac{dC}{dt} \tag{7}$$

where the proportionality constant ϕ_L is the CL quantum yield or efficiency, in photons emitted per molecule reacting. The CL efficiency is the product of the excitation quantum yield, ϕ_{EX}, in excited states produced per molecule reacting, and the emission quantum yield, ϕ_{EM}, in photons emitted per excited state. Efficiencies ranging from 10^{-15} to nearly 1 have been observed. Higher values are usually associated with BL reactions; by contrast, the conventional CL reactions usually employed for analytical purposes have typical efficiencies of 0.001–0.1.[12]

Although the lifetimes of the excited singlet states that are responsible for CL emission have typical durations in the nanoseconds region, the emission of light persists for as long as the CL reaction takes place. Thus, the photophysics of the electronically excited

products of the reaction and the kinetics of the reaction are involved in the CL emission. This justifies the prominent role assigned to CL methods among kinetic-based approaches. In fact, two of the more important factors to be considered in developing a CL method are the efficiency of the CL reaction, which ultimately affects sensitivity and detection limits, and the reaction kinetics, which dictates precision and throughput. In practice, both parameters are affected by experimental conditions including solvents, concentrations, pH, and reagent purity.

3 PHOTOLUMINESCENCE

In addition to the fact that all fluorescence and phosphorescence methods are intrinsically kinetic in nature, both luminescence processes have been applied to reaction rate methodology, but fluorescence-based kinetic determinations are much more numerous than those involving phosphorescence measurements. Overall, phosphorescence analysis has not been so widely used as fluorescence analysis, owing to the special requirements of the former, which originally called for cryogenic solvents and liquid nitrogen temperature, and later for the immobilization of phosphors adsorbed on solid matrices. However, the inception of micelle-stabilized room-temperature phosphorimetry (MSRTP),[13] which uses heavy atoms, has promoted the use of phosphorimetry as a determination tool.

3.1 Fluorescence

Fluorimetry has so far been the luminescence technique most frequently applied to reaction rate methodology. It is a competitive alternative to photometric and electrochemical measurements when fluorescent species are involved, as it generally affords better detection limits and selectivity, and wider dynamic ranges. Thus, detection limits can be two or three orders of magnitude lower than those obtained with photometric detection. Even if the chemical system involves no fluorescent species, fluorimetry can still be used as the detection system by coupling with a fluorogenic species in a secondary indicator reaction, as in a number of reported enzymatic determinations.[14] Thus, a generic enzymatic reaction can be represented by Equation (8):

$$S + E \longrightarrow P + E \tag{8}$$

where S is the substrate, E the enzyme and P the product, and any of them is fluorescent. An indicator reaction can be coupled, as in Equation (9):

$$P + R + E' \longrightarrow P' + R' + E' \tag{9}$$

where R′ is fluorescent and R is not, and E′ is the indicator enzyme. The variation of the fluorescence with time may be related to the concentration of S or E and, also, to the concentration of an activator or inhibitor of the enzymatic reaction. The scope of kinetic methodology is thus wider than that of the total change method, where the reaction is allowed to reach the equilibrium so only the substrate concentration can be determined.

Advances in instrumentation have also been applied to enzymatic kinetic analysis. Thus, an integrated microfabricated device that performs automated enzymatic analysis has been reported.[15] Electrokinetic flow was used to control the dilution and mixing of reagents, and the β-galactosidase-catalyzed hydrolysis of resorufin β-D-galactopyranoside, which is hydrolyzed to the fluorescent resorufin, was used as a model system for enzyme kinetic determination. The microchip design used is shown in Figure 2. An enzyme assay performed on the microchip within a 20-min period required only 200 pg of enzyme and 7.5 ng of substrate, thus reducing the amount of

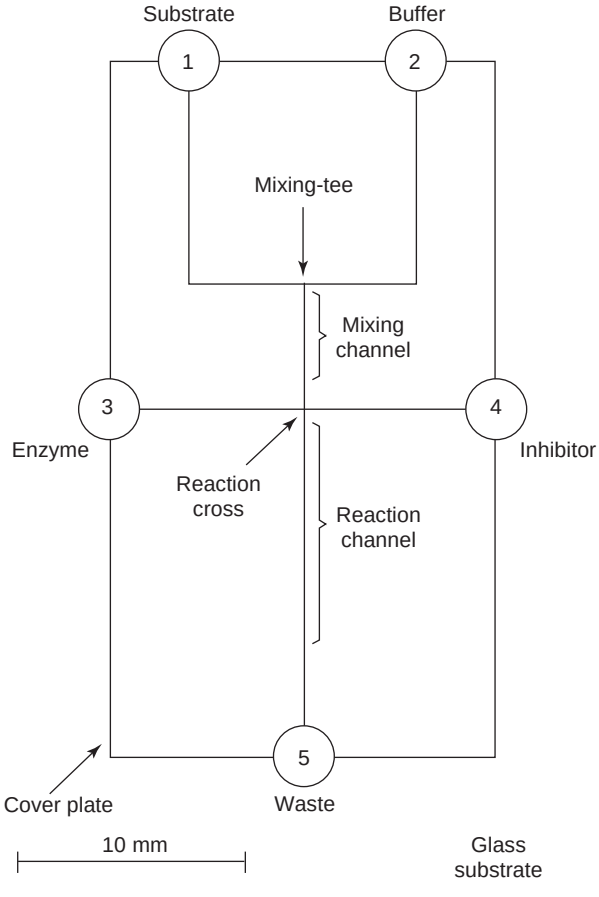

Figure 2 Diagram of the enzyme analysis chip. The channels terminate at reservoirs containing the indicated solutions. (Reproduced by permission of the American Chemical Society, from Hadd et al.[15])

reagent needed by four orders of magnitude relative to a conventional assay.

Methods for the determination of inorganic species based on fluorimetric reaction-rate measurements are relatively abundant.[5,6,16,17] Many are based on the use of redox reactions, where the inorganic species acts as a catalyst. This methodology has allowed the very sensitive determination of metal species such as iron and cobalt, which cannot be determined by other fluorimetric approaches, owing to their paramagnetism. Other determinations are based on activation or inhibition processes, using enzymatic reactions in some cases. The main shortcoming of many of these methods, particularly earlier ones, is that they were not applied to real samples. Also, the number of kinetic fluorimetric determinations for inorganic species reported in the last years has declined markedly. Kinetic methodology has scarcely been used in connection with complex formation reactions owing to their high rates, which are usually measured by using fast-mixing systems.

Fluorimetric determinations of organic compounds based on kinetic methodology generally involve uncatalyzed reactions. These reactions are as varied as the nature of the organic compounds themselves,[5,6] although oxidation and condensation reactions are the most commonly used. There has been a trend to automating these methods in recent years in various analytical fields, including clinical, pharmaceutical, environmental and food analyses. Although most kinetic fluorimetric methods utilize a reaction monitoring system that produces a single signal that is proportional to the reactant or product concentration, modern multidimensional instrumentation allows much more complete information about the chemical system to be derived. Thus, array detectors provide fluorescence spectra in a time short enough to be useful for extracting kinetic information with a view to monitoring multiple species, observing transient intermediates, obtaining information about species with overlapping spectra and correcting for background fluorescence.[2] Multidimensional detection can also be accomplished by combining two or more types of detectors. Thus, reaction rates can be corrected for the influence of temperature and pH during fluorimetric kinetic measurements by simultaneously monitoring fluorescence, pH and temperature.[18] A block diagram of the equipment used for this purpose is shown in Figure 3.

Long-wavelength fluorogenic reagents are being increasingly used to improve the selectivity of fluorimetric methods by avoiding potential spectral interferences of the sample matrix with analytical measurements, which are often encountered with conventional fluorescence reagents. The combined use of kinetic methodology and these reagents has been shown to boost selectivity through temporal and spectral discrimination of the analytical signal. This approach has been applied to the kinetic determination of proteins such as gliadins[19] and caseins[20] in foods, by reaction with a surfactant and suppression of the quenching effect of the surfactant on the fluorescence of the fluorophore. The increase in fluorescence intensity with time is directly related to the protein concentration.

The inception of computers in analytical chemistry has allowed new developments in data processing for kinetic fluorimetric determinations; applications, however, remain scant. For instance, the Kalman filter,[21] an algorithm for the recursive estimation of the parameters and states of a system from a series of noisy measurements, has been used to determine the parameters for data files with different rate constants, and has been applied to the fluorescent reactions of amino acids with phthalaldehyde.[22]

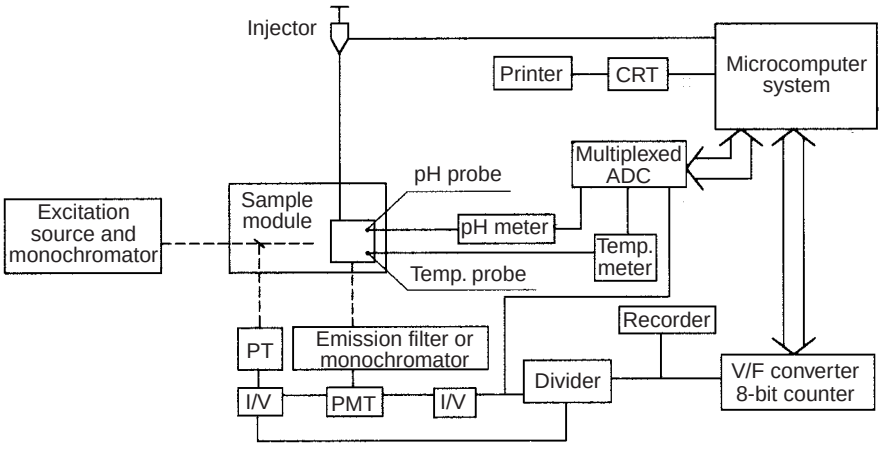

Figure 3 Block diagram of the equipment used for the simultaneous monitoring of fluorescence, pH and temperature. (I/V) current to voltage converter; (PMT) photomultiplier tube; (PT) reference vacuum phototube; (ADC) analog to digital converter; (CRT) cathode-ray tube screen. (Reproduced by permission of Elsevier, from Campli and Ingle.[18])

Dynamic fluorimetric detection has scarcely been used in combination with separation techniques. Recently, a reaction kinetic model combined with a number of theoretical and experimental variables was developed to evaluate the dynamic ranges and detection limits obtained for biotin as a function of the streptavidin concentration.[23] For this purpose, laser-induced fluorescence detection was used in postcapillary affinity detection for capillary zone electrophoresis.

3.2 Phosphorescence

Kinetic methodology has been used in phosphorimetry mainly to determine the lifetime of the excited triplet state, which is the analytical parameter in time-resolved phosphorimetry as shown in the following section. However, this approach has scarcely been used in analytical applications involving measuring the rate of formation of a phosphorescent system in solution using MSRTP. Similar to other analytical techniques, this can be done by recording the kinetic curve that reflects the variation of the phosphorescence with time and using its slope as the analytical parameter. This is especially useful when the phosphorescent signal is unstable or takes a long time to stabilize, which is very usual in phosphorescent systems and hinders application of phosphorescence methods to routine analyses. This approach has been applied to the determination of carbaryl[24] in the presence of sodium dodecyl sulfate (SDS) and the thallous ion. The advantages of this determination are noted in section 6.2. The stopped-flow (SF) mixing technique was required to measure the fast initial rate of the system.

The term ''kinetic phosphorimetry'' was coined in connection with a method for the determination of uranium,[25] based on laser excitation followed by temporal resolution of the phosphorescence signal. As noted in section 2, in the absence of side processes, the equation that describes the first-order kinetic decay of the excited species can be written as Equation (10):

$$\ln I_t = \ln I_0 - (k_p + k_q)t \qquad (10)$$

where I_t and I_0 are the phosphorescence intensities at time t and 0, respectively, and are proportional to the concentration of the emitting species, k_p is the rate constant for phosphorescent decay, and k_q is the rate constant for all other relaxation processes. The number of detected photons at any time t is proportional to the number of excited species. Thus, a linear fit of the phosphorescence intensity data in the above equation gives an intercept at time $t = 0$, $\ln I_0$, proportional to the number of excited species, irrespective of quenching effects. The luminescence from the analyte is related to its concentration using the intercept I_0 from the above expression in the calibration equation obtained with known analyte standards. The decay lifetime, τ, which is the reciprocal of $(k_p + k_q)$, corresponds to the negative reciprocal of the slope in the above equation. As a result, the linear fit of this equation allows the calculation of the concentration of the analyte and its lifetime. The use of a dye laser source is very appropriate for this type of spectroscopic measurement because of its high intensity, monochromatic radiation and the possibility of varying the excitation wavelength. Figure 4 schematically depicts the equipment required.

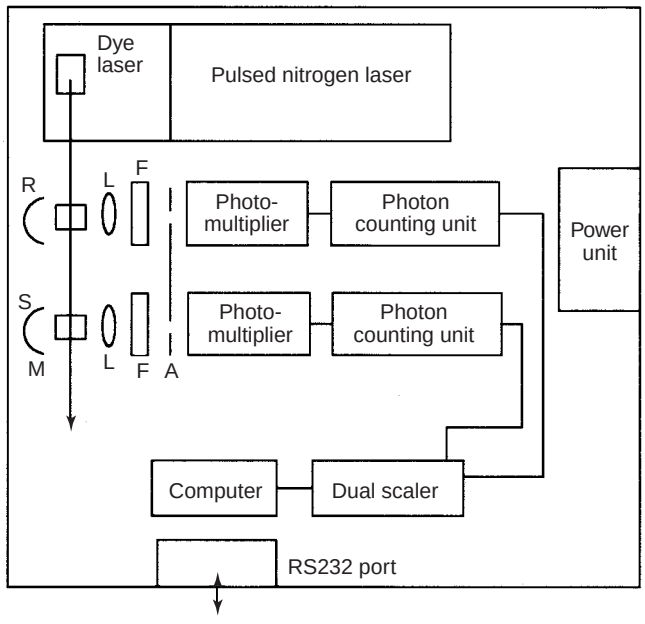

Figure 4 Schematic diagram of the kinetic phosphorescence analyzer: (R) reference cell; (S) sample cell; (L) focusing lens; (F) interference filter; (A) apertures to the detectors; (M) mirrors. (Reproduced by permission of the American Chemical Society, from Brina and Miller.[25])

The determination of uranium by using this methodology is based on the fact that the luminescence decay of an excited uranyl complex follows first-order kinetics,[25] so the intensity of the uranyl luminescence emission declines exponentially after excitation with each laser pulse and the logarithm of the luminescence intensity is a linear function of time. The luminescence intensity at time zero after termination of the laser pulse, the initial luminescence intensity, is proportional to the uranium concentration. The method has been applied to the determination of uranium in water, urine, milk and soil samples. Analyses of more complex biological samples have been limited because the organic components cause quenching of the luminescence associated with the uranyl ion. Quenching can compete with phosphorescence by shortening the excited-state lifetime and reducing the luminescence intensity, thus affecting the specificity,

sensitivity and accuracy of the method. However, analysis of these samples has been accomplished[26] by previously heating at 600 °C.

3.3 Time-resolved and Phase-resolved Luminescence

As noted in section 2, luminescence lifetimes can be determined by using time-resolved and phase-resolved methods.There is extensive literature about both approaches.[10,11] Briefly, in the time-resolved method, the sample is excited with a short-duration pulse of light, and the time-dependent decay of the fluorescence intensity is used to calculate the lifetime. In the phase-resolved method, the sample is excited with sinusoidally modulated light and the phase shift and demodulation of the emission, relative to the incident light, are used to calculate the lifetime. Each method has some advantages and disadvantages. The choice depends on several variables, including the nature of the sample, deviation of decay from ideality, and lifetime range desired.[7] However, only the time-resolved method allows the kinetic curve to be obtained. The resolution of spectrally similar fluorofors or phosphors based on differences in lifetime is suitable when instrumental or spectroscopic methods for simultaneous determinations fail, when it may be difficult to separate the components in the sample or when chemical treatment of the sample may increase analysis time or introduce contamination. It is very unlikely that two components will have the same lifetime as well as the same excitation and emission spectra.

The qualifier ''time-resolved'' has been also applied to a series of determinations involving no luminescence lifetime measurements. These methods are based on the use of chemical systems with relatively long luminescence lifetimes. Equilibrium measurements are obtained in the phosphorescence mode, using preset delay and gating times. Lanthanide chelates, which are the subject of the next section, have been widely used for this purpose. A very interesting analytical application of these chelates is their use as labels in time-resolved fluorescence immunoassays.[27]

3.4 Lanthanide-sensitized Luminescence

The high coordination number of lanthanide ions allows the formation of luminescent ternary complexes, where one of the ligands absorbs the excitation light and efficiently transfers the energy to the chelated ion, whereas the second ligand has a merely synergistic effect, removing water molecules, which are deleterious to lanthanide emission, from the coordination sphere of the ion. The analytical applications of lanthanide-sensitized luminescence are of great interest[28] since lanthanide chelates exhibit a special behavior as a result of the intramolecular energy transfer from the excited triplet state of the ligand to the emitting level of the central ion, which partly or completely circumvents the selectivity restrictions of luminescence methods. Thus, the large Stokes shifts and narrow emission bands of these chelates afford the spectral discrimination of the analytical signal, while their relatively long lifetimes enable temporal discrimination from analytical measurements made in the phosphorescence mode to avoid scatter, Raman and any fluorescence background signals. In addition, this time-resolved mode increases the analytical signal, as this is integrated over a longer period than in the fluorescence mode, which results in improved detection limits.[29,30]

This luminescence process can be considered a special type of sensitized phosphorescence. One difference between the use of lanthanide ions as energy acceptors instead of compounds such as biacetyl or 1,4-dibromonaphthalene is that, with the latter, the energy transfer is intermolecular; this accounts for the low efficiency relative to the former, which involves intramolecular energy transfer. One other difference of practical interest is that no oxygen removal is required. The most efficient energy acceptors among lanthanide ions are terbium(III) and europium(III), which form stable complexes with many organic compounds.[28]

Although the formation of lanthanide chelates is generally very fast and the luminescence signal rapidly stabilizes, it has been shown that the use of initial rate measurements instead of equilibrium measurements can improve the selectivity of a determination.[31] Figure 5

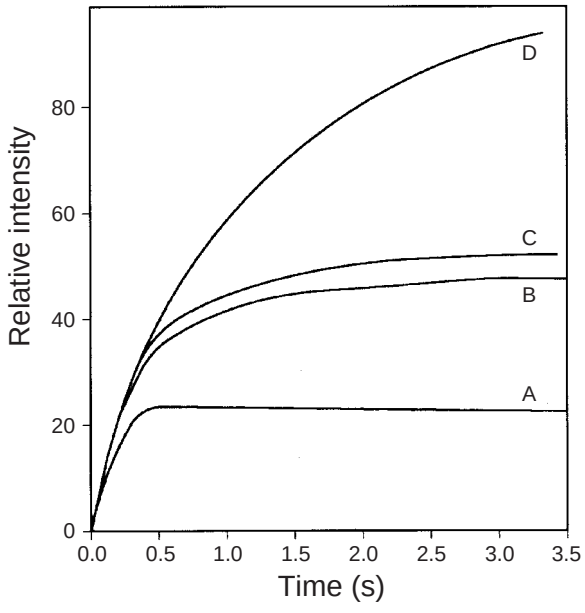

Figure 5 Kinetic curves obtained for propyl gallate (1 μg ml^{-1}): alone (A), and in the presence of octyl gallate (20 μg ml^{-1}) (B), lauryl gallate (10 μg ml^{-1}) (C), and gallic acid (100 μg ml^{-1}) (D). (Reproduced by permission of the Royal Society of Chemistry, from Panadero et al.[31])

shows the kinetic curves obtained for the terbium(III)-propyl gallate system in the presence of other gallate derivatives. As can be seen, the equilibrium signal is modified by these compounds, while the initial slope of the kinetic curves are unaffected. The usually high formation rate of lanthanide chelates precludes the use of the batch technique and justifies the use of the SF mixing technique, which, in addition, minimizes reactant manipulation and ensures a high throughput, as shown in section 6.2.

4 CHEMILUMINESCENCE

The CL process is probably the most markedly kinetic of the luminescent phenomena as the factors that affect the emission intensity are a combination of luminescence and chemical reaction rate aspects (see section 2). Thus, taking into account that CL measurements are dynamic in nature, because they are based on the detection of a transient light emission, all CL methods might be classified as kinetic methods. A large number of methods based on CL reactions have been reported in the last few years.

The analytical parameters most frequently employed in CL methods are the peak intensity and the integrated area under all or part of the light emission–time curve. In both instances, one must rigorously control the interval between initiation of the reaction and acquisition of data. However, formation and decay rates of CL reactions have been found to provide better precision and selectivity that conventional parameters, which, as shown in section 6.2, have been obtained by using SF mixing technique.[32] This section describes some general aspects about the analytical interest of the different CL processes, namely conventional CL, electrochemiluminescence and BL.

4.1 Conventional Chemiluminescence

The reactions used in CL methods can be direct or indirect,[33] depending on the origin of the CL. They involve the processes represented by Equations (11–14):

$$A + B \longrightarrow C^* \quad (11)$$

$$C^* \longrightarrow C + \text{light} \quad (12)$$

and

$$C^* + D \longrightarrow C + D^* \quad (13)$$

$$D^* \longrightarrow D + \text{light} \quad (14)$$

In the former case, the reaction produces the primary excited-state molecule, which is responsible for the subsequent light emission; in the latter process, known as a ''sensitized reaction'', the excited product is not the light emitter, but rather transfers its energy to a fluorescent acceptor, which then emits light. The use of sensitized CL has distinct advantages in the design of analytical systems, as optimum structural features can be designed and incorporated into separate chemical reactant and fluorophore molecules. The fluorophore is selected to provide the optimum compromise between excitation and emission wavelengths, chemical stability and quantum yield. This is in contrast to direct CL, where reactivity and fluorescence properties reside in a single molecule. Extensive literature on the different types of CL systems is available.[33]

The duration of CL reactions varies from very short (<1 s) to very long (>1 day). In any case, because the signal is transient, measurements of CL intensity are time-dependent, which dictates the choice of method for mixing the reactants. Batch luminometers are based on static systems where the sample is placed in a cuvette that is accommodated in the instrument and the CL reagent is added either manually (with a syringe) or via an automatic injector. The CL intensity is usually monitored over a preset time interval. This approach can be used when the CL reaction has a long lifetime, but it is unsuitable for reproducible monitoring of a fast CL reaction (<15 s). In the latter case, the use of flow systems allows the reactants to be mixed rapidly and reproducibly in front of the detector, as shown in section 6. The combination of CL detection and liquid chromatography (LC) is particularly attractive, but the sensitivity of CL detection and the selectivity of LC must be balanced against the incompatibility of the best mobile phase for separation and the best conditions for the CL reaction.[34]

4.2 Electrogenerated Luminescence

Electrogenerated chemiluminescence (ECL), also known as ''electroluminescence'' or ''electrochemiluminescence'',[35] is a process by which electrochemically generated reactants undergo high-energy electron transfer reactions in solution, to yield excited-state molecules. Usually, cation and anion radicals are produced from electron donor and acceptor molecules at an electrode surface, which is achieved by applying double step potentials to the electrode to give positive and negative potentials respectively. The radical species produced then undergo an electron transfer reaction to give the excited state of either or both reactants. Thus, the CL reaction takes place in the vicinity of an electrode and involves electrochemically generated active species. This technique retains the advantageous sensitivity and selectivity of conventional CL and affords greater control over the initiation, rate and course of the reaction, as this is produced by electrical stimulation. It provides a direct monitor of reaction rate, as the photon flux is proportional

to the rate of the oxidation or reduction that produces the excited species.

Although the ECL process has been studied for many years, it has only recently emerged as a useful tool for the determination of a variety of analytes.[36] The earliest research in this direction was aimed at studying the nature of the emitting state, the mechanisms by which it is produced and the efficiency with which the excited state is generated. Available methods are usually implemented in static electrochemical systems that are modified to include the optical measurement system, allow the reactants to be generated in situ and facilitate dismantling for cleaning. However, ECL detection systems for the flow injection (FI) and liquid chromatographic techniques have also been reported, as has the immobilization of ECL reagents to develop sensors and probes.[36]

The full potential of analytical ECL has not yet been realized, and further work is needed to overcome the problems posed by electrode fouling, to increase the stability and robustness of immobilized ECL reagents, and to improve flow cell designs, which are complicated by the incorporation of electrodes and an optical measurement device.

4.3 Bioluminescence

Bioluminescence is a type of CL in which visible radiation is emitted from living organisms or systems derived therefrom. BL reactions involve luciferase-catalyzed oxidation of luciferins, where luciferase is the enzyme and luciferin the substrate. The systems most widely used for analytical purposes involve firefly and bacterial luciferase.[33] The high sensitivity associated with BL detection has been exploited in the design of analytical methods with excellent detection limits. The basic kinetics of BL analysis, use of coupled enzyme reactions with a final luminescence step, amplification of luminescent reactions by enzymatic cycling and the use of BL in the assay of enzyme activity have been discussed.[37,38]

One recent application of kinetic methodology to BL analysis is the automatic determination of urea by using ATP-hydrolyzing urease, firefly luciferase and luciferin.[39] Light intensity was measured at 40 and 80 s after addition of the urease reagent. Also, a kinetic assay of glycerol for lipolysis studies was developed[40] by using photinus-luciferin 4-monooxygenase, luciferin and glycerol kinase. BL intensity was measured every 0.2 min for 1.2 min.

5 IMMUNOASSAY

Different luminescence techniques have been applied to the determination of species by using immunochemical reactions that have given rise to a number of homogeneous and heterogeneous immunoassays. Homogeneous immunoassay is known to be simpler and faster than heterogeneous immunoassay as it involves none of the incubation, separation and wash steps of the latter, but detection limits are generally higher, mainly as a result of the sample matrix background.

Many homogeneous immunoassays involving fluorimetric detection use fluorescein (λ_{ex} 492 nm, λ_{em} 517 nm) as a label. This compound has a narrow Stokes' shift, which can pose light-scattering problems; also, its fluorescence emission can be compromised with the background fluorescence from the sample matrix. However, this problem can be overcome by using fluorophores such as oxazines, thiazines or cyanines, which give emission bands more red-shifted than that of fluorescein, and measuring the initial rate of the immunochemical reaction in order to accomplish spectral and temporal discrimination, respectively, of the analytical signal as an alternative to the physical separation required in heterogeneous immunoassay. These compounds also exhibit a short fluorescence lifetime, which decreases the likelihood of nonradiative quenching processes. Despite their potential, the utility of these compounds as hapten labels in immunoassay has scarcely been studied because the covalent labeling efficiency of oxazine and thiazine dyes with carboxylic acid-containing analytes is low. However, the efficiency can be improved by using a bifunctional reagent such as a carbodiimide derivative. The combined use of kinetic methodology and an oxazine label, namely Cresyl Violet, has enabled the determination of 2,4,5-trichlorophenoxyacetic acid, a herbicide used to control shrubs and trees, in liquid food samples.[41] The high initial rate of the immunochemical reaction was measured by using the SF mixing technique.

One other homogeneous technique is fluorescence polarization immunoassay (FPIA), which is a competitive binding assay where the tracer (the antigen conjugate) binding with the antibody is directly measured without the need for a separation procedure. It relies on the fact that, when a molecule is excited by polarized light, the emitted light will also be polarized, provided the molecule does not rotate during the time elapsed between excitation and emission. Small fluorescent molecules rotate rapidly, and normally exhibit no fluorescence polarization; on binding to macromolecules, however, the rotation motion is slowed down and the fluorescence remains polarized. Thus, the tracer, which competes with the analyte for the antibody, will only exhibit polarized fluorescence (which will be inversely proportional to the analyte concentration) when bound to the antibody.

Fluorescence polarization is usually expressed as the degree of polarization P (Equation 15):

$$P = \frac{(A/B) - 1}{(A/B) + 1} \quad (15)$$

where A and B are the fluorescence intensities measured with the emission polarizer parallel and perpendicular, respectively, to the excitation polarizer. A variety of electronic and optical assemblies have been used[42] to perform fluorescence polarization measurements. Conventional FPIA involves measuring P when the formation of the tracer–antibody complex has reached or is close to equilibrium. However, measuring the variation of P with time (dP/dt) allows one to obtain the reaction rate of the tracer–antibody complex, which is inversely proportional to the analyte concentration. Kinetic methodology avoids or minimizes the static signal from the sample matrix, which is usually the main cause of the relatively high detection limits afforded by conventional FPIA. This has been demonstrated in the determination of therapeutic[43] and abuse[44] drugs, and also of atrazine.[45] Table 1 compares the detection limits obtained for some of these compounds by using equilibrium and kinetic measurements in FPIA.

The only requirement for application of kinetic methodology in FPIA is the availability of instrumentation enabling simultaneous measurements of the variation of A and B with time to be processed by a microcomputer in order to construct a polarization–time kinetic curve. A T-format spectrofluorimeter with two emission channels symmetrically arranged on both sides of the sample compartment and with the emission polarizers placed perpendicular to each other is a suitable choice (Figure 6). Conventional kinetic measurements (batch technique) are unsuitable in FPIA because the initial rate of the competitive antigen–antibody reaction is usually very high, but it can be obtained by using the SF mixing technique.

Both homogeneous and heterogeneous immunoassays have been developed using CL detection. Similarly to other immunoassay techniques, the CL homogeneous format is technically more simple, but it is subject to problems

Table 1 Comparison of detection limits obtained by using conventional and kinetic measurements in FPIA

Compound	Detection limit ($\mu g\,L^{-1}$)	
	Conventional FPIA	Kinetic FPIA
Nordiazepam	40	2.5
Imipramine	20	7.5
Morphine	20	6.0
D,L-Amphetamine	90	7.0
Benzoylecgonine[a]	30	5.0
11-nor-Δ^8-THC-COOH[b]	10	3.0
Atrazine	10	0.2

[a] A cocaine metabolite.
[b] 11-nor-Δ^8-tetrahydrocannabinol-9-carboxylic acid (a cannabinoid metabolite).

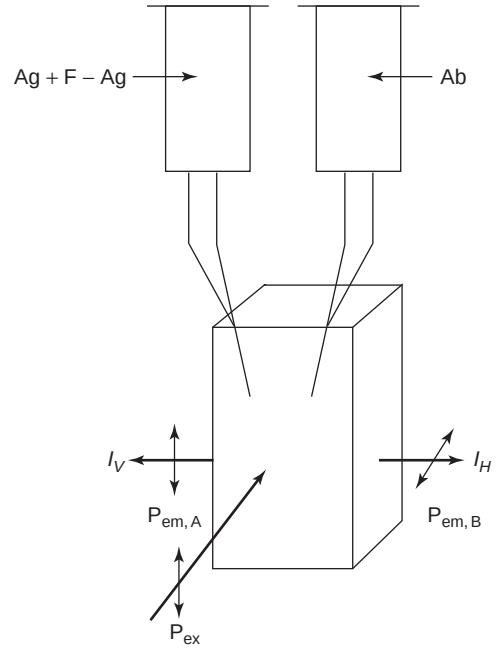

Figure 6 Scheme of the distribution of reactants in the SF mixing system and of the position of the polarizers in the T-format spectrofluorimeter. (Ag) analyte; (F–Ag) tracer; (Ab) antibody; (P_{ex}) excitation polarizer; (P_{em}) emission polarizer; (I_V and I_H) fluorescence intensities measured with the emission polarizer placed parallel and perpendicular, respectively, to the excitation polarizer.

arising from endogenous background interference, which are avoided in the CL heterogeneous format. As in other CL methods, the analyte concentration is usually related to the maximum emission intensity or the integrated light output over a fixed fraction of the total reaction time, although it has also been reported that the precision can be improved by measuring the rate of intensity change over a wide interval after mixing.[46]

A variety of reactions have been evaluated and shown to be suitable for use in CL immunoassay. Indeed, hundreds of reports describing such reactions have appeared in the literature.[47–49] CL immunoassays can be classified into three groups, as follows.

1. Those where the label can be a CL or BL substance, such as luminol or the enzyme luciferase. Typical problems with them include a considerable loss of CL efficiency and enzyme activity, respectively, on linking to other molecules such as antigens or antibodies.
2. Those where the label can be a catalyst or a cofactor for a CL reaction. One case in point is peroxidase, which is a catalyst for the luminol–peroxide reaction, and nicotinamide adenine dinucleotide (NAD), which is a cofactor for the BL reaction catalyzed by bacterial luciferase–oxidoreductase.

3. Products from labels that can be monitored using a luminescent reaction. One example is dehydrogenase enzyme labels that produce the reduced form of nicotinamide adenine dinucleotide (NADH). The product is monitored using the bioluminescent bacterial luciferase–oxidoreductase reaction.

6 AUTOMATION

Similarly to other kinetic determinations, automation has played a crucial role in luminescence-based kinetic determinations by facilitating application to routine analyses. For instance, automatic luminometers for BL determinations are now commercially available and can measure the maximum intensity or integrate the light output over a selection of fixed times. Some such luminometers have been specifically designed for measurement of ATP or nicotinamide adenine dinucleotide phosphate (reduced form) (NAD(P)H). This section briefly reviews the different automatic kinetic approaches based on luminescence detection, namely FI, SF and continuous addition of reagent (CAR).[50,51] Applications of these methods to drug analysis were recently reviewed.[52]

6.1 Flow Injection

Flow injection has proved extremely useful for luminescence-based kinetic determinations. The automated sample handling, controlled sample dispersion and precise time aspects of this approach are distinct advantages.[53] Among FI modes, SF FI[54] and sequential injection[55] are specially useful for kinetic-based determinations. In the SF FI mode, after a preset delay time, the pump is stopped and then re-started once the SF period has elapsed. This allows a selected portion of the dispersed sample zone to be arrested in the observation field of the detector. As physical dispersion virtually ceases when the carrier stream is stopped, the variation of the signal with time reflects the kinetics of the chemical reactions between the mutually dispersed components of the injected sample and carrier stream. This approach allows the reaction rate to be measured as it is not affected by the static background signal from the sample. Figure 7 shows a schematic diagram of the system (a), a portion of the dispersed sample zone (b) and the variation of the signal with time at three different analyte concentrations (c). The system is furnished with a timer (T) that is started when valve S is switched to its injection position.

The SF FI mode differs from the normal SF technique, which is described in the next section, in that (a) sample and reagent are not mixed at a high pressure, (b) the former does not allow fast reactions to be monitored and (c) real-time analyses (e.g. in an enzymatic reaction) are more difficult to accomplish because of the continuous nature of the flow. An attempt at overcoming the pitfalls of this technique was made by developing an automated micro SF continuous apparatus for serial measurements on enzyme reactions.[56] It uses two symmetrically arranged sets of sample injectors (FI system) and stops the carrier stream (SF system), which circulates at a high flow-rate, by means of a high-speed peristaltic pump. Equipped with suitable detectors, the apparatus was used as a real-time analyser for column chromatography.

Sequential injection[55] is a second-generation SF technique based on the sequential aspiration of sample and reagent zones through a selector valve into a holding

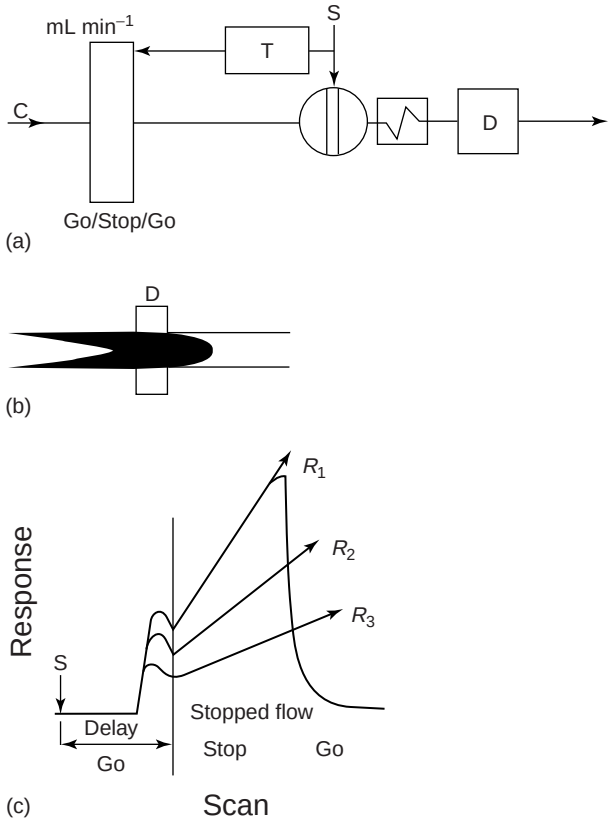

Figure 7 (a) Reaction rate measurements based on the SF method are performed in a manifold furnished with a timer, T, which is activated when valve S is turned to the injection position. After a pre-selected delay time, the pump is stopped and then re-started after the SF period has elapsed. This allows a selected portion of the dispersed sample zone (shown schematically in (b)) to be arrested to the observation field of the detector. (c) The resulting peak is composed of the following three sections: continuous flow showing the peak maximum as recorded during the delay period; the reaction rate curve recorded during SF; and the washout tail. Three injections with increasing analyte concentrations are shown, recorded from identical starting points and showing increasing reaction rates R_1, R_2 and R_3. (Reproduced by permission of the Royal Society of Chemistry, from Ruzicka and Christian.[54])

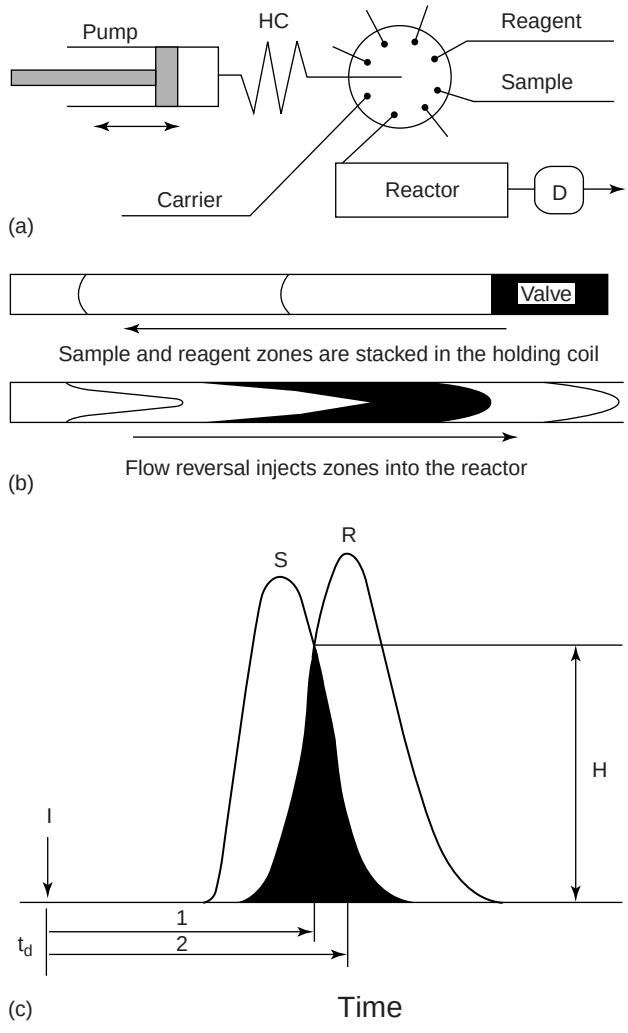

Figure 8 Sequential injection principle. (a) Flow scheme. (b) Structure of the stacked and injected zones. (c) Concentration profiles seen by the detector. (HC) holding coil; (D) detector; (S) sample; (R) reagent; (H) peak height; (I) injection point. (Reproduced by permission of Elsevier, from Ruzicka.[55])

coil, as shown in Figure 8. A stack of well-defined zones is thus obtained. The flow direction is then reversed, and the zones are injected into a reactor and detector. The flow reversal creates a composite zone in which sample and reagent zones penetrate each other. For kinetic determinations, the flow is then stopped and a portion of the composite zone is trapped inside the flow cell of the detector, which allows measurement of the reaction rate.

A procedure for obtaining kinetic data from FI peak profiles by correcting for the effects of dispersion was developed by Chung and Ingle.[57] The method acquires and stores the peak profiles of a nonreacting reference solution and a reacting sample solution. The dispersion coefficients obtained from the reference profile are used to normalize the profile of the sample solution. The utility of this method was demonstrated by adapting a kinetic fluorimetric method for the determination of aluminum with acid alizarin garnet R[58] and using the fixed-time method.

As noted in section 4.1, the shortcomings of batch luminometers can be circumvented by using FI systems. FI CL methods are simplified by reducing the number of reagents that need to be pumped separately to the reaction site, as reagents can be electrochemically produced from passive precursors in the carrier medium. A typical instrumental set-up that was applied to morphine determination[59] is shown in Figure 9. The sample is injected into one flowing stream and the CL reagent flows in the other stream. The two streams meet head-on at a T-piece, inside a light-tight casing, and then flow through a flat coil placed immediately in front of the detector. The coil serves to retain the solution in view of the detector whilst it is emitting most intensely. The system provides very rapid, reproducible mixing and emission intensities, and affords a high throughput.

One other choice in CL analysis is the use of immobilized systems. In this case, the CL intensity depends on both the kinetics of the CL reaction and on the efficiency of mass transfer processes involved in bringing the reactants into contact. Similarly to the use of enzymes, the principal advantages of BL reagents (e.g. firefly or bacterial luciferase) immobilized on solid supports are their reusability, which can significantly reduce reagent costs and improve stability, sensitivity and compatibility with automated continuous flow analyzers.[60] The use of ECL as detection system for LC and FI has given rise to the development of various types of ECL flow cells.[36]

6.2 Stopped-flow

The SF mixing technique is the automatic approach based on kinetic methodology most commonly used when fast reactions are involved. In applying this technique,

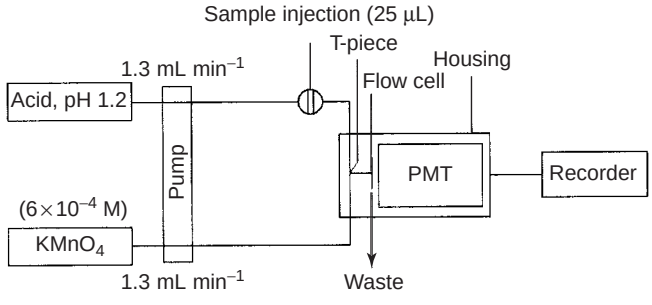

Figure 9 Flow injection manifold for morphine determination with chemiluminescence detection. (Reproduced by permission of the Royal Society of Chemistry, from Abbot et al.[59])

two reactants are driven at a high speed by means of two drive syringes through a mixing chamber, the flow of the reactants being abruptly stopped by using a third, stop syringe, and the extent of reaction being monitored by measuring a given property such as the luminescence emission. The most salient features of the SF technique are its ability to mix sample and reagent solutions automatically, the possibility of making measurements shortly after mixing, a high overall precision and minimization of potential interferences. Although this technique is specially suitable for fast reactions (it facilitates the study of the kinetics and mechanisms of these reactions and their use in developing kinetic determination methods), it has also proved useful for developing quantitative methods involving slower reactions as it minimizes reactant manipulation and affords a higher precision than do conventional kinetic approaches. These features make the SF technique particularly attractive to routine laboratories, which require automated instrumentation, high throughputs and the ability to use low solution volumes.[61]

The SF approach shows attractive features when applied to CL monitoring. Thus, mixing is very rapid, the emitting solution may be retained in the measuring cell for whatever time is desired and, unlike the FI approach, the intensity–time curve can be monitored, thus allowing kinetic measurements to be made (see Figure 10). Formation and decay rates of CL have been found to provide better precision and selectivity than conventional parameters such as the peak high intensity or the area under the emission light–time curve.[32] This approach was tested for the first time on the determination of hydrogen peroxide by its classical reaction with luminol catalyzed by cobalt,[32] and has proved useful for other determinations such as penicillins[62] and manganese.[63]

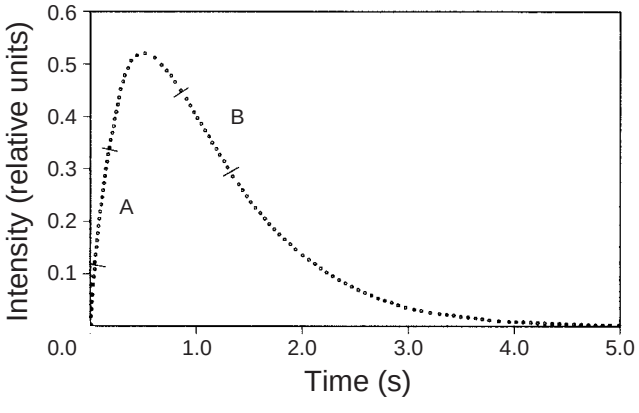

Figure 10 Intensity/time profile obtained in the SF determination of hydrogen peroxide. Zones A and B allow the formation and decay rates, respectively, to be evaluated. (Reproduced by permission of Elsevier, from González-Robledo et al.[32])

The potential of the SF mixing technique in combination with MSRTP was demonstrated by developing a kinetic method for carbaryl determination.[24] Compared with conventional MSRTP, where a phosphorescent system takes a relatively long time to reach equilibrium, the special features of the SF mixing technique allow the fast acquisition of kinetic data and provide a very simple means of accomplishing automation in phosphorescence routine analyses. This approach also avoids the two major constraints of solid-surface room-temperature phosphorimetry (SSRTP), namely the presence of a background signal from the solid supports used, which affects detection limits, and the critical preparation and measurement requirements, which influence the precision of the analytical results.

The term ''SF-fluoroimmunoassay'' has been applied to the use of this mixing technique in homogeneous determinations based on immunochemical reactions. This approach was used for the determination of the pesticide 2,4-dichlorophenoxyacetic acid in fruit juice samples.[64] The system was characterized by the very fast initial rate of the antigen–antibody reaction and the fluorescence quenching of the tracer (fluorescein-labeled pesticide) upon reaction with the antibody. The signal obtained at equilibrium was severely affected by the presence of the sample matrix, which is very common in fluorescence quenching methods, since the static signal is highly sensitive to the environmental conditions. This shortcoming, however, was avoided by measuring the initial rate of the system. SF-immunoassay was also applied to the development of a direct homogeneous method for the determination of coproporphyrin[65] in urine, based on the native fluorescence of this porphyrin and the use of the initial rate of the antigen–antibody reaction as analytical parameter, which was obtained by measuring the decrease in fluorescence intensity with time. The method is simpler than the competitive immunoassay, where a tracer is used, and avoids the separation step required in heterogeneous immunoassay.

6.3 Continuous Addition of Reagent

The CAR technique is a fairly recently reported open system approach based on the continuous addition of a reagent solution, at a constant rate, to another solution containing the species to be determined.[66] Analytical data obtained from second-order reactions are processed by using a mathematical treatment that relates kinetic parameters with the experimental variables. This treatment also allows the second-order rate constant of the system to be determined. The chief advantages of this technique are that the kinetic response curve can be readily obtained from both slow and fast reactions (for which the first-order condition is commonly used)

with straightforward instrumentation since the reaction is made second-order.[67] The typical instrumental set-up consists of: (1) an addition unit with an autoburette, a magnetic stirrer and a thermostated vessel (the reaction cell); (2) a detection system such as a luminescence detector; and (3) a computer unit for data acquisition and processing. The second-order kinetic profile of the signal versus time (or volume added) response includes a straight portion, the slope of which is proportional to the analyte concentration (maximum-rate method); this is equivalent to the initial-rate method used for pseudo first-order kinetic responses in the SF technique.

The CAR technique has been applied to the fluorimetric determination of morphine in urine[68] using the classical oxidative dimerization of this drug by alkaline potassium hexacyanoferrate(III) to form a highly fluorescent dimer (pseudomorphine). The conventional method poses problems arising from instability of the reaction products and undesirable interactions between them and the excess of reagent, which quench the fluorescence of pseudomorphine. However, the CAR technique avoids these problems since the reagent is continuously added from the autoburette, thereby increasing the fluorescence intensity.

The adaptation of the CAR technique to CL reactions has allowed some problems of the analytical application of these reactions to be solved. Thus, this approach is an effective means for suppressing background emission in peroxyoxalate CL reactions, which is ascribed to the formation and decomposition of CL reaction intermediates. This ability arises from the special way in which the sample and reagent are mixed in the CAR technique, in addition to its kinetic connotations. The analytical utility of the combination of the CAR technique with CL reactions has been demonstrated by applying the *bis*(2,4,6-trichlorophenyl)oxalate (TCPO)/hydrogen peroxide system to the determination of acepromazine in horse plasma samples.[69] The CAR technique has also been used as an effective means for enhancing the performance of energy transfer CL-based determinations involving the use of a sensitizer. Thus, the reaction between tertiary aliphatic amines and sodium hypochlorite in an alkaline medium containing fluorescein as sensitizer has been used to determine trimethylamine in fish samples.[70]

7 SIMULTANEOUS DETERMINATIONS

Although classical differential kinetic methods have been applied in simultaneous determinations with luminescent detection, recent investigations have chiefly been aimed at accomplishing mathematical discrimination by using complex chemometric methods or instrumental discrimination by use of multidetection devices. Procedures that exploit both kinetic and spectral differences have thus been developed. Also, automation is a general trend in the latest simultaneous determinations.

Among classical approaches, the proportional-equation method has been applied in conjunction with the SF mixing technique to the simultaneous determination of salicylic acid and diflunisal[30] by the combined use of kinetic and equilibrium measurements. The similar structures of these compounds result in mutual interferences with the determination of each other. Thus, the corresponding ternary anionic chelates formed with terbium(III) and ethylenediaminetetraacetic acid (EDTA) in an alkaline medium exhibit the same initial rate and equilibrium luminescence intensity. However, the presence of cetyltrimethylammonium bromide (CTAB) increases both parameters to a different extent, which has been ascribed to the formation of ion-association complexes between CTAB and the ternary anionic chelates formed by salicylate and diflunisal with terbium(III) and EDTA in the alkaline medium. Epinephrine and norepinephrine were simultaneously determined in urine[71] by using the trihydroxyindole method, which was simplified by implementing the SF technique.

Micellar media provide additional advantages that have scarcely been explored in simultaneous kinetic analyses with luminescence detection. Micelles can alter the apparent rate constant of two or more species with similar structure, interacting with a common reagent by both altering their intrinsic reactivity and, more generally, binding to a different extent to analytes. This effect has been applied to the simultaneous determination of pyridoxal and pyridoxal-5′-phosphate by using their cyanide-catalyzed oxidation in the presence of CTAB micelles.[72] The initial rate and the fluorescence intensity increment of these systems were used as analytical parameters and data were processed by using a system of two equations which was solved by multiple linear regression.

The logarithmic extrapolation method was applied to the simultaneous determination of thiamine and its pyrophosphate ester, cocarboxylase,[73] using the differential oxidation rate of these compounds with mercury(II).

The T-format spectrofluorimeter described above, furnished with an SF module, was used for the direct resolution of a mixture of two drugs, neomycin and promethazine (an antimicrobial and an antihistamine respectively), which are found together in some pharmaceutical preparations.[74] This was the first reported application of kinetic methodology based on reaction-rate measurements to a simultaneous fluorimetric determination using a dual-channel instrument. As the T-format configuration has two emission paths symmetrically

arranged on both sides of the sample compartment, it allows one to simultaneously monitor the development of the reaction of neomycin with *o*-phthaladehyde in the presence of *N*-acetylcysteine and the oxidation of promethazine by dissolved oxygen without the need for an oxidant such as hydrogen peroxide, since both reaction products have similar excitation wavelengths but different emission wavelengths. The absence of overlap in the emission maxima allows one to obtain kinetic data from each individual system, even if both possess similar rate constants and their reactions are fast. In this instance, both systems exhibited a disparate kinetic behavior: whereas neomycin reached equilibrium within only 3–4 s, promethazine took about 40 s. However, their high initial rates justified the use of the SF mixing technique. The method allowed the resolution of mixtures containing neomycin and promethazine in ratios between 8 : 1 and 1 : 9, and was applied to pharmaceutical analyses.

This approach was also applied to the simultaneous determination of the antibiotics ampicillin and tetracycline in milk.[75] The determination of ampicillin was based on its hydrolysis with penicillinase and reaction with mercuric chloride; that of tetracycline relied on the intramolecular energy transfer from the analyte to europium ions in the presence of thenoyltrifluoracetone. Owing to the differential kinetic behaviour of the two chemical systems involved, kinetic and equilibrium measurements were carried out to obtain quantitative data for penicillin and tetracycline, respectively.

The flexibility and simplicity of the FI approach has been exploited in the development of a number of differential kinetic methods. Thus, the differential catalytic activity of manganese(II) and iron(III) ions on the oxidation of salicylaldehyde thiosemicarbazone by hydrogen peroxide has been used for the simultaneous determination of these ions[76] with an FI configuration that splits the sample into two channels with different geometric and hydrodynamic characteristics and then recombines both channels prior to the detector. The differential residence times for the sample portions result in two peaks in which the contribution of each ion to the formation of the oxidation product is different.

Regarding chemometric methods, the Kalman filtering algorithm has been applied to the simultaneous time-resolved determination of antibiotics such as tetracyclines[77] and anthracyclines[78] in solution by formation of luminescent chelates with europium(III). An optosensing FI manifold involving a flow cell packed with a nonionic resin was used in both instances. Each determination utilized the differences in decay rates of the chelates adsorbed onto the resin and the Kalman filter to process decay data.

Artificial neural networks[79] have been used in combination with the CAR technique to resolve phenothiazine mixtures[80] by using CL reactions with a peroxyoxalate derivative. The CAR technique suppresses background emission from these CL reactions. The chemometric method allows the simultaneous determination of trimeprazine and methotrimeprazine, even though both compounds exhibited a similar kinetic behavior and synergism. Luminol CL induced by hexacyanoferrate(III) and enhanced by phenolic acids was used for the simultaneous determination of protocatechuic and caffeic acids.[81] The synergistic effects present required the use of surface response and least-squares matrix methodologies.

Kinetic fluorescence detection has been utilized for the determination of glycine and glutamine (previously partially separated on a thin-layer chromatography (TLC) plate).[82] The formation and decay of the intermediate fluorescent products formed between these species and *o*-phthalaldehyde was monitored with a charged-coupled device camera by taking sequential images of the TLC plate. A direct trilinear decomposition method was used to analyze the resulting third-order data, which consisted of fluorescence intensities as a function of elution distance, reaction time and sample number.

Table 2 Selected examples of luminescence-based kinetic determinations

Analytes	Luminescence process	Remarks	Refs.
Enzymes and enzymatic substrates	F	Many sensitive methods	14, 15
Inorganic ions	F	Catalysis, activation or inhibition	5, 6, 16, 17, 76
	CL	Catalysis, activation or inhibition	25, 26, 63
	P	Temporal resolution	36
Amino acids	F	Kalman filter	22
	F	Simultaneous determinations	82
Proteins	F	Long wavelength	19, 20
Pharmaceuticals	F	Immunoassay	43
	LSL	Simultaneous determinations	30, 75, 77, 78
	CL	SF	62
	F	ANN	80
Abused drugs	F	Immunoassay	44
	CL	FI	59
	F	CAR	68
Pesticides	F	Immunoassay	41, 45, 64
	P	MSRTP	24

F, fluorescence; P, phosphorescence; LSL, lanthanide-sensitized luminescence; ANN, artificial neural networks.

8 FINAL REMARKS

The analytical interest of luminescence-based kinetic determinations is the result of the high sensitivity and versatility of luminescence techniques together with the temporal discrimination of the analytical signal that is obtained by using kinetic methodology. Table 2 summarizes some selected examples of these determinations, which shows the wide application field of this approach.

ACKNOWLEDGMENTS

The author gratefully acknowledges financial support from Spain's DGICyT (Project PB96-0984).

LIST OF SYMBOLS

k_F	Fluorescence Rate Constant
k_I	Rate Constant for Internal Conversion for Lowest Excited Singlet State
k_{IS}	Rate Constant for Intersystem Crossing
k_M	Rate Constant for Intersystem Crossing between Lowest Triplet State and Ground State
k_P	Phosphorescence Rate Constant
k_Q	Rate Constant for Biomolecular Quenching
k_q	Rate Constant for Relaxation Processes other than Phosphorescence Decay
P	Degree of Polarization
ϕ_F	Fluorescence Quantum Yield
ϕ_P	Phosphorescence Quantum Yield
ϕ_T	Triplet Formation Quantum Yield
$[Q]$	Concentration of Quencher
τ_F	Observed Fluorescence Lifetime
τ_{F0}	Intrinsic Fluorescence Lifetime
τ_P	Observed Phosphorescence Lifetime
τ_{P0}	Intrinsic Phosphorescence Lifetime

ABBREVIATIONS AND ACRONYMS

ADC	Analog to Digital Converter
BL	Bioluminescence
CAR	Continuous Addition of Reagent
CL	Chemiluminescence
CRT	Cathode-ray Tube Screen
CTAB	Cetyltrimethylammonium Bromide
ECL	Electrogenerated Chemiluminescence
EDTA	Ethylenediaminetetraacetic Acid
FI	Flow Injection
FMN	Flavin Mononucleotide
FPIA	Fluorescence Polarization Immunoassay
LC	Liquid Chromatography
MSRTP	Micelle-stabilized Room-temperature Phosphorimetry
NAD	Nicotinamide Adenine Dinucleotide
NADH	Reduced Form of Nicotinamide Adenine Dinucleotide
NAD(P)H	Nicotinamide Adenine Dinucleotide Phosphate (Reduced Form)
PMT	Photomultiplier Tube
PT	Reference Vacuum Phototube
SDS	Sodium Dodecyl Sulfate
SF	Stopped-flow
SSRTP	Solid-surface Room-temperature Phosphorimetry
TCPO	*bis*(2,4,6-Trichlorophenyl)oxalate
TLC	Thin-layer Chromatography

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Flow-injection Techniques in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Luminescence in Environmental Analysis

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis • Proteins, Peptides, and Amino Acids Analysis in Food

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis • Phenoxy Acid and Other Acidic Pesticides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Pharmaceuticals and Drugs **(Volume 8)**
Antibiotics, Pharmaceutical Analysis of

Electroanalytical Methods **(Volume 11)**
Chemiluminescence, Electrogenerated

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence in Organized Assemblies • Fluorescence Lifetime Measurements, Applications of • Near-infrared Absorption/Luminescence Measurements • Phosphorescence Measurements, Applications of

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Differential Rate Determinations • Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics • Uncatalyzed Kinetic Determinations

REFERENCES

1. H. Theorell, A. Nygaard, 'Kinetics and Equilibria in Flavoprotein Systems, 1. A Fluorescence Recorder and its Application to a Study of the Dissociation of the Old Yellow Enzyme and its Resynthesis from Riboflavin Phosphate and Protein', *Acta. Chem. Scand.*, **8**, 877–888 (1954).
2. J.D. Ingle, Jr, M.A. Ryan, 'Reaction Rate Methods in Fluorescence Analysis', in *Modern Fluorescence Spectroscopy*, ed. E.L. Wehry, Plenum Press, New York, Vol. 3, 1981.
3. H.B. Mark, G.A. Rechnitz, 'Kinetics in Analytical Chemistry', Interscience, New York, 1968.
4. K.B. Yatsimirskii, 'Kinetic Methods of Analysis', Pergamon Press, Oxford, 1966.
5. D. Pérez-Bendito, M. Silva, 'Kinetic Methods in Analytical Chemistry', Ellis Horwood, Chichester, 1988.
6. H.A. Mottola, 'Kinetic Aspects of Analytical Chemistry', Wiley-Interscience, New York, 1988.
7. J.R. Lakowicz, 'Principles of Fluorescence Spectroscopy', Plenum Press, New York, 1983.
8. J.N. Demas, 'Excited-state Lifetime Measurements', Academic Press, New York, 1983.
9. R.J. Hurtubise, 'Phosphorimetry: Theory, Instrumentation, and Applications', VCH, New York, 1990.
10. J.R. Lakowicz, 'Topics in Fluorescence Spectroscopy', Plenum Press, New York, Vol. 1, 1991.
11. N. Boens, 'Pulse Fluorometry', in *Luminescence Techniques in Chemical and Biochemical Analysis*, eds. W.R.G. Baeyens, D.D. Keukeleire, K. Korkidis, Marcel Dekker, New York, 1991.
12. A.K. Campbell, 'Chemiluminescence: Principles and Applications in Biology and Medicine', Ellis Horwood, Chichester, 1988.
13. L.J. Cline Love, M. Skrilec, J.G. Habarta, 'Analysis by Micelle-stabilised Room-temperature Phosphorescence in Solution', *Anal. Chem.*, **52**, 754–759 (1980).
14. G.G. Guilbault, 'Practical Fluorescence', 2nd edition, Marcel Dekker, New York, 1990.
15. A.G. Hadd, D.E. Raymond, J.W. Halliwell, S.C. Stephen, J.M. Ramsey, 'Microchip Device for Performing Enzyme Assays', *Anal. Chem.*, **69**, 3407–3412 (1997).
16. M. Valcárcel, F. Grases, 'Fluorimetric Reaction-rate Methods of Inorganic Analysis: A Review', *Talanta*, **30**, 139–143 (1983).
17. T. Kawashima, S. Nakano, M. Tabata, M. Tanaka, 'Catalytic Reactions as Applied to Trace Analysis', *Trends Anal. Chem.*, **16**, 132–139 (1997).
18. G.L. Campli, D.J. Ingle, Jr, 'Monitoring and Compensating for Between-run Temperature and pH Variations in Kinetic-based Determinations', *Anal. Chim. Acta*, **224**, 275–287 (1989).
19. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Fluorimetric Determination of Gliadins in Foods', *Analyst*, **121**, 1133–1136 (1996).
20. M.P. Aguilar-Caballos, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Determination of Casein in Milk and Dairy Products by Long-wavelength Fluorescence Detection', *J. Agric. Food Chem.*, **46**, 4250–4254 (1998).
21. S.D. Brown, 'The Kalman Filter in Analytical Chemistry', *Anal. Chim. Acta*, **181**, 1–26 (1986).
22. M. Gul, S.C. Rutan, 'Determination of Initial Concentration of an Analyte by Kinetic Detection of the Intermediate Product in Consecutive First-order Reactions Using an Extended Kalman Filter', *Anal. Chem.*, **66**, 1513–1519 (1994).
23. J.A. Kelly, K.R. Reddy, C.S. Lee, 'Mechanistic Studies of Postcapillary Affinity Detection for Capillary Zone Electrophoresis Based on Biotin–Streptavidin System', *Anal. Chem.*, **69**, 5152–5158 (1997).
24. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Usefulness of the Stopped-flow Mixing Technique for Micelle-stabilized Room-temperature Liquid Phosphorimetry', *Anal. Chem.*, **66**, 919–923 (1994).
25. R. Brina, A.G. Miller, 'Direct Detection of Trace Levels of Uranium by Laser-induced Kinetic Phosphorimetry', *Anal. Chem.*, **64**, 1413–1418 (1992).
26. M.A. Hedaya, H.P. Birkenfeld, R.L. Kathren, 'A Sensitive Method for the Determination of Uranium in Biological Samples Utilizing Kinetic Phosphorescence Analysis (KPA)', *J. Pharm. Biom. Anal.*, **15**, 1157–1165 (1997).
27. E.P. Diamandis, T.K. Christopoulos, 'Europium Chelate Labels in Time-resolved Fluorescence Immunoassays and DNA Hybridization Assays', *Anal. Chem.*, **62**, 1149A–1157A (1990).
28. J. Georges, 'Lanthanide-sensitized Luminescence and Applications to the Determination of Organic Analytes', *Analyst*, **118**, 1481–1486 (1993).
29. Y.X. Ci, Y.Z. Liu, 'Selective Determination of DNA by its Enhancement Effect on the Fluorescence of the Eu^{3+}–Tetracycline Complex', *Anal. Chem.*, **67**, 1785–1788 (1995).
30. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Determination of Salicylic Acid, Diflunisal and their Mixture Based on Lanthanide-sensitized Luminescence', *Anal. Chim. Acta*, **329**, 135–141 (1996).
31. S. Panadero, A. Gómez-Hens, D. Pérez-Bendito, 'Kinetic Determination of Propyl Gallate in Edible and Cosmetic Oils with Sensitized Terbium(III) Luminescence Detection', *Analyst*, **120**, 125–128 (1995).

32. D. González-Robledo, M. Silva, D. Pérez-Bendito, 'Performance of the Stopped-flow Technique in Chemiluminescence Spectrometry Based on Direct Rate Measurements', *Anal. Chim. Acta*, **217**, 239–247 (1989).
33. K. Robards, P.J. Worsfold, 'Analytical Applications of Liquid-phase Chemiluminescence', *Anal. Chim. Acta*, **266**, 147–173 (1992).
34. T.A. Nieman, 'Analytical Chemiluminescence in Flow Systems: FIA and HPLC', in *Luminescence Techniques in Chemical and Biochemical Analysis*, eds. W.R.G. Baeyens, D.D. Keukeleire, K. Korkidis, Marcel Dekker, New York, 1991.
35. G.M. Greenway, 'Analytical Applications of Electrogenerated Chemiluminescence', *Trends Anal. Chem.*, **9**, 200–203 (1990).
36. A.W. Knight, G.M. Greenway, 'Occurrence, Mechanisms and Analytical Applications of Electrogenerated Chemiluminescence. A Review', *Analyst*, **119**, 879–890 (1994).
37. S.E. Brolin, P. Naeser, G. Wettermark, 'Design and Coupled Reactions for Simplification of Bioluminescence Analysis', *J. Biolumin. Chemilumin.*, **4**, 446–453 (1989).
38. S.E. Brolin, G. Wettermark, 'Analytical Applications of Luminescence – a Matter of Proper Kinetic Design and Recording', *J. Biochem. Biophys. Methods*, **22**, 1–18 (1991).
39. B. Naslund, L. Stahle, A. Lundin, B. Anderstam, P. Arner, J. Bergstrom, 'Luminometric Single Step Urea Assay Using ATP-Hydrolyzing Urease', *Clin. Chem.*, **44**, 1964–1973 (1998).
40. J. Hellmer, P. Arner, A. Lundin, 'Automatic Luminometric Kinetic Assay of Glycerol for Lipolysis Studies', *Anal. Biochem.*, **177**, 132–137 (1989).
41. M.P. Aguilar-Caballos, A. Gómez-Hens, D. Pérez-Bendito, 'Pesticide Determination by Stopped-flow Fluoroimmunoassay Using Cresyl Violet as Label', *Anal. Chim. Acta*, **381**, 147–154 (1999).
42. M.C. Gutiérrez, A. Gómez-Hens, D. Pérez-Bendito, 'Immunoassay Methods Based on Fluorescence Polarization', *Talanta*, **36**, 1187–1201 (1989).
43. A. Gaikwad, A. Gómez-Hens, D. Pérez-Bendito, 'Use of Stopped-flow Fluorescence Polarization Immunoassay in Drug Determinations', *Anal. Chim. Acta*, **280**, 129–135 (1993).
44. D. Pérez-Bendito, A. Gómez-Hens, A. Gaikwad, 'Direct Stopped-flow Fluorescence Polarization Immunoassay of Abused Drugs and their Metabolites in Urine', *Clin. Chem.*, **40**, 1489–1493 (1994).
45. B. Sendra, S. Panadero, S. Eremin, A. Gómez-Hens, 'Kinetic Determination of Atrazine in Foods Based on Stopped-flow Fluorescence Polarization Immunoassay', *Talanta*, **47**, 153–160 (1998).
46. A. Navas Díaz, F. García Sanchez, J. Lovillo, J.A. González García, 'Enhanced Chemiluminescence Kinetic ELISA of Dichlorprop Methyl Ester', *Anal. Chim. Acta*, **321**, 219–224 (1996).
47. L.J. Kricka, P.E. Stanley, G.H.G. Thorle, T.P. Whitehead, 'Analytical Applications of Bioluminescence and Chemiluminescence', Academic Press, London, 1984.
48. T.T. Ngo, 'Nonisotopic Immunoassay', Plenum Press, New York, 1988.
49. X.R. Zhang, W.R.G. Baeyens, A.M. García-Campaña, J. Ouyang, 'Recent Developments in Chemiluminescence Sensors', *Trends Anal. Chem.*, **18**, 384–391 (1999).
50. A. Townshend, 'Solution Chemiluminescence. Some Recent Analytical Developments', *Analyst*, **115**, 495–500 (1990).
51. M. Silva, 'Recent Strategies in Automated Reaction Rate Based Determinations', *Analyst*, **118**, 681–688 (1993).
52. D. Pérez-Bendito, A. Gómez-Hens, M. Silva, 'Advances in Drug Analysis by Kinetic Methods', *J. Pharm. Biomed. Anal.*, **14**, 917–930 (1996).
53. M. Valcárcel, M.D. Luque de Castro, 'Flow Injection Analysis. Principles and Applications', Ellis Horwood, Chichester, 1987.
54. J. Ruzicka, G.D. Christian, 'Flow Injection Analysis and Chromatography: Twins or Siblings', *Analyst*, **115**, 475–486 (1990).
55. J. Ruzicka, 'Second Coming of Flow-injection Analysis', *Anal. Chim. Acta*, **261**, 3–10 (1992).
56. K. Takahashi, S. Taniguchi, T. Kuroishi, K. Yasuda, T. Sano, 'Automated Micro Stopped-flow/Continuous-flow Apparatus for Serial Measurement of Enzyme Reactions and its Application as a Real-time Analyser for Column Chromatography', *Anal. Chim. Acta*, **220**, 13–21 (1989).
57. H.K. Chung, J.D. Ingle, Jr, 'Kinetic Data from Individual Peak Profiles Corrected for Dispersion in Flow Injection Analysis Applied to the Determination of Aluminum', *Anal. Chem.*, **62**, 2547–2552 (1990).
58. G.L. Campi, J.D. Ingle, Jr, 'Fluorimetric Reaction-rate Method for the Determination of Aluminum with Acid Alizarin Garnet R', *Anal. Chim. Acta*, **224**, 363–372 (1989).
59. R.W. Abbott, A. Townshend, R. Gill, 'Determination of Morphine by Flow Injection Analysis with Chemiluminescence Detection', *Analyst*, **111**, 635–640 (1986).
60. P.J. Worsfold, A. Nabi, 'Immobilized Bioluminescent Reagents in Continuous Flow Systems', *Trends Anal. Chem.*, **5**, 183–185 (1986).
61. A. Gómez-Hens, D. Pérez-Bendito 'The Stopped-flow Technique in Analytical Chemistry', *Anal. Chim. Acta*, **242**, 147–177 (1991).
62. S. Ventura, M. Silva, D. Pérez-Bendito, 'Stopped-flow Chemiluminescence Spectrometry to Improve the Determination of Penicillins Based on the Luminol–Iodine Reaction', *Anal. Chim. Acta*, **266**, 301–307 (1992).
63. A. Gaikwad, M. Silva, D. Pérez-Bendito, 'Selective Stopped-flow Determination of Manganese with Luminol in the Absence of Hydrogen Peroxide', *Anal. Chim. Acta*, **302**, 275–282 (1995).

64. E.G. Matveeva, M.P. Aguilar-Caballos, S.A. Eremin, A. Gómez-Hens, D. Pérez-Bendito, 'Use of Stopped-flow Fluoroimmunoassay in Pesticide Determination', *Analyst*, **122**, 863–866 (1997).
65. E.G. Matveeva, A.P. Savitsky, A. Gómez-Hens, 'Determination of Coproporphyrin in Urine by Using Stopped-flow Fluoroimmunoassay', *Anal. Chim. Acta*, **361**, 27–32 (1998).
66. M. Márquez, M. Silva, D. Pérez-Bendito, 'Direct Rate Measurements on Millisecond Reactions by Continuous Addition of Reagent for Routine Analysis', *Analyst*, **113**, 1733–1736 (1988).
67. A. Velasco, M. Silva, D. Pérez-Bendito, 'Processing Analytical Data Obtained from Second-order Reactions by Using Continuous Reagent Addition', *Anal. Chem.*, **64**, 2359–2365 (1992).
68. J. Cepas, M. Silva, D. Pérez-Bendito, 'Automated Kinetic–Spectrofluorimetric Method for the Determination of Morphine in Urine', *Analyst*, **118**, 923–927 (1993).
69. J. Cepas, M. Silva, D. Pérez-Bendito, 'Improved Peroxyoxalate Chemiluminescence-based Determinations by Use of Continuous Reagent Addition to Remove Background Emission', *Anal. Chem.*, **66**, 4079–4084 (1994).
70. M. Cobo, M. Silva, D. Pérez-Bendito, 'Chemiluminescence Measurement of Tertiary Amines Using the Continuous-addition-of-reagent Technique. Determination of Trimethylamine in Fish Samples', *Anal. Chim. Acta*, **356**, 51–59 (1997).
71. M.P. Llavero, S. Rubio, A. Gómez-Hens, D. Pérez-Bendito, 'Improved Trihydroxyindole Method for the Simultaneous Stopped-flow Spectrofluorimetric Determination of Epinephrine and Norepinephrine in Urine', *Anal. Chim. Acta*, **229**, 27–33 (1990).
72. F. Morales, D. Sicilia, S. Rubio, D. Pérez-Bendito, 'Application of Micellar Effects to the Simultaneous Determination of Pyridoxal and Pyridoxal-5′-Phosphate', *Anal. Chim. Acta*, **345**, 87–98 (1997).
73. V. González, S. Rubio, A. Gómez-Hens, D. Pérez-Bendito, 'Fluorimetric Determination of Thiamine and Cocarboxylase by a Differential Rate Method', *Anal. Lett.*, **21**, 993–1008 (1988).
74. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Use of a Stopped-flow/T-Format Spectrofluorimeter for Simultaneous Kinetic Analyses', *Anal. Chim. Acta*, **310**, 453–459 (1995).
75. B. Gala, A. Gómez-Hens, D. Pérez-Bendito, 'Simultaneous Determination of Ampicillin and Tetracycline in Milk by Using a Stopped-flow/T-Format Spectrofluorimeter', *Talanta*, **44**, 1883–1889 (1997).
76. F. Lázaro, M.D. Luque de Castro, M. Valcárcel, 'Sequential and Differential Catalytic–Fluorimetric Determination of Manganese and Iron by Flow Injection Analysis', *Anal. Chim. Acta*, **169**, 141–148 (1985).
77. F. Alava-Moreno, Y.M. Liu, M.E. Díaz-García, A. Sanz-Medel, 'Kalman Filtering-aided Time-resolved Solid-surface Room Temperature Phosphorimetry for Simultaneous Determination of Tetracyclines in Solution', *Mikrochim. Acta*, **112**, 47–54 (1993).
78. F. Alava-Moreno, M.J. Valencia-González, M.E. Díaz-García, 'Kalman Filtering-aided Time-resolved Solid-surface Room Temperature Phosphorimetry for Simultaneous Determination of Anthracyclines in Solution', *Analyst*, **123**, 691–694 (1998).
79. D. Pérez-Bendito, M. Silva, 'Recent Advances in Kinetometrics', *Trends Anal. Chem.*, **15**, 232–240 (1996).
80. C. Hervás, S. Ventura, M. Silva, D. Pérez-Bendito, 'Computational Neural Network for Resolving Non-linear Multicomponent Systems Based on Chemiluminescence-methods', *J. Chem. Inf. Comput. Sci.*, **38**, 1119–1124 (1998).
81. A. Navas Díaz, J.A. González García, 'Nonlinear Multicomponent Kinetic Analysis for the Simultaneous Stopped-flow Determination of Chemiluminescence Enhancers', *Anal. Chem.*, **66**, 988–993 (1994).
82. M. Gui, S.C. Rutan, A. Agbodjan, 'Kinetic Detection of Overlapped Amino Acids in Thin-layer Chromatography with a Direct Trilinear Decomposition Method', *Anal. Chem.*, **67**, 3293–3299 (1995).

Uncatalyzed Kinetic Determinations

Michael A. Koupparis
University of Athens, Greece

1 Theoretical Background **1**
1.1 Reaction Rate and Kinetic Equation 1
1.2 Determination of Reaction Orders and Rate Constants 2
2 Determination of a Single Component – Principles of Methods and Techniques **4**
2.1 Pseudo-first-order Conditions 4
2.2 Second-order Conditions 5
2.3 Multipoint Approaches 5
2.4 Techniques of Mixing and Monitoring 6
2.5 Detection Systems 7
3 Reaction Types **7**
3.1 Complexation Reactions 7
3.2 Redox Reactions 8
3.3 Coupled Reactions 8
3.4 Addition Reactions 9
3.5 Aromatic Substitution Reactions 9
3.6 Hydrolysis Reactions 10
3.7 Chemiluminescence Reactions 10
4 Analytes Determined **10**
4.1 Determination of Inorganic Species 10
4.2 Determination of Organic Species 10
5 Application Areas **10**
5.1 Clinical Analysis 10
5.2 Environmental Analysis 17
5.3 Food Analysis 17
5.4 Pharmaceutical Analysis 17
Abbreviations and Acronyms **17**
Related Articles **17**
References **18**

Analytical procedures in which the measurement step (chemical, physical or physicochemical) is influenced by a transient (kinetic) process, can strictly be classified as kinetic method. Thus, the majority of modern analytical methods (continuous-flow, time-resolved fluorimetric and chromatographic methods, among others) are kinetic in nature. Nevertheless, the term ''kinetic method'', for historical reasons, is incorrectly limited to methods based on direct or indirect measurements of the rate of a chemical reaction, which should be classified as ''reaction-rate methods''. Analytical methods based on signal measurements made under dynamic conditions compete in efficiency with static and equilibrium measurements. The growing development of kinetic methods is due to the need for quantification of microamounts of substances, increasing knowledge of reaction mechanisms, and breakthroughs in instrumentation and data handling and interpretation. The most popular classification of kinetic methods of analysis is based on the chemistry of reactions employed and distinguishes between catalytic (nonenzymatic, enzymatic and electrochemical), and noncatalytic methods. The latter are further categorized into those used to determine a single species or several components in mixtures (differential reaction-rate methods). This article deals with kinetic methods for the determination of single noncatalytic species. Although the majority of kinetic-based determinations depend on catalytic systems, some analytical applications using the rate of uncatalyzed reactions are of analytical interest, mainly those involving organic chemical species. Advances in instrumentation: (1) improved signal detection, (2) resolving of problems arising from time measurements especially in fast processes, and (3) availability of automated mixing devices and data processing (error-compensated methods, curve-fitting and predictive approaches) establish noncatalytic methods as powerful alternatives to equilibrium methods. The main field of application of noncatalytic reactions is organic analysis, unlike catalytic reactions where a metal ion usually acts as the catalyst.

1 THEORETICAL BACKGROUND[1–3]

1.1 Reaction Rate and Kinetic Equation

The kinetic determination of a noncatalytic species is based on the monitoring of its reaction rate by direct or indirect measurements of changes in the reactant or product concentration. The reaction rate is defined as the change in concentration of one of the components (reactant consumed or product formed) of the chemical system per unit time. From a practical analytical viewpoint, methods using uncatalyzed reaction rates are limited to those following pseudo-zero-order, first-order, and pseudo-first-order kinetics, since most of the chemical systems used involve bimolecular reactions and orders higher than two are rare and of no analytical applicability.

For a straightforward reaction of the type $A + R \longrightarrow P$, where A is the species of analytical interest (analyte), R is the added reagent and P is the product, the general

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

differential rate expression at time t will be given by Equation (1)

$$\text{rate} = \frac{\text{d[P]}}{\text{d}t} = -\frac{\text{d[A]}}{\text{d}t} = -\frac{\text{d[R]}}{\text{d}t} \quad (1)$$

in which the derivatives of [A] and [R] are negative because both species are consumed as the reaction proceeds. The reaction rate at time t depends on the concentrations of all the species involved and the pertinent kinetic equation can be derived as Equation (2)

$$\text{rate} = \frac{\text{d[P]}}{\text{d}t} = k[\text{A}][\text{R}] \quad (2)$$

where k is defined as the rate constant and expresses the reaction rate per unit concentration of the reactants. The reaction order is defined as the sum of the exponents of [A] and [R] in Equation (2), known as partial reaction orders; thus, the above reaction is second order.

In the case of one of the reactants (e.g. R) being present in a large excess (over 50-fold) relative to the other, then its concentration remains practically constant and can be included in the term k in Equation (2) and the reaction is considered to be pseudo-first-order in A or pseudo-zero-order in R, i.e. Equation (3)

$$\text{rate} = \frac{\text{d[P]}}{\text{d}t} = k_{\text{A}}[\text{A}] \quad (3)$$

where k_{A} is the pseudo-first-order rate constant. The experimental parameters for most kinetic methods are adjusted so that pseudo-first-order reaction conditions are established.

Instead of the concentrations in Equation (1) any measurable quantity of analytical signal (S) directly proportional to concentration (absorbance, fluorescence, etc.) can be used. Special equations are derived when the signal is not linearly related to the concentration (e.g. potentiometric measurements). The changes in reactant or product concentrations as a function of time can be monitored physically or chemically. Physicochemical techniques (spectrophotometry, fluorimetry, potentiometry, etc.) are preferable for reaction monitoring and provide a signal versus time curve. In the majority of kinetic methods only the initial section of the curve is used, i.e. when the reaction has only proceeded by 1–3%. This section is usually linear with a slope proportional to the analyte concentration (initial-rate approach).

1.2 Determination of Reaction Orders and Rate Constants

The kinetic study of a reaction is performed using the so-called *continuous* (in situ) or *discontinuous* (batch) approach. In a continuous kinetic study, an analytical parameter (signal) (absorbance, potential, fluorescence, etc.) related to the changing concentration of a reactant or product is continuously monitored by means of an analytical instrument connected to a recorder, printer or microcomputer. In contrast, in a discontinuous study (usually used for rather slow reactions and processes), aliquots of the reacting mixture are removed by the analyst at preselected time intervals and analyzed immediately or all together in a batch mode after quenching (killing) the reaction at the time of sampling. The analytical parameter – time curves can be used for the determination of reaction orders and rate constants, usually after transformation to concentration – time curves.

1.2.1 Determination of Reaction Order

The order of a reaction (or process) may be determined by the following methods.

1.2.1.1 Substitution Method The concentration data obtained in a kinetic study are substituted in the integrated form of the equations that describe the various reaction orders and the k values are calculated. The reaction is considered to be of that order that follows the equation that gives constant k values within the limits of experimental error.

A reaction order of 0 gives an integrated rate equation:

$$x = kt$$

and a half-life equation:

$$t_{1/2} = \frac{\alpha}{2k}$$

A reaction order of 1 gives an integrated rate equation:

$$\log \frac{\alpha}{(\alpha - x)} = \frac{k}{2.303}t$$

and a half-life equation:

$$t_{1/2} = \frac{0.693}{k}$$

A reaction order of 2 for [2A $\longrightarrow$ P] or [A + B $\longrightarrow$ P] with equal concentrations gives an integrated rate equation:

$$\frac{x}{\alpha(\alpha - x)} = kt$$

and a half-life equation:

$$t_{1/2} = \frac{1}{\alpha k}$$

A reaction order of 2 for [A + B $\longrightarrow$ P] with unequal concentrations gives an integrated rate equation:

$$\log \frac{\beta(\alpha - x)}{\alpha(\beta - x)} = \frac{(\alpha - \beta)k}{2.303}t$$

where α and β are the initial concentrations of reactants A and B, respectively, and x is the concentration of each species reacting in a time t.

1.2.1.2 Fitting (Graphical) Method Plots of a function of concentration C of the reactant versus t can also be used to ascertain the reaction order. Thus, if a plot of C versus t results in a straight line, the reaction is zero order. The reaction is first order if the plot of $\log C$ versus t yields a straight line and it is second-order if a plot of $1/C$ versus t results in a straight line (in the case of equal initial concentrations).

1.2.1.3 Half-life Method As shown above, the half-life is proportional to the quantity $1/(\alpha^{n-1})$, in which n is the order of the reaction (for a second-order reaction we assume that $\alpha = \beta$). Thus, if the same reaction is run at different initial concentrations, α_1 and α_2, the half-lives $t_{1/2(1)}$ and $t_{1/2(2)}$ are related according to Equation (4)

$$\frac{t_{1/2(1)}}{t_{1/2(2)}} = \left(\frac{\alpha_2}{\alpha_1}\right)^{n-1} \tag{4}$$

By taking the logarithms and rearranging we have Equation (5)

$$n = \frac{\log[t_{1/2(1)}/t_{1/2(2)}]}{\log(\alpha_2/\alpha_1)} + 1 \tag{5}$$

The half-lives are obtained by plotting C versus t at two different initial concentrations and reading the time at $\alpha_1/2$ and $\alpha_2/2$. By substituting these values and the initial concentrations in Equation (5), the order n is obtained directly. Alternatively, two concentrations during a single run may be taken as α_1 and α_2 and the half-lives determined in terms of α_1 and α_2.

1.2.1.4 Initial Reaction Rate (van't Hoff) Method In the previous methods 1–3 (sections 1.2.1.1–1.2.1.3), the reaction must be completed to a large extent. The so-called initial reaction-rate method consists of measuring the initial reaction rates v_0 for different initial concentrations of a reactant, while the initial concentrations of the other reactants are held constant. From Equation (6)

$$v = \text{rate} = \frac{\mathrm{d[P]}}{\mathrm{d}t} = k[\mathrm{A}]^a[\mathrm{B}]^b[\mathrm{C}]^c \tag{6}$$

it is concluded (since B and C are constant) that Equation (7) follows

$$\log v_0 = \log(\mathrm{d[P]}/\mathrm{d}t)_0 = \text{constant} + a \log [\mathrm{A}]_0 \tag{7}$$

The partial reaction order a, with respect to A, is obtained from the slope of the plot of the logarithms of initial rates versus the logarithms of the initial concentrations. In a similar way the partial reaction orders with respect to B and C can be determined by two further sets of kinetic runs.

In the above methods, instead of the concentration of the species of interest, an analytical parameter (signal) linearly related to the concentration can be used in calculations, since the units cancel.

1.2.2 Determination of Rate Constants

Once the order of the reaction has been determined, calculation of the rate constant can easily be achieved using the above integrated equations by substitution or least squares regression plotting. In the above initial reaction rate method, the antilogarithm of the constant term of Equation (7), which is referred to as the apparent or observed rate constant, k_{obs}, is equal to $k[\mathrm{B}]_0^b[\mathrm{C}]_0^c$ (Equation 6). Thus, the value of k is easily calculated from k_{obs} values using the known initial concentrations of B and C, and the already estimated partial orders b and c.

Since the majority of the analytical reactions used in kinetic analysis are first- or pseudo-first-order, two commonly used methods for the determination of first-order reaction rate constants from experimental (signal) data are presented.

1.2.2.1 Infinite-time Method The course of the reaction is followed by measuring some signal S, which is linearly related to the concentration of one of the reactants or the product. The signal varies with time according to Equation (8)

$$(S_\infty - S) = (S_\infty - S_0)\mathrm{e}^{-kt} \tag{8}$$

in which S_0, S, and S_∞ are the values of the measured signal at $t = 0$, t, and $t = \infty$, respectively. It is obvious that a plot of $\ln[(S_\infty - S)/(S_\infty - S_0)]$ versus t will be a straight line with a slope equal to $-k$ (or $-k_{\text{obs}}$). This method requires the completion of the reaction.

1.2.2.2 Guggenheim Method This method does not require completion of the reaction. If the readings of the analytical signal, $S_1, S_2, \ldots, S_i$ are made at times $t_1, t_2, \ldots, t_i$ and a second series of readings $S_1', S_2', \ldots, S_i'$ is made at the corresponding times $t_1 + \Delta t, t_2 + \Delta t, \ldots, t_i + \Delta t$ (where Δt is a constant of magnitude of the order of one half-life) then Equation (9) follows

$$(S_\infty - S_i') = (S_\infty - S_0)\mathrm{e}^{-kt_i} \tag{9}$$

and Equation (10)

$$(S_\infty - S_i') = (S_\infty - S_0)\mathrm{e}^{-k(t_i+\Delta t)} \tag{10}$$

After subtracting Equation (10) from Equation (9), obtaining the logarithms and assuming that S_∞, S_0, and Δt are constants, we have Equation (11)

$$\ln|P_i' - P_i| = \text{constant} - kt_i \tag{11}$$

from which by plotting the left-hand function versus t_i a straight line is obtained with a slope equal to $-k$ (or $-k_{obs}$). For the application of the Guggenheim method the reaction should obey first-order kinetics over at least two half-life periods.

The partial order in a given reactant will be unity if a plot of initial rate versus reactant concentration is a straight line; in fact, at $n = 1$, Equation (3) is representative of a straight line of zero intercept.

2 DETERMINATION OF A SINGLE COMPONENT – PRINCIPLES OF METHODS AND TECHNIQUES[1–8]

The determination of the analyte A based on the second-order reaction described by the rate Equation (1) can be performed under two limiting conditions: (1) in the case of R being in large excess (over a 50-fold) relative to the analyte (A), then the reaction will proceed under pseudo-first-order in respect to A or (2) in the case of concentration of R being equal to or less than 50 times the concentration of A, the reaction will proceed under second-order conditions.

The various approaches used in the determination of a single noncatalytic species are shown in Table 1 and discussed in detail below.

2.1 Pseudo-first-order Conditions

The general differential rate expression of a reaction under pseudo-first-order conditions in respect to the analyte, A, is Equation (12):

$$\text{rate} = -\frac{\mathrm{d[A]}}{\mathrm{d}t} = \frac{\mathrm{d[P]}}{\mathrm{d}t} = k[\mathrm{A}][\mathrm{R}]_0 = k_{\mathrm{A}}[\mathrm{A}] \quad (12)$$

where $k_{\mathrm{A}} = k[\mathrm{R}]_0$. Equation (12) is normally the basis for the kinetic determination of A under pseudo-first-order conditions.

Table 1 Analytical approaches used in kinetic determinations of single noncatalytic species

Pseudo-first-order conditions
Initial rate approach
Fixed-time approach
Variable-time approach
Signal-stat approach
Second-order conditions
Identical reactant concentrations
Unequal reactant concentrations
Multipoint approaches
Curve-fitting approach
Predictive approach
Error-compensated approach

In the case of initial-rate approach, the rate expression (Equation 12) can be written in incremental form as Equation (13):

$$\text{rate} = -\frac{\Delta[\mathrm{A}]}{\Delta t} = k_{\mathrm{A}}[\mathrm{A}]_0 \quad (13)$$

from which the concentration of the analyte A can readily be calculated. In this approach, the rate of signal change is measured (or computed) and related to the analyte concentration. Assuming that the analyte is the rate determining reactant (R is in excess), the relationship can be represented as Equation (14)

$$\text{rate} = \frac{\Delta S}{\Delta t} = bk[\mathrm{A}]_0 \quad (14)$$

where b is the sensitivity factor. Calibration standards are usually used to obtain a calibration constant representing the proportionality factor, bk. In such procedures, changes in any variable (e.g. temperature, pH) that would affect the rate constant would produce an error in the determined concentration.

The principles of fixed-time and variable-time approaches are based on Equation (15) which is the exponential expression of Equation (12):

$$\text{rate} = -\frac{\mathrm{d[A]}}{\mathrm{d}t} = k_{\mathrm{A}}[\mathrm{A}]_0 \exp(-kt) \quad (15)$$

Assuming that $t_2 = t_1 + \Delta t$, the integration of Equation (15) over an interval $t_2 - t_1$ results in Equation (16)

$$[\mathrm{A}]_0 = \frac{-\Delta[\mathrm{A}]}{\exp(-k_{\mathrm{A}}t_1)[1 - \exp(-k_{\mathrm{A}}\Delta t)]} \quad (16)$$

In the fixed-time approach t_1 and Δt are kept constant, thus, as stated from Equation (16) a linear relationship exists between $\Delta[\mathrm{A}]$ and $[\mathrm{A}]_0$. In a one-point/fixed-time approach, the detector signal, S_t, is measured at some fixed point in time after the reaction (process) is initiated. In such procedures applied to first-order reactions, the background signal, S_0, is assumed to be zero and analyte concentration is determined from a single proportionality of the form (Equation 17)

$$[\mathrm{A}]_0 = \frac{S_t}{b[1 - \exp(-k_{\mathrm{A}}t)]} \quad (17)$$

where b is the sensitivity factor. The calibration constant, $\{1/b[1 - \exp(-k_{\mathrm{A}}t)]\}$ is evaluated by calibration standards or calibration plots. The one-point option cannot compensate for background (blank) signals and it cannot detect unexpected changes in kinetic behavior. Any change in variables that would affect the rate constant would also affect the calibration constant and computed concentration. The one-point/fixed-time approach

is seldom used now because it does not compensate for changing backgrounds.

The two-point/fixed time option is perhaps the most widely used approach for kinetic determinations. In this approach, two signal values, S_1 and S_2, are measured at two points in time, t_1 and t_2, and the difference, $\Delta S_t = S_1 - S_2$, is used to compute the concentration. The mathematical description for a first-order process is easily shown to be Equation (18)

$$[A]_0 = \frac{\Delta S_\infty}{b} = \frac{\Delta S_t}{b[\exp(-k_A t_1) - \exp(-k_A t_2)]} \qquad (18)$$

In the most common applications of this approach, the calibration factor in the denominator of the term at the far right-hand side of the expression is determined by using calibration standards. This option compensates for background signals but, like the one-point option, cannot detect unexpected changes in kinetic behavior.

In the variable-time approach the quantity $\Delta[A]$ is kept constant. According to Equation (16) Δt and t_1 depend on $[A]_0$, therefore the analyte concentration $[A]_0$ under determination will never be linearly related to $1/\Delta t$. In this situation the error will be minimized if $[A]_1$ is almost equal to $[A]_0$ and therefore $\exp(-k_A t_1) \rightarrow 1$ and at short reaction times (small $\Delta[A]$ and $\Delta t \rightarrow 0$) Equation (16) can be linearized since $[1 - \exp(-k_A \Delta t)]$ can be simplified to $k\Delta t$ by expanding the exponential function as a Maclaurin series. In this option, the measurement objective is the time interval, Δt, required for the reaction (process) to proceed to a predetermined extent.

In the one-point/variable-time approach the quantity measured is the time interval between the start of a reaction and some predetermined point in the reaction, usually completion. An early example of this option is the so-called ''stop-watch method''.

In the two-point/variable-time approach, the measurement objective is the time interval between two predetermined signal values. The time interval is used to compute concentration. This option has some interesting features and has been used in automated kinetic potentiometric methods using a double switch system. A physicochemical analog involves a flow-injection (FI) approach in which the measured quantity is the time required for the detector signal to change between two predetermined points on a response profile (FI pseudo titrations). A plot of Δt versus $\log [A]_0$ is linear.

In the above approaches, knowledge of the relationship between concentration and the detector signal is not required and the dynamic range can be extended by control of reagent concentration.

In the signal-stat approach, the detector signal is used to indicate when a property (e.g. pH) has changed outside some predermined narrow limits and a reagent (e.g. acid or base) is added to return the property to within the desired limits. The quantity of reagent needed to maintain the measured property within the desired limits is measured and used to compute the analyte concentration. Therefore, the time-dependent addition of reagent reflects the time course of the reaction. Usually multiple additions of reagent are required so that a multiple response curve is available. The most common version of this approach is the pH-stat option but the concept should be applicable to the control of any reactant.

In conclusion, the fixed-time approach ensures wider dynamic concentration ranges than does the variable-time approach for first- or pseudo-first-order conditions.

2.2 Second-order Conditions

The rate of a reaction under second-order conditions is dependent on the concentrations of two reactants. The expression for the reaction rate is Equation (19)

$$\text{rate} = -\frac{d[A]}{dt} = -\frac{d[R]}{dt} = \frac{d[P]}{dt} = k[A][R] \qquad (19)$$

In the case of equal concentrations of the two reactants, ($[A]_0 = [B]_0$), Equation (19) becomes Equation (20):

$$\text{rate} = -\frac{d[A]}{dt} = k[A]^2 \qquad (20)$$

Integration of Equation (20) in the time interval zero and t results in the linear expression Equation (21):

$$\frac{1}{[A]_t} = \frac{1}{[A]_0} + kt \qquad (21)$$

from which the analyte concentration $[A]_0$ can be estimated directly.

In the case of an analytical reaction under second-order conditions in which the concentrations of the two reactants are different, integration of Equation (19) between $t_1 = 0$ and $t_2 = t$ results in the general second-order integrated rate expression (Equation 22):

$$\ln \frac{[R]_t}{[A]_t} = \ln \frac{[R]_0}{[A]_0} + ([R]_0 - [A]_0)kt \qquad (22)$$

A plot of the left-hand term $\ln([R]_t/[A]_t)$ versus time will be linear and the analyte concentration $[A]_0$ can be readily calculated from the intercept and the known value of $[R]_0$.

2.3 Multipoint Approaches

The wide availability of computers with high processing power which usually are used on-line with the analytical instruments of kinetic analysis has promoted the development of various powerful approaches based

on the efficient use of the full analytical information from the reaction curve. These include a curve-fitting approach, a predictive approach and an error-compensated approach.

A curve-fitting approach computes modified signals in order to obtain the best fit for a mathematical model describing the transient process. The least-square procedure is widely used for this purpose.

A predictive approach takes advantage of a large number of experimental data to calculate the best estimate for the signal change to be expected if the reaction system were allowed to reach equilibrium (S_∞), using linear or nonlinear least-squares or the Kalman filter. The object of the curve-fitting process would be to obtain best-fit values of S_0, ΔS_∞, and the rate constant, k. Thus, in a single data-processing step, it is possible to compensate intrinsically for variables that affect rate constants.

An error-compensated approach minimizes the marked dependence of the transient signal of the reaction system from experimental variables. Various mathematical algorithms or simultaneous measurements of analytical signals and the variable(s) affecting the system are used in order to establish pertinent corrections.

2.4 Techniques of Mixing and Monitoring

Kinetic methods are performed using instruments of various complexity, from conventional simple systems such as photometers, fluorimeters or potentiometers to highly sophisticated fully automated systems, capable of sample aliquoting, transferring to the detector, signal collecting and manipulating, and processing the results obtained.

Depending on the type of system concerned, several techniques are used for mixing of the reactants to start the analytical reaction.

2.4.1 Conventional Simple Techniques

Slow chemical reactions can be followed using simple spectrophotometric, fluorimetric or potentiometric systems, usually equipped with thermostatted observation cells, and manual mixing and/or magnetic or mechanical stirring.

2.4.2 Stopped-flow Technique

In the stopped-flow (SF) technique a prepared sample solution is mixed rapidly and efficiently with a prepared reagent solution by means of two driving syringes that are actuated manually or by a pneumatic device, through a mixing-chamber into a flow-cell which can also serve as the observation cell. The flow is stopped abruptly by using a third stopped syringe and the analytical signal (usually absorbance or fluorescence) is then recorded as a function of time. This technique is ideal for performing fast reactions with half-lives of a few milliseconds to several seconds. Its performance depends mainly on the dead time (mixing time).

2.4.3 Flow-injection technique

Flow-injection analysis (FIA) has been defined as ''information gathering from a concentration gradient formed from the injected, well-defined zone of a fluid, dispersed into a continuous unsegmented stream of a carrier''. Therefore, all chemical assays performed by FIA can be said inherently and necessarily to entail the utilization of chemical kinetics. Under FIA conditions it is possible to utilize masks which operate on a time-based mode, taking advantages of the differences in kinetic rates between the mask and the interferent or analyte. FIA can operate in a continuous-flow, SF, recycle flow or reversal flow mode. Reactions producing unstable intermediates can be easily adapted in the FIA technique.

In the continuous-flow mode a signal peak is obtained after a preselected time interval (fixed-time approach). If an external mixer is used a wide signal peak is obtained capable of being used in the two-point/variable-time approach. In the SF mode (FI/SF) the reacting mixture is stopped in the detector cell at time t_1 and the signal is monitored during a Δt time interval (reaction rate or slope approach). In the recycle flow mode, as denoted by its name, the reacting plug is passed several times through the flow detector in the same direction providing a multipeak response whose envelope represents the profile of the kinetic curve. In the reversal flow mode, the reacting plug is also passed several times through the flow detector but alternately in opposite directions, which provides a response similar to that of the recycle flow mode.

2.4.4 Continuous Addition of Reagents Technique

The continuous addition of reagents (CAR) technique is based on the continuous addition at a constant rate of one of the reaction components (usually the reagent) in the reaction cell equipped with a stirring device. The reaction curve (absorbance or fluorescence as function of time) is recorded simultaneously with the aid of an automated computer-controlled system.

2.4.5 Stat Technique

This technique, which is suitable for slow reactions, is based on the addition of microamounts of a reactant to the reaction cell at a known rate such that an analytical

parameter of the monitored reaction (pH, absorbance, luminescence, etc.) is maintained constant. Any deviation from this state because of the reaction processing is immediately compensated for by automatically adding an extra amount of the reactant. The rate of addition is used to estimate the reaction rate.

2.4.6 Centrifugal Mixing

This technique is used in parallel fast analyzers. Microvolumes of samples and reagent(s) are added in microwells on a removable rotating disk. The disk is placed in a centrifuge rotor. When the rotor spins, mixing is induced by the action of the centrifugal force, which finally displaces the mixed solution into the observation cell. The technique provides a great number of data during the observation time.

2.5 Detection Systems

The course of a reaction can be monitored by chemical or visual means, but mainly by instruments (photometric, fluorimetric, chemiluminometric, electroanalytical devices). A large number of reactions involve substances that absorb in the near-ultraviolet or visible region. Rapid scanning spectrometry (vidicon image detector, photodiode array) offers an attractive potential for kinetic methods of analysis. Ion selective electrodes (ISEs) are of special use in potentiometric techniques as applied to kinetic analysis on account of their usually high sensitivity to concentration changes and the freedom from optical interferences.

3 REACTION TYPES[1,6]

The various reactions used in kinetic-based noncatalytic determinations of single species can be categorized into several groups, which are described in detail below.

3.1 Complexation Reactions

This group includes both formation and substitution reactions, mainly applied to inorganic analytes.

3.1.1 Substitution Reactions

Kinetic methods based on substitution reactions involve either a ligand or a metal displacement from a starting complex:

- Ligand substitution reaction $ML_1 + L_2 \longrightarrow ML_2 + L_1$
- Metal substitution reaction $M_1L + M_2 \longrightarrow M_2L + M_1$

The ligand exchange between Cu(II) complex with 2-(2-thiazolyazo)-5-dimethylaminophenol and ethylenediaminetetraacetic acid (EDTA) is the basis of a spectrophotometric kinetic method for copper. The substitution of Cu(II) by Ga(III) in the copper–EDTA complex is the basis for a spectrophotometric kinetic method for Ga(III). The exchange between Th–DCTA (diaminohexanetetraacetic acid) complex and the attacking ligand arsenazo(III) can be used for the kinetic spectrophotometric determination of thorium.

The kinetic spectrophotometric determination of chloride using FI mixing and fixed-time approach is based on the following scheme of reactions, Equation (23) and (24):

$$2Cl^- + Hg(SCN)_2 \Longleftrightarrow HgCl_2 + 2SCN^- \quad (23)$$

$$SCN^- + Fe^{3+} \Longleftrightarrow Fe(SCN)^{2+} \quad (24)$$

The displacement of fluoride from its aluminum complexes by EDTA, citrate and α-hydroxyacids is the basis for their potentiometric kinetic determination using fluoride ISE.

3.1.2 Complex formation Reactions

This type of reaction is mainly used for the determination of a single metal ion or an organic substance acting as a ligand. The complexation of iron(III) with thiocyanate is the basis of a spectrophotometric kinetic method for Fe(III). Aluminum can be determined by an initial rate fluorimetric method based on its chelation with 2-hydroxy-1-naphthaldehyde *p*-methoxybenzoylhydrazone. *O*-Cresolphthalein complexone is used as the complexing agent for the spectrophotometric determination of calcium using the SF technique. The complexation of magnesium by calmagite in alkaline medium and in the presence of bis(aminoethyl)glycolether-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA) to mask calcium is the basis of a spectrophotometric kinetic method for magnesium.

A selective kinetic spectrophotometric method for zinc is based on its complexation with zincon which follows the differential demasking of the cyanide complexes with cyclohexanone, Equation (25) and (26):

$$[Zn(CN)_4]^{n-} + 4C_6H_{10}O + 4H_2O \longrightarrow Zn^{2+} + 4C_6H_{10}(OH)CN + 4OH^- \quad (25)$$

$$Zn^{2+} + \text{zincon} \Longleftrightarrow \text{blue complex} \quad (26)$$

The formation of molecular complexes has been used in organic kinetic analysis. The reaction of albumin with picrates monitored with a picrate ISE has been

proposed for the determination of albumin in serum. The formation of the characteristic yellow–orange complex of creatinine with alkaline picrate has been the basis of its spectrophotometric or potentiometric (using a picrate ISE) kinetic determinations.

Of special interest are the immunochemical reactions of antigen and haptens with their antibodies. Several drugs of abuse are determined using the fluorescence polarization technique. The specific precipitation of lipoprotein by heparin, calcium ions, EDTA and lipase is the basis of a turbidimetric kinetic method.

3.2 Redox Reactions

Redox reactions have been mainly used for the determination of anionic rather than cationic species in inorganic analysis.

The reduction of iodate to iodine by arsenite in strongly acidic medium is used in an initial-rate spectrophotometric SF method for arsenite. Iodide can be oxidized by perbromate to triodide complex, I_3^-, the spectrophotometric (353 nm) monitoring of which is the basis for the kinetic determination of perbromate (0.3–3 μg). The competitive equilibria of oxidation and complexation between the benzyl 2-pyridyl ketone 2-pyridylhydrazone, bromate and copper(II), nickel(II) and palladium(II) cations is the basis for the spectrophotometric kinetic method (initial rate approach) for these cations. The reduction of Mo(VI) by hydrazine in acidic medium can be used for the kinetic spectrophotometric determination of hydrazine.

The kinetic determination of iodine (0.6–7 ng mL^{-1}) is based on its reaction with acetone in alkaline medium to form iodoform and iodide which is monitored with an iodide ISE (initial rate approach) Equation (27):

$$(CH_3)_2CO + 3I_2 + 4OH^- \longrightarrow CHI_3 + CH_3COO^- + 3I^- + 3H_2O \qquad (27)$$

The oxidation of 4,8-diamino-1,5-dihydroxyanthraquinone-2,6-disulfonate in alkaline medium by manganese in the form of hydrated MnO_2 is the basis for a fluorimetric kinetic method for Mn.

The well-known Karl Fischer reaction for water is the basis for its kinetic SF spectrophotometric determination (1–10 mg mL^{-1}) using an acetate pyridine-free reagent, Equation (28):

$$CH_3SO_3I_2^- + H_2O \longrightarrow CH_3SO_4^- + 2I^- + 2H^+ \qquad (28)$$

Redox reactions are of special interest in organic analysis and rely on the oxidation of organic compounds by a strong oxidant such as permanganate, cerium(IV), iron(III), periodate, dichromate, hexacyanoferrate(III), peroxydisulfate, or hydrogen peroxide.

Periodate is a common oxidant for organic hydroxy compounds including phenols, chlorophenols, hydroxyacids (e.g. tartaric acid) and vicinal glycols (e.g. glycerol) (Malaprade reaction), Equation (29):

$$A + IO_4^- \Longleftrightarrow A \cdot IO_4^- \longrightarrow \text{products} \qquad (29)$$

where $A \cdot IO_4^-$ is an intermediate (periodate ester). Fixed-time (absorbance measurements) and variable-time (perchlorate or home-made periodate-selective electrode) methods can be used for the determination of these species and related substances such as carbohydrates (e.g. glucose).

Peroxydisulfate oxidation of organic carbon is the basis for an automated determination of organic carbon in water samples.

One interesting redox reaction is the reversible interconversion of ferriin/ferroin complexes. The 1,10-phenanthroline/iron(III) complex (ferriin) has been used to oxidize species of clinical (uric acid) and pharmaceutical interest such as analgesics (acetaminophen) and catecholamines, the reaction being monitored by measuring the initial rate change of the absorbance of the ferroin at 510 nm. The ligand 2,4,6-tris-(2-pyridyl)-*s*-triazine (TPTZ) has also been used.

The oxidation of various organic substances (carbamazepine, choline, acetylcholine) with Ce(IV) in acidic medium yields fluorecsent Ce(III) species used for their kinetic fluorimetric determination. Carbimazole is oxidized to a colored product using dichromate or molybdate at 90 °C.

The oxidation of various organic analytes (alcohols and carbohydrates) with dichromate in acidic medium can be monitored with a home-made dichromate ISE and used for their kinetic determination.

The formation of colored unstable free radicals yielded during the oxidation of phenothiazines (chlorpromazine, methotrimeprazine, thioridazine, etc.) with Fe(III) in acidic medium has been used for their spectrophotometric kinetic determination.

The Mn(II)-catalyzed oxidation of oxalate with permanganate has been used for a fast kinetic spectrophotometric determination of oxalate.

Bromosuccinimide oxidizes several organic compounds such as vitamins (pyridoxine, thiamine, ascorbic acid), being reduced to succinimide with the liberation of bromide which can be monitored with a bromide ISE.

3.3 Coupled Reactions

The most typical example of this type of reaction is the so-called Griess reaction, originally developed for the kinetic determination of nitrite. Nitrite in a weakly acidic solution reacts with an amine (usually

sulfanilic acid) to form the corresponding diazonium salt, which then couples with a naphthylamine derivative or naphthol. This reaction can be used for the kinetic determination of nitrate after reduction in a column (copperized cadmium), as well as for measuring other species such as sulfonamides (the Bratton–Marshall reaction), bromhexine, benzodiazepines (e.g. nitrazepam) and *N*-methylcarbamate pesticides, hydrolysis of which yields the corresponding benzophenones (which can be diazotized) and naphthols (useful for the coupling reaction), respectively. Nitroprusside liberates nitrite in alkaline medium and can be determined kinetically using the Griess reaction.

The well-known Berthelot reaction of ammonia is the basis for various kinetic methods for ammonia–nitrogen. Ammonia reacts with hypochlorite to form chloramine which then is coupled with phenol in two steps to form indophenol blue, Equations (30–32):

$$NH_3 + OCl^- \longrightarrow NH_2Cl + OH^- \quad (30)$$

$$2OCl^- + NH_2Cl + C_6H_5O^- \longrightarrow OC_6H_4NCl + 2Cl^- + OH^- + H_2O \quad (31)$$

$$OH^- + OC_6H_4NCl + C_6H_5O^- \longrightarrow OC_6H_4NC_6H_4O^- + H_2O + Cl^- \quad (32)$$

Aminoglycoside antibiotics (amicacin, kanamicin, tobramycin), through their primary amine group, participate in a cyclization reaction with *o*-phthalaldialdehyde and a thiol to form a 1-alkylthio-2-alkylisoindole fluorescent derivative which can be used for their kinetic determination.

Oxidative coupling reactions of *p*-phenylenediamines (PPD) with amines and phenols are widely used. PPD is oxidized to its quinone diimine (QDI) by hexacyanoferrates(III) or a similar oxidant in a weakly basic medium. QDI reacts with the amine or phenol compound to give leucoindamines (leucoindophenols), which are rapidly oxidized to colored indamines (indophenols) with the aid of a QDI molecule.

The reduced form of ascorbic acid (dehydroascorbic acid) obtained using Hg(II) reacts with *o*-phenylenediamine to form fluorescent quinoxaline. The oxidative condensation of *para*- and *meta*-substituted phenols with 1-nitroso-2-naphthol using Ce(IV) or Pb(IV) can be used for the kinetic spectrophotometric determination of phenols.

The so-called bromination reactions use bromine generated in situ from BrO_3^-/Br^- couple in an acidic medium. At low analyte concentrations, the rate of bromination and bromine generation are virtually the same, so methyl orange in the reaction medium will only be decolorized after the bromination process has been completed. The decolorization time will be proportional to the analyte (e.g. acetaminophen) concentration. Bromide can be determined in addition to phenol compounds, which are the typical substrates.

The bromination of the yellow indicator phenol red (PR) to bromothymol blue (BTB) at pH 4.6 using bromine produced by the oxidation of bromide with chloramine-T is the basis for a kinetic determination of bromide using FI mixing and the fixed-time approach, Equations (33) and (34):

$$2Br^- + H_3C{-}C_6H_4SO_2NCl^- + 2H^+ \longrightarrow Br_2 + H_3C{-}C_6H_4SO_2NH_2 + Cl^- \quad (33)$$

$$4Br_2 + PR \Longleftrightarrow BTB + 4H^+ + 4Br^- \quad (34)$$

3.4 Addition Reactions

Formation of addition compounds from species containing carbonyl groups has been exploited for the determination of organic compounds reacting with hydrogen sulfite ion in the presence of acetone, methanol or acetic acid. The addition reaction of sulfite with rosaniline at pH 11.8 has also been used for the kinetic spectrophotometric determination of sulfite. The addition reaction of hydrogen cyanide to carbonyl group (e.g. formaldehyde) monitored with a cyanide ISE has been used for its kinetic determination. Aromatic aldehydes have been determinated spectrophotometrically using their condensation reaction with barbituric acid. Primary and secondary amines (e.g. amphetamine, ephedrine) react with 1,2-naphthoquinone-4-sulfonate to form a colored product which is the basis for their kinetic determination. The reaction of amino acids with ninhydrin (a dihydroxyindanedione) at 80 °C is the basis for their kinetic photometric method. Tryptophan reacts with formaldehyde in alkaline solution to form a highly fluorescent product.

Urea reacts with the unstable diacetyl (produced in situ by the acidic hydrolysis of diacetyl monoxime) to form a yellow diazine derivative.

3.5 Aromatic Substitution Reactions

Several reagents capable to undergo aromatic substitution reactions with various substances owing reactive functional groups have been proposed for the kinetic determination of these substances. A characteristic example is the reaction of 1-fluoro-2,4-dinitrobenzene (FDNB) with amines, amino acids, phenols, thiols, hydrazines, and hydrazides. The reaction, through the formation of an intermediate complex (Meisenheimer complex)

Table 2 Selected noncatalytic determinations of inorganic single species

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Aluminum	Complex formation with fluoride (pH 3)	5–50 μM 20–200 μM	CF/SF with fluoride ISE detector (slope approach). Application in tea leaves
Aluminum	Complex formation with chrome azurol S, eriochrome cyanin R, pyrocathechol	0.060–30 μM	Spectrophotometric FIA. Estimation of kinetically labile Al in soils
Ammonia nitrogen	Berthelot reaction (phenol–hypochloride)	20–250	FI/SF system, multipoint reaction rate approach
Boric acid	Pseudotitration with mannitol in the presence of BTB indicator	2.5×10^{-4}–5×10^{-2} M	FIA spectrophotometric gradient system. Measurement of width of absorbance peaks (Δt vs $\log C$). Application in pharmaceutical formulations
Bromide	Reaction with chloramine-T to produce bromine and bromination of PR to bromophenol	1–10	FIA spectrophotometric system, fixed time approach. Application in waters, serum and pharmaceutical formulations of bromide salts of drugs and dissolution studies
Cerium(IV)	Oxidation of 4,8-diamino-1,5-dihydroxyanthraquinone-2,6-disulfonate	0.02–0.37	Fluorometric monitoring, initial-rate approach
Copper	Complexation with 5,5-dimethyl-1,3-cyclohexanedione bis(4-phenyl-3-thiosemicarbazone) in acidic medium	0.8–8	Spectrophotometric monitoring, Fixed-time and initial-rate approaches
Copper(II) or Zinc	Incorporation in coproporphyrin-I	$0–1 \times 10^{-5}$ M	Monitoring the decrease in fluorescence of free base porphyrin. Initial rate or fixed-time approach
Germanate	Formation of molybdenum germanate blue		Spectrophotometric monitoring
Hydrazine or sodium azide	Liberation of fluoride during the reaction with FDNB	$1–10 \times 10^{-4}$ M	Initial rate approach using fluoride ISE
Hypochlorite	CL reaction with luminol	0.0015–10	SF CL spectrometer. Formation rate, decay rate, peak height and peak area were used to construct calibration curves. Application in waters
Iodide	Reduction of iodate with hydrazine	region of 2	Monitoring with iodide ISE. Time required to generate sufficient iodide to cause the iodate–iodide reaction to become predominant is related to iodide concentration
Iron(III) and Thallium(III)	Oxidation transformation of 1,4-diamino-2,3-dihydroxy-anthraquinone	0.05–0.6 for Fe 0.05–0.4 for Th	Fluorometric monitoring, initial rate approach
Manganese(II)	Complexation with terephthal-monohydroxamic acid (pH 10.3)	1.25–7.25	Spectrophotometric monitoring, initial-rate approach. Application in steels
Mercury(II)	Liberation of cyanide from the reaction with 1,10-phenathroline and ferrocyanide	0.010–0.2 μM	Monitoring with Ag-coated piezoelectric quartz crystal for 4 min
Nickel	Compexation with xylenol orange in the presence of hexadecyl-trimethylammonium bromide	0–0.56	Spectrophotometric monitoring using a flow system. Application in nickel ores
Niobium	Complexation with 4-(2-pyridylazo)resorcinol in the presence of tartrate to mask tantalum(V)	0.5–20 μM	Spectrophotometric monitoring. Fixed-time approach
Nitrate	Liberation of nitrite using on-line reduction with copperized cadmium and subsequent diazotization and coupling with NEDD	0.025–3	Spectrophotometric SF monitoring, initial-rate and fixed-time approaches. Application in waters

Table 2 (*continued*)

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Nitrate	Liberation of nitrite after reduction with cadmium, diazotization with sulfanilamide and coupling to NEDD	2–250 μM	Spectrophotometric monitoring, fixed-time approach. Application to serum and urine
Nitroprusside	Liberation of nitrite in alkaline medium, diazotization with sulfadiazine and coupling reaction with *N*-(1-naphthyl)-ethylenediamine	0.1–100	Spectrophotometric monitoring initial rate approach. Application in pharmaceutical formulations
Phosphorus	Formation of molybdophosporic acid and reduction to phosphomolybdenum blue by ascorbic acid	(i) 20–100	(i) Spectrophotometric SF, reaction-rate approach (2–6 s). Application in feed samples
		(ii) 5–30	(ii) FI/SF system, multipoint reaction rate approach
Selenite	Reduction by ascorbic acid to elemental selenium	10^{-5} M level	Turbidimetric monitoring (460 nm), fixed-time procedure
Sulfide	Reduction of resazurin in the presence of Cu(II) and ethylenediamine	0.05–6	Spectrophotometric monitoring, fixed-time and slope approaches
Sulfite	Reduction of methyl green	0.05–25	Monitoring of decrease in absorbance, fixed-time approach
Sulfite	Reduction of: Fe(III)-1,10-phenanthroline Fe(III)-2,2′-bipyridyl	0.2–5	Spectrophotometric monitoring
Sulfite	Addition reaction with rosaniline (pH 11.8)	0.015	Spectrophotometric monitoring, slope approach. Application in beverages, biscuits and steamed buns
Sulfite	Oxidation with iodate producing iodide in acidic medium	5×10^{-5}–5×10^{-3} M	Reciprocal-time approach using iodide ISE
Technetium	Color formation with 1,3,5-triphenyl-Δ^2-pyrazoline	0.01–12	Initial-rate and fixed-time approaches
Thiocyanate	Inhibition of the ferroin – periodate reaction in acidic medium	20–200 $ng\,mL^{-1}$	Monitoring of decrease in absorbance, initial rate approach. Application in human serum and saliva
Vanadium(V)	Oxidation of 1-amino-4-hydroxyanthraquinone	100–530 $ng\,mL^{-1}$	Fluorimetric monitoring, initial rate method. Application in crude petroleum
Zinc	Complexation with zincon in the presence of cyanide to mask interfering metal ions and cyclohexanone to dissociate the zinc-cyanide complex	1–10	FIA spectrophotometric system. Fixed-time approach. Application in waters, alloys and insulin formulations

LOD, limit of detection; CF/SF, continuous flow/stopped flow; NEDD, *N*-(1-naphthyl)ethylenediamine dihydrochloride.

leads to the formation of a dinitrobenzene (DNB)-derivative (capable of spectrophotometric monitoring) and the release of fluoride ion (capable of potentiometric monitoring with a fluoride ISE) Equation (35):

$$\begin{aligned} RNH_2 + (NO_2)_2C_6H_3F &\Longleftrightarrow (NO_2)_2C_6H_3(F)NH_2R \\ &\longrightarrow (NO_2)_2C_6H_3NHR \\ &\quad + H^+ + F^- \end{aligned} \tag{35}$$

These rather slow reactions can be accelerated by cationic surfactants using micellar catalysis. Several spectrophotometric and potentiometric methods using manual or FI mixing have been proposed for a great number of substances of pharmaceutical interest.

The nucleophilic substitution reaction of trinitrobenzenesulfonate (TNBS) with amino acids and carbohydrates in an alkaline medium monitored with a TNBS home-made ISE has been used for the kinetic determination of the analytes, Equation (36):

$$\begin{aligned} RNH_2 + (NO_2)_3C_6H_2{-}SO_3^- &\longrightarrow (NO_2)_3C_6H_2{-}NHR \\ &\quad + SO_3^{2-} + H^+ \end{aligned} \tag{36}$$

3.6 Hydrolysis Reactions

Several hydrolysis reactions have been used for kinetic-based determinations. The alkaline hydrolysis of furazolidone, furaltadone, and methyl parathion can be used for their spectrophotometric kinetic determination. The oxidative (H_2O_2) alkaline hydrolysis of chlopyrifos is the basis for its kinetic determination. The alkaline hydrolysis of *p*-chloranil yields chloranilate which can be monitored with a chloranilate home-made ISE.

The acid degradation with heating of tetracycline and oxytetracycline to the anhydro derivatives is the basis of their spectrophotometric determination.

3.7 Chemiluminescence Reactions

While the main field of application of chemiluminescence (CL) reactions is catalytic processes (determination of inorganic catalysts and various inhibitors) and senzitization processes (by energy transfer and stabilization), some noncatalytic reactions are also of great interest on account of the high sensitivity they provide. Some characteristic examples of such reactions are that of hypochlorite with luminol for the determination of hypochlorite, the reaction of acetaminophen with Ce(IV) in acidic medium, and the reaction of ethanol with MnO_4^- in very acidic solutions.

4 ANALYTES DETERMINED[5,9–11]

4.1 Determination of Inorganic Species

Complex formation and redox reactions are mainly used for the determination of metals and nonmetals. Table 2 presents selected kinetic methods for inorganic species. The majority of these methods are based on photometric detection and the concentration ranges typically are in the micrograms per milliliter region.

4.2 Determination of Organic Species

Kinetic methods of analysis based on noncatalyzed reactions are very attractive in organic analysis, since they generally lack a catalytic activity. All types of reactions have been used in kinetic organic analysis and the majority of the reactions are monitored spectrophotometrically and fluorimetrically. Table 3 shows selected examples of kinetic methods employed in organic analysis.

5 APPLICATION AREAS

The increased automation of kinetic methods, especially in organic analysis, using a variety of commercial or home-made analyzers of various types (FI, SF, etc.) has resulted in a tremendous increase in the use of noncatalytic kinetic methods. The advantages of short analysis time, high precision and the absence of serious interferences make the application of uncatalyzed kinetic methods very attractive for routine analysis of real samples in a variety of areas. Clinical, environmental, food and pharmaceutical analysis are the areas where kinetic methods are applied with success and are discussed below.

5.1 Clinical Analysis

The following inorganic and organic species can be determined in serum and/or urine using kinetic methods based on noncatalytic reactions:

- **Serum:** Aluminum, bromide, calcium, magnesium, nitrate, phosphate, thiocyanate.
  Acetaminophen, ascorbic acid, carbamazepine, malonaldehyde. Albumin, creatinine, lipoprotein, total serum protein, urea, uric acid, hemoglobin, methemoglobin, reduced glutathione.
- **Urine:** Nitrate, phosphate.
  Amphetamines, ascorbic acid, morphine, sulfonamides. Catecholamines, creatinine, uric acid.

Some of the methods can be applied directly on the biological sample (urea, uric acid and phosphate) but others required treatment (protein precipitation, liquid–liquid extraction, solid-phase extraction (Sep-Pack cartridges) etc.). The detectability and sensitivity of the kinetic methods are suitable so that a rather low volume of sample is required (20–500 μL of serum). Most kinetic methods can be performed using automated clinical analyzers (centrifugal, batch, and flow analyzers).

5.2 Environmental Analysis

The application of kinetic methods in this area is extensive. The sample matrix (natural-, waste- and seawaters) is simple allowing direct determinations without any treatment. Water and air samples can very easily be analyzed using automated apparatus.

The following species, most of them being inorganic, can be determined in various environmental samples using noncatalytic kinetic methods:

- Air: carbon dioxide, manganese, nitrogen oxides.
- Natural waters: arsenite, hydrazine, hypochlorite, iodine, magnesium, manganese, nitrates, nitrite,

Table 3 Selected noncatalytic determinations of organic single species

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Acetaminophen	Reduction of Fe(III) in the presence of TPTZ	50–500	Spectrophotometric monitoring (593 nm), FI mixing, fixed-time approach. Application in drug formulations and dissolution studies. An alternative method uses SF mixing and initial rate approach ($20–200\,\mu g\,mL^{-1}$). In the presence of chlorpromazine the reaction is catalyzed allowing determination in the range $0.5–6\,\mu g\,mL^{-1}$ in serum after extraction
Acetaminophen	Reduction of Ce(IV) in acidic medium	1–10	Continuous flow mixing with chemiluminometric detector, fixed-time approach. Application in drug formulations
Acetaminophen (paracetamol)	Reaction with hexacyanoferrate(III), pH 8.5	0.5–15	FIA/SF fluorimetric system. Reaction rate approach (5.8 s). Application in pharmaceutical formulations
Albumin	Binding with picrate at pH 5.0	$10–70\,g\,L^{-1}$	Air-segmented continuous flow system with a picrate ISE flow detector. Fixed-time approach. Application in serum
Albumin	Reaction with bromcresol green	$0–68\,g\,L^{-1}$	Spectrophotometric monitoring, SF mixing, fixed-time (10 s) approach. Application in serum
Aldehydes (aromatic)	Condensation reaction with barbituric acid	$1 \times 10^{-6} - 5 \times 10^{-4}$ M	Spectrophotometric monitoring
Amines (primary and secondary)	Color formation by reacting with 1,2-naphthoquinone-4-sulfonate in the presence of Triton X-100, pH 10.5		Spectrophotometric monitoring, reaction rate (45–180 s) approach. Application in pharmaceutical formulations (amphetamine, ephedrine)
Amino acids	Reaction with ninhydrin at 80 °C	$(1–5) \times 10^{-5}$ M	Photometric monitoring at 570 nm. Absorbance versus time data from the kinetic region of reaction (1–3 half-lives) are fitted to a first -order model to predict total absorbance change at equilibrium
Amino acids and amines	Reaction with FDNB at pH 9 yielding fluoride ions	$1 \times 10^{-4} - 5 \times 10^{3}$ M	Monitoring with fluoride ISE. Initial-rate and fixed time approaches. Application in drug formulations
Amino acids, carbohydrates	Reaction with TNBS at pH 12	$3 \times 10^{-4} - 3 \times 10^{-3}$ M	Monitoring with TNBS ISE. Initial-rate or variable-time approaches
Ascorbic acid	Reaction with 2,6-dichlorophenol indophenol	$2 \times 10^{-7} - 2 \times 10^{-5}$ M	SF mixing, spectrophotometric monitoring (10 s)
Ascorbic acid	Rapid oxidation by mercury(II) to dehydroascorbic acid, which then reacts with *o*-phenylenediamine to form fluorescent quinoxaline	0.02–10	Spectrofluorimeter system with automated injector. Reaction rate measurement (20 s) after a delay time of 5 s. Application in drug formulations and juices
Ascorbic acid	Photochemical reaction with methylene blue	20–200	Various flow manifolds with irradiation of the reaction coil or the flow cell. Reaction rate and fixed time approaches. Application in drug formulations and fruit juices

(*continued overleaf*)

Table 3 (*continued*)

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Ascorbic acid	Reduction of toluidine blue	3–35	Spectrophotometric monitoring of decrease in absorbance, slope and fixed time approaches. Application in fruits and vegetables
Aspartame	Reaction with FDNB in the presence of CTAB micelles to yield colored DNB derivative	$(0.1–6) \times 10^{-4}$ M	Spectrophotometric monitoring (342 nm), FI/SF mixing, slope method. Application in beverages
Bromazepam	Complexation reaction with Fe(II) in acidic medium	$5 \times 10^{-4}–1.5 \times 10^{-3}$ M	Sequential injection technique, spectrophotometric monitoring (585 nm), fixed-time (100 s) approach. Application in drug formulations
Bromhexine hydrochloride	Coupling reaction of diazotized bromhexine derivative and NEDD	1.5–65	SF system, spectrophotometric monitoring at 495 nm. Initial reaction rate approach. Application to drug formulations without interference from antibiotics (penicillins and cephalosporins)
Carbamazepine	Oxidation with Ce(IV) in acidic medium	0.04–140	SF system with fluorimetric monitoring, initial rate approach. Application in serum after extraction
Carbimazole	Oxidation by dichromate or molybdate in acidic medium at 90 °C		Spectrophotometric measurement (585 or 650 nm), fixed-time (25 or 12 min) approach. Application in drug formulations
p-Chloranil	Alkaline hydrolysis to monochloranilate in a fast step and to chloranilate in a slow step	5–500	A chloranilate ISE monitors both anions and the potential change at the end of the first fast step is proportional to log *C* of *p*-chloranil. Application in fungicide preparation
Chlorpyrifos	Oxidative (H_2O_2) alkaline hydrolysis	$6.6 \times 10^{-6}–2 \times 10^{-2}$ M	FIA/SF system, spectrophotometric monitoring (328 nm), initial rate approach. Application in pesticide formulation
Choline and acetylcholine	Oxidation with Ce(IV) in acidic medium to yield fluorescent Ce(III)	$1.7 \times 10^{-5}–1 \times 10^{-3}$ M	Fluorimetric monitoring, reaction rate approach
Citric acid	Reaction with perbromate induced by Fe(II) at pH 4.75	96–960 μg	Monitoring with perbromate ISE. Fixed-time approach (6 min). Interference by tartaric and lactic acids
Corticosteroids	Reaction with tetrazolium blue in alkaline medium to form red formazan	mg dL^{-1} levels	Spectrophotometric monitoring (525 nm), SF mixing, slope approach after a delay time. Application in skin preparations
Creatinine	Jaffe reaction (alkaline picrate)	0.5–4 g L^{-1}	Monitoring with a picrate ISE. Fixed time (30 and 270 s). Application in serum and urine. Method free from optical interferences and bilirubin interference
Creatinine	Jaffe's reaction (picric acid in alkaline solution) at 37 °C	0.5–2 g L^{-1}	FIA, spectrophotometric monitoring, fixed-time approach. Application in urine. An alternative method uses FI/SF slope approach
Dehydroascorbic acid	Formation of mixture of methyl ketals by reacting with aqueous methanol-containing phosphate or citrate buffers. Ascorbic acid is masked using transition metal complexing agents	0–350 μM	Spectrophotometric monitoring at 346 nm. Methyl ketals were analyzed with fast atom bombardment mass spectrometry

Table 3 (*continued*)

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Diaquat	Charge-transfer complexation with cysteine	0.5–28	Spectrophotometric monitoring (495 nm), initial rate approach (60 s). Application in herbicides, waters, soils and potatoes
2,4-Dinitrophenyl-hydrazine	Monitoring of the decomposition of the unstable green product formed by reacting with piperidine in dimethyl sulfoxide	0–120	Spectrophotometric monitoring (640 nm)
Drugs of abuse (amphetamines, cocaine, cannabinoids)	Immunochemical reaction	20–300 $ng\,mL^{-1}$ (Amphetamine)	SF fluorescence polarization system, measuring variation of polarized fluorescence. Application in urine
EDTA, citrate	Displacement of fluoride from its aluminum complexes by EDTA, citrate and α-hydroxyacids	0.2–1.5 μmol EDTA, 0.1–1 μmol citrate	Monitoring of fluoride liberation with fluoride ISE. Initial rate ($\Delta E/\Delta t$) approach
Ethanol	Reaction with MnO_4^- in very acidic solutions	LOD 0.3% (v/v)	CL monitoring. Fitting of data with second order polynomial. Application in gin
Ethanol, other alcohols and carbohydrates	Reaction with dichromate in acidic medium (4.5 M H_2SO_4)	0.03–1.2 M	Monitoring with a dichromate ISE. Initial slope ($\Delta E/\Delta t$) approach. Application in beverages
Formaldehyde and hexamethylene-tetramine	Addition reaction of hydrogen cyanide to carbonyl group at pH 7.8. Hexamethylene-tetramine is determined after acid hydrolysis	60–300	Monitoring of rate of CN^- consumption with a cyanide ISE. Variable-time approach. Application in drug formulations
Furazolidone, furaltadone	Alkaline hydrolysis	1–20	Spectrophotometric monitoring, tangents approach. Application in pharmaceutical and feed products
Glucose	Reaction with periodate at pH 4.5	0.02–0.15 M	Monitoring with a perchlorate/periodate ISE. Variable time approach. Application to other carbohydrates
Glutathione (reduced)	Reaction of reduced glutathione with Cu(II) to yield Cu(I) which in turn reacts with neocuproine to form a colored complex	1.6–160 μM	Continuous addition of reagent (Cu(II)) technique. Spectrophotometric monitoring (455 nm). Application in blood
Glycerol	Reaction with periodate at pH 4.5	2.8×10^{-4}–0.12 M	Monitoring with periodate ISE. Slope or variable-time approaches. Application in soaps
Hb	Peroxidation of chlorpromazine by Hb	3–200 $mg\,L^{-1}$	Spectrophotometric monitoring (528 nm), reaction rate approach (3 min). Application in plasma and urine
α-Hydroxyacids	Reaction with periodate accelerated by Fe(II) or hexacyanoferrate(II)		Monitoring with periodate ISE. Fixed-time (90 s) approach
Isoniazid	Formation of hydrazone with 2-hydroxy-1-naphthaldehyde in acidic medium and consequent formation of a strongly fluorescent complex with scandium in weakly acidic medium	0–5	Fluorimetric monitoring. Fixed-time approach (1 min)

(*continued overleaf*)

Table 3 (*continued*)

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Maleimides	Reaction with hydroquinone and H_2O_2 in the presence of Cu(II)	0.5–15	Spectrophotometric monitoring, reaction rate approach (3 min)
Malonaldehyde	Reaction with thiobarbituric acid at 65 °C	1.1–50 $ng\,mL^{-1}$	Fluorimetric monitoring, reaction rate approach (400–500 s). Application in human serum
Malonaldehyde	Hantzsch reaction (with methylamine in the presence of propan-2-ol), 75 °C	0.5–2.8	Monitoring of fluorescence, reaction rate approach (1250–1450 s). Application in olive oil
mHb, Hb	Reaction with cyanide to produce CNmHb (monitoring of disappearance of mHb at 630 nm). To determine Hb the whole blood reacts with ferricyanide to produce mHb (monitoring of increase in absorbance at 630 nm)	5.6–56 $mmol\,L^{-1}$	SF (measurement time 250 ms) or centrifugal (measurement time 30 s) mixing. Absorbance data are fitted to first-order model by nonlinear regression method to calculate ΔA_∞ values
Methotrimeprazine, thioridazine	Reaction with Fe(III) in acidic solution at 50 °C	0.8–60 1.1–60	SF/diode-array system, reaction rate approach (5 s). Application in pharmaceutical formulations
Methyl parathion	Alkaline hydrolysis at 55 °C	2–30	Spectrophotometric monitoring at 400 nm. Initial rate and fixed time approaches. Application in formulations and surface waters
Morphine	Reaction with $Fe(CN)_6^{3-}$ at pH 10	15–925 $ng\,mL^{-1}$	Continuous addition of reagent method, fluorimetric monitoring. Application in urine after ion-exchange separation
Morphine	Reaction with FDNB at pH 9 at 35 °C to liberate fluoride ions	20–200 μM	Monitoring with fluoride ISE. Initial rate approach. Application in pharmaceutical formulations
Nitrazepam	Formation of azo dye with 1-naphthol after the acidic hydrolysis of the drug to 2-amino-7-nitrobenzophenone and diazotization	0.58–2.9	Continuous addition of reagent technique (addition of the coupling reagent at a rate of 35 $mL\,min^{-1}$ with stirring). Spectrophotometric monitoring at 600 nm. Application in drug formulations. Similar method for oxazepam
Nortriptylin	Reaction with 4-chloro-7-nitrobenzofurazan, pH 8.2 at 50 °C	0.12–60	SF system, spectrophotometric (470 nm) and fluorimetric monitoring. Reaction rate approach (30 s). Application in drug formulations
Organic peroxides and lipohydroperoxides	Mn(II)-catalyzed oxidation of 2-hydroxynaphthaldehyde thiosemicarbazone	0.08–1.6 μM	Fluorimetric monitoring, initial rate approach. Application in commercial oils
Oxalate	Reduction of permanganate catalysed by Mn(II) in acidic medium	0–20	Spectrophotometric monitoring of the first fast step of the reaction at 525 nm. SF mixing, initial rate approach
Phenolic (acetaminophen, isoxsuprine) and hydrazino (isoniazid) drugs	Reaction with FDNB in weakly alkaline medium to liberate fluoride catalyzed by CTAB micelles	$(1–50) \times 10^{-4}$ M	Automated FI system with a fluoride ISE in wall-jet configuration. Fixed-time approach (measurement of ΔE). Application in drug formulations
Phenols (*para* and *meta* substituted phenolic drugs)	Oxidative condensation with 1-nitroso-2-naphthol using Ce(IV) or Pb(IV) as oxidant	$(1–8) \times 10^{-4}$ M	Spectrophotometric monitoring (510, 540 nm), FI mixing, fixed-time approach. Application in drug formulations (acetaminophen, isoxsuprine, amoxicilline, etc.)

Table 3 (*continued*)

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Phenols (phenolic drugs)	Reaction with FDNB to yield fluoride ions in the presence of CTAB micelles	6×10^{-6}–7×10^{-4} M	Monitoring with a fluoride ISE, initial slope approach. Application in drug formulations (acetaminophen). An alternative method uses FI mixing and fixed-time approach
Phenothiazines	Oxidation by Fe(III) in very strong acidic medium to form unstable colored free radicals	10–250	Spectrophotometric monitoring (500–580 nm), FI mixing, fixed-time approach. Application in drug formulations and dissolution studies
1-PCC	Reaction with *p*-nitrobenzaldehyde and *o*-DNB. Starting with NaOH and quenching with ethanol after 45 min	0.1–0.7 μg	Spectrophotometric measurement (555 nm), fixed-time approach. Application in illicit samples
Salicylates, salicylamide, methyl salicylate	Trinder reaction (Fe(III) in weak acidic medium) yielding colored complex	25–1000	Spectrophotometric monitoring (540 nm), FI mixing, fixed-time approach. Application in drug formulations and dissolution studies
Sulfonamides	Azo dye formation between 1-naphthol and diazotized sulfonamide, pH 4.15	$(3–30) \times 10^{-6}$ M	Continuous addition of reagent (nitrite) technique. Spectrophotometric monitoring (470 nm), reaction rate approach. Application in drug formulations
Sulfonamides	Griess reaction (nitrosation and coupling with NEDD)	$(0.5–5) \times 10^{-5}$ M	SF mixing, photometric monitoring, variable time approach. Application in urine
Sulfonamides	Bratton–Marshall reaction (nitrosation and coupling with NEDD)	2–20	Spectrophotometric monitoring (545 nm), FI mixing, fixed-time approach. For the range 5×10^{-4}–5×10^{-3} M a FI pseudotitration approach is used. Application in serum, urine, feeds, drug formulations and dissolution studies
Tartaric acid	Reaction with periodate	0.4–120 μmol	Monitoring with a periodate ISE, initial rate ($\Delta E/\Delta t$) method. Application in drug formulations and impurities of citric acid
Tetracycline and oxytetracycline	Degradation with strong acids to the anhydro derivatives after a heating time of 30 min	8–40	Measurement of absorbance difference (fixed time approach) in comparison with alkaline solutions. Application in drug formulations
Theophylline	Formation of a 1 : 1 complex between the oxidation product of theophylline with Ce(IV) oxidant	1–250	SF mixing, fluorimetric monitoring, initial rate method (8 s). Application in drug formulations
Thiamine	Oxidation by Hg(II) in basic medium to form a fluorescent thiochrome	2×10^{-8}–1×10^{-4} M	Spectrofluorimeter with fixed-time digital ratemeter, initial rate approach. Application in vitamin and mineral preparations
Thiamine	Oxidation by ferricyanide in basic medium to fluorescent thiochrome	8.3×10^{-9}–7.5×10^{-6} M	Continuous addition of reagent technique. Fluorimetric monitoring, reaction rate approach. In an alternative method continuous flow mixing (fixed time approach) and chemiluminometric detection is used

(*continued overleaf*)

Table 3 (*continued*)

Analyte	Chemical system	Concentration range or LOD ($\mu g\,mL^{-1}$)	Comments
Thiols (thiolic drugs)	Reaction with FDNB in the presence of cationic micelles to yield fluoride ions		Monitoring with a fluoride ISE, fixed time approach. Application for thioglycolic acid in cosmetic formulations
Tryptophan	Reaction with formaldehyde at pH 10.8	$0-100\,nmol\,mL^{-1}$	Fluorimetric monitoring, fixed-time (60 s) approach. Application in foods and foodstuffs after alkaline hydrolysis with $Ba(OH)_2$
Urea	Reaction with biacetyl in acidic medium to yield a colored product in the presence of thiosemicarbazide and iron(III) to enhance and stabilize the color	0.5–15	SF system, spectrophotometric monitoring (530 nm). Initial rate approach. Application in serum
Uric acid	Analyte reduction of $Mn(CN)_5NO^{2-}$ at pH 10	$1-100\,\mu M$	Spectrophotometric monitoring (384 nm), tangent or fixed-time approach
Uric acid	Reaction with 1,1,3-tricyano-2-amino-prop-1-ene and H_2O_2, pH 9.2.	0.08–3	SF mixing, fluorimetric monitoring, Application in serum and urine
Uric acid	Reduction of Fe(III) in the presence of TPTZ		Photometric monitoring (593 nm), fixed-time approach (1 and 3 min). Application in undeproteinized serum
Vicinal glycols	Malaprade reaction (periodate)	0.14–7 mM	Monitoring with periodate ISE, variable-time approach
Vitamins (pyridoxin, thiamine, ascorbic acid, biotin)	Reaction with *N*-bromosuccinimide to yield bromide ions	$\mu g\,mL^{-1}$ levels	Monitoring with bromide ISE. Initial reaction rate approach. Application in drug formulations
Zineb (zinc-containing thiocarbamate-type pesticide and fungicide)	Complexation of zinc ions (produced from the acidic decomposition of zineb) with zincon	0.25–3	Continuous addition of reagent (zincon) technique. Spectrophotometric monitoring (620 nm), initial reaction rate approach. Application in zineb formulations and vine and olive leaves

mHb, methemoglobin; Hb, hemoglobin; PCC, piperidinocyclohexanecarbonitrile; CTAB, cetyltrimethylammonium bromide.

phosphate, silicate; carbaryl, carbofuran, diaquat, nitrilotriacetic acid.

- Sea and river waters: aluminum, bromide, calcium, magnesium, methyl parathion.
- Wastewaters: benzaldehyde, formaldehyde, mercury, nitrite.
- Soils and plants: aluminum (labile) (soils), carbaryl (vegetables), carbofuran (soils), diaquat (soils), zineb (vine and olive leaves).

5.3 Food Analysis

Several inorganic and organic species can be determined in a variety of food samples. Special treatment of the sample is required before the application of the kinetic method. The most frequently determined species are ascorbic acid, iron and phosphate.

The following species can be determined using uncatalyzed kinetic methods in food sample: calcium, iron, magnesium, crude protein, phosphorus, tryptophan, sulfonamides (animal feed), ascorbic acid (juices, green tea, potatoes, red pepper, spinach and strawberry), aluminum (spinach and tea leaves), aspartame (beverages), bromate (bread), butylated hydroxyanisole (vegetable oils), ethanol (gin and beverages), iron (wines), malonaldehyde (olive oil), nitrite (liver oil, cured meat), organic peroxides (oils), sucrose (milk and soft drinks), sulfite (beverages, biscuits, steamed buns), tryptophan (vegetables and fish).

5.4 Pharmaceutical Analysis

The quality control of drugs (assay of content, content uniformity, dissolution studies) requires a great number of determinations and automated kinetic methods are

attractive for this purpose. All kind of pharmaceutical formulations (capsules, injections, ointments, tablets, syrups) have been analyzed for several drugs using kinetic methods without serious interference from the excipients. Potentiometric kinetic methods using ISEs can be successfully applied in colored and cloudy sample solutions. The FIA has found tremendous use in routine pharmaceutical analysis using kinetic methods and has been successfully employed in dissolution studies of solid formulations (tablets and capsules), providing complete dissolution profiles.

The following drugs have been determined in pharmaceutical formulations: acetaminophen, aminoglycoside antibiotics, amoxicilline, amphetamine, ascorbic acid, benzocaine, bismuth, boric acid, bromazepam, bromhexine, bromide salts of drugs, carbimazole, catecholamines, chloride salts of drugs, chlorpromazine, corticosteroids, cyanocobalamin, ephedrine, furazolidone, furaltadone, hexamethylenetetramine, isoniazid, isoxsuprine, methotrimeprazine, morphine, nitrazepam, nitroglycerine, nitroprusside, nortriptylin, oxazepam, perphenazine, phenothiazines, procaine, salicylamide, sulfonamides, tartaric acid, tetracyclines, thiamine, thioridazine, theophylline, zinc.

ABBREVIATIONS AND ACRONYMS

BTB	Bromothymol Blue
CAR	Continuous Addition of Reagents
CF/SF	Continuous Flow/Stopped Flow
CL	Chemiluminescence
CTAB	Cetyltrimethylammonium Bromide
DCTA	Diaminohexanetetraacetic Acid
DNB	Dinitrobenzene
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Bis(aminoethyl)glycolether-*N*,*N*,*N*′,*N*′-tetraacetic Acid
FDNB	1-Fluoro-2,4-dinitrobenzene
FI	Flow-injection
FIA	Flow-injection Analysis
Hb	Hemoglobin
ISE	Ion Selective Electrode
LOD	Limit of Detection
mHb	Methemoglobin
NEDD	*N*-(1-Naphthyl)ethylenediamine Dihydrochloride
PCC	Piperidinocyclohexanecarbonitrile
PPD	*p*-Phenylenediamines
PR	Phenol Red
QDI	Quinone Diimine
SF	Stopped-flow
TNBS	Trinitrobenzenesulfonate
TPTZ	Tris-(2-pyridyl)-*s*-Triazine

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Automation in the Clinical Laboratory • Laboratory Instruments in Clinical Chemistry, Principles of

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction

Pharmaceuticals and Drugs **(Volume 8)**
Robotics and Laboratory Automation in Pharmaceuticals Analysis

Process Instrumental Methods **(Volume 9)**
Flow and Sequential Injection Analysis Techniques in Process Analysis

Electroanalytical Methods **(Volume 11)**
Ion-selective Electrodes: Fundamentals

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction • Catalytic Kinetic Determinations: Nonenzymatic • Data Treatment and Error Analysis in Kinetics • Differential Rate Determinations • Electrocatalysis-based Kinetics Determinations • Enzymatic Kinetic Determinations

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics • Luminescence-based Kinetic Determinations

REFERENCES

1. H.A. Mottola, *Kinetic Aspects of Analytical Chemistry*, New York, Wiley, 1988.
2. D. Perez-Bendito, M. Silva, *Kinetic Methods in Analytical Chemistry*, Chichester, Ellis Horwood, 1988.
3. M.D. Love, H.L. Pardue, 'Systematic Comparison of Data-Processing Options for Kinetic-based Single-Component Determinations of Non-catalysts', *Anal. Chim. Acta*, **299**, 195–218 (1994).
4. H.L. Pardue, 'Kinetic Aspects of Analytical Chemistry', *Anal. Chim. Acta*, **216**, 69–107 (1989).
5. C.E. Efstathiou, M.A. Koupparis, T.P. Hadjiioannou, 'Application of Ion-selective Electrodes in Reaction Kinetics and Kinetic Analysis', *Ion-selective Electrode Rev.*, **7**, 203–259 (1985).
6. H.A. Mottola, 'Kinetic Determination of Reactants Utilizing Uncatalyzed Reactions', *Anal. Chim. Acta*, **280**, 279–287 (1993).
7. E.H. Hansen, 'Exploiting Kinetic-based Flow-injection Methods for Quantitative Chemical Assays', *Anal. Chim. Acta*, **261**, 125–136 (1992).

8. M. Silva, 'Recent Strategies in Automated Reaction Rate Based Determinations', *Analyst*, **118**, 681–688 (1993).
9. H.A. Mottola, D. Perez-Bendito, H.B. Mark Jr, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **60**, 181R–200R (1988); **62**, 441R–461R (1990).
10. H.A. Mottola, D. Perez-Bendito, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **64**, 407R–428R (1992); **66**, 131R–162R (1994); **68**, 257R–289R (1996).
11. S.R. Crouch, T.F. Cullen, A. Scheeline, E.S. Kirkor, 'Kinetic Determinations and Some Kinetic Aspects of Analytical Chemistry', *Anal. Chem.*, **70**, 53R–106R (1998).

Liquid Chromatography: Introduction

John G. Dorsey
Florida State University, Tallahassee, USA

Liquid chromatography is more than likely *the* most used method of analysis for compounds in solution. Entire industries depend on the separation power of this modern analytical technique; walking through any pharmaceutical company in the world will show a liquid chromatograph on almost every bench! Yet chromatography by its nature can be thought of as a necessary evil. That is, it is a dilution technique which takes some length of time to occur. In other words, we apply a certain concentration of a sample to the top of a column, wait for seconds or minutes, and then try to detect a lower concentration of the components in the mixture! The rationale for this process is the elegant separations possible. If the other analytical techniques were selective enough, there would be no need for chromatographic separations. This is not likely to occur anytime soon, except for the simplest mixtures. It is still easier to detect a lower concentration of a pure component in a solvent stream than to measure it selectively in a complex mixture. Chromatography can also be thought of as a violation of the second law of thermodynamics. Mother Nature favors mixtures and we are asking her to let us perform a separation. The price she extracts is in either concentration or recovery, and we must pay in one currency or the other.

Since its invention in 1906 by Tswett, liquid chromatography has changed dramatically in terms of understanding and controlling the chemistry of the separation process, yet it has also remained virtually unchanged in its form. A translated paragraph from Tswett's first publication in a Russian botany journal is fascinating:

> ''If a petroleum ether solution of chlorophyll is filtered through a column of an adsorbent (I use mainly calcium carbonate which is stamped firmly into a narrow glass tube), then the pigments according to the elution sequence are resolved from top to bottom into various colored zones, since the stronger adsorbed pigments displace the weaker adsorbed ones and force them farther downwards. This separation becomes practically complete if, after the pigment solution has flowed through, one passes a stream of pure solvent through the adsorbent column. Like light rays in the spectrum, so the different components of a pigment mixture are resolved on the calcium carbonate column according to a law and can be estimated on it qualitatively and also quantitatively. Such a preparation I term a chromatogram and the corresponding method, the chromatographic method.''

We have instrumentalized the technique, miniaturized the technique, and have learned to control the chemistry of the mobile and stationary phases, yet in the simplest terms the technique as practiced today has changed very little from that earliest description.

The articles that follow encompass all of the important methods and concepts of this separation technique, and are authored by some of the world's leaders. From stationary phase chemistry to separation mechanisms to electrophoresis, thin-layer chromatography and supercritical fluid chromatography, we have attempted to cover all of the important areas. Entire volumes could be (and have been!) written on each of these subjects. By necessity we have limited the coverage to the basics while maintaining breadth.

Affinity Chromatography is one of the most selective of the separation methods, and can be thought of as ''digital'' chromatography. That is, a column is designed to retain a specific analyte, or very closely related analytes, while everything else in the sample is unretained. The analyte of interest is then ''released'' from the column and quantitated. While simple in description, the design of highly selective columns is very challenging!

The article **Biopolymer Chromatography** shows the power of the technique for the analysis of biological compounds. Biology and chemistry are merging, and the chemical analysis of complex biological systems has opened entire new frontiers to analytical chemists. Separations of polypeptides, proteins, and even cells are possible, but challenging. A protein will interact differently with a stationary phase in its native and denatured form, may denature upon contact with some stationary phases, and gives different retention times and peak shapes depending upon its conformation. In fact, column chromatography has been used to measure denaturation rate constants!

Capillary Electrophoresis is the newest high-resolution separation method, and grew very rapidly through the 1980s and 1990s. However, the sale of instruments has remained virtually flat since the introduction of commercial instrumentation in 1989, while the growth of publications has been almost exponential. Is this just a toy for academicians? It is a very important tool for the bio-tech industry, for DNA sequencing, and other specialized separations, but will likely never be the replacement for liquid chromatography that some originally thought.

The human body is chiral, and because of this both the study of life processes and the design of efficacious pharmaceutical compounds require efficient separations of racemic mixtures. The separation of optical isomers has been a great challenge for separation chemists. Yet that most sensitive of ''analytical methods'', the human body, can tell *immense* differences between the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

two forms. The thalidomide tragedy in Europe in the 1960s is but one example. While one enantiomer of this compound was efficacious against morning sickness, the other was a teratogen, and since the compound was made as a racemic mixture, hundreds of children were born seriously deformed. The article **Chiral Separations by High-performance Liquid Chromatography** describes what we have learned, and what is left to accomplish.

A fundamental article, **Column Theory and Resolution in Liquid Chromatography**, describes the fundamental processes that occur inside the column, and the terminology that is used to describe the quality of separations.

Most real separations cannot be performed with a single composition mobile phase. That is, the range of solutes is so broad that any given mobile phase will be too ''strong'' for some solutes, providing little or no resolution, and will be too ''weak'' for others, letting them move through the column too slowly, if at all. The article **Gradient Elution Chromatography** describes this most used solution to the ''General Elution Problem'', and shows both theory and methodology for speeding separations.

While most people think of chromatography for the separation of neutral organic molecules, there is also much need for the analysis of charged species; from simple inorganic ions in plating plants and waste-treatment processes, to charged biopolymers. The article **Ion Chromatography** discusses the development of this very important chromatographic method, both in terms of separations but also in terms of detection of compounds that may have no chromophore.

Capillary electrophoresis is an efficient method for the separation of charged species that migrate in an applied electric field. **Micellar Electrokinetic Chromatography** describes an elegantly simple means of applying this technique for the separation of neutral species. Simply put, if a charged micelle can be made to remain more or less stationary in the electroosmotically flowing buffer, then neutral species will be separated based on differential partitioning to the micelle! This development opened entire new fields of separations.

Liquid chromatography uses organic solvents as components of the mobile phase. Driving forces for reduction of solvent use include environmental concerns and simply the economics of purchasing. **Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography** describes the miniaturization of the technique. Other advantages besides solvent savings exist as well, and the technique will slowly move in this direction.

The chemistry of the interaction of the solute with the mobile and stationary phases is how these elegant separations are developed. **Normal-phase Liquid Chromatography** uses a polar stationary phase and nonpolar liquid mobile phases for the separation of polar solutes. While this was the first method described by Tswett, usage of this method has dwindled but it is still a very powerful separation technique.

Reversed Phase Liquid Chromatography is by far the most practiced liquid separation method. The name arose from the reversal in elution order of most solutes compared to normal phase separations, and here a nonpolar stationary phase is used with polar mobile phases, typically water with an added polar organic solvent. Reversed-phase liquid chromatography is *the* workhorse method of the pharmaceutical industry.

The design of selective stationary phases is what allows the rapid, efficient separations that are achievable today. **Silica Gel and its Derivatization for Liquid Chromatography** describes how these phases are prepared, and offers challenges for better design of robust, reproducible phases.

While most analysts think of mobile phases as either gases or liquids, an almost 40-year-old technique exists using a mobile phase of a fluid taken above its critical point. **Supercritical Fluid Chromatography** describes this method, and shows the advantages offered when a fluid (mobile phase) has the solvating power of liquids, and diffusion coefficients closer to gases.

While column chromatography is a very powerful and often practiced method, one of its major limitations is that it is a ''serial'' technique. That is, one sample at a time is all that can be analyzed. **Thin-layer Chromatography** describes a liquid phase separation method performed on a flat plate, so that parallel ''lanes'' are possible, and separations can be run in parallel. While offering *far* lower resolving power than column methods, this is still a very heavily practiced method for the separation of simple mixtures, or for screening large numbers of samples.

Liquid chromatography is used for more than chemical analysis. The technique can be performed on both laboratory preparative scales and on industrial scales. Here the goal is *throughput*, the production of the greatest amount of pure material in the shortest time, and while the fundamentals of the separation process are essentially the same, the goal mandates that the operational aspects are quite different. Usually the sample (feed) is grossly overloaded onto the column, and the technique is referred to as *nonlinear chromatography*. Since an interfacial partitioning process is the most used mechanism of separation, the experiment can also be turned around, and the chromatography column can be used to *measure* these partitioning processes. Both thermodynamic and kinetic information can be obtained from a chromatography column.

Still, it is chemical analysis that drives further development of this technique. We will learn to better control the separation, we will learn to do it faster and with better limits of detection, but I doubt if we deviate much from the description of Tswett!

Affinity Chromatography

David S. Hage
University of Nebraska, Lincoln, NE, USA

1 Introduction 1
1.1 Definition and Overview of Affinity Chromatography 1
1.2 History of Affinity Chromatography 1
2 Theory of Affinity Chromatography 2
3 Basic Components of Affinity Systems 3
3.1 Affinity Ligands 3
3.2 Support Materials 4
3.3 Immobilization Methods 5
3.4 Application and Elution Conditions 6
4 Types of Affinity Chromatography 8
4.1 Bioaffinity Chromatography 8
4.2 Immunoaffinity Chromatography 9
4.3 Dye-ligand Affinity Chromatography 12
4.4 Immobilized Metal Ion Affinity Chromatography 12
4.5 Analytical Affinity Chromatography 12
4.6 Miscellaneous Ligands and Methods 14
5 Future Developments 15
Acknowledgments 15
Abbreviations and Acronyms 15
Related Articles 15
References 16

Affinity chromatography is a type of liquid chromatography that makes use of biological-like interactions for the separation and specific analysis of sample components. This article discusses the basic principles behind the theory and use of this method. Various factors that are important in the development of an affinity chromatographic system are considered, such as the choice of affinity ligand, support material, immobilization method, and elution conditions. A survey of the different types of affinity chromatography is also presented; this includes bioaffinity chromatography, immunoaffinity chromatography, dye–ligand affinity chromatography, immobilized metal affinity chromatography, and analytical affinity chromatography. In addition, some areas of expected future growth are considered, such as the use of molecular imprints or aptamers as ligands in affinity columns, and the coupling of affinity chromatography with other analytical techniques.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Definition and Overview of Affinity Chromatography

Affinity chromatography is a liquid chromatographic technique that uses a well-defined binding agent as a stationary phase for the selective purification or analysis of sample components.[1–10] The retention of solutes in this method is generally based on the same types of specific, reversible interactions as those that occur in biological systems, such as the binding of an enzyme with a substrate or an antibody with an antigen. These interactions are used in affinity chromatography by immobilizing or adsorbing one of a pair of interacting molecules onto a solid support and placing this modified support into a column. The immobilized or adsorbed molecule that is used as the specific binding agent is known as the *affinity ligand*. The presence of this ligand results in a stationary phase that has the ability to bind to the complementary molecule in samples that are applied to the affinity column.

The wide range of selective ligands that are available in nature has made affinity chromatography a useful tool in areas such as biochemistry, pharmaceutical science, clinical chemistry, and environmental testing for the purification or analysis of compounds in complex sample mixtures. Because of the strong and relatively specific binding that characterizes many affinity ligands, solutes that are quantitated or purified by these ligands can often be separated with little or no interferences from other sample components. In many cases the solute of interest can be isolated in only one or two steps, with purification yields of one hundred- to several thousandfold being common.[2–5] Similar selectivity has been observed in numerous reports that have used affinity chromatography as an analytical tool for the detection and quantitation of specific compounds in such samples as serum, plasma, urine, foods, cell cultures, and water or soil extracts.[5,10,11]

1.2 History of Affinity Chromatography

The first reported use of affinity chromatography is generally recognized as having been in 1910, when Starkenstein used an insoluble starch support to purify the enzyme α-amylase.[12] There was some speculation in the mid-1920s about the value of using ''fixed partners'', or affinity ligands, as a general scheme for the isolation of biologically active compounds,[13] but there was little activity in this field for several decades after Starkenstein's work. In 1951, Campbell et al. were the first to use artificially immobilized ligands in an affinity column, when they attached antigens to diazotized *p*-aminobenzyl cellulose for the purification of antibodies.[14] After this work came several other applications, including reports

on the use of affinity chromatography for the purification of tyrosinase,[15] flavokinase,[16] and avidin.[17]

However, affinity chromatography did not see widespread use until the late 1960s. It was during this time period that there were several important developments that led to a greater flexibility and ease of use for affinity methods. One of these breakthroughs involved the introduction of better supports for affinity chromatography, such as beaded agarose.[18] Another development that came about at this time was a report on the CNBr method for coupling proteins and other ligands to affinity supports through primary amine groups.[19] Both of these advances were used in 1968 by Cuatrecasas, Wilchek, and Anfinsen to produce what are now considered the first ''modern'' affinity chromatographic separations.[20]

Beginning in the 1970s, affinity chromatography saw rapid growth as a purification tool for biochemicals,[1–10] but most of this work continued to employ soft gels like agarose for the affinity support.[5,10] These gels tended to limit the usefulness of affinity chromatography as an analytical tool because of their ability to work at only relatively low flow-rates and operating pressures which, in turn, resulted in fairly long run times.[5,10] More rapid separations were made possible by the development of alternative supports, such as the diol-bonded silica matrix reported in 1976 by Regnier and Noel.[21] These newer materials could withstand the same pressures and flow-rates as used in high-performance liquid chromatography (HPLC) but still provided a surface that could be modified for ligand attachment and that provided low nonspecific binding for most biological compounds.[5,10] In 1978, Ohlson et al. combined these HPLC-compatible materials with affinity ligands to produce a method known as high-performance affinity chromatography (HPAC).[22] It was this approach that now made it more feasible to use affinity chromatography as an analytical tool for the characterization of biological and non-biological samples.[5,10]

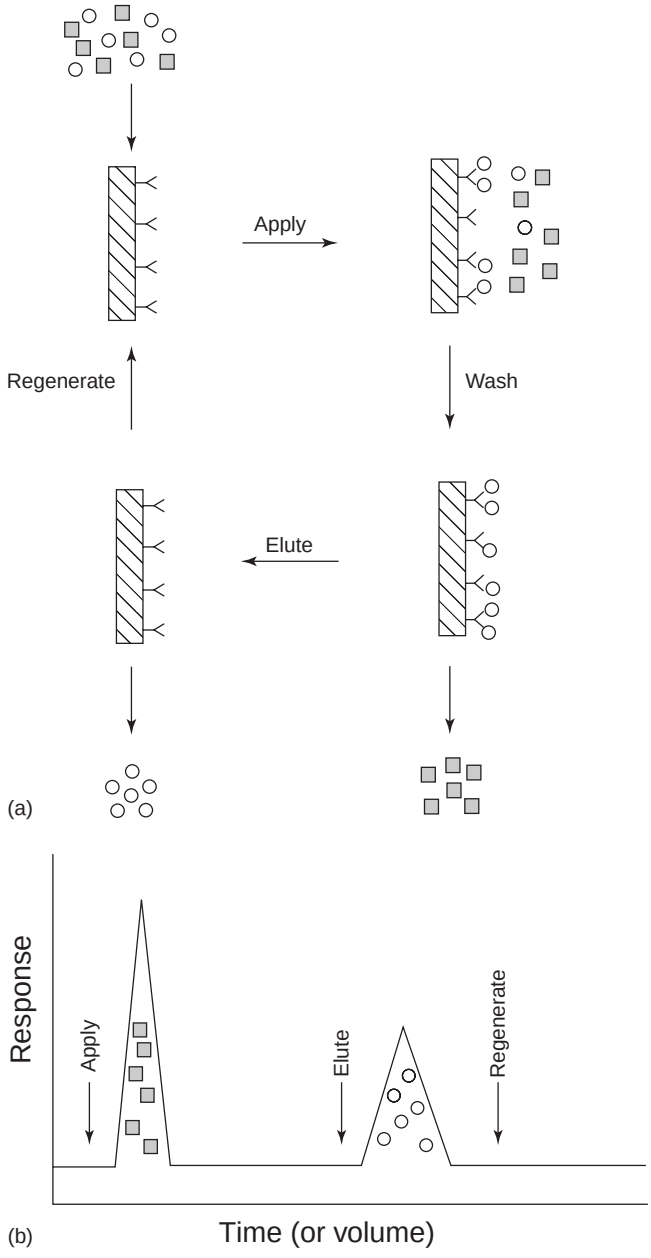

Figure 1 (a) Typical separation scheme for affinity chromatography and (b) a representative model chromatogram. The open circles represent the test analyte and the squares represent other, nonretained sample components. (Reproduced by permission of the Clinical Ligand Assay Society from D.S. Hage, *J. Clin. Ligand Assay*, **20**, 293–301 (1997).[48])

2 THEORY OF AFFINITY CHROMATOGRAPHY

The most common method used for performing affinity chromatography is a step-wise elution scheme, as shown in Figure 1. In this approach, a sample containing the compound of interest is injected onto the affinity column in the presence of a mobile phase that has the right pH, ionic strength, and solvent composition for solute-ligand binding. This solvent, which represents the weak mobile phase of the affinity column, is called the *application buffer*. As the sample passes through the column in the application buffer, compounds which are complementary to the affinity ligand will bind. However, owing to the high selectivity of this interaction, other solutes in the sample will tend to elute from the column nonretained. After all nonretained components have been washed from the column, the retained solutes are then eluted by applying a solvent that displaces them from the column or that promotes dissociation of the solute–ligand complex. This solvent, which represents the strong mobile phase for the column, is known as the *elution buffer*. As the solutes of

interest elute from the column, they are either measured directly or collected for later use. The application buffer is then reapplied to the system and the column is allowed to regenerate prior to the next sample injection.[5,10]

A number of factors are important in determining the retention and elution of a compound in an affinity column. These factors include the strength of the solute–ligand interaction, the amount of immobilized ligand present in the column, and the kinetics of solute–ligand association and dissociation. In the case where a solute (A) has single-site binding to a ligand (L), Equations (1–3) can be used to describe the interactions between the solute and ligand in an affinity column.

$$A + L \underset{k_d}{\overset{k_a}{\rightleftharpoons}} A - L \tag{1}$$

$$K_a = \frac{k_a}{k_d} \tag{2}$$

$$= \frac{\{A - L\}}{[A]\{L\}} \tag{3}$$

In these relationships, K_a is the association equilibrium constant for the binding of A with L, A–L is the resulting solute–ligand complex, [A] is the mobile phase concentration of A at equilibrium, and {L} or {A–L} represent the surface concentrations of the ligand or solute–ligand complex, respectively. The term k_a is the second-order association rate constant for solute–ligand binding and k_d is the first-order dissociation rate constant for the solute–ligand complex. At equilibrium, the retention of solute in the above system can be described by Equations (4) and (5),

$$k = \frac{K_a m_L}{V_M} \tag{4}$$

$$= \left(\frac{t_R}{t_M} \right. - 1 \tag{5}$$

where k is the retention factor for the injected solute, t_R is the solute's retention time, and t_M is the void time of column.[10,23] The term m_L represents the moles of active ligand in column and V_M is the column void volume.

Equations (4) and (5) indicate that the retention factor for a solute, or its retention time, will depend on both the strength of solute–ligand binding (i.e. the association equilibrium constant, K_a) and the amount of ligand in the column (i.e. m_L/V_M). This retention factor is often very large in the presence of the application buffers that are used in affinity chromatography. For instance, affinity columns that contain polyclonal antibodies as ligands (which generally have K_a values in the range of 10^8–$10^{10}\ M^{-1}$) can give rise to retention factors of 10^3 or greater under physiological conditions.[10] The only way in which solutes can be eluted from these columns in a reasonable amount of time is to lower the effective value of K_a for solute–ligand binding; in fact, this is the main role of the elution buffer for an affinity column.

Although the step gradient, or "on/off", elution mode shown in Figure 1 is the most common way of performing affinity chromatography, it is also possible to carry out affinity methods by using isocratic elution. This second approach can be employed if the value of a solute's association equilibrium constant (and retention factor) for the column are sufficiently small to allow elution on the minute-to-hour time scale. Fast association and dissociation kinetics are also desirable in this mode in order to allow a large number of solute–ligand interactions to occur as an analyte travels through the column, thus producing good plate heights and narrow peaks. This approach, sometimes called *weak-affinity chromatography*, is fairly easy to perform if a solute's K_a value is less than or equal to $10^4\ M^{-1}$, and has even been employed on some systems with association constants up to $10^6\ M^{-1}$.[24–26]

3 BASIC COMPONENTS OF AFFINITY SYSTEMS

3.1 Affinity Ligands

The key factor in determining the success of any affinity separation is the type of ligand that is used within the column. A number of ligands that are commonly used in affinity chromatography are listed in Table 1. Many of these ligands are of biological origin, but a wide range of natural and synthetic molecules of nonbiological origin can also be used. However, all of these compounds can be placed into one of two categories: (1) high-specificity ligands and (2) general, or group specific, ligands.[5,10]

The term *high-specificity ligands* refers to compounds which bind to only one or a few closely related molecules. This type of ligand is used in chromatographic systems where the goal is to analyze or purify a specific solute. Typical high-specificity ligands include antibodies (for binding antigens), substrates or inhibitors (for separating enzymes), and single-stranded nucleic acids (for the retention of a complementary sequence). As this list suggests, most high-specificity ligands tend to be biological compounds. These ligands also tend to have relatively large association constants for solutes and are generally eluted by a step gradient.

General, or group-specific, ligands are compounds that bind to a family or class of related molecules. These ligands are used in methods where the goal is to isolate a class of structurally similar solutes. General ligands can be of either biological or non-biological origin. Examples given in Table 1 include proteins A and G, lectins, boronates, triazine dyes, and immobilized metal chelates.

Table 1 Common ligands used in affinity chromatography

Type of ligand	Examples of retained compounds
High-specificity ligands	
Antibodies: Monoclonal or polyclonal antibodies, Fab fragments, $F(ab)_2$ fragments, or related antibody fragments	Complementary antigens, such as drugs, hormones, peptides, proteins, viruses, cell components, and pesticides
Enzyme-binding agents: Inhibitors, substrates, coenzymes, cofactors, anti-enzyme antibodies	Complementary enzymes
Nucleic acids: Single-stranded DNA or RNA, double-stranded DNA or RNA, nucleotides	Complementary nucleic acids, DNA/RNA-binding proteins or enzymes
General ligands	
Lectins: Concanavalin A, wheat germ agglutinin, jacalin, lectins from lentil, pea, peanut, and soybean, lectins 60 and 120 from *Ricinus communis*, phytohemagglutinins E_4 and L_4	Small sugars, polysaccharides, glycoproteins, glycolipids, glycoconjugates
Antibody-binding proteins: Protein A, protein G, protein A/G, protein B, mannan-binding protein	Intact antibodies and Fc fragments
Boronates: Boronic acid, phenylboronic acid	Sugars and polysaccharides, catechols, glycoproteins, glycolipids, glycoconjugates
Synthetic dyes: Cibacron Blue F3GA, Procion Blue MX-3B, Procion Blue MX-R, Procion Red HE-3B, thymol blue, phenol red	Dehydrogenases, kinases, albumin, α-fetoprotein, CoA-dependent enzymes, hydrolases, nucleases, polymerases, synthetases, transferases, immunoglobulin G
Metal chelates: Metal complexes of imino-diacetic acid, carboxymethylaspartic acid, tris-carboxymethylethylenediamine, tris(2-aminoethyl)amine, or dipicolylamine	Metal-binding amino acids, peptides, and proteins

Many of the compounds have weaker binding for solutes than that seen with high-specificity ligands; however, there are exceptions to this. For example, protein A has an equilibrium association of over $10^8\ M^{-1}$ for some types of antibodies.[27] Also, some molecules which are usually considered to be high-specificity ligands, such as antibodies, can be used to retain an entire class of solutes if they recognize a feature that is common to all of the desired analytes.[11,28]

3.2 Support Materials

Another important factor to consider in affinity chromatography is the material used to hold the ligand within the column. Table 2 lists some examples of common supports found in affinity columns. Ideally, this support material should have low nonspecific binding for sample components but should be easy to modify for ligand attachment. This material should also be stable under the flow-rate, pressure, and solvent conditions to be employed in the analysis or purification of samples. In addition, the support should be readily available and simple to use in method development.[1–10]

Table 2 Common support materials used in affinity chromatography[8,10]

Carbohydrate-based supports	Agarose and cross-linked agarose; cellulose and modified cellulose
Modified inorganic supports	Controlled-porosity glass; glass beads; silica
Synthetic organic-based supports	Agarose/acrylamide copolymers; azalactone beads (Emphaze™); dextran/acrylamide copolymers; hydroxylated polystyrene perfusion media (POROS™); polyacrylamide derivatives; polyethersulfone matrix; polymethacrylate derivatives

Depending on what type of support material is being used, affinity chromatography can be characterized as either a low- or high-performance technique. In *low-performance (or column) affinity chromatography*, the support is usually a large-diameter, nonrigid gel. Many of the carbohydrate-based supports and synthetic organic materials listed in Table 2 fall within this category. An example of a separation based on low-performance affinity chromatography is shown in Figure 2. The low back-pressure of these supports means that they can often be operated under gravity flow or with a peristaltic pump. This makes these gels relatively simple and inexpensive to use for affinity purifications. Disadvantages of these materials include their slow mass transfer properties and their limited stability at high flow-rates and pressures. These factors limit the usefulness of these supports in analytical applications, where both rapid and efficient separations are often desired.[5,10]

In HPAC, the support consists of small, rigid particles capable of withstanding the high flow-rates and/or pressures that are characteristic of HPLC.[5,10,22] Examples of affinity supports that are suitable for work under these conditions include modified silica or glass, azalactone beads, and hydroxylated polystyrene media. The mechanical stability and efficiency of these supports allows them to be used with standard HPLC equipment. Although the need for more sophisticated instrumentation does make HPAC more expensive to perform than low-performance affinity chromatography, the better speed and precision of this technique makes it the method of choice for analytical applications,[5,10,11] such as that shown in Figure 3.

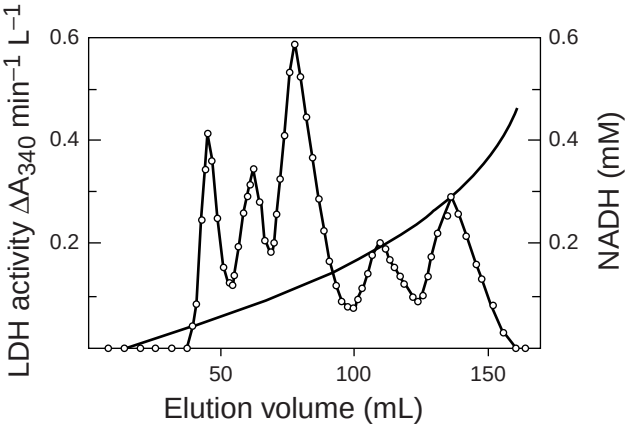

Figure 2 Example of a separation of lactate dehydrogenase (LDH) isoenzymes by low-performance affinity chromatography on an immobilized AMP-analog column and in the presence of a mobile phase gradient of increasing NADH concentration. (Reprinted from P. Brodelius, K. Mosbach, 'Separation of the Isoenzymes of Lactate Dehydrogenase by Affinity Chromatography using an Immobilized AMP-analogue', *FEBS Lett.*, **35**, 223–226, Copyright (1973), with permission from Elsevier Science.)

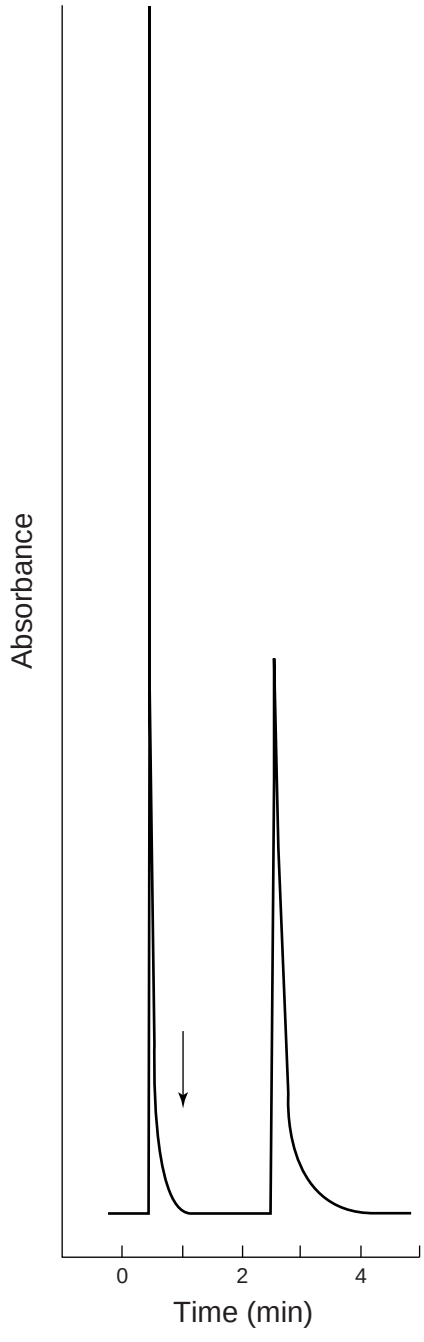

Figure 3 High-performance affinity chromatographic separation of concanavalin A (peak 2) from bovine serum albumin (peak 1) on an immobilized glucosamine column. The concanavalin A was eluted by applying at the arrow an elution buffer that contained methyl α-D-mannopyranoside. (Reprinted from R.R. Walters, 'High-performance Affinity Chromatography', *Trends Anal. Chem.*, **2**, 282–285, Copyright (1983), with permission from Elsevier Science.)

3.3 Immobilization Methods

A third item to consider in using affinity chromatography is the way in which the ligand is attached to the

solid support, or the *immobilization method*. Several techniques are available for this (see Table 3). For a protein or peptide, this generally involves coupling the molecule through free amine, carboxylic acid, or sulfhydryl residues present in its structure. Immobilization of a ligand through other functional sites (e.g. aldehyde groups produced by carbohydrate oxidation) is also possible. All immobilization methods involve at least two steps: (1) an *activation step*, in which the support is converted to a form that can be chemically attached to the ligand, and (2) a *coupling step*, in which the affinity ligand is attached to the activated support. With some techniques a third step, in which remaining activated groups are removed, is also required.[8,10]

The choice of immobilization method is important since it can affect the actual or apparent activity of the final affinity column. If the correct procedure is not used, a decrease in ligand activity can result from multisite attachment, improper orientation and/or steric hindrance (see Figure 4). *Multisite attachment* refers to the coupling of a ligand to the support through more than one functional group, which can lead to distortion and denaturation of the ligand's active region. This effect can be minimized by coupling the ligand through functional groups that occur in only a few places in its structure or by using a support with a low density of activated sites. *Improper orientation* can also lead to a loss in ligand activity but can be avoided by coupling the ligand through groups that are distant from its active region. If the actual location of the ligand's active site is not known or if no appropriate functional groups are available, it is sometimes possible to minimize this problem by empirically varying the procedure or conditions used for ligand attachment. *Steric hindrance* refers to the loss of ligand activity due to the presence of a nearby support or neighboring ligand molecules.[5] Steric hindrance produced by neighboring ligands can be minimized by using a low ligand coverage, while steric hindrance produced by the support can be reduced by adding a *spacer arm*, or tether, between the ligand and supporting material. The presence of a spacer arm is particularly important when using small ligands for the retention of large analytes. Examples of common spacer arms used in affinity chromatography include 6-aminocaproic acid, diaminodipropylamine, 1,6-diaminohexane, ethylenediamine and succinic anhydride.[8]

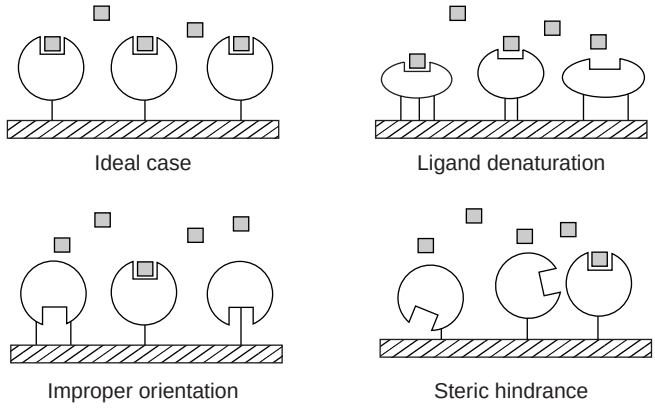

Figure 4 Illustration of multisite attachment, improper orientation, and steric hindrance effects in affinity chromatography.

Table 3 Immobilization methods for affinity chromatography[8,10]

Amine-reactive coupling techniques
Azalactone method (for Emphaze™); CNBr method; CDI method; divinylsulfone method; epoxy (bisoxirane) method; EDC method; FMP method; NHS method; Schiff base (reductive amination) method; tresyl chloride/tosyl chloride method
Sulfhydryl-reactive coupling techniques
Azalactone method (for Emphaze™); divinylsulfone method; epoxy (bisoxirane) method; iodoacetyl/bromoacetyl methods; maleimide method; pyridyl disulfide method; TNB-thiol method; tresyl chloride/tosyl chloride method
Carboxylate-reactive coupling techniques
EDC method
Hydroxyl-reactive coupling techniques
Cyanuric chloride method; divinylsulfone method; epoxy (bisoxirane) method
Aldehyde-reactive coupling techniques
Hydrazide method; Schiff base (reductive amination) method
Nucleic acid-coupling techniques
CDI method; carboxymethyl method (for cellulose); CNBr method; cyanuric chloride method; diazonium (azo) method; epoxy (bisoxirane) method; hydrazide method (for RNA); on-support synthesis; Schiff base (reductive amination) method (for RNA); UV irradiation (for cellulose)

CNBr, cyanogen bromide; CDI, *N*,*N*′-carbonyl diimidazole; EDC, ethyldimethylaminopropyl carbodiimide; FMP, fluoro methylpyridinium toluenesulfonate; NHS, *N*-hydroxysuccinimide ester; UV, ultraviolet.

3.4 Application and Elution Conditions

The application buffer is another parameter that must be considered in the use of an affinity column. Most application buffers in affinity chromatography are solvents that mimic the pH, ionic strength and polarity experienced by the solute and ligand in their natural environment. Any cofactors or metal ions required for solute–ligand binding should also be present in this solvent. Under these conditions, the solute will probably have its highest association constant for the ligand and its highest degree of retention on the column. The proper choice of the application buffer can also help to minimize any nonspecific binding due to undesired sample components. For example, coulombic interactions between

solutes and the support can often be decreased by altering the ionic strength and pH of the application buffer. In addition, surfactants and blocking agents may be added to the buffer to prevent nonspecific retention of solutes on the support or affinity ligand.

The activity of the immobilized ligand should be considered in determining how much sample can be applied to the affinity column with each use. A rough indication of the maximum possible column binding capacity can be made by assaying the total amount of ligand present or by actually measuring the ligand's binding activity. With some affinity systems, it is possible to see a large amount of nonretained solute during the application step even when the amount of injected sample is significantly less than known column binding capacity.[29] This phenomenon, known as the *split-peak effect*, is caused by the presence of slow adsorption and/or mass transfer kinetics within the column (see Figure 5). This effect has been reported with many types of affinity ligands,[29–33] but tends to occur with high-performance supports because of the more rapid flow-rates often used with these materials. Ways in which this effect can be minimized include reducing the flow-rate used for sample injection, increasing the column size, or placing a more efficient support within the column. In some cases, changing the immobilization method may help by providing a ligand with more rapid binding kinetics.[29]

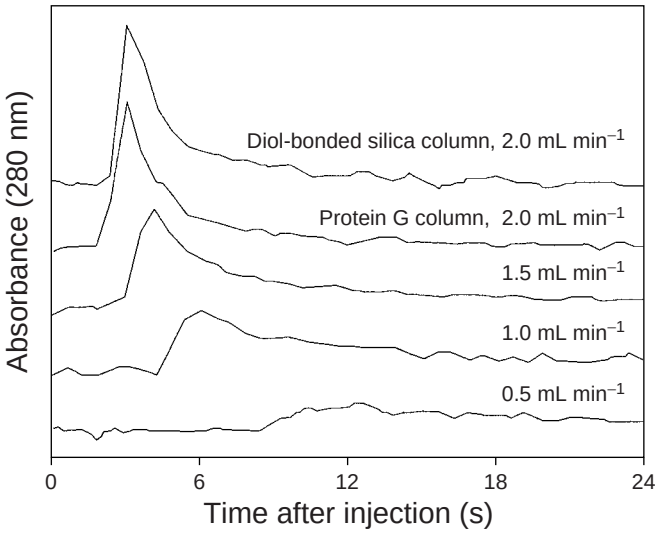

Figure 5 Nonretained (or free) fractions observed in split-peak measurements for the injection of rabbit immunoglobulin G onto a protein G affinity column at various sample application flow-rates. The top chromatogram shows the elution of rabbit immunoglobulin G on a column containing only diol-bonded silica, which has no binding for this analyte. (Reprinted from J.G. Rollag, D.S. Hage, *J. Chromatogr. A*, **795**, 185–198, Copyright (1998), with permission from Elsevier Science.[33])

The conditions used for removal of retained solutes is another item that should be considered in the design of an effective affinity separation. Just as the application conditions are selected to maximize specific solute–ligand interactions, the elution conditions are chosen to promote fast or gentle removal of solute from the column. The elution buffer used in affinity chromatography can be either a solvent that produces weak solute–ligand binding (i.e. a small association equilibrium constant) or a solvent that decreases the extent of this binding by using a competing agent that displaces solute from the column. These two approaches are known as *nonspecific elution* and *biospecific elution*, respectively.[5]

Biospecific elution is the gentler of these two elution methods since it is carried out under essentially the same solvent conditions as are used for sample application. This makes this approach attractive for purification work, where a high recovery of active solute is desired. Biospecific elution may be performed either by adding an agent to the eluting solvent that competes with the ligand for solute (i.e. *normal role elution*) or by adding an agent that competes with solute for ligand-binding sites (i.e. *reversed-role elution*). In both cases, retained solutes are eventually eluted from the column by displacement and mass action. The main advantage of biospecific elution is its ability to gently remove analyte from the column. The main disadvantages are the slow elution times and broad solute peaks that are typically produced by this approach. Other limitations in purification work include the need to remove the competing agent from the final fraction of eluted solute and, in analytical applications, the need to use a competing agent that does not produce a large background signal under the conditions used for analyte detection.

In nonspecific elution, the solute is eluted by changing the column conditions in order to weaken the interactions between any retained compounds and the immobilized ligand. This can be done by changing the pH, ionic strength, or polarity of the mobile phase. The addition of denaturing or chaotropic agents to the running buffer can also be used. This results in an alteration in the structure of the solute or ligand, leading to a lower association equilibrium constant and lower solute retention. Nonspecific elution tends to be much faster than biospecific elution in removing analytes from affinity columns. This results in sharper peaks, which in turn produces lower limits of detection and shorter analysis times. For these reasons, nonspecific elution is commonly used in analytical applications of affinity chromatography. This elution method can also be used in purifying solutes, but there is a greater risk of solute denaturation with this approach than there is with biospecific elution. Furthermore, care must be taken in nonspecific elution to avoid using conditions that are too harsh for the

column. If this is not considered, it may result in long column regeneration times or an irreversible loss of ligand activity.

4 TYPES OF AFFINITY CHROMATOGRAPHY

4.1 Bioaffinity Chromatography

Bioaffinity chromatography, also known as *biospecific adsorption*, is a name that is used to describe affinity methods that use a biological molecule as the affinity ligand. This was the first type of affinity chromatography that was developed and continues to represent the most diverse category of this technique. As stated earlier, the first use of bioaffinity chromatography was in 1910 when Starkenstein purified α-amylase using insoluble starch (a substrate for this enzyme) as both the support and affinity ligand.[12] Further examples of ligands that are used in bioaffinity chromatography are provided in Table 1.

The use of bioaffinity chromatography for enzyme purification, as originally reported by Starkenstein, continues to be an important and common use of this method. The importance of this field is illustrated by a review which cites several hundred applications in this area alone.[34] An example of an enzyme separation performed by bioaffinity chromatography is shown in Figure 2. Ligands that can be used for this purpose include enzyme inhibitors, coenzymes, substrates, and cofactors. For instance, nucleotide mono-, di-, and triphosphates can be used for the purification of various kinase enzymes, NAD has been used to isolate dehydrogenases, pyridoxal phosphate has been employed for the purification of tyrosine and aspartate aminotransferases, and RNA or DNA has been used for the collection of polymerases and nucleases.[4] Antibodies that specifically bind to certain enzymes are also commonly used in such purification schemes.[34,35]

Lectins represent a class of general ligands that are common in bioaffinity chromatography. The term *lectin affinity chromatography* is sometimes used to refer to the resulting separation method. Lectins are non-immune system proteins, generally of plant origin, that have the ability to recognize and bind certain types of carbohydrate residues.[36] Some of the most common lectins that are used in affinity chromatography include concanavalin A (which binds to α-D-mannose and α-D-glucose residues), wheat germ agglutinin (which binds to D-*N*-acetylglucosamines) and jacalin (which binds to α-D-galactose residues). Other lectins that are also used are listed in Table 4. These ligands are valuable in the separation of many carbohydrate-containing compounds, such as polysaccharides, glycoproteins (e.g. immunoglobulins or cell membrane proteins), glycolipids and other types of glycoconjugates.[1–4,8,10]

Table 4 Examples of carbohydrate residues that bind to lectins used in affinity chromatography[8]

Carbohydrate residue	Types of interacting lectins
D-*N*-Acetyl galactosamine	Lectin 60 from *Ricinus communis*, soybean lectin (Note: phytohemagglutinins E_4 and L_4 also bind to *N*-acetyl galactosamines)
α-D-Galactose	Jacalin (Note: peanut and soybean lectin also bind to D-galatose)
β-D-Galactose	*Ricinus communis* lectins 60 and 120
α-D-Glucose	Concanavalin A, lentil lectin, pea lectin
α-D-Mannose	Concanavalin A, lentil lectin, pea lectin

Another useful class of bioaffinity ligands are antibody-binding proteins. Two common examples of these are protein A, a bacterial cell wall protein from *Staphylococcus aureus*, and protein G, which is a cell wall protein from group G streptococci bacteria.[27,37,38] These ligands have the ability to bind to the constant region of many types of immunoglobulins, or antibodies. This makes them useful in antibody purification, particularly when dealing with IgG-class antibodies.[8] Protein A and protein G have their strongest binding to immunoglobulins at or near neutral pH but readily dissociate from these solutes when placed into a lower pH buffer. These two ligands differ in their ability to bind to antibodies from different species and classes.[27,39] Recombinant versions of both protein A and protein G are available commercially[8] along with a recombinant fusion protein, known as protein A/G, that blends the binding activities of these two ligands.[40]

Other examples of antibody-binding proteins include protein B and mannan-binding protein. Protein B is a b antigen protein found on the surface of group A streptococci bacteria. It is useful in binding to several types of IgA-class antibodies, such as human IgA_1, IgA_2 and secretory IgA. However, it does not bind to mouse IgA or IgAs from many other types of mammals.[8,41] Mannan-binding protein is found in the serum of various mammals[42] and has been found to bind IgM-class antibodies in mouse ascites fluid, making it of potential use for the isolation of IgM monoclonal antibodies.[43] The binding of antibodies to mannan-binding protein is calcium-dependent and the elution of bound IgM can be obtained by incorporating ethylene diamine tetraacetic acid (EDTA) as a calcium-complexing agent into the mobile phase.[8]

Nucleic acids and polynucleotides can act as either general or specific ligands in bioaffinity chromatography.[44,45]

For instance, as high-specificity ligands they can be used to purify DNA/RNA-binding enzymes and proteins or to isolate nucleic acids that contain a sequence which is complementary to the ligand.[46] As a group-specific ligand, an immobilized nucleic acid can be used to purify solutes that share a common nucleotide sequence. An example is the use of immobilized oligo(dT) for the isolation of nucleic acids containing poly(A) sequences. As mentioned earlier, another example of a general nucleic acid ligand is the use of nucleotide mono-, di-, and triphosphates for the purification of kinase enzymes.[4]

4.2 Immunoaffinity Chromatography

Immunoaffinity chromatography (IAC) is a special category of bioaffinity chromatography in which the ligand that is used is an antibody or antibody-related reagent.[10,11,35] Other methods that are sometimes included under the heading of IAC are those that use immobilized antigens or antibody-binding proteins for antibody purification.[8,11,47] The first known use of IAC was the work described earlier by Campbell et al. in 1951.[14] Many current applications are still based on the use of low-performance supports, particularly agarose.[35] However, recent work has also been performed using derivatized silica, glass, and perfusion media. The use of these supports along with an antibody or antigen ligand is referred to as *high-performance immunoaffinity chromatography* (HPIAC).[11,47,48]

The high selectivity of antibody–antigen interactions and the ability to produce antibodies against a wide range of solutes has made IAC a popular tool for biological purification and analysis. There are now literally hundreds to thousands of examples of immunoaffinity methods that have been developed for the isolation of antibodies, hormones, peptides, enzymes, recombinant proteins, receptors, viruses, and subcellular components.[34,35,47,49–53] The high association equilibrium constants of many antibodies requires that nonspecific elution be used with most immunoaffinity columns.[47] However, isocratic elution methods can also be used with low-affinity antibody systems.[24]

The use of IAC in analytical methods has also received a great deal of attention in recent years.[11,48] Table 5 summarizes some strategies that can be employed in these analyses and gives some examples of representative applications. The simplest format for using IAC in compound analysis involves the traditional ''on/off'' or step-gradient mode of affinity chromatography (see Figure 1). The advantages of using this approach in analytical applications, particularly when performed by HPIAC, include its relative simplicity, good precision, and potential for fast sample analysis. One requirement of the direct detection mode is that there must be some way to monitor the analyte as it leaves the column. For protein solutes, UV absorbance measurements at 210–215 nm or 280 nm are most commonly used. In some cases minor sample components have been monitored by using precolumn derivatization to place easy-to-detect labels (such as fluorescent tags) on sample solutes prior to injection. Another alternative is to collect fractions of the column eluent and later analyze these by a separate technique, such as an immunoassay or assay for biological activity.[11]

Immunoextraction is another approach by which immunoaffinity columns can be used for solute detection. In this method, an immunoaffinity column is used for the removal of a specific solute or group of solutes from a sample prior to determination by a second analytical method. This employs the same general operating scheme as shown in Figure 1, but now involves combining the immunoaffinity column either off-line or on-line with some other method for the actual quantitation of analytes. Off-line methods are generally the easiest and most common way for combining immunoaffinity columns with other analytical techniques.[11,54–57] This approach typically involves the use of antibodies that are immobilized onto a low-performance support that is packed into a small disposable syringe or solid-phase extraction cartridge. After conditioning the immunoaffinity column with the necessary application buffer or conditioning solvents, the sample is applied and undesired sample components are washed away. An elution buffer is then applied and the test analyte is collected as it is removed from the column. If desired, the collected fraction can be analyzed directly or first dried down and reconstituted in a solvent that is compatible with the analytical method to be used for quantitation (e.g. HPLC or gas chromatography (GC)).

The fact that off-line immunoextraction generally involves manual sample preparation does limit the potential speed and precision of this approach. To overcome this problem, a large number of recent reports have examined the use of on-line immunoextraction methods. The direct coupling of immunoextraction with HPLC has been of particular interest,[11,28] but some methods have also been reported that involve the coupling of on-line immunoextraction with GC[58] or capillary electrophoresis.[59–61] A typical column switching scheme for performing on-line immunoextraction with reversed-phase liquid chromatography (RPLC) is shown in Figure 6.[48] The general format for this type of system involves injecting the sample onto the immunoaffinity extraction column, with the nonretained components being allowed to go to a waste container. The immunoaffinity column is then switched on-line with a RPLC column and an elution buffer is applied to the immunoaffinity support to dissociate any retained analyte. As these analytes elute, they are captured and

Table 5 Examples of analytical applications of IAC[11]

Direct analyte detection
Anti-idiotypic antibodies, antithrombin III, bovine growth hormone, fibrinogen, fungal carbohydrate antigens, glucose tetrasaccharide, glutamine synthetase, granulocyte colony stimulating factor, group A-active oligosaccharides, human serum albumin, immunoglobulin G, immunoglobulin E, interferon, interleukin-2, lymphocyte receptors, β_2-microglobulin, tissue-type plasminogen activator, transferrin
Off-line immunoextraction
HPLC Methods: Aflatoxin, albuterol, benzodiazepines, cytokinins, fumonisin, human chorionic gonadotropin, ivermectin and avermectin, nortestosterone, ochratoxin A, oxytocin, phenylurea herbicides, sendai virus protein, trenbolone, triazine herbicides
GC Methods: Alkylated DNA adducts, chloramphenicol, dexamethasone, estrogens, flumethasone, nortestosterone, prostaglandins and thromboxanes
On-line immunoextraction
HPLC Methods: Aflatoxin M1, β-agonists, α1-antitrypsin, atrazine, atrazine metabolites, benzylpenicilloyl-peptides, bovine serum albumin, carbendazim, carbofuran, chloramphenicol, clenbuterol, cortisol, dexamethasone, diethylstilbestrol, digoxin, estrogens, hemoglobin, human epidermal growth factor, human growth hormone variants, interferon α-2, LSD and metabolites, lysozyme variants, 17β- and 17α-19-nortestosterone, phenytoin, propranolol, Δ9-tetrahydrocannabinol, tolubuterol, transferrin, 17β- and 17α-trenbolone
GC Methods: β-19-Nortestosterone and related steroids
Capillary Electrophoresis Methods: Cyclosporin, immunoglobulin E, insulin
Chromatographic immunoassays
Competitive binding immunoassays: Adrenocorticotropic hormone, α-amylase, atrazine/triazines, 2,4-dinitrophenyl lysine, human chorionic gonadotropin, human serum albumin, immunoglobulin G, isoproturon, testosterone, theophylline, thyroid stimulating hormone, thyroxine, transferrin, trinitrotoluene
Sandwich immunoassays: Anti-bovine IgG antibodies, human serum albumin, immunoglobulin G, interferon, parathyroid hormone
One-site immunometric assays: α-(Difluoromethyl)ornithine, 17-β-estradiol
Post-column immunodetection
Acetylcholinesterase, bovine growth hormone releasing factor, digoxin, digoxigenin, digoxin metabolites, human methionyl granulocyte colony stimulating factor

reconcentrated at the head of the RPLC column. After all solutes have left the immunoaffinity column, this column is then switched back off-line and regenerated by passing through the initial application buffer. Meanwhile, the RPLC column is developed with either an isocratic or gradient elution scheme involving the application of a mobile phase with an increased organic modifier content. As the solutes elute through the RPLC columns, they are then monitored and quantitated at an on-line detector.[11]

A third area that has received increasing attention in recent years has been the use of IAC to perform various types of *chromatographic* (or *flow-injection*) *immunoassays*.[11,48] This is particularly valuable in determining trace analytes that, by themselves, may not produce a readily detectable signal. The competitive binding assay is the most common format that has been used in performing these immunoassays. One way this can be done is to mix the sample with a labeled analyte analog (i.e. the ''label'') and simultaneously inject these onto an immunoaffinity column that contains a relatively small amount of antibody; this method is known as a *simultaneous injection competitive binding immunoassay*. An alternative format, known as a *sequential injection competitive binding immunoassay*, involves the application of only sample to the immunoaffinity column, followed later by a separate injection of the label (see Figure 7a). In either case, the analyte competes with the injected label for the antibody binding sites in the column, thus making the amount of label that elutes in either the retained or nonretained fractions an indirect measure of how much analyte was present in the original sample.

Sandwich immunoassays and one-site immunometric assays have also been performed as part of immunoaffinity systems.[11,48] In the sandwich immunoassay (Figure 7b), two different types of antibodies are used that each bind to the analyte of interest. The first of these two antibodies is attached to a solid-phase support and is

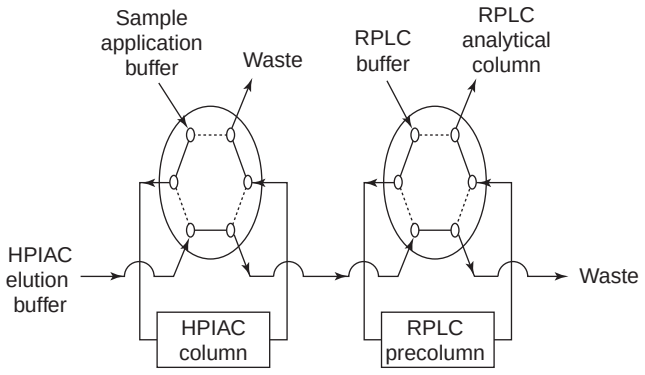

Figure 6 Column switching scheme for on-line immunoextraction in RPLC. (Reprinted with permission from D.S. Hage, J.G. Rollag, D.H. Thomas, 'Analysis of Atrazine and its Degradation Products in Water by Tandem High-performance Immunoaffinity Chromatography and Reversed-phase Liquid Chromatography', in *Immunochemical Technology for Environmental Applications*, eds. D.S. Aga, E.M. Thurman, ACS Press, Washington, DC, Chapter 10, Copyright (1997), American Chemical Society.)

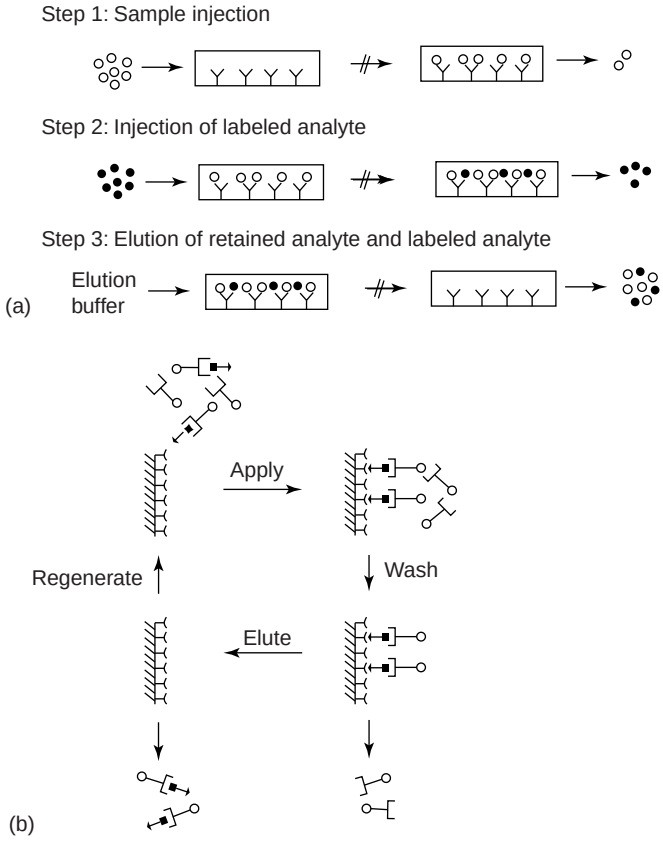

Figure 7 Schemes for chromatographic immunoassays based on (a) a sequential injection competitive binding format and (b) a sandwich immunoassay format. (Reprinted with permission from D.S. Hage, P.C. Kao, 'Theory of a Sequential Addition Competitive Binding Immunoassay Based on High-performance Immunoaffinity Chromatography', *Anal. Chem.*, **63**, 586–595, Copyright (1991), American Chemical Society, and D.S. Hage, D.H. Thomas, M.S. Beck, *Anal. Chem.*, **65**, 1622–1630, Copyright (1993), American Chemical Society.[32])

used to extract the analyte from samples. The second antibody contains an easily measured tag and is added in solution to the analyte either before or after this extraction; this second antibody serves to place a label onto the analyte, thus allowing the amount of analyte on the immunoaffinity support to be quantitated. In the one-site immunometric assay, the sample is first incubated with a known excess of labeled antibodies or Fab fragments that are specific for the analyte of interest. After binding between the sample analyte and antibodies has occurred, this mixture is then applied to a column that contains an immobilized analog of the analyte. This column serves to extract any antibodies or Fab fragments that are not bound to sample analyte. Meanwhile, those antibodies or Fab fragments that are bound to sample analyte will pass through the column in the nonretained peak. Detection is then performed by looking at the size of the nonretained peak for the labeled antibodies/Fab fragments.

Yet another way in which immunoaffinity columns can be used is for monitoring the elution of specific solutes from other chromatographic columns, a technique known as *post-column immunodetection*. This involves the use of a post-column reactor and an immobilized antibody or antigen column attached to the exit of an analytical HPLC column.[11,62] This can be performed by either using the on/off direct detection mode of IAC or chromatographic immunoassay formats based on one-site immunometric assays, sandwich immunoassays or competitive-binding immunoassays.[11,48] The one-site immunometric assay is the most common format for immunodetection.[11] The basic operation involves taking the analytical HPLC column eluent and combining this with a solution of labeled antibodies or Fab fragments that will bind the analyte of interest. The column eluent and antibody or Fab mixture is then allowed to react in a mixing coil and passed through an immunodetection column that contains an immobilized analog of the analyte. The antibodies or Fab fragments that are bound to the analyte will pass through this column and onto the detector, where they will provide a signal that is proportional to the amount of bound analyte. If desired, the immunodetection column can later be washed with an eluting solvent to dissociate the retained antibodies or Fab fragments; but a sufficiently high binding capacity is generally used so that a reasonably large amount of analytical column eluent can be analyzed before the immunodetection column must be regenerated.

4.3 Dye-ligand Affinity Chromatography

Another category of affinity chromatography is the technique known as *dye-ligand affinity chromatography*. The first use of dye-ligand affinity chromatography was described in 1974.[63] Since that time it has become an extremely popular tool for enzyme and protein purification,[1–4,8,10,63–65] with over 500 such compounds having been isolated by this technique.[5] In this method, a synthetic dye (like a triazine or triphenylmethane compound) is used as the immobilized ligand (see Figure 8). These dyes bind to the active site of a protein or enzyme by mimicking the structure of a substrate or cofactor for that biomolecule. The most common dye used for this purpose is Cibacron Blue F3G-A. Some other examples of dyes that are used as ligands in affinity chromatography include Procion Blue MX-3G or MX-R, Procion Red HE-3B, thymol blue and phenol red.[8]

The cibacron and procion dyes all have a chlorotriazine ring but have different side groups attached to this ring. For Cibacron Blue F3G-A, one of these side groups is an anthraquinone that reacts with enzymes that have a binding site for NAD, NADP or ATP. This property

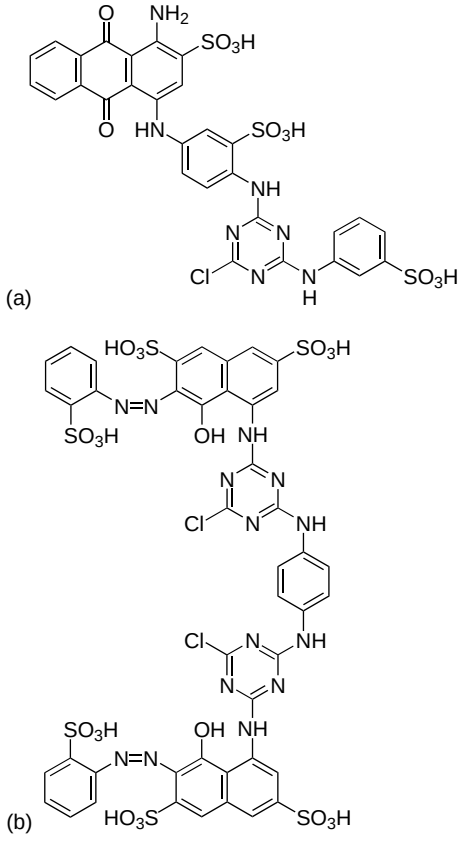

Figure 8 Structures of some common ligands used in dye-ligand affinity chromatography. (a) Cibacron Blue F3GA. (b) Procion Red HE3B.

makes this dye useful as a ligand for the purification of dehydrogenases, kinases, and other enzymes that require such cofactors. However, this same dye can also be used for a wide range of other enzymes and proteins. Examples include albumin, α-fetoprotein, CoA-dependent enzymes, hydrolases, immunoglobulin G, lipoproteins, nucleases, polymerases, synthetases and transferases.[64,65] The other dyes used in dye-ligand affinity chromatography have slightly different specificities, but can similarly be used for the purification of variety of enzymatic and nonenzymatic proteins.[8]

4.4 Immobilized Metal Ion Affinity Chromatography

In *immobilized metal ion affinity chromatography* (*IMAC*), or *metal chelate affinity chromatography*, the affinity ligand is a metal ion which is complexed with an immobilized chelating agent. This type of affinity chromatography was first described by J. Porath and co-workers in 1975.[66] It was initially used to separate proteins and peptides that contained amino acids with electron donor groups (such as histidine, tryptophan, or cysteine) with the proteins and peptides being retained owing to interactions between their amino acids and the immobilized metal chelate.[66,67] Since its discovery, a number of peptides, proteins, and amino acids have been purified by this method. An example of one such separation is shown in Figure 9.

Selectivity in IMAC is determined by the chelating group and the type of metal ion that is used to make up the affinity ligand. Iminodiacetic acid is the most common chelating agent employed. Other ligands that are used in IMAC include carboxymethylaspartic acid, tris-carboxymethylethylenediamine, tris(2-aminoethyl)amine and dipicolylamine.[66–68] Metal ions used with these chelating groups include Cu^{2+}, Zn^{2+}, Ni^{2+}, Co^{2+}, and Fe^{3+}.[68] In general, the stronger that the interaction is between the metal ion and the chelating group, the weaker the observed retention will be between the same immobilized metal and an injected protein or peptide.

4.5 Analytical Affinity Chromatography

Besides its use in separating and quantitating sample components, affinity chromatography can also be employed as a tool for studying solute–ligand interactions. This application of affinity chromatography is called *analytical*, or *quantitative, affinity chromatography*. Using this technique, information can be acquired regarding the equilibrium and rate constants for biological interactions, as well as the number and types of binding sites that are involved in these interactions.[23,69]

Information on the equilibrium constants for a solute–ligand system can be obtained by using either the method

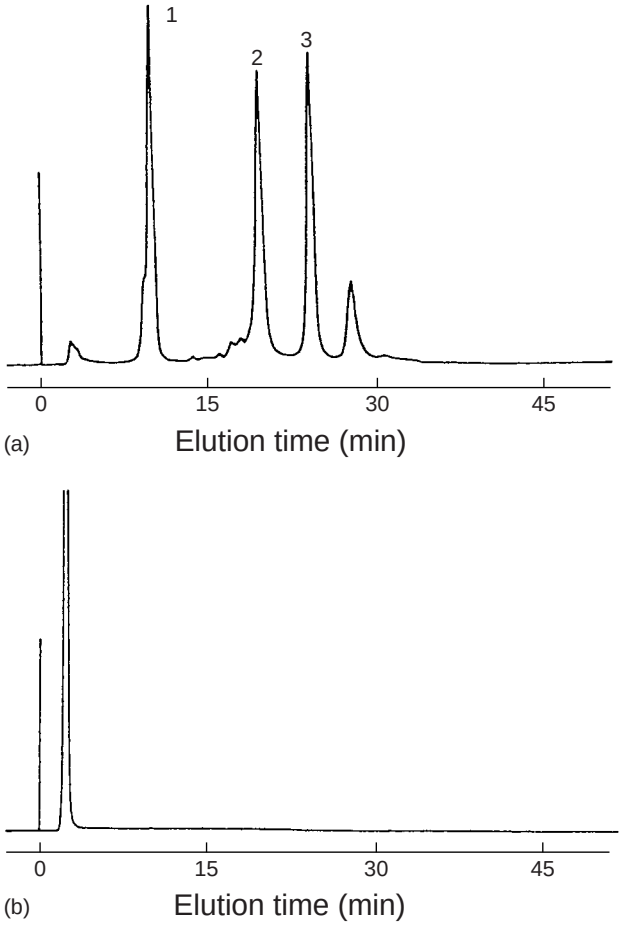

Figure 9 Injection of a mixture of ribonuclease A (1), transferrin (2), and carbonic anhydrase (3) onto an immobilized iminodiacetic acid column in (a) the presence of Zn^{2+} complexed to the column and (b) in the absence of Zn^{2+}. (Reprinted from Y. Kato, K. Nakamura, T. Hashimoto, 'High-performance Metal Chelate Affinity Chromatography of Proteins', *J. Chromatogr.*, **354**, 511–517, Copyright (1986), with permission from Elsevier Science.)

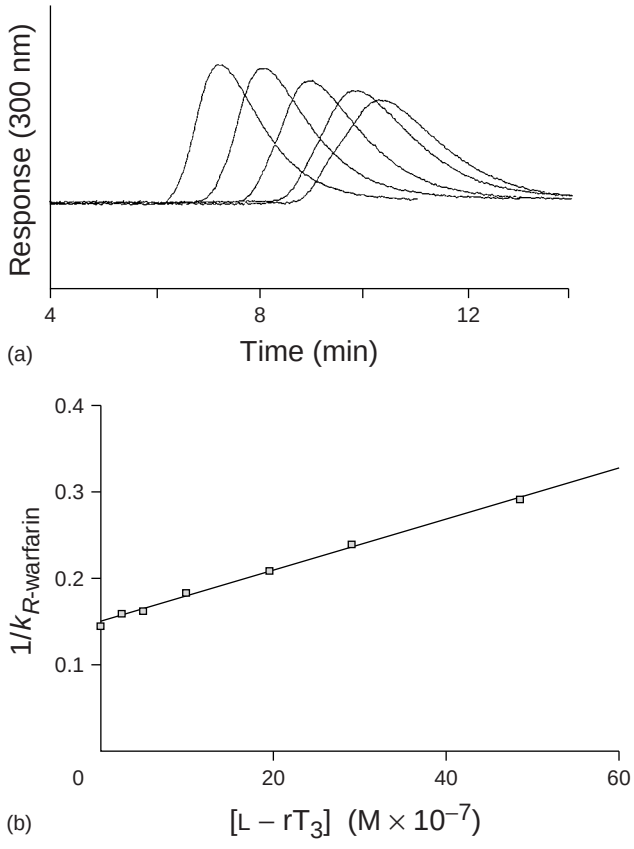

Figure 10 (a) Chromatograms and (b) relationship between the retention factor (k) and mobile phase additive concentration for zonal elution competitive binding experiments performed on an immobilized human serum albumin column using *R*-warfarin as the injected analyte and L-reverse triiodothyronine (L-rT_3) as a mobile phase additive. The mobile phase concentrations of L-rT_3 in (a) were (left-to-right) 1.90, 0.97, 0.49, 0.24 and 0 μM. (Reprinted from B. Loun, D.S. Hage, 'Characterization of Thyroxine–Albumin Binding Using High-performance Affinity Chromatography. II. Comparison of the Binding of Thyroxine, Triiodothyronine and Related Compounds at the Warfarin and Indole Sites of Human Serum Albumin', *J. Chromatogr. B*, **665**, 303–314, Copyright (1995), with permission from Elsevier Science.)

of *zonal elution* or *frontal analysis*. Zonal elution involves the injection of a small amount of solute onto an affinity column in the presence of a mobile phase that contains a known concentration of a competing agent. The equilibrium constants for binding of the ligand with the solute (and competing agent) can then be obtained by examining how the solute's retention changes with competing-agent concentration. A typical zonal elution experiment is shown in Figure 10. This technique was first used to study biological interactions by Andrews et al. in 1974[70] and by Dunn and Chaiken in 1975.[71] Since that time it has been used to examine a number of systems, such as enzyme–inhibitor binding, protein–protein interactions, and drug–protein binding.[5,10,23,69]

Frontal analysis is performed by continuously applying a known concentration of solute to an affinity column at a fixed flow-rate. The moles of analyte required to reach the mean point of the resulting breakthrough curve is then measured and used to determine the equilibrium constant for solute–ligand binding. This method was used first by Kasai and Ishii in 1975 to examine the interactions of trypsin with various peptide ligands.[72] It has since been used in many other applications.[23,69] One advantage of this approach over zonal elution is that it simultaneously provides information on both the equilibrium constants and the number of active sites involved in analyte–ligand binding. The main disadvantage of this method is the need for a larger quantity of solute then is required by zonal elution.

Information on the kinetics of solute–ligand interactions can also be obtained using affinity chromatography. A number of methods have been developed for this, including techniques based on band-broadening measurements,[73–77] the split-peak effect,[29,32,33] and peak decay analysis.[23] These methods are generally more difficult to perform than equilibrium constant measurements but represent a powerful means of examining the rates of biological interactions. Systems that have been studied by these techniques include the binding of lectins with sugars,[23,74,75] protein A or protein G with immunoglobulins,[29,33] antibodies with antigens,[23,32] and drugs or amino acids with human serum albumin.[76,77]

4.6 Miscellaneous Ligands and Methods

Besides the specific methods already discussed, a number of other ligands and techniques are often placed under the category of affinity chromatography. For example, boronic acid and its derivatives represent a class of general ligands that are usually considered to be affinity ligands. At a pH of 8 or higher, these ligands have the ability to form covalent bonds with compounds that contain cis-diol groups in their structure. This makes these ligands useful for the purification and analysis of many compounds which contain sugar residues, such as polysaccharides, glycoproteins, ribonucleic acids, and catechol-related compounds.[78–80]

Other methods that are related to affinity chromatography include *hydrophobic interaction chromatography* (*HIC*) and *thiophilic adsorption*. HIC is based on the interactions of proteins, peptides, and nucleic acids with short nonpolar chains on a support. This was first described in 1972[81,82] following work that examined the role of spacer arms on the nonspecific adsorption of affinity columns.[82] Thiophilic adsorption, also known as *covalent* or *chemisorption chromatography*, makes use of immobilized thiol groups for solute retention.[83] Applications of this method include the analysis of sulfhydryl-containing peptides or proteins and mercurated polynucleotides.[6]

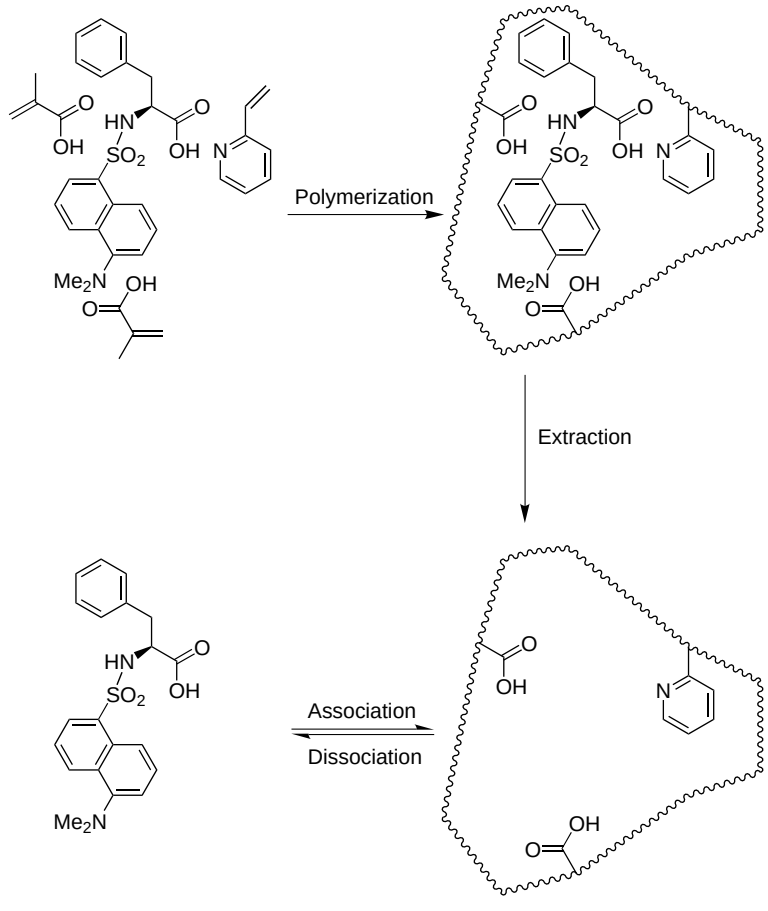

Figure 11 Preparation of a molecular imprint for dansyl-L-phenylalanine. (Reprinted with permission from D. Kriz, O. Ramstrom, K. Mosbach, *Anal. Chem.*, **69**, 345A–349A, Copyright (1997), American Chemical Society.[87])

5 FUTURE DEVELOPMENTS

A clear trend that has been present since the beginning of affinity chromatography has been the search for new and improved ligands to use in this technique. It has already been shown how a variety of biological and nonbiological compounds have been employed in affinity separations. In the future it is anticipated that an even greater variety of affinity stationary phases will be developed. One group of promising candidates are the aptamers, which are polymers of 8–120 nucleotides that have well-defined sequences and three-dimensional structures. It has been shown that a large number of aptamers can be randomly generated in an oligonucleotide library and then those ligands which bind to a given target solute can be selectively enriched for use in chemical sensors or other applications. The result is a ligand that is stable, easy to prepare and that has a relatively high affinity for the solute of interest.[84–86]

Another active area of new stationary phase development is in the field of molecular imprinting.[87,88] In this case, the affinity ligand is actually part of the surface or internal structure of the support used in the affinity column. Molecular imprints are usually made by combining the analyte of interest with a series of monomers that contain side chains capable of forming various interactions with the analyte. As these interactions take place, the monomers are fixed in position about the analyte by initiating their polymerization (see Figure 11). After the polymerization has occurred, the support is ground into a powder, the retained analyte is released by applying an appropriate solvent, and the imprinted support is placed into a column for use. In this way, an affinity support is created that has a known specificity and binding/elution properties.[87]

A second trend in affinity chromatography has been the increasing use of this method in tandem analytical techniques. For instance, Table 5 shows how IAC has already been combined, both on-line and off-line, with HPLC, GC and capillary electrophoresis.[11] Similar interfaces have been developed for other ligands (e.g. protein A, protein G, and boronic acid), but these have not seen nearly as much use as immunoaffinity columns in tandem techniques. One role that an affinity column can play in such a tandem system is to selectively extract an analyte or group of analytes from a sample prior to study by a second analytical method. The affinity column can also be used as a type of sensor to selectively monitor a given agent as it appears in the other analytical technique. The enhanced specificity that results in the overall method is a key reason why tandem affinity systems are receiving a good deal of attention in areas such as pharmaceutical analysis, food testing, and environmental monitoring.[11,28]

ACKNOWLEDGMENTS

This work was support by the National Institutes of Health under grant R01 GM44931.

ABBREVIATIONS AND ACRONYMS

CDI	*N*,*N*′-Carbonyl diimidazole
CNBr	Cyanogen Bromide
EDC	Ethyldimethylaminopropyl Carbodiimide
EDTA	Ethylene Diamine Tetraacetic Acid
FMP	Fluoro methylpyridinium toluenesulfonate
GC	Gas Chromatography
HIC	Hydrophobic Interaction Chromatography
HPAC	High-performance Affinity Chromatography
HPIAC	High-performance Immunoaffinity Chromatography
HPLC	High-performance Liquid Chromatography
IAC	Immunoaffinity Chromatography
IMAC	Immobilized Metal Ion Affinity Chromatography
LDH	Lactate Dehydrogenase
NHS	*N*-hydroxysuccinimide ester
RPLC	Reversed-phase Liquid Chromatography
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Clinical Chemistry **(Volume 2)**
Immunochemistry

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Immunoassay Techniques in Environmental Analyses

Nucleic Acids Structure and Mapping **(Volume 6)**
Aptamers

Peptides and Proteins **(Volume 7)**
Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Protein–Drug Interactions

Pesticides **(Volume 7)**
Immunochemical Assays in Pesticide Analysis

Liquid Chromatography **(Volume 13)**
Biopolymer Chromatography • Chiral Separations by High-performance Liquid Chromatography

REFERENCES

1. J. Turkova, *Affinity Chromatography*, Elsevier, Amsterdam, 1978.
2. W.H. Scouten, *Affinity Chromatography: Bioselective Adsorption on Inert Matrices*, John Wiley & Sons, New York, 1981.
3. H. Schott, *Affinity Chromatography: Template Chromatography of Nucleic Acids and Proteins*, Dekker, New York, 1985.
4. I. Parikh, P. Cuatrecasas, 'Affinity Chromatography', *Chem. Eng. News*, **63**, 17–29 (1985).
5. R.R. Walters, 'Affinity Chromatography', *Anal. Chem.*, **57**, 1099A–1114A (1985).
6. P. Mohr, K. Pommerening, *Affinity Chromatography: Practical and Theoretical Aspects*, Dekker, New York, 1985.
7. K. Jones, 'Affinity Chromatography – An Overview', *Anal. Proceed.*, **28**, 140–141 (1991).
8. G.T. Hermanson, A.K. Mallia, P.K. Smith, *Immobilized Affinity Ligand Techniques*, Academic Press, San Diego, 1992.
9. T.T. Ngo (ed.), *Molecular Interactions in Bioseparations*, Plenum Press, New York, 1993.
10. D.S. Hage, 'Affinity Chromatography', *Handbook of HPLC*, eds. E. Katz, R. Eksteen, P. Schoenmakers, N. Miller, Marcel Dekker, New York, Chap. 13, 1998.
11. D.S. Hage, 'Recent Advances in Analytical Applications of Immunoaffinity Chromatography', *J. Chromatogr. B*, **715**, 3–28 (1998).
12. E. Starkenstein, 'Ferment Action and the Influence Upon It of Neutral Salts', *Biochem. Z.*, **24**, 210–218 (1910).
13. V.A. Engel'gardt, L.L. Kiselev, R.S. Nezlin, 'Principle of a ''Fixed Partner'' as an Experimental Tool in Molecular Biology Research', *Monatsch. Chem.*, **101**, 1510–1570 (1970).
14. D.H. Campbell, E. Luescher, L.S. Lerman, 'Immunologic Adsorbents: I. Isolation of Antibody by Means of a Cellulose-Protein Antigen', *Proc. Natl. Acad. Sci. USA*, **37**, 575–582 (1951).
15. L.S. Lerman, 'Biochemically Specific Method For Enzyme Isolation', *Proc. Natl. Acad. Sci. USA*, **39**, 232–260 (1953).
16. C. Arsenis, D.B. McCormick, 'Purification of Liver Flavokinase by Column Chromatography on Flavine-Cellulose Compounds', *J. Biol. Chem.*, **239**, 3093–3097 (1964).
17. D.B. McCormick, 'Specific Purification of Avidin by Column Chromatography on Biotin-Cellulose', *Anal. Biochem.*, **13**, 194–198 (1965).
18. S. Hjerten, 'Chromatographic Separation According to Size of Macromolecules and Cell Particles on Columns of Agarose Suspension', *Arch. Biochem. Biophys.*, **99**, 466–475 (1962).
19. R. Axen, J. Porath, S. Ernback, 'Chemical Coupling of Peptides and Proteins to Polysaccharides by Means of Cyanogen Halides', *Nature (London)*, **214**, 1302–1304 (1967).
20. P. Cuatrecasas, M. Wilchek, C.B. Anfinsen, 'Selective Enzyme Purification by Affinity Chromatography', *Proc. Natl. Acad. Sci. USA*, **61**, 636–643 (1968).
21. F.E. Regnier, R. Noel, 'Glycerolpropylsilane Bonded Phases in the Steric Exclusion Chromatography of Biological Macromolecules', *J. Chromatogr. Sci.*, **14**, 316–320 (1976).
22. S. Ohlson, L. Hansson, P.-O. Larsson, K. Mosbach, 'High Performance Liquid Affinity Chromatography (HPLAC) and Its Application to the Separation of Enzymes and Antigens', *FEBS Lett.*, **93**, 5–9 (1978).
23. I.M. Chaiken (ed.), *Analytical Affinity Chromatography*, CRC Press, Boca Raton, FL, 1987.
24. D. Zopf, S. Ohlson, 'Weak-affinity Chromatography', *Nature*, **346**, 87–88 (1990).
25. M. Wikstroem, S. Ohlson, 'Computer Simulation of Weak Affinity Chromatography', *J. Chromatogr.*, **597**, 83–92 (1992).
26. B. Loun, D.S. Hage, 'Chiral Separation Mechanisms in Protein-based HPLC Columns. 1. Thermodynamic Studies of (R)- and (S)-Warfarin Binding to Immobilized Human Serum Albumin', *Anal. Chem.*, **66**, 3814–3822 (1994).
27. R. Lindmark, C. Biriell, J. Sjoequist, 'Quantitation of Specific IgG Antibodies in Rabbits by a Solid-phase Radioimmunoassay With ^{125}I-Protein A From *Staphylococcus aureus*', *Scand. J. Immunol.*, **14**, 409–420 (1981).
28. M. de Frutos, F.E. Regnier, 'Tandem Chromatographic Immunological Analyses', *Anal. Chem.*, **65**, 17A–25A (1993).
29. D.S. Hage, R.R. Walters, H.W. Hethcote, 'Split-peak Affinity Chromatographic Studies of the Immobilization-dependent Adsorption Kinetics of Protein A', *Anal. Chem.*, **58**, 274–279 (1986).
30. D.S. Hage, R.R. Walters, 'Dual Column Determination of Albumin and Immunoglobulin G in Serum by High-performance Affinity Chromatography', *J. Chromatogr.*, **386**, 37–49 (1987).
31. P.F. Ruhn, J.D. Taylor, D.S. Hage, 'Determination of Urinary Albumin Using High-performance Immunoaffinity Chromatography and Flow Injection Analysis', *Anal. Chem.*, **66**, 4265–4271 (1994).
32. D.S. Hage, D.H. Thomas, M.S. Beck, 'Theory of a Sequential Addition Competitive Binding Immunoassay

Based on High-performance Immunoaffinity Chromatography', *Anal. Chem.*, **65**, 1622–1630 (1993).
33. J.G. Rollag, D.S. Hage, 'Non-linear Elution Effects in Split-peak Chromatography. II. Role of Ligand Heterogeneity in Solute Binding to Columns with Adsorption-Limited Kinetics', *J. Chromatogr. A*, **795**, 185–198 (1998).
34. M. Wilchek, T. Miron, J. Kohn, 'Affinity Chromatography', *Methods Enzymol.*, **104**, 3–55 (1984).
35. G.J. Calton, 'Immunosorbent Separations', *Methods Enzymol.*, **104**, 381–387 (1984).
36. I.E. Liener, N. Sharon, I.J. Goldstein, *The Lectins: Properties, Functions and Applications in Biology and Medicine*, Academic Press, London, 1986.
37. P.L. Ey, S.J. Prowse, C.R. Jenkin, 'Isolation of Pure IgG, IgG_{2a} and IgG_{2b} Immunoglobulins From Mouse Serum Using Protein A-Sepharose', *Immunochem.*, **15**, 429–436 (1978).
38. L. Bjorck, G. Kronvall, 'Purification and Some Properties of Streptococcal Protein G, a Novel IgG-Binding Reagent', *J. Immunol.*, **133**, 969–974 (1984).
39. B. Aakerstrom, L. Bjoerck, 'A Physicochemical Study of Protein G, a Molecule with Unique Immunoglobulin G-Binding Properties', *J. Biol. Chem.*, **261**, 10240–10247 (1986).
40. M. Eliasson, A. Olsson, E. Palmcrantz, K. Wiberg, M. Inganaes, B. Guss, M. Lindberg, M. Uhlen, 'Chimeric IgG-Binding Receptors Engineered From Staphylococcal Protein A and Streptococcal Protein G', *J. Biol. Chem.*, **263**, 4323–4327 (1988).
41. E.L. Faulmann, J.L. Duval, M.D.P. Boyle, 'Protein B: A Versatile Bacterial Fc-Binding Protein Selective for Human IgA', *Bio/Techniques*, **10**, 748–750 (1991).
42. T. Kawaskai, N. Kawasaki, I. Yamashina, 'Mannose/N-Acetylglucosamine-binding Properties From Mammalian Sera', *Methods Enzymol.*, **179**, 310–321 (1989).
43. J.R. Nevens, A.K. Mallia, M.W. Wendt, P.K. Smith, 'Affinity Chromatographic Purification of Immunoglobulin M Antibodies Utilizing Immobilized Mannan-Binding Protein', *J. Chromatogr.*, **597**, 247–256 (1992).
44. H.W. Jarrett, 'Affinity Chromatography with Nucleic Acid Polymers', *J. Chromatogr.*, **618**, 315–339 (1993).
45. T.A. Goss, M. Bard, H.W. Jarrett, 'High-performance Affinity Chromatography of DNA', *J. Chromatogr.*, **508**, 279–287 (1990).
46. P.T. Gilham, 'The Synthesis of Polynucleotide–Celluloses and Their Use in the Fractionation of Polynucleotides', *J. Am. Chem. Soc.*, **86**, 4982–4985 (1990).
47. T.M. Phillips, 'High Performance Immunoaffinity Chromatography', *LC Mag.*, **3**, 962–972 (1985).
48. D.S. Hage, 'Chromatographic Approaches to Immunoassays', *J. Clin. Ligand Assay*, **20**, 293–301 (1997).
49. H. Ehle, A. Horn, 'Immunoaffinity Chromatography of Enzymes', *Bioseparation*, **1**, 97–110 (1990).
50. T.M. Phillips, 'Isolation and Recovery of Biologically Active Proteins by High-Performance Immunoaffinity Chromatography', *Recept. Biochem. Methodol.*, **14**, 129–154 (1989).
51. P. Bailon, S.K. Roy, 'Recovery of Recombinant Proteins by Immunoaffinity Chromatography', *ACS Symp. Ser.*, **427**, 150–167 (1990).
52. K.E. Howell, J. Gruenberg, A. Ito, G.E. Palade, 'Immunoisolation of Subcellular Components', *Prog. Clin. Biol. Res.*, **270**, 77–90 (1988).
53. M. Nakajima, I. Yamaguchi, 'Purification of Plant Hormones by Immunoaffinity Chromatography', *Kagaku to Seibutsu*, **29**, 270–275 (1991).
54. L.A. van Ginkel, 'Immunoaffinity Chromatography, its Applicability and Limitations in Multi-residue Analysis of Anabolizing and Doping Agents', *J. Chromatogr.*, **564**, 363–384 (1991).
55. N. Haagsma, C. van de Water, 'Immunochemical Methods in the Analysis of Veterinary Drug Residues', in *Analysis of Antibiotic Drug Residues in Food Products of Animal Origin*, ed. V.K. Agarwal, Plenum, New York, 81–97, 1992.
56. S.E. Katz, M. Siewierski, 'Drug Residue Analysis using Immunoaffinity Chromatography', *J. Chromatogr.*, **624**, 403–409 (1992).
57. S.E. Katz, M.S. Brady, 'High-performance Immunoaffinity Chromatography for Drug Residue Analysis', *J. Assoc. Off. Anal. Chem.*, **73**, 557–560 (1990).
58. A. Farjam, J.J. Vreuls, W.J.G.M. Cuppen, U.A.Th. Brinkman, G.J. de Jong, 'Direct Introduction of Large-volume Urine Samples into an On-line Immunoaffinity Sample Pretreatment–Capillary Gas Chromatography System', *Anal. Chem.*, **63**, 2481–2487 (1991).
59. T.M. Phillips, J.J. Chmielinska, 'Immunoaffinity Capillary Electrophoretic Analysis of Cyclosporin in Tears', *Biomed. Chromatogr.*, **8**, 242–246 (1994).
60. N.A. Guzman, 'Biomedical Applications of On-line Preconcentration–Capillary Electrophoresis using an Analyte Concentrator: Investigations of Design Options', *J. Liq. Chromatogr.*, **18**, 3751–3768 (1995).
61. L.J. Cole, R.T. Kennedy, 'Selective Preconcentration for Capillary Zone Electrophoresis using Protein G Immunoaffinity Capillary Chromatography', *Electrophoresis*, **16**, 549–556 (1995).
62. H. Irth, A.J. Oosterkamp, U.R. Tjaden, J. van der Greef, 'Strategies for On-line Coupling of Immunoassays to High-performance Liquid Chromatography', *Trends Anal. Chem.*, **14**, 355–361 (1995).
63. R.L. Easterday, J.M. Easterday, *Immobilized Biochemicals and Affinity Chromatography*, ed. R.B. Dunlap, Plenum, New York, 123, 1974.
64. Y.D. Clonis, A. Atkinson, C.J. Bruton, C.R. Lowe, *Reactive Dyes in Protein and Enzyme Technology*, Stockton, New York, 1987.
65. C.R. Lowe, J.C. Pearson, 'Affinity Chromatography on Immobilized Dyes', *Methods Enzymol.*, **104**, 97–113 (1984).

66. J. Porath, J. Carlsson, I. Olsson, G. Balfrage, 'Metal Chelate Affinity Chromatography: A New Approach to Protein Fractionation', *Nature*, **258**, 598–599 (1975).
67. M. Belew, J. Porath, 'Immobilized Metal Ion Affinity Chromatography: Effect of Solute Structure, Ligand Density and Salt Concentration on the Retention of Peptides', *J. Chromatogr.*, **516**, 333–354 (1990).
68. J. Porath, 'IMAC-Immobilized Metal Ion Affinity Based Chromatography', *Trends Anal. Chem.*, **7**, 254–259 (1988).
69. D.S. Hage, S.A. Tweed, 'Recent Advances in Chromatographic and Electrophoretic Methods for the Study of Drug-Protein Interactions', *J. Chromatogr. B*, **699**, 499–525 (1997).
70. P. Andrews, B.J. Kitchen, D. Winzor, 'Affinity Chromatography for the Quantitative Study of Acceptor–Ligand Interactions: Lactose Synthetic System', *Biochem. J.*, **135**, 897–900 (1973).
71. B.M. Dunn, I.M. Chaiken, 'Quantitative Affinity Chromatography: Determination of Binding Constants by Elution with Competitive Inhibitors', *Proc. Natl. Acad. Sci. USA*, **71**, 2382–2385 (1974).
72. K.-I. Kasai, S.-I. Ishii, 'Quantitative Analysis of Affinity Chromatography of Trypsin. New Technique for Investigation of Protein–Ligand Interaction', *J. Biochem.*, **77**, 261–264 (1975).
73. I.M. Chaiken, 'Quantitative Uses of Affinity Chromatography', *Anal. Biochem.*, **97**, 1–10 (1979).
74. A.J. Muller, P.W. Carr, 'Chromatographic Study of the Thermodynamic and Kinetic Characteristics of Silica-bound Concanavalin A', *J. Chromatogr.*, **284**, 33–51 (1984).
75. D.J. Anderson, R.R. Walters, 'Equilibrium and Rate Constants of Immobilized Concanavalin A Determined by High-performance Affinity Chromatography', *J. Chromatogr.*, **376**, 69–85 (1986).
76. J. Yang, D.S. Hage, 'Effect of Mobile Phase Composition on the Binding Kinetics of Chiral Solutes on a Protein-based High-performance Liquid Chromatography Column: Interactions of D- and L-Tryptophan with Immobilized Human Serum Albumin', *J. Chromatogr. A*, **766**, 15–25 (1997).
77. B. Loun, D.S. Hage, 'Chiral Separation Mechanisms in Protein-Based HPLC Columns. 2. Kinetic Studies on (R)- and (S)-Warfarin Binding to Immobilized Human Serum Albumin', *Anal. Chem.*, **68**, 1218–1225 (1996).
78. M. Glad, S. Ohlson, L. Hansson, M.-O. Mansson, K. Mosbach, 'High-performance Liquid Affinity Chromatography of Nucleosides, Nucleotides and Carbohydrates with Boronic Acid-substituted Microparticulate Silica', *J. Chromatogr.*, **200**, 254–260 (1980).
79. R.R. Maetas, J.R. Prieto, G.D. Kuehn, J.H. Hagemann, 'Polyacrylamide-Boronate Beads Saturated with Biomolecules: A New General Support for Affinity Chromatography of Enzymes', *J. Chromatogr.*, **189**, 225–231 (1980).
80. A.K. Mallia, G.T. Hermanson, R.I. Krohn, E.K. Fujimoto, P.K. Smith, 'Preparation and Use of a Boronic Acid Affinity Support for Separation and Quantitation of Glycosylated Hemoglobins', *Anal. Lett.*, **14**, 649–661 (1981).
81. R.J. Yon, 'Chromatography of Lipophilic Proteins on Adsorbents Containing Mixed Hydrophobic and Ionic Groups', *Biochem. J.*, **126**, 765–767 (1972).
82. S. Shaltiel, 'Hydrophobic Chromatography', *Methods Enzymol.*, **104**, 69–96 (1984).
83. K. Brocklehurst, J. Carlsson, M.P.J. Kierstan, E.M. Crook, 'Covalent Chromatography: Preparation of Fully Active Papain from Dried Papaya Latex', *Biochem. J.*, **133**, 573–584 (1973).
84. L.B. McGown, M.J. Joseph, J.B. Pitner, G.P. Vonk, C.P. Linn, 'The Nucleic Acid Ligand: A New Tool for Molecular Recognition', *Anal. Chem.*, **67**, 663A–668A (1995).
85. C. Turek, L. Gold, 'Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase', *Science*, **249**, 505–510 (1990).
86. A.D. Ellington, J.W. Szostak, 'In Vitro Selection of RNA Molecules that Bind Specific Ligands', *Nature*, **346**, 818–822 (1990).
87. D. Kriz, O. Ramstrom, K. Mosbach, 'Molecular Imprinting: New Possibilities for Sensor Technology', *Anal. Chem.*, **69**, 345A–349A (1997).
88. B. Sellergren, 'Noncovalent Molecular Imprinting: Antibody-like Molecular Recognition in Polymeric Network Materials', *Trends Anal. Chem.*, **16**, 310–319 (1997).

Biopolymer Chromatography

Christian G. Huber
Institute of Analytical Chemistry and Radiochemistry, Leopold-Franzens-University, Innsbruck, Austria

1 Introduction **2**
2 Separation Principles for Biopolymers **2**
2.1 Size-exclusion and Slalom Chromatography 3
2.2 Ion-exchange Chromatography 4
2.3 Normal-phase Chromatography 5
2.4 Reversed-phase Chromatography 5
2.5 Ion-pair Reversed-phase Chromatography 5
2.6 Hydrophobic Interaction Chromatography 6
2.7 Affinity Chromatography 6
2.8 Metal Interaction Chromatography 6
3 Stationary Phases for Biopolymer Separation **7**
3.1 Stationary Phase Support Materials and Functional Groups 7
3.2 Stationary Phase Morphologies with Enhanced Mass Transfer Properties 8
4 Biopolymer Detection **9**
4.1 Ultraviolet/Visible Absorbance and Fluorescence Detection 10
4.2 Electrochemical Detection 10
4.3 Radioactivity Detection 10
4.4 Pre- and Postcolumn Derivatization 11
4.5 Mass Spectrometry 12
4.6 Light-scattering Detection 12
5 Instrumental Requirements for Chromatographic Biopolymer Separations **12**
6 Separation of Peptides and Proteins **13**
6.1 Size-exclusion Chromatography 13
6.2 Ion-exchange Chromatography 13
6.3 Reversed-phase and Hydrophobic Interaction Chromatography 15
6.4 Metal Interaction and Affinity Chromatography 16
7 Separation of Nucleic Acids **17**
7.1 Size-exclusion Chromatography 17
7.2 Anion-exchange Chromatography 18
7.3 Ion-pair Reversed-phase Chromatography 19
7.4 Mixed-mode Chromatography 20
7.5 Affinity Chromatography 21
8 Separation of Oligo- and Polysaccharides **21**
8.1 Size-exclusion Chromatography 21
8.2 Normal- and Reversed-phase Chromatography 21
8.3 Ion-exchange Chromatography 22
8.4 Metal Interaction Chromatography 22
9 Modern Trends and Future Developments **23**
9.1 Miniaturization of Separation Systems 23
9.2 Multidimensional Separation of Biopolymers and Coupling to Mass Spectrometry 24
9.3 Capillary Electrochromatography 25
Abbreviations and Acronyms **25**
Related Articles **26**
References **26**

Biopolymers are biological macromolecules including peptides, proteins, nucleic acids, and polysaccharides whose physical and chemical properties differ considerably from those of small molecules. Biopolymer chromatography is a branch of liquid chromatography which is adapted to the specific requirements of biopolymers in order to obtain rapid and high-resolution chromatographic separations of these molecules. Because of the low diffusion coefficients of biopolymers, special stationary phase configurations and surface modifications are necessary to permit sufficiently fast mass transfer kinetics between the mobile and stationary phase. Exploiting different molecular properties of the biopolymers, such as size, charge, hydrophilicity, hydrophobicity, the ability to form complexes with metal ions, or biological function, the most popular liquid chromatographic modes for biopolymer separation are size-exclusion, ion-exchange, normal-phase, reversed-phase, ion-pair reversed-phase, metal-interaction, and affinity chromatography.

Because of their large size and the multitude of different functional groups present in biopolymers, more than one mechanism may be responsible for interaction with the stationary phase, resulting in enhanced selectivity due to mixed-mode interactions. Adsorption at multiple sites results in very steep elution isotherms as a function of the mobile phase solvent strength. Therefore, elution with a gradient of increasing solvent strength is usually applied in biopolymer chromatography. The choice of a suitable chromatographic environment is of special importance if

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the native confirmation of a biopolymer, which is essential for its biological activity, is to be preserved for subsequent experiments. Generally, conditions that come close to the physiological environment are most appropriate.

The most important detection techniques for biopolymers are ultraviolet/visible absorbance detection, fluorescence detection, electrochemical detection, mass spectrometry, and low-angle laser light scattering photometry. Miniaturization and multidimensional separation together with mass spectrometric detection allow the routine separation and identification of femtomole–attomole amounts of hundreds to thousands of different biopolymers contained in complex mixtures of biological origin.

1 INTRODUCTION

Biopolymers are naturally occurring macromolecules, primarily peptides, proteins, nucleic acids, and polysaccharides, that represent essential constituents of all living organisms.[1] Their molecular mass ranges from roughly 1000 to several million. Although biopolymers exhibit an immense structural and functional diversity, they share a common building principle where relatively few building blocks of different chemical structure are assembled into polymeric chains that constitute the three-dimensional structure of the biomacromolecule. Thus, an estimated number of 10^{11} naturally occurring proteins is built up by only 20 amino acids and the whole human genome comprising 3×10^9 nucleotides contains only four different nucleobases. The characterization and structural analysis of biopolymers is of utmost importance in all areas of biological sciences and represents the key to understanding their biological function.

Biopolymers are usually present in relatively complex mixtures, hence separation before structural investigation is obligatory. The two principal and most powerful methods to asses the composition of biological samples and to separate the individual components are liquid chromatography (LC)[2] and electrophoresis.[3] The extraordinary popularity of high-performance liquid chromatography (HPLC) for the separation and analysis of biopolymers can be attributed to a number of factors:

1. Excellent resolution can be achieved for both closely related and structurally disparate substances under a large variety of chromatographic conditions.
2. Chromatographic selectivity can be manipulated readily through changes in mobile phase composition and/or choice of a suitable stationary phase from a large number of commercially available stationary phases.
3. Even submicrogram amounts of biopolymers can be isolated with high recovery.
4. Rapid separation and quantitation can be carried out with high reproducibility and accuracy on a timescale ranging from a few seconds to hours.

In principle, the theory of LC for biopolymers is the same as that for small molecules (see **Column Theory and Resolution in Liquid Chromatography**). However, as the diffusion coefficient decreases with increasing solute molecular mass, 1–2 orders of magnitude can separate the diffusion rates of small and large molecules in the liquid phase (e.g. diffusion coefficient of 9×10^{-6} versus $0.6 \times 10^{-6}\,cm^2\,s^{-1}$ for alanine and hemoglobin in water, respectively). The slow diffusion characteristics of biopolymers are responsible for slow kinetics of adsorption/desorption at the stationary phase and secondary equilibria resulting in lower column efficiency relative to separations involving small molecules. On the other hand, the separation selectivity for biopolymers is often better than for small molecules allowing the application of gradient elution, which counterbalances the lower column efficiency and has the additional advantage of increasing the peak capacity (see **Gradient Elution Chromatography**).

In many cases, especially for proteins, the three-dimensional structure of a biopolymer is essential for its biological activity. Therefore, careful choice of the appropriate method and chromatographic conditions is necessary to preserve the three-dimensional structure during the chromatographic separation process if the separated material is to be used for subsequent investigations where biological activity is important. Generally, chromatographic conditions similar to the physiological environment, such as aqueous buffer solutions at neutral pH, are most appropriate for the preservation of biological activity. However, biopolymers in solution can have parts of the molecule as well as the whole molecule folded, resulting in a variety of different conformations. If the time required for interconversion of these various more or less folded species is comparable to the time of separation, the resulting peak profiles are quite wide and often misshapen, resulting in poor separation efficiency. Therefore, strongly denaturing conditions, e.g. extreme pH, high temperature, or organic solvents, help to improve the resolving power of chromatography as a result of higher chromatographic selectivity for unfolded or partially denatured forms than for the native form of biopolymers at the cost of loss of biological activity.

2 SEPARATION PRINCIPLES FOR BIOPOLYMERS

There are a great number of chromatographic methods available for the fractionation of biopolymers that are designed to sense differences in size, shape, flexibility,

charge, polarity, or biological function of the compounds to be separated.[4] It has been appreciated that the so-called noninteractive modes of chromatographic separation do not exhibit the same level of resolution as the interactive modes.[2] Consequently, the most successful separation techniques are those capable of selective interaction by probing the topography of biopolymers.

Size-exclusion chromatography (SEC) and slalom chromatography represent exceptional modes in that the separation of the sample components takes place without their interaction with the stationary phase and, therefore, they are called noninteractive modes of chromatography (Table 1). By contrast, the interactive modes are conveniently identified by the term of the particular interaction (Table 1). Thus, we distinguish between electrostatic interaction, where immobilized charges on the stationary phase interact with the charged biopolymer molecules, polar interaction, where permanent or induced dipoles interact with each other, and hydrophobic interaction, where the interaction is between hydrophobic moieties. Since charged, polar, and hydrophobic domains are all present on the surface of most biopolymers, these interactions are of general type. In metal interaction chromatography (MIC), biopolymers are retained by the semispecific complexation of an immobilized metal ion at the stationary phase. Involving electrostatic, polar, and dispersive interactions as well as steric factors of interaction at the same time, affinity chromatography (AC) exploits highly specific interactions characteristic of many biochemical systems.

Table 1 Principal chromatographic modes for biopolymer separation

Chromatographic mode	Acronym	Separation principle
Noninteractive modes of liquid chromatography:		
Size-exclusion chromatography	SEC	Differences in molecular size
Slalom chromatography	–	Differences in length and flexibility
Interactive modes of liquid chromatography:		
Ion-exchange chromatography	IEC	Electrostatic interactions
Normal-phase chromatography	NPC	Polar interactions
Reversed-phase chromatography	RPC	Dispersive interactions
Ion-pair reversed-phase chromatography	IPRPC	Electrostatic interactions
Hydrophobic interaction chromatography	HIC	Dispersive interactions
Affinity chromatography	AC	Biospecific interaction
Metal interaction chromatography	MIC	Complexation with an immobilized metal

Macromolecular solutes interact with the stationary phase in a different way from that of small molecules. Because of the large molecular dimensions of biopolymers, it can be expected that the geometry and distribution of charged, hydrophobic, polar, and complexing groups on their surface greatly influence the adsorption behavior. Several of such adsorption sites on the biopolymer surface make it highly probable that adsorption at the stationary phase will occur at multiple sites.[5] This characteristic, unique to macromolecules, is responsible for extremely steep elution isotherms resulting in an enormously rapid change in solute retention with small changes in the elution strength of the mobile phase, a phenomenon often referred to as the 'on–off' model.[4] According to this model, the solute initially sticks at the inlet of the column and at some critical solvent strength during a gradient run it is totally desorbed and rapidly elutes from the column (Figure 1).

2.1 Size-exclusion and Slalom Chromatography

SEC separates biopolymers on the basis of their size and is different from the interactive modes of chromatography in that any interaction between the analytes and the stationary phase has to be prevented. The size-exclusion effect is based on the fact that the stationary phase particles have pores with a defined distribution of pore diameters. Large molecules of dimensions larger than

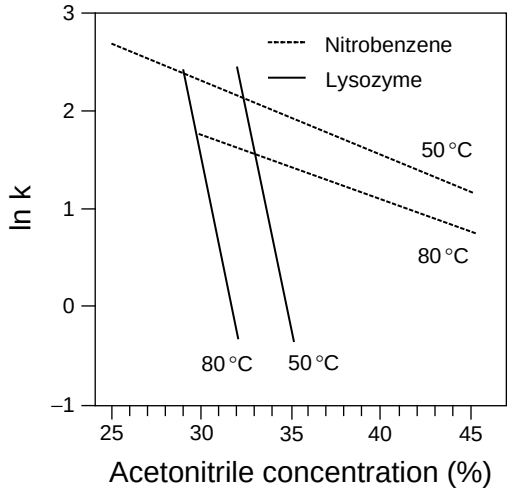

Figure 1 Plots of the logarithmic retention factor versus the concentration of acetonitrile in the RPC (reversed-phase chromatography) of lysozyme (M_r 16 951.5, solid lines) and nitrobenzene (M_r 123.11, dashed lines) on a PLRP-S macroporous, cross-linked polystyrene stationary phase from Polymer Laboratories. (Reproduced from H. Chen, C. Horváth, 'Rapid Separation of Proteins by Reversed Phase HPLC at Elevated Temperatures', *Anal. Methods Instrum.*, **1**, 213–222 (1993), Copyright John Wiley & Sons Limited. Reproduced with permission.)

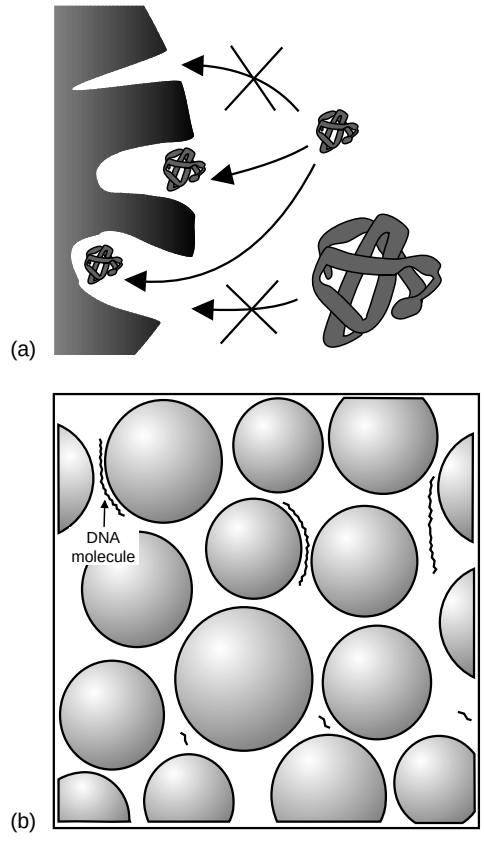

Figure 2 Principles of (a) SEC and (b) slalom chromatography.

the pore diameter are totally excluded from the internal volume and leave the column first (Figure 2a). For small molecules, all pore diameters are large enough to allow penetration of the whole internal volume. Molecules of intermediate size have access only to a portion of the internal volume but will be excluded from the smaller pores. Thus, in SEC, analytes are separated by molecular size.

The upper molecular mass limit of SEC is determined by the maximum pore diameter of the packing material and the shape of the biomacromolecule. For instance, approximately 20% of the pore volume of a stationary phase with 10 nm average pore diameter is accessible for proteins with a molecular mass of 50 000 whereas 50% of the pore volume of a stationary phase with 50-nm pores is accessible for proteins with a molecular mass of around 500 000. The molecular size of linear polymers of random coil structure, such as nucleic acids or linear polysaccharides, is closely related to the molecular mass. Proteins, however, exist in an ordered tertiary structure, which is much denser than that of random coils. Consequently, they can penetrate more of the pore volume than random coiled biopolymers of commensurate molecular mass. In Table 2, the molecular masses and sizes of some biopolymers are summarized.

Table 2 Molecular masses and sizes of some biopolymers

Biopolymer	M_r	Molecular dimensions (nm)
Cytochrome *c*	12 310	$2.5 \times 2.4 \times 3.7$[a]
Carbonic anhydrase	28 800	$4.7 \times 4.1 \times 4.1$[a]
Lactate dehydrogenase	146 200	$7.4 \times 7.4 \times 8.4$[a]
Ferritin	474 000	$8 \times 8 \times 8$[a]
DNA, 100 base pairs	69 400	$35 \times 2 \times 2$ (straight rod)
DNA, 800 base pairs	519 200	$60 \times 60 \times 60$ (random coil)[b]

[a] Determined from crystal structure.
[b] Calculated by the Kratky–Porod relation.[6]

Slalom chromatography has been developed for the separation of large, double-stranded DNA molecules (>1000 base pairs) that behave like a flexible random coil in aqueous solution.[7] When DNA molecules are applied to a column which is packed with hard and spherical beads, they are unfolded and extended owing to the laminar flow of solvent passing through the narrow openings between the closely packed particles. The separation is based on the fact that long DNA molecules (5000–50 000 base pairs) have to turn very fast during their passage through the narrow and tortuous channels between the packing particles (Figure 2b). The longer the DNA molecule, the more difficult it is to pass through the openings and, consequently, a separation according to size, where smaller DNA molecules elute first, is achieved. Important features of slalom chromatography are that DNA fragments do not interact with the column packing, that separation depends strongly on particle size and flow rate, and that the DNA fragments are totally excluded from the pores, and therefore pore size and pore volume are not important factors in determining the separation.

2.2 Ion-exchange Chromatography

Ion-exchange chromatography (IEC) separates molecules based on differences in their accessible surface charges (see **Ion Chromatography**). An ion-exchange stationary phase consists of positively charged (for anion exchange, see Figure 3a) or negatively charged (for cation exchange) functional groups immobilized on a suitable support material. The electrostatic interactions between the oppositely charged groups on the surface of the stationary phase and on the surface of the analytes are responsible for adsorption. Since biopolymers also contain nonpolar areas, relatively weak dispersive interactions with the stationary phase may provide additional retention.

Desorption takes place when the ionic strength of the mobile phase is increased by applying a gradient of increasing salt concentration. The salt concentration needed to displace the adsorbed biopolymers back into the mobile phase depends on the strength of the

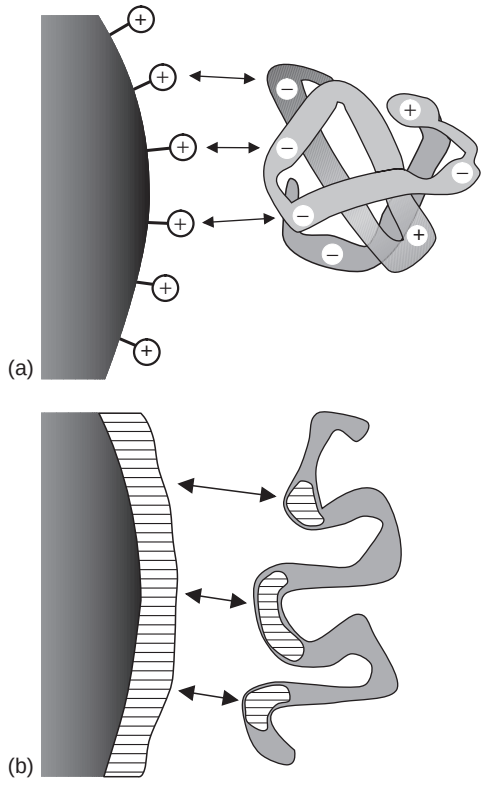

Figure 3 Principles of (a) IEC and (b) RPC. Shaded areas represent hydrophobic regions of the protein.

electrostatic interactions between the stationary phase and the biopolymer.

Consequently, the retention behavior of biopolymers in IEC depends strongly on their charge characteristics. The total net charge and the actual distribution of charges within a biomacromolecule critically affect both adsorption and elution. Whereas the phosphodiester groups of nucleic acids are dissociated over a broad pH range (2–14), the pH of the mobile phase strongly influences the charge characteristics of peptides and proteins. Oligosaccharides can be ionized only at pH values higher than 12 owing to the weak acidity of the hydroxyl groups. In addition to affecting the charge distribution within the molecule, the pH can also have an effect on the three-dimensional conformation of a biomacromolecule. Because of these phenomena, the effect of pH on the chromatographic selectivity can be profound and difficult to predict and an empirical approach to evaluate an ion-exchange separation over a range of pH values is necessary. The advantage of the ion-exchange mechanism is that under suitable pH conditions, relatively few or no structural changes in the biopolymers occur because the interaction between the stationary phase and the analyte is weak apart from the electrostatic interactions which are directed and specific.

2.3 Normal-phase Chromatography

In normal-phase chromatography (NPC), the stationary phase is more polar than the mobile phase (see **Normal-phase Liquid Chromatography**). The intermolecular interactions governing retention and selectivity in this mode are polar in origin, encompassing both hydrogen bonding, the extent of which depends upon the Lewis acidity or basicity of the interacting molecules, and dipole–dipole interaction, which is dependent on the dipole moments and polarizabilities of the molecules. The polar surface of a normal-phase stationary phase can either be obtained by derivatizing a support material with a suitable polar functional group, or it can be the original support, as in the case of silica gel. The retention of polar analytes increases with increasing hydrophilicity of the stationary phase and with decreasing hydrophilicity of the mobile phase, typically mixtures of organic solvents.

2.4 Reversed-phase Chromatography

Reversed-phase chromatography (RPC) separates biopolymers on the basis of differences in their hydrophobic properties (see **Reversed Phase Liquid Chromatography**). The stationary phase in RPC is a very hydrophobic, nonpolar surface, whereas the mobile phase is polar, usually a mixture of water and organic solvents. The hydrophobic interaction results from a tendency of nonpolar molecules to interact with each other rather than with water, resulting in the distribution of a hydrophobic solute between the hydrophobic hydrocarbon chains of the stationary phase and the mobile phase (Figure 3b). Reducing the polarity of the mobile phase through increasing the concentration of an organic solvent allows the hydrophobically bound analytes to desorb from the stationary phase, causing elution from a reversed-phase column. Under the conditions of RPC, i.e. the employment of hydroorganic solvents in combination with a strongly hydrophobic surface, proteins often lose their three-dimensional shape and denature, resulting in loss of biological activity, which cannot always be restored by putting the protein in a more physiological environment after separation. Nevertheless, the high resolution capability of this technique is a result of higher chromatographic selectivity for unfolded or partially denatured forms than for the native forms of the proteins, which makes RPC an important tool for high-resolution analytical separations.

2.5 Ion-pair Reversed-phase Chromatography

Ion-pair reversed-phase chromatography (IPRPC) is traditionally seen as a special branch of RPC because stationary phases and elution conditions typical for RPC are used. However, from a mechanistic point of

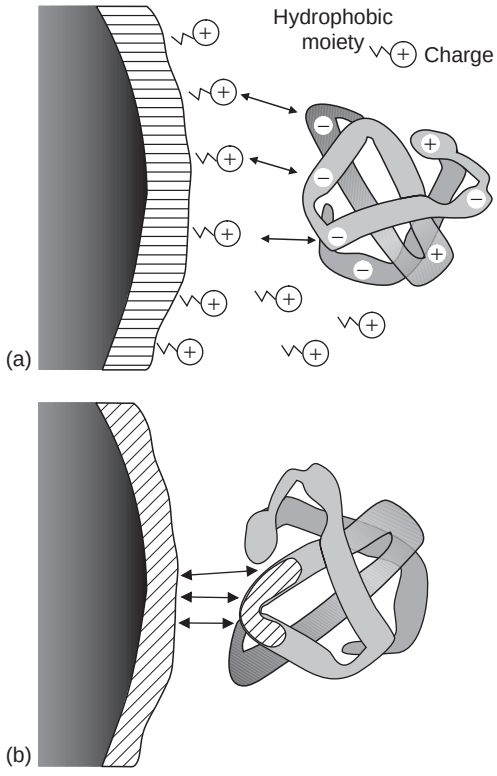

Figure 4 Principles of (a) IPRPC and (b) HIC. Shaded areas represent hydrophobic regions of the protein.

view, it is rather a branch of IEC because retention is governed by electrostatic interactions. In IPRPC, the phase system comprises a hydrophobic stationary phase and a hydroorganic eluent containing an amphiphilic ion that carries both charged and hydrophobic moieties. According to the electrostatic retention model,[8] the positively or negatively charged amphiphilic ions are adsorbed at the interface between the nonpolar stationary phase and the hydroorganic mobile phase, resulting in a surface potential that electrostatically interacts with solute molecules of opposite charge (Figure 4a). Elution of bound molecules is accomplished by increasing the concentration of an organic modifier, resulting in a decrease in surface potential because of desorption of amphiphilic ions from the stationary phase. Additional hydrophobic interactions between hydrophobic solutes and the hydrophobic surface of the stationary phase are possible in IPRPC.

2.6 Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography (HIC) separates biopolymers based on the hydrophobicity of their surface. Binding of biopolymers is induced by the addition of high salt concentration ($1–5\,\text{mol}\,\text{L}^{-1}$) whereas elution is accomplished by lowering the salt concentration (Figure 4b). The organization of the water molecules surrounding the analytes and the binding surface is the driving force for hydrophobic adsorption. When a hydrophobic region of a biopolymer binds to the surface of a mildly hydrophobic stationary phase, hydrophilic water molecules are effectively released from surrounding the hydrophobic areas, causing a thermodynamically favorable change in entropy. Because hydrophobic interaction is driven by entropy, temperature can have a strong effect.

Compared with RPC, the surface of the stationary phase in HIC is less hydrophobic so that no organic solvent is necessary for desorption of the analytes. Therefore, the three-dimensional structure of biopolymers is usually preserved during a separation by HIC. Ammonium sulfate, by virtue of its good salting-out properties and high solubility in water, is the most common salt used for HIC. The salt concentration required for the induced hydrophobic interaction can be modulated by changing the hydrophobicity of the stationary phase. More hydrophobic phases require less salt for binding. If the bonded phase is too apolar, the analytes do not elute when the salt is removed.

2.7 Affinity Chromatography

AC is based on biospecific interactions between a ligand covalently bound to the chromatographic packing and a target biopolymer in the sample solution (see **Affinity Chromatography**). The technique is almost exclusively utilized to purify proteins, although there are a few applications with peptides and nucleic acids. The forces involved in biospecific interaction include the same electrostatic and hydrophobic interactions that cause more nonspecific ion-exchange and hydrophobic binding. However, in the case of biospecific or affinity binding, the interacting groups are arranged on the two bonding molecules in a unique orientation. Moreover, hydrogen bonding also plays an important role. The two ligands thus fit together very much like a lock and key, with a high degree of specificity (Figure 5a). Elution from affinity media can be done under any conditions that disrupt the ligand–ligand interaction. The most common technique is to employ a shift to acidic pH (2–4), which induces dissociation because of protonation of acidic groups. Other elution techniques include the increase of ionic strength or the use of specific displacers such as the immobilized ligand or an analog in free solution. Affinity elution is usually in the form of a step gradient.

2.8 Metal Interaction Chromatography

MIC is a method for the separation of biopolymers which are capable of complexing with an immobilized

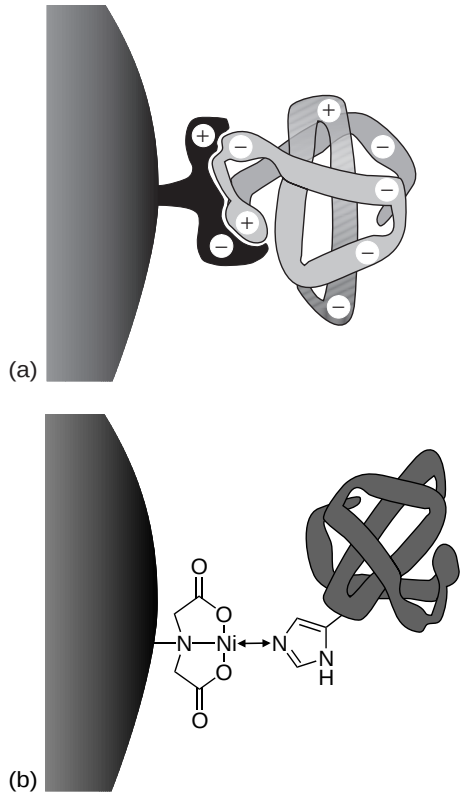

Figure 5 Principles of (a) AC and (b) MIC.

transition metal ion on the surface of a suitable support. The metal ion (typically Cu^{2+}, Zn^{2+}, Ni^{2+}, Co^{2+}, or Fe^{3+}) is usually bound to the support by means of a chelating ligand (typically iminodiacetate) such that one or more coordination sites are available for interaction with ligands that are part of biopolymers (Figure 5b). Typical functional groups in biopolymers capable of complexation with transition metal ions are imidazole, amino, and thiol groups present in the amino acids histidine, lysine, and cysteine (with Cu^{2+}, Ni^{2+}), phosphate groups in nucleic acids (with Fe^{3+}), and vicinal diols present in saccharides (with Cu^{2+}). Elution of complexed biopolymers is accomplished with gradients of increasing salt concentration or, even more effective, with increasing gradients of competitive ligands such as imidazole or phosphate.

3 STATIONARY PHASES FOR BIOPOLYMER SEPARATION

The choice of a suitable stationary phase is one of the first decisions to be made in the design of a chromatographic separation for a biopolymer mixture. Initially, it was assumed that the pool of packing materials commercialized for the separation of low molecular mass analytes could be easily transferred to the separation of biopolymers.[9] However, it soon became apparent that the handling of these substances in HPLC requires careful control of the morphology and the surface chemistry of the chromatographic packing material. The two major concerns that have to be taken into account with biopolymers are slow mass transfer because of the low diffusivities of biopolymers and multimodal interaction of biopolymers with the stationary phase. Usually it is advantageous to restrict the molecular interactions between analyte molecules and the stationary phase to one predominant interaction (single-mode interaction, Table 1), although other forces and mechanisms may play a role in determining the final separation (mixed-mode or multimode interaction). In many cases, however, these additional interactions are difficult to control and result in band broadening and asymmetric peak shapes.

For size-exclusion packing materials, the pore size distribution and a surface chemistry that prevents any interaction between the stationary phase and the analyte are the factors determining the success of separation. In the interactive modes of chromatography the driving forces of retention and selectivity are the molecular interactions between the analyte and the surface of the stationary phase. Additionally, the steric environment of the packing materials affects the molecular interactions. The selection of an optimal pore size for a particular sorbent is made on the basis that the solute molecular diameter (Table 2) must be a least one-tenth the size of the pore diameter of the packing material to avoid restricted diffusion of the solute and to allow the total surface of the sorbent material to be accessible. Therefore, compared with packings for separating small molecules, larger pore sizes of 15–100 nm or nonporous packing materials are needed for biopolymers.

3.1 Stationary Phase Support Materials and Functional Groups

Modern support materials used in HPLC are either derived from inorganic or organic polymers or are composites of these types. Inorganic supports comprise hydrous oxides (silica, alumina), phosphates (hydroxyapatite), silicates, porous glasses, and carbon materials. Usually, inorganic supports are hydrophilic (except for carbon, e.g. porous graphitic carbon) and possess acidic and/or basic functional groups. Cross-linked organic polymers exhibit a wide range of polarity, depending on the polymer backbone and the functional groups. The most prominent organic polymers include polyamide, polybutadiene, polysaccharides, polymethacrylates, poly(vinyl alcohol), and styrene–divinylbenzene copolymers. Silica-based packing materials are used in about 75% and

Table 3 Functionalities needed for different modes of chromatography

Chromatographic mode	Functional group	Remarks
SEC	Surface functionality is chosen such that no molecular interactions occur	Selectivity is provided by the pore-size distribution of the packing
Slalom	Surface functionality is chosen such that no molecular interactions occur	Selectivity is provided by the particle diameter of the packing combined with the flow rate
IEC	Sulfonic acid, carboxylic acid	For cation exchange
	Primary, secondary, tertiary, or quaternized amino groups	For anion exchange
NPC	–	Underivatized silica gel
	Hydroxyl-, diol-, amino-, cyano groups	Bound to silica gel or polymeric supports
RPC and IPRPC	–	Underivatized nonpolar polymers, e.g. PS/DVB
	Butyl, phenyl, octyl, or octadecyl groups	Bound to silica gel or polymeric supports
HIC	Methyl, butyl, or phenyl groups	–
AC	Epoxy, aldehyde, or aryl ester groups	Preactivated packings are commercially available; the specific affinity ligand is usually attached by the user
MIC	Iminodiacetate groups with complexed transition metals	Metals complexed: Ni(II), Co(II), Cu(II), Zn(II), Fe(III)

polymers in about 20% of all separations performed today.[10]

Silica gel is available with irregular- and spherical-shaped particles and with a wide variety of particle diameters, pore diameters, and pore volumes. Moreover, silica-based stationary phases are mechanically stable, provide excellent efficiency, and can be converted to bonded-phase materials by virtue of their reactive silanol groups (see **Silica Gel and its Derivatization for Liquid Chromatography**). However, silica has some limitations: it dissolves rapidly at pH > 9 and the acidic silanol groups cause tailing or irreversible adsorption of basic compounds.

The most popular polymeric support is polystyrene cross-linked with divinylbenzene [poly(styrene/divinylbenzene) (PS/DVB)]. Depending on the degree of cross-linking and the amount of porogen used in their production, polymeric packings are also available with a variety of particle diameters, porosities, and mechanical strengths. Polymeric packings are generally stable over a wider pH range than silica gel, and some can operate throughout the full pH 0–14 range. Polymers do not exhibit silanol activity, however hydrophobic or charged entities of the polymer backbone can be a source of additional interactions resulting in multimode adsorption.

Both types of support materials, inorganic oxides or organic polymers, can be surface functionalized by well-established methods[11,12] to obtain suitable stationary phases for the various chromatographic modes. Table 3 summarizes some of the most commonly used functionalities for the different modes of biopolymer chromatography.

A tremendous number of packing materials for different modes of chromatography are commercially available, and the market constantly changes with the introduction of new materials. Therefore, for the selection of a column for specific applications the reader is referred to comprehensive tables in recent monographs[13,14] or periodical reviews appearing in analytical journals.[10,15,16]

3.2 Stationary Phase Morphologies with Enhanced Mass Transfer Properties

Since slow mass transfer kinetics is often the limiting factor for speed and efficiency in biopolymer separations, the enhancement of intraparticular mass transport is particularly important for the rapid separation of large molecules having low diffusivities.[17,18] The different routes to enhance the mass transfer of biopolymers in stationary phases include (1) increasing the pore diameter and decreasing the particle diameter, (2) eliminating the support pores by using nonporous stationary phases, (3) decreasing the pore depth by using superficially porous stationary phases, (4) introducing pores of 600–800 nm diameter transversing the particle that cause the liquid to flow through the particle, and

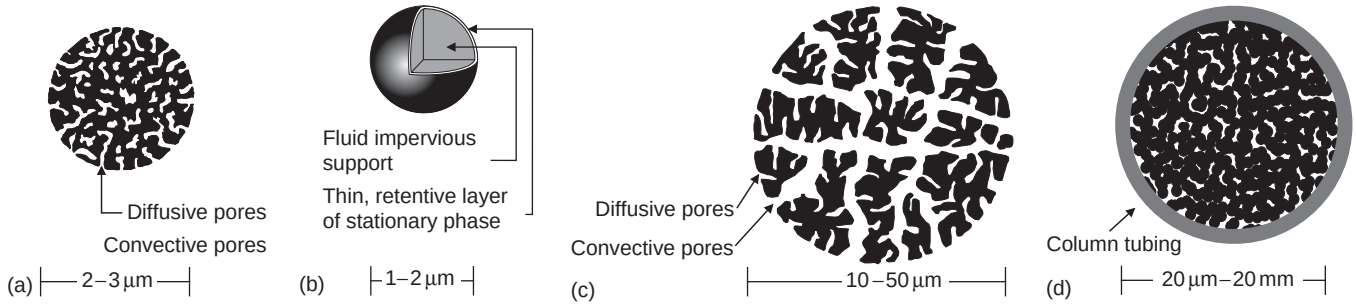

Figure 6 Schematic illustration of stationary phase configurations with enhanced mass transfer properties. (a) Macroporous packing; (b) nonporous packing; (c) bidisperse packing with macropores and gigapores; and (d) continuous chromatographic bed.

(5) using porous, continuous-bed polymers as the chromatographic bed.

The first attempts to adapt stationary phases for biopolymer separations applied small particles with large pores (macroporous particles)[9] that yield enhanced column efficiency by virtue of the relatively small intraparticular mass transfer resistance due to the short diffusion distances (Figure 6a). Another way to circumvent intraparticle diffusion is the complete elimination of the support pores.[19,20] Through the use of nonporous particles that lack any intraparticular pores, the only remaining particle-based diffusion limitations are in a thin layer of stagnant liquid at the surface of the particle (Figure 6b). The only limitation of nonporous sorbents is that they have approximately 1% of the surface area of porous sorbents and are, therefore, of limited loading capacity for preparative separations. This problem has been addressed by attaching a 1-μm thick layer of porous particles to the exterior surface of nonporous particles.[21] These poreshell particles with a total diameter of around 5 μm have pores of 1 μm depth or less and a loading capacity around 65% of that of a totally porous particle.

Another approach to the problem of stagnant mobile phase mass transfer is to cause the liquid to flow (or perfuse) through the particles. Introduction of 600–800-nm pores transversing the particle allows a small portion of the mobile phase to perfuse the porous support.[22] The surface area of perfusion chromatography supports is increased by a network of smaller, 50–150-nm pores of 1 μm or less in depth that interconnect the large throughpores (Figure 6c). Finally, the concept of continuous chromatographic beds, where the separation medium consists of a continuous rod of a rigid, macroporous polymer which has only internal porosity consisting of macropores and gigapores, has been developed to alleviate the problem of mass transfer (Figure 6d).[23,24] Because of the absence of intraparticular volume, all of the mobile phase is forced through the large pores in the separation medium, resulting in enhanced mass transport by convection.

4 BIOPOLYMER DETECTION

In this section, the aspects of detection important for biopolymers are emphasized.[14] There is no one single ideal detector for biopolymers, hence the chromatographer is forced to utilize several detection systems that all sense different molecular properties of the solute to be detected such as absorption of electromagnetic radiation, fluorescence, electrochemical activity, and molecular mass. The usual detector requirements also prevail for HPLC of high molecular mass biopolymers and include high linear dynamic range, high sensitivity, low limit of detection, low detector noise level, baseline stability, high signal-to-noise ratio, fast time constant, low peak dispersion, and the ability to quantitate with good detectability, reproducibility, and accuracy.[25] Moreover, an ideal detector system for biopolymer chromatography should be nondiscriminating in the sense of being able to detect with low limits of detection the variety of different compounds that are present in complex mixtures of biological origin. From another point of view, detection systems should yield information about the identity and structure of the detected biopolymer to permit the differentiation between biopolymers even of very similar structure. Another ideal property of a biopolymer detector is nondestruction, that is, the ability of samples to be recovered after detection to be used for subsequent analyses or biological testing. The most commonly used detection methods for biopolymers are ultraviolet/visible (UV/VIS) absorbance detection, fluorescence (FL) detection, electrochemical detection (ELCD), mass spectrometry (MS), and low-angle laser light scattering (LALLS) photometry. Although suffering from high limits of detection, refractive index detection is one of the few options for universal detection in LC and has to be used for biopolymers that lack any chromophore, fluorophore, or electroactive group.

4.1 Ultraviolet/Visible Absorbance and Fluorescence Detection

Spectrophotometric techniques are among the most widely applicable for the detection of biopolymers. The absorbance maximum for the peptide bond is around 185 nm, but most eluents used for chromatographic separations absorb light strongly below 200 nm. Therefore, peptides and proteins are usually detected at wavelengths between 210 and 220 nm. Nevertheless, a host of other ultraviolet (UV)-absorbing compounds present in biological samples may cause interferences with separation and quantitation. In this situation, alternative wavelengths that can be monitored include the regions of absorbance of the side chains of tryptophan and tyrosine (280 nm) or, less commonly, phenylalanine (254 nm), cystine (240 nm), or histidine (228 nm). Proteins containing prosthetic groups, such as metalloproteins, often may be selectively detected in the far-UV or visible (VIS) region, the wavelength being dependent on the oxidation state of the metal. Cytochrome *c* oxidase may be detected near 410 nm, whereas myoglobin absorbs at 405 nm, cytochromes at 550–560 nm, and hemoglobin at 415 nm.[26]

Nucleotides, the monomer units of nucleic acids, are UV active over the range 180–300 nm, with maxima at 180–200 and 250–280 nm. For the same reasons as described above for peptides and proteins, nucleic acids are normally detected at wavelengths between 250 and 260 nm. Polysaccharides are not ideal UV/VIS absorbers because of the absence of effective chromophores absorbing above 200 nm. Thus, underivatized polysaccharides have poor UV absorbance detection limits and other detection methods such as UV/VIS or FL detection after chemical derivatization or ELCD have to be used.

FL detection generally has very low limits of detection (10–1000 times lower than UV/VIS detection) but, as only a few percent of all compounds exhibit native FL, the method is highly selective. Non-fluorescing compounds can be tagged with fluorescent dyes by pre- or postcolumn derivatization. The native FL of tryptophan and tyrosine can be used for peptide and protein detection. By exciting at 220 nm and monitoring at 330 nm, tryptophan-containing peptides are detected with six times lower detection limits compared with UV absorbance detection. Native nucleic acids and polysaccharides usually do not contain any fluorophore so that FL detection of these compounds is possible only after derivatization with an appropriate fluorescent label (see section 4.4).

4.2 Electrochemical Detection

In all forms of ELCD the analytes are detected by the current generated at a working electrode by oxidation or reduction. The direct electrochemistry of biopolymers on noble metal electrodes frequently suffers from electrode fouling due to adsorption of oxidized species at the electrode surface. This problem has been addressed by the application of pulsed ELCD, where the electrode undergoes oxidative cleaning followed by reductive regeneration of the electrode surface through the application of a triple-pulse waveform. If the current is measured only at a specific voltage during the pulse, the technique is known as pulsed amperometric detection (PAD).

Amino acids that respond to ELCD include tyrosine, tryptophan, methionine, and cysteine/cystine with the electroactive groups phenol, indole, thioether, thiol, and disulfide, respectively. ELCD is, therefore, mainly used for selective determination of peptides containing such amino acids. A more general strategy for the ELCD of peptides is taking advantage of complexation reactions between copper(II) and amino acids at a copper electrode.[27] However, using ELCD for a wide range of peptides cannot currently be considered a general approach, since many peptides will not be detected.

Diffusion and proper orientation of the analyte to be oxidized or reduced at the electrode surface is of great importance in ELCD. The general use of ELCD for proteins is restricted by these two requirements, since proteins possess low diffusion coefficients and the electroactive groups may be buried inside the folded protein structure. Hence ELCD is restricted to certain specific proteins having special functionalities, e.g. heme groups,[28] accessible disulfide bridges,[29] or copper.[30]

Electrochemical oxidation of the purine bases at a carbon electrode can be used to determine nucleic acids.[31] The guanine and adenine bases can be readily distinguished owing to their different oxidation potentials (800 mV for guanine, 1080 mV for adenine versus Ag/AgCl). Nevertheless, like proteins, high molecular mass nucleic acids possess low diffusion coefficients, which eventually lead to poor sensitivity for larger nucleic acids.

PAD is the method of choice for detection of oligo- and polysaccharides with high sensitivity.[32] Aldehyde and terminal alcohol moieties in oligosaccharides can be oxidized at gold electrodes in alkaline media. A typical triple-pulse waveform for detection of oligosaccharides involves anodic oxidation of the analytes at a potential of 50–200 mV for 200–400 ms, oxidative cleaning of the electrode surface at 600–800 mV for 120–180 ms, and reduction of the electrode surface at −300 to 800 mV for 300–360 ms (Figure 7, all potentials versus Ag/AgCl).

4.3 Radioactivity Detection

Because of its very low detection limits, radioactivity (RA) detection of biopolymers is primarily applied in

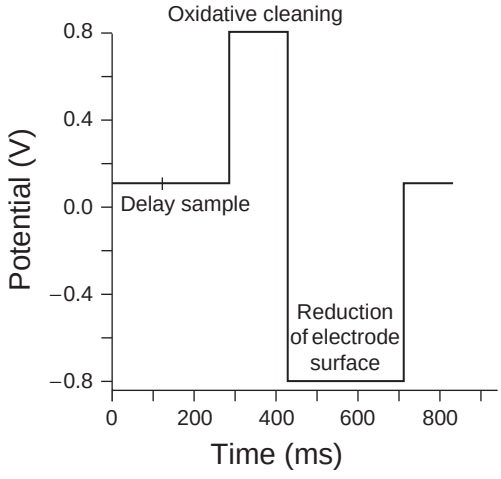

Figure 7 Pulsed waveform for ELCD of saccharides.

studies involving very small amounts of sample in complex matrices. As the natural RA of the elements contained in biopolymers is extremely low, chemical or enzymatic labeling with radioelements of high specific activity is obligatory. Radioisotopes emitting β-radiation such as ^{3}H, ^{14}C, ^{32}P, or ^{35}S are the most commonly used radioactive labels. Examples of radioactive labeling reactions are given in Table 4.

4.4 Pre- and Postcolumn Derivatization

A chemical derivatization reaction allows the conversion of biopolymers with poor UV/VIS absorptivity or electrochemical activity into derivatives that can be detected with higher sensitivity. Moreover, chemical derivatization often enhances the selectivity when only certain functional groups characteristic of a biopolymer react during the derivatization procedure. Derivatizations can be performed before (precolumn) and after (postcolumn) a chromatographic separation. Postcolumn techniques allow the detection of biopolymers after separation in their native state, whereas precolumn derivatization changes both detection and chromatographic behavior. Tagging of biopolymers with FL labels is the method of

Table 4 Derivatization reagents for UV/VIS, FL, ELCD, and RA detection of biopolymers

Reagent	Used for	Derivatization	Detection[a]	Refs.
o-Phthaldialdehyde	Peptides	Precolumn, postcolumn	FL, Ex 340 nm/Em 400 nm	33
Fluorescamine	Peptides, proteins	Precolumn, postcolumn	FL, Ex 390 nm/Em 490 nm	34
Indocyanine green	Proteins	Precolumn	FL, Ex 765 nm/Em 820–840 nm	35
[^{14}C]Acetic anhydride or [^{3}H]acetic anhydride	Proteins	Precolumn	RA, β^- 0.156 MeV RA, β^- 0.018 MeV	36
[γ-^{32}P]ATP/protein kinase	Peptides and proteins containing threonine or tyrosine	Precolumn	RA, β^- 1.71 MeV	37
Fluorescein	Nucleic acids	Precolumn	FL, Ex 450 nm/Em 520 nm	38
Phenylglyoxal	Nucleic acids containing guanosine	Postcolumn	FL, Ex 365 nm/Em 515 nm	39
Chloroacetaldehyde	Nucleic acids containing adenosine	Postcolumn	FL, Ex 254 nm/Em 400 nm	40
[γ-^{32}P]ATP/polynucleotide kinase	Nucleic acids	Precolumn	RA, β^- 1.71 MeV	41
Thymol	Reducing and nonreducing polysaccharides	Postcolumn	VIS, 500 nm	42
1-Phenyl-3-methyl-5-pyrazolone	Reducing polysaccharides	Precolumn	UV, 245 nm	43
2-Aminopyridine	Reducing saccharides	Precolumn	UV, 290 nm FL, Ex 320 nm/Em 400 nm	44
2-Cyanoacetamide	Reducing saccharides with a β-hydroxyl group	Postcolumn	UV, 270, 330 nm FL, Ex 331 nm/Em 383 nm ELCD	45, 46
Sodium [^{3}H]borohydride	Polysaccharides	Precolumn	RA, β^- 0.018 MeV	47

[a] Em, emission wavelength; Ex, excitation wavelength.

choice for derivatization because of the high sensitivity of FL detection (approximately 1000 times more sensitive than UV/VIS detection). Whereas FL derivatization of peptides, proteins and nucleic acids is necessary only when UV absorbance detection of the native compounds is not sensitive enough, derivatization of polysaccharides is very common owing to the lack of chromophoric or fluorogenic groups. Table 4 lists some examples of derivatization reagents used for UV/VIS, FL, ELCD, and RA detection of peptides, proteins, nucleic acids, and polysaccharides.

4.5 Mass Spectrometry

The molecular mass of a biopolymer is a very important parameter for its characterization and identification. Because most biopolymers are very polar and nonvolatile molecules, the classical MS methods such as electron ionization (EI) MS or chemical ionization (CI) MS are not applicable. However, the development of new, soft ionization techniques, such as fast atom bombardment (FAB),[48] electrospray ionization (ESI),[49] and matrix-assisted laser desorption/ionization (MALDI)[50] has made MS applicable as an LC detection system for biopolymers (see **Atmospheric Pressure Ionization Mass Spectrometry**).

While FAB ionization is confined to small biopolymers up to molecular mass (M_r) 15 000,[51] ESI enables the transfer of proteins larger than M_r 130 000[52] and of nucleic acids larger than M_r 7 000 000[53] into the gas phase for mass analysis (Table 5). Moreover, electrospray ionization mass spectrometry (ESIMS) greatly benefits from the multiple charging of biopolymers resulting in mass-to-charge ratios well within the mass range of most commercial mass spectrometers ($\leq$4000–6000 u) even for very large biomolecules. By virtue of its versatility for ionization of various biopolymers, broad mass range, high mass precision (typically 0.01%), and tolerance for high eluent flow rates, the ESI interface is the most commonly used to couple LC and MS. MALDI/MS is frequently applied to the off-line analysis of collected fractions after chromatographic separation, although prototypes of continuous-flow MALDI interfaces have been reported.[54]

4.6 Light-scattering Detection

A general property of macromolecular species is that they scatter light. The light intensity scattered by a biopolymer in solution at a certain angle of incidence depends on the molecular mass and concentration of the solute in the detector cell. LALLS detectors are capable of providing unique information about the molecular mass, radius of gyration, and diffusion coefficient of biopolymers in solution. To evaluate the molecular mass term, the LALLS detector is coupled in series to a concentration-dependent detector, such as the refractive index detector, and the response from both detectors is used to calculate the molecular mass without the need for external calibration standards.[55] LALLS detection is very useful for the characterization of biopolymer aggregates formed in solution,[56] information that cannot be accessed by MS.

5 INSTRUMENTAL REQUIREMENTS FOR CHROMATOGRAPHIC BIOPOLYMER SEPARATIONS

In principle, any liquid chromatograph can be used for biopolymer separations. All that is needed is a mobile phase delivery system, a sample injector, a separation column, a column thermostat, a detector (most commonly a UV absorbance detector), and a data system. Because the retentive properties of the individual components in a mixture of biopolymers can be very different, gradient elution capability is needed in most of the separation problems. The leaking of metal ions from the stainless-steel construction of the various parts of a conventional chromatograph may be a source of problems because many biopolymers interact with metal ions, resulting in irreversible adsorption or loss of biological activity of the analytes. Moreover, stainless steel rapidly corrodes with eluents containing halide salts such as those applied in IEC. To prevent this occurrence, all wetted parts of the so-called ''biocompatible'' liquid chromatographs are fabricated from inert materials such as titanium, glass, poly(tetrafluoroethylene) (PTFE), or poly(ether ether ketone) (PEEK).

Table 5 MS detection methods for biopolymers

Ionization technique	Mass range for proteins	Mass range for nucleic acids	Flow rate range ($\mu L\,min^{-1}$)
ESI	Up to M_r 150 000	Up to M_r 7 000 000	0.01–1000
CFFAB[a]	Up to M_r 15 000	Up to M_r 5000	0.5–10
MALDI	Up to M_r 400 000	Up to M_r 800 000	Off-line detection

[a] CFFAB, continuous-flow fast atom bombardment.

6 SEPARATION OF PEPTIDES AND PROTEINS

The distinction between a peptide and a protein is somewhat arbitrary, with peptides usually being defined as molecules containing 50 amino acids or less. Molecules with >50 amino acids usually have a stable three-dimensional structure in aqueous solution and are referred to as proteins. Secondary structures such as α-helix or β-sheet are generally absent for small peptides up to approximately 15 residues. However, the potential for defined secondary, tertiary, and quaternary structure increases with increasing chain length. Moreover, folding to internalize hydrophobic residues to stabilize the protein structure in aqueous solution is likely to become a significant conformational feature. This potential to form a variety of structures results in a large diversity of naturally occurring proteins and poses great challenges to separation techniques in terms of selectivity and resolution.

6.1 Size-exclusion Chromatography

SEC is frequently used as a first step in a purification scheme to separate the molecules of interest from those which have radically different size characteristics. It can also be used to change the buffer environment or for the desalting of a biopolymer sample. Finally, SEC can be implemented as a means of determining molecular parameters such as average molecular mass and hydrodynamic radius. Figure 8(a) illustrates a typical separation of proteins by SEC. The stationary phase used for this separation is SynChropak GPC 300, a controlled-porosity silica support having an average pore diameter of 30 nm with a glycerylpropylsilane bonded phase to prevent adsorptive protein retention.[57] High molecular mass calf thymus DNA, which is completely excluded from the pores, is used as a marker for the void volume (V_0) whereas the small dipeptide glycyltyrosine can penetrate the total pore volume and serves as a noninteractive marker for the total volume (V_t). Solutes with a molecular size between the two extremes are eluted at a volume (V_e) which is between V_0 and V_t. The distribution coefficient K_d specifies the fraction of V_t available to a molecule according to the relationship $V_e = V_0 + K_d V_t$. Under ideal size-exclusion conditions, the logarithm of the molecular mass is a linear function of the distribution coefficient and this relation can be used to define a calibration graph for protein molecular masses as a function of the elution volume (Figure 8b).

6.2 Ion-exchange Chromatography

Although RPC is the most commonly used chromatographic mode for the separation of peptides, IEC

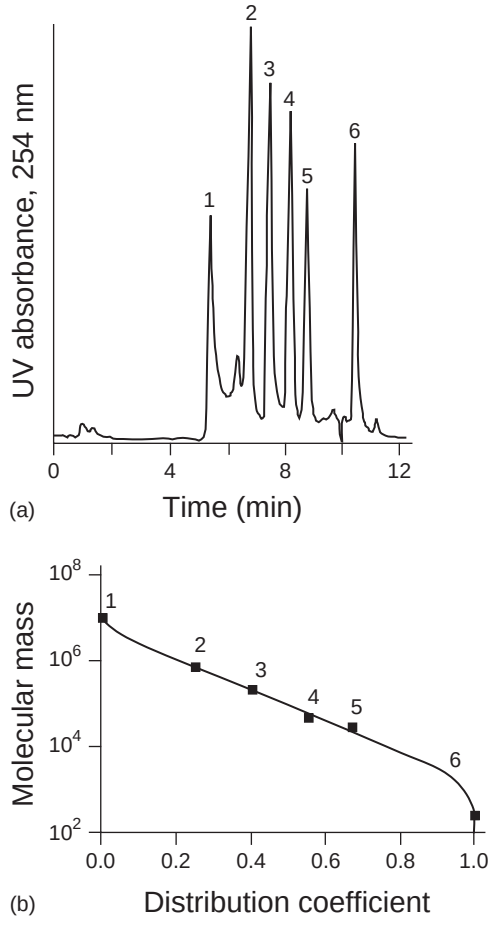

Figure 8 (a) SEC separation of proteins and (b) calibration graph for molecular mass determination of proteins. Column, SynChropak GPC 300 from Synchrom, 5 µm particle size, 30 nm pore size, 300 × 7.8 mm i.d.; mobile phase, 0.1 M potassium phosphate, pH 7.0; flow rate, 1 mL min^{-1}. Peak identification: 1, calf thymus DNA; 2, thyroglobulin; 3, β-amylase, 4, ovalbumin; 5, carbonic anhydrase; 6, glycyltyrosine. (Reprinted with permission from C.T. Mant, R.S. Hodges, *High-performance Liquid Chromatography of Peptides and Proteins*, CRC Press, Boca Raton, FL, 1991, Copyright Lewis Publishers, an imprint of CRC Press, Boca Raton, FL.)

represents the most promising alternative if RPC fails to separate the components of interest. At a pH near neutrality, anion-exchange chromatography is used for acidic peptides[58] and cation-exchange chromatography for basic peptides.[59] However, cation-exchange chromatography becomes applicable to peptides in general if a pH in the range 2.8–3.0 is employed, where peptide retention will be governed by the basic residues of lysine, arginine, and histidine, and the N-terminus. Figure 9 demonstrates that peptides are eluted with a gradient of increasing salt concentration in order of increasing absolute number of positively charged residues. The ability of anion-exchange chromatography to resolve peptides of the same net charge is attributable to some sort of

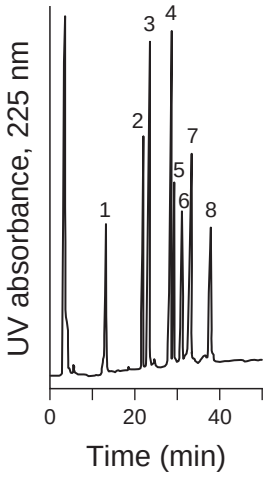

Figure 9 Separation of peptide standards by cation-exchange chromatography. Column, poly(2-sulfoethylaspartamide)silica, 5 μm particle size, 30 nm pore size, 200 × 4.6 mm i.d.; mobile phase, (A) 5 mM potassium phosphate, pH 3.0–25% acetonitrile, (B) 0.25 M potassium chloride in (A); linear gradient, 0–100% B in 40 min; flow rate, 0.7 mL min^{-1}. Peak identification: 1, oxytocin (1+); 2, [Arg^8]-vasopressin (2+); 3, somatostatin (3+), 4, substance P, free acid (3+); 5, substance P (3+); 6, bovine pancreatic polypeptide (5+), 7, anglerfish peptide (6+); 8, human neuropeptide Y (7+) (numbers in parentheses are net charge of the peptides at pH 3.0). (Reprinted from A.J. Alpert, P.C. Andrews, 'Cation-exchange Chromatography of Peptides on Poly(2-sulfoethylaspartamide) Silica', *J. Chromatogr.*, **443**, 85–96 (1988), Copyright 1988, with permission from Elsevier Science.)

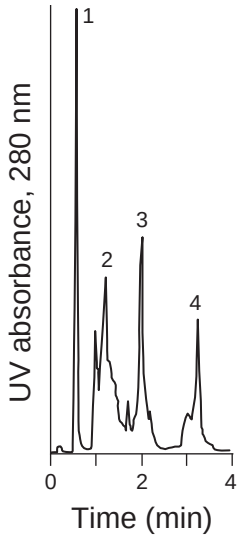

Figure 10 High-performance anion-exchange separation of proteins. Column, POROS Q/H, PS/DVB-based strong anion exchanger with bimodal pore size distribution from Perseptive Biosystems, 10 μm particle size, 50 × 4.6 mm i.d.; mobile phase, (A) 10 mM tris(hydroxymethyl)aminomethane (TRIS)–HCl, pH 8.0, (B) 0.35 M sodium chloride in (A); linear gradient, 0–100% B in 5 min; flow rate, 4 mL min^{-1}. Peak identification: 1, myoglobin; 2, conalbumin; 3, ovalbumin; 4, soybean trypsin inhibitor. (Reprinted from N.B. Afeyan, N.F. Gordon, I. Mazsaroff, L. Varady, S.P. Fulton, Y.B. Yang, F.E. Regnier, 'Flow-through Particles for the High-performance Liquid Chromatographic Separation of Biomolecules: Perfusion Chromatography', *J. Chromatogr.*, **519**, 1–29 (1990), Copyright 1990, with permission from Elsevier Science.)

mixed-mode interaction and/or to differences in the secondary structure of the peptides.

IEC is the method of choice for high-resolution separation of proteins in their native form without disrupting the three-dimensional structure. Figure 10 depicts the resolution of four standard proteins by anion-exchange chromatography on a stationary phase of the bimodal pore-size configuration (compare Figure 6c). A gradient of 0–0.35 M sodium chloride is applied to elute the proteins rapidly. The splitting of three of the four protein peaks into several incompletely resolved peaks and shoulders indicates the ability of IEC to distinguish between different isoforms of proteins, i.e. proteins of the same biological function but slightly different structure. The heterogeneity of ovalbumin (peak 3 in Figure 10), for instance, an M_r 43 000 glycophosphoroprotein consisting of 385 amino acids, is caused by oligosaccharides linked to asparagine 293 and two phosphorylation sites at serine 68 and serine 344.

The property of a protein which governs its retention in IEC is the net surface charge, which is highly pH related because of the presence of both weakly acidic and basic functional groups in proteins. The dependence of net charge of proteins on pH is a highly individual property of a protein and constitutes the basis for selectivity in IEC. At the isoelectric point, the net charge of a protein is zero and there should be no retention in IEC. However, experimental retention maps as a function of pH show that many proteins are retained on ion-exchange columns even when their net charge is zero (Figure 11), suggesting that the retention of proteins on ion-exchange stationary phases cannot be satisfactorily explained by the net charge concept alone. Deviations from this concept are mainly due to intramolecular charge asymmetry. Thus, regions of localized charges capable of interacting with oppositely charged functional groups on the surface of the stationary phase are present even at the isoelectric point.[60] In addition to the charge characteristics of the protein itself, the nature of the support surface and the displacing ion in the mobile phase are also important in determining the selectivity of an IEC separation system. The use of different displacing salts may result in retention variations of up to 100% and represents an easy and fast means of altering the selectivity of a chromatographic separation, as exemplified by the exchange of chloride by citrate as the displacing ion in the separation of soybean trypsin inhibitor and ovalbumin on a strong anion-exchange

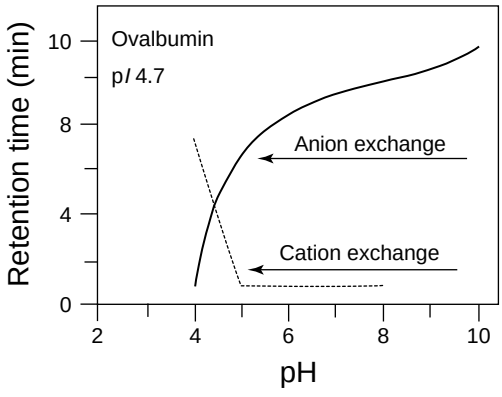

Figure 11 Experimental retention map for ovalbumin on a strong cation-exchange (Mono S from Pharmacia) and a strong anion-exchange (Mono Q from Pharmacia) column. Ovalbumin was eluted at 1 mL min^{-1} with a gradient of 0–0.5 M sodium chloride in 20 min. (Reprinted from W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, 'Retention Model for High-performance Ion-exchange Chromatography', *J. Chromatogr.*, **266**, 3–21 (1983), Copyright 1983, with permission from Elsevier Science.)

column, which entails an increase in the separation factor from 1.50 to 2.61.[60]

6.3 Reversed-phase and Hydrophobic Interaction Chromatography

The separation of peptide mixtures, particularly from proteolytic digests, is an important part of protein characterization. Peptide mapping by HPLC is routinely performed in the study of microheterogeneity, mutational variants, and isomers present in proteins. In the RPC of peptides and proteins, acids such as trifluoroacetic acid, phosphoric acid, hydrochloric acid, formic acid, and acetic acid are frequently used as mobile phase additives for the following reasons: (1) when silica-based stationary phases are employed, protonation of residual silanol groups on the support surface at low pH diminishes unwanted ionic interactions with the basic amino acid residues; (2) small organic acids form ion pairs with these basic sites and consequently increase the hydrophobic character of a peptide or protein; and (3) acidic solvents assist in the process of unfolding and denaturing the three-dimensional structure of a polypeptide, resulting in more homogeneous species that elute as sharper and more symmetric peaks. Trifluoroacetic acid is the most commonly used acid because it is volatile (unlike phosphoric acid), it does not corrode stainless-steel parts of the HPLC apparatus (unlike hydrochloric acid), and it is UV-transparent (unlike acetic and formic acid). The chromatogram of a tryptic digest of recombinant tissue plasminogen activator, a protein of 527 amino acids, on an octadecylated silica stationary phase using trifluoroacetic acid as mobile phase additive is shown in Figure 12.

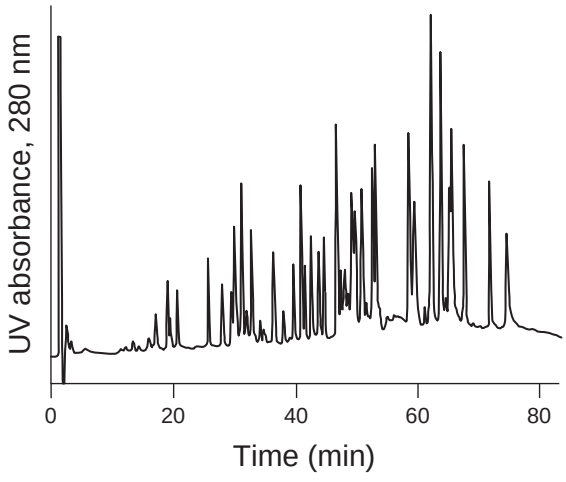

Figure 12 Reversed-phase separation of a tryptic digest of recombinant tissue plasminogen activator. Column, Nova-Pak C_{18} column from Waters, 5 µm particle size, 6 nm pore size, 150 × 3.9 mm i.d.; mobile phase, (A) 0.1% trifluoroacetic acid in water, (B) 0.08% trifluoroacetic acid in acetonitrile; linear gradient, 0–25% B in 50 min followed by 25–60% B in 35 min; flow rate, 1 mL min^{-1}; sample, tryptic map of reduced and carboxymethylated recombinant tissue plasminogen activator. (Reprinted from R.C. Chloupek, R.J. Harris, C.K. Leonhard, R.G. Keck, B.A. Keyt, M.W. Spellman, A.J.S. Jones, W.S. Hancock, 'Study of the Primary Structure of Recombinant Tissue Plasminogen Activator by Reversed-phase High-performance Liquid Chromatographic Tryptic Mapping', *J. Chromatogr.*, **463**, 375–396 (1989), Copyright 1989, with permission from Elsevier Science.)

The high selectivity and efficiency of the RPC of proteins is largely a consequence of the denaturing conditions applied (hydrophobic stationary phase, high concentration of organic solvent, low pH, elevated temperature). The potential efficiency of a 53 × 4 mm i.d. column packed with nonporous octadecylated PS/DVB is illustrated in Figure 13 by the high-resolution separation of 15 standard proteins in 13 min. The high selectivity of the separation system is reflected in the separation of the two β-lactoglobulins A and B, which differ by the substitution of only two amino acids in a total sequence of 162. β-Lactoglobulin B has a glycine at position 64 and an alanine at position 118, whereas β-lactoglobulin A contains aspartic acid and valine at the two respective positions.

The preparative purification of proteins by RPC suffers from poor recovery in terms of both total mass and biological activity. These observations are due to their strong interactions with the hydrophobic stationary phase and the strongly denaturing mobile phase conditions. HIC is an alternative chromatographic mode based on hydrophobic interactions, whereby the proteins are eluted from a weakly hydrophobic stationary phase by a decreasing salt gradient. Because the binding of the proteins to the stationary phase is weak compared with

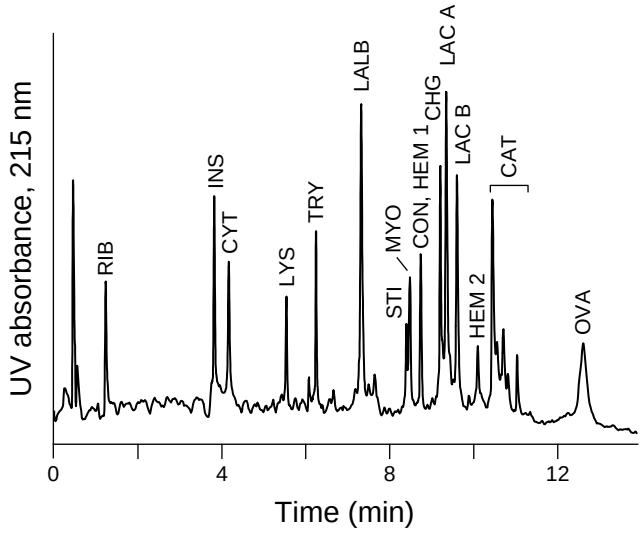

Figure 13 High-resolution RPC of proteins. Column, octadecylated PS/DVB from Transgenomic, 2.3 µm particle size, nonporous, 53 × 4 mm i.d.; mobile phase, (A) 0.1% trifluoroacetic acid–15% acetonitrile in water, (B) 0.1% trifluoroacetic acid–60% acetonitrile in water; linear gradient, 20–85% B in 15 min; flow rate, 0.56 mL min^{-1}; temperature, 80 °C. Peak identification: RIB, ribonuclease A; INS, insulin; CYT, cytochrome *c*; LYS, lysozyme; TRY, trypsin; LALB, α-lactalbumin; STI, trypsin inhibitor; MYO, myoglobin; CON, conalbumin; HEM, hemoglobin; CHG, chymotrypsinogen A; LAC A, β-lactoglobulin A; LAC B, β-lactoglobulin B; CAT, catalase; OVA, ovalbumin. (Reproduced from C.G. Huber, G. Kleindienst, G.K. Bonn, 'Application of Micropellicular Polystyrene/Divinylbenzene Stationary Phases for High-performance Liquid Chromatography Electrospray Mass Spectrometry of Proteins and Peptides', *Chromatographia*, **44**, 438–448 (1997), by permission of the authors.)

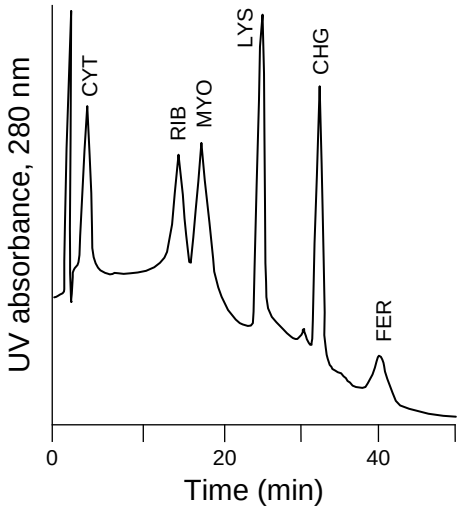

Figure 14 Separation of proteins by HIC. Column, polyethylene glycol 1000 covalently bound to 1,2-dihydroxy-n-propyl-silica, 5 µm particle size, 30 nm pore size, 100 × 4.6 mm i.d.; mobile phase, (A) 2.0 M ammonium sulfate in 0.1 M phosphate, pH 7.0, (B) 0.1 M phosphate, pH 7.0; linear gradient, 0–100% B in 40 min; flow rate, 1 mL min^{-1}; temperature, 25 °C. Peak identification: CYT, cytochrome *c*; RIB, ribonuclease A; MYO, myoglobin; LYS, lysozyme; CHG, chymotrypsinogen A; FER, ferritin. (Reproduced from Z. El Rassi, C. Horváth, 'Hydrophobic Interaction Chromatography of t-RNAs and Proteins', *J. Liq. Chromatogr.*, **9**, 3245–3268 (1986), by permission of Marcel Dekker Inc.)

RPC, and because organic solvents are generally not used, they may be eluted without conformational changes and retain their biological activity. The main variables controlling retention in HIC are hydrophobicity of the stationary phase, stationary phase ligand density and chain length, salt identity and concentration, gradient time and flow rate, pH, and temperature.[61] Figure 14 illustrates the HIC of proteins on a hydrophilic derivatized silica stationary phase. The proteins are adsorbed at the stationary phase with 2 M ammonium sulfate and elute as the salt concentration is gradually decreased.

6.4 Metal Interaction and Affinity Chromatography

Transition metals are known to form complexes with electron-rich compounds such as the amino acids cysteine, histidine, tryptophan, lysine, and tyrosine. The different number of such amino acids and the differences in their steric accessibility in proteins can be used for their chromatographic separation. The transition metal is usually immobilized on the surface of a chromatographic support by means of a chelating functionality such as iminodiacetic acid (Figure 15a).[62] Without metal, the iminodiacetic acid bonded stationary phase is expected to function as a cation exchanger. A typical chromatogram of a protein mixture separated under ion-exchange conditions at pH 5.0 is shown in Figure 15(b). However, the retention behavior of proteins is significantly altered upon chelation of Ni(II) (Figure 15c). It can be seen that the retention of most proteins increased and the selectivity changed considerably with the immobilized metal under otherwise identical conditions. In MIC, analytes are eluted with a gradient of increasing salt concentration, where the anion can be regarded as a competitive ligand for the coordination sites of the immobilized metal. The retention of proteins in MIC depends primarily on the number of ligands at the protein surface, the type and amount of immobilized metal, the strength of the protein–metal interaction, pH, and the type and concentration of displacing ion.

AC is the chromatographic method of highest specificity for proteins.[63] The separation system is usually designed such that only one target molecule is recognized in very complex mixtures by taking advantage of highly specific biochemical interactions. Therefore, chromatographic efficiency is normally not a big issue in AC. Since

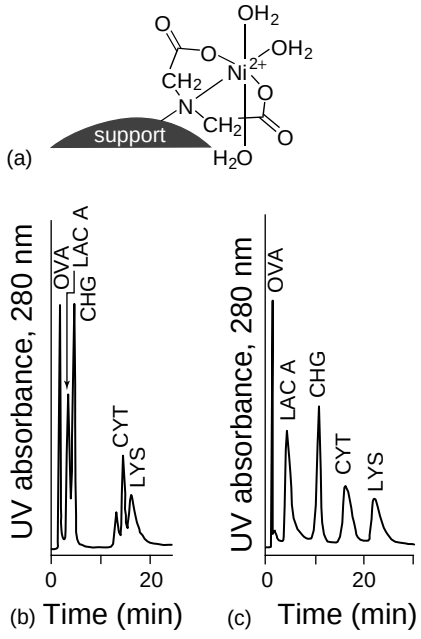

Figure 15 (a) Schematic illustration of the functional group in MIC and separation of proteins by (b) cation-exchange chromatography on iminodiacetate silica, and (c) MIC on Ni(II) iminodiacetate silica. Columns, unloaded (b) and Ni(II) loaded (c) iminodiacetate silica, 5 µm particle size, 30 nm pore size, 100 × 4.6 mm i.d.; mobile phase, (A) 25 mM phosphate, pH 5.0, (B) 0.5 M sodium chloride in (A); linear gradient, 16–100% B in 40 min; flow rate, 1 mL min^{-1}; temperature, 25 °C. Peak identification: OVA, ovalbumin; LAC A, β-lactoglobulin A; CHG, chymotrypsinogen A; CYT, cytochrome *c*; LYS, lysozyme. (Reprinted from Z. El Rassi, C. Horváth, 'Metal Chelate-interaction Chromatography of Proteins with Iminodiacetic Acid-bonded Stationary Phases on Silica Support', *J. Chromatogr.*, **359**, 241–253 (1986), Copyright 1986, with permission from Elsevier Science.)

the structure of the target molecule varies from application to application, commercially available AC media are usually preactivated supports on to which the specific ligand (antibody, receptor, antigen) is immobilized in the user's laboratory.

Figure 16(a) and (b) is a schematic illustration of the basic procedures involved in achieving a separation of proteins by AC. A sample which contains both the target molecule and impurities is loaded on the column. Adsorption of the target molecule occurs upon binding to the affinity ligand, while the impurities are washed through the column (Figure 16a). The target molecules are desorbed from the affinity stationary phase by one of the following methods (Figure 16b): (1) changing the pH, which induces the dissociation of the complex through protonation/deprotonation of basic/acidic sites involved in the biospecific interaction and/or through pH-induced conformational changes; (2) increasing the salt concentration, which weakens the electrostatic interactions between the target molecule and the affinity ligand; or (3) addition

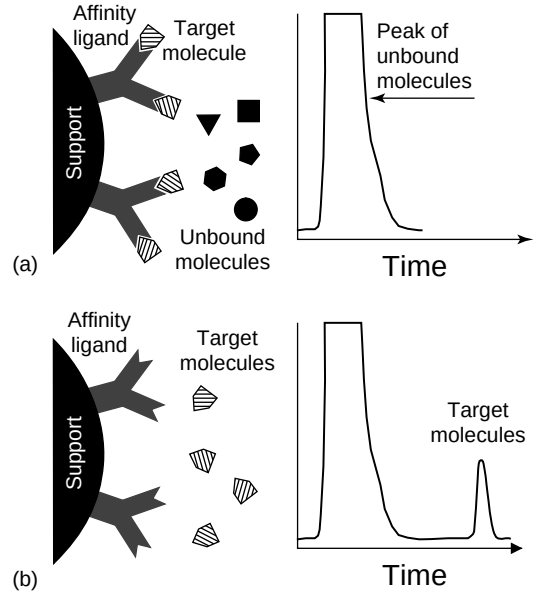

Figure 16 Schematic illustration of AC. (a) The target molecule is adsorbed biospecifically on the column, whereas all other molecules pass the column without retention. (b) The target molecule is released from the column upon dissociation of the affinity ligand–target molecule complex.

of a competitive ligand which strongly binds either to the target molecule or the affinity ligand.

7 SEPARATION OF NUCLEIC ACIDS

Nucleic acid structures can be divided into four different types, namely single-stranded oligonucleotides (10–100 nucleotides in length), double-stranded linear DNA (10 to more than 10 000 base pairs in length), supercoiled DNA (several thousand base pairs), and single-stranded RNA. Oligonucleotides are relatively small molecules where both the hydrophilic sugar–phosphate backbone and the hydrophobic nucleobases are exposed to the solution. In double-stranded DNA, a double-helical structure secludes the hydrophobic nucleobases inside the double helix, whereas the two helical sugar–phosphate chains spiral down the outside of the double-stranded structure, presenting a polyanionic and hydrophilic surface to the solvent. Supercoiled DNA is a special form of double-stranded DNA where closed circles of double-stranded DNA are wound around each other. Most single-stranded RNAs may be considered as having a globular structure with complementary sequences forming hairpin structures.

7.1 Size-exclusion Chromatography

Although SEC generally affords longer analysis times and lower resolution of oligonucleotides and DNA fragments

than IEC, mixed-mode chromatography, or IPRPC, it offers some advantages such as simple instrumentation, isocratic elution, great freedom of buffer choice, and elution in strict order of size. Resolution of mixtures of DNA fragments is achieved by selecting a stationary phase of a suitable pore-size distribution for the discrimination of molecular masses. The linear portion of a calibration graph generated by a polymer-based stationary phase containing >100-nm pores indicates (Figure 17) that this stationary phase is suitable for separating DNA fragments ranging in size from 100 to 7000 base pairs (one base pair has a mass of approximately 660).

7.2 Anion-exchange Chromatography

To fractionate polyanions such as nucleic acids, anion-exchange chromatography is the most obvious method. The negatively charged nucleic acids interact with the positively charged functional groups of the anion-exchange stationary phase and are desorbed by a gradient of increasing salt concentration. Figure 18 shows a separation of a hydrolysate of polydeoxyadenylic acid containing oligomers from the 17-mer through approximately the 100-mer on a nonporous, polymer based anion exchanger. Baseline separation is obtained for up to about the 32-mer and peaks appear for up to the 70-mer, the whole chromatogram taking only 20 min. Although the separation of homooligomers is strictly according to chain

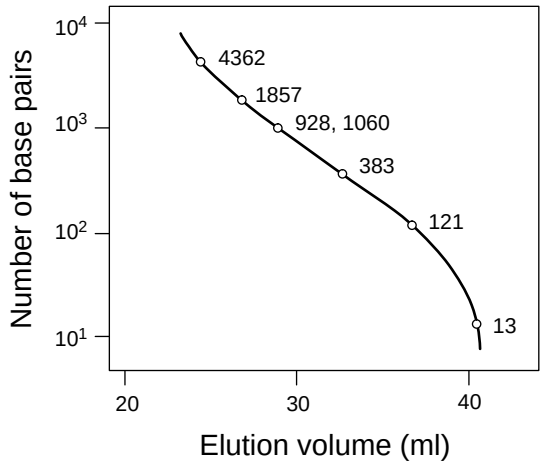

Figure 17 Relationship between elution volume and molecular mass in SEC of DNA fragments. Four TSKgel DNA-PW columns from TosoHaas connected in series, 10 µm particle size, >100 nm pore size, 300 × 7.8 mm i.d.; mobile phase, 0.1 M TRIS–HCl–0.3 M sodium chloride, pH 7.5; flow rate, 0.3 mL min^{-1}; sample, *Eco*RI-cleaved pBR322 DNA and *Bst*NI-cleaved pBR322 DNA. (Reprinted from Y. Kato, Y. Yamasaki, T. Hashimoto, T. Murotsu, S. Fukushige, K. Matsubara, 'Separation of Large DNA Restriction Fragments by High-performance Gel Filtration on TSKgel DNA-PW', *J. Chromatogr.*, **320**, 440–444 (1985), Copyright 1985, with permission from Elsevier Science.)

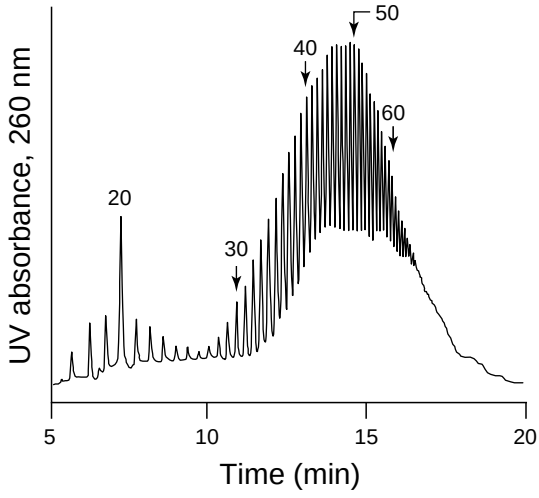

Figure 18 Anion-exchange separation of oligodeoxyadenylic acids. Two TSKgel DEAE NPR (diethylaminoethyl-derivatized hydrophilic polymer from TosoHaas) columns connected in series, 2.5 µm particle size, nonporous, 35 × 4.6 mm i.d.; mobile phase, (A) 20 mM TRIS–HCl, pH 9.0, (B) 1 M sodium chloride in (A); linear gradient, 25–100% B in 60 min; flow rate, 1 mL min^{-1}; temperature, 25 °C. Peaks are identified by the number of adenosine units in the oligomer. (Reprinted from Y. Kato, T. Kitamura, A. Mitsui, Y. Yamasaki, T. Hashimoto, T. Murotsu, S. Fukushige, K. Masubara, 'Separation of Oligonucleotides by High-performance Ion-exchange Chromatography on a Nonporous Ion Exchanger', *J. Chromatogr.*, **447**, 212–220 (1988), Copyright 1988, with permission from Elsevier Science.)

length, the base sequence of heterooligonucleotides has a significant impact in retention, most probably due to secondary interactions with the stationary phase and conformational influences. On a TSKgel DEAE NPR stationary phase, the retention of 12–18-mer oligonucleotides containing different bases increased in the order adenosine < thymidine < cytidine, whereas on a nonporous, quaternized PS/DVB anion exchanger the elution order was cytidine < adenosine < thymidine.[64]

The same stationary phase as used for the high-resolution separation of oligonucleotides can also be applied to the separation of double-stranded DNA fragments, as illustrated in Figure 19 by the fast separation of 22 DNA fragments of a 1-kilobase-pair ladder with fragments ranging in size from 75 to 12 216 base pairs. This separation demonstrates that IEC can separate DNA fragments with high resolution over an extremely broad size range in a fraction of the time that is needed for a size-exclusion separation (10 versus 130 min, compare Figure 17). The separation principle of anion-exchange chromatography implies that elution order is determined only by size because with double-stranded DNA, the hydrophobicity of the nucleobases is expected to have a minor influence on retention because they are secluded inside the double

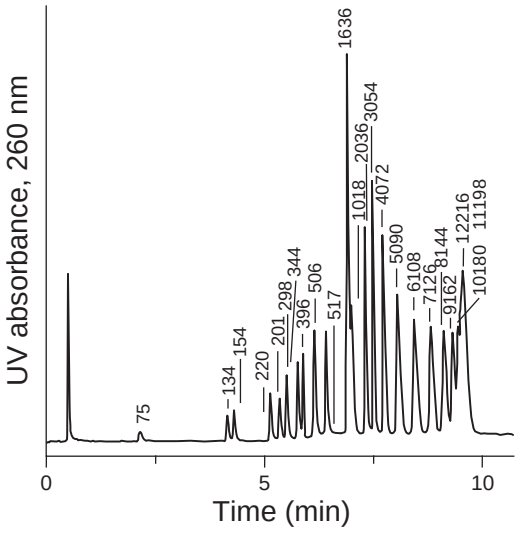

Figure 19 Anion-exchange separation of DNA fragments. Column, TSKgel DEAE NPR (diethylaminoethyl-derivatized hydrophilic polymer from TosoHaas), 2.5 μm particle size, nonporous, 75 × 4.6 mm i.d.; mobile phase, (A) 20 mM TRIS–HCl, pH 9.0, (B) 1 M sodium chloride in (A); linear gradient, 50–75% B in 15 min; flow rate, 1 mL min^{-1}; temperature, 25 °C. Peaks are identified by the number of base pairs in the DNA fragment. (Reprinted from C.G. Huber, 'Micropellicular Stationary Phases for High-performance Liquid Chromatography of DsDNA', *J. Chromatogr. A*, **806**, 3–30 (1998), Copyright 1998, with permission from Elsevier Science.)

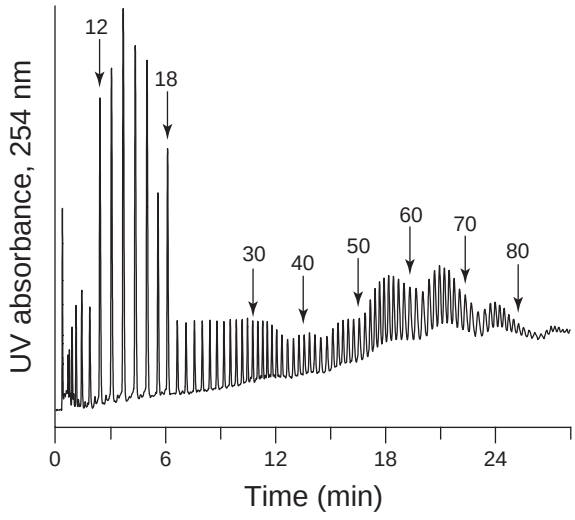

Figure 20 IPRPC separation of oligodeoxyadenylic acids. Column, octadecylated PS/DVB from Transgenomic, 2.3 μm particle size, nonporous, 50 × 4.6 mm i.d.; mobile phase, (A) 0.1 M triethylammonium acetate, pH 7.0, (B) 25% acetonitrile in (A); linear gradient, 22–33% B in 10 min, 33–36% B in 6 min, 36–45% B in 9 min; flow rate, 1 mL min^{-1}; temperature, 50 °C. Peaks are identified by the number of adenosine nucleotides in the oligomer. (Reproduced from C.G. Huber, E. Stimpfl, P.J. Oefner, G.K. Bonn, 'A Comparison of Micropellicular Ion-exchange and Reversed-phase Stationary Phases for HPLC of Oligonucleotides', *LC-GC*, **14**, 114–127 (1996), by permission of Advanstar Communications.)

helix. However, anion-exchange separation of double-stranded DNA fragments on a variety of different stationary phases has suffered from inversions in retention time as a function of chain length, preventing its use for size-accurate fragment identification (compare, e.g. the elution order of the 220 and 201 base-pair fragments or the 1636 and 1018 base-pair fragments in Figure 19). A more length-relevant separation of DNA fragments at the cost of lower resolution is feasible when a gradient of tetraethylammonium chloride is applied instead of sodium chloride.[65]

7.3 Ion-pair Reversed-phase Chromatography

Retention of nucleic acids is effected predominantly by electrostatic interactions both in anion-exchange chromatography and in IPRPC. However, the two chromatographic modes differ in the way in which the surface potential is generated at the stationary phase. Whereas anion exchangers have the charged functional groups covalently bound to the surface, the amphiphilic ions generating the surface potential in IPRPC are adsorbed through hydrophobic interactions on the surface of a reversed-phase stationary phase. The surface concentration of the amphiphile strongly depends on the concentration of organic modifier in the eluent, which offers a simple means of controlling the retention of the analytes in IPRPC. The most popular amphiphiles for the IPRPC of nucleic acids are trialkylammonium and tetraalkylammonium ions, where the alkyl chains vary from methyl to butyl. Common gradient formers are acetonitrile, methanol, and 2-propanol.

Figure 20 depicts the separation of a homologous series of oligodeoxyadenylic acids by IPRPC on a nonporous, octadecylated PS/DVB stationary phase. With a total separation time of 28 min, baseline separation is obtained up to about the 50-mer and peaks appear up to the 90-mer. Compared with anion-exchange chromatography, the influence of base composition on the retention behavior of oligonucleotides is more pronounced in IPRPC owing to additional hydrophobic interactions of the nucleobases with the hydrophobic stationary phase, the elution order of 4-mer homooligonucleotides being tetradeoxyguanosine < tetradeoxycytidine < tetradeoxyadenosine < tetradeoxythymidine.[64]

IPRPC is also suitable for the separation of double-stranded DNA fragments. Figure 21 shows the chromatogram of a mixture of four different restriction digests of the plasmids pBR322 and pBR328 containing 45 DNA fragments. Of these, 39 fragments are of different size and are separated into 33 at least partially resolved peaks. Compared with anion-exchange chromatography, IPRPC can separate DNA fragments up to a chain length of 2000

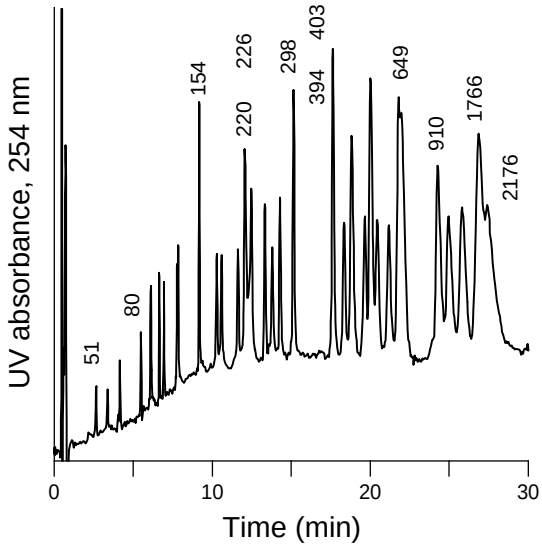

Figure 21 High-resolution IPRPC separation of DNA fragments. Column, octadecylated PS/DVB from Transgenomic, 2.3 µm particle size, nonporous, 50 × 4.6 mm i.d.; mobile phase, (A) 0.1 M triethylammonium acetate, pH 7.0, (B) 25% acetonitrile in (A); linear gradient, 38–54% B in 10 min, 33–36% B in 7 min, 54–62% B in 8 min, 62–71% B in 15 min; flow rate, 0.75 mL min^{-1}; temperature, 50 °C; sample, *Hae*III-cleaved pBR322 DNA, *Alu*I-cleaved pBR322 DNA, *Bgl*I-cleaved pBR328 DNA, and *Hinf*I-cleaved pBR328 DNA. Peaks are identified by the number of base pairs in the DNA fragment.

base pairs with higher resolution, whereas the former can better separate long DNA fragments (>2000 base pairs). In IPRPC, the retention of double-stranded nucleic acids is size dependent to a very high degree, which allows the calculation of fragment size from retention data with an accuracy of better than 3.2%.[66] Moreover, the correlation between size and retention time is almost linear up to 500 base pairs and then slowly levels off up to 2000 base pairs.

One important additional feature of IPRPC is the sequence-dependent retention of double-stranded DNA fragments at temperatures higher than 55 °C due to partial denaturation.[67] The temperature at which the separation of the two DNA strands begins is determined by the total number and extent of repetitive AT tracts in a DNA fragment and thus depends on the DNA sequence. The perturbation of the helical structure in partially denatured DNA fragments as a consequence of opening of the double strand results in decreased retention compared with completely double helical fragments of the same length because of a smaller contact area with the stationary phase.

7.4 Mixed-mode Chromatography

With mixed-mode chromatography, the separation of the analytes relies upon at least two different modes of interaction between the solute and the stationary phase. Although with many types of stationary phases the separation can be regarded as the result of mixed-mode processes, generally the additional modes of interaction are a minor effect and can be minimized through the appropriate choice of mobile phase parameters. In order to provide solute retention by mixed interactions, it is necessary that the stationary phase be multifunctional in character. In the case of nucleic acids, mixed electrostatic and hydrophobic interaction are most successfully employed. The two methods for the preparation of mixed-mode stationary phases for electrostatic and hydrophobic interaction are the introduction of hydrophobic moieties on to an ion-exchange packing (e.g. through alkylation of the amino groups of aminopropylsilica) or the introduction of sites for electrostatic interaction into a reversed-phase stationary phase (e.g. through adsorption of methyltrioctylammonium chloride on octadecyl silica).[68] Mixed-mode supports have proven to be very valuable for separating tRNAs by mixed-mode chromatography. Elution of tRNAs is performed with a combined gradient of increasing ammonium acetate concentration and increasing pH.

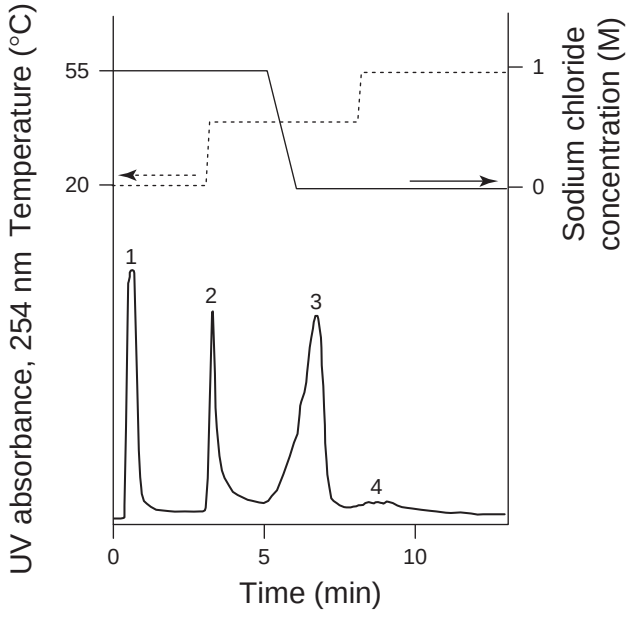

Figure 22 AC of polyadenosine mRNA from *Saccharomyces*. Column, silica gel, derivatized with $(dT)_{18}$, 7 µm particle size, 30 nm pore size, 30 × 4.6 mm i.d.; mobile phase, (A) 10 mM phosphate, pH 7.0–1 M sodium chloride, (B) 5 mM phosphate, pH 7.0; temperature and salt gradient as indicated; flow rate, 1 mL min^{-1}. Peak identification: see text. (Reprinted from T.A. Goss, M. Bard, H.W. Jarrett, 'High-performance Affinity Chromatography of Messenger RNA', *J. Chromatogr.*, **588**, 157–164 (1991), Copyright 1991, with permission from Elsevier Science.)

7.5 Affinity Chromatography

The AC of nucleic acids involves the highly selective base pairing of two complementary polynucleotide strands. By attaching single-stranded oligonucleotides to a chromatographic support, nucleic acids containing a sequence complementary to the immobilized sequence are retained by the column.[69] Figure 22 shows as an example the selective isolation of mRNA from *Saccharomyces* containing polyadenosine tracts at their 3′-ends. Using a silica stationary phase with an immobilized oligodeoxythymidine 18-mer [$(dT)_{18}$] and a combination of salt and temperature gradients, mRNA can be purified in as little as 8 min. The first, unretained peak contains rRNA and nonpolyadenylated mRNA and the second peak contains mRNA with short stretches of polyadenosine with less than 15 adenosine units. The third peak represents the bulk of the polyadenosine mRNA with tail lengths of 15–80 units. The fourth peak contains mRNA with long polyadenosine tails of about 80 units.

8 SEPARATION OF OLIGO- AND POLYSACCHARIDES

Oligo- and polysaccharides can have an enormous number of possible structures because of the various configurations possible for monosaccharides, their variable linkage forms, the possibility of α- and β- anomers at the glycoside linkage, and their different branching patterns. While the linkage of three different amino acids can result in only six different tripeptides, three different carbohydrates can theoretically form more than 700 trisaccharides. Furthermore, saccharides are diversified by the presence of nonglycosyl substituents. Oligosaccharides have covalent structures of 2–25 usually different monosaccharide residues joined by glycosidic linkages, resulting in linear, branched, or cyclic chains. Polysaccharides are polymeric forms of repeating mono-, di-, or tetrasaccharides. Naturally occurring polysaccharides may range in molecular mass from a few thousand to several million.

8.1 Size-exclusion Chromatography

SEC is the method of choice for the separation of high molecular mass polysaccharides.[32] With appropriate stationary phases of smaller pore size, it is also applicable to the separation of oligosaccharides, although possibly with lower resolution than that offered by IEC. With saccharides, having a high portion of hydroxyl and other polar groups, interaction with a polar stationary phase is likely to induce unwanted adsorption unless the ionic strength of the mobile phase is high enough (0.1–1 M) to exclude such interaction. SEC has found application in the carbohydrate field mainly as a means of fractionation and molecular mass distribution analysis of polysaccharides. Calibration is a general problem in SEC of polysaccharides because standard substances of known molecular mass are not readily available for all structural types of polysaccharides. However, advances in detection methods such as LALLS detection largely overcome this drawback and, furthermore, enable the extraction of information on polymer branching and conformation from size-exclusion data.

8.2 Normal- and Reversed-phase Chromatography

By virtue of their two different retention mechanisms, NPC and RPC are complementary separation techniques and should be used in conjunction for the isolation and analysis of oligosaccharides. A normal-phase separation of maltooligosaccharides is depicted in Figure 23. The substances are eluted in the order of their increasing molecular mass. Baseline separation up to the 17-mer is obtained on an aminopropylsilica stationary phase with a gradient of 25–75% water in acetonitrile.[70] The oligosaccharides are in the form of their 4-nitrophenylglycosides which are detected by their absorbance at 300 nm. Alternatively, the same sample can be separated by RPC on an octadecylsilica stationary phase (ODS-Hypersil from Shandon) with a gradient

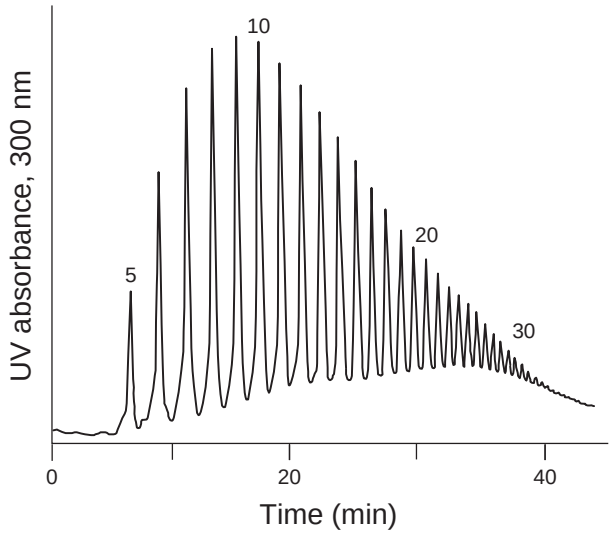

Figure 23 Separation of 4-nitrophenyl-derivatized α-D-maltooligomers by NPC. Column, Hypersil APS-2 (aminopropyl-modified silica gel) from Shandon, 5 μm particle size, 12 nm pore size, 250 × 4 mm; i.d. mobile phase, (A) acetonitrile, (B) water; linear gradient, 25–75% B in 60 min; flow rate, 1.5 mL min^{-1}. Numbers on peaks indicate number of glucose units. (Reprinted from C. Niemann, W. Saenger, R. Nuck, B. Pfannemüller, 'HPLC of 4-Nitrophenyl-α-D-malto-oligosaccharides', *Carbohydr. Res.*, **215**, 15–23 (1991), Copyright 1991, with permission from Elsevier Science.)

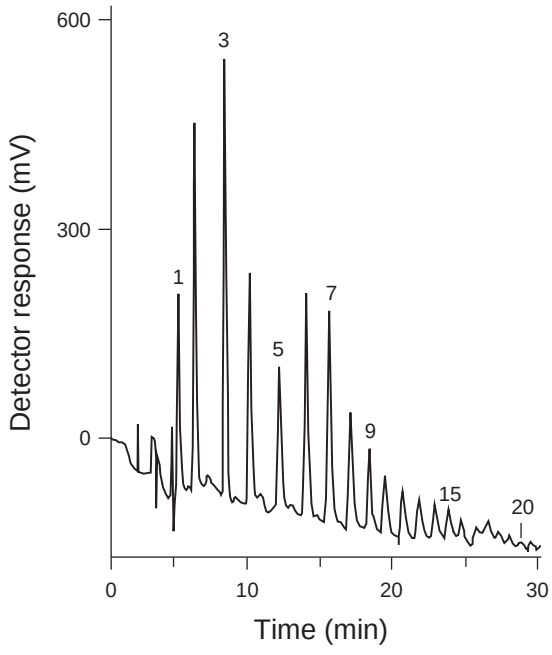

Figure 24 Separation of linear maltooligosaccharide oligomers of up to 21 glucose residues by anion-exchange chromatography with pulsed ELCD. Column, PS/DVB with trimethylammonium groups, 2.1 µm particle size, nonporous; mobile phase, (A) 50 mM sodium hydroxide, (B) 50 mM sodium hydroxide–1 M sodium acetate; linear gradient, 0–20% B in 20 min; flow rate, 0.5 mL min^{-1}; detection, PAD, $E_1 = 100$ mV ($t_1 = 300$ ms), $E_2 = 600$ mV ($t_2 = 120$ ms), $E_3 = -6$ mV ($t_3 = 300$ ms). Numbers on peaks indicate number of glucose units. (Reprinted from C. Corradini, D. Corradini, C.G. Huber, G.K. Bonn, 'Synthesis of a Polymeric-based Stationary Phase for Carbohydrate Separation by High-pH Anion-exchange Chromatography with Pulsed Amperometric Detection', *J. Chromatogr. A*, **685**, 213–220 (1994), Copyright 1994, with permission from Elsevier Science.)

from 4 to 90% acetonitrile in water with similar, but slightly lower, resolution.[70]

8.3 Ion-exchange Chromatography

Since carbohydrates are weak acids with pK_a values ranging from 12 to 14 they dissociate under highly alkaline conditions and can be separated directly at high pH by anion-exchange chromatography. Elution at high pH allows the separation of saccharides as their oxoanions. This approach cannot be used with silica-based stationary phases owing to their poor stability at high pH. Therefore, a range of polymeric, nonporous ion exchangers has been developed by Dionex. These stationary phases consist of 5–10-µm nonporous core particles on to which a layer of quaternary ammonium anion-exchange microbeads (<0.1 µm) is attached by electrostatic interactions, providing rapid mass transport, high ion-exchange capacity, and high chemical and mechanical stability. High-pH anion-exchange chromatography is uniquely able to resolve structural isomers of oligosaccharides derived from glycoproteins or complex biological samples. Detection of carbohydrates at high pH conditions poses a problem. Pre- and postcolumn derivatization methods have been largely replaced by PAD at gold electrodes, which works well under alkaline conditions. Moreover, PAD is characterized by low detection limits and relatively high specificity for compounds with hydroxyl groups. The high-pH anion-exchange separation of maltooligosaccharides on a nonporous PS/DVB stationary phase with covalently bound trimethylammonium groups is shown in Figure 24. The oligomers are eluted with a gradient of 0–0.2 M sodium acetate in 50 mM sodium hydroxide and baseline resolution of the oligomers is feasible at least up to the 20-mer.

8.4 Metal Interaction Chromatography

Separations of mono- and oligosaccharides on 5–10% cross-linked PS/DVB sulfonic acid cation exchangers in the hydrogen form are mainly due to the size-exclusion effect in the swollen polymer gels.[32] The degree of cross-linking of a resin is important because separation can only take place when the sample molecules penetrate at least partially into the matrix. For this reason, cation-exchange resins for the analysis of oligosaccharides usually have

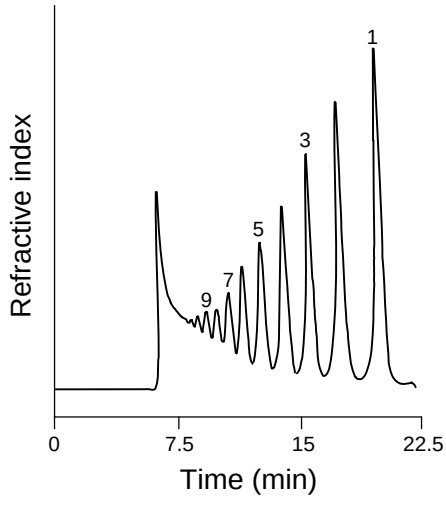

Figure 25 Separation of acid-hydrolyzed corn syrup on a silver-loaded cation-exchange resin. Column, Hamilton HC-50 5% cross-linked PS/DVB resin, 10–15 µm particle size, 71% silver form; mobile phase, water; flow rate, 0.55 mL min^{-1}; temperature, 85 °C. Numbers on peaks indicate number of glucose units. (Reprinted from H.D. Scobell, K.M. Brobst, 'Rapid High-resolution Separation of Oligosaccharides on Silver Form Cation-exchange Resins', *J. Chromatogr.*, **212**, 51–64 (1981), Copyright 1981, with permission from Elsevier Science.)

a lower degree of cross-linking than those optimized for the separation of monomers. Silver ions are known to form fairly stable monodentate complexes with most monosaccharides.[32] This complex formation also brings about greater retention of oligosaccharides on silver-loaded cation-exchange resins and allows more species to be resolved from the excluded high molecular mass polysaccharides. In the complex formation of saccharides with metal ions, the hydroxyl groups are thought to form coordinate bonds with the metal ion. The stability of the complexes depends both on the type of metal ion and on the configuration of the hydroxyl groups in the saccharides. The separation of oligosaccharides obtained by the hydrolysis of corn syrup is illustrated in Figure 25. Oligomers up to the 10-mer are separated on a 5% cross-linked cation exchanger that is loaded with 71% silver ions.

9 MODERN TRENDS AND FUTURE DEVELOPMENTS

9.1 Miniaturization of Separation Systems

One of the clear trends in science and technology is miniaturization, which is also a permanent goal in analytical chemistry, where the growing interest in analyzing minute samples of high complexity is the driving force for the rapid development of miniaturized techniques.[71] One of the most important innovations in the field of miniaturized separation has been the introduction of fused-silica capillaries, which gave birth to several miniaturized separation techniques including capillary gas chromatography (GC), capillary LC, capillary electrophoresis (CE), and capillary electrochromatography (CEC) (see **Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography**). Basically, four major advantages connected with the use of smaller dimensions in LC separation techniques can be specified: (1) better resolving power can be accomplished in a shorter time; (2) smaller samples can be analyzed; (3) on-line conjugation to MS is feasible; and (4) expenses connected with consumption and disposal of materials are decreased.

Miniaturization of column technology is usually the first step in the development of a miniaturized LC separation system. Today, four major types of microcolumns can be distinguished:[72] (1) open-tubular microcolumns; (2) partially packed capillary columns; (3) tightly packed capillary columns; and (4) continuous-bed capillary columns. Column inner diameters down to 5–20 μm have been realized with these types of capillary columns. The downscaling factor, which corresponds to the square of the ratio of the column diameters, for the reduction of the column inner diameter, e.g. from the traditional 4.6 mm to 25 μm, is 33 856, which means that all parameters related to the volume, including volumetric flow rate, injection volume, and detection volume, have to be divided by the downscaling factor. Thus, a capillary column of 25 μm i.d. is operated at a flow rate of around 30 nL min^{-1} and a few tenths of a nanoliter are injected, allowing the chromatographic analysis of attomole to zeptomole amounts of biopolymers.

However, the low flow rates and detection volumes represent considerable challenges to reliable instrumentation for miniaturized LC, and miniaturized versions of conventional instruments frequently fail to provide the required performance. Therefore, alternative principles for pumping and detecting devices, such as pumps based on thermal expansion[73] or electroosmosis,[74] and detection devices based on microelectrodes, laser light,[75] or MS[74] have been developed. One of the strongest supports in microtechnology is provided by microelectronics. Advances in manufacturing processes, namely silicon micromachining techniques, are the basis for miniaturization of liquid-phase separation techniques down to the submicrometer level. Figure 26 shows an electron micrograph of the inlet and part of the chromatographic bed of a microfabricated column which was machined by lithography on a silicon wafer. Hundreds of such column structures on a single wafer are possible based on the ease with which millions of submicrometer-sized structural features are micromachined into silicon wafers. This represents a great potential for multiplexing of miniaturized liquid-phase separation technology.

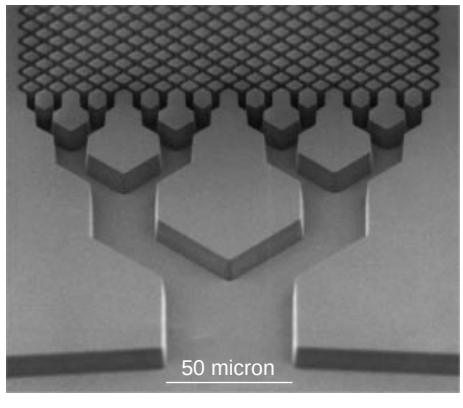

Figure 26 Scanning electron micrograph showing the column inlet and part of the stationary phase of a nanocolumn structure microfabricated on a quartz wafer. (Reprinted with permission from B. He, N. Tait, F.E. Regnier, 'Fabrication of Nanocolumns for Liquid Chromatography', *Anal. Chem.*, **70**, 3790–3797 (1998), Copyright 1998 American Chemical Society.)

9.2 Multidimensional Separation of Biopolymers and Coupling to Mass Spectrometry

Although single-stage chromatographic separation systems offer high selectivity and resolving power for biopolymers, the total peak capacity may not suffice to separate all the components of interest contained in complex mixtures of biological origin. A practical means of increasing the selectivity of a separation system is the use of two or more separation stages, resulting in the multidimensional separation of a biopolymer mixture.[76] The two criteria for multidimensional separation are that the sample components are displaced by two or more separation techniques based on different separation mechanisms and that the components separated by any single separation dimension must not be recombined in any further separation dimension. The two separation dimensions should ideally be orthogonal, i.e. separate the analyte according to different molecular properties. The key feature of a multidimensional system is that all analytes composing the initially injected sample are eventually transferred to all further separation stages.

In column LC, utilization of the total available separation selectivity would require a large number of secondary columns, so that all the sample cuts taken while eluting from the first dimension could be transferred to the second dimension. In practice, however, definite fractions of the separations taken in the first dimension are studied in the second dimension, a technique that is conventionally termed coupled-column LC. If the timescale for the first dimension is of the order of a few minutes for one eluting peak, whereas that for the second dimension is of the order of a few seconds per peak, analytes eluting from the first stage of separation can be sampled and analyzed in real time in the second stage without loss of the separation in the first dimension. The requirement for such a fast separation in the second dimension is ideally met by RPC. Therefore, the most frequently used combinations of separation dimensions for the multidimensional LC separation of protein mixtures are SEC/RPC and IEC/RPC.

MS is generally viewed as a spectroscopic technique yielding both the molecular mass and structural information from fragmentation and is used as a powerful detection method for the analytes separated by LC (see **Liquid Chromatography/Mass Spectrometry**). However, it can also be viewed as a separation technique distinguishing between different mass-to-charge ratios. Therefore, according to the definition given above, conjugation of LC and MS represents a multidimensional separation system. Two MS ionization techniques for biopolymers are considered to be most compatible with the flow rates identified with microscale liquid-phase separations: CFFAB and ESI. It is more difficult to couple MALDI directly to liquid-phase separations, because samples have to be mixed with a matrix and dried on a solid surface before insertion into the mass spectrometer. Therefore, MALDI MS is mainly used as off-line detection of fractions collected from LC.

Coupled reversed-phase HPLC with ESIMS has emerged as a very powerful tool for the identification of known protein sequences and also for the de novo sequencing of proteins. The most effective method is the combination of chemical or proteolytic digestion of a protein followed by reversed-phase separation and identification of the resulting peptide sequences by tandem MS. In tandem MS, ions of a single mass-to-charge ratio are selected for transmission through the first mass analyzer and into a cell for collision-induced dissociation, where the peptides fragment to create patterns characteristic of a specific amino acid sequence. These are subsequently detected by means of a second mass analyzer. With the help of a computer algorithm, the experimental patterns can be compared and matched with calculated patterns obtained from databases containing the sequences of all known proteins. This method allows the rapid and fully automated identification of the proteins in whole cell lysates from uninterpreted tandem mass spectra of tryptic peptides by comparison with sequences in a protein database.[77] Unknown protein sequences can be deduced from the sequence-specific pattern in the mass spectrum of the fragmented peptide.

A schematic diagram of an on-line multidimensional separation system for proteins is illustrated in Figure 27.

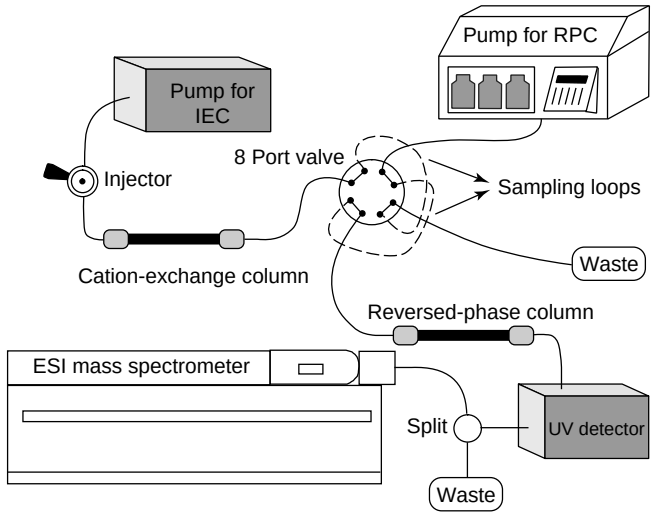

Figure 27 Schematic diagram of an on-line multidimensional separation system for proteins. (Reprinted with permission from G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, R.J. Anderegg, 'Comprehensive On-line LC/LC/MS of Proteins', *Anal. Chem.*, **69**, 1518–1524 (1997), Copyright 1997 American Chemical Society.)

The three applied separation dimensions involve cation-exchange chromatography, RPC, and MS. Sample components eluting from the cation-exchange column are sampled by an eight-port valve and subsequently injected on to a reversed-phase column. The eluate from RPC is split into a UV detector and an ESI mass spectrometer. One 120-min cation-exchange run is sampled by 48 individual 150-s duration reversed-phase runs and mass spectra from m/z 1000 to 2000 are recorded every 2 s. The peaks eluting from the cation-exchange column are approximately 6 min wide. With a time between the injection peak and the last-eluting peak of 100 min, a maximum peak capacity of 16 in the first dimension can be achieved. A peak width of about 5 s and an available elution time of 60 s give a peak capacity of 32 in the second dimension. Taking the inherent peak capacity of MS as 5, because it can identify at least this number of components in the m/z 1000–2000 mass range, the peak capacity of the entire three-dimensional system amounts to >2500, which is sufficient even for complicated biological mixtures.

9.3 Capillary Electrochromatography

CEC is an electrokinetic separation technique that is expected to combine the versatility offered by HPLC with the high plate efficiency of CE (see **Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography**).[78] The technique employs fused-silica capillaries packed with a stationary phase that has fixed charges at the surface. Thus, upon application of a high electric field, flow of eluent across the column is generated and maintained by electro-osmotic means. The diffuse part of the electrical double layer, which exists at all interfaces between a solid and an electrolyte solution, responds to the driving force of the electric field generating the flat flow profile characteristic for electroosmotically driven flow.

CEC promises much higher plate efficiencies than microcolumn HPLC with packed capillary columns. Furthermore, CEC offers a higher sample loading capacity and a wider range of selectivity than CE. The unique electro-osmotic flow field allows the use of long capillary columns packed with small particles, resulting in significant gains in peak capacity. CEC is applicable to the analysis of neutral and charged compounds. The sample components are separated on the basis of differences in their chromatographic partitioning between the mobile and the stationary phase and, if charged, also differences in their electrophoretic mobilities. The potential of CEC for the analysis of biopolymers lies in increased peak capacity owing to higher selectivity because of the two different separation mechanisms involved (chromatography and electrophoresis) and higher efficiency because of the flat electro-osmotic flow profile. Figure 28 shows the separation of standard proteins by gradient reversed-phase CEC in a monolithic capillary column.

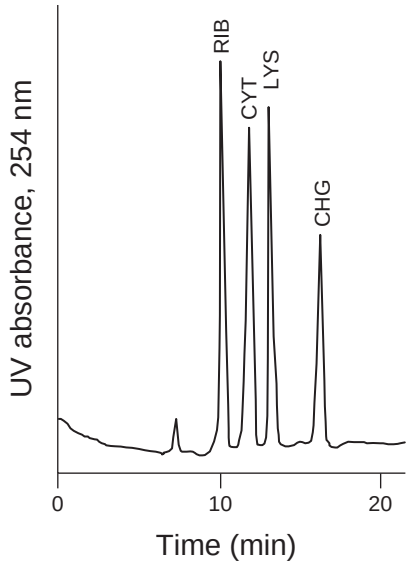

Figure 28 Reversed-phase gradient electrochromatographic separation of proteins on a continuous-bed stationary phase. Column, fused-silica tube filled with a porous methacrylamide–stearyl methacrylate–dimethyldiallylammonium chloride monolithic polymer, 80 × 0.05 mm i.d.; mobile phase, (A) 5 mM phosphate pH 2.0–5% acetonitrile, (B) 5 mM phosphate pH 2.0–80% acetonitrile; linear gradient, 0–100% B in 25 min; applied voltage, 5.5 kV. Peak identification: RIB, ribonuclease A; CYT, cytochrome *c*; LYS, lysozyme; CHG, chymotrypsinogen A. (Reprinted with permission from C. Ericson, S. Hjerten, 'Reversed-phase Electrochromatography of Proteins on Modified Continuous Beds Using Normal-flow and Counterflow Gradients. Theoretical and Practical Considerations', *Anal. Chem.*, **71**, 1621–1627 (1999), Copyright 1999 American Chemical Society.)

ABBREVIATIONS AND ACRONYMS

AC	Affinity Chromatography
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CFFAB	Continuous-flow Fast Atom Bombardment
CI	Chemical Ionization
EI	Electron Ionization
ELCD	Electrochemical Detection
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization Mass Spectrometry
FAB	Fast Atom Bombardment
FL	Fluorescence
GC	Gas Chromatography
HIC	Hydrophobic Interaction Chromatography

HPLC	High-performance Liquid Chromatography
IEC	Ion-exchange Chromatography
IPRPC	Ion-pair Reversed-phase Chromatography
LALLS	Low-angle Laser Light Scattering
LC	Liquid Chromatography
MALDI	Matrix-assisted Laser Desorption/ Ionization
MIC	Metal Interaction Chromatography
MS	Mass Spectrometry
NPC	Normal-phase Chromatography
PAD	Pulsed Amperometric Detection
PEEK	Poly(ether ether ketone)
PS/DVB	Poly(styrene/divinylbenzene)
PTFE	Poly(tetrafluoroethylene)
RA	Radioactivity
RPC	Reversed-phase Chromatography
SEC	Size-exclusion Chromatography
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • High-performance Liquid Chromatography of Biological Macromolecules

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

REFERENCES

1. *Ullmann's Encyclopedia of Industrial Chemistry*, VCH, Weinheim, 1996.
2. M.T.W. Hearn, *HPLC of Proteins, Peptides and Polynucleotides*, VCH, New York, 1991.
3. R. Westermeier, *Electrophoresis in Practice: a Guide to Methods and Applications of DNA and Protein Separations*, Wiley-VCH, Weinheim, 1997.
4. E. Heftman, *Chromatography*, 5th edition, Elsevier, Amsterdam, 1992.
5. X. Geng, F.E. Regnier, 'Retention Model for Proteins in Reversed-phase Liquid Chromatography', *J. Chromatogr.*, **296**, 15–30 (1984).
6. C. Heller, *Analysis of Nucleic Acids by Capillary Electrophoresis*, Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, 1997.
7. J. Hirabayashi, K.-I. Kasai, 'Size-dependent, Chromatographic Separation of Double-stranded DNA Which is Not Based on Gel Permeation Mode', *Anal. Biochem.*, **178**, 336–341 (1989).
8. A. Bartha, J. Stahlberg, 'Electrostatic Retention Model of Reversed-phase Ion-pair Chromatography', *J. Chromatogr. A*, **668**, 255–284 (1994).
9. K.K. Unger, R. Janzen, G. Jilge, 'Packings and Stationary Phases for Biopolymer Separations by HPLC', *Chromatographia*, **24**, 144–154 (1987).
10. R.E. Majors, 'Analytical HPLC Column Technology – The Current Status', *LC-GC Int. Suppl.*, 7–21 (1998).
11. K.K. Unger, *Porous Silica. Its Properties and Use as Support in Column Liquid Chromatography*, Elsevier, Amsterdam, 1979.
12. R. Arshady, 'Beaded Polymer Supports and Gels I: Manufacturing Techniques', *J. Chromatogr.*, **586**, 181–197 (1991).
13. Z. Deyl, I. Miksik, F. Tagliaro, E. Tesarova, *Advanced Chromatographic and Electromigration Methods in BioSciences*, Elsevier, Amsterdam, 1998.
14. E.D. Katz, *High-performance Liquid Chromatography: Principles and Methods in Biotechnology*, John Wiley & Sons, Chichester, 1996.
15. W.R. LaCourse, C.O. Dasenbrok, 'Column Liquid Chromatography: Equipment and Instrumentation', *Anal. Chem.*, **70**, 37R–52R (1998).
16. R.E. Majors, 'New Chromatography Columns and Accessories at the 1999 Pittsburgh Conference, Part 1', *LC-GC Int.*, **12**, 344–355 (1999).
17. S.K. Paliwal, M. De Frutos, F.E. Regnier, 'Rapid Separations of Proteins by Liquid Chromatography', *Methods Enzymol.*, **270**, 133–151 (1996).
18. H. Chen, C. Horváth, 'High-speed High-performance Liquid Chromatography of Peptides and Proteins', *J. Chromatogr. A*, **705**, 3–20 (1995).
19. K.K. Unger, G. Jilge, J.N. Kinkel, M.T.W. Hearn, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by High-performance Liquid Chromatography. II Performance of Nonporous Monodisperse 1.5 µm Silica Beads in the Separation of Proteins by Reversed-phase Gradient Elution High-performance Liquid Chromatography', *J. Chromatogr.*, **359**, 61–72 (1986).
20. K. Kalghatgi, C. Horváth, 'Micropellicular Sorbents for Rapid Reversed-phase Chromatography of Proteins and Peptides', in *Analytical Biotechnology – Capillary Electrophoresis and Chromatography*, eds. C. Horváth, J.G. Nikelly, ACS Symposium Series, American Chemical Society, Washington, DC, 163–180, Vol. 434, 1990.
21. J.J. Kirkland, 'Superficially Porous Silica Microspheres for the Fast High-performance Liquid Chromatography of Macromolecules', *Anal. Chem.*, **64**, 1239–1245 (1992).

22. N.B. Afeyan, N.F. Gordon, I. Mazsaroff, L. Varady, S.P. Fulton, Y.B. Yang, F.E. Regnier, 'Flow-through Particles for the High-performance Liquid Chromatographic Separation of Biomolecules: Perfusion Chromatography', *J. Chromatogr.*, **519**, 1–29 (1990).
23. S. Hjertén, Y.M. Li, J.L. Liao, J. Mohammad, K. Nakazato, G. Pettersson, 'Continuous Beds: High Resolving, Cost-effective Chromatographic Matrices', *Nature (London)*, **356**, 810–811 (1992).
24. Q.C. Wang, F. Svec, J.M.J. Fréchet, 'Macroporous Polymeric Stationary-phase Rod As Continuous Separation Medium for Reversed-phase Chromatography', *Anal. Chem.*, **65**, 2243–2248 (1993).
25. R.P.W. Scott, *Liquid Chromatography Detectors*, Elsevier, Amsterdam, 1986.
26. I.S. Krull, M.E. Szulc, S.-L. Wu, 'Principles of Biopolymer Detection in HPLC', *LC-GC Int.*, **6**, 414–418 (1993).
27. K. Stulik, V. Pacakova, G. Jokuszies, 'Analysis of Dipeptides by Reversed-phase High-performance Liquid Chromatography without Derivatization Using Amperometric Detection on a Copper Electrode', *J. Chromatogr.*, **436**, 334–337 (1988).
28. J.N. Ye, R.P. Baldwin, 'Catalytic Reduction of Myoglobin and Hemoglobin at Chemically Modified Electrodes Containing Methylene Blue', *Anal. Chem.*, **60**, 2263–2268 (1988).
29. J.A. Cox, T.J. Gray, 'Flow Injection Amperometric Determination of Insulin Based Upon its Oxidation at a Modified Electrode', *Anal. Chem.*, **61**, 2462–2464 (1989).
30. J. Ye, R.P. Baldwin, J.W. Schlager, 'LCEC Determination of the Copper Protein Ceruloplasmin in Human Serum at a Polyaniline Chemically Modified Electrode', *Electroanalysis*, **1**, 133–140 (1989).
31. D.P. Malliaros, M.J. DeBenedetto, P.M. Guy, T.P. Tougas, E.G.E. Jahngen, 'LC–EC Determination of Nucleotides and Nucleic Acids: Application to Enzyme Assays and the Analysis of DNA Fragments', *Anal. Biochem.*, **169**, 121–131 (1988).
32. Z. El Rassi, *Carbohydrate Analysis. High-performance Liquid Chromatography and Capillary Electrophoresis*, Elsevier, Amsterdam, 1995.
33. H. Nakazawa, 'Fluorimetric Determination of Rat Growth Hormone in Pituitary Cell Culture by High-performance Liquid Chromatography with Post-column Derivatization', *J. Chromatogr.*, **417**, 409–413 (1987).
34. V.K. Boppana, C. Miller-Stein, J.F. Politowski, G.R. Rhodes, 'High-performance Liquid Chromatographic Determination of Peptides in Biological Fluids by Automated Pre-column Fluorescence Derivatization with Fluorescamine', *J. Chromatogr.*, **548**, 319–327 (1991).
35. K. Sauda, T.I.N. Imasaka, 'Determination of Protein in Human Serum by High-performance Liquid Chromatography with Semiconductor Laser Fluorometric Detection', *Anal. Chem.*, **58**, 2649–2653 (1986).
36. R.C. Montelaro, R.R. Rueckert, 'Radiolabeling of Proteins and Viruses in Vitro by Acetylation with Radioactive Acetic Anhydride', *J. Biol. Chem.*, **250**, 1413–1421 (1975).
37. J.A. Beavo, P.J. Bechtel, E.G. Krebs, 'Preparation of Homogeneous Cyclic AMP-dependent Protein Kinase(s) and Its Subunits From Rabbit Skeletal Muscle', *Methods Enzymol.*, **38**, 299–308 (1974).
38. P.J. Oefner, C.G. Huber, F. Umlauft, G.N. Berti, E. Stimpfl, G.K. Bonn, 'High-resolution Liquid Chromatography of Fluorescent Dye Labeled Nucleic Acids', *Anal. Biochem.*, **223**, 39–46 (1994).
39. S. Yonekura, M. Iwasaki, M. Kai, Y. Ohkura, 'Determination of Guanine and Its Nucleotides in Human Erythrocytes by High-performance Liquid Chromatography with Postcolumn Fluorescence Derivatization Using Phenylglyoxal Reagent', *J. Chromatogr. B*, **654**, 19–24 (1994).
40. H. Fujimori, T. Sasaki, K. Hibi, M. Senda, M. Yoshioka, 'Direct Injection of Blood Samples into a High-performance Liquid Chromatographic Adenine Analyzer to Measure Adenine, Adenosine, and the Adenine Nucleotides with Fluorescence Detection', *J. Chromatogr.*, **515**, 363–373 (1990).
41. A.M. Maxam, W. Gilbert, 'Sequencing End-labeled DNA with Base-specific Chemical Cleavages', *Methods Enzymol.*, **65**, 499–560 (1980).
42. H. Engelhardt, P. Ohs, 'Trace Analysis of Sugars by HPLC and Post-column Derivatization', *Chromatographia*, **23**, 657–662 (1987).
43. S. Honda, E. Akao, S. Suzuki, M. Okuda, K. Kakehi, J. Nakamura, 'High-performance Liquid Chromatography of Reducing Carbohydrates as Strongly UV Absorbing and Electrochemically Sensitive 1-Phenyl-3-methyl-5-pyrazolone Derivatives', *Anal. Biochem.*, **180**, 351–357 (1989).
44. S. Hase, S. Hara, Y. Matsushima, 'Tagging of Sugars with a Fluorescent Compound, 2-Aminopyridine', *J. Biochem.*, **85**, 217–220 (1979).
45. S. Honda, Y. Matsuda, M. Takahashi, K. Kakehi, 'Fluorimetric Determination of Reducing Carbohydrates with 2-Cyanoacetamide and Application to Automated Analysis of Carbohydrates as Borate Complexes', *Anal. Chem.*, **52**, 1079–1082 (1980).
46. S. Honda, T. Konishi, S. Suzuki, 'Electrochemical Detection of Reducing Carbohydrates in High-performance Liquid Chromatography after Post-column Derivatization with 2-Cyanoacetamide', *J. Chromatogr.*, **299**, 245–251 (1984).
47. T. Endo, M. Hoshi, S. Endo, Y. Arata, A. Kobata, 'Structures of the Sugar Chains of a Major Glycoprotein Present in the Egg Jelly Coat of a Starfish, *Asterias amurensis*', *Arch. Biochem. Biophys.*, **252**, 105–12 (1987).
48. M. Barber, R.S. Bordoli, R.D. Sedgwick, A.N. Tyler, 'Fast Atom Bombardment of Solids (FAB): A New Ion Source for Mass Spectrometry', *J. Chem. Soc., Chem. Commun.*, 325–327 (1981).

49. C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, 'Electrospray Interface for Liquid Chromatographs and Mass Spectrometers', *Anal. Chem.*, **57**, 675–679 (1985).
50. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**, 2299–2301 (1988).
51. D.N. Nguyen, G.W. Becker, R.M. Riggin, 'Protein Mass Spectrometry: Applications to Analytical Biotechnology', *J. Chromatogr. A*, **705**, 21–45 (1995).
52. J.A. Loo, H.R. Udseth, R.D. Smith, 'Peptide and Protein Analysis by Electrospray Ionization–Mass Spectrometry and Capillary Electrophoresis–Mass Spectrometry', *Anal. Biochem.*, **179**, 404–412 (1989).
53. S.D. Fuerstenau, W.H. Benner, 'Molecular Weight Determination of Megadalton DNA Electrospray Ions Using Charge Detection Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1528–1538 (1995).
54. K.K. Murray, 'Coupling Matrix-assisted Laser Desorption/Ionization to Liquid Separations', *Mass Spectrom. Rev.*, **16**, 283–299 (1997).
55. I.S. Krull, R. Mhatre, J. Cunniff, 'Biopolymer Detection, Part II – Low-angle Laser Light Scattering', *LC-GC Int.*, **2**, 83–88 (1995).
56. R. Mhatre, I.S. Krull, 'Interfacing Gradient Elution Ion-exchange Chromatography and Low-angle Laser Light-scattering Photometry for Analysis of Proteins', *J. Chromatogr.*, **591**, 139–148 (1992).
57. K.M. Gooding, F.E. Regnier, 'Size Exclusion Chromatography', in *HPLC of Biological Macromolecules. Methods and Applications*, eds. K.M. Gooding, F.E. Regnier, Chromatographic Science Series, Marcel Dekker, New York, 47–75, Vol. 51, 1990.
58. M. Dizdaroglu, 'Weak Anion-exchange High-performance Liquid Chromatography of Peptides', *J. Chromatogr.*, **334**, 49–69 (1985).
59. C.T. Mant, R.S. Hodges, 'Separation of Peptides by Strong Cation-exchange High-performance Liquid Chromatography', *J. Chromatogr.*, **327**, 147–155 (1985).
60. W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, 'Retention Model for High-performance Ion-exchange Chromatography', *J. Chromatogr.*, **266**, 3–21 (1983).
61. L. Szepesy, G. Rippel, 'High-performance Hydrophobic Interaction Chromatography of Proteins', *LC-GC Int.*, **5**, 24–29 (1992).
62. J. Porath, 'Immobilized Metal Ion Affinity Chromatography', *Prot. Express. Purif.*, **3**, 263–81 (1992).
63. P.O. Larsson, M. Glad, L. Hansson, M.O. Mansson, S. Ohlson, K. Mosbach, 'High-performance Liquid Affinity Chromatography', *Adv. Chromatogr.*, **21**, 41–84 (1983).
64. C.G. Huber, E. Stimpfl, P.J. Oefner, G.K. Bonn, 'A Comparison of Micropellicular Ion-exchange and Reversed-phase Stationary Phases for HPLC of Oligonucleotides', *LC-GC*, **14**, 114–127 (1996).
65. W. Bloch, 'Anion-exchange Separation of Nucleic Acids', *Eur. Pat.*, 0 507 591 A2.
66. C.G. Huber, P.J. Oefner, G.K. Bonn, 'Rapid and Accurate Sizing of DNA Fragments by Ion-pair Reversed-phase Chromatography on Alkylated Nonporous Polystyrene/divinylbenzene Particles', *Anal. Chem.*, **67**, 578–585 (1995).
67. C.G. Huber, G.N. Berti, 'Detection of Partial Denaturation in AT-rich DNA Fragments by Ion-pair Reversed Phase Chromatography', *Anal. Chem.*, **68**, 2959–2965 (1996).
68. L.W. McLaughlin, 'Mixed-mode Chromatography of Nucleic Acids', *Chem. Rev.*, **89**, 309–319 (1989).
69. H.W. Jarrett, 'Affinity Chromatography with Nucleic Acid Polymers', *J. Chromatogr.*, **618**, 315–339 (1993).
70. C. Niemann, W. Saenger, R. Nuck, B. Pfannemüller, 'HPLC of 4-Nitrophenyl-α-D-malto-oligosaccharides', *Carbohydr. Res.*, **215**, 15–23 (1991).
71. M.D. Luque de Castro, C. Gamiz-Gracia, 'Miniaturization: a Well-defined Trend in Separation and Preconcentration Techniques', *Anal. Chim. Acta*, **351**, 23–40 (1997).
72. M. Novotny, 'Recent Advances in Microcolumn Liquid Chromatography', *Anal. Chem.*, **60**, 500A–510A (1988).
73. C. Ericson, S. Hjertén, 'Pump Based on Thermal Expansion of a Liquid for Delivery of a Pulse-free Flow Particularly for Capillary Chromatography and Other Microvolume Applications', *Anal. Chem.*, **70**, 366–372 (1998).
74. D. Figeys, R. Aebersold, 'Nanoflow Solvent Gradient Delivery From a Microfabricated Device for Protein Identifications by Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **70**, 3721–3727 (1998).
75. B. He, N. Tait, F.E. Regnier, 'Fabrication of Nanocolumns for Liquid Chromatography', *Anal. Chem.*, **70**, 3790–3797 (1998).
76. H.J. Cortes, 'Developments in Multidimensional Separation Systems', *J. Chromatogr.*, **626**, 3–23 (1992).
77. J.K. Eng, A.L. McCormack, J.R. Yates III, 'An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database', *J. Am. Soc. Mass. Spectrom.*, **5**, 976–989 (1994).
78. A.S. Rathore, C. Horváth, 'Separation Parameters Via Virtual Migration Distances in High Performance Liquid Chromatography, Capillary Zone Electrophoresis and Electrokinetic Chromatography', *J. Chromatogr. A*, **743**, 231–246 (1996).

Capillary Electrophoresis

Barbara A. Siles
University of Denver Research Institute, Denver, USA

1 Historical Background **1**
2 Principles of Capillary Electrophoresis **2**
2.1 Theory 2
2.2 Separation Parameters and Peak Shape 3
3 Instrumentation **5**
3.1 Introduction 5
3.2 The High-voltage Power Supply 5
3.3 The Separation Capillary 5
3.4 Sample Introduction 6
3.5 Modes of Detection 6
3.6 Microchip Capillary Electrophoresis 10
4 Modes of Capillary Electrophoresis **11**
4.1 Free-solution Capillary Electrophoresis 11
4.2 Micellar Electrokinetic Chromatography 12
4.3 Capillary Isoelectric Focusing 14
4.4 Size-sieving Capillary Electrophoresis 14
4.5 Nonaqueous Capillary Electrophoresis 17
4.6 Capillary Electrochromatography 18
5 Applications **19**
5.1 Amino Acids, Peptides, and Proteins 19
5.2 Nucleic Acids and Oligonucleotides 21
5.3 Carbohydrates 24
5.4 Chiral Separations 25
5.5 Inorganic Ions 26
Abbreviations and Acronyms **27**
Related Articles **28**
References **29**

Over the past two decades, capillary electrophoresis (CE) has emerged as a family of powerful and versatile separation techniques. The electrically driven flow profiles generated in CE provide, routinely, separation efficiencies of 10^5 theoretical plates. Although theory predicts that the achievable separation efficiencies will be greatest for macromolecules with small diffusion coefficients, CE has also been successful in the analysis of small molecules, such as inorganic ions. Strategies to increase separation efficiencies on-line have involved sample stacking and isotachophoretic techniques. The former promotes higher separation efficiencies based on differences in conductivity between the background electrolyte (BGE) and the sample buffer; the latter utilizes a discontinuous buffer system to create discrete solute zones. The instrumentation employed in CE is simple in design and is compatible with a variety of detection modes, including ultraviolet/visible (UV/VIS) absorbance, fluorescence, electrochemistry, mass spectrometry (MS), and chemiluminescence (CL). CE methodologies can be miniaturized on to a silicon wafer and integrated with sample preparation and chemical reactions. In addition to free-solution capillary electrophoresis (FSCE) utilizing a BGE and potentially a variety of BGE additives, several alternative separation modes have been developed and implemented. These include micellar electrokinetic chromatography (MEKC), capillary isoelectric focusing (cIEF), size-sieving capillary electrophoresis (SSCE), and capillary electrochromatography (CEC). Some of these separation modes have been adapted from the slab gel electrophoresis (SGE) format, or are reminiscent of separation mechanisms in liquid chromatography (LC). Currently, one of the greatest challenges in CE involves the maintenance of the capillary surface to provide reproducible electro-osmotic flow (EOF) and significantly reduce capillary wall interactions. According to the mode of separation employed and the characteristics of the solutes to be analyzed, both covalent and noncovalent (dynamic) capillary wall modifications have been demonstrated.

1 HISTORICAL BACKGROUND

Fundamental developments in the understanding of electrolysis, the electrical double layer, displacement electrophoresis, moving boundary electrophoresis, and other phenomena over the course of a 200-year period led to the emergence of CE as a family of powerful separation techniques. In 1967, Hjerten was the first to report free solution electrophoresis, conducted in 3-mm inner diameter (ID) capillary tubes.[1] The full advantages of this technique could not be realized until methods to introduce smaller quantities into the capillary tubes were developed to avoid sample overloading. In addition, the use of narrower bore capillary tubing (25–100 μm ID), with greater heat dissipation, would allow the use of higher field strengths (up to 900 V cm^{-1}) for faster and more efficient separations. In 1981, these advantages were realized by Jorgenson and Lukacs in what is considered the first report of modern FSCE or capillary zone electrophoresis (CZE).[2] Since then, other landmark developments have been made including the development of MEKC,[3] the development of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

microchip CE,[4] and the development of CEC.[5] We are now closing in on the second decade of modern CE separations, performed in narrow-bore fused-silica tubing. Currently, CE pervades many areas of analysis encompassing organic, inorganic, environmental, pharmaceutical, and biotechnological applications. Initially it was predicted that CE would replace modern LC in terms of solving difficult separation problems. However CE, being fundamentally different in its principle of separation, appears to complement separations performed using LC. Currently, commercial CE equipment is in its third generation, utilizing microtiter plates for high sample throughput and appropriate interfaces for MS detection.

2 PRINCIPLES OF CAPILLARY ELECTROPHORESIS

2.1 Theory

The electrophoretic mobility, μ_{ep}, of an ionic species migrating through a medium in an electrical field is directly proportional to electrical forces and inversely proportional to frictional forces acting on the ion (Equation 1):

$$\mu_{ep} \propto \frac{\text{electrical force } (F_e)}{\text{frictional force } (F_f)} \tag{1}$$

In an electrical field, ions are accelerated by the electrical force F_e, which is a function of field strength, E, and charge of the ion, q (Equation 2):

$$F_e = qE \tag{2}$$

The opposing frictional force acting on a spherical species is defined by Equation (3):

$$F_f = -6\pi\eta r v \tag{3}$$

where η is the viscosity of the electrophoretic medium, r is the radius of the solvated species, and v is its velocity. The ionic species reaches a steady state when electrical forces are balanced with frictional forces. Hence, equating Equations (2) and (3), electrophoretic mobility can be defined according to Equation (4):

$$\mu_{ep} = \frac{q}{6\pi\eta r} \tag{4}$$

Electrophoretic mobility is also the proportionality constant between the electrophoretic velocity and the applied field strength, E (Equation 5):

$$v_{ep} = \mu_{ep} E = \mu_{ep} \frac{V}{L_t} \tag{5}$$

where V is the applied voltage and L_t is the total capillary length. The apparent mobility, μ_{app}, of a species is the vector sum of its electrophoretic mobility, μ_{ep}, and the electro-osmotic mobility, μ_{eof}, due to the presence of EOF inside the capillary (Equation 6):

$$\mu_{app} = \mu_{ep} + \mu_{eof} \tag{6}$$

EOF refers to the migration of fluids along the length of the capillary. In a buffer-filled, fused-silica capillary, an electrical double layer is established at the interface of the electrolyte solution and the fused-silica capillary surface (Figure 1). The weakly acidic silanol groups at the capillary surface exist in either the protonated or the deprotonated form according to the following chemical equilibrium (Equation 7):

$$SiOH_{(s)} \rightleftharpoons SiO_{(s)}^{-} + H_{(aq)}^{+} \tag{7}$$

At buffer pH values greater than approximately 2, the capillary wall is negatively charged. Cations surrounded by their waters of hydration are strongly attracted to and partially neutralize the negatively charged silanol sites, forming the compact layer. A diffuse layer, consisting of

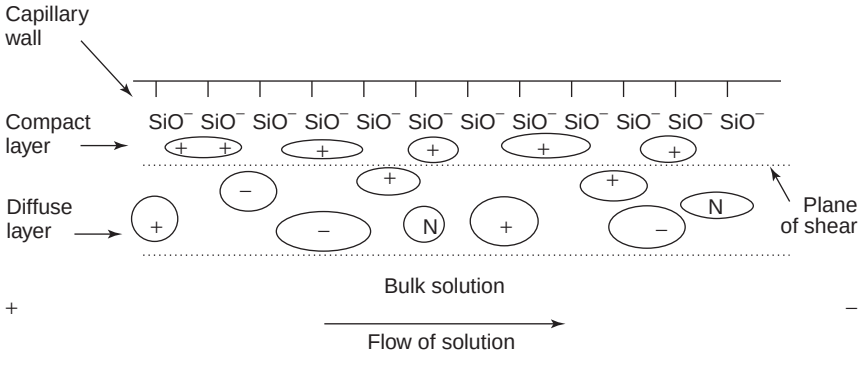

Figure 1 Electrical double layer at the capillary wall and creation of EOF (N = neutral solute). (Reproduced by permission of Academic Press, Inc., from A. Weston, P.R. Brown, 'Capillary Electrophoresis', *HPLC and CE: Principles and Practice*, Academic Press, San Diego, Chapter 4, 138, 1997.)

hydrated cations, anions, and neutral species, is formed adjacent to the compact layer, neutralizing the remaining negative charge. Under the influence of a longitudinally applied electrical field, the diffuse layer and the bulk solution migrate toward the negatively charged cathode, generating EOF. The zeta potential, ζ, is the potential formed across the two layers of cations at the capillary wall and is defined according to Equation (8):

$$\zeta = \frac{4\pi\delta e}{\varepsilon} \tag{8}$$

The zeta potential is a function of the thickness of the double layer, δ, the charge per unit of surface area, e, and the dielectric constant of the electrophoretic medium, ε. The thickness of the double layer is a function of buffer composition and is independent of the applied potential. The EOF velocity of a buffer system is given by Equation (9):

$$v_{\text{eof}} = \left(\frac{\varepsilon\zeta}{\eta} \quad E \right. \tag{9}$$

demonstrating that the magnitude of EOF may be modified by altering characteristics of the BGE, such as its pH, ionic strength, and viscosity. Note that EOF velocity is independent of the capillary dimensions.

Cations will migrate toward the negatively charged electrode (the cathode); anions will migrate toward the positively charged electrode (the anode) unless EOF is sufficient in magnitude to reverse the direction of their migration. Neutral species demonstrate no inherent electrophoretic mobility and hence migrate at the EOF velocity. Therefore, EOF can be determined for a particular system by introducing a neutral solute and observing its migration time, t, as given in Equation (10):

$$t = \frac{L_t L_{\text{d}}}{\mu_{\text{app}} V} \tag{10}$$

where L_t is the total capillary length, L_{d} is the effective capillary length from the injection end of the capillary to the detector, μ_{app} is the apparent mobility, and V is the applied voltage. EOF may be effectively eliminated by covalently or dynamically coating the capillary surface, and in that event $\mu_{\text{app}} = \mu_{\text{ep}}$.

2.2 Separation Parameters and Peak Shape

At the point of injection, a sample zone will be infinitesimally narrow. As migration through the capillary proceeds, a concentration gradient will be encountered as a function of time, t, and the distance migrated from the initial point, x, to generate a peak Gaussian in shape that can be described by Equation (11):[6]

$$C_{x,t} = \frac{k}{2(D\pi t)^{0.5}} \exp -\frac{(x - v_{\text{ep}}t)^2}{4Dt} \tag{11}$$

where $C_{x,t}$ is the solute concentration, k is a constant, D is the diffusion coefficient of the solute, and v_{ep} is the electrophoretic velocity of the solute. Integration of Equation (11) yields a mathematical expression that describes the resulting concentration–distance profile. The electrophoretic peak width is influenced directly by the shape of the solute profile within the capillary. Peak width is often measured as variance, σ, which is the standard deviation for a Gaussian distribution. Variance is related to the diffusion coefficient of the analyte, D, and the time, t, that the solute is allowed to diffuse from an infinitesimally narrow zone (Equation 12):

$$\sigma^2 = 2Dt \tag{12}$$

Combining Equations (10) and (12) yields a description of peak variance in relation to the apparent electrophoretic mobility of the solute (Equation 13):

$$\sigma^2 = \frac{2DL_{\text{d}}}{\mu_{\text{app}} E} \tag{13}$$

The theoretical plate height, H, is the proportionality constant between the variance of an electrophoretic peak and the distance that the corresponding solute zone has traveled through a capillary (Equation 14):

$$H = \frac{2Dt}{L_{\text{d}}} = \frac{2DL_t}{\mu_{\text{app}} V} \tag{14}$$

The concept of theoretical plate height was developed originally to describe separation efficiency in chromatographic techniques. Later, this concept was adapted by Giddings to electrophoretic techniques.[7] The van Deemter equation (Equation 15) summarizes the dependence of plate height on on-column band broadening phenomena in chromatographic (partitioning) processes:

$$H = A + \frac{B}{\mu_x} + C\mu_x \tag{15}$$

where A, B, and C are constants for a specified separation system, and μ_x is the eluent linear flow rate. The first term, A, describes band broadening due to multiple paths in a packed column. The second term, B/μ_x, represents band broadening due to longitudinal diffusion. The third term, $C\mu_x$, describes the contribution of resistance to mass transfer to band broadening due to the equilibration of an analyte between the stationary phase and the mobile phase. The multiple path term (A) and the resistance to mass transfer term ($C\mu_x$) do not contribute to band broadening in CE, as no stationary phase is present. Therefore, under ideal CE separation conditions, the only on-column contribution to band broadening is longitudinal diffusion. The electrically driven, nearly plug flow profiles produced in CE result in inherently higher

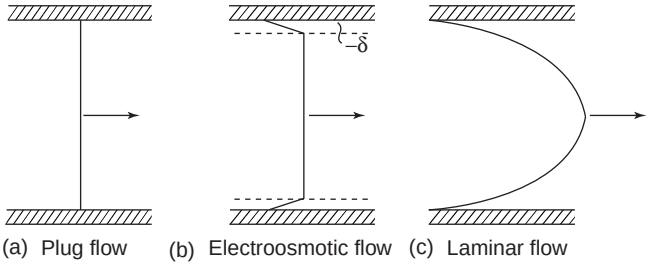

Figure 2 Various flow profiles. (Reproduced by permission of Academic Press, Inc., from A. Weston, P.R. Brown, 'Capillary Electrophoresis', *HPLC and CE: Principles and Practice*, Academic Press, San Diego, Chapter 4, 143, 1997.)

separation efficiencies than pressure-driven LC systems with laminar (parabolic) flow profiles (Figure 2a–c).

The total plate height, H_T, for an electrophoretic peak can be described as a sum of all plate height contributions (Equation 16):

$$H_T = \frac{\sigma_{tot}^2}{L_t} \tag{16}$$

The total variance for an electrophoretic peak, σ_{tot}^2, can be considered as the sum of dispersive phenomena on a solute zone in addition to the variance due to diffusion, σ_{dif}^2 (Equation 17):

$$\sigma_{tot}^2 = \sigma_{dif}^2 + \sigma_T^2 + \sigma_{inj}^2 + \sigma_{wall}^2 \tag{17}$$

Each factor of Equation (17) will now be defined. For a more detailed theoretical discussion, the reader is referred elsewhere.[8]

Parabolic temperature gradients across the capillary, caused by undissipated Joule heating, contribute to band broadening;[9] the variance associated with this phenomenon is approximated by Equation (18)[10] for capillaries with radii between 25 and 50 μm:

$$\sigma_T^2 = \frac{r^6 E^4 C_b^2 B^2 K^2 t}{1536 D k_b^2 T^4} \tag{18}$$

where r is the capillary radius, E is the applied field strength, C_b is the concentration of the BGE, B is a constant associated with the viscosity of the BGE, K is the conductance of the BGE, t is time, D is the diffusion coefficient of the solute, k_b is the thermal conductivity of the BGE, and T is absolute temperature. Equation (18) demonstrates the importance of employing narrow-bore separation capillaries in CE.

Injection of a sample zone greater than approximately 1 mm in length[11] promotes another form of peak variance, σ_{inj}^2, as given by Equation (19):[12]

$$\sigma_{inj}^2 = \frac{l_{inj}^2}{12} \tag{19}$$

where l_{inj} is the length of the injected (approximately rectangular) sample plug.

NonGaussian peak shapes may be caused also by solute adsorption to the internal capillary wall,[13] as given by Equation (20)[14] at a specified temperature:

$$\sigma_{wall}^2 = L_t \left\{ \frac{l^2}{12 L_t} + \frac{2D}{\mu_x} + \frac{k'\mu}{(1+k')^2} \left(\frac{r^2}{4D} + \frac{2}{k_d} \right) \right\} \tag{20}$$

where l is the length of the solute zone, L_t is the total capillary length, D is the diffusion coefficient of the solute, μ_x is the linear BGE flow rate, k' is the retention factor, r is the capillary radius, and k_d is the first-order rate constant for dissociation of the solute from the capillary surface. The solute–capillary wall interactions are usually solute-specific.

Mismatched conductivity between the sample zone and the BGE can alter the electrophoretic mobilities of the corresponding ions, generating nonGaussian peak shapes. In general, if the mobility of the BGE co-ion is higher than the mobility of the solute ion, peak ''tailing'' will be observed, whereas, the converse results in peak ''fronting''. Experimentally, actual peak shapes may deviate from these generalized predictions. Using a mathematical model, Gebauer and Bocek were able to predict peak symmetry as a function of BGE concentration and pH according to peak shape diagrams.[15]

In addition, laminar flow with parabolic-shaped flow profiles may be caused by siphoning between buffer reservoirs with uneven liquid levels; laminar flow is also present in the front end of a hydrodynamically introduced sample zone. Separation efficiency can be compromised further by the finite detection window volume, or coiling of the separation capillary, with greater band broadening observed for smaller capillary radii.[16]

The dimensionless parameter of efficiency, N, is related to the following previously defined parameters (Equation 21):

$$N = \frac{L_t}{H} = \frac{\mu_{app} V}{2D} \tag{21}$$

Hence the efficiency of a band in CE is a function of applied voltage and is not dependent on capillary length at a constant voltage. Higher applied voltages will yield faster analyses times and higher separation efficiencies, assuming that the Joule heating generated is dissipated sufficiently. In addition, theory predicts that separation efficiency will be high for a large solute with a small diffusion coefficient. Theoretical efficiencies (typically greater than 10^5) are greater generally than those determined experimentally due to the aforementioned extracolumn band broadening phenomena.

Resolution, R_s, is a dimensionless parameter that accounts for both analyte migration and peak dispersion in describing the degree of separation between two analytes (Equation 22):

$$R_s = \left(\frac{1}{4}\right) \frac{\Delta\mu_{app}}{\bar{\mu}_{app}} N^{1/2} \quad (22)$$

It follows that resolution may be increased by increasing the applied voltage. However, resolution increases by the factor $V^{1/2}$, hence the Joule heat produced at higher applied voltages will dictate a practical limit. Resolution may also be increased by decreasing EOF or reversing the flow of an analyte (counter to EOF). Since the analysis time, and hence band broadening, will increase as a result, an optimization trade-off is realized. The pH of the BGE can influence resolution according to the charge and mobility of individual analytes in addition to the relative mobilities of analytes. Relative analyte mobility is described by selectivity, α, for solutes i and j (Equation 23):

$$\alpha = \frac{\mu_{app,j}}{\mu_{app,i}} \quad (23)$$

where, by definition, α is always greater than 1. Alternatively, resolution may also be calculated directly from an electropherogram according to Equation (24):

$$R_s = \frac{2(t_i - t_j)}{w_{b,i} + w_{b,j}} \quad (24)$$

where t is the migration time of the analytes and w_b is peak width at the baseline of a Gaussian peak ($w = 4\sigma$) in time units.

3 INSTRUMENTATION

3.1 Introduction

A CE system is simple in design (Figure 3). The necessary components include two electrolyte reservoirs [a source (inlet) vial and a destination (outlet) vial], a separation capillary, a high-voltage power supply, a detector, and a data acquisition device for data collection and readout. In order to perform a sample ''injection'', the inlet buffer reservoir is replaced by the sample vial, and nanoliters of sample solution are introduced into the capillary either electrokinetically or hydrodynamically. In order to perform a CE separation, the sample vial is replaced by the inlet electrolyte reservoir and the power supply is turned on. It is imperative that the electrolyte levels of the inlet and outlet vials are equalized in height in order to eliminate siphoning effects. The capillary can be liquid-cooled or air-cooled in order to dissipate Joule heating and improve the reproducibility of analyses. A variety of detectors have been employed in CE as described in section 3.5.

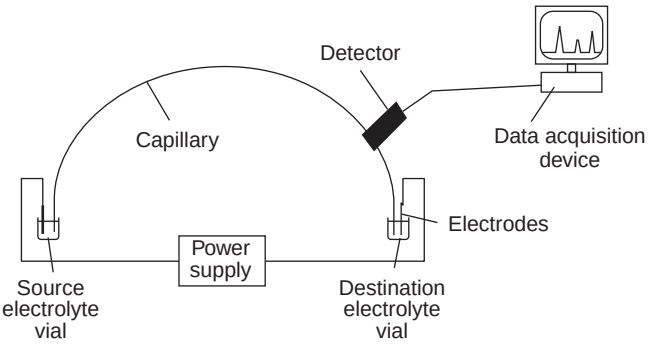

Figure 3 Schematic diagram of a CE instrument (not drawn to scale). (Reproduced by permission of Academic Press, Inc., from A. Weston, P.R. Brown, 'Instrumentation for Capillary Electrophoresis', *HPLC and CE: Principles and Practice*, Academic Press, San Diego, Chapter 6, 186, 1997.)

3.2 The High-voltage Power Supply

The electrical field and resulting EOF in CE are established using a direct current high-voltage power supply. In the normal polarity mode, the high-voltage cable is connected to the anode that resides in the electrolyte inlet reservoir; the ground cable is connected to the cathode that resides in the electrolyte outlet reservoir. The reversed polarity mode, where the anode and cathode of the system are switched, may be necessary for various separation mechanisms. The range of operation is typically from -30 to $+30$ kV, with a resulting operating current range of $\pm 300\,\mu$A. Access to the high-voltage electrode should be safety interlocked. Usually, a constant voltage is applied, although constant current and constant power modes can be applied as well. Field strength gradients can be applied to enhance separation performance.[17] Two 30-kV high-voltage power supplies can be used simultaneously, one connected to the anode and the other to the cathode, to apply up to 60 kV. High-voltage power supplies have also been adapted to run in the pulsed-field mode for the analysis of large molecules.[18]

3.3 The Separation Capillary

Fused silica is the most commonly used substrate for separation capillaries, providing both mechanical and chemical stability. The capillaries are coated externally with polyimide for enhanced strength and durability. A small (1-cm) section of the external polyimide coating is removed for alignment in an optical detector. Capillary IDs of 10–100 μm allow efficient dissipation of Joule heating produced during the application of high voltages. The

use of narrower bore capillaries can provide higher efficiencies, but limits the concentration sensitivity of optical detectors. Narrower bore capillaries are also susceptible to obstruction by microparticulates. Capillary lengths are typically 30–100 cm. Alternative capillary substrates and geometries have been explored briefly.[19,20] Shorter capillaries facilitate faster analysis times, but are more susceptible to sample overload. The ends of the capillary should be cleaved using a ceramic cutting tool by scoring the capillary in a single direction and pulling straight apart. This assures flat capillary edges that are advantageous especially during hydrodynamic sample introduction.

3.4 Sample Introduction

In CE, nanoliter sample volumes must be introduced efficiently and reproducibly. The two most common modes of sample introduction are hydrodynamic and electrokinetic. For hydrodynamic sample introduction, positive pressure is applied to the sample vial, negative pressure (vacuum) is applied to the outlet electrolyte vial, or the sample vial is elevated relative to the outlet reservoir in order to create a siphoning effect. Hydrodynamic sample introduction is nondiscriminatory, that is, the analyte and matrix components are introduced into the capillary in a manner that represents the original sample composition. If positive or negative pressure is applied, Equation (25) can be used to calculate the sample volume injected, V_{inj}:

$$V_{inj} = \frac{\Delta P d^4 \pi t_{inj}}{128 \eta L_t} \tag{25}$$

where ΔP is the pressure differential, d is the capillary ID, t_{inj} is the time of injection, η is the buffer viscosity, and L_t is the total capillary length. The main disadvantages of hydrodynamic sample introduction are that run-to-run reproducibilities are dependent on atmospheric pressure, the condition of the instrumentation, and the viscosity of the sample solution. Additionally, hydrodynamic sample introduction introduces a laminar flow profile to the front end of the sample zone which contributes to band broadening.

Alternatively, samples can be introduced into the capillary electrokinetically. In this mode of sample introduction, the inlet buffer reservoir is replaced by the sample vial and a high voltage is applied across the capillary for a short period of time. Both EOF and electrophoretic migration are responsible for introducing the sample into the capillary. The quantity of sample injected (Q), in moles, may be calculated according to Equation (26):

$$Q = (\mu_{ep} + \mu_{eof}) \frac{\kappa_B}{\kappa_S} E \pi r^2 C t \tag{26}$$

where μ_{ep} and μ_{eof} are the electrophoretic and electroosmotic mobilities, respectively, κ_B/κ_S is the ratio of conductivity of the buffer to the solute, E is the applied field strength, r is the internal capillary radius, C is the analyte concentration, and t is the time of injection. This method of sample introduction yields characteristically efficient peaks, but introduces sample bias as smaller, more mobile species in the sample matrix may be injected preferentially over larger, less mobile species. This in turn may lead to sample depletion over time and does not allow for accurate quantitation of sample components with different electrophoretic mobilities.

Sample stacking is a simple and universal way to increase simultaneously the peak efficiency and detection sensitivity. This is based upon conductivity differences between the solute zone and the BGE; typically, a 1 : 10 ratio of sample zone and BGE conductivities produces a Gaussian peak shape. When the separation potential is applied, the sample zone experiences a greater electrical field than the BGE owing to its lower conductance. Ions in the sample zone migrate rapidly until reaching the concentration boundary of the BGE. Positive species stack towards the front of the sample zone, negative species stack towards the back of the sample zone, and neutral species remain in the central portion of the sample zone. The higher conductivity of the BGE decreases the electrophoretic velocity of the sample species; the latter become concentrated in the narrow zone. Sample stacking has been applied in a variety of applications such as the analysis of neutral solutes in MEKC,[21] ultratrace environmental samples,[22] and DNA fragments.[23]

Trace analytes in dilute samples, especially those with complex matrices, may be concentrated up to 10 000-fold by the on-line coupling of isotachophoresis (ITP) with FSCE.[24] In ITP, analyte bands migrate at a constant velocity in an applied electrical field, positioned between a leading electrolyte and a terminating electrolyte of a discontinuous buffer system. The width of each discrete, ''stacked'' solute zone adjusts to maintain the same concentration as the leading electrolyte. Once the analyte bands reach the FSCE capillary, they migrate according to their velocity in a particular BGE and are separated. ITP preconcentration has been applied to the CE analysis of pharmaceutical samples,[25] the fractionation of human serum proteins from biological samples,[26] and chlorophenols in river and industrial wastewaters.[27]

3.5 Modes of Detection

The majority of detectors that have been applied in CE were adapted from LC. However, CE requires more stringent specifications owing to its narrower, lower volume solute zones. The CE detection method chosen for a particular application should reflect the

properties of the analyte and its sample matrix, in addition to the separation mode and BGE employed. The dynamic range, sensitivity, and selectivity of the detection method should be considered also. Detector response times of approximately 100 ms are adequate to preserve simultaneously the separation performance and detection sensitivity. For optical detection methods, on-column detection is used where a portion of the outer capillary coating is removed to create the detection window.

Many compounds possess chromophores that absorb light in the ultraviolet (UV) (190–350 nm) or visible (VIS) (350–700 nm) regions of the electromagnetic spectrum; alternatively, nonabsorbing compounds can be derivatized with UV/VIS-absorbing chromophores. As a result, the most common on-column detection mode is UV/VIS absorbance spectrophotometry. The widespread applicability of this detection mode is attributed also to its low maintenance and relatively low cost. Beer's law is observed theoretically for on-column absorbance (A) measurements (Equation 27):

$$A = \varepsilon b C \tag{27}$$

where ε is the molar absorptivity of the chromophore, b is the path length (in this case, the ID of the capillary), and C is the analyte concentration. Hence the narrow capillary dimensions limit the concentration sensitivity of this detection technique. Attempts have been made to enhance detection concentration sensitivity beyond the commonly observed range of 10^{-7}–10^{-5} M by increasing the optical path length, in "Z-cells"[28] or "bubble cells".[29] In the former, a section of the capillary is bent and absorbance is measured longitudinally down the central portion of the "Z"; in the latter, the ID of the capillary at the detection window is increased three to five times. The most common light source in UV detectors is a deuterium arc lamp, although low-pressure atomic vapor lamps can also be used. A tungsten–halogen lamp is employed commonly in the VIS range. Laser light sources have also been employed in order to enhance detection sensitivity.[30]

In a conventional, variable-wavelength UV/VIS absorbance detector, a monochrometor (single-wavelength selector containing a diffraction grating) and photon transducer (detector) are employed in a "forward" optics design. Alternatively, a polychrometer and a photodiode array (PDA) detector may be employed in a "reverse" optics design where the positions of the diffraction grating and the capillary window are reversed. The polychrometer allows the simultaneous detection of analyte absorbance at multiple wavelengths; the PDA detector can be advantageous in the confirmation of peak purity and in the identification of components in a mixture. Culbertson and Jorgenson increased the signal-to-noise ratio of a 1500-diode PDA by a factor of 85 by collecting and averaging electropherograms from individual diodes.[31] Fiber optics can be used instead of ball lenses, to collect and focus light from the capillary on to the detector. Additionally, a charge-coupled device (CCD) camera can be incorporated into the UV detection system where absorbance is recorded in two dimensions, as a function of both capillary distance and wavelength.[32]

UV/VIS absorbance detection may be applied more universally in the indirect mode.[33] A strongly UV/VIS-absorbing chromophore is added to the BGE to provide a stable background absorbance signal. Solute zones are detected as "negative" peaks owing to their displacement of the background chromophore. Typically, limits of detection are an order of magnitude poorer in the indirect mode of UV/VIS absorbance than the direct mode. To maintain a Gaussian peak shape, the electrophoretic mobilities of the analytes and the background chromophore should be comparable.

Fluorescence detectors are inherently more sensitive than UV/VIS absorbance detectors, making single-molecule detection in the VIS range a reality.[34] In CE, femtomolar limits of detection have been achieved over a wide range of path lengths (capillary diameters),[35] and linear dynamic ranges characteristically span several orders of magnitude. As molecular fluorescence is less common than UV/VIS absorbance, fluorescence detection is inherently more selective. In simple, on-column fluorescence detection, excitation radiation produced from a light source passes through a monochrometer and is focused on the capillary window. Fluorescent emission from the sample within the capillary then passes through a second monochrometer that is at a 90° angle to the excitation light path, and ultimately impinges upon a photomultiplier tube for signal detection. Intense light sources such as high-pressure mercury lamps and xenon pulse flash lamps have been applied. The use of Ar^+, He–Cd, or diode lasers as fluorescent light sources provides enhanced focusing capabilities and reduces stray light for enhanced limits of detection. However, wavelength selection can be limited with laser light sources and such high light intensities can cause photodegradation. In order to reduce the spectral background associated with the capillary walls, postcolumn[36] and sheath-flow[37] cell designs have been constructed. For further enhancement of detection limits, the fluorescent emission can be collected using a microscope objective[38] or an optical fiber. Wavelength-resolved fluorescence detection utilizing a PDA or a CCD camera has been applied to obtain qualitative analyte information without sacrificing sensitivity.[39] Time-resolved fluorescence measurements have been applied to discriminate against background Rayleigh and Raman scattering.[40] Numerous derivatization chemistries have been employed to impart fluorescence to nonfluorescent

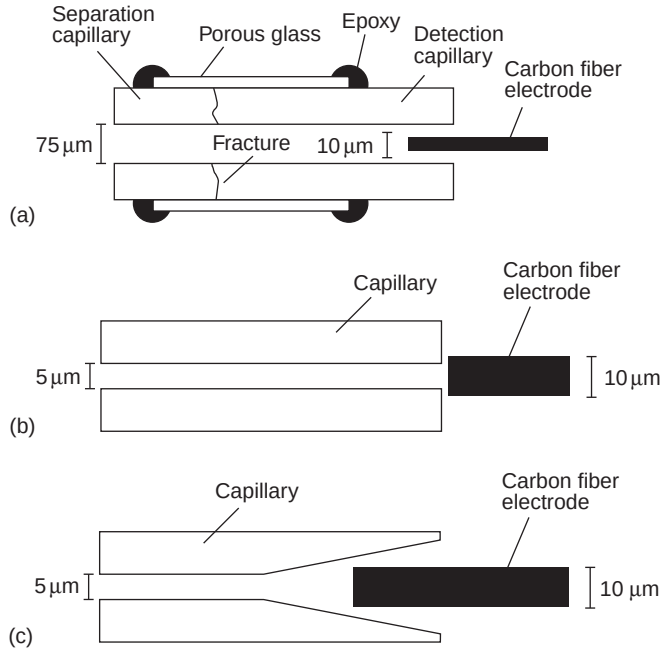

Figure 4 ED schemes for CE: (a) porous glass coupler; (b) end-column detection; (c) optimized end-column detection. (Reproduced with permission of the American Chemical Society from A.G. Ewing, J.M. Mesaros, P.F. Gavin, *Anal. Chem.*, **66**(9), 527A–537A (1994). Copyright 1994 American Chemical Society.)

molecules. Although precolumn derivatization is most common, reports have been published utilizing postcolumn detection.[41] Fluorescent molecules employed for this purpose include *o*-phthalaldehyde (OPA), naphthalenedialdehyde (NDA), fluorescein isothiocyanate (FITC), phenylthiohydantoin (PTH), and others.[42] Indirect fluorescence detection can be employed also in CE. Analogous to indirect absorbance detection, a fluorescent ion such as quinine sulfate is added to the BGE to provide a stable background signal.[43] As an analyte molecule passes through the capillary window, the fluorescent ion in the background is displaced, and a ''negative'' analyte signal is recorded.

Electrochemical detection (ED) is cost-effective and constructed easily on a scale that is compatible to CE. In contrast to optical detection methods used in CE, capillary electrophoresis/electrochemical detection (CE/ED) does not experience a loss in sensitivity upon miniaturization, and is applicable to almost any species that can be oxidized or reduced. Three-electrode configurations are common, with a saturated calomel or Ag/AgCl reference electrode, a platinum wire auxiliary electrode, and a working electrode. Limits of detection in the 10^{-19}–10^{-17} mol range have been reported for CE/ED.[44] In the past several years, advances have been made to address the major challenge of CE/ED: the isolation of the detection system from the electrophoretic potential.

For on-line detection, the electrochemical cell current (100 fA) is isolated from the electrophoretic current (100 μA) by creating a fracture or gap between the separation portion of the capillary and the detector portion of the capillary (Figure 4a). A grounded joint constructed from porous glass,[45] porous graphite,[46] or Nafion[47] surrounds the fracture or gap and eliminates the need to place the outlet end of the capillary in a buffer reservoir. The joints can be prone to leakage and breakage, and introduce band broadening. In the end-column format, the detection electrode may be micropositioned at the end of the separation capillary (Figure 4b).[48] To facilitate the placement of the electrode, and hence the precision of analysis, the capillary end can be etched chemically to a greater ID (Figure 4c).[49] Detection limits comparable to those reported for on-line detection are approached as the outer diameter of the electrode approaches the ID of the capillary.

Amperometry is the most sensitive and most commonly employed mode of CE/ED. The current arising from the oxidation or reduction of an analyte is measured when a constant potential is applied between the reference electrode and the working electrode. Chemically modified electrodes can be utilized to facilitate slow electron transfer kinetics.[50] Fouling of the electrode surface can be controlled with pulse amperometric detection.[51] Voltammetry, measuring current as the potential applied between the reference and working electrodes is scanned, can be employed to provide qualitative analyte information.[52] Potentiometric detection utilizing ion-selective electrodes has been applied less frequently in CE.[53]

Conductivity detection is an alternative, nearly universal mode of detection in CE that has been applied to the analysis of both organic and inorganic ions.[54] Detection limits are similar to those achieved using UV/VIS absorbance detection. A pair of inert electrodes is inserted into the sides of a capillary and aligned carefully in an on-, off-, or end-column arrangement. A small, constant current is applied between the two electrodes, and the conductivity of the BGE is measured. The presence of an ion in the BGE with a differing conductivity will be detected as a solute zone. As in amperometric ED, the challenge in designing a conductivity detector is the isolation of the electrophoretic potential from the electrochemical measurement. The on-column configuration may be more challenging to construct reproducibly, but resolution can be enhanced owing to the characteristically low detector dead volume. The end-capillary configuration has the potential disadvantage of extraband broadening, but the advantage of versatility in coupling with different capillaries. The suppressed conductivity detection mode achieves lower detection limits than the nonsuppressed mode. In the former, an ion-exchange

membrane is placed at the end of the separation capillary for suppression of the conductivity of the BGE.[55]

Unique applications of CE/ED include single cell analyses[56] and the coupling of CE/ED with microdialysis for the analysis of electrochemically active species in living animals.[47] Future advances point to the further development of electrochemical array detectors for the analysis of dynamic electrochemical and separation processes.[57]

Capillary electrophoresis/mass spectrometry (CE/MS) achieves effectively two-dimensional separations, combining the high separation efficiency of CE with the ability to determine mass and structural elements of a molecule. However, interfacing CE with MS detection presents significant challenges experimentally and instrumentally. First, restrictions are placed on the BGE; those containing alkali metals, surfactants, or size-sieving polymers can be particularly problematic. Acidic buffers of low ionic strength tend to yield high detection sensitivity;[58] organic cosolvents can also be beneficial as well.[59] Second, even though mass detection limits are in the low-femtomole range, concentration detection limits are in the 10^{-5}–10^{-4} M range.[60] To circumvent this problem via CE, sample stacking, ITP, and on-line sample preconcentration can be employed. To circumvent this problem via MS, alternative interface designs, mass analyzers, and detectors can be investigated.

Three different sample ionization processes have been coupled with CE: electrospray ionization (ESI), continuous-flow fast atom bombardment (CFFAB), and matrix-assisted laser desorption/ionization (MALDI). In ESI, eluate flows from the end of the capillary through a needle held at a potential of approximately 3 kV at atmospheric pressure.[61] Solvent is evaporated quickly by the surrounding flow of a hot gas, and molecules (typically less than 100 000 Da) are ionized prior to entering the MS analyzer. In fast atom bombardment (FAB), the sample is mixed with a glycerol-based matrix (to avoid freezing of the solvent in the low-pressure conditions) and deposited directly on a probe tip. Subsequent impact with high-energy (8–10 keV) species such as xenon molecules causes desorption of the sample into the gas phase. The continuous-flow version (CFFAB) has been demonstrated for molecules smaller than 3000 Da.[62] MALDI has been almost exclusively limited to off-line analyses in CE.

Common interfaces used in CE/MS include sheath flow, sheathless flow, and liquid junction (Figure 5a–c). In the sheath-flow design, a conducting, sheathing solution is added in a 100 : 1 ratio relative to the BGE exiting the capillary.[60] This is of utility for both ESI and CFFAB that require higher liquid flow rates than that provided by EOF (microliters versus nanoliters per minute). The sheathless-flow interface at low flow rates has been used

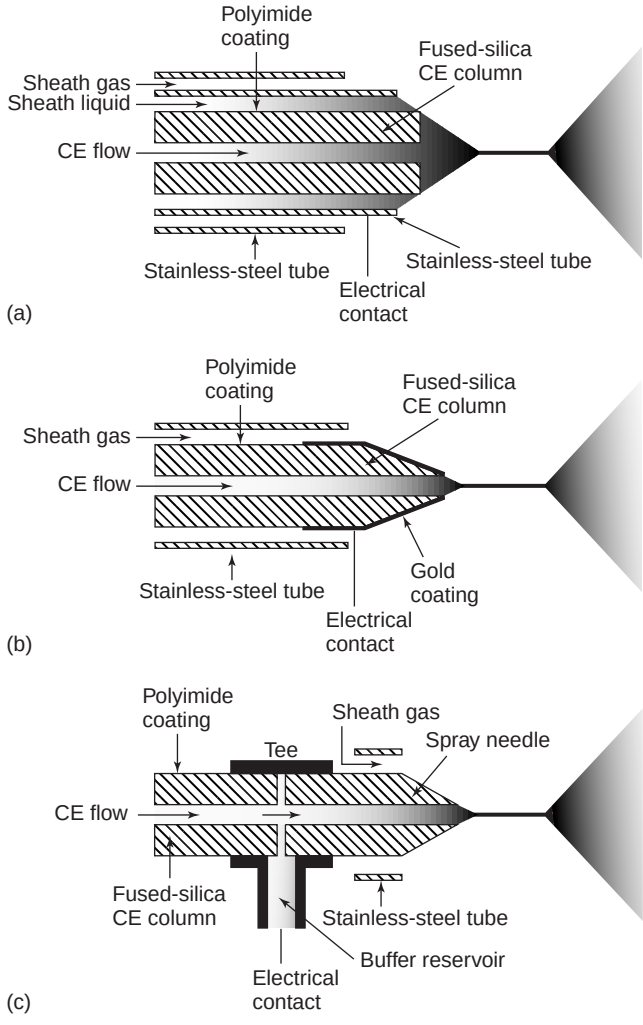

Figure 5 Different types of CE/MS interfaces: (a) sheath flow; (b) sheathless flow; and (c) liquid junction. (Reproduced by permission of John Wiley & Sons, Inc. from J.F. Banks, *Electrophoresis*, **18**, 2255–2266 (1997).)

in conjunction with ESI for enhanced sensitivity owing to a reduced level of background ions.[63] In the liquid-junction design, the separation capillary does not enter the mass spectrometer. The outlet end of the capillary and a transfer line are positioned end-to-end, separated by a 50-μm gap.[64] A reservoir of make-up fluid surrounds the gap to promote fluid transfer.

To enhance the sensitivity of analysis in CE/MS, selected-ion monitoring (SIM) is a viable technique. In SIM, the mass spectrometer detects only a small range of mass/charge ratios as opposed to a full spectrum of mass/charge ratios. In addition to quadrupole and magnetic sector mass analyzers, time-of-flight (TOF),[65] ion-trap[66] and Fourier transform (FT) ion cyclotron mass analyzers[67] have been used to improve sensitivity marginally. The FT mode of MS can offer additional enhancement in mass resolution capabilities.

The applications of CE/MS have been plentiful. Particularly successful applications have involved the determination of impurities in peptide and protein samples[68] and the determination of drug metabolites in urine.[69] Currently, the majority of instrumentation for CE/MS is in the developmental stage. As commercial instrumentation and interfaces between CE units and mass spectrometers progress, the scope and promise of CE/MS are expected to grow.

Alternative, less common detection modes have been explored in conjunction with CE. CL detection is analogous to fluorescence detection, but the analyte molecules are excited by a chemical reaction as opposed to a high-intensity light source. Detection limits for CL are intermediate between those for UV/VIS absorbance and fluorescence detection.[70] Refractive index (RI) detectors measure changes in the RI of the BGE as analytes migrate across the capillary window. Although RI detectors are universal in application, they demonstrate low selectivity and low concentration sensitivity, and are extremely prone to interferences.[71] Radionuclide detection offers high selectivity and sensitivity in association with CE, but can introduce obvious health hazards in the laboratory. Radiation from the first-order radioactive decay of a particular nuclide (such as ^{14}C, ^{11}C, ^{99}Tc, ^{32}P, or ^{18}F) is exposed to a scintillation device that converts the radiation into photons for detection.[72] More recently, Raman spectroscopy,[73] nuclear magnetic resonance (NMR) spectroscopy,[74] and light scattering detection[75] modes have been applied in CE.

3.6 Microchip Capillary Electrophoresis

Microfabricated separation devices are miniaturized analytical systems that integrate sample introduction, electrophoretic separation, and in some instances chemical reactions, on a planar chip or wafer (Figure 6a). These postage-stamp-sized devices can manipulate picoliter volumes of liquid electrokinetically without external valves or pumps through separation channels that are millimeters in length. Higher electrical field strengths (up to 2500 V cm^{-1}) relative to conventional CE may be applied due to enhanced Joule heat dissipation. As a result, higher separation efficiencies may be achieved and analyses may be performed nearly 100 times faster (less than 1 ms in some instances) (Figure 6b). Shorter analysis times minimize longitudinal diffusion and further reduce band broadening.

Insulating substrate materials, such as glass or fused quartz, can be microfabricated as CE microchips using standard photolithographic techniques and wet-chemical etching procedures. The glass or fused-quartz wafers are annealed thermally, polished mechanically, cleaned, and sputter-coated with a metal layer, typically gold and chromium. A positive photoresist layer, approximately 1–2 μm in thickness, is then spin-coated on the surface. A photomask is applied and the wafer is exposed to UV light, producing the desired channel pattern. The exposed metal layers are removed and channels are etched using an acid bath to reveal the channel pattern. Finally, the etched wafer and cover plate are cleaned, aligned, and thermally bonded to form capillary channels. Holes are drilled to allow access to the channels. The bonding process can be repeated to eliminate defects. Further

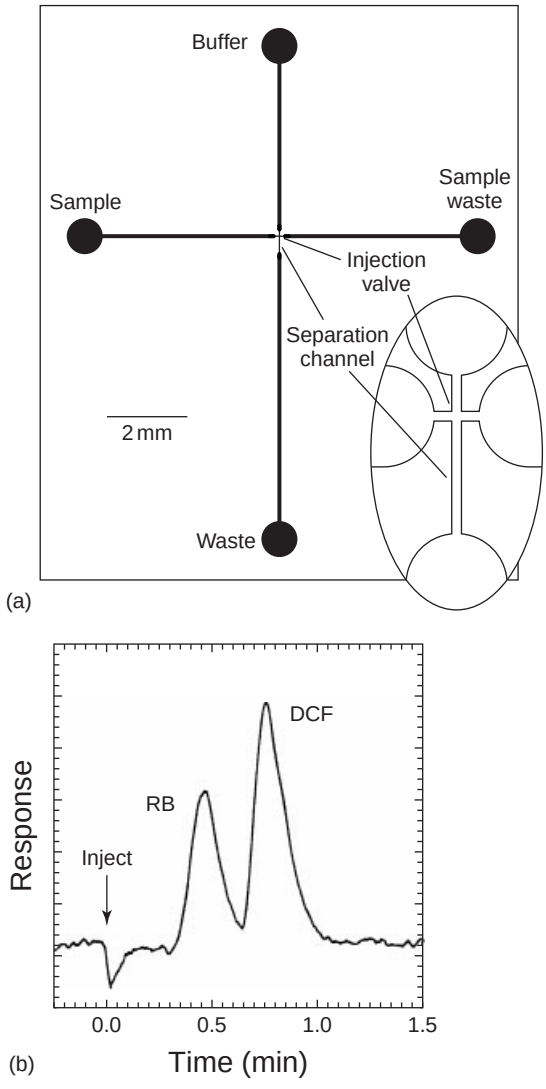

Figure 6 (a) Schematic of a microfabricated device used for high-speed electrophoretic separations. Inset: enlargement of the injection valve and separation channel. (b) High-speed electropherogram of Rhodamine B (RB) and dichlorofluorescein (DCF) resolved in 1 ms using a separation field strength of 53 kV cm^{-1} and a separation length of 200 μm. The start time is marked with an arrow at 0 ms. (Reproduced with permission of the American Chemical Society from S.C. Jacobson, C.T. Culbertson, J.E. Daler, J.M. Ramsey, *Anal. Chem.*, **70**(16), 3476–3480 (1998). Copyright 1998 American Chemical Society.)

details of the microfabrication process are provided elsewhere.[76]

Differences between substrates should be noted. The usage of glass results in more background noise.[77] Additionally, glass fluoresces at 325 nm, whereas quartz does not fluoresce. Despite their less desirable optical properties, glass substrates are used predominantly owing to their ease of micromachining and bonding. An alternative microchip substrate is a flexible poly(dimethylsiloxane) (PDMS) silicone elastomer that is molded to a microfabricated silicon master with a raised, etched channel pattern.[78] A solution consisting of Sylard 184 and its curing agent is poured over the wafer and allowed to cure for several hours. The PDMS slab is peeled away; holes are punched to allow access to the channels. The PDMS slab is then encased in a solid, rectangular, PDMS container. PDMS offers several benefits over glass or quartz substrates. The fabrication of PDMS microchips is more time and cost efficient. In the event that channels or holes of the device become clogged, the molded PDMS slab may be removed from the solid PDMS encasement, rinsed, dried, and reused. PDMS has a low RI of 1.430, lessening the magnitude of excitation light reflected in optical detection schemes. Overall, PDMS substrates are less advantageous electrically, thermally, and optically relative to glass or fused quartz.

Prior to sample injection, the channels are flushed hydrodynamically with a buffer solution using a syringe or pipet. Platinum electrodes are placed into the inlet and outlet reservoirs for the application of voltage. Samples are injected electrokinetically; automation of sample injection enhances the reproducibility of migration times and reduces sample diffusion.

Channel geometry influences the sample introduction process, and hence the overall performance of the CE microchip. Three basic components, the straight channel, the ''T'', and the cross, can be combined to produce simple to sophisticated channel patterns. The ''double-T'' injector provides high separation efficiency.[79] In this configuration, the channel connecting the sample and sample waste reservoirs intersects perpendicularly with the separation channel (Figure 6a). Potential is applied between the sample and sample waste reservoirs. Once the slowest moving analyte has entered the intersection, the potential is switched and applied across the separation channel only. The sample plug in the intersection is swept down the separation channel and is detected. The latter is referred to as ''floating injection'' and yields sufficient relative standard deviation (RSD) values of approximately 2%. This injection technique can allow leakage and eddy mixing of the sample at the channel intersection corners, distorting field lines. For more control of sample injection volume, Jacobson et al. also introduced ''pinched injection'', where the potential is applied simultaneously across both of the channels.[80] This injection technique yields higher peak efficiencies and improved RSD values (approximately 0.3%). More sophisticated injection techniques offering finer control of injection volumes and the possibility for sample stacking have also been demonstrated.[81]

Separation efficiencies attained for the initial microfabricated devices were relatively low (less than 100 000 theoretical plates). However, separation efficiencies continue to increase with improvements in wafer design. Alternative separation channel patterns have yielded efficiency values approaching 10^6 theoretical plates; these include the serpentine[82] and synchronized cyclic CE[83] (four short channels in a square arrangement) designs. In lower resolution, higher throughput purification applications, free-flow electrophoresis has been demonstrated.[84] Several detection methods have been used in conjunction with microchip CE, including UV/VIS absorbance,[85] laser-induced fluorescence (LIF),[86] ED,[87] and MS with an ESI interface.[88] LIF detection is the most widely employed mode of detection for microfabricated devices owing to its high sensitivity and selectivity. Confocal microscopy fluorescence detection offers the possibility of single-molecule detection.[89] The coupling of microchip CE devices with MS is still in its infancy.

Several different CE separation modes have been applied using microchip CE. MEKC of coumarin dyes using sodium dodecyl sulfate (SDS) as the surfactant was demonstrated.[90] The same coumarin dyes were separated also using open-tubular capillary electrochromatography (OTCEC) with octadecyl functionalities immobilized on the channel walls.[91] The increased separation efficiencies and decreased analysis times have allowed microchip CE to be applied to the DNA sequencing, genotyping, and the amplification of DNA fragments.[92] Capillary array electrophoresis devices that use up to 12 lanes simultaneously have been applied to the detection of HLA-H variants in the diagnosis of hemochromatosis.[93] Cell lysis and polymerase chain reaction (PCR) amplification, and size-sieving of DNA fragments have been achieved on a single microchip.[94] Research has been developed also to facilitate the detection of DNA mutations on a microchip device.[95]

4 MODES OF CAPILLARY ELECTROPHORESIS

4.1 Free-solution Capillary Electrophoresis

FSCE, or CZE, is the most commonly employed mode of CE owing to its versatility; the applications of FSCE are as widely varying as the composition of

the BGEs. The separation capillary is filled with a predominantly aqueous buffer solution; additives may be incorporated into the BGE to achieve the desired selectivity and resolution of separation. The BGE chosen should demonstrate good buffering capacity and low background current to suppress Joule heating effects. The pH of the BGE can influence both the charge of the analyte and the fused-silica capillary wall according to their pK_a values. Hence the EOF in the capillary is also influenced. Some buffer additives may influence the solute – capillary wall interactions. The compatibility of FSCE with the aforementioned detection modes (section 3.5) is limited only by interferences from the BGE components. In the normal polarity mode, charged species are separated according to their respective charge-to-mass ratios in an applied electrical field. Cations migrate by coulombic attraction toward the cathode, with the cation of the lowest charge-to-mass ratio migrating fastest. Conversely, anions migrate toward the anode, with those of the lowest charge-to-mass ratio migrating fastest. At a sufficient EOF within the capillary, the coulombic attraction of the anions for the anode will be overcome, and the anions will migrate toward the cathode. Overall, the migration order in FSCE is cations (from high to low charge-to-mass), followed by neutral species (unseparated), and finally anions (from low to high charge-to-mass).

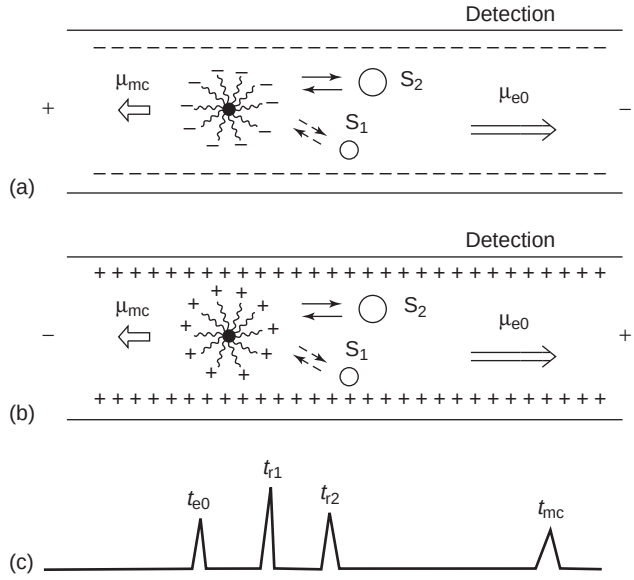

Figure 7 Migration of neutral solutes in MEKC using (a) anionic and (b) cationic pseudo-stationary phases. Separation of solutes S_1 and S_2 is achieved due to their differential partitioning into the pseudo-stationary phase. (c) The uncharged solutes are eluted within an elution window defined from t_{e0} to t_{mc}. (Reproduced by permission of John Wiley & Sons, Inc. from M.G. Khaledi, *High Performance Capillary Electrophoresis: Theory, Techniques, and Applications*, Wiley, New York, 78, 1998.)

4.2 Micellar Electrokinetic Chromatography

MEKC is one of only two modes of CE that allows the simultaneous separation of charged and neutral analytes. Since its introduction by Terabe et al. in 1984,[3] MEKC has grown dramatically to encompass hundreds of applications. The incorporation of micelles above their critical micelle concentration (CMC) in the BGE creates a pseudo-stationary phase. Differential partitioning of analytes with the micelles results in their differential migration through the capillary, and hence their separation. Since the separation mechanism involves both a partitioning process and electrophoretic migration, MEKC is considered a hybrid separation technique of FSCE and LC. However, MEKC offers several advantages over LC, including higher separation efficiencies, reduced analysis volumes and greater versatility in changing experimental parameters.

Micelles are spherical aggregates of surfactants with hydrophobic tail groups oriented toward the center of the micelle and polar head groups oriented toward the aqueous solution. The number of monomers in a micelle varies from surfactant to surfactant, and the presence of organic solvents in the separation buffer can disrupt micelle formation. The micelles will migrate through the capillary according to their apparent electrophoretic mobility, a function of both the charge-to-mass ratio of the surfactant and the EOF in the capillary. The migration of anionic micelles, such as SDS, opposes EOF in an uncoated capillary at normal system polarity. However, at sufficient EOF (pH > 6), the anionic micelles will migrate toward the cathode, past the detector (Figure 7a). The use of cationic surfactants, such as cetyltrimethylammonium bromide (CTAB), may require reversed system polarity. The cationic surfactant molecules adhere in a bilayer to negatively charged silanol groups at the capillary wall, rendering the capillary surface positively charged (Figure 7b). Nonionic and zwitterionic surfactants can also be utilized, but only in conjunction with charged surfactants. In addition to conventional surfactants, different types of organized media such as dendrimers[96] and ionic polymers[97] have been explored as psuedo-stationary phases in MEKC.

There exists a finite elution window (hence a finite peak capacity) in MEKC, between the time of migration for a neutral solute (such as methanol), t_{eo}, and the time of migration for a completely retained solute (such as the hydrophobic dye, Sudan III), t_{mc} (Figure 7c). For neutral analytes, the capacity factor, k', in MEKC is defined according to Equation (28):

$$k' = \frac{t_r - t_{e0}}{t_{e0}[1 - (t_r/t_{mc})]} \tag{28}$$

where t_r is the migration time of the solute. For charged analytes, Equation (28) must be adapted to include the time the analyte migrated in the BGE, t_0, (Equation 29):

$$k' = \frac{t_r - t_0}{t_0[1 - (t_r/t_{mc})]} \tag{29}$$

Note that as t_r approaches t_{mc}, Equations (28) and (29) are reduced to the capacity factor equation for chromatographic separations. Retention indices have been developed for MEKC, analogous to retention indices developed for gas chromatography (GC), to allow the normalization of different micellar systems.[98] The fundamental resolution equation in MEKC is given by Equation (30):

$$R = \left(\frac{N^{1/2}}{4}\right)\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k'_2}{1 + k'_2}\right)\left(\frac{1 - (t_r/t_{mc})}{1 + (t_{e0}/t_{mc})}k'_1\right) \tag{30}$$

where N is efficiency, α is separation selectivity (Equation 23), and k'_1 and k'_2 are the capacity factors for analytes 1 and 2, respectively. Note that resolution in MEKC decreases significantly as t_r approaches t_{mc}.

The pseudo-stationary phase used in MEKC affects analyte selectivity and resolution. Currently, there are many organized media to choose from; usually, the choice is made experimentally. In order to make a more informed selection, Yang and Khaledi have established linear solvation energy relationships,[99] taking into account the relative sizes, polarizability, hydrogen bond donating properties, and hydrogen bond accepting properties of the analytes and the various surfactant molecules. Increasing the micellar concentration above the CMC in the BGE can affect the size of the elution window, but has a negligible effect on analyte selectivity. Buffer additives such as organic solvents, glucose, and urea have been added to the BGE in MEKC in order to extend the elution window, while concomitantly reducing hydrophobic interactions between analytes and micelles and altering analyte selectivities. The incorporation of a cyclodextrin (CD) into the BGE introduces shape selectivity to the MEKC separation mechanism; this is especially useful in the analysis of polycyclic aromatic hydrocarbon (PAH) structural isomers.

SDS is by far the most commonly employed surfactant in MEKC. Owing to the strong hydrogen bond donating properties of SDS, this pseudo-stationary phase is applicable to a wide variety of analytes from caffeine metabolites[100] through DNA adducts[101] to flavanoids.[102] However, SDS micelles demonstrate a practical limit in the analysis of very hydrophobic compounds such as PAHs and steroids.[103] The incorporation of CDs into SDS micellar separation solutions has been shown to increase analyte selectivity (Figure 8a–c). For particularly hydrophobic analytes, bile salts such as sodium glycodeoxycholate and sodium taurocholate are more effective than SDS.[104,105] Fluorocarbon

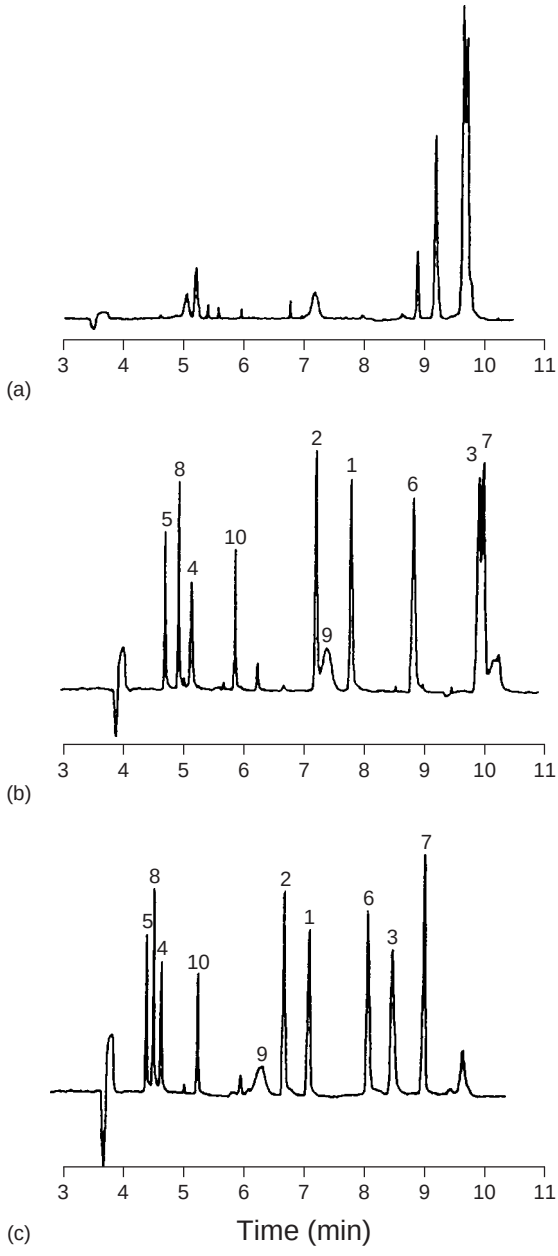

Figure 8 Effects of CDs on the MEKC separation of 10 estrogens. Buffers: 10 mM sodium borate (pH 9.2) containing 50 mM SDS and (a) 0 mM CD, (b) 20 mM β-CD and (c) 20 mM γ-CD. Capillary, 47 cm × 50 μm; voltage, 17 kV; pressure injection, 2 s; detection, absorption at 200 nm. Peak identification: (1) 17-β-estradiol; (2) 16-keto-17-β-estradiol; (3) 2-methoxyestradiol; (4) 2-hydroxyestradiol; (5) 4-hydroxyestradiol; (6) estrone; (7) 2-methoxyestrone; (8) 4-hydroxyestrone; (9) 16-α-hydroxyestrone; (10) estriol. (Reproduced from K.C. Chan, G.M. Muschik, H.J. Issaq, P.K. Siiteri, 'Separation of Estrogens by Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **690**, 149–154 (1995), Copyright 1995, with permission from Elsevier Science.)

micelles, such as lithium perfluorooctane sulfonate, are also applicable to the analysis of relatively hydrophobic solutes; these pseudo-stationary phases have been applied to a limited extent in MEKC, mostly owing to a lack of availability.[106] Anionic surfactants other than SDS have been applied in MEKC including *N*-D-gluco-*N*-methylalkanamide for amino acids (AAs)[107] and (*R*)- and (*S*)-*N*-dodecoxycarbonylvaline for chiral separations.[108] Cationic surfactants, such as CTAB, have been applied in the analysis of negatively charged species such as carboxylic acids[109] and nucleic acids.[110] Nonionic and zwitterionic micelles have been applied to the separation of AAs and polypeptides.[111] Mixed micellar separation systems have received a great deal of attention recently.[112] Major advantages include expanded elution windows and alternative selectivities to accommodate the separation of complex analyte mixtures. In contrast to conventional micelles, polymeric pseudo-stationary phases, which can tolerate up to 30% organic modifier in the BGE, have been applied to particularly hydrophobic solutes such as PAHs, fullerenes, and *N*-alkylphenones.[113] The use of vesicles or the simultaneous use of oppositely charged surfactants (SDS and dodecyltrimethylammonium bromide, for instance) has been shown to enhance greatly the elution window and increase solute selectivity.[114]

4.3 Capillary Isoelectric Focusing

Traditionally, the isoelectric focusing of proteins is carried out using SGE. The adaptation of this methodology to the capillary format, first demonstrated by Hjerten and Zhu in 1985, has revealed distinct advantages.[115] cIEF involves the separation of charged analytes, usually proteins or peptides, according to their respective isoelectric points (p*I*), which is a function of their surface charge. A stable pH gradient is established and maintained in the BGE by the addition of ampholytes, species that contain both basic and acidic functionalities. Analytes will migrate through the BGE until they encounter the pH that matches their p*I*. Ideally, a linear relationship between p*I* and relative analyte mobility is established (Figure 9a and b). When the focusing of analytes is complete, the current drops to approximately one-fifth of its original value. Nonideally, solutes may interact with the internal capillary surface. The capillary surface may be treated to control such interactions and also the mobility of solutes via EOF. Covalent capillary coatings have been utilized for this purpose.[116] Dynamic capillary wall modification has been achieved in uncoated capillaries by adding methylcellulose (MC) derivatives to the BGE.[117]

After the analyte bands have been focused, they must be detected. In the absence of EOF, two schemes are commonly utilized. The first scheme, chemical mobilization, involves buffer replacement. First, the analytes are focused according to their respective p*I* values. Once the system has reached a steady state, the inlet and outlet buffer reservoirs containing a highly acidic and a highly basic solution, respectively, are replaced with buffer reservoirs containing an additional counterion (typically Na^+ in the inlet and Cl^- in the outlet). The presence of the additional ions disrupts the established steady state and the system re-establishes equilibrium to maintain electroneutrality. The analyte molecules reacquire a net charge as the pH of each zone changes. The analyte bands are separated and migrate past the detector. Concomitantly, the current in the capillary increases according to the conductivity of the BGE. Owing to an inherent anodic or cathodic ''drift'' associated with cIEF, only analytes with p*I* values between approximately 5 and 9 may be analyzed in conjunction with buffer replacement.[118] Alternatively, negative pressure and voltage may be applied simultaneously in order to mobilize the focused analyte bands toward the cathode for detection.[119] The application of voltage is necessary to maintain focused analyte bands; otherwise, laminar flow introduced hydrodynamically promotes band broadening. Additionally for the detection of focused solutes in the absence of EOF, one can visualize the entire focused solute zone within the capillary using an array detector,[120] or move the capillary itself past a stationary detector.[121] Mobilizing the analyte bands past the detector in the presence of EOF results in faster analysis times. However, the variability of EOF according to the condition of the capillary surface, as well as the pH and ionic strength of the BGE, can lead to nonlinear functions of p*I* versus analyte mobility and restrict the useful analyte p*I* range to small intervals.[122] More recent innovations involve analyte zone fraction collection with off-line MALDI/TOF/MS detection[123] and cIEF performed on microchip devices.[124]

4.4 Size-sieving Capillary Electrophoresis

The size-dependent electrophoretic separation of proteins, DNA molecules, and carbohydrates in microbore capillaries offers several advantages over conventional SGE. Reduced quantities of samples, buffers, and gel matrices are required; quantitation is facilitated by on-line detection (typically LIF or UV/VIS absorbance); the system is more amenable to automation; and faster, more efficient separations are achievable at field strengths up to $500\ V\ cm^{-1}$. Owing to the anticonvective dimensions of microbore capillaries, less rigid matrices are required to separate biomolecules in the same size range. Capillary gel electrophoresis (CGE) utilizes chemically cross-linked gels and dynamic size-sieving capillary electrophoresis (DSCE) utilizes entangled polymer

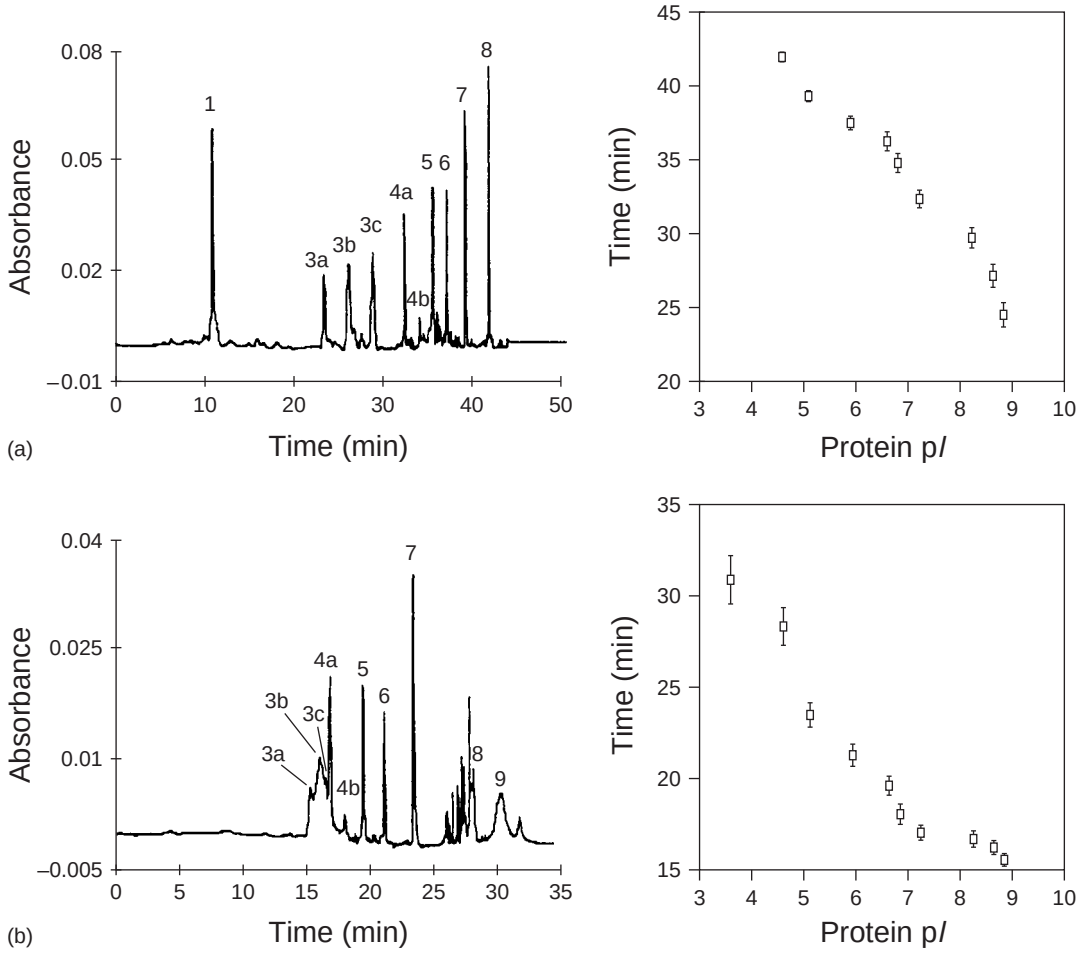

Figure 9 Electropherograms and pH gradient linearity plots for cIEF methods. (a) cIEF with chemical mobilization; (b) cIEF with hydrodynamic mobilization. Peak identification: (1) cytochrome *c*; (2) ribonuclease A; (3a–c) lentil lectin I–III; (4a, b) myoglobin I, II; (5) carbonic anhydrase I; (6) carbonic anhydrase II; (7) β-lactoglobulin; (8) trypsin inhibitor; (9) amyloglucosidase. (Reproduced with permission of the American Chemical Society from O. Hofmann, D. Che, K.A. Cruickshank, U.R. Muller, *Anal. Chem.*, **71**(3), 678–686 (1999). Copyright 1999 American Chemical Society.)

solutions. Currently, capillary arrays of up to 100 capillaries allow the separation of biomolecules in multiple ''lanes''.[125]

The electrophoretic mobilities of biomolecules with the same charge per molecular unit are effectively independent of molecular size. Therefore, the addition of a frictional component to the BGE is necessary to achieve size-dependent separations. Two models have been used to describe the separation of biomolecules in SGE, the Ogston sieving model[126] and the reptation model.[127] The former applies to the migration of nondeformable, approximately spherical biomolecules through a porous network. For species separated according to the Ogston sieving mechanism, a linear, semilogarithmic ''Ferguson plot'' can be constructed by plotting data according to Equation (31):

$$\log \mu_T = -\log \mu_0 - K_R(T) \qquad (31)$$

where μ_T is the corresponding electrophoretic mobility of the species at that gel concentration, μ_0 is the electrophoretic mobility of the species in free solution, the slope, K_R, is the retardation coefficient, and T is the percentage gel concentration. K_R is proportional to $(r + R_g)^2$, where r is the gel fiber radius and R_g is the radius of gyration of the species. In contrast, the reptation model describes the ''snake-like'' migration of relatively larger biomolecules through the separation matrix, with electrophoretic mobilities that are inversely related to molecular size. The biased reptation model[128] and the biased reptation model with fluctuations[129] have been used to account for the concomitant stretching and conformational fluctuations of much larger biomolecules at higher field strengths.

Chemically cross-linked gels employed in CGE have well-defined, rigid pore sizes and relatively high viscosities, resulting in characteristically high separation

efficiencies. Almost exclusively, polyacrylamide gels (PAGs) have been employed in CGE with acrylamide as the monomer and *N*,*N*′-methylenebisacrylamide as the cross-linking agent.[130] Gel properties are varied by adjusting the percentage of monomer (%T) and percentage of cross-linking agent (%C) during polymerization. For mechanical stability, the PAGs are chemically bonded to the capillary wall. Serious limitations of chemically cross-linked PAGs include (i) heat sensitivity leading to bubble formation with the elimination of EOF, (ii) sample contamination by carryover between analyses, (iii) severely limited lifetimes for individual capillaries (usually less than 10 analyses per capillary), and (iv) restriction of sample introduction to electrokinetic injection.[131]

In DSCE, low-viscosity polymer solutions are employed as size-sieving matrices at concentrations above or below the entanglement threshold of the polymer. These matrices possess a highly flexible, dynamic pore structure and may be replaced between analyses to enhance run-to-run reproducibilities. Linear polyacrylamide (LPA) solutions, with zero amount of cross-linking agent, have been used.[132] The use of derivatized agarose gels requires elevated capillary temperatures (i.e. above their gelling temperature) to facilitate the loading of the separation matrix and enhance the separation mechanism.[133,134] Other polymers have been used as size-sieving matrices in DSCE, including hydroxyethylcellulose (HEC),[135] MC,[136] hydroxypropylmethylcellulose (HPMC),[137,138] and poly(ethylene oxide) (PEO).[139] The mixing of different molecular weights of the same matrix and the mixing different matrices have resulted in enhanced separation performance.[139,140] More recently, Gao and Yeung introduced the use of polyvinylpyrrolidone (PVP) in uncoated capillaries for the separation of DNA fragments.[141]

Barron et al. demonstrated that a fully entangled network is not necessary for the dynamic size-sieving separation of DNA fragments.[142] Therefore, the Ogston sieving and reptation models that involve a porous network are insufficient descriptors of the separation mechanism in DSCE. As a result, the transient entanglement coupling mechanism was proposed, whereby large biomolecules undergo interactions with multiple size-sieving polymer molecules at any one time as they migrate through the capillary, yielding size-dependent separations (Figure 10). In DSCE, lower concentrations of higher molecular weight polymers are more effective in the separation of larger biomolecules, and high concentrations of lower molecular weight polymers are more effective in the separation of smaller biomolecules.

In DSCE, EOF within the capillary must be eliminated or effectively reduced such that the size-sieving matrix does not electrophorese from the capillary. The elimination of EOF also enhances run-to-run migration reproducibilities. To this end, both covalent and dynamic capillary coatings have been utilized. Alternatively, the surface silanols on the capillary wall may be fully protonated by treatment with 0.1 M HCl.[143] For the analysis of negatively charged biomolecules, it is necessary to use reversed system polarity with the negatively charged electrode (cathode) at the capillary inlet. Other experimental variables may be optimized, including applied field strength and capillary temperature. The migration velocity of biomolecules is directly proportional to the constant, applied field strength. Additionally, potential gradients[144] and pulsed electrical fields[145] have been applied to increase the separation efficiency and resolution and decrease the analysis time in biomolecular separations. Capillary temperature gradients may also be employed in DSCE; higher temperatures result in a decrease in viscosity and an increase in dynamic pore size.[17]

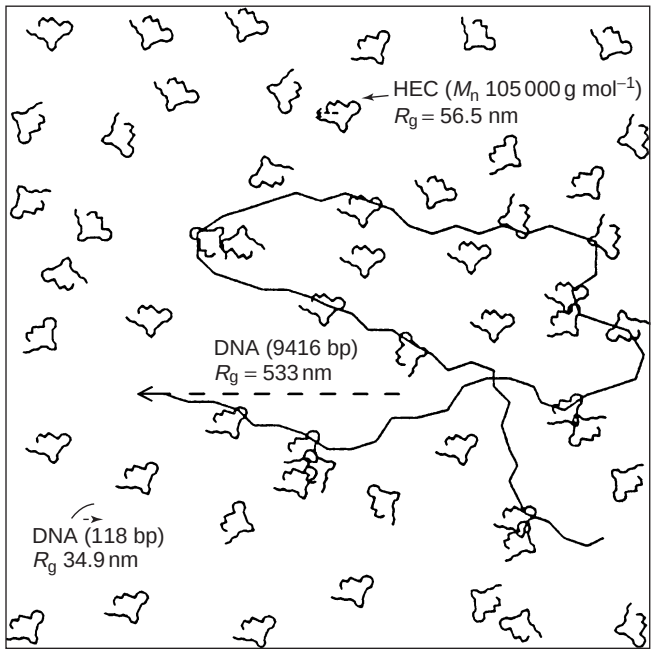

Figure 10 A schematic illustration of DNA motion in a dilute HEC solution. The approximate relative sizes of HEC and small and large DNA are shown; radius of gyration (R_g) values are also given. The 118 base pair DNA fragment is unlikely to entangle strongly with the HEC in dilute solution, whereas the 9146 base pair DNA fragment has a high probability of undergoing transient entanglement coupling with many HEC molecules. When transient entanglement coupling occurs, the larger DNA molecule must drag the uncharged HEC molecules along with it during electrophoresis, decreasing its electrophoretic mobility in a size-dependent manner. (Reproduced by permission of John Wiley & Sons, Inc. from A.E. Barron, H.W. Blanch, D.S. Soane, *Electrophoresis*, **15**(5), 597–615 (1994).)

4.5 Nonaqueous Capillary Electrophoresis

Aqueous-based buffer systems have been applied predominantly in CE separations owing to the many favorable characteristics of water and its well established chemical properties. In the mid-1980s, Walbroehl and Jorgenson were the first to report the use of predominantly nonaqueous solvents in CE.[146] Ten years later, Sahota and Khaledi demonstrated the potential of nonaqueous capillary electrophoresis (NACE) and provided systematic, comparative investigations to aqueous-based systems.[147] Nonaqueous BGEs can be particularly advantageous in the analysis of relatively hydrophobic compounds. Alternative separation selectivities may be achieved with the use of nonaqueous buffers relative to their aqueous counterparts (Figure 11a and b). In general, higher applied field strengths may be employed in NACE for faster, more efficient separations of improved detectability. The selection of appropriate nonaqueous solvents should include considerations of volatility, solubility of buffer salts and additives, toxicity, reactivity, cost, and flammability. Organic solvents of various chemical and physical properties, including *N*,*N*-dimethylformamide, formamide, methanol, ethanol, and acetonitrile, have been utilized successfully in NACE.

To a first approximation, electro-osmotic mobility is directly proportional to the ratio of dielectric constant to viscosity (ε/η) of a pure solvent.[148] A majority of nonaqueous solvents possess comparatively lower ε/η ratios than water, and as a result demonstrate decreased electro-osmotic mobilities. Simultaneously, the zeta potential at the capillary wall may be affected due to a change in the pK_a of the surface silanol groups; this further reduces electro-osmotic mobilities.[149] Nonaqueous BGEs of higher ionic strengths can be applied for higher sample solubility, sample stacking, and enhanced detectability. Lower currents are generated in nonaqueous solvents so that higher field strengths can be applied without detrimental Joule heating effects; Ohm's law plots show extended linearity in nonaqueous BGEs.[147] Analogously, the electrophoretic mobilities of analytes, and hence analyte selectivities, are influenced by the characteristic ε/η ratio of the solvent.[150] Organic solvents vary in their ability to stabilize ionic species in solution and accept or donate protons. Therefore, organic solvents alter the pK_a values of acidic and basic species; generally, the pK_a values of acidic species are more affected than basic species.[151] It is important to bear in mind that in nonaqueous systems, pH values of separation media and pK_a values of solutes must be replaced by the appropriate apparent values, pH* and pK_a^*, respectively.

Several modes of detection are compatible with NACE. Direct UV/VIS absorbance detection is most commonly employed, with detection limits generally in the 10^{-5} M range.[149] Limitations exist in the application of LIF detection to NACE owing to the potential of signal quenching in lower viscosity solvents, and also changes in fluorescent excitation and emission wavelengths according to solvent polarity.[148] The use of nonaqueous separation buffers in combination with solvent evaporation in electrospray ionization mass spectrometry (ESIMS) is particularly beneficial.[152] Detection limits for NACE in conjunction with ED as low as 10^{-9} M have been reported.[149]

NACE has been applied to several classes of charged solutes, including cationic pharmaceuticals and their metabolites,[153] metal ions, organic acids,[154] peptides,[147] carboxylic acids,[150] and anionic surfactants.[155] For the separation of neutral solutes, charged

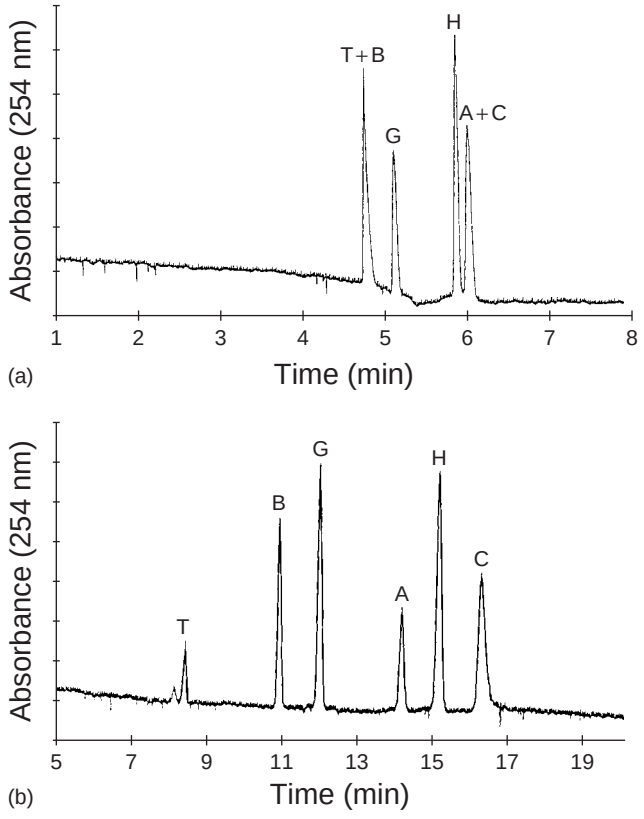

Figure 11 Comparison of aqueous versus NACE separation of anions in 75-μm ID poly(acryloylaminoethoxyethanol)-coated capillaries, 43 cm total length and 35 cm effective length. (a) Electropherogram of anions in 50 mM TRIS–acetate in water, pH 8.0; voltage, −15 kV. (b) Electropherogram of anions in 100 mM acetate in methanol, apparent pH 8.5; voltage, −12 kV. Peak identification: (T) *p*-toluenesulfonic acid; (B) benzoic acid; (G) *N*-acryloylglycine; (H) *p*-hydroxycinnamic acid; (A) *N*-acryloyl-γ-aminobutyric acid; (C) caffeic acid. (Reproduced from M. Chiari, E. Kenndler, 'Capillary Zone Electrophoresis in Organic Solvents: Separation of Anions in Methanolic Buffer Solutions', *J. Chromatogr. A*, **716**(1–2), 303–309 (1995), Copyright 1995, with permission from Elsevier Science.)

additives capable of interacting with the solutes must be employed. To this end, starburst dendrimers[96] and ionic polymers[113] have been utilized. The separation of PAHs by NACE requires charge-transfer or induced-dipole interactions provided by organic cations such as tropylium or phenyl-substituted pyrylium salts (tetrafluorborates and perchlorates).[156] Chiral separations have also been performed in NACE using CDs as the chiral selectors.[157] Such separations can be superior to their aqueous counterparts owing to the increased solubility of CDs in organic solvents and their reduced binding constants with extremely hydrophobic solutes.

4.6 Capillary Electrochromatography

CEC combines the high selectivity and versatility of LC stationary phases with the high efficiency, electrically driven flow of CE. Fused-silica capillaries are packed with stationary phase particles and the mobile phase is transported according to EOF. Neutral or charged solutes are separated on the basis of their differential partitioning interactions between the stationary phase and the mobile phase; charged solutes may also be separated according to their respective electrophoretic mobilities. In the absence of pressure-driven laminar flow, stationary phase particle diameters as small as 0.4 μm may be utilized. As a result, column efficiencies are usually an order of magnitude greater in CEC than in LC. The routine application of CEC to complex separation problems is currently limited by several factors, including the irreproducibility of EOF within the capillary, challenging column preparation, the production of Joule heat within the packed capillaries, and bubble formation within the capillary eliminating EOF. At present, it is unclear as to whether Joule heating within the capillary, inhomogeneous packing of the capillaries, or a combination of the two phenomena leads to bubble formation.

In CE, an electrical double layer forms at the solid/liquid interface of the capillary wall due to the presence of deprotonated silanol groups, producing EOF. In CEC, an electrical double layer is formed at the liquid/solid interface of each stationary phase particle. Theory dictates that EOF is independent of capillary ID if the latter is 20 times larger than the thickness of the electrical double layer.[158] However, the EOF in CEC is highly dependent on several experimental parameters, including characteristics of the capillary wall and stationary phase particles,[159] electrolyte concentration,[160] pH,[161] ionic strength, column temperature, and mobile phase organic modifier content.[162] The complex dependences of these parameters on the EOF result in the inherent irreproducibility of CEC. Additional on-column band broadening contributions exist relative to CE. Owing to the presence of a stationary phase in CEC, resistance to mass transfer and eddy diffusion are encountered in addition to longitudinal diffusion. In terms of off-column band-broadening processes, the column packing procedure and the presence of the frits can be problematic.[163] Rathore and Horvath have investigated discontinuities associated with flow velocity and electrical field strength in CEC.[164]

Column preparation in CEC, including frit formation and column packing, is considered to be an art at this time. The retaining frits are most commonly formed by the sintering of a plug of stationary phase particles, wetted with either water or potassium silicate solution,[165] or by in situ polymerization of a potassium silicate solution.[166] They must be mechanically stable and sufficiently porous, and must not promote significant solute band-broadening effects. Stationary phase particles may be packed by slurry techniques,[167] glass drawing,[168] the use of centripetal force,[169] or electrokinetic migration.[170]

Alternatively, OTCEC can be performed whereby the stationary phase is attached to the inner capillary walls or to channels of microchip devices.[91] The open-tubular format eliminates the eddy diffusion term of the van Deemter equation, but also reduces loadability and retention capabilities. OTCEC has been used in the separation of anions,[171] PAHs,[172] and proteins.[173]

A more recent innovation in CEC involves column construction via monolithic or sol–gel polymerization in situ to avoid the problems associated with inhomogeneous packing techniques and potentially problematic frit formation. For instance, *N*-isopropylacrylamide and 2-acrylamido-2-methylpropanesulfonic acid are copolymerized and cross-linked with *N*,*N*′-methyleneacrylamide.[174] Also, monolithic poly(butyl methacrylate–co-ethylene dimethacylate–co-2-acrylamido-2-methyl-1-propanesulfonic acid) capillary columns[175] have been synthesized and characterized in CEC. Sol–gel, monolithic fabricated OTCEC columns have been constructed,[176] forming a porous, high surface area layer on the capillary walls. Using an analogous in situ polymerization process, molecularly imprinted CEC columns may be fabricated. These have proven especially useful in chiral separation applications.[177]

The instrumentation employed for CEC resembles that of CE; however, buffer reservoirs may need to be pressurized in order to reduce bubble formation. Typical run conditions include the application of a potential of 10–50 kV across capillaries with dimensions of 0.5–1.5 m in length and 10–100 μm ID.[178] A detector window is created by removing a 1-cm section of the outer polyimide coating of the capillary. For in-column detection, the polyimide coating is removed from a packed portion of the capillary; for on-column detection, it is removed from an unpacked portion (Figure 12a). The former can lead to reduced sensitivity in optical detection modes

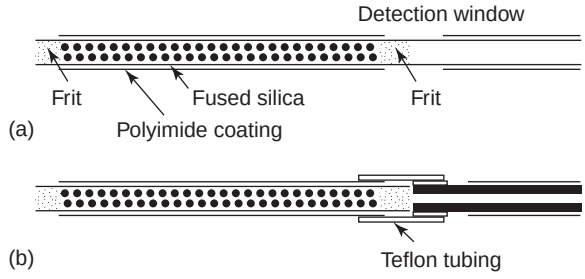

Figure 12 Packed capillary columns for electrochromatography as (a) an uncoupled and (b) a coupled system. (Reproduced from B. Behnke, E. Grom, E. Bayer, 'Evaluation of the Parameters Determining the Performance of Electrochromatography in Packed Capillary Columns', *J. Chromatogr. A*, **716**(1–2), 207–213 (1995), Copyright 1995, with permission from Elsevier Science.)

owing to background scattering from the particles; the latter can lead to the degradation of peak efficiency due to band broadening after passage through the end frit. Alternatively, a separate portion of capillary tubing containing a burned window may be coupled to the packed capillary using a Teflon® tubing sleeve (Figure 12b).[166] MS appears to be a viable detection scheme for packed CEC owing to the compatibility of mobile phase composition and flow rate (particularly for ESI interfaces) and the relatively higher sample capacities.[179]

A variety of stationary phase materials have been used in CEC to address chiral separations,[180] the analysis of pharmaceuticals,[181] and the resolution of PAH structural isomers.[182] Figure 13 shows the high-efficiency separation of 16 PAH structural isomers using CEC. Common particle diameters of stationary phase are 1.5–10 μm. The most common stationary phase material is octadecyl-derivatized silica. The choice of mobile phase not only affects analyte selectivity in CEC, but also directly influences the magnitude of EOF. Common organic modifiers in CEC for aqueous-based systems containing electrolyte include methanol, 2-propanol, and acetonitrile. Zare et al. demonstrated gradient elution CEC.[183] In this work, two capillaries were placed in different buffer reservoirs containing different mobile phase compositions. The two capillaries were connected to the separation capillary in a "T" formation. A separate power supply was connected to each mobile phase reservoir and the system was computer controlled to change the mobile phase composition as a function of time.

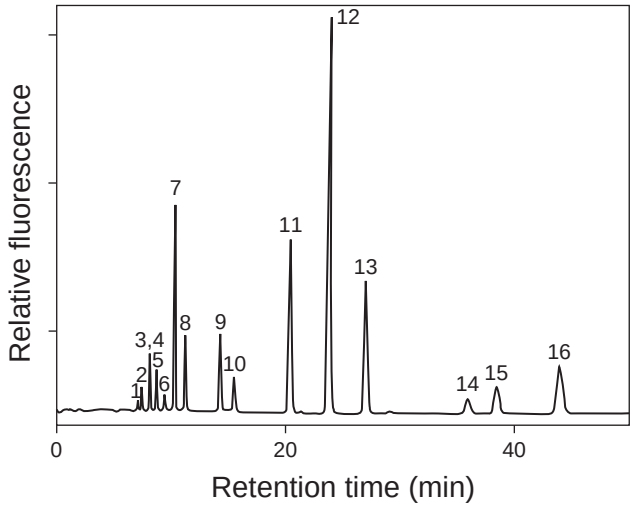

Figure 13 Electropherogram showing the capillary electrochromatographic separation of 16 PAHs. The column dimensions were 75 μm ID × 365 μm OD (33 cm packed length). The mobile phase consisted of 80% acetonitrile in 4 mM sodium borate solution. The applied voltage was 15 kV. Injection was performed electrokinetically at 5 kV for 5 s. PAH concentrations are approximately 10^{-6}–10^{-8} M. Peak identification: (1) naphthalene; (2) acenaphthylene; (3) acenaphthene; (4) fluorene; (5) phenanthrene; (6) anthracene; (7) fluoranthene; (8) pyrene; (9) benz[*a*]anthracene; (10) chrysene; (11) benzo[*b*]fluoranthene; (12) benzo[*k*]fluoranthene; (13) benzo[*a*]pyrene; (14) dibenz[*a*,*h*]anthracene; (15) benzo[*ghi*]perylene; (16) indeno[1,2,3-*cd*]pyrene. (Reproduced with permission of the American Chemical Society from C. Yan, R. Dadoo, H. Zhao, R. Zare, D. Rakestraw, *Anal. Chem.*, **67**(13), 2026–2029 (1995). Copyright 1995 American Chemical Society.)

5 APPLICATIONS

5.1 Amino Acids, Peptides, and Proteins

Most commonly, AA compounds are derivatized with fluorescent labels to enhance their detection in CE. In combination with LIF, detection levels into the zeptomole range have been achieved. The derivatization of charged AA functionalities alters the pK_a values and concomitantly electrophoretic mobilities; hence separation selectivity and resolution are affected. The majority of AA derivatization agents employed for detection in CE were developed originally for detection in LC.[184] These include 3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde (CBQCA), 4-(dimethylaminoazo)benzene-4′-sulfonyl (DABSYL) chloride, 9-fluorenylchloroformate (FMOC), PTH, NDA, OPA, FITC, dansyl chloride, and fluorescamine. In addition to FSCE, MEKC[185] and chiral CE[186] have been applied to the separation of AAs. Figure 14 shows an electropherogram of the MEKC separation of the 20 AAs labeled with PTH. According to the mode of separation employed, the BGE composition may vary greatly and must be optimized according to pH, percentage of organic modifier, and concentration of buffer additives.

The analysis of peptide mixtures is becoming increasingly important owing to their applications in the pharmaceutical industry and other fields. CE has proven

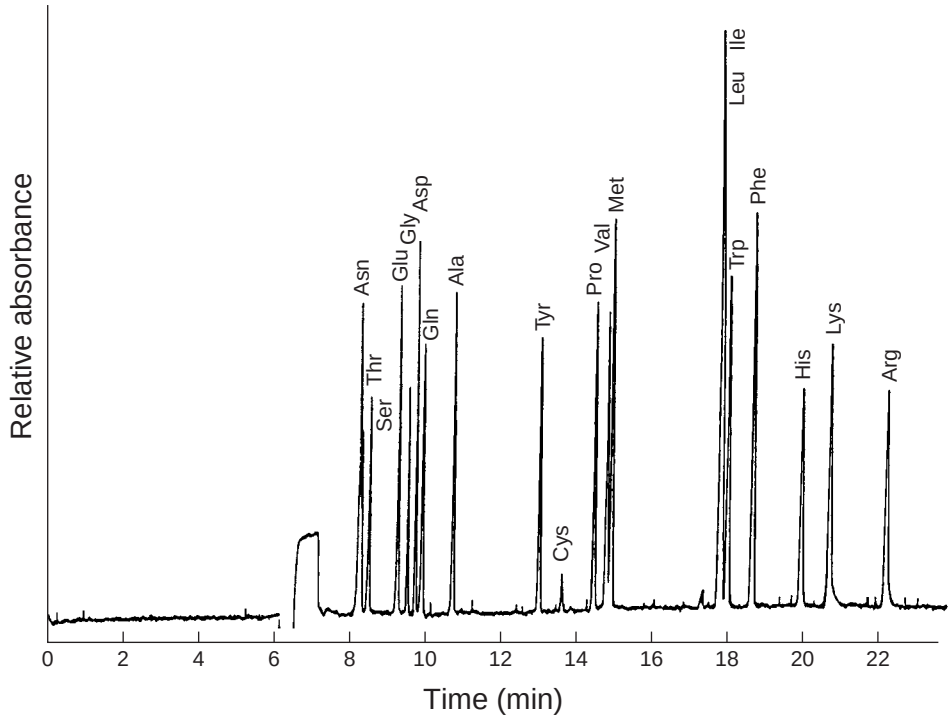

Figure 14 Electropherogram of PTH-labeled AAs with 25 mM phosphate buffer (pH 8.0) containing 50 mM SDS, at 27 kV applied voltage in a fused-silica capillary (70 cm × 50 μm ID). Detection by absorbance at 260 nm. (Reproduced by permission of John Wiley & Sons, Inc. from N.J. Kim, J.H. Kim, K.-J. Lee, *Electrophoresis*, **16**, 510–515 (1995).)

highly useful in the isolation, purification, protein digest mapping, and structural analysis of peptides. Several detection modes have been employed in conjunction with peptide separations in CE, including UV absorbance (both direct and indirect), LIF, ED, CL, radioactivity, and MS detection. Sample stacking, on-line preconcentration, and ITP preconcentration strategies have enhanced signal-to-noise ratios up to 1000-fold.[187]

For peptide separations in CE, temperature, capillary dimensions, applied field strength, and BGE composition should be optimized experimentally. Peptides can be separated according to differences in their terminal and side-chain pK_a values; however, specific pK_a values for peptides are rarely known. Owing to the charged functionalities of their AA constituents, the CE separation of peptides is influenced dramatically by the pH and ionic strength of the BGE. Additionally, the pH and ionic strength of the BGE alter the EOF in the capillary. The application of high ionic strength BGEs has been shown to enhance peptide separations; however, there is a practical limit due to the possibility of excessive Joule heating. Low-pH BGEs should be avoided as such conditions can cause analyte adsorption in uncoated fused-silica capillaries. In addition to the use of coated capillaries, buffer additives, such as metal ions, ion-pairing agents, and surfactants, have been employed to suppress capillary wall interactions.[188] These additives may also influence separation selectivities. The addition of organic solvents such as acetonitrile and aliphatic alcohols to the BGE alters the viscosity of the separation medium, and also the solubility and conformation of the peptides. For the enantiomeric resolution of peptides, chiral selectors such as CDs, bile salts, and chiral crown ethers have been utilized.[189] The impact of peptide hydrophobicity on separation in CE is unclear. Similarly, insufficient data exist concerning the effect of peptide secondary structure on separation in CE. Nonetheless, the *cis–trans* isomerization of peptides and the dependence of peptide secondary structure on hydrodynamic radius have been investigated.[190]

SGE is used extensively as a tool in protein chemistry. The transfer of this technology to the capillary format has offered several distinct advantages, including more rapid analyses with higher separation efficiencies, but at the same time has introduced new challenges. Interactions of structurally complex proteins with ionized silanol groups at the capillary wall can cause partial or complete denaturation of the proteins, generating spurious peaks and band broadening. Basic proteins are particularly problematic. Thus, regardless of the CE mode employed (FSCE, cIEF, or SSCE), the chemistry of the capillary surface is a critical experimental parameter. Depending on the concentration of the protein and its sample matrix, ITP preconcentration[191] or solid-phase

affinity purification[192] may be necessary. UV absorbance (190–280 nm) of peptide bonds or AA side-chain functionalities is the most common detection mode. LIF detection has been applied to a significantly lesser extent; both covalently bonded and noncovalently bound fluorescent dyes have been applied.[193] MS detection with a TOF analyzer has proven useful in the confirmation of molecular weights of proteins. Both MALDI/MS and ESIMS have been investigated; however, the former cannot be directly interfaced to the CE instrument at this time.

The strategy chosen for capillary surface modification in protein analysis should reflect the properties of the protein including its p*I* and post-translational modification chemistry. In high-pH buffers (above the p*I* for a protein), proteins are negatively charged, causing repulsion at the negatively charged wall. However, the secondary and tertiary structures of proteins are sensitive typically to pH extremes. Also, proteins have less charge diversity under such conditions, reducing the likelihood of their separation. Salts, amines (ranging in size from hydroxylamine[194] to chitosan[195]), surfactants, polyanions, and zwitterions can be added to the BGE to assist in protein separations. These additives compete for negatively charged sites at the capillary wall effectively to eliminate or reverse the direction of EOF. These additives can also form ion pairs with the proteins and alter their electrophoretic mobilities. Neutral polymers such as poly(vinyl alcohol) (PVA),[196] polyethyleneimine (PEI),[197] and cellulose derivatives[198] can be used to coat dynamically the capillary surface and nearly eliminate EOF. Figure 15(a) and (b) demonstrates the effects of a dynamic PVA coating versus a thermally immobilized PVA coating in the CE separation of five proteins. In addition, several chemistries have been proposed for the covalent modification of the capillary wall; the most frequently used covalent wall coating is polyacrylamide.[199] Alternative covalent coatings involve silanization chemistries, or the polymerization of cellulose derivatives, acrylates, and polyimines. Polymeric coatings tend to be more robust with respect to pH.[200]

A multitude of reports have appeared encompassing the separation of several classes of proteins in CE. Only a selected number of protein applications will be addressed here. FSCE has been used to separate charge variants of recombinant DNA-derived proteins,[201] conformational hemoglobin variants,[202] food proteins,[203] antigen–antibody complexes,[204] and microheterogeneous glycoproteins.[205] Lipoproteins may require the addition of surfactants or organic solvents to the BGE to enhance solubility.[206] cIEF has been used to separate monoclonal antibodies,[207] and in conjunction with immunoassays to resolve bound versus unbound antigen–antibody complexes.[208] In SDS/CGE or SDS/DSCE, proteins are denatured with 0.1% SDS and separated according to molecular weight, independent of conformation.[209] Even though this method offers advantages over the traditional sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE) technique, both techniques may be supplanted by MS detection in protein molecular weight determinations.

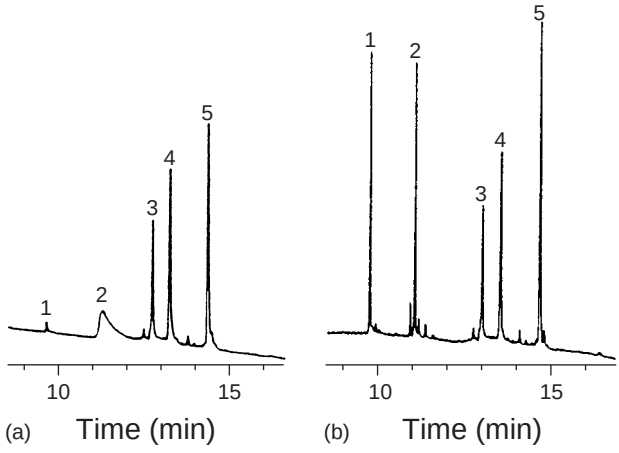

Figure 15 Effect of dynamic and permanent coating of PVA on the separation of proteins using (a) 0.05% (w/w) PVA as buffer additive and (b) thermally immobilized PVA. Sample: (1) cytochrome *c*, (2) lysozyme, (3) trypsin, (4) trypsinogen, and (5) α-chymotrypsinogen. Fused-silica capillary, 50 μm ID, 70 cm total length, 57 cm effective length; buffer, 50 mM sodium phosphate (pH 3.0), temperature, 20 °C; detection, UV absorbance at 214 nm; injection, 10 kV for 5 s; voltage, 30 kV (357 V cm^{-1}). (Reproduced with permission of the American Chemical Society from M. Gilges, M.H. Kleemiss, G. Schomburg, *Anal. Chem.*, **66**(13), 2038–2046 (1994). Copyright 1994 American Chemical Society.)

5.2 Nucleic Acids and Oligonucleotides

DNA molecules are negatively charged at pH values greater than approximately 4 and migrate accordingly in an applied electrical field. DNA separations are performed typically in aqueous TRIS–borate–EDTA buffer solutions, pH 8.3, to maximize solubility. In CE, the detection of DNA molecules can be achieved based on native UV absorbance at 260 nm, or by LIF. The latter enhances detection selectivity and sensitivity, but typically requires the labeling of DNA molecules with fluorescent molecules.

For the analysis of single-stranded DNA (oligonucleotides), a denaturant such as formamide or urea is added to the separation medium to prevent reannealing and secondary structure formation. The use of chemically cross-linked gel matrices in CGE offers single-base resolution,[210] but the use of low-viscosity, replaceable polymeric matrices in DSCE offers greater convenience and can provide comparable resolving power. In general, low molecular weight polymers are employed

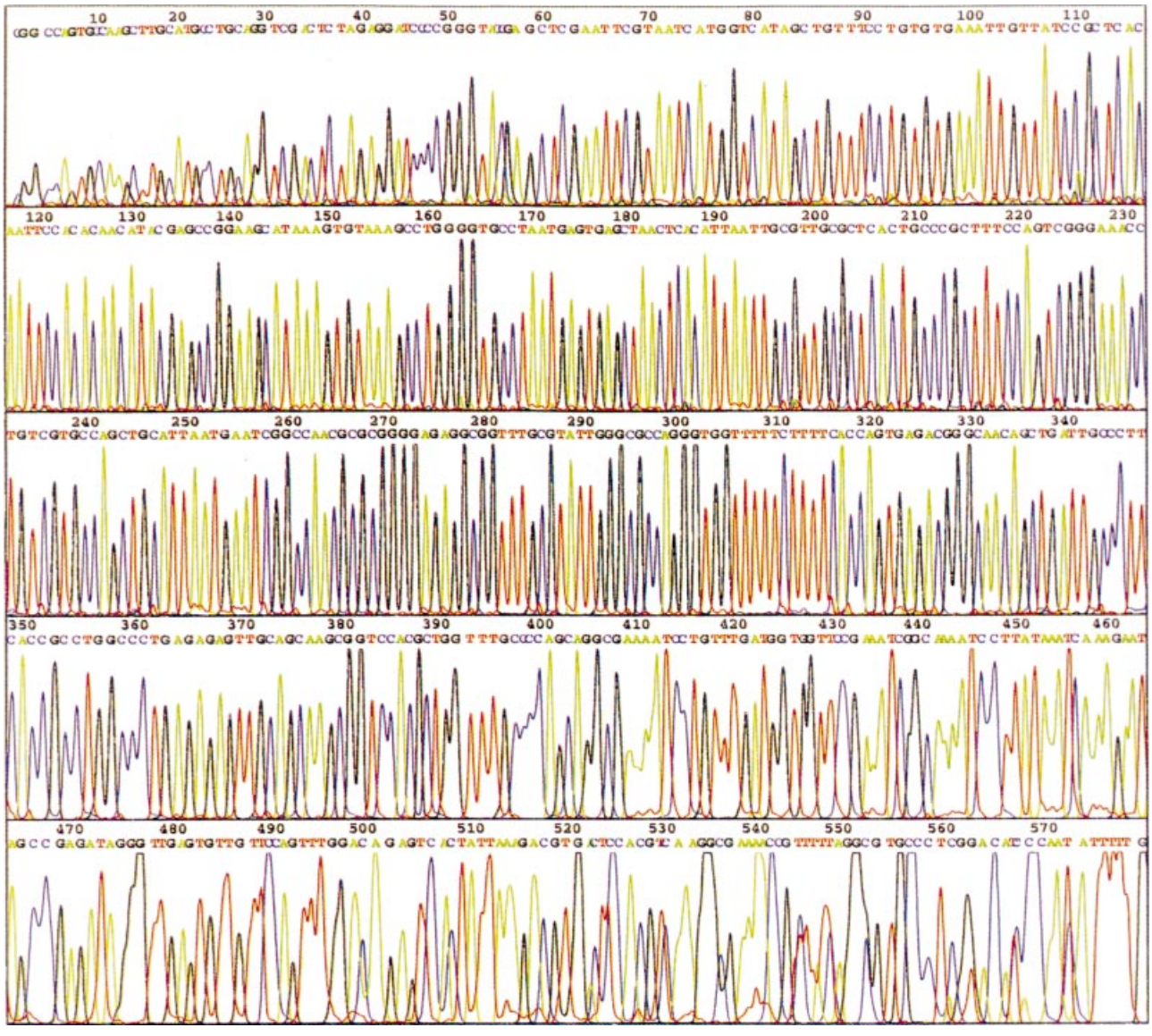

Figure 16 Four-color M13 DNA sequencing traces analyzed from a microchip CE device. Separation was performed on a 7-cm long channel with a 100-μm twin-T injector using 4% LPA as the separation medium at 40 °C. Separation was performed with a voltage of 160 V cm^{-1} and the detector was 6.5 cm from the injector. Only 2 μg of DNA template were employed per reaction, and 1 μL of the final reaction solution (33%) was loaded on the microchip device. The run was complete in 20 min. (Reproduced with permission of the American Chemical Society from S. Liu, Y. Shi, W.W. Ja, R.A. Mathies, *Anal. Chem.*, **71**(3), 566–573 (1999). Copyright 1999 American Chemical Society.)

at high concentrations for the separation of small oligonucleotides. Capillary affinity gel electrophoresis has been applied to the separation of oligonucleotides according to size and base-specific recognition.[211]

DNA sequencing of oligonucleotides is a rapidly growing application in CE.[212–214] The Sanger dideoxy base labeling chemistries, separation mechanisms, and methodologies currently employed in CE originated in SGE. The most desirable methodology involves a four-dye labeling scheme in a single capillary and the application of a ratio-based base-calling algorithm.[215] High ''read lengths'' and fast single-base resolution separations are necessary to attain the throughput necessary to accomplish the goals of the Human Genome Project. Although LPA is the most frequently used separation matrix for DNA sequencing applications, the advantages of PEO with various molecular weight ranges has been demonstrated.[216] Several formats for the multiplexing of capillaries in conjunction with DNA sequencing have been investigated, including capillary arrays of up to 100 capillaries, sheath-flow designs, and micromachined channels. Figure 16 shows four-color

DNA sequencing utilizing DSCE on a microfabricated (microchip) device. Chang and Yeung have implemented coaxial optical fibers for fluorescence excitation with an argon ion laser and detection using a CCD camera.[139] Extension to a greater number of capillaries in an array may require the use of a charge-injection device (CID). The CID is analogous to the CCD, as a solid-state imaging device, but provides more flexibility in detection schemes and less dark current for enhanced detection sensitivity.

The separation and detection of double-stranded DNA fragments (particularly PCR products and restriction fragments) have been facilitated by the incorporation of various fluorescent dyes in the separation matrix, such as oxazole yellow (YO) (commercially available as YO-PRO-1), oxazole yellow homodimer (YOYO), thiazole orange homodimer (TOTO), and ethidium bromide. These molecules intercalate between DNA base pairs to (i) stiffen the molecules, (ii) increase their positive charge, and (iii) increase their separation resolution.[217] For the separation of DNA fragments between 1000 base pairs and 1.6 megabase pairs in length, the applied field strength can be alternated in its direction and frequency as a function of time.[218] This periodic variation of the applied field strength, or field inversion, forces the DNA molecules to reorient continuously as they migrate through the capillary; the extent of this effect is a function of the DNA fragment length and the applied electrical field. Figure 17(a) and (b) shows the dramatic enhancement of the separation of the 1-kbp DNA ladder standard using field inversion versus the use of constant field in DSCE. Using video microscopy to monitor HEC size-sieving solutions, Morris et al. found that DNA molecules alternate between compact and extended conformations as the field strength is cycled.[219]

More recently, DSCE has been applied to the separation of DNA plasmid molecules: naturally occurring circular double-stranded DNA molecules that replicate independently of the genome.[220–223] Plasmids may adopt several conformations, including positively or negatively supercoiled, open-circular (either the relaxed form or the nicked form), and linear conformations. The electrophoretic mobility of DNA plasmids is highly dependent on the separation conditions employed, such as the BGE and the applied field strength. In addition, the majority of plasmid DNA separations have been performed in the presence of DNA intercalating dyes. As a result, differences in the migration order of conformers have been reported. DNA intercalators can provide high sensitivity in conjunction with LIF detection and enhanced electrophoretic separations of double-stranded DNA fragments. However, their presence may induce undesirable structural changes in plasmid DNA molecules, including stretching and the partial unwinding

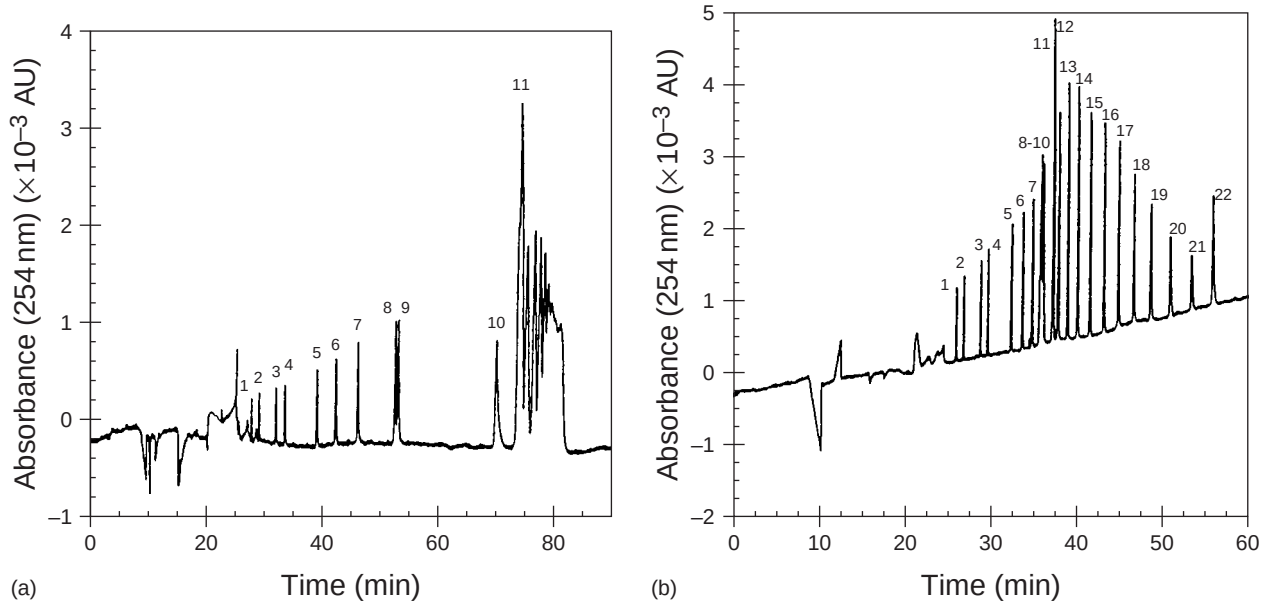

Figure 17 Comparison of field inversion CE and constant field separation. (a) Separation of the 1-kbp DNA ladder at a constant field of 81 V cm^{-1}. (b) Separation of the same sample with field inversion at 50 Hz with $E_f = 243$ V cm^{-1} and $E_b = 81$ V cm^{-1}. Conditions: 1 × TBE buffer containing 1% hydroxypropylcellulose, 10 μM ethidium bromide in a 100-μm ID coated capillary at 25 °C. Peak identification of DNA fragments in base pairs: (1) 134; (2) 154; (3) 201; (4) 220; (5) 298; (6) 344; (7) 396; (8) 506; (9) 517; (10) 1018; (11) 1636; (12) 2036; (13) 3054; (14) 4072; (15) 5090; (16) 6108; (17) 7126; (18) 8144; (19) 9162; (20) 10180; (21) 11198; (22) 12216. (Reproduced by permission of John Wiley & Sons, Inc. from C. Heller, C. Pakleza, J.L. Viovy, *Electrophoresis*, **16**, 1423–1428 (1995).)

of the DNA double helix, and also the reduction or reversal of the supercoiling orientation.

5.3 Carbohydrates

Carbohydrates are ubiquitous to numerous mechanisms in cellular biology; their structural diversity, including branching and optical activity, reflects their unique functions. FSCE, CGE, DSCE, MEKC, CEC, and chiral CE have been applied to the analysis of carbohydrates. Their separation and detection represent a challenge in CE owing to their charge neutrality (in some cases) and lack of absorption of UV or VIS light. Therefore, as analytes in CE, they often require chemical derivatization for the introduction of detectable chromophores and charged functionalities. Additionally, carbohydrates may vary widely in structural complexity, from simple mono-, oligo-, and polysaccharides, to glycopeptides, glycoproteins, and glycolipid oligosaccharides. Sample isolation and preparation techniques should be chosen judiciously as they may influence the CE separation.

The use of alkaline borate-containing BGEs enhances the UV absorbance detection of carbohydrates at 195 nm, and simultaneously introduces negative charges to the molecules through complexation. For AA-containing carbohydrates such as glycoproteins and glycopeptides, direct UV absorbance detection may be employed. LIF detection has proven to be a valuable tool in the analysis of fluorescently labeled carbohydrates, achieving single-molecule detection levels in some instances.[224] The inherently poor sensitivity characteristic of indirect detection for carbohydrates, either UV/VIS absorbance or fluorescence, can be enhanced significantly by the use of a laser light source. Both helium–cadmium lasers (325 or 442 nm) and argon ion lasers (488 nm) have been utilized for this purpose.[225] Both electrochemical[226] and RI detection[227] of carbohydrates have been demonstrated with femtomole detection levels and linearity over three orders of magnitude of concentration. MS detection with ESI sample introduction and various mass analyzers has allowed the simultaneous structural characterization of carbohydrates.[228] MALDI/MS in conjunction with fraction isolation by CE has proven useful in the elucidation of carbohydrate branching and molecular weight;[229] future applications may involve the direct coupling of MALDI/MS with CE.

In order to enhance detection sensitivity, and in some instances improve separation selectivity, carbohydrates have been derivatized chemically. Owing to varied and/or incomplete reactivities of native hydroxyl functionalities, reductive amination followed by reaction with an amine-selective reagent, such as 2-aminopyridine, generates negatively charged analytes in neutral and alkaline pH buffers with a suitable chromophore.[230] Unfortunately, contaminating amines can be derivatized, interfering with detection. The fluorescent label 8-aminonaphthalene-1,3,6-trisulfonic acid reacts with carbohydrates by Shiff base formation.[231] Figure 18(a–c) demonstrates the effects of various fluorescent labels on the separation of a dextran sample in FSCE. CBQCA has been used in conjunction with LIF detection and PAG-filled capillaries to achieve subattomole detection levels.[232]

Being either neutral or negatively charged, the adsorption of carbohydrates to the capillary wall is less problematic than with other macromolecules such as proteins. However, capillary surface modification has been shown to increase the resolution of carbohydrates.[230] Buffer additives such as polyamines, denaturants such as urea, or ion-pairing agents such as tetrabutylammonium phosphate[233] can invoke secondary equilibria effects and consequently alter the electrophoretic mobilities of the

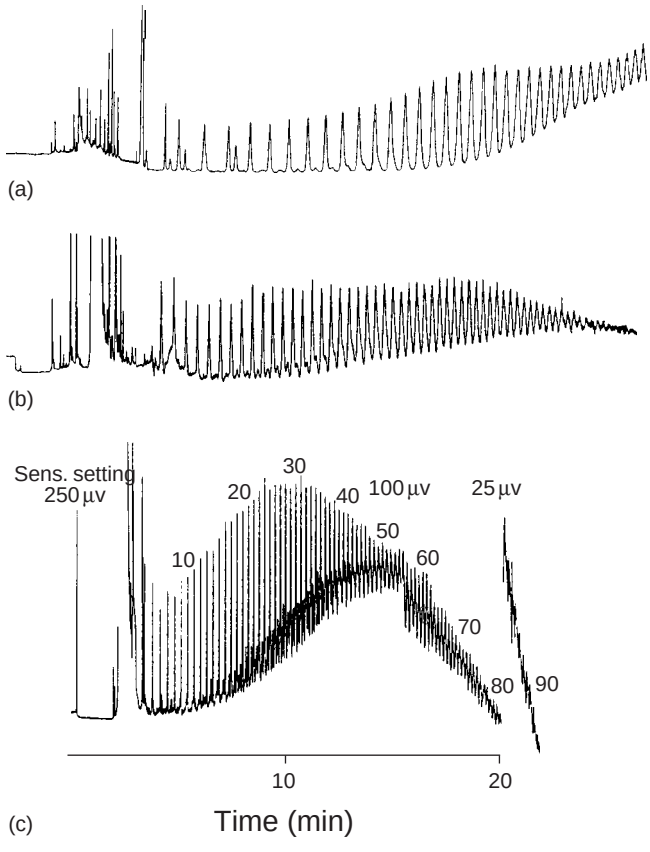

Figure 18 Influence of the fluorescent tag on the separation of a dextran standard with an average molecular weight of 18 300. The reagents used were (a) 2-aminopyridine, (b) 5-aminonaphthalene-2-sulfonate, and (c) 8-aminonaphthalene-1,3,6-trisulfonate. Conditions: voltage, $-500\,V\,cm^{-1}$ (10 µA); buffer, 0.1 M borate–TRIS at pH 8.65; capillary effective length, 35 cm. (Reproduced with permission of the American Chemical Society from M. Stefansson, M. Novotny, *Anal. Chem.*, **66**(7), 1134–1140 (1994). Copyright 1994 American Chemical Society.)

carbohydrate molecules. Ionic surfactants can bind to chemically derivatized celluloses,[231] and highly charged heparins can be partially neutralized by ion pairing with polycationic additives.[234] CDs can be used as buffer additives in CE to disrupt the undesirable formation of carbohydrate micelles in solution.[235]

Glycosoaminoglycan (GAG) and carboxylated or sulfated polysaccharide compounds are less difficult to separate and detect in CE owing to their inherent negatively charged sulfate groups and their absorption of UV light at 232 nm. The resolution and efficiency of these carbohydrates in FSCE and MEKC can be enhanced by suppressing EOF in the capillary with the use of an acidic BGE. In the size-sieving mode, DSCE is preferred to CGE owing to the large size of the carbohydrate molecules.[234] A highly viscous polyacrylamide solution was especially useful for the separation of the negatively charged, highly polydispersed polysaccharide hyaluronic acid.[236] Surfactant-containing BGEs can be applied in MEKC for the simultaneous separation of sulfated and nonsulfated GAGs.

The mapping, sequencing, and isomeric resolution of glycoprotein glycans have received attention recently. Varied substitution of oligosaccharide chains on the same polypeptide results in different glycoforms. Two- and three-dimensional CE separations may be necessary to resolve glycoprotein glycans.[237] Their sialic acid functionalities should be removed enzymatically prior to separation as these charged functionalities may be lost randomly during derivatization.[238] Peptides and proteins should be removed using alcohol precipitation or chromatographic techniques, or by the incorporation of surfactants in the BGE to simplify electropherograms.

5.4 Chiral Separations

Chiral enantiomers (racemates) exist as a mixture of compounds similar in structure but potentially dissimilar in function. Enzymatic reactions that occur in biological systems are stereospecific; a particular enantiomer may be preferred for maximum efficacy. Another enantiomer may be inactive or even highly toxic. Enantiomers of a chiral molecule display optical activity, differing only in the direction in which they rotate plane polarized light. In general, n chiral centers generate 2^n enantiomers. From a separation science standpoint, the resolution of chiral enantiomers can present a formidable challenge. In the past several decades, GC and, more commonly, LC have been applied to the separation of enantiomers. Since Zare et al. demonstrated the first chiral CE separation in 1985,[239] numerous reports of chiral separations have appeared in the CE literature. Chiral CE presents several advantages over chiral LC, including versatility in method development, reduced quantities of buffer volumes, chiral selectors and samples, and higher separation efficiencies. Additionally, CE may be used to measure chiral complexation constants and probe new chiral selectors for subsequent use in chromatographic applications.

Several classes of chiral selectors have been explored in CE; many of these chiral selectors were developed initially for chromatographic separation regimes. Chiral selectors differ in mechanism, hence they differ in enantioselectivity for various classes of compounds. Differential complexation of each enantiomer with a chiral selector that is dissolved in the BGE, or immobilized in a gel network or on a chromatographic support, is responsible for enantiomeric resolution. Key characteristics of CE chiral selectors include their stability and solubility in predominantly aqueous-based BGEs, low absorbance in the UV/VIS range, and multiple and varied chemical functionalities for a minimum of three noncovalent, stereospecific interactions with racemates.

Investigations have been performed in order to optimize the composition of the BGE for chiral separations. In general, the higher the binding constant between a chiral selector and an enantiomer, the lower is its required concentration in the separation system. Obviously, the pH of the BGE influences the degree of ionization of the analytes and the chiral selector, hence, the pH of the BGE influences enantioselectivity. The ionic strength of the BGE influences both the EOF and electrophoretic mobilities via the zeta potential at the capillary wall. If the ionic strength does not exceed levels that can cause additional Joule heating, sample stacking may be utilized. The presence of organic solvents in the BGE can alter binding constants and CMCs, extending the usefulness of chiral CE to more hydrophobic racemates.[240] Predominantly non-aqueous BGEs have also been used; the separation mechanisms between chiral selectors and racemates in nonaqueous buffer systems are not as well understood as in aqueous-based systems.[157] Organic cationic buffer reagents and cellulose-based buffer additives have been added to the BGE in order to reverse or suppress the EOF, thereby increasing enantiomeric resolution.

CDs and their derivatives are the most widely applicable class of chiral selectors. CDs are cyclic oligosaccharides with D-glucose units linked via α-(1,4′)-glycosidic bonds; α-, β-, and γ-CDs contain six, seven, and eight D-glucose units, respectively. The interior portion of each truncated-cone molecule is hydrophobic in nature, allowing the hydrophobic inclusion of enantiomers. Hydroxyl groups on the rim of each structure are polar in nature, promoting hydrogen bonding and dipolar interactions with enantiomers. The structure of β-CD is shown in Figure 19(a). Native CDs[241] or CDs derivatized with charged carboxyl, succinyl, and sulfate functionalities[242] have been applied. The use of derivatized CDs allows the

simultaneous separation of neutral and charged species, enhances resolution and selectivity, and alters the migration order of analytes. Electrostatic interactions may be introduced by changing the degree of ionization of the derivatized CD substituents according to the pH of the BGE. Similarly, both charged polysaccharides[243] (heparin, sulfated dextran, and sulfated chrondroitin) and neutral polysaccharides[244] (maltodextrins, maltooligosaccharides, dextrins, and dextrans) have been employed as chiral selectors in CE, albeit the binding constants of polysaccharides with racemates can be 1000-fold lower than for CDs. The macrocyclic chiral crown ether 18-crown-6-tetracarboxylic acid has been used to resolve organic and inorganic cationic enantiomers.[245] This polyether contains six oxygen atoms inside the ring for hydrogen bonding and dipolar interactions with racemates.

Armstrong et al. first demonstrated the use of macrocyclic antibiotics for chiral separations in CE.[246] The ''basket-like'' morphologies of these chiral selectors and their multitude of functionalities and stereogenic centers are responsible for their unprecedented enantioselectivities reported for hundreds of ionic and neutral racemates. Owing to their structural complexity, the optimization of experimental conditions using macrocyclic antibiotics involves a complex function of pH, ionic strength, and organic modifier content of the BGE. The chemical structure of the commonly employed multicyclic antibiotic chiral selector rifamycin B is shown in Figure 19(b). Recently, the macrocyclic alkaloid D-(+)-tubocurarine has been characterized and applied to the chiral separation of organic carboxylates.[247]

Proteins such as α_1-acid glycoprotein, bovine serum albumin (BSA), and orosomucoid have been used as chiral selectors in FSCE,[248] immobilized in a gel matrix, or covalently bonded to a solid support in CEC.[180] The secondary and tertiary structures of proteins influence their enantioselectivity, restricting the pH, ionic strength, and organic modifier content of the BGE that may be employed. Additionally, proteins can be problematic as chiral selectors owing to their strong noncovalent interactions with certain solutes and the capillary wall (reducing separation efficiencies), and also their background absorbance of UV light.

Finally, the combination of chiral CE and MEKC can generate a powerful separation tool. In chiral MEKC, natural or synthetic chiral surfactants are added to the BGE at concentrations above their CMCs. Natural, hydroxylated bile salts such as sodium cholate (Figure 19c) and its derivatives are ionized at relatively low pH, and are particularly useful for the resolution of aromatic racemates containing basic functionalities.[249] Synthetic chiral surfactants such as *N*-dodecoxycarbonylvaline (Figure 19d), with a polar AA ''head'' group and a hydrophobic ''tail'' group, are particularly useful for the separation of relatively hydrophobic enantiomers.[250] Polymer surfactants do not possess a CMC; therefore, they may be used in conjunction with higher percentages of organic modifier in the BGE, especially for hydrophobic enantiomers.

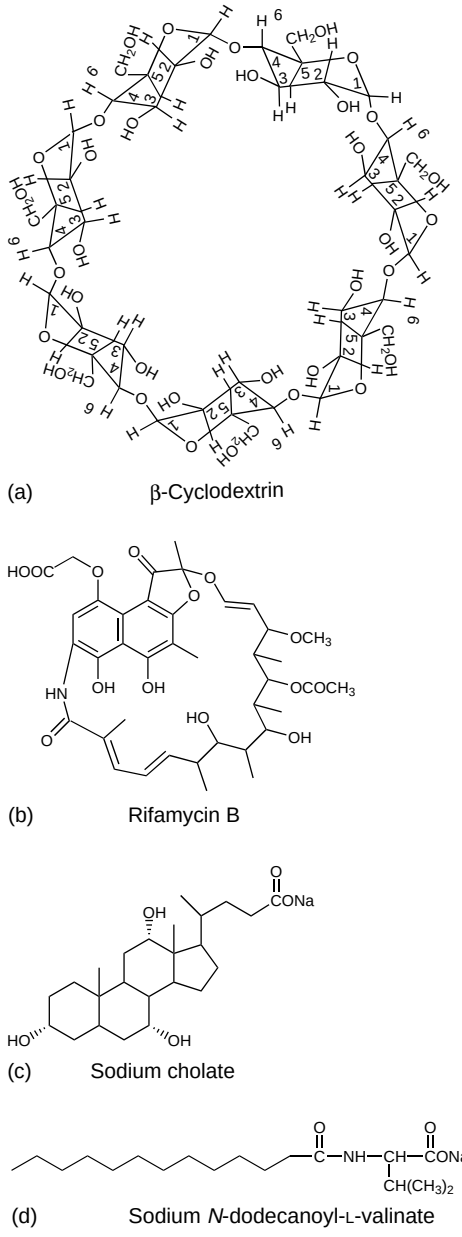

Figure 19 Selected chiral mobile phase additives. (Reproduced with permission of the American Chemical Society from T.J. Ward, *Anal. Chem.*, **66**(11), 634A–640A (1994). Copyright 1994 American Chemical Society.)

5.5 Inorganic Ions

Traditionally, ion-exchange chromatography, ion-selective electrodes, and atomic absorption spectroscopy have

been applied to the analysis of inorganic ions. More recently, FSCE with indirect UV/VIS absorbance detection has become an alternative method in the analysis of inorganic ions. Applications in the analysis of ions by indirect UV/VIS absorbance detection span several areas, including the pharmaceutical industry,[251] environmental samples,[252] and the paper industry.[253] A chromophore with the same charge and comparable electrophoretic mobility to the analyte ions is added to the BGE. Ideally, the analyte ions should exhibit minimal absorbance at the detection wavelength. The limit of detection is independent of chromophore concentration; however, chromophore concentrations of 1–10 mM are employed commonly to achieve linearity over three orders of magnitude of concentration.[254] The EOF should be in the same direction as the electrophoretic mobilities of the analytes of interest in order to reduce the time of analysis and enhance migration time reproducibilities.[255] Therefore, in the FSCE analysis of anions, a positively charged surfactant is added to the BGE in order to reverse the EOF. Analyte selectivity is influenced by the chemical properties of the surfactant and its concentration.[256]

Chromate is the most commonly employed chromophore, exhibiting absorbance from 200 to 400 nm. For enhanced detection sensitivity, the chromophore pyromellitic acid has been employed.[257] Other chromophores include naphthalene sulfates,[258] benzoate,[259] and ribonucleotides for polyphosphonates and polyphosphates.[260] Tetradecyltrimethylammonium bromide (TTAB) is the most commonly employed EOF reversal agent, but demonstrates limited detectability for bromide ion. Therefore, for samples containing bromide ion, tetradecyltrimethylammonium hydroxide has been employed.[261] The capillary walls may be dynamically or covalently coated to reverse the EOF.[262,263] Additionally, one or more co-ions may be added to the BGE for buffering purposes and to promote sample stacking.[264] Some inorganic anions such as chloride, bromide, iodide, chromate, thiocyanate, molybdate, tungstate, arsenate, nitrate, and nitrite are detectable in the low-UV region from 180 to 200 nm.[265] The direct UV/VIS absorbance detection of inorganic anions must involve a non-absorbing co-ion. For both indirect and direct UV/VIS absorbance detection of anions, limits of detection are approximately in the micrograms per liter range.

For the analysis of inorganic cations, the addition of EOF reversal agents to the BGE is not necessary. However, anionic and neutral complexing agents such as EDTA, glycolic acid, tartaric acid, citric acid, lactic acid, succinic acid, and 18-crown-6 can be added to alter analyte selectivities. Typically, the concentrations of these agents are a function of their electrophoretic mobilities and complexation constants. Imidazole is the most universal chromophore for cation analysis.[266] Direct

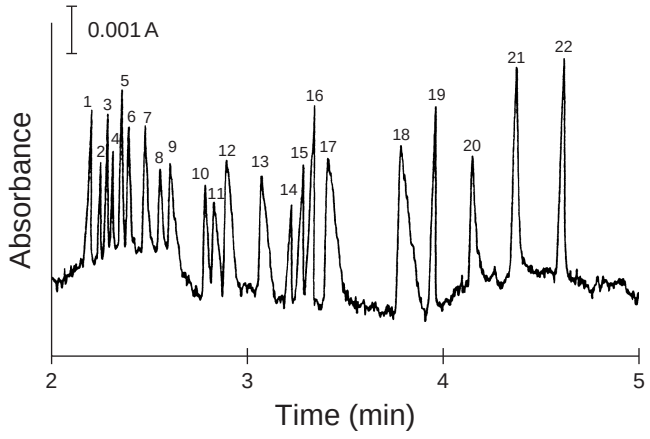

Figure 20 Electrophoretic separation of 22 anions and cations. Peak identification: (1) $S_2O_3^{2-}$; (2) Br^-; (3) Cl^-; (4) SO_4^{2-}; (5) NO_2^-; (6) NO_3^-; (7) WO_4^{2-}; (8) MoO_4^{2-}; (9) citrate; (10) maleate; (11) fumarate; (12) F^-; (13) HPO_4^{2-}; (14) Cs^+; (15) K^+; (16) NH_4^+; (17) HCO_3^-; (18) acetate; (19) Na^+; (20) Ca^{2+}; (21) Mg^{2+}; (22) Li^+. Conditions: indirect absorbance detection at 262 nm with peak reversal; time between injections, 80 s. (Reproduced with permission of the American Chemical Society from P. Kuban, B. Karlberg, *Anal. Chem.*, **70**(2), 360–365 (1998). Copyright 1998 American Chemical Society.)

UV/VIS absorbance detection may be used for cations after complexation with 8-hydroxyquinoline-5-sulfonic acid.[267] Quang and Khaledi developed a computer model to facilitate the multiparameter optimization of separations involving complexation agents. Their model could be used to simulate and predict the electrophoretic mobility of metal cations as a function of the pH and composition of the BGE.[268]

Anions and cations have been detected simultaneously using 4-aminopyridine as the chromophore; anions and cations approach a centrally located detection window from opposite directions of the same capillary (Figure 20).[269] In addition to direct and indirect UV/VIS absorbance, indirect fluorescence,[270] MS,[271] electrochemical,[272] and suppressed or non-suppressed conductivity[273] detection have been applied to the analysis of inorganic ions in CE. For conductivity detection, a tradeoff exists between sensitivity and symmetrical peak shape; distinct differences in conductivity but similar electrophoretic mobilities must exist for the analytes relative to the co-ion. Conductivity detection of inorganic ions can achieve relatively lower detection limits (nanograms per liter) in conjunction with sample stacking.

ABBREVIATIONS AND ACRONYMS

AA	Amino Acid
BGE	Background Electrolyte

BSA	Bovine Serum Albumin
CBQCA	3-(4-Carboxybenzoyl)-2-quinolinecarboxaldehyde
CCD	Charge-coupled Device
CD	Cyclodextrin
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CE/ED	Capillary Electrophoresis/Electrochemical Detection
CE/MS	Capillary Electrophoresis/Mass Spectrometry
CFFAB	Continuous-flow Fast Atom Bombardment
CGE	Capillary Gel Electrophoresis
CID	Charge-injection Device
cIEF	Capillary Isoelectric Focusing
CL	Chemiluminescence
CMC	Critical Micelle Concentration
CTAB	Cetyltrimethylammonium Bromide
CZE	Capillary Zone Electrophoresis
DABSYL	4-(Dimethylaminoazo)benzene-4′-sulfonyl
DSCE	Dynamic Size-sieving Capillary Electrophoresis
ED	Electrochemical Detection
EOF	Electro-osmotic Flow
ESI	Electrospray Ionization
ESIMS	Electrospray Ionization Mass Spectrometry
FAB	Fast Atom Bombardment
FITC	Fluorescein Isothiocyanate
FMOC	9-Fluorenylchloroformate
FSCE	Free-solution Capillary Electrophoresis
FT	Fourier Transform
GAG	Glycosoaminoglycan
GC	Gas Chromatography
HEC	Hydroxyethylcellulose
HPMC	Hydroxypropylmethylcellulose
ID	Inner Diameter
ITP	Isotachophoresis
LC	Liquid Chromatography
LIF	Laser-induced Fluorescence
LPA	Linear Polyacrylamide
MALDI	Matrix-assisted Laser Desorption/Ionization
MC	Methylcellulose
MEKC	Micellar Electrokinetic Chromatography
MS	Mass Spectrometry
NACE	Nonaqueous Capillary Electrophoresis
NDA	Naphthalenedialdehyde
NMR	Nuclear Magnetic Resonance
OPA	*o*-Phthalaldehyde
OTCEC	Open-tubular Capillary Electrochromatography
PAG	Polyacrylamide Gel
PAH	Polycyclic Aromatic Hydrocarbon
PCR	Polymerase Chain Reaction
PDA	Photodiode Array
PDMS	Poly(dimethylsiloxane)
PEI	Polyethyleneimine
PEO	Poly(ethylene oxide)
PTH	Phenylthiohydantoin
PVA	Poly(vinyl alcohol)
PVP	Polyvinylpyrrolidone
RI	Refractive Index
RSD	Relative Standard Deviation
SDS	Sodium Dodecyl Sulfate
SDS/PAGE	Sodium Dodecyl Sulfate/Polyacrylamide Gel Electrophoresis
SGE	Slab Gel Electrophoresis
SIM	Selected-ion Monitoring
SSCE	Size-sieving Capillary Electrophoresis
TOF	Time-of-flight
TOTO	Thiazole Orange Homodimer
TTAB	Tetradecyltrimethylammonium Bromide
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIS	Visible
YO	Oxazole Yellow
YOYO	Oxazole Yellow Homodimer

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 1)**
Capillary Electrophoresis in Detection of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Capillary Electrophoresis in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis • Cyanogen Chloride and Cyanogen Bromide Analysis in Drinking Water • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis

Environment: Water and Waste cont'd **(Volume 4)**
Organic Analysis in Environmental Samples by Capillary Electrophoresis

Nucleic Acids Structure and Mapping **(Volume 6)**
Capillary Electrophoresis of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis in Peptide and Protein Analysis, Detection Modes for • Capillary Electrophoresis of Peptides • Capillary Electrophoresis of Proteins and Glycoproteins • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Pulp and Paper **(Volume 9)**
Carbohydrates from Chemical Pulps: Characterization by Capillary Zone Electrophoresis

Liquid Chromatography **(Volume 13)**
Micellar Electrokinetic Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography

REFERENCES

1. S. Hjerten, 'Free Zone Electrophoresis', *Chromatogr. Rev.*, **9**, 122–219 (1967).
2. J.W. Jorgenson, K.D. Lukacs, 'Zone Electrophoresis in Open-tubular Glass Capillaries', *Anal. Chem.*, **53**(8), 1298–1302 (1981).
3. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open-tubular Capillaries', *Anal. Chem.*, **56**(1), 111–113 (1984).
4. D.J. Harrison, A. Manz, Z. Fan, H. Ludi, H.M. Widmer, 'Capillary Electrophoresis and Sample Injection Systems Integrated on a Planar Glass Chip', *Anal. Chem.*, **64**(17), 1926–1932 (1992).
5. J.H. Knox, I.H. Grant, 'Miniaturisation in Pressure and Electroendosmotically Driven Liquid Chromatography: Some Theoretical Considerations', *Chromatographia*, **24**, 135–143 (1987).
6. R. Weber, 'Concerning Theories of Electrophoretic Separations in Porous Tubes', *Helv. Chim. Acta*, **36**, 424–434 (1953).
7. J.C. Giddings, 'Generation of Variance, Theoretical Plates, Resolution, and Peak Capacity in Electrophoresis and Sedimentation', *Sep. Sci.*, **4**, 181–189 (1969).
8. F. Foret, L. Krivankova, P. Bocek, *Capillary Zone Electrophoresis*, VCH, Weinheim, 1993.
9. A. Cifuentes, W.T. Kok, H. Poppe, 'Capillary Electrophoresis using Air and Helium as Cooling Fluids', *J. Microcol. Sep.*, **7**(4), 365–374 (1995).
10. E. Grushka, R.M. McCormick, J.J. Kirkland, 'Effect of Temperature Gradients on the Efficiency of Capillary Zone Electrophoresis Separations', *Anal. Chem.*, **61**, 241–246 (1989).
11. F.E.P. Mikkers, F.M. Everaerts, Th.P.E.M. Verheggen, 'Concentration Distributions in Free Zone Electrophoresis', *J. Chromatogr.*, **169**(1), 1–10 (1979).
12. J.C. Sternberg, 'Extracolumn Contributions to Chromatographic Band Broadening', *Adv. Chromatogr.*, **2**, 205–269 (1996).
13. S.V. Ermakov, M.Y. Zhukov, L. Capelli, P.G. Righetti, 'Wall Adsorption in Capillary Electrophoresis: Experimental Study and Computer Simulation', *J. Chromatogr.*, **699**(1–2), 297–313 (1995).
14. D. McManigill, S.A. Swedberg, 'Factors Affecting Plate Height in High Performance Zonal Capillary Electrophoresis (HPZCE)', in *Techniques in Protein Chemistry*, ed. T. Hugli, Academic Press, San Diego, CA, 468, 1989.
15. P. Gebauer, P. Bocek, 'Predicting Peak Symmetry in Capillary Zone Electrophoresis: the Concept of the Peak Shape Diagram', *Anal. Chem.*, **69**(8), 1557–1563 (1997).
16. V. Kasicka, Z. Prusik, B. Gas, M. Stedry, 'Contribution of Capillary Coiling to Zone Dispersion in Capillary Zone Electrophoresis', *Electrophoresis*, **16**(11), 2034–2038 (1995).
17. A. Guttman, J. Horvath, N. Cooke, 'Influence of Temperature on the Sieving Effect of Different Polymer Matrices in Capillary SDS Gel Electrophoresis of Proteins', *Anal. Chem.*, **65**(3), 199–203 (1993).
18. J. Sudor, M.V. Notovny, 'Pulsed-field Capillary Electrophoresis: Optimizing Separation Parameters with Model Mixtures of Sulfonated Polystyrenes', *Anal. Chem.*, **66**, 2139–2147 (1994).
19. K.D. Lukacs, J.W. Jorgenson, 'Capillary Zone Electrophoresis: Effect of Physical Parameters on Separation Efficiency and Quantitation', *J. High Resolut. Chromatogr.*, **8**, 407–411 (1985).
20. T. Tsuda, J.V. Sweedler, R.N. Zare, 'Rectangular Capillaries for Capillary Zone Electrophoresis', *Anal. Chem.*, **62**(19), 2149–2152 (1990).
21. Z. Liu, P. Sam, S.R. Sirimanne, P.C. McClure, J. Grainger, D.G. Patterson, Jr, 'Field-amplified Sample Stacking in Micellar Electrokinetic Chromatography for On-column Sample Concentration of Neutral Molecules', *J. Chromatogr. A*, **673**(1), 125–132 (1994).
22. M.W.F. Nielen, 'Trace Enrichment of Environmental Samples in Capillary Zone Electrophoresis', *Trends Anal. Chem.*, **12**(9), 345–356 (1993).
23. S.M. Wolf, P. Vouros, 'Incorporation of Sample Stacking Techniques into the Capillary Electrophoresis CF-FAB Mass Spectrometric Analysis of DNA Adducts', *Anal. Chem.*, **67**(5), 891–900 (1995).
24. T. Hirokawa, A. Ohmori, Y. Kiso, 'Analysis of a Dilute Sample by Capillary Zone Electrophoresis with Isotachophoretic Preconcentration', *J. Chromatogr.*, **634**(1), 101–106 (1993).
25. T. Buzinkaiova, J. Sadecka, J. Polonsky, E. Vlasicova, V. Korinkova, 'Isotachophoretic Analysis of Some Antidepressants', *J. Chromatogr.*, **638**(2), 231–234 (1993).

26. G. Schmitz, C. Mollers, 'Analysis of Lipoproteins with Analytical Capillary Isotachophoresis', *Electrophoresis*, **15**(1), 31–39 (1994).
27. P. Praus, 'Possibilities of Simple Steam Distillation in Combination with Capillary Isotachophoresis for the Determination of Chlorophenols in River and Industrial Wastewaters', *Anal. Chim. Acta*, **302**(1), 39–44 (1995).
28. S.E. Moring, R.T. Reel, R.E.J. van Soest, 'Optical Improvements of a Z-shaped Cell for High-sensitivity UV Absorbance Detection in Capillary Electrophoresis', *Anal. Chem.*, **65**(23), 3454–3459 (1993).
29. D.H. Heiger, P. Kaltenbach, H.-J.P. Sievert, 'Diode Array Detection in Capillary Electrophoresis', *Electrophoresis*, **15**(10), 1234–1247 (1994).
30. Y. Xue, E.S. Yeung, 'On-column Double-beam Laser Absorption Detection for Capillary Electrophoresis', *Anal. Chem.*, **65**(15), 1988–1993 (1993).
31. C.T. Culbertson, J.W. Jorgenson, 'Lowering the UV Absorbance Detection Limit in Capillary Zone Electrophoresis using a Single Linear Photodiode Array Detector', *Anal. Chem.*, **70**(13), 2629–2638 (1998).
32. Z. Wu, W.G. Tong, 'Sensitive Absorbance Detection Method for Capillary Electrophoresis Based on Laser Wave-mixing', *J. Chromatogr. A*, **773**(1–2), 291–298 (1997).
33. W. Buchberger, S.M. Cousins, P.R. Haddad, 'Optimisation of Indirect UV Detection in Capillary Zone Electrophoresis of Low-molecular-mass Anions', *Trends Anal. Chem.*, **13**(8), 313–319 (1994).
34. D.C. Nguyen, R.A. Keller, J.H. Jett, J.C. Martin, 'Detection of Single Molecules of Phycoerythrin in Hydrodynamically Focused Flows by Laser-induced Fluorescence', *Anal. Chem.*, **59**(17), 2158–2161 (1987).
35. A.T. Timperman, K. Khatib, J.V. Sweedler, 'Wavelength-resolved Fluorescence Detection in Capillary Electrophoresis', *Anal. Chem.*, **67**(1), 139–144 (1995).
36. B. Nickerson, J.W. Jorgenson, 'Characterization of a Post-column Reaction-laser-induced Fluorescence Detector for Capillary Zone Electrophoresis', *J. Chromatogr.*, **480**(1), 157–168 (1989).
37. Y.-F. Cheng, N.J. Dovichi, 'Subattomole Amino Acid Analysis by Capillary Zone Electrophoresis and Laser-induced Fluorescence', *Science*, **242**(4878), 562–564 (1988).
38. L. Hernandez, R. Marquina, J. Escalona, N.A. Guzman, 'Detection and Quantification of Capillary Electrophoresis Zones by Fluorescence Microscopy', *J. Chromatogr.*, **502**(2), 247–255 (1990).
39. S. Carson, A.S. Cohen, A. Belenkii, M.C. Ruiz-Martinez, J. Berka, B.L. Karger, 'DNA Sequencing by Capillary Electrophoresis: use of a Two-laser-window Intensified Diode Array Detection System', *Anal. Chem.*, **65**(22), 3219–3226 (1991).
40. K.J. Miller, F.E. Lytle, 'Capillary Zone Electrophoresis with Time-resolved Fluorescence Detection using a Diode-pumped Solid-state Laser', *J. Chromatogr.*, **648**(1), 245–250 (1993).
41. S.D. Gilman, J.J. Pietron, A.G. Ewing, 'Post-column Derivatization in Narrow-bore Capillaries for the Analysis of Amino Acids and Proteins by Capillary Electrophoresis with Fluorescence Detection', *J. Microcol. Sep.*, **6**(4), 373–384 (1994).
42. R.T. Kennedy, J.W. Jorgenson, 'Quantitative Analysis of Individual Neurons by Open Tubular Liquid Chromatography with Voltammetric Detection', *Anal. Chem.*, **61**(5), 436–441 (1989).
43. M.N. Church, J.D. Spear, R.E. Russo, G.L. Klunder, P.M. Grant, B.D. Andresen, 'Transient Isotachophoretic Separations of Lanthanides with Indirect Laser-induced Fluorescence Detection', *Anal. Chem.*, **70**(13), 2475–2480 (1998).
44. A.G. Ewing, R.A. Wallingford, T.M. Olefirowicz, 'Capillary Electrophoresis', *Anal. Chem.*, **61**(4), 292A–303A (1989).
45. J. Bergquist, A. Tarkowski, R. Ekman, A. Ewing, 'Discovery of Endogenous Catecholamines in Lymphocytes and Evidence for Catecholamine Regulation of Lymphocyte Function via an Autocrine Loop', *Proc. Natl. Acad. Sci. USA*, **91**(26), 12912–12916 (1994).
46. Y.F. Yik, H.K. Lee, S.F.Y. Li, S.B. Khoo, 'Micellar Electrokinetic Chromatography of Vitamin B_6 with Electrochemical Detection', *J. Chromatogr.*, **585**(1), 139–144 (1991).
47. T.J. O'Shea, R.D. Greenhangen, S.M. Lunte, C.E. Lunte, M.R. Smyth, D.M. Radzik, N. Watanabe, 'Capillary Electrophoresis with Electrochemical Detection Employing an On-column Nafion Joint', *J. Chromatogr.*, **593**(1–2), 305–312 (1992).
48. W. Lu, R.M. Cassidy, A.S. Baranski, 'End-column Electrochemical Detection for Inorganic and Organic Species in High-voltage Capillary Electrophoresis', *J. Chromatogr.*, **640**(1–2), 433–440 (1993).
49. S. Sloss, A.G. Ewing, 'Improved Method for End-column Amperometric Detection for Capillary Electrophoresis', *Anal. Chem.*, **65**(5), 577–581 (1993).
50. T.J. O'Shea, S.M. Lunte, 'Chemically Modified Microelectrodes for Capillary Electrophoresis', *Anal. Chem.*, **66**(2), 307–311 (1994).
51. W. Lu, R.M. Cassidy, 'Evaluation of Ultramicroelectrodes for the Detection of Metal Ions Separated by Capillary Electrophoresis', *Anal. Chem.*, **65**(13), 1649–1653 (1993).
52. J. Wen, A. Baranski, R. Cassidy, 'Cyclic Voltammetric Detection in Capillary Electrophoresis with Application to Metal Ions', *Anal. Chem.*, **70**(13), 2504–2509 (1998).
53. T. Kappes, P.C. Hauser, 'Potentiometric Detection of Alkali and Alkaline Earth Metal Cations in Capillary Electrophoresis with Simplified Electrode Alignment and Enhanced Separation and Sensitivity', *Anal. Chem.*, **70**(13), 2487–2492 (1998).

54. C. Haber, R.J. Van Saun, W.R. Jones, 'Quantitative Analysis of Anions at ppb/ppt Levels with Capillary Electrophoresis and Conductivity Detection: Enhancement of System Linearity and Precision using an Internal Standard', *Anal. Chem.*, **70**(11), 2261–2267 (1998).
55. M. Harrold, J. Stillian, L. Bao, R. Rocklin, N. Avdalovic, 'Capillary Electrophoresis of Inorganic Anions and Organic Acids using Suppressed Conductivity Detection: Strategies for Selectivity Control', *J. Chromatogr. A*, **717**(1–2), 371–383 (1995).
56. J.A. Jankowski, S. Tracht, J.V. Sweedler, 'Assaying Single Cells with Capillary Electrophoresis', *Trends Anal. Chem.*, **14**(4), 170–176 (1995).
57. A.G. Ewing, J.M. Mesaros, P.F. Gavin, 'Electrochemical Detection in Microcolumn Separations', *Anal. Chem.*, **66**(9), 527A–537A (1994).
58. M.A. Moseley, J.W. Jorgenson, J. Shabanowitz, D.F. Hunt, K.B. Tomer, 'Optimization of Capillary Zone Electrophoresis/Electrospray Ionization for the Mass Spectrometry and Tandem Mass Spectrometry Analysis of Peptides', *J. Am. Soc. Mass Spectrom.*, **3**(4), 289–300 (1992).
59. E. Kenndler, 'Organic Solvents in Capillary Electrophoresis', in *Capillary Electrophoresis Technology*, ed. N.A. Guzman, Chromatography Science Series, Marcel Dekker, New York, Chapter 5, 161–186, Vol. 64, 1993.
60. M.A. Moseley, L.J. Deterding, K.B. Tomer, J.W. Jorgenson, 'Determination of Bioactive Peptides using Capillary Zone Electrophoresis/Mass Spectrometry', *Anal. Chem.*, **63**(2), 109–114 (1991).
61. R.D. Smith, C.J. Barinaga, H.R. Udseth, 'Improved Electrospray Ionization Interface for Capillary Zone Electrophoresis–Mass Spectrometry', *Anal. Chem.*, **60**(18), 1948–1952 (1988).
62. R.M. Caprioli, T. Fan, J.S. Cottrell, 'Continuous-flow Sample Probe for Fast Atom Bombardment Mass Spectrometry', *Anal. Chem.*, **58**(14), 2949–2954 (1986).
63. M.S. Kriger, K.D. Cook, R.S. Ramsey, 'Durable Gold-coated Fused Silica Capillaries for use in Electrospray Mass Spectrometry', *Anal. Chem.*, **67**(2), 385–389 (1995).
64. E.D. Lee, W. Muck, J.D. Henion, T.R. Covey, 'On-line Capillary Zone Electrophoresis–Ion Spray Tandem Mass Spectrometry for the Determination of Dynorphins', *J. Chromatogr.*, **458**(1), 313–321 (1988).
65. L. Fang, R. Zhang, E.R. Williams, R.N. Zare, 'On-line Time-of-flight Mass Spectrometric Analysis of Peptides By Capillary Electrophoresis', *Anal. Chem.*, **66**(21), 3696–3701 (1994).
66. R.S. Ramsey, S.A. McLuckey, 'Capillary Electrophoresis/Electrospray Ionization Ion Trap Mass Spectrometry using a Sheathless Interface', *J. Microcol. Sep.*, **7**(5), 461–469 (1995).
67. S.A. Hofstadler, J.H. Wahl, R. Bakhtiar, G.A. Anderson, J.E. Bruce, R.D. Smith, 'Capillary Electrophoresis Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Sustained Off-resonance Irradiation for the Characterization of Protein and Peptide Mixtures', *J. Am. Soc. Mass Spectrom.*, **5**(10), 894–899 (1994).
68. K. Tsuji, L. Baczynskyj, G.E. Bronson, 'Capillary Electrophoresis–Electrospray Mass Spectrometry for the Analysis of Recombinant Bovine and Porcine Somatotropins', *Anal. Chem.*, **64**(17), 1864–1870 (1992).
69. I.M. Johansson, R. Pavelka, J.D. Henion, 'Determination of Small Drug Molecules by Capillary Electrophoresis–Atmospheric Pressure Ionization Mass Spectrometry', *J. Chromatogr.*, **559**(1–2), 515–528 (1991).
70. R. Dadoo, A.G. Seto, L.A. Colon, R.N. Zare, 'End-column Chemiluminescence Detector for Capillary Electrophoresis', *Anal. Chem.*, **66**(2), 303–306 (1994).
71. C.Y. Chen, T. Demana, S.D. Huang, M.D. Morris, 'Capillary Zone Electrophoresis with Analyte Velocity Modulation. Application to Refractive Index Detection', *Anal. Chem.*, **61**(14), 1590–1593 (1989).
72. G. Westerberg, H. Lundqvist, F. Kilár, B. Långström, 'β^+-Selective Radiodetector for Capillary Electrophoresis', *J. Chromatogr.*, **645**(2), 319–325 (1993).
73. C.-Y. Chen, M.D. Morris, 'On-line Multichannel Raman Spectroscopic Detection System for Capillary Zone Electrophoresis', *J. Chromatogr.*, **540**(1–2), 355–363 (1991).
74. N. Wu, T.L. Peck, A.G. Webb, R.L. Magin, J.V. Sweedler, 'Nanoliter Volume Sample Cells for ^{1}H NMR: Application to On-line Detection in Capillary Electrophoresis', *J. Am. Chem. Soc.*, **116**(17), 7929–7930 (1994).
75. Z. Rosenzweig, E.S. Yeung, 'Laser-based Particle-counting Microimmunoassay for the Analysis of Single Human Erythrocytes', *Anal. Chem.*, **66**(10), 1771–1776 (1994).
76. S.C. Jacobson, R. Hergenröder, A.W. Moore, Jr, J.M. Ramsey, 'Precolumn Reactions with Electrophoretic Analysis Integrated on a Microchip', *Anal. Chem.*, **66**(23), 4127–4132 (1994).
77. K. Fluri, G. Fitzpatrick, N. Chiem, D.J. Harrison, 'Integrated Capillary Electrophoresis Devices with an Efficient Postcolumn Reactor in Planar Quartz and Glass Chips', *Anal. Chem.*, **68**(23), 4285–4290 (1996).
78. C.S. Effenhauser, G.J.M. Bruin, A. Paulus, M. Ehrat, 'Integrated Capillary Electrophoresis on Flexible Silicone Microdevices: Analysis of DNA Restriction Fragments and Detection of Single DNA Molecules on Microchips', *Anal. Chem.*, **69**(17), 3451–3457 (1993).
79. C.S. Effenhauser, A. Manz, H.M. Widmer, 'Glass Chips for High-speed Capillary Electrophoresis Separations with Submicrometer Plate Heights', *Anal. Chem.*, **65**(19), 2637–2642 (1993).
80. S.C. Jacobson, R. Hergenröder, L.B. Koutny, R.J. Warmack, J.M. Ramsey, 'Effects of Injection Schemes and Column Geometry on the Performance of Microchip Electrophoresis Devices', *Anal. Chem.*, **66**(7), 1107–1113 (1994).

81. S.C. Jacobson, J.M. Ramsey, 'Microchip Electrophoresis with Sample Stacking', *Electrophoresis*, **16**, 481–486 (1995).
82. S.C. Jacobson, J.M. Ramsey, 'Integrated Microdevice for DNA Restriction Fragment Analysis', *Anal. Chem.*, **68**(5), 720–723 (1996).
83. N. Burggaf, A. Manz, E. Verpoorte, C.S. Effenhauser, H.M. Widmer, N.F. de Rooji, 'A Novel Approach to Ion Separations in Solution: Synchronized Cyclic Capillary Electrophoresis', *Sens. Actuators B*, **20**, 103–110 (1994).
84. D.E. Raymond, A. Manz, H.M. Widmer, 'Continuous Sample Pretreatment using a Free-flow Electrophoresis Device Integrated onto a Silicon Chip', *Anal. Chem.*, **66**(18), 2858–2865 (1994).
85. Z. Liang, N. Chiem, G. Ocvirk, T. Tang, K. Fluri, D.J. Harrison, 'Microfabrication of a Planar Absorbance and Fluorescence Cell for Integrated Capillary Electrophoresis Devices', *Anal. Chem.*, **68**(6), 1040–1046 (1996).
86. P.M. Goodwin, W.P. Ambrose, R.A. Keller, 'Single-molecule Detection in Liquids by Laser-induced Fluorescence', *Acc. Chem. Res.*, **29**(12), 607–613 (1996).
87. A.T. Woolley, K. Lao, A.N. Glazer, R.A. Mathies, 'Capillary Electrophoresis Chips with Integrated Electrochemical Detection', *Anal. Chem.*, **70**(4), 684–688 (1998).
88. R.S. Ramsey, J.M. Ramsey, 'Generating Electrospray from Microchip Devices using Electroosmotic Pumping', *Anal. Chem.*, **69**(6), 1174–1178 (1997).
89. J.C. Fister, S.C. Jacobson, L.M. Davis, J.M. Ramsey, 'Counting Single Chromophore Molecules for Ultrasensitive Analysis and Separations on Microchip Devices', *Anal. Chem.*, **70**, 431–437 (1998).
90. A.W. Moore, Jr, S.C. Jacobson, J.M. Ramsey, 'Microchip Separations of Neutral Species via Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **67**(22), 4184–4189 (1995).
91. S.C. Jacobson, R. Hergenröder, L.B. Koutny, J.M. Ramsey, 'Open Channel Electrochromatography on a Microchip', *Anal. Chem.*, **66**(14), 2369–2373 (1994).
92. A.T. Woolley, R.A. Mathies, 'Ultra-high-speed DNA Sequencing using Capillary Electrophoresis Chips', *Anal. Chem.*, **67**(20), 3676–3680 (1995).
93. J. Cheng, M.A. Shoffner, K.R. Mitchelson, L.J. Kricka, P. Wilding, 'Analysis of Ligase Chain Reaction Products Amplified in a Silicon-glass Chip using Capillary Electrophoresis', *J. Chromatogr. A*, **732**(1), 151–158 (1996).
94. L.C. Waters, S.C. Jacobson, N. Kroutchinina, J. Khandurina, R.S. Foote, J.M. Ramsey, 'Microchip Device for Cell Lysis, Multiplex PCR Amplification, and Electrophoretic Sizing', *Anal. Chem.*, **70**(24), 5172–5176 (1998).
95. S.C. Jacobson, C.T. Culbertson, J.E. Daler, J.M. Ramsey, 'Microchip Structures for Submillisecond Electrophoresis', *Anal. Chem.*, **70**(16), 3476–3480 (1998).
96. N. Tanaka, F. Fukutome, K. Hosoya, K. Kimata, T. Araki, 'Polymer-supported Pseudo-stationary Phase for Electrokinetic Chromatography: Electrochromatography in a Full Range of Methanol–Water Mixtures with Alkyated Starburst Dendrimers', *J. Chromatogr. A*, **716**(1–2), 57–67 (1995).
97. H. Ozaki, A. Ichihara, S. Terabe, 'Micellar Electrokinetic Chromatography using High-molecular-mass Surfactants: Comparison Between Anionic and Cationic Surfactants and Effects of Modifiers', *J. Chromatogr. A*, **709**(1), 3–10 (1995).
98. E.S. Ahuja, J.P. Foley, 'A Retention Index for Micellar Electrokinetic Chromatography', *Analyst*, **119**(2), 353–360 (1994).
99. S. Yang, M.G. Khaledi, 'Chemical Selectivity in Micellar Electrokinetic Chromatography: Characterization of Solute–Micelle Interactions for Classification of Surfactants', *Anal. Chem.*, **67**(3), 499–510 (1995).
100. R. Guo, W. Thormann, 'Acetylator Phenotyping via Analysis of Four Caffeine Metabolites in Human Urine by Micellar Electrokinetic Capillary Chromatography with Multiwavelength Detection', *Electrophoresis*, **14**(5–6), 547–553 (1993).
101. S.D. Harvey, R.M. Bean, H.R. Udseth, 'High-resolution Separation and Detection of DNA Adducts of Benzo[*a*]-pyrene', *J. Microcol. Sep.*, **4**(3), 191–198 (1992).
102. W. Thormann, A. Minger, S. Molteni, J. Caslavska, P. Gebauer, 'Determination of Substituted Purines in Body Fluids by Micellar Electrokinetic Chromatography with Direct Sample Injection', *J. Chromatogr.*, **593**(1–2), 275–288 (1992).
103. J.G. Bumgarner, M.G. Khaledi, 'Mixed Micellar Electrokinetic Chromatography of Corticosteroids', *Electrophoresis*, **15**(10), 1260–1266 (1994).
104. R.O. Cole, M.J. Sepaniak, W.L. Hinze, J. Gorse, K. Oldiges, 'Bile Salt Surfactants in Micellar Electrokinetic Chromatography. Application to Hydrophobic Molecule Separations', *J. Chromatogr.*, **557**(1–2), 113–123 (1991).
105. E.S. Ahuja, J.P. Foley, 'Influence of Dodecyl Sulfate Counterion on Efficiency, Selectivity, Retention, Elution Range, and Resolution in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **67**(14), 2315–2324 (1995).
106. B. Ye, M. Hadjmohammadi, M.G. Khaledi, 'Selectivity Control in Micellar Electrokinetic Chromatography of Small Peptides using Mixed Fluorocarbon–Hydrocarbon Anionic Surfactants', *J. Chromatogr. A*, **692**(1–2), 291–300 (1995).
107. Y. Mechref, Z. El Rassi, 'Micellar Electrokinetic Chromatography with In-situ Charged Micelles. VI. Evaluation of Novel Chiral Micelles Consisting of Steroidal–Glycoside Surfactant–Borate Complexes', *J. Chromatogr. A*, **724**(1–2), 285–296 (1996).
108. K. Otsuka, J. Kawahara, K. Tatekawa, S. Terabe, 'Chiral Separations by Micellar Electrokinetic Chromatography with Sodium *N*-Dodecanoyl-L-Valinate', *J. Chromatogr.*, **559**(1–2), 209–214 (1991).

109. C. Bjergegaard, S. Michaelsen, H. Sorensen, 'Determination of Phenolic Carboxylic Acids by Micellar Capillary Chromatography and Evaluation of Factors Affecting the Method', *J. Chromatogr.*, **608**(1–2), 403–411 (1992).
110. A. Loregian, C. Scremin, M. Schiavon, A. Marcello, G. Palu, 'Quantitative Analysis of Ribonucleotides Triphosphates in Cell Extracts by High-performance Liquid Chromatography and Micellar Capillary Chromatography: a Comparative Study', *Anal. Chem.*, **66**(18), 2981–2984 (1994).
111. N. Matsubara, S. Terabe, 'Separation of 24 Dansylamino Acids by Capillary Electrophoresis with a Nonionic Surfactant', *J. Chromatogr. A*, **680**(1), 311–315 (1994).
112. C.P. Palmer, S. Terabe, 'A Novel Sulfate Polymer as a Pseudo-stationary Phase for Micellar Electrokinetic Chromatography', *J. Microcol. Sep.*, **8**(2), 115–121 (1996).
113. S. Yang, J.G. Bumgarner, M.G. Khaledi, 'Separation of Highly Hydrophobic Compounds in MEKC with an Ionic Polymer', *J. High Resolut. Chromatogr.*, **18**(7), 443–445 (1995).
114. M. Hong, B.S. Weekley, S.J. Grieb, J.P. Foley, 'Electrokinetic Chromatography using Thermodynamically Stable Vesicles and Mixed Micelles Formed from Oppositely Charged Surfactants', *Anal. Chem.*, **70**(7), 1394–1403 (1998).
115. S. Hjerten, M.-D. Zhu, 'Adaptation of the Equipment for High-performance Electrophoresis to Isoelectric Focusing', *J. Chromatogr.*, **346**(1), 265–270 (1985).
116. K.A. Cobb, V. Dolnik, M. Novotny, 'Electrophoretic Separations of Proteins in Capillaries with Hydrolytically Stable Surface Structures', *Anal. Chem.*, **62**(22), 2478–2483 (1990).
117. W. Thormann, J. Caslavska, S. Molteni, J. Chmelik, 'Capillary Isoelectric Focusing with Electroosmotic Zone Displacement and On-column Multichannel Detection', *J. Chromatogr.*, **589**(1–2), 321–327 (1992).
118. J.M. Hempe, R.D. Craver, 'Quantification of Hemoglobin Variants by Capillary Isoelectric Focusing', *Clin. Chem.*, **40**(12), 2288–2295 (1994).
119. S.-M. Chen, J.E. Wiktorowicz, 'Isoelectric Focusing by Free Solution Capillary Electrophoresis', *Anal. Biochem.*, **206**(1), 84–90 (1992).
120. J. Wu, J. Pawliszyn, 'Universal Detection for Capillary Isoelectric Focusing without Mobilization using a Concentration Gradient Imaging System', *Anal. Chem.*, **64**(2), 224–227 (1992).
121. T. Wang, R.A. Hartwick, 'Whole Column Detection in Capillary Isoelectric Focusing', *Anal. Chem.*, **64**(15), 1745–1747 (1992).
122. S. Molteni, H. Frischknecht, W. Thormann, 'Application of Dynamic Capillary Isoelectric Focusing to the Analysis of Human Hemoglobin Variants', *Electrophoresis*, **15**(1), 22–30 (1994).
123. D.P. Kirby, J.M. Thorne, W.K. Gotzinger, B.L. Karger, 'A CE/ESI/MS Interface for Stable, Low-flow Operation', *Anal. Chem.*, **68**(24), 4451–4457 (1996).
124. O. Hofmann, D. Che, K.A. Cruickshank, U.R. Muller, 'Adaptation of Capillary Isoelectric Focusing to Microchannels on a Glass Chip', *Anal. Chem.*, **71**(3), 678–686 (1999).
125. K. Ueno, E.S. Yeung, 'Simultaneous Monitoring of DNA Fragments Separated by Electrophoresis in a Multiplexed Array of 100 Capillaries', *Anal. Chem.*, **66**(9), 1424–1431 (1994).
126. A. Chrambach, D. Rodbard, 'Polyacrylamide Gel Electrophoresis', *Science*, **172**(3981), 440–451 (1971).
127. P.G. De Gennes, *Scaling Concept in Polymer Physics*, Cornell University Press, Ithaca, NY, 1979.
128. G.W. Slater, J. Noolandi, 'The Biased Reptation Model of DNA Gel Electrophoresis: Mobility vs. Molecular Size and Gel Concentration', *Biopolymers*, **28**(10), 1781–1791 (1989).
129. T. Duke, J.-L. Viovy, A.N. Semenov, 'Electrophoretic Mobility of DNA in Gels. I. New Biased Reptation Theory Including Fluctuations', *Biopolymers*, **34**(2), 239–247 (1994).
130. V. Dolnik, 'Capillary Gel Electrophoresis', *J. Microcol. Sep.*, **6**(4), 315–330 (1994).
131. R.S. Dubrow, 'Capillary Gel Electrophoresis', in *Capillary Electrophoresis: Theory and Practice*, Academic Press, San Diego, Chapter 5, 133–157, 1992.
132. P.D. Grossman, 'Electrophoretic Separation of DNA Sequencing Extension Products using Low-viscosity Entangled Polymer Networks', *J. Chromatogr. A*, **663**(2), 219–227 (1994).
133. D. Tietz, A. Chrambach, 'Analysis of Convex Ferguson Plots in Agarose Gel Electrophoresis by Empirical Computer Modeling', *Electrophoresis*, **7**(6), 241–250 (1986).
134. P. Bocek, A. Chrambach, 'Capillary Electrophoresis in Agarose Solutions: Extension of Size Separations to DNA of 12 kb in Length', *Electrophoresis*, **13**(1–2), 31–34 (1992).
135. A.E. Barron, D.S. Soane, H.W. Blanch, 'Capillary Electrophoresis of DNA in Uncross-linked Polymer Solutions', *J. Chromatogr. A*, **652**(1), 3–16 (1993).
136. W.A. MacCrehan, H.T. Rasmussen, D.M. Northrop, 'Size-selective Capillary Electrophoresis (SSCE) Separation of DNA Fragments', *J. Liq. Chromatogr.*, **15**(6–7), 1063–1080 (1992).
137. Y. Baba, N. Ishimaru, K. Samata, M. Tsuhako, 'High-resolution Separation of DNA Restriction Fragments by Capillary Electrophoresis in Cellulose Derivative Solutions', *J. Chromatogr. A*, **653**(2), 329–335 (1993).
138. P.E. Williams, 'The Analysis of DNA Restriction Fragments and Polymerase Chain Reaction Products by Dynamic Sieving Electrophoresis', *Methods: Compan. Methods Enzymol.*, **4**, 227–232 (1992).

139. H.-T. Chang, E.S. Yeung, 'Poly(ethylene oxide) for High-resolution and High-speed Separation of DNA by Capillary Electrophoresis', *J. Chromatogr. B*, **669**(1), 113–123 (1995).
140. B.A. Siles, D.E. Anderson, N.S. Buchanan, M.F. Warder, 'The Characterization of Heterogeneous Agarose/Hydroxyethyl Cellulose Matrices for the Separation of DNA Fragments using Capillary Electrophoresis', *Electrophoresis*, **18**, 1980–1989 (1997).
141. Q. Gao, E.S. Yeung, 'A Matrix for DNA Separation: Genotyping and Sequencing using Poly(vinylpyrrolidone) Solution in Uncoated Capillaries', *Anal. Chem.*, **70**(7), 1382–1388 (1998).
142. A.E. Barron, H.W. Blanch, D.S. Soane, 'A Transient Entanglement Coupling Mechanisms for DNA Separation by Capillary Electrophoresis in Ultradilute Polymer Solutions', *Electrophoresis*, **15**(5), 597–615 (1994).
143. E.N. Fung, E.S. Yeung, 'High-speed DNA Sequencing by using Mixed Poly(ethylene oxide) Solutions in Uncoated Capillary Columns', *Anal. Chem.*, **67**(13), 1913–1919 (1995).
144. A. Guttman, B. Wanders, N. Cooke, 'Enhanced Separation of DNA Restriction Fragments by Capillary Gel Electrophoresis using Field Strength Gradients', *Anal. Chem.*, **64**(20), 2348–2351 (1992).
145. Y. Kim, M.D. Morris, 'Pulsed Field Capillary Electrophoresis of Multikilobase Length Nucleic Acids in Dilute Methyl Cellulose Solutions', *Anal. Chem.*, **66**(19), 3081–3085 (1994).
146. Y. Walbroehl, J.W. Jorgenson, 'On-column UV Absorption Detector for Open-tubular Capillary Zone Electrophoresis', *J. Chromatogr.*, **315**(1), 135–143 (1984).
147. R.S. Sahota, M.G. Khaledi, 'Nonaqueous Capillary Electrophoresis', *Anal. Chem.*, **66**(7), 1141–1146 (1994).
148. J.L. Miller, M.G. Khaledi, 'Nonaqueous Capillary Electrophoresis', in *High-performance Capillary Electrophoresis*, ed. M.G. Khaledi, Wiley, New York, 525–578, 1997.
149. H. Salimi-Moosavi, R.M. Cassidy, 'Capillary Electrophoresis of Inorganic Anions in Nonaqueous Media with Electrochemical and Indirect UV Detection', *Anal. Chem.*, **67**(6), 1067–1073 (1995).
150. M. Jansson, J. Roeraade, '*N*-Methylformamide as a Separation Medium in Capillary Electrophoresis', *Chromatographia*, **40**(3–4), 163 (1995).
151. J.F. Coetzee, G.R. Padmanabhan, 'Dissociation and Homoconjugation of Certain Phenols in Acetonitrile', *J. Phys. Chem.*, **69**(9), 3193–3197 (1965).
152. A.J. Tomlinson, L.M. Benson, S. Naylor, 'Advantages of Nonaqueous Solvents in the Analysis of Drug Metabolites using CE and On-line CE/MS', *LC-GC*, **12**(2), 122–130 (1994).
153. A.J. Tomlinson, L.M. Benson, J.W. Gorrod, S. Naylor, 'Investigation of the In Vitro Metabolism of the H2-antagonist Mifentidine by On-line Capillary Electrophoresis/Mass Spectrometry using Non-aqueous Separation Conditions', *J. Chromatogr. B*, **657**(2), 373–381 (1994).
154. M. Chiari, E. Kenndler, 'Capillary Zone Electrophoresis in Organic Solvents: Separation of Anions in Methanolic Buffer Solutions', *J. Chromatogr. A*, **716**(1–2), 303–309 (1995).
155. H. Salimi-Moosavi, R.M. Cassidy, 'Application of Nonaqueous Capillary Electrophoresis to the Separation of Long-chain Surfactants', *Anal. Chem.*, **68**(2), 293–299 (1996).
156. J.L. Miller, M.G. Khaledi, D. Shea, 'Separation of Polycyclic Aromatic Hydrocarbons by Nonaqueous Capillary Electrophoresis using Charge-transfer Complexation with Planar Organic Cations', *Anal. Chem.*, **69**(6), 1223–1229 (1997).
157. F. Wang, M.G. Khaledi, 'Chiral Separations by Nonaqueous Capillary Electrophoresis', *Anal. Chem.*, **68**(19), 3460–3467 (1996).
158. C.L. Rice, R. Whitehead, 'Electrokinetic Flow in a Narrow Cylindrical Capillary', *J. Phys. Chem.*, **69**(11), 4017–4024 (1965).
159. T.M. Zimina, R.M. Smith, P. Myers, 'Comparison of ODS-modified Silica Gels as Stationary Phases for Electrochromatography in Packed Capillaries', *J. Chromatogr. A*, **758**(2), 191–197 (1997).
160. S.E. van den Bosch, S. Heemstra, J.C. Kraak, H. Poppe, 'Experiences with Packed Capillary Electrochromatography at Ambient Pressure', *J. Chromatogr. A*, **755**(2), 165–177 (1996).
161. M.M. Dittmann, G.P. Rozing, 'Capillary Electrochromatography – a High Efficiency Micro-separation Technique', *J. Chromatogr. A*, **744**(1–2), 63–74 (1996).
162. F. Lelièvre, C. Yan, R.N. Zare, P. Gareil, 'Capillary Electrochromatography: Operating Characteristics and Enantiomeric Separations', *J. Chromatogr. A*, **723**(1), 145–156 (1996).
163. C. Yan, R. Dadoo, H. Zhao, R. Zare, D.J. Rakestraw, 'Capillary Electrochromatography: Analysis of Polyaromatic Hydrocarbons', *Anal. Chem.*, **67**(13), 2026–2029 (1995).
164. A.S. Rathore, Cs. Horvath, 'Axial Nonuniformities and Flow in Columns for Capillary Electrochromatography', *Anal. Chem.*, **70**(14), 3069–3077 (1998).
165. R.J. Boughtflower, T. Underwood, C.J. Paterson, 'Capillary Electrochromatography – Some Important Considerations in the Preparation of Packed Capillaries and the Choice of Mobile Phase Buffers', *Chromatographia*, **40**(5–6), 329–335 (1995).
166. B. Behnke, E. Grom, E. Bayer, 'Evaluation of the Parameters Determining the Performance of Electrochromatography in Packed Capillary Columns', *J. Chromatogr. A*, **716**(1–2), 207–213 (1995).
167. R.J. Boughtfower, T. Underwood, J. Maddin, 'The Production of Packed Capillaries using a Novel

Pressurised Ultrasound Device', *Chromatographia*, **41**(7–8), 398–402 (1995).
168. J.H. Knox, I.H. Grant, 'Electrochromatography in Packed Tubes using 1.5 to 50 µm Silica Gels and ODS Bonded Silica Gels', *Chromatographia*, **32**(7–8), 317–328 (1991).
169. L.A. Colon, K.J. Reynolds, R. Alicea-Maldonado, A.M. Fermier, 'Advances in Capillary Electrochromatography', *Electrophoresis*, **18**, 2162–2174 (1997).
170. C. Yan, 'Electrokinetic Packing of Capillary Columns', US Patent 5 453 163, 1993.
171. W.D. Pfeffer, E.S. Yeung, 'Electroosmotically Driven Electrochromatography of Anions Having Similar Electrophoretic Mobilities by Ion Pairing', *J. Chromatogr. A*, **557**(1–2), 125–136 (1991).
172. T. Tsuda, K. Nomura, G. Nakagawa, 'Open-tubular Microcapillary Liquid Chromatography with Electroosmosis Flow using a UV Detector', *J. Chromatogr.*, **248**(1), 241–247 (1982).
173. J.J. Pesek, M.T. Matyska, 'Electrochromatography in Chemically Modified Etched Fused-silica Capillaries', *J. Chromatogr. A*, **736**(1–2), 255–264 (1996).
174. C. Fujimoto, Y. Fujise, E. Matsuzawa, 'Fritless Packed Columns for Capillary Electrochromatography: Separation of Uncharged Compounds on Hydrophobic Hydrogels', *Anal. Chem.*, **68**(17), 2753–2757 (1996).
175. E.C. Peters, M. Petro, F. Svec, J.M.J. Frechet, 'Molded Rigid Polymer Monoliths as Separation Media for Capillary Electrochromatography. 2. Effect of Chromatographic Conditions on the Separation', *Anal. Chem.*, **70**(11), 2296–2302 (1998).
176. Y. Guo, L.A. Colón, 'A Stationary Phase for Open Tubular Chromatography and Electrochromatography using Sol–Gel Technology', *Anal. Chem.*, **67**, 2511–2516 (1995).
177. L. Schweitz, L.I. Andersson, S. Nilsson, 'Capillary Electrochromatography with Predetermined Selectivity Obtained Through Molecular Imprinting', *Anal. Chem.*, **69**(6), 1179–1183 (1997).
178. A.L. Crego, A. González, M.L. Marina, 'Electrochromatography', *Crit. Rev. Anal. Chem.*, **26**(4), 261–304 (1996).
179. K. Schmeer, B. Behnke, E. Bayer, 'Capillary Electrochromatography–Electrospray Mass Spectrometry: a Microanalysis Technique', *Anal. Chem.*, **67**(20), 3656–3658 (1995).
180. S. Li, D.K. Lloyd, 'Direct Chiral Separations by Capillary Electrophoresis using Capillaries Packed with an α_1-Acid Glycoprotein Chiral Stationary Phase', *Anal. Chem.*, **65**(24), 3684–3690 (1995).
181. J.J. Pesek, M.T. Matyska, 'Separation of Tetracyclines by High-performance Capillary Electrophoresis and Capillary Electrochromatography', *J. Chromatogr. A*, **736**(1–2), 313–320 (1996).
182. S.A. Shamsi, C. Akbay, I.M. Warner, 'Polymeric Anionic Surfactant for Electrokinetic Chromatography: Separation of 16 Priority Polycyclic Aromatic Hydrocarbon Pollutants', *Anal. Chem.*, **70**(14), 3078–3083 (1998).
183. C. Yan, R. Dadoo, R.N. Zare, D.J. Rakestraw, D.S. Anex, 'Gradient Elution in Capillary Electrochromatography', *Anal. Chem.*, **68**(17), 2726–2730 (1996).
184. J.T. Smith, 'Developments in Amino Acid Analysis using Capillary Electrophoresis', *Electrophoresis*, **18**, 2377–2392 (1997).
185. Y. Mechref, J.T. Smith, Z. El Rassi, 'Micellar Electrokinetic Capillary Chromatography with In Situ Charged Micelles. VII. Expanding the Utility of Alkylglycoside–Borate Micelles to Acidic and Neutral pH for Capillary Electrophoresis of Dansyl Amino Acids and Herbicides', *J. Liq. Chromatogr.*, **18**(18–19), 3769–3786 (1995).
186. S. Terabe, Y. Miyashita, Y. Ishihama, O. Shibata, 'Cyclodextrin-modified Micellar Electrokinetic Chromatography: Separation of Hydrophobic and Enantiomeric Compounds', *J. Chromatogr.*, **636**(1), 47–55 (1993).
187. D.T. Witte, S. Nagard, M. Larsson, 'Improved Sensitivity by On-line Isotachophoretic Preconcentration in Capillary Zone Electrophoresis Determination of Peptide-like Solutes', *J. Chromatogr. A*, **687**(1), 155–166 (1994).
188. T. Kornfelt, A. Vinther, G.N. Okafo, P. Camilleri, 'Improved Peptide Mapping using Phytic Acid as Ion-pairing Buffer Additive in Capillary Electrophoresis', *J. Chromatogr. A*, **726**(1–2), 223–228 (1996).
189. H. Wan, L.G. Blomberg, 'Enantiomeric and Diastereomeric Separation of Di- and Tri-peptides by Capillary Electrophoresis', *J. Chromatogr. A*, **758**(2), 303–311 (1997).
190. F. Thunecke, A. Kalman, F. Kalman, S. Ma, A.S. Rathore, C. Horvath, 'Kinetic Study on the *Cis–Trans* Isomerization of Peptidyl–Proline Dipeptides', *J. Chromatogr. A*, **744**(1–2), 259–272 (1996).
191. N.J. Reinhoud, U.R. Tjaden, J. van der Greef, 'Automated On-capillary Isotachophoretic Reaction Cell for Fluorescence Derivatization of Small Sample Volumes at Low Concentrations Followed by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **673**(2), 255–266 (1994).
192. A.J. Tomlinson, L.M. Benson, N.A. Guzman, S. Naylor, 'Preconcentration and Microreaction Technology On-line with Capillary Electrophoresis', *J. Chromatogr. A*, **744**(1–2), 3–15 (1996).
193. T. Hara, J. Yokogi, S. Okamura, S. Kato, R. Nakajima, 'On-line Chemiluminescence Detection of Proteins Separated by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **652**(2), 361–367 (1993).
194. M.E. Legaz, M.M. Pedrosa, 'Effect of Polyamines on the Separation of Ovalbumin Glycoforms by Capillary Electrophoresis', *J. Chromatogr. A*, **719**(1), 159–170 (1996).

195. Y.J. Yao, S.F.Y. Li, 'Capillary Zone Electrophoresis of Basic Proteins with Chitosan as a Capillary Modifier', *J. Chromatogr. A*, **663**(1), 97–104 (1994).
196. M. Gilges, M.H. Kleemiss, G. Schomburg, 'Capillary Zone Electrophoresis Separations of Basic and Acidic Proteins using Poly(vinyl alcohol) Coatings in Fused Silica Capillaries', *Anal. Chem.*, **66**(13), 2038–2046 (1994).
197. J.K. Towns, F.E. Regnier, 'Polyethyleneimine-bonded Phases in the Separation of Proteins by Capillary Electrophoresis', *J. Chromatogr.*, **516**, 69–78 (1990).
198. M.H.A. Busch, J.C. Kraak, H. Poppe, 'Cellulose Acetate-coated Fused-silica Capillaries for the Separation of Proteins by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **695**(2), 287–296 (1995).
199. D. Schmalzing, C.A. Piggee, F. Foret, E. Carrilho, B.L. Karger, 'Characterization and Performance of a Neutral Hydrophilic Coating for the Capillary Electrophoretic Separation of Biopolymers', *J. Chromatogr. A*, **652**(1), 149–159 (1993).
200. M. Huang, J. Plocek, M.V. Novotny, 'Hydrolytically Stable Cellulose-derivative Coatings for Capillary Electrophoresis of Peptides, Proteins and Glycoconjugates', *Electrophoresis*, **16**(3), 396–401 (1995).
201. Y.J. Yao, K.C. Loh, S.F.Y. Li, 'Analysis of Recombinant Human Tumor Necrosis Factor Beta by Capillary Electrophoresis', *Electrophoresis*, **16**(4), 647–653 (1995).
202. M. Castagnola, I. Messana, L. Cassiano, R. Rabino, D.V. Rossetti, B. Giardina, 'The use of Capillary Electrophoresis for the Determination of Hemoglobin Variants', *Electrophoresis*, **16**(8), 1492–1498 (1995).
203. G.R. Paterson, J.P. Hill, D.E. Otter, 'Separation of β-Lactoglobulin A, B, and C Variants of Bovine Whey using Capillary Electrophoresis', *J. Chromatogr. A*, **700**(1–2), 105–110 (1995).
204. R.G. Nielsen, E.C. Rickard, P.F. Santa, D.A. Sharknas, G. Sittampalam, 'Separation of Antibody–Antigen Complexes by Capillary Zone Electrophoresis, Isoelectric Focusing and High Performance Size-exclusion Chromatography', *J. Chromatogr.*, **539**(1), 177–185 (1991).
205. M.J. Schmerr, K.R. Goodwin, 'Characterization by Capillary Electrophoresis of the Surface Glycoproteins of Ovine Lentiviruses Before and After Treatment with Glycosidic Enzymes', *J. Chromatogr. A*, **652**(1), 199–205 (1993).
206. A.Z. Hu, I.D. Cruzado, J.W. Hill, C.J. McNeal, R.D. Macfarlane, 'Characterization of Lipoprotein A by Capillary Zone Electrophoresis', *J. Chromatogr. A*, **717**(1–2), 33–39 (1995).
207. C. Schwer, 'Capillary Isoelectric Focusing: a Routine Method for Protein Analysis?', *Electrophoresis*, **16**(11), 2121–2126 (1995).
208. K. Shimura, B.L. Karger,'Affinity Probe Capillary Electrophoresis: Analysis of Recombinant Human Growth Hormone with a Fluorescent Labeled Antibody Fragment', *Anal. Chem.*, **66**(1), 9–15 (1994).
209. D. Wu, F.E. Regnier, 'Sodium Dodecyl Sulfate/Capillary Gel Electrophoresis of Proteins using Non-cross-linked Polyacrylamide', *J. Chromatogr. A*, **608**(1–2), 349–356 (1992).
210. M. Kamahori, H. Kambara, 'Characteristics of Single-stranded DNA Separation by Capillary Gel Electrophoresis', *Electrophoresis*, **17**(9), 1476–1484 (1996).
211. A. Muscate, F. Natt, A. Paulus, M. Ehrat, 'Capillary Affinity Gel Electrophoresis for Combined Size- and Sequence-dependent Separation of Oligonucleotides', *Anal. Chem.*, **70**(7), 1419–1424 (1998).
212. I. Kheterpal, R.A. Mathies, 'Capillary Array Electrophoresis: DNA Sequencing', *Anal. Chem.*, **71**(1), 31A–37A (1998).
213. S. Liu, Y. Shi, W.W. Ja, R.A. Mathies, 'Optimization of High-speed DNA Sequencing on Microfabricated Capillary Electrophoresis Channels', *Anal. Chem.*, **71**(3), 566–573 (1999).
214. O. Salas-Solano, E. Carrilho, L. Kotler, A.W. Miller, W. Goetzinger, Z. Sosic, B.L. Karger, 'Routine DNA Sequencing of 1000 Bases in Less Than One Hour by Capillary Electrophoresis with Replaceable Linear Polyacrylamide Solutions', *Anal. Chem.*, **70**(19), 3996–4003 (1998).
215. B.A. Johnson, S.G. McClain, E.R. Doran, G. Tice, M.A. Kirsch, 'Rapid Purification of Synthetic Oligonucleotides: a Convenient Alternative to High Performance Liquid Chromatography and Polyacrylamide Gel Electrophoresis', *BioTechniques*, **8**(4), 424–429 (1990).
216. L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B.H. Kent, L.E. Hood, 'Fluorescence Detection in Automated DNA Sequence Analysis', *Nature (London)*, **321**(6071), 674–679 (1986).
217. H. Zhu, S.M. Clark, S.C. Benson, H.S. Rye, A.N. Glazer, R.A. Mathies, 'High-sensitivity Capillary Electrophoresis of Double-stranded DNA Fragments using Monomeric and Dimeric Fluorescent Intercalating Dyes', *Anal. Chem.*, **66**, 1941–1948 (1994).
218. Y. Kim, M.D. Morris, 'Ultrafast High-resolution Separation of Large DNA Fragments by Pulsed-field Capillary Electrophoresis', *Electrophoresis*, **17**, 152–160 (1996).
219. X. Shi, R.W. Hammond, M.D. Morris, 'Dynamics of DNA During Pulsed Field Electrophoresis in Entangled and Dilute Polymer Solutions', *Anal. Chem.*, **67**(18), 3219–3222 (1995).
220. B.C. Courtney, K.C. Williams, Q.A. Bing, J.J. Schlager, 'Capillary Gel Electrophoresis as a Method to Determine Ligation Efficiency', *Anal. Biochem.*, **228**, 281–286 (1995).
221. Z. Nackerdien, S. Morris, S. Choquette, B. Ramos, D. Atha, 'Analysis of Laser-induced Plasmid DNA Photolysis by Capillary Electrophoresis', *J. Chromatogr. B*, **683**, 91–96 (1996).

222. R.H. Hammond, H. Oana, J.J. Schwinefus, J. Bonadio, R.J. Levy, M.D. Morris, 'Capillary Electrophoresis of Supercoiled and Linear DNA in Dilute Hydroxyethyl Cellulose Solution', *Anal. Chem.*, **69**(6), 1192–1196 (1997).
223. H. Oana, R.H. Hammond, J.J. Schwinefus, S.-C. Wang, M. Doi, M.D. Morris, 'High-speed Separation of Linear and Supercoiled DNA by Capillary Electrophoresis. Buffer, Entangling Polymer, and Electric Field Effects', *Anal. Chem.*, **70**(3), 574–579 (1998).
224. N.W.C. Chan, K. Stangier, R. Sherburne, D.E. Taylor, Y. Zhang, N.J. Dovichi, M.M. Palcic, 'The Biosynthesis of Lewis X in *Helicobacter Pylori*', *Glycobiology*, **5**(7), 683–688 (1995).
225. T.W. Garner, E.S. Yeung, 'Indirect Fluorescence Detection of Sugars Separated by Capillary Zone Electrophoresis with Visible Laser Excitation', *J. Chromatogr.*, **515**(1), 639–644 (1990).
226. L.A. Colon, R. Dadoo, R.N. Zare, 'Determination of Carbohydrates by Capillary Zone Electrophoresis with Amperometric Detection at a Copper Microelectrode', *Anal. Chem.*, **65**(4), 476–481 (1993).
227. A.E. Bruno, B. Krattiger, F. Maystre, H.M. Widmer, 'On-column Laser-based Refractive Index Detector for Capillary Electrophoresis', *Anal. Chem.*, **63**(23), 2689–2697 (1991).
228. G. Tuffal, A. Tuong, C. Dhers, F. Uzabiaga, M. Riviere, C. Picard, G. Puzo, 'Direct Evidence of Methylglucose Lipopolysaccharides/Palmitoyl-CoA Noncovalent Complexes by Capillary Zone Electrophoresis–Electrospray/Mass Spectrometry', *Anal. Chem.*, **70**(9), 1853–1858 (1998).
229. B. Stahl, M. Steup, M. Karas, F. Hillenkamp, 'Analysis of Neutral Oligosaccharides by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **63**(14), 1463–1466 (1991).
230. S. Honda, A. Makino, S. Suzuki, K. Kakehi, 'Analysis of the Oligosaccharides in Ovalbumin by High-performance Capillary Electrophoresis', *Anal. Biochem.*, **191**(2), 228–234 (1990).
231. M. Stefansson, M. Novotny, 'Separation of Complex Oligosaccharide Mixtures by Capillary Electrophoresis in the Open-tubular Format', *Anal. Chem.*, **66**(7), 1134–1140 (1994).
232. J. Liu, O. Shirota, M. Novotny, 'Capillary Electrophoresis of Amino Sugars with Laser-induced Fluorescence Detection', *Anal. Chem.*, **63**(5), 413–417 (1991).
233. E. Payan, N. Presle, N.F. Lapicque, J.Y. Jouzeau, K. Bordji, S. Oerther, G. Miralles, D. Mainard, P. Netter, 'Separation and Quantitation by Ion-association Capillary Zone Electrophoresis of Unsaturated Disaccharide Units of Chondroitin Sulfates and Oligosaccharides Derived from Hyaluronan', *Anal. Chem.*, **70**(22), 4780–4786 (1998).
234. M. Steffansson, J. Sudor, M. Hong, J. Chmelikova, J. Chmelik, M.V. Novotny, 'Aminodextran as a Migration Moderator in Capillary Gel Electrophoresis of Charged Polysaccharides', *Anal. Chem.*, **69**(18), 3846–3850 (1997).
235. Y.S. Yoo, Y.-S. Kim, G.J. Jhon, J. Park, 'Separation of Gangliosides using Cyclodextrin in Capillary Zone Electrophoresis', *J. Chromatogr. A*, **652**(2), 431–439 (1993).
236. M. Hong, J. Sudor, M. Stefansson, M.V. Novotny, 'High Resolution Studies of Hyaluronic Acid Mixtures Through Capillary Electrophoresis', *Anal. Chem.*, **70**(3), 568–573 (1998).
237. L.R. Zieske, D. Fu, S.H. Khan, R.A. O'Neill, 'Multidimensional Mapping of Pyridylamine-labeled N-linked Oligosaccharides by Capillary Electrophoresis', *J. Chromatogr. A*, **720**(1–2), 395–407 (1996).
238. F. Kilar, S. Hjertén, 'Separation of the Human Transferrin Isoforms by Carrier-free High-performance Zone Electrophoresis and Isoelectric Focusing', *J. Chromatogr.*, **480**(1), 351–357 (1989).
239. E. Gassmann, J.E. Kuo, R.N. Zare, 'Electrokinetic Separation of Chiral Compounds', *Science*, **230**(4727), 813–814 (1985).
240. S.A.C. Wren, R.C. Rowe, 'Theoretical Aspects of Chiral Separation in Capillary Electrophoresis. II. The Role of Organic Solvent', *J. Chromatogr.*, **609**(1–2), 363–367 (1992).
241. C. Kanz, T. Fleischmann, H.-P.E. Kohler, W. Giger, 'Separation of Chiral Biodegradation Intermediates of Linear Alkylbenzenesulfonates by Capillary Electrophoresis', *Anal. Chem.*, **70**(5), 913–917 (1998).
242. S.R. Gratz, A.M. Stalcup, 'Enantiomeric Separations of Terbutaline by CE with a Sulfated β-Cyclodextrin Chiral Selector: a Quantitative Binding Study', *Anal. Chem.*, **70**(24), 5166–5171 (1998).
243. H. Nishi, K. Nakamura, H. Nakai, T. Sato, 'Enantiomeric Separation of Drugs by Mucopolysaccharide-mediated Electrokinetic Chromatography', *Anal. Chem.*, **67**(14), 2334–2341 (1995).
244. H. Soini, M. Stefansson, M.-L. Riekkola, M.V. Novotny, 'Maltooligosaccharides as Chiral Selectors for the Separation of Pharmaceuticals by Capillary Electrophoresis', *Anal. Chem.*, **66**(20), 3477–3484 (1994).
245. M.G. Schmid, G. Gubitz, 'Capillary Zone Electrophoretic Separation of the Enantiomers of Dipeptides Based on Host–Guest Complexation with a Chiral Crown Ether', *J. Chromatogr. A*, **709**(1), 81–88 (1995).
246. D.W. Armstrong, K. Rundlett, G.L. Reid, III, 'Use of a Macrocyclic Antibiotic, Rifamycin B, and Indirect Detection for the Resolution of Racemic Amino Alcohols by CE', *Anal. Chem.*, **66**(10), 1690–1695 (1994).
247. U.B. Nair, D.W. Armstrong, W.L. Hinze, 'Characterization and Evaluation of D-(+)-Tubocurarine Chloride as a Chiral Selector for Capillary Electrophoretic Enantioseparations', *Anal. Chem.*, **70**(6), 1059–1065 (1998).

248. Y. Tanaka, N. Matsubara, S. Terabe, 'Separation of Enantiomers by Affinity Electrokinetic Chromatography using Avidin', *Electrophoresis*, **15**(6), 848–853 (1994).
249. R.O. Cole, M.J. Sepaniak, W.L. Hinze, 'Optimization of Binaphthyl Enantiomer Separations by Capillary Zone Electrophoresis using Mobile Phases Containing Bile Salts and Organic Solvent', *J. High Resolut. Chromatogr.*, **13**, 579–582 (1990).
250. J.R. Mazzeo, M.E. Swartz, E.R. Grover, 'A Resolution Equation for Electrokinetic Chromatography Based on Electrophoretic Mobilities', *Anal. Chem.*, **67**(17), 2966–2973 (1995).
251. M.E. Swartz, 'Capillary Electrophoretic Determination of Inorganic Ions in a Prenatal Vitamin Formulation', *J. Chromatogr.*, **640**(1–2), 441–444 (1993).
252. J.P. Romano, J. Krol, 'Capillary Ion Electrophoresis, an Environmental Method for the Determination of Anions in Water', *J. Chromatogr.*, **640**(1–2), 403–412 (1993).
253. D.R. Salomon, J. Romano, 'Applications of Capillary Ion Electrophoresis in the Pulp and Paper Industry', *J. Chromatogr.*, **602**(1–2), 219–225 (1992).
254. P. Jandik, G. Bonn, *Capillary Electrophoresis of Small Molecules and Ions*, VCH, New York, Chapter 3, 1993.
255. P. Jandik, W.R. Jones, A. Weston, P.R. Brown, 'Electrophoretic Capillary Ion Analysis: Origins, Principles, and Applications', *LC-GC*, **9**(9), 634–645 (1991).
256. T. Kaneta, S. Tanaka, M. Taga, 'Effect of Cetylmethylammonium Chloride on Electroosmotic and Electrophoretic Mobilities in Capillary Zone Electrophoresis', *J. Chromatogr.*, **653**(2), 313–319 (1993).
257. M.P. Harrold, M.J. Wojtusik, J. Riviello, P. Henson, 'Parameters Influencing Separation and Detection of Anions by Capillary Electrophoresis', *J. Chromatogr.*, **640**(1–2), 463–471 (1993).
258. S.A. Shamsi, N.D. Danielson, 'Naphthalenesulfonates as Electrolytes for Capillary Electrophoresis of Inorganic Anions, Organic Acids, and Surfactants with Indirect Photometric Detection', *Anal. Chem.*, **66**(21), 3757–3764 (1994).
259. S.M. Cousins, P.R. Haddad, W. Buchberger, 'Evaluation of Carrier Electrolytes for Capillary Zone Electrophoresis of Low-molecular-mass Anions with Indirect UV Detection', *J. Chromatogr.*, **671**(1–2), 397–402 (1994).
260. S.A. Shamsi, N.D. Danielson, 'Ribonucleotide Electrolytes for Capillary Electrophoresis of Polyphosphates and Polyphosphonates with Indirect Photometric Detection', *Anal. Chem.*, **67**(11), 1845–1852 (1995).
261. W.R. Jones, P. Jandik, 'Controlled Changes of Selectivity in the Separation of Ions by Capillary Electrophoresis', *J. Chromatogr.*, **546**(1–2), 445–458 (1991).
262. C. Stathakis, R.M. Cassidy, 'Effect of Electrolyte Composition in the Capillary Electrophoretic Separation of Inorganic/Organic Anions in the Presence of Cationic Polymers', *J. Chromatogr.*, **699**(1–2), 353–361 (1995).
263. P.J. Oefner, 'Surface-charge Reversed Capillary Zone Electrophoresis of Inorganic and Organic Anions', *Electrophoresis*, **16**(1), 46–56 (1995).
264. P. Doble, P.R. Haddad, 'Use of Electrolytes Containing Multiple Co-anions in the Analysis of Anions by Capillary Electrophoresis using Indirect Absorbance Detection', *Anal. Chem.*, **71**(1), 15–22 (1999).
265. F. Guan, H. Wu, Y. Luo, 'Sensitive and Selective Method for Direct Determination of Nitrite and Nitrate by High Performance Capillary Electrophoresis', *J. Chromatogr. A*, **719**(2), 427–433 (1996).
266. W. Beck, H. Engelhardt, 'Capillary Electrophoresis of Organic and Inorganic Cations with Indirect UV Detection', *Chromatographia*, **33**(7–8), 313–316 (1992).
267. A.R. Timerbaev, W. Buchberger, O.P. Semenova, G.K. Bonn, 'Metal Ion Capillary Zone Electrophoresis with Direct UV-detection: Determination of Transition Metals using an 8-Hydroxyquinoline-5-sulphonic Acid Chelating System', *J. Chromatogr.*, **630**(1–2), 379–389 (1993).
268. C. Quang, M.G. Khaledi, 'Prediction and Optimization of the Separation of Metal Cations by Capillary Electrophoresis with Indirect UV Detection', *J. Chromatogr.*, **659**(2), 459–466 (1994).
269. P. Kuban, B. Karlberg, 'Simultaneous Determination of Small Cations and Anions by Capillary Electrophoresis', *Anal. Chem.*, **70**(2), 360–365 (1998).
270. L. Gross, E.S. Yeung, 'Indirect Fluorometric Detection of Cations in Capillary Zone Electrophoresis', *Anal. Chem.*, **62**(5), 427–431 (1990).
271. J.W. Olesik, J.A. Kinzer, S.V. Olesik, 'Capillary Electrophoresis Inductively Coupled Plasma Spectrometry for Rapid Elemental Speciation', *Anal. Chem.*, **67**(1), 1–12 (1995).
272. A. Nann, E. Pretsch, 'Potentiometric Detection of Anions Separated by Capillary Electrophoresis using and Ion-selective Microelectrode', *J. Chromatogr.*, **676**(2), 437–442 (1994).
273. N. Avdalovic, C.A. Pohl, R.D. Rocklin, J.R. Stillian, 'Determination of Cations and Anions by Capillary Electrophoresis Combined with Suppressed Conductivity Detection', *Anal. Chem.*, **65**(10), 1470–1475 (1993).

Chiral Separations by High-performance Liquid Chromatography

Timothy J. Ward and Tanya M. Oswald
Millsaps College, Jackson, USA

1 Introduction **2**
1.1 General Considerations for Chromatographic Stationary Phases 3
1.2 General Considerations for Mobile Phase Parameters 3
2 Chiral Mobile Phase Additives **4**
3 Chiral Stationary Phase Based Separations **4**
4 Ligand Exchange **4**
5 Protein-based Chiral Stationary Phases **5**
5.1 Bovine Serum Albumin Chiral Stationary Phase 5
5.2 α_1-Acid Glycoprotein Chiral Stationary Phase 6
6 Carbohydrate-based Phases **7**
6.1 Cellulose- and Amylose-based Chiral Stationary Phases 8
7 Pirkle Chiral Stationary Phases **8**
7.1 Properties, Structures, and Applications 8
7.2 Separation Mechanism 9
8 Cyclodextrin Chiral Stationary Phases **10**
8.1 Properties and Structures 10
8.2 Applications 11
8.3 Separation Mechanism 12
9 The Macrocyclic Antibiotic Chiral Stationary Phases **13**
9.1 Properties and Structures 13
9.2 Applications 15
9.3 Separation Mechanism 15
10 Future Trends **16**
Acknowledgment **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

The word ''chiral'' is derived from the Greek word ''cheir'', which means hand. Chiral molecules are molecules that are related to each other in the same way that a left hand is related to a right hand. These molecules are mirror-images of each other and are nonsuperimposable. Chiral separations have been considered among the most difficult of all separations since enantiomers have identical chemical and physical properties in an achiral environment. In this chapter we will focus on techniques used in high-performance liquid chromatography (HPLC). Most chiral separations by HPLC are accomplished via direct resolution using a chiral stationary phase (CSP). In this technique a chiral resolving agent is bound or immobilized to an appropriate support to make a CSP, and the enantiomers are resolved by the formation of temporary diastereomeric complexes between the analyte and the CSP. Various types of CSPs have been developed, including ligand exchange, protein-based, carbohydrate-based, Pirkle-type, cyclodextrin-based, and macrocyclic antibiotic-based CSPs.

Ligand exchange phases are used with aqueous buffer mobile phases in which enantiomers are separated based on the differences in their charge and ionization constants. Limitations are that only ionized analytes can be separated using this technique and the copper-salt containing mobile phases used absorb in the ultraviolet (UV) region, decreasing detection sensitivity. Protein-based CSPs comprise a number of commercially available columns. These CSPs can be used in the reversed-phase mode with aqueous buffers and there are a limited number of variables to control in developing a separation method. Advantages of protein-based CSPs include low column capacity, limited solvent options and the inability to reverse the elution order of the analyte. The carbohydrate-based CSPs consist of derivatized cellulose and amylose phases and are generally used in the normal phase mode, with the exception of two derivatized phases which are conditioned for the reversed-phase mode. The main disadvantages of these phases are the limitations in pressure and solvent used since these phases are not covalently bonded but merely adsorbed on the silica. These phases may not be used with solvents of intermediate polarity, for example, methylene chloride, acetone, tetrahydrofuran, and acetonitrile. The Pirkle-type CSP typically uses nonpolar organic mobile phases such as hexane, with 2-propanol or ethanol as organic modifiers. Under these conditions, retention of the solutes decreases as the mobile phase polarity increases, following the normal phase mode behavior. The Pirkle-type columns are generally employed in separating compounds containing a π-acid or π-basic moiety, or both. The cyclodextrins can be used with either aqueous buffers or in the polar organic mode. Generally analytes separated using the cyclodextrins require formation of an inclusion complex with the cyclodextrin. Separation is most favorable when the analyte interacts with the mouth of the cyclodextrin molecule via hydrogen bonding to the functional groups present. The

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

macrocyclic antibiotics can be used in either the reversed-phase or the normal phase mode. Enantioselectivities in each mode have been shown to be different. Furthermore, the macrocyclic antibiotics can be derivatized to alter their selectivity. The bonded macrocyclic antibiotic CSPs resemble the protein-based CSPs in many ways. However, the macrocyclic antibiotic CSPs are more stable and have greater capacities than the protein-based CSPs. Numerous analytes are separated using this class of CSP.

1 INTRODUCTION

Chiral separations are important in many scientific disciplines such as chiral synthesis, mechanistic studies, catalysis, kinetics, biology, pharmacology, medicine, and biochemistry. The importance of stereochemistry has long been known, being first recognized more than 100 years ago by Pasteur and van't Hoff-LeBel.[1,2] In the early 1800s Pasteur achieved the first chiral resolution of a racemic mixture into optical isomers. Using no more elaborate equipment than a microscope and tweezers, he separated the ''right-handed'' crystals of ammonium tartaric acid from the ''left-handed''. The word ''chiral'' is derived from the Greek word ''cheir'', which means hand.[3] Chiral molecules are molecules that are related to each other in the same way that a left hand is related to a right hand; that is these molecules are mirror-images of each other and are nonsuperimposable (see **(1)** and **(2)**). The most common example of a chiral molecule is one that possesses a tetrahedral carbon (also known as a stereogenic center) with four different substituents. However, this is not the only requirement for a molecule to exist as chiral. Atoms other than carbon such as nitrogen, sulfur, phosphorus, etc. can also create chirality as well as molecules that lack a reflection symmetry. Structures **(3–6)** are examples of chiral molecules.

Molecules with different spatial arrangement of atoms may be classified as either diastereomers or enantiomers.[4] The term chiral and enantiomers are related in that enantiomers occur only with compounds whose molecules are chiral, that is, left- and right-handed. Diastereomers are stereoisomers which contain one or more asymmetric centers that are not enantiomers of each other. Two stereoisomers cannot be both diastereomers and enantiomers concurrently. Diastereomers have different chemical and physical properties. They can be distinguished from one another by a number of techniques that take advantage of these different properties. Enantiomers, on the other hand, have identical chemical and physical properties. The only difference is that enantiomers rotate plane-polarized light in opposite

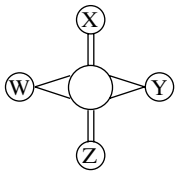

(1)

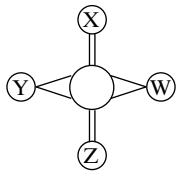

(2)

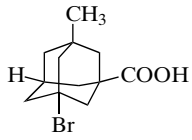

(3)

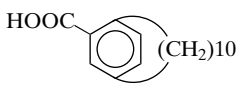

(4)

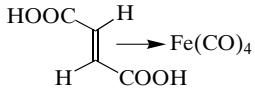

(5)

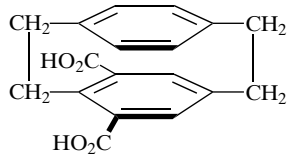

(6)

directions and interact differently in a chiral environment. Traditionally, chiral separations have been considered among the most difficult of all separations. This is because most conventional separation techniques such as distillation, crystallization, and liquid–liquid extraction, are based on differences in chemical and physical properties of the analytes. The separation of enantiomers is a challenging problem from both theoretical and experimental approaches.

The intense interest in enantiomeric separations emerges not only from pure scientific curiosity. Government guidelines regulating the development of new drugs mandate efficient methods for the analysis and preparation of enantiomerically pure compounds. Nature's well-known ability to produce and convert chiral compounds with remarkable stereospecificity has been fascinating and led to the study of stereochemical composition of chemical compounds and their pharmaceutical importance. The effect of chiral molecules on the reaction mechanism has long been studied by spectroscopists and organic chemists. Examples of enantiomeric resolutions (resolving stereochemical composition of chiral compounds) can be found in physiological reactions. For example, catecholamines epinephrine and norepinephrine, which are synthesized in the human body mostly by the adrenal medulla, are endogenous hormones that are important in the regulatory processes such as blood pressure, metabolism, and fight-or-flight response.[5] The biochemical synthesis of catecholamines begins with an amino acid tyrosine, which is converted to L-DOPA (3-(3,4-dihydroxyphenyl)alanine). L-DOPA is then converted by L-DOPA decarboxylase, a sterospecific enzyme affecting only the L-enantiomer, to dopamine.[6] Then, another stereospecific enzymatic hydroxylation converts

dopamine into the chiral compound norepinephrine and subsequently to epinephrine by N-methylation. L-DOPA is administered to patients suffering from Parkinson disease to increase the blood level of dopamine in these patients.[6,7] The importance of the correct configuration is also evident from the fact that the naturally occurring *R*-epinephrine is at least 20 times more active than its enantiomer. Likewise, *R*-(−)-isopropyl norepinephrine is approximately 800 times more effective at the β_2 receptor as a bronchodilator than its enantiomers.[2]

1.1 General Considerations for Chromatographic Stationary Phases

Since the two forms of enantiomers have identical chemical and physical properties in an achiral environment, separating chiral compounds requires a different approach. The new approach to obtain chiral separations is to create a ''chiral environment'', which provides more specific interactions between chiral selector-chiral analyte than is obtainable in conventional techniques. Chromatographic methods such as HPLC, thin-layer chromatography (TLC), and gas–liquid chromatography (LC), offer distinct advantages over conventional separation techniques.[8] These methods show promise in moderate-scale separations, offering substantial increases in efficiency. In this chapter we will focus on techniques used in HPLC.

There are four separation modes in HPLC.[9] These include liquid–liquid chromatography, liquid–solid chromatography, size-exclusion chromatography, and ion-exchange chromatography. The most important and commonly used mode in HPLC is liquid–liquid chromatography. The support generally employed is specially treated, size distribution-controlled silica. The separation occurs due to differences in retention of various compounds on the stationary phase. The stationary phase can be divided into two large categories: partition chromatography and adsorption chromatography. In partition chromatography the stationary phase used is a liquid, whereas in adsorption chromatography a solid is utilized as a stationary phase. The use of bonded organic phases for HPLC, which will be discussed later in this chapter, has blurred this distinction between partition and adsorption chromatography.

While there are various ways to separate chiral compounds, the common denominator among the available techniques is to create a chiral environment, in which the individual enantiomers can interact differently. It is therefore necessary to employ a chiral resolving agent (commonly called chiral selector) in separating chiral compounds. The use of chiral selectors can be divided into two broad categories: techniques that involve a chiral additive in the mobile phase and techniques that use a CSP, which is immobilized either directly in the column or immobilized on an appropriate support.[10] Another possibility is indirect separation where the analyte is derivatized prior to separation with a chiral reagent to produce diastereomers, which can be separated via achiral chromatographic methods. We will focus on direct resolution of enantiomers.

1.2 General Considerations for Mobile Phase Parameters

Mobile phase selection must be given careful consideration in HPLC. The degree of chemical purity should be seriously taken into consideration since solvent contamination can cause a variety of problems ranging from sensitivity, detection, and resolution, to name but a few. Particulates present in the solvent must be removed, for example by filtration. Also, when using aqueous solvent system such as buffers, bacterial contamination must be addressed. Furthermore, dissolved gases in the solvent can lead to problems with pumps and detectors. Dissolved oxygen, in particular, can cause detection problems, resulting in nonreproducible changes in absorbance when using a UV detector at low wavelengths (210 nm).[9] Additionally, it is crucial to select a mobile phase that is compatible with the stationary phase. Highly basic or acidic solvents, for example, can attack the silica support and destroy the packing of the column.

Other considerations in selecting a mobile phase in HPLC include solubility and polarity. The mobile phase selected must be able to dissolve the analyte and not allow the analyte to precipitate in the column. For example, a sample of protein dissolved in an aqueous buffer at a pH near its isoelectric point tends to be least soluble. Selecting a suitable composition of the mobile phase depends on the type of chiral reagent employed in the separation. When a chiral counterion is used such as in ion-pair chromatography, a mobile phase of low polarity such as methylene chloride is generally used to promote ion-pair formation.[8] On the other hand, when serum albumin is used as a chiral agent, the more appropriate choices for a mobile phase are aqueous solvents with phosphate buffers.

Viscosity is another factor to address when choosing between two mobile phases of similar polarity. Lower viscosity solvent tends to cause less pressure drop in the column, resulting in a greater degree of permeability through the stationary phase.[9] The less viscous mobile phase generally results in a faster separation time. The retention of the analyte in the column also can be increased or decreased by adjusting the concentration of the counterion or by adding a modifier. Adding a polar modifier such as 1-pentanol to the mobile phase will usually result in reduced stereoselectivity.

2 CHIRAL MOBILE PHASE ADDITIVES

One approach to provide a chiral enviroment is to add a chiral reagent to the bulk mobile phase.[11,12] Interestingly, many of the compounds used to synthesize the various CSPs were originally used as chiral mobile phase additives in TLC. The separation of chiral compounds by HPLC using a chiral selector in the mobile phase is based on the transient formation of diastereomeric complexes between the chiral selector and enantiomers of the analyte. The differences in the stability of the diastereomeric complexes, their solvation in the mobile phase, or binding of the diasteromeric complexes to the solid support, lead to differences in retention time. Generally, the enantiomer that forms the weakest diastereomeric complex will elute first, whereas the more stable complex will be retained in the column longer and will elute last.

The advantages of chiral mobile phase additives with respect to CSPs are the chiral selector added to the mobile phase can be easily changed; separations are less expensive since conventional achiral stationary phases are used; and a wider variety of chiral selectors are available as mobile phase additives than are available as CSPs. From a developmental standpoint, potential chiral selectors are much more readily investigated as a mobile phase additive than as a new CSP due to the difficulty and cost involved in synthesizing a CSP. Disadvantages associated with chiral mobile phase additives include interference with detection or recovery of the analyte, as well as impurity in the chiral additive can complicate the analytical separation.

In this chapter we will mainly focus on CSPs, their properties and structure, their applications, and the mechanism of separation.

3 CHIRAL STATIONARY PHASE BASED SEPARATIONS

Most chiral separations by HPLC are accomplished by direct resolution using a CSP. In this technique a chiral resolving agent is bound or immobilized to an appropriate support to make a CSP, and the enantiomers are resolved by the formation of diastereomeric complexes between the analyte and the CSP. Various types of CSP have been developed, including ligand exchange, the protein-based, carbohydrate-based, the Pirkle-type, cyclodextrin-based, and the macrocyclic antibiotic-based CSPs. Although each CSP is able to separate a variety of enantiomers, the application of each CSP is typically limited to structurally related chiral compounds. Table 1 gives an overview of the classes of HPLC CSPs.

4 LIGAND EXCHANGE

Among the earliest reports of chiral separations by LC were based on work done by Davankov et al.[13] using ligand exchange. Davankov et al. developed the first practical approach for resolution of amino acids in LC. These types of columns are available from Phenomenex (CA, USA), J.T. Baker (NJ, USA), and Regis Technologies, Inc. (IL, USA). Davankov et al. attached a chiral bidentate ligand, for example, proline or hydroxyproline, to a support, and added Cu^{2+} to the mobile phase as shown in Figure 1. Ligand exchange is a retention mode in which enantiomers are separated based on the differences in their charge and ionization constants.[14] Only ionized analytes can be separated by use of this technique. This type of chromatography

Table 1 Classes of HPLC CSPs

Chiral selector	Approximate year introduced	Typical mobile phase	Generally required analyte features
Ligand exchange	1971	Aqueous buffers	α-Hydroxy or α-amino acids near stereogenic center; some nonaromatic
Crown ethers	1975	Perchloric acid	Primary amines near stereogenic center; some non-aromatic
Pirkle phase	1979	Nonpolar organic; 2-propanol-hexane	π-Acid or π-basic moieties for charge transfer complex; hydrogen bonding; dipole stacking near stereogenic center
Protein phases	1982	Phosphate buffer	Aromatic near stereogenic center; organic acids/bases; cationic drugs
Cyclodextrin	1982	Aqueous buffers; polar organic	Strong inclusion complex or interaction between analyte hydrophobic moiety and mouth of cyclodextrin; hydrogen bonding near stereogenic center
Cellulosic/amylosic	1983	Nonpolar organic	Aromatic
Macrocyclic antibiotics	1994	Aqueous buffers; nonpolar and polar organic	Amines, amides, acids, esters; aromatic; hydrophobic moiety

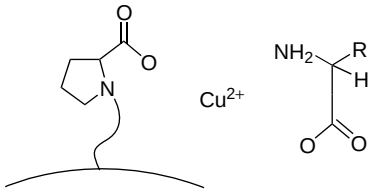

Figure 1 Structural diagram showing the interaction between the chiral selector and the analyte in the ligand exchange CSP.

involves the reversible formation of a metal complex by coordination of substrates, which act as ligands to the metal ion. Ligand exchange chromatography takes advantage of the ability of Cu^{2+} or other divalent metal ions to form diastereomeric complexes with chiral analytes. The complexes should be kinetically labile; that is, the association–dissociation time of the complexes should be faster than the chromatographic timescale. The separation of the enantiomers using this method is based on the difference in thermodynamic stabilities between the enantiomers and a divalent cation. The complex stability is highly dependent on the transition metal used. Cu^{2+}-complexes are generally more stable and are preferable for HPLC than Ni^{2+}- or Co^{2+}-complexes, which are used more prevalently in gas chromatography (GC).[15] Other racemates that can act as bidentate ligands, such as α-hydroxycarboxylic acids, can be resolved using this method. The selectivity and efficiency of this type of separation are greatly affected by pH, ionic strength of the mobile phase, and temperature.

Some of the drawbacks using the ligand-exchange approach is its limitation to chiral molecules that can coordinate selectively with the CSP. The dependence of chiral recognition on the formation of the diastereomeric complex imposes constraints on the proximity of the metal binding sites, usually either a hydroxy or an amine α to a carboxylic acid in the analyte. In addition, the copper-salt containing mobile phase absorbs in the UV region, decreasing detection sensitivity. It is interesting to note that the only commercially successful CSP for TLC is based on the ligand-exchange technique.

5 PROTEIN-BASED CHIRAL STATIONARY PHASES

Protein-based CSPs comprise a diverse and important class of CSPs for chiral separations. Within this class of CSP, there are a number of commercially available columns including α_1-acid glycoprotein (α_1-AGP), bovine serum albumin (BSA), human serum albumin, ovomucoid, and cellobiohydrolase. Protein CSPs are natural binders of small molecules and have numerous sites due to a variety of aromatic, basic, and acidic amino acids available to interact with the chiral analyte. These CSPs can be used in the reversed-phase mode and there are a limited number of variables to control in developing a separation method. Although the bonded-protein CSPs can be used to separate a wide variety of compounds, the applicability of each type of CSP is generally limited to a class of compounds and their related structures. For example, cellobiohydrolase CSP columns can most readily resolve primary amines and some secondary amines and amides whereas ovomucoid CSP is commonly employed in the separation of cyclic and secondary amines containing an aromatic group. The earlier versions of the protein-based CSP columns suffered from a lack of hardiness and longevity. They also tend to be more labile. The second generation of these columns showed improvements in stability and efficiency. Advantages of protein-based CSPs include low column capacity, limited solvent options and the inability to reverse the elution order of the analyte. The two most commonly employed protein CSPs are BSA and α_1-AGP. A summary of commercially available protein-based CSPs are listed in Table 2.[16]

5.1 Bovine Serum Albumin Chiral Stationary Phase

5.1.1 Properties and Structure

BSA is a globular protein having a molecular weight of 66 210.[17] The protein consists of 581 single chain amino acids, and 17 intrachain disulfide bridges, which connect the 34 half-cystines with the formation of 9 double loops.[18] BSA is a relatively acidic protein with an isoelectric point of 4.7. It is highly soluble in water but, like most globular proteins, it precipitates from solution at high salt concentrations. At pH 7.0 it has a net charge of −18. It exhibits hydrophobic character, and numerous examples of bindings of organic compounds to albumins have been reported.[19] Hydrophobic interaction has been shown to represent an important contribution to the total affinity of organic ligand to BSA. However, other

Table 2 Commercially available protein-based CSPs

Protein	Preferred analytes
BSA	Aromatic acids and anionic compounds
Human serum albumin	Aromatic acids and anionic compounds
α_1-AGP	Cyclic and secondary amines with aromatic group; some acids and neutral compounds
Ovomucoid	Cyclic and secondary amines with aromatic group; some acids and neutral compounds
Cellobiohydrolase	Primary amines and secondary amines; amides

interactions such as electrostatic interactions, hydrogen bonding, and charge-transfer contribute to the total affinity as well.

The first observation of the enantioselective properties of an albumin was made in 1958 by McMenamy and Oncley,[20] who demonstrated that the affinity for L-tryptophan exceeded that of D-tryptophan by a factor of 100 in a radioisotope labeling experiment and using a dialysis technique. This finding was later confirmed in 1973 by Stewart and Doherty when they demonstrated in an affinity chromatography on BSA immobilized to Sepharose that the D-enantiomer eluted under the conditions where L-enantiomer was retained.[21] Since then there have been extensive investigations of the chromatographic behavior of various racemic compounds under different mobile phase conditions. The BSA-silica columns for HPLC were introduced commercially in 1983.[22]

5.1.2 Applications

The mobile phases generally used in BSA columns are aqueous buffers with a limited pH range, mixed with organic modifiers such as 1-propanol. Retention and enantioselectivity properties on the BSA columns can be changed by varying parameters such as pH, ionic strength of the mobile phase, and temperature.[23] Various N-acylamino acids, amines, aromatic amino acids, sulfoxides and sulfoximine derivatives and related compounds, coumarin derivatives, and benzodiazepine derivatives have been shown to enantioresolve on BSA columns. The binding properties of BSA are highly dependent on the aqueous buffer system employed. This is partly due to the protein itself, that is, the overall charge and the conformation.[23] Consequently, it is possible to optimize enantioresolution by a proper selection of mobile phase parameters.

5.2 α_1-Acid Glycoprotein Chiral Stationary Phase

5.2.1 Properties and Structure

Human α_1-AGP has a molecular weight of approximately 41 000.[24,25] It is a single 181-amino acid peptide chain with five carbohydrate units, containing sialic acid, hexosamine, and neutral hexoses. The peptide chain comprises 55% of the molecular weight. α_1-AGP contains two cystine disulfide cross-linkages, which connect residues 5 and 147 and residues 72 and 164. In phosphate buffer, the glycoprotein has an isoelectric point of 2.7. Stability studies indicate that α_1-AGP column is very stable.[25] It can be used at elevated temperatures (25–80 °C), and can tolerate organic solvents such as pure 2-propanol. The columns also may be used over a wide pH range without denaturation. Storage of the columns can be done in a water–2-propanol mixture for 12 months with approximately 10% reduction in column retention characteristics.[25]

α_1-AGP column can be prepared by immobilization of the human plasma AGP on silica particles. The preparation is based on bonding the protein to diethylaminoethyl silica followed by a cross-linking procedure.[26] The resulting enamines from the cross-linking process are reduced to secondary amines with cyanoborohydride. The α_1-AGP column is commercially available under the trade name EnantioPac (LKB Products, Bromma, Sweden). Many of the advantages of the EnantioPac columns are associated with the reversed-phase mode of the column, which allows for direct injections as well as better regulation of retention and selectivity. Additionally, many basic and acidic drug substances without derivatization can be enantioresolved using the EnantioPac column. A second generation AGP column such as Chiral-AGP, on the other hand, uses covalent linkages to attach the protein to the silica support.[27,28] Chiral-AGP columns are thus more stable and resistant to hydrolytic attack. Chiral-AGP columns contain 5 μm silica particles instead of 10 μm as in EnantioPac, which gives Chiral-AGP columns a smaller pore volume and greater surface area, adding to their mechanical stability. Bonding capacity of the chiral phase as well as loading capacity of the solute are directly dependent on the amount of protein bound to the silica particles, which in turn depends on the bonding technique and the surface area of the silica support.

5.2.2 Applications

α_1-AGP CSP columns have wide application for the direct separation of many classes of pharmaceutical compounds, as well as amines, acids, and nonprotolytic compounds.[26,29] They are often used in the determination of enantiomers at low concentrations in biological fluids such as plasma and urine.[30,31] Concentration as low as 2 nmol L^{-1} plasma has been detected using a Chiral-AGP column and fluorescence detection. Enantioresolved acidic drugs on AGP columns include ibuprofen, ketoprofen, and naproxen. Basic drugs such as disopyramide, tropicamide, atropine, and homatropine have also been resolved on AGP columns.[32,33] In addition, AGP columns have been shown to baseline-resolve various other compounds including alprenolol, atenolol, ephedrine, and warfarin.

Concentration and properties of modifiers in the aqueous mobile phase play an important role in retention and selectivity of the solutes on AGP CSPs. Retention and enantioselectivity are found to decrease with increasing concentration of uncharged organic modifiers such as methanol, ethanol, 2-propanol, and acetonitrile.[27] This trend is not universal, however. For example, in the cases of mephenytoin and methylphenobarbital, chiral

resolution can be improved by adding uncharged modifier such as 2-propanol to the mobile phase. Enantioresolution of tertiary amines such as mepivacaine and bupivacaine, on the other hand, are not affected by the addition of organic modifier. Adding up to 8% propanol to the mobile phase results in no change in chiral resolution although the retention of these solutes decreases significantly under these conditions.[26]

Solute retention is generally found to increase when a modifier with the opposite charge is added to the solutes. Additions of same charge modifiers to the solutes tend to decrease retention time. The enantiomers of propiomazine, a fenthiazin derivative, which could not be resolved on an α_1-AGP column when using phosphate buffer at pH 7.55 as a mobile phase, was baseline resolved with a longer retention time upon adding 1 mM *N*,*N*-dimethyloctylamine (DMOA) to the mobile phase.[34] The changes observed in the selectivity are believed to result from the changes in α_1-AGP conformations, which in turn result from pH changes in the mobile phase. Changes in pH have been found to affect retention and selectivity for basic, acidic, and nonprotolytic compounds. While the separation of the weakly acidic hexobarbital and basic metoprolol increases with increasing mobile phase pH, enantioselectivity of 2-phenoxypropionic acid (pK_a 4.6) decreases at higher pHs.[31] Retention and enantioselectivity are also influenced by column temperature. Generally, retention, resolution, and separation factors decrease with increasing temperature. Separation efficiency, on the other hand, usually increases at higher temperatures.

5.2.3 Separation Mechanism

Although a great deal is known about the scientific basis of separation, that is, differences in binding between proteins and drug enantiomers, the mechanisms for enantiorecognition are not well understood. It is known that the native conformation of a protein can be changed by varying pH and adding organic modifier. The α_1-AGP molecule is believed to remain flexible in the column even after bonding with silica support. Hence, the addition of modifiers to the mobile phase is thought to affect the AGP molecule, inducing different binding properties.[26,31]

6 CARBOHYDRATE-BASED PHASES

Cellulose and amylose are among the most common natural polymers of this class of CSP. They are abundant in nature. In their native state, cellulose and amylose are poor chiral selectors. Improved chromatography and enantioselectivity of these chiral selectors can be achieved by derivatizing their hydroxyl groups with aromatic moieties through ester or carbamate linkages to decrease bulk polarity and provide additional interaction between the CSP and the solute.[35] The initial work on functionalized cellulose-based CSP was investigated in Europe, but it was Daicel, Ltd. (Tokyo, Japan) who commercialized a series of cellulose and amylose CSP.[16] A number of these commercialized stationary phases are shown in **(7–16)**. Structures **(7–10)** are derivatives of cellulosic carbamate CSPs. Structures **(11–14)** are derivatives of cellulosic ester CSPs. Structures **(15)** and **(16)** are coated CSPs consisting of derivatives of amylose. The 3,5-dimethylphenyl derivatives are the most widely used cellulose-based CSP (ChiraCel OD, Daicel) and amylose-based CSP (ChiraCel AD, Daicel), respectively.[16] Presently, there are nine cellulose derivative-based columns and two amylose-derivative CSPs produced by Daicel, Ltd. These CSPs are generally used in the normal phase mode, with the exception of two types which are conditioned for the reversed-phase mode. The molecular weight of a cellulose derivative affects chiral recognition and column efficiency. Cellulose and amylose columns have good capacity in the normal phase mode and are often used in the preparative-scale separations.[35]

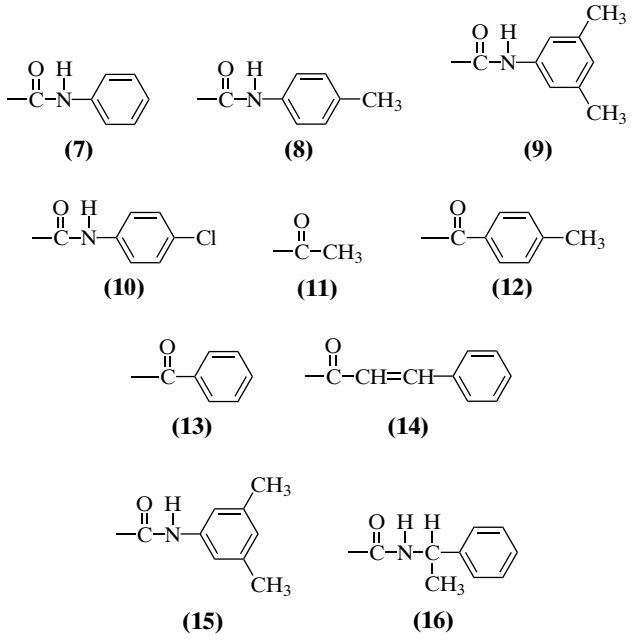

Derivatized cellulose and amylose are immobilized on a silica support. Mobile phases commonly employed are hexane: 2-propanol mixtures at various concentrations and occasionally more polar alcohols, such as methanol or ethanol, as well as aqueous-alcohol mixtures. These phases may not be used with solvents of intermediate polarity, for example, methylene chloride, acetone, tetrahydrofuran, and acetonitrile (over 1%). The main disadvantages of these phases are the limitations in

pressure and solvent used because these phases are not covalently bonded but merely adsorbed on the silica.

6.1 Cellulose- and Amylose-based Chiral Stationary Phases

6.1.1 Properties and Structures

Cellulose and amylose are naturally occurring polymers. They have highly organized structures, consisting of hydrogen-bonded chains of 1,4-linked D-(+)-glucose units. The glucose units on amylose are the same as the cyclodextrins and are connected via 1,4-β-linkages. However, the linkages in cellulose are 1,4-α-linkages.[12] Cellulosic and amylosic phases have been used extensively for chiral separations. The most frequently used cellulosic phase, in pure polymer form, is cellulose triacetate, which is commercially known as microcrystalline cellulose triacetate (MCA) (Daicel Chemical Industries, Ltd., Tokyo, Japan).[35] Cellulose triacetate is the product of heterogenous acetylation of microcrystalline cellulose.[17] Various cellulose ester derivatives on a silica gel support as HPLC columns are also commercially available (Daicel Chemical Industries, Ltd.). These columns appear to have good stability; however, because these chiral polymers are not attached covalently to the silica support, but are coated on silanized wide pore silica particles, caution must be taken in selecting mobile phase solvents.[36,37] Mobile phases such as hexane–2-propanol and alcohols can be used in these columns. The columns should be stored in hexane at low temperature (~4 °C).[16]

6.1.2 Applications

Cellulosic and amylosic phases have been successful in resolving structurally diverse compounds. MCA columns can be used to enantioresolve polar compounds such as amines, amides, imides, esters, and ketones.[38] Chiral selectivity seems to improve with a phenyl or a cycloalkyl group near the chiral center of the solute. Nonpolar solutes can also be separated in MCA columns, however, charged or dissociated compounds cannot be retained on these columns. Ethanol and methanol are the mobile phases generally used with these columns. The disadvantages of MCA columns include low efficiency and poor reproducibility. Additionally, a slow flow rate must be used with these columns. The Daicel columns are used for the separation of aromatic hydrocarbons with hydroxyl groups, carbonyls, sulfoxides, barbiturates, and β-blockers.[39,40]

6.1.3 Separation Mechanism

The chiral recognition sites on these carbohydrate-based CSPs are thought to be channels or grooves in the polymer matrix. The analytes are probably included in these grooves. This is evident by the enhanced enantiorecognition observed for many solutes when the steric bulk of the alcohol modifier is added to the mobile phase.[41,42] Chiral recognition appears to require the presence of an aromatic ring, for $\pi-\pi$ interactions, as well as polar sites of unsaturation and hydrogen bonding functionalities. Although the mechanism of separation by cellulose esters and carbamates is unclear, it is believed that chiral interactions result from the ester and urethane linkages.

7 PIRKLE CHIRAL STATIONARY PHASES

The first commercially available and successful CSP was developed by William Pirkle, an organic chemist with experience in chiral synthesis. The first generation of the ''Pirkle phases'', named after its creator, was introduced for LC in 1979.[43] The first generation Pirkle phase was based on the *N*-3,5-dinitrobenzoyl (DNB) moiety, which was immobilized on a silica support. Pirkle then introduced the concept that if a particular chiral selector is able to resolve certain analytes, it is more likely that a similar chiral selector to the original chiral selector could also be used to separate the analytes. Following this notion, a second generation of Pirkle phases, which were based on N-(2-naphthyl)-α-amino acids, was made.[44] Since these CSPs are produced synthetically, the stationary phase is often available in either absolute configuration.

7.1 Properties, Structures, and Applications

Using the above concept, Pirkle developed two commercially successful CSPs which are shown in **(17)** and **(18)**.[45] These phases consisted of N-(3,5-DNB) phenylglycine ionically **(17)** or covalently **(18)** attached to aminopropyl silica. These first Pirkle phases were very successful at enantioresolving compounds containing naphthyl moieties near the stereogenic center. Other phases based upon the DNB moiety were later developed **(19–21)**.[46] These included derivatives of leucine, ionically and covalently linked to an aminopropyl silica support. Since the DNB moiety is a π-acceptor, analytes for these CSPs should possess a π-donor group such as an aromatic ring with an alkyl, ether or amino substituent. These DNB phases have been successful at resolving a large number of compounds such as π-donor-substituted amines, alcohols, amino alcohols, and sulfoxides. While the covalently bonded version of these columns can be used with relatively polar mobile phases, the enantioselectivity and stability are much greater when using normal phase solvents. A second generation of Pirkle phases, which were based on N-(2-naphthyl)-α-amino acids were synthesized, providing reciprocity to the first DNB phases.[44] The

N-(2-naphthyl)-α-amino acids phases are π-donors, therefore analytes for these CSPs should possess a π-acceptor group. These columns showed selectivity toward analytes containing a 3,5-DNB group, such as 3,5-dinitrophenyl carbamates, ureas of chiral alcohols, and amines.

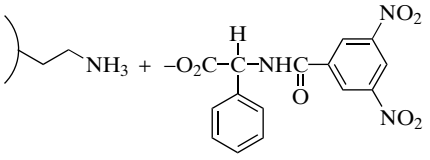

(17)

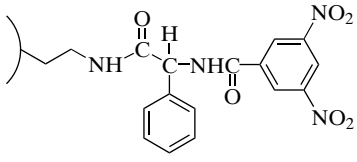

(18)

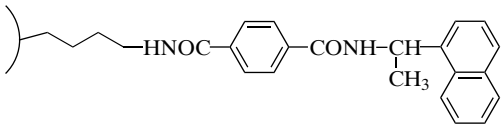

(19)

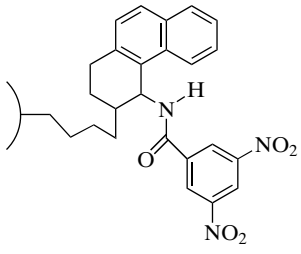

(20)

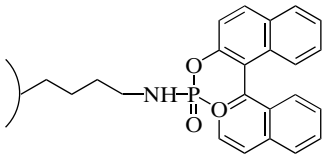

(21)

The most recent addition to the Pirkle phases is the commercial column marketed as the Whelk-O-1 phase (see **(22)**). The phase contains a chiral selector with a wedge-like chiral surface, with one edge containing a 3,5-DNB π-acidic moiety and the other edge comprising the π-basic tetrahydrophenanthrene ring system.[1,47] The two edges are linked together by the amide linkages. The presence of both the π-acid and π-basic, as well as the amide linkages, offers rigidity, and greater versatility than any other Pirkle phases. The Whelk-O-1 phase contains dipole stacking and hydrogen-bonding interaction sites, and imposes fewer structural constraints of the solutes and the mobile phase conditions. Because this chiral selector is small, it also promotes high bonded ligand densities in the stationary phase. This feature, coupled with the high enantioselectivity achieved with the phase, allows the Whelk-O-1 phase to be used in the preparative-scale separations.[1] Additionally, the chiral selector can interact with either the (*R*–*R*) or the (*S*–*S*) configuration of the solutes, giving the possibility of changing the elution order. This CSP has been shown to enantioresolve naproxen, warfarin, and its *p*-chloro analog, under the reversed-phase mode and nonaqueous reversed-phase condition.

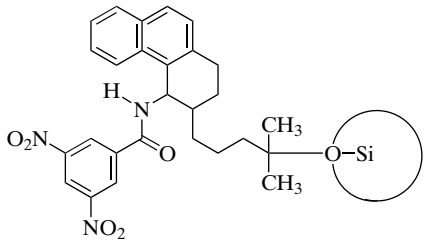

(22)

The mobile phases typically used with the Pirkle phases are the nonpolar organic mobile phases such as hexane, with 2-propanol or ethanol as organic modifiers. Under these conditions, retention of the solutes decreases as the mobile phase polarity increases, following the normal phase mode behavior. Since the mobile phase only weakly interacts with the CSP, and can be easily displaced by the solutes, specific enantiomeric interactions are promoted. The Pirkle phases are commonly used in the normal phase mode, although some have been used in the reversed-phase mode.

7.2 Separation Mechanism

Of all the commercially available CSPs for HPLC, the chiral recognition mechanism for the Pirkle phases is perhaps the best understood. Enantiorecognition on the Pirkle phases is believed to depend on complementary interactions between the chiral selector and the solutes.[44,47] The interactions involved are thought to include dipole–dipole interactions, π–π, hydrogen bonding, and steric hindrance as shown in Figure 2. These interactions contribute to the overall stability of the diastereomeric complexes formed between the individual enantiomer and the chiral selector on the stationary phase. The π–π interaction arises through the association of aromatic systems with complementary electron-withdrawing and electron-donating substituents

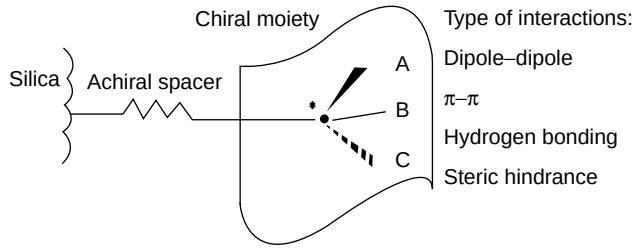

Figure 2 A schematic diagram showing enantiorecognition in the Pirkle phase due to hydrogen binding, dipole and charge-transfer interactions, and steric hindrance.

such as the nitro and alkyl groups, respectively. The aromatic system deficient in electrons is referred to as π-acid, in contrast to the electron-rich system, which is called π-base. In order to promote and enhance chiral recognition, functional groups are often introduced into the solute through achiral derivatization. The amines may be derivatized with 3,5-DNB chloride to create a π-acid aromatic group to the solute, thereby enhancing a diastereomeric complexation with a π-basic, (*R*)-*N*-(2-naphthyl)-alanine, chiral selector in the stationary phase.[44] Also, derivatizing the solute may have an additional advantage of increasing solute solubility.

It appears that three unique interactions from the stereogenic centers of the solutes and the chiral ligand in the stationary phase are necessary for enantiorecognition. This model describing three unique interactions is called the ''3-point interaction model''. It was first proposed by Dalgliesh.[48] Pirkle had recently restated the ''3-point rule'' as ''chiral recognition requires a minimum of three simultaneous interactions between the CSP and at least one of the enantiomers, with at least one of the interactions being stereochemically dependent''.[49] While no simple model can explain all the observations of chiral recognition, the three-point attachment model has greatly aided our understanding of the chiral recognition process.

8 CYCLODEXTRIN CHIRAL STATIONARY PHASES

Another general strategy to obtain optical resolutions employs the inclusion of the analyte into a chiral cavity. This strategy was first introduced by Dotsevi et al.[50] who used a crown ether bound to silica gel. Although the chiral crown ether phases were commercialized, for example, Diacel Crown-Pak, the majority of inclusion-based separations utilized bound cyclodextrin CSPs. Cyclodextrin CSPs immobilize the cyclodextrins to a solid support in such a way that the cyclodextrin is free to interact with solutes in solution. One of the earlier uses of cyclodextrins was as mobile phase modifiers in TLC.[51,52] Immobilizing cyclodextrin on a solid support yields several advantages over its use as a mobile phase modifier. For example, the solubility of β-cyclodextrin in the mobile phase is fairly low, and the efficiency obtained is often poor. Polymerized cyclodextrin gels were initially used as stationary phases in column chromatography but with limited success.[53,54] The use of cyclodextrins in HPLC was introduced when Fujimura et al.[55] and Kawagucki et al.[56] developed a method to attach different cyclodextrins to silica gel via ethylene diamine linkages. Unfortunately, these phases were hydrolytically unstable and unsuitable for commercialization. A new cyclodextrin packing, which contained no nitrogen linkages, was later introduced by Armstrong.[57] The new columns were thought to be more widely useful and commercially feasible. The nitrogen linkages were replaced by a 6–10 atom spacer. The linkages and the cyclodextrins were hydrolytically stable, and the stationary phases had increased loading and efficiency. This stable cyclodextrin-bonded phase is sold commercially under the trade name Cyclobond (Advanced Separation Technologies, Whippany, New Jersey). This stationary phase was the first chiral CSP to be developed for use in the reversed-phase mode and has been shown to successfully resolve diverse isomers formerly thought to be difficult to separate.[58–60]

8.1 Properties and Structures

Cyclodextrins are cyclic oligosaccharides naturally formed by the action of *Bacillus macerans* on starch.[61,62] They can also be produced by use of the enzyme cyclodextrin transglycosylase from starch.[63] There have been several procedures used to control the ratios of cyclodextrins produced. One is the addition of a substance to the reaction mixture to favor one form of cyclodextrin over another. For instance, the addition of 1-decanol and 1-nonanol favorably induces the formation of α-cyclodextrin.[64] In the presence of hexane or toluene, on the other hand, β-cyclodextrin is more strongly promoted. Both α- and β-cyclodextrins are simultaneously produced in the presence of 1-heptanol.[61,64] The native cyclodextrins contain 6 to 12 glucose units, which are connected to each other by α-(1,4) linkages. There are three commercially available cyclodextrin homologs: α-cyclodextrin (cyclohexamolyose), β-cyclodextrin (cycloheptamylose), and γ-cyclodextrin (cyclootamylose) as shown in **(23–25)**, respectively. The larger cyclodextrin homologs must be individually produced and isolated.

The Greek letter preceding the name of the cyclodextrin designates the number of glucose subunits: $\alpha = 6$, $\beta = 7$, and $\gamma = 8$. Each glucose unit on the cyclodextrin is chiral. Cyclodextrins are toroidal in shape with

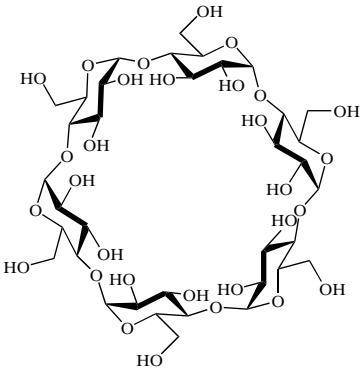

(23)

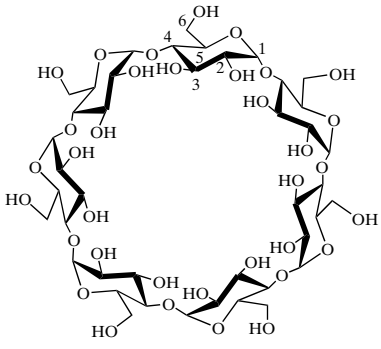

(24)

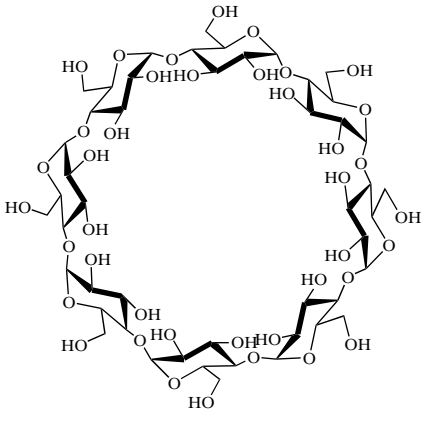

(25)

a relatively hydrophobic internal cavity. The 2-hydroxyl groups at the rim of the cyclodextrin cavity project in a clockwise direction. The cyclodextrins exterior is relatively hydrophilic due to the presence of the primary and secondary hydroxyl groups. The primary hydroxyl group on C-6 is free to rotate and can partially block the cyclodextrin cavity at the narrow end. The ability of cyclodextrins to form inclusion complexes with a variety of guest molecules and ions makes the cyclodextrins particulary useful for chiral separations (Figure 3).[65]

The cyclodextrins have a number of chemical and physical properties in common. Among these properties are:

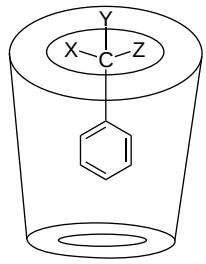

Figure 3 A diagram depicting a hydrophobic inclusion complex between a cyclodextrin cavity and a chiral analyte.

cyclodextrins are nonreducing; glucose is the only product of acid hydrolysis; molecular weights are always integral numbers of 162.1, the value for glucose; cyclodextrins are nontoxic; cyclodextrins do not appreciably absorb UV or visible light. The cyclodextrin molecules as a whole are hydrophilic making these CSPs completely compatible with aqueous mobile phases. Since cyclodextrin CSPs are compatible with aqueous mobile phases, method development is aided by the use of buffers and organic modifiers in adjusting column selectivity.

8.2 Applications

The most commonly used cyclodextrin in HPLC has been the native β-cyclodextrin **(24)**. A variety of water soluble–insoluble compounds can fit into the hydrophobic cavity, forming reversible host–guest inclusion complexes.[66] It is well known that molecules possessing an aromatic group such as naphthyl or biphenyl in close proximity to the stereogenic center are particularly amenable for separation.[63] The β-cyclodextrin column has shown excellent selectivity for amino acid derivatives. For example, the native β-cyclodextrin phase has been used to resolve a variety of racemates including dansylamino acids, β-naphthyl ester derivatives of amino acids, and β-naphthylamide. It is thought that underivatized amino acids are perhaps too small to bind tightly to the cyclodextrin cavity, and subsequently cannot be enantioresolved. When a dansyl group is incorporated in the amino acid, however, strong inclusion complexes between the cyclodextrin and the derivatized amino acid can be formed, and baseline resolution is achieved. Either the carboxyl group or the amino group on the amino acid can be derivatized. Derivatization of both groups appears to result in reduced chiral recognition. Enantiomeric derivatives of ferrocene, ruthenocene, and osmocene, as well as several racemic and diastereomeric 2,2-binaphthyldiyl crown ethers, have also been successfully resolved on the β-cyclodextrin phase.[67,68] It should be noted that although the crown ethers do not contain a chiral carbon, they still exist as enantiomers because of the staggered position of adjacent naphthyl rings.[68]

α-Cyclodextrin is suitable for separating small molecules such as tryptophan, phenylalanine, tyrosine, and their analogs. γ-Cyclodextrin has been used for the separation of di- and tripeptides.

Cyclodextrin-bonded phases provide a rapid and specific technique for the pharmacological evaluation of racemic drugs. Chiral drugs such as ketoprofen, mephenytoin, chlorpheniramine, and the barbiturates mephobarbital and hexobarbital, have been separated by the cyclodextrin column.[69] Many diastereomers, geometric isomers, and epimers, can be resolved using cyclodextrin phases as well. For example, *cis*- and *trans*-benzo[*a*]pyrene were completely separated on a γ-cyclodextrin column. In addition, cyclodextrin phases are of great utility in enantioresolving structural isomers, such as the ortho-, meta-, and para-isomers of nitroaniline, xylene, cresols, nitrophenols, and substituted benzoic acids.[70] An extremely sensitive test of optical purity can be achieved with the cyclodextrin columns. Detection of as little as 0.2% of one enantiomer in a racemic mixture has been shown to be feasible using the cyclodextrin column.[63]

Limitations with the selectivity of the native cyclodextrins led to the development of various functionalized cyclodextrin-based stationary phases. These derivatized cyclodextrin phases include acetylated, sulfated, 2-hydroxypropyl, 3,5-dimethylphenylcarbamoylated, and 1-naphthylethylcarbamoylated cyclodextrin.[1,16] Since each of the glucose molecules on a cyclodextrin provides three hydroxyl groups to which a substituent may be attached, there are multiple sites on the cyclodextrin available for derivatization. Typically, each β-cyclodextrin, which contains 21 hydroxyl groups, has a degree of substitution ranging from 3 to 10 derivatized groups per cyclodextrin molecule. The groups on these functionalized cyclodextrins enhance chiral recognition by a variety of means. In some cases, the groups mainly serve to extend the chiral cyclodextrin cavity and in others they may provide alternate interaction sites. For example, in the case of the naphthylethylcarbamoylated cyclodextrin the naphthyl ring provides a π-basic site while the carbamate linkage provides additional hydrogen bonding and dipolar interaction sites. The sulfate group on the sulfated cyclodextrin provides the potential for ion-pair formation while the two hydroxypropyl substituents add additional stereogenic centers to the cyclodextrin. One feature that is particularly advantageous with these functionalized cyclodextrins is that they often exhibit enantioselectivity in the hydroorganic reversed-phase as well as normal phase, and polar organic mobile phase conditions. This is unusual in that each mode can provide chiral separations for compounds, which are not resolved under any other types of mobile phase conditions.

Most chiral separations reported to date using the cyclodextrin columns have been accomplished in the reversed-phase mode using aqueous buffers with a small amount of organic modifiers. However, polar organic mobile phases have gained in popularity recently because of their ease of sample removal and reduced rate of column degradation.[1,71] In these cases, the solutes are thought to occupy the top of the cyclodextrin cavity like a ''lid'', rather than fit in the cavity because of the more nonpolar mobile phase. The commonly used mobile phases in the normal phase mode of separations are hexane–isopropanol. Separation in the normal phase mode is thought to be analogous to those of diol columns.

8.3 Separation Mechanism

The separation mechanism thought to be responsible for the chiral recognition observed with the cyclodextrin-bonded phases is based upon the formation of inclusion complex between the hydrophobic moiety of the solute and the hydrophobic cavity of the cyclodextrin.[63,69] This has been demonstrated by performing a separation in a normal phase mode using hexane–isopropanol mobile phase on a β-cyclodextrin column, a system by which the solute is excluded from the cyclodextrin cavity because a more hydrophobic mobile phase occupies the cavity. The inclusion complex formed between the solute and cyclodextrin cavity should also be a tight fit. This is apparent by the fact that better enantioresolution using a β-cyclodextrin column is obtained when the solute is about the size of biphenyl or naphthalene than when smaller or larger size solutes are used.[72] It is possible that smaller compounds may not be rigidly held in the cavity, resulting in a less stable diastereomeric complex and subsequently less successful separation. Moreover, the stereogenic center of the solute or the substituent attached to the stereogenic center of the solute should be able to interact with the rim of the cyclodextrin cavity in order to create strong interactions necessary for chiral separations.

The restricted conformational freedom of the 2- and 3-hydroxyl groups at the cyclodextrin rim is thought to play an important role for the chiral recognition properties in the cyclodextrins. This is evident by the modeling studies using (*R*)- and (*S*)-propanolol with β-cyclodextrin.[69] The (*R*)- and (*S*)-propanolol were placed inside the cyclodextrin cavity in an identical manner, and with the 3-hydroxyl group of the cyclodextrin lined up for an ideal hydrogen bonding with the hydroxyl groups on (*R*)- and (*S*)-propanolol as shown in Figure 4. An important difference which exists between the two complexes formed is at the secondary amine group. In the (*R*)-complex, the bond distances between the nitrogen on the amine group of the solute and the 2-

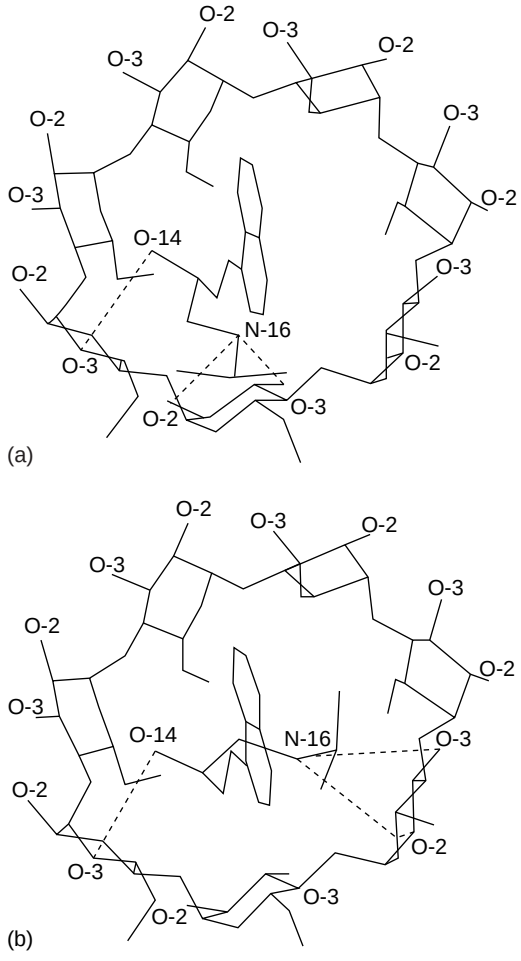

Figure 4 Computer projections of inclusion complexes of (a) *d*-propranolol and (b) *l*-propranolol in β-cyclodextrin from X-ray crystallographic data. Dotted lines represent potential hydrogen bonds. The configurations shown represent the optimal orientation of each isomer on the basis of the highest degree of hydrogen bonding and complexation. (Reprinted by permission from the American Association for the Advancement of Science.[69])

and 3-hydroxyl groups on the cyclodextrin are 0.33 and 0.28 nm, respectively. In contrast, the (*S*)-complex has the respective bond distances of 0.38 and 0.45 nm. The model suggests that (*R*)-propanolol is positioned more favorably for hydrogen bonding with the cyclodextrin, and can preferentially interact with the chiral selector in a way that the (*S*)-isomer cannot. This difference in their interaction results in enantiorecognition.

9 THE MACROCYCLIC ANTIBIOTIC CHIRAL STATIONARY PHASES

The macrocyclic antibiotics and their derivatives have formed one of the newest and perhaps more versatile class of chiral selectors in HPLC.[73] They are excellent chiral selectors, and can be bonded to silica gel via linkage chains using a variety of chemistries. They can be used in either the reversed-phase or the normal phase mode. Enantioselectivities in each mode have been shown to be different. Furthermore, the macrocyclic antibiotics can be derivatized to alter their selectivity. The bonded macrocyclic antibiotic CSPs resemble the protein-based CSPs in many ways. However, the macrocyclic antibiotic CSPs are more stable and have greater capacities than the protein-based CSPs. There are literally hundreds of macrocyclic antibiotics that have yet to be investigated. Their versatility in enantioresolving chiral analytes is becoming well known. Thus far, the two most successful structural types of antibiotics used for separations are *ansa* compounds and glycopeptides. These two classes of macrocyclic antibiotics will be the focus of our discussion.

9.1 Properties and Structures

The ansamycins include *ansa* compounds such as rifamycin B, rifamycin SV, rifamide, and rifampim. They are isolated from the fermentation broth of the bacteria *Nocardia mediterranei*.[74] These compounds contain a characteristic *ansa* structure, a ring structure or chromophore that is spanned by an aliphatic chain. As shown in Figure 5, the rifamycins differ from each other only in the type and the location of the substituents on their naphthohydroquinone ring. Of the rifamycins, rifamycin B is one of the more common members and has been utilized as a CSP for HPLC. Rifamycin B has nine stereogenic centers, four hydroxyl groups, one carboxylic acid moiety, and one amide bond. Rifamycin B has a molecular weight of 756 and is more soluble in ethanol and methanol

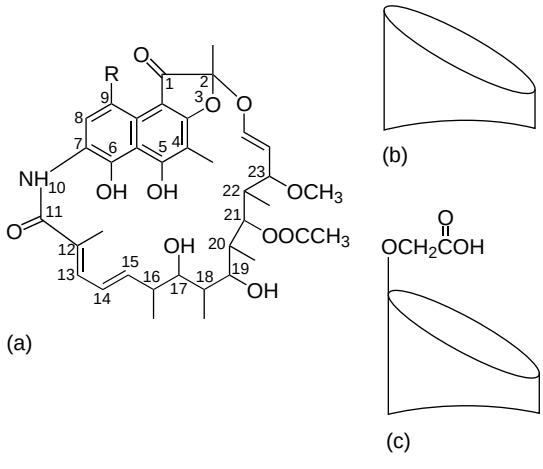

Figure 5 (a) Structure of rifamycin B (R = $-CH_2COOH$) and SV (R = $-OH$). Simplified schematic showing (b) rifamycin SV and (c) rifamycin B.

Table 3 Comparison of the physico-chemical properties of the macrocyclic antibiotics

Characteristic	Rifamycin B	Vancomycin	Thiostrepton	Teicoplanin
Molecular weight	755	1449	1665	1877
Stereogenic centers	9	18	17	23
No. of OH groups	4	9	5	15
No. of amine groups	0	2	1	1
No. of carboxylic acids	1	1	0	1
No. of amido groups	0	7	10	7
No. of aromatic groups	2	5	7	7
No. of methyl esters	1	0	0	0
pI	–	7.2	–	4.2

than in water. The substituent on rifamycin B is an oxyacetic acid ($-OCH_2COOH$), thus the carboxylic and the hydroxyl groups on rifamycin B are ionizable and rifamycin B can exist as a dibasic acid.

The glycopeptide macrocyclic antibiotics vancomycin, thiostrepton, and teicoplanin are produced by *Streptomyces orientalis*, *Streptomyces azureus*, and *Actinoplanes teichomyceticus*, respectively.[75,76] The glycopeptide macrocyclic antibiotics are soluble in water, buffers, and acidic aqueous solutions, and are less soluble at neutral pH. They are moderately soluble in polar aprotic solvents, such as dimethyl sulfoxide (DMSO) and dimethyl formamide (DMF), but are insoluble in most nonpolar organic solvents. The three glycopeptide macrocyclic antibiotics have many structural features and physico-chemical properties in common. Table 3 shows a comparison of the macrocyclic antibiotics.

Among the three glycopeptide macrocyclic antibiotics discussed, vancomycin **(26)** is the smallest in size (molecular weight 1449, 18 stereogenic centers) while thiostrepton **(27)** is of intermediate size (molecular weight 1665, 17 stereogenic centers), and teicoplanin **(28)** is the largest in size (molecular weight 1877, 23 stereogenic centers).[73,77,78] The macrocycles contain both ether and peptide linkages. The aglycon portion of vancomycin and teicoplanin consists of two chloro-substituted aromatic rings. In thiostrepton the analogous portion lacks a chloro-substituent. There is a carboxylic moiety on the aglycon of both vancomycin and teicoplanin. Thiostrepton contains in its aglycon portion five thiazole rings and one quinoline ring. The type and number of pendant carbohydrate moieties vary among the glycopeptide macrocyclic antibiotics as well. Vancomycin contains D-glucose and vancosamine. Teicoplanin, on the other hand, has three attached monosaccharides, two D-glucosamines and one D-mannose. Thiostrepton has no attached carbohydrate moieties.

In addition to the fused macrocyclic rings and the attached pendant carbohydrate moieties, the glycopeptide macrocyclic antibiotics have various analogous ionizable groups, which are thought to control their charge and enantioselective properties. For example, vancomycin and thiostrepton have a secondary amine unlike teicoplanin, which contains only a primary amine group.

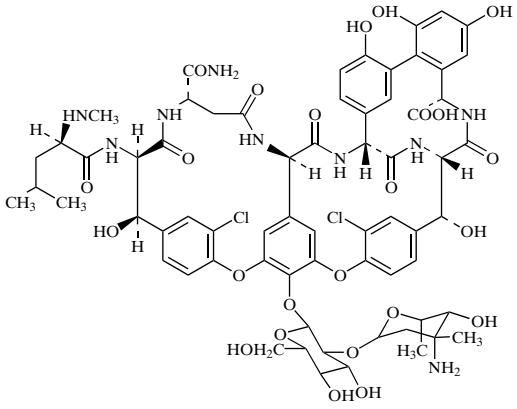

(26)

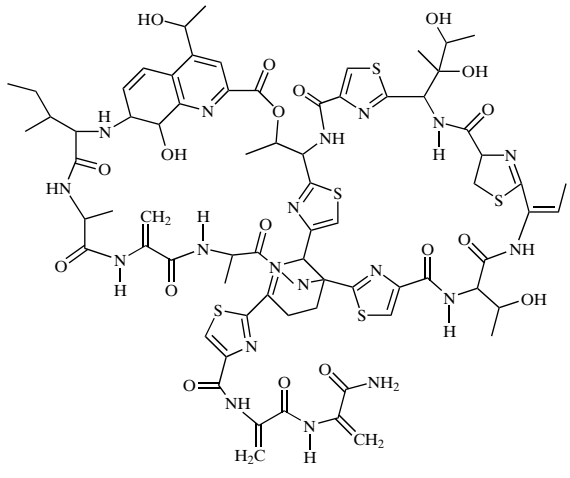

(27)

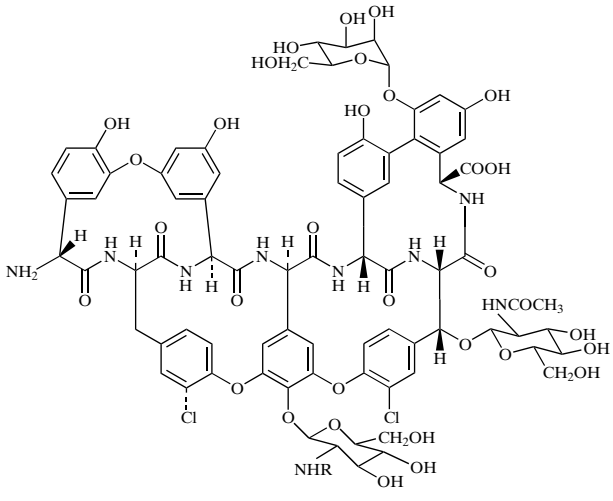

(28)

9.2 Applications

Currently, there are two types of glycopeptide-based CSPs available commercially: vancomycin and teicoplainin. The macrocyclic antibiotics can be attached to silica gel via epoxy-terminated organosilane, as has been used with the cyclodextrins, or by reacting the macrocyclic antibiotics with an isocyanate-terminated organosilane in anhydrous DMF. The macrocyclic antibiotics are multimodal CSPs. These CSPs can be useful in preparative-scale separations. They can be used effectively in the reversed-phase, normal phase, and polar organic modes. Different compounds are resolved in each mode.[77] When changing from one mode to another, there does not appear to be any deleterious effects to the stationary phases or any irreversible changes in the enantioselectivity. The macrocyclic antibiotics have been shown to exhibit excellent enantioselectivity for native amino acids, peptides, α-hydroxycarboxylic acids, and a variety of neutral analytes including cyclic amides and amines. Neutral molecules, for example, devrinol and 5-methyl-5-phenylhydantoin, seem to show considerable enantioselective retention on the macrocyclic antibiotic CSPs. The chiral selectors appear to be stable and the integrity of the CSP is excellent in all separation modes. For example, compounds resolved in reversed-phase mode using vancomycin-bonded phase column include bromacil, devrinol and coumachlor enantiomers whereas in normal mode the same column showed enantioselectivity for 3-[2-(2-bromoacetamido)acetamido]-PROXYL and *N*-(3,5-DNB)-α-methylbenzalamine.[77] Of the available modes, the reversed-phase mode is probably more widely employed when using the macrocyclic antibiotics as a CSP.

In comparison to the other two CSPs most often used in the reversed-phase mode, which are the protein columns and the cyclodextrin-based columns, the macrocyclic antibiotic stationary phases have similarities and differences with both of these classes. Similar to the protein columns, the macrocyclic antibiotics contain ionizable groups, and consequently their charge and possibly conformation can vary with the pH of the mobile phase. The effect of mobile phase pH, however, is not observed when using the cyclodextrin CSPs because the cyclodextrins remain neutral at pH 3–8.5, which is the pH range commonly used in reversed-phase HPLC. Like the protein phases and the cyclodextrin phases, it is not necessary to derivatize racemic analytes before separation on the macrocyclic antibiotic CSPs. Most macrocyclic antibiotics contain peptide bonds, some have carbohydrate moieties, and a few, such as vancomycin, have both features.

Optimization of reversed-phase separations can be achieved in a similar way as for cyclodextrin-based columns. Retention can be changed by varying the amount of organic modifier, such as acetonitrile and methanol, and the mobile phase pH. For example, using teicoplanin CSP, with 100% triethylammonium acetate pH 7 buffer, 5-methyl-5-phenylhydantoin has a $k > 4$, where k is the capacity factor, but with acetonitrile–1% triethylammonium acetate pH 7 buffer (60 : 40 v/v), k is nearly zero or the analyte elutes near the dead volume of the column.[78] A plot of retention vs mobile-phase composition shows the same U-shaped curve, which has been observed previously for chiral solutes on cyclodextrin-based columns and for proteins and peptides on reversed-phase columns. Typical normal phase retention behavior is found on a teicoplanin column. Retention and enantioselectivity are highest when the mobile phase consists mainly of the nonpolar component, such as hexane. Both retention and selectivity decrease as the amount of ethanol in the mobile phase is increased.

In addition, efficiency and selectivity can be affected by ionic strength, type of buffer, flow rate, and other additives. On a teicoplanin column under a normal phase mode, flow rate does not affect enantioselectivity or α, but does affect the separation efficiency. This is reflected by the inverse relationship between resolution (R_S) and flow rate. Decreasing the flow rate increases the resolution. However, decreasing the flow rate below 0.5 mL min^{-1} does not continue to produce an increase in resolution. It is believed that this is the result of poor stationary phase mass transfer. Temperature also has been shown to affect chiral separations using the macrocyclic antibiotic CSPs. Lowering the temperature appears to enhance enantioseparations by increasing α. Using the reversed-phase mode on a vancomycin CSP, the separation of *N*-carbamyl-D,L-phenylalanine increased α of 1.00 and R_S of 0 to α of 1.23 and R_S of 1.0 upon decreasing the temperature from 45 °C to 15 °C.[77] Lowering the temperature in general enhances resolution and selectivity but concurrently causes an increase in retention time.

9.3 Separation Mechanism

It is thought that the size, shape, geometric arrangement, as well as the functional groups present in the macrocyclic antibiotics, help to determine their enantioselective properties. In addition to the charge interactions, the retention behavior of organic analytes in the reversed-phase mode indicates the importance of hydrophobic interactions between the analyte and the macrocyclic antibiotics. However, whether or not there is a true inclusion complex between the chiral selector and the analyte as in the case of the cyclodextrins or just an association of the analyte with the hydrophobic pocket is not known. Hydrogen bonding is believed to be essential for enantiorecognition, particularly in acetonitrile mobile phases.

10 FUTURE TRENDS

The development and success of the various CSPs have had a profound impact not only on the field of analytical separations but also the pharmaceutical industries, and other industries that manufactured or marketed chiral compounds. CSPs have been beneficial to the understanding of molecular recognition, chromatographic theory, and mechanism. In the future, chiral separations have the potential to aid in our understanding of the chemistry behind biological processes, diseases, and probing intermolecular interactions. Mechanistic studies of chiral compounds as well as CSPs will undoubtedly continue in the future.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support by the National Institutes of Health (Grant R15 AI41182).

ABBREVIATIONS AND ACRONYMS

AGP	Acid Glycoprotein
BSA	Bovine Serum Albumin
CSP	Chiral Stationary Phase
DMF	Dimethyl Formamide
DMOA	Dimethyloctylamine
DMSO	Dimethyl Sulfoxide
DNB	Dinitrobenzoyl
DOPA	3-(3,4-Dihydroxyphenyl)alanine
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
MCA	Microcrystalline Cellulose Triacetate
TLC	Thin-layer Chromatography
UV	Ultraviolet

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Peptide Diastereomers, Separation of

Pharmaceuticals and Drugs **(Volume 8)**
Chiral Purity in Drug Analysis

REFERENCES

1. A.M. Stalcup, 'Chiral Separations', in *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edition, John Wiley & Sons, Inc., New York, 123–154, 1998.
2. S. Allenmark, *Chromatographic Enantioseparation: Methods and Applications*, 2nd edition, Ellis Horwood, New York, 1991.
3. M.G. Loudon, *Organic Chemistry*, 3rd edition, Benjamin/Cummings Publishing Company, Inc., California, 1995.
4. S. Ahuja (ed.), *Chiral Separations by Liquid Chromatography*, ACS Symposium Series, American Chemical Society, Washington, DC, 1991.
5. A.C. Guyton, J.E. Hall, *Textbook of Medical Physiology*, W.B. Saunders Company, Philadelphia, 1996.
6. D.E. Haines (ed.), *Fundamental Neuroscience*, Churchill Livingstone, New York, 1997.
7. L.S. Goodman, L.E. Limbird, P.B. Milinoff, A.G. Gilman, J.G. Hardman (eds.), *The Pharmacological Basis of Therapeutics*, 9th edition, McGraw Hill Text, New York, 1996.
8. S. Ahuja (ed.), *Chiral Separations: Applications and Technology*, American Chemical Society, Washington, DC, 1997.
9. T.J. Bruno, *Chromatographic and Electrophoretic Methods*, Prentice Hall, Englewood Cliffs, 1991.
10. T.J. Ward, 'Chiral Media', *Anal. Chem.*, **66**, 633A–640A (1994).
11. D.W. Armstrong, 'Optical Isomer Separations by Liquid Chromatography', *Anal. Chem.*, **59**, 84A–91A (1987).
12. K.L. Williams, A.M. Stalcup, 'Chromatography with Cyclodextrin-based Stationary Phases', in *Molecular Interactions in Bioseparations*, ed. T.T. Ngo, Plenum Press, New York, 189–202, 1993.
13. V.A. Davankov, A.A. Kurganov, A.S. Bocklov, in *Advances in Chromatography*, eds. J.C. Giddings, E. Grushka, J. Cazes, P.R. Brown, Marcel Dekker, New York, 71–116, Vol. 22, 1983.
14. B.A. Bidlingmeyer, *Practical HPLC Methodology and Applications*, John Wiley & Sons, Inc., New York, 1992.
15. V.A. Davankov, A.V. Semechkin, 'Ligand Exchange Chromatography', *J. Chromatogr.*, **141**, 313–353 (1977).
16. D.W. Armstrong, 'The Evolution of Chiral Stationary Phases for Liquid Chromatography', *LC–GC*, S20–S28 (1997).
17. T.J. Ward, 'Biopolymers Analytical Techniques', in *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edition, John Wiley & Sons, Inc., New York, 187–208, 1992.
18. Th. Peters, Jr, 'Serum Albumin', in *The Plasma Proteins*, ed. F.W. Putman, Academic Press, New York, 133–172, 1975.
19. M.C. Meyer, D.E. Guttman, 'The Binding of Drugs by Plasma Proteins', *J. Pharm. Sci.*, **57**(6), 895–901 (1968).
20. R.H. McMenamy, J.L. Oncley, 'The Specific Binding of L-tryptophan to Serum Albumin', *J. Biol. Chem.*, **233**, 1436 (1958).
21. E.J. Cohn, J.T. Edsall, *Proteins, Amino Acids and Peptides*, Reinhold Publishing Corp., New York, 1943.
22. S. Allenmark, B. Bomgren, H. Boren, 'Direct Resolution of Enantiomers by Liquid Affinity Chromatography

on Albumin-agarose Under Isocratic Conditions', *J. Chromatogr.*, **237**, 473–477 (1982).
23. S. Allenmark, 'Optical Resolution by Liquid Chromatography on Immobilized Bovine Serum Albumin', *J. Chromatogr.*, **9**(2&3), 425–442 (1986).
24. K. Schmid, 'Structure and Function of Glycopeptides', in *The Plasma Proteins*, ed. F.W. Putman, Academic Press, New York, 163–211, 1975.
25. G. Schill, I.W. Wainer, S.A. Barkan, 'Chiral Separation of Cationic Drugs on α_1-Acid Glycoprotein Bonded Stationary Phase', *J. Liq. Chromatogr.*, **9**(2&3), 641–646 (1986).
26. J. Hermansson, 'Liquid Chromatographic Resolution of Racemic Drugs Using a Chiral α_1-Acid Glycoprotein Column', *J. Chromatogr.*, **298**, 67–78 (1984).
27. Chiral-AGP, Application Note No. 1, Chrom Tech AB, Norsburg, Sweden, 1988.
28. J. Hermansson, 'Enantiomeric Separation of Drugs and Related Compounds Based on their Interaction with α_1-Acid Glycoprotein', *Trends Anal. Chem.*, **8**(7), 251–259 (1989).
29. J. Hermansson, G. Schill, 'Resolution of Enantiomeric Compounds by Silica-bonded Alpha$_1$-acid Glycoprotein', in *Chromatographic Chiral Separations*, eds. M. Zief, L.J. Crane, Marcel Dekker, New York, 245–281, Vol. 40, 1988.
30. M. Enquist, J. Hermansson, 'Separation and Quantitation of (*R*)-and (*S*)-atenolol in Human Plasma and Urine Using an α_1-AGP column', *Chirality*, **1**(3), 209–215 (1989).
31. B.A. Persson, K. Balmer, P.O. Lagerstrom, G. Schill, 'Enantioselective Determination of Metoprolol in Plasma by Liquid Chromatography on a Silica-bonded α_1-Acid Glycoprotein Column', *J. Chromatogr.*, **500**, 629–636 (1990).
32. E. Arvidsson, S.O. Jansson, G. Schill, 'Chiral Separations of Atropine and Homotropine on an α_1-Acid Glycoprotein-bonded Stationary Phase', *J. Chromatogr.*, **506**, 579–591 (1990).
33. J. Hermansson, M. Eriksson, 'Direct Liquid Chromatographic Resolution of Acidic Drugs Using a Chiral α_1-Acid Glycoprotein Column (Enantiopac)', *J. Liq. Chromatogr.*, **9**(2&3), 621–639 (1986).
34. J. Hermansson, 'Direct Liquid Chromatographic Resolution of Racemic Drugs by Means of α_1-Acid Glycoprotein as the Chiral Complexing Agent in the Mobile Phase', *J. Chromatogr.*, **316**, 537–546 (1984).
35. K. Hostettmann, A. Marston, M. Hostettmann, *Preparative Chromatography Techniques*, 2nd edition, Springer, Berlin, 1998.
36. Y. Okamoto, M. Kawashima, K. Yamamoto, K. Hatada, 'Chromatographic Resolution and Useful Chiral Packing Materials for High-performance Liquid Chromatographic Resolution: Cellulose Triacetate and Tribenzoate Coated on Macroporous Silica Gel', *Chem. Lett.*, **5**, 739–742 (1984).
37. Y. Okamoto, R. Aburatani, T. Kukumoto, K. Hatada, 'Chromatographic Resolution. XVII. Useful Chiral Stationary Phases for HPLC. Amylose Tris(3,5-dimethylphenylcarbamate) and Tris(3,5-dichlorphenylcarbamate) Supported on Silica Gel', *Chem. Lett.*, **9**, 1857–1860 (1987).
38. G. Blaschke, 'Chromatographic Resolution of Chiral Drugs on Polyamides and Cellulose Triacetate', *J. Liq. Chromatogr.*, **9**(2&3), 341–368 (1986).
39. D.W. Armstrong, S.M. Han, 'Evaluation of Dye–Micelle Binding Constants Using Diffusion Sensitive Band Broadening Effects', *Crit. Rev. Anal. Chem.*, **19**(3), 175 (1988).
40. M. Zief, L.J. Crane (eds.), *Chromatographic Chiral Separations*, Marcel Dekker, New York, Vol. 40, 1988.
41. G. Hesse, R. Hagel, 'A Complete Separation of a Racemic Mixture by Elution Chromatography on Cellulose Triacetate', *Chromatographia*, **6**, 277–278 (1973).
42. E. Francotte, R.M. Wolf, D. Lohmann, R. Mueller, 'Chromatographic Resolution of Racemates on Chiral Stationary Phases, I. Influence of Supermolecular Structure of Cellulose Triacetate', *J. Chromatogr.*, **347**, 25–31 (1985).
43. W.H. Pirkle, D.W. House, 'Chiral High-pressure Liquid Chromatographic Stationary Phases. 1. Separation of Enantiomers of Sulfoxides, Amines, Amino Acids, Alcohols, Hydroxyacids, Lactones, and Mercaptans', *J. Org. Chem.*, **44**, 1957–1960 (1979).
44. W.H. Pirkle, R. Dappen, 'Reciprocity in Chiral Recognition: Comparison of Several Chiral Stationary Phases', *J. Chromatogr.*, **404**, 107–115 (1987).
45. W.H. Pirkle, D.W. House, J.M. Finn, 'Broad Spectrum Resolution of Optical Isomers Using Chiral HPLC Bonded Phases', *J. Chromatogr.*, **192**, 143–158 (1980).
46. W.H. Pirkle, J.M. Finn, J.L. Schreiner, B.C. Hamper, 'A Widely Useful Chiral Stationary Phase for the High-performance Liquid Chromatography Separations of Enantiomers', *J. Am. Chem. Soc.*, **103**, 3964 (1981).
47. W.H. Pirkle, C.J. Welch, 'An Improved Chiral Stationary Phase for the Chromatographic Separation of Underivatized Naproxen Enantiomers', *J. Liq. Chromatogr.*, **15**, 1947 (1992).
48. C.E. Dalgliesh, 'The Optical Resolution of Aromatic Amino-acids on Paper Chromatograms', *J. Chem. Soc.*, **137**, 3940–3942 (1952).
49. W.H. Pirkle, T.C. Pochapsky, 'Considerations of Chiral Recognition Relevant to the Liquid Chromatographic Separation of Enantiomers', *Chem. Rev.*, **89**, 347–362 (1989).
50. G. Dotsevi, Y. Sogah, D.J. Cram, 'Chromatographic Optical Resolution through Chiral Complexation of Amino Ester Salts by a Host Covalently Bound to Silica Gel', *J. Am. Chem. Soc.*, **97**, 1259–1261 (1975).
51. D.W. Armstrong, 'Pseudophase Liquid Chromatography: Applications to TLC', *J. Liq. Chromatogr.*, **3**(6), 895–900 (1980).
52. W.L. Hinze, D.W. Armstrong, 'Thin Layer Chromatographic Separation of Ortho, Meta, and Para Substituted Benzoic Acids and Phenols with Aqueous Solutions of α-Cylcodextrin', *Anal. Lett.*, **13**, 1093–1104 (1980).

53. J.L. Hoffman, 'Chromatography of Nucleic Acid Components on Clyclodextrin Gels', *Anal. Biochem.*, **33**, 209–217 (1970).
54. A. Harada, M. Furue, S.I. Nozakura, 'Optical Resolution of Mandelic Acid Derivatives by Column Chromatography on Crosslinked Cyclodextrin Gels', *J. Polym. Sci.*, **16**, 189–196 (1978).
55. K. Fugimara, T. Ueda, T. Ando, 'Retention Behavior of Some Aromatic Compounds on Chemically Bonded Cyclodextrin Silica Stationary Phase in Liquid Chromatography', *Anal. Chem.*, **55**, 446–450 (1983).
56. Y. Kawagucki, M. Tanaka, M. Nakae, K. Funazo, T. Shono, 'Chemically Bonded Cyclodextrin Stationary Phases for Liquid Chromatographic Separation of Aromatic Compounds', *Anal. Chem.*, **55**, 1852–1857 (1983).
57. D.W. Armstrong, 'Bonded Phase Materials for Chromatographic Separations', Patent No. 4539399, 1985.
58. D.W. Armstrong, W. Demond, 'Cyclodextrin Bonded Phases for the Liquid Chromatographic Separation of Optical, Geometrical, and Structural Isomers', *J. Chromatogr. Sci.*, **22**, 415–441 (1984).
59. D.W. Armstrong, W. Demond, A. Alak, W.L. Hinze, T.E. Riehl, K.H. Bui, 'Liquid Chromatographic Separation of Diastereomers and Structural Isomers on Cyclodextrin-bonded Phases', *Anal. Chem.*, **57**, 234–237 (1985).
60. D.W. Armstrong, A. Alak, K.H. Bui, W. Demond, T. Ward, T.E. Riehl, W.L. Hinze, 'Facile Separation of Enantiomers, Geometrical Isomers, and Routine Compounds on Stable Cyclodextrin LC Bonded Phases', *J. Inclusion Phenom.*, **2**(3&4), 533–545 (1984).
61. M.L. Bender, M. Komiyama, *Cyclodextrin in Chemistry*, Springer-Verlag, Berlin, 1978.
62. W.L. Hinze, 'Applications of Cyclodextrins in Chromatographic Separations and Purification Methods', *Sep. Purif. Meth.*, **10**(2), 159–237 (1981).
63. T.J. Ward, D.W. Armstrong, 'Improved Cyclodextrin Chiral Phases: A Comparison and Review', *J. Liq. Chromatogr.*, **9**(2&3), 407–423 (1986).
64. J. Szejtli, *Cyclodextrins and their Inclusion Complexes*, Akademiai Kiado, Budapest, 1982.
65. D.W. Armstrong, 'Chiral Stationary Phases for High Performance Liquid Chromatographic Separation of Enantiomers: A Mini-review', *J. Liq. Chromatogr.*, **7**(Suppl. 2), 353–376 (1984).
66. D.W. Armstrong, W. DeMond, A. Alak, 'Liquid Chromatographic Separation of Diastereomers and Structural Isomers on Cyclodextrin-bonded Phases', *Anal. Chem.*, **57**, 234–237 (1985).
67. D.W. Armstrong, W. DeMond, B.P. Czech, 'Separation of Metallocene Enantiomers by Liquid Chromatography: Chiral Recognition via Cyclodextrin Bonded Phases', *Anal. Chem.*, **57**, 481–484 (1985).
68. D.W. Armstrong, T.J. Ward, A. Czech, B.P. Czech, R.A. Bartsch, 'Synthesis, Rapid Resolution, and Determination of Absolute Configuration of Racemic 2, 2′-binaphthyldiyl Crown Ethers and Analogs via β-Cyclodextrin Complexation', *J. Org. Chem.*, **50**, 5556–5559 (1985).
69. D.W. Armstrong, T.J. Ward, R.D. Armstrong, T.E. Beesley, 'Separation of Drug Stereoisomers by the Formation of β-Cyclodextrin Inclusion Complexes', *Science*, **232**, 1132–1135 (1986).
70. C.A. Chang, Q. Wu, L. Tan, 'Normal-phase High-performance Liquid Chromatographic Separations of Positional Isomers of Substituted Benzoic Acids with Amine and β-Cyclodextrin Bonded-phase Columns', *J. Chromatogr.*, **361**, 199–207 (1986).
71. D.W. Armstrong, S. Chen, C. Chang, S. Chang, 'A New Approach for the Direct Resolution of Racemic Beta Adrenergic Blocking Agents by HPLC', *J. Liq. Chromatogr.*, **15**, 545–556 (1992).
72. W.L. Hinze, 'Applications of Cyclodextrins in Chromatographic Separations and Purification Methods', *Sep. Purif. Meth.*, **10**(2), 159–237 (1981).
73. M.P. Gasper, A. Berthod, U.B. Nair, D.W. Armstrong, 'Comparison and Modeling Study of Vancomycin, Ristocetin A, and Teicoplanin for CE Enantioseparations', *Anal. Chem.*, **68**, 2501–2514 (1996).
74. D.W. Armstrong, K. Rundlett, G.L. Reid, III, 'Use of a Macrocyclic Antibiotics, Rifamycin B, and Indirect Detection for the Resolution of Racemic Amino Alcohols by CE', *Anal. Chem.*, **66**, 1690–1695 (1994).
75. H.M. Higgins, W.H. Harrison, G.M. Wild, H.R. Bungay, M.H. McCormick, 'Vancomycin, a New Antibiotic', *Antibiotics Annual*, 906–914 (1957–1958).
76. A.H. Hunt, R.M. Molloy, J.L. Occolowitz, G.G. Marconi, M. Debono, 'Structure of the Major Glycopeptide of the Teicoplanin Complex', *J. Am. Chem. Soc.*, **106**, 4891–4895 (1984).
77. D.W. Armstrong, Y. Tang, S. Chen, Y. Zhou, C. Bagwill, J.R. Chen, 'Macrocyclic Antibiotics as a New Class of Chiral Selectors for Liquid Chromatography', *Anal. Chem.*, **66**, 1473–1484 (1994).
78. D.W. Armstrong, Y. Lie, K.H. Ekorgott, 'A Covalently Bonded Teicoplanin Chiral Stationary Phase for HPLC Enantioseparations', *Chirality*, **7**, 474–497 (1995).

Column Theory and Resolution in Liquid Chromatography

John G. Dorsey
Florida State University, Tallahassee, USA

1 Introduction 1
2 Fundamentals of Chromatography 1
2.1 Thermodynamic Considerations 2
2.2 Kinetic Considerations 2
2.3 Resolution 6
2.4 Time Considerations 7
2.5 Peak Capacity 8
Abbreviations and Acronyms 8
Related Articles 8
References 8

1 INTRODUCTION

Chromatography is a general term applied to a wide variety of separation techniques based upon the sample partitioning between a moving phase, which can be a gas, liquid, or supercritical fluid and a stationary phase, which may be either a liquid or a solid. The discovery of chromatography is generally credited to Tswett[1] who in 1906 described his work on using a chalk column to separate pigments in green leaves. The term ''chromatography'' was coined by Tswett to describe the colored zones that moved down the column. The technique languished for years with only periodic spurts of development following innovations such as partition and paper chromatography in the 1940s, gas chromatography (GC) and thin-layer chromatography in the 1950s, and various gel or size-exclusion methods in the early 1960s.[2] Then in January, 1969, the *Journal of Gas Chromatography* officially changed its name and became the *Journal of Chromatographic Science*. This change reflected the renewed interest in the technique of liquid chromatography (LC) and officially signaled the beginning of the era of modern LC. This renewed interest in the oldest of chromatographic techniques was brought about both because of the successes and because of the failures of GC. On the one hand, GC provided a firm theoretical background upon which modern LC could build. However, this renewed interest in LC was being driven because of the inability of GC to handle thermally unstable or nonvolatile compounds. It has been estimated that fewer than 20% of organic compounds have sufficient volatility and thermal stability to successfully traverse a GC column. Admirers of LC also liked the selective interaction of its two chromatographic phases, and easy sample recovery because of its nondestructive detection methods and room temperature operation.

In the following years the technique was known by many acronyms. It was at various times designated high-pressure liquid chromatography, high-speed liquid chromatography, high-performance liquid chromatography (HPLC), high-priced liquid chromatography by cynical research directors, and finally modern liquid chromatography. The modifiers were all used to signify the difference in the technique brought about by small particle diameter stationary phases and the resulting high pressure needed to drive the mobile phase through the densely packed columns. These differences resulted in much faster separations with higher resolution than traditional column chromatography. Through the 1970s the technique grew with astounding speed, and for nine consecutive years from the middle of the 1970s to the early 1980s the technique led all other analytical instruments in growth rate in an annual survey conducted by the *Industrial Research and Development* magazine.

A discussion of the history of LC is beyond the scope of this volume; an excellent treatment has been published by Ettre.[3] However, some perspective on the beginnings of modern LC is useful for an understanding of the technique as now practiced. The first comprehensive publication on the technique was a book edited by Kirkland,[4] which grew from a short course offered by the Chromatography Forum of the Delaware Valley. This then laid the foundation for the definitive book, now in its second edition, *Introduction to Modern Liquid Chromatography*.[2]

During the development of modern LC, advances have been driven by both instrumentation and chemistry. The technique as now practiced produces elegant separations that have been developed through chemical manipulation of the stationary and mobile phases, which are then effected and detected by modern instrumentation.

2 FUNDAMENTALS OF CHROMATOGRAPHY

Chromatography is a separation method which employs two phases, one stationary and one mobile. As a mixture of analytes being carried through the system by the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

mobile phase (m) passes over and through the stationary phase (s), the individual components (X) of the mixture equilibrate or distribute between the two phases:

$$X_m \longleftrightarrow X_s$$

The corresponding thermodynamic distribution coefficient K is defined as the concentration of component (X) in the stationary phase divided by its concentration in the mobile phase, as shown in Equation (1).

$$K = \frac{[X]_s}{[X]_m} \tag{1}$$

The use of a typical equilibrium constant K in chromatographic theory indicates that the system can be assumed to operate at equilibrium. As the analyte (X) proceeds through the system, it partitions between the two phases and is retained in proportion to its affinity for the stationary phase. At any given time, a particular analyte molecule is either in the mobile phase, moving at its velocity, or in the stationary phase and not moving at all. The individual properties of each analyte control its thermodynamic distribution and retention, and result in differential migration of the components in the mixture; the basis of the chromatographic separation. The effectiveness of the separation, however, is a function of *both* thermodynamics and kinetics.

2.1 Thermodynamic Considerations

In order to understand the thermodynamic considerations of the separation process, we need to look at the basic equations which describe the retention of a solute and relate the parameters of retention time, retention volume, and retention factor to the thermodynamic solute distribution coefficient.

Figure 1 shows a model chromatogram for the separation of two compounds. The time required for the elution of a retained compound is given by t_r, the retention time, and is equal to the time the solute spends in the moving mobile phase plus the time spent in the stationary phase, not moving. If a solute is unretained and has no interaction with the stationary phase, it will elute at t_0, the hold-up time or dead time, which is the same time required for the mobile phase solvent molecules to traverse the column.

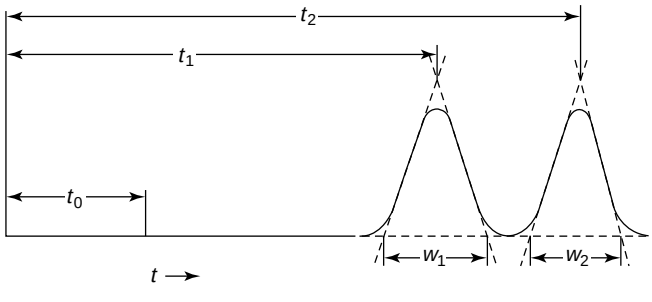

Figure 1 A model chromatogram for the separation of two compounds, **1** and **2**, by column chromatography.

Another fundamental measure of retention is retention volume, V_r. It is sometimes preferable to record values of V_r rather than t_r since t_r varies with the flow rate F, while V_r is independent of F. If we wish to describe the volume of the mobile phase required to elute a retained compound, then the retention volume is the product of the retention time and the mobile phase flow rate, as shown in Equation (2).

$$V_r = t_r F \tag{2}$$

V_0 is a measure of the total volume of space available to the mobile phase in the system and correlates to t_0 as the *volume* necessary to elute an unretained compound. This is shown in Equation (3):

$$V_0 = t_0 F \tag{3}$$

The most commonly used retention parameter in HPLC is the retention factor, k'. While the distribution coefficient, K, describes the concentration ratios, the retention factor, k', as shown in Equation (4), is the ratio of the amount of solute in each phase.

$$k' = \frac{\text{moles of X in stationary phase}}{\text{moles of X in mobile phase}} = \frac{[X]_s V_s}{[X]_m V_m} = \frac{K V_s}{V_m} \tag{4}$$

Equation (4) shows that k' is directly proportional to V_s, and so k' for each solute changes with the stationary phase loading on the silica support.

The retention factor is conveniently measured from retention parameters, as shown in Equation (5), since k' also describes the amount of time the solute spends in each phase.

$$k' = \frac{t_r - t_0}{t_0} = \frac{V_r - V_0}{V_0} \tag{5}$$

Rearrangement of Equation (5) gives Equation (6):

$$V_r = V_0(1 + k') \tag{6}$$

The retention volume can then be related to the distribution coefficient by substituting Equation (4) for k' in Equation (6), where $V_m = V_0$. This gives Equation (7):

$$V_r = V_0 + K V_s \tag{7}$$

2.2 Kinetic Considerations

It is obvious from Figure 1 that the chromatographic separation of components in a mixture is dependent on

two factors; the difference in retention times of two adjacent peaks, or more precisely the difference between peak maxima, and the peak widths. It was shown in the preceding discussion that the retention of a solute is a thermodynamic process controlled by the distribution coefficient and the stationary phase volume. The peak width, or band broadening, on the other hand is a function of the kinetics of the system.

Early papers on chromatographic theory described the technique in terms similar to distillation or extraction. The Nobel Prize winning work by Martin and Synge[5] in 1941 introduced liquid–liquid (or partition) chromatography and the accompanying theory that became known as the *plate theory*. Although the plate theory was useful in the development of chromatography, it contained several poor assumptions. The theory assumed that equilibration between phases was instantaneous, and that longitudinal diffusion did not occur. Furthermore, it did not consider dimensions of phases or flow rates. An alternative to the plate theory which came into prominence in the 1950s was the so-called "*rate theory*". The paper which has had the greatest impact was the one published by van Deemter et al.[6] They described the chromatographic process in terms of kinetics and examined diffusion and mass transfer. The popular van Deemter plot resulted and the rate theory has become the backbone of chromatographic theory.

Several processes, which will be discussed shortly, contribute to the overall width of the band, and the contribution from each process can be described in terms of the variance (σ^2) of the chromatographic peak, which is the square of its standard deviation (σ).

The most common measure of efficiency of a chromatographic system is the plate number, N. Because the concept originated from the analogy with distillation, it was originally called the number of theoretical plates contained in a chromatographic column. This is not a useful analogy since a chromatographic column does not contain "plates" but the terminology has become universally adopted and the original definition has persisted. The expression for the efficiency of a column is given by Equation (8).

$$N \equiv \left(\frac{t_r}{\sigma_t}\right)^2 \tag{8}$$

where σ_t is the standard deviation in time units; the parameters t_r and σ_t must be measured in the same units so the number N is dimensionless.

For Gaussian peaks we can express σ in terms of peak width where the width of the peak at the baseline (w_b) is equal to 4σ (see Figure 2). N is then given by Equation (9):

$$N = \left(\frac{t_r}{\sigma_t}\right)^2 = 16\left(\frac{t_r}{w_b}\right)^2 \tag{9}$$

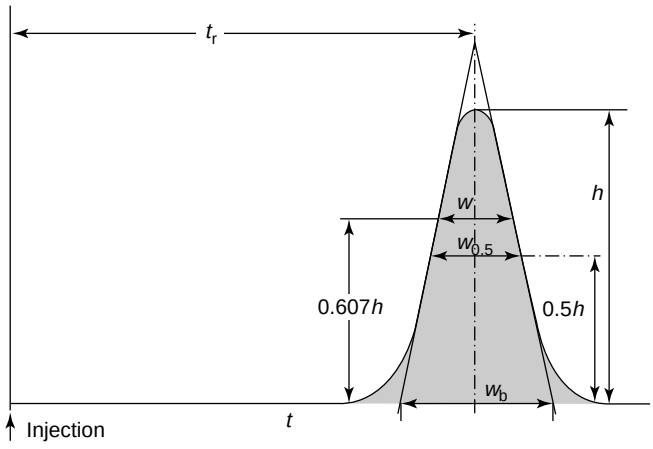

Figure 2 Evaluation of a chromatographic peak for the calculation of column efficiency.

Because measurements depending on w_b rely on the accurate positioning of tangents to the chromatographic peak, another common calculation of N uses the more easily measured peak width at half height, $w_{0.5}$, where $w_{0.5} = 2.354\sigma$, as shown in Equation (10):

$$N = \left(\frac{t_r}{\sigma_t}\right)^2 = 5.54\left(\frac{t_r}{w_{0.5}}\right)^2 \tag{10}$$

The assumption for the measurement of peak widths in Equations (8–10) is that the peak is Gaussian. Unfortunately, few peaks are truly Gaussian and in general, for asymmetrical peaks, N, calculated by a Gaussian-based equation, increases the higher up on the peak the width is measured. Figure 3 shows the

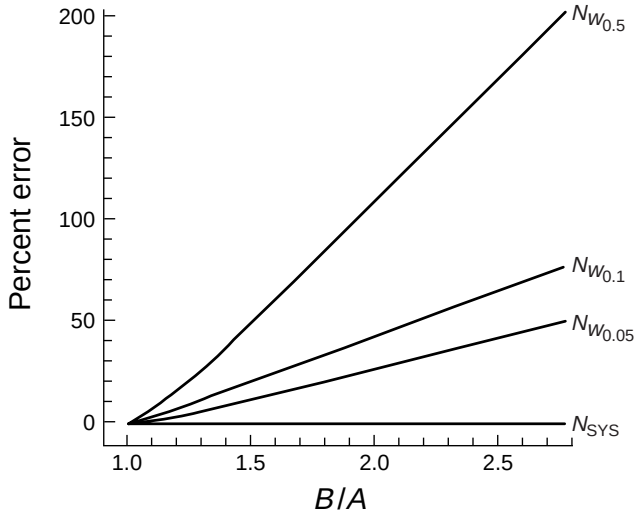

Figure 3 Plot of percentage error in plate count determinations for asymmetrical peaks as a function of peak height at which the width is measured. The quantities B and A are illustrated in Figure 4. (Adapted from Foley and Dorsey.[7])

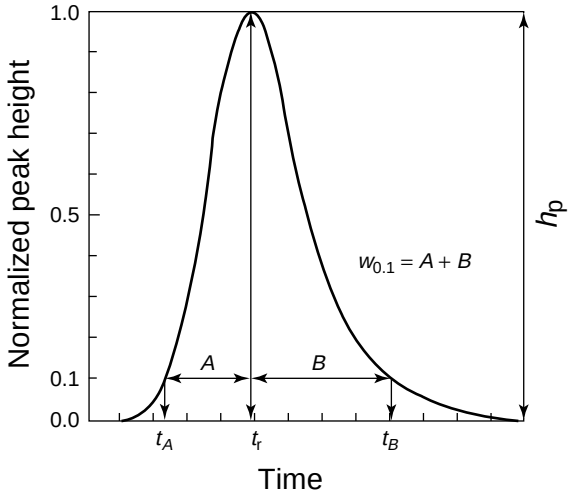

Figure 4 Plot of normalized peak height vs. time.

error in plate count determinations for asymmetrical peaks as a function of the peak height at which the width is measured. A more accurate approach to efficiency measurement, shown in Equation (11), has been presented by Foley and Dorsey[7] which takes into account the peak asymmetry.

$$N = \frac{41.7(t_r/w_{0.1})^2}{1.25 + (B/A)} \tag{11}$$

Here the peak width is measured at 10% of peak height and the asymmetric ratio or tailing factor is calculated as the ratio of B/A (see Figure 4). For a symmetrical peak, the B/A value will be 1, and the tailed peaks will have values greater than 1.

The plate number depends on the column length, making comparisons among columns difficult unless they are all of the same length. Another measure of efficiency which removes this dependency is given by the plate height, as shown in Equation (12):

$$H = \frac{L}{N} \tag{12}$$

where L is the length of the column. H can thus be thought of as the length of the column that contains one plate. For the highest efficiencies, the goal is to attain maximum N and small H values. We will see that H is proportional to the diameter of the stationary phase particles, so a better measure of expressing efficiency is the reduced plate height, h, given by Equation (13):

$$h = \frac{H}{d_p} \tag{13}$$

where d_p is the stationary phase particle diameter. The reduced plate height is a dimensionless number (H and d_p being measured in the same units) and this calculation allows us to compare efficiencies of columns with different particle diameters. A well operated chromatographic system should have h values of 2 to 5.

As mentioned earlier, the extent of band broadening, and thus the efficiency of a chromatographic column, is dependent on several contributing processes. The four important band broadening phenomena to be considered are: (i) resistance to mass transfer (rmt); (ii) eddy diffusion (ed); (iii) longitudinal diffusion (long); and (iv) extra-column effects (ex). Each of these independently would produce a Gaussian band. The variances of each band broadening process are additive to give the overall variance of the system, as shown in Equation (14):

$$\sigma_t^2 = \sigma_{rmt}^2 + \sigma_{ed}^2 + \sigma_{long}^2 + \sigma_{ex}^2 \tag{14}$$

2.2.1 Resistance to Mass Transfer

As solute molecules interact with the column packing they continually transfer into and out of the stationary phase. Resistance to mass transfer relates to the rate at which the molecules exchange between phases and may be the dominant cause of band spreading.

If there were no flow through the column, there would be an equilibrium distribution of the analyte molecules between the mobile and stationary phases according to its partition coefficient. However, since there is flow, molecules in the mobile phase are swept downstream creating a condition of nonequilibrium in the immediately adjacent stationary phase. In order to restore the system to equilibrium, some molecules left behind in the stationary phase must desorb and enter the mobile phase, and some that were carried ahead must leave the mobile phase and sorb onto the stationary phase. This mass transfer takes a finite time and results in band broadening. The faster the mass transfer occurs, the less will be the effect. Both the stationary and mobile phase play a role in determining the rate of mass transfer. A discussion of all the contributing factors is given by Giddings,[8] but three of the most important are the diffusion coefficient of the solute in the mobile phase (D_m), the stationary phase particle diameter (d_p), and mobile phase flow velocity (v). The relationship is shown by Equation (15):

$$\text{rate of mass transfer} \propto \frac{d_p^2 v}{D_m} \tag{15}$$

A molecule is in constant motion in solution. The larger the volume of mobile phase the molecule is in, the longer it will take to diffuse to a phase boundary. In chromatographic columns, the interstitial volume between particles is of the order of the same size as the particles themselves. Small particle diameter packings

therefore result in a much higher rate of contact and a greater rate of mass transfer. A high rate of diffusion of the solute in the mobile phase also increases the rate of contact by decreasing the time for the molecule to move between stationary phase particles. For this reason, many chromatographers choose low viscosity solvents as eluents and/or operate the system at elevated temperatures to increase the diffusion coefficient of the solute in the mobile phase.

The contribution to band broadening by flow velocity is obvious, since the faster the mobile phase is flowing, the farther molecules will be swept downstream before they exchange phases.

2.2.2 Eddy Diffusion

The velocity of the liquid moving through a column may vary significantly across the diameter of the column depending on the bed structure. The velocity of solutes being carried through the column by the mobile phase will therefore fluctuate between wide limits, and the total distance traveled by individual molecules will also vary. These variations are a result of different flow paths taken as a function of least resistance. Since the average velocity of the solute determines its retention time, these random fluctuations result in band broadening. There is still much disagreement about quantitation of this effect, but a generally accepted proportionality is given by Equation (16):

$$\text{eddy diffusion} \propto d_p V^{1/3} \tag{16}$$

The smaller the stationary phase particle diameter the closer will be the packing. This minimizes the variation in flow paths available to the molecules and reduces band broadening. High flow velocities affect the band broadening by causing molecules traveling though open pathways to move more rapidly than those in narrow pathways. The simple theory of eddy diffusion assumes that a particle will remain in a single flowpath. In practice, this is not the case since there is nothing to stop it from diffusing laterally from one flow path to another. This process, called ''coupling'', averages out the two flowpaths and reduces the band broadening so that the final band width, although still greater than the initial band width, is less than if coupling had not occurred. So, rather than a direct relationship to flow velocity, as theory would predict, eddy diffusion is more nearly proportional to velocity to the 1/3 power.

2.2.3 Longitudinal Diffusion

As solute bands move through a column, diffusion of molecules into the surrounding solvent occurs from the region of higher concentration to the region of lower concentration in proportion to the diffusion coefficient, D_m, according to Fick's law. The faster the mobile phase moves, the less time the zone is in the column, the less time there is for diffusion, and the lower the band broadening due to diffusion. The relationship for longitudinal diffusion, D_{long}, is given by Equation (17):

$$D_{long} \propto \frac{D_m}{V} \tag{17}$$

In principle, longitudinal diffusion may occur in both mobile and stationary phases, but because the rate is so small in the stationary phase, this factor can be neglected.

Combination of the three terms (resistance to mass transfer, eddy diffusion and longitudinal diffusion) gives the well known van Deemter expression shown in Equation (18):

$$H = A + \frac{B}{v} + Cv \tag{18}$$

where the A term is due to eddy diffusion, the B term is due to longitudinal diffusion, and the C term is due to resistance to mass transfer. The familiar van Deemter plot is shown in Figure 5. This equation is also often written with reduced plate height and velocities, and is then known as the Knox equation.

2.2.4 Extra-column Effects

Band broadening can occur in other parts of the chromatographic system as well as in the column. Contributions to this extra-column broadening may come from the injector, the detector flow cell, and the connecting tubing. Slow time constants of detectors and recorders may also contribute. These extra-column effects are more severe for early, narrow peaks in the chromatogram than for later, broader peaks.

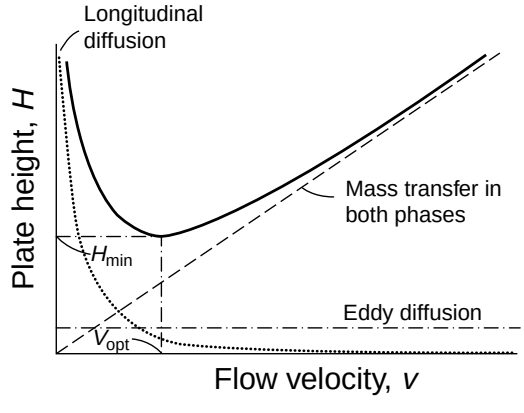

Figure 5 Van Deemter plot applied to LC.

2.3 Resolution

We have seen that the effective separation of two peaks is a function of both thermodynamic and kinetic effects. A common measure of the separation, the resolution, is the ratio of the distance between peak maxima and the average peak width.

$$R_s = \frac{\text{peak separation}}{\text{average peak width}} \quad \begin{matrix}\text{(thermodynamic)} \\ \text{(kinetic)}\end{matrix} \tag{19}$$

In order to obtain peak separation, one of the components in a mixture must be more selectively retained than another, or we may say the column showed selectivity toward the components in the mixture. Selectivity is measured by the separation factor or relative retention (α) and may be given by any of relationships shown in Equation (20):

$$\alpha_{2,1} = \frac{K_2}{K_1} = \frac{k'_2}{k'_1} = \frac{t'_{r2}}{t'_{r1}} = \frac{t_{r2} - t_m}{t_{r1} - t_m} \tag{20}$$

The relationship between the separation factor and the distribution coefficients K_2 and K_1 of the two components (2) and (1) emphasizes the thermodynamic basis of the separation.

Figure 6 illustrates the influence on resolution of peak separation and peak width. Figure 6(a) shows two components poorly resolved because of inadequate selectivity and large band widths. Figure 6(b) shows the components with the same band widths but good resolution as a result of improved selectivity. Figure 6(c) also demonstrates good resolution, but here narrow band widths are the contributing factor even though the selectivity is unchanged. The quantitative measure of resolution is given by Equation (21).

$$R_s = \frac{2(t_{r2} - t_{r1})}{w_1 + w_2} \tag{21}$$

In practice, it is seldom necessary to calculate the value of resolution. More frequently, the chromatographer simply looks at the shape of the peaks and estimates the R_s value from the peak shape. Of course, the concentrations and/or responses of the components in a mixture are rarely equal. Figure 7 illustrates the significance of the resolution value for two components with different relative concentrations.

The chromatographic control of resolution is a function of several factors. The fundamental resolution equation is shown in Equation (22):

$$R_s = \left(\frac{N^{1/2}}{4}\right)\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k'}{k' + 1}\right) \tag{22}$$

where $\alpha = k'_2/k'_1$.

The plate number N (column factor) can be increased by lengthening the column, changing the packing particle diameter, or optimizing flow rate. However, improving resolution by increasing N is expensive in time. Doubling the column length doubles the elution time, solvent consumption, and pressure while only increasing R_s by 1.4. Likewise, reducing the particle diameter increases the resolution but may exceed the maximum allowable pressure.

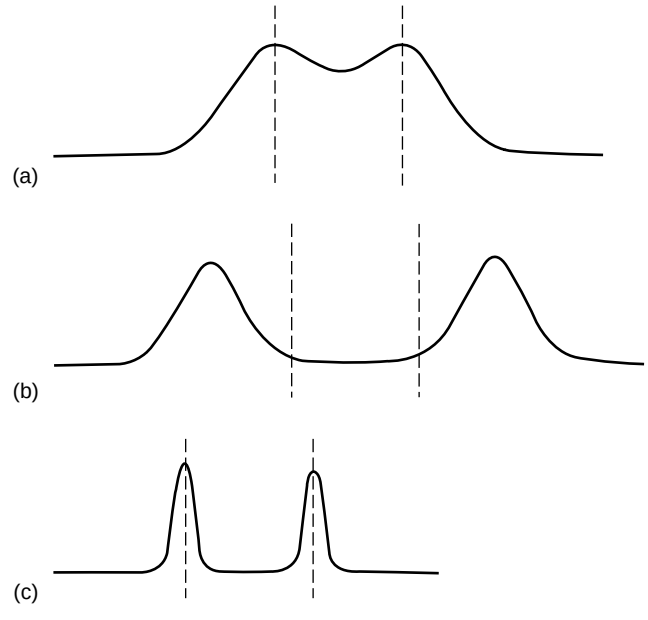

Figure 6 Two-component chromatograms to illustrate the importance of peak-to-peak separation and peak widths on separation.

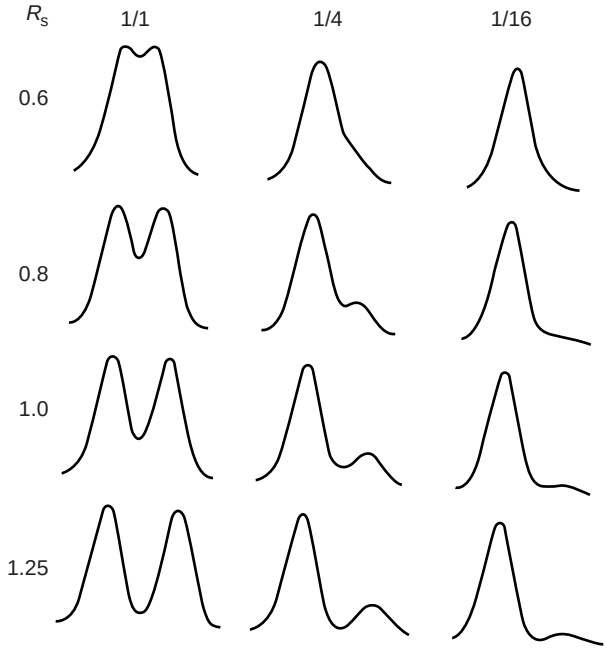

Figure 7 Significance of resolution values for two components with different relative concentrations.

The selectivity α (thermodynamic factor) can be increased by changing columns to a different stationary phase or by imposing secondary equilibria through changes in mobile phase pH or the addition of complexing agents to the mobile phase, for example. In order to discuss the effect of selectivity changes, it is necessary to rearrange Equation (22) and solve for N, as shown in Equation (23):

$$N_{\text{req}} = 16R_s^2 \left(\frac{\alpha}{\alpha - 1}\right)^2 \left(\frac{k' + 1}{k'}\right)^2 \qquad (23)$$

If we assume $R_s = 1.0$ and $k' = 2$, then

$$\alpha = 1.01 \qquad N_{\text{req}} = 367\,000$$
$$\alpha = 1.05 \qquad N_{\text{req}} = 16\,000$$

Changing α to 1.05 gives us a 96% reduction in required column efficiency which translates to a 96% reduction in column length, analysis time, solvent consumption, and solvent disposal. Changing selectivity is certainly the most efficient way to improve resolution. However, the cost of changing α is in method development time.

Increasing the retention factor, k', can also increase resolution, particularly for early eluting compounds. This is done primarily by going to a weaker mobile phase. Looking at Equation (23) again, for $R_s = 1.0$, and $\alpha = 1.05$,

$$k_2' = 0.1 \qquad N_{\text{req}} = 854\,000$$
$$k_2' = 1.0 \qquad N_{\text{req}} = 28\,000$$
$$k_2' = 10.0 \qquad N_{\text{req}} = 8500$$

While a 97% decrease in required column efficiency is calculated for changing k' from 0.1 to 1.0, another 70% decrease occurs for a k' change from 1.0 to 10.0. It is clear from Equation (22) that as k' becomes large, the resolution is unaffected by small changes in k'. Changing k' to improve resolution thus has an upper limit of effectiveness at approximately $k' = 10$. This is illustrated in Figure 8.

2.4 Time Considerations

While any of the above methods will improve resolution, they are far from equal in terms of their effects on time. Recalling that $t_r = t_0(1 + k')$, and $H = L/N$, and realizing $t_0 = L/v$, where v is the average flow velocity, then $t_0 = HN/v$. So, $t_r = (HN/v)(1 + k')$ and rearranging gives Equation (24):

$$N = \frac{t_r v}{H} \frac{1}{1 + k'} \qquad (24)$$

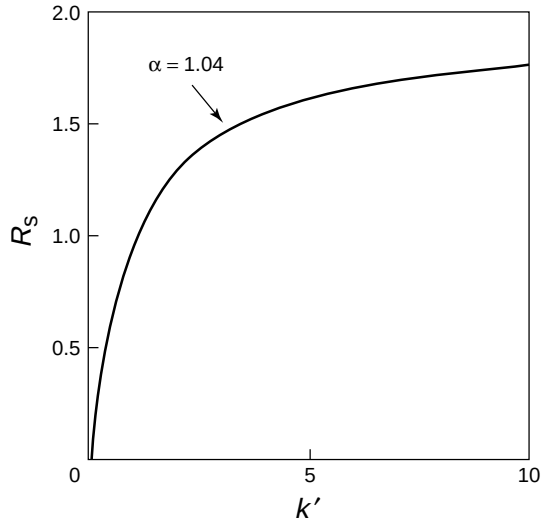

Figure 8 Resolution of a pair of peaks as a function of k'. Plate count N is 10 000 and α is 1.04.

Setting Equation (24) equal to Equation (23) and solving for t_r gives Equation (25):

$$t_r = 16R_s^2 \left(\frac{\alpha}{\alpha - 1}\right)^2 \frac{(k' + 1)^3}{k'^2} \frac{H}{v} \qquad (25)$$

We can see that to double the resolution by increasing N (holding α, k', H and v constant) requires a four-fold increase in time.

In the previous section we saw how changes in selectivity can improve resolution while decreasing analysis time. If we assume $N = 10\,000$ and $k' = 2$, and further if we could manipulate chromatographic conditions such that k_2' would remain unchanged as we change selectivity, then by Equation (22) a change in α from 1.01 to 1.05 would increase resolution from 0.2

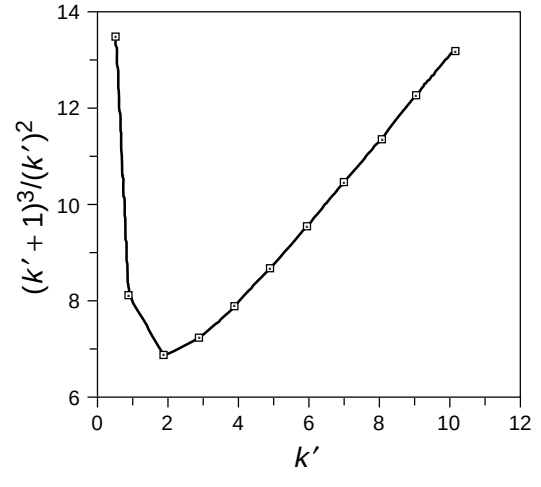

Figure 9 Plot of $(k' + 1)^3/(k')^2$ vs k'.

to 0.8. We can see from Equation (25) then that with this improvement in resolution and the stated change in selectivity (holding H and v constant), that retention time would decrease by a factor of $t_{r2}/t_{r1} = 0.9$, or 10%.

Figure 9 shows that the $(k' + 1)^3/(k')^2$ term in Equation (25) goes to infinity at both large and small values of k'. Thus, in the effective range of k' values from 1 to 10 (holding H and v constant, and assuming $\alpha = 1.05$ and $N = 10\,000$) the resolution would improve by a factor of 1.8 with a reduction in analysis time of 66%.

We know we can reduce analysis time by shortening the column, but we have a required efficiency necessary to separate the components in a mixture with the desired resolution. Assuming a well operating chromatograph yielding a reduced plate height of three, 10 000 plates can be generated with 10 µm diameter particles in 30 cm, with 5 µm particles in 15 cm, or with 3 µm particles in 9 cm. Assuming a solute with a retention factor of 2, the retention volume for these columns would be 9.0, 4.5 and 2.7 mL, respectively, and the peak base width would be 360, 180 and 108 µL, respectively.

2.5 Peak Capacity

Peak capacity (PC), defined in Equation (26), is the maximum number of components in a mixture that can be resolved with a resolution of one between the inert peak and the last peak.

$$\mathrm{PC} = 1 + \left(\frac{N^{1/2}}{4}\right) \ln(1 + k') \qquad (26)$$

As samples become more complex, the ability of a particular separation method to resolve all components decreases. A statistical study of component overlap has shown that a chromatogram must be approximately 95% vacant to provide a 90% probability that a given component of interest will appear as an isolated peak.[9] This is shown graphically in Figure 10 where the probability of separation is plotted as a function of the system peak capacity for cases where the number of components varies from 10 to 70. A thorough discussion of resolving power is given by Martin et al.[10]

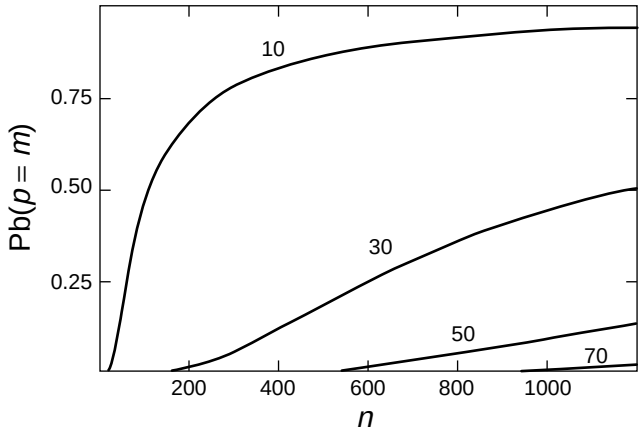

Figure 10 Plots of the probability that all (m) sample components will be separated, Pb($p = m$), as a function of the system peak capacity, n, for the representative cases where $m = 10, 30, 50$ and 70. (Adapted from Martin et al.[10])

ABBREVIATIONS AND ACRONYMS

GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography

RELATED ARTICLES

Food **(Volume 5)**
Liquid Chromatography in Food Analysis

Pesticides **(Volume 7)**
High-performance Liquid Chromatography Methods in Pesticide Residue Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Biopolymer Chromatography • Capillary Electrophoresis • Chiral Separations by High-performance Liquid Chromatography • Gradient Elution Chromatography • Ion Chromatography • Micellar Electrokinetic Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography • Supercritical Fluid Chromatography • Thin-layer Chromatography

REFERENCES

1. M. Tswett, 'On a New Category of Adsorption Phenomena and Their Application to Biochemical Analysis', *Ber. Dtch. Bot. Geo.*, **24**, 385 (1906).
2. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley-Interscience, New York, 1979.
3. L.S. Ettre, 'Evolution of Liquid Chromatography: A Historical Overview', in *High Performance Liquid Chromatography: Advances and Perspectives*, ed. Cs. Horvath, Academic Press, New York, Vol. 1, 1980.

4. J.J. Kirkland (ed.), *Modern Practice of Liquid Chromatography*, Wiley-Interscience, New York, 1971.
5. A.J.P. Martin, R.L.M. Synge, *Biochem. J.*, **35**, 1358 (1941).
6. J.J. van Deemter, F.J. Zuiderweg, A. Klinkenberg, *Chem. Eng. Sci.*, **5**, 271 (1956).
7. J.P. Foley, J.G. Dorsey, *Anal. Chem.*, **55**, 730 (1983).
8. J.C. Giddings, *Unified Separation Science*, Wiley-Interscience, New York, 1991.
9. J.M. Davis, J.C. Giddings, *Anal. Chem.*, **55**, 418 (1983).
10. M. Martin, D.P. Herman, G. Guiochon, *Anal. Chem.*, **58**, 2200 (1986).

Gradient Elution Chromatography

J.W. Dolan and L.R. Snyder
LC Resources Inc., Walnut Creek, CA, USA

1 Introduction **1**
1.1 Characteristic Features 1
1.2 Advantages of Gradient Elution 2
1.3 Linear versus Nonlinear Gradients 4
1.4 Equipment Used for Gradient Elution 4
2 Theory **5**
2.1 Band Migration During Gradient Elution 5
2.2 Linear Solvent Strength Model 6
2.3 Retention and Bandwidth in Gradient Elution 6
2.4 Resolution in Gradient Elution 7
3 Selecting Conditions for Gradient Elution **7**
3.1 Effects of Gradient Steepness, Range, and Shape 7
3.2 Control of Band Spacing 10
3.3 Effects of Column Dimensions, Particle Size and Flow Rate 12
3.4 Computer Simulation 13
4 "Nonideal" Effects in Gradient Elution **15**
4.1 Nonequilibration of the Column 15
4.2 Gradient Delay and Rounding 15
4.3 Equipment and Baseline Problems 15
List of Symbols **17**
Abbreviations and Acronyms **18**
References **18**

Elution chromatography can be carried out in either isocratic or gradient modes. In isocratic elution, the mobile-phase composition is held constant during separation of the sample, e.g. 60% v acetonitrile–water. In gradient elution, the mobile-phase composition will be varied during sample separation, e.g. changing from 0 to 100% v acetonitrile–water. Gradient elution requires special chromatographic equipment, as well as somewhat greater care on the part of the operator, but it has important advantages for many separations. Thus, in isocratic elution (Figure 1a), sample peaks tend to "bunch up" at the beginning of the chromatogram (often with decreased resolution) and to broaden at the end of the chromatogram (with reduced detection sensitivity). Gradient elution (Figure 1b), on the other hand, provides a more even spacing of peaks, similar widths throughout the chromatogram, and often a shorter run time. For these and other reasons, gradient elution is preferred for the separation of many samples.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

1.1 Characteristic Features

Gradient elution represents an alternative to isocratic separation and it is therefore useful to begin with a comparison of the two techniques. In isocratic elution, sample retention can be controlled by varying the composition of a mobile phase composed of a "weak" solvent A and a "strong" solvent B. An example is shown in Figure 2, for the separation of the *o*-dialkylphthalate sample of Figure 1 with a reversed-phase C_{18} column and mobile phases composed of water (A) and acetonitrile (B). An increase in % B from 50% B in Figure 2(a) to 70% B in Figure 2(b) results in decreased sample retention and a shorter run time. In isocratic separation as in Figure 2, the separation of any two adjacent peaks is critically dependent on % B. As the concentration of B increases and sample retention decreases, this in turn results in (1) reduced resolution, R_s, of the two bands, (2) decreased bandwidth and better detection sensitivity (taller peaks), and (3) a shorter run time as measured by the retention time, t_R, of the later-eluting band. For 50% B as mobile phase (Figure 2a), the separation of early bands C_2–C_4 is acceptable ($R_s > 1.5$), but later bands C_8–C_{10} require a long time for their elution from the column and are wider and therefore less easily detected. For 70% B as mobile phase (Figure 2b), later bands C_8–C_{10} leave the column in a much shorter time and are narrower and better detected; however, the separation of early bands C_2 and C_3 is poor ($R_s = 0.6$).

The isocratic separation of the *o*-dialkyl phthalate sample of Figures 1(a) and 2 is never completely satisfactory, regardless of the choice of mobile phase % B. The reason is that early bands require a weaker solvent (50% B in Figure 2a), while later bands benefit from a stronger solvent (70% B). The advantage of gradient elution for samples such as this is that each component of the sample can be eluted under conditions of optimum retention. Thus, in gradient elution all sample components are retained initially at the column inlet, because the starting mobile phase is usually quite weak and all bands initially have large values of k, the isocratic retention factor. As

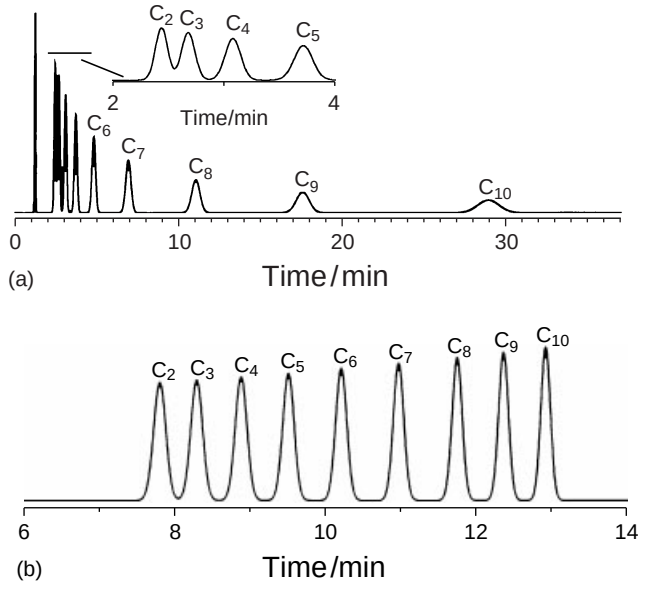

Figure 1 Separation of a mixture of *o*-dialkyl phthalates by reversed-phase high-performance liquid chromatography (HPLC). Sample: alkyl homologs from dimethyl (C_2) to dipentyl (C_{10}). Conditions: 25 × 0.46-cm C_{18} column, acetonitrile (B)–water (A) mobile phases, 35 °C, 2 mL min^{-1}. (a) Isocratic separation with 60% B, (b) gradient separation: 0–100% B in 12 min. Computer simulations[1] based on data of Snyder et al.[2] with $N = 2000$.

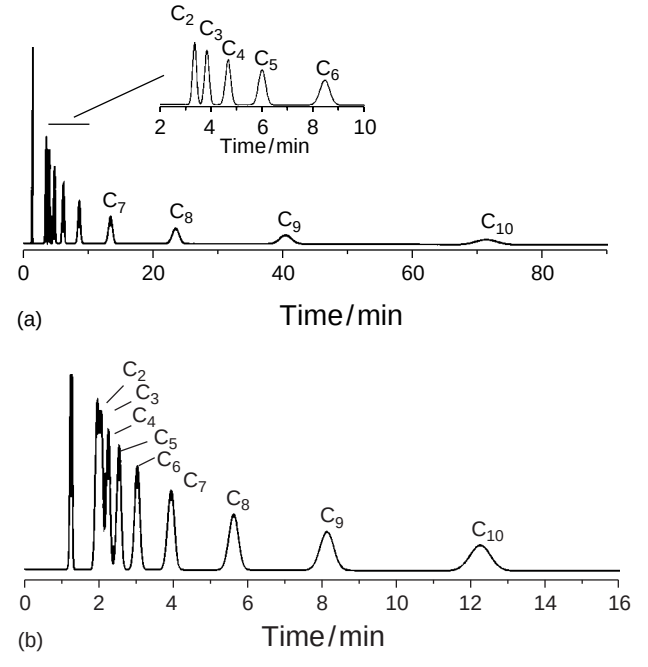

Figure 2 Isocratic separation of *o*-dialkyl phthalate bands C_2 and C_3 as a function of % B. (a) 50% B, (b) 70% B. Other conditions as in Figure 1.

the mobile phase becomes progressively stronger during the separation (% B increases), early bands begin to move through the column and are separated under near-optimal conditions (lower % B). Later bands move through the column at a later time under conditions (higher % B) that are also better for separation. As a result, seen in the example of Figure 1(b), it is possible by means of gradient elution to achieve good resolution for both early and late eluting bands, similar bandwidths and comparable ease of detection, and a shorter separation time.

The mobile phase gradient can be described in terms of the dependence of % B on time during the separation, as illustrated in Figure 3 for several gradient shapes. In most cases, the use of linear gradients as in Figure 3(a) is preferable; linear gradients will be assumed in the following discussion unless noted otherwise. The gradient elution separation can be regarded as a series of small isocratic steps as in Figure 3(e). As a result, the basic principles of isocratic and gradient elution are similar, and it will prove useful to describe gradient separations in terms of ''corresponding'' isocratic separations, where the sample and all experimental conditions are the same except for the change in % B with time in gradient elution. For a more detailed discussion of gradient elution, see Jandera and Churacek,[3] Snyder et al.[4–6]

1.2 Advantages of Gradient Elution

Isocratic elution is a simpler procedure than gradient elution, and for this reason it is usually the preferred technique. However, there are several situations where gradient elution can be a better choice, as discussed below.

1.2.1 Samples with Widely Different Retention

If the sample of Figure 2(b) were limited to the later-eluting five components (C_6–C_{10}), the resulting separation with 70% B as mobile phase would be quite satisfactory, i.e. good resolution of all bands, good detectability for all bands, and a reasonable run time (14 min). The problem with the separation of the total sample (C_2–C_{10}) is that early bands (C_2–C_4) are weakly retained and later bands (C_8–C_{10}) are strongly retained. In isocratic elution it is desirable that all sample bands be eluted with a retention factor k that falls within a narrow range of values: e.g. $1 < k < 10$. When this is not possible, as in the example of Figures 1 and 2, no isocratic mobile phase can provide a ''good'' separation of the sample. This is the most common reason for using gradient elution and it is sometimes referred to as the general elution problem. We will refer to samples such as that in Figures 1 and 2 as having a retention range that is too wide for isocratic elution.

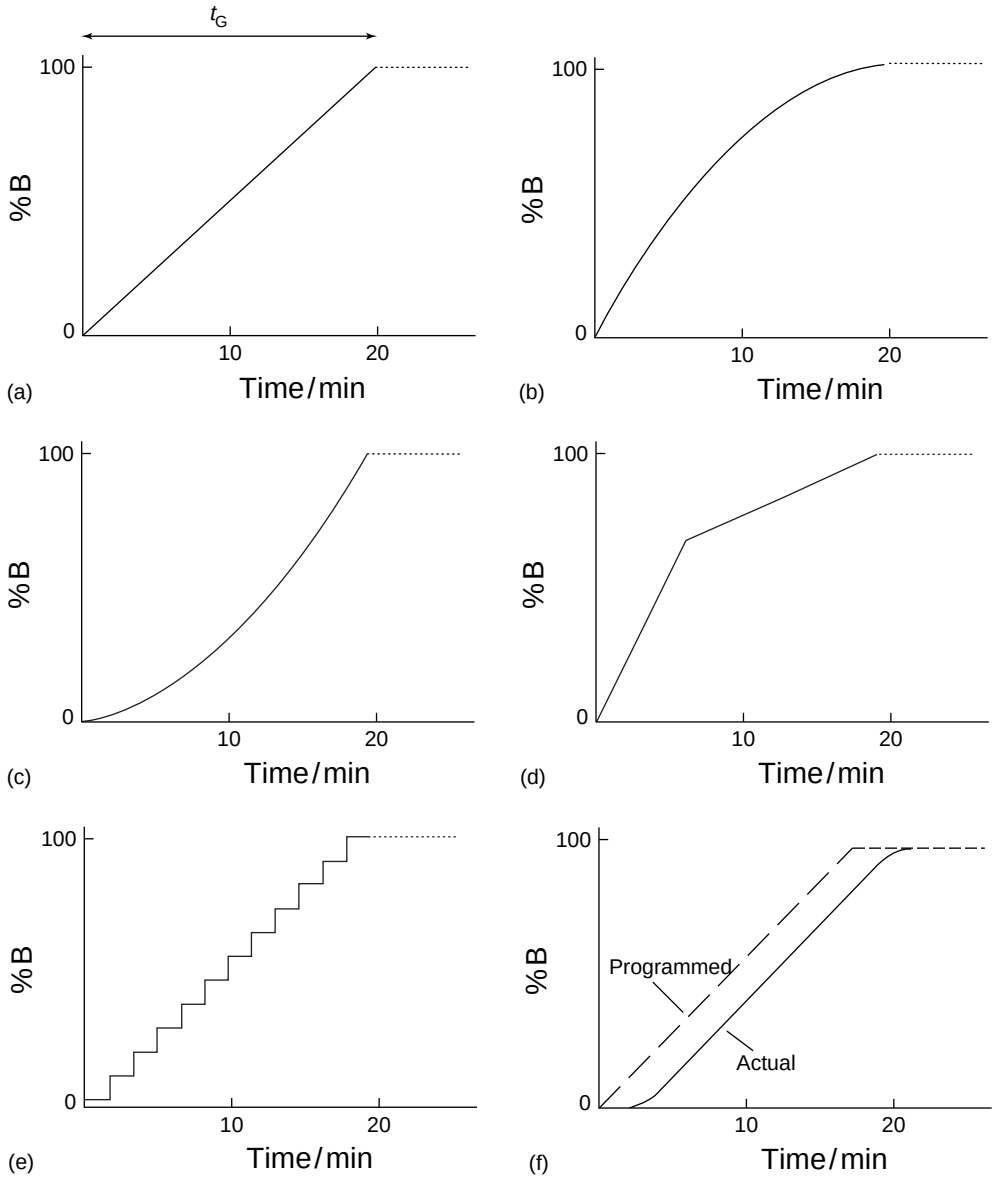

Figure 3 Different gradient shapes. (a) Linear, (b) convex, (c) concave, (d) segmented, (e) step gradient, (f) gradient delay and rounding. See text for additional details.

1.2.2 "Dirty" Samples

Many samples contain minor amounts of strongly retained components that are not of interest to the analyst. Examples include extracts of plant or animal tissue, environmental samples, and synthetic organic reaction mixtures. The repeated isocratic separation of such samples can result in a gradual accumulation of these "late eluters" on the column, with corresponding changes in sample retention, loss of column efficiency, and degradation of the baseline. For this reason, a frequent cleansing of the column may be required, as by washing the column with a strong solvent (e.g. 100% B). When gradient elution is used for the analysis of such samples, the column will be cleaned automatically by the strong mobile phase at the end of the gradient.

1.2.3 "Scouting" Runs During Method Development

The use of an initial gradient separation can be a valuable start to the development of a final HPLC method. Conditions for optimized retention can be estimated prior to the first experimental run (see section 3.1), such that most sample components are partially resolved and some idea of sample complexity can be formed. Furthermore, minor components that elute either early or late are

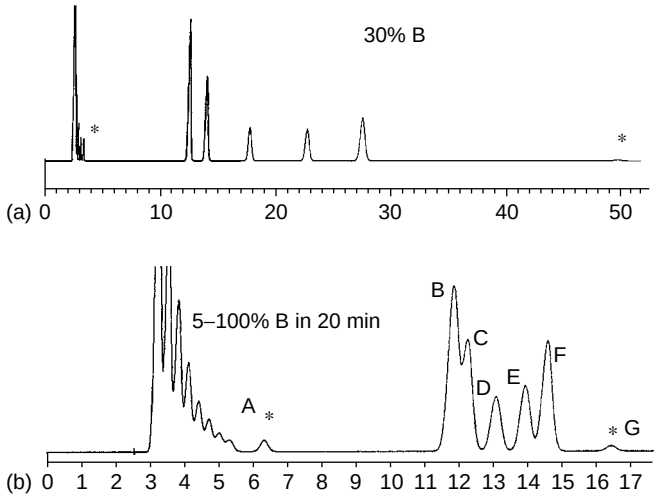

Figure 4 Easier detection of minor bands (∗) with gradient elution (b) versus isocratic separation (a). Hypothetical chromatograms.

often overlooked during isocratic separation, whereas the presence of such components is much more obvious in a gradient run (see Figure 4 and note peaks A and G marked by an asterisk). Once an initial gradient run has been obtained, it is possible to estimate reliably whether further experiments (and the final method) should be carried out under isocratic or under gradient conditions. Figure 4, which involves a gradient time $t_G = 20$ min (see Figure 3a) can be used to illustrate this procedure. As a first example, assume that only the major peaks B–F are of interest. Next, determine the difference in retention times for the first (B) and last (F) peaks, equal to $14.5 - 11.8 = 2.7$ min. Multiply this by the gradient rate (95% per 20 min $= 4.75\%$ min^{-1}). If the elution range (2.7 min $\times$ 4.75% min^{-1} $= 12.8\%$) is less than 25%, isocratic elution will be preferred. If this ratio is larger, then gradient elution generally will be advantageous. In this latter example, isocratic elution is recommended. Alternatively, assume that peaks A–G are of interest. For this case, the retention time difference is $16.3 - 6.3 = 10$ min, and the elution range is 10 min $\times$ 4.75% min^{-1} $= 47.5\%$. For the separation of peaks A–G, gradient elution would be a better choice. For samples where isocratic elution is preferred, the initial gradient run can also provide an estimate of isocratic % B for good sample retention ($1 < k < 10$),[6] thus avoiding a need for trial-and-error adjustments of % B in later runs.

1.3 Linear versus Nonlinear Gradients

Figure 3(a–e) illustrates different gradient shapes that modern equipment is able to deliver, as desired by the user. While linear gradients (Figure 3a) are generally preferred, nonlinear gradients can sometimes reduce separation time, improve sample resolution, and/or increase detection sensitivity. Segmented gradients (as in Figure 3d) are generally preferred with respect to other nonlinear gradient shapes (e.g. Figure 3b and c), because they are easier to optimize for a given sample. A common application of segmented gradients is the use of a very steep end-segment (or a step gradient) to 100% B in order to clean the column before the next injection. Less often, segmented gradients are useful in optimizing different parts of the chromatogram (see section 3.2). Thus, early peaks may prefer a steep gradient, while later peaks are better separated with a flat gradient (or vice versa).

1.4 Equipment Used for Gradient Elution

An HPLC system equipped for gradient elution must provide a means for creating gradients of various shapes, as in Figure 3. Typically, this is achieved by an electronic controller that varies the flow of two solvents A and B into a mixing chamber (“gradient mixer”), from which the mobile phase A–B passes into the column. This is illustrated in Figure 5(a) for a so-called “high pressure” mixing system. Solvents A and B are contained in reservoirs that feed separate pumps. The electronic controller can be programmed by the user to specify the desired gradient in terms of starting % B, final % B, gradient time t_G (or % B min^{-1}), and shape. The total flow rate is usually maintained constant during the gradient. Once the gradient is programmed and separation is begun, the controller determines the relative flow of solvents A and B to the mixer as a function of time. This in turn generates gradients as in Figure 3(a–e). Typically, sample injection initiates the start of the gradient.

There are two general designs of gradient equipment, corresponding to the high-pressure mixing system of Figure 5(a), or the low-pressure mixing system of Figure 5(b). In the latter system, the gradient is formed before the pump and only one pump is required to deliver the gradient to the column. High- and low-pressure systems are each widely used and are equally capable of reliable operation.

High-pressure mixing systems have a simple design, with the flow rate of the two pumps varying to provide the desired proportion of solvents A and B. Mixing takes place on the high-pressure side of the pump, minimizing outgassing problems (see section 4.3) and usually giving a smaller dwell volume (see section 4.2).

Low-pressure mixing systems require only one pump and may accommodate up to four solvents in many designs. Because solvents are mixed at atmospheric pressure, outgassing can compromise pump reliability

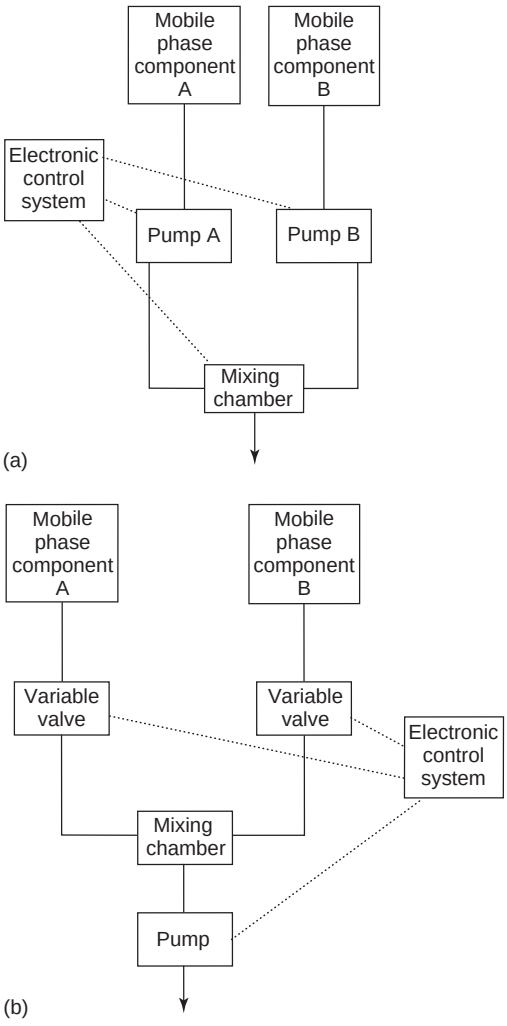

Figure 5 Equipment for gradient elution. (a) High-pressure mixing, (b) low-pressure mixing. (Reprinted with permission from Poole and Schuette.[7])

and the passage of the mixed solvents through the pump usually results in a larger dwell volume.

2 THEORY

Gradient elution involves additional experimental options when compared with isocratic separation, and for this reason is a more complicated technique to use and to understand. Fortunately, it is possible to describe gradient elution in virtually the same way as isocratic elution. Since most chromatographers are familiar with isocratic elution, it is therefore possible to achieve a similar understanding of gradient elution with little extra effort. In the following discussion, mobile phase composition as defined by % B will also be referred to as ϕ, the volume-fraction of B in the mobile phase A–B; $\phi = 0.01 \times (\%\ \mathrm{B})$.

2.1 Band Migration During Gradient Elution

In gradient elution the sample is typically (but not always) strongly retained at the beginning of the gradient, i.e. as for isocratic elution with the starting mobile phase. As the mobile phase becomes stronger during the gradient (larger ϕ or % B), sample retention decreases (as in Figure 2) and eventually the least-retained band X in the sample begins to migrate through the column. This process is illustrated in Figure 6 for the first (X) and last (Z) bands in the chromatogram. The solid curve labeled "X" shown in Figure 6(a) shows the migration of the band from the column inlet to outlet during the gradient. The accompanying dashed curve describes the instantaneous isocratic retention k during the gradient, i.e. the value of k for a mobile phase composition ϕ at the same location in the column occupied by band X. Because ϕ increases during the gradient, k decreases, and band migration accelerates as a function of time. At some time t_x (equal to the retention time t_R of the band), compound X leaves the column and appears as the first band in the chromatogram.

In gradient elution it is possible to define an average retention factor k^*, corresponding to an isocratic value k. Likewise, it is possible to define an average mobile phase composition $\phi = \phi^*$. Values of k^* and ϕ^* correspond to values of k and ϕ when the band has migrated halfway through the column ("band migration" = 0.5 in Figure 6). In the example of Figure 6, $k^* = 2$ for band X. Other sample components will migrate through the column at later times, as illustrated by the last band Z

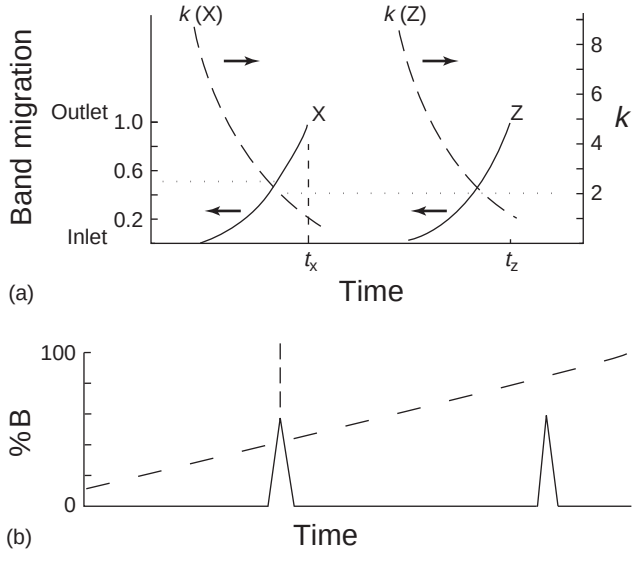

Figure 6 Band migration during gradient elution. (a) Solid curve, band migration; dashed curve, instantaneous value of k. (b) Resulting chromatogram showing first (X) and last (Z) peaks. See text for details. (Reprinted with permission from Snyder et al.[6])

in Figure 6. These later-eluting bands can be thought of as remaining stationary at the head of the column (very large isocratic k) until a sufficiently strong mobile phase arrives, causing them to move through the column. Because of the similar migration pattern (solid curves in Figure 6) for different sample components during gradient elution, values of k^* generally will be approximately equal for each band and therefore bandwidths in gradient elution tend to be similar for all sample components (see Figure 2b).

2.2 Linear Solvent Strength Model

Reversed-phase retention usually can be approximated by Equation (1)

$$\log k = \log k_w - S\phi \tag{1}$$

Here, $\log k$ varies linearly with mobile phase composition ϕ; k_w refers to the extrapolated value of k for water as mobile phase ($\phi = 0$), and S is a constant for a particular sample compound when only ϕ varies. If a linear gradient is chosen (as in Figure 3a), Equation (2) applies

$$\phi = \phi_o + (\phi_f - \phi_o)\frac{t}{t_G} \tag{2}$$

where ϕ_o and ϕ_f refer to the initial and final ϕ values, and t is time. Combination of Equations (1) and (2) then gives Equation (3)

$$\log k = (\log k_w - S\phi_o) - \frac{S(\phi_f - \phi_o)}{t_G}t = A - Bt \tag{3}$$

Here, A and B are constants for a sample compound, and $\log k$ varies linearly with time t. We will refer to gradient separations which can be approximated by Equation (3) as linear solvent strength (LSS) gradients. Gradients of this type can be described in the same terms used for isocratic elution, and for this reason LSS gradients provide a convenient basis for understanding gradient separations. The LSS model can also provide a more qualitative (but still useful) description of other HPLC methods, e.g. ion-exchange or normal-phase chromatography. For a more detailed discussion of LSS theory, see Snyder and Dolan.[8]

2.3 Retention and Bandwidth in Gradient Elution

The average retention of a band in reversed-phase gradient elution is given for LSS gradients as Equation (4)

$$k^* = \frac{0.87 t_G F}{V_m \Delta\phi S} \tag{4}$$

where F is flow rate, V_m is the column dead volume, and $\Delta\phi$ equals $(\phi_f - \phi_o)$, the change in ϕ during the gradient. For reversed-phase HPLC, an average value of $S \approx 4$ can be assumed for compounds with molecular weights < 500 Da.[9] Just as in isocratic elution it is desirable for $1 < k < 10$, so in gradient elution it is advantageous for $1 < k^* < 10$. However, in gradient elution with a linear gradient the value of k^* will be similar for every sample band and an intermediate value of k^*, e.g. $k^* = 5$, is a good initial choice.

Prior to any experimental runs, preferred starting conditions for a gradient separation can be selected, such that $k^* \approx 5$. As an example, assume that a column and flow rate are selected for the same reasons as in isocratic separation (to provide acceptable column back-pressure and a reasonable run time), e.g. a 15×0.46 cm C_{18} column at 2.0 mL min^{-1}. For columns with an internal diameter of 0.46 cm, the column dead volume V_m can be approximated by 0.1 times the column length in centimeters, or 1.5 mL in this example. Similarly, for a sample that has not previously been separated by HPLC, it is wise to use a full-range gradient (e.g. 0–100% B) so that $\Delta\phi = 1.0$. Full-range gradients can provide reasonable separation for samples that are either weakly or strongly retained. Referring to Equation (4), the only variable so far not specified is gradient time t_G; Equation (4), with $S \approx 4$, can be rearranged to solve for the necessary gradient time to achieve a value of $k^* = 5$, Equation (5):

$$t_G \approx \frac{20\, V_m \Delta\phi}{F} \tag{5}$$

For the column ($V_m = 1.5$ mL), flow rate (2.0 mL min^{-1}), and gradient range ($\Delta\phi = 1$) assumed in this example, a gradient time of 15 min should result in $k^* \approx 5$. This would be a good choice of conditions for the first experiment in HPLC method development.

Baseline bandwidth, W, in isocratic separation is given from the definition of column plate number, N, as, Equation (6)

$$W = \left(\frac{16}{N}\right)^{1/2} t_R = 4t_o \frac{k+1}{N^{-1/2}} \tag{6}$$

For gradient separations, a similar equation can be derived from the LSS model,[8] where k in Equation (6) is replaced by the value of k in gradient elution when the band leaves the column: $k^*/2$. This gives for gradient elution, Equation (7)

$$W = 4t_o \frac{(k^*/2)+1}{N^{-1/2}} \tag{7}$$

Thus, bandwidth in gradient elution is very similar to that in isocratic separation, except that for ''corresponding conditions'' ($k = k^*$), bandwidths in gradient elution are predicted to be smaller. However, Equation (7)

ignores the fact that a gradient tends to compress the sample band further, because the band tail leaves the column in a higher % B than the band front. On the other hand, this effect (which tends to sharpen bands in gradient versus isocratic elution) is compensated by another phenomenon, ''anomalous band broadening'' as discussed by Snyder and Dolan.[8] As a result, Equation (6) is in most cases an adequate approximation for bandwidth in gradient elution.

2.4 Resolution in Gradient Elution

Resolution in gradient elution is conveniently expressed by a relationship that can be derived from the LSS model, Equation (8):

$$R_s = \frac{1}{4}(\alpha - 1)N^{1/2}\frac{k^*}{k^* + 1} \qquad (8)$$

Equation (8) is of identical form as for isocratic elution (where k replaces k^* in Equation 8). Here, α is the separation factor, and both α and N in Equation (8) are average quantities measured when the band is at the column midpoint. Equation (8) is useful for gradient elution method development, where separation can be systematically improved by successive changes in sample retention, k^*, separation selectivity as measured by α, and column efficiency N. This is further illustrated in section 3.

3 SELECTING CONDITIONS FOR GRADIENT ELUTION

HPLC method development consists of experiments that can be used to select conditions for adequate separation, a process that is sometimes referred to as *optimizing* the separation. A detailed discussion of method development for gradient elution is given by Snyder et al.,[6] where it is shown that essentially the same procedure can be used for both gradient and isocratic methods. Sample retention as measured by k^* or k is the most easily optimized and usually is adjusted first. Separation selectivity as measured by α is often of critical importance and usually is considered next. Changes in selectivity (α) are essential whenever bands overlap in the chromatogram. Changes in column efficiency (N) are achieved by varying column length, particle size, and flow rate, and usually are carried out after optimizing sample retention k^* and separation selectivity α.

3.1 Effects of Gradient Steepness, Range, and Shape

Changes in gradient steepness, range, and shape mainly affect sample retention k^*.

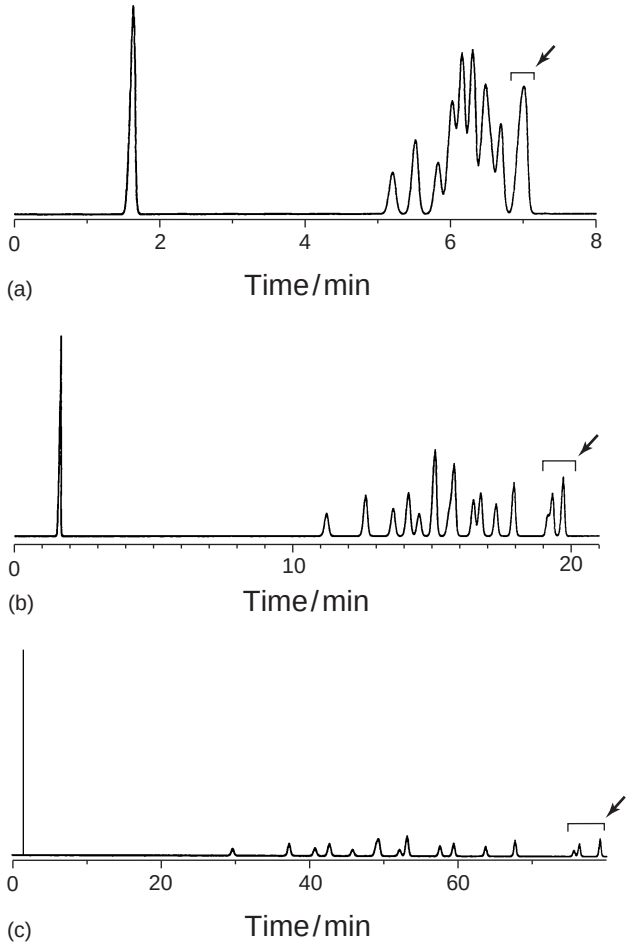

Figure 7 Effect of gradient time (t_G) and steepness (% B min^{-1}) on separation. Herbicides sample, 25 × 0.46-cm C_{18} column, methanol–water gradients. (a) 0–100% B in 5 min, 20% min^{-1}, $k^* = 0.7$; (b) 0–100% B in 20 min, 5% min^{-1}, $k^* = 3$; (c) 0–100% in 100 min, 1% min^{-1}, $k^* = 14$. (Reprinted with permission from Snyder et al.[6])

3.1.1 Gradient Steepness

Gradient steepness is defined as the change in % B per minute during the gradient (or $\Delta\phi/t_G$). From Equation (4), k^* is seen to vary with $t_G/\Delta\phi$, i.e. inversely as gradient steepness. As k increases in isocratic separation (Figure 2), overall separation (R_s) is seen to improve, run time increases, and bands become wider so that detection sensitivity decreases. Exactly the same results are expected in gradient elution when k^* increases. An increase in k^* is most conveniently effected by increasing gradient time, t_G, equivalent to decreasing gradient steepness. This is illustrated in Figure 7 for the separation of a herbicide sample, where the gradient time is changed from 5 min (a) to 20 min (b) and 100 min (c), corresponding to a change in k^* from 0.7 to 3 to 14. It is apparent in Figure 7 that sample resolution increases from Figure 7(a) to Figure 7(c); this is better seen for the

three bands at the end of the chromatogram (bracket and arrow). For the five-minute gradient Figure 7(a), the three bands overlap completely. When the gradient time is increased to 20 min Figure 7(b), the last band is separated to baseline, and the other two bands are partly resolved. For the 100-minute gradient Figure 7(c), all three bands are adequately resolved. At the same time, when gradient time is increased, all bands (except the solvent peak at 1.5 min) broaden and, because band area is constant, corresponding peak heights decrease (reduced detection sensitivity).

We can summarize Figure 7 as follows. Resolution is inadequate for the five-minute gradient in Figure 7(a), where $k^* = 0.7$; similarly, for the 100-minute gradient of Figure 7(c), where $k^* = 14$, run time is excessive and detection sensitivity is poor. These results are typical of separations where $k^* < 1$ or $k^* > 10$. In general, an intermediate gradient time will be optimum, and the 20-minute gradient of Figure 7(b), for which $k^* = 3$, comes close in this respect. However, three band pairs are poorly resolved, so an increase in gradient time to 50 min ($k^* = 7$) is a better choice, as seen in Figure 8(a). Although bands 7/8 are separated with marginal resolution ($R_s = 1.1$), this is a reasonable result at this stage in method development. It should be possible to improve the separation of bands 7/8 by a later change in separation selectivity (see section 3.2).

3.1.2 *Gradient Range*

Once an optimum value of gradient steepness is determined as in Figure 8(a) (2% B min^{-1}), a change in gradient range (values of ϕ_o and ϕ_f) is the next step. The primary goal of changing the initial and final % B of the gradient is to save time. Thus, in the separation of Figure 8(a), there is an empty space in the chromatogram (devoid of peaks) between 4 and 19 min which represents wasted time. This wasted time can be reduced by an increase in the initial value of % B (ϕ_o). Figure 8(b) shows the result of an increase in initial % B from 0 to 40%. When changing the gradient range in order to reduce run time, it is important to keep gradient steepness (2% B min^{-1} in this example) constant to keep k^* constant. So gradient time is reduced from 50 to 30 min in the example of Figure 8(b). The overall appearance of the chromatogram is nearly identical between the two runs of Figure 8(a) and (b), which is typical when changing gradient range while maintaining gradient steepness constant. A slight loss of resolution (from $R_s = 1.1$ to 1.0) is seen for critical band-pair 7/8, but this is acceptable. The result of a further increase in initial % B to 60% is shown in Figure 8(c), with a further decrease in gradient time to 20 min. However, now the initial % B of the gradient is too large; early bands are pushed together with a serious loss

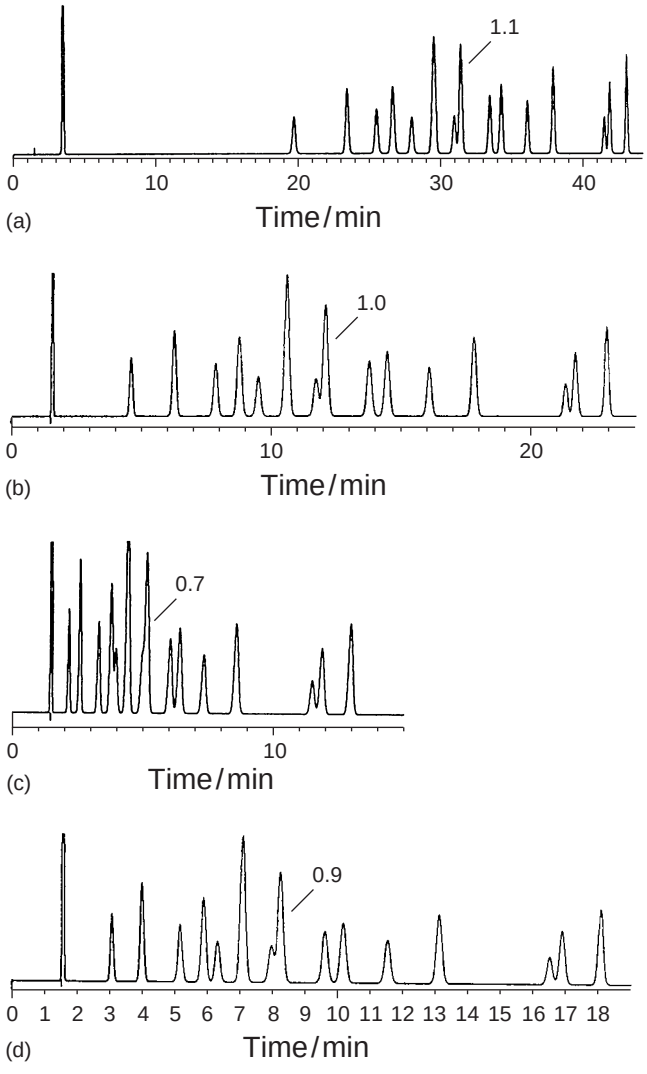

Figure 8 Effect of initial % B (ϕ_o) on gradient separation. Herbicides sample, conditions (other than initial % B) as in Figure 7. (a) 0–100% B in 50 min (2% min^{-1}), (b) 40–100% B in 30 min (2% min^{-1}), (c) 60–100% B in 20 min (2% min^{-1}), (d) 50–100% B in 25 min (optimum). (Reprinted with permission from Snyder et al.[6])

of resolution ($R_s = 0.7$). A more reasonable compromise is shown in Figure 8(d), where 50% B as initial mobile phase results in marginal resolution ($R_s = 0.9$), but run time is reduced to only 25 min. The choice of Figure 8(d) as ''best'' assumes that later changes in separation selectivity will be able to restore the resolution of bands 7/8 to an acceptable value (e.g. $R_s > 1.5$).

The last band in Figure 8(d) leaves the column before the gradient has ended, which represents a further opportunity to reduce wasted time and shorten the run time. This is illustrated for the same separation in Figure 9(a), where the end of the gradient is indicated by an arrow. A general goal in selecting the final % B for the

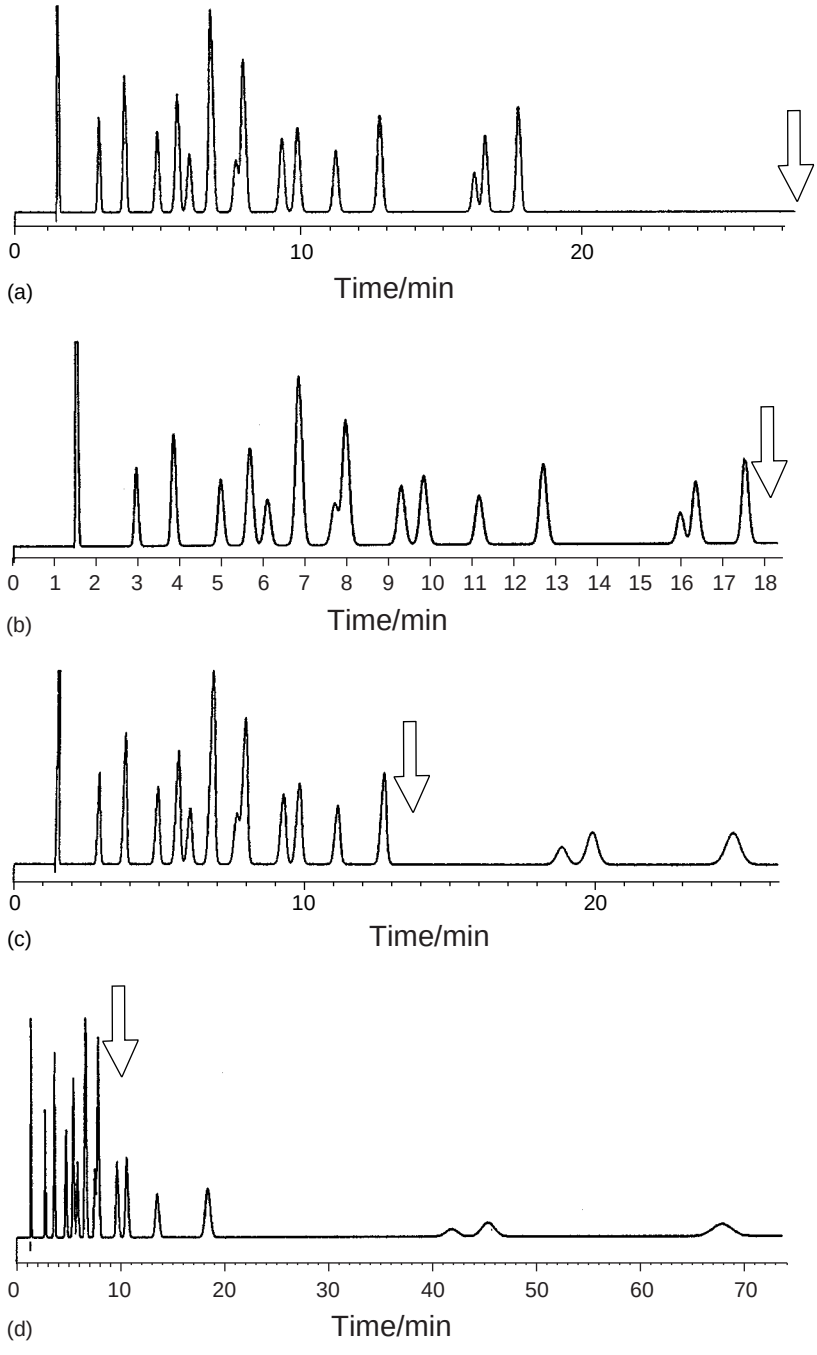

Figure 9 Effect of final % B (ϕ_f) on gradient separation. Herbicides sample, conditions (other than final % B and gradient time) as in Figure 8(d). (a) 50–100% B in 25 min (2% min^{-1}), (b) 50–80% B in 15 min (2% min^{-1}), (c) 50–70% B in 10 min (2% min^{-1}), (d) 50–60% B in 5 min (2% min^{-1}). (Reprinted with permission from Snyder et al.[6])

gradient is to terminate the gradient at the time the last band is eluted. In the present case, the last band elutes when the mobile phase leaving the column is 80% B, suggesting a gradient range of 50–80% B and a gradient time of 15 min. This separation is shown in Figure 9(b). At this point, the run time of the original gradient (0–100% B in 50 min) has been shortened from 43 min (Figure 8a) to 18 min (Figure 9b), with only a minor loss in resolution. If the final % B is further reduced to 70% (Figure 9c) or 60% (Figure 9d), the result is an increase in run time and a broadening of later bands with loss in detection sensitivity, as a result of elution of bands after the gradient ends. These bands show isocratic retention behavior, as seen by broader peaks as retention increases. Elution of

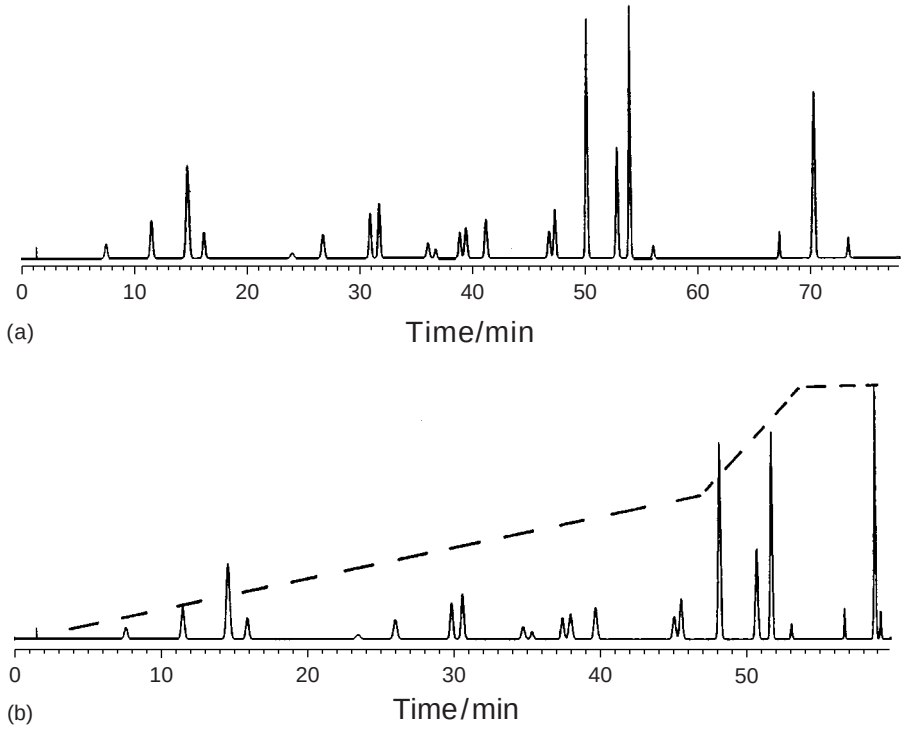

Figure 10 Gradient separations of a 22-component peptide sample (tryptic digest of recombinant human growth hormone); acetonitrile–water gradients with 0.1% trifluoroacetic acid (TFA); 1.0 mL min^{-1}; 40 °C. (a) 0–47% B in 74 min; (b) 2/32/47% B in 0/48/56 min. (Reprinted with permission from Snyder et al.[6])

sample bands after the gradient is always disadvantageous and should be avoided where possible.

3.1.3 Gradient Shape

Linear gradients are generally preferred and should always be tried first. After gradient retention has been adjusted as in Figures 7–9, a change in gradient shape can sometimes be used to improve the separation further.[6] A typical example is shown in Figure 10, where the linear gradient in Figure 10(a) results in a rather long run time, and later bands are widely spaced in the chromatogram (i.e. are overseparated). In cases like this, an increase in gradient steepness toward the end of the chromatogram can result in faster elution of later bands without an unacceptable decrease in resolution. This is illustrated in Figure 10(b), where run time has been reduced from 74 to 60 min, without compromising the separation of later bands. Because gradient steepness is the same in Figure 10(a) and (b) prior to 48 min (the breakpoint in the gradient of Figure 10b), the separation of peaks (#1–15) eluting before 48 min is identical in Figure 10(a) and (b).

A change in gradient steepness also can be used to improve sample resolution, when separation selectivity varies with gradient steepness. This is illustrated in the example of Figure 11, discussed further in the following section.

3.2 Control of Band Spacing

After gradient retention has been optimized as in Figures 7–9, the next step is a change in separation selectivity as a means of avoiding overlapping bands. With only minor changes, the same approach can be used in both gradient and isocratic elution in order to optimize selectivity and band spacing. Table 1 lists some of the variables that are commonly used to change band spacing. Because there are several ways to change selectivity and because improvements in band spacing are unpredictable, optimizing band spacing (α) usually requires more effort than optimizing either retention (k^*) or column efficiency (N).

Some variables (first column in Table 1) affect every kind of sample, whereas other variables (second column of Table 1) only affect samples that are ionizable, i.e. containing acids and/or bases. The ability of a change in the variables of Table 1 to change band spacing (and improve resolution) decreases generally in going down each list;[10] i.e. B-solvent type > column type > gradient steepness > temperature, and pH > ion-pairing > buffer or additive concentration. Also, ionizable samples generally show larger changes in band spacing when a particular variable is changed. On the other hand, the use of gradient steepness and temperature to change selectivity is particularly convenient and is free from

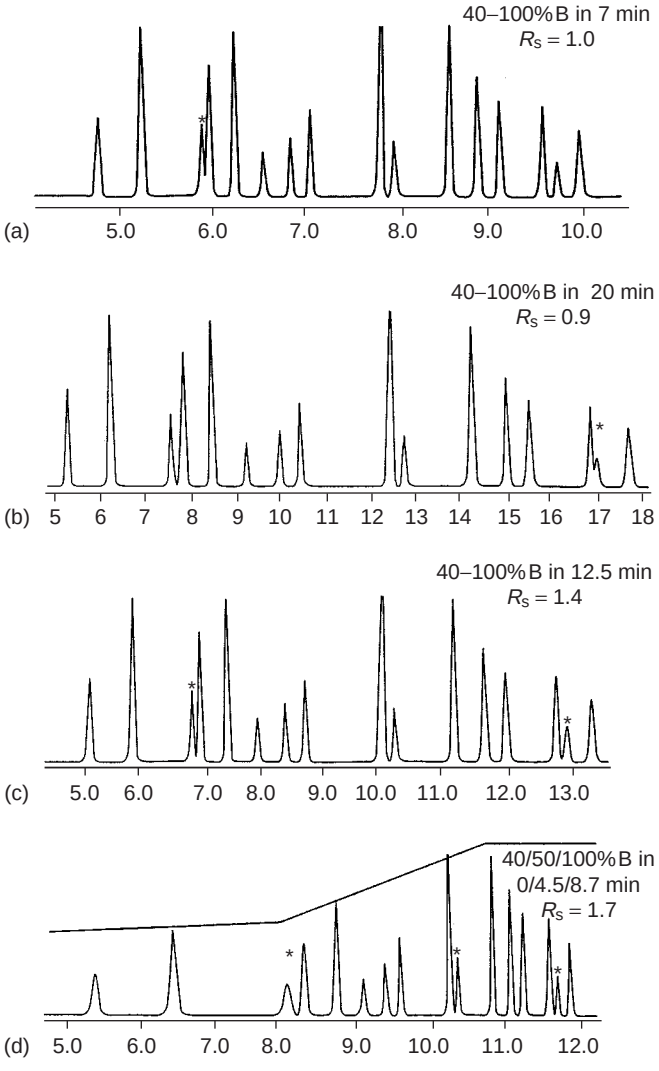

Figure 11 Gradient separation of a polycyclic aromatic hydrocarbon (PAH) sample as a function of gradient steepness. Sample: 16 compounds ranging from naphthalene to indenopyrene. Conditions: 15 × 0.46-cm Supelco LC-PAH column; acetonitrile–water gradients; 2.0 mL min^{-1}; 35 °C. (a) 40–100% B in 7 min, $R_s = 1.0$; (b) 40–100% B in 20 min, $R_s = 0.9$; (c) 40–100% B in 12.5 min, $R_s = 1.4$; (d) 40/50/100% B in 0/45/8.7 min, $R_s = 1.7$. (Reprinted with permission from Snyder et al.[6])

various problems that can arise during the control of selectivity.[10] For this reason, the simultaneous variation of both temperature and gradient steepness is especially useful for controlling selectivity during gradient elution method development (see section 3.4).

Changes in gradient steepness, as in Figure 7, are usually the first step in gradient method development. During these experiments, changes in band spacing may be observed which can be exploited for optimizing selectivity. This is illustrated in Figure 11 for the separation of a mixture of 16 PAHs of environmental concern.

Table 1 Separation variables commonly used to change band spacing (α)

Any sample	Ionizable samples
B-solvent type	Mobile phase pH
Column type	Ion-pairing
Fine-tune t_G (vary k^*)	Buffer or additives concentration
Temperature	

Figure 11(a) and (b) shows separations for gradient times of 7 and 20 min. In the steeper gradient of Figure 11(a), band-pair 3/4 is ''critical'', i.e. is the least resolved. In the shallower gradient of Figure 11(b), band-pair 14/15 is critical. This suggests that an intermediate gradient time will provide a better overall separation, as is seen in Figure 11(c) for a gradient time of 12.5 min. Because early bands 3/4 prefer a shallower gradient, while later bands 14/15 prefer a steeper gradient, an even better separation is predicted for a segmented gradient which is shallow initially and steep at the end. This is confirmed in the separation of Figure 11(d), where all bands are separated to baseline ($R_s \geq 1.7$). This example illustrates one of the (few) unique differences between isocratic and gradient elution. In isocratic elution, because experimental conditions cannot be changed during the separation, maximum resolution for the critical band-pair(s) usually represents a compromise for the overall sample, as in Figure 11(c) for this gradient separation. However, because conditions can be changed during gradient elution, it is possible to optimize conditions separately for early-eluting with respect to late-eluting bands.

Figure 12 shows a similar change in band spacing for the separation of a mixture of substituted phenols. The separation in Figure 12(a) with an acetonitrile–water gradient provides baseline separation of all compounds except bands 8/9. A change in B-solvent from acetonitrile to methanol in Figure 12(b) results in the separation of band-pair 8/9, but with some loss in resolution of bands 2 and 3. In Figure 12(c), the use of a gradient that begins with 20% acetonitrile–water and ends with 100% methanol provides the best overall separation, similar to the example of Figure 11(d).

A final example of selectivity control in gradient elution is shown in Figure 13. Here, five synthetic peptides are separated with acetonitrile–water gradients, with a variable amount of TFA added to the mobile phase as a combined buffer and ion-pairing agent. A change in the concentration of an ion-pairing agent can provide very large changes in band spacing, when the sample contains compounds whose molecular charge varies widely (as in the present case, where the band numbering in Figure 13 indicates the net positive charge on each compound). An examination of these six chromatograms shows that the addition of 0.05% TFA to the mobile phase provides the

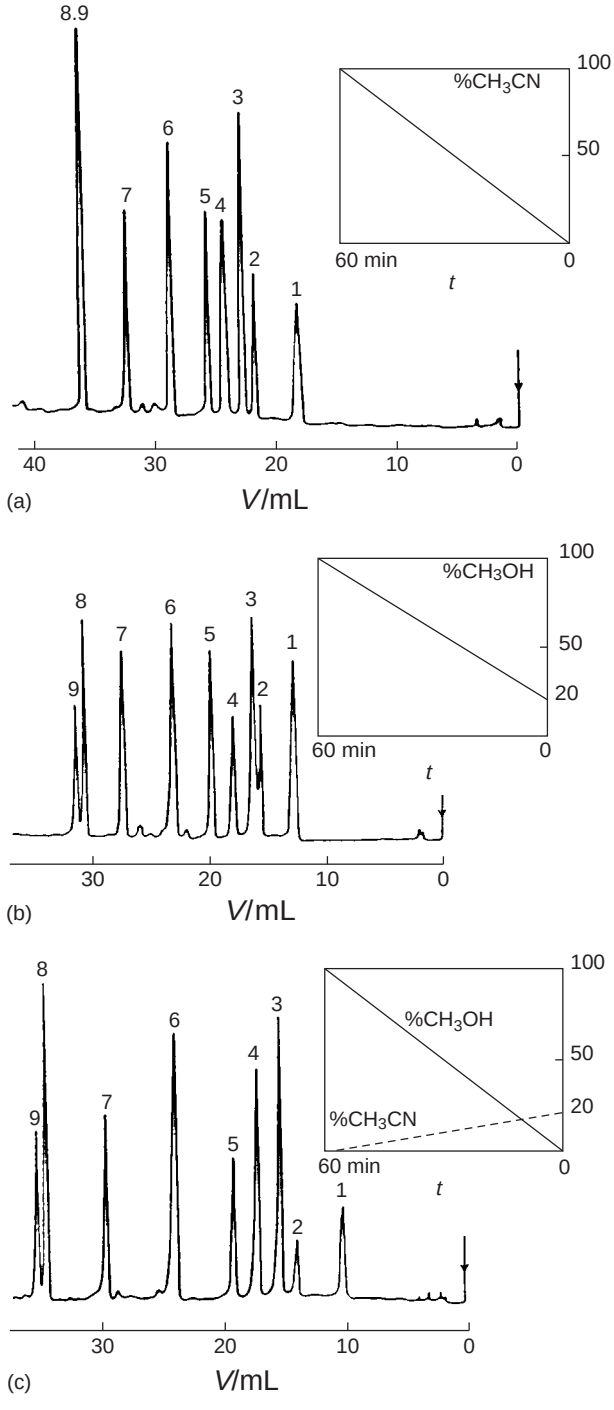

Figure 12 Effect of B-solvent on separation of a mixture of phenols. Conditions: 30 × 0.42-cm C_{18} column; gradients as in figure; 1.0 mL min^{-1}; ambient. See text for details. (Reprinted from Jandera et al.[11] *J. Chromatogr.*, Copyright (1981) with permission of Elsevier Science.)

best overall separation. Because bands 1 and 2a are the critical pair in this chromatogram, it can be predicted that a slight increase in TFA (to 0.06 or 0.07%) should provide an even better resolution for the critical band pair.

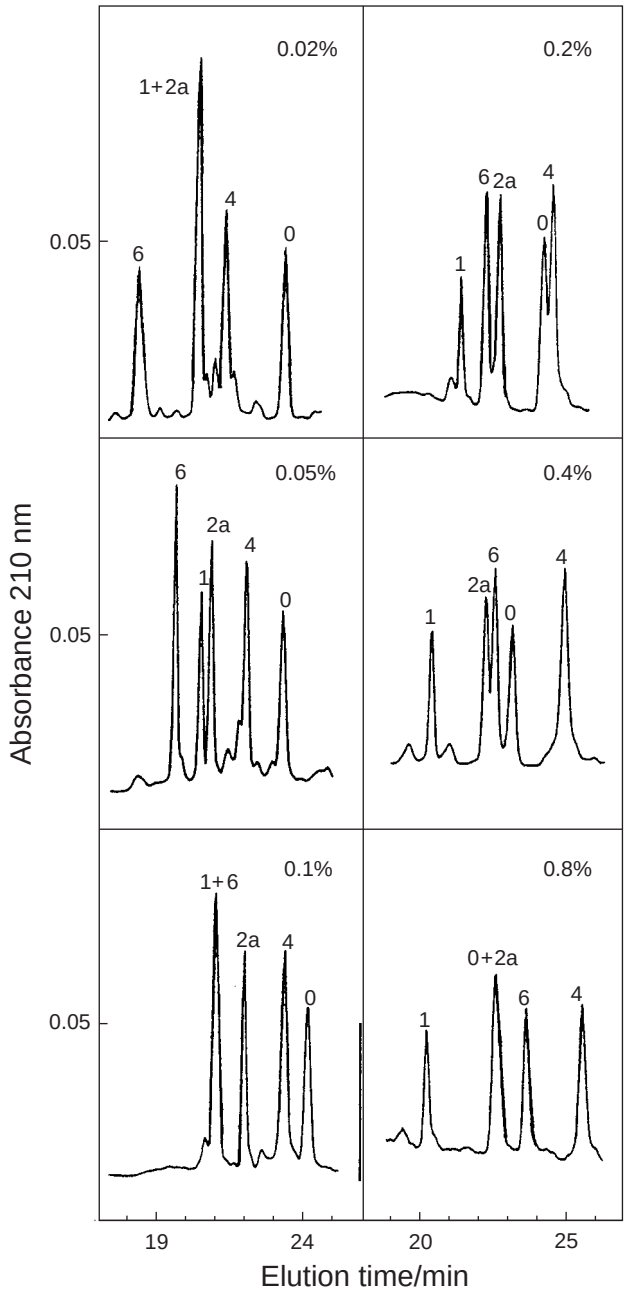

Figure 13 Separation of basic peptides as a function of TFA concentration. Conditions: 25 × 0.46-cm C_{18} column; A-solvent, water plus indicated %-TFA; B-solvent, acetonitrile plus indicated %-TFA; gradient from 0% B at 1% min^{-1}; 1.0 mL min^{-1}; 26 °C. (Reprinted with permission from Guo et al.[12] *J. Chromatogr.*, Copyright (1987) with permission of Elsevier Science.)

3.3 Effects of Column Dimensions, Particle Size and Flow Rate

Varying the column plate number N in either isocratic or gradient method development is usually best delayed until after retention (k or k^*) and selectivity (α) have been

optimized. At this point in method development, N can be increased for an increase in sample resolution. Also, since an increase in N usually requires a longer run time, a decrease in N can be used to save separation time, if sample resolution is greater than necessary. A change in N can be achieved by a change in "column conditions": column length L, particle size of the column packing d_p, or flow rate F. An increase in N usually results from an increase in L or a decrease in d_p or F. (Changes in N with F are more pronounced for larger (e.g. 10-μm d_p) particle packings, and may be of little practical benefit for 3–5-μm d_p materials.)

In isocratic separation, a change in L or F has no effect on values of k or α, so optimizing column conditions can be carried out independently of changes in k or α. That is, further improvements in an isocratic separation by varying L or F can be made without worrying about any corresponding changes in k or α. Because k^* in gradient elution depends on both column size (V_m) and F (Equation 4), changes in k^* and α can result when varying L or F. This is illustrated in Figure 14, for the separation of a peptide sample (tryptic digest of the protein myoglobin). Entire chromatograms are shown in Figure 14(a) for two different flow rates (0.5 and 1.5 mL min^{-1}), other conditions being the same. Expanded portions of each chromatogram (brackets, arrows) are reproduced in Figure 14(b) for easier visibility. It is seen in Figure 14(b) that the change in flow rate from 0.5 to 1.5 mL min^{-1} results in major changes in band spacing: coelution of bands 5 and 5a at 1.5 mL min^{-1}, coelution of bands 6 and 6a at 0.5 mL min^{-1}, and a reversal of the retention of bands 6b and 7. Changes in selectivity such as this, as a result of varying F or L, would not occur in isocratic separation.

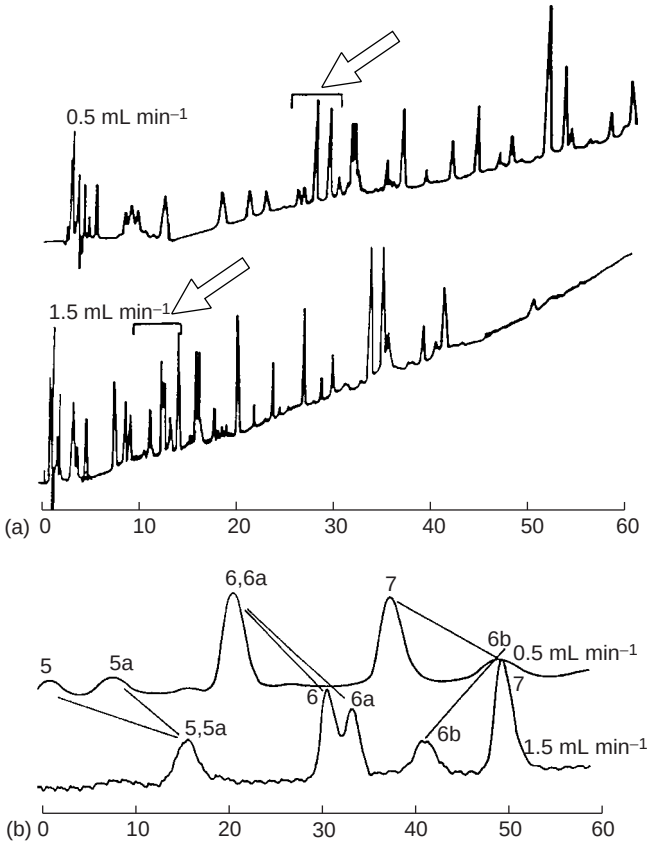

Figure 14 Effect of flow rate on selectivity in gradient elution. Sample: peptides from tryptic digest of myoglobin. Conditions: 8 × 0.62-cm C_8 column; 10–70% acetonitrile–water gradient in 60 min (0.1% TFA added to each solvent); 35 °C; flow rates as indicated. (a) Entire chromatograms; (b) expansions of parts of each chromatogram from (a) (see arrows). (Reprinted with permission from Glajch et al.[13])

Because a change in column conditions generally is delayed until after band spacing has been optimized, further changes in band spacing when optimizing N are undesirable. Changes in α similar to those of Figure 14 can be avoided when changing column conditions by maintaining k^* constant. Referring to Equation (4) (and noting that V_m is proportional to L), if k^* is to remain constant while varying L or F, then it is necessary to maintain ($t_G F/L$) constant. This can be achieved by varying gradient time in inverse proportion to F and in direct proportion to L.

3.4 Computer Simulation

The theory of reversed-phase gradient elution has now been developed to a point where it is possible to predict retention times, t_R, and bandwidths, W, as a function of experimental conditions, when the parameters of Equation (1) (k_w and S) are known. Thus, Equation (9)

$$t_R = \frac{t_o}{1.15k^*} \log\left(\frac{2k_o}{k^*} + 1\right) + t_o + t_D \qquad (9)$$

Here, t_o is the column dead time, k_o is the value of k at the start of the gradient, and t_D is the equipment hold-up or "dwell" time (see section 4.2). Similar predictions of gradient retention time can be made for ion-exchange separation.[14,15] For the case of reversed-phase gradient elution, two experimental runs suffice to determine k_w and S for each compound in the sample.[8] If these values are entered into computer simulation software (Snyder et al.,[2] and Snyder et al.[6] page 448–452), it is then possible to predict separation for

Table 2 Predictions of gradient separation by means of computer simulation (Snyder et al.[2] and Snyder et al.[6] 448–452). Based on two experimental runs, the changes in conditions listed can be predicted

Gradient time	Column length and diameter
Initial and final % B	Particle size
Gradient shape	Flow rate

various changes in conditions as summarized in Table 2. Thus, a small number of experiments during method development allow the prediction of a large number of separations, leading to the speedy optimization of the various conditions of Table 2. Computer simulation is also useful for studying the general consequences of different changes in conditions, as in the examples (computer simulations) of Figures 7–9.

Computer simulation has recently been extended[16,17] to allow the simultaneous optimization of temperature and gradient steepness, in addition to the other conditions of Table 2. This is illustrated in Figure 15 for the separation of a mixture of 13 commercial herbicides. Predictions of separation as in Figure 15 require four experimental runs, where two different temperatures T and gradient times t_G are used: 30 and 40 °C and 40 and 120 min in this example. Following entry of the resulting data (retention times, band areas, experimental conditions) into the computer simulation program, various predictions as in Table 2 are possible. Usually it is desirable to request a resolution map, which plots critical resolution (for the least resolved band

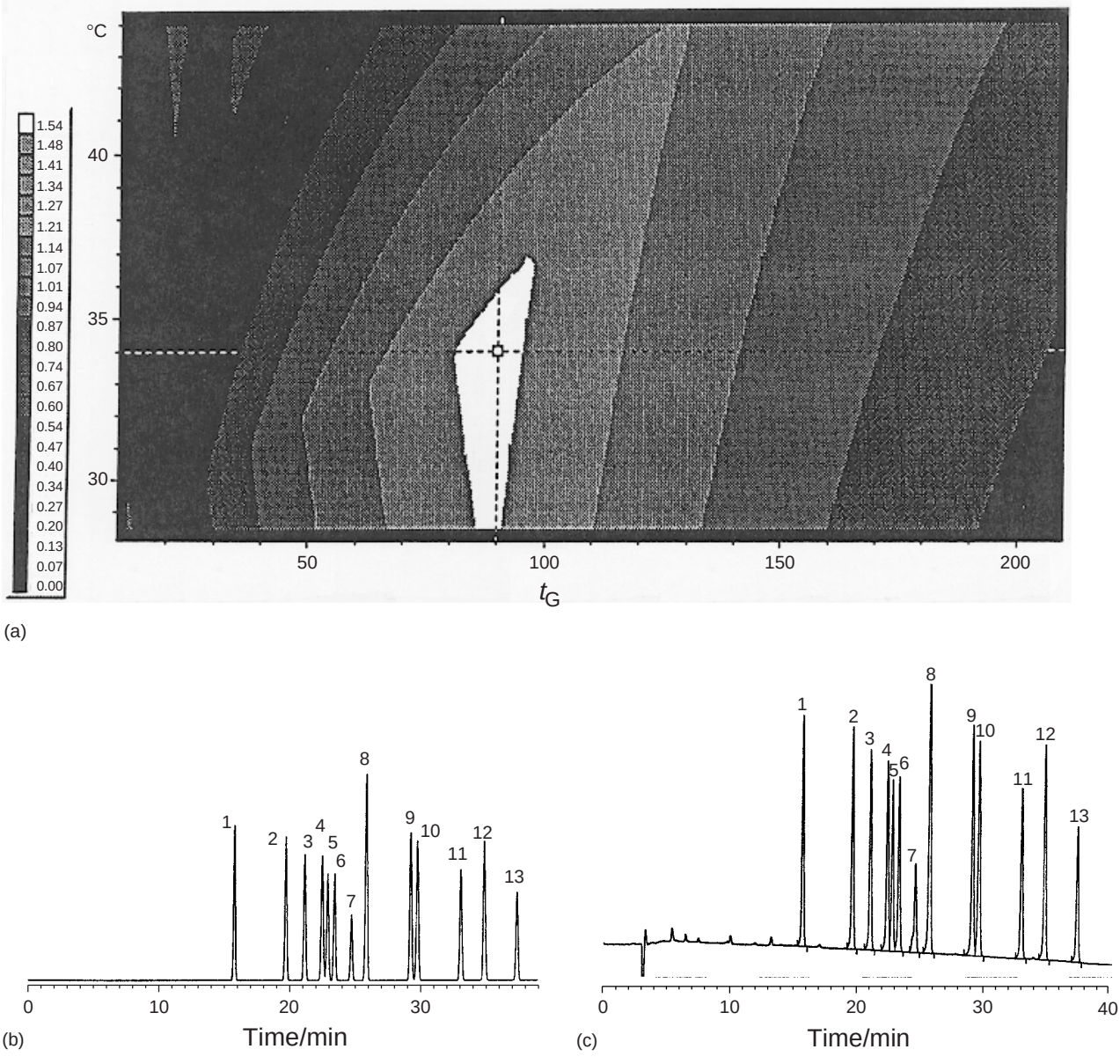

Figure 15 Computer simulation for the optimization of the separation of 13 herbicides. Conditions: 25 × 0.46-cm C_{18} column, 20–95% acetonitrile–buffer gradients, 1.0 mL min^{-1}. (a) Resolution map; (b) predicted chromatogram for 30 °C and 80-min gradient; (c) experimental chromatogram for conditions of (b). (Reprinted from Dolan et al.[16] *J. Chromatogr.*, Copyright (1998) with permission from Elsevier Science.)

pair) as a function of gradient time and temperature in this example. Figure 15(a) shows the resulting resolution map, with maximum resolution indicated for 34 °C and a 90-min gradient time. The predicted separation is shown in Figure 15(b), and the experimental confirmation of this prediction is given in Figure 15(c). Retention time predictions by means of computer simulation are typically accurate to better than 1%, as in the example of Figure 15.

4 "NONIDEAL" EFFECTS IN GRADIENT ELUTION

Apart from its somewhat greater complexity, as summarized in sections 1–3, gradient elution is subject to some additional problems which are absent (or uncommon) in the case of isocratic elution. The effective use of gradient elution for practical applications requires an awareness of these "non-ideal" effects (see Snyder et al.[6] for further details).

4.1 Nonequilibration of the Column

When the mobile phase is changed in isocratic elution, it is necessary to flush the column with a sufficient volume of the new mobile phase to equilibrate the column. However, once the column is equilibrated, samples can be successively injected without further concern for column equilibration. In the case of gradient elution, the mobile phase at the end of each separation differs from the beginning mobile phase for the next sample injection. This requires that the column be flushed with a sufficient volume of the initial mobile phase to achieve equilibration of the column, prior to injection of the next sample. Although column equilibration typically can be achieved by flushing with 5–10 column volumes of the starting mobile phase, the exact intersample flush volume must be experimentally confirmed for each gradient method. Equilibration volume can be checked by selecting a starting volume (e.g. 5 column volumes) and making a run. Then the equilibration volume is increased (e.g. to 10 column volumes) and the run is repeated. The equilibration volume will be satisfactory when further increases in volume cause no change in sample retention (poor equilibration generally appears as shorter retention times, especially for early eluting peaks). Because of the need for column equilibration between each sample injection, total run time in gradient elution usually is somewhat greater than for isocratic separation. However, extremely fast gradient separations (e.g. 1–2 min run times) are nevertheless possible.[18]

Another practical consequence of column equilibration in gradient elution regards the choice of the initial % B. At very high water content (e.g. >97%), the bonded phase on reversed-phase columns tends to undergo a configuration change or "collapse". Resolvation of the phase following collapse can be slow when compared to normal column equilibration, further extending the turnaround time for routine operation. The magnitude of this equilibration problem varies with the stationary phase and sample type. It is prudent to use ≥5% B as the starting solvent for gradient elution to avoid this phenomenon. Thus, most workers prefer 5–100% B gradients over 0–100% B.

4.2 Gradient Delay and Rounding

The design of the equipment used for gradient elution (Figure 5) results in some distortion of the gradient delivered to the column, as illustrated in Figure 3(f). Thus, following the combination of A and B solvents, the resulting mobile phase must pass through the gradient mixer, various connecting lines and filters, the autosampler, and (in the case of low-pressure gradient mixing) the pump. As a result of this movement of mobile phase through the system, the gradient will be delayed and to some extent dispersed or rounded. This means that the programmed gradient can differ significantly from the actual gradient. Because these differences are peculiar to each equipment design, gradient methods tend to be less transferable than isocratic methods.

The main effect of equipment differences on separation is usually the result of differences in the hold-up or "dwell" volume, i.e. the volume swept out by the mobile phase as it passes through the mixer onto the column. If the equipment dwell volume is known, it is possible to predict its effect on separation by means of computer simulation (see section 3.4). In most cases, minor changes in gradient conditions (initial % B, gradient steepness) can be used to adjust for differences in dwell volume.

4.3 Equipment and Baseline Problems

The added complexity of the equipment used for gradient elution can create additional problems that are less common or nonexistent with isocratic systems. In practice, however, many workers use the same equipment for isocratic as for gradient separations because of the convenience of on-line mixing. The potential problems outlined here can be overcome, for the most part.

4.3.1 Mobile Phase Outgassing

When organic solvent/water blends are prepared, the resulting mixture usually has a lower capacity for dissolved air than the pure solvents. When solvents are mixed on-line to form the gradient, outgassing can occur. If these bubbles pass through the pump, check-valve failure, pressure fluctuations, and flow variation can occur. Outgassing tends to be more of a problem

for low-pressure mixing systems because the solvents are mixed at (or below) atmospheric pressure. High-pressure mixing tends to compress the air, keeping it in solution, so fewer bubble problems are experienced. For both types of equipment, degassing, as with helium sparging, will give more reliable operation.

Sometimes air bubbles are generated when the mobile phase returns to atmospheric pressure at the end of the column. If these bubbles pass through the detector, baseline disturbances result. In addition to mobile phase degassing, the use of a back-pressure regulator after the detector usually will alleviate detector bubble problems.

4.3.2 Proportioning Problems

The actual gradient reaching the column differs from that programmed into the system controller. As illustrated in Figure 3(f), rounding at the ends of the gradient occurs because of the washout characteristics of the equipment. Further distortion of the gradient can occur if the equipment is not operating properly. An easy way to check this is to place a UV (ultraviolet)-absorbing solute in the B-reservoir and run a gradient with a short piece of tubing replacing the column. For example, water can be used for the A solvent and 0.1% acetone in water as the B solvent, with a 0–100% B gradient run over 20 min. If the gradient is straight in the middle, as in Figure 3(f), the equipment appears to be operating properly. If the gradient is bowed or an offset is observed mid-gradient, the system needs to be serviced.

Proportioning accuracy can be checked by programming a series of isocratic steps, as in Figure 3(e), for example 10% steps over a 0–100% B range, using the acetone–water mobile phase. The steps should approach stair steps and be even in size, as in Figure 16(b). If uneven or distorted steps are observed, as in Figure 16(a), equipment service is required. Most manufacturers specify proportioning accuracy of ±0.5–1.0% B; from a practical standpoint, ±1.0% (as in Figure 16b) is satisfactory if the steps are even and well-shaped.

4.3.3 Baseline Drift

Baseline drift in gradient elution is caused by changes in temperature or mobile phase absorbance. Temperature drift should not be a problem if the column is operated in a column oven – this is a wise choice for isocratic methods as well, because retention and selectivity are influenced by temperature.

Baseline drift in temperature-controlled systems may be observed when the A and B solvents have different UV absorbance (assuming UV detection). This is illustrated in Figure 17. Two general trends are observed. First, the organic solvents used for reversed-phase usually have more absorbance than water and the absorbance

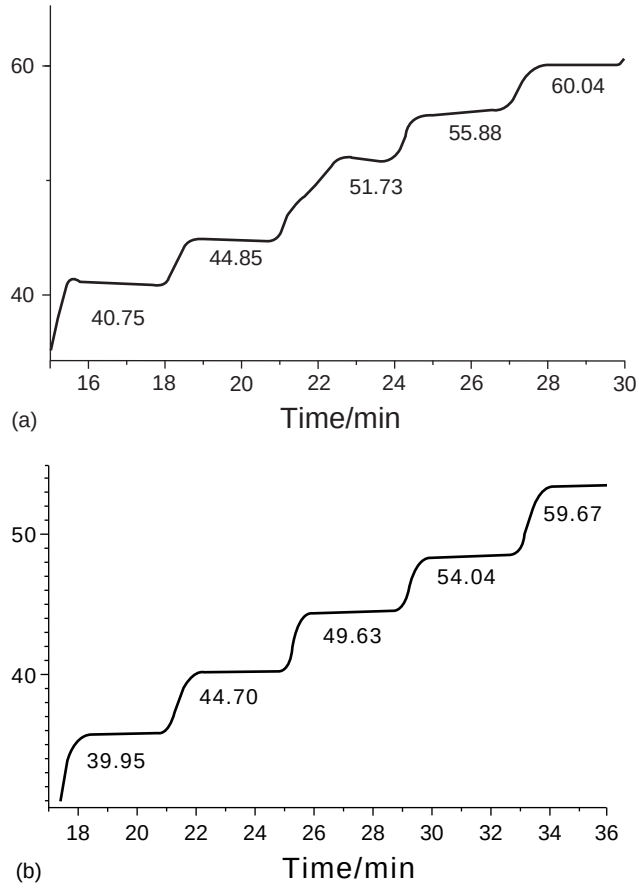

Figure 16 Sample results for step test of pump proportioning accuracy for (a) poor pump performance and (b) acceptable pump performance. True % B values shown below each step. See text for details.

varies with the solvent. This is seen for the larger drift with the water–methanol gradient of Figure 17(a) when compared to water–acetonitrile (Figure 17b). Second, as the detection wavelength is reduced, the drift is magnified (compare Figure 17c and 17a). As long as the baseline drift is acceptable (i.e. the baseline stays on-scale), most data systems can successfully integrate peaks, so quantitation should be satisfactory. The lower background absorbance of acetonitrile at low wavelengths (Figure 17a versus Figure 17b) makes it the solvent of choice for low-wavelength gradient applications. An additional advantage of acetonitrile over methanol for gradient elution is its lower viscosity when mixed with water. This results in substantially lower column back pressure, allowing higher flow rates and thus shorter run times.

4.3.4 Baseline Artifacts

One problem with gradient elution that is not experienced with isocratic separation is the presence of background peaks in the gradient. A blank gradient always should

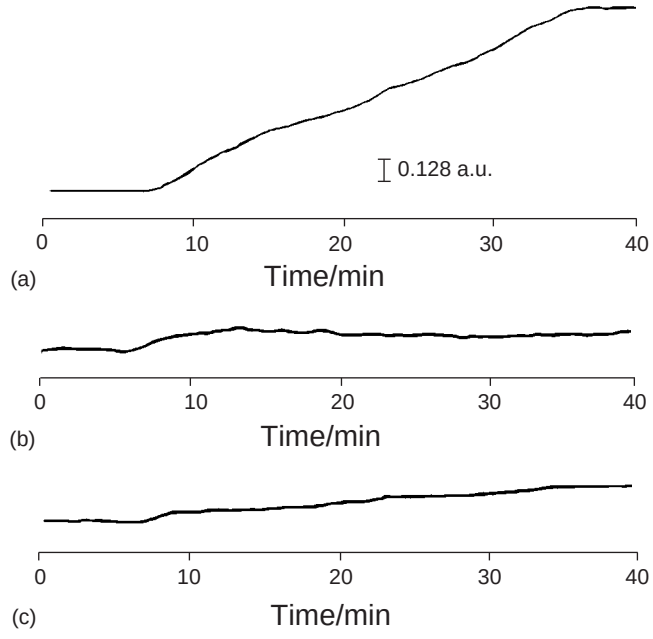

Figure 17 Absorbance profiles for water–organic gradients; wavelengths as noted. (a) Water versus methanol, 215 nm; (b) Water versus acetonitrile, 215 nm; (c) Water versus methanol, 254 nm. (Data from the *HPLC Solvent Reference Manual*.[19] Reproduced courtesy of Mallinckrodt Baker Inc.)

be run prior to a series of samples to determine the magnitude of background peaks. In extreme cases, a blank baseline can appear as shown in Figure 18(a). The observed peaks represent contaminants in the water phase that collect at the head of the column during equilibration and then elute off as if they were sample components when the gradient is run. Use of a better quality of water (Figure 18b) usually will remove most or all of the unwanted peaks. In particularly difficult cases, on-line cleanup[20,21] can be used to advantage. Baseline artifacts will be minimized if HPLC-grade solvents and additives are used.

4.3.5 Buffer Precipitation

Precipitation of buffers in the HPLC system can be a problem, particularly when acetonitrile is used as the organic phase, owing to the poor solubility of buffer salts in high percent-organic solutions. Methanol usually is less of a problem than acetonitrile. Although a buffer may be soluble in a bulk solution, for example in 80% acetonitrile, the mixing characteristics of HPLC systems can cause buffer to be mixed directly with 100% acetonitrile under conditions that will precipitate the buffer. Precipitated buffers are difficult to remove and can foul check valves and column frits. A four-fold approach to check for and avoid such problems usually is successful. First, add the buffer phase drop-wise to the organic phase in a test tube.

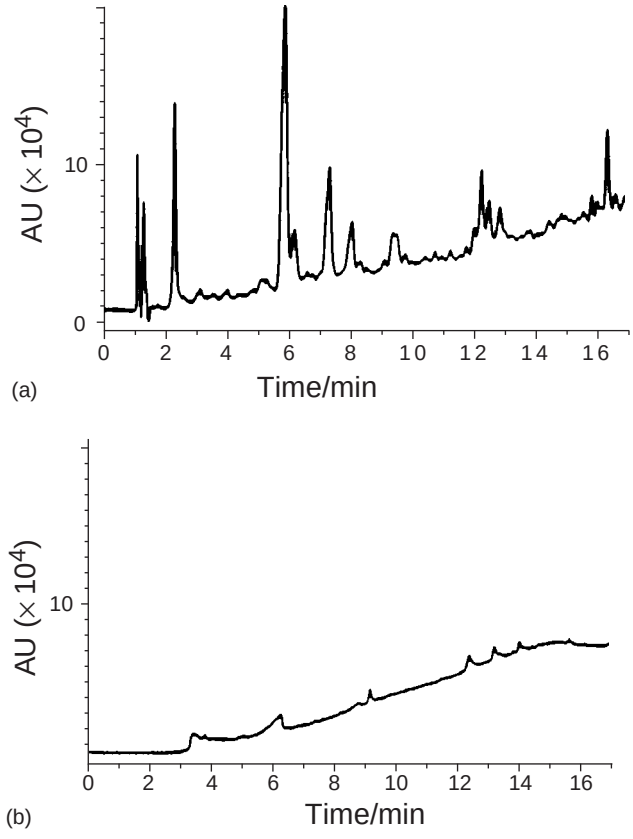

Figure 18 Effect of water purity on gradient baselines. Contaminated water (a) and purified water (b). 0–83% acetonitrile–water gradients at 255 nm in 13 min followed by hold; 15 × 0.46-cm C_{18} column, 1.5 mL min^{-1}, 2.3 mL dwell volume.

If any cloudiness is observed when the buffer contacts the organic, it is likely to happen in the HPLC system as well. Second, keep the buffer concentration low (e.g. $\leq$25 mM). Third, premixing the buffer and organic may be helpful. For example, mix 5% buffer in the B reservoir and 5% B-solvent in the A-reservoir. Finally, avoid 100% acetonitrile gradients. Often running a 5–80% acetonitrile–buffer gradient will give the desired gradient conditions without the potential for buffer precipitation as the gradient nears 100% acetonitrile.

LIST OF SYMBOLS

A, B	refers to weak (A) and strong (B) solvents that comprise the mobile phase in reversed-phase HPLC
C_2, C_3, etc.	dimethyl, methylethyl, etc. *o*-phthalates
d_p	column packing particle size (μm)
F	flow rate (mL min^{-1})
k	isocratic retention factor

k^*	gradient retention factor (Equation 4)
k_o	isocratic value of k for a solute at start of gradient
k_w	isocratic value of k for water as mobile phase ($\phi = 0$)
L	column length (cm)
N	column plate number (Equation 6)
R_s	resolution of two adjacent bands, equal to difference in retention times divided by W
S	solute parameter defined in Equation (1)
t	time after start of gradient
t_D	equipment hold-up or ''dwell'' time (min); $t_D = V_D/F$
t_o	column dead time (min); $t_o = V_m/F$
t_G	gradient time (min)
t_R	retention time (min)
V_D	equipment hold-up or ''dwell'' volume (mL)
V_m	column dead volume (mL)
W	baseline bandwidth (min)
α	separation factor, equal to ratio of k-values for two adjacent bands
ϕ	volume fraction of B in mobile phase A–B
ϕ^*	value of ϕ for mobile phase at column mid-point; corresponds to k^*
ϕ_o, ϕ_f	values of ϕ at start and end of gradient

ABBREVIATIONS AND ACRONYMS

HPLC	High-performance Liquid Chromatography
LSS	Linear Solvent Strength
PAH	Polycyclic Aromatic Hydrocarbon
TFA	Trifluoroacetic Acid
UV	Ultraviolet

REFERENCES

1. M.A. Quarry, R.L. Grob, L.R. Snyder, 'Measurement and Use of Retention Data from High Performance Gradient Elution. Correction for ''Nonideal'' Processes Originating Within the Column'. *J. Chromatogr.*, **285**, 19–51 (1984).
2. L.R. Snyder, J.W. Dolan, D.C. Lommen, 'DryLabR Computer Simulation for HPLC Method Development'. *J. Chromatogr.*, **485**, 65–89, 91–112 (1989).
3. P. Jandera, J. Churacek, *Gradient Elution in Column Liquid Chromatography*, Elsevier, Amsterdam, 1985.
4. L.R. Snyder, 'Gradient Elution', in *High-performance Liquid Chromatography. Advances and Perspectives*, ed. Cs. Horvath, Academic Press, New York, 207–316, Vol. 1, 1980.
5. L.R. Snyder, M.A. Stadalius, 'HPLC Separations of Large Molecules. A General Model', in *High-performance Liquid Chromatography. Advances and Perspectives*, ed. Cs. Horvath, Academic Press, New York, 195–312, Vol. 4, 1986.
6. L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2nd edition, Wiley-Interscience, New York, 350–401, 1997.
7. C.F. Poole, S.A. Schuette, *Contemporary Practice of Chromatography*, Elsevier, Amsterdam, 366, 1984.
8. L.R. Snyder, J.W. Dolan, 'The Linear-solvent-strength Model of Gradient Elution', *Adv. Chromatogr.*, **38**, 115–187 (1998).
9. L.R. Snyder, J.W. Dolan, 'Initial Experiments in HPLC Method Development. I. Use of a Starting Gradient Run', *J. Chromatogr., A*, **721**, 1 (1996).
10. L.R. Snyder, 'Changing Reversed-phase HPLC Selectivity: Which Variables should be Tried First?', *J. Chromatogr., B*, **689**, 105–115 (1997).
11. P. Jandera, J. Chracek, H. Colin, 'Gradient Elution in Liquid Chromatography. XIV. Theory of Ternary Gradients in Reversed-phase Liquid Chromatography', *J. Chromatogr.*, **214**, 35–46 (1981).
12. D. Guo, C.T. Mant, R.S. Hodges, 'Effect of Ion-pairing Reagents on the Prediction of Peptide Retention in Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **386**, 205–222 (1987).
13. J.L. Glajch, M.A. Quarry, J.F. Vasta, L.R. Snyder, 'Separation of Peptide Mixtures by Reversed-phase Gradient Elution. Use of Flow Rate Changes for Controlling Band Spacing and Improving Resolution', *Anal. Chem.*, **58**, 280–285 (1986).
14. T. Sasagawa, Y. Sakamoto, T. Hirose, T. Yoshida, Y. Kobayashi, Y. Sato, 'Prediction of Retention Times in Ion-exchange Chromatography', *J. Chromatogr.*, **485**, 533–540 (1989).
15. Y. Baba, M.K. Ito, 'Optimization of Gradients in Anion-exchange Chromatography of Oligonucleotides Using Computer-assisted Retention Prediction and a High-performance Liquid Chromatography Simulation System', *J. Chromatogr.*, **485**, 647–655 (1989).
16. J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, D.C. Locke, D.L. Saunders, L. Van Heukelem, T.J. Waeghe, 'A Computer-assisted Strategy for HPLC Method Development. I. Simultaneous Variation of Temperature and Gradient Steepness', *J. Chromatogr., A*, **803**, 1–31 (1998).
17. J.W. Dolan, L.R. Snyder, D.L. Saunders, L. Van Heukelem, 'A Computer-assisted Strategy for HPLC Method Development. II. Simultaneous Changes in Temperature and Gradient Steepness Combined with Change in One or More Other Variables', *J. Chromatogr., A*, **803**, 33–50 (1998).

18. J.B. Li, J. Morawski, 'Strategies for Faster Gradient Chromatography', *LC/GC Mag.*, **16**, 468–476 (1998).
19. *HPLC Solvent Reference Manual*, J.T. Baker, Phillipsburg, NJ, 7–2, graphs A1, A2, and A4, 1985.
20. P-.L. Zhu, L.R. Snyder, J.W. Dolan, 'Improved Baselines in Gradient Elution', *J. Chromatogr., A*, **718**, 429–435 (1996).
21. J.W. Dolan, J.R. Kern, T. Culley, 'Water Problems – A Case Study', *LC/GC*, **14**, 202–208 (1996).

Ion Chromatography

Donald J. Pietrzyk
University of Iowa, Iowa City, USA

1 Introduction 1
2 Basic Principles of Ion Exchange 3
2.1 Ion-exchange Equilibrium 3
2.2 Properties of Ion Exchangers 3
3 Ion Exchangers in Ion Chromatography 5
3.1 Polymeric Ion Exchangers 5
3.2 Latex-based Ion Exchangers 7
3.3 Silica-based Ion Exchangers 7
3.4 Other Exchangers Used in Ion Chromatography 8
4 Ion Exchange Separations 8
5 Suppressed Ion Chromatography 9
5.1 Suppression by an Ion Exchanger 9
5.2 Membrane Suppression in Ion Chromatography 11
5.3 Other Suppression Strategies in Ion Chromatography 12
6 Eluents for Suppressed Ion Chromatography 13
7 Nonsuppressed Ion Chromatography 15
8 Eluents for Nonsuppressed Ion Chromatography 15
9 Detection in Ion Chromatography 16
10 Applications of Ion Chromatography 17
11 Ion Exclusion Chromatography 18
12 Ion Interaction Chromatography 20
Abbreviations and Acronyms 22
Related Articles 23
References 23

Ion chromatography (IC), which is an important methodology within high-performance liquid chromatography (HPLC), is a chromatographic separation/detection strategy that is applied to the determination of inorganic and organic analyte anions and cations. The methodology consists of three main components: (1) a low-capacity ion-exchange column, where the separation occurs; (2) a suppressor system, which reduces conductivity due to the mobile phase electrolyte; and (3) a conductivity detector, which is used to detect the analyte ions. Anionic analytes are separated on an anion-exchange column and analyte cations are separated on a cation-exchange column. IC, which is routinely applied in water, environmental, health, and food analysis, is a rapid, sensitive, and accurate method for the trace determination of complex mixtures of analyte ions, particularly analyte anions, where no other method offers similar advantages and scope of application.

Today, the term IC defines a broader group of methodologies applicable to the separation of analyte ions and now also defines the following strategies. Nonsuppressed IC of analyte anions and cations is possible. In addition, analyte ions can be separated with an ion interaction reagent (ion pair reagent) as a mobile phase additive and a reversed stationary phase column rather than an ion-exchange column; this is often called ion interaction chromatography (IIC). Ion exclusion chromatography (IEC), which is a separation of weakly dissociated acid and base analytes on an anion exchanger, is also an important methodology within the field of IC.

1 INTRODUCTION

The concept of what is now called IC was first described by Small, Stevens, and Bauman in 1975.[1] Their paper, which is a classic work in the development of modern liquid column chromatography, describes methodology for the separation of inorganic ions on ion exchangers and detection of the separated ions by conductivity. The keys to their success were twofold. First, a low-capacity ion exchanger was used for the analytical column to effect the separation. Separation occurs because of differences in the way that individual analyte ions and electrolyte ions of similar charge in the mobile phase, which are required for the separation, compete electrostatically for the ion-exchange sites on the ion exchanger. Second, a high-capacity exchanger was used as a suppressor column in series following the analytical column to reduce the effects of mobile phase background electrolyte on conductance but still maintain the analyte in the column effluent as an electrolyte so that it could be detected with sensitivity by conductivity. Separations were accomplished at favorable column efficiencies, low detection limits (DLs), and at fast analysis times. Other separation methods or analytical methodologies, including ion-exchange separation procedures, available at that time, did not offer these advantages. The suppression strategy is essentially a form of postcolumn chemistry, that is, the column effluent containing the mobile phase and the analyte are altered so that the system permits conductivity detection of only the analyte with high sensitivity.

In the work of Small et al.[1] mixtures of inorganic anions, such as Cl^-, Br^-, I^-, SO_4^{2-}, and CO_3^{2-}, and cations, such as Na^+, K^+, Rb^+, Cs^+, NH_4^+, Mg^{2+}, and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Ca^{2+}, were completely separated on anion and cation exchangers, respectively, in 10–15 min at analyte levels as low as 1 ppm depending on the analyte and the mixture. For example, Na^+ and K^+ were readily determined in urine, blood serum, and fruit juices while anions were determined in treated and untreated water. Often the samples could be handled in a dilute and shoot strategy, that is, the sample is diluted and an aliquot of this solution is introduced into the column for the separation. Thus, the sample matrix did not have to be removed from the sample prior to analysis. It was also possible through pH control to separate simple organic amines as cations and organic acids as anions. For the first time mixtures of simple inorganic and organic anions and cations could be separated and determined quickly, often in minutes, and at trace to ultratrace levels often in parts per million to sub-parts per million quantities.

Small et al.[1] referred to the methodology as ''Conductive Chromatography''. However, as the significance of this work became recognized and was commercially developed this separation methodology quickly became known as IC. The introduction and subsequent development of IC is of tremendous practical consequence in quantitative analytical chemistry and applications of liquid column chromatography. This is illustrated by Figure 1(a) and 1(b) where ion chromatograms for the separation of test samples of inorganic anions and cations respectively, are listed using current state-of-the-art ion chromatographic mobile-phase conditions, columns, and instrumentation. The chromatography is reproducible and peak area and peak height are proportional to analyte concentration; thus, unknown analysis through calibration is accurate, precise, and sensitive. Furthermore, analysis time is remarkably short. No other analytical method offers such a rapid, accurate, and yet relatively simple procedure for the trace analysis of these kinds of analyte ions.

IC was quickly recognized as one of the foremost methodologies for the determination of simple inorganic and organic analyte ions. It quickly became a routine analytical method that was applied to the determination of water quality and analysis of environmental, biological fluid, food, and industrial samples and related products, by-products, and materials. IC can be performed manually at the bench with individual samples, it can be automated, and it can be used in on-line process monitoring. IC measurements in the field are also possible. Many IC procedures are approved by regulatory agencies, for example the United States Environmental Protection Agency (USEPA), and are invaluable in industrial, health, food, electronic, and environmental analytical laboratories. At present IC procedures often serve as the standard to which other newly developed analytical methodology is compared particularly when considering

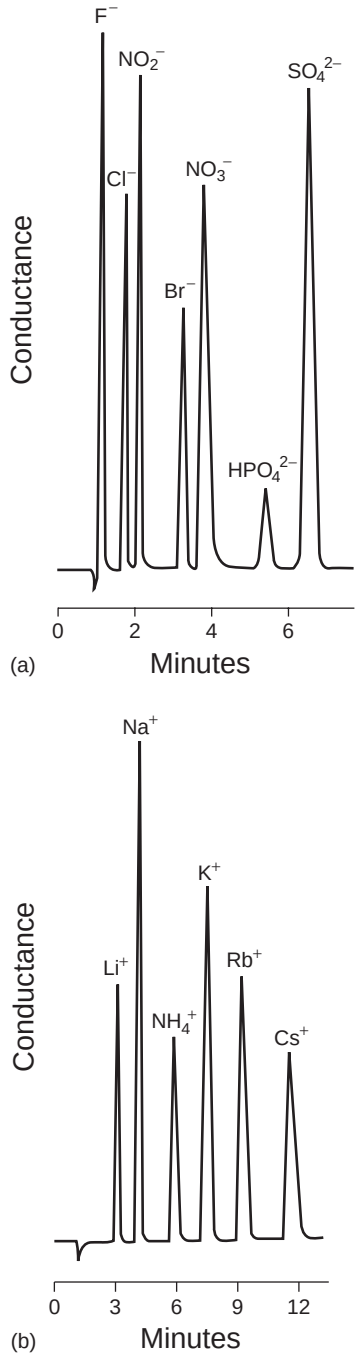

Figure 1 ICs for the separation of anions (a) and cations (b). (a) An aqueous 1.7 mM $NaHCO_3$/1.8 mM Na_2CO_3 mobile phase at 2.0 mL min^{-1} with 50 µL injection of a sample of 3 ppm F^-, 4 ppm Cl^-, 10 ppm NO_2^-, 10 ppm Br^-, 20 ppm NO_3^-, 10 ppm HPO_4^{2-}, and 25 ppm SO_4^{2-}, suppressed conductivity detection, and a latex-based anion exchanger (IonPac AS4A). (b) An aqueous 30 mM HCl mobile phase at 1.0 mL min^{-1} with 50 µL injection of a sample of 5 ppm Li^+, 5 ppm Na^+, 10 ppm NH_4^+, 10 ppm K^+, 20 ppm Rb^+, and 30 ppm Cs^+, suppressed conductivity detection, and a latex-based cation exchanger (Fast Sep Cation 1). (Reproduced by permission of Wiley-VCH from J. Weiss, *Ion Chromatography*, 82 and 174.[2])

the analysis of simple inorganic and organic anions and cations.

The scope of IC as demonstrated by Small et al.[1] has broadened and the modern use of the terminology in chromatography means more than the ion-exchange separation of analyte ions and their detection by conductance following suppression of mobile-phase electrolyte. The following sections, while focusing on suppressed IC and its advances, also briefly introduces other advances in IC, namely improvements in ion-exchange column technology, detection, and ion interaction reagents and advances in nonsuppressed IC, IIC and IEC. A more detailed discussion of the principles and applications of IC, nonsuppressed IC, IIC, and IEC is available in several excellent reference books.[2–5]

Clearly, successful application of IC and related strategies to provide solutions to important separation problems rests on four major factors. For a particular separation problem these are: (1) selection of the correct ion-exchange column based on an understanding of the underlying properties of the exchanger, (2) being aware of the scope of the eluting conditions relative to the analyte mixture and the problem at hand, (3) carrying out a systematic manipulation of the first two factors to establish the required resolution of the analyte ions within the column and to obtain detector compatibility at the desired DL, and (4) handling of the chromatographic data to solve the problem at hand.

2 BASIC PRINCIPLES OF ION EXCHANGE

2.1 Ion-exchange Equilibrium

To understand why analyte ion separations occur in IC it is necessary to understand ion-exchange chromatography or the separation of analyte ions on a column containing an ion exchanger. A typical ion exchanger is an insoluble solid substance containing fixed charge centers (ionogenic groups) and exchangeable counterions of opposite charge associated with the ionogenic groups. These counterions can be exchanged for a stoichiometrically equivalent amount of other ions of the same sign when the ion exchanger is in contact with an electrolyte solution. Cation exchangers exchange cations and anion exchangers exchange anions.

The cation-exchange process, where H^+ and Na^+ are competing for the exchange site by electrostatic attraction, can be represented by

$$R^-H^+_{(s)} + Na^+_{(aq)} = R^-Na^+_{(s)} + H^+_{(aq)} \qquad (1)$$

In Equation (1) R^- is the solid insoluble exchanger containing a negatively charged ionogenic group, (s) is the solid form, and (aq) is the aqueous form. For anion exchange, where OH^- and Cl^- are competing for the anion-exchange site by electrostatic attraction, the process is represented by Equation (2)

$$R^+OH^-_{(s)} + Cl^-_{(aq)} = R^+Cl^-_{(s)} + OH^-_{(aq)} \qquad (2)$$

In this case R^+ is the solid insoluble exchanger containing a positively charged ionogenic group. Many naturally occurring and synthetic materials are capable of exhibiting ion-exchange properties but those that are of the most interest in IC are ion exchangers that are specially synthesized for IC applications.

2.2 Properties of Ion Exchangers

2.2.1 Ion-exchange Selectivity

The equilibrium constants, K, for reactions (Equations 1 and 2), which are given by expressions (Equations 3 and 4), respectively, are selectivity constants because each specifies the relative preference of the ion exchanger for one ion over the other.

$$K_H^{Na} = \frac{(R^-Na^+)_s[H^+]_{aq}}{(R^-H^+)_s[Na^+]_{aq}} \qquad (3)$$

$$K_{OH}^{Cl} = \frac{(R^+Cl^-)_s[OH^-]_{aq}}{(R^+OH^-)_s[Cl^-]_{aq}} \qquad (4)$$

By designating one ion as a standard, a selectivity scale can be established for other ions of similar charge relative to the standard ion. For cation exchange H^+ is the standard while for anion exchange the standard is usually OH^- or Cl^-. Table 1 lists cation- and anion-exchange selectivities for a typical strong acid cation and strong base anion exchanger, respectively, where the exchanger matrix is polymer based. It should be noted that selectivity will be influenced not only by the ion exchanger ionogenic group but also by the exchanger matrix to which the ionogenic

Table 1 Anion and cation ion exchange selectivities[a]

Anions	Selectivity	Cations	Selectivity
OH^-	0.09	Li^+	0.9
F^-	0.09	H^+	1.0
$C_2H_3O_2^-$	0.17	Na^+	1.5
HCO_2^-	0.22	NH_4^+	2.0
$H_2PO_4^-$	0.25	K^+	2.5
HCO_3^-	0.32	Rb^+	2.6
Cl^-	1.0	Cs^+	2.7
NO_2^-	1.2	Mg^{2+}	2.5
Br^-	2.8	Ca^{2+}	3.9
NO_3^-	3.8	Sr^{2+}	5.0
HSO_4^-	4.1	Zn^{2+}	2.7
Phenolate	5.2	Cu^{2+}	2.9
I^-	8.7	Ni^{2+}	3.0

[a] See Gjerde and Fritz pps. 65 and 66.[4]

group is bound. In general, when all other factors are equal ion exchangers exhibit higher binding (increased selectivities) for multivalent ions, for ions with decreased hydrated radius, and for ions with increased polarizability.

2.2.2 *Reversibility*

Equations (1) and (2) are equilibrium reactions that are stoichiometric and reversible. Typically, equilibrium is reached quickly. Consider Equation (1) where Na^+ and H^+ are competing for the exchanger anionic site. If the exchanger is placed in an aqueous environment and H^+ and Na^+ (and a suitable anion) are present in equal molar quantities, equilibrium is reached according to the selectivity given by the equilibrium constant in Equation (3). On the other hand if the cation exchanger in the H^+ form is put in a column and a Na^+ solution (for example aqueous NaCl) is continuously passed through the column, H^+ will be continuously removed from the column and will be replaced by the Na^+ until all the H^+ has been replaced. When all the H^+ is removed the Na^+ concentration entering the column will equal the Na^+ leaving the column. If the Na^+ solution is washed out of the column and an aqueous HCl solution is now passed through the column, the exchanger will be converted back to the H^+ form. Thus, Equation (1) can be driven in either direction by either a batch or column operation. Furthermore, the equilibrium can be driven by a cation that has a lower selectivity than the other by using the cation at a higher concentration or by continued passage of the cation solution through the column. This is possible because of mass action.

Binding is quantitative and stoichiometric. Selectivity also determines the order of ion removal from the exchanger and corresponds to the ion elution order if several different analyte ions are present. Since the exchange process usually occurs rapidly, column experiments can be carried out efficiently and reproducibly.

2.2.3 *Rate of Ion Exchange*

Equilibrium is reached rapidly because rates of exchange are usually high. Exchange rates are controlled by diffusion within the exchanger particle, if ionogenic groups within the particle are accessible, and by film diffusion at the exchanger particle surface. Physical properties of the exchanger (see section 2.2.6 in the following) and the type of ionogenic group will influence exchange rates. While ion exchangers are available with capacities as high as 3–5 mM g^{-1}, most ion exchangers used in IC have capacities at or below 200 μM g^{-1}.

2.2.4 *Ion-exchange Capacity*

Ion-exchange capacity is defined by the quantity of exchangeable ions per unit amount of ion exchanger. When an ion exchanger is used in a column to bring about analyte separation due to differences in analyte exchange selectivity, only a small portion of the available exchange capacity should be employed to prevent a column analyte mass overload. Exchange capacity is also sensitive to ion-exchanger physical properties (see section 2.2.6 in the following) because these properties influence the accessibility of the exchange sites.

2.2.5 *Ionogenic Group*

Ionogenic groups or ion-exchange sites are classified as being either strong or weak acid or base groups. Strong acid and base groups are highly dissociated and the exchangers containing these groups are like insoluble strong electrolytes. These groups are $-SO_3^-H^+$ and $-NR_3^+OH^-$ and exchange cations and anions, respectively. Common weak acid ionogenic groups are $-CO_2H$, $-PO_3H_2$, aromatic $-OH$, and aromatic $-SH$ while weak base ionogenic groups are $-NH_2$, $-NHR$, and $-NR_2$. These are like insoluble weak electrolytes and their ability to participate in ion exchange as cation and anion exchangers, respectively, depends on dissociation of the group and therefore on their ionization constant and the pH of the environment. The ionogenic groups can be chemically bound to an organic polymer matrix or chemically bound to silica or other oxides.

2.2.6 *Physical Properties*

Ion exchangers developed for IC applications must satisfy the physical properties demanded by modern liquid chromatography (LC) in addition to providing the desired ion-exchange properties. The exchangers must be available as small, physically strong, uniform size, spherical microparticles that can be reproducibly packed into an efficient column.

Ion exchangers based on an organic polymeric matrix will expand and contract when placed in water or a polar solvent. A highly polar medium favors swelling while a less polar environment favors contraction. Both the type of solvent and the presence of an electrolyte and its concentration influence swelling and contraction. Increased cross-linking in the organic polymer matrix increases physical strength, reduces swelling and solubility, and affects the accessibility of the exchanger sites within the organic polymer and therefore the available exchange capacity. Bead fracture can occur because of rapid swelling or contraction of the ion exchanger. Higher cross-linking will reduce bead fracture and increase particle stability. Rapid swelling and contraction can cause discontinuity in the column packing since these processes will occur in the column when solvent and/or electrolyte concentration of the mobile phase are changed abruptly.

Cross-linking, swelling/contraction, and related properties are more specific to organic polymeric ion exchangers and are generally of less concern with pellicular, latex, or silica-based ion exchangers.

The polymerization strategy used to prepare the polymeric-based ion exchanger will also influence its properties. Polymeric ion exchangers can be prepared that are either gel-like or macroporous. The former have low surface areas and are not truly porous. Their porous-like property is obtained only when the polymer is swollen. In contrast the macroporous polymer has a high surface area and has high porosity. This is because it is a spherical particle that is made up of many ultramicrogel particles joined together. Its porosity is therefore not dependent on swelling of the bead but more so on the space that is present between the attached microgel beads.

The organic polymeric, pellicular, latex, and silica-based ion exchangers are commercially available as spherical particles of uniform microparticle size. Bulk lots of the exchangers can be purchased to pack columns in the laboratory. However, commercially packed columns are used most often because they are available commercially and provide excellent column efficiency and reproducibility.

2.2.7 Stability

A useful ion exchanger in IC must withstand the usual rigors of modern LC and mobile phases routinely used in IC. Bead fracture, swelling/contraction, and reactivity towards eluents can cause serious limitations in IC applications. Organic polymeric, pellicular, and latex-based ion exchangers are stable throughout the pH range of 1 to 14. Silica-based ion exchangers, in contrast, are stable within the range of pH 2 to 9. The ionogenic groups are generally stable; however, the $-R_3N^+X^-$ ionogenic group is the least stable particularly when X^- is OH^- and when the exchanger is dry.

2.2.8 Ion-exchanger Matrix

Ion exchangers for IC are organic polymeric, latex, pellicular, or silica-based and the nature of the matrix can have a strong influence on the chromatographic result. Syntheses of these four types of exchangers are briefly described in the next section.

For an ion exchanger with an organic polymeric matrix its composition, cross-linking, and porous or gel-like properties influence the accessibility of the ionogenic groups, the rate of exchange, and compatibility with mixed and/or nonaqueous solvents. The matrix can also interact with nonionic materials, which may be present in the analyte sample, and over time can alter the column performance. Organic analyte ions that are lipophilic may also interact with the matrix through hydrophobic interactions in addition to ion exchange. Pellicular ion exchangers with a polymeric core can exhibit similar properties but to a much lesser degree.

Analyte hydrophobic interactions are not encountered with pellicular ion exchangers with a silica core or silica bonded-phase ion exchangers. However, these exchangers have a limited useful pH range and can have residual silanol (−SiOH) sites. These sites are weakly acidic and will exhibit cation-exchange properties. The latex ion exchangers typically do not exhibit hydrophobic matrix-retention interactions but they can exhibit both anion and cation exchange depending on their synthesis. A more detailed discussion of the properties of ion exchangers is available in several excellent reference books.[6,7]

3 ION EXCHANGERS IN ION CHROMATOGRAPHY

The major types of anion and cation exchangers used in IC are organic polymeric-based, latex-based, and silica bonded-phase exchangers. These exchangers for IC applications are low in exchange capacity, typically less than 200 $\mu eq\ g^{-1}$, and most often have ionogenic sites on the outer particle surface rather than on the interior of the particle.

3.1 Polymeric Ion Exchangers

Most organic polymeric-type ion exchangers are derived from polystyrene–divinylbenzene (PSDVB) copolymers. Other polymer types used are polyvinyl and polymethacrylate polymers.

PSDVB is prepared by reaction of styrene **(1)** and divinylbenzene **(2)** using the appropriate solvent, catalyst, and reaction conditions according to Scheme 1.

The amount of divinylbenzene used in the synthesis determines the amount of cross-linking in the copolymer and influences the exchanger properties as outlined earlier. The anion exchanger, often referred to as a strong base anion exchanger, is prepared by chemically bonding a quaternary ammonium group to the PSDVB copolymer by chloromethylation of the copolymer followed by reaction with a tertiary amine. This is shown in Scheme 2, where the PSDVB copolymer is represented by the benzene ring. While the R groups can be varied, they typically are methyl groups and the anion exchanger is frequently supplied in the Cl^- form rather than in the OH^- form because the latter is less stable particularly when dry.

The cation exchanger, often referred to as a strong acid cation exchanger, is prepared by a sulfonation procedure.

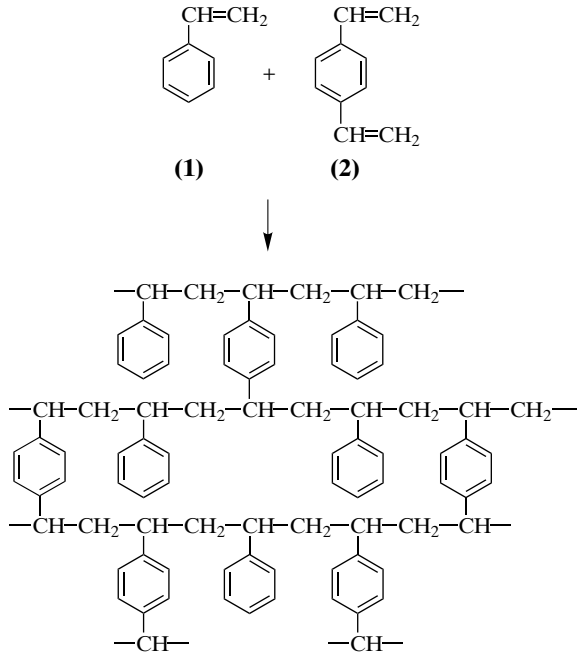

Cross-linked styrene-divinylbenzene copolymer

Scheme 1

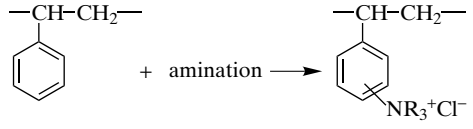

Scheme 2

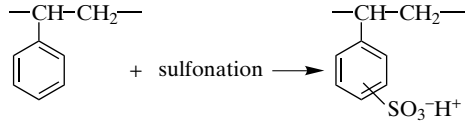

Scheme 3

In this case a $-SO_3^-H^+$ ionogenic group is bound to the PSDVB copolymer as shown in Scheme 3,
where the benzene ring represents the PSDVB copolymer. Usually, commercially available columns packed with the strong acid cation exchanger are supplied in the H^+ or Na^+ form. Both are equally stable forms.

For both the PSDVB strong acid and base cation and anion exchangers for IC the ionogenic groups are primarily located on the surface of the spherical copolymer beads. For this reason ion-exchange capacity is low but mass transfer for exchange is more favorable. High-capacity ion exchangers, where exchange groups are located on the surface and within the bead, are commercially available. These are usually not used in IC but are useful for IEC. The surface-covered exchangers are preferred because of better exchange kinetics and a more rapid attainment of equilibrium. Other kinds of ionogenic groups exhibiting ion-exchange properties have been prepared but are not necessarily available with an ion-exchange capacity suitable for IC. Several of these are listed in Table 2. It is important to note that all of these have weak acid or base properties and dissociation of the groups, and therefore the ion-exchange properties are pH dependent. The strong acid and base ion exchangers in contrast are highly dissociated like strong insoluble, electrolytes and will exhibit ion-exchange properties throughout the pH range.

Table 2 Common ionogenic groups

Strong acid	Weak acid
$-SO_3^-H^+$[a]	$-CO_2H$,[a] $-PO_3H_2$,[a] phenolic-OH
Strong base	Weak base[a]
$-NR_3^+OH^-$[a]	$-NH_2$, $-NRH$, $-NR_2$

[a] Commercially available for IC.

Ion exchangers that contain the PSDVB matrix will swell in polar solvents and in the presence of electrolyte. Sharp changes in their concentrations can cause rapid and excessive swelling or contracting and can lead to bead fracture. The PSDVB-type exchangers that are of the gel type owing to the polymerization procedure, will typically contain 2–16% divinylbenzene as the cross-linking agent to inhibit swelling/contracting. This gel-type PSDVB, however, is not truly porous-like as a sponge and its porous-like property results from expansion (or contraction) between the copolymer chains. The gel-type ion exchanger, thus, will have an apparent porosity and surface area which are typically very low even when swollen, that is, dependent on the extent of the bead swelling.

A macroporous (also called porous, isoporous, or macroreticular) PSDVB copolymeric-based ion exchanger can be made via modifications in the polymerization strategy. The copolymer for the ion exchanger is still derived from styrene and divinylbenzene, (see Scheme 1) is highly cross-linked ($\gg$16%), and yields a network of submicrosphere gel beads joined together to form a larger, spherical, single bead. The high porosity and surface area, which are much less dependent on bead swelling/contraction, are more permanent because the interstitial spaces between the microgel beads are largely responsible for the porosity. Pore diameters and surface area can exceed 10 nm and 100 $m^2\,g^{-1}$, respectively. The gel-type-derived ion exchanger, in contrast, has only a fraction of this porosity and surface area depending on swelling/contraction. The macroporous-based ion exchangers are physically strong and have favorable ion-exchange rates and mass transfer provided exchange capacity is on the surface particularly when mixed solvents are used. Macroporous ion exchangers with high ion-exchange capacities can be made but the macroporous ion exchangers usually used in IC are low capacity.

3.2 Latex-based Ion Exchangers

Latex-based anion and cation exchangers for IC applications are commercially available. These are pellicular in nature, that is, the ionogenic groups are primarily located on the outer surface of the particle.[2] For the anion exchanger a surface-sulfonated PSDVB particle is the matrix and this is covered by aminated latex microparticles (0.1 μm) which are held to the core primarily by electrostatic forces and to a lesser extent by van der Waals forces. This general structure of the latex anion exchanger is shown in Figure 2(a). The quaternary ammonium groups are in large excess compared to the sulfonic acid groups and the net effect is that anion exchange occurs at the bead surface.

The cation latex-based ion exchangers can be prepared in a similar fashion except that the PSDVB copolymeric surface is first aminated and this is subsequently covered with sulfonated latex microparticles. Since the sulfonated groups are in excess, it functions as a cation exchanger. This type of latex-based cation exchanger, however, suffers from poor reproducibility and chemical stability. A more reliable latex-based cation exchanger is prepared by taking a sulfonated PSDVB particle, coating it with aminated latex microparticles as an inner layer and coating this layer with a layer of sulfonated latex microparticles. A typical structure for the cation latex-based exchanger is shown in Figure 2(b). In this case the sulfonated groups on the outer layer due to the sulfonated latex microparticles are in excess and cation exchange occurs at this sulfonated surface.

Properties of the two exchangers can be altered by changes in the matrix and the ratio of the aminated or the sulfonated latex to the pellicular sulfonated or aminated groups. In general, cation and anion retention order is the same for both the PSDVB and latex-based strong-acid and base cation and anion exchangers. However, selectivity differences for the latex-based exchangers can be modified, but usually not reversed. The cross-linking in the PSDVB core can be changed, the particle size of the latex particles can be changed, and the number of charged sites on the core versus the number in the latex microparticles can be altered. Consequently, resolution of anions or cations in IC can be improved by modification of one or more of these properties. A different particle core material can also be used. For example, acrylate-based latex anion exchangers have been prepared and are commercially available. These are often preferred for the separation of anions of high anion-exchange selectivity, such as phosphate and sulfate, because their retention on this anion exchanger is much less. Because of these modifications several different latex anion and cation exchangers have been developed for the separation of specific kinds of mixtures and are commercially available from Dionex Corporation. Furthermore, because both aminated and sulfonated groups are present in the latex-based exchangers, these exchangers are capable of exhibiting both anion- and cation-exchange properties simultaneously even though one group is in excess over the other.[9]

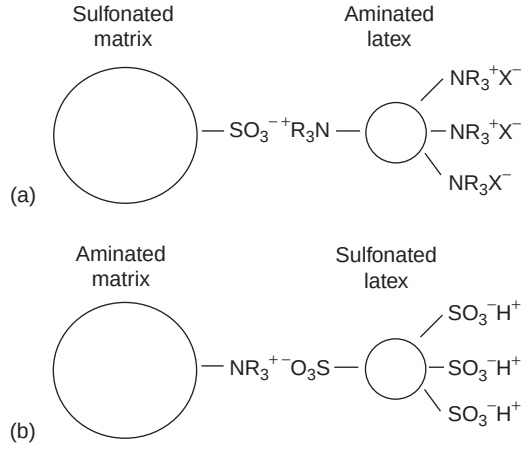

Figure 2 Latex-based ion exchangers, (a) anion exchanger, (b) cation exchanger.

3.3 Silica-based Ion Exchangers

Introduction of the bonded stationary phase, where the stationary phase is chemically bonded to silica (also now to alumina and zirconia) to HPLC was a milestone event in the development of HPLC. While the major thrust in the initial studies with bonded phases focused on the development of reversed stationary phases and to a lesser extent normal stationary phases, research also emerged on the synthesis of bonded-phase ion exchangers. Those of interest in IC are low in exchange capacity and are usually prepared by silanization chemistry.

Typically, silica-based ion exchangers are prepared by silanization of the silica silanol (Si−OH) sites with chlorosilane reagents that contain the ionogenic group, or groups that will undergo additional reactions to introduce the ionogenic groups. Strong acid ($-SO_3^-H^+$) and strong base ($-NR_3^+X^-$) silica-based cation and anion exchangers are commercially available as are weak acid ($-CO_2H$) and base ($-NH_2$) ion exchangers. The commercially available silica-based bonded-phase ion exchangers typically have high surface areas ($>200\,m^2\,g^{-1}$), are physically strong, do not swell/contract as solvent or electrolyte is changed, and provide high column efficiencies. However, they have a limited pH range of about pH 2 to 9 because of reactivity of the silica and the bonded group towards acid and base. They are commercially available as packed columns of uniform spherical microparticles of 3, 5, or 10 μm. Exchange capacities can be varied in the synthesis and in general those that are

commercially available and used in IC are low in capacity although their capacity is greater than the typical IC polymer and latex-based ion exchangers.

A pellicular-type silica-based ion exchanger is formed by depositing the ion exchanger matrix as a thin layer onto the surface of an inert core particle which is spherical, microparticle, and uniform in size. The ionogenic group is then attached to the matrix. Strong acid and base and weak acid and base ion exchangers have been made this way and several are commercially available. Typical core materials are a glass bead or a nonporous silica bead while the matrix layer is typically PSDVB or another polymer that provides reaction centers for the bonding of the ionogenic groups. One important weak acid pellicular-type cation exchanger with a silica core is an exchanger that has a poly(butadiene–maleic acid) copolymer covering and weak acid carboxy ($-CO_2H$) ionogenic groups. This type of cation exchanger is particularly useful for the simultaneous separation of alkali metals and alkaline earths (see Figure 3) with short analysis times and with a single eluent. This kind of separation is not easily accomplished with $-SO_3^-H^+$ polymer or latex-based cation exchangers.

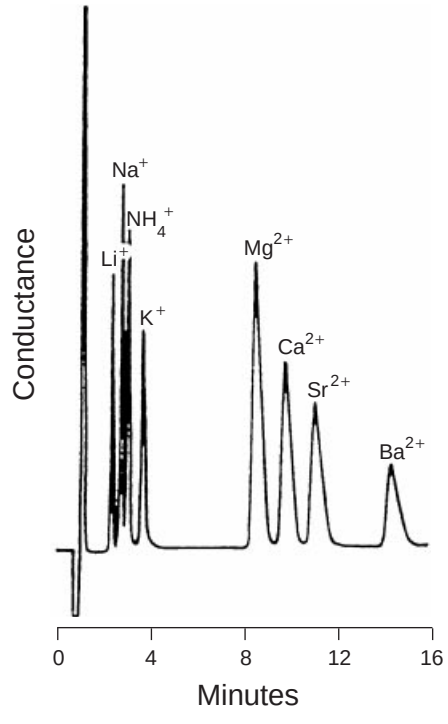

Figure 3 Separation of analyte cations on a butadiene–maleic acid polymer-based cation exchanger. An aqueous 5.0 mM tartaric acid mobile phase at 1.0 mL min^{-1} with 10 µL injection of a sample 1 ppm Li^+, 5 ppm Na^+, 5 ppm NH_4^+, 10 ppm K^+, 10 ppm Mg^{2+}, 10 ppm Ca^{2+}, 20 ppm Sr^{2+}, 20 ppm Ba^{2+}, and direct conductivity detection. (Reproduced by permission of Wiley-VCH from J. Weiss, *Ion Chromatography*, 176.[2])

3.4 Other Exchangers Used in Ion Chromatography

Other stationary phases used in IC are chelating ion exchangers and certain inorganic oxides and salts. Most chelating exchangers will contain acidic or basic groups that can participate in ion exchange in addition to the chelating group. The major interaction, however, is coordination between the analyte, usually a cation analyte, and the coordination center of the chelating exchanger. Chelating exchangers can be highly selective and this property can be predicted from known stability constants for complex formation between the analyte cation and the ligand in solution. Chelating ion exchangers are used more often for isolating and concentrating trace levels of cations rather than for the resolution of complex mixtures of cations. The high selectivity and stability of complex formation favors the former while slow exchange kinetics is a major handicap to the latter application. Few chelating ion exchangers for IC are commercially available.

Inorganic oxides and salts can participate in cation and/or anion exchange because they also exhibit weak acid and base properties. The ion-exchange property and ion-exchange capacity are therefore pH dependent. The exchange behavior of oxides also depends on many other environmental factors including buffer and inert electrolyte components and concentration in the mobile phase, pretreatment, heat treatment, ageing, method of preparation, crystallinity, and surface area. Furthermore, few are available as uniform spherical microparticles that meet high column-efficiency requirements. Notable exceptions are alumina, silica, and zirconia. Their uniqueness in IC applications comes from the fact that inorganic analyte anion and cation ion-exchange selectivities, and consequently analyte elution order, on the oxides and salts are often different to those obtained with the $-SO_3^-H^+$ and $-NR_3^+X^-$ cation and anion exchangers, respectively. Unfortunately, the stability of the inorganic oxides and salts as exchangers in an aqueous environment is not the same as the stability and reproducibility offered by the polymeric and latex-based ion exchangers.

4 ION EXCHANGE SEPARATIONS

Ion exchange is the main separation strategy that is used to separate ionic analytes. The chromatographic system consists of a separation or analytical column packed with an anion exchanger to separate anions or a cation exchanger to separate cations. The mobile phase will contain an electrolyte that provides a competing ion (anion for anion exchange and cation for cation exchange) of suitable ion-exchange selectivity (see Table 1) and concentration. When the mixture of analyte ions is introduced and carried into the column by

the mobile phase, the analyte ions are electrostatically attracted to the ionogenic groups. They are subsequently released to the mobile phase by competitive electrostatic attraction between the competing ions provided by the mobile-phase electrolyte and the analyte ions. Separation occurs because the analytes migrate through the column at different rates because of differences in their ion-exchange selectivities. The separation is completed by detection of the analyte ions as they emerge from the column. Absorption, fluorescence, and electrochemical detection are common detection modes in HPLC, but these detection strategies have limitations in direct detection of the ion-exchange separation of simple inorganic and organic anions and cations because most of these analytes are detector inactive. Conductivity, however, should be a universal detector for ionic analytes as most would have high ionic equivalent conductances. Unfortunately, conductivity detection in ordinary ion-exchange separations is not very sensitive because the analyte ion conductivity signal-to-noise ratio (S/N) is very small. In a typical ion-exchange separation the mobile phase itself is highly conducting because of the electrolyte required in the mobile phase.

Consider the case where an analyte mixture of Cl^-, Br^-, and I^- are being separated on an anion-exchange separator column capable of delivering high column efficiency. An aqueous NaOH mobile phase of suitable concentration is used and detection is by conductivity. The chromatographic set-up is shown in Figure 4. The sample is introduced as a small aqueous aliquot of sodium salts and the NaOH mobile phase carries the sample into the anion-exchange column. Anion exchange occurs according to the exchange equilibrium listed in Equation (5)

$$\mathbf{S}\text{–}NR_3^+OH^- + X^- = \mathbf{S}\text{–}NR_3^+X^- + OH^- \qquad (5)$$

where **S** is the exchanger matrix and X^- is Cl^-, Br^-, and I^-. Because of the anion ion-exchange selectivities (see Table 1) and the electrostatic competition between OH^- and the analyte anions towards the anion-exchange site, Cl^- travels the fastest and I^- the slowest and resolution starts to occur in the column as shown in Figure 4. When the analytes as the Na^+ salts emerge from the column as separated chromatographic zones they do so in the presence of the NaOH mobile phase. They subsequently pass into the conductivity detector where conductance of the flowing stream is determined. Because the conductivity for the NaOH is very high, the conductance signal caused by the presence of the three analytes is very small, S/N for the analyte anions is not favorable, and detection is poor. This is illustrated in the chromatogram in Figure 4, where the peaks correspond to the Na^+ salts of the analyte anions. If the NaOH was absent in the background, conductivity for the background would be very low and the three analyte signals would be very high. Thus, S/N would be very favorable and detection would be very sensitive. This major problem was solved by the suppressed IC strategy developed by Small et al.[1]

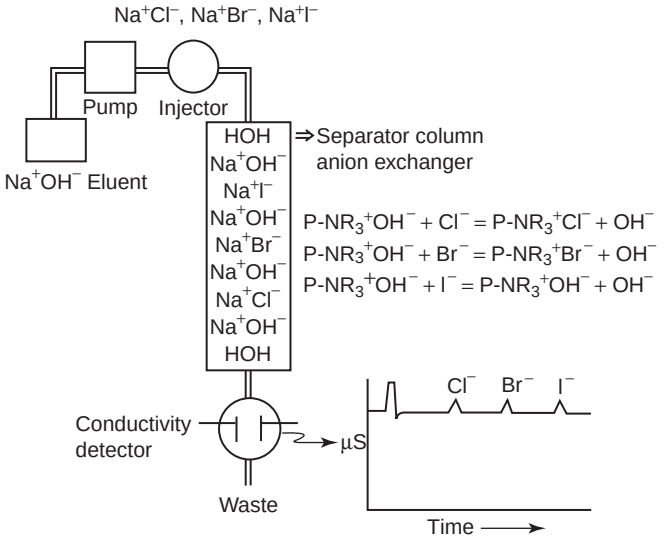

Figure 4 Separation of halides on an anion exchanger without postcolumn suppression.

5 SUPPRESSED ION CHROMATOGRAPHY

5.1 Suppression by an Ion Exchanger

In the suppressed IC strategy developed by Small et al.[1] a mixture of analyte anions or cations are separated by high-performance ion exchange and the analytes are detected by conductivity. In their strategy the key to a favorable S/N and sensitive detection is the removal or reduction of the mobile-phase background conductance due to the mobile-phase electrolyte. To separate analyte anions a low-capacity strong base anion-exchange column (separator or analytical column) is used for the separation and a high-capacity strong acid cation-exchange column connected in series serves as the suppressor column. For the separation of analyte cations the system is reversed or the analytical column is a strong acid low-capacity cation exchanger and the suppressor column is a high-capacity strong base anion exchanger. The basic instrumental set-up for anion and cation suppressed IC is shown in Figure 5(a) and (b), respectively.

Consider the nonsuppressed Cl^-, Br^-, and I^- anion-exchange separation illustrated in Figure 4. When this separation is carried out by an IC suppression strategy the effluent from the separator column is passed first

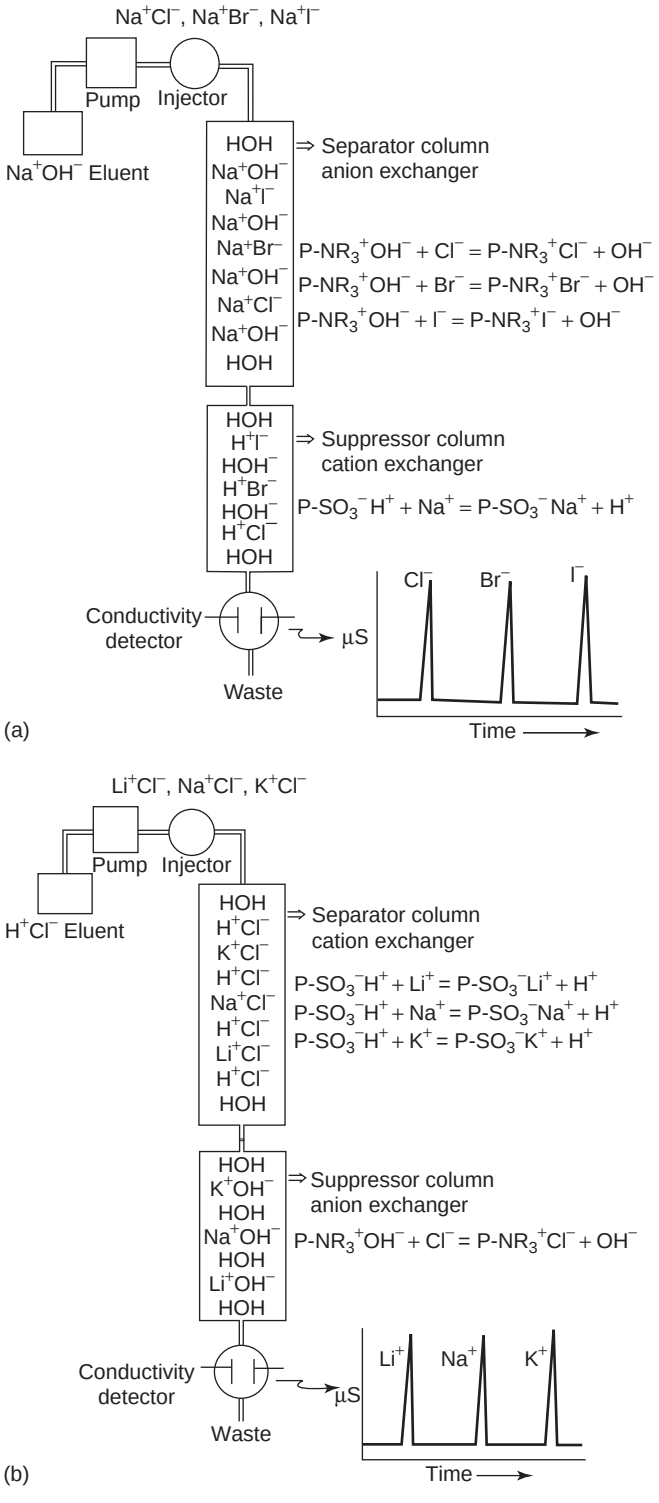

Figure 5 Separation of anions (a) and cations (b) with post-column suppression. P stands for a polymeric matrix of the exchanger.

into the suppression column of a strong acid cation exchanger charged in the H^+ form as shown in Figure 5(a) and then through the conductivity detector. In the suppression column NaOH, NaCl, NaBr, and NaI are converted to HOH, HCl, HBr, and HI, respectively, according to the exchange equilibrium summarized by Equation (6).

$$\mathbf{S}\text{–}SO_3^-H^+ + Na^+ = \mathbf{S}\text{–}SO_3^-Na^+ + H^+ \qquad (6)$$

Thus, when the effluent from the suppressor column passes through the conductivity detector the background is water and the analytes are the corresponding acids. Conductivity detection is therefore very sensitive because of the absence of the background conductance, S/N is high, and DLs are significantly reduced compared to the absence of the suppression column. The conductivity signals are also enhanced because the analytes are present as the acids rather than the Na^+ salts in the detection step. The ionic equivalent conductance for H^+ is about 7 times that of Na^+ which enhances the conductivity signal. This internal signal amplification is another advantage of the suppression system.

For the separation of analyte cations, for example the separation of Li^+, Na^+, and K^+ as Cl^- salts (see Figure 5b), a strong acid, low-capacity cation exchanger is used for the separator column and a strong base, high-capacity anion exchanger in the OH^- form is used for the suppressor column. An HCl mobile phase is used for the separation of the cation analytes. The separation occurs because of the differences in cation selectivities for the three analyte cations (see Table 1) and competition between the cations and H^+ for the cation-exchange sites. The exchange equilibria are summarized by Equation (7)

$$\mathbf{S}\text{–}SO_3^-H^+ + M^+ = \mathbf{S}\text{–}SO_3^-M^+ + H^+ \qquad (7)$$

where M^+ is Li^+, Na^+, and K^+. If the effluent from the separator column is passed directly into the conductivity detector in Figure 5(b), the detector would detect LiCl, NaCl, and KCl in that order on an HCl background. Conductance of the background is very high and conductance change due to the analytes would be small. However, if the effluent is passed into the suppression ion-exchange column, which is a high-capacity strong base anion exchanger charged in the OH^- form, the HCl background electrolyte and the alkali metal chloride analytes are converted to HOH and LiOH, NaOH, and KOH, respectively, as summarized by Equation (8)

$$\mathbf{S}\text{–}NR_3^+OH^- + Cl^- = \mathbf{S}\text{–}NR_3^+Cl^- + OH^- \qquad (8)$$

Thus, detection of each analyte by the conductivity detector is very sensitive because detection of the analyte takes place in an electrolyte-free background. Furthermore, the signal is enhanced because the analytes are detected as the OH^-, which has a much higher ionic equivalent conductance compared to Cl^-.

This separation/detection strategy was quickly recognized as a superior method for the rapid, accurate, and sensitive separation and determination of simple inorganic and organic ionic analytes. Analysis times and DLs obtained by this strategy were not possible by other methods.

The mobile-phase electrolytes used in Figure 5(a) and (b) are strong acid and base electrolytes, respectively, and when suppression is employed background conductance due to the electrolyte is overcome. Both eluents are modestly weak in eluent strength, which can be increased only by an increase in the concentration electrolyte. If a stronger eluent is required, which would be needed for example to separate highly retained analyte anions, an electrolyte that provides an anion of higher selectivity must be used (see Table 1). This is possible with suppression IC by using salts of weak acids and bases for the mobile-phase electrolyte in anion and cation separations, respectively. In the suppression step these electrolytes are converted to weak acids and bases, respectively. Because the weak acids and bases are poorly dissociated (K_a and K_b are therefore an important variable) their contribution to background conductance is low, S/N is favorable, and conductivity detection is still sensitive.

The suppression strategy outlined in Figure 5(a) and (b) had several limitations. DLs, while being favorable, were still limited by the analyte peak broadening that occurs in the separator column and even more so in the suppression column. Reduction of the broadening and an increased resolution was subsequently addressed by the development of new ion exchangers specifically designed for the separator column for IC and new, more versatile mobile-phase/electrolyte mixtures to separate the analyte anions and cations. A second recognized problem was that the suppression column's exchange capacity is exhausted in long-term operation, even though it has a high capacity. Therefore, continuous operation of this suppression IC system is not possible. The IC system would have to be shutdown either to replace or to regenerate the suppression column.

5.2 Membrane Suppression in Ion Chromatography

Development of a membrane suppressor[2,8] was subsequently shown to overcome these two major problems of the high-capacity ion-exchange suppressor column, namely, peak broadening in the suppressor and inability to carry out continuous IC owing to the required suppressor regeneration. A schematic representation of an anion membrane suppressor that would be used to suppress background electrolyte in the separation of anions is shown in Figure 6(a). The central region where the eluent flows is microvolume in size and is sandwiched by two sulfonated cation-exchange membranes. A regeneration chamber is on opposite sides of the membranes and the entire device is encased in a small leakproof unit.

Consider the IC separation shown in Figure 5(a) except that the ion-exchange suppression column is replaced by an anion membrane suppressor. As the column effluent from the anion-exchange separator column passes through the microvolume eluent chamber of the membrane suppressor, Na^+ is replaced by H^+ at the membrane. Simultaneously, a H_2SO_4 solution is continuously passed through the regeneration chamber and the Na^+ is then replaced by H^+. Thus, the suppression is continuous, band broadening is minimal, and conductivity DLs are significantly lower compared with those of the ion-exchange column suppression which was used in Figure 5(a). A 1–10 ppm DL (lower is possible) is often routine depending on the injection volume, analyte anion, and the separation conditions.

A cation membrane suppressor is also available and its schematic representation is shown in Figure 6(b). It functions much like the anion membrane suppressor except that the membrane is a strong base cation-exchange membrane charged in the OH^- form and the regeneration solution is usually an aqueous $R_4N^+OH^-$ solution.

The membrane suppressor has low internal volume and can be used continuously to reduce background conductance to almost negligible levels depending on the mobile-phase electrolyte and its concentration. It is efficient, rapid, reproducible, solvent resistant, and is compatible with gradient elution. Maintenance is low and only occasional cleaning is needed. The membrane suppressor is also readily adapted for on-line operation for automated IC in industrial applications.

The commercially available membrane suppressor has also undergone stages of improvement. The key improvements are the following. First the internal volume of the suppressor is now markedly reduced without sacrificing suppression properties and is currently referred to as a micromembrane suppressor (MMS). The MMS provides better DLs, better mass transfer across the membrane, and is better suited to smaller diameter analytical ion-exchange columns. A second key advance is the replacement of the suppressor membrane regeneration process by a continuous electrolytic regeneration process.[2,10] This modification is illustrated in Figure 7. In this case electrolysis of water is used to generate H^+ at the anode which is used to regenerate the anion MMS continuously, while for the cation MMS, OH^- produced at the cathode is used for the regeneration. The extra pump, the delivery system, and the H_2SO_4 solution are no longer required. Regeneration is simplified, continuous, more efficient (which also allows reduction in the internal volume of the suppressor), and is applicable to a wide range of mobile-phase electrolytes in both isocratic and gradient modes.

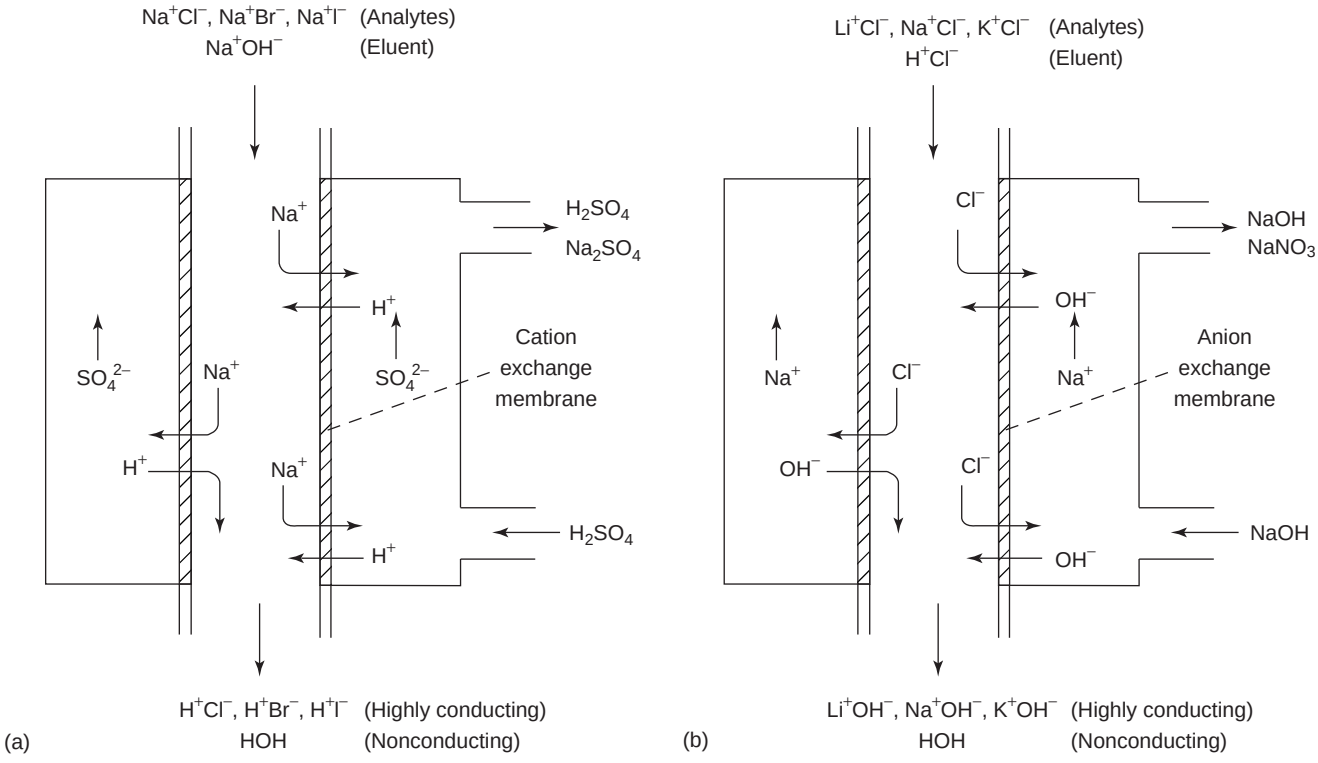

Figure 6 Membrane suppressors for anion (a) and cation (b) separations. (Reproduced by permission of CRC Press from H.F. Walton and R.D. Rocklin, *Ion Exchange in Analytical Chemistry*, 1990.)

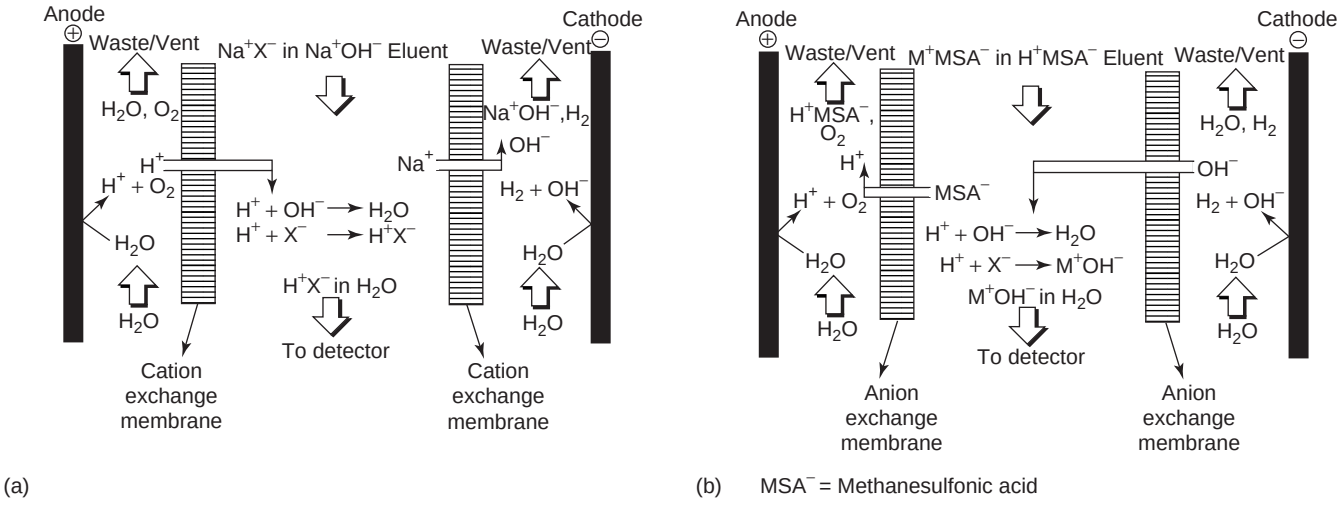

Figure 7 Electrolytic regeneration of a membrane suppressor. (Reproduced by permission of International Scientific Communications, from A. Henshall, S. Rabin, J. Statler, J. Stillian, *Am. Lab.*, **24**(Nov), 20R (1992).[10])

5.3 Other Suppression Strategies in Ion Chromatography

Other kinds of postcolumn suppression strategies have been described in the literature. However, none of these offer the sensitivity, simplicity, versatility in applications, and excellent reproducibility offered by a MMS. Nonmembrane suppression procedures include precipitation of the mobile-phase electrolyte but not the analyte in a suppressor ion-exchange column[11] or chelation[12] of the electrolyte in the suppression step. Another suppression procedure is to pump a slurry of small ion-exchange particles into the analytical column effluent following the separation and prior to conductivity

detection.[13] For example, if anions are being separated the separator column is an anion exchanger and colloidal (about 7 μm) strong acid cation-exchange particles in the H^+ form are continuously combined with the column effluent. Reactions describing the separation and the conversion of the analyte and the mobile-phase electrolyte to the H^+ form in the suppression step are the same (see Equations 5 and 6). Thus, the mobile-phase stream is converted into a low-conducting background while enhancing the analyte anion signal. The presence of the colloidal ion exchanger does not interfere in the conductivity detection.

Postcolumn suppression by an ion-exchange column can also be done effectively and efficiently. This configuration differs from the earlier ion-exchange suppressor unit described by Small et al.[1] in the following way. This system,[14] which is commercially available from Alltech Associates, Inc., includes two ion-exchange columns as the suppressor with either one of the two in operation at any given time. If anions are being separated the separator column is still an anion exchanger and the two suppression columns are strong acid cation-exchanger columns of high exchange capacity, small in size (14 mm × 7.5 mm), and low in internal volume. For cation-exchange separations the separator column is a cation exchanger and the two suppression columns are anion exchangers. The reactions describing the anion and cation separations and suppression are the same as described earlier. The system is continuous because when one suppressor column is being used the other is being regenerated and equilibrated simultaneously. When the first suppression column is exhausted, which is indicated by a color change of an indicator incorporated into the exchanger, a valve switches the separator column effluent to the second suppressor column and the first is regenerated and equilibrated while the second is used for suppression. The regeneration in this two-column suppressor strategy is also done by continuous electrolysis of water. Continuous separations are therefore possible. Also, the suppressor need not be large so that band broadening and reduction in S/N are avoided and DLs are low.

MMSs have applications other than mobile-phase electrolyte suppression in IC. For example, the suppressor can be applied to sample cleanup prior to IC analysis.[15] Another practical application is to use the suppressor to remove NaOH (up to 48%) from a concentrated NaOH solution prior to trace IC anion analysis in the NaOH solution.[16] The MMS can also be used to enhance detection of organic analytes that are separated on reversed-stationary-phase columns when employing a mobile phase of high ionic strength. When these analytes cannot be detected by absorbance or fluorescence, the mobile-phase ionic strength electrolyte can be converted to a weakly conducting species with the MMS. Thus, the analytes can then be detected by conductivity with good sensitivity. For example, anionic surfactants are separated on reversed stationary phases with good resolution only if the mobile phase is at a high ionic strength. The mobile-phase ionic strength electrolyte, which is a salt of a weak acid, is suppressed by an MMS which allows a sensitive conductometric detection of the anionic surfactant.[17]

6 ELUENTS FOR SUPPRESSED ION CHROMATOGRAPHY

Eluent strength is determined by the mobile-phase electrolyte concentration and the electrolyte anion for the IC separation of anions and electrolyte cation for the IC separation of cations. A higher concentration or an ion of higher ion-exchange selectivity yields a stronger eluent.

Table 3 lists several common electrolytes that are used in the suppressed IC separation of anions. Eluent strength, in general, will increase from top to bottom although differences may be small depending on the electrolyte mobile-phase concentration. The key factors to optimize are: the mobile-phase electrolyte anion, the electrolyte concentration, and selection of the anion exchanger. If an electrolyte containing a weak acid anion is used, S/N, and hence DL, will depend on dissociation of the weak acid that is formed in the suppression step.

The useful electrolyte cations for suppressed analyte cation IC separations are few in number. If the suppression step involves OH^-, hydroxide precipitation depending on the electrolyte cation can interfere in suppressor and conductivity-detector performance. Many cation analytes, particularly multivalent cations, have high, similar cation-exchange selectivities and require a strong eluting condition. Because of the similarity in selectivities resolution is often poor even with a strong eluent electrolyte. This problem can be overcome in some cases by using a carboxylate-type cation exchanger or

Table 3 Common electrolytes used in mobile phases in suppressed IC

Separation of anions	Separation of cations
$Na_2B_4O_7$	HCl
NaOH	HNO_3
$NaHCO_3$	H_2SO_4
Na_2CO_3	Isonicotinic acid
$NaHCO_3/Na_2CO_3$	Picolinic acid
Sodium phenolate	DAP
Sodium *p*-cyanophenolate	HCl/DAP
Disodium adipate	
Disodium succinate	
Tyrosine	

DAP, 2,3-Diaminopropionic acid.

more often by including a complexing agent in the mobile phase. Thus, formation constants for the complexes that form between the complexing agent and the multivalent cations will determine the cation elution order. Typical complexing agents used in the mobile phase in addition to an electrolyte are ethylenediaminetetraacetic acid (EDTA) and citrate, tartrate, 2-hydroxybutyrate, and pyridine-2,6-dicarboxylate salts.

Alkali metal, other univalent and alkaline earth cations can be separated with suppression using strong mineral acids, alkyl sulfonic acids, and protonated forms of organic bases such as aniline, ethylenediamine, phenylenediamine, and DAP as the mobile-phase electrolyte. The eluent strength of the latter three are also pH dependent since they are polyprotic acids.

The electrolysis of water, which is the basis for electrochemical regeneration of the MMS, offers other advantages in IC separations. This strategy can be used in IC separations not only to prepare a highly purified mobile phase but also to increase electrolyte concentration electrically in the mobile phase for a gradient application. For example, a carbonate-free NaOH mobile phase can be prepared via water electrolysis and used for IC anion separation.[18,19] Removal of the carbonate significantly lowers the DL for the separated anions.

A new version of IC, called *ion reflux*, also employs water electrolysis in a unique way to generate the mobile phase.[19,20] Analytes are still separated by ion exchange and an electrolyte solution is still the eluent, but it is pure water that is pumped into the system. In ion reflux the electrolyte solution is generated in situ via electrolysis and either analyte anions or cations can be separated.

Figure 8 illustrates an ion-reflux unit for the IC separation of analyte anions. The unit is composed of a source of pure water, a pump, a sample injector, an anion-exchange separator column where the separation of the analyte anions takes place after the sample is injected into the ion-reflux unit, and a generator system where the eluent is formed. This generator system contains two cation exchangers, one is charged in the K^+ form and is placed at the top while the bottom one is charged in the H^+ form as shown in Figure 8. The cathode is in a cathode chamber and is separated from the K^+ cation exchanger by a cation-exchange membrane (shown as the black region in Figure 8). Cations will pass through the membrane but anions do not owing to Donnan exclusion. When the potential is applied and pure water is flowing through the unit, electrolysis of the water will occur at the electrodes with OH^- as KOH (now the eluent electrolyte) forming at the cathode and H^+ at the anode. The H^+ produced at the anode helps to move the K^+ through the cation-exchange membrane to form the KOH eluent. However, a steady state occurs in the generator because the eluent KOH and the separated

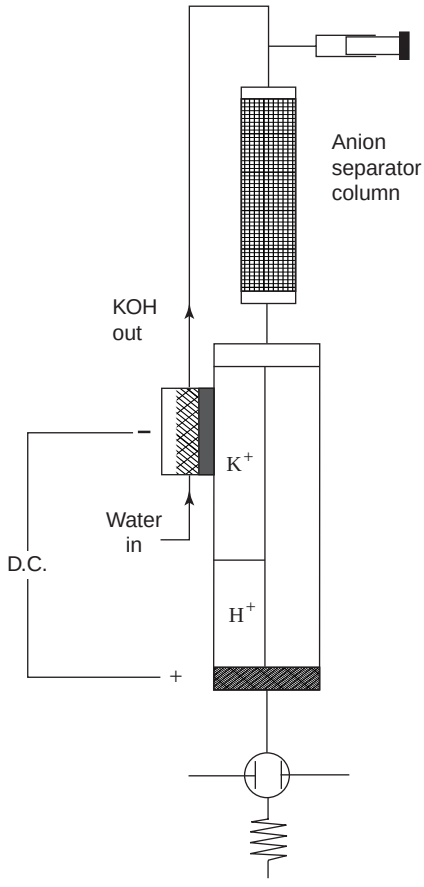

Figure 8 Schematic representation of an ion-reflux unit for the IC separation of analyte anions. Only key components are shown and not drawn to scale. (Reproduced by permission of American Chemical Society, from H. Small, J. Riviello, *Anal. Chem.*, **70**, 2205, (1998).[20])

analyte anions, after passing through the anion separator column, eventually reach the K^+/H^+ cation-exchanger boundary of the generator part of the unit. At this point the KOH comes into contact with the H^+-form cation exchanger and is converted to HOH and the K^+-form cation exchanger. Thus, this is both the suppression step and a recharging of the K^+ cation exchanger. In addition the analyte anions are converted into the acids which eventually pass through the conductivity detector. Since OH^- and H^+ are produced in equivalent amounts during the electrolysis, the K^+/H^+ boundary remains fixed when the steady state is established since K^+ or H^+ are not lost in the operation assuming the sample anions are introduced as the acids. Thus, ion reflux is water in and water plus the separated analyte anions out.

The elution process can be either isocratic or gradient, even though pure water is pumped into the unit. When the current is held constant, a fixed known eluent concentration is generated and the elution is isocratic. If the current is gradually increased in a

known, reproducible manner, then a known, reproducible concentration gradient in OH^- as KOH is produced.

For the separation of cations in an ion-reflux unit the separating column is a cation exchanger and the other two exchangers in the generator part of the ion reflux unit are anion exchangers. The top anion exchanger is charged in either the Cl^- or $CH_3SO_3^-$ form and is separated from the electrode by an anion-exchange membrane while the exchanger at the bottom is charged in the OH^- form.[20]

7 NONSUPPRESSED ION CHROMATOGRAPHY

In nonsuppressed IC, analyte anions are separated usually on a strong base-type anion exchanger, and analyte cations on a strong acid-type cation exchanger. A typical IC set-up for nonsuppressed separation of anions, for example, is similar to that shown in Figure 4. If conductivity detection is used no attempt is made to reduce background conductivity by removal of the background electrolyte. The influence of background conductance on S/N and therefore DL is overcome by altering other factors in the ion-exchange process.

Fritz et al.[21,22] exploited two features of ion exchange to improve conductivity detection in the absence of suppression. First, an ion exchanger was used that is very low in ion-exchange capacity but still capable of providing a high efficiency separation. Typically, this is in the range of $7-100\,\mu eq\,g^{-1}$. Thus, mobile-phase electrolyte concentration required for the separation will be very low and therefore mobile-phase background conductance will be very low. The second modification is to use a mobile-phase electrolyte that provides an eluent anion (for analyte anion separations) or cation (for analyte cation separations) with a high ion-exchange selectivity and an ionic equivalent conductance that differs significantly from the analyte ions. When these two adjustments in the methodology are made and a quality conductivity detector is used that has excellent conductivity offset capabilities and good temperature control, conductivity change due to the analyte gives a good S/N and therefore a favorable DL. Improvements in conductivity-detector electronic suppression of conductivity background and temperature control also contribute to the increased capabilities of nonsuppressed IC.

When the ion-exchange column exchange capacity and the mobile phase electrolyte and its concentration are optimized and a quality conductance detector with good electronic offset, temperature control, and low internal cell volume is used, DLs by nonsuppressed IC often approach those obtained by suppressed IC. Since the exchange capacity is very low, care is needed to avoid column overload with the sample. A major advantage offered by nonsuppressed IC is that this approach is compatible with other types of detection in addition to conductivity. This broadens the scope of applications and sensitivity of nonsuppressed IC. Both direct and indirect absorption, fluorescence, and electrochemical detection methodologies can be used where applicable. In addition nonsuppressed IC is compatible with postcolumn reactions to convert the analytes into products that absorb, fluoresce, or are electrochemically active.

8 ELUENTS FOR NONSUPPRESSED ION CHROMATOGRAPHY

Table 4 lists several common electrolytes employed as mobile-phase electrolytes for the separation of anions. These represent a wide range of eluent strength but are not listed in order of eluent strength. Several have low limiting ionic equivalent conductances, particularly the salts of organic acids, and dilute solutions of these are compatible with conductivity detection. Others are more compatible with other detection strategies. Those that are multiprotic can be changed in eluent strength not only by concentration but also by pH which will determine the charge of the mobile-phase anion.

Borate salts and EDTA are mobile-phase electrolytes whose eluent strength can be varied over a wide range. Borate forms complexes with neutral and acidic polyhydroxy compounds such as sugars and tartaric or gluconic acid, respectively. The charge of the complex, which will determine the eluent strength, depends on the borate concentration, the polyhydroxy compound and its concentration, their ratio, and the pH. At a higher pH a mononegative 1:1 complex forms, which is a weak

Table 4 Common eluent electrolytes for nonsuppressed ion chromatographic separation of anions

Monoprotic acids	Multiprotic acids
Formic acid	HCO_3^-/CO_3^{2-}
Acetic acid	Borate/hydroxy compounds
Benzoic acid and derivatives	*o*-Phthalic acid
Benzenesulfonic acid and derivatives	Tartaric acid
Naphthylamine sulfonic acid and derivatives	Fumaric acid
Alkane sulfonic acids	1,3,5-Tricarboxybenzene
Mineral acids	EDTA
	Sulfobenzoic acid
	Sulfophthalic acid

eluent, low in conductivity, and therefore suitable for conductivity detection. EDTA is a tetraprotic acid and will be present in the mobile phase as a -1 to -4 anion depending on the pH. It is a strong eluent anion as a multivalent anion and is suitable for the elution of highly retained anions, such as polyphosphate analyte anions.

While a strong base, such as a NaOH or KOH mobile phase is easily used in analyte anion separations by suppressed IC, these strong base mobile phases are not suitable for nonsuppressed IC with conductivity detection because the OH^- has such a high ionic equivalent conductance. A salt of a weak acid is a much better mobile-phase electrolyte. A mobile phase containing a strong base, however, is preferred when separating anions of low anion-exchange selectivity or anions derived from very weak acids, such as HCN or H_3BO_3, because OH^- provides a mobile phase of low eluent strength. If a strong base is used, typically a detection method other than conductivity is used.

Electrolytes for the nonsuppressed IC separation of analyte cations are fewer in number compared to nonsuppressed anion IC. A strong mineral acid mobile phase, such as HCl or HNO_3, can be used but is incompatible with conductivity detection, if low DLs are required, because of high ionic equivalent conductance caused by H^+. Buffered solutions of organic bases, such as benzylamine, pyridine, aniline, pyridinecarboxylic acid, and isonicotinic acid are used to obtain the protonated base as the mobile phase electrolyte cation. Ethylenediamine with tartaric acid or α-hydroxybutyric acid, oxalic acid/citric acid mixtures, and pyridine-2,6-dicarboxylic acid eluents are also used and provide both a strong eluent cation and complexing conditions. If multiprotic bases are used, such as ethylenediamine, phenylenediamine, DAP, or polyamines, than multivalent mobile-phase cations are obtained that are much higher in eluent strength depending on mobile-phase pH and pK_b values for the base. Quaternary ammonium (R_4N^+) salts, which can be strong mobile-phase cations depending on the structure of R, can be used but these also can undergo interaction with the cation exchanger matrix particularly if it is a PSDVB polymeric-based cation exchanger. Mono and divalent inorganic cations provide a wide range of eluent strength. However, they are not usually employed with conductivity detection but are applicable to other detection modes. For example, Cu^{2+} or Ce^{3+} are both strong eluent cations ($Ce^{3+} > Cu^{2+}$) and are good mobile-phase electrolytes for the separation and resolution in the nonsuppressed separation of mono and divalent cations on a cation exchanger. Both Cu^{2+} and Ce^{3+} absorb, and thus, detection by indirect absorbance detection is possible. The Ce^{3+} also fluoresces and sensitivity is better with Ce^{3+}, if an indirect fluorescence detection mode is employed.

9 DETECTION IN ION CHROMATOGRAPHY

Most detection strategies used in LC are applicable to IC separations.[1–5,7] However, conductivity detection is the detection method of choice in virtually all suppressed IC separations since the suppression strategy overcomes the limiting effects of a mobile-phase conducting background (see Figure 5a and b). The conductance cell contains microelectrodes; consequently, cell volume is very small often corresponding to a few nanoliters and cell contribution to band broadening is small. Response is rapid, reproducible, sensitive (about 5×10^{-9} g mL^{-1}), and has a linear dynamic range of over three orders of magnitude provided the conductivity cell temperature is constant. While conductivity detection is also widely applied in nonsuppressed IC, other detection strategies are equally as important.

Spectroscopic detection methodology is second to conductivity detection in IC separations and is frequently used with nonsuppressed IC separations. Many inorganic and organic analyte ions absorb in the UV (ultraviolet) or visible region and a direct detection of absorbance change due to change in analyte ion concentration is straightforward. Fluorescence on the other hand is more selective since fewer analyte ions, particularly inorganic analyte ions, will fluoresce. DLs are specific to the analyte ion and will depend on its molar absorptivity for the former detection mode and the fluorescence yield in the latter.

Both absorption and fluorescence detection can be used in a direct or indirect mode. In the direct detection mode, absorbance or fluorescence of the analyte ion is determined in the presence of little or no absorbance or fluorescence in the background. As analyte ion concentration increases, its absorbance or fluorescence increases resulting in the analyte band appearing as a positive peak in the chromatogram. The peak height and peak area are proportional to analyte ion concentration and this correlation is the basis for calibration and subsequent determination of unknown samples of the same analyte ion. In the indirect detection mode the mobile-phase electrolyte absorbs or fluoresces and the analyte ion does not. Since the analyte ion competes with the electrolyte ion of the same charge for the ion-exchange site, a decrease in absorbance or fluorescence, which is due to this exchange competition, is observed on an absorbing or fluorescent background. Thus, negative detection peaks for the separated analyte ions are obtained in the chromatogram whose peak heights and

areas are proportional to each analyte ion concentration. If analyte anions are being separated, the mobile-phase electrolyte must provide an absorbing or fluorescing anion, while for cation separations the electrolyte cation must absorb or fluoresce. Indirect detection strategies are also feasible when using ion interaction reagents as illustrated later. The essential criteria that must be met for calibration, and the handling of the data in these direct and indirect detection strategies in quantitative determination applications is reviewed elsewhere.[2]

Atomic spectroscopic detection strategies are useful in IC separations but limited to those kinds of analyte ions that can be atomized into the vapor state following the separation. Thus, both atomic emission and absorption methodologies are selective rather than universal detection strategies. These modes of detection have been shown to be both sensitive and reliable, where applicable, in IC separations. Of particular interest is detection by inductively coupled plasma mass spectrometry (ICPMS) since both qualitative identification and quantitative determination is possible with this detection system.

Another detection strategy that is vital in IC separations, particularly in the detection of inorganic ions, is a postcolumn reaction that produces a species that absorbs, fluoresces, or is photoluminescent. In this strategy the column effluent containing the separated analyte ions is combined with reagents postcolumn in a controlled way and the conditions are adjusted to produce the detector-active species. For inorganic analyte cations the reaction is often one that involves a formation of a detector-active complex. A successful postcolumn reaction must be reproducible and therefore careful control of the reaction conditions is vital. Typically, reaction time, temperature, pH, mixing rate, and reagent concentrations are parameters that must be optimized according to the chemical reaction and these details are discussed elsewhere.[1–5] Table 5 lists several postcolumn reagents that are routinely used for postcolumn detection in the IC separation of inorganic analyte anions and cations. Since detection is determined by the chemical reaction, selectivity can be incorporated into the detection step by using a reagent for the postcolumn reaction that is selective in its chemistry.

10 APPLICATIONS OF ION CHROMATOGRAPHY

IC is an indispensable analytical technique for the determination of common inorganic and organic analyte ions. As indicated in Figure 1 the method is rapid, reproducible, has excellent DLs and is applicable to trace quantities, can be insensitive to sample matrices, is applicable to complex mixtures, and is accurate and precise. For anion analysis IC is the method of choice for both simple and complex mixtures and has for the most part replaced chemical and other instrumental methods for anion analysis. For cation analysis IC is a companion method to emission and absorption spectroscopic methodologies, but IC does have the advantage of being more applicable to simultaneous multication analysis.

IC is vital to water analysis and frequently is the basis for officially approved methods. It is a routine method in studying drinkable water and rain water. Any industry, such as environmental, clinical, agricultural, pulp and paper, mining, chemical, pharmaceutical, power plant, electronic, and others, that depends on water quality will utilize IC in its analytical laboratories. Numerous examples are in the literature demonstrating the importance of IC in water quality studies. For example, Figure 9(a) and (b) illustrate anion analysis of drinking water and rain water, respectively, where the separation is complete in minutes, while Figure 9(c) shows the separation of the transition metals in an industrial wastewater sample. Suppressed conductivity is used in Figure 9(a) and (b). In Figure 9(c) the transition metals are detected via a postcolumn reaction. Following the separation on the cation exchanger the metal ions are

Table 5 Selected postcolumn reactions used for detection in IC

Analyte	Postcolumn reagent	Reaction	Detection
Inorganic anions	$Fe(ClO_4)_3$	Fe^{3+}-anion complexes	Absorption, 300–340 nm
Phosphorus oxy anions	Mo^{5+}/Mo^{4+}	Phosphomolybdate	Absorption, 830 nm
Phosphorus oxy anions	Al^{3+}-morin	Al^{3+}-anion complexes	Fluorescence, 480 nm
Transition and heavy metals	PAR	Complexes	Absorption, 490–530 nm
Lanthanides, alkaline earth metals	Arsenazo III	Complexes	Absorption, 570 nm
Transition metals	Dithizone	Complexes	Absorption, 590 nm
Zr^{4+}, rare earths, Sc^{3+}, Al^{3+}, Zn^{2+}, Cd^{2+}, Ca^{2+}, Mg^{2+}	8-Oxine	Complexes	Fluorescence, 530 nm
Al^{3+}	Tiron	Complex	Absorption, 310 nm

PAR = 4-(2-pyridylazo)-resorcinol, morin = 2′,3,4′,5,7-pentahydroxyflavone, arsenazo III = 2,7-bis(2-arsonophenylazo)-1,8-dihydroxynaphthalene-3,6-disulfonic acid, dithizone = diphenylthiocarbazone, 8-oxine = 8-hydroxyquinoline, tiron = 4,5-dihydroxy-1,3-benzenedisulfonic acid.

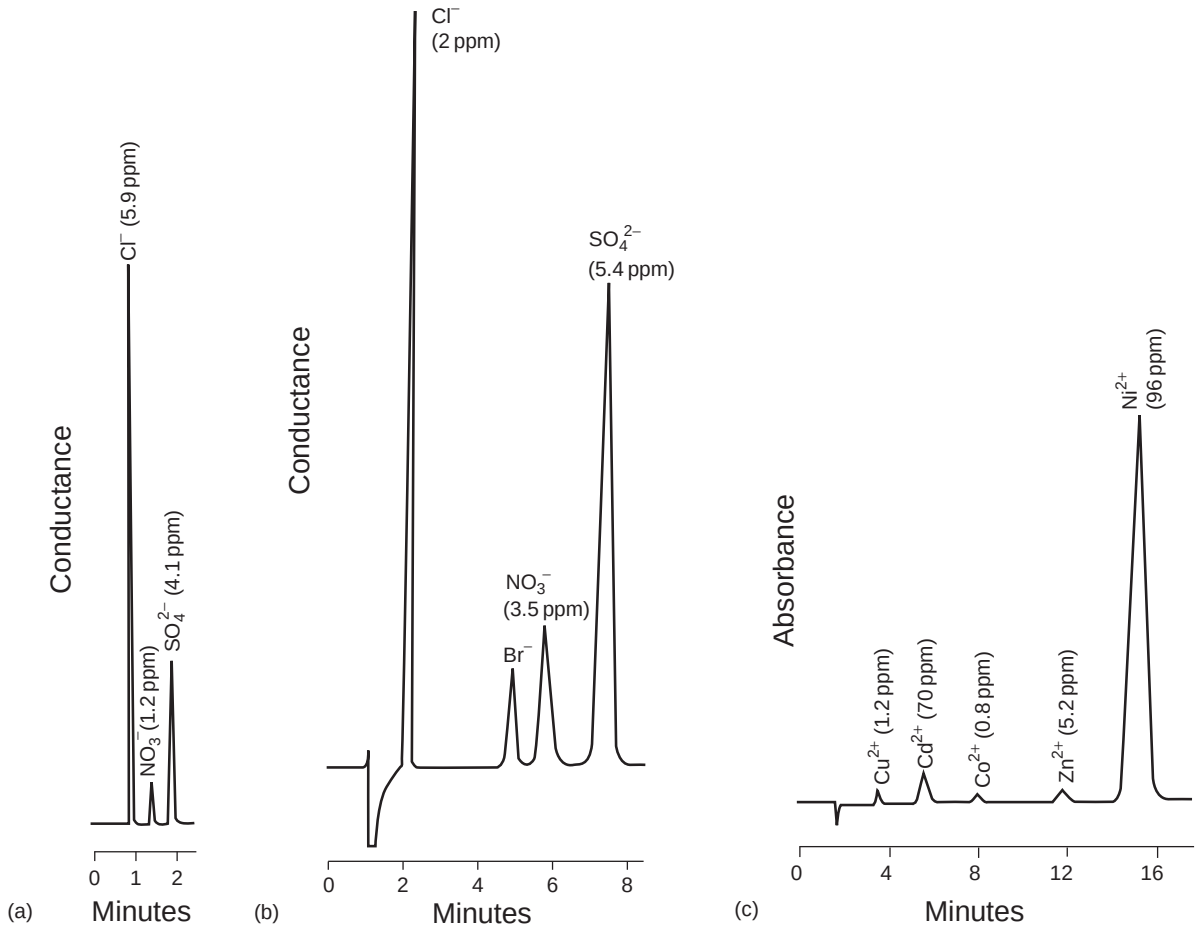

Figure 9 IC anion analysis of (a) drinking water and (b) rain water and cation analysis of (c) industrial wastewater. (a) An aqueous 0.15 mM $NaHCO_3$, 2.0 mM Na_2CO_3 mobile phase at 2.0 mL min^{-1} with 25 µL injection on a Fast-Sep anion-exchange column and suppressed conductivity detection. (b) An aqueous 2.8 mM $NaHCO_3$, 2.2 mM Na_2CO_3 mobile phase at 2.0 mL min^{-1} with 50 µL injection on a IonPac AS4 anion-exchange column and suppressed conductivity detection. (c) An aqueous 50 mM $H_2C_2O_4$, 95 mM LiOH mobile phase at 1.0 mL min^{-1} with 50 µL injection on a IonPac CS5 cation-exchange column with detection at 520 nm after a postcolumn reaction with PAR. (Reproduced by permission of Wiley-VCH from J. Weiss, *Ion Chromatography*, 349, 352, and 357.[2])

combined with PAR and the metal-ion–PAR complexes are detected by absorption at 520 nm.

The excellent resolving power of IC is demonstrated in Figure 10 where 36 analyte anions are separated in about 30 min. Figure 10 also shows that IC is applicable: (1) to the separation of the more highly retained multivalent anions, (2) to organic anions such as the simple carboxylic acids which are important in the food and health sciences, and (3) that gradient elution is readily carried out with suppressed IC using conductivity detection. Although not discussed here gradient elution is also a viable elution strategy with nonsuppressed IC.[2–5]

IC is applicable to ultratrace analysis as shown in Figure 11(a) and (b). Often in these trace-analysis applications concentration of the analyte is required and the IC instrumentation is modified to include this operation. An example of IC trace analysis is illustrated in Figure 11(a), where the IC chromatogram for the determination of SO_4^{2-} in pure NaCl is shown. The Cl^- is removed by a precolumn and the SO_4^{2-} is detected after separation on an anion exchanger by suppressed conductivity. If required, Br^- and NO_3^- can also be determined but absorption detection would be more sensitive. Figure 11(b) illustrates the IC chromatogram for the determination of transition metals in 50% NaOH. The trace metals are concentrated on a precolumn, then separated by cation exchange, and subsequently detected postcolumn by absorption after reaction with PAR.

11 ION EXCLUSION CHROMATOGRAPHY

IEC is an ion-exchange method which allows the separation of ionic analytes from nonionic analytes and

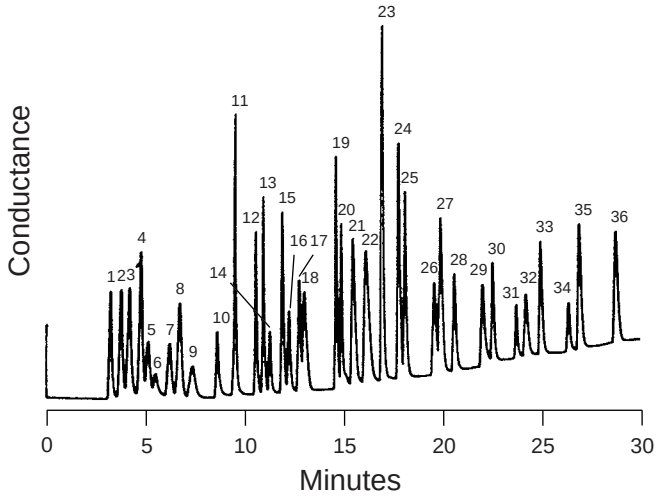

Figure 10 IC separation of a complex mixture of inorganic and organic anions with a gradient elution and suppressed conductivity detection. An aqueous mobile phase of 0.75 mM NaOH for 5 min and gradient to 85 mM NaOH in 30 min at 1.0 mL min^{-1} with a sample that was 10 ppm in each analyte except where noted, containing (1) F^- (1.5 ppm), (2) α-hydroxybutyrate, (3) $C_2H_3O_2^-$, (4) glycolate, (5) butyrate, (6) gluconate, (7) α-hydroxyvalerate, (8) formate (5 ppm), (9) valerate, (10) pyruvate, (11) monochloroacetate, (12) BrO_3^-, (13) Cl^- (3 ppm), (14) galacturonate, (15) NO_2^- (5 ppm), (16) glucoronate, (17) dichloroacetate, (18) trifluoroacetate, (19) phosphite, (20) selenite, (21) Br^-, (22) NO_3^-, (23) SO_4^{2-}, (24) $C_2O_4^{2-}$, (25) selenate, (26) α-ketoglutarate, (27) fumarate, (28) phthalate, (29) oxalacetate, (30) PO_4^{3-}, (31) AsO_5^{3-}, (32) CrO_4^{2-}, (33) citrate, (34) isocitrate, (35) *cis*-aconitate, (36) *trans*-aconitate on a Dionex AS5A anion-exchange column with suppressed conductivity detection. (Reprinted by permission of CRC Press from H.F. Walton, R.D. Rocklin, *Ion Exchange in Analytical Chemistry*, 68.[7])

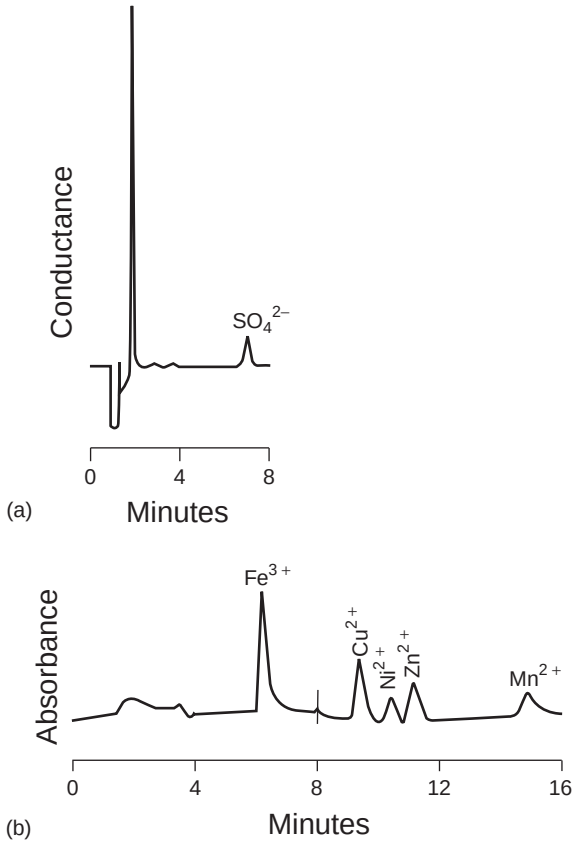

Figure 11 IC in trace analysis, (a) determination of sulfate in pure sodium chloride and (b) determination of transition metals in 50% sodium hydroxide. (a) An aqueous 1.7 mM $NaHCO_3$, 1.8 mM Na_2CO_3 mobile phase at 2.0 mL min^{-1} with 50 µL of a 1% solution injection on a IonPac AS4A anion-exchange column with suppressed conductivity detection. (b) An aqueous 6.0 mM pyridine-2,6-dicarboxylic acid, 10 mM NaOH, 40 mM $NaC_2H_3O_2$, 50 mM $HC_2H_3O_2$ mobile phase at 1.0 mL min^{-1} with a 5 mL sample concentrated to a µL quantity on an IonPac CS5 cation-exchange column with detection at 520 nm after a postcolumn reaction with PAR. (Reproduced by permission of Wiley-VCH from J. Weiss, *Ion Chromatography*, 435 and 437.[2])

is most often employed in the separation of mixtures of weak acids or bases. Ordinary high-capacity ion exchangers, usually the PSDVB type exchangers, are used. Weak acid analytes are separated on strong acid ion exchangers and weak bases are separated on strong base ion exchangers. Instrumentation for HPLC is used in IEC and detection is often by conductivity with or without suppression or by absorbance if the analytes are chromophoric.

While ion exchangers are used as the stationary phase in IEC, an exchange of ions is not responsible for the separation as the analytes travel through the column in the mobile phase.[2–4] Consider the case where the weak acid acetic acid is the analyte, the mobile phase is an aqueous HCl solution, and the stationary phase is a PSDVB-$SO_3^-H^+$ cation exchanger. A simplified schematic illustrating what happens in the column is shown in Figure 12. The system is composed of three main components: one is the solid PSDVB polymeric network, the second is the stationary phase represented by the hydrated $-SO_3^-H^+$ sites, and the third is the mobile phase traveling around and between the exchanger particles. A Donnan-like membrane is established through which certain species only are able to penetrate. Neutral molecules penetrate and ions similar in charge to the co-ions (H^+ in Figure 12) pass into the exchanger resin phase, but ions of similar charge to the ionogenic group, for example $C_2H_3O_2^-$ in Figure 12, do not. The amount of ionized and neutral undissociated forms of the analyte will depend on the analyte's dissociation constant. Strong acid (and base) analytes, which are completely dissociated, are electrostatically repulsed and travel with a volume of mobile phase that corresponds to the column void volume. Partially dissociated analytes, such as acetic acid in Figure 12,

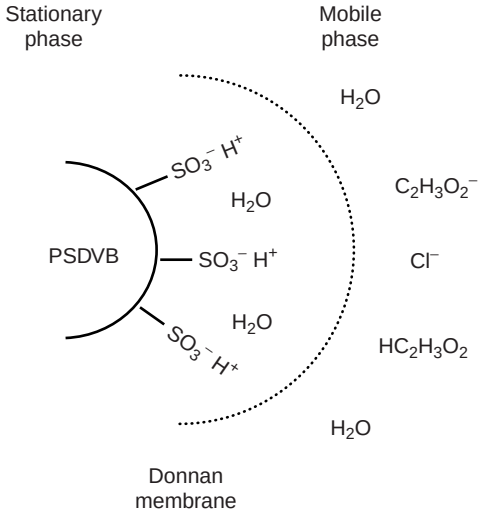

Figure 12 Schematic description of ion exclusion.

penetrate into the resin phase and are therefore retarded in their movement through the column. Elution of the partially dissociated analytes will often correspond to their ionization constants; that is the least dissociated travels the slowest through the column.

While dissociation is a key factor which influences analyte movement through the column in IEC, other factors are also significant depending on the analyte and the ion exchanger used for the IEC separation. Other principal contributing factors influencing analyte retention and elution order are interaction of the analyte with the exchanger matrix (adsorption), analyte size (size exclusion), and polar interactions between the analyte and the exchanger ionogenic group. Because of these interactions, which can be significant, certain neutral molecules can also be retained and resolved on ion exchangers.

IEC is widely used for the separation of weak organic acids. The role of the mobile phase is to dissolve the sample and act as a carrier of the analytes through the column. Often water is the mobile phase. Since dissociation is a key factor in IEC, ionization of weak acid analytes can be suppressed to enhance retention and resolution by including a mineral acid, for example HCl or H_2SO_4, in the mobile phase. Also, acetonitrile or alcohols are added to the mobile phase to increase analyte solubility, influence dissociation, and/or to alter weak acid analyte–exchanger matrix interactions.

IEC is widely used to separate complex mixtures of weak organic acids. In addition IEC is a good method for separating carbohydrates, aldehydes, ketones, and alcohols as neutral molecules. Because of these kinds of application IEC is invaluable as a separation method in the food/beverage industry and in the health sciences.

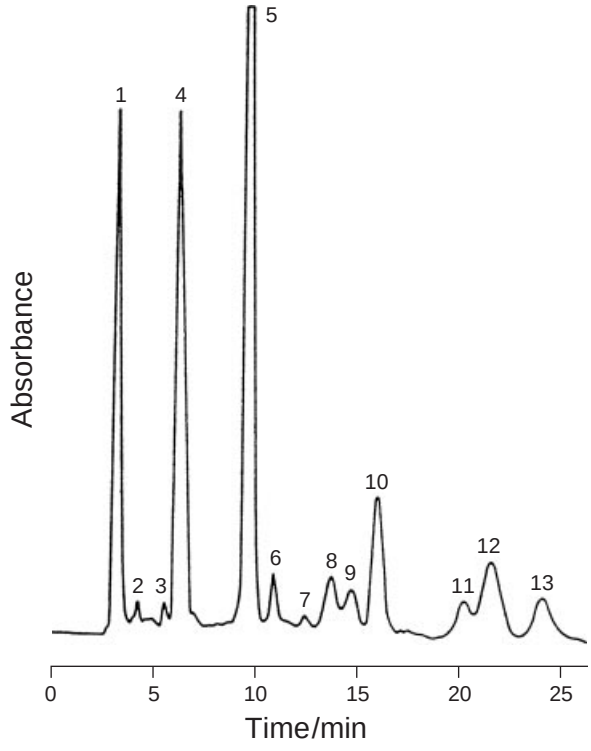

Figure 13 IEC separation of organic acids in a human urine sample. A 1:9 MeOH:H_2O 10 mM H_2SO_4 mobile phase and an Interaction ORH 801 cation-exchange column with detection at 254 nm. Analytes are: (1) oxalic acid, (2) oxaloacetic acid, (3) α-ketoisovaleric acid, (4) ascorbic acid and α-keto-β-methyl-*n*-valeric acid, (5) β-phenylpyruvic acid, (6) uric acid, (7) α-ketobutyric acid, (8) homoprotocatechuic acid, (9) unknown, (10) unknown, (11) hydroxyphenylacetic acid, (12) *p*-hydroxyphenylacetic acid, and (13) homovanillic acid. (Reproduced from P.R. Haddad, P.E. Jackson, *Ion Chromatography Principles and Applications*, 215. Copyright (1990) with permission of Elsevier Science.[3])

Figure 13 illustrates an IEC separation on a commercially available high-capacity cation exchanger of organic acids that are present in a human urine sample.

12 ION INTERACTION CHROMATOGRAPHY

Charge on an organic analyte sharply reduces its retention on a reversed stationary phase often leading to low retention and poor resolution on the reversed stationary phase. The scope of reversed-phase HPLC can be broadened when a lipophilic ion with a charge opposite to the analyte ion charge is included in the mobile phase. Increased retention, selectivity, resolution, and even an increased number of detection options are often realized. This separation strategy to enhance retention and resolution is particularly important in the application

of reversed stationary phases to the separation of mixtures of biological molecules containing acidic or basic groups.

These advantages are also realized when a lipophilic ion opposite in charge is added to the mobile phase to increase retention and resolution in the separation and determination of inorganic and simple organic anions and cations on a reversed stationary phase. This strategy, which is often called IIC or ion pair chromatography, is a useful alternative separation methodology to IEC. To separate analyte anions, a lipophilic cation, such as a tetraalkylammonium salt (for example, $R_4N^+Cl^-$) is added to the mobile phase, while to separate analyte cations, a lipophilic anion, such as a long alkyl-chain sulfonic acid (for example $RSO_3^-H^+$), is added to the mobile phase. Several equilibrium steps contribute to the overall retention of the analyte ion, thus several column and mobile-phase parameters must be optimized and carefully controlled to obtain reproducible separation results. These include pH if analytes and the lipophilic ion are weak acids or bases, buffer components and concentration, ionic strength, selectivity of counterions, mobile-phase organic solvent/water ratio, type of organic solvent, hydrophobicity of the R group in the lipophilic ion, concentration of the lipophilic ion, and the type of reversed stationary phase used.[23,24]

Either silica reversed stationary phases or PSDVB reversed-phase adsorbents are suitable stationary phases. The former offers the major advantages of better mass transfer, better column efficiency, and a better and reproducible stationary-phase surface. The latter offers pH stability throughout the pH range, often higher analyte retention, and larger analyte loading capacity.[2,3,23,24]

Much research has gone into establishing a retention model that is consistent with all the experimental results. For a mobile phase that is largely aqueous and buffered, a lipophilic salt that readily dissociates and is well below its critical micelle concentration and modest in ionic strength, analyte-ion retention on the reversed stationary phase can be described by the ion interaction model. When the mobile phase containing an appropriate concentration of the lipophilic ion, buffer, and ionic strength electrolyte in a mixed organic modifier/water solvent are passed through the reversed stationary phase column, an equilibrium amount of the lipophilic ion and its counterion are maintained on the reversed stationary-phase surface as a double layer. A schematic illustration of the key elements of this model for the case where $R_4N^+C^-$ is the mobile phase lipophilic ion that is continuously passed through the column is shown in Figure 14. When analyte anions represented by X^- in Figure 14 enter the column they compete with C^- in the diffuse secondary layer for the R_4N^+ sites. Differential migration of the different analyte anions occurs because of the preference the R_4N^+ shows for one analyte anion over another compared to C^- just as if it were an anion-exchange selectivity. The similarity is such that the elution order for inorganic analyte anions follows the selectivity for strong base anion exchangers listed in Table 1.

Retention of analyte cations would be described by a similar model except that the lipophilic reagent would be anionic, for example $RSO_3^-H^+$. The primary layer on the reversed stationary phase is RSO_3^- and the countercation in the secondary diffuse layer is H^+ or some other countercation provided by the mobile-phase buffer or ionic strength electrolyte. The analyte cations are therefore in competition with these cations in the diffuse layer.

Suppression methods can be used with analyte anion and cation separations by IIC to enhance conductivity detection. The MMS used when separating anions must contain an organic solvent-resistant membrane that is permeable to the lipophilic cation, for example R_4N^+, while the MMS for cation IIC separations must contain an organic solvent-resistant membrane permeable to the lipophilic anion, for example, RSO_3^-. Both are available commercially. Conductivity detection can be used with nonsuppressed IIC but DLs are usually not as good as those obtained by suppressed IIC. Figure 15 illustrates the chromatogram for the separation of inorganic anions on a C_{18} reversed stationary phase with a quaternary ammonium salt as the mobile-phase lipophilic ion additive and nonsuppressed direct conductivity detection.

Direct absorbance or fluorescence detection is also a viable and widely used detection mode in nonsuppressed IIC when analytes absorb or fluoresce. Indirect absorption or fluorescence detection in IIC also provide good DLs when analyte ions do not absorb or fluoresce. In one approach the counterion, C^- or C^+ (see Figure 14), to the lipophilic ion that is used for analyte anion or cation separations, respectively, absorbs or fluoresces.[25] As the analyte anion or cation band

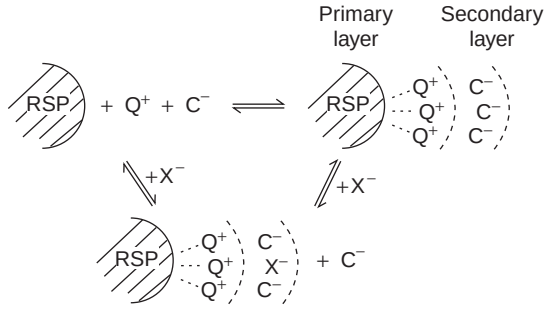

Figure 14 Schematic description of IIC. RSP, reversed stationary phase; Q^+, quaternary ammonium salt; C^-, counteranion; X^-, analyte anion.

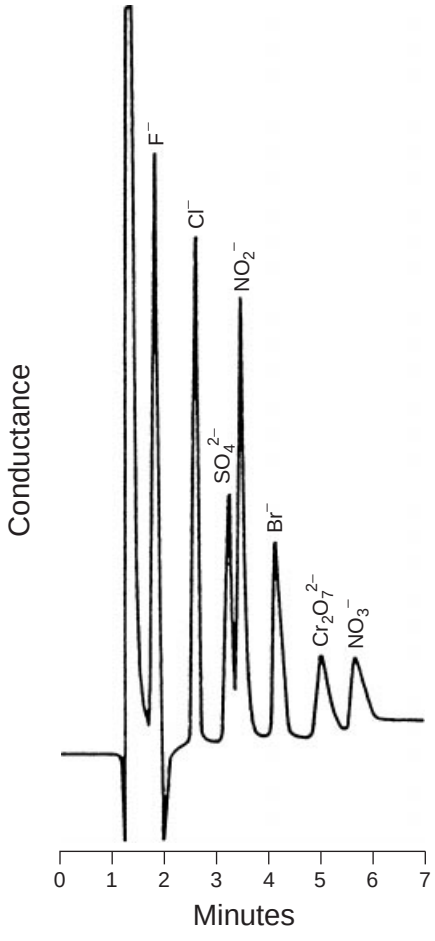

Figure 15 IIC of inorganic anions on a reversed stationary phase. A 2.0 mM tetrabutylammonium hydroxide, 50 mM phosphate pH = 6.7 buffered mobile phase at 2.0 mL min^{-1} with 20 µL injection on a Lichrosorb RP-18 reversed stationary phase column and direct conductivity detection. Each analyte is about 1000 ppm. (Reproduced from P.R. Haddad, P.E. Jackson, *Ion Chromatography Principles and Applications*, 170. Copyright (1990) with permission of Elsevier Science.[3])

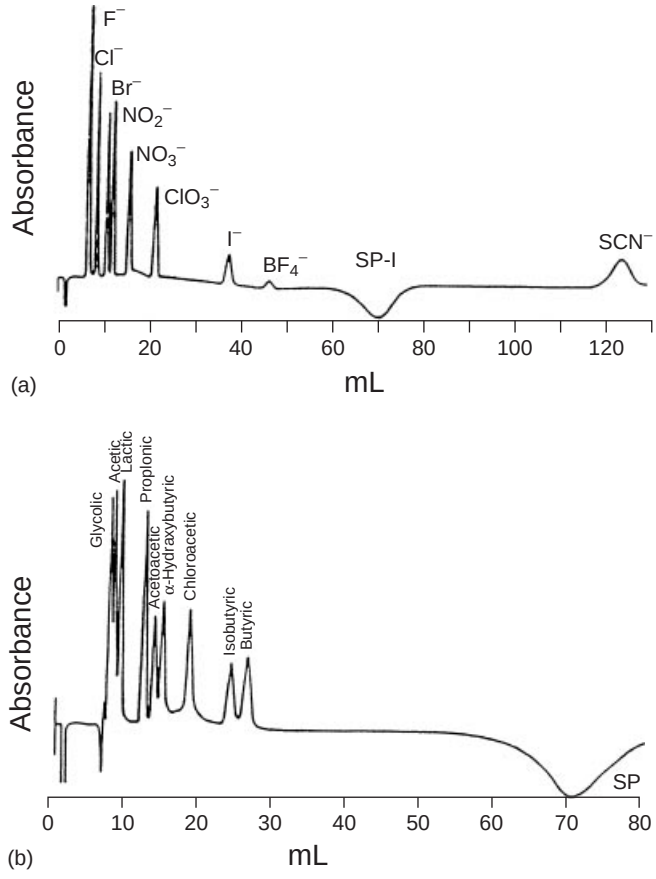

Figure 16 IIC separation of (a) inorganic and (b) organic anions on a reversed stationary phase with indirect absorbance detection. (a) An aqueous 0.1 mM Fe(1,10-phenthroline)$_3SO_4$, 0.1 mM succinate, pH = 6.1 buffered mobile phase at 1.0 mL min^{-1} and a PRP-1 reversed stationary phase. Each analyte was about 1 µg and indirect absorbance detection was at 510 nm. (b) Same as (a) except 0.1 mM Fe(1,10-phenthroline)$_3$ $(ClO_4)_2$ 0.1 mM succinate, pH = 6.1 mobile phase. SP = system peak. (Reprinted by permission from P.G. Rigas, D.J. Pietrzyk, *Anal. Chem.*, **58**, 2226 (1986) and *Anal. Chem.*, **59**, 1388 (1987). Copyright American Chemical Society.[26,27])

moves down the column an equivalent amount of C^- or C^+ in the band is displaced by the analyte ion causing the absorbance or fluorescence in the band to be less relative to the background absorbance or fluorescence. Thus, the analyte peaks appear as negative chromatographic peaks. In a second approach the lipophilic ion itself absorbs or fluoresces.[26,27] Figure 16(a) and (b) show a separation of inorganic and organic anions, respectively, on a reversed stationary phase using the $[Fe(II)-(1,10\text{-phenanthroline})_3]^{2+}$ complex as the mobile phase lipophilic ion. As the analyte anions travel through the column an ionic strength difference in the analyte band relative to the background causes the amount of the complex to be greater in the analyte band. Thus, the analyte peaks appear as positive chromatographic peaks in the chromatogram.

ABBREVIATIONS AND ACRONYMS

DAP	2,3-Diaminopropionic Acid
DL	Detection Limit
EDTA	Ethylenediaminetetraacetic Acid
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IEC	Ion Exclusion Chromatography
IIC	Ion Interaction Chromatography
LC	Liquid Chromatography
MMS	Micromembrane Suppressor
PAR	4-(2-Pyridylazo)-resorcinol

PSDVB	Polystyrene–divinylbenzene
S/N	Signal-to-noise Ratio
USEPA	United States Environmental Protection Agency
UV	Ultraviolet

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis

Clinical Chemistry **(Volume 2)**
Atomic Spectrometry in Clinical Chemistry • Nucleic Acid Analysis in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Heavy Metals Analysis in Seawater and Brines • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Mercury Analysis in Environmental Samples by Cold Vapor Techniques

Field-portable Instrumentation **(Volume 4)**
Field-based Analysis of Organic Vapors in Air

Food **(Volume 5)**
Liquid Chromatography in Food Analysis

Forensic Science **(Volume 5)**
Capillary Ion Electrophoresis in Forensic Science

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Peptides and Proteins **(Volume 7)**
Chromatography of Membrane Proteins and Lipoproteins

Pharmaceuticals and Drugs **(Volume 8)**
Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

REFERENCES

1. H. Small, T.S. Stevens, W.C. Bauman, 'Novel Ion Exchange Chromatographic Method Using Conductimetric Detection', *Anal. Chem.*, **47**, 1801 (1975).
2. J. Weiss, *Ion Chromatography*, 2nd edition, VCH Publishers, Weinheim, 1995.
3. P.R. Haddad, P.E. Jackson, *Ion Chromatography Principles and Applications*, Journal of Chromatography Library-Volume 46, Elsevier Science, Amsterdam, 1990.
4. D.T. Gjerde, J.S. Fritz, *Ion Chromatography*, 2nd edition, Huthig Publishing, Heidelberg, 1987.
5. R.E. Smith, *Ion Chromatography Applications*, CRC Press, Boca Raton, FL, 1988.
6. F. Helfferich, *Ion Exchange*, McGraw-Hill, New York, 1962.
7. H.F. Walton, R.D. Rocklin, *Ion Exchange in Analytical Chemistry*, CRC Press, Boca Raton, FL, 1990.
8. D.J. Pietrzyk, S.M. Senne, D.M. Brown, 'Anion-Cation Separations on a Mixed-bed Ion-exchange Column With Indirect Detection', *J. Chromatogr.*, **546**, 101 (1991).
9. J. Stillian, 'An Improved Suppressor for Ion Chromatography', *LC Mag.*, **3**(9), 802 (1985).
10. A. Henshall, S. Rabin, J. Statler, J. Stillian, 'A Recent Development in Ion Chromatography Detection: The Self Regenerating Suppressor', *Am. Lab.*, **24**(Nov), 20R (1992).
11. J.D. Lamb, L.D. Hansen, G.G. Patch, F.R. Nordmeyer, 'Iodate-Suppressed Lead Eluent for Ion Chromatographic Determination of Cations', *Anal. Chem.*, **53**, 749 (1981).
12. H. Sato, A. Miyanaga, 'Background Suppression by Chelation in the Ion-exchange Chromatographic Separation of Anions', *Anal. Chem.*, **61**, 122 (1989).
13. D.T. Gjerde, J.V. Benson, 'Suspension Postcolumn Detection Method for Liquid Chromatography', *Anal. Chem.*, **62**, 612 (1990).
14. R. Saari-Nordhaus, J.M. Anderson, Jr, 'Ion Chromatographic Analysis of Anions Using a Solid-phase Chemical Suppressor', *Am. Lab.*, **26**(Jan), 28C (1994).
15. P.R. Haddad, S.L. Laksana, R.G. Simons, 'Electrodialysis for Clean-up of Strongly Alkaline Samples in Chromatography', *J. Chromatogr.*, **640**, 135 (1993).
16. A. Siriraks, J. Stillian, 'Determination of Anions and Cations in Concentrated Bases and Acids by Ion Chromatography', *J. Chromatogr.*, **640**, 151 (1993).
17. D. Zhou, D.J. Pietrzyk, 'Liquid Chromatographic Separation of Alkanesulfonate and Alkyl Sulfate Surfactants: Effect of Ionic Strength', *Anal. Chem.*, **64**, 1003 (1992).
18. D.L. Strong, P.K. Dasgupta, K. Friedman, J.R. Stillian, 'Electrodialytic Eluent Production and Gradient Generation in Ion Chromatography', *Anal. Chem.*, **63**, 480 (1991).
19. H. Small, Y. Liu, N. Avdalovic, 'Electrically Polarized Ion-exchange Beds in Ion Chromatography: Eluent Generation and Recycling', *Anal. Chem.*, **70**, 3629 (1998).
20. H. Small, J. Riviello, 'Electrically Ion-exchange Beds in Ion Chromatography: Ion Reflux', *Anal. Chem.*, **70**, 2205 (1998).

21. D.T. Gjerde, G. Schmuckler, J.S. Fritz, 'Anion Chromatography with Low Conductivity Eluents. II', *J. Chromatogr.*, **187**, 35 (1980).
22. J.S. Fritz, D.T. Gjerde, R.M. Becker, 'Cation Chromatography with a Conductivity Detector', *Anal. Chem.*, **52**, 1519 (1980).
23. Z. Iskandarani, D.J. Pietrzyk, 'Ion Interaction Chromatography of Ionorganic Anions on a Poly(styrene–divinylbenzene) Adsorbent in the Presence of Tetraalkylammonium Salts', *Anal. Chem.*, **54**, 2427 (1982).
24. R.M. Cassidy, S. Elchuk, 'Dynamically Coated Columns for the Separation of Metal Ions and Anions by Ion Chromatography', *Anal. Chem.*, **54**, 1558 (1982).
25. B.A. Bildingmeyer, C.T. Santasassia, F.V. Warren, Jr, 'Ion Pair Chromatographic Determination of Anions Using an Ultraviolet Absorbing Co-ion in the Mobile Phase', *Anal. Chem.*, **59**, 1843 (1987).
26. P.G. Rigas, D.J. Pietrzyk, 'Liquid Chromatographic Separation and Indirect Detection of Inorganic Anions Using Iron(II) 1, 10-Phenanthroline as a Mobile Phase Additive', *Anal. Chem.*, **58**, 2226 (1986).
27. P.G. Rigas, D.J. Pietrzyk, 'Liquid Chromatographic Separation and Indirect Detection of Inorganic Anions Using Iron(II) 1, 10-Phenanthroline as a Mobile Phase Additive', *Anal. Chem.*, **59**, 1388 (1987).

Micellar Electrokinetic Chromatography

Ute Pyell
University of Marburg, Marburg, Germany

1 Introduction **1**
2 Theory of Separation **2**
2.1 Separation Principles 2
2.2 Separation by Distribution 3
2.3 Separation by Electrophoresis 4
3 Resolution Optimization **5**
3.1 Resolution of a Peak Pair 5
3.2 Adjustment of Retention Factors 6
3.3 Widening of the Migration Window 8
3.4 Selectivity 9
3.5 Efficiency 11
3.6 Computer-assisted Resolution Optimization 11
3.7 Sample Stacking/Sweeping 12
4 Application **12**
4.1 Sample Preparation and Figures of Merit 12
4.2 Analysis of Drugs 13
4.3 Analysis of Food and Beverages 14
4.4 Miscellaneous 15
5 Separation Carrier Concept **15**
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Micellar electrokinetic chromatography (MEKC) is a highly efficient miniaturized separation technique that is performed with the same experimental set-up that is used with capillary electrophoresis (CE). While the separation principle of CE is based on differences in the electrophoretic mobility of the solutes to be separated (only charged species can be separated), in MEKC separation is also based on differences in the distribution constants between two phases migrating at different velocity due to electrokinetic effects. The second phase is introduced into the separation electrolyte by adding an ionic surfactant (anionic or cationic) at a concentration above the critical micelle concentration (CMC). Aggregates of surfactant monomers (micelles) are present in the separation electrolyte. These aggregates have an electrophoretic mobility and are able to incorporate solutes or interact strongly with species dissolved in the surrounding medium.

MEKC extends the applicability of CE to neutral solutes. In addition, MEKC has been proven to be superior to CE in the separation of very complex mixtures of solutes with similar electrophoretic mobility, because of the various factors that are available for the manipulation of separation selectivity. Currently, MEKC can be regarded as a separation technique with a similar scope to that of reversed-phase high-performance liquid chromatography (RPHPLC), having advantages over the latter concerning the efficiency of the separation system, separation speed, cost and tolerance to matrix constituents, as in MEKC no stationary phase is present that can be fouled by irreversible sorption of matrix constituents of the sample. Owing to the small injection volume (a few nanoliters) and small detection volume that is necessary to avoid intolerable instrumental band broadening, current MEKC instrumentation restricts the applicability of MEKC to the determination of those compounds in complex mixtures that are present in concentrations sufficiently high with respect to the limits of detection that can be obtained with such instrumentation.

1 INTRODUCTION

MEKC as a miniaturized separation technique for both charged *and* noncharged species was initially developed by Terabe et al.[1,2] A theory of the separation principles involved is based mostly on investigations employing the anionic surfactant sodium dodecyl sulfate (SDS) as the micelle-forming agent in the separation electrolyte.[1–5] Otsuka et al.[4] have shown that in place of SDS cationic surfactants [here dodecyltrimethylammonium bromide (DTAB)] can also be successfully employed in MEKC. Their results showed that the separation selectivity in MEKC can be altered easily by variation of the structure of the added micelle-forming agent.

Since these pioneering studies, MEKC has matured into a separation technique that can be employed for the quantitation of noncharged solutes in complex matrices or for the improvement of the separation selectivity in CE. Comprehensive introductions to MEKC are those by Vindevogel and Sandra[6] and Terabe.[7] More recent reviews on current state-of-the-art in MEKC are those by Khaledi[8] and Muijselaar et al.[9]

The instrumentation used in MEKC is identical to that employed in CE (see **Capillary Electrophoresis**). Separation takes place in a fused-silica capillary with an inner diameter $< 100\,\mu m$ and a length mostly varying between 20 and 100 cm. The capillary is filled with the separation electrolyte and is immersed at both

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

ends in vessels filled with the same electrolyte. A high voltage (up to 35 kV) is applied between two electrodes immersed in the same vessels. The sample is injected directly into the first segment of the capillary. In order to avoid instrumental band broadening, detection is effected in a short segment of the capillary, mostly using photometric or fluorimetric detection. The migration of analyte zones is caused exclusively by electrokinetic effects. Hydrostatic liquid flow in the capillary must be avoided. The temperature of the capillary must be controlled, and this can be done either with a thermostated liquid flowing in a jacket around the separation capillary or by placing the capillary in a thermostated chamber.

2 THEORY OF SEPARATION

2.1 Separation Principles

In MEKC, the separation electrolyte contains the background electrolyte (mostly two inorganic salts forming a buffer system) and an ionic surfactant (anionic or cationic) at a concentration above the CMC dissolved in water or a water–organic solvent mixture. According to the theory of zone electrophoresis, during separation the field strength inside the capillary remains constant along the capillary, assuming that the dissolved analytes do not contribute significantly to the charge transport inside the capillary. Thus the field strength can be calculated from the ratio of the voltage (ca. 10–35 kV) to the total length of the capillary (ca. 20–100 cm).

A fraction of the surfactant monomers (depending on the total surfactant concentration and the CMC) dissolved in the separation electrolyte forms aggregates (micelles), which are charged in the case of ionic surfactants. In the case of the most popular surfactant in MEKC, SDS, the mean aggregation number in pure water at room temperature is 64.[10] There is a distribution of the aggregation number and this distribution is dependent on the temperature and on the composition of the surrounding medium.

As a result of electro-osmosis in bare fused-silica capillaries, the liquid in the capillary migrates with constant velocity towards the cathode. The direction of the electro-osmotic flow (EOF) is reversed in the case of a cationic surfactant in the separation electrolyte, because of the formation of hemimicelles at the capillary surface. The micelles are charged, thus have an electrophoretic mobility and migrate in the opposite direction to the EOF. Whether they are finally transported in the same direction as the EOF (but with lower velocity) or show an apparent velocity (apparent velocity of the micelles $\boldsymbol{v}_{MC}$ = electrophoretic velocity $\boldsymbol{v}_{ep}$ + electro-osmotic velocity $\boldsymbol{v}_{eo}$, addition of vectors) in the opposite direction depends on the relation of the absolute electrophoretic velocity $|\boldsymbol{v}_{ep}|$ of the micelles to the absolute velocity $|\boldsymbol{v}_{eo}|$ of the EOF. In all instances, $\boldsymbol{v}_{eo}$ and $\boldsymbol{v}_{ep}$ of the micelles have opposite directions provided that neither $\boldsymbol{v}_{eo}$ nor $\boldsymbol{v}_{ep}$ equals zero.

Dissolved compounds can be present either in the micellar phase or in the surrounding phase. Hence they are distributed between two phases migrating at different velocities (see Figure 1). In the theory of MEKC it is assumed that the velocity of the micelles is not altered by the incorporation of a solute. Consequently, a neutral solute will migrate with an apparent velocity identical with $\boldsymbol{v}_{eo}$ as long as it is dissolved in the surrounding phase and will migrate with $\boldsymbol{v}_{MC}$ as long as it is incorporated into a micelle. Its apparent migation velocity $\boldsymbol{v}_{mig}$ is between $\boldsymbol{v}_{eo}$ and $\boldsymbol{v}_{MC}$ depending on the distribution constant K_{MS} between these two phases (see Figure 2). K_{MS} here is defined as the concentration of solute in the micellar medium divided by the concentration of this solute in the surrounding medium with regard to a section of the capillary. A solute with $K_{MS} = 0$ (i.e. a very polar

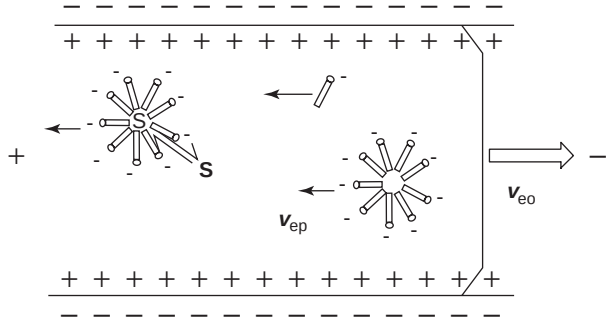

Figure 1 Scheme illustrating separation mechanism in MEKC (anionic surfactant, normal elution mode).

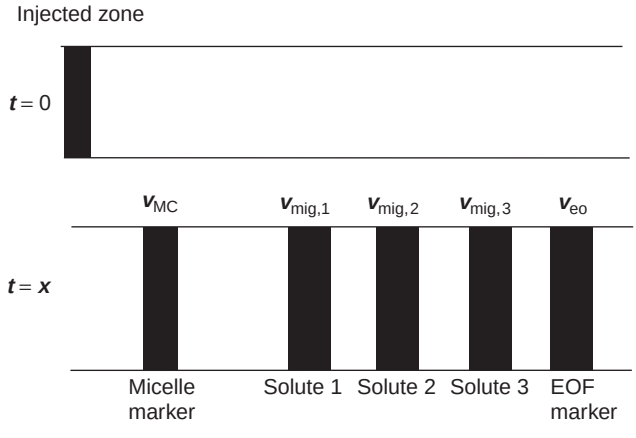

Figure 2 Development of solute zones in a separation capillary, normal elution mode, neutral solutes, K_{MS}(solute 1) > K_{MS}(solute 2) > K_{MS}(solute 3).

compound) can therefore be used to trace $\boldsymbol{v}_{eo}$, and a solute with $K_{MS} \rightarrow \infty$ (i.e. a very nonpolar compound) can be employed to trace $\boldsymbol{v}_{MC}$.

In the case of solutes with an effective electrophoretic mobility different from zero (ions or solutes that are in a dissociation equilibrium with a charged species, i.e. acids or bases), separation is based not only on differences in K_{MS} but also on differences in the effective electrophoretic mobilities μ_{ep} of the solutes. In this case, separation can be regarded as electrophoresis modified by the presence of micelles.

2.2 Separation by Distribution

In MEKC it is possible to separate neutral solutes that differ only in K_{MS}. Figure 3 shows the separation of a mixture of 24 explosives (mostly nitrotoluenes) within 11 min. Separation by distribution between phases of different velocity is analogous to the separation process in chromatography. In analogy with chromatography, the phase surrounding the micellar phase is termed the mobile phase and the micellar phase is termed the pseudostationary phase, taking into consideration that only in rare cases is this phase virtually immobilized (in this case $\boldsymbol{v}_{eo} = -\boldsymbol{v}_{ep}$ of the micelles). In the normal elution mode $|\boldsymbol{v}_{MC}| < |\boldsymbol{v}_{eo}|$. In this mode, the micellar phase is retarded with respect to the surrounding, mostly aqueous phase.

For determining the retention factor k in MEKC, it has to be taken into consideration that the solute is transported not only by the mobile phase but also by the pseudostationary phase.

The retention factor k for a neutral solute is defined according to Equation (1):

$$k = \frac{t_{stat}}{t_{mob}} = \frac{t_R - t_{mob}}{t_{mob}} \tag{1}$$

where t_{stat} = fraction of time present in the stationary phase, t_{mob} = fraction of time present in the mobile phase and t_R = retention time. It is important to note that t_{mob} does not equal the elution time of a nonretarded marker and it has to be calculated from Equation (2):

$$\boldsymbol{v}_{mig} = \frac{t_{mob}}{t_R}\boldsymbol{v}_{eo} + \frac{t_R - t_{mob}}{t_R}\boldsymbol{v}_{MC} \tag{2}$$

Taking into consideration that $\boldsymbol{v}_{mig}$, $\boldsymbol{v}_{eo}$ and $\boldsymbol{v}_{MC}$ can be replaced by t_R/L, t_0/L and t_{MC}/L, respectively [t_0 = elution time of a nonretarded marker ($K_{MS} \rightarrow 0$), t_{MC} = elution time of a marker of the velocity of the micelles ($K_{MS} \rightarrow \infty$) and L = length of the capillary to the detector], Equations (1) and (2) can be transformed by rearrangement into Equation (3):

$$k = \frac{t_R - t_0}{t_0(1 - t_R/t_{MC})} \tag{3}$$

In order to obtain k, markers must be available that make it possible to determine t_0 and t_{MC}. These markers are often thiourea or methanol for the measurement of t_0 and timepidium bromide or quinine hydrochoride (for anionic micelles) or Sudan III for the measurement of t_{MC}.

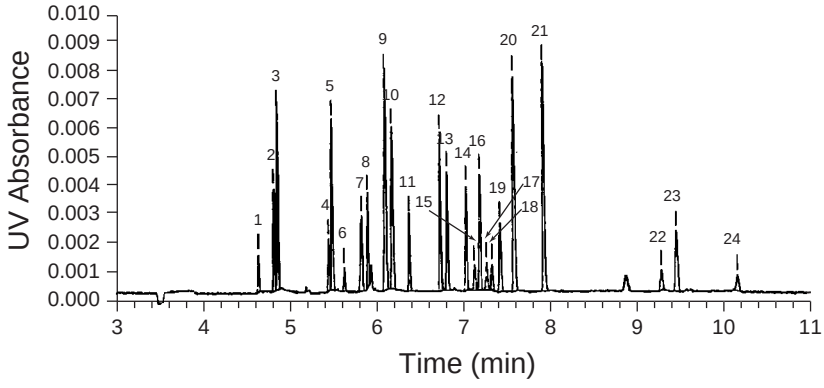

Figure 3 MEKC separation of a mixture of 24 nitroaromatic compounds. Composition of separation electrolyte: 2.5 mmol L^{-1} disodium tetraborate, 12.5 mmol L^{-1} boric acid, 50 mmol L^{-1} SDS, pH = 8.70; photometric detection, λ = 230 nm; U = 25 kV. Capillary dimensions: 50 µm inner diameter, 600 mm total length. (1) 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane, (2) 1,3,5-trinitro-1,3,5-triazacyclohexane, (3) 1,3,5-trinitrobenzene, (4) 1,4-dinitrobenzene, (5) 1,3-dinitrobenzene, (6) nitrobenzene, (7) 2,4-dinitro-6-methylphenol, (8) 1,2-dinitrobenzene, (9) 3-methyl-4-nitrophenol and 2,4,6-trinitrotoluene, (10) 2-methyl-3-nitroaniline, (11) 2,4,6-*N*-tetranitro-*N*-methylaniline, (12) 2-methyl-5-nitroaniline, (13) 2,4-dinitrotoluene, (14) 2,6-dinitrotoluene, (15) 2-nitrotoluene, (16) 3,4-dinitrotoluene, (17) 4-nitrotoluene, (18) 3-nitrotoluene, (19) 2,3-dinitrotoluene, (20) 2-amino-4,6-dinitrotoluene, (21) 4-amino-2,6-dinitrotoluene, (22) 1,5-dinitronaphthalene, (23) diphenylamine, (24) Sudan III. [Reproduced from E. Mussenbrock, W. Kleiböhmer, 'Separation Strategies for the Determination of Residues of Explosives in Soil Using Micellar Electrokinetic Capillary Chromatography', *J. Microcol. Sep.*, **7**(2), 107–116 (1995), Copyright 1995 John Wiley & Sons Inc. Reprinted by permission of John Wiley & Sons, Inc.]

Chen et al.[11] have pointed out that in the presence of organic modifiers in the separation electrolyte, the determination of t_{MC} with a single hydrophobic compound as marker is unreliable. Therefore, the determination of t_{MC} based upon a homologous series has been proposed.[12] Kuzdzal et al.[13] proposed a procedure that can simultaneously determine t_0 and t_{MC} based on the retention times of several molecules in a homologous series. They employed the homologous series of methyl to *n*-pentyl *p*-hydroxybenzoates. In their approach, errors associated with the selection of appropriate markers can be avoided.

In the normal elution mode, t_R for a neutral solute can only have values between $t_0 (k = 0)$ and $t_{MC} (k \to \infty)$; the time span between t_0 and t_{MC} is also called the migration or elution window. This elution window limits seriously the peak capacity obtainable in MEKC. Often the ratio t_0/t_{MC} is used as a descriptor of the elution window, as it reflects the velocity ratio of the two phases. As t_0/t_{MC} decreases, the elution window increases.

If $\boldsymbol{v}_{eo} = 0$, then the analytes have to be transported by the micelles to the detector. Hence the EOF is not a prerequisite for MEKC. If $|\boldsymbol{v}_{ep}|$ of the micelles $> |\boldsymbol{v}_{eo}|$, then those solutes with $k > -t_{MC}/t_0$ have to be transported by the micelles to the detector.[6] For these solutes, the elution order is reversed with respect to the elution order obtained in the normal elution mode (the solute with the highest K_{MS} elutes first). The elution mode in which the solutes are transported with the micelles to the detector has therefore been termed reversed-migration micellar electrokinetic chromatography (RMMEKC). In contrast to the normal elution mode, in RMMEKC the elution window is not limited.

All the above conclusions are valid only in the case of neutral, noncharged solutes with an effective electrophoretic mobility $\mu_{ep} = 0$. If μ_{ep} is not negligible, then the apparent velocity of this solute in the mobile phase will be different from $\boldsymbol{v}_{eo}$. Hence Equation (2) has to be replaced by Equation (4):

$$\boldsymbol{v}_{mig} = \frac{t_{mob}}{t_R}(\boldsymbol{v}_{eo} + \mu_{ep}\boldsymbol{E}) + \frac{t_R - t_{mob}}{t_R}\boldsymbol{v}_{MC} \qquad (4)$$

where $\boldsymbol{E}$ = electric field strength.

The electrophoretic mobility has to be determined in a separate experiment with a separation electrolyte similar to that employed in MEKC without the presence of micelles. If μ_{ep} is known, it is possible to determine the true retention factor from the data already discussed for neutral solutes, replacing Equation (3) by Equation (5):[14]

$$k = \frac{t_R - t_{0M}}{t_{0M}(1 - t_R/t_{MC})} \qquad (5)$$

where t_{0M} = migration time of a solute in the absence of micelles.

It is important to note that t_{0M} is different from t_0. The magnitude t_{0M} has to be calculated from μ_{ep} of the solute, the field strength $\boldsymbol{E}$, the electro-osmotic velocity $\boldsymbol{v}_{eo}$ obtained with the micellar phase and the length $\boldsymbol{L}$ to the detector, according to Equation (6):

$$t_{0M} = \frac{\boldsymbol{L}}{\mu_{ep}\boldsymbol{E} + \boldsymbol{v}_{eo}} \qquad (6)$$

It should be noted, however, that, in MEKC, mobility data obtained with CE experiments have to be used with caution for calculating retention factors of charged solutes. Differences between true and calculated retention factors can arise from neglecting a fundamental difference between the composition of the mobile phase in MEKC and the electrolyte in CE, namely the presence of the surfactant monomer in the mobile phase at a concentration equal to the CMC in the case of MEKC.[15]

2.3 Separation by Electrophoresis

In the case of solutes with non-negligible μ_{ep}, the separation of two solutes is possible either if these solutes differ in the distribution constant K_{MS} or in μ_{ep}. Hence the addition of a surfactant to the separation electrolyte can be a valuable tool to effect the separation of charged solutes, if the selectivity obtained by electrophoresis is not sufficient, and also when electrophoresis remains the dominating effect.

From the viewpoint of electrophoresis, MEKC can be regarded as electrophoresis, with the micellar phase as an "interaction agent" influencing the effective mobility of the solutes.

For qualitative analysis (peak assignment), it can be useful to employ the pseudoeffective mobility $\tilde{\mu}_{ep}$[16] as a descriptor of a solute. This magnitude can be calculated from the elution time t_R of the solute and the elution time t_0 of a non-retained marker (Equation 7):

$$\boldsymbol{v}_{mig} = \frac{\boldsymbol{L}}{t_R} = \boldsymbol{v}_{eo} + (\tilde{\mu}_{ep}\boldsymbol{E}) = \frac{\boldsymbol{L}}{t_0} + (\tilde{\mu}_{ep}\boldsymbol{E}) \qquad (7)$$

The advantage of $\tilde{\mu}_{ep}$ calculated according to Equation (7) over k calculated according to Equation (5) is that neither t_{MC} nor the true effective mobility μ_{ep} have to be determined.

Ghowsi et al.[17] termed $\tilde{\mu}_{ep}$ the effective electrophoretic mobility of the solute. They defined the effective migration mobility of the solute $\tilde{\mu}_s$ as the summation of the effective electrophoretic mobility of the solute $\tilde{\mu}_{ep}$ and the electro-osmotic mobility μ_{eo} (Equation 8):

$$\tilde{\mu}_s = \tilde{\mu}_{ep} + \mu_{eo} \qquad (8)$$

It is important to note that also a neutral compound has a pseudoeffective mobility if k is different

from zero. For a neutral solute the effective electrophoretic mobility as defined in Equation (7) can therefore be described in terms of the retention factor k and the electrophoretic mobility of the micelle $\mu_{ep,MC}$[17] (Equation 9):

$$\tilde{\mu}_{ep} = \left(\frac{k}{1+k} \quad \mu_{ep,MC} \right. \tag{9}$$

For charged and non-charged solutes, $\tilde{\mu}_{ep}$ can be related to the retention time t_R via Equation (10):

$$t_R = \frac{LL_{tot}}{(\tilde{\mu}_{ep} + \mu_{eo})V} \tag{10}$$

where L = length to the detector, L_{tot} = total length and V = voltage. This equation is identical with that usually employed in CE and illustrates the impact of μ_{eo} and other physical parameters on the migration time of a solute.

3 RESOLUTION OPTIMIZATION

3.1 Resolution of a Peak Pair

For deriving the resolution equation for a pair of neutral solutes in the normal elution mode, Equation (3) has to be used instead of the term $k = (t_R - t_0)/t_0$ used in conventional chromatography. The resulting equation was derived by Terabe et al.[2] [Equation 11]:

$$R = \frac{\sqrt{N}}{4} \times \frac{\alpha - 1}{\alpha} \times \frac{k_2}{1 + k_2} \times \frac{1 - t_0/t_{MC}}{1 + (t_0/t_{MC})k_1} \tag{11}$$

where N = number of theoretical plates, α = selectivity factor and k_1 and k_2 = retention factors of the first and second solute (according to the elution order) ($\alpha = k_2/k_1$).

In contrast to conventional liquid chromatography (LC), the resolution is dependent not only on the chromatographic efficiency, the selectivity and the retention factor, but also on the ratio t_0/t_{MC}, which in the normal elution mode is smaller than unity. Consequently, in MEKC in the normal elution mode at identical N, α, k_1, and k_2, the resulting resolution is always smaller than that obtainable with an immobilized stationary phase. The impact of t_0/t_{MC} on the resolution in comparison with conventional chromatography ($t_0/t_{MC} = 0$) is illustrated in Figure 4, where $f(k)$ is plotted against k, approximating $k_1 = k_2 = k$. The term $f(k)$ corresponds to the last two terms of Equation (11). It is also called the retention term and consists of the two terms containing the retention factor (Equation 12):

$$f(k) = \frac{k}{1+k} \times \frac{1 - t_0/t_{MC}}{1 + (t_0/t_{MC})k} \tag{12}$$

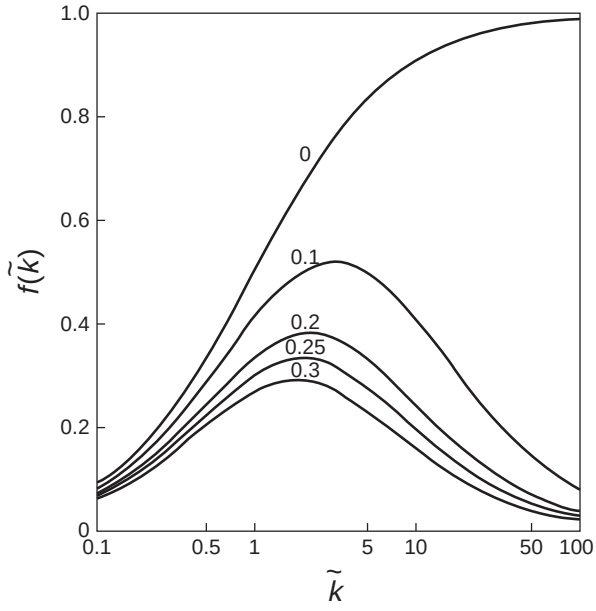

Figure 4 Dependence of $f(k)$ on the retention factor k for different values of t_0/t_{MC} (given on each curve). (Reprinted with permission from S. Terabe, K. Otsuka, T. Ando, *Anal. Chem.*, **57**(4), 834–841 (1985), Copyright 1985 American Chemical Society.)

With Figure 4, Terabe et al.[2] have also shown that the resolution does not increase steadily with increase in the mean retention factor for the solutes to be separated (provided that N, α and t_0/t_{MC} are constant), but reaches a maximum. By differentiation of Equation (12) with respect to k and setting the resulting expression equal to zero, Foley[18] calculated that maximum resolution is obtained for a mean retention factor $\bar{k} = \sqrt{t_{MC}/t_0}$.

In the case of charged solutes, the resolution is affected by both differences in the effective electrophoretic mobilities of the solutes and differences in the retention factor. A resolution equation applicable to charged solutes was derived by Corstjens et al.[19] This equation relates the resolution for a peak pair (normal elution mode) to the capacity factors and the t_{M0} values (see Equations 5 and 6). It can be used to gain insight into the separation mechanisms involved.

Zhang et al.[20] presented a classification scheme that is restricted to the separation of neutral solutes. They defined the phase velocity ratio P_r according to Equation (13):

$$P_r = \frac{t_{MC}}{t_0} \tag{13}$$

With this introduced magnitude, Equation (3) can be rewritten as Equation (14):

$$t_R = t_0(k+1)\frac{P_r}{k + P_r} \tag{14}$$

It is important to note that Equation (14) is valid for any value of P_r. P_r can be negative based on the definition that the velocity is positive when the migration is towards the negative electrode and negative when it is towards the positive electrode.[2] Hence a negative t_{MC} or t_0 indicates that the micellar phase or the surrounding medium migrates towards the positive electrode. The same holds true for the migration time t_R. The absolute value of t_R indicates the elution time of a solute, *if the voltage is suitable*.

According to the value of P_r, MEKC has been classified by Zhang et al. into eight cases, that can be simplified to six cases: (1) $P_r = 0$, (2) $0 < P_r < 1$, (3) $P_r = 1$, (4) $1 < P_r < \infty$, (5) $P_r \to \infty$ or $-\infty$ and (6) $-\infty < P_r < 0$.

The last term of Equation (11) corresponds to the column availability A_{co} introduced by Zhang et al.[20] (Equation 15):

$$A_{co} = \frac{P_r - 1}{P_r + k_1} \tag{15}$$

The normal elution mode corresponds to case (4). In this case $0 < A_{co} < 1$ (see Figure 4). Case (2) would represent the inverted normal elution mode, if a physical realization were possible. Both phases migrate in the same direction, but the micelles migrate faster than the EOF. Case (5) corresponds to conventional chromatography with a mobile and a stationary phase. In this case $A_{co} = 1$. Case (1) corresponds to conventional chromatography if the micellar phase is regarded as the mobile phase and the surrounding phase as the stationary phase. As the micelles are in this case the nonpolar phase, this would correspond to normal-phase LC. In case (3), no separation is obtained, as both phases migrate with the same velocity, $A_{co} = 0$. In case (6), the column availability can be larger than unity. However, in case (6) the migration direction of a solute (towards the cathodic or the anodic end) depends on the retention factor. Case (6) has therefore also been termed the restricted elution mode for $k < -t_{MC}/t_0$.[6] For $k > -t_{MC}/t_0$ it is identical with RMMEKC defined in section 2.2.

3.2 Adjustment of Retention Factors

As depicted in Figure 4, one optimization strategy for the separation of non-charged solutes with MEKC in the normal elution mode is to bring the retention factors of the solutes to be separated into the optimum range. This can be done by:

- variation of the pseudostationary phase/mobile phase phase ratio;
- employing a different surfactant;
- adding a modifier to the separation electrolyte that influences the distribution constant K_{MS}.

The retention factor k is related to the pseudostationary phase/mobile phase phase ratio by Equation (16):

$$k = K_{MS}\frac{V_{mic}}{V_{sur}} \tag{16}$$

where V_{mic} = volume of micellar phase and V_{sur} = volume of surrounding medium. V_{mic} can be determined from the total concentration of surfactant c_{tot} minus the concentration of surfactant present as monomers (this corresponds to the CMC) and the molar volume $\tilde{V}_{surf}$ of surfactant present in the micellar phase[2] (Equation 17):

$$\frac{V_{mic}}{V_{sur}} = \frac{\tilde{V}_{surf}(c_{tot} - CMC)}{1 - \tilde{V}_{surf}(c_{tot} - CMC)} \tag{17}$$

If $c_{tot} - CMC$ is small, $1 - \tilde{V}_{surf}(c_{tot} - CMC)$ can be approximated by unity. Hence there is a linear relationship between the total surfactant concentration and the retention factor[2] (Equation 18):

$$k = K_{MS}\tilde{V}_{surf}(c_{tot} - CMC) \tag{18}$$

By employing Equation (3) for the calculation of k for neutral solutes from experimental data, the validity of the assumptions made for Equation (18) can be verified experimentally.[2] In Figure 5, the dependence of the retention factor for *p*-nitrotoluene on the total concentration of the surfactant SDS in the separation electrolyte is given for different concentrations of urea as modifier. In all cases straight lines are obtained. The intersection of the regression line with the x-axis at $k = 0$ corresponds to the CMC.

The practical range of surfactant concentration is limited. The lower usable concentration limit is determined by the need to generate a micellar phase, hence it must be

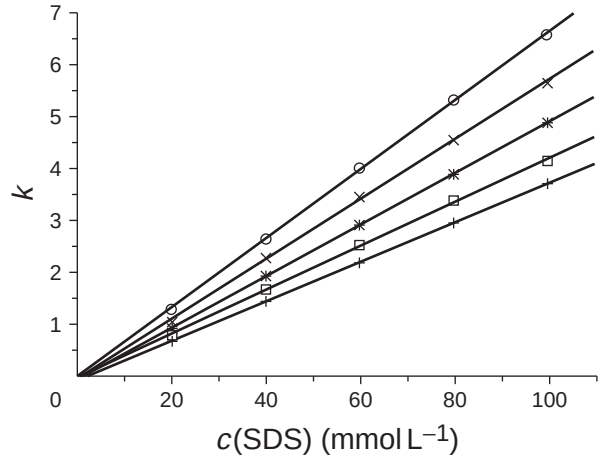

Figure 5 Retention factor k for *p*-nitrotoluene as a function of the concentration of the surfactant SDS in the separation electrolyte. c(Urea) = (○) 0; (×) 1; (∗) 2; ( ) 3; (+) 4 mol L^{-1}. Background electrolyte: 30 mmol L^{-1} disodium tetraborate.

considerably above the CMC. A too low volume of micellar phase might cause band broadening due to overload of the pseudostationary phase. The upper concentration limit is given by the solubility of the surfactant in the surrounding phase (precipitation of surfactant must be avoided) and by limitations concerning the electric current tolerated by the experimental set-up. The addition of an ionic surfactant to the separation electrolyte increases considerably its specific electric conductivity. Consequently, instrumental parameters such as the inner diameter of the capillary, total length of capillary and efficiency of the thermostatting system can have an impact on the upper concentration limit.

In the case of non-polar analytes it might be impossible to bring the retention factors of the solutes to be separated into the optimum range only by varying the total surfactant concentration. In these cases it is possible to use a modifier, mostly an organic solvent[21,22] or urea,[23,24] or urea and an organic modifier simultaneously.[25] The modifier reduces the distribution constant K_{MS}. In Figure 5, the slope of the regression line decreases with increasing content of urea in the separation electrolyte. As the x-axis intercept corresponds to the CMC and there is virtually no dependence of the CMC on the modifier concentration, assumptions are corroborated that in the case of urea as modifier there is only a small impact of the urea concentration on the CMC of SDS. Only a few data are available on the impact of the modifier on the CMC or on the micelle structure (i.e. mean aggregation number, distribution of aggregation number, shape). There is an indication that the addition of urea and of acetonitrile (below a critical concentration) does not alter the CMC of SDS, largely in electrolytes that are usually employed in MEKC.[22,24] Consequently, it can be assumed in those cases that the phase ratio can be approximated to be constant independent of the modifier content.

The large impact of organic modifiers on the retention factors is illustrated in Figure 6(a) and (b), comparing the MEKC separation of dansylated alkylamines effected with a separation buffer containing no organic modifier and effected under identical experimental conditions with a separation buffer containing a volume concentration of acetonitrile of 15% (v/v) and a concentration of urea of 4.1 mol L^{-1}.[25] The concentration of the surfactant SDS was kept constant at 20 mmol L^{-1}. The retention factor of dansylated 1-aminohexane can be reduced (at a constant surfactant concentration) from 185 obtained under the conditions of Figure 6(a) to 1.26 with a separation buffer containing a concentration of acetonitrile of 15% (v/v) and a concentration of urea of 5.0 mol L^{-1}.[25]

The retention factors of solutes can also be drastically reduced by adding cyclodextrins to the separation electrolyte. Terabe et al.[26] termed this variant of MEKC

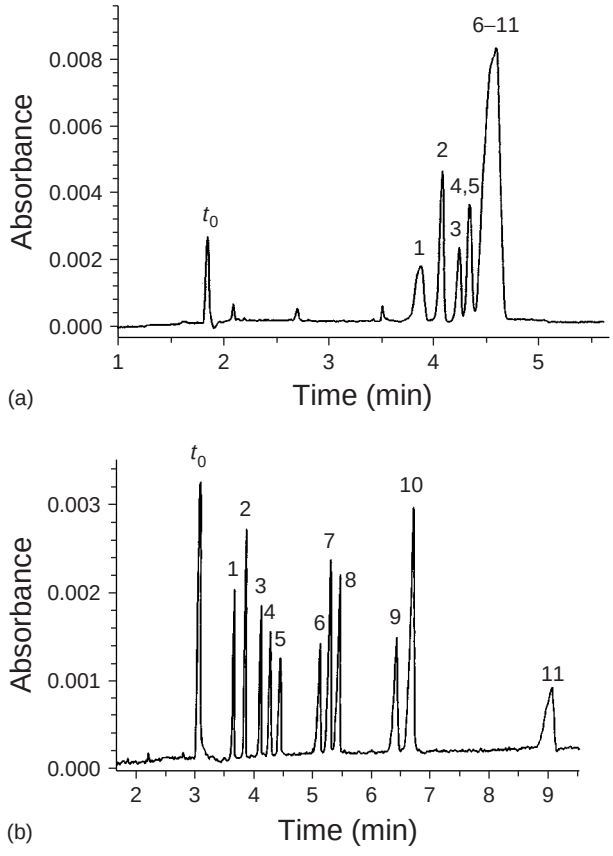

Figure 6 Separation of dansylated aliphatic amines (a) without addition of organic modifiers and (b) with 15% acetonitrile and 4.1 mol L^{-1} urea in the separation electrolyte. Composition of separation electrolyte: 10 mmol L^{-1} disodium tetraborate, 10 mmol L^{-1} boric acid, 20 mmol L^{-1} SDS; photometric detection, $\lambda = 254$ nm; $U = 25$ kV. Capillary dimensions: 75 µm inner diameter, 465 mm total length. (1) 2-aminoethanol, (2) methylamine, (3) ethylamine, (4) morpholine, (5) 2-aminopropane, (6) diethylamine, (7) 1-amino-2-methylpropane, (8) 1-aminobutane, (9) 1-amino-3-methylbutane, (10) 1-aminopentane, (11) 1-aminohexane. (Reprinted from U. Bütehorn, U. Pyell, 'Simultaneous Use of Urea and Acetonitrile as Organic Modifiers for Optimization of Resolution in Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **792**, 157–163 (1997), Copyright 1997, with permission from Elsevier Science.)

cyclodextrin-modified micellar electrokinetic chromatography (CDMEKC). With this method they were even able to separate chlorinated benzene congeners and other highly hydrophobic solutes.

A third possibility for varying the retention factor is the selection of the surfactant used. The micelles or multimers of bile salts (cholates and taurocholates) show a solubilizing power that is usually much smaller than that of SDS. They have therefore been employed for the separation of very nonpolar solutes.[27] Their chiral properties have also been used for enantiomeric resolution.[28]

3.3 Widening of the Migration Window

The considerations outlined in section 3.1 have shown the strong impact of the ratio t_{MC}/t_0 (also termed the phase velocity ratio, P_r) on the column availability A_{co}, and consequently the resolution obtainable [Equation 19]:

$$R = \frac{\sqrt{N}}{4} \times \frac{\alpha - 1}{\alpha} \times \frac{k_2}{1 + k_2} A_{co} \qquad (19)$$

Equation (19) corresponds to Equation (11).

In the normal elution mode for neutral solutes, t_R is restricted to the migration or elution window with the limits t_0 and t_{MC}. From Equation (11) and Figure 4 it can be deduced that in this mode the resolution will be improved if the ratio t_0/t_{MC} is lowered. The limited migration window limits also the peak capacity of the separation system (peak capactity = maximum number of substances that can be separated by a system). For rapid separations, however, a small resolution window [corresponding to a steep gradient in gradient elution high-performance liquid chromatography (HPLC)] can also have advantages.

An overview on parameters controlling the elution window in MEKC with an anionic surfactant (mostly SDS) as micelle-forming agent was given by Muijselaar et al.[29] These parameters include buffer pH, ionic strength, capillary surface modifications, alkyl chain length of the surfactant, surfactant concentration and organic modifiers. Principally, two effects can be used to widen the migration window: (1) increase in the electrophoretic velocity of the micelles and (2) reduction of the electro-osmotic velocity. Although effect (1) has advantages over (2) as it widens the migration window without prolonging the hold-up time of the system, thus increasing the analysis time only slightly, there are difficulties in its physical realization. Hence effect (2) is mostly employed for improving the migration window. It should be noted, however, that by addition of urea to a separation electrolyte containing SDS it is possible to widen the migration window significantly, while t_0 is only marginally increased.[23,24] Other organic modifiers (e.g. methanol and acetonitrile) also widen the migration window.[29,30] These modifiers, however, also have a large impact on the retention factors of the solutes. They are, therefore, predominantly used for adjusting retention factors.

When a separation electrolyte is used that contains an anionic surfactant as micelle-forming agent, the electrokinetic potential and consequently the electro-osmotic velocity is not influenced by the formation of hemimicelles on the negatively charged fused-silica surface. In this case the electro-osmotic mobility is largely dependent on the pH, whereas the electrophoretic mobility of the micelles can be expected to be virtually independent of the pH.[29] Thus in MEKC with an anionic surfactant the pH is a very powerful tool for regulating t_{MC}/t_0 at the expense of increasing the hold-up time. Otsuka and Terabe[31] have shown that with SDS as micelle-forming agent the direction of the apparent velocity of the micelles is changed at pH 5.0. Whereas at high pH the absolute electrophoretic mobility $|\mu_{ep}|$ of the micelles is smaller than the absolute electro-osmotic mobility $|\mu_{eo}|$ corresponding to the normal elution mode, at low pH $|\mu_{ep}|$ is larger than $|\mu_{eo}|$ corresponding to the restricted elution mode or reversed elution mode.[31,32]

Janini et al.[33] have shown that coating of the inner surface of the fused-silica capillary with a linear polyacrylamide can be used to reduce drastically the electro-osmotic mobility. In this case the reversed elution mode can also be obtained at higher pH.

When a cationic surfactant is used as micelle-forming agent in MEKC, μ_{eo} is reversed by the hemimicelles that are formed by the surfactant monomers on the fused-silica surface. Hemimicelles are aggregates of surfactant monomers that are adsorbed on the solid at the solid–liquid interface.[34]

Currently, cationic surfactants are rarely employed in MEKC separations, partly because in the normal elution mode the migration window is generally much smaller when using an alkyltrimethylammonium bromide as cationic surfactant than the migration window that can be obtained with SDS as the micelle-forming agent.[4] Cationic surfactants, however, have a complementary selectivity to other anionic surfactants.[35,36] Moreover, there is indication that cationic surfactants can have advantages over anionic surfactants concerning repeatability in routine analysis.[37] However, means of improving the migration window in MEKC with cationic surfactants have scarcely been investigated so far.

The pH dependence in the presence of a cationic surfactant differs strongly from that observed with an anionic surfactant. Crosby and ElRassi[38] determined the dependence of μ_{eo} on pH with a separation electrolyte containing hexadecyltrimethylammonium bromide (CTAB). Under the conditions of their measurements, the width of the migration window remained virtually unchanged as the pH was varied between 4.5 and 9.0.

In order to widen the migration window, additives (e.g. inorganic salts with divalent metal cations) to the separation electrolyte can be used that are selectively adsorbed on the fused-silica surface, competing with the surfactant monomers for positions in the electric double layer.[39,40] The result is a reduction in the electro-osmotic mobility μ_{eo}. If μ_{eo} is decreased while the electrophoretic mobility μ_{ep} of the micelles is maintained, the result will be an improvement in the migration window. In Figure 7(a)

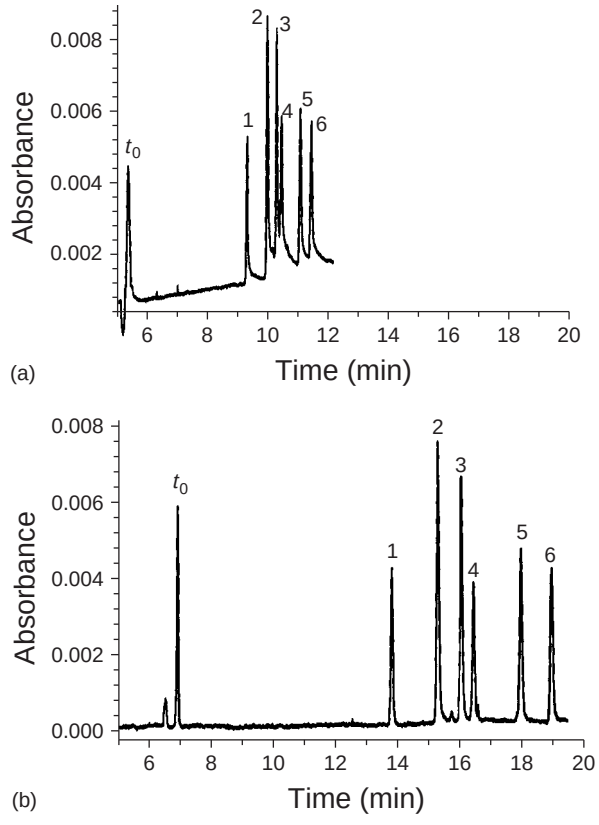

Figure 7 Separation of nitrotoluenes with a separation electrolyte containing (a) 0 and (b) 30 mmol L^{-1} $CaCl_2$. Composition of separation electrolyte: 10 mmol L^{-1} sodium acetate, 10 mmol L^{-1} acetic acid, 40 mmol L^{-1} TTAB, pH = 4.6; photometric detection, $\lambda = 254$ nm; $U = 25$ kV. Capillary dimensions: 75 µm inner diameter, 565 mm total length. (1) 2,4,6-trinitrotoluene, (2) 2,5-dinitrotoluene, (3) 2,4-dinitrotoluene, (4) 2,6-dinitrotoluene, (5) 3,4-dinitrotoluene, (6) 2,3-dinitrotoluene. (Reprinted from A. Dworschak, U. Pyell, 'Widening of the Elution Window in Micellar Electrokinetic Chromatography with Cationic Surfactants. I. Selection of Surfactant, Variation of pH and Addition of Organic Modifiers or Inorganic Metal Salts', *J. Chromatogr. A*, **848**, 387–400 (1999), Copyright 1999, with permission from Elsevier Science.)

and (b) the effect of the addition of $CaCl_2$ to the separation electrolyte on the separation of nitrotoluenes by MEKC with tetradecyltrimethylammonium bromide (TTAB) is shown. While the retention factors are kept virtually constant, there is a large improvement in the resolution due to deceleration of the EOF, resulting in an increase in the ratio t_{MC}/t_0 and improvement of the column availability factor.

3.4 Selectivity

A major advantage of MEKC is its flexibility. The composition of the separation electrolyte and, in particular, the pseudostationary phase can easily be changed by rinsing the capillary. Hence MEKC is more flexible than a chromatographic technique with an immobilized stationary phase. The composition of the mobile phase and the composition of the pseudostationary phase can easily be adapted to the respective separation problem.

In MEKC, the selectivity for separations of neutral solutes is mainly dependent on the surfactant system used. Various types of surfactants have been used in MEKC over the past 10 years: charged surfactants (anionic or cationic), non-ionic and zwitterionic surfactants (see Table 1). In order to separate noncharged solutes, charged surfactants must be used. Surfactants with a zero net charge can be used along with an ionic surfactant as selectivity modifier forming mixed micelles with the charged moities. Nonionic and zwitterionic surfactants can also be used alone to modify the selectivity for the separation of charged solutes.

Table 1 Surfactants employed in MEKC and their CMC determined in distilled water at ambient temperature

Surfactant	CMC (mmol L^{-1})
Anionic surfactants:	
Sodium decyl sulfate	33[a]
SDS	8.2[a]
Sodium tetradecyl sulfate	2.1[a]
Sodium hexadecyl sulfate	0.45[a]
Sodium *N*-lauroyl-*N*-methyl-β-alaninate (ALE)	9.8[b] (40 °C)
Sodium *N*-lauroyl-*N*-methyltaurate (LMT)	8.7[b] (35 °C)
Lithium perfluorooctanesulfonate (LIPFOS)	6.7[c]
Bile salts:	
SC	13[d]
Sodium deoxycholate (SDC)	10[d]
Sodium taurocholate (STC)	10[d]
Sodium taurodeoxycholate (STDC)	6[d]
Cationic surfactants:	
DTAB	16[e] (25 °C)
TTAB	3.6[e] (25 °C)
CTAB	0.92[e] (25 °C)
Nonionic surfactant:	
Brij® 35	0.1[d]

[a] Data from E.A.G. Aniansson, S.N. Wall, M. Almgren, H. Hoffmann, I. Kielmann, W. Ulbricht, R. Zana, J. Lang, C. Tondre, 'Theory of the Kinetics of Micellar Equilibria and Quantitative Interpretation of Chemical Relaxation Studies of Micellar Solutions of Ionic Surfactants', *J. Phys. Chem.*, **80**(9), 905–918 (1976).

[b] Data from S. Takeda, S. Wakida, M. Yamane, K. Higashi, S. Terabe, 'Effect of the Polar Groups of Anionic Surfactant on Migration Behavior in Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **781**, 11–16 (1997).

[c] Data from Khaledi.[8]

[d] Data from A. Roda, A.F. Hofmann, K.J. Mysels, 'The Influence of Bile Salt Structure on Self-association in Aqueous Solutions', *J. Biol. Chem.*, **258**(10), 6362–6370 (1983).

[e] Data from Crosby.[38]

SC, sodium cholate; Brij® 35, polyoxyethylene(23)dodecyl ether.

The fundamental importance of selectivity for resolution is represented by the presence of the selectivity factor α in Equation (11). For neutral solutes the choice of surfactant is the most important consideration in optimizing selectivity. Poole and Poole[41] have shown that the experimental parameters surfactant concentration, buffer concentration, buffer pH, temperature and choice of buffer have little influence on the selectivity (for neutral solutes and SC as surfactant). There is some alteration in the selectivity with organic modifiers in the separation electrolyte. However, the choice of organic solvent (5%, v/v) as modifier (methanol, acetonitrile, tetrahydrofuran, propan-2-ol) seems to be virtually without influence on the selectivity.

For rapid method development in MEKC, it is therefore highly desirable to have a set of surfactants with complementary selectivity. The following constraints have to be respected: (1) their Krafft points must be lower than ambient temperature; (2) their CMC in the separation electrolyte (different from that in pure water) should not exceed 20–30 mmol L^{-1}; (3) their solubility in aqueous buffers at ambient temperature must be sufficient. These constraints can mostly be fulfilled by only one member of a homologous series with varying alkyl chain length (see Table 1).

Yang and Khaledi[42] and Yang et al.[43] employed linear solvation energy relationships (Kamlet–Taft solvatochromic model) in order to make an attempt to classify surfactants according to the type of solute–surfactant interaction. They concluded that hydrogen-bonding interactions play a major role in providing different chemical selectivity among the five surfactant systems that they investigated.

Poole and Poole[41] employed the solvation parameter model, that is in use in areas of retention modeling in chromatography, in order to characterize the surfactant selectivity in MEKC. In this model the solvation process is described in terms of fundamental intermolecular interactions with solute or solvent properties determined from equilibrium processes. Employing data obtained by applying this model, they suggested a working list for surfactants for selectivity optimization in MEKC: SDS, SC, lithium perfluoro-octanesulfonate, *N*-dodecanoyl-*N*-methyltaurine and TTAB.[44]

Muijselaar et al.[45] introduced the retention index concept for application in MEKC. They defined the retention index I of a solute according to Equation (20):

$$I = 100z + 100\frac{\log\left(\frac{t_s - t_0}{t_{MC} - t_s}\right) - \log\left(\frac{t_z - t_0}{t_{MC} - t_z}\right)}{\log\left(\frac{t_{z+1} - t_0}{t_{MC} - t_{z+1}}\right) - \log\left(\frac{t_z - t_0}{t_{MC} - t_z}\right)} \quad (20)$$

where z = number of carbon atoms in a member of a homologous series eluted before the solute, t_s = elution time of a solute, t_0 = hold-up time, t_{MC} = migration time of the micelles, t_z = elution time of a neighboring member of a homologous series (eluted before the solute) and t_{z+1} = elution time of a neighboring member of a homologous series (eluted after the solute). The retention index defined according to Equation (20) is not dependent on the phase ratio, i.e. it is independent of the surfactant concentration. Alkylbenzenes and alkyl aryl ketones can be used as retention index standards in MEKC. The retention index is useful as a parameter for peak identification. In analogy with the Rohrschneider–McReynolds scale in gas chromatography (GC), the retention index can be also employed for the characterization of pseudostationary phases.[46]

Fine tuning of the selectivity can be performed by employing mixed micelles. To this end, Little and Foley[47] employed a mixed micellar system consisting of SDS and Brij® 35. The variation of selectivity with composition for a mixed-micellar buffer has been studied by employing the solvation parameter model.[48] The change in selectivity associated with the addition of Brij® 35 to separation electrolytes containing SDS has been reported by Rasmussen et al.[49] The main disadvantage of this approach is an unfavorable narrowing of the migration window in the normal elution mode caused by the integration of the nonionic surfactant into the anionic micelles, thus reducing the electrophoretic mobility of the micelles. As can be seen from Equation (11), this effect (increase in t_0/t_{MC}) can counteract the resolution improvement through increase in α.

A general overview on techniques that can be employed to manipulate the selectivity in MEKC was given by Terabe.[50] These techniques comprise the choice of the surfactant and the use of mixed micelles, but also additives such as ion-pair reagents, urea, organic solvents and metal ions. For ionizable solutes a strong impact on selectivity is given by the pH of the separation electrolyte. An extensive investigation on the migration behavior of acidic solutes has been presented by Khaledi et al.[51] Their phenomenological approach assumes that the parameters, retention factor and electrophoretic mobility, of an acidic solute are the weighted average of the parameter of the solute in the dissociated and the undissociated forms. By variation of the pH, the dissociation equilibrium is influenced, and thus also the effective retention factor and the effective electrophoretic mobility. Khaledi et al. developed equations describing the relationships between the two migration parameters – retention factor and mobility – on the two experimental parameters: pH and surfactant concentration. The same approach can be employed for an understanding of the dependence of the migration behavior of basic solutes on pH.[52]

If basic solutes are separated in MEKC with an anionic surfactant or acidic solutes are separated in MEKC with a

cationic surfactant, not only the acid–base equilibrium but also ion-pair formation equilibrium between the charged species and the monomer surfactants and interaction between the charged species and the micellar phase have to be taken into consideration. Ionic species exhibit a strong interaction with a surfactant having the opposite charge and have a measurable retention factor due to interaction with the oppositely charged micellar phase.[53]

For the separation of acidic solutes (partly or fully dissociated) or basic solutes (partly or fully protonated), the selectivity can be completely altered (at constant pH) if a cationic surfactant is used instead of an anionic surfactant.[4] Nishi et al.[5] have demonstrated that the selectivity can also be effectively manipulated in MEKC with an anionic surfactant (separation of carboxylic acids and cationic and zwitterionic substances) by addition of an ion-pair reagent (tetraalkylammonium bromide). By addition of the ion-pair reagent to the separation electrolyte containing SDS as micelle-forming agent, the migration time of cationic solutes is decreased and the migration time of anionic solutes, in contrast, is increased at about constant electro-osmotic velocity.

3.5 Efficiency

The efficiency of a chromatographic system is quantified by the number of theoretical plates per column, N. A high efficiency is important for the resolution of closely related compounds [see Equation 11], for rapid separations and for obtaining a sufficiently high peak capacity. MEKC can be characterized as a separation method with inherently high efficiency – one of its main advantages over conventional chromatographic techniques. The number of theoretical plates per capillary (length to the detector 50–70 cm) obtained in MEKC for retarded ($k = 1–5$) neutral solutes under routine conditions is about 200 000–300 000.[54,55]

Terabe et al.[54] have studied in detail for MEKC the band-broadening phenomena that occur in the separation capillary during a run. These phenomena have to be divided into two groups: the extra-column and the in-column effects. Extra-column effects comprise the injection volume, the cell volume of the detection part of the capillary, the time constant of the detector and the sampling rate of the data system or the response of the recorder pen. With suitable injected plug length and adequate in-column detection, extra-column effects can be kept low.

The causes of in-column band broadening can be ascribed to five mechanisms: longitudinal diffusion, sorption–desorption kinetics, intermicelle mass transfer, radial temperature gradient effects and electrophoretic dispersion of the micelles (heterogeneity in micelle mobilities). At slow velocity of the solute zones, longitudinal diffusion is the dominant factor, whereas at higher velocities sorption–desorption kinetics and heterogeneity in micelle mobilities become significant factors. For separations in the normal elution mode, Terabe et al.[54] recommend working with an electro-osmotic velocity $>1\,\mathrm{mm\,s^{-1}}$ and keeping the retention factors to <5.

The influence of the surfactant concentration on efficiency is controversially discussed in the literature.[54,56,57] Yu et al.[57] did not obtain a significant dependence of N on the SDS concentration of the separation electrolyte for moderately retarded solutes in the concentration range $15–100\,\mathrm{mmol\,L^{-1}}$. They showed that it is possible in MEKC to obtain an efficiency that is limited only by the injected sample plug length and longitudinal diffusion. At very high electro-osmotic velocity, factors corresponding to the resistance in mass transfer known for conventional chromatography are dominant, thus resulting in a drastic decrease in N with increasing electro-osmotic velocity.[55,57,58]

In order to use the high efficiency obtainable in MEKC, band broadening due to inadequate buffer leveling has to be avoided. If the buffer levels in the inlet and the outlet vessel are not identical, a laminar flow in the capillary independent of the EOF will be induced (siphoning effect). In theoretical considerations, a perfect plug-like flow is assumed. This assumption is not fulfilled in the case of siphoning effects. Siphoning effects can result in a dramatic decrease in efficiency. The same holds true for band broadening due to solute–wall interactions. Muijselaar et al.[29] reported an intolerable decrease in efficiency when employing coated capillaries in MEKC. Band-broadening effects due to solute–wall interactions can also be due to fouling of the capillary.

3.6 Computer-assisted Resolution Optimization

Separations in MEKC are characterized by a large number of parameters and therefore empirical method development can be very difficult and time-consuming. Further complications can arise from the mutual interaction of these parameters affecting the separation. Computer-assisted resolution optimization schemes can be regarded as a tool in method development that gives guidelines for selecting the significant parameters, and that makes it possible to find suitable conditions after a minimum of initial experiments. Although for HPLC several optimization schemes have been published, these approaches at least require substantial modification before they can be used to optimize an MEKC separation.

An exhaustive and instructive introduction into this area was given by Khaledi et al.[59] In a review by Corstjens et al.,[60] typical features of optimization

strategies applicable to MEKC are discussed. A more recent review on optimization strategies developed for MEKC was given by Khaledi.[8]

Approaches based on physicochemical models have to be distinguished from approaches based on empirical models. Although the use of physicochemical models is always preferred, since an in-depth understanding of the underlying equilibria can be achieved, such an approach is not always feasible, as numerous experiments can be required to determine the relevant constants in the equations with sufficient accuracy.

Smith and Khaledi[61] and Quang et al.[62] developed a strategy to model the migration behavior on the basis of physicochemical parameters estimated from model fitting on the basis of few initial experiments. With their approach, the rapid optimization of pH and surfactant concentration is possible. The linear relation between the retention factor and the concentration of surfactant [see Equation 18 and Figure 5] was used by Pyell and Bütehorn[24] to develop a strategy for the optimization of the surfactant concentration (one-parameter optimization) on the basis of only two initial experiments. This approach has been extended to the simultaneous optimization of the surfactant and the modifier concentration for urea[63] and acetonitrile[22] as modifier.

For certain parameters (e.g. temperature), physicochemical models are unavailable. In this case, empirical (also called statistical) models are useful in minimizing the number of experiments required to find suitable conditions. In addition to factor-based methods,[64] a simplex algorithm,[65] the overlapping resolution mapping scheme[66] and the iterative regression strategy[67] have been applied in MEKC. Wiedmer et al.[68] optimized the composition of a separation electrolyte containing SDS and SC by using empirical quadratic regression models.

3.7 Sample Stacking/Sweeping

Because of the small detection cell volume, which has to be chosen to avoid radial temperature gradient effects due to Joule heating (limitation of the inner diameter of the separation capillary) and to avoid extra-column band broadening (limitation of the length of the detection cell), MEKC methods often exhibit detection limits that are one or more orders of magnitude lower than in corresponding HPLC methods. The disadvantages due to the limitations of the detection volume, however, can possibly be overcome by on-line sample concentration techniques, which are not applicable in HPLC.

One on-line technique applicable in MEKC is stacking. Sample stacking was initially developed for charged analytes in CE. It is based on differences in the electric conductivity of bordering zones, resulting in differences in the effective migration mobilities $\tilde{\mu}_s$ of the solutes, following from the Kohlrausch regulating function. If the sample solution has a conductivity much lower than that of the separation electrolyte, the electric field strength in the sample zone will be significantly higher than in other parts of the capillary. As the ions move across this boundary that separates regions of different electric field strength, they will be focused in a zone narrower than the initially injected sample zone. Sample stacking techniques developed for CE can be also applied for the MEKC separation of charged solutes.

According to Equation (9), a neutral compound also has a pseudoeffective mobility if k is different from zero. Hence sample stacking is possible with neutral solutes in MEKC, provided that the solutes exhibit an effective migration mobility $\tilde{\mu}_s$ due to interaction with a pseudostationary phase during the stacking process.[69–72]

A second mechanism suitable for on-line concentration of neutral analytes in MEKC is sweeping. Sweeping is a term that was coined by Quirino and Terabe, defined as 'the picking and accumulation of analytes by the pseudostationary phase that fills the sample zone during application of voltage'.[73,74] It occurs whenever the sample matrix is devoid of pseudostationary phase and it does not need a lower electric conductivity in the sample solution than in the separation electrolyte. Sweeping can be compared to injection zone focusing processes occuring in conventional chromatography, when the sample solution is injected into the separation column. As has been shown by Quirino and Terabe,[73,74] in the reversed elution mode this concentration effect is effective, in particular, with solutes with very high retention factors. An on-line concentration effect that has been reported by Gilges[75] for the determination of impurities in a drug substance by MEKC in the normal elution mode is obviously also due to sweeping.

Theoretically, it should be possible to combine stacking and sweeping by preparing the sample in a solution with low electric conductivity containing no pseudostationary phase. The full potential of on-line concentration techniques in MEKC has not yet been pushed to its limit. On-line concentration techniques will play a key role in extending the applicability of MEKC to trace analysis and making this technique competitive in comparison with HPLC.

4 APPLICATION

4.1 Sample Preparation and Figures of Merit

There is no stationary phase in MEKC that can be fouled by irreversible sorption of matrix constituents, as the

separation electrolyte, containing the pseudostationary phase, is exchanged by rinsing procedures after each run. However, irreversible sorption of matrix constituents on the inner capillary wall must be avoided, as it results in an irreproducibility of migration times depending on the status of the capillary surface. In order to avoid blocking of the capillary by injected particles, samples should be filtered with a membrane filter prior to injection.

Because of the absence of a stationary phase and solubilization of matrix constituents by the micellar phase, direct sample injection is possible in many cases. In these cases, the highly efficient separation system must provide sufficient separation of the solutes of interest from interfering matrix constituents. In the case of direct sample injection of 'dirty' samples, often specific rinsing procedures between the runs have to be applied in order to maintain the repeatability of migration times.[76]

With proper rinsing procedures and suitable thermostating of the capillary, the relative standard deviation of the retention times (intra-day) for detectable solutes in real samples is generally <1%.[76–79] In practice, because of electrode reactions, buffer depletion can occur, resulting in a drift of the pH of the separation electrolyte and other parameters. This problem, however, can easily be overcome by a frequent change of the separation electrolyte. If the composition of the sample solution deviates strongly from that of the separation electrolyte (e.g. much higher content of organic modifier in the sample than in the separation electrolyte), peak splitting, retention time shifts and peak deformations (fronting, tailing, broadening) can result.[16,78,80] Consequently, mismatch of the composition of the sample solution with the composition of the separation electrolyte has to be avoided or must be taken into account for the preparation of standard solutions.

In some publications, figures of merit obtainable with an HPLC method are compared with those that can be obtained with an MEKC method: determination of nitroaromatic compounds in soil,[79] phospholipids in lecithin,[81] phenolic compounds in spiritous liquor,[77] impurities in a pharmaceutical compound[75] and pesticides in aqueous matrix.[82] In all these comparative studies photometric detection was employed. Although higher than those in HPLC, the limits of detection obtained in MEKC were fit for purpose.

Data obtained by HPLC agreed well with those given by MEKC.[77,79,81] The precision of the determination expressed as relative standard deviation for the determined peak area is strongly dependent on the method parameters. With an internal standard (in order to eliminate imprecision in sample injection), relative standard deviations for the peak areas normalized on the peak area of an internal standard of <1% can be obtained.[78] With real samples (coffee, spiritous liquor) and suitable rinsing of the capillary between runs, relative standard deviations for the peak areas determined for consecutive runs (without internal standard) as low as 1–1.5% have been reported. In other studies, a significantly higher relative standard deviation for the (absolute) peak areas was obtained for the MEKC method in comparison with HPLC.[79,81] However, the MEKC method can offer advantages over conventional HPLC methods in the routine analysis of real samples because of the higher peak capacity, higher efficiency and a selectivity different from that obtained in RPHPLC with conventional stationary phases: improved selectivity, higher accuracy due to baseline separation of adjacent peaks, separation of very closely related compounds (higher information content) and short run times.[75,81]

4.2 Analysis of Drugs

Nishi and Terabe[83] showed in a review on drug analysis by MEKC that most types of analysis in this field can be

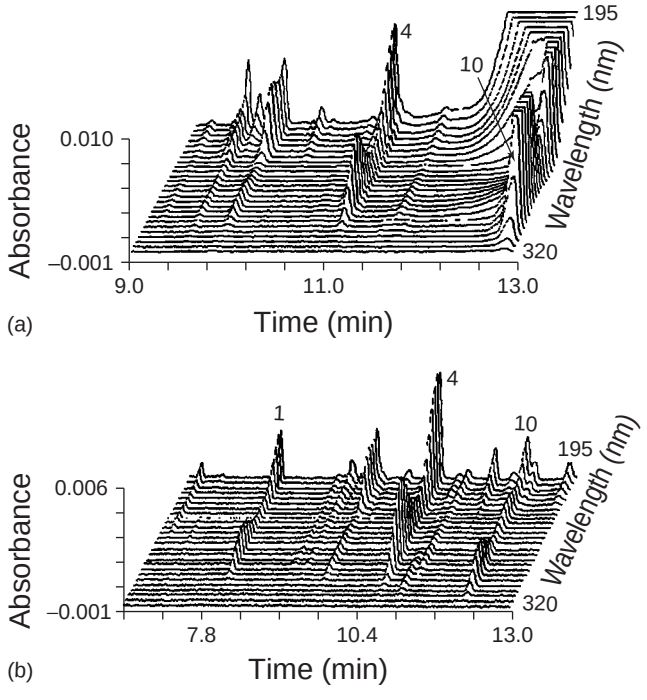

Figure 8 Three-dimensional plot for the MEKC analysis of (a) a directly injected and (b) a cleaned-up serum sample from a patient receiving theophylline pharmacotherapy. Peak assignment: (1) theobromine (internal standard); (4) theophylline; (10) uric acid. Composition of separation electrolyte: 6 mmol L^{-1} disodium tetraborate, 10 mmol L^{-1} disodium phosphate, 75 mmol L^{-1} SDS, pH = 9; detection with photodiode detector; $U = 25$ kV. (Reprinted from W. Thormann, A. Minger, S. Molteni, J. Caslavska, P. Gebauer, 'Determination of Substituted Purines in Body Fluids by Micellar Electrokinetic Capillary Chromatography with Direct Sample Injection', *J. Chromatogr.*, **593**, 275–288 (1992), Copyright 1992, with permission from Elsevier Science.)

performed by using MEKC. These applications include purity testing of drugs, assay of drugs, determination of drugs in biological fluids, including the direct injection method, and separation of complex mixtures. Nishi and Terabe[83] highlighted that MEKC is, in particular, a powerful method for the separation of complex mixtures such as natural products and crude drugs because of its high peak capacity.

MEKC with an ionic surfactant allows the direct injection of proteinaceous fluids, such as serum. Most of the work on the determination of drugs in biological fluids by MEKC with direct injection has been done so far with SDS as surfactant in the normal elution mode. The proteins are solubilized by SDS and elute (as a very broad band) after uric acid. With human serum or plasma, several drugs including theophylline[84] (see Figure 8a and b) were shown to elute in an interference-free window in front of uric acid, allowing these substances to be analyzed by direct sample injection. It is also possible to analyze drugs in saliva and, in special cases, also in urine. An introduction to this very interesting field of application of MEKC has been given by Thormann.[85]

Another important area of application of MEKC in drug analysis is direct enantiomer separations. These can be achieved either by employing chiral surfactants or with chiral additives to the separation electrolyte such as cyclodextrins in the presence of achiral micelles.[86]

4.3 Analysis of Food and Beverages

In its present state, MEKC is not a routine method in food analysis. The methods in this field that have been developed so far, however, are very promising and show that MEKC can compete favorably with HPLC. The same advantages as outlined for the analysis of drugs by MEKC (high peak capacity, feasible direct injection,

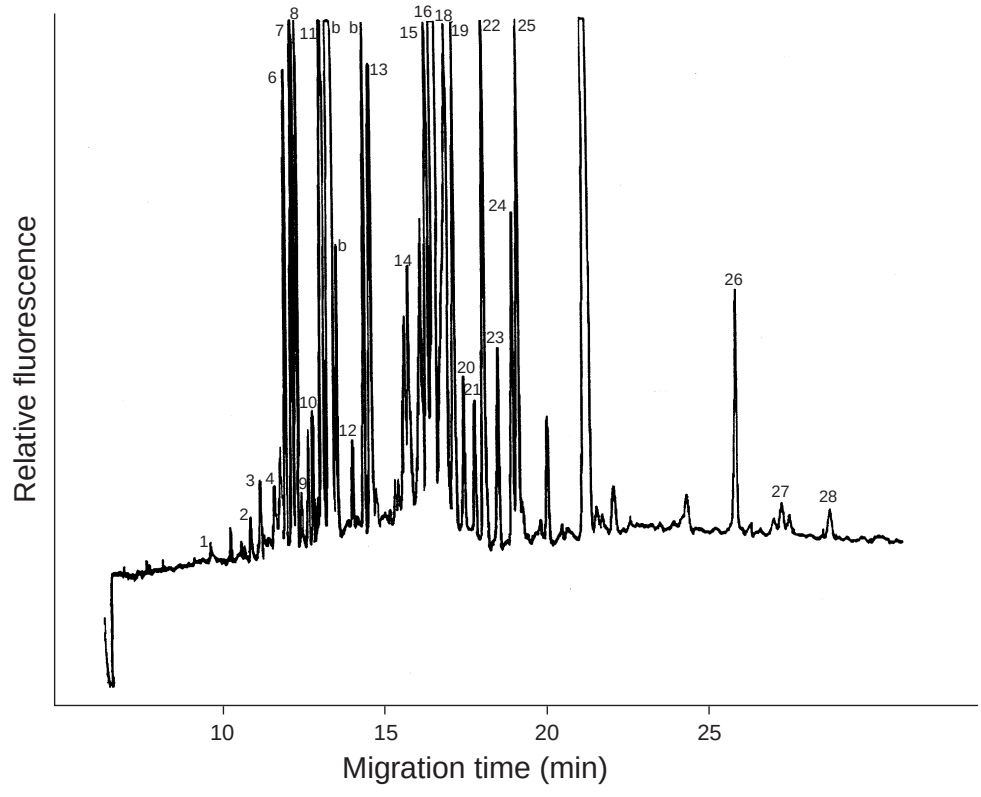

Figure 9 Analysis of a French red wine after derivatization reaction of the sample with fluorescein isothiocyanate. Analysis by MEKC with LIF detection, $\lambda_{ex} = 488$ nm. Composition of separation electrolyte: 100 mmol L^{-1} boric acid, pH = 9.3 adjusted with NaOH, 20 mmol L^{-1} SDS; $U = 20$ kV. Capillary dimensions: 50 µm inner diameter, 750 mm total length. (1) putrescine, (2) arginine, (3) lysine, (4) histamine, (5) ornithine, (6) histidine, (7) ethanolamine, (8) ethylamine, (9) tyramine, (10) β-phenylethylamine, (11) ammonia, (12) spermidine, (13) spermine, (14) threonine, (15) leucine + isoleucine, (16) proline, (17) phenylalanine + valine, (18) tryptophane, (19) asparagine, (20) tyrosine, (21) serine, (22) alanine, (23) cysteine, (24) taurine, (25) glycine, (26) glutamic acid, (27) aspartic acid, (28) cysteic acid, (b) by-product of derivatization. (Peaks 5 and 17 are not identified in this wine.) (Reprinted from G. Nouadje, N. Simeon, F. Dedieu, M. Nertz, Ph. Puig, F. Couderc, 'Determination of Twenty Eight Biogenic Amines and Amino Acids During Wine Aging by Micellar Electrokinetic Chromatography and Laser-induced Fluorescence Detection', *J. Chromatogr. A*, **765**, 337–343 (1997), Copyright 1997, with permission from Elsevier Science.)

short run times) hold true for the analysis of food and beverages.

Nouadje et al.[87] were able to quantitate 28 biogenic amines and amino acids in red wines by MEKC with laser-induced fluorescence (LIF) detection of the fluorescein thiocarbamate derivatives (see Figure 9). Because of the high peak capacity in MEKC no clean-up steps were required. Bjergegaard et al.[88] developed an MEKC method for the determination of naturally occurring phenolic carboxylic acids (cinnamic and benzoic acid derivatives) in samples prepared directly from dietary fibers and other plant materials. According to Bjergegaard et al., the MEKC method developed is highly effective in comparison with HPLC and GC methods. Szücs et al.[81] determined phospholipids in lecithins by MEKC with photometric detection and concluded that the MEKC method developed was preferable over the International Union of Pure and Applied Chemistry (IUPAC) HPLC method.

4.4 Miscellaneous

As most of the applications of MEKC reported so far are with photometric detection, there is a serious limitation on the applicability of MEKC to trace and ultratrace components in environmental samples. In spite of these limitations, the high peak capacity of MEKC methods can have advantages over conventional techniques in the analysis of very complex samples. Song et al.[89] reviewed the application of MEKC on the analysis of environmental pollutants.

In this context, methods to lower the detection limits are highly desirable. The following approaches are suitable for lowering detection limits in MEKC: on-line concentration (see section 3.7), off-line concentration during sample preparation and improving the sensitivity of detection. The sensitivity of detection can be improved by using detection cells with extended optical pathlength or by application of an alternative detection mode offering a better sensitivity than the conventional detection technique.

With suitable off-line concentration, MEKC can be employed for mono- and mulitresidue analysis of herbicides in tap water at the micrograms per liter level.[90] Van Bruijnsvoort et al.[91] developed a method for the determination of chlorophenols by MEKC with amperometric detection. The detection limits obtained were of the order of $10\,\mu g\,L^{-1}$, one to three orders of magnitude lower than with photometric detection. Luong et al.[92] designed an off-column amperometric cell for the determination of explosives after determination by MEKC. Such a detection system offered a 10-fold detection limit improvement over photometric detection for most of the explosives tested. Penmetsa et al.[93] have shown that with proper precolumn derivatization techniques, LIF detection in combination with MEKC can provide detection limits for herbicides as low as $10\,ng\,L^{-1}$. Thermal lens detection can also offer advantages over photometric detection. Seidel et al.[94] employed MEKC for the separation of nitrophenol pesticides and obtained one to two orderes of magnitude lower detection limits with thermal lens detection than with conventional photometric detection.

5 SEPARATION CARRIER CONCEPT

According to the separation carrier concept developed by Terabe,[7] MEKC is a special form of electrokinetic chromatography (EKC) when micelles are the pseudostationary phase. EKC is a technique that uses electrokinetic phenomena to perform a chromatographic separation. Hence in EKC, generally, charged and neutral solutes can be separated. A prerequisite of a chromatographic separation is a two-phase system, with phases of different migration velocity. This prerequisite can be fulfilled by adding a charged substance, that can incorporate solute molecules, to the separation electrolyte while performing CE. This charged substance, which can incorporate solute molecules, is usable as the 'second phase' (the 'first phase' is the surrounding bulk phase) in EKC. This 'second phase', migrating at a velocity different to that of the 'first phase', has been termed a separation carrier.

Not only surfactant micelles but also microemulsions, polymer ions, polymerized surfactants, dendrimers, resocarenes, charged cyclodextrins and other classes of substances, have been used as separation carriers in EKC. An overview of the different modes of EKC based on a classification of separation carriers was given by Nishi and Terabe[83] and Khaledi.[8] Charged cyclodextrins have found numerous applications in enantiomeric separations.[86] The interest in polymeric phases arises from the fact that they have a stable structure even when surrounded by a phase with a high content of organic solvent. The second advantage of these phases is the nonexistence of dissolved monomers at a concentration corresponding to the CMC of the surfactant. In EKC with an ionic surfactant (MEKC), these dissolved monomers increase considerably the specific conductivity of separation electrolytes.

However, considerations made for micelles as a separation carrier cannot always be directly transferred to the above-mentioned separation carriers. A constraint of Equation (2) and subsequent equations is the assumption that the velocity of the separation carrier is not altered by the incorporation of a solute molecule. At the present stage, it is not clear whether this constraint is fulfilled with all the separation carriers listed above.

ACKNOWLEDGMENTS

The help of U. Bütehorn and A. Dworschak in preparing some figures of the manuscript is gratefully acknowledged.

ABBREVIATIONS AND ACRONYMS

Brij® 35	Polyoxyethylene(23)dodecyl Ether
CDMEKC	Cyclodextrin-modified Micellar Electrokinetic Chromatography
CE	Capillary Electrophoresis
CMC	Critical Micelle Concentration
CTAB	Hexadecyltrimethylammonium Bromide
DTAB	Dodecyltrimethylammonium Bromide
EKC	Electrokinetic Chromatography
EOF	Electro-osmotic Flow
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LIF	Laser-induced Fluorescence
MEKC	Micellar Electrokinetic Chromatography
RMMEKC	Reversed-migration Micellar Electrokinetic Chromatography
RPHPLC	Reversed-phase High-performance Liquid Chromatography
SC	Sodium Cholate
SDS	Sodium Dodecyl Sulfate
TTAB	Tetradecyltrimethylammonium Bromide

RELATED ARTICLES

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography

REFERENCES

1. S. Terabe, K. Otsuka, K. Ichikama, A. Tsuchia, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open Tubular Capillaries', *Anal. Chem.*, **56**(1), 111–113 (1984).
2. S. Terabe, K. Otsuka, T. Ando, 'Electrokinetic Chromatography with Micellar Solution and Open-tubular Capillary', *Anal. Chem.*, **57**(4), 834–841 (1985).
3. K. Otsuka, S. Terabe, T. Ando, 'Electrokinetic Chromatography with Micellar Solutions. Retention Behavior and Separation of Chlorinated Phenols', *J. Chromatogr.*, **348**, 39–47 (1985).
4. K. Otsuka, S. Terabe, T. Ando, 'Electrokinetic Chromatography with Micellar Solutions, Separation of Phenylhydantoin-amino Acids', *J. Chromatogr.*, **332**, 219–226 (1985).
5. H. Nishi, N. Tsumagari, S. Terabe, 'Effect of Tetraalkylammonium Salts on Micellar Electrokinetic Chromatography of Ionic Solutes', *Anal. Chem.*, **61**(21), 2434–2439 (1989).
6. J. Vindevogel, P. Sandra, *Introduction to Micellar Electrokinetic Chromatography*, Hüthig, Heidelberg, 1992.
7. S. Terabe, 'Micellar Electrokinetic Chromatography', in *Capillary Electrophoresis Technology*, ed. N.A. Guzman, Chromatographic Science Series, Marcel Dekker, New York, 65–87, Vol. 64, 1993.
8. M.G. Khaledi, 'Micelles as Separation Media in High-performance Liquid Chromatography and High-performance Capillary Electrophoresis: Overview and Perspective', *J. Chromatogr. A*, **780**, 3–40 (1997).
9. P.G. Muijselaar, K. Otsuka, S. Terabe, 'Micelles as Pseudostationary Phases in Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **780**, 41–61 (1997).
10. E.A.G. Aniansson, S.N. Wall, M. Almgren, H. Hoffmann, I. Kielmann, W. Ulbricht, R. Zana, J. Lang, C. Tondre, 'Theory on the Kinetics of Micelle Equilibria and Quantitative Interpretation of Chemical Relaxation Studies of Micellar Solutions of Ionic Surfactants', *J. Phys. Chem.*, **80**(9), 905–919 (1976).
11. N. Chen, S. Terabe, T. Nakagawa, 'Effect of Organic Modifier Concentration on Electrokinetic Migrations in Micellar Electrokinetic Chromatography', *Electrophoresis*, **16**, 1457–1462 (1995).
12. M.M. Bushey, J.W. Jorgenson, 'Separation of Dansylated Methylamine and Dansylated Methyl-D_3-amine by Micellar Electrokinetic Capillary Chromatography with Methanol-modified Mobile Phase', *Anal. Chem.*, **61**(5), 491–493 (1989).
13. S.A. Kuzdzal, J.J. Hagen, C.A. Monning, 'Simultaneous Determination of the Electroosmotic Flow and Pseudo-stationary Phase Migration Time in Electrokinetic Chromatography', *J. High Resolut. Chromatogr.*, **18**(7), 439–442 (1995).
14. M.G. Khaledi, S.C. Smith, J.K. Strasters, 'Micellar Electrokinetic Capillary Chromatography of Acidic Solutes: Migration Behavior and Optimizations Strategies', *Anal. Chem.*, **63**(17), 1820–1830 (1991).
15. P.G. Muijselaar, H.A. Claessens, C.A. Cramers, 'Migration Behavior of Monovalent Weak Acids in Micellar Electrokinetic Chromatography. Mobility Model Versus Retention Model', *J. Chromatogr. A*, **765**, 295–306 (1997).

16. M.T. Ackermans, F.M. Everaerts, J.L. Beckers, 'The Pseudo-effective Mobility as a Parameter for Screening', *J. Chromatogr.*, **585**, 123–128 (1991).
17. K. Ghowsi, J.P. Foley, R.J. Gale, 'Micellar Electrokinetic Capillary Chromatography Theory Based on Electrochemical Parameters: Optimization for Three Modes of Operation', *Anal.Chem.*, **62**(24), 2714–2721 (1990).
18. J.P. Foley, 'Optimization of Micellar Electrokinetic Chromatography', *Anal. Chem.*, **62**(13), 1302–1308 (1990).
19. H. Corstjens, H.A.H. Billiet, J. Frank, K.C.A.M. Luyben, 'Equation for the Description of the Resolution of Charged Solutes in Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **753**, 121–131 (1996).
20. C.X. Zhang, Z.P. Sun, D.K. Ling, 'Micellar Electrokinetic Capillary Chromatography Theory Based on Conventional Chromatography', *J. Chromatogr. A*, **655**, 309–316 (1993).
21. K. Otsuka, M. Higashimori, R. Koike, K. Karuhaka, Y. Okada, S. Terabe, 'Separation of Lipophilic Compounds by Micellar Electrokinetic Chromatography with Organic Modifiers', *Electrophoresis*, **15**, 1280–1283 (1994).
22. U. Bütehorn, U. Pyell, 'Optimization of Resolution in Micellar Electrokinetic Chromatography via Computer-aided Variation of Concentrations of Sodium Dodecyl Sulfate and Acetonitrile as Modifier', *J. Chromatogr. A*, **772**, 27–38 (1997).
23. S. Terabe, Y. Ishihama, H. Nishi, T. Fukuyama, K. Otsuka, 'Effect of Urea Addition in Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **545**, 359–368 (1991).
24. U. Pyell, U. Bütehorn, 'Optimization Strategies in Micellar Electrokinetic Capillary Chromatography. One Parameter Optimizations of the Concentration of Sodium Dodecyl Sulfate and the Concentration of Modifiers (Urea and Glucose)', *Chromatographia*, **40**(3/4), 175–184 (1995).
25. U. Bütehorn, U. Pyell, 'Simultaneous Use of Urea and Acetonitrile as Organic Modifiers for Optimization of Resolution in Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **792**, 157–163 (1997).
26. S. Terabe, Y. Miyashita, O. Shibata, E.R. Barnhart, L.R. Alexander, D.G. Patterson, B.L. Karger, 'Separation of Highly Hydrophobic Compounds by Cyclodextrin-modified Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **516**(1), 23–31 (1990).
27. T. Kaneta, T. Yamashita, T. Imasaka, 'Separation of Polycyclic Aromatic Hydrocarbons by Micellar Electrokinetic Chromatography with Laser Fluorescence Detection', *Anal. Chim. Acta*, **299**(1), 371–375 (1995).
28. S. Terabe, M. Shibata, Y. Miyashita, 'Chiral Separation by Electrokinetic Chromatography with Bile Salt Micelles', *J. Chromatogr.*, **480**, 403–411 (1989).
29. P.G.H.M. Muijselaar, H.A. Claessens, C.A. Cramers, 'Parameters Controlling the Elution Window and Retention Factors in Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **696**, 273–284 (1995).
30. J. Gorse, A.T. Balchunas, D.F. Swaile, M.J. Sepaniak, 'Effects of Organic Mobile Phase Modifiers in Micellar Electrokinetic Capillary Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **11**(8), 554–559 (1988).
31. K. Otsuka, S. Terabe, 'Effect of pH on Electrokinetic Velocities in Micellar Electrokinetic Chromatography', *J. Microcol. Sep.*, **1**(3), 150–154 (1989).
32. H.T. Rasmussen, H.M. McNair, 'Influence of pH on Elution Behavior in Micellar Electrokinetic Capillary Chromatography', *J. High Resolut. Chromatogr.*, **12**(9), 635–636 (1989).
33. G.M. Janini, G.M. Mischik, H.J. Issaq, 'Micellar Electrokinetic Chromatography in Zero-electroosmotic Flow Environment', *J. Chromatogr. B*, **683**, 29–35 (1996).
34. D.W. Fuerstenau, 'Streaming Potential Studies on Quartz in Solutions of Aminium Acetates in Relation to the Formation of Hemimicelles at the Quartz–Solution Interface', *J. Phys. Chem.*, **60**(7), 981–985 (1956).
35. S.K. Poole, C.F. Poole, 'Characterization of Surfactant Selectivity in Micellar Electrokinetic Chromatography', *Analyst*, **122**(3), 267–274 (1997).
36. S. Yang, J.G. Bumgarner, M.G. Khaledi, 'Chemical Selectivity in Micellar Electrokinetic Chromatography. II. Rationalization of Elution Patterns in Different Surfactant Systems', *J. Chromatogr. A*, **738**, 265–274 (1996).
37. D.E. Burton, L.L. Powell, X. Xi, 'Reproducibility of the Determination of Caffeine in Coffee by Micellar Electrokinetic Capillary Chromatography', *J. Microcol. Sep.*, **6**(1), 5–10 (1994).
38. D. Crosby, Z. ElRassi, 'Micellar Electrokinetic Capillary Chromatography with Cationic Surfactants', *J. Liq. Chromatogr.*, **16**(9/10), 2161–2187 (1993).
39. A. Dworschak, U. Pyell, 'Widening of the Elution Window in Micellar Electrokinetic Chromatography with Cationic Surfactants. I. Selection of Surfactant, Variation of pH and Addition of Organic Modifiers or Inorganic Metal Salts', *J. Chromatogr. A*, **848**, 387–400 (1999).
40. A. Dworschak, U. Pyell, 'Widening of the Elution Window in Micellar Electrokinetic Chromatography with Cationic Surfactants. II. Cationic Additives and Modifiers of the Electroosmotic Flow', *J. Chromatogr. A*, **855**, 669–679 (1999).
41. S.K. Poole, C.F. Poole, 'Characterization of Surfactant Selectivity in Micellar Electrokinetic Chromatography', *Analyst*, **122**(3), 267–274 (1997).
42. S. Yang, M.G. Khaledi, 'Chemical Selectivity in Micellar Electrokinetic Chromatography: Characterization of Solute–Micelle Interactions for Classification of Surfactants', *Anal. Chem.*, **67**(3), 499–510 (1995).
43. S. Yang, J.G. Bumgarner, M.G. Khaledi, 'Chemical Selectivity in Micellar Electrokinetic Chromatography II.

Rationalization of Elution Patterns in Different Surfactant Systems', *J. Chromatogr. A*, **738**, 265–274 (1996).
44. C.F. Poole, S.K. Poole, M.H. Abraham, 'Recommendations for the Determination of Selectivity in Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **798**, 207–222 (1998).
45. P.G.H.M. Muijselaar, H.A. Claessens, C.A. Cramers, 'Application of the Retention Index Concept in Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **66**(5), 635–644 (1994).
46. P.G.H.M. Muijselaar, H.A. Claessens, C.A. Cramers, 'Characterization of Pseudostationary Phases in Micellar Electrokinetic Chromatography by Applying Linear Solvation Energy Relationships and Retention Indexes', *Anal. Chem.*, **66**(5), 635–644 (1994).
47. E.L. Little, J.P. Foley, 'Optimization of the Resolution of PTH-amino Acids Through Control of Surfactant Concentration in Micellar Electrokinetic Capillary Chromatography: SDS vs Brij® 35/SDS Micellar Systems', *J. Microcol. Sep.*, **4**(2), 145–154 (1992).
48. S.K. Poole, C.F. Poole, 'Variation of Selectivity with Composition for a Mixed-micellar Buffer in Micellar Electrokinetic Chromatography', *J. High Resolut. Chromatogr.*, **20**(3), 174–178 (1997).
49. H.T. Rasmussen, L.K. Goebel, H.M. McNair, 'Optimization of Resolution in Micellar Electrokinetic Chromatography', *J. High Resolut. Chromatogr.*, **14**(1), 25–28 (1991).
50. S. Terabe, 'Selectivity Manipulation in Micellar Electrokinetic Chromatography', *J. Pharm. Biomed. Anal.*, **10**(10–12), 705–715 (1992).
51. M.G. Khaledi, S.C. Smith, J.K. Strasters, 'Micellar Electrokinetic Capillary Chromatography of Acidic Solutes: Migration Behavior and Optimization Strategies', *Anal. Chem.*, **63**(17), 1820–1830 (1991).
52. J.K. Strasters, M.G. Khaledi, 'Migration Behavior of Cationic Solutes in Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **63**(21), 2503–2508 (1991).
53. T. Kaneta, S. Tanaka, M. Taga, H. Yoshida, 'Migration Behavior of Inorganic Anions in Micellar Electrokinetic Capillary Chromatography Using a Cationic Surfactant', *Anal. Chem.*, **64**(7), 798–801 (1992).
54. S. Terabe, K. Otsuka, T. Ando, 'Band Broadening in Electrokinetic Chromatography with Micellar Solutions and Open-tubular Capillaries', *Anal. Chem.*, **61**(3), 251–260 (1989).
55. U. Pyell, U. Bütehorn, 'Optimization Strategies in Micellar Electrokinetic Capillary Chromatography. Optimization of the Temperature of the Separation Capillary', *Chromatographia*, **40**(1/2), 69–77 (1995).
56. M.J. Sepaniak, R.O. Cole, 'Column Efficiency in Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **59**(3), 472–476 (1987).
57. L. Yu, T.H. Seals, J.M. Davis, 'Reexamination of Dependence of Plate Number on SDS Concentration in Micellar Electrokinetic Chromatography', *Anal. Chem.*, **68**(23), 4270–4280 (1996).
58. J.M. Davis, 'Random-walk Theory of Nonequilibrium Plate Height in Micellar Electrokinetic Capillary Chromatography', *Anal. Chem.*, **61**(21), 2455–2461 (1989).
59. M.G. Khaledi, C. Quang, R. Sahota, J.K. Straasters, S.C. Smith, 'Controlling Migration Behavior in Capillary Electrophoresis: Optimization Strategies for Method Development', in *Capillary Electrophoresis Technology*, ed. N.A. Guzman, Chromatographic Science Series, Marcel Dekker, New York, 187–260, Vol. 64, 1993.
60. H. Corstjens, H.A.H. Billiet, J. Frank, K.C.A.M. Luyben, 'Optimisation of Selectivity in Capillary Electrophoresis with Emphasis on Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **715**, 1–11 (1995).
61. S.C. Smith, M.G. Khaledi, 'Prediction of the Migration Behavior of Organic Acids in Micellar Electrokinetic Chromatography', *J. Chromatogr.*, **632**, 177–184 (1993).
62. C. Quang, J.K. Strasters, M.G. Khaledi, 'Computer-assisted Modeling, Prediction, and Multifactor Optimization in Micellar Electrokinetic Chromatography of Ionizable Compounds', *Anal. Chem.*, **66**(10), 1646–1653 (1994).
63. U. Pyell, U. Bütehorn, 'Optimization of Resolution in Micellar Electrokinetic Chromatography via Computer-aided Simultaneous Variations of Concentrations of Sodium Dodecyl Sulfate and Urea as Modifier', *J. Chromatogr. A*, **716**, 81–95 (1995).
64. J. Vindevogel, P. Sandra, 'Resolution Optimization in Micellar Electrokinetic Chromatography: Use of Plackett–Burman Statistical Design for the Analysis of Testosterone Esters', *Anal. Chem.*, **63**(15), 1530–1536 (1991).
65. M. Castagnola, D.V. Rosetti, L. Cassiano, L. Rabino, G. Nocca, B. Giardina, 'Optimization of Phenylhydantoinamino Acid Separation by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr.*, **638**, 327–334 (1993).
66. Y.F. Yik, S.F.Y. Li, 'Resolution Optimization in Micellar Electrokinetic Chromatography: Use of Overlapping Resolution Mapping Scheme in the Analysis of Dinitrophenyl Derivatized Amino Acids', *Chromatographia*, **35**(9–12), 560–566 (1993).
67. H. Corstjens, A.E.E. Oord, H.A.H. Billiet, J. Frank, K.C.A.M. Luyben, 'Optimization of Selectivity in Capillary Zone Electrophoresis and Micellar Electrokinetic Capillary Chromatography Using the Iterative Regression Strategy', *J. High. Resolut. Chromatogr.*, **18**(9), 551–558 (1995).
68. S.K. Wiedmer, J.H. Jumppanen, H. Haario, M.-L. Riekkola, 'Optimization of Selectivity and Resolution in Micellar Electrokinetic Capillary Chromatography with a Mixed Micellar System of Sodium Dodecyl Sulfate and Sodium Cholate', *Electrophoresis*, **17**(12), 1931–1937 (1996).
69. Z. Liu, P. Sam, S.R. Sirimanne, P.C. McClure, J. Graininger, D.G. Patterson, Jr, 'Field-amplified Sample Stacking

in Micellar Electrokinetic Chromatography for On-column Sample Concentration of Neutral Molecules', *J. Chromatogr. A*, **673**(1), 125–132 (1994).
70. K.R. Nielsen, J.P. Foley, 'Zone Sharpening of Neutral Solutes in Micellar Electrokinetic Chromatography with Electrokinetic Injection', *J. Chromatogr. A*, **686**, 283–291 (1994).
71. J.P. Quirino, S. Terabe, 'On-line Concentration of Neutral Analytes for Micellar Electrokinetic Chromatography. 3. Stacking with Reverse Migrating Micelles', *Anal. Chem.*, **70**(1), 149–157 (1998).
72. J.P. Quirino, S. Terabe, 'On-line Concentration of Neutral Analytes for Micellar Electrokinetic Chromatography. 5. Field-enhanced Sample Injection with Reverse Migrating Micelles', *Anal. Chem.*, **70**(1), 149–157 (1998).
73. J.P. Quirino, S. Terabe, 'Exceeding 5000-Fold Concentration of Dilute Analytes in Micellar Electrokinetic Chromatography', *Science*, **282**(20), 465–468 (1998).
74. J.P. Quirino, S. Terabe, 'Sweeping of Analyte Zones in Electrokinetic Chromatography', *Anal. Chem.*, **71**(8), 1683–1644 (1999).
75. M. Gilges, 'Determination of Impurities in an Acidic Drug Substance by Micellar Electrokinetic Chromatography', *Chromatographia*, **44**(3/4), 191–196 (1997).
76. D.E. Burton, L.L. Powell, X. Xi, 'Reproducibility of the Determination of Caffeine in Coffee by Micellar Electrokinetic Capillary Chromatography', *J. Microcol. Sep.*, **6**(1), 5–10 (1994).
77. T. Watanabe, A. Yamamoto, S. Nagai, S. Terabe, 'Micellar Electrokinetic Chromatography as an Alternative to High-performance Liquid Chromatography for Separation and Determination of Phenolic Compounds in Japanese Spiritous Liquor', *J. Chromatogr. A*, **793**, 409–413 (1998).
78. U. Bütehorn, U. Pyell, 'Micellar Electrokinetic Chromatography as a Screening Method for the Analysis of Vanilla Flavourings and Vanilla Extracts', *J. Chromatogr. A*, **736**, 321–332 (1996).
79. W. Kleiböhmer, K. Camman, J. Robert, E. Mussenbrock, 'Determination of Explosives Residues in Soils by Micellar Electrokinetic Capillary Chromatography and High-performance Liquid Chromatography. A Comparative Study', *J. Chromatogr.*, **638**(2), 349–356 (1993).
80. H.J. Crabtree, I.D. Ireland, N.J. Dovichi, 'Effect of Acetonitrile in the Sampling Solution on the Analyte Peak Shape in Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **669**, 263–267 (1994).
81. R. Szücs, K. Verleysen, G.S.M.J.E. Duchateau, P. Sandra, B.G.M. Vandeginste, 'Analysis of Phospholipids in Lecithins. Comparison Between Micellar Electrokinetic Chromatography and High-performance Liquid Chromatography', *J. Chromatogr. A*, **738**, 25–29 (1996).
82. A. Farran, S. Ruiz, C. Serra, M. Aguilar, 'Comparative Study of High-performance Liquid Chromatography and Micellar Electrokinetic Capillary Chromatography Applied to the Analysis of Different Mixtures of Pesticides', *J. Chromatogr. A*, **737**, 109–116 (1996).
83. H. Nishi, S. Terabe, 'Micellar Electrokinetic Chromatography. Perspectives in Drug Analysis', *J. Chromatogr. A*, **735**, 3–27 (1996).
84. W. Thormann, A. Minger, S. Molteni, J. Caslavska, P. Gebauer, 'Determination of Substituted Purines in Body Fluids by Micellar Electrokinetic Capillary Chromatography with Direct Sample Injection', *J. Chromatogr.*, **593**, 275–288 (1992).
85. W. Thormann, 'Capillary Electrophoresis and Electrokinetic Capillary Chromatography of Drugs in Body Fluids', in *Capillary Electrophoresis Technology*, ed. N.A. Guzman, Chromatographic Science Series, Marcel Dekker, New York, 693–704, Vol. 64, 1993.
86. S. Terabe, K. Otsuka, H. Nishi, 'Separation of Enantiomers by Capillary Electrophoretic Techniques', *J. Chromatogr. A*, **666**, 295–319 (1994).
87. G. Nouadje, N. Simeon, F. Dedieu, M. Nertz, Ph. Puig, F. Couderc, 'Determination of Twenty Eight Biogenic Amines and Amino Acids During Wine Aging by Micellar Electrokinetic Chromatography and Laser-induced Fluorescence Detection', *J. Chromatogr. A*, **765**, 337–343 (1997).
88. C. Bjergegaard, S. Michaelsen, H. Sørensen, 'Determination of Phenolic Carboxylic Acids by Micellar Electrokinetic Capillary Chromatography and Evaluation of Factors Affecting the Method', *J. Chromatogr.*, **608**, 403–411 (1992).
89. L. Song, Z. Xu, J. Kang, J. Cheng, 'Analysis of Environmental Pollutants by Capillary Electrophoresis with Emphasis on Micellar Electrokinetic Chromatography', *J. Chromatogr. A*, **780**, 297–328 (1997).
90. G. Dinelli, A. Bonetti, P. Catizone, G.C. Galletti, 'Separation and Detection of Herbicides in Water by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. B*, **656**, 275–280 (1994).
91. M. van Bruijnsvoort, S.K. Sanghi, H. Poppe, W.Th. Kok, 'Determination of Chlorophenols by Micellar Electrokinetic Chromatography with Electrochemical Detection', *J. Chromatogr. A*, **757**, 203–213 (1997).
92. J.H.T. Luong, A. Hilmi, A.L. Nguyen, 'Development of Electrokinetic Capillary Electrophoresis Equipped with Amperometric Detection for Analysis of Explosive Compounds', Poster Presented at the 23rd International Symposium on High Performance Liquid Phase Separations and Related Techniques, Granada, Spain, May–June, 1999.
93. K.V. Penmetsa, R.B. Leidy, D. Shea, 'Herbicide Analysis by Micellar Electrokinetic Capillary Chromatography', *J. Chromatogr. A*, **745**, 201–208 (1996).
94. B.S. Seidel, W. Faubel, H.J. Ache, 'Thermal Lens Detection System for Capillary Electrophoresis', *J. Biomed. Opt.*, **2**(3), 326–331 (1997).

Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography

Patrick T. Vallano, Gabriela S. Chirica, and Vincent T. Remcho
Oregon State University, Corvallis, USA

1 Introduction 1
2 History 2
2.1 Evolution of Microscale Liquid Chromatography 2
2.2 Capillary Electrophoresis and Its Influence on Microscale Liquid Chromatography 5
2.3 Crossroads of Capillary Electrophoresis and Liquid Chromatography: Micellar Electrokinetic Capillary Chromatography and Capillary Electrochromatography 6
3 Theory 6
3.1 Efficiency and Its Effect on Resolution in Chromatography 6
3.2 Advantages of Column Miniaturization 8
3.3 Extra-column Effects 10
3.4 Flow Profile and Efficiency: Advantages of Electro-osmotic Flow 11
4 Instrumentation 13
4.1 Columns for Microscale Liquid Phase Separations 13
4.2 System Components: Microscale Liquid Chromatography 15
4.3 System Components: Capillary Electrochromatography 16
4.4 Mass Spectrometry Coupling 17
5 Optimization of Analysis 18
5.1 Selectivity Optimization 18
5.2 Gradient Separation and Temperature Programming 19
5.3 Ultra High-pressure Micro Liquid Chromatography 19
6 Applications 19
6.1 Micro Liquid Chromatography 19
6.2 Capillary Electrochromatography 20
7 Future Perspectives 21
Abbreviations and Acronyms 22
Related Articles 23
References 23

Miniaturization is a general trend in analytical chemistry. In the field of separation science, this trend has led to the development of microscale techniques based on high-performance liquid chromatography (HPLC) and electrophoresis. Two such techniques, microscale high-performance liquid chromatography (or micro LC) and capillary electrochromatography (CEC) are the subjects of this article. Micro LC is a general term that encompasses HPLC separations in which the column diameter is less than 1.0 mm. CEC is a hybrid of micro LC and capillary electrophoresis (CE). Relative to the conventional scale technique, micro LC possesses several advantages, namely higher efficiency, often shorter analysis times, reduced solvent consumption and higher mass sensitivity. Micro LC is uniquely suited for analysis of very small sample volumes (e.g. biological cells). It is this type of analysis, a formidable challenge, which has propelled the technique's continued development. CEC most commonly employs conventional HPLC particles packed into fused silica capillary columns. A potential gradient is applied along the capillary; the resulting electro-osmotic flow (EOF) then propels the mobile phase (and analytes) through the column. In CEC, the electric field (generated by a power supply) takes the place of pressure (generated by a pump) in an HPLC separation. One of the most important advantages of CEC relative to micro LC or CE is enhanced resolution. High resolution in CEC arises from the combination of high separation efficiency and selectivity inherent in the technique. CEC, however, is a relatively new technique and its ability to become a ''workhorse'' in the laboratory has not yet been demonstrated.

1 INTRODUCTION

HPLC is one of the most popular and widely applied analytical techniques in use today. Among the many reasons for the technique's popularity are its versatility and wide range of applicability. In its most common form, the modern analytical scale HPLC column has an internal diameter (ID) of 4.6 mm. However, the trend toward miniaturization common throughout analytical chemistry has led to the continued development of smaller diameter columns for HPLC. Today, HPLC in columns having diameters of less than 1.0 mm has evolved into a closely related family of techniques, collectively referred to as *micro LC*. These include open tubular (OT) LC (liquid

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

chromatography), packed capillary LC and microbore LC. Since the pioneering work of several research groups in the late 1960s and early 1970s, micro LC has grown into an important analytical technique offering several benefits over its conventional scale counterpart. Higher separation efficiencies, decreased analysis time, reduced solvent consumption and increased mass sensitivity are some of the advantages that have propelled the growth of this technique.

Owing to its high resolving power and quantitative capability, microscale LC is ideally suited to analyzing complex, volume-limited samples, for example the contents of individual biological cells. One limitation of micro LC imposed by the small column dimensions, however, is reduced sample loading capacity, which may result in decreased sensitivity relative to conventional LC.

CEC can be viewed as an outgrowth of micro LC. A relatively youthful technique, CEC represents a marriage of packed capillary LC and CE, a separation technique that itself has grown markedly since the early 1980s. In its most common form, CEC employs columns identical to those used in packed capillary LC (i.e. 50–250 μm ID fused silica capillaries packed with reversed-phase, silica-based particles). However, rather than using a pump to drive the mobile phase, bulk flow arises in CEC from electroosmosis, which is induced by an electric field applied across the capillary. For neutral species, the separation mechanism is identical to that in reversed-phase HPLC (i.e. partitioning between the mobile and stationary phases). Charged species, however, will be subject to electrophoresis as well as partitioning, providing a dual separation mechanism that may be employed to increase, or at least alter, selectivity. EOF is intrinsically advantageous for mobile phase propulsion in chromatography owing to its favorable flow dynamics, which result in decreased peak broadening relative to pressure-driven flow. This leads to much higher separation efficiencies in CEC than are obtainable in capillary LC. Higher efficiencies translate into greater resolving power. This combined with the advantages afforded by the small column dimensions (e.g. increased mass sensitivity, short analysis times, low sample volume requirement) make CEC an extremely powerful separation technique.

In contrast to packed capillary LC, the high efficiency of CEC affords high peak capacity, which is useful in analyzing mixtures containing many components. For this reason, CEC is especially suitable for analyzing highly complex samples, such as those involving biological or environmental matrices. As might be expected with a newly developed technique, a disadvantage of CEC is reproducibility.

2 HISTORY

2.1 Evolution of Microscale Liquid Chromatography

The development timeline of micro LC parallels that of ''conventional'' HPLC. Both were challenged and inspired by impressive achievements in the field of gas chromatography (GC). Technological developments in electronics and materials science, coupled with developments in theoretical descriptions of basic chromatographic processes, have driven both techniques toward increasingly better performance.

2.1.1 Semimicro High-performance Liquid Chromatography

The germ of HPLC was realized in 1941, when Martin and Synge published a classic paper in which they described liquid–liquid partition chromatography.[1] They stated that in order to obtain fast separations, small particles and high pressures are required. Their suggestion was realized only in the late 1960s, after the pioneering work of Hamilton et al.[2] in 1960 and Huber et al.[3] in 1967. Most successes of GC such as pellicular and porous packing, temperature programming and especially the use of capillary columns (microbore columns in the present nomenclature) found immediate use in LC. Note that the term microbore as adopted here refers to columns of 1 mm ID. Several names have in the past been used for these columns, such as narrow bore, microbore and small-bore.[4] The various sizes and nomenclature of HPLC columns are presented in the flow chart depicted in Figure 1.

A number of papers published in the late 1960s considered the use of high-pressure LC in the microscale format, primarily with columns having IDs of 1 and 2 mm.[5] Horvath et al.[6] published the first microbore HPLC chromatogram in 1967, which illustrated the use of pellicular packing for the separation of ribonucleoside–monophosphoric acids in about 30 min.

In 1969 Kirkland[7] reiterated the feasibility of the technique by separating urea herbicides in less than 6 min with a 2.0 mm ID column. In 1979 Scott and Kucera[8] showed experimentally that very high theoretical plate numbers (750 000) could be achieved with a microbore column. Five years later, Menet, Gareil and Rosset[9] announced that efficiencies as high as one million plates could be obtained by using a 22 m long microbore column operated at 800 bar inlet pressure.

The ''high speed'' expectation has already been met. Likewise, other advantages of microbore columns such as good resolution, minimal sample volume and low solvent consumption have been realized. In fact, for a while the

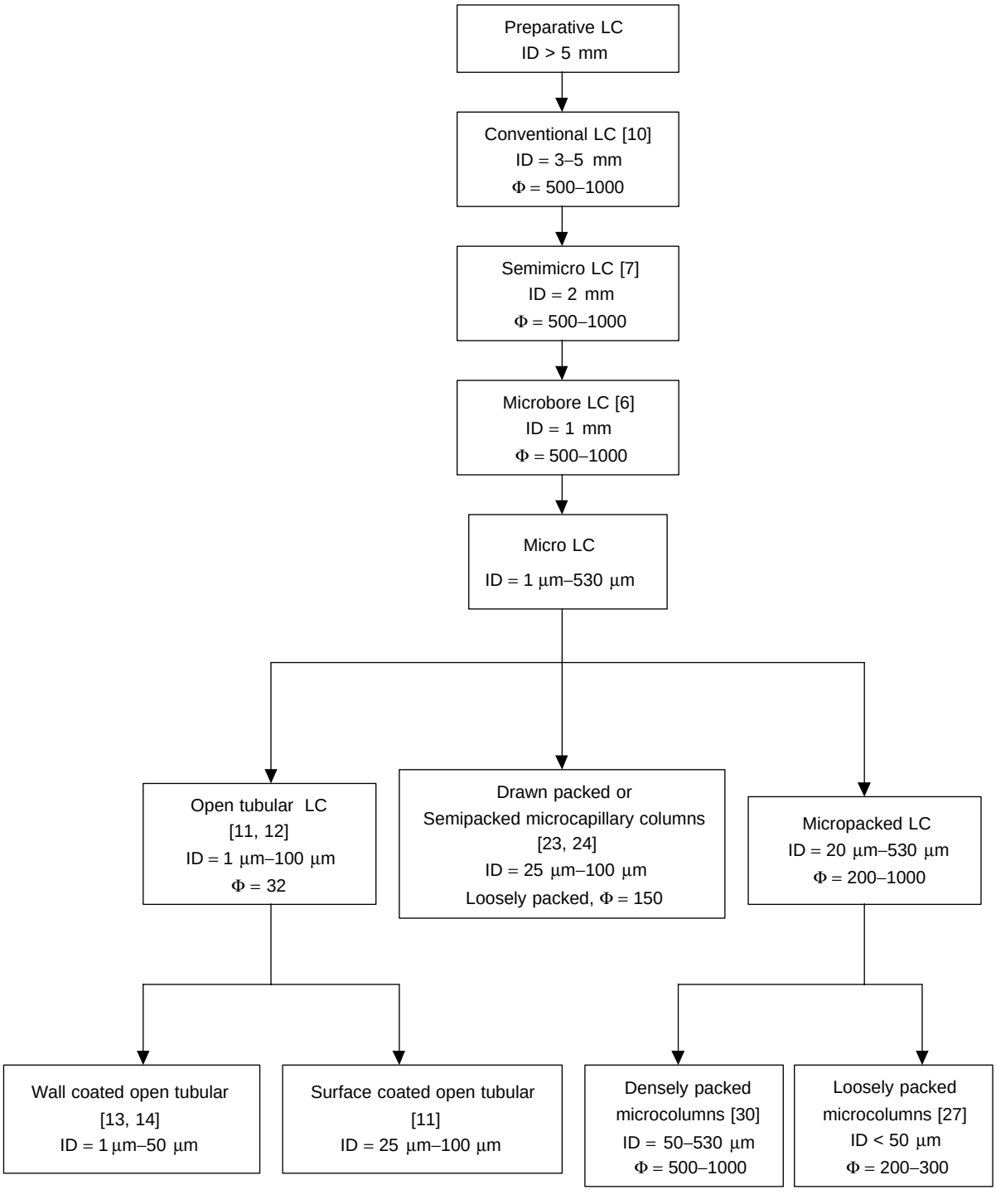

Figure 1 Classification of HPLC columns according to ID.

term high-pressure or high-speed LC was synonymous with microbore LC.

In 1970 DeStefano and Beachell[10] used columns with larger diameters (8–11 mm) and achieved efficiencies that at the time were better than those obtained with microcolumns. The improvement is caused by the so-called ''infinite diameter effect'' that occurs when a sample is injected centrally into a wide column. The result is a decrease in band broadening since the ''wall effects'' due to permeability inhomogeneity across the column are diminished. Since then wider bore columns have been the preference and have come to be the convention in HPLC.

2.1.2 *Open Tubular High-performance Liquid Chromatography*

The first LC separation in an OT column was obtained in 1970.[11] A soft glass capillary of 0.23–0.30 mm ID was treated with alkaline solution to generate a silica gel layer and then used to separate dansyl amino acids. Because of the large column diameter, the resolution was poor and the separation required 5 h. Tsuda and Novotny[12] published in 1978 their analysis of band broadening phenomena for unretained species indicating the potential for high-resolution separations with open tubular liquid chromatography (OTLC). In 1978 Ishii applied the results of earlier theoretical studies, which indicated that smaller IDs provide higher efficiencies, and used a microcapillary tube of 60 μm ID. He demonstrated that both physically coated[13] and chemically bonded[14] stationary phases were able to resolve the components of a mixture by LC.

The high efficiency attainable in OTLC was demonstrated by the achievement of 2 800 000 theoretical plates with a 27.5 m long column of 32 μm ID.[15]

For OT capillary columns, large numbers of theoretical plates (>250 000) are not unusual in GC and over the years it has been proven that efficiencies such as this could also be achieved in LC. OT columns have the highest permeability of all column formats, allowing them to produce larger numbers of theoretical plates per unit time and unit pressure drop than packed columns (PCs).

Skill and experience are necessary to obtain OT columns with a uniform stationary phase layer. Covalent bonding of the stationary phase on the walls of the column offered a more uniform and stable coating, which made solvent gradient elution and temperature programming possible. On the other hand, the covalently bonded stationary phase consists of a monomolecular layer having inherently low-analyte capacity; unfortunately OT columns usually require a high-loading capacity to achieve adequate detection limits (DLs).

To achieve maximum efficiency in OTLC requires the use of narrow ID columns, typically 10 μm and below. Peak volumes generated in columns of this size are on the order of 0.1 nL and were incompatible with the detectors available at that time[16] owing to excessive extra-column peak broadening. Unfortunately, this issue limited the use of such columns.

Approaches to increase the sample capacity of the column involved the use of support materials coated on the walls of the capillaries. Supports such as silica gel, porous silica microspheres or organic polymers (polystyrene, polyacrylates)[17] were studied.

The development of fused silica capillary columns in 1979 by Dandeneau and Zerenner[18] revolutionized miniaturized separation techniques. As opposed to the fragile glass columns and surface inhomogeneities of metal columns, the surface of fused silica is smooth, more inert, and the outside polyimide coating makes it flexible and much easier to work with. Before long, fused silica became the material of choice in both the OT and packed capillary format.

The first to investigate LC in fused silica capillaries was Frank Yang, who derivatized columns of different IDs with octadecylsiloxane. The best results were obtained with the 30 μm ID capillary.[19] In the same paper he introduced in situ or on-column detection, which could easily be performed after removing the polyimide coating from a small section of the capillary (2–10 mm). This design effectively eliminated any extra-column volume contribution from the detection cell.

The challenge posed by microbiology on the practical limits of microanalysis drove forward the development of detection systems, such that even smaller ID columns might be used. Their manufacture was not an easy task. In 1988 fused silica capillaries of 2 μm ID were obtained,[20] but the high purity intrinsic to fused silica makes it extremely difficult to increase its surface area by etching in order to enhance column loadability. Therefore, materials used early in the development of LC, such as soda lime glass or borosilicate glass, were etched with acidic and basic solutions to give higher inner surface areas.[21,22] Subsequent chemical bonding of the stationary phases provided the desired selectivity of the OT column.

The requirement of highly specialized instrumentation derived from the limited column loadability narrows the applicability of OTLC.

2.1.3 *Drawn-packed Capillaries*

An interesting alternative solution to the small loadability of OT columns and high-flow resistance parameters of PCs is offered by drawn-packed microcapillary columns, also known as semi-packed microcapillary columns.

These columns were widely used in GC. In 1978 Tsuda and Novotny achieved the first liquid-phase separations in drawn-PCs of 50–200 μm ID, ten times smaller then those currently used in GC.[23]

The manufacturing process consists of packing 1–2 mm ID tubing with sorbent material and subsequently drawing it to the desired diameter using a glass drawing machine. The application of high temperatures during the drawing process results in the impression of some packing material on the walls of the tubing conferring the ''loose'' packing structure specific to this type of column. Unfortunately, such high temperatures preclude the use of derivatized packing materials. Nevertheless, it is possible to prepare various chemically bonded stationary phases in situ, by passing the silane reagent through the capillary.[24] Efficient separations were obtained with a 42 μm ID column of 29.4 m length.[25] Theoretical plate

numbers varied from 400 000 to 880 000 but the analysis required more than 3 h.

2.1.4 Micropacked Capillaries

The low sample capacity of OT columns encouraged scientists to return to the study of PCs for LC. Interestingly, the packed capillary column format actually minimizes the previously described ''wall effects''. For a certain particle diameter, as the walls of the capillary are closer together the overall permeability across the column is more homogeneous.[26,27] This results in a decreased range of flow velocities, smaller pressure drop and more effective dissipation of the frictional heat generated in the column. Thus, in contrast to conventional columns, micropacked capillaries facilitate the use of longer columns, resulting in higher theoretical plate numbers. However, some of the problems encountered with narrow bore columns (such as extra column effects and difficulties in obtaining uniform packing) had to be addressed.

In 1979 Scott and Kucera introduced a low-dead-volume detection cell of 1 μL which significantly improved detection sensitivity.[8] Numerous other geometries were proposed, but on-column detection remains the most popular for fused silica capillaries in spite of the path length limitation it imposes.

In 1969 Knox and Parcher studied the interrelationship between column ID, particle size and plate height.[28] The research, continued by Wilson and McNair[29] for capillary columns, indicated that higher efficiencies can be obtained for Knox–Parcher ratios ρ, defined as the column diameter divided by the diameter of the packing material, of less than five because they generate a more ordered packing structure.

In addition to the general trend toward smaller diameter particles, the design of the packing material itself was continually improved with respect to uniformity in particle and pore sizes. From the small glass spheres coated with a layer of silica gel or microspheres of silica,[7] to the present day material (porous homogeneous covalently bonded microparticles of 2–10 μm), the evolution of packing materials has led to increased column efficiencies. However, conventional (4.6 mm ID) columns packed with the smaller size particles develop an extremely large pressure drop. Capillary columns present one practical solution to this problem. Efficiencies of 100 000 theoretical plates can routinely be attained when columns of 1 m length containing 3 or 5-μm particles are used.[30]

Column packing was extensively studied and methods such as balanced density slurry and vacuum pulled slurry were commonly used.[4] In the first procedure the packing material was mixed with a solvent of similar density (tetrachloromethane), such that the slurry obtained was less susceptible to aggregate formation during packing. High pressures (700–1000 bar) were used to force the slurry into the capillary, the result being a homogeneous, well-packed chromatographic bed. For glass, PTFE (polytetrafluoroethylene) and other tubing materials that could not withstand such high pressures, the packing method proposed by Ishii,[31] which used vacuum to pull the slurry into the capillary, was appropriate. After packing, a compression step was required to compact the bed.

The flexibility of fused silica and availability of polymeric ferrules that permit high-pressure operation greatly facilitated the development of improved packing techniques. Novotny et al. modified the slurry packing method by using higher density slurries, which were sonicated during packing to maintain homogeneity. A less viscous solvent (acetonitrile) was used as both slurry and displacement fluid.[30] The use of acetonitrile also decreased the time required to pack and depressurize the column. It must be stressed that the optimum slurry and/or displacement solvent may vary with packing materials which have different surface chemistries.

Considering the performance of micropacked capillaries coupled with the versatility of the packing materials already available for conventional LC, it is widely considered that the micropacked capillary format of micro LC has the greatest potential for continued development and application.

2.2 Capillary Electrophoresis and Its Influence on Microscale Liquid Chromatography

As early as 1877, Helmholz studied the effect of the application of an electric field across a glass capillary filled with a salt solution.[32] He determined that the wall of the capillary acquired a negative charge, and that while voltage was applied, positively charged particles associated with the wall moved toward the cathode. This nondiscriminatory transport process was termed electroosmosis or electroendosmosis.

The movement of charged particles in solutions under the influence of an electric field, the other electric-field driven transport mechanism, called electrophoresis, was intensively studied by electrochemists. Not until 1937 did electrophoretic migration catch the interest of a separation scientist. Tiselius exploited the differences in electrophoretic mobilities of proteins to separate them, and UV (ultraviolet) photography to detect them.[33] In these studies, it soon became obvious that resistive heating of the solution caused severe band broadening and the appended cooling systems could not effectively eliminate the problem. In spite of this limitation, several separation techniques, including free zone electrophoresis, isotachophoresis, isoelectric focusing, all of which are

based on electrophoretic transport, have become important analytical tools, especially in the fields of biology and biochemistry.

Column miniaturization became fashionable in the early 1960s and with good reason. In 1970 Hjerten[34] introduced surface coating of the walls of small capillaries to minimize electroosmosis, which resulted in essentially pure electrophoretic migration of species. The benefits introduced by column downsizing, such as increased heat dissipation, sensitive real-time detection, applicability to samples in limited amount and ease of automation were notable. The significant increase in efficiency revealed in 1981 by Jorgenson and Lukacs[35] provided a major breakthrough in the field of electrophoresis.

A term was required for the new technique and CE was an obvious choice. The ever-increasing number of practical applications that span from small organic and inorganic molecules, to bio and synthetic polymers, drugs, food additives, and so on, established the name ''high-performance CE''.

Before long, several other variations of CE were developed. Some of the methods that stemmed from CE adopted the electric-field driven migration of the sample and solvents as a bulk transport mechanism. This is the point at which CE, which owed its rapid progress to the ready application of available chromatography advances, began to reward its predecessor.

2.3 Crossroads of Capillary Electrophoresis and Liquid Chromatography: Micellar Electrokinetic Capillary Chromatography and Capillary Electrochromatography

In 1984 Terabe et al. proposed an interesting separation technique which he referred to as micellar electrokinetic capillary chromatography (MEKC).[36] In this technique an anionic surfactant (SDS (sodium dodecyl sulfate) is widely used) is added to the buffer of the CE system at a concentration higher than its critical micelle concentration, such that it forms transient micelles. During a run the neutral components of the sample move towards the cathode carried by the EOF. Given the negative charge of the micelles, the direction of electrophoretic migration is toward the anode. However, in uncoated capillaries the magnitude of the EOF vector is usually larger than the electrophoretic component of the micelle. Therefore, the micelles migrate toward the cathode, although at a lower speed than the bulk solution. If an analyte partitions into the micelle, it moves during the period of association at the micelle velocity. Different affinities for the micelles, hence different partition coefficients into the ''pseudostationary phase'' allow for the separation to occur. Thus, MEKC is a chromatographic technique because the separation mechanism is based on distribution equilibria.

Although the invention of MEKC greatly widened the applicability of CE, it did not broaden the technique to the point of universal utility. Analytes that partition into the micelles are sparingly soluble in water and only small amounts of organic modifiers could be added to enhance solubility without perturbing the formation of micelles. Techniques that allow the use of high concentrations of organic solvents would perhaps be more versatile.

In 1974 Pretorius et al. studied electroosmosis in packed glass tubes and observed that irregularity in the packed bed did not affect the chromatographic performance.[37] High efficiencies were obtained for unretained species, indicating the potential of the technique. In 1981, Jorgenson and Lukacs replaced the high-pressure pump of micro LC with EOF in 170-μm ID, 68-cm long capillaries packed with 10-μm reversed-phase particles.[35] The column generated over 20 000 theoretical plates corresponding to reduced plate heights of 1.9 and 2.5. At the time, the typical optimum reduced plate height for PCs was 2.5, so these values demonstrate the potential for ultrahigh efficiencies for electro-driven separation techniques. This application of electroosmosis for bulk transport in micro LC would later be called CEC.

In 1988, Knox described the theory of CEC operation,[38] but the method was not widely researched before a new paper was published in 1991.[39] There, Knox employed both slurry packed capillaries and drawn-packed capillaries in a comprehensive study of the effect of particle size on EOF velocity. He found that particle size had essentially no effect on velocity and electrolyte concentration as low as 0.002 M resulted in a relatively fast analysis with high efficiency. An impressive 500 000 plates m^{-1} and reduced plate heights of 1.3 were obtained for the slurry PC.

3 THEORY

3.1 Efficiency and Its Effect on Resolution in Chromatography

In LC, separation is achieved by the differential migration of solutes through the column, which arises due to the solutes' partitioning behavior between the stationary phase and flowing solvent (mobile phase). The quantitative measure of the separation of any two components is resolution (R_s). Resolution is an important chromatographic feature of merit because it is a measure of the ability of the system to distinguish between closely spaced components in a separation. The goal in chromatography is to obtain adequate resolution of the components of interest in a minimum amount of time.

Separations in which many components must be resolved are often the most challenging. This is typically the case when the species to be analyzed are present in a complex matrix. To obtain adequate separations with these samples generally requires high resolution. To meet these challenges, there is an ever-present need to seek ways in which chromatographic resolution can be maximized while at the same time minimizing analysis time.

The master resolution equation expresses resolution as a function of three basic chromatographic parameters: capacity factor, selectivity and efficiency

$$R_s = \frac{1}{4}\left(\frac{k'}{1+k'}\right)\left(\frac{\alpha-1}{\alpha}\right)\sqrt{N} \qquad (1)$$

The first term in Equation (1) is capacity factor (k') dependent. The capacity factor for a given solute is a measure of its affinity for the stationary phase (it is proportional to the solute's thermodynamic partition ratio). If all other variables are held constant, a solute with a lower value of k' will migrate through the column faster (i.e. have a shorter retention time) than a solute with a greater k'.

The second term in the equation is selectivity, or separation factor, (α) dependent. Selectivity, which is measured for a pair of peaks, is simply the ratio of their capacity factors. Selectivity values are by definition greater than or equal to one and indicate differences in the affinities of analytes for the stationary phase. Capacity factor and selectivity are dependent upon the thermodynamics of the solute–stationary phase interaction.

The third term in the master resolution equation is efficiency (N) dependent. Efficiency is related to the narrowness of the chromatographic peak and is expressed as the number of theoretical plates. Unlike capacity factor and selectivity, efficiency is chiefly affected by physical, rather than chemical processes that occur as the solute passes through the column. These processes include kinetic phenomena such as molecular diffusion, slow mass transfer and others that result in broadening of peaks as they migrate through the column. Efficiency is quantified by Equation (2):

$$N = \frac{t_r^2}{\sigma^2} \cong 16\left(\frac{t_r}{W_b}\right)^2 \qquad (2)$$

where t_r is the retention time of the peak, σ^2 is the peak variance (the square of the standard deviation σ) in units of time and W_b is the peak width at baseline in time units.

From Equations (1) and (2) it is evident that peak broadening degrades resolution. It is desirable, then, to minimize peak broadening by understanding and minimizing (as much as it is possible) its causes.

A more useful parameter with which to assess chromatographic peak dispersion is the height equivalent to a theoretical plate or plate height (H), which is given by Equation (3):

$$H = \frac{\sigma^2}{L} \equiv \frac{L}{N} \qquad (3)$$

where σ^2 is the peak variance and L is the column length. Plate height is a measure of peak broadening per unit column length. Assuming Gaussian peaks, which under most conditions is a reasonably valid assumption, plate height may be viewed as the length of column that "contains" one theoretical plate (Equation 3).

Greater values of H correspond to broader peaks. Because plate height is a measure of peak broadening per unit length of column, it is often a more useful feature of merit than N in that it allows direct comparisons between columns of different lengths. In PC chromatography, plate height is often expressed in its reduced form, h, ($h = H/d_p$) in which it is normalized to the packing particle diameter (d_p).

Let us now examine the phenomena that contribute to H. According to chromatographic rate theory, peak broadening arises from various kinetic factors that occur as the solute zone passes through the column. The following well-known relationship, which derives from chromatographic rate theory, expresses plate height as the sum of individual contributions, Equation (4):

$$H = A + \frac{B}{\mu} + C_s\mu + C_m\mu \qquad (4)$$

In this general form of the equation, the A term represents the plate height contribution due to eddy diffusion and the B term represents axial molecular diffusion. The contributions of slow solute mass transfer in the stationary and mobile phases are contained in the C_s and C_m terms, respectively. Mobile phase linear velocity is given by μ. It is common in chromatography to depict the dependence of H on μ graphically in the form of a Van Deemter plot.

The eddy diffusion or A term in Equation (5) arises primarily from the multiple pathways in a packed bed through which solute molecules can flow. A molecule traveling a less tortuous path will travel faster relative to the average condition. Conversely, a molecule traveling in a more difficult path will fall behind with respect to the average. The resultant chromatographic peak, a distribution of arrival times about a mean value (the peak apex), becomes broadened. It has been shown that the magnitude of the A term is proportional to the diameter of the packing particle:

$$A \propto d_p \qquad (5)$$

Axial, or longitudinal molecular diffusion, represented by the B term, is the plate height contribution caused

by diffusion of solute molecules along the column axis. The B coefficient is proportional to the solute's diffusion coefficient in the mobile phase. The slower the velocity of the solute zone through the column, the more time is available for diffusion to occur, thus there is an inverse relationship between mobile phase velocity and the magnitude of this contribution. In liquids, the diffusion coefficients of solutes are small (typically on the order of $10^{-5}\,cm^2\,s^{-1}$). For this reason, the effect of axial diffusion on plate height is minimal under typical operating conditions.

A finite amount of time is required for solute molecules to enter and leave the stationary phase. The rate of this exchange has an inverse effect on plate height (slower exchange results in broader peaks). The C_s term of the plate height equation represents this contribution to plate height. As might be expected, the value of C_s is related to the thickness of the stationary phase layer on the particle surface. In modern LC, the stationary phase film is typically formed as a monolayer on the particle surface. Because of this, slow mass transfer in the stationary phase is usually a minor contributor to plate height.

The contribution to plate height of slow mass transfer kinetics in the mobile phase is included in the C_m term. In even the most well packed LC column, point to point differences in mobile phase velocity exist. For example, velocity differences arise from the intrinsic nature of laminar (pressure-driven) flow. Consider for instance that in any microscopic flow channel within the packed bed (e.g. in the packing interstices), the flow velocity is at some maximum value at the channel center and zero at the particle surface owing to friction. Between these extremes, flow velocity varies as a function of distance from the center. Molecules near the center of the channel will outdistance those near the particle surface, broadening the zone. Diffusion of solute molecules into and out of the various velocity regimes serves to counteract the zone broadening effect. However, as discussed above, diffusion coefficients in liquids are small and so this velocity nonhomogeneity effect contributes to plate height. Slow solute diffusion in the mobile phase results in additional zone broadening effects when porous packings are employed. Stagnant mobile phase mass transfer involves pools of nonflowing mobile phase deep inside the pores of the packing material. Solute molecules that diffuse into these pools fall behind with respect to the average and the peak broadens. Because diffusion is the sole mechanism by which solute molecules enter and leave the pools, the slow diffusion rates of solutes in liquids exacerbate zone broadening due to this process.

Not surprisingly, the C_m term scales in inverse proportion to the solute's diffusion coefficient in the mobile phase (D_m). The proportionality in Equation (6) shows the dependence of C_m on particle size and the diffusion coefficient:

$$C_m \propto \frac{d_p^2}{D_m} \tag{6}$$

3.2 Advantages of Column Miniaturization

The ongoing evolution of HPLC has led to the development of ever-smaller column diameters (<1.0 mm) since the mid 1980s. The impetus for this miniaturization has been the numerous advantages afforded by miniaturized columns. These advantages include:

- improved column permeability, resulting in enhanced efficiency
- decreased solvent consumption
- increased mass sensitivity
- more efficient heat transfer
- less packing material required
- more facile interfacing with mass spectrometry (MS).

Let us now examine each of these advantages in greater detail.

3.2.1 Column Permeability

The permeability of a chromatographic column is a measure of its resistance to flow and is related to the dimensionless flow resistance parameter ϕ, Equation (7):

$$\phi = \frac{\Delta P d_p^2}{\mu L \eta} \tag{7}$$

where ΔP is the pressure drop, d_p is the particle diameter, L is the column length, η is the mobile phase viscosity and μ is the mobile phase linear velocity. Rearrangement of Equation (7) shows a direct proportionality between ΔP and ϕ. With a fixed set of operational conditions (i.e. fixed d_p, L and η), the pressure drop required to obtain a particular linear velocity is dependent upon ϕ. This is relevant in LC because there is an upper limit to the pressure attainable by conventional instrumentation (usually no higher than 400 bar). This in turn places limits on L and d_p, which affect efficiency, and μ, which affects analysis time. Within these limits, optimizing the separation usually involves a compromise between efficiency and speed.

A useful parameter in assessing the performance characteristics of LC columns that incorporates pressure drop and analysis time is separation impedance ($E^\dagger$). This dimensionless parameter introduced by Bristow and Knox[40] is a measure of column efficiency per unit time and per unit pressure drop, and is defined by Equation (8):

$$E^\dagger = \left(\frac{t_0}{N}\right)\left(\frac{\Delta P}{N}\right)\left(\frac{1}{\eta}\right) = \left(\frac{H}{d_p}\right)^2 \phi \tag{8}$$

Table 1 Values of h, ϕ and $E^{\dagger}$ for various column diameters

Column type	h_{min}	ϕ	$E^{\dagger}$
Narrow bore and conventional PCs ($d_c/d_p > 10$)	2	500–1000	2000–4000
Packed capillary columns ($d_c/d_p \approx 2.5$)	2	150	600
OT columns	0.8	32	20

Adapted from Knox.[41]
h = reduced plate height, ϕ = flow resistance parameter, $E^{\dagger}$ = separation impedance.

Table 2 Comparison of volumetric flow rates for various column diameters operated at fixed linear velocity (approximately three times optimum velocity)

Column diameter (mm)	Volumetric flow rate ($\mu L\,min^{-1}$)
4.60	1400
1.00	66
0.530	19
0.100	0.66
0.050	0.16

Assumed conditions: $\varepsilon_{tot} = 0.7$; $\mu = 2.0\,mm\,s^{-1}$.

where t_0 is the elution time of an unretained solute; other variables are defined previously. Lower values of $E^{\dagger}$ correspond to better column performance. Values of $E^{\dagger}$ and ϕ for various column diameters are shown in Table 1. The lower $E^{\dagger}$ values of miniaturized columns mean that these columns can yield more theoretical plates in less time and with a lower pressure drop relative to conventional bores.

The enhanced permeability of miniaturized columns allows for further increases in efficiency by making possible the use of smaller packing particles and/or longer columns. Recall from Equations (4–6) that plate height is related to particle size through the A and C_m terms of the plate height equation and efficiency is proportional to column length (Equation 3). In spite of this, conventional columns are constrained by excessive pressure drop, which limits usable lengths and particle sizes. Most 4.6-mm columns are 150–250 mm long and employ 5 µm particles. Alternatively, the use of smaller particles can result in reduced analysis times because when particle size is reduced, a given plate number is achievable with a shorter length of column.

3.2.2 Solvent Consumption

Decreasing column diameter results in a substantial reduction in the volumetric flow rate required for a given linear velocity. Volumetric flow rate (F) through a PC is given by Equation (9):

$$F = \frac{\pi d_c^2 \varepsilon_{tot}\,\mu}{4} \tag{9}$$

where d_c is the column diameter and ε_{tot} is the total column porosity. The dependence of F on the square of column diameter translates into significant reductions in solvent consumption on miniaturization. Table 2 compares volumetric flow rates for various column diameters. For example, downsizing to a 1.0-mm ID from a conventional 4.6 mm will reduce solvent consumption over 20-fold. Decreased solvent consumption is economical (reduced purchase and disposal costs) as well as environmentally responsible.

3.2.3 Mass Sensitivity

Chromatographic separations are intrinsically processes of dilution: an initial sample plug of concentration C is diluted in the mobile phase and reaches the detector as concentration C' (where $C' < C$). Because the majority of LC detectors are concentration-based, a reduction in sensitivity results. To understand why this process is ameliorated upon column miniaturization, we will define peak volume (V_p). Assuming a Gaussian peak, Equation (10):

$$V_p = FW_b = \left(\frac{\mu \varepsilon_{tot} \pi d_c^2}{4}\right) W_b \tag{10}$$

where W_b is the peak width at baseline in units of time. V_p is simply the volume over which the peak elutes. According to Equation (10), peak volumes are proportional to the square of column diameter; thus values of V_p for microcolumns are substantially smaller than for conventional 4.6-mm columns.

The concentration of solute at the peak maximum is inversely proportional to V_p.[4] Qualitatively, this relationship can be envisioned as follows. A given mass of solute is diluted in a volume V_p yielding a concentration C. If the same mass is diluted in a smaller volume V_p', the resulting concentration C' will exceed C in proportion to the volume difference. As indicated in Table 3, for a fixed mass of solute injected, column miniaturization can result in substantial increases in sensitivity with concentration-based detectors (e.g. UV, fluorescence).

3.2.4 Other Advantages

Decreased cross-sectional area allows for more efficient heat transfer to and from the column interior. Relative to conventional bore columns, miniaturized columns are capable of more rapid equilibration at different temperatures, which makes possible the use of temperature

Table 3 Typical peak volumes and relative sensitivity of various column diameters

Column diameter (mm)	Peak volume V_p (μL)	Relative sensitivity[a]
4.60	350	1.0
1.00	17	20
0.530	4.8	75
0.100	0.17	2000
0.050	0.040	8500

Assumed conditions: $L = 250\,\text{mm}$; $k' = 2$; $N = 10\,000$; $\mu = 2.0\,\text{mm s}^{-1}$; $\varepsilon_{tot} = 0.7$.
[a] This calculation is based on a fixed mass of solute injected and a concentration-based detector.

programming as in capillary GC. In addition, these columns can more effectively dissipate heat generated by the pressure drop, minimizing temperature gradients that can cause peak broadening. Smaller volume columns require less packing material, which is beneficial if costly ''exotic'' materials are used. This is also the case for mobile phases.

The low volumetric flow rates of miniaturized columns simplify interfacing to MS. Columns may be interfaced directly using thermospray, electrospray and other means, obviating the need for flow splitting, which reduces sensitivity.

3.3 Extra-column Effects

Not all peak broadening occurs within the chromatographic column. In addition to the phenomena accounted for in the plate height equation (Equation 5), which occur as the solute band passes through the column, the final peak width is affected by so-called extra-column effects. These include contributions from the sample injector, the injection volume, connecting tubing, detector cell volume and the detector time constant. Well-designed chromatographs are constructed such that extra-column effects comprise a small fraction ($\leq$10%) of the final peak width. One disadvantage of column miniaturization is that the system becomes more susceptible to extra-column effects as column diameter decreases.

Peak volume (V_p) was defined in Equation (10) as the peak width at baseline in units of volume. For Gaussian peaks, the final peak volume observed at the detector V_{obs} can be expressed in terms of the in-column V_p and extra-column V_{ec} contributions, Equation (11):[42]

$$(V_{obs})^2 = (V_p)^2 + (V_{ec})^2 \tag{11}$$

Dividing through by V_p^2 gives the relative increase in V_{obs} from extra-column effects, Equation (12):

$$\left(\frac{V_{obs}}{V_p}\right)^2 = 1 + \left(\frac{V_{ec}}{V_p}\right)^2 \tag{12}$$

According to this equation, if the extra-column contribution to the final peak volume, V_{obs}, is to be limited to 10%, then the extra-column volume must not exceed 46% of the column peak volume. This poses less of a problem with the relatively large peak volumes encountered with conventional bore columns. However, because peak volumes are proportional to the square of column diameter, to maintain the same fraction (10%) of extra-column broadening in micro LC requires much more stringent control of extra-column volumes. In such cases, extra-column volumes must be significantly reduced in order to minimize the contribution to the final peak volume. As a consequence, conventional HPLC instrumentation (i.e. injector, tubing and detector) must be modified in order to accommodate smaller diameter columns, which is a disadvantage of micro LC. Table 4 lists maximum allowable values for injection volume, connecting tubing length and detector cell volume for various column diameters.

In addition to the dilution and mixing that occurs within extra-column volumes, a contribution to band broadening may arise from the detection electronics. The speed with which the detector responds to a change in

Table 4 Maximum allowable extra-column volumes for various column diameters

Column diameter (mm)	Peak volume (μL)	Injection volume[a] (μL)	Tubing length[b] (cm)	Tubing length[c] (cm)	Cell volume[d] (μL)
4.60	350	40	31	N/A	15
1.00	17	1.9	0.071	940	0.73
0.530	4.8	0.54	0.0040	270	0.17
0.100	0.17	0.019	N/A	9.4	–[e]
0.050	0.040	0.0045	N/A	2.2	–[e]

Assumed conditions: $\varepsilon_{tot} = 0.7$; $\mu = 2.0\,\text{mm s}^{-1}$; $N = 10\,000$; $L = 250\,\text{mm}$; $k' = 2$; $D_m = 10^{-5}\,\text{cm}^2\,\text{s}^{-1}$.
[a] Allows 5% peak broadening.
[b] Allows 3% peak broadening, connective tubing diameter = 250 μm.
[c] Allows 3% peak broadening, connective tubing diameter = 50 μm.
[d] Allows 3% peak broadening.
[e] Assumes on-column detection with a negligible extra-column volume contribution.

solute concentration is affected by its time constant, τ. The time constant, which typically can be adjusted within a range of values, is useful as a filter for high-frequency noise in the chromatogram. Longer time constants slow the detector response providing enhanced signal filtering. Problems arise when the chromatographic peaks are very narrow, as in high efficiency separations. In these cases, slow detector response (long time constants) will result in additional peak broadening. It can be shown[43] that the maximum allowable time constant for a given fraction of peak broadening is proportional to $1/\sqrt{N}$. Thus, detectors employed in high efficiency separations require more rapid detection electronics.

3.4 Flow Profile and Efficiency: Advantages of Electro-osmotic Flow

EOF originates near the fused-silica capillary wall (or packing particle surface) owing to a charge imbalance that exists at the surface–liquid interface. The surfaces of fused-silica tubing and silica-based packing materials bear a net negative charge owing to the presence of ionized silanol moieties ($Si-O^-$). When exposed to a solution containing oppositely charged ions (e.g. the mobile phase), the surface will attract counterions to preserve electroneutrality. The attracted counterions form a two-part layer at the surface, known as an electrical double layer. The Stern layer consists of immobile counterions tightly affixed to the surface. A second layer, the Gouy or diffuse layer, extends some distance from the surface. The counterions present in this diffuse region are mobile and freely exchange with counterions in the bulk solution. The electrical double layer is depicted in Figure 2. Upon application of an electric field, counterions in the diffuse region begin to migrate toward the electrode of opposite polarity (generally the cathode). Because these migrating ions are solvated and because of frictional forces, the bulk solution is dragged toward the respective electrode. Electroosmosis is particularly advantageous as a driving force for mobile phase flow in chromatography because it generates considerably less peak broadening relative to pressure-driven flow. This is mainly because of the more uniform velocity profile of EOF. In fact, the potential to combine the selectivity of HPLC with the favorable flow dynamics of electroosmosis has been a driving force in the development of CEC. There are three main advantages of EOF over pressure-driven flow for mobile phase propulsion in chromatography: (1) EOF flow velocity is radially uniform (across flow channel), (2) EOF flow velocity is independent of channel diameter, and (3) EOF requires no pressure drop. Let us explore each of these points in greater detail.

Consider a capillary tube of radius r_c. In pressure-driven (or laminar) flow, the flow velocity at any radial

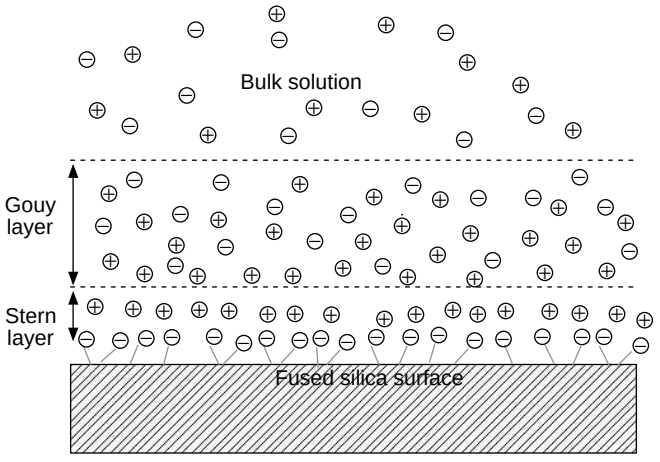

Figure 2 Schematic representation of the electrical double layer near the surface of fused-silica capillary tubing.

point in the tube $u(r)$ is a function of distance from the center as given by the Hagen–Poiseuille Equation (13):

$$u(r) = \frac{\Delta P}{4\eta L}(r_c^2 - r) \tag{13}$$

where r is the distance from the tube center. The resultant flow velocity profile is parabolic in shape. There exists a distribution of velocities over the tube cross-section. Recall from the discussion of the mobile phase mass transfer contribution to plate height (section 3.1), that such velocity inhomogeneity leads to an increase in plate height. More specifically, this zone-broadening phenomenon is known as the transchannel effect.[44]

Consider now EOF through the same capillary tube. EOF velocity (v_{eo}) at any radial point in the tube is given by Equation (14):

$$v_{eo} = \frac{\varepsilon\zeta E}{4\pi\eta}\left(1 - \frac{I_0(\kappa_d a)}{I_0(\kappa_d r)}\right. \tag{14}$$

where ε is the permittivity of mobile phase, ζ is the zeta potential, E is the electric field strength, η is the mobile phase viscosity, κ_d is the reciprocal of the thickness of the double layer, a is the distance from the center of the tube and I_0 is a zero-order Bessel function of the first kind.

Typically, the double layer thickness is quite small (on the order of 1 nm) and the term in parentheses reduces to one, yielding the following expression, Equation (15):

$$\langle v_{eo}\rangle = -\frac{\varepsilon\zeta E}{4\pi\eta} \tag{15}$$

where $\langle v_{eo}\rangle$ indicates the average flow velocity in the tube. Equation (15), which is valid in the absence of double-layer overlap, shows that EOF velocity is independent of radial position in the flow channel. Thus, the flow profile is flat over most of the capillary cross-section. Figure 3

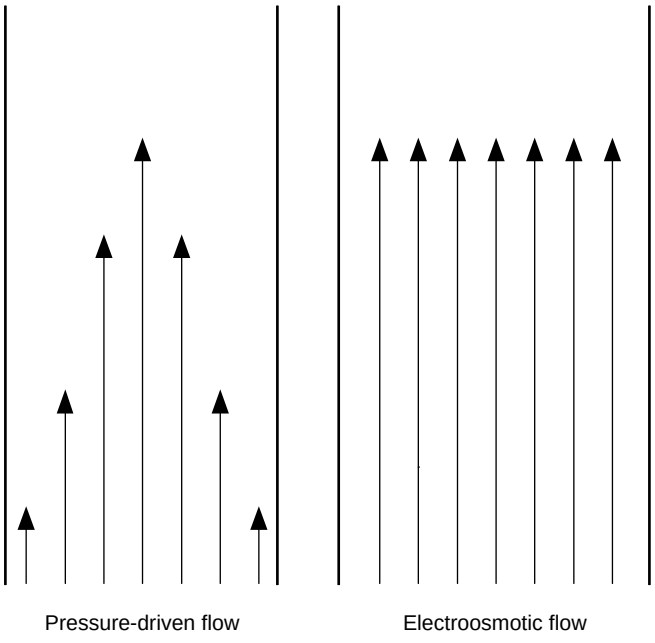

Figure 3 Radial flow velocity profiles of pressure and electrically driven flow through open tubes. Arrows depict velocity vectors.

compares the radial profiles of pressure and electrically driven flow. This uniformity in flow velocity minimizes transchannel zone broadening, resulting in a lower C_m term in the plate height equation.

In packed chromatographic columns, there invariably exist local differences in packing density which give rise to interstitial flow channels of varying diameter. With pressure-driven flow, the squared dependence of velocity on channel diameter (Equation 13) gives rise to a distribution of velocities in these channels. As in the transchannel effect discussed previously, point to point differences in flow velocity contribute to zone broadening. The plate height increment due to this phenomenon is part of the eddy diffusion (A) term of the plate height equation.

In contrast, EOF velocity, which is essentially independent of channel diameter, results in a much more uniform distribution of velocities in spite of differences in local packing density (depicted in Figure 4). We can expect the improved velocity distribution of electrically driven flow to result in a lower A term in the plate height equation.

The third advantage of EOF is that no pressure drop is required. This is important from the standpoint of efficiency because it permits the use of smaller diameter particles. Plate height, as discussed in section 3.1, is strongly dependent upon particle diameter and decreases with a decrease in d_p. In HPLC this advantage is offset by the inverse square dependence of pressure drop on d_p (Equation 7). The maximum pressure attainable by most

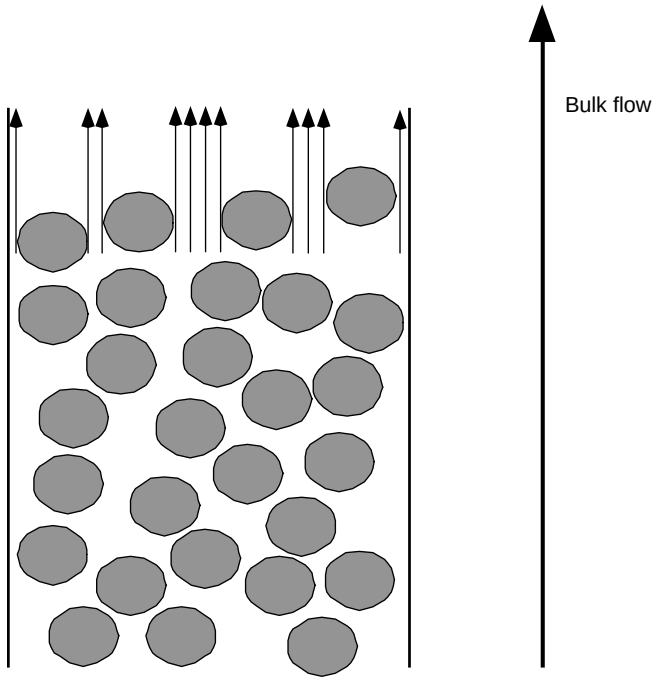

Figure 4 Schematic depicting EOF through a packed chromatographic column. Arrows represent velocity vectors, which are uniform through channels of varying diameter.

Table 5 Comparison of achievable efficiency and peak capacity in micro LC and CEC with various particle sizes

Particle diameter (μm)	CEC		micro LC	
	Plates per column	Peak capacity	Plates per column	Peak capacity
5	59 000	147	40 000	121
3	98 000	189	47 000[a]	130
1.5	200 000	269	24 000[b]	94

Assumed conditions: linear velocity = 1.0 mm s^{-1}; column length = 50 cm; high-pressure limit (micro LC) = 400 bar; $k'_{max} = 10$; Reduced plate heights: $h_{cec} = 1.7$, $h_{\mu-LC} = 2.5$.
[a] Column length limited to 35 cm owing to high-pressure limit.
[b] Column length limited to 9 cm owing to high-pressure limit.

commercial LC pumps is about 400 bar, which in most cases limits particle diameter to not less than 3 μm. With no such limit, CEC separations regularly employ 1.5-μm particles and the potential exists to employ much smaller particles.[45]

Reduced plate height values obtained for the same columns under electrically and pressure-driven conditions have been shown to be consistently lower under the former,[39] a direct result of the improved flow dynamics of EOF. Greater efficiency, combined with the absence of pressure drop gives CEC the potential to generate total column plate numbers that are substantially higher than micro HPLC.

An important feature of merit for a chromatographic technique is peak capacity, which is defined as the number of peaks with unit resolution that can be contained in the column. Peak capacity, then, is a measure of the applicability of a separation technique to highly complex samples (i.e. those containing large numbers of peaks). In comparison to capillary GC, peak capacity in HPLC is limited by the relatively low efficiency of the technique. In contrast, the high separation efficiency of CEC provides increased peak capacity. Table 5 compares typical column efficiency and peak capacity values achievable in CEC and capillary HPLC.

4 INSTRUMENTATION

Although capillary columns offer numerous advantages, their full exploitation is limited by the technological ability to miniaturize the instrumentation properly.

4.1 Columns for Microscale Liquid Phase Separations

4.1.1 Column Performance

All types of packing material offered for conventional HPLC can be used in micro LC. However, submicron size packing materials have been synthesized to increase column performances, considering the fact that pressure drop is no longer a limitation in the micro HPLC format. Moreover, the CEC stationary phase must be able to sustain EOF; thus charged groups need to be incorporated in the packing particles in the case of PCs, or on the capillary surface in OT columns.

Undoubtedly, OT columns offer the highest separation efficiency and highest permeability. Numerous papers have indicated that 1 million plates can routinely be obtained[17] with reduced plate heights on the order of 1.

In micro LC, drawn-packed capillary columns provide efficiencies of 100 000 plates m^{-1} if very long columns are prepared and if long analysis times are acceptable. Such performances are comparable to the packed capillaries, so both types of columns were considered in micro HPLC.

However, in CEC mode the drawn-packed capillaries showed efficiencies of 100 000 plates m^{-1}, considerably lower than the 500 000 plates m^{-1} obtained with the packed capillaries.[39] Since the manufacturing of drawn-packed capillaries is much more elaborate and the columns are fragile compared to the fused-silica ones, these columns are no longer used for CEC.

Packed capillary columns of 100–300-μm ID routinely offer 100 000 plates per meter in micro LC with 5-μm particle diameters (C_{18}). The reduced plate heights obtained vary between 2.2 and 3, close to the theoretical limit of two accepted for pressure-driven liquid chromatographic systems.[40] Kennedy and Jorgenson[27] demonstrated that columns of ID smaller than 50 μm may yield even smaller reduced plate heights, $h_{min} = 1.5–1.8$. For these columns the separation impedance ranges from 1300 to 2500; thus longer columns can be used and extremely high separation efficiencies are possible.[27,46]

Jorgenson explained this effect in a later paper,[47] in which he indicated that microcolumns and conventional columns contain two regions of packing, which are differentiated by their permeability. In the core of the column, the more tightly packed region comprises the first region, while the second region, loosely packed, is near the walls of the capillary.

In CEC two notable benefits are provided by EOF: the flat profile, which as mentioned earlier provides increased efficiency per unit length, and the potential for use of very long columns, which derives from the absence of flow resistance as encountered with conventional pumps. Thus, in CEC reduced plate heights of 1.5 are easily obtainable and total plate counts in excess of 150 000 are possible.

Pressure-driven flow is at present more reproducible than EOF. Column-to-column irreproducibility (RSD (relative standard deviation) larger than 4%) is more significant than run-to-run irreproducibility (RSD currently 1–2%). Some analysts consider that this may be caused by column-to-column irreproducibility in packing, frit fabrication and column surface chemistry.

4.1.2 Column Preparation

Columns in both the OT and packed capillary format are commercially available from companies such as LC packings, Unimicro, Micro-Tech, Keystone and others. However, in most research laboratories the columns are home built using fused-silica tubing available from Polymicro Technologies (Phoenix, AZ, USA) and others using procedures similar to those outlined below.

In OT columns the stationary phases are attached to the walls of the capillary by covalent bonding or by polymer coating. Both procedures require an etching step which activates the silanol groups of the fused silica surface, such that most of them can participate in stationary phase bonding and will not interfere with the retention mechanism. The chemical bond in the first column type, commonly known as a monolayer column, is formed between anchoring groups and molecules that will serve as, and determine, the selectivity of the stationary phase.

Polymer coatings are often preferred to the monomolecular phases because the thick layer of stationary phase ensures an increased sample capacity and solute retention. The in situ procedure comprises the introduction into the column of a mixture containing the functional monomer (often acrylate), initiator, cross-linker and solvent. During the polymerization process,

which can be initiated thermally or via UV radiation, the solvent is evaporated and the oligomers are left immobilized on the surface.[48] If suitable anchoring groups are attached to the surface before introducing the polymerization mixture, the polymeric network is covalently bonded onto the surface and the column is more stable. For polyacrylate phases, used in both micro LC[17] and CEC,[49] the selectivity factors vary from capillary to capillary within 1 and 2% RSD. Prior to use, the columns are flushed with the appropriate mobile phase.

The quality and lifetime of PCs is directly dependent on the homogeneity of the packed bed. The method most often used for preparing a PC is the slurry packing technique. Dry packing methods, which claim shorter packing times and similar column performances, have also been used for both polar and nonpolar stationary phases.[50]

In a typical slurry packing procedure, the packing material and about 1 mL of solvent (selected for compatibility with the packing material; for reversed-phase, acetonitrile and water often yield the best results) are thoroughly mixed and sonicated, then transferred to a stainless steel reservoir. The reservoir contains a metal frit and is connected on one side to a high-pressure pump that is filled with a displacement solvent. The capillary to be packed, the outlet side of which contains a retaining frit (made of silica, Teflon™, glass wool or metal) is fitted to the opposite side of the reservoir. In order to ensure the homogeneity of the slurry, a tiny magnetic stir bar placed into the reservoir can be used for mixing. Also the reservoir can be placed in an ultrasonic bath, and sonicated during packing. These features are essential to minimize particle aggregation and settling out of the slurry during packing. Pressure is applied and increased gradually up to about 250 bar. When the column has been packed to the desired length, the system is depressurized and the column can be flushed and conditioned with the mobile phase that is to be used for the analysis.

Alternative packing procedures use different means to force the packing material from the slurry reservoir into the capillary. During electrokinetic packing, a potential ramp (2–30 kV) is applied across the capillary and the slurry reservoir such that the charged particles electromigrate into the column.[51] The centripetal-based packing technique uses the centripetal force created by spinning a rotary reservoir onto which numerous capillaries are mounted.[52] Both techniques are suitable for mass production of packed capillaries and have been shown to produce columns offering performance characteristics equal to columns made using more conventional means.

4.1.2.1 The ''Frit Issue'' Before mounting the capillary in the slurry reservoir, the outlet end must be fitted with a porous ''plug'', a frit, which will allow the solvent to pass through while the particles stack to form a packed bed. Thus, frits must be able to withstand the high pressures used during the packing procedure. In micro LC, glass wool, metal and silicate frits are used. For CEC, EOF should preferably be supported by the frit, so silicate frits are typically used. They are easier to make[53] and can provide porosities very similar to those of packed beds. However, the porosity and surface structure of each of these frits is often sufficiently different from the packed bed to result in peak broadening in micro LC and CEC and gas bubble formation in CEC. Also, instability of the current and baseline signals might occur. Moreover, the columns used for CEC designed for on-column detection must have an additional frit at the column outlet, to keep the bed in place during use. These frits can be obtained by packing porous silica particles and heating the outlet area with a highly focused source of heat such as a Ni–Cr coil or a fiber optic fusion splicer. This will partially adhere the silica spheres together to form a frit; the side effect of applying the heat is the total combustion of the polyimide coating, which renders the column extremely fragile. Various compositions for silicate frit preparations have been tried,[53] but none seems to solve the problem completely.

4.1.3 Perspectives

Recent advances in column fabrication include the development of nanosized particles, refinement of chiral selectivity of stationary phases, as well as refinements in the design and fabrication of packing materials specific for CEC separations.

Another trend considers the need to address the frit issue and the best answer seems to be the fritless column. New column designs involve a continuous highly cross-linked porous matrix that can incorporate selective sites, as in the case of ''monolithic'' columns, or can surround the packing material and hold it in place during the operation of ''entrapped'' columns.

Numerous papers published recently have studied the synthesis and performance of monolithic columns.[54–56] The in situ polymerization of an organic or silicate-based monomer within a capillary column acting as a mold produces a continuous rod of rigid highly porous polymer. For CEC the cross-linked polymer is often provided with charged groups (e.g. sulfonic acid groups) able to sustain the EOF. The main advantages offered by silicate matrices over the synthetic organic polymers include resistance to swelling, stability over time and readily available silanol groups for operating in the CEC mode. On the other hand the synthetic polymers offer a wide variety of functionalities and the possibility to operate at a wider range of pH.

The silicate-entrapped columns are made by trapping the packing particles in a network of silica. Therefore,

aside from the robustness afforded the column, the silicate entrapment provides ionizable silanol groups, which result in an increase in the flow velocity and in radial homogeneity of the electric field. Faster separations and modestly increased efficiencies per unit time can be obtained.[57]

4.2 System Components: Microscale Liquid Chromatography

4.2.1 Pumps

The overall accuracy and precision of any HPLC system is strongly dependent on the ability of the solvent delivery system to deliver stable flow rates. The very small volumes (microliters and nanoliters per minute) required by miniaturized columns place an even more stringent demand on micro LC pumping system. Reciprocating piston pumps, which are the standard in conventional HPLC, lack reproducibility at flow rates below about $0.1\,\text{mL}\,\text{min}^{-1}$. For this reason, these pumps are rarely used in micro LC without a flow splitting device.

The majority of micro LC applications employ syringe pumps. This type of pump, a schematic of which is shown in Figure 5, consists of syringe of fixed volume driven by a stepper motor. Syringe pumps, such as those available from Isco (Lincoln, NE, USA) and Carlo Erba (Italy) are capable of delivering precisely metered flow rates, typically within the range of $0.1–1000\,\mu\text{L}\,\text{min}^{-1}$. Advantages over reciprocating piston pumps include better reproducibility at low flow rates and pulse-free operation. The main disadvantage of the syringe pump, a limited volume of solvent, is less problematic with the very low flow rates encountered with micro-columns.

4.2.2 Injectors

In general, the sample injection system in micro LC must be capable of introducing very small volumes (e.g. nanoliters) onto the column with a high degree of accuracy and precision. In addition, the injection system must contribute minimally to extra-column peak broadening.

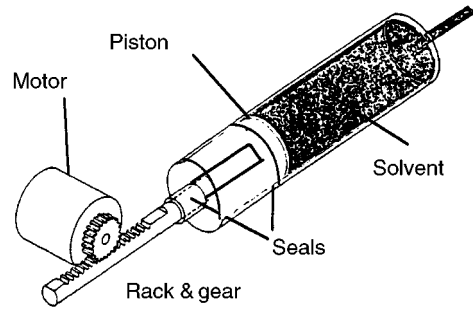

Figure 5 Schematic of a syringe pump.

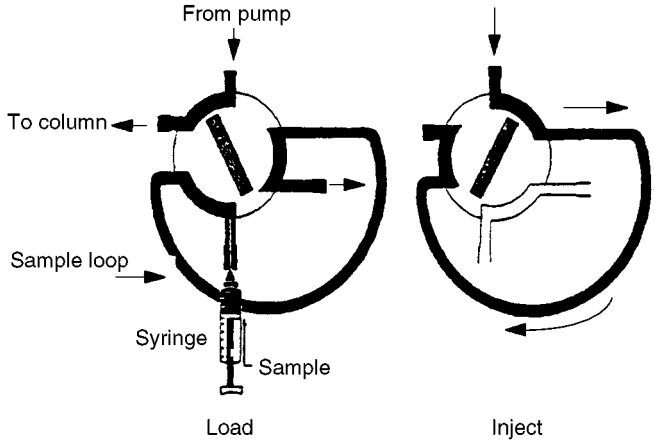

Figure 6 Schematic of an injection valve.

The most common means of sample injection in micro LC are microvalve injectors, such as those produced by Valco (Houston TX, USA) and Rheodyne (Cotati CA, USA). In this type of device, depicted schematically in Figure 6, an internal loop of fixed volume controls the amount of sample injected. Injection volume is varied by changing the sample loop. Injection loops are available commercially from a number of manufacturers in volumes ranging from 60 nL to 1 μL.

The flow geometry and low volume of these injectors permits clean and thorough sweeping of the loop volume, which minimizes peak broadening and tailing.

4.2.3 Detectors

A variety of detection schemes have been employed in micro LC, the most common of which include UV, fluorescence, electrochemical, and mass spectrometric detection. Of these, the optical methods, UV and fluorescence, are the most widely used. With optical detection, the flow cell design is vital in controlling extra-column band broadening; specifically, cell volumes must be extremely small. Ideally, the cell volume should be less than one-tenth of the peak volume;[42] however, decreasing cell volume will generally decrease the detection path length, which reduces sensitivity in accordance with the Beer–Lambert relationship. Thus, optimizing cell design generally involves a compromise between sensitivity and efficiency.

Optical detection in capillary LC can be performed using the capillary column itself as the flow cell. In this mode, known as on-column detection, a small section of the polyimide coating of the capillary is removed downstream from the packed bed, rendering it UV transparent. Incident radiation is directed through a slit, passes through the capillary and strikes a photodetector

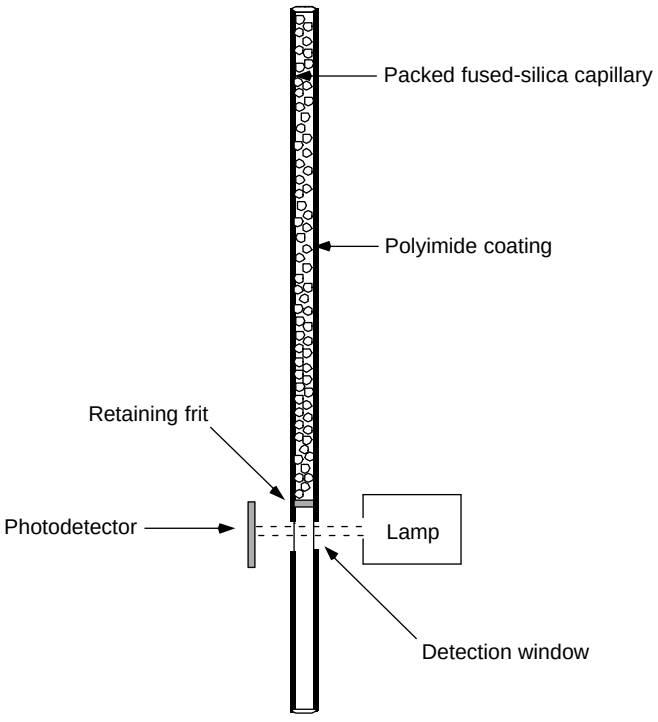

Figure 7 Schematic of on-column detection with a packed fused-silica capillary column.

situated opposite the entrance slit (depicted in Figure 7). Advantages of on-column detection include its simplicity and minimal extra-column volume contribution. It is, however, limited by the low sensitivity imposed by the short detection path length.

Alternatively, micro flow cells can be employed with longer path lengths to provide increased sensitivity while contributing only very small volumes to the system. One such cell configuration, available commercially from LC Packings (Amsterdam, the Netherlands), employs a Z-shaped fluid path designed to increase detection path length.

Some limitations of micro flow cells are cost (relative to on-column detection) and problems with coupling to the chromatographic column (e.g. leaks, dead volumes).

Electrochemical detection is particularly attractive for use in micro LC. Used primarily in voltammetric and amperometric modes, electrochemical detection is sensitive, selective, amenable to miniaturization and, in contrast to optical detection, suffers no loss of sensitivity with miniaturized cell volumes. (For a recent review of electrochemical detection in miniaturized separation techniques see Ewing et al.[58]). DLs in the femtomole range have been obtained with voltammetric detection using a carbon fiber microelectrode coupled to a capillary LC column.[59]

Alternatively, on-column detection can be performed with fused-silica capillary columns. On-column detection

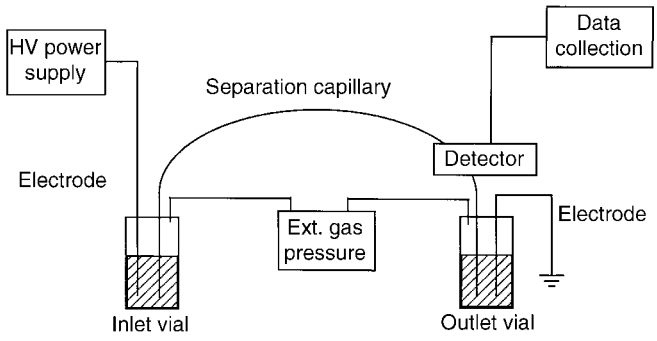

Figure 8 Block diagram of a CEC system.

is advantageous primarily owing to its simplicity and minimal extra-column volume contribution. A drawback of this approach, however, is reduced sensitivity due to a substantially reduced detection path length.

4.3 System Components: Capillary Electrochromatography

A block diagram of a typical CEC system is illustrated in Figure 8. The instrument consists of a high voltage (HV) power supply capable of applying a high potential (usually up to 30 kV) and a pair of platinum wire electrodes. The separation capillary is immersed in the inlet and outlet mobile phase reservoirs, forming a complete circuit. In PC CEC, it is usually necessary to pressurize the mobile phase reservoirs in order to minimize the formation of gas bubbles, which can drastically affect EOF.

Sample introduction in CEC is usually achieved via electrokinetic injection. In this technique, the capillary inlet is first removed from the mobile phase vial, immersed into the sample and a voltage applied. The resultant EOF draws a given amount of sample into the capillary. The capillary inlet is then reimmersed in the mobile phase vial and the separation begun. In contrast to capillary LC, in which a known volume of sample is injected by means of a valve, the injection volume in CEC must be determined indirectly from the injection time and the EOF velocity at the voltage used. Because of this, the injection precision is dependent upon the stability of the EOF. Not surprisingly, injection to injection reproducibility tends to be lower in CEC than in capillary LC.

CEC instrumentation is virtually identical to that used in CE, with the exception of external pressurization, a minor modification. This is an added advantage for CEC in that both separation techniques can be performed on the same instrumentation.

Optical detection in CEC is generally in the on-column format. This requires that a detection window be made (as in capillary LC) by removing a small section of the polyimide cladding. As in capillary LC, low detection sensitivity imposed by short cell path lengths is an issue

in CEC. Specially designed detection cells including ''Z-cells'', ''bubble cells'' and the High Sensitivity Detection Cell® marketed by Hewlett–Packard provide increased path lengths, which improve DLs.

Low sensitivity, a major disadvantage of CE, is less a problem in CEC owing to increased sample capacity. The packed bed provides a high surface area of stationary phase on which sample can be accommodated. As a result, CEC allows for larger volume injections without compromising efficiency.

The most common detection scheme in CEC is on-column UV. Alternatively, on-column laser induced fluorescence (LIF) is a highly sensitive technique that provides some compensation for the short detection path lengths. DLs as low as 10^{-11} M were obtained for a series of polycyclicaromatic hydrocarbons (PAHs) using CEC/LIF.[60] As discussed in the next section, mass spectrometric detection has been combined with CEC with excellent results.

4.4 Mass Spectrometry Coupling

4.4.1 Micro Liquid Chromatography/Mass Spectrometry

Interfacing a mass spectrometer to a chromatograph can provide a wealth of analytical information. In addition to mass-selective detection, a mass spectrometer can yield structural information that can be used to identify compounds as they elute from the column. In spite of the advantages gained by combining the techniques, interfacing LC to MS has been hampered by practical problems. In contrast to GC/MS, it is only since about 1985 that LC/MS has grown into a widespread technique, with the advent of improved interface and ionization source technology, in particular electrospray ionization (ESI).

The chief problem in coupling LC to a mass spectrometer is the incompatibility that results from the relatively large volumes of effluent from the LC column and the vacuum requirement of the mass spectrometer. Ideally, the LC effluent should be vaporized and the solvent vapor pumped out of the system prior to the gas-phase solutes entering the environment of the mass spectrometer. Eliminating the large volumes of solvent vapor generated by 4.6-mm columns is difficult. Consequently, reducing the volumetric flow through the LC system facilitates coupling of the systems. Thus, with respect to flow rates, micro LC is a more natural match for MS than conventional LC. This, in addition to the inherent advantages of miniaturized columns described earlier, has led to the emergence of micro LC/MS as a powerful analytical tool. (For a thorough review of the theory and practice of capillary LC/MS, see Tomer et al.[61])

Currently one of the most popular means of interfacing micro LC to MS is via the ESI source. In the typical ESI arrangement, the sample solution passes through a metal capillary tube, to which is applied a potential on the order of 3–5 kV with respect to a grounded counter electrode a short distance from the tip. Electrostatic repulsion results in the ejection of a stream of charged droplets from the capillary tip. Ionized sample molecules, at atmospheric pressure, are desolvated with the aid of a stream of warm nitrogen gas. The gas phase sample ions enter the low-pressure environment of the mass spectrometer through a series of skimmers. Because no sample heating is required, ESI is well suited for application to large, thermally labile molecules (such as many biomolecules). A soft ionization source, ESI yields predominantly molecular ions. For structural elucidation, ESI is often combined with triple quadrupole mass analyzers in which collision-induced dissociation (CID) is employed to induce fragmentation. Typical solution flow rates for ESI are $1–10\,\mu L\,min^{-1}$. Pneumatically assisted ESI or ion spray, in which a nebulizing gas is employed to assist droplet formation, can accommodate higher flow rates, up to $100–300\,\mu L\,min^{-1}$. These flow rates are directly compatible with miniaturized HPLC columns; flow splitting of the column effluent, which reduces sensitivity, is usually not required.

The high mass sensitivity of micro LC facilitates the analysis of volume-limited samples. For this reason, capillary LC/MS is especially useful for biological samples. In particular, capillary LC/ESI/MS has been utilized extensively in protein and peptide analysis.[62–67] The high resolving power of capillary LC is used to separate complex mixtures of peptide digests, which are subsequently sequenced using tandem MS. One of the most important advantages of the capillary LC format in these sequencing analyses is that only very small amounts of starting material are required. Sequences of peptides have been obtained via capillary LC/ESI/MS with as little as 800 fmol of sample.[64]

Capillary LC/ESI/MS has been employed to study the various peptides that bind to class I major histocompatibility complex molecules.[63] Recognition of these bound peptides by cytotoxic T lymphocytes is an important component of the immune response of the body. Peptide fractions were separated by reversed-phase capillary HPLC and sequence information was obtained by tandem MS using a triple quadrupole instrument. Peptide sequences were obtained with as little as 900 fmol of sample.

Improvements in ESI detection sensitivity gained by reducing column diameter have been demonstrated by Banks.[65] Using a time-of-flight mass spectrometer, measured signal-to-noise (S/N) ratios increased 428-fold upon decreasing column diameter from 2.1 mm to 0.1 mm. A mass DL of 2.5 fmol was obtained for cytochrome-C using the 0.1-mm capillary. Figure 9 is a separation of a mixture containing five proteins (25 fmol each) in less than 7 min using this instrument.[65] These results underscore the power of capillary LC/MS.

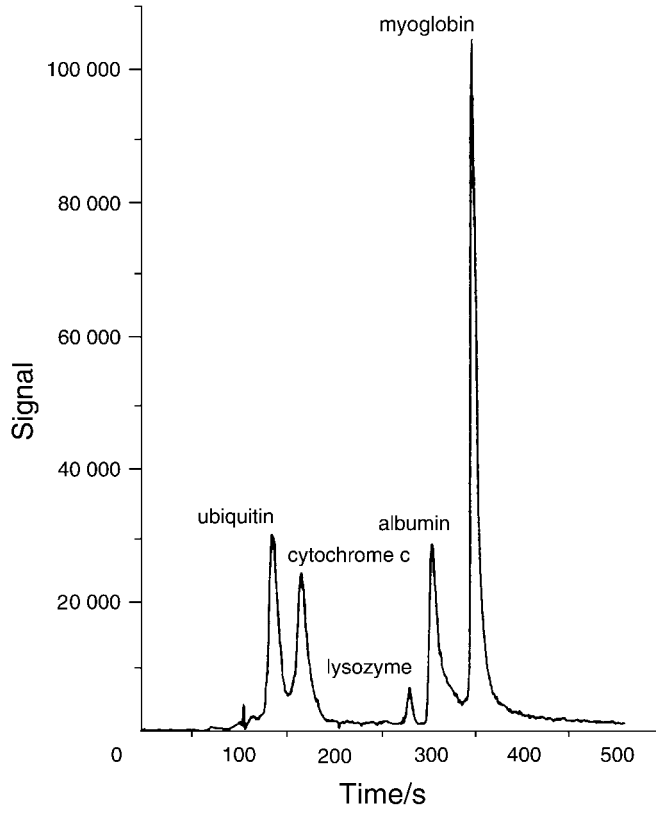

Figure 9 Total ion chromatogram of a protein mixture (25 fmol each) obtained with a 100-μm ID HPLC column. (Adapted from Banks, *J.Chromatogr., A*, **786**, 67–73.[65] Copyright (1997) with permission of Elsevier Science.)

4.4.2 Capillary Electrochromatography/Mass Spectrometry

The volumetric flow rates of PC CEC (typically $0.5–2\,\mu L\,min^{-1}$) are also ideal for coupling to MS, in particular ESI. For the analysis of neutral solutes, coupling CEC to MS is more straightforward than MEKC because the latter requires relatively high concentrations of surfactants, which in general are incompatible with most MS interfaces.

Since the mid-1990s, several research groups have been active in developing CEC/MS. In 1994, a 50-μm capillary packed with 3-μm particles was coupled to a mass spectrometer using a fast atom bombardment (FAB) interface.[68] The system was used to analyze a series of steroids. In this instrument, the separation capillary was connected to the MS by means of a transfer capillary, which degraded the chromatographic efficiency of the system owing to the extra-column volume contribution of the transfer line. Schmeer et al.[69] coupled 100-μm ID CEC columns directly to an ESI/MS without the need for a transfer capillary. The outlet end of the separation capillary was positioned directly inside the inner steel needle of the ESI source, eliminating the postcolumn peak broadening that occurs with transfer lines. Although supplementary pressure was employed to stabilize the flow rates, bulk flow was generated mainly by EOF. The instrument was used to separate peptide mixtures at picomole levels.

A few of the other applications of CEC/MS to date are the analysis of textile dyes,[70] pharmaceuticals[71] and DNA adduct mixtures.[72]

CEC/MS is a technique with much promise. The ability to analyze very small sample volumes and the high peak capacity of CEC in combination with the sensitivity and selectivity of MS yields a powerful microanalytical technique.

5 OPTIMIZATION OF ANALYSIS

In order to achieve the ''optimum'' solution to a particular separation problem, numerous aspects must be considered: analyte structure and matrix complexity, nature of the stationary phase, type of column (capillary or conventional, OT or PC), detection sensitivity and operating conditions (such as mobile phase composition, temperature and pressure).

5.1 Selectivity Optimization

The trends in stationary-phase evolution for micro LC and CEC are similar to those in conventional LC: development of stationary phases able to affect chiral separations or screen components of combinatorial libraries or other complex mixtures.

An interesting approach involves the use of molecular imprinted polymers (MIPs). Specific functional monomers interact through covalent or noncovalent interactions with complementary groups of a template molecule. In a subsequent polymerization step cross-linking polymers react with the monomers and create a macroporous matrix. Extraction of the template leaves cavities complementary to the imprint molecule in both shape and chemical functionality. Prepared in bulk, followed by subsequent grinding and sieving, or prepared in situ, resulting in MIP monolithic columns, or produced as spherical beads, the molecularly imprinted polymers provide extremely selective materials for HPLC and electro-driven separation techniques.

Much effort has been invested in designing MIPs suitable for CEC since the increased efficiency specific to this technique is expected to minimize to an extent the band broadening often encountered in separations using MIPs. Promising results have been obtained in the OT/CEC format where baseline separation of D- and L-dansyl phenylalanine was achieved.[73] Other successful approaches consider the macroporous monoliths, which

demonstrated commercial potential by providing the enantioseparation of phenylalanineanilide[74] and β-adrenergic antagonists propranolol and metoprolol.[75]

5.2 Gradient Separation and Temperature Programming

When separating analytes having substantially different capacity factors, the peaks that elute early tend to be poorly resolved while those that appear later in the chromatogram are broadened, a phenomenon known as the ''general elution problem''. Solutions to the general elution problem generally employ establishment of a gradient in one (or more) of the following: solvent strength (gradient elution), viscosity (temperature programming) flow velocity (flow programming), and so on. In gradient elution the strength of the mobile phase is continuously increased during a run, while in temperature programming the viscosity of the solvent decreases as the temperature of the capillary situated in an oven is gradually raised. In CEC increasing flow velocity can be accomplished readily by programming a gradual increase in the applied potential.

The generation of reproducible controlled solvent gradients in capillary columns is not an easy task considering the small flow rates (submicroliter per minute) that must be delivered. In CEC the problem becomes even more complex for instrumental design considerations, since in addition to the delivery of a solvent gradient, a high voltage must be applied simultaneously, and from a chromatographic point of view, because variations in solvent composition significantly affect EOF velocity.

One of the first approaches considered the supply of a pressure-driven gradient in parallel with application of high electric field.[76] In order to take advantage of the high efficiency offered by CEC, the separation must be performed using solely EOF. Therefore, the most embraced approaches consider designs that use HPLC pumps to deliver mobile-phase gradients to the inlet of the separation column.[77,78] Mixtures of drugs,[78] PTH (parathyroid hormone)-aminoacids (from the Edman degradation of a peptide) and steroid hormones[77] were resolved using a mobile-phase splitting device.

Another type of gradient elution system proposed for CEC provides a pure electroosmotically driven flow.[79] It is composed of two fused silica capillaries each connected to a programmable high-voltage power supply and to a different mobile phase reservoir. The applied voltages are varied during the run such that as the flow of one mobile phase increases (as the corresponding potential increases) the amount of the other mobile phase decreases (the applied potential decreases) in a manner that ensures a constant resultant flow. The outlets of the solvent delivering capillaries are attached through a T connector to the inlet of the separation column. System feasibility was proven when 16 PAHs were separated in less than 90 min.

As opposed to conventional LC columns, capillary columns have a small heat capacity, which confers another advantage: the use of temperature programming. The simplicity of the instrument requirements (an oven with temperature control) allows the consideration of temperature programming as a replacement for or augmentation of gradient elution for thermally stable analytes. A comparative study[80] was performed on reversed-phase packed capillaries for separations of alkylbenzenes at temperatures varying from 30 to 80 °C and mobile phases of 60–80% acetonitrile in water. The results indicate that a 5 °C change in column temperature is equivalent to 1% change in acetonitrile concentration. Thus, temperature programming can be used alone or in conjunction with gradient elution to tune the separation of closely eluting compounds finely.

The issue of temperature programming is a bit more complicated in CEC since, besides retention factors, changes in temperature have a considerable effect on the viscosity of the mobile phase, thus on the EOF. Moreover, when working with relatively volatile solvents it is desirable to operate at decreased temperature in order to avoid bubble formation, which has a dramatic effect on the EOF. However, it has been shown that by raising the temperature 25–50 °C the enhancement of mass transport provides a three-fold increase in separation speed.[80]

5.3 Ultra High-pressure Micro Liquid Chromatography

Since the mid 1980s numerous aspects of micro LC and CEC instrumentation have been studied, and pressure or voltage limits once considered fixed have been pushed further in an attempt to increase the performance of miniaturized systems.

Jorgenson, MacNair and Lewis have indicated a possible new direction in micro LC development, the use of extremely high pressure.[81] It has long been recognized that small particles can significantly decrease plate height, though they simultaneously generate large back pressures which impede the use of long columns and increase analysis time by dictating lower flow rates. By using pressures as high as 1400 bar (20 000 psi), Jorgenson et al. obtained reduced plate heights of 1.4 for 66-cm long packed capillaries, accompanied by a significant reduction in the analysis time.

6 APPLICATIONS

6.1 Micro Liquid Chromatography

Micro LC has been employed in a variety of different applications. A number of textbooks and literature

reviews[4,82] present thorough discussions of the various analytical problems to which micro LC has been applied. In this space, only a small sample of these is presented; the interested reader is referred to the cited literature for a more detailed listing.

An interesting application of micro LC that exploits the technique's small volume requirement, is the analysis of single biological cells. In 1992, Jorgenson et al. published a method in which the catecholamines norepinephrine and epinephrine were determined in a single bovine adrenomedullary cell.[83] These authors employed 42 and 50-μm ID fused-silica capillary columns packed with 5-μm C_{18}-coated silica particles. Amperometric detection, which was employed using a carbon fiber microelectrode, yielded DLs in the attomole range. A chromatogram obtained from a single adrenomedullary cell is shown in Figure 10. This research group also published a method for the analysis of single neurons using OT LC in 1989.[59]

Separations of enantiomers are becoming increasingly important and present tough challenges to analysts. One approach to chiral separations in LC is to use a conventional stationary phase and add a chiral selector (e.g. cyclodextrins) to the mobile phase. One of the disadvantages of this approach is that these additives are often quite costly. However, the very low column volumes and solvent consumption rates of miniaturized columns facilitate the use of expensive additives because only a very small amount is required. In a paper by Takeuchi et al.,[84] capillary LC with a cyclodextrin additive was employed to obtain separations of 1-ferrocenylethanol and dansyl-phenylalanine enantiomers. In this method, 350-μm ID capillaries were packed with 5-μm ODS (octadecylsilane)-5 particles. Various amounts of methylated β-cyclodextrin were added to the aqueous acetonitrile mobile phase. A separation of ferrocenylethanol enantiomers is shown in Figure 11.

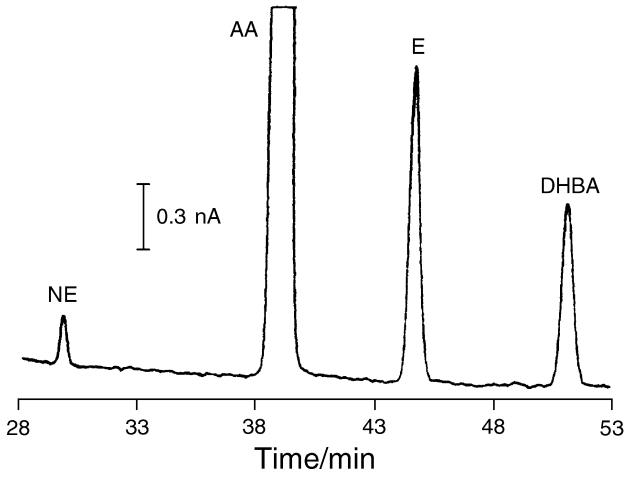

Figure 10 Chromatogram from a single bovine adrenomedullary cell obtained with a packed capillary HPLC column and electrochemical detection. NE is norepinephrine, AA is ascorbic acid (added to samples as an antioxidant), E is epinephrine and DHBA is (3,4-dihydroxybenzyl)amine (internal standard). These peaks correspond to 13.4 and 149 fmol of NE and E, respectively. (Adapted with permission from Cooper et al.[83] Copyright (1992) American Chemical Society.)

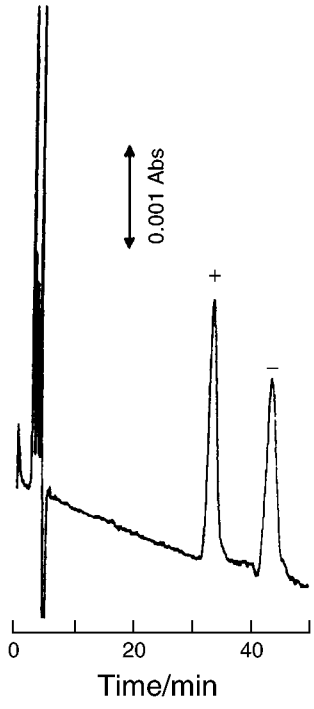

Figure 11 Capillary LC separation of ferrocenylethanol enantiomers obtained with 350-μm ID capillaries and a β-cyclodextrin-based chiral selector added to the mobile phase. (Adapted from Hu et al. *Anal. Chim. Acta*, **295**, 173–179.[84] Copyright (1994) with permission of Elsevier Science.)

A short listing of micro LC applications is presented in Table 6.

6.2 Capillary Electrochromatography

CEC is a relatively new technique that is receiving a great deal of attention from analytical chemists. For this reason, the number of publications on CEC is growing rapidly, as are its applications. A review by Bartle et al. includes a comprehensive listing of CEC applications up to the end of 1996.[90] In addition to those discussed below, a few applications of CEC are presented in Table 7.

Published in 1995, the separation of 16 PAHs shown in Figure 12 demonstrates the separation power of CEC.[60] PAHs are environmental contaminants classified by the US Environmental Protection Agency as "priority" pollutants. This separation was performed with a 75-μm capillary packed with mixture

Table 6 Applications of micro LC

Compounds	Column ID (μm)	Stationary phase	Detection	Comment	Ref.
β-Lactam antibiotics	320	7–8-μm perfusion	ESI/MS	Analyzed in milk	85
Prostaglandins; steroids	240	3-μm ODS	Fluorescence	Analyzed in plasma	87
Bile acids	260	5-μm	Fluorescence	Analyzed in serum	4
Diltiazem and metabolites	320	5-μm C_{18}	UV	Analyzed in plasma	88
Retinoids	180	3-μm C_{18}	Electrochemical	Analyzed in serum	89
Neurotransmitters; amino acids	15–20	Dimethyloctadecylsilane (OTLC)	Electrochemical	Single neuron analysis	20

Table 7 Applications of CEC

Compounds	Column ID (μm)	Stationary phase	Detection	Comment	Ref.
Tricyclic antidepressants	50	3-μm ODS	UV	$>8 \times 10^6$ plates m^{-1}	91
Triglycerides	100	3-μm ODS	UV	$h < 2$	92
PAHs	75	3-μm ODS	LIF	$1.0–1.5 \times 10^5$ plates obtained	60
Hydroxybenzoates	25	Polymethacrylate (OT/CEC)	UV	In situ polymerization of stationary phase	49
Tetracyclines	50	C_{18}	UV	Stationary phase covalently linked to wall	93
Corticosteroids	50	3-μm ODS	UV	Analyzed in urine and plasma	94

(90 : 10) of 3-μm ODS and 1-μm nonporous silica. On-column LIF detection yielded mass DLs on the order of 10^{-19} mol. Efficiencies of 100 000–150 000 plates m^{-1} were obtained.

In 1997 Euerby et al. published a paper on the applications of CEC in pharmaceutical analysis.[86] Here, this group reported separations of a variety of compound classes including peptides, nucleotides, steroids, diuretics and acidic drugs. Figure 13 illustrates a separation of four commonly used diuretics performed on a commercially available 3-μm CEC Hypersil™ C-18 capillary. Efficiencies ranging from 149 000 to 344 000 plates m^{-1} were obtained.

An interesting application of CEC was recently reported in which inorganic ions were separated by ion exchange capillary electrochromatography (IECEC).[95] Anions such as I^-, IO_3^- and TcO_4^- are often present in nuclear waste materials in locations such as the US Department of Energy site at Hanford, Washington (USA). For this reason, the analysis of such ions constitutes an important environmental application. In this IECEC method, silica-based particles modified with an anion exchanger, 5-μm Nucleosil™ SB, were packed into 75-μm fused-silica capillaries. UV detection was employed to analyze iodide, iodate and perrhenate (ReO_4^- was used as a surrogate for TcO_4^-). Direct comparisons of the IECEC method with capillary zone electrophoresis (CZE) revealed a 20-fold lower DL for the former owing to enhanced sample capacity. These nuclear waste sites are an increasingly important environmental concern, as leaks from underground storage containers are damaging to the environment. Thus, the need exists for improved analytical techniques with which to monitor these sites. IECEC may provide an alternative to CZE for analyzing low levels of inorganic ions leaching from such sites.

7 FUTURE PERSPECTIVES

In spite of the many advantages afforded by miniaturized chromatographic techniques, it is quite doubtful that micro LC and/or CEC will ever completely supplant conventional scale HPLC as the workhorse in the typical analytical laboratory. A more realistic view is that these techniques are complementary in nature. To be sure, capillary LC will continue to play a central role in biological analyses, especially in protein sequencing. In general, the technique is likely to be employed in cases where sample is volume limited. CEC is currently undergoing a period of intense growth; the number of publications has increased dramatically since 1995. The high separation efficiencies obtained in practice, in agreement with theory, are encouraging. Also promising is the fact that many of the CEC separations reported thus far have employed typical HPLC packings, which will simplify method transfer between the techniques. The future of CEC is most likely in applications where high peak capacity is required (e.g. very complex samples), and/or in rapid highly parallel, automated, perhaps chip-based analysis.

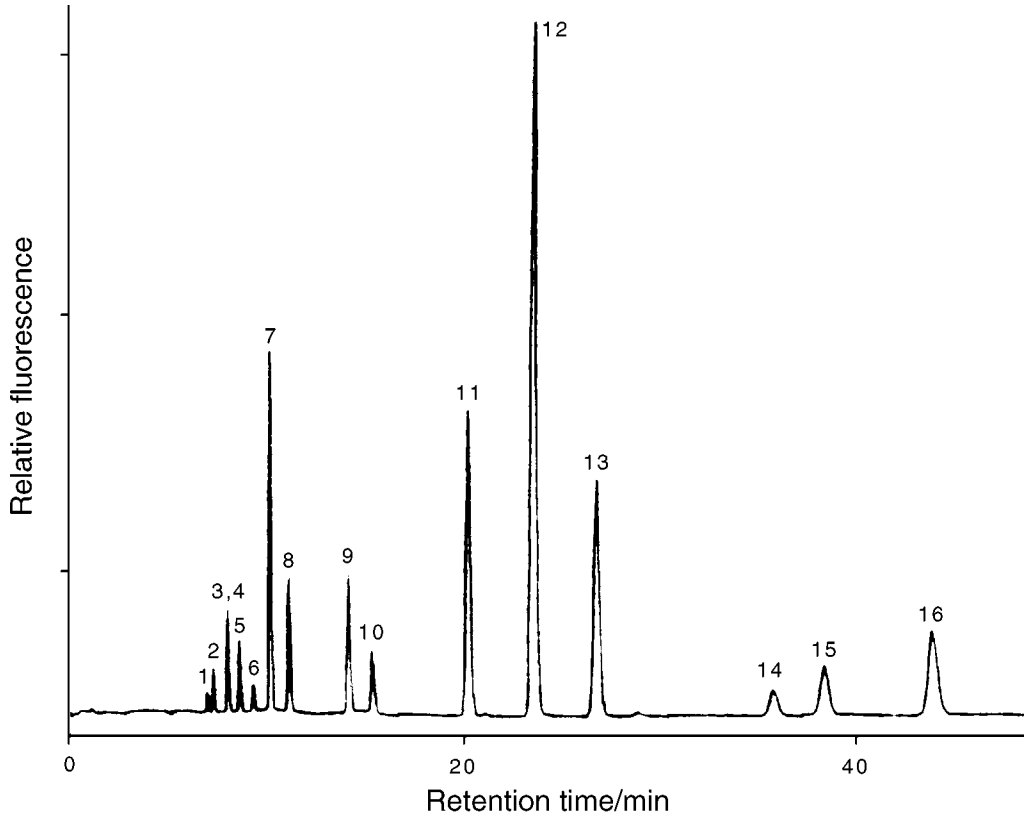

Figure 12 CEC separation of 16 PAHs obtained with a 75-μm ID capillary packed with 3-μm C_{18} particles. The peaks are identified as follows: naphthalene (1), acenaphthylene (2), acenaphthene (3), fluorene (4), phenanthrene (5), anthracene (6), fluoranthene (7), pyrene (8), benz[*a*]anthracene (9), chrysene (10), benzo[*b*]fluoranthene (11), benzo[*k*]fluoranthene (12), benzo[*a*]pyrene (13), dibenz[*a*, *h*]anthracene (14), benzo[*ghi*]perylene (15) and indeno-[1,2,3-*cd*]pyrene (16). (Adapted with permission from Yan et al.[60] Copyright (1995) American Chemical Society.)

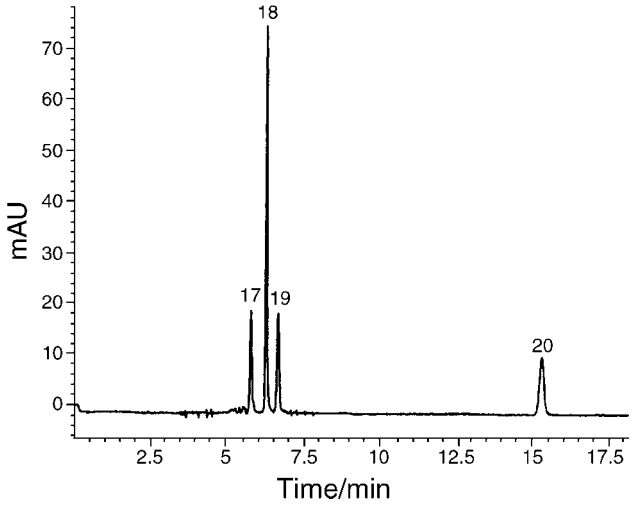

Figure 13 CEC separation of four diuretic drugs obtained with a 50-μm ID capillary packed with 3-μm Hypersil™ C-18. Peaks are labeled as follows: hydrochlorothiazide (17), chlorthalidone (18), hydroflumethiazide (19) and bendroflumethiazide (20). (Adapted from Euerby et al., *J. Microcolumn Sep.* Copyright (1997) by permission of J. Wiley & Sons.)

ABBREVIATIONS AND ACRONYMS

CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CID	Collision-induced Dissociation
CZE	Capillary Zone Electrophoresis
DL	Detection Limit
EOF	Electro-osmotic Flow
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
ID	Internal Diameter
IECEC	Ion Exchange Capillary Electrochromatography
LC	Liquid Chromatography
LIF	Laser Induced Fluorescence
MEKC	Micellar Electrokinetic Capillary Chromatography
micro LC	Microscale Liquid Chromatography
MIP	Molecular Imprinted Polymer

MS	Mass Spectrometry
ODS	Octadecylsilane
OT	Open Tubular
OTLC	Open Tubular Liquid Chromatography
PAH	Polycyclicaromatic Hydrocarbon
PC	Packed Column
PTFE	Polytetrafluoroethylene
PTH	Parathyroid Hormone
RSD	Relative Standard Deviation
SDS	Sodium Dodecyl Sulfate
S/N	Signal-to-noise
UV	Ultraviolet

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis of Peptides • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis

Liquid Chromatography **(Volume 13)**
Capillary Electrophoresis • Column Theory and Resolution in Liquid Chromatography • Micellar Electrokinetic Chromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Atmospheric Pressure Ionization Mass Spectrometry • Liquid Chromatography/Mass Spectrometry

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods

REFERENCES

1. A.J.P. Martin, R.L.M. Synge, 'A New Form of Chromatogram Employing Two Liquid Phases. I. A Theory of Chromatography. II. Application to the Microdetermination of the Higher Monoaminoacids in Proteins', *Biochem J.*, **35**, 1358–1368 (1941).
2. P.B. Hamilton, D.C. Bogue, R.A. Anderson, 'Ion-Exchange Chromatography of Amino Acids. Analysis of Diffusion (Mass Transfer) Mechanisms', *Anal. Chem.*, **32**, 1782–1792 (1960).
3. J.F.K. Huber, J.A.R.L. Hulsman, 'A Study of Liquid Chromatography in Columns', *Anal. Chim. Acta*, **38**(1–2), 305–313 (1967).
4. D. Ishii (ed.), *Introduction to Microscale High-performance Liquid Chromatography*, VCH, New York, 1988.
5. J.N. Done, J.H. Knox, J. Loheac, *Applications of High-speed Liquid Chromatography*, John Wiley & Sons, New York, 1974.
6. C.G. Horvath, B.A. Preiss, S.R. Lipsky, 'Fast Liquid Chromatography: An Investigation of Operating Parameters and the Separation of Nucleotides on Pellicular Ion Exchangers', *Anal. Chem.*, **39**(12), 1422–1428 (1967).
7. J.J. Kirkland, 'High-Speed Liquid Chromatography with Controlled Surface Porosity Supports', *J. Chromatogr. Sci.*, **7**, 7–12 (1969).
8. R.P.W. Scott, P. Kucera, 'Mode of Operation and Performance Characteristics of Microbore Columns for Use in Liquid Chromatography', *J. Chromatogr.*, **169**, 51–71 (1979).
9. H.G. Menet, P.C. Gareil, R.H. Rosset, 'Experimental Achievement of One Million Theoretical Plates with Microbore Liquid Chromatographic Columns', *Anal. Chem.*, **58**(11), 1770–1773 (1984).
10. J.J. DeStefano, H.C. Beachell, 'Performance of Large Diameter Liquid Partition Analytical Columns Packed with Controlled Surface Porosity Support', *J. Chromatogr. Sci.*, **8**, 434–438 (1970).
11. G. Nota, G. Marino, V. Buoncuore, A. Ballio, 'Liquid–Solid Chromatography with Open Glass Capillary Columns. Separation of 1-dimethylamino-5-naphthylsulfonylamino Acids', *J. Chromatogr.*, **46**(1), 103–106, (1970).
12. T. Tsuda, M. Novotny, 'Band-broadening Phenomena in Microcapillary Tubes under the Conditions of Liquid Chromatography', *Anal. Chem.*, **50**(4), 632–634 (1978).
13. K. Hibi, D. Ishii, I. Fujishima, T. Takeuchi, T. Nakanishi, 'Studies of Open Tubular Micro-capillary Liquid Chromatography; I The Development of Open Tubular Micro Capillary Liquid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, (July), 21–27 (1978).
14. T. Tsuda, K. Hibi, T. Nakanishi, T. Takeuchi, D. Ishii, 'Studies of Open Tubular Micro-capillary Liquid Chromatography; II. Chemically Bonded Octadecylsilane Stationary Phase', *J. Chromatogr.*, **158**, 227–232 (1978).
15. R. Tijseen, J.P.A. Bleumer, A. Smit, 'Microcapillary Liquid Chromatography in Open Tubular Columns with Diameters of 10–50 μm. Potential Application to Chemical Ionization Mass Spectrometric Detection', *J. Chromatogr.*, **218**, 137–165 (1981).
16. H. Poppe, 'Some Reflections on Speed and Efficiency of Modern Chromatographic Methods', *J. Chromatogr., A*, **778**, 3–21 (1997).
17. R. Swart, J.C. Kraak, H. Poppe, 'Recent Progress in Open Tubular Liquid Chromatography', *Trends Anal. Chem.*, **16**(6), 332–342 (1997).
18. R.D. Dandeneau, E.H. Zerenner, 'An Investigation of Glasses for Capillary Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **2**(June), 351–356 (1979).
19. F.J. Yang, 'Fused Silica Open Tubular Column for Liquid Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **3**(11), 589–590 (1980).

20. J.W.Jorgenson, R.T.Kennedy, R.L. St Claire, III, J.G. White, P.R. Dluzneski, 'Open Tubular Liquid Chromatography and the Analysis of Single Neurons', *J. Res. Natl. Bur. Stand. (US)*, **93**(3), 403–406 (1988).
21. R.L. St. Claire, D.M. Dohmeier, J.W. Jorgenson, 'Preparation of Reversed Phase Open Tubular Liquid Chromatographic Columns from Borosilicate Glass' *J. Microcolumn Sep.*, **3**, 303–309 (1991).
22. D.M. Dohmeier, J.W. Jorgenson, 'Evaluation of Etched Borosilicate Glass Capillary Columns in Reversed Phase Open Tubular Liquid Chromatography', *J. Microcolumn Sep.*, **3**, 311–317 (1991).
23. T. Tsuda, M. Novotny, 'Packed Microcapillary Columns in High Performance Liquid Chromatography', *Anal. Chem.*, **50**(2), 271–275 (1978).
24. Y. Hirata, M. Novotny, T. Tsuda, D. Ishii, 'Packed Microcapillary Columns with Different Selectivities for Liquid Chromatography', *Anal. Chem.*, **51**(11), 1807–1809 (1979).
25. T. Tsuda, I. Tanaka, G. Nakagawa, 'Packed Microcapillary Columns with 10-μm Silica Gel for Liquid Chromatography', *Anal. Chem.*, **56**(8), 1249–1252 (1984).
26. W.H. Wilson, H.M. McNair, Y.F. Maa, K.J. Hyver, 'The Effect of the Knox–Parcher Ratio in Microcolumn LC', *J. High Resolut. Chromatogr.*, **13**(1), 18–21 (1990).
27. R.T. Kennedy, J.W. Jorgenson, 'Preparation and Evaluation of Packed Capillary Liquid Chromatography Columns with Inner Diameters from 20 to 50 μm', *Anal. Chem.*, **61**(10), 1128–1135 (1989).
28. J.H. Knox, J.F. Parcher, 'Effect of Column to Particle Diameter Ratio on the Dispersion of Unsorbed Solutes in Chromatography', *Anal. Chem.*, **41**(12), 1599–1606 (1969).
29. W.H. Wilson, 'Packed Capillary Columns for Liquid Chromatography', Ph.D. Dissertation, Virginia Polytechnic Institute & State University, 1990.
30. J.C. Gluckman, A. Hirose, V.L. McGuffin, M. Novotny, 'Performance Evaluation of Slurry Packed Capillary Columns for Liquid Chromatography', *Chromatographia*, **17**(6), 303–309 (1983).
31. D. Ishii, K. Hibi, K. Asai, T. Jonokuchi, 'Studies of Micro High-performance Liquid Chromatography II. Application to Gel Permeation Chromatography of Techniques Developed for Micro High-performance Liquid Chromatography', *J. Chromatogr.*, **1**, 147–154 (1978).
32. H.V. Helmholtz, *Wiedemanns Ann. (Ann. Phys. Leipzig)*, **7**, 337 (1877).
33. A. Tiselius, *Trans. Faraday Soc.*, **33**, 524 (1937).
34. S. Hjerten, 'Free Zone Electrophoresis. Theory Equipment, and Applications', *Methods Biochem. Anal.*, **18**, 55–79 (1970).
35. J.W. Jorgenson, K.D. Lukacs, 'High-resolution Separations Based on Electrophoresis and Electroosmosis', *J. Chromatogr.*, **218**, 209–216 (1981).
36. S. Terabe, K. Otsuka, K. Ichikawa, A. Tsuchiya, T. Ando, 'Electrokinetic Separations with Micellar Solutions and Open Tubular Capillaries', *Anal. Chem.*, **56**(1), 111–113 (1984).
37. V. Pretorius, B. Hopkins, J. Schieke, 'Electro-osmosis a New Concept for High-speed Liquid Chromatography', *J. Chromatogr.*, **99**, 23 (1974).
38. J.H. Knox, 'Thermal Effects and Band Spreading in Capillary Electro-Separation', *Chromatographia*, **26**, 329–337 (1988).
39. J.H. Knox, I.H. Grant, 'Electrochromatography in Packed Tubes Using 1.5 to 50 μm Silica Gels and ODS Bonded Silica Gels', *Chromatographia*, **32**(7–8), 317–328 (1991).
40. P.A. Bristow, J.H. Knox, 'Standardization of Test Conditions for HPLC Columns', *Chromatographia*, **10**(6), 279–288 (1977).
41. J.H. Knox, 'Theoretical Aspects of LC with Packed and Open Small-bore Columns', *J. Chromatogr. Sci.*, **18**, 453–461 (1980).
42. L. Snyder, J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley Interscience, New York, 1979.
43. M. Martin, C. Eon, G. Guiochon, 'Study of the Pertinency of Pressure in Liquid Chromatography II. Problems in Equipment Design', *J. Chromatogr.*, **108**, 229–241 (1975).
44. J.C. Giddings, *Dynamics of Chromatography*, Marcel Dekker, New York, 1965.
45. D. Li, V.T. Remcho, 'Perfusive Electroosmotic Transport in Packed Capillary Electrochromatography: Mechanism and Utility', *J. Microcolumn Sep.*, **9**, 389–397 (1997).
46. K.E. Karlsson, M. Novotny, 'Separation Efficiency of Slurry Packed Liquid Chromatography Microcolumns with Very Small Inner Diameters', *Anal. Chem.*, **60**(17), 1662–1665 (1988).
47. S. Hsieh, J.W. Jorgenson, 'Preparation and Evaluation of Slurry Packed Liquid Chromatography Microcolumns with Inner Diameters from 12 to 33 μm', *Anal. Chem.*, **68**(7), 1212–1217 (1996).
48. G. Schomburg, 'Polymer Coating of Surfaces in Column Liquid Chromatography and Capillary Electrophoresis', *Trends Anal. Chem.*, **10**(5), 163–169 (1991).
49. Z.J. Tan, V.T. Remcho, 'Preparation and Evaluation of Bonded Linear Polymethacrylate Stationary Phase for Open Tubular Capillary Electrokinetic Chromatography', *Anal. Chem.*, **69**(4), 581–586 (1997).
50. G. Crescentini, A.R. Mastrogiacomo, 'Liquid Chromatography Capillary Columns Dry-Packed with Polar and Nonpolar Stationary Phases', *J. Microcolumn Sep.*, **3**, 539–545 (1991).
51. C. Yan, US Patent 5,453,163, 'Electrokinetic Packing of Capillary Columns', 26 September 1995.
52. A.M. Fermier, L.A. Colon, 'Packing Columns for Electrochromatography', poster presented at the 20th International Symposium of High Performance Liquid Phase Separations and Related Techniques (HPLC '96) San Francisco, California, 16–20, June 1996.

53. B. Behnke, E. Grom, E. Bayer, 'Evaluation of the Parameters Determining the Performance of Electrochromatography in Packed Capillary Columns', *J. Chromatogr., A*, **716**, 207–213 (1995).
54. E.C. Peters, M. Petro, F. Svec, M.J. Frechet, 'Molded Rigid Polymer Monoliths as Separation Media for Capillary Electrochromatography', *Anal. Chem.*, **69**(17), 3646–3649 (1997).
55. A. Palm, M. Novotny, 'Macroporous Polyacrylamide/Poly(ethylene glycol) Matrixes as Stationary Phases in Capillary Electrochromatography', *Anal. Chem.*, **69**(22), 4499–4507 (1997).
56. C. Ericson, J.L. Liao, K. Nakazato, S. Hjerten, 'Preparation of Continuous Beds for Electrochromatography and Reverse-phase Liquid Chromatography of Low-molecular Mass Compounds', *J. Chromatogr., A*, **767**, 33–41 (1997).
57. G. Chirica, V.T. Remcho, 'New Columns Designed for Capillary Electrochromatography', *Electrophoresis*, **20**, 50–56 (1998).
58. A.G. Ewing, J.M. Mesaros, P.F. Gavin, 'Electrochemical Detection in Microcolumn Separations', *Anal. Chem.*, **6**(9), 527A–535A (1994).
59. R.T. Kennedy, J.W. Jorgenson, 'Quantitative Analysis of Individual Neurons by Open Tubular Liquid Chromatography with Voltammetric Detection', *Anal. Chem.*, **61**(5), 436–441 (1989).
60. C. Yan, R. Dadoo, H. Zhao, R.N. Zare, D.J. Rakestraw, 'Capillary Electrochromatography: Analysis of Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **67**, 2026–2029 (1995).
61. K.B. Tomer, M.A. Moseley, L.J. Deterding, C.E. Parker, 'Capillary Liquid Chromatography/Mass Spectrometry', *Mass Spectrom. Rev.*, **13**, 431–457 (1994).
62. E.C. Huang, J.D. Henion, 'Packed Capillary Liquid Chromatography/Ion Spray Mass Spectrometry Determination of Biomolecules', *Anal. Chem.*, **63**(7), 732–739 (1991).
63. R.A. Henderson, A.L. Cox, K. Sakaguchi, E. Apella, J. Shabanowitz, D.F. Hunt, V.H. Engelhard, 'Direct Identification of an Endogenous Peptide Recognized by Multiple HLA-A2.1-Specific Cytotoxic T Cells', *Proc. Natl. Acad. Sci. USA*, **90**, 10275–10279 (1993).
64. P.R. Griffin, J.A. Coffman, L.E. Hood, J.R. Yates, III, 'Structural Analysis of Proteins by Capillary HPLC Electrospray Tandem Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **111**, 131–149 (1991).
65. J.F. Banks, 'Protein Analysis by Packed Capillary Liquid Chromatography with Electrospray Ionization and Time-of-Flight Mass Spectrometry Detection', *J. Chromatogr., A*, **786**, 67–73 (1997).
66. D.F. Hunt, J.E. Alexander, A.L. McCormack, P.A. Martino, H. Michel, J. Shabanowitz, N. Sherman, M.A. Moseley, J.W. Jorgenson, K.B. Tomer, *Mass Spectrometric Methods for Protein and Peptide Sequence Analysis, Techniques in Protein Chemistry II*, Academic Press, New York, 441–453, 1991.
67. J.R. Yates, III, A.L. McCormack, A.J. Link, D. Schieltz, J. Eng, L. Hays, 'Future Prospects for the Analysis of Complex Biological Systems Using Micro-column Liquid Chromatography–Electrospray Mass Spectrometry', *Analyst*, **121**, 65R–76R (1996).
68. D.B. Gordon, G.A. Lord, D.S. Jones, 'Development of Packed Capillary Column Electrochromatography/Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 544 (1994).
69. K. Schmeer, B. Behnke, E. Bayer, 'Capillary Electrochromatography–Electrospray Mass Spectrometry: A Micro-analysis Technique', *Anal. Chem.*, **67**(20), 3656–3658 (1995).
70. G.A. Lord, D.B. Gordon, L.W. Tetler, C.M. Carr, 'Electrochromatography–Electrospray Mass Spectrometry of Textile Dyes', *J. Chromatogr., A*, **700**(1–2), 27–33 (1995).
71. S.J. Lane, R. Boughtflower, C. Paterson, T. Underwood, 'Capillary Electrochromatography/Mass Spectrometry: Principles and Potential for Application in the Pharmaceutical Industry', *Rapid Commun. Mass Spectrom.*, **9**, 1283–1287 (1995).
72. J. Ding, P. Vouros, 'Capillary Electrochromatography and Capillary Electrochromatography–Mass Spectrometry for the Analysis of DNA Adduct Mixtures', *Anal. Chem.*, **69**(3), 379–384 (1997).
73. J.Z. Tan, V.T. Remcho, 'Molecular Imprint Polymers as Highly Selective Stationary Phases for Open Tubular Liquid Chromatography and Capillary Electrochromatography', *Electrophoresis*, **19**, 2055–2060 (1998).
74. J.M. Lin, T. Nakagama, K. Uchiyama, X.Z. Wu, T. Hobo, 'Capillary Electrochromatographic Separation of Amino Acid Enantiomers with Molecular Imprinted Polymer as Chiral Recognition Agents', *Fresenius' J. Anal. Chem.*, **357**, 130 (1997).
75. L. Schweitz, L.I. Andersson, S. Nilsson, 'Capillary Electrochromatography with Predetermined Selectivity Obtained through Molecular Imprinting', *Anal. Chem.*, **69**(6), 1179–1183 (1997).
76. B. Behnke, E. Bayer, 'Pressurized Gradient Electro-high-performance Liquid Chromatography', *J. Chromatogr.*, **680**(1), 93–98 (1994).
77. C. Huber, G. Choudhary, C. Horvath, 'Capillary Electrochromatography with Gradient Elution', *Anal. Chem.*, **69**(21), 4429–4436 (1997).
78. M.R. Taylor, P. Teale, 'Gradient Capillary Electrochromatography of Drug Mixture with UV and Electrospray Ionization Mass Spectrometric Detection', *J. Chromatogr., A*, **768**, 89–95 (1995).
79. C.Yan, R.Dadoo, R.N.Zare, D.J.Rakestraw, D.S.Anex, 'Gradient Elution in Capillary Electrochromatography', *Anal. Chem.*, **68**(17), 2726–2730 (1996).
80. M.H. Chen, C. Horvath, 'Temperature Programming and Gradient Elution in Reversed-phase Chromatography with Packed Capillary Columns', *J. Chromatogr., A*, **788**, 57–61 (1997).

81. J.E. MacNair, K.C. Lewis, J.W. Jorgenson, 'Ultrahigh-pressure Reversed-phase Liquid Chromatography in Packed Capillary Columns', *Anal. Chem.*, **69**(6), 983–989 (1997).
82. J.P.C. Vissers, H.A. Claessens, C.A. Cramers, 'Microcolumn Liquid Chromatography: Instrumentation, Detection and Applications', *J. Chromatogr., A*, **779**, 1–28 (1997).
83. B.R. Cooper, J.A. Jankowski, D.J. Leszczyszyn, R.M. Wightman, J.W. Jorgenson, 'Quantitative Determination of Catecholamines in Individual Cells by Reversed-phase Microcolumn Liquid Chromatography with Electrochemical Detection', *Anal. Chem.*, **64**(6), 691–694 (1992).
84. R. Hu, T. Takeuchi, J-Y. Jin, T. Miwa, 'Separation of Enantiomers by Microcolumn Liquid Chromatography with Methylated β-Cyclodextrin as Mobile Phase Additive', *Anal. Chim. Acta*, **295**, 173–179 (1994).
85. R. Straub, M. Linder, R.D. Voyksner, 'Determination of β-Lactam Residues in Milk Using Perfusive-Particle Liquid Chromatography Combined with Ultrasonic Nebulization Electrospray Mass Spectrometry', *Anal. Chem.*, **66**, 3651–3658 (1994).
86. M.R. Euerby, D. Gilligan, C.M. Johnson, S.C.P. Roulin, P. Myers, K.D. Bartle, 'Applications of Capillary Electrochromatography in Pharmaceutical Analysis', *J. Microcolumn Sep.*, **9**, 373–387 (1997).
87. K.E. Karlsson, D. Wiesler, M. Alasandro, M. Novotny, '7-(Chlorocarbonyl) methoxy-4-methylcoumarin: A Novel Fluorescent Reagent for the Precolumn Derivatization of Hydroxy Compounds in Liquid Chromatography', *Anal. Chem.*, **57**, 229–234 (1985).
88. B. Malavasi, V. Ascalone, 'Suitability of Automated Capillary and Micro Liquid Chromatography for Routine Determination of Drugs in Human Plasma Samples from Clinical Pharmacokinetics Investigations', *J. High Resolut. Chromatogr.*, **19**, 503–510 (1996).
89. J.J. Hagen, K.A. Washco, C.A. Monnig, 'Determination of Retinoids by Reversed-phase Capillary Liquid Chromatography with Amperometric Electrochemical Detection', *J. Chromatogr., B, Biomedical Applications*, **677**(2), 225–231 (1996).
90. M.M. Robeson, M.G. Cikalo, P. Myers, M.R. Euerby, K.D. Bartle, 'Capillary Electrochromatography: A Review', *J. Microcolumn Sep.*, **9**, 357–372 (1997).
91. N.W. Smith, M.B. Evans, 'The Efficient Analysis of Neutral and Highly Polar Pharmaceutical Compounds Using Reversed-phase and Ion-exchange Electrochromatography', *Chromatographia*, **41**, 197 (1995).
92. P. Sandra, A. Dermaux, V. Ferraz, M.M. Dittmann, G. Rozing, 'Analysis of Triglycerides by Capillary Electrochromatography', *J. Microcolumn Sep.*, **9**, 409–419 (1997).
93. J.J. Pesek, M.T. Matyska, 'Separation of Tetracyclines by High-performance Capillary Electrophoresis and Capillary Electrochromatography', *J. Chromatogr., A*, **736**, 313–320 (1996).
94. M.R. Taylor, P. Teale, S.A. Westwood, D. Perrett, 'Analysis of Corticosteroids in Biofluids by Capillary Electrochromatography with Gradient Elution', *Anal. Chem.*, **69**(13), 2554–2558 (1997).
95. D. Li, H.H. Knobel, V.T. Remcho, 'Capillary Zone Electrophoresis and Ion-exchange Capillary Electrochromatography: Analytical Tools for Probing the Hanford Nuclear Site Environment', *J. Chromatogr., B*, **695**, 169–174 (1997).

Normal-phase Liquid Chromatography

William T. Cooper
Florida State University, Tallahassee, USA

1 Introduction **1**
1.1 Normal-phase Liquid Chromatography 1
1.2 Normal-phase Liquid Chromatography Compared with Reversed-phase Liquid Chromatography 2
2 Fundamentals of Retention **2**
2.1 Liquid–Solid Chromatography: the Adsorption–Displacement Model 2
2.2 Localization and Secondary Solvent Effects 3
3 Stationary Phases **4**
3.1 Solid Adsorbents 4
3.2 Bonded Phases 4
4 Mobile Phases **5**
4.1 Common Solvents Used in Normal-phase Liquid Chromatography 5
4.2 Solvent Strength 5
4.3 Solvent Selectivity 6
4.4 Aqueous Mobile Phases 8
5 Optimization Strategies **8**
5.1 Mobile Phase Optimization 8
5.2 Stationary Phase Choices 9
6 Techniques **9**
6.1 Isocratic Separations 9
6.2 Gradient Elution 9
6.3 Column Switching, Stationary Phase Programming, and Two-dimensional Liquid Chromatography 11
7 Applications **12**
7.1 Isomers 12
7.2 Environmental 12
7.3 Biomedical and Pharmaceutical 13
Acknowledgments **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **14**

Normal-phase liquid chromatography (NPLC) is a technique that uses columns packed with polar stationary phases combined with nonpolar or moderately-polar mobile phases to separate the components of mixtures. The rate at which individual solutes migrate through NPLC columns is primarily a function of their polarity. Less polar solutes move the fastest and therefore exit the column and are detected first, followed by solutes of increasing polarity which move more slowly. However, polarity can sometimes play a secondary role relative to a solute's ability to experience a specific interaction with active sites on the stationary phase surface. The importance of these specific solute–stationary phase interactions in NPLC gives it some unique advantages over the more widely practised reversed-phase liquid chromatography (RPLC) technique. RPLC utilizes nonpolar stationary phases and aqueous-based polar mobile phases, and the elution order of solutes in a mixture is related to their hydrophobicity, not polarity; more polar solutes move the fastest and appear first, followed by solutes of decreasing polarity. RPLC is useful for separating mixtures in which components differ in molecular weight and/or water solubility. However, NPLC is preferred for many separation problems, including those in which the water solubility of sample compounds is limited. In addition, NPLC is a better technique for separating compounds that differ in the number or character of functional groups and is particularly useful for separating many types of isomers.

1 INTRODUCTION

1.1 Normal-phase Liquid Chromatography

The extent to which solutes are slowed by interactions with the stationary phase as they move through a liquid chromatography (LC) column is broadly referred to as retention. In NPLC retention is the result of a complicated set of specific molecular interactions, including interactions between the solute and stationary phase, mobile phase and stationary phase, and solute and mobile phase. In NPLC the stationary phase is usually more polar than the flowing mobile phase and the order of retention generally tracks the polarity of components in a mixture: less polar first, followed by solutes of increasing polarity.

The absolute magnitude of retention in a particular NPLC column, as in most LC techniques, is normally controlled by the ''solvent strength'' of the mobile phase. Note that in LC, the mobile phase is often referred to as the solvent. An increase in the solvent strength of the mobile phase results in faster migration of solutes and a general decrease in solute retention times.

In NPLC, solute and mobile phase molecules compete for polar active sites on the stationary phase surface. Changes in mobile phase polarity are thus

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

used to adjust solvent strength and optimize retention times. With modern liquid chromatographs solvent strength is normally controlled by combining a nonpolar weak solvent such as hexane with moderately polar strong solvents such as isopropanol. In binary (or in ternary/quaternary) mobile phases, the weaker solvent is usually designated the ''A'' solvent, while the additional solvent(s) are known as modifiers and designated the B, C, etc. solvents. Generally, increasing the modifier content in the mobile phase increases solvent strength and thus decreases retention.

The relative positions of and spacing between solutes in a normal-phase liquid chromatogram can be optimized through changes in both solvent strength and ''solvent selectivity''. Solvent selectivity refers to the totality of molecular interactions a solvent is capable of, including dispersion, dipole–dipole, and acid–base and hydrogen-bonding interactions. In NPLC, varying the magnitude (solvent strength) or nature (solvent selectivity) of the solvent–stationary phase interactions by changing the mobile phase modifier will affect solute retention. For example, resolution enhancement can often be accomplished by changing from a hexane/isopropanol mobile phase to hexane/dichloroethane, or by changing the amount of a very strong polar modifier such as methyl-*tert*-butyl ether (MTBE) in a hexane/MTBE binary mixture.

1.2 Normal-phase Liquid Chromatography Compared with Reversed-phase Liquid Chromatography

By the mid-1970s, the ability to attach virtually any functional group, polar or nonpolar, to silica particles and make bonded stationary phases led very quickly to the rapid development of RPLC. In contrast to NPLC, RPLC utilizes nonpolar bonded stationary phases (e.g. octadecylsilane (ODS), often abbreviated C_{18}) and aqueous-based polar mobile phases. Retention is based largely on a solute's hydrophobicity, and elution order is thus generally the reverse of that in NPLC: more polar solutes first, followed by solutes of decreasing polarity. RPLC proved to be applicable to a wide variety of separation problems and now represents about 80% of all high-performance liquid chromatography (HPLC) applications. It is useful for separating mixtures in which components differ in molecular weight and/or water solubility.

However, NPLC is still a useful approach for many separation problems, including those in which the water solubility of sample compounds is limited. In addition, NPLC is the preferred mode for separating compounds that differ in the number or character of functional groups, and is particularly useful for separating many types of isomers. Polar bonded phases useful for NPLC are now available, providing a wider array of stationary phase choices than traditional liquid–solid chromatography (LSC) adsorbents such as silica and alumina. Because of the specificity of retention in NPLC, the use of such polar bonded phases in the normal phase mode actually provides a wider variety of optimization choices than are possible with reversed-phase columns.

2 FUNDAMENTALS OF RETENTION

2.1 Liquid–Solid Chromatography: the Adsorption–Displacement Model

The most successful models of retention in LSC are those of Snyder et al.[1–4] and Soczewinski.[5] Both models describe retention as resulting from competitive adsorption between solute and solvent (mobile phase) molecules for active sites on the stationary phase surface. Each assumes a surface monolayer of solvent and/or solute molecules, but Snyder's model further assumes a homogeneous surface so that adsorption energies of both solute and solvent molecules are constant.

The stoichiometry of solute–solvent competition in Snyder's model can be expressed by Equation (1):

$$X_m + nS_a \rightleftharpoons X_a + nS_m \tag{1}$$

where m and a refer to solute (X) and solvent (S) in the mobile and adsorbed phases, respectively. The coefficient n represents the ratio of areas occupied by solute and solvent molecules, respectively, on the adsorbent surface and reflects the fact that a solute molecule must displace all solvent molecules that restrict its access to an adsorption site. The Soczewinski model is similar to Snyder's, but it does not assume a homogeneous surface. Rather, the surface is considered to be energetically heterogeneous, with adsorption occurring entirely at high-energy sites, leading to discrete, one-to-one complexes of the form (Equation 2):

$$X_m + n'S\text{–}A^* \rightleftharpoons n'S_m + X\text{–}A^* \tag{2}$$

where A^* is an active surface site and n' refers to the number of substituents on a solute molecule that are capable of simultaneously interacting with the active sites. Equation (2) thus takes into account the possibility of multisite attachment. By assuming a binary mobile phase consisting of a weak, nonpolar solvent (e.g. hexane) and a strong, polar solvent modifier (e.g. isopropanol), the adsorption of the weak solvent can be ignored and retention predicted by Equation (3):

$$\log k = \text{constant} - n \log N_b \tag{3}$$

Snyder has pointed out a more convenient form of Equation (3) that is equally reliable (Equation 4):

$$\log k = \text{constant} - n \log(\%B) \tag{4}$$

In Equations (3) and (4), k is the capacity factor and N_b the mole fraction of the strong solvent in the mobile phase. The capacity factor is a relative parameter which normalizes solute retention volume (V_r) or retention time (t_r) in terms of column void volume (V_m) or void time (t_0) (Equation 5):

$$k = \frac{V_r - V_m}{V_m} = \frac{t_r - t_0}{t_0} \tag{5}$$

Equations (3) and (4) predict that a plot of $\log k$ against $\log N_b$ or $\log(\%B)$ should yield a straight line. In the Snyder model, the slope of this line should be the ratio of the molecular areas of solute and solvent, whereas Soczewinski's model predicts that the slope is the number of strongly adsorbing substituent groups on the solute.

2.2 Localization and Secondary Solvent Effects

Neither the Snyder nor Soczewinski model is entirely satisfactory as presented for describing retention in NPLC, particularly when bonded stationary phases are used. Such normal bonded phase (NBP) LC columns do contain strongly adsorbing active sites (Soczewinski model), but solute molecular area, not just polar substituents, is known to play an important role in competitive adsorption (Snyder model). In addition, neither model accounts for so-called secondary solvent effects. These effects, which arise from solute–solvent interactions in both the mobile and adsorbed phases, lead to some very useful changes in retention. A rigorous description of retention in NPLC, and thus separation optimization, must therefore include some discussion of the highly specific solute–stationary phase, solvent–stationary phase, and solute–solvent interactions encountered with the types of polar surfaces and molecules commonly used in NPLC. Detailed reviews of these effects can be found in Snyder[6] and Dorsey and Cooper.[7]

Localization refers to the phenomenon which occurs when the strength of strong polar interactions (e.g. dipole–dipole, hydrogen bonding) between solute or solvent molecules and active sites on the stationary phase surface exceed the thermal energy of solute or solvent. When this happens, the solute/solvent molecules are preferentially located in the vicinity of the active site, rather than delocalized in a surface monolayer. The competition between solutes and solvent molecules for localized positions on the stationary phase leads to delocalization of previously localized molecules.

Localization phenomena give rise to some of the most interesting and useful retention behavior in NPLC. For example, it has been established that, at least for solid adsorbents such as silica and alumina, solvent localization is more important in determining mobile phase selectivity than a solvent's acidity, basicity or dipolarity. Snyder has presented a solvent optimization strategy that is largely based on these localization effects.[8] It should be noted that this delocalization competition occurs between solutes and the polar modifier, since the weaker solvent (normally hexane) does not localize.

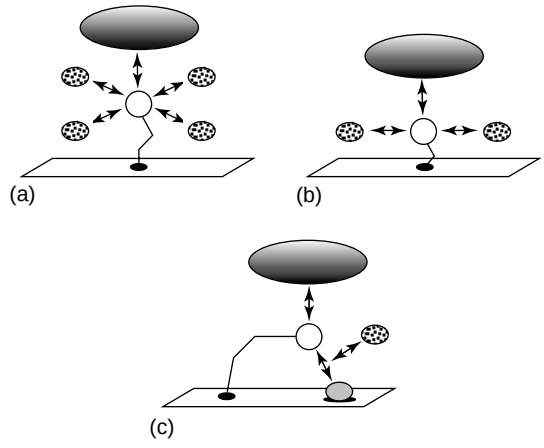

Figure 1 Representations of various delocalization processes in NPLC. (a) Site-competition delocalization. (b) Restricted-access delocalization. (c) Mixed-site delocalization. Active sites on the stationary phase surface are represented by small, clear circles; polar solute molecules by large, shaded spheroids; polar solvent modifier molecules by stippled circles; and residual silanols by shaded circles on the surface. (Reproduced by permission of the American Chemical Society from J.G. Dorsey, W.T. Cooper, *Anal. Chem.*, **66**, 857A–867A (1994).)

There are actually a number of different types of localization commonly encountered in NPLC, and these are depicted in Figure 1.[7] Site-competition delocalization occurs when localizing polar solutes and modifier molecules have unrestricted access to the active surface site. In this case, an increase in modifier content in the mobile phase will weaken solute localization and retention will decrease to a greater extent than predicted from a solvent strength value determined with nonlocalizing solutes. For example, MTBE, like most basic, localizing solvents, produces a stronger mobile phase than predicted when used with silica and certain bonded phase columns. Correcting for such an effect is straightforward because the solvent strength of the modifier can be adjusted. However, this must be done on a specific solvent–stationary phase basis. The necessary solvent strengths and/or corrections for the most widely used normal-phase columns and polar modifiers have been tabulated.[8,9]

Restricted-access delocalization (Figure 1b), which occurs when polar modifier and solute molecules compete for localization on rigid, fixed surface-active sites where

access is highly limited, is more difficult to predict and quantify. When the modifier content is low it will localize and disrupt solute localization and the effective solvent strength of the modifier will be relatively large. However, with increasing coverage of the adsorbent surface, a point is reached at which adsorbed modifier molecules interfere with further modifier localization. The effective solvent strength of the modifier thus decreases with increasing surface coverage. Since this decrease in modifier solvent strength is non-linear, a simple modification to modifier solvent strength is not suitable. Rather, a delocalization function is required which can be used to adjust the modifier solvent strength.[10–12]

A localization effect that is unique to bonded-phase adsorbents is mixed-site delocalization. This occurs because unreacted, ''residual'' silanol groups (Si–OH) on the underlying silica support can influence the character of the bonded phase through hydrogen bonding with its polar functional group. This hydrogen bonding can be disrupted by the presence of strongly localizing modifiers (Figure 1c). This mixed-site delocalization effect is most pronounced in weaker, less polar bonded phase columns such as cyano-silica.[13]

Secondary solvent effects arise from specific, polar interactions between solute and solvent molecules. While they are normally considered ''secondary'' in their importance on retention relative to solute–solvent–stationary phase interactions, they often provide an effective means of altering retention selectivity. There are three different types of solvent effects observed in NPLC. The first and most straightforward occurs in the mobile phase when there is a specific polar interaction between solutes and the mobile phase modifier. This type of solvent effect, which is not balanced by a corresponding effect in the stationary phase, decreases retention of the interacting solute and is thus considered a negative effect.

A second type of solvent effect is observed when there is a strong affinity between the stationary phase and polar modifiers, resulting in an excess concentration of modifier in the surface adsorbed layer. If a solute experiences a strong interaction with this modifier, and the concentration of modifier is higher in the stationary phase than in the mobile phase, the solute–solvent interaction will increase retention. This second type is thus considered a positive effect because it increases retention.

The third type of solvent effect is also due to excess modifier in the stationary phase, but in this case the modifier is so strongly adsorbed it is sterically hindered from interacting with an adsorbed solute which is nonetheless experiencing a strong interaction with the modifier in the mobile phase. This unbalanced interaction decreases retention and is thus a negative effect.

3 STATIONARY PHASES

3.1 Solid Adsorbents

Silica gel is by far the most popular stationary phase for LSC. Its rigidity provides structural stability at the high pressures used in HPLC and it can be synthesized with a wide variety of particle sizes, surface areas, and pore diameters. For analytical separations, 5–10 μm porous silica is most commonly used, but larger silica with less porosity can be obtained.

In addition to its advantageous physical attributes, silica provides a very polar surface with a high concentration of active sites. The active sites on silica are Si–OH groups (silanols), and most chromatographic-grade silicas have Si–OH concentrations of $\sim 8\,\mu mol\,m^{-2}$. Unfortunately, not all silanols are identical, which leads to a significantly heterogeneous surface and, in many cases, ''peak tailing'' of highly polar solutes. For this reason, small amounts of water are often added to NPLC mobile phases to deactivate the higher-energy silanols.

Silanols are weakly acidic (pK_a 5–7), and thus care in the choice of the type of silica is necessary if basic solutes are to be separated. So-called Type A silicas contain a high concentration of highly-acidic ''free'' silanols and are thus not appropriate stationary phases for strongly basic solutes unless special conditions are employed. Type B silicas are dominated by less acidic geminal and associated silanols. Columns made from these Type B silicas have been shown to be useful for normal phase separations of highly polar and basic solutes.[14]

For samples that are strongly basic or acid-sensitive, alumina is the adsorbent of choice. Alumina is weakly basic, and strongly acidic and base-sensitive compounds cannot normally be chromatographed on it. Retention of neutral solutes is similar on alumina and silica, but alumina lacks many of the physical attributes of silica that have led to silica's prominence. Thus, unless precluded by its acidic character, silica is normally the LSC stationary phase of choice.

Other adsorbents which have been suggested as possible alternatives to silica and alumina include zirconia and titania. Both of these solids appear to be more basic than silica, much like alumina.[15] While each has some unique chromatographic properties, neither adsorbent is widely used today.

3.2 Bonded Phases

Chemically-bonded stationary phases have two distinct advantages over solid adsorbents for NPLC. First, derivatization of silica gel with a polar organic functional group provides a much more homogeneous surface than bare silica. Even though steric hindrance by adjacent ligands normally prevents complete reaction

of all surface silanols, a bonded phase packing is much more homogeneous than bare silica and deactivation procedures such as addition of water to the mobile phase are no longer necessary.

A second, more subtle, but ultimately more significant advantage of NPLC bonded stationary phases is that they allow chromatographic selectivity to be enhanced through adjustment of both the stationary and mobile phases. Because retention in NPLC is governed primarily by competitive solute–stationary phase and solvent–stationary phase interactions, changing the column as well as the mobile phase can optimize absolute and relative retention of solutes in a mixture. This is in contrast to RPLC, where the stationary phase, while not a completely passive participant in retention, is nevertheless much less important than the mobile phase.

Virtually all NPLC bonded phases are prepared by reacting a monochloroalkylsilane with silanols on the surface of silica gel. This reaction results in an alkylsilane attached to the silica surface via a silyl ether bond (Si–O–Si). The desired functional group used to produce the specific character of the bonded phase is located at the end of the alkyl chain opposite from the ether bond. The alkyl chain is referred to as the ''spacer''; in NPLC these spacers are normally only 2–3 carbon atoms in length.

The most widely-used NPLC bonded phases are cyanopropyl-silica (cyano), aminopropyl-silica (amino), and 1,2-dihydroxypropyl-silica (diol). In addition, a number of other functional groups have been used to produce highly selective NBP columns (e.g. phenylpropyl columns are useful for some highly specific applications). However, the cyano-, amino-, and diol-silica columns are now the most popular. This is due in great part to their widely different retention characteristics and moderate polarity.

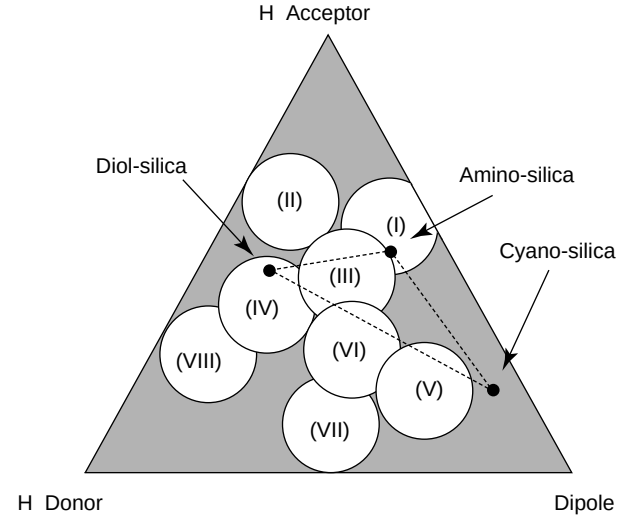

Figure 2 Stationary phase selectivity triangle formed by the combination of diol-, amino- and cyano-silica NBP columns.

The diol column is acidic, the amino column moderately basic, and the cyano column a moderate dipole–dipole interactor. These three columns together provide virtually all the retention characteristics necessary for efficient optimization of sample resolution. It has even been suggested that these three phases form a ''stationary phase selectivity triangle'' similar to the solvent selectivity triangle used to pick mobile phase solvents (Figure 2).[4,16] In chromatography, a selectivity triangle is used as a guide for producing the most dramatic changes in chemical characteristics of mobile and/or stationary phases during optimization experiments. Cyano, diol and amino columns appear to provide these dramatic differences in retention characteristics.

4 MOBILE PHASES

4.1 Common Solvents Used in Normal-phase Liquid Chromatography

A wide variety of organic solvents have been used as mobile phases in NPLC, ranging in polarity from fluorocarbons to methanol. Liquids with low boiling points and low viscosities are preferred. In addition, it should be noted that in normal-phase chromatography the mobile phase should be less polar than all sample components. In recent years there has been a tendency to choose from a rather limited number of available solvents, making solvent optimization more efficient. Hexane (or pentane) has become the nonpolar solvent of choice, although isooctane is sometimes used. For very nonpolar samples, fluorocarbons may be required.

Most NPLC separations require a mobile phase more polar than hexane. While this can be accomplished with a single solvent, it is actually more convenient to combine hexane with a much stronger solvent and adjust the relative proportions of the two solvents to obtain the correct solvent strength. In such binary mixtures, hexane is the weak solvent (normally designated the A solvent) and the polar solvent the strong solvent (the B solvent). The strong solvent is also frequently referred to as the ''modifier''. Solvents commonly used as modifiers include chloroform, methylene chloride, dichloroethane, ethyl acetate, MTBE, tetrahydrofuran (THF), and alcohols such as methanol or isopropanol. When very strong (i.e. very polar) mobile phases are required, hexane can be omitted and one of the polar modifiers used alone or in combination with another modifier.

4.2 Solvent Strength

Retention in NPLC is determined by competitive adsorption between solute and solvent molecules at the stationary phase surface. Thus, the greater the affinity of

the solvent for the stationary phase, the greater is its solvent strength. However, because of the polar interactions involved in NPLC and the specific localization effects that result, a solvent's apparent strength can vary from column to column. There is thus no absolute scale of solvent strength in the normal-phase mode that is useful for all the different columns available. However, a number of empirical scales are available for predicting and/or controlling retention. In LSC the adsorption solvent strength parameter ε^0 is most useful. This parameter is defined as the adsorption energy per unit cross-sectional area of solvent. Table 1 contains ε^0 values for commonly used normal-phase solvents.[8] For normal bonded-phase LC, empirically-determined solvent strengths have been published for dichloromethane, chloroform and MTBE when used with cyano, amino and diol columns.[9] These solvent strengths are summarized in Table 2.

As noted, it is much more convenient to mix solvents of different polarity in the correct proportions than to find a single solvent of just the right solvent strength and selectivity. This involves combining a nonpolar solvent with a polar solvent which alone would be too strong. These two solvents are then blended empirically to give the correct solvent strength and optimum range of solute capacity factors. Unfortunately, solvent strength in NPLC is not a linear function of the volume percent of modifier in the mobile phase, but is described by Equation (6).

$$E_{ab} = E_a + \frac{\log[N_b 10^{\alpha n_b (E_b - E_a)} + 1 - N_b]}{\alpha n_b} \qquad (6)$$

Here, E_{ab} is the solvent strength of the mixture, E_a and E_b the pure solvent strengths of solvents a and b, respectively, N_b the mole fraction and n_b the molecular area of the modifier, and α an adsorbent activity factor normally taken as unity in modern analytical columns. However, Equation (6) can be greatly simplified if hexane is the weak solvent (A) and the fraction of modifier (B) is not too low (Equation 7).

$$E_{ab} = E_b + \frac{\log N_b}{\alpha n_b} \qquad (7)$$

Careful examination of Equation (6) or (7) reveals that binary solvent strengths increase rapidly with increasing modifier content at low modifier levels and less dramatically at higher modifier levels.

It is sometimes necessary to change the polar B solvent to optimize solvent selectivity (see next section). When carrying out such selectivity optimization experiments it is convenient to maintain a constant solvent strength E_{ab}. This can be accomplished by using Equation (6) or (7) and the appropriate solvent strengths and physical parameters. Alternatively, published solvent-strength nomographs provide the same information for silica[17] and cyano or amino bonded-phase columns.[8]

Table 1 Solvent strength values of some common normal-phase solvents when used with solid adsorbents[a]

Solvent	Solvent strength, ε^0	
	Silica	Alumina
Heptane	0.01	0.01
Hexane	0.01	0.01
Isooctane	0.01	0.01
1-Chlorobutane	0.20	0.26
Chloroform	0.26	0.40
Dichloromethane	0.32	0.42
Ethyl acetate	0.38	0.58
THF	0.44	0.57
Acetonitrile	0.50	0.65
Methanol	~0.7	0.95

[a] Reproduced from L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 1997, with the permission of Wiley-Interscience, New York.

Table 2 Solvent strength values of some common normal-phase solvents when used with bonded-phase columns[a]

Solvent	Pure solvent strength		
	Cyano	Amino	Diol
Chloroform	0.106	0.143	0.097
Dichloromethane	0.120	0.141	0.096
MTBE	0.049	0.124	0.071

[a] Reproduced with permission from P.L. Smith, W.T. Cooper, 'Retention and Selectivity in Cyano-, Amino- and Diol-silica Normal Bonded Phase HPLC Columns', *J. Chromatogr.*, **410**, 249–265 (1987).

4.3 Solvent Selectivity

Once the correct solvent strength has been established (i.e. a solvent strength that produces sample capacity factors in the range of 0.5–20[8]), it is frequently necessary to further enhance sample resolution through optimization of solvent selectivity. Solvent selectivity refers to the totality of molecular interactions a solvent is capable of, including dispersion, dipole–dipole, and acid–base and hydrogen-bonding interactions. In NPLC, these molecular interactions are exploited in both the stationary and mobile phases. In the stationary phase solvent localization plays a critical role in determining retention selectivity, particularly when silica columns are used. Specific molecular interactions in the mobile phase generally play a lesser role in controlling selectivity and are thus referred to as secondary solvent effects. However, in the weaker, less polar bonded phase columns (cyano, amino, diol), these secondary solvent effects

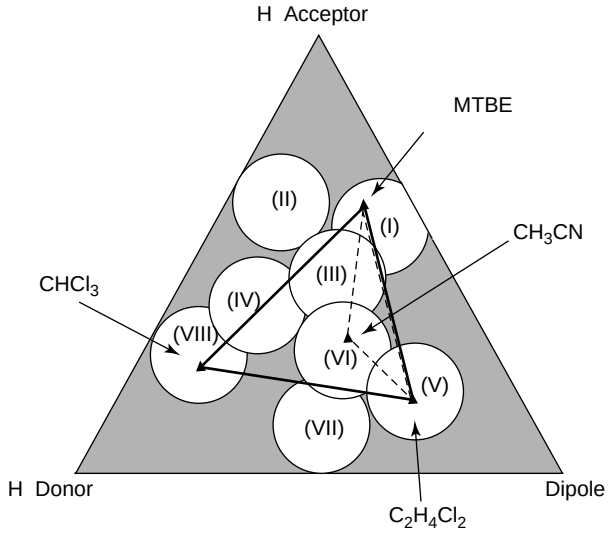

Figure 3 Positions and ''inner selectivity triangle'' (———) of common NPLC modifiers useful with bonded-phase columns. Positions and ''inner selectivity triangle'' (- - - - -) of NPLC modifiers useful with solid adsorbent columns.

can sometimes provide selectivity enhancement not obtainable through changes in solvent localization alone.

The fundamental rule of mobile phase selectivity optimization in LC is that the greatest change in selectivity occurs when molecular interactions between solute, solvent and stationary phase are dramatically changed. These dramatic changes in solvent selectivity can be visualized using the selectivity triangle[18,19] which positions solvents in a triangle based on the relative importance of their proton acceptor, proton donor, and dipole–dipole characteristics. The most dramatic changes in solvent characteristics occur when a solvent from one apex of the triangle (e.g. proton donors) is replaced by one from another apex (e.g. proton acceptors). This has led to the use of three modifiers for solvent selectivity optimization in LC. Since these three modifiers represent regions of the selectivity triangle, they are sometimes referred to as the ''inner triangle''. In NPLC, the inner triangle would normally be formed with chloroform (proton donor), MTBE or THF (proton acceptor), and dichloromethane or dichloroethane (dipole). The solvent selectivity triangle with the normal-phase inner triangle superimposed is depicted in Figure 3. This ''inner triangle'' has been shown to produce large selectivity changes in cyano, amino and diol bonded-phase columns.[9,16]

Solvent selectivity in silica columns, however, is dominated by the ability of a solvent to localize. Furthermore, because of the acidic nature of the active sites on silica (silanols), basic localizing solvents show different selectivities than nonbasic localizing solvents. These observations led to an alternative selectivity triangle which emphasizes

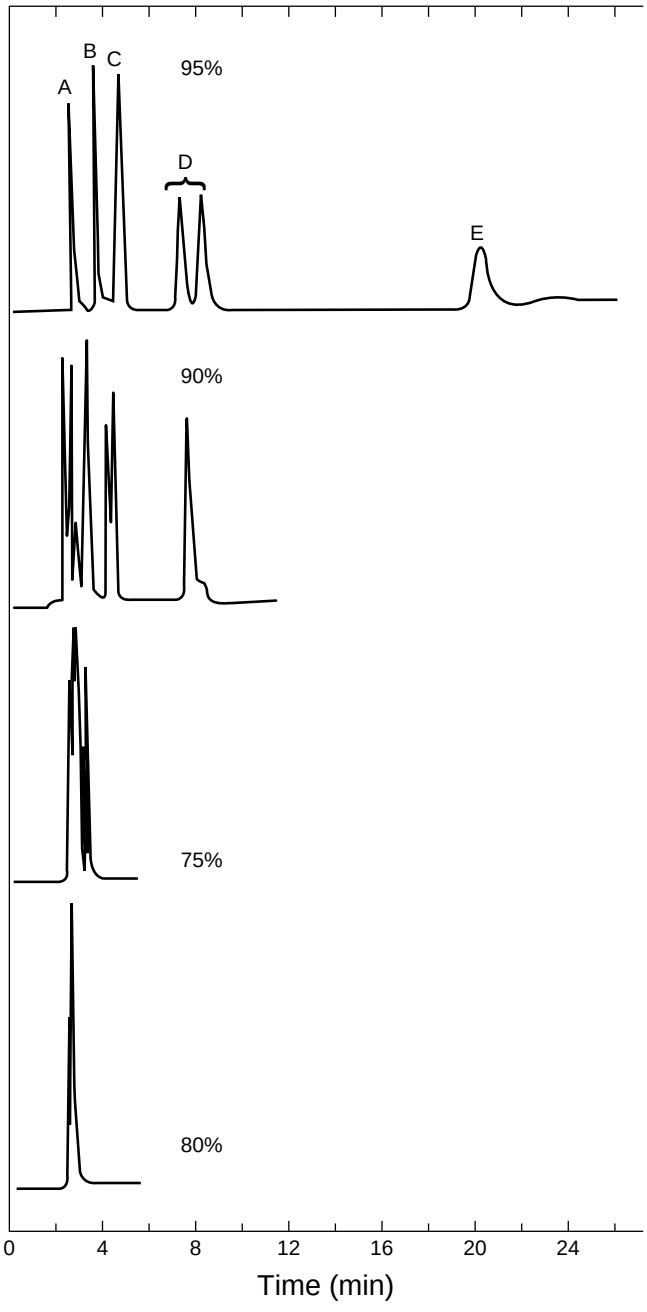

Figure 4 Normal-phase separation of cyclopeptide mixture on a polyhydroxyethyl column by hydrophilic interaction chromatography (HILIC). Mobile phase 15 mM tetraethylammonium phosphate (pH 2.8) with indicated percent acetonitrile included for each separation. (Reproduced by permission of Elsevier Science Publishers from A. Alpert, *J. Chromatogr.*, **499**, 177–196 (1990).)

localization effects and includes a basic localizing solvent (MTBE), a nonbasic localizing solvent (acetonitrile or ethyl acetate), and a nonlocalizing solvent (methylene chloride or ethylene dichloride).[20] This localization-based inner selectivity triangle is also included in Figure 3.

4.4 Aqueous Mobile Phases

While most NPLC mobile phases are made of either polar organic solvents or mixtures of a polar and nonpolar solvent, there are some applications where water is an effective additive. It was noted earlier that silica columns are the strongest NPLC stationary phases and are characterized by a wide variety of silanol adsorption sites. Because water binds very strongly to these silanols, small variations in the water content of the mobile phase can lead to dramatic changes in solute retention. These variations in retention can be largely prevented by purposefully adding small amounts of water to the mobile phase.

Aqueous-based mobile phases are also appropriate for very hydrophilic samples that do not easily dissolve in organic solvents. Frequently, special columns are used with aqueous mobile phases resulting in a unique type of normal-phase separation referred to as HILIC. Mobile phase strength is varied by adding a water-soluble organic solvent. Unlike ''normal'' NPLC, however, increasing the polar organic modifier content leads to an overall *increase* in solute retention. This behavior is demonstrated in Figure 4, which contains chromatograms of a cyclopeptide mixture separated on an HILIC column with various water/acetonitrile mobile phases.[21]

5 OPTIMIZATION STRATEGIES

5.1 Mobile Phase Optimization

Mobile phase optimization begins by establishing the correct solvent strength: that is, a solvent strength that yields solute capacity factors in the range of 0.5–20. As noted previously, solvent strength is normally set by mixing a weak solvent with a strong polar modifier. Hexane is the weak solvent of choice unless a very strong mobile phase is required. In such cases a more polar weak solvent such as methylene chloride or MTBE is more appropriate. Solvent strength is optimized through a series of experiments in which the composition of the polar modifier is increased, or alternatively with a gradient elution experiment. Because of the importance of localization effects in NPLC, solvent strength optimization can frequently provide the necessary retention selectivity.

If an appropriate solvent strength and selectivity cannot be found with a single modifier, a change in modifier type is required. This involves choosing a second polar modifier from a different region of the selectivity triangle (Figure 3). This modifier is added to the weak solvent at a level which produces the same optimized solvent strength as the original mixture. The correct level of new modifier can be estimated via Equation (6) or (7), or by using published nomographs.[8,17] In this way the original overall range of sample capacity factors which was optimized in the solvent strength experiments should remain roughly constant but the relative positions of sample components might be altered through localization and secondary solvent effects. If this second solvent mixture still does not provide adequate resolution of all sample components, a third can be tried, again keeping solvent strength constant. Here, the solvent modifier should come from a corner of the selectivity triangle not represented in the first two solvent mixtures.

With silica columns the inner solvent selectivity triangle that emphasizes localization effects produces the most dramatic changes in retention selectivity.[8,19] The dominance of these localization effects in determining the selectivity of polar bonded-phase columns has not yet been established, however. Significant differences in the selectivities of bonded-phase columns have been accomplished with the traditional inner triangle that emphasizes proton donor, proton acceptor and dipole characteristics.[9,16] Thus, the appropriate strong solvents to use when changing modifier type for selectivity enhancement appear to depend on the column type.

As an example of such a mobile phase optimization, imagine that an amino column is chosen to separate the components in a particular sample. Initial solvent strength experiments with hexane and chloroform then determine that a 90 : 10 (v/v) mixture of hexane–chloroform gives the correct capacity factor range. From the data in Dorsey and Cooper[7] the solvent strength of this composition is estimated to be 0.044. A change in selectivity is then accomplished by changing to a hexane/MTBE mixture, and the data[7] indicate that a 80 : 20 hexane–MTBE composition also has a solvent strength of 0.044. Finally, the third solvent mixture would be hexane–dichloromethane at a 90 : 10 ratio, since the solvent strengths of chloroform and dichloromethane are approximately equal when used with an amino column.

A more elaborate mobile phase optimization scheme has been proposed and shown to be effective in developing NPLC separations of very complex mixtures.[22] This approach again begins by establishing the correct solvent strength using a binary mobile phase, either through incremental, decreasing changes in the amount of strong solvent, or through a ''reversed-gradient'' experiment. Then, a series of seven isocratic experiments are carried out in which the three modifiers from an inner triangle are mixed in a systematic way, all the while maintaining an approximately constant solvent strength. In the example cited the separation of a mixture of substituted naphthalenes was developed. Since a silica column was used, the modifiers were chosen from the ''localization'' inner triangle: MTBE (basic localizing),

acetonitrile (nonbasic localizing) and methylene chloride (nonlocalizing). Initial experiments indicated that a 42% : 58% hexane–methylene chloride mixture yielded the appropriate solvent strength. This solvent mixture was then used for the first isocratic experiment. Subsequent experiments with 96 : 4 hexane–MTBE and 87 : 10 : 3 hexane–methylene chloride–acetonitrile followed. Note that the third mixture required methylene chloride as a co-solvent for hexane and acetonitrile. Experiment 4 mixed the mobile phases used in experiments 1 and 2: 77 : 22 : 1 hexane–methylene chloride–MTBE. Similarly, experiment 5 mixed the mobile phases used in experiments 2 and 3 (92 : 5 : 2 : 2 hexane–methylene chloride–MTBE–acetonitrile), while experiment 6 mixed the mobile phases from experiments 1 and 3 (69 : 30 : 1 hexane–methylene chloride–acetonitrile). Finally, experiment 7 mixed all three binary mobile phases used in the initial three experiments: 89 : 9 : 1 : 1 hexane–methylene chloride–MTBE–acetonitrile.

5.2 Stationary Phase Choices

The introduction of bonded phases for use in NPLC allows stationary phase optimization in parallel with mobile phase optimization. It has been suggested that the cyano, amino and diol columns represent a stationary phase selectivity triangle.[4,16] Not only do these columns provide different retention characteristics, they also can induce very significant and unique secondary solvent effects due to preferential adsorption of mobile phase modifiers. Thus, for particularly challenging separations, mobile phase optimizations superimposed on changes in stationary phase selectivity may be required. Figure 5 demonstrates how a change in column alone (silica to a diol bonded phase) can provide adequate separation of a mixture of phenols. Figure 5 also points out the dramatic changes in solvent strength of modifiers – in this case methanol – when used with different stationary phases.

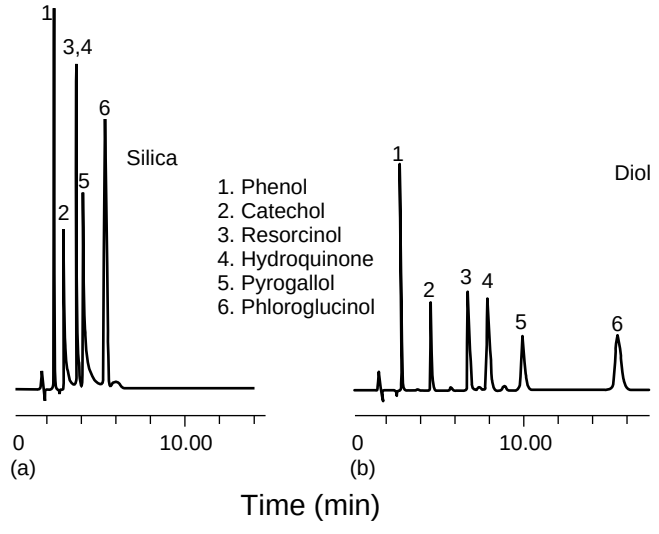

Figure 5 Resolution enhancement through stationary phase optimization. Separation of phenols on bare silica (a) and diol-silica (b). Both columns are 4.6 i.d. × 250 mm, 5 µm particles. Mobile phase: 85 : 15 (v/v) isooctane–ethanol, flow rate 2.0 mL min^{-1}. (Reproduced by permission of YMC, Inc.)

6 TECHNIQUES

6.1 Isocratic Separations

Isocratic separations utilize a single mobile phase composition to elute sample components, in contrast to gradient elution where the mobile phase composition is changed during the course of a separation. The overwhelming majority of LC separations are done isocratically, which is preferred for a number of reasons. First, most mobile phase solvents are not completely transparent to spectroscopic detectors and thus changes in mobile phase composition produce detector ''drift''. In addition, after a gradient elution separation is completed the mobile phase composition must be returned to its original value. This is not as simple as might seem, since the column must also be re-equilibrated with the original solvent mixture, a time- and solvent-consuming procedure.

The introduction of bonded phases for use in the normal-phase mode has greatly increased the opportunities for isocratic NPLC separations. With judicious solvent and/or stationary phase optimization, most sample mixtures can be adequately resolved with a single mobile–stationary phase combination. In addition, many of the benefits of gradient elution can be realized using the column switching and stationary-phase programming techniques discussed below.

6.2 Gradient Elution

In spite of the most exhaustive isocratic optimization efforts, some samples are so complex and contain solutes with such a wide range of sample capacity factors at any mobile phase composition that gradient elution is required. This is known as the ''general elution problem''. At some intermediate value of solvent strength there is insufficient resolution of early-eluting peaks, with extremely long retention of later-eluting peaks. Decreasing solvent strength can increase resolution of the early-eluting peaks, but the result is unacceptably long retention times of the rest of the sample mixture components. Conversely, the strongly-retained solutes can be eluted in a reasonable time by increasing the mobile phase solvent strength, but at the expense of the resolution of the early peaks.

The solution to this general elution problem is to change the capacity factors of solutes during the course of a separation by changing mobile phase composition. This is normally accomplished by using a binary mixture of a weak and strong solvent and increasing the relative proportion of strong solvent as the chromatogram proceeds. A gradient elution separation thus begins with a weak binary solvent mixture that yields capacity factors for early peaks which are sufficiently high for the necessary resolution. Later peaks clearly have capacity factors which are very large and they do not move very far down the column in the first stages of the gradient. However, as the mobile phase increases in strength, the capacity factors of more strongly-retained compounds decrease to the point that they begin to move and eventually elute from the column.

There are a number of very positive benefits that result from gradient elution. The first and most obvious is that a wide range of solutes differing greatly in molecular properties can be separated in a single chromatogram. In addition, since the movement of any solute through the column is significant only when some minimum capacity factor has been reached, most solutes experience roughly the same ''average'' capacity factor during their elution. This means that solute bands elute with approximately equal peak widths. Finally, the gradient elution procedure actually produces a gradient of mobile phase solvent strengths along the column. This means that the back side, or tail, of a solute band will be in a stronger solvent environment than the front. This effect tends to sharpen the bands of polar, strongly-retained solutes which would otherwise be broad due to specific interactions with the stationary phase surface. This last feature of gradient elution is particularly useful in NPLC because of the polar stationary phases commonly used.

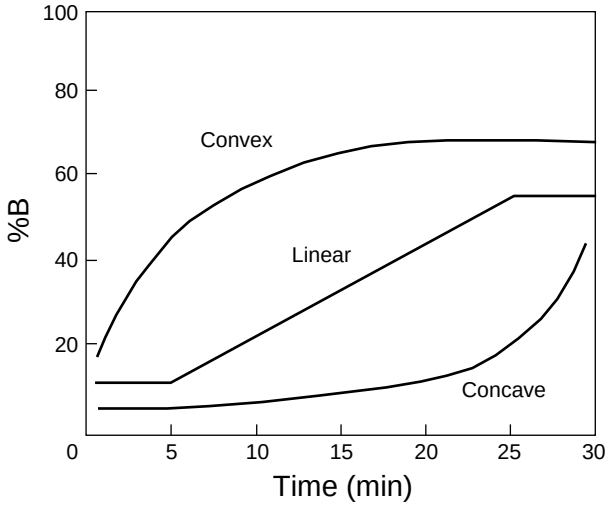

Figure 6 Typical gradients used in HPLC.

There are two important considerations when developing a gradient separation: the shape of the gradient and the rate at which the solvent strength of the mobile phase is changed. The most common gradients are depicted in Figure 6. Note that the gradient is usually defined in terms of the change in the amount of strong (B) solvent in the mobile phase. Linear gradients are the most popular, but modern liquid chromatographs are capable of producing a variety of alternative gradients. A convex gradient, which produces a slower rate of change of mobile phase composition at the end of the chromatogram, may be called for if a linear gradient produces adequate resolution of early-eluting peaks but inadequate resolution of later peaks. This situation is sometimes encountered in environmental and petroleum applications when, for example, polynuclear aromatic hydrocarbons are separated by both number of rings and then number of substituents on each ring. Because the number of possible substituents increases with number of aromatic rings, greater resolution is actually needed at the end of the chromatogram, not the beginning.

Of more interest in NPLC are concave gradients in which the rate of change of mobile phase composition is slower in the early stages of the chromatogram, increasing as the chromatogram develops. This type of gradient is often more appropriate for NPLC because of the nonlinear relationship between solvent strength and % modifier (Equation 7). Such a gradient comes much closer to producing a gradient that is linear in solvent strength.

In addition to choosing a gradient type, the optimum rate at which mobile phase composition is changed must be determined. The rate at which the strong modifier is added to the mobile phase is referred to as the gradient steepness and is normally expressed as $\Delta\,\%\mathrm{B\ min^{-1}}$. In the linear gradient depicted in Figure 6 the modifier content changes from 10% to 50% over a period of 20 min: a gradient steepness of $2\%\ \mathrm{min^{-1}}$. Because different modifiers have different pure solvent strengths, they thus produce different solvent strength gradients at fixed steepnesses. Determining the optimum gradient steepness for a particular gradient elution separation normally requires a trial-and-error approach, although Snyder, Kirkland and Glach[8] have provided some guidelines.

One final word about gradients is necessary and is suggested in Figure 6, where none of the gradients begin at 0% B. Because the polar modifiers in NPLC are preferentially adsorbed by nonpolar stationary phases, the first exposure of the stationary phase to the modifier normally produces a rather significant imbalance in mobile and stationary phase concentrations of B. This will continue until the stationary phase is saturated with B such that the proper equilibrium of B between mobile and stationary phases is reached. Furthermore, this imbalance

is most pronounced at very low (<1–5%B) modifier contents. Thus, to produce ''smooth'' gradients, it is always advisable to begin the gradient at some fixed level of modifier in order to saturate the stationary phase.

6.3 Column Switching, Stationary Phase Programming, and Two-dimensional Liquid Chromatography

For all of its advantages, gradient elution LC does have some drawbacks, as noted previously. These drawbacks have led to an interest in exploiting changes in stationary phase strength and selectivity to accomplish the same chromatographic effect produced by gradient elution. Again, the fundamental principal involved is changing the capacity factors of solutes during the chromatogram. NPLC has the advantage over RPLC in this regard because of the ability to change retention mechanisms by changing columns. This is not routinely possible in RPLC where hydrophobic forces dominate retention in all columns.

Column switching refers to the use of multiple chromatographic columns and switching valves to optimize the separation of very complex mixtures. It is a routine practice in gas chromatography and is finding increasing popularity for LC separations. The fundamental advantage of using multiple columns in this way is that isocratic mobile phases can be employed.

Stationary phase programming is one of the most useful forms of column switching and is analogous to gradient elution. It is normally accomplished by combining columns of different retentive strength or activity with a switching valve. The first column has a significantly lower activity than the second. A sample mixture is injected into the first column and rapidly-moving bands made up of nonpolar components are not separated by this first column but are diverted by the switching valve to the second. Once these fast-moving bands have been loaded onto the second column, the valve is switched so that the slower-moving, more strongly retained solutes, which in principal will be separated by the first column, are directed to the detector. After detection of this first set of polar compounds, the valve is returned to its original position, directing solvent back through column 2. The first, less-polar group of solutes is then eluted. Thus, the retention order of solutes 1–6, where 1 was the least polar and 6 the most polar, might be:

$$4 < 5 < 6 \ll 1 < 2 < 3$$

A column switching technique that is more powerful than stationary phase programming and which is made available by the unique characteristics of different NBP columns is two-dimensional (2-D) chromatography. 2-D chromatography is normally thought of as involving sample displacement (elution) in two dimensions that are physically orthogonal to each other. However, if the two displacements are also chemically orthogonal, then the maximum amount of separation space is available. When the two displacements are physically orthogonal the total peak capacity of the system (the maximum number of peaks that can be resolved in a given time window) is the sum of the individual peak capacities of each displacement (Equation 8).

$$n_T = n_1 + n_2 \tag{8}$$

However, when the two displacements are also chemically orthogonal, the overall peak capacity of the system approaches the product of the individual peak capacities (Equation 9).[23]

$$n_T = n_1 n_2 \tag{9}$$

Figure 7 is a schematic representation of a serially coupled NPLC system that can be used to achieve such 2-D separations.[24] The system uses two bonded phase columns that had previously demonstrated different retention selectivities.[9] It also includes two six-port valves. In the ''load loop 1'' position, loading valve L directs effluent from column 1 to sampling loop 1 while at the same time pumping solvent 2 through loop 2 and into the injection valve I. The injection valve I simply directs effluent from loop 1 to waste and solvent 2 through column 2 and detector 2. Here, solvents 1 and 2 are identical. Once the appropriate peak(s) have been loaded in loop 1, the load and injector valves are rotated and the loop–column configurations reversed. Loop 2 is now being filled with the effluent from column 1 while the contents of loop 1 are injected onto column 2.

Figure 8 contains the resulting 1-D (one-dimensional) and 2-D chromatograms using an identical 70 : 30 (v/v)

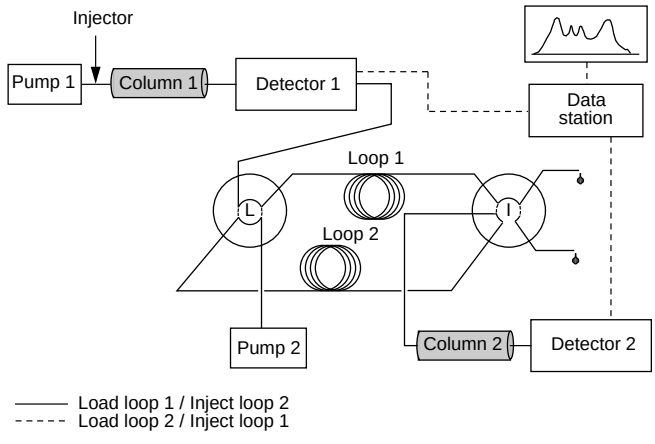

Figure 7 Schematic representation of system used for on-line 2-D LC. Column 1: aminopropylsilica; column 2: 1,2-dihydroxypropylsilica (diol). (See Hsu[24] for further details.)

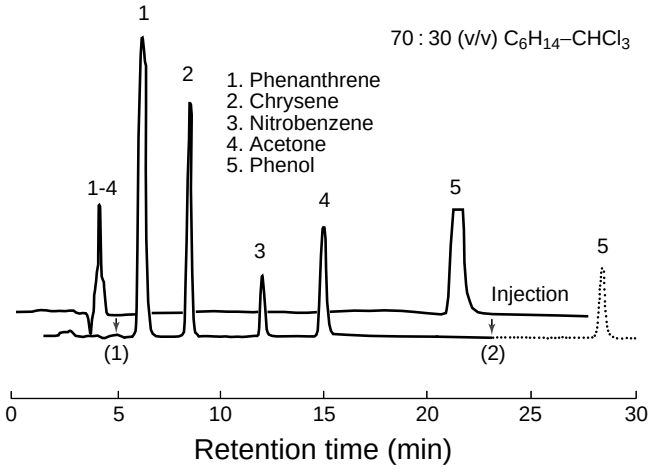

Figure 8 2-D NPLC separation of substituted aromatic compounds. Arrows indicate times at which valve configuration in Figure 7 is changed. At first arrow, effluent from column 1, which is stored in loop 1, is injected onto column 2. At second arrow, effluent from column 1 stored in loop 2 is injected onto column 2. NOTE that response characteristics of detectors 1 and 2 are not identical.

mixture of hexane–chloroform in both columns. Note that the 1-D separation of the sample mixture on the amino column (signal produced by detector 1) demonstrates a classic general elution problem. A 1-D separation on the diol column is also not sufficient because chrysene and phenol co-elute. However, by using these two columns in a chemically orthogonal way, all components in the mixture are well-resolved isocratically in less than 30 min.

7 APPLICATIONS

Following are some examples of separation problems that are uniquely suited to NPLC. NPLC is particularly useful in the environmental, biomedical and pharmaceutical fields, since many analytical procedures in these areas begin with extraction of solutes from a polar matrix into a lipophilic solvent. One of the major advantages of NPLC over RPLC is that a solvent changeover is not necessary after such an extraction. For RPLC analyses, solutes in the lipophilic extract must be moved to a polar solvent suitable for the reversed-phase mode.

7.1 Isomers

As noted in the Introduction, NPLC is the preferred mode for separating achiral isomers. The ability of isomeric molecules containing multiple polar substituents to interact with the fixed absorption sites on NPLC columns and experience localization/delocalization effects depends on the geometry of molecules and the positions of the polar groups. Figure 9 provides a dramatic example of the superior effectiveness of NPLC over RPLC for separating a series of retinol isomers.[25]

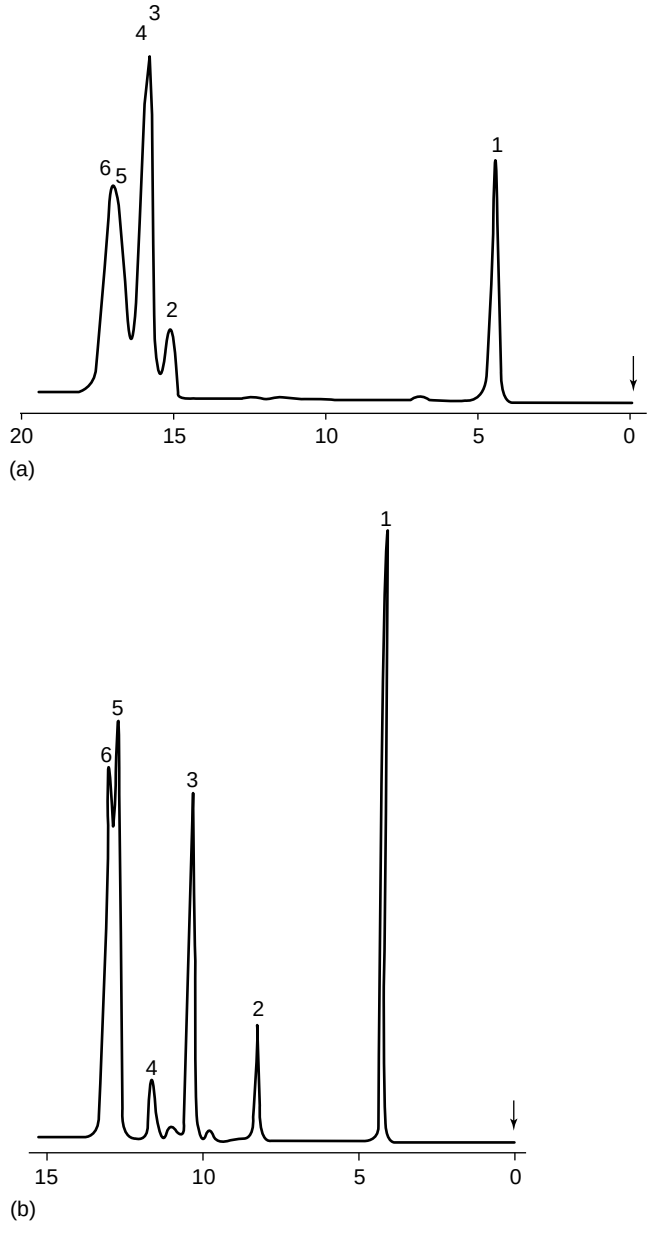

Figure 9 Comparison of RPLC (a) with NPLC (b) separations of retinol isomers (bands 2–6). (Reproduced by permission of Elsevier Science Publishers from B. Stancher, F. Zonta, *J. Chromatogr.*, **234**, 244–248 (1982).)

7.2 Environmental

The analysis of environmental contaminants in water, soils and foodstuffs is now one of the most common procedures carried out by regulatory agencies.

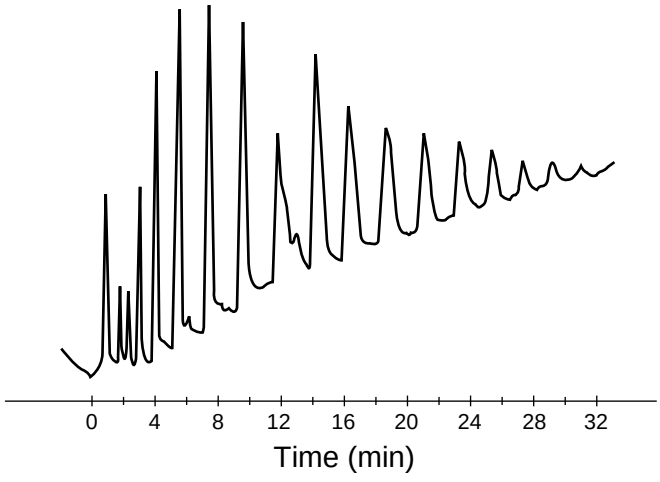

Figure 10 Gradient elution NPLC separation of Triton-X nonionic surfactants. Column: aminopropylsilica, 4.6 i.d. × 250 mm, 5 μm particles. Mobile phase: A = hexane–2-propanol, 90 : 10 (v/v); B = 2-propanol–water, 90 : 10 (v/v). Gradient: 0–3 min, 97%A/3%B; 53 min, 47%A/53%B at $\Delta 1\%$B min^{-1}. (Reproduced by permission of Supelco, Inc.)

Many of these contaminants are only sparingly soluble in water and are thus extracted from environmental matrices with hydrophobic organic solvents. Resolution of individual components in these extracts is facilitated if the extract can be transferred directly to the separation column. This makes NPLC very attractive for such analyses since these lipophilic extraction solvents are normally compatible with NPLC solvents. Indeed, there are now many descriptions in the literature of automated, on-line extraction and LC separation systems.

Figure 10 demonstrates just such a combination of lipophilic extraction and NPLC for environmental purposes. The nonionic surfactant mixture Triton-X-100 was extracted from water onto a solid-phase microextraction fiber, then desorbed with 97 : 3 (v/v) hexane–isopropanol. The Triton-X mixture in that extract was then separated by gradient elution NPLC using an amino column. The wide range of molecular weights in this Triton-X mixture is a common feature of samples that require gradient elution techniques.

7.3 Biomedical and Pharmaceutical

There are many molecules of biological and pharmaceutical interest that are not only lipophilic, but also require separations based on subtle structural features which are most easily distinguished by NPLC. Fat-soluble vitamins are a good example, since it is very difficult if not impossible to get them dispersed in a reversed-phase solvent. Figures 11 and 12 demonstrate the utility of a normal-phase amino column for such separations. In Figure 11

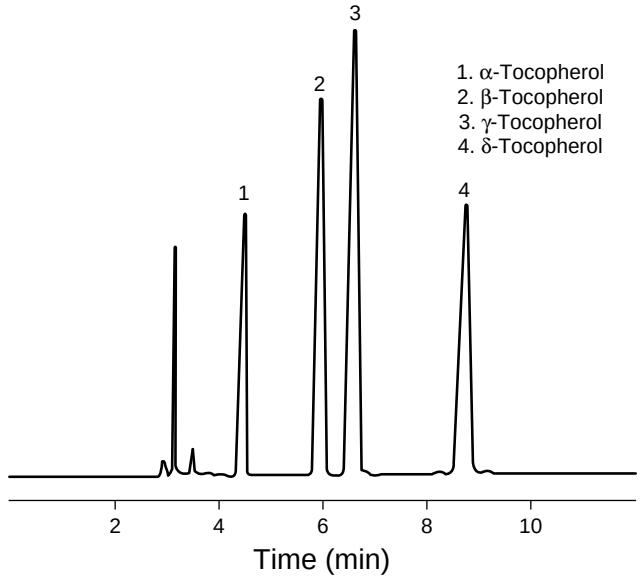

Figure 11 Separation of tocopherols on an aminopropylsilica bonded phase column. Mobile phase: 70 : 30 (v/v) hexane–ethyl acetate, flow rate 1.0 mL min^{-1}. (Reproduced by permission of Supelco, Inc.)

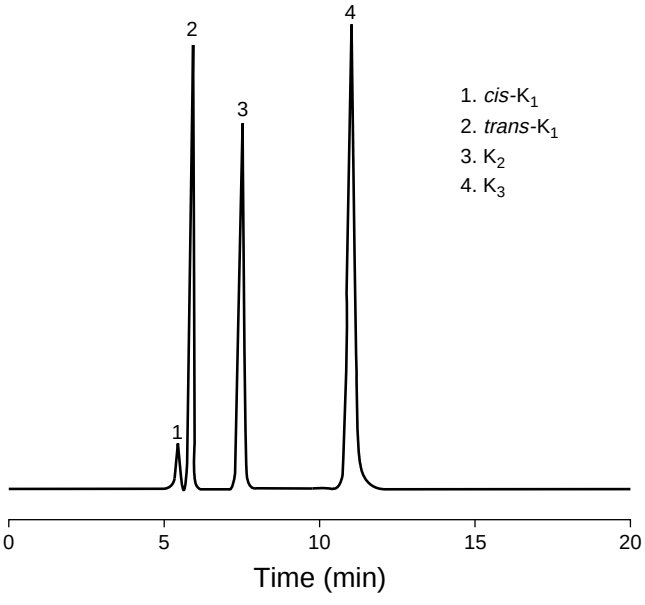

Figure 12 Separation of Vitamin K isomers on an aminopropylsilica bonded phase column. Mobile phase: 99 : 1 (v/v) hexane–ethyl acetate, flow rate 1.5 mL min^{-1}. (Reproduced by permission of Supelco, Inc.)

various tocopherols (Vitamin D) are easily resolved in less than 10 minutes isocratically. The strength of NPLC for isomer separations is further demonstrated in the chromatogram in Figure 12 in which Vitamin K isomers are cleanly resolved.

While RPLC is normally the choice for separating and purifying peptide mixtures, it is not always the best choice

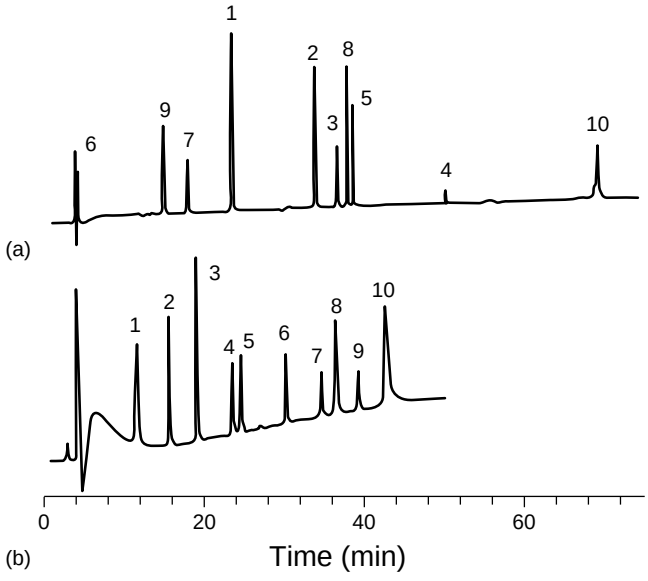

Figure 13 Comparison of chromatograms of peptides separated on (a) TSK®-ODS reversed-phase and (b) TSK®-Amide normal-phase columns. Chromatogram (a) obtained with 83.3 min linear gradient of acetonitrile–water (0.1% TFA) from 5 : 95 to 55 : 45. Chromatogram (b) obtained with a 70 min linear gradient of water(0.6% TFA)–acetonitrile from 3 : 97 to 45 : 55. (Reproduced by permission of the American Chemical Society from T. Yoshida, *Anal. Chem.*, **69**, 3038–3043 (1997).)

for hydrophilic peptides. Such peptides can frequently be better separated using special normal-phase columns and aqueous-based mobile phases. Figure 13 contains chromatograms of a mixture of peptides separated on reversed-phase and normal-phase TSK® gel columns.[26] The TSK®-Amide normal-phase column clearly provides better resolution and efficiency in less time than the TSK®-ODS reversed-phase column.

ACKNOWLEDGMENTS

The author would like to acknowledge the contributions of the following former graduate students in further developing the science of NPLC: Patty Smith, Scott Boudreau, Leo Hsu and Tim Keefe. He is also grateful for financial support from the US Environmental Protection Agency, Office of Research and Development.

ABBREVIATIONS AND ACRONYMS

HILIC	Hydrophilic Interaction Chromatography
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
LSC	Liquid–Solid Chromatography
MTBE	Methyl-*tert*-butyl Ether
NBP	Normal Bonded Phase
NPLC	Normal-phase Liquid Chromatography
ODS	Octadecylsilane
RPLC	Reversed-phase Liquid Chromatography
THF	Tetrahydrofuran
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Environment: Water and Waste **(Volume 3)**
Detection and Quantification of Environmental Pollutants

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Phenols Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples

Food **(Volume 5)**
Liquid Chromatography in Food Analysis

Pesticides **(Volume 7)**
Chiral Pesticides and Polychlorinated Biphenyl Congeners in Environmental Samples, Analysis of • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Polymers and Rubbers **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization

Liquid Chromatography **(Volume 13)**
Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Reversed Phase Liquid Chromatography • Silica Gel and its Derivatization for Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

REFERENCES

1. L.R. Snyder, *Principles of Adsorption Chromatography*, Marcel Dekker, New York, 1968.

2. L.R. Snyder, T.C. Schunk, 'Retention Mechanism and the Role of the Mobile Phase in Normal-phase Separations on an Amino-bonded-phase Column', *Anal. Chem.*, **54**, 1764–1772 (1982).
3. L.R. Snyder, J.L. Glajch, J.J. Kirkland, 'Theoretical Basis for Systematic Optimization of Mobile Phase Selectivity in Liquid–Solid Chromatography. Solvent–Solute Localization Effects', *J. Chromatogr.*, **218**, 299–326 (1981).
4. L.R. Snyder, 'Liquid–Solid Chromatography: New Insights into Retention on Bonded-phase Packings', *LC Magazine*, **1**, 478–486 (1983).
5. E. Soczewinski, 'Solvent Composition Effects in Thin-layer Chromatography Systems of the Type Silica Gel–Electron Donor Solvent', *Anal. Chem.*, **41**, 179–183 (1969).
6. L.R. Snyder, 'Mobile Phase Effects in Liquid–Solid Chromatography', in *High-performance Liquid Chromatography: Advances and Perspectives*, ed. C. Horvath, Academic Press, New York, 157, Vol. 3, 1983.
7. J.G. Dorsey, W.T. Cooper, 'Retention Mechanisms of Bonded-phase Liquid Chromatography', *Anal. Chem.*, **66**, 857A–867A (1994).
8. L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2nd edition, Wiley-Interscience, New York, 1997.
9. P.L. Smith, W.T. Cooper, 'Retention and Selectivity in Cyano-, Amino- and Diol-silica Normal Bonded Phase HPLC Columns', *J. Chromatogr.*, **410**, 249–265 (1987).
10. J.E. Eberle, R.L. Grob, P.E. Antle, L.R. Snyder, 'Simplified Description of High-performance Liquid Chromatography Separations under Overload Conditions Based on the Craig Distribution Model I. Computer Simulations for a Single Elution Band Assuming a Langmuir Isotherm', *J. Chromatogr.*, **384**, 25–44 (1987).
11. J.E. Eberle, R.L. Grob, P.E. Antle, L.R. Snyder, 'Simplified Description of High-performance Liquid Chromatography Separations under Overload Conditions Based on the Craig Distribution Model I. Computer Simulations for Two Co-eluting Bands Assuming a Langmuir Isotherm', *J. Chromatogr.*, **405**, 1–29 (1987).
12. C.W. Hsu, W.T. Cooper, 'Solvent–Stationary Phase Interactions in Normal Bonded Phase High Performance Liquid Chromatographic Columns: I. Investigation of System Peaks in Amino Bonded Phase Columns', *J. Chromatogr.*, **603**, 63–71 (1992).
13. S.P. Boudreau, P.L. Smith, W.T. Cooper, 'HGSC – A New Approach to Inverse Chromatography', *Chromatography*, **2**(5), 31–35 (1987).
14. J.J. Kirkland, C.H. Dilks, J.J. DeStefano, 'Normal-phase High-performance Liquid Chromatography with Highly Purified Porous Silica Microspheres', *J. Chromatogr., A*, **635**, 19–30 (1993).
15. A. Kurganov, U. Truedinger, T. Isaeva, K. Unger, 'Native and Modified Alumina, Titania and Zirconia in Normal and Reversed-phase High-performance Liquid Chromatography', *Chromatographia*, **42**, 217–222 (1996).
16. P.L. Smith, W.T. Cooper, 'Extended Solubility Parameters and Stationary Phase Selectivities of Amino-, Cyano-, and Diol-silica Normal Bonded Phase Liquid Chromatography Columns', *Chromatographia*, **25**, 55–60 (1988).
17. V.R. Meyer, M.D. Palamareva, 'New Graph of Binary Mixture Solvent Strength in Adsorption Liquid-chromatography', *J. Chromatogr.*, **641**, 391–395 (1993).
18. L.R. Snyder, 'Classification of the Solvent Properties of Common Liquids', *J. Chromatogr. Sci.*, **16**, 223–234 (1978).
19. L.R. Snyder, P.W. Carr, S.C. Rutan, 'Solvatochromically Based Solvent-selectivity Triangle', *J. Chromatogr. A*, **656**, 537–547 (1993).
20. J.L. Glajch, J.J. Kirkland, L.R. Snyder, 'Practical Optimization of Solvent Selectivity in Liquid Solid Chromatography Using a Mixture-design Statistical Technique', *J. Chromatogr.*, **238**, 269–280 (1982).
21. A. Alpert, 'Hydrophilic-interaction Chromatography for the Separation of Peptides, Nucleic Acids and Other Polar Compounds', *J. Chromatogr.*, **499**, 177–196 (1990).
22. L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2nd edition, Wiley-Interscience, New York, 282, 1997.
23. J.C. Giddings, 'Two-dimensional Separations: Concepts and Promise', *Anal. Chem.*, **56**, 1258A–1270A (1984).
24. C.W. Hsu, 'The Investigation of Solvent–Stationary Phase Interaction in Normal Bonded Phase High Performance Liquid Chromatography Columns', PhD Dissertation, Florida State University, 143, 1992.
25. B. Stancher, F. Zonta, 'Comparison between Straight and Reversed Phases in the High-performance Liquid-chromatographic Fractionation of Retinol Isomers', *J. Chromatogr.*, **234**, 244–248 (1982).
26. T. Yoshida, 'Peptide Separation in Normal Phase Liquid Chromatography', *Anal. Chem.*, **69**, 3038–3043 (1997).

Reversed Phase Liquid Chromatography

W. John Lough
Institute of Pharmacy & Chemistry, University of Sunderland, UK

1 Introduction 1
2 Theory and Retention Mechanism 2
3 Experimental Variables 3
3.1 Chromatographic Resolution 3
3.2 Selectivity Manipulation Using Mobile Phase pH 4
3.3 Control of Retention 5
3.4 Stationary Phase Selectivity 6
3.5 Mobile Phase Optimization 7
4 Perspective and Future Developments 8
Abbreviations and Acronyms 8
Related Articles 8
References 8

Reversed-phase liquid chromatography (RPLC) is an elution procedure used in liquid chromatography (LC) in which the mobile phase is significantly more polar than the stationary phase, e.g. a microporous silica-based material with chemically bonded alkyl chains. Typically, the mobile phase is a mixture of an aqueous component, often containing a buffer to control pH, and a water-miscible polar organic solvent. This most frequently used mode of LC is suitable for the analysis of compounds with a very wide range of polarities, using experimental variables such as mobile-phase pH and ratio of organic to aqueous mobile phase components to manipulate selectivity and retention to separate sample components from one another.

1 INTRODUCTION

RPLC is a mode of chromatography in which there is a polar mobile phase and a nonpolar stationary phase. In principle this applies to any chromatographic system in which the mobile phase is more polar than the stationary phase. However in practice there must be a marked difference in polarity between the phases for significant chromatographic retention to be obtained. Accordingly, the most common form of RPLC is that in which the mobile phase is a mixture of an aqueous component, often containing a buffer to control pH, and a water-miscible polar organic solvent and the stationary phase is an alkyl derivatized silica (see **Silica Gel and its Derivatization for Liquid Chromatography**).

RPLC was first developed in the mid- to late-1970s as a direct consequence of needs arising from the shortcomings of normal phase LC. Normal phase LC itself utilizing adsorption chromatography had been an advance on partition chromatography.[1] It might not have preceded RPLC had it not been for the fact that polar materials such as silica and alumina could be produced as porous microparticles while at the time there were no materials available readily to prepare nonpolar porous microparticles. However this situation had to be addressed as the aforementioned shortcomings of normal phase LC became more apparent (see **Silica Gel and its Derivatization for Liquid Chromatography**):

- The silica surface was heterogeneous, containing acidic isolated silanol groups, H-bonded silanol groups, geminal silanol groups, relatively nonpolar Si−O−Si bonds, metal impurity activated silanol groups, ion-exchange $Si-O^-$ sites and adsorbed water; this could give rise to poor efficiency arising from mixed retention mechanisms or at worst very bad peak tailing for basic compounds caused by interaction with the more strongly acidic sites.
- Polar analytes could have very long retention times or, even worse, not be soluble in the mobile phase.

The latter issue was the most telling since much of the momentum for the development of LC was provided by the necessity to analyze drugs in a rapidly expanding pharmaceutical industry. Since drugs must get into the body, which contains a substantial proportion of water, they generally tend to be polar ionizable compounds. Further, they are often presented as salts or in aqueous solutions. Accordingly it became essential that suitable nonpolar stationary phases should be developed in order that polar mobile phases might be used.

In the early days of RPLC the market was dominated by columns packed with Waters' C_{18}-μBondapak, a 10-μm irregular material in which residual silanol groups had not been end-capped (see **Silica Gel and its Derivatization for Liquid Chromatography**). This material enjoyed enduring commercial success but stationary-phase technology soon progressed to the point in the early 1980s that 5-μm spherical end-capped C_{18} materials with a narrow particle size range were the norm. This was still very much the case in the mid-1990s despite a number of subsequent useful developments in the intervening period, including the advent of 3-μm particles, base-deactivated silicas and more stable protected C_{18} materials (sections 3.4 and 4). Similarly there have been many exciting developments in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

miniaturized LC but in a typical routine quality control laboratory the only concession to saving on solvent use and shortening run times might be the use of a 10-cm or 15-cm column rather than a 25-cm column and only occasionally the use of a 2-mm or 3-mm rather than 4.6-mm inner diameter (ID) column diameter.

A system that would be fairly characteristic of the usual practice of RPLC would therefore be a 150 × 4.6-mm ID stainless steel column packed with a bed of 5-μm spherical end-capped C_{18} porous microparticles with a narrow particle size range using a mobile phase of methanol and an aqueous phosphate buffer. While there are several organic solvents that might be used as alternatives to methanol, most can be ruled out on the grounds of high cost, high toxicity, high viscosity or not being transparent to the detector, usually ultraviolet (UV), being used. Therefore acetonitrile is the only other frequently used organic component of the mobile phase. Tetrahydrofuran (THF) is also used but usually in proportions in the order of 5–20% along with acetonitrile or methanol in order to ''fine tune'' selectivity. Phosphate is a popular buffer because it controls pH well over a wide part (2.5–4.5 and 6.5–7.5) of the pH range between 2.5–7.5 over which silica-based columns are usually considered to be stable. The main disadvantage of its use is that the inorganic salt can precipitate out of the mobile phase if the proportion of the organic component is too high. This is particularly a problem for acetonitrile. Its other disadvantage is that it is not volatile. This is becoming much more of an issue with the increasing use of mass spectrometric detection in LC. Even if an assay is being developed for use with UV detection, mass spectrometry (MS) compatability might still have to be borne in mind since it might become necessary to use MS to identify peaks of unknown identity in the chromatograms obtained. As the instrumentation required to carry out isocratic LC is simpler than that required to carry out gradient LC (see **Gradient Elution Chromatography**) every effort will usually be made to develop isocratic RPLC methods. However for some analyses of complex mixtures in which the compounds present have a wide range of polarities, the use of gradient RPLC may be unavoidable. The determination of structurally related impurities in a drug substance is one such example.

It has already been stated that the desirability of being able to dissolve polar analytes in a polar mobile phase was one of the driving forces for the development of RPLC. This remains one of the main advantages of RPLC. Nonetheless it should not be overlooked that it also has the advantage of being applicable to the analysis of compounds having a very wide range of polarities. The other main advantage is the number of variables that can be used to manipulate selectivity and retention to achieve a desired chromatographic resolution. These advantages together account for the fact that RPLC is so versatile and effective that it is used for a very large proportion of all LC analyses.

The disadvantages of RPLC are few indeed. There are not too many analyte classes for which another mode of chromatography is needed in order to obtain a separation that could not be obtained by RPLC. There are also only a few instances where RPLC is not used for other reasons of a more practical nature. For example wide pore packing materials may be used to carry out RPLC of proteins but another mode of chromatography will frequently be preferred to RPLC because of the danger of the proteins being denatured by the stationary phase or the presence of too high a concentration of the organic component of the mobile phase.

2 THEORY AND RETENTION MECHANISM

Like most other modes of chromatography, retention in RPLC may be described by either a solvent interaction model or a solvent competition model. However it is most commonly interpreted by the application of solvophobic theory[2] which borrows elements from both the solvent interaction and the solvent competition model. The stationary phase is regarded as being saturated with molecules of the organic component, i.e. the most strongly eluting component of the mobile phase. Retention by the stationary phase arises from repulsion of hydrophobic regions of the analyte molecule by the water molecules which ''drives'' the analyte onto the stationary phase or at least into the layer of the organic component adsorbed on its surface (Figure 1). The extent to which the situation resembles the solvent interaction or solvent competition

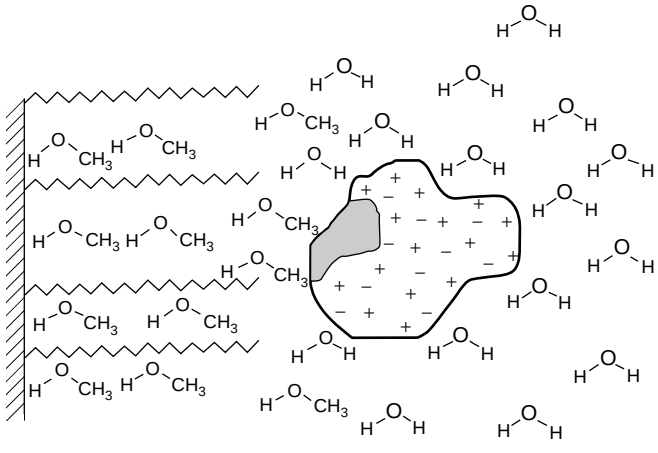

Figure 1 Retention in RPLC on an alkyl-bonded silica arising from repulsion of nonpolar regions of analyte molecules by the water molecules of the mobile phase.

model might depend on the extent to which this hydrophobic repulsion is moderated by specific analyte–organic solvent or analyte–stationary phase interactions. The frequency with which these are a significant influence is discussed later (sections 3.4 and 3.5).

However, since retention is mainly related to analyte hydrophobicity, the 1-octanol–water partition coefficient ($\log P$), or more particularly $\log P_{app}$ (where P_{app} is the distribution coefficient at a particular pH or apparent partition coefficient; it will be less than P if some of the analyte is ionized at the pH in question) to take into account the degree of ionization at the mobile phase pH, of the analyte can be used either qualitatively or quantitatively to assess how well an analyte will be retained. Similarly in attempting to rationalize the effect experimental variables have on retention it is helpful to think in terms of how a change in the mobile phase will affect the proportion of water molecules in the mobile phase and thereby how it will affect the degree to which the analyte is repelled towards the stationary phase. In the same vein it is helpful to think in terms of how any change of stationary phase might affect the volume of the organic solvent adsorbed on the stationary phase and thereby how its capacity to ''accept'' analyte molecules is affected.

3 EXPERIMENTAL VARIABLES

3.1 Chromatographic Resolution

The whole point of separative methods is to obtain specificity in an analytical method by resolving the analyte from all other detectable components of the sample being analyzed. In RPLC, as with other modes of LC, chromatographic resolution depends on efficiency, retention and selectivity, Equations (1–5):

$$\text{Efficiency, } N = 16\left(\frac{t_r}{w_b}\right)^2 \quad (1)$$

$$\text{Capacity factor, } k' = \frac{t_r - t_0}{t_0} \quad (2)$$

$$\text{Selectivity, } \alpha = \frac{k'_2}{k'_1} \quad (3)$$

$$\text{Resolution, } R_s = \frac{t_{r2} - t_{r1}}{0.5(w_{b2} + w_{b1})} \quad (4)$$

$$R_s = \underbrace{\frac{1}{4}\sqrt{N}}_{\text{efficiency factor}} \cdot \underbrace{\left(\frac{\alpha - 1}{\alpha}\right.}_{\text{selectivity factor}} \cdot \underbrace{\left(\frac{k'_1}{k'_1 + 1}\right.}_{\text{retention factor}} \quad (5)$$

where t_{rx} is the retention time for peak x (r is peak number, e.g. 0 for unretained, 1 for first, 2 for second etc.), t_0 is the retention time for an unretained analyte and w_{bx} is the baseline width for peak x; all of these must be measured in the same units.

Given the general use of 5-μm particles, for good reasons, for example the practical problems associated with the use of particles with lower diameter than this and the fact that there is a square root dependency of resolution on efficiency (Equation 5), means that there is less incentive to overcome these practical problems, and column efficiency is not often used as a variable to improve resolution. Occasionally a chromatographic method may involve operation at elevated temperature, with faster mass transfer resulting in improved efficiency. However the exploitation of elevated temperature in RPLC is more likely to involve simply thermostatting at 5 °C or more above ambient temperature in order to allow strict control of temperature for reproducible chromatography.

Most of the variables in RPLC affect retention but changes in retention have limited value in improving resolution because (i) an increase in run time is not desirable under normal circumstances, (ii) above $k' \sim 4$ increases in retention have limited effectiveness in improving resolution (Figure 2), and (iii) most importantly, most of the variables which change the retention of one sample component will change the retention of all the other components in the same direction (section 3.3). Having said this, it is apparent from Figure 2 that for analytes with k' in the range 0–2, an increase in k' will be highly effective in bringing about an improvement in resolution.

By a process of elimination it should be apparent that improving selectivity is the most useful means of improving resolution and this is indeed the case (Figure 3). The fact that increasing selectivity is not so effective for increasing resolution when selectivity is already greater than ~2 is of little consequence

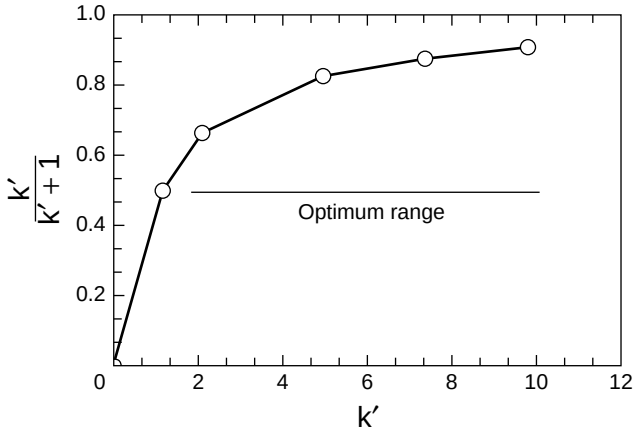

Figure 2 Variation of retention factor, and therefore resolution, since retention factor $k'/(k' + 1)$ is proportional to R_s.

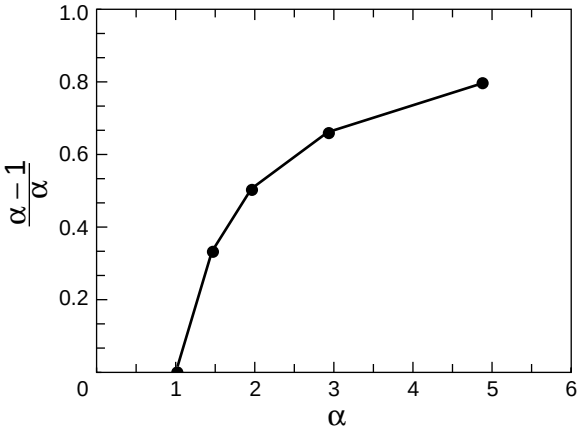

Figure 3 Variation of selectivity factor, and therefore resolution, with selectivity. $(\alpha - 1)/\alpha$ is proportional to R_s.

because in this range separations would be very easy to achieve in any case. It has already been stated that most of the variables which change the retention of one sample component will change the retention of all the other components in the same direction. Clearly, to improve selectivity it would be best to use an experimental variable that gave rise to increases in retention for some analytes and decreases in retention for others. This is why the use of changes in mobile phase pH is so important in RPLC method development.

3.2 Selectivity Manipulation Using Mobile Phase pH

pH control is very important in RPLC of ionizable compounds, Equation (6) for acids and Equation (7) for bases:

Acids

$$\underset{\text{low pH}}{\text{RH}} \Longleftrightarrow \underset{\text{high pH}}{\text{R}^- + \text{H}^+} \qquad (6)$$

(sulfonic acids, carboxylic acids, phenols, coumarins, sulfonamides, barbiturates, β-diketones, etc.)

Bases

$$\underset{\text{low pH}}{\text{RH}^+} \Longleftrightarrow \underset{\text{high pH}}{\text{R} + \text{H}^+} \qquad (7)$$

(amines)

Irrespective of retention and selectivity issues, it is important that the mobile phase pH is such that an ionizable analyte is either fully ionized or fully un-ionized. For anything in between this situation, if the ionization/deionization rates are not rapid on the chromatographic timescale then chromatographic efficiency will be reduced, possibly accompanied by the development of peak asymmetry. Even if the equilibrium is rapidly reversible, problems might arise. A slight change in the preparation of the mobile phase might introduce a slight change in pH which would bring about a slight change in the position of the equilibrium and thereby a shift in retention. In other words an analytical method based on such chromatography would not be robust.

The decision on whether the ionizable analyte should be chromatographed in its fully ionized form or its fully un-ionized form depends on whether it is required for the analyte to be well retained or otherwise. Clearly the ionized form, being charged, will be the more polar form (Equations 6 and 7). Being more polar it will be less repelled by water molecules and therefore less well retained in RPLC. Accordingly if it is required to reduce the retention of an acid, the mobile phase pH must be increased. Similarly if it is required to reduce the retention of a base, the mobile phase pH must be decreased. The fact that a pH change will affect the retention of an acid and a base in the opposite direction is the reason why it is the most important variable in RPLC method development in the sense that it is able to bring about a change in selectivity.

It is necessary then in RPLC method development to be aware of the pH values at which an analyte will be fully ionized or un-ionized. The degree of ionization of an ionizable analyte at a particular pH may be determined by the well-known Henderson–Hasselbach equation, Equations (8) and (9):

$$\%\ \text{Ionized acid} = \frac{100}{1 + \textit{anti}\log(\text{p}K_a - \text{pH})} \qquad (8)$$

$$\%\ \text{Ionized base} = \frac{100}{1 + \textit{anti}\log(\text{pH} - \text{p}K_a)} \qquad (9)$$

This can be simplified for the situation when the analyte is 50% ionized and so on for 90%, 99%, to develop a so-called ''rule of thumb'', see Table 1.

Therefore for an ionizable analyte to be fully ionized or fully un-ionized the mobile phase pH must be at least three pH units higher or lower than the pK_a of the analyte. While the pK_a of the analyte may be measured, there is a need to know the pK_a of all the other sample components because they will have to be resolved from the analyte. As it is impractical to carry out pK_a measurements for all sample components, it is vital in RPLC for the analyst to be able to recognise functional groups in an analyte and to be aware of their likely pK_a values. This is very much the case, for example, in the analysis of drugs. As has already been stated (section 1), compounds of pharmaceutical interest are usually polar and ionizable.

Table 1 Rule of thumb for ionization of acids and bases[a]

pH	% Ionized[d]	
Acids[b] (e.g. CH_3CO_2H, $pK_a = 4.8$)		
1.8	0.1	↑OH ↑pH ↑% ionized
2.8	1	
3.8	10	$CH_3COOH + OH \longrightarrow CH_3COO + H_2$
4.8	50	
5.8	90	
6.8	99	
7.8	99.9	
Bases[c] (e.g. $H_2N-CH_2-CH_2OH$, pK_a $(NH_3^+) = 9.4$)		
6.4	99.9	
7.4	99	↑H^+ ↓pH ↑% ionized
8.4	90	
9.4	50	$H_2N-CH_2-CH_2OH + H_3O^+ \longrightarrow H_3N^+-CH_2-CH_2OH + H_2O$
10.4	10	
11.4	1	
12.4	0.1	

[a] Adapted from lecture handout. C.J. Edwards, University of Sunderland.
[b] For almost complete ionization, the pH of the environment should be at least three units above the pK_a of the acid.
[c] For almost complete ionization, the pH of the environment should be at least three units below the pK_a of the conjugate acid of the base.
[d] The values for percent ionization shown above are only approximately accurate. Values may be calculated using the Henderson–Hasselbach equation.

In fact they can often contain more than one acidic or basic functional group. Illustrations of the types of functional group found in some common drugs are shown in structures **(1)** to **(8)**.

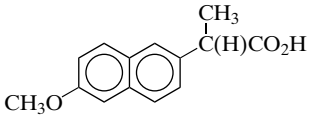

(1)

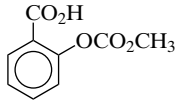

(2)

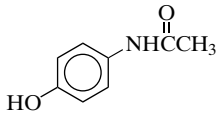

(3)

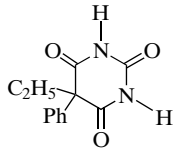

(4)

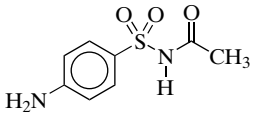

(5)

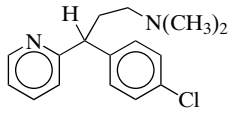

(6)

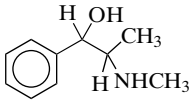

(7)

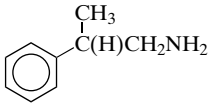

(8)

pK_a values associated with ionizable functional groups in drug molecules **(1)** naproxen, carboxylic acid pK_a 4.2, **(2)** aspirin, arylcarboxylic acid pK_a 3.5, **(3)** paracetamol, phenol pK_a 9.5, **(4)** phenobarbitone, imide pK_a 7.4, **(5)** sulfacetamide, arylsulfonimide pK_a 5.4, primary arylamine pK_a 1.8, **(6)** chlorpheniramine, pyridine ring pK_a 4.0, tertiary amine pK_a 9.2, **(7)** ephedrine, secondary amine pK_a 9.6, **(8)** amphetamine, primary amine pK_a 9.9.

3.3 Control of Retention

While the manipulation of pH is invaluable with respect to improving selectivity between ionized and un-ionized analytes, the role of other experimental variables in changing retention should not to be underestimated. As has been observed (Figure 2), increasing retention is very effective for increasing resolution when $k' < 2$. At higher retention, increasing retention is a less effective means of improving resolution and in any case it would not be advisable to go on increasing analysis time when it is already excessive.

The effect on k' of changing various experimental variables is shown in Table 2. These effects can be rationalized by the solvophobic theory. For example by increasing the proportion of methanol in the mobile phase or switching from methanol to acetonitrile the mobile phase is made less polar. Being more hydrophobic it is less likely to give rise to repulsion of a hydrophobic region

Table 2 Effect on k′ of experimental variables in RPLC

Variable	Acid	Base	Neutral
pH 2.5 → 7.5	↓	↑	little effect
% methanol ↑	↓	↓	↓
$C_{18} \rightarrow C_8 \rightarrow C_2$	↓	↓	↓
Ionic strength ↑	↑	↑	↑
Methanol → acetonitrile → THF	↓	↓	↓
Add DMOA	↑	slight ↓	little effect
Column strength ↑	no effect	no effect	no effect
Flow rate ↑	no effect	no effect	no effect

DMOA (*N*,*N*-dimethyloctylamine) is a ''masking agent'' introduced into the mobile phase at levels of ~0.0005 M to prevent the interactions between basic analytes and residual silanol groups on the stationary phase which lead to peak tailing.

of the analyte and ''drive'' it onto the stationary phase. Hence the mobile phase is more strongly eluting and k′ is reduced. This holds so long as there is a ''pure'' reversed-phase retention mechanism based on hydrophobicity and there are no specific analyte–solvent interactions. Methanol can take part in hydrogen-bonding interactions, acetonitrile can take part in dipole–dipole interactions and THF can act as an electron pair donor. If such interactions between solvent and analyte take place they will hold up the analyte in the mobile phase more than would be the case and hence give rise to a reduced retention time.

It is perhaps less clear on the basis of solvophobic theory why alklyl-bonded silica phases with longer alkyl chains are more retentive. The greater the ''carbon loading'' of the stationary phase the greater will be its capacity at saturation for the organic component of the mobile phase. Effectively then the stationary phase has a greater ''volume'' and is therefore more retentive.

The effect of masking agents such as DMOA (see **Silica Gel and its Derivatization for Liquid Chromatography**) on retention has been less of a concern because they are being used less with the increasing use of base-deactivated silaceous reversed-phase packing materials. However it remains an issue since many methods that were developed years ago are still in use. The interactions of basic analytes with residual silanols that the masking agent is designed to block contribute to the overall degree of retention. Therefore when the blocking takes place, k′ is reduced. Generally silanol interactions do not contribute significantly to the retention of neutral and acidic analytes. However, acidic analytes in their ionized form may enter into a hydrophobic ion-pair with net zero charge with the masking agent and hence be more retained.

While changes in flow rate and column length affect retention time they clearly do not affect k, as the elution time of an unretained analyte would be changed in the same way. Any changes to resolution that they bring about will arise from their effect on column efficiency.

3.4 Stationary Phase Selectivity

As mentioned already in the context of the use of masking agents and the occurrence of solvent–analyte interactions, the retention mechanism in RPLC will not always be exclusively based on hydrophobicity. The most common additional contribution to retention from stationary phases is that from the interaction of basic analytes with residual silanol groups on the stationary phase, this being more of a problem for more acidic isolated silanol groups or those activated by the presence of metallic impurities than for silanols which are H-bonded to neighboring silanols. While these interactions affect peak shape and retention of basic analytes they would not have a profound affect on the selectivity exhibited towards a series of basic analytes. The same may be said for stationary phase specific interactions exhibited by functional groups (e.g. cyano-, phenyl-, diol, nitro-) present in medium polarity RPLC stationary phases. Studies with a wide range of model analytes have suggested that retention orders correlate fairly well with those found on alkyl-bonded silicas.[3]

Retention in the few stationary phases that are not silica-based is also dependent on hydrophobicity. This is especially the case for the styrene–divinylbenzene copolymer (PRP-1) (Scheme 1). There appear to be no specific interreactions with the aromatic rings on the stationary phase surface and the degree of correlation between retention and hydrophobicity is such that log k′ values may be used as a measure of analyte partition coefficients.

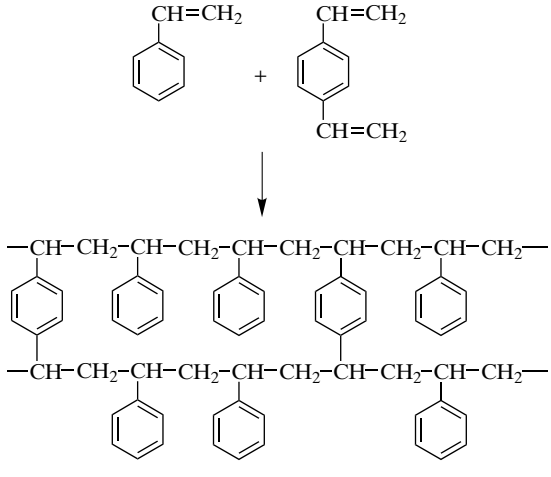

Scheme 1 Preparation of PRP-1.

Another nonsilaceous packing material for RPLC is the porous graphitic carbon phase, Hypercarb®. Like PRP-1, Hypercarb® can be used at extremes of pH and is more retentive than octadecylsilyl silicas. In fact it is much more retentive even than PRP-1 and this is a property that may be usefully exploited. A good illustration of this is the use of Hypercarb® in microbore LC columns. One of the advantages of microbore LC is that it may be used to obtain increased mass sensitivity. However gains in this direction are compromised by extra-column band-broadening effects arising in particular from the injection volume. As Hypercarb® is so retentive it is an easy matter to find a weakly eluting solvent for the sample solution so that when large volumes of sample solution are injected onto the column an on-column sample focusing effect takes place.

While the fact that Hypercarb® is highly retentive and may be used at extremes of pH is of interest, what really sets this material apart is its unique selectivity. Hydrophobicity has an influence on retention but there is a strong contribution from π-stacking effects and there is often strong retention of rigid planar analyte molecules which are able to interact strongly with the flat planar graphite surface (Figure 4). Accordingly Hypercarb® is the only material used in RPLC that has sufficiently different selectivity from that of alkyl-bonded silicas for it to be claimed to have ''orthogonal'' selectivity from such phases.

3.5 Mobile Phase Optimization

As retention in RPLC is so dependent on hydrophobicity of the analyte, mobile phase and stationary phase, method development in RPLC ought to be fairly simple with retention times of analytes being fairly easy to predict. The issue of course is complicated by the contributions to retention of specific solvent and stationary phase interactions which are difficult to predict a priori. Therefore, rather than using expert systems which use information on analyte, stationary phase and mobile phase properties, it is much more common to choose a column likely to be suitable for the analyte and to carry out computer-aided automated mobile phase optimization in which predictions are based on a few initial experiments with the mobile phase conditions then being refined. One of the most popular packages used for RPLC method development is DryLab in which isocratic and gradient conditions are predicted on the basis of just two initial experiments which are usually gradient runs. This allows the analyst quickly to arrive at suitable mobile phase conditions under which the analytes have the desired retention times. There is a range of more sophisticated software packages available which are designed to work in conjunction with a specific manufacturer's instrumentation to allow totally unattended fine tuning of the separation.[4]

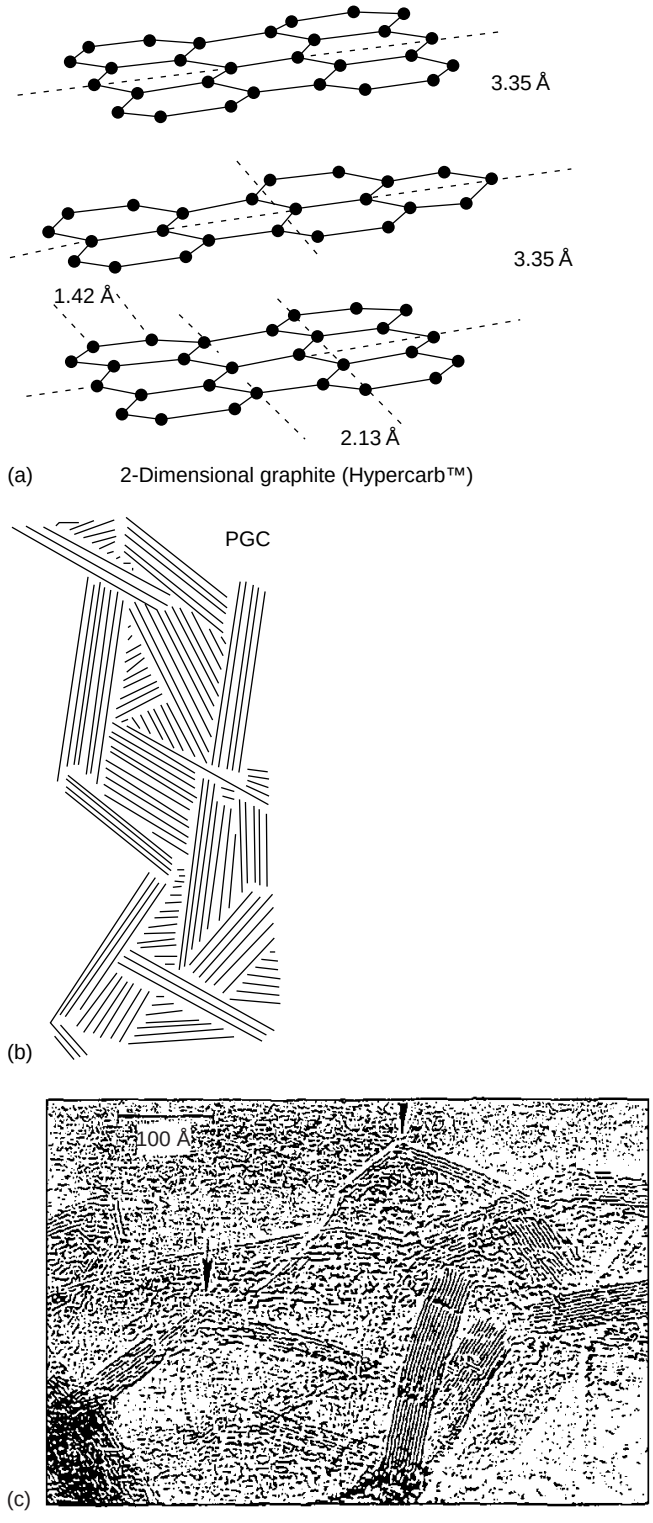

Figure 4 Structure (a), surface cross-section (b) and plan-view (c) of the porous graphitic carbon stationary phase, Hypercarb®.

4 PERSPECTIVE AND FUTURE DEVELOPMENTS

Despite the availability of other modes of LC and alternative techniques such as capillary electrophoresis and capillary electrochromatography, RPLC will continue to be the most frequently used means of carrying out a separative method for the foreseeable future. It is versatile, provides adequate resolution and convenient conditions for the vast majority of applications and, most importantly, because there are vast sums of money invested into the use of RPLC, it will not be replaced until the unlikely event of a new LC mode or alternative technique coming along which is significantly better than RPLC across the board.

Developments in new stationary phases will continue but already much has been done to exploit materials other than silica and to produce silica-based phases which are usable over a wider pH range and/or may be made more reproducible and/or give much reduced peak tailing with basic analytes. As for other modes of LC there will be further moves in RPLC towards the use of columns with reduced dimensions. This is because economic and environmental considerations are now more important. Also, with the increasing use of mass spectrometric detection for RPLC there is less need for complete resolution of all peaks so that it is possible to use shorter columns with faster flow rates for increased sample throughput. One of the main advantages of RPLC is that it may be used for a wide range of analyte polarities. However this often leads to the necessity for solvent gradient methods when the sample is a complex mixture containing analytes with very different polarities. This requires more sophisticated instrumentation and usually involves long analysis times even not allowing for the time required for the column to be reconditioned with the starting mobile phase. The elimination of such lengthy procedures is one of the few remaining research targets in RPLC.

ABBREVIATIONS AND ACRONYMS

DMOA	*N*,*N*-Dimethyloctylamine
ID	Inner Diameter
LC	Liquid Chromatography
MS	Mass Spectrometry
PRP-1	Styrene–Divinylbenzene Copolymer
RPLC	Reversed-phase Liquid Chromatography
THF	Tetrahydrofuran
UV	Ultraviolet

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

REFERENCES

1. C.M. Riley, 'Modes of Chromatography', in *High Performance Liquid Chromatography – Fundamental Principles and Practices*, eds. W.J. Lough, I.W. Wainer, Chapman and Hall, London, 36–78, 1995.
2. C. Horvath, W. Mellander, I. Molnar, 'Solvophobic Interactions in Liquid Chromatography with Non-polar Stationary Phases', *J. Chromatogr.*, **125**, 129–156 (1976).
3. P.J. Slonecker, X. Li, T.H. Ridgeway, J.G. Dorsey, 'Informational Orthogonality of Two-dimensional Chromatographic Separations', *Anal. Chem.*, **68**, 682–689 (1996).
4. W.J. Lough, I.W. Wainer, 'Method Development and Quantitation', in *High Performance Liquid Chromatography – Fundamental Principles and Practices*, eds. W.J. Lough, I.W. Wainer, Chapman and Hall, London, 143–167, 1995.

Silica Gel and its Derivatization for Liquid Chromatography

Uwe D. Neue
Waters Corporation, Milford, MA, USA

1 Introduction **1**
2 Properties of Silica **1**
2.1 Preparation of Porous Silicas 2
2.2 Physical Properties of Silicas 2
2.3 Chemical Properties 6
3 Bonded Phases **6**
3.1 Derivatization Techniques 6
3.2 Monofunctional Phases 8
3.3 Multifunctional Phases 8
3.4 Other Approaches 10
3.5 Properties of Bonded Phases 10
4 Coated Phases **16**
4.1 Preparation Techniques for Coated Phases 17
4.2 Properties of Coated Phases 19
5 Summary **19**
Abbreviations and Acronyms **19**
Related Articles **20**
References **20**

In this article, we review the preparation and use of silica and its derivatives in liquid chromatography (LC). First, the properties of silica that are important for its application in LC are examined, then bonded phases and coated phases based on silica are considered. The current state-of-the-art of the preparation of bonded phases and coated phases is surveyed, and the properties that are relevant for the chromatographer are discussed. Owing to the popularity of reversed-phase liquid chromatography (RPLC), most of the emphasis is placed on phases used in this technique, but other chromatographic techniques are not ignored. A detailed characterization of reversed-phase packings is provided.

1 INTRODUCTION

Bonded phases based on silica gel as well as other silica derivatives constitute the most widely used stationary phases in high-performance liquid chromatography (HPLC). On one hand, this is due to the superior hardness of silica compared with organic gels, which enables the use of the high pressures required to achieve high resolving power at short analysis times. On the other hand, the ability to modify the surface of silica via silanization has contributed significantly to its widespread applicability. While the chemical stability of zirconia or titania is significantly better than that of silica, the surface chemistry of these oxides is more complicated and cannot be manipulated as easily as the surface chemistry of silica. Therefore, the true secret of the success of silica lies in the silanization chemistry.

In this article, the physical and chemical properties of silicas used as sorbents in HPLC will be reviewed first. This discussion includes a brief description of the newer technologies such as nonporous particles and continuous rods. Then silica-based bonded phases will be described in substantial detail: preparation techniques, properties of the bonded phases and general characterization techniques, and the properties and use of specific bonded phases. Owing to the widespread use of reversed-phase chromatography, most of this section is dedicated to C_{18} and C_8 bonded phases; however, we will not neglect specific topics such as special phases designed for carbohydrate separations or protein separations. The final section is dedicated to phases that are coated onto silica. These phases may or may not be attached to the silica surface. We will review the preparation techniques as well as the properties of these phases. It should be noted that some of the stationary phases of this type can also be prepared on other inorganic carriers, but a wider discussion is outside the scope of this article.

2 PROPERTIES OF SILICA

What are the properties of silica that make it desirable for use as a sorbent in HPLC? Chromatographic retention is proportional to the ratio of accessible surface area to packed bed volume. In addition, the amount of analyte that can be injected onto a column is also a function of the surface area in a column. Since the analysis should be insensitive to the amount injected onto a column, the surface area of a packing should be maximized. To achieve a maximum ratio of surface area to volume, the packing should be porous. However, the surface needs to be accessible to the analytes of interest, which puts a lower limit on the pore size and an upper limit on the specific surface area of a packing. The freedom in the manipulation of pore size and surface area is one of the strengths of silica.

In contrast to sorbents based on an organic backbone, silica does not swell nor shrink depending on the solvent in

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

which it is used. Physical properties such as pore volume, pore size distribution and surface area are permanent properties of the packing. Similarly, the permeability of a packed bed is independent of the solvent used. Both effects represent significant advantages in the solvent compatibility of silica-based packings compared with organic polymeric packings. The hardness of the packings also allows the use of higher pressures and smaller particle sizes than are possible with organic polymeric packings.

The third distinguishing property of silica is the ease of modification of the surface using silanization techniques. Silanization allows the design of ''reversed'' phases and many other bonded phases that modify the properties of the native silica substantially and therefore enable specific separation techniques that are not possible with other packings. Most of the discussion in this article is dedicated to the description of the surface modification of silica.

In the following paragraphs, the physical properties of silica and packed beds are first discussed in more detail. Then the chemical properties of silica are examined.

2.1 Preparation of Porous Silicas

In order to frame this discussion, we will start with a brief review of the preparation of porous silica.[1] The common starting point in the preparation is a silicic acid monomer, such as sodium silicate or one of the silicic acid esters, such as tetramethoxysilane. These monomers are polymerized to a specific molecular weight or size. This prepolymerized silica, which may be in the form of a sol, is then further agglomerated to form the final particles. The agglomeration is carried out in the presence of porogens, which are additives that are used to influence the specific pore volume and, together with the size distribution of the prepolymerized sol, the pore size distribution of the silica. In all modern preparation techniques for HPLC-grade silica, spherical particles are generated in this process. The most commonly used technique for the preparation of the silica particles is suspension polymerization, but a spray drying process can be employed as well. The type and purity of the raw materials significantly influence the properties of the chromatographic sorbent. This will be discussed in more detail in section 2.3 which discusses chemical properties of silica, and section 3.5.2 which covers the chromatographic tests for the properties of reversed-phase bonded phases.

The retentive properties of a silica for specific analytes depend on the access of the analytes to its surface and therefore on its pore size distribution. The pore size distribution of a silica prepared in the manner described above depends on the size distribution of the primary silica sol and the specific pore volume. If one wants to change the average pore size of a packing, two techniques are available. The first technique is manipulation of the average size of the sol before preparation of the silica particle. The second is the use of an Ostwald ripening technique to increase the size of the structural units that form the backbone of the silica particle. This second technique is employed after preparation of the particle and is mostly utilized in the production of silica gels with a large pore size, which are commonly used for the chromatography of high-molecular-weight compounds such as polymers and biopolymers. In this technique the silica is subjected to elevated temperature and pressure in the presence of water in an autoclave to achieve the desired pore size distribution. Under these conditions, both the structure of the silica backbone and the pores grow in size. Additionally, if carried out properly, the pore volume remains largely unaffected. Silicas prepared by both techniques can be distinguished easily microscopically (Figures 1 and 2).

2.2 Physical Properties of Silicas

2.2.1 Pore Size, Pore Volume, Surface Area

The pore size distribution, pore volume and surface area are the physical characteristics of a silica that are responsible for its retention properties. Porous silicas used for the chromatography of small molecules have an average pore size of around 10 nm (commonly 5 nm to 15 nm) and a specific surface area between $100\,m^2\,g^{-1}$ and $400\,m^2\,g^{-1}$. Silicas with an average pore size smaller than about 5 nm exhibit restriction to mass transfer in the pores for typical analytes. As a result, chromatographic peaks are broader than they should be. Silicas with a

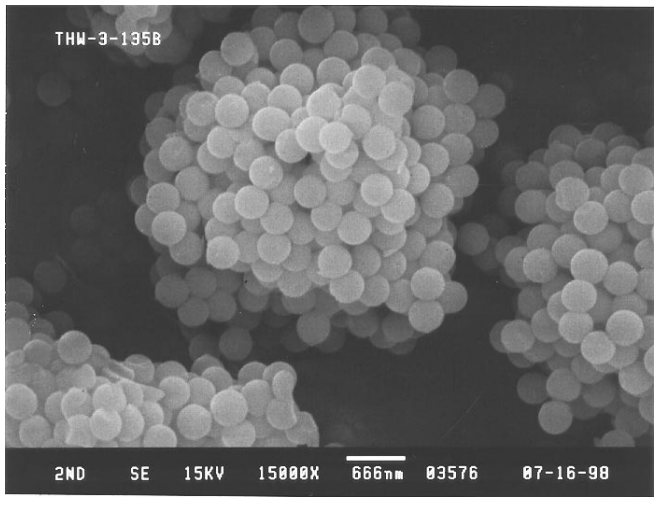

Figure 1 Silica prepared by agglomeration of a silica sol. The agglomeration of the primary sol particles used to form the final particle can be seen clearly. Material: Zorbax™ PSM 3000 (Hewlett Packard, USA).

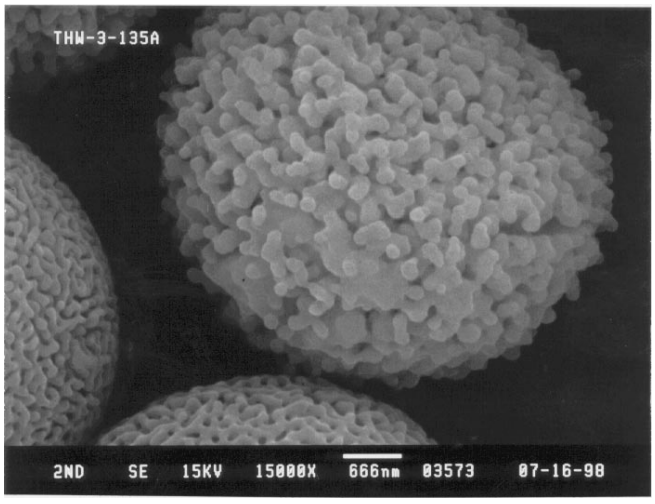

Figure 2 Silica prepared using Ostwald ripening for pore enlargement. The structure reminiscent of corals is typical for a pore-enlarged silica. Material: Nucleosil® 4000–7 (Macherey-Nagel, Germany).

significantly larger pore size have a smaller surface area and therefore less retention for typical analytes. However, as the molecular weight of the analytes increases, the pore size of the most appropriate packing increases as well. The chromatography of proteins is commonly carried out using silicas with an average pore size of 30 nm or more.

The *specific* surface area and the *specific* pore volume are the respective values normalized per gram of packing. They are commonly reported in the manufacturers' literature. While the surface area of a packing is responsible for the retention, it should not be assumed that the retentivity of a silica is proportional to the surface area alone. It should be remembered that the retentivity of a silica is determined by the phase ratio of a packing, which is the ratio of surface area to fluid volume in a packed bed. The phase ratio β can be calculated by Equation (1):

$$\beta = (1 - \varepsilon_i)\frac{A_{sp}}{V_{sp} + (1/\rho_{sk})} \quad (1)$$

Table 1 Characteristic physical properties of some commercial HPLC-grade silicas[a]

Silica	Specific surface area ($m^2\ g^{-1}$)	Specific pore volume ($mL\ g^{-1}$)	Average pore size (nm)	Average pore size (nm) (from Equation 2)	Phase ratio ($m^2\ mL^{-1}$)
Nucleosil® 50	600	0.60	5	3	340
LiChrosorb® 60	550	0.70	6	4	285
LiChrospher® 60	650	0.95	6	4.5	280
ProntoSIL 60	450	0.75	6	5	225
Nucleosil® 100	350	0.80	10	7	170
Partisil®	350	0.80	8.5	7	170
Hypersil® 100	300	0.70	10	7	155
Symmetry®	335	0.90	10	8	150
Chemcosorb®	360	1.00	10	8.5	150
LiChrospher® 100	420	1.25	10	9	145
Kromasil®	340	1.00	10	9	140
µBondapak™	330	1.00	12.5	9	135
YMC Pro™	335	1.06	12.5	9.5	130
Platinum™	210	0.50	10	7.5	130
ProntoSIL 120	300	1.00	12	10	125
Resolve™	200	0.50	9	7.5	125
Waters Spherisorb®	200	0.50	8	7.5	125
Inertsil®	320	1.10	15	10	125
Delta-Pak™ 100	300	1.00	10	10	125
LiChrosorb® 100	300	1.00	10	10	125
Nova-Pak®	120	0.30	6	7.5	95
Hypersil®	170	0.65	12	11.5	90
Hypersil® BDS	170	0.65	13	11.5	90
ProntoSIL 200	200	1.10	20	16.5	80
Discovery™	200	1.00	19	15	80
ProntoSIL 300	150	0.90	30	18	70
Vydac®	90	0.60	30	20	50
Delta-Pak™ 300	125	1.00	30	24	50
Symmetry300™	115	0.85	26	22	50
Hypersil® 300	90	0.70	30	23	45

[a] The source of this information is the column manufacturer's literature. No guarantee of the accuracy of the information can be made.

where ε_i is the interstitial porosity of the packed bed, with a typical value of 0.4 for spherical particles. A_{sp} is the specific surface area of the packing and V_{sp} is its specific pore volume. ρ_{sk} is the skeletal density of the packing, which is 2.2 g cm^{-3} for silica. As we can see, silicas with large specific surface areas and large specific pore volumes have a similar phase ratio, and therefore similar retentivity, to silicas with small specific surface areas and small specific pore volumes. To obtain a true impression of the retentivity of a packing, it is therefore important to calculate the phase ratio rather than considering the specific surface area of a packing alone. This has been done in Table 1 for a range of commercial packings, for which the relevant information is available.

The three parameters, average pore size d_{po}, pore volume V_{sp} and surface area A_{sp} are interdependent: if two are fixed, the third one is determined as well (p. 41 of Unger[1]). This can be expressed by the formula in Equation (2):

$$A_{sp} \approx 3\frac{V_{sp}}{d_{po}}. \quad (2)$$

This approximate formula can be used to estimate the average pore size of a packing. This is quite useful, since there are no commonly accepted conventions for the determination of the average pore size.

The specific pore volume of a silica is the key factor that determines the hardness or strength of a silica. Packings with a small pore volume are stronger than packings with a large pore volume. However, all silicas with a specific pore volume of up to about 1 mL g^{-1} are quite suitable for HPLC applications and can be subjected to column inlet pressures in excess of 40 MPa, which is the common pressure limit of most HPLC instrumentation.[2]

2.2.2 *Particle Size*

The particle size of a packing determines the chromatographic efficiency and the backpressure of columns packed with these particles. The most commonly used particle size is 5 µm for analytical separations, but 10 µm particles are still employed, while the application of columns packed with 3-µm particles is steadily increasing. Columns packed with fully porous particles down to a particle size of 1.5 µm are commercially available, but their usefulness has not yet been sufficiently explored. For preparative applications, larger particle sizes are frequently employed owing to the pressure limitation of the chromatography instrument on the one hand and the lower demand on the separation power of the column on the other.

Pressure limitations also play a role in the use of analytical columns. Commonly, the preferred operating range is at pressures less than 20 MPa, but most columns are stable at pressures exceeding twice this value. The pressure limitations are related to the equipment rather than the columns. The pressure needed to use a column at a particular flow rate increases with increasing column length. Also, at a given column length, the pressure is inversely proportional to the square of the particle size. This puts a general limit on the ratio of column length to particle size.

It is worthwhile exploring the properties of column length and particle size a little further. Generally[2] the maximum plate count of a column depends simply on the ratio of column length to particle size. Similarly, the shortest analysis time that can be reached at a given pressure limit is also simply a function of the ratio of column length to particle size. Columns with a larger length to particle size ratio achieve a higher plate count, but are slower. Columns with a smaller length to particle size ratio are faster, but have a lower plate count maximum. The particle size, together with the mobile phase viscosity and the molecular weight of the analyte, determines the analysis time at which the optimum column performance is achieved. It occurs at a slower analysis time for larger particles than for smaller particles. The analysis time at which an equal plate count maximum is achieved decreases with the square of the particle size at a constant length to particle size ratio. However, the pressure limitations of the instrumentation set a limit to both the plate count and the speed of a particular analysis, even if this optimization process is followed. An additional limit to column design has been pointed out by Halász and colleagues.[3] The pressure needed to pump the mobile phase through the column generates a significant amount of heat, which measurably affects retention for particles of 3 µm or less.

Nearly all HPLC packings have a fairly wide particle size distribution, most commonly extending to ±25% around the mean. In addition, what is given as the average particle size depends on the measurement technique[2] and no standard technique has been established. Therefore the ''5-µm'' particles of one manufacturer are not necessarily identical to the ''5-µm'' particles of another. However, since the backpressure of a column depends on the particle size of the packing, this observation can be turned around and the average particle size d_p can be determined from the backpressure Δp of a column.[3] The procedure requires knowledge of the flow rate F, the viscosity of the mobile phase η and the column dimensions, (Equation 3):

$$d_p = \sqrt{1000\frac{F\eta L}{\pi r^2 \Delta p}} \quad (3)$$

This very practical procedure puts the determination of the particle size in the hands of the HPLC user.

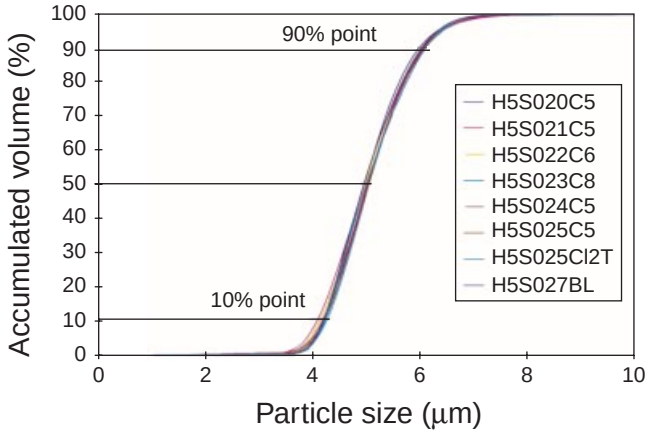

Figure 3 Particle size distribution of a modern HPLC packing. The characteristic values are the center of the distribution at the 50% point and the ratio of the 90% point to the 10% point, which is a measure of the width of the distribution.

Both the efficiency of a column and the column backpressure depend on the width of the particle size distribution. The permeability of a column, and thus its backpressure, depends cleanly on the square of the population-average particle size. However, column performance as measured by the *C*-term of the van Deemter equation,[4] seems to depend on the volume-average particle size.[2] Since the difference between population-average particle size and volume-average particle size decreases with the width of the distribution, narrower particle size distributions are preferred over broader distributions.

The value commonly used for an assessment of the width of a particle size distribution is the ratio of the 90% point to the 10% point, often of the volumetric distribution (Figure 3). For today's high-quality packings, this value may be around 1.5 or smaller.

2.2.3 Porous Versus Nonporous Particles

Nonporous particles with a small particle size (down to 1.5 μm) have now become commercially available. Due to the low permeability of packed beds prepared from such a small particle size, the columns used are quite short. Fast analyses can still be achieved with reasonable column performance, albeit at high backpressure. However, a major drawback of nonporous particles compared with fully porous particles is the significantly lower surface area per column volume. The amount that can be injected on a column without peak distortion depends on the surface area of the packing. For small molecules, the mass loadability of a packed bed prepared from nonporous particles is over three orders of magnitude lower than the loadability of the same bed prepared from fully porous particles. In addition, significant differences in the mobile-phase composition are required to obtain the same retention with nonporous and with porous particles. For the analysis of small molecules, the use of nonporous particles does not provide any benefits compared with porous particles.

However, the situation changes when the retention chromatography of macromolecules is considered, such as proteins and nucleic acids. For unrestricted access of molecules of a very large molecular weight to the surface of the packing, the pore size of the packing needs to be very large. Under these circumstances, the surface area available for interaction is not much larger for a fully porous packing and a nonporous packing, and quite good performances have been achieved[5–9] using nonporous particles. For analytical separations of macromolecules, the use of nonporous particles makes sense. A review by Lee[9] deals with the separation of proteins using nonporous sorbents.

2.2.4 Continuous Structures

The newest development in the design of separation devices are continuous rod columns.[10–12] In this case, the chromatographic bed consists of a continuous porous structure with a dual pore size, containing micropores that provide the surface for chromatographic retention and macropores that provide the flow channels, like the interstitial space in a packed bed. The advantage of such a continuous structure lies in the fact that the size of the macropores can to some degree be designed independently of the size of the structural units that contain the micropores. Consequently, structures can be designed that provide a higher performance at a given permeability or a higher permeability at a given performance than packed beds can provide.

At the time of writing (1998), the practical limitations of continuous rod columns are not known, although just as in a packed bed, the structural uniformity of the rod plays a significant role in its performance. For bonded phases, uniform derivatization of the chromatographic surface needs to be achieved. While the capabilities and limitations of packed bed technology are reasonably well understood, it is not clear if the continuous rod technology gives similar freedom in the design and manufacturing of chromatographic beds of widely varying aspect ratios and widely varying surface chemistry.

To evaluate continuous rod structures (or other chromatographic devices) and compare them to packed beds, we can follow the thought process outlined by Neue on page 73 of his book.[2] The chromatographic structure is treated as a black box, whose internal makeup is unknown. From the permeability of the device, the particle size equivalent of a packed bed can be calculated (Equation 3). The height equivalent to a theoretical plate

(HETP) is then measured as a function of the linear velocity. Then, the particle size equivalent is used to create the reduced plate height and the reduced velocity, just as if the device were a packed bed. In this way, all chromatographic devices are compared with each other using exactly the same framework: separation power, speed of analysis and pressure.

2.3 Chemical Properties

The chemical properties of silica vary somewhat with the preparation technique and the purity of the raw materials used in the preparation.[1] Nevertheless, a range of general statements can be made that are universally applicable.

The surface of a chromatographic silica is covered with silanol groups, which are responsible for the polar adsorption properties of the silica, and with siloxane bridges. Several different types of silanol can be distinguished, structures **(1)** and **(2)**: single or lone silanols, bridged or vicinal silanols which emerge from neighboring silicon atoms and are joined by a hydrogen bridge and geminal silanols, which are attached to the same silicon atom. Silanols are weakly acidic adsorption sites. They are also the sites at which further chemical modification of the silica is initiated.

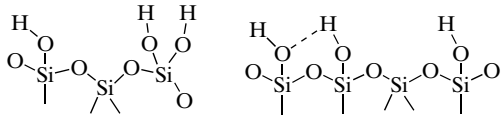

(1) **(2)**

A fully hydroxylated silica contains about $8\,\mu mol\,m^{-2}$ of silanols, or about five silanol groups per square nanometer. In water, the silanols dissociate at pH values higher than about 3. Therefore, silica is a weak cation exchanger. Consequently, the surface of a silica may contain cations unless it has been treated with an acid. If waterglass (sodium silicate) is used as a raw material, the surface of the silica may be occupied by sodium ions. Cations on the surface of the silica can be removed fairly easily by treatment with acids. However, if the cations are incorporated in the matrix, they can not be removed. Apart from sodium, aluminum or iron are common contaminants of classical silicas. The metal ions in the silica matrix affect the activity, or better said the acidity, of the surface silanols. This in turn can lead to undesirable interactions with analytes, especially in the chromatography of analytes with basic functional groups in reversed-phase chromatography. Therefore, newer high-purity silicas were developed in the late 1980s and early 1990s, based on organic silanes such as tetramethoxysilane as raw materials. Reversed-phase packings based on high-purity silicas have rapidly gained acceptance owing to their superior chromatographic performance with basic analytes. An excellent review of the influence of silica structure and impurities has been published by Cox.[13]

When silica is heated, neighboring silanols condense and form siloxane bridges, leaving behind lone silanols. At temperatures higher than about 800 °C, most of the silanols have been removed and the structure of the silica starts to collapse, resulting ultimately in sintering of the particles. At lower temperatures, the siloxane formation is reversible and silica treated at elevated temperature can be rehydroxylated. The preparation of many silicas requires the addition of an organic polymer to influence the pore volume and this polymer is often removed by treatment at elevated temperature. Since the high-temperature treatment removes silanols, rehydroxylation of the silica surface is an important step in the preparation of many silicas. The activity of a silica depends on the details of its temperature history.

3 BONDED PHASES

Bonded phases can be broadly grouped into two categories, nonpolar bonded phases used for reversed-phase applications and bonded phases with polar functional groups that can be used in normal-phase chromatography, hydrophilic-interaction chromatography, size-exclusion chromatography, or ion-exchange chromatography. Owing to the popularity of reversed-phase chromatography, most of the work on the preparation and characterization of bonded phases has focused on this class of packings. In this section, we will first concentrate on the different approaches used to derivatize silica via silanization techniques. Classical techniques as well as some of the newer techniques will be discussed. Then the properties of specific bonded phases, their characterization and their use will be examined.

3.1 Derivatization Techniques

The preparation of a bonded phase commonly starts with a clean fully hydroxylated silica. Silanization of the surface can be carried out with chlorosilanes, alkoxysilanes, aminosilanes or other reactive silanes. The classical approach to the preparation of reversed-phase packings uses chlorosilanes in the presence of a scavenger of the hydrochloric acid by-product (Scheme 1). Inexpensive organic bases such as pyridinium or triethylamine are commonly used to remove the hydrochloric acid.

Bonded phases with polar functional groups are usually prepared using alkoxysilanes. Two approaches towards this reaction have been reported in the literature. In

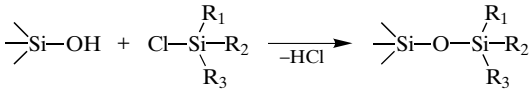

Scheme 1

the first approach (Scheme 2), the alkoxysilane is reacted directly with the silica surface in a hydrophobic solvent such as toluene.[14] The side product of this reaction is an alcohol, which can subsequently react with surface silanols to form an alkoxy-derivatized silica. Therefore, the alcohol needs to be removed from the reaction mixture.

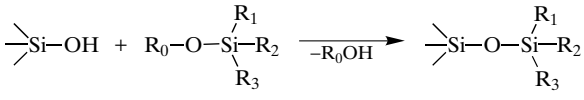

Scheme 2

The second approach towards polar bonded phases is the derivatization of the silica in an aqueous medium. In this approach, the silane polymerizes initially and the reaction with the surface occurs in a secondary step. This is shown schematically below; the secondary step is simplified as the formation of a trisilanol and its condensation with the silica surface (Scheme 3):

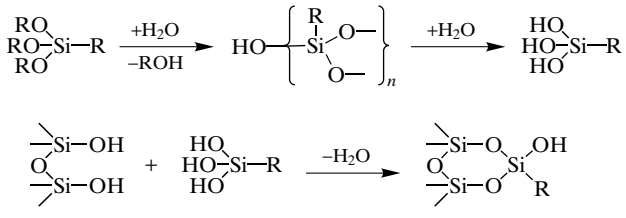

Scheme 3

Despite the complexity of this reaction, good results have been obtained with it. For example, the reaction of silica with glycidoxypropylsilane results in materials that are suitable for aqueous size-exclusion chromatography owing to a low level of interaction with analytes in aqueous mobile phases.

Other derivatization techniques have been reported as well. A high-temperature derivatization with hexamethylcyclotrisiloxane[15] has been patented. The quality of the surface modification appears to be quite good. Also, other reactive silanes can be used.

The level of the surface derivatization can be assessed through elemental analysis. Manufacturers of stationary phases often report typical values for the carbon content of silica-based bonded phases. However, a better picture of the nature of the surface is obtained when the surface coverage which expresses the molar amount of ligand attached to the surface per unit surface area, is calculated. The surface coverage χ for a monofunctional silane is calculated in Equation (4):

$$\chi = \frac{\%C}{100A_{sp}n_c12\left(1 - \frac{\%C}{100}\frac{MW-1}{n_c12}\right.} \qquad (4)$$

where $\%C$ is the carbon content obtained from elemental analysis, A_{sp} is the specific surface area of the underivatized silica, MW is the molecular weight of the ligand and n_C is the number of carbon atoms in the bonded phase molecule. Equation (4) takes into account the weight gain of the sample due to the incorporation of the bonded phase. For bi- and trifunctional silanes, the weight gain correction is not quite as straightforward. Probably the best approach is to assume that the bonded phase is attached to the surface via two Si−O bonds or to measure the actual distribution using ^{29}Si-CP/MAS/NMR (cross-polarization/magic angle spinning/nuclear magnetic resonance).

The silane used for the derivatization can be a monofunctional **(3)**, difunctional **(4)** and **(5)** or trifunctional silane **(6)** and **(7)**. Depending on this characteristic, the silane can bond to the surface with one or two chemical bonds. The formation of a triple bond with the surface can be excluded for steric reasons. However, with multifunctional silanes the formation of siloxane bridges to neighboring ligands is possible as well.

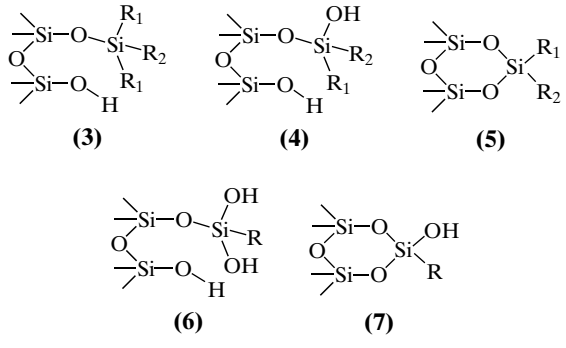

When a monofunctional silane is reacted with the silica surface, the surface coverage is limited by steric hindrance. Only less than about 4 μmol m^{-2} of a monofunctional silane fit onto the surface.[14] For difunctional silanes, a larger surface coverage can be achieved. However, it seems that the removal of surface silanols is still incomplete.[16–18] For every silane that binds to the surface, only 1.5 silanols are on the average removed from the surface. With trifunctional silanes, the number of silanols consumed per bound silane is limited to two anyway; therefore, every trifunctional ligand that is attached to the surface generates at least one new silanol group.

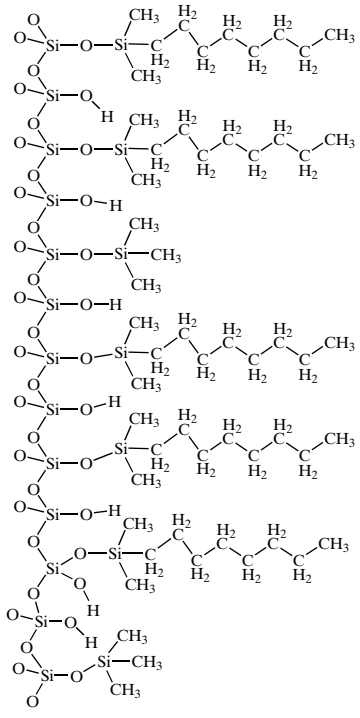

(8)

While an additional removal of silanols should not be impossible, the literature information is inconclusive.

Further removal of silanols from reversed-phase packings after the primary bonding procedure has taken place can be accomplished by a process called ''endcapping''.[1] In this case, an active trimethylsilane is used to bond to silanols that were either formed in the primary reaction (multifunctional silanes) or did not react due to steric hindrance. Commonly, trimethylchlorosilane, hexamethyldisilazane, mixtures of both or other reactive trimethylsilanes have been used.[19–21] Recently, high-temperature endcapping reactions have been explored.[22,23] Generally, the endcapping process is quite effective and the removal of silanols is measurable. However, complete removal of silanols is impossible for steric reasons.[24,25] **(8)** is a schematic representation of the surface of a typical reversed-phase packing, a fully endcapped C_8 bonded phase with a high level of derivatization. Note that there are about as many silanols left on the surface as have been reacted with either the C_8 reagent or the endcapping reagent. Even for highly derivatized packings, ''residual'' silanols contribute to the retention properties of the packing.

3.2 Monofunctional Phases

Monofunctional phases are derived from silanes that can react with the silica surface through one single bond only. Many reversed-phase stationary phases are based on monofunctional silanes. The largest advantage of these phases is the reproducibility of the preparation. When multifunctional silanes are reacted with the silica surface in the presence of small amounts of water, additional silanol groups are created through the hydrolysis of the reactive group. As a consequence, the reaction can proceed further than in the absence of water. This does not occur when monofunctional silanes are reacted with the silica surface. The high reproducibility of the preparation of the stationary phases based on monofunctional silanes has been demonstrated in the literature.[26,27]

A disadvantage of the use of monofunctional silanes is the ease with which the single bond to the surface is cleaved. Especially at pH values below 2, the weakest point of the attachment of the bonded phase is the Si–O–Si linkage. However, the nature of the ligand influences the stability of the bonded phase at low pH. Kirkland et al. have described[28,29] a group of bonded phases that improve this problem: they use ligands with large bulky side chains that sterically protect the point of attachment of the ligand to the surface, **(9)** and **(10)**.

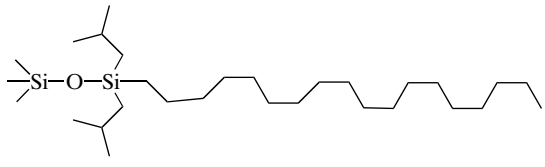

(9)

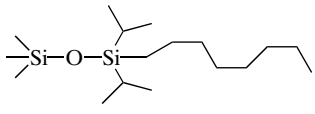

(10)

The improvement in hydrolytic stability at low pH is excellent compared with simple monofunctional phases with methyl groups as the side chains. However, the larger steric hindrance also results in a lower coverage, exposing more silanols. Therefore, the adsorptive properties of these phases are inferior to densely coated classical phases, especially around neutral pH, leading to tailing for basic analytes.

3.3 Multifunctional Phases

The classical approach towards the preparation of multifunctional phases uses di- and trifunctional silanes that can bond to the surface via multiple bonds. Improvements in hydrolytic stability result at low pH. At high pH, however, the dissolution of the silica itself determines the hydrolytic stability of a phase, and therefore the type of bonded phase plays only a secondary role.

Generally, the reproducibility of the preparation of multifunctional phases is inferior to monofunctional phases and the bonding reactions need to be controlled more closely. A key factor in the preparation of multifunctional phases is the influence of trace amounts of water on the reaction. For chlorosilanes, the presence of water in the reaction medium creates a polymerization reaction, which influences the nature and the amount of ligand attached to the surface. On the other hand, water acts as a catalyst in the reaction of alkoxysilanes with silica and is an important ingredient in the preparation of bonded phases prepared from alkoxysilanes.[16,30]

Most phases designed for applications other than reversed-phase chromatography are based on multifunctional silanes. Typical examples are the propylamino phase or the diol phase. Therefore, we need to realize that the presence of unprotected and unshielded silanols plays a role in the adsorptive properties of these phases. This will be discussed in more detail in the section that deals with specific bonded phases.

An alternative approach to the design of stationary phases has been discussed by Kirkland et al.[28] The ligands that he attached to the surface contain two reactive silanes linked by a short hydrocarbon chain **(11)**:

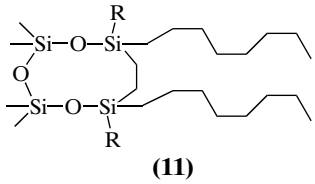

(11)

It is expected that the hydrolytic stability of such phases is improved over classical bonded phases, if indeed both silicons are attached to the surface. The details of this reaction are still unknown.

Another more recent approach is the attempt to form initially a monolayer of a trifunctional silane on the silica surface followed by bonding this monolayer to the surface.[31,32] However, the chromatographic evaluation of these materials in the literature is only sketchy. An evaluation in our own laboratory[33] demonstrated an excellent chemical stability of the packing, but the silanol activity was very high compared with modern HPLC packings. A key assumption underlying the ''horizontal'' polymerization technique is the flatness of the surface, which is not given for porous packings.

The use of multifunctional silanes allows for a high coating level on the surface. Sander and Wise have shown[34,35] that the relative retention and therefore the resolution of polynuclear aromatic hydrocarbons depends on the density of the C_{18} chains of a reversed-phase packing (Figure 4). Therefore, C_{18} bonded phases with a high coating level based on

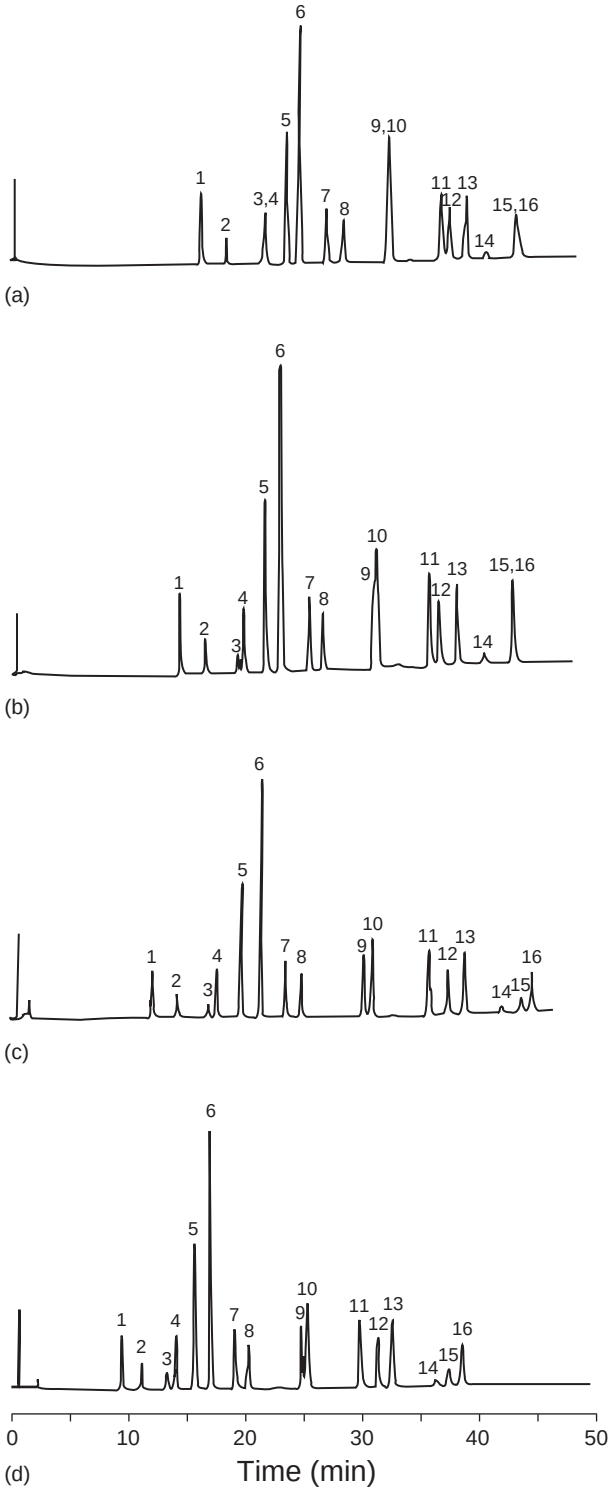

Figure 4 Separation of polynuclear aromatic hydrocarbons on C_{18} phases with different coating levels. (a) 6 nm pores, $\alpha_{TBN/BaP} = 1.32$; (b) 10 nm pores, $\alpha_{TBN/BaP} = 1.24$; (c) 15 nm pores, $\alpha_{TBN/BaP} = 0.84$; (d) 30 nm pores, $\alpha_{TBN/BaP} = 0.85$ (BaP, benzo[*a*]pyrene; TBN, 1,2-3,4-5,6-7,8-tetrabenzonaphthalene.) (Reproduced from Sander and Wise, *J. Chromatogr. A*,[34] Copyright (1993) with permission from Elsevier Science.)

trifunctional silanes have found a specific application niche in the environmental analysis of polynuclear hydrocarbons.

Along the same lines, a polymeric C_{30} stationary phase[36] has been developed that is capable of separating the geometric isomers of common carotenoids. Comparisons with traditional C_{18} phases demonstrated the superior resolving power of the C_{30} bonded phase for this class of compounds.[37]

3.4 Other Approaches

There are a few interesting approaches to the design of special stationary phases that have not yet been covered in this discussion. One is the design of immobilized artificial membranes (IAMs),[38–41] consisting of a monolayer of phospholipid covalently bonded to the surface of a silica (Figure 5). The IAM derivatized silica is a chemically stable chromatographic packing that emulates the exterior of a biological cell membrane. The logarithm of the retention factor of drugs measured on an IAM column correlates well with their permeability through Caco-2 cells[42] or their absorption by inverted rat intestine.[43] Therefore, fast chromatographic methods can be used to estimate the drug permeability through biological membranes. Recently, IAM columns have been used for rapid screening of drug–membrane interactions in drug discovery.[44]

Another interesting technology is the use of reversed-phase packings designed with an external hydrophilic surface. These packings, called internal-surface reversed-phase or restricted access packings,[45,46] are designed for the chromatographic analysis of drugs directly from blood plasma. The internal hydrophobic surface on the inside of the pores of the packing is responsible for the retention of drugs and metabolites. However, proteins interact only with the external hydrophilic surface and are excluded from the internal surface owing to their size. The binding of proteins to the external hydrophilic surface is weak and they are eluted without difficulty. On classical reversed-phase packings, proteins would stick to the external surface and rapidly clog the column. Therefore, improved column life is possible with restricted-access packings, but the internal surface can still become contaminated with lower-molecular-weight contaminants that can penetrate the pores.

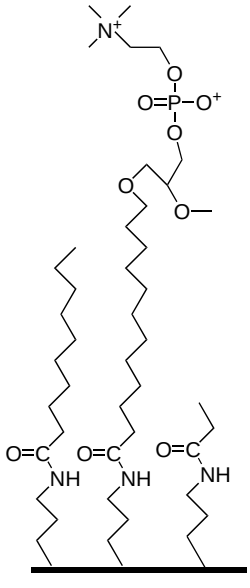

Figure 5 IAM. The phase designated ^{ether}IAM.PC$^{C10/C3}$ by Ong et al.[38] is shown.

In one approach,[45] a hydrophobic tripeptide is bonded to a diol-derivatized silica with a standard pore size. Subsequently, the peptide is clipped off the external surface of the packing with an enzyme that is unable to penetrate the smaller pores of the packing. Consequently, the external surface becomes hydrophilic, showing little interaction with serum proteins, while the internal surface is responsible for the reversed-phase retention of the smaller analytes. In the second approach,[46] a standard hydrophobic surface is created first via a standard silanization technique. Then, the packing is immersed in an aqueous acid. Surface tension does not permit the penetration of the aqueous acid into the hydrophobic pores, but the coating external to the pores of the packing is cleaved off. In the final step, the external surface is coated with a diol silane to prevent the interaction of proteins with the external surface.

In order to create an analog to lipid membranes for separation with molecular recognition, a comb-shaped bonded phase was devised.[47] The ligand was synthesized by telomerization of 3-thiopropyltrimethoxysilane and octadecylacrylate and subsequently attached to the surface of silica. This bonded phase showed a remarkable temperature dependence of the retention factor and of the relative retention for aromatic hydrocarbons, indicating a phase transition of the immobilized ligand.

3.5 Properties of Bonded Phases

In this section, the specific properties of bonded phases and their use will be examined. Since the largest use of columns is for reversed-phase chromatography, this application area will be discussed first. Included in the discussion is the description of evaluation techniques for reversed-phase packings and the reproducibility of the preparation technique. Then we will discuss normal phase chromatography, followed by a brief review of other application areas.

3.5.1 Phases for Reversed-Phase Chromatography

The most commonly used phase for reversed-phase chromatography is the C_{18} phase, in which an octadecylsilane **(12)** is attached to the silica. This is followed by the closely related C_8 phase which uses an octylsilane **(13)**.

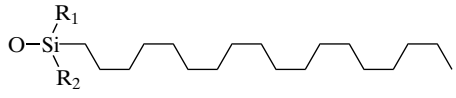

(12)

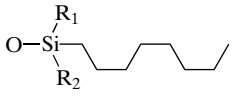

(13)

The retention of hydrophobic analytes in a given mobile phase is generally larger for a C_{18} phase than for a C_8 phase, and generally proportional to the ratio of the amount of stationary phase to the mobile phase volume, similar to the phase ratio discussed earlier. Therefore, C_{18} phases with a low surface coverage may exhibit retention properties similar to highly coated C_8 phases. For more polar analytes, the situation is more complicated: polar analytes may penetrate the bonded-phase layer only partially and therefore interact with only part of the bonded phase.[48,49] In such a case, the retentivity of a C_{18} and a C_8 phase may be very similar. Conversely, polar, especially basic analytes may be more strongly retained on packing materials with a high activity of the silanol groups. Therefore, the assumption that a C_{18} packing exhibits more retention than a C_8 packing should be treated with caution.

Most C_{18} and C_8 packings are based on silica with a pore size of around 100 Å (1 Å = 10^{-10} m). This pore size is marginal for peptide separations and too small for the separation of proteins. For this application area, packings with a pore size of 300–500 Å have been designed. Commonly, a gradient is used from 5% or 10% acetonitrile to a high percentage of acetonitrile, with an acidic buffer based on trifluoroacetic acid. Some other mobile phase options are available as well, but are only rarely used.

Reversed-phase packings with an incorporated polar functional group are closely related to the classical C_8 and C_{18} phases. Phases in this group are SymmetryShield™ RP_8 **(14)** and SymmetryShield™ RP_{18} from Waters, USA,[50] Discovery™ RP-AmideC16 and its predecessor Supelcosil™ ABZ + **(15)** from Supelco, USA,[51,52] and the Prism™ column from Keystone Scientific, USA.[53] The common factor among these phases is the fact

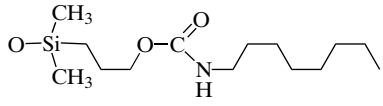

(14)

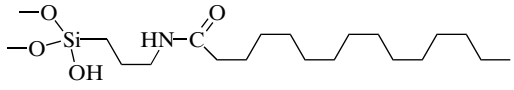

(15)

that a polar functional group is incorporated into the hydrophobic bonded phase.

The incorporation of the polar functional group adds several advantageous properties to these packings. Most important, it suppresses the influence of silanol groups on the retention of basic analytes much more effectively than any other approach used. As a consequence, excellent peak shapes are obtained for basic compounds that exhibit tailing even on the best C_8 or C_{18} packings. Another consequence of the suppression of the influence of silanols is the fact that phases with an incorporated polar functional group exhibit different selectivities from classical C_8 and C_{18} packings, while the overall retentivity is similar to comparable C_8 or C_{18} packings. The third advantage of these phases is the fact that they do not exhibit ''hydrophobic collapse'' as do highly coated classical phases. Hydrophobic collapse is a the name given to a phenomenon observed for the most deactivated reversed-phase packings in highly aqueous mobile phases: suddenly the retentivity of the packing is lost when the flow through the column is stopped for a moment. This occurs because the polar mobile phase is driven out of the pores of a hydrophobic packing owing to surface tension effects,[54] specifically the high contact angle between water and the hydrophobic surface. Since the incorporated polar functional group reduces the contact angle with water and makes the surface more ''wettable'', this phenomenon does not occur.

3.5.2 Evaluation Techniques for Reversed-phase Packings

Reversed-phase packings exhibit a wide range of retention properties that vary with the nature of the analyte. This has created significant interest in a universal characterization of the retention properties of reversed-phase packings. A better understanding of the selectivity of reversed-phase packings would help a user in the selection of a packing.

Several column tests are available that characterize relevant properties of reversed-phase packings. A test developed by Sander and Wise[34,35] is designed to characterize the shape selectivity of a packing, which depends

on the chain spacing of the bonded phase. The selectivity for the shape of the analytes is especially important in the analysis of polynuclear aromatic hydrocarbons. The test mixture, which is available as a National Institute of Standards and Technology (NIST) standard reference material, comprises three aromatic hydrocarbons, BaP, phenanthro[3,4-*c*]phenanthrene (PhPh), and TBN. The elution order of these compounds depends on the coating level, and thus the chain spacing, of a C_{18} bonded phase. For packings with a surface coverage under 3.5 μmol m^{-2}, the elution order is BaP ≤ PhPh < TBN. These packings are considered to be monomeric packings. For polymeric packings with a higher surface coverage, the elution order is PhPh < TBN ≤ BaP. The relative retention between TBN and BaP has been found to vary from 0.6 to 2.2 for a range of commercial packings. This test is useful for the characterization of the density of the surface coating of a packing and can be used to predict the properties of a packing for the separation of aromatic hydrocarbons.

Tests developed by Engelhardt and co-workers[55–57] are primarily designed to test the hydrophobicity and the silanol activity of a reversed-phase packing. The test mixture varied slightly over the history of the test, but consists fundamentally of hydrophobic probes, toluene and ethylbenzene, a mildly acidic probe, phenol, and basic probes such as aniline, *N*,*N*-dimethylaniline or toluidines. The mobile phase is a mixture of 49% methanol and 51% water (w/w). The relative retention between the basic samples and toluene can be used to assess the silanol activity of a packing. High values indicate a high silanol activity. In addition, an increase in tailing for the basic compounds is an indicator of silanol activity. Furthermore, the co-elution of the toluidine isomers is characteristic of a silica with low silanol activity.

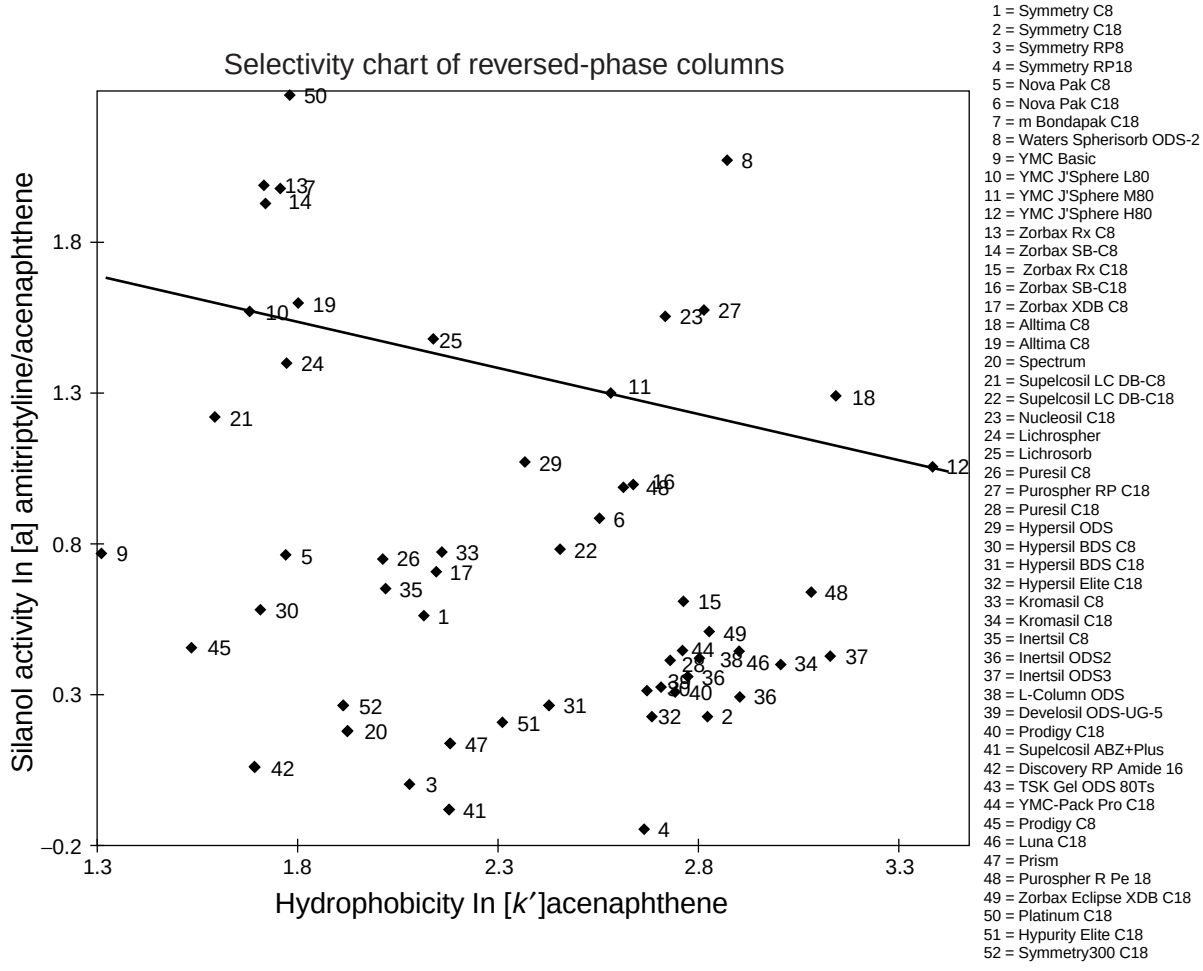

Figure 6 Selectivity chart of commercial C_8 and C_{18} packings. The *x*-axis is the natural logarithm of the hydrophobicity of the packing, measured by the retention factor of acenaphthene. The *y*-axis is the natural logarithm of the silanol activity, measured by the relative retention of amitriptyline and acenaphthene at pH 7.00. (The graph is based on data provided by B.A. Alden, Waters Corp.)

A derivative of the Engelhardt test has been developed by Neue and Serowik et al.[2,58,59] The principle of both tests is similar: a mixture of neutral, acidic and basic analytes is used to characterize the overall properties of a packing. The hydrophobicity of a packing is measured via the retention of an aromatic hydrocarbon such as toluene or acenaphthene and the silanol activity is measured using basic analytes such as tricyclic antidepressants or propranolol. These basic analytes exhibit an especially strong interaction with silanols, which makes it possible to run this test in a buffered mobile phase consisting of 65.0% methanol and 35.0% of a 20 mM phosphate buffer at pH 7.00. This test, which originally was designed to test the batch-to-batch reproducibility of the same packing, can also be used to classify different packings. In Figure 6 the results are shown for a multitude of packings. The hydrophobicity is obtained from the retention factor of acenaphthene and the silanol activity is derived from the relative retention of amitriptyline and acenaphthene. Packings with a low hydrophobicity are found on the left side of the graph, packings with a high hydrophobicity on the right side. Packings with a low silanol activity are found on the bottom of the chart, unendcapped packings with a high silanol activity on the top of the chart. Note that both axes are logarithmic! C_8 packings are commonly found on the left side of the chart and C_{18} packings on the right side. Generally, packings based on high-purity silicas with a high coating level cluster together in the lower part of the chart. A line can be drawn between three different packings that are based on the same silica but differ from each other in the C_{18} coating: YMC J'Sphere L80, YMC J'Sphere M80 and YMC J'Sphere H80. This demonstrates the general validity of the chart: at low coating (YMC J'Sphere L80), more silanols are available for interaction, resulting in a high silanophilic activity together with a low hydrophobicity; at high coating (YMC J'Sphere H80), the silanol activity decreases and the hydrophobicity increases. Packings with an incorporated polar functionality are found at the very bottom of the chart, indicating that this design of the bonded phase indeed reduces the influence of silanols on the retention of reversed-phase packings.

Similar tests were developed by Tanaka et al.[60] He proposed use of the retention factor of amylbenzene in 80% methanol as a measure of the number of alkyl chains bonded to the silica and the relative retention between amylbenzene and butylbenzene as a measure of the hydrophobicity of a packing. In analogy to the procedure used by Sander and Wise, the relative retention between triphenylene and *o*-terphenyl is used as a measure of the steric selectivity of a packing. The relative retention between caffeine and phenol measured in 30% methanol/water is used as a measure of the hydrogen bonding capacity of the packing, which depends on the amount of underivatized silanols on the surface of the packing. In addition, the relative retention of benzylamine and phenol is measured at acidic and neutral pH with 30% methanol, as a measure of the ion-exchange sites at both pH values. Since the silanol sites are also the ion-exchange sites, one would expect a reasonable correlation between the relative retention of caffeine/phenol and the relative retention of benzylamine/phenol at neutral pH. However, a buffered mobile phase provides a more controlled environment that is independent of the state of the surface silanols left from the prior history of the packing. The differences between the amount of ion exchange measured at acidic pH and at neutral pH is a measure of the acidity of the surface silanols and is therefore related to the temperature history and the purity of the silica.

The evaluation techniques discussed above have all been used to characterize the retention properties of different packings, but a large scale comparison between the different methods is not yet available. However, smaller scale comparisons reveal a reasonable correlation between some aspects of the different tests. The test developed by Neue and Serowik was originally designed as a batch-to-batch reproducibility test. Therefore, its reliability, reproducibility and simplicity are notable. Tanaka's procedure involves several separate tests, which increases the complexity, but allows the measurement of features that escape the Neue/Serowik test. The main strengths of the Engelhardt test are the selection of test compounds that are readily available to anybody, and a very large body of data.

3.5.3 Batch-to-batch Reproducibility

An important aspect of a packing is the reproducibility of the separations obtained when a new column packed with the same type of packing is used. In general, the reproducibility is impeccable if the same batch of packing is used. However, when different preparations of the same packing are compared with each other, differences in the retention properties can be observed. From some manufacturers, a sufficient part of the specifications of the packing is available to allow a comparison of the reproducibility. A key element is the surface coverage of a packing, which extends beyond the carbon content. For packings for which this information is not available, the user can get an estimate from the specification of the carbon content and the specific surface area, using Equation (4). Examples of the allowable ranges of surface coverage (and therefore reproducibility) are 2.75–3.75 $\mu mol\,m^{-2}$ for Luna™ $C_{18}(2)$ (Phenomenex, USA) or a range of 3.09–3.31 for Symmetry™ C_{18} (Waters Corp., USA). Unfortunately, the information given by

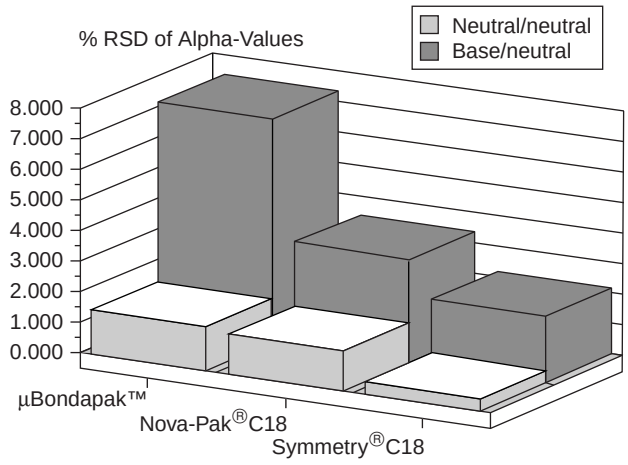

Figure 7 Batch-to-batch reproducibility of three commercial packings. The relative standard deviations (RSDs) for two characteristic parameters are shown: in the foreground, the batch-to-batch reproducibility of the relative retention between two neutral compounds; in the background, the same data are shown for a pair of basic analyte and neutral analyte.

many manufacturers is insufficient to make a judgement on the reproducibility of a packing. However, the characterization tools described in the last section can also be used to test the reproducibility of the preparation of a packing, if proper controls of the mobile-phase composition and column temperature are implemented.[26,27] The best way to assess the reproducibility of a packing is to look at the reproducibility of both a compound pair of very similar properties and a compound pair of widely different properties. An example of the first group is the relative retention of neutral, hydrophobic analytes, such as acenaphthene and toluene. An example of the second group is the relative retention between a basic analyte, such as amitriptyline and a neutral analyte, such as acenaphthene. These data are shown in Figure 7 for three commercial packings. The data were collected over a period of several years.

3.5.4 Other Reversed-phase Packings

Owing to the broad range of available properties, most reversed-phase applications are accomplished with either C_{18} or C_8 columns. Nevertheless, other types of stationary phases have found their niche. The most popular types are phenyl columns **(16)** and cyano columns **(17)**, whose structures are shown below.

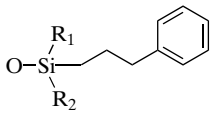

(16)

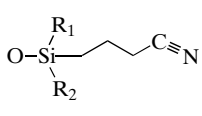

(17)

There are several possibilities for the link between the phenyl group and the silicon attached to the surface. Unfortunately, the information given by manufacturers of phenyl columns is often too vague to establish the way in which a particular phenyl column was designed. For some of the more popular phenyl phases, the structure is known: Nova-Pak Phenyl™ (Waters Corp., USA) is based on a propylphenyl ligand, Zorbax Phenyl™ packings (Hewlett-Packard, USA) use an ethylphenyl ligand, and µBondapak Phenyl™ (Waters Corp., USA) employs an α-methylphenylethyl ligand. Selectivity differences between phenyl phases and classical C_8 and C_{18} phases have been observed, but it is not clear if they can be attributed to the phenyl group or simply to a difference in surface coverage and chain density compared with C_8 and C_{18} packings. Before the introduction of the packings with an incorporated polar functional group, phenyl packings were the tool used by chromatographers to achieve a different selectivity for the separation.

In the early years of chromatography, cyano phases were often employed when different selectivities were sought. However, experience has driven most users away from reversed-phase applications of cyano columns. The hydrolytic stability is about three orders of magnitude lower for CN phases compared with C_{18} phases. In addition, the mechanical stability of CN columns is significantly inferior to other reversed-phase columns. Owing to the multitude of better options, reversed-phase applications of CN columns have been declining.

Very short chains are commercially available as well, but the nomenclature of the designation may not be clear. A phase designated C_3 by one manufacturer may be identical to a C_1 phase of another manufacturer and the bonded phase is based on a trimethylsilane. These very short chains have a better hydrolytic and mechanical stability than CN phases and can be used with advantage. Short chains, commonly C_4 **(18)**, are used for the reversed-phase separation of proteins. In this case, the pore size of the silica is larger, typically 300 Å or 500 Å.

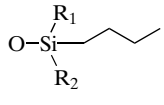

(18)

3.5.5 Bonded Phases for Other Types of Chromatography

While the predominant use of bonded phases is in the world of reversed-phase chromatography, other application areas should not be overlooked. A few phases, commonly abbreviated as cyano **(19)**, amino **(20)**, diol **(21)**, and nitro **(22)** are used in normal phase

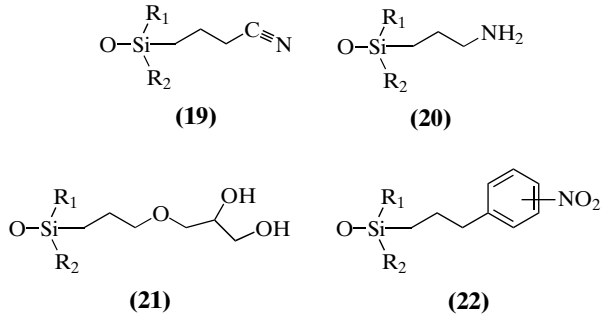

chromatography. The diol and amino phases are mostly utilized in highly specific application areas.

Most polar bonded phases are based on trifunctional silanes. On one hand, this adds to the hydrolytic stability, on the other hand the influence of a large population of silanols on the separation should be considered.

Compared with silica columns, bonded phases are less sensitive to the water content of the mobile phase, which improves the reproducibility of a normal-phase separation. In addition, significantly different selectivities can be achieved. However, manipulation of the selectivity of a separation is rarely a problem in normal-phase chromatography, since a wide range of mobile phase options is available.[2]

Cyano packings are the least polar of the packings used in normal-phase chromatography. While there is little concern about the hydrolytic stability of CN phases in normal-phase chromatography, care should still be exercised with respect to the mechanical stability, especially when solvents of intermediate polarity are used to wash the columns in a clean-up procedure.

Nitro phases, or better nitrophenyl phases, have been used in the normal-phase separation of aromatic hydrocarbons. The separation mechanism is influenced by the formation of $\pi-\pi$ bonds between the analytes and the bonded phase. The consequence is a group separation of the aromatic hydrocarbons based on the number of rings. Nitro phases with more than one nitro group per ring have also been designed. This increases the strength of the $\pi-\pi$ interaction. Occasionally, a reversed-phase application of nitro phase is found as well.

While applications of diol columns in normal phase chromatography are comparatively rare, this stationary phase nevertheless presents an interesting polar alternative to silica, since it is less sensitive than silica to the water content of the mobile phase. However, the primary application area of diol columns is in aqueous size-exclusion chromatography, especially of proteins. High coating levels largely suppress the interaction of proteins with the underlying silica, provided a mobile phase with a salt concentration between 0.1 and 0.5 mol L^{-1} is used. At lower salt concentrations, the influence of residual silanols on retention becomes apparent, at higher salt concentration, hydrophobic interactions distort the separation. Figure 8 shows a gel filtration separation of protein standards using a diol column. Another, less commonly used application area of diol columns is hydrophilic interaction chromatography, using acetonitrile–water mixtures with a high acetonitrile content for the separation of very polar compounds.

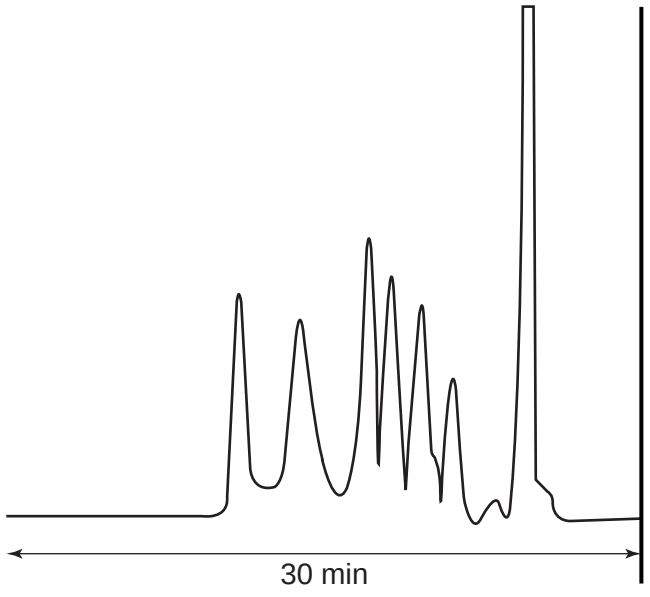

Figure 8 Size-exclusion separation of protein standards on a diol column. Column: Waters Protein-Pak 300SW, 7.8 mm × 300 mm. Mobile phase: 0.1 M K_2HPO_4, pH 7.0; flow rate: 1.0 mL min^{-1}; detection: 280 nm (chromatogram courtesy of Waters Corporation).

In normal-phase chromatography, amino columns are another interesting alternative to underivatized silica. A special application is related to the main application of nitrophenyl packings: the class separation of aromatic hydrocarbons. The separation patterns are similar for both columns: in both cases, the analytes are separated into groups of common ring size. However, the primary application area for amino columns is in hydrophilic interaction chromatography for the separation of carbohydrates (Figure 9). The separation mechanism[61] for the sugar separations is a partitioning mechanism between a mobile phase with a low water content and the stationary phase with an enriched water content. The main advantage of the amino column over other polar packings (such as diol) is the fact that amino columns do not separate the sugar anomers, while other polar phases do. The reason for this is the fact that the local basic environment promotes the interconversion of the anomers during the chromatographic run. Therefore, all sugars elute as single, albeit somewhat broad peaks. It should be noted though that silica-based amino phases

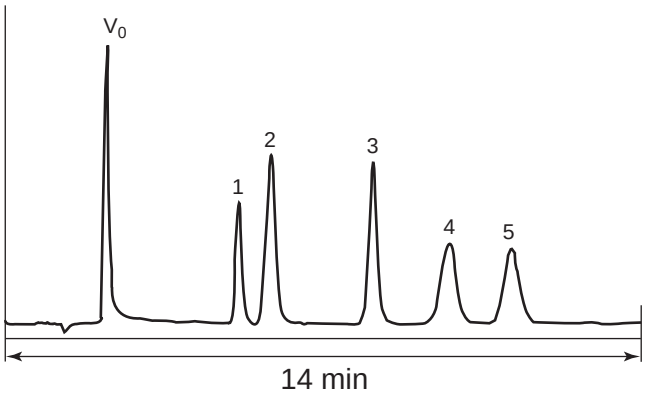

Figure 9 Separation of sugars using an aminopropyl bonded phase. Column: Waters High-Performance Carbohydrate Column. Mobile phase: acetonitrile/water 75/25 v/v (chromatogram courtesy of Dorothy Phillips, Waters Corporation).

are not completely stable in an aqueous environment: the high concentration of amino groups in the pores of the packing creates a high pH value, which leads to a partial dissolution of the silica and the bonded phase. This occurs until a sufficient number of acidic silanol groups is available to reduce the local pH. Therefore, amino columns for the separation of sugars are commonly pretreated to reach a stable arrangement.

Another group of stationary phases based on silica are the ion exchangers **(23)** and **(24)**. Both strong cation and strong anion exchangers are commercially available. The anion exchanger can be prepared by direct synthesis or by a secondary reaction of a dimethylaminopropyl silane bonded phase with a methylating reagent. Two types of cation exchangers are possible: one is based on a purely aliphatic chain, the other is based on an aromatic sulfonic acid. Similarly, quarternary amines based on an aromatic core group in the bonded phase are possible as well.

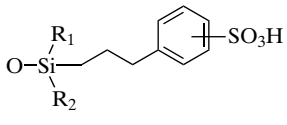

(23)

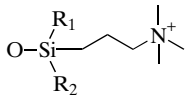

(24)

The application literature is not as extensive as the application literature using classical ion exchangers based on an organic backbone. This is mostly due to the limitations in the pH range imposed by the silica backbone. In addition, ion-pair chromatography[62] based on well-defined reversed-phase packings has become a more versatile substitute for ion-exchange separations of small molecules.

However, ion-exchange chromatography remains one of the primary tools in the separation of biomacromolecules. The special phases for this application fall into the category of coated phases and are discussed in the next section.

4 COATED PHASES

In this section, the properties of silica-based coated phases are discussed. The coating consists of a physically adsorbed or chemically bonded and/or crosslinked polymer layer that modifies the adsorption properties of the substrate by superimposing its own properties. Properly executed, it should be possible to combine the advantageous properties of the coating with the mechanical strength of the silica. Note that many of the stationary phases discussed in this section can be prepared on inorganic substrates other than silica. The advantage of other substrates lies in the improved pH stability compared with silica: silica dissolves with an appreciable rate at pH values exceeding 9 at room temperature, therefore substrates like alumina, titania, or zirconia have marked advantages. Nevertheless, most work has been done using silica as the base material.

In general, very few phases with a polymeric surface layer have been commercially successful. Prevailing applications include packings designed for the chromatography of macromolecules like proteins[63–65] and packings for special applications, e.g. ion chromatography.[66–68] The reason for this limited success is the inferior mass-transfer properties of coated phases for small molecules, which is due to the fact that the polymeric layer can be at least partially penetrated by small molecules and that diffusion in the polymeric layer is slow. This feature is shared with fully organic packings. The problem either becomes irrelevant, if no packing with equivalent selectivity and good mass transfer is available, or disappears, if the molecular weight of the analytes is sufficiently large, as is the case for proteins.

Many different polymers have been immobilized on silica. Petro and Berek[68] list some 80 different types of polymer–silica combination. The polymeric layer may be bonded to the surface, or cross-linking may keep it in place without attachment to the surface. If the polymer is not generated in situ, an important consideration in the preparation is the access of the polymer to the pores of the packing. Therefore, four different preparation procedures for polymer-coated packings can be differentiated:[70]

1. polymerization or polycondensation of monomers without bonding to the surface
2. polymerization or polycondensation of monomers with bonding to the surface
3. immobilization of prepolymers without attachment to the surface

4. immobilization of prepolymers with attachment to the surface.

In the following, these different approaches are described briefly.

4.1 Preparation Techniques for Coated Phases

4.1.1 Polymerization or Polycondensation of Monomers Without Bonding to the Surface

In this approach, the monomers are first coated into the pores or onto the surface of the silica. Then, the polymerization is initiated. Since it is not difficult to generate a uniform coating of the surface with the monomer(s), such a process should then result in a uniform polymeric coating of the packing as well. However, results obtained with this technique[71,72] were rather disappointing, when small analytes were investigated: for thin layers, the silica matrix was poorly shielded by the polymeric layer and influenced the retention of basic analytes, while thick layers resulted in an inferior mass transfer and broad peaks for small analytes. The results are shown in Table 2. The influence of silanols on the retention of basic analytes can be measured by the relative retention between propranolol and acenaphthene; as the coating level increases, the relative retention becomes smaller, i.e. the relative influence of silanols on the retention drops. However, the mass-transfer term, measured with toluene as sample, increases as the coating level increases. On a C_{18} column based on the same silica, the mass-transfer term for toluene was only 3 ms. This clearly exemplifies the general problem of polymeric coatings on silica or other inorganics: polymeric coatings result in inferior mass transfer and thus inferior plate counts for small molecules.

4.1.2 Polymerization or Polycondensation of Monomers with Bonding to the Surface

When at least one of the monomers used in the polymerization is attached to the surface, better adhesion of the resulting polymer to the surface is expected and a more conformal coating should be achievable. Packings prepared using this approach have been reported by Engelhardt and co-workers[72,73] and are commercially available under the trade name Polyencap™ (Bischoff GmbH, Germany). The packings are prepared by first bonding vinylsilane to the silica surface, then coating the resulting surface with an acrylate monomer, which is then polymerized along the surface. Generally, good mass-transfer properties have been reported for these packings, but owing to the low coreactivity of vinylsilane with the monomers, a clear picture of the nature of the resulting surface is not available. In this procedure, only 50–80% of the monomer was incorporated in the packing and chemically anchored to the surface, demonstrating the low coreactivity of the vinylsilane with the monomers.

Table 2 Properties of 10 μm Spherisorb® silica with divinylbenzene polymer layers

Nominal layer thickness	Coating %	α-Propranolol/ acenaphthene	Mass transfer term (ms)
1/4	2.56	2.3	17.8
1/2	5.2	2.2	15.1
1	10.1	1.9	21.8
1.5	14.6	1.6	18.4
2	18.1	1.3	26.9

The method used for the preparation of a range of commercial silica derivatives used for the separation of biopolymers (PolyLC, USA) is related to this approach. In this procedure,[63–65] silica of the correct pore size is first derivatized with aminopropyltriethoxysilane in toluene. The resulting amino-bonded phase is then treated with poly(succinimide) at room temperature in tetrahydrofuran (THF). This phase can then be used to prepare a range of derivatives, e.g. a poly(aspartic acid) derivative by reaction with β-alanine as a weak cation exchanger, a poly(2-aminoethyl aspartamide) derivative from ethylenediamine as a weak anion exchanger, poly(alkyl aspartamide) derivatives for hydrophobic interaction chromatography,[64] a poly(2-sulfoethyl aspartamide) derived from taurine as a strong cation exchanger,[65] and a poly(2-hydroxyethyl aspartamide) derivative for hydrophilic interaction chromatography.[74] These packings have been used extensively in the separation of proteins and peptides. While data for the mass-transfer behavior of these packings for small molecules are not available, the quality of the gradient separations of proteins and peptides is impressive.

Another successful group of packings in this category is the group of ''tentacle'' ion exchangers (Merck, Germany) which are also employed in the chromatography of biopolymers.[75] The difference between classical ion exchangers and the tentacle type is the location of the ion-exchange groups on the packing. For classical ion exchangers, the ion-exchange groups are located on the internal surface of the packing. For tentacle ion exchangers, they are located on linear polymer chains grafted to the support surface. The free access to the ion-exchange groups improves the interaction with biopolymers, while avoiding distortions of the three-dimensional conformation of the analytes. Consequently, significantly different selectivities have been observed compared with classical ion exchangers and improved chromatographic performance has been reported. However, as for any other packing for the

chromatography of biopolymers, a quantitative measurement of the mass-transfer properties[2] of the tentacle packings is not available, which makes an objective judgement of the quality of the packing difficult.

4.1.3 Immobilization of Prepolymers Without Attachment to the Surface

In the third approach, a low-molecular-weight polymer is coated onto the surface and then immobilized through chain extension or crosslinking. Examples of this approach have been pursued by Schomburg et al.[76] Of course, whenever procedures are used that can generate a nonbonded coating on silica, other inorganics can be substituted with advantage. Poly(butadiene)-coated alumina exhibits an improved pH stability relative to silica-based sorbents. However, while no concrete data on column efficiency or mass transfer are available in the literature, a decrease in column performance relative to classical phases has often been observed. However, this is not an impedance for specific applications. Schomburg's group also coated silica with a copolymer of butadiene and maleic acid to create a cation exchanger.[77] The most popular application of this packing is the simultaneous separation of alkali and alkaline earth ions in a single isocratic chromatographic run (Figure 10), which is not achievable with any other packing.

The coating of silica with polybutadiene[78] or with poly(methyloctylsiloxane)[79] has been investigated in detail. In the latter study, it was demonstrated that the surface area decreases linearly with the specific loading of polymer onto the silica and that this behavior is inconsistent with a uniform film coating the surface of the packing. In the former study, a broad arsenal of tools ranging from nitrogen sorption to solid-state NMR was employed to demonstrate the same feature: ''polybutadiene coatings on silica do not result in a homogeneous polymer film but in an inhomogeneous loading...'', and ''an increasing polymer loading leads to increasingly filled pores instead of a thicker polymer film''.[78] Until now, the problem of generating a uniform polymeric film on the surface of a sorbent without bonding to the surface remains unsolved.

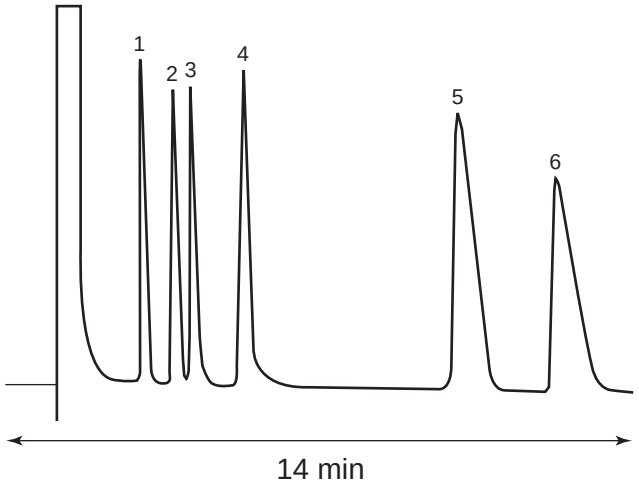

Figure 10 Separation of alkali and alkaline earth ions on poly(butadiene-maleic acid) coated silica. Column: IC-Pak C M/D, 3.9 mm × 150 mm, Waters Corporation. Eluent: 0.1 mM ethylenediaminetetraacetic acid, 3.0 mM HNO_3; flow rate: 1 mL min^{-1}; sample: 1. lithium, 2. sodium, 3. ammonium, 4. potassium, 5. magnesium, 6. calcium; detection mode: conductivity (chromatogram courtesy of Waters Corporation).

For the separation of proteins, polyethyleneimine was coated onto silica and crosslinked.[80] This durable anion exchanger has been successfully employed in the separation of many biopolymers. It could be argued that the polyethyleneimine is attached to the surface via an ion-exchange interaction and that this packing belongs to the category discussed in the next paragraph.

Both techniques have been used to coat small nonporous silica spheres as well. The principle behind this approach is to take advantage of the more rapid mass transfer of nonporous particles for the fast separation of macromolecules. For example, polybutadiene was coated onto nonporous silica spheres of 1.7 μm diameter.[81] The small particle diameter together with the fact that the packing was nonporous allowed for rapid gradient separations of peptides. In an earlier publication,[82] the coating of nonporous silica spheres with polyethyleneimine was investigated. Very rapid separations of proteins and nucleotides were obtained.

4.1.4 Immobilization of Prepolymers with Attachment to the Surface

An example of the immobilization of a prepolymer with attachment to the surface of silica has been published by Petro et al.[83] The purpose for the design of such a phase was the size-exclusion chromatography of proteins. The silica is first reacted with γ-glycidoxypropyltriethoxysilane in dry toluene. The resulting product is then further derivatized with aqueous solutions of dextran or ethylenediamine derivatives of dextran at various pH values in water. Unfortunately, a significant amount of interaction with silanols remained on the packing, requiring still higher salt concentrations (0.25 M) to elute proteins in size-exclusion mode.

In another example,[84] polyvinylpyrrolidone was coated onto large pore silicas. First, a copolymer of vinylmethyldiethoxysilane and vinylpyrrolidone was created. Then this polymer was attached to the silicas via the active silane groups. Finally, the material was endcapped using a standard process. Both size-exclusion chromatography and hydrophobic interaction chromatography of

proteins were possible with this packing. However, the packing loses attractiveness owing to the necessity of following through with the hydrolytically weak endcapping step.

4.2 Properties of Coated Phases

Most investigations of silica-based coated phases have been focused on the preparation of materials for the separation of biomolecules. This is largely due to the fact that most phases in this category suffer from restricted mass transfer for small molecules and therefore inferior column performance. The most promising category for small molecules appears to be the second category, where the preparation involves a polymeric surface coating from monomers with a known attachment to the surface.

For proteins and other large molecules, the same principle has also yielded the most promising results. The long-standing success of the phases described by Alpert[63–65,74] as well as the newer success of the tentacle phases support this approach. Nevertheless, one needs to differentiate between reality and marketing claims. In a fundamental study, Ratnayake and Regnier[85] studied the protein binding capacity of ion exchangers prepared by the second principle. They concluded that the mechanism of adsorption is similar between tentacle-type ion exchangers and surface-bound ion exchangers. Among other things, the *Z*-number, which is a characteristic value for the number of the interaction sites between ion exchanger and analyte, was similar for both types of sorbents. The increased capacity of tentacle-type ion exchangers compared with surface-bound ion exchangers is therefore simply due to an increased number of available exchange sites owing to the larger surface area of the tentacle-type packings.

The other issue for the separation of macromolecules is the plate count achievable with different packings. Unfortunately, the techniques that allow a determination of the plate height under gradient conditions[3] have not yet been applied to these packings and therefore an objective comparison is not yet possible. It might be expected that the situation is similar to the situation for small molecules: nonporous packings exhibit a higher plate-count performance, but a lower capacity than porous packings. However, the differences are expected to be much smaller owing to the necessity of taking into account larger pore sizes for macromolecules compared with smaller molecules. With respect to the performance of coated phases compared with classical bonded phases, a lesser sensitivity of the HETP for macromolecules on the details of the surface coating can be expected. The success of the Alpert phases and the tentacle phases supports this expectation.

5 SUMMARY

The classical derivatization of silica has become the main tool for the preparation of stationary phases for classical HPLC, i.e. the separation of small molecules. The influence of the properties of the silica and the stationary phase are well understood. Nevertheless, progress continues to be made. The success in the 1990s of high-purity silica as the carrier of bonded phases is the dominant example. On a smaller scale, bonded phases for reversed-phase chromatography with incorporated polar functionality are now more widely used owing to their improved peak shape compared with classical packings. The newest development for the analysis of small molecules are the continuous rod structures, whose success will be determined in the future.

Until now, silica-based stationary phases have played a more limited role in the HPLC separations of biopolymers. Some innovative approaches have been applied and progress has been made, but silica-based phases are only one of many tools used in the separation of biomolecules. The more promising approaches use a polymeric coating of the silica to combine the benefits of the strength of silica with the biocompatibility of the polymeric coating.

In addition to the information presented in this article, the literature contains supplemental or related specific information. Dorsey et al.[86] review fundamental developments in the LC literature from October 1995 to October 1997. Leonard[87] gives a general review of new packing materials for protein chromatography, including polymer-based and silica-based packings. Interested readers are referred to these review articles.

ABBREVIATIONS AND ACRONYMS

BaP	Benzo[*a*]pyrene
CP/MAS/NMR	Cross-polarization/Magic Angle Spinning/Nuclear Magnetic Resonance
HETP	Height Equivalent to a Theoretical Plate
HPLC	High-performance Liquid Chromatography
IAM	Immobilized Artificial Membrane
LC	Liquid Chromatography
NIST	National Institute of Standards and Technology
PhPh	Phenanthro[3,4-*c*]phenanthrene
RPLC	Reversed-phase Liquid Chromatography

RSD	Relative Standard Deviation
TBN	1,2-3,4-5,6-7,8-Tetrabenzonaphthalene
THF	Tetrahydrofuran

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis

Environment: Water and Waste **(Volume 3)**
Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Food **(Volume 5)**
Liquid Chromatography in Food Analysis

Peptides and Proteins **(Volume 7)**
Chromatography of Membrane Proteins and Lipoproteins • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Hydrophilic-interaction Chromatography in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pesticides **(Volume 7)**
High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Eluent Additives and the Optimization of High-performance Liquid Chromatography Procedures • Gas and Liquid Chromatography, Column Selection for, in Drug Analysis

Polymers and Rubbers **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization • Size-exclusion Chromatography of Polymers

Process Instrumental Methods **(Volume 9)**
Chromatography in Process Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Affinity Chromatography • Biopolymer Chromatography • Chiral Separations by High-performance Liquid Chromatography • Column Theory and Resolution in Liquid Chromatography • Gradient Elution Chromatography • Ion Chromatography • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography • Normal-phase Liquid Chromatography • Reversed Phase Liquid Chromatography

Mass Spectrometry **(Volume 13)**
Liquid Chromatography/Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
High-performance Liquid Chromatography Nuclear Magnetic Resonance

REFERENCES

1. K.K. Unger, *Porous Silica*, Elsevier Scientific, Amsterdam , 1979.
2. U.D. Neue, *HPLC Columns, Theory, Technology, and Practice*, Wiley-VCH, New York, 1997.
3. R. Endele, I. Halász, K. Unger, 'Influence of the Particle Size (5–35 μm) of Spherical Silica on Column Efficiencies in HPLC', *J. Chromatogr.*, **99**, 377 (1974).
4. J.J van Deemter, F.J. Zuiderweg, A. Klinkenberg, 'Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Nonideality in Chromatography', *Chem. Eng. Sci.*, **5**, 271–289 (1956).
5. D.C. Lommen, L.R. Snyder, 'Fast HPLC Separations with Small Nonporous Particles', *LC-GC*, **11**, 222–232 (1993).
6. G. Jilge, K.K. Unger, U. Esser, H.J. Schäfer, G. Rathgeber, W. Müller, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by HPLC', *J. Chromatogr.*, **476**, 37–48 (1989).
7. J.J. Kirkland, 'Superficially Porous Silica Microspheres for the Fast HPLC of Macromolecules', *Anal. Chem.*, **64**, 1239–1245 (1992).
8. K. Kaghatgi, Cs. Horváth, 'Rapid Analysis of Proteins and Peptides by Reversed-phase Chromatography', *J. Chromatogr.*, **398**, 335–339 (1987).

9. W.-C. Lee, 'Protein Separation using Non-porous Sorbent', *J. Chromatogr., B*, **699**, 29–45 (1997).
10. H. Minakuchi, K. Nakanishi, N. Soga, N. Ishizuka, N. Tanaka, 'Effect of Skeleton Size on the Performance of Octadecylsilylated Continuous Porous Silica Columns in RPLC', *J. Chromatogr., A*, **762**, 135–146 (1997).
11. H. Minakuchi, K. Nakanishi, N. Soga, N. Ishizuka, N. Tanaka, 'Effect of Domain Size on the Performance of Octadecylsilylated Continuous Porous Silica Columns in RPLC', *J. Chromatogr., A*, **797**, 121–131 (1998).
12. N. Ishizuka, H. Minakuchi, K. Nakanishi, N. Soga, N. Tanaka, 'Designing Monolithic Double-pore Silica for HPLC', *J. Chromatogr., A*, **797**, 133–137 (1998).
13. G.B. Cox, 'The Influence of Silica Structure on Reversed-phase Retention', *J. Chromatogr., A*, **656**, 353–367 (1993).
14. G.E. Berendsen, L. de Galan, 'A Geometrical Model for Chemically Bonded TMS and PDS Phases', *J. Liq. Chromatogr.*, **1**, 403–426 (1978).
15. Y. Sudo, T. Takahata, US Patent 5134110, 1992.
16. P. Orth, 'Alkoxysilylierte Kieselgele für die HPLC', Dissertation, Universität des Saarlandes, 1979.
17. H. Schmidt, 'Herstellung, Eigenschaften und Einsatzmöglichkeiten von Umkehrphasen in der Hochdruckflüssigkeitschromatographie', Dissertation, Universität des Saarlandes, 1978.
18. R.K. Gilpin, M.E. Gangoda, 'Chromatographic and NMR Investigations of the Solvation and Mobility of Bonded Alkyl Phases', *J. Chromatogr. Sci.*, **21**, 352–361 (1983).
19. A. Pryde, 'Chemically Bonded Stationary Phases – Preparation and Application to High Speed Liquid Chromatography', *J. Chromatogr. Sci.*, **12**, 486–498 (1974).
20. R.E. Majors, M.J. Hopper, 'Studies of Siloxane Phases Bonded to Silica Gel for Use in HPLC', *J. Chromatogr. Sci.*, **12**, 767–778 (1974).
21. K. Karch, I. Sebestian, I. Halász, 'Preparation and Properties of Reversed Phases', *J. Chromatogr.*, **122**, 3–16 (1976).
22. Y. Sudo, 'End-capping of Octadecylsilylated Silica Gels by High-temperature Silylation', *J. Chromatogr., A*, **737**, 139–147 (1996).
23. Y. Sudo, 'Optimization of End-capping of Octadecylsilylated Silica Gels by High-temperature Silylation', *J. Chromatogr., A*, **757**, 21–28 (1997).
24. A. Nahum, Cs. Horváth, 'Surface Silanols in Silica-bonded Hydrocarbonaceous Stationary Phases, I. Dual Retention Mechanism in Reversed-phase Chromatography', *J. Chromatogr.*, **203**, 53–63 (1981).
25. K.E. Bij, Cs. Horváth, W.R. Melander, A. Nahum, 'Surface Silanols in Silica-bonded Hydrocarbonaceous Stationary Phases, II. Irregular Retention Behavior and Effect of Silanol Masking', *J. Chromatogr.*, **203**, 65–84 (1981).
26. M. Kele, G. Guiochon, 'Repeatability and Reproducibility of Retention Data and Band Profiles on Reversed-phase Liquid Chromatography Columns: I. Experimental Protocol', *J. Chromatogr., A*, **830**, 41–54 (1999).
27. M. Kele, G. Guiochon, 'Repeatability and Reproducibility of Retention Data and Band Profiles on Reversed-phase Chromatography Columns: II. Results Obtained with Symmetry™ C_{18} Columns.', *J. Chromatogr., A*, **830**, 55–79 (1999).
28. J.J. Kirkland, J.L. Glajch, R.D. Farlee, 'Synthesis and Characterization of Highly Stable Bonded Phases for HPLC Column Packings', *Anal. Chem.*, **61**, 2–11 (1989).
29. J.J. Kirkland, C.H. Dilts, J.E. Henderson, 'Technologies for an Improved C_{18} Stationary Phase in Reversed-phase HPLC Separations', *LC-GC*, **11**, 290–297 (1993).
30. H. Engelhardt, P. Orth, 'Alkoxy Silanes for the Preparation of Silica-based Stationary Phases with Bonded Polar Functional Groups', *J. Liq. Chromatogr.*, **10**, 1999–2022 (1987).
31. M.J. Wirth, R.W.P. Fairbank, 'Preparation of Mixed C_{18}/C_1 Horizontally Polymerized Chromatographic Phases', *J. Liq. Chromatogr.*, **19**, 2799–2810 (1996).
32. M.J. Wirth, R.W.P. Fairbank, H.O. Fatunmbi, 'Mixed Self-assembled Monolayers in Chemical Separations', *Science*, **275**, 44–47 (1997).
33. T.H. Walter, B.A. Alden, private communication.
34. L.C. Sander, S.A. Wise, 'Shape Selectivity in Reversed-phase Liquid Chromatography for the Separation of Planar and Non-planar Solutes', *J. Chromatogr., A*, **656**, 335–351 (1993).
35. L.C. Sander, S.A. Wise, 'Influence of Stationary Phase Chemistry on Shape Recognition in Liquid Chromatography', *Anal. Chem.*, **67**, 3284–3292 (1995).
36. L.C. Sander, K.E. Sharpless, N.E. Craft, S.A. Wise, 'Development of Engineered Stationary Phases for the Separation of Carotenoid Isomers', *Anal. Chem.*, **66**, 1667–1674 (1994).
37. C. Emenhiser, L.C. Sander, S.J. Schwartz, 'Capability of a Polymeric C_{30} Stationary Phase to Resolve *cis-trans* Carotenoid Isomers in Reversed-phase Liquid Chromatography', *J. Chromatogr., A*, **707**, 205–216 (1995).
38. S. Ong, H. Liu, X. Qui, G. Bhat, C. Pidgeon, 'Membrane Partition Coefficients Chromatographically Measured Using Immobilized Artificial Membrane Surfaces', *Anal. Chem.*, **67**, 755–762 (1995).
39. S. Ong, H. Liu, C. Pidgeon, 'Immobilized-artificial-membrane Chromatography: Measurements of Membrane Partition Coefficient and Predicting Drug Membrane Permeability', *J. Chromatogr., A*, **728**, 113–128 (1996).
40. C. Pidgeon, US Patent 4 927 879, 1990; US Patent 4 931 498, 1990.
41. C. Pidgeon, J.M. Stevens, S. Otto, C. Jefcoate, C. Marcus, 'Immobilized Artificial Membrane Chromatography: Rapid Purification of Functional Membrane Proteins', *Anal. Biochem.*, **194**, 163–173 (1991).
42. P. Artursson, J. Karlsson, 'Correlation between Oral Drug Absorption in Humans and Apparent Drug Permeability Coefficients in Human Intestinal Epithelial (Caco-2) Cells', *J. Biochem., Biophys. Res. Commun.*, **175**, 880–885 (1991).

43. L.S. Schanker, D.J. Tocco, B.B. Brodie, C.A.M. Hogben, 'Adsorption of Drugs from the Rat Small Intestine', *J. Pharmacol. Exp. Ther.*, **123**, 81–88 (1958).
44. G.W. Caldwell, J.A. Masucci, M. Evangelisto, R. White, 'Evaluation of the Immobilized Artificial Membrane Phosphatidylcholine', *J. Chromatogr., A*, **800**, 161–169 (1998).
45. I.H. Hagestam, T.C. Pinkerton, 'Internal Surface Reversed-phase Silica Supports for Liquid Chromatography', *Anal. Chem.*, **57**, 1757–1763 (1985).
46. K. Kimata, R. Tsuboi, K. Hosoya, N. Tanaka, T. Araki, 'Method for the Preparation of Internal-surface Reversed-phase Packing Materials Starting from Alkylsilylated Silica Gels', *J. Chromatogr.*, **515**, 73–84 (1990).
47. H. Ihara, S. Nagaoka, H. Tanaka, S. Sakaki, C. Hirayama, 'Lipid Membrane Analogue-immobilized Silica Gels for Separation with Molecular Recognition', *J. Liq. Chromatogr. Rel. Technol.*, **19**, 2967–2984 (1996).
48. H. Hemmetsberger, W. Maasfeld, H. Ricken, 'The Effect of Chain Length of Bonded Organic Phases in Reversed-phase HPLC', *Chromatographia*, **9**, 303 (1976).
49. H. Hemmetsberger, P. Behrensmeyer, J. Henning, H. Ricken, 'Reversed-phase HPLC: Effect of the Structure of the Chemically Bonded Hydrocarbon Ligand on Retention and Selectivity', *Chromatographia*, **12**, 71–76 (1979).
50. J.E. O'Gara, B.A. Alden, T.H. Walter, J.S. Petersen, C.L. Niederländer, U.D. Neue, 'Simple Preparation of a C_8 HPLC Stationary Phase with an Internal Polar Functional Group', *Anal. Chem.*, **67**, 3809–3813 (1995).
51. T.L. Ascah, B. Feibush, 'Novel, Highly Deactivated Reversed-phase for Basic Compounds', *J. Chromatogr.*, **506**, 357–369 (1990).
52. T.L. Ascah, K.M.R. Kallury, C.A. Szafranski, S.D. Corman, F. Liu, 'Characterization and HPLC Evaluation of a New Amide-functionalized Reversed-phase Column', *J. Liq. Chrom. Rel. Technol.*, **19**(17&18), 3049–3073 (1996).
53. K.J. Duff, 'Improved HPLC Separations Using Intrinsically Base-deactivated Columns', *American Laboratory*, February, 32AA–32JJ (1998).
54. T.H. Walter, P. Iraneta, M. Capparella, 'Observations on the Wetting of Reversed-phase HPLC Packings', Presentation at HPLC'97, Birmingham, UK, 1997.
55. H. Engelhardt, M. Jungheim, 'Comparison and Characterization of Reversed Phases', *Chromatographia*, **29**, 59–68 (1990).
56. B. Dreyer, 'Neue Methoden zur Klassifizierung und Charakterisierung von Umkehrphasen für die HPLC', Dissertation, Universität des Saarlandes, 1984.
57. H. Engelhardt, M. Arangio, T. Lobert, 'A Chromatographic Test Procedure for Reversed-phase HPLC Column Evaluation', *LC-GC Internat.*, December, 803–812 (1997).
58. E. Serowik, U.D. Neue, unpublished results, 1985.
59. U.D. Neue, D.J. Phillips, T.H. Walter, M. Capparella, B.A. Alden, R.P. Fisk, 'The Relationship between the Quality of a Reversed-phase Column and the Quality of a Pharmaceutical Analysis', *LC-GC*, **12**(6), 468–479 (1994).
60. K. Kimata, K. Iwaguchi, S. Onishi, K. Jinno, R. Eksteen, K. Hosoya, M. Araki, N. Tanaka, 'Chromatographic Characterization of Silica C_{18} Packing Materials. Correlation between a Preparation Method and Retention Behavior of Stationary Phase', *J. Chromatogr. Sci.*, **27**, 721–728 (1989).
61. U.D. Neue, C.H. Phoebe, 'HPLC of Derivatized and Non-Derivatized Oligosaccharides: A Review', in *A Laboratory Guide to Glycoconjugate Analysis, BioMethods*, ed. P. Jackson, Birkhäuser Verlag, Basel, **9**, 1–22, 1997.
62. J.H. Knox, R.A. Hartwick, 'Mechanism of Ion-pair Liquid Chromatography of Amines, Neutrals, Zwitterions and Acids using Anionic Hetaerons', *J. Chromatogr.*, **204**, 3–21 (1981).
63. A.J. Alpert, 'Cation-exchange HPLC of Proteins on Poly-(aspartic acid)-silica', *J. Chromatogr.*, **266**, 23–37 (1983).
64. A.J. Alpert, 'High-performance Hydrophobic-interaction Chromatography of Proteins on a Series of Poly(alkyl aspartamide)-silica', *J. Chromatogr.*, **359**, 85–97 (1986).
65. A.J. Alpert, P.C. Andrews, 'Cation-exchange Chromatography of Peptides on Poly(2-sulfoethyl aspartamide)-silica', *J. Chromatogr.*, **443**, 85–96 (1988).
66. G. Schomburg, P. Kolla, M.W. Läubli, 'Ion Chromatography of Alkali and Alkaline Earth Metal Ions using a Silica-based Polymer-coated Stationary Phase', *Int. Laboratory*, **4**, 40–48 (1989).
67. M.W. Läubli, B. Kampus, 'Cation Analysis on a New Poly(butadiene-maleic acid)-based Column', *J. Chromatogr., A*, **706**, 99–102 (1995).
68. M.W. Läubli, B. Kampus, 'Selectivity Enhancement on a Poly(butadiene-maleic acid)-coated Cation Phase Induced by Ethylene Oxide-based Complexing Agents', *J. Chromatogr., A*, **706**, 103–107 (1995).
69. M. Petro, D. Berek, 'Polymers Immobilized on Silica Gel Stationary Phases for Liquid Chromatography', *Chromatographia*, **37**(9/10), 549–561 (1993).
70. M. Hanson, A. Kurganov, K.K. Unger, V.A. Davankov, 'Polymer-coated Reversed-phase Packings in HPLC', *J. Chromatogr., A*, **656**, 369–380 (1993).
71. U.D. Neue, M. Mauß, unpublished results, 1987.
72. M. Mauß, 'Polymerbelegte Stationäre Phasen für die Ionenaustausch-Chromatographie', Dissertation, Universität des Saarlandes, 1988.
73. H. Engelhardt, H. Löw, W. Eberhardt, M. Mauß, 'Polymer Encapsulated Stationary Phases: Advantages, Properties and Selectivities', *Chromatographia*, **27**(11/12), 535–543 (1989).
74. A.J. Alpert, 'Hydrophilic-interaction Chromatography for the Separation of Peptides, Nucleic Acids and other Polar Compounds', *J. Chromatogr.*, **499**, 177–196 (1990).
75. W. Müller, 'New Ion Exchangers for the Chromatography of Biopolymers', *J. Chromatogr.*, **510**, 133–140 (1990).

76. U. Bien-Vogelsang, A. Degge, H. Figge, J. Köhler, G. Schomburg, 'Syntheses of Stationary Phases for Reversed-phase LC Using Silanization and Polymer Coating', *Chromatographia*, **19**, 170–179 (1984).
77. P. Kolla, J. Köhler, G. Schomburg, 'Polymer-coated Cation-exchange Stationary Phases on the Basis of Silica', *Chromatographia*, **23**, 465–472 (1987).
78. M. Hanson, B. Eray, K. Unger, A.V. Neimark, J. Schmidt, K. Albert, E. Bayer, 'A Model for Polybutadiene Coatings on Porous Silica', *Chromatographia*, **35**(7/8), 403–409 (1993).
79. K.E. Collins, M.L.M.M. Granja, R.G. Pereira Filho, T.A. Anazawa, I.C.S.F. Jardim, 'The Spatial Distribution of Poly(methyloctylsiloxane) Deposited by Solvent Evaporation in the Pores of HPLC Silica Particles', *Chromatographia*, **45**, 99–103 (1997).
80. A.J. Alpert, F.E. Regnier, 'Preparation of a Porous Microparticulate Anion-exchange Chromatography Support for Proteins', *J. Chromatogr.*, **185**, 375–392 (1979).
81. M. Hanson, K.K. Unger, G. Schomburg, 'Non-porous Polybutadiene-coated Silicas as Stationary Phases in Reversed-phase Chromatography', *J. Chromatogr.*, **517**, 269–284 (1990).
82. G. Jilge, K.K. Unger, U. Esser, H.J. Schäfer, G. Rathgeber, W. Müller, 'Evaluation of Advanced Silica Packings for the Separation of Biopolymers by HPLC', *J. Chromatogr.*, **476**, 37–48 (1989).
83. M. Petro, P. Gemeiner, D. Berek, 'Dextran-grafted Silica Gel for High-performance Size-exclusion Chromatography of Proteins', *J. Chromatogr., A*, **665**, 37–45 (1994).
84. A. Kurganov, Yu. Puchkova, V. Davankov, F. Eisenbeiss, 'Polyvinylpyrrolidone-coated Silica Packings for Chromatography of Proteins and Peptides', *J. Chromatogr., A*, **663**, 163–174 (1994).
85. C.K. Ratnayake, F.E. Regnier, 'Study of Protein Binding to a Silica Support with a Polymeric Cation-exchange Coating', *J. Chromatogr., A*, **743**, 15–23 (1996).
86. J.G. Dorsey, William T. Cooper, B.A. Siles, J.P. Foley, H.G. Barth, 'Liquid Chromatography: Theory and Methodology', *Anal. Chem.*, **70**, 591R–644R (1998).
87. M. Leonard, 'New Packing Materials for Protein Chromatography', *J. Chromatogr., B*, **699**, 3–27 (1997).

Supercritical Fluid Chromatography

Huimin Yuan
Merck Research Laboratories, Merck & Co., Inc., Rahway, USA

Susan V. Olesik
The Ohio State University, Columbus, USA

1 Introduction **1**
2 Theory and Principles of Supercritical Fluid Chromatography **2**
2.1 Properties of Supercritical Fluids 2
2.2 Why Supercritical Fluid Chromatography? 3
2.3 Closely Related Chromatographic Techniques and Unified Chromatography 4
2.4 Capillary Supercritical Fluid Chromatography Versus Packed Column Supercritical Fluid Chromatography 4
3 Instrumentation **5**
3.1 Pump and Mixer 5
3.2 Sample Injection 6
3.3 Column and Oven 6
3.4 Pressure Control 6
3.5 Detection 6
4 Stationary Phases for Supercritical Fluid Chromatography **7**
5 Mobile Phases for Supercritical Fluid Chromatography **7**
5.1 Pure Fluids 7
5.2 Binary or Ternary Mobile Phases 8
6 Perspectives and Future Development **9**
Abbreviations and Acronyms **9**
Related Articles **9**
References **9**

Supercritical fluid chromatography (SFC) employs fluids that are raised above their critical pressure (P_c) and critical temperature (T_c) as mobile phases. Carbon dioxide, CO_2, is the supercritical fluid most commonly used as a mobile phase for SFC. It is capable of solvating mostly nonpolar analytes. Cosolvents, such as methanol and acetonitrile, are added to increase the polarity of the analytes that can be analyzed by SFC. SFC is most commonly practiced using a normal-phase retention mechanism. In other words, the stationary phase is typically more polar than the mobile phase and the retention order is in increasing order of polarity. The advantages of SFC over liquid chromatography (LC) include higher efficiency separations and faster speed of analysis. The advantage that SFC has over gas chromatography (GC) is that SFC can efficiently separate thermally labile compounds. The disadvantage of SFC is that the polarity of the mobile phase is limited. Highly polar solutes are not soluble in supercritical fluids, even when modifiers are added. SFC is commonly used to separate petroleum distillates and by-products, moderately polar pesticides and herbicides, and more recently for the separation of chiral compounds, such as pharmaceutical drugs.

1 INTRODUCTION

SFC employs fluids at temperatures and pressures above their critical temperatures (T_c) and critical pressures (P_c). When used as chromatographic mobile phases, supercritical fluids are able to solvate moderately polar analytes that are difficult to volatize in GC, but are still compatible with GC detectors. Whenever analytes could be separated by either LC or by SFC, the latter provides higher efficiency, shorter analysis time, lower pressure drop across the column and wider choices of detectors. Of the numerous supercritical fluids studied, carbon dioxide has found the widest application in SFC due to a number of advantages, including mild critical parameters, inertness, environmental friendliness, low cost and easy access to high purity. With solvating power equivalent to that of pentane or hexane under supercritical conditions, pure carbon dioxide cannot solvate and elute high polarity analytes very easily. Recent developments in the applications of polar modifiers to the carbon dioxide mobile phase in packed column supercritical fluid chromatography (pSFC) have established the place of SFC in modern chromatographic sciences. Other high fluidity mobile phases are also under study.[1–3] These include subcritical, near-critical, enhanced-fluidity liquids and solvating gases.

In this chapter, the fundamental merits of using supercritical fluids and other high fluidity mobile phases in chromatography, the relationship of SFC to GC and high-performance liquid chromatography (HPLC), instrumentation of SFC and common mobile and stationary phases are discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 THEORY AND PRINCIPLES OF SUPERCRITICAL FLUID CHROMATOGRAPHY

2.1 Properties of Supercritical Fluids

A supercritical fluid corresponds to a condition where the compound is compressed above its critical pressure (P_c) and heated above its critical temperature (T_c) (Figure 1). Supercritical fluids have properties that are similar to those of both gases and liquids. The density of a supercritical fluid is close to that of a liquid, while its viscosity approaches that of a gas, and diffusion coefficients of supercritical fluids are typically intermediate between those of gases and liquids (Table 1). The phase change between gas and liquid is accompanied by abrupt physical property changes. However, the transition to and from the supercritical state is continuous and no abrupt changes in physical properties occur when passing through the critical point.[4] Figure 2 shows the approximate regions occupied by mobile phases used in GC, SFC and LC on a diffusion coefficient, D_m, vs density, ρ, plot.[5] Also, Figure 2 clearly shows the large difference in diffusion coefficients between gases and liquids.

Because supercritical fluids are highly compressible, the properties of supercritical fluids vary with pressure. For example, the density of supercritical CO_2 at approximately 40 °C can be varied from 0.25 to 0.8 g cm^{-3} by changing the pressure from 75 to 300 bar.[6] Accordingly, the solvent strength of a supercritical fluid can be varied by changing the pressure, which also means that analyte solubility is controlled by pressure variation. Similarly,

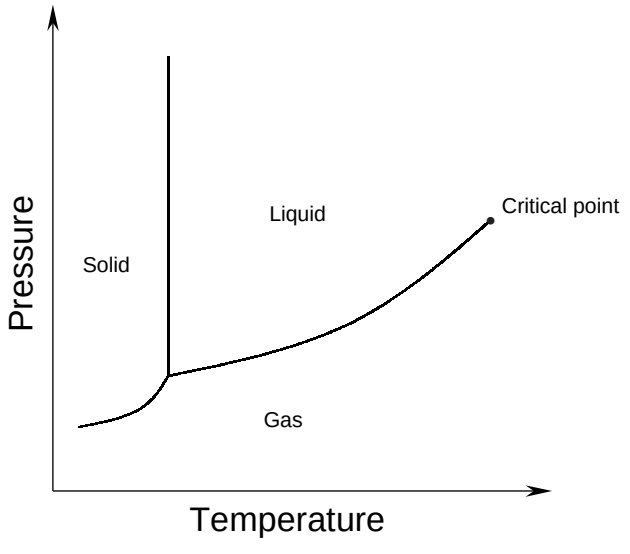

Figure 1 Phase diagram for a pure substance. The regions for the three phase states are labeled. When temperature and pressure are above the critical point of the substance, a supercritical state is obtained.

Table 1 Typical physical properties of gases, supercritical fluids and liquids

	Density (g mL^{-1})	Viscosity (cP)	Self-diffusion coefficients (cm^2 s^{-1})
Gas (30 °C, 1 atm)	$0.6–2 \times 10^{-3}$	$1–3 \times 10^{-2}$	0.1–0.4
Supercritical fluid	0.2–0.9	$1–9 \times 10^{-2}$	$0.2–0.7 \times 10^{-3}$
Liquid (30 °C, 1 atm)	0.6–1.6	0.2–3	$0.2–2 \times 10^{-5}$

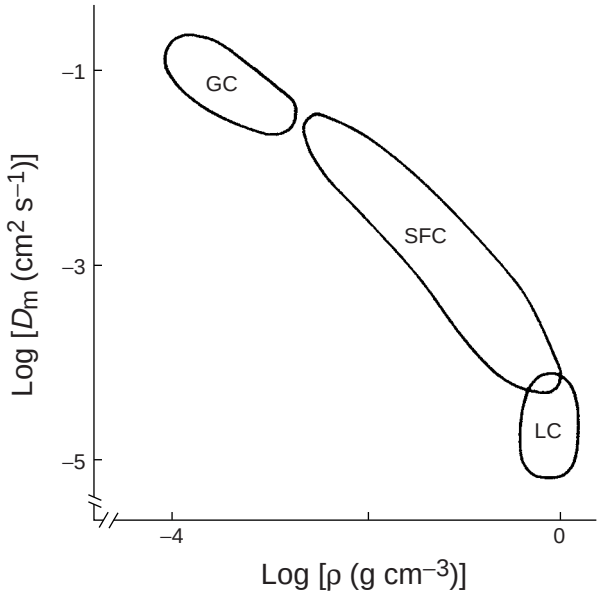

Figure 2 Schematic illustration of the relationship between the diffusion coefficient (D_m) and density (ρ) of mobile phases in GC, SFC and LC. The relatively defined regions occupied by the mobile phases of GC and LC are bridged by the elongated region for SFC. (Reproduced from Schoenmakers and Uunk[5] by permission of Marcel Dekker, Inc.)

mass transport properties, such as viscosity and diffusion rates, vary with pressure. The precise control of solvent properties by fine tuning the pressure of supercritical solvents is unsurpassed by any other phase.

Carbon dioxide, CO_2, is the most commonly used supercritical fluid. Its low critical temperature and pressure (31.3 °C and 72.8 atm) allow most separations to be undertaken near room temperature. CO_2 also has low toxicity and is environmentally benign. CO_2 is also highly compatible with a large range of detectors. However, the solvent strength of CO_2 is low. Even under high pressures, polar analytes will not dissolve in supercritical CO_2.

The solvent strengths of supercritical fluids were experimentally measured by monitoring the shift of the ultraviolet/visible (UV/VIS) absorbance maxima for a number of solvatochromic dyes.[7–10] The solvent strength of CO_2 was shown to be close to that of pentane

or hexane.[6] Despite the low polarity of supercritical CO_2, it remains the dominant mobile phase for SFC in applications to date due to the merits referred in the previous section. Furthermore, polarity enhancement of CO_2-based mobile phases is accomplished by the addition of polar modifiers and/or additives.

2.2 Why Supercritical Fluid Chromatography?

The major chromatographic advantages of using supercritical fluids are lower viscosity and higher diffusivities and concomitant higher efficiency, increased speed of analysis and lower pressure drop, as compared to LC. The connection between fast mass transport properties and improved chromatographic performance is delineated in the following sections.

2.2.1 Band Dispersion as a Function of Mobile Phase Velocity

The van Deemter equation will be used to correlate chromatographic efficiency with the mobile phase velocity. An uncoupled van Deemter equation[11] for packed columns in LC is given by Equation (1):

$$H = A + \frac{B}{u} + Cu \tag{1}$$

where H is the plate height (or band dispersion), u is the linear velocity of the mobile phase, and A, B and C are velocity-independent parameters that are determined by the properties of the mobile and stationary phases. The A term is the contribution from multiple flow paths of a packed column, the B term is the contribution to band dispersion from longitudinal (along the flow axis) diffusion and the C term represents the combined obstruction of mass transfer in the mobile phase and the stationary phase or C_m and C_{sm}. The B and C terms can be further expressed by Equations (2) and (3):

$$B = \gamma D_m \tag{2}$$

$$C = \frac{f(k')}{D_m} \tag{3}$$

where γ is the obstruction factor, k' is the retention factor and D_m is the average diffusion coefficient of the solute in the column. For incompressible mobile phases, D_m is independent of pressure, while for compressible mobile phases, such as in GC and SFC, D_m is influenced by flow rate when column pressure is a function of flow rate. More complicated expressions have been used to approximate the relationship between the average column pressure and the flow rate for compressible mobile phases.[12] Nevertheless the results from the more complicated model were comparable to the conclusions shown below. For the sake of simplicity, the independence of D_m on the column flow rate is assumed.

By taking the first derivative of Equation (1) with respect to u and setting the derivative to zero, the optimal mobile phase velocity, u_{opt}, for the minimum plate height, H_{min}, is obtained, as shown in Equations (4) and (5):

$$u_{opt} = D_m \sqrt{\frac{\gamma}{f(k')}} \tag{4}$$

and

$$H_{min} = A + \sqrt{2\gamma \times f(k')} \tag{5}$$

Equation (4) shows that u_{opt} is proportional to D_m; thus, the mobile phase with higher solute diffusivity gives higher u_{opt}. Since the diffusion coefficient of a solute in a supercritical fluid is larger than that in liquids, the u_{opt} using supercritical mobile phases is higher than using a liquid solvent. At high velocities the plate height is dominated by the C term, and is inversely proportional to D_m. Higher D_m values result in a small C term, which means smaller band dispersion and higher efficiency in the high velocity range. Equation (5) indicates that H_{min} (or maximum efficiency) is independent of D_m or the mobile phase properties. In other words, the maximum efficiency using a particular column is irrelevant to the type of chromatography, i.e. in principle GC, SFC and LC generate the same efficiency at their individual optimal flow rates. However, practically, the optimal linear velocity, u_{opt}, for LC using a typical column is small. Thus, the routine practice of HPLC is at much higher linear velocity than its optimum value, which results in lowered efficiency. At a high mobile phase velocity, the efficiency decreases in the order of GC, SFC and LC. In summary, for samples previously only amenable to HPLC, SFC provides higher efficiency and shorter analysis time, while for samples just beyond the scope of GC, SFC analysis is not compromised severely on efficiency and analysis time.

2.2.2 Effect of Viscosity on Chromatographic Performance

Equation (6) shows the relationship between the pressure drop and the mobile phase viscosity as described by Darcy's law:[13]

$$\Delta P = \frac{\phi \eta u L}{d_p^2} \tag{6}$$

where ϕ is a dimensionless flow resistance parameter, η is the viscosity of the mobile phase, u is the average linear velocity, and L and d_p are the column length and the particle diameter. Darcy's law indicates that a lower pressure drop across the column can be obtained by decreasing

the viscosity of the mobile phase. Therefore, for the same column and flow rate, SFC has a lower pressure drop than LC. The lower pressure drop across a column in SFC allows longer columns or more columns to be coupled together to obtain higher efficiency and better separation.

2.3 Closely Related Chromatographic Techniques and Unified Chromatography

The properties of a fluid,[4,14,15] i.e. density, viscosity, diffusion coefficient and solvating power, change continuously near the critical point. Therefore, chromatographic systems using other portions of the phase diagram that are near the critical point may have merit as well. Subcritical fluid chromatography (chromatography done under liquid state conditions that are near the critical point) often has similar objectives and uses the same instrumentation as SFC. Sometimes, subcritical fluids are preferred to supercritical fluids when the critical temperatures of the mobile phases are too high, or the analytes are thermally labile. For example, because chiral separations are often enthalpy controlled, the selectivity of the separation is improved by decreasing the temperature of the separation (i.e. using subcritical conditions).

Other approaches have also been taken to generate mobile phases with low viscosities and high diffusivities. These include high-temperature liquid chromatography (HTLC) and enhanced-fluidity liquid chromatography (EFLC). In high temperature HPLC the column temperature is raised to lower the mobile phase viscosity. For example, at 60 °C the viscosity of water is about half of that at 20 °C.[16] HTLC using heated water (up to 210 °C and 20 bar)[17,18] and acetonitrile (up to 150 °C)[19] has been reported. The lower viscosity at high temperatures is accompanied by higher vapor pressure and the column outlet pressure was increased above atmospheric pressure to prevent the mobile phase from boiling. Enhanced-fluidity HPLC involves the addition of a low-viscosity liquid, such as liquid CO_2 or liquid fluoroform, to organic solvents. This greatly lowers the viscosity of the solvent while maintaining a solvent strength similar to that of the pure organic solvent.[20,21]

While SFC fits between GC and LC, HTLC and EFLC are techniques that provide a continuum between SFC and LC, and solvating gas chromatography (SGC)[22] is an effort to bridge the gap between GC and SFC. In SGC, where capillary SFC instrumentation is commonly used, high pressures are exerted at the column inlet to maintain a supercritical, subcritical or liquid state of the mobile phase, but low pressures (slightly higher than atmospheric) at the column outlet result in phase transitions down the length of the column. Examples of SGC include the use of carbon dioxide[23] and ammonia as mobile phases.[23]

In all these studies, a common and key point is to maintain the column outlet pressure at a pressure greater than atmospheric pressure. Once the outlet pressure is controlled, gases can be compressed to supercritical fluid conditions or liquid state conditions. In fact, the mobile phase can essentially span between liquid and gas phases depending on the application without the necessity to distinguish the exact state in practice. This prompted the concept of ''unified chromatography''.[24,25]

2.4 Capillary Supercritical Fluid Chromatography Versus Packed Column Supercritical Fluid Chromatography

Capillary supercritical fluid chromatography (cSFC) experienced rapid development in the 1980s, while pSFC did not rival and prevail over cSFC until the early 1990s. Although packed capillary columns are also used,[26] cSFC will be used in this text to refer to open tubular capillary column SFC.

cSFC and pSFC differ in instrumentation and application. The column pressure of cSFC is usually maintained through the upstream pump and the after-column fixed restrictor. Higher flow rates are associated with higher column pressures. Owing to the low ratio between the stationary phase volume to column void volume, cSFC has a lower pressure drop per unit column length, which allows long column lengths to be used to achieve high efficiency. cSFC has high peak capacity and is well suited to analyze complex mixtures.[27–29] In cSFC a single pump is used to deliver a neat mobile phase and pressure programming is common. Direct coupling of GC detectors, such as flame ionization detectors (FIDs), is advantageous for high sensitivity measurements. However, insufficient solvating power of the mobile phase (CO_2 normally) even at high pressures precludes it from analyzing moderate or highly polar compounds.

pSFC is primarily used for analysis of polar analytes that are beyond the scope of cSFC through the use of polar modifiers. It has a larger ratio of the stationary phase volume to column void volume, leading to higher retention and necessitating the use of stronger eluent, such as polar modifiers. Compared with cSFC, pSFC has a higher flow rate and consequently a high pressure drop. The pressure gradient along the column results in gradients of mobile phase density, diffusion coefficient of the solute and flow rate along the column. Although pressure drops can be greater than 100 atm, no serious compromise on column efficiency is observed.[30–32] pSFC gives shorter analysis time and higher efficiency per unit time. The large mass flow permits the usage of a double pump system for preparing binary or ternary mobile phases. It also makes independent control of flow rate and

pressure possible, which is critical in decoupling the two for method development and optimization. The column pressure is controlled by a (electronically controlled) pressure regulator placed after the column. Polar modifier and/or additive, supplied from the second pump, are normally used to enhance the eluting power and solubility of analytes. pSFC uses LC columns, which provides a wide stationary phase selection. Inherently, pSFC exhibits higher sensitivity, reproducibility and sample loadability. However, pSFC cannot use some GC detectors, such as FIDs, due to the presence of the modifier. Recent advances in technology have made modern packed column supercritical fluid chromatographs commercially available, which has greatly promoted the development and popularity of pSFC.

3 INSTRUMENTATION

The main components of the SFC chromatograph include pump, injector, oven, column, pressure control device and detector. The instrumentation used for cSFC and pSFC is not the same. Schematic block diagrams for chromatographs used in pSFC and cSFC are shown in Figure 3.

3.1 Pump and Mixer

The chromatograph for cSFC usually contains a single syringe pump for delivering neat or premixed mobile phase, while that for pSFC often contains a high pressure reciprocating pump and an additional microflow high pressure pump for delivering the modifier and additive. The syringe pump in cSFC typically delivers mobile phase fluid at a rate of 1 to 80 $\mu L\,min^{-1}$, whereas, the reciprocating pump in pSFC works in a range of 0.5 to 10 $mL\,min^{-1}$ and the microflow pump delivers modifiers in the range of 1–20% of the fluid.[33–36] The mobile phase, e.g. CO_2, is normally from pressurized cylinders. For precise delivery of the fluid, compressibility compensation has to be included in the pumping process. For pSFC, owing to the large flow rate, a reciprocating pump is required for uninterrupted delivery of the mobile phase. Precise control of temperature at the pump head is essential for compressibility compensation.[37] Because the compressing process is exothermic, the pump head is usually well chilled using a circulating bath or thermoelectric (Peltier) cooler.

For pSFC, the pressurized fluid and modifier are dynamically mixed at the high pressure mixer. A pulse damper is usually placed in line to smooth the fluid pulsation produced by reciprocating pumps.

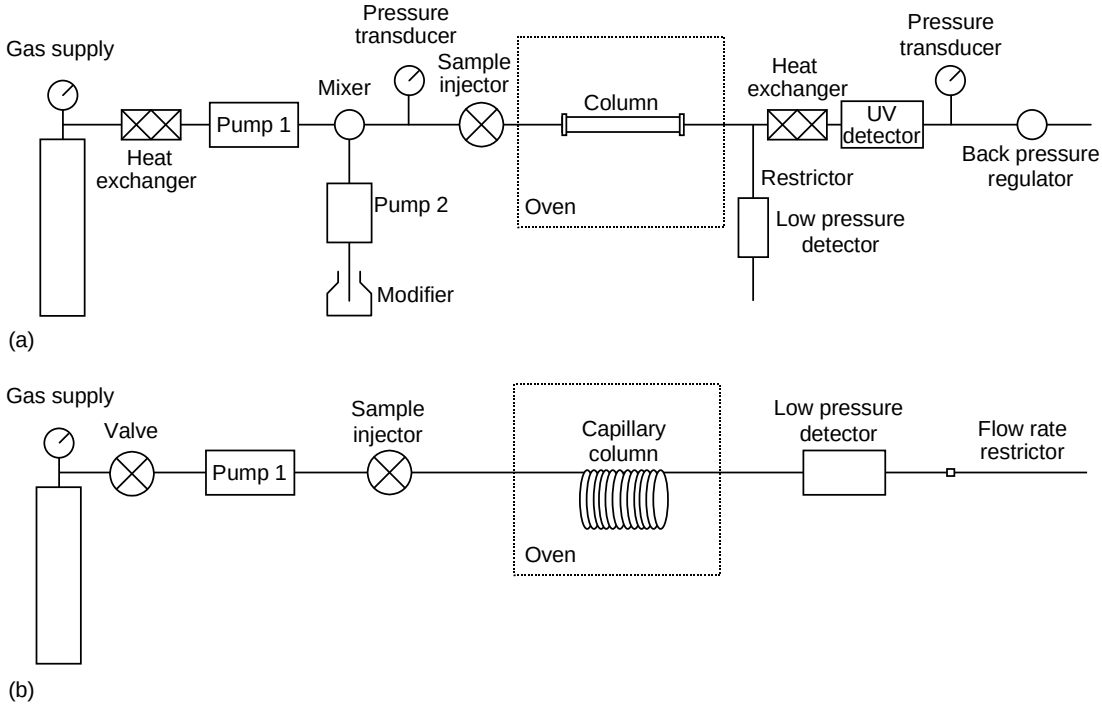

Figure 3 Schematic diagrams showing the instrumental set-up for (a) pSFC and (b) cSFC. In pSFC, two pumps and an online mixer are normally used to deliver a mixture of mobile phases. The column outlet pressure is maintained by the back pressure regulator. To use GC detectors, flow splitting before the back pressure regulator is necessary. In cSFC, a single pump is normally used to deliver pure or premixed mobile phase. The column pressure is controlled at the inlet by the pump, resulting in the coupling of column pressure and flow rate.

3.2 Sample Injection

The amount of sample that can be introduced depends on the type of SFC. In cSFC, the injection volume is about 0.5 nL. To obtain reproducible injections for such a small volume, an injection valve with a split mechanism is necessary. Currently, two designs are commonly used, the split/splitless[33,38] and the timed-split.[36,39] Both allow injection of only a fraction of the sample loaded in the internal loop.

In pSFC a standard LC rotary injector with loop size of 5 to 20 μL can be used. When a large loop size is used, special precautions should be taken when switching the valve from injection position to loading position to minimize band dispersion because the expansion of the high-pressure fluid is vigorous. Since SFC is usually normal-phase chromatography (see discussion later), an injection solvent of weaker polarity than the mobile phase is preferred in order to minimize interference of the injection solvent with the analyte retention. When injection solvents are used that are too strong, analytes tend to flow down the column with the solvent plug for a significant fraction of the column length instead of interacting with the stationary phase. This leads to peak widening and poor selectivity. Thus, when polar solvents (e.g. water or methanol) cannot be avoided, the injection volume should be minimized. When large volumes of diluted solute are to be used, preconcentration or adsorption of the solute onto a precolumn or guard column mounted at the loop position is useful.

3.3 Column and Oven

The column temperature of cSFC is maintained using GC-type ovens. pSFC uses LC-type ovens. An upper temperature limit of 200 °C is normally sufficient. In fact, a significant proportion of the published pSFC experiments are run at subcritical conditions. Furthermore, the capability of the pSFC oven to lower temperature to subzero ranges is beneficial for enhancing selectivity of the separation, particularly for chiral separations.[40,41]

The columns for cSFC are typically made of fused silica with diameter of 0.025 to 0.1 mm and length of 1 to 35 m. The stationary phases are discussed in detail in a later section. The columns for pSFC (including microbore columns) are the same as those used in LC with dimensions of 0.5 to 4.6 mm inner diameter, 3 to 25 cm long and are normally made of stainless steel. Typical pSFC columns are packed with totally porous silica particles 3 to 10 μm in size.

3.4 Pressure Control

The mobile phase flow rate in cSFC is maintained typically by placing a small diameter open tube at the end of the column.[42] The pressure of the cSFC system is controlled by the outlet pressure of the pump. A pressure gradient, frequently used in cSFC, is generated by a programmed flow rate gradient. However, decreased column efficiency will result if the flow rates become too high. Independent pressure control and flow rate control in cSFC by using a pressure regulator with very small dead volume to accommodate the mass flow is still too technically challenging. However, precise independent pressure control is possible and has been implemented in pSFC by placing a pressure regulator after the column. An electromagnetically controlled needle or piston inside the regulator controls the closing and opening of an orifice in conjunction with the flow rate to maintain a specified pressure.

3.5 Detection

Generally speaking, SFC is superior to HPLC in the range of compatible detection methods, because most detectors used in GC can also be used in SFC, particularly for CO_2-based mobile phases. The applicability of a specific detector, though, depends on the nature of the analyte, flow rate and composition of the mobile phase.

The compatibility of SFC with the FID is one of its strengths. As a universal detection method with high sensitivity and wide dynamic range ($\sim 10^5$), FID is the detector of choice in cSFC, when normally no organic modifier is used and flow rate after mobile phase expansion is compatible with FID. However, for pSFC, GC detectors usually cannot be used directly due to the large mass flow and the use of combustible organic modifiers. To lower the mass flow, a flow splitter through a restrictor at a ''zero-dead volume'' tee is generally employed,[1] which does not interfere with the back pressure regulators. Two polar modifiers that do not produce FID response are water and formic acid. Use of subcritical water as mobile phase at 80 to 175 °C in the analysis of small alcohol compounds, hydroxyphenols and amino acids with flame ionization detection was reported by Miller and Hawthorne.[17] Other low-pressure detectors commonly used in SFC include the thermionic ionization detector (TID),[43,44] the flame photometric detector (FPD),[45,46] nitrogen–phosphorus detector (NPD),[47,48] the electron-capture detector (ECD),[38,49,50] and the chemiluminescence detector.[51,52]

An alternative universal detector to the FID, which is compatible with mobile phase modifiers, is the evaporative light-scattering detector (ELSD).[53] It is less sensitive and has a narrow dynamic range ($\sim 10^3$). Encouraging results using ELSD in SFC have been reported.[54,55]

UV absorption detectors are the most commonly used detection method in pSFC for analytes with UV-chromophores. However, differing from HPLC, the flow

cell must be capable of withstanding pressures up to 400 bar, because it is placed in-line before the pressure regulator. Other high pressure detection methods that are commonly used in SFC include fluorescence detection. Owing to the high column temperature usually used in SFC, to prevent interference of the UV or fluorescence detection from refractive index fluctuations during the cooling process, a heat exchanger placed before the detector is often used.[1]

Mass spectrometry (MS) affords a highly sensitive and selective means of detection. Numerous methods of coupling SFC to MS have been published.[56–61] Thermospray, electrospray, particle beam and atmospheric pressure chemical ionization (APCI) have attracted the most attention.[51] Atmospheric pressure ionization (electrospray and APCI) is currently undergoing the fastest growth, because it allows independent optimization of SFC and MS.[51]

Other detection methods that are increasingly used in SFC include Fourier transform infrared (FTIR) spectroscopy,[62,63] nuclear magnetic resonance (NMR),[64,65] and electrochemical[66] and amperometric detectors.[67,68]

4 STATIONARY PHASES FOR SUPERCRITICAL FLUID CHROMATOGRAPHY

The stationary phases for cSFC and pSFC are different. The stationary phases in cSFC are often those also used in GC, but they are typically further crosslinked when used for SFC. The film thickness of the stationary phase is typically 0.1 to 3 μm. The most commonly used stationary phases are polysiloxanes.[69] The polarity of the stationary phase is varied by using polysiloxanes with different substituent groups, among which methyl, phenyl and cyano are common choices. Combinations of these groups in different ratios satisfy various polarity needs. For example, dimethylpolysiloxane (100% methyl), phenylmethylpolysiloxane (50% phenyl, 50% methyl), cyanopolysiloxane (100% cyano) and cyanopropylmethylphenylmethylsiloxane are common stationary phases used in cSFC.

The stationary phases for pSFC are similar to those in HPLC. Porous silica is the most common particle for pSFC. The silica packing consists of silica particles (3–10 μm diameter) with 50–500 Å pore sizes and 50–500 $m^2 g^{-1}$ surface area. The polarity of the porous silica is often modified through the attachment of a variety of groups to the organosilane groups. Common functional groups include methyl, phenyl, diphenyl, naphthyl, octadecyl, octyl, diol, cyanopropyl and aminopropyl groups. Unreacted silanol groups are deactivated by end-capping with silylation reagents (e.g. trimethylmethoxysilane). However, residual silanol groups always exist. Active sites on the surface of the particle interact strongly with polar analytes when neat CO_2 is used as the mobile phase, causing poor peak shapes in the chromatogram.[70–73] The addition of polar modifiers into the mobile phase effectively improves the chromatographic band shape as well as lowers the retention of the analyte.[74] Mobile phase modifiers are discussed in detail in the next section. Coating the silica particles with a thin layer of organic polymer is another approach to produce a stationary phase for SFC. Polymer coated silica stationary phases showed improved peak shapes, but still a polar modifier is necessary for effective elution of polar analytes.[1] Totally polymer-based particles, such as those made from polystyrene–divinylbenzene are also used. However, peak tailing is very common with polymeric surfaces as well. In addition, polymer-based particles often swell when exposed to supercritical mobile phases and exhibit less temperature and pressure (mechanical) resistance.[1]

For most of the stationary phases used in pSFC, the apparent mechanism of separation is normal-phase chromatography.[1] For HTLC and EFLC, where CO_2 is not the primary mobile phase, other modes of chromatography, such as reversed-phase, can be easily implemented.

5 MOBILE PHASES FOR SUPERCRITICAL FLUID CHROMATOGRAPHY

5.1 Pure Fluids

Pure fluid mobile phases provide simplicity of operation and often better detector compatibility. Therefore, whenever possible, pure fluids should be considered first. Pure CO_2 is the common first choice. It has been used in the analysis of nonpolar and moderately polar analytes. CO_2 has no permanent dipole moment but has a relatively large quadrupole moment allowing some interactions with polar analytes.[75] CO_2 is suitable for cSFC where a pressure gradient is commonly used to change the retention of solutes. Unlike GC, a temperature gradient is generally not applicable, since under similar pressure, increasing temperature decreases the density and solvent strength of the mobile phase, and therefore increases retention. However, the solvating power of CO_2 is quite limited even under high pressures.

Other pure solvents are also used in SFC. Low-molecular-weight alkanes,[76–78] such as ethane, propane, pentane and hexane, have been used. They exhibit stronger solvent strength than CO_2 and provide different selectivity toward compounds such as polycyclic aromatic hydrocarbons.[71] Their flammability and incompatibility

with FID greatly reduce their popularity. Nitrous oxide has the polarity and critical parameters similar to those of CO_2. It was used in SFC in the early stages of the technique, but no particular advantage was found,[71] while its explosive propensity when used for samples with high organic content[79] deterred further applications.[80] Sulfur hexafluoride is primarily used in the analysis of hydrocarbons.[81] Its compatibility with FID is not as good as carbon dioxide, but it has a wider transparent window for infrared (IR) detection. Trifluoromethane (fluoroform) has a permanent dipole moment and has a lower T_c (26.2 °C) and P_c (48.0 atm) than CO_2.[82] It is also not involved in ozone destruction.[83] It exhibits greater selectivity in the SFC analysis of polycyclic aromatic hydrocarbons than CO_2, nitrous oxide, ethane and $CClF_3$[73] and has stronger eluotropic power than CO_2 and methanol-modified CO_2.[84] The use of other hydrofluorocarbons in SFC is also under study.[85]

The choices for pure polar solvents for SFC are quite limited. Water, despite its unmatched low cost and environmental friendliness, has high critical parameters (374.1 °C and 217.6 atm). Chromatography using subcritical water has been reported.[18] Ammonia, a highly polar solvent with critical temperature and pressure of 132.4 °C and 112.0 atm, has been used as the chromatographic mobile phase in quite a few studies under supercritical or subcritical conditions[4,86,87] as well as for SGC.[24] However, the corrosive and toxic nature prevents it from being widely investigated.

Other than CO_2, none of the pure solvents examined are likely to find wide use. Their applications are preferred only when markedly better selectivity occurs or when CO_2 fails (e.g. because of reactivity toward the sample). Within the foreseeable future, CO_2 and CO_2 with polar modifiers will still dominate the SFC studies.

5.2 Binary or Ternary Mobile Phases

The addition of low percentages (e.g. 1–20%) of polar modifiers into CO_2 is an efficient means to enhance the solvent strength of mobile phase. However, the use of modifiers compromises the compatibility with most GC detectors. For instance, even a small amount of a combustible modifier (e.g. methanol) generates background signal in the FID. Only water and formic acid are exceptions.

The solvent strength enhancement of polar modifiers is disproportionate to its composition in the mixture. For example, the addition of 5% methanol to CO_2 increases the solvent strength to slightly greater than 25% of that of pure methanol.[88] To further improve the solvent strength of the mobile phase or the peak shape of the analyte, a third highly polar component is sometimes added at low concentrations (generally 0.1–1%). For example bases, water, and acids have been added to the modifier to improve performance.

The effects of adding polar modifiers or additives are multifaceted. The modifier can increase the solubility of the analyte, enhance the eluting power of the mobile phase and improve the chromatographic peak shape of the analyte. The solubility enhancement generally follows the solvent strength of the modifier, whereas, other effects depend on the interactions between the mobile phase–solute, mobile phase–stationary phase and solute–stationary phase pairs.[89] For example, the addition of methanol to CO_2 did not improve the elution of chrysene from an octadecylsilyl column, while the same methanol–CO_2 mixtures dramatically decreased the retention factor of chrysene on a sulfonic acid column.[1] The ability of modifiers to improve poor peak shapes is caused by the blocking of ''active sites''[90,91] (residual silanol groups on silica which provide secondary retention mechanisms) on the stationary phase. Other potentially useful interactions between the modifier and the solute include ion-pairing with the analyte[92] and suppression of analyte ionization.[93,94] The mechanism by which the modifier and solute interact is still a controversial issue. As a general rule, additives with similar functionality to the solute are most effective in improving the elution profile.[75]

The addition of modifiers or additives also increases the critical parameters (T_c and P_c) of the resultant mixtures. The degree of change depends on the critical parameters of the modifier or additive and the proportion of modifier added. Methanol–CO_2 mixtures, the most frequently employed system, show moderate increases in T_c and P_c, e.g. CO_2 with 20% methanol exhibits T_c and P_c of 72.2 °C and 76.7 atm, respectively.[95] When binary and ternary mobile phases are used, the chromatographic separation is often actually at subcritical mobile phase conditions.

The phase behavior of the binary or tertiary mobile phase must be known to avoid phase separation in the column and therefore to achieve high efficiency separations. Phase diagrams for commonly used mixtures of mobile phases are available, e.g. methanol–CO_2,[96,97] methanol–CHF_3,[98] methanol–water–CO_2[92,93] and methanol–water–CHF_3,[94] and tetrahydrofuran–CO_2.[99]

Binary and ternary mixtures for SFC can be prepared as premixed or through online mixing. The premixing of solvents provides accurate control of the composition and requires simple chromatographic design. Premixed CO_2 and methanol mixtures or other commonly used modifiers are commercially available in cylinders. However, due to the difference in the vapor pressure of the components, compositional change can occur with usage. For example, a more than twofold change in retention time was reported using premixed methanol–CO_2 throughout the

usage of the cylinder, due to presumably an increase in methanol concentration in the cylinder with time.[100] Online mixing requires two high pressure pumps and accurate control of the delivery ratio from the pumps, but it allows a quick change in modifier concentration and use of a modifier gradient.

Enhanced-fluidity liquids are low viscosity fluids that target the insufficient solvating power of supercritical CO_2 in chromatography and extraction. Instead of adding a polar modifier to CO_2, EFLC adds liquid CO_2 or other low viscosity liquids to common organic solvents.[21,22,101,102] The column pressure is maintained at a value high enough to ensure that the liquid is a single-phase system. The fluidity enhancement observed from the addition of liquid CO_2 to common liquids is significant. For example, by adding 40% CO_2 to methanol the viscosity decreased by approximately 50% while the solvent strength remains similar to that of pure methanol. While most CO_2-based SFC is restricted in normal-phase chromatography, EFLC has been applied to normal,[103] reversed-phase,[22,97] and size-exclusion[95,104] chromatographies. In fact, EFLC can be viewed as bridging the gap between SFC and HPLC.

6 PERSPECTIVES AND FUTURE DEVELOPMENT

Although SFC was first demonstrated in 1960,[105] its major development and recognition of its importance occurred during the early 1990s. Currently, SFC is a standard analytical method alongside GC and LC in many laboratories. It tends to replace normal-phase LC. SFC has found diverse applications in all types of chromatographic analysis traditionally accomplished by GC and HPLC. Numerous reviews and monographs are available on the applications of SFC.[1,2,30,75,106] SFC has been used for the analysis of common chemicals, petroleum-based compounds, pharmaceuticals, food components, natural products, biologically active compounds, agricultural chemicals, synthetic polymers and environmental contaminants.

In the coming years, further applications of analytical and preparative-scale SFC for the separation of chiral pharmaceuticals are expected. Compared to chiral GC, SFC not only extends the sample applicability but also allows separations at lower temperature to prevent potential racemization. For chiral samples amenable to HPLC analysis, SFC provides higher efficiency, a critical parameter in increasing enantiomeric resolution, and shorter analysis time. Enantiomeric separation at cryogenic temperatures using SFC potentiates higher enantiomeric selectivity at much faster speeds compared to HPLC. SFC is also advantageous in preparative chiral separation and solvent removal.

Studies of the mechanism of modifier or additive continue. As the idea of ''unified chromatography'' prevails, more investigations are expected in emerging areas, such as SGC, EFLC and HTLC.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
cSFC	Capillary Supercritical Fluid Chromatography
ECD	Electron-capture Detector
EFLC	Enhanced-fluidity Liquid Chromatography
ELSD	Evaporative Light-scattering Detector
FID	Flame Ionization Detector
FPD	Flame Photometric Detector
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-performance Liquid Chromatography
HTLC	High-temperature Liquid Chromatography
IR	Infrared
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NPD	Nitrogen–Phosphorus Detector
pSFC	Packed Column Supercritical Fluid Chromatography
SFC	Supercritical Fluid Chromatography
SGC	Solvating Gas Chromatography
TID	Thermionic Ionization Detector
UV/VIS	Ultraviolet/Visible

RELATED ARTICLES

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Instrumentation of Gas Chromatography

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

REFERENCES

1. T.A. Berger, *Packed Column SFC*, The Royal Society of Chemistry, Loughborough, 1995.
2. T.L. Chester, J.D. Pinkston, D.E. Raynie, 'Supercritical Fluid Chromatography and Extraction', *Anal. Chem.*, **70**, 301R–319R (1998).
3. T.L. Chester, 'Chromatography From the Mobile-phase Perspective', *Anal. Chem.*, 165A–169A (1997).

4. H.H. Lauer, D. McManiglll, R D. Board, 'Mobile-phase Transport Properties of Liquefied Gases in Near-critical and Supercritical Fluid Chromatography', *Anal. Chem.*, **55**, 1370–1375 (1983).
5. P.J. Schoenmakers, L.G.M. Uunk, 'Mobile Phase and Stationary Phases for Supercritical Fluid Chromatography', in *Advances in Chromatography*, eds. J.C. Giddings, E. Gruska, P.R. Brown, Marcel Dekker, New York, 1–80, Vol. 30, 1989.
6. S. Angus, B. Armstrong, K.M. deReuck (eds.), *International Thermodynamic Tables of the Fluid State. Carbon Dioxide*, Pergamon Press, New York, Table 5, 1976.
7. J.F. Deye, T.A. Berger, A.G. Anderson, 'Nile Red as a Solvatochromic Dye for Measuring Solvent Strength in Normal Liquids and Mixtures of Normal Liquids with Supercritical and Near Critical Fluids', *Anal. Chem.*, **62**, 615–622 (1990).
8. J.A. Hyatt, 'Liquid and Supercritical Carbon Dioxide and Organic Solvents', *J. Org. Chem.*, **49**, 5097–5101 (1984).
9. M.E. Sigman, S.M. Lindley, J.E. Leffler, 'Supercritical Carbon Dioxide: Behavior of π^* and β Solvatochromic Indicators in Media of Different Densities', *J. Am. Chem. Soc.*, **107**, 1471–1472 (1985).
10. C.R. Yonker, S.L. Frye, D.R. Kalkwarf, R.D. Smith, 'Characterization of Supercritical Fluid Solvents Using Solvatochromic Shifts', *J. Phys. Chem.*, **90**, 3022–3026 (1986).
11. J.H. Knox, J.F. Parcher, 'Effect of Column to Particle Diameter Ratio on the Dispersion of Unsorbed Solutes in Chromatography', *Anal. Chem.*, **41**, 1599–1606 (1969).
12. Y. Shen, Y. Yang, M.L. Lee, 'Fundamental Consideration of Packed-capillary GC, SFC, and LC Using Nonporous Silica Particles', *Anal. Chem.*, **69**, 628–635 (1997).
13. A.E. Scheidegger, *The Physics of Flow Through Porous Media*, Macmillan Publishing, New York, 68, 1960.
14. T.A. Berger, 'Density of Methanol–Carbon Dioxide Mixtures at Three Temperatures – Comparison with Vapor–Liquid-equilibria Measurements and Results Obtained from Chromatography', *J. High Resolut. Chromatogr.*, **14**, 312–316 (1991).
15. T.A. Berger, W.H. Wilson, J.F. Deye, 'Analysis of Carbamate Pesticides by Packed-column Supercritical-fluid Chromatography', *J. Chromatogr. Sci.*, **32**, 179–184 (1994).
16. J.H. Perry, *Chemical Engineers Handbook*, 4th edition, McGraw-Hill, New York, 1963.
17. D.J. Miller, S.B. Hawthorne, 'Subcritical Water Chromatography with Flame Ionization Detection', *Anal. Chem.*, **69**, 623–627 (1997).
18. R.M. Smith, R.J. Burgess, 'Superheated Water as an Eluent for Reversed-phase High-performance Liquid Chromatography', *J. Chromatogr.*, **785**, 49–55 (1997).
19. R. Trones, A. Tangen, W. Lund, T. Greibrokk, 'Packed Capillary High-temperature Chromatography Coupled to Inductively Coupled Plasma Mass Spectrometry', *J. Chromatogr. A*, **835**, 105–112 (1999).
20. Y. Cui, S.V. Olesik, 'High-performance Liquid-chromatography Using Mobile Phases with Enhanced Fluidity', *Anal. Chem.*, **63**, 1812–1819 (1991).
21. S.T. Lee, S.V. Olesik, 'Comparison of Enhanced Fluidity and Elevated-temperature Mobile Phases in Reversed-phase High-performance Liquid-chromatography', *Anal. Chem.*, **66**, 4498–4506 (1994).
22. Y. Shen, M.L. Lee, 'Separation of Polar Compounds by Solvating Gas Chromatography with Carbon Dioxide Mobile Phase', *Chromatographia*, **46**, 587–592 (1997).
23. Y. Shen, M.L. Lee, 'Solvating Gas Chromatography Using Ammonia as Mobile Phase', *J. Chromatogr. A*, **806**, 345–348 (1998).
24. D. Ishii, T. Takeuchi, 'Unified Fluid Chromatography', *J. Chromatogr. Sci.*, **27**, 71–74 (1989).
25. D. Tong, K.D. Bartle, A.A. Clifford, R.E. Robinson, 'Unified Chromatograph for Gas Chromatography, Supercritical Fluid Chromatography and Micro-liquid Chromatography', *Analyst*, **120**, 2461–2467 (1995).
26. Y. Shen, M.L. Lee, 'Packed Column Supercritical Fluid Chromatography Using Neat Carbon Dioxide', in *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*, eds. K. Anton, C. Berger, Marcel Dekker, Inc., New York, 125–160, 1998.
27. M. Novotny, S.R. Springston, P.A. Peaden, J.C. Fjeldsted, M.L. Lee, 'Capillary Supercritical Fluid Chromatography', *Anal. Chem.*, **53**, 407A–414A, 1981.
28. M. Novotny, D. Ishii (eds.), *Microcolumn Separations, Columns, Instrumentation and Ancillary Techniques*, Elsevier, Amsterdam, Vol. 30, 1985.
29. T.A. Berger, *Packed Column SFC*, RSC Monograph Series, Royal Society of Chemistry, Cambridge, UK, 56, 1995.
30. W. Li, A. Malik, M.L. Lee, 'Pressure Drop Effects in Packed Capillary Column Supercritical Fluid Chromatography', *J. Chromatogr. A*, **758**, 117–123 (1997).
31. C. Bouigeon, D. Thiebaut, M. Caude, 'Long Packed Column Supercritical Fluid Chromatography: Influence of Pressure Drop on Apparent Efficiency', *Anal. Chem.*, **68**, 3622–3630 (1996).
32. T.A. Berger, W.H. Wilson, 'Packed Column Supercritical Fluid Chromatography with 220 000 Plates', *Anal. Chem.*, **65**, 1451–1455 (1993).
33. M.L. Lee, B. Xu, E.C. Huang, N.M. Djordjevic, H.-C.K. Chang, K.E. Markides, *J. Microcolumn Sep.*, **1**, 7–13 (1989).
34. T.A. Berger, T. Greibrokk, 'Instrumentation for SFC and SFE', *Chromatogr.: Princ. Pract.*, **2**, 107–138 (1999).
35. C.F. Poole, S.A. Schuette, *Contemporary Practice of Chromatography*, Elsevier, Amsterdam, 1984.
36. K. Anton, 'Supercritical Fluid Chromatography: Instrumentation', in *Encyclopedia of Analytical Science*, ed. A. Townshend, Academic Press, San Diego, Vol. 8, 1995.

37. T.A. Berger, 'Chemical Basis for the Instrumentation Used in Packed Column Supercritical Fluid Chromatography', in *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*, eds. K. Anton, C. Berger, Marcel Dekker, Inc., New York, 19–58, 1998.
38. B.E. Richter, N.L. Porter, D.E. Knowles, D.W. Later, M.R. Anderson, E.R. Campbell, 'Reproducibility in Capillary Supercritical Fluid Chromatography – Comparison of Injection Techniques', *J. High Resolut. Chromatogr., Chromatogr. Commun.*, **11**, 29–32 (1988).
39. J. Kohler, A. Rose, G. Schomburg, 'Instrumentation for SFC Systems – Different Sampling and Restriction Designs', *J. High Resolut. Chromatogr., Chromatogr. Commun.*, **11**, 191–197 (1988).
40. C. Wolf, W.H. Pirkle, 'Enantioseparations by Subcritical Fluid Chromatography at Cryogenic Temperatures', *J. Chromatogr. A*, **785**, 173–178 (1997).
41. K.L. Williams, L.C. Sander, 'Enantiomer Separations on Chiral Stationary Phase in Supercritical Fluid Chromatography', *J. Chromatogr. A*, **785**, 149–158 (1997).
42. C.F. Poole, S.K. Poole, *Chromatography Today*, Elsevier, Amsterdam, 640, 1993.
43. B.E. Richter, J.G. Wangsgaard, J.T. Swanson, D.J. Bornhop, M.R. Andersen, 'Gas-chromatographic Detectors in SFC', *J. Chromatogr. Sci.*, **27**, 303–308 (1989).
44. P.A. David, M. Novotny, 'Characterization of the Nitrogen and Phosphorus Thermionic Detector Response in Capillary Supercritical Fluid Chromatography', *Anal. Chem.*, **61**, 2082–2086 (1989).
45. S.V. Olesik, L.A. Pekay, E.A. Paliwoda, 'Characterization and Optimization of Flame Photometric Detection in Supercritical Fluid Chromatography', *Anal. Chem.*, **61**, 58–65 (1989).
46. L.A. Pekay, S.V. Olesik, 'Supercritical Fluid Chromatography Flame Photometric Detection – Determination of High Molecular-weight Compounds', *Anal. Chem.*, **61**, 2616–2624 (1989).
47. P.A. David, M. Novotny, 'Separation of Alpha-keto Acids by Capillary Supercritical Fluid Chromatography as Their Quinoxalinol Derivatives', *J. Chromatogr.*, **452**, 623–629 (1988).
48. J.G.J. Mol, B.N. Zegers, H. Lindeman, U.A.Th. Brinkman, 'Packed-capillary Supercritical Fluid Chromatography of Pesticides Using Phosphorus-selective Detection', *Chromatographia*, **32**, 203–210 (1991).
49. H-C.K. Chang, L.T. Taylor, 'The Performance of Electron-capture Detection After Capillary Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **28**, 29–33 (1990).
50. J.T.B. Strode, III, L.T. Taylor, 'Optimization of Electron-capture Detector When Using Packed-column Supercritical Fluid Chromatography with Modified Carbon Dioxide', *J. Chromatogr.*, **723**, 361–369 (1996).
51. W.T. Foreman, C.L. Shellum, J.W. Birks, R.E. Sievers, 'Supercritical Fluid Chromatography with Sulfur Chemiluminescence Detection', *J. Chromatogr.*, **465**, 23–33 (1989).
52. M.T. Combs, M. Ashraf-Khorassani, L.T. Taylor, 'Optimization of Chemiluminescent Nitrogen Detection for Packed-column Supercritical Fluid Chromatography with Methanol-modified CO_2', *Anal. Chem.*, **69**, 3044–3048 (1997).
53. J.T.B. Strode, III, L.T. Taylor, K. Anton, N. Pericles, M. Bach, 'Packed Column Supercritical Fluid Chromatography/Evaporative Light-scattering Detection', in *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*, eds. K. Anton, C. Berger, Marcel Dekker, Inc., New York, 97–124, 1998.
54. A.J. Berry, E.D. Ramsey, M. Newby, D.E. Games, 'Applications of Packed-column SFC Using Light-scattering Detection', *J. Chromatogr. Sci.*, **34**, 245–253 (1996).
55. J.T.B. Strode, III, L.T. Taylor, T.A. van Beek, 'Supercritical Fluid Chromatography of Ginkgolides A, B, C and J and Bilobalide', *J. Chromatogr. A*, **738**, 115–122 (1996).
56. M.T. Combs, M. Ashraf-Khorassani, L.T. Taylor, 'Packed Column Supercritical Fluid Chromatography/Mass Spectroscopy: a Review', *J. Chromatogr. A*, **785**, 85–100 (1997).
57. J.D. Pinkston, T.L. Chester, 'Putting Opposites Together, Guidelines for Successful SFC/MS', *Anal. Chem.*, 650A–656A, 1995.
58. J.K. Broadbent, B.S. Martincigh, M.W. Raynor, L.F. Salter, R. Moulder, P. Sjoberg, K.E. Markides, 'Capillary Supercritical Fluid Chromatography Combined with Atmospheric Pressure Chemical Ionization Mass Spectrometry for the Investigation of Photoproduct Formation in the Sunscreen Absorber 2-ethylhexyl-*p*-methoxycinnamate', *J. Chromatogr. A*, **732**, 101–110 (1996).
59. M. Tuomola, M. Hakala, P. Manninen, 'Determination of Androstenone in Pig Fat Using Packed Column Supercritical Fluid Chromatography/Mass Spectrometry', *J. Chromatogr. B*, **719**, 25–30 (1998).
60. R. Pal, M. Juhas, A. Stumpf, 'Detailed Analysis of Hydrocarbon Groups in Diesel Range Petroleum Fractions with On-line Coupled Supercritical Fluid Chromatography/Gas Chromatography/Mass Spectrometry', *J. Chromatogr. A*, **819**, 249–257 (1998).
61. D.G. Morgan, K.L. Harbol, N.P. Kitrinos, Jr, 'Optimization of a Supercritical Fluid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry Interface Using an Ion Trap and Two Quadrupole Mass Spectrometers', *J. Chromatogr. A*, **800**, 39–49 (1998).
62. S.H. Smith, S.L. Jordan, L.T. Taylor, J. Dwyer, J. Willis, 'Packed Column Supercritical Fluid Chromatography/Mobile Phase Elimination Fourier Transform Infrared

Spectrometry Employing Modified Fluids', *J. Chromatogr. A*, **764**, 295–300 (1997).
63. J. Yang, P.R. Griffiths, 'Separation and Identification of Sulfanilamides by Capillary Supercritical Fluid Chromatography/Fourier Transform Infrared Spectroscopy', *J. Chromatogr. A*, **785**, 111–119 (1997).
64. K. Albert, 'Supercritical Fluid Chromatography/Proton Nuclear Magnetic Resonance Spectroscopy Coupling', *J. Chromatogr. A*, **785**, 65–83 (1997).
65. U. Braumann, H. Hendel, S. Strohschein, M. Spraul, G. Krack, R. Ecker, K. Albert, 'Separation and Identification of Vitamin A Acetate Isomers by Supercritical Fluid Chromatography/1H NMR Coupling', *J. Chromatogr. A*, **761**, 336–340 (1997).
66. S.F. Dressman, A.M. Simeone, A.C. Michael, 'Supercritical Fluid Chromatography with Electrochemical Detection of Phenols and Polyaromatic Hydrocarbons', *Anal. Chem.*, **68**, 3121–3127 (1996).
67. S.R. Wallenborg, K.E. Markides, L. Nyholm, 'Development of an Amperometric Detector for Packed Capillary Column Supercritical Fluid Chromatography', *Anal. Chem.*, **69**, 439–445 (1997).
68. S.R. Wallenborg, K.E. Markides, L. Nyholm, 'Oxidative and Reductive Amperometric Detection of Phenol and Nitroaromatic Compounds in Packed Capillary Column Supercritical Fluid Chromatography', *J. Chromatogr. A*, **785**, 121–128 (1997).
69. J.K. Haken, 'Developments in Polysiloxane Stationary Phases in Gas Chromatography', *J. Chromatogr.*, **300**, 1–77 (1984).
70. B.W. Wright, R.D. Smith, 'Investigation of Polar Modifiers in Carbon-dioxide Mobile Phases for Capillary Supercritical Fluid Chromatography', *J. Chromatogr.*, **355**, 367–373 (1986).
71. S.M. Fields, K.E. Markides, M.L. Lee, 'Retention in Capillary Supercritical Fluid Chromatography with Carbon-dioxide Mobile Phases Containing Polar Modifiers', *J. Chromatogr.*, **406**, 223–235 (1987).
72. Ph. Morin, M. Caude, R.J. Rosset, 'Supercritical and Subcritical Fluid Chromatography Fourier-transform Infrared Spectrometry of Medium Polar Compounds on Polymeric Styrene Divinylbenzene Packed-columns', *J. Chromatogr.*, **407**, 87–108 (1987).
73. C.R. Yonker, D.G. McMinn, B.W. Wright, R.D. Smith, 'Effect of Temperature and Modifier Concentration on Retention in Supercritical Fluid Chromatography', *J. Chromatogr.*, **396**, 19–29 (1987).
74. A.L. Blilie, T. Greibrokk, 'Modifier Effects on Retention and Peak Shape in Supercritical Fluid Chromatography', *Anal. Chem.*, **57**, 2239–2242 (1985).
75. D.E. Stogryn, A.P. Stogryn, 'Molecular Multipole Moments', *Mol. Phys.*, **11**, 371–393 (1966).
76. B.W. Wright, H.T. Kalinoski, R.D. Smith, 'Investigation of Retention and Selectivity Effects Using Various Mobile Phases in Capillary Supercritical Fluid Chromatography', *Anal. Chem.*, **57**, 2823–2829 (1985).
77. S.R. Springston, M. Novotny, 'Mobile-phase Solute Mass Transfer in Supercritical Fluid Chromatography', *Anal. Chem.*, **56**, 1762–1766 (1984).
78. D. Leyendecker, D. Leyendecker, F.P. Schmitz, E. Klesper, 'Comparison of Eluents in Supercritical Fluid Chromatography', *J. Liq. Chromatogr.*, **10**, 1917–1947 (1987).
79. D.E. Raynie, 'Warning Concerning the Use of Nitrous-oxide in Supercritical-fluid Extraction', *Anal. Chem.*, **65**, 3127–3128 (1993).
80. T.A. Berger, 'Separation of Polar Solutes by Packed Column Supercritical Fluid Chromatography', *J. Chromatogr. A*, **785**, 3–33 (1997).
81. H.E. Schwartz, R.G. Brownlee, 'Hydrocarbon Group Analysis of Gasolines with Microbore Supercritical Fluid Chromatography and Flame Ionization Detection', *J. Chromatogr.*, **353**, 77–93 (1986).
82. D.R. Lide (ed.), *Handbook of Chemistry and Physics*, 72nd edition, CRC Press, Boca Raton, FL, 1991–1992.
83. A.R. Ravishankara, A.A. Turnipseed, N.R. Jenson, S. Barone, M. Mills, C.J. Howard, S. Solomon, 'Do Hydrofluorocarbons Destroy Stratospheric Ozone?', *Science*, **263**, 71–75 (1994).
84. J.A. Blackwell, L.E. Schallinger, 'Functional Group Selectivity Using Fluoroform as a Mobile Phase for Supercritical Fluid Chromatography', *J. Microcolumn Sep.*, **6**, 475–482 (1994).
85. J.A. Blackwell, L.E. Schallinger, 'Hydrofluorocarbon and Perfluorocarbon Mobile Phases for Capillary Supercritical Fluid Chromatography', *J. Microcolumn. Sep.*, **6**, 551–556 (1994).
86. J.C. Giddings, M.N. Myers, L. Mclaran, R.A. Keller, 'High Pressure Gas Chromatography of Nonvolatile Species', *Science*, **162**, 67–73 (1968).
87. J.C. Kuei, K.E. Markides, M.L. Lee, 'Supercritical Ammonia as Mobile Phase in Capillary Chromatography', *J. High Resolut. Chromatogr. Chromatogr. Commun.*, **10**, 257–262 (1987).
88. T.A. Berger, J.F. Deye, 'Use of Solvatochromic Dyes to Correlate Mobile Phase Solvent Strength to Chromatographic Retention', in *Supercritical Fluid Technology*, eds. F.V. Bright, M.E.P. McNally, ACS Symposium Series 488, American Chemical Society, Washington, DC, 132–142, 1992.
89. J.M. Levy, W.M. Ritchey, 'Investigations of the Use of Modifiers in Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **24**, 242–248 (1986).
90. P.J. Schoenmakers, F.C.C.J.G. Verhoeven, H.M. van der Bogaert, 'Application of Supercritical Fluid Chromatography to the Analysis of Liquid Crystal Mixtures', *J. Chromatogr.*, **371**, 121–134 (1986).
91. T.A. Berger, J.F. Deye, 'Role of Additives in Packed-column Supercritical Fluid Chromatography – Suppression of Solute Ionization', *J. Chromatogr.*, **547**, 377–392 (1991).

92. W. Steuer, J. Baumann, F. Erni, 'Separation of Ionic Drug Substances by Supercritical Fluid Chromatography', *J. Chromatogr.*, **500**, 469–479 (1990).
93. T.A. Berger, J.F. Deye, 'Effect of Basic Additives on Peak Shapes of Strong Bases Separated by Packed-column Supercritical Fluid Chromatography', *J. Chromatogr. Sci.*, **29**, 310–317 (1991).
94. A. Giorgetti, N. Pericles, H.M. Widmer, K. Anton, P. Datwyler, 'Mixed Mobile Phases and Pressure Programming in Packed and Capillary Column Supercritical Fluid Chromatography – a Unified Approach', *J. Chromatogr. Sci.*, **27**, 318–324 (1989).
95. SFsolver software, ISCO, Lincoln, NE.
96. S.T. Lee, T.S. Reighard, S.V. Olesik, 'Phase-diagram Studies of Methanol–H_2O–CO_2 and Acetonitrile–H_2O–CO_2 Mixtures', *Fluid Phase Equilibria*, **122**, 233–241 (1996).
97. T.S. Reighard, S.T. Lee, S.V. Olesik, 'Determination of Methanol–CO_2 and Acetonitrile–CO_2 Vapor–Liquid Phase-equilibria Using a Variable-volume View Cell', *Fluid Phase Equilibria*, **123**, 215–230 (1996).
98. J. Zhao, S.V. Olesik, 'Phase-diagram Studies of Methanol–CHF_3 and Methanol–H_2O–CHF_3 Mixtures', *Fluid Phase Equilibria*, **154**, 261–284 (1999).
99. H. Yuan, S.V. Olesik, 'Comparison of Enhanced-fluidity and Elevated Temperature Mobile Phases for High-performance Size-exclusion Chromatography', *J. Chromatogr. A*, **785**, 35–48 (1997).
100. J. Via, L.T. Taylor, F.K. Schweighardt, 'Experimental Determination of Changes in Methanol Modifier Concentration in Premixed Carbon Dioxide Cylinders', *Anal. Chem.*, **66**, 1459–1461 (1994).
101. H. Yuan, S.V. Olesik, 'Comparison of Reversed-phase HPLC Separation Using Carbon-dioxide and Fluoroform for Enhanced-fluidity Liquid Mobile Phases', *Anal. Chem.*, **70**, 1595–1603 (1998).
102. Q. Sun, Olesik, 'Chiral Separations Performed by Enhanced-fluidity Liquid Chromatography on a Macrocyclic Antibiotic Chiral Stationary Phase', *Anal. Chem.*, **71**, 2139–2145 (1999).
103. S.T. Lee, S.V. Olesik, 'Normal-phase High-performance Liquid Chromatography Using Enhanced-fluidity Liquid Mobile Phases', *J. Chromatogr. A*, **707**, 217–224 (1995).
104. H. Yuan, I. Souvignet, S.V. Olesik, 'High Performance Size Exclusion Chromatography Using Enhanced-fluidity Liquids', *J. Chromatogr. Sci.*, **35**, 409–416 (1997).
105. E. Klesper, A.H. Cirwin, D.A. Turner, 'High Pressure Gas Chromatography Above Critical Temperatures', *J. Org. Chem.*, **27**, 700–701 (1960).
106. K. Anton, C. Berger (eds.), *Supercritical Fluid Chromatography with Packed Columns: Techniques and Applications*, Marcel Dekker, Inc., New York, 1998.

Thin-layer Chromatography

Joseph Sherma
Lafayette College, Easton, USA

1 Introduction 1
2 Sample Preparation 2
3 Stationary Phases 3
4 Mobile Phases 3
5 Application of Samples and Standards 4
5.1 Sample Application as Spots 4
5.2 Sample Application as Bands 4
5.3 Automated Sample Application 4
6 Chromatogram Development 4
6.1 Capillary Flow Thin-layer Chromatography 5
6.2 Forced Flow Planar Chromatography 7
7 Zone Detection 7
8 Documentation of Chromatograms 8
9 Zone Identification 8
10 Quantitative Analysis 8
10.1 Nondensitometric Methods 8
10.2 Densitometric Evaluation 9
11 Special Techniques 10
11.1 Thin-layer Chromatography Combined with Spectrometric Methods 10
11.2 Direct Coupling of Thin-layer Chromatography and High-performance Liquid Chromatography 11
11.3 Thin-layer Radiochromatography 11
11.4 Rod Thin-layer Chromatography with Flame Ionization Detection 12
12 Applications of Thin-layer Chromatography 12
Abbreviations and Acronyms 12
Related Articles 13
Further Reading 13

In thin-layer chromatography (TLC), the sample solution is applied as a spot or band on the origin of the layer spread on a support (the plate). After evaporation of the sample solvent, the plate is placed in a sealed chamber containing a solvent chosen as the mobile phase. Development occurs as the mobile phase moves through the layer, and the components of the sample move at different rates to create the separation. The plate is removed from the chamber, and the separated zones are detected by physical or chemical methods, identified by comparison of their migration to standard zones on the same plate, and quantified by visual or instrumental methods based on measurement of zone sizes and intensities. A great variety of compounds can be separated, including organic, inorganic, biological, and chiral. Environmental, pharmaceutical, biomedical, and food samples are among the sample types commonly analyzed by TLC. Advantages of TLC include rapid analysis time because many samples can be analyzed simultaneously, low solvent usage on a per-sample basis, a high degree of accuracy and precision for instrumental TLC, and sensitivity in the nanogram or picogram range. Coupling of TLC with other analytical methods such as high-performance liquid chromatography (HPLC), mass spectrometry (MS), and Fourier transform infrared spectrometry (FTIR) provides enhanced opportunities for sample analysis. Disadvantages of TLC include application to only nonvolatile compounds, limited resolution capability (separation numbers or peak capacities of 10–50), and the absence of fully automated systems, although the individual steps of the technique can be automated.

1 INTRODUCTION

TLC is a type of liquid chromatography in which the stationary phase takes the form of a layer on a glass, aluminum, or plastic support. The term ''planar chromatography'' is often used for both TLC and paper chromatography (PC) because each employs a planar stationary phase rather than a column. PC, which utilizes plain, modified, or impregnated paper (cellulose) as the stationary phase, involves many of the same basic techniques as TLC, but it has not evolved into an efficient, sensitive, quantitative, instrument-based analytical method and has many disadvantages relative to TLC. PC will not be covered in this article.

Originally developed in 1951 by J.G. Kirchner and colleagues, later standardized by E. Stahl and colleagues, and still widely practiced today, classical, capillary-action TLC is an inexpensive, simple method requiring little instrumentation, which is used for separation of simple mixtures and for qualitative identification or semiquantitative, visual analysis of samples. By contrast, modern TLC (usually termed high-performance thin-layer chromatography, HPTLC), which began around 1975 with the introduction by Merck of high-efficiency, commercially precoated plates, is an instrumental technique carried out on efficient, fine-particle layers. HPTLC is capable of producing fast, high-resolution separations and quantitative results with accuracy and precision rivaling those of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

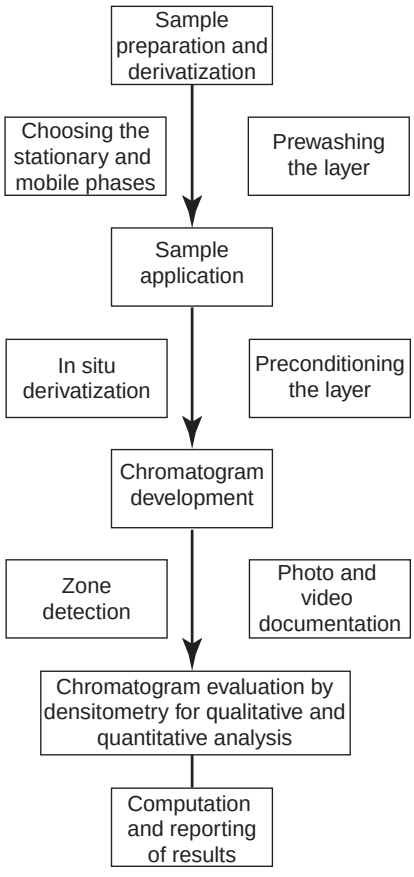

Figure 1 Schematic diagram of the steps in a thin-layer chromatographic analysis. (Adapted with permission from a figure supplied by D. Jaenchen, Camag, Muttenz, Switzerland.)

gas chromatography (GC) and HPLC. It also has many advantages over these methods.

TLC is an off-line process in which the various stages (Figure 1) are carried out independently. The advantages of this arrangement, using an open, disposable layer compared with an on-line column process such as HPLC, include the possibility of separating up to 70 samples and standards simultaneously on a single plate, leading to high-throughput, low-cost analyses, and the ability to construct calibration curves from standards chromatographed under the same conditions as the samples; analyzing samples with minimum sample preparation without fear of irreversible contamination of the column or carryover of fractions from one sample to another, which can occur in HPLC with sequential injections; and analyzing a sample by use of multiple separation steps and static postchromatographic detection procedures with various universal and specific visualization reagents, which is possible because all sample components are ''stored'' on the layer without chance of loss. TLC is highly selective and flexible because of the great variety of layers that is available commercially. It has proven to be as sensitive as HPLC in many analyses, and solvent usage per sample is extremely low.

TLC and HPTLC plates are usually developed by capillary flow of the mobile phase without pressure in ascending or horizontal modes, but forced flow methods, in which the mobile phase is driven through the layer by pumping under pressure or by centrifugal force, are also used. In capillary flow TLC, the migration speed of the mobile phase decreases with the square of the solvent migration distance.

Limited separation efficiency is a disadvantage of TLC. The maximum number of theoretical plates (N) for HPTLC is around 5000, compared with around 15 000 for HPLC, while the separation number (the number of spots that can be separated over the distance of the run with a resolution of unity) is around 15 in TLC compared with 200 in HPLC. This limited efficiency is a result of the capillary flow mechanism over a restricted migration distance; under forced flow conditions, TLC separation efficiency is significantly improved. An apparent disadvantage of TLC is that, although many of the individual steps have been automated and on-line coupling with other chromatographic and spectrometric techniques has been achieved, complete automation of TLC has not been realized. However, changing the off-line nature of TLC to an on-line closed system with complete automation would eliminate many of the advantages stated above. It has been shown that more samples can be processed with a stepwise automated TLC workstation in an 8 h workday than with a fully automated HPLC system in 24 h.

Analytical TLC differs from preparative layer chromatography (PLC) in that larger weights and volumes of samples are applied to thicker (0.5–2 mm) and sometimes larger layers in the latter method, the purpose of which is the isolation of 10–1000 mg of sample for further analysis.

This article presents the techniques and instrumentation of modern TLC and illustrates how it can complement HPLC and GC in a contemporary analytical laboratory. A reading list is presented at the end from which additional, detailed information on the topics covered, as well as many others, can be obtained.

2 SAMPLE PREPARATION

Sample preparation procedures for TLC are similar to those for GC and HPLC. In order for the solution to be spotted, it must be sufficiently concentrated that the analyte can be detected in the applied volume, and pure enough so that it can be separated as a discrete, compact spot or zone. The solvent in which the sample is dissolved must be suitable in terms of viscosity, volatility, and ability

to wet the layer, and not have the potential for unintended predevelopment during sample application.

Relatively pure samples or their concentrated extracts can often be directly spotted for TLC analysis. If the analyte is present in low concentration in a complex sample, solvent extraction, clean-up (purification), and concentration procedures must precede TLC. Because layers are not reused, it is often possible to apply cruder samples than could be injected into a GC or HPLC column, including samples with irreversibly sorbed impurities. However, impurities that co-migrate with the analyte, adversely affect its detection, or distort its zone (i.e. cause streaking or trailing) must be removed prior to TLC. Carryover of material from one sample to another is not a problem as it is in on-line column methods involving sequential injections, so sample preparation is often simpler for TLC than for these other chromatographic methods.

Common clean-up procedures include liquid–liquid extraction, column chromatography, desalting, and deproteinization. Solid-phase extraction (SPE) using small, disposable columns or membrane disks has recently become widely used for isolation and clean-up of samples prior to TLC analysis. Supercritical fluid extraction (SFE) has been directly coupled to TLC for analysis of solid samples or solutions loaded on glass fiber filters. A derivative of the analyte can be formed in solution prior to spotting, or in situ at the origin by overspotting of a reagent, in order to improve resolution or detection. Special plates with preadsorbent zones serve for sample clean-up by retaining some interfering substances.

3 STATIONARY PHASES

TLC and HPTLC plates are commercially available in the form of precoated layers supported on glass, plastic sheets, or aluminum foil. HPTLC plates are smaller (10×10 cm or 10×20 cm), have a thinner (0.1–0.2 mm) more uniform layer composed of smaller-diameter particles (5–6 μm), and are developed over shorter distances (3–7 cm) compared with classical 20×20 cm TLC plates, which have a 0.25-mm-thick layer of 12–20 μm particle size and are developed for 10–12 cm. Optimal development distances are the point beyond which increased resolution of zones is offset by diffusion effects. Compared with TLC, HPTLC provides better separation efficiency and lower detection limits.

The choice of the layer and mobile phase is made in relation to the nature of the sample. Normal- or straight-phase adsorption TLC on silica gel with a less polar mobile phase, such as chloroform–methanol, is the most widely used mode. Lipophilic C-18, C-8, and C-2 bonded silica gel phases with a polar aqueous mobile phase, such as methanol–water, are used for reversed-phase TLC. Other precoated layers include alumina, magnesium silicate (Florisil), polyamide, cellulose, ion exchangers, and chemically bonded phenyl, amino, cyano, and diol layers. The last three bonded phases can function with multimodal mechanisms, depending on the composition of the mobile phase. Silica gel can be impregnated with various solvents, buffers, and selective reagents to improve separations. Chiral plates composed of a reversed-phase layer impregnated with copper acetate and a chiral selector, (2*S*,4*R*,2′*RS*)-4-hydroxy-1-(2-hydroxydodecyl)proline, can be used to separate enantiomers through a ligand exchange mechanism. Preparative layers are thicker than analytical layers to provide higher sample capacity.

Layers are cleaned by predevelopment with or immersion in methanol prior to sample application, especially for quantitative and preparative applications. The most important manufacturers of TLC plates are Merck, Whatman, Analtech, and Macherey-Nagel, and literature from these companies should be consulted for details of availability, properties, usage, and applications.

4 MOBILE PHASES

Unlike GC, in which the mobile phase (carrier gas) is not a factor in the selectivity of the chromatographic system, the mobile phase in liquid chromatography, including TLC, exerts a decisive influence on the separation. In HPLC, the analyte passes through the on-line detector in the presence of the mobile phase, so solvents must be chosen not only to provide the required resolution but also to not absorb at the ultraviolet (UV) detection wavelength. In TLC the mobile phase is removed (evaporated) before the zones are detected, therefore a wider variety of solvents can be used to prepare mobile phases compared with HPLC.

The mobile phase is usually a mixture of two to five different solvents selected empirically using trial and error guided by prior personal experience and literature reports of similar separations. In addition, various systematic mobile phase optimization schemes involving solvent classification (selectivity) and the elutropic series (strength) patterned after HPLC have been described, most notably the PRISMA model based on Snyder's solvent classification system as developed by Nyiredy. For normal phase (silica gel), the following 10 solvents from Snyder's eight selectivity groups are used for exploratory TLC of the mixture: diethyl ether (group I), isopropanol and ethanol (II), tetrahydrofuran (III), acetic acid (IV), dichloromethane (V), ethyl acetate and dioxane (VI),

toluene (VII), and chloroform (VIII). Hexane (solvent strength = 0) is used to adjust the R_f values within the optimum range (0.2–0.8), if necessary. Between two and five solvents are then selected for construction of the PRISMA model that leads to identification of the optimized mobile phase. A similar procedure is followed for reversed-phase TLC (e.g. C_{18} bonded phase layer) using mixtures of methanol, acetonitrile, and/or tetrahydrofuran with water (solvent strength = 0).

TLC is usually carried out with a single mobile phase, rather than a mobile phase gradient which is often used in HPLC. Equilibration between the mobile phase and the layer occurs gradually during TLC development, and the mobile phase composition can change because different constituents migrate through the layer at different rates (solvent demixing). This leads to solvent gradients along the layer during ''isocratic'' TLC, the formation process of which is very different from the intentional, well-controlled gradients in a fully equilibrated HPLC column. Mobile phase selection for automated multiple development (AMD) will be described below in section 5.

5 APPLICATION OF SAMPLES AND STANDARDS

Application of small, precisely positioned initial zones from sample and standard solutions without damaging the layer surface is critical for achieving maximum resolution and reliable quantitative analysis. The volumes applied and the method of application depend on the type of analysis to be carried out (qualitative or quantitative), the layer (TLC or HPTLC), and the detection limit.

5.1 Sample Application as Spots

For TLC, 0.5- to 5-μL volumes are usually applied manually with a micropipet to produce initial zones with diameters in the 3–4 mm range. For HPTLC, initial zones in the form of spots with a maximum diameter of 1–2 mm are applied from a 100-nL to 1-μL fixed-volume, self-loading glass capillary pipet, which is held in a rocker-type mechanized spotting device (the Nanomat) that mechanically controls its positioning and brings the capillary tip into gentle and uniform contact with the layer to discharge the solution without damaging the layer. The volume precision of this applicator is below 1% (relative standard deviation, RSD). Variable volumes can be applied with a micrometer-controlled syringe.

5.2 Sample Application as Bands

Sample application in the form of bands is advantageous for high-resolution separations and precise (1% RSD) quantitative scanning densitometry using the aliquot technique (scanning with a slit of one-half to two-thirds the length of the applied band). Narrow, homogeneous sample bands of controlled length [1 mm (spot) to 190 mm] can be applied by use of a spray-on device (the Linomat) in which the plate is mechanically moved right to left in the x-direction beneath a fixed syringe from which 2–100 μL of sample (5–490 μL for preparative separations) is sprayed by an atomizer operating with a controlled nitrogen gas pressure. The bands are exactly positioned according to a coordinate system, which facilitates automated scanning after chromatogram development, or overspraying of samples with a reagent for in situ prechromatographic derivatization or with spiking solutions for validation of quantitative analysis by the standard-addition method. The ability to apply larger volumes to an HPTLC plate without loss of resolution lowers the determination limits with respect to the concentration of the solution, which aids in trace analysis. Complex, impure samples can often be successfully quantified, but only if bands are applied rather than spots.

In addition to band formation by instrumental means, sharp bands are also produced when samples are applied manually with a micropipet as diffuse vertical streaks to special plates containing a preadsorbent or concentrating zone. These plates, as well as the Linomat, facilitate quantitative analysis by allowing different volumes of the same standard solution to be applied to produce the densitometric calibration curve, rather than the same volume of a series of standards when spots are applied.

5.3 Automated Sample Application

Figure 2 shows the Automatic TLC Sampler III, which is a fully automated, computer-controlled device for sequential application of samples and standards from a rack of vials through a steel capillary as spots by contact transfer or as bands by a spray-on technique. The speed, volume, and pattern of standard and sample application are controllable, and a programmable rinse cycle can eliminate cross-contamination. Analyses performed with this applicator combined with densitometric chromatogram evaluation controlled by the same computer conform to good laboratory practice (GLP) and good manufacturing practice (GMP) standards.

6 CHROMATOGRAM DEVELOPMENT

Development times are typically in the range of 3–60 min, depending on the layer, mobile phase, and development method chosen. However, development time does not significantly influence the overall analysis time per

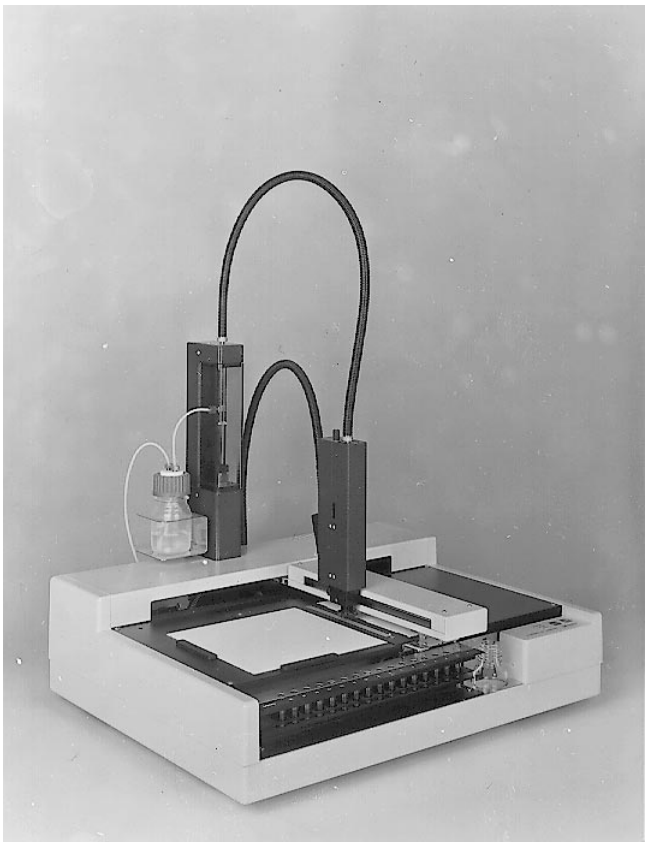

Figure 2 Automatic TLC Sampler. (Photograph supplied by D. Jaenchen, Camag, Muttenz, Switzerland.)

sample because many samples and standards can be chromatographed simultaneously.

6.1 Capillary Flow Thin-layer Chromatography

6.1.1 Ascending Development

The results of TLC are strongly dependent upon the environmental conditions during development, such as small changes in mobile phase composition, temperature, humidity, and the size and type of the chamber and its solvent vapor saturation conditions.

In the classical method of linear, ascending development TLC and HPTLC, the developing solvent is contained in a large-volume, covered glass tank (N-tank). The spotted plate is inclined against an inside wall of the tank with its lower edge immersed in the developing solvent below the starting line, and the solvent begins to rise immediately through the initial zones because of the capillary flow. As the mobile phase ascends, the layer interacts with the vapor phase in the tank. The space inside the tank is more or less equilibrated with solvent vapors, depending on the presence or absence of a mobile phase-soaked paper liner and the period of time the tank is allowed to stand before the plate is inserted. The use of a paper liner to provide saturation of the tank atmosphere with solvent vapors can be advantageous or detrimental, depending on the composition of the mobile phase.

More reproducible conditions can be established in an S-chamber (sandwich chamber), which consists of the TLC plate, a spacer of 3 mm thickness, and a cover- or counter-plate that is either blank glass or a solvent-soaked TLC plate. These parts are clamped together, so that the bottom 2 cm of the layer is uncovered, and are placed in a trough containing the mobile phase. Interaction between the layer, dry or wetted, and the gas phase is largely suppressed in an S-chamber, and reproducibility of the separation is improved.

The twin-trough chamber is an N-chamber modified with an inverted V-shaped ridge on the bottom dividing the tank into two sections, which allow development with only 4–20 mL of solvent, depending on the plate size, on one side, and easy pre-equilibration of the layer with vapors of the mobile phase or another conditioning liquid (e.g. a sulfuric acid–water mixture to control humidity) or volatile reagent on the other side.

A computer-controlled automatic developing chamber has been designed to provide reproducible isocratic one-dimensional (1-D) ascending development without operator attention. Parameters for preconditioning, N- or S-tank configuration, solvent migration distance, and drying with filtered cold or heated air are programmable via a key pad. The developing process is monitored, and elapsed time and running distance are continually displayed. Ten complete developing programs can be stored in computer memory.

6.1.2 Horizontal Development

The horizontal developing chamber permits simultaneous development from opposite edges to the middle of 60 sample spots on a 20×10 cm HPTLC plate, or 30 samples from one end to the other. The developing solvent, held in narrow troughs, is carried to the layer through capillary slits formed between the trough walls and glass slides. The chamber is covered with a glass plate during pre-equilibration and development and can be operated in N-type or S-type configurations, including humidity control by placing an appropriate sulfuric acid–water mixture in a conditioning tray. Use of the horizontal developing chamber allows separation conditions to be efficiently standardized, and only low amounts of mobile phase are required.

6.1.3 Continuous Development

The short bed continuous development (SB/CD) chamber is used for continuous development, which leads to

improved resolution of zones with low migration rates because of the larger effective separation distance. Four glass ridges and the back wall of the chamber support the plate at five different angles of inclination, each of which allows increasing lengths of the layer to protrude out of the top of the chamber. Solvent continually evaporates from the external layer at the top and is replaced by additional solvent drawn from the chamber at the bottom.

6.1.4 Gradient Elution Thin-layer Chromatography Combined with Automated Multiple Development

AMD generally involves 10–30 individual linear ascending developments of an HPTLC plate (usually silica gel) carried out in an AMD instrument, which includes an N-type chamber equipped with connections for feeding and releasing mobile phase and pumping a gas phase in and out, storage bottles for pure solvents and waste, a gradient mixer, syringes for measuring solvent volumes, and sensors for measuring migration distances. The developments are performed in the same direction with a stepwise mobile phase gradient that becomes progressively weaker (i.e. less polar) over distances that increase by 1–5 mm for each stage. The solvent is completely removed from the chamber and the layer is dried under vacuum applied by a pump for a preselected time after each development. The layer is then preconditioned with the vapor phase of the next batch of fresh solvent, which is fed into the chamber before the next incremental run.

The solvent strength may be changed for each development, or several stages may be carried out with the same solvent before changing its strength. The repeated movement of the solvent front through the chromatographic zones causes them to become compressed into narrow bands (width about 1 mm) during AMD, leading to peak capacities of more than 50 over a separation distance of 80 mm. Typical ''universal gradients'' for AMD are produced from methanol or acetonitrile (polar), dichloromethane, diisopropyl ether or *tert*-butylmethyl ether (medium polarity) and hexane (nonpolar). The central or basis solvent and the nonpolar solvent have the greatest effect on selectivity. By superimposing the densitogram of a chromatogram with a matched-scale diagram of the gradient, required modifications of the solvent system can be predicted.

Complex mixtures containing compounds with widely different polarities can be separated by AMD on one chromatogram, in which sharply focused zones migrate different distances according to their polarities. Zone widths are independent of migration distance and size of the starting zone, leading to high resolution and the ability to resolve relatively large samples for trace analysis. Migration distances of individual components are largely independent of the sample matrix, and detection limits are improved because of the highly concentrated zones that are produced.

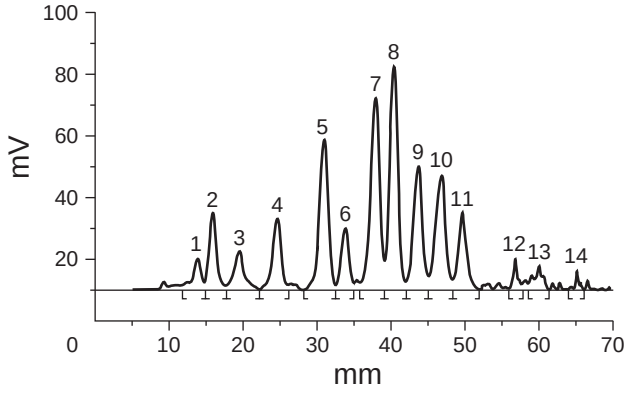

Figure 3 Densitogram of 11-component mixture of saccharides (230–500 ng) separated by AMD on a phosphate-buffered amino-bonded silica gel layer using an acetonitrile–acetone–water gradient (30:30:40 to 20:15:65 v/v/v) and 15 development steps of increasing length. 1, maltohexose; 2, maltopentose; 3, maltotetrose; 4, maltotriose; 5, maltose; 6, sucrose; 7, glucose; 8, fructose; 9, xylose; 10, rhamnose; 11, desoxyribose. (Photograph supplied by D. Jaenchen, Camag, Muttenz, Switzerland.)

Figure 3 illustrates a densitometer scan of a chromatogram produced by gradient AMD.

6.1.5 Two-dimensional Development

Two-dimensional (2-D) TLC involves spotting the sample in one corner of the layer, developing (ascending or horizontal) with the first mobile phase, drying the plate, and developing at a 90° angle with a second mobile phase having a diverse, complementary separation mechanism (selectivity). Computer simulation has been used to optimize 2-D separations based on 1-D data. Resolution (spot capacity) in 2-D TLC is greatly improved compared with 1-D TLC because sample components are resolved over the entire area of the layer, and is usually superior to that obtained by HPLC. Resolution by 2-D forced flow planar chromatography (FFPC) (see below) exceeds 2-D capillary flow TLC because spot diffusion is smaller. Disadvantages of 2-D TLC include a limit of one sample per plate and the time required for two developments and intermediate drying. In addition, quantitative calibration or qualitative identification standards cannot be developed in parallel on the same plate at the same time. Improved video densitometers may allow more reliable quantitative 2-D TLC in the future than is possible today with the commonly used slit-scanning densitometers.

6.2 Forced Flow Planar Chromatography

The mobile phase can migrate through the layer by capillary action, as is the case in the procedures described above, or under the influence of forced flow. FFPC has theoretical advantages relative to capillary flow, including independent optimization of mobile-phase velocity, higher efficiency, lower separation time, and use of solvents that do not wet the layer, but it requires specialized, complex commercial instrumentation. Forced flow is produced by mechanically pumping solvent through a sealed layer (overpressured layer chromatography, OPLC) or by spinning a glass rotor covered with sorbent around a central axis to drive the solvent from the center to the periphery of the layer by centrifugal force (rotation planar chromatography, RPC). Separations can be accomplished with a dry layer (off-line FFPC), but the closed system arrangement also allows the separation to be started after the layer is equilibrated with the mobile phase, similar to the situation in HPLC (on-line FFPC).

In RPC, samples are applied to the rotating stationary phase near the center, and centrifugal force plus capillary action drives the mobile phase through the sorbent from the center to the periphery of the plate. Up to 72 samples can be applied for analytical separations, and in situ quantification is possible. One circular sample is applied for micropreparative and preparative separations, which can be carried out off- and on-line. Various chambers are used for RPC, which differ mainly in the volume of the vapor space. The major commercial instruments for RPC are the Chromatotron 7924, CLC-5, and Rotachrom Model P. Although analytical applications have been proposed, RPC appears to be most useful for preparative applications.

OPLC combines many advantages of classical TLC and HPLC. Two commercial development chambers available for OPLC are the Chrompres 10 and Chrompres 25, which operate at pressures up to 10 and 25 atm respectively. The higher pressure of the Chrompres 25 allows the use of more viscous mobile phases and/or higher mobile phase velocities in the linear development mode. The chambers consist of a bottom support block and an upper block with a poly(methyl methacrylate) support plate, fixed in an external frame. The plate is placed on the surface of the bottom block, and the chamber is closed. A pump produces a water cushion against a flexible plastic membrane, which covers the layer and eliminates any vapor space, resulting in more reproducible separations under controlled conditions. A second pump forces mobile phase through the layer at a constant flow rate in the range of $1-12\,\mathrm{mL\,min^{-1}}$, leading to higher separation efficiency than is possible with capillary flow. A mobile-phase outlet is located in the upper block. To carry out linear chromatography, the layer must be specially prepared by scraping the edges and treating with a polymer sealant to eliminate leaks during development, and by cutting mobile-phase inlet and solvent outlet channels at appropriate positions. The Chrompres instruments allow linear development distances of 18 or 36 cm on a 20×40 cm TLC plate in the on-line mode, plus off-line circular development, for which no plate preparation is required for a maximum separation distance of 9 cm. Special plate preparation is required for circular or anticircular development over longer distances. Linear and circular OPLC are suitable for development of several plates simultaneously (parallel connected OPLC), or serial combination of plates is possible (long-distance OPLC). Different types of layers can be employed for a single analysis in these multilayer OPLC methods. The Personal OPLC BS 50 is a newer commercial OPLC instrument. Although OPLC has been the most used FFPC method, its reported practical applications have been quite limited.

High-pressure planar liquid chromatography (HPPLC) is a radial OPLC method that has been described but not widely used because of the unavailability of a commercial instrument. A single sample is applied to the center of the layer, and the mobile phase is delivered to the same point using up to 200 bar external pressure. In-flow sample injection is also possible. The method allows transfer of capacity factor (k') data to HPLC as well as coupling of HPLC and planar chromatography.

7 ZONE DETECTION

After development with the mobile phase, the plate is dried in a fumehood and heated to completely evaporate the mobile phase. Separated compounds are detected on the layer by their natural color, natural fluorescence, quenching of fluorescence, or as colored, UV-absorbing, or fluorescent zones after reaction with an appropriate reagent (post-chromatographic derivatization). Although dependent upon the particular analyte and the detection method chosen, sensitivity values are generally in the nanogram range for absorbance and picogram range for fluorescence.

Layers are frequently heated after applying the detection reagent in order to accelerate the reaction upon which detection is based. Heating is carried out with a hair drier in a fumehood, in an oven, or with a TLC plate heater. The plate heater, which contains a 20×20 cm flat, even heating area, a grid to facilitate proper positioning of TLC and HPTLC plates, programmable temperature between 25 and 200 °C, and digital display of the actual temperature, provides the most consistent heating conditions.

Compounds that are naturally colored are viewed directly on the layer in daylight, while compounds with

native fluorescence are viewed as bright zones on a dark background under UV light. Viewing cabinets incorporating shortwave (254 nm) and longwave (366 nm) UV lamps are available for inspecting chromatograms in an undarkened room.

Compounds that absorb around 254 nm, particularly those with aromatic rings and conjugated double bonds, can be detected on an ''F-layer'' containing a phosphor or fluorescent indicator (often zinc silicate). When irradiated with 254 nm UV light, absorbing compounds diminish (quench) the uniform layer fluorescence and are detected as dark violet spots on a bright background.

Universal or selective chromogenic and fluorogenic liquid detection reagents are applied by spraying or dipping the layer. Various types of aerosol sprayers and dip tanks are available for manual operation. For safety purposes, spraying is carried out inside a laboratory fumehood or commercial TLC spray cabinet with a blower (fan) and exhaust hose. Dipping is usually the best method for reproducible results, and the most uniform dip application of reagents can be achieved by using a battery-operated chromatogram immersion instrument, which provides selectable, consistent vertical immersion and withdrawal speeds of 30–50 mm s^{-1} and immersion times of 1–8 s for plates with 10 or 20 cm heights. This mechanized dipping device can also be used for prewashing TLC plates prior to initial zone application, impregnation of layers with reagents that improve resolution or detection prior to initial zone application and development, and for postdevelopment impregnation of chromatograms containing fluorescent zones with a fluorescence enhancer and stabilizer, such as paraffin. A few detection reagents (HCl, sulfuryl chloride, iodine) can be transferred uniformly to the layer as vapors in a closed chamber.

8 DOCUMENTATION OF CHROMATOGRAMS

TLC separations are best documented by photography or video recording. Commercial systems for photodocumentation contain instant and conventional film cameras and associated lighting accessories for photography of colored, fluorescent, and fluorescence-quenched zones on TLC plates in shortwave, midrange, and longwave UV light and in visible light. A commercial video documentation system includes visible and UV lighting, a video or digital camera with zoom lens, monitor, and video printer. The system can also be extended to include data storage on a personal computer. A cabinet is available to house the photographic and video instruments for use in an undarkened room.

The latest technology for documentation of TLC plates involves computer scanning of the spots on the plate. The equipment required includes a computer, scanner, and monochrome or color printer. Computer scanning can be used only for visible spots, and not for those that are fluorescent or quench fluorescence.

9 ZONE IDENTIFICATION

The identity of TLC zones is obtained in the first instance by comparison of R_f values between samples and reference standards chromatographed on the same plate, where R_f equals the migration distance of the center of the zone divided by the migration distance of the mobile-phase front. Identity is more certain if a selective chromogenic reagent yields the same characteristic color for sample and standard zones, or if an R_f match is obtained in at least two TLC systems with diverse mechanisms, e.g. silica gel normal phase and C-18 bonded silica gel reversed-phase. Comparison of standard and sample in situ ultraviolet/visible (UV/VIS) absorption or fluorescence emission spectra, obtained by using the spectral mode of a densitometer, can also aid identification, but these spectra often contain inadequate structural information for complex mixtures. Identity can be confirmed by application of combined TLC-spectrometry methods described in section 11.1.

10 QUANTITATIVE ANALYSIS

10.1 Nondensitometric Methods

Quantification of thin-layer chromatograms can be performed after manually scraping off the separated zones of samples and standards and elution of the substances from the layer material with a strong, volatile solvent. The eluates are concentrated and analyzed by use of spectrometry, GC, HPLC, or some other sensitive microanalytical method. This method of quantification is laborious and time-consuming, and the difficulty in recovering samples and standards uniformly is a major source of error. Although its importance has declined relative to densitometry, the indirect scraping and elution quantification method is still widely used, e.g. for some drug assays according to the US Pharmacopoeia.

Direct TLC semiquantitative analysis can be performed by visual comparison of sample spot intensities with the intensities of reference spots developed simultaneously on the same layer. For this comparison, the bracketing method is used in which standard spots with concentrations equal to, greater than, and less than the expected

sample concentration are placed on either side of duplicate sample spots. The concentrations of samples and standards should lie within the linear response range of the detection method. The use of TLC for compliance screening of drug products is an important example of this approach.

10.2 Densitometric Evaluation

Most modern HPTLC quantitative analyses are performed by in situ measurement of the absorbance or fluorescence of the separated zones in the chromatogram tracks using an optical densitometic scanner with a fixed sample light beam in the form of a rectangular slit. The length and width of the slit are selectable for optimized scanning of zones with different dimensions. The densitometer measures the difference between the optical signal from a zone-free background area of the plate and from the calibration standards and sample zones. With automated zone application, precision ranging from 1% to 3% RSD is typical for densitometric analyses.

The plate is mounted on a moveable stage controlled by a stepping motor, which allows each chromatogram to be scanned, usually in the direction of development. Single-beam geometry is most often used and gives excellent results when high-quality plates and mobile phase produce compact, well-separated zones. A schematic diagram of a single-beam scanner arranged for measurement of absorption in the reflectance mode is shown in Figure 4. A tungsten–halogen lamp is used as the source for scanning colored zones in the 400–800 nm range (visible absorption) and a deuterium lamp for scanning UV-absorbing zones directly or as quenched zones on F-layers in the 190–400 nm range. The monochromator used with these continuous wavelength sources can be a quartz prism or, more often, a grating. The detector is a photomultiplier or photodiode. For fluorescence scanning, a high-intensity xenon or mercury vapor lamp is used as the source, the optimum excitation wavelength is selected by the monochromator, and

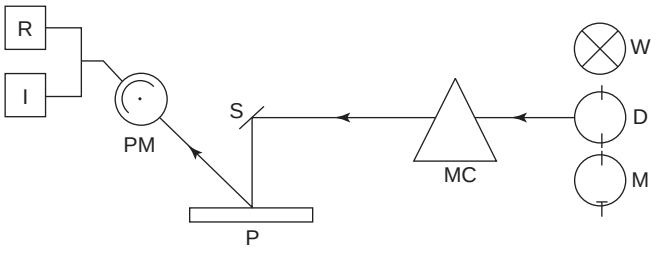

Figure 4 Schematic diagram of the optical path of commercially available densitometers for absorption scanning. R, recorder; I, integrator; PM, photomultiplier detector; P, plate; S, mirror; MC, monochromator; W, tungsten incandescent lamp; D, deuterium lamp; M, mercury lamp.

a cut-off filter is placed between the plate and detector to block the exciting UV radiation and transmit the visible emitted fluorescence. Simultaneous measurement of reflection and transmission, or transmission alone, can be carried out by means of a second detector positioned on the opposite side of the plate. Ratio-recording double-beam scanners, which can correct for background disturbances and drift caused by fluctuations in the source and detector, are equipped with a chopper to alternate the light beam between sample and reference paths before reaching the detector (double beam in time). For dual-wavelength, single-beam scanning, which will correct for matrix scattering of the absorbed light by subtracting out the (presumably equal) scattering at a nonabsorbed wavelength, a light beam is selected by a mirror and passes through two separate monochromators to isolate the two different wavelengths. The two beams are alternated by a chopper and recombined into a single beam representing a difference signal at the detector. Zig-zag scanning with a small spot of light is possible with scanners that have two independent stepping motors to move the plate in the x- and y-axes. Computer algorithms integrate the maximum absorbance measurements from each swing, which corresponds to the length of the slit, to produce a distribution profile of zones having any shape. Disadvantages of scanning with a moving light spot include problems with data processing, lower spatial resolution for HPTLC, and unfavorable error propagation. Some densitometers have the ability to rotate the plate while scanning for measurement of circular and anticircular chromatograms.

Most modern scanners have a computer-controlled motor-driven monochromator that allows automatic recording of in situ absorption and fluorescence excitation spectra. These spectra can aid compound identification by comparison with co-chromatographed standards or stored standard spectra, can test for identity by superimposition of spectra from different zones on a plate, and check zone purity by superimposition of spectra from different areas of a single zone. The spectral maximum determined from the in situ spectrum is usually the optimal wavelength for scanning standard and sample areas for quantitative analysis.

The scanner is connected to a recorder, integrator, or computer. A computer with software designed specifically for TLC is most common for data processing and automated control of the scanning process in modern instruments. With a fully automated system, the computer can carry out the following functions: data acquisition by scanning all plate tracks following a preselected geometric pattern with control of all scanning parameters; automated peak searching and optimization of scanning for each fraction located; multiple wavelength scanning to find, if possible, a common wavelength at which all substances can be quantified or the wavelength of maximum absorbance

of each compound, to optically resolve fractions incompletely separated by TLC, and to identify fractions by comparison with stored data through pattern recognition; baseline location and correction; computation of peak areas and/or heights of samples and co-developed standards and processing of the analog raw data to quantitative digital results, including calculation of calibration curves by linear or polynomial regression, interpolation of sample concentrations, statistical analysis of reproducibility, and presentation of a complete analysis report; and storage of raw data on disk for later reintegration, calibration, and evaluation with different parameters.

Because of light scattering from the sorbent particles, a simple, well-defined mathematical relationship between the amount of analyte and the light signal has not been found. Curves relating absorption signal (peak height or area) and concentration or weight of standards on the layer are usually nonlinear, especially at higher concentrations, and do not pass through the origin. Modern integrators and computer software programs can routinely perform linear or polynomial regression of the calibration data, depending on which is most suitable. Fluorescence calibration curves are generally linear and pass through the origin, and analyses based on fluorescence are more specific and 10–1000 times more sensitive than those employing absorbance. Because of these advantages, compounds that are not naturally fluorescent are derivatized before or after chromatography to allow them to be scanned in the fluorescence mode if an appropriate reagent is available. However, absorbance in the 190–300 nm UV range has mostly been used for densitometric analyses.

Validation procedures for quantitative analysis are in some aspects very similar to those for HPLC and GC, with additional considerations related to procedural aspects specific to TLC. Protocols for validation of TLC results, especially for pharmaceutical analysis, are available in the literature. Some densitometers include automatic instrument validation programs in their software.

Video densitometers (image processors) are an alternative to the optical/mechanical scanners. Video densitometry is based on electronic point-scanning of a stationary plate using an instrument composed of UV and visible light sources, a charge coupled device (CCD) camera with zoom capabilities, and a computer with imaging and evaluation software. Video scanners have advantages including rapid data collection and storage, simple design with virtually no moving parts, easy operation, and the ability to quantify 2-D chromatograms, but they have not yet been shown to have the required capabilities to replace slit scanning densitometers. Current video scanners can function only in the visible range to measure colored, fluorescence quenched, or fluorescent spots. They lack the spectral selectivity and accuracy based on the ability to scan with monochromatic light of selectable wavelength throughout the visible and UV range (190–800 nm) that is inherent in classical densitometry, and they cannot record in situ spectra.

11 SPECIAL TECHNIQUES

11.1 Thin-layer Chromatography Combined with Spectrometric Methods

11.1.1 Mass Spectrometry

The identity of TLC zones can be confirmed by MS analysis. A zone can be scraped out of the layer, eluted from the sorbent, and transferred to the mass spectrometer as a discrete sample. Alternatively, the sample is not separated from the scraped sorbent and both are introduced into the spectrometer source on a direct insertion probe. Desorption methods include volatilization from a heated probe tip, laser ablation, or particle beam sputtering. Ionization techniques that have been used are electron impact (EI) for stable, volatile analytes, while less volatile or unstable analytes can be ionized by chemical ionization (CI), secondary ion mass spectrometry (SIMS), fast atom bombardment (FAB) or liquid secondary ion mass spectrometry (LSIMS) after mixing the sorbent with a suitable liquid matrix (e.g. glycerol). An alternative instrumental approach uses probes or plate scanners specifically designed to measure separated TLC zones. A sophisticated device for 2-D imaging of TLC plates using LSIMS has been described, as well as applications of TLC/tandem MS (TLC/MS/MS) to a variety of compound classes.

11.1.2 Infrared Spectrometry

Zones can also be confirmed by combining TLC with FTIR. Transfer techniques involve solvent extraction of the zone from the layer to an infrared (IR)-transparent pellet or powder (KBr). The most common in situ method is diffuse reflectance infrared detection (DRIFT), which has been used with silica, alumina, cellulose, and reversed-phase layers for identifications at the 1 μg level. Background absorption by the sorbent (3700–3100 and 1650–800 cm^{-1} for silica gel) must be subtracted from the sample spectrum; the resulting difference spectrum of the analyte may not match a solution spectrum in terms of absorption band positions and shapes because of interactions between the compound and sorbent. Sensitivity limits are high (ca. 1–10 μg per spot), but quantification is possible for compounds that do not absorb UV radiation or are not detectable by derivatization reagents.

11.1.3 Raman Spectrometry

In situ normal Raman spectrometry (RS) has a relatively high sensitivity of around 0.5–10 μg for measurement of zones on a TLC plate, but spectra resemble those measured on solids in published collections and are suitable for identification of unknown substances.

Surface enhanced Raman spectrometry (SERS) has sensitivity in the nanogram or picogram range. The layer is treated with a colloidal silver solution, and it is then excited with an Ar-ion or He–Ne laser. Unknown spectra must be compared to a collection of spectra of reference compounds measured on similar layers in the user's laboratory.

11.1.4 X-ray Fluorescence Spectrometry

TLC and X-ray fluorescence spectrometry (XRF) have been combined without an interface to provide direct, nondestructive visualization of elements in zones on the layer. Precoated silica gel and cellulose plates were developed, dried, wrapped in polyethylene film, and taped onto the sample stage of an X-ray analytical microscope equipped with an unfiltered Rh-target X-ray guide tube having an aperture set to analyze a spot around 100 μm in diameter. Longitudinal scanning obtains the elemental image along the length of the chromatogram.

TLC/XRF has been applied to the detection of inorganic metal ions, organometallics (porphyrins), halogens in pesticides and related compounds, and P and S in nucleotides and amino acids, respectively. Organic compounds containing electronegative elements can be selectively detected by individual elemental imaging, such as phenols containing chlorine, bromine, or iodine. Minimum detectable element concentrations ranged from 1 to 200 nmol μL^{-1} spotted (0.13–25 nmol mm^{-2} of layer surface).

11.2 Direct Coupling of Thin-layer Chromatography and High-performance Liquid Chromatography

TLC and HPLC can be combined into a powerful multidimensional method with enhanced separation efficiency because the former is generally used with a normal phase mechanism and samples are separated by distance, while the latter is most often reversed-phase and samples are separated by time. On-line coupling is possible by depositing HPLC column effluent onto the surface of a moving plate. The deposited HPLC fractions serve as the origin for further TLC separation with different selectivity.

The amount of HPLC effluent reaching the plate must be minimized so that overloading does not reduce resolution and the layer material is not washed away. This is done by using effluent-splitting or microbore columns coupled with a solvent elimination interface. The Linomat and Automatic TLC Sampler (section 5) have been modified and connected to the column outlet for spray-on application of HPLC effluent. With the combination of HPLC and TLC/AMD, peak capacities of around 500 can be accomplished.

HPLC and TLC have been coupled with spectrometric methods to achieve powerful detection and identification systems. Methods combined with HPLC/TLC have included FTIR, SERS, and fluorescence spectrometry. TLC has been interfaced on-line with GC, supercritical fluid chromatography (SFC), and high-speed countercurrent chromatography in a manner analogous to HPLC.

11.3 Thin-layer Radiochromatography

Location and quantification of separated radioisotope-labeled substances on a thin layer requires the use of autoradiography, zonal analysis, or direct scanning with a radiation detector.

Autoradiography involves exposure of X-ray or photographic film to emissions from radioisotope zones to produce an image on the film. After exposure and development of the film, the radioisotopes are visible as dark spots, which can be quantified by measurement of their optical densities with a scanning densitometer.

Zonal analysis involves scraping radioactive zones from the plate, placing the sorbent in counting vials, adding scintillation fluid or cocktail to elute the radioactive components, and liquid scintillation counting.

The first radiation detectors for direct measurement of radioactivity on TLC plates were the radioscanners, which utilized a mechanically driven windowless gas-flow Geiger counter. Linear analyzers are superior to radioscanners because the components of one entire track of the plate can be simultaneously measured without mechanical scanning, and the instrument can be programmed to move the detector head automatically to any chosen position. Linear analyzers incorporate a position-sensitive windowless gas-flow proportional counter as the detector. The fill gas (e.g. argon–methane) is ionized when radioactive emissions from the TLC zones enter the detector, producing electrons. The electron pulses are detected electronically and stored in computer memory to provide a digital image of the distribution of radioactivity on the layer. Multiwire proportional counters (Digital Autoradiograph or microchannel array detector) are 2-D detectors based on a measuring principle similar to the linear analyzer, but they can detect all areas of radiation from a 20×20 cm layer simultaneously without moving a detector head. Bioimaging/phosphor imaging analyzers represent the latest and most advantageous technology for measuring radioactive TLC zones. The layer is exposed to a phosphor imaging plate that accumulates and stores irradiating radioactive energy from the zones.

The plate is then inserted into an image-reading unit and scanned with a fine laser beam. Luminescence is emitted in proportion to the intensity of the recorded radiation, collected by a photomultiplier tube, and converted to electrical energy to produce an instrumental readout. Resolution, linear range, and sensitivity are equal to, or better than, those of the other detection methods.

Thin-layer radiochromatography is especially important in drug and pesticide metabolism studies in plants, animals, and humans and in studies of the fate of labeled chemicals in the environment. The methods described above are mainly for qualitative detection, but some degree of quantitative information can be obtained.

11.4 Rod Thin-layer Chromatography with Flame Ionization Detection

Separation on thin, cylindrical quartz rods coated with sintered silica gel (Chromarods) has been combined directly with flame ionization detection (FID) for analysis of compounds lacking a chromophore functional group that can facilitate optical densitometry, such as hydrocarbons and lipids. The sample is applied to the bottom of the rod, and several rods are placed in a support frame, then the frame is placed in a development chamber. After development, solvent is removed from the rods by oven drying, and individual rods are automatically passed through the detector at constant speed. The organic compounds separated on the rods are ionized by the flame of the FID, and the ions generate an electric current proportional to the amount of each compound. The data obtained are processed in a manner similar to GC with the same detector.

12 APPLICATIONS OF THIN-LAYER CHROMATOGRAPHY

TLC can provide rapid, low-cost qualitative analyses and screening in order to obtain information, such as sample stability, purity, and uniformity, and to follow the course of a reaction, while instrumental HPTLC can provide accurate and reproducible (1–3% RSD) quantitative results. Samples that are difficult to prepare can be analyzed readily, and detection is especially flexible in the absence of the mobile phase and with a variety of parameters.

TLC has been applied in virtually all areas of analysis, including chemistry, biochemistry, biology, industry, agriculture, environment, food, pharmaceuticals, clinical studies, natural products, toxicology, forensics, plant science, bacteriology, parasitology, and entomology. Compound types to which TLC has been successfully applied include pharmaceuticals and drugs, neutral and polar lipids, organic acids, carbohydrates, amino acids, peptides, phenols, indoles, purines, steroids, synthetic and natural dyes, vitamins, inorganics, and chiral compounds. High and low (trace) levels of analytes can be detected, identified, and quantified.

Discussion of specific applications of TLC is beyond the scope of this article. The reading list provides sources of information on the analysis of all types of compounds and samples, as well as discussions on theory, techniques, and instrumentation. Especially recommended for information on applications are the Sherma–Fried *Handbook of Thin Layer Chromatography*, the Fried–Sherma *Practical Thin Layer Chromatography* book, the biennial reviews by Sherma in *Analytical Chemistry*, and the Camag "CBS". The *Journal of Planar Chromatography–Modern* TLC, published in six issues per year by the Research Institute for Medicinal Plants (Budakalasz, Hungary) in cooperation with Springer Hungarica, Budapest, exclusively features research papers, short communications, and reviews in the fields of analytical and preparative TLC and HPTLC.

ABBREVIATIONS AND ACRONYMS

AMD	Automated Multiple Development
CCD	Charge Coupled Device
CI	Chemical Ionization
DRIFT	Diffuse Reflectance Infrared Detection
EI	Electron Impact
FAB	Fast Atom Bombardment
FFPC	Forced Flow Planar Chromatography
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared Spectrometry
GC	Gas Chromatography
GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
HPLC	High-performance Liquid Chromatography
HPPLC	High-pressure Planar Liquid Chromatography
HPTLC	High-performance Thin-layer Chromatography
IR	Infrared
LSIMS	Liquid Secondary Ion Mass Spectrometry
MS	Mass Spectrometry
OPLC	Overpressured Layer Chromatography
PC	Paper Chromatography
PLC	Preparative Layer Chromatography
RPC	Rotation Planar Chromatography
RS	Raman Spectrometry
RSD	Relative Standard Deviation
SB/CD	Short Bed Continuous Development
SERS	Surface Enhanced Raman Spectrometry
SFC	Supercritical Fluid Chromatography

SFE	Supercritical Fluid Extraction
SIMS	Secondary Ion Mass Spectrometry
SPE	Solid-phase Extraction
TLC	Thin-layer Chromatography
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
XRF	X-ray Fluorescence Spectrometry
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Planar Chromatography in Clinical Chemistry

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Pharmaceuticals and Drugs **(Volume 8)**
Planar Chromatography in Pharmaceutical Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Infrared Spectroscopy **(Volume 12)**
Theory of Infrared Spectroscopy

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Wavelength-dispersive X-ray Fluorescence Analysis

FURTHER READING

K. Bauer, L. Gros, W. Sauer, *Thin Layer Chromatography – An Introduction*, EM Science, Darmstadt, 1991.

Camag Bibliography Service 'CBS', *Camag Scientific Inc.*, Wilmington, NC (literature abstracts in 38 classifications of TLC have been published twice a year since 1965; available in printed form and on CD-ROM).

T. Cserhati, E. Forgacs, 'Trends in Thin Layer Chromatography: 1997', *J. Chromatogr. Sci.*, **35**, 383–391 (1997).

T. Cserhati, E. Forgacs, 'Hyphenated Techniques in Thin Layer Chromatography', *J. AOAC Int.*, **81**, 329–332 (1998).

G. Esser, D. Klockow, 'Detection of Hydroperoxides in Combustion Aerosols by Supercritical Fluid Extraction Coupled to Thin Layer Chromatography', *Mikrochim. Acta*, **113**, 373–379 (1994).

K. Ferenczi-Fodor, Z. Vegh, Z. Pap-Sziklay, 'Validation of Quantitative Planar Chromatographic Analysis of Drug Substances. 1: Definitions and Practice in TLC', *J. Planar Chromatogr.-Mod. TLC.*, **6**, 198–203 (1993).

B. Fried, J. Sherma (eds.), *Practical Thin Layer Chromatography – A Multidisciplinary Approach*, CRC Press, Boca Raton, FL, 1996.

B. Fried, J. Sherma, *Thin Layer Chromatography – Techniques and Applications*, 4th edition, Marcel Dekker, New York, 1999.

F. Geiss, *Fundamentals of Thin Layer Chromatography*, Dr. Alfred Huethig Verlag, Heidelberg, 1987.

N. Grinberg (ed.), *Modern Thin Layer Chromatography*, Marcel Dekker, New York, 1990.

D. Jaenchen, 'Thin Layer (Planar) Chromatography', in *Handbook of Instrumental Techniques for Analytical Chemistry*, ed. F. Settle, Prentice Hall PTR, Upper Saddle River, NJ, 221–239, 1997.

H. Jork, W. Funk, W. Fischer, H. Wimmer, *Thin Layer Chromatography*, VCH, Weinheim, Vol. 1a, 1990.

H. Jork, W. Funk, W. Fischer, H. Wimmer, *Thin Layer Chromatography*, VCH, Weinheim, Vol. 1b, 1994.

H. Kalasz, L.S. Ettre, M. Bathori, 'Past Accomplishments, Present Status, and Future Challenges of Thin Layer Chromatography', *LC-GC*, **15**, 1044, 1046–1050 (1997).

I. Klebovich, J. Szunyog, I. Hazai, 'TLC-Digital Autoradiography (DAR) for the Analysis of Biological Samples. A Newly Developed Rapid Tool for Studying Drug Metabolism', *J. Planar Chromatogr.-Mod. TLC*, **10**, 399–405 (1997).

O. Klein, T. Clark, 'The Advantages of a New Bio-Imaging Analyzer for Investigation of the Metabolism of ^{14}C-Radiolabelled Pesticides', *J. Planar Chromatogr.-Mod. TLC*, **6**, 368–377 (1993).

R.J. Maxwell, A.R. Lightfield, 'An Improved Method for Efficient Predevelopment of HPTLC Plates', *J. Planar Chromatogr.-Mod. TLC*, **12**, 109–113 (1999).

S. Nyiredy, 'Planar Chromatography', in *Chromatography*, 5th edition, part A, ed. E. Heftmann, Elsevier, New York, 109–150, 1992.

M. Ohnishi-Kameyama, T. Nagata, 'An Application of X-ray Fluorescence Spectrometry to the Study of Thin Layer Chromatography', *Anal. Chem.*, **70**, 1916–1920 (1998).

C.F. Poole, S.K. Poole, 'Thin Layer Chromatography', in *Chromatography Today*, Elsevier, New York, 649–734, 1991.

C.F. Poole, S.K. Poole, 'Instrumental Thin Layer Chromatography', *Anal. Chem.*, **66**, 27A–37A (1994).

C.F. Poole, S.K. Poole, 'Multidimensionality in Planar Chromatography', *J. Chromatogr. A*, **703**, 573–612 (1995).

B. Renger, 'Contemporary Thin Layer Chromatography in Pharmaceutical Quality Control', *J. AOAC Int.*, **81**, 333–340 (1998).

J.K. Rozylo, R. Siembida, A. Jamrozek-Manko, 'A New Simple Method for Documentation of TLC Plates', *J. Planar Chromatogr.-Mod. TLC*, **10**, 225–228 (1997).

J. Sherma, 'Comparison of Thin Layer Chromatography and Liquid Chromatography', *J. AOAC Int.*, **74**, 435–437 (1991).

J. Sherma, 'Modern Thin Layer Chromatographic Pesticide Analysis Using Multiple Development', *J. AOAC Int.*, **75**, 15–17 (1992).

J. Sherma, 'Modern High Performance Thin Layer Chromatography', *J. AOAC Int.*, **77**, 297–306 (1994).

J. Sherma, 'Thin Layer Chromatography in Environmental Analysis', *Reviews in Analytical Chemistry*, **XIV**(2), 75–142 (1995).

J. Sherma, 'Thin Layer Chromatography in Food Analysis', *Food Testing and Analysis*, **2**(2), 39–44 (1996).

J. Sherma, 'Thin Layer Chromatography', in *Encyclopedia of Pharmaceutical Technology*, eds. J. Swarbrick, J.C. Boylan, Marcel Dekker, New York, 81–106, Vol. 15, 1997.

J. Sherma, 'Planar Chromatography', *Anal. Chem.*, **70**, 7–26 (1998). (Previous biennial reviews of TLC by J. Sherma are contained in issue 12 of this journal each even-numbered year dating back to volume 42, 1970.)

J. Sherma, B. Fried (eds.), *Handbook of Thin Layer Chromatography*, 2nd edition, Marcel Dekker, New York, 1996.

G.W. Somsen, W. Morden, I.D. Wilson, 'Planar Chromatography Coupled with Spectroscopic Techniques', *J. Chromatogr. A*, **703**, 613–665 (1995).

G. Szepesi, 'Some Aspects of the Validation of Planar Chromatographic Methods in Pharmaceutical Analysis', *J. Planar Chromatogr.-Mod. TLC.*, **6**, 187–197 (1993).

J. Touchstone, *Practice of Thin Layer Chromatography*, 3rd edition, Wiley-Interscience, New York, 1992.

C. Weins, H.E. Hauck, 'Advances and Developments in Thin Layer Chromatography', *LC-GC*, **14**, 456, 458, 460–461, 463–464 (1996).

Mass Spectrometry: Overview and History

O. David Sparkman
California, USA

1 Introduction 1

2 History 4
2.1 Beginning to the 1940s 7
2.2 The Birth of Organic Mass Spectrometry 10
2.3 Gas Chromatography/Mass Spectrometry – A Tremendous Success 11
2.4 Liquid Chromatography/Mass Spectrometry – A Slow Start – But a Fastcomer in the Backstretch 15
2.5 Desorption/Ionization 17
2.6 Types of Mass (m/z) Analyzers 22
2.7 Inorganic Mass Spectrometry 31

3 The Future of Mass Spectrometry 32

4 Epilog 33

Abbreviations and Acronyms 34

Note 34

Appendix 1: Important Worldwide Web URLs 35

Appendix 2: Bibliography 36
Introductory Books 37
Reference Books 38
Technique-oriented Books 40
Interpretation Books 44
Books of Historical Significance 45
Collections of Mass Spectra in Hard Copy 46
Mass Spectrometry, Gas Chromatography/Mass Spectrometry, and Liquid Chromatography/Mass Spectrometry Journals 47
Monographs 47
Personal Computer Mass Spectrometry Abstract Sources 48
Integrated Spectral Interpretation Books 48
Software 48

Appendix 3: Mass Spectral Interpretation Data 51
Proof of a Molecular Ion Peak – $M^{+\bullet}$ 51

References 53

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Mass spectrometry (MS) is the study of matter through the determination of the abundance and the mass-to-charge ratio (m/z) of ions in the gas phase. The data of MS are mass spectra. The mass spectrum is a plot of relative intensities (which represent the ion abundances) on the ordinate versus the m/z values of the ions on the abscissa. The tools of MS are mass spectrometers. A mass spectrometer consists of a sample-inlet system, a device for getting ions into the gas phase (the ion source), a method of separating the ions according to their m/z values (the m/z or mass analyzer), and a means of detecting the ions after they have been separated (the detector). The ions are separated and detected in the absence of other substances (i.e. in a high vacuum). Therefore, another significant component of a mass spectrometer is the vacuum system (Figure 1). In the mass spectrometers of the latter half of the 20th century and those of the 21st century, sample introduction, ionization, ion separation and detection, and data analysis are all done through the use of computers, which have become as integral to the mass spectrometer as any of its other components.

There are no limits to the types of matter that can be studied with MS. MS was originally used to determine the existence of the stable isotopes of the elements in the periodic table.[1] [The word *isotope* was suggested by Frederick Soddy (1877–1956) as a student and collaborator of Ernest Rutherford at McGill University in Montreal, Canada, in 1913. Isotopes are different forms of the same element that have the same atomic number, but differ in their relative atomic mass due to a difference in the number of neutrons present in the nucleus of the atom. The word is derived from the Greek words *isos* (equal) and *topos* (places), 'having the same place' in the periodic table.] MS has been used as a preparative technique [e.g. the separation of ^{235}U from uranium[2]]. It is also used for the identification and quantitation of various organic substances from the simplest gases such as methane and halomethanes to complex biomolecules such as proteins, oligonucleotides, and noncovalent complexes. Today, MS plays an important role not only in organic and biochemistry but also in inorganic chemistry such as the determination of metal contaminants in silicone wafers, drinking water, soils, industrial waste, etc. Mass spectrometers are not only found in analytical laboratories but also inside the helmets of space suits (to determine the level of gases that may pose a hazard), in tanks and other battlefield vehicles and on ships and aircraft (to detect the presence of chemical and biological warfare agents), and as more conventional field-portable instrumentation (for use at crime scenes, hazardous

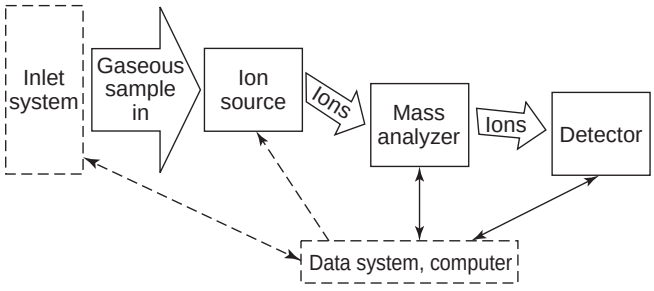

Figure 1 The concept of MS is to form ions from a sample, separate the ions based on their mass-to-charge ratio (this can be considered to be the same as the mass because the ion has only a single charge in most cases), and measure the abundance of the ions. In modern MS instrumentation used in environmental analyses, all of the functions (ionization separation of the ions, rate of data acquisition, detection of the ions, and storage of the data) are under computer control. Gaseous molecules are ionized in the ion source to form molecular ions. Some of the resulting molecular ions will fragment. The molecular and fragment ions are accelerated into a mass analyzer (mass analyzer and mass filter are used interchangeably). By various processes, ions of differing m/z values pass through the mass analyzer one at a time to reach the detector. When the ions strike the detector, they are converted into an electrical signal which, in turn, is converted into a digital response that can be stored by the computer. (Reproduced by permission from O.D. Sparkman, 'Mass Spectrometry', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, Wiley, New York, 1998.)

material spills, industrial waste sites, chemical-plant production locations, etc.). MS is used to characterize and analyze surfaces such as metallic and nonmetallic skins on the wings of modern aircraft or the coatings on papers specially prepared for use with laser printers.

Two factors make MS one of the most powerful analytical chemistry techniques in use today. The first is the simplicity of MS. The analytes are characterized by "... the mass of their molecules and the masses of the pieces of their molecules".[3] Unlike the spectroscopic techniques that rely on the absorption of specific energies of electromagnetic radiation, there is no knowledge of esoteric characteristic wavelengths of absorption or resonance frequency of atoms required to deal with resultant data. The second factor is that MS yields more information about the composition and structure of a molecule from less sample than any other analytical technique. MS is different from the other various spectrometries in that it deals with analytes one molecule at a time rather than the behavior of the total sample. The only negative aspect of MS is that the sample is consumed by the analysis [with the exception of Fourier transform mass spectrometry (FTMS)]; however, the sample size required is small enough to make this particular disadvantage insignificant. MS has been used to determine the mass of a complete protein from attomole quantities[4] and to quantitate unambiguously identified subpicogram amounts of the pesticide malathion on orange peel.[5] It is used to determine the age of artifacts [isotope-ratio mass spectrometry (IR MS)], contamination of the surface of metal and composite airplane wings [secondary ion mass spectrometry (SIMS)], and the components that give fresh-baked bread its delightful aroma. The list of applications is endless.

Another important distinguishing characteristic of MS is that it deals with the mass of the isotopes of the individual elements rather than the atomic weight of the elements. See Note 1 at the end of the article.

The field of MS was originally referred to as *mass spectroscopy* by Sir Francis William Aston in 1920 (English Nobel Laureate in Chemistry, 1922; 1877–1945; one of the founding fathers of MS),[6] who called his instrument a *mass spectrograph* because it had an arrangement of the electric and magnetic fields, used to separate the ions, that was analogous to that of an achromatic set of prisms without lenses, which produced a spectrum of lines such as an optical spectrograph. *Mass spectroscopy* became the all-encompassing term for the use of mass spectrographs (instruments that focused ions on to a plane for detection on a photographic plate or an array detector) and mass spectrometers (instruments that focus ions to a point for electrical detection one m/z value at a time) as well as the studies of isotopic abundances, precise mass determination, analytical chemical use, appearance potential, etc. However, the term *mass spectroscopy* is too broad and is no longer recommended. The preferred term for techniques involved with the measurement of ions according to their m/z values and their abundances is *MS*.

Over the last half-century, the field of MS has been extensively reviewed. Other than from an historical perspective, only the more recent general references are of value.[7–9] This article reviews MS from an historical perspective as opposed to the technology of the instrumentation and the utility of the applications.

The field of MS is discussed in terms of ion-separation methods, ionization techniques, or applications. All of these are extensively covered either in this section of the Encyclopedia or in one of the Applications sections. An important aspect of the successful use of an analytical technique is a clear understanding of the method by which the data are acquired. It is therefore worthy of mention as to sources of extensive and definitive information on m/z analyzers [magnetic-sector,[10] transmission-quadrupole (TQ),[11] quadrupole ion-trap (QIT),[12] time-of-flight (TOF),[13] and Fourier transform ion cyclotron resonance (FTICR)[14]]. These references provide a seminal understanding of the

physics of ion-separating techniques employed by these individual analyzers.

The literature sources for the understanding of ionization techniques are somewhat less specific. The important aspect of electron ionization (EI), which is the technique most often used with gas chromatography/mass spectrometry (GC/MS), is understanding the fragmentation mechanisms that result in the mass spectra. This understanding is achievable through a series of excellent references[15] or courses available from the American Chemical Society (ACS)[16] or the American Society for Mass Spectrometry (ASMS).[17] An important aspect of chemical ionization (CI), which is the technique most often used to obtain molecular weight information of a compound whose EI mass spectrum does not exhibit a molecular ion peak or that has some degree of volatility but is best separated by liquid chromatography/mass spectrometry (LC/MS), is the choice of the appropriate ionization compound[18] (reagent). In order to obtain EI and CI spectra of some compounds, it may be necessary to form volatile derivatives.[19]

The desorption/ionization (DI) technique of electrospray (ES) is still evolving. There are numerous descriptions of the mechanism involved in getting ions from the solution into the gas phase.[20] The other current popular DI technique, matrix-assisted laser desorption/ionization (MALDI), requires less of an understanding of the ionization mechanisms but more of an appreciation for the sample preparation[21] (the matrix). Other DI techniques such as fast atom bombardment (FAB), field desorption (FD), SIMS, and desorption chemical ionization (DCI) will be reviewed in this article and/or other articles in the Encyclopedia.

At this time, there is not a good single reference for the general technique of LC/MS; however, because of its maturity, GC/MS methods and techniques have been well elaborated.[22] MS receives extensive coverage in both the fundamentals review and the applications review

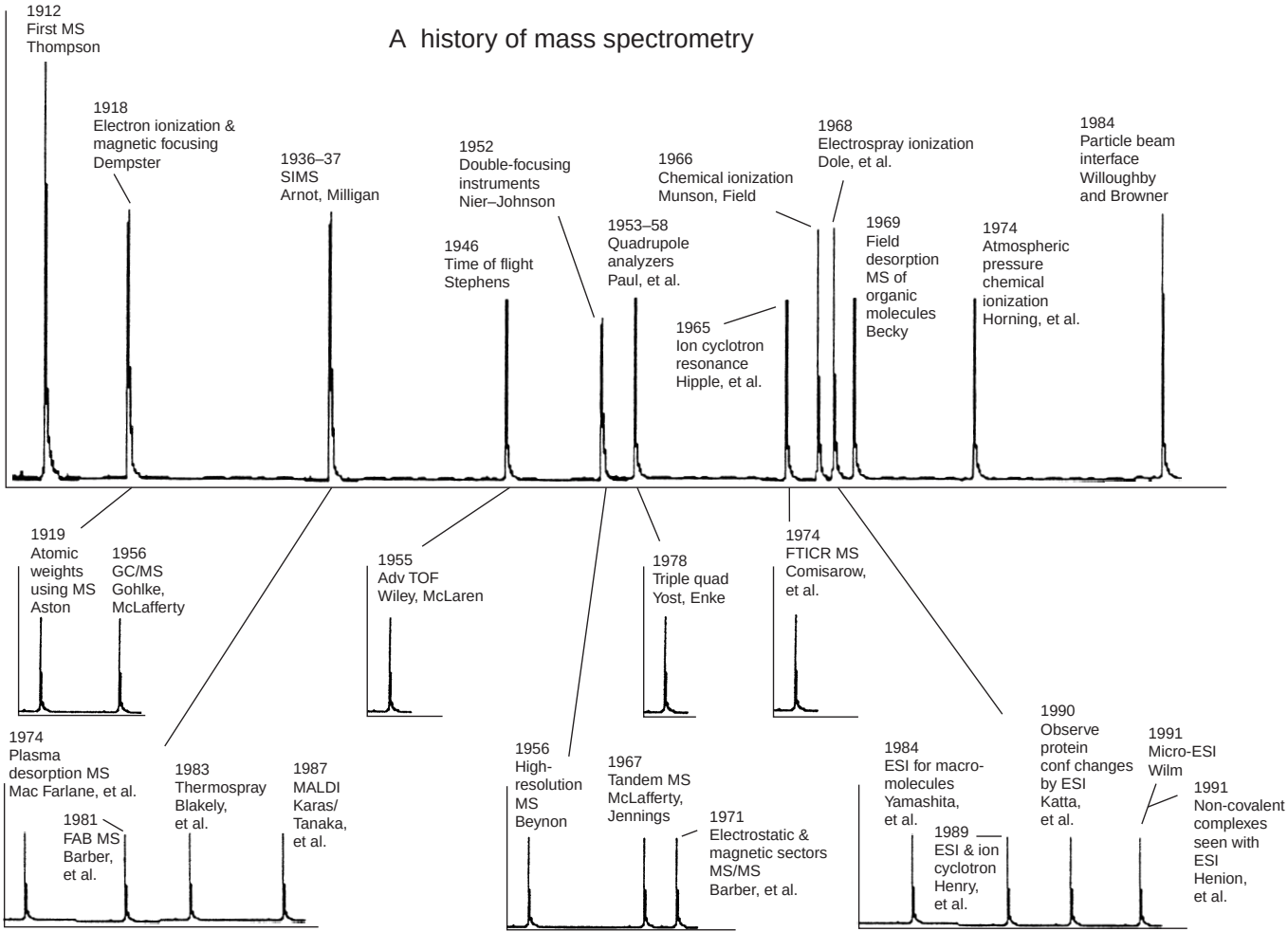

Figure 2 A partial time line of the history of MS. (Adapted with permission from Siuzdak.[23])

issues, which appear in alternating years, of *Analytical Chemistry*.

2 HISTORY

The history of MS is graphically chronicled in Figure 2, which is an adaptation from Gary Siuzdak's History of MS page.[23] Siuzdak's Web site also contains literature references for many of these events.

MS had its embryogenesis with the report of the German physicist Eugene Goldstein (1850–1930) at the Berlin Observatory that luminous rays in a discharge tube containing gases at low pressure emerged in straight lines from holes in a perforated metal disk used as a cathode. He called the rays Kanalstrahlen (canal rays).[24] This report has been referred to as the first MS paper[10] although the terms *mass spectrum* and *mass spectrograph* were suggested by Aston ca. 1920,[6] and *mass spectrometer* was a term first used by two well-known early pioneers of MS, William R. Smythe (US scientist) and Josef Heinvich Elizabeth Mattauch (Austrian physicist), ca. 1926.[6]

In 1895, the suggestion of the French physicist Jean Baptiste Perrin (1870–1942) that these rays were associated with a positive charge was confirmed by the German physicist and 1911 Nobel Laureate Wilhelm Wien (1862–1928).[25,26] Wien showed that these rays were deflected in a magnetic field and that their behavior could be studied by the combined effects of magnetic and electric fields.

The first instrumentation for the measurement of masses of charged particles was described in a lecture by Sir Joseph John Thomson entitled 'Cathode Rays' on April 30, 1887, at the Friday Evening Discourses in Physical Science held at the Royal Institution in London[27–29] (Figure 3a and b). This particular event was Thomson's announcement of the discovery of the electron for which he was awarded the 1906 Nobel Prize in Physics. Thomson called the Kanalstrahlen rays *rays of positive electricity*. Thomson became aware of the potential for these as tools in the analysis of matter and published the first MS book: *Rays of Positive Electricity and Their Application to Chemical Analyses*.[30] Thomson's discovery and initial instrument were followed by the development of a sector-based (magnetic and electric) mass spectrograph ca. 1911, which resulted in Francis William Aston (one of Thomson's collaborators at the Cavendish Laboratory) being awarded the 1922 Nobel Prize in Chemistry (Figure 4). This instrument employed velocity focusing. Aston's 1920 paper, 'The Constitution of Atmospheric Neon',[31] is considered the first paper in applied MS.[21]

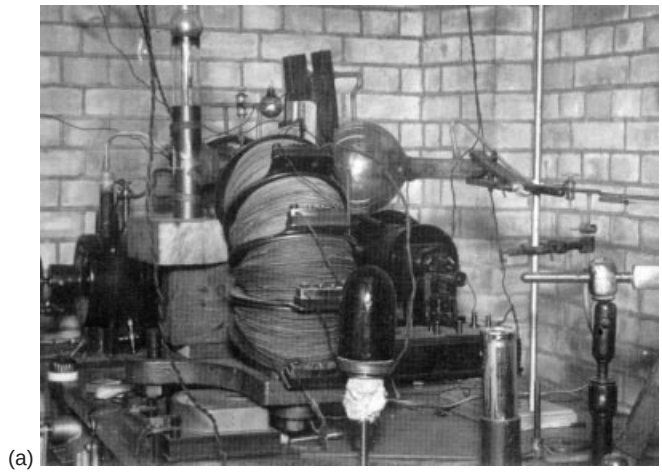

Figure 3 (a) Photograph of the positive-ray apparatus used by Thomson.[30a] (b) A schematic representation of the parabola mass spectrograph built by Thomson that includes: A, the gas inlet; B, the anode; C, the discharge tube; D, the port to the vacuum system; E, the cathode; F, the magnetic shields; G, the water jacket for cooling; H, the insulators; and J, the photographic plate used to detect the ions.[6]

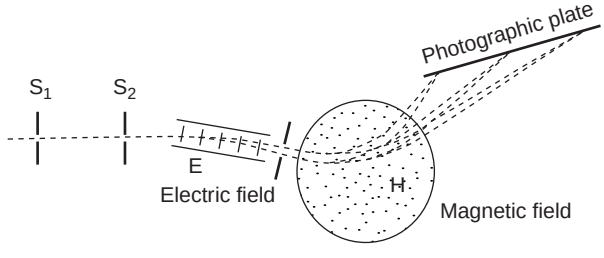

Figure 4 A schematic representation of Aston's velocity-focusing mass spectrograph.[45]

During the same period that Aston was pursuing the mass spectrograph, the Canadian–American physicist Arthur Jeffery Dempster (1886–1950), in an independent effort, developed a single magnetic-sector instrument that employed direction focusing of constant energy

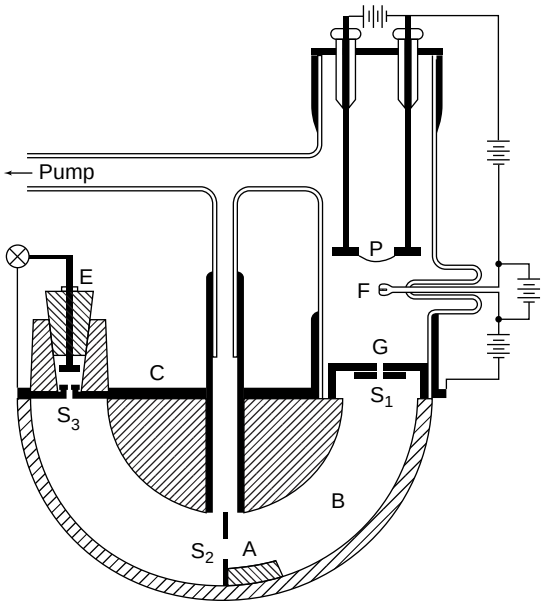

Figure 5 An illustration of Dempster's direction-focusing mass spectrometer that includes E, the ion-detection device; G, the glass ion-source housing containing the filament F that was heated to produce electrons that bombarded the platinum sample holder P; and B, the analyzer section surrounded by the magnet. A potential difference between the sample holder and the first slit (S_1) provided the ion acceleration.[6]

ions at the University of Chicago. Whereas Aston's instrument provided accurate m/z values, Dempster's instrument provided accurate ion abundances (Figure 5). This distinction between velocity focusing and direction focusing was the difference between mass spectrographs and mass spectrometers, respectively.

In the decade that followed (1911–1925), MS developed as a distinct field mainly as a result of these three pioneers. In the late 1920s, Dempster proposed the combination of direction focusing and velocity focusing with W. Bartky.[32] The first double-focusing mass spectrometers began to appear ca. 1934 with the Dempster, Mattauch–Herzog, and Bainbridge–Jordan geometries (the word "geometry" in MS refers to the arrangement of the electric and magnetic sectors between the ion source and detector in the mass spectrometer) (see section 2.6.1).

In these formative years, MS was used to determine the isotopes of the elements, to study atomic theory and thermodynamics, and to determine metabolic pathways through the analysis of biological substances generated with stable isotope-labeled precursors. The only other books published on MS were by Aston.[1,33] Although these books (two editions each) describe advances that had appeared in the literature, they concentrated on the primary use of MS as a tool for the determination of the mass of the isotopes of the elements. It was Aston's determination that there was more than one stable isotope of oxygen that led to the difference in the physical and chemical definitions of the atomic mass unit (amu). Prior to this time, there had been a single definition of the atomic mass unit, which was that 1 amu was equal to 1/16 the atomic weight of oxygen. Aston defined the amu as 1/16 the mass of the most abundant naturally occurring stable isotope of oxygen, ^{16}O. This definition resulted in a factor of 1.000275 difference between the standards, which was resolved by the replacement of the amu with the unified atomic mass unit (u or, in current terms, Da, the dalton) in 1962. Although unsubstantiated, it has been reported that Aston once said, 'Once all the isotopes of the elements in the periodic table had been determined, there would no longer be a need for MS'.

As the users of these early mass spectrometers (always custom built) began to complain about the presence of ions that did not correspond to their analytes, the organic chemists realized this was coming from organic impurities in the samples. This led to a growing interest on the part of the organic chemist. Commercial instruments began to appear, and MS became an essential tool in the petroleum industry. The 1940s saw the birth of organic MS. Not all organic compounds were suitable for mass spectral analysis. Since ions were separated in the gas phase, they had to be formed in the gas phase in this early instrumentation. This meant that the analyte had to be volatile and nonthermally labile.

While the development of the atomic bomb by the United States was underway, many of the developments in MS became shrouded in secrecy owing to the key role that it played in this development and the actual implementation of the bomb (Figure 6a and b).[2]

In the 1950s, a great deal of research took place on MS instrumentation. This led to several different ways of separating the ions according to their m/z values. Although the sector-based instruments (mass spectrometers are usually described by the way which ions are separated according to their m/z values) dominated the analytical chemistry laboratories, the desire for lower cost, simpler to use instruments led to the discovery and development of the TOF mass spectrometer, followed by the TQ, the ion cyclotron resonance (ICR), and eventually the QIT mass spectrometer. The pioneering efforts of this period led to the instruments of predominant use today. This same era was the beginning of the commercial MS industry with manufacturers appearing in the UK, Germany, and Japan to join those already established in the USA.

The gas chromatograph (GC), developed ca. 1952 by A. James and A. Martin,[34] was interfaced to the mass spectrometer in the mid-1960s.[35,36] This development allowed mixtures of organic analytes to be separated into relatively contaminant-free individual compounds that were delivered to the mass spectrometer in the gas phase.

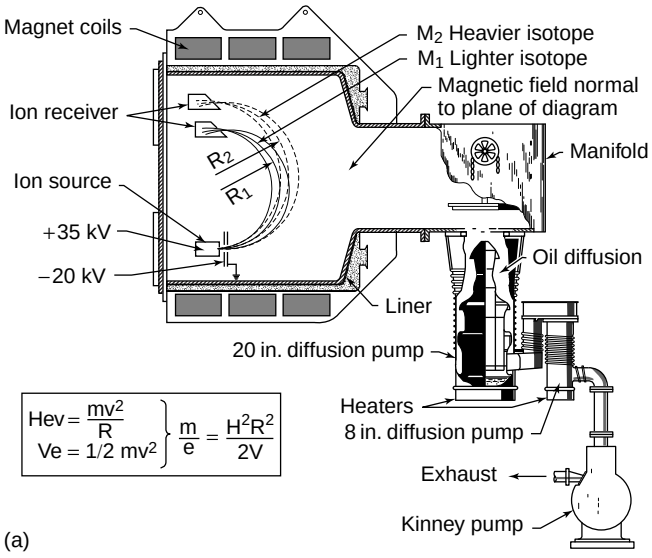

(a)

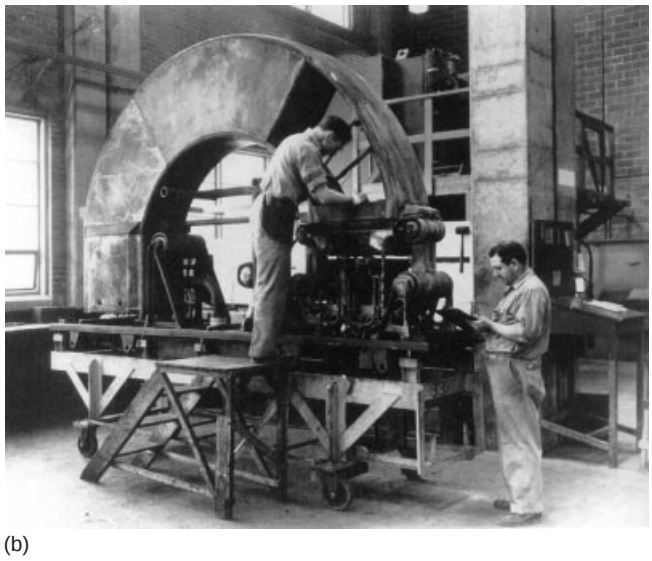

(b)

Figure 6 (a) An illustration of the Calutron's second-stage β units used at Oak Ridge National Laboratory to produce weapons-grade ^{235}U from the first stage. The origin of the name for this preparative instrument is 'CalU' from the University of California, Berkeley, and the Greek suffix 'tron', which means instrument – Calutron. ORNL Drawing 42951.[2] (b) Photograph of two workers cleaning one of the Calutron's vacuum system tank liners at the Y-12 Electromagnetic Separation Plant, previously known as the Clinton Engineering Works.[2]

GC/MS was limited in its usefulness by the plethora of data produced until the development of the minicomputer later in that same decade. The 1960s were also the period that MS of organic compounds began to be explained by organic reaction mechanisms, which allowed for a more detailed interpretation of the mass spectral data resulting from the EI of an organic compound and the prediction of fragment ions based on the presence of elements and substructures. New ionization techniques such as chemical and field ionization (FI) were also developed in the 1960s to be able to directly determine the molecular weight of compounds whose EI mass spectra did not have a molecular ion peak.

The gas chromatograph–mass spectrometer (GC-MS) became one of the most widely used analytical instruments by organic chemists in the 1970s,[37] and thus began the eventual dominance of the TQ mass spectrometer over the sector-based instrument that had ruled for the first 40 years of MS. With its lower price, more rapid spectral-acquisition ability, and ease of use compared with the sector-based instruments, the TQ mass spectrometer was ideally suited for GC/MS. In the late 1960s and early 1970s, new computer techniques were developed that resulted in easier methods of dealing with the GC/MS data.[38] The triple-quadrupole m/z analyzer was developed to determine additional structural information on the fragment ions produced by EI and CI.[39] The role of positive ions had dominated the MS of organic compounds; however, in the 1970s, the discovery of the formation of negative ions under CI conditions led to new levels of sensitivity in GC/MS.[40]

The high-performance liquid chromatograph (HPLC)[41] was interfaced to the mass spectrometer in the 1970s. This marriage was not as successful as that of the GC and the mass spectrometer. Owing to the volatile nonthermally labile requirement of the analyte (even though the analyte could be separated from the mobile phase of the HPLC), ionization was not always possible. The liquid chromatograph–mass spectrometer (LC-MS) remained a laboratory curiosity until the development of DI techniques. At the end of the 20th century, the LC-MS had become an even more prevalent instrument than the GC-MS. Part of the reason for this emerging dominance is the ability to form ions of analytes in the liquid chromatograph (LC) eluates using the atmospheric pressure ionization (API) technique of ES. The technique of LC/MS has also been popularized through the other API technique of atmospheric pressure chemical ionization (APCI). Unlike ES, APCI forms the ions in the gas phase. Although the ions must still be in the gas phase for their separation according to mass-to-charge ratio to take place, analytes that may not survive the rigorous temperature conditions of the GC required for their separation can pass through the LC process without harm and still have ions that end up in the gas phase. The API LC/MS techniques do not provide the same level of structural information as can be obtained from the gas-phase EI process; however, structural information can be obtained through techniques of mass spectrometry/mass spectrometry (MS/MS) and using in-source collisionally activated dissociation (CAD), which are also fairly recent developments.

The late 1980s and the 1990s constitute the period during which the DI technique of MALDI developed and became a routine tool. ES is also referred to as a DI technique. These two DI techniques allow for transfer of ions formed in solution (ES) or in a solid (MALDI) to the gas phase for m/z analysis. The DI techniques have eliminated the volatile nonthermally labile requirements for analytes allowing for the characterization of synthetic organic polymers and biopolymers by MS. In this same time period, inductively coupled plasma (ICP) ionization for inorganic substances in solution was linked with MS. In addition, other DI techniques such as SIMS has become a routine tool in the analysis of both organic and inorganic substances on the surfaces of substances important to the material scientist. These last two techniques have given MS a rebirth in the field of inorganic analyses.

Since the determination of stable isotopes of the elements of the periodic table by MS, the widest impact of the analytical technique has been in organic chemistry. MS still plays an important role in inorganic chemistry and geochemistry, but the major discoveries and uses are in the organic field.

2.1 Beginning to the 1940s

At the time Thomson reported his parabolic instrument using electrical and magnetic fields, he referred to the particles that are known as *electrons* as *corpuscles* (small negatively charged particles from which atoms were built up). This made Thomson the first mass spectrometrist to generate neologisms (unfortunately, a trait not uncommon to modern mass spectrometrists) that later fell into disuse (*rays of positive electricity* and *corpuscles*). With the collaboration of Aston, Thomson's apparatus was improved to the extent that these *rays of positive electricity* (ions of positive charge) could be separated according to their mass-to-charge ratios. Their joint efforts led to the 1910 improvements of the Thomson positive-ray tube, which was used in the 1912 discovery of the two stable isotopes of neon. This report was the first of nonradioactive isotopes[31] and had a profound effect on the development of atomic and nuclear theory.

After World War I, Aston returned to the Cavendish Laboratory, continued to collaborate with Thomson, and dedicated his career to the separation of the two isotopes of neon and the proof of their existence beyond doubt. This work produced continued improvements of Thomson's apparatus and culminated in the development of the mass spectrograph. Thomson is the discoverer of MS, and Aston is the developer of the mass spectrometer.

In the early part of the 20th century, the technique of positive-ray analysis took two different paths: (1) that pursued by Aston, which was the accurate determination of mass; and (2) that pursued by Dempster (in the USA), which was more suited for the determination of the abundance of the ions. Aston's instrument results in the ions striking a photographic plate that is on a focal plane. Aston called this a mass spectrograph and also coined the term *mass spectrum*. The Dempster instrument focused the ions of one m/z value at a time on a point so that they could be converted to an electrical signal that is then amplified and measured. Through the ability to record accurate ion abundances, Dempster was able to show that seven out of every 1000 uranium atoms were the isotope with a nominal mass of 235. Thomson, Aston, and Dempster (Figure 7a–c) are considered to be the founding fathers of MS.[10]

Aston foretold the impact of MS on atomic theory in his 1922 Nobel address where he forecast both nuclear bombs and power plants based on the future use of the energy in an atom's nucleus.

Building on the developments of the founding fathers of MS, others began to advance the field in the period just prior to World War II. During this period, Dempster developed EI. [EI was originally called electron impact. This is a process by which gas-phase molecules at a pressure of $>10^{-3}$ Torr are ionized by a beam of electrons, produced by a hot wire (filament), that have been accelerated by 70 V (i.e. 70 eV).] EI is used in modern mass spectrometers where analytes are in the gas phase. There has been more study of EI processes and fragmentation than all others types of mass spectral ionization techniques combined. All of the large archives of mass spectral data contain EI spectra obtained at 70 eV. Dempster also developed the spark-source and vacuum-vibrator techniques used in the analysis of inorganics in solids.

During this pre-World War II period and after the significant developments of the Aston/Thomson/Dempster era, most of the advances in MS were made in physics. A series of seminal articles were presented by Roboz (p. 491).[10] These papers began to appear in the early 1920s and continued through the mid-1930s. After that time, there were few references, if any, to MS because of its strategic importance to the war effort both in the area of organic chemistry and nebular weapon development.

One important aspect of the use of MS came about through the isolation of heavy water and the discovery of deuterium by Harold Clayton Urey (1893–1981), a US chemist at Columbia University, New York, in 1934. This accomplishment resulted in Urey's receipt of the Nobel Prize for Chemistry in 1934. David Rittenberg, one of the doctoral students in Urey's laboratory at the time of this discovery, studied some of the thermodynamic properties of compounds containing these newly discovered isotopes. Shortly after receiving his PhD in 1934, Rittenburg joined the Biochemistry Department at Columbia University, where he remained until his

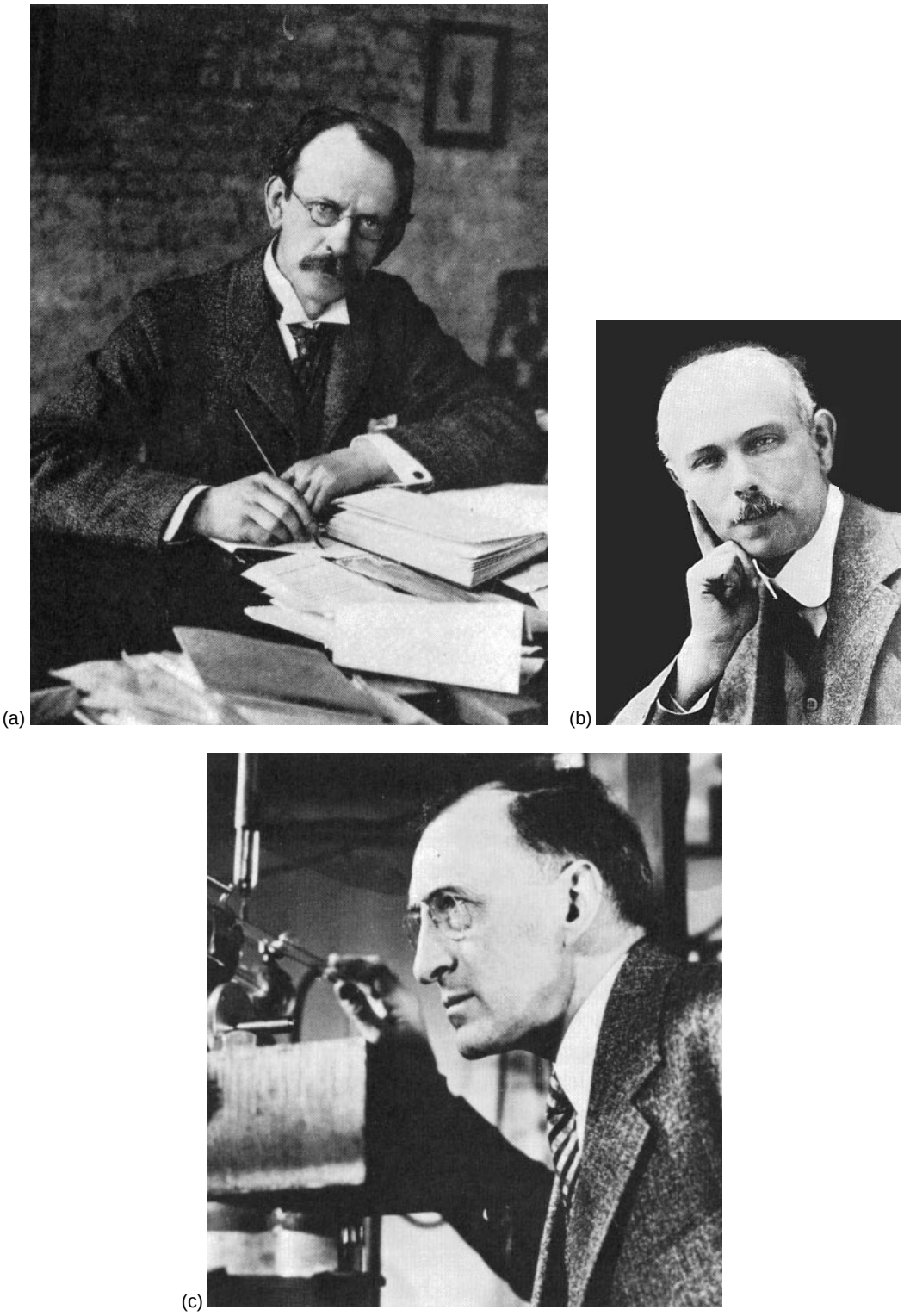

Figure 7 (a) Sir Joseph John Thomson, English physicist and 1906 Nobel Laureate in Physics, 1856–1940. (Courtesy of Dr. Vernon H. Dibeler.[6]) (b) Francis William Aston, English physicist, student of Thomson, and 1922 Nobel Laureate in Chemistry, 1877–1945. (Courtesy of Dr Edward Wichers.[6]) (c) Arthur Jeffery Dempster, Canadian–American Physics Professor, University of Chicago, 1886–1950. (Courtesy of the University of Chicago.[10])

death in 1970. He and a departmental colleague, Rudolf Schoenheimer, began a study in 1935 of intermediate metabolism using stable isotopes. They were the first to use deuterium labeling to investigate the metabolism of amino and fatty acids. This original work of Rittenburg and Schoenheimer resulted in the 'dynamic state of body constituents' concept, which said that the apparent static state of many constituents results from rapid opposing reactions at nearly equal rates.[42] Over the next 36 years, Rittenburg published over 200 papers. Many of the specifics of using stable isotope-labeled compounds in the determination of metabolic pathways is presented by Richard Caprioli,[43] the last person to receive a PhD under the direction of Rittenburg.

The ability to use the difference in mass between an analyte and its analog labeled with a stable isotope of nitrogen, carbon, oxygen, and/or hydrogen has resulted in one of the primary uses of MS in the biological sciences. The labeled material can be used to determine which compounds resulted from the metabolism of an ingested drug or nutrient. Labeled analogs can be used as internal standards, which result in lower detection limits owing to finite analyte-absorptive properties of a particular analysis. As internal standards, labeled analogs of an analyte will behave chemically and physically the same as the analyte. This means reduced errors in quantitation due to internal standard effects. The most difficult step in the use of stable labeled isotopes is the preparation of the label compound. These selectively labeled compounds have been extensively used to determine the mechanisms of fragmentation. One of the pioneers in this area was the Swedish medical scientist Einar Stenhagen (1911–1973), who received several patents on the synthesis of labeled compounds of biological interest. He was also granted a patent for the jet separator, the most popular device for the interface of high-flow packed columns and large-bore, capillary GC columns [530–750 μm inner diameter (ID)], which have high optimum flow rates (15–30 mL min^{-1}) to a mass spectrometer.[44]

In spite of the extensive contributions by Rittenburg based on the use of MS, he only receives token mention[10,45,46] in some of the early references on MS and none in others.[6,47] This lack of reference is one of the reasons for the importance of the two early books edited by George Waller at Oklahoma State University.[43,44]

It was not until the 1940s that commercial instruments began to appear. All the instruments used in isotope-labeling studies were built as one-of-a-kind research curiosities by their user or colleagues of these users. The first of these commercial instruments was the Consolidated Engineering Corporation's (Pasadena, CA) CEC 21-101. Until that time, mass spectrometers were custom built by researchers such as Dempster. Among this group of early innovators of the instrumentation were Dempster, Smythe, Mattauch (in cooperation with Richard Herzog in Germany ca. 1934–1936), and the American physicist Alfred Otto Carl Nier (1911–1994), who was a key figure in the Manhattan project. One of the two most popular geometries for a double-focusing instrument was a result of this era, the Mattauch–Herzog mass spectrograph, with ion optics worked out by Herzog[48] and built by Mattauch in Germany.[49] The other is the Neir–Johnson geometry developed by Neir and Edger G. Johnson, a master's degree graduate student, after World War II, ca. 1948.[50]

One of the hindrances to the development of MS was the crude state of knowledge about the production of a vacuum. The low-pressure requirement of MS was such that poor vacuum technology resulted in hours of frustration. One very significant improvement that came out of the war research into MS was advances in vacuum systems.

The Consolidated Engineering company was the first commercial manufacturer of mass spectrometers. This pioneer mass spectrometer manufacturer was founded by Herbert Hoover, Jr, in 1937 as the engineering and manufacturing subsidiary of the United Geophysical Company to develop a mass spectrometer to locate oil deposits by detecting hydrocarbon gases emitting from the ground. The company had close connections with the California Institute of Technology and the petroleum industry. The initial project was a failure due to the ubiquitous nature of methane in ground-seepage gas. However, the instrument was useful as a replacement for a distillation analysis used to detect light-end hydrocarbon gas mixtures in the production of aviation fuel.[51] The company was later renamed the Consolidated Electrodynamic Corporation because of restrictions of the use of the words 'engineering' and 'engineer' in the state of Texas to those that were licensed by the state as engineers. The company was then sold to Bell & Howell and later to DuPont. While under the control of DuPont, the manufacture and support of mass spectrometers was terminated.

The CEC 21-101 had an m/z range of 16–72, and it was able to produce a spectrum over this range of a mixture of C_1–C_5 hydrocarbons in 20 min. This initial design was refined into several subsequent models culminating in the CEC 104, which had a unit-resolving power over its entire m/z range to 700. The last of these instruments was manufactured in 1965. The last instrument was taken out of service by Exxon in 1997.

One of the hindrances to the advancement of MS after the introduction of the CEC 21-101 was a requirement in the CEC sales agreement demanding that all data made available be made available to CEC.[52] This requirement was not well received in the competitive petroleum industry and led to an all-glass instrument

that was produced by Westinghouse Electric Corporation (the Type LV, a 90° sector instrument as opposed to the CEC 180° instrument) in collaboration with Humble Oil and Refining and several other oil companies. These Westinghouse instruments were some of the first used outside of the temples of MS (University of Minnesota, Princeton, Harvard, University of Chicago, and Cal Tech) for the study of organic compounds. After the end of the war, CEC changed its information-sharing policy and went on to dominate the field of commercial MS through the 1960s. The end of the 1940s saw the beginning of the second era of MS.

It is interesting to note that the CEC 21-101 mass spectrometer was copied from Nier's 1937 instrument, which had an analyzer copied from Dempster's 1918 instrument. Nier built this instrument while studying with Kenneth T. Bainbridge at Harvard University. Unlike Dempster's 180° instrument, the Nier mass spectrometer had an ion source with both electric and magnetic fields. This instrument represented the birth of commercialization of academic research in MS.

2.2 The Birth of Organic Mass Spectrometry

Until World War II, MS had only a small role in organic chemistry. At the end of the war, advancements in MS had been largely confined to improvements in instruments to separate isotopes of elements and how to make these instruments suitable for the preparation of production quantities of isotopes that played a major role in atomic weapons production. There were only three books devoted to the subject of MS – one authored by Thomson[30] and two by Aston.[1,33] All three of these books had two editions. The first book to appear after World War II was G.P. Barnard's *Modern Mass Spectrometry* in 1953.[45] This book was similar to those of Aston and Thomson in that it concentrated on the instrumentation and ion behavior. Treatment of applications was somewhat broader than that of Thomson or Aston, but still only of footnote quality and quantity. And keeping with that philosophy, there is only a brief mention of Rittenburg's use of MS.

Although MS continued to be primarily the privy study of the physicist, the role of the organic chemist began to emerge with a high degree of visibility. Several companies (CEC in the USA, AEI in the UK, Atlas in Germany, and Hitachi and JEOL in Japan) began the development of commercial instruments after World War II primarily for use in the analysis of organic compounds. This proliferation of manufacturers led to more users and a desire for people to discuss common experiences and problems. According to Seymour Meyerson, the first MS meeting, the symposium on mass, infrared (IR), and ultraviolet (UV) spectrometry,[53] was held in the USA and organized by Rubber Reserve Company and the Petroleum Administration for War. The first review of mass literature appeared in *Analytical Chemistry* in January 1949.[54] In this review, John Hipple of the National Bureau of Standards (NBS), Washington, DC, a Division of the United States Department of Commerce,[55] pointed out that in *Chemical Abstracts* there had been 11 references to MS in 1943, 15 in 1944, 17 in 1945, 26 in 1946, and 40 in 1947. Hipple's review had 165 citations as compared with that of Alma Burlingame in the same type of review published in 1998,[9] which had 1551 citations divided into nine categories: Overview (5), Scope (173), Innovative Techniques and Instrumentation (364), IR MS (89), High-power Lasers in MS (51), Dissociation by Low-intensity IR Radiation (18), Polymers (61), Peptides and Proteins (264), and Oligonucleotides and Nucleic Acids (166).

September 6–8, 1951, were the dates for a symposium on MS in physics research, held in Washington, DC, as one of 12 symposia commemorating the semicentennial of the then NBS [now the National Institute of Standards and Technology (NIST)]. John Hipple organized this symposium in cooperation with the Office of Naval Research.[56] Of the 36 presentations, only one pertained to the use of MS in organic chemistry: 'Mass Spectra of Heavy Hydrocarbons', by M.J. O'Neal, Shell Oil Company (Houston, TX). This presentation included spectra of aliphatic hydrocarbons and aliphatic hydrocarbons with aromatic substitutions. The remainder of the presentations revolved around instrumentation and ion physics.

Similar conferences were organized in the UK by the MS Panel of the Institute of Petroleum in 1950 and 1953. The proceedings of the second of these meetings, held October 29–31, 1953, at the Institute of Electrical Engineering in London with 200 people in attendance, were published.[57] This conference consisted of five half-day sessions, with a total of 26 presentations. A commercial exhibition associated with the conference had four companies displaying computing equipment: British Tabulating Co., Ltd., Elliot Bros. (London), Ltd., Ferranti, Ltd., and International Business Machines (United Kingdom), Ltd., and two companies displaying mass spectrometers, Metropolitan–Vickers Electrical Company, Ltd., and Vacuum Industries Applications, Ltd. This conference, as could be guessed by the name of the organizer, was much more related to the use of MS in organic analyses.

The proliferation of mass spectrometer manufacturers led to the organization of 'user meetings' where people using the same types of instruments could gather and discuss common problems and tell of their new discoveries. One such user meeting organized jointly by CEC and General Electric Corporation at the 1952

meeting (third meeting) of the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Pittsburgh, PA), sponsored by The Society for Analytical Chemists of Pittsburgh and The Spectroscopy Society of Pittsburgh, became the E-14 Committee (an educational committee) of the American Society for Testing and Materials (ASTM). Proceedings of the annual E-14 meeting have been published since 1961. At the 1969 E-14 Committee meeting in Dallas, TX, the membership formed the ASMS, which is the largest organized group of mass spectrometrists in the world today. After 1969, the published proceedings were under the ASMS name. See Note 2 at the end of the article.

As an outgrowth of the two meetings organized by the MS Committee of the Institute of Petroleum in the UK, an international MS meeting, the International MS Conference, began in 1958. There have been 14 of these triennial meetings. The first six meetings were held in the UK. Beginning with the seventh meeting (Florence, Italy), each succeeding meeting has been held in a different European country. The 15th International MS Conference is scheduled for Barcelona, Spain, in 2000.

The instrument manufacturers' user meetings are still significant, and there were at least 10 companies holding these meetings at the 1999 annual ASMS Conference on MS and Allied Topics.

In the early 1960s, John Beynon in the UK (1960) followed by Klaus Biemann in the USA (1962) published two of the classical books on MS of organic compounds: *MS and its Applications to Organic Chemistry*[46] and *MS: Organic Chemical Applications*.[47] These books were quickly followed by a series of books showing that mass spectral data could be related to mechanisms of organic reactions.[58] In addition to these important works pertaining to organic MS came another series of general-interest books from both American and British authors. Two of the most important were those authored in the US by Robert W. Kiser (Kansas State University, Lawrence, KS)[6] and John Roboz (Air Reduction Company, Inc., Murray Hill, NJ).[10] A book edited by George Waller (Oklahoma State University) was published in 1972,[59] *Biochemical Applications of Mass Spectrometry*. This book is the first book devoted to MS and biochemistry.

Through cooperation between those interested in MS for applications in organic chemistry and biochemistry such as Biemann and Beynon and the several instrument companies that were vying for the market, a number of improvements were made in the magnetic-sector mass spectrometer. These improvements included more rapid acquisition rates and the development of mass accuracy measurements that allowed the assignment of empirical formula to ions whose mass was determined to the nearest 10 millimass units. During these early phases of organic MS, the instruments were used as separation devices for mixtures such as the paraffin, olefin, naphtha, and aromatic (PONA) analyses used by the petroleum industry, in the determination of isotopically labeled compounds purified from complex biological mixtures, and in the determination of the structure of purified compounds.

The period after World War II also saw developments in nonmagnetic-field technologies to separate the ions. These developments resulted in the TOF, TQ, and QIT mass spectrometers. Details of these developments can be found in section 2.6.

2.3 Gas Chromatography/Mass Spectrometry – A Tremendous Success

Almost from the appearance of the first paper on gas chromatography, the organic mass spectrometrist saw the potential advantage of separating a complex mixture into nearly pure components followed by mass and structural analysis with no need for human intervention. Although the technique of GC/MS had the greatest potential for those in the biological science, the first published report of this interface that involved dilute analytes in an above-atmospheric-pressure gas being introduced into an instrument that operates at pressures as low as 10^{-10} Torr [pressures in high vacuum are referred to in Torr by most manufacturers of mass spectrometers. 1 Torr is equivalent to 1 mm of mercury or pascal (Pa); 1 Torr = 133.322 Pa; 1 Pa $\approx$ 0.0075 Torr; 1 atm = 760 Torr or 1.013×10^5 Pa (exactly, by definition)] came from the bulk chemical industry – Dow Chemical in Midland, MI.[35,36] This first effort used a TOF mass spectrometer because of its rapid spectral acquisition rate compared with a sector-based instrument and an oscilloscope fitted with a Polaroid camera to record the spectra. Less than 1% of the GC eluate was put into the mass spectrometer.

Three major hurdles had to be overcome in order for GC/MS to be an executable analytical technique: (1) the problem with the amount of gas expelled from the GC and the requirement for high vacuum in the mass spectrometer; (2) the need for rapid mass spectral acquisition to compensate for the ever-changing partial pressure of the analyte; and (3) the tremendous amount of data collected during a single GC/MS analysis.

Dealing with the pressure difference and the low analyte concentration in the GC eluate was undertaken by a number of different investigators. This work resulted in several different devices that are known as GC/MS separators. The most popular was the jet separator invented, and later patented, by Einar Stenhagen (Swedish medical scientist) and perfected by Ragnar Ryhage.[60] In addition, Jack Watson, while a PhD student of Klaus Biemann at Massachusetts Institute of Technology (Cambridge, MA), and Duane Littlejohn and Peter Llewellyn, at the

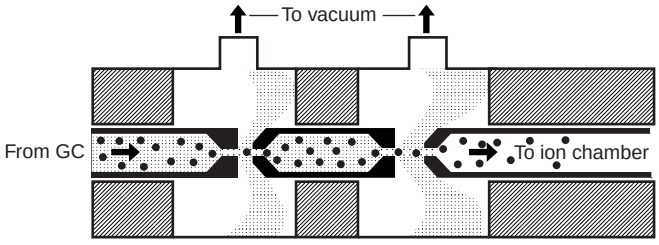

Figure 8 A graphic illustration of a two-stage jet separator. (Reproduced by permission from O.D. Sparkman, 'Mass Spectrometry', in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, Wiley, New York, 1998.)

Varian Research Center in Palo Alto, CA, also developed separators.[61,62] The Llewellyn separator enjoyed some commercial success but not to the same extent as that of the jet separator (Figure 8). As the use of capillary columns (50–250 μm ID GC columns and flow rates of 1–2 mL min^{-1} where the stationary phase is coated on the inner walls) replaced the packed columns of the 1960s and 1970s (2–5 mm ID GC columns and flow rates of 20–30 mL min^{-1} where the stationary phase is a high-boiling liquid coated on diatomaceous earth particles packed inside of the column), the need for these devices used to enrich the concentration of the analyte in the GC eluate passing into the mass spectrometer as well as reducing the total volume of gas introduced to the mass spectrometer was no longer needed. Jet separators are still used with larger diameter capillary columns (530–750 μm ID), but the use of these columns is becoming less as new preconcentration techniques are developed.

Although the use of the TOF mass spectrometer appeared to resolve the problem of rapid data acquisition, there were other problems associated with this type of m/z analyzer that kept it from becoming a part of a successful commercial GC-MS. With the potential for GC/MS, the first efforts were put into interfacing existing instruments to the GC. None of these endeavors were very successful. As more investigation took place, it was found that GC/MS was as different from either gas chromatography or MS as these two techniques were from one another. The first commercial instrument built as a GC-MS was the LKB 9000 produced in Sweden by LKB Instruments in the mid-1960s. Other manufacturers produced a mass spectrometer with a GC inlet, but the LKB 9000 was neither a GC with a mass spectrometer as the detector nor a mass spectrometer with a GC inlet. LKB pioneered the technology required to remove the carrier gas from the GC eluate from packed columns (enriching the flow with analyte prior to introduction into the mass spectrometer: jet separator). The LKB 9000 was a single-focusing magnetic-sector instrument. To optimize it for GC/MS, special considerations were given to the vacuum system and the scanning of the magnetic field. In addition,

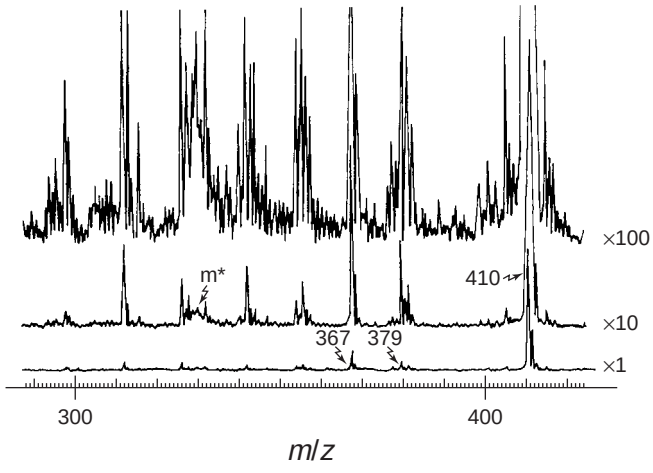

Figure 9 An illustration of mass spectral data acquired with a three-channel light-beam oscillographic recorder.

emphasis was put on the detector system to allow for the rapid acquisition of data required for GC/MS analyses. The first prototype LKB 9000 was delivered to George Waller at Oklahoma State University (Stillwater, OK) in 1964. The LKB 9000 did not address the issue of the amount of data that is acquired in a GC/MS analysis. This instrument used a light-beam oscillographic recorder to record the mass spectra. Spectra were manually selected for recording (Figure 9).

With the development of the minicomputer in the mid-1960s, it became possible to deal with the vast amount of data generated in a single GC/MS analysis. Before that time, acquiring spectra over a range of 200–250 m/z units every 2–3 s in a GC analysis that lasted for 15–30 min produced as many as 225 000 data points (mass spectral peaks in multiple spectra) that had to be dealt with manually. Data were typically acquired by using a multichannel analog recorder that increased the amount of data and the complexity in converting it to usable information. The paper used to record these data was usually light-sensitive and had a short life before the recorded images would fade into oblivion especially if exposed to fluorescent or sun light. This volume-of-data problem greatly reduced the number of samples that could be analyzed with a reasonable amount of effort. The minicomputer allowed for the automated collection of data. The generation of several different types of chromatograms was made possible by computer processing the m/z and intensity data recorded by the data system (DS). Sums of intensities for all the m/z values in each spectrum (total ion current) could be plotted versus the spectrum numbers to produce a reconstructed total ion current chromatogram. Intensities of specific m/z values, sums of m/zvalues, or ranges of m/z values could be plotted versus spectrum numbers to produce

reconstructed 'mass chromatograms'. A DS typically could display bar-graph mass spectra (unit resolution and intensity data) for the data from individually acquired spectra. These bar-graph spectra would generally be normalized to the most intense peak (base peak) and could have peaks due to ions from substances other than the analyte (background) deleted. Even with these primitive abilities (by today's standards), the number of samples that could be handled in a week increased by orders of magnitude.

GC/MS began to become a routine technology with the introduction of the TQ GC/MS instrument. The first commercial TQ GC/MS system was developed by Finnigan Corporation (San Jose, CA), now part of ThermoQuest. According to Bob Finnigan, one of the founders of Finnigan Corporation, in an *Analytical Chemistry* A-pages article,[37] the first Finnigan 1015 GC/MS instrument (prototype) was delivered to Bert Halpern and Walt Reynolds at the Stanford Medical School Department of Genetics in January 1968. Two other prototypes were delivered to John Amy at Purdue University and Carl Djerassi at Stanford University. Based on the feedback from these initial installations, many improvements were made in the Model 1015, especially in the single-jet separator design that had been given to Finnigan by Stenhagen (the Ryhage jet separator used in the LKB 9000 was a dual-stage design), which took almost 2 years to perfect.

Bob Finnigan became aware of the power of GC/MS in combination with the results he saw from the DS designed and built by Reynolds and Tim Coburn at Stanford.[63] Although Hites and Biemann had earlier reported a system to deal with the acquired data from a GC/MS system,[38] the Reynolds and Coburn system for the TQ mass spectrometer along with a similar system for magnetic-sector instruments developed by Jack Holland and Chuck Sweeley at Michigan State University[64] were the first DSs that actually controlled the mass spectrometer and also stored the data. The Stanford GC/MS DS led Finnigan to enter into an agreement with a new company, Systems Industries, Inc. (Sunnyvale, CA), to commercialize the Stanford system and become the first commercial GC/MS DS – the System 150. Finnigan accepted its first order for a computerized GC/MS system in 1968 from Evan Horning at the Baylor College of Medicine (Houston, TX).

Initially, the System 150 was sold only by Finnigan Corporation. The data consisted of a Digital Equipment Corporation (DEC) PDP-8 minicomputer with 4K words (12-bit words) of core memory, a magnetic tape system (a LINC, Lincoln Laboratories at MIT, tape) designed specifically for the PDP-8, a standard ASR 32 teletype for input and output of text information as well as the ability to load programs through a punched paper tape, and a Houston Instruments digital plotter for the output of bar-graph mass spectra normalized to the most intense peak and reconstructed chromatograms (total ion current and mass). The DS had an interface that would convert digital instructions into analog signals used to tell the mass spectrometer which m/z range to scan, the time interval for each scan, the total analysis time, the magnitude of the radiofrequency (RF) and direct current (DC) voltages for each individual m/z value, and other factors associated with data as well as taking the analog data from the mass spectrometer that represented m/z values (based on calibration with a compound that produced ions of known m/z values) and ion abundances. These analog signals from the mass spectrometer were converted to digital values that were used to display the acquired data as spectra or reconstructed chromatograms. The computer and magnetic tape, along with the interface, were fitted into an electronics rack that was 19 in wide, 60 in high, and 30 in deep. The System 150 went on to be a general-purpose GC/MS DS sold by Systems Industries as well as several manufacturers of GC/MS instrumentation on an original equipment manufacturer (OEM) basis.

The System 150's capabilities were expanded and refined through the use of third-party software such as the Biemann computerized library search of mass spectral databases[65] and use of the Tektronic's electronic storage terminal for graphical display (both developed at the Battelle Columbus Laboratories), and data analysis and acquisition software specific to sector-based instruments developed at Michigan State University. The System 150, later marketed by the French GC/MS manufacturer Nermag (originally Riber) as the RDS 400, also took advantage of new hardware features such as disk storage that replaced the slower and lower capacity magnetic tape systems. The Riber 400 remained a reliable and competitive GC/MS DS through the mid-1980s. Because all of the software was developed using the PDP-8's assembly code and the system was capable of either acquiring or processing data, but not both at the same time on a single computer, it was very fast. It has only been in the last 2–3 years that GC/MS data processing has achieved the speed of the System 150/Riber 400, and this has been due to the very fast computers with extremely large amounts of random access memory (RAM).

There were two additional third-party DS companies that appeared in the late 1970s: INCOS (acronym for Instrument Company Operating System) and Teknivent Corporation. The INCOS DS was originally designed as a retrofit to sector-based instruments. The INCOS used a Data General minicomputer with a custom disk operating system called IDOS (another acronym) that allowed for simultaneous data acquisition and analysis. The data analysis features were unique with some rather spectacular graphics displayed on the Tektronic's graphics

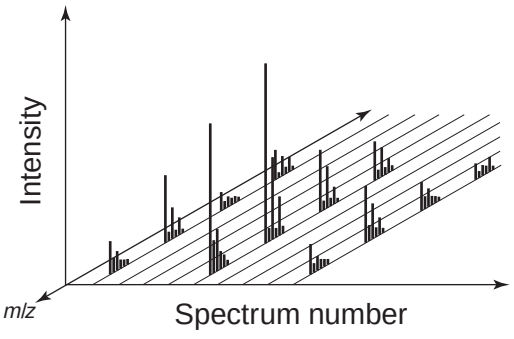

Figure 10 A partial display of the three dimensions of chromatographic/mass spectrometric data.

terminal. These graphics included a three-dimensional plot of the spectra along the z-axis, a spectrum number on the x-axis, and an m/z intensity on the y-axis (Figure 10). The INCOS DS introduced a new era in the comparison of a spectrum of an unknown substance with those contained in a mass spectral database. The INCOS library search algorithm became the basis for most commercial DSs used with current instrumentation. INCOS was sold to Finnigan Corporation in the late 1970s. Finnigan quickly adapted the system for use with its line of TQ mass spectrometers. The INCOS DS gave Finnigan a significant advantage in its second generation of GC/MS instrumentation – the 4000 series.

The INCOS DS remained an exclusive of the Data General minicomputer long after many manufacturers had shifted DS developments to the IBM PC-type computer. This unusual longevity is rumored to have resulted from a settlement between Data General and Finnigan over the fact that the INCOS' unique disk-operating system, IDOS, was an infringement of copyrights held by Data General on disk-operating systems.

Teknivent was another company that joined System Industries/Riber and INCOS in competition for the after-market DS business. Teknivent developed a series of DSs that were based first on various minicomputers and then on the IBM-type microcomputers (PCs). They were one of the first companies to introduce a GC/MS DS on the PC using the IBM PCXT ca. 1983. Teknivent was the only instrument company to use the IBM OS2 operating system, which allowed for control of multiple gas chromatograph/mass spectrometer systems from a single computer. A significant number of these DSs were sold through the late 1980s and early 1990s as replacements for minicomputer systems used with the Riber R10-10, Finnigan 3000 and 4000, and INCOS 50 series of GC/MS instrumentation as well as many of the Hewlett-Packard instruments that used UNIX and Pascal-based DSs. This St Louis, MO-based company ceased doing business around 1996, and its assets were acquired by ProLab Resources (Madison, WI), a company that had been functioning as a third-party software developer for GC/MS data analysis systems.

Although none of these original third-party DS manufacturers exist today (with the exception of Teknivent within ProLab Resources), the concept of after-market DSs is still very strong. Among the companies at the end of the second millennium, there is one that produces a system primarily for sector-based instruments – MSS in Manchester, UK. MSS also has a system for quadrupole instruments. Los Gatos Circuits (LGC, San Jose, CA) markets a DS for some models of TQ instruments that is produced by SysNet Corporation (Scotts Valley, CA). In addition, Shrader Analytical (Detroit, MI) has produced a system that is OEM to JEOL USA and sold as an after-market system for sector-based instruments.

Over the next few years, following the introduction of Finnigan's first instrument, the field of GC/MS grew at an exponential rate largely fueled by a very active market in environmental chemistry. The Finnigan Model 1015/System 150 was the first commercial GC/MS DS and was introduced at the Fall ACS meeting in New York City in 1969. A short time later, Hewlett-Packard introduced an instrument also based on the TQ mass spectrometer, which helped to legitimize the TQ technique for organic MS. Although the growth has been nothing short of spectacular for the past 30 years, the start was slow with Finnigan selling only one Model 1015/System 150 in the 18 months after the New York introduction.

Hewlett-Packard and Finnigan dominated the GC/MS market, largely owing to the simplicity of operation of the TQ and the continued advancement of the DS. The LKB 9000, which had dominated the GC/MS market especially in biotechnology, in the last half of the 1960s fell into rather quick obscurity owing to the lack of a computerized DS. The major European and Japanese mass spectrometer manufacturers of the early 1970s concentrated on magnetic-sector mass spectrometers that had GCs as inlets. These efforts were never very successful because of the higher cost of the sector instruments, and insufficient emphasis was placed on the DS. The only other serious attempt at a magnetic-sector-based GC/MS system was made by DuPont with an instrument using a fixed-field magnet and a variable accelerating voltage. DuPont had acquired the mass spectrometer pioneer CEC from Bell & Howell in a last attempt to make the company competitive in the GC/MS marketplace, but it failed, largely owing to problems of source arcing at the high voltages required (in excess of 12 500 V) to look at ions below m/z 43.[66]

Hewlett-Packard and Finnigan continued to emphasize the DS as being as significant as the mass spectrometer; however, Hewlett-Packard began to look at the technique

as one for the chromatographer rather than the mass spectrometrist. This approach led to the development of the benchtop GC-MS, which has dominated the market since 1982 with the introduction of the HP 5970. This benchtop GC-MS was followed by the introduction of an IBM PC-based DS for the Finnigan benchtop ITD™ 700 in 1984. Hewlett-Packard introduced the PC-based DS for the 597X family of instruments (5970, 5971, 5972, and 5973) in 1989.[67] Currently, Hewlett-Packard has more GC/MS units in service than all other types of mass spectrometers combined, and the TQ m/z analyzer has long ago replaced the magnetic-sector m/z analyzer as the mass spectrometer of choice for GC/MS.

After Finnigan and Hewlett-Packard became dominant in the field of GC/MS, other companies have developed popular products such as the Shimadzu QP 5050 and the Perkin-Elmer Qmass and Turbomass, both based on the transmission quadrupole. Finnigan introduced the ITD™ 700 in 1984 based on the QIT m/z analyzer. Through a technology purchase and an initial OEM agreement, Varian advanced Finnigan's internal ionization QIT to the point of being a dominant player in today's GC/MS market. More recently, LECO Corporation has introduced a rapid acquisition GC/MS system based on time-of-flight mass spectrometry (TOFMS) technology developed at Michigan State University that will acquire 1000 spectra per second over an m/z range of 1000 and will allow for greatly reduced analysis times. Micromass/Waters introduced a GC/TOFMS system at the 1999 Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Orlando, FL) that has the conventional EI and CI sources in addition to an FI source. This instrument has the capability of exact mass measurements. The magnetic-sector-based mass spectrometer has returned to the GC/MS field with a reverse-geometry, double-focusing, exact-mass-measurement 'benchtop' instrument introduced in the mid-1990s – the JEOL GCmate.

Another contributing factor to the success of GC/MS was the development of CI,[68] a technique that imparted far less energy to the analyte molecule when ionized in the gas phase, in the mid-1960s. CI is a process where a reagent gas present in the mass spectrometer's ion source at a relatively high pressure [0.1–1 Torr (in the internal ionization quadrupole ion-trap (I^2QIT) mass spectrometer, this type of CI is carried out in a time domain at reagent gas partial pressure of $\sim 10^{-5}$ Torr)] is ionized by a beam of high-energy electrons. Through this initial EI of the reagent gas and subsequent reactions of the resulting ions with reagent gas molecules, reagent ions are formed (Figure 11) that will then ionize the analyte through one of several mechanisms (Figure 12) resulting from ion–molecule reactions. In many cases, this will produce an ion that is directly related to the intact molecule (e.g. a protonated molecule). This information can be complementary to an EI mass spectrum, especially when there is no molecular ion peak in the EI spectrum.

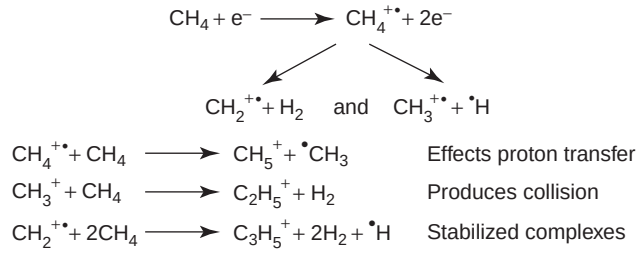

Figure 11 EI of methane followed by fragmentation of the molecular ion. The methane molecular and fragment ions then react with molecules of methane (ion/molecule reactions) to produce various types of methane reagent ions used to bring about the CI of an analyte.

The technique of MS/MS, developed in the mid-1970s, has been very important in GC/MS.[69] The triple quadrupole using an RF-only TQ position between two traditional scanning TQ m/z analyzers[39] or tandem-in-time analyses with the QIT m/z analyzer[70] has allowed for more and more accurate information about analytes that can be ionized in the gas phase using GC/MS.

2.4 Liquid Chromatography/Mass Spectrometry – A Slow Start – But a Fastcomer in the Backstretch

LC/MS is a more-encompassing term than GC/MS. In GC/MS, samples are introduced into the mass spectrometer as analytes in the gas phase from either a packed or capillary column (or one of the several iterations of the capillary column), where they are ionized. In LC/MS, analytes can enter the mass spectrometer in solution from an analytical (usually 4.6 mm ID) or capillary ($\leq$2 mm ID) LC column via flow injection, or through direct infusion of a liquid that flows because of a pump [e.g. microelectrospray (microES)] or due to electro-osmotic flow [e.g. nanoelectrospray (nanoES)], or as components from capillary electrochromatography or capillary electrophoresis. In some cases, these analytes are ionized in the mass spectrometer, and in others (e.g. ES or one of its adaptations: microES or nanoES), they enter the mass spectrometer as solvated ions. Whereas in GC/MS the chromatographic mobile phase is an inert single component (usually helium or hydrogen) and the only concern is its contribution to the pressure inside the ion source, the mobile phase in LC/MS is often a complicated mixture of solvents and nonvolatile substances, some of which may be present to react purposefully with the analytes to improve chromatographic separations but interfere with the formation of gas-phase ions. All of this makes LC/MS a much more complex issue than ionization of gas-phase analytes.

Charge transfer (the reagent ion, missing an e^-, takes an e^- from the analyte)

$$CH_4^{+\bullet} + RH \longrightarrow RH^{+\bullet} + CH_4 \qquad M^{+\bullet}$$

Proton transfer (most common when the analyte molecule has a higher proton affinity than the reagent gas. The analyte takes H^+ from the reagent ion)

$$CH_5^+ + RH \longrightarrow RH_2^+ + CH_4 \qquad [M+1]^+$$
$$C_2H_5^+ + RH \longrightarrow RH_2^+ + C_2H_4$$

Hydride abstraction (the reagent ion has a high hydride affinity, the ability to remove a H^- from the analyte molecule)

$$CF_3^+ + RH \longrightarrow R^+ + CF_3H \qquad [M-1]^+$$
$$C_2H_5^+ + RH \longrightarrow R^+ + C_2H_6$$

Collision-stabilized complexes (occurs when the PA of the analyte and reagent gas are comparable. The reagent ion becomes attached to the analyte. When methane is used, this series, $[M+1]^+$, $[M+29]^+$, and $[M+41]^+$, is a very good confirmation of molecular weight)

$$C_2H_5^+ + RH \longrightarrow (C_2H_5 : RH)^+ \qquad [M+29]^+$$

$$C_3H_5^+ + RH \longrightarrow (C_3H_5 : RH)^+ \qquad [M+41]^+$$

Figure 12 Types of ion/molecule reactions that result from positive-ion CI.

GC/MS was an established technology within the first 5 years after the introduction of the first commercial instrument – the Finnigan Model 1015/System 150 in 1969. By the end of the 1970s, GC/MS dominated the field of MS, and the TQ mass spectrometer was an accepted technology. LC/MS has been a different story. The interface posed a more difficult problem. Dealing with a gas load of 20–30 mL min^{-1} (typical carrier gas flow rates for packed columns used with early GC/MS instruction) was very different from 550 mL min^{-1} of gas (at atmospheric pressure and room temperature) produced from a flow of 1 mL min^{-1} of acetonitrile eluting from an LC column. It was also important to the early developers of LC/MS that the spectra be produced by EI. The EI spectra requirement meant that the analyte had to be in the gas phase at some point; therefore, the technique would have been limited to analytes that were not in an ionic form in the LC eluate.

The earliest attempts were made by Patrick Arpino and Fred McLafferty with a technique called 'direct liquid introduction' (DLI)[71] (Figure 13). These attempts were carried out by using EI and the relatively new technique of CI. Attempts were made to use the volatilized mobile phase of the LC process as a reagent for a CI process.

Following attempts at the DLI (which were never very successful from a commercial standpoint) came the moving-belt interface (Figure 14), pioneered by William McFadden at Finnigan Corporation.[72] This technique did not offer a great deal more usable data than DLI; however, it was commercially much more successful because the ion-source pumping requirements were much less. The limitation of a somewhat volatile analyte remained.

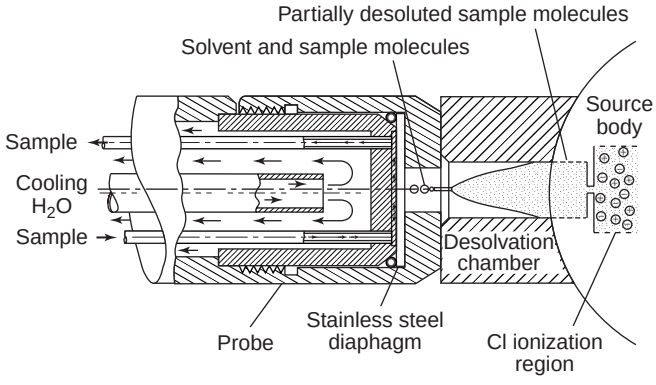

Figure 13 A schematic illustration of a direct-liquid-inlet probe for LC/MS. (Courtesy of Hewlett-Packard (now Agilent Technologies), Palo Alto, CA.)

A technological success for an interface between a LC and a mass spectrometer was the original work on API done by Evan Horning's group at the Baylor College of Medicine (Houston, TX) in the mid-1970s.[73] This technique involved the formation of ions of the intact analyte through a CI process that was cried out at atmospheric pressure. At the time of this development, not much acceptance was forthcoming because only molecular weight information was generated. The customary fragmentation data used for structure elucidations was missing. Horning called this technique 'API'. API is the term now applied to the two techniques involved with

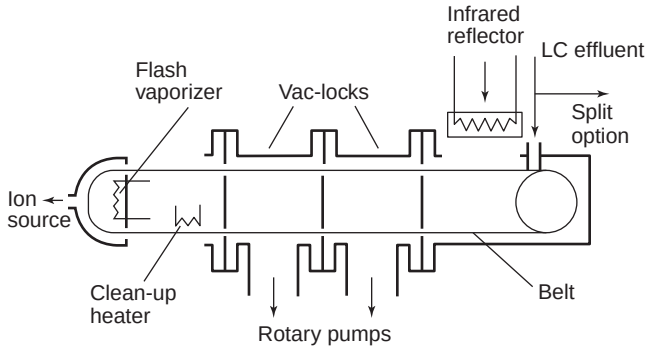

Figure 14 An illustration of the moving-belt LC/MS interface. (Courtesy of Finnigan Corporation (now ThermoQuest Corporation), San Jose, CA.)

getting ions into the gas phase at atmospheric pressure: APCI and ES.

The APCI technique requires the analyte to be volatile enough to allow for a gas-phase ion–molecule reaction to take place, which means that the analyte cannot be in the ionic phase in the LC eluate. This volatility requirement, coupled with the fact that for the most part only molecular weight information resulted, led to an initial limited use. The technique required a lot of pumping that added to the cost and complexity of commercial instrumentation. APCI did not become popular until after the development of the triple-quadrupole mass spectrometer, which allowed for MS/MS of these ions representing the intact molecule, thereby producing structural information about the analyte. APCI also became a good companion for ES because both were atmospheric techniques and used similar interfaces to the mass spectrometer.

Today, APCI and ES dominate LC/MS instrumentation. The only other LC/MS technique that was commercially viable at the end of the 20th century was the particle beam interface developed by Ross Willoughby and Robert Browner in the mid-1980s.[74] This technique allows for acquisition of EI and CI mass spectra of analytes in the gas phase after a partial desolvation using a heated drying gas and momentum separator. Clumps of moist analyte molecules are propelled onto a heated plate located in the ion source of the mass spectrometer. These moist clumps striking the heated plate result in the volatilization of the analyte molecules, allowing for their subsequent ionization in the gas phase. Although this technique can yield more structural information about the analyte, it has much poorer sensitivity than ES or APCI and has the limitation of the volatile requirement for the analyte.

During the mid-1980s, two different independent approaches were made to LC/MS that used the DI technique FAB by Y. Ito[75] and Richard Caprioli.[76] These techniques were called 'frit-flow FAB' and 'continuous-flow FAB', respectively. Again, these were labor-intensive techniques and were far from allowing for routine analyses. However, at the time, this was the only technique that allowed for the analysis of nonvolatile thermally labile LC analytes such as polysaccharides, peptides, and oligonucleotides. These iterations of FAB have been replaced by ES.

Another abortive attempt at LC/MS was a technique called 'thermospray'. Thermospray was pioneered by Marvin Vestal,[77] currently at PE Biosystems and at VesTech and the University of Houston at the time this technique was developed, in the early 1980s. The original form of this technique (filament-on) was similar to the DLI. Later, a technique was serendipitously developed called 'filament-off' for the analytes that were in an ionic form in the LC eluate when ion current was observed after a failure of the filament. Even though the thermospray technique was also labor intensive and fraught with many difficulties, it had a reasonable commercial success in that investigators were desperate for tools to analyze substances separated by LC.

Beginning in the mid-to-late 1980s with publications by John Fenn[78] and Jack Henion,[79] the technique of ES started to evolve. After Fenn, Mathias Mann, and C.K. Meng published an article entitled 'Protons or Proteins – A Beam's Beam'[80] in 1988, ES and its companion technique of APCI underwent a logarithmic growth in improvement, numbers of applications, and advancements in the commercial instrumentation. ES went from being a research-laboratory curiosity to one of the most widely used tools in biotechnology.

Techniques such as continuous-flow fast atom bombardment (CF-FAB), currently being developed, will take LC/MS into expanded areas of use in organic chemistry and biotechnology.

2.5 Desorption/Ionization

DI is the term applied to the processes by which mass spectrometrists '... make elephants fly'.[81] DI is a technique in which ions are formed in other than the gas phase, and the ions are then desorbed from the matrix where they are formed into the gas phase. Currently, the two most widely used and popular DI techniques are MALDI and ES. However, although not as widely employed, the techniques of FI and FD and also FAB are still in use. A new TOF-based GC/MS system was introduced by Micromass at the 1999 Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Orlando, FL), which had an FI source available as an option.[82] It is still common to find abstracts in almost every issue (once every 2 weeks) of *CA Selects Plus MS* (a publication of the Chemical Abstracts

Service of the ACS) on the use of FAB or CF-FAB in LC/MS applications.

Another important DI technique that is still widely used is DCI. This probe technique is especially suitable for use with TQ mass spectrometers. Vinca Promacvich, organic analytical facilities manager in the Department of Chemistry, Columbia University, says, ''When these organic chemistry students bring samples ('pure' or mixtures for GC/MS analysis), the first thing I do is try them [the samples] with DCI. If I do not get usable spectra, then I will not take the chance of ruining a GC column or contaminating the ion source by trying the analysis of something that is probably a crystallized inorganic salt rather than the suspected reaction product.''

2.5.1 Field Desorption and Field Ionization

The first DI technique to be seriously considered in MS was FD. FD is an ionization technique by which analyte ions are desorbed from the surface of one of the two electrodes used to produce an electrical field. This technique was the first to allow the determination of the mass of analytes that were nonvolatile or thermally labile. FD uses specially prepared sample emitters that allow for the production of protonated molecules (MH^+). Because single-charge ions are produced, the mass limitations at the time of the development were determined by the maximum m/z value of the mass spectrometer. Formation of ions via an electrical field had its origin with field-ion microscopy developed in 1951 by E.W. Müller in Germany. The first developments involving FD were by R. Gomer and M.G. Ingrahm.[83,84]

The technique of FI of gas-phase analytes was an outgrowth of the development and study of FD. FI is a technique, rivaling CI, that results in a high abundance of ions representing the intact molecule. Unlike CI, which produces even-electron protonated molecules (MH^+), FI produces odd-electron molecular ions ($M^{+\bullet}$). In FI, ionization of an analyte molecule in the vapor phase takes place in an electrical field (107–108 V cm^{-1}) maintained between two sharp points or edges of two electrodes. The technique of FI was developed by H.D. Beckey at the Institut für Physikalische Chemie der Universität Bonn (Bonn, Germany) in 1957.[85]

Becky has gone on to show additional developments in both the FD and FI techniques. His 1977 book *Principles of FI and FD MS* is considered the seminal reference for these techniques. According to a statement made by Ronald D. Macfarlane, developer of the Cf DI technique, ''FD opened the door for biological MS by demonstrating feasibility.''

FD is carried out with sample deposited from solution on tungsten filaments that have been activated by growing a dense matrix of needles of graphite on its surfaces. The process of preparing the filaments and introducing the sample without disturbing this delicate forest of needles is an art form in and of itself. Although FD was used to a large extent as the only technique for nonvolatile thermally labile analytes from its development in the 1960s, it was largely replaced with other less-tedious techniques such as FAB, DCI, and now MALDI and ES.

2.5.2 Plasma Desorption (^{252}Cf Desorption/Ionization)

The next DI technique to be developed was that pioneered by Ronald D. Macfarlane at Texas A&M. For the most part, the technique of ^{252}Cf DI (Cf DI) [or plasma desorption (PD)] remained a laboratory curiosity through its useful life. The dependence of PD on TOFMS was not sufficient to prevent the last US manufacturer of these instruments (CVC Corporation) from discontinuing their manufacture in the late 1970s. There was a single commercial attempt at a PD instrument in the 1980s by Bio-Ion in Uppsala, Sweden. The instrument was later marketed by Kratos Analytical, UK. Bio-Ion was later acquired by Applied Biosystems (now PE Biosystems).

The primary reason for the lack of popularity of Cf DI was the radioactive nature of the ionization source, which presented a significant safety and disposal problem for a number of laboratories.

PD was first reported in 1974 with results obtained from the nonvolatile compound sodium acetate. This report predicted the use of PD for larger nonvolatile molecules such as biopolymers.[86] Two years later, Macfarlane's group reported the analysis of a peptide with a molecular weight of 1881 Da.[87]

PD is accomplished by putting the analyte sample on one side of a nickel or aluminized Mylar® plate. On the other side of the plate is a 10 μCi ^{252}Cf source. Each fission event results in two products of nearly equal mass, each having energies of 80–120 MeV propelled in opposite directions. One triggers the start of the detector and the other rips through the sample holder to desorb intact adduct ions of the analyte molecule ($[M + H]^+$, $[M + Na]^+$, etc.). The ions are captured by the ion optics of a specially designed TOFMS instrument and are accelerated into its flight tube. Mass spectra are accumulated over a period of several hours depending on the signal strength of each event and the required signal-to-background for the analysis.

Macfarlane also developed the method of ES for the deposition of the sample on the PD sample holder.[88] Peter Roepstorff (Odense University, Denmark) became an outspoken proponent of the technique in support of Bio-Ion.[89]

The DI technique that showed mass spectra could be obtained from large peptides and proteins as well as other biologically significant substances such as

oligonucleotides was Cf DI, which was the genesis of bioanalytical MS. This DI technique is no longer in use, even in university research laboratories.

2.5.3 *Fast Atom Bombardment*

In an effort to use MS for the analysis of proteins, investigators attempted to get results by layering the purified protein onto a metal surface and then used a SIMS technique (see section 2.7.3) to desorb ions from the surface of this open-face sandwich. This was not very effective because the energy absorbed by the sample often resulted in sample decomposition. When secondary ions were produced using lower-energy bombarding ions to avoid damage to the top layer of the analyte and prolong the secondary ion current, the analytical ion current was significantly reduced resulting in poor detection limits. Also, the SIMS technique resulted in an ion current of a very short duration, which also affected the detection limit.

In a further attempt to obtain better results from the 'sputtering' mass spectral ionization techniques, Mickey (Michael) Barber in the Department of Chemistry at the University of Manchester Institute of Science and Technology in the mid-1970s began to look at the possibility of using accelerated atoms to bombard the sample as opposed to ions.[90] The decision to use atoms as opposed to ions was born out of the desire to use a double-focusing mass spectrometer that required ion acceleration of ~8 kV. This technique was effective but, for the targeted organic analytes, suffered from the same short duration of ion current that had been observed with SIMS. When Barber's group undertook this research, they had set the goal of developing a new ion source that met the following criteria: (1) produces ions at room temperature, thereby eliminating the necessity for volatilizing the analyte prior to ionization; (2) the ionization process should result in both the abundance of molecular ions to allow for molecular weight determination and fragment ions that can be used to deduce the analyte's structure; (3) the mass spectrum resulting from this ionization process should have a sufficient lifetime to allow for techniques such as metastable ion detection, which can be time-consuming in a double-focusing mass spectrometer; (4) the detection limit should be such that structural information can be obtained from as little as 1 μg of analyte; and (5) the ionization technique should have the potential for easily and reproducibly generating mass spectra of analytes that have proved problematic or impossible by other types of ionization.

After half a decade of research and based on curious observations of sample contaminants by pumping fluids (e.g. Apiezon oils, Santovac 5, and Convalex 10) and glass–glass vacuum sealants such as siloxane stopcock greases,[91] a way to prolong the ion current was found

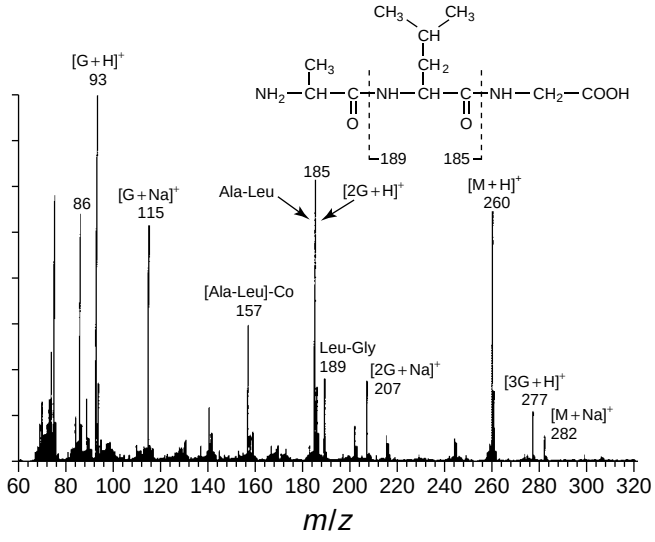

Figure 15 FAB mass spectrum of the tripeptide alanylleucylglycerine (molecular weight 259) produced with 7 KeV xenon atoms. While smaller than the peptide originally analyzed by Barber et al.[92] (Tyr-Gly-Gly-Phe-Met, molecular weight 573), this spectrum illustrates the principle of FAB. Note the peaks representing ions that result from sodium addition to the analyte molecule. (Courtesy of B.L. Ackermann in the NIH Mass Spectrometry Facility at Michigan State University, East Lansing, MI.)

by mixing the analyte with a low-vapor-pressure viscous solvent. The first material found to be satisfactory for this technique was glycerine (glycerol) used to obtain the mass spectrum of a small peptide (methionine enkephalin, Figure 15).[92]

The genesis of FAB is well chronicled in an *Analytical Chemistry* A-pages article that appeared in 1982.[93] FAB was easily adaptable to existing TQ and double-focusing mass spectrometers although the best results were obtained using the sector-based instruments. FAB met all of Barber's objectives for a new ionization technique and was readily adopted by the MS community. A Boolean search of MEDLINE (http://www.infotrieve.com/freemedline/) for 'FAB' yielded 6 papers in 1981, 69 in the period 1981–1984, and over 325 in the period 1985–1989. This last number is far from the total number of publications that appeared using this powerful technique during that period. FAB also found immediate applications in the technique of MS/MS using both TQ and reverse-geometry double-focusing instruments (see section 2.6). FAB differed from either FD or PD in that it was easy to prepare sample, easy to use, and spectra could be obtained in a reasonable period rather than the hours or days required for PD. As soon as fragmentation spectra were being obtained from peptides, schemes for their interpretation were developed. FAB allowed for the more

realistic characterization of microorganisms by MS that was pioneered by Catherine Fenselau at the University of Maryland, Baltimore County.[94]

As a variation of FAB, the technique of liquid secondary ion mass spectrometry (LSIMS) was developed. Instead of the beam of neutral atoms being fired at the mixture of analyte and carrier, a beam of ions is used. For the most part, the same results are obtained. The LSIMS technique imparts more energy to the analyte. As described above under LC/MS, FAB proved to be a useful technique with dynamic separations as well as one that allowed for the analysis of purified samples. FAB was still in use at the end of the 20th century; however, it has largely been replaced by MALDI and ES.

2.5.4 Desorption Chemical Ionization

As MS developed, techniques evolved to introduce sample in all three physical states of matter (gas, liquid, and solids). Some solids had sufficient volatility that results could be obtained by putting the material on the end of a probe placing the probe tip in the mass spectrometer's ion source and heating the probe to produce a vapor that was ionized. In some cases, this proved to produce better quality spectra than could be obtained by volatilizing the sample outside of the ion source.[95] In 1973, Fred McLafferty introduced the technique of placing a sample on a probe and heating it while subjecting it to an ion plasma produced in a CI source by showing a spectrum of a nonderivatized peptide (VAAF) with a large protonated molecule $(M + H)^+$ peak and other peaks that were used to determine the sequence.[96] This variation of the technique used to introduce samples on a direct-insertion probe (or solids probe) has been called 'direct CI', 'in-beam', 'direct exposure', and 'desorption CI'.[97] The last has been the accepted name for a number of years.

Desorption electron ionization (DEI) has also been used. The probe used for DCI and DEI is called a 'direct-exposure probe'. The direct-exposure probe is distinguished from the solids probe in that the sample is placed on the outer surface of the probe, whereas the sample is placed in a hollow cavity at the tip of the solids probe. This hollow cavity is often similar to a short (5–10 mm long) melting-point capillary tube.

Today's direct-exposure probes are constructed with a filament that can be rapidly heated. The physical shape, material of construction, and treatment of these filaments all contribute to the results obtained. A sample is dissolved in an appropriate volatile solvent, a drop of the solantoin is placed on the filament so that the filament is fully and evenly coated, and the solvent is then evaporated (Figure 16). The filament on the end of the probe is passed through the vacuum lock into the ion source of the mass spectrometer. A current is passed through the filament, which results in rapid heating. DCI takes place when a CI plasma is present, and DEI results when only a beam of electrons is used for ionization.

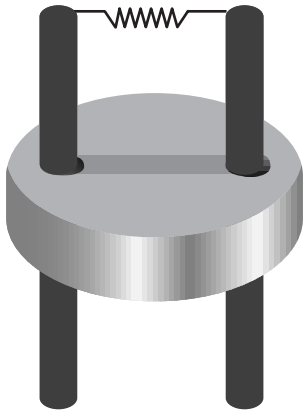

Figure 16 Filament used with the DCI probe of the Nermag R10-10 series of TQ mass spectrometers. (Courtesy of Scientific Instrument Services, Inc., Ringoes, NJ.)

The exact mechanism of this ionization has never been clearly explained. Evidence has been shown that indicates the intact analyte is desorbed from the surface of the filament, and ionization takes place in the gas phase.[98] This evidence supported an approach that used rapid heating. However, the difference in ion current obtained with different reagent gases indicated that the ionization was taking place on the surface of the filament, and the ions were then desorbed into the gas phase.[99] DCI is especially applicable to use with TQ mass spectrometers. It is unclear whether it was a combination of filament physical design and manufacturing material or factors pertaining to ion-source potentials, but results obtained using TQ instruments appeared to be significantly better than those obtained with sector-based instruments. DCI will produce mass spectra from nonderivatized sugars and other organic compounds that are thermally labile.

DCI produced spectra for many compounds that had been unsuccessfully ionized by FD. One of the more significant examples was the mass spectrum of brevetoxin B obtained by Koji Nakanishi at Columbia University in the mid-1980s on a Nermag R10-10 TQ mass spectrometer (Figure 17).[100] The Nermag R10-10's DCI probe used the coiled filaments shown in Figure 16. These filaments were constructed of a thoriated tungsten. This instrument, using this particular design, produced some of the best DCI spectra obtained on a commercial instrument. This particular technology is still commercially available through ThermoQuest's Automass Multi combination GC/LC/MS system.

DCI was developed and matured in the same time frame as FAB. Both techniques were amenable for

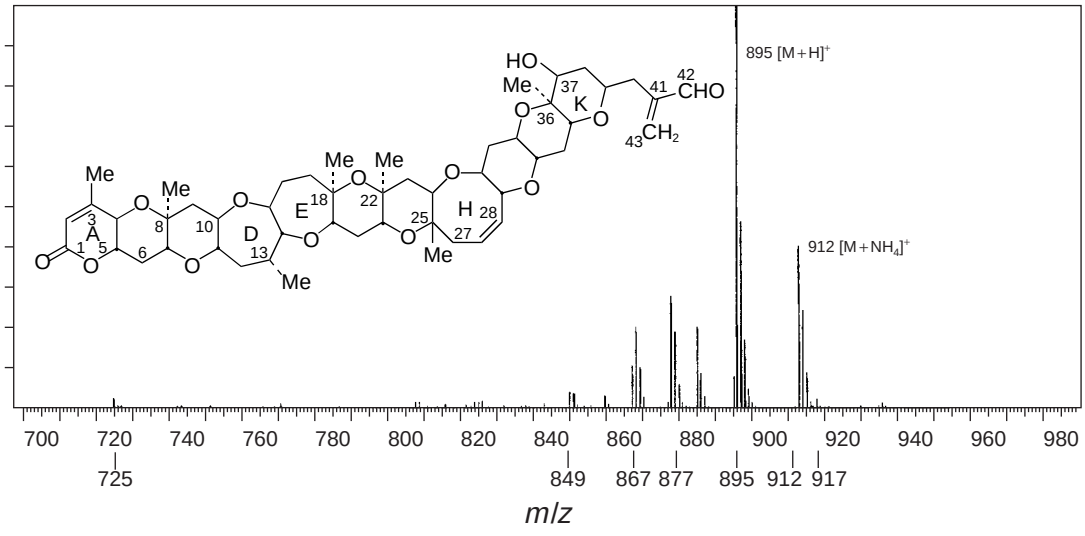

Figure 17 DCI of mass spectrum of brevetoxin B obtained by Koji Nakanishi at Columbia University in the mid-1980s on a Nermag R10-10 TQ mass spectrometer. (Courtesy of Koji Nakanishi.)

compounds that were thermally labile and/or nonvolatile. DCI has been a very good technique for use with TQ mass spectrometers, which are more limited in their maximum m/z detectability than double-focusing instruments. FAB is the preferred technique for proteins and peptides, whereas DCI works nicely with 1–5 saccharide sugars and coordination complexes.

2.5.5 Matrix-assisted Laser Desorption/Ionization

MALDI is a technique by which an analyte is mixed with an energy-absorbing matrix. Analytes are usually of high molecular weight (>2 kDa), and the matrix is usually of a much lower molecular weight (~0.5 kDa or less). Each is separately introduced into a shallow well on a flat plate in solution, where the two components are mixed through the natural mixing of the solutions. The solvent is evaporated leaving a solid mixture of matrix and analyte. This mixture is then subjected to a blast from a laser [both IR and UV lasers have been used]. The matrix molecules absorb the energy of the laser shot (Figure 18a–c). The energized matrix molecules are then responsible for the formation of protonated molecules that are passed into the gas phase, partly due to the ablation of the sample caused by the impact of the laser. This technique was found to produce ions of intact molecules that were mostly single-charge ions, but there is also the possibility of double- and triple-charge ions. The mass spectrum can also contain peaks representing dimers and trimers of the analyte (not very common). MALDI was developed by two German scientists (Franz Hillenkamp and Michael Karas) working at the Institute für medizinische Physik und Biophysik, Universität Münster (Münster, Germany).[101]

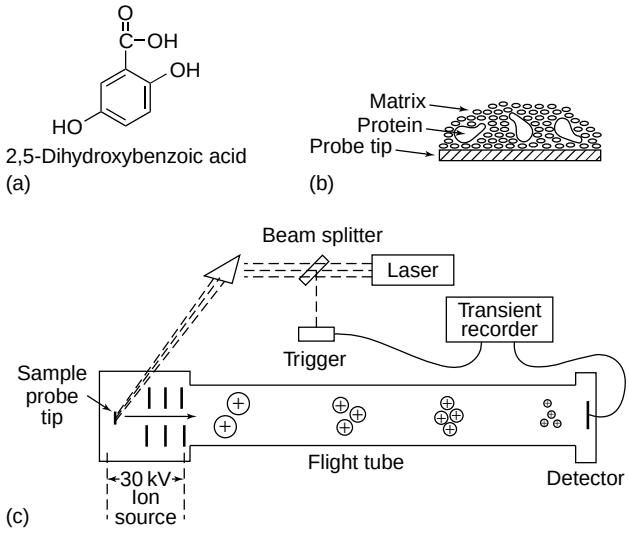

Figure 18 An illustration of the MALDI principle: (a) the chemical structure of a common matrix molecule for MALDI; (b) an artist's conception of the cocrystallized decomposition of a mixture of matrix molecules and protein analyte on a probe tip; (c) illustration of the operating principles and a typical configuration of a MALDI instrument with a TOF mass spectrometer.[7]

MALDI allows the detection of ions representing intact proteins and other biopolymers that had masses in the tens to hundreds of kilodalton range. The fact that MS could be used to determine the mass of such substances was the good news. The bad news was that most of the mass spectrometers currently in use at the time MALDI was developed could not be used to separate ions of such high mass-to-charge ratios as experienced in this type of analysis. The upper m/z limit of the sector-based mass

spectrometer is a function of the maximum magnetic field that could be produced and still be stable. The instruments that rely on RF (TQ and QIT mass spectrometers) are limited by the maximum amplitude that could be obtained for a given frequency before the wave will break up. The only mass spectrometer that was not affected by one of these limits of the physics involved with the ion separation is the TOF mass spectrometer. Unfortunately, TOFMS instruments had fallen into the realm of an historical curiosity by the time MALDI was developed. Just a few years earlier, the only manufacturer of TOFMS instruments had stopped their production. There was little or no interest on the part of research laboratories in the USA or Europe in developments of TOF. There was still a fairly active interest in TOFMS in what was then the Soviet Union (now Russia and some of the countries just to the west of Russia). The power and advantages of MALDI resulted in reassurances in TOFMS to the point that many new applications have been developed for this MS technique. At the end of the 20th century, TOFMS was not only important to MALDI MS but also found uses in inductively coupled plasma mass spectrometer (ICP-MS), GC/MS, ES and APCI LC/MS, and SIMS used in surface analyses.

FAB was the gleaming new technology that dominated MS throughout the 1980s. MALDI has been the technique of the 1990s and, along with ES, has largely replaced FAB even though *Analytical Chemistry* did publish a review of commercially available FAB instrumentation as late as October 1997.[102]

2.5.6 *Electrospray*

Unlike all the previous types of DI, ES is unique in that the mass spectrometer is not necessarily involved in the ionization process. All the other DI techniques involve the formation of ions in the condensed phase by some process in the ion source of the mass spectrometer followed by the desorption of these ions into the gas phase. In ES, the analyte enters the mass spectrometer in solution already in an ionic state. Although the exact mechanism of ES is still widely debated,[20] it is generally agreed that solvated ions in charged droplets are desolvated in the interface of the mass spectrometer, which is why some people do not call the process ES 'ionization'. If ionization is a process of forming ions, then ES should not be called ionization; if ionization is a process by which ions are put into the gas phase, then ES is an ionization process.

The technique of ES, popularized in the last decade of the 20th century, has a long history that goes back to the 1930s.

2.6 Types of Mass (m/z) Analyzers

The part of the mass spectrometer used to separate gas-phase ions according to their m/z values is the mass analyzer. Mass analyzer is the traditional terminology. This term was satisfactory when used with instruments that existed before the 1990s because almost all encountered ions that had only a single charge. The term mass and m/z were used interchangeably, which led to labeling the x-axis of spectra with terms that were units of mass. After the popularization of ES, the multiple-charge ion became more commonplace; and, while always problematic, labeling the x-axis of a mass spectrum with unit-of-mass terms became a potential for conveying incorrect information. As an example, the DS accompanying an LC/MS instrument featuring an ES interface introduced in 1996 had the x-axis of all the spectral displays labeled 'amu'. Amu was not only an incorrect term (owing to its replacement with the symbol u in 1962) for mass, but the use of any term or symbol for a unit of mass on an axis that should have been labeled m/z, especially for spectra that contained peaks primarily representing multiple-charge ions, was also unconscionable.

Mass (m/z) analyzers can be divided into two broad categories: (1) those that in some way isolate ions of individual m/z values from a beam – beam-type instruments; and (2) those that store ions of all m/z values and detect ions through some process of single m/z isolation – traps. Magnetic-sector, double-focusing, TQ, and TOF mass spectrometers are beam-type instruments. QIT (both external and internal ionization variations) and ICR mass spectrometers are traps.

2.6.1 *Magnetic-sector Mass Spectrometer*

Magnetic-sector mass spectrometers use only a magnetic field to separate ions according to their m/z values (Figure 19). These instruments are referred to as single-focusing mass spectrometers. They are capable of separating ions that differ in one m/z unit over a range from 1 to $\sim$700 m/z. These types of instruments have been used as residual gas analyzers (RGAs) for analysis of elements, as isotope-ratio mass spectrometers, and as general analytical mass spectrometers for organic compounds. As one of the first mass spectrometers designed specifically for use in GC/MS, the LKB 9000 was based on a single-focusing magnetic sector. Owing to the inability of these instruments to acquire sequential mass spectra rapidly because of the hysteresis inherent in electromagnets and the complexity of the electronics resulting in a high initial cost and difficulty of operation, these types of instruments quickly dropped from commercial popularity for use in hyphenated techniques such as GC/MS when the TQ GC-MS was introduced in the late 1960s.

The explosion of new mass spectrometer manufacturers that came about after World War II resulted in several popular models of single-focusing magnetic-sector instruments. These instruments included the AEI (name

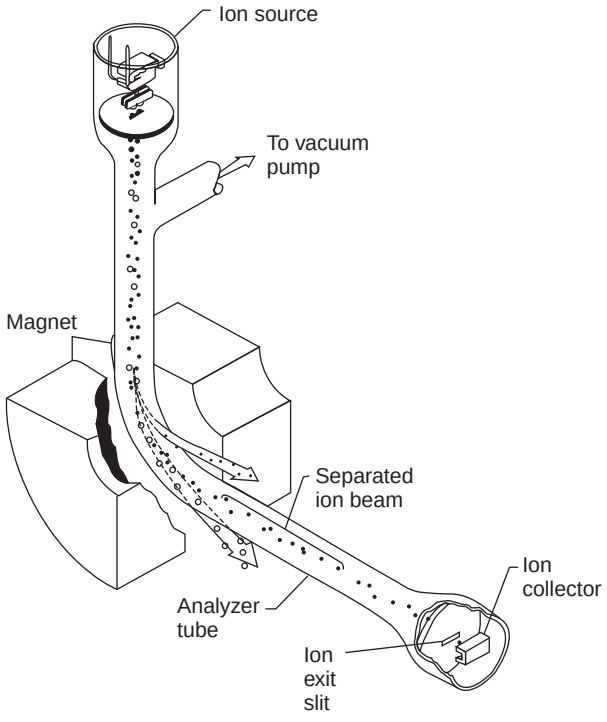

Figure 19 A graphic illustration of a single-focusing magnetic-sector mass spectrometer.

of Metropolitan Vickers parent company, Associated Electrical Industries, after 1960; see below) MS-2, MS-3 and MS-5; the MAT-Atlas CH-4 and CH-5; and the CEC 21-103C and 21-130.

2.6.2 Double-focusing Mass Spectrometer

Double-focusing mass spectrometers use a magnetic field to select ions based on their m/z values and an electric field to select ions based on their energy. These instruments became the workhorse of MS from the 1930s through the end of the 1970s. These instruments are capable of separating ions with very small differences in m/z values allowing for the determination of the elemental composition of the ion based on these millimass measurements. CEC was the first commercial manufacturer of double-focusing mass spectrometers beginning before World War II. After World War II, manufacturers and instrument designs sprung up in the UK (Metropolitan Vickers delivered its first instrument in 1946 in Manchester, UK); Germany [Mes und Analysen-Technik (MAT) founded in 1947 in Bremen, Germany, as a subsidiary of Atlas-Werke AG]; and Japan (JEOL and Hitachi having existed as analytical instrument manufacturers before the war). At the beginning of the 21st century, a spinoff from Met Vickers (VG Analytical, now Micromass, Manchester, UK, a wholly owned subsidiary of Waters Corporation), JEOL, and the current owner of MAT (ThermoQuest) are the only commercial manufacturers of double-focusing mass spectrometers. Of these three manufacturers, JEOL is the only company that continues to have this type of mass spectrometer as its only mass spectrometer offering.

Names of the developers of the ion-optics systems used in double-focusing mass spectrometers are often associated with the instrument [e.g. Neir–Johnson (Figure 20), Mattauch–Herzog (Figure 21), and Bainbridge–Jordan].

The technique of MS/MS was developed as a result of forced metastable[103] ion formation between the two sectors of the double-focusing mass spectrometer. By placing (Figure 22) the magnetic sector as the first field following the ion source (reverse geometry) as opposed to the traditional configuration of having the electric sector as the first field (forward geometry), routine MS/MS experiments could be carried out.[104] This MS/MS technique remained popular until the development of the triple-quadrupole mass spectrometer (two TQ m/z analyzers separated by a third quadrupole operated in the RF-only mode used as a collision cell). MS/MS in the double-focusing mass spectrometer is an important selling feature of the JEOL LCMate and GCMate at the beginning of the third millennium. Although the TQ instruments remain dominant in MS/MS, the QIT is gaining in importance to this technique.

Many of the often-cited scientists[6,10] in MS had a great deal to do with the development of double-focusing

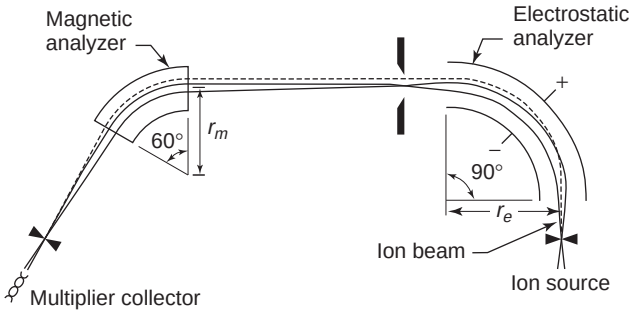

Figure 20 The ion optics of a Nier–Johnson double-focusing mass spectrometer with forward geometry.

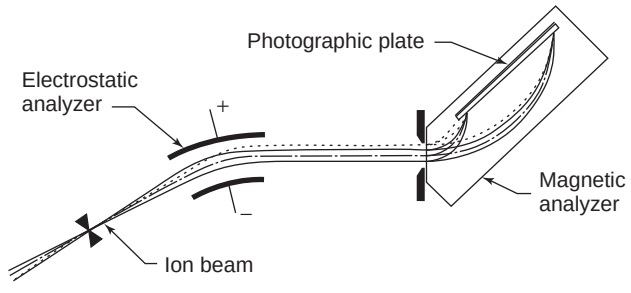

Figure 21 The ion optics of a Mattauch–Herzog double-focusing mass spectrometer with forward geometry.

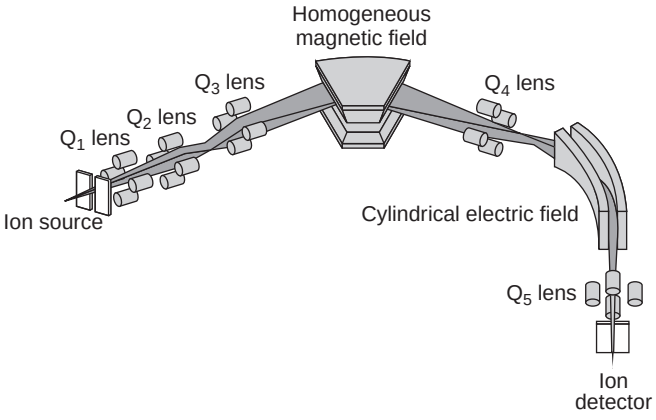

Figure 22 A graphic representation of the ion optics of the JEOL GCMate/LCMate illustrating reverse geometry. (Courtesy of JEOL USA, Inc., Peabody, MA.)

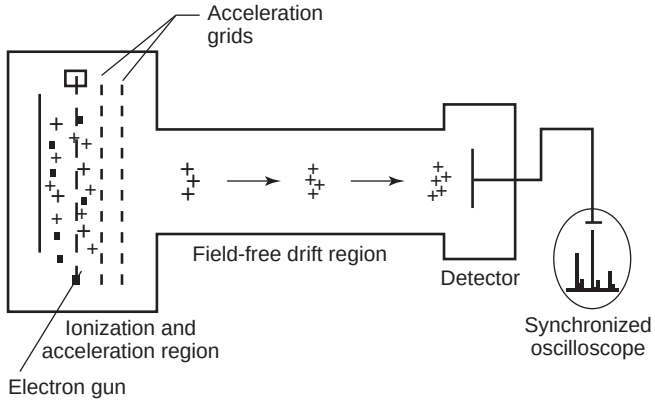

Figure 23 An illustration of the principle of TOFMS.

instruments. They were the people who were involved with applications development and ion optics. However, a number of other factors that took place behind the scenes were just as important. One of these factors includes the development of microprocessor electronics, which reduced the need for fine-tuning of the instrument by highly skilled operators. Another important factor was the development of the laminated magnet in the 1970s. This development helped to overcome the hysteresis problems with magnets that were the limiting factor in scan speed. The double-focusing mass spectrometer became a commercially viable instrument after World War II largely because of the advances in vacuum technology that grew out the use of the mass spectrometer as a preparative tool in the development of the atomic bomb.[2]

2.6.3 Time-of-flight Mass Spectrometer

The next gas-phase ion-separation technique in MS was involved with the development of the time-of-flight mass spectrometer (TOF-MS). Ions of different m/z values accelerated from a region such as an ion source into an evacuated tube will have different velocities, and therefore these ions will reach the end of this evacuated region at different times. By separating the times at which ion current is observed at a detector placed at the end of this evacuated region, it is possible to obtain a mass spectrum. Ions of the lowest m/z will reach the detector first (Figure 23).

The earliest report of experimentation with the concept of TOF for gas-phase ion separation was by William R. Smythe (US scientist) and Josef Heinvich Elizabeth Mattauch (Austrian physicist) in 1932.[105] According to an account by Robert W. Kiser in his book *Introduction to Mass Spectrometry and its Applications*,[6] the history of the early development of the TOF-MS involved a number of people.

While at the University of Pennsylvania (Philadelphia, PA), W.E. Stephens reported on the concept of the TOF-MS in an abstract appearing in the American Physical Society's program in 1946.[106] Some 7 years later, Stephens published a paper with M.M. Wolff with the same title as Stephens's 1946 abstract, 'A Pulsed Mass Spectrometer with Time Dispersion', which described the instrument they had built based on the TOF principle.[107] The seminal paper on TOFMS, 'TOF Mass Spectrometer with Improved Resolution', was published in 1955 by two researchers working for Bendix Aviation Corporation Research Laboratories in Detroit, MI.[108] This paper provides a detailed treatment of ion focusing in the TOF-MS. The Wiley–McLaren paper was considered to be of enough importance that it was republished in the *Journal of Mass Spectrometry* in 1997.[109]

According to a citation in Robert Cotter's book *TOFMS: Instrumentation and Applications in Biological Research* (ACS, Washington, DC, 1997), "... it was estimated that in 1962, one-third of the mass spectrometers in use in the United States were TOF instruments".[110] In his 1965 book, Kiser[6] states that the bulk of the US market had been provided by CEC (approximately 700 instruments). Considering these two statements, it is possible to believe that by the mid-1960s, only 10 years after Wiley and McLaren's paper, there were over 300 TOF mass spectrometers in use in the USA, and most of them had been manufactured by what later became known as the Bendix Corporation. In the mid-1960s, the future for the TOF-MS looked very bright. However, this was before the commercialization of the TQ mass spectrometer.

The Bendix Corporation was sold to Consolidated Vacuum Corporation, later known as CVC Products (Rochester, NY). Sale of TOF mass spectrometers declined to the level that CVC discontinued their manufacture sometime in the late 1970s to early 1980s. The popularity of the TOF-MS went through a period

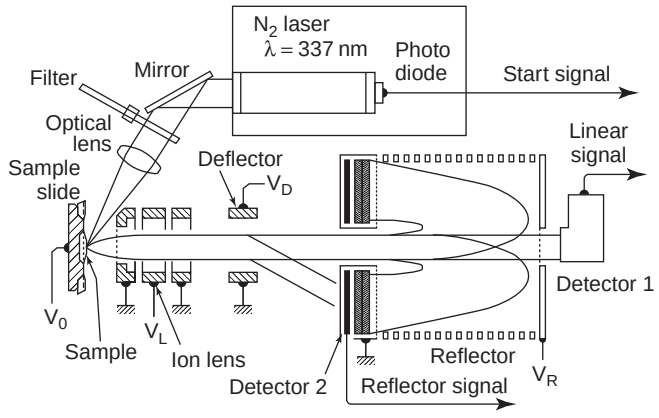

Figure 24 Ion optics of the Kratos Kompact TOFMS instrument with a curved-field reflectron for improved ion focusing over a broad range of m/z values.

of rapid decline until the development of MALDI by Franz Hillenkamp and Michael Karas[101] in the late 1980s. In the decade that followed, the renewed interest in the TOF-MS resulted in an explosion of new developments, which included greatly improved resolving power, very rapid acquisition of data, and new ionization techniques such as Aviv Amirav's supersonic beams.[111] Beginning with the introduction of commercial MALDI TOF mass spectrometers in the mid-1990s followed by the explosion of TOF instrumentation in LC/MS, the TOF-MS has re-established itself as a significant factor in the industry (Figure 24). LECO Corporation (St. Joseph, MI) is a newcomer to MS through the acquisition of a GC/MS TOF system (Pegasus II) and an LC/MS TOF system (Jaguar) where both instruments represent new paradigms in their respective hyphenated techniques. Micromass and PE Biosystems both have an LC/MS product using TOF technology and an MS/MS hybrid instrument (TQ TOF). These instruments, which became widely used at the end of the 20th century, are just the genesis of the new TOF-MS. Within the next decade, the TOF-MS could supersede the TQ as the most used instrument in MS.

2.6.4 Transmission-quadrupole Mass Spectrometer

The TQ mass spectrometer uses a quadrupole field produced by a combination of DC and RF voltages that are applied to four conducting surfaces that have a hyperbolic cross-section. The first TQ mass spectrometers used four opposing round rods arranged parallel to one another to create this field. The TQ is often referred to as a 'mass filter' because it can be set to allow an ion of a single m/z value to pass through it while removing ions of all other m/z values from an ion beam. Quadrupole technology (which includes the TQ and monopole mass filters and the QIT) was first explored by Wolfgang Paul and colleagues at the University of Bonn (Bonn, Germany) in the early 1950s.[112] The initial paper on the quadrupole technology was followed by a detailed description of the theory.[113] Paul shared half of the 1989 Nobel Prize in Physics with the German-born American physicist Hans Georg Dehmelt, the developer of the Penning (magnetic) trap, 'for the development of ion-trap techniques'. The quadrupole ion trap was an outgrowth of his study of quadrupole fields. The monopole was commercialized in the 1960s by General Electric Corporation but never became very popular. The seminal reference on the quadrupole technology is Peter Dawson's *Quadrupole MS and its Applications* (Elsevier: Amsterdam, 1976; reprinted by the American Institute of Physics, Woodbury, NY, 1995).

The TQ mass spectrometer was very well suited for use with the emerging field of GC/MS. This m/z analyzer requires a much lower accelerating voltage EI and CI ion source than the sector instruments. It was smaller in physical size and much more suitable to rapid scanning and electronic control than the sector instruments. The TQ has a practical capability of producing unit-resolution spectra over an m/z range from 0 to ~4000.

The well-established mass spectrometer manufacturers of the time (which also included VG Analytical, a company that began some years after the others) were all committed to the magnetic-sector and double-focusing technology. Two companies new to MS (Finnigan Corporation, Sunnyvale, CA, now part of ThermoQuest Corporation, Austin, TX; and the California Instrument Division of Hewlett-Packard Corporation, Palo Alto, CA, now part of the HP spinoff Agilent Technologies, Palo Alto, CA) were the first to introduce GC/MS instrumentation based on the TQ mass spectrometer and are equally responsible for success of this type of mass spectrometer. At the end of the 20th century, there were more TQ mass spectrometers in use than all other types of mass spectrometers combined. In 1999, more GC/MS and LC/MS instruments using a TQ m/z analyzer were purchased than all other types of m/z analyzers combined.

An important note about the TQ that might have fallen from the pages of history if not for an article by Robert F. Finnigan on the history of TQ is that while Paul and his colleagues pursued this technology, Richard Post at the University of California Lawrence Berkeley Laboratory carried out independent research in the 1950s on the same technology.[37] Post did not publish or apply for patents on his findings. Other than information contained in his personal notebooks, the only record of his work is a University of California Radiation Laboratory Report (UCRL 2209) published in 1953.

Robert Finnigan (whose background was electrical engineering) actually began his quest for a TQ GC-MS at a process instrument and controls group that

he and a physicist colleague, P.M. Mike Uthe (from Lawrence Livermore National Laboratory and then Stanford Research Institute, SRI), founded for Electronic Associates Inc. (EAI, Long Branch, NJ), an analog computer company.[37,114] From the founding of this group within EAI in 1964 until 1967, Finnigan and his colleagues produced over 500 RGA instruments using the TQ. TQ mass spectrometers continue to have extensive use as RGAs. This RGA experience with EAI gave the founding group of Finnigan Corporation a solid background in TQ technology. Along with several others from EAI and the then general manager of Varian's chromatography business, Varian Aerograph, T.Z. Chu started Finnigan Corporation in January 1967. The goal of this company was to produce an easy-to-use, low-cost GC-MS based on the TQ mass spectrometer.

The first Finnigan GC/MS instruments sold for about $35 000, which was far less than any of the sector-based instruments. The TQ mass spectrometer was particularly suited to the newly emerging digital minicomputers. Finnigan Corporation, through a personal relation between Bob Finnigan and Ed Zschau (Stanford Business School Professor and early Silicon Valley entrepreneur) formed a relationship with Zschau's new company, System Industries, by which SI commercialized a DS based on the Digital Equipment Corporation's PDP-8 computer that had been in the Genetics Department at Stanford's Medical School.[63] Although the GC-MS was a low-priced item, adding the $50 000 for the DS produced a rather pricey item. Finnigan Corporation introduced the first GC-MS DS at the Fall ACS meeting in New York in 1969. This instrument was delivered to Evan Horning's group at the Baylor College of Medicine (Houston, TX).[37]

By the early 1970s, computerized GC/MS systems based on the TQ mass spectrometer were established analytical instruments. However, the $150 000 price for these instruments ($450 000 in the mid-1990s) prevented their wide use. Hewlett-Packard's (HP) efforts in the early days of TQ GC/MS have not been as well chronicled as those of Finnigan Corporation. HP's early efforts did not result in the same initial impact as Finnigan's, but the subsequent efforts of HP led to the affordability of the TQ-based GC-MS DS. Other companies played a key role with respect to the TQ mass spectrometer in the decade of GC/MS (1970s) and the decade of MS/MS (1980s). These companies are SCIEX (Toronto, Canada, now selling its instruments through PE Biosystems, Foster City, CA), Extranuclear Laboratories (Pittsburgh, PA, later known as Extrel, now operating as ABB Extrel and no longer providing analytical instrument systems), Riber/RiberMAG/Nermag (a French company still in France and owned by ThermoQuest), and VG Micromass. The only double-focusing instrument manufacturer at the time Finnigan and HP introduced their first instruments that were able to embrace successfully the TQ technology was VG Analytical (now operating as Micromass).

In addition to the continued manufacture of TQ m/z analyzers for use in GC/MS and LC/MS by ThermoQuest and Agilent Technologies, high-quality analytical instruments used for these purposes are also available from Waters/Micromass, Shimadzu, Perkin-Elmer Analytical Instruments (a division of EG&G), PE Biosystems/SCIEX, Hitachi, and others. The TQ is also used in RGAs and ICP mass spectrometers.

2.6.5 Quadrupole Ion-trap Mass Spectrometer

The QIT mass spectrometer was born out of the same Paul research that produced the transmission quadrupole; however, the QIT did not become a commercially viable instrument until the mid-1980s. The QIT uses a three-dimensional quadrupole field to store ions of multiple m/z values as opposed to filtering out ions of all m/z values except those of a selected value from an ion beam. The QIT consists of two end-caps electrically isolated on either side of a ring electrode placed between the two end-caps (Figure 25). This three-dimensional device has a hyperbolic cross-section.

The primary premise of any mass spectrometer has always been that the ions had to be separated in an area of low pressure. When ions were introduced into or formed inside the QIT maintained at traditional mass spectrometer operating pressures of 10^{-6}–10^{-10} Torr, it

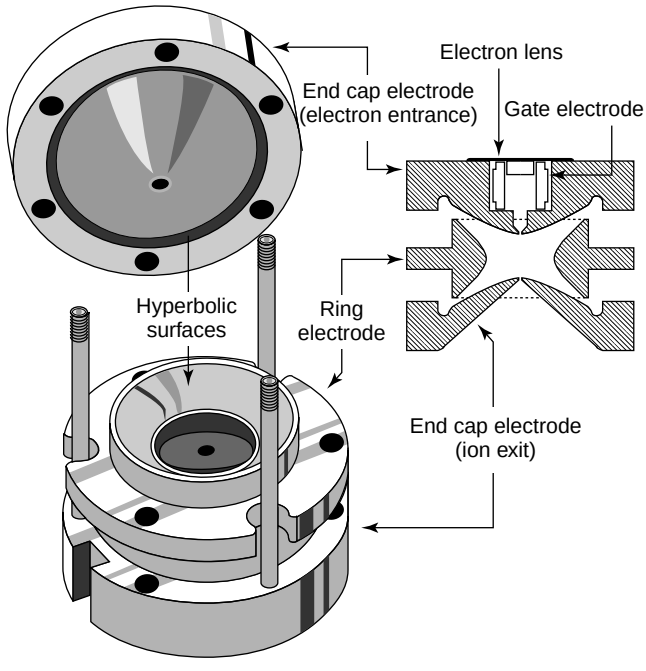

Figure 25 An illustration of the I^{2}QIT mass spectrometer.

was found that while storage was very efficient, separating ions that differed by no more than an integer m/z value was not possible. Serendipitously, George Stafford and co-workers at Finnigan Corporation (San Jose, CA) found that by using a light buffer such as helium to maintain the trap at about 10^{-3} Torr produced excellent unit resolution and improved sensitivity.[115] This late-1982 discovery coupled with Stafford's earlier announcement of the mass-selective instability scan allowed the commercialization of the I^2QIT instrument. Ray March, in the preface to *Particle Aspects of Ion Trap Mass Spectrometry, Vol. 1: Fundamentals of Ion Trap Mass Spectrometry* (CRC, Boca Raton, FL, 1995), states that Stafford's mass-selective instability scan is as significant as Paul's work.

Stafford's initial breakthroughs of buffer gas use and the mass-selective scan mode led Finnigan to announce the first commercial GC-MS based on a QIT – the ITD™ 700 (an I^2QIT device) – at the 1983 Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Atlantic City, NJ). The plan had been to begin shipping instruments by mid-1983; however, shipments did not begin until December 1984. One of the problems that delayed shipment of the ITD™ 700 was mis-mass assignment for the molecular ion peak of some analytes that were unrelated to space-charge effects or ion/molecule reactions (both related to pressure, explained below). This mis-mass assignment was compound (ion) specific. The final solution resulted in the 'stretched' trap. This stretched trap did not result in a homogeneous quadrupole field, which had previously been thought to be a requirement for proper ion separation,[116] but was found as the only way to overcome this particular problem.

Even with many of the initial problems of bringing the I^2QIT GC-MS to market overcome in an intensive development cycle as described by John Syka[115] (ThermoQuest, San Jose, CA), the instrument met with a mixed reaction regarding its usefulness as a GC/MS instrument by those who received these first units. Because packed GC columns were predominant among GC users (especially in the USA), the interface between the GC and the mass spectrometer was a device that split the GC eluate with one part in 10 going to the mass spectrometer and the remainder going to atmosphere as a discard (the open-split interface). This split arrangement often resulted in a higher than desired amount of analyte in the trap. The amount of analyte was dependent on its initial concentration in the GC eluate. Ionization took place in the trap for a fixed time period (settable by the operator). If the amount of analyte was too high and a larger number of ions than could be stored in separated orbitals were produced, a space-charge mass spectrum would result with clusters of peaks on either side of the nominal m/z value peaks. There was no way to prevent this space-charge effect because the partial pressure of the analyte could not be controlled nor the ionization varied as a function of analyte partial pressure.

Unlike the beam-type mass spectrometers (sector, TOF, and TQ), which detected ions a few microseconds after their formation, the QIT stored ions for long periods (during the ionization time) in an area with analyte molecules. Depending on the ions formed by the analyte, this could lead to ion/molecule reactions that produced mass spectra that exhibited much higher than normal $[M+1]^+$ peaks. This phenomenon (sometimes referred to as self-CI) is analyte dependent. The most dramatic example of self-CI can be observed in the analysis of 2-octane (MW 128). This compound can produce a mass spectrum with essentially a single peak at m/z 129, which represents the protonated molecule. These ion–molecule reactions between an analyte and fragments of its molecular ions can easily be avoided in modern QIT instruments.

The problems with space-charge effects and self-CI resulted in a lot of criticism directed toward the ITD™ 700. On the other hand, the good sensitivity of this instrument was unprecedented. Users found detection limits lower than they had ever believed possible. Further development by Finnigan led to a modification in data acquisition that produced a variable ionization period based on the partial pressure of analyte. This development went a long way toward correcting the space-charge problems. Another significant factor about this particular fix was that it involved only a software modification. No additional hardware was required. Existing users could have their space-charge problems addressed with a software upgrade. Finnigan had entered into an OEM agreement with Perkin-Elmer Corporation (Norwalk, CT) whereby Perkin-Elmer sold the ITD™ 700 to customers that were not disclosed to Finnigan. While Finnigan upgraded all of its customers to the next generation of the I^2QIT software (ITD™ 800), unfortunately, Perkin-Elmer did not, thereby resulting in many dissatisfied initial ion-trap users.

In 1989, Finnigan sold rights to use their ion-trap patents to Varian Associates' Analytical Instrument Division (now Varian Corporation, Palo Alto, CA). This assignment of rights to Varian was followed by the introduction of the third-generation I^2QIT GC/MS system (Varian's Saturn and Finnigan's ITS40). By this time, the capillary GC column had replaced the packed column, and a number of significant hardware and software advancements had been made by Finnigan. One of the most significant was the ability to apply an RF voltage, which was at half the frequency of that applied to the ring electrode, to the two end-caps. This feature known as axial modulation greatly increased the storage capability of the trap, further reducing space-charge problems.[117]

The I^2QIT went on to demonstrate other advantages such as low-pressure CI (reagent gas partial pressure of 10^{-5} Torr compared with 0.1–1 Torr in beam-type instruments), MS/MS in time and MS^n, and selected ion storage. Some 16 years after the introduction of the first commercial QIT mass spectrometer, a GC/MS instrument using external ionization is available from ThermoQuest; an internal ionization GC is still produced as several models by Varian worldwide, and Hitachi manufactures and sells an internal ionization GC-MS in Japan; ThermoQuest has two models of a second-generation API (ES and APCI) LC-MS; and Hitachi and BrukerDALTONICS (in a joint marketing agreement with Agilent Technologies) produce an ES QIT LC-MS. Because of the multiple iterations of ion isolation followed by CAD, the QIT is of particular interest to LC/MS. For this same reason and the extreme good sensitivity, QIT technology is also very popular with GC/MS users. The QIT GC-MS is the instrument of choice for the analysis of air pollutants. The best reference for understanding the physics of the QIT is *Quadrupole Storage MS* by R.E. March and R.J. Hughes (Wiley, New York, 1989).

2.6.6 Ion Cyclotron Resonance and Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

The ICR mass spectrometer is the other type of ion trap – magnetic as opposed to quadrupole. Whereas the quadrupole ion trap is referred to by the name of its original developer (the Paul trap), the ICR is called the Penning trap, a name given it by its developer, Hans G. Dehmelt.[118] Dehmelt's work with the development of the ICR, which began in 1955 at the University of Washington (Seattle, WA), led to his sharing half of the 1989 Nobel Prize in Physics with Wolfgang Paul. The ICR is a combination of the magnetron and the Penning ion source. Unlike the omegatron that uses a fixed-field magnet, the magnetron/Penning mass spectrometer uses a variable-field magnet.

The Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) was developed by Allen Marshall and Melvin B. Comisarow when they applied Fourier transform algorithms to the mass separation technique of ICR in 1974.[119] The application of the Fourier transforms to detect ions separated by the ICR technique improved it to the point that the technique became of significant value to analytical chemistry. This instrument is capable of extremes in high-resolving power (Figure 26) and has proved to be of great value in ES LC/MS where the number of charges on an ion can easily be determined by counting the number of isotope peaks within one m/z unit.

The FTICR-MS requires the use of a superconducting magnet, which, depending on the magnet's field strength, can add significantly to the cost. The higher the field

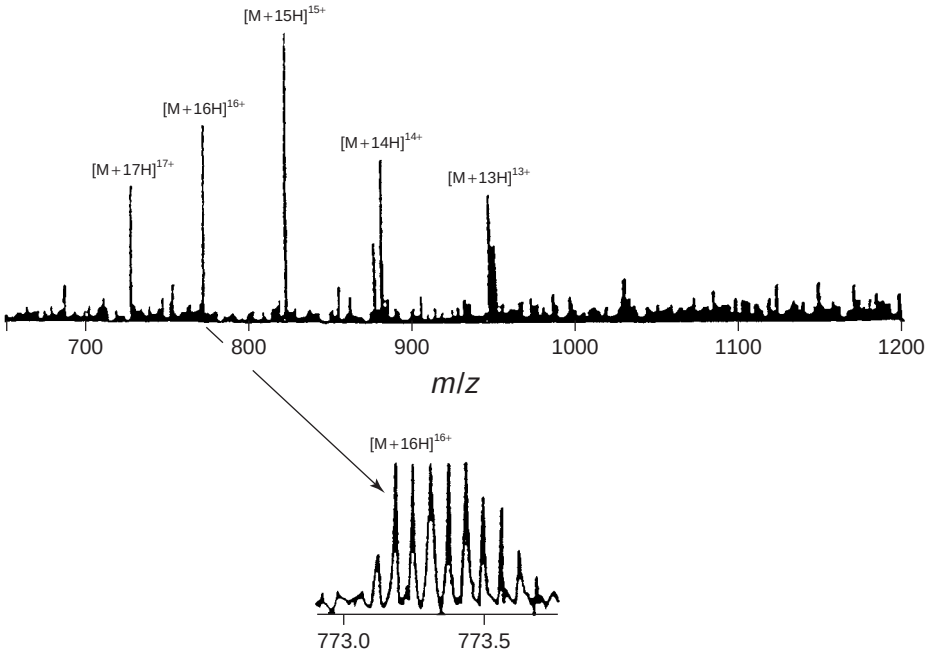

Figure 26 An example of the utilization of the high resolving power of FTICRMS in determining the number of charge states on an ES ion. The number of isotope peaks that can be observed in a single integer m/z unit is equal to the number of the charges on the ion. (Reproduced by permission from D.H. Williams, 'Applications of Chemical Techniques, NMR, Mass Spectrometry in Structural and Molecular Recognition Studies', American Chemical Society 1994 short course *Frontiers in Organic Chemistry*, Stanford University, Palo Alto, CA.)

strength, the higher is the m/z range of the instrument. These instruments also require a very skilled operator and need a lot of maintenance. In a 1998 article, Alan Marshall stated, ''... more than 235 [Fourier transform ion cyclotron resonance mass spectrometry (FTICRMS) instruments] installations by 1998''.[120] In all likelihood, most of these instruments were installed within the past 5 years and are in use with ES interfaces for the analysis of proteins.

Over the 24 years since Marshall and Comisarow first published their findings, three companies commercially manufactured these instruments: Nicolet Instrument Corporation (Madison, WI), Bruker-Franzen Analytik GmbH (Bremen, Germany), and Ionspec Corporation (Irvine, CA). The lowest advertised price for an FTICR-MS in 1999 was approximately $350 000. Nicolet originally sold its patents and FTICR-MS business to Extrel (previously known as Extra Nuclear); Extrel was then acquired by Waters Corporation; Waters then sold the FTICR-MS business to Finnigan Corporation. Through all of these ownerships, instruments continued to be developed and manufactured in Madison, WI. In late 1998, Finnigan (ThermoQuest) closed all operations in Madison and began the development of a new FTICR-MS in their Bremen, Germany, factory. This new instrument has yet to be announced, leaving only Bruker and Ionspec as manufacturers of this system. The FTICR-MS began to gain a great deal of popularity for protein analyses by ES and MALDI. With the introduction of the TOF-MS and TOF hybrids, the FTICR-MS has lost some of its luster. The TOF instruments have sufficient resolving power to provide the needed information, and even at the high prices of the hybrid instruments are less costly than the FTICR-MS.

2.6.7 Tandem Mass Spectrometry

In his book *Introduction to Mass Spectrometry*, J. Throck Watson[7] states, ''From the stability standpoint there are roughly three types of ions produced in the [ion] source: (a) those that are stable for 100 μs or longer and thus reach the source intact, (b) those that decompose immediately (less than 10^{-7} s) after formation, and (c) those that decompose 1–100 μs after formation.'' Watson goes on to define these three ions as *stable*, *unstable*, and *metastable*, respectively. These metastable ions had been obscured in the earliest mass spectrometers of Thomson and Aston. The peaks produced by metastable ions on the photographically recorded mass spectrum were different from those produced by stable ions in that they were more diffuse. Thomson said these peaks were a result of 'secondary' rays and Aston called these peaks 'bands' in the mass spectrum.[33] The metastable ion (m^*) represents the precursor ion (m_1) and its resulting fragment (product ion, m_2) with the relation $m^* = (m_2)^2/m_1$. For a long time, these broad peaks in mass spectra interfered with the quantitation of individual compounds in hydrocarbon mixtures; therefore, the instruments were tuned to avoid their appearance. As time progressed, it became obvious that not only did these bands result from decomposition of ions labeled as metastable; but, most of the time, ion decomposition in a mass spectrometer resulted from high-energy collisions of stable ions with background gas molecules in the mass spectrometer, which led to Fred McLafferty coining the term 'CAD'.[104] [Graham Cooks coined the term 'collision-induced decomposition' (CID) in his 1978 book *Collision Spectroscopy* (Plenum, New York). Today, the CID abbreviation is used for collision-induced dissociation. Both CAD and CID are correct and are used interchangeably.]

In 1966, T.W. Shannon and McLafferty showed that by forcing decomposition of a stable ion, isomeric ions could be characterized.[121] McLafferty cites several other examples in the 'Introduction to Tandem Mass Spectrometry' chapter in *Tandem Mass Spectrometry*, an early and important reference to this technology.[104]

The acronym coined for the technique of bringing about a decomposition of a stable ion by a forced collision with neutral gas molecules was MS/MS. William F. Haddon defined the acronym as 'MS/MS' in a symposium organized by Michael L. Gross presented at the University of Nebraska (Lincoln, NE, November 3–5, 1976).[122] The term was later defined by McLafferty and Frank M. Bockhoff as 'mass separation/mass spectral characterization' by analogy to GC/MS.[123] However, MS/MS is the definition of MS/MS that is used today.

After the developments of CI and FD, it became obvious that in order to get structural information from low-energy ions representing the intact analyte molecule that were produced by soft ionization techniques, some method of ion fragmentation would be necessary. As the possibility of CAD became more of a reality, new instrumental methods would be required. Much of the early significance came out of the work by Graham Cooks and his colleagues at Purdue. The early work can be found in *Collision Spectroscopy*,[124] a 1978 book that Cooks edited, and his 1973 book *Metastable Ions*.[103] Cooks also published an *Analytical Chemistry* A-pages article in the January 1978 issue entitled 'Direct Analysis of Mixtures by Mass Spectrometry', which described the potential power of what would become known as MS/MS.[69] (Figure 27)

Two significant hardware developments were reported in the mid- to late-1970s that propelled the field of MS/MS: (1) the adaptation of conventional double-focusing mass spectrometers to have ions pass from an ion source into a magnetic sector, then through a collision cell, and finally through an electric sector, thus allowing for

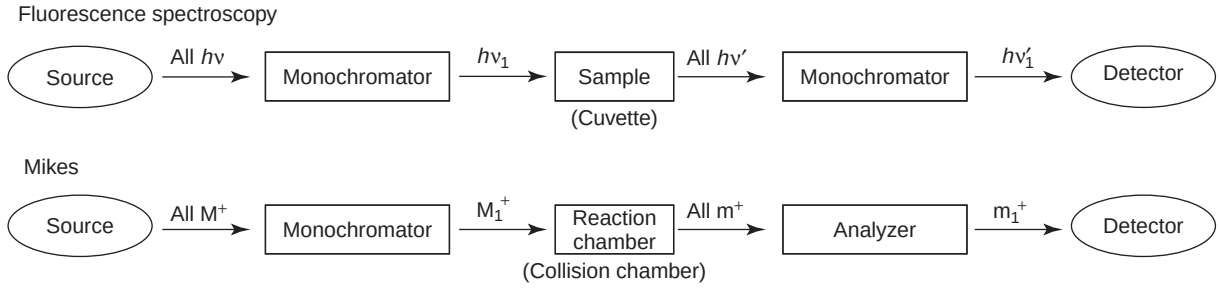

Figure 27 Prior to the development of the triple-quadrupole mass spectrometer for MS/MS, the technique of mass analyzed ion kinetic energy spectrometry (MIKES) in double-focusing mass spectrometers was the technique used to obtain product-ion mass spectra of isolated ions from an initial ionization and/or fragmentation. Graham Cooks made the analogy of tandem MS/tandem spectroscopy as shown in this figure.[69]

the determination of the mass of the precursor ion and the product ion by Micky Barber and R.M. Elliot[125] in 1964; and (2) the development of the tandem-quadrupole mass spectrometer (the triple quadrupole) by Rick Yost and Chris Enke at Michigan State University (East Lansing, MI)[126] using technology originally described by D.C. McGilvery and J.D. Morrison at La Trobe University (Bundoora, Australia).[127] See Note 3 at the end of the article.

The about-to-be explosion of MS/MS was exemplified at the 1980 Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Atlantic City, NJ) with the introduction of three triple-quadrupole instruments (Finnigan Corporation, Extranuclear Laboratories, Inc., and SCIEX, Inc. – the TAGA 6000), a reverse-geometry double-focusing instrument by VG Micromass, Ltd. – the ZAB-2F (Cheshire, UK), and a triple-sector (electric–magnetic–electric) instrument custom-built for Mike Gross at the University of Nebraska with a price tag of $400 000–600 000.[128] Twenty years later, SCIEX (marketed through PE Biosystems), Finnigan (as part of ThermoQuest), and VG Micromass (with a triple-quadrupole instrument as opposed to its original reverse-geometry double-focusing instrument and as a subsidiary of waters) are all still strong producers of MS/MS instrumentation.

The 1980s was definitely the decade of the triple-quadrupole m/z analyzer. As the quadrupole ion trap developed as a mass spectrometer, its role in MS/MS was obvious. The QIT mass spectrometer began the era of tandem-in-time as opposed to tandem-in-space (Figure 28). The concept of tandem-in-time allows for multiple iterations of ion isolation and CAD fragmentation (MS^n). Although the quadrupole has helped to advance the technique of MS/MS, it is not a replacement for tandem-in-space, which still has an important role in the area of common-neutral-loss analysis and common product-ion scanning (Figure 29). Neither of these types of analyses are amenable for use with the tandem-in-time technique.

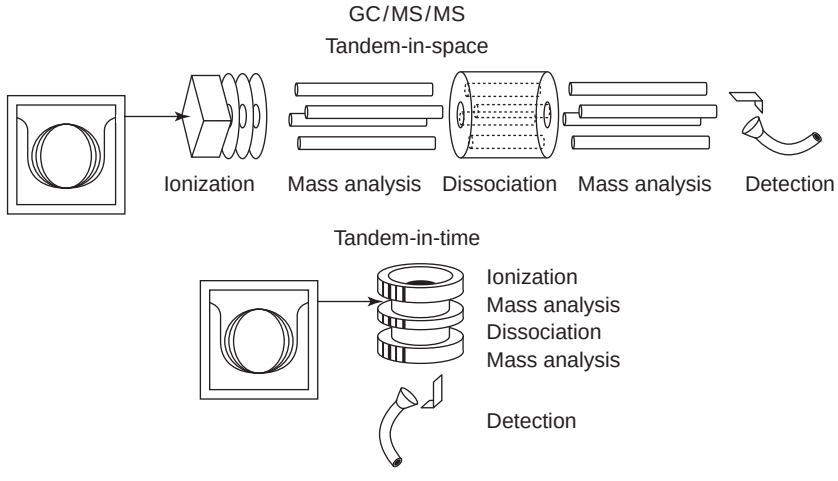

Figure 28 The concept of MS/MS as tandem MS in space vs tandem MS in time. (Adapted from J.V. Johnson, R.A. Yost, P.E. Kelly, D.C. Bradford, 'Tandem-in-space and Tandem-in-time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps', *Anal. Chem.*, **62**, 2162–2172 (1992).)

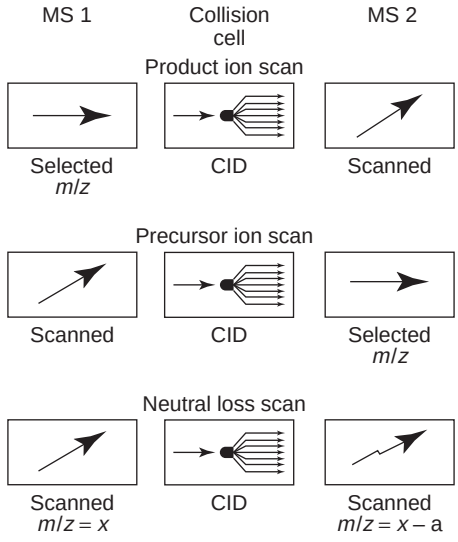

Figure 29 Illustration of three modes of MS/MS. (Reproduced by permission from E. de Hoffmann, 'Tandem Mass Spectrometry: A Primer', *J. Mass Spectrom.*, **31**, 129–137 (1996).)

2.7 Inorganic Mass Spectrometry

MS was originally used to determine the isotopes of the elements. It found some of its first applications in elemental analysis. The only preparative applications of MS have been in the area of purifying isotopes of elements. In recent years, MS has played a more publicized role in the analysis of organic compounds. However, like the solid analytical technique that it is, MS has advanced significantly in inorganic analyses. Surface analysis used to characterize papers used in laser printers and metal and composite skins of aircraft and materials put into outer space have advanced because of the continued development of SIMS. The ICP ion source for MS allows for the detection of trace elemental impurities in semiconductor manufacturing and the monitoring of inorganic pollutants in drinking and wastewaters.

2.7.1 Glow Discharge Mass Spectrometry

Glow discharge (GD) is the oldest ionization technique used in MS. John Aston gives a detailed account of his experiences with GD in the 1942 second edition of his book entitled *Mass Spectra and Isotopes*.[33] This technique was soon replaced by the spark-source technique in MS. In his 1965 book,[6] Robert Kiser states, ''The gaseous discharge ion source is largely of historical interest . . .''. John Roboz gives it about the same amount of space in his 1968 book.[10]

Although GD is best known as an ionization source for MS, it can be used in conjunction with other analytical techniques. GD has experienced a renaissance of interest in the last 20–30 years, largely due to the work of Willard W. Harrison at the University of Florida (Gainesville, FL). An *Analytical Chemistry* A-pages article by Harrison, 'GD Techniques in Analytical Chemistry', appeared in 1990 that described the use of GD with atomic absorption spectroscopy (AAS), atomic emission spectroscopy (AES), and atomic fluorescence spectroscopy (AFS) in addition to MS.[129]

One of the advantages that GD has over the other popular technique used for the analysis of inorganic samples (ICP) is that direct sampling of solids can be accomplished by GD. ICP requires that samples be in solution and nebulized or introduced as very fine aerosols. These aerosols can be produced from laser ablation or electrothermal vaporization. However, these intermediate sample techniques add to the cost and complexity of the method.

Another significant difference in these two inorganic MS techniques is the difference in the source pressure. The GD source operates at pressures similar to those used with CI in GC/MS, whereas the ICP source is an API source. A brief comparison of the operational parameters of the two techniques is described by Harrison.[130] This paper also includes additional references to other sources of comparison.

A search for articles in all ACS journals over the period 1996–1999 that contained 'GD' in their titles resulted in a list of 24. All but four of these were in *Analytical Chemistry*, and five were authored by Harrison. A similar search for articles containing 'ICP' resulted in a brief list of about a dozen articles related to MS. It does appear that GD as an ion source has had a higher degree of interest over the last few years; however, there are far more manufacturers of ICP mass spectrometers than there are of GD instruments. In his 1990 review of GD,[129] Harrison stated (without comment on the type of analyzer used) that commercial instruments were available from Extrel and VG Micromass. In all likelihood, the Extrel instrument was a TQ m/z analyzer and the VG Micromass instrument was based on a double-focusing sector instrument. In his 1999 paper, he states, ''Although most commercial GD MS instruments are coupled with double-focusing magnetic-sector units, the most recent developments in GD MS have been performed using quadrupole and TOF systems''.

Like ICP-MS, a common problem in glow discharge mass spectrometry (GD MS) is interferences from isobars of specific analytes due to polyatomic clusters of the analyte atoms and the analyte atoms with the inert gas (often Ar) used to create the plasma. A great deal of research has been conducted that has resulted in overcoming this problem in the ICP mass spectrometer, but has not been surmounted in the GD mass spectrometer. The ability to analyze solids is significant, but may not yield as much information or

the type of information that can be gained from SIMS. One of the reasons that ICP-MS dominates over GD MS is because much of the analytical MS involved with inorganic sample deals with analytes in solution.

2.7.2 Inductively Coupled Plasma Mass Spectrometry

In the 'Introduction' chapter to *ICP Spectrometry and its Applications*, Steve J. Hill, Editor (Sheffield Academic: Sheffield, UK; and CRC LLC, Boca Raton, FL, 1999), Stanley Greenfield and Michael Foulkes state that the first record of an ICP was made by J.J. Thomson in 1891.[131] Like GD, ICP is applicable to AAS, AES, and AFS in addition to MS.

Unlike GD, ICP did not become a mass technique until the mid-1970s, which was partly due to the requirement that the ICP be produced at atmospheric pressure. ICP had to wait until problems with API MS could be resolved. In 1974, Alan A. Gary, University of Surrey (Guild, Surrey, UK) published the use of a plasma jet as the ion source for a TQ mass spectrometer.[132] Gary continued to publish on this technique; however, in 1980 he took part in a feasibility study that changed his ion source to an ICP. The results of this study showed that elemental and isotopic compositions of solutions could be obtained with relatively little sample preparation.[133] This study was done with Robert S. Houk, who is at the US Department of Energy Ames Laboratory and the Department of Chemistry, Iowa State University (Ames, IA). Houk continues to be one of the leading researchers in ICP-MS.

Shortly after the Houk/Gary paper appeared, commercial instruments began to be introduced. The last half of the 1990s saw a number of new instruments and several significant technology advances. The 1997 and 1999 first-place *Analytical Instrument Industry (AII) Report* (Dr Gordon Wilkinson, Editor) Editors' and Writers' Award for the top three new instruments introduced went to instruments with ICP-MS technology advancements. The 1997 award went to Micromass (Manchester, UK) for the Platform-ICP (a TQ ICP mass spectrometer that used 'ICP-HEX-MS' to eliminate interferences from argon ions and reduce the energy dispersion of ions).[134] The 1999 award went to Perkin-Elmer (Wilton, CT) and MDS SCIEX (Concord, Ontario, Canada) for the ELAN 6100 DRC (dynamic reaction cell) ICP/MS system.[135] The PE/SCIEX DRC addressed the same problems as Micromass' ICP-HEX-MS, but used a different technology.[136]

Another significant advancement for ICP-MS was the introduction of a benchtop instrument by Hewlett-Packard (now Agilent Technologies) in 1997 through a joint venture with a Japanese developer. This product has been very successful and has brought ICP-MS within reach of many laboratories. Since 1994, every Pittsburgh Conference has seen the introduction of one or two new ICP-MS instruments. The ICP-MS technology is interfaced to double-focusing magnetic-sector TQ and TOF mass spectrometers. LECO Corporation (St. Joseph, MI) and GBC Scientific Equipment Pty. Ltd. (Dandenong, Victoria, Australia) introduced TOF-based ICP-MS instruments at the 1998 Pittsburgh Conference (New Orleans, LA). ICP-MS appears to have a good future in inorganic MS analyses.

2.7.3 Secondary Ion Mass Spectrometry

SIMS was an early development in MS. The study of surfaces by 'sputtering', although not thoroughly understood, was reviewed by H.S.W. Massey and E.S.H. Burhop in 1952.[137] They reported that in 1931 K.S. Woodcock at the Ryerson Physical Laboratory, University of Chicago (Chicago, IL), observed secondary-ion formation when a surface was bombarded with an ion beam.[138] According to Charles M. Judson,[139] Richard Herzog filed a German patent application for a SIMS instrument in 1942. In 1951, R.H. Plumlee reported on ion-sputtering techniques as being employed by RCA Laboratories (Princeton, NJ) for the analysis of solids in a symposium entitled 'Mass Spectroscopy in Physics Research' held September 6–9, 1951, at the NBS (now NIST).[140] R.E. Honig, also an employee of RCA Laboratories, is credited with much of the early work that proved productive in this field.[141,142]

SIMS involves the formation of ions on or near the surface by bombarding the sample with a beam of mass-selected ions. The resulting secondary ions are desorbed into the gas phase. This technique was used to study the impurities in material such as the aluminum used for airplane wings and germanium wafers used in early solid-state electrical devices. SIMS and its variants are widely used in the analysis and study of surfaces of all types of material – from papers used in laser printers, to microprocessor devices, to the wings of aircraft manufactured from many new polymeric composites.

As an interesting point, the abbreviation for SIMS is a pronounced acronym with an obvious 's' at the end, whereas the abbreviation for the technique used in chromatography/MS, selected ion monitoring (SIM), should not be used as an acronym because of the potential for confusion of these two entirely different techniques. Unfortunately, because of its ease of use, many people say '. . . analyzed by SIM' or '. . . a SIM analysis' rather than using the proper form '. . . analyzed by S-I-M' or '. . . an S-I-M analysis'. It is also proper to say each word when talking about either technique, but never use SIM as an acronym.

3 THE FUTURE OF MASS SPECTROMETRY

From its beginning with Thomson's 1897 report of the discovery of the electron (using what is considered

by many as the first mass spectrometer – the parabola mass spectrograph) to the present day status, the mass spectrometer has been and continues to be probably the most widely used analytical instrument. In all likelihood, MS technology has had the greatest impact on society of any analytical technique. Its nearest rival, nuclear magnetic resonance (NMR) spectroscopy, while enjoying a wide degree of use, has not had the impact of MS. The commercialization of and advances in (1) MALDI and ES instruments for the analysis of nonvolatile and thermally labile organics, (2) SIMS used in the characterization of organic and inorganic surfaces, (3) the specialty uses of MS such as RGAs that are used in space helmets to monitor patient health and medication performance, and in semiconductor manufacturing, (4) IR MS, and (5) ICP-MS have made MS a common technique for all fields of science.

The first 50 years were dominated by the physicists and their quest for the understanding of the ion optics of the magnetic-based single- and double-focusing instruments. As a result of MS's role in the development of the atomic bomb in World War II, the underlying principles of the sector-based mass spectrometer were well understood as we headed into the postwar years. The 1950s saw the embrace of the mass spectrometer by the organic chemist, the development of TOF technology, and the embryo of the quadrupole field by the physicist. The 1960s resulted in increased usage of MS by the organic chemist with the development of new ionization techniques such as CI and FD. The 1960s were also a period for the advances of MS instruments as RGAs and as a technique for the characterization of surfaces.

The 1970s were the decade of GC/MS and an entirely new paradigm for the analysis of complex mixtures of organic compounds. GC/MS led to new ways of looking at drinking and waste waters and drug metabolites, and for solving forensic puzzles, to name just a few of the new abilities. These analysis enhancements were possible because of the commercialization of the TQ, which radically changed the field of organic MS. Unfortunately, the 1970s also saw the beginning of the temporary decline of interest in the TOF mass spectrometer. The 1980s were the decade of MS/MS, and the 1990s were characterized by DI, which led to more developments in the area of LC/MS and the reemergence of the TOF mass spectrometer. During all three of these decades, uses of MS outside organic chemistry expanded, such as the popularity of ICP-MS using both the TQ and the double-focusing mass spectrometer.

As we move into the 21st century, MS takes on even more of a significance in science than it has had in the last decade. The MS literature is growing at a rapid rate with six English language journals devoted solely to the field. The ACS *Analytical Chemistry* journal is at least 25% MS, and Elsevier's *Journal of Chromatography A* and *Journal of Chromatography B* have as many articles related to the use of MS as they do to pure chromatographic techniques. By 1999, the extended abstracts, which comprised the proceedings of the ASMS annual meeting on MS and allied topics, had grown so large that it was no longer practical to publish a print edition. Therefore, after 38 years (the first time the meeting proceedings were published was 1961), ASMS discontinued publishing the proceedings of its meeting as a print issue and began sole distribution on CD-ROM.

Although yet to be established, it is possible that the next decade of MS will be the era of imaging. At the 1999 ASMS meeting (Dallas, TX), there were a number of reports on the use of MS to image the distributions of molecular-specific matter on the surfaces of materials such as contact lenses and in tissue slices from rat brain as well as polymeric surfaces that are used in composite aircraft wings.[143]

The advances that MS has made in (1) GC/MS and LC/MS, (2) surface analyses, inorganic analyses in environmental sciences and semiconductor manufacturing, to name only two, (3) residual gas analyses used in process monitoring and medicine, (4) detection of biological and chemical warfare agents, and (5) analyses of biopolymers with MALDI will continue at rates equal to those of the last three decades.

One thing is clear: In the first quarter of the 20th century, the inventions of Joe Thomson and John Aston in Manchester, UK, had as profound an effect on mankind as they did on the field of analytical chemistry.

4 EPILOG

It is important to remember that 'What you know about a subject is not nearly as significant as knowing where to find out what you need to know about that subject'. To that end, Appendix 1 of this article is a listing of important Web sites in 1999 that will help you locate much of the information you may need about MS. Appendix 2 is a bibliography of books on MS that have been published through 1999. Appendix 3 comprises a series of tables with helpful information that will be valuable in the interpretation of mass spectra of organic substances.

The MS section of this Encyclopedia contains articles that deal directly with techniques. In addition, there are articles that pertain to specific areas of analyses and analytes that also contain information on the use of MS. All of these articles, along with the referenced books in Appendix 2, will be invaluable in a concentrated study of any aspect of this exciting field of science.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
ACS	American Chemical Society
AES	Atomic Emission Spectroscopy
AFS	Atomic Fluorescence Spectroscopy
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
ASMS	American Society for Mass Spectrometry
ASTM	American Society for Testing and Materials
CAD	Collisionally Activated Dissociation
CF-FAB	Continuous-flow Fast Atom Bombardment
CI	Chemical Ionization
CID	Collision-induced Dissociation
DC	Direct Current
DCI	Desorption Chemical Ionization
DEI	Desorption Electron Ionization
DI	Desorption/Ionization
DLI	Direct Liquid Introduction
DS	Data System
EI	Electron Ionization
EPA	United States Environmental Protection Agency
ES	Electrospray
FAB	Fast Atom Bombardment
FD	Field Desorption
FI	Field Ionization
FTICR	Fourier Transform Ion Cyclotron Resonance
FTICRMS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
FTICR-MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometer
FTMS	Fourier Transform Mass Spectrometry
GC	Gas Chromatograph
GC/MS	Gas Chromatography/Mass Spectrometry
GC-MS	Gas Chromatograph–Mass Spectrometer
GD	Glow Discharge
GD MS	Glow Discharge Mass Spectrometry
HPLC	High-performance Liquid Chromatograph
ICP	Inductively Coupled Plasma
ICP-MS	Inductively Coupled Plasma Mass Spectrometer
ICR	Ion Cyclotron Resonance
ID	Inner Diameter
INCOS	Instrument Company Operating System
I^2QIT	Internal Ionization Quadrupole Ion-trap
IR	Infrared
IR MS	Isotope-ratio Mass Spectrometry
LC	Liquid Chromatograph
LC/MS	Liquid Chromatography/Mass Spectrometry
LC-MS	Liquid Chromatograph–Mass Spectrometer
LSIMS	Liquid Secondary Ion Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/Ionization
microES	Microelectrospray
MIKES	Mass Analyzed Ion Kinetic Energy Spectrometry
MS	Mass Spectrometry
MS/MS	Mass Spectrometry/Mass Spectrometry
nanoES	Nanoelectrospray
NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
OEM	Original Equipment Manufacturer
PD	Plasma Desorption
PFTBA	Perfluorotributylamine
PONA	Paraffin, Olefin, Naphtha, and Aromatic
QIT	Quadrupole Ion-trap
RAM	Random Access Memory
RF	Radiofrequency
RGA	Residual Gas Analyzer
SIM	Selected Ion Monitoring
SIMS	Secondary Ion Mass Spectrometry
TOF	Time-of-flight
TOFMS	Time-of-flight Mass Spectrometry
TOF-MS	Time-of-flight Mass Spectrometer
TQ	Transmission-quadrupole
UV	Ultraviolet

NOTE

1. MS measures the abundance of ions according to their mass-to-charge ratios. Therefore, the abscissa of a mass spectrum is generally labeled with the units of mass-to-charge, not mass. The exception to this is the labeling of the abscissa of a mass spectrum constructed from the charge deconvolution of data with a series of multiple-charge ions. The units of mass are either unified atomic mass units with the symbol 'u' or daltons with the symbol 'Da'. The use of the symbol 'amu' for mass was discontinued in 1962.

   The amu symbol for a unit of mass was used when the standard for mass was based on oxygen-16. Physicists reported mass based on the most abundant naturally occurring stable isotope of oxygen [^{16}O, established by Francis William Aston (1877–1945) in 1929 after his discovery that oxygen was composed of three different isotopes, two of which had a higher mass (^{17}O and ^{18}O) than the most abundant isotope], which was assigned an exact mass of 16 (1 amu = 1/16 of the mass of ^{16}O). This definition was the basis of the *physical atomic mass scale*.

Chemists used amu to define a unit of mass as 1/16 the atomic weight (the average atomic mass) of oxygen [officially established in 1905 on the suggestion of the Belgian chemist Jean Servais Stas (1813–1891)]. This definition was the basis for the *chemical atomic mass scale*. The two scales differed by a factor of 1.000275 (physical > chemical).

An atomic weight is the weighted average of the masses of the naturally occurring stable isotopes of an element, and oxygen has three such isotopes. The atomic weight of oxygen was an absolute value of 16 by definition. The chemical atomic mass scale made the determination of the atomic weight of newly discovered elements easy by forming their oxides.

To eliminate the ambiguity between the physical and chemical standards, the standard of a single mass unit as 1/12 the most abundant naturally occurring stable isotope of carbon (^{12}C) was adopted in 1960 by the International Union of Physicists at Ottawa and in 1961 by the International Union of Chemists at Montreal. This standard is based on the independent recommendations of D.A. Olander and A.O. Nier in 1957. The symbol for the *unified atomic mass unit* was established as u, NOT μ, which appears in the *ASMS Guidelines* and in the 4th edition of *Interpretation of Mass Spectra* (McLafferty, 1993).

Prior to the oxygen standards, the basis for atomic mass had been set ca. 1805 by John Dalton (1766–1844) as 1 for the lightest element, hydrogen. In 1815, the Swedish scientist Jöns Jacob Berzelius (1779–1848), set the atomic weight (relative atomic mass) of oxygen to 100 in his table of atomic masses; however, the Berzelius standard of mass was not accepted. Stas's recommendation of setting the oxygen-16 standard allowed hydrogen to retain a mass close to 1, thereby keeping Dalton's scale somewhat intact.

2. When formed, ASMS had three vice-presidents: Programs, Arrangements, and Data & Standards. The vice-president of Data & Standards was not elected by the ASMS membership, but was the elected ASTM E-14 Committee Chair. In 1974, the ASMS membership changed the role of the E-14 Committee Chair to that of a non-elected Director-at-Large. The role of the E-14 Committee Chair in ASMS changed again in 1974 when the ASMS membership voted to no longer have a member of the E-14 as an automatic member of the ASMS Board of Directors. This position was replaced by a third Member-at-Large for Measurements and Standards to be elected by the ASMS membership. From that time forward, reports of E-14 meetings no longer appeared in the proceedings of the annual ASMS Conference on MS and Allied Topics. This change was made by ASMS because the interest of the ASMS membership was far broader than that of the E-14 Committee, and the interest of the E-14 Committee represented a small fraction of the diverse MS interest of ASMS. The year after the E-14 Committee Chair's removal from the ASMS Board of Directors, the ASTM abolished the official standing of the committee. Note: the above information is taken from reports of the annual meetings of ASMS from 1972 to 1998 as published in the Society's annual proceedings.
3. As the MS/MS technique developed, the term 'parent' was used to describe the precursor ion. The use of this word resulted in some confusion because organic mass spectrometrists had traditionally called the molecular ion a 'parent ion'. The product ion of a CAD was called the 'daughter ion'. The parent–daughter terminology led one group to coin the term DADI (pronounced daddy) as an acronym for direct analysis of daughter ions.[144] Today, this parent–daughter terminology is discouraged owing to its perceived offensive nature by some people. Therefore, the term DADI would no longer be appropriate.

APPENDIX 1: IMPORTANT WORLDWIDE WEB URLs

1. **MS Resources on the Internet**
   http://www.base-peak.wiley.com
2. **ASMS Links to eJASMS**
   http://www.asms.org
3. **ACS**
   http://www.acs.org
4. ***J. Am. Soc. Mass Spectrom. On-Line***
   http://www-east.elsevier.com/webjam/menu.html
5. **ChemWeb.com**
   http://www.chemweb.com/home
6. ***MEDLINE and Evaluated MEDLINE***
   http://www3.infotrieve.com/medline/infotrieve/ and http://www.biomednet.com/db/medline
7. **The Association of Biomolecular Resource Facilities**
   http://www.abrf.org
8. **Ross C. Willoughby – *A Global View of LC/MS***
   Chem-Space Associates http://www.lcms.com
9. **INTERLOGIX™ Chromatography.Net™**
   HPLC, GC, IC, ion exchange, GPC, CE, CEC, etc. column database
   http://www.chromatography.net
10. **Prospector**
    http://www.prospector.ucsf.edu/mshome3.0.4.htm

11. **Protein information Retrieval On-line World Wide Web Lab**
    http://www.prowl.rockefeller.edu/resource.htm
12. **Service and Supplies**
    Scientific Instrument Services
    http://www.sisweb.com
13. **Out-of-Print MS Books**
    http://www.addall.com/used/index.html or http://www.bookfinder.com
14. **In-Print Books**
    http://www.sisweb.com
    http://www.barnesandnoble.com or http://www.amazon.com
15. **MS Significant News Groups**
    sci.techniques.mass-spec
    sci.techniques.spectroscopy
    sci.chem.analytical

APPENDIX 2: BIBLIOGRAPHY

(Reprinted with permission from O.D. Sparkman, *Mass Spectrometry Desk Reference*, Global View, Pittsburgh, PA, 2000)

Preparation of any such compilation is thwarted with the problem that by the time it reaches the intended audience, it will be out of date. Although this is the case with this current effort, the generation of this collection of book titles was, to some degree, inspired by two such collections that are found in books authored by Roboz in 1968 (reference 77) and Kiser in 1965 (reference 81). Both of these collections have proved to be invaluable in personal studies of MS. More recent collections of book titles in MS appeared in McLafferty and Tureček (Interpretation 6) and de Hoffman et al. (Introductory 8). Both of these collections, like the Roboz and Kiser collections, are dated, even though both were prepared in the 1990s.

The old adage, ''You can't judge a book by its cover!'' is truer today than ever before. Although there has been a number of outstanding MS books that address the new technologies developed at the end of the twentieth century, the literature of MS has been cursed with a plethora of recently published titles that suffer from bad technical content, little or no competent copyediting, and/or amateurish and sloppy production. These problems are due to consolidations in the publishing industry resulting in fewer publishers, a perceived need for new titles to replace older ones and to provide information on a subject that has exploded in the last decade with new ionization techniques, and the lack of desire and ability of authors to research their subject and avoid the self-gratification of generating an avalanche of neologisms and technical errors. These types of problems were practically nonexistent in books on MS before 1990.

There is a well-written and organized general text on MS (Introductory 4) that suffers from a less than adequate production effort. In this particular book, there is a lack of consistency in the fonts used for the presentation of reaction schemes, figures are improperly imported from electronic submissions, and promised graphics on the inside cover are replaced with an easy-to-lose insert. This book is the singularly best modern text on MS partially because its thousands of journal references all have titles. It is a shame that the publisher did not make the effort to do a better job of the presentation. This is one of the reasons why you see more self-publishing (Introductory 2 and Technique 10).

Another book, produced by a major publisher, consistently uses the word ''spectra'' (the plural form) for the word ''spectrum'' (the singular form) (Technique 6). As a corollary to ''you can't judge a book by its cover'', ''you can't necessarily judge a book by its title''. A book was published with the title *Understanding Mass Spectra – A Basic Approach* (Technique 3) when the title *Understanding Mass Spectra of Illicit Drugs* would have been more reflective of the contents and the complex spectra presented. There are also problems with poorly translated foreign-language books such as one originally written in French and translated into English by a person whose native language was Chinese (Introductory 8); reviewed in *J. Am. Soc. Mass Spectrom.*, **8**, 1193–1194 (1997). Another problem with translated books is the delay in information. By the time one of these books reaches its intended audience, the information is dated. This is especially true with the rapid change that is taking place in LC/MS and MS and biotechnology.

Most of the books on MS come from the chemists who use the technique. Chemistry, unlike biochemistry and the biological sciences, has been slow to recognize the importance of the titles of journal articles in cited references. Most of the MS books published in the last 10 years (and nearly all those previously published) do not include titles with cited journal articles. That is why a special effort has been made to mark books that do use titles in cited journal articles in this bibliography. These books are noted with an asterisk (*).

This bibliography is presented in 11 segments. Book-related segments include **Introductory**, **Reference**, **Technique-oriented**, **Interpretation**, **Historical Significance**, **Collections of Mass Spectra in Hardcopy**; and **Integrated Spectral Interpretation**. Non-book-related segments include **MS**; **GC/MS**; and **LC/MS Journals**; **Personal Computer MS Abstract Sources**; **Software**; and **Monographs**. These segments are included because of their importance in finding information on MS.

There is no inclusion of specific articles from journals. Through the 1960s, such listings of journal articles were published in various forms – often by MS instrumentation manufacturers. However, since the development of comprehensive abstracts systems and their electronic availability, these printed listings no longer have much value. The **Personal Computer MS Abstract Sources** section is a guide to information on current journal articles.

There is no inclusion of Internet search engines such as Medline and Chemical Abstracts Service. Some of these search engines are found on the **Important Worldwide Web URLs** in **Appendix 1**.

There has been some duplication of book titles. For example, the Watson book (Introductory 4 and Interpretation 4) covers two areas of MS. This book is excellent for the interpretation of EI spectra, as well as an introductory book. All books are listed in chronological order by section.

Introductory Books

The **Introductory Books** section contains titles of introductory books on MS, GC, and LC. If you are going to own only one MS book, it should be *Introduction to Mass Spectrometry*, 3rd edition (Introductory 4). This is a comprehensive book that touches on all the current technology. This book provides an excellent understanding of EI fragmentation mechanisms (which is essential to the understanding of all fragmentation in MS); and, of great importance, all of the journal articles referenced have titles. Another important book in this section is *Pushing Electrons: a Guide for Students of Organic Chemistry* (Introductory 3). This is a student workbook and instills a good understanding of moving electrons within organic ions and molecules, which is also essential to the understanding of mass spectral fragmentation.

MS Fundamentals (Introductory 9) is not a book. It is a multimedia training tool consisting of a computer-based training program, video, and book. This entry is also listed in the **Software** section. To those who work with the TQ mass spectrometer, this multimedia presentation is indispensable. In order to get the most from any analytical instrument, you must have a thorough understanding of the technology. This multimedia presentation will result in that understanding, even in those who have little technical background.

The **Introductory Books** section has three *Analytical Chemistry by Open Learning* titles (Introductory 1, 10, and 16). These books are programmed-learning texts and are very good for self-study. Two of the entries in this section are video courses produced by the ACS (Introductory 15 and 19). Neither is recommended because the MS course is dated, and the GC course has too little emphasis on capillary columns.

1. J. Barker, *Mass Spectrometry: Analytical Chemistry by Open Learning*, 2nd edition, ed. D.J. Ando, Wiley, Chichester, 1999. (R. Davis, M.J. Frearson, 1st edition, 1987).
2. R.L. Cunico, K.M. Gooding, T. Wehr, *Basic HPLC and CE of Biomolecules*, Bay Bioanalytical Laboratory, Richmond, CA, 1998.
3. D.P. Weeks, *Pushing Electrons: a Guide for Students of Organic Chemistry*, 3rd edition, Saunders College, Fort Worth, TX, 1998.
4. *J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, Philadelphia, 1997.
5. H.M. McNair, J.M. Miller, *Basic Gas Chromatography*, Wiley, New York, 1997.
6. G. Siuzdak, *Mass Spectrometry for Biotechnology*, Academic Press, San Diego, CA, 1996.
7. R.A.W. Johnstone, M.E. Rose, *Mass Spectrometry for Chemists and Biochemists*, 2nd edition, Cambridge University Press, Cambridge, 1996.
8. E. de Hoffmann, J. Charette, V. Stroobant, *Mass Spectrometry: Principles and Applications*, Masson, Paris, 1996; Wiley, New York, 1996 (original French language edition, *Spectrométrie de Masse*, Masson, Paris, 1994).
9. *MS Fundamentals: a Multimedia Training Course*, SAVANT, Fullerton, CA, 1995.
10. I.A. Fowlis, *Gas Chromatography: Analytical Chemistry by Open Learning*, 2nd edition, Wiley, Chichester, 1995.
11. D.H. Russell (ed.), *Experimental Mass Spectrometry*, Plenum Press, New York, 1994.
12. J.V. Hinshaw, L.S. Ettre, *Introduction to Open-tubular Column Gas Chromatography*, Advanstar, Cleveland, OH, 1993.
13. L.S. Ettre, J.V. Hinshaw, *Basic Relationships of Gas Chromatography*, Advanstar, Cleveland, OH, 1993.
14. J.R. Chapman, *Practical Organic Mass Spectrometry: a Guide for Chemical and Biochemical Analysis*, Wiley, New York, 1993.
15. J.T. Watson, *Introduction to Mass Spectrometry*, ACS Video Courses, American Chemical Society, Washington, DC, 2 tapes, 130, 1993.
16. S. Lindsay, *High Performance Liquid Chromatography: Analytical Chemistry by Open Learning*, 2nd edition, Wiley, Chichester, 1992 (1st edition, 1987).
17. *D.M. Desiderio (ed.), *Mass Spectrometry: Clinical and Biomedical Applications*, Plenum Press, New York, Vols. 1 and 2, 1992.
18. F.W. Karasek, R.E. Clement, *Basic Gas Chromatography–Mass Spectrometry*, Elsevier, New York, 1988.
19. H.M. McNair, *Basic Gas Chromatography*, ACS Video Courses, American Chemical Society, Washington, DC, 1988.
20. J.H. Beynon, A.G. Brenton, *Introduction to Mass Spectrometry*, University of Wales Press, Swansea, 1982.
21. E. Johnson, R. Stevenson, *Basic Liquid Chromatography*, Varian, Palo Alto, CA, 1978.

22. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley, New York, 1978.
23. H.M. McNair, E.J. Bonelli, *Basic Gas Chromatography*, 5th edition, Varian, Palo Alto, CA, 1969.

Reference Books

The citations in this section are meant to be general references on MS. In addition, there is a reference to LC method development. One of the most important entries in this section is the 'MS' article in the biennial Fundamental Reviews issue of *Analytical Chemistry*. This review began in 1949 with the article authored by John A. Hipple and Martin Shepherd (NBS, now the NIST) with 165 citations. The latest of these reviews (reference 5) had 1551 citations divided into nine categories: Overview (5), Scope (173), Innovative Techniques and Instrumentation (364), Isotope-ratio MS (89), High-power Lasers in MS (51), Dissociation by Low-intensity IR Radiation (18), Polymers (61), Peptides and Proteins (624), and Oligonucleotides and Nucleic Acids (166). The 'MS' articles in the Fundamental Reviews issues of *Analytical Chemistry* have had Alma L. Burlingame as their primary author since 1972 with 14 consecutive articles.

An important aspect of using a mass spectrometer is 'becoming one' with the instrument. In order to accomplish this, you must have an in-depth working understanding of the instrument. The multimedia training program distributed by SAVANT, *MS Fundamentals* (Introductory 9 and Software 23), is an excellent aid to gaining this understanding of the TQ mass spectrometer. In addition, the Dawson book (Reference 63) is a seminal reference for the TQ and the QIT. The March and Hughes book (Reference 35) is the seminal reference for the QIT, and the Cotter book (Reference 12) is a likewise reference for the TOF mass spectrometer. Care must be taken with respect to the Cotter title because he has edited a book (Reference 20) that does not provide the detailed understanding of TOFMS. The same is true for the March and Hughes book because of a similar title edited by March and Todd (Reference 17). There is no good reference for the current technology of magnetic-sector mass spectrometers. The best material on understanding the fundamentals of a magnetic-sector and/or double-focusing instrument is the Roboz book (Reference 77). The best understanding of the workings of FTMS is found in an article, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer', authored by Allen Marshall et al. in *Mass Spectrom. Rev.*, **17**(1), 1–36 (1998).

Working with mass spectrometers often requires the derivatization of analytes to obtain the best results. This section has three good books that aid in this task (References 25, 53, and 78). In addition, there are two other books that do not primarily pertain to MS techniques: *Basic Vacuum Practice* from Varian (Reference 36) and *The Mass Spec Handbook of Service* by Manura (Reference 24).

It should be noted that several books of the 1960s and early 1970s have been reprinted in the past few years (References 63, 82, and 84) by the American Institute of Physics and the ASMS. This is not a new practice. Several titles were reprinted by a publisher other than the original publisher several years after the first printing. ASMS plans to reprint several other titles of both historical and technological significance over the next few years.

1. *S.J. Hill (ed.), *ICP Spectrometry and its Applications*, Sheffield Academic, Sheffield, 1999.
2. K.W. Shepard (ed.), *Heavy Ion Accelerator Technology*, Eighth International Conference (AIP Conference Proceedings), American Institute of Physics, New York, Vol. 473, 1999.
3. *O.D. Sparkman, (p. 2604), G.J. Wells, C.K. Huston, (p. 2662), F. Adams et al., (p. 2650), in *Encyclopedia of Environmental Analysis and Remediation*, ed. R.A. Meyers, Wiley, New York, Vol. 4, 1998.
4. E.J. Karjalainen, A.E. Hesso, J.E. Jalonen, U.P. Karjalainen (eds.), *Advances in Mass Spectrometry*, Proceedings of the 14th International Conference, 1997, Tampere, Finland, Elsevier, Amsterdam, 1998 (also available as a CD ROM).
5. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', in the Fundamental Reviews issue of *Anal. Chem.*, **70**(16), (1998).
6. C. Tuniz, J.R. Tuniz, D. Fink Bird (eds.), *Accelerator Mass Spectrometry: Ultrasensitive Analysis for Global Science*, CRC Press, Boca Raton, FL, 1998.
7. H.E. Grove, *From Hiroshima to the Iceman: the Development and Applications of Accelerator Mass Spectrometry*, American Institute of Physics, New York, 1998.
8. A. Montaser (ed.), *Inductively Coupled Plasma Mass Spectrometry*, VCH, Berlin, 1998.
9. L.T. Platzner, K. Habfast, A. Walder, A. Goetz, *Modern Isotope Ratio Mass Spectrometry*, Wiley, New York, 1997.
10. G. Holland, S.D. Tanner (eds.), *Plasma Source Mass Spectrometry: Development and Applications*, International Conference on Plasma Source Mass Spectrometry, American Chemical Society, Washington, DC, 1997.
11. C.L. Wilkins, 'Mass Spectrometry', in *Handbook of Instrumental Techniques for Analytical Chemistry*, ed. F.A. Settle, Prentice Hall, Upper Saddle River, NJ, Section V, 1997.
12. R.J. Cotter, *Time-of-flight Mass Spectrometry: Instrumentation and Applications in Biological Research*, American Chemical Society, Washington, DC, 1997.
13. L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2nd edition, Wiley, New York, 1997.

14. A.E. Ashcroft, *Ionization Methods in Organic Mass Spectrometry*, RSC Analytical Spectroscopy Monographs, Royal Society of Chemistry, Cambridge, 1997.
15. T. Baer, C.-Y. Ng, I. Powis (eds.), *Large Ions: Their Vaporization, Detection and Structural Analysis*, Wiley, New York, 1996.
16. A. Townsend et al. (ed.), *Encyclopedia of Analytical Science*, Academic Press, San Diego, CA, 1995.
17. R.E. March, J.F.J. Todd (eds.), *Practical Aspects of Ion Trap Mass Spectrometry*, CRC Press, Boca Raton, FL, Vols. I–III, 1995.
18. P.K. Ghosh, *International Series of Monographs on Physics, 90: Ion Traps*, Oxford Science, New York, 1995.
19. I. Corndies, G.y. Horváth, K. Vékey (eds.), *Advances in Mass Spectrometry*, Proceedings of the 13th International Conference, 1995, Budapest, Elsevier, Amsterdam, 1995.
20. R.J. Cotter (ed.), *Time-of-flight Mass Spectrometry*, ACS Symposium Series 549, American Chemical Society, Washington, DC, 1994.
21. E.W. Schlag (ed.), *Time-of-flight Mass Spectrometry and its Applications*, Elsevier, Amsterdam, 1994.
22. T. Matsuo, R.M. Caprioli, M.L. Gross, Y. Seyama (eds.), *Biological Mass Spectrometry: Present and Future*, Wiley, New York, 1994.
23. A. Benninghoven, R. Shimizu, H.W. Werner, Y. Nihei, *Secondary Ion Mass Spectrometry: SIMS IX*, Wiley, New York, 1994.
24. J.J. Manura, C.W. Baker (eds.), *The Mass Spec Handbook of Service*, Scientific Instrument Services, Ringoes, NJ, Vol. 2, 1993.
25. K. Blau, J. Halket (eds.), *Handbook of Derivatives for Chromatography*, Wiley, New York, 1993.
26. A. Vertes, R. Gijbels, F. Adams, *Laser Ionization Mass Analysis*, Wiley, New York, 1993.
27. P.G. Kistemaker, N.M. Nibbering (eds.), *Advances in Mass Spectrometry*, Proceedings of the 12th International Conference, Amsterdam, 26–30 August 1991, Elsevier, Amsterdam, Vol. 12, 1992.
28. K.G. Standing, W. Ens (eds.), *Methods and Mechanisms for Producing Ions from Large Molecules*, Plenum Press, New York, 1991.
29. K.E. Jarvis, A.L. Gray, R.S. Houk, *Handbook of Inductively Coupled Plasma Mass Spectrometry*, Chapman Hall, London, 1991.
30. A.G. Marshall, F.R. Verdun, *Fourier Transforms in NMR, Optical and Mass Spectrometry: a User's Handbook*, Elsevier, Amsterdam, 1990.
31. K.E. Jarvis, A.L. Gray, J.G. Williams (eds.), *Plasma Source Mass Spectrometry*, Proceedings of the Third Surrey Conference on Plasma Source Mass Spectrometry, Royal Society of Chemistry, London, 1990.
32. E. Constantin, A. Schnell, (M.H. Chalmers, translator), *Mass Spectrometry*, Ellis Horwood, Chichester, 1990 (original French language edition, *Spectrométrie de Masse*, Tec & Doc, France).
33. D.M. Lubman (ed.), *Lasers in Mass Spectrometry*, Oxford University Press, Oxford, 1990.
34. H.L.C. Meuzelaar, T.L. Isenhour (eds.), *Computer-enhanced Analytical Spectroscopy*, Plenum Press, New York, Vol. 2, 1990.
35. *R.E. March, R.J. Hughes, *Quadrupole Storage Mass Spectrometry*, Wiley, New York, 1989.
36. Varian Vacuum Products Division, *Basic Vacuum Practice*, 2nd edition, Varian, Palo Alto, CA, 1989.
37. A. Benninghoven (ed.), *Ion Formation from Organic Solids: Mass Spectrometry of Involatile Materials*, Wiley, New York, 1989.
38. N. Heinrich, H. Schwarz, in *Ion and Cluster Ion Spectroscopy and Structure*, ed. J.P. Maier, Elsevier, Amsterdam, 1989.
39. R.G. Wilson, F.A. Stevie, C.W. Magee, *Secondary Ion Mass Spectrometry: A Practical Handbook for Depth Profiling and Bulk Impurity Analysis*, Wiley, New York, 1989.
40. L. Prokai, *Field Desorption Mass Spectrometry*, Marcel Dekker, New York, 1989.
41. B.S. Middleditch, *Analytical Artifacts: GC, MS, HPLC, TLC, and PC*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 44, 1989.
42. M.V. Buchanan (ed.), *Fourier Transform Mass Spectrometry*, American Chemical Society, Washington, DC, 1987.
43. N.A.B. Gray, *Computer-Assisted Structure Elucidation*, Wiley, New York, 1986.
44. J.H. Futrell (ed.), *Gaseous Ion Chemistry and Mass Spectrometry*, Wiley, New York, 1986.
45. H.E. Duckworth, R.C. Barber, V.S. Venkalasubramanian, *Mass Spectroscopy*, 2nd edition, Cambridge University Press, Cambridge, 1986.
46. F.A. White, G.M. Wood, *Mass Spectrometry: Applications in Science and Engineering*, Wiley, New York, 1986.
47. T.D. Mark, G.H. Dunn, *Electron Impact Ionization*, Springer, Berlin, 1985.
48. J.H. Beynon, M.L. McGlashan (eds.), *Current Topics in Mass Spectrometry and Chemical Kinetics*, Heyden, London, 1982.
49. A. Benninghoven, *Secondary Ion Mass Spectrometry: Sims III*, Springer, Berlin, 1982.
50. P. de Mayo (ed.), *Rearrangements in Ground and Excited States*, Academic Press, New York, Vols. 1–3, 1980.
51. C. Merritt, Jr, C.N. McEwen (eds.), *Practical Spectroscopy Series, Mass Spectrometry: Part B*, Marcel Dekker, New York, Vol. 3, 1980.
52. C. Merritt, Jr, C.N. McEwen (eds.), *Practical Spectroscopy Series, Mass Spectrometry: Part A*, Marcel Dekker, New York, Vol. 3, 1979.
53. D.R. Knapp, *Handbook of Analytical Derivatization Reactions*, Wiley-Interscience, New York, 1979.
54. M.T. Bowers (ed.), *Gas Phase Ion Chemistry*, Academic Press, New York, Vols. 1 and 2, 1979, Vol. 3, 1984.
55. J.L. Franklin (ed.), *Benchmark Papers in Physical Chemistry and Chemical Physics. Ion–Molecule Reactions.*

Part I. The Nature of Collisions and Reactions of Ions with Molecules and Ion–Molecule Reactions. Part II. Elevated Pressures and Long Reaction Times, Dowden, Hutchingson & Ross, Stroudsburg, PA, Vol. 3, 1979.
56. P.J. Ausloose (ed.), *Kinetics of Ion–Molecule Reactions*, Nato Advanced Studies Institute Series, Plenum Press, New York, 1979.
57. M.L. Gross (ed.), *High Performance Mass Spectrometry: Chemical Applications*, ACS Symposium Series 70, American Chemical Society, Washington, DC, 1978.
58. B.J. Millard, *Quantitative Mass Spectrometry*, Heyden, London, 1978.
59. R.G. Cooks (ed.), *Collision Spectroscopy*, Plenum Press, New York, 1978.
60. K. Levsen, *Fundamental Aspects of Organic Mass Spectrometry*, Verlag Chemie, Weinheim, 1978. This is Vol. 4 of a series entitled *Progress in Mass Spectrometry Fortschritte der Massenspektrometrie*, ed. H. Budzikiewicz, Band 1: M. Hesse, *Indolakaloide*, Teil 1 (Text), Teil 2 (Spektren). Vol. 2: S.E. Drewes, *Chroman and Related Compounds*. Band 3: M. Hesse, H.O. Bernhard, *Alkaloide (ausser Indol-, Triterpen- und Steroidalkaloide)*. The publication dates of these three previous volumes is not know, nor is it known if there are subsequent volumes.
61. J.R. Majer, *The Mass Spectrometer*, Taylor and Francis, Bristol, PA, 1977.
62. B.J. Gudzinowicz, M.J. Gudzinowicz, H.F. Martin, *Fundamentals of Integrated GC/MS* (in three parts). *Part III: The Integrated GC/MS Analytical System*, Marcel Dekker, New York, 1977.
63. P.H. Dawson (ed.), *Quadrupole Mass Spectrometry and its Applications*, Elsevier, Amsterdam, 1976 (reprinted by American Institute of Physics, Woodbury, NY, 1995).
64. B.J. Gudzinowicz, M.J. Gudzinowicz, H.F. Martin, *Fundamentals of Integrated GC/MS* (in three parts). *Part II: Mass Spectrometry*, Marcel Dekker, New York, 1976.
65. B.J. Gudzinowicz, M.J. Gudzinowicz, H.F. Martin, *Fundamentals of Integrated GC/MS* (in three parts). *Part I: Gas Chromatography*, Marcel Dekker, New York, 1976.
66. T.A. Lehman, M.M. Bursey, *Ion Cyclotron Resonance Spectrometry*, Wiley, New York, 1976.
67. S.G. Lias, P.J. Ausloos, *Ion–Molecule Reactions: Their Role in Radiation Chemistry*, American Chemical Society, Washington, DC, 1975.
68. G.R. Waller (ed.), *Biochemical Applications of Mass Spectrometry*, Wiley-Interscience, New York, 1972 (G.R. Waller, O.C. Dermer, 1st Supplement, 1980).
69. D.H. Williams, I. Howe, *Principles of Organic Mass Spectrometry*, 2nd edition, McGraw-Hill, London, 1972 (1st edition, 1964).
70. A. Maccoll (ed.), *Mass Spectrometry*, MTP International Review of Science, Butterworths, London, 1972.
71. J.L. Franklin (ed.), *Ion–Molecule Reactions*, Plenum Press, New York, Vols. 1 and 2, 1972.
72. G.W.A. Milne (ed.), *Mass Spectrometry: Techniques and Applications*, Wiley-Interscience, New York, 1971.
73. D.H. Williams (ed.), *Mass Spectrometry*, Chemical Society, London, Vol. 1, 1971, Vol. 2, 1973.
74. F.H. Field, J.L. Franklin, *Electron Impact Phenomena and the Properties of Gaseous Ions*, revised edition, Academic Press, New York, 1970 (1st edition, 1957).
75. E.W. McDaniel, V. Čhermák, A. Dalgarno, E.E. Ferguson, L. Friedman, *Ion–Molecule Reactions*, Wiley-Interscience, New York, 1970.
76. P.F. Knewstubb, *Mass Spectrometry and Ion–Molecule Reactions*, Cambridge University Press, Cambridge, 1969.
77. J. Roboz, *Introduction to Mass Spectrometry Instrumentation and Techniques*, Wiley, New York, 1968.
78. A.E. Pierce, *Silylation of Organic Compounds*, Pierce Chemical, Rockford, IL, 1968.
79. F.A. White, *Mass Spectrometry in Science and Technology*, Wiley, New York, 1968.
80. P.J. Ausloos (ed.), *Ion–Molecule Reactions in the Gas Phase*, ACS Advances in Chemistry Series 58, American Chemical Society, Washington, DC, 1966.
81. R.W. Kiser, *Introduction to Mass Spectrometry and its Application*, Prentice-Hall, Englewood Cliffs, NJ, 1965.
82. K. Biemann, *Mass Spectrometry: Organic Chemical Applications*, McGraw-Hill, New York, 1962 (reprinted by ASMS, 1998).
83. R.I. Reed, *Ion Production by Electron Impact*, Academic Press, London, 1962.
84. J.H. Beynon, *Mass Spectrometry and its Applications to Organic Chemistry*, Elsevier, Amsterdam, 1960 (reprinted by ASMS, 1999).

Technique-oriented Books

The books listed in this section pertain to specific techniques of MS and hyphenated chromatography/MS techniques. In some cases, the books are specific to certain types of analytes. Books of this type began to appear at the end of the 1960s (only one listing before 1970, Technique 91). Slightly more than half of these books (47 out of 90) have copyright dates in the last decade. Unfortunately, the last three or four years have seen a proliferation of very poorly written books with little or no copyediting and poor quality production.

One of the titles in this section (Technique 6), as well as one in the **Interpretation Books** section (Interpretation 3), is not of much merit. The *JASMS* reviews of both these books expressed an opinion that they would be more harmful to the reader than helpful [Technique 6: *J. Am. Soc. Mass Spectrom.*, **10**, 364–367 (1999); Interpretation 3: *J. Am. Soc. Mass Spectrom.*, **9**, 852–854 (1998)].

Many of the technique-oriented books are made obsolete by changing technology within a few years or months of their publication. A good example is the 1990 Yergey book (Technique 39) on LC/MS. Because of the massive advancements in technology that have taken place in the last 10 years, this book had little relevance to

the technique within 2 years of its publication. The book still has a great deal of value in that it provides good information on how to perform the chromatographic separations required in LC/MS, and it has an excellent set of journal-article references that contain titles. The problem of dated material can be especially significant with foreign-language books that are translated into English. Unfortunately, in some cases, publishers do not indicate that a book has been translated. It is only through careful research that the foreign-language routes of a book can be established. This is the case with Technique 2.

There are two books (Techniques 10 and 21) that are of particular value to the LC/MS and GC/MS practitioner, respectively. These two books have a great deal of practical information on the running of different types of analyses and are good aids in the decision-making process about how to proceed with a particular sample. The Willoughby book (Technique 10) has very useful information in deciding whether to use a contract laboratory or perform the analysis in-house. The Willoughby book was reviewed in *J. Am. Soc. Mass Spectrom.*, **10**, 78–79 (1999), as was the Kitson book, *J. Am. Soc. Mass Spectrom.*, **9**, 294–295 (1997). Care must be taken with respect to the unfortunate similarity between the titles of the Kitson book (Technique 21) and that of Technique 6.

Books that are edited works rather than having a single author generally do not get my approval. These edited editions often end up looking like a 'camel' (a horse designed by a committee). This is more true of books published in the last two decades than those published before that time. There is one notable exception in Technique 16 by Cole, reviewed in *J. Am. Soc. Mass Spectrom.*, **8**, 1191–1192 (1997). This book is an excellent reference for those working in ES. The second edition of the Niessen book (Technique 5) is also a good reference for ES and also other LC/MS techniques. The single negative aspect of both of these books is that they do not include titles with journal citations.

There are two important references on environmental GC/MS that should be reviewed by anyone working in this area (Techniques 71 and 73). Although both of these books were written in the era of the packed column, the fundamentals of environmental analyses and the United States Environmental Protection Agency (EPA) tune criteria are covered in detail.

Another reference of the packed-column era is the McFadden book (Technique 86). This book, along with the Karasek book (Introductory 18), is very useful to those starting in GC/MS.

If you are using CI (either APCI or CI under the conditions normally encountered in GC/MS), you need the Harrison book (Technique 33). Unlike the Yinon book (Technique 27), this book is a second edition and is labeled as such.

In looking at new titles of technique-orientated books, care must be taken to know when the title is nothing more than the collection of a series of articles from a journal or a bound issue of a journal. The value of such books is often less than their extremely high selling price.

1. *R. Willoughby, E. Sheehan, *A, Global View of MS/MS*, Global View, Pittsburgh, PA, 2000.
2. P. Gerhards, U. Bons, J. Sawazki, J. Szigan, A. Wertmann, *GC/MS in Clinical Chemistry*, Wiley, Chichester, 1999.
3. R.M. Smith, (K.L. Busch, tech. ed.), *Understanding Mass Spectra: a Basic Approach*, Wiley, New York, 1999.
4. G. Montaudo, R.P. Lattimer, *Mass Spectrometry of Polymers*, CRC Press, Boca Raton, FL, 1999.
5. W.M.A. Niessen, *Liquid Chromatography–Mass Spectrometry*, 2nd edition, Chromatographic Science Series, Marcel Dekker, New York, Vol. 79, 1998.
6. M. McMaster, C. McMaster, *GC–MS: A Practical User's Guide*, Wiley, New York, 1998.
7. W.M.A. Niessen, R.D. Voyksner (eds.), *Current Practice in Liquid Chromatography–Mass Spectrometry*, reprinted from *J. Chromatogr. A*, Elsevier, Amsterdam, Vol. 794, 1998.
8. W. Ens, K.G. Standing, I.V. Chernushevich (eds.), *New Methods for the Study of Biomolecular Complexes*, NATO ASI Series (Series C), Mathematical and Physical Science, Proceedings of the NATO Advanced Research Workshop on New Methods for the Study of Molecular Aggregates, The Lodge at Kananaskis Village, Alberta, Canada, 16–20 June 1996, Kluwer, Boston, MA, Vol. 510, 1998.
9. *B.S. Larsen, C.N. McEwen (eds.), *Mass Spectrometry of Biological Materials*, Marcel Dekker, New York, 1998.
10. *R. Willoughby, E. Sheehan, S. Mitrovich, *A Global View of LC/MS*, Global View, Pittsburgh, PA, 1998.
11. *E.M. Thurman, M.S. Mills, *Solid Phase Extraction: Principles and Practice*, Wiley, New York, 1998.
12. R.A. Meyers (ed.), *Encyclopedia of Environmental Analysis and Remediation*, Wiley, New York, 1998.
13. D. Briggs, I.M. Ward, S. Suresh, D.R. Clarke (eds.), *Surface Analysis of Polymers by XPS and Static SIMS*, Cambridge University Press, Cambridge, 1998.
14. R.M. Caprioli, A. Malorni, G. Sindona (eds.), *Selected Topics in Mass Spectrometry in the Biomolecular Sciences: a Tutorial*, NATO ASI Series (Series C), Mathematical and Physical Science, Altavilla-Milicia (PA) Italy, 7–18 July 1996, Kluwer, Boston, MA, Vol. 504, 1997.
15. R.B. Cole (ed.), *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications*, Wiley, New York, 1997.
16. R.P. Newton, T.J. Walton (eds.), *Proceedings of the Phytochemical Society of Europe, 40: Applications of Modern Mass Spectrometry in Plant Science Research*, Clarendon Press, Oxford, 1997.
17. W.S. Hancock, *New Methods in Peptide Mapping for the Characterization of Proteins*, CRC Press, Boca Raton, FL, 1996.

18. D. Barcelo (ed.), *Applications of LC/MS in Environmental Chemistry*, Journal of Chromatography Library, Elsevier, Amsterdam, Vol. 59, 1996.
19. A.L. Burlingame, S.A. Carr (eds.), *Mass Spectrometry in the Biological Sciences*, Humana Press, Totowa, NJ, 1996.
20. *J.R. Chapman (ed.), *Protein and Peptide Analysis by Mass Spectrometry*, Humana Press, Totowa, NJ, 1996.
21. *F.G. Kitson, B.S. Larsen, C.N. McEwen, *Gas Chromatography and Mass Spectrometry: A Practical Guide*, Academic Press, San Diego, CA, 1996.
22. R.M. Caprioli, A. Malorni, G. Sindona (eds.), *Mass Spectrometry in the Biomolecular Sciences: a Tutorial*, NATO ASI Series (Series C), Mathematical and Physical Science, Lacco Ameno, Ischia, Italy, June 23–July 5, 1993, Kluwer, Boston, MA, Vol. 475, 1996.
23. B.L. Karger, W.S. Hancock (eds.), *Methods in Enzymology, Vols. 270 and 271, High Resolution Separation and Analysis of Biological Macromolecules, Part A: Fundamentals and Part B: Applications*, Academic Press, San Diego, CA, 1996 (two separate books).
24. *T.W. Boutton, S.-I. Yamasaki (eds.), *Mass Spectrometry of Soils*, Marcel Dekker, New York, 1996.
25. *J.M. Walker (ed.), *The Protein Protocols Handbook*, Humana Press, Totowa, NJ, 1996.
26. A.P. Snyder (ed.), *Biochemical and Biotechnology Applications of Electrospray Ionization Mass Spectrometry*, ACS Symposium Series 619, American Chemical Society, Washington, DC, 1995.
27. J. Yinon (ed.), *Forensic Applications of Mass Spectrometry*, CRC Press, Boca Raton, FL, 1995.
28. C. Fenselau (ed.), *Mass Spectrometry for the Characterization of Microorganisms*, ACS Symposium Series 549, American Chemical Society, Washington, DC, 1994.
29. B. Ardrey (ed.), *Liquid Chromatography/Mass Spectrometry*, VCH, New York, 1993.
30. *R.C. Murphy, *Handbook of Lipid Research, No. 7, Mass Spectrometry of Lipids*, Plenum Press, New York, 1993.
31. W.M.A. Niessen, J. van der Greef, *Liquid Chromatography–Mass Spectrometry*, Chromatographic Science Series, Marcel Dekker, New York, Vol. 58, 1992.
32. M.L. Gross (ed.), *Mass Spectrometry in the Biological Sciences: a Tutorial*, NATO ASI Series, Kluwer, Boston, MA, 1992.
33. *A.G. Harrison, *Chemical Ionization Mass Spectrometry*, CRC Press, Boca Raton, FL, 1992.
34. J. St. Pyrek, 'Mass Spectrometry in the Chemistry of Natural Products', in *Recent Advances in Phytochemistry, Modern Phytochemical Methods*, eds. N.H. Fischer et al., Plenum Press, New York, Ch. 6, Vol. 25, 1991.
35. A.W. Czanderna, D.M. Hercules, *Ion Spectroscopies for Surface Analysis*, Plenum Press, New York, 1991.
36. E.R. Hilf (ed.), *Mass Spectrometry of Large Non-volatile Molecules for Marine Organic Chemistry*, World Scientific, River Edge, NJ, 1990.
37. *A. Fox, S.L. Morgan, L. Larsson, G. Odham (eds.), *Analytical Microbiology Methods: Chromatography and Mass Spectrometry*, Plenum Press, New York, 1990.
38. A.L. Burlingame, J.A. McCloskey (eds.), *Biological Mass Spectrometry*, Proceedings of the 2nd International Symposium of Mass Spectrometry in Health and Life Sciences, San Francisco, CA, August 27–31, 1989, Elsevier, Amsterdam, 1990.
39. *A.L. Yergey, C.G. Edmonds, I.A.S. Lewis, M.L. Vestal, *Liquid Chromatography/Mass Spectrometry: Techniques and Applications*, Plenum Press, New York, 1990.
40. J.M. Halket, M.E. Rose, *Introduction to Benchtop GC/MS*, HD Science, Stapleford, 1990.
41. C.H. Suelter, J.T. Watson (eds.), *Methods of Biochemical Analysis, Biomedical Applications in Mass Spectrometry*, Wiley-Interscience, New York, Vol. 34, 1990.
42. J.A. McCloskey (ed.), *Methods in Enzymology, Mass Spectrometry*, Academic Press, San Diego, CA, Vol. 193, 1990.
43. C.N. McEwen, B.S. Larsen (eds.), *Practical Spectroscopy Series: Mass Spectrometry of Biological Materials*, Marcel Dekker, New York, 1990.
44. M.A. Brown (ed.), *Liquid Chromatography/Mass Spectrometry: Applications in Agricultural, Pharmaceutical and Environmental Chemistry*, ACS Symposium Series 420, American Chemical Society, Washington, DC, 1990.
45. R.M. Caprioli (ed.), *Continuous-flow Fast Atom Bombardment Mass Spectrometry*, Wiley, New York, 1990.
46. D.M. Desiderio (ed.), *Mass Spectrometry of Peptides*, CRC Press, Boca Raton, FL, 1990.
47. SCIEX, *The API Book*, SCIEX, Division of MDS Health Group, Mississauga, Ontario, 1990.
48. J.W. Dolan, L.R. Snyder, *Troubleshooting LC Systems*, Humana Press, Totowa, NJ, 1989.
49. T.R. Ashe, K.V. Wood (eds.), *Novel Techniques in Fossil Fuel Mass Spectrometry*, ASTM, Philadelphia, PA, 1989.
50. J.M. Farrar, W.H. Saunders, *Techniques for the Study of Ion–Molecule Reactions* (Techniques of Chemistry), Wiley, New York, Vol. 20, 1988.
51. C.J. Biermann, G.D. McGinnis (eds.), *Analysis of Carbohydrates by GLC and MS*, CRC Press, Boca Raton, FL, 1988.
52. F. Adams, *Inorganic Mass Spectrometry*, Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications, Wiley, New York, 1988.
53. S.-T.F. Lai, *Gas Chromatography/Mass Spectrometry Operation*, Realistic Systems, East Longmeadow, MA, 1988.
54. K.L. Busch, G.L. Glish, S.A. McLuckey (eds.), *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*, VCH, New York, 1988.
55. C.J. McNeal (ed.), *The Analysis of Peptides and Proteins by Mass Spectrometry*, Proceedings of the 4th Texas Symposium, College Station, TX, April 17–20, 1988, Wiley, New York, 1988.

56. J. Gilbert (ed.), *Applications of Mass Spectrometry in Food Science*, Elsevier, London, 1987.
57. E. Heinzle, M. Reuss (eds.), *Mass Spectrometry in Biotechnological Process Analysis and Control*, Plenum Press, New York, 1987.
58. J.D. Rosen (ed.), *Applications of New Mass Spectrometry Techniques in Pesticide Chemistry*, Chemical Analysis: a Series of Monographs on Analytical Chemistry and its Applications, ed. J.D. Winefordner, Ed. Emeritus, Kolthoff, IM, Wiley-Interscience, New York, Vol. 91, 1987.
59. *H.F. Linskens, J.F. Jackson (eds.), *Modern Methods of Plant Analysis: Gas Chromatography/Mass Spectrometry*, New Series, Springer, Berlin, Vol. 3, 1986.
60. C.J. McNeal (ed.), *Mass Spectrometry in the Analysis of Large Molecules*, Wiley, Chichester, 1986.
61. S.J. Gaskell (ed.), *Mass Spectrometry in Biomedical Research*, Wiley, Chichester, 1986.
62. T. Aczel (ed.), *Mass Spectrometric Characterization of Shale Oils*, ASTM, Philadelphia, PA, 1986.
63. P.A. Lyon (ed.), *Desorption Mass Spectrometry: are SIMS and FAB the Same?*, American Chemical Society, Washington, DC, 1985.
64. F.W. Karasek, O. Hutzinger, S. Safe (eds.), *Mass Spectrometry in Environmental Sciences*, Plenum Press, New York, 1985.
65. S. Facchetti (ed.), *Mass Spectrometry of Large Molecules*, Elsevier, Amsterdam, 1985.
66. G.M. Message, *Practical Aspects of Gas Chromatography/Mass Spectrometry*, Wiley, New York, 1984.
67. G. Odham, L. Larsson, P.-A. Mardh (eds.), *Gas Chromatography/Mass Spectrometry: Applications in Microbiology*, Plenum Press, New York, 1984.
68. F.W. McLafferty (ed.), *Tandem Mass Spectrometry*, Wiley-Interscience, New York, 1983.
69. *S.I. Goodman, S.P. Markey, *Diagnosis of Organic Academias by Gas Chromatography–Mass Spectrometry, Laboratory and Research Methods in Biology and Medicine*, Alan R. Liss, New York, Vol. 6, 1981.
70. H.L.C. Meuzelaar, J. Haverkamp, S.D. Hileman, *Pyrolysis Mass Spectrometry of Biomaterials*, Elsevier, Amsterdam, 1980.
71. W.L. Budde, J.W. Eichelberger, *Organics Analysis Using Gas Chromatography/Mass Spectrometry*, Ann Arbor Science, Ann Arbor, MI, 1979.
72. B.S. Middleditch (ed.), *Practical Mass Spectrometry*, Plenum Press, New York, 1979.
73. L.H. Keith (ed.), *Identification and Analysis of Organic Pollutants in Water*, Ann Arbor Science, Ann Arbor, MI, 1979.
74. A.F. Frigerio (ed.), *Recent Developments in Mass Spectrometry in Biochemistry and Medicine*, Proceedings of the 5th International Symposium on Mass Spectrometry in Biochemistry and Medicine, Rimini, Italy, June 1978, Plenum Press, New York, Vol. 2, 1979.
75. *J.P. Payne, J.A. Bushman, D.W. Hill (eds.), *Medical and Biological Applications of Mass Spectrometry*, Academic Press, London, 1979.
76. D.G. Land, H.E. Nursten, *Progress in Flavour Research*, Applied Science, London, 1979.
77. B.J. Millard, *Quantitative Mass Spectrometry*, Heyden, London, 1978.
78. A.F. Frigerio, E.L. Ghisalberti (eds.), *Mass Spectrometry in Drug Metabolism*, Proceedings of the International Symposium on Mass Spectrometry in Drug Metabolism, Mario Negri Institute for Pharmacological Research, Milan, Italy, June 1976, Plenum Press, New York, 1977.
79. B.J. Gudzinowicz (ed.), *Analysis of Drugs and Metabolites by Gas Chromatography: Mass Spectrometry, Respiratory Gases, Ethyl Alcohol, and Related Toxicological Materials*, Vol. 1, 1977; *Hypnotics, Anticonvulsants, and Sedatives,* Vol. 2*; Antipsychotics, Antiemetics, and Antidepressant Drugs,* Vol. 3*; Central Nervous System Stimulants,* Vol. 4*; Analgesics, Local Anaesthetics, and Antibiotics*, Vol. 5, 1978; *Cardiovascular, Antihypertensive, Hypoglycemic, and Tiered-related Agents*, Vol. 6, 1979; subtitle unknown, Vol. 7, 1980, Marcel Dekker, New York.
80. H.D. Beckey, *Principles of Field Ionization and Field Desorption Mass Spectrometry*, Pergamon Press, New York, 1977.
81. Y. Masada, *Analysis of Essential Oils by Gas Chromatography and Mass Spectrometry*, Halsted Press, New York, 1976 (© 1976, Hirokawa, Japan).
82. Z.V. Zaretskii, *Mass Spectrometry of Steroids*, Wiley, New York, 1976.
83. *A.F. Frigerio, E.L. Ghisalberti (eds.), *Mass Spectrometry in Biochemistry and Medicine*, Monographs of the Mario Negri Institute for Pharmacological Research, Milan, Raven Press, New York, 1974.
84. R. Haque, F.J. Biros (eds.), *Mass Spectrometry and NMR Spectroscopy in Pesticide Chemistry*, Plenum Press, New York, 1974.
85. R.G. Cooks, J.H. Beynon, R.M. Caprioli, G.R. Lester, *Metastable Ions*, Elsevier, New York, 1973.
86. W. McFadden, *Techniques of Combined Gas Chromatography/Mass Spectrometry: Applications in Organic Analysis*, Wiley-Interscience, New York, 1973.
87. E. Costa, B. Holmstedt (eds.), *Gas Chromatography–Mass Spectrometry in Neurobiology*, Raven Press, New York, 1973.
88. A.J. Ahearn (ed.), *Trace Analysis by Mass Spectrometry*, Academic Press, New York, 1972.
89. C.E. Melton, *Principles of Mass Spectrometry and Negative Ions*, Marcel Dekker, New York, 1970.
90. A.L. Burlingame, N. Castagnoli (eds.), *Topics in Organic Mass Spectrometry*, Wiley-Interscience, New York, 1970.
91. L.S. Ettre, W.H. McFadden (eds.), *Ancillary Techniques of Gas Chromatography*, Wiley, New York, 1969.

Interpretation Books

The seminal book for the interpretation of EI mass spectra is the McLafferty book (Interpretation 6). This book is extremely valuable, but may be too advanced for a beginner trying to self-teach. The beginner should try to start with either the McLafferty 2nd edition book (Interpretation 8) or the book by Shrader (Interpretation 1), both of which have been out of print for some time but can be found on the used-book Web sites.

The books by Budzikiewicz et al. (Interpretations 19, 23, and 24) were the first books written using mechanisms in organic reactions to describe the fragmentation of energetic ions produced by EI. The information in these books still has a great deal of relevance to the subject. All of the interpretation books listed in this section pertain primarily to odd-electron molecular ions formed by the EI process. The ions formed in LC/MS are predominately protonated molecules, which are even-electron ions. Even-electron ions fragment to produce other even-electron ions, which requires the breaking of more than one bond in the ion. Many of the neutral losses in even-electron ion fragmentation and odd-electron ion fragmentation are the same.

One important note on Interpretation 3: just as book titles can be similar (Techniques 6 and 22), people can have the same or similar names that can result in confusion. The author of Interpretation 3, Terrence A. Lee, a Department of Chemistry faculty member at Middle Tennessee State University in Murfreesboro, TN, should not be confused with Terry Lee, a noted researcher in MS of biological substances at the Beckman Research Institute, City of Hope/Division of Immunology in Duarte, CA.

1. S.R. Shrader, *Introductory Mass Spectrometry*, 2nd edition, Shrader Laboratories, Detroit, MI, 1999 (originally published by Allyn and Bacon, Boston, MA, 1971).
2. J. Barker, *Mass Spectrometry: Analytical Chemistry by Open Learning*, 2nd edition, ed. D.J. Ando, Wiley, Chichester, 1999 (R. Davis, M.J. Frearson, 1st edition, 1987).
3. T.A. Lee, *A Beginner's Guide to Mass Spectral Interpretation*, Wiley, Chichester, 1998.
4. *J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, Philadelphia, 1997.
5. J.S. Splitter, F. Tureček (eds.), *Applications of Mass Spectrometry to Organic Stereochemistry*, VCH, New York, 1994.
6. F.W. McLafferty, F. Tureček, *Interpretation of Mass Spectra*, 4th edition, University Science, Mill Valley, CA, 1993.
7. Q.N. Porter, *Mass Spectrometry of Heterocyclic Compounds*, 2nd edition, Wiley-Interscience, New York, 1985.
8. F.W. McLafferty, R. Venkataraghavan, *Mass Spectral Correlations*, 2nd edition, American Chemical Society, Washington, DC, 1982.
9. B. Sklarz (ed.), *Mass Spectrometry of Natural Products*, plenary lecturers presented at the International Mass Spectrometry Symposium on Natural Products, Rehovot, Israel, 28 August–2 September, 1977, Pergamon Press, Oxford, 1978.
10. K. Levsen, *Fundamental Aspects of Organic Mass Spectrometry*, Verlag Chemie, Weinheim, 1978.
11. D.C. DeJongh, *Interpretation of Mass Spectra*, ACS Audio Series, American Chemical Society, Washington, DC, 1975 (6 audio cassette tapes, 158 pp.).
12. F.W. McLafferty, *Interpretation of Mass Spectra*, 2nd edition, Benjamin, Reading, MA, 1973.
13. M.G. Hamming, N.G. Foster, *Interpretation of Mass Spectra of Organic Compounds*, Academic Press, New York, 1972.
14. H.C. Hill, *Introduction to Mass Spectrometry*, 2nd edition, Heyden, London, 1972 (1st edition, 1966).
15. J. Seibl, *Massenspektrometrie*, Akademische Verlagsgesellschaft, Frankfurt, 1970.
16. R. Brymner, J.R. Penney (eds.), *Mass Spectrometry*, Chemical Publishing, New York, 1969.
17. J.H. Beynon, R.A. Saunders, A.E. Williams, *The Mass Spectra of Organic Molecules*, Elsevier, Amsterdam, 1968.
18. A.A. Polyakova, R.A. Khmel'nitskii, (J. Schmorak, translator), *Introduction to Mass Spectrometry of Organic Compounds*, Israel Program for Scientific Translations, Jerusalem, Israel, 1968 (original Russian language edition, *Vvedenie v Mass Spektrometriyu Organicheskikh Soedinenii*, Khimya, Moscow, 1966).
19. H. Budzikiewicz, C. Djerassi, D.H. Williams, *Mass Spectrometry of Organic Compounds*, Holden-Day, San Francisco, CA, 1967.
20. R.I. Reed, *Applications of Mass Spectrometry to Organic Chemistry*, Academic Press, New York, 1966.
21. G. Spiteller, *Massenspektrometrische Strukturanalyse Organischer Verbindungen*, Verlag Chemie, Weinheim, 1966.
22. A. Quayle, R.I. Reed, 'Interpretation of Mass Spectra', in *Interpretation of Organic Spectra*, ed. D.W. Mathieson, Academic Press, New York, 1965.
23. H. Budzikiewicz, C. Djerassi, D.H. Williams, *Structure Elucidation of Natural Products by Mass Spectrometry, Alkaloids, Vol. I, Steroids, Terpenoids, Sugars, and Miscellaneous Natural Products*, Holden-Day, San Francisco, CA, Vol. II, 1964.
24. H. Budzikiewicz, C. Djerassi, D.H. Williams, *Interpretation of Mass Spectra of Organic Compounds*, Holden-Day, San Francisco, CA, 1964.
25. F.W. McLafferty (ed.), *Mass Spectra of Organic Ions*, Academic Press, New York, 1963.
26. J.H. Beynon, A.E. Williams, *Mass and Abundance Tables for Use in Mass Spectrometry*, Elsevier, New York, 1963.
27. F.W. McLafferty, 'Mass Spectrometry', in *Determination of Organic Structures by Physical Methods*, eds. F.C. Nachod, W.D. Phillips, Academic Press, New York, Vol. II, 1962.

Books of Historical Significance

* Note that historical references are in the Bibliography section of the Kiser book (Reference 81) and in the Information and Data chapter of the Roboz book (Reference 77).

Based on an 1886 paper [*Berl. Ber.*, **39**, 691 (1886)] by Eugene Goldstein (German physicist, 1850–1930) reporting the discovery of luminous rays emerging as straight lines from holes in a metal disc used as a cathode in a discharge tube (he called the rays Kanalstrahlen – canal rays) and the confirmation by Wilhelm Carl Werner Otto Fritz Franz Wien (German Nobel Laureate in Physics, 1911; 1864–1928) that Jean Baptiste Perrin's (French physicist, 1870–1942) 1895 postulation that the rays were associated with positive charge by studying their deflection in electric and magnetic fields [*Verh. Phys. Ges.*, **17**, 1898 (1898); *Ann. Phys.*, **65**, 440 (1898); *Ann. Phys. Leipzig*, **8**, 224 (1902)], the field of MS developed into a science between 1911 and 1925. This was due to the results of the experiments conducted by the three founding fathers of MS: Joseph John Thomson (English Nobel Laureate in Physics, 1906; 1856–1940); Francis William Aston (English Nobel Laureate in Chemistry, 1922; 1877–1945; Aston was an associate of Thomson in the Cavendish Laboratory in Manchester, England); and Arthur Jeffery Dempster (1886–1950, Canadian–American physics professor, University of Chicago).

In his 1968 book, Roboz (Reference 77) lists 20 selected papers for those wanting to learn the history of MS through original references (Chapter 14, page 490). Of these papers, five were authored by Aston, four by Dempster, and two by Thomson. Another three were authored by William R. Smythe (US scientist) and two by Kenneth Bainbridge (US physicist, 1904–1906, Director of the Trinity test – the first test explosion of the atomic bomb), who also were early pioneers in MS.

1. E.A. Davis, I.J. Falconer, *J.J. Thomson and the Discovery of the Electron*, Taylor & Francis, London, 1997.
2. P.F. Dahl, *Flash of the Cathode Rays: A History of JJ Thomson's Electron*, American Institute of Physics, Philadelphia, PA, 1997.
3. F.C. Nachod, J.J. Zuckerman, E.W. Randall (eds.), *Determination of Organic Structures by Physical Methods*, Academic Press, New York, Vol. 6, 1976.
4. A series of books (*Mass Spectrometry in Biochemistry and Medicine, Vols. 1 and 2, Advances in Mass Spectrometry in Biochemistry and Medicine*, 1974, 1975; *Mass Spectrometry in Drug Metabolism*, 1976, Vols. 1, 2, 6, 7 and 8, *Recent Developments in Mass Spectrometry in Biochemistry and Medicine*, 1977–1980 and 1982), ed. A. Frigerio, published by Plenum and Elsevier, as the proceedings of a meeting organized by the Mario Negri Institute for Pharmacological Research in Milan, Italy.
5. Biennial Specialist Periodical Reports – *Mass Spectrometry, a Review of the Recent Literature Published between July 19XX and June* $19XX + 2$, Vol. 1, (1968–1970), 1971; Vol. 2, (1970–1972), 1973, ed. D.H. Williams; Vol. 3, (1972–1974), 1975; Vol. 4, (1974–1976), 1977; Vol. 5, (1976–1978), 1979; Vol. 6, (1978–1980), 1981; Vol. 7, (1980–1982), 1984, ed. R.A.W. Johnstone; Vol. 8, (1982–1984), 1985; Vol. 9 (1984–1986), 1987; Vol. 10, (1986–1988), 1989, ed. M.E. Rose, Royal Society of Chemistry (formerly Chemical Society, Cambridge).
6. The proceedings the seven European Time-of-flight Symposia have been published by Heyden, London, eds. D. Price, J.E. Williams, *Time-of-flight Mass Spectrometry* (1st), 1969; all subsequent proceedings entitled *Dynamic Mass Spectrometry*, Vol. 1, 1970 (2nd); Vol. 2, 1971 (3rd); Vol. 3, 1972 (4th), ed. D. Price, Vol. 4, 1976 (5th), eds. D. Price, J.F.J. Todd, Vol. 5, 1978 (6th); Vol. 6, 1981 (7th).
7. H.S.W. Massey, E.H.S. Burhop, H.B. Gilbody, *Electronic and Ionic Impact Phenomena*, 2nd edition, *Electron Collisions with Atoms*, Vol. I, 1969; *Electron Collisions with Molecules – Photoionization*, Vol. II, 1969; *Slow Collisions of Heavy Particles*, Vol. III, 1971; *Recombination and Fast Collisions of Heavy Particles*, Vol. IV, 1974; *Slow Positron and Muon Collisions – Notes on Recent Advances*, 1974, Oxford University Press, London.
8. H. Kientiz, *Massenspektrometrie*, Verlag Chemie, Weinheim, 1968.
9. R.I. Reed (ed.), *Modern Aspects of Mass Spectrometry*, Proceedings of the 2nd NATO Advanced Study Institute of Mass Spectrometry on Theory, Design, and Applications, July 1966, University of Glasgow, Plenum Press, New York, 1968.
10. E.W. Blauth, *Dynamic Mass Spectrometers* (translated from German), Elsevier, Amsterdam, 1966.
11. R. Jayaram, *Mass Spectrometry: Theory and Applications*, Plenum Press, New York, 1966.
12. W.L. Mead (ed.), *Advances in Mass Spectrometry*, Pergamon Press, New York, Vol. 3, 1966.
13. G.P. Thomson, *J.J. Thomson and the Cavendish Laboratory in His Day*, Doubleday, New York, 1965.
14. R.I. Reed (ed.), *Mass Spectrometry*, Proceedings of the 1st NATO Advanced Study Institute of Mass Spectrometry on Theory, Design, and Applications, Academic Press, London, 1965.
15. E.W. McDaniel, *Collision Phenomena in Ionized Gases*, Wiley, New York, 1964.
16. C.A. McDowell (ed.), *Mass Spectrometry*, McGraw-Hill, New York, 1963 (reprinted by Robert E. Krieger, Huntington, NY, 1979).
17. R.M. Elliot (ed.), *Advances in Mass Spectrometry*, Pergamon Press, New York, Vol. 2, 1963.
18. J.D. Waldron (ed.), *Advances in Mass Spectrometry*, Pergamon Press, New York, Vol. 1, 1959.
19. H.E. Duckworth, *Mass Spectroscopy*, University Press, Cambridge, 1958.

20. G.R. Rieck, *Einführung in die Massenspektroskopie* (translated from Russian), VEB Deutscher Verlag der Wissenschaften, Berlin, 1956.
21. L.B. Loeb, *Basic Processes of Gaseous Electronics*, University of California Press, Berkeley, CA, 1955 (reprinted in 1960 with Appendix I).
22. A.J.B. Robertson, *Mass Spectrometry: Methuen's Monographs on Chemical Subjects*, Wiley, New York, 1954.
23. J. Blears, (Chairman, Mass Spectrometry Panel, Institute of Petroleum), *Applied Mass Spectrometry*, a report of a conference organized by the Mass Spectrometry Panel of the Institute of Petroleum, London, 29–31 October 1953, Institute of Petroleum, London, 1954.
24. G.P. Barnard, *Modern Mass Spectrometry*, Institute of Physics, London, 1953.
25. J.A. Hipple, L.T. Aldrich, A.O.C. Nier, V.H. Dibeler, F.L. Mohler, R.E. O'Dette, H. Odishaw, H. Sommer, (Mass Spectroscopy Committee), *Mass Spectrometry in Physics Research*, National Bureau of Standards Circular 522, United States Government Printing Office, Washington, DC, 1953.
26. H. Ewald, H. Hintenberger, *Methoden und Anwendunyngen der Massenspektroskopie*, Verlag Chemie, Weinheim, 1952; English translation by USAEC, Translation Series AEC-tr-5080, Office of Technical Service, Washington, DC, 1962.
27. H.S.W. Massey, E.H.S. Burhop, *Electronic and Ionic Impact Phenomena*, Oxford University Press, London, 1952 (2nd printing, 1956).
28. H.S.W. Massey, *Negative Ions*, 2nd edition, Cambridge University Press, Cambridge, 1950 (1st edition, 1933).
29. N.F. Mott, H.S.W. Massey, *The Theory of Atomic Collisions*, 2nd edition, Oxford University Press, London, 1949.
30. F.W. Aston, *Mass Spectrometry and Isotopes*, 2nd edition, Edward Arnold, London, 1942 (1st edition, 1933).
31. F.W. Aston, *Isotopes*, 2nd edition, Edward Arnold, London, 1924 (1st edition, 1922).
32. J.J. Thomson, *Rays of Positive Electricity and their Application to Chemical Analysis*, 2nd edition, Longmans Green, London, 1921 (1st edition, 1913).

Collections of Mass Spectra in Hard Copy

There have been many collections of mass spectra that have come and gone. In a 1985 monograph (*A Guide to, and Commentary on, the Published Collection and Literature of Mass Spectral Data*) published by VG Analytical (the UK MS company now known as Micromass/Waters), 33 separate collections were referenced. In 1974 and 1978, the ASMS published the 1st and 2nd editions of *A Guide to Collections of Mass Spectral Data*. These editions include 24 and 30 references, respectively. None of these collections has been lost. They have all been consolidated into either the NIST98 Mass Spectral Database or the Wiley Registry of Mass Spectral Data, or both. Some of these collections are not currently available in an electronic format, or the electronic format is only of abbreviated spectra that range from a minimum of 16 to a maximum of 50 mass spectral peaks (Collections 6). Hard-copy volumes are somewhat less valuable than electronic versions. The Cornu collection is even less valuable because it is a tabular listing of the 10 most intense peaks.

1. H.L.J. Makin, D.J.H. Trafford, J. Nolan, *Mass Spectra and GC Data of Steroids: Androgens and Estrogens*, Wiley, Chichester, 1999.
2. R. Newman, M.W. Gilbert, K. Lothridge, *GC/MS Guide to Ignitable Liquids*, CRC Press, Boca Raton, FL, 1998.
3. ‡J.C. Vickerman, D. Briggs, A. Henderson (eds.), *The Wiley Static SIMS Library*, Wiley, New York, 1996.
4. ‡R.P. Adams, *Identification of Essential Oil Compounds by Gas Chromatography/Mass Spectrometry*, Allured, Carol Stream, IL, 1995 (1211 spectra).
5. ‡T. Mills III, J.C. Roberson, *Instrumental Data For Drug Analysis*, 2nd edition, CRC Press, Boca Raton, FL, Vols. 1–7, 1993 (originally published by Elsevier, New York, 1987–1992).
6. ‡K. Pfleger, W.W. Maurer, A. Weber, *Mass Spectral and GC Data of Drugs, Pollutants, Pesticides and Metabolites*, 2nd edition, 3-volume set, VCH, New York, 1992.
7. R.A. Hites, *CRC Handbook of Mass Spectra of Environmental Contaminants*, 2nd edition, CRC Press, Boca Raton, FL, 1992 (533 spectra).
8. F.W. McLafferty, D.B. Stauffer, *Important Peak Index of the Registry of Mass Spectral Data*, 3 volumes, Wiley, New York, 1991.
9. *The Eight Peak Index of Mass Spectra*, 4th edition, Royal Society of Chemistry, Cambridge, 1991.
10. ‡F.W. McLafferty, D.B. Stauffer, *The Wiley/NBS Registry of Mass Spectral Data*, 7 volumes, Wiley, New York, 1989 (133 000 spectra).
11. E.A. Stemmler, R.A. Hites, *Electron Capture Negative Ion Mass Spectra of Environmental Contaminants and Related Compounds*, VCH, New York, 1988 (361 spectra).
12. †S.R. Heller, G.W.A. Milne, L.H. Gevantman, *EPA/NIH Mass Spectral Data Base, Supplement 2, 1983*, National Standard Reference Data System, National Bureau of Standards, Department of Commerce, Washington, DC, 1983 (6557 spectra).
13. †Sunshine et al., *CRC Handbook of Mass Spectra of Drugs*, CRC Press, Boca Raton, FL, 1981 (1208 EI spectra and 628 CI spectra).
14. †B.S. Middleditch, S.R. Missler, H.B. Hines, *Mass Spectrometry of Priority Pollutants*, Plenum Press, New York, 1981 (114 spectra).
15. †S.R. Heller, G.W.A. Milne, *EPA/NIH Mass Spectral Data Base, Supplement 1, 1980*, National Standard

‡ Also available in electronic format.
† Out of print.

Reference Data System, National Bureau of Standards, Department of Commerce, Washington, DC, 1980 (8807 spectra).

16. †‡S.R. Heller, G.W.A. Milne, *EPA/NIH Mass Spectral Data Base*, 4 volumes and index, National Standard Reference Data System, National Bureau of Standards, Department of Commerce, Washington, DC, 1978 (23 556 spectra).
17. †A. Cornu, R. Massot, *Compilation of Mass Spectral Data*, 2nd edition, 2 volumes, Heyden, Philadelphia, PA, 1975 (10 000 spectra).
18. †S. Safe, O. Hutzinger, *Mass Spectrometry of Pesticides and Pollutants*, CRC Press, Cleveland, OH, 1973 (275 spectra).

Mass Spectrometry, Gas Chromatography/Mass Spectrometry, and Liquid Chromatography/Mass Spectrometry Journals

Articles containing information on MS can be found in many different scientific journals in addition to those listed below. This is a listing of journals that are specific to MS (Journals 1, 2, 3, 4, 5, 6, 8, 9, 15, and 16), that pertain to a specific analytical technique (Journals 10 and 11), or that pertain to general chemistry (Journals 7, 12, 13, and 14). Some of the journals have complimentary subscriptions (Journals 12, 13, and 14), whereas other journals have annual subscription rates of thousands of dollars (Journals 1, 4, 6, 10, and 11). Some journals have reasonable society membership rates (Journals 2 and 7). The more expensive journals often have reasonable individual subscription prices (Journals 1 and 4).

In addition to review and research articles, most of these journals also provide reviews of software, books, and other items of interest to the mass spectrometrist. The exceptions are the proceedings of meetings (Journals 16 and 17) and list and/or abstract sources (Journals 8 and 9).

One of the interesting features of *Journal of Mass Spectrometry* (Journals 1) is a section entitled 'Current Literature in MS' that appears at the end of every issue. This is a bibliography of articles published over the past 6–8 weeks. It is divided into 11 major sections with the Biology/Biochemistry section subdivided into four additional categories. At the end of each volume, all the listings for the year are made available in a Microsoft® Access format that can be searched electronically.

1. *Journal of Mass Spectrometry*, Wiley, Chichester (formerly *Organic Mass Spectrometry*, incorporating *Biomedical Mass Spectrometry*).
2. *Journal of the American Society for Mass Spectrometry*, Elsevier, New York.
3. *European Mass Spectrometry*, IM Publications, Chichester.
4. *Rapid Communications in Mass Spectrometry*, Wiley, Chichester.
5. *Mass Spectrometry Reviews*, Wiley, New York.
6. *International Journal of Mass Spectrometry and Ion Processes*, Elsevier, New York.
7. *Analytical Chemistry*, American Chemical Society, Washington, DC.
8. *CA Selects Plus: Mass Spectrometry*, American Chemical Society, Washington, DC.
9. *Mass Spectrometry Bulletin*, Royal Society of Chemistry, Cambridge.
10. *Journal of Chromatography A*, Elsevier, Amsterdam.
11. *Journal of Chromatography B*, Elsevier, Amsterdam.
12. *American Laboratory*, ISC, Shelton, CT.
13. *LC/GC*, Advanstar, Eugene, OR.
14. *Spectroscopy*, Advanstar, Eugene, OR.
15. *Proceedings of the nth ASMS Conference on Mass Spectrometry and Allied Topics*, published annually, American Society for Mass Spectrometry, Santa Fe, NM (ASTM E14 Committee Meetings began in 1952 – published annually from 1961 to 1969, ASTM E14 Committee Meeting Proceedings – ASMS began in 1970 with the 18th Conference Proceedings).
16. *Advances in Mass Spectrometry*, Proceedings of the Triennial International Mass Spectrometry Conference. Vols. 1, 2, and 3 (1958, 1961, and 1964) can be found in the **Books of Historical Significance** section and Vols. 12, 13, and 14 (1991, 1994, and 1997) in the **Reference Books** section of this bibliography. There are Vols. 4–11 for meeting years 1967, 1970, 1973, 1976, 1979, 1982, 1985, and 1988. The first six meetings were held in the UK. Beginning with the 7th meeting (Florence, Italy), each succeeding meeting has been held in a different European country. The next meeting is scheduled for Barcelona, Spain, in 2000.

Monographs

All the citations in this section are from VG Instruments/Micromass. All instrument manufacturers publish application notes; however, Micromass (and its preceding companies) is the only manufacturer that has published this type of general-topic monograph. These monographs are like review articles found in *Mass Spectrometry Reviews* or the Special Features section of the *Journal of Mass Spectrometry*. These monographs are not as well referenced as the articles in the two journals mentioned, but do provide a good overview of the subject. Such promotional material is of benefit to those wanting to get a quick understanding of a topic, and it is hoped that more of this type of material will be forthcoming.

Another example of the ready-reference-material approach to information dissemination is found in the Siuzdak book (Introductory 6). This book was written to provide a quick understanding of MS to biotechnology executives who have to make financial decisions about MS instrumentation and facilities.

1. M.E. Rose, *Modern Practice of Gas Chromatography/Mass Spectrometry*, VG Monographs in Mass Spectrometry, No. 1, VG Instruments, Altrincham.
2. F.A. Mellon, *Liquid Chromatography/Mass Spectrometry*, VG Monographs in Mass Spectrometry, No. 2, VG Instruments, Altrincham.
3. M.R. Clench, *A Comparison of Thermospray, Plasmaspray, Electrospray and Dynamic FAB*, VG Monographs in Mass Spectrometry, No. 3, VG Instruments, Altrincham.
4. J.H. Scrivens, K. Rollins, *Tandem Mass Spectrometry*, VG Monographs in Mass Spectrometry, No. 4, VG Analytical, Fisons Instruments, Altrincham.
5. B. Ardrey, *Mass Spectrometry in the Forensic Sciences*, VG Monographs in Mass Spectrometry, No. 5, VG Analytical, Fisons Instruments, Altrincham.
6. J.-P. Hsu, *High and Low Resolution GC/MS in Environmental Sciences*, VG Monographs in Mass Spectrometry, No. 6, VG Analytical, Fisons Instruments, Altrincham.

Personal Computer Mass Spectrometry Abstract Sources

1. *Current Contents® on CD-ROM, Physical, Chemical and Earth Sciences*, Institute for Scientific Information, Philadelphia, PA; FREE Demo available.
2. *Analytical Abstracts on CD-ROM*, Royal Society of Chemistry, Cambridge, or SilverPlatter Information, Norwood, MA; FREE 30-day trial subscription.
3. *CA Surveyor: Mass Spectrometry and Applications*, Chemical Abstracts Service, American Chemical Society, Washington, DC; FREE Demo available.
4. *LC/MS Update* (1991–current), *GC/MS Update Part A: Environmental* (1991–1996), and *GC/MS Update Part B: Biomedical, Clinical, Drugs*, includes forensics (1991–current), HD Science, Newport, Wilmington, DE, or HD Science, Nottingham.
5. *The PC Version of the Mass Spectrometry Bulletin*, Royal Society of Chemistry, Cambridge.
6. Annual Collections of the 'Current Literature in Mass Spectrometry', in *Journal of Mass Spectrometry*, (1995–1998), Wiley, Chichester.

Integrated Spectral Interpretation Books

General integrated spectral interpretation books include information on the interpretation of proton NMR, IR, and mass spectra, as well as how to use UV data in conjunction with the previous three spectral techniques. Some books also include a section on ^{13}C-NMR. These books are good for an overview of the subject, but do not provide the in-depth MS interpretational information.

In addition to the books listed below, the Mathieson book (Interpretation 22) is also an integrated book in that it includes separate sections on NMR and IR, in addition to the one on MS.

1. P. Crews, M. Jaspars, J. Rodriquez, *Organic Structure Analysis*, 1st edition, Oxford University Press, Oxford, 1998.
2. J.B. Lambert, H.F. Shurvell, D.A. Lightner, R.G. Cooks, *Organic Structural Spectroscopy*, Prentice Hall, Upper Saddle River, NJ, 1998.
3. R.M. Silverstein, F.X. Webster, *Spectrometric Identification of Organic Compounds*, 6th edition, Wiley, New York, 1998 (1st edition, 1963; 2nd edition, 1967, eds. R.M. Silverstein, G.C. Bassler; 3rd edition, 1974; 4th edition, 1981; 5th edition, 1991, eds. R.M. Silverstein, G.C. Bassler, T.C. Morrill).
4. L.M. Harwood, T.D.W. Claridge, *Introduction to Organic Spectroscopy*, 1st edition, Oxford University Press, New York, 1997.
5. M. Hesse, H. Meier, B. Zeeh, (A. Linden, M. Murray, translators), *Spectroscopic Methods in Organic Chemistry*, 1st edition, Georg Thieme, New York, 1997.
6. D.L. Pavia, G.M. Lampman, G.S. Kriz, *Introduction to Spectroscopy: a Guide for Students of Organic Chemistry*, 2nd edition, Saunders College, Orlando, FL, 1996 (1st edition, 1979).
7. L.D. Field, S. Sternhell, J.R. Kalman, *Organic Structures from Spectra*, 2nd edition, Wiley, Chichester, 1995 (1st edition, 1986).
8. K. Feinstein, *Guide to Spectroscopic Identification of Organic Compounds*, CRC Press, Boca Raton, FL, 1995.
9. D.H. Williams, I. Fleming, *Spectroscopic Methods in Organic Chemistry*, 5th edition, McGraw-Hill, London, 1995 (1st edition, 1966).
10. C. Jones, B. Mulloy, A.H. Thomas (eds.), *Spectroscopic Methods and Analyses (NMR, Mass Spectrometry, and Metalloprotein Techniques), Methods in Molecular Biology*, Humana Press, Totowa, NJ, Vol. 17, 1993.
11. W. Kemp, *Organic Spectroscopy*, 3rd edition, W.H. Freeman, New York, 1991.
12. W. Fresenius, J.F.K. Huber, E. Pungor, G.A. Rechnitz, W. Simon, T.S. West (eds.), *Tables of Spectral Data for Structure Determination of Organic Compounds*, 2nd English edition, translated from the German edition by K. Biemann, Springer, Berlin, 1989.
13. T.N. Sorrell, *Interpreting Spectra of Organic Molecules*, University Science, Mill Valley, CA, 1988.
14. F. Scheinmann (ed.), *An Introduction to Spectroscopic Methods for the Identification of Organic Compounds, Mass Spectrometry, Ultraviolet Spectroscopy, Electron Spin Resonance Spectroscopy, NMR (Recent Developments), Use of Various Spectral Methods Together, and Documentation of Molecular Spectra*, Pergamon Press, Oxford, Vol. 2, 1974.
15. D.W. Mathieson (ed.), *Interpretation of Organic Spectra*, Academic Press, New York, 1965.

Software

There are several programs that are available as self-training. Those programs developed by the UK company

Cognitive Solutions (Software 10–13) are somewhat like English roast beef. They are intellectually nutritious; however, they fail to excite the experiential palate. Equivalent titles available from SAVANT (Software 14–22) will hold the user's interest in a much more conductive manner for learning. Just as the *MS Fundamentals* program is a seminal tool in the development of the understanding of the TQ mass spectrometer, *Fundamentals of GC/MS* is one of the better instrument-user software packages developed. This program was initially developed and sent out several times for review to a number of people who are involved in training on various aspects of GC/MS and in the development of training programs. The result is what is assured to become an award-winning effort.

All of the training programs from SAVANT and Cognitive Solutions were developed in Interactive ToolBook, a powerful tool for the development of training programs. All have tests built into the programs to allow the user to evaluate the results of the training.

The two volumes of *SpectraBook* are also based on the Interactive ToolBook platform. These two programs each contain data on 50 separate compounds: mass, proton NMR, ^{13}C-NMR and IR spectra as well as the structure, molecular weight based on the atomic weights of each compound's elements, physical properties, and several synonyms. Help files are provided to assist the user in developing desired interpretational skills. Another nice feature is the ability to display what properties result in specific spectral peaks. Placing the mouse pointed on a labeled mass spectral peak and holding down the left mouse button will result in a display of the mechanism(s) that produced the ion represented by that peak. Similar displays are provided for the other types of spectra.

It is unfortunate that the author of *SpectraBook* (programs copyrighted in 1990 and 1992) did not take more care to be correct in some of the presentations such as the use of *m*/*z* as the symbol for mass-to-charge (inappropriately written Mass to Charge on the abscissa of mass spectra) ratio instead of the *m*/*e* symbol, which was replaced in the 1970s. The indicated shift of pairs of electrons in the displayed mechanism for beta cleavage resulting from a gamma-hydrogen shift does not instill confidence in the accuracy of instruction.

The programs published by ChemSW (Software 1–9) are very well thought out and provide utilities that are not found in the data-system software for most, if not all, commercially available instruments. The titles of these programs are self-explanatory.

The two programs associated with the Wiley and NIST Mass Spectral Databases (Software 31 and 33) are widely available from a number of different sources. A number of GC/MS and LC/MS programs now provide the NIST Mass Spectral Search Program as the search routine used with their proprietary instrument software. Both of the programs are capable of reading most, if not all, commercially available instrument data formats.

1. †*CESAR™ Capillary Electrophoresis Simulation for Application Research*, ChemSW, Fairfield, CA.
2. *GC-SOS™ Gas Chromatography Simulation and Operation Software, Ver. 5*, ChemSW, Fairfield, CA.
3. J.T. Bernert, Jr, (Quadtech Associates), *Mass Spec Calculator™, Ver. 3*, and *Mass Spec Calculator™ Pro*, ChemSW, Fairfield, CA.
4. T. Junk, *GC and GC/MS File Translator™ Professional*, ChemSW, Fairfield, CA.
5. T. Junk, *GC and GC/MS File Manager™*, ChemSW, Fairfield, CA.
6. J.T. Bernert, Jr, (Quadtech Associates), *Mass Differential Analysis Tools*, ChemSW, Fairfield, CA.
7. *Protein Tools™*, ChemSW, Fairfield, CA.
8. *HPLC Optimization™*, ChemSW, Fairfield, CA.
9. *GPMAW™ General Protein Mass Analysis for Windows*, ChemSW, Fairfield, CA.
10. *Interactive Training Program™ *Gas Chromatography*, Cognitive Solutions, Glasgow, a.k.a. Softbooks out of the United States, distributed by ChemSW, Fairfield, CA.
11. ‡Interactive Training Program™ *High Performance Liquid Chromatography*, Cognitive Solutions, Glasgow, a.k.a. Softbooks out of the United States, distributed by ChemSW, Fairfield, CA.
12. ‡Interactive Training Program™ *Advanced Gas Chromatography*, Cognitive Solutions, Glasgow, a.k.a. Softbooks out of the United States, distributed by ChemSW, Fairfield, CA.
13. ‡S. Davis, (HD Technologies), Interactive Training Program™ *Mass Spectrometry*, Cognitive Solution, Glasgow, a.k.a. Softbooks out of the United States, distributed by ChemSW, Fairfield, CA.
14. ‡D. Saunders, *Introduction to Gas Chromatography*, SAVANT†, Fullerton, CA.
15. ‡D. Saunders, *Fundamentals of Gas Chromatography/Mass Spectrometry*, SAVANT, Fullerton, CA.
16. ‡*Introduction to High Performance Liquid Chromatography*, SAVANT, Fullerton, CA.
17. ‡*Method Development in High Performance Liquid Chromatography*, SAVANT, Fullerton, CA.
18. ‡*High Performance Liquid Chromatography Equipment*, SAVANT, Fullerton, CA.
19. ‡*Troubleshooting High Performance Liquid Chromatography*, SAVANT, Fullerton, CA.
20. ‡*Separation Modes of High Performance Liquid Chromatography*, SAVANT, Fullerton, CA.
21. ‡*HPLC Calculation Assistant & Reference Tables*, SAVANT, Fullerton, CA.
22. ‡*Identification & Quantification for HPLC*, SAVANT, Fullerton, CA.

† Sloane Audio Visuals for Analysis and Training.

‡ Uses Asymetrix ToolBook Runtime.

23. ‡M. Hart, *MS Fundamentals*, Hewlett-Packard, Palo Alto, CA; distributed by SAVANT, Fullerton, CA.
24. ‡P.F. Schatz, *SpectraBook*, Vol. 1, 1990 and Vol. 2, 1992, Falcon Software, distributed by SAVANT, Fullerton, CA.
25. *DryLab*, LC Resources, Lafayette, CA.
26. *Introduction to CE*, LC Resources, Lafayette, CA.
27. J. Figueras, *Mass Spec*, Ver. 3.0, Trinity Software, Plymouth, NH.
28. *ACD/SpecManager: MS Module*, Advanced Chemistry Development, Toronto.
29. *Mass Frontier*, ThemoQuest/HighChem, San Jose, CA.
30. *MASSTransit*, Palisade, Newfield, NY.
31. *Benchtop PBM* with *Wiley Registry of Mass Spectral Data, Ver. 6 or 6N, or Select*, Palisade, Newfield, NY.
32. D. Dahl, *SIMION 3D*, Ver. 6.0, Ion and Electron Optics Program, Scientific Instrument Services, Ringoes, NJ.
33. *NIST Mass Spectral Search Program for Windows*, Ver. 1.7, and *NIST/EPA/NIH Mass Spectral Database* with *AMDIS, Automated Mass Spectral Deconvolution and Identification System, Ver. 2.0*, National Standard Reference Data System, National Institute of Standards and Technology. Gaithersburg, MD.
34. F. Antolasic, *Wsearch*, Freeware program for data analysis of a variety of different GC/MS data file formats, Mass Spectrometry Facility, Department of Applied Chemistry, RMIT University, Melbourne (http://minyos.its.rmit.edu.au/~rcmfa/).

Web Sites for Software Companies

ChemSW	http://www.chemsw.com
LC Resources	http://www.lcresources.com
NIST	http://www.nist.gov/srd/analy.htm
Palisade	http://www.palisade.com
Scientific Instrument Services	http://www.sisweb.com
Trinity Software	http://www.trinitysoftware.com
Advanced Chemistry Development	http://www.acdlabs.com
HighChem	http://www.highchem.com

Table 1 Element/isotopic mass and abundance data

Type	Element	Symbol	Integer mass[a]	Exact mass[b]	Percent abundance	$X+1$ factor[c]	$X+2$ factor[d]
X	Hydrogen	H	1	1.0078	99.99		
		D or ^{2}H	2	2.0141	0.01		
$X+1$	Carbon	^{12}C	12	12.0000	98.91		
		^{13}C	13	13.0034	1.1	$1.1n_{C}$	$0.0060n_{C}^{2}$
$X+1$	Nitrogen	^{14}N	14	14.0031	99.6		
		^{15}N	15	15.0001	0.4	$0.37n_{N}$	
$X+2$	Oxygen	^{16}O	16	15.9949	99.76		
		^{17}O	17	16.9991	0.04	$0.04n_{O}$	
		^{18}O	18	17.9992	0.20		$0.20n_{O}$
X	Fluorine	F	19	18.9984	100		
$X+2$	Silicon	^{28}Si	28	27.9769	92.2		
		^{29}Si	29	28.9765	4.7	$5.1n_{Si}$	
		^{30}Si	30	29.9738	3.1		$3.4n_{Si}$
X	Phosphorus	P	31	30.9738	100		
$X+2$	Sulfur	^{32}S	32	31.9721	95.02		
		^{33}S	33	32.9715	0.76	$0.8n_{S}$	
		^{34}S	34	33.9679	4.22		$4.4n_{S}$
$X+2$	Chlorine	^{35}Cl	35	34.9689	75.77		
		^{37}Cl	37	36.9659	24.23		$32.5n_{Cl}$
$X+2$	Bromine	^{79}Br	79	78.9183	50.5		
		^{81}Br	81	80.9163	49.5		$98.0n_{Br}$
X	Iodine	I	127	126.9045	100		

[a] The integer mass of the most abundant[e] naturally occurring stable isotope of an element is the element's *nominal mass*. The nominal mass of an ion is the sum of the nominal masses of the elements in its empirical formula (e.g. $C_3H_6O^{+\bullet}$ has a nominal mass of 58).
[b] The exact mass of the most abundant[e] naturally occurring stable isotope of an element is the element's *monoisotopic mass*. The monoisotopic mass of an ion is the sum of the monoisotopic masses of the elements in its empirical formula (e.g. $C_3H_6O^{+\bullet}$ has a monoisotopic mass of 58.0417).
[c] Assume $X = 100\%$; X represents the relative intensity of the first peak in a cluster of peaks corresponding to isotopic variants of a given ion.
[d] The factor is multiplied by the number (n) of atoms of the element present to determine the magnitude of the intensity contribution for a given isotope. For example, the contribution at $X+1$ due to ^{15}N for an ion containing three nitrogens would be $0.37 \times 3 = 1.11$ relative to 100 at X.
[e] This may not always be the lowest mass naturally occurring stable isotope of the element, as is the case with the elements in this table. The lowest mass isotope of Hg is 196 and the nominal mass isotope is 202, seventh from the lowest mass isotope.

APPENDIX 3: MASS SPECTRAL INTERPRETATION DATA

EI fragmentation data are given in Tables 1 and 2, Figures 30 and 31, and Tables 3 and 4.

Proof of a Molecular Ion Peak – $M^{+\bullet}$

1. If a compound is known, the molecular ion has a mass-to-charge ratio (m/z) value equal to the sum of the atomic masses of the most abundant isotope of

Table 2 Common fragment ion

29	HCO^+	Aldehydes
30	$CH_2NH_2^+$	Amines
31	$H_2C{=}\overset{+}{O}H$	Alcohols
	CH_3O^+	Methyl esters
43	$C_3H_7^+$	Propyl ion
	CH_3CO^+	Acetyl groups
47	CH_3S^+	Sulfides
49	CH_2Cl^+	Chloro compounds
55	$C_4H_7^+$	Alkyl groups
57	$C_4H_9^+$	Alkyl groups
	$C_2H_5CO^+$	Acylium ion
58	$(H_2C\overset{H}{\diagup}COCH_3)^{+\bullet}$	Ketones with a γ-hydrogen
59	$(COOCH_3)^+$	Methyl esters
61	$CH_3\overset{OH}{C}{=}O^+H$	Esters of high molecular weight alcohols
70		Pyrrolidines
74	$(CH_2{=}\overset{OH}{C}{-}OCH_3)^{+\bullet}$	Methyl esters with a γ-hydrogen
77	$C_6H_5^+$	Aromatics
78	$C_5H_4N^+$	Pyridines and alkyl pyrroles
80		Pyrroles
80 (82)	$HBr^{+\bullet}$	Bromo compounds
81		Furans
	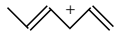	Aliphatic chain with two double bonds
83	$C_6H_{11}^+$	Cyclohexanes or hexenes
	$CHCl_2^+$	Chloro compounds
85		Tetrahydropyranyl ethers
88	$CH_3OCO{=}CH$ $\overset{+}{N}H_2$	Amino acid esters
91	$C_7H_7^+$	Aromatic hydrocarbons with side chains
92	$C_7H_8^+$	Benzyl compounds with a γ-hydrogen
95	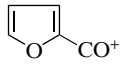	Furyl−CO−X
97		Alkyl thiophenes
99	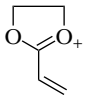	Ethylene ketals of cyclic compounds (steroids)
104	$C_8H_8^{+\bullet}$	Alkyl aromatics
105	$C_6H_5CO^+$	Benzoyl compounds
	$C_8H_9^+$	Aromatic hydrocarbons
106	$CH_2{=}\overset{+}{N}H_2$	Amino benzyl
107	$C_7H_7O^+$	Phenolic hydrocarbons
117	$C_9H_9^+$	Styrenes
128	HI^+	Iodo compounds
130	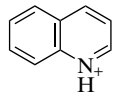	Indoles
131	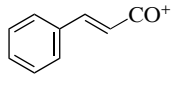	Cinnamates
149	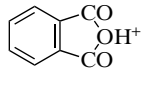	Dialkyl phthalates (rearrangement)

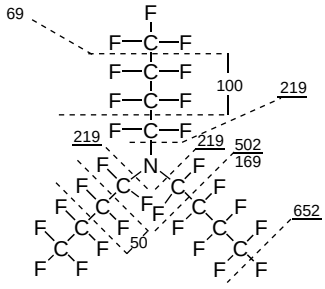

Figure 30 Fragments produced by the EI of perfluorotributylamine (PFTBA) that are used for the calibration of the mass spectrometer's m/z scale. The molecular weight of PFTBA (FC-43), used to calibrate the m/z scale of TQ and QIT mass spectrometers operated in the EI mode, is 671. The following is an explanation of the origin of some of the peaks observed in its EI mass spectrum:

$$\begin{array}{lll} 671 & 671 & 671 \\ -207 & -507 & -57 \\ \hline 464 & 164 & 614 \\ & & 169 \\ -50 & +100 & -38 \\ \hline 414 & 264 & 131 \end{array} \quad \begin{array}{l} (3\times 69) \\ \\ (2\times 19) \end{array} \quad \begin{array}{l} (3\times 169) \\ \\ \\ \end{array} \quad \begin{array}{l} (3\times 19) \\ \\ (2\times 19) \end{array}$$

Table 3 Chlorine and bromine isotopic abundance ratios

Atoms of ClBr	X	$X+2$	$X+4$	$X+6$	$X+8$	$X+10$
Cl	100	32.5				
Cl_2	100	65.0	10.6			
Cl_3	100	97.5	31.7	3.4		
Cl_4	76.9	100	48.7	0.5	0.9	
Cl_5	61.5	100	65.0	21.1	3.4	0.2
Cl_6	51.2	100	81.2	35.2	8.5	1.1
ClBr	76.6	100	24.4			
Cl_2Br	61.4	100	45.6	6.6		
Cl_3Br	51.2	100	65.0	17.6	1.7	
$ClBr_2$	43.8	100	69.9	13.7		
Cl_2Br_2	38.3	100	89.7	31.9	3.9	
Cl_3Br_2	31.3	92.0	100	49.9	11.6	1.0
$ClBr_3$	26.1	85.1	100	48.9	8.0	
Cl_2Br_3	20.4	73.3	100	63.8	18.7	2.0
Br	100	98.0				
Br_2	51.0	100	49.0			
Br_3	34.0	100	98.0	32.0		
Br_4	17.4	68.0	100	65.3	16.0	

each element that comprises the molecule (assuming the ion is a single-charge ion).

2. The nominal molecular weight of a compound, or the m/z value for the molecular ion, is an even number for any compound containing only C, H, O, S, Si, P, and the halogens.

   Fragment ions derived via homolytic, heterolytic, or sigma-bond cleavage from these molecular ions (even m/z) have an odd m/z value and an even number of electrons.

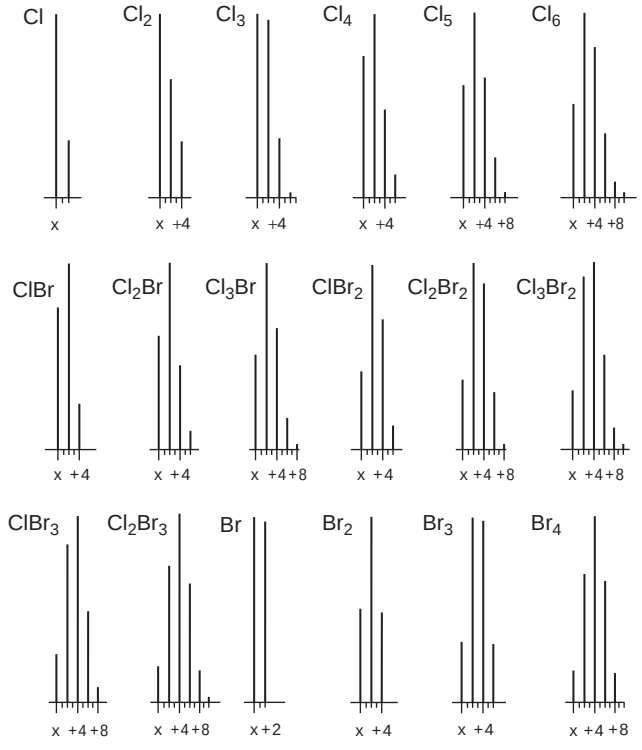

Figure 31 Graphical representation of relative isotope peak intensities for any given ion containing the indicated number of chlorine and/or bromine atoms. Numeric values are shown in Table 3. Steps to determine elemental composition based on isotope peak ratios:

1. Determine the nominal m/z value peak (peak at lowest m/z value, above which other peaks can be attributed to isotopic multiplicity or background).
2. Assign the $X+2$ elements, except oxygen.
3. Assign the $X+1$ elements. (Remember to normalize $X+1$ to X, if necessary.)
4. Balance the mass.
5. Assign the atoms of oxygen.
6. Balance the mass.
7. Assign the X elements.
8. From the empirical formula, determine the number of rings plus double bonds.
9. Propose a possible structure.
10. Does it make sense?

Fragment ions derived from these molecular ions (even m/z) via expulsion of neutral components (e.g. H_2O, CO, ethylene) have an even m/z value and an odd number of electrons.

3. *Nitrogen rule*: a compound containing an odd number of nitrogen atoms – in addition to C, H, O, S, Si, P, and the halogens – has an odd molecular weight.

   Molecular ions of these compounds fragment via homolytic, heterolytic, or sigma-bond cleavage to produce ions of an even m/z value unless the nitrogen atom is lost with the neutral radical.

Table 4 Common neutral losses

M – 1	Loss of hydrogen radical	M – $^{\bullet}H$
M – 15	Loss of methyl radical	M – $^{\bullet}CH_3$
M – 29	Loss of ethyl radical	M – $^{\bullet}CH_2CH_3$
M – 31	Loss of methoxyl radical	M – $^{\bullet}OCH_3$
M – 43	Loss of propyl	M – $^{\bullet}CH_2CH_2CH_3$
M – 45	Loss of ethoxyl	M – $^{\bullet}OCH_2CH_3$
M – 57	Loss of butyl radical	M – $^{\bullet}CH_2CH_2CH_2CH_3$
M – 2	Loss of hydrogen	M – H_2
M – 18	Loss of water	M – H_2O
M – 28	Loss of CO or ethylene	M – CO or M – C_2H_4
M – 32	Loss of methanol	M – CH_3OH
M – 44	Loss of CO_2	M – CO_2
M – 60	Loss of acetic acid	M – CH_3CO_2H
M – 90	Loss of silanol: $HOSi(CH_3)_3$	M – $HOSi(CH_3)_3$

An even number of nitrogen atoms in a compound results in an even nominal molecular weight.

4. The molecular ion peak must be the highest m/z value of any significant (nonisotope or nonbackground) peak in the spectrum. Corollary: the highest m/z value peak observed in the mass spectrum need not represent a molecular ion.
5. The peak at the next lowest m/z value in the mass spectrum must not correspond to the loss of an impossible or improbable combination of atoms.
6. No fragment ion may contain a larger number of atoms of any particular element than the molecular ion.

Courses on the interpretation of mass spectra and techniques of MS are offered by the Continuing Education Department of the ACS: http://www.acs.org/education/profdevl/short.html.

REFERENCES

1. F.W. Aston, *Isotopes*, 2nd edition, Edward Arnold, London, 1924 (1st edition, 1923).
2. A.L. Yerger, A.K. Yergy, 'Preparative Scale Mass Spectrometry: a Brief History of the Calutron', *J. Am. Soc. Mass Spectrom.*, **8**, 943–953 (1997).
3. F.W. McLafferty, in *MS Fundamentals*, *a Multimedia Training Course*, Savant, Fullerton, CA, 1995.
4. S.J. Gaskell, 'Electrospray: Principles and Practice, Special Feature: Tutorial', *J. Mass Spectrom.*, **32**, 677–688 (1997).
5. S. Schachterle, R.D. Brittain, J.D. Mills, 'Analysis of Pesticide Residues in Food Using Gas Chromatography–Tandem Mass Spectrometry with a Benchtop Ion Trap Mass Spectrometer', *J. Chromatogr. A*, **683**, 185–193 (1994).
6. R.W. Kiser, *Introduction to Mass Spectrometry and its Application*, Prentice-Hall, Englewood Cliffs, NJ, 1965.
7. J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, Philadelphia, 1997.
8. E. De Hoffmann, J. Charette, V. Stroobant, *Mass Spectrometry: Principles and Applications*, Masson, Paris, 1996; original French language edition: *Spectrométrie de Masse*, Masson, Paris, 1996; Wiley, New York, 1996.
9. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', in the Fundamental Reviews issue of *Anal. Chem.*, **70**(16), 1998 (this review began in 1949; since sometime before 1956, it has appeared every 2 years on the even year).
10. J. Roboz, *Introduction to Mass Spectrometry Instrumentation and Techniques*, Wiley, New York, 1968.
11. P.H. Dawson (ed.), *Quadrupole Mass Spectrometry and its Applications*, American Institute of Physics, Woodbury, NY, 1995; originally published by Elsevier, Amsterdam, 1976.
12. R.E. March, R.J. Hughes, *Quadrupole Storage Mass Spectrometry*, Wiley, New York, 1989.
13. R.J. Cotter, *Time-of-flight Mass Spectrometry: Instrumentation and Applications in Biological Research*, American Chemical Society, Washington, DC, 1997.
14. A.G. Marshall, C.L. Hendrickson, G.S. Jackson, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer', *Mass Spectrom. Rev.*, **17**, 1–36 (1998).
15. The following are recommended sources on the interpretation of EI mass spectra: (a) F.W. McLafferty, F. Turecek, *Interpretation of Mass Spectra*, 4th edition, University Science, Mill Valley, CA, 1993; (b) H. Budzikiewicz, C. Djerassi, D.H. Williams, *Mass Spectrometry of Organic Compounds*, Holden-Day, San Francisco, 1967; (c) J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, Philadelphia, 1997; (d) S.R. Shrader, *Introductory Mass Spectrometry*, Allyn and Bacon, Boston, MA, 1971; (e) F.W. McLafferty, *Interpretation of Mass Spectra*, 2nd edition, Benjamin, Reading, MA, 1973.
16. American Chemical Society, Department of Continuing Education, Washington, DC, http://www.acs.org/education/professional/short.html.
17. American Society for Mass Spectrometry, Santa Fe, NM, http://www.asms.org.
18. A.G. Harrison, *Chemical Ionization Mass Spectrometry*, CRC, Boca Raton, FL, 1992.
19. The following are recommended sources for derivative formation: (a) K. Blau, J. Halket, *Handbook of Derivatives for Chromatography*, Wiley, New York, 1993; (b) D.R. Knapp, *Handbook of Analytical Derivatization Reactions*, Wiley-Interscience, New York, 1979; (c) A.E. Pierce, *Silylation of Organic Compounds*, Pierce Chemical, Rockford, IL, 1968.
20. The following are references for the process of electrospray: (a) R.B. Cole (ed.), *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation*, and

Applications, Wiley, New York, 1997; (b) C.G. Enke, 'A Predictive Model for Matrix and Analyte Effects in Electrospray Ionization of Singly-charged Ionic Analytes', *Anal. Chem.*, **69**, 4885–4893 (1997); (c) J.B. Fenn, J. Rosell, C.K. Meng, 'In Electrospray Ionization, How Much Pull Does an Ion Need to Escape its Droplet', *J. Am. Soc. Mass Spectrom.*, **8**, 1147–1157 (1997).

21. R. Zenobi, R. Knochenmuss, 'Ion Formation in MALDI Mass Spectrometry', *Mass Spectrom. Rev.*, **17**, 337–366 (1998).
22. F.G. Kitson, B.S. Larsen, C.N. McEwen, *Gas Chromatography and Mass Spectrometry: A Practical Guide*, Academic Press, San Diego, CA, 1996.
23. G. Siuzdak, 'A Tandem History of Mass Spectrometry', ⟨http://masspec.scripps.edu/hist.html⟩.
24. E. Goldstein, *Berl. Ber.*, **39**, 69 (1886) (title not available).
25. W.C. Wien, *Ann. Phys., Leipzig*, **8**, 224 (1902) (title not available).
26. W.C. Wien, *Verh. Phys. Ges.*, **17**, 1898 (1898) (title not available).
27. J.J. Thomson, 'Cathode Rays', in *Physical Science*, eds. W.L. Bragg, G. Porter, Applied Science, London, 1903.
28. J.J. Thomson, *Electrician*, **39**, 104 (1897) (title not available).
29. J.J. Thomson, *Philos. Mag.*, **44**, 293 (1897) (title not available).
30. J.J. Thomson, *Rays of Positive Electricity and their Application to Chemical Analysis*, 2nd edition, Longmans Green, London, 1921 (1st edition, 1913). Thomson's life and research have been chronicled in three different biographies: (a) E.A. Davis, I.J. Falconer, *J.J. Thomson and the Discovery of the Electron*, Taylor & Francis, London, 1997; (b) P.F. Dahl, *Flash of the Cathode Rays – A History of J.J. Thomson's Electron*, Institute of Physics, Philadelphia, PA, 1997; (c) G.P. Thomson, *J.J. Thomson and the Cavendish Laboratory in His Day*, Doubleday, New York, 1965.
31. F.W. Aston, 'The Constitution of Atmospheric Neon', *Philos. Mag., 6th Ser.*, **39**, 449 (1920).
32. W. Bartky, A.J. Dempster, 'Paths of Charged Particles in Electric and Magnetic Fields', *Phys. Rev.*, **33**, 1019 (1929).
33. F.W. Aston, *Mass Spectrometry and Isotopes*, 2nd edition, Edward Arnold, London, 1942 (1st edition, 1933).
34. A.T. James, A.J.P. Martin, 'Gas–Liquid Partition Chromatography: The Micro Separation and Micro-estimation of Volatile Acids from Formic Acid to Dodecanoic Acid', *Biochem. J.*, **50**, 670 (1952). Note: there were several other papers published by James and Martin at about this same time for which titles are not available: (a) A.T. James, A.J.P. Martin, *Biochem. J. Proc.*, **48**, vii (1951); (b) A.T. James, A.J.P. Martin, *Analyst*, **77**, 915 (1952); (c) A.T. James, A.J.P. Martin, G.H. Smith, *Biochem. J.*, **52**, 238 (1952); (d) A.T. James, *Biochem. J.*, **52**, 242 (1952).
35. R.S. Gohlke, 'Time-of-flight Mass Spectrometry and Gas–Liquid Partition Chromatography', *Anal. Chem.*, **31**, 535–541 (1959).
36. R.S. Gohlke, F.W. McLafferty, 'Early Gas Chromatography/Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 367–371 (1993).
37. R.E. Finnigan, 'Quadrupole Mass Spectrometers: from Development to Commercialization', *Anal. Chem.*, **66**, 969A–975A (1994).
38. Two important papers appeared that dealt with the analysis of GC/MS data: (a) R.A. Hites, K. Biemann, 'A Computer-compatible Digital Data Acquisition System for Fast-scanning, Single-focusing Mass Spectrometers', *Anal. Chem.*, **39**, 965–970 (1967); (b) R.A. Hites, K. Biemann, 'Mass Spectrometry–Computer System Particularly Suited for Gas Chromatography of Complex Mixtures', *Anal. Chem.*, **40**, 1736–1739 (1968).
39. R.A. Yost, C.G. Enke, 'Triple Quadrupole Mass Spectrometry for Direct Mixture Analysis and Structure Elucidation', *Anal. Chem.*, **51**, 1251A–1264A (1979).
40. D.F. Hunt, G.C. Stafford, Jr, F.W. Crow, J.W. Russell, 'Pulsed Positive Negative CI', *Anal. Chem.*, **48**, 2098–2105 (1976).
41. L.S. Ettre, 'Pittsburgh Conference – A Personal Appraisal', *LC-GC*, **17**, 156–168 (1999).
42. R. Schoenheimer, D. Rittenberg, *J. Biol. Chem.*, **127**, 285 (1939) (title not available).
43. R.M. Caprioli, 'Use of Stable Isotopes', in *Biochemical Applications of Mass Spectrometry*, ed. G.R. Waller, Wiley-Interscience, New York, 1972.
44. G.R. Waller, O.C. Dermer (eds.), *Biochemical Applications of Mass Spectrometry, First Supplement*, Wiley-Interscience, New York, 1980.
45. G.P. Barnard, *Modern Mass Spectrometry*, Institute of Physics, London, 1953.
46. J.H. Beynon, *Mass Spectrometry and its Applications to Organic Chemistry*, Elsevier, Amsterdam, 1960; reprinted by American Society for Mass Spectrometry, Santa Fe, NM, 1999.
47. K. Biemann, *Mass Spectrometry: Organic Chemical Applications*, McGraw-Hill, New York, 1962; reprinted by American Society for Mass Spectrometry, Santa Fe, NM, 1998.
48. J. Mattauch, R.Z. Herzog, *Phys.*, **89**, 786 (1934) (title not available).
49. J. Mattauch, *J. Phys. Rev.*, **50**, 617, 1089 (1936) (title not available).
50. (a) S.J. Pachuta, 'Alfred O. Nier: an Appreciation', *J. Mass Spectrom.*, **30**, 401, 402 (1995); (b) A.O. Neir, 'The Development of a High Resolution Mass Spectrometer: a Reminiscence', *J. Am. Soc. Mass Spectrom.*, **2**, 447–452 (1991).
51. C. Judson, M. Grayson, 'A Workhorse Comes to CHF: Exxon Donates a CEC 21-103C Mass Spectrometer', *Chem. Heritage*, **16**(2), 2–21 (1998).

52. F.H. Field, 'Letter to the Editor', *Mass Spectrom. Rev.*, **13**, 99–101 (1994).
53. S. Meyerson, 'From Black Magic to Chemistry: the Metamorphoses of Organic Mass Spectrometry', *Anal. Chem.*, **66**, 960A–964A (1994).
54. J.A. Hipple, M. Shephard, 'Mass Spectrometry', *Anal. Chem.*, **21**, 32–36 (1949).
55. Now the National Institute of Standards and Technology (NIST), Gaithersburg, MD.
56. J.A. Hipple, L.T. Aldrich, A.O. Nier, V.H. Dibeler, F.L. Mohler, R.E. O'Dette, H. Odishaw, H. Sommer, Mass Spectroscopy Committee, Mass Spectrometry in Physics Research, National Bureau of Standards Circular 522, United States Government Printing Office, Washington, DC, 1953.
57. J. Blears, Chairman, Mass Spectrometry Panel, Institute of Petroleum, Applied Mass Spectrometry, a report of a conference organized by the Mass Spectrometry Panel of the Institute of Petroleum, held in London, 29–31 October 1953, Institute of Petroleum, London, 1954.
58. The following are the original references to organic mechanisms in fragmentation of EI mass spectra: (a) H. Budzikiewicz, C. Djerassi, D.H. Williams, *Mass Spectrometry of Organic Compounds*, Holden-Day, San Francisco, 1967; (b) G. Spiteller, *Massenspektrometrische Strukturanalyse Organischer Verbindungen*, Verlag Chemie, Weinheim, 1966; (c) A. Quayle, R.I. Reed, 'Interpretation of Mass Spectra', in *Interpretation of Organic Spectra*, ed. D.W. Mathieson, Academic Press, New York, 1965; (d) H. Budzikiewicz, C. Djerassi, D.H. Williams, *Structure Elucidation of Natural Products by Mass Spectrometry: Vol. I, Alkaloids; Vol. II, Steroids, Terpenoids, Sugars, and Miscellaneous Natural Products*, Holden-Day, San Francisco, 1964; (e) H. Budzikiewicz, C. Djerassi, D.H. Williams, *Interpretation of Mass Spectra of Organic Compounds*, Holden-Day, San Francisco, 1964; (f) F.W. McLafferty (ed.), *Mass Spectra of Organic Ions*, Academic Press, New York, 1963.
59. G.R. Waller (ed.), *Biochemical Applications of Mass Spectrometry*, Wiley-Interscience, New York, 1972.
60. R. Ryhage, 'MS as a Detector for GC', *Anal. Chem.*, **36**, 759–764 (1964).
61. J.T. Watson, K. Biemann, 'High-resolution MS with GC', *Anal. Chem.*, **36**, 1135–1137 (1964); J.T. Watson, K. Biemann, 'High-resolution MS of GC Effluents', *Anal. Chem.*, **37**, 844–851 (1965).
62. P.M. Llewellyn, D.P. Littlejohn, Presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, 1966.
63. W.E. Reynolds, V.A. Bacon, J.C. Bridges, T.C. Coburn, B. Halpern, J. Lederberg, E.C. Levinthal, E. Steed, R.B. Tucker, 'A Computer Operated Mass Spectrometry System', *Anal. Chem.*, **42**, 1122–1129 (1970).
64. C.C. Sweeley, B.D. Ray, W.I. Wood, J.F. Holland, 'On-line Digital Computer System for High-speed Single Focusing Mass Spectrometry', *Anal. Chem.*, **42**, 1505–1516 (1970).
65. R.A. Hites, K. Biemann, 'Identification of Mass Spectra by Computer Searching a File of Known Spectra', *Anal. Chem.*, **43**, 681–691 (1971).
66. R.M. Smith, K.L. Busch (tech. eds.), *Understanding Mass Spectra: A Basic Approach*, Wiley-Interscience, New York, 1999.
67. J. Fjeldsted, J.-L. Truche, 'Evolution of the MS as a Tool for Chromatography', *Am. Lab.*, October, 33–39 (1989).
68. (a) B. Munson, F.H. Field, 'Chemical Ionization Mass Spectrometry', *J. Am. Chem. Soc.*, **88**, 2621–2630 (1966); (b) B. Munson, 'CI-MS: 10 Years Later', *Anal. Chem.*, **49**, 772A–778A (1977).
69. R.W. Kondrat, R.G. Cooks, 'Direct Analysis of Mixtures by Mass Spectrometry', *Anal. Chem.*, **50**, 81A–92A (1978).
70. R.E. March, 'An Introduction to Quadrupole Ion Trap Mass Spectrometry', *J. Mass Spectrom.*, **32**, 351–369 (1997).
71. P.J. Arpino, M.A. Baldwin, F.W. McLafferty, 'Liquid Chromatography/Mass Chromatography. Continuous Monitoring', *Biomed. Mass Spectrom.*, **1**, 80–82 (1974).
72. W.H. McFadden, H.L. Schwartz, S. Evans, 'Direct Analysis of Liquid Chromatographic Effluents', *J. Chromatogr.*, **122**, 389–396 (1976).
73. This work was pioneered by Evan Horning's group at Baylor University, Houston, Texas: (a) E.C. Horning, D.I. Carroll, I. Dzidic, K.D. Haegele, M.G. Horning, R.N. Stillwell, 'Atmospheric Pressure Ionization (API) Mass Spectrometry; Solvent-mediated Ionization of Samples Introduced in Solution and a Liquid Chromatograph Effluent Stream', *J. Chromatogr. Sci.*, **12**, 725–729 (1974); (b) E.C. Horning, D.I. Carroll, I. Dzidic, K.D. Haegele, M.G. Horning, R.N. Stillwell, 'Liquid Chromatography–Mass Spectrometry Computer Analytical Systems; Continuous-flow Systems Based on Atmospheric Pressure Ionization Mass Spectrometry', *J. Chromatogr.*, **99**, 13–21 (1974); (c) D.I. Carroll, I. Dzidic, K.D. Haegele, R.N. Stillwell, E.C. Horning, 'Atmospheric Pressure Ionization Mass Spectrometry: Corona-discharge Ion Source for Use in Liquid Chromatography–Mass Spectrometry Computer Analytical System', *Anal. Chem.*, **47**, 2369–2373 (1975); (d) E.C. Horning, D.I. Carroll, I. Dzidic, K.D. Haegele, S.N. Lin, C.U. Oertli, R.N. Stillwell, 'Development and Use of Bioanalytical Systems Based on Mass Spectrometry', *Clin. Chem.*, **23**, 13–21 (1977); (e) E.C. Horning, M.G. Horning, D.I. Carroll, I. Dzidic, R.N. Stillwell, 'New Picogram Detection System Based on a Mass Spectrometer with an External Ionization Source at Atmospheric Pressure', *Anal. Chem.*, **45**, 936–943 (1973); (f) D.I. Carroll, I. Dzidic, R.N. Stillwell, M.G. Horning, E.C. Horning, 'Subpicogram Detection System for Gas-phase Analysis Based upon Atmospheric Pressure

Ionization (API) Mass Spectrometry', *Anal. Chem.*, **46**, 706–710 (1974).
74. R.C. Willoughby, R.F. Browner, 'Monodisperse Aerosol Generating Interface for Combining Liquid Chromatography with Mass Spectrometry', *Anal. Chem.*, **56**, 2625 (1984).
75. Y. Ito, T. Takeuchi, D. Ishii, M. Goto, 'Direct Coupling of Micro High-performance Liquid Chromatography with Fast Atom Bombardment Mass Spectrometry', *J. Chromatogr.*, **346**, 161–166 (1985).
76. R.M. Caprioli, T. Fan, J.S. Cottrell, 'Continuous-flow Sample Probe for Fast Atom Bombardment Mass Spectrometry', *Anal. Chem.*, **58**(14), 2949–2953 (1986).
77. C.R. Blakely, M.L. Vestal, 'Thermospray Interface for Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **55**, 750–754 (1983).
78. C.M. Whitehouse, J.B. Fenn, M. Yamashita, 'An Electrospray Ion Source for Mass Spectrometry of Fragile Organic Species', Proceedings of the 32nd Annual Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, USA, May 27–June 1, 1984, American Society for Mass Spectrometry, Santa Fe, NM, 188, 189, 1984.
79. A.P. Bruins, T.R. Covey, J.D. Henion, 'Ionspray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 2642–2646 (1987).
80. C.K. Meng, M. Mann, J.B. Fenn, 'Of Protons or Proteins – A Beam's Beam', *Z. Phys.*, **D10**, 361–368 (1988).
81. Attributed to Professor Patrick R. Jones (Chairman Chemistry) in a post-sabbatical seminar presented at the University of the Pacific, Stockton, CA, in the Fall of 1990.
82. O.D. Sparkman, 'Mass Spectrometry PittCon'99', *J. Am. Soc. Mass Spectrom.*, **10**, 457–465 (1999).
83. (a) R. Gomer, M.G. Ingrahm, *J. Chem. Phys.*, **22**, 1279 (1954) (title not available); (b) R. Gomer, M.G. Ingrahm, *J. Am. Chem. Soc.*, **77**, 500 (1955) (title not available).
84. H.D. Beckey, *Principles of Field Ionization and Field Desorption Mass Spectrometry*, Pergamon Press, New York, 1977.
85. H.D. Beckey, 'Field Ionization Mass Spectrometry', in *Advances in Mass Spectrometry*, ed. R.M. Elliott, Pergamon Press, Oxford, 1–24, Vol. 2, 1963.
86. D.F. Torgerson, R.P. Skowronski, R.D. Macfarlane, 'New Approach to the Mass Spectroscopy of Nonvolatile Compounds', *Biochem. Biophys. Res. Commun.*, **60**, 616–621 (1974).
87. R.D. Macfarlane, D.F. Torgerson, 'Californium-252 Plasma Desorption Mass Spectrometry', *Science*, **191**, 920–925 (1976).
88. C.J. McNeal, R.D. Macfarlane, E.L. Thurston, 'Thin Film Deposition by the Electrospray Method for Californium-252 Plasma Desorption Studies of Involatile Molecules', *Anal. Chem.*, **58**, 1303–1307 (1979).
89. P. Roepstorff, 'Plasma Desorption Mass Spectrometry: Principles and Applications to Protein Studies', in *Mass Spectrometry in the Biological Sciences: a Tutorial*, ed. M.L. Gross, Kluwer, Boston, MA, 1992.
90. M. Barber, J.C. Vickerman, *Surf. Defect Prop. Solids*, **5**, 162 (1976) (title not available).
91. M. Barber, R.S. Bordoli, R.D. Sedgwick, A.N. Tyler, 'Fast Atom Bombardment of Solids (FAB)', *J. Chem. Soc., Chem. Commun.*, 325–327 (1981).
92. M. Barber, R.S. Bordoli, G.V. Garner, D.B. Gordon, R.D. Sedgwick, A.N. Tyler, 'Fast-atom-bombardment Mass Spectra of Enkephalins', *Biochem. J.*, **197**(2), 401–404 (1981).
93. M. Barber, R.S. Bordoli, G.J. Elliott, R.D. Sedgwick, A.N. Tyler, 'Fast Atom Bombardment Mass Spectrometry', *Anal. Chem.*, **54**, 645A–657A (1982).
94. C. Fenselau, 'Mass Spectrometry for Characterization of Microorganisms: an Overview', in *Mass Spectrometry for the Characterization of Microorganisms*, ACS Symposium Series 541, ed. C. Fenselau, American Chemical Society, Washington, DC, 1994.
95. R.I. Reed, 'Some Problems in Mass Spectrometry', in *Mass Spectrometry*, ed. R.I. Reed, Academic Press, London, 401, 1965, and references to this article.
96. M.A. Baldwin, F.W. McLafferty, *Org. Mass Spectrom.*, **7**, 1353 (1973).
97. R.J. Cotter, 'Mass Spectrometry of Nonvolatile Compounds – Desorption from Extended Probes', *Anal. Chem.*, **52**, 1598A–1606A (1980).
98. R.J. Beuhler, E. Flanagan, L.J. Greene, L. Friedman, 'Proton Transfer MS of Peptides', *J. Am. Chem. Soc.*, **96**, 3990 (1974).
99. G. Devant, Personal Communication from Nermag SA, Rueil Malmaison, 1979.
100. Y.Y. Lin et al., *J. Am. Chem. Soc.*, **108**, 514 (1986).
101. (a) M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp, 'MALDI', *Int. J. Mass Spectrom. Ion Processes*, **78**, 53–68 (1987); (b) F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**, 1193A–2003A (1991).
102. C. Henry, 'FABMS: Still FABulous', *Anal. Chem.*, **69**, 625–627 (1997).
103. R.G. Cooks, J.H. Beynon, R.M. Caprioli, G.R. Lester, *Metastable Ions*, Elsevier, New York, 1973.
104. F.W. McLafferty (ed.), *Tandem Mass Spectrometry*, Wiley-Interscience, New York, 1983.
105. W.R. Smythe, J.H.E. Mattauch, 'A New Mass Spectrometer', *Phys. Rev.*, **40**, 429–433 (1932).
106. W.E. Stephens, 'A Pulsed Mass Spectrometer with Time Dispersion', *Phys. Rev.*, **69**, 691 (1946).
107. M.M. Wolff, W.E. Stephens, 'A Pulsed Mass Spectrometer with Time Dispersion', *Rev. Sci. Instrum.*, **24**, 616–617 (1953).

108. W.C. Wiley, I.H. McLaren, 'Time-of-flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, **26**, 1150–1157 (1955).
109. M. Karas, 'Special Feature: Historical: Comments on the Reprint of ''Time-of-flight Mass Spectrometer with Improved Resolution'' ', W.C. Wiley, I.H. McLaren, *Rev. Sci. Instrum.*, **26**, 1150 (1955), *J. Mass Spectrom.*, **32**, 1–3 (1997).
110. T. Hiroshi, *Kage Kagaku Dzassi*, **67**, 1769 (1964) (title not available).
111. T. Shahar, S. Dagan, A. Amirav, 'Laser Desorption Fast Gas Chromatography–Mass Spectrometry in Supersonic Molecular Beams', *J. Am. Soc. Mass Spectrom.*, **9**, 628–637 (1998).
112. W. Paul, H. Steinwedel, *Z. Naturforsch., Teil A*, **8**, 448 (1953) (title not available).
113. W. Paul, M. Raether, 'Das Elektrische Massenfilter', *Z. Phys.*, **140**, 262–271 (1955).
114. T.Z. Chu, R.E. Finnigan, 'Thirty Years of a Mass Spectrometry Company', *Am. Lab.*, October, 90s–109s (1998).
115. G.C. Stafford, Jr, P.E. Kelley, J.E.P. Syka, W.E. Reyonolds, J.F.J. Todd, 'Ion Trap Technology', *Int. J. Mass Spectrom. Ion Processes*, **60**, 85–98 (1984).
116. J.E.P. Syka, 'Commercialization of the Quadrupole Ion Trap', in *Particle Aspects of Ion Trap Mass Spectrometry: Fundamentals of Ion Trap Mass Spectrometry*, eds. R.E. March, J.F.J. Todd, CRC Press, Boca Raton, FL, Vol. 1, 1995.
117. O.D. Sparkman, 'New Generation Ion Trap GC/MS Technology Axial Modulation', GC/MS Varian Application Note, No. 8, 1990.
118. Development of the magnetic ion trap: (a) H. Dehmelt, 'Atomic Phosphorus Paramagnetic Resonance Experiment', *Phys. Rev.*, **99**, 527 (1955); (b) H. Dehmelt, 'Paramagnetic Resonance Reorientation of Atoms and Ions Aligned by Electron Impact', *Phys. Rev.*, **103**, 1125 (1956); (c) H. Dehmelt, 'Slow Spin Relaxation of Optically Polarized Sodium Atoms', *Phys. Rev.*, **105**, 1487 (1957); (d) H. Dehmelt, 'Modulation of a Light Beam by Processing Absorbing Atoms', *Phys. Rev.*, **105**, 1924 (1957).
119. First reports of FTICR MS: (a) M.B. Comisarow, A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **25**, 282, 283 (1974); (b) M.B. Comisarow, A.G. Marshall, 'Frequency-sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **26**, 489, 490 (1974); (c) M.B. Comisarow, A.G. Marshall, 'Selective-phase Ion Cyclotron Resonance Spectroscopy', *Can. J. Chem.*, **52**, 1997–1999 (1974).
120. A.G. Marshall, C.L. Hendrickson, G.S. Jackson, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer', *Mass. Spectrom. Rev.*, **17**, 1–36 (1998).
121. T.W. Shannon, F.W. McLafferty, *J. Chem. Soc.*, **88**, 5021 (1966) (title not available).
122. W.F. Haddon, 'Organic Trace Analysis Using Direct Probe Sample Introduction and High Resolution Mass Spectrometry', in *High Performance Mass Spectrometry: Chemical Applications*, ed. M.L. Gross, ACS Symposium Series 70, American Chemical Society, Washington, DC, 97–119, 1978.
123. F.W. McLafferty, F.M. Bockhoff, 'Separation/Identification Systems for Complex Mixtures Using Mass Separation and Mass Spectral Characterization', *Anal. Chem.*, **50**, 69–78 (1978).
124. R.G. Cooks (ed.), *Collision Spectroscopy*, Plenum Press, New York, 1978.
125. M. Barber, R.M. Elliott, 'Comparison of Metastable Spectra from Single and Double Focusing Mass Spectrometers', in *Proceedings of the 12th Annual Conference on Mass Spectrometry and Allied Topics*, Montreal, Canada, June 7–12, 1964, ASTM, Philadelphia, PA, 150–157, 1964.
126. R.A. Yost, C.G. Enke, 'Selected Ion Fragmentation with Tandem Quadrupole Mass Spectrometer', *J. Am. Chem. Soc.*, **100**, 2274, 2275 (1978).
127. D.C. McGilvery, J.D. Morrison, 'A Mass Spectrometer for the Study of Laser-induced Photodissociation of Ions', *Int. J. Mass Spectrom. Ion Phys.*, **28**, 81–92 (1978).
128. T.H. Maugh, II, 'Separations by MS Speed Up Simplified Analysis', Reprint Series, *Science*, **209**, 675–677 (1980).
129. W.W. Harrison, C.M. Barshick, J.A. Klingler, P.H. Ratliff, Y. Mei, 'Glow Discharge Techniques in Analytical Chemistry', *Anal. Chem.*, **62**, 943A–949A (1990).
130. W. Hang, Y. Xiaomei, D.M. Wayne, J.A. Olivares, W.W. Harrison, V. Majidi, 'Glow Discharge Source Interfacing to Mass Analyzers: Theoretical and Practical Considerations', *Anal. Chem.*, **71**, 3231–3237 (1999).
131. J.J. Thomson, 'On the Discharge of Electricity Through Exhausted Tubes without Electrodes', *Philos. Mag.*, **32**, 322–335 (1891).
132. A.L. Gray, 'Plasma Source for Mass Analysis', *Proc. Soc. Anal. Chem.*, **11**, 182, 183 (1974).
133. R.S. Houk, V.A. Fassel, G.D. Flesch, H.J. Svec, A.L. Gray, C.E. Taylor, 'Inductively Coupled Argon Plasma as an Ion Source for Mass Spectrometric Determination of Trace Elements', *Anal. Chem.*, **52**, 2283–2289 (1980).
134. O.D. Sparkman, 'Mass Spectrometry: PittCon'99', *J. Am. Soc. Mass Spectrom.*, **10**, 457–465 (1999).
135. O.D. Sparkman, 'Mass Spectrometry: PittCon'97', *J. Am. Soc. Mass Spectrom.*, **8**, 573–579 (1997).
136. E.R. Denoyer, S.D. Tanner, U. Voelkopf, 'A New Dynamic Reaction Cell for Reducing ICPMS Interferences Using Chemical Reaction', *Spectroscopy*, **14**(2), 43–54 (1999).
137. H.S.W. Massey, E.H.S. Burhop, *Electronic and Ionic Impact Phenomena*, Oxford University Press, London, 1952 (2nd printing, 1956).
138. K.S. Woodcock, 'The Emission of Negative Ions Under the Bombardment of Positive Ions', *Phys. Rev.*, **38**, 1696 (1931).

139. C.M. Judson, 'An Early History of Mass Spectrometry Ionization Methods', University of Kansas, Lawrence, KS (appeared on the Scripps Institute Mass Spectrometry Web site, URL: http://masspec.scripps.edu/jud.html).
140. R.H. Plumlee, 'Mass Spectrometric Studies in Solids', in *Mass Spectrometry in Physics Research*, National Bureau of Standards Circular 522, United States Government Printing Office, Washington, DC, 229–234, 1953.
141. R.E. Honig, 'The Application of Mass Spectrometry to the Study of Surfaces by Sputtering', in *Advances in Mass Spectrometry*, ed. J.D. Waldron, Pergamon Press, London, 162–171, Vol. 1, 1959.
142. R.E. Honig, 'Mass Spectrometric Studies of Solid Surfaces', in *Advances in Mass Spectrometry*, ed. R.M. Elliott, Pergamon Press, London, 25–37, Vol. 2, 1963.
143. Presentations from the Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics, Dallas, TX, June 13–17, 1999, American Society for Mass Spectrometry, Santa Fe, NM, 1999: (a) P. Kingshoot, H. St. John, S. McArthur, K. McLean, C. Sutton, H. Griesser, 'Surface-MALDI Investigation of Protein Adsorption to Contact Lens Surfaces', Thursday Poster (049) in Section C: Laser Desorption and Ionization; (b) M. Stoeckli, P. Chaurand, R. Caprioli, 'Application of MALDI MS Imaging of Biological Samples', Tuesday Poster (264) in section G: Proteins and Peptides.
144. K.H. Maurer, C. Brunee, G. Kappus, K. Habfast, U. Scroder, P. Schulz, 'Direct Analysis of Daughter Ions Arising from Metastable Decompositions', in Proceedings of the 19th Annual Conference on Mass Spectrometry and Allied Topics, Atlanta, GA, May 2–7, 1971; presentation K9, no extended abstract presented, American Society for Mass Spectrometry, Santa Fe, NM, 1971.

Artificial Intelligence and Expert Systems in Mass Spectrometry

Ronald C. Beavis
Proteometrics LLC, New York, NY, USA

Steven M. Colby
Scientific Instrument Services, Inc., Ringoes, NJ, USA

Royston Goodacre
University of Wales, Aberystwyth, UK

Peter de B. Harrington
Ohio University, Athens, OH, USA

James P. Reilly
Indiana University, Bloomington, IN, USA

Stephen Sokolow
Bear Instruments, Santa Clara, CA, USA

Charles W. Wilkerson
Los Alamos National Laboratory, Los Alamos, NM, USA

1 Introduction **2**
1.1 Definitions of Artificial Intelligence and Expert Systems 2
1.2 Growth in Technology 2
1.3 Article Summary 2
2 Brief History of Computers in Mass Spectrometry **2**
2.1 Introduction 2
2.2 Early Devices 2
2.3 Instrument Design 3
2.4 Computerization 3
2.5 Brief Introduction to Artificial Intelligence and Expert Systems 3
2.6 Brief Overview of Artificial Intelligence and Expert Systems in Mass Spectrometry 4
3 Mass Spectrometry Data Systems **4**
3.1 Introduction 4
3.2 Fundamental Tasks of a Data System 5
3.3 Requirements for Operating Systems 6
3.4 Impact of Continuing Advances in Computers on Mass Spectrometry Data Systems 6
3.5 Programmability 7
4 Biological Applications **9**
4.1 Protein Sequence Determination 9
4.2 Database Search Strategies 9
4.3 Nucleotide Databases 9
4.4 Protein Modification Analysis 10
4.5 Use with Differential Displays 11
4.6 Alternate Splicing 11
5 Mass Spectrometry Applications of Principal Component and Factor Analyses **12**
5.1 Introduction 12
5.2 Selected History 12
5.3 Introductory Example 13
5.4 Theoretical Basis 14
5.5 Related Methods and Future Applications 17
5.6 Reviews and Tutorials 18
5.7 Acknowledgments 18
6 Artificial Neural Networks **18**
6.1 Summary 18
6.2 Introduction to Multivariate Data 18
6.3 Supervised Versus Unsupervised Learning 18
6.4 Biological Inspiration 19
6.5 Data Selection 20
6.6 Cluster Analyses with Artificial Neural Networks 21
6.7 Supervised Analysis with Artificial Neural Networks 23
6.8 Applications of Artificial Neural Networks to Pyrolysis Mass Spectrometry 26
6.9 Concluding Remarks 28
7 Optimization Techniques in Mass Spectrometry **29**
7.1 Introduction 29
7.2 Time-of-flight Mass Spectroscopy Mass Calibration 29
Abbreviations and Acronyms **31**
Related Articles **31**
References **32**

This article provides a brief introduction to aspects of mass spectrometry (MS) that employ artificial intelligence

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

(AI) and expert system (ES) technology. These areas have grown rapidly with the development of computer software and hardware capabilities. In many cases, they have become fundamental parts of modern mass spectrometers.

Specific attention is paid to applications that demonstrate how important features of MS are now dependent on AI and ESs. The following topics are specifically covered: history, MS data systems, biological applications, artificial neural networks (ANNs), and optimization techniques.

1 INTRODUCTION

1.1 Definitions of Artificial Intelligence and Expert Systems

This article covers the application of AI and ESs as applied to the techniques of MS. ESs are methods or programs by which a fixed set of rules or data is used to control a system, analyze data, or generate a result. In contrast, AI is associated with the higher intellectual processes, such as the ability to reason, discover meanings, generalize, or learn. In relation to MS, AI is generally limited to cases wherein ANNs are employed to learn or discover new patterns or relationships between data. Reviews of AI and ESs are available.[1,2]

1.2 Growth in Technology

The growth in MS has been spurred by improvements in software sophistication and computer capabilities. The ability of computing systems to both collect and analyze data has grown very rapidly since the 1970s. The most important improvements have been in calculation speed of the machines, their ability to store large amounts of data very quickly, and their size. These improvements have allowed processes such as multitasking during data acquisition, where the computer both collects data and controls the instrument operation, and automated spectral matching, where large volumes of data are quickly analyzed.

The improvements in computer technology have resulted in an increase in the performance and types of mass spectrometers available. For example, instruments with even the simplest types of mass analyzers are now computer controlled. This has dramatically increased the stability, reproducibility, and capabilities of these devices. It is now possible to perform a tandem mass spectrometry (MS/MS) experiment where the data collection parameters are changed on the millisecond timescale in response to the data collected.[3] This allows a library search to be performed, possible match candidates to be experimentally tested, and a positive identification to be made, all during the elution of a chromatography peak.

The development of computers has also allowed the use of new types of MS. For example, the data generated by Fourier transform mass spectrometry (FTMS), pyrolysis, and electrospray MS would be very difficult if not impossible to collect and analyze without high speed computers.

1.3 Article Summary

This article includes sections on the history of computers in MS, MS data systems, biological applications, MS applications of principal component analysis (PCA) and factor analysis (FA), ANNs, and optimization techniques in MS. This article does not include a discussion of the use and development of libraries of electron impact ionization data or of peak deconvolution and component identification based on these libraries. Reviews of these topics are available.[4,5]

2 BRIEF HISTORY OF COMPUTERS IN MASS SPECTROMETRY

2.1 Introduction

Digital computers are now an indispensable part of most analytical instruments. There are many reasons for this pervasive presence. Perhaps most important is the ability of computers to perform repetitive tasks without variation (in the absence of hardware failure), which is critical to reproducible and defensible experimental results. Further, properly designed and implemented computer control/data systems maximize instrument and laboratory efficiency, resulting in higher sample throughput, faster results to the end-user, and increased profitability (either in terms of publications or currency) for the laboratory. As a technique that arguably provides more chemical and structural information per unit sample than any other, MS has been employed in a variety of environments over its long history. The evolution of the mass spectrometer from a fundamental research tool for the elucidation of atomic and molecular properties to a benchtop turnkey instrument, in large measure parallels the evolution of both discrete and integrated electronic devices and computational hardware and software.

In this article no attempt is made to tabulate an exhaustive list of historical references to the application of computers in MS, but rather selected citations are presented to provide a flavor of the development in the field. There is one monograph dedicated to computers in MS,[6] and the topic is given treatments ranging from cursory to complete in a variety of books on mass spectrometric techniques.

2.2 Early Devices

Early mass analyzers were spatially dispersive instruments, or mass spectrographs,[7] utilizing static magnetic or DC (direct current) electric fields to perturb the

trajectories of accelerated ions. At the time of their development (ca. 1910–1920) photographic plates were placed in the focal plane of the spectrograph, and after exposure to the ion beam an image was developed and the resulting data analyzed. Quantitative analyses were effected by comparing the amount of exposure on a film produced when the unknown sample was determined to calibrated plates developed after measuring known amounts of reference materials. This technique is still in use today in certain specialized (and somewhat archaic) applications. In the 1930s and 1940s detectors based on direct measurement of ion beam flux (such as the Faraday cup and electron multiplier), were introduced. Such detectors are single-channel transducers, and require that slits be positioned (in a dispersive instrument) to limit the exposure of the detector to a single mass at any given time. The signal is then amplified and recorded as a function of some independent variable (such as magnetic field strength, or the ion accelerating voltage) that is proportional to the mass-to-charge ratio (m/z) of the ions in the sample.

With the introduction of electronic detectors, it became practical to couple detector output to a digital computer via some type of interface. For low-intensity signals, such as measurement of discrete ions, pulse-counting techniques are employed. As this is inherently a digital process, transmission of data to a computer is relatively straightforward. Larger signals, characterized by significant and measurable detector currents, employ analog-to-digital converters (ADCs) prior to storage and manipulation on the computer.

2.3 Instrument Design

2.3.1 Time-of-flight

Time-of-flight (TOF) mass spectrometers were first developed in 1932,[8] but the most familiar design, which forms the basis of current instruments, was described by Wiley and McLaren in 1955.[9] Accurate measurement of ion time of arrival at the detector is key to achieving optimum resolving power and mass accuracy with this instrument. Prior to the introduction of computer data acquisition, oscillographic recording was required, with manual postprocessing.

2.3.2 Quadrupole

The quadrupole mass filter was first described in 1958.[10] The advantages of this instrument include small size, low ion energy (volts rather than kilovolts for dispersive and TOF instruments), modest production costs, and the ability to quickly scan through a wide range of m/z values. As a result, this design has become by far the most popular variety of mass spectrometer. A related mass analyzer, the quadrupole ion trap, was not widely developed until the 1970s.[11] Like the linear quadrupole mass filter, the ion trap is small, inexpensive, and robust. Both of these devices rely on the application of concerted radiofrequency (RF) and DC fields in order to define conditions under which ions have stable trajectories in the instrument.

2.3.3 Ion Cyclotron Resonance

The ion cyclotron resonance (ICR) mass spectrometer, first reported in 1968,[12] relies on the absorption of RF energy and the natural precession of charged particles in the presence of a magnetic field for mass separation. Nominal resolving power is obtained in this instrument when operated in a continuous scanning mode, where the RF frequency is slowly swept and energy is absorbed when ions in the cell are resonant with the excitation. The most common incarnation of ICR is often referred to as FTMS,[13] and spectral information is extracted from the digitally recorded decay and dephasing of ion orbits after a pulsed application of RF energy. This approach allows for significantly improved resolving power (1000-fold improvement) over the scanning experiment.

2.4 Computerization

As a result of the widespread availability of minicomputers in the late 1960s, and microcomputers in the 1970s and 1980s, automation of mass spectrometer control, tuning, data acquisition, and data processing became practical. The reduction in both size and cost of computational engines, with a concomitant increase in processing power, cannot be overemphasized in the development of automated data systems for mass spectrometers (and other analytical instrumentation). Certainly, the widespread implementation of gas chromatography/mass spectrometry (GC/MS) would have been significantly delayed without the availability of reasonably priced quadrupole mass spectrometers and minicomputer-based data acquisition and processing equipment. The operation of FTMS would be nearly impossible without the involvement of computers.

2.5 Brief Introduction to Artificial Intelligence and Expert Systems

Almost from the beginning of the digital computing era, both hardware and software engineers have been interested in developing computing tools that can monitor their environments, and subsequently make decisions and/or carry out actions based on rules either known a priori – from programming – or deduced – as a result of iterative observation/decision/feedback experiences. Such computational devices may be called ‘expert systems’, or may be said to operate based on ‘artificial

intelligence'. It is certainly beyond the scope of this article to provide a complete history of AI and ES, but there are a multitude of both books and research articles related to this topic.[14,15] Today, many parts of our world are monitored, and in some cases controlled, by automated, computerized equipment. In an effort to make these devices more responsive and efficient, many of them employ embedded ESs of varying degrees of sophistication. Programming languages such as LISP and PROLOG have been developed specifically to facilitate the development of software to implement AI and ES. The combination of powerful hardware, innovative algorithms, and capture of years of expert knowledge has allowed instruments to become increasingly independent of operator interaction, reducing the possibility for error and allowing the scientist to concentrate on the interpretation of the processed data and the formulation of new experiments.

2.6 Brief Overview of Artificial Intelligence and Expert Systems in Mass Spectrometry

In the world of MS, AI and ES tools are used in three primary areas: optimization and control of the performance of the mass spectrometer itself, collection of the detector signal as a function of m/z, and analysis of the data.

2.6.1 Spectrometer Control

There are many instrumental parameters that need to be adjusted and held at an optimum value for best spectrometer performance. Initially, the instrument must be tuned, i.e. brought to a state in which peak intensity, peak shape, and mass calibration are all within acceptable limits. This is accomplished by introducing a known compound, such as PFTBA (perfluorotributylamine), into the spectrometer that produces a variety of well-characterized fragments over the mass range of interest, and adjusting (in an optimized fashion) the various instrument parameters to achieve the desired level of performance. Computers are almost invariably used to perform this task, because the adjustable parameters are often highly interrelated (repeller voltage, ion focusing lens potentials, electron multiplier voltage, mass scan rate, ion storage time, chemical ionization reagent gas pressure, time delay for ion extraction, etc.). Techniques such as simplex optimization are used to efficiently locate in parameter space the best-tune conditions. After tuning is complete, the computer can then monitor all of the vital signs of the instrument during operation, and alert the spectrometrist of marginal performance conditions, and even recommend appropriate interventions, before data quality is affected.

2.6.2 Data Collection

In almost all data systems, the operator uses the computer to define the scope of the measurements to be made. Subsequently, the computer sets instrument parameters to control, for example, the speed of data collection, the mass range to be recorded, and other instrument type-dependent variables. As the experiment is performed, the computer records the detector signal via either a direct digital interface (for counting experiments) or an ADC. Correlation of the detector signal with the corresponding m/z condition is accomplished through a mass-axis calibration routine. Depending on the mass spectrometer type, this may be a DC, RF, or time reference.

2.6.3 Data Analysis

After the data have been collected, their chemical information must be extracted and interpreted. There has been a significant amount of development in the area of data analysis software since the first report of such use in 1959.[16] In this early work, a system of simultaneous linear equations were used to convert raw peak areas to normalized analyte mole fractions. A 17-component sample required 0.5–3 min of computing time for processing. Today, mixtures with nearly an order of magnitude more analytes can be reduced in less time, providing significantly more information than simply peak quantitation. In addition to quantifying analytes, mass spectrometer data systems routinely provide identification of species from their mass spectral fingerprints. One of the earliest examples of the application of AI to mass spectral interpretation was the work of Djerassi et al.[17] A LISP (a list processing language)-based code, DENDRAL, was developed and subsequently applied to a variety of analyte classes. Most mass spectrometrists are familiar with spectral libraries, ranging from the print version of the so-called eight-peak index[18] to the most modern computerized systems. The latter use intelligent peak-searching and pattern-matching algorithms to provide the operator with the most likely identities of species in a spectrum.

3 MASS SPECTROMETRY DATA SYSTEMS

3.1 Introduction

Since the mid-1970s the programming of mass spectral data systems has changed enormously. Although the basic tasks of an MS data system are fundamentally the same now as they were in the 1970s, many of the numbers involved have become substantially larger. In addition, developing mass spectral technologies such as FTMS have placed very heavy demands on the acquisition process.

Spectrum libraries have become larger. Analyses of large complex molecules (i.e. peptides) may consume a great deal of computer resources. Fortunately, the changes in computer and operating system technologies since the 1970s have been even more staggering than the changes in MS.

Section 3.2 defines the basic tasks of a MS data system. Section 3.3 describes the requirements imposed on the computers and operating systems that aspire to perform these tasks. Section 3.4 examines some of the specifics of how changes in computer technology have affected mass spectral data systems. Section 3.5 treats the subject of programmability. As the number of MS algorithms proliferate, the need for a data system to be customizable (i.e. programmable) has become ever more important – if users cannot define their own ways of collecting and analyzing data, unlimited computer power may be useless. Practical examples from actual data systems are presented, to show that the concerns of a programmer are often quite different from those of a chemist.

3.2 Fundamental Tasks of a Data System

The tasks of an MS data system are often neatly divided into instrument control, acquisition of data to a storage medium, and analysis of the data. The division is, of course, not really so simple. The collection of data depends significantly on simultaneous instrument control and the analysis of the collected data may be fed back into the instrument control. For example, in the process of tuning an instrument, the software may vary a variety of different parameters, each time collecting and assessing some data before trying a new set of conditions. In this case there is a feedback loop that involves control, acquisition, and analysis. The feedback must be very tightly orchestrated to be useful.

3.2.1 Instrument Control

The task of instrument control has several aspects – routine operation, instrument protection, tuning, and diagnostic programs. During routine operation many voltages must be set or scanned, and as much instrument status as possible must be read from the instrument. This status information may be stored with the data. It may be used to keep temperatures stable within the instrument by running PID (Proportional–Integral–Differential) loops on heaters. Or, it may be used to protect the instrument. For example, a sudden rise in pressure may indicate a leak and some voltages should be turned off. If mass peaks are saturated, perhaps the detector voltage should be decreased, or a warning message should be shown on the computer screen. The process of tuning and diagnostic programs, each in their own way, are microcosms of the entire MS data system. Those experienced in designing MS data systems have learned that it is advantageous to first write the diagnostic programs, basing them on very small and easily understood modules. These will, after all, be needed for the first evaluation of the instrument. It is then possible to base the ordinary operation of the instrument on these same modules. Doing so tends to provide the entire system with a relatively good structure. This bottom-up modular structure also makes it easy to add elementary operations (e.g. when adding new hardware) and higher-level operations can almost always be defined as combinations of the elementary processes.

3.2.2 Data Collection

The task of data collection is fundamentally important. Today's computer operating systems are multitasking and therefore capable of running several processes at once. If the mass spectrometer is connected to a chromatograph or other time-dependent sample-introduction device, then the data collection must have priority over all other operations. A disaster can result if some data are missed. To guard against this, an MS data system may use more than one processor, dedicating at least one processor to data collection.

3.2.3 Data Analysis

Analysis of the collected data includes the following items:

- Conversion of raw (e.g. profile or Fourier-transform) data to mass peaks.
- Data display for the chemist.
- Enhancement of the data by background subtraction or other means.
- Use of the area under chromatogram peaks or other MS data to compare unknowns with standards and to achieve quantitative results.
- Library searching.
- Report generation.

A modern data analysis program includes other more advanced topics, which are covered elsewhere in this article; even the elementary operations listed above have many variations. Data systems must be flexible enough to allow the user to perform the operations in exactly the way and order required, hence the importance of programmability. The control, collection and analysis are all achieved through a user interface. This element of the data system determines the ways in which the user is able to enter information and communicate with the system. Section 3.4 looks at how changes in operating systems have affected the user interface and hence the ease of using mass spectral data systems.

It should be noted that, from the programmer's point of view, the design of an easy-to-use user interface is generally a much harder and time-consuming part of the programmer's task than implementing all of the chemical algorithms. The user interface includes the display of data and instrument status, as well as input devices such as menus and buttons that allow the user to control the system.

The display must respond to real changes in instrument status in a timely manner. For example, suppose that in the process of tuning an instrument the user is manually increasing a voltage setting by clicking a button on the screen. If nothing happens to the status display for more than a second, the user is very likely to click on the button again to accelerate the change in the system. This is simply because faster computer response time has naturally led to greater user impatience. However, overclicking can result in overshooting an optimum setting and this makes instrument adjustment almost impossible. Therefore, a crucial task of the data system to reflect the real-time status of the instrument.

3.3 Requirements for Operating Systems

As noted above, data collection must never fail. As the operating system used by a chemist is almost certainly a multitasking system, it is necessary to ensure that the highest possible priority is given to the data collection task. It must not be possible for other tasks to usurp the precious time required by the data collection procedures. This is the overriding concern in the selection of an operating system. For this reason Windows NT is a much more appropriate choice than Windows 95, for MS data systems. Several other operations also require high priority because they cannot be interrupted, such as those that involve delicate timing or real-time feedback.

If multiple processors are used, other requirements must be considered. For example, if an embedded processor in the instrument communicates with the data system over a serial or parallel line, it is important that the instrument be plug-and-play; that is, both sides should disconnect cleanly when the cable is disconnected and reconnect automatically when the cable is reconnected. If the embedded processor is depending on the data system for control and the connection is broken, the embedded processor should go into a standby state for safety purposes.

Most instrument manufacturers have chosen to base their data systems on PCs running Microsoft operating systems. A brief survey of 22 instrument manufacturers found that 18 of them were using a version of Microsoft Windows. Others used OS/2, and operating systems from Hewlett-Packard, Sun, and Apple.

3.4 Impact of Continuing Advances in Computers on Mass Spectrometry Data Systems

The most obvious improvements in computers have been the dramatic increases in speed and in the size of computer memories and storage. The forefathers of today's data systems were developed on home-built computers using Intel chipsets or on systems produced by Data General, Digital Equipment, Commodore, or Apple (section 2). These systems typically had 16–64 kB of ram and sometimes included a 5 or 10 MB disk. Since the 1970s the availability of memory and storage has increased by over three orders of magnitude. Execution times have also increased, albeit to a lesser extent. For example, library searches are now four to eight times faster.

Operations that require large arrays of data and massive amounts of arithmetic have benefited most from the improvements in hardware design. These improvements have also made it much easier to implement algorithms. Previously, developers had to implement programming tricks to handle very large arrays of data. Activities such as library searches required extensive coding in order for their execution to be completed in a reasonable amount of time. Today even more advanced and thorough searches can be implemented with a few lines of C code. These advantages also apply to algorithms written by the user of the data system (if a programming language is available – see section 3.5).

Networks are beginning to have a major impact on data systems. Local networking provides a great advantage by giving the user a wide variety of high-capacity storage options. The internet allows easier transfer of data and results, but has found only limited use in instrument control. In both cases security issues are a major concern. Although most laboratory management systems provide security features, such as passwords, etc. the proper set-up and administration of these controls is required. This may be beyond the resources of some laboratories and is clearly an added cost.

The current operating systems have had a significant impact on the standardization of user interfaces. In the first mass spectral data systems, each had different ways to enter commands or click with a mouse. It was therefore a major challenge to instruct users on how to use a data system. In some cases the operator had to develop significant programming skills to use the system. In current user interfaces many operations, such as cut and paste, are standardized on the computer. As these are performed in the same way in most computer programs, everyone who has worked with a computer is well-versed in the art of using menus and mouse clicks to interact with a computer program. The fact that a large majority of data systems are based on Windows makes this even more true. Chemists now have a much easier time learning

to use new data systems because they already have a good idea of how the user interface will work. This standardization has produced the one drawback, in that many programs now look the same and it is becoming a challenge for programmers to make their systems unique and original.

3.5 Programmability

As discussed above, many aspects of modern mass spectral data systems require that they be programmable (or customizable). Every system is limited to have a finite number of built-in operating modes and algorithms. The chemist, therefore, needs to have the ability to mix modes and tailor algorithms to suit experimental objectives. The programmer who writes the data system is not able to anticipate which aspects of an algorithm the user may wish to vary, so ultimately the user needs to be able to program functions into the system. This section describes the elements that a system must include, to be truly programmable.

First the user needs a language to write algorithms in. The language needs to incorporate basic arithmetic and common math functions. It also needs to have program flow control elements such as loop and logic structures ('if', 'while', and 'repeat'). The user needs to be able to use predefined variables such as 'first_mass', 'last_mass', 'detector_voltage'. They also need to control MS operations with built-in commands such as 'Do_one_scan', 'filament_on', 'filament_off'. The language must have built-in feedback so that decisions can be based on the state of the instrument or the nature of the data. Functions such as 'Source_temperature' or 'Manifold_pressure' can serve this purpose. The most advanced systems include functions such as 'Intensity of mass 278 in the last data-set' or 'Mass of the biggest peak in the last data-set' that return facts about the data.

The language should to be able to perform all control, collection, and analysis steps. It ought to be possible to run more than one process at once, so that the system can collect one set of data while analyzing another, and perhaps reporting on a third. For good laboratory practice, it is important to have functions to write any sort of information into a file. This will ensure that every dataset has enough information stored within it to show exactly how it was acquired. It also allows diagnostic programs to keep track of instrument performance over any period of time.

The feedback functions in the language can be used for a wide variety of algorithms. For example, in the arena of safety, the chemist can specify the actions to be taken if a temperature or pressure gets too high. Alternatively, the chemist could write a tuning loop that sets a voltage, collects a scan of data, and reads back information about a peak.

Section 3.5.1 includes a number of illustrative examples. The procedures are written in a pseudocode quite similar to an actual programming language. The first example shows the optimization of data collection by timing acquisition. The second is part of an autotune algorithm. The third is a higher-level procedure for automatic quantitation, meant to run continuously in the background.

3.5.1 Example 1: Timed Acquisition

One can increase the amount of analytically relevant information by only collecting data that is appropriate for the retention time. The following routine is for an MS/MS instrument that does single reaction monitoring of several different reactions, 219–69 for the first two minutes, 512–69 for the next two minutes and 131–69 thereafter:

```
start_collection
srm(219,69)
while retention_time < 2 : scan : end
srm(512,69)
while retention_time < 4 : scan : end
srm(131,69)
while retention_time < 10 : scan : end
end_collection
```

The functions referred to have the following meaning:

- srm(m1,m2) means set the instrument to monitor the reaction m1–m2.
- scan means collect one scan of data.

3.5.2 Example 2: Tuning

This is an example of a tuning algorithm called 'optimize_lens'; it's one argument specifies which lens to tune. While tuning, the system collect raw data. For these data, 'height' refers to the height of the biggest peak in the dataset. As before, 'scan' means collect one scan of data. The items 'biggest_area' and 'best_lens' are temporary variables. The goal of the procedure is to find an optimum value of a lens.

```
optimize_lens(n)
  biggest_height = 0
  for lens(n) = −100 to 0 in steps of 1
    scan
    if height >= biggest_height
      biggest_height = height
      best_lens = lens(1)
    end
  end
  lens(n) = best_lens
```

When this is done, lens *n* will have been optimized.
Such a routine may be built into a higher level-routine:

```
optimize_all_lenses
  optimize_lens(1)
  optimize_lens(2)
  optimize_lens(3)
  etc . . .
```

This process may be abstracted to as high a level as required.

3.5.3 Example 3: Automatic Quantitation

If the data system is designed properly, rules can be defined to run continuously in the background. Here is an example of a high-level algorithm that provides automatic updating of a quantitation list when the chemist changes the calculations for one of the compounds in the list. For example, suppose the user has collected several data files, including analytes and internal and external standards. They have quantitated a set of compounds in these data files, using mass chromatograms to obtain an area for each unknown or standard. The areas and concentrations of the standards are used to create a response curve. The areas of unknowns are used, in conjunction with the response curve, to calculate the unknown concentrations. One now has a list of areas and quantities for each compound, along with the information on how they were computed. If the user were to change the area of one of the standard compounds by changing the parameters that went into its calculation, we would like to see the amounts of all related peaks change correspondingly. Here is an example of a procedure that performs this operation.

```
Repeat-forever
  if some_compound_changed
    for r = 1 to number_of_response_points
      c = external_standard(r)
      c1 = internal_standard(r)
      response_x(r) = compound_area(c)/
        compound_area(c1)
      response_y(r) = compound_amount(c)/
        compound_amount(c1)
    end
    for c = 1 to number_of_compounds
      compute_amount(c)
    end
  end
  Sleep_one_second
end
```

The functions referred to have the following meanings:

- some_compound_changed — set to 'True' if any one of the compounds in the list changed area or amount, which means that 'compound_area' or 'compound_amount' changed for that compound.
- number_of_response_points — the number of points in the response list.
- external_standard(r) — the compound number of the external standard at position r in the response list.
- internal_standard(r) — the compound number of the internal standard at position r in the response list.
- compound_area(c) — the area under the chromatogram for compound c.
- compound_amount(c) — the calculated or given amount of compound c.
- number_of_compounds — the number of compounds in the list.
- compute_amount(c) — computes the amount of compound c from its area and the response list.
- sleep_one_second — prevents the procedure from hogging the system – there is no need to check more than once a second that the user has changed the data.

This procedure checks whether some compound has changed area or amount (changed by the user). If so, it recalculates the response curve by filling in each point on the response curve from the areas and the amounts of the appropriate compounds. Then, for each compound, it computes the amount of that compound ('compute_amount' uses the response curve). If the display of data is responsive to changes in the data, the user will see all areas and amounts change as soon as one value is changed. In section 3.2 an example was given of the necessity of a close link between data and display; this procedure is another example.

To keep the code simple, this example assumes that there is only one response list involved. However, it is easy to extend the code to a system that includes several response lists.

These examples give an indication of how programmable a data system can be. The challenge for the

designers of data systems is to balance flexibility with simplicity for the sake of the chemist who is content with the basic operation of the system. MS is not a trivial task and operating a mass spectral data system is likely to remain a challenging task as the functionality of MS Data systems continues to expand. Hopefully, the user interface, which is what makes it possible to use all this functionality, will keep pace.

4 BIOLOGICAL APPLICATIONS

4.1 Protein Sequence Determination

MS has long had as a goal the ability to determine the sequence within polymeric biologically important molecules, such as DNA and proteins. There have been notable advances in this area in the period 1990–1999.[19–24] However, the goal of developing a simple yet general method for rapidly sequencing these molecules by MS has remained elusive.

Fortunately, alternative approaches have been introduced that take advantage of the large amount of DNA and RNA sequence information that has been generated by genome sequencing projects and which is currently stored in databases. Using this nucleic acid sequence information, it is possible to determine whether the results of a mass spectrometric experiment correspond to a sequence in a database. If such a correspondence exists, it is no longer necessary to sequence the protein (or corresponding RNA) by MS or other means – the answer can be simply read from the database. If the database information is incomplete, it can serve as a starting point for other studies, greatly reducing the experimental work required for the determination of the full sequence.

4.1.1 Peptide Cleavage and Extraction

All protein sequence identification experiments begin with the creation of a set of smaller oligopeptide molecules from the intact protein. The patterns generated from these oligopeptides are then used to search nucleotide sequence databases. These oligopeptides (frequently referred to simply as 'peptides') are produced by the action of a reagent that cleaves the protein's peptide bond backbone at sequence-specific sites, such as peptide bonds that are adjacent to a limited set of amino acids. Peptide digesting enzymes, such as trypsin or endopeptidase Lys-C, are commonly used for this purpose. Reactive amino acids, particularly cysteine residues, are protected with chemical reagents that prevent them from modification during the process.

4.1.2 Dataset Generation – Mass Spectrometry, Matrix-assisted Laser Desorption/Ionization and Electrospray Ionization

Once the oligopeptides have been generated, the masses of all of the peptides generated from a protein can be measured at once, using matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI) ion sources, mounted on a variety of different types of mass analyzers. Analysis using a MALDI ion source is currently the most common method, but the use of sophisticated deconvolution will make ESI a viable option. Proteins produce patterns containing 10–1000 isotopic peak clusters, depending on the sequence of a particular protein. Each peak cluster represents a particular peptide sequence.

Alternatively, the ions corresponding to an individual peptide from a protein digestion can be isolated, either using chromatography or MS/MS techniques. The resulting ions can then be fragmented in a gas phase collision cell producing a pattern of masses characteristic of the sequence of the original peptide (MS/MS or MS/MS/MS, i.e. MS^n). This pattern can be used to search databases, using the accumulated knowledge of the preferred gas-phase peptide bond cleavage rate constants. The resulting pattern is strongly affected by the time elapsed between collision and measurement of the product ion distribution, so different rules must be applied for different types of MS/MS analyzers.

4.2 Database Search Strategies

The data sets generated by mass spectrometric experiments can be compared to the nucleotide sequence information present in databases in several ways. All of these methods share some common features. In order to compare sequences, the chemical reactions involved in producing the cleaved peptides are simulated, producing a theoretical set of peptides for each known protein sequence in the database. This simulation can either be done during the search process or a specialized database consisting of the peptides resulting from a particular cleavage and protection chemistry can be prepared in advance. The simulations are then compared to the experimental data, either using specialized correlation functions or using multiple-step discrete pattern matching. This comparison is done by assuming that sequences that correspond to the experimental data set will contain a set of peptides with masses that agree with the experimental data, within some experimental error.

4.3 Nucleotide Databases

Databases of complete gene sequences can be searched as though they were protein sequence databases. The

existence of known start codons and intron/exon assignments allows the use of, either MS or MS^n patterns. Nucleotide databases that contain incomplete sequence information, such as the database of expressed sequence tags (dBEST),[25] present special challenges. In this type of database, there are six possible peptide sequences for each nucleotide sequence and each must be searched independently. The short length of the sequences makes the use of MS-only data impractical; these databases require the use of MS^n fragmentation patterns.

4.3.1 Annotated Protein Databases

Dedicated protein sequence databases that store annotated oligopeptide translations of nucleic acid sequences are the best databases for any MS-related search strategy. The annotations in the database indicate what is known about post-translational modification of the protein, allowing the chemical cleavage simulation to be performed more accurately than is possible using nucleotide information alone. The number of protein sequences in this type of database is still very limited – annotation is time-consuming and only possible when detailed experimental results are available for a particular sequence.

4.3.2 Confirmation and Scoring Results

The results of comparing a set of experimental masses to a sequence database usually results in the identification of a number of candidate sequences that match to some extent with the experimental data. The task of distinguishing random matches from the 'correct' match has been approached in a number of ways. The simplest scoring system involves counting the number of masses that agree within a given error and reporting the sequence with the most matches as being the best candidate sequence. This approach is very simplistic and frequently deceptive. More sophisticated scoring schemes involve appraising pattern matches on the following criteria:

- sequence coverage – the fraction of the candidate protein represented by the experimental masses;
- sequence distribution – the pattern of matched peptides in the candidate protein;
- mass deviation – the pattern of experimental mass deviations from the simulation values;
- statistical significance – the likelihood that the match could have occurred at random.

Research into the appropriate scoring scheme for MS and MS^n match scoring is still ongoing. The most successful of scoring systems will be the basis for the next generation of fully automated protein identification instruments.

Currently, none of the protein identification algorithms make use of AI or algorithm training methods. The Profound algorithm is currently the closest to using AI – it uses a Bayesian statistical approach to evaluating data sets, allowing for the unbiased evaluation of search results and for the detection of multiple sequences in a single MS data set.[26]

4.4 Protein Modification Analysis

MS may have limitations in the determination of protein sequences de novo, but it is very well suited to the detection of chemical modifications of a known sequence. The detection of these modifications is very dependent on good software as there is too much information for manual data reduction. The general strategy is very similar to that used to identify proteins, a process that grew out of the standard practice for finding modifications. The general strategy is as follows: determination of the intact protein molecular mass; cleavage to peptides; generation of mass spectra; and automated, detailed comparison of the MS data set with a known sequence.

4.4.1 Peptide Cleavage and Extraction

The cleavage and protection chemistry available for detection of modifications is much broader than that used in protein identification experiments. Any proteolytic enzyme, chemical cleavage or protection method can be used, depending on the type of modification sought. Popular endoproteinase enzymes are trypsin, endoproteinases Lys-C and Asp-N, and *Staph*. V8 proteinase.[27] Exopeptidases, such as carboxypeptidase A, B, and P can also be useful for generating C-terminal sequencing ladders for smaller peptides.[28] Unlike the protein identification procedure, it is very useful to follow a time course of protein cleavage, as the dynamics of proteolysis can provide valuable clues to the identity and location of modifications. Chemical cleavage reagents, such as cyanogen bromide, iodosobenzoic acid and hydroxylamine, can be used in place of enzymes. These reagents are less popular than enzymes, because of their propensity for producing complicating modifications in the sequence through side-reactions.

4.4.2 Generation of Mass Spectroscopy Datasets

Mass spectroscopy datasets are collected in the same way as for protein identification experiments. Typically, a number of experiments are run, using different cleavage reagents with different and complementary specificity. For example, both a trypsin and endoproteinase Asp-N digest would be both run, taking several time points during the reaction to reconstruct its time course. All of the data collected is stored for later analysis.

Datasets for MS^n can be prepared that greatly assist analysis in the detection of common modifications, such as phosphorylation or disulfide cross-linking. These modifications produce characteristic fragmentation signals following gas-phase collisions. The most popular method for collecting this type of specialized data is directly coupling the output from high-performance liquid chromatography (HPLC) to an MS/MS device (such as a triple quadrupole or a ion trap analyzer) and flagging spectra that contain these characteristic signals.

4.4.3 Comparison with Sequence

Once a dataset has been assembled, it must be compared with the results that would be expected from the predicted amino acid sequence. For a simple enzymatic cleavage experiment on a protein that has 30 theoretical cleavage sites (N) and no cystines, there are approximately 450 possible product peptides. The complexity of the task of examining a dataset for each of the possible products and locating the peaks that do not agree is clearly too time-consuming and error prone to be performed manually.

The majority of data is analyzed using automated systems to assist the investigator – no system that performs a complete and reliable analysis is currently available. Modern analysis is performed by first determining the mass of a peak in the MS dataset and searching a sequence for a peptide with a mass that is within a user-defined error of the experimental value. The dataset can be a single mass spectrum containing all of the cleaved peptides or an HPLC/MS dataset that contains thousands of individual spectra, each of which will contain zero or more of the peptides, depending on the chromatographic conditions.

The best analysis systems use a multifactorial fuzzy-logic-based approach to analyzing the data. The entire dataset is interrogated and individual matches rated with respect to all of the other assignments. Peptides with the same mass (within the allowed error) are assigned based on the kinetics of the cleavage reaction, as inferred by the fuzzy logic rules. Peaks that can be assigned by mass, but which are unlikely based on the entire data set, are flagged for further examination and confirmation. These flagged peaks, as well as those that could not be assigned are then iterated through a selection of known modifications and the complete sequence assignment process repeated. The fuzzy logic assignments depend on the entire data set so the change of value in the simulated experiment requires a complete reexamination for the assignments.

Once this iterative process is finished, the results can be projected back onto the theoretical sequence, with each assignment flagged and color coded so that interesting portions of the sequence are displayed. This process is particularly effective if the three-dimensional structure of the protein is known, where the peptides can be located in a structure diagram shown in a stereoscopic display.

4.5 Use with Differential Displays

Differential displays are a particularly useful tool in current cell biology. They consist of some type of high-resolution protein separation system, such as two-dimensional gel electrophoresis, and a signal detection process such as affinity or silver staining. A cell challenged in various ways will produce displays that differ as the protein complement being expressed in the cell changes. By overlaying displays, spots that change are apparent. The challenge is then to determine what protein has appeared (or disappeared or changed positions).

The techniques described in sections 4.1–4.3 can be applied to these displays. By excising interesting areas of the separation bed and extracting the protein components in various ways, the protein sequence can be rapidly identified. A new generation of automated differential display devices utilizing MS as a protein identification system is currently being designed. These instruments will replace the current practice of manual sample preparation and mass analysis, although the protein identification algorithms will remain the same. The fully automated instruments will probably perform best on data derived from species with known genomes.

4.6 Alternate Splicing

When a eukaryotic organism translates its DNA into RNA in the nucleus (the primary transcript), the transfer RNA is usually edited before it is exported out of the nucleus as transfer RNA for transcription into a peptide chain. This editing process, generally referred to as RNA splicing, involves the removal of portions of the RNA that do not code for peptide sequence (exons), leaving the portions that do code for sequence and transcription regulatory functions (introns). In multicellular organisms with differentiated cell and tissue types – which includes all animals and plants – different exons can be spliced into the transfer-RNA in different cell types, resulting in different protein sequences that originate from the same gene. These different proteins that originate from the same gene are called ‘alternate splices’. The regions of genomic DNA that will be deleted or included can be predicted with some accuracy for the most likely transfer-RNA product; however, the alternate forms cannot be predicted in advance and they must be discovered experimentally.

Protein identification-type experiments are ideally suited to the rapid identification of alternately spliced proteins. In order to distinguish alternate splicing from proteolytic processing, the existing generation of protein recognition algorithms will need to include a method for searching and scoring multiple gaps using the genomic sequence as a starting point. By using predicted exon/intron divisions, it should be possible to search the possible DNA-to-RNA translation sequences to determine whether an alternate splice form is present in a particular differential display. Such a search is beyond the capabilities of the current generation of software: they all require an accurate RNA translation. However, with the introduction of AI-type training capabilities, it should be possible to apply the most sophisticated of the current algorithms to this problem.

5 MASS SPECTROMETRY APPLICATIONS OF PRINCIPAL COMPONENT AND FACTOR ANALYSES

5.1 Introduction

PCA calculates an orthogonal basis (i.e. coordinate system) for sets of mass spectra for which each axis maximizes the variation of the spectral dataset. Each axis is represented as a vector that relates the linear dependence of the mass spectral features (i.e. m/z variables). Typically, the new coordinate system has a reduced dimensionality. The PCA procedure allows the scientist to compress the data, remove noise, and discover the underlying or latent linear structure of the data.

FA rotates the principal components away from directions that maximize variance towards new chemically relevant directions; it allows scientists to resolve underlying pure components in mixtures, build classification models, and determine mass spectral features that relate to specific properties such as concentration or class.

5.2 Selected History

When computers were interfaced with mass spectrometers, numerical calculations could be used to simplify the data. A brief and somewhat selective history follows. The PCA technique was developed for the study of psychological measurements that are inherently complicated by many difficult-to-control factors.[29] These factors can be attributed to the different environmental, behavioral, or genetic influence on the human subjects who are evaluated. Some method was needed that would determine which factors were important and which factors were correlated.

The earliest applications of PCA in analytical chemistry determined the number of underlying components in mixtures. Specifically, for optical measurements, a mixture could be effectively modeled by a linear combination of the spectra of the pure components. The number of pure components of the mixture would correspond to the rank of the data matrix. The rank of a matrix of optical spectra of mixtures was computed using Gaussian elimination.[30,31] The application of FA to solving problems in chemical analysis was pioneered by Malinowski et al.[32,33]

The applications of PCA and FA to gas chromatography (GC) and MS first occurred in the 1970s. Initially, FA was employed to study the relationships between chemical structure and GC retention indices.[34–37] Then PCA was demonstrated as a tool for deconvolving overlapping GC peaks.[38] Next, FA was applied to 22 isomers of alkyl benzenes to assist the interpretation of fragmentation pathways and as a method for compressing the mass spectra to lower dimensionality.[39,40] The FA method was used for classifying mass spectra.[41]

The coupling of multichannel detection, specifically MS to GC, allowed PCA and FA to resolve overlapping components of GC/MS peaks.[42,43] The target transform FA method was automated for GC/MS analysis.[44]

FA was initially applied to solving problems of overlapping peaks in GC/MS. Soon it was realized that FA was a useful tool for the analysis of complex mixtures such as biological (bacteria, proteins, and hair) and geological (coal, atmospheric particles, and kerogen) samples. These complex samples were all amenable to pyrolysis mass spectrometry (PyMS).[45] The discriminant and FA were applied to various biological samples.[46] An unsupervised graphical rotation method was developed and applied to geological samples.[47] Canonical variates analysis (CVA)[48] was used to take advantage of measurement errors furnished by replicate spectra and was combined with rotation for mixtures of glycogen, dextran, and bovine serum albumin,[49] and has become one of the methods of choice for the analysis of MS fingerprints from bacteria.[50] The FA method was demonstrated as an effective tool for analysis of smoke particles by PyMS.[49] A related method that exploits PCA for classification is soft independent modeling for class analogies (SIMCA).[51]

Other techniques that benefited from FA and PCA are laser ionization mass spectrometry (LI/MS),[52] fast atom bombardment mass spectrometry (FAB/MS),[53] electrospray MS,[54] and secondary ion mass spectrometry (SIMS).[55] In the SIMS work, cluster analysis was used to help align high-resolution mass measurements into optimized columns of the data matrix, which was evaluated using PCA.

Table 1 The number of hydrocarbon spectra in the data set with respect to class and carbon number

Hydrocarbon class	Carbon number					Total
	4	5	6	7	10	
Diene	16	40	52	56	33	197
Alkene	12	17	60	61	37	187
Alkane	8	14	28	31	62	143
Total	36	71	140	148	132	527

5.3 Introductory Example

A brief demonstration of PCA and FA is presented with accompanying graphs. A data set of mass spectra was obtained from the Wiley Registry of Mass Spectra, 5th edition,[245] that comprised spectra of hydrocarbons that were alkane, alkene, or diene. This data matrix is exemplary because the MS fragmentation patterns are easy to predict. These data were part of a larger project that built classification models for identifying spectra of plastic recycling products.[56] The data matrix was composed of 527 spectra and 95 columns that correspond to m/z values. The m/z values ranged from 50 to 144. Typically, if all the spectra have no mass peaks at a specified m/z, this column is excluded from the data matrix. Table 1 gives the design of the hydrocarbon data set.

The principal components were calculated by singular value decomposition (SVD)[57] in a Win32 program that was written in C++. The analysis of these data required less than 5 s on a 300 MHz PC computer with 128 MB of random access memory and operating under Windows 98 in console mode.

The spectra were preprocessed by normalizing to unit vector length and centering the spectra about their mean spectrum before the PCA. Figure 1 gives the eigenvalues with respect to the component number. The eigenvalues measure the variance spanned by each eigenvector. For intricate data sets, the eigenvalues typically asymptotically approach zero. The relative variance of each eigenvalue is calculated by dividing the eigenvalue by the total variance of the data matrix. The total variance is easily obtained as the sum of the eigenvalues. From this calculation, the first two principal components account for approximately half the variance in this data set.

Examination of the mass spectral scores on the first two components in Figure 2 shows that the spectra tend to cluster by class (i.e. degree of unsaturation). The first component has the largest range of values and is responsible for separating the spectra in order of diene, alkene, and alkane. This component can be investigated further using the variable loadings in Figure 3. This graph shows the principal component plotted with respect to m/z, so that key spectral features may be investigated.

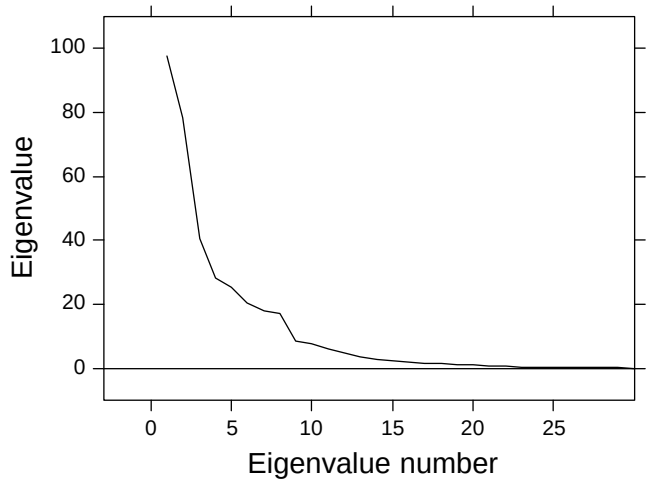

Figure 1 Eigenvalues plotted as function of the number of components for a set of 527 mass spectra with 95 variables.

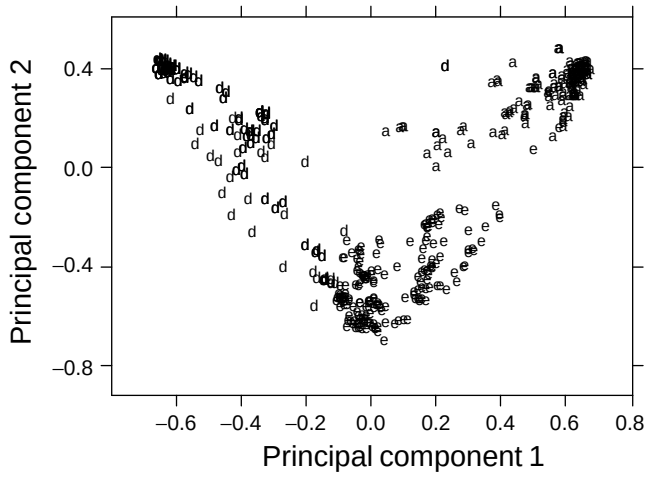

Figure 2 Observation scores of hydrocarbon mass spectra on the first two principal components, 47% of the cumulative variance: a, alkanes; d, dienes; e, alkenes.

The principal components point in mathematically, but not necessarily chemically, relevant directions. Target transform FA was used to rotate 13 principal components that spanned 95% of the variance in directions that correlate with the specific structural classes of the spectra. Figures 4–6 give the rotated factors for the diene, alkene, and alkane classes. Notice that the periodicity of the fragmentation pattern is precisely as one would expect for these sets of data. The alkenes follow a pattern of carbon number times 14, the dienes follow the same pattern except shifted to two less, and the alkanes shifted by two more. The shifts account for the change in mass of the molecule by the loss of two hydrogen atoms for each degree of unsaturation.

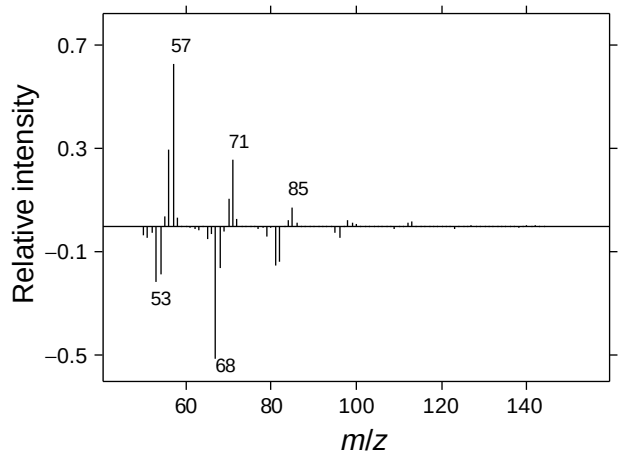

Figure 3 Variable loadings for the first principal component of the mass spectra dataset, 21% of the cumulative variance.

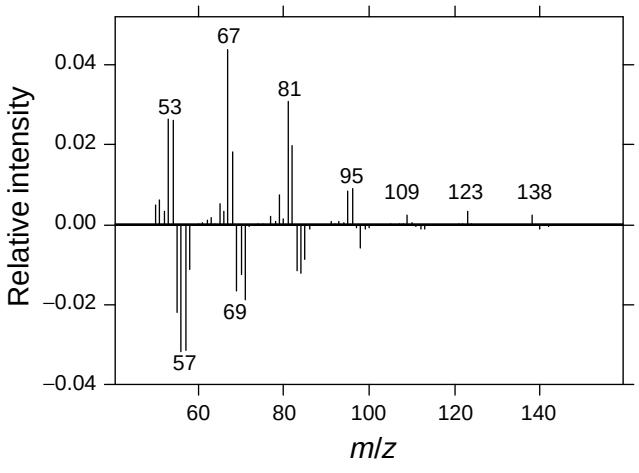

Figure 4 The target-transformed factor for dienes obtained from a set of 13 principal components that spanned 95% of the cumulative variance.

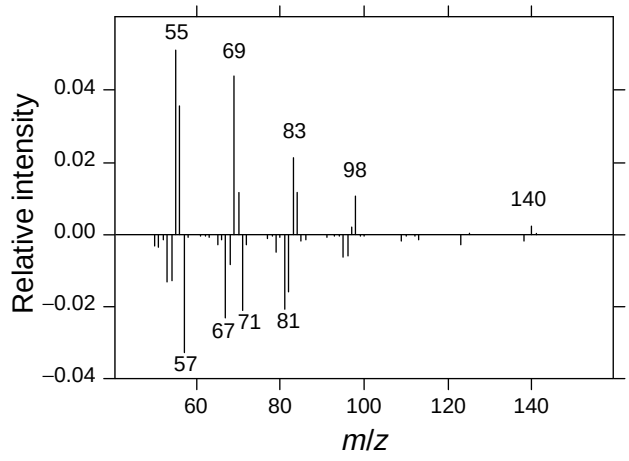

Figure 5 The target-transformed factor for alkenes obtained from a set of 13 principal components that spanned 95% of the cumulative variance.

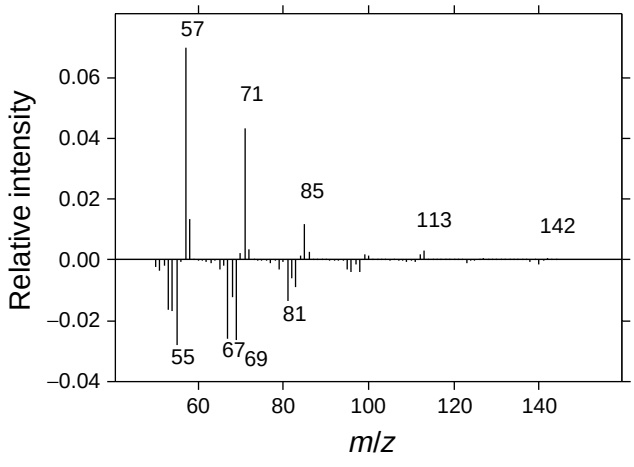

Figure 6 The target-transformed factor for alkanes obtained from a set of 13 principal components that spanned 95% of the cumulative variance.

5.4 Theoretical Basis

5.4.1 Principal Component Analysis

Typically, data are arranged into a matrix format so that each row corresponds to a mass spectrum and each column to a measurement at a specific m/z value. This matrix is designated as **D**. The PCA method is mathematically based on eigenvectors or eigenanalysis.

The method decomposes a set of data into two sets of matrices. The matrices are special in that the columns point in directions of major sources of variation of the data matrix. These vectors are eigenvectors. (*Eigen* is the German word for characteristic.) Because these vectors already point in a direction inherent to the data matrix, they will not change direction when multiplied by the data matrix. This property is referred to as the eigenvector relationship and is defined as Equations (1) and (2):

$$\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{1}$$

$$\mathbf{D}\mathbf{D}^{\mathrm{T}}\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{2}$$

where $\mathbf{D}^{\mathrm{T}}\mathbf{D}$ is a square symmetric matrix that characterizes the covariance of the columns of the data set **D**, $\mathbf{v}_i$ is eigenvector i that is in the row-space of **D**, and λ_i is eigenvalue i. In Equation (2), $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ is a square symmetric matrix that characterizes the covariance of the rows of the data set and $\mathbf{u}_i$ is in the column-space of the **D**.

Besides maximizing the variance, the sets of eigenvectors form orthogonal bases. This property can be expressed by Equations (3) and (4),

$$\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I} \tag{3}$$

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I} \tag{4}$$

for which **V** is a matrix comprising row-space eigenvectors ($\mathbf{v}_i$) and **U** is a matrix comprising column-space eigenvectors ($\mathbf{u}_i$). The identity matrix **I**, comprises values of unity along the diagonal and values of zero for all other matrix elements. The relationship given in Equations (3) and (4) is important because it shows that the transpose of an orthogonal matrix is equal to its inverse.

For large sets of data, computing the covariance matrix is time-consuming. A method that is precise and fast for computing both sets of eigenvectors is SVD:[58]

$$\mathbf{D} = \mathbf{USV}^{\mathrm{T}} \tag{5}$$

From Equation (5) **D** can be decomposed into the two matrices of eigenvectors and a diagonal matrix **S** of singular values (Equation 6):

$$\lambda_i = s_i^2 \tag{6}$$

The singular values w_i are equal to the square root of the eigenvalues, which leads to another important property that is given by Equation (7):

$$\mathbf{D}^n = \mathbf{US}^n\mathbf{V}^{\mathrm{T}} \tag{7}$$

This relationship is important because any power of **D** can be calculated by decomposing the matrix, raising the diagonal matrix of singular values to the *n*th power, and reconstructing the matrix. A useful power is negative unity, because $\mathbf{D}^{-1}$ can be used for calculating calibration models. Furthermore, pseudoinverses can be calculated from singular or ill-conditioned data matrices by reconstructing using only the components (i.e. vectors) that correspond to singular values above a threshold.

The other important element of PCA is the concept of a principal component. Because the row-space and column-space eigenvectors are orthogonal and hence independent, the number of eigenvectors will always be less than the dimensionality (i.e. minimum of the number of rows or columns) of **D**. The number of nonzero eigen or singular values gives the mathematical rank *r* of **D**. The rank gives the number of underlying linear components of a matrix. However, besides mathematical rank, there are physical and chemical ranks. The physical rank gives all the sources of variances that are associated with the physics of obtaining the measurement including noise. These variances may correspond to background or instrumental components. The chemical rank corresponds to variances related to the chemical components of interest. Therefore, the mathematical rank is the number of components with eigenvalues greater than zero. The physical rank corresponds to eigenvalues greater than a threshold that characterizes the indeterminate error of making the measurement. The chemical rank is typically the smallest and corresponds to the number of chemical components, when the variances of the data follow a linear model.

Typically, the components that are less than either the physical or chemical ranks are referred to as principal components. The components that correspond to the smaller eigenvalues are referred to as secondary components. Secondary components usually characterize noise or undesirable variances in the data. The determination of the correct number of principal components *r* is important. If the number of principal components is too small then characteristic variances will be removed from the data. If the number of principal components is too large then noise will be embedded in the components as well as signal. There are several methods to evaluate the calculation of the correct number of principal components.

One of the simplest methods is to reconstruct the data **D** using subsets of the eigenvectors. When the reconstructed data resemble the original data within the precision of the measurement, then the proper number of principal components has been obtained. An empirical approach determines the minimum of the indicator function (IND), which is not well understood, but furnishes reliable estimates of the chemical rank.[59]

There are three key pieces of information furnished by PCA. The first is the relationship of the variance that is spanned by each component. Plotting the eigenvalues as a function of components, gives information regarding the distribution of information in the data. The eigenvalues also convey information regarding the condition number of the data matrix. The condition number is obtained by dividing the largest eigenvalue by the smallest. This condition number can be used to assess the error bounds on a regression model[60] and as a means to evaluate selectivity.[61]

This approach was what made PCA useful for assessing the number of analytical components contained in a GC peak. This methodology is still used; however, it is referred to as window or evolving factor analysis (EFA). Instead of processing the spectra contained in a chromatographic peak, a window (i.e. a user-defined dataset) can be moved along the chromatogram. The chemical rank is evaluated and gives the number of chemical components in the window.

The second piece of information is furnished by the observation scores. Score plots display the distribution of spectra or rows of **D** in a lower dimension graph. The scores of the first two components provide a two-dimensional window that maximizes information content of the spectra. If the rows are ordered with respect to time, the observation scores give trajectories of the changes that occur in the data over time (Equation 8):

$$\mathbf{o}_i = \mathbf{d}_i\mathbf{V} = \mathbf{u}_i\mathbf{S} \tag{8}$$

for which $\mathbf{o}_i$ is a row vector of the *i*th observation score of spectrum *i* ($\mathbf{d}_i$). This may be calculated by multiplying

a spectrum or the *i*th row of **D** by the matrix of principal components. The observation scores can be calculated directly for the results of SVD by multiplying the matrix of singular values **W** by the *i*th row of the column-space eigenvectors **U**. Plots of the observation scores are also referred to as the Karhunen–Loève plots. These plots allow clustering of the data to be visualized.

The final piece of information is yielded by the variable loadings, which indicate the direction that the row-space eigenvectors point. The variable loadings show the importance of each variable for a given principal component. Plots of variable loadings can be examined for characteristic spectral features. They also are used together with the observation score plots to see which spectral features are responsible for separating objects in the score plots.

In some instances, the data matrix **D** can be modified so that the principal components point in directions that are more chemically relevant. These modifications to **D** are referred to collectively as preprocessing. Typically, the spectra are mean-centered, which refers to subtracting the average spectrum from each spectrum in the dataset. This centers the distribution of spectra about the origin. If the data are not mean-centered, the first principal component will span the variance characterized by the overall distance of the data set from the origin.

In some cases, the spectra are normalized so as to remove any variations related to concentrations. Normalization scales the rows of **D**, so that each row is weighted equally. Mathematically, normalizing the spectra to unit vector length will achieve this equalized weighting. For spectra that vary linearly with concentration, the concentration information is manifested in the vector length of the spectrum. Other methods of normalization include normalizing to a constant base peak intensity (i.e. maximum peak of unity) or to a constant integrated intensity (i.e. sum of peaks of unity).

The data may be scaled so that the variables or columns of **D** are weighted equally. Scaling is important for mass spectra, because peaks of higher mass that tend to convey more information, have smaller intensities, and tend to be less influential in the calculation of principal components.

Autoscaling gives each variable or column of data equal weight. This method of scaling is useful when the noise or the signals are not uniformly distributed across the mass range. For this method of preprocessing, each column of **D** is divided by its standard deviation. The problem with autoscaling is that variables that convey noise only are given equal weight with those that convey signal. A better approach is to scale the data by the experimental errors for each variable. Experimental error can be measured as the standard deviation of replicate spectra. The variances of these standard deviations can be added for different samples to calculate an estimate of the experimental error. The experimental error avoids the diminution of the signals during scaling.

An alternative to scaling is transformation. In some cases the data may be converted to the square root or logarithm to give greater weight to smaller features. A useful method for preprocessing mass spectra is through modulo compression.[62]

5.4.2 *Canonical Variates Analysis*

For supervised classification, a useful method related to PCA is CVA,[63] which is also applied with discriminant (function) analysis.[64] The CVA method is not usually performed on the original feature space (mass spectra) because the mass spectra have colinear variables or too many variables for CVA. This problem may be resolved by compressing the data, such as using principal component scores[52] or by calculating the pseudo-inverse of the covariance matrix.[65] The canonical variates (CVs) are principal components that are calculated from a matrix that is related to Fisher variance and analysis of variance. In the traditional method, two covariance matrices are calculated. The first matrix characterizes the covariance of the class means about the grand mean of the data. The second matrix characterizes the variation of the spectra about their class means.

The CVA approach uses PCA twice in the calculation. First, SVD is used to compute the pseudo-inverse of the within-groups sum of squares matrix ($\mathbf{SS}_{\mathrm{w}}^{+}$). The CVs are the variable loadings obtained from PCA applied to **R**, which is obtained by Equation (9):

$$\mathbf{R} = \mathbf{SS}_{\mathrm{b}}\,\mathbf{SS}_{\mathrm{w}}^{+} \tag{9}$$

for which $\mathbf{S}_{\mathrm{b}}$ is the between-class sum of squares matrix and $\mathbf{S}_{\mathrm{w}}^{+}$ is the pseudo-inverse of the within-class sum of squares matrix ($\mathbf{S}_{\mathrm{w}}$). These are calculated by Equations (10) and (11),

$$\mathbf{SS}_{\mathrm{b}} = \sum_{i=1}^{N_{\mathrm{c}}} N_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^{\mathrm{T}} \tag{10}$$

$$\mathbf{SS}_{\mathrm{w}} = \sum_{i=1}^{N_{\mathrm{c}}} \sum_{j=1}^{N_i} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ji} - \bar{\mathbf{x}}_i)^{\mathrm{T}} \tag{11}$$

for which N_{c} is the number of classes, N_i is the number of spectra in the *i*th class, $\bar{\mathbf{x}}_i$ is the class mean, and the $\bar{\mathbf{x}}$ is the global mean. The rank of **R** will be equal to the number of classes less one (e.g. $N_{\mathrm{c}} - 1$), because a degree of freedom is lost by centering the spectra about the global mean and the product of two matrices can not exceed the minimum rank of the two multiplier matrices.[66] The CVs are a basis set of orthogonal vectors that maximize the separations of the classes (i.e. maximize the distance among the means

and minimize the distance of the spectra from their class means). Thus the principle of CVA is similar to PCA but, because the objective of CVA is to maximize the ratio of the between-group to within-group variance, a plot of the first two CVs displays the best two-dimensional representation of the class separation.

5.4.3 Factor Analysis

PCA yields variable loadings that are characteristic for the data matrix. The variable loadings are meaningful with respect to maximizing variances. For other applications it is useful to rotate these loadings in other directions that pertain to specific problems. Once the principal components are rotated, the technique is referred to as FA. Rotations are either oblique or orthogonal. The orthogonal rotations maintain the linear independence of the principal components and the basis. Oblique rotations remove the constraint of linear independence and therefore model more closely physical and chemical phenomena. These methods calculate a square matrix $\mathbf{T}$ of coefficients that rotate the components with a dimensionality of r. For which r is the number of principal components. Typically, the column-space components or observation scores are rotated in the forward direction and the row-space components or variable loadings are rotated in the reverse direction using $\mathbf{T}^{-1}$. The rotation matrices can be computed by numerical optimization of a variety of objective functions or they can be rotated graphically until they resemble a target. For orthogonal rotation, the most popular objective function is Varimax.[67] This rotation method seeks to increase the magnitude of the observation scores on a single component and reduce the scores magnitude on all other components.

Target transformation calculates a transformation matrix that rotates the row-space and column-space eigenvectors or components in directions that agree with a target vector. Typically, the targets are a set of properties that may correlate with the objects, and the transformation matrix is calculated by regression. These transformation matrices may be calculated using the eigenvectors from SVD (Equations 12 and 13),

$$\mathbf{T} = \mathbf{U}^{\mathrm{T}}\mathbf{X} \tag{12}$$

$$\widehat{\mathbf{X}} = \mathbf{UT} \tag{13}$$

for which $\mathbf{X}$ is composed of columns of targets, $\mathbf{T}$ is the transformation matrix, and $\widehat{\mathbf{X}}$ is the estimated target matrix. The loadings can be rotated by regressing the matrix of variable loadings $\mathbf{V}$ onto the target matrix $\mathbf{T}$ that has r rows and the number of columns equals the number of target vectors (Equation 14):

$$\widehat{\mathbf{Y}} = \mathbf{VT}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1} \tag{14}$$

In some cases, it is advised to use the pseudo-inverse of $\mathbf{T}$, because the inner product of $\mathbf{T}$ may be ill-conditioned or singular. The factor variable loadings for the targets are estimated by $\widehat{\mathbf{Y}}$.

5.5 Related Methods and Future Applications

5.5.1 Calibration

There are various methods to exploit the properties of eigenvectors to accomplish calibration. Calibration furnishes models that predict properties from data such as mass spectra. The most common use for calibration is to construct models that estimate the concentration of components in complex samples by their mass spectra.

Principal component regression (PCR) uses the observation scores for computing the regression model. The advantage of this approach is that for MS data in many cases $\mathbf{D}$ is underdetermined (i.e. more m/z measurements than spectra). Because the observation scores will equal the chemical rank, the number of variables are reduced and regression by inverse least squares becomes possible.

A related method uses SVD to calculate the pseudoinverse $\mathbf{D}^{+}$. The SVD regression is computationally more efficient than PCR, but is mathematically equivalent. A very effective method for many problems is partial least squares (PLS). This calculates common column-space eigenvectors between the independent block (i.e. $\mathbf{D}$) and dependent block (i.e. $\mathbf{Y}$) of data. The PLS method was initially developed in the field of econometrics. Both PLS and PCA are described in a tutorial;[68] PLS has been enhanced to handle multiway or higher-order data.[69]

Quantitative analysis of complex binary and tertiary biochemical mixtures analyzed with PyMS[70] showed that, of the latent variable PCR and PLS methods, the best technique was PLS, a finding to be found generally by other studies.[71,72]

5.5.2 Multivariate Curve Resolution

The same FA methods that were initially applied to peaks of GC/MS data have evolved so that they can be applied to the entire chromatographic runs. These methods start with a set of principal components. The components are rotated by a method known as alternating least squares (ALS). The key is to apply mathematical constraints such as non-negativity (no negative peaks) and unimodality (a spectrum will appear in only one peak of a chromatogram).

Curve resolution provides a means to enhance the spatial or depth resolution of ion measurements of surfaces or could be exploited to examine changes in electrospray mass spectra as a function of changing solvent conditions. Curve resolution will continue to exploit PCA and FA

to detect impure chromatographic peaks and mathematically resolve the overlapping components.

EFA and window factor analysis (WFA) use the eigenvalues to model the change in concentrations of components in the data matrices. The eigenvalues can be combined to form initial concentration profiles that are regressed onto the data. The concentration profiles and extracted spectra are refined using ALS with constraints.

5.5.3 Multiway Analysis

The entire chromatographic mass spectral data matrix **D** is only the beginning. If several chromatographic runs are used to characterize a chemical process or if multidimensional MS matrices of data are collected, a tensor or cube of data would be obtained. Using methods based on the Tucker model,[73] the higher-order sets of data can be decomposed into vectors or planes of principal components. A method related to the Tucker model is PARAFAC.

5.6 Reviews and Tutorials

Malinowski's monograph is an excellent resource for PCA and FA.[74] Tutorials on FA and related methods can be found in the literature – the philosophical basis of PCA and FA,[75] EFA,[76] and target transform FA.[77] Multivariate curve resolution applied to chromatography with multichannel detection has been published as a tutorial[78] and reviewed specifically for GC/MS.[79] Tutorials of the multiway PCA method PARAFAC[80] and PLS[68] are also useful entry points into these methods. The text by Martens and Næs on multivariate calibration thoroughly describes PLS.[81]

5.7 Acknowledgments

Tricia Buxton, Guoxiang Chen, and Aaron Urbas are thanked for their help with preparing this section. Thomas Isenhour and Kent Voorhees are thanked for their help with searching the literature. The introductory example data set was initially prepared by Peter Tandler.

6 ARTIFICIAL NEURAL NETWORKS

6.1 Summary

The availability of powerful desktop computers in conjunction with the development of several user-friendly packages that can simulate ANNs has led to the increase in adoption of these 'intelligent' systems by the analytical scientist for pattern recognition. The nature, properties and exploitation of ANNs with particular reference to MS is reviewed.

6.2 Introduction to Multivariate Data

Multivariate data consist of the results of observations of many different characters (variables) for a number of individuals (objects).[82,83] Each variable may be regarded as constituting a different dimension, such that if there are n variables each object may be said to reside at a unique position in an abstract entity, referred to as n-dimensional hyperspace. In the case of MS, these variables are represented by the intensities of particular mass ions. This hyperspace is necessarily difficult to visualize, and the underlying theme of multivariate analysis (MVA) is thus simplification[84] or dimensionality reduction, which usually means that we want to summarize a large body of data by means of relatively few parameters, preferably the two or three that lend themselves to graphical display, with minimal loss of information.

6.3 Supervised Versus Unsupervised Learning

Conventionally the reduction of the multivariate data generated by MS[85–87] has normally been carried out using PCA;[84,88–90] the PCA technique is well-known for reducing the dimensionality of multivariate data while preserving most of the variance, and the principal component scores can easily be plotted and clusters in the data visualized.

Analyses of this type fall into the category of unsupervised learning (Figure 7a), in which the relevant multivariate algorithms seek clusters in the data.[90] Provided that the data set contains standards – of known origin and relevant to the analyses – it is evident that one can establish the closeness of any unknown samples to a standard, and thus effect the identification of the former. This technique is termed 'operational fingerprinting' by Meuzelaar et al.[91]

Such methods, although in some sense quantitative, are better seen as qualitative as their chief purpose is merely to distinguish objects or populations. More recently, a variety of related but much more powerful methods, which are most often referred to within the framework of chemometrics, have been applied to supervised analysis of multivariate data (Figure 7b). In these methods, one seeks to relate the multivariate MS inputs to the concentrations of target determinands, i.e. to generate a quantitative analysis, essentially via suitable types of multidimensional curve fitting or linear regression analysis.[83,92–96] Although nonlinear versions of these techniques are increasingly available,[97–103] the usual implementations of these methods are linear in scope. However, a related approach to chemometrics, which is inherently nonlinear, is the use of ANNs.

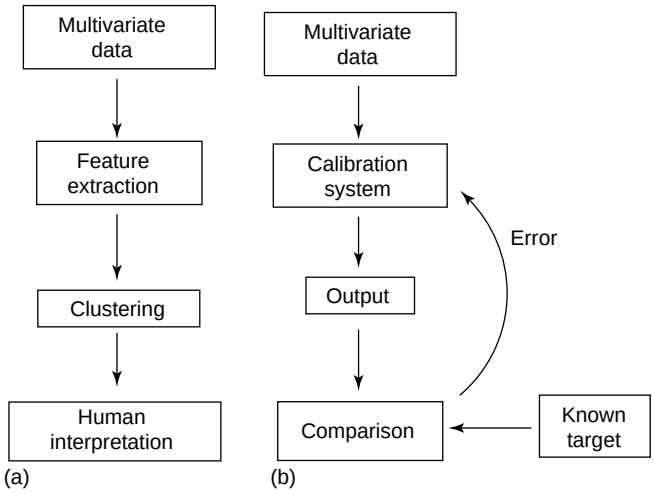

Figure 7 (a) Unsupervised learning – when learning is unsupervised, the system is shown a set of inputs (multivariate MS data) and then left to cluster them into groups. For MVA this optimization procedure is usually simplification or dimensionality reduction; this means that a large body of data (the inputs) are summarized by means of a few parameters with minimal loss of information. After clustering the results then have to be interpreted. (b) Supervised learning – when the desired responses (targets) associated with each of the inputs (multivariate data) are known then the system may be supervised. The goal of supervised learning is to find a model that will correctly associate the inputs with the targets; this is usually achieved by minimizing the error between the known target and the model's response (output).

6.4 Biological Inspiration

ANNs are biologically inspired; they are composed of processing units that act in a manner that is analogous to the basic function of the biological neuron (Figure 8). In essence, the functionality of the biological neuron consists of receiving signals, or stimuli, from other cells at their synapses, processing this information, and deciding (usually on a threshold basis) whether or not to produce a response, that is passed onto other cells. In ANNs these neurons are replaced with very simple computational units which can take a numerical input and transform it (usually via summation) into an output. These processing units are then organized in a way that models the organization of the biological neural network, the brain.

Despite the rather superficial resemblance between the ANN and biological neural network, ANNs do exhibit a surprising number of the brain's characteristics. For example, they learn from experience, generalize from previous examples to new ones, abstract essential characteristics from inputs containing irrelevant data, and make errors (although this is usually because of badly chosen training data.[83,104–109]) All these traits are considered more characteristic of human thought than of serial processing by computers. These systems offer the

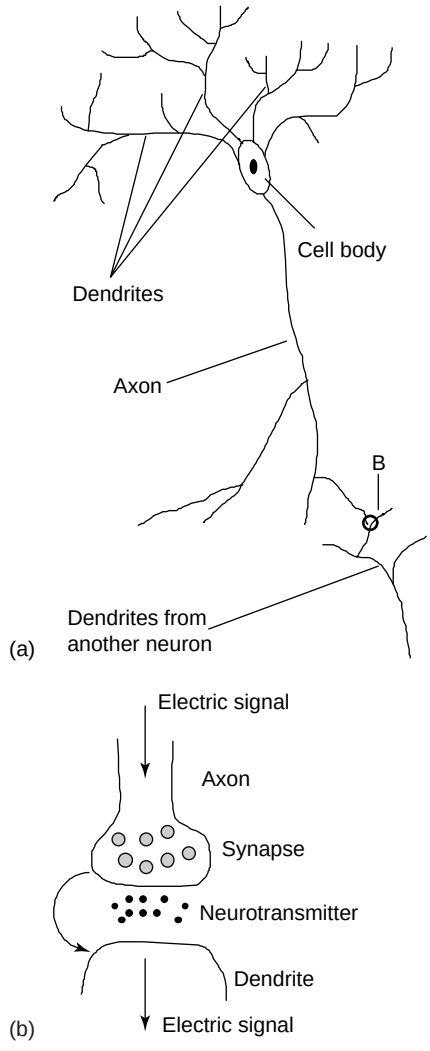

Figure 8 (a) A very simplified scheme of a biological neuron. Dendrites are filaments that extend from the cell body to other neurons where they receive signals at connection points called synapses, these dendrites then provide inputs to the cell body, the axon sends output signals. (b) The synapse is a special connection which can be strengthened or weakened to allow more, or less, signal through.

mass spectrometrist the capability of performing pattern recognition on very complex uninterpretable (at least to the naked eye) multivariate data.

For a MS analytical system, there are some mass spectra that have desired responses, which are known (i.e. the concentration of target determinands). These two types of data (the representation of the objects and their responses in the system) form pairs, which are usually called inputs and targets. The goal of supervised learning is to find a model or mapping that will correctly associate the inputs with the targets (Figure 7).

Thus the basic idea in these supervised learning neural network techniques is that there are minimally four

datasets to be studied, as follows. The training data consist of (a) a matrix of s rows and n columns in which s is the number of objects and n the number of variables (these are the normalized ion intensities at a particular mass-to-charge ratio for MS), and (b) a second matrix of target data, again consisting of s rows and typically 1 or two columns, in which the columns represent the variable(s) whose value(s) it is desired to know (these are the result(s) wanted) and which for the training set have actually been determined by some existing, benchmark method. This variable may be the concentration of a target determinand, and is always paired with the patterns in the same row in (a). The test data also consist of two matrices, (c) and (d), corresponding to those in (a) and (b) above, but the test set contains different objects. As the name suggests, this second pair is used to test the accuracy of the system; alternatively they may be used to cross-validate the model. That is to say, after construction of the model using the training set (a, b) the test data (c) (these may be new spectra) are then passed through the calibration model so as to obtain the model's prediction of results. These may then be compared with the known, expected responses (d).

6.5 Data Selection

As in all other data analysis techniques, these supervised learning methods are not immune from sensitivity to badly chosen initial data,[106] and so good modelling practice must be adopted.[109] Therefore, the exemplars for the training set must be carefully chosen; the adage is 'garbage in – garbage out'. It is known[83,104,105,107–109] that neural networks (and other supervised learning methods such as PLS) can over-fit data. For example, an over-trained neural network has usually learnt perfectly the stimulus patterns it has seen but cannot give accurate predictions for unseen stimuli, i.e. it is no longer able to generalize. For supervised learning methods accurately to learn and predict the concentrations of determinands in biological systems, or to identify new observations as being from something previously seen, the model must obviously be calibrated to the correct point. The reality is that in extension to normal chemometric practices detailed above the data should be split into three sets: (a) data used to calibrate the model; (b) data employed to cross-validate the model; and (c) spectra whose determinand concentration, or identities, were unknown and used to test the calibrated system. During calibration, the models would be interrogated with both the training and the cross-validation set and the error between the output seen and that expected calculated, thus allowing two calibration curves for the training and cross-validation sets to be drawn. When the error on the cross-validation data was lowest, the system will be deemed to have reached the best generalization point and then may be challenged with input stimuli whose determinand concentrations, or identities, are really unknown.

An alternative approach is to determine an acceptable error for the model, and construct ANN models that fit within this error. For many classification problems a relative root mean square error of calibration (RRMSEC) of 10% is usually sufficient (Equation 15):

$$\mathrm{RRMSEC} = \sqrt{\frac{\sum_{i=1}^{s}\sum_{j=1}^{p}(y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^{s}\sum_{j=1}^{p}(y_{ij} - \bar{y}_j)^2}} \qquad (15)$$

for which RRMSEC is the relative root mean square standard error of calibration, $\hat{y}_{ij}$ is the neural network output of unit j and training object $\boldsymbol{I}$, y_{ij} is the corresponding target value. There are p outputs for the neural network model and s training objects about the class mean $\hat{y}_i$.

Many neural network models are often overly optimized and do not generalize well, even though monitoring sets or cross-validation are used. The caveat is that the prediction and training sets must be well designed and representative of the specific problem. With poorly designed training and prediction sets the neural networks will model hidden experimental factors that correlate with the target properties. A second problem occurs when the prediction set is used to configure the network – if the prediction set matches the training set too well the ANN model will overtrain, and if the prediction set is too dissimilar the network will undertrain.

Latin-partitioning is a useful experimental design tool for evaluating neural network models, the method constructs training and prediction set pairs for which each target object in the data is present once and only once in the prediction sets.[110] The method randomly partitions the data so that each target pattern is represented in the sets with equal proportionality. This method is important because the composition of the prediction set is a major source of variation for evaluating neural network models and by including all objects in the prediction set it will not be biased.

For quantitative determinations it is also imperative that the objects fill the sample space. If a neural network is trained with samples in the concentration range from 0% to 50% it is unlikely to give accurate estimates for samples whose concentrations are greater than 50%; that is to say, the network is unable to extrapolate.[109] Furthermore for the network to provide good interpolation it needs to be

trained with a number of samples covering the desired concentration range.[111]

6.5.1 Sensitivity Analysis of Artificial Neural Network Models

Sensitivity analysis may be used to probe ANN models and can lead to an understanding as to why they do not predict or generalize well. Furthermore, interpreting the sensitivities may lead to an understanding of causal relationships between input and output variables. Kowalski and Faber proposed methods on the quantitative measurement of sensitivities.[112] Ebhart et al.[113] proposed three comparable methods to calculate the mean square sensitivity, absolute value average sensitivity, and maximum sensitivity. Similarly, Howes and Crook[114] studied the three types of input influence, namely general influence, specific influence, and potential influence, on the network output. Most of these studies considered the effects of weight matrix in multilayer perceptrons (MLPs) models. Choi and Choi[115] defined the sensitivity of input perturbations as the ratio of output error to the standard deviation of each input perturbation, which involves complex weight calculations. Kovalishyn et al.[116] have proposed several sensitivity measurements to be used with cascade-correlation networks (CCNs) for variable selection. The sensitivity was measured by connection weights, or the second derivative of the error function with respect to the neuron weight. It was shown in this paper, the sensitivities measured from their definition were not stable with the dynamic growth of the network, and also sensitive to addition of noise. In Ikonomopoulos' study of importance of input variables with the wavelet adaptive neural network, the sensitivity of input variables was estimated by the ratio of the standard deviations of the prediction and the altered input.[117] It was found that with this method, sensitivity measurements were highly correlated with the input perturbation. Sung derived the sensitivity coefficient matrix for a backpropagation (BP) neural network with two hidden layers.[118] Other sensitivity analysis methods based on the input magnitude and functional measurements have also been proposed.[119,120]

Because neural network models are fundamentally nonlinear, the sensitivity will depend on the input values from which they are calculated. Harrington et al.[121] proposed using partial derivatives of the neural network output with respect to the input. They compare the sensitivity of the average input for each class with the average sensitivity of each class. This method provides a method for detecting input variables that are modeled by higher-ordered functions in the neural network model, and provides a quantitative measure of the input variables contribution for each target output. Weight vectors were not directly involved in the sensitivity measurement.

6.6 Cluster Analyses with Artificial Neural Networks

These analyses fall into the category of unsupervised learning, in which the relevant multivariate algorithms seek clusters in the data.[90] Recently there has been an interest in the use of neural computation methods, which can perform exploratory data analyses on multivariate data, the most commonly used are feature or self-organizing maps (SOMs) and auto-associative artificial neural networks (AAANNs).

6.6.1 Self-organizing Maps

These provide an objective way of classifying data through self-organizing networks of artificial neurons.[122–124] These neural networks are also referred to as Kohonen ANNs, after their inventor.[125] The SOM algorithm is very powerful and is now extensively used for data mining, representation of multidimensional data and the analysis of relationships between variables.[126]

The SOMs used to analyze mass spectra typically consist of a two-dimensional network of neurons arranged on a rectangular grid;[127–129] although a variety of output arrays (Figure 9) and hence neighborhoods are possible.

Consider the situation where a square two-dimensional Kohonen output layer is used (Figure 9b). Each neuron

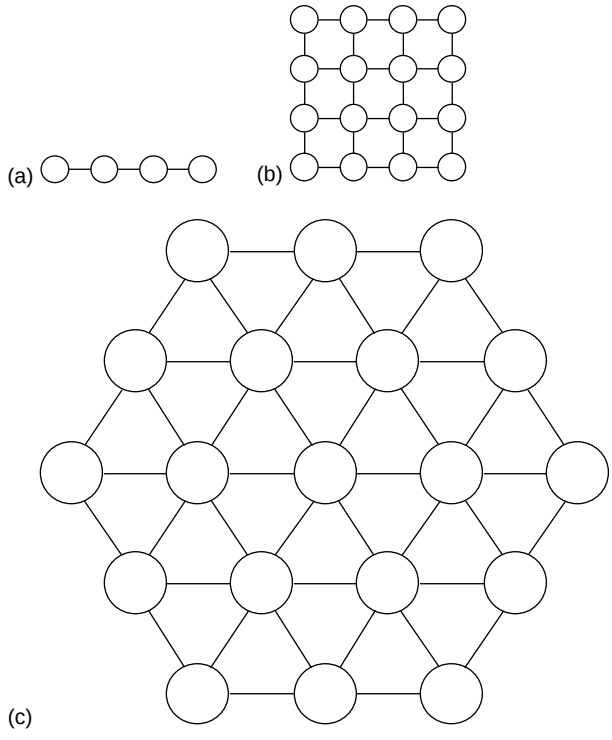

Figure 9 Commonly used SOM structures: (a) one-dimensional array, (b) two-dimensional rectangular network, (c) two-dimensional hexagonal network. The lines represent the neighborhoods.

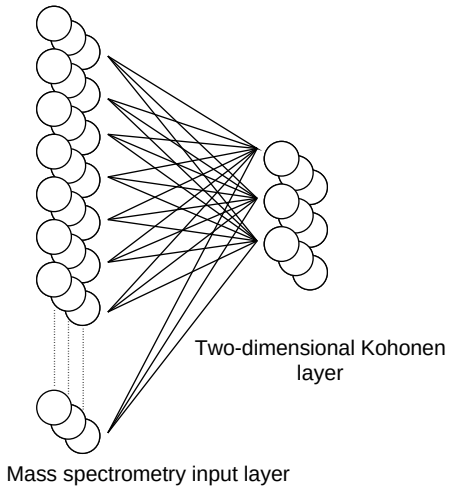

Figure 10 A simplified SOM. Nodes in the two-dimensional Kohonen layer are interconnected with each other (only a few are shown in this figure), such that an activation node also tends to activate surrounding nodes (Figure 3). The mass spectral data are applied to the input layer which activates a node or group of neighboring nodes in the Kohonen layer (represented here as having 3×3 nodes).

is connected to its eight nearest neighbors on the grid. The neurons store a set of weights (a weight vector), each of which corresponds to one of the inputs in the data. (Figure 10 shows a diagrammatic representation of a SOM). Thus, for mass spectral data consisting of 150 quantitative ion intensity measurements at particular m/z charges, each node stores 150 weights in its weight vector. Upon presentation of a mass spectrum (represented as a vector consisting of the 150 ion counts) to the network each neuron calculates its activation level. A node's activation level is defined as Equation (16):

$$\sqrt{\sum_{i=0}^{n}(\text{weight}_i - \text{input}_i)^2} \tag{16}$$

This is simply the Euclidean distance between the points represented by the weight vector and the input vector in n-dimensional space. Thus a node whose weight vector closely matches the input vector will have a small activation level, and a node whose weight vector is very different from the input vector will have a large activation level. The node in the network with the smallest activation level is deemed to be the winner for the current input vector.

During the training process the network is presented with each input pattern in turn, and all the nodes calculate their activation levels as described above. The winning node and some of the nodes around it are then allowed to adjust their weight vectors to match the current input vector more closely. The nodes included in the set, which are allowed to adjust their weights are said to belong to the neighborhood of the winner. The size of the winner's neighborhood is varied throughout the training process. Initially all of the nodes in the network are included in the neighborhood of the winner but, as training proceeds, the size of the neighborhood is decreased linearly after each presentation of the complete training set (all the mass spectra being analyzed), until it includes only the winner itself. The amount by which the nodes in the neighborhood are allowed to adjust their weights is also reduced linearly throughout the training period.

The factor, which governs the size of the weight alterations is known as the learning rate and is represented by α. The iterative adjustments to each item in the weight vector (where δw is the change in the weight) are made in accordance with Equation (17):

$$\delta w_i = -\alpha(w_i - i_i) \tag{17}$$

This is carried out for $i = 1$ to $i = n$, where in this case $n = 150$. The initial value for α is 1 and the final value is 0.

The effect of the learning rule (weight update algorithm) is to distribute the neurons evenly throughout the region of n-dimensional space populated by the training set.[122,123,125] This effect is displayed in Figure 11, which shows the distribution of a square network over an evenly populated two-dimensional square input space. The neuron with the weight vector closest to a given input pattern will win for that pattern and for any other input patterns that it is closest to. Input patterns that allow the same node to win are then deemed to be in the same group, and when a map of their relationship is drawn a line encloses them. By training with networks of increasing size, a map with several levels of groups or contours can be drawn. However, these contours may sometimes cross which

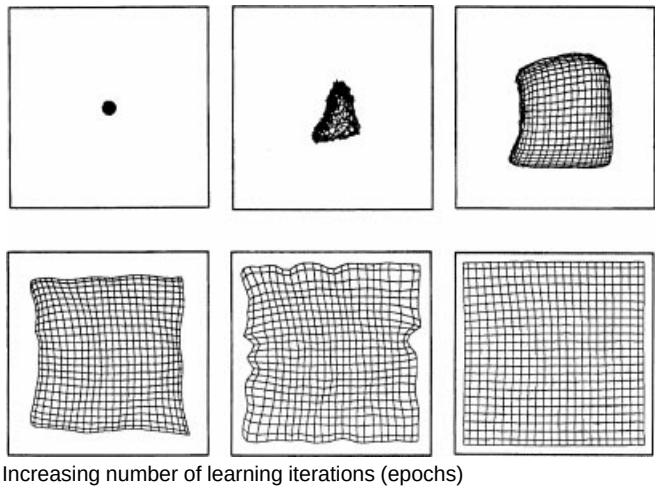

Figure 11 Representations of square networks distributed across an evenly distributed square input space.

appears to be due to failure of the SOM to converge to an even distribution of neurons over the input space.[130] Construction of these maps allows close examination of the relationships between the items in a training set.

A relatively recent paper by Belic and Gyergyek[131] compared several neural network methodologies for the recognition of MS data, based on simulated mass spectra samples and concluded that SOMs could be recommended for practical use in MS recognition. Somewhat earlier, however, Lohninger and Stancl[132] first showed that SOMs were better than *k*-nearest neighbor clustering for the discrimination of (real) mass spectra of steroids from eight distinct classes of chemical compounds. Goodacre et al. have also exploited SOMs successfully to carry out unsupervised learning on pyrolysis mass spectra, and hence the classification of canine *Propionibacterium acnes* isolates,[127] *P. acnes* isolated from man,[133] and plant seeds.[128]

6.6.2 Auto-associative Artificial Neural Networks

AAANNs are a neural network-based method again for unsupervised feature extraction and were pioneered by Kramer.[134,135] They consist of five layers containing processing nodes (neurons or units) made up of a layer of *x* input nodes, *x* output nodes (exactly the same as used in the input layer), and three hidden layers containing (in the example shown in Figure 12) 7, 3 and 7 nodes respectively; this may be represented as an *x*-7-3-7-*x* architecture. Adjacent layers of the network are fully interconnected, and the algorithm used to train these neural networks is the standard BP.[136–138] As these neural networks are auto-associative in nature, that is to say, during training the input and output layer are presented with identical multivariate data, a key feature of these networks is the data compression in the middle (third) bottle-neck layer of three nodes. The second and fourth layers each consist of nodes that map and demap the multivariate data, allowing feature extraction in the bottle-neck layer. Moreover this is equivalent to nonlinear PCA.[134,135] After training, each of the multivariate data used to train the AAANN is applied in turn to the input layer and the overall activation on the three nodes in the bottle-neck layer calculated. Plots of the activations of the nodes in the bottle-neck layer therefore allow clusters to be found in the data.

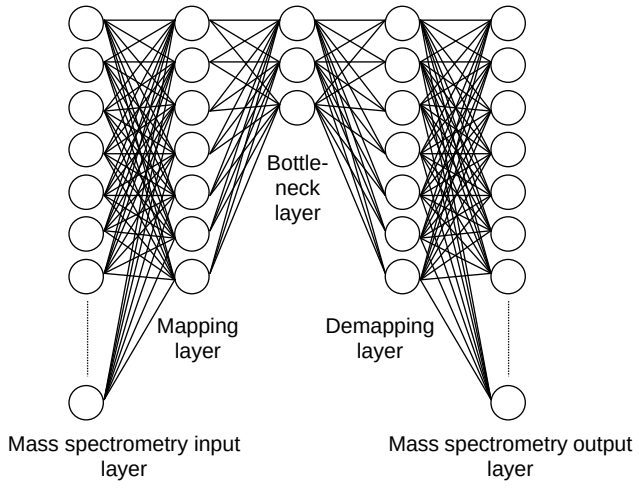

Figure 12 Architecture of an AAANN consisting of five layers. In the architecture shown, adjacent layers of the network are fully interconnected. The input and output layers are presented with identical mass spectral data. A key feature of the auto-associative network is the data compression in the middle (third) bottle-neck layer of three nodes. The second and fourth layers each consist of seven nodes and these map and demap the mass spectra allowing feature extraction in the bottle neck layer.

With reference to MS these AAANNs have only been applied to PyMS, to effect exploratory cluster analyses for the classification of plant seeds[128] and for the authentication of animal cell lines.[139] In the latter study this method of nonlinear PCA was particularly useful because the clusters observed were comparable with the groups obtained from the more conventional statistical approaches of hierarchical cluster analysis. This approach could detect the contamination of cell lines with low numbers of bacteria and fungi, and may plausibly be extended for the rapid detection of mycoplasma infection in animal cell lines.

Elsewhere, within spectroscopy, AAANNs have been used to reduce the dimensionality of the infrared (IR) spectra of polysaccharides and hence extract spectral features due to polysaccharides,[140] to detect plasmid instability using on-line measurements from an industrial fermentation producing a recombinant protein expressed by *Escherichia coli*,[141] and for knowledge extraction in chemical process control.[142] An optimal associative memory (OAM) was developed for removing backgrounds from mid-IR spectra.[143,144] The memory stores reference spectra and generates a best-fit reference spectrum when a sample IR scan is input. This method was extended to a fuzzy optimal associative memory (FOAM) by implementing fuzzy logic to near-IR spectra and was applied to calibration models of glucose in bovine plasma.[145]

6.7 Supervised Analysis with Artificial Neural Networks

As discussed above when the desired responses (targets) associated with each of the inputs (spectra) are known, the system is referred to as supervised. ANNs are very powerful at finding a mapping that will correctly associate mass spectra with known targets; these targets may be the identity of something, or be the quantitative amount of a substance. The two most exploited of the neural computational methods for these purposes are MLPs,

	Encoding in output layer				
	1	2	3	4	5
i	1	0	0	0	0
ii	0	1	0	0	0
iii	0	0	1	0	0
iv	0	0	0	1	0
v	0	0	0	0	1

Figure 13 Binary encoding the five nodes in the output layer on a MLP or radial basis function trained to classify one of five substances i–v.

using standard BP of error, and radial basis function neural networks (RBFs).

In MLPs and RBFs that are to be trained for identification purposes the training data used to calibrate the model (as detailed above) consist of (a) a matrix of s rows and n columns in which s is the number of objects and n the number of variables, and (b) a second matrix, again consisting of s rows and the same number of columns as there are classes to be identified. For identification these s rows are binary encoded as shown in Figure 13; these are the result(s) wanted and for which the training set have actually been determined by classical identification methods, and are always paired with the patterns in the same row in (a). Once trained, new input data can be passed through these ANNs, and the identities read off easily because a tabular format is employed in the classification encoding. Alternatively, for quantification purposes the output node (or nodes) would encode the amount of the substance(s) (in a mixture) that had been analyzed.

The following texts and books are recommended introductory texts to ANNs.[104,106,108,123,136,137,146–154] The following section briefly describes the salient features of both MLPs and RBFs.

6.7.1 Multilayer Perceptrons

The structure of a typical MLP is shown in Figure 14(a) and consists of three layers: MS data as the input layer, connected to an output layer encoded for identification or quantification purposes, via a single hidden layer. Each of the input nodes are connected to the nodes of the hidden layer using abstract interconnections (connections or synapses). These connections each have an associated real value, termed the weight (w_i), that scales the input (i_i) passing through them, this also includes the bias (ϑ), which also has a modifiable weight. Nodes sum the signals feeding to them (Net; Equation 18):

$$\begin{aligned}\text{Net} &= i_1w_1 + i_2w_2 + i_3w_3 + \cdots + i_iw_i + \cdots + i_nw_n \\ &= \sum_{i=1}^{n} i_iw_i + \vartheta \end{aligned} \tag{18}$$

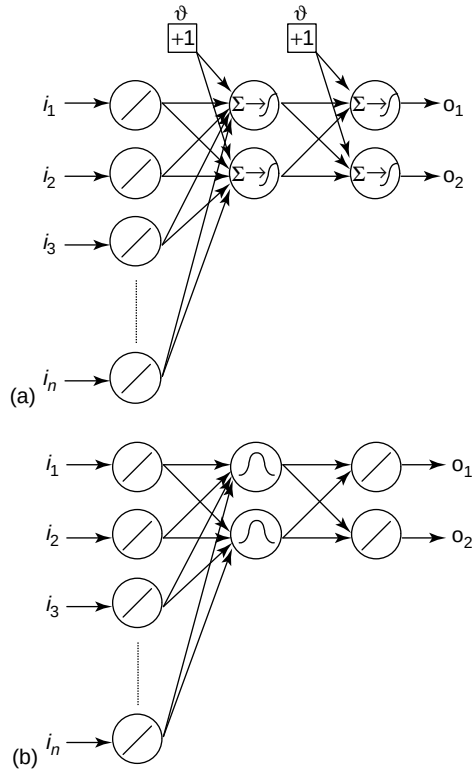

Figure 14 (a) A MLP neural network consisting of an input layer connected to two nodes in the output layer by a hidden layer (also containing two nodes). In the architecture shown, adjacent layers of the network are fully interconnected although other architectures are possible. Nodes in the hidden and output layers consist of processing elements which sum the input applied to the node and scale the signal using a sigmoidal logistic squashing function; also shown in squares is the bias. (b) RBFs consisting of an input layer connected to two nodes in the output layer by a hidden layer (also containing two nodes). The hidden layer consists of radially symmetric Gaussian functions.

The sum of the scaled inputs and the node's bias, are then scaled to lie between 0 and +1 by an activation function to give the nodes output (Out); this scaling is typically achieved using a sigmoidal (logistic) squashing function (Equation 19):

$$\text{Out} = \frac{1}{(1 + \exp^{-\text{Net}})} \tag{19}$$

These signals (Out) are then passed to the output nodes which sums them and in turn squashes this summation by the above sigmoidal activation function; the product of this node is then fed to the outside world.

For the training of the MLP the algorithm used most often is standard BP.[104,136–138,155] Briefly when the input is applied to the network, it is allowed to run until an output is produced at each output node. The differences between the actual and the desired output, taken over the entire training set are fed back through the network in

the reverse direction to signal flow (hence BP) modifying the weights as they go. This process is repeated until a suitable level of error is achieved.

One reason that MLPs are so attractive for the analysis of multivariate (spectral) data is that it has been shown mathematically[156–161] that an MLP neural network consisting of only one hidden layer, with an arbitrarily large number of nodes, can learn any, arbitrary (and hence nonlinear) mapping of a continuous function to an arbitrary degree of accuracy.

Counter-propagation neural networks use a Kohonen hidden layer that is coupled to a Grossberg output layer. These networks are hybrid in that the output layer trains by supervised delta learning and the input is unsupervised. These networks were the early precursors to the radial basis function networks. Harrington and Pack[162] modified the counterpropagation training algorithm so that both hidden and output layers were concomitantly optimized.

Training MLP networks is a very inefficient process, because all processing units are adjusted simultaneously. In addition, the number of hidden units and layers must be configured before training. The CCN developed by Fahlman and Lebiere[163] overcomes these limitations. They add hidden units as needed to reduce the calibration error. The CCN only adjusts a single processing unit or neuron at a time, and therefore trains faster than BP networks.

The problem with the perceptron model is that typically if the spectra form clusters in the data space, there are an infinite set of weight vectors orientations that will furnish zero calibration errors regardless of the dimensionality of the input data. A solution to this problem is to constrain the perceptron model. Temperature constraints that originated with the minimal neural network (MNN),[164] were implemented in BP neural networks with a single global temperature,[165] and local temperature constraints for individual perceptrons in the CCNs.[166] The temperature relates to the thermodynamic temperature employed in other methods such as simulated annealing (reference), and controls the magnitude of the weight vector length (i.e. **w** in Equation 3). The networks are trained so that the magnitude of first derivative of the objective function (e.g. error in BP and covariance in CCN) is maximized with respect to temperature. This constraint ensures that the derivative of the weight vector is large, to facilitate the training rate and the output of the perceptron remains continuous, which improves the reproducibility and the generalization capability of the networks.

6.7.2 Radial Basis Functions

By contrast, RBFs are hybrid neural networks encompassing both unsupervised and supervised learning.[108,167–173] They are also typically three-layer neural networks and, in essence, the sigmoidal squashing function is replaced by nonlinear (often either Gaussian or Mexican hat) basis functions or kernels (Figure 14b). The kernel is the function that determines the output of each node in the hidden layer when an input pattern is applied to it. This output is simply a function of the Euclidean distance from the kernel centre to the presented input pattern in the multidimensional space, and each node in the hidden layer only produces an output when the input applied is within its receptive field; if the input is beyond this receptive field the output is 0. This receptive field can be chosen and is radially symmetric around the kernel centre. Between them the receptive fields cover the entire region of the input space in which a multivariate input pattern may occur; a diagrammatic representation of this is given in Figure 15, where a two-dimensional input is mapped by eight radially symmetric basis functions. This is a fundamentally different approach from the MLP, in which each hidden node represents a nonlinear hyperplanar decision boundary bisecting the input space (Figure 15a). Thus RBFs have the advantage over gradient descent MLPs in that they have the ability to learn any arbitrary nonlinear mapping of a discontinuous function to an arbitrary degree of accuracy.[108,155,167]

The outputs of the RBF nodes in the hidden layer are then fed forward via weighted connections to the nodes in the output layer in a similar fashion to the MLP, and each output node calculates a weighted sum of the outputs from the nonlinear transfer from the kernels in the hidden layer. The only difference is that the output nodes of an RBF are normally linear, whilst those of the MLP more typically employ a sigmoidal or logistic (nonlinear) squashing function.

Thus in the RBF training proceeds in two stages. Stage 1 involves unsupervised clustering of the input data, typically using the K-means clustering algorithm[90,172,174] to divide the high-dimensional input data into clusters. Next, kernel centers are placed at the mean of each cluster of data points. The use of K-means is particularly convenient because it positions the kernels relative to the density of the input data points. Next the receptive field is determined by the nearest neighbor heuristic where r_j (the radius of kernel j) is set to the Euclidean distance between w_j (the vector determining the centre for the jth RBF) and its nearest neighbor (k), and an overlap constant (Overlap) is used (Equation 20):

$$r_j = \text{Overlap} \times \min(\|w_j - w_k\|) \tag{20}$$

where $\|\cdots\|$ denotes a vector norm, or Euclidean distance. The overlap that often gives best results is where the edge of the radius of one kernel is at the centre of its nearest neighbor.[170]

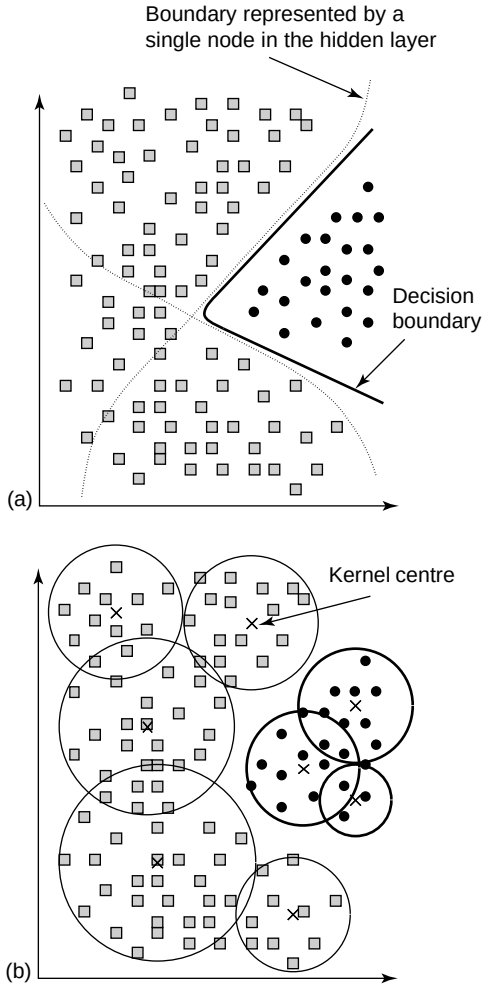

Figure 15 (a) Typical decision boundary for a classification problem created between two data classes by a MLP with two nodes in the hidden layer, for two input nodes. Each hidden node represents a nonlinear hyperplanar boundary and the node(s) in the output layer interpolate this to form a decision boundary. (b) The same classification problem modeled by eight radially symmetric basis functions. The width of each kernel function (referred to as its receptive field) is determined by the local density distribution of training examples.

The output from nodes in the hidden layer is dependent on the shape of the basis function and the one used was that of the Gaussian. Thus this value (R_j) for node j when given the ith input vector (i_i) can be calculated by (Equation 21):

$$R_j(i_i) = e^{-(i_i^2/r_j^2)} \tag{21}$$

Stage 2 involves supervised learning using simple linear regression. The inputs are the output values for all n basis functions ($R_1 - R_n$) for all the training input patterns to that layer ($i_1 - i_n$), and the outputs are identities binary encoded as shown in Figure 13. More recently, Walczak and Massart have used PLS as the linear regression method.[101] Wan and Harrington used SVD regression and reported a self-configuring RBF network that optimizes the number of kernel functions.[175]

6.8 Applications of Artificial Neural Networks to Pyrolysis Mass Spectrometry

PyMS involves the thermal degradation of nonvolatile complex molecules (such as bacteria) in a vacuum causing their cleavage to smaller, volatile fragments, separable by a mass spectrometer on the basis of their m/z.[91] The PyMS method allows the (bio-)chemically-based discrimination of microbial cells (and other organic material) and produces complex biochemical fingerprints (i.e. pyrolysis mass spectra) which are distinct for different bacteria. It is the automation of the instrumentation and ease of use that has lead to the widespread exploitation of PyMS as a taxonomic tool for whole-organism fingerprinting.[86,176] The analytically useful multivariate data (Figure 16) are typically constituted by a set of 150 normalized intensities versus m/z in the range 51 to 200,

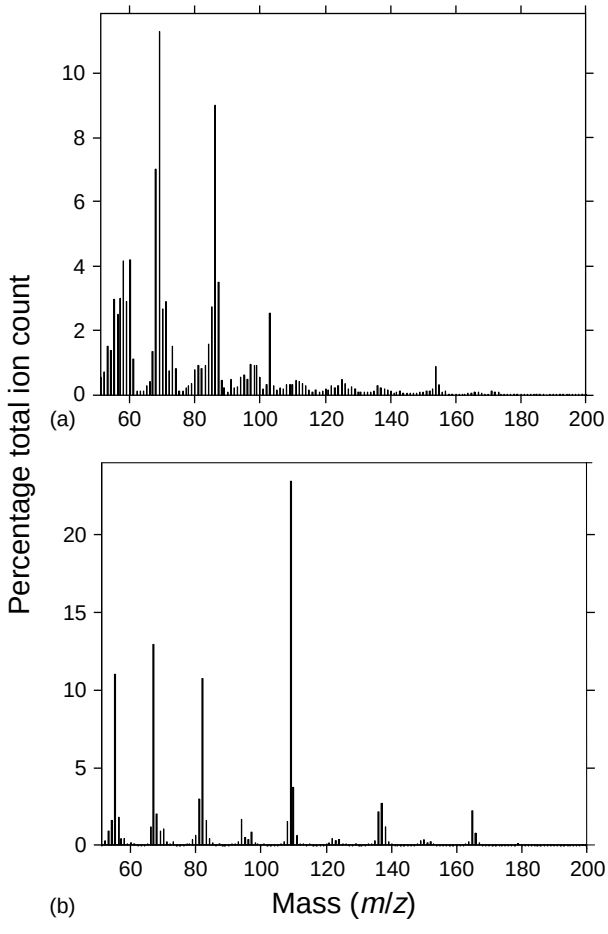

Figure 16 Pyrolysis mass spectra: (a) the bacterium *Bacillus cereus* DSM 31; (b) the simple (bio)chemical caffeine.

and these are applied to the nodes on the input layers of ANNs.

The first demonstration of the ability of ANNs to discriminate between biological samples using their pyrolysis mass spectra was for the qualitative assessment of the adulteration of extra virgin olive oils with various seed oils;[177,178] in this study, which was performed double-blind, neural networks were trained with the spectra from 12 virgin olive oils, coded '1' at the output node, and with the spectra from 12 adulterated oils, which were coded '0'. All oils in the test were correctly identified; in a typical run, the virgins were assessed with a code of 0.99976 ± 0.000146 (range $0.99954 - 1.00016$) and the adulterated olive oils in the test set with a code of 0.001079 ± 0.002838 (range $0.00026 - 0.01009$). This permitted their rapid and precise assessment, a task which previously was labour intensive and very difficult. It was most significant that the traditional unsupervised MVA of PCA, DFA and HCA failed to separate the oils according to their virginity or otherwise, but rather discriminated them on the basis of their cultivar (that is to say, the biggest difference in the mass spectra was due to the type of olive tree that the fruit came from, rather than the adulterant).

The use of PyMS with MLPs for the analysis of foodstuffs is becoming widespread[179] and has been investigated for identifying the geographical origin of olive oils,[180] for the characterization of cocoa butters[181,182] and milk,[183,184] classification of pork backfat[185] and European wines,[186] for differentiating between industrially made vinegar 'Aceto Balsamico di Modena' and traditionally produced vinegar 'Aceto Balsamico Tradizionale di Modena e di Reggio Emilia',[187] for detecting the adulteration of orange juice,[188] and for detecting caffeine in coffee, tea and cola drinks.[189]

Several studies have also shown that this combination of PyMS and MLPs are also very effective for the rapid identification of a variety of bacterial strains of industrial, clinical and veterinary importance.[87,190] For example, this approach has allowed the propionibacteria isolated from dogs to be correctly identified as human *Propionibacterium acnes*,[127] for detecting *Escherichia coli* isolates which produced verocytotoxins,[191] for distinguishing between *Mycobacterium tuberculosis* and *M. bovis*,[192] and for identifying streptomycetes recovered from soil,[193–195] oral abscess bacteria,[196] and fungi belonging to the genus *Penicillium* which were associated with cheese.[197] An example of the highly discriminatory nature of MLPs is that one can even use them to differentiate between methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*;[198] the relevant difference is an alteration in a single penicillin-binding protein.[199,200] Similarly, MLPs can be used to discriminate the physiological difference between sporulated and vegetative *Bacillus* species,[201] and differentiating the verocytotoxin production status in *Escherichia coli*.[191]

RBF have been rather less widely applied to the analysis of mass spectral data. Kang et al.[202,203] have exploited RBFs to detect physiological changes in industrial fermentations of *Streptomyces* species, whereas Goodacre et al.[204] have used RBFs for the identification of common infectious agents associated with urinary tract infection from their MS, IR, and Raman spectra.

An analytical expression for the derivatives of the entropy function for MNNs was derived.[205] This expression was implemented for classifying pyrolysis MS/MS data and the results were compared to those obtained from a BP/ANN.[206] The MNNs differ from BP/ANNs in that they use localized processing and build classification trees with branches composed of multiple processing units. A global entropy minimization may be achieved at a branch by combining the processing logic using principles from fuzzy set theory. Weight vectors are adjusted using an angular coordinate system and gradients of the fuzzy entropy function. The branches are optimal with respect to fuzziness and can accommodate nonlinearly separable or ill-conditioned data. The most significant advantage of the MNNs is that relations among the training data and the mechanism of inference may be directly observed; thus rule-based classification trees have been constructed from the mass spectral daughter ions to discriminate between diesel smoke, dry yeast, *Escherichia coli*, MS-2 coliphage, grass pollen, *Bacillus subtilis*, fog oil, wood smoke, aldolase and *Bacillus globigii*.[206]

6.8.1 *Classification and Qualitative Analysis of Mass Spectra*

ANNs may be used to construct classification models from mass spectra. Once the classification models are built they may be used to rapidly screen large collections of mass spectra. The earliest application of perceptron models was applied to substructure recognition from mass spectra.[207] MLPs were first employed for recognizing functional groups from large collections of mass spectra by Curry and Rumelhart.[208] Werther et al.[209] demonstrated that classifiers based on RBF were better at recognizing functional groups than soft independent models for class analogies (SIMCA[210]), K-nearest neighbors,[211] and linear discriminant analysis[212] from mass spectra.[209] The combination of separation stages to mass spectrometers, such as chromatographic and electrophoretic, furnishes large collections of mass spectra. A fuzzy rule-building expert system (FuRES) was applied to screening GC/MS data of plastic recycling products.[213] The ES was capable

of classifying each mass spectral scan of the chromatographic run by degree of unsaturation (i.e. alkane, alkene, and diene) and furnished a separate chromatogram for each of the three classes. A FuRES has also been used to classify pyrolysis mass spectra[214] and IR spectra.[215]

Temperature-constrained cascade correlation networks (TCCCNs) were applied to the screening of GC/MS pesticide data. Substructures and toxicity ANN models were built for organophosphorus pesticides and applied to screening GC/MS data.[216] The TCCCN was applied to recognizing substructures of carbamates pesticides, and reported the Latin-partition method for evaluating ANN models.[110]

6.8.2 *Quantitative Analysis with Artificial Neural Networks*

All the above studies have been classification problems but perhaps the most significant application of ANNs to the analysis of MS data is to gain accurate and precise quantitative information about the chemical constituents of microbial samples. For example, it has been shown that it is possible using this method to measure the concentrations of binary and tertiary mixtures of cells of the bacteria *Bacillus subtilis*, *Escherichia coli*, and *Staphylococcus aureus*.[107,149,217] With regard to biotechnology, the combination of PyMS and ANNs can be exploited to quantify the amount of mammalian cytochrome b_5 expressed in *Escherichia coli*,[218] and to measure the level of metabolites in fermentor broths.[87,219] In related studies *Penicillium chrysogenum* fermentation broths were analyzed quantitatively for penicillins using PyMS and ANNs,[220] and this approach has also been used to monitor *Gibberella fujikuroi* fermentations producing gibberellic acid[221] and quantify the level of clavulanic acid produced by *Streptomyces clavuligerus*,[203] and to quantify the expression of the heterologous protein α2-interferon in *Escherichia coli*.[222] These and related chemometric approaches have been extended to work with a variety of high dimensional spectroscopic methods,[223] including those based on IR,[222,224] Raman,[221,225] dielectrics,[226] and flow cytometric measurements.[227]

6.8.3 *Instrument Reproducibility*

For MS to be used for the routine identification of microorganisms new (spectral) fingerprints must be able to be compared to those previously collected. However, the major problem with most analytical instruments is that long-term reproducibility is poor and interlaboratory reproducibility abysmal, and so the biochemical or genetic fingerprints of the same material analyzed at two different times are different. Because of the uncertainties over the long-term reproducibility of the PyMS system (defined as over 30 days), PyMS has really been limited within clinical microbiology to the typing of short-term outbreaks where all micro-organisms are analyzed in a single batch.[86,228]

After tuning the instrument, to correct for drift one would need to analyze the same standards at the two different times and use some sort of mathematical correction method. This could simply be subtracting the amount of drift from new spectra collected; however, this assumes that the drift is uniform (linear) with time, which is obviously not the case. This method also relies on the variables (characters) being void of noise, which is also not the case. An alternative method would be to transform the spectra to look like the spectra of the same material previously collected using a method that was (a) robust to noisy data and (b) able to perform nonlinear mappings. ANNs carry out nonlinear mappings, while being able to map the linearities, and are purported to be robust to noisy data. These mathematical methods are therefore ideally suited to be exploited for the correction of mass spectral drift.

Goodacre and Kell[229,230] have found that neural networks can be used successfully to correct for instrumental drift; identical materials were analyzed by PyMS at dates from 4 to 20 months apart, but neural network models produced at earlier times could not be used to give accurate estimates of determinand concentrations or bacterial identities. Calibration samples common to the two datasets were run at the two times, and ANNs set-up in which the inputs were the 150 new calibration masses and the outputs were the 150 calibration masses from the old spectra. Such associative nets could thus be used as signal-processing elements to effect the transformation of data acquired one day to those that would have been acquired on a later date. A further study[231] has shown that one can also affect calibration transfer between laboratories using this approach. These results show clearly that for the first time PyMS can be used to acquire spectra which could be compared to those previously collected and held in a database. It should seem obvious that this approach is not limited solely to PyMS but is generally applicable to any analytical tool which is prone to instrumental drift (which cannot be compensated for by tuning).

6.9 Concluding Remarks

Within MS the move from a stare-and-compare approach to the analysis of highly dimensional multivariate data necessitates the use of chemometrics; particularly when quantitative information is sought from biological systems. The application of ANNs for quantitative and qualitative analyses is becoming more accepted within MS, especially because these neural computational methods clearly present themselves as extremely powerful and valuable tools for the analysis of complex data.

7 OPTIMIZATION TECHNIQUES IN MASS SPECTROMETRY

7.1 Introduction

Parameter optimization is often required in MS. It can be employed in the design of the instrument, tuning of conditions during operation, or calibration of mass scales in data analysis. Good sources of information on general optimization techniques are readily available.[232–235] Therefore, a specific example is provided, illustrating how optimization is advancing MS capability. This example is the application of simplex optimization to mass calibration in TOF instruments.

The simplex method is one of the most popular of several mathematical techniques for optimizing multivariate systems. Developed in 1947, it is composed of successive tests and variation of independent parameters until the system is determined to be unbounded or optimized.[234] Whereas the original method used a graphical representation, current methods rely on high-speed computing. The advances in computers allow the determination of optimal conditions in complicated systems involving a large number of independent parameters.

7.2 Time-of-flight Mass Spectroscopy Mass Calibration

Because of its simplicity and unlimited mass range, TOF instrumentation is particularly well-suited for the analysis of MALDI and electrospray-generated macromolecular ions. Recent advances in TOF technology have facilitated the attainment of high resolution in MALDI/TOF experiments,[236–239] but the method's utility critically depends on its ability to measure ion masses accurately. An ion's flight time can be expressed as an exact function of its mass if information about its formation time, location, and initial velocity is available. Most often, this relationship is expressed to some level of approximation. To zeroth order, the TOF is given by Equation (22):

$$\text{TOF} = k \times \text{mass}^{1/2} \tag{22}$$

This is accurate only when ions are formed with zero initial velocity. Non-zero initial velocities introduce significant complications in the relationship between TOF and mass.

It is currently popular to express TOF as a multiterm polynomial function of mass.[240–242] In this expansion the $\text{mass}^{1/2}$ term is a dominant contributor, but additional terms are needed to produce accurate results. An infinite series is needed to reach arbitrary accuracy. In the polynomial curve-fitting approach, mass spectra of a variety of known calibrant samples are recorded, ion flight times are measured and the coefficients in the polynomial relating TOF to mass are derived using a least-squares minimization routine. When other samples of known mass are introduced into the instrument and their masses are derived using the polynomial relationship, mass accuracies in the parts per million realm have been demonstrated.[243] Nevertheless, to achieve high-quality results with this method, it is necessary that the masses to be measured are near those of calibrants and it is best if they are bracketed by the latter. Attempts to extrapolate the mass calibration over a range that extends beyond that of the calibrants generally leads to poor results.

As an alternative to fitting TOF to an arbitrary polynomial function of mass, it is possible to use elementary physics to calculate the flight times of ions based on instrumental voltages, distances and other operating parameters. Besides not using arbitrary parameters, this approach incorporates the correct physical relationship between TOF and mass, and thus it should extrapolate more accurately into mass regions that are far from the calibrants. Although the merits of this approach are easy to envisage, the stumbling block to its use may be equivalently obvious. For ions initiating their trajectories in a TOF instrument with nonzero velocities, being accelerated in more than one field, drifting in a field-free region and, possibly being postaccelerated or decelerated, the relationship between TOF and mass contains a number of parameters whose exact values are not known well enough to provide the basis for an accurate mass calibration equation. For example, suppose that a singly charged lysozyme ion having a mass of 14 306 Da is accelerated to 20 keV total energy and drifts through a 1.00 m long field-free flight tube. The overall flight time is calculated to be 67 147.2 ns. An error in the high voltage of 10 V leads to a flight time shift of 15 ns. Likewise, an error in the drift tube length of 0.3 mm corresponds to a time shift of 18 ns. These numbers are equivalent to mass errors of 6.5 and 7.8 Da or 460 and 550 parts per million, respectively. Fortunately, accurate values for these imprecisely known instrument parameters can be derived through a simplex optimization procedure. This leads to both a mass calibration equation and a computationally accurate description of the instrument.

7.2.1 Use of the Simplex Algorithm

The minimization used in the simplex algorithm involves computing residual errors between an array of experimental and calculated flight times. The algorithm reiteratively optimizes the instrument parameters in order to minimize the difference between experimental flight times and those calculated using values of the instrument parameters. Any residual error function may be used in the minimization routine. For example, the sum of the squares of differences between experimental and calculated flight times has been used. Two of the input parameters needed by the Simplex algorithm are the characteristic length vector and the fit tolerance. Vector components are typically

matched to measurement uncertainties. The fit tolerance represents the desired conditions for termination of the optimization and is based on expected error between experimental and optimized calculations. Too small a value increases the iterative requirements of the calculation; too large a number causes the simplex navigation to terminate before a minimum is found.

The Simplex calculation constrains ion behavior to physically meaningful values as it is based on exact electrostatic equations. The significance of this can best be demonstrated through direct comparison with the curve-fitting approach to mass calibration. Theoretical flight times for 101 ions having masses between 100 and 1100 Da are initially computed using the exact electrostatic TOF expression. When the resulting flight times are fitted to an optimized three- or five-term polynomial function of mass, the residual errors are rather small, as displayed by the curves in Figure 17. Note that the scale of this graph is rather expanded. However, when the Simplex algorithm is employed for the same purpose, the calibration is seen to be noticeably improved. Although curve-fitting with an *n*-term polynomial exactly matches the theoretical data at the *n* points, discrepancies can be noted between these points. Furthermore the exact electrostatic calculations and the fitted polynomial diverge significantly at masses on the low and high ends of the calibration range. The quality of results obtained by curve-fitting does vary

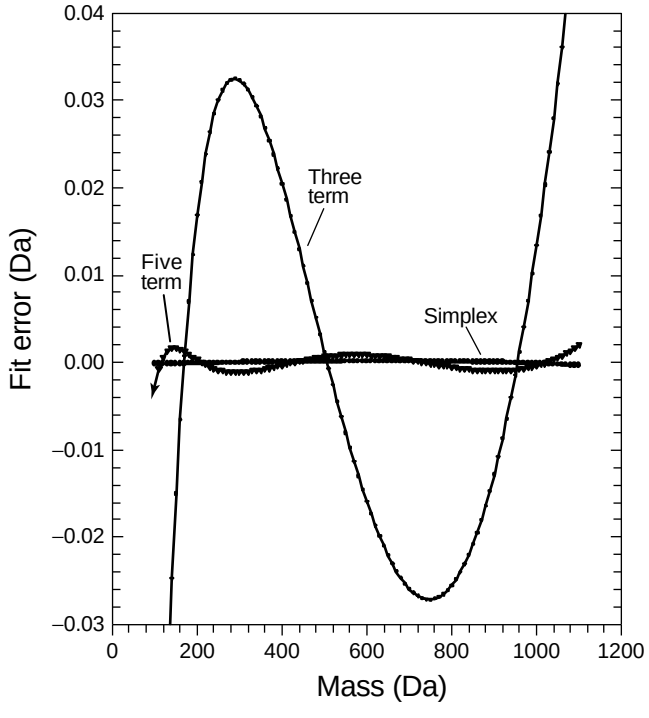

Figure 17 Comparison of polynomial curve fitting and simplex optimization mass calibration approaches when applied to theoretical data.

depending on the polynomial function form and in some cases can match that attained by Simplex. However, the general trends just noted are still observed.

An alternative way to compare the capabilities of Simplex-based mass calibration with the curve-fitting approach is to examine experimental data. A particularly interesting case involves a sample containing a mixture of alkylthioate-coated gold nanoparticles. The challenging aspect of this calibration problem is that gold nanoparticles are laser-desorbed and ionized without being incorporated into a matrix.[244] Consequently, the resulting ions have a velocity distribution that is not characteristic of MALDI ions. In fact, gold nanoparticle ions have lower desorption velocities than their MALDI counterparts. Unfortunately, in the 10–20 kDa mass range, the most familiar mass calibration standards are MALDI-generated protein ions. Thus, an inherent incompatibility between sample and calibrants exists. To deal with this, the parameters chosen for simplex optimization must be carefully chosen. In general, the best ones to optimize are those that are subject to the largest measurement error. Extraction voltage, the ion drawout pulse delay, and the length of the flight tube are all obvious choices. Optimization of this collection of parameters normally yields an accurate mass calibration for a mixture of peptides for which a typical initial ion velocity is $600\,m\,s^{-1}$. However, for the gold sample, the Simplex optimization would not converge when ions were assigned this initial velocity. If the initial gold nanoparticle ion velocity is changed to $100\,m\,s^{-1}$, the algorithm does converge, leading to RMS mass errors for about 30 ion peaks of 0.24 Da. This corresponds to 17 ppm error, which is respectable considering the velocity differences between MALDI-generated protein ions and gold ions. It is noteworthy that in this example, the Simplex algorithm provided more than just a means to perform mass calibration. It also yielded information about the average velocity with which the nanoparticles desorbed.

7.2.2 Mass Calibration Extrapolation

As noted above, calibration of most mass spectra is performed by surrounding the peaks of interest with a good set of known standards. It may seem unreasonable to expect a calibration method to remain accurate in mass regions extrapolated beyond the range of the calibrants. However, this is one of the virtues of the simplex approach. Six peaks in the gold nanoparticle mass spectrum were used to calibrate the spectrum. As displayed in Figure 18, a five-term polynomial curve fit established an acceptable relationship between masses and flight times for ions in the 13–14 kDa range. The simplex approach performed comparably well within this range. However, at masses below 13 kDa or above 14 kDa,

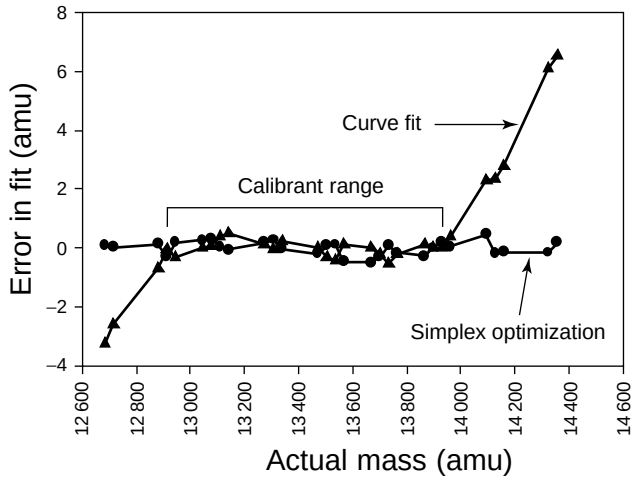

Figure 18 Comparison of polynomial curve fitting and simplex optimization mass calibration approaches when applied to experimentally measured gold nanoparticle mass spectra.

the high-order polynomial curve clearly changes slope and errors relative to the exact TOF calibration and the more realistic Simplex curve increase dramatically.

7.2.3 Conclusions

In summary, a simplex approach to calibrate MALDI/TOF mass spectra appears to be both robust and easily executed. It yields results that are comparable to those obtained with polynomial curve fitting for routine applications, and it excels under more difficult situations such as when samples and calibrants are fundamentally different, when calibrant peaks are incorrectly assigned, or when the mass range of interest must be extended beyond that of the calibrants. It should therefore be of utility in a wide variety of applications.

ABBREVIATIONS AND ACRONYMS

AAANN	Auto-associative Artificial Neural Network
ADC	Analog-to-digital Converter
AI	Artificial Intelligence
ALS	Alternating Least Squares
ANN	Artificial Neural Network
BP	Backpropagation
CCN	Cascade-correlation Network
CVA	Canonical Variates Analysis
CVs	Canonical Variates
dBEST	Database of Expressed Sequence Tags
DC	Direct Current
EFA	Evolving Factor Analysis
ES	Expert System
ESI	Electrospray Ionization
FA	Factor Analysis
FAB/MS	Fast Atom Bombardment Mass Spectrometry
FOAM	Fuzzy Optimal Associative Memory
FTMS	Fourier Transform Mass Spectrometry
FuRES	Fuzzy Rule-building Expert System
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ICR	Ion Cyclotron Resonance
IND	Indicator Function
IR	Infrared
LI/MS	Laser Ionization Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/Ionization
MLP	Multilayer Perceptron
MNN	Minimal Neural Network
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MVA	Multivariate Analysis
OAM	Optimal Associative Memory
PCA	Principal Component Analysis
PCR	Principal Component Regression
PFTBA	Perfluorotributylamine
PLS	Partial Least Squares
PyMS	Pyrolysis Mass Spectrometry
RBF	Radial Basis Function Neural Networks
RF	Radiofrequency
RRMSEC	Relative Root Mean Square Error of Calibration
SIMCA	Soft Independent Modeling for Class Analogies
SIMS	Secondary Ion Mass Spectrometry
SOM	Self-organizing Map
SVD	Singular Value Decomposition
TCCCN	Temperature-constrained Cascade Correlation Network
TOF	Time-of-flight
WFA	Window Factor Analysis

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications • Pyrolysis Gas Chromatography in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Pyrolysis Techniques in the Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Pulp and Paper **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Electron Ionization Mass Spectrometry • Literature of Mass Spectrometry • Time-of-flight Mass Spectrometry

REFERENCES

1. P.H. Winston, *Artificial Intelligence*, 2nd edition, Addison-Wesley, London, 1984.
2. E. Rich, *Artificial Intelligence*, McGraw-Hill, New York, 1983.
3. Application Report 002, 'Townsend Discharge Nitric Oxide CI GC/MS for Hydrocarbon Analysis of the Middle Distillates', Bear Instruments, Santa Clara, California, 1999.
4. S.E. Stein, 'An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data', *J. Am. Soc. Mass Spectrom.*, **10**(8), 770–781 (1999).
5. P. Ausloos, C.L. Clifton, S.G. Lias, A.I. Mikaya, S.E. Stein, D.V. Tchekhovskoi, O.D. Sparkman, V. Zaikin, Damo Zhu, 'The Critical Evaluation of a Comprehensive Mass Spectral Library', *J. Am. Soc. Mass Spectrom.*, **199**(10), 287–299 (1999).
6. J.R. Chapman, *Computers in Mass Spectrometry*, Academic Press, London, 1978.
7. H.E. Duckworth, R.C. Barber, V.S. Venkatasubramanian, *Mass Spectroscopy*, Cambridge University Press, 1990.
8. W.R. Smythe, J. Mattauch, 'A New Mass Spectrometer', *Phys. Rev.*, **40**, 429–433 (1932).
9. W.C. Wiley, I.H. McLaren, 'Time-of-flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Inst.*, **26**, 1150–1157 (1955).
10. W. Paul, H.P. Reinhard, U. von Zahn, 'Das elektrische Massenfilter als Massenspektrometer und Isotopentrenner', *Z. Phys.*, **152**, 143–182 (1958).
11. R.F. Bonner, G. Lawson, J.F.J. Todd, 'Ion–Molecule Reaction Studies with Quadrupole Ion Storage Trap', *Int. J. Mass Spectrom. Ion Phys.*, **197**, 197–203 (1972/73).
12. J.D. Baldeschwieler, H. Benz, P.M. Llewellyn, 'Ion–Molecule Reactions in an Ion Cyclotron Resonance Mass Spectrometer', *Advances in Mass Spectrometry*, ed. E. Kendrick, Institute of Petroleum, London, Vol. 4, 113–120, 1968.
13. M.B. Comisarow, A.G. Marshall, 'Theory of Fourier Transform Ion Cyclotron Resonance Mass Spectroscopy. I. Fundamental Equations and Low Pressure Line Shape', *J. Chem. Phys.*, **64**, 110–119 (1976).
14. S. Savory (ed.), *Artificial Intelligence and Expert Systems*, Halsted Press, Chichester, 1988.
15. R.A. Edmunds, *The Prentice Hall Guide to Expert Systems*, Prentice Hall, Englewood Cliffs, NJ, 1988.
16. J.M. Gillette, 'Mass Spectrometer Data Reduction Program for IBM 650', *Anal. Chem.*, **31**(9), 1518–1521 (1959).
17. J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield, C. Djerassi, 'Applications of Artificial Intelligence for Chemical Inference. I. The Number of Possible Organic Compounds. Acyclic Structures Containing C, H, O and N', *J. Am. Chem. Soc.*, **11**, 2973–2976 (1968).
18. Mass Spectrometry Data Centre; Imperial Chemical Industries, *Eight Peak Index of Mass Spectra: The Eight Most Abundant Ions in 66,720 Mass Spectra, Indexed by Molecular Weight, Elemental Composition and Most Abundant Ions*, The Centre, Royal Society of Chemistry, Nottingham, UK, 1983.
19. D.J. Pappin, 'Peptide Mass Fingerprinting Using MALDI-TOF Mass Spectrometry', *Methods Mol. Biol.*, **64**, 165–173 (1997).

20. R.S. Brown, J.J. Lennon, 'Sequence-specific Fragmentation of Matrix-assisted Laser-desorbed Protein/Peptide Ions', *Anal. Chem.*, **67**, 3990–3999 (1995).
21. J. Qin, B.T. Chait, 'Identification and Characterization of Posttranslational Modifications of Proteins by MALDI Ion Trap Mass Spectrometry', *Anal. Chem.*, **69**, 4002–4009 (1997).
22. A. Shevchenko, I. Chenushevich, W. Ens, K.G. Standing, B. Thomson, M. Wilm, M. Mann, 'Rapid 'de novo' Peptide Sequencing by a Combination of Nanoelectrospray, Isotope Labeling and A Quadrupole/Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **11**, 1015–1024 (1997).
23. A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, 'Linking Genome and Proteome by Mass Spectrometry: Large Scale Identification of Yeast Proteins from Two Dimensional Gels', *Proc. Nat. Acad. Sci. USA*, **93**, 14 440–14 445 (1996).
24. J.R. Yates, 'Mass Spectrometry. From Genomics to Proteomics', *Trends Genet.*, **16**, 5–8 (2000).
25. M.S. Boguski, T.M. Lowe, C.M. Tolstoshev, 'dbEST – Database for ''Expressed Sequence Tags'' ', *Nat. Genet.*, **4**, 332–333 (1993).
26. W. Zhang, B.T. Chait, 'ProFound – An Expert System for Protein Identification Using Mass Spectrometric Peptide Mapping Information', *Anal. Chem.*, **72**, in press (2000).
27. R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser Desorption Ionization Mass Spectrometry of Proteins', *Methods Enzymol.*, **270**, 519–551 (1996).
28. D.H. Patterson, G.E. Tarr, F.E. Regnier, S.A. Martin, 'C-terminal Ladder Sequencing Via Matrix-assisted Laser Desorption Mass Spectrometry Coupled with Carboxypeptidase Y Time-dependent and Concentration-dependent Digestions', *Anal. Chem.*, **67**, 3971–3978.
29. H. Hotelling, 'Analysis of a Complex of Statistical Variables into Principal Components', *J. Educat. Psychol.*, **24**, 417–441 (1933).
30. R.M. Wallace, 'Analysis of Absorption Spectra of Multicomponent Systems', *J. Phys. Chem.*, **64**, 899–901 (1960).
31. R.M. Wallace, S.M. Katz, 'A Method for the Determination of Rank in the Analysis of Rank in the Analysis of Absorption Spectra of Multicomponent Systems', *J. Phys. Chem.*, **68**(12), 3890–3892 (1964).
32. P.T. Funke, E.R. Malinowski, D.E. Martire, L.Z. Pollara, 'Application of Factor Analysis to the Prediction of Activity Coefficients of Nonelectrolytes', *Sep. Sci.*, **1**(6), 661–676 (1966).
33. P.H. Weiner, E.R. Malinowski, A.R. Levinstone, 'Factor Analysis of Solvent Shifts in Proton Magnetic Resonance', *J. Phys. Chem.*, **74**(26), 4537–4542 (1970).
34. P.H. Weiner, D.G. Howery, 'Factor Analysis of Some Chemical and Physical Influences in Gas–Liquid Chromatography', *Anal. Chem.*, **44**(7), 1189–1194 (1972).
35. P.H. Weiner, C.J. Dack, D.G. Howery, 'Retention Index-structure Relationships for Alcohols Using Factor Analysis', *J. Chromatogr.*, **69**, 249–260 (1972).
36. P.H. Weiner, J.F. Parcher, 'Prediction of Some Physical Properties of Organic Molecules by Factor Analysis of Gas Chromatographic Retention Indices', *Anal. Chem.*, **45**(2), 302–307 (1973).
37. S. Wold, K. Andersson, 'Major Components Influencing Retention Indices in Gas Chromatography', *J. Chromatogr.*, **80**, 43–59 (1973).
38. D. Macnaughtan, Jr, L.B. Rogers, G. Wernimont, 'Principal-component Analysis Applied to Chromatographic Data', *Anal. Chem.*, **44**(8), 1421–1427 (1972).
39. R.W. Rozett, E.M. Petersen, 'Methods of Factor Analysis of Mass Spectra', *Anal. Chem.*, **47**(8), 1301–1308 (1975).
40. J.B. Justice, Jr, T.L. Isenhour, 'Factor Analysis of Mass Spectra', *Anal. Chem.*, **47**(13), 2286–2288 (1975).
41. R.W. Rozett, E.M. Petersen, 'Classification of Compounds by the Factor Analysis of Their Mass Spectra', *Anal. Chem.*, **48**(6), 817–825 (1976).
42. G.L. Ritter, S.R. Lowry, T.L. Isenhour, C.L. Wilkins, 'Factor Analysis of the Mass Spectra of Mixtures', *Anal. Chem.*, **48**(3), 591–595 (1976).
43. F.J. Knorr, J.H. Futrell, 'Separation of Mass Spectra of Mixtures by Factor Analysis', *Anal. Chem.*, **51**(8), 1236–1241 (1979).
44. P.C. Tway, L.J. Cline Love, H.B. Woodruff, 'A Totally Automated Data Acquisition/Reduction System for Routine Treatment of Mass Spectrometric Data by Factor Analysis', *Anal. Chim. Acta*, **117**, 45–52 (1980).
45. H.L.C. Meuzelaar, J. Haverkamp, F.D. Hileman, *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*, Elsevier, Amsterdam, 1982.
46. W. Windig, J. Haverkamp, P.G. Kistemaker, 'Interpretation of Sets of Pyrolysis Mass Spectra by Discriminant Analysis and Graphical Rotation', *Anal. Chem.*, **55**(1), 81–88 (1983).
47. W. Windig, H.L.C. Meuzelaar, 'Nonsupervised Numerical Component Extraction from Pyrolysis Mass Spectra of Complex Mixtures', *Anal. Chem.*, **56**(13), 2297–2303 (1984).
48. H.J.H. MacFie, C.S. Gutteridge, 'Comparative Studies on some Methods for Handling Quantitative Data Generated by Analytical Pyrolysis', *J. Anal. Appl. Pyrolysis*, **4**, 175–204 (1982).
49. L.V. Vallis, H.J. MacFie, C.S. Gutteridge, 'Comparison of Canonical Variates Analysis with Target Rotation and Least-squares Regression as Applied to Pyrolysis Mass Spectra of Simple Biochemical Mixtures', *Anal. Chem.*, **57**(3), 704–709 (1985).
50. J.T. Magee, 'Whole-organism Fingerprinting', in *Handbook of New Bacterial Systematics*, eds. M. Goodfellow, A.G. O'Donnell, Academic Press, London, 383–427, 1993.

51. G. Blomquist, E. Johansson, B. Soderstrom, S. Wold, 'Data Analysis of Pyrolysis–Chromatograms by Means of SIMCA Pattern Recognition', *J. Anal. Appl. Pyrolysis*, **1**, 53–65 (1979).
52. F.R. Di Brozolo, R.W. Odom, P.d.B. Harrington, K.J. Voorhees, 'Organic Polymer Analysis by Laser Ionization Mass Spectrometry and Pattern Recognition Techniques', *J. Appl. Polym. Sci.*, **41**, 1737–1752 (1990).
53. M. Lamberto, M. Saitta, 'Principal Component Analysis in Fast-atom-bombardment Mass-spectrometry of Triacylglycerols in Edible Oils', *J. Am. Oil Chem. Soc.*, **72**, 867–871 (1995).
54. R. Goodacre, J.K. Heald, D.B. Kell, 'Characterization of Intact Microorganisms Using Electrospray Ionization Mass Spectrometry', *FEMS Microbiol. Lett.*, **176**, 17–24 (1999).
55. P. Zheng, P.B. Harrington, A. Craig, R. Felming, 'Cluster Analysis for Variable Alignment of High Resolution Data', *Anal. Chim. Acta*, **310**, 485–492 (1995).
56. P.J. Tandler, J.A. Butcher, H. Tao, P.B. Harrington, 'Chemometric Analysis of Plastic Recycling Products', *Anal. Chim. Acta*, **312**(3), 231–244 (1995).
57. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, 'Numerical Recipes in C', in *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, 59–70, 1994.
58. M.B. Seaholtz, R.J. Pell, K.E. Gates, 'Comments on the Power Method', *J. Chemom.*, **4**, 331–334 (1990).
59. E.R. Malinowski, 'Determination of the Number of Factors and the Experimental Error in a Data Matrix', *Anal. Chem.*, **49**(4), 612–617 (1977).
60. J.H. Kalivas, P. Lang, 'Condition Numbers, Iterative Refinement and Error Bounds', *J. Chemom.*, **3**, 443–449 (1989).
61. J.H. Kalivas, P.M. Lang, 'Interrelationships Between Sensitivity and Selectivity Measures for Spectroscopic Analysis', *Chemom. Intell. Lab. Syst.*, **32**, 135–149 (1996).
62. L.R. Crawford, J.D. Morrison, 'Computer Methods in Analytical Mass Spectrometry Empirical Identification of Molecular Class', *Anal. Chem.*, **40**(10), 1469–1474 (1968).
63. R.H. Lindeman, P.F. Merenda, R.Z. Gold, *Introduction to Bivariate and Multivariate Analysis*, Scott, Foresman & Co., New York, 444, 1980.
64. B.F.J. Manly, *Multivariate Statistical Methods: A Primer*, Chapman & Hall, London, 1994.
65. C. Cai, P.B. Harrington, 'Linear Discriminant Classification of Fourier and Wavelet Compressed IMS Data', *Appl. Spectrosc.*, (submitted).
66. W.J. Dixon, *Biomedical Computer Programs*, University of California Press, Los Angeles, 1975.
67. H.F. Kaiser, 'The Varimax Criterion for Analytic Rotation in Factor Analysis', *Psychometrika*, **23**, 187–200 (1958).
68. P. Geladi, B.R. Kowalski, 'Partial Least-squares Regression: A Tutorial', *Anal. Chim. Acta*, **185**, 1–17 (1986).
69. S. Wold, P. Geladi, K. Esbensen, J. Ohman, 'Principal Components- and PLS-analyses Generalized to Multiway (Multi-order) Data Arrays', *J. Chemom.*, **1**, 41–56 (1987).
70. R. Goodacre, M.J. Neal, D.B. Kell, 'Rapid and Quantitative Analysis of the Pyrolysis Mass Spectra of Complex Binary and Tertiary Mixtures Using Multivariate Calibration and Artificial Neural Networks', *Anal. Chem.*, **66**, 1070–1085 (1994).
71. T.J. McAvoy, H.T. Su, N.S. Wang, M. He, J. Horvath, H. Semerjian, 'A Comparison of Neural Networks and Partial Least-squares for Deconvoluting Fluorescence Spectra', *Bioltechnol. Bioeng.*, **40**, 53–62 (1992).
72. W.P. Carey, K.R. Beebe, E. Sanchez, P. Geladi, B. Kowalski, 'Chemometric Analysis of Multisensor Arrays', *Sens. Actuators*, **9**, 223–234 (1986).
73. L. Tucker, 'Some Mathematical Notes on Three-mode Factor Analysis', *Psychometrika*, **31**, 279–311 (1966).
74. E.R. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, John Wiley, New York, 1991.
75. S.D. Brown, 'Chemical Systems Under Indirect Observation: Latent Properties and Chemometrics', *Appl. Spectrosc.*, **49**(12), 14A–31A (1995).
76. H.B. Keller, D.L. Massart, 'Evolving Factor Analysis', *Chemom. Intell. Lab. Syst.*, **12**, 209–224 (1992).
77. P.K. Hopke, 'Target Transformation Factor Analysis', *Chemom. Intell. Lab. Syst.*, **6**, 7–19 (1989).
78. J. Toft, 'Evolutionary Rank Analysis Applied to Multidetectional Chromatographic Structures', *Chemom. Intell. Lab. Syst.*, **29**(2), 189–212 (1995).
79. F. Brakstad, 'The Feasibility of Latent Variables Applied to GC–MS Data', *Chemom. Intell. Lab. Syst.*, **29**, 157–176 (1995).
80. R. Bro, 'PARAFAC. Tutorial and Applications', *Chemom. Intell. Lab. Syst.*, **38**(2), 149–171 (1997).
81. H. Martens, T. Næs, *Multivariate Calibration*, John Wiley, Chichester, 1989.
82. H. Mark, *Principles and Practice of Spectroscopic Calibration*, John Wiley & Sons, New York, 1991.
83. H. Martens, T. Næs, *Multivariate Calibration*, John Wiley, Chichester, 1989.
84. C. Chatfield, A.J. Collins, *Introduction to Multivariate Analysis*, Chapman & Hall, London, 1980.
85. C.S. Gutteridge, L. Vallis, H.J.H. MacFie, 'Numerical Methods in the Classification of Microorganisms by Pyrolysis Mass Spectrometry', in *Computer-assisted Bacterial Systematics*, eds. M. Goodfellow, D. Jones, F. Priest, Academic Press, London, 369–401, 1985.
86. J.T. Magee, 'Whole-organism Fingerprinting', in *Handbook of New Bacterial Systematics*, eds. M. Goodfellow, A.G. O'Donnell, Academic Press, London, 383–427, 1993.

87. R. Goodacre, D.B. Kell, 'Pyrolysis Mass Spectrometry and its Applications in Biotechnology', *Curr. Opin. Biotechnol.*, **7**, 20–28 (1996).
88. I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
89. D.R. Causton, *A Biologist's Advanced Mathematics*, Allen and Unwin, London, 1987.
90. B.S. Everitt, *Cluster Analysis*, Edward Arnold, London, 1993.
91. H.L.C. Meuzelaar, J. Haverkamp, F.D. Hileman, *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*, Elsevier, Amsterdam, 1982.
92. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufmann, *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1988.
93. R.G. Brereton, *Multivariate Pattern Recognition in Chemometrics*, Elsevier, Amsterdam, 1992.
94. S.D. Brown, S.T. Sum, F. Despagne, B.K. Lavine, 'Chemometrics', *Anal. Chem.*, **68**, R21–R61 (1996).
95. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier, Amsterdam, 1997.
96. B.K. Lavine, 'Chemometrics', *Anal. Chem.*, **70**, R209–R228 (1998).
97. S. Wold, N. Kettaneh-Wold, B. Skagerberg, 'Nonlinear PLS Modelling', *Chem. Intell. Lab. Sys.*, **7**, 53–65 (1989).
98. A. Höskuldsson, 'Quadratic PLS Regression', *J. Chemom.*, **6**, 307–334 (1992).
99. V.M. Taavitsainen, P. Korhonen, 'Nonlinear Data Analysis with Latent Variable', *Chem. Intell. Lab. Sys.*, **14**, 185–194 (1992).
100. S. Wold, 'Nonlinear Partial Least-squares Modeling. 2. Spline Inner Relation', *Chem. Intell. Lab. Sys.*, **14**, 71–84 (1992).
101. B. Walczak, D.L. Massart, 'The Radial Basis Functions – Partial Least Squares Approach as a Flexible Non-linear Regression Technique', *Anal. Chim. Acta*, **331**, 177–185 (1996).
102. A. Berglund, S. Wold, 'INLR, Implicit Non-linear Latent Variable Regression', *J. Chemom.*, **11**, 141–156 (1997).
103. R.A. Heikka, K.T. Immonen, P.O. Minkkinen, E.Y.O. Paatero, T.O. Salmi, 'Determination of Acid Value, Hydroxyl Value and Water Content in Reactions Between Dicarboxylic Acids and Diols Using Near-infrared Spectroscopy and Non-linear Partial Least Squares Regression', *Anal. Chim. Acta*, **349**, 287–294 (1997).
104. P.D. Wasserman, *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York, 1989.
105. R. Goodacre, D.B. Kell, 'Rapid and Quantitative Analysis of Bioprocesses Using Pyrolysis Mass Spectrometry and Neural Networks – Application to Indole Production', *Anal. Chim. Acta*, **279**, 17–26 (1993).
106. J. Zupan, J. Gasteiger, *Neural Networks for Chemists: An Introduction*, VCH, Weinheim, 1993.
107. R. Goodacre, M.J. Neal, D.B. Kell, 'Rapid and Quantitative Analysis of the Pyrolysis Mass Spectra of Complex Binary and Tertiary Mixtures Using Multivariate Calibration and Artificial Neural Networks', *Anal. Chem.*, **66**, 1070–1085 (1994).
108. C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.
109. D.B. Kell, B. Sonnleitner, 'GMP – Good Modelling Practice: An Essential Component of Good Manufacturing Practice', *Trends Biotechnol.*, **13**, 481–492 (1995).
110. C. Wan, P.B. Harrington, 'Screening GC–MS Data for Carbamate Pesticides with Temperature Constrained-Cascade Correlation Networks', *Anal. Chim. Acta*, (in Press).
111. R. Goodacre, A.N. Edmonds, D.B. Kell, 'Quantitative Analysis of the Pyrolysis Mass Spectra of Complex Mixtures Using Artificial Neural Networks – Application to Amino Acids in Glycogen', *J. Anal. Appl. Pyrol.*, **26**, 93–114 (1993).
112. B.R. Kowalski, K.F. Faber, 'Comments on a Recent Sensitivity Analysis of Radial Base Function and Multilayer Feed-forward Neural Network Models', *Chem. Intell. Lab. Syst.*, **34**, 293–297 (1996).
113. R.C. Ebhart, I. Cloete, J.M. Zurada, 'Determining the Significance of Input Parameters Using Sensitivity Analysis', *Lecture Notes Computer Science*, **930**, 382–388 (1995).
114. P. Howes, N. Crook, 'Using Input Parameter Influences to Support the Decision of Feedforward Neural Networks', *Neurocomputing*, **24**, 191–206 (1999).
115. J.Y. Choi, C.H. Choi, 'Sensitivity Analysis of Multilayer Perceptron with Differentiable Activation Functions', *IEEE Trans. Neural Networks*, **3**, 101–107 (1992).
116. V.V. Kovalishyn, I.V. Tetko, A.I. Luik, V.V. Kholodovych, A.E.P. Villa, D.J. Livingstone, 'Neural Network Studies. 3. Variable Selection in the Cascade-correlation Learning Architecture', *J. Chem. Inf. Comput. Sci.*, **38**, 651–659 (1998).
117. A. Ikonomopoulos, 'Wavelet Decomposition and Radial Basis Function Networks for System Monitoring', *IEEE Trans. Neural Sci.*, **45**, 2293–2301 (1998).
118. A.H. Sung, 'Ranking Importance of Input Parameters of Neural Networks', *Expert System with Applications*, **15**, 405–411 (1998).
119. T.D. Gedeon, 'Data Mining of Inputs: Analyzing Magnitude and Functional Measures', *Int. J. Neural Syst.*, **8**, 209–218 (1997).
120. M. Koda, 'Stochastic Sensitivity Analysis Method for Neural-network Learning', *Int. J. Syst. Sci.*, **26**, 703–711 (1995).
121. P.B. Harrington, C. Wan, A. Urbas, 'Generalized Sensitivity Analysis of Neural Network Models', *J. Am. Chem. Soc.*, (submitted).

122. R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, Massachusetts, 1990.
123. J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, California, 1991.
124. T. Kohonen, *Self-organizing Maps*, ed. S.S.i.I. Sciences, Springer, Berlin, Heidelberg, New York, Vol. 30, 1997.
125. T. Kohonen, *Self-organization and Associative Memory*, Springer-Verlag, Berlin, 1989.
126. E. Oja, S. Kaski, *Kohonen Maps*, Elsevier, Amsterdam, 1999.
127. R. Goodacre, M.J. Neal, D.B. Kell, L.W. Greenham, W.C. Noble, R.G. Harvey, 'Rapid Identification Using Pyrolysis Mass Spectrometry and Artificial Neural Networks of *Propionibacterium acnes* Isolated from Dogs', *J. Appl. Bacteriol.*, **76**, 124–134 (1994).
128. R. Goodacre, J. Pygall, D.B. Kell, 'Plant Seed Classification Using Pyrolysis Mass Spectrometry with Unsupervised Learning; the Application of Auto-associative and Kohonen Artificial Neural Networks', *Chem. Intell. Lab. Sys.*, **34**, 69–83 (1996).
129. R. Goodacre, A.C. McGovern, N. Kaderbhai, E.A. Goodacre, 'Chemometric Analyses with Self Organizing Feature Maps: A Worked Example of the Analysis of Cosmetics Using Raman Spectroscopy', in *Kohonen Maps*, eds. E. Oja, S. Kaski, Elsevier, Amsterdam, 335–347, 1999.
130. E. Erwin, K. Obermayer, K. Schulten, 'Self-organizing maps: Ordering, Convergence Properties and Energy Functions', *Biol. Cyber.*, **67**, 47–55 (1992).
131. I. Belic, L. Gyergyek, 'Neural Network Methodologies for Mass Spectra Recognition', *Vacuum*, **48**, 633–637 (1997).
132. H. Lohninger, F. Stancl, 'Comparing the Performance of Neural Networks to Well Established Methods of Multivariate Data Analysis – the Classification of Mass Spectral Data', *Fresenius J. Anal. Chem.*, **344**, 186–189 (1992).
133. R. Goodacre, S.A. Howell, W.C. Noble, M.J. Neal, 'Subspecies Discrimination Using Pyrolysis Mass Spectrometry and Self-organizing Neural Networks of *Propionibacterium acnes* Isolated from Normal Human Skin', *Zbl. Bakt. – Int. J. Med. M.*, **284**, 501–515 (1996).
134. M.A. Kramer, 'Non Linear Principal Components Analysis Using Auto-associative Neural Networks', *AIChE J.*, **37**, 233–243 (1991).
135. M.A. Kramer, 'Autoassociative Neural Networks', *Comput. Chem. Eng.*, **16**, 313–328 (1992).
136. D.E. Rumelhart, J.L. McClelland, *Parallel Distributed Processing, Experiments in the Microstructure of Cognition*, PDP Research Group, MIT Press, Cambridge, MA, Vols. I and II, 1986.
137. P.J. Werbos, *The Roots of Back-propagation: From Ordered Derivatives to Neural Networks and Political Forecasting*, John Wiley, Chichester, 1994.
138. Y. Chauvin, D.E. Rumelhart, *Backpropagation: Theory, Architectures, and Applications*, Erlbaum, Hove, UK, 1995.
139. R. Goodacre, D.J. Rischert, P.M. Evans, D.B. Kell, 'Rapid Authentication of Animal Cell Lines Using Pyrolysis Mass Spectrometry and Auto-associative Artificial Neural Networks', *Cytotechnol.*, **21**, 231–241 (1996).
140. S.P. Jacobsson, 'Feature Extraction of Polysaccharides by Low-dimensional Internal Representation Neural Networks and Infrared Spectroscopy', *Anal. Chim. Acta*, **291**, 19–27 (1994).
141. G. Montague, J. Morris, 'Neural Network Contributions in Biotechnology', *Trends Biotechnol.*, **12**, 312–324 (1994).
142. D.R. Kuespert, T.J. McAvoy, 'Knowledge Extraction in Chemical Process Control', *Chem. Eng. Commun.*, **130**, 251–264 (1994).
143. B. Wabuyele, P.B. Harrington, 'Optimal Associative Memory for Baseline Correction of Infrared Spectra', *Anal. Chem.*, **66**, 2047–2051 (1994).
144. B. Wabuyele, P.B. Harrington, 'Quantitative Comparison of Optimal and Bidirectional Associative Memories for Background Prediction of Infrared Spectra', *Chemom. Intell. Lab. Syst.*, **29**, 51–61 (1995).
145. B. Wabuyele, P.B. Harrington, 'Fuzzy Optimal Associative Memories', *Appl. Spectrosc.*, **50**, 35–42 (1996).
146. W.G. Baxt, 'Application of Artificial Neural Networks to Clinical Medicine', *The Lancet*, **346**, 1135–1138 (1995).
147. A. Cawsey, *The Essence of Artificial Intelligence*, Prentice Hall, London, 1998.
148. R. Dybowski, V. Gant, 'Artificial Neural Networks in Pathological and Medical Laboratories', *Lancet*, **346**, 1203–1207 (1995).
149. R. Goodacre, M.J. Neal, D.B. Kell, 'Quantitative Analysis of Multivariate Data Using Artificial Neural Networks: A Tutorial Review and Applications to the Deconvolution of Pyrolysis Mass Spectra', *Zbl. Bakt. – Int. J. Med. M.*, **284**, 516–539 (1996).
150. S. Haykin, *Neural Networks*, Macmillan, New York, 1994.
151. B.D. Ripley, 'Neural Networks and Related Methods for Classification', *J. Roy. Stats. Soc. Ser. B*, **56**, 409–437 (1994).
152. B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, 1996.
153. M.D. Richard, R.P. Lippmann, 'Neural Network Classifiers Estimate Bayesian *a Posteriori* Probabilities', *Neural Computation*, **3**, 461–483 (1991).
154. P.K. Simpson, *Artificial Neural Systems*, Pergamon Press, Oxford, 1990.
155. S.S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan, New York, 1994.
156. G. Cybenko, 'Approximation by Superposition of a Sigmoidal Function', *Mathematical Control Signals Systems*, **2**, 303–314 (1989).

157. K. Funahashi, 'On the Approximate Realization of Continuous-mappings by Neural Networks', *Neural Networks*, **2**, 183–192 (1989).
158. K. Hornik, M. Stinchcombe, H. White, 'Multilayer Feedforward Networks are Universal Approximators', *Neural Networks*, **2**, 359–368 (1989).
159. K. Hornik, M. Stinchcombe, H. White, 'Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks', *Neural Networks*, **3**, 551–560 (1990).
160. H. White, 'Connectionist Nonparametric Regression – Multilayer Feedforward Networks Can Learn Arbitrary Mappings', *Neural Networks*, **3**, 535–549 (1990).
161. H. White, *Artificial Neural Networks: Approximation and Learning Theory*, Blackwell, Oxford, 1992.
162. P.B. Harrington, B.W. Pack, 'FLIN: Fuzzy Linear Interpolating Network', *Anal. Chim. Acta*, **277**, 189–197 (1993).
163. S.E. Fahlman, C. Lebiere, *The Cascade-correlation Learning Architecture*, Carnegie Mellon University, Pittsburgh, 1991.
164. P.B. Harrington, 'Fuzzy Rule-building Expert Systems: Minimal Neural Networks', *J. Chemom.*, **5**, 467–486 (1991).
165. P.B. Harrington, 'Temperature Constrained Backpropagation Neural Networks', *Anal. Chem.*, **66**, 802–807 (1994).
166. P.B. Harrington, 'Temperature-constrained Cascade Correlation Networks', *Anal. Chem.*, **70**, 1297–1306 (1998).
167. D.S. Broomhead, D. Lowe, 'Multivariable Functional Interpolation and Adaptive Networks', *Complex Systems*, **2**, 312–355 (1988).
168. J. Moody, C.J. Darken, 'Fast Learning in Networks of Locally-tuned Processing Units', *Neural Computation*, **1**, 281–294 (1989).
169. R. Beale, T. Jackson, *Neural Computing: An Introduction*, Adam Hilger, Bristol, 1990.
170. A. Saha, J.D. Keller, 'Algorithms for Better Representation and Faster Learning in Radial Basis Functions', in *Advances in Neural Information Processing Systems*, ed. D. Touretzky, Morgan Kaufmann Publishers, San Mateo, CA, 482–489, Vol. 2, 1990.
171. J. Park, I.W. Sandberg, 'Universal Approximation Using Radial Basis Function Networks', *Neural Computation*, **3**, 246–257 (1991).
172. D.R. Hush, B.G. Horne, 'Progress in Supervised Neural Networks – What's New Since Lippmann', *IEEE Signal Processing Magazine*, **10**, 8–39 (1993).
173. M.F. Wilkins, C.W. Morris, L. Boddy, 'A Comparison of Radial Basis Function and Backpropagation Neural Networks for Identification of Marine Phytoplankton from Multivariate Flow Cytometry Data', *Comput. Appl. Biosci.*, **10**, 285–294 (1994).
174. R.O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, (1973).
175. C. Wan, P.B. Harrington, 'Self-configuring Radial Basis Function Neural Network for Chemical Pattern Recognition', *J. Chem. Inf. Comput. Sci.*, **39**, 1049–1056 (1999).
176. R. Goodacre, 'Characterization and Quantification of Microbial Systems Using Pyrolysis Mass Spectrometry: Introducing Neural Networks to Analytical Pyrolysis', *Microbiol. Eur.*, **2**, 16–22 (1994).
177. R. Goodacre, D.B. Kell, G. Bianchi, 'Neural Networks and Olive Oil', *Nature*, **359**, 594–594 (1992).
178. R. Goodacre, D.B. Kell, G. Bianchi, 'Rapid Assessment of the Adulteration of Virgin Olive Oils by Other Seed Oils Using Pyrolysis Mass Spectrometry and Artificial Neural Networks', *J. Sci. Food Agric.*, **63**, 297–307 (1993).
179. C. Guillou, M. Lipp, B. Radovic, F. Reniero, M. Schmidt, E. Anklam, 'Use of Pyrolysis Mass Spectrometry in Food Analysis: Applications in the Food Analysis Laboratory of the European Commissions' Joint Research Centre', *J. Anal. Appl. Pyrol.*, **49**, 329–335 (1999).
180. G.J. Salter, M. Lazzari, L. Giansante, R. Goodacre, A. Jones, G. Surricchio, D.B. Kell, G. Bianchi, 'Determination of the Geographical Origin of Italian Extra Virgin Olive Oil Using Pyrolysis Mass Spectrometry and Artificial Neural Networks', *J. Anal. Appl. Pyrol.*, **40/41**, 159–170 (1997).
181. E. Anklam, M.R. Bassani, T. Eiberger, S. Kriebel, M. Lipp, R. Matissek, 'Characterization of Cocoa Butters and Other Vegetable Fats by Pyrolysis Mass Spectrometry', *Fresenius J. Anal. Chem.*, **357**, 981–984 (1997).
182. B.S. Radovic, M. Lipp, E. Anklam, 'Classification of Cocoa Butters Using Pyrolysis Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**, 783–789 (1998).
183. R. Goodacre, 'Use of Pyrolysis Mass Spectrometry with Supervised Learning for the Assessment of the Adulteration of Milk of Different Species', *Appl. Spectrosc.*, **51**, 1144–1153 (1997).
184. M.A.E. Schmidt, B.S. Radovic, M. Lipp, G. Harzer, E. Anklam, 'Characterization of Milk Samples with Various Whey Protein Contents by Pyrolysis–Mass Spectrometry (Py–MS)', *Food Chem.*, **65**, 123–128 (1999).
185. J.L. Berdague, C. Rabot, M. Bonneau, 'Rapid Classification of Backfat Samples Selected According to Their Androstenone Content Pyrolysis–Mass Spectrometry', *Sciences Des Aliments*, **16**, 425–433 (1996).
186. L. Montanarella, M.R. Bassani, O. Breas, 'Chemometric Classification of some European Wines Using Pyrolysis Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1589–1593 (1995).
187. E. Anklam, M. Lipp, B. Radovic, E. Chiavaro, G. Palla, 'Characterization of Italian Vinegar by Pyrolysis Mass Spectrometry and a Sensor Device ('Electronic Nose')', *Food Chem.*, **61**, 243–248 (1998).
188. R. Goodacre, D. Hammond, D.B. Kell, 'Quantitative Analysis of the Adulteration of Orange Juice with

Sucrose Using Pyrolysis Mass Spectrometry and Chemometrics', *J. Anal. Appl. Pyrol.*, **40/41**, 135–158 (1997).
189. R. Goodacre, R.J. Gilbert, 'The Detection of Caffeine in a Variety of Beverages Using Curie-point Pyrolysis Mass Spectrometry and Genetic Programming', *Analyst*, **124**, 1069–1074 (1999).
190. R.G.W. Kenyon, E.V. Ferguson, A.C. Ward, 'Application of Neural Networks to the Analysis of Pyrolysis Mass Spectra', *Zentralblatt fur Bakteriologie–Int. J. Med. Microbiol. Virol. Parasitol. Infect. Dis.*, **285**, 267–277 (1997).
191. P.R. Sisson, R. Freeman, D. Law, A.C. Ward, N.F. Lightfoot, 'Rapid Detection of Verocytotoxin Production Status in *Escherichia coli* by Artificial Neural Network Analysis of Pyrolysis Mass Spectra', *J. Anal. Appl. Pyrol.*, **32**, 179–185 (1995).
192. R. Freeman, R. Goodacre, P.R. Sisson, J.G. Magee, A.C. Ward, N.F. Lightfoot, 'Rapid Identification of Species within the *Mycobacterium tuberculosis* Complex by Artificial Neural Network Analysis of Pyrolysis Mass Spectra', *J. Med. Microbiol.*, **40**, 170–173 (1994).
193. J. Chun, E. Atalan, A.C. Ward, M. Goodfellow, 'Artificial Neural Network Analysis of Pyrolysis Mass Spectrometric Data in the Identification of *Streptomyces* Strains', *FEMS Microbiol. Lett.*, **107**, 321–325 (1993).
194. J. Chun, E. Atalan, S.B. Kim, H.J. Kim, M.E. Hamid, M.E. Trujillo, J.G. Magee, G.P. Manfio, A.C. Ward, M. Goodfellow, 'Rapid Identification of *Streptomycetes* by Artificial Neural Network Analysis of Pyrolysis Mass Spectra', *FEMS Microbiol. Lett.*, **114**, 115–119 (1993).
195. J.S. Chun, A.C. Ward, S.O. Kang, Y.C. Hah, M. Goodfellow, 'Long-term Identification of Streptomycetes Using Pyrolysis Mass Spectrometry and Artificial Neural Networks', *Zentralblatt fur Bakteriologie–Int. J. Med. Microbiol. Virol. Parasitol. Infect. Dis.*, **285**, 258–266 (1997).
196. R. Goodacre, S.J. Hiom, S.L. Cheeseman, D. Murdoch, A.J. Weightman, W.G. Wade, 'Identification and Discrimination of Oral Asaccharolytic *Eubacterium* spp. Using Pyrolysis Mass Spectrometry and Artificial Neural Networks', *Cur. Microbiol.*, **32**, 77–84 (1996).
197. T. Nilsson, M.R. Bassani, T.O. Larsen, L. Montanarella, 'Classification of Species in the Genus *Penicillium* by Curie Point Pyrolysis Mass Spectrometry Followed by Multivariate Analysis and Artificial Neural Networks', *J. Mass Spectrom.*, **31**, 1422–1428 (1996).
198. R. Goodacre, P.J. Rooney, D.B. Kell, 'Discrimination Between Methicillin-resistant and Methicillin-susceptible *Staphylococcus aureus* Using Pyrolysis Mass Spectrometry and Artificial Neural Networks', *J. Antimicr. Chemother.*, **41**, 27–34 (1998).
199. P.E. Reynolds, C. Fuller, 'Methicillin Resistant Strains of *Staphylococcus aureus*; Presence of Identical Additional Penicillin Binding Protein in all Strains Examined', *FEMS Microbiol. Lett.*, **33**, 251–254 (1986).
200. D.M. O'Hara, P.E. Reynolds, 'Antibody Used to Identify Penicillin Binding Protein 2' in Methicillin Resistant Strains of *Staphylococcus aureus* (MRSA)', *FEBS Lett.*, **212**, 237–241 (1987).
201. R. Goodacre, B. Shann, R.J. Gilbert, É.M. Timmins, A.C. McGovern, B.K. Alsberg, D.B. Kell, N.A. Logan, 'The Detection of the Dipicolinic Acid Biomarker in *Bacillus* Spores Using Curie-point Pyrolysis Mass Spectrometry and Fourier Transform Infrared Spectroscopy', *Anal. Chem.*, **72**, 119–127 (2000).
202. S.G. Kang, R.G.W. Kenyon, A.C. Ward, K.J. Lee, 'Analysis of Differentiation State in *Streptomyces albidoflavus* SMF301 by the Combination of Pyrolysis Mass Spectrometry and Neural Networks', *J. Biotechnol.*, **62**, 1–10 (1998).
203. S.G. Kang, D.H. Lee, A.C. Ward, K.J. Lee, 'Rapid and Quantitative Analysis of Clavulanic Acid Production by the Combination of Pyrolysis Mass Spectrometry and Artificial Neural Network', *J. Microbiol. Biotechnol.*, **8**, 523–530 (1998).
204. R. Goodacre, É.M. Timmins, R. Burton, N. Kaderbhai, A.M. Woodward, D.B. Kell, P.J. Rooney, 'Rapid Identification of Urinary Tract Infection Bacteria Using Hyperspectral, Whole Organism Fingerprinting and Artificial Neural Networks', *Microbiology*, **144**, 1157–1170 (1998).
205. P.B. Harrington, 'Minimal Neural Networks: Differentiation of Classification Entropy', *Chem. Intell. Lab. Syst.*, **19**, 143–154 (1993).
206. P.D. Harrington, 'Minimal Neural Networks – Concerted Optimization of Multiple Decision Planes', *Chem. Intell. Lab. Syst.*, **18**, 157–170 (1993).
207. P.C. Jurs, B.R. Kowalski, T.L. Isenhour, 'Computerized Learning Machines Applied To Chemical Problems – Molecular Formula Determination From Low Resolution Mass Spectrometry', *Anal. Chem.*, **41**, 21–27 (1969).
208. B. Curry, D.E. Rumelhart, 'MSNet: A Neural Network That Classifies Mass Spectra', *Tetrahedron Comput. Methodol.*, **3**, 213 (1990).
209. W. Werther, H. Lohninger, F. Stancl, K. Varmuza, 'Classification of Mass Spectra – A Comparison of Yes/no Classification Methods for the Recognition of Simple Structural Properties', *Chemom. Intell. Lab. Syst.*, **22**, 63–76 (1994).
210. G. Blomquist, E. Johansson, B. Soderstrom, S. Wold, 'Data Analysis of Pyrolysis-Chromatograms by Means of SIMCA Pattern Recognition', *J. Anal. Appl. Pyrolysis*, **1**, 53–65 (1979).
211. R. Todeschini, 'K-nearest Neighbor Method: The Influence of Data Transformations and Metrics', *Chemom. Intell. Lab. Syst.*, **6**, 213–220 (1989).
212. R.H. Lindeman, P.F. Merenda, R.Z. Gold, *Introduction to Bivariate and Multivariate Analysis*, Scott, Foresman & Co., New York, 1980.

213. P.J. Tandler, J.A. Butcher, H. Tao, P.B. Harrington, 'Chemometric Analysis of Plastic Recycling Products', *Anal. Chim. Acta*, **312**, 231–244 (1995).
214. B.K. Alsberg, R. Goodacre, J.J. Rowland, D.B. Kell, 'Classification of Pyrolysis Mass Spectra by Fuzzy Multivariate Rule Induction – Comparison with Regression, K-nearest Neighbor, Neural and Decision-tree Methods', *Anal. Chim. Acta*, **348**, 389–407 (1997).
215. B.K. Alsberg, W.G. Wade, R. Goodacre, 'Chemometric Analysis of Diffuse Reflectance–Absorbance Fourier Transform Infrared Spectra Using Rule Induction Methods: Application to the Classification of *Eubacterium* Species', *Appl. Spectrosc.*, **52**, 823–832 (1998).
216. C. Cai, P.B. Harrington, 'Prediction of Substructure and Toxicity of Pesticides with Temperature Constrained-cascade Correlation Network from Low-resolution Mass Spectra', *Anal. Chem.*, **71**, 4134–4141 (1999).
217. É.M. Timmins, R. Goodacre, 'Rapid Quantitative Analysis of Binary Mixtures of *Escherichia coli* Strains Using Pyrolysis Mass Spectrometry with Multivariate Calibration and Artificial Neural Networks', *J. Appl. Microbiol.*, **83**, 208–218 (1997).
218. R. Goodacre, A. Karim, M.A. Kaderbhai, D.B. Kell, 'Rapid and Quantitative Analysis of Recombinant Protein Expression Using Pyrolysis Mass Spectrometry and Artificial Neural Networks – Application to Mammalian Cytochrome b_5 in *Escherichia coli*', *J. Biotechnol.*, **34**, 185–193 (1994).
219. R. Goodacre, S. Trew, C. Wrigley-Jones, M.J. Neal, J. Maddock, T.W. Ottley, N. Porter, D.B. Kell, 'Rapid Screening for Metabolite Overproduction in Fermentor Broths Using Pyrolysis Mass Spectrometry with Multivariate Calibration and Artificial Neural Networks', *Biotechnol. Bioeng.*, **44**, 1205–1216 (1994).
220. R. Goodacre, S. Trew, C. Wrigley-Jones, G. Saunders, M.J. Neal, N. Porter, D.B. Kell, 'Rapid and Quantitative Analysis of Metabolites in Fermentor Broths Using Pyrolysis Mass Spectrometry with Supervised Learning: Application to the Screening of *Penicillium chrysogenum* Fermentations for the Overproduction of Penicillins', *Anal. Chim. Acta*, **313**, 25–43 (1995).
221. A.C. McGovern, D. Broadhurst, J. Taylor, R.J. Gilbert, N. Kaderbhai, M.K. Winson, D.A. Small, J.J. Rowlard, D.B. Kell, R. Goodacre, 'Monitoring of Complex Industrial Bioprocesses for Metabolite Concentrations Using Modern Spectroscopies and Machine Learning: Application to Gibberellic Acid Production', *Biotechnol. Bioeng.*, (submitted).
222. A.C. McGovern, R. Ernill, B.V. Kara, D.B. Kell, R. Goodacre, 'Rapid Analysis of the Expression of Heterologous Proteins in *Escherichia coli* Using Pyrolysis Mass Spectrometry and Fourier Transform Infrared Spectroscopy with Chemometrics: Application to α2-interferon Production', *J. Biotechnol.*, **72**, 157–167 (1999).
223. A.D. Shaw, M.K. Winson, A.M. Woodward, A.C. McGovern, H.M. Dowey, N. Kaderbhai, D. Broadhurst, R.J. Gilbert, J. Taylor, E.M. Timmins, B.K. Alsberg, J.J. Rowland, R. Goodacre, D.B. Kell, 'Rapid Analysis of High-dimensional Bioprocesses Using Multivariate Spectroscopies and Advanced Chemometrics', in *Advances in Biochemical Engineering/Biotechnology*, ed. T. Scheper, Springer-Verlag, Berlin, 83–114, Vol. 66, 2000.
224. M.K. Winson, R. Goodacre, A.M. Woodward, É.M. Timmins, A. Jones, B.K. Alsberg, J.J. Rowland, D.B. Kell, 'Diffuse Reflectance Absorbance Spectroscopy Taking in Chemometrics (DRASTIC). A Hyperspectral FT–IR-based Approach to Rapid Screening for Metabolite Overproduction', *Anal. Chim. Acta*, **348**, 273–282 (1997).
225. A.D. Shaw, N. Kaderbhai, A. Jones, A.M. Woodward, R. Goodacre, J.J. Rowland, D.B. Kell, 'Noninvasive, On-line Monitoring of the Biotransformation by Yeast of Glucose to Ethanol Using Dispersive Raman Spectroscopy and Chemometrics', *Appl. Spectrosc.*, **53**, 1419–1428 (1999).
226. A.M. Woodward, A. Jones, X.Z. Zhang, J. Rowland, D.B. Kell, 'Rapid and Non-invasive Quantification of Metabolic Substrates in Biological Cell Suspensions Using Non-linear Dielectric Spectroscopy with Multivariate Calibration and Artificial Neural Networks. Principles and Applications', *Bioelectrochem. Bioenerg.*, **40**, 99–132 (1996).
227. H.M. Davey, D.B. Kell, 'Flow Cytometry and Cell Sorting of Heterogeneous Microbial Populations – the Importance of Single Cell Analyses', *Microbiol. Rev.*, **60**, 641–696 (1996).
228. M. Goodfellow, 'Inter-strain Comparison of Pathogenic Microorganisms by Pyrolysis Mass Spectrometry', *Binary – Comp. Microbiol.*, **7**, 54–60 (1995).
229. R. Goodacre, D.B. Kell, *Composition Analysis*, UK Patent, 1995; International Patent #WO 96/42058 of 27 December, 1996; US5946640 of 31 August, 1999.
230. R. Goodacre, D.B. Kell, 'Correction of Mass Spectral Drift Using Artificial Neural Networks', *Anal. Chem.*, **68**, 271–280 (1996).
231. R. Goodacre, É.M. Timmins, A. Jones, D.B. Kell, J. Maddock, M.L. Heginbothom, J.T. Magee, 'On Mass Spectrometer Instrument Standardization and Interlaboratory Calibration Transfer Using Neural Networks', *Anal. Chim. Acta*, **348**, 511–532 (1997).
232. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vettering, *Numerical Recipes*, Cambridge University Press, Cambridge, 1986.
233. W. Spendley, G.R. Hext, F.R. Himsworth, *Technometrics*, **4**, 441–461 (1962).
234. M.J.D. Powell, *Comput J.*, **7**, 155–162 (1965).
235. J.A. Nelder, R. Mead, *Comput. J.*, **7**, 308–313 (1965).

236. S.M. Colby, T.B. King, J.P. Reilly, *Rapid Commun. Mass Spectrom.*, **8**, 865–868 (1994).
237. R.S. Brown, J.J. Lennon, *Anal. Chem.*, **67**, 1998–2003 (1995).
238. R.M. Whittal, L. Li, *Anal. Chem.*, **67**, 1950–1954 (1995).
239. M.L. Vestal, P. Huhasz, S.A. Martin, *Rapid Commun. Mass Spectrom.*, **9**, 1044–1050 (1995).
240. C.C. Vera, R. Zubarev, H. Ehring, P. Hakansson, B.U.R. Sunqvist, *Rapid Comm. Mass Spectrom.*, **10**, 1429–1432 (1996).
241. P. Juhasz, M.L. Vestal, S.A. Martin, *J. Am. Soc. Mass Spectrom.*, **8**, 209–217 (1997).
242. M. Vestal, P. Juhasz, *J. Am. Soc. Mass Spectrom.*, **9**, 892–911 (1998).
243. R.D. Edmonson, D.H. Russell, *J. Am. Soc. Mass Spectrom.*, **7**, 995–1001 (1996).
244. R.J. Arnold, J.P. Reilly, *J. Am. Chem. Soc.*, **120**, 1528–1532 (1998).

Atmospheric Pressure Ionization Mass Spectrometry

Matthew A. Tarr, Junhua Zhu, and Richard B. Cole
University of New Orleans, New Orleans, USA

1 Atmospheric Pressure Chemical Ionization/Mass Spectrometry **1**
1.1 Atmospheric Pressure Chemical Ionization Design and Function 2
1.2 Practical Use 7
1.3 Applications 11
2 Electrospray Ionization Mass Spectrometry **13**
2.1 Fundamental Mechanism of Electrospray Ionization 15
2.2 Instrumentation 18
2.3 The Electrospray Ionization Mass Spectrum 21
2.4 Applications 25
3 Summary **26**
Abbreviations and Acronyms **26**
Related Article **26**
References **27**

The atmospheric pressure ionization (API) techniques serve to transform neutral analytes into ionic species that may then be analyzed by mass spectrometry (MS). Ionization takes place under ambient pressure (~1 atm), and formed ions are subsequently directed into the low-pressure regions of the mass spectrometer for mass analysis. The API techniques have offered notable successes in enabling ionization of nonvolatile compounds contained in solution that are not amenable to ''classical'' mass spectrometric ionization techniques such as electron ionization and conventional chemical ionization. Under favorable API conditions, neutral analytes may be converted to ionic forms in extremely high efficiencies approaching 100%. Fundamental aspects and applications of the two most widely used API techniques, atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI), are discussed at length in this article.

1 ATMOSPHERIC PRESSURE CHEMICAL IONIZATION/MASS SPECTROMETRY

APCI is a process by which initially dissolved solutes are ionized, after desolvation, through gas-phase ion–molecule reactions at atmospheric pressure. The resulting ions are then mass analyzed in an appropriate mass spectrometer. APCI generally utilizes continuous introduction of liquid samples at flow rates of 0.2–2 mL min^{-1}, and the technique is often used to interface liquid chromatography (LC) with MS. APCI can accommodate a wide range of solvents and solutes, and is one of the most widely applicable methods for the ionization of dissolved analytes. Molecular or quasi-molecular ions (e.g. $M^{+\bullet}$, $M^{-\bullet}$, $[M + H]^+$, and $[M - H]^-$) as well as adduct ions are typically formed in APCI. Generally, there is only slight to moderate fragmentation. To some extent, the degree of fragmentation can be adjusted, and the type of molecular or quasi-molecular ion can be altered by varying the ionization conditions. APCI can be divided into three general steps: nebulization/desolvation, ionization, and extraction. Each of these steps will be discussed in more detail below.

API was first developed in the early 1970s.[1–4] Since then, extensive reviews have appeared on various API methods.[5–7] The number of publications on APCI has increased steadily through the 1990s. The range of applicability of atmospheric pressure chemical ionization/mass spectrometry (APCI/MS) lies between that of gas chromatography/mass spectrometry (GC/MS) and electrospray ionization/mass spectrometry (ESI/MS). GC/MS is generally applicable to relatively low-boiling, thermally stable compounds with fairly low polarities and low molecular weights. On the other end of the spectrum, ESI/MS typically produces high signals for compounds, including large molecules, that exist in ionic forms in solution. APCI is applicable to all compounds between these two groups: slightly polar to highly polar with intermediate sizes and a broad range of boiling points. Furthermore, thermally labile compounds are generally amenable to APCI, except for a few highly unstable species. Figure 1 presents a qualitative depiction of the useful polarity ranges for various MS sample introduction methods.

The two major requirements for APCI are that a compound be soluble in a solvent and the compound must undergo some gas-phase reaction resulting in ionization. Because APCI can utilize most solvents and because ionization reactions can be adjusted by altering the conditions, APCI is widely applicable to most analytes. The major application of APCI is in interfacing LC with MS. Liquid chromatography/atmospheric pressure chemical ionization/mass spectrometry (LC/APCI/MS) analyses

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

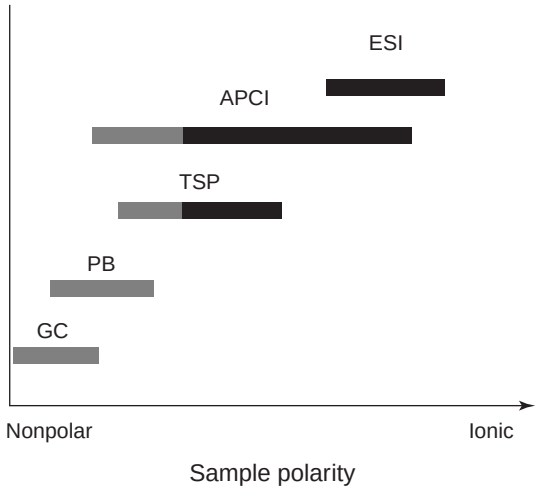

Figure 1 Applicability of MS inlet systems as a function of sample polarity. TSP, thermospray; PB, particle beam. (Reproduced with permission from Herderich et al.[8])

have been reported for environmental samples (pesticides, polycyclic aromatic hydrocarbons (PAHs), etc.), for agricultural and food samples (pesticides, vitamins, pigments, flavorings, fats, etc.) and for biomedical samples (drugs and metabolites, steroids, biomolecules, etc.), along with other applications.

The remainder of this article will detail the design, principles, and function of APCI interfaces. Practical considerations for the use of APCI will also be discussed, and examples of actual applications will be presented. These examples will illustrate experimental conditions used for the ionization of various compounds, and will indicate the capabilities and limitations of APCI.

1.1 Atmospheric Pressure Chemical Ionization Design and Function

APCI involves three primary steps: conversion of the analyte to a gas-phase species (nebulization and desolvation); ionization of the analyte through gas-phase ion–molecule reactions at atmospheric pressure; and extraction of the ions into the mass spectrometer with simultaneous exclusion of most of the nonionized species.

In some cases, the sample is already in the gas phase, and conversion is not necessary. Examples include direct sampling of atmospheric gases[3,9] and the introduction of gas chromatographic effluents into the APCI source. Despite the utility of these applications, APCI has arisen primarily as a means of interfacing high-performance liquid chromatography (HPLC) to MS. These two methods are inherently incompatible due to the high density of liquids used in HPLC compared to the low-density vacuum required for MS. Therefore, a suitable interface is needed to ionize analytes and to eliminate the high volume of solvent molecules present. Efforts to interface HPLC with MS date to around the early 1980s, when moving-belt interfaces[10] and direct liquid introduction were first developed.[11] Although these devices succeeded in producing mass spectra from HPLC effluents, they were not practical. A number of devices were designed subsequently that relied on nebulization of the liquid followed by desolvation and ionization processes. These devices include thermospray,[12] electrospray,[13] particle beam,[14] and APCI. Today, APCI and electrospray predominate the interfaces used for the introduction of liquid samples into mass spectrometers. A schematic of an APCI interface is presented in Figure 2. The details of each section of this interface will be discussed below.

1.1.1 Nebulization

The first step in API is nebulization of the liquid. In this step, the liquid is converted to small droplets, generally through interaction with a high-velocity gas with simultaneous heating. By breaking the liquid into small droplets and providing sufficient heat, the liquid sample is converted almost entirely into gas-phase species. A typical design for an APCI nebulizer consists of a fused silica capillary containing the liquid flow surrounded by a concentric, high-velocity flow of a nebulizing gas. In addition, a sheath gas surrounding the nebulizer may be utilized. Typically nitrogen is used as the nebulizer gas. For the interface in Figure 2, the nebulizer manifold is heated to 500–700 °C, with a maximum temperature of 1000 °C. Although these temperatures may seem high enough to cause significant thermal decomposition of analytes, the bulk of the energy is consumed in vaporization of the solvent. The actual temperature experienced by an analyte molecule is therefore significantly lower, and thermal degradation is generally not significant. Operation of the interface is compatible with common reversed-phase HPLC solvents, such as methanol, acetonitrile, and water. More discussion on solvent systems will be given in section 1.2.1.

The APCI nebulizer is quite similar to the design of thermospray nebulizers, which rely strictly on heating to induce vaporization of solvent and analytes usually in the presence of a volatile buffer additive. However, in thermospray, ionization of the analyte typically proceeds via interaction with the buffer salt and requires no supplemental electron process (such as a corona discharge as used in APCI).

1.1.2 Ionization

Following nebulization, any remaining condensed-phase droplets are evaporated in the heated region just after the nebulizer. Subsequent to this desolvation region is where the ionization process occurs. Ionization is

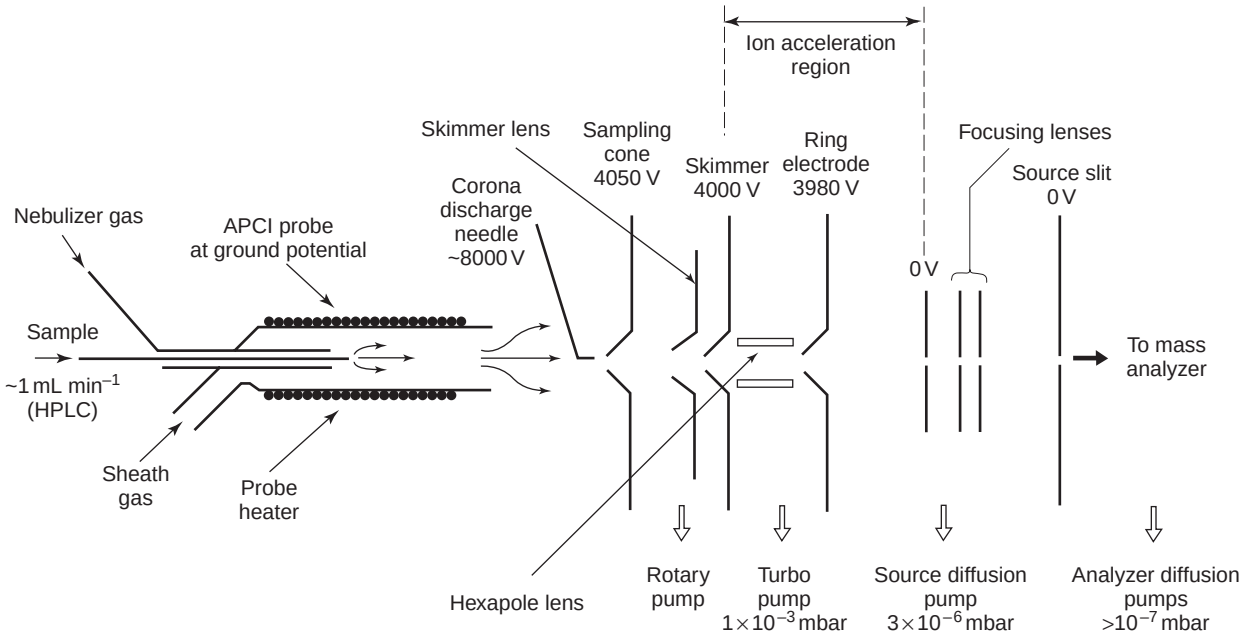

Figure 2 Diagram of an APCI interface. (Reproduced with permission from Micromass Autospec Operator's Manual.)

initiated by a corona discharge, although early instrument designs relied on a radioactive β-particle source instead of a corona discharge.[1,2,15] The corona discharge is produced by a needle held at high voltage within the gas stream. The high potential between the needle and the sampling cone results in the emission of electrons that are accelerated in the field and initiate a gas discharge. For nitrogen, which is the most commonly used nebulizer gas, the discharge results in the following initial ionization processes (Equations 1–6) in the presence of water:[5]

$$N_2 + e \longrightarrow N_2^{+\bullet} + 2e \quad (1)$$

$$N_2^{+\bullet} + 2N_2 \longrightarrow N_4^{+\bullet} + N_2 \quad (2)$$

$$N_4^{+\bullet} + H_2O \longrightarrow H_2O^{+\bullet} + 2N_2 \quad (3)$$

$$H_2O^{+\bullet} + H_2O \longrightarrow H_3O^{+} + HO^{\bullet} \quad (4)$$

$$H_3O^{+} + H_2O + N_2 \longrightarrow H^{+}(H_2O)_2 + N_2 \quad (5)$$

$$H^{+}(H_2O)_{n-1} + H_2O + N_2 \longrightarrow H^{+}(H_2O)_n + N_2 \quad (6)$$

In addition, species such as $N^{+\bullet}$ and $N_3^{+\bullet}$ are formed, and in the presence of oxygen, $O_2^{+\bullet}$ and various nitrogen oxide ions are also formed.[5] The exact nature of the ions formed is dependent on the gaseous species present. In fact, addition of reagents to the liquid phase or nebulizer gas can be used to alter the ions present. This topic will be discussed in more detail below.

The ion–molecule reactions observed in APCI are in many ways similar to those observed for chemical ionization. However, because chemical ionization is conducted at much lower pressures (e.g. $\sim 10^{-3}$ Torr), the reaction products are typically governed by kinetics, i.e. the most rapidly formed products are observed in highest abundances. In contrast, the APCI process generally results in the formation of the most thermodynamically stable product. Because APCI is carried out at atmospheric pressure, numerous gas-phase collisions occur, allowing equilibrium to be approached. The time allowed for equilibration can be altered by adjusting the position of the corona discharge needle. When the corona is relatively far from the first sampling aperture, equilibrium is generally achieved. If the corona is initiated close to the sampling aperture, much less time is available for equilibration, and kinetic products tend to predominate. This phenomenon occurs because the gas rapidly expands into the low-pressure region behind the sampling aperture. Once expansion occurs, collisions are minimized and only unimolecular decompositions take place, so that equilibration can no longer occur. Therefore, the closer the discharge is to the sampling aperture, the less time is available at atmospheric pressure to obtain equilibrium. The profile of ions observed for a corona discharge in nitrogen at different distances from the sampling aperture is illustrated in Figure 3. A clear change in the predominant ions is observed, corresponding to a shift from kinetic products to thermodynamic products.

With the continuous presence of solvent- or nebulizer-gas-reactive ions in the source region, introduction of an analyte leads to ion–molecule reactions between the analyte and the ions present. The widespread applicability of APCI to many compounds is the result of the ability to find reaction conditions that produce

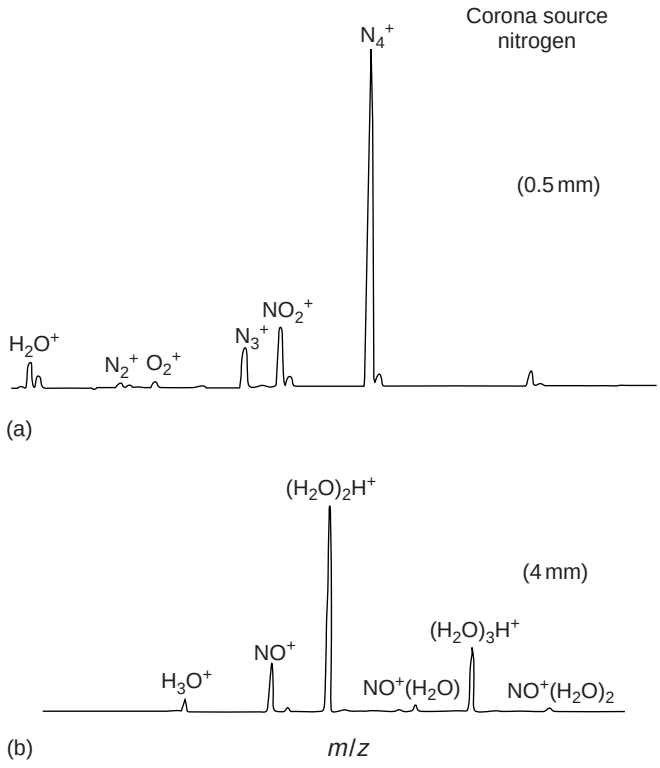

Figure 3 Ions observed for nitrogen in a corona discharge with the corona needle held at different distances from the sampling aperture: (a) 0.5 mm, ions are not thermally equilibrated; (b) 4 mm, ions are thermally equilibrated. (Reproduced with permission from Carroll et al.[5])

ionic forms of the analyte. A number of reaction mechanisms have been described for the formation of both positive and negative ions. Positive ion formation proceeds by several processes, including: proton transfer, charge transfer, hydride abstraction, cation addition, and ipso-substitution.[5] Negative ions are also formed by a number of processes, including: electron attachment, proton abstraction, reaction with oxygen or superoxide ions, and halide ion attachment.[5] A description of each process is given below.

Proton transfer is one of the commonly observed mechanisms for positive ion formation. This process is essentially a gas-phase acid–base reaction, and therefore the reaction can be characterized by the gas-phase basicity (GPB) and gas-phase acidity (GPA) of the two species involved. In theory, GPB is defined as the free energy, ΔG, for the process (Equations 7 and 8):[16]

$$M + H^+ \longrightarrow [M + H]^+ \quad (7)$$

$$\Delta G = \Delta H - T\Delta S \quad (8)$$

The more negative the value of the free energy, the more favorable is the forward reaction. The free energy is a combination of the energy change due to enthalpy change (ΔH) and entropy change ($T\Delta S$). Commonly, proton affinity (PA) is used instead of GPB, where $PA = -\Delta H$. In this manner, a larger PA value indicates a higher propensity to capture protons (i.e. Equation 7 is more exothermic). In general the entropy term is quite small compared to the enthalpy term, and is therefore ignored. Most organic compounds have a PA in the range 550–1000 kJ mol^{-1}.[16] The PA values for several common reagent species are given in Table 1.

Equation (7), however, does not generally occur in APCI because free protons are not usually present. Instead, transfer of a proton between two species occurs by a reaction such as that shown in Equation (9):

$$M + RH^+ \longrightarrow MH^+ + R \quad (9)$$

where M represents the gas-phase analyte and RH^+ represents a gas-phase protonated reagent molecule present in the source. Equation (9) can be written as the sum of forward and reverse GPB reactions (Equations 10 and 11):

$$M + H^+ \longrightarrow MH^+ \text{ (forward GPB reaction)} \quad (10)$$

$$RH^+ \longrightarrow R + H^+ \text{ (reverse GPB reaction)} \quad (11)$$

Equation (9) will occur spontaneously if M is a stronger gas-phase base than R. This condition can be approximated by $PA_M > PA_R$, and indicates that the sum of Equations (10) and (11) is energetically favorable. If the difference in GPBs is large (i.e. large negative ΔG), then the products are likely to have high internal energies,[16] possibly leading to fragmentation. Another parameter that is often used to determine the acid–base properties of a gas-phase molecule is the GPA. The GPA is the free energy change for the reaction depicted in Equation (12):

$$M \longrightarrow [M - H]^- + H^+ \quad (12)$$

Table 1 PA for several compounds

Compound	PA (kJ mol^{-1})
H_2O	697
HCOOH	748
CH_3OH	761
CH_3CN	787
CH_3COOH	796
$(CH_3)_2CO$	823
Tetrahydrofuran	831
NH_3	854
$(CH_3)_2NH$	923
$(CH_3)_3N$	942
$(CH_3CH_2)_3N$	972
$(CH_3CH_2CH_2)_3N$	979

Data from van Baar.[16]

Table 2 GPA for several compounds

Compound	GPA ($kJ\,mol^{-1}$)
HBr	1331
HCl	1372
HCOOH	1415
HNCO	1415
CH_3COOH	1429
HCN	1438
$CHCl_3$	1461
$(CH_3)_2CO$	1514
CH_3CN	1528
HF	1530
CH_2Cl_2	1535
$(CH_3)_2CHOH$	1543
CH_3OH	1565
H_2O	1607
NH_3	1657

Data from van Baar.[16]

The GPAs for several compounds are listed in Table 2.

Unlike low-pressure chemical ionization (e.g. $\sim 10^{-3}$ Torr), in APCI the internal energy resulting from a chemical reaction can be dissipated by collisions with neutral species present at atmospheric pressure, such as N_2.[16] In general, the chemical ionization processes occurring in APCI are somewhat selective. Therefore, depending upon the reagent ions present in the source, a molecule may or may not be ionized. This property may be either a potential advantage or perhaps a disadvantage. The advantage comes in the possibility of selectively ionizing compounds of interest but not ionizing other species. The disadvantage is that an analyte species may not be effectively ionized without special conditions. Despite this limitation, APCI is generally applicable to a wide range of compounds under the appropriate conditions.

Charge transfer ionization proceeds via the pathway shown in Equation (13):

$$M + R^{+\bullet} \longrightarrow M^{+\bullet} + R \qquad (13)$$

This reaction will proceed spontaneously if the ionization potential for M is less than that for R. The use of aprotic reagents (e.g. CO_2, Ar) can favor charge transfer by inhibiting proton transfer, which is generally a competing process in the presence of a protic solvent.[16] In addition, when present, protonated cluster ions such as $H^+(H_2O)_2$ do not generally undergo charge transfer reactions.[5] Some reagent species are capable of both proton transfer and charge exchange, resulting in the formation of both MH^+ and $M^{+\bullet}$ in the ion source. Reagents such as water, methanol, ethanol, methane, ethane, propane, and isobutane favor proton transfer reactions.[5]

Hydride abstraction also results in the formation of positive ions by the reaction shown in Equation (14):

$$M + R^+ \longrightarrow [M - H]^+ + RH \qquad (14)$$

Ionization by addition generally involves the formation of adduct ions. For example, $[M + CH_3]^+$ ions were observed for hydroxy-1,4-naphthoquinone in the presence of methanol[17] and $[M + Na]^+$ ions and $[M + NH_4]^+$ ions have been observed.[18,19]

Ipso-substitution reactions, as illustrated in Equation (15), have been described for aromatic compounds:[5]

$$C_6H_5Cl + NH_3{}^{+\bullet} \longrightarrow C_6H_5NH_3{}^+ + Cl^{\bullet} \qquad (15)$$

A thorough review of positive ion reagents for low-pressure chemical ionization has been published.[20] Although this review is not completely applicable to APCI because solvent molecules may predominate over added reagents, it still provides a useful background on gas-phase reactions for numerous reagents.

In addition to the cation formation reactions listed above, negative ion formation can also occur, but product anions must be detected separately from positive ions. Electron attachment is probably the simplest mechanism of negative ion formation, and can be considered a charge transfer mechanism as shown in Equation (16):

$$M + e^- \longrightarrow M^{-\bullet} \qquad (16)$$

The resulting open-shell (odd-electron) ions generally have a higher tendency toward fragmentation than closed-shell (even-electron) ions.[16] Molecules with high electron affinities are likely to form ions by this mechanism. Several hydroxy-substituted PAHs exhibit $M^{-\bullet}$ ions,[17] but electronegative species such as halogenated organics are the most likely candidates for electron capture.

Although chlorinated organic compounds are expected to form $M^{-\bullet}$ ions through electron capture, trichloroethylene (TCE) does not exhibit such ions by APCI. Instead, the halide attachment ions are seen.[9] These ions take the form $Cl^-(TCE)_n$, and are formed through gas-phase attachment of Cl^-.

Proton transfer, which is a common route for positive ion formation, can also result in the formation of negative ions via proton abstraction. These reactions (Equations 17 and 18) are similar to those forming positive ions, except that the neutral analyte molecule loses a proton to the reagent molecule or ion:

$$M + R \longrightarrow [M - H]^- + RH^+ \qquad (17)$$

$$M + R^- \longrightarrow [M - H]^- + RH \qquad (18)$$

The $[M - H]^-$ ion is commonly found in negative ion spectra. A few examples include chloro- and nitrophenols,[21,22] hydroxy PAHs,[17] and several classes of pesticides.[18]

Reaction with superoxide (Equations 19 and 20) is another possible route to the formation of negative ions:

$$M + O_2^{-\bullet} \longrightarrow M^{-\bullet} + O_2 \quad (19)$$

$$M + O_2^{-\bullet} \longrightarrow [M - H]^- + HO_2^{\bullet} \quad (20)$$

These are simply charge exchange and proton transfer reactions involving superoxide. Oxygen or superoxide can also react with chlorinated aromatic analytes to yield $[M - Cl + O]^-$ ions.[5]

A review on fragmentation of even-electron organic anions is available.[23] This review focuses on low-pressure ionization methods, but it still provides information that may be useful in determining the fragmentation of $[M - H]^-$ and other even-electron negative ions formed by APCI.

1.1.3 Ion Extraction

Once ions are formed in the region of the corona discharge, these ions may interact with other molecules before being extracted into the low-pressure regions of the mass spectrometer. As mentioned above, the high pressure in the ion source region generally results in thermalization of ions, i.e. excess internal energy resulting from the ionization process is often dissipated through very-low-energy collisions, resulting in stabilization of the ions. In addition to this stabilization effect, multiple collisions favor thermodynamically stable products. Therefore, ions that are not stable undergo further reaction. In general, the distribution of ions sampled by the mass spectrometer consist of the most thermodynamically favored products appearing in the highest amounts, followed by smaller and smaller quantities of decreasingly favored products.

Beyond these thermodynamic considerations, extraction of the ions into the low-pressure region can also affect the distribution of ions. The aim of the extraction process is to transfer as many analyte ions as possible into the mass analyzer whilst maintaining the vacuum of the mass spectrometer. For the most part this involves selectively pumping away neutrals and transporting ions. Two-stage differential pumping systems are typically used to achieve this goal. Referring to Figure 2, once the ions are formed in the region of the corona discharge, the ions and neutrals are transported through the sampling cone. Only a small aperture is present in this cone, and therefore a majority of the gases are diverted to waste. The aperture is generally of submillimeter diameter. The back side of the sampling cone is maintained under a vacuum of around 1 mbar (via a rotary pump as illustrated in Figure 2). The pressure drop across the cone results in a supersonic expansion of gases from the high-pressure region into the low-pressure region.

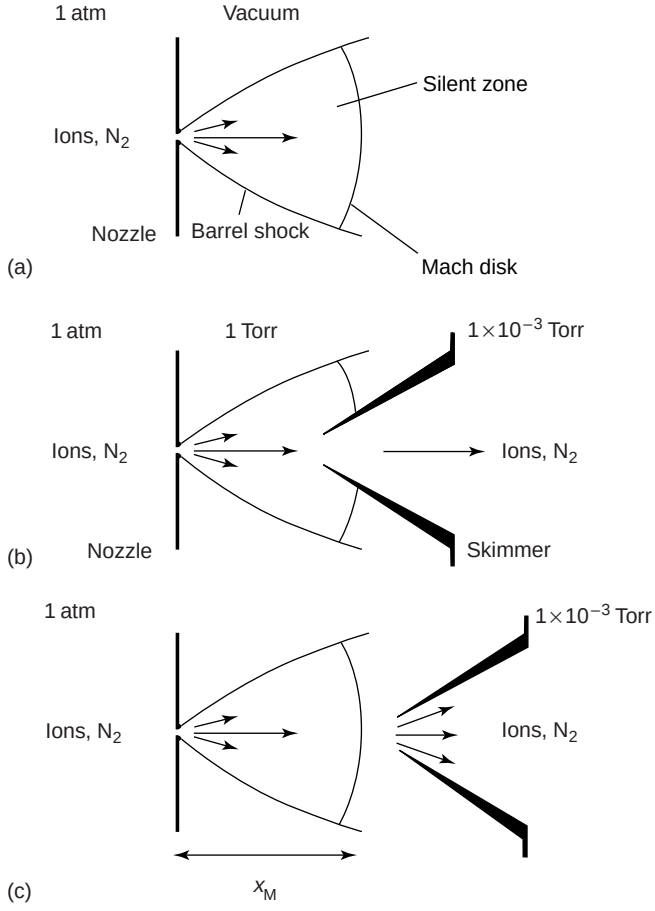

Figure 4 Expansion of gas and ions through an aperture into a vacuum: (a) basic principle; (b) skimmer placed within the supersonic expansion zone, resulting in a collimated beam; (c) skimmer placed past the Mach disk. (Reproduced with permission from Bruins.[7])

The expansion consists of a silent zone bounded by the barrel shock and the Mach disk,[7] as illustrated in Figure 4. Within the expansion zone, the molecules travel in relatively straight lines and, owing to the adiabatic expansion, ions are effectively cooled. The cooling results in stabilization of ions, most notably those with high internal energies, resulting in decreased fragmentation. In addition, the cooling can inhibit the break-up of solvent and analyte clusters, resulting in higher cluster ion signals.[5]

After the sampling cone, the pressure is further reduced by a skimmer: another cone with a small aperture. If the skimmer is placed within the free jet expansion, sampling of the ions is efficient due to the directionality of the ions within the expansion.[7] Beyond the barrel shock and the Mach disk, more random motion is observed, resulting in less-efficient collection. However, cluster ions formed in the free jet expansion (e.g. with water molecules) are often dissociated by collisions occurring in the Mach

disk.[7] In Figure 2, the region behind the skimmer is maintained at around 1×10^{-3} mbar by a turbo pump. An additional orifice then allows ions to pass into a region of lower pressure (10^{-6} mbar) maintained by a diffusion or turbomolecular pump. Finally, the ions are passed into the analyzer region (10^{-7} mbar) via the source slit. Also shown in Figure 2 are a number of focusing devices, including electrostatic lenses, a hexapole, and even the sampling cone and skimmer. These optics maintain the path of the ions into the mass analyzer and prevent them from being pumped away with the neutrals.

The formation of cluster ions, especially from solvent and water, is a common problem in APCI. Use of a counter-current bath gas can reduce the abundance of cluster ion signals.[7] Another method of minimizing cluster ion signals is to use the region between the sampling cone and the skimmer as a preanalyzer collision cell. By applying a small voltage between the sampling cone and skimmer lens (for the configuration represented in Figure 2), analyte ions are accelerated. Because the pressure is still relatively high, energetic collisions with background gas result. At low accelerating voltages, cluster ions are dissociated to give single ions without causing fragmentation of the ion. At higher voltages (100–200 V for the Figure 2 design), sufficient energy is imparted to the ions to cause collision-induced dissociation (CID). Such fragmentation events can be useful for structural determination and for more fundamental studies on fragmentation processes. Examples of in-source CID are illustrated in Figures 5 and 6 for *N*-(palmitoyl)sphingosine and 2,4-dinitrophenol.

1.2 Practical Use

In this section, important information for the use of APCI with commercial systems will be presented. The choice of solvent systems, the selection of solvent delivery systems, common positive and negative ion reagent systems, dynamic range and detection limit, APCI operation, and calibration will be discussed.

1.2.1 Solvents and Solvent Delivery Systems

APCI systems typically operate in the range 0.2–2 mL min^{-1}, although the optimum conditions may occur in a narrower range of flow rates. A high-quality HPLC pump is recommended for APCI work. The pump, however, must be capable of pulse-free operation. Otherwise, signal variations in the total ion current will be observed due to fluctuations in the intensity of background and analyte signals as the flow to the nebulizer is varied. Generally, these fluctuations are observed as regular variations in the signal, usually corresponding to the frequency of the pump piston(s). Pumps with flow compensation are preferred (maintaining constant flow under varying pressures and viscosities). Commercial pulse dampeners are available and can be added to a pump exhibiting unacceptable pulsation. As an alternative to reciprocating HPLC pumps, syringe pumps offer pulse-free operation. Syringe pumps are more commonly used for low-flow-rate applications. Typical HPLC injection valves can be used for introducing samples into the flow stream. As with any other HPLC application, the introduced sample should be miscible with the solvent system. For flow injection work, the use of special pulseless injection valves is necessary. With an ordinary valve, a pressure pulse results when the sample is introduced into the flow stream. This pulse results in a large signal spike in the mass spectrometer. In flow injection work, the spike appears at the same time as the analyte, causing significant interference. For HPLC applications, a pulse-free valve is not necessary because the analytes generally elute well after the initial pressure pulse.

The choice of solvents and flow rates is also generally set by the separation being performed. However, adjustment of the separation conditions may be necessary to optimize the APCI process. Both normal-phase (nonpolar mobile phase, polar stationary phase) and reversed-phase (polar mobile phase, nonpolar stationary phase) chromatography are compatible with APCI. Although early applications of high-performance liquid chromatography/atmospheric pressure chemical ionization/mass spectrometry (HPLC/APCI/MS) used normal-phase chromatography,[2,4] more recent applications, especially those investigating analytes in aqueous-based systems, generally rely on reversed-phase separations. Typical solvent systems use water–methanol or water–acetonitrile mixtures. The use of volatile buffers or ion pair agents is also common, and gradient systems are widely used. A few examples of isocratic and gradient mobile phases are listed in Table 3.

Although gradient elution is often used with HPLC/APCI/MS analysis, the use of limited gradient ranges is recommended to reduce possible desolvation problems. If a solvent composition change of greater than 25% is required, the following strategies can be employed to avoid problems. Postcolumn addition of organic solvent can be utilized to allow easier desolvation in the APCI interface. This approach generally limits the maximum HPLC flow rate to about 0.8 mL min^{-1}, but allows the postcolumn addition of a calibration compound (as indicated in Figure 7). If a large gradient is necessary and it is impractical to add flow postcolumn, the temperature of the nebulizer/desolvation chamber should be ramped to give higher temperatures at high aqueous content and lower temperatures at high organic content.

The background ions observed in APCI will vary as a function of the mobile phase composition. Background

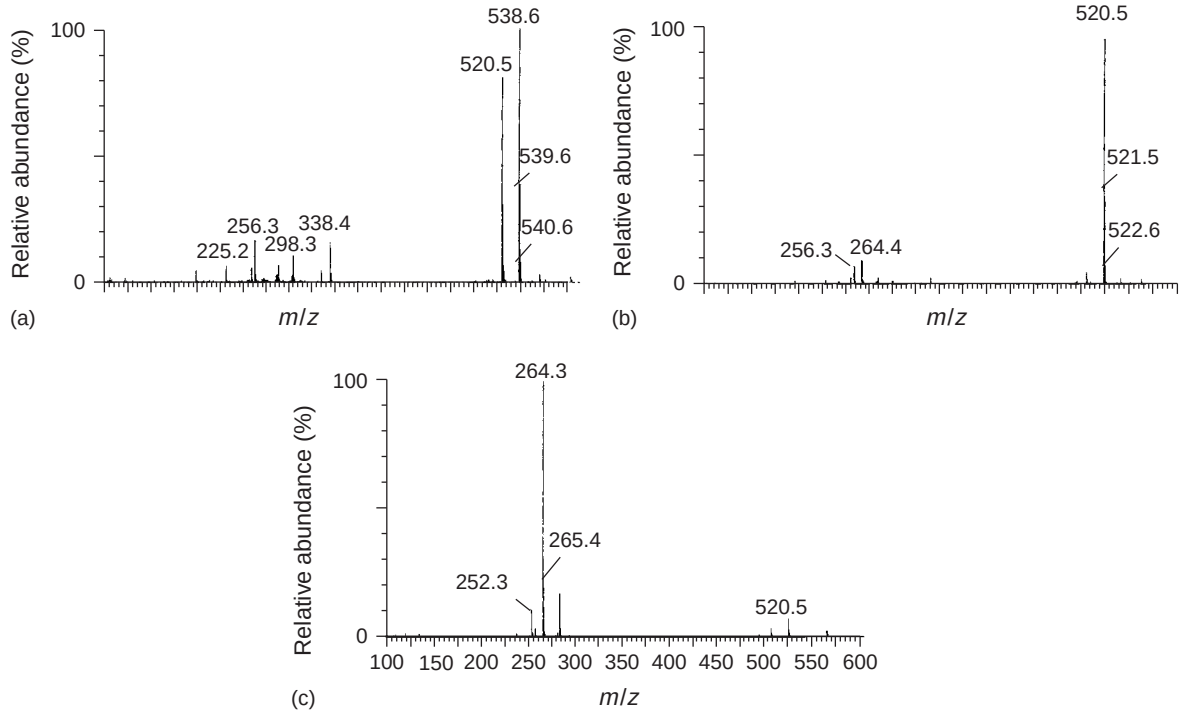

Figure 5 APCI mass spectra of *N*-(palmitoyl)sphingosine with sampling cone–skimmer voltages of: (a) 15 V; (b) 30 V; (c) 45 V. The degree of fragmentation increases as the voltage increases due to in-source CID. (Reproduced with permission from Couch et al.[24])

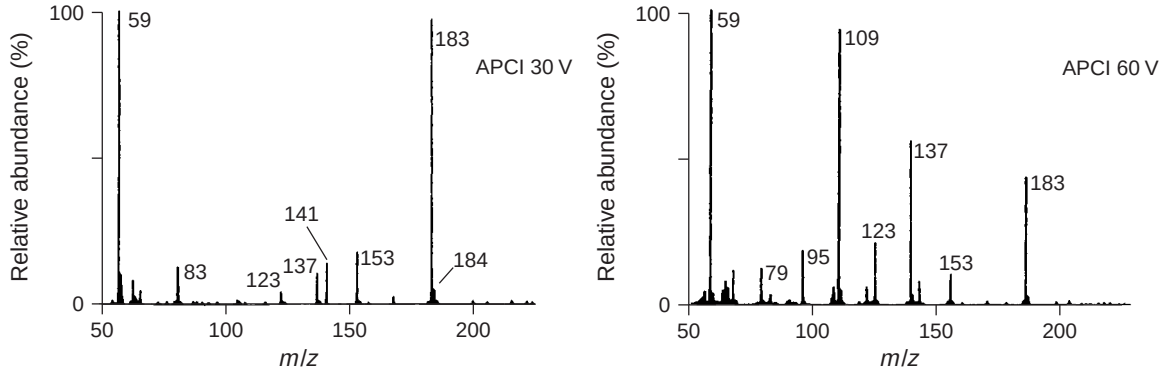

Figure 6 APCI mass spectra of 2,4-dinitrophenol with sampling cone–skimmer voltages of 30 V and 60 V. The degree of fragmentation increases as the voltage increases due to in-source CID. (Reproduced with permission from Puig et al.[22])

ions for a water–methanol–formic acid mobile phase in positive and negative ionization modes are shown in Figure 8. In addition to changes in background ions, the degree of solvent adduct formation also varies with solvent composition. Lower degrees of solvent adduct formation have been observed with water–acetonitrile mixtures than with water–methanol mixtures.

Overall, HPLC/APCI/MS is well suited for the analysis of a wide array of compounds, ranging from nonpolar to polar, and encompassing thermally labile and thermally stable compounds. However, sometimes a compromise must be made between optimum separation conditions and optimum APCI conditions. An example is the practical limitations to gradient elution as described above. In addition, only relatively volatile buffers are acceptable. Finally, ionization of an analyte depends on its gas-phase reactivity (via reaction pathways described in section 1.1.2). In general, compounds with high GPBs will be ionized efficiently in the positive ion mode. Compounds exhibiting significant GPAs will produce abundant negative ions. Careful selection of mobile phases and/or the use of low levels of modifiers can

Table 3 Examples of isocratic and gradient mobile phases used for HPLC/APCI/MS

Application	Solvent A	Solvent B	Mobile phase A : B and flow rate	Ref.
Reserpine	Water	Methanol	30 : 70, 1 mL min^{-1}	58
Hormones	Water	Methanol	50 : 50 to 10 : 90 in 15 min, 1 mL min^{-1}	
Chloro- and nitro-phenolics in water	0.05% Acetic acid (aq)	Acetonitrile	50 : 50, 1 mL min^{-1}	21
Hydroxypolycyclic aromatic hydrocarbons	10 mM Formic acid/ammonium formate, pH 3 (aq)	Methanol	50 : 50, 1 mL min^{-1}	17
Retinol and retinyl palmitate in human serum	Methanol–water–acetic acid (50 : 50 : 0.5)	Methanol–methyl *t*-butyl ether–acetic acid (50 : 50 : 0.5)	70 : 30 to 10 : 90 in 30 min	25
Priority pesticides in environmental waters	0.1 M Ammonium acetate (aq) adjusted to pH 4.5 with acetic acid	Methanol	70 : 30 to 40 : 60 in 20 min, then to 30 : 70 in 20 min, then to 10 : 90 in 5 min, 0.8 mL min^{-1}	18
PAHs	Water	Acetonitrile	40 : 60 to 0 : 100 in 30 min, hold for 5 min, then to 100% dichloromethane in 40 min, hold for 25 min, 0.2 mL min^{-1}	59
Triazine and phenylurea herbicides in water	Water–methanol (95 : 5)	Water–methanol (5 : 95)	80 : 20 to 50 : 50 in 5 min or 95 : 5 to 50 : 50 in 5 min, hold for 2 min, 0.5 mL min^{-1}	60
Priority phenolics in water	Water (1% acetic acid)	Methanol (1% acetic acid)	75 : 25 to 0 : 100 in 37 min, 1 mL min^{-1}	22

improve ionization efficiencies or can allow for selective ionization of compounds of interest. Selection of appropriate conditions can be based on GPA or GPB, ionization potential, and other thermodynamic terms. However, experience with specific systems is probably just as, if not more, useful. Examples of specific ionization systems will be discussed in section 1.3.

One example of a good match between the separation and APCI requirements is the use of acetic acid in the mobile phase for the determination of retinol and retinyl palmitate in human serum. In this application, the acetic acid served as an ion pair reagent in the separation, and as a proton donor for ionization of the analytes.[25]

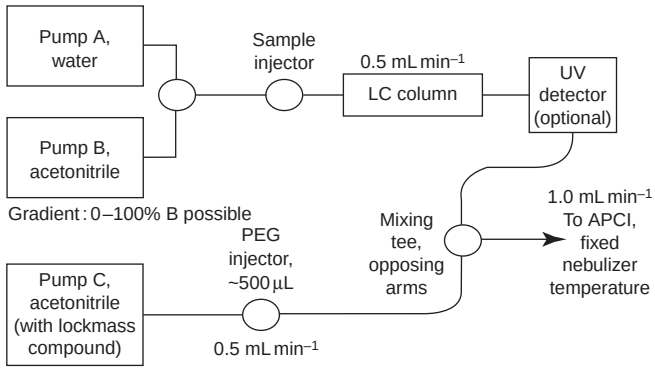

Figure 7 Diagram of HPLC set-up for gradient separation with large water content. The less polar solvent (with a mass calibration standard) is added postcolumn. (Reproduced with permission from Micromass Autospec Operator's Manual.)

1.2.2 Atmospheric Pressure Chemical Ionization Operation

The first task in any MS experiment is to obtain an ion beam. In order to do so, both ion formation and ion transport into and through the mass spectrometer must be achieved. Typically, APCI interfaces are combined with magnetic sector or quadrupole mass spectrometers. Although the details of adjusting the mass spectrometer parameters will vary by instrument type, the adjustment of the APCI source itself is similar for both types of spectrometers.

For tuning of the mass spectrometer, it is often possible to observe water cluster ions. In fact, such clusters can often be seen without any solvent flow due to trace amounts of water vapor in the gas stream or residual water vapor in the interface. With the corona needle at

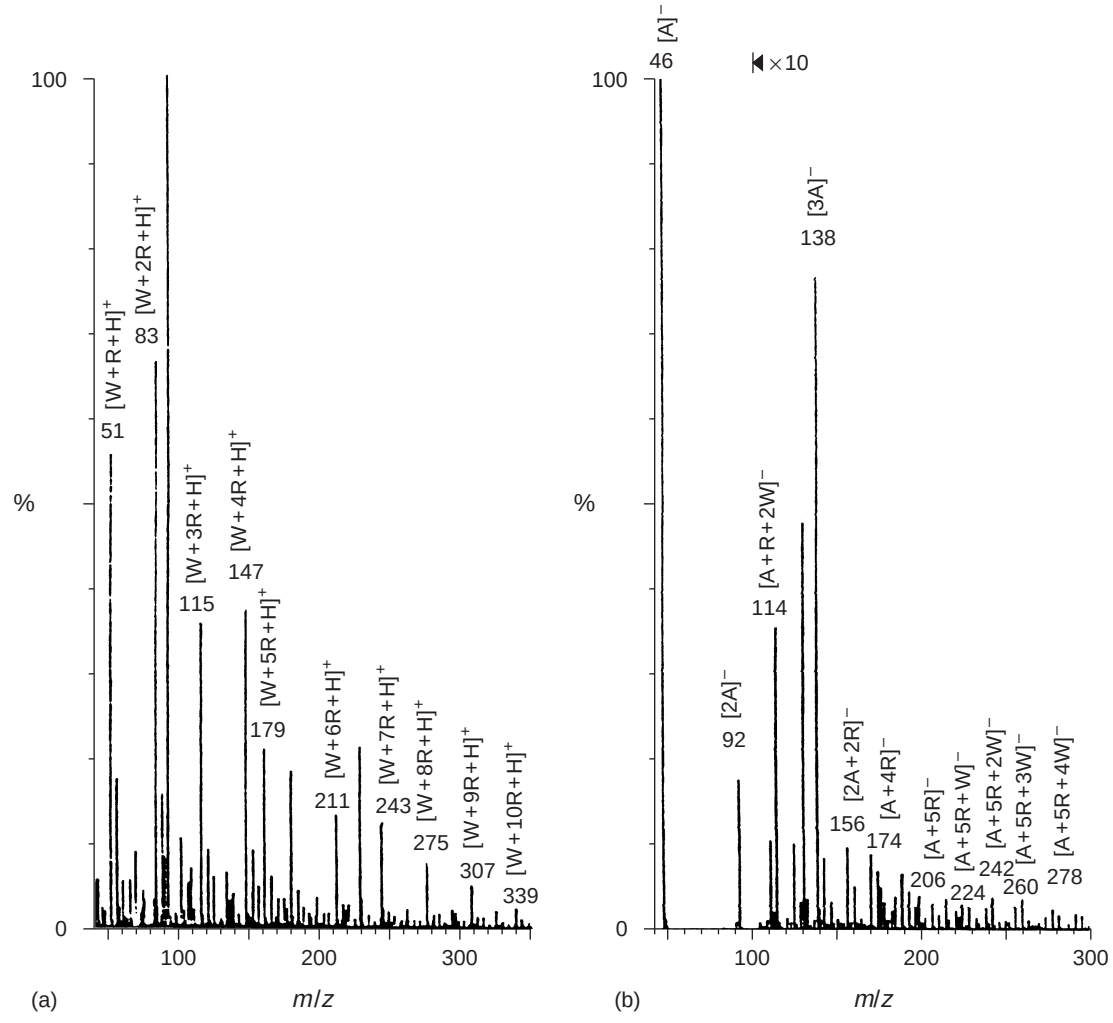

Figure 8 APCI background mass spectra due to mobile phase (W = water, R = methanol, A = formic acid) in (a) positive and (b) negative ion modes. (Reproduced with permission from Galceran and Moyano.[17])

high voltage, water cluster ions of the form $[(H_2O)_nH]^+$ may be found at various m/z ratios, including 55, 73, and 91. Larger clusters with m/z as high as ~1200 may also be observed. Although solvent cluster ions may be used for mass calibration, their intensities may be variable, which may make it difficult to use them for routine mass calibration. The formation of water clusters is often dependent on the nebulizer temperature, with more clusters formed at low temperature. However, thermally labile analytes may not survive high nebulizer temperatures.

Tuning and mass calibration can often be performed using the same species. In addition to solvent cluster ions, several added reagents may also be used for these purposes. Optimization of the APCI interface requires a constant flow of a suitable compound to the interface over a period of several minutes. Flow injection methods are not suitable for this purpose, and direct introduction of a tuning solution is usually necessary. For positive ions, a useful tuning solution is 3-picoline (0.005% v/v) in 1:1 water–acetonitrile. This solution produces $[M + H]^+$ and $[M + \text{acetonitrile}]^+$ adduct ions of 3-picoline at m/z 94 and 135. The high volatility of 3-picoline allows for continuous infusion of this solution without causing source contamination. Alternatively, pyridine may be used at the same concentration, yielding an $[M + H]^+$ ion at m/z 80. Pyridine, however, is somewhat more hazardous than 3-picoline. Caffeine at $1\,\text{ng}\,\mu\text{L}^{-1}$ in 1:1 water–acetonitrile yields $[M + H]^+$ and $[M + \text{acetonitrile}]^+$ ions at m/z 195 and 236. Source contamination is generally not a problem with this solution, although slightly longer washout times may be necessary in order to perform trace analyses near these masses. Finally, polyethylene glycol (PEG; 0.001% v/v PEG-200) in 1:1 water–acetonitrile may also be used for tuning on positive ions. PEG generally requires long washout times for complete removal from the interface.

Table 4 Calibration compounds for APCI

Compound	Mass range	Usual base peak	Comment
Sodium dodecyl sulfate	90 to ~4000	265	Negative ions
Ivory (commercial AES mix)	90 to ~800	265	Negative ions
PEG 600 diacid	~200 to ~1200	~700	Negative ions
PEG 200–1000	Varies	Varies	
Polypropylene glycol (425, 1000, 2000, 3000, 4000)	~200 to 4000	Varies	
CsI	133 to ~7000		

From Micromass Autospec Operator's Manual.

For negative ion tuning, solutions of organic acids are typically suitable. For example, 1% acetic acid in 1:1 water–acetonitrile or 1:1 water–methanol yields a strong signal with little risk of source contamination. Anthraquinone at $1\,\mathrm{ng}\,\mu\mathrm{L}^{-1}$ provides a strong signal at m/z 207, and *p*-nitrophenol yields a strong signal at m/z 138.

Table 4 lists some additional calibration compounds for APCI work. For low mass resolution, mass calibration can be performed prior to analyte analysis, with complete removal of the calibrant prior to analyte introduction. In contrast, high-resolution work requires continuous introduction of the calibrant simultaneously with the analyte. For chromatographic work, the calibrant is usually added postcolumn, as illustrated in Figure 7.

1.2.3 Detection Limit and Dynamic Range Considerations

Detection limits for APCI are highly dependent on the analyte and the conditions. Of course, optimization of the source parameters for the analyte is necessary to achieve optimum results. As with any other mass spectrometric technique, the best detection limits are achieved when performing selected ion monitoring (SIM). In this technique, only the mass of the most abundant ion is observed. The result is integration of the analytical signal for the full measurement period, whereas scanning the full mass range results in integration of a single ion for only a fraction of the total acquisition time. The disadvantage of SIM is that only information about one ion (or sometimes several) is recorded. As a result, fragmentation patterns cannot be recorded and compounds of other, perhaps unanticipated, m/z will not be observed. SIM measurements often are carried out for several ions instead of a single ion. In this manner, expected fragment ions can be observed to verify compound identity, and multiple analytes can be measured. In addition, SIM measurements can be programmed to correspond to elution times of the analytes. For example, in a chromatographic analysis of A and B, when compound A elutes, a series of ions corresponding to A are monitored. Similarly, when compound B elutes at a later time, a different series of ions corresponding to B are monitored.

Absolute detection limits for APCI have been reported to be as low as 2 pg (~7 fmol).[26] Concentration detection limits of 10 ppt (parts per trillion) have been reported.[27] More typical detection limits are in the low parts per billion range, or for absolute detection limits in the 100-pg to low-nanogram range. Detection limits for APCI are primarily dependent on the ionization efficiency of the analyte (which can be close to 100%) and the transmission efficiency of the ions to the detector. In addition to optimization of the source parameters, selecting the best reagent gas (usually governed by the mobile phase) and the best mode of operation (positive or negative ion mode) is likely to improve detection limits.

Linear ranges reported for APCI calibration curves are generally quite narrow. Reported linear calibration ranges are of the order of 1–4 orders of magnitude. Nonlinearity at higher analyte concentrations may be caused by depletion of the reagent ions by the presence of large amounts of analyte molecules. In addition, the presence of other compounds in the sample matrix (e.g. additional constituents, additives, contaminants) can alter the ionization efficiency by competing for reagent ions. Co-elution of compounds in chromatographic separations can therefore pose serious problems, either by degradation of detection limits or by causing nonlinear calibrations or matrix effects. The presence of certain mobile phase constituents, such as buffers, can also degrade detection limits if the additive is a significant consumer of reagent ions.

1.3 Applications

The number of research publications on APCI has increased steadily since the early 1990s. In 1997, over 100 journal articles were published on the topic, accounting for about 20–30% of all chemical ionization articles. The increase in publications employing APCI also corresponds to broader application of the technique to many different sample types. A comprehensive review of all APCI publications will not be presented here; however, a survey of the recent literature will be presented to illustrate the breadth of applications and to provide specific examples of conditions used in actual APCI analyses. APCI has been applied to the analysis of food products, pharmaceuticals, environmental samples, biological and medical samples, plant materials, fullerenes, catalysts, industrial products, as well as other sample types.

Food and pharmaceutical analyses include the determination of flavors and nutrients, and the determination of compounds having pharmaceutical value. Melatonin was analyzed by liquid chromatography/atmospheric pressure chemical ionization/tandem mass spectrometry (LC/APCI/MS/MS) in various commercially available formulations.[28] An ion trap mass spectrometer was used, and linear response ranges were 0.5–5 ng on-column. The mobile phase was 20 : 80 acetonitrile–water (1% acetic acid, 10 mM ammonium acetate). Compound identity was verified by monitoring the $[M + H]^+$ ion and a single daughter ion after CID of the quasimolecular ion. Volatile nitrosamines in dry sausage were determined by LC/APCI/MS/MS, and evidence was found for the formation of nitrosamines during ripening of dry sausages by reaction of residual nitrite with biogenic amines.[29] A methanol–water gradient was used for separation, and $[M + H]^+$ ions were observed. Isoflavones, especially from soy, can reportedly lower the risk of several chronic diseases. In order to investigate such benefits, a method for isoflavonoid analysis was developed utilizing LC/APCI/MS/MS.[30] Specificity in isoflavonoid identification was achieved by CID using a triple quadrupole mass spectrometer. The technique was demonstrated to have broad applications for phytoestrogen analysis. Analysis of food-derived fats has also been performed with APCI techniques, including triacylglycerols from milk fat[31] and from vegetable and berry oils.[32–35] These analyses utilized either supercritical fluid chromatography (SFC) or HPLC with interfacing to an APCI mass spectrometer. For SFC, methanol vapor was added to the ionization chamber as a reagent species.[31] Under these conditions, $[M + H]^+$ and $[M - RCOO]^+$ ions of triacylglycerols were observed. These ions were used to define the molecular weight and fatty acid constituents, respectively. Supercritical fluid chromatography/atmospheric pressure chemical ionization/tandem mass spectrometry (SFC/APCI/MS/MS) analysis of triacylglycerols was achieved using an unmodified commercial interface.[34] A triple quadrupole mass spectrometer was used, and ionization was achieved by introduction of vapors of methanol, isopropanol, water, or ammonium hydroxide in methanol. The triacylglycerols formed $[M + H]^+$ and $[M - RCOO]^+$ ions, except with added ammonium hydroxide, where $[M + NH_4]^+$ ions were observed instead. The abundance of the $[M + H]^+$ ions increased with increasing degree of unsaturation of the triacylglycerol, and the abundance of the $[M - RCOO]^+$ ion depended on the regiospecific distribution of the fatty acid moiety and the number of double bonds. Mottram et al.[35] utilized nonaqueous reversed-phase HPLC/APCI/MS with propionitrile as the mobile phase. This method also yielded $[M + H]^+$ and $[M - RCOO]^+$ ions, allowing for regiospecific analysis of the triacylglycerols. SFC/APCI/MS has been utilized for analysis of *Cannabis* by forensic laboratories.[36] This study used CO_2 with 2–7% methanol as mobile phase. No additional reagent was added for ionization. Protonated molecules were observed for each of the four important analytes. Increased cone voltages were used to perform in-source CID to allow for the additional structural information needed to verify compound identity. Carotenoids are important biological pigments that exhibit provitamin A activity, function as antioxidants, are possible inhibitors of carcinogenesis, may enhance immune response, and may prevent cardiovascular disease. HPLC/APCI/MS has been applied successfully to the analysis of this class of compounds.[37] Carotenoids exhibited $M^{+\bullet}$ and/or $[M + H]^+$ and $[MH - 18]^+$ ions in positive mode and $M^{-\bullet}$ and/or $[M - H]^-$ ions in negative mode. The type of ion observed was solvent dependent, with polar solvents increasing the abundance of protonated carotenoids and nonpolar solvents favoring the formation of radical molecular ions.

Environmental applications of APCI often involve analysis of polar compounds that cannot be analyzed by gas chromatography. Several pesticides and degradation products have been analyzed by HPLC/APCI/MS in both positive and negative ion modes.[18,38] Separations used mobile phase gradients of methanol–0.1 M ammonium acetate (pH 4.5) or acetonitrile–water. Of the 27 compounds studied, 24 yielded positive ions and 13 yielded negative ions. Most compounds gave $[M + H]^+$ or $[M - H]^-$ ions, although a few compounds exhibited $[M + Na]^+$ or $[M - CH_3]^-$ ions. With a cone voltage of 20 V, slight to moderate fragmentation was observed. Increasing the cone voltage to 40 V resulted in significant fragmentation, often resulting in complete loss of molecular weight information. The linear dynamic range was two orders of magnitude or less. Substituted phenols have also been determined in water by APCI techniques.[21] Phenols were separated using a mobile phase of acetonitrile–0.05% acetic acid (aq) (50 : 50). Postcolumn addition of 100 mM dimethylamine in acetonitrile–water (50 : 50) was used to improve ionization efficiency in negative ion mode. With the extraction voltage set at −20 V, $[M - H]^-$ was the base peak for all compounds studied. Increasing the extraction voltage to −50 V resulted in significant fragmentation, although the pseudomolecular ion was still observed. For these compounds, detection limits for APCI were better than those observed by ESI. Other reports have also indicated $[M - H]^-$ ions for chloro-, nitro-, and other substituted phenols.[22,39] PAHs and metabolites have also been studied by APCI.[26,40] Polar metabolites of benzo[*a*]pyrene were determined in the negative ion mode. Both $M^{-\bullet}$ and $[M - H]^-$ ions were observed, and the linear dynamic range

was 0.1–10 ng. The metabolism of nitrofluoranthenes by fungi was studied using LC/APCI/MS.[40] Acetonitrile–aqueous ammonium acetate (50 : 50) was used as the mobile phase. Low cone voltages were used to achieve molecular weight information and high cone voltages were used to produce in-source CID for improved structural analysis. Additional reports on azo dyes, pesticides, and naturally occurring pesticides are available.[41,45]

Biological and medical studies make up a large percentage of published APCI papers. Negative APCI was used to analyze bacteriohopanepolyols, which are abundant membrane constituents in certain bacteria.[46] Dichloromethane was added postcolumn and was used to promote the formation of chloride adduct ions. Primarily single chloride addition ions ($[M + Cl]^-$) were observed. Dextruxins (cyclic hexadepsipeptides) produced by fungi are insecticidal, phytotoxic, antiviral, nematocidal, and immunomodulating compounds. LC/APCI/MS with in-source CID was used to screen for dextruxins.[47] Acetonitrile–water (56 : 44) was used as the mobile phase and mass spectra were collected in the positive ion mode by employing a double-focusing mass spectrometer. Retinol and retinyl palmitate were analyzed in human serum by HPLC/APCI/MS.[25] APCI, in contrast to ESI, yielded a linear response over four orders of magnitude. Both compounds gave the same base peak, resulting from loss of water or loss of palmitic acid, respectively. Sampatrilat, a developmental hypertension and congestive heart failure medication, was analyzed by high-performance liquid chromatography/atmospheric pressure chemical ionization/tandem mass spectrometry (HPLC/APCI/MS/MS).[48] Despite the general applicability of APCI to analyses of polar compounds, this dicarboxylic acid could not be ionized without prior derivatization of the carboxylic acid groups with BF_3–methanol. The resulting carboxylic acid ester was ionized in positive ion mode to yield $[M + H]^+$. Sphingolipids, glycerophospholipids, and plasmalogens were analyzed from bovine brain, human serum, and porcine lens using HPLC/APCI/MS.[49] A quaternary gradient was used with isopropanol, hexane, methanol, and water, with a constant ammonium hydroxide concentration of 0.66%. Depending on the species, several different ion types were observed, including $[M + H]^+$, $[M + NH_4]^+$, and $[M + Na]^+$. The degree of fragmentation also varied from compound to compound. Determination of steroids has been conducted by APCI/MS.[50] Gradients of methanol–1% acetic acid (aq) or acetonitrile–1% acetic acid (aq) were used, and mass spectra were dominated by $[M + H]^+$ and $[MH + nH_2O]^+$ ions. In addition, at lower capillary and vaporizer temperatures, adduct ions such as $[MH + CH_3OH]^+$, $[MH + 2CH_3OH]^+$, and $[MH + CH_3CN]^+$ were also observed. Presumably due to the low PAs of the steroids, decreased sensitivity was observed in the presence of acetonitrile, propanol, or ammonium acetate in the mobile phase. Monoglucuronides of vitamin D_2 in rat bile were analyzed by APCI using both positive and negative ion modes.[51] The underivatized compound exhibited $[M - H]^-$ ions with little fragmentation, but no ionization could be achieved in positive ion mode. The methyl ester derivative yielded predominantly $[M + NH_4]^+$ ions with some $[M + H]^+$ also observed. Separations utilized acetonitrile–water or methanol–water mixtures with 65 mM ammonium acetate. Numerous other studies on biological and biomedical applications of APCI are available.[24,36,50,52–57] For a summary of the APCI applications mentioned here, see Table 5.

2 ELECTROSPRAY IONIZATION MASS SPECTROMETRY

ESI is a technique that ultimately transfers ions from a solution that is emerging from a metal capillary held at high voltage, into the gas phase. Often the ions observed in the mass spectrum were predisposed to exist in ionic form in solution, either as salts or via association of neutral molecules with small cations or anions present in solution. Owing to the fact that most chemical and biochemical processes deal with compounds existing in solution, ESI has led to a wide variety of applications in chemistry, biochemistry, pharmaceuticals, and environmental science. Other mass spectrometric ionization techniques are less successful for generating intact gaseous ions from large, highly polar, nonvolatile species.

Electrospray was first introduced by Dole et al. in 1968[61] as a method to charge synthetic polymer molecules. In 1984, Yamashita and Fenn[62,63] successfully coupled an electrospray ion source to a single quadrupole mass spectrometer. Independently and at the same time in the Soviet Union, Aleksandrov et al.[64] were developing an ESI time-of-flight mass spectrometer. The early ESI/MS achievements of Fenn, especially the demonstration that large biological molecules could be analyzed on a mass analyzer of limited m/z range because of the ability to multiply charge analyte molecules, led to the rapid popularization of this technique. This ability to obtain extensive multiple charging of analyte molecules distinguishes ESI from other mass spectrometric ionization techniques. Since its advent, much attention has been directed toward examining the fundamental mechanisms of ion formation in ESI, as well as the design and improvement of instrumentation, the optimization of experimental parameters for specific analytical applications, and its interfacing to HPLC[65,66] and capillary electrophoresis.[67,68]

Table 5 Summary of selected APCI applications

Compound	Molecular or pseudomolecular ion(s)	Refs.
Food and pharmaceutical applications		
Melatonin	$[M + H]^+$	28
Volatile nitrosamines	$[M + H]^+$	29
Isoflavones	$[M + H]^+$, $[M - H]^-$	30
Triacylglycerols	$[M + H]^+$, $[M - RCOO]^+$, $[M + 18]^+$	31–35
Cannabinoids	$[M + H]^+$	36
Carotenoids	$M^{+\bullet}$, $[M + H]^+$, $[MH - 18]^+$, $M^{-\bullet}$, $[M - H]^-$	37
Environmental applications		
Organophosphorus pesticides	$M^{+\bullet}$, $[M + H]^+$, $[M + Na]^+$, $[M - CH_3]^+$, $[M - H]^-$, $[M - CH_3]^-$, $[(CH_3O_2)PS_2]^-$[a]	18, 38
Triazine pesticides	$[M + H]^+$	18
Diazine pesticide	$[M + H]^+$, $[M - H]^-$	18
Phenylurea pesticide	$[M + H]^+$, $[M + Na]^+$, $[M - H]^-$	18
Chlorophenoxy acid pesticides	$[M + 14]^+$, $[M - H]^-$	18
Phenolic pesticides	$[M - H]^-$	18
Thiocarbamate pesticide	$[M + H]^+$	18
Substituted phenols	$[M - H]^-$	21, 22, 39
PAHs	$M^{+\bullet}$, $M^{-\bullet}$, $[M - H]^-$	26, 40
Biological and biomedical applications		
Bacteriohopanepolyols	$[M + Cl]^-$	46
Dextruxins (cyclic hexadepsipeptides)	$[M + H]^+$	47
Retinol, retinyl palmitate	$[M + H]^+$	25
Sampatrilat, carboxylic acid ester	$[M + H]^+$	48
Sphingolipids, glycerophospholipids, plasmalogens	$[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$	49
Steroids	$[M + H]^+$, $[MH + nH_2O]^+$, $[MH + nCH_3OH]^+$, $[MH + CH_3CN]^+$	50
Monoglucuronides of vitamin D_2	$[M - H]^-$, $[M + NH_4]^+$,[b] $[M + H]^+$[b]	51

[a] This fragment was the only ion observed for malathion in negative ion mode.
[b] Observed only for methyl ester derivative.

A variety of molecules, ranging in polarity from ionic, to polar, to even nonpolar compounds, can be detected via ESI/MS. Included in this list[69] are: species that are inherently charged in solution, such as organic or inorganic salts; polar neutral species exhibiting association/dissociation of small ions in solutions (e.g. H^+, Na^+, NH_4^+, etc.) such as multiply protonated/deprotonated proteins; and nonpolar species that undergo electrochemical oxidation/reduction at the ESI capillary (inherent to the positive/negative ion ESI process), such as PAHs.[70–72] A restriction is that all analytes must dissolve in a solvent exhibiting moderate conductivity;[73,74] methanol and water mixtures are most commonly used for ESI/MS.

Compared to other desorption/ionization methods, such as liquid secondary ion MS,[75] electrospray is a ''softer'' ionization method, hence a lower degree of fragmentation is inherent to the ionization process. In ESI/MS, the internal energies of desorbed, charged analyte molecules are dissipated during the desolvation stage when loosely bound solvent molecules dissociate from the final charged cluster. Because of the gentle nature of the transfer from solution into the gas phase, the three-dimensional structures of ions in solution are relatively unperturbed upon transfer into the gas phase; it appears to be the least perturbed of all mass spectrometric ionization techniques. This feature allows the examination of aspects of three-dimensional structure that are inaccessible by other MS approaches, including the propensity for hydrogen/deuterium exchange[76] (indicative of molecular conformation and intramolecular binding), as well as characteristics of weakly bound noncovalent interactions (e.g. enzyme–substrate, protein–protein, protein–cofactor, protein–metal ion interactions.[77,78]).

Despite the rapidly growing number and types of applications of ESI/MS in various areas, questions still exist regarding the precise mechanism of ion formation. A more complete understanding of the underlying mechanism may indicate new directions for future uses of ESI/MS.

2.1 Fundamental Mechanism of Electrospray Ionization

The production of gas-phase ions from solution via electrospray can be separated into four key steps:

- electrochemical processes occurring at the ESI capillary;
- charged droplet formation at the tip of the ESI capillary;
- droplet evaporation and Coulomb fission;
- formation of gas-phase ions from small, highly charged droplets.

The following sections deal with these steps in the order in which they occur.

2.1.1 Electrochemical Processes Occurring at the Electrospray Ionization Capillary

Electrospray is an ionization technique that employs a high voltage to produce an excess of charged species of one polarity in solution. These charged species serve to initiate liquid break-up and they provide a means to impart charge onto solution-phase analytes that ultimately appear as gas-phase ions in the mass spectrum. A high voltage (typically about 3 kV in positive ion mode, or about −3 kV in negative ion mode) is applied to the metal capillary through which the sample solution is driven via pressure. The exit of the capillary is usually located about 1–3 cm from the sampling aperture (counter-electrode) in the ESI source. The value of the electric field, E_c, in the air at the capillary tip can be estimated from Equation (21)[86,95] if the counter-electrode is large and planar compared to the working electrode:

$$E_c = \frac{2V_c}{[r_c \ln(4d/r_c)]} \quad (21)$$

where r_c is the capillary outer radius, V_c is the applied electric potential, and d is the distance from the capillary tip to the counter-electrode. The electric field, which is proportional to V_c, typically has a value of $\sim 10^6 - 10^7$ V m^{-1}. In response to the imposed electric field, ionic species emerging from the ESI capillary undergo electrophoretic movement. In the positive ion mode, anions migrate in the direction of the ESI capillary held at high positive potential, whereas cations migrate in the opposite direction, toward the meniscus of the emerging liquid. The force of the electric field acting upon positive ions is counteracted by the surface tension of the liquid, resulting in the formation of a dynamic cone of liquid referred to as a ''Taylor cone''.[80] Because of the high voltage difference between the ESI capillary and the counter-electrode (ESI capillary positive for positive ion ESI/MS), oxidation reactions occur at the metal–solution interface of the ESI capillary, such as H^+ production from water, as shown in Equation (22):[81–83]

$$2H_2O = O_2 + 4H^+ + 4e^- \; E^0 \text{ (vs SHE)} = 1.229\,\text{V} \quad (22)$$

Conversely, in the negative ion mode, electrophoretic movement of anions and cations is reversed relative to the above discussion, and reduction reactions occur at the metal ESI capillary. In order to sustain a continuous production of charged droplets of one polarity, it is imperative to have an electrochemical mechanism whereby ions are continually produced at the ESI capillary. Similarly, to maintain a fixed voltage at the counter-electrode, where there is a continuous arrival of charged species of one polarity, it is necessary to perform a second electrochemical reaction. In the positive ion mode, requiring oxidation at the ESI capillary, the second reaction must be a reduction at the counter-electrode. Logically, the counter-electrode reaction in the negative ion mode is an oxidation. When this second reaction occurs, it may be thought of as completing the electrical circuit of the ESI source, which has been described thereby as a special type of electrolytic cell (Figure 9) in which electrolysis maintains the charge balance to allow continuous production of charged droplets.[81,82,84,85]

The electrospray source has been further likened to a controlled-current electrolytic cell.[82] This implies that in positive ion experiments, in order to supply a certain

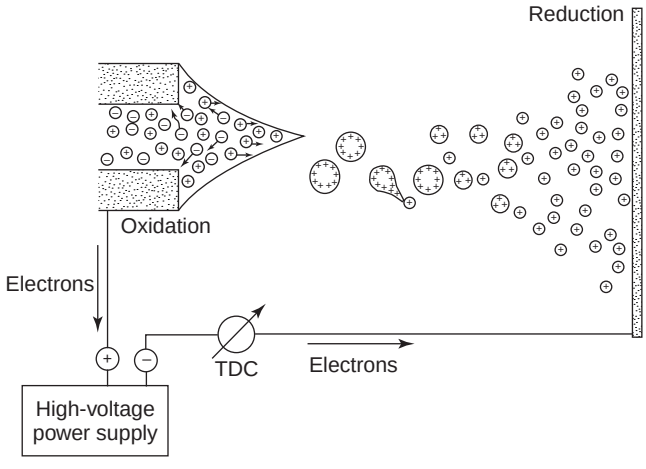

Figure 9 Schematic diagram of charged droplet generation in ESI in the positive ion mode. Oxidation processes occur at the electrospray capillary to create an excess of positive charges. The solution containing an excess of positive charge is attracted to the counter-electrode, in opposition to the surface tension of the liquid, resulting in the formation of a stable ''Taylor cone''. Liquid break-up occurs at the tip of the cone, forming positively charged droplets; solvent evaporation then leads to uneven fission and generation of smaller ''offspring'' droplets from which gas-phase ions are ultimately formed. (Reproduced with permission from Kebarle and Tang.[81])

ESI current, the oxidation process that occurs at the ESI capillary is that characterized by the lowest oxidation potential, as determined by the solution components, and the materials present in the capillary. If the arrival rate of these most electroactive species to the metal–solution interface is insufficient to produce the required current, the next most easily oxidized species will undergo electron removal, and so on, until an adequate current is achieved.

The observation of Zn^{2+} ions in the positive ion ESI mass spectrum of a solution sprayed through a zinc metal ESI capillary[84] was given as evidence of the inherent electrochemical nature of ESI. More recently, Van Berkel et al.[83] reported pH changes of as much as 4 pH units during the ESI process. The pH lowering in the positive ion mode was attributed to the oxidation of water to produce H^+ in the employed aqueous solution when an inert platinum ESI capillary was used in place of a stainless steel capillary. This latter evidence provides proof that solution-phase oxidation/reduction reactions are inherent to the ESI process. Corrosion (oxidation) of certain components of stainless steel capillaries was proposed to serve as a type of redox ''buffer'', thus attenuating the pH change. Moreover, pH changes could be avoided at the platinum ESI capillary when an easily oxidized species (e.g. KI as redox buffer[83]) was added to suppress the oxidation of water. pH changes attributable to the electrochemical reaction of water may have a profound effect upon the appearance of the mass spectrum, especially when low flow rates[83] are employed (such as in the nanospray mode) and electrochemical reaction products are undergoing only minimal dilution.

2.1.2 *Charged Droplet Formation at the Tip of the Electrospray Ionization Capillary*

The above-described electrochemical redox processes lead to the continuous production of charged droplets of a given polarity (Figure 9), as determined by the polarity of the applied potential. In the positive ion mode, as the excess positive charges accumulate near the end of the capillary, they come under the influence of the imposed electric field. The positive ions emerging from the capillary are thus attracted to the relatively negative ''pole'' that is the counter-electrode. This attraction causes the liquid to elongate in the direction of the counter-electrode. However, the surface tension of the liquid serves to provide an opposing force to hold the solution together. The result of these two opposing forces is the formation of a dynamic cone of liquid at the tip of the stainless-steel capillary, i.e. the ''Taylor cone''.[80] At the apex of the Taylor cone a ''jet'' of liquid emerges, wherein the charge density is sufficiently high to overcome the surface tension, and droplet break-up results. This process produces a fine spray of charged droplets that head toward the counter-electrode owing to a combination of the potential gradient and a weak pressure gradient. The droplets carry excess positive charges and stability is gained upon droplet break-up because the repulsive force between like charges is spread over a larger surface area. Of course, in the negative ion mode, ESI droplets contain excess negative charges.

According to Pfeifer and Hendricks,[86] the radius of the generated charged droplets will increase with increasing flow rate or decreasing conductivity. Therefore, the use of low flow rates and high-conductivity solutions will result in the production of very small droplets. The stream of charged droplets emerging from the ESI capillary can be thought of as the ''ESI current'', and its magnitude is identical to the current generated by electrochemical reactions at the ESI capillary. The steadiness of the ESI current (e.g. as measured at the counter-electrode) is an indicator of the stability of the generated spray. The measured current also provides a quantitative estimate of the total number of ions that could be converted into gas-phase ions.

2.1.3 *Droplet Evaporation and Coulomb Fission*

After the charged droplets are produced at the tip of the electrospray capillary, heating induces evaporation of high-volatility solvent molecules. As the solvent evaporates, the size of the droplets becomes smaller but the charge on the droplets remains constant. Therefore, the electrostatic repulsion increases until it reaches the so-called ''Rayleigh stability limit'' where the force of electrostatic repulsion between like charges becomes greater than the surface tension force holding the droplet together. The droplet becomes unstable when the Rayleigh limit is reached, as defined by the Rayleigh equation (Equation 23):[87]

$$q_R = 8\pi(\varepsilon_0 \gamma R^3)^{1/2} \tag{23}$$

where q_R is the excess charge on the droplet of radius R, γ is the surface tension, and ε_0 is the permittivity of the vacuum.

As the Rayleigh stability limit is approached, the droplets undergo what is referred to as ''Coulomb fission'', which is essentially a break-up of the droplet into smaller ''offspring droplets''. Figure 10 is a schematic depiction of the Coulomb fission process that generates offspring droplets.[81] The driving force behind this process is that the repulsive force of the like charges is alleviated if the charges can be spread over a larger surface area; this is made possible only if the droplet is sheared. Droplet fission into two particles of nearly equal mass and charge (i.e. ''even fission'') is not observed. Instead, the charged droplets are not static entities, but

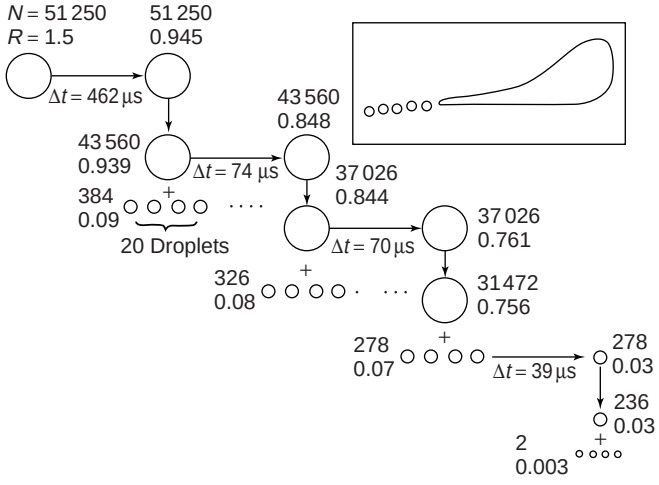

Figure 10 Schematic diagram depicting the Coulomb fission process. The droplet at the extreme left is one that has just left the Taylor cone: N = number of elementary charges on the droplet; R = radius of droplet in micrometers. The Δt values indicate the time required for droplet shrinkage to a size where fission occurs. (Reproduced with permission from Kebarle and Tang.[81])

distort alternately from oblate to prolate shapes.[88] Such distortions stimulate disruptions in which a ''stream'' of much smaller droplets (''uneven fission'') is emitted. The combined total of offspring droplets formed during an uneven fission event carry off about 2% of the mass and 15% of the charge of the parent droplet,[88] with a monodisperse distribution of offspring droplet sizes having radii roughly one-tenth that of the parent droplets. The offspring droplets will thus have higher charge-to-mass ratios than the parent droplets that produced them. Upon further shrinkage of the offspring droplets, the uneven fission process may repeat itself a second and, perhaps, a third time before gas-phase ions are generated.

The above description indicates that the initial charge and droplet size are very important in determining when Coulomb fission will occur. The initial radius of the droplets that leave the capillary depends upon the flow rate and the conductivity of the solution. The size of droplets increases approximately in proportion to (flow rate)$^{2/3}$, as given by Fernandez de la Mora et al.[89] hence, low flow rates clearly produce smaller droplets. Higher solution conductivities also lead to smaller initial droplet sizes, which may result in more efficient evaporation of solvent.

2.1.4 *Formation of Gas-phase Ions from Small, Highly Charged Droplets*

Gas-phase ions originate from small, highly charged droplets after a series of Coulomb fission events. The exact mechanism whereby ions are transferred from the final droplets into the gas phase has not been clarified completely, mainly because the processes that occur in these final droplets are very difficult to probe. Two principal mechanisms have been put forward to explain the process.

The first mechanism, proposed in initial form by Dole,[61] predicts that a series of fission events ultimately results in the production of extremely small droplets bearing one or more excess charges but only a single analyte molecule. Upon evaporation of the last few solvent molecules, the available charges will find their way to the most stable sites on the analyte molecule, leading to the production of a gas-phase ion. This description is referred to as the *charged residue model*. Dole argued that if the original concentration is low enough, a succession of Coulomb fission events would produce droplets so small that they contained only one analyte molecule. As the last of the solvent evaporates, the analyte molecule would retain the ''residual'' droplet charge, and thus form a ''free'' ion in the gas phase.

The second mechanism, proposed by Iribarne and Thomson,[90,91] purports that direct emission of solvated ions from the droplets is responsible for the production of gas-phase ions after the radii of the droplets decrease to a given size. This process is called the *ion evaporation mechanism*. In this mechanism, Iribarne and Thomson argued that the charged droplets are so small (radius < 10 nm) that solvent evaporation could make the electric field on the surface sufficiently high to emit solvated ions from the solution phase into the gas phase before the Rayleigh limit was reached. They proposed that at a certain intermediate stage, the high charge density on the charged droplet surface was sufficient to lift a charged analyte molecule, surrounded by a few solvent molecules, from the droplet surface into the ambient gas.

A fundamental difference between the charged residue mechanism and the ion evaporation mechanism is in how an analyte molecule becomes separated from other solute molecules in the droplet,[92] although in either mechanism, both solvent evaporation and Coulomb repulsion are the driving forces leading to the formation of gas-phase ions. In the ion evaporation mechanism, separation is realized by the ion desorption/emission of a single solvated analyte ion carrying some of the droplet's charge, thus alleviating Coulomb repulsion to some degree upon departure. In the charged residue mechanism, the separation comes from the repetitive fission events that produce the ultimate small charged droplets that contain only one analyte molecule (the gas-phase ion precursor). However, it is difficult to distinguish the phenomenon of desorption of a single solvated ion from a parent droplet, from that of a Coulomb fission event that produces an ultimate droplet that contains only one analyte molecule. Moreover, any

prediction of a difference in the number of charges or the degree of solvation found in a desorbed ion produced according to the ion evaporation mechanism, or those found in an ultimate droplet produced by the charged residue mechanism, are likely to be quite subtle. Nevertheless, a growing consensus has emerged that at least the ESI-generated ions of very large molecules are formed according to a charged residue scenario.[93–95]

2.2 Instrumentation

Initial charged droplet formation in ESI occurs at atmospheric pressure but all mass analyzers for MS will operate correctly only under high vacuum. Therefore, an early challenge of electrospray source design was to couple efficiently an atmospheric pressure spray to the high-vacuum mass analyzer environment in order to maximize gas-phase ion production and transmission. Thus, the ESI mass spectrometer must: optimize the efficiency of liquid break-up; maximize the generation of gas-phase ions; and maximize the transport of ions through the mass spectrometer.

2.2.1 Electrospray Ionization Source Construction and Operation

A schematic diagram of an ESI source[96] is presented in Figure 11. The high-voltage arrangement required to initiate the ESI process is usually obtained in one of two ways: either the high voltage is applied to the metal ESI capillary with the counter-electrode being the low-potential end; or the metal capillary is held at or near ground while the counter-electrode is the high-voltage end. The source and sprayer are rarely both connected to high-voltage power supplies. A consideration that may influence the preferred arrangement is whether or not on-line separation techniques such as capillary electrophoresis are to

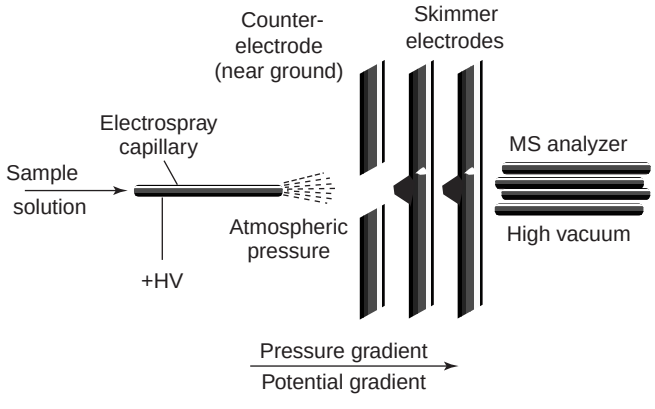

Figure 11 Schematic diagram of a generic ESI source. (Reproduced with permission from Gaskell.[96])

be performed. Conventional capillary electrophoresis is usually conducted by applying 20–30 kV between cathode and anode, and separation is based upon differing electrophoretic mobilities in the capillary in response to the imposed electric field. In capillary electrophoresis/MS experiments, if the source vial (anode) is held at 20 kV, while the cathode is set at the high voltage required for electrospray, say 4 kV, the voltage drop across the capillary will be reduced to 16 kV and a longer separation time may result.

The required operation of ESI sources at high voltages creates very high charge densities at the sharp edges of the metal capillary tip; hence electron emission can occur, leading to initiation of a gas discharge (corona discharge) between the ESI capillary and the counter-electrode. This type of discharge is difficult to avoid completely, and it is especially prevalent in the negative ion mode and when the conductivity of analyte solutions is low. Electrical discharge should be minimized because it causes instability in charged droplet production (spray instability) and can change the appearance of mass spectra. To do so, the electric field is normally operated at the minimum value required to initiate formation of the Taylor cone and concomitant gas-phase ion production. Electron-scavenging gases, such as oxygen or SF_6,[97] have been added to the ESI source chamber and chlorinated solvents, such as chloroform,[98] have been added to sample solutions to reduce the problem.

2.2.2 Position of the Electrospray Capillary

Gas-phase ions are generated from charged droplets after solvent evaporation and Coulomb fission. The droplets, formed at the tip of the electrospray capillary, can be produced with a distribution of sizes and number of excess charges. The optimum positioning of the ESI capillary relative to the counter-electrode (Figure 12a) will depend upon several factors. First of all, for a given voltage difference between the ESI capillary and counter-electrode, the electric field will drop off as the distance between the two is increased. On the other hand, the capillary must not be too close to the counter-electrode because droplets may not be sufficiently desolvated upon entering the mass spectrometer. Of course, larger droplets are likely to be more susceptible to incomplete desolvation. Off-axis positioning of the ESI capillary can offer increased sensitivity, whereas diagonal positioning appears to be more effective for pneumatically assisted electrospray (Figure 12c) because it affords higher stability and no loss in sensitivity. Both seem to help prevent contamination of the sampling orifice. The position of the sprayer, therefore, plays an important part in determining the quality of obtained mass spectra, including sensitivity. For pneumatically (or

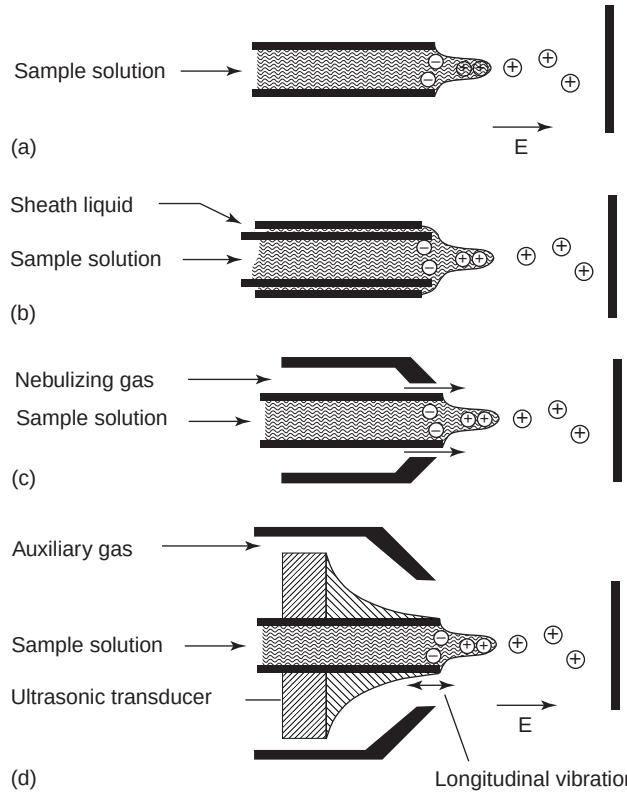

Figure 12 Schematic diagram of aerosol formation by: (a) ''pure'' electrospray; (b) electrospray with liquid sheath; (c) pneumatically assisted electrospray; (d) ultrasonically assisted electrospray. (Reproduced with permission from Bruins.[110])

ultrasonically) assisted electrospray, the ''shearing'' of the emerging liquid is more dependent on the gas flow (or the ultrasonic vibration; Figure 12d) than the electric field. The optimum capillary position is typically closer to the sampling orifice, and good signals are obtained over a narrower region.

2.2.3 Nebulization at High Flow Rate

High-flow-rate electrospray is especially designed for the purpose of on-line coupling of liquid separation techniques such as HPLC to ESI/MS.[66,99,100] It has been established[101] that low ESI flow rates result in high ion desorption rates, which implies high sensitivity in detection. The practical upper limit to flow rate for ''pure'' electrospray (i.e. no use of nebulizing gas) is $\sim$10–20 μL min^{-1}, depending on the composition of the solvent. In these microliter per minute flow regimes, the ESI technique is characterized as ''concentration sensitive'' (i.e. signal strength is considered to be proportional to concentration), whereas at extremely low flow rates of tens of nanoliters per minute, where ionization efficiency can approach 100%, ESI can be described as ''mass flux sensitive''[69] (i.e. response is proportional to the absolute quantity or ''mass'' of material present).

To meet the demands for the use of the ESI mass spectrometer as a detector for LC separations, special consideration had to be given to ESI source design, with the aim of increasing the inlet flow tolerance to accommodate the high flow rates (e.g. up to 2 mL min^{-1}) used with conventional HPLC columns. The most common approach used to handle high solution flow rates (e.g. 20 μL min^{-1}–2 mL min^{-1}) is to add a counter-current of ''bath gas''[99] or pneumatically assisted nebulization,[100] as shown in Figure 12(c), to facilitate liquid break-up. This type of device relies upon the shearing force of nitrogen being delivered at high pressure to the region where solution is emerging from the ESI capillary. Both gas and liquid are forced out of a small orifice, resulting in aerosol formation.

Ultrasonic nebulizers (Figure 12d) have also been employed to assist with liquid break-up at higher flow rates. In those devices, it is the ultrasonic vibrations that assist in shearing the liquid emerging from the ESI capillary. Pneumatically assisted nebulizers and ultrasonic nebulizers are particularly useful for sample solutions of high aqueous content, including mobile phases used in reversed-phase LC. However, both approaches must still rely on an imposed electric field for droplet charging.

At flow rates above about 200 μL min^{-1}, efficient droplet evaporation is absolutely imperative. If it is not achieved, ionization efficiency, and thus sensitivity, can be seriously impaired. Often the easiest approach to overcoming an exceedingly elevated elution rate from an HPLC column is to resort to ''postcolumn splitting'',[102] where a certain proportion of the effluent exiting the column is diverted away from the mass spectrometer. In theory, this approach will not result in sensitivity loss because at high flow rates ESI/MS is considered to be a ''concentration sensitive'' technique. Alternatively, additional thermal energy can be provided to enhance droplet evaporation.[103] Three commercial approaches have been employed for this purpose: heating the nebulizing gas; introducing a desolvation region consisting of a long, heated metal capillary; and introducing resistive heating externally to various regions of the ion source. In order to desolvate effectively, one must combat the inherent cooling property of the gas undergoing expansion as it enters the mass spectrometer. This requires a significant amount of heat input to raise the temperature by several hundred degrees.[69] Although most of the thermal energy is consumed during desolvation, some risk of thermal degradation may be introduced for certain compounds that are highly thermally labile.

2.2.4 Nanoflow Regime

Both the number of charges per droplet and the radius of the initial droplet are dependent on the solution flow rate. As rationalized by Wilm and Mann,[101] the higher surface-to-volume ratio of smaller droplets allows a higher efficiency of desorption of charged analyte molecules. Moreover, a smaller-aperture capillary tip gives a narrower dispersion of sprayed droplets and hence less sample is lost due to inefficient transfer into the mass spectrometer. For these reasons, the overall ionization efficiency for a given analyte can increase by as much as two orders of magnitude[101] when the flow rate is decreased from the conventional microliter per minute range to $\sim 20\,\text{nL}\,\text{min}^{-1}$. The latter flow rate in the so-called "nanoflow regime" (i.e. $\sim 10-100\,\text{nL}\,\text{min}^{-1}$) yields droplet sizes of approximately 200 nm.[96,104] The improved ionization efficiency has resulted in reports of extremely low limits of detection, such as $5\,\text{amol}\,\mu\text{L}^{-1}$ for methionine–enkephalin[105] and $32\,\text{amol}\,\mu\text{L}^{-1}$ for neurotensin.[106]

When analyzing a small volume of a very valuable sample, as is often the case when performing trace analyses of samples of biological origin, low flow rates also allow for reduced sample consumption and extended acquisition times. This combination can afford the maximum signal-to-noise ratio via signal averaging over a long time period. Long acquisition times can also facilitate tuning and preparation for more complicated experiments such as tandem MS studies. Extensive development of nanoflow regime sprayers occurred in the mid- to late-1990s. The challenge was to achieve stable electrospray signals at the very low flow rates. The approach used to accomplish this was to reduce the diameter of the sprayer aperture, because lower flow rates require smaller diameters for optimum stability. Theoretical treatment[104] gives an approximate value of $20\,\text{nL}\,\text{min}^{-1}$ for a finely drawn borosilicate capillary having a 1–3 μm i.d. aperture that is coated with silver[107] or gold.[108] The sputter-coated metal serves to provide electrical contact to the capillary tip. Liquid break-up at the capillary exit is attributable exclusively to the electrical stress due to the imposed high voltage (i.e. no nebulization assistance is employed). Therefore, the flow rate is determined (and regulated) via the field strength, the solvent viscosity and surface tension, the electrolyte concentration (conductivity), and the capillary tip diameter.[69] It is possible to work with 100% aqueous solvent[109] at low flow rates ($\sim 20\,\text{nL}\,\text{min}^{-1}$). For such a low-diameter capillary aperture, blockage (clogging) is a potential problem, so solutions need to be filtered before use.

2.2.5 Transport into the Vacuum Region

As the ions are transferred from the atmospheric pressure region into progressively lower-pressure regions of the mass spectrometer, the "spray plume" expands and a strong cooling effect results. Remaining solvent–ion aggregates follow a straight line along the axis of the nozzle, whereas gas molecules far away from the axis have a random motion and will be pumped away. Immediately beyond the cone orifice, a barrel shock wave exists (Figure 13) where many collisions occur between ions and gas molecules.[110] Inside this shock wave is the so-called "silent zone", i.e. the region between the cone and the Mach disk, where the ions and gas molecules move at the same speed and direction into the vacuum system, while experiencing a strong cooling effect. The location of the Mach disk is described[110–112] in relation to the cone (according to Equation 24), where X_M is the distance beyond the cone:

$$X_M = 0.67 D_0 \left(\frac{P_0}{P_1} \right)^{1/2} \tag{24}$$

where D_0 is the diameter of the cone orifice, P_0 is the pressure on the entrance side of the cone, and P_1 is the pressure on the vacuum side of the cone. The entrance of the skimmer (typically $\sim 1.4-0.6$ mm in diameter) immediately following the cone ($\sim 0.3-0.5$ mm in diameter) should be positioned so that sampling of the molecular beam occurs inside the silent zone of the free jet expansion (Figure 13b), where the transmission of ions through the skimmer is most efficient.

Accompanying expansion into the vacuum, of course, is considerable cooling and hence a tendency for condensation, especially if water molecules are present. However, the condensation can be inhibited via the use of a so-called "curtain gas" that is directed across the front of the cone (Sciex), or a counter-current flow of drying gas that is sometimes heated (Analytica of Branford).[99] The pressure in the region between cone and skimmer is around 10^{-2} Torr; thus the mean free path of movement of an ion is sufficiently long that acceleration of ions between the cone and skimmer will cause energetic collisions between ions and residual gas molecules, resulting in fragmentation. The level of fragmentation can be controlled to some degree by varying the potential difference between the cone and the skimmer (effectively changing the velocity of the ions undergoing collision).[113] This CID process, referred to as "in-source CID" or "nozzle-skimmer CID", should not be confused with CID occurring in a collision cell or trapped ion region of a tandem mass spectrometer. Mild (low-energy) "in-source CID" can be used to break up the remaining cluster ions. For each type of counter-electrode, be it a cone-shaped orifice or a tube, the build-up of contaminants on the counter-electrode over time can block the orifice, thus necessitating frequent cleaning; well-designed systems allow for easy removal and cleaning.

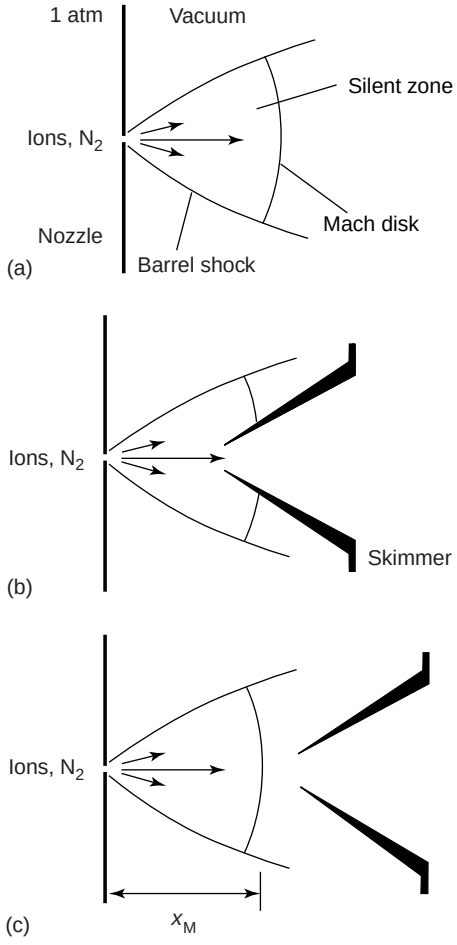

Figure 13 "Free-jet" expansion of neutral gas and ions into a vacuum region: (a) basic principle and nomenclature; (b) geometry with skimmer penetrating into the "silent zone"; (c) skimmer located past the Mach disk. (Reproduced with permission from Bruins.[110])

2.3 The Electrospray Ionization Mass Spectrum

The above discussion has centered upon a description of how ions are transferred from solution into the gas phase during the ESI process. We now turn to a consideration of factors that influence the appearance of the ESI mass spectrum. Two characteristics that we will examine are: the factors that determine relative intensities of peaks in the mass spectrum; and those that influence the distribution of charge states when multiple charging is occurring. The ability to form ions characterized by very high charge states is unique, among ionization methods, to ESI.

2.3.1 Signal Suppression

Characterization of the relationship between ion abundance (i.e. signal response) and analyte concentration is an important factor in the understanding of any technique of solution analysis. Even under the best of circumstances, the solution containing even a purified analyte that is subjected to electrospray conditions cannot be considered to be a "single electrolyte" system. Electrochemical reactions occurring at the ESI capillary alter the solution composition, and impurities can enter from a variety of sources, including via reagent-grade solvents (e.g. methanol, water, acetonitrile) that often contain $\sim 10^{-5}$ M Na^+, and via contamination of the capillary line and fittings leading all the way up to the spray tip. Moreover, electrolytes are often added to sample solutions to improve the production of specific types of ions, e.g. acid is commonly added to promote the formation of MH^+, whereas base is sometimes added to increase $[M - H]^-$ production in negative ion ESI. In considering a solution containing analyte A in the presence of added electrolyte B and impurity C, the total electrospray current, I_{total}, is given by the sum of the signals given by the three constituents $I_A + I_B + I_C$ (Equation 25).[114,115] The signal from the analyte, I_A, is given by Equation (26):

$$I_{total} = I_A + I_B + I_C \tag{25}$$

$$I_A = \frac{fpI_{total}k_A[A^+]}{k_A[A^+] + k_B[B^+] + k_C[C^+]} \tag{26}$$

where k_A, k_B, and k_C are the corresponding rate constants for desorption of analyte A, added electrolyte B and impurity C, f is the fraction of the total charges on all droplets that are transferred to gas-phase ions, and p is the ion-sampling efficiency, i.e. the fraction of formed gas-phase ions that reach the detector. The characteristics of the specific species undergoing desorption determine the value of the desorption rate constants. Species with relatively high surface activities will have higher desorption rate constants, and will thus exhibit an increased tendency to produce signal.[114,115]

From Equation (26) it can be seen that the intensity of the peak for analyte A is determined not only by its desorption rate constant and concentration but also by the concentrations and characteristics of the electrolytes present, be they intentionally added or entering as impurities. This description is consistent with experimental results.[81] In practical terms, the implication is that electrolytes present in solution effectively suppress analyte signals, and hence solutions containing high concentration electrolytes should be avoided.[116] Analyte signal suppression from competing analytes turns out to be one of the largest shortcomings of ESI/MS and it can be a prohibitive factor for the successful analyses of complex mixtures, especially when the compounds of interest are present at trace levels.

2.3.2 Linearity and Dynamic Range

Even in the presence of competing electrolytes, the signal produced by a given analyte exhibits a linear dependence

on concentration over a certain ''linear dynamic range'' that can extend from the lower limit of detection (often below 10^{-12} M) up to $\sim 10^{-5}$ M.[110,117] Over this range, in favorable cases, the analyte signal increases linearly with concentration. Above the upper limit of $\sim 10^{-5}$ M, it is postulated that insufficient charge is present to create ionic species out of all analyte molecules.[115] As elaborated above, signal response is affected by competition for available charge with other solution species. Sensitivity is also a function of solvent composition, flow rate, and characteristics of the analyte itself, such as its surface activity. Highly surface-active analytes have lower limits of detection. Of course, the ability to create ionic form(s) of the analyte in solution is also important, whether through protonation/deprotonation, adduct formation with small solution cations/anions, or ionization via electrochemical oxidation/reduction processes. Different chemical structures will have different rate constants for ion desorption. For the purposes of quantitation, knowledge of the linear dynamic range can be of key importance. The expansion of the linear dynamic range is of practical significance for the pharmaceutical industry and many other fields. As noted above, low flow rates can increase the overall ionization efficiency and thereby improve detection sensitivities, allowing the attainment of detection limits in favorable cases in the attomole (10^{-18} mol) range.[105,106]

2.3.3 Charge State Distributions

Among all the mass spectrometric ionization techniques, the ability to form multiply charged analyte ions of very high charge states is unique to the ESI process. Multiple charging in the positive ion mode usually occurs via multiple protonations of analyte ions, such as proteins. In the negative ion mode, multiple deprotonation (i.e. multiple H^+ dissociations) is usually occurring in order to produce anions of high charge states. The result is that the mass spectrum of a ''pure'' compound exhibits a series of peaks that are all representative of intact molecules that differ in the number of charges that they carry. Adjacent peaks in the characteristic ''envelope'' of ions constituting the mass spectrum of a protein or other multiply charged molecule (Figure 14) progressively contain one less charge as one moves toward higher m/z values in the spectrum.

The appearance of the envelope, including the range of charge states observed, and the most intense peak in the mass spectrum, as well as the relative intensities of other peaks, depend upon a combination of factors. Parameters that affect the so-called ''charge state distribution'' of a given analyte include characteristics of the molecule such as structure, conformation, and concentration. In addition, solution conditions such as equilibrium considerations and solvent polarity, as well as experimental parameters particular to a given mass spectral acquisition, will influence the distribution of charge states observed in the ESI mass spectrum.

2.3.3.1 Effect of Analyte Structure An obvious structural feature that has a profound affect on the charge state distribution is the number of available sites for protonation/deprotonation on the analyte molecule. Early reports on the multiple charging phenomenon correlated the number of basic sites on protein molecules to the number of charges (protons) that could attach and were retained on analyte molecules in obtained ESI mass spectra.[118–121] Another structural feature that has a strong influence on the propensity for charge attachment onto proteins is the tertiary structure (conformation) of

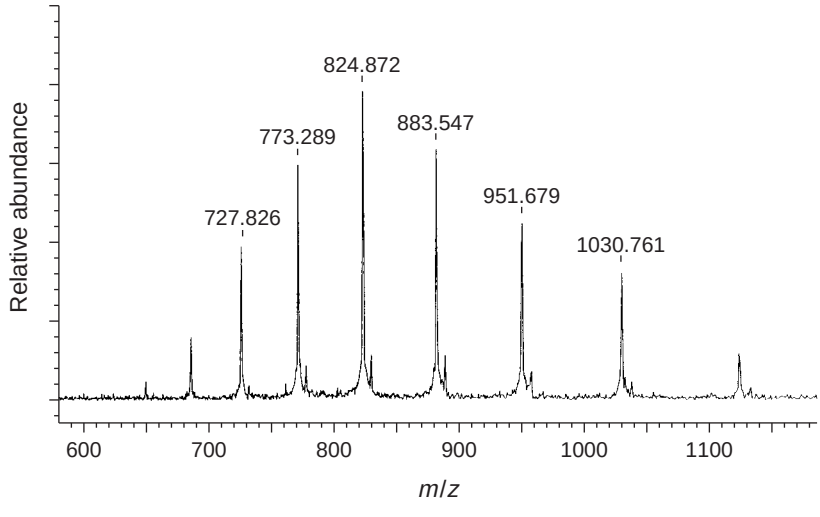

Figure 14 Positive ion ESI mass spectrum of equine cytochrome *c*. All peaks are representative of intact cytochrome *c* molecules, with different charge states (varying numbers of protons attached).

the protein.[119,122] A native protein with a tightly folded conformation (e.g. a protein that has multiple disulfide bonds) may ''bury'' some of the ionizable sites in a relatively hydrophobic interior of the folded molecule. A site which otherwise might carry a charge is thus rendered neutral, and protonation/deprotonation on peripheral sites of the molecule might contribute further to shielding the charge attachment/departure at the interior site. This would result in a shift in the charge state distribution toward lower values. Indeed, the observed maximum charge states are much lower than predicted by the number of basic sites for certain proteins containing multiple cysteine–cysteine disulfide bonds.[119]

For a protein that has undergone denaturation (via heat, or via chemical reduction of the disulfide bond), the biopolymer can become fully extended. Because sites that were previously shielded to charging become exposed, such denaturation often results in a shifting of observed charge states toward higher values.[119,121–124] One example is hen egg-white lysozyme, which shows a maximum charge state of 14; however, upon the addition of dithiothreitol, the maximum charge state observed in the ESI mass spectrum shifts to a value of 20. This latter number of attached protons corresponds to the total number of basic sites on the protein.[119]

2.3.3.2 Effect of Analyte Concentration A varying concentration of the analyte can also change the charge state distribution.[125] At higher concentrations there are more analyte molecules competing for a limited number of excess charges in a droplet, which implies that the charge state distribution will shift toward lower values as compared to lower-concentration solutions.[126] This analysis implies that different charge-state ions of the same analyte may have different response vs concentration characteristics, including different linear dynamic ranges and even different slopes in the linear range. Thus, appropriate caution is due when evaluating the concentration dependences of signals arising from multiply charged analytes.

A mathematical treatment (see Equation 27) has been used to evaluate how many excess charges are available per analyte molecule as a function of analyte concentration, where N/N_0 represents the ratio of the total number of excess charges (N) to the total number of analyte molecules present (N_0) in the droplets:[125]

$$\frac{N}{N_0} = \frac{I/e}{ACV_f} \tag{27}$$

Here, I represents the current leaving the ESI capillary. When this is divided by e, the elemental charge, one obtains the number of excess charges available. To obtain the total number of analyte molecules present in all droplets, one calculates the product (ACV_f), where A is Avogadro's number (6.023×10^{23} molecules mol^{-1}), C is the analyte concentration (mol L^{-1}), and V_f is the solution flow rate into the mass spectrometer (L s^{-1}). It was determined that as the analyte concentration was progressively raised, the value of N/N_0 (representing the number of charges available per analyte molecule) decreased, as did the average observed charge state in the ESI mass spectrum.[125]

2.3.3.3 Effect of Solution-phase Properties Multiply charged ions observed in ESI/MS arise from molecules that undergo either protonation/deprotonation or cation/anion attachment in the initial ''neutral solution'' or during the ESI process. Therefore, it is reasonable to presume that solution-phase equilibria considerations, which are affected by the choice of solvent, pH, and electrolyte concentration, will have an effect on the charge state distribution of generated gas-phase ions.

The characteristics of the solvent influence the formation of gas-phase ions in a variety of ways. For example, the onset voltage for electrospray (i.e. the minimum potential required to form the Taylor cone) will increase with the surface tension of the solution, the spray current will increase with solution conductivity, and the initial droplet size will increase with solvent viscosity.[73] Higher solvent polarity also shifts the dissociation equilibrium to favor the formation of higher charge-state ions in solution (see Equations 28 and 29), thereby favoring the formation of more highly charged ions in the gas phase.

$$AB_2 = AB^- + B^+ \tag{28}$$

$$AB^- = A^{2-} + B^+ \tag{29}$$

Both of these dissociation equilibria shift to the right as the solvent polarity is *increased*, which has the effect of increasing the concentrations of species of higher charge states. Raising the solvent polarity was shown to produce higher charge states of analyte ions in ESI/MS experiments conducted in both the negative ion[127] and the positive ion[128] modes.

Another solution property that will affect the acid–base equilibrium dramatically is solution pH. For the determination of proteins, Chowdhury et al.[122] observed that the lowering of pH below a critical threshold (corresponding to the acidity where denaturation takes place) resulted in a sudden dramatic shift in charge states observed in the ESI mass spectrum. As outlined above, the protein was capable of acquiring a significantly higher level of charge as soon as it became unfolded, but separate experiments revealed that for the relatively rigid structures of the peptides bradykinin and gramicidin S only minor shifts in the charge state distributions occurred as the hydronium ion concentration was varied by seven orders of magnitude.[129] This latter result clearly revealed the

discrepancy between solution-phase equilibria and charge states observed in ESI mass spectra.

Evidence has emerged that electrolytes present in solution, especially the counter-ions associated with charged sites on the molecule, can have an important influence on the observed distribution of charge states of gas-phase ions. Mirza and Chait[130] reported that the different nature of counter-ions present would shift the analyte charge state to different degrees. When present, the following anions were observed to shift charge state distributions to lower values, but by progressively decreasing amounts: $CCl_3COO^- > CF_3COO^- > CH_3COO^- \sim Cl^-$. This type of shift, however, was corroborated[116] to be quite variable and dependent upon the level of interaction between the counter-ion and the analyte under investigation, with certain electrolyte–analyte combinations being quite resistant to any shift.

Fenselau and co-workers[131] examined positive and negative ion ESI mass spectra of myoglobin at varying pH. They reported that for basic solutions (pH 10), where minimal solution-phase protonation was occurring, in positive ion ESI an envelope of multiply charged ions with a maximum charge state of +14 was observed, corresponding to multiply protonated myoglobin. This result is not expected because at pH 10 myoglobin exists virtually exclusively in anionic form in solution. Moreover, at an acidic pH of 3, in the negative ion mode an envelope of multiply deprotonated myoglobin molecules was observed, having a maximum charge state of −11, even though myoglobin exists virtually completely in cationic form in solutions of pH 3.

In order to explain the production of positively charged proteins in basic solutions to which nitrogen bases had been added, Le Blanc et al.[132] proposed that the multiply protonated proteins were desorbed with the nitrogen bases attached. These complexes might then dissociate in the gas phase, resulting in the partitioning of available charge (protons) between the polypeptide and the nitrogen base. This explanation was used to rationalize an observed shift in charge states toward lower values in the ESI mass spectrum obtained from a solution containing a nitrogen base as compared to that of an otherwise identical solution devoid of base. More recently, Boyd and co-workers[133] have rationalized the appearance of protonated molecules (MH^+) in ESI mass spectra obtained from highly basic solutions where initial solution protonation of analyte molecules was negligible. The solution pH was raised via addition of tetramethylammonium hydroxide, an electrolyte that cannot be considered to be a source of protons. A mechanism was proposed whereby tetramethylammonium cations react with water (solvent) to liberate protons from water. These protons of solvent origin attach to analyte molecules as desolvation reaches completion, thus accounting for the multiply protonated molecules that arise from highly basic solutions where initial solution protonation is negligible.

2.3.3.4 Effect of Instrumental Conditions Obtained charge state distributions in ESI/MS are also affected, either subtly or substantially, by instrumental factors such as ion source configuration, counter-current drying gas and cooling gas flow rates, as well as the cone voltage. The design of the ion source will determine the thermal energy and time available for desolvation.[134] In the desolvation step, solvent can depart carrying some of the available excess charges. This process can effectively shift the charge state distributions of analyte ions toward lower values. When the flow rate of heated counter-current drying gas was increased, thus increasing the droplet evaporation rate, charge states shifted to higher values.[93] Moreover, an increase in the source temperature would have a similar effect and can therefore increase charge state values.

Another instrumental factor that exerts a large influence on charge state distributions is the cone voltage, which determines the level of ''in-source'' CID. For gas-phase ions, the cone serves both as an electrostatic lens and as an accelerator of ions. The velocity of ions in the region just after the cone is determined largely by the voltage drop between the cone and the ensuing skimmer. An elevated cone voltage results in more energetic collisions between ions and the neutral gas molecules present, which induces a higher degree of decomposition. Collisions are most efficient at producing dissociation when the pressure is low enough such that an ion has a significant mean free path of travel before colliding into a gas molecule, yet high enough such that the probability of collision is still significant. An elevated cone voltage implies a high translation energy for the ions, resulting in a high uptake of internal energy upon collision. Higher charge-state ions of a given molecule will be accelerated to a greater degree than lower charge-state counterparts, and it is likely that higher charge-state ions will also undergo desolvation at a slower rate. A higher charge state also implies a higher propensity toward reaction, because the number of reactive charge sites is increased and the Coulomb energy is higher. These factors result in an increased tendency for higher charge-state ions to lose charges (via transfer to residual solvent, other solution components, or residual gases). Thus, an increasing cone voltage leads to a shift toward lower charge-state ions in the ESI mass spectrum.[113]

2.3.3.5 Effect of Gas-phase Processes The last factor to be described here that affects the charge state distribution is proton transfer reactions that occur in the gas phase. McLuckey et al.[135,136] demonstrated that dimethylamine would react with multiply charged

proteins to lower charge states via proton abstraction. Winger et al.[134] also reported that the intentional addition of gaseous water could dramatically shift the distribution of charge states toward lower values. Ogorzalek Loo et al.[137] and Cassady et al.[138] also reported a decreasing reactivity of multiply charged protein ions for lower charge states. Williams and co-workers[139–141] offered an experimental approach to quantify the Coulomb energy in multiply charged molecules. The Coulomb energy was calculated based on the differences in apparent gas-phase basicities of individual charge states of multiply protonated peptides, proteins,[142] and diaminoalkanes;[143] calculated Coulomb energies increased exponentially with the charge state.

2.4 Applications

The inherent feature of producing gaseous ions from ions in the solution phase has led to a wide variety of application areas, especially in the biological sciences. Given below are brief summaries of the major application areas of ESI/MS (see Table 6).

2.4.1 Electrospray Ionization Mass Spectrometry of Peptides and Proteins

ESI/MS has been applied successfully to the determination of primary protein structures, including molecular weight determinations and peptide mapping, via tandem MS. In addition, protein conformation[77,78,119,121,122,144,145] and a variety of noncovalent protein complexes,[77,78] such as protein–peptide, protein–metal, protein–small molecule, protein–subunit, and protein–nucleic acid, have been studied intensively.

2.4.2 Electrospray Ionization Mass Spectrometry of Nucleic Acids

The determination of molecular weights allows one to deduce the compositions of nucleic acids having lengths up to 14-mers, or up to 25-mers if one has access to a greater degree of mass accuracy (e.g. using Fourier transform ion cyclotron resonance/MS). Oligonucleotide mapping/sequencing and fragmentation have also been studied intensively by electrospray ionization/tandem mass spectrometry (ESI/MS/MS).[146–150] Protocols have been developed for the detection of modified nucleosides, such as those arising from endonuclease digests of RNA[151] and DNA.[152,153] Noncovalent interactions of nucleic acids have also been studied successfully by ESI/MS,[148] including nucleic acid–nucleic acid, nucleic acid–protein, and nucleic acid–small molecule interactions. ESI/MS has been applied to the determination of products from polymerase chain reactions of DNA samples.[154,155]

Table 6 Summary of applications of ESI/MS

Compounds	Positive ions	Negative ions
Proteins and peptides	$[M + nH]^{n+}$	$[M - nH]^{n-}$
Nucleic acids	$[M + nH]^{n+}$	$[M - nH]^{n-}$
Carbohydrates and lipids	$[M + nH]^{n+}$	$[M - nH]^{n-}$ $[M + Cl]^{-}$
Pharmacokinetics	$[M + nH]^{n+}$	$[M - nH]^{n-}$
Inorganics and organometallics	$ML_x{}^{n+}$	$ML_x{}^{n-}$

2.4.3 Electrospray Ionization Mass Spectrometry of Carbohydrates and Lipids

Both derivatized and underivatized carbohydrates and lipids can be investigated by ESI/MS.[156–158] Neutral oligosaccharides have been observed as sodiated ions rather than protonated ions in the positive ion mode, due to the stronger tendency for Na^+ addition. Other metal cations have also been observed to form adducts in positive ion ESI/MS, such as Ca^{2+} and Mg^{2+} for oligosaccharides,[159] and Ca^{2+} for glycolipids.[160] However, in the negative ion mode, glycoconjugates display the unusual features of chloride attachment in the presence of chlorinated solvents,[161,162] and $C_4H_9O_2^-$ attachment in the presence of methanol.[161] Acidic glycoconjugates such as gangliosides,[156] sulfated Lewis-type oligosaccharides,[163] phospholipids,[164] acylglycerols,[165] lipid A,[166,167] and glycosylphosphatidylinositols[168] have also been investigated in both positive and negative ion modes; however, negative ion ESI/MS generally exhibits higher sensitivity for acidic glycoconjugates.[157,169]

2.4.4 Electrospray Ionization Mass Spectrometry of Drug Metabolites and Pharmacokinetics

ESI/MS has been widely used to determine drug metabolites in biological fluids when a mixture of radioactive isotope-labeled and normal parent drugs is employed.[170,171] A mass profile, containing both radiolabeled and non-radiolabeled pseudomolecular ions, reveals all the compounds representing metabolites of the parent drug derived from biodegradation and biosynthesis processes. LC with ESI/MS[66] and capillary electrophoresis with ESI/MS[68,79,172] are well-established techniques for the isolation and determination of drug metabolites, as well as pharmacokinetics.[173]

2.4.5 Electrospray Ionization Mass Spectrometry of Inorganic and Organometallic Complexes

Study of inorganic and organometallic complexes by ESI/MS has also received a high level of attention.[174–179]

Many inorganic or organometallic compounds can form ions in solution that are amenable to ESI determination, such as solvated metal ions,[174,175] metal oxides,[180] metal oxo ions,[181] polyoxoanions,[182] organoarsenics,[183] and lanthanide metal ions.[176,184] ESI/MS also presents a new research area for coordination chemistry, namely the study of gas-phase metal complexes that have been studied previously only in solution phase. The inherent soft ionization character of ESI allows the structural determination of a variety of complexes of metals with various entities such as proteins and peptides, amino acids, nucleic acids, carbohydrates, as well as the characterization of supermolecular complexes.

3 SUMMARY

The API techniques of APCI and ESI represent valuable and convenient means of creating gas-phase ions out of analytes dissolved in solution. Coupling of these ionization techniques to virtually all types of mass spectrometers (quadrupole, magnetic sector, time-of-flight, quadrupole ion trap, Fourier transform ion cyclotron resonance) is at an advanced state. Both techniques are extremely well suited for on-line coupling to liquid chromatographs, and it is fair to say that these are the two preferred mass spectrometric ionization methods for LC/MS experiments. Moreover, the types of compounds for which one would tend to choose LC as a preferred separation method (i.e. moderately polar, thermally labile compounds) often have the types of molecules that are ionized most efficiently by APCI and ESI. Thus, both APCI/MS and ESI/MS are considered to offer extremely high sensitivities for the analysis of polar molecules, with ESI performing somewhat better for very large, highly polar, and ionic compounds, whereas APCI offers certain advantages for lower-polarity analytes. Both APCI and ESI are considered to be ''soft'' ionization techniques that yield predominantly ionic forms of intact molecules. ESI is so gentle that it allows the investigation of weak (non-covalent) solution-phase interactions with the least perturbation during the transfer into the gas phase. These examples prove that both APCI and ESI are established, indispensable tools in the MS arsenal.

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
APCI/MS	Atmospheric Pressure Chemical Ionization/Mass Spectrometry
API	Atmospheric Pressure Ionization
CID	Collision-induced Dissociation
ESI	Electrospray Ionization
ESI/MS	Electrospray Ionization/Mass Spectrometry
ESI/MS/MS	Electrospray Ionization/ Tandem Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
GPA	Gas-phase Acidity
GPB	Gas-phase Basicity
HPLC	High-performance Liquid Chromatography
HPLC/APCI/MS	High-performance Liquid Chromatography/ Atmospheric Pressure Chemical Ionization/ Mass Spectrometry
HPLC/APCI/MS/MS	High-performance Liquid Chromatography/ Atmospheric Pressure Chemical Ionization/ Tandem Mass Spectrometry
LC	Liquid Chromatography
LC/APCI/MS	Liquid Chromatography/ Atmospheric Pressure Chemical Ionization/ Mass Spectrometry
LC/APCI/MS/MS	Liquid Chromatography/ Atmospheric Pressure Chemical Ionization/Tandem Mass Spectrometry
MS	Mass Spectrometry
PA	Proton Affinity
PAH	Polycyclic Aromatic Hydrocarbon
PEG	Polyethylene Glycol
SFC	Supercritical Fluid Chromatography
SFC/APCI/MS/MS	Supercritical Fluid Chromatography/ Atmospheric Pressure Chemical Ionization/ Tandem Mass Spectrometry
SIM	Selected Ion Monitoring
TCE	Trichloroethylene

RELATED ARTICLE

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History

REFERENCES

1. E.C. Horning, M.G. Horning, D.I. Carroll, I. Dzidic, R.N. Stillwell, 'New Picogram Detection System Base on a Mass Spectrometer with an External Ionization Source at Atmospheric Pressure', *Anal. Chem.*, **45**(6), 936–943 (1973).
2. E.C. Horning, D.I. Carroll, I. Dzidic, K.D. Haegele, M.G. Horning, R.N. Stillwell, 'Liquid Chromatograph–Mass Spectrometer–Computer Analytical Systems: A Continuous-flow System Based on Atmospheric Pressure Ionization Mass Spectrometry', *J. Chromatogr.*, **99**, 13–21 (1974).
3. D.I. Carroll, I. Dzidic, R.N. Stillwell, M.G. Horning, E.C. Horning, 'Subpicogram Detection System for Gas-phase Detection Based upon Atmospheric Pressure Ionization (API) Mass Spectrometry', *Anal. Chem.*, **46**, 706–710 (1974).
4. D.I. Carroll, I. Dzidic, R.N. Stillwell, K.D. Haegele, E.C. Horning, 'Atmospheric Pressure Ionization Mass Spectrometry: Corona Discharge Ion Source for Use in Liquid Chromatograph–Mass Spectrometer–Computer Analytical System', *Anal. Chem.*, **47**, 2369 (1975).
5. D.I. Carroll, I. Dzidic, M.G. Horning, R.N. Stillwell, 'Atmospheric Pressure Ionization Mass Spectrometry', *Appl. Spectrosc. Rev.*, **17**, 337–406 (1981).
6. R.K. Mitchum, W.A. Korfmacher, 'Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **55**(14), 1485A–1499A (1983).
7. A.P. Bruins, 'Mass Spectrometry with Ion Sources Operating at Atmospheric Pressure', *Mass Spectrom. Rev.*, **10**, 53 (1991).
8. M. Herderich, E. Richling, R. Roscher, C. Schneider, W. Schwab, H.U. Humpf, P. Schreier, 'Application of Atmospheric Pressure Ionization HPLC–MS–MS for the Analysis of Natural Products', *Chromatographia*, **45**, 127–132 (1997).
9. A. Dono, C. Paradisi, G. Scorrano, 'Abatement of Volatile Organic Compounds by Corona Discharge. A Study of the Reactivity of Trichloroethylene under Atmospheric Pressure Ionization Conditions', *Rapid Commun. Mass Spectrom.*, **11**, 1687–1694 (1997).
10. N.J. Alcock, C. Eckers, D.E. Gaues, M.P.L. Gaues, M.S. Lant, M.A. McDowall, M. Rossiter, R.W. Smith, S.A. Westwood, H.Y. Wong, *J. Chromatogr.*, **251**, 165–174 (1982).
11. P.J. Arpino, G. Guiochon, 'Optimization of the Instrumental Parameters of a Combined Liquid Chromatograph–Mass Spectrometer Coupled by an Interface for Direct Liquid Introduction', *J. Chromatogr.*, **251**, 153–164 (1982).
12. C.R. Blakley, J.J. Carmody, M.L. Vestal, 'A New Soft Ionization Technique for Mass Specrometry of Complex Molecules', *J. Am. Chem. Soc.*, **102**, 5931–5933 (1980).
13. M. Yamashita, J.B. Fenn, 'Electrospray Ion Source. Another Variation on the Free-jet Theme', *J. Phys. Chem.*, **88**, 4451–4459 (1984).
14. R.C. Willoughby, R.F. Browner, 'Monodisperse Aerosol Generation Interface for Combining Liquid Chromatography with Mass Spectroscopy', *Anal. Chem.*, **56**, 2626–2631 (1984).
15. Dzidic, D.I. Carroll, R.N. Stillwell, E.C. Horning, 'Comparison of Positive Ions Formed in Nickel-63 and Corona Discharge Ion Sources Using Nitrogen, Argon, Isobutane, Ammonia and Nitric Oxide as Reagents in Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **48**(12), 1763–1768 (1976).
16. B.L.M. van Baar, 'Ionization Methods in LC–MS and LC–MS–MS (TSP, APCI, ESP and CF–FAB)', in *Applications of LC–MS in Environmental Chemistry*, ed. D. Barcelo, Elsevier, Amsterdam, 71–133, 1996.
17. M.T. Galceran, E. Moyano, 'Determination of Hydroxy Polycyclic Aromatic Hydrocarbons by Liquid Chromatography–Mass Spectrometry: Comparison of Atmospheric Pressure Chemical Ionization and Electrospray', *J. Chromatogr. A*, **731**, 75–84 (1996).
18. C. Aguilar, I. Ferrer, F. Borrull, R.M. Marce, D. Barcelo, 'Comparison of Automated On-line Solid-phase Extraction Followed by Liquid Chromatography Mass Spectrometry with Atmospheric Pressure Chemical Ionization and Particle Beam Mass Spectrometry for the Determination of a Priority Group of Pesticides in Environmental Waters', *J. Chromatogr. A*, **794**(1/2), 147–163 (1998).
19. M. Sakairi, Y. Kato, 'Multi-atmospheric Pressure Ionization Interface for Liquid Chromatography–Mass Spectrometry', *J. Chromatogr. A*, **794**, 391–406 (1998).
20. M. Vairamani, U.A. Mirza, R. Srinivas, 'Unusual Positive Reagents in Chemical Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **9**, 235–258 (1990).
21. O. Jauregui, E. Moyano, M.T. Galceran, 'Liquid Chromatography Atmospheric Pressure Ionization Mass Spectrometry for the Determination of Chloro- and Nitrophenolic Compounds in Tap Water and Seawater', *J. Chromatogr. A*, **787**(1/2), 79–89 (1997).
22. D. Puig, D. Barcelo, I. Silgoner, M. Grasserbauer, 'Comparison of Three Different Liquid Chromatography–Mass Spectrometry Interfacing Techniques for the Determination of Priority Phenolic Compounds in Water', *J. Mass Spectrom.*, **31**(11), 1297–1307 (1996).
23. J.H. Bowie, 'The Fragmentations of Even-electron Organic Negative Ions', *Mass Spectrom. Rev.*, **9**(3), 349–379 (1990).
24. L.H. Couch, M.I. Churchwell, D.R. Doerge, W.H. Tolleson, P.C. Howard, 'Identification of Ceramides in Human Cells Using Liquid Chromatography with Detection by Atmospheric Pressure Chemical Ionization–Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 504–512 (1997).

25. R.B. van Breemen, D. Nikolic, X.Y. Xu, Y.S. Xiong, M. van Lieshot, C.E. West, A.B. Schilling, 'Development of a Method for Quantitation of Retinol and Retinyl Palmitate in Human Serum Using High-performance Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. A*, **794**(1/2), 245–251 (1998).
26. R. Koeber, R. Niessner, J.M. Bayona, 'Comparison of Liquid Chromatography Mass Spectrometry Interfaces for the Analysis of Polar Metabolites of Benzo[*a*]pyrene', *Fresenius' J. Anal. Chem.*, **359**, 267–273 (1997).
27. V.F. Fredline, P.J. Taylor, H.M. Dodds, A.G. Johnson, 'A Reference Method for the Analysis of Aldosterone in Blood by High-performance Liquid Chromatography Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry', *Anal. Biochem.*, **252**, 308–313 (1997).
28. F.M. Xie, P. Wong, K. Yoshioka, R.G. Cooks, P.J. Kissinger, 'Determination of Melatonin in Commercially Available Products by LCEC and LC/MS/MS', *Liq. Chromatogr. Relat. Technol.*, **21**, 1273–1282 (1998).
29. S. Eerola, I. Otegui, L. Saari, A. Rizzo, 'Application of Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry and Tandem Mass Spectrometry to the Determination of Volatile Nitrosamines in Dry Sausages', *Food Addit. Contam.*, **15**, 270–279 (1998).
30. S. Barnes, L. Coward, M. Kirk, J. Sfakianos, 'HPLC Mass Spectrometry Analysis of Isoflavones', *Proc. Soc. Exp. Biol. Med.*, **217**, 254–262 (1998).
31. P. Laakso, P. Manninen, 'Identification of Milk Fat Triacylglycerols by Capillary Supercritical Fluid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Lipids*, **32**, 1285–1295 (1997).
32. P.J. Laakso, 'Characterization of alpha- and gamma-Linolenic Acid Oils by Reversed-phase High-performance Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Am. Oil Chemists Soc.*, **74**, 1291–1300 (1997).
33. P. Manninen, P. Laakso, 'Capillary Supercritical Fluid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry of gamma- and alpha-Linolenic Acid Containing Triacylglycerols in Berry Oils', *Lipids*, **32**(8), 825–831 (1997).
34. P. Manninen, P.J. Laakso, 'Capillary Supercritical Fluid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry of Triacylglycerols in Berry Oils', *Am. Oil Chemists Soc.*, **74**, 1089–1098 (1997).
35. H.R. Mottram, S.E. Woodbury, R.P. Evershed, 'Identification of Triacylglycerol Positional Isomers Present in Vegetable Oils by High Performance Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 1240–1252 (1997).
36. B. Backstrom, M.D. Cole, M.J. Carrott, D.C. Jones, G. Davidson, K. Coleman, 'A Preliminary Study of the Analysis of Cannabis by Supercritical Fluid Chromatography with Atmospheric Pressure Chemical Ionisation Mass Spectroscopic Detection', *Sci. Justice*, **37**(2), 91–97 (1997).
37. R.B. van Breemen, 'Liquid Chromatography Mass Spectrometry of Carotenoids', *Pure Appl. Chem.*, **69**(10), 2061–2066 (1997).
38. S. Lacorte, G. Jeanty, J.L. Marty, D. Barcelo, 'Identification of Fenthion and Temephos and their Transformation Products in Water by High-performance Liquid Chromatography with Diode Array Detection and Atmospheric Pressure Chemical Ionization Mass Spectrometric Detection', *J. Chromatogr. A*, **777**(1), 99–114 (1997).
39. D. Puig, I. Silgoner, M. Grasserbauer, D. Barcelo, 'Part-per-trillion Level Determination of Priority Methyl-, Nitro-, and Chlorophenols in River Water Samples by Automated On-line Liquid/Solid Extraction Followed by Liquid Chromatography Mass Spectrometry Using Atmospheric Pressure Chemical Ionization and Ion Spray Interfaces', *Anal. Chem.*, **69**(14), 2756–2761 (1997).
40. J.V. Pothuluri, D.R. Doerge, M.I. Churchwell, P.P. Fu, C.E. Cerniglia, 'Fungal Metabolism of Nitrofluoranthenes', *J. Toxicol. Environ. Health A*, **53**(2), 153–174 (1998).
41. C. Rafols, D. Barcelo, 'Determination of Mono- and Disulfonated Azo Dyes by Liquid Chromatography Atmospheric Pressure Ionization Mass Spectrometry', *J. Chromatogr. A*, **777**(1), 177–192 (1997).
42. M. Takino, T. Yamagami, S. Daishima, 'Simultaneous Determination of *n*-Methylcarbamate Pesticides by LC/Atmospheric Pressure-ionization MS', *Buseki Kagaku*, **46**, 555–561 (1997).
43. X.J. Zang, E.K. Fukuda, J.D. Rosen, 'Method for the Determination of Veratridine and Cevadine, Major Components of the Natural Insecticide Sabadilla, in Lettuce and Cucumbers', *J. Agric. Food Chem.*, **45**(5), 1758–1761 (1997).
44. K.A. Barnes, R.J. Fussell, J.R. Startin, M.K. Pegg, S.A. Thorpe, S.L. Reynolds, 'High-performance Liquid Chromatography/Atmospheric Pressure Chemical Ionization Mass Spectrometry with Ionization Polarity Switching for the Determination of Selected Pesticides', *Rapid Commun. Mass Spectrom.*, **11**, 117–123 (1997).
45. A. Oubina, D. Puig, J. Gascon, D. Barcelo, 'Determination of Pentachlorophenol in Certified Wastewaters, Soil Samples and Industrial Effluents Using ELISA and Liquid Solid Extraction Followed by Liquid Chromatography', *Anal. Chim. Acta*, **346**(1), 49–59 (1997).
46. P.A. Fox, J. Carter, P. Farrimond, 'Analysis of Bacteriohopanepolyols in Sediment and Bacterial Extracts by High Performance Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**(10), 609–612 (1998).

47. A. Jegorov, V. Havlicek, P. Sedmera, 'Rapid Screening of Dextruxins by Liquid Chromatography Mass Spectrometry', *J. Mass Spectrom.*, **33**(3), 274–280 (1998).
48. R.F. Venn, B. Kaye, P.V. Macrae, K.C. Saunders, 'Clinical Analysis of Sampatrilat, a Combined Renal Endopeptidase and Angiotensin-converting Enzyme Inhibitor – I: Assay in Plasma of Human Volunteers by Atmospheric-pressure Ionisation Mass Spectrometry Following Derivatisation with BF_3–Methanol', *J. Pharm. Biomed. Anal.*, **16**(5), 875–881 (1998).
49. W.C. Byrdwell, 'Dual Parallel Mass Spectrometers for Analysis of Sphingolipid, Glycerophospholipid and Plasmalogen Molecular Species', *Rapid Commun. Mass Spectrom.*, **12**(5), 256–272 (1998).
50. Y.C. Ma, H.Y. Kim, 'Determination of Steroids by Liquid Chromatography Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **8**(9), 1010–1020 (1997).
51. K. Shimada, K. Mitamura, I. Nakatani, 'Characterization of Monoglucuronides of Vitamin D-2 and 25-Hydroxyvitamin D-2 in Rat Bile Using High-performance Liquid Chromatography–Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr.*, **690**(1/2), 348–354 (1997).
52. M.J. Bogusz, R.D. Maier, M. Erkens, S. Driessen, 'Determination of Morphine and its 3- and 6-Glucuronides, Codeine, Codeine-Glucuronide and 6-Monoacetylmorphine in Body Fluids by Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *J. Chromatogr. B*, **703**(1/2), 115–127 (1997).
53. C.X. Gao, K.H. Bui, A.S. Martz, N.C. LeDonne, M.C. Dyroff, 'Determination of Zd9583, a Thromboxane Receptor Antagonist, in Human Plasma and Urine by Liquid Chromatography Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry', *J. Chromatogr. B*, **695**(2), 299–307 (1997).
54. M. Zell, C. Husser, R. Erdin, G. Hopfgartner, 'Simultaneous Determination of a Potassium Channel Opener and its Metabolite in Rat Plasma with Column-switching Liquid Chromatography Using Atmospheric Pressure Chemical Ionisation', *J. Chromatogr. B*, **694**(1), 135–143 (1997).
55. K.A. Hutton, D.E. Games, 'The Analysis of Beta-blockers by Capillary Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **11**, 735–744 (1997).
56. F. Susanto, H. Reinauer, 'Screening and Simultaneous Quantitative Measurement of Six Sulfonylureas in Serum by Liquid Chromatography Mass Spectrometry with Atmospheric-pressure Chemical-ionization (APCI LC/MS)', *Fresenius' J. Anal. Chem.*, **357**(8), 1202–1205 (1997).
57. D. Koga, T. Santa, T. Fukushima, H. Homma, K. Imai, 'Liquid Chromatographic–Atmospheric Pressure Chemical Ionization Mass Spectrometric Determination of Anandamide and its Analogs in Rat Brain and Peripheral Tissues', *J. Chromatogr. B*, **690**(1/2), 7–13 (1997).
58. K. Nojima, S. Fujimaki, R.C. Hertsens, T. Morita, 'Application of Liquid-chromatography Atmospheric-pressure Chemical-ionization Mass-spectrometry to a Sector Mass-spectrometer', *J. Chromatogr. A*, **712**(1), 17–19 (1995).
59. A.L. Lafleur, K. Taghizadeh, J.B. Howard, J.F. Anacleto, M.A. Quillim, 'Characterization of Flame-generated C-10 to C-160 Polycyclic Aromatic Hydrocarbons by Atmospheric-pressure Chemical-ionization Mass-spectrometry with Liquid Introduction via Heated Nebulizer Interface', *J. Am. Soc. Mass Spectrom.*, **7**(3), 276–286 (1996).
60. A.C. Hogenboom, W.M.A. Niessen, U.A.T. Brinkman, 'Rapid Target Analysis of Microcontaminants in Water by On-line Single-short-column Liquid Chromatography Combined with Atmospheric Pressure Chemical Ionization Ion-trap Mass Spectrometry', *J. Chromatogr. A*, **794**(1/2), 201–210 (1998).
61. M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice, 'Molecular Beams of Macroions', *J. Chem. Phys.*, **49**, 2240–2249 (1968).
62. M. Yamashita, J. Fenn, 'Electrospray Ion Source. Another Variation on the Free-jet Theme', *J. Phys. Chem.*, **88**, 4451–4459 (1984).
63. M. Yamashita, J. Fenn, 'Negative Ion Production with the Electrospray Ion Source', *J. Phys. Chem.*, **88**, 4671–4675 (1984).
64. M.L. Aleksandrov, L.N. Gall, V.N. Krasnov, V.I. Nikolaev, V.A. Pavlenko, V.A. Shkurov, *Dokl. Akad. Nauk. SSSR*, **277**, 379 (1984).
65. R.M. Caprioli, 'On-line LC/MS and CE/MS', in *Biological Mass Spectrometry: Present and Future*, eds. T. Matsuo, R.M. Caprioli, M.L. Gross, John Wiley & Sons, New York, 1994.
66. R.D. Voyksner, 'Combining Liquid Chromatography with Electrospray Mass Spectrometry', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 323–341, 1997.
67. R.D. Smith, K.J. Light-Wahl, B.E. Winger, D.R. Goodlett, 'Electrospray Ionization', in *Biological Mass Spectrometry: Present and Future*, eds. T. Matsuo, R.M. Caprioli, M.L. Gross, John Wiley & Sons, New York, 1994.
68. J.C. Severs, R.D. Smith, 'Capillary Electrophoresis–Electrospray Ionization Mass Spectrometry', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 343–382, 1997.
69. T. Covey, 'Analytical Characteristics of the Electrospray Ionization Process', in *Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry*, ed. A.P. Snyder, American Chemical Society, Washington, DC, 21–59, 1995.

70. G.J. Van Berkel, S.A. McLuckey, G.L. Glish, 'Electrochemical Origin of Radical Cations Observed in Electrospray Ionization Mass Spectra', *Anal. Chem.*, **64**, 1586–1593 (1992).
71. J.F. Anacleto, M.A. Quilliam, R.K. Boyd, J.B. Howard, A.L. Lafleur, T. Yadav, 'Charge Transfer Ion Spray Liquid Chromatography/Mass Spectrometry Analysis of Fullerenes and Related Compounds from Flame Generated Materials', *Rapid Commun. Mass Spectrom.*, **7**, 229 (1993).
72. X. Xu, W. Lu, R.B. Cole, 'On-line Probe for Fast Electrochemistry/Electrospray Mass Spectrometry. Investigation of Polycyclic Aromatic Hydrocarbons', *Anal. Chem.*, **68**, 4244–4253 (1996).
73. D.P.H. Smith, 'The Electrohydrodynamic Atomization of Liquids', *IEEE Trans. Ind. Appl.*, **22**, 527–535 (1986).
74. I. Hayati, A.I. Bailey, Th.F. Tadros, 'Investigations into the Mechanism of Electrohydrodynamic Spraying of Liquids', *J. Colloid Interface Sci.*, **117**, 205–230 (1987).
75. J.L. Jones, A.R. Dongre, A. Somogyi, V.H. Wysocki, 'Sequence Dependence of Peptide Fragmentation Efficiency Curves Determined by Electrospray Ionization/Surface-induced Dissociation Mass Spectrometry', *J. Am. Chem. Soc.*, **116**, 8368–8369 (1994).
76. D. Suckau, Y. Shi, S.C. Beu, M.W. Senko, J.P. Quinn, F.W. Wampler, F.W. McLafferty, 'Coexisting Stable Conformations of Gas Protein Ions', *Proc. Natl. Acad. Sci. USA*, **90**, 790–793 (1993).
77. J.A. Loo, 'Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **16**, 1–23 (1997).
78. J.A. Loo, R.R.O. Loo, 'Electrospray Ionization of Peptides and Proteins', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 385–419, 1997.
79. A.P. Bruins, T.R. Covey, J.D. Henion, 'Ion Spray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 2642–2646 (1987).
80. G.I. Taylor, 'Disintegration of Water Drops in an Electric Field', *Proc. R. Soc. London, Ser. A*, **280**, 383–397 (1964).
81. P. Kebarle, L. Tang, 'From Ions in Solution to Ions in the Gas Phase', *Anal. Chem.*, **65**, 972A–986A (1993).
82. G.J. Van Berkel, 'The Electrolytic Nature of Electrospray', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 65–105, 1997.
83. G.J. Van Berkel, F. Zhou, J.T. Aronson, 'Changes in Bulk Solution pH Caused by the Inherent Controlled-current Electrolytic Process of an Electrospray Ion Source', *Int. J. Mass Spectrom. Ion Proc.*, **162**, 55–67 (1997).
84. A.T. Blades, M.G. Ikonomou, P. Kebarle, 'Mechanism of Electrospray Mass Spectrometry. Electrospray as an Electrolysis Cell', *Anal. Chem.*, **63**, 2109–2114 (1991).
85. G.J. Van Berkel, F. Zhou, 'Characterization of an Electrospray Ion Source as a Controlled-current Electrolytic Cell', *Anal. Chem.*, **67**, 2916–2923 (1995).
86. R.J. Pfeifer, C.D. Hendricks, 'Parametric Studies of Electrohydrodynamic Spraying', *AIAA J. 1968*, **6**, 496–502 (1968).
87. Lord Rayleigh, 'On the Equilibrium of Liquid Conducting Masses Charged with Electricity', *Philos. Mag.*, **14**, 184 (1882).
88. A. Gomez, K. Tang, 'Charge and Fission of Droplets in Electrostatic Sprays', *Phys. Fluids*, **6**, 404–414 (1994).
89. J. Fernandez de la Mora, J. Navascues, F. Fernandez, J. Rosell-Llompart, 'Generation of Submicron Monodisperse Aerosols in Electrosprays', *J. Aerosol Sci.*, **21**, 5673–5676 (1990).
90. J.V. Iribarne, B.A. Thomson, 'On the Evaporation of Small Ions from Charged Droplets', *J. Chem. Phys.*, **64**, 2287–2294 (1976).
91. B.A. Thomson, J.V. Iribarne, 'Field Induced Ion Evaporation from Liquid Surfaces at Atmospheric Pressure', *J. Chem. Phys.*, **71**, 4451–4463 (1979).
92. J.B. Fenn, J. Rosell, T. Nohmi, F.J. Banks, Jr, 'Electrospray Ion Formation: Desorption versus Desertion', in *Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry*, ed. A.P. Snyder, American Chemical Society, Washington, DC, 60–80, 1995.
93. J.B. Fenn, 'Ion Formation from Charged Droplets: Roles of Geometry, Energy, and Time', *J. Am. Soc. Mass Spectrom.*, **4**, 524–535 (1993).
94. G. Wang, R.B. Cole, 'Solvation Energy and Gas-phase Stability Influences on Alkali Metal Cluster Ion Formation in Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **70**, 873–881 (1998).
95. P. Kebarle, Y. Ho, 'On the Mechanism of Electrospray Mass Spectrometry', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 3–63, 1997.
96. S.J. Gaskell, 'Electrospray: Principles and Practice', *J. Mass Spectrom.*, **32**, 677–688 (1997).
97. F.M. Wampler, A.T. Blades, P. Kebarle, 'Negative Ion Electrospray Mass Spectrometry of Nucleotides: Ionization from Water Solution via SF_6 Discharge Suppression', *J. Am. Soc. Mass Spectrom.*, **4**, 289–295 (1993).
98. R.B. Cole, A.K. Harrata, 'Charge-state Distribution and Electric-discharge Suppression in Negative-ion Electrospray Mass Spectrometry Using Chlorinated Solvents', *Rapid Commun. Mass Spectrom.*, **6**, 536–539 (1992).
99. C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, 'Electrospray Interface for Liquid Chromatographs and Mass Spectrometers', *Anal. Chem.*, **57**, 675–679 (1985).

100. E.C. Huang, T. Wachs, J.J. Couboy, J.D. Henion, 'Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **62**, 713A–725A (1990).
101. M. Wilm, M. Mann, 'Analytical Properties of the Nanoelectrospray Ion Source', *Anal. Chem.*, **68**, 1–8 (1996).
102. T.R. Covey, E.C. Huang, J.D. Henion, 'Structural Characterization of Protein Tryptic Peptides via Liquid Chromatography/Mass Spectrometry and Collision-induced Dissociation of their Doubly Charged Molecular Ions', *Anal. Chem.*, **61**, 1193 (1991).
103. M.G. Ikonomou, P. Kebarle, 'A Heated Electrospray Source for Mass Spectrometry of Analytes from Aqueous Solutions', *J. Am. Soc. Mass Spectrom.*, **5**, 791–799 (1994).
104. M. Wilm, M. Mann, 'Electrospray and Taylor-cone Theory, Dole's Beams of Macromolecules at Last?', *Int. J. Mass Spectrom. Ion Proc.*, **136**, 167–180 (1994).
105. M.R. Emmett, R.M. Caprioli, 'Micro-electrospray Mass Spectrometry: Ultra-high-sensitivity Analysis of Peptides and Proteins', *J. Am. Soc. Mass Spectrom.*, **5**, 605–613 (1994).
106. P.E. Andren, M.R. Emmett, R.M. Caprioli, 'Micro-electrospray: Zeptomole/Attomole per Microliter Sensitivity for Peptides', *J. Am. Soc. Mass Spectrom.*, **5**, 867–869 (1994).
107. J.H. Eahl, D.C. Gale, R.D. Smith, 'Sheathless Capillary Electrophoresis–Electrospray Ionization Mass Spectrometry Using 10 μm I.D. Capillary: Analysis of Tryptic Digests of Cytochrome *c*', *J. Chromatogr.*, **659**, 217–222 (1993).
108. M.S. Kriger, K.D. Cook, R.S. Ramsey, 'Durable Gold-coated Fused Silica Capillaries for Use in Electrospray Mass Spectrometry', *Anal. Chem.*, **67**, 385–389 (1995).
109. M. Acedo, G. Tarrason, J. Piulats, M. Mann, M. Wilm, R. Eritja, 'Preparation of Oligonucleotide–Dexamethasone Conjugates', *Bioorg. Med. Chem. Lett.*, **5**, 1577–1580 (1995).
110. A.P. Bruins, 'ESI Source Design and Dynamic Range Considerations', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 107–136, 1997.
111. J.B. Anderson, R.P. Andres, J.B. Fenn, 'High Intensity and High Energy Molecular Beams', in *Advances in Atomic and Molecular Physics*, eds. D.R. Bates, I. Estermann, Academic Press, New York, Chapter 8, 345–389, 1965.
112. R. Campargue, 'Progress in Overexpanded Supersonic Jets and Skimmed Molecular Beams in Free-jet Zone of Silence', *J. Phys. Chem.*, **88**, 4466–4474 (1984).
113. G. Wang, R.B. Cole, 'Solution, Gas-phase, and Instrumental Parameter Influences on Charge-state Distributions in Electrospray Ionization Mass Spectrometry', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 137–174, 1997.
114. L. Tang, P. Kebarle, 'Effect of the Conductivity of the Electrosprayed Solution on the Electrospray Current. Factors Determining Analyte Sensitivity in Electrospray Mass Spectrometry', *Anal. Chem.*, **63**, 2709–2715 (1991).
115. L. Tang, P. Kebarle, 'Dependence of Ion Intensity in Electrospray Mass Spectrometry on the Concentration of Analytes in the Electrosprayed Solution', *Anal. Chem.*, **65**, 3654–3668 (1993).
116. G. Wang, R.B. Cole, 'Effect of Solution Ionic Strength on Analyte Charge State Distributions in Positive and Negative Ion Electrospray Mass Spectrometry', *Anal. Chem.*, **66**, 3702–3708 (1994).
117. D.R. Zook, A.P. Bruins, 'On Cluster Ions, Ion Transmission and Linear Dynamic Range Limitations in Electrospray Ionization Mass Spectrometry', *Int J. Mass Spectrom. Ion Proc.*, **162**, 129–147 (1997).
118. J.A. Loo, H.R. Udseth, R.D. Smith, *Biomed. Environ. Mass Spectrom.*, **29**, 419–427 (1990).
119. J.A. Loo, C.G. Edmonds, H.R. Udseth, R.D. Smith, 'Effect of Reducing Disulfide-containing Proteins on Electrospray Ionization Mass Spectra', *Anal. Chem.*, **62**, 693–698 (1990).
120. T.R. Covey, R.F. Bonner, B.I. Shushan, J. Henion, 'The Determination of Protein, Oligonucleotide and Peptide Molecular Weights by Ion-spray Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **2**, 249–256 (1988).
121. R.D. Smith, J.A. Loo, R.R. Ogorzalek Loo, M. Busman, H.R. Udseth, 'Principles and Practice of Electrospray Ionization Mass Spectrometry for Large Polypeptides and Proteins', *Mass Spectrom. Rev.*, **10**, 359–451 (1991).
122. S.K. Chowdhury, V. Katta, B.T. Chait, 'Probing Conformational Changes in Proteins by Mass Spectrometry', *J. Am. Chem. Soc.*, **112**, 9012–9013 (1990).
123. J.C.Y. Le Blanc, D. Beauchemin, K.W.M. Siu, R. Guevremont, S.S. Berman, 'Thermal Denaturation of Some Proteins and its Effect on their Electrospray Mass Spectra', *Org. Mass. Spectrom.*, **26**, 831–839 (1991).
124. U.A. Mirza, S.L. Cohen, T. Chait, 'Heat-induced Conformational Changes in Proteins Studied by Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **65**, 1–6 (1993).
125. G. Wang, R.B. Cole, 'Mechanistic Interpretation of the Dependence of Charge State Distributions on Analyte Concentrations in Electrospray Ionization Mass Spectrometry', *Anal. Chem.*, **67**, 2892–2900 (1995).
126. S.K. Chowdury, V. Katta, B.T. Chait, *Rapid Commun. Mass Spectrom.*, **4**, 81–87 (1990).
127. R.B. Cole, A.K. Harrata, 'Solvent Effect on Analyte Charge State, Signal Intensity, and Stability on Negative Ion Electrospray Mass Spectrometry; Implications for

the Mechanism of Negative Ion Formation', *J. Am. Soc. Mass. Spectrom.*, **4**, 546–556 (1993).
128. G. Wang, R.B. Cole, 'Effects of Solvent and Counter-ion on Ion Pairing and Observed Charge States of Diquaternary Ammonium Salts in Electrospray Ionization Mass Spectrometry', *J. Am. Soc. Mass. Spectrom.*, **7**, 1050–1058 (1996).
129. G. Wang, R.B. Cole, 'Disparity between Solution-phase Equilibria and Charge State Distributions in Positive-ion Electrospray Mass Spectrometry', *Org. Mass. Spectrom.*, **29**, 419–427 (1994).
130. U.A. Mirza, B.T. Chait, 'Effect of Anions on the Positive Ion Electrospray Ionization Mass Spectra of Peptides and Proteins', *Anal. Chem.*, **66**, 2898–2904 (1994).
131. M.A. Kelly, M.M. Vestling, C.C. Fenselau, P.B. Smith, 'Electrospray Analysis of Proteins: A Comparison of Positive-ion and Negative-ion Mass Spectra at High and Low pH', *Org. Mass Spectrom.*, **27**, 1143–1147 (1992).
132. J.C.Y. Le Blanc, J. Wang, R. Guevremont, K.W.M. Siu, 'Electrospray Mass Spectra of Protein Cations Formed in Basic Solutions', *Org. Mass. Spectrom.*, **29**, 587–593 (1994).
133. B.A. Mansoori, D.A. Volmer, R.K. Boyd, '"Wrong-way-round" Electrospray Ionization of Amino Acids', *Rapid Commun. Mass Spectrom.*, **11**, 1120–1130 (1997).
134. B.E. Winger, K.J. Light-Wahl, R.D. Smith, 'Gas-phase Proton Transfer Reactions Involving Multiply Charged Cytochrome *c* Ions and Intensity under Thermal Conditions', *J. Am. Soc. Mass Spectrom.*, **3**, 624–630 (1992).
135. S.A. McLuckey, G.J. Van Berkel, G.L. Glish, 'Reactions of Dimethylamine with Multiply Charged Ions of Cytochrome *c*', *J. Am. Chem. Soc.*, **112**, 5668–5670 (1990).
136. S.A. McLuckey, G.L. Glish, G.J. Van Berkel, 'Charge Determination of Product Ions Formed from Collision-induced Dissociation of Multiply Charge Protonated Molecules via Ion/Molecular Reactions', *Anal. Chem.*, **63**, 1971–1978 (1991).
137. R.R. Ogorzalek Loo, J.A. Loo, H.R. Udseth, J.L. Fulton, R.D. Smith, 'Protein Structural Effects in Gas Phase Ion/Molecule Reactions with Diethylamine', *Rapid Commun. Mass Spectrom.*, **6**, 159–165 (1992).
138. C.J. Cassady, J. Wronka, G.H. Kruppa, F.H. Laukien, 'Deprotonation Reactions of Multiply Protonated Ubiquitin Ions', *Rapid Commun. Mass Spectrom.*, **8**, 394–400 (1994).
139. P.D. Schnier, D.S. Gross, E.R. Williams, 'Electrostatic Forces and Dielectric Polarizability of Multiply Protonated Gas-phase Cytochrome *c* Ion Probed by Ion/Molecular Chemistry', *J. Am. Chem. Soc.*, **117**, 6747–6757 (1995a).
140. D.S. Gross, E.R. Williams, 'Experimental Measurement of Coulomb Energy and Intrinsic Dielectric Polarizability of a Multiply Protonated Peptide Ion Using Electrospray Ionization Fourier Transfer Mass Spectrometry', *J. Am. Chem. Soc.*, **117**, 883–890 (1995).
141. E.R. Williams, 'Proton Transfer Reactivity of Large Multiply Charged Ions', *J. Mass Spectrom.*, **31**, 831–842 (1998).
142. P.D. Schnier, D.S. Gross, E.R. Williams, 'On the Maximum Charge State and Proton Transfer Reactivity of Peptides and Protein Ions Formed by Electrospray Ionization', *J. Am. Soc. Mass Spectrom.*, **6**, 1086–1097 (1995).
143. D.S. Gross, S.E. Rodriguez-Cruz, S. Bock, E.R. Williams, 'Measurement of Coulomb Energy and Dielectric Polarizability of Gas-phase Diprotonated Diaminoalkanes', *J. Phys. Chem.*, **99**, 4034–4038 (1995).
144. R. Bakhtiar, S.A. Hofstadler, R.D. Smith, 'Electrospray Ionization Mass Spectrometry', *J. Chem. Ed.*, **73**, A118–A123 (1996).
145. B.N. Pramanik, P.L. Bartner, U.A. Mirza, Y. Liu, A.K. Ganguly, 'Electrospray Ionization Mass Spectrometry for the Study of Non-covalent Complexes: An Emerging Technology', *J. Mass Spectrom.*, **33**, 911–920 (1998).
146. S.A. McLuckey, G.J. Van Berkel, G.L. Glish, 'Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides', *J. Am. Soc. Mass Spectrom.*, **3**, 60–70 (1992).
147. S.A. McLuckey, G. Vaiolyanathan, S. Habibi-Gondarzi, 'Charge vs. Neutral Loss from Multiply Charged Oligonucleotide Anions', *J. Mass Spectrom.*, **30**, 1222 (1995).
148. E. Nordhoff, F. Kirpekar, P. Roepstorff, 'Electrospray Mass Spectrometry of Nucleic Acids', *Mass Spectrom. Rev.*, **15**, 67–138 (1996).
149. J. Boschenok, M.M. Sheil, 'Electrospray Tandem Mass Spectrometry of Nucleotides', *Rapid Commun. Mass Spectrom.*, **10**, 144–149 (1996).
150. P.F. Crane, 'Electrospray Ionization Mass Spectrometry of Nucleic Acids and Their Constituents', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 421–457, 1997.
151. J.A. Kowalak, S.C. Pomerants, P.F. Crain, J.A. McCloskey, 'A Novel Method for the Determination of Post-transcriptional Modification in RNA by Mass Spectrometry', *Nucleic Acids Res.*, **21**, 4577 (1993).
152. P. Janning, W. Schrader, M. Linscheid, 'A New Mass Spectrometric Approach to Detect Modifications in DNA', *Rapid Commun. Mass Spectrom.*, **8**, 1035–1040 (1994).
153. W. Schrader, M. Linscheid, 'Determination of Styrene Oxide Adducts in DNA and DNA Components', *J. Chromatogr.*, **717**, 117–125 (1995).
154. Y. Naito, K. Ishikawa, Y. Koga, T. Tsuneyoshi, H. Terinuma, R. Arakawa, 'Molecular Mass Measurement of Polymerase Chain Reaction Products Amplified from Human Blood DNA by Electrospray Ionization Mass

Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1484 (1995).
155. D.S. Wunschel, K.F. Fox, A. Fox, J.E. Bruce, D.C. Muddiman, R.D. Smith, 'Analysis of Double-strand Polymerase Chain Reaction Products from the *Bacillus cereus* Group by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 29 (1996).
156. B.B. Reinhold, S.-Y. Chan, S. Chan, V.N. Reinhold, 'Profiling Glycosphingolipids Structural Detail: Periodate Oxidation, Electrospray, Collision Induced Dissociation, and Tandem Mass Spectrometry', *Org. Mass Spectrom.*, **29**, 736–74 (1994).
157. V.N. Reinhold, B.B. Reinhold, C.E. Costello, 'Carbohydrate Molecular Weight Profiling, Sequencing, Linkage, and Branching Data: ESI–MS and CID', *Anal. Chem.*, **67**, 1772–1784 (1995).
158. Y. Ohashi, 'Electrospray Ionization Mass Spectrometry of Carbohydrates and Lipids', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 459–498, 1997.
159. A. Fura, J.A. Leary, 'Determination of Ca^{2+}- and Mg^{2+}-coordinated Branched Trisaccharide Isomers: An Electrospray Ionization and Tandem Mass Spectrometry Study', *Anal. Chem.*, **65**, 2805–2808 (1993).
160. G. Siuzdak, K. Ichikawa, T.J. Caulfield, B. Munoz, C. Wong, K.C. Nicolau, 'Evidence of Ca^{2+}-dependent Carbohydrate Association Through Ion Spray Mass Spectrometry', *J. Am. Chem. Soc.*, **115**, 2877–2881 (1993).
161. T. Ii, Y. Ohashi, Y. Nagai, 'Observation of a Novel Adduct Ion in Negative Ion Electrospray Mass Spectrometry of Natural Glycolipids', *Org. Mass Spectrom.*, **28**, 927–928 (1993).
162. J. Zhu, Y. Li, S. Li, R.B. Cole, 'Structural Characterization of Gangliosides Isolated from Mullet Milt Using Electrospray Ionization Tandem Mass Spectrometry', *Glycobiology*, **9**, 985–993 (1999).
163. T. Ii, Y. Ohashi, T. Ogawa, Y. Nagai, 'Negative-ion Fast Atom Bombardment and Electrospray Ionization Tandem Mass Spectrometry for Characterization of Sulfated and Sialyl Lewis-type Glycosphingolipids', *Glycoconj. J.*, **13**, 273–283 (1996).
164. L. Silvestro, R.B. Cole, Scappaticci, D. Libertucci, L. Biancone, G. Camussi, 'Development of a High-performance Liquid Chromatography–Mass Spectrometric Technique with an Ion Spray Interface for the Determination of Platelet-activating Factor (PAF) and Lyso-PAF in Biological Samples', *J. Chromatogr.*, **647**, 261–269 (1993).
165. K.L. Duffin, J.D. Henion, J.J. Shieh, 'Observation of Noncovalent Enzyme–Substrate and Enzyme–Product Complexes by Ion-spray Mass Spectrometry', *J. Am. Chem. Soc.*, **113**, 7818–7819 (1991).
166. A.K. Harrata, L.N. Domelsmith, R.B. Cole, 'Electrospray Mass Spectrometry for Characterization of Lipid A from *Enterobacter agglomerans*', *Biol. Mass Spectrom.*, **22**, 59–67 (1993).
167. S. Chan, V.N. Reinhold, 'Detailed Structural Characterization of Lipid A: Electrospray Ionization Coupled with Tandem Mass Spectrometry', *Anal. Biochem.*, **218**, 63–73 (1994).
168. C.A. Redman, B.N. Green, J.E. Thomas-Oates, V.N. Reinhold, M.A. Ferguson, 'Analysis of Glycosylphosphatidylinositol Membrane Anchors by Electrospray Ionization–Mass Spectrometry and Collision Induced Dissociation', *Glycoconj. J.*, **11**, 187–193 (1994).
169. D. Garozzo, G. Impallomeni, E. Spina, B.N. Green, T. Hutton, 'Linkage Analysis in Disaccharides by Electrospray Mass Spectrometry', *Carbohydr. Res.*, **221**, 253–257 (1991).
170. L. Weidolf, T. Covey, 'Studies on the Metabolism of Omeprazole in the Rat Using Liquid Chromatography/Ionspray Mass Spectrometry and the Isotope Cluster Technique with [^{34}S]Omeprazole', *Rapid Commun. Mass Spectrom.*, **6**, 192–196 (1992).
171. P. Wheelan, R.C. Murphy, 'Metabolism of Leukotriene B4 in Cultured Hepatoma Cells', *Arch. Biochem. Biophys.*, **321**, 381–389 (1995).
172. I.M. Johansson, R. Pavelka, J.D. Henion, 'Determination of Small Drug Molecules by Capillary Electrophoresis–Atmospheric Pressure Ionization Mass Spectrometry', *J. Chromatogr.*, **559**, 515–528 (1991).
173. G.K. Poon, 'Drug Metabolism and Pharmacokinetics', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 499–525, 1997.
174. A.T. Blades, P. Jayaweera, M.G. Ikonomou, P. Kebarle, 'Studies of Alkaline Earth and Transition Metal M^{2+} Gas Phase Ion Chemistry', *J. Chem. Phys.*, **92**, 5900–5906 (1990).
175. A.T. Blades, P. Jayaweera, M.G. Ikonomou, P. Kebarle, 'Ion–Molecular Clusters Involving Doubly-charged Metal Ions (M^{2+})', *Int. J. Mass Spectrom. Ion Proc.*, **102**, 251–267 (1990).
176. V. Katta, S.K. Chowdhury, B.T. Chait, 'Hydrogen/Deuterium Exchange Electrospray Ionization Mass Spectrometry: A Method for Probing Protein Conformational Changes in Solution', *J. Am. Chem. Soc.*, **112**, 5348–5349 (1990).
177. X. Xu, S.P. Nolan, R.B. Cole, 'Electrochemical Oxidation and Nucleophilic Addition Reactions of Metallocenes in Electrospray Mass Spectrometry', *Anal. Chem.*, **66**, 119–125 (1994).
178. R. Colton, A. D'Agostino, J.C. Traeger, 'Electrospray Mass Spectrometry Applied to Inorganic and Organometallic Chemistry', *Mass Spectrom. Rev.*, **14**, 79–106 (1995).

179. C.L. Gatlin, F. Turecek, 'Electrospray Ionization of Inorganic and Organometallic Complexes', in *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, ed. R.B. Cole, John Wiley & Sons, New York, 527–570, 1997.
180. Z.L. Cheng, K.W.M. Siu, R. Guevremont, S.S. Berman, 'Solvent-derived Metal Oxides in Electrospray Mass Spectrometry of Metal Salt Solutions', *Organic Mass Spectrom.*, **27**, 1370–1376 (1992).
181. T.C. Lau, J. Wang, K.W.M. Siu, R. Guevremont, *J. Chem. Soc. Chem. Commun.*, 1487–1488 (1994).
182. J. Le Quan Toui, E. Muller, 'Electrospray Mass Spectrometry in a Magnetic Sector Instrument Applied to Heteropolyacids', *Rapid Commun. Mass Spectrom.*, **8**, 692–694 (1994).
183. K.W.M. Siu, G.J. Gardner, S.S. Berman, 'Atmospheric Pressure Chemical Ionization and Electrospray Mass Spectrometry of Some Organoarsenic Species', *Rapid Commun. Mass Spectrom.*, **2**, 69–71 (1988).
184. I.I. Stewart, G. Horlick, 'Electrospray Mass Spectra of Lanthanides', *Anal. Chem.*, **66**, 3983–3993 (1994).

Chemical Ionization Mass Spectrometry: Theory and Applications

Burnaby Munson
University of Delaware, Newark, USA

1 Introduction **1**
2 Fundamentals of Ion/Molecule Reactions **2**
2.1 Historical Background 2
2.2 Incidental Observations of Ion/Molecule Reactions 3
2.3 Reactions of Ions in Methane 3
2.4 Ion Chemistry of CH_5^+ and $C_2H_5^+$ 3
2.5 Thermochemical Kinetics 4
2.6 Instrumentation 5
3 Methane as a Chemical Ionization Reagent Gas **5**
3.1 Ionic Reactions 5
3.2 Effects of Experimental Parameters 7
4 Other Chemical Ionization Reagent Gases **8**
4.1 Isobutane 8
4.2 Ammonia 8
4.3 Tetramethylsilane 10
4.4 Other Reagent Gases 11
5 Charge Exchange Reagents **11**
5.1 Introduction 11
5.2 Helium 11
5.3 Benzene, Chlorobenzene 12
5.4 Nitric Oxide 12
6 Negative Ion Chemical Ionization **13**
6.1 Introduction 13
6.2 Negative Ion/Molecule Reactions 13
6.3 Electron Capture 14
7 Atmospheric Pressure Ionization **15**
8 Prospects **16**
Abbreviations and Acronyms **17**
Related Articles **17**
References **18**

Chemical ionization mass spectrometry (CIMS) is a technique for forming ions of the compound of interest (analyte, A) by ion/molecule reactions from reactant ions of a reagent gas that is generally present in a much greater abundance than the analyte. The reactant ions are generally produced by electron ionization (EI) of the reagent gas. The ions produced by EI often react with the large excess of the reagent gas to form the actual reagent ions that react with the analyte. CIMS is performed with both positively and negatively charged reactant ions. The most common ion/molecule reactions in CIMS are proton transfer (which forms AH^+ or $[A-H]^-$ ions), hydride transfer (which forms $[A-H]^+$ ions), charge or electron transfer (which forms $A^{+\bullet}$ or $A^{-\bullet}$ions), and adduct formation or attachment (which forms $[A+R]^+$ or $[A+R]^-$ ions). Fragment ions from decompositions of these AH^+, $A^{\pm\bullet}$, $[A-H]^{\pm}$, and $[A+R]^{\pm}$ ions are frequently observed. The extent of fragmentation can be controlled by the choice of reagent gas used and can be predicted to some extent from ionic thermochemical data. Collisionally stabilized electron capture at high pressures (to form $A^{-\bullet}$ ions) is often classified as chemical ionization (CI). The most common use of CIMS in analytical mass spectrometry is to obtain simplified mass spectra of compounds, often one species spectra, which can be used for quantitative analysis of mixtures. CIMS can be performed with any type of mass spectrometer (quadrupole, magnetic, time-of-flight (TOF), Fourier transform ion cyclotron resonance (FTICR), ion trap) and CIMS capabilities are routinely available on many commercial instruments. With ion trap or FTICR mass spectrometers, it is possible to select specific reactant ions. It is generally considered that CI and EI sensitivities are approximately the same in the positive ion mode and that electron capture CI at high pressures for compounds with high electronegativities gives a much greater sensitivity than other EI or CI techniques. CIMS has the same general limitations as electron ionization mass spectrometry (EIMS) on the volatility and thermal stability of the compound being analyzed. However, direct insertion of the sample into the source of the mass spectrometer allows the analysis of relatively involatile and thermally unstable compounds.

1 INTRODUCTION

CIMS is an analytical application of ion/molecule reactions. This analytical technique is an outgrowth of studies of ion/molecule reactions at high (~2 torr) pressures in the source of a mass spectrometer, when it was observed that the distribution of ions in methane was very sensitive to very small amounts of water and ethane.[1] The sensitivity of ionic composition in gases to trace impurities was established many years ago in drift velocity experiments.[2] Subsequent experiments on mixtures of CH_4 with small amounts (~1%) of other compounds showed the rapid reactions of CH_5^+ and $C_2H_5^+$ (the major ions in methane) to form ions characteristic of the added compounds.[3,4]

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

CI was introduced shortly thereafter with a comparison of the ions produced from ion/molecule reactions of CH_5^+ and $C_2H_5^+$ with compounds (their CI spectra) and the EI spectra of these compounds.[5] The CI spectra showed enhanced abundances of ions in the molecular weight region, $[M \pm H]^+$, compared with the EI spectra, $[M^{+\bullet}]$, which aided in the molecular weight determination and identification of these compounds. The formation of the fragment ions could generally be explained by proton or hydride transfer followed by simple cleavage or rearrangement decompositions, and quantitative analysis was demonstrated for a simple mixture of alkanes.[5] The use of CIMS received a big boost a few years later by work demonstrating the usefulness of the technique for the analysis of complex molecules of biomedical interest: ''The biologically important ethanolamine, ephedrine, is a substituted benzyl alcohol and thus the absence of a molecular ion in its EI mass spectrum is not surprising. The CI mass spectrum on the other hand could stand alone as a complete structure proof.''[6]

One common use of CIMS is illustrated by the analysis of cholesterol esters of fatty acids. The EI mass spectra of cholesterol esters are very complex, containing many fragment ions with roughly equal abundances and no $M^{+\bullet}$ ions. The *i*-C_4H_{10} CI mass spectra, on the other hand, were reported to be essentially two species spectra: the cholesteryl ion at *m*/*z* 369 and the protonated acids.[7] Thus, one could easily identify the components in a mixture without prior separation. The simplicity of *i*-C_4H_{10} CI mass spectra and the high likelihood of obtaining MH^+ ions for each compound in a mixture were combined with collisionally induced dissociations (CID) to characterize components in a complex mixture (nutmeg) without chromatographic separations.[8]

One illustration of the selectivity of the technique is shown in an analysis of polymer additives by vaporization from a polypropylene sample directly into the source of a mass spectrometer using a CH_4/1% NH_3 mixture as the reagent gas.[9] In these experiments the polar high molecular weight additives were detected as MH^+ or $[M + NH_4]^+$ ions and hydrocarbons from the polymer were not detected. The methane/ammonia spectra generally contained only one major ion and quantitative analyses could be obtained from the integrated area as the polymer additive was driven from the sample by heating. The additives could be identified from a small set of likely compounds by the molecular weight and CH_4 CI or EI spectra could be obtained for further confirmation.

It is generally considered that the sensitivities in positive and negative ion CI are approximately the same, and approximately the same as the sensitivities for the same compounds in EI because the rate constants for positively and negatively charged ions with neutral molecules are generally comparable in magnitude.[10,11] Negative ion CIMS by definition refers to negative ions produced by reactions of negatively charged reactant ions with the analyte; however, the mechanism of formation of the negative ions is not always well established and electron capture negative ion spectra are obtained at high pressures with compounds of high electron affinities. Negative ion sensitivities for certain compounds containing highly electronegative groups have been reported to be 100–1000 times as high as positive ion sensitivities.[12]

2 FUNDAMENTALS OF ION/MOLECULE REACTIONS

2.1 Historical Background

Ion/molecule reactions were observed, but not necessarily recognized as such, early in the development of mass spectrometry, primarily because of the difficulties in obtaining sufficiently low pressures that collisions between the ions and the large excess of un-ionized molecules would not occur within the ionization region (source) of the mass spectrometer. An ion at *m*/*z* 3 was reported by Thomson from bombardment of many solids with cathode rays and although he assigned the species as H_3^+, he did not develop a mechanism for its formation and perhaps considered H_3 to be a stable compound.[13]

Subsequently Hogness and Lund reported an extensive study on the variation of the relative abundances of ions in H_2 as the pressure was increased at an essentially constant reaction time.[14] These data showed a decrease in the relative abundance of $H_2^{+\bullet}$ and a consequent increase in the relative abundance of H_3^+ with increasing pressure. These observations definitively established the following reactions (Equations 1 and 2) for the formation of H_3^+,

$$H_2 + e^- \longrightarrow H_2^{+\bullet} + 2e^- \quad (1)$$

$$H_2^{+\bullet} + H_2 \longrightarrow H_3^+ + H^\bullet \quad (2)$$

These pressure plots represent the type of experiments that were done to study ion/molecule reactions in many of the early investigations into ion/molecule reactions. Any ion whose relative abundance increased with increasing pressure was at least partially formed by ion/molecule reactions. If the reactions could be studied to a large extent of conversion (as in this case), any ion whose relative abundance decreased with increasing pressure reacted in an ion/molecule reaction.

H_3^+ had one important use in the mass spectrography for the determination of the accurate atomic weight of deuterium. ''The packing fraction of H^2, which is clearly of the greatest interest in connection with the structure

of its nucleus, has been measured ... between the line of helium and that of the triatomic molecule $H^1H^1H^2$ obtained from the discharge in hydrogen enriched by distillation.''[15]

2.2 Incidental Observations of Ion/Molecule Reactions

There were few studies of ion/molecule reactions during the 1930s and 1940s and the primary research emphasis was on ionization processes. The vacuum technology was improved so that pressures and frequencies of collisions of ions with molecules in the ionization region were much lower. Pyrolysis products from the hot wire filaments no longer contaminated the source. An ion at m/z 19, H_3O^+, was observed in experiments with water, and pressure studies showed that it was formed by a second-order process, that is, I(19)/I(18) increased with increasing pressure of water in the mass spectrometer (as noted in section 2.1 for H_3^+ and $H_2^{+\bullet}$).[16] The H_2O system contains two abundant ions from EI, $H_2O^{+\bullet}$ and HO^+; so that a simple pressure study over a small pressure range would not establish which of the two ions was the reactant. Appearance potential measurements indicated that the minimum energy for the formation of H_3O^+ was approximately the same as the ionization potential (IP) of H_2O and significantly below the minimum energy for the formation of HO^+ from water.[16] Consequently, the reactions for the formation of H_3O^+ are Equations (3) and (4):

$$H_2O + e^- \longrightarrow H_2O^{+\bullet} + 2e^- \quad (3)$$

$$H_2O^{+\bullet} + H_2O \longrightarrow H_3O^+ + HO^\bullet \quad (4)$$

The use of ionization and appearance potential measurements has been critical in the determination of reactant/product sequences for ion/molecule reactions.

Analytical mass spectrometry and commercial mass spectrometers were developed in the 1940s and used extensively in the petroleum industry for the quantitative analysis of mixtures of light hydrocarbons.[17,18] Pressures in these instruments were maintained at low values and the experimental conditions were generally standardized to eliminate ion/molecule reactions as complications for quantitative analyses and the determination of isotopic abundances.

Occasional observations of ion/molecule reactions were made, which demonstrate the care that was used in most experiments to eliminate these processes. The determination of ^{13}C in organic compounds is generally done by measuring the I(45)/I(44) ratio in CO_2 after combustion of the samples. It was noted that if H_2O was not eliminated from the combustion products the I(45)/I(44) ratio was abnormally large and increased with increasing sample size: clear indications of ion/molecule reactions.[19] As mass spectrometers became more common and were being used for qualitative analysis and identification of compounds in addition to quantitative analyses of known mixtures, ions were observed at masses larger than the presumed molecular weights of the compounds being introduced. These observations were useful for the determination of molecular weights of polar compounds (such as esters and ethers) whose EI spectra contained essentially no $M^{+\bullet}$ ions.[20,21] Formation of MH^+ ions from reactions of fragment ions of the sample with the neutral sample molecules is now often called 'self-CI'.

2.3 Reactions of Ions in Methane

The modern work on ion/molecule reactions is continuous from the experiments in the 1950s. Three groups were heavily involved in this work: Tal'roze in the Soviet Union; Stevenson and Schissler at Shell Oil; and Field, Franklin, and Lampe at Humble Oil (Esso).[22–24] Tal'roze and Ljubimova's experiments showed that the ratio, I(17)/I(16), increased with increasing pressure of methane in the source from the correct isotopic value of 1.1% to ~3% at the highest pressures used. In addition, the appearance potential for CH_5^+ and the IP of CH_4 were the same within experimental precision.[22] Consequently, the following reaction sequence was established, (Equations 5 and 6):

$$CH_4 + e^- \longrightarrow CH_4^{+\bullet} + 2e^- \quad (5)$$

$$CH_4^{+\bullet} + CH_4 \longrightarrow CH_5^+ + CH_3^\bullet \quad (6)$$

Rate constants were measured for Equation (6) and some simple ion/molecule reactions, including the other major ion/molecule reaction in methane, (Equation 7):

$$CH_3^+ + CH_4 \longrightarrow C_2H_5^+ + H_2, \quad (7)$$

and it was observed that the reactions which could be observed at the low source pressures (~0.001 mmHg) and short times (microseconds) that were accessible at that time were all very fast, with rate constants $\sim 1 \times 10^{-9}$ cm^3 mol^{-1} s^{-1} ($\sim 6 \times 10^{20}$ L mol^{-1} s^{-1}).[23,24] These rate constants correspond to reactions which occur at essentially every collision.

2.4 Ion Chemistry of CH_5^+ and $C_2H_5^+$

The experiments in the 1950s that re-established the occurrence of fast ion/molecule reactions were generally done with conventional mass spectrometers that had been modified to allow operation at slightly higher pressures. However, the pressures were not sufficiently high that significant conversions of reactant ions to product ions could be seen. Subsequently, high-pressure mass spectrometers were developed which could operate at source pressures of a few torr.

It was shown with ultra high purity methane that neither CH_5^+ nor $C_2H_5^+$ reacted at a significant rate with methane to produce new product ions.[1] Experimentally, the relative abundances of CH_5^+ and $C_2H_5^+$ reached constant values as the pressure was increased. Subsequently, the nonreactivity of CH_5^+ and $C_2H_5^+$ with methane was confirmed by ion cyclotron resonance (ICR) experiments at very low pressures and long reaction times.[25] A kinetic analysis shows that the rate expressions contain a term that is the product of concentration of neutral molecules and the ionic reaction time. CIMS experiments at high pressures and short times and Fourier transform ion cyclotron resonance/mass spectrometry (FTICR/MS) experiments at low pressures and long reaction times have approximately the same value for this product, $[X]t \sim 10^{-6}$ torr sec.

Since neither CH_5^+ nor $C_2H_5^+$ reacted with methane, it was possible to study the reactions of CH_5^+ and $C_2H_5^+$ with other compounds by adding a small amount (<1%) of the other compounds to methane and observing the effects on the relative abundances of CH_5^+ and $C_2H_5^+$. As noted in section 2.1, any ion that reacts by an ion/molecule reaction will decrease in relative abundance as the pressure (and number of collisions) increases. Consequently, reaction of CH_5^+ or $C_2H_5^+$ is indicated by a maximum in the relative abundance of either ion as the pressure of a mixture of constant composition is increased and the product ions can easily be identified by mass. The kinetic analysis of this system is essentially the same as that of the consecutive radioactive decay scheme, $A \rightarrow B \rightarrow C$. Analogous experiments were done at the low pressures and long times characteristic of ICR experiments with similar results.[25]

It was shown that CH_5^+ reacted by proton transfer, acting as a Brønsted acid. There was often extensive decomposition of the AH^+ ions of the added substances because CH_5^+ is a strong acid (CH_4 is a weak base) and the AH^{+*} ions are formed with sufficient internal energy that they may decompose. $C_2H_5^+$ reacted by both proton transfer (a less exothermic reaction than proton transfer from CH_5^+) and by hydride transfer, acting as a Lewis acid.[3,4,25,26] Some addition ions were observed with unsaturated compounds, but the dominant product ions were MH^+, $[M-H]^+$, and some fragment ions. Proton transfer and hydride transfer are the dominant reactions in CH_4 CI.

2.5 Thermochemical Kinetics

Since the ions from the analyte are formed by ion/molecule reactions, the rate constants for these ion/molecule reactions are critical in determining the sensitivity and selectivity in CI mass spectrometry. One can model the CI process as a reaction (Equation 8) of a reactant ion, R^+, with the analyte, A, to give products, P_i^+:

$$R^+ + A \longrightarrow \sum P_i^+ + \sum N_i \tag{8}$$

The rate expression for the formation of all the product ions is given in the usual way as Equation (9):

$$\frac{d \sum[P_i^+]}{dt} = k[R^+][A] \tag{9}$$

For a small extent of conversion of R^+ to product ions, P_i^+, (which is generally true in most CI experiments), one may approximate $[R^+]$ with $[R^+]_0$, the initial concentration of the reactant ion. Since the initial concentration of product ions is zero, one obtains the following simple relationship (Equation 10) between the concentrations of product ions, $[P_I^+]$, and reactant ions, $[R^+]_0$, and [A], with the rate constant, k, and reaction time, Δt:

$$\sum[P_I^+] = k[A][R]_0 \Delta t \tag{10}$$

Consequently, one can see that a large rate constant will give a large current of analyte ions and a sensitive analysis and that a very small (or zero) rate constant will produce a few ions or no product ions at all. The sensitivities for CI reagent ions can be correlated with the rate constants for the reactions and the rate constants can be predicted to some extent from theoretical models.[10]

It is one of the major postulates of ion/molecule reactions that endothermic reactions have an activation energy and will be slow and probably not observable under the normal CI conditions of a few collisions of the reactant ion with the analyte molecules. Consequently, one can achieve selectivity in CIMS by choosing a reactant ion which will not react in an exothermic reaction with one analyte (or class of analytes).

For proton transfer reactions, (Equation 11):

$$AH^+ + B \longrightarrow BH^+ + A, \tag{11}$$

one can show that the heat of reaction is given by the difference in proton affinities (PA) (or base strengths) of A and B (Equations 12 and 13):

$$\Delta H^\circ(11) = PA(A) - PA(B) \tag{12}$$

$$PA(X) = \Delta H_F^\circ(H^+) + \Delta H_F^\circ(X) - \Delta H_F^\circ(XH^+) \tag{13}$$

That is, the proton transfer reaction follows the usual rules of acid–base reactions: in an exothermic reaction, the stronger base accepts the proton from the conjugate acid of the weaker base.

For hydride transfer reactions, (Equation 14):

$$R^+ + NH \longrightarrow RH + N^+, \tag{14}$$

one can show that the heat of reaction is determined by the difference in hydride affinities (HA) of the two ions,

$$\Delta H^\circ(14) = \mathrm{HA}(\mathrm{N}^+) - \mathrm{HA}(\mathrm{R}^+) \quad (15)$$

$$\mathrm{HA}(\mathrm{X}^+) = \Delta H_\mathrm{F}{}^\circ(\mathrm{X}^+) + \Delta H_\mathrm{F}{}^\circ(\mathrm{H}^-) - \Delta H_\mathrm{F}{}^\circ(\mathrm{XH}) \quad (16)$$

That is, in an exothermic reaction, the stronger Lewis acid abstracts a hydride ion from the conjugate base of the weaker Lewis acid.

From the acid strength of a protonating reactant ion, RH^+, or the proton affinity of the conjugate base, R, and the proton affinity of the analyte A, one can determine whether the CI reaction,

$$\mathrm{RH}^+ + \mathrm{A} \longrightarrow \mathrm{AH}^+ + \mathrm{R}, \quad (17)$$

will be exothermic and, therefore, most probably fast, or endothermic and, therefore, most probably slow (or essentially zero). Similarly from the Lewis acidities of the reactant ion, R^+ and the analyte ion, A^+, one can determine whether the CI reaction,

$$\mathrm{R}^+ + \mathrm{AH} \longrightarrow \mathrm{A}^+ + \mathrm{RH}, \quad (18)$$

will be exothermic and fast or endothermic and slow. Some data on PAs and HAs are available for simple species, and one can often make reasonable guesses from chemical intuition about more complex molecules.

2.6 Instrumentation

CIMS can be done with any type of mass spectrometer. The ion source must be much tighter than a conventional EI source to allow high pressures in the ionization region and low pressures in the analyzer region. Small electron entrance and ion exit apertures are essential. All other connections must be very tight to reduce the leakage of gas from the source. The reactions that produce the analyte ions occur in the source of the mass spectrometer and are not dependent on the type of mass analysis that is used to detect the ions. Magnetic, TOF, and quadrupole mass spectrometers have been used in high pressure CIMS. Magnetic instruments operate with the source of the mass spectrometer at a potential that is well above (or below, for negative ions) ground. The pressure region for most CI experiments (a few tenths of a torr) is the region where electric discharges are the most common. However, by using appropriate restrictors in the inlet line for the CI reagent gas, a sufficiently large pressure drop can be achieved that discharges do not occur from the source through the reagent gas to ground. Very good differential pumping is necessary to allow the source to operate at high pressures without degrading the resolution of the instrument or causing significant amounts of collisionally induced decompositions in the flight path of the ions. CI, however, is a standard addition for modern commercial magnetic mass spectrometers.

Quadrupole mass spectrometers are somewhat more forgiving about pressure problems since the sources of the instruments are not at high voltage. The pathlength of the ions in the analyzer region is also shorter and the resolution is generally not sufficiently large that collisional broadening of the peaks will create problems in mass detection.

Ion trap mass spectrometers have been used for CI studies in a lower pressure region with success. FTICR mass spectrometers have also been used for CI studies in the ultra low pressure region. Both of these instruments have the advantage that the CI reactant ions can be individually selected, in contrast with high pressure instruments. That is, using an ion trap or FTICR mass spectrometer, one could obtain CI mass spectra from CH_5^+ only, whereas with the high pressure instruments, one can only obtain CI spectra using both CH_5^+ and $C_2H_5^+$ ions.

3 METHANE AS A CHEMICAL IONIZATION REAGENT GAS

3.1 Ionic Reactions

Methane was the first CI reagent gas used and it is still used when one wants a nonselective reagent gas that will ionize all samples. The dominant ions in the EI mass spectrum of methane are $CH_4^{+\bullet}$ at m/z 16 (~50%) and CH_3^+ at m/z 15 (~40%) and $CH_2^{+\bullet}$ at m/z 14 (~10%). These ions react with methane to form a set of ions (CH_5^+, $C_2H_5^+$, $C_3H_5^+$) that do not react with methane but react with the analyte (Equations 6, 7, 19 and 20):

$$\mathrm{CH_4}^{+\bullet} + \mathrm{CH_4} \longrightarrow \mathrm{CH_5}^+ + \mathrm{CH_3}^\bullet \quad (6)$$

$$\mathrm{CH_3}^+ + \mathrm{CH_4} \longrightarrow \mathrm{C_2H_5}^+ + \mathrm{H_2}, \quad (7)$$

$$\mathrm{CH_2}^{+\bullet} + \mathrm{CH_4} \longrightarrow \mathrm{C_2H_3}^+ + \mathrm{H_2} + \mathrm{H}^\bullet \quad (19)$$

$$\mathrm{C_2H_3}^+ + \mathrm{CH_4} \longrightarrow \mathrm{C_3H_5}^+ + \mathrm{H_2} \quad (20)$$

The dominant reaction of CH_5^+ is proton transfer, generally with extensive dissociation because CH_4 is a very weak base (or CH_5^+ is a very strong gas-phase acid). The dissociation processes are the loss of H_2 and the loss of stable neutral fragments. Some of the initially protonated species may be stabilized by collision Equations (21–24):

$$\mathrm{CH_5}^+ + \mathrm{A} \longrightarrow \mathrm{AH}^{+*} + \mathrm{CH_4} \quad (21)$$

$$\mathrm{AH}^{+*} \longrightarrow [\mathrm{A} - \mathrm{H}]^+ + \mathrm{H_2} \quad (22)$$

$$\mathrm{AH}^{+*} \longrightarrow [\mathrm{A} - \mathrm{Z}]^+ + \mathrm{ZH} \quad (23)$$

$$\mathrm{AH}^{+*} + \mathrm{CH_4} \longrightarrow \mathrm{AH}^+ \quad (24)$$

$C_2H_5^+$ will undergo similar dissociative proton transfer reactions, although the extent of decomposition of AH^{+*} will be less because $C_2H_5^+$ is a weaker Brønsted acid than CH_5^+ (C_2H_4 is a stronger base than CH_4). In addition, $C_2H_5^+$ can react as a Lewis acid by H^- abstraction, and the intermediate complex can be stabilized by collisions Equations (25–27):

$$C_2H_5^+ + A \longrightarrow [A - H]^+ + C_2H_6 \quad (25)$$

$$C_2H_5^+ + A \longrightarrow [A + C_2H_5]^{+*} \quad (26)$$

$$[A + C_2H_5]^{+*} + CH_4 \longrightarrow [A + C_2H_5^+] \quad (27)$$

$C_3H_5^+$ undergoes similar reactions to form $[M \pm H]^+$ and $[M + C_3H_5^+]$ ions. Adduct ions are formed from both $C_2H_5^+$ and $C_3H_5^+$ in small amounts and with compounds containing sites of unsaturation.

Consequently, the CH_4 CI mass spectra of organic compounds are expected to contain $[M \pm H]^+$ and fragment ions and not the $M^{+\bullet}$ ions that are observed in EI spectra. Although the fragment ions are likely to be similar in mass and structure for both EI and CI mass spectra, they are produced by entirely different reactions. Only for compounds with very low IPs, IP(A) < IP($C_2H_5^\bullet$), such as amines, are $M^{+\bullet}$ ions observed from charge exchange reactions of $C_2H_5^+$.

The CH_4 CI mass spectra of *n*-alkanes give abundant $[M - H]^+$ ions and approximately equal amounts of lower alkyl ions.[27] It was suggested that this product distribution resulted from random attack at the different methylene sites followed by localized decomposition. The CH_4 CI mass spectra of branched alkanes contain $[M - H]^+$ ions of lower abundance than those observed for *n*-alkanes, and fragment ions that identify the site of the branch. The relative abundances, of $[M - H]^+$ ions in the CH_4 CI mass spectra are much greater than the relative abundances of $M^{+\bullet}$ ions in the EI spectra and make molecular weight identification or carbon number determination of the compounds much easier.

ICR experiments showed that both CH_5^+ and $C_2H_5^+$ formed $[M - H]^+$ and fragment ions although the $[M - CH_3]^+$ ions were formed only by dissociative proton transfer from CH_5^+. There was generally good agreement between the ICR experiments at very low pressures and the high pressure experiments, but a few differences were observed which could be attributed to collisional stabilization of ions in the high pressures of the standard CI experiments.[28]

The CH_4 CI mass spectra of low molecular weight aliphatic esters are relatively similar.[29] The relative abundances of $[M + H]^+$ ions in these spectra are larger than the relative abundances of $M^{+\bullet}$ ions in their EI mass spectra. $[M - H]^+$ ions are also observed. Acyl ions, characteristic of the acid group, are generally formed. If the alkyl group in the ester is C_2 or higher, a rearrangement decomposition to give the protonated acid, $RCOOH_2^+$ is observed. Rearrangement decompositions are frequently observed in CI spectra, as well as simple cleavage reactions at the protonated site. In addition to $[M + H]^+$ ions, many of the esters give low abundance $[M + C_2H_5^+]$ and $[M + C_3H_5^+]$ ions. The presence of three ions at *m/z* values of X, X + 28, and X + 40 indicates MW = (X − 1). Alkyl fragment ions from the alcohol portion of the ester are generally observed for esters of the larger alcohols. The spectra of the esters are moderately complex but the major fragment ions can be explained in terms of simple decomposition and rearrangement reactions. The spectra appear to result from predominant attack at the carboxyl group.

Studies on the CH_4 CI mass spectra of methyl esters of long chain fatty acids indicate that $[M + H]^+$ ions are abundant peaks for the esters of the C_8–C_{22} acids.[30,31] The abundances of $[M + H]^+$ ions in the CH_4 CI spectra are much greater than the abundances of $M^{+\bullet}$ ions in the EI spectra of the corresponding methyl ester. The relative abundance of $[M - H]^+$ ions increases continually as the length of the acid chain increases. Increases were observed in the $[M - H]^+/[M + H]^+$ ratios with increasing acid chain length.[30,31] Similar observations were reported for the H_2 CI spectra of methyl esters of fatty acids.[31] These observations indicate that although attack is predominantly at the polar carboxyl group for the short chain compounds, hydride abstraction from the aliphatic chain becomes competitive when the polarizability of the hydrocarbon portion of the molecule is sufficiently large.

The CH_4 CI mass spectrum of decanol contains no $[MH]^+$ ions: H_2O is easily lost from MH^+ to form $C_{10}H_{21}^+$ as the base peak in the spectrum. Lower carbon number alkyl ($C_xH_{2x+1}^+$) and alkenyl ($C_xH_{2x-1}^+$) ions are also observed. Significant amounts of $[M - H]^+$ ions are also observed.[5] The two high mass ions, $[M - H]^+$ and R^+, differ by 16 mass units and allow easy assignment of the molecular weight. Internal hydrogen bonding was indicated in the CH_4 CI mass spectra of long chain α, w-diols by the presence of relatively abundant $[M + H]^+$ ions, even for the C_{46} diol.[32] The relative abundance of $[M - H]^+$ ions increased with increasing carbon number for these diols, and hydride abstraction apparently occurred randomly from the different methylene groups. These observations also indicate that attack occurs at both the polar substituent (the source of the dipole moment in the molecule) and the nonpolar, but polarizable chain.

Methane CI spectra of several classes of compounds are reviewed in Harrison's monograph on CIMS.[10]

3.2 Effects of Experimental Parameters

The CH_4 CI mass spectra obtained are essentially independent of pressure over a wide range of pressures if the methane reagent gas is sufficiently pure and the introduction system is sufficiently clean. However, small amounts of impurities (frequently water in the inlet lines for the reagent gas or solvents from the sample from a probe) can alter the reactant ions and hence the spectra. It is always advisable to examine the reactant ions during the course of CIMS experiments, particularly with highly reactive ions like CH_5^+ and $C_2H_5^+$ that will react with most polar impurities to form reactant ions that are weaker acids and will cause significantly less fragmentation. Since there are no standard compilations of CI spectra for comparison, differences in abundances of fragment ions caused by different reactant ions in what is supposed to be pure CH_4 will probably not cause significant problems. However, for quantitative studies, a change in reactant ions from $CH_5^+/C_2H_5^+$ to H_3O^+ or $CH_3OH_2^+$ (from impurities in a sample introduced through a probe) can cause a significant change in the relative abundance of MH^+ ions and greatly affect the accuracy of the experiments.[33]

Neither the electron energy nor the electron current has any significant effect on CH_4 CI spectra. This lack of effect is expected since the ions of the analyte are produced by reactions of ions that are themselves produced by ion/molecule reactions in methane. Any excitation energy in the initially formed CH_3^+ or $CH_4^{+\bullet}$ that might be carried over into the $C_2H_5^+$ or CH_5^+ ions would most probably be lost through collisions with methane before reacting with the analyte. Increasing the electron energy generally increases the total ion current and CI sensitivity. The higher energy electrons have a greater penetration depth into the source of the mass spectrometer and consequently increase the concentration of ions in the volume from which the ions most efficiently leave the source. In addition more ions can be produced from 250 eV electrons than from the traditional 70 eV electrons used in EI experiments.

Some (but not all) mass spectrometers (mostly magnetic instruments) have a repeller field in the source to improve the efficiency of the ion extraction. Many of the newer instruments have no added fields in the CI source. The CH_4 CI spectra are essentially independent of repeller field within the source as long as the voltage is of the order of 10 V or less. Very high repeller fields can cause decomposition of the ions within a high-pressure source.[34] Using a modified source with grids to maintain an electric field gradient within the source, one can decompose ions within the source to obtain high-pressure collisional activation (or decomposition) of sample ions.[35]

The CH_4 (and other) CI spectra can depend on sample size. If the sample size is sufficiently small, then the only ions that are produced are those from reaction of the reactant ions with the analyte. If the sample size is sufficiently large (not always easy to establish), sample ion/sample molecule reactions will occur and these reactions may alter the spectra significantly. Sample ion/sample molecule reactions will generally not be noticed even if they occur because systematic observations of spectra with changing concentrations are seldom made. Clear evidence of sample ion/sample molecule reactions is shown by the presence in CI mass spectra of protonated dimer ions, MHM^+, at large sample concentrations. These ions are often observed from probe samples where it is impossible to control the sample pressure. These ions will generally create no difficulties in structure or molecular weight determination, as long as one recognizes their occurrence. Examination of CI spectra across a chromatographic peak or during heating of a probe sample can show that sample ion/sample molecule reactions are occurring. CH_4 CI spectra of polyethylene glycol (PEG) oligomers and PEG ethers obtained by gas chromatography/chemical ionization mass spectrometry (GC/CIMS) showed that the $[M + H]^+$ ions were strongly sample-size dependent and suggested that sample ion/sample molecule reactions might be a common occurrence in CIMS.[36]

CH_4 CI spectra are temperature dependent.[29,37] In general, the higher mass ions, $[M \pm H]^+$ decrease in relative abundance as the temperature is increased. One would expect such a phenomenon since the internal energies of the ions and, therefore, their rates of decomposition increase with increasing temperature. If consecutive decompositions occur, the abundances of the intermediate ions may pass through a maximum as the temperature is increased. The stable lower-mass ions almost always increase in relative abundance as the temperature is increased.

The total ion current and sample ion current in CH_4 CI (and with all reagent gases) will pass through maxima as the pressure is increased.[38] The pressure corresponding to the maximum in sensitivity will vary among instruments since it depends on the source geometry and the pumping capacity of the system. The reaction time and the extent of ion/molecule reactions increase with increasing pressure; therefore, the analyte ions will increase in relative abundance with increasing pressure. However, there are losses from ion scattering in the region immediately outside the source as well as in the mass analyzer region that also increase with increasing pressure. The result of these two competing processes is a maximum in sensitivity with increasing pressure. The maximum in the sensitivity does not, however, correspond to conditions for the greatest extent of ion/molecule reactions.

4 OTHER CHEMICAL IONIZATION REAGENT GASES

4.1 Isobutane

As mentioned in section 2.1, one of the major uses of CIMS is the production of simplified spectra of analytes, for qualitative analysis and for quantitative analysis of mixtures. The CH_4 CI mass spectra frequently contain many fragment ions because CH_5^+ and $C_2H_5^+$ are strong Brønsted acids, and proton transfer to the analyte, A, produces excited AH^{+*} ions which decompose. Use of lower-energy (more weakly acidic) reactant ions will produce spectra with less fragmentation. A commonly used weakly acidic system is i-C_4H_{10}.

The dominant ion produced by EI of i-C_4H_{10} is the propyl ion, $C_3H_7^+$, at m/z 43. The propyl ions react rapidly with i-C_4H_{10} to form t-$C_4H_9^+$ at m/z 57 (Equations 28 and 29). The t-$C_4H_9^+$ ion is a much weaker acid than CH_5^+ or $C_2H_5^+$ because the proton affinity of i-C_4H_8 is significantly higher than the proton affinity of CH_4 or C_2H_4: 802.1 kJ/mol, compared with 543.5 kJ/mol and 680.5 kJ/mol.[39]

$$i\text{-}C_4H_{10} + e^- \longrightarrow s\text{-}C_3H_7^+ + CH_3^\bullet + 2e^- \quad (28)$$

$$s\text{-}C_3H_7^+ + i\text{-}C_4H_{10} \longrightarrow t\text{-}C_4H_9^+ + C_3H_8 \quad (29)$$

Because the t-butyl ion is a much weaker protonating agent (Brønsted acid) than CH_5^+ or $C_2H_5^+$, less energy is transferred in proton transfer to form AH^{+*} and there is much less fragmentation in i-C_4H_{10} CI spectra than in CH_4 CI spectra.[10,40] The t-butyl ion is also a weaker Lewis acid than $C_2H_5^+$.[39] Adduct ions, $[M + C_4H_9]^+$, are observed in the i-C_4H_{10} CI spectra of certain compounds.

Most i-C_4H_{10} CI spectra of complex organic molecules contain abundant $[M + H]^+$ ions, often with very little or no fragmentation. Consequently, i-C_4H_{10} CI has been used to identify compounds in mixtures from sets of likely possibilities. The i-C_4H_{10} CI spectra of many barbiturates and other commonly used drugs were sufficiently different and easily recognizable that these drugs could be identified in complex mixtures by vaporizing the samples directly into the source of a mass spectrometer (without the time of a chromatographic separation), even from gastric contents in drug overdose cases.[41,42] Quantitative analyses of mixtures were also achieved with i-C_4H_{10} CI by vaporizing the samples directly into the source of the mass spectrometer. Calibration curves and an internal standard were needed.[43]

Because both fragment ions and MH^+ ions are observed in reasonable abundances, i-C_4H_{10} CI has been successfully used as a stereochemical probe for cyclic diols and their derivatives.[44,45] The i-C_4H_{10} CI spectra of isomers of adamantanediol dimethyl ethers and dihydroxyprotoadamantanes for which internal hydrogen bonding was possible in the protonated molecules, MH^+, contained MH^+ ions as the most abundant peak. Internal hydrogen bonding was possible for all three isomers of the 2,4-adamantanediol diacetates, and the i-C_4H_{10} CI spectra were virtually indistinguishable; therefore the nature of the derivative is critically important in determining the stereochemistry of ring systems.[45]

The i-C_4H_{10} CI spectra are sensitive to the pressure of the i-C_4H_{10} reagent gas and, consequently, the spectra will not be highly reproducible from one instrument to another in the extent of fragmentation because the pressures are seldom measured in most commercial instruments.[30,36] If any pressures are reported for CI reagent gases, the values are often the ionization gauge pressures well removed from the ionization source itself. Although the source pressure and ionization gauge pressure are proportional to each other, the ionization gauge pressure does not provide a reproducible measure of source pressure. In general, the extent of fragmentation decreases with increasing pressure of the i-C_4H_{10} reagent gas.

The pressure dependence of i-C_4H_{10} CI spectra results partly from the collisional stabilization of the (probably slightly) excited analyte ions, AH^{+*}, produced by proton transfer from t-$C_4H_9^+$. However, the major cause of the pressure dependence of the spectra is the change in the composition of the reactant ions with increasing pressure. The s-propyl ion, $C_3H_7^+$, is a stronger Brønsted acid and stronger Lewis acid than t-$C_4H_9^+$; consequently, the analyte ions produced by reactions from $C_3H_7^+$ will contain more internal energy and dissociate more extensively than analyte ions produced by reactions from t-$C_4H_9^+$.

As noted above in Equations (28) and (29), the propyl ions are the major ion in the EI mass spectrum of i-C_4H_{10}, and these propyl ions (m/z 43) react to form t-butyl ions (m/z 57) as the pressure is increased. The ratio, I(57)/I(43) is a sensitive increasing function of i-C_4H_{10} pressure and can be used as a method of obtaining reproducible, even though not accurately known, pressures. A method has been proposed for the calculation of source pressures under CI conditions from the ionic distributions.[46] Realistic estimates of pressures in the CI source or the reactant ion spectrum are seldom given with the presentation of CI spectra.

4.2 Ammonia

Another commonly used proton transfer reagent gas system is NH_3. Ammonia is a stronger base than i-C_4H_8 and, consequently, NH_4^+ should be an even more selective reactant ion than t-$C_4H_9^+$.[39] The reactant ion spectrum in NH_3 is a complex function of temperature, pressure, and reaction time. The dominant ions under normal CI conditions are NH_4^+ at m/z 18 and $NH_4^+ \cdot NH_3$

at m/z 35 with lesser amounts of the more highly solvated ammonium ions.[47,48] Because the reactant ion distribution is very sensitive to the experimental conditions of temperature and pressure, the NH_3 CI spectra are also sensitive to these conditions.

Early work on NH_3 CIMS showed the expected selectivity.[48–50] Very basic compounds, such as amines, amides, and α,β-unsaturated ketones, gave $[M + H]^+$ and their solvated $[M + NH_4^+]$ or $[M + H + NH_3]^+$ ions. Less basic compounds such as ketones, esters, ethers, anhydrides, and polyhydroxy compounds reacted by adduct formation to give $[M + NH_4]^+$ and lesser amounts of $[M + NH_4 + NH_3]^+$ ions. Alkanes, alkenes, aromatic hydrocarbons, and alcohols gave no sample ions with small sample sizes. Substitution ions were also observed, $[M + NH_4 - H_2O]^+$ that have the same nominal mass as $M^{+\bullet}$.

The mechanism for the formation of sample ions is complex since several reactions are possible, depending on the PA of the sample (Equations 30–36):

$$NH_4^+ + A \longrightarrow AH^+ + NH_3 \quad (30)$$

$$AH^+ + NH_3 \Longleftrightarrow ANH_4^+ \quad (31)$$

$$NH_4^+ + A \Longleftrightarrow ANH_4^+ \quad (32)$$

$$ANH_4^+ + NH_3 \Longleftrightarrow ANH_4^+ \cdot NH_3 \quad (33)$$

$$NH_4^+ \cdot NH_3 + A \longrightarrow AH^+ + 2NH_3 \quad (34)$$

$$NH_4^+ \cdot NH_3 + A \longrightarrow ANH_4^+ + NH_3 \quad (35)$$

$$NH_4^+ + XOH \longrightarrow [X + NH_4 - H_2O]^+ \quad (36)$$

Equations (30) and (31) represent proton transfer to a basic compound, PA(A) > PA(NH_3), followed by solvation, and possibly by additional solvation, Equation (33), at high pressures and low temperatures. This reaction sequence will give ions at masses $X + 1$, $X + 18$ and $X + 35$ in ratios that depend on the pressure of NH_3 and the temperature. Equation (32) represents equilibrium (or at least a reversible addition) of NH_4^+ with a compound that is less basic than ammonia and cannot accept a proton from direct reaction. This adduct ion can also solvate further by Equation (33). Equations (34) and (35) show the competing reactions of the solvated ammonium ion with the analyte: Equation (34) will occur with very basic compounds and Equation (35) will occur with less basic compounds. Equation (36) is observed with certain hydroxy-containing compounds. Despite these complications, NH_3 CI spectra are frequently used and an extensive review is available.[48]

It was reported many years ago that useful spectra could be obtained from CH_4/NH_3 mixtures containing only small amounts of NH_3(~1–3%).[51] In these mixtures, the dominant reactant ion was NH_4^+ and there was no $NH_4^+ \cdot NH_3$. Consequently, the reactions were much simpler. By altering the concentrations of NH_3 and CH_4 in mixture one can obtain spectra with essentially no fragmentation, or spectra with reasonable fragmentation if CH_5^+ and $C_2H_5^+$ ions are present as reactant ions as well as NH_4^+.

With CH_4/NH_3 mixtures containing a low concentration of NH_3, one can observe $[M + NH_4]^+$ ions for compounds which cannot be detected using pure NH_3 as the reagent gas. It was shown from high pressure and ion cyclotron resonance/mass spectrometry (ICR/MS) experiments that an additional reaction involving the $[M + NH_4]^+$ adducts occurred to remove sample containing ions from the spectrum and therefore reduce the sensitivities of polar, but not very basic compounds.[52,53]

$$[M + NH_4]^+ + NH_3 \longrightarrow NH_4^+ \cdot NH_3 + M \quad (37)$$

In order for the reaction (Equation 37) to be fast, it must be exothermic, and will be observed, therefore, with those compounds that are less basic than ammonia and solvate the ammonium ion less strongly than does NH_3, $D(M{-}NH_4)^+ < D(NH_3{-}NH_4^+)$.

Experiments have been done using a mixture of CH_4 with 5% of a 1/1 $^{15}NH_3/^{14}NH_3$ mixture to aid in identifying the composition of the sample ions, since ions at X and X + 17 can be $[M + H]^+$ and $[M + NH_4]^+$ or $[M + NH_4]^+$ and $[M + NH_4 + NH_3]^+$.[54] At the pressures of this mixture used in these experiments, the reactant ion spectrum consisted of essentially equal abundances of $^{14}NH_4^+$ and $^{15}NH_4^+$ at m/z 18 and 19, with small amounts of the CH_5^+ and $C_2H_5^+$ ions as well. A sample ion containing NH_4^+ appeared as two peaks of approximately the same intensity. The spectrum of hexyl acetate contained a low abundance ion at m/z 145 and two ions of essentially equal intensity at m/z 162 and 163. Obviously, the species are $[M + H]^+$ at m/z 145 from reactions of $CH_5^+/C_2H_5^+$ and $[M + NH_4]^+$ at m/z 162/163.

One of the first uses of NH_3 in CIMS was for the differentiation among primary, secondary, and tertiary amines from a comparison of the spectra obtained with NH_3 and with ND_3.[55] In these experiments, all of the active hydrogens of the amine are exchanged with the ND_3; therefore, there is a distinct mass shift in the spectra depending on the number of labile hydrogens in the molecule. The NH_3 CI spectra of 2,6-dimethylaniline and the isomeric *N*-ethyl aniline each contained two peaks at m/z 122 and 139, $[M + H]^+$ and $[M + H + NH_3]^+$. The ND_3 CI spectra of 2,6-dimethylaniline contained two peaks at m/z 125 $(CH_3)_2C_6H_3ND_3^+$ and 145, $(CH_3)_2C_6H_3ND_3^+.ND_3$. However, the ND_3 CI spectrum of *N*-ethyl aniline contained two peaks at m/z 124 and 144 because there was one less exchangeable hydrogen

in the molecule. The use of labeled CI reagent gases (ND_3, CH_3OD, D_2O) to determine the number of active hydrogens in molecules in biologically important molecules is well established.[56]

4.3 Tetramethylsilane

Many different gases and liquids have been used as CI reagent gases, although methane, isobutane, and ammonia are the most common. Another reagent that has been used is tetramethylsilane, $(CH_3)_4Si$, the silicon analog of neo-pentane. Although this reagent gas has not been routinely used for analysis, there are some interesting chemical reactions that have been observed with this reagent gas and the trimethylsilyl ion is analogous to the *t*-butyl ion. Tetramethylsilane is a liquid, but its vapor pressure is sufficiently high that it can be used with commercial gas introduction systems. It is advisable to keep the liquid tetramethylsilane in a container of water to reduce the pressure variations in the source because of temperature variations in the room.

The dominant ion produced by EI of $(CH_3)_4Si$ is the trimethylsilyl ion, $(CH_3)_3Si^+$, the silicon analog of $(CH_3)_3C^+$, the *t*-butyl ion. Since $(CH_3)_2Si{=}CH_2$ is not a stable species, its heat of formation and proton affinity have been determined by ion/molecule reactions rather than conventional thermochemistry.[39] The ionic chemistry of $(CH_3)_4Si$ is very different from that of *i*-C_4H_{10} or of *i*-C_5H_{12} and, by inference of $(CH_3)_4C$.[57,58] The *t*-butyl ion does not add to saturated hydrocarbons, but the trimethylsilyl ion does add to tetramethylsilane to form $[Si_2(CH_3)_7]^+$ as a major product ion. There are other, less abundant product ions in high pressures of tetramethylsilane, including an ion at *m/z* 91, $(CH_3)_3SiOH_2^+$, formed from reactions with impurities of water in the inlet system. Because $(CH_3)_3Si^+$ is both a weak Brønsted and a weak Lewis acid, the major reaction with polar compounds is adduct formation to give $[M + (CH_3)_3Si]^+$ ions and there are only slow reactions with hydrocarbons.[59–61] The ever-present impurity ion, $(CH_3)_3SiOH_2^+$ at *m/z* 91, reacts by proton transfer to give $[M + H]^+$ ions or by displacement to give $[M + (CH_3)_3Si]^+$ ions.

Tetramethylsilane has been used as a CI reagent either as a pure substance or diluted with methane, nitrogen, or isobutane. Dilution with methane reduced problems associated with source contamination by the reagent gas, increased the sensitivity for samples, and simplified the reagent ion spectrum by reducing the abundance of $(CH_3)_7Si_2^+$ ions.[60,61]

In experiments derived from the technique used in the mass spectral analysis of biomedical samples (using a mixture of labeled and unlabeled samples), a 1/1 mixture of $(CH_3)_4Si$ with $(CD_3)_4Si$ was used to give equal abundances of $[M + (CH_3)_3Si]^+$ and $[M + (CD_3)_3Si]^+$ ions for the compounds. The adduct ions could easily be identified by the two peaks of equal abundance separated by 9 mass units.[61]

Tetramethylsilane CIMS can be used to characterize alcohols, since adduct ions are formed. Early ICR experiments showed both $[M + (CH_3)_3Si]^+$ and $(CH_3)_3SiOH_2^+$ ions are products from reactions of $(CH_3)_3Si^+$ with ethyl and higher alcohols having β-hydrogens.[62] The early high-pressure CI experiments showed only adduct ions.[59] From examination of the tetramethylsilane spectra of alcohols across the chromatographic peaks in GC/CIMS experiments, it was observed that the ratio, $I\{[M + 73]^+\}/I\{[CH_3)_3SIOH_2]^+\}$, went through a maximum across each chromatographic peak.[63] These observations and ICR experiments showed that the reaction was the two-step sequence (Equations 38a, 38b and 39):

$$(CH_3)_3Si^+ + ROH \longrightarrow R^+ + (CH_3)_3SiOH \quad (38a)$$

$$(CH_3)_3Si^+ + ROH \longrightarrow (CH_3)_3SiOH_2^+ + [R{-}H] \quad (38b)$$

$$(CH_3)_3SiOH_2^+ + ROH \longrightarrow (CH_3)_3SiOHR^+ + H_2O \quad (39)$$

Primary, secondary, and tertiary alcohols could easily be distinguished. The relative sensitivities for the alcohols were essentially independent of molecular weight and temperature (120–260 °C).

Relative molar sensitivities in GC/CIMS using tetramethylsilane as the reagent gas could be described reasonably well from theoretical models for rate constants of ion/molecule reactions.[64] Significant differences were observed in the temperature dependence of several oxygenated compounds. For compounds for which sample ions were formed by the two-step mechanism, Equations (38) and (39), there was essentially no change in sensitivity with increasing temperature. However, there was a significant decrease in sensitivity with increasing temperature for those compounds for which the $[M + (CH_3)_3Si]^+$ ions were formed by collisional stabilization of the adduct ion.[64]

The selectivity of $(CH_3)_3Si^+$ ions for the identification and quantitative analysis of polar compounds in complex mixtures was demonstrated by an analysis of oxygenated compounds in gasoline.[65] The oxygenated compounds could be unambiguously identified and quantified from selected ion traces in GC/CIMS experiments. Relative standard deviations for short term replications were ~5%.

The association reactions and ionic chemistry of $(CH_3)_3Si^+$ have been reviewed recently.[66]

4.4 Other Reagent Gases

Many other compounds have been used as reagent gases for CIMS studies, although few have been extensively used.[67–70]

5 CHARGE EXCHANGE REAGENTS

5.1 Introduction

In the original work (and the patent) on CIMS, all reactions involved proton or hydride transfer or adduct formation. However, subsequently electron transfer or charge exchange reactions became considered as part of CIMS, (Equation 40):

$$R^{+\bullet} + A \longrightarrow A^{+\bullet} + R \qquad (40)$$

Using the fundamental postulate of ion/molecule reactions, that only exothermic reactions will be fast, one can calculate the thermochemistry of this electron or charge transfer reaction from the IPs (energies) of R and A, (Equation 41):

$$\Delta H^{\circ}(40) = \mathrm{IP(A)} - \mathrm{IP(R)} \qquad (41)$$

Consequently, if the IP of the analyte, A, is less than the IP of the reactant, R, then the charge exchange reaction will be exothermic and, in all probability, fast.

In addition, the greater the exothermicity of the reaction, the greater the internal energy in the $A^{+\bullet *}$ ion and the greater the extent of dissociation or fragmentation observed in the charge exchange CI spectrum. Consequently, if one wishes to detect all compounds in a mixture, one will choose a reagent gas that has a high IP (one of the rare gases) and it is likely that there will be abundant fragment ions in the spectra. Conversely, if one wishes selective detection or very little fragmentation, one would choose a reagent gas with a low IP, but with an IP that is larger than the IP of the sample to be analyzed.

There is an added complication that many compounds (including practically all polar compounds) will react with themselves to form other ions (generally protonated species). Consequently, the choice of reagents for charge exchange is more limited than the choice for proton or hydride transfer in high-pressure mass spectrometry. In addition, the distribution of the reagent ions is sensitive to impurities, particularly water. Consequently, care must be taken to keep the gas inlet lines scrupulously dry. It is probably realistic that the ^{13}C isotope ratios will be unreliable because of the formation of MH^+ ions from hydrogen-containing ions among the reactant ions.[71]

The mechanisms for the formation of fragment ions in charge exchange CIMS are very different from those in proton or hydride transfer since the decomposing species are very different: $A^{+\bullet}$ decomposes after charge exchange and AH^+ decomposes after proton transfer. The decomposition mechanisms after charge exchange are essentially the same as those after EI. There is a major difference between charge exchange and EI, however. In charge exchange, the $A^{+\bullet}$ ions have a well-defined internal energy: IP(R) − IP(A). In EI, the $A^{+\bullet}$ ions have a distribution of energies and decompose at different rates.

Charge exchange reactions are the initial ion/molecule reactions involved in atmospheric pressure ionization (API) or atmospheric pressure chemical ionization (APCI). The $N_2^{+\bullet}$ ions (produced by direct ionization) react rapidly with O_2 because $IP(N_2) > IP(O_2)$ and $O_2^{+\bullet}$ ions (formed by direct ionization and charge exchange) react with samples and or impurities in the gas to produce other ions which may be sample ions or which may react with sample ions. $O_2^{+\bullet}$ charge exchange has been shown to be a very sensitive technique for the detection of impurities in air.[72]

5.2 Helium

On the basis of the discussion above, He would be the worst possible charge exchange reagent gas because its IP (24.6 eV) is much greater than the IPs of practically all compounds of interest (~8–10 eV).[73] He is the most common carrier gas in gas chromatography (GC) and is obviously a reagent gas to try, although there have been few experiments using He for the reasons given above.

GC/CIMS experiments using He as the carrier gas were used to detect SF_6 at the ppb level in air.[74] The mass spectrum of SF_6 obtained under these conditions was surprisingly similar to the EI spectrum. Neither the EI spectrum nor the ''He charge exchange'' spectrum contained $M^{+\bullet}$ ions at m/z 146. Both spectra contained $(M - F)^+$ as the base peak. The ''He charge exchange'' spectrum contained 108 as 11% of the base peak and 89 at 68% of the base peak. The EI spectrum shows 108 at 9–11% of the base peak and 89 as 20–26% of the base peak.[75] Although this spectrum shows somewhat more fragmentation, the extent of fragmentation is not nearly as large as one would expect for the amount of internal energy in a simple molecule: (24.6–15.3) eV.[73]

It is likely that processes other than charge exchange by $He^{+\bullet}$ produce the spectrum, or at least make significant contributions to it. Some collisional stabilization of the highly excited $SF_6^{+\bullet}$ occurs that will reduce the extent of fragmentation. In addition, the initial high-energy electrons lose energy in collisions and produce more electrons that may be able to ionize the sample molecules. Since the ionization cross-section of He is much smaller than ionization cross-sections of other molecules, direct

EI may well be occurring.[76] Experiments in our laboratory showed spectra of compounds from ''He charge exchange'' that contained more fragmentation than the EI spectra, but which were sufficiently like the EI spectra to be recognizable. Consequently, ''He charge exchange'' CIMS may be useful.

5.3 Benzene, Chlorobenzene

The more common use of charge exchange reagents is for selective detection of certain classes of compounds and not others. One obvious difference among classes of compounds is the difference in IPs between saturated and unsaturated compounds. The IPs of unsaturated compounds are consistently lower than the IPs of similar saturated compounds.[73]

Chlorobenzene, (C_6H_5Cl, IP = 9.04 eV) was found to be a selective reagent gas for the detection of aromatic hydrocarbons in the presence of alkanes and cycloalkanes.[77] The $C_6H_5Cl^{+\bullet}$ ion is the dominant species among the reactant ions at ''high'' pressures (~0.04 torr) and essentially the only sample ions observed for chlorobenzene CI of alkylbenzenes and alkylnaphthalenes are the molecular ions. In these experiments there was an apparent decrease in the relative sensitivities of the aromatic hydrocarbons with increasing molecular weight that was attributed to the electron multiplier. To a first approximation, the sensitivities of the aromatic hydrocarbons were essentially independent of molecular weight. Complex mixtures of gasoline or fuel oils could be quantitatively analyzed using chlorobenzene CI without chromatographic separation.

Benzene (C_6H_6, IP = 9.25 eV) has also been used for selective detection in GC/CIMS studies. By recording the ion current of the $C_6H_6^{+\bullet}$ during a GC/CIMS experiment (Reactant Ion Monitoring), one could see that $C_6H_6^{+\bullet}$ ions reacted with unsaturated compounds in mixtures but not with saturated compounds.[78,79] Again, benzene CI mass spectra contain essentially no fragment ions. Quantitative studies using benzene GC/CIMS showed that the relative molar sensitivities of a series of alkylbenzenes (C_8–C_{12}) and olefins (C_5–C_9) were essentially constant and the relative molar sensitivities of alkanes (C_5–C_{12}) were less than 1% of these values.[80] Calculations of rate constants (molar sensitivities) from theories of ion/molecule reaction rate constants predict only a small variation in the relative rate constants for charge exchange of $C_6H_6^{+\bullet}$ with alkylbenzenes of increasing molecular weight.[10] Similar conclusions would be noted for charge exchange from $C_6H_5Cl^{+\bullet}$.

Benzene or chlorobenzene charge exchange has an advantage over low-voltage EI, the commonly used technique for mass spectrometric analyses of aliphatic/aromatic hydrocarbon mixtures. The relative molar sensitivities of substituted alkylbenzenes are independent of structure for benzene CIMS and very strongly dependent on structure for low-voltage EIMS.[81] Benzene or chlorobenzene charge exchange CIMS may be more sensitive than low-voltage EIMS, but definitive experiments have not been performed.

5.4 Nitric Oxide

Although the ion/molecule reactions with NO as a CI reagent gas are moderately complex, nitric oxide has been used extensively as a CI reagent gas.[10,67] The major ionic reactions which can occur with NO^+ are the following (Equations 42–44):

$$NO^+ + M \longrightarrow M^{+\bullet} + NO^{\bullet} \quad (42)$$

$$NO^+ + M \longrightarrow [M + NO]^+ \quad (43)$$

$$NO^+ + M \longrightarrow [M - H]^+ + HNO \quad (44)$$

In addition X^- abstraction (X = HO, Cl) reactions have been observed.[10,50,67]

Esters and ketones give $[M + NO]^+$ ions, as long as IP(M) > IP(NO); aldehydes can be differentiated from ketones because aldehydes give abundant $[M - H]^+$ ions rather than $[M + NO]^+$ adduct ions; alkanes give essentially only $[M - H]^+$ ions.[50] The $[M + NO]^+/[M^+]$ ratio for a series of substituted benzenes gave a good correlation with the Hammett substituent constant, $\sigma_p{}^+$.[82] Analysis of the data for many substituted benzenes showed that NO^+ has an ''effective IP'' of ~8.7 eV for charge exchange reactions, rather than the observed value of 9.3 eV.[10]

The mass spectrometric analysis of alcohols has traditionally been difficult. EI, CH_4 CI, and i-C_4H_{10} CI have been used with some success and tetramethylsilane CI can also differentiate among primary, secondary, and tertiary structures. The CI techniques are improvements over the EI technique since some sample ions containing the correct number of C-atoms are observed as noted in sections 3.1 and 4.1. The NO CI spectra of primary, secondary, and tertiary alcohols were relatively easily distinguishable.[83] Primary alcohols generally gave NO CI spectra which contained $[M + NO]^+$, $[M - H]^+$, and some $[M - 3]^+$ ions; secondary alcohols gave NO CI spectra which contained abundant $[M - H]^+$ and $[M - OH]^+$ ions; and tertiary alcohols gave predominantly $[M - OH]^+$ ions. The $[M - 3]^+$, and $[M - 2 + NO]^+$ ions were attributed to oxidation of ions by NO, since the abundance of these ions vanished with the use of dilute mixtures of NO in N_2. Alkanes also gave ions from ionic reactions of sample ions with the reagent gas, NO.

The presence of the $[M - 3]^+$ and $[M - 2 + NO]^+$ ions and other unexpected ions show one of the difficulties of working with NO as a reagent gas: oxidation of ions. In

addition, NO also rapidly oxidized the hot wire filament used to generate electrons for ionization of the reagent gas. One solution to the problem of rapid oxidation of the filament that produces the ionizing electrons was to use a discharge source that does not have a filament.[84] Although the NO CI spectra of hydrocarbons are not single-species spectra, the extent of fragmentation is relatively small so that there is little overlap in the masses of spectra of the different classes of compounds of approximately the same molecular weight. Complex mixtures of hydrocarbons (C_6–C_{20}) were analyzed using GC/CIMS with NO as the reactant gas. Complete chromatographic separation was not obtained because of the complexity of the mixtures, but type analyses by carbon number were obtained from the NO CI mass spectra.[85]

Another way to increase the lifetime of the filament and also to reduce the oxidation reactions by NO is to use dilute mixtures of NO in other reagent gases – N_2 and H_2 were successfully used.[86,87] The exact composition of the mixtures was determined by the source pressures which were used, but these reagent mixtures allowed the reactions of NO^+ to be observed without the oxidation processes.

Another solution to the oxidation problems with pure NO was to use N_2O as the reagent gas.[88] Nitrous oxide does not oxidize the hot filaments rapidly. The high-pressure mass spectrum of N_2O contains about 33% NO^+ ions and 63% $N_2O^{+\bullet}$ ions, independent of the pressure of nitrous oxide. The $N_2O^{+\bullet}$ ions react predominantly by dissociative charge exchange ($IP(N_2O) = 12.9$ eV) to give characteristic fragment ions and the NO^+ ions undergo the reactions described above. No oxidation products like $[M-3]^+$ ions were observed; but the extent of fragmentation was much higher using this reagent gas than using pure NO.

6 NEGATIVE ION CHEMICAL IONIZATION

6.1 Introduction

Negative ions were observed in the very early gaseous discharge experiments (H^-, $O^{-\bullet}$, Cl^-, $C^{-\bullet}$, $C_2^{-\bullet}$) in abundances that were sometimes comparable to the positive ions.[13] However extensive work and analytical applications of negative ion mass spectrometry are relatively recent. Historically, the mass spectrometers (both research and commercial analytical) were designed for the detection of positive ions. At the very low pressures of most mass spectrometric experiments, the efficiency for the production of positive ions is some 100–1000 times as large as the efficiency for production of negative ions.[89] Electron capture of 70 eV electrons is very inefficient. In addition, the negative ion spectra that were obtained contained predominantly low-mass ions that were not helpful in identification of compounds.

The dominant reactions for production of negative ions from collisions with electrons are given in Equations (45–47):

$$AB + e^- \longrightarrow AB^{-\bullet} \quad (45)$$

$$AB + e^- \longrightarrow B^- + A \quad (46)$$

$$AB + e^- \longrightarrow B^- + A^+ + e^- \quad (47)$$

Equation (45) is electron capture and is efficient only if AB has a high electron affinity (EA) and if the kinetic energy of the electrons is very low, approximately thermal energy. Equation (46), dissociative electron capture, occurs with electrons with kinetic energies of a few electron volts and will produce negative ions for atoms or small groups with large EA. Equation (47), pair production, occurs with high-energy electrons (greater than the IPs of the molecules) and again produces negative ions of atoms or small groups with high electron affinities.

Negative CIMS should probably be divided into two general categories, one in which the negative ions of the samples are produced by ion/molecule reactions and the other in which the negative ions of the samples are produced by collisionally stabilized electron capture. The mechanisms for the formation of the negative ions are different for the two processes, but both processes may occur under the same conditions and such a distinction is not always made.

6.2 Negative Ion/Molecule Reactions

The rate constants for ion/molecule reactions of many negative ions are essentially the same as the rate constants for positive ions, $\sim 10^{-9} cm^3\, mol^{-1}\, s^{-1}$, as long as the reactions are slightly exothermic.[11] The reactions of the negative reactant ions should, therefore, be analogous to the reactions of positive reactant ions in CIMS. The reactions of the negative ions are frequently as in Equations (48–50):

$$R^- + SH \longrightarrow S^- + RH \quad (48)$$

$$R^- + SH \longrightarrow SH^- + R \quad (49)$$

$$R^- + SH \longrightarrow SHR^- \quad (50)$$

The thermochemistry of Equation (48) is determined by the relative base strengths of R^- and S^-. If R^- is a stronger base than S^-, the reaction is exothermic and likely to be fast. Comparisons can be made of the gas-phase acidities of the anions from a standard compilation to predict whether or not a particular reaction will occur.[90] The

thermochemistry of Equation (49) is determined by the difference in EAs of R and SH. If EA(SH) > EA(R), the reaction will be exothermic and fast. Adduct formation, Equation (50), is likely only if Equations (48) and (49) are both endothermic and the species is polar.

A frequently used negatively charged CI reactant ion is HO^-. It is necessary to produce HO^- ions by indirect means, and mixtures of N_2O with H_2, CH_4, or alkanes have been successful for generating ionic distributions that contain mostly HO^-.[91,92] The compositions of the mixtures depend on the pressures, but conditions can be achieved so that the reactant ion spectrum contains predominantly HO^- ions. The process for producing the HO^- ions (Equations 51 and 52) involves the formation of $O^{-\bullet}$ from N_2O by low energy electrons produced from the high pressures of the reagent gas and subsequent abstraction of $H^\bullet$ by $O^{-\bullet}$.

$$e^- + N_2O \longrightarrow O^{-\bullet} + N_2 \quad (51)$$

$$O^{-\bullet} + XH \longrightarrow HO^- + X^\bullet \quad (52)$$

The dominant reaction of HO^- with acids, esters, ketones, alcohols, thiols, alkenes, alkynes, alkylbenzenes, and alkylnaphthalenes is proton transfer to give primarily $[M - H]^-$ ions. Some fragment ions are produced and some oxidation products, $[M - 3]^-$, are observed, but the spectra of many compounds are simple. Since proton transfer from alkanes and cycloalkanes to HO^- is not observed, it was possible to analyze complex petroleum mixtures for aromatic compounds without separation from the alkanes and cycloalkanes.[92] There appeared to be a variation of a factor of 2–3 among the relative sensitivities of alkyl aromatic hydrocarbons. At high pressures, there are additional ions in the spectra of alkylbenzenes and olefins that result from reactions of the $[M - H]^-$ ions with N_2O.[91]

The HO^- CI spectra of a series of steroids and cholesteryl esters proved useful for identification.[93] There were abundant $[M - H]^-$ ions in these spectra which allowed easy determination of the molecular weights. The acid anion is a major fragment ion from nucleophilic displacement, (Equation 53):

$$HO^- + RCOOR' \longrightarrow RCOO^- + HOR' \quad (53)$$

This reaction is not the same as the nucleophilic reaction in solution that cleaves the acyl oxygen bond to give RO^-.

The HO^- CI spectra of the cholesteryl esters may be compared with the *i*-C_4H_{10} CI (*t*-$C_4H_9^+$) spectra discussed earlier.[7] The significant abundance of $[M - H]^-$ ions from reactions with HO^- and the absence of $[M + H]^+$ ions from reactions with *t*-$C_4H_9^+$ suggest that proton abstraction from the sample in negative ion CI imparts less energy to the sample ions than proton transfer to the sample in positive ion CI. The CI spectra of a series of terpene alcohols gave mostly $[M - OH]^+$ ions from reactions with *t*-$C_4H_9^+$ and predominantly $[M - H]^-$ ions from reactions with HO^-.[94] However, in this case it is likely that the $[M - OH]^+$ ions are formed by HO^- abstraction by *t*-$C_4H_9^+$, rather than dissociative proton transfer. The *i*-C_4H_{10} CI spectra of a series of terpene esters contained mostly alkyl ions and essentially no $[M + H]^+$ ions while the HO^- spectra contained significant amounts of $[M - H]^-$ ions, although extensive displacement to give $RCOO^-$ ions was observed. The HO^- spectra of methyl esters of long chain fatty acids contain predominantly $[M - H]^-$ ions with smaller amounts of $RCOO^-$ and $[M-H-CH_3OH]^-$ fragment ions.[95] The *i*-C_4H_{10} CI spectra of methyl esters of long chain fatty acids contain $[M + H]^+$ as the base peak with increasing amounts of $[M - H]^+$ ions as the chain length increases.[30] Neither reactant ion has a clear cut advantage over the other.

In quadrupole (but not magnetic) mass spectrometers, ions of identical m/z but opposite charge pass through the quadrupole field and exit at the same spot. Consequently, by rapidly alternating the potential on the source of the quadrupole and having one detector for positive ions and another for negative ions, one can detect both positive and negative ions essentially simultaneously under CI conditions.[96] Any CI reagent can be used. O_2 would give charge exchange in both positive and negative ion modes from reactions of $O_2^{+\bullet}$ or $O_2^{-\bullet}$. $O^{-\bullet}$ will react primarily by proton abstraction from the samples. There are no stable negative ions in CH_4 or NH_3; therefore, the negative ions formed in these experiments would be formed by electron capture.

6.3 Electron Capture

Among the earliest experiments in negative ion CIMS were those done with CH_4 as the CI reagent as for the analysis of chlorinated hydrocarbons.[97] Since there are no stable negative ions in any reasonable concentrations in high pressures of methane after EI, any sample ions should be produced by capture of thermal electrons. In these experiments there were molecular ion clusters, abundant low molecular weight halogen-containing ions (Cl^-, $Cl_2^{-\bullet}$, H_2OCl^-, H_2Cl^-) and abundant adduct ions at $[M + Cl]^-$ and other masses. The spectra were useful, although complex, and were interpreted in terms of electron capture to give some $M^{-\bullet}$ ions and abundant Cl^- (and other fragment) ions that reacted with the samples to give the adduct ions (sample ion/sample molecule reactions, as discussed earlier). Subsequent experiments showed that with small sample sizes (obtained by GC/CIMS) of chlorinated

hydrocarbons and their derivatives, electron capture and dissociative electron capture were the dominant processes and that the adduct ions were formed when larger sample concentrations were present, as would be the case for probe introduction of samples.[98]

This technique of producing low-energy electrons by collisions with methane as the reagent gas is sometimes called ''methane enhanced'' negative CI.[98] The spectra are sensitive to trace amounts of O_2 and small amounts of oxygen were sometimes added to obtain reproducible spectra. The addition of oxygen to methane can form HO^- and other negatively charged ions; so the mixtures will give both electron capture and negative ion molecule reactions.[99] The methane negative ion CIMS of several halogenated benzenes gave predominantly the halogen ions from dissociative electron capture. Use of the CH_4/O_2 mixture gave spectra with greater abundances of higher molecular weight ions: $[M - H]^-$ or $[M + O - Cl]^-$ or $[M + O - Br]^-$ ions.

The main advantage of electron capture negative ion CIMS is the greatly enhanced sensitivity for compounds with high electron affinities.[96] Using CH_4 as the reagent gas to provide thermal electrons for electron capture and no negative ions for ion/molecule reactions, major differences in sensitivities were observed for negative ions over positive ions: factors of 100 to 1000 for compounds like 9,10-anthraquinone and phthalic anhydride or pentafluorobenzonitrile or 2,4,6-trinitrotoluene. Extremely high sensitivities and low detection limits could be obtained using electron capture negative CI with highly electronegative derivatives (like pentafluoropropionyl, CF_3CF_2COO-) of biologically important molecules.[100] There are continued efforts to detect smaller and smaller amounts of biologically important materials. Detection of $\sim 30 \times 10^{-18}$ g of material has been reported from gas chromatography/mass spectrometry (GC/MS) experiments using electron capture negative CI of an appropriately derivatized compound.[101]

Negative ion CI mass spectra are temperature sensitive, as are positive ion CI mass spectra (section 3.1). In general, the higher mass ions decompose to give lower mass fragments. In extreme cases, a temperature increase from ~100 to ~200 °C can cause the virtually complete fragmentation of $[M - H]^-$ ions.[94] Similarly, in electron capture negative CI, an increase in temperature can cause fragmentation of $M^{-\bullet}$ or thermal detachment of the electron from $M^{-\bullet}$ and greatly decrease the sensitivity of analyses.[102]

A major use of electron capture negative CI is in the analysis of polyhalogenated environmental pollutants.[103] The technique is extremely sensitive for highly halogenated compounds, but the sensitivities will vary widely for compounds of different degrees of halogenation. The spectra are sensitive to experimental conditions, as discussed above (pressure and purity of the moderating gas, temperature, and sample size) but reproducible differences have been observed which allow differentiation of isomers. The sensitivity factors are also strongly dependent on the experimental parameters. The $[M - 19]^-$ or $[M - Cl + O]^-$ ions are good indicators of small amounts of oxygen as an impurity in the reagent gas.

The fundamentals of electron capture negative CI have been investigated, but the system is complex and the results are not in agreement with expectations from simple models.[104] There is a general increase in electron capture response (or sensitivity) with increasing EA for many compounds, but there are also several unexpected and unexplained exceptions. Secondary processes of the initially formed $M^{-\bullet}$ ions (dissociation and thermal electron detachment) that are very sensitive to molecular structure are important in determining the overall sensitivities (or response factors) for the compounds. Very short calculated times for electron thermalization show clear advantages for CO_2 or CH_4 as reagent (moderator) gases. Values for rate constants for electron capture for many compounds were essentially the same at 10 torr and at ~1 atm.

7 ATMOSPHERIC PRESSURE IONIZATION

API or APCI was developed in the early 1970s by Horning et al. at the time that they were evaluating the precision and accuracy of mass measurement of plasma chromatography.[105–107] The source of ionizing radiation was ^{63}Ni on gold foil in the ionization chamber (a β-emitter that was the radiation source used in plasma chromatography). A 25 μm pinhole separated the source at atmospheric pressure from the quadrupole analyzer at $\sim 10^{-5}$ torr. Both positive and negative ions could be detected with the system. High purity N_2 (exact value not specified) was additionally purified with a 13X molecular sieve column to remove water and organic impurities.

Because the reactions are occurring at atmospheric pressure (many more collisions of ions with molecules than in a standard CI source) there were problems associated with adsorbed impurities in the inlet lines, in the source, and from previous samples. However, baking the ionization region and working with carefully purified materials reduced these difficulties.

The ionic reactions in an API (APCI) source can be quite complex, depending on the purity of the reagent gas. For extremely high purity N_2, the reactions are the following (Equations 54 and 55):

$$N_2^{+\bullet} + N_2 + N_2 \longrightarrow N_4^{+\bullet} + N_2 \qquad (54)$$

$$N_4^{+\bullet} + S \longrightarrow S^{+\bullet} + 2N_2 \qquad (55)$$

Consequently, one might expect to observe molecular ions and fragments from charge exchange reactions. Somewhat more realistically, it is impossible to remove all of the water from the system; so the following reactions are likely (Equations 56–58):

$$N_4^{+\bullet} + H_2O \longrightarrow H_2O^{+\bullet} + 2N_2 \quad (56)$$

$$H_2O^{+\bullet} + H_2O \longrightarrow H_3O^+ + HO^\bullet \quad (57)$$

$$H_3O^+ + nH_2O + N_2 \longrightarrow H_3O^+(H_2O)_n + N_2 \quad (58)$$

The degree of solvation of H_3O^+ depends on the concentration of water in the nitrogen reagent gas. Consequently, one might expect to observe MH^+ ions with basic compounds and little or no ionization with nonpolar species. If the samples are introduced in a solvent, then $N_4^{+\bullet}$ will react with the solvent to form ions that will then react with the sample that is necessarily present at a much lower concentration than the solvent. API mass spectra are strongly dependent on the concentrations of all compounds in the ionization source.

The observations that attracted the greatest attention were the extreme sensitivity of the method for basic compounds and the virtual absence of response for nonpolar or non-basic compounds. A strong correlation has been established between API sensitivities and the PA or gas-phase basicities of samples: an increase in sensitivity with increasing proton affinity to a limiting value for compounds with PA larger than a certain value (that would depend on the strength of the protonating agent).[108] With a substantial amount of water in the reagent gas, it was observed that the sensitivities of compounds showed a similar correlation with aqueous basicities.[108]

It was immediately obvious that combinations of API with GC and liquid chromatography would be both useful and relatively straightforward instrumentally. Such combinations are commercially available and routinely used for many analyses.[109–111]

Collisionally induced decomposition (CID), or mass spectrometry/mass spectrometry (MS/MS), has been used extensively with CIMS for identifying, characterizing, or quantifying targeted compounds by decomposing the MH^+ or $M^{+\bullet}$ ions of each species that are produced from low energy CI reactant ions.[8] Similarly, MS/MS has been combined with API or APCI to individual components, generally separated by liquid chromatography or GC.[110] The analysis of unresolved mixtures can be done with API as was done with CI.

Exchange reactions of ions with labeled polar compounds (such as D_2O, ND_3, CH_3OD) have been used to identify the number of labile hydrogens in molecules.[55,56] Such studies can be done in API experiments. If polar solvents are used in liquid chromatography, the sample ions are often solvated with solvent molecules and a small potential and a drying gas are used in the intermediate pressure region to desolvate the sample ions by ion/molecule collisions. Introducing a labeled gas into this region allows the exchange of labile protons to be observed.[112] Again, continuing the combination of techniques, one can apply MS/MS to these exchanged ions to interpret the CID processes.

8 PROSPECTS

CIMS is a ''mature'' technique: one that is well established and frequently used as an incidental technique in analytical problem solving. Although the basic phenomena are well understood, not all of the details of the reactions or physical processes (ion/electron recombination, ion/ion recombination, radical reactions, surface processes) are understood or easily explained. Additional work on the fundamentals of ion/molecule reactions is needed to explain the analytical observations more completely. Ion/molecule reactions are involved in sample ion production from other mass spectrometric techniques, including fast atom bombardment (FAB), electrospray ionization (ESI), and matrix-assisted laser desorption ionization (MALDI). The extent of involvement of ion/molecule reactions in these techniques is not well understood or always recognized.

CIMS is an analytical application of ion/molecule reactions. As the instrumentation for the study of ion/molecule reactions continues to improve, the potential for the analytical applications of the ion/molecule reactions becomes greater. FTICR and ion trap mass spectrometers have become more available and more easily used for the study of ion/molecule reactions. With these instruments, it is possible to study reactions of carefully chosen reactant ions with samples for selective detection of certain molecular properties (IP, EA, PA, stereochemistry) as well as reactions of sample ions with different neutral compounds. These ion/molecule reactions can be used for characterization of increasingly complex or heavier molecules or to detect more subtle differences in structures of complex molecules.[69,113] A few measurements of rate constants for H/D exchange have been observed for peptides and proteins, which were interpreted in terms of the structures of the peptides. Reactions of multi-functional ions with samples or reactions of sample ions with multi-functional compounds may be useful for the determination of the stereochemistry of isomeric species.

The drift velocities or mobilities of ions in gases have been measured for many years and the mobilities used to determine the interaction potentials between

these ions and the gases.[2,114] These measurements were initially limited to small ions, but subsequently a technique called ''plasma chromatography'' was developed for larger molecules as an analytical technique.[114] The mobility of an ion depends on both its mass and its size or shape and the variation with shape or molecular type hindered the acceptance of plasma chromatography for the analysis of small molecules, although the technique is extremely sensitive.

There has been a resurgence of interest in, and improved instrumentation for, ion mobility measurements and the technique is now called ''ion mobility spectrometry''.[115] Small differences in reduced mobilities of small molecules (mobilities recalculated to standard conditions) have been interpreted as resulting from different shapes (or sizes) of the ions and demonstrate the potential of ion mobility spectrometry for differentiation of isomers. Quantitative analyses have been demonstrated and the instruments are relatively portable for fieldwork.

The term ''ion chromatography'' is also used to describe the use of ion mobility measurements for determination of properties of ions. The groups doing ''ion chromatography'' experiments work on different systems and apply a more theoretical approach to the use of the reduced mobility measurements than those doing ion mobility spectrometry, but the techniques are very similar (if not identical). Recent work in ion chromatography has been directed toward the determination of the conformations of gas-phase ions of macromolecules from comparisons of experimental reduced mobilities with mobilities calculated from potential functions and assumed ionic geometries.[116,117]

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
CI	Chemical Ionization
CID	Collisionally Induced Dissociations
CIMS	Chemical Ionization Mass Spectrometry
EA	Electron Affinity
EI	Electron Ionization
EIMS	Electron Ionization Mass Spectrometry
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FTICR	Fourier Transform Ion Cyclotron Resonance
FTICR/MS	Fourier Transform Ion Cyclotron Resonance/Mass Spectrometry
GC	Gas Chromatography
GC/CIMS	Gas Chromatography/Chemical Ionization Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
HA	Hydride Affinities
ICR	Ion Cyclotron Resonance
ICR/MS	Ion Cyclotron Resonance/Mass Spectrometry
IP	Ionization Potential
MALDI	Matrix-assisted Laser Desorption Ionization
MS/MS	Mass Spectrometry/Mass Spectrometry
PA	Proton Affinities
PEG	Polyethylene Glycol
TOF	Time-of-flight

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of • Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Food **(Volume 5)**
Flavor Analysis in Food

Forensic Science **(Volume 5)**
Ion Mobility Spectrometry in Forensic Science • Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Coupled Liquid Chromatographic Techniques in Molecular Characterization

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. F.H. Field, M.S.B. Munson, 'Reactions of Gaseous Ions. XIV. Mass Spectrometric Studies of Methane at Pressures up to 2 Torr', *J. Am. Chem. Soc.*, **87**, 3289–3294 (1965).
2. L.B. Loeb, *Fundamental Processes of Electrical Discharges in Gases*, John Wiley & Sons, New York, Chap 2, 1939.
3. M.S.B. Munson, F.H. Field, 'Reactions of Gaseous Ions. XV. Methane + 1% Ethane and Methane + 1% Propane', *J. Am. Chem. Soc.*, **87**, 3294–3299 (1965).
4. M.S.B. Munson, F.H. Field, 'Reactions of Gaseous Ions. XVI. Effects of Additives on Ionic Reactions in Methane', *J. Am. Chem. Soc.*, **87**, 4242–4247 (1965).
5. M.S.B. Munson, F.H. Field, 'Chemical Ionization Mass Spectrometry. I. General Introduction', *J. Am. Chem. Soc.*, **88**, 2621–2630 (1966).
6. H.M. Fales, H.A. Lloyd, G.W.A. Milne, 'Chemical Ionization Mass Spectrometry of Complex Molecules. II. Alkaloids' *J. Am. Chem. Soc.*, **92**, 1590–1597 (1970).
7. T. Murata, S. Takahashi, T. Takeda, 'Chemical Ionization – Mass Spectrometry. II. Application to Sterol Esters', *Anal. Chem.*, **47**, 577–580 (1975).
8. D.V. Davis, R.G. Cooks, 'Direct Characterization of Nutmeg Constituents by Mass Spectrometry/Mass Spectrometry', *J. Agric. Food Chem.*, **30**, 495–504 (1982).
9. P. Rudewicz, B. Munson, 'Determination of Additives in Polypropylene by Selective Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **58**, 358–361 (1986).
10. A.G. Harrison, *Chemical Ionization Mass Spectrometry*, 2nd edition, CRC Press, Boca Raton, FL, 1992.
11. Y. Ikezoe, S. Matsuoka, M. Takebe, A. Viggiano, *Gas Phase Ion/Molecule Reaction Rate Constants Through 1986*, Maruzen , Tokyo, 1987.
12. H. Budzikiewicz, 'Negative Chemical Ionization (NCI) of Organic Compounds', *Mass Spectrometry Reviews*, **5**, 345–380 (1986).
13. Sir J.J. Thomson, *Rays of Positive Electricity and Their Application to Chemical Analysis*, Longmans, Green; London, 116–122, 1913.
14. T.R. Hogness, E.G. Lunn, 'The Ionization of Hydrogen by Electron Impact as Interpreted by Positive Ray Analysis', *Phys. Rev.*, **26**, 44–55 (1925).
15. F.W. Aston, *Mass-spectra and Isotopes*, Longmans, Green, London, 101, 1933.
16. M.M. Mann, A. Hustrulid, J.T. Tate, 'The Ionization and Dissociation of Water and Ammonia by Electron Impact', *Phys. Rev.*, **58**, 340–347 (1940).
17. H.W. Washburn, H.F. Wiley, S.M. Rock, 'The Mass Spectrometer as an Analytical Tool', *Ind. Eng. Chem., Anal. Ed.*, **15**, 541–547 (1943).
18. H.F. Washburn, H.F. Wiley, S.M. Rock, C.E. Berry, 'Mass Spectrometry', *Ind. Eng. Chem., Anal. Ed.*, **17**, 74–81 (1945).
19. J.J. Mitchell, R.H. Perkins, F.F. Coleman, 'Secondary Processes of Ion Production in the Mass Spectrometer', *J. Chem. Phys.*, **16**, 835–836 (1948).
20. F.W. McLafferty, 'Mass Spectrometric Analysis Aliphatic Ethers', *Anal. Chem.*, **29**, 1782–1789 (1957).
21. J.H. Beynon, *Mass Spectrometry and its Applications to Organic Chemistry*, Elsevier, London, 275–279, 1960.
22. V.L. Tal'roze, A.K. Ljubimova, 'Secondary Processes in the Ion Source of a Mass Spectrometer' *J. Mass Spectrom.*, **33**, 502–504 (1998). Translated from *Doklady Akad. Nauk. S.S.S.R.*, **86**, 909–912 (1952).
23. D.P. Stevenson, D.O. Schissler, 'Rate of the Gaseous Reactions, $X^+ + YH = XH^+ + Y$', *J. Chem. Phys.*, **23**, 1353–1354 (1955).

24. F.H. Field, J.L. Franklin, F.W. Lampe, 'Reactions of Gaseous Ions. I. Methane and Ethylene', *J. Am. Chem. Soc.*, **79**, 2419–2429 (1957).
25. S. Ghaderi, P.S. Kulkarni, E.B. Ledford, Jr, C.L. Wilkins, M.L. Gross, 'Chemical Ionization in Fourier Transform Mass Spectrometry', *Anal. Chem.*, **53**, 428–437 (1981).
26. M.S.B. Munson, F.H. Field, 'Reactions of Gaseous Ions. XVII. Methane + Unsaturated Hydrocarbons', *J. Am. Chem. Soc.*, **91**, 3413–3418 (1969).
27. F.H. Field, M.S.B. Munson, D.A. Becker, 'Chemical Ionization Mass Spectrometry Paraffin Hydrocarbons' in *Advances in Chemistry Series*, American Chemical Society, Washington, DC, **58**, 167–192, 1966.
28. R.P. Clow, J.H. Futrell, 'Ion Cyclotron Resonance Study of the Mechanism of Chemical Ionization. Mass Spectroscopy of Selected Hydrocarbons Using Methane Reagent Gas', *J. Am. Chem. Soc.*, **94**, 3748–3755 (1972).
29. M.S.B. Munson, F.H. Field, 'Chemical Ionization Mass Spectrometry. II. Esters', *J. Am. Chem. Soc.*, **88**, 4337–4345 (1966).
30. J.J. Evans, 'Chemical Ionization of Methyl Esters', Ph.D. thesis, University of Delaware, Newark, DE, 1998.
31. C.W. Tsang, A.G. Harrison, 'Effect of Chain Length on the Chemical Ionization Mass Spectra of Methyl *n*-Alkanoates', *J. Chem. Soc. Perkin Trans.*, **2**, 1718–1722 (1975).
32. I. Dzidic, J.A. McCloskey, 'Influence of Remote Functional Groups in the Chemical Ionization Mass Spectra of Long-chain Compounds', *J. Am. Chem. Soc.*, **93**, 4955–4956.
33. G.J. Kallos, V. Caldecourt, J.C. Tou, 'Matrix Effects in Quantitation and Identification by Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **54**, 1313–1316 (1982).
34. J. Michnowicz, B. Munson, 'Studies in Chemical Ionization Mass Spectrometry', *Org. Mass Spectrom.*, **4**, 481–492 (1970).
35. K. Blom, B. Munson, 'High Pressure Collisional Activation Mass Spectrometry', *J. Am. Chem. Soc.*, **105**, 3793–3799 (1983).
36. A. Onigbinde, 'Mass Spectrometric Analysis of Oligomeric Polyethylene Glycols, Polyethylene Glycol Monoalkyl Ethers, and Polyethylene Glycol Dimethyl Ethers', Ph.D. thesis, University of Delaware, Newark, DE, 1998.
37. D.F. Hunt, C.N. McEwen, 'Chemical Ionization Mass Spectrometry Studies – VII: Deuterium Labeled Decanes', *Org. Mass Spectrom.*, **7**, 441–448 (1973).
38. R.P. Morgan, E.J. Hayward, G. Steel, 'On the Evaluation of Ion Source Conditions for Optimum Sensitivity of Chemical Ionization', *Org. Mass Spectrom.*, **14**, 627–628 (1979).
39. E.P. Hunter, S.G. Lias, 'Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update', in NIST Chemistry Webbook, NIST Standard Reference Data Base, Number 69, eds. W.G. Mallard, P.J. Linstrom, March, 1998, NIST, Gaithersburg, MD, (http://webbook.nist.gov)
40. B. Munson, 'Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **43**, 28A–43A (1971).
41. H.M. Fales, G.W.A. Milne, T. Axenrod, 'Identification of Barbiturates by Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **42**, 1432–1435 (1970).
42. G.W.A. Milne, H.M. Fales, T. Axenrod, 'Identification of Dangerous Drugs by Isobutane Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **43**, 1815–1820 (1971).
43. R.J. Weinkam, H. Lin, 'Direct Mixture Analysis by Probe Insertion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **51**, 972–975 (1979).
44. B. Munson, B.L. Jelus, F. Hatch, T.K. Morgan, Jr, R.K. Murray, Jr, 'Stereochemical Effects in the Mass Spectra of 2-Hydroxy-, 5-Hydroxy-, and 2,5-Dihydroxyprotoadamantanes', *Org. Mass Spectrom.*, **15**, 161–165 (1980).
45. L. D'Haenens, C.C. Van de Sande, M. Schelfaut, 'The Chemical Ionization Mass Spectra of 2,4-Adamantanediol Diacetates and Dimethyl Ethers', *Org. Mass Spectrom.*, **17**, 265–268 (1982).
46. R.A. Hancock, R. Walder, H. Weigel, 'Chemical Ionization Mass Spectrometry: A Method of Estimating Reagent Gas Pressure in the Ion Source Based on Kinetic Data', *Org. Mass Spectrom.*, **14**, 507–511 (1979).
47. P. Rudewicz, B. Munson, 'Analysis of Complex Mixtures of Ethoxylated Alcohols by Probe Distillation/Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **54**, 674–679 (1986).
48. J.B. Westmore, M.M. Alauddin, 'Ammonia Chemical Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **5**, 381–465 (1986).
49. I. Dzidic, J.A. McCloskey, 'Chemical Ionization Mass Spectrometry Using Ammonia Reagent Gas: Selective Protonation of Conjugated Ketones', *Org. Mass Spectrom.*, **6**, 939–940 (1972).
50. D.F. Hunt, 'Reagent Gases for Chemical Ionization Mass Spectrometry', in *Advances in Mass Spectrometry*, ed. A.R. West, Applied Science Publishers (Institute of Petroleum), Barking, UK, 517–522, Vol. 6, 1974.
51. D.E. Smith, J.S. Smith, D. Jerolamon, A.F. Weston, D. Richton, E.J. Brozowski, 'Industrial Applications of Chemical Ionization with the Ammonium Ion', Presented at the 26th ASMS Conference on Mass Spectrometry, St. Louis, MO, 1978.
52. P. Rudewicz, B. Munson, 'Effect of Partial Pressure on the Sensitivities for Oxygenated Compounds in Ammonia Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **58**, 2903–2907 (1986).
53. R.B. Cody, 'Separation of the Reagent Ions from the Reagent Gas in Ammonia Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **61**, 2511–2515 (1989).
54. W.V. Ligon, Jr, H. Grade, 'Chemical Ionization Mass Spectrometry Utilizing an Isotopically Labeled Reagent Gas', *J. Am. Soc. Mass Spectrom.*, **5**, 596–598 (1994).

55. D.F. Hunt, C.N. McEwen, R.A. Upham, 'Chemical Ionization Mass Spectrometry. II. Differentiation of Primary, Secondary, and Tertiary Amines', *Tetrahedron Lett.*, **47**, 4539–4542 (1971).
56. Y.Y. Lin, L.L. Smith, 'Active Hydrogen by Chemical Ionization Mass Spectrometry', *Biomed. Mass Spectrom.*, **6**, 15–18 (1979).
57. L. Klevan, B. Munson, 'Gaseous Ionic Reactions in Tetramethylsilane', *Int. J. Mass Spectrom. Ion Physics*, **13**, 261–268 (1974).
58. J.M. McGuire, B. Munson, 'Comparison of Isopentane and Isobutane as Chemical Ionization Reagent Gases', *Anal. Chem.*, **57**, 680–683 (1985).
59. T.J. Odiorne, D.J. Harvey, P. Vouros, 'Chemical Ionization Mass Spectrometry Using Tetramethylsilane', *J. Phys. Chem.*, **76**, 3217–3220 (1972).
60. D. Clemens, B. Munson, 'Selective Reagents in Chemical Ionization Mass Spectrometry: Tetramethylsilane', *Org. Mass Spectrom.*, **20**, 368–373 (1985).
61. R.N. Stillwell, D.I. Carroll, J.G. Nowlin, E.C. Horning, 'Formation of Trimethylsilyl Molecular Adduct Ions in Desorption Chemical Ionization Mass Spectrometry of Nonvolatile Organic Compounds', *Anal. Chem.*, **55**, 1313–1318 (1983).
62. I.A. Blair, G. Phillipou, J.H. Bowie, 'Ion-Cyclotron Resonance Studies of Alkylsilyl Ions. I The Reactions Between Alcohols and the Trimethylsilyl Cation', *Austr. J. Chem.*, **32**, 59–64 (1979).
63. R. Orlando, F. Strobel, D.P. Ridge, B. Munson, 'Selective Reagents in Chemical Ionization Mass Spectrometry: Tetramethylsilane with Aliphatic Alcohols', *Org. Mass Spectrom.*, **22**, 597–605 (1987).
64. R. Orlando, B. Munson, 'Comparison of Theoretical and Experimental Relative Sensitivities in Gas Chromatography/Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **60**, 1815–1818 (1988).
65. R. Orlando, B. Munson, 'Trimethylsilyl Ions for Selective Detection of Oxygenated Compounds in Gasoline by Gas Chromatography/Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **58**, 2788–2791 (1986).
66. J.A. Stone, 'Gas-phase Association Reactions of Trimethylsilylium ($(CH_3)_3Si^+$) with Organic Bases', *Mass Spectrom. Rev.*, **16**, 25–49 (1997).
67. M. Vairamani, U.A. Mirza, R. Srnivas, 'Unusual Positive Ion Reagents in Chemical Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **9**, 235–258 (1990).
68. E.P. Burrows, 'Dimethyl Ether Chemical Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **14**, 107–115 (1995).
69. J.S. Brodbelt, 'Analytical Applications of Ion/Molecule Reactions', *Mass Spectrom. Rev.*, **16**, 91–109 (1997).
70. S. Prabhakar, M. Vairamani, 'Acetone Chemical Ionization Mass Spectrometry', *Mass Spectrom. Rev.*, **16**, 259–281 (1997).
71. N. Einolf, B. Munson, 'High Pressure Charge Exchange Mass Spectrometry', *Int. J. Mass Spectrom. Ion. Physics.*, **9**, 141–160 (1972).
72. M.E. Cisper, A.W. Garrett, D. Cameron, P.H. Hemberger, 'Analysis of Polar Organic Compounds Using Charge Exchange Ionization and Membrane Introduction Mass Spectrometry', *Anal. Chem.*, **68**, 2097–2101 (1996).
73. S.G. Lias, 'Ionization Energy Evaluation' in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, eds. W.G. Mallard, P.J. Linstrom, March 1998, NIST, Gaithersburg, MD. (http://webbook.nist.gov)
74. J.P. Hoppesch, R.C. Domingo, 'Picogram Detection of Sulfur Hexafluoride by High Pressure Charge Exchange Mass Spectrometry', *Anal. Lett.*, **8**, 839–848 (1975).
75. E. Stenhagen, S. Abrahamsson, F.W. McLafferty (eds.), 'Atlas of Mass Spectral Data', Interscience, New York, 816, Vol. 2, 1969.
76. F.W. Lampe, J.L. Franklin, F.H. Field, 'Cross-sections for Ionization by Electrons', *J. Am. Chem. Soc.*, **79**, 6129–6132 (1957).
77. L.W. Sieck, 'Determination of Molecular Weight Distribution of Aromatic Components in Petroleum Products by Chemical Ionization Mass Spectrometry with Chlorobenzene as Reagent Gas', *Anal. Chem.*, **55**, 38–41 (1983).
78. F. Hatch, B. Munson, 'Reactant Ion Monitoring for Selective Detection in Gas Chromatography/Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **49**, 731–733 (1977).
79. S.C. Subba Rao, C. Fenselau, 'Evaluation of Benzene as a Charge Exchange Reagent', *Anal. Chem.*, **50**, 511–515 (1978).
80. C. Allgood, Y.C. Ma, B. Munson, 'Quantitation Using Benzene in Gas Chromatography/Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **63**, 721–725 (1991).
81. G.F. Crable, G.L. Kearns, M.S. Norris, 'Low Voltage Mass Spectrometric Sensitivities of Aromatics' *Anal. Chem.*, **32**, 13–17 (1960).
82. S. Daishima, Y. Iida, F. Kanda, 'Evaluation of Nitric Oxide Chemical Ionization Mass Spectra of Substituted Benzenes', *Org. Mass Spectrom.*, **26**, 486–492 (1991).
83. D.F. Hunt, T.M. Harvey, W.C. Brumley, J.F. Ryan, III, J.W. Russell, 'Nitric Oxide Chemical Ionization Mass Spectrometry of Alcohols', *Anal. Chem.*, **54**, 492–496 (1982).
84. D.F. Hunt, C.N. McEwen, T.M. Harvey, 'Positive and Negative Chemical Ionization Mass Spectrometry Using a Townsend Discharge', *Anal. Chem.*, **47**, 1730–1734 (1975).
85. I. Dzidic, H.A. Petersen, P.A. Wadsworth, H.V. Hart, 'Townsend Discharge Nitric Oxide Chemical Ionization Gas Chromatography/Mass Spectrometry for Hydrocarbon Analysis of Middle Distillates', *Anal. Chem.*, **64**, 2227–2232 (1992).

86. B. Jelus, B. Munson, C. Fenselau, 'Charge Exchange Mass Spectra of Trimethylsilyl Ethers of Biologically Important Compounds: An Analytical Technique', *Anal. Chem.*, **46**, 729–730 (1974).
87. R. Chai, A.G. Harrison, 'Mixtures of Nitric Oxide with Hydrogen and Methane as Chemical Ionization Reagent Gases', *Anal. Chem.*, **55**, 969–971 (1983).
88. C.W. Polley, Jr, B. Munson, 'Nitrous Oxide as Reagent Gas for Positive Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **55**, 754–757 (1983).
89. C.E. Melton, 'Negative Ion Mass Spectra', in *Mass Spectrometry of Organic Ions*, ed. F.W. McLafferty, Academic Press, New York, Chap. 4, 1963.
90. J.E. Bartmess, 'Negative Ion Energetics Data' in *NIST Chemistry WebBook, NIST Standard Reference Data Base Number 69*, eds. W.G. Mallard, P.J. Linstrom, NIST, Gaithersburg, MD, March, 1998 (http://webbook.nist.gov)
91. A.L.C. Smit, F.H. Field, 'Gaseous Anion Chemistry. Formation and Reactions of OH^-; Reactions of Anions with N_2O; OH^- Negative Chemical Ionization', *J. Am. Chem. Soc.*, **99**, 6471–6483 (1977).
92. L.W. Sieck, K.R. Jennings, P.D. Burke, 'Hydroxyl Anion Chemical Ionization Screening of Liquid Fuels', *Anal. Chem.*, **51**, 2232–2235 (1979).
93. T.A. Roy, F.H. Field, Y.Y. Lin, L.L. Smith, 'Hydroxyl Ion Negative Chemical Ionization Mass Spectra of Steroids', *Anal. Chem.*, **51**, 272–278 (1979).
94. A.P. Bruins, 'Negative Chemical Ionization Mass Spectrometry in the Determination of Components in Essential Oils', *Anal. Chem.*, **51**, 967–972 (1979).
95. M. Cheung, A.B. Young, A.G. Harrison, '$O^{-\bullet}$ and OH^- Chemical Ionization of Some Fatty Acid Methyl Esters and Triacylglycerols', *J. Am. Soc. Mass Spectrom.*, **5**, 553–557 (1994).
96. D.F. Hunt, G.C. Stafford, Jr, F.W. Crow, J.W. Russell, 'Pulsed Positive Negative Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **48**, 2098–2104 (1976).
97. R.C. Dougherty, J. Dalton, F.J. Biros, 'Negative Chemical Ionization Mass Spectra of Polycyclic Chlorinated Insecticides', *Org. Mass Spectrom.*, **6**, 1171–1181 (1972).
98. E.A. Stemmler, R.A. Hites, 'Methane Enhanced Negative Ion Mass Spectra of Hexachlorocyclopentadiene Derivatives', *Anal. Chem.*, **57**, 684–692 (1985).
99. F.W. Crow, A. Bjorseth, K.T. Knapp, R. Bennett, 'Determination of Polyhalogenated Hydrocarbons by Glass Capillary Gas Chromatography/Negative Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **53**, 619–625 (1981).
100. C.N. Kenyon, P.C. Goodley, D.J. Dixon, J.O. Whitney, K.D. Faull, J.D. Barchas, 'Direct Liquid Inlet LC/MS and Negative Chemical Ionization MS', *Am. Lab.*, 38–49, Jan. (1983).
101. P. Wang, V. Murugaiah, B. Yeung, P. Vouros, R.W. Giese, '2-Phosphoglycolate and Glycolate-electrophore Detection, Including Detection of 87 Zeptomoles of the Latter by Gas Chromatography/Electron-capture Mass Spectrometry', *J. Chromatog. A*, **721**, 289–296 (1996).
102. B.J. Miwa, W.A. Garland, P. Blumenthal, 'Determination of Flurazepam in Human Plasma by Gas Chromatography/Electron Capture Negative Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **53**, 793–797 (1981).
103. V.S. Ong, R.A. Hites, 'Electron Capture Mass Spectrometry of Organic Environmental Contaminants', *Mass Spectrom. Rev.*, **13**, 259–283 (1994).
104. W.B. Knighton, L.J. Sears, E.P. Grimsrud, 'High-pressure Electron Capture Mass Spectrometry', *Mass Spectrom. Rev.*, **14**, 327–343 (1995).
105. E.C. Horning, M.G. Horning, D.I. Carroll, I. Dzidic, R.N. Stillwell, 'New Picogram Detection System Based on a Mass Spectrometer with an External Ionization Source at Atmospheric Pressure', *Anal. Chem.*, **45**, 936–943 (1973).
106. G.W. Griffin, I. Dzidic, D.I. Carroll, R.N. Stillwell, E.C. Horning, 'Ion Mass Assignments Based on Mobility Measurements', *Anal. Chem.*, **45**, 1204–1209 (1973).
107. D.I. Carroll, I. Dzidic, R.N. Stillwell, J.G. Horning, E.C. Horning, 'Subpicogram Detection System for Gas Phase Analysis Based on Atmospheric Pressure Ionization (API) Mass Spectrometry', *Anal. Chem.*, **46**, 706–710 (1974).
108. J. Sunner, G. Nicol, P. Kebarle, 'Factors Determining Relative Sensitivity of Analytes in Positive Mode Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **60**, 1300–1307 (1988).
109. D.I. Carroll, I. Dzidic, E.C. Horning, R.N. Stillwell, 'Atmospheric Pressure Ionization', *Appl. Spectr. Rev.*, **17**, 337–406 (1981).
110. J.D. Henion, B.A. Thomson, P.H. Dawson, 'Determination of Sulfa Drugs in Biological Fluids by Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **54**, 451–456 (1982).
111. R.K. Mitchum, W.A. Korfmacher, 'Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **55**, 1485A–1499A (1983).
112. M.E. Hemling, J.J. Conboy, J.F. Bean, M. Mentzer, S.A. Carr, 'Gas Phase Hydrogen/Deuterium Exchange in Electrospray Ionization Mass Spectrometry as a Practical Tool for Structure Elucidation', *J. Am. Soc. Mass Spectrom.*, **5**, 434–442 (1994).
113. M.K. Green, C.B. Lebrilla, 'Ion/Molecule Reactions as Probes of Gas-phase Structures of Peptides and Proteins', *Mass Spectrom. Rev.*, **16**, 53–71 (1997).
114. T.W. Carr (ed.), *Plasma Chromatography*, Plenum Press, New York, 1984.
115. G.A. Eiceman, Z. Karpas, *Ion Mobility Spectrometry*, CRC Press, Boca Raton, FL, 1994.
116. G. von Helden, T. Wyttenbach, M.T. Bowers, 'Conformation of Macromolecules in the Gas Phase: Use of

Matrix-assisted Laser Desorption Methods in Ion Chromatography', *Science*, **267**, 1483–1485 (1995).
117. G. Von Helden, T. Wyttenback, M.T. Bowers, 'Inclusion of a MALDI Source in the Ion Chromatography Technique: Conformational Information on Polymer and Biomolecular Ions', *Int. J. Mass Spectrom. Ion Proc.*, **146/147**, 349–364 (1995).

Discrete Energy Electron Capture Negative Ion Mass Spectrometry

James A. Laramée
Midwest Research Institute, Kansas City, USA

Robert B. Cody
JEOL USA, Inc, Peabody, USA

Max L. Deinzer
Oregon State University, Corvallis, USA

1 Introduction **1**
2 Background **2**
3 Electron–Molecule Interactions **3**
4 Electron Monochromator **6**
4.1 Designs 6
4.2 Performance 7
4.3 Analytical Information 9
4.4 Electron Capture Negative Ion Mass Spectrometry of Compound Classes 12
4.5 Mechanisms in Electron Capture Negative Ion Mass Spectrometry 21
5 Conclusions and Future Directions **23**
Abbreviations and Acronyms **23**
Related Articles **24**
References **24**

Standard electron capture negative ion mass spectrometry (ECNIMS) has long been known as a sensitive and selective method for the analysis of molecules with positive electron affinities. Applications of this technique for the analysis of environmental contaminants, particularly of halogenated compounds such as polychlorobiphenyls (PCBs), are common. However, the requirement for a moderating gas in the ion source to lower the energy of the electrons into the range required for resonance electron capture generates new experimental variables that are difficult to control. Differences in the moderating gas pressure cause changes in the electron energy distribution and in the resulting ion abundances, whereas small fluctuations in temperature often cause major differences in ion abundance ratios. Even more serious are the uncontrollable reactions between ions, molecules, and radicals within the high-pressure ion source region that give unpredictable spurious product ions, making spectral interpretation difficult and leading to wide variations of negative ion spectra from different instruments. Because of these difficulties, the advantages in sensitivity and specificity of ECNIMS for analysis of electronegative compounds often have been unrealized.

Better results can be obtained by having direct control over the electron energies. Control over electron energies can be achieved using a trochoidal electron monochromator, which allows the operator to tune to or scan a range of electron energies necessary for resonance electron capture. With this device some of the inherent difficulties encountered with ECNIMS can be avoided and reliable spectra for a broader range of compounds becomes possible. The ability to tune to a desired electron energy or to scan over a range of electron energies makes another dimension of analytical information available for identifying compounds. Quadrupole and sector instruments have been equipped with electron monochromators and the results are very promising. A clear advantage of these units is the reproducibility of the data they produce, the potential for three-dimensional spectra for easier interpretation of the results, and the availability of a new dimension of analytical information. These instruments have not shown the sensitivity inherent in the ECNIMS method. Improvements of about two or three orders of magnitude in sensitivity will be required before these instruments can compete with existing technology for trace analysis of environmental residues or other electron-capturing compounds. An improvement in the sensitivity of this magnitude is well within reasonable expectations and it should not be long before gas chromatography/electron monochromator mass spectrometry (GC/EMMS) becomes a standard technique for detecting environmental compounds, explosives and substances of abuse.

1 INTRODUCTION

ECNIMS is one of the most sensitive and specific analytical methods available when applied to the right kind of samples. Compounds that respond well to this technique are those that capture low-energy electrons and form molecular radical anions or dissociative products consisting of radicals and anions. These compounds readily undergo resonance capture of slow electrons because of the presence of electronegative groups or atoms and/or extended conjugated π-systems. Aromatic nitro, carbonyl, phosphate, and halogenated compounds are among the strongest electron absorbers, and for this reason the ECNIMS technique has been favored whenever possible for the analysis of environmental compounds where sensitivity and specificity are of paramount importance.

Although the advantages of ECNIMS are well established, there are equally well-known disadvantages to the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

method, including a lack of reproducible data particularly for quantitative studies and the general lack of information on mechanisms of ion formation. Efforts to remedy the problems of reproducibility have met with mixed success, largely because of the difficulty of working with low-energy electrons and the presence of a reagent gas in the ion source that not only generates slow electrons but also adds positive ions to the ion plasma that can react to form spurious ion–ion or ion–molecule reaction products. The potential to solve this problem is embodied in the electron monochromator, which can deliver discrete low-energy electrons in the absence of reagent gases. The electron monochromator has been available for several decades and the use of these devices in the ion source of mass spectrometers was demonstrated almost from the outset. Many of the reproducibility problems associated with ECNIMS can be avoided by using electron monochromator mass spectrometry (EMMS). The specificity of compound detection is enhanced further through the additional dimension of analytical information gained from resonance capture of electrons of discrete energy. ENMS instruments are now commercially available from JEOL USA, Inc and provide an alternative to the standard gas moderated ECNIMS ionization sources. Direct formation of negative ions provides information that can be used to identify target compounds and distinguish isomers. Because there is no reagent gas and no bipolar plasma, the electron monochromator provides highly reproducible negative ion mass spectra and avoids artifacts that can be observed in conventional ECNIMS.

Meanwhile GC/EMMS, when used in conjunction with high-level ab initio computational methods, can provide information on the mechanism of negative ion formation that is difficult to obtain by other methods. Much of the information available on energies of resonance states comes from studies carried out by electron transmission spectroscopy (ETS). This is a powerful technique for studying electronic and vibrational energy levels of electron-absorbing molecules, even for molecules that fail to form stable anions. But the technique gives no information on the nature of the anions being measured. The advantage of EMMS is that the masses of the ions formed are recorded along with the electron attachment energies. Thus, ETS measurements and the data from EMMS are complementary. In this account, emphasis is placed on what can be learned about the mechanistic aspects of resonance electron capture ion-forming reactions in the gas phase using EMMS.

2 BACKGROUND

The production of analytically useful negative ions in mass spectrometry has a history as long as that for chemical ionization mass spectrometry. The terminology ''negative chemical ionization mass spectrometry'' (NCIMS) or ''electron capture negative chemical ionization mass spectrometry'' (ECNCIMS) reveals the origin of the method but the terminology itself is somewhat misleading with respect to the process involved in forming negative ions. ECNIMS is a preferred title for this technique because it is free of any suggestions of how the slow electrons necessary for resonance electron capture processes are produced.

Standard ECNIMS has long been known as a sensitive and selective method for the analysis of electronegative compounds,[1,2] and excellent texts[3] and reviews[4–7] have been written about the subject. It is the preferred technique, for example, in the analysis of PCBs and other halogenated compounds. Because of the inherent sensitivity of the method, an important application of ECNIMS has been the analysis of tissue and biological fluid samples. Polychlorophenols and biphenyls in human adipose tissues,[8] tetrachloroterephthalates in human urine samples,[9] a variety of chlorinated compounds in human seminal fluids,[10] polychorinated dibenzofurans, dibenzo-*p*-dioxins and styrenes in fish,[11] and pentachlorophenol and related chlorinated compounds in human organs[12] are just a few examples of analyses that have been carried out by ECNIMS.

To generate the slow electrons required for resonance electron capture, a reagent gas, variously called moderating gas or buffer gas, must be introduced into the ion source. The reagent gas, usually methane or isobutane, acts to remove kinetic energy from the accelerated electrons that are emitted from the filament. A combination of ionization of the reagent gas molecules (Equation 1) that reduces the electron energies to around 30 eV,[13] plus kinematic processes resulting from inelastic collisions with reagent gas molecules,[14] further reduces the electron energies to thermal or epithermal levels. The pressure of the buffer gas, which is usually in the range 53.3–106.6 Pa (0.4–0.8 Torr), determines the final electron energies and their distributions. Typically, the source is operated so that about 50% of the electrons have energies close to 0 eV.[6]

$$2CH_4 + 2e^- \longrightarrow CH_4^{+\bullet} + CH_3^+ + H^\bullet + 2e^-(1^\circ) + 2e^-(2^\circ) \quad (1)$$

The presence of the reagent gas in the ion source, however, introduces a number of complications so that good control over the ionization process is difficult to achieve. Different gas pressures cause changes in the electron energies and in the resulting relative ion intensities,[15–18] and small fluctuations in temperature cause major differences in ion abundance ratios.[19–21] Variations in ion transmission efficiencies of the extraction optics of

different instruments also affect ion abundance ratios.[22] Various ions and neutral molecules in the source can react to give unpredictable spurious product ions,[23–25] making it difficult to interpret the spectra. Wide variations in electron energies between different instruments may be a factor responsible for the poor reproducibility of negative ion spectra from different laboratories.[26] Because of these difficulties, the advantages in sensitivity and specificity for the analysis of electronegative compounds often are compromised or lost entirely[22] and the method is often avoided, especially for any kind of semiquantitative purposes. Nevertheless, ECNIMS would be a powerful tool if the electron energies could be controlled and the ion–ion and molecule–ion interactions avoided. EMMS provides control over the electron energies and also avoids the reagent gas that causes spurious product ions. Most importantly, the use of EMMS provides a new dimension of analytical information to characterize compound classes, which cannot be readily achieved by standard ECNIMS methods.

3 ELECTRON–MOLECULE INTERACTIONS

The salient treatises on negative ion resonances (NIR) can be found in the two comprehensive volumes by Christophorou.[27,28] Excellent reviews on the subject have also been published.[7,29–31] Only the basics of electron–molecule interactions are mentioned here.

There are three processes leading to the formation of gas-phase negative ions that can be observed by mass spectrometry, and they result from the interaction of neutral molecules with electrons. Ion pair formation occurs when electrons, of energies equal to at least the ionization potential plus electron affinity of the neutral fragment, i.e. electrons of energies 10–15 eV, interact with neutral molecules to give excited states that dissociate into positive and negative ions (Equation 2). The cross-section for this process increases roughly linearly over a threefold range with increasing electron energies.[32] However, there is no particular advantage to this process in mass spectrometric analysis and as a result it has found little practical use.

$$AB + e^-(\sim 15\,eV) \longrightarrow (AB^*) \longrightarrow A^+ + B^- + e^- \quad (2)$$

The potential energy surfaces for the formation of useful negative ions are shown in Figure 1. When an electron of low energy (0–10 eV) interacts with an electron-capturing molecule AB in the gas phase, it can undergo: nonresonant elastic scattering; inelastic scattering, which leaves the molecule in an excited state (Equation 3):

$$AB + e^-(0–10\,eV) \longrightarrow AB^* + e^- \quad (3)$$

or resonant attachment, wherein a temporary or transient negative ion ($AB^{-\bullet}$) or ($M^{-\bullet}$) is produced (Equations 4–6). If the kinetic energy of the incident electron goes solely into nuclear motion of the neutral molecule, long-lived molecular radical anions ($M^{-\bullet}$) (Figure 1) can be produced by nuclear-excited Feshbach resonance, with lifetimes that can be equal to or exceed the microsecond range (Equation 5)[33] or the time-frame necessary for detection by mass spectrometry. In this type of resonance, there is no electronic excitation; the energy of the captured electron goes into molecular vibrations. Because of this excess energy, molecular ions formed by Feshbach resonance can decay by autodetachment, i.e. by loss of the electron and return to the neutral state (Equation 4), and the ion therefore would not be detectable by mass spectrometry. One of the purposes of the reagent or buffer gas in the ion source is to stabilize the molecular ion by removing excess energy via collisions.

$$AB + e^-(0–10\,eV) \longrightarrow (AB^{-\bullet})^* \longrightarrow AB + e^- \quad (4)$$

$$AB + e^-(\sim 0\,eV) \longrightarrow (AB^{-\bullet})^* \longrightarrow AB^{-\bullet} \quad (5)$$

$$AB + e^-(0.1–10\,eV) \longrightarrow A^\bullet + B^- \quad (6)$$

A long lifetime (T_a) also arises if the anion's internal energy can be distributed over a large number of vibrational degrees of freedom (N). A reasonable correlation between the lifetime and the electron affinity (EA) was found for monosubstituted nitrobenzenes (Equation 7):[36]

$$T_a^{1/(N-1)} \propto EA \quad (7)$$

This relationship was tested further by inclusion of the electron attachment rate constant k_a (Equation 8) to produce the molecular radical anion:[7,36]

$$(k_a T_a)^{1/(N-1)} \propto EA \quad (8)$$

The correlation was not improved further and it was concluded that regardless of wide variations in k_a values there is no convincing correlation with the electron affinities nor, for that matter, any relationship with dipole moments or structural changes caused by the formation of the radical anion.[35] The authors conclude that the initial step in resonance electron capture is a very poorly understood process.[7]

Trapping of the incident electron by the molecule in its ground state can produce short-lived transient molecular radical anions with lifetimes that are generally less than the nanosecond range or as short as that of a single vibration.[33] This type of electron–molecule interaction leads to shape resonances that are responsible for the production of most of the anion and radical fragments (Equation 6) observed in ECNIMS.

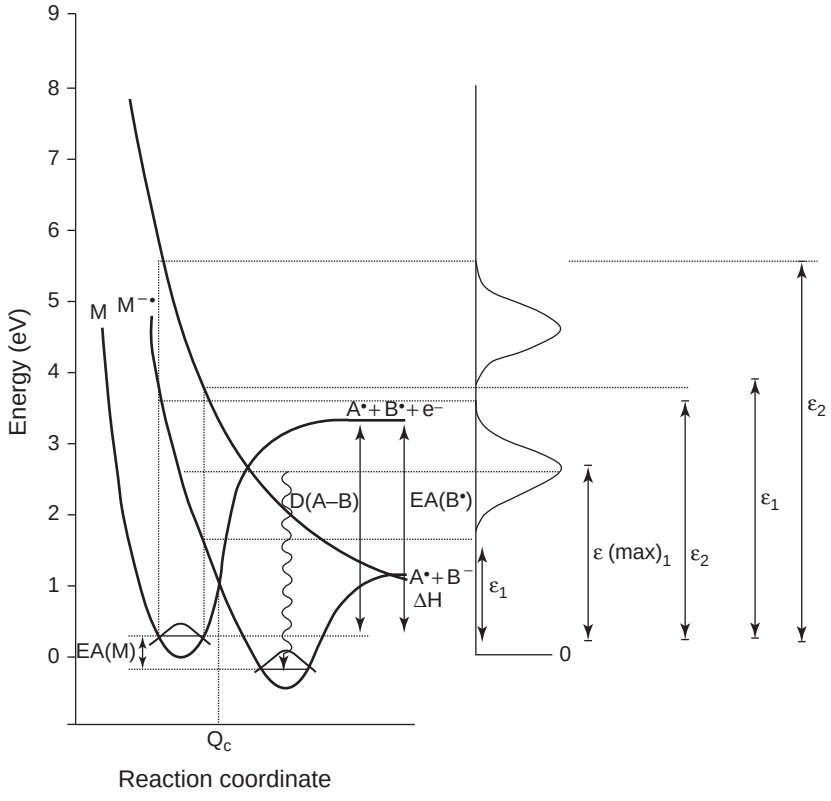

Figure 1 Hypothetical one-dimensional Born–Oppenheimer energy diagram showing the relationship between the capture of an electron by a neutral molecule (AB) and the production of a molecular radical anion ($M^{-\bullet}$) and fragment anion (B^-) via dissociative electron capture or metastable ion decomposition.

The overall process leading to $A^\bullet$ and B^- is shown as a purely repulsive potential energy surface (Figure 1), even though a shallow attractive state is present. The temporary or transient negative ions result from the attractive polarization between the neutral molecule and electron, and the repulsive force from the relative motion of the two bodies.[33,34] A shallow well is created, hence the terminology ''shape resonance''. The electron attaches and detaches by tunneling thorough the angular momentum barrier, and the height and thickness of this barrier determine the lifetime of the temporary negative ion. Orbital symmetry, resonance energy, and ion lifetimes are interrelated.[34] Radical anions involving σ^* orbitals are short-lived and generally produced with electrons of energies greater than 5 eV, as is the case for saturated hydrocarbons, whereas electron capture into low-lying antibonding π^* orbitals have much higher angular momentum barriers to electron tunneling. These anions may be sufficiently long-lived to observe vibrational fine structure in ETS experiments.[31,34]

The electron affinities of neutral molecules are an important criterion in determining whether or not stable anions can be formed. The electron affinity is the energy difference between the electronic ground states of the neutral molecule and radical anion, each in its lowest vibrational mode. By convention, molecules that capture electrons and produce radical anions that lie energetically below the ground state of the neutral molecule have positive electron affinities, and those that form radical anions lying energetically above the neutral molecule have negative electron affinities.[3] The former ions are sufficiently stable to be detected by mass spectrometry under normal ECNIMS operating conditions. Benzene, for example, has an electron affinity of -1.2 eV[37,38] but the molecular ion is not observed by ECNIMS, whereas hexafluorobenzene has a positive electron affinity of $\sim$0.4 eV[37] and produces an intense molecular ion peak with electrons of energies $\sim$0.03 eV.

A variety of different experimental procedures, theoretical methods, and semi-empirical approaches have been used to estimate electron affinities, but the values obtained are often in poor agreement with each other.[37] Relationships between electron affinities and NIR states or lowest unoccupied molecular orbital (LUMO) energies (Equation 9) can be useful within a given series or class of compounds.[37,39] The EA/ε_{LUMO} relationship requires a constant, δ: the reorganization energy

correction term to account for the charge density readjustment when an electron is added to the molecule. Its value lies in the range 0–3 eV. A scaling constant ξ is also required:

$$EA = -\xi\varepsilon_{\mathrm{LUMO}} - \delta \tag{9}$$

The lower the LUMO energies, the more likely the neutral molecule will capture an electron to form a stable molecular radical anion. Conjugation and electron-withdrawing groups on molecules lower the LUMO energies and favor molecular ion formation.

Dissociative electron capture processes (Equation 6) are distinctly separate events that generally do not result from decomposition of a stable molecular ion. Although metastable ions do occur in negative ion mass spectrometry, the resonances for dissociative capture are energetically distinct from those that produce the molecular ions.[39] As noted above, transient molecular radical anions produced via shape resonances are involved in dissociative electron capture events. Anions produced by dissociative electron capture (Equation 6) have another channel available for dissociation with a rate that is competitive with autodetachment. The dissociative electron capture cross-section, σ_{da}, is dependent on the cross-section for electron capture, σ_0, as well as on the time that it takes for dissociation to occur, T_{d}, versus the autodetachment lifetime, T_{a} (Equation 10):[33]

$$\sigma_{\mathrm{da}} = \sigma_0 \exp\left(\frac{-T_{\mathrm{d}}}{T_{\mathrm{a}}}\right. \tag{10}$$

When temporary negative ions are formed along repulsive potential energy surfaces, the ion lifetimes are short, i.e. 10^{-13}–10^{-15} s. Most polyatomic molecules fall into this category. However, lifetimes extending to the microsecond range occur when the resonances are close to 0 eV, i.e. via Feshbach resonances. In this case the ions may undergo metastable decompositions (Figure 1), which involves redistribution of the energy and often molecular rearrangements.[30] Although dissociative electron capture processes do not require stabilizing collisions,[33] an optimum reagent gas pressure is required for maximum ion production.[7,15] This is because the dissociative electron capture process, like the molecular ion-forming process, is a resonance phenomenon.

Dissociative capture fragment ions may result from several resonance states. In addition to shape resonances, there are core-excited resonances in which electron attachment is accompanied by promotion of the molecule to an electronically excited state.[33] As a result of this promotion of the core electron, the core itself becomes positively polarized, which provides an attractive potential for the electron being captured.

The minimum energy required for the formation of any anion is the appearance energy, ε_1. Within the Born–Oppenheimer approximation, i.e. no nuclear motion when the electron is captured, the energy ε_1 is the difference in energy between the ground state, when the neutral molecule is at the relaxed classical turning point for nuclear motion, and the anionic surface (Figure 1). The Franck–Condon principle requires that transitions to the anionic surface occur only within the energy range ε_1–ε_2. The position and shape of the anion yield curve for B^- are simply a reflection of the neutral molecule's wave function bounded by the slope of the anionic dissociative surface within the Franck–Condon region.

The dissociation limit or minimum heat of reaction (ΔH_0) is the difference between the bond dissociation energy (D) of AB and the electron affinity (EA) of $B^{\bullet}$ (Equation 11):

$$\Delta H_0 = D(\mathrm{A{-}B}) - EA(\mathrm{B}^{\bullet}) \tag{11}$$

Unless there is electronic or vibrational predissociation in the precursor ion,[27] the dissociation limit lies below the portion of the anionic surface that is located within the Franck–Condon region. It is apparent that the relatively modest energy requirements for dissociative electron capture result from the high electron affinity of $B^{\bullet}$. Thus, a large number of molecules in their electronic and vibrational ground states undergo dissociative electron capture processes despite ε_1 values (0–2 eV) that are much smaller than the bond dissociation energies D(A–B). Dissociative processes for chlorinated compounds from ground electronic states, for example, may involve resonances that are less than 1 eV. Fluoroorganics generally require higher electron energies because of the larger carbon–fluorine bond dissociation energies. But threshold energies ε_1 for fluoride ion production may be significantly smaller than the bond dissociation energies because of the high electron affinity ($EA = 3.45$ eV) of the fluorine atom.[37] The appearance energy ε_1 for fluoride ion from CF_2Cl_2, for example, is 1.8 eV[40] and the carbon–fluorine bond dissociation energy D(C–F) is 4.93 eV. These values give $\Delta H_0 = 1.48$ eV. In addition to these low-energy pathways, most molecules have dissociative capture channels that involve higher energy states. Ions formed by electron capture can possess excess energy ε_{x} (Equation 12). Thus, the fluoride ion produced from CF_2Cl_2 possesses 0.32 eV of excess energy, which is partitioned into translational and rotational energy or held internally as vibrational energy.[40]

$$\varepsilon_1 = \Delta H_0 + \varepsilon_{\mathrm{x}} \tag{12}$$

Negative ions produced with electrons of energies <1 eV are generally more intense than those produced

from capture of higher energy electrons. Generally, the cross-section for electron capture to form temporary negative ions depends on the relative thickness and height of the angular momentum barrier through which the electrons must tunnel, and these are greater for resonances close to 0 eV. The steepness of the repulsive potential surface also dictates the lifetime for autodetachment T_a (Equation 10), which occurs only within the boundary of the neutral molecule potential energy surface. Once the internuclear distance is large enough to have passed the crossover point, Q_c (Figure 1), autodetachment no longer occurs[33,62] and the temporary negative ion proceeds to the product species.

4 ELECTRON MONOCHROMATOR

4.1 Designs

Electrons from a hot filament possess an energy spread of 0–2 eV according to a Maxwell–Boltzmann distribution. By varying the filament potential, standard instruments can be used to generate slow electrons in the energy range 0–10 eV that are necessary for resonance electron capture. Such procedures have been used, for example, to study the NIR spectra of mixed anhydrides formed from fatty acids and trifluoroacetic acid[41] and a variety of myristic acid derivatives and their methyl esters.[42]

Investigators who were interested in obtaining accurate values for the ionization potentials of atoms and simple molecules, however, recognized early on that the energy spread (±5 eV) of electrons from a hot filament was unacceptable for this purpose. Hence, the need arose for sources with electron energy bandwidths of <0.1 eV. This need led to the development of a variety of electron monochromator designs long before their general utility in mass spectrometry could be appreciated. All early monochromators were designed to produce positive ions, and the production of thermal electrons was not particularly important. One early design utilized a magnetic field of about 10 G to disperse the electrons according to their energies.[43] It was, however, difficult to maintain a uniform magnetic field of this strength in the presence of stray magnetic fields. The electrostatic parallel-plate electron-energy selector[44,45] was simple in design, but massive room-sized Helmholz coils were necessary to shield out stray magnetic fields. A 127° cylindrical electrostatic deflector[46–48] was one of the more popular early electron monochromator designs, because the foci of electron trajectories are located at the field boundaries. With this device, electron beams of 0.1 μA and energy bandwidths of <0.1 eV became available in the energy range 0–50 eV.[47] But space charging problems in these monochromators resulted in the addition of grids and electron-absorbing surfaces referred to as "electron velvet". An early 180° hemispherical electrostatic deflector[49] avoided the use of slits in the entrance and exit planes, instead using electron lenses at high potentials. Space charging was thus reduced without using coatings and grids, although the strong deceleration of the electron velocities created its own set of undesirable effects. This instrument produced homogeneous electron beams with energy spreads of <0.005 eV but no improvement in beam brightness.

In all of these designs, the intensities of the electron beams produced were very weak and efforts to increase beam brightness could only be achieved at the expense of energy bandwidths that were much greater than those that would have been predicted on the basis of the filament temperatures.[50] Moreover, the performance of electrostatic analyzers is unpredictable when gases introduced into the target chamber modify the contact potentials of the electron optic elements.[51] Irregular beam transmission is also a problem when the target region is not shielded against stray electrostatic and magnetic fields. At low electron energies, this problem becomes even worse.

Hemispherical analyzers provide two-dimensional focusing and have a resolving power almost twice that of cylindrical analyzers of the same radius, but they must be operated at 1–3 eV.[52] A completely different approach for producing zero-energy electrons, referred to as reversal electron attachment detection,[53,54] involves focusing an electron beam into an electrostatic mirror to a plane where the electron translation is reversed, at which point the longitudinal electron energy is zero. Non-zero-energy electrons are captured along the longitudinal axis at some point beyond the reversal plane. The method was developed for ECNIMS, but it is difficult to obtain true energy monochromation and an electron beam that can produce significant abundance of ions with it.

As early as 1907, Thomson[55] recognized the advantages of using crossed electric (**E**) and magnetic (**B**) fields for determining the m/z values of positive and negative ions generated under a variety of experimental conditions. The $\mathbf{E} \times \mathbf{B}$ field was configured with the initial beam trajectory oriented perpendicular to both fields. It was recognized that crossed electric and magnetic fields have perfect focusing that depends only on the m/z value and not on the initial velocity or direction of the charged particles entering the analyzer.[56] By 1955 Fox had introduced a new electron gun that used electrodes to create retarding potential differences, and its original purpose was to reduce the energy spread of electrons in axial magnetic fields.[57] These devices were initially used to determine ionization potentials of rare gases,[58] but soon afterwards an electron monochromator was described[59] that utilized the insights of Fox to confine a low-energy

electron beam by an axial magnetic field and Thomson's crossed electric and magnetic field to disperse simultaneously the electrons according to their kinetic energies. Electron monochromators utilizing the basic $\mathbf{E} \times \mathbf{B}$ field principle have gained some popularity for resonance electron capture studies in recent years.[27,29,30,61–63]

In a typical monochromator design (Figure 2), electrons emitted from the filament (a) are focused by electrodes (b) into a deflection region (c) defined by the crossed magnetic (**B**) and electric fields (**E**). The electrons within these crossed fields undergo trochoidal motion[56] along the equipotential surfaces of the magnetic field lines according to their energies. The monochromatic electron beam emerges through an aperture from the deflection region (c) and is focused by electrodes (d) into the ion chamber (g). A small electric field (**E**) of about 0.4 V cm^{-1} in the deflection region (c) is established between a pair of deflecting plates oriented perpendicular to the electron beam, and a magnetic field (**B**) of about 130 G is oriented perpendicular to **E**. The drift velocity ($\mathbf{v}_d$) (Equation 13)

$$\mathbf{v}_d = \frac{\mathbf{E} \times \mathbf{B}}{B^2} \tag{13}$$

governs the electron energy (ε) by displacing the electron beam of initial velocity υ_0 a distance D within the crossed field region of length L. The final electron kinetic energy is determined by the potential difference between the filament center and the ion reaction chamber. The electron energy can be controlled by varying **B**, **E** or υ_0. The last one, which is controlled by the filament potential, is most convenient. This design is very successful because the electrons are confined without chromatic loss in the magnetic field, regardless of their energies. The axial magnetic field provides constant electron beam focusing throughout the energy range (0–60 eV), thus eliminating focal adjustments during operation.[63] In addition to quadrupole analyzers,[27,29,30,61–63] the electron monochromator has been adapted to sector instruments.[64–66]

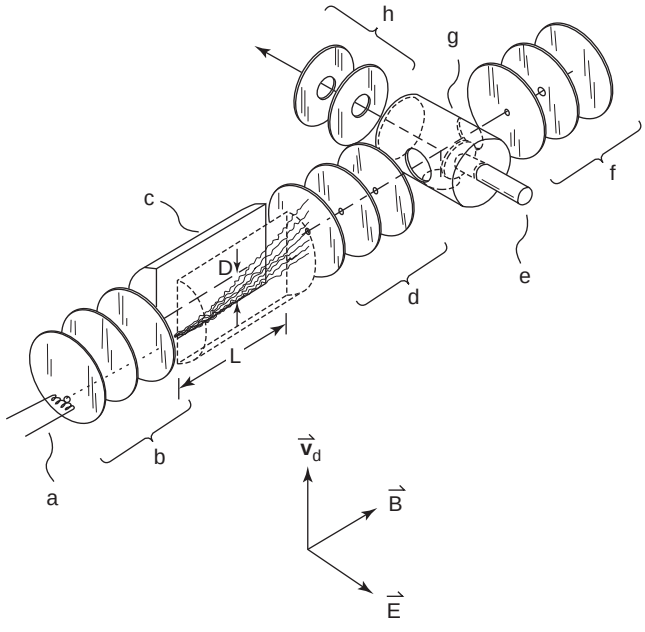

Figure 2 Diagram of a trochoidal electron monochromator: filament (a); collimating and focusing lenses (b); deflection region (c) (note: opposing dee shown with dashed lines to allow for clarity of electron beam path); exit lenses (d); anion repeller (e); electron collector electrodes (f); reaction chamber (g); and ion extraction optics (h). Note the trochoidal motion of the electron beams in the region of electric dees.[62]

4.2 Performance

4.2.1 Sensitivity

High analytical sensitivity in mass spectrometry begins with bright electron beams in the ion source. The prototype electron monochromator adapted to a quadrupole analyzer[61] transmits an electron beam of 2 μA with an energy spread of ±0.1 eV full width at half-maximum (fwhm). The usual trade-off between beam current and resolution is observed. By reducing the exit aperture size of the monochromator (Figure 2) as well as the electric field strength, the resolution can be improved. At ±0.07 eV (fwhm) energy distribution, the transmitted current is 0.5 μA. With an electron energy setting of 0.03 eV, 50 pg of hexachlorobenzene effluent from a gas chromatograph can be detected with a signal-to-noise ratio of ≥20. The Pierce element[67] used in the prototype for the sector instrument yields electron beams of 430 μA at 0.03 eV kinetic energy and gives a benchmark sensitivity of 1 pg of hexachlorobenzene.[66] Further refinements by JEOL USA, Inc have produced detection limits of 10 fg at a signal-to-noise ratio of 30. The reasons for this lack of sensitivity are not clear at this time.

4.2.2 Collisional Stabilization

The moderating gas introduced into the ion source for generating slow electrons also acts to remove the excess energy from the molecular ions and stabilizes them towards autodetachment (Equation 4).[6,7,36] Because the EMMS instrument provides direct control over the energy of the ionizing electron beam, the moderating gas is not needed to thermalize the electrons, and collisional stabilization of the ions can be studied independent of the electron energies.

The lowest pressure (c. 6.67×10^{-2} Pa or 0.5 mTorr) attainable in the ion source when a gas chromatograph is used is limited by the residual gas flowing from the capillary column.[68] Thus, a pressure change of helium from 6.67×10^{-2} Pa to 1.33–1.60 Pa (0.5 mTorr to 10–12 Torr)

Table 1 Effect of helium buffer gas on collisional stabilization

Chemical	Molecular ion intensity		Fragment ion intensity	
	6.67×10^{-2} Pa (0.5 mTorr)	1.33–1.60 Pa (10–12 mTorr)	6.67×10^{-2} Pa (0.5 mTorr)	1.33–1.60 Pa (10–12 mTorr)
Hexachlorobenzene	190	3100	–[c]	–[c]
Paraoxon[a]	54	970	–[c]	–[c]
TCDD[b]	0	26	6420 (Cl^-)	6360 (Cl^-)
Ethiofencarb	–[c]	–[c]	215 (m/z 167$^-$)	164 (m/z 167$^-$)

[a] 1,2,3,4-Tetrachlorodibenzo-*p*-dioxin.
[b] *O*, *O*-Diethyl-*O*-4-nitrophenylphosphate.
[c] Not measured.

causes about a 20-fold enhancement in the molecular ion intensities (Table 1). The effect of this pressure difference on the dissociative attachment ions is essentially zero (Table 1) or, more likely, a small decrease in ion signal because of ion scattering.[68] Further increase in helium pressure to 3.33 Pa (25 mTorr) raises the molecular ion intensities by only an additional 1.5-fold (Figure 3).

Different buffer gases, of course, have different effects on the stabilization of the molecular ions (Figure 3). Gases possessing internal degrees of freedom have a greater stabilizing effect. With the helium gas flowing from the gas chromatograph creating a pressure of 1.33–1.60 Pa (10–12 mTorr), added methane at about 1.1 Pa (8 mTorr) causes a further threefold increase in the intensity of the molecular ion of hexachlorobenzene. Similar but smaller increases in the molecular ion intensities are observed when isobutane, ethylene, carbon dioxide and hydrogen are added. Each of these gases shows an optimum pressure for ion stabilization, after which the intensities decrease because of ion scattering.

4.2.3 Electron Flux Variation with Added Buffer Gases

Ideally the electron flux should remain constant over the entire electron energy range (0–10 eV) of interest. In ECNIMS this aspect is entirely overlooked because it is impossible to obtain a constant flux of electrons at different energies without changing the reagent gas pressure. Typically, the electron energy distribution is ~50% near 0 eV, 5% between 0 and 0.15 eV, and 12% between 0.15 and 0.45 eV. With the EMMS system the problem is not nearly as severe. Experiments have shown that the electron current varies by a maximum of ±30% (~0.1 μA eV^{-1}) over the energy range 0–3.5 eV (Figure 4). The addition of carbon dioxide as a buffer gas

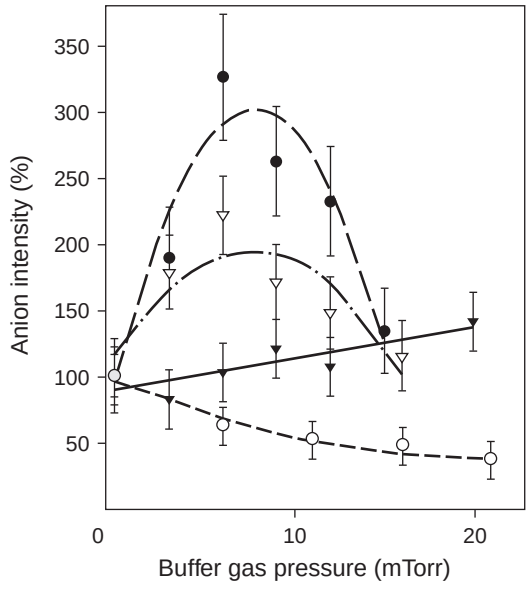

Figure 3 The effect of buffer gas pressure on 2, 2′, 3, 4, 4′, 5, 6-pentachlorobiphenyl radical anion intensities: (●) $M^{-\bullet}$, with methane buffer gas; (▽) $M^{-\bullet}$, with ethylene buffer gas; ( ) $M^{-\bullet}$, with helium buffer gas; (○) Cl^-, with methane buffer gas.[68]

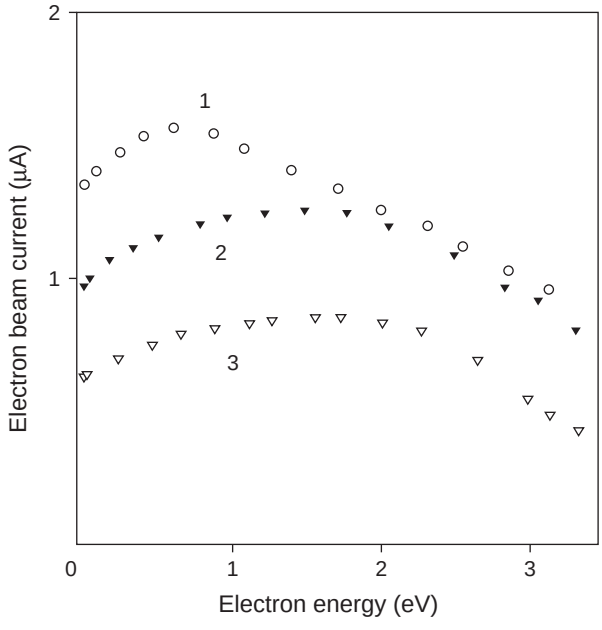

Figure 4 Ion current dependence on ion source pressure with carbon dioxide at: (1) 0 Pa; (2) 2.13 Pa (16 mTorr); (3) 5.46 (41 mTorr).[68]

reduces this variation to about ±22% at pressures of 2.13 and 5.46 Pa (16 and 41 mTorr). But carbon dioxide in the ion source also reduces the absolute current by 30% at 2.13 Pa (16 mTorr) and by another 40% when the pressure is increased to 5.46 Pa (40 mTorr). Carbon dioxide is among the more effective gases for inelastic scattering of electrons. Most carrier gases for gas chromatography, particularly helium, would not cause as great a change in the electron flux. In general, these are insignificant losses in the electron beam.

Thermalization of the electron energies is also likely when buffer gases are introduced. In carbon dioxide, for example, the electron thermalization rate constant k_e is 5×10^{-9} cc s^{-1}, which is competitive with quenching of the ion excess energy $k_q = 1 \times 10^{-10}$ cc s^{-1}.[69] If buffer gases are needed for stabilization of molecular ions, the potential problem of resonance peak shifts due to electron thermalization can be countered by calibrating the electron energy with the buffer gas present. The effect of high pressures on electron energies as well as on ion stabilization is potentially more serious when samples are introduced through the batch inlet, because the amount of sample vaporized into the ion source cannot be controlled and the pressure is likely to vary unpredictably.

4.2.4 Electron Energy Ramping

The capture of electrons by molecules with positive electron affinities is a resonance process, and different anions require different electron energies for maximum signal production. To obtain complete negative ion spectra of compounds, broad distributions of electron energies must be available. When samples are analyzed with EMMS the electron beam must be ramped continuously and rapidly over the entire electron energy range required to produce the mass spectra. In studies with hexafluorobenzene[70] the signal intensity was found to be 90% of the maximum value at a slew rate of 250 eV s^{-1} over a 10-eV electron energy range. At this rate 50 scans can be recorded of a 2-s-wide gas chromatographic effluent peak. The more difficult problem is to scan the mass range of the compound in this time-frame. There are perhaps several analyzers that could be used for this purpose, but the most practical will likely be the time-of-flight analyzer.

4.3 Analytical Information

4.3.1 Discrimination between $M^{-\bullet}$ and $^{13}C(M-H)^-$ by Resonance Electron Energies

Resonance electron energies can be used as an additional element of analytical resolution, this advantage has not been fully recognized in the past. For example, the mass-resolving power required to separate $M^{-\bullet}$ from $(M-H)^-$ containing one carbon-13 at mass 215 is 4.8×10^4 by conventional mass spectrometry. The herbicide atrazine (2-chloro-4-ethylamino-6-isopropylamino-*S*-triazine) has an ion with *m/z* 215 that could be due to either $M^{-\bullet}$ or $^{13}C(M-H)^-$. Two ε_{max} values of 0.38 and 1.78 eV appear in the electron energy scan of peak *m/z* 215.[61,62] The latter peak is due to the $^{13}C(M-H)^-$ ion because the instrument also transmits a peak at *m/z* 214 due to the $^{12}C(M-H)^-$ ion with the same ε_{max} value.

Previous reports attributing energy scan peaks with ε_{max} values of 3.8 and 7.2 eV to molecular ion resonances for nitrobenzene[61,71] are, in fact, due to the $^{13}C(M-H)^-$ ion.[72] The peaks with maxima around 7 eV reported for molecular ions of polyaromatic hydrocarbons[61] may also be due to the $^{13}C(M-H)^-$ ion.

4.3.2 Two- and Three-dimensional Electron Energy-mass Spectra

NIRs for all resonance and dissociative electron capture processes can be plotted as a function of electron energy versus the ion current.[73] Such plots represent the total electron capture cross-section or probability of electron capture $[P(e)]$ at any given electron energy (e) for the compound under study. Two- and three-dimensional presentations of the data[73] readily show distinguishing spectral features for each compound as shown here for *o*-butylbenzyl phthalate (Figure 5) and *o*-diethyl phthalate (Figure 6). Overlapping resonances can be deconvoluted with standard non-linear curve-fitting algorithms. Automated scanning of both the mass/charge ratio and the electron energies has not been achieved, but it should be possible using ion storage devices or time-of-flight mass analyzers.

4.3.3 Isotope Effects

Apparent chloride ion isotope effects have been reported.[74] More recently, the analysis of chlorine-37-enriched compounds 2[^{37}Cl],3- and 1[^{37}Cl],4-dichlorodibenzo-*p*-dioxin by standard ECNIMS methods using methane buffer gas was found to result in 1–3% greater loss of the lighter isotope.[75] Using EMMS the loss of chloride ion from each of the positions in the two compounds was found to be nearly equally probable.[75,76] Lower chlorine-37 ion yields in ECNIMS experiments may be due to a lower diffusion coefficient of the heavier isotope in the presence of buffer gas, which gives the ion more time in the ion source to undergo ion–ion or ion–wall reactions.[62]

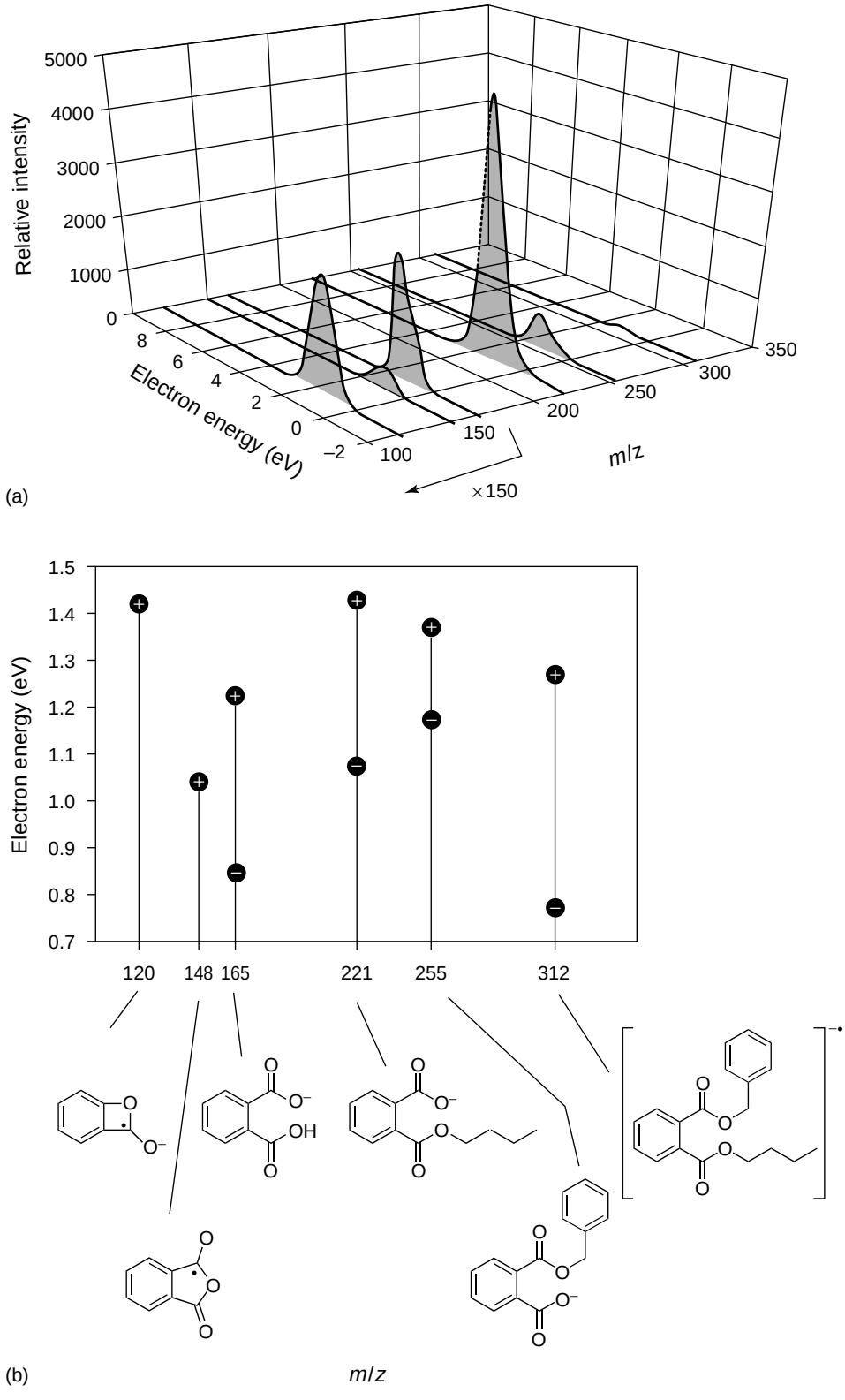

Figure 5 Two- and three-dimensional plots of *o*-butylbenzyl phthalate electron capture negative ion spectra: (a) three-dimensional electron energy versus ion mass/charge versus ion relative intensity; (b) two-dimensional electron energy versus ion mass/charge spectrum.[77]

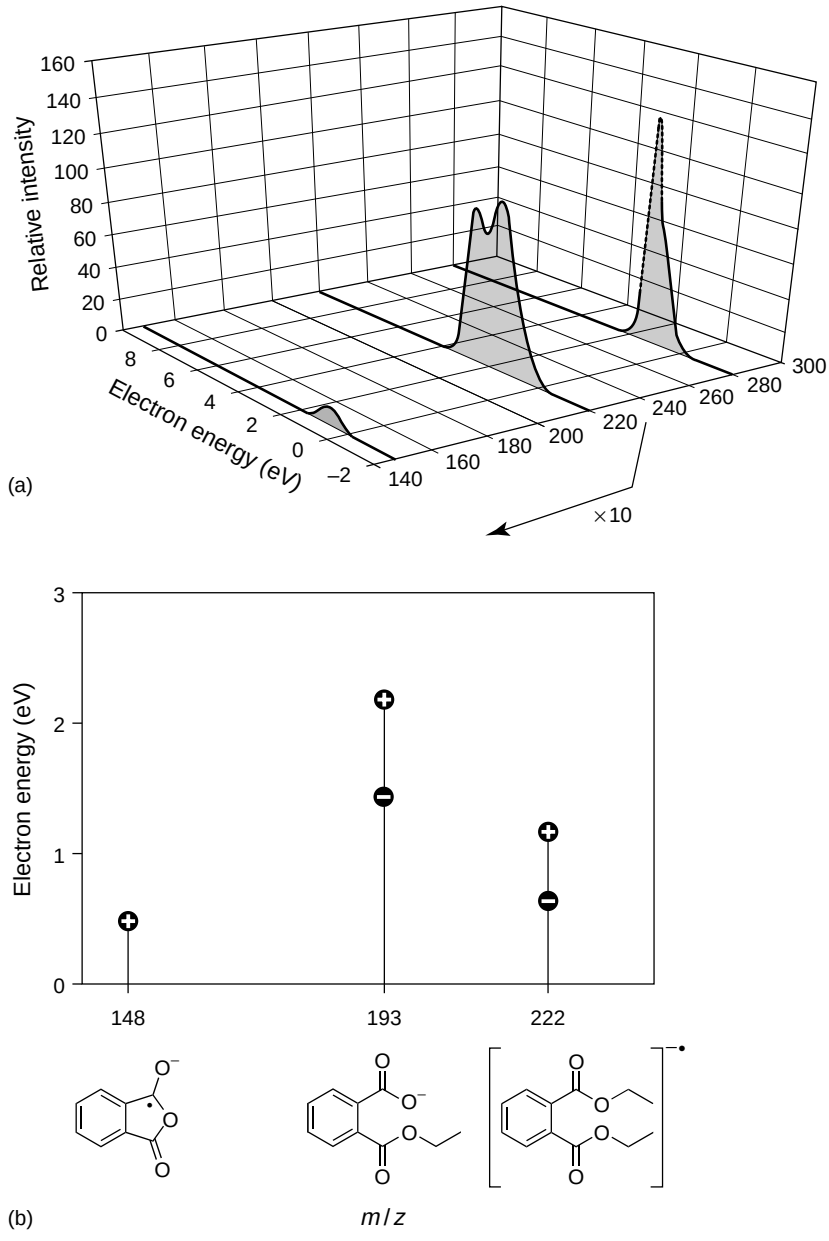

Figure 6 Two- and three-dimensional plots of *o*-diethyl phthalate electron capture negative ion spectra: (a) three-dimensional electron energy versus ion mass/charge versus ion relative intensity; (b) two-dimensional electron energy versus ion mass/charge spectrum.[77]

4.3.4 Electron Energy-dependent Regioselectivity

The EMMS system is ideally suited to study the regiochemistry of dissociating anions with respect to their electron attachment energies. Regiospecific labeling of compounds with stable isotopes and mass selection of the appropriate isotope provides a method for studying the energetics of regiochemical dissociation, as well as a potentially powerful method to follow the fate of compounds in metabolism and bioremediation studies.

The regioselective loss of the nitro groups from *p*-$^{15}NO_2$-labeled trinitrotoluene (TNT)[70] occurs with electrons of 0.15 eV for the *p*-nitro group, whereas the loss of the *o*-nitro groups occurs at higher electron energies. The analysis of regiospecifically chlorine-37-labeled polychlorodibenzo-*p*-dioxins (PCDDs) showed that the more highly substituted ring is the more likely one to lose a chloride ion (Table 2). This is not an unexpected result, because there is steric relief in the resulting aromatic radical and some stabilization from the remaining chlorines.

Table 2 Total chloride ion loss (%)[a] from labeled positions of regiospecifically labeled PCDDs at various electron energies

Chlorodibenzo-*p*-dioxin	0.03 eV	0.50 eV	1.00 eV	1.25 eV	2.00 eV
1,2[^{37}Cl]-Di-	56.9 ± 1.1	60.4 ± 0.4	58.2 ± 0.8		58.0 ± 3.6
1[^{37}Cl]3-Di-	51.1 ± 2.2	50.5 ± 1.3	48.2 ± 1.9	48.2 ± 2.0	50.2 ± 2.2
1[^{37}Cl]4-Di-	50.3 ± 1.1	50.4 ± 0.7	50.3 ± 1.0		51.3 ± 4.3
2[^{37}Cl]3-Di-	50.1 ± 1.2	50.0 ± 0.8	50.4 ± 1.0		48.7 ± 1.8
1[^{37}Cl]7,8-Tri-	31.2 ± 1.4	31.3 ± 0.6	29.4 ± 0.7		28.2 ± 1.2
2[^{37}Cl]7,8-Tri-	23.7 ± 2.1	21.6 ± 2.1	21.0 ± 1.8		21.2 ± 4.5
1,2,3[^{37}Cl]-Tri-	41.9 ± 2.6	43.5 ± 0.5	45.8 ± 0.7		50.9 ± 1.8
1[^{37}Cl]2,3,4-Tetra-	19.1 ± 1.3	19.0 ± 1.2	19.8 ± 1.4		27.5 ± 1.5
1[^{37}Cl]2,3,7,8-Penta-	43.8 ± 1.3	35.3 ± 1.6	32.2 ± 3.0		
1,2[^{37}Cl]3,7,8-Penta-	20.9 ± 1.8	22.2 ± 1.8	22.2 ± 2.9		
1,2,3[^{37}Cl]7,8-Penta-	21.5 ± 1.6	19.8 ± 1.6	27.5 ± 2.4	32.9 ± 3.3	
1[^{37}Cl]2,3,4,7,8-Hexa-	33.5 ± 0.8	32.2 ± 0.7	25.1 ± 2.2	25.9 ± 2.2	
1,2[^{37}Cl]3,4,7,8-Hexa-	13.7 ± 0.8	12.8 ± 1.4	14.1 ± 1.4		
1,2,3,4,7[^{37}Cl]8-Hexa-	2.2 ± 1.1	5.8 ± 1.6	7.6 ± 1.9	9.9 ± 1.9	

[a] The method of calculation takes into account the natural abundance of the chlorine isotopes at the unlabeled positions.

These studies also showed an electron energy dependence for the regioselective loss of the chloride ion under dissociative electron capture conditions. The best examples for this regiochemical dependence are provided by the data for 1,2,3,7,8-pentachloro- and 1,2,3,4,7,8-hexachlorodioxins, both of which show the highest loss of chlorine from the 1-position with electrons of energy 0.03 eV and less chlorine loss from this position with electrons of higher energies. For the pentachlorodioxin isomer, more chlorine at the 3-position is lost at higher electron energies, whereas more chlorine is lost from the 7-position of the hexachlorodioxin isomer. In almost every example the chlorines at the 2-position are least sensitive to changes in the energies of the ionizing electrons. The explanation for the regiochemistry lies in the relative magnitude of the antibonding orbital densities in the electronic states where the electron becomes localized.[76] However, an analysis of transition-state models of the transient radical anion is needed for any reliable explanations. An analysis of the transition states by open-shell calculations would be most helpful, but it is difficult to find sufficiently flexible basis sets in which the wavefunction does not degenerate to the neutral molecule and free electron.[34]

Studies with regiospecifically labeled compounds by standard ECNIMS methods showed that regioselective chloride ion loss closely matched those results observed at 0.03 eV with EMMS.[62,76] The results were consistent with expectations for the interaction of a high population of thermal electrons under reagent gas thermalization conditions. Comparisons of this type are potentially useful for developing an understanding of the ionization process under standard ECNIMS conditions.

4.4 Electron Capture Negative Ion Mass Spectrometry of Compound Classes

4.4.1 Pesticides

4.4.1.1 Organophosphates Spectra of phosphorothioates produced with EMMS are generally rich in fragment ions but the spectra of phosphates are much less so.[73,77] A comparison of the experimental electron attachment energies for phosphorothioate molecular ions with those obtained by a linear scaling relationship based on calculated LUMO energies shows surprisingly good agreement (Table 3).

A common fragment observed for the phosphorothioates listed in Table 3 is the loss of an ethyl radical (Scheme 1).[77] The production of $(M - Et)^-$ ions occurs at electron energies that are typically higher than those that produce the molecular ions, and most phosphorothioates have two resonance states from which these ions are formed. The current view (see discussion in section 4.5) for the formation of these ions is that the ionizing electron is captured into a π-antibonding orbital of the benzene ring and, through mixing of the π^* state with a σ^* state involving a phosphate oxygen, the ethyl radical cleaves. Only through this π^*–σ^* interaction can such groups be cleaved by a symmetry-allowed process.

Phosphorothioates exhibit the formation of both phenolic and thiophenolic fragments under electron capture conditions in the EMMS system. These ions are frequently produced from several resonance states over a relatively broad range (0.3–9 eV) of energies. The thiophenolic fragments must be formed after a rearrangement, perhaps through a spiro intermediate, whereas the production of the phenoxide ion could

Table 3 Observed and predicted electron attachment energies (*AE*) based on the LUMO energies calculated by the B3LYP/6-31G*//HF/6-31G* level of theory for *O*,*O*-diethyl *O*-4(X) phenylphosphorothioates

X =	Calculated LUMO energy (ε_{v_0})	Predicted *AE* (eV)[a]	Observed *AE* (eV)[b]
NO_2	−2.23271	0.15	0.24
CN	−1.36902	0.40	0.31
Br	−0.57797	0.63	0.66
Cl	−0.58233	0.63	0.67
F	−0.50097	0.65	0.69
Me	−0.19756	0.74	0.79
OMe	−0.29116	0.71	0.72
H	−0.22259	0.73	0.71

[a] Based on the scaling relationship $AE_{\text{Calc.}} = 0.288\varepsilon_{v_0} + 0.997$.

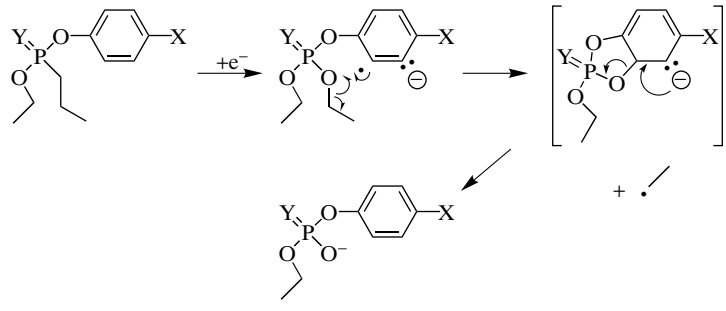

Scheme 1

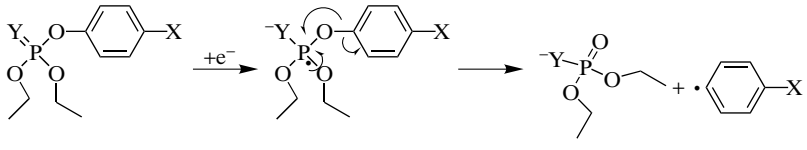

Scheme 2

occur via a simple cleavage between the phosphorus and oxygen atoms.[73] The electron energies for the formation of these ions are different, hence, the mechanisms for their formation must be different as well. A simple cleavage reaction that produces phenoxide ions, in contrast to the rearrangement mechanism that forms thiophenoxide ions, is supported by the observation that the former occurs with electrons of higher energies.[73,77] Simple cleavage reactions generally are higher energy processes.

Phosphate and thiophosphate anions are frequently the most intense ions in the EMMS spectra of organophosphorothioates. These ions are produced over a relatively broad range (0.63–4.91 eV) of electron energies. The mechanism of production of these ions may be different from those that produce the $(M - Et)^-$, thiophenoxide and phenoxide ions. The association of the electron with the π^*-orbital of the phosphate group and π^*–σ^* mixing that allows cleavage to produce the phosphate group and phenyl radicals (Scheme 2) is supported by ab initio calculations.[77]

A number of other common ions are observed in this series of compounds, including $(M-Et-X)^-$, $(M-OEt)^-$, and $(M-OEt-X)^-$ ions, many of which are of lower intensities. Some unusual ions observed for organophosphorothioates containing a *p*-nitro group are $C_3H_2NO^-$ and $C_5H_4NO^-$ ions.[73] The relatively large numbers of ions produced, as well as the aromatic cleavages observed for the nitro compounds, may be due to the close spacing of many low-lying orbitals, resulting in a large amount of orbital mixing that gives rise to complex fragmentations.[27,29]

4.4.1.2 Carbamates Three common negative ions have been observed for all the carbamates studied.[62] The molecular ions generally arise from one resonance state, but may involve two states (Table 4). There are no obvious characteristics about these molecules that

Table 4 Electron energies, ε_{max}, necessary for ion formation from carbamate pesticides[a]

Carbamate	Chemical name	$\varepsilon_{(M^{-\cdot})}$ (eV)	$\varepsilon_{(74)^-}$ (eV)	$\varepsilon_{(M-58)^-}$ (eV)
Oxamyl	Methyl *N′*, *N′*-dimethyl-*N*-{(methylcarbamoyl)oxy}-1-thiooxamidate	0.22	0.25	0.29, 2.79, 6.16
Carbaryl	1-Naphthyl-*N*-methylcarbamate	0.43, 1.78	0.41, 3.06	1.66, 2.74, 6.45
Aldicarb	2-Methyl-2-(methylthio)propionaldehyde *O*-(methylcarbamoyl)oxime	0.46	0.61	0.69
Methiocarb	3,5-Dimethyl-4-(methylthio)phenyl methylcarbamate	0.21	0.70, 3.49	1.30, 2.77, 5.91
Methomyl	*S*-Methyl-*N*-{(methylcarbamomoyl)oxy}thioacetimidate	0.43, 2.03	0.90	0.93, 3.08, 5.90
Terbutol	2,6-Di-*tert*-butyl-*p*-tolylmethylcarbamate	0.25	0.93	0.22, 2.47
Carbofuran	2,3-Dihydro-2,2-dimethyl-7-benzofuranyl methylcarbamate	0.23, 1.99	1.15	0.53, 3.08
Ethiofencarb	2-(Ethylthiomethyl)phenyl methylcarbamate	0.34	1.17	0.42, 2.44, 5.62
Dioxacarb	2-(1,3-Dioxolan-2-yl)phenyl-*N*-methylcarbamate	0.29	1.18	0.92, 2.80
Promecarb	3-Methyl-5-isopropylphenyl-*N*-methylcarbamate	0.24, 1.86	1.18	1.82, 2.84, 6.00
Baycarb	2-(1-Methylpropyl)phenyl methylcarbamate	0.64	1.24	1.35, 2.68, 6.07
Aminocarb	4-(Dimethylamino)-3-methylphenol methylcarbamate	0.29	1.31	1.66, 2.88, 6.24
MIPC	2-(1-Methylethyl)phenyl methylcarbamate	0.29, 1.75	1.45	1.25, 2.93, 6.14
Propoxur	2-(1-Methylethoxy)phenol methylcarbamate	0.10, 1.59	1.57	1.74, 2.78, 5.71
Mexacarbate	4-Dimethylamino-3,5-xylyl-*N*-methylcarbamate	0.64	1.72	1.38, 2.77, 5.91

[a] Data from Laramée et al.[62]

would distinguish those that form molecular ions from one state versus those that form the ions from two states.

A common carbamate anion (*m*/*z* 74) and an $(M - 58)^-$ ion that arises through loss of the acylamido radical $(CH_3NHC{=}O)^{\bullet}$ from the parent molecule make up the important dissociative attachment ions. The energies for these processes are different, but the energies for the first NIRs tend to follow each other, i.e. carbamates with higher $\varepsilon_{(74)^-}$ also have higher $\varepsilon_{(M-58)^-}$ values. In this series the carbamate ions (*m*/*z* 74) generally arise from only one NIR, whereas the $(M - 58)^-$ ions result from several states. Multiple resonances are generally associated with ions that have a larger number of channels for deposition of the excess energy.[27,29]

4.4.1.3 sym-*Triazine Herbicides* Numerous adduct ions, including $(M + 1)^-$, $(M + 2)^{-\bullet}$, $(M + 14)^{-\bullet}$ and $(M + 28)^{-\bullet}$ and, for compounds containing chlorine, $(M + 35)^-$, are reported to have significant intensities in standard ECNIMS spectra of *S*-triazine herbicides.[78] The abundances of these ions vary widely depending on the source temperature, pressure and source cleanliness. There are no adduct ions when these compounds are analyzed with the EMMS system. Although not many *S*-triazines have been analyzed, the information that is available suggests that there may be considerable differences in the number of NIRs that produce specific fragment ions. The molecular ions for atrazine (2-chloro-4-ethylamino-6-isopropylamino-*S*-triazine) and ametryne (2-(ethylamino)-4-isopropylamino-6-methyl-*S*-triazine), for example, arise from two states of similar energies, but the $(M - H)^-$ ion of the former arises from one state (ε_{max} 1.97 eV) and that of the latter arises from four states in the electron energy range 0.35–9.20 eV. Such diverse electron energy scan patterns could be useful for distinguishing these compounds.

4.4.2 Environmental Compounds

Standard ECNIMS spectra of environmental compounds using reagent gases for electron thermalization have been published elsewhere.[79] The negative ion spectra include those of haloalkanes, halo- and nitro-aromatics, PCDDs, polychlorodibenzofurans, polychorobiphenyls, alkyl phthalates and many others, and the interested reader is referred to this excellent source.[79] The following discussions are limited to those compounds for which data have been obtained using EMMS.

4.4.2.1 Polychlorodibenzo-p-*dioxins* The electron capture negative ion mass spectra recorded with EMMS for PCDDs show only a few peaks.[62,75,76] Higher chlorinated PCDDs show a single resonance for the molecular ion peak in the electron energy range 0.12–0.20 eV. Dichlorodibenzo-*p*-dioxins and some trichlorodibenzo-*p*-dioxins show no molecular ions. When the electron energy is scanned over the range 0–10 eV with the mass set to record the chloride ion, two broad peaks are observed for all PCDDs with maxima around 0 and 4 eV (Figure 7). The low-energy peak appears to be a composite of at least two peaks.

4.4.2.2 Polychlorobiphenyls PCBs have been analyzed with GC/EMMS by monitoring the chloride ion peak produced with electrons of 0.03 eV energy.[62] The lower chlorinated PCBs show no molecular ions when analyzed with EMMS. Some pentachloro-[77] and even

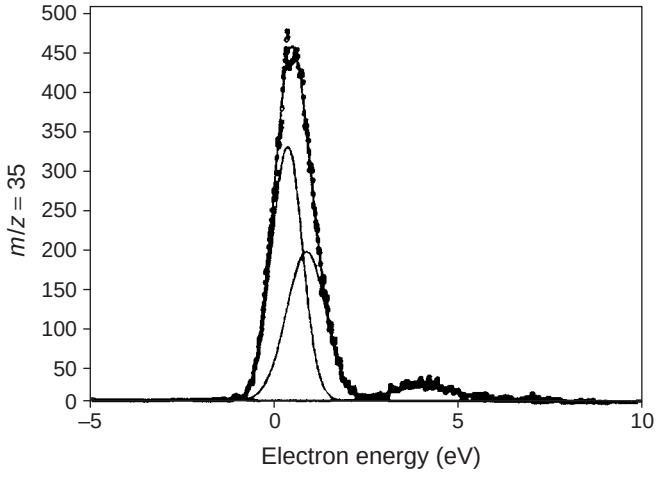

Figure 7 Electron energy scan for Cl^- loss from 1,7,8-trichlorodibenzo-*p*-dioxin under electron capture negative ion conditions using the EMMS system.

tetrachoro-PCB isomers[62,77] may show molecular ions, but generally at least six chlorines are necessary on the molecule before a molecular ion is observed. It is also of interest that 3,3′,5,5′-tetrachloro-PCB displays a molecular ion (ε_{max} 0.44 eV), whereas PCBs with chlorine on the ortho positions generally do not. Molecular orbital (MO) calculations at the PM3 level show that the benzene rings of the 3,3′,5,5′-isomer are coplanar with a low-lying π-orbital throughout the entire molecule.[62,77] The PCB isomers with chlorines in the ortho position act more like two independent benzene rings and require more chlorines in either ring for the molecular ion to be observed.

The PCBs substituted in the ortho positions produce mainly chloride ions in EMMS, with electron energies for the first negative ion states ranging from 0.34 to 1.07 eV[77] for the ortho-substituted PCBs. In addition to the molecular and chloride ions, PCBs may also produce $(M-Cl)^-$, $(M-Cl_2)^{-\bullet}$ and $(M-H)^{-\bullet}$.[62,79]

4.4.2.3 Chloronaphthalenes Few electron energy scans have been recorded for chloronaphthalenes, but the potential for isomer distinction is suggested by the different profiles for the molecular ions from 1,4,6,8- and 1,4,7,8-tetrachloronaphthalenes. Broad and nonsymmetrical[62] peaks are observed with maxima around 0.4 and 0.6 eV.

4.4.3 Organofluorine Compounds

4.4.3.1 Fluoroarenes Fluorinated compounds have been studied extensively with the EMMS system.[27,28,31,60,80,81] The behavior of fluorinated compounds is not always predictable. Some distinguishing features in the spectra of fluorinated compounds include the general absence of molecular ions, the much higher than average electron energies required for production of the dissociative capture ions, and the very low cross-sections for dissociative electron capture in comparison to other compounds. Only hexafluorobenzene among the fluorinated aromatic compounds forms a long-lived detectable molecular ion,[81] and this ion is formed with electrons of very narrow energy spread near 0 eV. Isomeric fluorobenzenes undergo dissociative capture with electrons of energies in the range 3–10 eV. Most states lie between 8 and 10 eV.[81] The fluoride ion is observed mainly in this region, whereas the $(M-H)^-$ ion is formed in the energy range 2–4 eV. Many of the isomers yield fluoride ions by ion pair formation, as shown by the continuous increase in product ion with increasing energies above 14–16 eV.

The NIRs for dissociative electron attachment of isomeric di-, tri-, and tetrafluorobenzenes are distinctly different for each isomer. Thus, 1,2,4,5-tetrafluorobenzene has resonances for F^-, $C_6HF_2^-$, and $C_6HF_4^-$ that are as much as 1 eV lower than those required for the production of ions from 1,2,3,5-tetrafluorobenzene.[81] The symmetrical isomer 1,3,5-trifluorobenzene, in comparison to 1,2,4-trifluorobenzene, shows a simpler mass/electron energy map and the resonance peaks are more symmetrical.

4.4.3.2 Unsaturated Cyclic and Alicyclic Fluoro Compounds The negative ion spectra of the structural isomers hexafluoro-1,3-butadiene, hexafluoro-2-butyne and hexafluorocyclobutene show significant differences in the abundance of various ions in the energy range 0–14 eV. Most of the ions above 2 eV are weak in intensity, as expected, but single peaks for the molecular ions, CF_3^- and C_3F^-, and the clearly resolved peaks for fluoride ion and $C_3F_3^-$ from several different negative ion states were recorded over the entire energy range.[80] In the lower electron energy region, hexafluoro-2-butyne and hexafluorocyclobutene show molecular ions but hexafluoro-1,3-butadiene does not. On the other hand, hexafluoro-1,3-butadiene produces an intense fluoride ion at 2 eV, a weaker one at 1 eV and a $C_4F_4^-$ ion at 1 eV, whereas hexafluoro-2-butyne produces only a weak fluoride ion at 2 eV and hexafluorocyclobutene produces neither ions. These results point out that the three structural isomers might possibly be differentiated by their electron energy scans.

Electron transmission spectra of the fluoroethylene series show that increasing fluorine substitution results in higher electron attachment energies, indicating that there is increased destabilization of the π^* states.[82] Solution-phase reduction indicates just the reverse, i.e. that it is easier to reduce the more highly

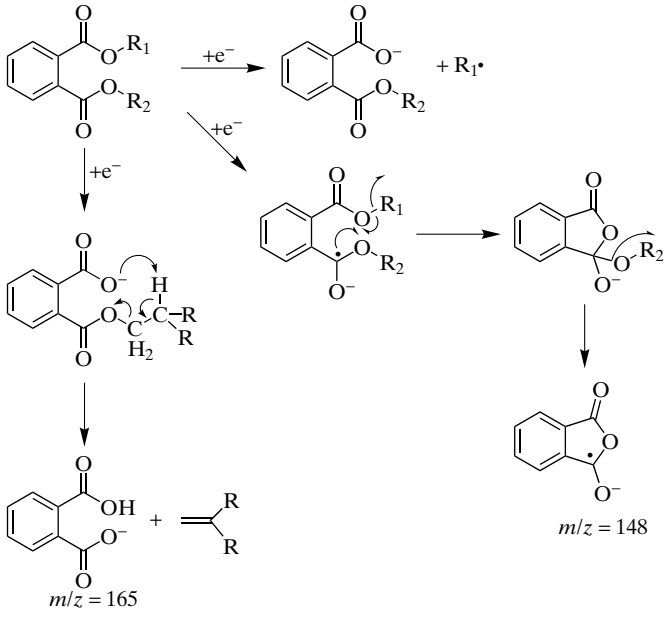

Scheme 3

fluorinated compounds. The results are rationalized on the basis that solution-phase reduction involves the geometries of the relaxed anion, whereas gas-phase electron–molecule interactions involve neutral molecule geometries.

4.4.3.3 Chlorofluoroalkanes Chlorofluoromethanes yield a series of negative ions over the electron energy range 0–6 eV.[60] The common ions from CCl_3F, CCl_2F_2 and $CClF_3$ are F^-, Cl^- and ClF^-. Chloride ion is produced in greatest relative abundance from dissociative electron attachment, even where statistical considerations would dictate otherwise. For example, the relative intensities of the Cl^- and F^- peaks from CCl_3F, CCl_2F_2, and $CClF_3$ are in the ratio 200 : 7.25 : 5. According to semi-empirical modified intermediate neglect of differential overlap (MINDO) calculations, the lowest σ^* MO involves a C–Cl bond.[29] The bond order for C–Cl decreases with the accommodation of the extra electron, which leads to a direct electronic dissociation to give the chloride ion. On the other hand, the bond order for C–F does not decrease upon capture of a low-energy electron. Typically in the case of mixed halogens, the electron goes to the σ-orbital and associates with the halogen having the greater "electron capacity",[27,29] which increases on going down the periodic table. This accounts for the tendency of these molecules to undergo dissociative electron capture with loss of a fluorine rather than a chlorine to give $CClF_2^-$, CCl_2F^- and CCl_3^- from the respective parent molecules. Chlorofluoromethanes with relatively more fluorines have higher appearance energies for production of the ions.

4.4.4 Organic Chemicals

4.4.4.1 Phthalates The interaction of low-energy electrons with phthalic anhydride[83] and arylphthalides[84,85] has been reported. More recently, a study of one *meta*- and seven *ortho*-phthalates was conducted in detail using GC/EMMS.[86] Two NIRs are generally observed for each ion, with the molecular radical anion maxima appearing at electron energies ~0.6 and ~1.1 eV. Higher energy electrons produce $(M - R_1)^-$, $(M - R_1 - R_2 + H)^-$ and phthalic anhydride ions (Scheme 3). These fragmentation pathways almost always are associated with their own unique resonances. In several instances the fragmentation involves the same ε_{max} value as that which produces the molecular ion. This indicates metastable ion decomposition. *m*-Dimethyl phthalate yields a molecular ion with electrons of energies 0.19 and 0.87 eV, but no fragment ions.

4.4.4.2 Sulfonamides Sulfonamide antibiotics are a class of compounds structurally related to sulfanilamide. These compounds are used extensively in pharmacy and agriculture as antimicrobial agents. Molecular ions are observed for the largest of the sulfonamides, i.e. sulfabenzamide, sulfamethazine, and sulfisomidine. The major negative ion product from all underivatized sulfonamides in spectra taken using EMMS is the 4-aminobenzene sulfonate anion.[86] Standard ECNIMS experiments, on the other hand, produce two major ions: the sulfonate ion, *m/z* 156; and an ion with one less hydrogen, *m/z* 155. Apparently, the reagent gas participates in the abstraction of a hydrogen atom from

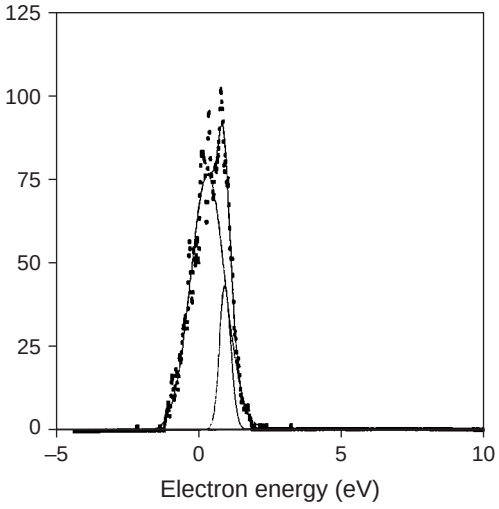

Figure 8 Electron energy scan of the *m/z* 156 fragment ion of sulfabenzamide.[77]

the parent compound or from the benzene sulfonate ion. Deconvolution of the main electron energy scan peak (Figure 8) suggests that there are two resonances responsible for the formation of the sulfonate anion from sulfonamides (Table 5). The electron energy required to form this ion lies in the range 0.15–0.70 eV for the first peak and 0.70–1.1 eV for the second peak.

The effects of electropositive and electronegative derivatizing groups on the negative ion spectra are clearly shown in this series of compounds. Silylation of sulfonamides results in fewer ions in the spectra. Only four of the derivatives show any ions at all (Table 5) and the spectra are generally weak when compared with those of the underivatized compound. Weak spectra are also observed under standard ECNIMS conditions. Trifluoroacetylation of the analinium nitrogen, on the other hand, yields derivatives that give much more intense signals under both standard ECNIMS conditions or when recorded with the EMMS system. The formation of fluorinated derivatives frequently gives good signals but no distinguishing features in the spectra of the different compounds, because the strongest peak in the spectrum is due to the derivatizing group anion. The most intense peaks in the EMMS spectra of trifluoroacetamide derivatives of sulfonamides, however, result from cleavage of the sulfonamide bond to give the 4-trifluoroacetylaminosulfonate anion (*m/z* 252). Deconvolution of the broad electron energy scan peak indicates that there are two resonances – one around 0.1 eV and another closer to 1 eV – for nearly all the trifluoroacetylated sulfamides (Table 5). Very weak resonance peaks can also be discerned in the range 3–4 eV for most of these derivatives.

4.4.4.3 Chloroethylenes The most abundant ion produced from chloroethanes and chloroethylenes under ECNIMS conditions is the chloride ion.[33] Other ions formed are $(M-Cl)^-$ and Cl_2^-. The electron energy plots for chloroethylenes show some interesting patterns in the resonance energies. For example, chloride ion loss generally occurs at higher resonance energies when chlorine substitution is lower (Table 6). This pattern is just the opposite to that observed for the ETS spectra of fluoroethylenes,[82] assuming that fluoride ion is the product from interaction of the electron with the parent molecule. There are too few examples for a general conclusion about geometrical isomerism, but geminal

Table 5 Electron attachment energies (*AE*) for fragment ions from sulfonamides and their derivatives under electron capture conditions in an EMMS system

Name	Parent *m/z*	*AE* (eV)	Silylated derivative *m/z*	*AE* (eV)	Trifluoroacetyl derivative *m/z*	*AE* (eV)
Sulfabenzamide	156	0.23, 0.85	155	0.62	252	0.13, 0.88, 4.72
	276	0.68	270	0.38, 0.76		
	199	0.48, 0.95				
Sulfacetamide	156	0.15, 0.70	155	0.80, 0.60	252	0.03, 0.81, 1.91
	199	0.03, 0.62	270	0.03, 0.69		
Sulfamethoxypyridazine	156	0.29, 0.77			252	0.12, 1.05, 3.18
Sulfameter	156	0.40, 0.91			252	0.05, 0.52, 3.52
	46	0.39			221	0.03, 0.91, 2.45
Sulfadiazene	156	0.30, 0.97	155	0.14, 0.51	252	0.12, 0.77, 3.58
			270	0.18, 0.55		
Sulfamethazine	156	0.63, 1.11			252	0.09, 0.62, 3.03
	198	0.10, 0.98				
	278	0.50				
Sulfaisomidine	156	0.47, 1.04	155	0.24, 0.51	252	0.12, 1.01, 3.6
	278	0.03	270	0.09, 0.43		
Sulfaisoxazole	156	0.32, 0.79			252	0.29, 2.42
Sulfamethoxazole	156	0.31, 0.88			252	0.31, 2.65

Table 6 Electron energy maxima for the production of chloride ion from chloroethylenes

Chloroethylene	$\varepsilon_{max}(i)$[a] (eV)
Tetrachloroethylene	0.19(24), 0.53
Trichloroethylene	0.19(19), 0.75
1,1-Dichloroethylene	0.19(10), 0.75
cis-1,2-Dichloroethylene	1.35
trans-1,2-Dichloroethylene	0.95

[a] Relative chloride ion current in arbitrary units.

substitution of chlorines (Table 6) seems to result in a common resonance appearing around 0.19 eV.

4.4.4.4 Other studies An interesting mechanistic study involving methanol, deuteriomethanol, and allyl alcohol was undertaken with EMMS to examine hydrogen scrambling during dissociative electron capture.[29] The calculated minimum heats of reaction (ΔH_0) for producing OH^-, O^- and CH_3O^- are 2.1, 2.4 and 2.9 eV respectively. Experimentally, the methoxide ion shows three resonance peaks above 5.5 eV, whereas hydroxide and oxide ions show only a single resonance peak with $\varepsilon_{max} = 10.5$ eV. Labeling studies show scrambling between OH^- and OD^-, but none for the production of methoxide ion, which indicates direct O–D bond cleavage. The MINDO results indicate high-lying orbitals with mixed σ^* states. These results have been interpreted in terms of a core-excited resonance at 10.5 eV to account for the hydrogen scrambling experiments. The OH^- and $C_3H_5O^-$ ions from allyl alcohol occur with $\varepsilon_{max} = 1.7$ eV. Calculations by the MINDO method indicate that a π^*-orbital is involved in the production of these anions. Ethylene captures electrons near 1.8 eV, which is about the same as that for allyl alcohol.

Resonance electron capture processes are important in a variety of other scientific areas. The electron capture detector used in gas chromatography, for example, depends very much on the gas-phase ion chemistry that can be elucidated by studies carried out with EMMS and vice versa. Thus, the pulsed electron capture detector and its hybrid, the photodetachement-modulated electron capture detector, are among the few tools available to study the relative rates of ion–molecule reactions, ion–ion recombination reactions, photodetachment reactions, thermal electron attachment and detachment reactions.[87] The domain of particle bombardment mass spectrometric methods is a potentially important area in which resonance electron capture processes may play a role. The analysis of dihydroxybenzenes[88] and the anticancer natural product paclitaxel,[89] for example, yield radical anion adducts with nitro-containing matrices, such as 3-nitrobenzyl alcohol, by fast atom bombardment mass spectrometry. The degree of radical anion formation from oligonucleotides[90] under liquid secondary ion mass spectrometry conditions depends on the chain length, and the major fragmentations in carbamoylmorpholine-[91] and carbamate-linked oligonucleotides[92] also are postulated to arise from radical anions resulting from the capture of electrons. In these cases electron-capturing matrices, such as 3-nitrobenzyl alcohol, quench the apparent electron capture reactions and only preformed $(M - H)^-$ ions are observed. Recent reports on the neutralization of electrosprayed protein/peptide cations and their reionization with thermal electrons[93,94] provide another example of the importance of the resonances of the electron capture process in areas outside of the traditional volatile chemical arena. In each of these examples, analysis of the compounds with EMMS should reveal some interesting relationships between the ions and the resonance energies required to produce them.

4.4.5 Derivatization Reactions

The ECNIMS methods can be extended to analyze compounds with active hydrogens by simply derivatizing them with electrophores. Derivatizing groups containing large numbers of fluorine atoms are very popular because of the excellent sensitivity that results when electrons are captured by these groups under ECNIMS conditions. For example, perfluoroacyl, pentafluorobenzyl, pentafluorobenzoyl, and (pentafluorophenyl)methanimine derivatives of chorophenols and choroanilines[95] produce analyte-specific anions under ECNIMS conditions. Hammett linear free energy relationships have been found useful for predicting a variety of properties, including the relative intensity of the molecular ion and the direction of fragmentation. 3,5-*Bis*-trifluoromethylbenzoyl derivatives of mono- and dihydroxy-steroids have been prepared to give improved sensitivity and chromatographic behavior in ECNIMS studies.[96] Derivatization of active hydrogens by electrophores is particularly important in biomedical applications[97,98] and the analysis of drugs.[99] Pentafluorobenzyl chloroformate, for example, has been used to derivatize amino acids in biological fluids for subsequent detection by ECNIMS in the femtomolar range.[100] DNA oxidative damage quantified via the pentafluorobenzyl-*tert*-butyldimethylsilyl derivative of 2-methylglycerate released from 2-thymidine glycol can be detected with just 41 amol injected onto the column by gas chromatography/electron capture negative ion mass spectrometry (GC/ECNIMS).[101] Pentafluorobenzylation of L-tryptophan and its acidic metabolites in studies on infectious and inflammatory diseases allows their detection in the femtogram range by GC/ECNIMS[102] and in the low picogram range by liquid chromatography/electron capture negative ion mass spectrometry (LC/ECNIMS) using particle beam transfer.[103]

The simplicity and effectiveness of derivatization reactions, together with the inherent specificity and sensitivity of ECNIMS methods, should become even more powerful when the analyses are performed with EMMS because of the extra analytical dimension available. The effectiveness of the analysis of sulfonamides after derivatization has already been noted (Table 5). The Schotten–Baumann reaction[104] involves a very simple procedure that can be used to derivatize the active hydrogens of hydroxyl, sulfhydryl, amino, and phosphonic acid groups (Equation 14) by shaking the compounds of interest with pentafluorobenzoyl chloride in aqueous media at room temperature.[105] Three-dimensional plots of these derivatives obtained with GC/EMMS are relatively simple but sufficiently different that they can be used for rapid screening of compounds that might be more difficult to analyze by standard EMMS methods (Figure 9).

$$\mathrm{RZH} + \mathrm{C_6F_5COCl} \longrightarrow \mathrm{C_6F_5COZR} + \mathrm{HCl}$$
$$\mathrm{ZH} = -\mathrm{OH}, -\mathrm{SH}, -\mathrm{NH}, -\mathrm{NH_2}, -\mathrm{P(O)OH} \quad (14)$$

Examples of other potentially important applications of these methods include the analysis of fatty acids, which are characteristic structural components of bacterial lipolylpolysaccharides, tell-tale long-chain fatty alcohols from mycobacterial infections, including tuberculosis, and small amino acid neurotransmitters. Dipicolinic acid is a chemical unique to bacterial spores that produces strong negative ion signals. The negative ion mass spectra of spores are, however, complicated by the presence of extraneous ions whose appearance is tied to the degree of sporulation in a given culture, and to slight changes in the ion source pressure and temperature. The appearance of spurious ions in ECNIMS is a common problem that is exacerbated by the introduction of whole bacteria into the mass spectrometer. These artifacts were eliminated when the electron monochromator was used to supply electrons with energies specific for electron capture by dipicolinic acid. Derivatized *Bacillus globigi* spores produced an intense peak of signal-to-noise greater than 20 corresponding to 300 spores.[106]

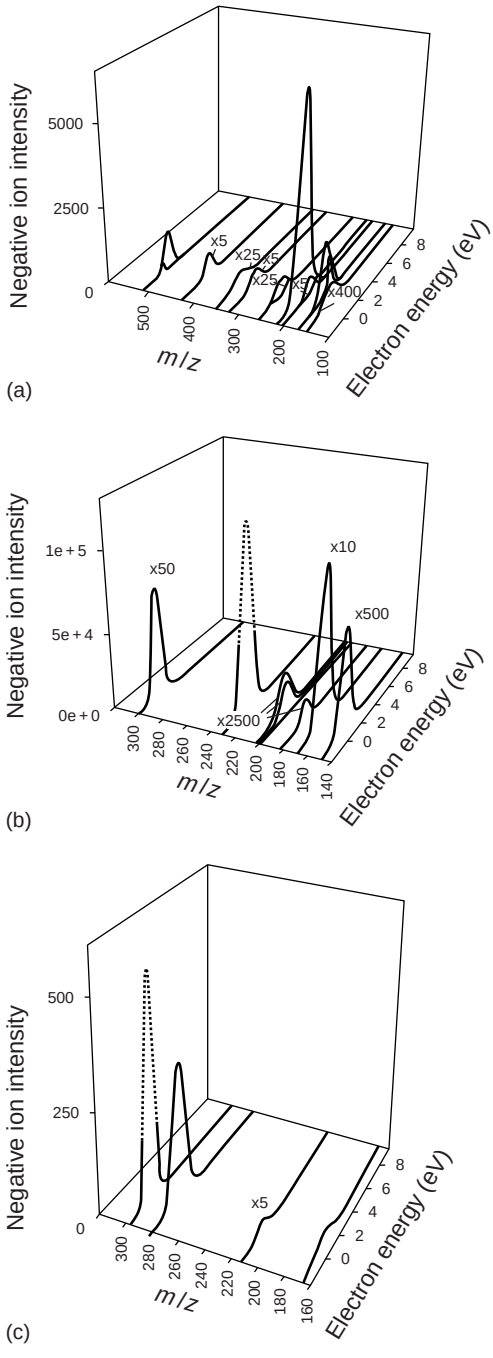

Figure 9 Pentafluorobenzoyl derivatives of: (a) 2,2-thiodiethanol; (b) *N*, *N*-dimethylamino-2-ethanthiol; (c) pinacolyl alcohol.

4.4.6 Explosives

An area that can benefit from using EMMS is explosives detection. The extra dimension of information provided by resonance electron energies should help greatly in the early detection of these compounds. In many cases, it may be possible to identify common explosives by simply scanning the electron energy for the nitro (*m*/*z* 46) anion (Figure 10).[62,107,108]

TNT possess two types of $C-NO_2$ bonds: a resonance peak for loss of the *p*-nitro group near 0 eV and a broad peak centered around 3.5 eV. The nitrate ester, pentaerythritol tetranitrate (PETN), contains an $O-NO_2$ bond and shows just a single resonance peak around 0 eV for loss of the nitro group. 1,3,5-Trinitro-1,3,5-triazocyclohexane (RDX) contains nitramine bonds ($N-NO_2$) and shows ε_{max} values of 0.3, 4.5, and 9.4 eV. The experimentally observed resonances for loss of the nitro groups in these three compounds are distinctive and they could easily be used to identify them even

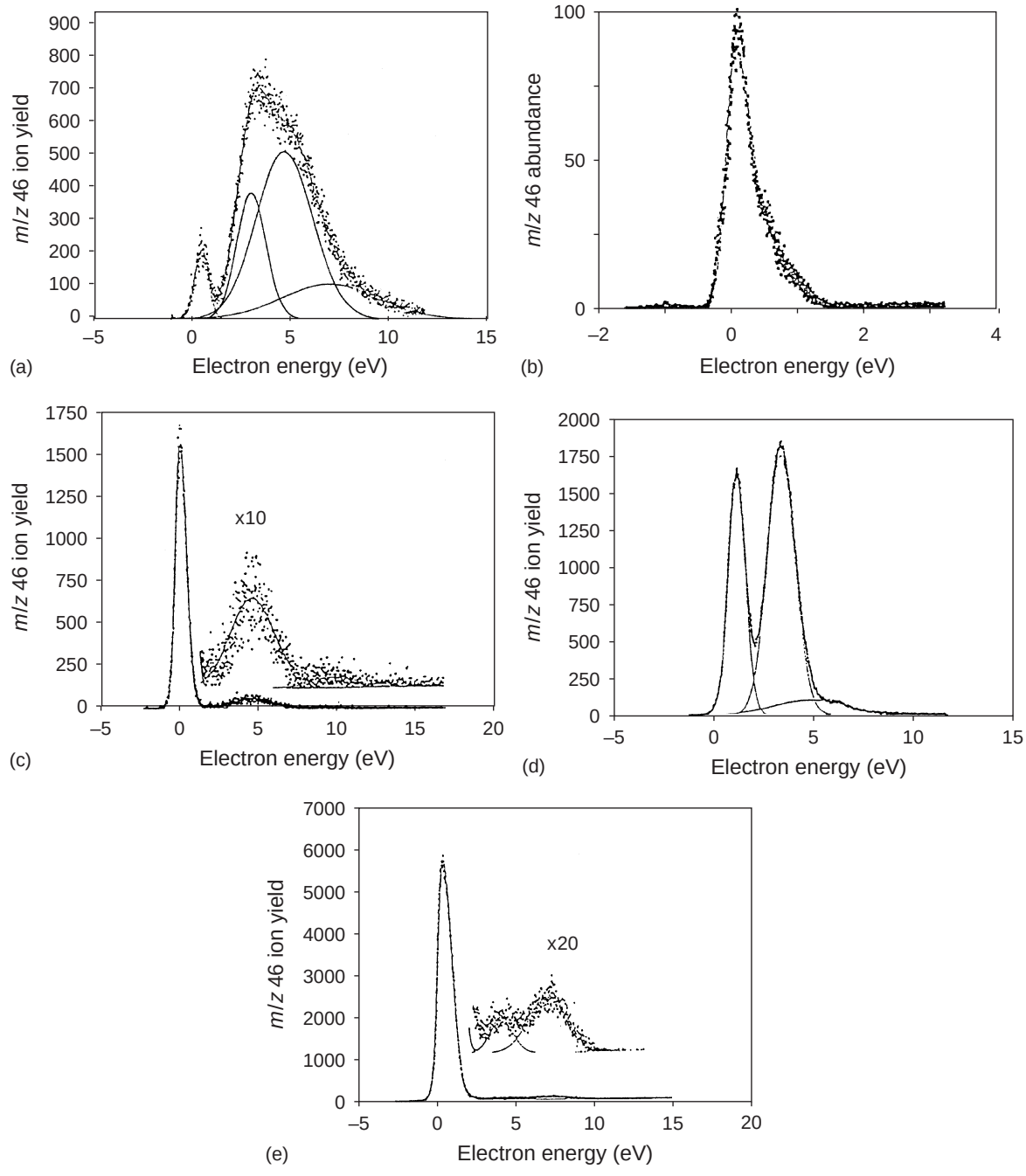

Figure 10 Electron energy scans of the nitro group anion of: (a) 2,4,6-trinitrotoluene (TNT); (b) pentaerythritol tetranitrate (PETN); (c) 1,3,5-trinitro-1,3,5-triazocyclohexane (RDX); (d) nitrobenzene; (e) nitropropane.[62]

without any other mass spectral information or gas chromatography retention times.

These explosives can also be distinguished from some common nitro-containing compounds on the basis of the electron energy scans. Nitrobenzene, a component of domestic products, for example, is easily distinguished from RDX, PETN, and TNT just by the nitro anion resonances around 1 and 4 eV. Nitrobenzene also shows several strong resonances for the phenyl anion at 3.6 and 6.0 eV[61] and one for the molecular ion around 0.06 eV.[61,72] 2-Nitropropane, a combustion product, has three observable resonances at 0.68, 4.56, and 7.88 eV. These values are sufficiently different from those of the explosives and especially of RDX that it is unlikely that they would be identified incorrectly. The peak shapes are also quite characteristic and could be used for

identification. If mass measurements are included, then positive identifications are greatly enhanced, because most nitro-containing compounds also show resonance peaks for the molecular ion and other fragments. Portable EMMS systems for explosives detection would require ''sniffing devices'' for collection and concentration of the vapors. When time is not the primary consideration, GC/EMMS could be used for the extra dimension and certainty of identification.

4.5 Mechanisms in Electron Capture Negative Ion Mass Spectrometry

The interpretation of mechanisms in ECNIMS has been largely based on models wherein a molecular radical anion is formed and, through classical electron pushing, fragmentations could be postulated. The distinction between fragmentation that occurs in the ion source and metastable decompositions that occur after the ion has been accelerated has been recognized from the outset.[5] Metastable ion decompositions may involve cleavages with or without molecular rearrangements or composite decompositions that involve both simple cleavage and rearrangement followed by cleavage.[5] Collision-induced dissociations can be used to enhance the fragmentation processes for parent ions with low internal energies. Mechanistic studies wherein decompositions take place from even-electron ions such as $(M-H)^-$ can be produced by secondary ion mass spectrometry or fast atom bombardment mass spectrometry[109] and used for understanding fragmentations apart from the electron capture process itself. Radical anions may also undergo competing ion/molecule reactions that include electron transfer, proton transfer, hydrogen atom abstraction, and nucleophilic substitution processes,[110] from which much can be learned about the relative acidities, electron affinities, reactivities, structures and whether such species behave more as anions or as radicals. A particularly interesting and useful process is the dissociative electron attachment to N_2O, which yields O^- ions that in turn react with organic compounds to yield anions, radicals or radical anions.[111,112]

The mechanisms involved in product formation in dissociative electron capture, however, may not be quite as simple as once thought, because the states that produce observable molecular ions are often, if not usually, different from those that produce the fragment ions. The production of stable molecular ions that are long-lived enough to observe in the mass spectrometer almost certainly requires the presence of π-orbitals, whereas fragmentation may involve either π- or σ-orbitals. In the following sections some details on mechanisms of resonance electron capture processes that can be extracted from combined studies involving MO calculations with high-level basis sets are described, along with data from an EMMS instrument.

4.5.1 Electronic States

There have been a number of studies showing linear relationships between calculated virtual orbital energies using Koopmans' theorem[113] and the experimental resonance electron attachment energies[31,114] or electron affinities[115] of organic molecules. Strong correlations between electron attachment energies for aromatic compounds obtained from ETS data and the virtual orbital energies calculated by Gaussian basis sets have been reported.[116] An extension of this approach with data for PCDDs was also highly successful.[76] Assuming neutral ground-state geometries for both the molecular ions and chloride ions, a correlation was found between π virtual orbital energies calculated by the B3LYP/D95//B3LYP/D95 level of theory and the ε_{max} values for the resonances of the PCDDs obtained with EMMS and those for substituted benzenes from ETS (Figure 11). These correlations suggest that all negative-ion-forming processes for this class of compounds result from the π^* state. The loss of the chloride ion directly from the π^* state is a symmetry-forbidden process. Hence, there must be $\pi^*-\sigma^*$ mixing in the transition state of the dissociating PCDD ion.

Through-space symmetry-allowed interactions between π^*- and σ^*- orbitals have been proposed for dissociative attachment of chloride ion from chloronorbornyl[117] and other chlorinated multicyclic ring[118] systems. The loss of chloride ions from planar molecules such as vinyl chloride or chlorobenzenes in which the chlorine is attached directly to the olefinic system have

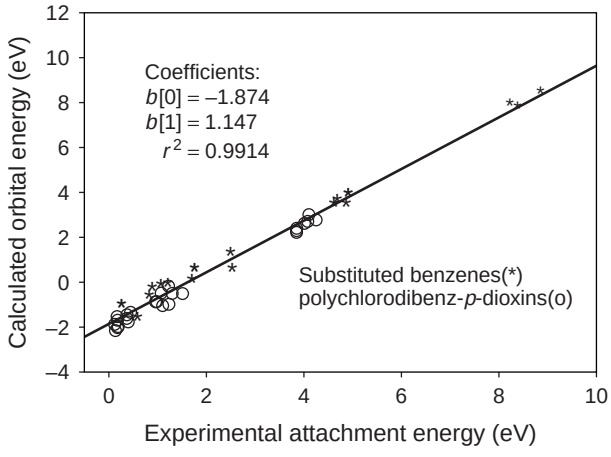

Figure 11 Correlation between calculated virtual orbital energies and experimental electron attachment energies for aromatic compounds. Aromatic compounds are benzonitrile, benzene, chlorobenzene, phenol, styrene and toluene. The dioxins are those shown in Table 2.[76]

been postulated to involve out-of-plane distortions[119] of the carbon–chlorine bond or a ''bent-bond'' model[120] in the transition state. Studies with halobenzyl halides have shown strong $\pi^*-\sigma^*$ mixing as the phenylic carbon–halogen bond stretches.[121] Thus, it is concluded that whenever an aromatic system is available, the dissociative attachment chemistry has a high probability of involving a π^* state.

The resonance for chloride ion loss from PCDDs around 4 eV (Figure 7) has a precedent in the halobenzyl halide series,[121] where resonances were observed around 4.5 eV. This peak is considerably more intense than the peaks resulting in phenylic chloride ion loss.[121] A diffuse totally symmetrical s-orbital of the parent singlet excited state (π, π^*) was proposed. This state is symmetrical with respect to the aromatic ring and a transition to the σ^* state is symmetry-allowed, thereby yielding a relatively intense peak.

4.5.2 Steric Requirements

The dissociative electron capture events that lead to the production of $(M-R)^-$ fragments from *o*-phthalates is driven by steric interactions that force the ester functionalities into noncoplanar orientations with the benzene ring.[86] As the alkyl substituents become larger, the electron energy necessary to promote fragmentation decreases. Semi-empirical calculations, in fact, show that the optimum geometry for *m*-dimethyl phthalate has both ester groups coplanar with the benzene ring, whereas the ester groups are displaced by an average of 85° and 11° from coplanarity for the two groups in the *o*-phthalates. *m*-Dimethyl phthalate, as noted earlier, yields only a molecular ion by resonance electron capture.

Because the ionizing electrons are captured by the π-orbitals of the benzene ring, $\pi^*-\sigma^*$ mixing is necessary for fragmentation. The alkyl oxygens of *o*-phthalates are orthogonal to the plane of the benzene ring, thereby allowing efficient mixing of the σ^*-orbital of the oxygen with the π^*-orbital of the benzene ring. In *m*-phthalates the alkyl oxygens lie in the plane of the benzene ring where through-space orbital mixing is not possible and fragmentation cannot occur. The long-range $\pi^*-\sigma^*$ orbital mixing observed for the *o*-phthalates that results in the loss of alkyl radicals is similar to the interaction reported for the loss of the benzyl halide in halobenzylhalides.[121] The cross-section for the halide ion arising from the benzylic position is several orders of magnitude greater than that for the halide ion arising from the phenylic position.[121] The freely rotating benzylic C–X bond facilitates through-space $\pi^*-\sigma^*$ interaction, giving high yields of halide ion, whereas out-of-plane distortion of the phenylic C–Cl bond is necessary for $\pi^*-\sigma^*$ mixing and bond cleavage. The rigid constraints imposed by the geometry result in a much lower yield of phenylic halide ion.

An exact analogy for this mechanism was observed for the dissociative loss of the benzyl radical from benzylbutyl phthalate, which is consistent with the ionizing electron becoming localized in the benzyl group.[86] This system does not fit the expected trend for steric crowding displayed by the *o*-dialkyl phthalates where the ionizing electron must first be associated with the benzene ring. These distinct π systems, wherein one system fails to fit the trend of steric crowding during dissociative electron capture, provide added support for the separate roles of electronic and steric factors in dissociative electron capture processes.

4.5.3 Hammett Relationships

Relationships between the ε_{max} values and Hammett substituent constants for organophosphorothioates have been observed.[62,77] An excellent correlation has been demonstrated for the electron energies required to produce $(M - Et)^-$ with Hammett substituent constants (Figure 12). The fluoro compound, however, does not fit the correlation. This compound is also the only one that has an electron energy maximum for the production of $(M - Et)^-$ at precisely the value (0.69 eV) required to form the molecular ion. The ethyl radical in this case must, therefore, arise from decomposition of a metastable ion. The effect of electron-withdrawing substituents on the electron energy required for production of $(M - Et)^-$ is both small and in the direction expected for pairing of the incipient nucleophilic oxygen radical with the unpaired phenyl electron (Scheme 1).

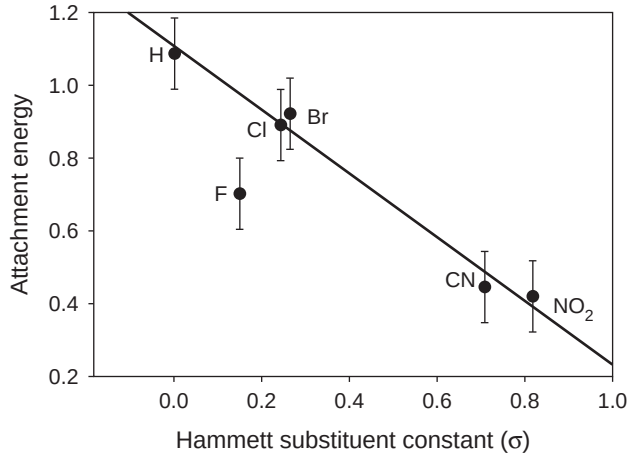

Figure 12 Relationship between Hammett substituent constants and electron attachment energies required to produce $(M - Et)^-$.[77]

5 CONCLUSIONS AND FUTURE DIRECTIONS

The superior sensitivity and selectivity of ECNIMS for electron-capturing compounds continues to attract a large amount of interest. Compounds that absorb low-energy electrons by resonance processes include chlorinated aromatics, chlorinated aliphatics, nitro compounds, polyaromatic hydrocarbons, and organophosphates. These compounds, which may be toxic and of public health concern, are often present in trace amounts. The development of a simple, sensitive, and reliable analytical system for their detection is a compelling goal. Despite the advantages of ECNIMS, there is a basic weakness in the method, which is its unreliability for quantifying small amounts of chemicals. The EMMS system has the potential to remedy this problem because of the elimination of the troublesome reagent gas from the ion source. Data gathered over the past few years indicate that reproducibility of analytical results from this system is better than it is with standard ECNIMS methods. Electron monochromators have been adapted to quadrupole and sector instruments with excellent results and a new source design has led to improved sensitivity over an earlier prototype.

The versatility of this system could be enhanced greatly by interfacing the electron monochromator to an analyzer that would allow simultaneous recording of electron energies and ion masses. The added dimension of analytical information from such a system would provide for more reliable identification of compounds and the three-dimensional print-outs would facilitate interpretation of the data. An electron monochromator interfaced to a time-of-flight analyzer would result in a system that would yield three-dimensional data, as well as open up opportunities for analyzing metastable ions and carrying out ''fast'' gas chromatography/mass spectrometry for greater efficiency in analyzing trace residues of a large number of compounds in mixtures. Pattern recognition and neural networking could also be used to improve the efficiency of identification of compounds in mixtures. And the analysis of chemicals with functionalities that can be derivatized with electrophores will greatly expand the range of applications of EMMS instruments. In addition to the obvious applications in gas chromatography/mass spectrometry, the electron monochromator could play a crucial role in the analysis of biopolymers. For example, the neutralization of electrosprayed protein and peptide cations and their reionization to negative ions with thermal electrons would be useful in analyzing phosphorylated, sulfated, or otherwise modified peptides. The electron monochromator could also be used to advantage for ionizing post-desorbed neutral molecules in matrix-assisted laser desorption/ionization plumes. Sensitivity enhancement for the detection of oligonucleotides and the identification of peptides with electron-capturing functionalities are among the expected benefits.

Another potential advantage of using EMMS analyzers is that the NIRs can be predicted for unknown compounds within a given class by knowing their virtual orbital energies from high-level Gaussian calculations. These methods have proven reliable for correlating the experimental electron attachment energies of aromatic compounds with their calculated π virtual orbital energies. The resonance attachment energies for electron-capturing aliphatic compounds containing only σ-orbitals should also be predictable. The combination of data from ETS, EMMS and high-level MO calculations will help greatly in providing a better understanding of the mechanisms of ion formation and decomposition. Correlations between the energies of electrons required to produce negative fragment ions from metastable ions[5,122] with data from linked scanning or tandem mass spectrometric analysis will also bring about a better understanding of negative ion gas phase chemistry. The range of applications for elucidating negative ion fragmentations with these instrumental and theoretical methods is extensive and certainly goes well beyond the gas phase. For example, these methods could help in gaining a better understanding of the ionization process in particle beam desorption methods, the effects of solvents on electrochemical and anionic reactions, the photochemical decomposition of exciplexes and the repair mechanism of thymidine dimers by photolyase.

ABBREVIATIONS AND ACRONYMS

ECNCIMS	Electron Capture Negative Chemical Ionization Mass Spectrometry
ECNIMS	Electron Capture Negative Ion Mass Spectrometry
EMMS	Electron Monochromator Mass Spectrometry
ETS	Electron Transmission Spectroscopy
fwhm	Full Width at Half-maximum
GC/ECNIMS	Gas Chromatography/Electron Capture Negative Ion Mass Spectrometry
GC/EMMS	Gas Chromatography/Electron Monochromator Mass Spectrometry
LC/ECNIMS	Liquid Chromatography/Electron Capture Negative Ion Mass Spectrometry
LUMO	Lowest Unoccupied Molecular Orbital
MINDO	Modified Intermediate Neglect of Differential Overlap

MO	Molecular Orbital
NCIMS	Negative Chemical Ionization Mass Spectrometry
NIR	Negative Ion Resonance
PCB	Polychlorobiphenyl
PCDD	Polychlorodibenzo-*p*-dioxin
PETN	Pentaerythritol Tetranitrate
RDX	1,3,5-Trinitro-1,3,5-triazocyclohexane
TNT	Trinitrotoluene

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Chemical Weapons Chemicals Analysis **(Volume 1)**
Verification of Chemicals Related to the Chemical Weapons Convention • Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Chemical Weapons Chemicals Analysis cont'd **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Sampling, Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Dioxin-like Compounds, Screening Assays • Explosives Analysis in the Environment • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Nitroaromatics, Environmental Analysis of • Polychlorinated Biphenyls Analysis in Environmental Samples • Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors

Food **(Volume 5)**
Pesticides, Mycotoxins and Residues Analysis in Food

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Industrial Hygiene **(Volume 6)**
Sensors in the Measurement of Toxic Gases in the Air • Spectroscopic Techniques in Industrial Hygiene

Nucleic Acids Structure and Mapping **(Volume 6)**
Polycyclic Aromatic Compounds Mapping

Pesticides **(Volume 7)**
Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis • Multiclass, Multiresidue Analysis of Pesticides, Strategies for • Pesticides (New Generation) and Related Compounds, Analysis of • Phenyl- and Sulfonylurea Herbicides: Single Class, Multiresidue Analysis of

Pesticides cont'd **(Volume 8)**
s-Triazine Herbicides and their Transformation Products, Multi-residue Analysis of

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices • Metals, Nitrogen and Sulfur in Petroleum Residue, Analysis of

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry

REFERENCES

1. D.F. Hunt, 'Electron Capture Negative Ion Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **50**, 1781–1784 (1978).
2. R.C. Dougherty, 'Negative Ion Mass Spectrometry', *Anal. Chem.*, **53**, 625A–636A (1981).
3. C.E. Melton, *Principles of Mass Spectrometry and Negative Ions*, Marcel Dekker, New York, 1970.
4. H. Budzikiewicz, 'Mass Spectrometry of Negative Ions', *Angew. Chem. Int. Ed. Engl.*, **20**, 624–637 (1981).
5. J.H. Bowie, 'The Formation and Fragmentation of Negative Ions Derived from Organic Molecules', *Mass Spectrom. Rev.*, **3**, 161–207 (1984).

6. V.S. Ong, R.A. Hites, 'Electron Capture Mass Spectrometry of Organic Environmental Contaminants', *Mass Spectrom. Rev.*, **13**, 259–283 (1994).
7. W.B. Knighton, L.J. Sears, E.P. Grimsrud, 'High-pressure Electron Capture Mass Spectrometry', *Mass Spectrom. Rev.*, **14**, 327–343 (1995).
8. D.W. Kuehl, M.J. Whitaker, R.C. Dougherty, 'Micromethods for Toxic Residue Screening by Negative Chemical Ionization Mass Spectrometry', *Anal. Chem.*, **52**, 935–940 (1980).
9. Y. Tondeur, R.C. Dougherty, 'New Screening Methods for Acidic Toxic Substances Using Negative Chemical Ionization Mass Spectrometry. Tetrachloroterephthalate in Human Urine', *Environ. Sci. Technol.*, **15**, 216–219 (1981).
10. R.C. Dougherty, K. Piotrowska, 'Screening by Negative Chemical Ionization Mass Spectrometry from Environmental Contamination with Toxic Residues: Application to Human Urines', *Proc. Natl. Acad. Sci. USA*, **73**, 1777–1781 (1976).
11. R.C. Dougherty, M.J. Whitaker, L.M. Smith, D.L. Stalling, D.W. Kuehl, 'Negative Chemical Ionization Studies of Human and Food Chain Contamination with Xenobiotic Chemicals', *Environ. Health Perspect.*, **36**, 113–118 (1980).
12. S.L. Wagner, L.R. Durand, R.D. Inman, U. Kiigemagi, M.L. Deinzer, 'Residues of Pentachlorophenol and Other Chlorinated Contaminants in Human Tissues: Analysis by Electron Capture Gas Chromatography and Electron Capture Negative Ion Mass Spectrometry', *Arch. Environ. Contam. Toxicol.*, **21**, 596–606 (1991).
13. C.E. Klotz, 'Energy Deposition Mechanisms', in *Fundamental Processes in Radiation Chemistry*, ed. P. Ausloos, John Wiley and Sons, New York, 1–57, 1968.
14. E.D. Pellizzari, 'Electron Capture Detection in Gas Chromatography', *J. Chromatogr.*, **98**, 323–361 (1974).
15. J.A. Laramée, B.C Arbogast, M.L. Deinzer, 'Electron Capture Negative Ion Chemical Ionization Mass Spectrometry of 1,2,3,4-Tetrachlorodibenzo-*p*-dioxin', *Anal. Chem.*, **58**, 2907–2912 (1986).
16. J.A. Laramée, B.C. Arbogast, M.L. Deinzer, 'Negative Ion Mass Spectrometry of Polychlorodibenzo-*p*-dixoins: Correlations Between Observed Fragmentation and Calculated Total Internal Energies', *Anal. Chem.*, **60**, 1937–1943 (1988).
17. I.K. Gregor, M. Guilhaus, 'Comparative Study of the Influence of the Electron Energy Moderating Gas in Electron Attachment Reactions under NCI Conditions', *Int. J. Mass Spectrom. Ion Phys.*, **56**, 167–176 (1984).
18. D.R. Zook, W.B. Knighton, E.P. Grimsrud, 'Effect of Buffer Gas and Pressure Variations on the Formation of Br_2^- in Reactions of Thermal Electrons with Dibrominated Hydrocarbons and Fluorocarbons', *Int. J. Mass Spectrom. Ion Proc.*, **104**, 63–80 (1991).
19. E.A. Stemmler, R.A. Hites, 'A Systematic Study of Instrumental Parameters Affecting Electron Capture Negative Ion Mass Spectra', *Biomed. Environ. Mass Spectrom.*, **15**, 659–667 (1988).
20. J. Greaves, J.G. Bekesi, J. Roboz, 'Halogen Anion Formation by Polybrominated Compounds in Negative Chemical Ionization Mass Spectrometry', *Biomed. Environ. Mass Spectrom.*, **9**, 406–410 (1982).
21. W.A. Garland, B.J. Miwa, 'The $[M-1]^-$ Ion in the Negative Chemical Ionization Mass Spectra of Diazepam and Nordiazepam', *Biomed. Environ. Mass Spectrom.*, **10**, 126–129 (1983).
22. V.S. Ong, R.A. Hites, 'Why are Electron Capture Negative Ion Mass Spectra Not Reproducible? an Ion Source Problem', *J. Am. Soc. Mass Spectrom.*, **4**, 270–277 (1993).
23. D. Stöckl, H. Budzikiewicz, 'Negative Chemical Ionization Spectra of Aromatic Acids and Esters or Are Negative Chemical Ionization Spectra of Use Only If One Knows What One Is Looking For?', *Org. Mass Spectrom.*, **17**, 470–474 (1982).
24. D. Stöckl, H. Budzikiewicz, 'Molecule/Radical Reactions Prior to Ionization under Negative Chemical Ionization Conditions', *Org. Mass Spectrom.*, **17**, 376–381 (1982).
25. E.A. Stemmler, R.A. Hites, B. Arbogast, W.L. Budde, M.L. Deinzer, R.C. Dougherty, J.W. Eichelberger, R.L. Foltz, C. Grimm, E.P. Grimsrud, C. Sakashita, L.J. Sears, 'Interlaboratory Comparison of Methane Electron Capture Negative Ion Mass Spectra', *Anal. Chem.*, **60**, 781–787 (1988).
26. B. Arbogast, W.L. Budde, M.L. Deinzer, R.C. Dougherty, J. Eichelberger, R.D. Foltz, C.C. Grimm, R.A. Hites, C. Sakashita, E. Stemmler, 'Interlaboratory Comparison of Limits of Detection in Negative Chemical Ionization Mass Spectrometry', *Org. Mass Spectrom.*, **25**, 191–196 (1990).
27. L.G. Christophorou (ed.), *Electron–Molecule Interactions and their Applications*, Academic Press, Orlando, FL, Vol. 1, 1984.
28. L.G. Christophorou (ed.), *Electron–Molecule Interactions and their Applications*, Academic Press, Orlando, FL, Vol. 2, 1984.
29. T. Oster, A. Kühn, E. Illenberger, 'Gas Phase Negative Ion Chemistry', *Int. J. Mass Spectrom. Ion Proc.*, **89**, 1–72 (1989).
30. O. Ingólfsson, F. Weik, E. Illenberger, 'The Reactivity of Slow Electrons with Molecules at Different Degrees of Aggregation: Gas Phase, Clusters and Condensed Phase', *Int. J. Mass Spectrom. Ion. Proc.*, **155**, 1–68 (1996).
31. P.D. Burrow, G.A. Gallup, I.I. Fabrikant, K.D. Jordan, 'Dissociative Attachment Studies of Halogen Containing Molecules: Problems, Applications and Challenges', *Aust. J. Phys.*, **49**, 403–423 (1996).
32. A.G. Harrison, *Chemical Ionization Mass Spectrometry*, CRC Press, Boca Raton, 23, 1983.

33. L.G. Christophorou, D.L. McCorkle, A.A. Christodoulides, 'Electron Attachment Processes', in *Electron–Molecule Interactions and their Applications*, ed. L.G. Christophorou, Academic Press, Orlando, FL, 477–617, Vol. 1, 1984.
34. K.D. Jordan, P.D. Burrow, 'Resonances in Electron–Molecule Scattering, van der Waals Complexes, and Reactive Chemical Dynamics', in *American Chemical Society Symposium Series*, ed. D.G. Truhlar, ACS Press, Washington, DC, 165–182, Vol. 9, 1984.
35. J.P. Johnson, D.L. McCorkle, L.G. Christophorou, J.G. Carter, 'Long-lived Parent Negative Ions formed via Nuclear-excited Feshbach Resonances', *J. Chem. Soc., Faraday Trans. II*, **71**, 1742–1751 (1975).
36. W.B. Knighton, R.S. Mock, D.S. McGrew, E.P. Grimsrud, 'Resonance Electron Capture Rate Constants for Substituted Nitrobenzenes', *J. Phys. Chem.*, **98**, 3770–3776 (1994).
37. A.A. Christodoulides, D.L. McCorkle, L.G. Christophorou, 'Electron Affinities of Atoms, Molecules, and Radicals', in *Electron–Molecule Interactions and their Applications*, ed. L.G. Christophorou, Academic Press, Orlando, FL, 423–641, Vol. 2, 1984.
38. L. Sanche, G.J. Schulz, 'Electron Transmission Spectroscopy: Resonances in Triatomic Molecules and Hydrocarbons', *J. Chem. Phys.*, **58**, 479–493 (1973).
39. T. Kunii, H. Kuroda, 'Ionization Potentials and Electron Affinities of Carbo- and Heterocyclic C-conjugated Molecules', *Theor. Chim. Acta*, **11**, 97–106 (1968).
40. E. Illenberger, H.-U. Scheunemann, H. Baumgärtel, 'Negative Ion Formation in CF_2Cl_2, CF_3Cl and $CFCl_3$ following Low-energy (0–10 eV) Impact with Near Monoenergetic Electrons', *Chem. Phys.*, **37**, 21–31 (1979).
41. Y.N. El'kin, V.G. Boinov, A.A. Usol'tsev, 'Detection of Fatty Acids in the Form of their Mixed Anhydrides with Trifluoroacetic Acid by the Methods of Mass Spectroscopy and Resonance Electron Capture', *J. Anal. Chem.*, **49**(5), 443–446 (1994).
42. V.G. Voinov, Y.N. El'kin, 'Mass Spectrometric Study of Resonance Electron Capture by Molecules of 3-Substituted Myristic Acids', *Int. J. Mass Spectrom. Ion Proc.*, **113**, 241–246 (1992).
43. W.B. Nottingham, 'Ionization and Excitation in Mercury Vapor Produced by Electron Bombardment', *Phys. Rev.*, **55**, 203–219 (1939).
44. D.A. Hutchison, 'An Electrostatic Parallel Plate Electron-energy Selector for Determinations of Critical Ionization Potentials by Electron Impact', *Adv. Mass Spectrom.*, **2**, 527–539 (1963).
45. S.N. Foner, B.H. Nall, 'Structure in the Ionization near Threshold of Rare Gases by Electron Impact', *Phys. Rev.*, **122**, 512–524 (1961).
46. E.M. Clarke, 'Ionization Probability Curves Using an Electron Selector', *Can. J. Phys.*, **32**, 764–774 (1954).
47. P. Marmet, L. Kerwin, 'An Improved Electrostatic Electron Selector', *Can. J. Phys.*, **38**, 787–796 (1960).
48. C.E. Brion, D.C. Frost, C.A. McDowell, 'Mass-spectrometric Studies of the Rare Gases Using an Electron Velocity Selector', *J. Chem. Phys.*, **44**, 1034–1038 (1966).
49. J.A. Simpson, 'High Resolution, Low Energy Electron Spectrometer', *Rev. Sci. Instrum.*, **35**, 1699–1704 (1964).
50. C.E. Kuyat, J.A. Simpson, 'Electron Monochromator Design', *Rev. Sci. Instrum.*, **38**, 103–111 (1967).
51. J.H. Parker, Jr, R.W. Warren, 'Kelvin Device to Scan Large Areas for Variations in Contact Potential', *Rev. Sci. Instrum.*, **33**, 948–950 (1962).
52. K. Maeda, G.P. Semeluk, F.P. Lossing, 'A Two-stage Double Hemispherical Electron Energy Selector', *Int. J. Mass Spectrom. Ion Phys.*, **1**, 395–407 (1968).
53. M.T. Bernius, A. Chutjian, 'Pulsed, High-current, In-line Reversal Electron Attachment Detector', *J. Appl. Phys.*, **66**, 2783–2788 (1989).
54. S. Boumsellek, A. Chatjian, 'Pulsed, Gridded Electron Reversal Ionizer', *Rev. Sci. Instrum.*, **64**(5), 1135–1139 (1993).
55. J.J. Thomson, 'On Rays of Positive Electricity', *Philos. Mag. S.*, **13**, 561–578 (1907).
56. W. Bleakney, J.A. Hipple, Jr, 'A New Mass Spectrometer with Improved Focusing Properties', *Phys. Rev.*, **53**, 521–529 (1938).
57. R.E. Fox, W.M. Hickam, D.J. Grove, T. Kjeldaas, Jr, 'Ionization in a Mass Spectrometer by Monoenergetic Electrons', *Rev. Sci. Instrum.*, **26**(12), 1101–1107 (1955).
58. R.E. Fox, W.M. Hickam, T. Kjeldaas, Jr, 'Ionization Probability Curves for Krypton and Xenon near Threshold', *Phys. Rev.*, **89**(3), 555–558 (1953).
59. A. Stamatovic, G.J. Schulz, 'Trochoidal Electron Monochromator', *Rev. Sci. Instrum.*, **39**(11), 1752–1753 (1968).
60. A. Stamatovic, G.J. Schulz, 'Characteristics of the Trochoidal Electron Monochromator', *Rev. Sci. Instrum.*, **41**, 423–427 (1970).
61. J.A. Laramée, C.A. Kocher, M.L. Deinzer, 'Application of a Trochoidal Electron Monochromator/Mass Spectrometer System to the Study of Environmental Chemicals', *Anal. Chem.*, **64**, 2316–2322 (1992).
62. J.A. Laramée, P. Mazurkiewicz, V. Berkout, M.L. Deinzer, 'Electron Monochromator–Mass Spectrometer Instrument for Negative Ion Analysis of Electronegative Compounds', *Mass Spectrom. Rev.*, **15**, 15–42 (1996).
63. P. Cloutier, L. Sanche, 'A Trochoidal Spectrometer for the Analysis of Low-energy Inelastically Backscattered Electrons', *Rev. Sci. Instrum.*, **60**, 1054–1060 (1989).
64. M.V. Muftakhov, Y.V. Vasil'ev, É.R. Nazirov, V.A. Mazunov, 'Electron Monochromator for Negative Ion Source of Mi-1201 Mass Spectrometer', *Instrum. Exp. Tech.*, **32**(2), 443–445 (1989).

65. V.A. Mazunov, Y.U. Vasil'ev, M.V. Muftakhov, V.I. Khvostenko, 'Reproducibility of Results in Mass Spectrometric Analysis with Resonance Electron Capture', *J. Anal. Chem. USSR*, **44**(11), 1571–1576 (1989).
66. J.A. Laramée, R.B. Cody, W.B. Hermannsfeldt, M.L. Deinzer, 'Tunable Energy Electron Monochromator Interface to a Sector Instrument', *Proc. 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, May 12–16, 1996.
67. J.R. Pierce, *Theory and Design of Electron Beams*, Van Nostrand, New York, 167–187, 1949.
68. V.D. Berkout, P.H. Mazurkiewicz, M.L. Deinzer, 'Collisional Stabilization of Negative Ions Produced by a Monoenergetic Electron Beam', *Rapid Commun. Mass Spectrom.*, **13**, 1850–1854 (1999).
69. L.J. Sears, E.P. Grimsrud, 'Elimination of Unexpected Ions in Electron Capture Mass Spectrometry using Carbon Dioxide Buffer Gas', *Anal. Chem.*, **61**, 2523–2528 (1989).
70. J.A. Laramée, M.L. Deinzer, 'Capillary Gas Chromatographic Introduction of Environmental Compounds in a Trochoidal Electron Monochromator/Mass Spectrometer', *Anal. Chem.*, **66**, 719–724 (1994).
71. K. Jäger, A. Henglein, 'Negative Iononen Durch Electronen Stoss aus Organischen Nitroverbindungnen Äthylnitrit und Äthylnitrat. *Z. Naturforsch.*, **22a**, 700–704 (1967).
72. R.B. Cody, J.A. Laramée, V. Berkout, M.L. Deinzer, '$^{13}C(M-H)^-$ Interferences as Revealed by the Electron Monochromator–Mass Spectrometer', *Proc. 46th ASMS Conference on Mass Spectrometry and Allied Topics*, 1998, Orlando, FL, 1245, May 31–June 4, 1998.
73. J.A. Laramée, P.C.H. Eichinger, P. Mazurkiewicz, M.L. Deinzer, 'Analysis of Organophosphate Pesticides by a Trochoidal Electron Monochromator Mass Spectrometer System', *Anal. Chem.*, **67**, 3476–3481 (1995).
74. D. Zakett, R.G.A. Flynn, R.G. Cooks, 'Chlorine Isotope Effects in Mass Spectrometry by Multiple Reaction Monitoring', *J. Phys. Chem.*, **82**, 2359–2362 (1978).
75. V.D. Berkout, J.A. Laramée, M.L. Deinzer, 'Electron Energy-dependent Regioselective Loss of Chloride Anions from Polychlorodibenzo-*p*-dioxins', *Proc. 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, May 12–16, 1996.
76. V.D. Berkout, P.H. Mazurkiewicz, M.L. Deinzer, 'Electron Energy Dependence of Regioselective Chloride Ion Loss from Polychlorodibenzo-*p*-dioxins. Relationship Between Resonance Electron Energies and Virtual Orbital Energies', *J. Am. Chem. Soc.*, **121**(11), 2561–2570 (1999).
77. P.H. Mazurkiewicz, 'Analysis of Electrophilic Compounds Using an Electron Monochromator–Mass Spectrometer System', *PhD Thesis*, Oregon State University, 27–88, 1998.
78. L.Q. Huang, M.J.I. Mattina, 'Methane-enhanced Negative Chemical Ionization Mass Spectrometry of *s*-Triazines', *Biomed. Environ. Mass Spectrom.*, **18**, 828–835 (1989).
79. E.A. Stemmler, R.A. Hites, *Electron Capture Negative Ion Mass Spectra of Environmental Contaminants and Related Compounds*, VCH Publishers, New York, 1988.
80. S. Süzer, E. Illenberger, H. Baumgärtel, 'Negative Ion Mass Spectra of Hexafluoro-1,3-butadiene, Hexafluoro-2-butene, and Hexafluorocyclobutene. Identification of the Structural Isomers', *Org. Mass Spectrom.*, **6**, 292–293 (1984).
81. H.-P. Fenzlaff, E. Illenberger, 'Low Energy Electron Impact on Benzene and the Fluorobenzenes, Formation and Dissociation of Negative Ions', *Int. J. Mass Spectrom. Ion Proc.*, **59**, 185–202 (1984).
82. K.D. Jordan, P.D. Burrow, 'Temporary Anion States of Polyatomic Hydrocarbons', *Chem. Rev.*, **87**, 557–588 (1987).
83. A. Hadjiantoniou, L.G. Christophorou, J.G. Carter, 'Long-lived Parent Negative Ions Formed via Nuclear-excited Feshbach Resonances', *J. Chem. Soc., Faraday Trans. II*, **69**, 1704–1712 (1973).
84. B.G. Zykov, Yu.V. Vasil'ev, V.S. Fal'ko, A.N. Lachinov, V.I. Khvostenko, N.G. Gileva, 'Dynamics of the Electron Structure of Phthalide Derivatives at the Interaction with Low-energy Electrons', *JEPT Lett.*, **64**(6), 439–443 (1996).
85. Yu V. Vasil'ev, B.G. Zykov, V.S. Fal'ko, A.N. Lachinov, V.I. Khvostenko, N.G. Gileva, 'Resonant Capture of Low-energy Electrons by Molecules of Phthalide Derivatives', *Synth. Met.*, **84**, 975–976 (1997).
86. P.H. Mazurkiewicz, M.L. Deinzer, 'Steric Effects on Gas Phase Negative Ion Resonances', *J. Am. Chem. Soc.*, **121**(14), 3421–3427 (1999).
87. E.P. Grimsrud, 'Gas-phase Ion Chemistry within the Electron Capture Detector', *Mass Spectrom. Rev.*, **10**(6), 467–489 (1992).
88. K.P. Madhusudanan, 'Formation of Radical Anion Adducts between Dihydroxybenzenes and the Matrix under Fast Atom Bombardment', *J. Mass Spectrom.*, **33**, 173–182 (1998).
89. K.P. Madhusudanan, 'Formation of Radical Anion Adduct between Paclitaxel and Matrix under Fast-atom Bombardment', *Rapid Commun. Mass Spectrom.*, **11**, 295–297 (1997).
90. J.A. Laramée, B. Arbogast, M.L. Deinzer, 'Evidence for Radical Anion Formation During Liquid Secondary Ion Mass Spectrometry Analysis of Oligonucleotides and Synthetic Oligomeric Analogs: a Deconvolution Algorithm for Molecular Ion Region Clusters', *Anal. Chem.*, **61**, 2154–2160 (1989).
91. J.A. Laramée, B. Arbogast, M.L. Deinzer, E.P. Stirchak, D.D. Weller, 'Negative Ion Liquid Secondary Ion Mass Spectrometry of Antiparallel *N*-Carbamoylmorpholine-linked Nucleic Acid Oligomers: Evidence

for Fragmentation from Molecular Radical Ion Precursors', *Org. Mass Spectrom.*, **25**, 33–38 (1990).
92. J.A. Laramée, B. Arbogast, E.P. Stirchak, D.D. Weller, M.L. Deinzer, 'Negative Ion Liquid Secondary Ion Mass Spectrometry of Carbamate-linked Oligodeoxynucleosides. II', *Org. Mass Spectrom.*, **25**, 219–224 (1990).
93. R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, 'Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process', *J. Am. Chem. Soc.*, **120**, 3265–3266 (1998).
94. R.A. Zubarev, N.A. Kruger, E.K. Fridriksson, M.A. Lewis, D.M. Horn, B.K. Carpenter, F.W. McLafferty, 'Electron Capture Dissociation of Gaseous Multiply-charged Proteins is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity', *J. Am. Chem. Soc.*, **121**, 2857–2862 (1999).
95. T.M. Trainor, P. Vouros, 'Electron Capture Negative Ion Chemical Ionization Mass Spectrometry of Derivatized Chlorophenols and Chloranilines', *Anal. Chem.*, **59**, 601–610 (1987).
96. S. Murray, D. Watson, '*Bis*-Trifluoromethylbenzoyl Derivatives for Steroid Analysis by Gas Chromatography Electron Capture Negative Ion Chemical Ionization Mass Spectrometry', *J. Steroid Biochem.*, **25**(2), 255–260 (1986).
97. D.A. Durden, 'An Evaluation of the Negative Ion Mass Spectra of Electron-capturing Derivatives of the Biogenic Trace Amines. I. Phenylethylamine', *Biol. Mass Spectrom.*, **20**, 367–374 (1991).
98. D.A. Durden, B.A. Davis, A.A. Boulton, 'Quantification of Plasma Phenylethylamine by Combined Gas Chromatography/Electron Capture Negative Ion Mass Spectrometry of the *N*-Acetyl-*N*-pentafluorobenzoyl Derivative', *Biol. Mass Spectrom.*, **20**, 375–381 (1991).
99. S. Murray, D. Watson, D.S. Davies, 'Bistrifluoromethylaryl Derivatives for Drug Analysis by Gas Chromatography Electron Capture Negative Ion Chemical Ionization Mass Spectrometry. Application of the Measurement of (*N*-Dicyclopropylmethyl)amino-2-oxazoline in Plasma', *Biomed. Environ. Mass Spectrom.*, **12**(5), 230–237 (1985).
100. J.T. Simpson, D.S. Torok, J.E. Girard, S.P. Markey, 'Analysis of Amino Acids in Biological Fluids by Pentafluorobenzyl Chloroformate Derivatization and Detection by Electron Capture Negative Ionization Mass Spectrometry', *Anal. Biochem.*, **233**(1), 58–66 (1996).
101. D.B. Maritsin, S.P. Markey, 'Assessment of DNA Oxidative Damage by Quantification of Thymidine Glycol Residues using Gas Chromatoraphy/Electron Capture Negative Ionization Mass Spectrometry', *Anal. Biochem.*, **241**(1), 35–41 (1996).
102. D.B. Naritsin, R.L. Boni, S.P. Markey, 'Pentafluorobenzylation Method for Quantification of Acidic Trytophan Metabolites using Electron Capture Negative Ion Mass Spectrometry', *Anal. Chem.*, **67**(5), 863–870 (1995).
103. R.L. Boni, J.T. Simpson, D.B. Naritsin, K. Saito, S.P. Markey, 'Quantification of L-Tryptophan and L-Kynurenine by Liquid Chromatography/Electron Capture Negative Ion Chemical Ionization Mass Spectrometry', *Biol. Mass Spectrom.*, **23**(1), 27–32 (1994).
104. C. Schotten, 'Ueber die Oxidation dis Piperidins', *Berichte*, **17**, 2544–2547 (1884); E. Baumann, 'Ueber eine Einfache Methode der Darstellung von Benzoësäureäthern', *Berichte*, **19**, 3218–3222 (1886).
105. J.A. Laramée, V. Berkout, S. Eby, M.L. Deinzer, 'Analysis of Materials Related to Chemical Warfare Agents using the Electron Monochromator/Mass Spectrometer Instrument', *Proc. 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, 465, June 1–5, 1997.
106. M.B. Beverly, K.J. Voorhees, R.B. Cody, J.A. Laramée, E. Phillips, 'Analysis of Dipicolinic Acid in Bacterial Spores by Electron Monochromator – Mass Spectrometry', *Proc. 47th ASMA Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, June 13–17, 1999.
107. J.A. Laramée, M.L. Deinzer, 'Tunable Energy (0.03 to 30 eV) Electron Capture Negative Ion Mass Spectrometry', *Proc. 5th International Conference on Explosives*, Washington, DC, Dec. 4–8, 1995.
108. J.A. Laramée, M.L. Deinzer, 'Dry CI – The Use of Monoenergetic Electrons for Analysis of Explosives, Pesticides and Environmental Chemicals', *Proc. 40th ASMS Conference on Mass Spectrometry and Allied Topics*, Washington, DC, 995–996, May 31–June 5, 1992.
109. J.H. Bowie, 'The Fragmentations of Even-electron Organic Negative Ions', *Mass Spectrom. Rev.*, **9**, 349–379 (1990).
110. M. Born, S. Ingemann, N.M.M. Nibbering, 'Formation and Chemistry of Radical Anions in the Gas Phase', *Mass Spectrom. Rev.*, **16**, 181–200 (1997).
111. Y. Guo, J.J. Grabowski, 'Mechanistic Insight into the Reactions of $O^{-\cdot}$ with Aromatic Compounds and the Synthesis of Didehydroaromatic anions in the Gas Phase', *Int. J. Mass Spectrom. Ion Proc.*, **117**, 299–326 (1992).
112. J. Lee, J.J. Grabowski, 'Reactions of the Atomic Oxygen Radical Anion and the Synthesis of Organic Reactive Intermediates', *Chem. Rev.*, **92**, 1611–1647 (1992).
113. T. Koopmans, 'Über die Zuordnung von Wellenfunktionen und Eigenewerten zu den Einzelnen Elektronen Eines Atoms', *Physica*, **1**, 104–113 (1934).
114. A.E. Howard, S.W. Staley, 'Negative Ion States of Three- and Four-membered Ring Hydrocarbons. Studied by Electron Transmission Spectroscopy', in *Resonances*, ed. D.G. Tuhlar, American Chemical Society, Washington, DC, 183–192, 1984.
115. N. Heinrich, W. Koch, G. Frenking, 'On the Use of Koopmans' Theorem to Estimate Negative Electron Affinities', *Chem. Phys. Lett.*, **124**(1), 20–25 (1986).

116. S.W. Staley, J.T. Strnad, 'Calculation of Energies of π^* Negative Ion Resonance States by the Use of Koopmans' Theorem', *J. Phys. Chem.*, **98**, 116–121 (1994).
117. D.M. Pearl, P.D. Burrow, J.J. Nash, H. Morrison, K.D. Jordan, 'Dissociative Attachment as a Probe of Intramolecular Electron Transfer', *J. Am. Chem. Soc.*, **115**, 9876–9877 (1993).
118. D.M. Pearl, P.D. Burrow, J.J. Nash, H. Morrison, D. Nachtigallova, K.D. Jordan, 'Dissociative Attachment as a Probe of the Distance Dependence of Intramolecular Electron Transfer', *J. Phys. Chem.*, **99**, 12 379–12 381 (1995).
119. K.L. Stricklett, S.C. Chu, P.D. Burrow, 'Dissociative Attachment in Vinyl and Allyl Chloride, Chlorobenzene and Benzyl Chloride', *Chem. Phys. Lett.*, **131**(3), 279–284 (1986).
120. P.K. Freeman, R. Srinivasa, J.-A. Campbell, M.L. Deinzer, 'The Photochemistry of Polyhaloarenes. 5. Fragmentation Pathways in Polychlorobenzene Radical Ions', *J. Am. Chem. Soc.*, **108**, 5531–5536 (1986).
121. C. Bulliard, M. Allan, E. Haselbach, 'Intramolecular Competition of Phenylic and Benzylic CX Bond Breaking in Dissociative Electron Attachment to Dihalotoluenes', *J. Phys. Chem.*, **98**, 11 040–11 045 (1994).
122. J.A. Laramée, Y.S. Chang, B.C. Arbogast, M.L. Deinzer, 'Relationship of Molecular Radical Anion Abundance and Calculated Lowest Unoccupied Molecular Orbital Energies for Polychlorinated Dibenzofurans and Dibenzodioxins in Electron Capture Negative Ion Mass Spectrometry: Evidence for Negative Metastable Ions', *Biomed. Environ. Mass Spectrom.*, **17**, 63–67 (1988).

Electron Ionization Mass Spectrometry

J. Throck Watson
Michigan State University, East Lansing, USA

1 Introduction 1
2 General Appearance of the Data 2
3 Impact of Stable Isotopes 2
4 Strategy for Data Interpretation 4
5 General Features of Mass Spectral Data 4
6 Cursory Preliminary Examination of Mass Spectral Data 5
7 Detailed Examination/Rationalization of Fragmentation 8
8 Representative Mass Spectra of Selected Classes of Compounds 11
8.1 Hydrocarbons 11
8.2 Alcohols 12
8.3 Esters of Carboxylic Acids 13
8.4 Amines 14
9 Summary 14
Abbreviations and Acronyms 15
Related Articles 15
References 15

Electron ionization (EI) is the basis for one of the most effective mass spectrometric means for identification of a given organic compound. The identification is based on an indication of the molecular weight (MW) coupled with a fragmentation pattern that serves as a chemical fingerprint for the compound. Of the many ionization techniques available in mass spectrometry, EI is one of the most harsh in terms of promoting fragmentation of the molecular ion. The analyst can take advantage of this extensive fragmentation by using some of the fundamental principles of organic mechanistic chemistry and physical chemistry for the purposes of rationalizing the fragmentation pathways that characterize various structural features in an organic molecule. Four basic mechanisms of fragmentation are considered in this chapter; most of them operate on an odd-electron ion, although some pathways will be applicable to even-electron ions as well. The multiplicity of peaks that represent a given ion allows one to estimate the elemental composition of an ion. This information on composition, coupled with a rationalization of a given fragmentation pathway, often enables the structure of a given organic compound to be deduced.[1–3]

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Mass spectrometers respond to the mass-to-charge, not the mass, of a particle. Thus, the essential first step in any mass spectrometric analysis is to convert the neutral molecules of the analyte to ions or charged particles that can be affected by applied electric or magnetic fields. In EI, the molecules of the analyte must be present in the vapor phase so that they can interact effectively with energetic electrons at reduced pressure. The vacuum is essential so that a reasonable flux of free electrons can be generated and focused into a region of interaction with the analyte molecules. The bombarding electrons are made available, by convention, with 70 eV of kinetic energy. Ionization of the analyte molecule is not necessarily achieved by direct impact by the electron, but rather by interaction of the bombarding electron with the electron cloud of the molecule in a near-miss trajectory. In this way, some of the kinetic energy of the bombarding electron is absorbed by the analyte molecule; a wide range of energies can be transferred to the analyte molecule, but typically an average of 10–12 eV of energy is absorbed by the molecule.

As indicated in Scheme 1, the analyte molecule absorbs approximately 10 eV, which exceeds the ionization potential of the molecule, and thus it expels one of its own electrons to become a molecular ion. Note that the molecular ion thus formed is an odd-electron species.

$$M + e^-_{70\,eV} \longrightarrow M^{+\bullet} + e^-_{60\,eV} + e^-_{thermal}$$

Scheme 1

The nascent molecular ion formed in the ionization process illustrated in Scheme 1 has considerable excess energy. Typically, the ionization potential of an organic molecule is around 9 eV; thus, the nascent molecular ion formed in the ionization process would in this case have an excess of 1 eV (absorption of 10 eV, minus 9 eV lost during the ionization process). As the molecular ion is in a vacuum, in which the mean free path is on the order of 1 m, it is not likely that it will be able to give up the excess energy through a collision. Thus, the nascent molecular ion will tend to fragment chemical bonds in an effort to give up the excess energy. The bonds that are broken during the fragmentation process are not necessarily the weakest bonds in the molecule, but rather those bonds that lead to the production of a more stable ion (perhaps

an even-electron ion and a reasonably stable radical species) as described in greater detail in a later section.

2 GENERAL APPEARANCE OF THE DATA

The mass spectrum is a data record of the abundance of various ions in the mass spectrometer as a function of their mass-to-charge value. These data are generally available in graphical form such as the bar graph shown in Figure 1. By convention, the ordinate at the left represents the relative intensity of the peaks in the mass spectrum; the peak representing the most abundant ion in the mass spectrum is called the base peak and is assigned a value of 100%. The ordinate on the right-hand side of the bar graph indicates the percentage of total ionization; that is, the abundance of any given ion relative to the sum of the abundances of all the ions. The abscissa is the value of the mass-to-charge on the ion as represented by the symbol m/z where m represents the mass of the particle in daltons and z represents the number of charges on the ion.

The mass of an ion is given in daltons where a dalton is equal to one-twelfth the mass of ^{12}C. An atomic mass unit is represented by the symbol ''u''; this is the same as a dalton (symbol Da). The symbol ''amu'' is no longer used as, prior to 1961, amu was used by chemists to represent one-sixteenth the average mass of oxygen and was used by physicists to represent the one-sixteenth the mass of ^{16}O. The symbol u is now recommended by IUPAC as the unified atomic mass unit and is assigned the value of one-twelfth the mass of ^{12}C.

The chemical fingerprint of 4-decanone is manifest in the mass spectrum shown in Figure 1. The peak at m/z 156 represents the molecular ion; note that it occurs at an even value of mass-to-charge. The peak at m/z 43 is the base peak in this mass spectrum. Note that most of the fragment ion peaks occur at an odd value of m/z, namely, m/z 113, m/z 99, m/z 71, m/z 43, m/z 4l, m/z 39, m/z 29, and m/z 27. Two of the peaks in the ''fragment ion region'' occur at an even value of mass-to-charge, namely, m/z 86 and m/z 58.

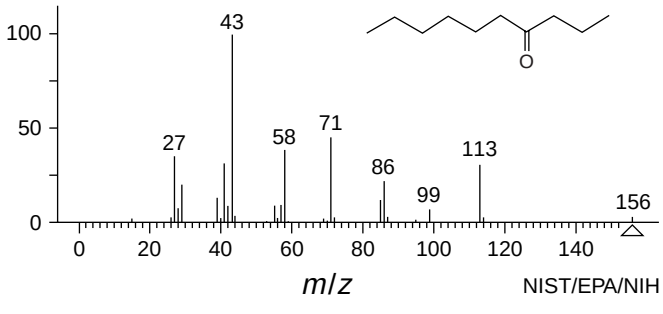

Figure 1 Mass spectrum of 4-decanone.

It is important to realize that ions exist within the mass spectrometer itself and that the relative concentration of the ions in the mass spectrometer is indicated by their abundance. Peaks in the mass spectrum represent the ions in the mass spectrometer. The abundances of the ions in the mass spectrometer are represented by the intensities of the corresponding peaks in the mass spectrum. Thus, in the mass spectrum of 4-decanone (Figure 1), the ions of m/z 43 are the most abundant, as indicated by the peak at m/z 43 being the most intense peak in the mass spectrum.

The peaks at m/z 156, m/z 86, and m/z 58 represent odd-electron ions. The peaks in the mass spectrum at m/z 113, m/z 99, m/z 71, m/z 43, and m/z 27 represent even-electron ions, as described in a later section.

3 IMPACT OF STABLE ISOTOPES

To an extent limited by the resolution of the mass spectrometer, mass spectrometry is a technique for determining the mass of isotopes as opposed to the atomic weight of an atom.

The issue of average mass of an element (i.e. atomic weight) versus the mass of individual isotopes of an element can be made by examining the mass spectrum of an atom of bromine. As can be found in any elementary text on chemistry, the average atomic weight of bromine is 80 Da. However, as can be seen in the mass spectrum of a bromine atom as shown in Figure 2, there is no peak at m/z 80 – there is only a peak at m/z 79, and another at 81. The peak at m/z 79 represents the nominal mass of bromine, whereas the peak at m/z 81 represents the isotope peak, which for bromine (and chlorine) occurs two mass units above the nominal mass.

The nominal mass of an element is the mass of the most abundant stable isotope, as rounded to an integer value. The accepted convention in low-resolution EI mass spectrometry is to refer to the nominal mass of an ion regardless of whether it represents a molecular or a fragment ion.

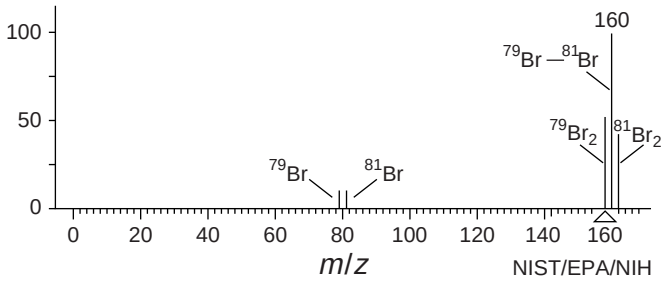

Figure 2 Mass spectrum of atomic and molecular bromine.

The multiplicity of isotope peaks when more than one atom of a given element is present in an ion is shown by the peaks at m/z 158, 160, and 162 in Figure 2, which represents the mass spectrum of diatomic bromine. The nominal mass of molecular bromine is 158 Da. The peak at m/z 158 represents those molecules of bromine that contain both atoms as the isotope of ^{79}Br. The peak at m/z 160 represents those bromine molecules containing one ^{79}Br isotope combined with one ^{81}Br isotope, the most probable case. The peak at m/z 162 represents those bromine molecules in which both bromine atoms are present as the ^{81}Br isotope. Note that the peak at m/z 158 for the bromine molecule is not the most intense in the isotope cluster, even though it represents the nominal mass of the molecule. This is because of the unusually high abundance of the heavy isotope of Br.

The distinctive multiplicity of isotope peaks for two bromine atoms is present in the mass spectrum of any ion containing these two atoms, as indicated in the spectra shown in Figures 3 and 4. The peak for the nominal mass of the molecular ion has shifted to m/z 172 in Figure 3, a shift of 14 Da representing the mass of a methylene unit (a carbon and two hydrogens). The multiplicity of the peaks at m/z 172, 174, and 176 are again in the ratio of 1 : 2 : 1 as they were in Figure 2 because these isotope peaks represent the presence of two bromine atoms.

The mass spectrum of 5,6-dibromo-8-hydroxyquinoline is presented in Figure 4, which shows a peak at m/z 301

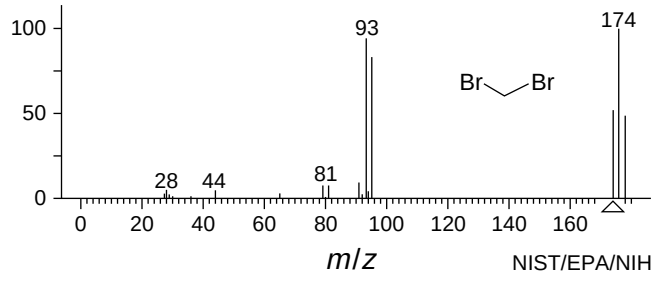

Figure 3 Mass spectrum of dibromomethane.

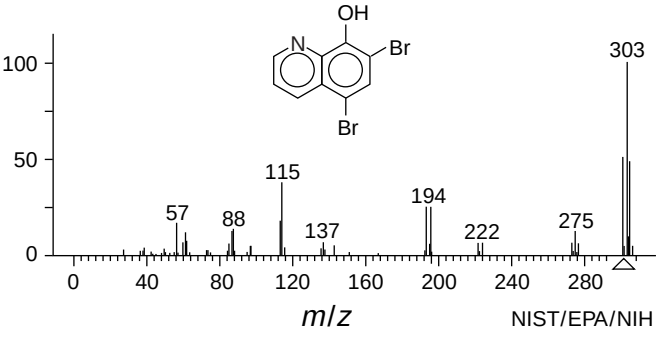

Figure 4 Mass spectrum of 5,6-dibromo-8-hydroxyquinoline.

for the nominal molecular ion. Again, the distinctive multiplicity of isotope peaks for two bromines is apparent with the peaks at m/z 301, 303, and 305 (as well as at m/z 273, 275, and 277) present in the ratio of 1 : 2 : 1.

The impact of the natural abundance of stable isotopes on the appearance of the mass spectral data has been introduced with bromine, an element that has the most abundant stable isotopes of those encountered in organic mass spectrometry. The impact of most of the other elements is much more subtle, but quite important from the standpoint of interpretation. The appearance of carbon can be noted in Figure 3 by the peaks at m/z 173, 175, and 177. The appearance of isotope peaks for carbon are much more apparent in Figure 4 where now the appearance of peaks at m/z 302, 304, and 306 are much more apparent. Note that the isotope peaks for carbon, which are due to the presence of ^{13}C, occur at one mass unit higher than the nominal mass of carbon (^{12}C), whereas for bromine the isotope peak occurred at two mass units above the nominal mass of bromine. Thus, bromine is called an "$X + 2$" element, where X refers to the nominal mass of the element and 2 refers to the number of mass units above the nominal mass that corresponds to the mass of the isotope. Therefore, carbon is an "$X + 1$" element because the isotope occurs one mass unit above the nominal mass.

The importance of isotope peaks in the interpretation of mass spectral data is that they provide limits to the elemental composition that should be considered for a given ion. The natural abundance of ^{13}C in ^{12}C is 1.1%. Thus, one can correlate the intensity of the peak at $X + 1$ with the number of carbon atoms present in the ion. For example, in Figure 3 the peak at m/z 173 is 1.1% of that at m/z 172, and that at m/z 175 is 1.1% of that at m/z 174, etc. In Figure 4, the intensity of the peak at m/z 302 is 9.9% of that at m/z 301, and that at m/z 304 is 9.9% of that at m/z 303, etc. The intensity of the isotope peaks at $X + 1$ in Figure 4 are more intense than those in Figure 3 because there are nine carbon atoms in dibromomethane and, thus, the probability of having at least one ^{13}C present in a molecule of dibromohydroxyquinoline is nine times greater than that in a molecule of dibromohydroxyquinoline, as illustrated in Figure 4 compared to Figure 3. That is to say, in a molecule of dibromohydroxyquinoline, the probability of having one ^{13}C present in the molecule is equal to 9 times 1.1% = 9.9%.

The natural abundances of stable isotopes of other elements that may be present in a given ion are listed in Table 1. The abundances of these stable isotopes at $X + 1$ and $X + 2$ are additive, and thus the contributions from ^{15}N and ^{13}C and ^{17}O all occur in the same $X + 1$ peak in the mass spectrum of a given molecule. However,

Table 1 Integer (nominal) masses and natural isotopic abundance for selected elements

Element	Symbol	Integer mass	Abundance
Hydrogen	H	1	99.99
	[D or ^{2}H]	2	0.01
Carbon	^{12}C	12	98.9
	^{13}C	13	1.1
Nitrogen	^{14}N	14	99.6
	^{15}N	15	0.4
Oxygen	^{16}O	16	99.76
	^{17}O	17	0.04
	^{18}O	18	0.20
Fluorine	F	19	100
Sulfur	^{32}S	32	95.02
	^{33}S	33	0.76
	^{34}S	34	4.22
Chlorine	^{35}Cl	35	75.77
	^{37}Cl	37	24.23
Bromine	^{79}Br	79	50.5
	^{81}Br	81	49.5
Iodine	I	127	100

as Table 1 shows, the contributions from ^{15}N and ^{17}O to $X+1$ are substantially less than the contribution of ^{13}C at $X+1$, thus, the intensity most often observed at $X+1$ will be due in large measure to carbon.

Some interesting and characteristic features of organic mass spectrometry can be noted from a cursory examination of the data in Table 1. Note that for all of the elements commonly encountered in organic mass spectrometry, the lightest isotope is always the most abundant. This feature is barely achieved by bromine for which the abundance of the lighter isotope is 50.5%. As indicated in the previous paragraphs, the natural abundance of ^{13}C is 1.1%, leaving 98.9% of the atoms present as ^{12}C. Whereas this number may seem small, compare it with the natural abundance of deuterium in nature, which is about one part in 10 000. Thus, even though carbon and hydrogen are present in large quantities in most organic molecules, the contribution from carbon to the stable isotope abundance is about two orders of magnitude greater than that from hydrogen. Note also in Table 1 that some elements have no stable isotopes, such as fluorine and iodine.

4 STRATEGY FOR DATA INTERPRETATION

Before any mass spectral data are interpreted in great detail, it is worthwhile determining whether the general quality of the data warrants a detailed investigation. For example, one should examine the data to be certain no artifactual peaks are present or that there are no gross errors in the mass scale calibration, or that there are no peaks that might represent superimposed mass spectra. A detailed examination of the mass spectra will give major consideration to at least four major classes of fragmentation that are described in the next sections.

In general, there are two major reasons for a detailed interpretation of the mass spectra: (a) to predict the likely appearance of peaks at given m/z values should the putative structure be present in the sample, or (b) to use the mass spectral information as a justification for the presence of a given structure in the sample. Either one of these reasons for interpreting the data rely heavily on the general appearance of the mass spectrum and on the mechanistic fragmentation processes described in the following sections. These are the language and nomenclature of mass spectrometry.

5 GENERAL FEATURES OF MASS SPECTRAL DATA

The relative abundance of the molecular ion in the mass spectrum can give a clue as to the type of compound that is undergoing analysis. Because the molecular ion of the compound is the mother of all other ions in the mass spectrum, its relative stability or its resistance to fragmentation suggests some features about its structural makeup. For example, a polyaromatic hydrocarbon resists fragmentation because, in general, more than two bonds must be broken to generate any fragment ions. Thus, the molecular ion peak in the mass spectrum of a polyaromatic hydrocarbon might represent a third or more of all the ion current represented by all the ions in the complete mass spectrum. This kind of information is borne out in Table 2 which shows an approximate relationship between the structure of the analyte and the

Table 2 Percentages of total ionization represented by molecular ion peak ($M^{+\bullet}$) versus compound type

Compound type	Examples	Approximate percentage of total ion current for $M^{+\bullet}$
Heterocyclic/polyaromatic	Estrone, anthracene	10–40
Aliphatic hydrocarbons, ketones, esters	Hexadecane, 4-decanone	0.5–2
Alphatic alcohols, highly branched compounds, alkylhalides	*n*-Decanol, chlorodecane, 2,2,4-Trimethylpentone	<0.01

fraction of the total ion current in the mass spectrum that represents the molecular ion.

For most of the compounds analyzed by mass spectrometry, only about 1% of the molecular ions will survive and be detected as such. Or, said differently, 99% of the molecular ions will decompose into fragment ions that will be represented by fragment ion peaks in the mass spectrum. In spite of the fact that 1% of the total ion current would seem to be a low number, such a molecular ion would generate a readily distinguishable peak in the mass spectrum. The compounds that create problems for the analyst when determining the identity of an unknown are the classes of compounds that typically generate a molecular ion that represents less than 0.1% of the total ion current. Alcohols are an infamous class of compounds from this point of view, in that they do not dependably provide a discernible peak in the mass spectrum that represents the molecular ion.

The nitrogen rule is a very important concept that is helpful in the interpretation of mass spectral data. The nitrogen rule states that any molecule containing carbon, hydrogen, oxygen, halogens, phosphorus, and an odd number of nitrogens will have an odd MW. Such a molecule containing all said elements, but with an even number of nitrogens (including zero) will have an even MW. This is easily demonstrated by simply considering the MW of methane (16 Da) and ammonia (17 Da). A corollary to the nitrogen rule is that any fragment ion that retains an even number of nitrogens will occur at an odd value of mass-to-charge ratio, whereas a fragment ion retaining an odd number of nitrogens will occur at an even value of mass-to-charge ratio, providing that said fragment ions are even-electron species (i.e. that they were formed by expelling a radical from an odd-electron precursor ion).

The fragmentation pattern present in a mass spectrum is established by the molecular ion and certain other fragment ions undergoing a fragmentation process in which logical groups of atoms are lost as a radical or a neutral molecular species. The number of atoms expelled in the fragmentation process must be reasonably consistent with valence considerations of the elements involved. For example, one would not expect a fragmentation process to proceed with the expulsion of a carbon atom or the loss of 12 Da from a given ion. Similarly, one would not expect the loss of 14 Da from a given precursor ion, which would suggest the loss of a nitrogen atom or the loss of a methylene group. This is because neither of these species is sufficiently stable to warrant serious consideration as a decomposition product. Similarly, the loss of 5 Da would also not be expected, as this would suggest the loss of five hydrogens. In summary, one would not expect a molecular ion to lose a group of atoms whose sum would correspond to more than three but less than 15 Da. However, losses of 1, 2 (for H_2), 15, 29, or 31 Da would be considered reasonable losses as these could easily correspond to $^{\bullet}H$, H_2, $^{\bullet}CH_3$, $^{\bullet}C_2H_5$, or $^{\bullet}OCH_3$, respectively.

6 CURSORY PRELIMINARY EXAMINATION OF MASS SPECTRAL DATA

A careful preliminary examination of the mass spectra may save the considerable frustration that might derive from attempts to interpret poorly calibrated or contaminated mass spectral data. The first major objective in interpreting a mass spectrum is to select a peak in the mass spectrum as a candidate for the molecular ion peak. In the circumstance where it is known that the analyte of interest contains an odd number of nitrogens, and the mass spectrum indicates a peak at highest mass that occurs at an even value of mass-to-charge, it should be suspected that something is wrong. Either the wrong sample has been analyzed, namely one containing a compound that contains an even number of nitrogens and thus would have an even MW, or the m/z scale has not been properly calibrated on the mass spectrometer. This is an example of the case in which the nitrogen rule helps to determine the possibility of bad mass spectral data.

The second step in the general scheme of data interpretation is to test the validity of a candidate molecular ion by determining whether it undergoes reasonable losses in a fragmentation process. If a fragment ion peak is found in the mass spectrum that indicates the candidate molecular ion loses 15 Da, you will have some confidence in the selection of the molecular ion because the loss of 15 Da is a reasonable loss. However, a peak 13 mass units below the peak of the candidate molecular ion should give reason for great concern as 13 Da is not a reasonable loss. This observation casts doubt on the selection of a correct peak for the molecular ion or suggests a contaminated mass spectrum in which the peak 13 mass units below the putative molecular ion represents an impurity.

With the advent of the combined use of gas chromatography and mass spectrometry, the likelihood of obtaining a contaminated mass spectrum or one that consists of the superimposed mass spectra of two or more compounds is relatively low. Such mixed spectra can be obtained, of course, from the mass spectrum of unresolved compounds or one represented by a shoulder of a contaminant peak in the chromatogram. It is also possible that the column bleed from the gas chromatograph could contaminate the mass spectrum; however, the column bleed will be constant and it can be subtracted from the mass spectrum and/or easily recognized and ignored, as shown in the following examples.

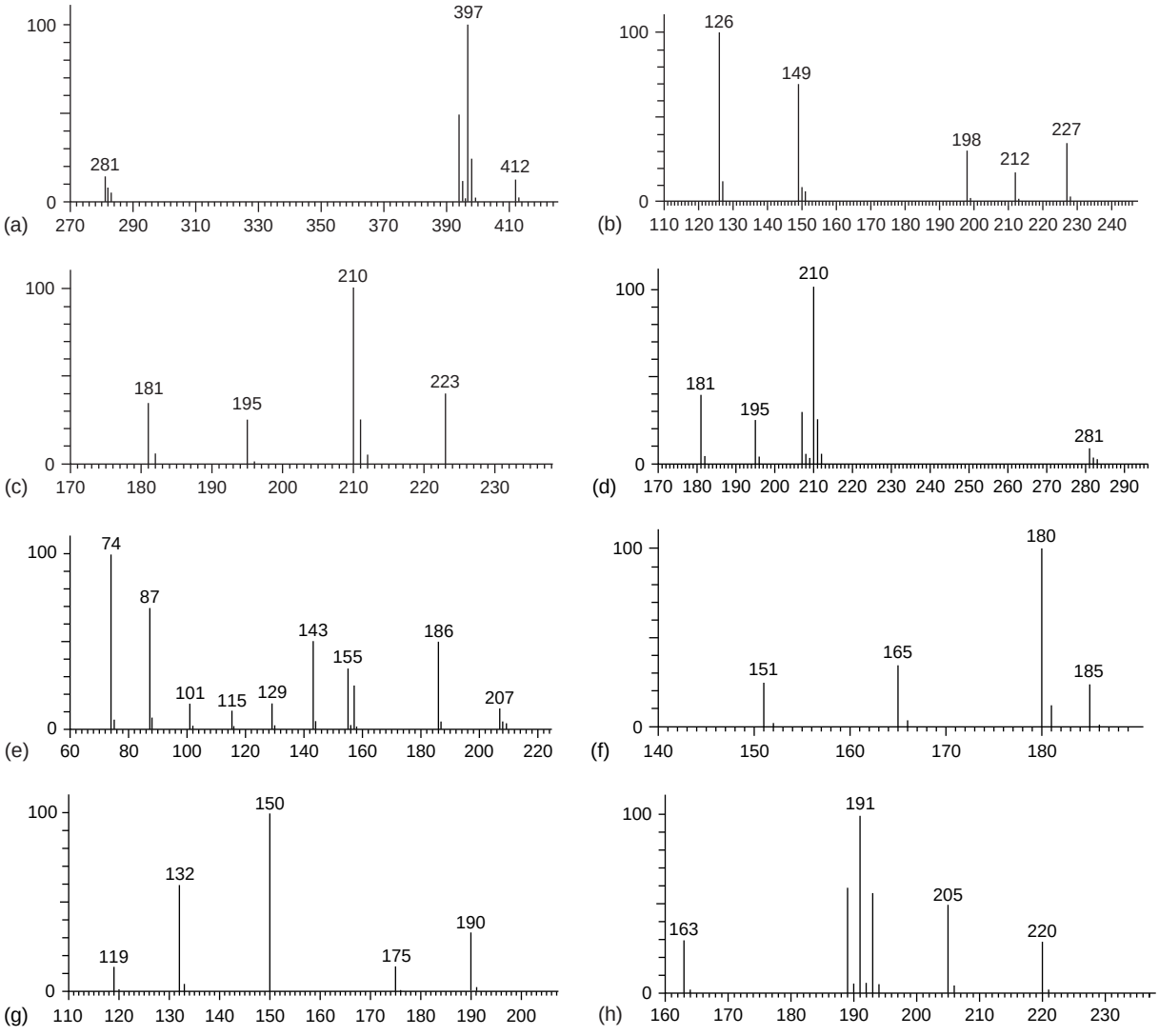

Figure 5 Abbreviated mass spectra of unknown compounds.

Now consider the mass spectra in Figure 5 from the point of view of making a preliminary examination of the data. In this cursory examination, assess which peak would represent the molecular ion, whether any suspicious peaks are present in the mass spectrum, and which peaks might be due to contaminants, etc. It would be helpful to do this as an exercise against which to compare the following commentary on each of the mass spectra, to help recognize certain details.

In Figure 5(a) the peak at m/z 412 is a good candidate for the molecular ion as it is consistent with the nitrogen rule for a compound containing no nitrogens, i.e. the peak at m/z 412 occurs at an even value. This candidate peak at m/z 412 is corroborated as a molecular ion peak by the presence of the peak at m/z 397, which could represent a fragment ion formed by the expulsion of a methyl radical, a reasonable loss. Furthermore, the peak at m/z 394 occurs 18 u lower than the peak at m/z 412 and represents the reasonable loss of water from the molecular ion. The peak at m/z 281 is likely to be background arising from column bleed from a siloxane column.

In Figure 5(b) the peak at m/z 227 is a candidate for a molecular ion peak in that it is a peak at highest m/z and, if it does indeed represent a molecular ion, it would represent a compound containing an odd number of nitrogens. The candidacy of the peak at m/z 227 as a molecular ion peak is strengthened by observing a peak at m/z 212 which could represent a fragment ion formed by the ion of m/z 227 losing a methyl group, a reasonable loss. The peak at m/z 198 also represents a reasonable loss from the candidate molecular ion at m/z 227; this would correspond to the loss of 29 Da as an ethyl radical. The peak at m/z 149 could be a part of the mass spectrum which has a molecular ion peak at

m/*z* 227, although seeing a peak at *m*/*z* 149 often suggests contamination of the sample with phthalates, a ubiquitous contaminant in most organic samples these days. The peak at *m*/*z* 126 also could be related to the compound of MW 227 Da in that it occurs at an even mass unit, which would suggest a fragment ion retaining the nitrogen.

In Figure 5(c) there is a suspicious peak at *m*/*z* 223; the strange feature is that this peak has no isotope peak at *m*/*z* 224. As all other peak clusters represented in this brief spectrum do show isotope peaks, it is likely that the peak at *m*/*z* 223 represents a noise spike in the particular spectrum and should be discounted unless it occurs in consecutive mass spectra. In any case, some explanation should be made for the fact that it does not have an isotope peak, and until such a reasonable explanation can be made, it should be ignored. Discounting the peak at *m*/*z* 223, the next peak of highest mass is at *m*/*z* 210, which would be a good candidate for the molecular ion of a compound containing no nitrogen. The suggestion that *m*/*z* 210 is a molecular ion peak is strengthened by observing a peak at *m*/*z* 195 which would correspond to an ion resulting from the loss of a methyl group from the compound. The peak at *m*/*z* 181 also strengthens the candidacy of the peak at *m*/*z* 210 as being a molecular ion in that the peak at *m*/*z* 181 probably represents the loss of an ethyl radical from the unknown compound.

Figure 5(d) contains peaks at *m*/*z* 207 and 281 that are in the expected ratio of 3 : 1 as would be present in the column bleed background from a siloxane stationary phase in GC/MS (gas chromatography/mass spectrometry) applications. Ignoring these peaks as background gives the same mass spectrum as occurs in unknown spectrum of Figure 5(c).

In Figure 5(e) the peak at highest mass is at *m*/*z* 207. Again, this peak is probably due to column bleed background, a feature that would be corroborated if it were possible to see whether there is a peak at *m*/*z* 281. In the meantime, the peak at *m*/*z* 207 can be ignored, and thus the peak at *m*/*z* 186 would be considered a candidate for the molecular ion peak for a species containing no nitrogens. The peaks at *m*/*z* 157 and 155 would correspond to the losses of 29 and 31 Da, respectively, which corroborate the candidacy of the peak at *m*/*z* 186 as a molecular ion peak. Other peaks continuing down the spectrum occurring at odd mass are consistent with the nitrogen rule except for the peak at *m*/*z* 74. This occurs at even *m*/*z* thereby suggesting an odd-electron ion which could either be due to a contaminant in the sample or, if proved to be a fragment ion, would suggest a rearrangement of the molecular ion of 186 Da. In fact, the peak at *m*/*z* 74 is taken to be a strong indication of the presence of the methyl ester of a fatty acid, the peak at *m*/*z* 74 being due to rearrangement about the carbomethoxyl group involving shift of a γ hydrogen with β-cleavage.

Figure 5(f) shows a peak at *m*/*z* 185 at high mass with another peak at *m*/*z* 180. The difference in mass between 185 and 180 is good reason to be suspicious, as such an increment does not correspond to a reasonable loss. Thus, the peak at *m*/*z* 185 does not represent the molecular ion of the same molecule that would have a fragment ion at *m*/*z* 180. This observation suggests that this unknown spectrum is that of a mixture of compounds with the peak at *m*/*z* 185 representing one compound and the peak at *m*/*z* 180 representing an ion of yet another compound. The peak at *m*/*z* 165 would appear to be related to the latter compound in that it probably corresponds to the loss of a methyl group from the compound of *m*/*z* 180. In any case, some effort should go into an assessment of the analytical procedure before much effort is put into interpreting the complete mass spectrum, as apparently some mixture of spectra is present.

Figure 5(g) has a peak at *m*/*z* 190 that would be a reasonable candidate for the molecular ion peak. It occurs at an even value of *m*/*z* suggesting the possibility of no nitrogen. Further, its candidacy as a molecular ion peak is corroborated by the presence of a peak at *m*/*z* 175, corresponding to the loss of a methyl group, which is a reasonable loss. The peak at *m*/*z* 150 is suspicious in that it has no accompanying isotope peak at *m*/*z* 151; some verification that this is a legitimate peak and not a noise spike should be pursued before much effort goes into interpretation of the whole spectrum. The peak at *m*/*z* 132 could be a molecular ion of a contaminant that contains no nitrogen or it could be a rearrangement fragment ion of the species of *m*/*z* 190, although the elimination of 58 Da is somewhat unusual. The peak at *m*/*z* 119 could be a fragment ion of the species of *m*/*z* 190 in that the peak at *m*/*z* 119 occurs at odd mass and is thereby consistent with the nitrogen rule.

In Figure 5(h) the peak at highest mass at *m*/*z* 220 is a candidate for the molecular ion peak of a species containing no nitrogen; this is corroborated by the peak at *m*/*z* 285 which would correspond to the loss of a methyl group, a reasonable loss of 15 Da. The next cluster of peaks is interesting in that it has the multiplicity of isotope peaks expected for an ion of nominal mass 189 containing two bromines. If this is true, the spectrum is indeed suspect as this observation would indicate the presence of at least two compounds, one containing bromine and one not containing bromine, the latter probably of MW 220 Da. The basis for this realization is that the peak at *m*/*z* 189 probably represents a fragment ion of a compound containing no nitrogen; however, it could be the molecular ion of a compound containing an odd number of nitrogens and two bromines, but is surely not related to the species at *m*/*z* 220 which does

not contain bromine. Thus, before any further time is invested in the interpretation of these data, it is important to determine whether this spectrum had been obtained under conditions where a substantial impurity was present or whether it had been obtained by GC/MS of species that were unresolved from one another.

7 DETAILED EXAMINATION/ RATIONALIZATION OF FRAGMENTATION

There are two general reasons for a detailed investigation of the fragmentation processes of molecular ions. One general use is to anticipate or predict the ways in which a given known compound is likely to fragment; for this purpose, try to anticipate peaks in the mass spectrum by which to recognize the presence of said compound in the sample. The other general use of fragmentation schemes is to interpret an unknown mass spectrum for purposes of deducing the structure of the compound present in the sample. The latter is much more difficult than the former, but both rely heavily on the use of the fragmentation schemes outlined below as these are the language of mass spectrometrists in describing structure elucidation.

As the molecular ion is the precursor of most of the fragment ions observed in the mass spectrum, some consideration should be given to the general form most often encountered for the molecular ion for certain classes of compounds. As described at the beginning of this article, the molecular ion is formed by expulsion of an electron from the neutral molecule to form an odd-electron molecular ion. An important question to contemplate is: which electron is expelled from the molecule?

For molecules containing a heteroatom (i.e. an atom other than a carbon or a hydrogen), one of the nonbonding electrons is most likely to be expelled in formation of the odd-electron molecular ion. The reason for this is that it is easier to remove an electron from a nonbonding orbital than it is from a σ-bonding orbital. In a hydrocarbon (namely a compound in which there is no heteroatom), the only choice is to remove an electron from a σ-bond unless there is a site of unsaturation, in which case it is likely that the electron would be removed from a π-orbital). In principle, any of the electrons in a molecule is available for expulsion in formation of the molecular ion. However, from an energetic point of view it requires less energy to expel the electron having the least binding energy, and thus most of the electrons that are lost in a molecule containing a heteroatom will come from the nonbonding orbital.

Because molecules that do contain a heteroatom also contain many bonds that are made up of σ-bonds, it is possible to have some of the molecular ions result from the expulsion of an electron from a sigma bond; however, these will be in the minority because they are expensive to produce in that it takes more energy to remove an electron from a σ-bond than it does from a nonbonding orbital. An example of ionization by removal of an electron from a nonbonding orbital is given in Scheme 2, which shows on the right the molecular ion as an odd-electron species with the "+" on the oxygen indicating the site from which the electron was lost. The same molecule is shown schematically in Scheme 3 which illustrates the process under which a molecular ion could be produced by σ-bond ionization; note that the electron deficiency in the molecular ion on the right of Scheme 3 occurs in the σ-bond between the R and the first methylene group. Although both molecular ions produced in Scheme 2 and Scheme 3 would be detected at the same mass-to-charge value, the one in Scheme 2 having the electron deficiency and the odd electron on the oxygen is most likely to represent the majority of the cases. In this article, (as in most mass spectrometry journals and books), the form of the molecular ion when a heteroatom is present is written as shown on the right-hand side of Scheme 2.

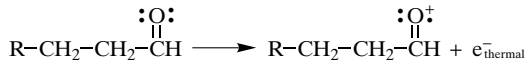

Scheme 2

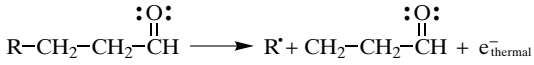

Scheme 3

The fragmentation process of any ion is driven by the thermodynamics involved in forming a yet more stable ion and/or a more stable radical by breaking a chemical bond. The energetics driving the fragmentation process can be represented as in Scheme 4 by imagining that the free energy of ion A plus the free energy of radical B are less than the free energy associated with the precursor molecular ion – in which case the process would be driven by the energy savings in going to the right. If the process given in Scheme 4 is favorable (i.e. a negative value for the change in free energy, ΔG), a peak will be seen in the mass spectrum that represents the ion A^+. The important feature to remember is that the bond in the molecular ion that breaks is not necessarily the weakest bond, but rather a bond that, when broken, leads to either a much more stable radical and/or a much more stable ion than the precursor ion.

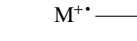

Scheme 4

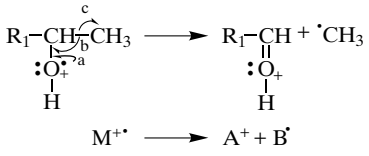

Scheme 5

Homolytic cleavage, as introduced schematically in Scheme 5, is frequently used to rationalize specific bond cleavages observed in the fragmentation of molecular ions of molecules containing heteroatoms. As can be seen in Scheme 5, the electron deficiency or radical site exists on the heteroatom, an oxygen in this case, in the molecular ion. In homolytic fission, it is the odd electron that initiates the process. The pairing of electrons leads to a more favorable or stable situation from an energetic point of view. Thus, in homolytic fission, a bond α to the atom adjacent to the atom containing the odd electron is broken in an effort to share yet another single electron with the existing odd electron that occurs on the oxygen in this case. In Scheme 5 a singly barbed arrow indicates the movement of a single electron. Thus, the single barbed arrow **a** shows the movement of the radical site, whereas the singly barbed arrow **b** shows the movement of one electron out of a two-electron bond to join the two electrons in a new bond between the oxygen and a carbon in species A^+. The species A^+ is now an even-electron ion, in this case, an oxonium ion, which is quite stable. The other product of this homolytic cleavage process is a radical species, $^{\bullet}B$, which has an odd electron. Thus, it may seem that the fragmentation was of no benefit as one of the species still has an odd electron. However, it must be remembered that if enough energy is lost by forming the very stable oxonium ion, then the fragmentation process was beneficial in an overall sense.

The process represented in Scheme 5 is called homolytic fission or cleavage because the bond that undergoes cleavage allows each of the two carbons that were connected by the two electrons to leave the fragmentation process with an electron; in this way, homolysis is an equitable dissociation process. The fragmentation process represented in Scheme 5 is also called α-cleavage because it involves cleavage of the bond α to the atom that is adjacent to the atom having the original problem with the odd electron, namely, the oxygen in this case.

Heterolytic cleavage is represented in Scheme 6 where, once again, the site of electron deficiency and an odd electron are on the heteroatom. In heterolytic fission, it is the charge site, rather than the radical site, that initiates the fragmentation process. Heterolytic cleavage differs from homolytic cleavage, in that heterolytic cleavage involves the movement of both electrons from a covalent bond as opposed to the movement of only a single electron from a covalent bond in homolytic fission. As is indicated in Scheme 6, the positive charge on the oxygen induces the movement of both electrons from the adjacent carbon–oxygen bond to move to the oxygen, and away from the carbon atom. This process results in a positive charge developing on the carbon atom as it now has

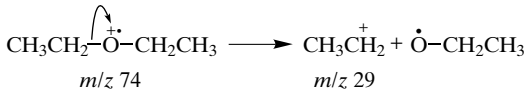

Scheme 6

only three devoted electrons around it, rather than the preferred number of four devoted electrons or eight shared electrons; thus, the ethyl ion will be represented by a peak at *m/z* 29 in the spectrum. In heterolytic fission, the movement of electrons is indicated by a doubly barbed arrow to represent the movement of two electrons, and the charge appears to move from its initial site on the heteroatom.

A rearrangement of a given ion involves the cleavage of at least two bonds, either in a concerted process or in a stepwise process resulting in the elimination of a neutral molecular species. In Scheme 7 the movement of two pairs of electrons is shown in concert, breaking two bonds in a rearrangement to expel an alkene, thereby forming a new even-electron ion of *m/z* 31 from an even-electron ion of *m/z* 59. A two-step rearrangement process is shown in Scheme 8 in which the odd-electron molecular ion first undergoes a movement of a hydrogen atom through a six-membered ring to form the distonic ion shown as species B. This then undergoes α-cleavage as driven by the radical site now on the γ-carbon to cleave a carbon–carbon bond leading to the formation and expulsion of an alkene, species D. Notice that the rearrangement process here formed a new odd-electron ion, C, from the odd-electron molecular ion.

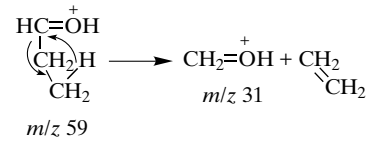

Scheme 7

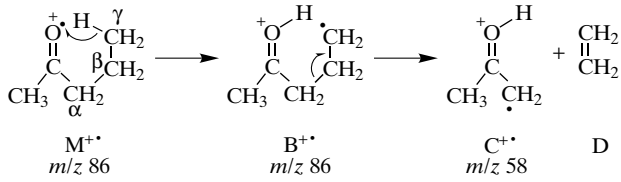

Scheme 8

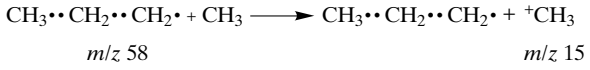

Scheme 9

Cleavage of a σ-bond can occur when ionization of the molecule occurs by removal of an electron from a σ-bond. Whereas σ-bond ionization can occur in any molecule, it is illustrated in Scheme 9 in the form of a hydrocarbon molecular ion (these kinds of molecules have no choice but to be ionized by σ-bond ionization). As can be seen on the molecular ion, the electron that was removed from the molecule was originally part of a pair of electrons between the terminal carbon on the right and the adjacent carbon such that the σ-bond here is now only a one-electron bond; thus, it is weaker than the other bonds in the molecule. If the molecular ion is to fragment, it will surely fragment between carbons 2 and 1, as this is now the weakest bond at this stage because it is only a one-electron bond. If the electronic structure of the molecular ion is as represented in Scheme 9, then cleavage of this carbon–carbon bond will leave the charge on the methyl group (detected to give a peak at m/z 15) and the odd electron will escape as a propyl radical (no signal, as the radical has no charge on it) as shown. If the other electron in the ionized bond had been removed during formation of the molecular ion, and this same bond were to fragment, then the charge would be on a propyl ion (m/z 43), and the odd electron would reside on the methyl group. Because there is nothing to distinguish one σ-bond from another in a hydrocarbon, the ionization of hydrocarbons causes a random distribution of molecular ions, having lost an electron to form a different σ-bond in each of several different forms of the molecular ion. The result is a mass spectrum in which there are peaks every 14 mass units, as shown subsequently in the mass spectrum of a long-chain hydrocarbon.

The mass spectrum of 4-decanone (Figure 1) is worthy of considerable attention as it contains peaks which represent ions formed by each of the four fragmentation processes introduced above. A cursory examination of the peaks in the mass spectrum indicate that, as is consistent with the nitrogen rule, the MW of this non nitrogen-containing molecule is even with a molecular ion peak shown at m/z 156. The peaks at lower mass-to-charge values mostly occur at odd values with the exception of the peaks at m/z 86 and m/z 58. Peaks occurring at an even value of mass-to-charge for a molecule containing an even number of nitrogens represent odd-electron ions. This is a clue that rearrangements must be possible in this molecule to allow the molecular ion, an odd-electron species, to eliminate a molecular species that, in turn, requires the formation of a different or new odd-electron species. Consecutive stages of two different rearrangements are indicated in Scheme 10 to rationalize the formation of ions represented by the peak at m/z 86 and that at m/z 58. Note in the beginning of Scheme 10 that the odd-electron on the oxygen in the molecular ion initiates the first rearrangement from the longer of the two available γ-hydrogen-containing side chains to eliminate an alkene to generate the species of m/z 86. The distonic ion shown at the upper right in Scheme 10 for the ion of m/z 86 can be rewritten as the radical cation shown in the lower right, which again can initiate, with its radical site, a rearrangement involving the propyl side chain to eliminate an alkene, namely ethylene, to form the distonic odd-electron ion of m/z 58.

Because the rearrangements illustrated in Scheme 10 have been studied so extensively since the early 1960s by F.W. McLafferty,[4] this general rearrangement has affectionately taken on the name of McLafferty rearrangement. However, this overall rearrangement process, which involves γ-hydrogen migration coupled with cleavage of the β-bond, was originally described by Nicholson,[5] who compared photolytic and EI decomposition products of aliphatic ketones.

Recognizing that the ketone contains a heteroatom, namely the oxygen, and that it is the most probable site of

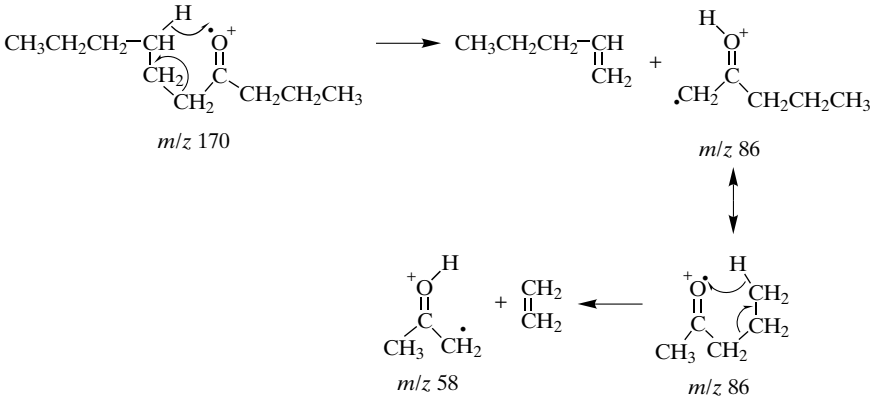

Scheme 10

ionization, it is expected that homolytic cleavage near the heteroatom would lead to the fragmentation illustrated in Scheme 11. This scheme illustrates that there can be α-cleavage on either side of the carbonyl, leading to the production of an even-electron acylium ion of m/z 113 in one case and m/z 71 in the other, both of which are represented by significant peaks in the mass spectrum.

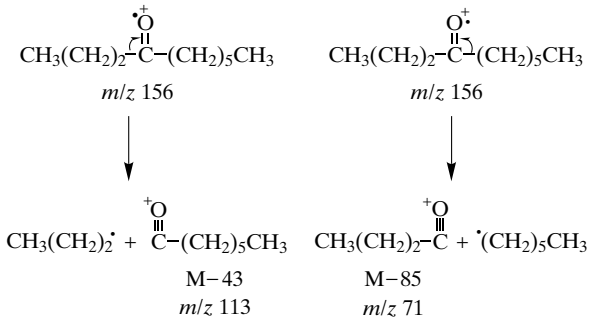

Scheme 11

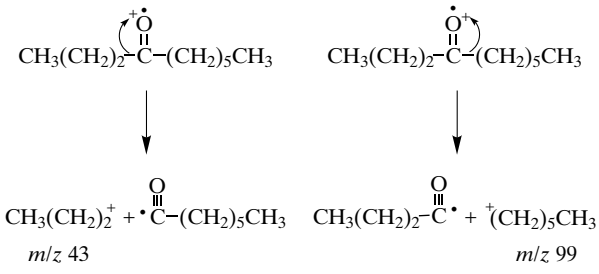

Scheme 12

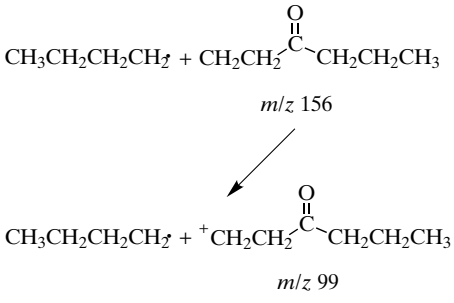

Scheme 13

Heterolytic fission is also a reasonable possibility and, as shown by the reactions in Scheme 12, the positive charge can induce heterolytic fission on either side of the carbonyl to generate ions of m/z 43 or m/z 99, both of which are represented by peaks in the mass spectrum.

Most of the peaks in the mass spectrum of 4-decanone (Figure 1) have been rationalized through the formation of ions involving three of the four fragmentation mechanisms. However, the peak at m/z 99 has not yet been explained, but σ-bond ionization followed by σ-bond cleavage is a rational route of fragmentation, as illustrated in Scheme 13.

8 REPRESENTATIVE MASS SPECTRA OF SELECTED CLASSES OF COMPOUNDS

8.1 Hydrocarbons

The mass spectra of aliphatic hydrocarbons are somewhat monotonous in that they contain a peak every 14 u, as illustrated in Figure 6. The molecular ion peak occurs at an even value of mass-to-charge, whereas all of the major fragment ion peaks occur at an odd value of mass-to-charge, consistent with the nitrogen rule. Because there is no structural feature in the aliphatic carbon backbone that distinguishes one σ-bond from another, the molecular ions of these hydrocarbons consist of a random mixture of σ-bond-ionized molecular ions, each of which can expel an alkyl radical to give a peak every 14 u. It is important to realize that the peaks occur every 14 u because the molecular ion has the option of expelling homologous alkyl radicals, that is, those differing by 14 Da, depending on which of the many carbon–carbon σ-bonds was the site of ionization; the peaks do not represent the loss of 14 Da (which would correspond to a methylene group) from one fragment to the next.

The mass spectra of aromatic hydrocarbons are vastly different in appearance from those of the aliphatic hydrocarbons as can be seen from a comparison between Figure 7 and Figure 6. The molecular ion of a polyaromatic hydrocarbon is quite resistant to fragmentation as more than one bond must be broken to produce a fragment ion. Thus, as indicated in the mass spectrum of anthracene (Figure 7), the molecular ion peak represents nearly half of all the ion current in the mass spectrum.

The mass spectra of other aromatic hydrocarbons, such as the phenyl-containing hydrocarbons, also have characteristic features. For example, if the phenyl ring is connected to the rest of the molecule via an aliphatic carbon chain, the aromatic component is frequently lost

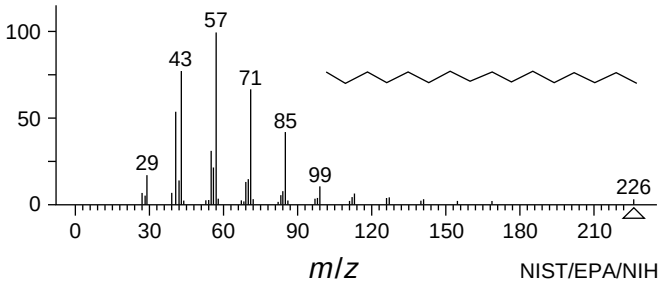

Figure 6 Mass spectrum of hexadecane.

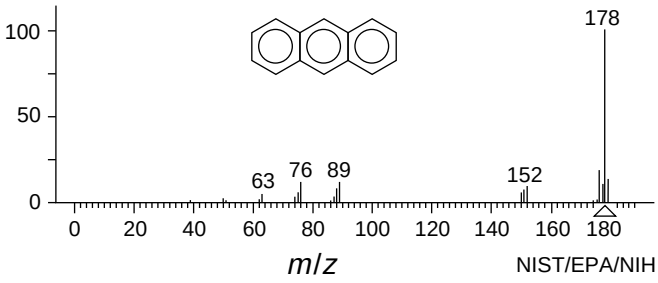

Figure 7 Mass spectrum of anthracene.

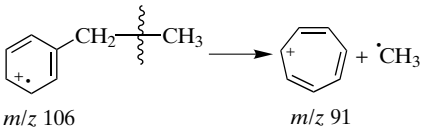

Scheme 14

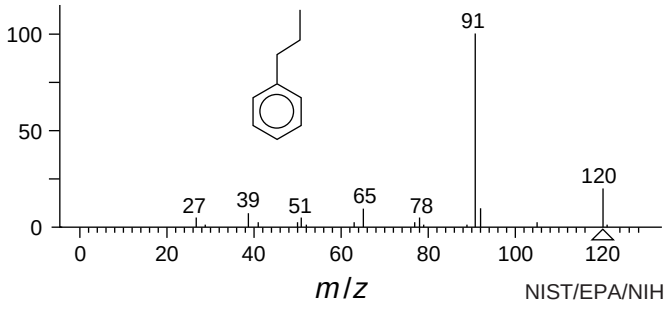

Figure 8 Mass spectrum of propylbenzene.

as a fragment ion, although not due to cleavage directly at the phenyl ring, but rather one methylene unit away as indicated in Scheme 14. Formation of the tropylium ion is favorable as this even-electron species is remarkably stable and is represented by an intense peak in the mass spectrum at *m*/*z* 91 as indicated in Figure 8.

8.2 Alcohols

Aliphatic alcohols pose a serious problem for the mass spectrometrist in that these compounds do not reliably provide a discernible peak for the molecular ion. Thus, the interpretation of the mass spectrum of an alcohol is frequently in jeopardy because there is no indication of the MW of the alcohol. On the other hand, aliphatic alcohols do produce a characteristic peak at *m*/*z* 31, which is a good indicator for the presence of these compounds.

Aliphatic alcohols undergo characteristic fragmentation to generate an ion of *m*/*z* 31 as illustrated in Scheme 15. Note that as usual the radical site is on the heteroatom, namely the oxygen in this case, which promotes homolytic fission to generate the oxonium ion of *m*/*z* 31. Although a peak at *m*/*z* 31 in the mass spectrum of an aliphatic alcohol is rarely the base peak, a peak of any intensity at *m*/*z* 31 is a good indicator of the presence of an alcohol.

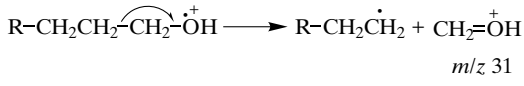

Scheme 15

The mass spectrum of *n*-decanol shows the characteristic features of an aliphatic alcohol (Figure 9). Namely, there is no reliable indication of the molecular ion; note that the MW of *n*-decanol is 158 Da, but there is no discernible peak at *m*/*z* 158 in the mass spectrum. The peak at *m*/*z* 31 is certainly not very intense, but its presence is a reliable indicator of an alcohol. The principal peaks in the mass spectrum are those that represent fragmentation of the hydrocarbon backbone of the aliphatic alcohol.

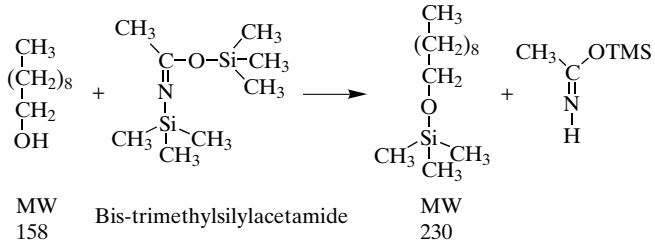

Scheme 16

Recognizing that aliphatic alcohols are not well behaved from the standpoint of providing readily interpretable mass spectra, it is frequently desirable to convert the alcohol to another compound which is better behaved during EI mass spectrometry. Conversion of aliphatic alcohols to trimethylsilyl (TMS) ethers, as shown in Scheme 16 for *n*-decanol, is frequently used to characterize a given aliphatic alcohol. As will be seen in the following paragraphs, the TMS ether gives peaks in the mass spectrum from which one can determine the MW of the ether, and thus the MW of the parent alcohol,

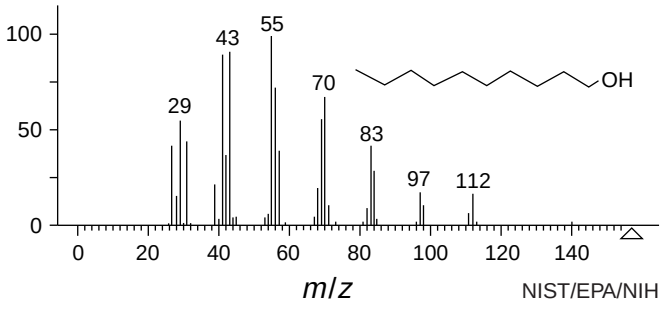

Figure 9 Mass spectrum of *n*-decanol.

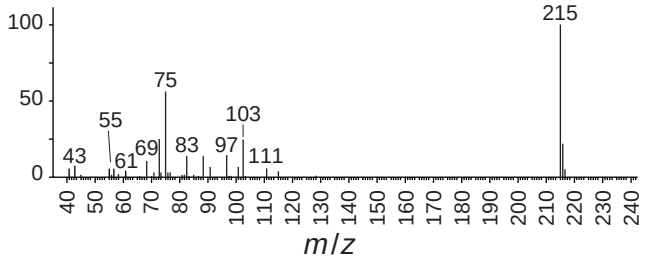

Figure 10 Mass spectrum of the TMS of *n*-decanol.

in addition to learning about other structural features of the alcohol.

Whereas TMS ethers sometimes show a peak in the mass spectrum for the molecular ion, it is clear from examining the mass spectrum of the TMS ether of decanol in Figure 10 that this is not always the case. Nevertheless, the substantial peak at *m*/*z* 215 in Figure 10 gives an indirect indication of the MW of the alcohol because the peak at *m*/*z* 215 merely represents the loss of a methyl group, which is not part of the parent structure of the alcohol. The trained analyst would know the characteristic fragmentation behavior of the TMS ether derivative in that they readily lose a methyl radical, as shown in Scheme 17 and, with this knowledge, the peak at high mass which occurs at an odd value of mass-to-charge would be recognized as such a fragment ion.

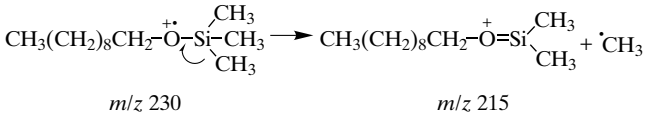

Scheme 17

In summary, by converting aliphatic alcohols to aliphatic ethers, the parent structure of the alcohol is preserved and is represented by high-mass peaks in the mass spectrum of the corresponding ether. The mass spectrum of an ether may show either peaks for both the molecular ion and a fragment ion corresponding to the loss of methyl group, or a peak at high mass corresponding to the loss of the methyl group only; either situation allows the analyst to obtain the critically important information concerning the MW of the alcohol. Note that in the chemical reaction in Scheme 16, the hydrogen on the alcohol group is substituted for a TMS group; the TMS group has a mass of 73 Da, but it replaces a hydrogen which has a mass of 1 Da, thus the shift in mass is $73 - 1 = 72$ Da.

The mass spectra of aromatic alcohols (phenols), are just the opposite in appearance from those of the aliphatic alcohols. Note in the mass spectrum of dichlorophenol in Figure 11, that the molecular ion peak is the base peak.

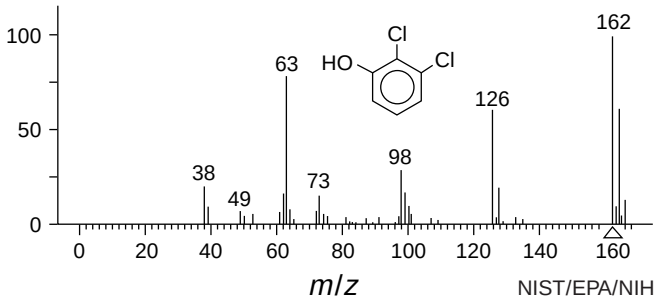

Figure 11 Mass spectrum of dichlorophenol.

This indicates that the molecular ion of phenols is very stable, resisting fragmentation. Note also in Figure 11 that the isotope peaks *m*/*z* 164 and *m*/*z* 166 are in the ratio expected for two chlorine atoms being present in the molecule. This attribute, as well as the fact the phenols are characterized by intense molecular ion peaks, have been used to advantage in environmental chemistry analyses for several years.

8.3 Esters of Carboxylic Acids

The methyl esters of long-chain carboxylic acids undergo characteristic rearrangement involving migration of a γ-hydrogen with β-cleavage as shown in Scheme 18, which leads to an odd-electron fragment ion of *m*/*z* 74. This ion

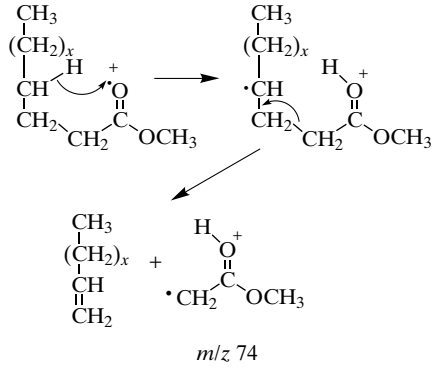

Scheme 18

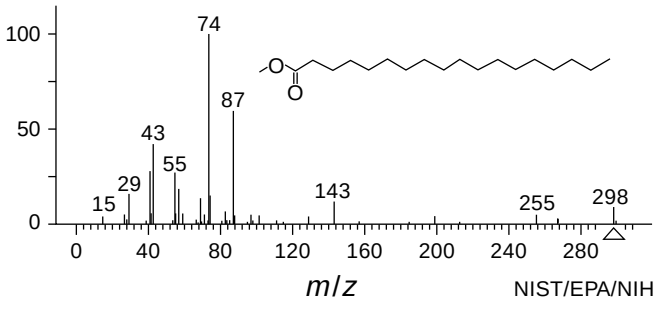

Figure 12 Mass spectrum of methylstearate.

is represented in the mass spectrum of methylstearate, Figure 12, by the base peak. Note also in Figure 12 that not only does the molecular ion peak occur at an even value of mass-to-charge, namely at *m*/*z* 298, thereby representing an odd-electron species, but also the rearrangement fragment ion occurs at an even value of mass-to-charge (*m*/*z* 74). The rearrangement ion of *m*/*z* 74 is also an odd-electron species as can be seen from the electron accounting shown in Scheme 18. Thus, in accordance with the nitrogen rule, odd-electron ions for compounds containing an even number of nitrogens occur at an even value of mass as verified in Figure 12 and also earlier in Figure 1, which is the mass spectrum of 4-decanone that had two peaks at an even mass-to-charge ratio, namely those at *m*/*z* 58 and at *m*/*z* 86. The rearrangement ion of *m*/*z* 74 for methyl esters may not always be represented by the base peak in the mass spectrum as it is in Figure 12; however, seeing a peak at *m*/*z* 74 regardless of its intensity should be suggestive to the analyst of the presence of a methyl ester of a carboxylic acid.

8.4 Amines

The fragmentation of an amine is strongly influenced by the "+•" on the nitrogen in the molecular ion. A cursory inspection of the aliphatic secondary amine in Figure 13 illustrates agreement of the nitrogen rule with the mass spectrum of a compound containing an odd number of nitrogens. Note that the molecular ion peak occurs at an odd mass-to-charge value, namely *m*/*z* 101. All the major fragment ion peaks occur at an even value of mass-to-charge value as it will be shown that the ion represented by these peaks retain the nitrogen atom. The fragmentation scheme illustrated in Scheme 19 explains losses that correspond to the loss of an ethyl radical followed by a rearrangement of the even-electron ion of *m*/*z* 72 to generate another odd-electron ion of *m*/*z* 30. One should not get the impression from Scheme 19 that all the even-electron ions of *m*/*z* 72 undergo further rearrangement to eliminate the alkene, because many

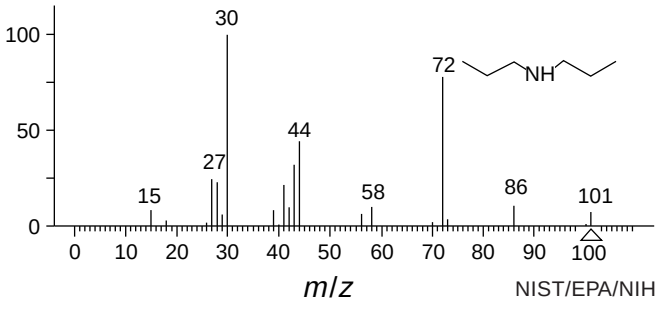

Figure 13 Mass spectrum of propylpropanamine.

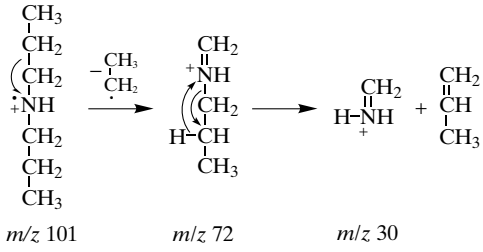

Scheme 19

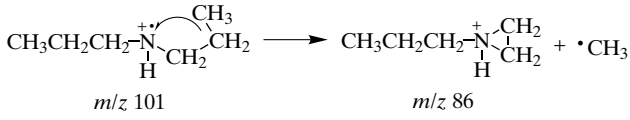

Scheme 20

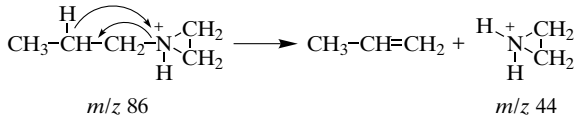

Scheme 21

of the ions of *m*/*z* 72 are sufficiently stable to survive long enough to reach the mass analyzer and are detected as a fragment ion having lost only an ethyl radical from the molecular ion.

The peak at *m*/*z* 86 is consistent with the loss of a methyl radical, which can be rationalized by homolytic cleavage as illustrated in Scheme 20.

The peak at *m*/*z* 44 represents an even-electron fragment ion that can be rationalized by the fragment ion of *m*/*z* 86 undergoing a rearrangement, as illustrated in Scheme 21 to eliminate an alkene.

9 SUMMARY

This article shows that the fragmentation pattern or the mass spectrum of various classes of compounds contain unique features that help identify the compound type. In addition, four different classes of fragmentation have been presented, and examples shown of the way in which these fragmentation schemes can help rationalize the observed mass spectral data to correlate with or suggest a particular structural feature of the molecule. Consistency of various rules, such as the nitrogen rule, allows one to recognize the integrity of the mass spectral data. The impact of stable isotopes as manifested by the multiplicity of isotope peaks in the mass spectrum allows one to deduce a reasonable estimate of the elemental composition of the ions in the mass spectrum. The elemental composition of the ions as indicated by isotope peaks, and the fragmentation of the molecule after ionization provide a chemical fingerprint of the compound

that is invaluable from the standpoint of structural elucidation or selective detection of the compound by mass spectrometry.

ABBREVIATIONS AND ACRONYMS

EI	Electron Ionization
GC/MS	Gas Chromatography/Mass Spectrometry
MW	Molecular Weight
TMS	Trimethylsilyl

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Environment: Water and Waste **(Volume 4)**
Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Mass Spectrometry **(Volume 13)**
Gas Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

REFERENCES

1. Q.N. Porter, *Mass Spectrometry of Heterocyclic Compounds*, 2nd edition, J. Wiley and Sons, New York, 1985.
2. J.S. Splitter, F. Turecek, *Applications of Mass Spectrometry to Organic Stereochemistry*, VCH, New York, 1994.
3. J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott Raven, 1997.
4. F.W. McLafferty, F. Turecek, *Interpretation of Mass Spectra*, 4th edition, University Science Books, Mill Valley, California, 1993.
5. A.J.C. Nicholson, 'Photochemical Decomposition of Aliphatic Methyl Ketones', *Trans. Faraday Soc.*, **50**, 1067–1073 (1954).

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

Alan G. Marshall, Christopher L. Hendrickson, and George S. Jackson
Florida State University, Tallahassee, FL, USA

1 Introduction **2**
2 Ion Cyclotron Motion: Ion Cyclotron Orbital Frequency, Radius, Velocity, and Energy **2**
3 Excitation and Detection of an Ion Cyclotron Resonance Signal **5**
3.1 Azimuthal Dipolar Single-frequency Excitation 5
3.2 Azimuthal Dipolar Single-frequency Detection 6
3.3 Broadband Excitation 7
3.4 Broadband Detection, Detection Limit 9
4 Ion-neutral Collisions **9**
5 Effects of Axial Confinement of Ions in a Trap of Finite Size **10**
5.1 Axial Ion Oscillation Due to the z Component of the Electrostatic Trapping Potential 10
5.2 Radial Ion Magnetron Rotation Due to Combination of the Magnetic Field and the Radial Component of the Electrostatic Trapping Potential 12
5.3 Mass Calibration 13
6 Quadrupolar Excitation, Axialization, and Ion Remeasurement **14**
7 Effect of Trap Size and Shape on Dipolar and Two-dimensional Quadrupolar Excitation **15**
8 Mass Resolving Power, Mass Resolution, and Mass Accuracy **16**
9 Upper Mass and Energy Limit(s) **17**
9.1 Trap Dimension Mass Limit 17
9.2 Trapping Potential Mass Limit 17
9.3 Trapping Potential Energy Limit; Space Charge and the Peak Coalescence Mass Limit 18
10 Ion Sources **19**
10.1 Internal Ionization 19
10.2 External Ionization 19
10.3 Chromatographic Interfaces 19
11 Tandem Mass Spectrometry Methods **20**
12 Advantages of High Magnetic Fields **20**
13 Fourier Transform Spectroscopy Aspects **21**
14 Relation to Paul (Quadrupole) Ion Trap **22**
15 Selected Applications **22**
15.1 Elemental Composition from Accurate Mass Measurement 22
15.2 Detection Limit for Biological Analysis 22
15.3 High Mass 22
15.4 Isotopic Amplification for Unit Mass Accuracy of Biomacromolecules 23
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **24**
Appendix 1: Physical Constants and Precise Elemental Masses **25**
Physical Constants 25
Precise Masses of Various Common Elements 26
Further Reading **26**
Books and Journal Articles on Fourier Transform Ion Cyclotron Resonance 26
Special Journal Issues on Fourier Transform Ion Cyclotron Resonance 26
Compilations 26
Fourier Transform Techniques 26
Early History of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry 26
Ion Cyclotron Resonance Excitation 26
Ion Cyclotron Resonance Signal Generation and Detection 26
Ion Cyclotron Resonance Line Shape at Low, Intermediate, and High Pressure 26
Stored Waveform Inverse Fourier Transform Excitation 27
Penning Traps 27
External Ion Injection 27
Ion–Molecule Chemistry from Fourier Transform Ion Cyclotron Resonance Mass Spectrometry 27
Quadrupolar Axialization 27
Two-dimensional Ion Cyclotron Resonance 27

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

High Mass Fourier Transform Ion Cyclotron Resonance Mass Spectrometry 27
Capillary Electrophoresis Fourier Transform Ion Cyclotron Resonance Mass Spectrometry 27
Reviews 27
References 29

This review offers an introduction to the principles and generic applications of Fourier transform ion cyclotron resonance/mass spectrometry (FTICR/MS), directed to readers with no prior experience with the technique. The fundamental Fourier transform ion cyclotron resonance (FTICR) phenomenon is explained from a simplified theoretical treatment of ion behavior in idealized magnetic and electric fields. The effects of trapping voltage, trap size and shape, and other nonidealities are then manifested mainly as perturbations that preserve the idealized ion behavior modified by appropriate numerical correction factors. Topics include: effect of ion mass, charge, magnetic field, and trapping voltage on ion cyclotron frequency; excitation and detection of ion cyclotron resonance (ICR) signals; mass calibration; mass resolving power and mass accuracy; upper mass limit(s); dynamic range; detection limit, strategies for mass and energy selection for multiple mass spectroscopy (MS^n); ion axialization, cooling, and remeasurement; and means for guiding externally formed ions into the ion trap. The relationship of FTICR/MS to other types of Fourier transform (FT) spectroscopy and to the Paul (quadrupole) ion trap is described. The article concludes with selected applications, an appendix listing accurate fundamental constants needed for ultrahigh-precision analysis, and an annotated list of selected reviews and primary source publications describing various FTICR/MS techniques and applications in fuller detail.

1 INTRODUCTION

FTICR/MS or FTMS is widely practiced, with more than 235 installations worldwide by 1998. In the 24 years since its inception,[1–4] FTICR/MS has been the sole or principal subject of three books, four journal special issues, and more than 60 review articles (see Further Reading). However, the various prior descriptions and reviews are distributed among dozens of primary publications with different formalism and notation, and those papers are directed mainly at FTICR/MS practitioners. Here, an introduction is given to the principles and generic applications of FTICR/MS, at the minimum depth needed to explain the concepts, directed to those without prior experience in the field.

Fortunately, it turns out that many fundamental aspects of FTICR can be understood from very simple idealized models. First, ion cyclotron frequency, radius, velocity, and energy as a function of ion mass, ion charge, and magnetic field strength follow directly from the motion of an ion in a spatially uniform static magnetic field. Second, ion cyclotron motion may be rendered spatially coherent (and thus observable) by application of a spatially uniform radiofrequency (RF) electric field (excitation) at the same frequency as, i.e. resonant with, the ion cyclotron frequency. The ICR (time-domain) signal results from induction (detection) of an oscillating image charge on two conductive infinitely extended opposed parallel electrodes. A frequency-domain spectrum, convertible to a mass-domain spectrum, is obtained by Fourier transformation of the digitized time-domain ICR signal. Third, confinement of ions by application of a three-dimensional axial quadrupolar direct current (DC) electric field shifts the ion cyclotron frequency, whereas excitation and detection remain essentially linear (i.e. doubling the excitation amplitude doubles the detected ICR signal) but with a reduced proportionality constant. A simple mass calibration rule follows from this treatment. Thus, FTICR/MS may be performed in essentially the same way in ion traps of widely different shape (e.g. cubic, cylindrical). Fourth, collisions broaden the ICR signal in a simple way, and actually make it possible to cool and compress an ion packet for improved detection (and even multiple remeasurement). Fifth, although FTICR/MS has been coupled to virtually every type of ion source, most ion sources work best (or at least most conveniently) outside the magnet. Thus, several methods have been developed to guide externally generated ions into an ion trap inside a high-field magnet. Finally, the above features may be combined in various experimental event sequences (much like pulse sequences in FTNMR (Fourier transform nuclear magnetic resonance) spectroscopy) to perform tandem mass spectrometry (MS/MS, or MS^n).

2 ION CYCLOTRON MOTION: ION CYCLOTRON ORBITAL FREQUENCY, RADIUS, VELOCITY, AND ENERGY

An ion moving in the presence of a spatially uniform magnetic field, $\mathbf{B} = -B_0\mathbf{k}$ (i.e. the z-axis is defined as the direction opposite to B) is subject to a force given (in SI units) by Equation (1),

$$\text{force} = \text{mass} \times \text{acceleration} = m\frac{d\mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B} \quad (1)$$

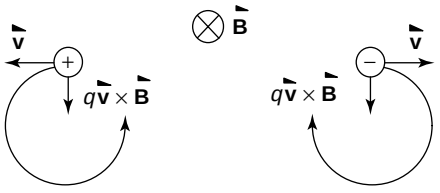

Figure 1 Ion cyclotron motion. The path of an ion moving in the plane of the paper is bent into a circle (see text) by the inward-directed Lorentz magnetic force produced by a magnetic field directed perpendicular to the plane of the paper. Note that positive and negative ions orbit in opposite senses. (Reproduced with permission from Marshall and Grosshans, 1991.[5])

in which m, q, and $\mathbf{v}$ are ionic mass, charge, and velocity, and the vector cross-product term means that the direction of the magnetic component of the Lorentz force is perpendicular to the plane determined by $\mathbf{v}$ and $\mathbf{B}$. If the ion maintains constant speed (i.e. no collisions), then the magnetic field bends the ion path into a circle of radius r (Figure 1).

Let $v_{xy} = \sqrt{v_x^2 + v_y^2}$ denote the ion velocity in the xy plane (i.e. the plane perpendicular to $\mathbf{B}$). Because angular acceleration, $|\mathrm{d}\mathbf{v}/\mathrm{d}t| = v_{xy}^2/r$, Equation (1) then becomes Equation (2)

$$\frac{mv_{xy}^2}{r} = qv_{xy}B_0 \tag{2}$$

However, the angular velocity w (in radians per second) about the z-axis is defined by Equation (3),

$$w = \frac{v_{xy}}{r} \tag{3}$$

so that Equation (2) becomes Equation (4),

$$mw^2r = qB_0wr \tag{4}$$

or simply

$$w_c = \frac{qB_0}{m} \quad \text{(SI units)} \tag{5}$$

Equation (5) can be expressed as Equation (6),

$$\nu_c = \frac{w_c}{2\pi} = \frac{1.535611 \times 10^7 B_0}{m/z} \tag{6}$$

where ν_c is in hertz, B_0 in tesla, m in unified atomic mass units, and z in multiples of elementary charge.

Equation (5) is the celebrated cyclotron equation, in which the unperturbed ion cyclotron frequency is denoted as w_c. A remarkable feature of Equation (5) is that all ions of a given mass-to-charge ratio m/q, have the same ICR frequency, independent of their velocity. This property makes ICR especially useful for mass spectrometry (MS),

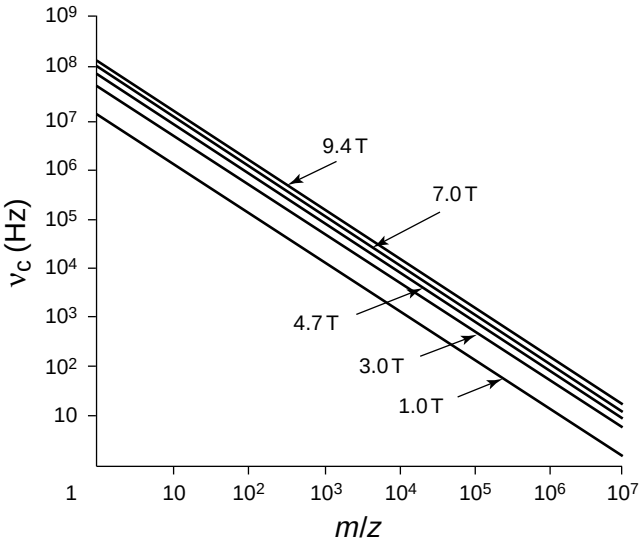

Figure 2 The ICR orbital frequency, $\nu_c = w_c/2\pi$, (in hertz), as a function of ionic mass-to-charge ratio, m/z, in daltons per multiple (z) of the elementary electron charge, at each of five magnetic field strengths: 1.0, 3.0, 4.7, 7.0 and 9.4 T. Note that ICR frequencies for ions in the usual chemical mass-to-charge ratio range ($\sim 15 \leq m/z \leq 10\,000$) typically lie between a few kilohertz and a few megahertz.

because translational energy focusing is not essential for precise determination of m/q.

Several useful conclusions follow directly from Equation (5). First, at a representative static magnetic field value of 7.0 T (at which the corresponding proton nuclear magnetic resonance (NMR) Larmor frequency would be 300 MHz), ICR frequencies for ions formed from typical molecules range from a few kilohertz to a few megahertz (Figure 2), a very convenient range for commercially available electronics.

Rearrangement of Equation (2) yields the ion cyclotron orbital radius of an ion of velocity v_{xy} (Equation 7),

$$r = \frac{mv_{xy}}{qB_0} \quad \text{(SI units)} \tag{7}$$

or

$$r = \frac{1.036427 \times 10^{-8}(m/z)v_{xy}}{B_0} \tag{8}$$

In Equation (8) r is in meters, v_{xy} in meters per second, B_0 in tesla, m in unified atomic mass units, and z in multiples of elementary charge. For example, the average x–y translational energy of an ion in equilibrium with its surroundings at a temperature T (in kelvin), is given by

$$\frac{m\langle v_{xy}^2\rangle}{2} \approx kT \tag{9}$$

in which k is the Boltzmann constant. Solving Equation (9) for v_{xy}, and substituting back into Equation (7),

yields Equation (10),

$$r = \frac{1}{qB_0}\sqrt{2mkT} \quad \text{(SI units)} \tag{10}$$

which gives

$$r = \frac{1.336\,510 \times 10^{-6}}{zB_0}\sqrt{mT} \tag{11}$$

In Equation (11) r is in meters, B_0 in tesla, m in unified atomic mass units, T in kelvin, and z in multiples of elementary charge.

It is quickly confirmed that at room temperature, a singly charged ion of $m = 100$ u in a magnetic field of 3 T (i.e. 30 000 G) has an ICR orbital radius of ~0.08 mm. At room temperature, a singly charged ion of mass 10 000 u has an ICR orbital radius of only ~0.8 mm, whereas a singly charged ion of 50 000 u has an ICR orbital radius of 1 cm. Thus, ions of thermal energy formed from all but the largest molecules are confined by the magnetic field to conveniently small orbital radii for ICR excitation and detection (Figure 3).

Conversely, we can compute the velocity and translational energy of an ion excited (by as yet unspecified means) to a larger-than-thermal orbital radius. Rearrangement of Equation (7) gives Equation (12),

$$v_{xy} = \frac{qB_0 r}{m} \quad \text{(SI units)} \tag{12}$$

which may also be expressed as Equation (13):

$$v_{xy} = \frac{9.64853 \times 10^7 B_0 r}{m/z} \tag{13}$$

where v_{xy} is in meters per second, r in meters, B_0 in tesla, m in unified atomic mass units, and z in multiples of elementary charge.

From Equation (12), a singly charged ion of 100 u excited to an ICR orbital radius of 1 cm in a magnetic field of 3 T has a translational velocity v_{xy} of 2.97×10^4 m s^{-1}, corresponding to a translational energy of 434 eV (Equations 14 and 15):

$$\text{kinetic energy} = \frac{mv_{xy}^2}{2} = \frac{q^2 B_0^2 r^2}{2m} \quad \text{(SI units)} \tag{14}$$

or

$$\text{kinetic energy} = \frac{4.824\,265 \times 10^7 z^2 B_0^2 r^2}{m} \tag{15}$$

where the kinetic energy is in electron volts, r in meters, B_0 in tesla, m in unified atomic mass units, and z in multiples of elementary charge. Thus, ions can be heated to high translational energy even in a relatively small container, and then induced to break into smaller fragments by collision with neutral gas molecules (see below). The dependence of the ion kinetic energy on ICR orbital radius and magnetic field strength is shown in Figure 4.

Finally, it is instructive to note that the excited ion of 100 u of the preceding paragraph travels a distance of about 30 km during a 1 s observation period! That's basically why ICR offers potentially much higher mass

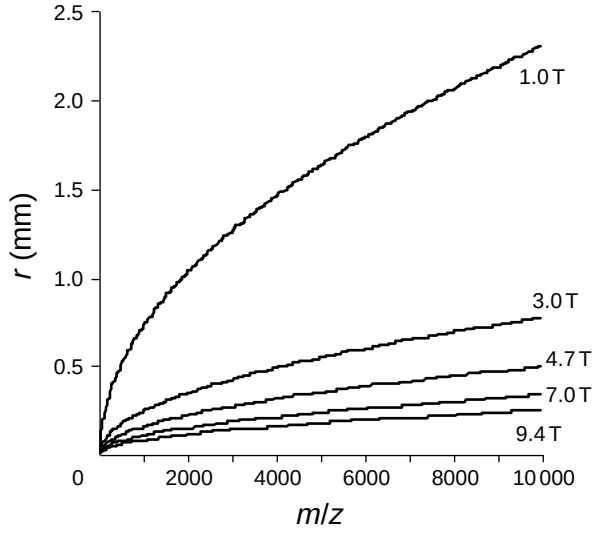

Figure 3 The ICR orbital radius r (Equation 10), versus ionic mass-to-charge ratio, m/z (in unified atomic mass units per multiple of elementary electronic charge) at each of five representative magnetic field strengths: 1.0, 3.0, 4.7, 7.0, and 9.4 T at 298K. Note that even relatively heavy ions are confined to conveniently small-radius orbits by such magnetic fields.

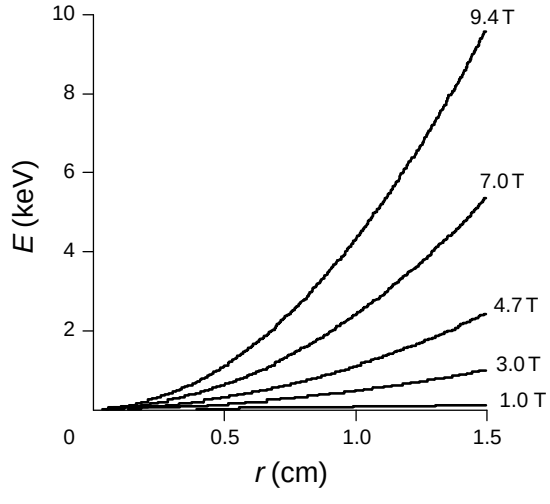

Figure 4 Ion translational (kinetic) energy (Equations 14 and 15) as a function of ICR orbital radius, at each of five common magnetic field strengths: 1.0, 3.0, 4.7, 7.0, and 9.4 T, for an ion of $m/z = 100$. Note that ions may be accelerated to relatively high energy (≥1 keV, as in collision-induced dissociation (CID) experiments) while still confined to relatively small (~1 cm) orbital radii (compare with Figure 3).

resolution than even a several-meter long magnetic-sector beam instrument.

3 EXCITATION AND DETECTION OF AN ION CYCLOTRON RESONANCE SIGNAL

3.1 Azimuthal Dipolar Single-frequency Excitation

Ion cyclotron motion is not by itself useful. Virtually all applications are therefore based on excitation produced by applying a spatially uniform electric field oscillating (or rotating) at or near the cyclotron frequency of ions of a particular m/z value. Excitation is used in three ways in FTICR/MS (Figure 5): (a) to acclerate ions coherently to a larger (and thus detectable) orbital radius; (b) to increase ion kinetic energy above the threshold for ion dissociation and/or ion–molecule reaction; and (c) to accelerate ions to a cyclotron radius larger than the radius of the ion trap (see below), so that ions are removed, i.e. ejected, from the instrument.

Next let us apply an azimuthal (i.e. in a plane perpendicular to **B**) spatially uniform electric field $\mathbf{E}(t)$, oscillating sinusoidally with time along the $\pm y$ direction:

$$\mathbf{E}(t) = E_0 \cos w_c t\mathbf{j} \tag{16}$$

In Equation (16) and below, **i** and **j** denote unit vectors along the x- and y-axes. For now, we shall suppose that

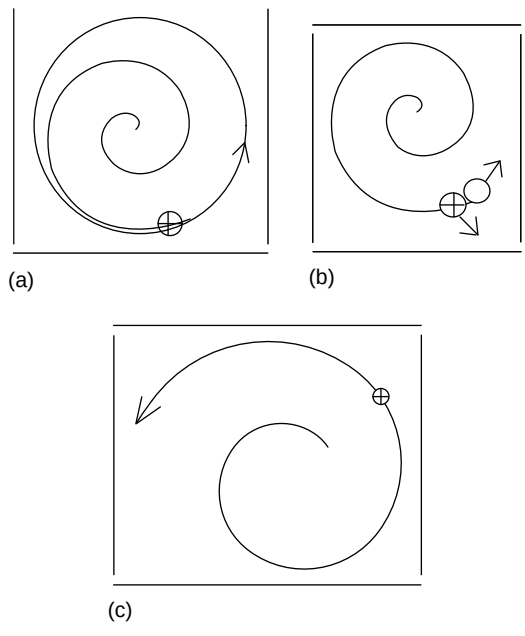

Figure 5 Uses for ion cyclotron excitation. (a) Acceleration of ions to form a spatially coherent packet at a detectable orbital radius. (b) Increase in ion kinetic energy to above the threshold for collision-activated dissociation (CAD) or reaction. (c) Ejection of ions of a given mass-to-charge ratio.

E_0 is generated by applying $+V_0$ and $-V_0$ volts to two opposed infinitely extended parallel conductive flat plates separated by a distance d (meters), so that Equation (17)

$$E_0 = \frac{2V_0}{d} = \frac{V_{\text{p-p}}}{d} \tag{17}$$

in which $V_{\text{p-p}}$ is the peak-to-peak voltage difference between the two plates (Figure 6).

This linearly-polarized electric field may be analyzed into two counter-rotating components, $\mathbf{E}_{\text{L}}(t)$ and $\mathbf{E}_{\text{R}}(t)$ (Equation 18),

$$\mathbf{E}(t) = \mathbf{E}_{\text{L}}(t) + \mathbf{E}_{\text{R}}(t) \tag{18}$$

in which the right-hand terms are given by Equations (19) and (20):

$$\mathbf{E}_{\text{R}}(t) = \frac{E_0}{2} \cos wt\mathbf{j} + \frac{E_0}{2} \sin wt\mathbf{i} \tag{19}$$

$$\mathbf{E}_{\text{L}}(t) = \frac{E_0}{2} \cos wt\mathbf{j} - \frac{E_0}{2} \sin wt\mathbf{i} \tag{20}$$

The RF electric field component $\mathbf{E}_{\text{R}}(t)$, rotating in the same sense (namely, counterclockwise in the xy-plane) and at the same frequency (i.e. in resonance with) the ion of interest pushes that ion continuously forward in its orbit, as shown in the top diagram of Figure 7. The electric field component $\mathbf{E}_{\text{L}}(t)$, rotating in the opposite sense as the ion, is off-resonance by $2w$ and has virtually no significant net effect after several cycles of the excitation frequency.

An ion absorbs power, $A(t)$, according to the dot product (Equation 21):

$$A(t) = \text{force} \times \text{velocity} = q\mathbf{E}(t) \times v_{xy} \tag{21}$$

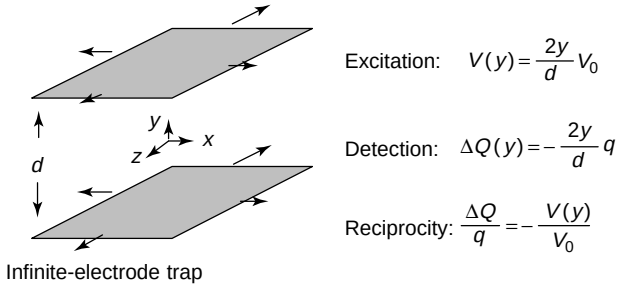

Figure 6 Excitation voltage difference and detection induced charge difference in FTICR/MS, demonstrated for two infinitely extended parallel flat electrodes located at $y = \pm d/2$ m away from the z-axis. If a potential $V_0 = V_{\text{p-p}}/2$ is applied to the upper electrode and $-V_0$ to the lower electrode, the potential, $V(y)$, anywhere between the electrodes is $2V_0 y/d$. If a point charge q is located between the same electrodes (in the absence of any applied voltage), then the difference, $\Delta Q(y)$, between the charge induced on the upper and lower electrodes is $-2qy/d$. It turns out that the relationship, $\Delta Q/q = -V/V_0$, is true (by the principle of reciprocity, also known as Earnshaw's theorem[6]) for opposed electrodes of arbitrary shape.

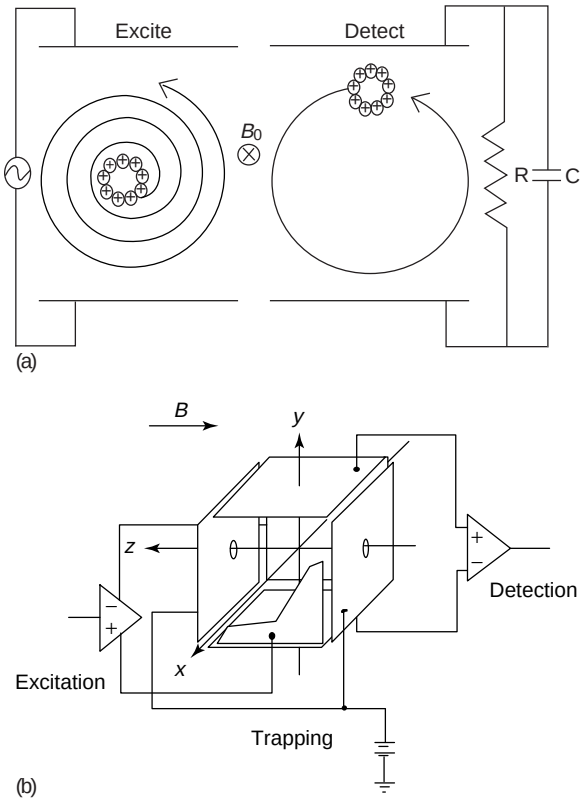

Figure 7 Incoherent ion cyclotron orbital motion ((a) left) is converted to coherent (and therefore detectable) motion ((a) right) by the application of a rotating electric field which rotates in the same sense and at the ICR frequency of the ions of a given m/z value. The electronic circuitry is shown in (b).

For a positive (or negative) ion (initially at rest) subjected to oscillating resonant excitation (Equations 16 and 17) for a period T_{excite}, it is straightforward to show from Equation (21) that the postexcitation ion cyclotron radius r is given by Equations (22) and (23):[7,8]

$$r = \frac{E_0 T_{\text{excite}}}{2B_0} \tag{22}$$

or

$$r = \frac{V_{\text{p-p}} T_{\text{excite}}}{2dB_0} \quad \text{(SI units)} \tag{23}$$

A particularly delightful feature of Equation (22) is that the postexcitation ion cyclotron orbital radius is independent of m/z! Thus, all ions of a given m/z range can be excited to the same ICR orbital radius, by application of an RF electric field whose magnitude is constant with frequency – i.e. no mass discrimination results from excitation, in the limit of perfectly spatially uniform RF electric excitation field. For example, an ion of arbitrary m/z in a magnetic field of 7.0 T can be excited to a radius of 0.72 cm in 1.0 ms by a constant RF resonant oscillating voltage of ± 1 V ($2V_{\text{p-p}}$) applied to infinitely extended parallel flat electrodes 2 cm apart. Thus, ions can be excited to a detectably useful ICR orbital radius (i.e. to a cyclotron radius large compared to the ion initial room-temperature thermal cyclotron radius) by a relatively small RF voltage.

From Equations (22) and (23), and remembering that $w_c = qB_0/m$, the ion kinetic energy immediately following azimuthal dipolar resonant excitation for a period T_{excite}, is given by Equations (24) and (25):

$$\text{kinetic energy}_{\text{postexcite}} = \frac{mw_c^2 r^2}{2} = \frac{q^2 E_0^2 (T_{\text{excite}})^2}{8m} = \frac{q^2 V_{\text{p-p}}^2 (T_{\text{excite}})^2}{8d^2 m} \tag{24}$$

$$\text{kinetic energy}_{\text{postexcite}} = \frac{1.206\,07 \times 10^7 z^2 V_{\text{p-p}}^2 (T_{\text{excite}})^2}{d^2 m} \tag{25}$$

in which the kinetic energy is in electron volts, $V_{\text{p-p}}$ is in volts, T_{excite} is in seconds, d is in meters, m is in unified atomic mass units and z is in multiples of elementary charge. For example, for an ion of 100 u, excited by ± 1 V applied for 1 ms to infinitely extended parallel flat electrodes 2 cm apart, the postexcitation ion kinetic energy is 1.2 keV. Thus, it is relatively easy to generate a high collision energy (see below) in an ion cyclotron. Equations (24) and (25) show that, for a given excitation electric field amplitude and duration, the postexcitation ion kinetic energy is independent of magnetic field strength B_0. However, the postexcitation cyclotron radius (Equation 22) varies inversely with B_0, so that for a given postexcitation ion cyclotron radius, the postexcitation ion energy increases as B_0^2.[7] The postexcitation ion kinetic energy as a function of ion mass, is shown in Figure 8, for either fixed postexcitation radius at several different magnetic field values, or fixed excitation amplitude but several different excitation periods.

3.2 Azimuthal Dipolar Single-frequency Detection

It is important to recognize that ICR orbital motion does not by itself generate an observable electrical signal (namely, a net difference between the charge induced in two opposed parallel electrodes). At its instant of formation in (or injection into) the ion trap, the phase of each ion's orbital motion is random – i.e. an ion may start its cyclotron motion at any point around either circle shown in Figure 1. Thus, for an ensemble of ions, any charge induced in either of two opposed detector plates will be balanced, on average, by an equal charge induced by an ion whose phase is 180° different (i.e. an ion located on the far side of the same orbit), so that the net difference in detected charge between the two plates is zero. Moreover, the cyclotron radius of thermal ions is too small to induce a detectable signal, even if all ions of a

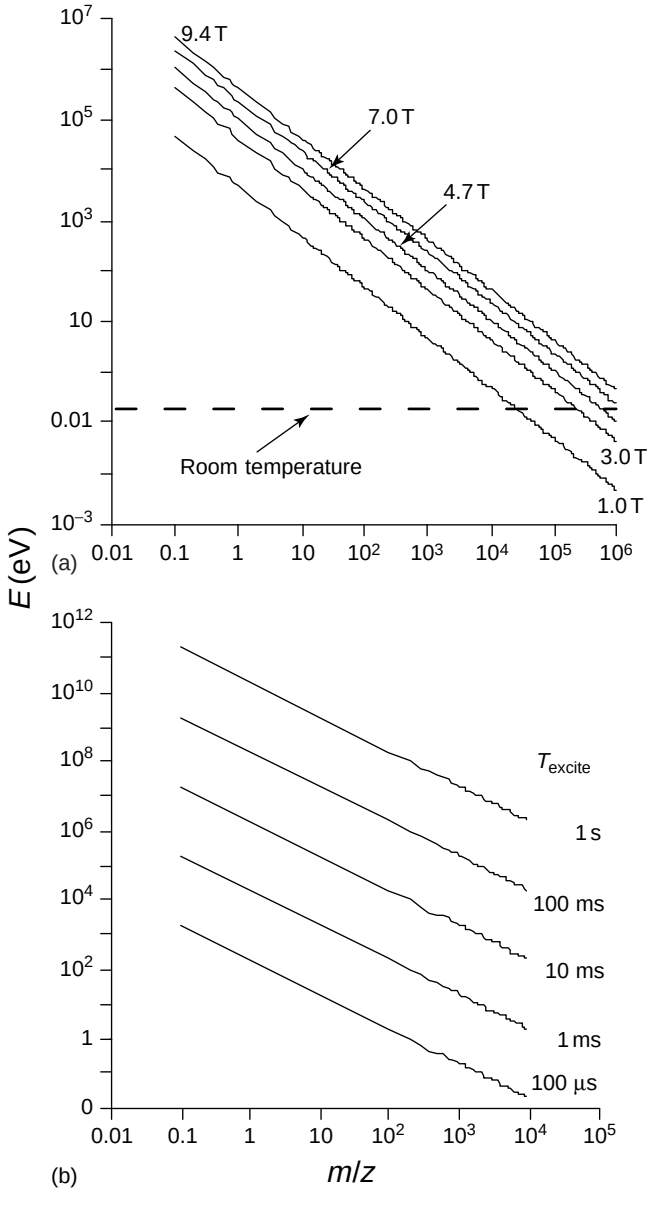

Figure 8 Postexcitation ion kinetic energy as a function of ion mass. (a) Fixed postexcitation cyclotron radius of 1 cm, for each of five magnetic field strengths: 1.0, 3.0, 4.7, 7.0, and 9.4 T. (b) Fixed excitation electric field = 1 $V_{p\text{-}p}$ applied between electrodes 2 cm apart, for each of five excitation durations.

given mass-to-charge ratio had the same cyclotron phase. Thus, resonant excitation, as just described above, creates spatial coherence (and thus a detectable signal) from an ion packet whose cyclotron orbits are (say) initially centered on the z-axis, by increasing the ion cyclotron radius moving the ion initial packet significantly off-axis (see Figure 5a).

The difference ΔQ in image charge on two opposed infinitely extended parallel flat conductive plates, induced by an ion of charge q between the plates, is given by Equation (26) (see Figure 6),

$$\Delta Q = -\frac{2qy}{d} \tag{26}$$

The ICR signal is proportional to the induced current (Equation 27),[9–13]

$$\frac{\mathrm{d}\Delta Q}{\mathrm{d}t} = -\frac{2q(\mathrm{d}y/\mathrm{d}t)}{d} \tag{27}$$

and is thus independent of magnetic field strength. Moreover, because the ion y-velocity component, $\mathrm{d}y/\mathrm{d}t$, increases linearly with ion cyclotron radius (Equation 7), the ICR signal increases linearly with ion cyclotron postexcitation radius. Also, the detected signal increases linearly with the ion charge, so that ICR is increasingly more sensitive for multiply charged ions (as from electrospray ionization (ESI)). Linearity is especially important for two reasons. First, because the ICR signal varies linearly with ion cyclotron postexcitation radius (which in turn depends linearly on the product of the excitation voltage amplitude and the duration), the ICR response at any frequency is proportional to the excitation spectral magnitude at that frequency. (This idea is used in the next section to show how frequency-sweep or off-resonant or stored waveform inverse Fourier transform (SWIFT) excitation works.) Second, a FT of the time-domain ICR response gives the same absorption spectrum that would otherwise have been obtained by measuring power absorption while sweeping infinitely slowly across the m/z range.[14] In addition, the superposition principle implies that the signals from any number of ions of arbitrary m/z values simply add at the detector; thus, ions over a wide m/z range can be detected simultaneously. The previous two statements combine to constitute the multichannel advantage of pulsed excitation followed by Fourier transformation to yield a spectrum of N data points in $1/N$ the time it would take to scan the spectrum one channel at a time.[14] We shall next examine broadband excitation and defer further consideration of the ICR signal until we have discussed the effect of ion-neutral collisions.

3.3 Broadband Excitation

So far, we have examined the response of an ion in a static spatially homogeneous magnetic field, $\mathbf{B} = -B_0\mathbf{k}$, to a single-frequency resonant (at w_c) oscillating electric field, $\mathbf{E}(t) = E_0 \cos w_c t\mathbf{j}$. However, the very act of turning the excitation on and off generates excitation at other frequencies, as shown in Figure 9. For example, the frequency-domain spectrum $E(\nu)$, of single-frequency excitation at frequency w_c, applied steadily for T_{excite}

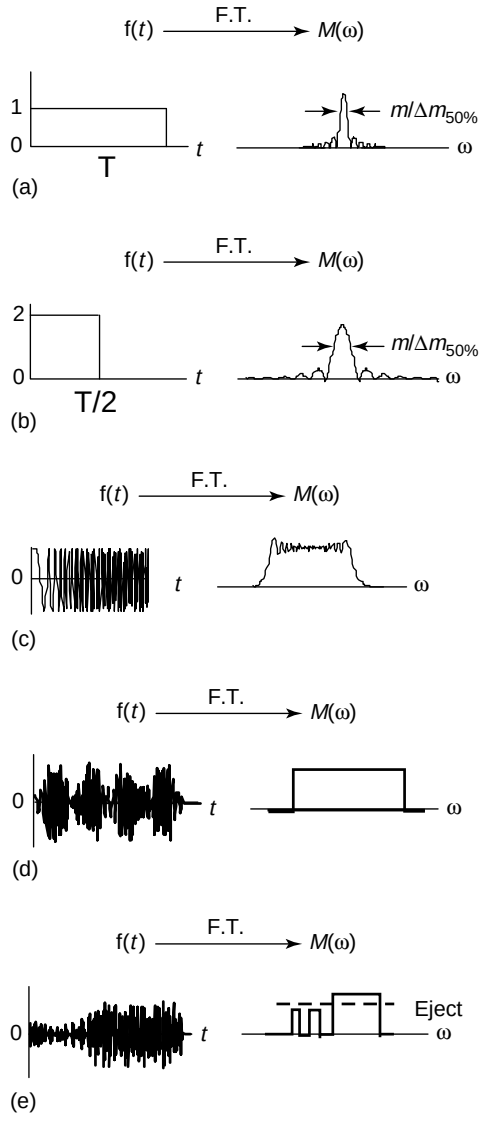

Figure 9 Time-domain (left) and frequency-domain (right) excitation waveforms: (a), (b) rectangular pulses; (c) frequency-sweep (chirp); (d), (e) SWIFT waveforms.

seconds, is given by Equation (28):[14]

$$E(\nu) = E_0 \frac{\sin(2\pi\nu_c T_{\text{excite}})}{2\pi\nu_c} \tag{28}$$

Thus, although the applied excitation may be precisely centered at a frequency w, which differs from w_c, turning the excitation voltage on at time zero and off at time T_{excite}, effectively broadens the spectral range of the excitation (to a bandwidth of $\sim 1/T_{\text{excite}}$ hertz). The longer the time-domain signal duration, the narrower is its corresponding frequency-domain spectral width. Thus, if an off-resonance excitation at frequency ν is sufficiently brief that $(1/T_{\text{excite}}) \geq 2\pi|\nu - \nu_c|$ it can still excite ions of cyclotron frequency w_c. This general result defines the ultimate precision with which ions can be mass-selected by their ICR frequencies. Fortunately, because ion cyclotron frequencies are so high (typically kilohertz to megahertz), excitation for as short as one second is still exquisitely mass selective (for instance about 10 ppm for ions of $m/z = 1000$ at 7.0 T).

Equation (28) shows that single-frequency excitation of duration T_{excite} can excite ions whose ion cyclotron frequencies span a range of $\sim 0.1/T_{\text{excite}}$ Hz to near-uniform cyclotron radii. (As the ICR signal magnitude is proportional to the postexcitation ICR orbital radius, which is in turn proportional to excitation magnitude, we seek a flat excitation magnitude spectrum in order to produce a mass spectrum whose peak areas accurately reflect the relative abundances of ions of different m/z.) However, Figure 9(a) and 9(b) shows that doubling the excitation frequency-domain bandwidth while maintaining the frequency-domain magnitude requires halving the time-domain excitation duration and doubling the time-domain amplitude. Thus, given that a single-frequency resonant excitation voltage of ± 1 V (between electrodes about 2 cm apart) is needed to excite ions to a detectable cyclotron radius of 1 cm in 1.4 ms (based on Equation 22), excitation of ions anywhere within a bandwidth ~ 1 MHz to near-uniform cyclotron radii will require single-frequency excitation of ~ 0.1 μs (i.e. about one-tenth of the width of the sinc function of Equation 28), corresponding to $\sim \pm 10\,000$ V (i.e. inconveniently large) RF excitation voltage. Therefore, broadband ion cyclotron excitation is commonly performed by frequency-sweep (i.e. chirp) excitation (Figure 9c),[1–4,14,15] which achieves excitation of relatively flat magnitude over a broad frequency range by use of relatively low excitation voltage (see section 7).

The disadvantages of chirp excitation are evident in Figure 9(c) – nonuniform excitation amplitude (translating to variation in apparent ion relative abundances) across the spectrum, and limited mass selectivity at the start and end frequencies of the sweep. Fortunately, a very general excitation mode is made possible by the highly linear excitation–response of ICR. The idea is simply to reverse the process, and begin by defining the desired excitation profile in the mass domain, converting it to a frequency-domain excitation spectrum, and then performing an inverse FT to generate the desired time-domain excitation waveform,[16,17] as shown in Figure 9(d) and 9(e). In this way, one can create mass spectral windows within which none (or all) of the ions are excited or ejected, thereby providing complete broadband mass selectivity for MS/MS or for enhanced dynamic range[18] by selective ejection of highly abundant ions. SWIFT achieves the flattest and most frequency-selective excitation magnitude spectrum theoretically possible for a given time-domain excitation period. The principles and

uses for SWIFT excitation (for both Penning (ICR) and Paul (quadrupole)) ion traps have been reviewed.[19]

3.4 Broadband Detection, Detection Limit

A coherently orbiting ion packet induces a differential current between two opposed detection plates and may be modeled as a current source. The receiver plates and wiring that connect the ion trap to the detection preamplifier have an inherent resistance and capacitance in parallel (see Figure 7).[9,10,12,13] Experimentally, at typical ICR frequencies (>10 kHz) the signal-to-noise ratio is independent of cyclotron frequency. However, at sufficiently low frequency (<10 kHz), the signal varies directly with frequency. Therefore, throughout most of the frequency range excited by a broadband chirp or SWIFT waveform in a standard FTICR/MS experiment, the detected signal-to-noise ratio reflects the relative current differential induced on the detection plates. Furthermore, the detection limit (namely, the minimum number N of ions that may be detected from an undamped signal in a single 1 s acquisition period to yield a signal-to-noise ratio of 3 : 1) may be calculated from Equation (29),

$$N = \frac{CV_{\mathrm{d(p\text{-}p)}}}{qA_1(r)} \tag{29}$$

in which C is the capacitance of the detection circuit, $V_{\mathrm{d(p\text{-}p)}}$ is the peak-to-peak amplitude of the detected voltage (calibrated for a given spectrometer), and $A_1(r)$ is a coefficient that is approximately proportional to r and may be determined graphically.[20] For example, for typical operating parameters, namely a detection circuit capacitance of 50 pF, $V_{\mathrm{d(p\text{-}p)}}$ of 3×10^{-7} V, and $A_1(r) = 0.5$ (i.e. the ion is excited to approximately half of its maximal cyclotron radius), an observed signal-to-noise ratio of 3 : 1 corresponds to a detection limit of about 187 ions.

4 ION-NEUTRAL COLLISIONS

For ions of a single m/z value, the ICR frequency-domain response in the absence of collisions looks just like Equation (28), and is obtained by Fourier transformation of a digitized time-domain signal to give a sinc function spectrum that looks just like Equation (14), except that T_{excite} is replaced by the data acquisition period T_{acq}. The simplest way to treat ion-neutral collisions is to model their effect as a frictional damping force in the overall ion Equation (30) of motion:

$$\text{force} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} - f\mathbf{v} \tag{30}$$

This model[21,22] leads to an exponential damping (Equation 31) of the time-domain ICR signal for N_{i} ions of ICR frequency ν_{i},

$$s_{\mathrm{i}}(t) \propto N_{\mathrm{i}} \exp\left(\frac{-t}{\tau_{\mathrm{i}}}\right) \cos(2\pi\nu_{\mathrm{i}}t) \tag{31}$$

whose FT yields a Lorentzian frequency-domain magnitude-mode spectrum (Equation 32):

$$M(w) \propto \frac{\tau}{\sqrt{1 + (w - 2\pi\nu_{\mathrm{i}})^2\tau^2}} \tag{32}$$

Thus (Figure 10), the FTICR spectral line shape approaches the sinc shape in the limit of zero pressure (i.e. $\tau \gg T_{\mathrm{acq}}$, so that there are essentially no ion-neutral collisions during data acquisition), and approaches the Lorentzian shape in the high-pressure limit ($\tau \ll T_{\mathrm{acq}}$) that the signal has damped nearly to zero during the data acquisition period. These line shapes define mass resolution and mass resolving power, as discussed below. The factor τ may be defined as in Equation (33),[22]

$$\frac{1}{\tau} = \frac{m_{\mathrm{neutral}}}{m_{\mathrm{ion}} + m_{\mathrm{neutral}}}\nu_{\mathrm{collision}} \tag{33}$$

in which $\nu_{\mathrm{collision}}$ is the number of ion–molecule collisions per second. In order to calculate $\nu_{\mathrm{collision}}$, one must first determine the rate constant for ion–molecule collisions.

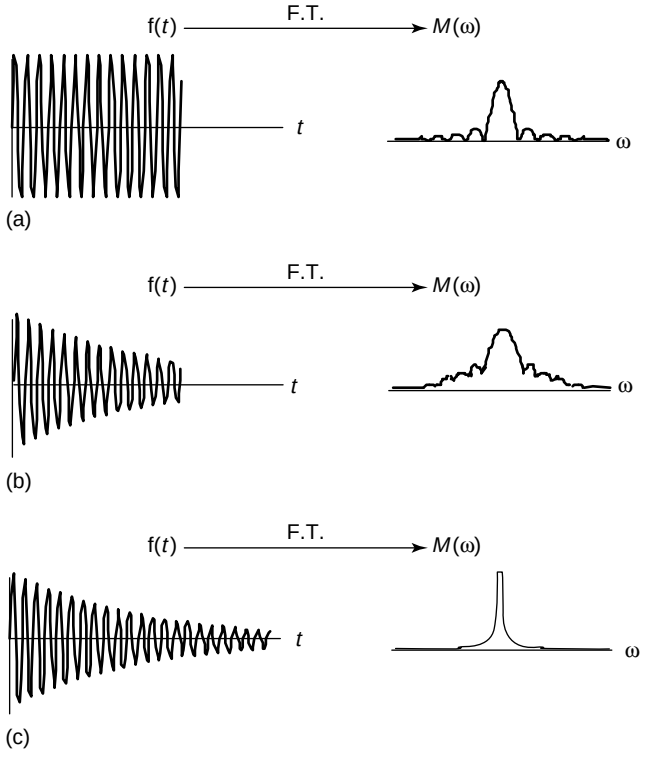

Figure 10 Simulated time-domain ICR signals (left) and frequency-domain magnitude spectra (right) for: (a) low-pressure, $\tau \gg T_{\mathrm{acq}}$; (b) intermediate-pressure, $\tau \approx T_{\mathrm{acq}}$; and (c) high-pressure, $\tau \ll T_{\mathrm{acq}}$ limits.

At thermal velocity, an ion and a neutral molecule interact through the ion-induced dipole (Langevin) potential (Equations 34 and 35),[23]

$$U(r) = -\frac{\alpha' q^2}{8\pi\varepsilon_0 r^4} \quad \text{(SI units)} \tag{34}$$

$$U(r) = -\frac{\alpha' q^2}{2r^4} \quad \text{(cgs units)} \tag{35}$$

in which q is the ion charge, α' is the isotropic polarizability of the neutral, r is the distance between the two bodies, and the ion is modeled as a point charge. From this interaction potential and the Langevin model the ion–neutral collision rate constant may be calculated (Equations 36 and 37):

$$k = \sqrt{\frac{\pi\alpha' q^2}{\varepsilon_0 \mu}} \quad \text{(SI units)} \tag{36}$$

$$k = \sqrt{\frac{4\pi^2\alpha' q^2}{\mu}} \quad \text{(cgs units)} \tag{37}$$

At this juncture, a note about units is in order. The parameter k is almost always reported in cgs units. In cgs units, the elementary charge q is $4.806\,53 \times 10^{-10}$ statcoulombs (electrostatic units or esu), α' is in cubic centimeters, and μ (the reduced mass of the ion and molecule) is in grams. Thus, in order to obtain the collision frequency for one ion, the pressure must be computed in number of molecules per cubic centimeter. For example, suppose that the pressure inside the Penning trap is 1×10^{-8} torr, corresponding ($PV = nRT$) to a neutral concentration of 3.24×10^8 molecules per cubic centimeter at a temperature of 298.15K. For an ion of 100 u colliding with nitrogen (N_2) molecules, the neutral polarizability, $\alpha' = 1.7403 \times 10^{-24}\,\text{cm}^3$.[24] From these values, Equation (37) yields an ion–neutral collision rate constant, $k = 6.61 \times 10^{-10}\,\text{cm}^3\,\text{s}^{-1}$. If we now multiply this number by the neutral concentration, we obtain $\nu_{\text{collision}} \approx 0.2$ collisions per second per ion.

Recently, it has been pointed out that the collisional model just described derives from a Langevin (ion–induced dipole) collision model.[25] Although the Langevin model offers a good description for room-temperature ions, a hard-sphere collision model is much more appropriate for the much higher velocities of ions during FTICR data acquisition. Equations (29) and (30) then take the form of Equation (38):[26]

$$\text{force} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} - f\mathbf{v}^2 \tag{38}$$

Instead of exponential damping of the ion velocity with time, the ion velocity, $v(t)$, is now represented by Equation (39),

$$v(t) = \frac{v_0}{1 + v_0[m_{\text{neutral}}/(m_{\text{neutral}} + m_{\text{ion}})]N\sigma_{\text{hard sphere}}t} \tag{39}$$

in which N is the neutral number density, $\sigma_{\text{hard sphere}}$ is the hard-sphere collision cross-section, and m_{neutral} and m_{ion} are the neutral and ion masses. Although there is no analytical FT of Equation (33), it can be transformed numerically to yield a frequency-domain line shape that has a narrow width at half-maximum peak height, but is very broad at the base. At present, FTICR time-domain signals are typically weighted (windowed, apodized) so as to suppress the broad base of the true hard-sphere line shape; however, the natural line shape may prove useful for future quantitation of ion–neutral collision frequencies (e.g. for comparison to ion mobilities measured in beam instruments).

5 EFFECTS OF AXIAL CONFINEMENT OF IONS IN A TRAP OF FINITE SIZE

Up to now, we have considered only idealized ion behavior in a spatially uniform **B** field and an orthogonal spatially uniform oscillating **E** field. Here we show how those results are modified by the need to perform the experiment in a cell of finite dimensions. First, a static magnetic field applied along the z direction effectively confines ions in the x and y directions according to the cyclotron motion described above. However, ions are still free to escape along the z axis (i.e. along or opposed to the magnetic field direction). In order to prevent such escape, it is usual to apply a small (~1 V) electrostatic potential (ideally three-dimensional axial quadrupolar, see below) to each of two end cap electrodes positioned at $z = \pm a/2$ from the center of the cell. Second, because the electrodes that generate the electrostatic trapping potential and the excitation (or detection) electric potentials are necessarily finite in size, the resultant electric field is nonlinear (and reduced in magnitude) from that from infinitely extended opposed parallel flat conductive plates. We next show briefly how these two perturbations affect the FTICR experiment.

5.1 Axial Ion Oscillation Due to the z Component of the Electrostatic Trapping Potential

The optimal one-dimensional potential for axial confinement of ions would be a simple quadratic potential, $\Phi(z) \propto (z^2/2)$, because the gradient of that potential yields a simple harmonic restoring force, so that ions oscillate sinusoidally back and forth between the end cap electrodes, at a frequency independent of ion z position. However, in three dimensions, Laplace's equation,

$\nabla^2\Phi(x, y, z) = 0$, requires the addition of terms in x^2 and y^2. Hans Dehmelt and Wolfgang Paul, in work that laid the basis for their shared Nobel prize in 1989, exploited the experimentally convenient (i.e. cylindrically symmetric) three-dimensional axial quadrupolar electrostatic trapping potential of Equation (40),

$$\Phi(x, y, z) = V_{\text{trap}} \quad \gamma + \frac{\alpha}{2a^2}(2z^2 - x^2 - y^2) \qquad (40)$$

or Equation (41)

$$\Phi(r, z) = V_{\text{trap}} \quad \gamma + \frac{\alpha}{2a^2}(2z^2 - r^2) \qquad (41)$$

in which $r = \sqrt{x^2 + y^2}$ is the radial position of the ion in the x–y plane, a is a measure of trap size, and γ and α are constants that depend on the trap shape.[27] The shape of the potential of Equations (40) and (41) is graphically evident from a plot of its isopotential surfaces (Figure 11).

Values of γ and α for several ICR trap shapes (Figure 12) in common use are shown in Table 1. Methodology for computing γ and α for orthorhombic, tetragonal, and cylindrical traps of arbitrary aspect ratio is available.[44] The dipolar excitation fields for cubic and cylindrical (unit aspect ratio) traps are essentially identical, if the excitation voltage is about 1.4 times higher for the cylindrical trap; similarly, the detected signal from the cylindrical trap is about 0.7 of that for a cubic trap whose corners just touch the inner walls of the comparable cylindrical trap. (We have chosen a cubic trap inscribed (rather than circumscribed) relative to a cylinder, because both traps then each have the largest possible equal diameter, and thus are equally separated from the inner (cylindrical) wall of the enveloping vacuum chamber.) The hyperbolic trap is near-perfect for trapping potential, but is very nonlinear for dipolar excitation. Although capacitive coupling of the three-cylinder open trap improves its dipolar excitation linearity only slightly for ions in the $z = 0$ midplane, the capacitively coupled design provides near-linear excitation from one end of the trap to the other, and therefore virtually eliminates z ejection. Also, although β_{quad} for the matrix-shimmed trap is not as large as for the cylindrical traps, it closely approximates the ideal β_{quad} of 8/3 (i.e. the limiting value for an infinitely extended tetragonal trap). The infinity trap improves dipolar excitation, but not (in the absence of an additional switching network) dipolar detection, and requires that quadrupolar excitation be performed with two-plate rather than four-plate excitation.[45]

From Equations (40) and (41) it is straightforward to solve the Equation (42) of ion z motion,

$$\text{axial force} = m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -q\nabla\Phi(x, y, z) \qquad (42)$$

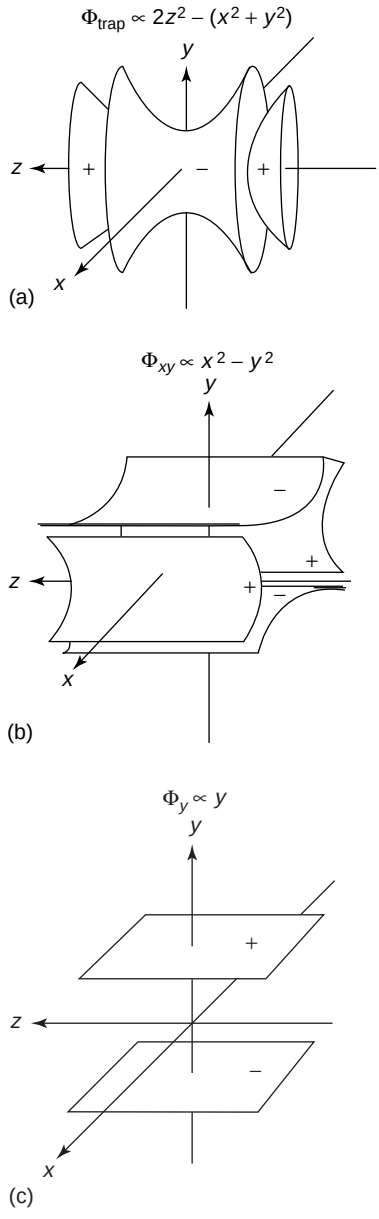

Figure 11 Plots of isopotential surfaces: (a) trapping, (b) two-dimensional quadrupolar, and (c) dipolar potentials.

to obtain an ion z position that oscillates sinusoidally with time (Equations 43–45)

$$z(t) = z(0)\cos(2\pi\nu_z t) \qquad (43)$$

in which

$$\nu_z = \frac{1}{2\pi}\sqrt{\frac{2qV_{\text{trap}}\alpha}{ma^2}} \quad \text{(SI units)} \qquad (44)$$

or

$$\nu_z = 2.210\,88 \times 10^3\sqrt{\frac{zV_{\text{trap}}\alpha}{ma^2}} \qquad (45)$$

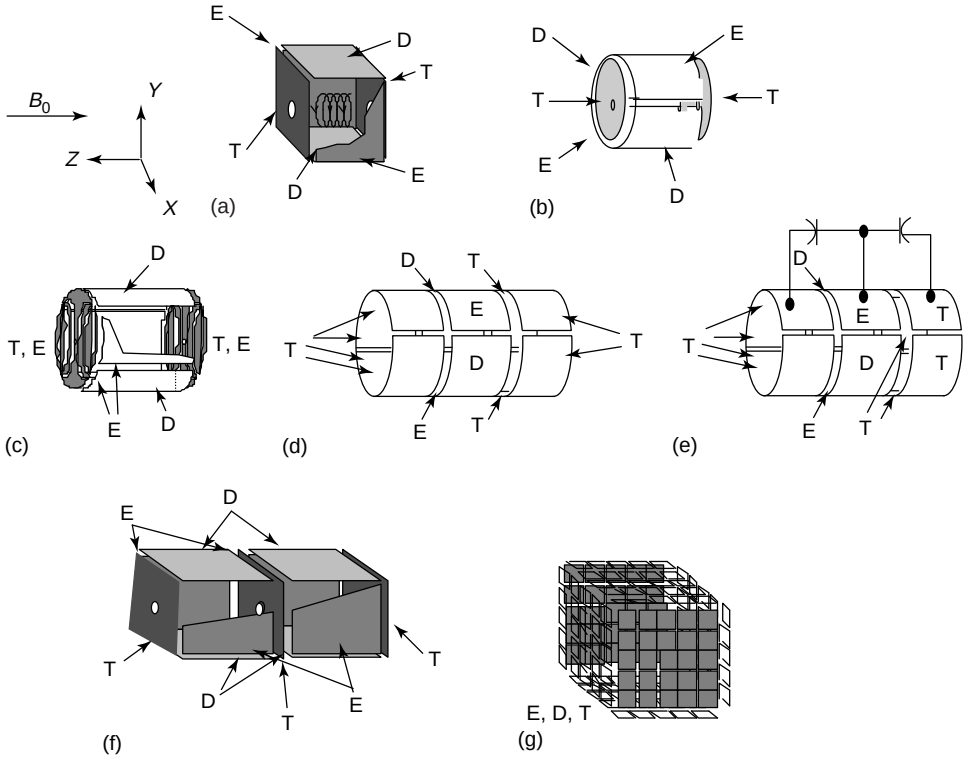

Figure 12 ICR ion trap configurations. (E = excitation, D = detection, T = end cap (trapping): (a) cubic;[28,29] (b) cylindrical;[1–4,30–33] (c) end caps segmented to linearize excitation potential (infinity trap);[34] (d) and (e) open-ended, without or with capacitive RF coupling between the three sections;[35–37] (f) dual;[38] and (g) matrix-shimmed.[11]

Table 1 Trapping (γ, α in Equations 40 and 41), azimuthal dipolar excitation (β_{dipolar}, Equations 56 and 57) and two-dimensional azimuthal quadrupolar excitation (β_{quad}, Equations 61 and 62) scale factors for traps shapes in common use in FTICR/MS

Trap shape	End cap separation	Excitation electrode separation	γ	α	β_{dipolar}	β_{quad}
Ideal	NA	NA	0.500 00	4.000 00	1.000 00	2.666 67
Cube[39]	a	$d = a$	0.333 33	2.773 73	0.721 67	2.773 73
Cylinder[31,32,40]	a	d	0.278 7	2.840 4	0.808 18	3.255 22
Infinity	a	d	0.278 7	2.840 4	≈0.900[a]	NA[b]
Open Uncoupled[c]	a	d	0.145 27	3.867 98	0.867 38	3.305 27
Open Coupled[c]	a	d	0.145 27	3.867 98	0.896 99	3.348 00[d]
Hyperbolic[42,43]	a	$d = \sqrt{2}a$	0.444 03	4.090 5	0.664 83	2.6469
Matrix-shimmed cube	a	d	0.500 00	3.925 4	≈1[e]	2.6264

[a] Approaches the value of β_{dipolar} for an infinitely long cylindrical trap.
[b] Cannot perform traditional four-plate quadrupolar excitation without additional circuitry and switching; however, two-plate quadrupolar excitation may be used instead.[41]
[c] Three consecutive right circular cylinders, each of unit aspect ratio, capacitively coupled or not. In this case, a refers to the length of the central cylinder.
[d] Requires capacitive coupling of the detection electrodes as well as the excitation electrodes.
[e] Up to a radius of $a/4$ (i.e. halfway from the trap central axis to the side of the trap).

where ν_z is in hertz, V_{trap} in volts, a in meters, m in unified atomic mass units, and z in multiples of elementary charge. For example, for $V_{\text{trap}} = 1$ V in a cubic trap of side length $a = 2.54$ cm, an ion of $m/z = 1000$ will oscillate at a trapping frequency of 4580 Hz.

5.2 Radial Ion Magnetron Rotation Due to Combination of the Magnetic Field and the Radial Component of the Electrostatic Trapping Potential

The DC trapping potential of Equations (40) and (41) also produces a radial force (Equation 46):

$$\text{radial force} = qE(r) = \frac{qV_{\text{trap}}\alpha}{a^2}r \tag{46}$$

The radial electric field acting on the ion produces an outward-directed electric force which opposes (and thus has opposite sign with respect to) the inward-directed Lorentz magnetic force from the applied magnetic field. We can now combine Equation (46) with Equation (5) to obtain the equation for ion motion subject to a static $\mathbf{B} = B_0\mathbf{k}$ field and a three-dimensional axial quadrupolar electrostatic potential of Equation (41), giving Equation (47):

$$\text{force} = mw^2r = qB_0wr - \frac{qV_{\text{trap}}\alpha}{a^2}r \tag{47}$$

or

$$w^2 - \frac{qB_0w}{m} + \frac{qV_{\text{trap}}\alpha}{ma^2} = 0 \tag{48}$$

Note that Equation (48) is quadratic in w, but that w is independent of r! That is the great advantage of the three-dimensional axial quadrupolar DC trapping potential – namely, each of the resultant ion motional frequencies (see below) is independent of ion position inside the trap. Solving Equation (48) for w, gives two natural rotational frequencies (in radians per second) in place of the original unperturbed cyclotron frequency that is observed in the absence of DC trapping potential (Equations 49–52):

$$w_+ = \frac{w_c}{2} + \sqrt{\left(\frac{w_c}{2}\right)^2 - \frac{w_z^2}{2}} \quad \text{(reduced cyclotron frequency)} \tag{49}$$

$$w_- = \frac{w_c}{2} - \sqrt{\left(\frac{w_c}{2}\right)^2 - \frac{w_z^2}{2}} \quad \text{(magnetron frequency)} \tag{50}$$

in which

$$w_z = \sqrt{\frac{2qV_{\text{trap}}\alpha}{ma^2}} \quad \text{(trapping oscillation frequency, in SI units)} \tag{51}$$

and

$$w_c = \frac{qB_0}{m} \quad \text{(unperturbed cyclotron frequency, in SI units)} \tag{52}$$

The three natural ion motional modes (cyclotron rotation, magnetron rotation, and trapping oscillation) are shown in Figure 13, and their relative frequencies as a function of ion m/z and DC trapping potential are shown in Figure 14.[46] The magnetron and trapping frequencies are usually much less than the cyclotron frequency, and generally are not detected (except as small sidebands when the ion trap is misaligned with the magnet axis and/or ion motional amplitudes approach the dimensions of the trap).[5,47,48]

5.3 Mass Calibration

Equation (49) shows that the imposition of a quadrupolar DC trapping potential reduces the ion cyclotron orbital frequency, because the radially outward-directed magnetic field effectively reduces the magnetic field strength. In the absence of electric space charge and trapping potentials, measurement of the frequency of a single ion of known mass would serve to calibrate the magnetic field strength, from which the m/z values of all other ions in the mass spectrum could be computed from Equation (5). However, the introduction of trapping potential changes the ICR mass–frequency relationship to the form shown in Equation (49), from which Equation (53) can be derived.[49] The factors A and B are constants obtained by fitting a particular set of ICR mass

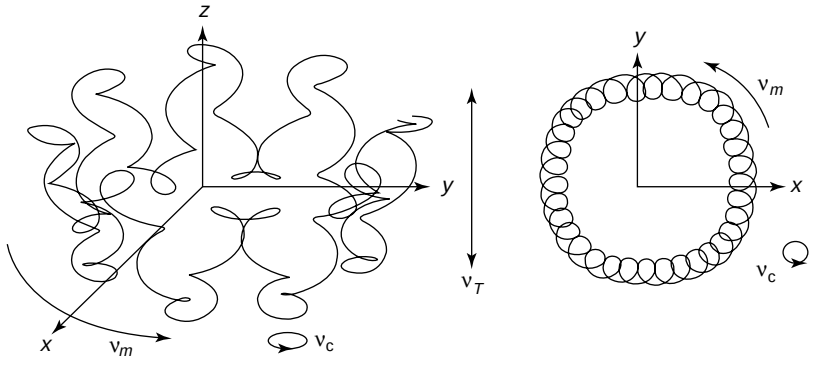

Figure 13 Ion motion in a 2 in cubic Penning trap in a perfectly homogeneous magnetic field of 3 T for an ion of m/z 2300, for 10 V trapping voltage. The three natural motional frequencies and amplitudes have relative magnitudes given by $w_+ = 4.25w_z$, $w_z = 8.5w_-$, $\rho_- = 4_{z_{\max}} = 8\rho_+$. Note that the field produced by a cubic trap is not perfectly quadrupolar as manifested by the shape of the magnetron orbit (not a perfect circle) in the xy plane. The magnetic field points in the negative z direction.

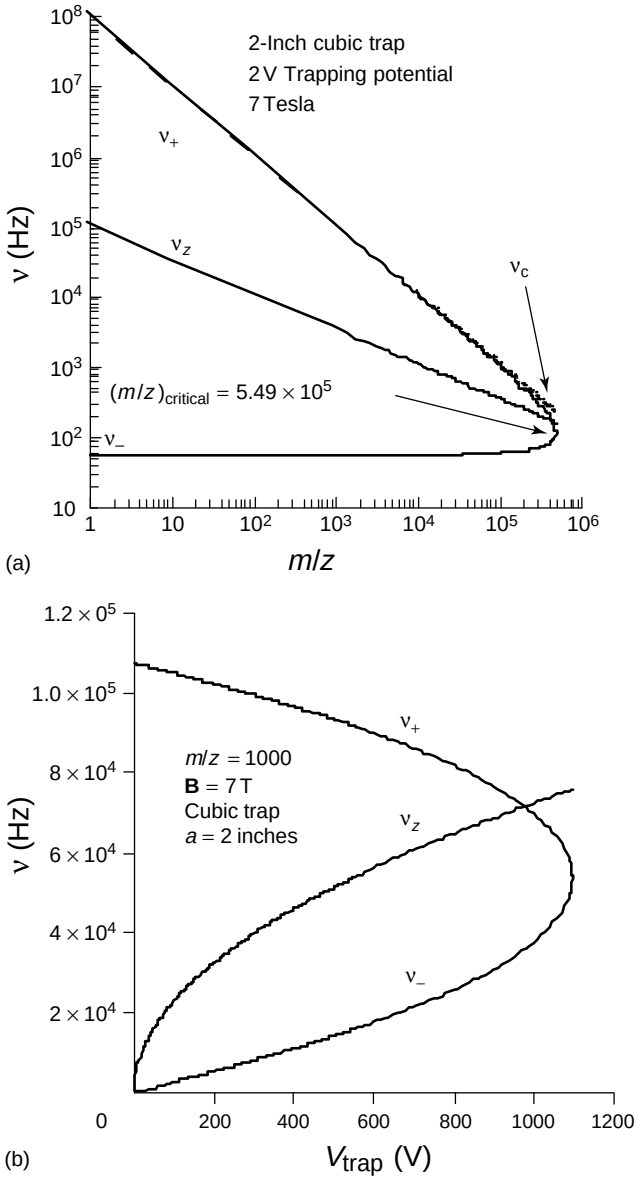

Figure 14 (a) Mass-to-charge ratio (m/z) dependence of the frequency of each of the three ion natural motions (cyclotron rotation, magnetron rotation, and axial oscillation) in an ICR (Penning) ion trap (see Equations 49–51). At m/z values well below the critical m/z, the reduced cyclotron frequency ν_+ is approximately inversely proportional to m/z, whereas the axial frequency ν_z is approximately inversely proportional to the square root of m/z, and magnetron frequency is nearly independent of m/z. At the critical m/z value, $\nu_+ = \nu_- = \nu_z/\sqrt{2} = \nu_c/2$ (see Equation 68). At or above the critical m/z value, ions are no longer trapped. (Reproduced with permission from Guan and Marshall.[7]) (b) Natural motional frequencies as a function of DC trapping potential. Note the maximum trapping potential above which ions are not confined radially.

spectral peak frequencies for ions of at least two known m/z values to Equation (53):

$$\frac{m}{z} = \frac{A}{\nu_+} + \frac{B}{\nu_+^2} \tag{53}$$

Strictly speaking, Equations (49–53) should be valid only in the single-ion limit (i.e. no Coulomb interactions between ions). In practice, however, FTICR/MS experiments are usually performed with sufficiently few ions that the space charge perturbation is small; moreover, the perturbation affects both calibrant and analyte ions. This is basically why ICR mass calibration works so well, even when many ions are present. Internal calibration (i.e. calibrant ions are present along with analyte ions in the same sample) typically improves mass accuracy by a factor of at least three compared to external calibration (i.e. calibration is performed on a separate sample from the analyte). External calibration works best when calibrant ions are excited to the same cyclotron radius as analyte ions and when the number of ions in the trap is the same for both experiments.

6 QUADRUPOLAR EXCITATION, AXIALIZATION, AND ION REMEASUREMENT

So far, it has been shown that a three-dimensional axial quadrupolar electrostatic potential (see Figure 11a) is optimal for trapping ions for ICR, and that a one-dimensional linear RF electric potential (see Figure 11c), which produces a spatially uniform RF electric field, is optimal for ICR excitation/detection. We now discuss the purpose of a two-dimensional azimuthal quadrupolar RF potential (see Figure 11b). It can be shown[50] that such a potential (oscillating at the unperturbed ion cyclotron frequency $w_c = qB_0/m$) periodically interconverts magnetron and cyclotron motions. As a matter of fact, it has been shown that such an experiment is formally analogous to resonant excitation in a two-level spin one-half magnetic resonance experiment: conversion from pure magnetron to pure cyclotron motion corresponds to population inversion by a π-pulse.[51]

In the presence of collisions, the ion magnetron radius increases slowly with time, whereas the ion cyclotron radius decreases rapidly with time. Thus, if the magnetron radius (i.e. the radial position of the center of an ion cyclotron orbit) is converted to the cyclotron radius (by two-dimensional azimuthal quadrupolar excitation at frequency w_c), then collisions will rapidly damp the cyclotron motion to zero, and ions will relax toward the central axis of the ion trap. The interconversion frequency is given by[52]

$$w_{\text{interconvert}} = \frac{3(2.666\,67)qV_{\text{quad}}}{2md^2(w_+ - w_-)} \quad \text{(SI units)} \tag{54}$$

or

$$\nu_{\text{interconvert}} = \frac{9.77601 \times 10^6 z V_{\text{quad}}}{md^2(\nu_+ - \nu_-)} \tag{55}$$

In Equations (54) and (55) the frequencies are in hertz, V_{quad} in volts, d in meters, m in unified atomic mass units, and z in multiples of elementary charge. The term $\pm V_{\text{quad}}$ is the voltage applied to electrodes (separated by d meters), and 2.666 67 is an ideal scaling factor corresponding to an infinitely extended tetragonal trap (β_{quad}, see next section), in the configuration shown in Figure 11(b). For example, for ions of $m/z = 1000$, application of two-dimensional azimuthal quadrupolar excitation of ± 1 V amplitude to electrodes separated by 2.54 cm in the configuration of Figure 11(b), at the unperturbed ion cyclotron frequency (107 500 Hz at 7.0 T), will interconvert magnetron and cyclotron motion at a frequency of ~141 Hz. Thus, if those ions have zero initial cyclotron radius at time zero, then their magnetron motion will be completely converted to cyclotron motion in $0.5\,\nu_{\text{interconvert}}$ or about 3.5 ms.

The conversion of magnetron motion to cyclotron motion, followed by collisional damping of the cyclotron radius has come to be known as quadrupolar axialization.[52] By shrink-wrapping a packet of ions initially distributed widely in radius (for instance as a result of ion initial formation or injection into the ion trap), axialization improves virtually every aspect of FTICR performance: mass resolving power and mass accuracy; cooling of ion internal energy; mass selectivity for MS^n; CID efficiency; transfer efficiency in dual-trap experiments; and ion remeasurement efficiency. Ion remeasurement[53–56] is especially useful for improving signal-to-noise ratio (and reducing the ICR detection limit)[57] and for high-resolution MS^n.[58] Finally, quadrupolar axialization combined with nondestructive ICR detection enhances user-interactive FTICR/MS.[59]

7 EFFECT OF TRAP SIZE AND SHAPE ON DIPOLAR AND TWO-DIMENSIONAL QUADRUPOLAR EXCITATION

Equations (22–25) give the ion cyclotron radius r, and translational (kinetic) energy, following single-frequency resonant azimuthal dipolar excitation by a spatially uniform RF electric field. To a good approximation, those equations also apply for similar excitation in a finite-size ion trap, by inclusion of a scaling factor β. (Equations 56–60):

$$r = \frac{\beta_{\text{dipolar}} E_0 T_{\text{excite}}}{2B_0} \tag{56}$$

or

$$r = \frac{\beta_{\text{dipolar}} V_{\text{p-p}} T_{\text{excite}}}{2dB_0} \quad \text{(SI units)} \tag{57}$$

or, for broadband frequency-sweep dipolar excitation,

$$r = \frac{\beta_{\text{dipolar}} V_{\text{p-p}} \sqrt{1/\text{sweep rate}}}{2dB_0} \tag{58}$$

and

$$\begin{aligned} \text{kinetic energy}_{\text{postexcite}} &= \frac{\beta^2_{\text{dipolar}} q^2 E_0^2 (T_{\text{excite}})^2}{8m} \\ &= \frac{\beta^2_{\text{dipolar}} q^2 V^2_{\text{p-p}} (T_{\text{excite}})^2}{8d^2 m} \quad \text{(SI)} \end{aligned} \tag{59}$$

or

$$\begin{aligned} \text{kinetic energy}_{\text{postexcite}} &= \frac{1.206\,07 \times 10^7 \beta^2_{\text{dipolar}} z^2 \times V^2_{\text{p-p}} (T_{\text{excite}})^2}{d^2 m} \\ &= \frac{1.206\,07 \times 10^7 \beta^2_{\text{dipolar}} z^2 V^2_{\text{p-p}} \times (1/\text{sweep rate})}{d^2 m} \end{aligned} \tag{60}$$

in which energy is in electron volts, $V_{\text{p-p}}$ is in volts, T_{excite} is in seconds, d is in meters, m is in unified atomic mass units, and z is in multiples of elementary charge. For example, for single-frequency resonant excitation at $10\,V_{\text{p-p}}$ for 400 μs in a cubic Penning trap whose excitation plates are 2 cm apart, in a 7 T magnet field, the postexcitation cyclotron radius is 1.03 cm. Alternatively, ions having a wide range of mass-to-charge ratios may be excited to the same radius by frequency-sweep excitation (spanning the ICR frequencies of interest) of amplitude, $V_{\text{p-p}} = 123$ V, at a sweep rate of 1.00×10^9 Hz s^{-1}. It is worth noting that the magnitude mode spectral peak height is proportional to $1/\sqrt{\text{sweep rate}}$.

Values of β_{dipolar} for several ICR traps in common use are shown in Table 1. Methodology for computing β_{dipolar} for orthorhombic, tetragonal, and cylindrical traps of arbitrary aspect ratio is available.[44]

Similarly, Equations (54) and (55) for the interconversion frequency for azimuthal two-dimensional quadrupolar excitation are rescaled as Equations (61) and (62). Values of β_{quad} for several ICR traps in common use are shown in Table 1, and formulas for computing β_{quad} for orthorhombic, tetragonal, and cylindrical traps of arbitrary aspect ratio have been published.[44]

$$w_{\text{interconvert}} = \frac{3q\beta_{\text{quad}} V_{\text{quad}}}{2md^2(w_+ - w_-)} \quad \text{(SI units)} \tag{61}$$

or

$$\nu_{\text{interconvert}} = \frac{2.303\,42 \times 10^7 z \beta_{\text{quad}} V_{\text{quad}}}{md^2(\nu_+ - \nu_-)} \tag{62}$$

where frequencies are in hertz, V_{quad} in volts, d in meters, m in unified atomic mass units, and z in multiples of elementary charge.

Ideal scaling factors for all three potentials (dipolar excitation, quadrupolar excitation, and quadrupolar trapping) are closely approached by the potentials generated by the matrix-shimmed trap (see Figure 12). For example, $\beta_{dipolar} \approx 1$ for the matrix-shimmed trap. The matrix-shimmed trap consists of a cubic trap in which each side is segmented into a 5×5 grid for a total of 150 electrodes. A near-perfect potential is then created by applying the appropriate voltage to each of these electrodes. The matrix-shimmed trap is an example of inverse logic and is discussed in the ion trap review paper by Guan and Marshall.[11] Although the matrix-shimmed trap is near-perfect in trapping, dipolar excitation, and quadrupolar excitation potential shape, its actual construction necessarily adds so much capacitance that its excitation and detection efficiency are unacceptably low.

8 MASS RESOLVING POWER, MASS RESOLUTION, AND MASS ACCURACY

The first derivative of Equation (5) with respect to m yields Equation (63),

$$\frac{dw_c}{dm} = \frac{-qB_0}{m^2} = -\frac{w_c}{m} \tag{63}$$

and hence Equation (64):

$$\frac{w_c}{dw_c} = -\frac{m}{dm} \tag{64}$$

It is usual in all types of FT spectroscopy to define resolution as the full width of a spectral peak at half-maximum peak height; namely, $\Delta w_{50\%}$ or $\Delta m_{50\%}$ for frequency-domain or mass-domain FTICR spectra, because a valley just begins to appear between two peaks of equal height and shape when they are separated by slightly more than $\Delta w_{50\%}$ or $\Delta m_{50\%}$. The resolving power is then defined as $w/\Delta w_{50\%}$ or $m/\Delta m_{50\%}$, respectively. Equation (64) shows that the frequency resolving power and mass resolving power in ICR/MS are the same (except for a minus sign).

Because the frequency of an FTICR/MS peak is approximately qB_0/m, the experimental mass resolution in Equation (64) can be expressed as Equation (65):

$$\frac{m}{\Delta m_{50\%}} = -\frac{qB_0}{m\Delta w_{50\%}} \quad \text{(SI units)} \tag{65}$$

It is useful to evaluate ICR mass resolution for either $T \gg \tau$ (high-pressure limit) or $T \ll \tau$, where T is the time-domain acquisition period and τ is the time-domain exponential damping constant (Equation 31). The results are shown in Table 2. In the low-pressure limit, the time-domain ICR signal persists, undamped, throughout the acquisition period T. In the high-pressure limit, the ICR time-domain signal damps essentially to zero during the acquisition period.

Table 2 The FTICR mass resolution $\Delta m_{50\%}$, and mass resolving power $m/\Delta m_{50\%}$, in which $\Delta m_{50\%}$ is the full magnitude-mode FTICR/MS peak width at half-maximum peak height.[60] The factor T_{acq} is time-domain ICR signal acquisition period, and τ is the collisional damping constant (Equation 31)

Parameter	Low pressure ($T_{acq} \ll \tau$)	High pressure ($T_{acq} \gg \tau$)
$\Delta m_{50\%}$	$\frac{7.583\,m^2}{qB_0T_{acq}}$	$\frac{2\sqrt{3} \times m^2}{qB_0\tau}$
	(SI units)	
$\Delta m_{50\%}$	$\frac{7.859 \times 10^{-8} m^2}{zB_0T_{acq}}$	$\frac{3.590 \times 10^{-8} m^2}{zB_0\tau}$
	(m/z in unified atomic mass units per elementary charge)	
$\frac{m}{\Delta m_{50\%}}$	$\frac{0.132qB_0T_{acq}}{m}$	$\frac{qB_0\tau}{2\sqrt{3} \times m}$
	(SI units)	
$\frac{m}{\Delta m_{50\%}}$	$\frac{1.274 \times 10^7 zB_0T_{acq}}{m}$	$\frac{2.785 \times 10^7 zB_0\tau}{m}$
	(m/z in unified atomic mass units per elementary charge)	

In the low-pressure limit, the corresponding FTICR mass spectral peak width is independent of m/z. Thus, as m/z increases, the peak width remains relatively constant, but the peaks are closer together, because ICR frequency varies inversely with m/z. That is why (see Equation 64 and Table 2), for a given ion–neutral collision frequency, ICR mass resolving power (at constant B_0, for a given T_{acq}) varies inversely with m/z, as shown in Figure 15.

It is worth noting that Figure 15 applies to other magnet-based mass separators (notably magnetic sector mass spectrometers). However, in magnetic sector MS, it is usual to vary the magnetic field strength in order to bring ions of a given m/z into focus. In that case, magnetic sector mass resolving power might follow one of the dashed lines in Figure 15 (i.e. if FTICR and magnetic sector mass resolving power are the same at m/z 1000, then the sector resolving power will remain constant throughout the m/z range, whereas the FTICR resolving power will increase with decreasing m/z). Thus, FTICR/MS achieves the highest mass resolution possible for magnet-based mass analysis, by operating at maximum magnetic field strength throughout the mass range. (A fixed-field magnet also produces a spatially more homogeneous field, thereby further increasing the mass resolving power.)

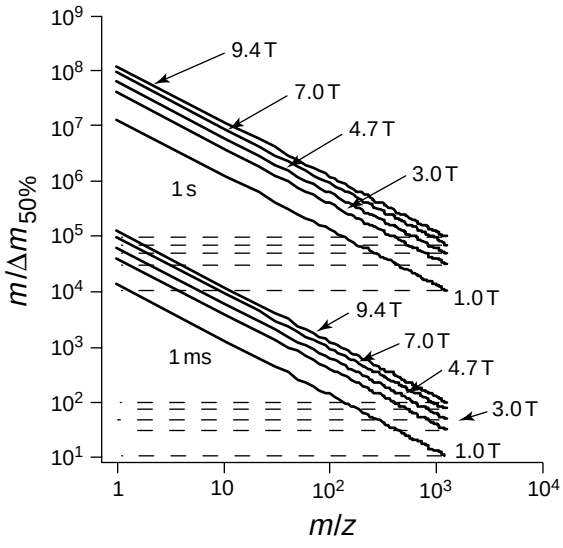

Figure 15 The FTICR theoretical mass resolving power at 1.0, 3.0, 4.7, 7.0, and 9.4 T as a function of mass-to-charge ratio m/z, in unified atomic mass units per elementary charge, in the low-pressure limit (i.e. no ion-neutral collisions during the detection period), for $T_{acq} = 1$ ms or 1 s, and a fixed ion–neutral collision frequency. Although the mass resolving power decreases with increasing m/z, the mass resolution can be spectacularly high if the time-domain acquisition period is long enough. The dashed lines indicate mass resolving power for magnetic field-swept instruments (e.g. magnetic sectors).

9 UPPER MASS AND ENERGY LIMIT(S)

9.1 Trap Dimension Mass Limit

Because an ion trap must contain side electrodes in order to provide for excitation and/or detection of the ICR signal, it is clear that the ultimate upper mass limit is the mass at which the ion cyclotron radius of a thermal ion reaches the radius of the trap. Equation (10) gives the cyclotron radius of an ion of average thermal translational energy kT, from which we can solve for the highest mass ion for which the cyclotron radius is less than the trap radius $d/2$. Rewriting Equation (10) we find that at 7.0 T, the upper mass limit for a singly charged room-temperature ion would be about 5.89 MDa in a trap of 1 in cross-sectional radius (Equations 66 or 67):

$$m_{upper} = \frac{q^2 B_0^2 r_{max}^2}{2kT} \quad \text{(SI units)} \tag{66}$$

or

$$m_{upper} = \frac{5.598 \times 10^{11} z^2 B_0^2 r_{max}^2}{T} \tag{67}$$

where r is in meters, B_0 in tesla, m in unified atomic mass units, T in kelvin, and z in multiples of elementary charge. Of course, the actual upper mass limit will be smaller, because ions must begin with an ICR orbital radius much smaller than that of the trap in order that a detectable coherent signal be generated by exciting their cyclotron orbital motion to a larger radius for detection.

9.2 Trapping Potential Mass Limit

Equation (66) is valid only in the absence of electric field (from either the applied electrostatic trapping potential or Coulomb repulsions between ions). In the presence of a three-dimensional axial quadrupolar electrostatic trapping potential, Figure 14 shows the magnetron, reduced cyclotron, and trapping oscillation frequencies as a function of m/z. Analysis of Equations (49–52) shows that the magnetron and reduced cyclotron frequencies converge to a common value, $w_+ = w_- = w_c/2 = qB_0/2m$, at the so-called critical m/z,[49] namely, when Equation (68) holds:

$$\left(\frac{w_c}{2}\right)^2 = \frac{w_z^2}{2} \tag{68}$$

or

$$m_{critical} = \frac{qB_0^2 a^2}{4V_{trap}\alpha} \quad \text{(SI units)} \tag{69}$$

$$m_{critical} = \frac{1.20607 \times 10^7 z B_0^2 a^2}{2V_{trap}\alpha} \tag{70}$$

In Equations (69) and (70) a is in meters, B_0 in tesla, m in unified atomic mass units, V_{trap} in volts, and z in multiples of elementary charge.

For $m/z > m_{critical}$, ion cyclotron motion is no longer stable, and the ion spirals outward until it is lost from the trap. For example, at 7.0 T, for $V_{trap} = 1$ V applied to a cubic trap ($\alpha = 2.773\,73$) of $a = 2.54$ cm, $m_{critical} = 274\,000\ z$. Figure 16 shows the dependence of $m_{critical}$ on trapping voltage, at each of five magnetic field strengths. (The actual upper mass limit in each case is slightly lower if we specify that (say) 99% of thermal ions fit within the 1 in separation between the detector electrodes.) Figure 16 also shows the desirability of cooling ions so that the trapping potential may be lowered, thereby increasing the upper mass limit.

9.3 Trapping Potential Energy Limit; Space Charge and the Peak Coalescence Mass Limit

Obviously, ions with translational energy higher than the potential well depth (which in turn scales linearly with the trapping potential applied to each end cap electrode) can escape axially from the trap. That problem may be solved by increasing the trapping voltage; however, it is also possible to trap too many ions. In fact, when more than about 10 000 ions are present in the trap, it becomes

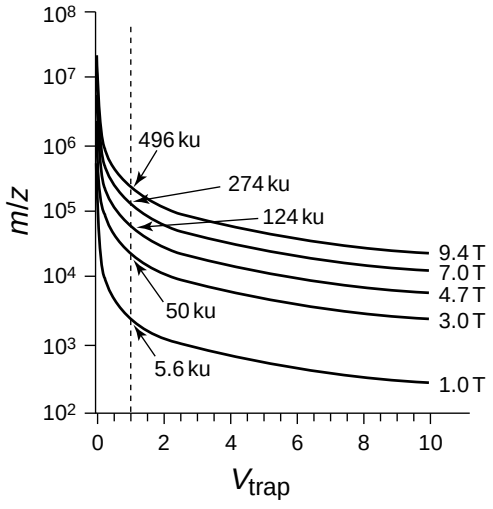

Figure 16 Upper mass limit for a singly charged ion in a cubic Penning trap, for each of five common magnetic field strengths: 1.0, 3.0, 4.7, 7.0, and 9.4 T. At $V_{trap} = 1$ V, each arrow denotes the highest m/z for which an ion has a stable orbit, independent of ion initial velocity (temperature).

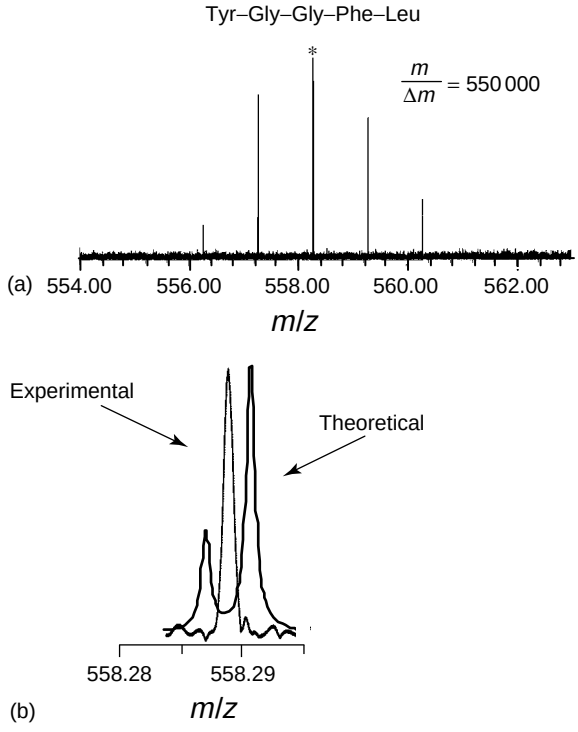

Figure 17 Coulomb-mediated peak coalescence. (a) Matrix-assisted laser desorption/ionization (MALDI) FTICR magnitude-mode mass spectrum of quasimolecular ions, $(M + H)^+$, of partially deuterated leucine enkephalin. Although mass resolving power greater than 550 000 has been achieved (the experimental peak in (b)), the $^{12}C^2H/^{13}C^1H$ doublet is not resolved (the theoretical doublet computed from two magnitude-mode Lorentzians at a resolving power of $m/\Delta m_{50\%} \approx 550\,000$ in (b)), due to the Coulomb interaction of ions of two nearly identical mass-to-charge ratios. (Reproduced with permission from Pasa-Tolic et al.[62])

necessary to consider the static and dynamic effects of ion–ion repulsions. It can be shown that in a spatially uniform static electromagnetic field, Coulomb repulsions between ions do not affect the ICR orbital frequency of the ion packet.[61] However, in the spatially nonuniform electromagnetic field of a typical ion trap, Coulomb repulsions can shift and broaden FTICR mass spectral peaks, by pushing like-charge ions apart into regions of different applied external electric or magnetic field.

Peak coalescence (i.e. observation of a single FTICR mass spectral peak whose width is narrower than the separation between the cyclotron frequencies of ions of two different m/z) occurs when two ion clouds have sufficiently large ion populations and very similar mass-to-charge ratios (Figure 17). Theoretically, for two very long cylindrical ion clouds whose single-ion cyclotron radius R_c and ion cloud radius ρ_c are initially overlapping, the ion clouds coalesce if the difference in cyclotron frequency, $w_2 - w_1$, between the two clouds is (Equation 71)

$$\Delta w_c < \frac{N_1 q_1 + N_2 q_2}{2\pi^2 \varepsilon_0 L B_0 R_c s_{eff}} \tag{71}$$

in which ε_0 is the vacuum permittivity, $N_1 q_1$ or $N_2 q_2$ is the total charge contained in each cylinder, L is the length of cylinder, and B_0 is the magnetic field induction. $s_{eff} \approx 1.04\rho_c$ is calculated numerically, based on an initial separation between the two ion clouds of the order of the ion cloud radius ρ_c.[63] Thus, the tendency for two ion clouds to coalesce to a single experimentally observable ion cyclotron frequency varies directly with the number of ions in each cloud, and inversely with magnetic field strength and ion cyclotron orbital radius. Furthermore, one can determine the maximum molecular weight at which coalescence first occurs from Equation (72) or (73),[64]

$$m_{max} \approx 3.2 B_0 \sqrt{\frac{\varepsilon_0 L R_c \rho_c \Delta m}{N_{ave}}} \quad \text{(SI units)} \tag{72}$$

or

$$m_{max} \approx 2.333\,27 \times 10^5 B_0 \sqrt{\frac{L R_c \rho_c \Delta m}{N_{ave}}} \tag{73}$$

where L, R_c, and ρ_c are in centimeters, and m in unified atomic mass units. The factor $N_{ave} = (N_1 + N_2)/2$ is the average number of ions in the two clouds. Equation (72) leads to a particularly useful theoretical prediction, namely, that the highest mass at which ions whose masses differ by 1 Da begin to coalesce is about $10\,000 B_0$ Da.[64] Although based on a highly idealized theoretical model

(and therefore not to be taken too literally), Mitchell's result provides good insight into the variables that affect peak coalescence.

10 ION SOURCES

10.1 Internal Ionization

Successful application of the FTICR/MS technique requires ionization and trapping of the species of interest. For volatile compounds these processes may be achieved by leaking the corresponding neutral vapor into the volume of the ion trap (either through a valve or by applying less volatile substances to a probe surface, positioning the probe adjacent to the ion trap, and heating to vaporize the substance) and ionizing in situ (known as internal ionization). The most common internal ionization technique is electron ionization (EI) (photoionization (PI) is also in use). In EI, an electron beam of controllable energy and current (typically 70 eV and $\sim$1 μA) is directed through the center of the ion trap for a specified time period. The interaction between the electrons and neutral molecules can result in either ejection of a valence electron to form positive ions or (for very low energy electrons) electron capture to form negative ions. The electron beam can also be used to perform chemical ionization (CI), in which a suitable reagent is ionized by the electron beam and then undergoes proton or electron transfer to form a chemically ionized analyte. PI is typically performed in the same manner but with a suitable light source (i.e. laser, arc lamp, etc.) directed along the central axis of the ion trap (although not required, ions are almost always formed or injected along the central trap axis for optimal FTICR performance). These techniques were the first to be implemented with FTICR mass analysis, and have provided ions for a voluminous number of studies relating to ion–molecule reaction pathways, ion structure and energetics, kinetics, and compound identification.[65–70]

10.2 External Ionization

Although internal ionization methods encompass a wide range of experiments, it is often desirable to couple FTICR mass analysis with ion sources that form ions outside the trap, mainly because nonvolatile substances that are not amenable to heating can be effectively desorbed into the gas phase and ionized (e.g. metals, peptides, proteins, oligonucleotides, oligosaccharides, lipids, synthetic polymers, etc.). These external (to the ICR trap) ionization methods may be divided into two categories. In the first, the nonvolatile substance is deposited on a probe surface and the probe is positioned adjacent to the ion trap. A desorption/ionization event follows and the resulting ions can be directed toward the trap and captured. Examples include laser desorption/ionization (or laser desorption of neutrals followed by EI[71] and MALDI,[72–74] as well as many others.[75–79] The second class of external ionization techniques operates at such high pressure (>1 torr) that the ion source must be separated from the ion trap by several stages of differential pumping in order to achieve the necessarily low pressure ($<1 \times 10^{-8}$ torr) required for optimal mass analysis. These methods include EI,[80,81] cluster sources,[82] and high-pressure sources.[83,84] The multiple pumping stages usually require that the source be located outside the solenoidal ICR magnet (although there is an exception[85]) so that ions formed by these methods are injected through the magnetic fringe field by a suitable ion guide (electrostatic Einzel lenses,[31,32] multipole RF ion guides,[86] and DC wire ion guides[12,13] have been used). Although more complex, the ion guides have worked well enough that many ionization techniques that can be located inside the magnet have also been implemented outside the magnet,[87,88] usually for convenience in accessing the ion source.

10.3 Chromatographic Interfaces

The FTICR/MS technique is capable of powerful mixture analysis due to its high mass range and ultrahigh mass resolving power. However, in many cases it is still desirable to couple a chromatographic interface to the mass spectrometer for sample purification, preconcentration, and mixture separation. The first separation technique to be coupled with FTICR was gas chromatography (GC).[89] The GC effluent is admitted into the ion trap region of the vacuum chamber and ions are formed with an electron beam. More recent efforts have focused on liquid chromatography (LC)[90–92] and capillary electrophoresis (CE),[93] both coupled to the mass spectrometer by use of EI. These couplings have produced exciting new advances in the analysis of biopolymers present in complex biological matrices.[90,94]

11 TANDEM MASS SPECTROMETRY METHODS

MS/MS consists of activation of primary (parent) ions, resulting in dissociation or reaction followed by mass analysis of the resulting secondary (product) ions.[145] In an FTICR instrument, activation is commonly achieved by the following: collisions of ions with neutrals (CAD or CID)[95] or surfaces (surface-induced dissociation (SID));[54,55,72,73,96] ultraviolet photodissociation

(UVPD);[97] multiphoton infrared photodissociation (IRMPD);[98–100] or blackbody infrared dissociation (BIRD).[101] For CAD (CID) of macromolecular ions, the most popular method is sustained off-resonance irradiation (SORI),[102] although very low-energy (VLE)[103] and multiple excitation collisional activation (MECA)[104] are also available. Figure 18 shows that SORI, VLE, and MECA each provide for periodic excitation of ion cyclotron radius (and thus ion translational energy) to an adjustable maximum value; SORI tends to be preferred, probably because it is simplest to implement and tune.

Photodissociation of large biomolecules has also proved effective for biomolecule characterization. Typically infrared (IR; 10.6 μm) laser photons are used for slow heating and fragments similar to those obtained by CAD are produced.[98] One advantage of IRMPD is that gas pulses are not necessary, so that high-resolution FTICR detection can occur quickly after dissociation (no pumpdown is required). An alternative to laser photodissociation takes advantage of the blackbody radiation produced by a heated vacuum chamber (BIRD).[105,106] Ions are confined in the hot ICR trap (10–1000 s), where ion structure- and temperature-dependent interaction with the blackbody radiation occurs. A unique feature of BIRD is that the ion cloud assumes a known (i.e. Boltzmann) and controllable internal energy distribution. Thus, information on dissociation energetics and mechanisms can be obtained from the temperature dependence of the unimolecular dissociation rate constants.[105,106]

Finally, SID has been investigated as a probe of ion structure by FTICR/MS. The advantage of SID is the potential for high internal energy deposition with a relatively narrow distribution. However, results to date indicate that collection of product ions is difficult and that unique structural information is rarely obtained.[54,55,72,73,96,107]

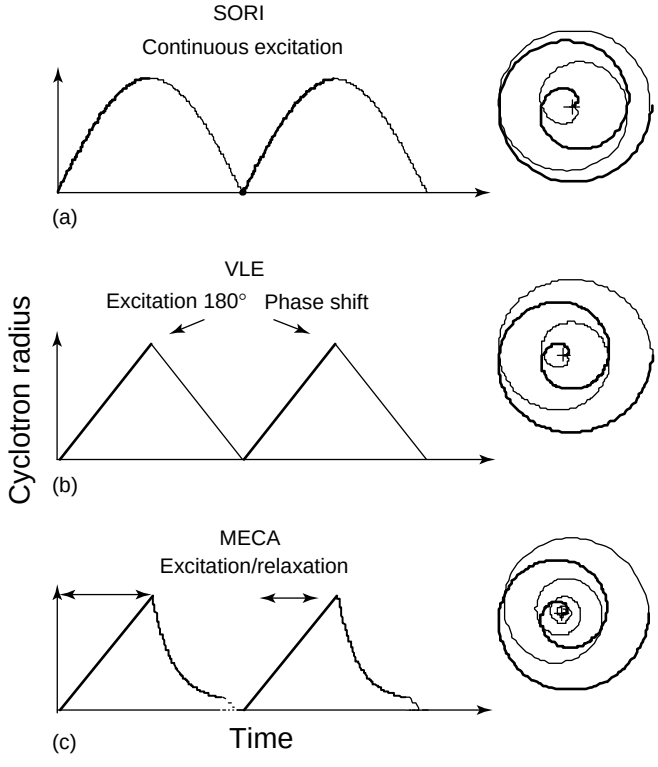

Figure 18 Time evolution of ion cyclotron radius (left) and ion *xy* trajectory (right) for each of three ICR techniques for ion multiple activation based on repeated single-frequency dipolar excitation for CID. (a) In SORI,[102] ions of a selected m/z ratio are alternately excited and de-excited due to the difference between the excitation frequency and the ion cyclotron frequency. (b) In VLE CID,[103] ions are alternately excited and de-excited by resonant excitation whose phase alternates bimodally between 0° and 180°. (c) In MECA,[104] ions are resonantly excited and then allowed to relax by collisions. (Reproduced with permission from Guan and Marshall.[7])

12 ADVANTAGES OF HIGH MAGNETIC FIELDS

It was noted above that mass resolving power in FTICR/MS increases linearly with increasing applied magnetic field induction **B** for fixed ion–neutral collision frequency. In fact, seven other FTICR primary performance parameters theoretically also increase linearly (data acquisition speed, upper mass limit for peak coalescence) or quadratically (upper mass limit due to trapping potential, maximum ion kinetic energy, maximum number of trapped ions, maximum ion trapping duration, two-dimensional FTICR mass resolving power) with increasing **B** (Figure 19). Two-dimensional FTICR/MS, such as two-dimensional nuclear Overhauser effect NMR spectroscopy, provides an automated method for determining ion–molecule reaction pathways, kinetics, and equilibria simultaneously for all primary ions in a mixture.[108–110] The origin of (and conditions for) the magnetic field dependence of each of these parameters have been collected and discussed.[64,111] These fundamental advantages lead to corollary improvement in other FTICR performance parameters, such as signal-to-noise ratio, dynamic range, mass accuracy, ion remeasurement efficiency, and mass selectivity for MS/MS. Finally, these various advantages may be exploited in combination, so as to produce even higher enhancement in a particular parameter; for instance, the signal-to-noise ratio can improve by more than a factor of $\mathbf{B}^2$ if the mass resolving power is fixed at the same value as at lower magnetic field.[112]

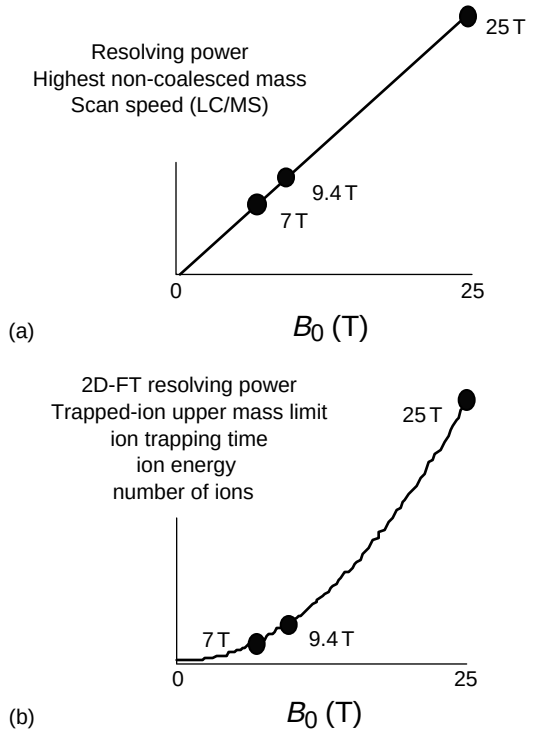

Figure 19 Performance parameters of FTICR/MS as a function of applied static magnetic field induction B_0. (a) Parameters that increase linearly with increasing magnetic field. (b) Parameters that increase quadratically with increasing magnetic field.

13 FOURIER TRANSFORM SPECTROSCOPY ASPECTS

The FTICR/MS technique shares many common conceptual and data reduction features with other types of FT spectroscopy, as has been discussed in depth.[14] The homology between FTICR and FTNMR is deep and virtually complete.[113] The most unique advantage of FTICR as a mass analyzer is that the ion mass-to-charge ratio is experimentally manifested as a frequency. Because frequency can be measured more accurately than any other experimental parameter, ICR/MS therefore offers inherently higher resolution (and thus higher mass accuracy) than any other type of mass measurement. The introduction of FT techniques to ICR/MS[1–4] brought not only the Fellgett (multichannel, opening the exit slit) advantages of increased speed (factor of 10 000), or increased sensitivity (factor of 100), but also the advantages of fixed magnetic field rather than swept field,[9,10] namely increased mass resolution (factor of 10 000), and increased mass range (factor of 500). Applications deriving from these advantages include determination of chemical formulas (particularly in complex mixtures), detection limit in the attomole range, and multistage MS^n.

FT manifestations in ICR include: Nyquist sampling and foldover;[114] fast FT; zero-filling,[115] windowing or apodization;[14] deconvolution;[15,116,117] oversampling;[118] and two-dimensional Hadamard[54–56] or Fourier MS/MS.[110] In addition, various non-FT methods for obtaining a frequency-domain (and thus mass-domain) spectrum from a time-domain interferogram ICR signal include: the Hartley transform (a way of performing a FT on real-only data);[119,120] the Bayesian maximum entropy method (MEM);[121,122] and linear prediction.[59,123,124]

However, several aspects of FTICR data reduction differ from those of FT interferometry or FTNMR spectroscopy. Notably, the phasing of FTICR spectra over a broad frequency range is difficult;[125] thus, spectra are usually reported in magnitude mode rather than absorption mode.[1–4] As a result, resolving power is lower than for absorption mode by a factor ranging from about $\sqrt{3}$ (Lorentzian peak shape) to 2 (sinc peak shape),[60] and new peak-fitting algorithms have been developed.[126] Apodization of magnitude-mode spectra has been performed by means of magnitude-mode multiple-derivative techniques.[127] Moreover, FTNMR is conducted exclusively in heterodyne-mode, whereas FTICR/MS is often performed by sampling the ICR signal directly without heterodyning. Due to the need for massively large datasets (several megawords of time-domain data) to take advantage of potentially ultrahigh mass resolving power over a wide mass range, data clipping (to as low as 1 bit/word) has been demonstrated, with only modest distortion of the FT spectrum.[128] Finally, the ICR time-domain signal typically exhibits non-exponential damping;[26] hence, it is common to apodize the time-domain signal with a window function (e.g. Blackman-Harris) that preferentially damps the initial portion of the time-domain signal prior to Fourier transformation.[14]

14 RELATION TO PAUL (QUADRUPOLE) ION TRAP

The close relation between Penning (ICR) and Paul (quadrupole) ion traps is perhaps most evident from the 1989 Nobel Prize, which was shared by Dehmelt and Paul, the two most famous developers of the two techniques. Many trapped-ion techniques first introduced in FTICR have since been adapted to the Paul trap, including: CID of trapped ions;[129] mass-selective ion ejection;[130] frequency-sweep excitation;[131] stored-waveform excitation/ejection;[7,132–134] FT detection of an induced image current;[135,136] and the interfacing of various ion sources to the mass analyzer.

15 SELECTED APPLICATIONS

15.1 Elemental Composition from Accurate Mass Measurement

A powerful advantage of FTICR over other mass analyzers is accurate mass capability. For singly charged ions of below 700 Da, a unique elemental composition can be assigned directly from the measured mass if ~1 ppm mass accuracy can be achieved.[19,137] An example is shown in Figure 20, in which EI FTICR mass analysis of raw diesel fuel feedstock resolves ~500 singly charged ion masses in the range 90–300 Da. The ultrahigh resolving power afforded by FTICR/MS reveals several different species at each nominal mass. With proper mass calibration (see Equation 53), the elemental composition of each species can be assigned unambiguously based on sub-ppm mass accuracy. Elemental composition is critical in such analyses because it is important to identify and monitor sulfur- and nitrogen-containing species as they are removed during the fuel purification process.

15.2 Detection Limit for Biological Analysis

A fundamental limit of FTICR broadband image current detection is that typically about 100 ions of a given mass-to-charge ratio are required to induce a measurable signal. Ion counting (destructive-detection) mass spectrometers are therefore inherently more sensitive. However, the FTICR detection limit can still be spectacularly low. With CE coupling, Hofstadler et al. acquired hemoglobin mass spectra from a single red blood cell (containing about 450 amol of hemoglobin).[94] An even lower detection limit has been achieved, where CE/FTICR mass spectra have been observed from sub-attomole protein samples.[138] Further, the same researchers used a CAD spectrum from 9 amol of carbonic anhydrase and database searching to unambiguously identify the protein in spite of N-terminal acetylation. Although lower detection limits have been obtained with other mass analyzers, the unique combination afforded by accurate mass measurement, ultrahigh resolution, and nondestructive detection allowing for MS^n in combination with a very low detection limit makes FTICR extremely attractive for biological analysis.

A key feature common to these impressive results is the CE coupling. However, a major disadvantage of CE sample introduction is the requirement of a fairly concentrated analyte solution. The LC/FTICR method is an attractive alternative for the very dilute solutions often encountered in biological samples. Figure 21 shows a selected-ion chromatogram and mass spectra from LC/FTICR mass analysis of a three-component peptide mixture.[139] Each of those peptides was present at 500 pM concentration in a 100 mM salt aqueous matrix! Sample loading onto a specially designed nano-LC column/microelectrospray tip facilitates sample desalting, preconcentration and gradient separation.[140]

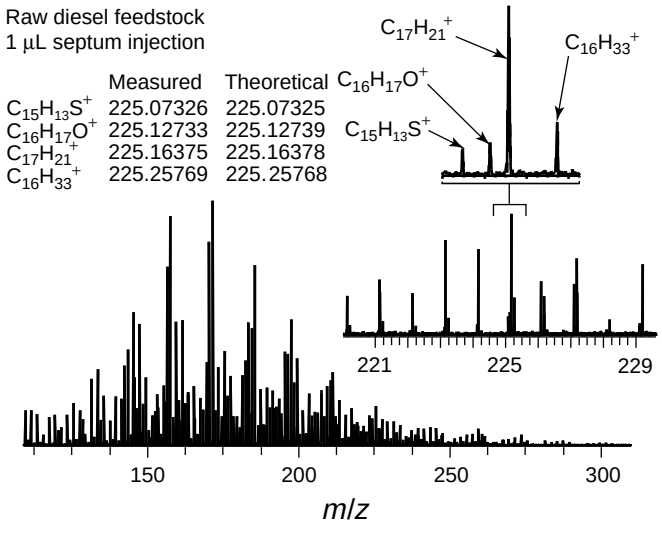

Figure 20 Low-energy electron impact ionization 5.6 T FTICR mass spectrum, showing about 500 resolved peaks, for a 1 μL septum injection of raw diesel feedstock. The inset mass window shows resolution of four ions of different elemental composition at the same nominal mass, each identified with mass accuracy below 0.3 ppm at a mass resolving power $m/\Delta m_{50\%} \approx 100\,000$. (Data kindly provided by R. Rodgers.)

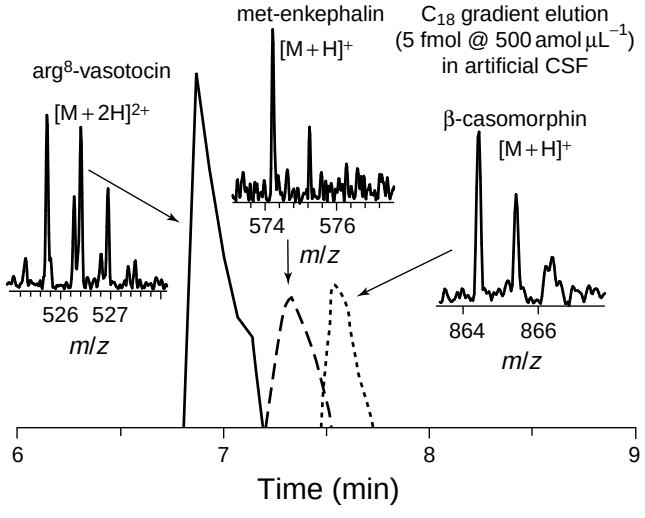

Figure 21 True gradient elution of 5 fmol each of an equimolar mixture of three peptides: Arg^8-vasotocin, methionine enkephalin, and β-casomorphin. Samples were dissolved in artificial cerebrospinal fluid at 500 amol μL^{-1} each and loaded (10 μL) onto a C_{18} packed micro-ESI needle. The reconstructed ion chromatogram is shown below, with ESI/FTICR mass spectra of individual peptide spectra shown as insets. (Reproduced from Emmett et al.[139])

15.3 High Mass

Multiple-charging inherent to ESI has revolutionized the mass analysis of large molecules.[141] Typically the highest-mass species carry the greatest number of charges, thereby yielding ions with mass-to-charge ratio $500 < m/z < 2500$, especially favorable for FTICR detection even at very high ion mass. The combination of ESI with FTICR detection has yielded striking results in the analysis of large ions. Unit mass resolution and accurate mass analysis are now routine for proteins as large as 67 kDa[112,142,143] and have been achieved for two 112 kDa proteins,[144] the highest-mass molecules for which unit mass resolution has been achieved (Figure 22). In that work, the masses of two chondroitinase enzymes were determined with 3 Da accuracy, information that helps in obtaining FDA approval for use of the enzymes as drugs. Finally, trapping and detection of a single 100 MDa DNA ion with around 30 000 charges has been achieved.[146]

15.4 Isotopic Amplification for Unit Mass Accuracy of Biomacromolecules

Several related problems arise from the presence of multiple isotopes in biological macromolecular ions. First, a typical electrospray-ionized protein can take on about one proton per kilodalton, by protonation of available arginine and lysine residues. Because mass analyzers in general (and ICR in particular) separate ions based on mass-to-charge ratio (m/z, in which m is ion mass in unified atomic mass units and z is charge in multiples of the elementary charge), the first problem is how to determine charge independent from mass. Fortunately, if the isotopic envelope can be resolved to better than 1 u (Figure 22a), adjacent peaks differ by $m/z = 1/z$; thus, the ion charge may be obtained simply as the reciprocal of the spacing between two adjacent peaks in the m/z spectrum.[147] Because only FTICR/MS can resolve such multiplets for macromolecular ions of more than a few kilodaltons, FTICR/MS is the method of choice for mass analysis of such species.

A second (more vexing) problem is that even if m/z for each of the resolved isotopic peaks can be determined to ppm accuracy, the molecular weight can still be in error by a whole dalton or more, because (except for the lowest-mass monoisotopic species – namely, the unique elemental composition for which all carbons are ^{12}C, all nitrogens are ^{14}N, all oxygens are ^{16}O, all sulfurs are ^{32}S, etc.), it is necessary to match the observed isotopic abundance distribution to that predicted for, for instance, a protein of average amino acid composition.[148] However, if the relative abundances are just a bit in error, or if the unknown protein differs in composition

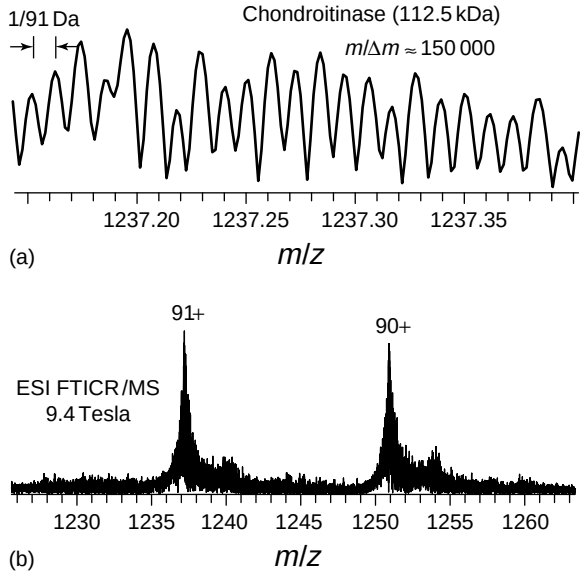

Figure 22 The ESI/FTICR mass spectra of chondroitinase I. (a) Mass scale-expansion showing unit mass resolution of the isotopic distribution of the $z = 91$ charge state. (Data kindly provided by N. Kelleher and described in detail elsewhere.[14]) (b) Heterodyne data for SWIFT-isolated ions, $1226 < m/z < 1273$, with external ion accumulation,[90] from 10 co-added time-domain signals; the peaks at m/z 1240 and 1254 correspond to an unidentified adduct of ~260 Da.

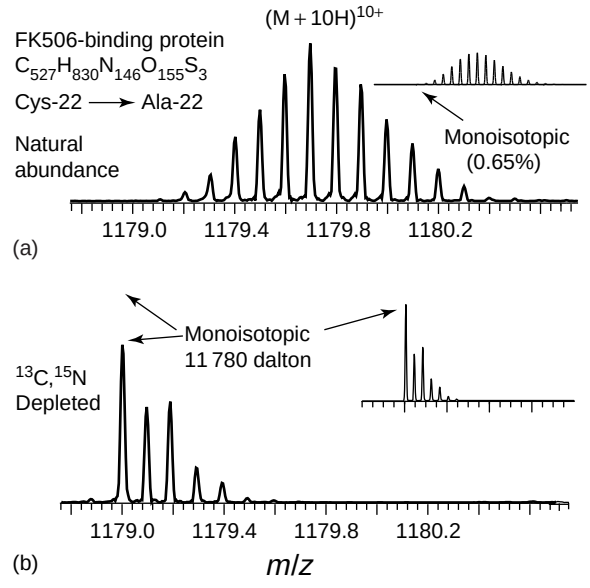

Figure 23 The ESI/FTICR mass spectra (9.4 T) of a mutant (C22A) FK506-binding protein. (a) Natural-abundance isotopic distribution (~98.89% ^{12}C; ~99.63% ^{14}N). (b) Isotopic distribution for the same protein grown on a medium with 99.95% ^{12}C and 99.99% ^{14}N. Insets: Isotopic distributions calculated (same vertical scale) from the chemical formula for natural-abundance (a) and ^{13}C, ^{15}N doubly depleted (b) FK506 binding protein. (Reproduced with permission from Marshall et al.[149])

from the average protein in the database, then the estimated molecular weight can be in error by one or more daltons.

A simple solution to the above problem is to express a protein from a minimal medium containing ^{13}C-depleted glucose and ^{15}N-depleted ammonium sulfate. An example is shown in Figure 23. Both simulated and experimental electrospray FTICR mass spectra show that double-depletion of ^{13}C and ^{15}N effectively narrows and shifts the isotopic distribution to the left, so that the monoisotopic species is now prominent and easily identified.[149] This technique promises to extend the upper mass limit for protein MS by about an order of magnitude, and is optimally exploited only by FTICR/MS.[149] By measuring protein mass accurately to within 1 Da, it becomes possible to: count the number of disulfide bridges ($-S-S-$ is 2 Da lighter than two $-SH$); identify deamidation ($-NH_2$ is 1 Da lighter than $-OH$); identify post-translational modifications and noncovalent adducts; and identify proteome components directly without prior two-dimensional gel separation.

ACKNOWLEDGMENTS

This work was supported by grants from NSF (CHE-93-22824), N.I.H. (GM-31683), the NSF National High Field FTICR Mass Spectrometry Facility (CHE-94-13008 and CHE-99-09502), Florida State University, and the National High Magnetic Field Laboratory. This article has been reproduced with the permission of John Wiley & Sons Inc., from *Mass Spectrometry Reviews*, **17**(1), 1–35 (1998).

ABBREVIATIONS AND ACRONYMS

BIRD	Blackbody Infrared Dissociation
CAD	Collision-activated Dissociation
CE	Capillary Electrophoresis
CI	Chemical Ionization
CID	Collision-induced Dissociation
DC	Direct Current
EI	Electron Ionization
ESI	Electrospray Ionization
FT	Fourier Transform
FTICR	Fourier Transform Ion Cyclotron Resonance
FTICR/MS	Fourier Transform Ion Cyclotron Resonance/Mass Spectrometry
FTNMR	Fourier Transform Nuclear Magnetic Resonance
GC	Gas Chromatography
ICR	Ion Cyclotron Resonance
IR	Infrared
IRMPD	Multiphoton Infrared Photodissociation
LC	Liquid Chromatography
MALDI	Matrix-assisted Laser Desorption/ Ionization
MECA	Multiple Excitation Collisional Activation
MEM	Maximum Entropy Method
MS	Mass Spectrometry
MS^n	Multiple Mass Spectroscopy
MS/MS	Tandem Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PI	Photoionization
RF	Radiofrequency
SID	Surface-induced Dissociation
SORI	Sustained Off-resonance Irradiation
SWIFT	Stored Waveform Inverse Fourier Transform
UVPD	Ultraviolet Photodissociation
VLE	Very Low-energy

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • Mass Spectrometry in Structural Biology • Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis • Proteoglycan and Acidic Polysaccharide Analysis

Clinical Chemistry **(Volume 2)**
Laboratory Instruments in Clinical Chemistry, Principles of • Nucleic Acid Analysis in Clinical Chemistry

Coatings **(Volume 2)**
Coatings Analysis: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Gas Chromatography with Selective Detectors for Amines

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Nuclear Magnetic Resonance for Environmental Monitoring • Organic Analysis in Environmental Samples by Capillary Electrophoresis •

Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Soil Instrumental Methods • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Volatile Organic Compounds in Groundwater, Probes for the Analysis of • Water Analysis: Organic Carbon Determinations

Food **(Volume 5)**
Proteins, Peptides, and Amino Acids Analysis in Food

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Mass Spectrometry for Forensic Applications

Industrial Hygiene **(Volume 6)**
Sampling and Recovery Techniques for the Determination of Gases and Vapors in Air

Nucleic Acids Structure and Mapping **(Volume 6)**
DNA Structures of Biological Relevance, Studies of Unusual Sequences • Mass Spectrometry of Nucleic Acids • PNA and Its Applications • Polycyclic Aromatic Compounds Mapping • Sequencing Strategies and Tactics in DNA and RNA Analysis

Peptides and Proteins **(Volume 7)**
Separation and Analysis of Peptides and Proteins: Introduction • Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis • Posttranslational Oxidative Modifications of Proteins • Protein–Drug Interactions • Protein–Oligonucleotide Interactions • Proteolytic Mapping

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Petroleum Residues, Characterization of

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Chemometrics **(Volume 11)**
Signal Processing in Analytical Chemistry

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation

Radiochemical Methods **(Volume 14)**
Mass Spectrometry of Long-lived Radionuclides

APPENDIX 1: PHYSICAL CONSTANTS AND PRECISE ELEMENTAL MASSES

Physical Constants[150]

Elementary charge
$= q = 1.602\,177\,33 \times 10^{-19}$ coulombs $\pm$ 0.30 ppm
$= 4.806\,53 \times 10^{-10}$ statcoulombs or esu
so that 1 eV
$= 1.602\,177\,33 \times 10^{-19}$ J
Boltzmann constant
$= k = 1.380\,658 \times 10^{-23}\ \mathrm{J\,K^{-1}} \pm 8.5$ ppm
Molar gas constant
$= R = 8.314\,510\ \mathrm{J\,mol^{-1}\,K^{-1}} \pm 8.4$ ppm
Atomic mass unit
$= \mathrm{u} = \mathrm{Da} = 1.660\,540\,2 \times 10^{-27}\ \mathrm{kg} \pm 0.59$ ppm
Avogadro constant
$= N_A = 6.022\,136\,7 \times 10^{23}\ \mathrm{mol^{-1}} \pm 0.59$ ppm
Electron mass
$= 9.109\,389\,7 \times 10^{-31}\ \mathrm{kg} \pm 0.59$ ppm
Speed of light
$= c = 2.997\,924\,58 \times 10^{8}\ \mathrm{m\,s^{-1}}$
Conversions: 1 eV/c^2
$= 1.782\,662\,696 \times 10^{-36}$ kg
$= 1.073\,543\,836 \times 10^{-9}$ u $\pm$ 0.3 ppm

Precise Masses of Various Common Elements[151]

^{1}H = 1.007 825 032 14 u $\pm$ 0.000 35 μu
(one standard deviation)
^{2}H = 2.014 101 777 99 u $\pm$ 0.000 36 μu
^{3}H = 3.016 049 267 5 u $\pm$ 0.0011 μu
^{3}He = 3.016 029 309 70 u $\pm$ 0.000 86 μu
^{4}He = 4.002 603 249 7 u $\pm$ 0.0010 μu
^{13}C = 13.003 354 837 8 u $\pm$ 0.0010 μu
^{14}C = 14.003 241 988 4 u $\pm$ 0.0040 μu

$^{14}N = 14.003\,074\,005\,24\,u \pm 0.000\,86\,\mu u$
$^{15}N = 15.000\,108\,898\,44\,u \pm 0.000\,92\,\mu u$
$^{16}O = 15.994\,914\,622\,1u \pm 0.0015\,\mu u$
$^{17}O = 16.999\,131\,501\,u \pm 0.22\,\mu u$
$^{18}O = 17.999\,160\,419\,u \pm 0.9\,\mu u$
$^{20}Ne = 19.992\,440\,175\,9\,u \pm 0.0020\,\mu u$
$^{23}Na = 22.989\,769\,675\,u \pm 0.23\,\mu u$
$^{28}Si = 27.976\,926\,532\,7\,u \pm 0.0020\,\mu u$
$^{31}P = 30.973\,761\,512\,u \pm 0.20\,\mu u$
$^{32}S = 31.972\,070\,690\,u \pm 0.12\,\mu u$
$^{34}S = 33.967\,866\,831\,u \pm 0.11\,\mu u$
$^{39}K = 38.963\,706\,861u \pm 0.3\,\mu u$
$^{40}Ar = 39.962\,383\,123\,2\,u \pm 0.0030\,\mu u$

FURTHER READING

Books and Journal Articles on Fourier Transform Ion Cyclotron Resonance

B. Asamoto, *Spectroscopy*, **3**, 38–46 (1988). (Reviews industrial applications.)

M.V. Buchanan (ed.), *Fourier Transform Mass Spectrometry: Evolution, Innovation, and Applications*, American Chemical Society, Washington, DC, 1987. (Book with 11 contributed chapters on FTICR principles and applications.)

Special Journal Issues on Fourier Transform Ion Cyclotron Resonance

H.L. Pardue (ed.), 'Special Issue: Fourier Transform Mass Spectrometry: Fundamental Aspects and Analytical Applications', *Anal. Chim. Acta*, **178** (1985).

M.B. Comisarow, N.M.M. Nibbering (eds.), 'Special Issue: Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **72** (1986).

C.L. Wilkins (ed.), 'Special Issue: Fourier Transform Mass Spectrometry', *Trends Anal. Chem.*, **13**, 223–251 (1994).

A.G. Marshall (ed.), 'Special Issue: Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **137/138** (1996).

Compilations

K.-P. Wanczek, *Int. J. Mass Spectrom. Ion Proc.*, **60**, 11–60 (1984). (Comprehensive ICR literature review to that date.)

C.L. Wilkins, A.K. Chowdhury, L.M. Nuwaysir, M.L. Coates, *Mass Spectrom. Rev.*, **8**, 67–92 (1989). (Inclusive review of FT-ICR applications from 1986–88.)

T. Dienes, S.J. Pastor, S. Schurch, J.R. Scott, J. Yao, S. Cui, C.L. Wilkins, *Mass Spectrom. Rev.*, **15**, 163–211 (1996). (Comprehensive ICR literature review from 1992 to mid-1996.)

Fourier Transform Techniques

A.G. Marshall, F.R. Verdun, *Fourier Transforms in NMR, Optical, and Mass Spectrometry: a User's Handbook*, Elsevier, Amsterdam, 1990. (Reasonably comprehensive description of analog and digital FT manipulations as applied to spectroscopy.)

Early History of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

M.B. Comisarow, A.G. Marshall, *J. Mass Spectrom.*, **31**, 581–585 (1996). (Review of the first FTICR/MS experiments.)

A.G. Marshall, *Acc. Chem. Res.*, **18**, 316–322 (1985). (Historical review of the early development of FTICR mass spectrometry.)

M.L. Gross, D.L. Rempel, *Science*, **226**, 261–268 (1984). (Review of early FTICR experiments.)

Ion Cyclotron Resonance Excitation

L. Schweikhard, A.G. Marshall, *J. Am. Soc. Mass Spectrom.*, **4**, 433–452 (1993). (Dipolar/quadrupolar, axial/cyclotron/magnetron excitation.)

Ion Cyclotron Resonance Signal Generation and Detection

M.B. Comisarow, *J. Chem. Phys.*, **69**, 4097–4104 (1978). (Fundamental treatment of ICR signal generation in an infinitely extended ICR ion trap.)

P.A. Limbach, P.B. Grosshans, A.G. Marshall, *Anal. Chem.*, **65**, 135–140 (1993). (Determination of number of trapped ions in a cubic trap.)

Ion Cyclotron Resonance Line Shape at Low, Intermediate, and High Pressure

A.G. Marshall, M.B. Comisarow, G. Parisod, *J. Chem. Phys.*, **71**, 4434–4444 (1979). (Fundamental relations between relaxation and FTICR spectral line shape.)

Stored Waveform Inverse Fourier Transform Excitation

S. Guan, A.G. Marshall, *Int. J. Mass Spectrom. Ion Processes*, **157/158**, 5–37 (1996). (Comprehensive review.)

Penning Traps

L.S. Brown, G. Gabrielse, 'Geonium Theory. Physics of a Single Electron or Ion in a Penning Trap', *Rev. Mod. Phys.*, **58**, 233–311 (1986). (Classic reference for the motions of an ion in an ideal Penning trap.)

S. Guan, A.G. Marshall, *Int. J. Mass Spectrom. Ion Processes*, **146/147**, 261–296 (1995). (Comprehensive review.)

V.H. Vartanian, J.S. Anderson, D.A. Laude, *Mass Spectrom. Rev.*, **14**, 1–19 (1995). (Good discussion of open cylindrial traps.)

External Ion Injection

A.G. Marshall, L. Schweikhard, *Int. J. Mass Spectrom. Ion Processes*, **118/119**, 37–70 (1992). (Reviews methods for

guiding ions from an external ion source through the magnetic fringe field to the ICR Penning trap.)

Ion–Molecule Chemistry from Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

B.S. Freiser, *Acc. Chem. Res.*, **27**, 353–360 (1994). (Review of selected topics in organometallic ion chemistry studied by FTICR.)

M.K. Green, C.B. Lebrilla, *Mass Spectrom. Rev.*, **16**, 53–71 (1997). (Review of ion–molecule reactions as probes of gas phase structures of peptides and proteins.)

Quadrupolar Axialization

S. Guan, H.S. Kim, A.G. Marshall, M.C. Wahl, T.D. Wood, X. Xiang, *Chem. Rev.*, **8**, 2161–2182 (1994). (Comprehensive review of theory and applications.)

Two-dimensional Ion Cyclotron Resonance

S. Haebel, T. Gäumann, *Int. J. Mass Spectrom. Ion Proc.*, **144**, 139–166 (1995). (Reviews Hadamard and Fourier transform methods for sorting out parent/product ion connections for all parent ions at once.)

High Mass Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

M.V. Buchanan, R.L. Hettich, *Anal. Chem.*, **65**, 245A–259A (1993). (Review of FTICR mass spectrometry of high-mass biomolecules.)

C.L. Holliman, D.L. Rempel, M.L. Gross, *Mass Spectrom. Rev.*, **13**, 105–132 (1994). (General discussion of problems associated with high m/z ions in FTICR.)

M.W. Senko, F.W. McLafferty, *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 763–785 (1994). (Review of mass spectrometry of large biomolecules.)

Capillary Electrophoresis Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

S.A. Hofstadler, J.C. Severs, R.D. Smith, F.D. Swanek, A.G. Ewing, *J. High Resolut. Chromatogr.*, **19**, 617–621 (1996). (History and 'state-of-the-art' for CE FTICR MS.)

Reviews

Books

M.V. Buchanan (ed.), *Fourier Transform Mass Spectrometry: Evolution, Innovation, and Applications*, American Chemical Society, Washington, DC, 1987.

A.G. Marshall, F.R. Verdun, *Fourier Transforms in NMR, Optical, and Mass Spectrometry: a User's Handbook*, Elsevier, Amsterdam, 1990.

B. Asamoto, R.C. Dunbar, *Analytical Applications of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*, VCH, New York, 1991.

Journal Special Issues

H.L. Pardue, 'Special Issue: Fourier Transform Mass Spectrometry: Fundamental Aspects and Analytical Applications', *Anal. Chim. Acta*, **178** (1985).

M.B. Comisarow, N.M.M. Nibbering (eds.), 'Special Issue: Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **72** (1986).

C.L. Wilkins, 'Special Issue: Fourier Transform Mass Spectrometry', *Trends Anal. Chem.*, **13**, 223–251 (1994).

A.G. Marshall, 'Special Issue: Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **137/138**, (1996).

Review Articles

Early History of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

M.B. Comisarow, 'Fourier Transform Ion Cyclotron Resonance Spectroscopy', in *Fourier, Hadamard, and Hilbert Transforms in Chemistry*, ed. A.G. Marshall, Plenum, New York, 125–146, 1982.

A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Acc. Chem. Res.*, **18**, 316–322 (1985).

M.B. Comisarow, A.G. Marshall, 'The Early Development of Fourier Transform Ion Cyclotron Resonance (FT-ICR) Spectroscopy', *J. Mass Spectrom.*, **31**, 581–585 (1996).

A.G. Marshall, 'Ion Cyclotron Resonance Mass Spectrometry: a Brief History', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, Wiley, Chichester, 486–489, Vol. 1, 1996.

Later Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Developments

K.-P. Wanczek, 'ICR Spectrometry – A Review of New Developments in Theory, Instrumentation, and Applications. I 1983–1986', *Int. J. Mass Spectrom. Ion Proc.*, **95**, 1–38 (1989).

B.S. Freiser, 'Methods for Determining Metal–Ligand and Metal–Metal Bond Energies Using FTMS', in *Bonding Energetics in Organometallic Compounds*, ed. T.J. Marks, American Chemical Society, Washington, DC, 55–69, Vol. 428, 1990.

J.P. Speir, G.S. Gorman, I.J. Amster, 'Laser Desorption, Chemical Ionization, and Laser Desorption/Chemical Ionization Applications with FTMS', in *Mass Spectrometry in the Biological Sciences: a Tutorial*, ed. M.L. Gross, Kluwer, The Netherlands, 199–212, 1992.

A.G. Marshall, L. Schweikhard, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Technique Developments', *Int. J. Mass Spectrom. Ion Processes*, **118/119**, 37–70 (1992).

N.M.M. Nibbering, 'Principles and Applications of FTICR Mass Spectrometry', *Analyst*, **117**, 289–293 (1992).
J.T. Brenna, W.R. Creasy, J.A. Zimmerman, 'Laser Fourier Transform Mass Spectrometry for Polymer Characterization', *Amer. Chem. Soc. Symp. Ser.*, **236**, 129–154 (1993).
M.V. Buchanan, R.L. Hettich, 'Characterization of Large Biomolecules by Fourier Transform Mass Spectrometry', *Anal. Chem.*, **65**, 245A–259A (1993).
S. Guan, H.S. Kim, A.G. Marshall, M.C. Wahl, T.D. Wood, X. Xiang, 'Shrink-wrapping an Ion Cloud for Higher-performance Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Chem. Rev.*, **94**, 2161–2182 (1994).
F.W. McLafferty, 'High-resolution Tandem FT Mass Spectrometry above 10 kDa', *Acc. Chem. Res.*, **27**, 379–386 (1994).
S. Guan, A.G. Marshall, 'Ion Traps for FT-ICR/MS: Principles and Design of Geometric and Electric Configurations', *Int. J. Mass Spectrom. Ion Processes*, **146/147**, 261–296 (1995).
I.J. Amster, 'A Tutorial on Fourier Transform Mass Spectrometry', *J. Mass Spectrom.*, **31**, 1325–1337 (1996).
A.G. Marshall, S. Guan, 'Advantages of High Magnetic Field for FT-ICR Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 1819–1823 (1996).
A.G. Marshall, 'Ion Cyclotron Resonance and Nuclear Magnetic Resonance Spectroscopies: Magnetic Partners for Elucidation of Molecular Structure and Reactivity', *Acc. Chem. Res.*, **9**, 307–316 (1996).
S. Guan, A.G. Marshall, 'Stored Waveform Inverse Fourier Transform (SWIFT) Ion Excitation in Trapped-ion Mass Spectrometry: Theory and Applications', *Int. J. Mass Spectrom. Ion Processes*, **157/158**, 5–37 (1996).
T. Dienes, S.J. Pastor, S. Schürch, J.R. Scott, J. Yao, S. Cui, C.L. Wilkins, 'Fourier Transform Mass Spectrometry–Advancing Years (1992–Mid 1996)', *Mass Spectrom. Rev.*, **15**, 163–211 (1996).
M.K. Green, C.B. Lebrilla, 'Ion-molecule Reactions as Probes of Gas Phase Structures of Peptides and Proteins', *Mass Spectrom. Rev.*, **16**, 53–71 (1997).

Other Reviews Cited in the Above Articles

A.G. Marshall, M.B. Comisarow, 'Fourier and Hadamard Transform Methods in Spectroscopy', *Anal. Chem.*, **47**, 491A (1975).
A.G. Marshall, M.B. Comisarow, 'Fourier Transform Methods in Spectroscopy', *J. Chem. Educ.*, **52**, 638–641 (1975).
A.G. Marshall, M.B. Comisarow, 'Multichannel Methods in Spectroscopy', in *Transform Techniques in Chemistry*, ed. P.R. Griffiths, Plenum, New York, 39–68, 1978.
C.L. Wilkins, 'Fourier Transform Mass Spectrometry', *Anal. Chem.*, **50**, 493A–498A (1978).
C.L. Wilkins, M.L. Gross, 'FTMS for Analysis', *Anal. Chem.*, **53**, 1661A–1668A (1981).
A.G. Marshall, 'Advantages of Transform Methods in Chemistry', in *Fourier, Hadamard, and Hilbert Transforms in Chemistry*, ed. A.G. Marshall, Plenum, New York, 1–43, 1982.
C.L. Johlman, R.L. White, C.L. Wilkins, 'Applications of FTMS', *Mass Spectrom. Rev.*, **2**, 389–415 (1983).
A.G. Marshall, 'Transform Techniques in Chemistry', in *Physical Methods of Modern Chemical Analysis*, ed. T. Kuwana, Academic Press, New York, 57–135, Vol. 3, 1983.
M.L. Gross, D.L. Rempel, 'Fourier Transform Mass Spectrometry', *Science*, **226**, 261–268 (1984).
K.-P. Wanczek, 'ICR Spectrometry – A Review', *Int. J. Mass Spectrom. Ion Proc.*, **60**, 11–60 (1984).
M.B. Comisarow, 'Fundamental Aspects and Applications of Fourier-transform Ion-cyclotron Resonance Spectrometry', *Anal. Chim. Acta*, **178**, 1–15 (1985).
B.S. Freiser, 'Investigation of the Reactions of Metal Ions and their Clusters in the Gas Phase by Laser-ionization FTMS', *Talanta*, **32**, 697–708 (1985).
A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Acc. Chem. Res.*, **18**, 316–322 (1985).
A.G. Marshall, 'Fourier Transform Methods in Spectroscopy', in *IUCCP Symposium*, ed. A.G. Marshall, Texas A&M University Press, College Station, TX, 111–134, 1985.
A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. Potential for Biomedical Applications', in *Mass Spectrometry in the Health and Life Sciences*, ed. A.L. Burlingame, Elsevier, Amsterdam, 265–286, Vol. 24, 1985.
D.A. Laude, Jr, C.L. Johlman, R.S. Brown, D.A. Weil, C.L. Wilkins, 'FTMS: Recent Instrument Developments and Applications', *Mass Spectrom. Rev.*, **5**, 107–166 (1986).
A.G. Marshall, L. Chen, A.T. Hsu, S.L. Mullen, T.L. Ricca, I. Santos, R.E. Shomo, II, T.-C.L. Wang, C.R. Weisenberger, 'Fourier Transform Ion Cyclotron Mass Spectrometry: New Theoretical and Experimental Developments', in *Advances in Mass Spectrometry*, ed. J.F.J. Todd, Wiley, New York, 927–928, Vol. 1985B, 1986.
A.G. Marshall, 'Fourier Transform Mass Spectrometry', in *Spectroscopy in the Biomedical Sciences*, ed. R.M. Gendreau, CRC Press, Boca Raton, FL, 87–105, 1986.
N.M.M. Nibbering, 'Experiments with Ions in a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer', *Comments At. Mol. Phys.*, **18**, 223–234 (1986).
D.H. Russell, 'An Evaluation of Fourier Transform Mass Spectrometry for High Mass Applications', *Mass Spectrom. Rev.*, **5**, 167–189 (1986).
C.S. Hanson, M.E. Castro, D.H. Russell, D.F. Hunt, J. Shabanowitz, 'Fourier Transform Mass Spectrometry of Large ($m/z > 5000$) Biomolecules', in *Fourier Transform Mass Spectrometry: Evolution, Innovation, and Applications*, ed. M.V. Buchanan, American Chemical Society, Washington, DC, 100–115, Vol. 359, 1987.
B. Asamoto, 'FTMS for Industrial Problem Solving', *Spectroscopy*, **3**, 38–46 (1988).
M.P. Chiarelli, M.L. Gross, 'FTMS for Chemical Analysis: a Brief Status Report', in *Analytical Applications of Spectroscopy*, eds. C.S. Creaser, A.M.C. Davies, Royal Society of Chemistry, London, 262–273, 1988.
R.B. Cody, 'An Overview of MS/MS and FTMS', *Analusis*, **16**, 30–36 (1988).

B.S. Freiser, 'Fourier Transform Mass Spectrometry', in *Techniques for the Study of Ion Molecule Reactions*, eds. J.M. Farrar, W.H. Saunders, Jr, Wiley, New York, 61–118, Vol. 20, 1988.

C.D. Hanson, E.L. Kerley, D.H. Russell, 'Recent Developments in Experimental FT/ICR', in *Treatise in Analytical Chemistry*, 2nd edition, ed. J.D. Winefordner, Wiley, New York, 117–187, Vol. 11, 1988.

N.M.M. Nibbering, 'Gas-phase Reactions of Organic Anions', *Adv. Phys. Org. Chem.*, **24**, 1–55 (1988).

B.S. Freiser, 'Applications of FT/ICR Spectrometry to the Study of Transition Metal-containing Ions in the Gas Phase', *Chemtracts – Anal. Phys. Chem.*, **1**, 65–109 (1989).

S. Ghaderi, 'Analysis of Ceramics by Laser Ablation FTMS', *Ceram. Trans.*, **5**, 73–86 (1989).

J.R. Gord, B.S. Freiser, 'Separation of Experiments in Time and Space Using Dual-cell FT/ICR/MS', *Anal. Chim. Acta*, **225**, 11–24 (1989).

A.G. Marshall, 'Analytical Capabilities and Applications of FT/ICR Mass Spectrometry', *Adv. Mass Spectrom.*, **11A**, 651–668 (1989).

N.M.M. Nibbering, 'A View on Some Recent Developments in Mass Spectrometry', *Adv. Mass Spectrom.*, **11A**, 101–125 (1989).

P. Sharpe, D.E. Richardson, 'Applications of FT/ICR/MS in Coordination Chemistry', *Coord. Chem. Rev.*, **93**, 59–85 (1989).

C.L. Wilkins, A.K. Chowdhury, L.M. Nuwaysir, M.L. Coates, 'FTMS: Current Status', *Mass Spectrom. Rev.*, **8**, 67–92 (1989).

D.A. Laude, Jr, J.D. Hogan, 'High Performance FTMS for Chemical Analysis', *Technisches Messen*, **57**, 155–159 (1990).

D.M. Lubman (ed.), *Lasers in Mass Spectrometry*, Oxford University Press, New York, 1990.

A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Spectroscopy*, **5**, 30 (1990).

J.E. Campana, 'Laser Probe Mass Spectrometry', in *Proc. SPIE Applied Spectroscopy in Material Science*, International Society for Optical Engineering, Bellingham, WA, 138–149, 1991.

K. Eller, H. Schwarz, 'Organometallic Chemistry in the Gas Phase', *Chem. Rev.*, **91**, 1121–1177 (1991).

A.G. Marshall, P.B. Grosshans, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: the Teenage Years', *Anal. Chem.*, **63**, 215A–229A (1991).

L.M. Nuwaysir, C.L. Wilkins, 'Matrix-assisted Laser Desorption by FTMS', in *Proc. SPIE Applied Spectroscopy in Material Science*, International Society for Optical Engineering, Bellingham, WA, 112–123, 1991.

R.C. Dunbar, 'Infrared Radiative Cooling of Gas-phase Ions', *Mass Spectrom. Rev.*, **11**, 309–339 (1992).

C.B. Jacoby, C.L. Holliman, M.L. Gross, 'FTMS: Features, Principles, Capabilities, and Limitations', in *Mass Spectrometry in the Biological Sciences: A Tutorial*, ed. M.L. Gross, Kluwer, Dordrecht, 93–116, 1992.

C. Köster, M.S. Kahr, J.A. Castoro, C.L. Wilkins, 'Fourier-transform Mass Spectrometry', *Mass Spectrom. Rev.*, **11**, 495–512 (1992).

L. Schweikhard, G.M. Alber, A.G. Marshall, 'FT/ICR of Highly-Charged Atomic Ions', *Phys. Scr.*, **46**, 598–602 (1992).

L. Schweikhard, A.G. Marshall, 'Excitation Modes for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 433–452 (1993).

C.L. Holliman, D.L. Rempel, M.L. Gross, 'Detection of High Mass-to-charge Ions by FTMS', *Mass Spectrom. Rev.*, **13**, 105–132 (1994).

V.H. Vartanian, J.S. Anderson, D.A. Laude, 'Advances in Trapped Ion Cells for FTICRMS', *Mass Spectrom. Rev.*, **14**, 1–19 (1995).

B.S. Freiser, 'Gas Phase Metal Ion Chemistry', *J. Mass Spectrom.*, **31**, 703–715 (1996).

REFERENCES

1. M.B. Comisarow, A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **25**, 282–283 (1974).
2. M.B. Comisarow, A.G. Marshall, 'Frequency-sweep Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **26**, 489–490 (1974).
3. M.B. Comisarow, A.G. Marshall, 'Selective-phase Ion Cyclotron Resonance Spectroscopy', *Can. J. Chem.*, **52**, 1997–1999 (1974).
4. M.B. Comisarow, A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Spectroscopy Method and Apparatus', *USA Patent* No. 3 937 955, issued 14 February, 1976.
5. A.G. Marshall, P.B. Grosshans, 'Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: the Teenage Years', *Anal. Chem.*, **63**, 215A–229A (1991).
6. W.T. Scott, 'Who was Earnshaw?', *Am. J. Phys.*, **27**, 418–419 (1959).
7. S. Guan, A.G. Marshall, 'Stored Waveform Inverse Fourier Transform (SWIFT) Ion Excitation in Trapped-ion Mass Spectrometry: Theory and Applications', *Int. J. Mass Spectrom. Ion Processes*, **157/158**, 5–37 (1996).
8. L. Schweikhard, A.G. Marshall, 'Excitation Modes for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 433–452 (1993).
9. M.B. Comisarow, 'FT/ICR Spectroscopy', *Adv. Mass Spectrom.*, **7**, 1042–1046 (1978).
10. M.B. Comisarow, 'Signal Modeling for Ion Cyclotron Resonance', *J. Chem. Phys.*, **69**, 4097–4104 (1978).
11. S. Guan, A.G. Marshall, 'Ion Traps for FT-ICR/MS: Principles and Design of Geometric and Electric Configurations', *Int. J. Mass Spectrom. Ion Processes*, **146/147**, 261–296 (1995).
12. P.A. Limbach, P.B. Grosshans, A.G. Marshall, 'Experimental Determination of the Number of Trapped Ions,

Detection Limit, and Dynamic Range in FT/ICR/MS', *Anal. Chem.*, **65**, 135–140 (1993).
13. P.A. Limbach, A.G. Marshall, M. Wang, 'An Electrostatic Ion Guide for Efficient Transmission of Low Energy Externally Formed Ions into a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer', *Int. J. Mass Spectrom. Ion Proc.*, **125**, 135–143 (1993).
14. A.G. Marshall, F.R. Verdun, *Fourier Transforms in NMR, Optical, and Mass Spectrometry: a User's Handbook*, Elsevier, Amsterdam, 1990.
15. A.G. Marshall, D.C. Roe, 'Theory of Fourier Transform Ion Cyclotron Resonance Mass Spectroscopy: Response to Frequency-sweep Excitation', *J. Chem. Phys.*, **73**, 1581–1590 (1980).
16. A.G. Marshall, T.L. Ricca, T.-C.L. Wang, 'Tailored Excitation for Trapped Ion Mass Spectrometry', *USA Patent* No. 4 761 545, issued 2 August, 1988.
17. A.G. Marshall, T.-C.L. Wang, T.L. Ricca, 'Tailored Excitation for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Chem. Soc.*, **107**, 7893–7897 (1985).
18. T.-C.L. Wang, T.L. Ricca, A.G. Marshall, 'Extension of Dynamic Range in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry via Stored Waveform Inverse Fourier Transform Excitation', *Anal. Chem.*, **58**, 2935–2938 (1986).
19. S. Guan, A.G. Marshall, S.E. Scheppele, 'Resolution and Chemical Formula Identification of Aromatic Hydrocarbons Containing Sulfur, Nitrogen, and/or Oxygen in Crude Oil Distillates', *Anal. Chem.*, **68**, 46–71 (1996).
20. P.B. Grosshans, A.G. Marshall, 'Theory of Ion Cyclotron Resonance Mass Spectrometry: Resonant Excitation and Radial Ejection in Orthorhombic and Cylindrical Ion Traps', *Int. J. Mass Spectrom. Ion Proc.*, **100**, 347–379 (1990).
21. J.L. Beauchamp, 'Theory of Collision-broadened ICR Spectra', *J. Chem. Phys.*, **46**, 1231–1243 (1967).
22. D. Wobschall, J.R. Graham, Jr, D.P. Malone, 'ICR and the Determination of Collision Cross Sections', *Phys. Rev.*, **131**, 1565–1571 (1963).
23. J.I. Steinfeld, J.S. Francisco, W.L. Hase, *Chemical Kinetics and Dynamics*, Prentice Hall, Englewood Cliffs, NJ, 1989.
24. R.C. Weast, *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 1989–1990.
25. P. Langevin, 'Une Formule Fondamentale de Theorie Cinetique', *Ann. Chim. Phys.*, **5**, 245 (1905).
26. S. Guan, G.-Z. Li, A.G. Marshall, 'Effect of Ion-neutral Collision Mechanism on Trapped-ion Equation of Motion: a New Mass Spectral Line Shape for High-mass Trapped Ions', *Int. J. Mass Spectrom. Ion Processes*, **167/168**, 185–194 (1998).
27. H. Dehmelt, 'Ion Traps', *Rev. Mod. Phys.*, **62**, 525–530 (1990).
28. M.B. Comisarow, 'Cubic Trapped-ion Cell for Ion Cyclotron Resonance', *Int. J. Mass Spectrom. Ion Phys.*, **37**, 251–257 (1981).
29. M.B. Comisarow, 'Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Adv. Mass Spectrom.*, **8**, 1698–1706 (1980).
30. J.L. Elkind, F.D. Weiss, J.M. Alford, R.T. Laaksonen, R.E. Smalley, 'Fourier Transform Ion Cyclotron Resonance Studies of H_2 Chemisorption on Niobium Cluster Cations', *J. Chem. Phys.*, **88**, 5215–5224 (1988).
31. P. Kofel, M. Allemann, H. Kellerhals, K.-P. Wanczek, 'Coupling of Axial and Radial Motions in ICR Cells During Excitation', *Int. J. Mass Spectrom. Ion Proc.*, **74**, 1–12 (1986).
32. P. Kofel, M. Allemann, H. Kellerhals, K.-P. Wanczek, 'Time-of-flight ICR Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **72**, 53–61 (1986).
33. S.H. Lee, K.-P. Wanczek, H. Hartmann, 'A New Cylindrical Trapped Ion ICR Cell', *Adv. Mass Spectrom.*, **8B**, 1645–1649 (1980).
34. P. Caravatti, M. Allemann, 'RF Shim by Trap Segmentation', *Org. Mass Spectrom.*, **26**, 514–518 (1991).
35. S.C. Beu, D.A. Laude, Jr, 'Elimination of Axial Ejection during Excitation with a Capacitively Coupled Open Trapped-ion Cell for FTICRMS', *Anal. Chem.*, **64**, 177–180 (1992).
36. S.C. Beu, D.A. Laude, Jr, 'Open Trapped Ion Cell Geometries for FT/ICR/MS', *Int. J. Mass Spectrom. Ion Proc.*, **112**, 215–230 (1992).
37. G. Gabrielse, L. Haarsma, S.L. Rolston, 'Open-endcap Penning Traps for High Precision Experiments', *Int. J. Mass Spectrom. Ion Processes*, **88**, 319–332 (1989).
38. D.P. Littlejohn, S. Ghaderi, 'Mass Spectrometer and Method', *USA Patent* No. 4 581 533, issued 8 April, 1986.
39. D.L. Rempel, S.K. Huang, M.L. Gross, 'Relation of Signal Sensitivity and Ion z-motion in Cubic Cells. Theory and Implication for Ion Kinetic Studies', *Int. J. Mass Spectrom. Ion Processes*, **70**, 163–184 (1986).
40. D.W. Mitchell, A.L. Rockwood, R. Chen, M.G. Sherman, R.D. Smith, 'Theoretical Investigations of Frequency Shifts Caused by Electrostatic Trapping Field Inhomogeneities in Ion Cyclotron Resonance Mass Spectrometry', in *Proceedings 42nd American Society Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, American Society Mass Spectrometry, Chicago, IL, 729, 1994.
41. C.L. Hendrickson, J.J. Drader, D.A. Laude, Jr, 'Simplified Application of Quadrupolar Excitation in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **6**, 448–452 (1995).
42. R.D. Knight, 'The General Form of the Quadrupole Ion Trap Potential', *Int. J. Mass Spectrom. Ion Proc.*, **51**, 127–131 (1983).
43. W.W. Yin, M. Wang, A.G. Marshall, E.B. Ledford, Jr, 'Experimental Evaluation of a Hyperbolic Ion Trap for

FT/ICR/MS', *J. Am. Soc. Mass Spectrom.*, **3**, 188–197 (1992).
44. G. Jackson, J.D. Canterbury, S. Guan, A.G. Marshall, 'Linearity and Quadrupolarity of Tetragonal and Cylindrical Penning Traps of Arbitrary Length-to-width Ratio', *J. Am. Soc. Mass Spectrom.*, **8**, 283–293 (1997).
45. G.S. Jackson, C.L. Hendrickson, B.B. Reinhold, A.G. Marshall, 'Two-plate vs. Four-plate Azimuthal Quadrupolar Excitation for FT-ICR Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **165/166**, 327–338 (1997).
46. L. Schweikhard, J. Ziegler, H. Bopp, K. Lützenkirchen, 'The Trapping Condition and a New Instability of the Ion Motion in the Ion Cyclotron Resonance Trap', *Int. J. Mass Spectrom. Ion Processes*, **141**, 77–90 (1995).
47. M. Allemann, H.-P. Kellerhals, K.-P. Wanczek, 'Sidebands in the ICR Spectrum and their Application for Exact Mass Determination', *Chem. Phys. Lett.*, **84**, 547–551 (1981).
48. D. Mitchell, S. Delong, D. Cherniak, M. Harrison, 'z-Axis Oscillation Sidebands in FT/ICR Mass Spectra', *Int. J. Mass Spectrom. Ion Processes*, **91**, 273–282 (1989).
49. E.B. Ledford, Jr, D.L. Rempel, M.L. Gross, 'Space Charge Effects in Fourier Transform Mass Spectrometry. Mass Calibration', *Anal. Chem.*, **56**, 2744–2748 (1984).
50. G. Bollen, R.B. Moore, G. Savard, H. Stolzenberg, 'Quadrupole Excitation for Magnetron-to-cyclotron Conversion', *Appl. Phys.*, **68**, 4355–4374 (1990).
51. S. Guan, 'Generation of Optimal Excitation Pulses for Two Energy Level Systems Using an Inverse FT Method', *J. Chem. Phys.*, **96**, 7959–7964 (1992).
52. S. Guan, H.S. Kim, A.G. Marshall, M.C. Wahl, T.D. Wood, X. Xiang, 'Shrink-wrapping an Ion Cloud for Higher-performance Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Chem. Rev.*, **94**, 2161–2182 (1994).
53. J.P. Speir, G.S. Gorman, C.C. Pitsenberger, C.A. Turner, P.P. Wang, I.J. Amster, 'Remeasurement of Ions Using Quadrupolar Excitation Fourier Transform Ion Cyclotron Resonance Spectrometry', *Anal. Chem.*, **65**, 1746–1752 (1993).
54. E.R. Williams, K.D. Henry, F.W. McLafferty, 'Remeasurement of Heavy Ions', *J. Am. Chem. Soc.*, **112**, 6157–6162 (1990).
55. E.R. Williams, K.D. Henry, F.W. McLafferty, J. Shabanowitz, D.F. Hunt, 'Surface-induced Dissociation of Peptide Ions in FT/MS', *J. Am. Soc. Mass Spectrom.*, **1**, 413–416 (1990).
56. E.R. Williams, S.Y. Loh, F.W. McLafferty, R.B. Cody, 'Hadamard Transform Measurement of Tandem FT Mass Spectra', *Anal. Chem.*, **62**, 698–703 (1990).
57. T. Solouki, J.A. Marto, F.M. White, S. Guan, A.G. Marshall, 'Attomole Biomolecule Mass Analysis by Matrix-assisted Laser Desorption/Ionization Fourier Transform Ion Cyclotron Resonance', *Anal. Chem.*, **67**, 4139–4144 (1995).
58. T. Solouki, L. Pasa-Tolic, G.S. Jackson, S. Guan, A.G. Marshall, 'High-resolution Multistage MS, MS^2, MS^3 Matrix-assisted Laser Desorption/Ionization FT-ICR Mass Spectra of Peptides from a Single Laser Shot', *Anal. Chem.*, **68**, 3718–3725 (1996).
59. S. Guan, A.G. Marshall, 'Two-way Conversation with a Mass Spectrometer: Non-destructive Interactive Mass Spectrometry', *Anal. Chem.*, **69**, 1–4 (1997).
60. A.G. Marshall, M.B. Comisarow, G. Parisod, 'Relaxation and Spectral Line Shape in Fourier Transform Ion Cyclotron Resonance Spectroscopy', *J. Chem. Phys.*, **71**, 4434–4444 (1979).
61. D.J. Wineland, H.G. Dehmelt, 'Line Shifts and Widths of Axial, Cyclotron and G-2 Resonances in Tailored, Stored Electron (Ion) Cloud', *Int. J. Mass Spectrom. Ion Phys.*, **16**, 338 (1975).
62. L. Pasa-Tolic, Y. Huang, S. Guan, H.S. Kim, A.G. Marshall, 'Ultrahigh-resolution Matrix-assisted Laser Desorption/Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Peptides', *J. Mass Spectrom.*, **30**, 825–833 (1995).
63. D.W. Mitchell, R.D. Smith, 'Cyclotron Motion of Two Coulombically Interacting Ion Clouds with Implications to FT-ICR Mass Spectrometry', *Phys. Rev. E*, **52**, 4366–4386 (1995).
64. D.W. Mitchell, R.D. Smith, 'Prediction of a Space Charge Induced Upper Molecular Mass Limit Towards Achieving Unit Mass Resolution in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Mass Spectrom.*, **31**, 771–790 (1996).
65. B. Asamoto, R.C. Dunbar, *Analytical Applications of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*, VCH, New York, 1991.
66. M.P. Chiarelli, M.L. Gross, 'FTMS for Chemical Analysis: a Brief Status Report', in *Analytical Applications of Spectroscopy*, eds. C.S. Creaser, A.M.C. Davies, Royal Society of Chemistry, London, 262–273, 1988.
67. K. Eller, H. Schwarz, 'Organometallic Chemistry in the Gas Phase', *Chem. Rev.*, **91**, 1121–1177 (1991).
68. B.S. Freiser, 'Fourier Transform Mass Spectrometry', in *Techniques for the Study of Ion Molecule Reactions*, eds. J.M. Farrar, W.H. Saunders, Jr, Wiley, New York, 61–118, Vol. 20, 1988.
69. B.S. Freiser, 'Gas Phase Metal Ion Chemistry', *J. Mass Spectrom.*, **31**, 703–715 (1996).
70. N.M.M. Nibbering, 'Gas-phase Ion/Molecule Reactions as Studied by FT/ICR', *Acc. Chem. Res.*, **23**, 279–285 (1990).
71. D.M. Lubman (ed.), *Lasers in Mass Spectrometry*, Oxford University Press, New York, 1990.
72. J.A. Castoro, C. Köster, C. Wilkins, 'Matrix-assisted Laser Desorption/Ionization of High-mass Molecules by Fourier-transform Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 239–241 (1992).

73. J.A. Castoro, L.M. Nuwaysir, C.F. Ijames, C.L. Wilkins, 'Comparative Study of Photodissociation and Surface-induced Dissociation by Laser Desorption Mass Spectrometry', *Anal. Chem.*, **64**, 2238–2243 (1992).
74. R.L. Hettich, M.V. Buchanan, 'Matrix-assisted Laser Desorption Fourier Transform Mass Spectrometry for the Structural Examination of Modified Nucleic Acid Constituents', *Int. J. Mass Spectrom. Ion Proc.*, **111**, 365–380 (1991).
75. I.J. Amster, J.A. Loo, J.J.P. Furlong, F.W. McLafferty, 'Cesium Ion Desorption Ionization with Fourier Transform Mass Spectrometry', *Anal. Chem.*, **59**, 313–317 (1987).
76. M.E. Castro, D.H. Russell, 'Cesium Ion Desorption Ionization with Fourier Transform Mass Spectrometry', *Anal. Chem.*, **56**, 578–581 (1984).
77. N.C. Hill, P.A. Limbach, R.E. Shomo, II, A.G. Marshall, A.D. Appelhans, J.E. Delmore, 'Fast Neutral Beam Ion Source Coupled to a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer', *Rev. Sci. Instrum.*, **62**, 2612–2617 (1991).
78. B.H. Wang, I.J. Amster, F.W. McLafferty, I.B. Brown, 'Metal-vapor Vacuum Arc as a Primary Ion Source for Secondary Ion Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **100**, 51–61 (1990).
79. E.R. Williams, F.W. McLafferty, 'Plasma Desorption FT/ICR/MS', *J. Am. Soc. Mass Spectrom.*, **1**, 427–430 (1990).
80. K.D. Henry, F.W. McLafferty, 'Electrospray Ionization with Fourier-transform Mass Spectrometry. Charge State Assignment from Resolved Isotopic Peaks', *Org. Mass Spectrom.*, **25**, 490–492 (1990).
81. K.D. Henry, E.R. Williams, B.-H. Wang, F.W. McLafferty, J. Shabanowitz, D.F. Hunt, 'Fourier-transform Mass Spectrometry of Large Molecules by Electrospray Ionization', *Proc. Natl. Acad. Sci. USA*, **86**, 9075–9078 (1989).
82. J.M. Alford, P.E. Williams, D.J. Trevor, R.E. Smalley, 'Metal Cluster ICR. Combining Supersonic Metal Cluster Beam Technology with FT-ICR', *Int. J. Mass Spectrom. Ion Proc.*, **72**, 33–51 (1986).
83. C.E.C.A. Hop, T.B. McMahon, G.D. Willett, 'Determination of Bond Dissociation Energies via Energy-resolved CID in an FT/ICR Spectrometer', *Int. J. Mass Spectrom. Ion Proc.*, **101**, 191–208 (1990).
84. P. Kofel, T.B. McMahon, 'A High Pressure External Ion Source for FT/ICR Spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **98**, 1–24 (1990).
85. S.A. Hofstadler, D.A. Laude, Jr, 'Trapping and Detection of Ions Generated in a High Magnetic Field Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **3**, 615–623 (1992).
86. R.T. McIver, Jr, R.L. Hunter, W.D. Bowers, 'Coupling a Quadrupole Mass Spectrometer and a Fourier Transform Mass Spectrometer', *Int. J. Mass Spectrom. Ion Processes*, **64**, 67–77 (1985).
87. J.A. Marto, S. Guan, A.G. Marshall, 'Wide Mass Range Axialization for High-resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Externally Generated Ions', *Rapid Commun. Mass Spectrom.*, **8**, 615–620 (1994).
88. R.T. McIver, Jr, Y. Li, R.L. Hunter, 'FTMS Method for High Resolution Matrix-assisted Laser Desorption', *Int. J. Mass Spectrom. Ion Processes*, **132**, L1–L7 (1994).
89. G.S. Larsen, J. Wronka, D.P. Ridge, 'An Ion Cyclotron Resonance Spectrometer as a Gas Chromatographic Detector. The Effect of Continuous Trapping on Performance', *Int. J. Mass Spectrom. Ion Processes*, **72**, 73–84 (1986).
90. M.W. Senko, C.L. Hendrickson, M.R. Emmett, S.D.-H. Shi, A.G. Marshall, 'External Accumulation of Ions for Enhanced Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **8**, 970–976 (1997).
91. C.C. Stacey, G.H. Kruppa, C.H. Watson, J. Wronka, F.H. Laukien, J.F. Banks, C.M. Whitehouse, 'Reverse-phase Liquid Chromatography/Electrospray-ionization Fourier-transform Mass Spectrometry in the Analysis of Peptides', *Rapid Commun. Mass Spectrom.*, **8**, 513–516 (1994).
92. G.W. Stockton, J.T. Meek, W.G. Millen, R.S. Wayne, 'An External Ion Source FT-ICR Mass Spectrometer for Ultrahigh-mass Accuracy Measurements and High-performance Liquid Chromatographic Sample Introduction', in *FT-ICR/MS: Analytical Applications of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*, ed. B. Asamoto, VCH, New York, 235–272, 1991.
93. S.A. Hofstadler, J.H. Wahl, J.E. Bruce, R.D. Smith, 'On-line Capillary Electrophoresis with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Chem. Soc.*, **115**, 6983–6984 (1993).
94. S.A. Hofstadler, J.C. Severs, R.D. Smith, F.D. Swanek, A.G. Ewing, 'Analysis of Single Cells with Capillary Electrophoresis Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 919–922 (1996).
95. R.B. Cody, R.C. Burnier, B.S. Freiser, 'Collision-induced Dissociation with Fourier Transform Mass Spectrometry', *Anal. Chem.*, **54**, 96–101 (1982).
96. C.F. Ijames, C.L. Wilkins, 'Surface-induced Dissociation by FT/ICR/MS', *Anal. Chem.*, **62**, 1295–1299 (1990).
97. W.D. Bowers, S.-S. Delbert, R.L. Hunter, R.T. McIver, Jr, 'Fragmentation of Oligopeptide Ions using Ultraviolet Laser Radiation and Fourier Transform Mass Spectrometry', *J. Am. Chem. Soc.*, **106**, 7288–7289 (1984).
98. D.P. Little, J.P. Speir, M.W. Senko, P.B. O'Cornnor, F.W. McLafferty, 'Infrared Multiphoton Dissociation of Large Multiply-charged Ions for Biomolecule Sequencing', *Anal. Chem.*, **66**, 2809–2815 (1994).

99. D.M. Peiris, M.A. Cheeseman, R. Ramanathan, J.R. Eyler, 'Infrared Multiple Photon Dissociation Spectra of Gaseous Ions', *J. Phys. Chem.*, **97**, 7839–7843 (1993).
100. R.L. Woodlin, D.S. Bomse, J.L. Beauchamp, 'Multiphoton Dissociation of Molecules with Low Power Continuous Wave Infrared Laser Radiation', *J. Am. Chem. Soc.*, **100**, 3248–3250 (1978).
101. R.C. Dunbar, T.B. McMahon, 'Activation of Unimolecular Reactions by Ambient Blackbody Radiation', *Science*, **279**, 194–197 (1998).
102. J.W. Gauthier, T.R. Trautman, D.B. Jacobson, 'Sustained Off-resonance Irradiation for CAD Involving FTMS. CAD Technique that Emulates Infrared Multiphoton Dissociation', *Anal. Chim. Acta*, **246**, 211–225 (1991).
103. K.A. Boering, J. Rolfe, J.I. Brauman, 'Control of Ion Kinetic Energy in Ion Cyclotron Resonance Spectrometry: Very-low-energy Collision-induced Dissociation', *Rapid Commun. Mass Spectrom.*, **6**, 303–305 (1992).
104. S.A. Lee, C.Q. Jiao, Y. Huang, B.S. Freiser, 'Multiple Excitation Collisional Activation in Fourier-transform Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 819–821 (1993).
105. W.D. Price, P.D. Schnier, R.A. Jockusch, E.F. Strittmatter, E.R. Williams, 'Unimolecular Reaction Kinetics in the High-pressure Limit without Collisions', *J. Am. Chem. Soc.*, **118**, 10 640–10 644 (1996).
106. W.D. Price, P.D. Schnier, E.R. Williams, 'Tandem Mass Spectrometry of Large Biomolecule Ions by Blackbody Infrared Radiative Dissociation', *Anal. Chem.*, **68**, 859–866 (1996).
107. R.A. Chorush, D.P. Little, S.C. Beu, T.D. Wood, F.W. McLafferty, 'Surface-induced Dissociation of Multiply-protonated Proteins', *Anal. Chem.*, **67**, 1042–1046 (1995).
108. S. Haebel, M.-E. Walser, T. Gäumann, 'High Front-end Resolution Collision-induced Dissociation in FT-ICR Mass Spectrometry', *Int. J. Mass Spectrom. Ion Proc.*, **151**, 97–115 (1995).
109. P. Pfändler, G. Bodenhausen, J. Rapin, M.-E. Walser, T. Gäumann, '2D FT/ICR/MS', *J. Am. Chem. Soc.*, **110**, 5625–5628 (1988).
110. C.W. Ross, III, S. Guan, P.B. Grosshans, T.L. Ricca, A.G. Marshall, 'Two-dimensional Fourier Transform Ion Cyclotron Resonance Mass Spectrometry/Mass Spectrometry with Stored-waveform Ion Radius Modulation', *J. Am. Chem. Soc.*, **115**, 7854–7861 (1993).
111. A.G. Marshall, S. Guan, 'Advantages of High Magnetic Field for FT-ICR Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 1819–1823 (1996).
112. M.W. Senko, C.L. Hendrickson, L. Pasa-Tolic, J.A. Marto, F.M. White, S. Guan, A.G. Marshall, 'Electrospray Ionization FT-ICR Mass Spectrometry at 9.4 Tesla', *Rapid Commun. Mass Spectrom.*, **10**, 1824–1828 (1996).
113. A.G. Marshall, 'Ion Cyclotron Resonance and Nuclear Magnetic Resonance Spectroscopies: Magnetic Partners for Elucidation of Molecular Structure and Reactivity', *Acc. Chem. Res.*, **9**, 307–316 (1996).
114. M. Wang, A.G. Marshall, 'High-resolution Multiple-ion Simultaneous Monitoring by Means of Multiple-foldover Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **60**, 341–344 (1988).
115. M.B. Comisarow, J. Melka, 'Error Estimates for Finite Zero-filling in Fourier Transform Spectroscopy', *Anal. Chem.*, **51**, 2198–2203 (1979).
116. A.G. Marshall, 'Convolution Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **63**, 515–518 (1979).
117. Z. Zhang, S. Guan, A.G. Marshall, 'Enhancement of the Effective Resolution of Mass Spectra of High-mass Biomolecules by Maximum-entropy Based Deconvolution to Eliminate the Isotopic Natural Abundance Distribution', *J. Am. Soc. Mass Spectrom.*, **8**, 659–670 (1997).
118. G.M. Alber, A.G. Marshall, 'Effect of Sampling Rate on Fourier Transform Spectra: Oversampling is Overrated', *Appl. Spectrosc.*, **44**, 1111–1116 (1990).
119. C.P. Williams, A.G. Marshall, 'Hartley Transform Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **61**, 428–431 (1989).
120. C.P. Williams, A.G. Marshall, 'Hartley/Hilbert Transform Spectroscopy: Absorption-mode Resolution with Magnitude-mode Precision', *Anal. Chem.*, **64**, 916–923 (1992).
121. J.E. Meier, A.G. Marshall, 'Bayesian Versus Fourier Spectral Analysis of Ion Cyclotron Resonance Time-domain Signals', *Anal. Chem.*, **62**, 201–208 (1990).
122. J.E. Meier, A.G. Marshall, 'Pure Absorption Mode Spectra from Bayesian Maximum Entropy Analysis', *Anal. Chem.*, **63**, 551–560 (1991).
123. S. Guan, A.G. Marshall, 'Linear Prediction Cholesky Decomposition vs. Fourier Transform Spectral Analysis for Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **69**, 1156–1162 (1997).
124. J.F. Loo, M.D. Krahling, T.C. Farrar, 'Accurate Ion Abundance Measurements in ICR/MS by Linear Prediction', *Rapid Commun. Mass Spectrom.*, **4**, 297–299 (1990).
125. E.C. Craig, I. Santos, A.G. Marshall, 'Dispersion vs. Absorption (DISPA) Method for Automatic Phase Correction of Fourier Transform Ion Cyclotron Resonance Mass Spectra', *Rapid Commun. Mass Spectrom.*, **1**, 33–37 (1987).
126. A. Serreqi, M.B. Comisarow, 'Frequency Interpolation of Discrete, Apodized Magnitude Line Shapes', *Appl. Spectrosc.*, **41**, 288–295 (1987).
127. H.S. Kim, A.G. Marshall, 'Magnitude-mode Multiple-derivative Spectra for Resolution Enhancement without Loss in Signal-to-noise Ratio in Fourier Transform Mass Spectrometry', *J. Mass Spectrom.*, **30**, 1237–1244 (1995).

128. A.T. Hsu, T.L. Ricca, A.G. Marshall, 'Clipped Representations of Fourier Transform Ion Cyclotron Resonance Mass Spectra', *Anal. Chim. Acta*, **178**, 27–41 (1985).
129. J.N. Louris, J.S. Brodbelt-Lustig, R.G. Cooks, G.L. Glish, G.J. van Berkel, S.A. McLuckey, 'Ion Isolation and Sequential Stages of Mass Spectrometry in a Quadrupole Ion Trap Mass Spectrometer', *Int. J. Mass Spectrom. Ion Proc.*, **96**, 117–137 (1990).
130. D.E. Goeringer, W.B. Whitten, J.M. Ramsey, S.A. McLuckey, G.L. Glish, 'Theory of High-resolution Mass Spectrometry Achieved via Resonance Ejection in the Quadrupole Ion Trap', *Anal. Chem.*, **64**, 1434–1439 (1992).
131. S.A. McLuckey, D.E. Goeringer, G.L. Glish, 'Selective Ion Isolation/Ejection Over a Broad Mass Range in the Quadrupole Ion Trap', *J. Am. Soc. Mass Spectrom.*, **2**, 11–21 (1991).
132. S. Guan, A.G. Marshall, 'Stored Waveform Inverse Fourier Transform (SWIFT) Axial Excitation/Ejection for Quadrupole Ion Trap Mass Spectrometry', *Anal. Chem.*, **65**, 1288–1294 (1993).
133. R.K. Julian, Jr, R.G. Cooks, 'Broad Band Excitation in the Quadrupole Ion Trap Mass Spectrometer Using Shaped Pulses with the Inverse Fourier Transform', *Anal. Chem.*, **65**, 1827–1833 (1993).
134. M.H. Soni, R.G. Cooks, 'Selective Injection and Isolation of Ions in Quadrupole Ion Trap Mass Spectrometry Using Notched Waveforms Created Using the Inverse Fourier Transform', *Anal. Chem.*, **66**, 2488–2496 (1994).
135. M. Soni, V. Frankevich, M. Nappi, R.E. Santini, J.W. Amy, R.G. Cooks, 'Broad-band Fourier Transform Quadrupole Ion Trap Mass Spectrometry', *Anal. Chem.*, **68**, 3314–3320 (1996).
136. J.E.P. Syka, W.J. Fies, Jr, 'Fourier Transform Quadrupole Mass Spectrometer and Method', *USA Patent* No. 4 755 670, Issued, 1988.
137. R.A. Zubarev, P. Hakansson, B. Sundqvist, 'Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements', *Anal. Chem.*, **68**, 4060–4063 (1996).
138. G.A. Valaskovic, N.K. Kelleher, F.W. McLafferty, 'Attomole Protein Characterization by Capillary Electrophoresis-mass Spectrometry', *Science*, **273**, 1199–1202 (1996).
139. M.R. Emmett, F.M. White, C.L. Hendrickson, S.D.-H. Shi, A.G. Marshall, 'Application of Micro-electrospray Liquid Chromatography Techniques to FT-ICR MS to Enable High Sensitivity Biological Analysis', *J. Am. Soc. Mass Spectrom.*, **9**, 333–340 (1998).
140. M.R. Emmett, R.M. Caprioli, 'Micro-electrospray Mass Spectrometry: Ultra-high-sensitivity Analysis of Peptides and Proteins', *J. Am. Soc. Mass Spectrom.*, **5**, 605–613 (1994).
141. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, 'Electrospray Ionization for Mass Spectrometry of Large Biomolecules', *Science*, **246**, 64–71 (1989).
142. J.P. Speir, M.W. Senko, D.P. Little, J.A. Loo, F.W. McLafferty, 'High-resolution Tandem Mass Spectra of 37–67 kDa Proteins', *J. Mass Spectrom.*, **30**, 39–42 (1995).
143. T.D. Wood, L.H. Chen, N.L. Kelleher, D.P. Little, G.L. Kenyon, F.W. McLafferty, 'Direct Sequence Data from Heterogeneous Creatine Kinase (43 kDa) by High-resolution Tandem Mass Spectrometry', *Biochemistry*, **34**, 16 251–16 254 (1995).
144. N.L. Kelleher, M.W. Senko, M.M. Siegel, F.W. McLafferty, 'Unit Resolution Mass Spectra of 112 kDa Molecules with 3 Da Accuracy', *J. Am. Soc. Mass Spectrom.*, **8**, 380–383 (1997).
145. F.W. McLafferty, 'High-resolution Tandem FT Mass Spectrometry above 10 kDa', *Acc. Chem. Res.*, **27**, 379–386 (1994).
146. R. Chen, X. Cheng, D.W. Mitchell, S.A. Hofstadler, Q. Wu, A.L. Rockwood, M.G. Sherman, R.G. Smith, 'Trapping, Detection, and Mass Determination of Coliphage T4 DNA Ions of 10^8 Da by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Anal. Chem.*, **67**, 1159–1163 (1995).
147. M.W. Senko, S.C. Beu, F.W. McLafferty, 'Automated Assignment of Charge States from Resolved Isotopic Peaks for Multiply Charged Ions', *J. Am. Soc. Mass Spectrom.*, **6**, 52–56 (1995).
148. M.W. Senko, S.C. Beu, F.W. McLafferty, 'Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions', *J. Am. Soc. Mass Spectrom.*, **6**, 229–233 (1995).
149. A.G. Marshall, M.W. Senko, W. Li, M. Li, S. Dillon, S. Guan, T.M. Logan, 'Protein Molecular Weight to 1 Da by ^{13}C, ^{15}N Double-depletion and FT-ICR Mass Spectrometry', *J. Am. Chem. Soc.*, **119**, 433–434 (1997).
150. E.R. Cohen, B.N. Taylor, 'The Fundamental Physical Constants', *Physics Today*, **49**, BG9–BG13 (1996).
151. G. Audi, A.H. Wapstra, 'The 1995 Update to the Atomic Mass Evaluation', *Nucl. Phys. A*, **595**, 409–425 (1995).

Gas Chromatography/Mass Spectrometry

Peter T. Palmer
San Francisco State University, San Francisco, USA

1 Introduction **1**
2 Gas Chromatography/Mass Spectrometry Instrumentation **2**
2.1 Sample Introduction Systems 3
2.2 Gas Chromatographs 4
2.3 Mass Analyzers 6
3 Data Analysis **10**
3.1 Identification 10
3.2 Quantitation 11
3.3 Pattern Recognition, Neural Nets, and Artificial Intelligence Methods 11
4 Practical Considerations **11**
4.1 Gas Chromatography 12
4.2 Mass Spectrometry 12
5 Perspective and Future Developments **13**
Acknowledgments **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **15**

Gas chromatography/mass spectrometry (GC/MS) is perhaps the most widely used analytical technique for the analysis of complex mixtures. The power of this technique stems from the ''marriage'' of two powerful techniques into a hyphenated method. Gas chromatography (GC) is used to separate the components of a mixture based on their differing affinities for the stationary phase on the column. Mass spectrometry (MS) serves not only to detect and quantify the various species eluting off the column but also provides definitive information as to their identity. The most common GC/MS instruments employ a gas chromatograph with a split/splitless injector, capillary column, and either a quadrupole or quadrupole ion trap as a mass analyzer. A high degree of automation and a powerful data system are required for handling the large amounts of three-dimensional (3-D) data (time, mass-to-charge ratio, and intensity) that can be generated. The technique typically requires that the analytes in question have sufficient volatility to be amenable to GC. It is very sensitive, with detection limits typically in the picogram or even sub-picogram range. It can be highly selective and allows for incredible flexibility with respect to the type and quality of the mass spectral data; the user may choose from a number of ionization modes and scan modes, and control mass range and mass resolution, and many other experimental parameters. Analysis speeds are relatively slow due to the time required for adequate chromatographic separation, but can be reduced to an order of a few minutes using fast GC techniques. GC/MS instrumentation is usually intended for benchtop operation in the laboratory, although a number of field-portable instruments have been developed. The technique requires significant user expertise, knowledge of various means for compound identification and quantitation, and instrument tuning, maintenance, and upkeep.

1 INTRODUCTION

GC/MS is perhaps the most successful of all the so-called hyphenated techniques.[1] GC separates compounds on the basis of their differing affinities for the stationary phase. MS records the separation process and three dimensions of information: time, mass-to-charge ratio (m/z) and intensity. Identification of individual compounds is usually achieved through the combination of a retention time match with a known standard and a match of the experimental electron ionization (EI) mass spectrum against a library spectrum. Quantitation is usually performed by ratioing the response of a selected quantitation or ''quan'' ion in the sample to that of a standard of known concentration. The combination of GC with MS represents the most powerful and widely used technique for the analysis of complex mixtures. Its primary applications include the analysis of drugs, explosives, pesticides, petroleum products, and pollutants in air, water and soil; and its uses encompass such diverse areas as clinical chemistry, environmental science, food and beverage analysis, forensic science, industrial hygiene, and process monitoring.

GC/MS necessitates that the target compounds of interest be amenable to GC analysis (i.e. relatively low molecular weight, nonlabile, thermally stable, reasonably volatile or derivatizable). The sensitivities of MS detectors are either comparable with or better than many other GC detectors, with the notable exceptions being the electron capture detector (ECD) and other highly selective GC detectors. Selectivity in GC/MS analyses is outstanding and in many cases can provide conclusive evidence for the presence of a compound due to the combination of the orthogonal information from GC and the structural information from MS. Selectivity and sensitivity can be further enhanced through the use of selective chemical ionization (CI) techniques and/or selected ion monitoring (SIM)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and mass spectrometry/mass spectrometry (MS/MS) scan modes. The technique is mature, robust, and the method of choice for an incredibly wide variety of applications.

From an historical perspective, some of the early success of GC/MS can be traced to a number of advances. These include the interfacing of a gas chromatograph to a quadrupole mass analyzer, the development of the first computerized data acquisition and mass spectral library searching software, and the acceptance of GC/MS as the method of choice for the 600 method series by the United States Environmental Protection Agency (USEPA).[2–4] In the past decade, GC/MS has evolved from a technique that had been mainly used by scientists with training in MS to a more general-purpose GC detector requiring minimal operator expertise. The main reasons for this are the proliferation of low-cost, benchtop GC/MS instruments equipped with powerful, user-friendly software to facilitate instrument control and data analysis. The ever-growing number, and types, of GC/MS applications has and will continue to drive the evolution of sample introduction systems, improved analytical figures of merit, ''smart'' autotuning procedures and diagnostic routines, and highly automated chemical analyses.

The difficulties posed by deciding on the proper context for a discussion on GC/MS are perhaps best represented by Eiceman et al. who asked ''Is gas chromatography an inlet for mass spectrometry, or is mass spectrometry a detector for gas chromatography?''[5] A number of books on MS detail the many facets associated with GC/MS.[6–12] Fundamental and application reviews describing recent advances in GC,[5] MS,[13] and their applications[14] are published biannually in *Analytical Chemistry*. Primary research articles are provided in a wide array of scientific journals too numerous to list here. For the purposes of brevity, this article will discuss a limited number of topics most relevant to the theory and practice of GC/MS. The individual instrumental components required for GC/MS will be described. These include sample introduction systems, gas chromatographs, and mass spectrometers. Data analysis including compound identification, quantitation, and artificial intelligence-based (AI) systems will be discussed. The article concludes with a discussion of practical considerations and future prospects.

2 GAS CHROMATOGRAPHY/MASS SPECTROMETRY INSTRUMENTATION

A simplified schematic diagram of a GC/MS instrument is shown in Figure 1. It may include a specialized sample introduction system for various applications (e.g. air concentrator for analysis of volatile organics in air, purge and trap module for analysis of purgable organics in water). Other components include a gas chromatograph, mass analyzer, and control computer. Each of these components is discussed here from the standpoint of requisite instrumentation. This discussion is limited to those issues most germane to GC/MS analyses, specifically sample introduction systems, GC injectors and columns; mass analyzers, ionization modes, and scan

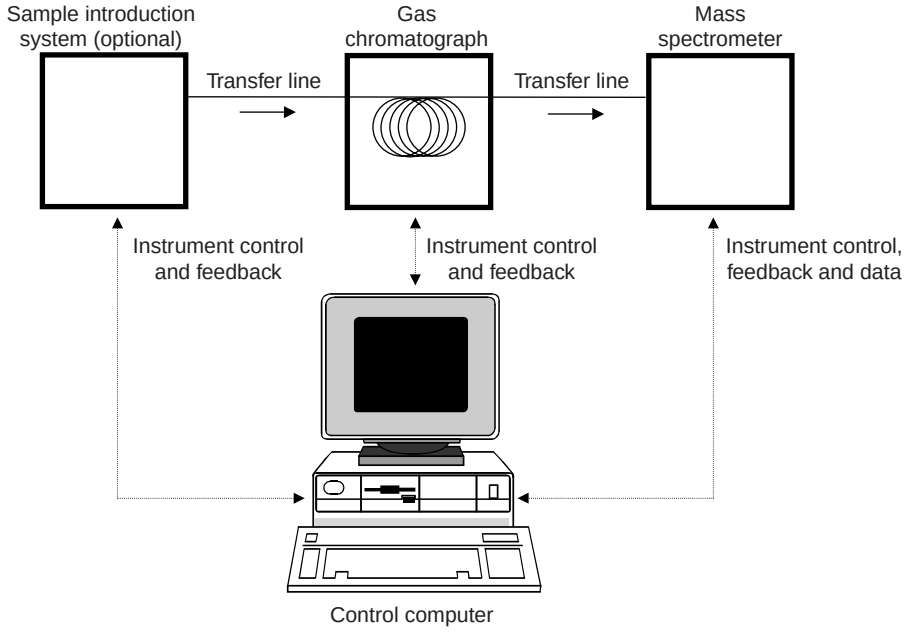

Figure 1 Block diagram of a GC/MS system.

modes; and field-portable instruments. An appropriate but slightly dated review of commercial benchtop GC/MS instrumentation is provided elsewhere.[15]

2.1 Sample Introduction Systems

In most applications, the sample workup procedure results in injection of a sample containing the analytes of interest on to the head of the column for subsequent GC/MS analysis. Examples of this might include a gas sample, biological extract, or a soil sample extract. In some cases, the sample requires special treatment or derivatization prior to injection on to the head of the column. This section discusses the more widely used sample introduction systems for GC/MS analyses.

2.1.1 Air Concentrators

Increased interest is being paid to monitoring the types and concentrations of a variety of volatile organic compounds (VOCs). The revisions to the Clean Air Act in 1990 and increased interest in the sources, transport, and fate of a variety of anthropogenic and biogenic hydrocarbons have similarly led to a large number of ongoing efforts to collect data on the types and concentrations of these compounds.[16,17] It should be noted that GC coupled with some detector other than a mass spectrometer is often entirely appropriate for analyses of specific target compounds in well-characterized samples. To give a couple of examples, GC coupled with flame ionization detection (FID) is widely used for the analysis of various hydrocarbons in gasoline, and GC coupled with ECD provides a very sensitive, selective means for monitoring halogenated species such as dioxins, polychlorinated biphenyls (PCBs), and pesticides. Such one-dimensional (1-D) GC detectors are usually unable to provide conclusive information as to the identity of unknown species in uncharacterized samples. In these cases, MS is a better detector due to the wealth of structural information it can provide.

Air samples can be collected by a variety of means. For whole air samples, Tedlar bags or stainless steel canisters are employed. For higher molecular weight VOCs (typically containing six or more carbons), a sorbent trap is used, with the sorbent agent being selected on the basis of its ability to trap the VOCs of interest. Once collected, the samples are analyzed by standard USEPA methods. Details on the sorbent-trap-based methods are provided in USEPA TO-1 and TO-2, whereas the Tedlar bag or canister-based methods are covered by USEPA TO-14.[18]

The analysis for VOCs in air samples involves a number of discrete steps. In the case of samples collected in Tedlar bags or canisters, a known volume of the sample is passed through a cryogenically cooled internal trap, which traps the VOCs while passing the bulk constituents of air (i.e. nitrogen and oxygen). In the case of samples collected on sorbent traps, the VOCs are thermally desorbed on to the cryogenically cooled internal trap. Next, the VOCs are thermally desorbed on to the head of a column. In some cases, water and carbon dioxide removal steps are employed during this process to prevent clogging of the transfer lines with ice and improve chromatographic resolution of early eluting compounds. When employing a standard bore (0.25 mm internal diameter (ID)) column, a final cryofocusing step must be used to focus VOCs on to the head of the column for subsequent desorption, to minimize zone broadening and maintain chromatographic resolution.

In years past, numerous research groups developed their own air concentrators using a relatively simple system that included a sample loop which was immersed in liquid nitrogen to trap the air sample and then rapidly heated to desorb the VOCs on to the column. In more recent years, a number of commercial air concentrators have become available. The more popular models provide separate inlets for introduction of samples and internal standards; include separate modules for removal of both water and carbon dioxide; utilize heated, specially treated (silylated or silco-grade) stainless steel for the entire sample path; and are highly automated. Nevertheless, these devices are expensive, require constant attention to ensure proper operation (e.g. ensuring that background levels are low, there is minimal carryover, etc.), and often necessitate ancillary equipment (e.g. cryofocusing module, canister cleaning system, sorbent trap conditioner, etc.). Used properly, they can provide part-per-trillion (ppt) level detection limits, minimal band broadening, and reliable results.

2.1.2 Liquid Purge and Trap

Liquid purge and trap GC is a mature, widely used technique for routine water quality testing, the analysis of off-odors in foods and beverages, and other applications. The heart of this technique is the concentration of purgable organic compounds from a headspace above a liquid sample on to a sorbent trap, which is subsequently heated to desorb and release these concentrated organics on to the head of a GC column. The bulk of these analyses are based on USEPA methods 601, 602, 501, and 502.[19] A number of vendors offer automated liquid purge and trap devices, ranging from compact units that can be mounted directly on to the GC to larger units that can handle multiple samples. As with air concentators, these devices work well for their intended applications, provide a high degree of automation, and can achieve sub-part-per-billion (ppb) level detection limits.

2.1.3 Solid-phase Microextraction

In the past, routine analyses of extractable organic compounds in aqueous-based samples and soil samples usually required time-consuming and arduous Soxhlet, liquid–liquid, and liquid–solid extraction techniques. The development of new solid-phase extraction (SPE) and the more recent solid-phase microextraction (SPME) techniques have greatly changed the way these analyses are performed. SPE techniques are probably more accurately characterized as sample preparation techniques which are used prior to GC/MS analysis and hence are not discussed further here. Although SPME is likewise a sample preparation technique, SPME cartridges can be used to extract the compounds of interest and then directly inject the extract into the GC/MS instrument. Hence their discussion is deemed appropriate here as a sample introduction system for GC/MS.

SPME is appropriate for the analysis of either volatile or semivolatile compounds in various matrices. It has been used for a wide variety of applications including analysis of VOCs in air and water; pesticides in water, fruits, and vegetables; and flavors and off-odors in foods and beverages. The technique was pioneered by Pawliszyn in the early 1990s[20] and has since been commercialized by a number of vendors. The basic principle of SPME is the adsorption of analytes of interest on to a fused silica fiber coated with a porous solid phase that selectively retains the analytes of interest. The SPME device is constructed to enable both extraction and injection and its use involves a number of discrete steps. First, an appropriate SPME fiber is chosen to extract the compounds of interest. Since the SPME device can be regenerated and used for multiple analyses, it is appropriate to first condition the SPME fiber to remove any potential contaminants or carryover. The SPME device is then exposed to the sample for a predefined period of time. The SPME device is removed from the sample and ''injected'' into the GC, where the elevated temperatures of the GC injector desorb the analyte off the fiber and on to the column.

Continued development of new fiber coatings for SPME will serve to extend the applicability of this technique. Although many vendors of SPME equipment offer application notes describing its use for a particular application, the user should note that development of new SPME methods may require significant effort to determine extraction time, extraction efficiencies, desorption temperature, and other parameters. In most cases, the large ID of the SPME needle necessitates frequent septum replacement. Alternately, a special injection valve designed specifically for SPME injections can be used to obviate the use of septa. The SPME technique is fast, selective, simple, relatively inexpensive, and represents an attractive option over conventional extraction techniques.

2.2 Gas Chromatographs

A wide variety of analytical instrument companies market gas chromatographs. Technological advances in recent years have focused on the development of a wider variety of injection techniques, columns, and other instrument refinements. In the past, the use of packed columns with their resulting inherently high flow rates of carrier gas often required a specialized interface between the gas chromatograph and mass spectrometer (e.g. jet separator, open-split interface). With the proliferation of standard bore (0.25 mm ID) chromatographic columns with an optimal flow rate of the order of $1\,mL\,min^{-1}$, the most common interface involves introducing the effluent from the GC directly into the ion source region of the mass spectrometer. As a result, jet separators and other specialized GC/MS interfaces have become less important for many analyses. These topics are detailed in various texts[6–12] and are not discussed further here.

2.2.1 Injectors

Choosing among the various injection techniques requires knowledge of the type of sample, the analytes of interest, their concentrations, and the application. The three major GC injection techniques are split, splitless, and on-column. The most common type of GC injector is the split/splitless, with which both split and splitless injections are possible. These injection techniques are illustrated in Figure 2. The split technique involves the injection of the sample on to the injector with the split valve held in the on position. This places only a fraction of the sample on the column, with that fraction defined by the ratio of the flow on to the column divided by the flow out the split vent. The splitless technique is similar to the split technique, with the exception being that the split valve is programmed to be in the off position for the approximately 30 s it typically takes for the sample to be flushed out of the injector and on to the column. The split technique is used for more concentrated samples to dilute the amount of solvent and sample flowing on to the column, whereas the splitless technique is most often employed for trace level analyses.

On-column injection of large volumes (i.e. 100 μL) has received increased attention.[21] One of the attractive features of this technique is that it obviates the requirement for solvent evaporation in the sample workup procedures. The solvent is vaporized in a short section of precolumn and then vented in a split port. In some cases, the injector is held at ambient temperatures during injection and then temperature-programmed to improve the efficiency of solvent removal. A review of large-volume injection techniques is provided by Mol et al.[22]

The recent development of pressure-programming of the carrier gas is of particular importance for GC/MS.

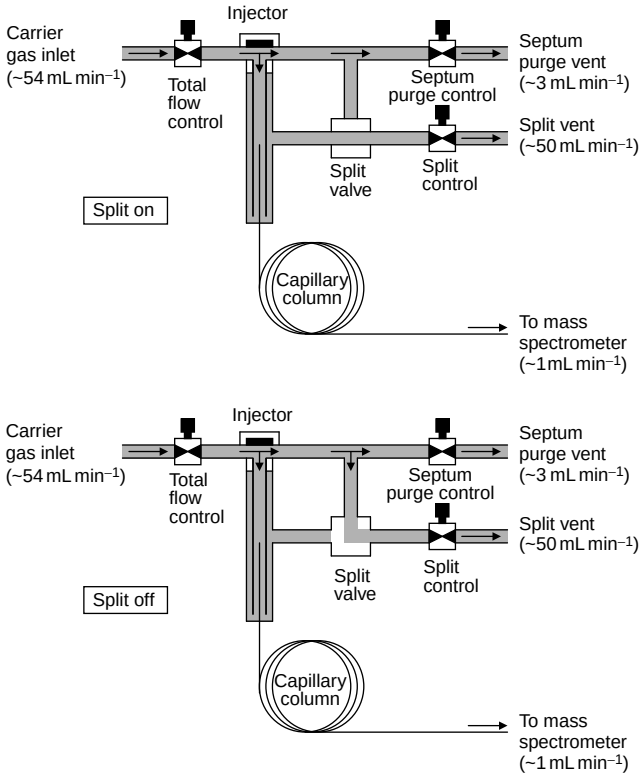

Figure 2 Illustration of split and splitless injection techniques.

In normal injectors, the flow rate through the column decreases as the temperature increases. Although the pumping systems in most mass spectrometers can handle the low carrier-gas flows provided by a range of columns, the quadrupole ion trap instrument requires a nominal 1 mL min^{-1} flow of helium to serve as a buffer gas and improve both sensitivity and mass resolution.[23] Hence, in conventional injectors, the user must live with the fact that the flow rate will not be constant over the course of a temperature-programmed elution and that this may affect the chromatography and quality of the resulting mass spectra. A pressure-programming option for the GC allows the user to choose from one of two modes of operation. In constant-flow mode, the head pressure on the column is programmed to maintain a constant flow out of the end of the column. In pressure-programmed mode, the carrier-gas pressure is increased over the course of an analysis so as to speed up the elution. Both of these modes have their utility in GC/MS analyses, and this development allows the user control over one more parameter to effect the desired outcome of an analysis.

2.2.2 Columns

Since the first reports of high-resolution capillary columns in the 1960s, development of new column technology has been evolutionary rather than revolutionary. The majority of columns used for GC/MS analyses are based on some combination of divinylbenzene and polysiloxane-based polymer for the stationary phase. Column manufacturers offer a wide array of column stationary phases, lengths, diameters, and film thicknesses. An in-depth review of chromatographic theory, solid supports, liquid phases, and other details is provided elsewhere.[5] This section will briefly describe new developments in column technologies.

One salient development has been specialized low-bleed columns for GC/MS analyses. These columns provide in some cases up to an order of magnitude reduction in column bleed at elevated temperatures. This is illustrated in Figure 3, which plots the reconstructed total ion current (TIC) as a function of time for both a standard column and a new low-bleed DB-5 column. Since detection limits are ultimately related to signal-to-noise ratios, reduction of column bleed is an importance advance that will provide improved performance, especially for analytes eluting at higher temperatures.

Another development is new porous layer open tubular (PLOT) columns. These columns have a solid-support stationary phase bonded on to the walls of the column. The separation process is effected by partitioning or by adsorption of analytes on to this stationary phase. These columns are particularly effective for separating highly volatile, low-boiling compounds such as permanent gases,[24] low-molecular-weight hydrocarbons,[25] and halocarbons.[26] As most of these columns still have a minimum ID of 0.53 mm, the resulting large carrier-gas flows may require the use of an open-split interface or a jet separator interface between the gas chromatograph and the mass spectrometer.

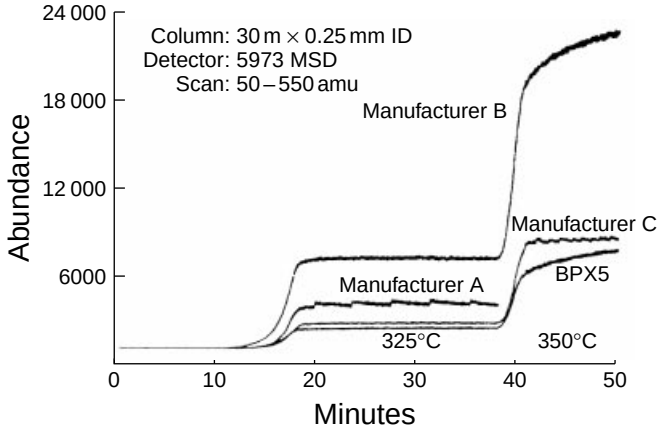

Figure 3 TICs showing column bleed with conventional and low-bleed columns. (Reproduced from D.R. Di Feo, Jr, 'A Technical Tourniquet for Capillary GC Column Bleed', *Todays Chemist at Work*, 39–44, (September 1997), by permission of ACS Publications.)

One related development germane to this discussion is that of transfer line GC/MS. This technique was originally demonstrated by Hale and Yost.[27] It is based on resistive heating of a short metal-coated column. It has been mainly used for analyses of organics in soil[28] and air.[29]

2.2.3 Fast Gas Chromatography

The field of fast GC is described in several recent reviews.[5,30,31] Truly portable gas chromatographs with low power consumption and small footprints have been available for years.[32–34] The use of fast GC in conjunction with MS is particularly valuable insofar as it can quickly separate complex mixtures and provide orthogonal information via chromatographic separation in time prior to mass spectrometric detection.[35] Successful application of fast GC/MS requires both narrow bandwidth injections and a mass analyzer/data system capable of acquiring data at rates high enough to preserve the information content of the chromatogram. The major impediment to field applications of fast GC/MS involves the inherent difficulties associated with making the mass analyzer and its accompanying vacuum system truly field-portable.[36]

2.3 Mass Analyzers

The tremendous versatility of MS is due in a large part to the wide variety of mass analyzers that are available. The different types of mass spectrometers are usually characterized on the basis of what physical property they employ to separate ions of differing m/z. In a time-of-flight (TOF) instrument, ions are separated on the basis of their differing velocities down a flight tube. In a magnetic sector instrument, ions are separated on the basis of their differing curvatures through a magnetic field. In a quadrupole mass spectrometer and a quadrupole ion trap, ions are separated on the basis of their stability in the presence of a radiofrequency (RF) electric field. In Fourier transform ion cyclotron resonance (FTICR) or Fourier transform mass spectrometry (FTMS) instruments, ions are separated on the basis of their differing frequencies of oscillation and the resulting image currents these provide. The fact that there are so many different types of mass analyzers serves only to strengthen the technique as a whole, as each has its inherent advantages and each has utility for specific applications. Although the term ion trap can be considered to apply to any device that is capable of trapping ions (e.g. quadrupole ion trap, FTICR, or FTMS), the term ion trap will be used exclusively to refer to the quadrupole ion trap from this point forward.

Detection of ions in most mass spectrometers is accomplished through the use of an electron multiplier. This device amplifies the current from an ion beam by a factor as high as 10^7. The resulting signal is typically passed through a current-to-voltage converter and then an analog-to-digital converter. Although the resulting signal is represented by the data system as the intensity or abundance, this is an arbitrary unit that has no relation to the actual number of ions detected. For negative-ion mass spectra, a conversion dynode is required to convert negative ions to positive ions prior to detection by the electron multiplier.

Quadrupole and ion trap mass spectrometers comprise more than 75% of all mass spectrometers sold. This percentage is certainly higher if the group is further limited to only GC/MS instruments, most likely because these particular mass analyzers provide the best combination of sensitivity, performance, size, and ease of use.[3] Hence, although GC used in conjunction with dual-sector mass spectrometers and FTMS instruments provides the high mass resolution required for certain applications, such applications are far from routine and the discussion here will be limited to the unit mass resolution GC/MS instruments.

Quadrupole mass analyzers offer the primary advantage of ruggedness and ease of use. Although they are relatively free from the ion/molecule problems associated with ion traps, they are not capable of implementing MS/MS experiments, and switching from EI to CI mode requires changing of the ion volumes used in the source. Ion traps are capable of performing both EI and CI experiments without requiring changes in source hardware (in internal source-type instruments only), and can implement MS/MS and MS^n (n stages of mass spectrometry) scan modes. It should be noted that CI experiments in an ion trap typically require much lower pressures of reagent gas than conventional beam-type mass spectrometers (i.e. 0.001 Pa versus 10–100 Pa). Negative CI experiments on an ion trap require the use of an external ion source, as electrons do not have a sufficiently long residence time to permit electron capture to occur in the trap. The implementation of MS experiments in a ion trap instrument is more a function of time than space, insofar as a quadrupole mass analyzer has separate regions for ionization, mass analysis, and detection, whereas each of these steps is performed in a distinct time slice in an ion trap. This has some important implications with respect to ion/molecule reactions, in that under certain conditions and given sufficient time, an ion can undergo charge exchange (CE) with neutral species in the ion trap. This can lead to protonation of analyte ions by fragment ions (which typically occurs under space charge conditions), unexpected ions, and erroneous library search results. These effects have been well studied and should provide minimal problems given proper selection of experimental parameters.[37]

With respect to sensitivity, there are relatively minor differences between the detection limits of various mass spectrometers used in GC/MS instruments. While vendors continually tout their own instruments as the most sensitive, it is often difficult to compare the performance of different mass analyzers when the vendors use different compounds for assessing their sensitivity and developing their specifications (e.g. methyl stearate, hexachlorobenzene, decafluorobenzophenone). Selectivity can be enhanced through the use of selective ionization techniques and/or more specific MS scan modes. Many mass spectrometers are now available in benchtop size and costs continue to decrease. Not surprisingly, it is the features of the mass spectrometer that often play a more important role in choosing an appropriate GC/MS instrument.

2.3.1 Ionization Modes

The most common ionization modes used in GC/MS analyses are EI, CI, and electron capture–negative chemical ionization (ECNCI). A more detailed discussion of each of these ionization modes is provided elsewhere.[38] Although all of these ionization modes are accessible with the use of an external ionization source for the ion trap, the use of ECNCI is precluded with internal ionization ion traps. It should also be noted that the use of CI in quadrupole and external ion trap instruments often requires switching to a ''closed'' ion volume to enable higher CI reagent gas pressures. In some applications, the user should consider different ionization modes and even different CI reagent gases to improve the selectivity of the analysis. The following section discusses each of these ionization modes, how they affect the types and relative intensities of ions in the resulting mass spectra, and their advantages and disadvantages for GC/MS applications.

EI is the most common ionization mode. The reactions involved are shown in Scheme 1.

$M + e^- \Longrightarrow M^{\bullet+*} + 2e^-$	Ionization
$M^{\bullet+*} \Longrightarrow M^{\bullet+}$	Formation of molecular ion
$M^{\bullet+*} \Longrightarrow A^+ + B^\bullet$	Decomposition of molecular ion to form ion and radical
$M^{\bullet+*} \Longrightarrow C^{\bullet+} + D$	Decomposition of molecular ion to form ion and neutral loss

Scheme 1 Reactions associated with EI.

EI typically employs 70 eV electrons to ''knock'' an electron off a sample molecule M and create the $M^{\bullet+*}$ ion. This process results in an ion population with a distribution of energies which can subsequently undergo collisional or vibrational deactivation to form a molecular ion $M^{\bullet+}$, fragmentation to form the ion A^+ and radical $B^\bullet$, or rearrangement to form the ion $C^{\bullet+}$ with a neutral loss of molecule D. The result is a mass spectrum that includes a wide range of fragment ions that can serve as a ''fingerprint'' for that compound. This is illustrated in Figure 4(a), which shows an EI mass spectrum of α-pinene, a common monoterpene. Although EI is perhaps the simplest and easiest ionization mode to implement, it ''spreads'' the signal from a compound over a large number of fragment ions, which may result in diminished precursor ion counts and hence reduced sensitivity for MS/MS applications.

CI is the next most common ionization technique. In the broadest sense, CI encompasses a range of techniques based on ion molecule reactions, including positive chemical ionization (PCI), ECNCI, and CE, but the term is most often used to refer to the most common form of CI which is PCI. CI is typically employed to provide an increased molecular ion signal from which a compound's molecular weight may be deduced. An illustration of the major reactions associated with the use of methane as a CI reagent gas is shown in Scheme 2.

$CH_4 + e^- \Longrightarrow CH_4^{\bullet+} + 2e^-$	Electron ionization of reagent gas
$CH_4^{\bullet+} + CH_4 \Longrightarrow CH_5^+ + CH_3^\bullet$	Ion/molecule reaction to produce reagent gas ion
$M + CH_5^+ \Longrightarrow MH^+ + CH_4$	Transfer of proton from reagent gas ion to sample molecule

Scheme 2 Reactions associated with CI.

If the pressure of the methane reagent gas is significantly greater than that of the sample molecule M, methane is preferentially ionized by an EI process to form the ion $CH_4^{\bullet+}$. This cation then undergoes a proton transfer reaction with a neutral methane molecule to produce the reagent ion CH_5^+, which in turn undergoes another proton transfer reaction with the sample molecule M to form a protonated molecule MH^+.

The exothermicity of this last reaction is governed by the difference in proton affinity (PA) between the reagent gas and sample molecule. If the PA of the sample molecule is less than that of the reagent gas, ionization will not take place. Selecting reagent gases with increasingly larger PAs than the sample molecule will increase the extent of fragmentation in the resulting mass spectra. These features are often exploited to selectively ionize and control the amount of energy imparted in the sample molecule and hence fragmentation in the resulting mass spectrum. This is illustrated in Figures 4(b) and 4(c), which show the methane and acetonitrile CI mass spectra of α-pinene. In the case of acetonitrile CI, the difference in PAs is smaller than with methane CI, with a much more intense MH^+ ion at m/z 137 and a simpler mass spectrum.

Most routine applications of CI typically involve the use of methane as a reagent gas, which has a PA low

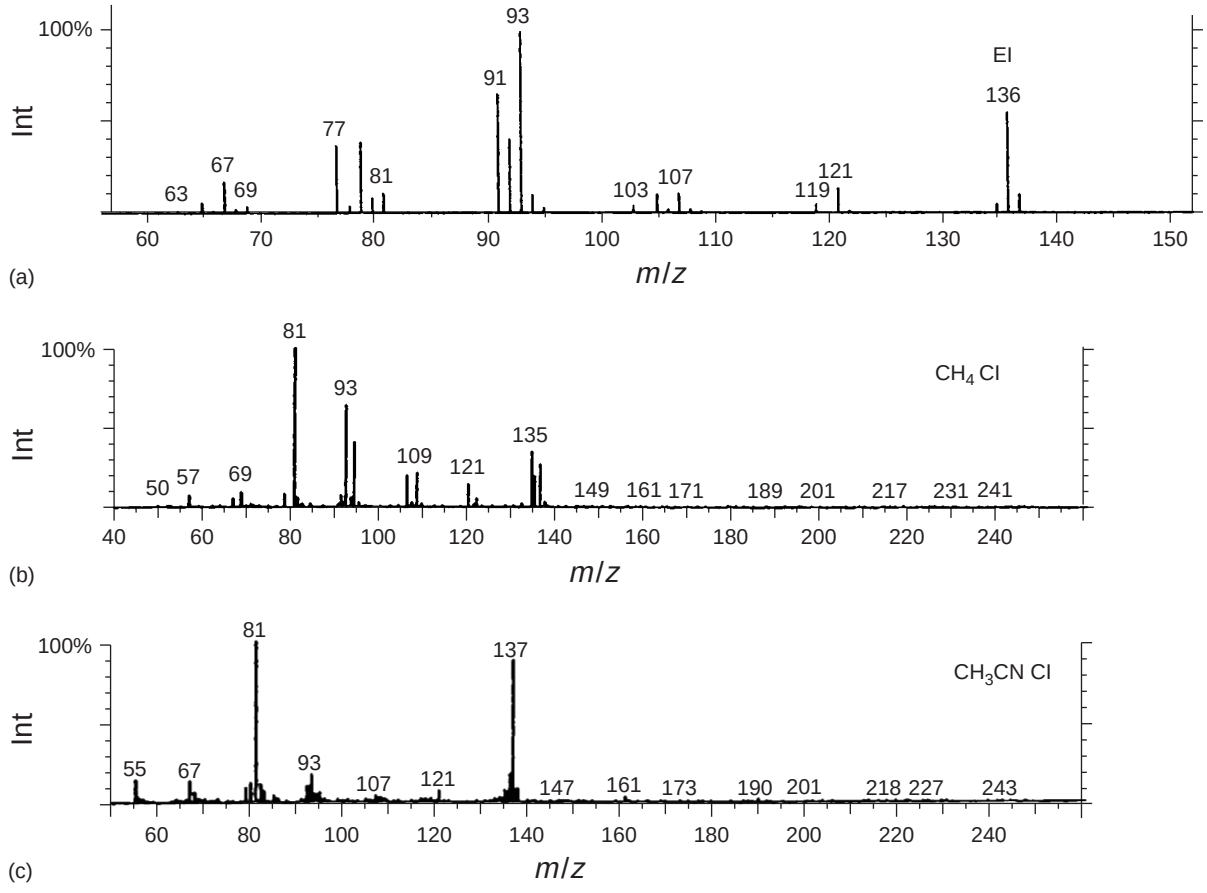

Figure 4 (a) EI, (b) methane CI, and (c) acetonitrile CI mass spectra of α-pinene.

enough to protonate nearly any organic compound. Many applications take advantage of differing PAs of the components in a mixture by using a reagent gas with a PA higher than methane to provide more selective ionization of analytes in question. The use of CI in conjunction with MS/MS is quite common, since MS/MS often requires a large molecular ion signal for subsequent isolation and fragmentation. Here, the reagent gas is chosen to concentrate the signal for a sample into fewer ions and provide a larger initial precursor ion count for a subsequent MS/MS experiment. The most common reagent gases in order of increasing PA include methane, isobutane, water, methanol, acetonitrile, and ammonia. More extensive tabulations of PAs of various reagent gases are provided elsewhere.[38]

In ECNCI, a negatively charged molecular ion is formed when a sample molecule captures an electron. These electrons can be generated thermally from a filament or produced as a by-product of gas-phase ion/molecule reactions. Since electrons formed inside an ion trap are rapidly ejected by the RF field, the use of this ionization technique necessitates an external ionization source for the ion trap. Given that rate constants for electron capture are typically two orders of magnitude greater than those for ion/molecule reactions, ECNCI is inherently more sensitive than conventional CI techniques. ECNCI is particularly well suited for organic compounds containing halogen atoms, nitro groups, and aromatic rings.

2.3.2 Scan Modes

Given the wide variety of mass analyzers and scan modes, it is often difficult for the novice to understand the implementation and fragmentation pathways followed in specific MS experiments. Lehman[39] and Kondrat[40] provide useful pictograms for representing such information. Nearly all mass analyzers allow the user several scan modes to choose from. These range from the relatively nonspecific MS and SIM modes to the more selective tandem MS (MS/MS, MS/MS/MS, and MS^n) modes. In the case of tandem quadrupole mass spectrometers, all three MS/MS scan modes are accessible. This includes product-ion, parent-ion, and neutral-loss scans. In the case of quadrupole ion traps, only product-ion scan modes are accessible in commercial instruments. This

section discusses each of these scan modes and their use in GC/MS applications.

Mass spectrometers are most commonly operated in conventional MS scan mode, in which the mass analyzer is set to acquire a mass spectrum across a user-specified range of m/z values. MS mode is typically used when attempting to obtain sample spectra for subsequent library searching and tentative identification. SIM mode involves setting the mass analyzer to acquire intensity data across specific m/z value(s). It is most commonly used in conjunction with GC/MS to improve detection limits for specific target compounds and the resulting plot of the intensity of that selected ion as a function of time is referred to as a SIM mass chromatogram. As a more definitive identification of a particular compound typically requires the presence of more than one characteristic m/z value for that compound, a SIM application may require monitoring of more than one ion. On GC/tandem quadrupole MS instruments, parent and neutral-loss scan modes are useful when screening for particular target compound classes. Hunt and Shabanowitz[41] provide an excellent summary of the use of MS/MS scan modes to identify specific target compounds.

The proliferation of relatively low cost GC/MS instruments based on the use of ion trap mass analyzers with the MS/MS capability has been shown to be useful for a variety of mixture analysis problems. In these applications, the selectivity of MS/MS is used in conjunction with GC to simplify target compound analysis. MS/MS involves isolation and storage of a specific precursor ion, fragmentation of that ion via collision-induced dissociation (CID), and acquisition of a mass spectrum of the product ions. The ion trap enables additional flexibility in the ways in which isolation and CID can be performed. Isolation techniques include apex, consecutive or two-step, and broadband techniques such as random noise and SWIFT (stored waveform inverse Fourier transform).[42–46] CID techniques can be achieved via single frequencies (resonant excitation) and broadband techniques (nonresonant excitation).[47–50] In contrast to conventional tandem mass spectrometers which require a separate mass analyzer for each stage of MS (tandem-in-space), the ion trap can perform ionization, isolation, CID, and mass analysis within one analyzer (tandem-in-time). This means that the ion trap can be used to perform two (MS/MS), three (MS/MS/MS), and unlimited (MS^n) stages of MS. In practice, two stages of MS usually suffice for selective identification of a target compound. Although parent and neutral-loss scan modes have been demonstrated on an ion trap, they have yet to be implemented on any commercial ion trap instruments.

The primary reason to use an MS/MS scan mode in GC/MS analyses is that it simplifies the resulting chromatogram for subsequent target compound analysis. Detection limits in MS/MS scan mode are usually better than MS due to reduced noise levels. They may not be as low as SIM as the CID process in an MS/MS experiment may not result in 100% conversion of a precursor ion to a product ion, especially for highly conjugated ions. However, MS/MS is a much more reliable means for target compound identification and quantitation, as a precursor–product ion combination is inherently more characteristic of a particular compound than a single ion from either MS or SIM modes. MS/MS is not a panacea for each and every target compound. Lightweight compounds, such as CH_4 and NO_2, produce few fragment ions and hence cannot be monitored via MS/MS. Structural isomers, such as xylenes and monoterpenes, produce nearly identical MS/MS spectra and hence cannot be differentiated from one another unless prior chromatographic separation is achieved. These cases, however, are the exceptions, and MS/MS is becoming more widely accepted for selective monitoring of target compounds in complex mixtures.

2.3.3 Field-portable Mass Spectrometers

The development of field-portable GC/MS instrumentation is receiving increased attention. The driving force behind this is the ever-growing number of applications requiring rapid, in situ monitoring by instrumentation with minimal space, weight, and power requirements. Several companies have developed and marketed small gas chromatographs and these instruments have been used widely for years. The use of fast GC implies smaller diameter columns and reduced sample capacities, and in some cases this may result in poorer detection limits. Fast GC also results in peaks that are very narrow as measured on the chromatographic timescale, and hence necessitates a detector with a very fast response time, or, in the case of an MS detector, fast scanning speeds. Only a few companies have developed field-portable MS and GC/MS instruments. Bruker's Mobile Environmental Mass Spectrometer can be equipped with a number of sampling modules for air, water, and soil analysis.[51–55] While this system has a proven ability for on-site analysis, it is mobile rather than truly portable. More recently, efforts have focused on developing smaller instrumentation that can be carried into the field on a truck or ported to a site by one or more individuals; and a human-portable quadrupole-based GC/MS system has been developed.[56–58] Other groups have developed miniaturized sector,[59] TOF,[60,61] and ion cyclotron resonance-type mass spectrometers.[62,63] Recent reviews discuss requirements for field-deployable monitors[64] and detailed specifications for commercial field-portable GC and GC/MS instruments.[65]

3 DATA ANALYSIS

GC/MS instruments usually come equipped with data systems provided by the manufacturer. Usually, this is done so as to enable both instrument control and data analysis. Since instrument control routines are obviously specific to each instrument, they are not discussed further here other than to mention that data systems increasingly automate more routine tasks such as ''smart'' autotuning and diagnostics. Instrument manufacturers have been repeatedly urged by users to integrate routines into their data systems to enable GC/MS data to be saved in the JCAMP-DX (Joint Committee on Atomic and Molecular Physics–Data eXchange) format that is becoming the standard format to facilitate data interchange. Other non-instrument-specific data systems have been developed for the sole purpose of data analysis. These data systems may include enhanced features such as library search routines, and chemical structures for each compound in the library. They are useful in larger MS laboratories insofar as they free the user from having to learn and utilize the different data systems associated with each instrument.

One of the most common GC/MS data analysis procedures involves library searching of experimental mass spectra against reference or library spectra. In the past, instrument manufacturers integrated such routines into their data systems using their own libraries and search routines. More recently, both vendors and users have begun to standardize on the use of mass spectral libraries and search routines provided by the National Institute of Standards and Technology (NIST). NIST continues to add new spectra to their library, performs quality control and oversight over new and existing data, and has continually improved the algorithms for library searching. At the most recent count, the NIST mass spectral library contained more than 129 000 EI mass spectra of over 107 000 different compounds. The NIST library now represents the most widely used mass spectral library in the world. The NIST search software also includes tools for deconvoluting GC/MS data, mass spectral interpretation, and chemical substructure identification.

In recent years, more attention has been paid to developing and formalizing good laboratory practices (GLPs) as related to the application of MS. A recent report by Boyd et al. discusses efforts towards standardization of instruments, quality assurance/quality control (QA/QC) procedures, and performance requirements and is very worthwhile reading for any practitioner of GC/MS.[66] Another worthwhile report by Baldwin et al. discusses criteria for identification, quantitation, and determining detection limits.[67]

3.1 Identification

Mass spectral libraries used in target compound analyses may be larger, more comprehensive libraries such as the NIST library or smaller, more customized libraries used for specific applications. The former are typically encoded into electronic format and are available from NIST, Wiley/NBS (National Bureau of Standards),[68] and EPA/NIH (Environmental Protection Agency/National Institute of Health)[69] (note that these last two have been largely supplanted by the NIST library). The latter include specialized compilations of mass spectra in hardcopy form of various compound classes including steroids;[70] oil components;[71] pollutants, pesticides, and metabolites;[72] environmental contaminants;[73] drugs;[74] priority pollutants;[75] and ECNCI spectra.[76]

Commercially available data systems use one of two basic search algorithms in computing the degree of similarity between experimental and library spectra. These are the dot-product metric (used in Finnigan and Varian data systems) and the probability metric used in the probability-based matching (PBM) system (used in Hewlett-Packard data systems). The search can be bounded by various constraints, such as molecular weight and elemental composition. The result of the search is a ranked list which includes some metric to describe the degree of match between experimental and library spectra. Various library search algorithms utilize a number of data-system-specific terms to describe this degree of match. The ''match factor'' (used in NIST library search software) or ''purity'' metric (used in Finnigan and Varian data systems) attempts to take into account the degree of match between the experimental and library spectra. This metric is computed by a dot-product of a peak-by-peak comparison of the two spectra. The ''reverse match factor'' (NIST library search software) or ''fit'' metric (used in Finnigan and Varian data systems) attempts to fit the library spectrum into the experimental mass spectrum. Peaks that are in the experimental spectrum but are not in the library spectrum are disregarded in evaluating the degree of match. This type of search has the advantage of being able to determine multiple components in an experimental spectrum and/or to disregard any contaminant or background peaks in the sample spectrum. An ''rfit'' or ''reverse fit'' (used in Finnigan and Varian data systems) attempts to fit the experimental mass spectrum into the library spectrum. Given the confusing terminology and different means by which various algorithms compute the degree of match, users are urged to consult the documentation for their data system for more detailed information as to how this is achieved.

The result of a library search is no more than a ranked list of the most similar library mass spectra.

Sparkman has summarized the use and misuse of library searching[77] and users are reminded that a library search is meant to assist the user in identification and that the top ranking hit is not necessarily proof positive of the identification. Erroneous results may be due to errors in the experimental mass spectrum, errors in the reference mass spectra, or the simple fact that the library does not contain that spectrum.

USEPA criteria for positive identification of a target compound usually call for a retention time match with a known standard as well as simultaneous maximization of three major ions for that compound within a specified time window.[67] More commonly, data systems are configured to identify a target compound on the basis of both a retention time match and a library search match above a certain preset match factor. It should be noted that recent development of retention time locking software minimizes drift of retention times due to instrument or parameter modifications. When the mass spectrometer is operated in SIM mode, positive identification requires both retention time match as well as the presence of one or more quantitation or ''quan'' ions within a specific time window. For operation in MS/MS mode, the USEPA has yet to develop formal criteria for positive identification of target compounds via GC/MS/MS. Although the American Society for Mass Spectrometry's (ASMS) workshops have discussed this issue,[67] the selection of a suitable precursor–product ion combination is left to the user.

3.2 Quantitation

External standards are most often used for quantitation of a limited number of samples. Here, working standards of the target compounds are prepared from appropriate sources (e.g. pure liquids, dilute stock solutions, or gas standards). These standards are analyzed under the same conditions as the samples. The quan ion is chosen to reflect an ion that has high abundance and, in the cases of potential interference, is selective for that compound. A calibration curve representing the quan ion intensity versus concentration is constructed and used to compute concentrations in the samples.

Internal standards are used to account for short-term drift in instrument response in quantitation. The process is the same as with external standards except that standards and samples are ''doped'' with a constant amount of a carefully selected internal standard. The calibration curve therefore plots the ratio of quan ion intensity for the target compound to the quan ion intensity for the internal standard. To determine the concentrations of large numbers of target compounds in a single analysis, several internal standards may be employed to more accurately reflect the different chromatographic behaviors of the target compounds. For the most accurate quantitation, the internal standard is an isotopically labeled form of the analyte in question. This means of quantitation is usually referred to as an isotopic dilution method. In this case, the internal standard behaves almost identically to the target compound, will elute at nearly the same retention time, and will account for any analyte losses during sample preparation.

The use of powerful quantitation software in modern GC/MS data systems can greatly facilitate and automate the identification and quantitation processes. This software typically has a steep learning curve associated with it, requires time to enter parameters for searching and quantitation, and hence is typically applied in routine, high-volume analyses. Despite the power of automated identification and quantitation software, misinterpretation of data can sometimes occur and users should remember that these routines do not remove the need for proper review of the data and results.

3.3 Pattern Recognition, Neural Nets, and Artificial Intelligence Methods

The ''heyday'' of the development of techniques based on artificial intelligence (AI) for automated mass spectral interpretation is past. Over-zealous predictions of expert system and AI methods have yet to meet user expectations. Although both the Dendral Project[78,79] and the Self-training Interpretive and Retrieval System (STIRS)[80] have demonstrated their utility for structure elucidation from MS data, these software tools never enjoyed wide acceptance by users. Nevertheless, a wide variety of multivariate data analysis, pattern recognition, and neural network software continue to show their utility in diagnosing problems with GC,[81] correcting for mass spectral drift,[82] and dealing with the large amount of information contained in GC/MS data.[83,84] The use of such techniques for GC/MS data analysis continues to evolve, and advances are slowly being incorporated into data systems.[85]

4 PRACTICAL CONSIDERATIONS

As with any analytical technique, the most useful teaching method for learning the many and sundry details associated with the practice of GC/MS is hands-on experience. The purpose of this section is to briefly touch upon some of the more important considerations associated with optimizing both the gas chromatograph and mass spectrometer to obtain the desired results.

4.1 Gas Chromatography

Of critical importance for GC/MS analyses is developing the appropriate chromatographic conditions to achieve the desired separation. This job is made easier by the wide range of stationary phases, column lengths, IDs, and film thicknesses now available. A quick perusal of any chromatography catalog shows a wealth of information regarding the various options for analysis of specific compounds of interest for an incredibly wide variety of applications. Once the column is selected, the user must then decide upon a temperature program and carrier gas flow rate. For highly volatile species, subambient GC temperatures may be required to effect separation of low-boiling compounds or alternately a longer column can be chosen. For more polar species, some chromatographic columns have higher upper temperature limits, thus enabling eluting of these higher boiling species. Rather than using best guess or trial and error methods for deciding upon appropriate parameters for a GC separation, the user can employ modeling,[86] neural network,[87] or simplex software[88] to optimize conditions.

In the past, most GC columns had sufficiently large IDs that they provided gas flows that were too large for direct coupling to a mass spectrometer. Hence, devices such as open-split interfaces, jet separators, and even membrane separators were employed to reduce the gas flows and in some cases provide an extra enrichment step prior to introduction into the source region of the mass analyzer. These devices are rapidly becoming obsolete, as manufacturers devise mass analyzers and/or pumping systems that are more forgiving of the wide range of gas flows from microbore (<0.25 mm ID), standard bore (0.25 mm ID), and megabore (0.53 mm ID) columns.

Furthermore, users are becoming more familiar with simple techniques to provide easier switching of columns. Rather than connecting the end of a column directly to the source region of the mass spectrometer, many users use a short section of deactivated fused silica transfer line between the column and mass analyzer. This practice means that the user does not need to vent the mass spectrometer when changing columns, which saves both time and effort. But it also necessitates a means for connecting the column to the transfer line. Past practice here was to use low-dead-volume stainless steel fittings with appropriate ferrules. These devices were difficult to connect properly, heavy enough to place stress on the column and transfer line, and required sufficient experience to adequately judge when the nuts were sufficiently tightened to minimize leaks. Polyimide-coated glass press-fit connectors are a much better option for these connections, since they are easy to use, lightweight, and provide an excellent leak-tight seal so long as the column is cut properly and inserted correctly.

4.2 Mass Spectrometry

One of the most problematic aspects involved with the practice of GC/MS is the fact that the instrument requires a vacuum system. A surprising feature of many GC/MS instruments is that they do not include a pressure sensor or do not even permit one to be attached to the vacuum manifold without substantial modifications. The operating pressure of the mass spectrometer is a critical piece of information that is highly useful in gauging the status of the instrument. With a quick glance at the pressure, the user can determine whether the system is operating nominally or if pressures are excessive. Although the pressure can sometimes be inferred from the instrument's diagnostic routines (i.e. turbo pump speed, which is usually expressed as a percentage of normal operating speed), this is certainly no substitute for an actual pressure readout. For users purchasing a GC/MS system, it is highly recommended that the instrument be configured with some sort of pressure gauge.

Leak detection remains as one of the more problematic aspects to operating a GC/MS instrument. In performing any maintenance or service to the mass analyzer, the user runs the risk of inducing a leak into the vacuum system and hence either reducing the performance of the instrument or preventing its operation. If the leak is minor enough so that the instrument can be operated without risk of damage to the filament and multiplier, the mass spectrometer itself can be used as a leak detector. Here, the user would direct a stream of argon or some other low-molecular-weight hydrochlorofluorocarbon (HCFC) at suspected leak sites while concomitantly observing the mass spectral response at the appropriate m/z range.

In an ion trap instrument, the relative ratios of air/water peaks can be used to deduce the presence of either excessive water levels or air leaks. Here, reactions between ions and neutrals are the cause of the observed response. When there are high water levels in the ion trap, the reactions shown in Scheme 3 occur.

$$H_2O + e^- \Longrightarrow H_2O^{\bullet +} + 2e^- \quad \text{EI process resulting in formation of } m/z\ 18$$

$$H_2O^{\bullet +} + H_2O \Longrightarrow H_3O^+ + HO^{\bullet} \quad \text{Ion molecule reaction resulting in formation of } m/z\ 19$$

Scheme 3 Production of water ions in an ion trap.

Hence, the large intensity ratios of m/z 19 (H_3O^+) to 18 ($H_2O^{\bullet +}$) indicate excessive water levels in the trap. When air levels are nominal, the ratio of m/z 28 ($N_2^{\bullet +}$) to 32 ($O_2^{\bullet +}$) is usually approximately 3 to 1. Large levels of nitrogen contamination in the carrier gas are indicated by higher ratios of m/z 28 to 32. When there is significant air in the ion trap, the reactions shown in Scheme 4 occur.

$$N_2 + e^- \Longrightarrow N_2^{\bullet +} + 2e^-$$ EI process resulting in formation of m/z 28

$$N_2^{\bullet +} + O_2 \Longrightarrow N_2 + O_2^{\bullet +}$$ Charge exchange process resulting in formation of m/z 32

Scheme 4 Production of air ions in an ion trap.

Large intensities of m/z 32 relative to 28 indicate the presence of a minor air leak. The presence of a broad peak centered around m/z 32 indicates the presence of a major air leak. The reason in both cases is that the energetics of the CE process favor oxygen molecules retaining the bulk of the charge under multiple collision conditions.[38]

Deducing the source of a major leak is often problematic. Although it is hoped that the source of such a leak could be deduced by sound or by rechecking the connections associated with the last modifications done to the system, this does not work in all cases. Such problems are best approached systematically, with the user working backwards from the instrument, checking that the pumps are in working order, and checking all connections to the vacuum manifold, and the connections between the column and the transfer line to the mass spectrometer.

Chemical background is also of concern for trace level analyses. This background can be due to septum bleed, column bleed, pump oil, and trace contaminants in the carrier gas stream. The user must pay close attention to choosing the appropriate quality components to provide the background levels necessary for the analysis in question. The use of silco-grade steel in the manufacturing of mass spectrometer components has been shown to provide reduced background levels as well as facilitating analyses of more polar species.

GC/MS instruments increasingly come equipped with powerful diagnostic routines to check the performance of individual components. Users can use these to debug problems and trace them to specific sources, whether it is a power supply, specific printed circuit board, or component. These instruments also come equipped with autotune procedures that automate processes such as multiplier gain check, mass calibration, and other common operations. Although the user should take full advantage of these in day-to-day operations, they are no substitute for proper evaluation of instrument performance by a human expert. USEPA procedures commonly call for the instrument to be tuned daily if necessary to ensure that EI mass spectra of bromofluorobenzene or some other reference compound meet predefined performance criteria. As a final suggestion, users are urged to maintain a logbook to document instrument use, performance, and maintenance procedures such as oil changes.

5 PERSPECTIVE AND FUTURE DEVELOPMENTS

The future of GC/MS should see further progress in improvement of GC performance (i.e. faster elutions, lower bleed columns, wider variety of columns), MS performance (i.e. lower detection limits, wider mass ranges, faster scanning speeds), data systems (totally automated tuning, analysis, interpretation and quantitation), and further miniaturization for specialized field applications. Ion traps will most likely be the mass analyzer of choice for GC/MS and GC/MS/MS applications due to their numerous high-performance features.[89,90] The ion trap is small, relatively simple, and inexpensive. It is recognized as one of the most sensitive mass spectrometers currently available. It has excellent experimental versatility and is capable of collecting EI, CI, and MS/MS data. This tandem MS capability is particularly valuable for targeted compound analysis, in which additional stages of MS can be used to tailor the selectivity of the analysis to the compound and matrix of interest. High mass analysis,[91,92] high mass resolution,[93] and even the Fourier transform technique have been demonstrated on an ion trap.[94,95] Recent work shows that the ion trap can be operated using air instead of helium as a buffer gas while maintaining unit mass resolution up to m/z 200.[96] It remains in the hands of instrument manufacturers to incorporate these advances into commercial GC/MS instruments. Collectively, these features make the ion trap well suited for a host of applications and the future should see ion traps gradually replacing quadrupole mass analyzers in these applications.

Although MS/MS and its utility for target compound analysis and structure elucidation has been known for years, it has only recently been exploited on a more routine basis in conjunction with GC. Some of the limiting factors in this have been the high cost of tandem MS instruments and the lack of progress in the development of computerized databases of MS/MS spectra. Although there have been some efforts in this area,[97,98] the problems associated with differing MS/MS spectra from different instruments and the dependence of such spectra on operating conditions (e.g. CID gas pressure, collision energy, etc.),[99,100] and the lack of further championing of such efforts by an organization such as NIST, have hampered efforts in this area. If a user wants to employ MS/MS for target compound analysis, there is no database or tabulated information on precursor–product ion combinations. Although such information can in some cases be gleaned from the literature,[41] more often the user must evaluate a compound's fragmentation patterns. The continued development and eventual acceptance of MS/MS as a technique will require the development of an MS/MS database.

But without question, data systems represent the greatest area for progress. The improvements in data system abilities to facilitate instrument control and data analysis has been a major reason behind the acceptance of MS as a more routine detector for GC, especially in the hands of nonexperts. Future instruments should provide even more powerful software for ''autotuning'', data interpretation, and diagnostics. Finally, ''smart'' selection of parameters for MS/MS experiments (i.e. selecting precursor–product ion combinations, tuning for precursor ion selection and CID processes) will enhance the utility of GC/MS/MS experiments. These advancements will continue to promote an even wider acceptance and use of GC/MS in the future.

ACKNOWLEDGMENTS

The author acknowledges NASA contract no. NAS 9-19410 for financial support and O. David Sparkman for his helpful comments and suggestions.

ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
ASMS	American Society for Mass Spectrometry
CE	Charge Exchange
CI	Chemical Ionization
CID	Collision-induced Dissociation
ECD	Electron Capture Detector
ECNCI	Electron Capture–Negative Chemical Ionization
EI	Electron Ionization
FID	Flame Ionization Detection
FTICR	Fourier Transform Ion Cyclotron Resonance
FTMS	Fourier Transform Mass Spectrometry
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GLP	Good Laboratory Practice
HCFC	Hydrochlorofluorocarbon
ID	Internal Diameter
MS	Mass Spectrometry
MS^n	*n* Stages of Mass Spectrometry
MS/MS	Mass Spectrometry/Mass Spectrometry
NIST	National Institute of Standards and Technology
PA	Proton Affinity
PBM	Probability-based Matching
PCB	Polychlorinated Biphenyl
PCI	Positive Chemical Ionization
PLOT	Porous Layer Open Tubular
ppb	Part-per-billion
ppt	Part-per-trillion
QA/QC	Quality Assurance/Quality Control
RF	Radiofrequency
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
STIRS	Self-training Interpretive and Retrieval System
SWIFT	Stored Waveform Inverse Fourier Transform
TIC	Total Ion Current
TOF	Time-of-flight
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
1-D	One-dimensional
3-D	Three-Dimensional

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Solid-phase Microextraction in Environmental Analysis • Soxhlet and Ultrasonic Extraction of Organics in Solids • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Trace Organic Analysis by Gas Chromatography with Selective Detectors • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 4)**
Field-based Analysis of Organic Vapors in Air

Field-portable Instrumentation cont'd **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Industrial Hygiene **(Volume 5)**
Chromatographic Techniques in Industrial Hygiene

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons: Gas Chromatography Procedures for On-line and Off-line Analysis • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction • Column Technology in Gas Chromatography • Data Reduction in Gas Chromatography • Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography • Liquid Phases for Gas Chromatography • Sample Preparation for Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. T. Hirschfeld, 'The Hyphenated Method', *Anal. Chem.*, **52**, 297A–303A (1980).
2. R.E. Finnigan, 'Quadrupole Mass Spectrometers', *Anal. Chem.*, **66**, 969A–975A (1994).
3. M.S. Story, 'The Development of Science and Technology for Quadrupole Mass Spectrometry', *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, 36, May 12–16, 1996, American Society for Mass Spectrometry, Santa Fe, NM.
4. T.Z. Chu, R. Finnigan, 'Thirty Years of a Mass Spectrometer Company', *Am. Lab.*, 90S–109S (October 1998).
5. G.A. Eiceman, H.H. Hill, J. Gardea-Torresdey, 'Gas Chromatography', *Anal. Chem.*, **70**, 321R–339R (1998).
6. J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippencott-Raven, New York, 1997.
7. J.R. Chapman, *Practical Organic Mass Spectrometry*, 2nd edition, Wiley, New York, 1993.
8. R.A. Johnstone, M.E. Rose, *Mass Spectrometry for Chemists and Biochemists*, 2nd edition, Cambridge University Press, Cambridge, 1996.
9. K.L. Busch, T.A. Lehman, *Guide to Mass Spectrometry*, VCH, Weinheim, 1996.
10. E. DeHoffmann, J. Charette, V. Stroobant, *Mass Spectrometry: Principals and Applications*, Wiley, New York, 1996.
11. F.G. Kitson, B.S. Larsen, C.N. McEwen, *Gas Chromatography and Mass Spectrometry*, Academic Press, San Diego, 1996.
12. W. Duenges, *Practical Introduction in GC/MS Analysis with Quadrupoles: Basics and Applications*, Huethig, Heidelburg, 1996.
13. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', *Anal. Chem.*, **70**, 647R–716R (1998).
14. Various authors, 'Applications Reviews', *Anal. Chem.*, **69**, 1R–328R (1997).
15. F. Wach, 'Benchtop GC/MS Instruments', *Anal. Chem.*, **66**, 927A–930A (1994).
16. M.J. Molina, F.S. Rowland, 'Stratospheric Sink of Chlorofluoromethanes: Chlorine Atom-catalyzed Destruction of Ozone', *Nature*, **249**, 810–814 (1974).
17. H.B. Singh, *Composition, Chemistry, and Climate of the Atmosphere*, Van Nostrand Reinhold, New York, 1995.
18. R.M. Riggin, W.T. Winberry, N.V. Tilley, *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, EPA document 600/4-84-041, Environmental Monitoring Systems Laboratory, Research Triangle Park, 1986.
19. *Methods for the Determination of Organic Compounds in Drinking Water*, EPA document 600/4-88/039, Environmental Monitoring Systems Laboratory, Research Triangle Park, 1991.
20. J. Pawliszyn, *Solid Phase Microextraction: Theory and Practice*, Wiley-VCH, New York, 1997.
21. K. Grob, 'Injection Technique in GC', *Anal. Chem.*, **66**, 1009A–1019A (1994).
22. H.G. Mol, M. Althuizen, H.G. Jenssen, C.A. Cramers, U.A. Brinkman, *J. High Resolut. Chromatogr.*, **19**, 69–79 (1996).
23. G.C. Stafford, P.E. Kelley, J.E.P. Syka, W.E. Reynolds, J.F.J. Todd, 'Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology', *Int. J. Mass Spectrom. Ion Processes*, **60**, 85–98 (1984).
24. D.W. Armstrong, G.L. Reid, J. Luong, *Curr. Sep.*, **15**, 5–11 (1996).
25. H. Yun, M.L. Lee, K.E. Markides, *J. Microcolumn Sep.*, **7**, 207–212 (1995).

26. D.W. Armstrong, G.L. Reid, M.P. Gasper, 'Halocarbon Separations of a New GSC-PLOT Column', *J. Microcolumn Sep.*, **8**, 83–87 (1996).
27. M.E. Hale, R.A. Yost, 'Compact Gas Chromatograph Probe for Gas Chromatography/Mass Spectrometry Utilizing Resistively Heated Aluminum-clad Capillary Columns, *Anal. Chem.*, **61**, 2410–2416 (1989).
28. W.H. McClennan, N.S. Arnold, K.A. Roberts, H.L. Meuzelaar, J.S. Lighty, E.R. Lindgren, 'Fast, Repetitive GC/MS Analysis of Thermally Desorbed Polycyclic Aromatic Hydrocarbons (PAHs) from Contaminated Soils', *Combust. Sci. Technol.*, **74**, 297–309 (1990).
29. N.S. Arnold, W.H. McClennan, H.L. Meuzelaar, 'Vapor Sampling Device for Direct Short Column Gas Chromatography/Mass Spectrometry Analyses of Atmospheric Vapors', *Anal. Chem.*, **63**, 299–304 (1991).
30. M. Klemp, A. Peters, R. Sacks, 'High Speed GC Analysis of VOCs: Sample Collection and Inlet Systems', *Environ. Sci. Technol.*, **28**, 369A–376A (1994).
31. R. Sacks, M. Akard, 'High Speed GC Analysis of VOCs: Tunable Selectivity and Column Selection', *Environ. Sci. Technol.*, **28**, 428A–433A (1994).
32. R.R. Reston, E.S. Kolesar, 'Silicon-micromachined Gas Chromatography System Used to Separate and Detect Ammonia and Nitrogen Dioxide – Part I: Design, Fabrication and Integration of the Gas Chromatography System', *J. Microelectromech. Syst.*, **3**, 134–146 (1994).
33. E.B. Overton, K.R. Carney, US Patent 95-563558, November 28, 1995.
34. E.U. Ehrmann, H.P. Dharmasena, K. Carney, E.B. Overton, 'Novel Column Heater for Fast Capillary Gas Chromatography', *J. Chromatogr. Sci.*, **34**, 533–539 (1996).
35. W.H. McClennen, N.S. Arnold, H.L.C. Meuzelaar, 'Field Portable Hyphenated Instrumentation – the Birth of the Tricorder?', *Trends Anal. Chem.*, **13**, 1–8 (1994).
36. W.H. McClennen, C.L. Vaughn, P.A. Cole, S.N. Sheya, D.J. Wager, T.J. Mott, J.P. Dworzanski, N.S. Arnold, H.L.C. Meuzelaar, 'Roving GC/MS: Mapping VOC Gradients and Trends in Space and Time', *Field Anal. Chem. Tech.*, **1**, 109–116 (1996).
37. K.J. Hart, S.A. McLuckey, G.L. Glish, 'Reaction of Analyte Ions with Neutral Chemical Ionization Gas', *J. Am. Soc. Mass Spectrom.*, **3**, 549–557 (1992).
38. A.G. Harrison, *Chemical Ionization Mass Spectrometry*, CRC Press, Boca Raton, 1992.
39. W.D. Lehman, 'Pictograms for Experimental Parameters in Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **8**, 756–759 (1997).
40. R.W. Kondrat, R.G. Cooks, 'Direct Analysis of Mixtures by Mass Spectrometry', *Anal. Chem.*, **50**, 81A–92A (1978).
41. D.L. Hunt, J. Shabanowitz, 'Scheme for Direct Analysis of Organics in the Environment by Tandem Mass Spectrometry', *Anal. Chem.*, **57**, 525–537 (1985).
42. D.E. Goeringer, K.G. Asano, S.A. McLuckey, D. Hoekman, S.W. Stiller, 'Filtered Noise Field Signals for Mass-selective Accumulation of Externally Formed Ions in a Quadrupole Ion Trap', *Anal. Chem.*, **66**, 313–318 (1994).
43. G. Wells, C. Huston, 'Field-modulated Selective Ion Storage in a Quadrupole Ion Trap', *J. Am. Soc. Mass Spectrom.*, **6**, 928–935 (1995).
44. M.H. Soni, R.G. Cooks, 'Selective Injection and Isolation of Ions in Quadrupole Ion Trap Mass Spectrometry Using Notched Waveforms Created Using the Inverse Fourier Transform', *Anal. Chem.*, **66**, 2488–2496 (1994).
45. S.A. McLuckey, D.E. Goeringer, G.L. Glish, 'Selective Ion Isolation/Rejection Over a Broad Mass Range in the Quadrupole Ion Trap', *J. Am. Soc. Mass Spectrom.*, **2**, 11–21 (1991).
46. C.E. Ardanaz, P. Traldi, U. Vettori, J. Kavka, F. Guidugli, 'The Ion Trap Mass Spectrometer in Ion Structure Studies. The Case of $[M - H]^+$ Ions from Chalcone', *Rapid Commun. Mass Spectrom.*, **5**, 5–10 (1991).
47. S. Schachterle, 'A Comparison of Excitation Methods for Collision Induced Dissociation in the Ion Trap Mass Spectrometer', *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics*, 714, May 29–June 3, 1994, American Society for Mass Spectrometry, Santa Fe, NM.
48. R.K. Julian, R.G. Cooks, 'Broad-band Excitation in the Quadrupole Ion Trap Mass Spectrometer Using Shaped Pulses Created by the Inverse Fourier Transform', *Anal. Chem.*, **65**, 1827–1833 (1993).
49. S.A. McLuckey, D.E. Goeringer, G.L. Glish, 'Collisional Activation with Random Noise in Ion Trap Mass Spectrometry', *Anal. Chem.*, **64**, 1455–1460 (1992).
50. J.V. Johnson, C. Basic, R.E. Pedder, B.L. Kleintop, R.A. Yost, 'Energetics and Collision-induced Dissociation Efficiencies in the Quadrupole Ion Trap Mass Spectrometer', in *Practical Aspects of Ion Trap Mass Spectrometry: Volume 3–Chemical, Environmental, and Biomedical Applications*, eds. R.E. March, J.F.L. Todd, CRC Press, Boca Raton, 257–312, 1995.
51. G. Baykut, B. Nolke, H.P. Vetters, G. Weiss, 'Quick Environmental Analysis Using a Mobile and Modular GC/MS System', *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics*, 283, May 29–June 3, 1994, American Society for Mass Spectrometry, Santa Fe, NM.
52. G. Baykut, G. Weiss, H.P. Brekenfeld, H.P. Vetters, G. Heinen, J. Franzen, 'Modular Inlet Systems and Sampling Devices for Mobile Environmental Mass Spectrometry', *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*, 645, May 31–June 5, 1992, American Society for Mass Spectrometry, Santa Fe, NM.
53. S. Billets, V.A. Ecker, D.J. Chaloud, 'Results of a Field Evaluation of the Bruker Mobile Mass Spectrometer

Under the SITE Demonstration Program', *Proceedings of the 39th ASMS Conference on Mass Spectrometry and Allied Topics*, 1651–1652, May 19–24, 1991, American Society for Mass Spectrometry, Santa Fe, NM.

54. P. Kowalski, F.H. Laukien, J. Wronka, 'Demonstration of Mobile Mass Spectrometry Through the EPA Superfund Innovative Technology Evaluation Program', *Proceedings of the 39th ASMS Conference on Mass Spectrometry and Allied Topics*, 665–666, May 19–24, 1991, American Society for Mass Spectrometry, Santa Fe, NM.
55. G. Xyrafas, A. Robbar, J. Wronka, 'Development of Methods for the Fast On-site Identification and Quantitation of Volatile and Semivolatile Pollutants Using a Fieldable Mass Spectrometer', *Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics*, 627–628, June 3–8, 1990, American Society for Mass Spectrometry, Santa Fe, NM.
56. W.D. Du, N.S. Arnold, P.A. Cole, B. Watteyne, D.T. Urban, H.L.C. Meuzelaar, 'The Next Horizon in Portable GC/MS for Air Monitoring Applications', *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*, 439, May 31–June 4, 1993, American Society for Mass Spectrometry, Santa Fe, NM.
57. N.S. Arnold, D.T. Urban, R.L. Watteyne, P.C. Cole, H.L.C. Meuzelaar, 'Remotely Operated, Field Portable GC/MS System for Monitoring Hazardous Atmospheric Vapors', *Proceedings of the 39th ASMS Conference on Mass Spectrometry and Allied Topics*, 671–672, May 19–24, 1991, American Society for Mass Spectrometry, Santa Fe, NM.
58. D.T. Urban, N.S. Arnold, H.L.C. Meuzelaar, 'Development of a Man-portable Transfer Line GC/MS System for Direct On-site Monitoring of Environmental Vapors', *Proceedings of the 38th ASMS Conference on Mass Spectrometry and Allied Topics*, 615–616, June 3–8, 1990, American Society for Mass Spectrometry, Santa Fe, NM.
59. M.D. Sinha, A.D. Tomassian, 'Miniature Gas Chromatograph/Mass Spectrometer', *NASA Tech Briefs*, **17**, 50 (1993).
60. L.I. Grace, D.M. Chambers, S.W. Thomas, B.D. Andresen, 'Progress in the Development of a Transportabler QUISTOR/reTOF-MS for Air Monitoring', *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, 139, May 12–16, 1996, American Society for Mass Spectrometry, Santa Fe, NM.
61. B.D. Andresen, J.D. Eckels, J.F. Kimmons, D.W. Myers, US Patent 5,525,799, June 11, 1996.
62. M.C. Prieto, D. Dietrich, B. Keville, D. Hopkins, 'Development of a Briefcase-size Ion Cyclotron Resonance Mass Spectrometer', *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, 474, June 1–5, 1997, American Society for Mass Spectrometry, Santa Fe, NM.
63. D. Dietrich, 'Mass Spectrometer is Lightweight and Portable', *R & D Magazine*, Cahners Publishing, Vol 10, 10, 1995.
64. R.E. Berkey, 'Field Deployable Monitors for Volatile Organic Compounds in Air', *Process Control and Quality*, **5**, 159–164 (1993).
65. C. Henry, 'Taking the Show on the Road: Portable GC and GC/MS', *Anal. Chem.*, **69**, 195A–200A (1997).
66. R.K. Boyd, J.D. Henion, M. Alexander, W.L. Budde, J.D. Gilbert, S.M. Musser, C. Palmer, E.K. Zurek, 'Mass Spectrometry and Good Laboratory Practices', *J. Am. Soc. Mass Spectrom.*, **7**, 211–218 (1996).
67. R. Baldwin, R.A. Bethem, R.K. Boyd, W.L. Budde, T. Cairns, R.D. Gibbons, J.D. Henion, M.A. Kaiser, D.L. Lewis, J.E. Matusik, J.A. Sphon, R.W. Stephany, R.K. Trubey, '1996 ASMS Workshop: Limits to Confirmation, Quantitation, and Detection', *J. Am. Soc. Mass Spectrom.*, **8**, 1180–1190 (1997).
68. F.W. McLafferty, D.B. Stauffer, *The Wiley/NBS Registry of Mass Spectral Data*, Wiley & Sons, NY, 1991.
69. S.R. Heller, G.W.A. Milne, L.H. Gevantman, *EPA/NIH Mass Spectral Database, Supplement 2, 1983*, National Bureau of Standards, US Government, Washington, DC, 1983.
70. H.L.J. Malkin, D.J.H. Trafford, *Mass Spectra and GC Data of Steroids: Androgens and Estrogens*, Wiley & Sons, Chichester, 1999.
71. R.P. Adams, *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*, Allured, Carol Springs, 1995.
72. K. Pfleger, W.W. Maurer, A. Weber, *Mass Spectral and GC Data of Drugs, Pollutants, Pesticides and Metabolites*, 2nd edition, VCH Publishers, NY, 1992.
73. R.A. Hites, *CRC Handbook of Mass Spectra of Environmental Contaminants*, 2nd edition, CRC Press, Boca Raton, 1992.
74. R.A. Sunshine, *CRC Handbook of Mass Spectra of Drugs*, CRC Press, Boca Raton, 1981.
75. B.S. Middleditch, S.R. Missler, H.B. Hines, *Mass Spectrometry of Priority Pollutants*, Plenum Publishing, NY, 1981.
76. E.A. Stemmler, R.A. Hites, *Electron Capture Negative Ion Mass Spectra of Environmental Contaminants and Related Compounds*, VCH Publishers, NY, 1988.
77. O.D. Sparkman, 'Evaluating Electron Ionization Mass Spectral Library Search Results', *J. Am. Soc. Mass Spectrom.*, **7**, 313–318 (1996).
78. R.K. Lindsay, G.B. Buchanan, E.A. Feigenbaum, J. Lederburg, *Applications of Artificial Intelligence for Organic Chemistry – The Dendral Project*, McGraw-Hill, NY, 1980.
79. A. Barr, E.A. Feigenbaum, *The Handbook of Artificial Intelligence*, Heuristech Press, Stanford, Volume II, 1982.

80. F.W. McLafferty, D.B. Stauffer, 'Retrieval and Interpretative Computer Programs for Mass Spectrometry', *J. Chem. Inf. Comput. Sci.*, **25**, 245–252 (1985).
81. J.W. Elling, S. Lahari, J.P. Luck, R.S. Roberts, S.I. Hruska, K.L. Adair, A.P. Levis, R.G. Timpany, J.J. Robinson, 'Hybrid Artificial Intelligence Tools for Assessing GC Data', *Anal. Chem.*, **69**, 410A–415A (1997).
82. R. Goodacre, D.B. Kell, 'Correction of Mass Spectral Drift Using Artificial Neural Networks', *Anal. Chem.*, **68**, 271–280 (1996).
83. B. Curry, D.E. Rumerlhart, 'MSnet: A Neural Network that Classifies Mass Spectra', *Tetrahedron Comput. Methodol.*, **3**, 213–237 (1990).
84. M. Sarker, W.G. Glen, L.B. Yin, W.J. Dunn, D.R. Scott, S. Swanson, 'Comparison of SIMCA Pattern Recognition and Library Search Identification of Hazardous Compounds from Mass Spectra', *Anal. Chim. Acta*, **257**, 229–238 (1992).
85. K. Varmuza, F. Stancl, H. Lohninger, W. Werther, *Chemom. Intell. Lab. Syst.*, **31**, 225–230 (1996).
86. C.E. Loope, C.R. Vargo, *Am. Environ. Lab.*, **7**, 50–53 (1995).
87. W. Guo, P. Zhu, H. Brodowsky, 'Study for Optimization of Chromatographic Conditions by Means of Artificial Neural Networks', *Talanta*, **44**, 1995–2001 (1997).
88. Q.S. Wang, C.S. Zhu, B.W. Yan, *Chin. J. Chem.*, **13**, 169–174 (1995).
89. R.G. Cooks, G.L. Glish, S.A. McLuckey, R.E. Kaiser, 'Ion Trap Mass Spectrometry', *Chem. Eng. News*, 26–41 (May 25, 1991).
90. P.T. Palmer, C.M. Wong, R.A. Yost, N.A. Yates, T.M. Griffin, 'Advanced Automation of Ion Trap Mass Spectrometry: New Opportunities for Real-time Autonomous Analysis', in *Computer-Assisted Analytical Spectroscopy*, ed. S.D. Brown, Wiley & Sons, NY, 25–60, 1996.
91. R.E. Kaiser, J.N. Louris, J.W. Amy, R.G. Cooks, 'Extending the Mass Range of the Quadrupole Ion Trap Using Axial Modulation', *Rapid Commun. Mass Spectrom.*, **3**, 225–229 (1989).
92. R.E. Kaiser, R.G. Cooks, G.C. Stafford, J.E. Syka, P.H. Hemberger, 'Operation of a Quadrupole Ion Trap Mass Spectrometer to Achieve High Mass/Charge Ratios', *Int. J. Mass Spectrom. Ion Processes*, **106**, 79–115 (1991).
93. R.E. March, F.A. Londry, G.J. Wells, 'Mass Resolution in a Quadrupole Ion Trap', *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*, 790, May 31–June 5, 1992, American Society for Mass Spectrometry, Santa Fe, NM.
94. M. Soni, V. Frankevich, M. Nappi, R.E. Santini, J.W. Amy, R.G. Cooks, 'Broad-band Fourier Transform Quadrupole Ion Trap Mass Spectrometry', *Anal. Chem.*, **68**, 3314–3320 (1996).
95. E.R. Badman, J.M. Wells, H.A. Bui, R.G. Cooks, 'Fourier Transform Detection in a Cylindrical Quadrupole Ion Trap', *Anal. Chem.*, **70**, 3545–3547 (1998).
96. S.A. Lammert, J.M. Wells, 'Experimental Investigations into the Performance of Ion Traps Using Air Versus Helium as the Buffer Gas', *Rapid. Commun. Mass Spectrom.*, **10**, 361–371 (1996).
97. W.R. Davidson, J.E. Fulford, 'The Creation of Mass Spectral Libraries for Quadrupole-based Mass Spectrometers', *Proceedings of the 31st ASMS Conference on Mass Spectrometry and Allied Topics*, 559, May 8–13, 1983, American Society for Mass Spectrometry, Santa Fe, NM.
98. F.W. McLafferty, A. Hirota, M.P. Barbalas, 'Library of Collisional Activation Mass Spectra', *Org. Mass Spectrom.*, **15**, 327–328 (1983).
99. R.I. Martinez, 'Instrument-independent CAD Spectral Databases: Absolute Cross-section Measurements in QQQ Instruments', *J. Res. Natl. Inst. Std. Technol.*, **94**, 229–237 (1989).
100. R.I. Martinez, 'Instrument-independent Tandem Mass Spectrometry Database for XQQ Instruments: The Dynamical Prerequisites', *Rapid. Commun. Mass Spectrom.*, **1**, 272–297 (1990).

High-resolution Mass Spectrometry and its Applications

Keith R. Jennings
University of Warwick, Coventry, UK

1 Introduction **1**
1.1 Definitions of Resolving Power and Resolution 1
1.2 Need for High-resolving Power 2
2 Instrumentation **3**
2.1 Magnetic Deflection Instruments 3
2.2 Quadrupole Mass Analyzer 5
2.3 Quadrupole Ion Trap 6
2.4 Time-of-flight Mass Analyzers 7
2.5 Fourier Transform Ion Cyclotron Resonance Instruments 10
3 Applications **12**
3.1 Applications Involving Determination of Elemental Composition 12
3.2 High-resolution Electrospray Ionization Mass Spectrometry 13
4 Concluding Remarks **14**
Abbreviations and Acronyms **14**
Related Articles **14**
References **14**

With the exception of the ^{12}C isotope of carbon which has a mass arbitrarily set equal to exactly 12 Da, all isotopes have unique nonintegral masses. The elemental composition of a sample can therefore often be inferred from an accurate measurement of the relative molecular mass of the sample. This requires that the mass resolving power (RP) of the instrument in use must be sufficiently high for the peak on which the mass measurement is based to be resolved from peaks arising from other components of closely similar mass. For samples of relative molecular mass below 500, an RP of 20 000 is often adequate, but for samples such as peptides and proteins an RP of several hundred thousand is often required.

The modes of operation, strengths and weaknesses of the several different types of mass analyzer currently available for the production of mass spectra of sufficiently high resolution are reviewed. Particular emphasis is placed on the use of mass spectrometry in the detection and identification of samples of biological origin for which its high sensitivity and specificity of information provided are of great importance. This article concludes with a brief review of the uses of high-resolution mass spectrometry, in which the difficulties encountered in working with samples of very high relative molecular mass are discussed and an account is presented of the information to be gained from time-dependent studies of H/D interchange at high resolution.

1 INTRODUCTION

1.1 Definitions of Resolving Power and Resolution

The *RP* of a mass spectrometer is a measure of the ability of the instrument to separate ions of closely similar mass-to-charge ratio, m/z, and to produce a spectrum of a given *resolution*. Because it is somewhat subjective to decide whether or not two overlapping peaks are just resolved, several different definitions of RP have been suggested and it is useful to consider three of the most commonly used.

The RP of magnetic deflection instruments is usually expressed in terms of the ''10% valley'' definition or the ''5% adjacent peak contribution'' definition which are equivalent. In such an instrument operated with narrow source and collector slits, peaks are triangular. They are arbitrarily said to be just resolved either when the valley between two peaks of equal height is 10% of the height of the peaks or when each peak contributes 5% of the peak height to the valley between the peaks, as shown in Figure 1.

The RP is then given by the nominal mass of the ions, m, divided by the difference in their masses, Δm. Since the possibility is very remote of finding two such peaks in a spectrum, it is normal to make use of a single peak. The peak due to the ion of mass m is observed on an oscilloscope screen alternately at full accelerating voltage V and at a lower accelerating voltage $V - \Delta V$. The lower voltage is chosen to move the peak along the mass scale by Δm such that when the peaks are viewed alternately, the valley between the peaks is 10% of the peak height. The magnitude of Δm is often expressed in parts per million, leading to an alternative definition of RP of 10^6 $(\Delta m/m)$.

The above definition is rarely used as a measure of the RP of other types of mass spectrometer since the peak shapes they produce are often closer to Lorentzian rather than triangular. The full width at half maximum (FWHM) definition is used. This is given by the mass of the ion, m, giving rise to the peak, divided by the full width of the peak in mass units at half the maximum height of the peak, Δm. Defined in this way, the RP $m/\Delta m$

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

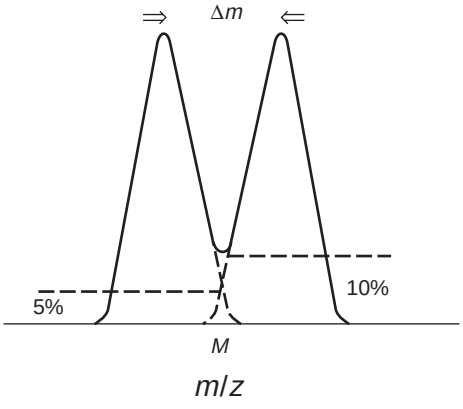

Figure 1 10% Valley (5% adjacent peak contribution) definition of RP normally used for spectra given by magnetic deflection instruments.

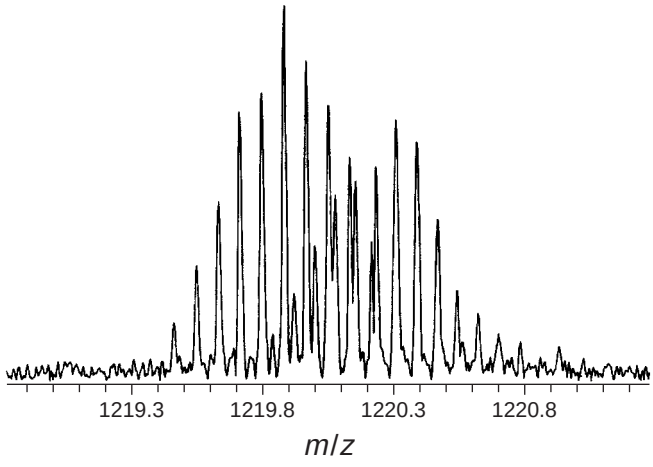

Figure 2 Small section of the high-resolution ESI spectrum of the B protein of soluble methane monooxygenase in the region of m/z 1220. (Reproduced by permission of H. Dalton, P.J. Derrick, A.J.R. Heck, K.R. Jennings, unpublished work.)

is approximately twice that based on the 10% valley definition. The FWHM definition has the advantage that it is largely independent of peak shape and it can be measured even when the baseline is noisy as a result of unresolved low intensity peaks arising from ''chemical noise''.

The terms ''low resolution'' and ''high resolution'' are somewhat subjective but low-resolution spectra are usually those obtained under conditions such that unit mass resolution is obtained up to about m/z 2500. Quadrupole instruments are unusual in that they are normally operated at unit mass RP, i.e. an RP of 500 at m/z 500, 1500 at m/z 1500, and so on. Since their upper limit for transmission of ions is about m/z 3000–4000, however, they are intrinsically low-RP instruments.

High-resolution spectra are usually obtained at an RP sufficiently high to resolve at least lower mass isobaric ions so that the masses of the components may be accurately measured. In some circumstances, therefore, a spectrum obtained at a resolution of 2000 may be considered a high-resolution spectrum for low mass species, but for high mass biological samples an RP in excess of 100 000 may be needed to solve a specific problem.

1.2 Need for High-resolving Power

At all but the lowest masses, there are often several ions of different isotopic composition that have the same nominal mass; these are known as isobaric ions. At low-RP, peaks due to isobaric ions are not resolved, so that the recorded peak is made up of contributions from two or more ionic species. If it is assumed that the peak is due only to one component and an attempt is made to measure the accurate mass of the ion giving rise to the peak, a weighted average that does not correspond to the mass of any single isotopic composition is obtained. Hence for accurate mass determinations, it is essential to work at an RP that separates all likely isobaric ions so that the measurement may be carried out on a peak due to a single component. For example, even at a nominal mass of 28 Da, isobaric ions commonly found are CO^+, N_2^+ and $C_2H_4^+$, the accurate masses of which are 27.9949, 28.0061 and 28.031 Da, respectively. The difference in mass, Δm, between CO^+ and N_2^+ is 0.0112 Da, so that the RP required to separate them is 28/(0.0112) or 2500. At higher masses, there are many more possible isobaric ions, but, in general, only a few are likely to be present in a given sample and so to interfere with the ion of interest. It is usual to work with an RP just sufficient to separate ions that may cause interference since working at a higher RP normally reduces sensitivity. At masses greater than about 500 Da, it becomes increasingly difficult to separate all possible isobaric ions, but, in practice, relatively few contaminants are likely to be present, so that an RP of 10 000–20 000 is often adequate. For very complex systems, however, an RP of several hundred thousand may be necessary in order to ensure that the peak of interest is resolved. An example of a high-resolution electrospray ionization (ESI) spectrum of the B protein of methane monooxygenase is illustrated in Figure 2. This shows overlapping sets of isotope peaks of two multiply charged ions which in a simple quadrupole instrument would appear as a single peak at m/z 1220.

When the molecular weight of a sample exceeds the RP of a mass spectrometer at that value of m/z, it is no longer possible to resolve individual isotope peaks such as those due to the presence of varying numbers of ^{13}C, ^{15}N and ^{34}S isotopes. The mass of

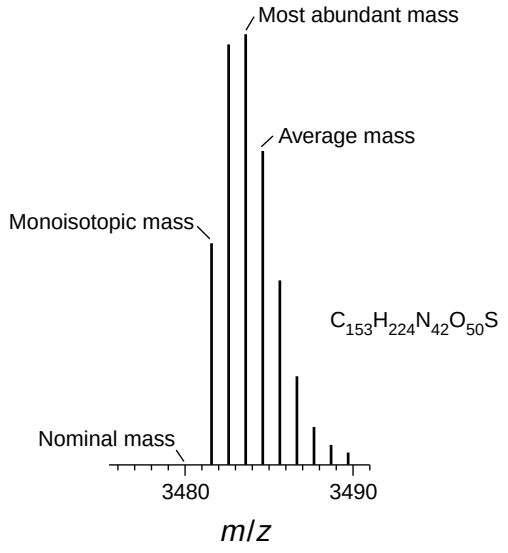

Figure 3 Nominal, monoisotopic and average molecular masses. (Adapted by permission of the American Chemical Society, from J. Yergy, D. Heller, R.J. Cotter, C. Fenselau, *Anal. Chem.*, **55**, 353 (1983).)

molecular ions derived from measurements on unresolved peaks is therefore the *average* molecular mass rather than the *monoisotopic* molecular mass derived from measurements at high RP. It is important to make this distinction since for high-mass biological samples, the difference between these two masses may be several daltons. It is also important to note that the mass of the monoisotopic peak is significantly higher than the nominal mass for such samples because of the high number of hydrogen atoms present (H = 1.007825). For example, the nominal molecular mass of human insulin is 5802 Da, the monoisotopic molecular mass is 5804.6 Da, and the average mass is 5808.6 Da. Another example of this is illustrated in Figure 3.

A particular problem for which high RP is essential is the study of the incorporation of deuterium atoms by H/D exchange into proteins. If, for example, a protein of molecular weight 50 000 Da incorporates a single deuterium atom, thereby increasing its molecular weight by 1 Da, a RP of 50 000 is required to resolve the peaks of the deuterated and undeuterated ions even if the ion is multiply charged as in ESI. In general, the number of hydrogen atoms that exchange varies, so that a series of peaks results and their resolution requires an RP equal to the molecular weight of the protein. In the ESI spectra of many proteins, the many possible isotope peaks are not resolved in a low-resolution spectrum because if the ions carry n charges, the peaks are separated by only $1/n$ of a unit on the m/z scale. This is discussed more fully in a later section.

2 INSTRUMENTATION

2.1 Magnetic Deflection Instruments

Ions of mass-to-charge ratio m/z, where z is the number of electronic charges e, when subjected to an accelerating voltage V acquire a velocity v, where

$$Vze = \frac{mv^2}{2} \tag{1}$$

When they enter a magnetic field of strength B perpendicular to the direction of the ion beam, the ions describe a circular path of radius r such that

$$Bzev = \frac{mv^2}{r} \quad \text{or} \quad r = \frac{mv}{Bze} \tag{2}$$

For a constant value of B, the radius described is therefore proportional to the momentum of the ions. Combining Equations (1) and (2) gives the normal Equation (3) for a magnetic deflection instrument:

$$\frac{m}{z} = \frac{B^2r^2e}{2V} \tag{3}$$

In practice, whichever ionization method is employed, a beam of ions emanating from the source slit in the xz plane which have a range of energies also have an angular divergence in the y direction, normally given the value of 2α at the source slit. A magnetic sector is described as ''single focusing'' because it focuses a monoenergetic beam of ions of a given mass diverging from the source slit and so gives ''direction focusing''. It does not, however, focus ions at a given point if they have different energies (i.e. velocities), so it does not provide ''velocity focusing'' and its RP is therefore reduced. High-resolution magnetic sector instruments employ an electric sector to reduce the energy spread because it brings to focus a diverging monoenergetic beam of ions. The focusing properties of magnetic and electric sectors are illustrated in Figure 4. Certain geometrical arrangements of an electric and a magnetic sector are termed ''double focusing'' in that the energy dispersion of one sector is equal and opposite to that of the other and direction focusing is maintained throughout the instrument. These are usually fitted with variable α- and β-slits to allow the angular divergence of the beam and the energy bandpass of the instrument to be varied. Two of the most popular geometries are based on variations of those known as the Nier–Johnson[1] and Mattauch–Herzog[2,3] geometries and these have formed the basis of a number of commercial instruments. They are illustrated in Figure 5.

The performance of double-focusing instruments is determined primarily by the size of aberrations from second-order terms that arise from such factors as field inhomogeneities and fringing fields. These aberrations

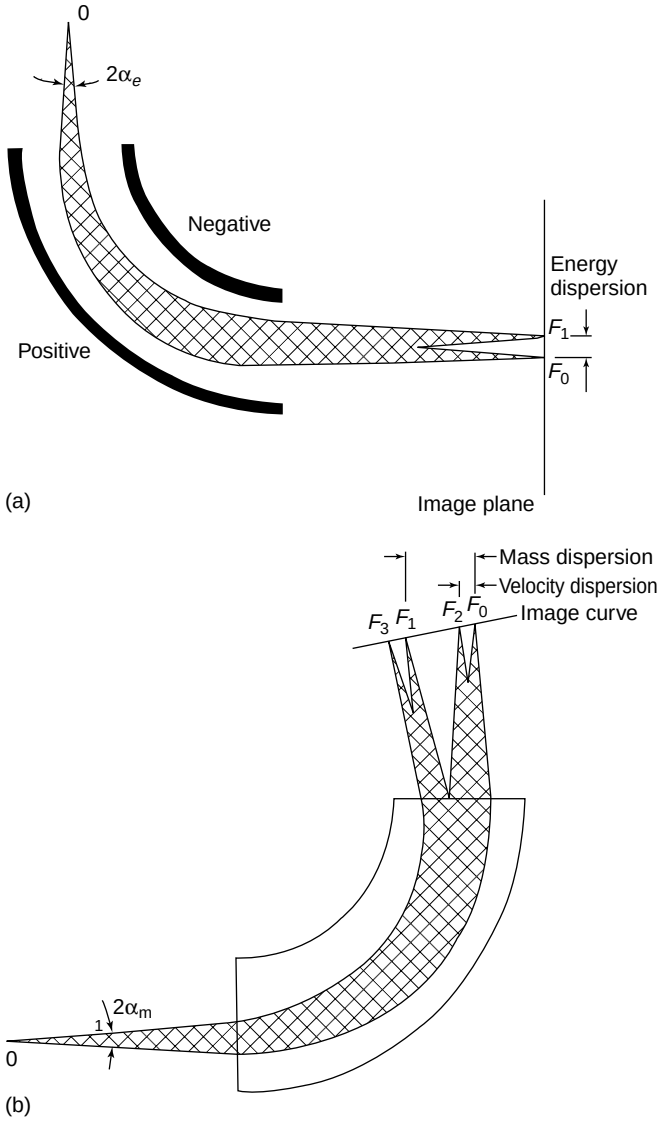

Figure 4 Focusing of ion beams by (a) electric and (b) magnetic sectors. (Reproduced by permission of Academic Press, from Jennings and Dolnikowski.[5])

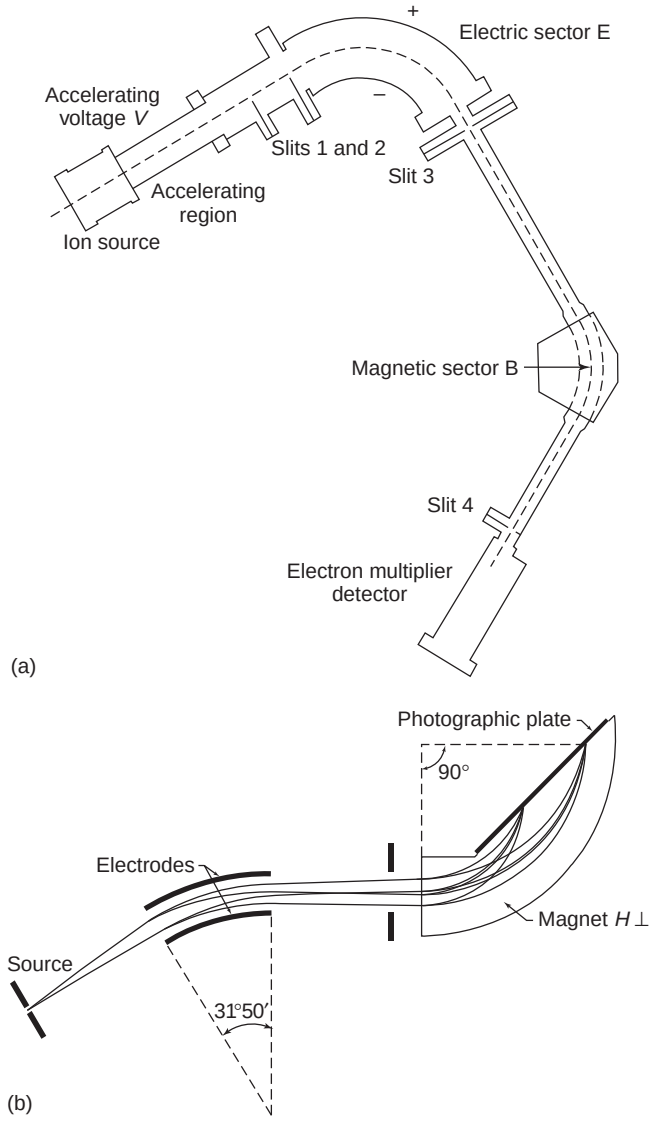

Figure 5 (a) Nier–Johnson geometry, (b) Mattauch–Herzog geometry. ((a) Adapted with permission from H.A. Strobel, W.R. Heineman, *Chemical Instrumentation: A Systematic Approach*, 3rd edition, Wiley-Interscience, New York, 689, 1989; (b) Reproduced by permission of the McGraw-Hill Companies, from C.A. McDowell, ed., *Mass Spectrometry*, McGraw-Hill, New York, 19, 1963.)

are normally reduced by using electrostatic lenses to optimize performance, e.g. hexapole lenses to correct for image rotation and curvature. High RP is maintained, especially if the collector slit length is reduced. In some cases, further improvement in performance has been obtained by the use of non-normal entry into the magnetic sector to provide some focusing in the z direction (the direction of the magnetic field). The use of tapered pole faces that produce small radial field gradients has a focusing action in one plane and divergence in another, but when the magnetic sector is made up of several segments there is an overall focusing effect. This improves the high-mass performance without increasing the size of the instrument. Such instruments are capable of RPs in excess of 100 000 and until the development of commercial Fourier transform ion cyclotron resonance (FTICR) instruments, double-focusing magnetic deflection instruments provided the highest RP commercially available.[4]

Two-sector instruments have been widely used for tandem mass spectrometry work, either in the investigation of products of metastable transitions or of collision-induced decomposition of ions. By making use of a scan of the accelerating voltage V at constant E and B in a

forward geometry instrument (EB), all precursors of a given product ion are observed. Similarly, in a reverse geometry instrument (BE), a scan of E at constant V and B gives a spectrum of all products of a given precursor ion. The release of translational energy in the fragmentation reactions degrades the resolution of such spectra and the RP attainable is often only 100–200. Higher RP of the product ions is obtained by use of the linked scan in which V is held constant and B and E are scanned simultaneously such that the ratio B/E has a constant value determined by the mass of the precursor ion. In a typical 30-cm radius double-focusing instrument, the product ion resolution is about 1000 but the resolution for precursor ion selection is only in the region of 350.[5]

Higher resolution tandem mass spectra require the use of a multisector instrument such as a four-sector instrument of either EBEB or BEBE geometry.[6] In principle, both the first and second double-focusing instrument can be operated at high RP, so that both the precursor and product ions can be selected precisely. In practice, one or other of the analyzers is operated at a RP that is just sufficient for the problem under investigation because the sensitivity of the experiment is otherwise very low. Since it is usually the precursor ion that is selected at high RP, hybrid tandem instruments have been used in which the second analyzer is a quadrupole analyzer with an RP of about 1500.[7]

2.2 Quadrupole Mass Analyzer

The quadrupole mass analyzer consists of four parallel rods, separation $2r_0$, of hyperbolic cross-section, though properly spaced rods of circular cross-section give a similar performance.[8] Opposite pairs of rods are connected electrically and a direct current (DC) voltage U and radiofrequency (RF) voltage $V_0 \cos(wt)$ are applied to the rods. The transmission characteristics are normally described in terms of the two parameters, a and q where, Equations (4–6)

$$a = \frac{8eU}{mr_0^2w^2} \tag{4}$$

$$q = \frac{4eV_0}{mr_0^2w^2} \tag{5}$$

so that

$$\frac{a}{q} = \frac{2U}{V_0} \tag{6}$$

For ions of a given m/z, certain values of a and q lead to stable oscillations so that these ions are transmitted; all others are lost to the rods or the walls of the vacuum system. This is illustrated in Figure 6,[9] which shows that if q is greater than 0.91, there are no stable trajectories and no ions are transmitted.

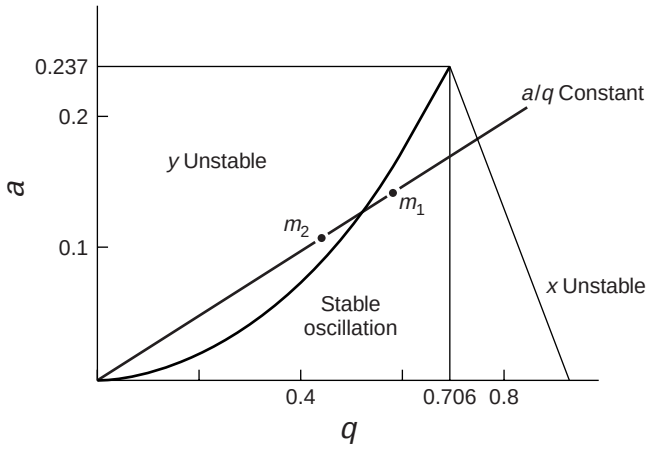

Figure 6 Quadrupole stability diagram illustrating conditions for stable and unstable oscillations. (Reproduced with permission of J. Wiley & Sons, Inc., from J.R. Chapman, *Practical Organic Mass Spectrometry*, 2nd edition, J. Wiley & Sons, New York, 10, 1983.)

Optimum resolution is obtained when $a/q = 0.167$ and ions of a given m/z are transmitted when

$$\frac{m}{z} = \frac{0.136V_0}{r_0^2f^2} \tag{7}$$

where V_0 is in volts, r_0 in centimeters, and $f(= w/2\pi)$ is in megahertz. The instrument is usually operated at unit mass RP, i.e., throughout the scan, ions of mass m are resolved from those of mass $m+1$ so that the RP is equal to m. The increase in RP with increase in m/z is due to the fact that the heavier ions move more slowly through the analyzer and so experience more oscillations than the lighter ions. Typically, $r_0 = 0.4$ cm and $f = 1$ MHz, so that for $m/z = 10\,000$, and hence for an RP of 10 000, $V_0 = 11.76$ kV, which is usually considered to be impracticable. In practice, the effects of fringing fields become greater as the value of m/z rises, since the time spent in these fields is proportional to $m^{1/2}$ and the transmission falls rapidly as m/z increases above about 1500. These effects are reduced by the use of a circular aperture at the entrance to the analyzer, so reducing radial divergence of the ions, and by the use of short RF-only pre- and postfilters to steer ions through the fringing fields. More generally, the RP is given by Equation (8):

$$\text{RP} = \left(\frac{\text{length of quadrupole field}}{\text{radius}}\right)^2 \left(\frac{V_0}{E_{\text{tr}}}\right. \tag{8}$$

where E_{tr} is the injection energy of the ions. If the upper practical limit of V_0 has been reached, a reduction in E_{tr} will increase the RP but at the cost of reducing transmission due to increased losses arising from radial translational energy.

Similarly, lengthening the quadrupole field is of limited value because of imperfections in the quadrupolar field leading to increased ion losses. Although the performance of commercial versions of these instruments has improved greatly during the 1980s and 1990s, the RP obtainable from a quadrupole mass analyzer rarely exceeds 2000. Higher RP can be obtained by reducing the operating frequency from the usual value of 1.2–1.5 MHz to a much lower value. Operating at 683 kHz, the upper mass range was extended to a little over 8500 at unit mass resolution and by working at fairly high pressure in the quadrupole collision cell collisional focusing resulted in there being only a small reduction in transmission.[10]

2.3 Quadrupole Ion Trap

The first description of the ion trap was by Paul in 1956[11] but a commercial version was not available until nearly thirty years later when it was marketed by Finnigan-MAT as the ion trap detector. Ion traps were initially developed as low-resolution low-mass detectors for use with gas chromatographs but since their introduction their performance has been improved considerably and several commercial versions are now available. As its name suggests, ions are stored within the trap and detection is accomplished by ejecting ions of a given mass-to-charge ratio to strike a conversion dynode before amplification of the signal by an electron multiplier. A mass spectrum is obtained by scanning an RF voltage as described below but ions which are not being ejected are stored rather than being lost as in the scanning of a magnetic sector or quadrupole instrument. A general introduction to the properties of ion traps has been given[12] and an account of instrumental developments is contained in March.[13]

Although the ion trap is related to a quadrupole mass filter, the two instruments have very different properties and modes of operation. The ion trap can be considered to be generated by rotating the hyperbolic rod electrodes of a quadrupole analyzer about an axis perpendicular to the z-axis (the direction in which the ions travel) and passing through a point half way along the length of two opposing rods. The rods through which the axis passes form circular endcap electrodes and the other pair join up to form a doughnut-shaped ring electrode, as illustrated in Figure 7(a). The system is axially symmetrical and, ideally, the surfaces should be hyperbolic for maximum performance. Unlike the quadrupole analyzer, the ion trap is able to confine ions for long periods in a small volume close to the center of the trap. Ions may be generated within the trap or injected through an endcap electrode, after which they may undergo reactions prior to mass analysis.

There are several methods of operating an ion trap; the type and magnitude of the potentials applied to the

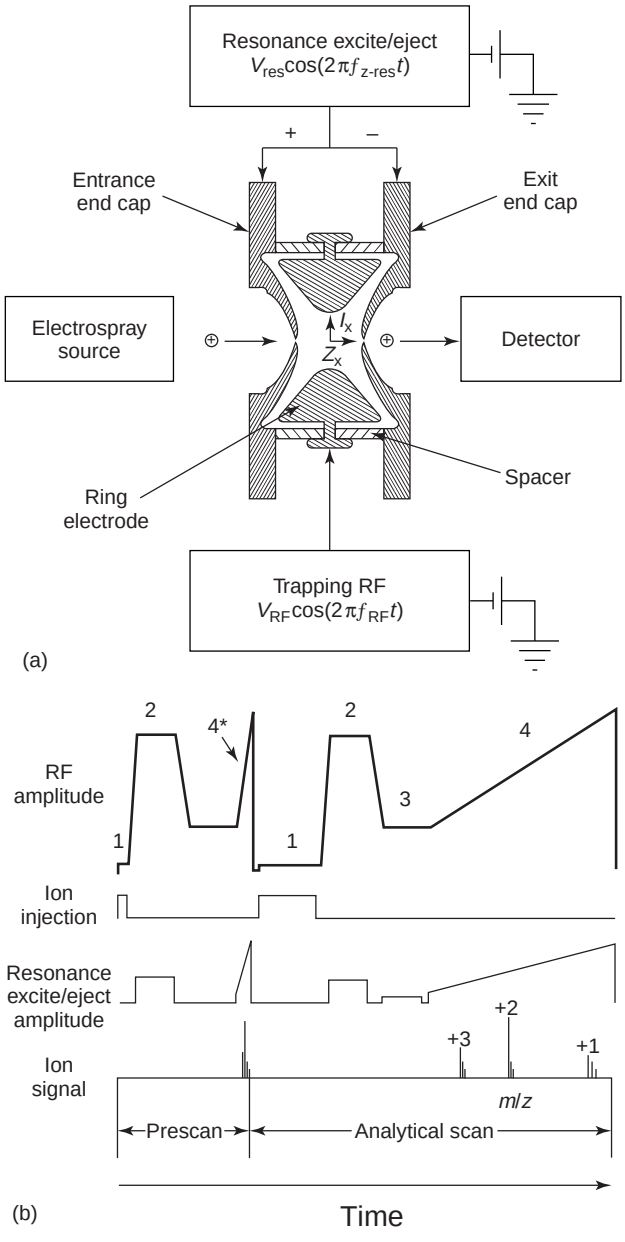

Figure 7 (a) Quadrupolar ion trap and (b) typical scanning sequence; 1 ion injection, 2 isolation, 3 excitation/reaction, 4 mass analysis. (Reproduced with permission of J. Wiley & Sons, Inc., from R.B. Cole, ed., *Electrospray Ionisation Mass Spectrometry*, J. Wiley & Sons, New York, Chap. 7, 1997.)

electrodes vary with the type of experiment to be carried out. When used to produce a conventional analytical mass spectrum, a fixed RF voltage of frequency about 1 MHz with no DC voltage is applied to the ring electrode. All ions above a particular minimum m/z, determined by the amplitude of the RF voltage, are stored and in principle there is no upper mass limit to the ions stored. If the amplitude of the RF voltage is increased, the low mass cut-off increases and ions are ejected in sequence through

an endcap electrode and detected externally. The upper mass limit is controlled by the maximum RF voltage that can be applied (typically 6 kV, peak-to-peak) and at the end of the voltage scan the RF voltage is removed and all remaining ions are lost to the walls or are swept out of the trap with the bath gas. The process is then repeated. This mode of operation is termed mass-selective axial instability ejection. Improved performance is obtained when a pressure of about 10^{-3} torr of helium is present in the ion trap, collisionally cooling the ions and forcing them into the center of the trap.

A further increase in RP is obtained by applying to the endcap electrodes a very low amplitude (e.g. 6 V, peak-to-peak) RF voltage of about half the frequency of the main RF voltage during the period when the latter is being ramped to produce an analytical scan. This results in ions of a particular m/z being removed from the main cloud of ions immediately prior to their being ejected. Free from space charge effects, they are concentrated into a very tight bunch of ions just before ejection.

If a DC voltage is applied to the endcap electrodes together with the RF voltage applied to the ring electrode, the ion trap now has both a low-mass and a high-mass cut-off. By careful control of the voltages, a small mass range of ions or ions of a chosen m/z can be isolated in the trap prior to its reactions being studied. Several alternative methods of controlling ion motion within the trap are based on using the main RF voltage on the ring electrode to confine the ions and an additional very low amplitude RF voltage, known as the ''tickle voltage'', being applied to the endcap electrodes. If a single frequency or band of frequencies is applied and the duration and amplitude of the tickle voltage is controlled, resonant excitation of one or more ions over a chosen mass range may be effected, leading either to collision-induced decomposition of the ions or to their selective ejection. This may be used either to generate a mass spectrum or to remove unwanted ions prior to studying the reactions of those ions remaining in the trap.

The equations of motion for ions in the ion trap are similar to those that describe the motion of ions in a quadrupole mass analyzer. The parameters a_z and q_z are defined in a similar manner for motion in the z, or axial, direction and a plot of a_z against q_z again produces regions of stability and instability, Equations (9) and (10):

$$a_z = \frac{16eU}{[m(r_0^2 + 2z_0^2)w^2]} \tag{9}$$

$$q_z = \frac{8eV_0}{[m(r_0^2 + 2z_0^2)w^2]} \tag{10}$$

where the symbols are as defined previously. A typical ion trap and scanning sequence is illustrated in Figure 7.

The complex equations of motion are derived for a pure quadrupolar trapping field, but in practice better performance is achieved by the use of ''stretched'' geometry in which the separation of the end caps is about 11% greater than that calculated based on these equations.[14] This modification of the geometry causes a significant shift in the resonant frequency for ions of a given m/z but this is readily accounted for by the dependence of the frequency on the value of r_0.

Typical commercial ion trap instruments have RPs similar to or somewhat higher than those of quadrupole instruments. When operated in the mass-selective axial instability mode with axial modulation, the mass resolution obtainable for singly charged ions up to m/z 650 is approximately $3 \times m$. Most instruments use an RF amplitude scanning rate of 5555 units of m/z per second, which is equivalent to approximately 186 μs per unit of m/z. If the rate of scanning the RF voltage is considerably reduced, e.g. by a factor of 10^4 or 10^5, very high RPs can be attained.[15] For example, when the rate of scanning of a standard commercial instrument was reduced from 5555 to 0.80 units of m/z per second, a mass RP in excess of a million was obtained at m/z 615, and even higher RPs have been obtained at still lower scanning speeds.[13] More usefully, rather smaller reductions in the scanning rate can give very useful increases in RP. Accurate mass measurements at high resolution are still difficult because of variation of conditions in the ion trap causing multiple scans to give individual mass measurements of an ion spanning 0.2–0.3 Da.[16] The accuracy can be improved considerably if (i) there is a phase-locked relationship between the main RF field and that of the axial modulation field[17] and (ii) line-locking is used to minimize the noise level on the main RF field.[18] When both of these were employed, the accuracy of mass measurement was better than 5 ppm at m/z 264 and mass RPs in excess of 50 000 were achieved. It seems probable that RPs considerably above 2000 will be more common when one or more of the above techniques are employed more widely.

2.4 Time-of-flight Mass Analyzers

In principle, the time-of-flight mass analyzer is a very simple device. Ions at rest are subjected to a very short high voltage pulse of V volts that sends the ions down a tube a distance D to an electron multiplier with very high time resolution. Alternatively, since time-of-flight instruments are now used primarily with pulsed matrix-assisted laser desorption ionization (MALDI) sources, the accelerating voltage V may be applied continuously as the pulses of ions are generated by the pulsed laser, the pulse width of which is typically less than a few nanoseconds. Reviews of developments in time-of-flight instrumentation have been published.[19–21]

Whichever method is employed, all singly charged ions have, in principle, the same energy, Ve, so that from Equation (1), ions of mass m have a velocity v given by Equation (11):

$$v = \left(\frac{2Ve}{m}\right)^{1/2} \tag{11}$$

Hence, for ions with masses m_1 and m_2 having velocities of v_1 and v_2,

$$\frac{m_1}{m_2} = \left(\frac{v_2}{v_1}\right)^{1/2} \tag{12}$$

in other words v is proportional to $1/m^{1/2}$. By recording ion current as a function of time, a mass spectrum is obtained because the times t_1 and t_2 taken by ions of masses m_1 and m_2 to reach the detector are given by Equation (13):

$$\frac{t_1}{t_2} = \frac{v_2}{v_1} = \left(\frac{m_1}{m_2}\right)^{1/2} \tag{13}$$

From this equation, it is clear that $m = at^2$, so that $\mathrm{d}m = 2at$ and the RP is given by Equation (14):

$$\frac{m}{\mathrm{d}m} = \frac{t}{2\,\mathrm{d}t} \tag{14}$$

The RP may therefore be increased by increasing t, e.g. by increasing D or by reducing V. Although a high value of V is usually required for reasonable sensitivity at higher mass, it leads to a marked increase in the energy spread of the ions caused by collisional processes that occur in the expanding plume of ions. It also increases the apparent spread in the times of formation of the ions. A reduction in $\mathrm{d}t$, the minimum time interval that can be measured at the detector, also improves RP. If ions of mass m (Da) and charge ze are subjected to an accelerating voltage of V (volts), the velocity of the ions in meters per second is given by Equation (15):

$$v = 1.39 \times 10^4 \left(\frac{zV}{m}\right)^{1/2} \tag{15}$$

If $V = 20\,000\,\mathrm{V}$ and $m = 5000\,\mathrm{Da}$, $v = 2.78 \times 10^4\,\mathrm{m\,s^{-1}}$ for a singly charged ion, so that for a flight path of 1.25 m, $t = 4.5 \times 10^{-5}\,\mathrm{s}$. If the minimum time interval that can be measured is 1 ns, the RP attainable is, in principle, 22 500, but in practice the performance is substantially below this for a variety of reasons.

Implicit in the above description is the assumption that all ions are formed at rest within a plane perpendicular to their direction of flight down the analyzer tube. These conditions are not fulfilled and ions leave the source with a spread of energies. If a large number of ions is formed, space charge effects cause the ion cloud to expand and increase the range of velocities of ions within the cloud, so that RP falls as laser power increases. As a result, the RP of a simple time-of-flight instrument is typically no more than 400 at m/z 1500 and decreases rapidly at higher masses. The time resolution of the detector can be quite modest, e.g. 100 MHz, i.e. $\mathrm{d}t$ is 10 ns, in such instruments.

The earliest approach adopted to improve performance was the use of a ''reflectron'', a repulsive electrostatic field, often known as an electrostatic or ion mirror, that reverses the direction of motion of the ions so that they are collected close to the source, as illustrated in Figure 8.[22,23] A single uniform reflecting field gives only first-order time focusing of the energy spread of the ions. Improved performance is obtained by the use of a double reflecting field in which ions are retarded in the first field and reflected in the second field, enabling second order focusing to be achieved. The mirror may consist of one or two high transmission grids to produce a single- or two-stage reflector together with a series of annular concentric plates. In order to avoid edge effects and other grid-induced field distortions, the field is gridless, being defined simply by the annular plates through which the ions may pass. This essentially doubles the flight path, D, which is advantageous, but the main function of the reflectron is to try to correct the effect of a spread of translational energy that arises during the ionization process by ''bunching'' of ions of the same m/z ratio. Faster ions penetrate further than ions of average velocity into the electrostatic field so that they follow a slightly longer flight path in travelling from the source to the detector. Conversely, ions of below average velocity penetrate less far into the field and so travel a shorter path length. Typically, the use of a reflectron increases the RP of an instrument which has a 1-m flight path from

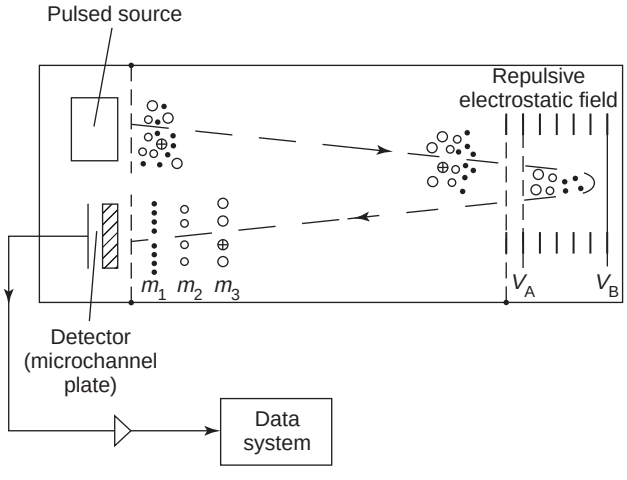

Figure 8 Schematic of a typical reflectron time-of-flight instrument. (Adapted by permission of Elsevier Scientific Publications, from B.A. Mamyrin, *Int. J. Mass Spectrom. Ion Processes*, **131**, 1 (1994).)

about 400 to about 2000 at m/z 1500. Although this is a very useful increase in RP, neither the reflectron nor a high accelerating voltage can correct for ions formed at different times, e.g. in chemical reactions in the ablation plume.

Much greater improvements in RP can be obtained, however, if the method of ion extraction into the flight tube is improved. There have been two developments in source design that have tackled the problem of ion production in the source. Instead of the use of continuous extraction of ions from the MALDI source, a number of related techniques, known variously as time-lag focusing, pulsed delayed extraction, post source pulsed focusing, and so on have been described (see for example Vestal et al.[24]) and which resemble the technique first described by Wiley and McLaren in 1955.[25] Source designs based on these have been incorporated into commercial instruments since about 1997 onwards. In one version of time-lag focusing,[26] a voltage pulse of 2–3 kV that is a few nanoseconds wide is applied to extract ions formed on the source repeller plate from the source, and after a certain time lag or delay they encounter the full accelerating voltage of 20–25 kV as shown in Figure 9.

During the delay, newly formed ions of a given m/z are repelled from the repeller plate and from each other, so that they travel away from the plate into the field-free region with a broad distribution of velocities. With the repeller plate and first focus plate held at the same potential, the ions with the higher axial velocities travel further from the repeller plate than those ions with lower velocities, creating a distribution of ions within the field-free region. After a delay of about 0.5–5 μs, the short high-voltage extraction pulse is applied to the source repeller plate; more energy is imparted to ions close to the repeller plate than to those which are closer to the focus plate. The amplitude of the pulse or the delay time is adjusted so that ions close to the repeller catch up with the ions nearer the focus plate as they traverse the flight tube so that they reach the detector at the same time. These adjustments are m/z dependent and higher voltages and longer delay times are required to focus ions of higher m/z.

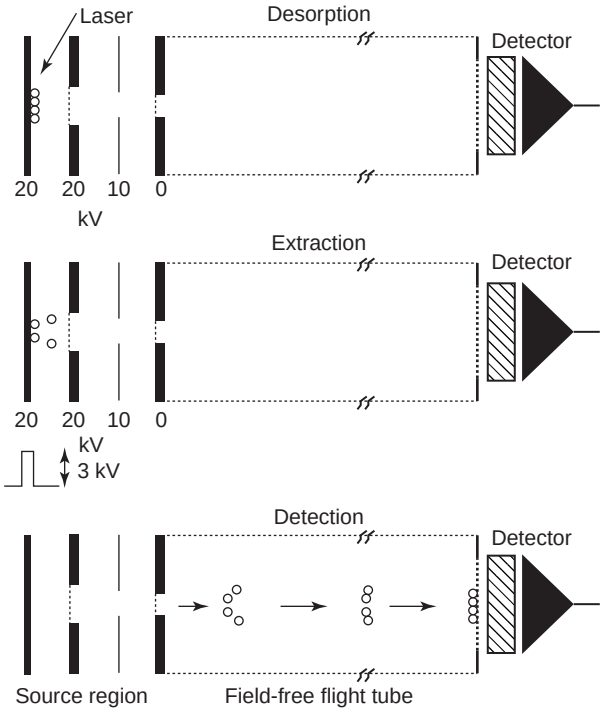

Figure 9 Illustration of the principle of time-lag focusing. (Reproduced with permission of International Scientific Publications from Whittal and Li.[26])

Linear instruments of this type with a flight path of about 1 m are capable of resolving carbon isotope peaks at m/z 3000. Although the RP decreases with increase in m/z, some commercial instruments with longer flight paths can attain an RP of about 5000 at m/z 2500. When combined with a reflectron, which gives further focusing of ions of a given m/z, RPs in excess of 10 000 are obtained up to m/z 10 000. To obtain the best results from such instruments, a very fast response detector is required with a minimum measurable time difference of 1 ns or less and variations in the precise times of firing the laser and starting data collection (''jitter'') must be minimized. Although the focusing achieved is m/z-dependent, a broad mass range of ions can be observed at higher RP and the source can be adjusted to give optimum RP at the ions of the particular m/z of interest. An added advantage is that the accuracy of mass measurement is also improved and errors as low as 100 ppm up to m/z 30 000 Da are attainable. An example of a spectrum obtained by the use of this type of instrumentation is shown in Figure 10.

An alternative technique is to use orthogonal extraction of ions from the source.[27–30] This is particularly useful when the time-of-flight analyzer is the second mass analyzer in a tandem instrument for which time-lag focusing is less easy to use.[31] Even after collimation, the energy spread of the ion beam in the direction of the beam emanating from the collision cell is large but the energy spread in the plane perpendicular to the beam is very small. Ions pass into the source of a time-of-flight instrument and are extracted into the flight tube by means of a large voltage pulse applied perpendicular to the direction of the beam. When used in conjunction with a reflectron, instruments of this type can give mass RPs in excess of 10 000 at m/z 2500 and because of their high duty cycle they are intrinsically much more sensitive than a scanning instrument. Although this is well below the highest RP given by magnetic sector and FTICR instruments, the high sensitivity and very high m/z range of a time-of-flight

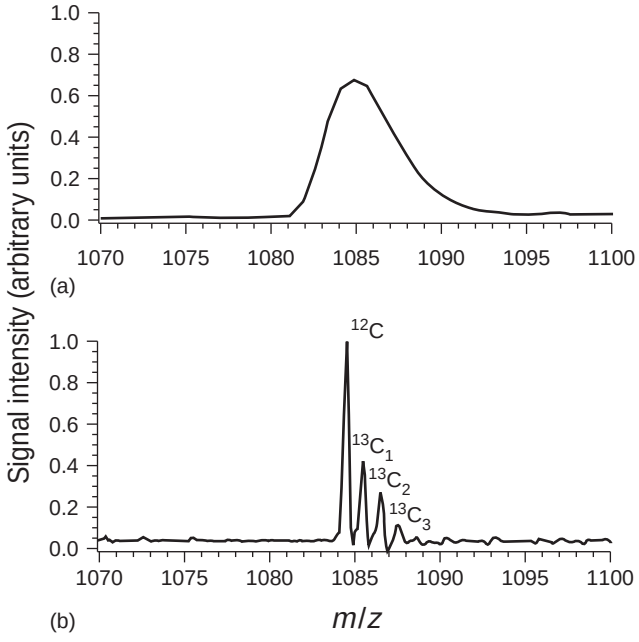

Figure 10 Illustration of the effectiveness of time-lag focusing coupled with the use of a reflectron; (a) a spectrum obtained from a linear instrument without these features; (b) a spectrum from the same sample obtained from an instrument incorporating both features. (Reproduced with permission of International Scientific Publications from Whittal and Li.[26])

instrument makes this a very useful RP when working with samples that give ions above m/z 10 000.

When a reflectron time-of-flight instrument is used to study products of ions that decompose in the flight tube after acceleration from the source (post-source decay), the resolution with which products may be observed is good only over a limited mass range. This is because the range of energies with which the ions are formed causes them to be focused at different points in the flight tube. A full spectrum of products is usually obtained by looking at several sections of the mass range in turn and then combining them to form a complete spectrum. In a tandem ToF/ToF instrument, the first reflectron improves precursor mass resolution and much higher resolution of product ions over a much wider mass range may be achieved by the use of a curved reflecting field.[32]

2.5 Fourier Transform Ion Cyclotron Resonance Instruments

The technique of ion cyclotron resonance (ICR) mass spectrometry was introduced in the 1960s in the form of a low-resolution device and was used primarily to investigate the mechanisms of ion–molecule reactions of low mass species. FTICR mass spectrometry was first described by Comisarow and Marshall in 1974[33,34] and since then it has been developed into a technique that is capable of very high-mass resolution, far outstripping that obtainable by any other type of mass analyzer.[35,36]

Cells of various geometries have been described but the principles of operation of an ICR instrument will be illustrated by reference to the operation of the cubic cell as illustrated in Figure 11.[37]

Detection of ions of mass-to-charge ratio m/z is based on the absorption of RF power at the ICR frequency, w_c (radians s^{-1}), which is approximately given by Equation (16):

$$w_c = \frac{zeB}{m} \tag{16}$$

where z is the number of electronic charges e, and e is expressed in coulombs, B is the magnetic induction in tesla and m is the mass of the ion in kilograms. Since w_c can be measured with an accuracy of about

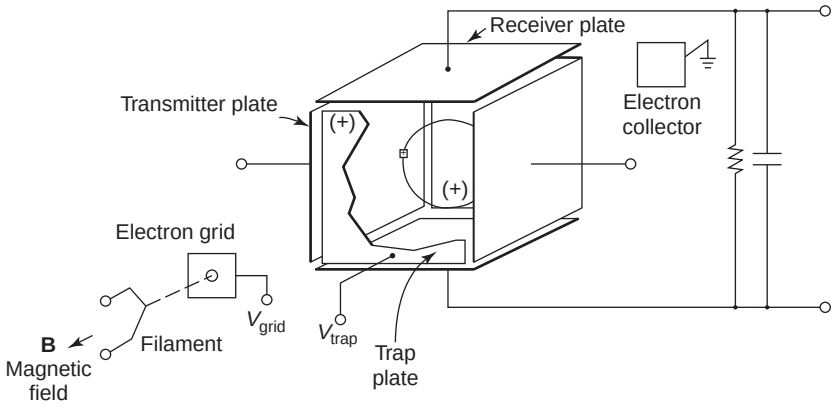

Figure 11 Schematic drawing of the basic ICR cubic cell illustrating the excitation, detection and trapping of ions in the cell. (Reproduced with permission of Elsevier Scientific Publications, from M.L. Gross, ed., *Mass Spectrometry in Biological Sciences: A Tutorial*, NATO ASI Series C, 353, 1992.)

1 part in 10^9, this in principle provides the basis for very high-mass resolution and accuracy of mass measurement. Furthermore, ions are not destroyed on detection, so that their concentrations can be monitored as a function of time, allowing a variety of multiple resonance experiments to be performed. Typical values of B for commercial instruments range from 3.0 to 9.4 T and mass RPs of well over a million have been obtained for low mass ions and of several hundred thousand for multiply charged biomolecules with m/z in the region of 1000–2000.

In the basic cubic cell shown in Figure 11, ions are generated by electron ionization by passage of an electron beam through the cell along the z-axis. In most applications, ions are generated either by MALDI or ESI in an external source and are introduced into the ICR cell for reaction or analysis. Once inside the cell, the motion of the ions is extremely complicated and three different types of oscillation can be distinguished. In the presence of the strong magnetic field, ions execute cyclotron motion about the z-axis in the xy plane but they are free to move along the z-axis. A small trapping voltage, usually in the region of 1 V, of the same sign as the charge on the ions to be trapped, is therefore applied to the two plates that are in the xy plane perpendicular to the magnetic field. This effectively prevents loss of ions along the z-axis but also modifies the motion of the ions so that three types of oscillation have to be considered:

1. cyclotron motion about the z-axis
2. magnetron motion due to the precessing of the cyclotron motion about the z-axis
3. z oscillation in which ions oscillate between the trapping plates.

The angular frequencies of these three types of motion in a cubic cell are given by Equations (17–19):

$$\text{Cyclotron} \quad w_c = \frac{[zB + (z^2B^2 - 4mzE_0)^{1/2}]}{2m} \tag{17}$$

$$\text{Magnetron} \quad w_m = \frac{[zB - (z^2B^2 - 4mzE_0)^{1/2}]}{2m} \tag{18}$$

$$z\text{-Trapping} \quad w_T = \left(\frac{2}{d}\right)\left(\frac{zV_T}{m}\right)^{1/2} \tag{19}$$

where $E_0 = 2\alpha V_T/a^2$, α is a geometrical factor equal to 1.386 for a cubic cell and a is the length of the cell. V_T is the trapping voltage applied to plates of separation d. Equation (17) therefore gives a better estimate than the simple Equation (16) for the relationship between w_c and m, but space charge effects which vary with the number of ions in the cell cause a slight reduction in w_c for a given ion, so that Equation (17) is only approximate. In addition, the effects of space charge and trapping voltages are slightly different for cells with different geometries.

Ions are excited by the application of an electrical signal of frequency w_c across the upper and lower plates of the cubic cell. The radius to which the ions are excited is independent of their mass but is a function of the amplitude and duration of irradiation. Radiation of constant amplitude over the frequency range required to excite all ions of interest is required for constant sensitivity over the whole mass range. The excitation of the ions causes an RF current across the plates which is proportional to the number of ions of the particular mass-to-charge ratio that are being excited. In FTICR, a ''chirp'' or short burst of RF radiation of constant amplitude during which the frequency is scanned rapidly was originally used to excite the ions. This has the disadvantage of exciting the motion of the ions along the z-axis, the extent of which is mass dependent. As a result, alternative methods of excitation have been employed, of which stored waveform inverse Fourier transform (SWIFT) excitation is increasingly used.[38] This has the advantage of being able to eject ions in a mass selective manner so that unwanted ions can be removed from the cell, thereby reducing space charge effects.

The maximum theoretical mass RP obtainable in an ICR cell is given by Equation (20):

$$R = \text{constant} \times \frac{zeB}{mP} \tag{20}$$

where P is the pressure within the cell. The RP therefore is predicted to fall linearly as m/z and P rise but to rise linearly with increase in magnetic field strength. The highest RP is therefore obtained with instruments having high field magnets (7.0 or 9.4 T) and an operating pressure in the region of 10^{-8} torr. In practice, for multiply charged ions of biological samples with $m/z > 2000$, the RP falls off more rapidly than is predicted by Equation (18) owing to the increased effects of nonideal electric fields. More complex cell designs attempt to overcome this problem but the design usually requires a compromise to be made between RP and efficiency of ion trapping. Because the effects of ion trapping voltages and space charge vary with cell design, the precise relationship between w_c and m/z is difficult to establish. Consequently, although very high RP can be obtained, the accuracy of mass measurement based simply on the measurement of w_c without calibration is difficult to predict.

One of the problems is that on excitation, the ions move away from the cell axis and have a wide spatial distribution caused by the increased radius of the cyclotron and magnetron motions of the ions, leading to reduced mass RP. Various ion cooling techniques have been used to concentrate the ions into a tight spatially confined packet

prior to injecting them into the ICR cell along the z-axis. A buffer gas rapidly reduces the amplitudes of cyclotron and z-trapping oscillations but increases the amplitude of magnetron motion with deleterious effects. This motion increases as the pressure, ion density and trapping field increase and in extreme cases can lead to the loss of ions. Even under less extreme conditions, the coherence of the ion cloud is reduced, resulting in decreased RP and sensitivity, especially for ions with a higher mass-to-charge ratio.

A promising development, first used in Penning traps, is quadrupolar excitation which makes use of irradiation by means of an RF electric field of two-dimensional azimuthal quadrupolar symmetry.[39] This has the property of converting the troublesome magnetron motion into cyclotron motion. Since cyclotron motion is rapidly removed by collisions of ions with a buffer gas, the use of quadrupolar excitation and a buffer gas effectively cool all three types of ion motion so that the ions are confined to the cell axis. The technique is therefore often referred to as quadrupolar axialization.[40,41] It produces both a dramatic increase in RP and sensitivity and also an increase in efficiency of ion trapping, so that it is possible to measure the ion current repeatedly with an efficiency of remeasurement approaching unity. This not only allows the detection of samples at attomole levels but also allows sensitive time-dependent experiments to be carried out at high resolution.

The dynamic range of the basic cubic cell ICR instrument is about 10^3 because with fewer than 10^3 ions in the cell, ion statistics problems begin to appear and with more than 10^6 ions in the cell, space charge effects become an increasing problem. For high RP operation, fewer than 10^5 ions are usually required. The techniques of SWIFT and quadrupolar axialization largely overcome this problem because SWIFT can be used selectively to eject unwanted ions, reducing space charge effects, and quadrupolar axialization allows remeasurement of relatively few ions under well-controlled conditions.

As a result of these improvements in controlling the ion motion within the ICR cell, the FTICR instrument now offers the highest RP at the highest sensitivity of any commercially available mass spectrometer. As larger magnetic field strengths become available, performance will continue to increase because many of the desirable features of this type of instrument improve either linearly or quadratically with the strength of the magnetic field. The increased size of the magnets will require that ions made in external sources can be transferred with high efficiency over the increased distance between the source and ICR cell by means of an octopole ion guide.

3 APPLICATIONS

3.1 Applications Involving Determination of Elemental Composition

If the RP is sufficiently high to separate all components having the same nominal mass, i.e. the ''chemical noise'' is removed, an accurate mass measurement is frequently sufficient to provide an elemental composition of an ion which, in many circumstances, is sufficient to identify the component unequivocally. One of the major driving forces for the development of high-resolution electron ionization mass spectrometry at the beginning of the 1960s was the need in the oil and petroleum industries to identify sulfur-containing compounds present in complex mixtures of hydrocarbons. A major difficulty was differentiating between hydrocarbons and sulfur-containing compounds of the same nominal mass in which three ^{12}C atoms had been replaced by one ^{32}S and four ^{1}H atoms, increasing the mass by only 0.003371 Da. Even with an RP of 100 000 which was attainable when using a large double-focusing magnetic sector instrument, it is possible to resolve ions of this type only up to about 340 Da, and only then if it is assumed that no other ions interfere.

The wider applicability of this technique was quickly recognized by synthetic chemists and one of the most common uses of high-resolution mass spectrometry is in verifying that the mass of a newly synthesized compound is in agreement with that calculated from isotopic masses. In the absence of any appreciable background spectrum, an accuracy of ± 5 ppm is usually deemed to be sufficient to verify the elemental composition of a compound. Similarly, the accurate mass measurement of a fragment ion can indicate its elemental composition and hence that of the neutral species lost, but in this case high RP is required because different fragmentation reactions may lead to ions of different elemental composition that have the same nominal mass. A straightforward example of the need for this is in demonstrating that certain peaks in the mass spectra of ketones are doublets arising from both alkyl and acyl ions. The m/z 57 ion in the mass spectrum of 3-heptanone, n-$C_4H_9COC_2H_5$, is a mixture of the ions $C_4H_9^+$ and $C_2H_5CO^+$, the accurate masses of which are 57.0704 and 57.0364 Da, respectively. Since the mass difference is 0.0340 Da, the peaks are fully resolved at a relatively low RP of about 1600 but similar pairs of ions at higher masses require increasingly high RP to separate them.

A somewhat more complicated use of high RP is in the identification of a component in a complex mixture. Here, there may be many more components that can potentially interfere with the observation of the peak of interest and the RP employed must be sufficiently high to cope with this. A simple example of this type of application is the detection of traces of the potent

carcinogen dimethylnitrosamine, $(CH_3)_2NNO$, for which the accurate molecular weight is 74.0479 Da. In foodstuffs, this may often occur together with methyl acetate, which has an accurate molecular weight of 74.0367 Da, a difference in mass of 0.0112 Da. An RP of approximately 6600 is required to separate these peaks, allowing high-resolution single ion monitoring to be used to detect dimethylnitrosamine directly from the effluent of a gas chromatograph.

A further example of the use of high RP to assist in compound identification is the detection of 2,3,7,8-tetrachlorodibenzodioxin, the most abundant ion of which, in the molecular ion region, has a nominal mass of 322 Da. This is due to the ion $^{12}C_{12}H_4{}^{35}Cl_3{}^{37}Cl$, for which the accurate mass is 321.8936 Da. A number of other ions of chlorinated species have closely similar masses, the closest being those of composition $^{12}C_{13}H_6O^{35}Cl_2{}^{37}Cl_2$ and of mass 321.9114 Da, a mass difference of 0.0178 Da. In order to resolve the peaks from these two ions, an RP in excess of 18 000 is required together with high stability of the mass calibration of the instrument.[42]

3.2 High-resolution Electrospray Ionization Mass Spectrometry

The use of high RP to enable the elemental composition of an ion to be determined as described above becomes decreasingly reliable as m/z rises, and above m/z 700–800 the number of different possible isotopic compositions rises rapidly. Frequently, many of these can be excluded on other grounds, so that the method is still of value in showing that the measured mass is consistent with the proposed elemental composition, though it is rarely possible to determine this unequivocally. The two major ionization methods introduced in the mid-1980s, MALDI and ESI, were initially interfaced to low-resolution time-of-flight and quadrupole mass analyzers and were used to obtain approximate molecular weights of sample of 10 000–50 000 Da to ±0.1%. It was not until they were used in conjunction with an FTICR mass analyzer and later with an ion trap that interest in new applications of high-resolution mass spectrometry revived in about 1990.

Peaks observed in a low-RP ESI spectrum are generally unresolved clusters of isotope peaks. In the absence of any interfering peaks, these can be used to give an estimate of the average or ''chemical'' molecular weight of a sample provided that the number of charges carried by the ion can be determined. If an ion carries z charges, the separation of peaks on the m/z scale is $1/z$ units. Hence, for an ion of nominal mass 10 000 Da carrying 10 charges, $m/z = 1000$ and the separation between peaks on the m/z scale is 0.1 unit. The RP needed to separate these peaks is 1000/0.1 or 10 000, which is readily attainable in an ESI/FTICR instrument. This method of using high RP to separate isotopic peaks, to determine the number of charges from the separation of the peaks and so to measure masses to an accuracy of less than 1 ppm has now been accomplished in a number of laboratories. This is particularly valuable in determining the products of collision-induced decomposition induced by sustained off-resonance irradiation (SORI). In the case of the +10 charge state of ubiquitin at m/z 857, McLafferty et al. have obtained an RP in excess of 2 000 000[43] and fragments of masses in the range 1000–7000 Da have been measured to an accuracy of better than 0.1 ppm.[44] The isotopic resolution of ions of mass above 30 000 Da initially proved difficult but this was first accomplished by McLafferty's group when they resolved the +43 charge state of electrosprayed albumin with an RP of 166 000.[45]

It is important to note, however, that above about 30 000 Da, the measured mass is not the mass of a monoisotopic ion. Each such ''peak'' consists of a large number of superimposed peaks due to ions of the same nominal mass containing different numbers of ^{13}C, ^{15}N, ^{34}S, and so on, heavy isotopes. In general, the RP required to separate these peaks is appreciably higher than that available and the measured mass is therefore an average value rather than that of a single ion. The monoisotopic peak due to ions consisting entirely of light isotopes is not observed because the probability of forming such a molecule is vanishingly small. The position of this peak can be estimated by matching the experimentally observed isotopic abundances to those calculated for a protein of average amino acid composition, but slight errors in the measured abundances may lead to errors of more than 1 Da.[46] An alternative approach is to express a protein from a medium that contains ^{13}C-depleted glucose and ^{15}N-depleted ammonium sulfate, thereby considerably increasing the probability of forming the monoisotopic molecule and so observing directly the peak due to this ion. This significantly increases the upper mass limit at which the accuracy of mass measurement is within 1 Da and facilitates the identification of variants and various types of adduct.[47]

The gas phase conformations of proteins can be probed by observing the rate and extent of H/D interchange between D_2O and multiply charged protein ions contained in the cell of an FTICR instrument. An early example is shown in Figure 12 of electrosprayed equine cytochrome-c in the presence of D_2O in which the isotopic envelope of the 11+ ion is shown as a function of time.[48,49] In general, several different rates of exchange were observed, reflecting the ease with which certain groups of protons could be exchanged. These, in turn, reflected the gas phase conformations of the proteins, exchange being more rapid and extensive for more open conformations of the protein, so that changes in conformation can be followed by changes in the extent

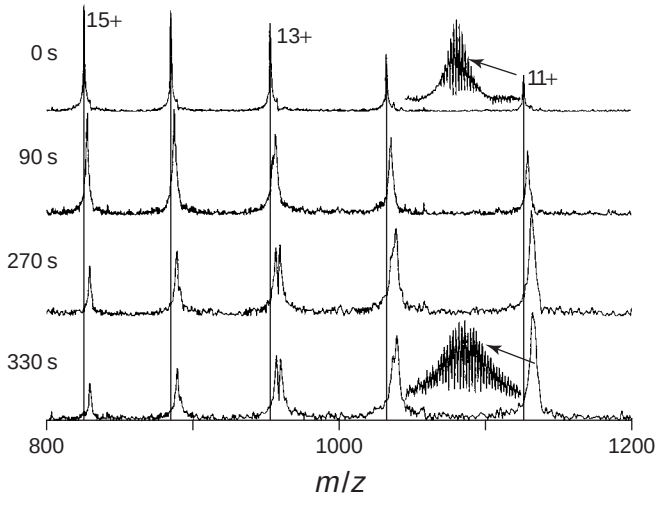

Figure 12 FTICR spectrum of electrosprayed equine cytochrome-c in the presence of D_2O. (Reproduced with permission of the American National Academy of Sciences, from Suckau et al.[49])

of H/D interchange. If the gas phase data are assumed to reflect accurately the relative concentrations of species in solution, the rate and extent of H/D interchange in solution can also be followed by high-resolution mass spectrometry. Subsequently, by using enzymatic digests and tandem mass spectrometry, the position of the H/D exchange within the protein may be determined. The RP required to follow H/D exchange in a protein is not less than unit mass resolution in the region of the molecular ion of the protein, i.e. typically 30 000–50 000.

4 CONCLUDING REMARKS

After a period during which the importance of high RP and accuracy of mass measurement appeared to decrease, the application of mass spectrometry to high molecular weight biological molecules has brought a resurgence of interest in these topics. Instrumental improvements in time-of-flight and ICR instruments now offer good high-resolution performance at relatively high masses. These will be increasingly important in studies of detailed mechanisms of reactions of peptides and proteins and high-resolution mass spectrometry will doubtless play an important role in such studies.

ABBREVIATIONS AND ACRONYMS

DC	Direct Current
ESI	Electrospray Ionization
FTICR	Fourier Transform Ion Cyclotron Resonance
FWHM	Full Width at Half Maximum
ICR	Ion Cyclotron Resonance
MALDI	Matrix-assisted Laser Desorption Ionization
RF	Radiofrequency
RP	Resolving Power
SORI	Sustained Off-resonance Irradiation
SWIFT	Stored Waveform Inverse Fourier Transform

RELATED ARTICLES

Peptides and Proteins **(Volume 7)**
High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis

Mass Spectrometry **(Volume 13)**
Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. E.G. Johnson, A.O. Nier, 'Angular Aberrations in Sector Shaped Electromagnetic Lenses for Focusing Beams of Charged Particles', *Phys. Rev.* (2nd Ser.), **91**, 10–17 (1953).
2. J. Mattauch, R. Herzog, 'New Mass Spectrograph', *Z. Physik.*, **89**, 786–795 (1934).
3. J. Mattauch, 'A Double-focusing Mass Spectrograph and the Masses of N^{15} and O^{18}', *Phys. Rev.* (2nd Ser.), **50**, 617–623 (1936).
4. J.S. Cottrell, R.J. Greathead, 'Extending the Mass Range of a Sector Mass Spectrometer', *Mass Spectrom. Rev.*, **5**, 215–247 (1986).
5. K.R. Jennings, G.G. Dolnikowski, 'Mass Analyzers', in *Methods in Enzymology*, ed. J.A. McCloskey, Academic Press, New York, 37–61, Vol. 193, 1990.
6. M.L. Gross, 'Tandem Mass Spectrometry: Multisector Magnetic Instruments', in *Methods in Enzymology*, ed. J.A. McCloskey, Academic Press, New York, 131–153, Vol. 193, 1990.
7. R.A. Yost, R.K. Boyd, 'Tandem Mass Spectrometers: Quadrupole and Hybrid Instruments', in *Methods in Enzymology*, ed. J.A. McCloskey, Academic Press, New York, 154–200, Vol. 193, 1990.
8. J.B. Farmer, 'Types of Mass Spectrometers', in *Mass Spectrometry*, ed. C.A. McDowell, McGraw-Hill, New York, 7–44, 1963.
9. R.W. Kiser, *Introduction to Mass Spectrometry and its Applications*, Prentice Hall, Englewood Cliffs, 88, 1965.

10. B.A. Collings, D.J. Douglas, 'An Extended Mass Range Quadrupole for Electrospray Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **162**, 121–127 (1997).
11. W. Paul, 'Electromagnetic Traps for Charged and Neutral Particles', *Angew. Chem., Int. Ed. Engl.*, **29**, 739–748 (1990).
12. R.E. March, 'An Introduction to Ion Trap Mass Spectrometry', *J. Mass Spectrom.*, **32**, 351–369 (1997).
13. R.E. March, 'Advances in Quadrupole Ion Trap Mass Spectrometry: Instrument Development and Applications', *Adv. Mass Spectrom.*, **14**, 241–278 (1998).
14. J. Louris, J. Schwartz, G. Stafford, J. Syka, D. Taylor, 'The Paul Ion Trap Mass Selective Instability Scan: Trap Geometry and Resolution', Proceedings of the 40th ASMS Conference on Mass Spectrometry and Allied Topics, Washington, DC, 1003–1004, 1992.
15. F.A. Londry, G.J. Wells, R.E. March, 'Enhanced Mass Resolution in a Quadrupole Ion Trap', *Rapid Commun. Mass Spectrom.*, **7**, 43–45 (1993).
16. C.D. Cleven, K.A. Cox, R.G. Cooks, M.E. Bier, 'Mass Shifts due to Ion/Ion Interactions in a Quadrupole Ion-trap Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **8**, 451–454 (1994).
17. R.K. Julian, Jr, H. Reiser, R.G. Cooks, 'Large Scale Simulation of Mass Spectra Recorded with a Quadrupole Ion Trap Mass Spectrometer', *Int. J. Mass Spectrom. Ion Processes*, **123**, 85–96 (1993).
18. F.A. Londry, R.E. March, 'Systematic Factors Affecting High Mass-resolution and Accurate Mass Assignment in a Quadrupole Ion Trap', *Int. J. Mass Spectrom. Ion Processes*, **144**, 87–103 (1995).
19. M. Guilhaus, 'Principles and Instrumentation in Time-of-flight Mass Spectrometry', *J. Mass Spectrom.*, **30**, 1519–1532 (1995).
20. D.H. Russell, R.D. Edmondson, 'High-resolution Mass Spectrometry and Accurate Mass Measurements with Emphasis on the Characterization of Peptides and Proteins by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *J. Mass Spectrom.*, **32**, 263–276 (1997).
21. C.G. Enke, 'The Unique Capabilities of Time-of-flight Mass Analysers', *Adv. Mass Spectrom.*, **14**, 197–219 (1998).
22. B.A. Mamyrin, V.I. Karataev, D.V. Schmikk, V.A. Zagulin, 'The Mass-reflectron, a New Nonmagnetic Time-of-flight Mass Spectrometer with High Resolution', *Sov. Phys. JETP*, **37**, 45–48 (1973).
23. P.R. Vlasak, D.J. Beussman, Q. Ji, C.G. Enke, 'Methods for the Design of Broad Energy Range Focusing Reflectrons', *J. Am. Soc. Mass Spectrom.*, **7**, 1002–1008 (1996).
24. M.L. Vestal, P. Juhasz, S.A. Martin, 'Delayed Extraction Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1044–1050 (1995).
25. W.C. Wiley, I.H. McLaren, 'Time-of-flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, **26**, 1150–1157 (1955).
26. R.M. Whittal, L. Li, 'Time Lag Focusing MALDI/TOF Mass Spectrometry', *Int. Laboratory*, **28**, 41–47 (1998).
27. J.H.J. Dawson, M. Guilhaus, 'Orthogonal-acceleration Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **3**, 155–159 (1989).
28. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An Inductively-coupled Plasma-time-of-flight Mass Spectrometer for Elemental Analysis. Part II: Direct Current Quadrupole Lens System for Improved Performance', *J. Am. Soc. Mass Spectrom.*, **6**, 400–410 (1995).
29. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An Inductively-coupled Plasma-time-of-flight Mass Spectrometer for Elemental Analysis. Part III: Analytical Performance', *J. Am. Soc. Mass Spectrom.*, **6**, 411–420 (1995).
30. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'Isotope Ratios and Abundance Sensitivity Obtained with an Inductively Coupled Plasma-time-of-flight Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **6**, 920–927 (1995).
31. R.H. Bateman, M.R. Green, G. Scott, E. Clayton, 'A Combined Magnetic Sector-time-of-flight Mass Spectrometer for Structural Determination Studies by Tandem Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1227–1233 (1995).
32. T.J. Cornish, R.J. Cotter, 'A Curved-field Reflectron for Improved Energy Focusing of Product Ions in Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 1037–1040 (1993).
33. M.B. Comisarow, A.G. Marshall, 'Fourier Transform Ion Cyclotron Resonance Spectroscopy', *Chem. Phys. Lett.*, **25**, 282–283 (1974).
34. M.B. Comisarow, A.G. Marshall, 'The Early Development of Fourier Transform Ion Cyclotron Resonance (FTICR) Spectroscopy', *J. Mass Spectrom.*, **31**, 581–585 (1996).
35. I.J. Amster, 'Fourier Transform Mass Spectrometry', *J. Mass Spectrom.*, **31**, 1325–1337 (1996).
36. A.G. Marshall, C.L. Hendrickson, M.R. Emmett, 'Recent Advances in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Adv. Mass Spectrom.*, **14**, 221–239 (1998).
37. C.B. Jacoby, C.L. Holliman, M.L. Gross, 'Fourier Transform Mass Spectrometry: Features, Principles, Capabilities and Limitations', *NATO ASI Series C*, **353**, 93–116 (1992).
38. A.G. Marshall, T.-C.L. Wang, T.L. Ricca, 'Tailored Excitation for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *J. Am. Chem. Soc.*, **107**, 7893–7897 (1985).
39. G. Bollen, R.B. Moore, G. Savard, H. Stolzenberg, 'The Accuracy of Heavy-ion Mass Measurements using

Time-of-flight–Ion Cyclotron Resonance in a Penning Trap', *J. Appl. Phys.*, **68**, 4355–4374 (1990).

40. L. Schweikhard, S. Guan, A.G. Marshall, 'Quadrupolar Excitation and Collisional Cooling for Axialization and High Pressure Trapping of Ions in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **120**, 71–83 (1992).
41. S. Guan, A.G. Marshall, 'Two-way Conversation with a Mass Spectrometer: Nondestructive Interactive Mass Spectrometry', *Anal. Chem.*, **69**, 1–4 (1997).
42. J.E. Campana, D.M. Schoengold, L.C. Butler, 'An Environmental Reference Materials Program 1. Dioxin Performance Evaluation Materials', *Chemosphere*, **18**, 169–176 (1989).
43. S.C. Beu, M.W. Senko, J.P. Quinn, F.M. Wampler, III, F.W. McLafferty, 'Fourier-transform Electrospray Instrumentation for Tandem High-resolution Mass Spectrometry of Large Molecules', *J. Am. Soc. Mass Spectrom.*, **4**, 557–565 (1993).
44. M.W. Senko, J.P. Speir, F.W. McLafferty, 'Collisional Activation of Large Multiply Charged Ions using Fourier Transform Mass Spectrometry', *Anal. Chem.*, **66**, 2801–2808 (1994).
45. J.P. Spier, M.W. Senko, D.P. Little, J.A. Loo, F.W. McLafferty, 'High Resolution Tandem Mass Spectrometry of 37–67 kDa Proteins', *J. Mass Spectrom.*, **30**, 39–42 (1995).
46. M.W. Senko, S.C. Beu, F.W. McLafferty, 'Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions', *J. Am. Soc. Mass Spectrom.*, **6**, 229–233 (1995).
47. A.G. Marshall, M.W. Senko, W. Li, M. Li, S. Dillon, S. Guan, T.M. Logan, 'Protein Molecular Weights to 1 Da by ^{13}C and ^{15}N Double Detection and FT-ICR Mass Spectrometry', *J. Am. Chem. Soc.*, **119**, 433–434 (1997).
48. D.A. Laude, E. Stevenson, J.M. Robinson, 'Electrospray Ionization/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry', in *Electrospray Ionization Mass Spectrometry. Fundamentals, Instrumentation and Applications*, ed. R.B. Cole, J. Wiley & Sons, New York, 291–319, 1997.
49. D. Suckau, Y. Shi, S.C. Beu, M.W. Senko, J.P. Quinn, F.M. Wampler, III, F.W. McLafferty, 'Coexisting Stable Conformations of Gaseous Protein Ions', *Proc. Natl. Acad. Sci. USA*, **90**, 790–793 (1993).

Inorganic Substances, Mass Spectrometric in the Analysis of

Howard E. Taylor
US Geological Survey, Boulder, USA

1 Introduction 1
2 Mass Spectrometric Analysis of Inorganic Materials – Historical Perspectives 2
3 Techniques Used for Inorganic Mass Spectrometry 2
3.1 Electron Ionization Mass Spectrometry 2
3.2 Spark Source Mass Spectrometry 3
3.3 Thermal Ionization Mass Spectrometry 3
3.4 Inductively Coupled Plasma Mass Spectrometry 4
3.5 Glow Discharge Mass Spectrometry 4
4 Description of Mass Spectrometers 5
4.1 Magnetic Sector Mass Spectrometers 5
4.2 Quadrupole Mass Spectrometers 6
4.3 Time-of-flight Mass Spectrometers 7
4.4 Ion Trap Mass Spectrometers 7
5 Detectors 8
6 Sample Introduction Techniques for Inductively Coupled Plasma Mass Spectrometry Applications 8
6.1 Solutions 8
6.2 Solids 10
6.3 Gases 11
7 Comparison of Characteristics of Inductively Coupled Plasma Mass Spectrometry with Inductively Coupled Plasma Atomic Emission Spectrometry 12
Abbreviations and Acronyms 12
Related Articles 13
Further Reading 13

Mass spectrometry is an analytical technique used to measure the composition of a substance by isolating specific analyte components according to their individual atomic or molecular mass-to-charge ratios. Inorganic mass spectrometry is specifically used to determine the elemental and isotopic composition of the material being analyzed. The techniques are capable of the measurement of a range of concentrations from major components to ultratrace constituents.

Several instrumental approaches are used to separate and measure the abundance of component ions formed from the sample. These techniques include electron ionization mass spectrometry (EIMS), inductively coupled plasma mass spectrometry (ICPMS), thermal ionization mass spectrometry (TIMS) and spark source mass spectrometry (SSMS). These techniques utilize a variety of mass spectrometers including ion trap, quadrupole, magnetic sector and time-of-flight mass analyzers, depending on the type of sample being analyzed and the desired quality of the results. Often sample introduction techniques can be utilized to enhance the capabilities for solving specific analytical chemistry problems.

1 INTRODUCTION

Mass spectrometry is an analytical technique used to measure the composition of a substance by isolating specific analyte components according to their individual atomic or molecular mass-to-charge ratio (m/z). In order for these components to be separated from each other for measurement, they are converted from their native form (indigenous form in the original sample) into a stream of ions, prior to an isolation process using a mass spectrometric technique. The mass spectrometer is a device that is used to separate each of the ions, measure their relative m/z (to identify the nature of the component qualitatively), and finally measure their ion current, which is proportional to the analyte concentration.

Several approaches are used to convert gas, liquid and solid samples to ions for subsequent mass spectrometric measurement. These techniques, which are identified as EIMS, TIMS, SSMS, ICPMS and glow discharge mass spectrometry (GDMS), are used for the solution of specific analytical chemistry problems. In addition, various types of sample introduction techniques, including pneumatic nebulization, electrothermal vaporization, laser ablation and spark ablation, are used in specific analytical chemistry applications.

Instrumental methods used to separate the ions according to their m/z, or effectively to serve as a mass filter, employ approaches based upon a variety of techniques. These techniques use principles such as (1) the angular dispersion of individual m/z ions in an ion beam as it transverses through a magnetic field; (2) the differences in time required for ions of different m/z to drift a fixed distance after reaching constant acceleration; (3) the use of a tuned radiofrequency (RF) electrostatic

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

field to reject all m/z ions that have spiral paths other than the one that resonates, thereby functioning as a mass filter; and finally (4) the use of electrostatic lenses to trap ions within an RF field, followed by the controlled release of specific m/z ions by variation of the lens' potential. Combinations of ionization procedures and spectrometric techniques result in analytical instruments suitable for the measurement of major, minor and trace concentrations of elements in a variety of materials. In addition, many of these combinations are used for the measurement of stable isotopic abundances, which are required for stable isotope dilution quantitation, isotopic tracer studies and isotope geochronology.

2 MASS SPECTROMETRIC ANALYSIS OF INORGANIC MATERIALS – HISTORICAL PERSPECTIVES

The technique of mass spectrometry was first introduced by J.J. Thomson in 1913. Thomson employed the principle to identify the isotopes of neon. He used a combination of electric and magnetic fields to separate ions of ^{20}Ne and ^{22}Ne originating from an electrical gas discharge. The first identifiable mass spectrograph was designed in the early 1920s by F.W. Aston. In these early years mass spectrometry was primarily used for the measurement of elemental stable isotopes, for which Aston was awarded the Nobel prize. Shortly after winning the prize, Aston incorrectly predicted that the future usefulness of the technique would be limited. However, in the early 1940s, mass spectrometry found use in the field of organic analytical chemistry for molecular structure identification and hydrocarbon analysis.

In the 1950s, the first mass spectrometer, which was designed specifically for inorganic analytical chemistry, utilized an RF spark discharge as an ionization source. The primary advantage of this technique, in addition to the ability to measure stable isotopes of analyte elements, was the capability to determine elements at sub-parts per million concentration levels for both easy and difficult to ionize atoms, and its ability to perform analysis of solid samples directly, without the need for sample dissolution.

The first use of a plasma as an ion source for inorganic mass spectrometric analysis was reported by A.L. Gray in 1975. He used a DC (direct current) capillary arc for analyte ionization. This was rapidly followed by the use of an inductively coupled plasma (ICP) for sample atomization and ionization, patterned after the popular atomic emission spectrometric approach that was being extensively used in many laboratories for trace element analysis.

3 TECHNIQUES USED FOR INORGANIC MASS SPECTROMETRY

3.1 Electron Ionization Mass Spectrometry

EIMS is used in inorganic analysis for the determination of noble and permanent gases such as H_2, O_2, and N_2. In addition, other components such as carbon and sulfur can be determined after oxidation, by the measurement of the ionized oxide molecules, CO_2 and SO_2. The instrument uses a gas inlet system coupled to a conventional magnetic sector mass spectrometer, sometimes equipped with multiple detectors to provide simultaneous multiconstituent or multiisotope determinations. The sample in a gaseous form is introduced through a controlled leak gas inlet system (see Figure 1) into the ionization source, where collision of analyte species occurs with a beam of accelerated electrons. This collision of electrons, which typically are accelerated to an energy of approximately 70 eV, with components of the sample causes ionization of the sample atoms or molecules. These ionized atoms or molecules are then separated according to their m/z by the mass analyzer and measured by the detection system of the spectrometers.

In addition to permanent gases, selected metals can also be determined as gaseous species. Mercury, because it has a high vapor pressure at room temperature, can be measured directly in its elemental state. Other metals such as selenium can be determined when bonded to volatile organic compounds. For example, the isotopic composition of calcium has been determined using volatile chelate compounds formed with trifluoracetylacetone. Other metals such as chromium and nickel (using dithiocarbamate) can be determined in a similar manner.

This technique provides an ideal approach for the high precision determination of isotope ratios of constituents. In addition, the ability to measure isotopic composition

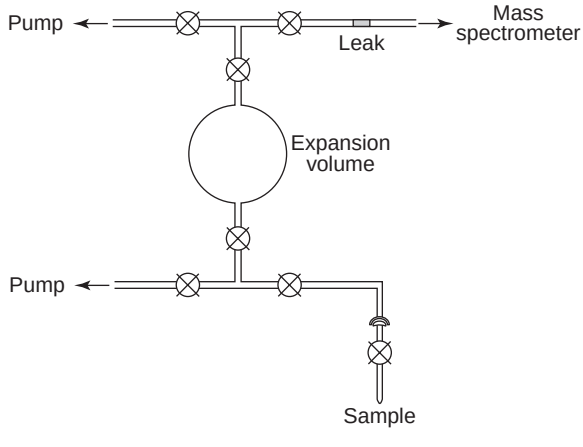

Figure 1 Gas inlet system for EIMS.

allows high accuracy isotope dilution quantitation to be performed.

3.2 Spark Source Mass Spectrometry

SSMS is a technique that was designed for the direct trace elemental analysis of solid samples requiring a minimum of sample preparation. In 1934, A.J. Dempster introduced a spark source, created by an RF (1-MHz) oscillating potential of several kilovolts, applied to electrodes formed from the sample material. If the sample is electrically nonconducting, it is mixed with a conducting binder. This spark initially vaporizes and atomizes the components of the sample, followed by the immediate ionization of the atomized elements.

In 1954, N.B. Hannay described the first practical analytical instrument utilizing the spark source. Modern analytical instrumentation uses the spark source to create ions of the components of the sample, which by electrostatic acceleration form an ion beam whose composition is representative of the original sample. This composite ion beam is separated into individual analyte ion beams by a double-focusing mass spectrograph utilizing the high resolution Mattauch–Herzog geometry. The resolved ion beams are focused on a nearly flat focal plane, where an ion sensitive emulsion on a photographic plate is used as a detection system. The photographic plate offers the combination of an integrating detector, which provides high sensitivity with truly simultaneous multielement detection. A fixed position electron multiplier can also provide ''electrical detection'', which by controlled variation of the magnetic field provides mass scanning capability.

This technique provides the capability for the direct determination of about 70 elements in solid material, without the need for extensive sample preparation, with a sensitivity of within a factor of 2–3 for all elements. Detection limits in the 1–100 ppb range can be achieved with a few milligrams of sample. Calibration curves demonstrate linearity of over five orders of magnitude, ranging from major components down to trace constituents on a single analysis. Although the technique offers many unique analytical capabilities, it does not provide high measurement reproducibility. The combination of the limited stability of the spark discharge and its ability to atomize the sample, the inhomogeneity of the distribution of the analyte elements in the sample and the variations in the ion beam due to changes in the electrode gap during sample vaporization limit the precision of ion current measurements to ±10–20% relative standard deviation (RSD) depending on the integration time. This relatively poor reproducibility limits the absolute accuracy that can be achieved in routine operation. The use of isotope dilution quantitation for selected elements overcomes this limitation. Because of this problem with reproducibility, multielement SSMS is often used as a semiquantitative trace element survey technique.

3.3 Thermal Ionization Mass Spectrometry

Sample in the form of a liquid solution of chloride or nitrate salts is deposited on the surface of a refractory metal filament, usually made of tantalum, rhenium, tungsten or thoriated tungsten (to lower its work function). The filament is resistively heated to over 1000 °C by passing an electrical current through it. The heating process sequentially vaporizes the analyte followed by immediate thermal ionization at the surface of the metal filament due to the low work function of the filament metal (Figure 2). Table 1 shows the respective melting temperatures and work functions of the typical filament metals. Thermal sources can consist of a single, double or triple filament. Single filament ion sources are used both to vaporize and ionize the analyte element at the same filament surface. With the double and triple filament ion sources, sample vaporization occurs from the first filament followed by adsorption of the gaseous

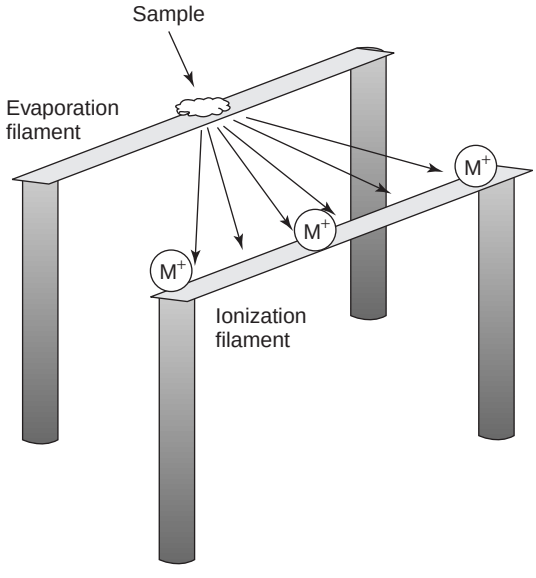

Figure 2 Diagram of a thermal ionization filament source.

Table 1 Composition, melting points (°C) and work functions (eV) of common filament materials used in TIMS

Filament material	Melting point (°C)	Work function (eV)
Platinum	1772	5.13
Tantalum	2996	4.30
Rhenium	3180	4.98
Tungsten	3410	4.58
Tungsten (thoriated)	~3410	2.7

sample on the second or third filament, where ionization takes place by transfer of an electron from the analyte to the filament metal.

Elements that have a first ionization potential of <7 eV are directly ionized by this technique with high efficiency. This includes elements such as alkalis, alkaline earths, actinides and rare earths. Other elements, such as 1st and 2nd row transition elements that have first ionization potentials greater than 7 eV, often require matrix modifications to enhance their ionization efficiency. For example, silica gel is added for the analysis of elements such as cadmium, copper, lead, silver and zinc. In addition, negative ions can also be formed for elements such as the halogens, selenium, sulfur and tellurium, by transfer of an electron from the filament metal to the analyte atom.

Ions produced by this thermal ionization process are separated and quantitatively measured by a mass spectrometer, hence the name TIMS. Since predominately singly charged ions are formed, the mass spectrum is relatively simple. Instruments utilizing quadrupole mass spectrometers or medium resolution single-focusing magnetic sector mass spectrometers are used for separation of the ions formed by this method. This approach can produce very high precision isotope ratio measurements, which can be effectively utilized for isotopic measurement applications such as geochronology or quantitation employing the isotope dilution procedure.

3.4 Inductively Coupled Plasma Mass Spectrometry

The ICP is a high electron collisional temperature (10 000 K) inert gas (usually argon) medium, sustained by a high-power (1–2 kW) RF electromagnetic field. The plasma is confined in a quartz torch, sustained by the electromagnetic field, which is coupled to the plasma gas via an external antenna (load coil), thereby eliminating artifacts originating from internal electrode materials (Table 2).

The ICP has been extensively used as an atomic emission source for trace element analytical chemistry. It has the unique features of being able to atomize efficiently essentially all elements in the periodic table, including the refractory (high-boiling-point) elements. The ability to form predominately singly charged ions of all the elements, with low probability of the formation of doubly charged ions (owing to the 15.76-eV ionization energy of argon, which is less than the 2nd ionization potential of most elements), results in a simple mass spectrum. The chemical inertness of the noble gas medium reduces the prevalence of interference effects.

Table 2 Typical operating conditions for an argon ICP ionization source

Power	1–2 kW
Argon flow rates	
coolant	15 L min^{-1}
auxiliary	0–2 L min^{-1}
nebulizer	1 L min^{-1}
Sampling depth[a]	14–18 mm

[a] Distance above the load coil.

The sample, in liquid form, is introduced into the plasma as a gas-transported aerosol, produced by pneumatic nebulization. When entering the plasma the aerosol is initially desolvated, followed almost immediately by thermal atomization. As the atoms are transported into the higher energy regions of the plasma, ionization takes place, primarily by a collision mechanism with accelerated argon ions and electrons from the plasma. The analyte ions are mechanically extracted from the plasma through the orifice of a sampling cone made of nickel or platinum. Ions extracted from the plasma form an ion beam, which is further refined by the orifice of an internal skimmer cone. This ion beam, which has a composition representative of the original sample, passes into the mass analyzer where individual ion beams of each of the components are isolated and sequentially transported to the electron multiplier detector.

Other sample introduction techniques such as laser ablation, electrothermal vaporization, or direct insertion devices provide a method for nonliquid samples to be quantitatively introduced into the plasma. Each of these techniques offers special characteristics, which can be utilized for the solution of specific analytical chemical problems.

Most ICPMS instruments employ a quadrupole mass analyzer for isolation of individual ion currents. This equipment provides basic unit mass resolution capability, which is usually sufficient for most analytical applications (Figure 4). For the solution of specialized analysis problems, the coupling of the ICP with a high-resolution magnetic sector mass spectrometer provides the capability for analyte ions to be separated from isobaric interferences in the mass spectrum, thereby allowing high accuracy determinations.

3.5 Glow Discharge Mass Spectrometry

Because of their simplicity and stability, low-pressure DC and RF discharges were used as ionization sources in early mass spectrometers. The modern technique of GDMS consists of a sample enclosed in a glow discharge cell (see Figure 3). An inert gas, usually argon or neon, at low pressure (0.1–100 mbar) is used to form a glow discharge (low-pressure plasma). For DC glow discharges, the sample must be an electrically conducting material or a nonconductor mixed with a conductor, such as graphite,

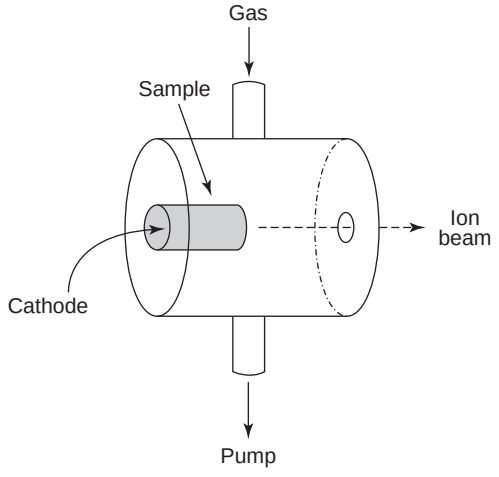

Figure 3 Diagram of a glow discharge cell.

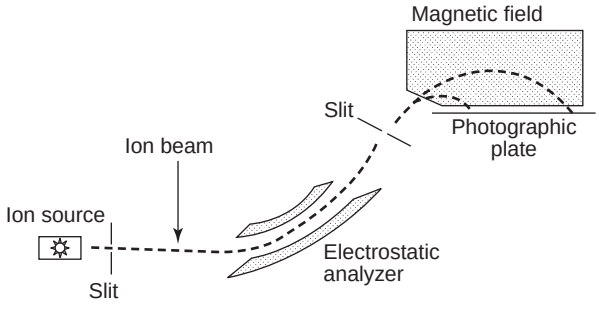

Figure 4 Diagram of a Mattauch–Herzog geometry double-focusing mass spectrometer.

as a binder. RF glow discharges can analyze nonconductor samples directly.

The sample is usually configured as the cathode in the glow discharge cell, with the anode forming the housing of the cell. The surface of the sample is sputtered by the high-energy argon ions of the inert gas ions of the discharge, which vaporizes and atomizes the analyte species into the cell. After atomization, the analyte atoms are ionized by a collisional mechanism similar to that employed in atmospheric pressure plasmas. Ions are removed from the glow discharge cell through an extraction orifice. An ion beam is formed from the ions extracted from the cell. This beam is accelerated and focused by ion lenses and directed into the mass spectrometer. Both magnetic sector and quadrupole mass analyzers are commonly used with glow discharge sources.

GDMS has been extensively utilized for the determination of trace elements in pure metals and alloys (i.e. steel, high-duty nickel alloys, gold, aluminum, etc.) and in various semiconductor materials (i.e. silicon, gallium arsenide, etc.) High sensitivities produce detection limits for most elements in the parts per trillion range.

4 DESCRIPTION OF MASS SPECTROMETERS

The designs of mass spectrometers used in inorganic chemistry applications are basically the same as those used in organic mass spectrometry. Variations of three basic types of mass spectrometer are employed depending upon the specific application: the combination of magnetic and electrostatic fields for separation of ions (magnetic/electrostatic sector), the time separation of ions (time-of-flight), and the use of resonance mass filters (quadrupole/ion traps).

4.1 Magnetic Sector Mass Spectrometers

The classical type of mass spectrometer, the magnetic sector instrument, as the name suggests, employs a magnetic field to separate various m/z ions from each other, so that they can be measured independently. It is based on the principle that a charged particle, traveling at a constant velocity, perpendicular to the lines of force of a magnetic field will be deflected at an angle proportional to the mass and charge on the particle. Equation (1) describes the mathematical relationship between the accelerating voltage (V), mass (m) and charge (z) of an ion and its radius of curvature (r) as it passes through a fixed magnetic field (H)

$$r = \frac{(2Vm/z)^{1/2}}{H} \tag{1}$$

From this equation, it is apparent that the radius of curvature of a specific m/z ion is proportional to the square root of the mass of the ion. Therefore, an ion beam of a mixture of m/z ions will be separated into multiple ion beams that can be resolved for independent measurement.

A single-focusing magnetic sector instrument is the simplest mass spectrometer. It utilizes an ion beam accelerated to a constant velocity passing through a magnetic field. For multielement determinations, the resolved analyte ion beams are simultaneously detected by a photographic emulsion, or for single ion monitoring, the specific ion beam is isolated by an exit slit and measured with an electron multiplier detector.

A double-focusing mass spectrometer combines the previously described magnetic sector with an electrostatic analyzer. The electrostatic analyzer serves as an energy-focusing device to improve the energy distribution of the ion beam. Ions of a common m/z but with different energies can be focused with an electrostatic analyzer. This arrangement provides higher resolution (i.e. the ability to separate ions of similar mass). Double-focusing mass spectrometers can be configured with the

electrostatic analyzer either in front of or behind the magnetic sector.

When the electrostatic analyzer is positioned behind the magnetic sector, the instrument is named the ''reverse Nier–Johnson geometry''. This arrangement provides for focusing for only one m/z ion at a time. Therefore, it is only suitable for the use of a single m/z ratio detection system. Multiple ions are detected individually by scanning the magnetic field to position different m/z ions on the detector sequentially in time.

The ''Mattauch–Herzog geometry'' spectrometer has the magnetic analyzer positioned behind the electrostatic analyzer (see Figure 4). This type of arrangement permits double focusing for all masses simultaneously, providing the ideal geometry for multielement detection using a photographic emulsion.

The primary characteristic of magnetic sector instruments is their ability to operate at high resolution. Double-focusing mass spectrometers can easily operate at a resolution of 10 000. This means that the instrument has the ability to separate an ion with an m/z of 250 from one with an m/z of 250.025. This capability allows the isobaric interference of a molecular ion to be resolved from the analyte ion. For example, the resolution that is required for the separation of the $^{40}Ar^{16}O^{+}$ and $^{40}Ca^{16}O^{+}$ ions from the $^{56}Fe^{+}$ analyte ion, at the nominal mass of 56, is approximately 4000. Most mass spectrometers that cannot operate at this resolution will not be able to analyze iron accurately at m/z 56 in an argon plasma with calcium present.

4.2 Quadrupole Mass Spectrometers

The quadrupole mass analyzer is a device that essentially rejects all m/z ions except the unique one that is permitted to pass through the tuned RF and DC electrical fields, thereby serving as a mass filter. The quadrupole device consists of four precisely machined cylindrical rods arranged symmetrically parallel to each other (see Figure 5). The central space between the four rods is aligned concentric to the ion beam, which is accelerated and focused by electrostatic ion lenses. The ion beam is a composite of a mixture of ions representative of all of the components of the sample. As this mixture of ions with different m/z pass through this center space and travel parallel to the length of the quadrupole rods, only a single m/z ion species is permitted to traverse this space unimpeded and exit the opposite end where it is focused on an analog cup or electron multiplier detector (Figure 6).

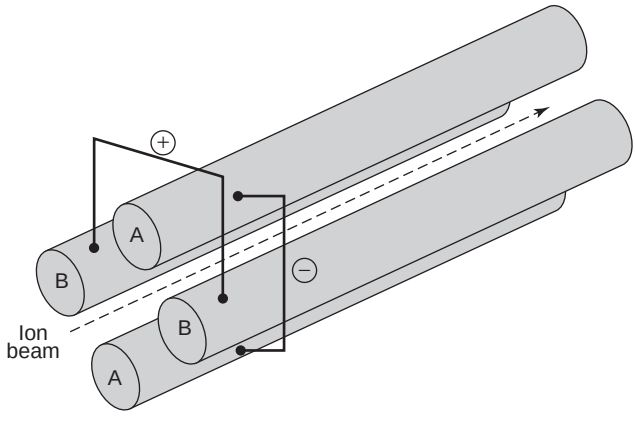

Figure 5 Diagram of a quadrupole mass spectrometer.

The ion isolation process involves the application of both DC and RF potentials to opposite rods in the quadrupole rod set. The electromagnetic field created by these potentials on the rods interacts with the beam of mixed ion species. In this applied field, each m/z ion is deflected in a spiral path as it traverses through the space between the rods. All ions, except the one with an m/z tuned to the applied fields, will be deflected into the quadrupole rods. The single unique m/z ion which is tuned to the applied field is allowed to pass through the rods and on to the detector.

Ions with different m/z values can be scanned by continuously varying the field between the quadrupole rods. This field can be changed in a uniform fashion by altering either the applied DC potential or the frequency of the RF voltage. Thus by continuously varying the applied potentials on the quadrupole rods as a function of time, the m/z values of the ions passing through the spectrometer are changed in a uniform fashion. This results in a scan of the mass spectrum of the ion beam. A spectrum obtained from a quadruple mass analyzer is shown in Figure 7. Scanning provides a suitable method for rapid multielement qualitative analysis. Peaks representing specific m/z ions are characteristic of components in the sample. Measuring the intensity of these peaks can be used to estimate the concentration of the specific component by comparison with a known concentration standard.

Another mode of operation, sometimes called ''peak hopping'' can be used to perform quantitative analysis of a specific group of one or more elements. This technique involves predetermining the necessary applied potentials to the quadrupole rods to permit the appropriate m/z ion to pass. These calibrated potentials can be

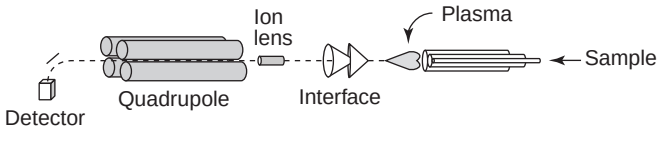

Figure 6 Diagram of an ICP mass spectrometer with a quadrupole mass analyzer.

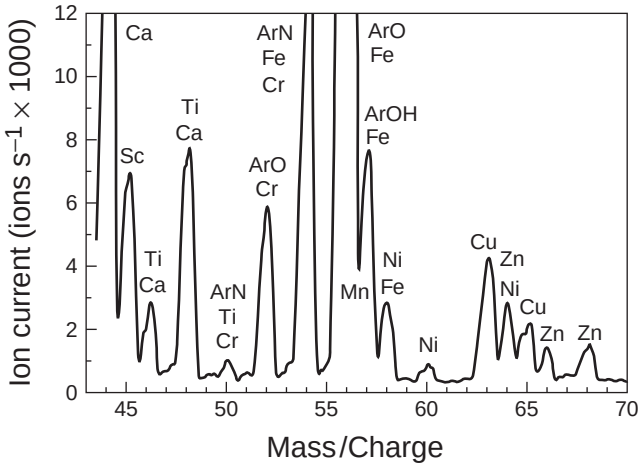

Figure 7 Mass spectrum obtained with a quadrupole mass analyzer.

rapidly changed under computer control. By sequentially changing these applied potentials with a preprogrammed routine, different analyte ions can be sequentially selected for measurement. Each of these ions beams are integrated for a specific period of time, which can then be used for quantitation of the specific analyte elements.

In addition to providing higher sensitivity and precision, the peak-hopping mode of operation is generally considered to be the most desirable for quantitative analysis. It provides the most time-efficient approach for trace element analysis, because time is not wasted measuring regions of the spectrum that are not of interest. Peak hopping is also the method of choice for isotope ratio measurements with a quadrupole mass spectrometer.

4.3 Time-of-flight Mass Spectrometers

The time-of-flight instrument is a very simple mass analyzer that does not depend on a magnetic or electrostatic field to separate or isolate various m/z ions. Ions, which are accelerated to approximately constant kinetic energy, are allowed to drift down a flight tube 1–2 m in length. From the kinetic energy equation (Equation 2), velocities (v) of ions vary inversely with the square root of mass (m)

$$KE = \tfrac{1}{2}mv^2 \qquad (2)$$

As they traverse down the drift tube, ions of different m/z reach the detector at the end of the tube at different times (i.e. are temporally resolved).

In order for this system to work, the ions must be pulsed into the spectrometer. As a pulsed group of ions enter the spectrometer, they are each time-resolved and measured as individual masses. Different m/z ions are separated in time by tens to hundreds of microseconds. Therefore, ions across the entire mass range are essentially measured simultaneously. Sequential pulsed groups of ions represent replicate measurements of the same sample.

Short analysis times are an important advantage of time-of-flight mass analyzers. Complete analysis of short-duration discrete transient sample introduction is possible. Therefore, in applications such as laser ablation or electrothermal vaporization, time-of-fight mass analyzers can be used with great advantage. In addition, this approach has the potential to provide much better precision in isotope ratio measurements.

On increasing the path length and focusing the kinetic energy of various m/z ions by using an ion reflector, which is essentially a repellent electrostatic well at the end of the flight tube, the direction of the drift of ions is reversed 180° (see Figure 8). By using this arrangement resolutions of the order of 3000 can be achieved.

4.4 Ion Trap Mass Spectrometers

An ion trap mass analyzer is based on principles somewhat related to a quadrupole mass filter. The system functions by trapping and storing a group of ions by means of an RF field in a space between a central doughnut-shaped ring electrode and two end-cap electrodes (see Figure 9). At certain RF potentials, all ions above a specific m/z are trapped within the space between the electrodes. As the RF and DC potentials are increased, ions of increasing

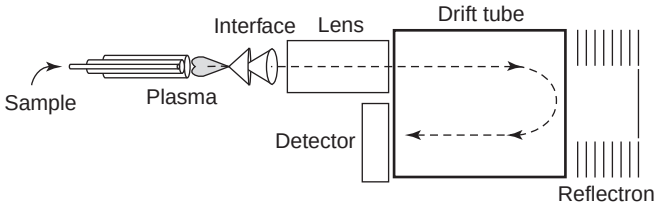

Figure 8 Diagram of a time-of-flight mass spectrometer.

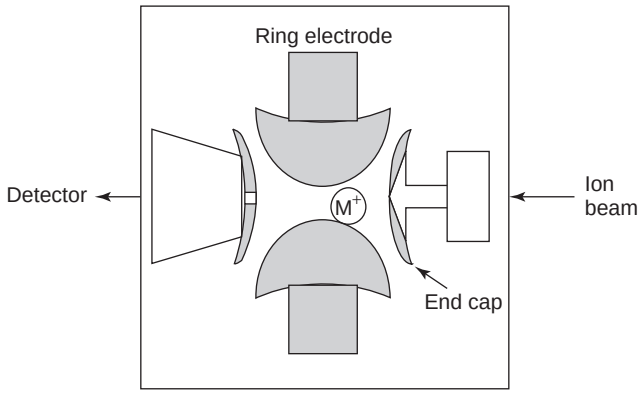

Figure 9 Diagram of an ion trap mass analyzer.

m/z come into resonance with the trapping signal and are ejected through the end-cap electrode. These selected m/z ions are then detected by an electron multiplier. As with the quadrupole mass analyzer, the mass spectrum can be scanned by continuously varying the RF and DC potentials.

The ion trap normally operates at a higher working pressure than other mass analyzers (typically 10^{-3} mbar). This feature greatly facilitates the coupling of the ion trap to atmospheric pressure ion sources.

By careful selection of operating conditions, ions of a specific m/z can be rejected from the ion trap, while other m/z ions are preferentially retained. This feature is particularly valuable when the ion trap mass analyzer is used with inert gas (Ar) plasmas. The matrix element ions (Ar^+) are continuously ejected from the trap while analyte ions are retained.

5 DETECTORS

The photographic plate with an ion sensitive emulsion was used extensively in the early days of mass spectrometry. It was used as the detector of choice for SSMS, where it offered the advantage of relatively uniform sensitivity for all elements, multimass detection (i.e. simultaneous measurement of multiple mass ions on a single aliquot of sample), and the ability to integrate a signal for a long period of time, which yielded high sensitivity resulting in very low detection limits. Unfortunately, the requirement of manual chemical development of the photoplate, coupled with limitations in precision (5–10%) and accuracy caused by plate inhomogeneity and the nonlinear response, reduced the popularity of this detector when more desirable alternatives were developed.

There are basically two types of detector that are employed with modern mass spectrometers: (1) the Faraday cup or analog detector and (2) the electron multiplier detector. The Faraday cup is the simplest of the two detectors. It is nothing more than a metal collector, coated with a low-work-function material such as BeO, GaP or CsSb, which converts an ion beam into an electrical current. A conventional low-noise DC amplifier then amplifies this electrical current. Advantages of the Faraday cup detector are its high stability, no mass discrimination and a virtually indefinite lifetime (no fatigue of the detector with use). Because only a few electrons are generated for each ion that collides with the coated collector, this detector is only suitable for high-intensity ion beam measurements, precluding its use for trace analysis.

The electron multiplier is used to measure the ion current of low-intensity ion beams. It consists of a series of dynodes with successively higher positive applied potentials. The dynodes are geometrically arranged so that an ion colliding with the first stage expels several electrons, which in turn are accelerated towards the second stage. When these electrons collide with the second dynode surface, more electrons are emitted which are accelerated towards the third stage dynode. A cascading effect results in a current at the final stage which has been multiplied by approximately a factor of 10^6. This high gain and low dark current (without the requirement for detector cooling) makes the electron multiplier suitable for pulse counting applications. Electron multipliers have a finite operating life, which is proportional to the total ion flux or accumulated charge on the initial stage. Fatigue of the low work potential surface of the detector caused by high ion current exposure can also result in a loss of stability of the system. Therefore, care must be exercised in protecting the detector from overexposure. Specialized detectors which combine analog and electron multiplier devices can be used effectively to increase the dynamic working range of the spectrometer.

6 SAMPLE INTRODUCTION TECHNIQUES FOR INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY APPLICATIONS

Sample introduction techniques developed for atomic emission plasma spectrometry are directly applicable for ICPMS measurements. Three major types of samples are encountered: (1) solutions, (2) solids and (3) gases. Each of these will be discussed in some detail.

6.1 Solutions

Solutions are usually analyzed by a nebulization process, which converts the liquid sample into a finely divided aerosol (liquid droplets). This aerosol has a composition identical to the original sample. The droplets formed by the nebulization process are size classified by routing the gas stream containing the aerosol through a spray chamber or across an impinger bead. This process removes the larger size droplets and allows the finer ones to be transported via the gas stream to the plasma torch, where it is introduced into the plasma for further processing. The size classification process significantly improves the stability of the plasma by reducing ''flickering'' caused by the presence of large-size solvent droplets.

Several types of nebulizers are used depending on the specific requirements of the analytical application. The most common type is the pneumatic nebulizer, of which there are several different designs. All pneumatic

nebulizers share a common feature – they use the force of a flowing gas to create droplets from the liquid sample.

The crossflow nebulizer consists of two capillary tubes, usually made of glass or quartz, positioned at right angles to each other so that gas flowing through one capillary aspirates solution from the other capillary tube, generating an aerosol at the tip. As with other pneumatic nebulizers, the flowing gas that produces the droplets is used to transport the aerosol through the spray chamber and into the plasma.

A concentric-type nebulizer has similar performance characteristics to the crossflow nebulizer. It also consists of two capillary tubes, one that is slightly smaller is positioned inside the larger one in a coaxial configuration. Gas is directed through the outside capillary, aspirating sample solution through the center capillary. As with the crossflow, the aerosol is formed at the tip of the capillaries. Although both of these nebulizers will naturally aspirate liquid when gas is passing through them at an appropriate flow rate, pumping the solution using a peristaltic pump will usually produce a more uniform aerosol, while eliminating the possibility of variations in flow rate as a function of sample viscosity. Sample uptake rate for these types of nebulizers is typically about $1\,mL\,min^{-1}$. Typical detection limits for selected elements are listed in Table 3.

A special configuration of the concentric nebulizer, called the microconcentric nebulizer or the direct injection nebulizer, is used for the analysis of small quantities of sample. This nebulizer is mounted at the base of the plasma torch in such a way that aerosol droplets formed by the nebulizer are directly injected into the base of the plasma, without the use of a spray chamber. The diameter of the sample capillary is much smaller than that of a conventional concentric nebulizer, resulting in smaller size droplet formation and sample flow rates of the order of $100\,\mu L\,min^{-1}$. In addition to low-volume sample consumption, when constructed from appropriate materials, this nebulizer provides for the direct analysis of corrosive substances.

Table 3 Comparison of typical detection limits of ICPMS for selected elements using pneumatic nebulization, ultrasonic nebulization and electrothermal vaporization ($ng\,L^{-1}$)

Element	Mass	Pneumatic nebulizer	Ultrasonic nebulizer	Electrothermal vaporizer[a]
As	75	50	1	0.2
Sb	121	50	0.5	0.05
Sn	120	50	7	0.1
Pb	208	50	0.3	0.1
Au	197	5	4	0.1
U	238	1	0.2	0.05
Hg	202	100	8	1

[a] Based on 30-μL sample volume.

A third type of pneumatic nebulizer is based on the design described by Babington. The principle of operation of this nebulizer is based on gas being forced through a small orifice over which the sample solution is passed. As the gas passes through the orifice, its shear force causes nebulization of the liquid at the orifice interface. This type of nebulizer requires a peristaltic pump to deliver the sample to the orifice, so it will not naturally aspirate sample. Although the measurement sensitivity of analytes determined using this nebulizer is similar to the other pneumatic nebulizers, it offers a significant advantage in not requiring sample to be processed through a small diameter capillary or orifice, precluding the possibility of partial clogging that can seriously affect the accuracy of the determination. Several variations of this nebulizer include the slot nebulizer, the cone spray nebulizer, the fritted disk nebulizer and the grid nebulizer. Each of these nebulizers has slightly different performance characteristics but they all function similarly.

Solid material in a suspended particulate form can be analyzed directly using the Babington-type nebulizer. This process is called slurry nebulization. As long as the maximum particle size is $<8\,\mu m$ diameter, the particulate is quantitatively entrained in the aerosol droplets produced by the nebulizer. This aerosol is transported through the spray chamber and into the plasma. A suitable dispersing agent (surfactant) such as Triton-X100 is added to the sample to keep the particulate matter suspended. The primary advantage of this approach is to facilitate the analysis of solid materials without the need to perform chemical dissolution sample preparation, which may result in loss or contamination of analyte constituents.

In addition to pneumatic nebulizers, there are other types of nebulizer that are used for converting liquid samples to aerosols. One of the most popular is the ultrasonic nebulizer, which works on the principle that small quantities of liquid sample deposited on the surface of a high-frequency vibrating transducer crystal will produce an aerosol. The transducer is made to vibrate by the application of a 1-MHz RF signal from an oscillator. This type of nebulizer produces an aerosol with very fine-size droplets. Because of the fine particle size, the transport efficiency of this aerosol is very high, resulting in a complete elemental analysis on a very small quantity of sample. Sample is typically pumped onto the surface of the transducer through a capillary tube at the rate of about $100\,\mu L\,min^{-1}$.

Ultrasonic nebulizers are usually operated with a desolvating apparatus. This desolvation process consists of passing the gas stream, containing the sample aerosol, through a heated chamber to evaporate the solvent. The mixture of dry aerosol and vaporized solvent is then passed through a chilled condenser where the solvent

is condensed and separated from the gas stream as a liquid. The dry sample particles in the gas steam are then transported to the plasma where they are vaporized. Because there is minimal solvent transported into the plasma after desolvation, more energy is available for dissociation and ionization of the analyte elements. Detection limits using this nebulizer are improved by a factor of 10–20 over those achieved by conventional pneumatic nebulization (see Table 3).

Another type of sample introduction device for the trace analysis of solutions is the electrothermal vaporizer. Aliquots of the sample are placed onto the surface of a metal filament, carbon rod or the interior surface of a graphite tube. An electric current is passed through the device, gradually heating it resistively to evaporate the solvent, leaving the sample residue on the surface of the device. The electrical current is then rapidly increased to heat the vaporizer to the boiling point of the constituents in the sample. The vaporized sample material is then transported in the gas stream flowing through the vaporizer into the plasma.

The instrument signal response produced by this technique is transient in nature. This means that a finite amount of analyte is introduced into the plasma over a specific period of time. During the period of time the analyte is resident in the plasma, the ions produced are sampled and the corresponding ion current is measured by the mass spectrometer. A plot demonstrating this transient signal is shown in Figure 10. Quantitation is performed by measuring either the peak height or the peak area of this transient signal. Peak area usually provides the best sensitivity and precision. Transient signal processing greatly complicates simultaneous multielement determinations. Because different analytes have a variety of analysis characteristics, such as variable volatility, it is often difficult to select a compromise set of operating conditions when processing transient signals.

High sensitivity is obtained by this technique, with detection limits approximately an order of magnitude superior to those obtained with ultrasonic nebulization

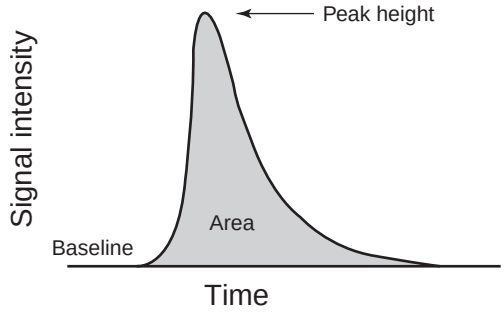

Figure 10 Plot of transient signal.

(see Table 3). Typical size of sample deposited on the electrothermal vaporizer is in the 1–200-μL range, making this technique suitable for limited sample applications.

6.2 Solids

The direct introduction of solid sample into the plasma offers some significant advantages over the traditional chemical dissolution of solid materials followed by liquid sample introduction. For solid sampling techniques, minimal sample preparation is required. This approach not only saves time for the routine analysis of large numbers of samples, but it reduces the possibility of contamination of analyte species from other samples or from dissolution reagents. Because no solvent is present for direct solid analysis, molecular species such as ArO^+ are present at low concentrations and therefore are of minimal significance. Also, where appropriate for specific applications, some solid sampling techniques provide spatial resolution of the sample, providing lateral or depth profiling information.

In addition to the slurry nebulization technique previously described, solids can be introduced into the plasma by several techniques. These techniques include: direct insertion devices, electrothermal vaporization, arc or spark ablation, and laser ablation. Each of these techniques has unique characteristics that can solve specific trace analysis problems advantageously.

A direct sample insertion technique involves placing a powdered sample on a probe, which is inserted directly into the base of the plasma. The sample is ground sufficiently finely for milligram-size aliquots to be representatively sampled. The probe usually consists of a graphite rod with a cup formed at the end to hold the sample. Some types of apparatus employ a mechanical device to reproducibly insert the probe axially into the base of the plasma through the bottom of the plasma torch assembly. While the probe is being inserted, the sample is sequentially vaporized, decomposed, atomized and ionized. This is an extremely efficient means of introducing solid samples into the plasma.

Detection limits for this technique are easily achieved at the nanogram per gram level for most elements. Detection limits are quite variable depending on operating conditions, types of sample and element volatility. Often, thermochemical reagents are added to the sample to improve the analysis characteristics of refractory or carbide-forming elements such as aluminum, uranium and zirconium. Precision is of the order of 3–10% RSD. This relatively poor precision is attributed to sample inhomogeneity, which is exacerbated by the small sample size used by this technique.

Electrothermal vaporization, previously discussed under solution techniques, can also be used for direct

solid sample introduction. It is not a very popular technique because of the many problems associated with its utilization. Maximum sample size is restricted to about 1 mg. As with the direct insertion device, this size of sample presents problems in its ability to achieve statistically representative subsampling. This is reflected in the relatively poor precision (10% RSD) that can be obtained by using the technique. In addition to poor precision, poor accuracy is also characteristic of this technique because of nonrepresentative sampling for trace constituents and the lack of appropriate calibration standards.

Like the direct insertion device, electrothermal vaporization (for both liquid and solid samples) produces transient signals. It suffers similar difficulties to the direct insertion device. Coupling the electrothermal vaporization device with a time-of-flight mass spectrometer provides a very fast data acquisition technique that could produce many multielement measurements over the timeframe of the transient signal. This approach greatly improves the multielement analytical capabilities of electrothermal vaporization.

Arc and spark ablation consist of the application of an electrical current between the surface of an electrically conducting sample and a counter electrode. This ablation process produces a plume of volatilized and aerosol particulate sample that is transported by a carrier gas to the plasma for atomization and ionization. The main difference between an arc and a spark is the timescale of the discharge. The spark is a high-voltage (1 kV), low-current (mA), high-frequency discharge with a pulse width of between 1 μs and 1 ms, which consumes a small quantity of the surface of the sample (1% or less). Arcs, on the other hand, are low-voltage (40 V DC), high-current (10 A) discharges that are continuous in nature and can consume up to 50% of the sample. Depending on the operating conditions and the nature of the sample, material can be either directly vaporized or sputtered.

Since the spark technique works with a relatively large sample and operates in a continuous mode, a steady state rather than transient signal is obtained. This is analogous to constant introduction of a solution from a nebulizer, greatly simplifying data processing. This technique finds popular application in the trace analysis of pure metals and alloy samples. In addition, nonconducting materials, such as powdered geological rock samples, are analyzed by mixing with graphite and pressing into briquettes.

Finally, laser ablation has been developed into the most versatile solid sampling technique. Energy in the form of a focused infrared or ultraviolet laser beam is directed onto a solid sample surface, where a discrete portion is sputtered and vaporized. The sample specimen is housed in an enclosed cell with the laser beam focused on it through a window. As with the spark ablation system, the dry aerosol and/or sample vapor produced by the laser interacting with the specimen is transported to the plasma for ionization. By carefully controlling the operating characteristics of the laser beam (i.e. laser power, wavelength, repetition rate, etc.), the size and depth of the ablation crater can be controlled (5–60 μm diameter). This permits a spatially discrete amount of material to be sampled, resulting in a transient signal. By using an optical positioning apparatus, such things as grain boundaries or inclusions can be selectively analyzed. Repetitive sampling at a specific location can provide depth-profiling analysis of the sample.

Average composition or bulk analysis of the sample can be performed by rastering the laser beam across a fixed area of the sample surface. This approach provides an improvement in trace element sampling statistics and minimizes problems of sample inhomogeneity. Although the sensitivity for trace analysis by laser ablation ICPMS is not as good as that obtained by sample dissolution followed by pneumatic nebulization, the absolute detection limits are far superior.

Precision can be improved greatly by fusion of the sample with a flux such as lithium metaborate, forming a glass with a homogeneous composition and a flat surface for laser sampling. This approach also provides a convenient way of preparing calibration standards. A series of fusions are prepared with differing concentrations of the analyte elements. These fusions can then be used to produce calibration curves for quantitative analysis.

The laser ablation approach provides a powerful method for the analysis of solid samples requiring minimal sample preparation. It is especially useful for the analysis of materials that are difficult to dissolve, such as refractory ceramics or minerals. It also provides the unique ability to perform microspatial trace analysis of complex sample materials that are essentially impossible to analyze by other techniques.

6.3 Gases

Sample in the gaseous state is the easiest to introduce into the plasma for analysis, because it can be added directly to the plasma argon carrier gas. Generally the scope of analysis is limited to a single element or a small group of elements with similar chemical characteristics. Chemical reactions in solution are often used to generate analyte species in a gaseous form that are transported to the plasma from the reaction vessel. These reactions can be used to isolate the analyte species from the matrix of the sample and can also serve to provide a level of preconcentration.

The simplest chemical reaction is the reduction of mercury compounds to elemental mercury, which has a high vapor pressure at room temperature. Reducing

agents such as stannous chloride in hydrochloric acid or sodium borohydride quantitatively reduce dissolved inorganic species to the elemental mercury state. This mercury vapor is separated from solution and transported in the argon stream to the plasma for ionization.

Conversion of analyte species to molecular hydride compounds, which are volatile at room temperature, is also a suitable method for the determination of selected elements. Hydride generation is an effective technique for isolation and enhancement of the detection capability for elements such as antimony, arsenic, germanium, selenium, tellurium and tin. The most common reducing agent is sodium borohydride in acidic solution. The gaseous hydride compound is separated from the aqueous solution by sparging with argon gas or by using a gas–liquid separator. After separation, the hydride gas is transported to the plasma via the argon gas carrier stream, where atomization and ionization take place.

Gas chromatography coupled to an ICP has also been used for the determination of various volatile organometallic compounds. The compounds are separated in the chromatographic column and the effluent from the column is mixed into the argon gas carrier stream for introduction into the plasma. This approach is particularly important for speciation applications.

7 COMPARISON OF CHARACTERISTICS OF INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY WITH INDUCTIVELY COUPLED PLASMA ATOMIC EMISSION SPECTROMETRY

The ICP has been effectively used as an atomic emission spectroanalytical source for trace multielement analysis for many years. The configuration and operating characteristics of the plasma are essentially identical regardless of whether it is being used for inductively coupled plasma atomic emission spectrometry (ICPAES) or atomic mass spectrometry (ICPMS). Both techniques offer advantageous characteristics for analytical applications. Because the plasma is essentially the same, the sample introduction techniques described above can be effectively used by either approach.

The primary difference between the two techniques is that with ICPMS ions generated by the high-temperature plasma are sampled with an extraction cone and measured by the mass spectrometer. Comparatively, with the ICPAES technique, atoms and ions in the high-energy plasma are excited and emit optical radiation, which is focused with a lens or mirror onto the entrance slit of the spectrometer, where specific characteristic wavelengths are isolated and measured.

Although both techniques provide the ability to determine essentially all elements in the periodic table, ICPMS provides the additional capability of measurement of specific stable isotopes. Both techniques provide a wide linear dynamic measurement range, covering three to five orders of magnitude. Use of multiple emission wavelengths with the ICPAES technique allows the analyst to select the spectral line with the most appropriate intensity to provide optimal signal-to-noise ratios (i.e. low-intensity lines for measurement of high concentrations and high-intensity lines for measurement of low concentrations). Although this approach is usually not available with ICPMS (relatively few uninterfered isotopes are available for analytical measurements), sophisticated multiple detector instrumentation permits a wide dynamic range signal to be processed, which extends the working range over many orders of magnitude (eight or more). With either technique, concentration of elements ranging from major components of the sample to ultratrace constituents can be measured essentially simultaneously on a single sample aliquot. Generally detection limits for ICPMS ($1–50\,\mathrm{ng\,L^{-1}}$) are one to two orders of magnitude superior to those obtained for ICPAES.

A serious problem with ICPAES is the potential for spectral interference, which can have a serious impact on the accuracy of the analysis. Spectral overlap between high-concentration components in the sample and the trace analyte emission lines can preclude or degrade the quality of analytical measurements. In addition, molecular band emission spectra and continuum background radiation can have a serious impact on accuracy and detection limits. The relatively simple mass spectrum obtained with inorganic mass spectrometry with very low backgrounds, including the low prevalence of polyatomic molecular species and the relatively few isobaric interferences, can be alleviated with the use of high-resolution mass spectrometers, mitigating a similar problem in ICPMS.

The combination of high sensitivity (low detection limits), fewer interferences, excellent accuracy and good precision, combined with the ability to make independent stable isotope measurements, makes ICPMS the most powerful tool for trace elemental analysis. The use of numerous sample introduction techniques provides the versatility to solve most of the analytical chemistry problems facing the modern analytical chemist.

ABBREVIATIONS AND ACRONYMS

DC	Direct Current
EIMS	Electron Ionization Mass Spectrometry
GDMS	Glow Discharge Mass Spectrometry

ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
RF	Radiofrequency
RSD	Relative Standard Deviation
SSMS	Spark Source Mass Spectrometry
TIMS	Thermal Ionization Mass Spectrometry

RELATED ARTICLES

Coatings **(Volume 2)**
Microscopy of Coatings

Environment: Trace Gas Monitoring **(Volume 3)**
Laser Mass Spectrometry in Trace Analysis

Environment: Water and Waste **(Volume 3)**
Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis

Atomic Spectroscopy **(Volume 11)**
Glow Discharge Optical Spectroscopy and Mass Spectrometry • Laser Ablation in Atomic Spectroscopy

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Electron Ionization Mass Spectrometry • Isotope Ratio Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Time-of-flight Mass Spectrometry

Nuclear Methods **(Volume 14)**
Elemental Analysis by Isotope Dilution

FURTHER READING

F. Adams, R. Gijbels, R. Van Grieken (eds.), *Inorganic Mass Spectrometry*, Chemical Analysis Series, John Wiley and Sons, New York, 404, Vol. 95, 1988.

A.J. Ahearne (ed.), *Trace Analysis by Mass Spectrometry*, Academic Press, New York, 1972.

A.R. Date, A.L. Gray (eds.), *Applications of Inductively Coupled Plasma Mass Spectrometry*, Blackie, London, 254, 1989.

W.W. Harrison, B.L. Benz, 'Trace Element Analysis by Glow Discharge Mass Spectrometry', in *Trace Element Analytical Chemistry in Medicine and Biology*, eds. P. Bratter, P. Schramel, Walter de Gruyer, New York, 1980.

S. Hill (ed.), *Inductively Coupled Plasma Spectrometry and Its Applications*, Sheffield Academic Press, CRC Press, Boca Raton, FL, 119–199, 1999.

G. Holland, A.N. Eaton (eds.), *Applications of Plasma Source Mass Spectrometry*, Royal Society of Chemistry, London, 222, 1991.

G. Holland, A.N. Eaton (eds.), *Applications of Plasma Source Mass Spectrometry II*, Royal Society of Chemistry, London, 243, 1993.

K.E. Jarvis, A.L. Gray, R.S. Houk, *Handbook of Inductively Coupled Plasma Mass Spectrometry*, Blackie, Glasgow, 380, 1992.

A. Montaser (ed.), *Inductively Coupled Plasma Mass Spectrometry*, Wiley–VCH, New York, 964, 1998.

J. Roboz, *Mass Spectrometry*, Interscience, New York, 398, 1968.

Isotope Ratio Mass Spectrometry

H. Roy Krouse
University of Calgary, Canada

1 Introduction **1**
2 Basic Magnetic Sector Isotope Ratio Mass Spectrometer **4**
2.1 Vacuum System 4
2.2 Source 6
2.3 Gas Source 6
2.4 Thermal Ionization Source 8
2.5 Analyzer Region 11
2.6 Collector Design and Measurement of Ion Currents 12
2.7 Automation of Magnetic Sector Mass Spectrometers 15
3 Continuous Flow Isotope Ratio Mass Spectrometry **16**
3.1 Basic Concept 16
3.2 Bulk Sample Isotope Analysis 17
3.3 Multielement Isotope Analysis 17
3.4 Compound-specific Isotope Analysis 18
3.5 Chemical Reactors 18
3.6 Laser Probe 18
4 Secondary Ion Mass Spectrometers **19**
5 Inductively Coupled Plasma Mass Spectrometers **19**
6 Accelator Mass Spectrometers **20**
7 Other Isotope Ratio Mass Spectrometers **21**
8 Applications **22**
8.1 Introduction 22
8.2 Concentration Measurements 22
8.3 Radioactive Decay and Nuclear Reactions 22
8.4 Isotopic Tracers 23
8.5 Isotopic Fractionation 23
8.6 Extraterrestrial Applications 24
Abbreviations and Acronyms **25**
Related Article **25**
References **25**

With the separation of two of the three stable isotopes of neon by Sir J.J. Thomson in 1913, mass spectrometers have continued to be the instrument of choice for measuring ratios of isotope abundances. The first magnetic sector mass spectrometer specifically dedicated to this task was designed by A.O. Nier in 1947 and featured simultaneous collection of ion currents. Many diverse instruments have evolved from this design. Decades of development of sample preparation techniques have realized isotope ratio data from many elements in solids, liquids, and gases. Coupling of devices such as gas chromatographs, combustion apparati, and laser probes to mass spectrometers in concert with computer control, have realized features such as unattended operation, compound-specific isotope analyses and isotopic data spatially resolved over distances of a few micrometers on solid surfaces.

Today, applications of isotope ratio mass spectrometry (IRMS) embrace many disciplines with topics such as paleodiets, food adulteration, paleoclimatology, migration of birds and animals, pollutant tracing, ore and oil deposits, meteorology, energy expenditure of animals, including astronauts, and the origin of the universe.

1 INTRODUCTION

The term 'isotope' is derived from the Greek words *isos* and *topos* to denote the 'same place' in the periodic table.[1] Thus, isotopes are atoms whose nuclei contain the same number of protons but different numbers of neutrons. Isotopes are designated by a superscript which denotes the nucleon number (protons plus neutrons), placed to the upper left (upper right in earlier literature) of the symbol for the element, e.g. ^{1}H, ^{2}H, ^{3}H for the three isotopes of hydrogen. In contrast to the practice with other elements, hydrogen's isotopes have been dubbed hydrogen (or protium), deuterium, and tritium and are often given the symbols H, D, and T.

Only 21 elements have one stable isotope, whereas element 50, Sn, possesses the most, 10. Of nearly 400 naturally occurring isotopes, 271 are stable or have half-lives many orders of magnitude larger than the age of our Earth. Their existence and abundances are determined by nuclear stability principles, including the Oddo–Harkins classification (Table 1) based on even versus odd numbers of protons and neutrons in the nucleus. Isotopes differ in fundamental physical properties such as mass, nuclear spin, and nuclear stability. Although isotopes of a given element participate in equivalent chemical reactions, they do so with different rate constants because of differences in isotopic vibrational frequencies and zero point energies of molecules. Many physical processes are also mass dependent and hence isotopically selective.

Alteration of isotope abundances is termed isotope fractionation or an isotope effect. These terms have been used interchangeably but there is a tendency to identify

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Frequency of occurrence of stable isotopes

Proton number \ Neutron number	Even	Odd
Even	160	56
Odd	50	5

the former with differences in isotope compositions among components in a system and the latter with a mass-dependent process.

The term 'isotope fractionation factor', α, is defined by Equation (1),

$$\alpha_{A-B} = \frac{R_A}{R_B} \tag{1}$$

where R is the ratio of the heavy to light isotope abundances (D/H, $^4He/^3He$, etc.) in compounds A and B. The species A and B may also be different phases of the same molecule, such as liquid and gaseous H_2O.

Because of the difficulty of determining absolute isotope abundances, relative abundance ratios are reported using δ-values (historically called 'del' or 'delta values' with a decided preference for the latter to be consistent with the Greek symbol). The δ-values for the stable isotopes of hydrogen are defined (Equation 2) as

$$\delta D(\text{in parts per thousand}) = \left(\frac{[D/H]x}{[D/H]s} - 1 \quad \times 1000 \right. \tag{2}$$

where x and s denote the sample and an internationally accepted standard, respectively. The term D/H is the ratio of the numbers of deuterium and protium isotopes. The scale is in parts per thousand, ‰, and written as 'permil', 'per mil', 'per mille', and other language-dependent alternatives. Positive and negative δ-values mean that the sample is enriched and depleted, respectively, in the heavier isotope, relative to the standard.

There are similar δ-value definitions for other elements; for example, $\delta^{18}O$ in terms of $[^{18}O/^{16}O]$, $\delta^{34}S$ in terms of $[^{34}S/^{32}S]$, etc.

The δ-value comprises a ratio of ratios and is consistent with the measurement strategy of IRMS. The objective is to compare a sample with a calibrated laboratory standard under identical conditions so that measurement biases cancel in the numerator and denominator of the δ-value definition. The precision of δ-value determinations can be better than ±1‰ for hydrogen and better than ±0.1‰ for other elements. For agreement among different laboratories, a variety of reference materials have been prepared and distributed by the National Institute of Standards and Technology (NIST) in the USA and the International Atomic Energy Agency (IAEA) in Vienna.[2] Most researchers use these reference materials to calibrate working standards and to ensure that data are reported on scales consistent with other laboratories.

The fractionation factor is related to δ-values by Equation (3):

$$\alpha_{A-B} \approx \frac{1000 + \delta_A}{1000 + \delta_B} \tag{3}$$

Isotope fractionation can be identified with reversible exchange reactions and unidirectional kinetic effects. An exchange reaction can be generally written as Equation (4),

$$aA_1 + bB_2 \longleftrightarrow aA_2 + bB_1 \tag{4}$$

where A and B are as defined above, a and b are the number of moles of A and B, and subscripts 1 and 2 denote the lighter and heavier isotopic species, respectively. The equilibrium constant for this reaction is given by Equation (5):

$$K = \frac{(A_2/A_1)^a}{(B_2/B_1)^b} \tag{5}$$

Bringing compounds together does not ensure that exchange will occur or that isotopic equilibrium will be established. Equilibrium for oxygen isotope exchange between CO_2 and H_2O is realized in a few hours at room temperature,[3] whereas oxygen isotope exchange between SO_4^{2-} and H_2O has a half-time of 10^5 years under ambient neutral solution conditions.[4] The heavier isotopes are usually favored in the compound in which the element is in the higher valence state (lower zero-point energy) and in the more condensed phases. Using methods of statistical thermodynamics, isotopic partition function ratios can be calculated and used to evaluate equilibrium constants as a function of temperature.[5,6] Additional to their importance in nature and commercial isotope enrichment operations, exchange reactions are used in analyses of isotope compositions. For example, O- and H-isotope compositions of water can be based on measurements with CO_2 and H_2 that have been equilibrated with H_2O samples.[3,7]

A kinetic isotope effect (KIE) is identified with a unidirectional process and is expressed in terms of two competing reactions (Equation 6):

$$A_1 \xrightarrow{k_1} P_1 \quad \text{and} \quad A_2 \xrightarrow{k_2} P_2 \tag{6}$$

where A and P correspond to reactant and product. Although the term has been used with a variety of physical, chemical, and biological conversions, in a stricter sense it is defined as the ratio of the isotopic rate constants k_1/k_2 in one step of a chemical reaction. Lighter isotopic bonds have higher vibrational frequencies corresponding to higher ground-state energies and therefore lower

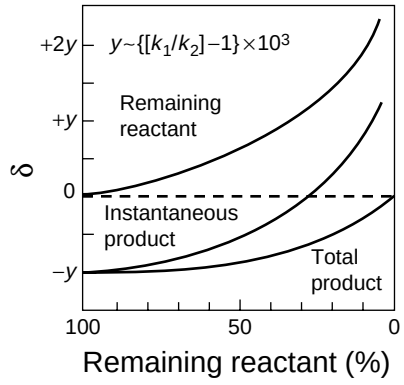

Figure 1 Isotopic behavior of components in a one-step first-order reaction with a KIE k_1/k_2.

energy requirements for rupture. Transition state theory provides a basis for calculation of k_1/k_2 values.[8]

As an isotopically selective reaction proceeds, the isotopic compositions of components in the system change depending upon whether it is open or closed to reactant and product. For a one-step first-order process illustrated in Figure 1, the equations derived by Lord Rayleigh for distillation of mixed liquids can be applied. With the lighter isotope favored in the product, the remaining reactant becomes more relatively enriched in the heavier isotope as the conversion proceeds. The product formed at any instant has a δ-value which is approximately $[(k_1/k_2) - 1] \times 10^3$ lower than that of the residual reactant. The instantaneous curve in Figure 1 would apply to a system with the product escaping upon its formation. If all the product is retained in the system, its isotopic composition at 100% conversion is the same as that of the initial reactant as shown by the total product curve in Figure 1.

Many approximations to the curves in Figure 1 have been derived.[9,10] For example, changes in the isotopic composition of the remaining reactant can be described by Equation (7) or Equation (8),

$$\frac{A_1}{A_2} = \left(\frac{A_1}{A_2}\right)_{\text{initial}} f^{(\alpha-1)} \tag{7}$$

$$\delta = \delta_{\text{initial}} f^{(\alpha-1)} \tag{8}$$

where f is the fraction of unconverted reactant.

Mass-dependent isotope fractionation is also associated with physical phenomena including diffusion, evaporation, melting, absorption, heat capacity, and crystallization. Not only can isotopic abundance data be used to better understand these processes, but the fractionation may be relevant to the design of instrumentation used to measure isotope abundances (sections 2.3.2 and 2.4). Stable isotope abundances of an element may be also substantially altered by nuclear reactions (section 7).

In principle, there are many ways to determine isotope abundances. Relatively simple concepts such as density and heat capacity determinations are difficult to effect in practice for many reasons including the requirement of very pure substances. As molecular vibrational and rotational frequencies are isotopically dependent, some applications such as ^{15}N-labeling studies routinely use optical emission spectroscopy.[11] Nuclear magnetic resonance (NMR) spectroscopy has the distinct advantage of being able to determine the isotope compositions of specific sites in molecules. Although there has been considerable success with D/H measurements,[12] many environmentally important isotopes such as ^{16}O and ^{32}S have zero nuclear spin and therefore are not NMR responsive.

The instrument most widely used for isotope abundance determinations is the mass spectrometer. Historically, the challenge was the discovery of stable isotopes with Sir J.J. Thomson separating two of the three stable isotopes of neon in his positive ray apparatus in 1913.[13] A few years later, Aston[14] and Dempster[15] built improved mass spectrographs to discover and determine the abundances of many isotopes. Aston identified the third stable isotope of He and in his lifetime discovered 212 naturally occurring isotopes.[16]

From the principles of isotope fractionation outlined above, it is seen that it is highly desirable to measure ratios of isotope abundances (referred to as isotope ratios) in specimens. The first magnetic sector mass spectrometer specifically dedicated to isotope ratio determinations was a 60° magnetic analyzer (MA) designed by Nier.[17,18] Many diverse instruments have evolved from this design.[19–21]

Prior to the 1960s, most stable isotope mass spectrometers were constructed by individual researchers throughout the world. Understandably, current commercial units are expensive because of the overhead requirements for manufacturing such specialized equipment. The financial stability of companies willing to undertake the task was constantly challenged and most changed ownership over time. Finnigan-MAT, Germany (Atlas-MAT), and Vacuum Generators, UK (Micromass, Fisons), account for the bulk of the installations globally. Earlier instruments in the USA included many produced by Nuclide and a few solid source units by AVCO. In other countries, instruments were offered by Thomson (France) and Hitachi (Japan). Institutions in the former USSR had mass spectrometers produced within the Soviet Academy of Science. Europa Scientific, UK, which emerged in recent years, has specialized in continuous flow (CF) instruments (section 3).

Specific commercial instruments will not be discussed as such but rather the fundamental concepts underlying components of isotope ratio mass spectrometers.

2 BASIC MAGNETIC SECTOR ISOTOPE RATIO MASS SPECTROMETER

A magnetic sector IRMS consists of a vacuum system, source, MA region, ion collector assembly, and ion detection electronics.

2.1 Vacuum System

Flight tubes of mass spectrometers are operated at high vacuum ($\sim 10^{-8}$ torr) to minimize collisions of the ion beams with ambient molecules. (1 torr is approximately the pressure due to a 1 mm column of mercury and equates to 133.3 Pa or 1.3 mbar.) Such low pressures can be realized with a number of pumping systems.[22–25]

Most systems require a roughing pump. Roughing pumps are usually oil-sealed mechanical units capable of realizing pressures in the 10^{-3} torr region. During their rotation, a portion of low-pressure gas in the system is trapped at its inlet, compressed, and expelled to the atmosphere (Figure 2a). A trap (e.g. molecular sieve) is placed at its input where it serves as a fore or backing pump for high-vacuum pumps.

There are four general types of high-vacuum pumps which operate on different principles: vapor-stream or diffusion pumps; turbo-molecular pumps; chemical/ion pumps; and cryogenic pumps. Diffusion pumps have boiling heavy liquid reservoirs (Figure 2b). The vapor is directed through jets to impart momentum to molecules in the system so that they move towards the inlet of the fore pump. Subsequently, the vapor is condensed by a water-cooling jacket and returned to the reservoir. The working fluid must be thermally stable, chemically nonreactive, and have a low vapor pressure. Mercury used in earlier systems required a liquid N_2 trap at the pump's inlet to prevent its condensation in the mass spectrometer. Since the early 1980s, polyphenyl ethers have been used successfully in diffusion pumps and do not require liquid N_2 trapping.

Turbomolecular pumps (Figure 2c) consist of a metal housing containing a rotor and stator with obliquely slotted discs. They can be considered as turbine engines operating in reverse. The rotor is driven at speeds as high as 60 000 rpm and molecules are forced to each end of the port. Stator–Rotor disc pairs act as pumping stages in series so that gas in the system at pressures of the order of 10^{-9} torr is compressed to 10^{-2} torr and introduced to a backing pump. The pressure ratio increases with the molecular weight of the gas so pumping of H_2 and He is much slower. Lubrication of the rotor bearings is a challenge. The use of oil introduces organic vapors and cooling of the bearings is necessary. The situation has been improved by using ceramic bearings and lubrication can be avoided entirely by magnetic levitation of the rotor.

Chemical pumps form a chemical bond between the pumped gas and an absorbent. In a typical ion pump (Figure 2d), a stainless steel grid is placed between two titanium cathode plates. A potential difference of about 3 kV is applied to the electrodes which sets up a cold-cathode discharge. Positive ions impacting on the cathode sputter titanium onto surfaces of the anode grid. A magnetic field ($\sim$0.1 T) normal to the electrode sandwich causes electrons to move in spiral paths, thereby increasing the ionization efficiency. Ion pumps are quiet, have low power consumption and, after starting, do not require a backing pump. The ion current is a measure of the pressure, whereas other systems require a vacuum gauge. The pump fails when the cathode has become saturated with buried atoms so the lifetime (order of 5 years at 10^{-8} torr) is inversely proportional to the pressure. The anode and cathode can be electrolytically cleaned and reused a few times. During operation, buried atoms can be released from the cathode with subsequent bombardment and chemically inactive gases such as Ar are particularly prone to release.

In one version of a cryogenic pump or cryopump, an internal element cooled by circulating liquid He, called a cryopanel, condenses gases in the system. Effective pumping requires that the built-up solid molecular layers possess a low vapour pressure and high sticking probability for incident molecules. The pumping element may be enclosed in a liquid N_2 cooled baffle so that it does not receive radiation from the inner walls of the system (Figure 2e). The cryopanel can be regenerated by warming and pumping with a fore pump. A fore pump is also used before starting the cryopump, to avoid a thick condensation layer from pumping air at atmospheric pressure. Cryopumps have much higher speeds for a given size in comparison to other high-vacuum pumps. The species He, H_2, and Ne are difficult to condense. Steps may be required to isolate the mass spectrometer from vibrations caused by the circulation of liquid He.

Sorption pumps (Figure 2f) differ from chemical and cryopumps in that molecules are sorbed as a monolayer on charcoal or zeolite (molecular sieve) pellets usually immersed in a liquid-N_2 trap. The sorbed molecules are released by mild heating. These pumps can be used in place of mechanical backing pumps during the start-up of ion pumps, thereby eliminating organic contamination in systems.

Realizing high vacuum is not only dependent upon the pumping system but also upon the plumbing. Piping to pumps should have diameters as large and lengths as short as possible. Mass spectrometer tube assemblies are usually constructed with nonmagnetic stainless steel and inert gas-welded joints. The latter should be constructed so that pockets cannot form, where trapped gases are

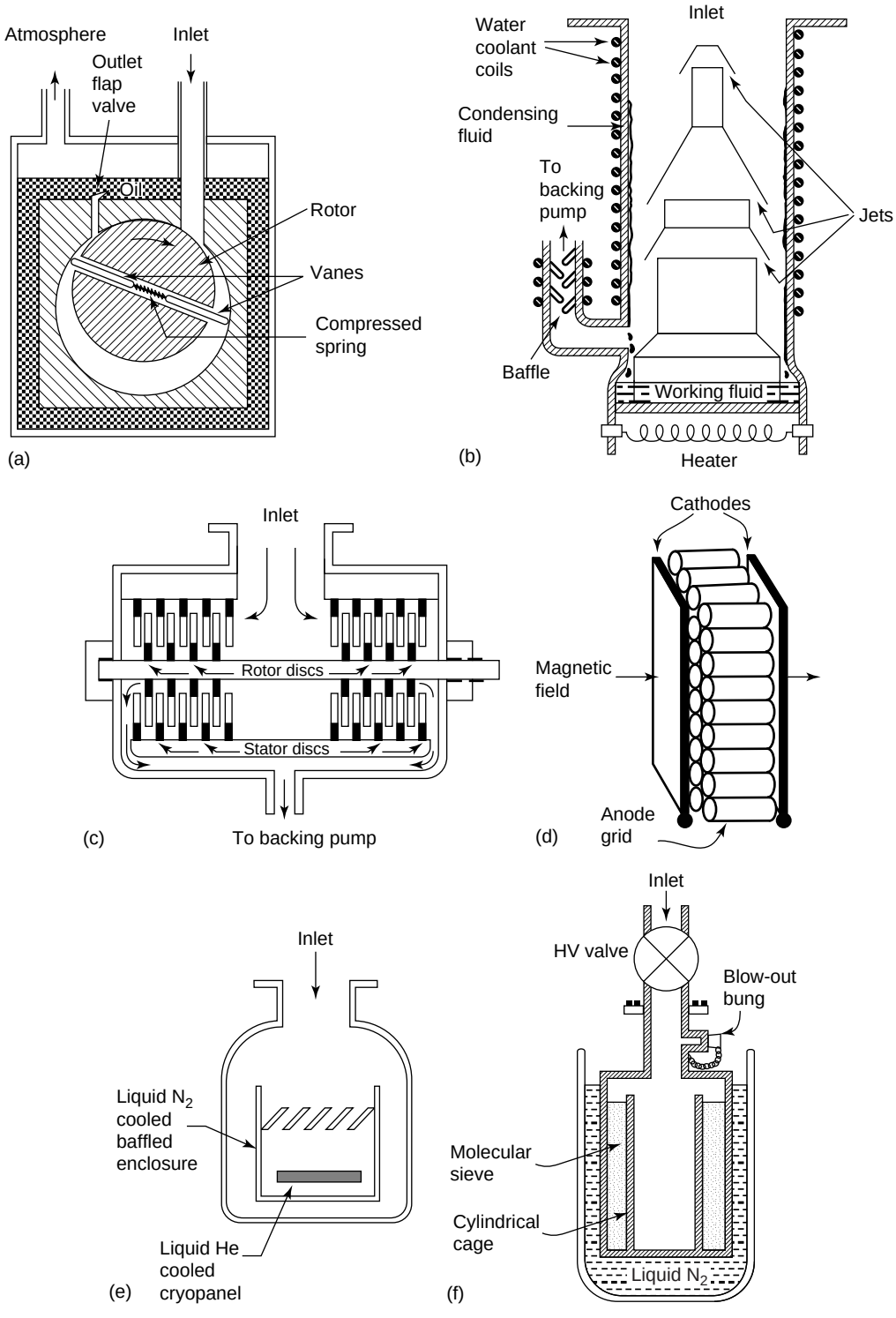

Figure 2 Vacuum pumps: (a) oil sealed rotary vane mechanical pump, (b) diffusion pump, (c) turbomolecular pump, (d) interior of diode ion pump, (e) cryopump, (f) sorption pump.

difficult to pump out. Coupling of sections is usually done with flanges with knife-edge annuli pushing into copper gaskets. If reactive gases such as SO_2 are involved, gold O-rings compressed between flat flanges can be used. For flanges that are frequently uncoupled, Viton® O-rings are another possibility.

Electrical insulators inside the source and collector as well as feedthroughs should be made of glass or ceramic

materials, rather than of polymers which may outgas over long periods.

Dependent upon the sample introduction system, pressures in the source region can greatly exceed 10^{-6} torr. Hence, it is desirable to separate it from the flight tube and collector; the only necessary connection being the source exit slit. In this way, the former may be pumped with a fast system such as a turbomolecular pump and the latter maintained at 10^{-8} torr with perhaps a modest ion pump.

2.2 Source

Mass spectrometer sources serve two functions: generation of ions from samples in solid, gas, or liquid states; and formation of an ion beam that is directed down the flight tube into the analyser region. Many combinations of nonmagnetic stainless steel plates with slits or other features have been used to form ion beams.

The ionization region is usually placed at a potential of 2–10 kV above ground. A variation of a Nier-type source is shown in Figure 3.[26] The plate below the ionization region is termed the drawing out plate (DOP) and its potential is such that some equipotential surfaces curve upwards into the ionization region. The corresponding electric field lines curve inwards as they go downwards towards the slit of the DOP. Thus ions from a cross-sectional area in the ionization region larger than that of the DOP slit, pass through the latter.

Two half-plates below the DOP are usually at slightly different potentials to bend the ion beam so that transmission through the grounded exit slit is maximized. The DOP–half-plate combination may have simple slits or shaped elements such as the Nier thick lens shown in Figure 3.

In simple designs, consecutive grounded plates with different slit widths may serve to geometrically focus or collimate the ion beam. In other designs, there may be a collimator plate with slit held at a potential intermediate between the half-plates and grounded exit slit (Figure 3).

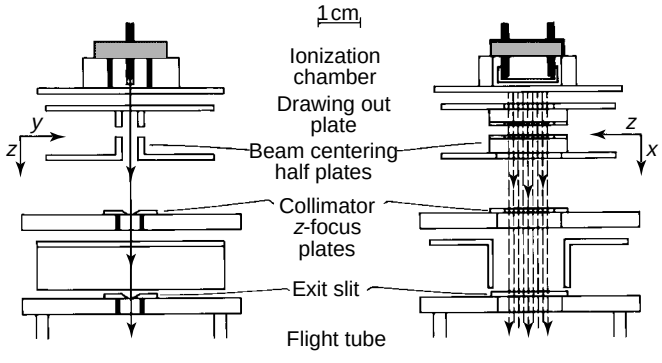

Figure 3 Single-filament thermal ionization (TI) mass spectrometer source featuring a Nier thick lens and z-focus plates. A gas source can be made by replacing the filament housing with an electron impact (EI) ionization chamber (Figure 4). Electrical insulation spacers between plates are not shown.

Another option shown in Figure 3, is a pair of z-focus plates which reduce spreading of the length of the ion beam's cross-section. Slightly different potentials on these plates can shift the ion beam along the exit slit length-wise (referred to as tilt) to maximize transmission.

To a first approximation, an ion current leaving the source is proportional to the flux of the corresponding molecular species entering the ionization region. Change in the proportionality constant with sample flux is referred to as nonlinearity. A perfectly linear source is unobtainable because of phenomena such as space charge distribution which depend on ion density. Nonlinearity in IRMS is not limited to the source, and isotope abundance data should be obtained with the same major ion current for all samples.

In the source, an ion acquires kinetic energy given by Equation (9):

$$Vq = \tfrac{1}{2}mv^2 \tag{9}$$

where V is the potential difference between the position where the ion is formed and the exit plate which is usually at ground, q is the charge, m is the mass, and v is the speed.

2.3 Gas Source

Gases may be ionized in a number of ways (e.g. ultraviolet radiation), but the preferred method is by electron ionization (EI) (also termed electron bombardment) because of its higher efficiency.

2.3.1 Electron Impact Assembly

A typical EI assembly developed by Nier and others,[17,18] is shown in Figure 4. Electrons emitted by a filament are accelerated by a potential difference of 50–100 V and pass as a beam into an ionization box, referred to as the case or shield. Gases passing through the opening shown at the top of the case in Figure 4, collide with the electron beam and are ionized. The electron beam exits the case and enters a trap held at a potential of about 20 V above that of the case. The filament electronics supply uses the trap current (~200 μA) in a feedback loop to control the filament current and hence electron emission.

During electron bombardment, the majority of the produced ions are positive. Not only are valence electrons knocked from molecules, but ion fragments are also produced dependent upon the electron energy. Increasing the potential difference between the filament and case increases the total ion current through a maximum and also tends to produce more fragments, as shown for GeF_4

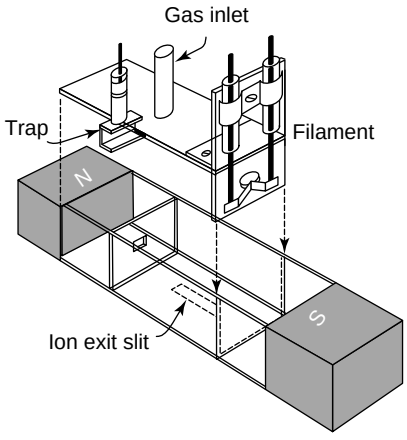

Figure 4 EI ionization chamber for gases.

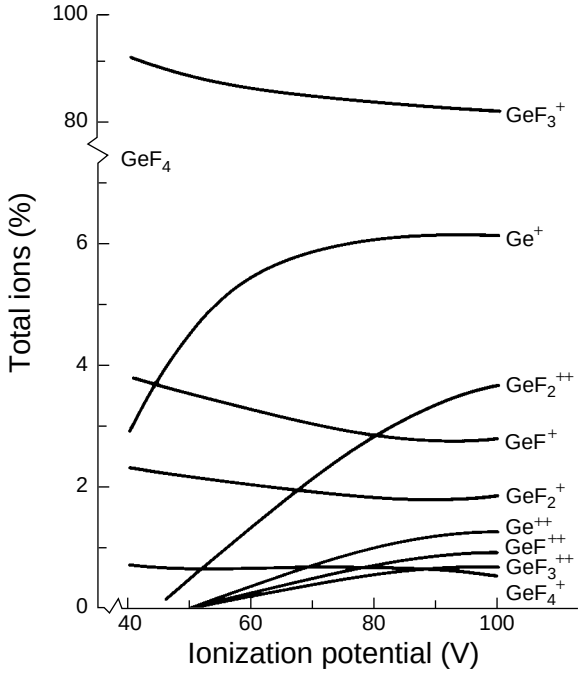

Figure 5 Percentage yields of various ion species as a function of electron energy during electron bombardment of GeF_4. (Unpublished results, H.M. Brown (1961).)

in Figure 5. Interestingly for GeF_4 and other fluoride gases such as SiF_4, SF_6, SeF_6, and TeF_6, removal of one electron is much less probable than release of a F atom. With SO_2, the maximum SO_2^+ production is of the order of 60% of the total ions. As by suitable choice of electron energy the SO^+ production can exceed 40%, some workers use that species for sulfur isotope abundance determinations especially if the IRMS has low resolution.

Application of a magnetic field along the direction of the electron beam causes electrons diverging from the path to follow tight helices. This can improve the ionization efficiency by a factor of 10 and the narrowing of the beam results in ions being formed more or less along an equipotential surface. The ions leave through a slit shown at the bottom of the case in Figure 4 along the electric field lines created by the DOP (section 2.2). Some EI assemblies feature a repeller above the electron beam which can be adjusted a few volts positive or negative with respect to the case to maximize expulsion of ions.

For some gases such as UF_6, adsorption and precipitation of solids on the inner walls of the case have been addressed with a more open case construction.[27] Reactions between molecules and ions may require special attention. For example, during EI ionization of H_2, H_3^+ ions are formed which are not resolved from the HD^+ ion beam in a typical low-resolution IRMS. The $[H_3^+]/[H_2^+]$ value has a linear dependence on $[H_2^+]$ and hence the H_2 pressure but also varies with the repeller voltage. Through measurements at different pressures, the H_3^+ contribution can be determined for the particular operating conditions.

Because the EI source is ideally suited for gases and even vapors emitted from solids, it is the most widely used in mass spectrometry (MS). It has also been modified for special purposes such as an open ion source in space flights.

2.3.2 Gaseous Sample Introduction

A distinct advantage of a gas source IRMS is sample introduction without breaking the vacuum. Samples should be prepared as pure as possible. In some cases, certain impurities cannot be tolerated. For example, N-isotope analyses with $^{14}N_2^+$ and $^{14}N^{15}N^+$ are adversely effected by even small amounts of CO_2 or CO because of $^{12}C^{16}O^+$ and $^{13}C^{16}O^+$ contributions at masses 28 and 29. A further requirement might be isotopic consistency of some elements in the gas being analyzed. For S-isotope analysis of SO_2, $^{34}S^{16}O_2^+$ and $^{32}S^{18}O^{16}O^+$ and other species are unresolved at mass 66. Hence, the oxygen isotope composition of SO_2 should be invariant from sample to sample.

In some applications where high sensitivity is required, an IRMS is operated in the static mode. The entire gas sample is admitted and all pumping of the mass spectrometer suspended. The minimum amount of gas required depends upon its ionization efficiency, the percentage of produced ions reaching the collector (transmission efficiency), and the mode of ion detection. Measurements have been made with fewer than 10^6 atoms of a given isotope.[28]

It is more usual for a gas-source IRMS to be operated in the dynamic mode. The gas is introduced through a leak and the mass spectrometer continuously pumped.

Gas admission to the source through a pin-hole leak is described by Graham's law of diffusion whereby the

flow is proportional to the inverse of the square root of the mass. This is known as molecular or effusive flow and the mean free path of the molecules is many times greater than the diameter of the hole. It also happens that unionized gas diffuses effusively out of the ionization chamber. Hence there is a concellation of mass discrimination effects and the ion currents leaving the source closely represent the isotopic composition of the sample. The disadvantage of a molecular leak is the change in isotopic composition of the sample over time. This can be minimal if a small fraction of a given sample is consumed during analysis. It is more challenging to maintain reference gas reservoirs which must be large and frequently monitored isotopically. The low pressure differential across a molecular leak realizes its main advantage, small sample requirement.

The alternate extreme is a long capillary leak (order of 0.1 mm internal diameter by 0.5 m length) with the mean free path of molecules in the sample being many times shorter than the dimensions of the leak. The flow is termed viscous and is independent of mass of the molecules (Poiseuille's law). In practice, passage of gas through a capillary is complicated with viscous flow on the reservoir end, molecular flow into the ionization chamber, complex flow between the ends, and back-diffusion. The situation is improved by careful pinching of the capillary close to the end entering the mass spectrometer and having a sample pressure of ~20 torr.[29] Conditions between the molecular and viscous extremes are termed 'Knudsen flow', in recognition of his mathematical expressions relating flow to pressure, temperature, mass, and tube dimensions. As the mass dependence of flow is highly dependent upon sample pressure, this region should be avoided.

In traditional IRMS, the pressure requirement for viscous flow limited the minimal sample size to ~1 mg of the element of interest. For some studies such as trace gases in ice, this is unsatisfactory. There have been moderately successful attempts to carry out traditional IRMS with smaller samples. One approach is the use of very small sample reservoirs (~500 μL). As the flow conditions depend on the total pressure, viscous flow can be maintained for small samples by additions of inert gas. Higher-sensitivity detectors are then used to measure the smaller ion currents produced from the sample gas.[30]

Another approach has been to freeze the prepared gas into a microcontainer and use a pusher gas to act as a piston to introduce the sample into the mass spectrometer.[31] The challenge is to realize isotopic analyses of the sample before there is substantial mixing with the pusher gas. The latter can be the same chemical compound as that being analyzed. Then the change in isotope composition of the gas mixture with time can be used to more accurately calculate the isotope composition of the sample.

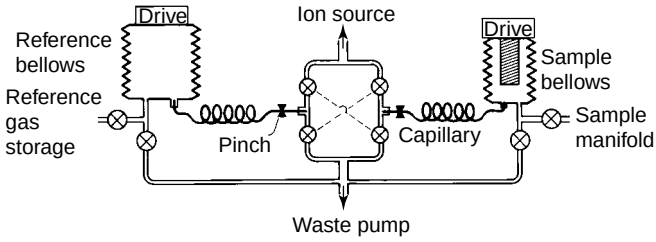

Figure 6 Diagram of a dual inlet system for gas-source IRMS.

To minimize problems associated with source nonlinearity, it is desirable to compare a sample and reference gas under identical conditions. For this purpose, dual-inlet (DI) systems were developed in the 1940s.[32] In Figure 6, the central valves are cross-connected so that one pair is open while the other is closed. In this way, the sample gas can enter the ion source while the reference gas is pumped away by a waste line, or vice versa. There are various valve designs including electromagnetically operated ground glass joints[33] and pneumatically actuated Nupro valves. Switching can also be accomplished by alternately blocking the low-pressure ends of the leaks with Teflon® diaphragms[34] or modified Nupro valves. Although pressure builds up in the blocked capillary, the increase in flow upon opening subsides within a few seconds. This approach does not require a waste line.

The pressures of the reference and sample gas are matched by varying the volume of the stainless steel bellows in Figure 6, either manually or by stepper motors. To accommodate small quantities of gases, the sample bellows is smaller. As bellows have a limited compression distance, a cylindrical block inserted in the sample bellows provides a larger fractional range of volume change. In an earlier environmentally questionable design, columns of Hg were raised or lowered in vertical glass tubes to adjust the pressures. The U-tube Hg manometers used to measure sample pressure have been replaced today by electronic capacitance manometers.

If sample size is not limited, DI/IRMS is the instrument of choice because it realizes the best precision. However, there is the requirement that gases for analyses be prepared in very pure form without isotope fractionation. This can prove time-consuming and involve expensive reagents.

2.4 Thermal Ionization Source

During heating of a salt-coated filament, positive and negative ions may be emitted as well as neutral particles. This has been termed thermionic emission surface ionization or thermal ionization (TI), most frequently implemented as thermal ionization mass spectrometry (TIMS).

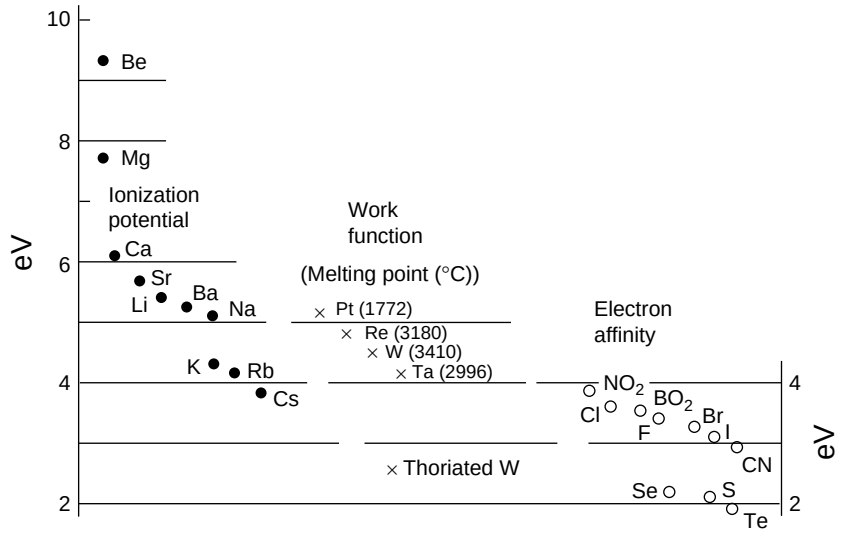

Figure 7 Ionization potentials, work functions, melting points, and electron affinities for selected elements and molecules.

2.4.1 Thermal Ionization

The efficiency of the emission of positive ions has been theoretically described by the Langmuir–Saha equation,

$$N^+ = N_0 \exp\left(\frac{W - e\mathrm{IP}}{kT}\right) \tag{10}$$

In Equation (10) N^+ is the number of positive ions, N_0 is the number of neutral particles, e is the electronic charge, W is the work function of the filament material in eV, IP is the ionization potential of the species being ionized, k is the Boltzmann constant and T the absolute temperature, kT being in units of electron volts.

For negative ions, Equation (10) is altered to Equation (11),

$$N^- = N_0 \exp\left(\frac{\mathrm{EA} - W}{kT}\right) \tag{11}$$

where EA is the electron affinity of the species in eV.

2.4.2 Filament Design

Using typical pure metal filament materials, there are only a few cases for which the exponent in Equation (10) is positive (Figure 7). Easily shaped platinum with the highest work function, is used extensively, the drawback being its low melting point. Zone-refined rhenium with a slightly lower work function but much higher melting point is the material of choice. It turns out that many elements, notably the alkalis, alkaline earths, rare earths and transuranium elements, can be successfully run by positive ion thermal ionization mass spectrometry (PTIMS).[35]

To optimize negative ion production, a filament material should be chosen with a work function lower than the electron affinity of the sample. This cannot be realized with pure metals (Figure 7). However, lower work functions can be obtained by depositing thin layers of certain salts (La, Ba, Ca, Th) on metals and consequently the isotopic analyses of many elements has been satisfactorily accomplished by negative ion thermal ionization mass spectrometry (NTIMS).[21,36–42]

Filaments are usually spot welded to stainless steel or other metal posts set in a ceramic button (Figure 8). Alignment pins on the button are used to plug the filament assembly into the ionization chamber of a TIMS (Figure 3). With single filaments, higher temperatures promote ion production but the sample is more rapidly evaporated. The evaporation and ionization processes can be separated by employing two filaments in close proximity (Figure 8).[43] One filament, coated with the salt, is heated at a lower temperature and evaporates the sample in a controlled manner. The second hotter bare filament ionizes the evaporated species. A further development is the triple filament source in which the two filaments coated with sample are placed symmetrically about the central ionizing filament (Figure 8). For some salts, a reference and sample can be placed on the two ionization filaments and alternately analyzed in the double filament mode.

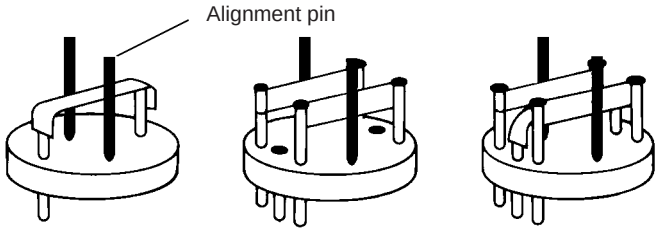

Figure 8 Single-, double-, and triple-filament assemblies.

It turns out that Equations (10) and (11) are at best, qualitative guidelines. Boat-shaped single filaments or even placing of sample in a dimple[43] on a single filament can achieve ionization efficiencies approaching those of triple-filament assemblies. Further, a variety of additives such as silica gel/H_3PO_4[45] have enhanced ionization for elements with high ionization potentials (>7 eV).[20,21]

A disadvantage of TIMS is the mass-dependent fractionation occurring during evaporation/ionization.[20,21,46] To a first approximation, the evaporation rate is proportional to the inverse of mass and the change in isotopic composition of remaining sample over time can be described by a Rayleigh distillation equation (Figure 1). One approach is to attempt to analyze all of the sample but this could require several hours. In practice, the phenomenon is quite complex and various equations have been derived to extract true isotope abundance ratios from temporally varying ion currents.[46,47]

Mass-dependent fractionation may be reduced through the use of multiple filaments. For example, when $^6Li^+$ and $^7Li^+$ are generated with a single filament, the kinetic effect is $7^{1/2}/6^{1/2}$. However, when $LiNO_3$ is evaporated intact from a sample filament and the molecules dissociated to produce Li^+ by the ionization filament, the mass dependence ratio is smaller, $69^{1/2}/68^{1/2}$.[48]

To minimize the effects of instrumental nonlinearity (section 2.2), the major ion current should be maintained as constant as possible throughout a run and from sample to sample. This also requires that filament construction and sample deposition be reproduceable.

Along with its drawbacks including the time requirements for sample preparation and ion current measurement, TIMS has many favorable features. The ions are nearly monoenergetic (~0.2 eV spread) and samples in the range 10^{-6}–10^{-9} g can be analyzed. A large number of elements can be analyzed by PTIMS or NTIMS and some by both. Measurements of two elements with one sample loading is also possible.[49]

A disadvantage is the requirement to open the instrument to insert the filament assembly. Through the use of a gate valve above the source housing, it is possible to expose only a small volume to the atmosphere (Figure 9). The filament assembly attached to a probe can be pushed into and pulled from the ionization chamber. In Figure 9, the probe is just above its position during operation. For removal:

1. pull probe outwards until filaments are just above the gate valve;
2. close the gate and fore line valves;
3. open the bleed valve and complete withdrawal.

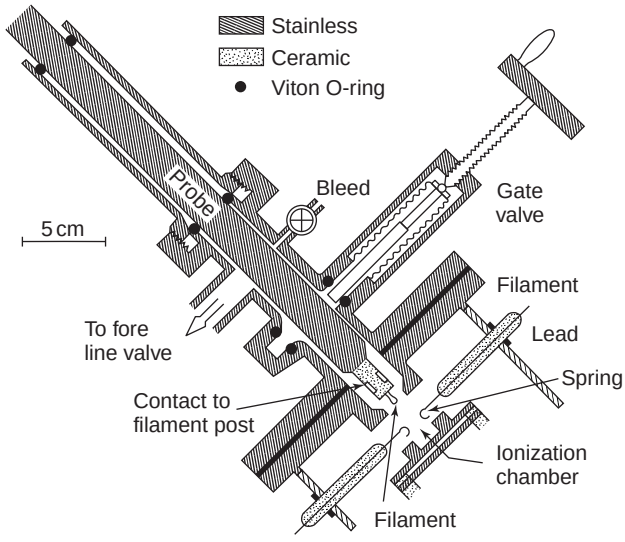

Figure 9 Vacuum lock for introducing and removing filament assemblies from a TIMS with minimal temporary reduction of vacuum. It is similar to a device on older AVCO instruments.

For insertion:

1. push probe inwards until filaments are just above the gate valve;
2. close bleed valve and open fore line valve;
3. at lowest fore line pressure, close fore line valve;
4. open gate valve and push probe inwards until the filament is in the ionization chamber.

To minimize opening of the source region, most commercial TIMSs are equipped with a rotatable carousel (Vacuum Generators, UK) or ferris wheel or turret (Finnigan MAT 262) on which up to 21 filament assemblies can be attached.

2.4.3 Isotope Spiking

Spiking is the addition of known amounts of one or more isotopes of the element of interest to a sample. One application is the determination of low concentrations of elements in a specimen. When the spiked sample is analyzed isotopically and the known spike additions accounted for, the numbers of each isotopic species in the specimen can be calculated.[20,21] This technique is called isotope dilution mass spectrometry (IDMS).[50]

A double spike is a known combination of two separated isotopes and its addition to a sample can be used to calculate systematic discrimination during isotope analysis.[51] Double spiking has become routine for many elements in TIMS. Addition of the spike at an early stage of chemical digestion of samples can also account for mass-dependent isotope fractionation during extraction of the element. For this reason, spiking should be used more frequently with other IRMS methods.

2.5 Analyzer Region

A magnetic field exerts a force on a moving ion which is perpendicular to both its velocity and the magnetic field lines (Equation 12):

$$\overrightarrow{F} = q\vec{v} \times \overrightarrow{B} \tag{12}$$

where $\overrightarrow{F}$ is the force, q the charge, $\vec{v}$ the velocity, and $\overrightarrow{B}$ the magnetic flux density. In most mass spectrometer designs, $\overrightarrow{B} \perp \vec{v}$ and the ion travels along a circular path. The force provides the centripetal acceleration (Equation 13):

$$F = qvB = \frac{mv^2}{r} \tag{13}$$

Rearrangement gives Equation (14):

$$mv = Bqr \tag{14}$$

showing that magnetic fields discriminate particles according to their momenta.

Combining Equations (9) and (14) gives

$$\frac{m}{q} = \frac{B^2r^2}{2V} \tag{15}$$

which defines the radius of the path in the magnetic field for a given ion species. By taking the differential of each side of Equation (15), it can be shown that

$$\frac{m}{\Delta m} = \frac{r}{2\Delta r} \tag{16}$$

The quantity on the left-hand side of Equation (16) is known as the resolving power. The mass resolution depends on other factors including source and collector slit widths.

A monoenergic ion beam focused at the entry of a uniform magnetic field is refocused after a deflection of 180°.[15] Hence semicircular magnets formed the basis of a number of early isotope ratio mass spectrometers.[35,52] By the 1930s, theory was developed for focusing with magnetic sector fields[20] and sector magnets gained favor because of their smaller size and hence lower costs.[17,18]

By analogy many terms associated with light are used in ion optics. For normal incidence, the object focal point (source slit), image focal point (collector slit) and center of curvature lie on a straight line for a normal geometry sector magnetic field (Figure 10a). This is termed directional focusing. There is only one path where the ions are deflected 90° for the 90° sector magnetic field shown in Figure 10(a). Ions traveling with different momenta at larger and smaller radii are deflected less and more than 90°, respectively. They are nevertheless focused on a curved focal plane.

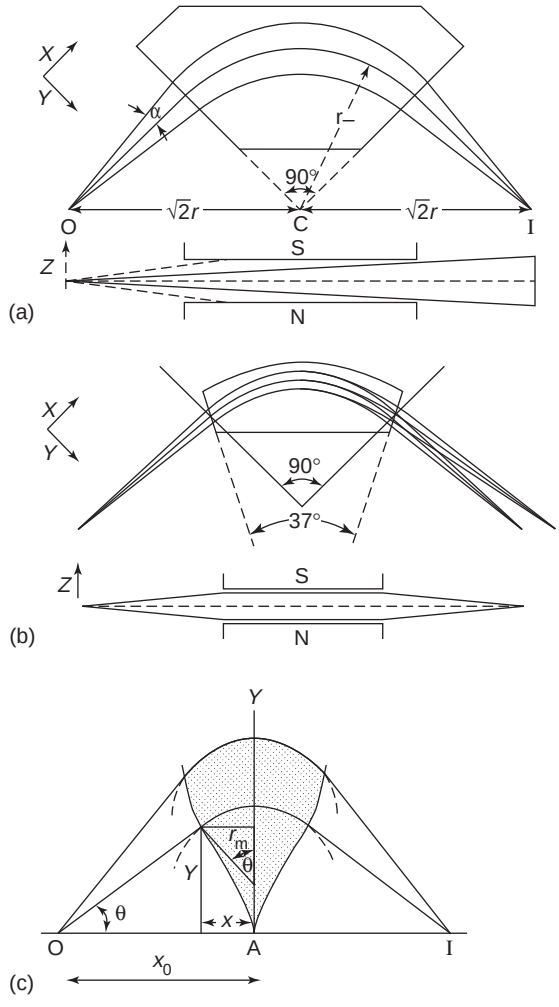

Figure 10 Directional focusing in homogeneous symmetrical magnetic fields: (a) 90° sector magnet normal geometry, (b) 90° deflection, extended geometry, (c) perfect focusing in a homogeneous magnetic field.[53]

Theoretical treatment of the normal geometry case predicts good focusing for small angles of divergence, α. Terms containing α^2, α^3, etc. are neglected so the conditions are referred to as first-order focusing. Perfect focusing can be accomplished by appropriate curving of the pole pieces[53] as shown in Figure 10(c). For the ion beams to cross the axis of symmetry AY, Equation (17) holds:

$$Y = x\frac{x_o - x}{(r_m^2 - x^2)^{1/2}} \tag{17}$$

As the curved shape is hard to manufacture, parts of the curve are approximated by circles or straight lines to achieve second-order focusing. For the latter, first- and second-order focusing requires (Equation 18),

$$\tan\varepsilon = -\frac{1}{2}\tan\frac{\phi_m}{2} \tag{18}$$

where ε is the angle between the actual and normal incidence and ϕ_m is the magnetic field angle. The object distance l from the source focal point to the magnetic field boundary is given by Equation (19):

$$l = \frac{2}{3} r_m \cot \frac{\phi_m}{2} \tag{19}$$

An example of this geometry used by Finnigan MAT and others is shown in Figure 10(b). Note that the center of curvature is no longer in the plane of the source and collector slits and the angle of deflection is not equal to the angle of the magnetic field. This is referred to as extended geometry. A bonus with this geometry is focusing in the axial direction termed z-focusing (compare Figures 10a and b).

A variety of shapes have been used in symmetrical and nonsymmetrical arrangements to achieve first- and second-order focusing. Slightly rounded pole pieces used on the Vacuum Generators' PRISM generate a collector focal plane that is nearly normal to the central ion beam. In the asymmetric Europa Scientific 2020, heavier ions such as CO_2^+ are deflected through 120° and the image distance is shorter than the object distance. For hydrogen isotopes, mass 2, H_2^+ follows the same path but mass 3, HD^+, is deflected through a side arm at a smaller angle and its image distance is longer than the object distance.[54] Another factor considered in designs is that the magnetic field extends slightly beyond the pole pieces. Inhomogeneous magnetic fields have also been used to achieve higher-order focusing.

When the magnetic field or source potential difference is slowly changed, ion currents are swept across the collector slit, generating peaks on a chart recorder or oscillographic display. For an ion beam of uniform cross-section, the ratio of the width of the flat topped peak to its base is $(C - S)/(C + S)$ where C and S are the widths of the collector slit and the thickness of the ion beam ($\sim$ source slit width), respectively (Figure 11). In practice, peaks are rounded on the tops and have tailing at the baseline. Factors contributing to these departures from the ideal case are the response time of the amplifier, imperfections in the slits, and the energy spread of the ion beam. A problem arises when an isotope is so low in abundance that it cannot be detected in the tail of the adjacent ion current. The term 'abundance sensitivity' is used to describe how well an instrument can detect a minor ion current. It is the ratio of the ion current at mass M over its contribution at mass $M + 1$. For sector magnetic field instruments, abundance sensitivities approaching 10^5 are achieved at mass 100. By using two consecutive MAs with a slit and ion acceleration after the first, the abundance sensitivity can be increased to 10^8 (see section 2.6.4).[26,35]

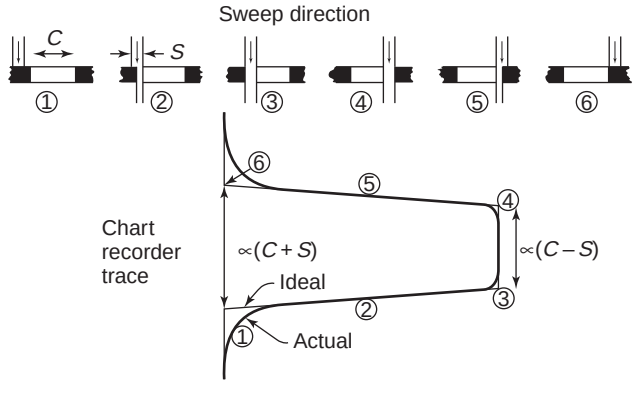

Figure 11 Generation of an ion current peak on a recorder as a uniform ion beam of thickness S is swept across a collector slit of width C.

The magnetic sector analyzer is said to be single focusing because it is direction focusing and requires a uniform energy ion beam. Directional focusing can also be realized in the radial electrostatic field that exists between concentric cylindrical condenser plates, i.e. electrostatic analyzer (ESA). It turns out that certain combinations of electrostatic and magnetic analyzers can be used whereby the velocity dispersion in one field is counterbalanced by that in another.[35,55] Such geometry is required for sources with a wide energy spread (see Figure 16). A compact version with a mass of less than 5 kg has been built for space exploration applications.[56]

2.6 Collector Design and Measurement of Ion Currents

Flat-top peaks (Figure 11) are desirable even if the mass spectrometer is not used in a scanning mode, as the ion current entering the collector will remain constant with small variations in the source potential difference and/or magnetic field. This requires that the collector slit be wider than the thickness of the ion beam. This reduces the instrument's resolution R which, for a symmetrical MA, is given by

$$R = \frac{\Delta m}{m} = \frac{S + C}{r_m} \tag{20}$$

Equation (20) neglects a term in the numerator dependent upon the beam's divergence angle and applies to the condition of the bases of ideal peaks just touching. In practice, the inverse of R, the resolving power, is used. Thus, the minimal separation of peak of mass 100 from that of 101 is stated as a resolving power of 100. To account for tailing, resolution specifications are based on criteria for the valley between two comparable peaks, usually 5–10% of the peak height.

For many lighter elements, isotopic ion currents are in the range 10^{-9}–10^{-12} A. For other elements, the minor ion current may be lower than 10^{-15} A. Different

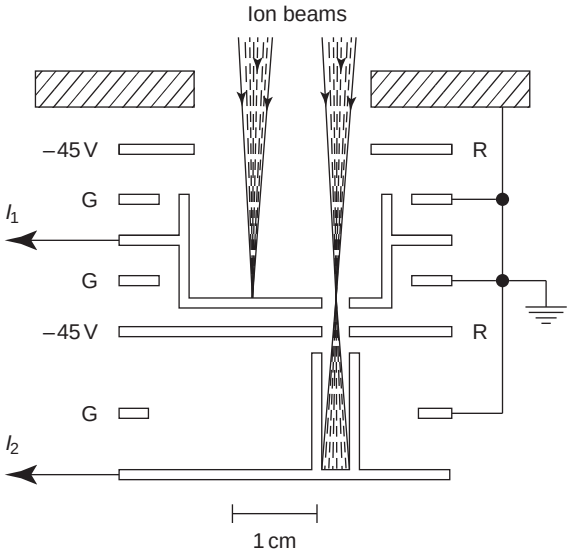

Figure 12 Diagram of Faraday cup-based collector assembly for simultaneous collection of two ion currents. R and G are repeller and guard plates, respectively. Electrical insulation spacers between plates are not shown.

measurement strategies are required to cover the range of currents.

2.6.1 Faraday Cup

A Faraday cup is usually a rectangular metal box with one open end into which the ion current enters (Figure 12). The box is connected electrically to the input of an amplifier with a very high impedence ($\sim 10^{15}\,\Omega$) input. The amplifier measures the electron current going to the box to neutralize incoming positive ions (or electron current leaving the box if negative ions are collected).

When ions hit surfaces of the box, some secondary electrons are emitted, giving an incorrect reading of the ion current. Secondary electron emission is suppressed by having deep boxes which may have inner walls coated with materials such as carbon. A repeller plate held at a potential of about $-45\,V$ with respect to ground, is also placed in front of the entrance to force secondary electrons to return to the box (Figure 12). Grounded metal guard plates or washers are placed between the repeller and the Faraday cup to prevent electrical leakage between them. Permanent magnets may also be used in the collector to prevent escape of secondary electrons.

Amplifiers operate with grid leak resistors of the order of 10^9–$10^{11}\,\Omega$, a current gain of $\sim 10^7$ and a voltage gain near unity. In earlier instruments, thermionic electrometer tubes were used.[17,18] They were subject to drift and, during the 1950s, commercial vibrating reed electrometers (Cary, USA; Vibron and Keithley, UK) were used because of their stability. In these devices, an input DC (direct current) potential is applied through a resistor to a capacitor with vibrating plates, thereby generating an AC (alternating current) signal proportional to the input. The AC signal is in turn amplified, converted back to DC and applied as negative feedback to cancel the input voltage (null method). The amount of negative feedback is measured. Today, relatively inexpensive integrated circuit amplifiers with MOSFET (metal oxide silicon field effect transistor) inputs have sufficiently high impedance and current stabilities of better than 10^{-15} A in typical operation.

Amplification as described above is based on Ohm's law. Another approach is to charge a capacitor for a fixed time and measure the voltage across it (Equation 21):[57,58]

$$V = \frac{Q}{C} = \frac{\int I\,\mathrm{d}t}{C} \tag{21}$$

where I is the current, Q is the accumulated charge $\int I\,\mathrm{d}t$ and C is the capacitance.

2.6.2 Electron Multiplier

Emission of secondary electrons, which was undesirable for the Faraday cup detector, is the basis of operation of the electron multiplier.[35] Ions arriving at the collector slit are accelerated to more forceably collide with the conversion dynode which emits secondary electrons (Figure 13). These are in turn accelerated to the next dynode to generate more secondary electrons. The dynodes are constructed of layered metals (Be on Cu, Mg on Ag). They are also curved to focus the electrons on subsequent dynodes in the chain. With as many as 14 stages, the current gain can be greater than 10^5.

Compact versions of electron multipliers include the channel electron multiplier (CEM). This is a hollow tube with inner semiconducting surfaces. The secondary electrons possess transverse velocity components and strike the inner wall but they are also accelerated axially by a potential of $\sim 2\,kV$ applied to the ends of the tube.

Electron multipliers may be used for current amplification or for generating pulses from individual incident ions at low currents. For pulse counting, the Daly detector[59] has found favor. Initial attempts to use scintillators to convert impinging ions into light pulses for amplification by photomultiplier tubes failed because the ion bombardment damaged the scintillator. Therefore, secondary electrons rather than direct ion beams were directed to the scintillator. The Daly detector is more proficient in distinguishing ion counts from background signal (known as dark current). A Daly detector may be preceded by a modest electron multiplier and currents as low as 10^{-20} A have been measured by pulse counting. Higher currents

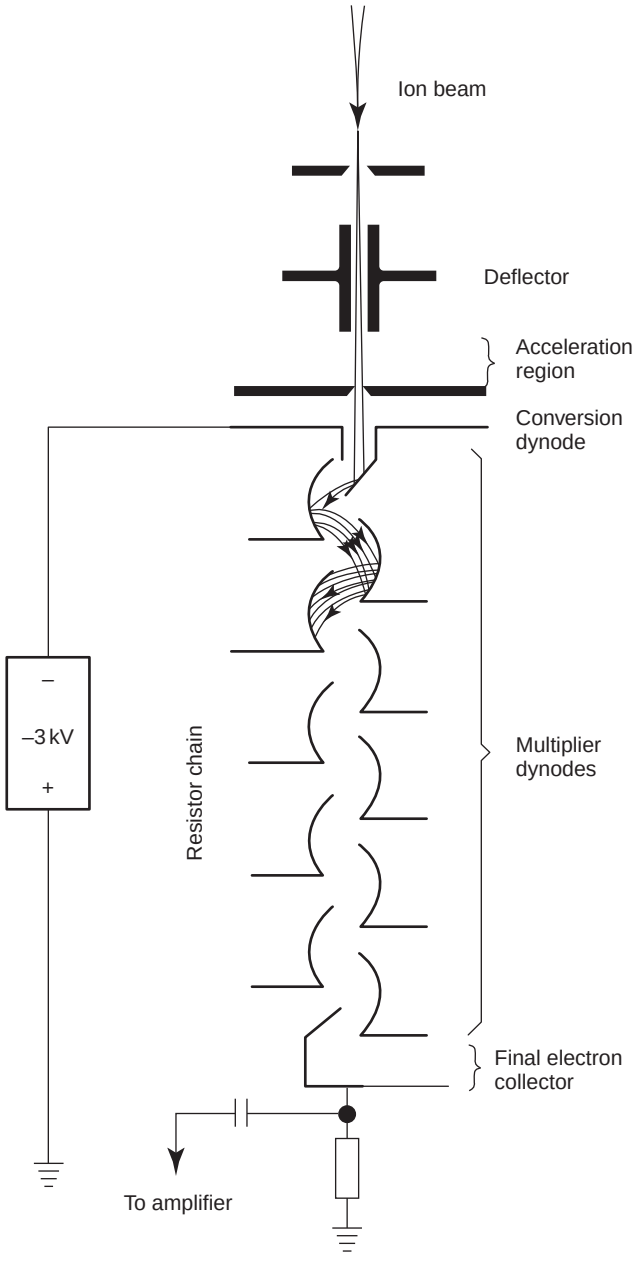

Figure 13 Diagram of secondary electron multiplier (SEM).

produce more losses because of counter dead time and a maximum ($\sim 10^{-10}$ A) is reached whereby the detector does not respond with distinct pulses.

2.6.3 Scanning

For an element containing many isotopes, the tendency has been to scan the mass spectrum and record ion peaks (see Figure 11). Scanning by changing the source potential difference (high voltage (HV)) is fast but the energies for different isotopic ions also change. This mode permits the use of permanent magnets for cost savings and stability. Scanning the magnet current is slower because of hysteresis which can be improved by the use of laminated magnet construction. Tracing of entire peaks is not required once their shape is optimized. Jump-scanning can be used whereby ion currents are measured at a few points on the peak tops and on the baseline between peaks.

Scanning with measurement of one peak at a time incurs errors due to changes in ion production – while measuring peak B, previously measured peak A might be higher or lower than when it was recorded. Some sources have larger fluctuations in ion production than others. With more stable electronics and faster scan rates, the situation has improved. However, scanning is inefficient for data acquisition as other ion currents are ignored when one peak or baseline is being measured.

2.6.4 Simultaneous Ion Collection

Collection of two or more ion currents with constant electric and magnetic fields (simultaneous collection) was established early in the history of IRMS.[32,33,60] Data collection approaches 100% efficiency and fluctuations in ion production increase or decrease all ion currents proportionally.

An assembly suitable for collecting two isotopic ion currents from either N_2^+, CO_2^+, SO^+, or SO_2^+ is shown in Figure 12. A minor ion current (heavier isotopic species for the above ions) passes through to the narrow deep Faraday cup while the major ion current plus perhaps another minor ion current is collected in the larger box.

A triple collector with a central narrow slit and two outer wide slits over three deep Faraday cups is frequently used with the above ions. It has the advantage that masses 44, 45, and 46 can be collected for CO_2^+, providing carbon and oxygen isotope data simultaneously. The occurrence of $^{12}C^{17}O^{16}O^+$ and $^{13}C^{16}O_2^+$ at mass 45 and other combinations at mass 46 necessitate corrections.[61–63] Triple collection has been used with other elements, such as $^{76}SeF_5^+$, $^{80}SeF_5^+$, and $^{82}SeF_5^+$,[64] and even more collectors (multiple collectors (MCs)) are available in some instruments, e.g. 12 Faraday cups in the Nu plasma ICP (inductively coupled plasma)/IRMS (section 5). MCs are not limited to Faraday cups. Electron multipliers in the current amplification or pulse counting modes can be combined with each other or with Faraday cups. For gas source IRMS, the combination of dual inlet and simultaneous collection provides the parameters for the definition of δ-values (Equation 2).

One device used in multicollector systems for resolving very low ion currents buried in an adjacent peak tail, is the retarding potential quadrupole (RPQ) such as the RPQ II option on Finnigan MAT 262 instruments. The RPQ operates on the low-mass side of the major ion

peak where the ions possess lower kinetic energy due to collisions in the ionization region and/or along the flight path. For Th isotopes, the major ion current, $^{232}Th^+$ is collected in a Faraday cup. However, some $^{232}Th^+$ ions arrive at the same position as that of $^{230}Th^+$ which is about 1.5×10^5 times less abundant. An RPQ is placed where the ^{230}Th ion current impinges and an energy threshold is set to stop the lower energy $^{230}Th^+$. The more energetic $^{230}Th^+$ ions pass through into an electron multiplier to be converted to pulses. Another approach used in the ISOLAB-54 is to have the minor beam filtered by an ESA placed after the multicollector.

2.6.5 *Simultaneous Collection Electronics*

Null methods were used in earlier measurements of simultaneously collected ion currents.[32,33,60] In the version shown in Figure 14, the output voltage of the major ion current electrometer amplifier is larger than that of the minor current amplifier (despite the latter's larger feedback resistor). The output of the major current amplifier is placed across a decaded put-and-take potentiometer. The put-and-take feature means that if a resistance, say 100 Ω, is added above the central wiper, the same resistance is removed below the wiper so that the total resistance remains the same, typically 10 kΩ. The total resistance between the ground and the central wiper is measured to four or five significant figures by reading numbers on dials for each decade. The voltage at the center tap is adjusted to match that of the full output of the minor current amplifier and the two voltages are applied to a zero-centered chart recorder, giving trace R for the reference gas (Figure 14). When a sample with a different isotope composition enters the mass spectrometer, the signal to the recorder is no longer zero, giving trace X. The distance between R and X is measured with a ruler. The chart paper is calibrated in per mil units using

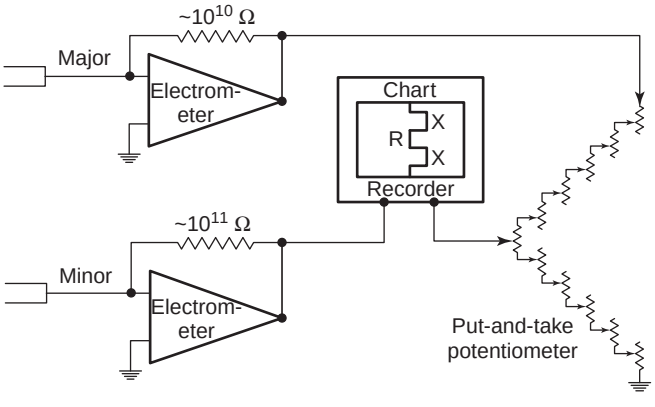

Figure 14 Circuitry for null method of determining the ratio of two isotopic ion currents.

different settings of the put-and-take potentiometer with the reference gas.

In an alternative circuit, the wiper is connected through a high-ohm resistance to the input of the minor current amplifier.[33,35,60] The potentiometer is adjusted so that the corresponding inverse feedback current cancels the minor ion current signal. The output of the minor current amplifier goes to the chart recorder.

During the early 1960s, digital conversion of ion currents was applied to IRMS.[65] The timebase frequency in a commercial integrating digital voltmeter (DVM) was replaced by that of a voltage-to-frequency (V to F) converter on the output of the major ion current amplifier. The output of the minor ion current amplifier went to the input of the voltmeter where V to F conversion also occurred. This configured the DVM as a digital ratiometer and its output to a printer produced a number which was proportional to the ratio of the two ion currents.

There are inexpensive ways to obtain a number proportional to the ratio of ion currents. Whereas five digits are required if full amplifier outputs are used, the put-and-take potentiometer can be retained and a cheaper three-digit DVM used to measure departures from null conditions. Inexpensive components can also be used with the charged capacitor method.[66]

Today, the outputs of simultaneous collection MOSFET operational amplifier integrated circuits undergo V to F conversion followed by inputs to counters (or integrated circuit ADCs (analog to digital converters)) and interfacing to a computer. The computer calculates the ratios of ion currents and ultimately the δ-values of samples.

2.7 Automation of Magnetic Sector Mass Spectrometers

Early IRMS was manually intensive with the operator continually adjusting filament currents on TIMS instruments, scanning and measuring peak heights. The DI/IRMS operator switched the inlet system, checked that reference and sample major ion currents were matched, and measured deflections on a chart recorder. Today, these operations are fully automated.

Through optical coupling to provide the necessary electrical insulation, the filament of a TIMS can be controlled through feedback from the major ion current amplifier. Magnetic fields are stabilized and measured by Hall probes.[67] Scanning constitutes selection of magnetic field values and even masses (rather than the former settings of magnet currents) and entering them into a computer.

In DI/IRMS, a value can be chosen for the major ion current. Through feedback from the collector or pressure measuring devices, stepper motor-driven bellows (see Figure 6) are periodically compressed or relaxed to maintain the selected current. The adjustments are rapid

because the position of the bellows is known to the computer along with values for the actual ion current and the selected current. The software calculates the number of steps and the direction that the motor must move.

Rather than evaluate corrections for each sample, two or three reference gases can be analyzed as samples at various times during a DI/IRMS run. At the end of the run, measured values for the references can be compared to accepted values, thereby providing correction factors to apply to other samples.

Gas samples attached to a manifold can be automatically introduced into the bellows, pressures adjusted to give the selected major ion current, and pumped away after analyses. A number of sample preparation lines including isotopic equilibration of CO_2 or H_2 with H_2O[68] and generation of CO_2 from carbonates reacted with H_3PO_4[69] or combustion of organic matter can be attached to an IRMS and fully automated.

3 CONTINUOUS FLOW ISOTOPE RATIO MASS SPECTROMETRY

3.1 Basic Concept

The minimum sample size for traditional IRMS is equivalent to ~1 mg of the element of interest because of the gas pressure required to meet viscous flow conditions. Further, the analysis usually consumes less than 1% of the prepared sample, so that most of the prepared sample is wasted. The sample size requirement can be reduced to the microgram range using CF/IRMS, also called isotope ratio monitoring (IRM).[70–82]

In this scheme, the prepared gas is swept through the mass spectrometer source with a carrier gas, usually He. Different ion currents corresponding to isotopic species are simultaneously collected, as with a traditional MC/IRMS. However, instead of measuring currents as such, they are integrated throughout the time that the sample is in the source. In other words, the total charge, Q, generated for each isotopic ion species is collected in each Faraday cup and measured. The He carrier ensures that viscous flow conditions are met.

An overview of a CF/IRMS system is shown in Figure 15(a). A prepared sample is flushed through a long quartz capillary to a split interface. The region between the two concentric cylinders of the interface is open to the ambient atmosphere. Therefore, it should be placed in a fume hood or other venting system for noxious gases. The interface is required because the total flow of gas into the ion source must be less than 0.5 mL min^{-1} and the required flow through some preparation devices can be 10^2 times higher. Thus, sample is also wasted with

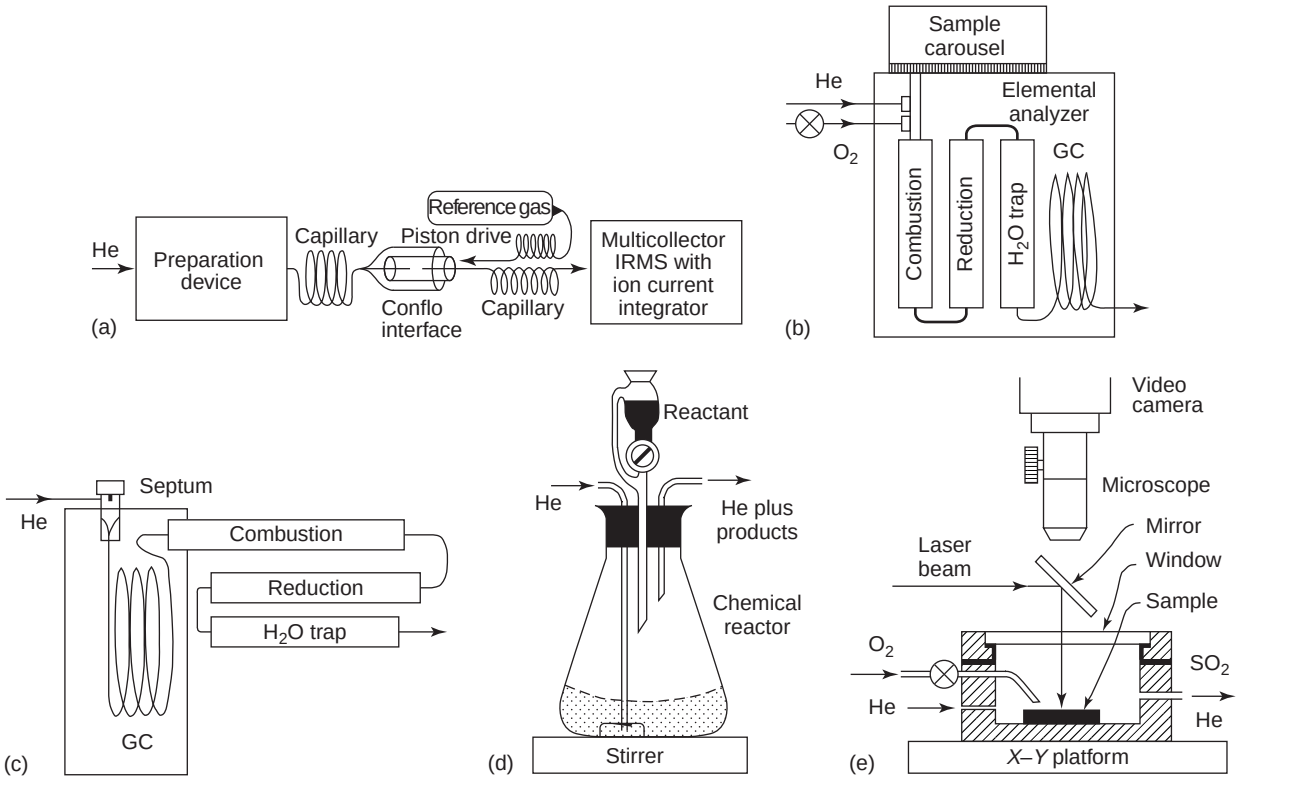

Figure 15 Diagram of (a) CF/IRMS system and various preparation devices, (b) elemental analyser, (c) compound-specific analyzer, (d) chemical reactor, and (e) laser probe.

CF/IRMS but, despite the loss, the minimum required mass of the element of interest is reduced by a factor of at least 10^2 compared to DI/IRMS. The split gas passes through another long quartz capillary (MS sniffer) into the mass spectrometer.

Nonlinearity of the mass spectrometer poses problems for CF/IRMS because the ion currents are transient. The shape of the I versus t plot for the processed sample is approximately Gaussian, whereas injected reference gases produce a nearly rectangular peak. If possible, peaks with the same area should be generated for all samples and the reference gas. One approach is to analyze a sample gas first and use the area of its major peak with electronic–mechanical feedback to adjust the volume of reference gas subsequently injected.

Clearly the benefits of CF/IRMS are small sample size requirements, high throughput, simplified preparative chemistry, the ability to measure the amount of prepared gas, and lower analytical costs. Precision is not as good as with DI/IRMS, but more than adequate for most applications.

A number of preparation devices have been attached to CF/IRMS instruments. Some have been fully automated, whereas others have been operated manually. These devices have also been used in DI/IRMS with cryogenic trapping of prepared gas and larger raw sample size.

3.2 Bulk Sample Isotope Analysis

The most frequently used preparation device with CF/IRMS is a combustion system such as in an elemental analyzer (EA), CNS analyzer, or automatic nitrogen, sulfur, and carbon analyzer (ANSCA). Although some are built in-house, commercial units such as Carlo Erba have carousel (autosampler) features for loading 50 or more raw samples. The procedure is termed bulk sample isotope analysis (BSIA) or bulk composition IRM. Submilligram quantities of the sample weighed into small tin cups (~3 mm diameter by 5 mm deep), are dropped into a vertical combustion tube (Figure 15b). Simultaneously, a pulse of O_2 is admitted into the He stream. The furnace is maintained at ~1100 °C, whereas the exothermic reaction of oxidation of tin should realize a much higher temperature (~1600 °C). Addition of solid oxidants to the cup is discouraged because the increased mass means increased heat capacity and probably a lower average reaction temperature. However, oxidants such as CuO, Cr_2O_3, and Co_3O_4 are placed in the bottom of the combustion tubes, sometimes as coatings on chemically inert spheres to avoid impeding gas flow. The combustion furnace is followed by a reduction furnace with Cu shavings at 600 °C to reduce nitrogen oxides to N_2 and remove surplus O_2. Sometimes oxidation and reduction zones are included in one tube. A packed gas chromatography (GC) column separates the products of combustion, N_2, CO_2, H_2O, and SO_2. SO_2 and H_2O are often chemically removed, as C and/or N isotopes are measured in most applications. However, analyses of SO_2 (S isotopes)[73,77] and H_2O (H isotopes)[78] have also been carried out. With the latter, there is the problem of resolving HD^+ from $^3He^+$ and tailing of the $^4He^+$ peak which is ~10^6 times higher. Possible solutions include high resolution of the peaks at mass three, use of an RPQ, or use of another carrier gas. As the He flow in BSIA is ~100 mL min^{-1}, a high split ratio is required of about 1 : 200.

For routine S isotope measurements on pure sulfide and sulfate minerals, it can prove expedient to remove the GC as SO_2 tends to stick to walls in vacuum systems. The reduction furnace should be retained to reduce any product SO_3 to SO_2.

Bulk $\delta^{18}O$ analyses have been attempted with some success by adapting an EA for pyrolysis of organic materials, and inorganic compounds including carbonates, sulfates, and H_2O.[79–83] One problem is cogeneration of CO_2 and CO and production of the latter is encouraged by reaction at high temperatures (~1300 °C) in a carbon-rich environment. Quartz combustion tubes should not be used because they exchange O isotopes with CO at high temperatures.[84]

3.3 Multielement Isotope Analysis

The split interface in Figure 15(a) was dubbed a conflo interface by Finnigan MAT. In the standard conflo, a piston-driven capillary with He flow can be inserted additional to the sample inlet and MS sniffer capillaries. During measurements on N_2 and CO_2 prepared from the same sample with an EA, the CO_2 can be diluted by He permitting the same amplifier gains as for N_2 and avoiding overpressure in the source. In the triple conflo version, two more piston-driven capillaries are added for insertion of N_2 and CO_2 isotopic references. The sequence of analysis is:

1. With the source HV set for N_2^+, momentarily insert and withdraw the N_2 reference capillary and measure Q_{29}/Q_{28}.
2. Q_{29}/Q_{28} is measured for N_2 from the EA.
3 The source HV is adjusted for CO_2^+ and the He dilution capillary inserted.
4. Q_{45}/Q_{44} and Q_{46}/Q_{44} are measured for CO_2 from the EA.
5. The He dilution capillary is withdrawn (optional). The CO_2 reference capillary is momentarily inserted and withdrawn for Q_{45}/Q_{44} and Q_{46}/Q_{44} measurement.

In addition to inserting the reference gas capillaries for each sample analyzed, raw material references can be randomly loaded into the autosampler.

The above system can be expanded to include SO_2. However, when the typical isothermal GC in an EA is adjusted to adequately separate N_2 and CO_2, the SO_2 emerges much later. Hence researchers tend to carry out $\delta^{34}S$ measurements separately with a shorter GC column to save time.

3.4 Compound-specific Isotope Analysis

In compound-specific isotope analysis (CSIA), a specimen is separated into different chemical compounds and suitable gases prepared from the separates for isotopic analyses.[70,74–76] Light HC (hydrocarbon) gases may be separated on a packed column GC. Heavier organic molecules require capillary GC techniques, although high-precision liquid chromatography has also been used.[85] The latter requires special transport of separated compounds into a combustion furnace by a moving Ni wire. The He carrier gas is introduced at the entrance to the furnace. A GC set-up for CSIA is shown in Figure 15(c). The separated compounds pass through a combustion furnace producing CO_2 and N_2, followed by a reduction furnace for converting nitrogen oxides to N_2. To minimize broadening of the peak of prepared gas reaching the IRMS, the furnace tubes have a small internal diameter ($\sim$1 mm). Thin wires of Pt, Ni, and Cu are inserted in the combustion tube and oxidized to give NiO and CuO which in turn oxidize the separated compounds. Water is removed by immersing a trap in a dry ice bath with a risk of plugging the capillary or by a Nafion® hydroscopic capillary section. If N isotope measurements are conducted, the CO_2 is frozen out in a liquid N_2 trap.

Introduction of reference gases at the interface is more complex. It can be done before and after the chromatographic trace and perhaps when there is sufficient time, between successive GC peaks. However, as the amounts of sample vary with the specific compound, injection of different amounts of the reference is desirable to correct for nonlinearity in the IRMS. Satisfactory calibration may require diverting gas prepared from some specific compounds in order to insert the reference gas. An alternative is to apply known additions of standard mixtures to the raw sample.

Another problem is the isotopic fractionation in a GC column. For example, ^{13}C-containing species elute sooner than ^{12}C-containing species. With light gases, the separation is sufficient that there is time to collect all isotopic charges for a given compound. With heavier molecules, overlapping of peak tails requires monitoring of the changing isotopic ion current ratios and curve fitting.[86] (This is the origin of the term 'isotopic ratio monitoring'.)

For determination of $\delta^{13}C$ and $\delta^{15}N$ values of trace atmospheric gases (CO_2, N_2O, CH_4, etc.), preconcentration steps are required prior to GC separation.[87] About 100 mL of air is needed for firm analyses of these gases, whereas 100 L would be required to process them for DI/IRMS.

3.5 Chemical Reactors

This category includes many possibilities whereby a suitable gas for isotopic analyses is prepared. It could constitute a manifold of many reactors where H_2 and CO_2 are isotopically equilibrated with H_2O for δD and $\delta^{18}O$ determinations. Mixed-carbonate minerals may be treated with H_3PO_4 to release CO_2 sequentially from calcite, dolomite, and siderite in about 1 h, 2 days, and 2 months, respectively at room temperature. By monitoring the isotopic composition of the CO_2 evolved, the isotopic composition of individual minerals and their percentage occurrence in the sample can be ascertained. By increasing the temperature during the run, the analytical time is shortened. Individual reactors (Figure 15d) have been used to study carbon KIE during decarboxylation reactions and sulfur KIE associated with SO_2 evolution when SO_2Cl_2 is used as a chlorination agent. In principle, quite small microbiological and chemical reactors can be used to provide detailed measurements of isotope effects associated with HC oxidation, methanogenesis, denitrification, and other processes.

3.6 Laser Probe

Laser beams have been used for decrepitation of crystals to release gases[88] or as a heat source to promote chemical conversions over a limited area (down to $\sim$10 μm diameters) on a solid surface.[89–102] They have also been used with powdered reactant mixtures for routine sample preparation. Whereas many units are stand alone or attached to DI/IRMS systems, their use with CF/IRMS is growing.[101,102]

For some materials, laser beam energy is not absorbed effectively at the target spot. The material may be highly reflective, highly transparent, easily fractured, or conduct heat away from the spot too quickly. In the case of the latter, yields of gases are higher at the edges than near the centre of the specimen's surface. In the former, the solid angle for heat conduction is π steradians, but 2π steradians for the latter. The choice of wavelength also influences the yields. The Nd : YAG laser operates in the infrared at 1.06 μm and has proved particularly suited to the production of SO_2 from certain sulfide minerals (PbS, ZnS, FeS, $FeCuS_2$, Ag_2S, Sb_2S_3). Its frequency can be doubled or quadrupled (ultraviolet) for other

applications. A CO_2 laser (10.6 μm) has a frequency which closely matches a fundamental vibrational mode in CO_3^{2-} and hence the energy absorption should be similar among different carbonate minerals.

Isotope fractionation during laser-assisted sample preparation can be substantial. Mineral dependent shifts in $\delta^{34}S$ values (as much as 5‰ lower) during SO_2 preparation result during evaporation, condensation of unreacted vapor, and chemical oxidation.[95,96]

A laser probe attachment suitable for preparation of SO_2 from sulfide minerals is shown in Figure 15(e). The mirror that reflects the Nd:YAG laser beam ($\sim 10^7$ W m^{-2}) towards the sample is transparent to visible light. Hence the specimen is viewed with a video camera connected to a TV monitor with an option of up to 500 × magnification of the surface relief by a contour synthesizer. Injections of O_2 is coincident with the laser beam pulses (~1 s duration). A dust filter should be placed at the exit of the chamber to avoid plugging the capillary. Excess O_2 can be consumed with a furnace containing Cu or the sample can be purified by passage through a GC column.

Material (ZnSe, BaF_2) for the window above the reaction chamber is chosen on the bases of transmission efficiency at the wavelength of the laser beam and its resistance to chemical attack. BaF_2 is desirable during fluorination of silicates to release O_2 for $\delta^{18}O$ analyses[93,94] or for preparation of SF_6 for $\delta^{34}S$ determinations.[97,98]

The sample holder sits on a stage which can be moved in the x and y (and even z) directions either manually or by computer controlled stepper motors.

A CF/IRMS CO_2 laser attachment has been recently developed to volatilize carbonaceous materials with the evolved gases subsequently combusted to produce CO_2 for $\delta^{13}C$ measurements.[102]

4 SECONDARY ION MASS SPECTROMETERS

In secondary ion mass spectrometry (SIMS), secondary ions are produced at the surface of a sample bombarded with focused primary ions (O^{2+}, O^-, N^{2+}, or Ar^+ from a duoplasmatron;[103] Cs^+ or Rb^+ by TIMS; or Ga^+ or In^+ from a liquid metal field emission source). Samples are cast into a thick (~5 mm) epoxy disk. One surface of the disk is ground and polished flat to expose the sample. A thin gold coating is then applied by vapor deposition to prevent static charge build-up during bombardment by the primary ion beam.

Initially, secondary ions were analyzed by quadrupole mass analysis (QMA) or time-of-flight (TOF) mass spectrometry (MS) (section 7). A few isotope ratio measurements of heavier elements have been made with these ion probes but it became clear that double focusing mass spectrometer optics were required to provide sufficient resolution (Figure 16). Multicollection of ions is also possible with the latter. Ion currents are measured with electron multipliers usually in the counting mode. Commercial instruments include the ISOLAB-54, Cameca IMS-1270, and SHRIMP (sensitive high-resolution ion microprobe) developed at the Australia National University.[104–106] These instruments have emphasized in situ U/Pb dating of zircons and have found more than one event in a crystal. However, they have also been used with lighter elements.[105,106] One problem with SIMS is the matrix-dependent large shifting of δ-values (up to 100‰ for light elements). Reproduceability of the order of a few ‰ can be obtained by alternately bombarding the sample and a number of standards containing minerals that occur in the sample. Positive features of SIMS are high resolution analyses of surfaces and, to some extent, measurements of variations in composition and isotope abundances with depth. The primary beam can also be electrostatically deflected to conveniently scan the sample's surface.

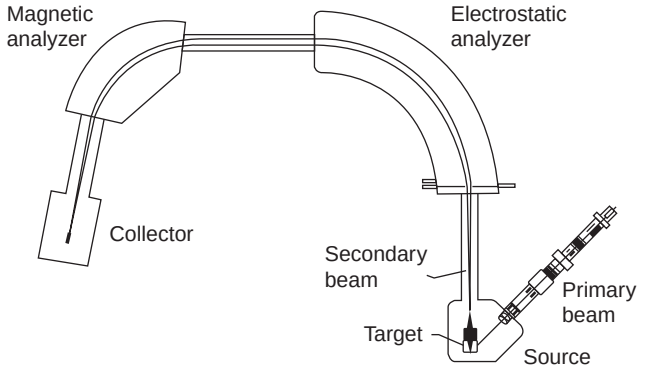

Figure 16 Diagram of a SIMS set-up. For the SHRIMP II, the dimensions are 85°, 1.7 m radius for electrostatic analyser and 72.5°, 1 m radius for magnetic analyser.[104]

5 INDUCTIVELY COUPLED PLASMA MASS SPECTROMETERS

As with SIMS, inductively coupled plasma mass spectrometry (ICP/MS) was initially used for chemical analyses and was quadropole mass spectrometry (QMS) and TOF/MS based. Commercial units including the Perkin-Elmer Elan 5000, Hewlett-Packard HP4500, and Spectromass 2000 have been examined for performance in isotope ratio measurements.[107] Although acceptable for many investigations, such instruments do not produce flat-topped peaks and cannot be used for multiple collection. Precise IRMS required incorporation of MC double

focusing mass spectrometers. Commercial units include the VG Plasma 54 with nine Faraday collectors and one secondary electron multiplier (SEM),[108] and Nu Instruments Plasma with 12 Faraday cups and three SEMs.

The ICP source contains a plasma torch and conical extraction plates (Figure 17). The torch has an outer SiO_2 tube ~18 mm outer diameter and 100 mm long. A coil of three turns of water-cooled Cu tubing around one end of this tube is connected to a radiofrequency generator (~1 kW, 27.12 MHz). Two inner concentric SiO_2 tubes are ring-sealed to the outer tube at the end opposite to the coil (not shown). The innermost tube ends with a 1.5 mm capillary just short of the coil (shown) to introduce the sample to the plasma region. Argon, introduced through side arms (not shown), cools the inner wall of the outer tube ($>10\,L\,min^{-1}$) and serves as a carrier ($\sim 1\,L\,min^{-1}$) for the sample. The sample and carrier gas are excited by a Tesla coil and an intense plasma (8000 K) is generated with energy supplied by the radiofrequency coil. Samples can be introduced into the plasma by nebulization, vaporization, and laser ablation. Only small particles and droplets (<4 μm diameter) are conveyed by the carrier gas into the plasma region. Under plasma conditions, most ions are atomic. They are extracted by a water-cooled nickle cone (sampler) with an orifice of ~1 mm diameter pointed at the plasma region (Figure 17). A more pointed skimmer cone lies beyond the sampler. The space between the cones is pumped to remove the large flux of gas which passed through the sampler, thereby serving as an intermediate stage in reducing the pressure from ~1 atm in the plasma region down to high-vacuum conditions suitable for the analyzer region. A quadrupole lens system can be used to change the cross-section of the extracted ion beam from a circular to narrow rectangular shape suitable for slits of a double-focusing mass spectrometer. Another quadrupole lens system after the ESA and MA can be used to alter the dispersion of the ion beams so that they enter consecutive Faraday cups in the mass range chosen for measurement.

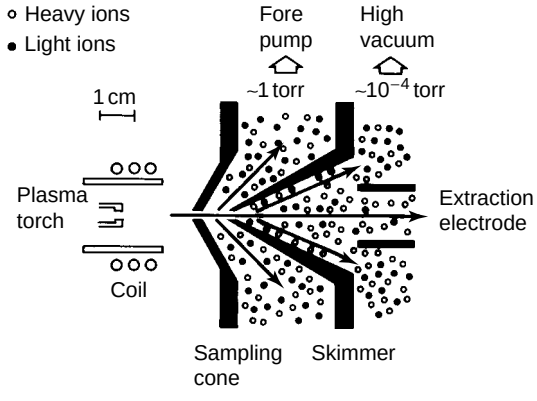

Figure 17 Diagram of mass fractionation occurring in an ICP/MS source.[21,107]

The ICP/IRMS technique has some inherent problems. Although the bulk of the ions are atomic, amounts of some molecular ions generated from the atmosphere and carrier gas are sufficient to cause interference. Thus $^{16}O_2^+$ is isobaric with $^{32}S^+$, $^{40}Ar_2^+$ with $^{80}Se^+$, etc. From the viewpoint of absolute isotope ratio abundances, mass discrimination in ICP/IRMS can be very large, e.g. ~4% per mass unit with Se.[107] Mass discrimination as qualitatively illustrated in Figure 17 is attributed to at least two mechanisms. One is space charge effects, whereby positive ions leaving the skimmer undergo mutual repulsion which deflects lighter ions preferentially. The second mechanism is nozzle separation, occurring in the supersonic gas jet in the region between the plasma torch and the skimmer. Both processes favor the heavier isotopes in the central transmitted beam. Fortunately, MC/ICP/MS has the capability of addressing some of its own problems. The mass discrimination can be evaluated using isotopic reference materials or by also collecting ion currents from elements which do not display significant natural isotopic fractionation. Thus by collecting ions from ^{203}Tl and ^{205}Tl, together with ^{204}Pb, ^{205}Pb, ^{206}Pb, and ^{208}Pb, mass-bias corrections can be made to the measured Pb-isotope abundances. Isobaric interference corrections can also be made; e.g. contribution of ^{204}Hg to the ^{204}Pb peak can be calculated if ^{202}Hg is also collected.

The ICP/IRMS technique is clearly an emerging tool with many problems that are being aggressively addressed. Attractive features are the ability to work with impure solutions with concentrations of the element of interest <1 ppm. Analysis time is a few minutes with ~1 μg of sample consumed. Isotopic reference solutions can be injected frequently to mimic DI/IRMS.

6 ACCELATOR MASS SPECTROMETERS

Accelerator mass spectrometry (AMS) uses accelerators (tandem, cyclotron, linear) in many modes to accomplish the measurements of abundances of very rare isotopes (below 10^{-12} percent).[109–113] Despite the size of these instruments and high capital/operating costs, there are at least 28 AMS laboratories worldwide. The scheme was first used in a cyclotron to measure ^3He.[109] The analytical objective is removal of interfering masses (isobars) during charge-changing collisions at energies of a few MeV. In some cases the negative ions of interfering isotopes are not stable; examples are ^{14}N, ^{26}Mg, ^{36}Ar, and ^{129}Xe during ^{14}C, ^{26}Al, ^{36}Cl, and ^{129}I analysis, respectively. In other situations, various approaches[110] such as energy loss rate are used for removal of undesirable isobars.

In practice, a number of geometries have been devised to deal with specific problems. Combinations of electric and magnetic fields are placed between the source and accelerator. A frequently used source is a Cs gun, which directs primary ions onto a solid surface. Initially, neutral species are sputtered, but soon build-up of Cs on the surfaces lowers the work function and promotes release of negative ions (section 2.4.1) in microamp quantities. The ions are moderately accelerated and pass through ESA and MAs to the accelerator. There the ions are accelerated to several MeV energy and lose electrons by passing through a foil or colliding with a gas. Molecular ions decompose, and atomic ions emerge with charges of +3 or higher. After the accelerator, combinations of ESA and MA fields ensure that remaining unwanted ions do not reach the detector.

The AMS method has been used mainly for ^{14}C dating of old materials, the sample size (~1 mg) being much smaller than that required for radioactive counting (~1 g). In this application, tandem accelerators with voltages around ~1.5 MV can be used (low-energy AMS). With heavier isobaric pairs, higher-voltage accelerators are required (high-energy AMS). A crucial aspect is that the cross-section for small-angle scattering of ions of residual gas decreases dramatically with increasing ion energy. Operation in the MeV range increases the abundance sensitivity by several orders of magnitude over that realized in the keV range employed with conventional tandem analyzers.

7 OTHER ISOTOPE RATIO MASS SPECTROMETERS

Numerous other mass spectrometer designs have been used for IRMS, of which a few will be described briefly. Some have unique characteristics which makes them ideal for specific tasks. Most have limitations that are overcome by MC/IRMS.

QMS can operate with a variety of sources for IRMS[114] (sections 4 and 5). Ions are directed along the axis of four symmetrically placed parallel cylindrical metallic rods (approximating hyperbolic surfaces). Through combinations of radiofrequency and DC voltages applied to opposing rods, ions oscillate at right angles to the axis with trajectories described by the Mathieu equation.[20] Scanning can be done by changing either the radiofrequency or DC voltages to permit transmission of certain ions. Advantages of the instrument are: small size, rapid scanning, resolving power up to ~8000, electric control of the resolving power, and linear operation up to pressures $\sim 10^{-4}$ torr, which in turn embraces a large dynamic range of ion currents.

A TIMS/QMA device, known as the thermal ionization quadrupole (THQ) offered by Finnigan MAT, combines the MAT 261 source with a Balzers quadrupole mass analyzer (Type QMA 150) and has a mass range of 500 amu. Its performance with negative ions of B, Cl, Br, and I has been compared to magnetic sector IRMS data.[115] The precision for the latter on average was about three times better than for the former for isotope ratio and IDMS measurements.

Gas-source QMS set-ups, such as the Balzers GAM 500, have been used for rare gases, metallic gases such as UF_6, and even coupled to an EA for $\delta^{13}C$ and $\delta^{15}N$ measurements.[114]

Another interesting IRMS device is the omegatron developed in the late 1940s.[116] It operates on the cyclotron principle. Ions are formed along an electron beam (~1 μA) that is parallel to a magnetic field and axial to two parallel nonmagnetic plates. When a radiofrequency voltage is applied to the plates, ions which are consistent with the cyclotron frequency (Equation 22),

$$f = \frac{Bq}{2\pi M} \tag{22}$$

follow an outward spiral path of a few thousand revolutions to arrive at a collector.

The analyzer region can be quite small and a useful instrument can be constructed with a permanent magnet. However, the dependence of the resolution on B^2 has promoted the use of large magnets. Another interesting property is the dependence of resolution on the inverse of the mass of the ion.

One problem with the omegatron is collisions over the long ion path, leading to reduction of ion currents. Traces of hydrogen in the system promote hydride ion species. H_3^+/HD^+ ratios of 2 have been observed[117] compared to <0.1 in typical EI sources. KrH^+ species have been found to comprise as much as 40% of the Kr^+ spectrum.[117] Consequently, the omegatron has been seldom used for IRMS analyses of gases, although it has some unique desirable features.

There are applications which require very high resolution at low masses, such as interference of H ions during $^4He/^3He$ analyses. The necessary resolution for separating H_3^+ and HD^+ is ~2000 and for HD^+ and $^3He^+$, ~700. The latter was realized with a QMS device[114] and omegatrons[117,118] and the former with an omegatron.[117]

TOF MS has been applied to IRMS, notably with SIMS sources. Such an instrument can be a drift tube of length L.[119] At one end, short bursts of ions are generated and travel down the tube. A given ion species arrives at the

collector, at time t later given by Equation (23):

$$t = \frac{Lm^{1/2}}{(2Vq)^{1/2}} \tag{23}$$

where V is the source HV. Gating of the pulse and the collector are accomplished by means such as electrostatic grids to avoid interrupting the ionization and collector electronics.

One disadvantage of the above mass spectrometers for IRMS, the inability to carry out MC, is being addressed by increasingly more sensitive detectors and faster scanning.

Another mass spectrometer for which MC is possible is the cycloidal mass spectrometer.[120] It is based on the fact that a uniform magnetic field crossed with a uniform electric field possesses perfect double focusing in a plane normal to the magnetic field. If an ion leaving the source entered only a magnetic field, the path would be a circle. However, if it travels initially against an applied electric field, its velocity component in that direction is reduced to zero and then increased in the opposite direction. The radius of the path increases with increasing speed and the path is a cycloid. Problems with the instrument include space charge build-up and difficulty in constructing a uniform electric field. However, these have been addressed to build a reliable MC isotope ratio mass spectrometer for H isotopes.[121]

8 APPLICATIONS

8.1 Introduction

IRMS has been developed for all relevant elements.[21] Although initially in the domain of the physical scientists, because of their ability to develop the instrumentation, IRMS practitioners were joined by earth scientists as commercial instruments became available. The ranks have been gradually augmented by archaeologists, biologists, space scientists, forensic scientists, and industrial and medical researchers. Consequently the collective published articles on IRMS applications number in the tens of thousands and only a few examples will be mentioned.

Multi-element IRMS has been particularly revealing in certain applications. The isotope fractionation of some elements is dominated by one or two mechanisms. Variability in H/D and $^{18}O/^{16}O$ ratios is due mainly to differences in the vapor pressures of isotopic species of water. The KIE in bacterial reduction is responsible for large variations in $^{34}S/^{32}S$ and $^{15}N/^{14}N$ ratios. Isotopic selectivity during photosynthesis is a major factor in altering $^{13}C/^{12}C$ ratios, although the lower flux of bacterial generation of methane is accompanied by much larger KIEs. The spread in natural $^{11}B/^{10}B$ ratios is the result of isotopic exchange between boric acid and borate ion in aqueous solutions. For other elements such as Ar and Pb, isotopic abundance variability is due to their radiogenic isotopes (isotopes produced by radioactive transformations).

For mass-dependent fractionation, a relationship exists between δ-values and the masses of the isotopes of a given element (Equation 24):

$$\frac{\delta_3}{\delta_2} = \frac{m_3 - m_1}{m_2 - m_1} \tag{24}$$

where '2' and '3' are two heavier isotopes whose δ-values are referred to the lighter isotope '1'. It is desirable to measure as many isotope ratios as feasible as departures from this relationship identify mass-independent processes such as radiogenic isotope production and/or contamination in prepared samples.[64] This concept is also used to correct T and ^{14}C data for mass-dependent isotope fractionation.

Applications can be grouped under four general categories: concentration measurements by isotope dilution, radioactive decay and nuclear reactions, isotope tracing, and isotope fractionation. These are not totally independent. The effectiveness of an isotopic tracer can be compromised if it is subjected to extensive isotopic fractionation. A suitable natural tracer may be generated by isotopic fractionation.

8.2 Concentration Measurements

Isotope dilution analysis (section 2.4.3) has been widely used to measure low concentrations of elements in a variety of materials including Cd in polyethylene,[122] S in steel,[123] Kr in the atmosphere,[124] and I in food.[115] In round-robin studies involving several methods, IDMS results have proved the most consistent, regardless of the IRMS instrumentation, and closest to certified values.[50] An intriguing application in zoology is the measurement of total body water (TBW).[125] If known additions of D_2O and $H_2{}^{18}O$ are given to a subject (doubly labeled water (DLW) technique), it is possible to measure their energy expenditure.[126] This technique has been applied to creatures ranging from humming birds to astronauts.

8.3 Radioactive Decay and Nuclear Reactions

Whereas radioactive counting techniques are better for measuring the half-lives of very rapidly decaying nuclides, IRMS has proved competitive over a wide range of lifetimes.[46] Measurements can be made with either the parent or daughter isotope. For short half-lives (such as $^{85}Kr \sim 10$ years[127]) measurements are made with the parent, whereas measurements of the daughter, often by ID, are used for long lived nuclei (e.g. $^{87}Rb \sim 4.5 \times 10^{10}$ years[128]).

Radioactive decay is the basis for a number of age-dating methods (geochronology) for our earth and meteorites.[129–132] Geochronological isotope clocks include $^{238}U \rightarrow {}^{206}Pb$, $^{235}U \rightarrow {}^{207}Pb$, $^{232}Th \rightarrow {}^{208}Pb$, $^{87}Rb \rightarrow {}^{87}Sr$, $^{40}K \rightarrow {}^{40}Ar$, $^{187}Re \rightarrow {}^{187}Os$, $^{147}Sm \rightarrow {}^{143}Nd$, $T \rightarrow {}^{3}He$, and $^{14}C \rightarrow {}^{14}N$. Measurements of the latter (5730 years) is based on the isotopic composition of the parent by AMS (section 6). Perhaps the most famous study was ^{14}C dating of threads from the Turin shroud. For T decay (12.26 years), IDMS measurements can be made on accumulated ^{3}He in closed bottles over several months.

At nuclear reactor installations, IRMS is employed to monitor fuel burn-up with $^{235}U/^{238}U$ ratios[46] and IDMS measurements of Gd isotopes,[133] and to determine consumption of control rods ($^{10}B/^{11}B$, $^{115}Cd/^{114}Cd$), neutron absorption cross-sections, and neutron flux. The IRMS of isotopes of many elements including H, Li, He, Sr, U, and Pu has been used in phases of development of nuclear weapons and global monitoring of their testing. The technique has also been used extensively to study the Oklo natural reactor in Gabon, West Africa.[134,135]

8.4 Isotopic Tracers

Isotopic tracer or label applications include planned additions of artificially or naturally enriched isotopes and scenarios where tracers enter a system naturally or by inadvertent industrial additions.

The most widely used artificially enriched isotope tracer is ^{15}N in studies of plant nutrition, soil N transformation, N fixation by plants and bacteria, and animal health.[136,137] A recent medical breakthrough in routine use is the noninvasive detection of *Heliobacter pylorii*, a bacterium that causes ulcers. It can be detected by having the subject consume ^{13}C-labeled urea which the organism converts to NH_3 and CO_2. Marked increase in $\delta^{13}C$ value of breath CO_2 after ~30 min is evidence of the bacterial conversion.

An example of using the spread in natural isotope abundances for isotope labeling was the use of SO_4^{2-} with high $\delta^{34}S$ and $\delta^{18}O$ values in artificial rain to study S conversions in soil cores with initially lower δ-values.[138] The sour (H_2S-rich) gas industry often emits SO_2 with $\delta^{34}S$ values much higher than those of pre-industrial environmental acceptors. This constitutes a large-scale tracer experiment for studying the fate of atmospheric S in surface and groundwater, soil, plants, and animals.[139] A similar example of industrial labeling of emissions is the Sr isotope composition of fly-ash from coal-fired power plants.

During terrestrial growth, carbon isotopes in plants are fractionated to different extents by three photosynthetic mechanisms; C_3 or Calvin–Benson ($\delta^{13}C \sim -35$ to -31‰), C_4 or Hatch–Slack ($\delta^{13}C \sim -14$ to -10‰), and crassulacean acid metabolism (CAM; $\delta^{13}C \sim -20$ to -10‰).[140–142] As the shifts in $\delta^{13}C$ values are relatively small in food chains,[140,142] photosynthetic C-isotope fractionation has produced natural tracers for many uses. With $\delta^{13}C$ values from bone collagen, archaeologists can ascertain the onset of maize (C_4) cultivation in a society that previously hunted animals in a C_3 environment. Adulteration of maple syrup (C_3) or apple juice (C_3) with corn syrup (C_4) or cane sugar (C_4) has been detected by $\delta^{13}C$ values and the evidence used by government prosecutors in court trials.

The isotopes of C, N, and S have distinct ranges of δ-values for marine and continental environments.[143] This provides estimates of fluxes of organic matter from both environments into estuarine food webs[144] and insights into interesting topics such as the migration of whales.[145]

The difference in $\delta^{13}C$ values for synthetic and natural compounds has been used for detecting food adulteration and the presence of performance-enhancing drugs such as testosterone in athletes, horses, and greyhounds. Changes in the $\delta^{13}C$ of atmospheric CO_2 recorded in situ and in proxy records such as tree rings and trapped CO_2 in glacier cores have been linked to increasing atmospheric CO_2 content and the increasing combustion of fossil fuels,[146] most of which have been ultimately derived from C_3 photosynthesis.

8.5 Isotopic Fractionation

Applications based on isotope fractionation per se span a wide range of disciplines. Laboratory studies of isotope fractionation are directed towards the understanding of processes and developing methods of isotope separation.

Solid state diffusion mechanisms can be identified on the basis of the isotopic dependence of the diffusion coefficients.[48] Molecular flow diffusion of UF_6 has been used for enriching isotopes of U. Electrolysis of water leaves the residual solution more enriched in D and T. Measurements of the fractionation has led to optimization for production of heavy water (D_2O). The process is also used to enrich T for radioactive counting.

Measurements of the equilibrium constants for isotope exchange reactions[147] are used to design isotopic enrichment schemes. The temperature dependence of the equilibrium constants is used for a variety of geothermometers.[148–150]

Measurement of the KIE during laboratory chemical[64,151–153] and microbial[152–155] conversions provides information about reaction kinetics and means of interpreting natural isotope fractionation.

The world's meteorology system is a complex distillation apparatus for isotopic species of H_2O because of

their differences in vapor pressures. Isotope data for precipitation at many stations in the Northern Hemisphere fall along the global meteoric water line (GMWL).[156,157]

The averages and variability of the δ-values depend upon location and are determined by factors such as latitude, temperature, topography, and storm path. For H_2O that has undergone substantial evaporation, the slope for Equation (25) decreases to ~5 and the intercept varies with location:

$$\delta D = \delta^{18}O + 11‰ \quad (25)$$

The δD and/or $\delta^{18}O$ value of H_2O has been the basis of research topics in many disciplines including rate of lateral mixing of rivers[158] and stratified lakes,[159] climatic records in ice[160] and tree rings,[161] meteoric water intrusion into gas and oil reservoirs,[162] and turnover of body H_2O in animals.[163]

The $\delta^{18}O$ value of H_2O may be shifted substantially through exchange with carbonate and silicate minerals. Such water–rock interactions and associated microbial conversions have led to the extensive field of isotope hydrology or isotope hydrogeology. In this field, IRMS elucidates the origins of H_2O and ions, flow paths, mixing, chemical/biological processes, and changes in the ocean over geological time.[159,164–171]

The IRMS technique has been used extensively to study topics in mineral and ore deposition including origin of metals and anions (e.g. S^{2-}), biological and chemical processes, thermometry, and chronology. Investigations have covered ore deposition temperatures ranging from below 100 °C with considerable bacterial involvement[172] up to magmatic conditions where minerals such as diamonds form.[173,174]

Investigations of fossil fuel deposits using IRMS have addressed many questions concerning origin of CH_4,[175–178] gas–oil–rock correlations,[177] reservoir connections,[175] H_2S in gas,[179,180] S in coal,[181,182] in situ biodegradation,[162] and chemical reactions occurring during recovery.[183]

Whereas isotopic tracers play a major role in biological research (section 8.4), measurements of isotopic fractionation per se are also very important. Some processes such as SO_4^{2-} assimilation and S^{2-} oxidation are accompanied by little isotope fractionation. Hence, apportioning of sources of S to plants and animals can usually be done with confidence.[139,184,185] However under S stress, plants emit H_2S with a high attending sulfur KIE and IRMS of foliage can detect this process.[139] This is analogous to large KIEs accompanying bacterial reduction (Equation 26):

$$SO_4^{2-} \longrightarrow H_2S, NO_3^- \longrightarrow N_2, CO_2 \longrightarrow CH_4, \text{ etc.} \quad (26)$$

Although isotopic shifts between trophic levels in food webs are small,[186,187] they are nevertheless important in terms of the factors which alter them[137,141–143] and accurate dietary reconstruction. Variations in isotopic selectivity during photosynthesis tend to be small for a given species, but the small differences can be related to temperature, humidity, and other factors.[137]

The term 'stable isotope geochemistry' is commonly used[150,167] to embrace most of the aforementioned topics. Terms such as 'environmental isotopes' have become part of the jargon.

8.6 Extraterrestrial Applications

It is fitting that IRMS should be used to examine the very processes (stellar nucleosynthesis) that generated isotopes and determined their initial abundances.[188] The field is called 'isotopic cosmochemistry' which could be considered as also embracing isotopic geochemistry.

Meteorites comprise samples of extraterrestrial material with large ranges of chemical composition. Some have experienced high-temperature metamorphism, whereas others such as the carbonaceous chondrites have a low-temperature history. Their chemical/physical heterogeneity seems to have been preserved since their formation around 4.6×10^9 years ago. Four decades of IRMS of meteorites[189,190] evolved into emphasis on fragments (submillimeter) and mineral separates. These studies have revealed many isotope abundance variations of nucleosynthetic origin.[190] The finding of these isotopic anomalies illustrates the importance of measuring abundances for three or more isotopes of an element (section 8.1). The IRMS of meteorites has shown that the solar nebula was not completely homogenized prior to formation of the planets. Extra material, the exotic component, was added possibly from a nearby supernova explosion.

The next major step in IRMS of extraterrestrial materials was retrieval of samples from lunar missions.[191] The primordial isotopic heterogeneities evident in meteorites, have not been found on the earth and moon presumeably because of thermal obliteration. Material on the moon's surface has been classed as crystalline rocks representing deep seated lithosphere, brecciated rocks associated with meteorite impacts, and dust/fines highly influenced by solar wind bombardment. The O- and S-isotope compositions showed less variation than those of earth. Spallation effects (interactions with cosmic rays) led to increases in $\delta^{15}N$ and $\delta^{13}C$ in the lunar fines. As these elements occur in relatively low concentrations, contamination during treatment is a serious issue.

Detection of isotopic anomalies at heavy masses requires large earth-based equipment. However, miniature IRMS devices[56] transported on unmanned flights have provided informative data for light elements in the

atmospheres of Mars,[192,193] Venus,[194,195] and Halley's comet.[196]

Thus, applications of IRMS appear to be limitless, ranging from studying the functioning of micro-organisms to trying to comprehend cosmic scale events. This should not be surprising as STABLE ISOTOPES ARE EVERYWHERE!

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
ADC	Analog to Digital Converter
AMS	Accelerator Mass Spectrometry
ANSCA	Automatic Nitrogen, Sulfur, and Carbon Analyzer
BSIA	Bulk Sample Isotope Analysis
CEM	Channel Electron Multiplier
CF	Continuous Flow
CSIA	Compound-specific Isotope Analysis
DC	Direct Current
DI	Dual-inlet
DLW	Doubly Labeled Water
DOP	Drawing Out Plate
DVM	Digital Voltmeter
EA	Elemental Analyzer
EI	Electron Impact
ESA	Electrostatic Analyzer
GC	Gas Chromatography
GMWL	Global Meteoric Water Line
HC	Hydrocarbon
HV	High Voltage
IAEA	International Atomic Energy Agency
ICP	Inductively Coupled Plasma
ICP/MS	Inductively Coupled Plasma Mass Spectrometry
IDMS	Isotope Dilution Mass Spectrometry
IRM	Isotope Ratio Monitoring
IRMS	Isotope Ratio Mass Spectrometry
KIE	Kinetic Isotope Effect
MA	Magnetic Analyzer
MC	Multiple Collector
MOSFET	Metal Oxide Silicon Field Effect Transistor
MS	Mass Spectrometry
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
NTIMS	Negative Ion Thermal Ionization Mass Spectrometry
PTIMS	Positive Ion Thermal Ionization Mass Spectrometry
QMA	Quadrupole Mass Analysis
QMS	Quadropole Mass Spectrometry
RPQ	Retarding Potential Quadrupole
SEM	Secondary Electron Multiplier
SIMS	Secondary Ion Mass Spectrometry
TBW	Total Body Water
THQ	Thermal Ionization Quadrupole
TI	Thermal Ionization
TIMS	Thermal Ionization Mass Spectrometry
TOF	Time-of-flight
V to F	Voltage-to-frequency

RELATED ARTICLE

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History

REFERENCES

1. F. Soddy, 'Radioactivity', *Chem. Soc. Ann. Rep.*, **7**, 256–286 (1910).
2. International Atomic Energy Agency, 'Reference and Intercomparison Materials for Stable Isotopes of Light Elements', Proc. Consultants Mtg., IAEA, Vienna, Dec 1–3, 1993. IAEA-TECDOC-825, 1995.
3. S. Epstein, T.K. Mayeda, 'Variations in the $^{18}O/^{16}O$ Ratio in Natural Waters', *Geochim. Cosmochim. Acta*, **4**, 213–224 (1953).
4. H. Chiba, H. Sakai, 'Oxygen Isotope Exchange Between Dissolved Sulfate and Water at Hydrothermal Temperatures', *Geochim. Cosmochim. Acta*, **49**, 993–1000 (1985).
5. H.C. Urey, 'The Thermodynamic Properties of Isotopic Substances', *J. Chem. Soc.*, 562–581 (1947).
6. J. Bigeleisen, M.G. Mayer, 'Calculation of Equilibrium Constants for Isotopic Exchange Reactions', *J. Chem. Phys.*, **15**, 261–267 (1947).
7. T.B. Coplen, J.D. Wildman, J. Chen, 'Improvements in the Gaseous Hydrogen–Water Equilibration Technique for Hydrogen Isotope Ratio Analysis', *Anal. Chem.*, **63**, 910–912 (1991).
8. J. Bigeleisen, 'The Relative Reaction Velocities of Isotopic Molecules', *J. Chem. Phys.*, **17**, 675–678 (1949).
9. N. Nakai, M.L. Jensen, 'The Kinetic Isotope Effect in the Bacterial Reduction and Oxidation of Sulfur', *Geochim. Cosmochim. Acta*, **28**, 1893–1912 (1964).
10. C.E. Rees, 'A Steady-state Model for Sulphur Isotope Fractionation in Bacterial Reduction Processes', *Geochim. Cosmochim. Acta*, **37**, 1141–1162 (1973).

11. R. Fiedler, G. Proskch, 'The Determination of Nitrogen-15 by Emission and Mass Spectrometry in Biochemical Analysis: A Review', *Anal. Chem. Acta*, **78**, 1–62 (1975).
12. G.J. Martin, X.Y. Sun, C. Guillon, M.L. Martin, 'NMR Determination of Absolute Site-specific Natural Isotope Ratios of Hydrogen in Organic Molecules, Analytical and Mechanistic Applications', *Tetrahedron*, **41**, 3285–3296 (1985).
13. J.J. Thomson, 'Rays of Positive Electricity', *Proc. Roy. Soc. (London), A*, **89**, 1–20 (1913).
14. F.W. Aston, 'A Positive Ray Spectrograph', *Phil. Mag.*, **38**, 707–714 (1919).
15. A.J. Dempster, 'A New Method of Positive Ray Analysis', *Phys. Rev.*, **11**, 316–325 (1918).
16. F.W. Aston, *Mass Spectra and Isotopes*, 2nd edition, Edward Arnold and Co., London, 1942.
17. A.O. Nier, 'A Mass Spectrometer for Routine Isotope Abundance Measurements', *Rev. Sci. Inst.*, **11**, 212–216 (1940).
18. A.O. Nier, 'A Mass Spectrometer for Isotope and Gas Analysis', *Rev. Sci. Inst.*, **18**, 398–411 (1947).
19. H. Birkenfeld, G. Haase, H. Zahn, 'Massenspektrometrische Isotopen Analyse', *Physikalisch-Chemische Trenn und Messmethoden 2*, ed. E. Krell, VEB, Deutscher Verlag der Wissenschaften, Berlin, 1969.
20. H.E. Duckworth, R.C. Barber, V.S. Venkatasubramanian, *Mass Spectroscopy*, 2nd edition, Cambridge Univ. Press, 1986.
21. I.T. Platzner, with contributions by K. Habfast, A.J. Walder, A. Goetz, *Modern Isotope Ratio Mass Spectrometry*, John Wiley & Sons, 1997.
22. L.G. Carpenter, *Vacuum Technology*, 2nd edition, Adam Hilger Ltd., Bristol, 1983.
23. B.J. Hopkins, 'Ultra-high Vacua', *Contemp. Phys.*, **9**, 115–135 (1968).
24. H.A. Steinherz, P.A. Redhead, 'Ultrahigh Vacuum', *Sci. Amer.*, **206**, 78–93 (1962).
25. D. Alpert, 'Production and Measurement of Ultra-high Vacuum', *Handbuch der Physik*, **12**, 609–663 (1958).
26. L.A. Dietz, C.F. Pachuki, J.C. Sheffield, A.B. Hance, L.R. Hanrahan, 'Further Development of the Two-stage Mass Spectrometer for Isotopic Analysis of Uranium', *Anal. Chem.*, **32**, 1276–1278 (1960).
27. C. Brunee, 'A New Mass Spectrometer for Precision Measurement of the $^{235}U/^{238}U$ Isotopic Ratio of UF_6', *Adv. Mass Spectrom.*, **2**, 230–243 (1963).
28. J.H. Reynolds, 'High Sensitivity Mass Spectrometer for Noble Gas Analysis', *Rev. Sci. Instrum.*, **27**, 928–934 (1956).
29. R.E. Halstead, A.O. Nier, 'Gas Flow Through the Mass Spectrometer Viscous Leak', *Rev. Sci. Instrum.*, **21**, 1019–1021 (1950).
30. J.M. Hayes, D.J. Des Marais, D.W. Peterson, D.A. Schoeller, S.P. Taylor, 'High Precision Stable Isotope Ratios from Microgram Samples', *Adv. Mass Spectrom.*, **7A**, 475–480 (1978).
31. S. Halas, H.R. Krouse, 'Isotopic Analysis of Nanomole Gas Samples by Means of Dynamic Flow Mass Spectrometry', *Rev. Sci. Instrum.*, **54**, 437–443 (1983).
32. C.R. McKinney, J.M. McCrea, S. Epstein, H.A. Allen, H.C. Urey, 'Improvements in Mass Spectrometers for the Measurement of Small Differences in Isotope Abundance Ratios', *Rev. Sci. Inst.*, **21**, 724–730 (1950).
33. R.K. Wanless, H.G. Thode, 'A Mass Spectrometer for High Precision Isotope Ratio Determinations', *J. Sci. Instrum.*, **30**, 395–399 (1953).
34. S. Halas, 'An Automatic Inlet System with Pneumatic Changeover Valves for Isotope Ratio Mass Spectrometer', *J. Phys. E*, **12**, 418–420 (1979).
35. M.G. Inghram, R.J. Hayden, *A Handbook of Mass Spectroscopy*, Nat. Acad. Sci. – Nat. Res. Council Nuclear Science Series No 14, 1954.
36. U.S. Klötzli, 'Negative Thermal Ionization Mass Spectrometry: A New Approach to Boron Isotope Geochemistry', *Chem. Geol.*, **101**, 111–122 (1992).
37. N.G. Hemming, G.N. Hanson, 'A Procedure for the Isotopic Analysis of Boron by Negative Thermal Ionization Mass Spectrometry', *Chem. Geol.*, **114**, 147–156 (1994).
38. R.A. Creaser, A. Papanastassiou, G.J. Wasserburg, 'Negative Thermal Ion Mass Spectrometry of Osmium, Rhenium, and Iridium', *Geochim. Cosmochim. Acta*, **55**, 397–401 (1991).
39. J. Völkenberg, M. Köppe, K.G. Heumann, 'Tungsten Isotope Ratio Determination by Negative Thermal Ionization Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **107**, 361–368 (1991).
40. M. Wachsmann, K.G. Heumann, 'Negative Thermal Ionization Mass Spectrometry of Main Group Elements, Part 1. 5th Group: Phosphorous, Arsenic, and Antimony', *Int. J. Mass Spectrom. Ion Processes*, **108**, 75–86 (1991).
41. M. Wachsmann, K.G. Heumann, 'Negative Thermal Ionization Mass Spectrometry of Main Group Elements, Part 2. 6th Group: Sulfur, Selenium, and Tellurium', *Int. J. Mass Spectrom. Ion Processes*, **114**, 209–220 (1992).
42. M. Unger, K.G. Heumann, 'Determination of NO_2 by Negative Thermal Ionization Mass Spectrometry', *Int. J. Mass Spectrom. Ion Phys.*, **48**, 373–376 (1983).
43. M.G. Inghram, W.A. Chupka, 'Surface Ionization Source Using Multiple Filaments', *Rev. Sci. Instrum.*, **24**, 518–520 (1953).
44. J.H. Chen, G.J. Wasserburg, 'Isotopic Determination of Uranium in Picomole and Subpicomole Quantities', *Anal. Chem.*, **53**, 2060–2067 (1981).
45. A.E. Cameron, D.H. Smith, R.H. Walker, 'Mass Spectrometry of Nanogram-size Samples of Lead', *Anal. Chem.*, **41**, 525–526 (1969).
46. P. De Bièvre, 'Accurate Isotope Ratio Mass Spectrometry: Some Problems and Possibilities', *Adv. Mass Spectrom.*, **7A**, 395–447 (1978).

47. L.J. Moore, E.F. Heald, J.J. Filliben, 'An Isotopic Fractionation Model for the Multiple Filament Thermal Ion Source', *Adv. Mass Spectrom.*, **7A**, 448–474 (1978).
48. O.W. Johnson, H.R. Krouse, 'Isotopic Mass Dependence of Li Diffusion in Rutile', *J. Appl. Phys.*, **37**, 668–670 (1966).
49. S.K. Sahoo, A. Masuda, 'Simultaneous Measurement of Lithium and Boron Isotopes as Lithium Tetraborate Ion by Thermal Ionization Mass Spectrometry', *Analyst*, **120**, 335–339 (1995).
50. K.G. Heumann, 'Isotope Dilution Mass Spectrometry', *Adv. Mass Spectrom.*, **12**, 575–592 (1992).
51. R.D. Russell, 'The Systematics of Double Spiking', *J. Geophys. Res.*, **76**, 4949–4955 (1971).
52. H.G. Thode, R.L. Graham, J.A. Ziegler, 'A Mass Spectrometer and the Measurement of Isotope Exchange Factors', *Can. J. Res.*, **B23**, 40–47 (1945).
53. L. Kerwin, 'Improved Magnetic Focusing of Charged Particles', *Rev. Sci. Instrum.*, **20**, 36–41 (1949).
54. S.J. Prosser, 'A Novel Magnetic Sector Mass Spectrometer for Isotope Ratio Determination of Light Gases', *Int. J. Mass Spectrom. Ion Phys.*, **125**, 241–266 (1993).
55. A.O. Nier, 'The Determination of Isotopic Masses and Abundances by Mass Spectrometry', *Zeit. Elektrochemie*, **58**, 559–567 (1954).
56. A.O. Nier, 'Small General Purpose Double Focusing Mass Spectrometer', *Rev. Sci. Instrum.*, **31**, 1127–1132 (1960).
57. H.A. Straus, 'A New Mass Spectrograph and the Isotopic Constitution of Nickel', *Phys. Rev.*, **59**, 430–433 (1941).
58. M.C. Jackson, W.A.P. Young, 'A Capacitive Integration System for the Precise Measurement of Isotopic Ratios in a Mass Spectrometer', *Rev. Sci. Instrum.*, **44**, 32–34 (1973).
59. N.R. Daly, 'Scintillation Type Mass Spectrometer Ion Detector', *Rev. Sci. Instrum.*, **31**, 264–267 (1960).
60. A.O. Nier, E.P. Ney, M.G. Inghram, 'A Null Method for the Comparison of Two Ion Currents in a Mass Spectrometer', *Rev. Sci. Instrum.*, **18**, 294–297 (1946).
61. H. Craig, 'Isotopic Standards for Carbon and Oxygen and Correction Factors for Mass Spectrometric Analysis of Carbon Dioxide', *Geochim. Cosmochim. Acta*, **12**, 133–149 (1957).
62. P. Deines, 'Mass Spectrometer Correction Factors for the Determination of Small Isotopic Composition Variations of Carbon and Oxygen', *Int. J. Mass Spectrom. Ion Phys.*, **4**, 283–295 (1970).
63. W.G. Mook, P.M. Grootes, 'The Measuring Procedure and Corrections for the High-precision Mass-spectrometric Analysis of Isotopic Abundance Ratios, Especially Referring to Carbon, Oxygen, and Nitrogen', *Int. J. Mass Spectrom. Ion Phys.*, **12**, 273–298 (1973).
64. K. Rashid, H.R. Krouse, 'Selenium Isotopic Fractionation During SeO_3^{2-} Reduction to Se^o and H_2Se', *Can. J. Chem.*, **63**, 3195–3199 (1985).
65. H. McCullough, H.R. Krouse, 'Application of Digital Recording to Simultaneous Collection in Mass Spectrometry', *Rev. Sci. Instrum.*, **36**, 1132–1134 (1965).
66. S. Halas, Z. Skorzyski, 'An Inexpensive Device for Digital Measurements of Isotopic Ratios', *J. Phys. E., Sci. Instrum.*, **13**, 346–349 (1980).
67. S. Halas, J. Sikora, 'Precise m/e Monitoring in a Magnetically Scanned Mass Spectrometer by Using Two Hall Probes', *J. Phys. E: Sci. Instrum.*, **20**, 662–664 (1987).
68. J. Horita, A. Ueda, K. Mizukami, I. Takatori, 'Automatic D and ^{18}O Analysis of Multi-water Samples Using H_2- and CO_2 Water Equilibration Methods with a Common Equilibration Set-up', *Appl. Radiat. Isot.*, **40**, 801–805 (1989).
69. J.M. McCrea, 'The Isotopic Chemistry of Carbonates and a Paleotemperature Scale', *J. Chem. Phys.*, **18**, 849–857 (1950).
70. D.E. Matthews, J.M. Hayes, 'Isotope-ratio-monitoring Gas Chromatography–Mass Spectrometry', *Anal. Chem.*, **50**, 1465–1473 (1978).
71. A. Barrie, J. Bricout, J. Koziet, 'Gas Chromatography–Stable Isotope Ratio Analysis at Natural Abundance Levels', *Biomed. Mass-spectrom.*, **11**, 583–588 (1984).
72. P.A. Freedman, E.C.P. Guillyon, E.J. Jumeau, 'Design and Application of a New Instrument for the GC–Isotope Ratio MS', *Amer. Lab.*, **20**, 114–119 (1984).
73. F. Pichlmayer, K. Blochberger, 'Isotopenhäufigkeitsanalyse von Kohlenstoff, Stickstoff und Schwefel Mittels Gerätekopplung Elementaranalysator-massenspektrometer', *Frensenius Z. Anal. Chem.*, **331**, 1961–1201 (1988).
74. D.A. Merritt, J.M. Hayes, 'Nitrogen Isotopic Analysis by Isotope-ratio-monitoring Gas Chromatography/Mass Spectrometry', *J. Amer. Soc. Mass Spectrom.*, **5**, 387–397 (1994).
75. M. Schoell, J.M. Hayes (eds.), 'Compound-specific Isotope Analysis in Biogeochemistry and Petroleum Research', Proc. Amer. Chem. Soc. Symp., 1192, San Francisco, *Org. Geochem.*, **21**, 561–827 (1994).
76. S.A. Macko, 'Compound-specific Approaches Using Stable Isotopes', in *Stable Isotopes in Ecology and Environmental Science*, eds. K. Lajtha, R.M. Michener, Blackwell Scientific Publishers, Oxford, 241–247, 1994.
77. A. Giesemann, H.J. Jäger, A.-L. Norman, H.R. Krouse, W.A. Brand, 'On-line Sulfur-isotope Determination Using an Elemental Analyser Coupled to a Mass Spectrometer', *Anal. Chem.*, **66**, 2816–2819 (1994).
78. A. Barrie, S.J. Prosser, 'Automated Analysis of Light-element Stable Isotopes by Isotope Ratio Mass Spectrometry', in *Mass Spectrometry of Soils*, eds. T.W. Boutton, S-I. Yamasaki, Marcel Dekker, 1–46, 1996.

79. S.J. Prosser, C.M. Scrimgeour, 'High-precision Determination of $^2H/^1H$ in H_2 and H_2O by Continuous Flow Isotope Ratio Mass Spectrometry', *Anal. Chem.*, **67**, 1992–1997 (1995).
80. R.A. Werner, B.E. Kornexl, A. Roßman, H-L. Schmidt, 'Online Determination of $\delta^{18}O$-Values of Organic Substances', *Anal. Chem.*, **319**, 159–164 (1996).
81. G.D. Farquhar, B.K. Henry, J.M. Styles, 'A Rapid On-line Technique for Determination of Oxygen Isotope Composition of Nitrogen-containing Organic Matter and Water', *Rapid Comm. Mass Spectrom.*, **11**, 1554–1560 (1997).
82. J. Koziet, 'Isotope Ratio Mass Spectrometric Method for On-line Determination of Oxygen-18 in Organic Matter', *J. Mass Spectrom.*, **32**, 103–108 (1997).
83. N.J. Loader, W.M. Buhay, 'Rapid Catalytic Oxidation of CO to CO_2 on the Development of a New Approach to On-line Oxygen Isotope Analysis of Organic Matter', *Rapid Commun. Mass Spectrom.* (in press).
84. H. Sakai, H.R. Krouse, 'Elimination of Memory Effects in $^{18}O/^{16}O$ Determination in Sulfate', *Earth Planet. Sci. Lett.*, **11**, 369–373 (1971).
85. R.J. Caimi, J.T. Brenna, 'High Precision Liquid Chromatography–Combustion Isotope Ratio Mass Spectrometry (Lcc–IRMS)', *Anal. Chem.*, **65**, 3497–3500 (1993).
86. W.A. Brand, 'Isotope Ratio Mass Spectrometry: Precision from Transient Signals', in *Advances in Mass Spectrometry*, eds. E.J. Karajalainen, A.E. Hesso, J.E. Jalonen, U.P. Karjalainen, **14**, 655–680, 1998.
87. W.A. Brand, 'Precon: A Fully Automated Interface for the Pre-GC Concentration of Trace Gases in Air for Isotopic Analysis', *Isotope Environ. Health Stud.*, **31**, 277–284 (1995).
88. G.J. Megru, 'Isotopic Analyses of Rare Gases with a Laser Microprobe', *Science*, **157**, 1555–1556 (1967).
89. M.D. Powel, T.K. Kyser, 'Analysis of $\delta^{13}C$ and $\delta^{18}O$ in Calcite, Dolomite, Rhodochrosite Using a Laser Extraction-system', *Chem. Geol.*, **94**, 55–66 (1991).
90. J.A.D. Dickson, P.C. Smalley, A. Raheim, D.E. Stijfhoorn, 'Intracrystalline Carbon and Oxygen Isotope Variations in Calcite Revealed by Laser Microsampling', *Geol.*, **18**, 809–811 (1990).
91. I.A. Franchi, E.K. Gibson, Jr, I.P. Wright, C.T. Pillinger, 'Nitrogen Isotopes by Laser Probe Extraction', *Lunar. Planet. Sci. Abst.*, **16**, 248–249 (1985).
92. I.A. Franchi, I.P. Wright, E.K. Gibson, Jr, C.T. Pillinger, 'The Laser Microprobe: A Technique for Extracting Carbon, Nitrogen, and Oxygen from Solid Samples for Isotopic Measurements', *J. Geophys. Res.*, **91**, D514–D524 (1986).
93. Z.D. Sharp, 'A Laser-based Microanalytical Technique for the in situ Determination of Oxygen Isotope Ratios of Silicates and Oxides', *Geochim. Cosmochim. Acta*, **54**, 1353–1357 (1990).
94. D. Mattey, C. Macpherson, 'High-precision Oxygen Isotope Microanalysis of Ferromagnesium Minerals by Laser-fluorination', *Chem. Geol.*, **105**, 305–318 (1993).
95. D.E. Crowe, J.W. Valley, K.L. Baker, 'Micro-analysis of Sulfur-isotope Ratios and Zonation by Laser Microprobe', *Geochim. Cosmochim. Acta*, **54**, 2075–2092 (1990).
96. S.P. Kelley, A.E. Fallick, 'High Precision Spatially Resolved Analysis of $\delta^{34}S$ in Sulphides Using a Laser Extraction Technique', *Geochim. Cosmochim. Acta*, **54**, 883–888 (1990).
97. G. Beaudoin, B.E. Taylor, 'High Precision and Spatial Resolution Sulfur Isotope Analysis Using Miles Laser Microprobe', *Geochim. Cosmochim. Acta*, **58**, 5055–5063 (1994).
98. D. Rumble, T.C. Hoering, J.M. Patin, 'Preparation of SF_6 for Sulfur Isotope Analysis by Laser Heating Sulfide Minerals in the Presence of F_2 Gas', *Geochim. Cosmochim. Acta*, **57**, 4499–4512 (1993).
99. L.M. Jones, A.R. Taylor, D.L. Winter, S.D. Hurst, G.W. Keen, 'The Use of the Laser Microprobe for Sample Preparation in Stable Isotope Mass Spectrometry', *Terra Cognita*, **6**, 263–267 (1986).
100. T.K. Kyser, 'Micro Analytical Techniques in Stable Isotope Geochemistry', *Can. Mineral.*, **33**, 261–278 (1995).
101. Z.D. Sharp, T.E. Cerling, 'A Laser GC–IRMS Technique for in situ Stable Isotope Analyses of Carbonates and Phosphates', *Geochim. Cosmochim. Acta*, **60**, 2909–2916 (1996).
102. M.E. Wieser, W.A. Brand, 'A Laser Extraction/Combustion Technique for in situ $\delta^{13}C$ Analysis of Organic and Inorganic Materials', *Rapid Commun. Mass Spectrom.*, **13**, 1–8 (1999).
103. F.M. Bacon, 'Gas Discharge Ion Source. I. Duoplasmatron', *Rev. Sci. Instrum.*, **49**, 427–434 (1978).
104. J.R. De Laeter, A.K. Kennedy, 'A Double Focusing Mass Spectrometer for Geochronology', *Int. J. Mass Spectrom.*, **178**, 43–50 (1998).
105. C.S. Eldridge, W. Compston, I.S. Williams, R.A. Both, J.L. Walshe, H. Ohmoto, 'Sulfur Isotope Variations in Sediment Hosted Massive Sulfide Deposits as Determined Using the Ion Microprobe SHRIMP I.I., An Example from the Rammelsberg Ore Body', *Econ. Geol.*, **83**, 443–449 (1988).
106. M.A. McKibben, C.S. Eldridge, 'Microscopic Sulphur Isotope Variations in Ore Minerals from the Viburnum Trend, Southeast Missouri: a SHRIMP Study', *Econ. Geol.*, **90**, 228–245 (1995).
107. K.G. Heumann, S.M. Gallus, Gunther Rälinger, Jochen Vogl, 'Precision and Accuracy in Isotope Ratio Measurements by Plasma Source Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **13**, 1001–1008 (1998).
108. A.J. Walder, 'Advanced Isotope Ratio Mass Spectrometry II: Isotope Ratio Measurement by Multiple

Collector Inductively Coupled Plasma Mass Spectrometry', in *Modern Isotope Ratio Mass Spectrometry*, ed. I.T. Platzner, John Wiley & Sons, 83–108, 1997.
109. L.W. Alvarez, R. Cornog, '^{3}He in Helium', *Phys. Rev.*, **56**, 37 (1939).
110. A.E. Litherland, K.W. Allen, E.T. Hall (eds.), *Ultra-high-sensitivity Mass Spectrometry with Accelerators*, The Royal Society, London, 1987.
111. D. Elmore, F.M. Phillips, 'Accelerator Mass Spectrometry for Measurement of Long Lived Radioisotopes', *Science*, **236**, 543–550 (1987).
112. K.H. Purser, R.B. Liebert, C.J. Russo, 'MACS: An Accelerator Based Radioisotope Measuring System', *Radiocarbon*, **27**, 794–803 (1980).
113. W. Wölfli, H. Polach, H.H. Anderson, 'Accelerator Mass Spectrometry', *Nucl. Instrum. Meth.*, **233**, 91–488 (1984).
114. A. Goetz, 'Advanced Isotope Ratio Mass Spectrometry III: Quadrupole Isotope Ratio Mass Spectrometry', in *Modern Isotope Ratio Mass Spectrometry*, ed. I.T. Platzner, John Wiley & Sons, 109–133, 1997.
115. K.G. Heumann, W. Schindlmeier, H. Zeininger, M. Schmidt, 'Application of an Economical and Small Thermal Ionization Mass Spectrometer for Accurate Anion Trace Analyses', *Fresenius Z. Anal. Chem.*, **320**, 457–462 (1985).
116. J.A. Hipple, H. Sommer, H.A. Thomas, 'A Precise Method for Determining the Faraday by Magnetic Resonance', *Rhys. Rev.*, **75**, 1877–1878 (1949).
117. P.V. Amerl, H.R. Krouse, S. Halas, P. Irwin, 'The Omegatron: An Improved Radio Frequency Isotope Ratio Mass Spectrometer', *RMZ-Materials and Geoenvironment* (formerly Rudarsko-Metalurski Zbornik), special issue, **45**, 5–7 (1998).
118. U. Engelmann, M. Glugla, R.-D. Penzhorn, H.J. Ache, 'Application of an Omegatron Type High Resolution Mass Spectrometer for the Analysis of Mixtures of Hydrogen and Helium Isotopes', *Nucl. Inst. Methods Phys. Res.*, **A302**, 345–351 (1999).
119. W.C. Wiley, I.H. McLaren, 'Time-of-Flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, **26**, 1150–1157 (1955).
120. W. Bleakney, J.A. Hipple, 'A New Mass Spectrometer with Improved Focusing Properties', *Phys. Rev.*, **53**, 521–524 (1938).
121. S. Halas, 'Dual Collector Cycloidal Mass Spectrometer for Precision Analysis of Hydrogen Isotopes', *Int. J. Appl. Rad. Isotopes*, **36**, 957–960 (1985).
122. A. Lamberty, P. De Bièvre, A. Goetz, 'CBNM-International Measurement Evaluation Programme (IMEP), IMEP-2: Cd in Polyethylene', *Fresenius J. Anal. Chem.*, **345**, 310–313 (1993).
123. W.R. Kelly, L. Chen, J.W. Gramlich, K.E. Hehn, 'Determination of Sulphur in Low Sulphur Steels by Isotope Dilution Thermal Ionisation Mass Spectrometry', *Analyst*, **115**, 1019–1024 (1990).
124. J.R. Walton, A.E. Cameron, R.L. Walker, T.L. Hebble, 'Determination of the Abundance of Krypton in the Earth's Atmosphere by Isotope Dilution Mass Spectrometry', *Int. J. Mass Spectrom. Ion Phys.*, **12**, 439–453 (1973).
125. D. Halliday, A.G. Miller, 'Precise Measurement of Total Body Water Using Trace Quantities of Deuterium Oxide', *Biomed. Mass Spectrom.*, **4**, 82–87 (1977).
126. N. Lifson, W.S. Little, D.G. Levitt, R.M. Henderson, 'D_2 ^{18}O Method for CO_2 Output in Small Animals and Economic Feasibility in Man', *J. Appl. Physiol.*, **39**, 657–664 (1975).
127. R.K. Wanless, H.G. Thode, 'The Mass Spectrometer Determination of the Half-life of ^{85}Kr', *Can. J. Phys.*, **31**, 517–519 (1953).
128. C.C. McMullen, K. Fritze, R.H. Tomlinson, 'The Half Life of Rubidium-87', *Can. J. Phys.*, **44**, 3033–3038 (1966).
129. G. Faure, *Principles of Isotope Geology*, John Wiley & Sons, 1977.
130. D. York, R.M. Farquhar, *The Earth's Age and Geochronology*, Pergamon Press, 1972.
131. B.R. Doe, *Lead Isotopes*, Springer-Verlag, 1970.
132. D.J. De Paulo, *Neodynium Isotope Geochemistry: An Introduction*, Springer-Verlag, 1988.
133. R. Würtz, Dietmar Bege, 'Determination of Burn-up Behaviour of Gadolinium-containing Nuclear Fuel by Isotope Dilution Mass Spectrometry', *Adv. Mass Spectrom.*, **7A**, 499–503 (1978).
134. International Atomic Energy Agency, 'Proc. Int. Symp. On The Oklo Phenomenon', Libreville, IAEA-SM-204, IAEA, Vienna, 1975.
135. R. Hagemann, M. Lucas, G. Nief, E. Roth, 'Measures Isotopiques du Rubidium et du Strontium de la Mineralisation de l'Uranium du Reacteur Naturel Fossile d'Oklo', *Earth Planet. Sci. Lett.*, **23**, 170–176 (1974).
136. International Atomic Energy Agency, 'Stable Isotopes in Plant Nutrition, Soil Fertility and Environmental Studies', IAEA, Vienna, 1991.
137. E. Wada, T. Yoneyama, M. Minagawa, T. Ando, B.D. Fry (eds.), *Stable Isotopes in the Biosphere*, Kyoto Univ. Press, 1995.
138. B. Mayer, P. Fritz, J. Prietzel, H.R. Krouse, 'The Use of Stable Sulfur and Oxygen Isotope Ratios for Interpreting the Mobility of Sulfate in Aerobic Forest Soils', *Appl. Geochem.*, **10**, 161–173 (1995).
139. H.R. Krouse, V.A. Grinenko (eds.), *Stable Isotopes: Natural and Anthropogenic Sulphur in the Environment*, SCOPE 43, John Wiley & Sons, 1991.
140. D.C. Coleman, B. Fry (eds.), *Carbon Isotope Techniques*, Academic Press, 1991.
141. J. Lloyd, G.D. Farquhar, '^{13}C Discrimination during CO_2 Assimilation by the Terrestrial Biosphere', *Oecologia*, **99**, 201–215 (1994).

142. H. Griffiths (ed.), *Stable Isotopes: Integration of Biological, Ecological, and Geochemical Processes*, Bios Scientific Publishers, 1998.
143. M.J. Schoeninger, M.J. DeNiro, 'Nitrogen and Carbon Isotope Composition of Bone Collagen from Marine and Terrestrial Animals', *Geochim. Cosmochim. Acta*, **48**, 625–639 (1984).
144. B.C. Peterson, R.W. Howarth, R.H. Garritt, 'Multiple Stable Isotopes used to Trace the Flow of Organic Matter in Estuarine Food Webs', *Science*, **227**, 1361–1363 (1985).
145. D.M. Schell, S.M. Saupe, N. Haubenstock, 'Natural Isotope Abundances in Bowhead Whale (*Balaena mysticetus*) Baleen: Markers of Aging and Habitat usage', in *Stable Isotopes in Ecological Research*, eds. P.W. Rundel, J.R. Ehleringer, K.A. Nagy, Springer-Verlag, 260–269, 1988.
146. C.D. Keeling, W.G. Mook, P. Tans, 'Recent Trends in the $^{13}C/^{12}C$ Ratio of Atmospheric Carbon Dioxide', *Nature*, **277**, 121–123 (1979).
147. I. Friedman, J.R. O'Neil, 'Compilation of Stable Isotope Fractionation Factors of Geochemical Interest', *Data of Geochemistry*, ed. M. Fleischer, 6th edition, US Geological Survey Prof. Paper 440-KK, US Geological Survey, Denver, 1977.
148. H.C. Urey, S. Epstein, C. McKinney, J. McCrae, 'Method for Measurement of Paleotemperatures', *Bull. Geol. Soc. Amer.*, **59**, 1359–1360 (1948).
149. Y. Kajiwara, H.R. Krouse, 'Sulfur Isotope Partitioning in Metallic Sulfide Systems', *Can. J. Earth Sci.*, **8**, 1397–1408 (1971).
150. H.P. Taylor, Jr, J.R. O'Neil, I.R. Kaplan (eds.), *Stable Isotope Geochemistry: A Tribute to Samuel Epstein*, Geochem. Soc. Spec. Pub. No. 3, 1991.
151. L. Melander, W.H. Saunders, *Reaction Rates of Isotopic Molecules*, John Wiley & Sons, 1980.
152. C.J. Collins, N.S. Bowman (eds.), *Isotope Effects in Chemical Reactions*, Van Nostrand Reinhold, New York, 1970.
153. D.R. Van Stempvoort, H.R. Krouse, 'Controls of $\delta^{18}O$ in Sulfate: Review of Experimental Data and Application to Specific Environments', in *Environmental Geochemistry of Sulfide Oxidation*, eds. C.N. Alpers, D.W. Blowes, American Chemical Society, 446–480, 1994.
154. I.R. Kaplan, S.C. Rittenberg, 'Microbiological Fractionation of Sulphur Isotopes', *J. Gen. Microbiol.*, **34**, 195–212 (1964).
155. L.A. Chambers, P.A. Trudinger, 'Microbiological Fractionation of Stable Sulfur Isotopes: A Review and Critique', *Geomicrobiol. J.*, **1**, 249–293 (1979).
156. W. Dansgaard, 'Stable Isotopes in Precipitation', *Tellus*, **16**, 436–468 (1964).
157. K. Rozanski, L. Araguás-Araguás, R. Gonfiantini, 'Isotopic Patterns in Modern Global Precipitation', in *Climate Change in Continental Isotopic Records*, eds. P.K. Swart, K.C. Lohman, J. McKenzie, S. Savin, Geophys. Mono. 78, American Geophysical Union, Washington, DC, 1–36, 1993.
158. H.R. Krouse, J.R. Mackay, 'Application of H_2O^{18}/H_2O^{16} Abundances to the Problem of Lateral Mixing in the Liard–Mackenzie River System', *Can. J. Earth Sci.*, **8**, 1107–1109 (1971).
159. M.O. Jeffries, H.R. Krouse, M.A. Shakur, S.A. Harris, 'Isotope Geochemistry of Stratified Lake 'A', Ellesmere Island, N.W.T., Canada', *Can. J. Earth Sci.*, **21**, 1008–1017 (1984).
160. C. Lorius, J. Jouzel, C. Ritz, L. Merlivat, N.I. Barkov, Y.S. Korotkevich, V.M. Kotlyakov, 'A 150 000 Year Climatic Record from Antarctic Ice', *Nature*, **316**, 591–596 (1985).
161. C.J. Yapp, S. Epstein, 'Climatic Implications of D/H Ratios of Meteoric Water over North America (9500–22 000 B.P.) as Inferred from Ancient Wood Cellulose C-H Hydrogen', *Earth Planet. Sci. Lett.*, **34**, 333–350 (1977).
162. N.J.L. Bailey, H.R. Krouse, C.R. Evans, M.A. Rogers, 'The Alteration of Crude Oil by Waters and Bacteria; Evidence from Geochemical and Isotope Studies', *AAPG Bull.*, **57**, 1276–1290 (1973).
163. M.A. Katzenberg, H.R. Krouse, 'Application of Stable Isotopes in Human Tissues to Problems in Identification', *Can. Soc. Forensic Sci. J.*, **22**, 7–19 (1989).
164. H. Moser, W. Rauert, *Isotopenmethoden in der Hydrologie. Lehrbuch der Hydrogeologie, Band 8*, Gebrueder Borntraeger, Berlin, Stuttgart, 1980.
165. I. Clark, P. Fritz, *Environmental Isotopes in Hydrogeology*, Lewis Publishers, 1997.
166. C. Kendall, J.J. McDonnell (eds.), *Isotope Tracers in Catchment Hydrology*, Elsevier, 1998.
167. J. Hoefs, *Stable Isotope Geochemistry*, 4th edition, Springer, 1997.
168. P. Fritz, J. Ch. Fontes (eds.), *Handbook of Environmental Isotope Geochemistry, 1, The Terrestrial Environment*, A, Elsevier Scientific Publishing, 1980.
169. G.E. Claypool, W.T. Holser, I.R. Kaplan, H. Sakai, I. Zak, 'The Age Curves of Sulfur and Oxygen Isotopes in Marine Sulfate and Their Mutual Interpretation', *Chem. Geol.*, **28**, 199–260 (1980).
170. H. Strauss, 'The Isotopic Composition of Sedimentary Sulfur Through Time', *Palaeogeography, Palaeoclimatology, Palaeoecology*, **132**, 97–118 (1997).
171. W.H. Burke, R.E. Denison, E.A. Hetherington, R.B. Koepnick, H.R. Nelson, J.B. Otto, 'Variation of Seawater $^{87}Sr/^{86}Sr$ Throughout Phanerozoic Time', *Geology*, **10**, 516–519 (1982).
172. T.K. Kyser (ed.), *Stable Isotope Geochemistry of Low Temperature Fluids*, Mineralogical Association of Canada, Toronto, 13, 1987.
173. J.W. Valley, H.P. Taylor, Jr, J.R. O'Neil (eds.), *Stable Isotopes in High Temperature Geological Processes*,

Reviews in Mineralogy, 16, Mineralogical Society of America, 1986.
174. E.M. Galimov, 'Isotopic Fractionation Related to Kimberlite Magmatism and Diamond Formation', *Geochim. Cosmochim. Acta*, **55**, 1697–1708 (1991).
175. D. Emery, A. Robinson with contributions from A. Alpin, C. Smalley, *Inorganic Geochemistry: Applications to Petroleum Geology*, Blackwell Scientific Publications, London, 254, 1993.
176. A.N. Fuex, 'The use of Stable Carbon Isotopes in Hydrocarbon Exploration', *J. Geochem. Explor.*, **7**, 155–188 (1977).
177. M. Schoell, 'Recent Advances in Petroleum Isotope Geochemistry', *Org. Geochem.*, **6**, 645–663 (1984).
178. M. Schoell, 'Multiple Origins of Methane in the Earth', *Chem. Geol.*, **71**, 1–10 (1988).
179. H.R. Krouse, 'Sulphur Isotope Studies and their Role in Petroleum Exploration', *J. Geochem. Explor.*, **7**, 189–211 (1977).
180. H.R. Krouse, C.A. Viau, L.S. Eliuk, A. Ueda, S. Halas, 'Chemical and Isotopic Evidence of Thermochemical Sulphate Reduction by Light Hydrocarbon Gases in Deep Carbonate Reservoirs', *Nature*, **333**, 415–419 (1988).
181. J.W. Smith, B.D. Batts, 'The Distribution and Isotopic Composition of Sulfur in Coal', *Geochim. Cosmochim. Acta*, **38**, 121–133 (1974).
182. F.T. Price, Y.N. Shieh, 'The Distribution and Isotopic Composition of Sulfur in Coals from the Illinois Basin', *Econ. Geol.*, **74**, 1445–1461 (1979).
183. I. Hutcheon, H.J. Abercrombie, H.R. Krouse, 'Inorganic Origin of Carbon Dioxide During Low Temperature Thermal Recovery of Bitumen: Chemical and Isotopic Evidence', *Geochim. Cosmochim. Acta*, **54**, 165–171 (1990).
184. B. Fry, R.S. Scalan, J.K. Winters, P.L. Parker, 'Sulphur Uptake by Salt Grasses, Mangroves and Seagrasses in Anaerobic Sediments', *Geochim. Cosmochim. Acta*, **46**, 1121–1124 (1982).
185. B. Fry, H. Gest, J.M. Hayes, 'Sulphur Isotopic Compositions of Deep Sea Hydrothermal Vent Animals', *Nature*, **306**, 51–52 (1983).
186. M.J. DeNiro, S. Epstein, 'Influence of Diet on the Distribution of Carbon Isotopes in Animals', *Geochim. Cosmochim. Acta*, **42**, 495–506 (1978).
187. M. Minigawa, E. Wada, 'Stepwise Enrichments of ^{15}N along Food Chains: Further Evidence and the Relation Between $\delta^{15}N$ and Animal Age', *Geochim. Cosmochim. Acta*, **48**, 1135–1140 (1984).
188. E.M. Burbidge, G.R. Burbidge, W.A. Fowler, F. Hoyle, 'Synthesis of Elements in Stars', *Rev. Mod. Phys.*, **29**, 547–650 (1957).
189. H. Craig, S.L. Miller, G.J. Wasserburg (eds.), *Isotopic and Cosmic Chemistry*, North Holland, 1964.
190. R.N. Clayton, 'Isotopic Anomalies in the Early Solar System', *Ann. Rev. Nucl. Part. Sci.*, **28**, 501–522 (1978).
191. Proceedings of the Sixth Lunar Science Conference, 2, Chemical and Isotopic Studies', *Geochim. Cosmochim. Acta, Suppl.*, **6**, 1203–2426 (1975).
192. A.O. Nier, M.B. McElroy, Y.L. Yung, 'Isotopic Composition of the Martian Atmosphere', *Science*, **194**, 68–70 (1976).
193. T. Owen, K. Biemann, D.R. Rushneck, J.E. Biller, D.W. Howarth, A.L. Lafleur, 'The Composition of the Atmosphere at the Surface of Mars', *J. Geophys. Res.*, **82**, 4635–4639 (1977).
194. J.H. Hoffman, R.R. Hodges, M.B. McElroy, T.M. Donahue, M. Koplin, 'Composition and Structure of the Venus Atmosphere: Results from Pioneer Venus', *Science*, **205**, 49–52 (1979).
195. T.M. Donahue, J.H. Hoffman, R.D. Hodges, A.J. Watson, 'Venus was Wet: A Measurement of the Ratio of Deuterium to Hydrogen', *Science*, **216**, 630–633 (1982).
196. S. Wyckoff, 'Comets: Clues to the Early History of the Solar System', *Earth Sci. Rev.*, **30**, 125–174 (1991).

Liquid Chromatography/Mass Spectrometry

Frederick E. Klink
Scientific Training and Marketing, San Ramon, USA

1 Introduction and History 1
1.1 Development of Solvent Elimination/Gas-phase Ionization Interfacing for Liquid Chromatography/Mass Spectrometry 4
1.2 Development of Desorption/Ionization Techniques for Liquid Chromatography/Mass Spectrometry 4
2 Liquid Chromatography Considerations for Mass Spectrometry 5
3 Modern Methods of Interfacing Liquid Chromatography Effluent with Mass Spectrometry 7
3.1 Particle Beam Interface 7
3.2 Atmospheric Pressure Ionization Interfaces 8
4 Collisionally Activated Dissociation in Liquid Chromatography/Mass Spectrometry 13
5 Sample and Solvent Considerations in Method Development 15
5.1 Spectral Interference 15
5.2 System Compromise 15
5.3 Adduct Formation 15
5.4 Ion Suppression 16
Abbreviations and Acronyms 16
Related Articles 16
References 17

*Liquid chromatography/mass spectrometry (LC/MS) is the combination of high-performance liquid chromatography (HPLC) and mass spectrometry (MS) into a single, continuous-flow analytical system. HPLC separates analytes from their matrix and from one another in a flowing system on a time basis. These separated analytes are removed from the solvent stream and introduced to the MS system through one of several types of available interfaces which facilitate ionization of the analyte molecules. MS is an analytical technique which separates these analyte ions on the basis of their mass-to-charge ratio (*m/z*). Data collected from LC/MS is a series of mass spectra taken over time as the chromatographic effluent exits the column and is analyzed by the mass spectrometer. The three interface types in common usage are particle beam (PB), atmospheric pressure chemical ionization (APCI) and electrospray interface (ESI). These interfaces differ in the characteristics of the mass spectra that they produce, the types of chemical compounds for which they are best suited and the interferences to which they are subject. With the combined capability of two of these interfaces, APCI and ESI, LC/MS is able to address nearly any type of organic molecule from small molecules to biological macromolecules. ESI has ushered in a revolution in biochemical research on large molecules although the technology itself is not completely understood and is subject to a variety of interferences. APCI is limited to smaller molecules and is most at home in routine use for very high-throughput drug and metabolite analyses. The PB interface, although not as flexible as the other two, is able to produce electron ionization (EI) spectra which can be searched against standard, commercial libraries. ESI and APCI are not able to produce EI spectra but, through the use of collisionally activated dissociation (CAD), are able to produce usable spectra in both single-stage and tandem mass spectrometry (MS/MS).*

1 INTRODUCTION AND HISTORY

Scientists using LC (liquid chromatography) techniques have long wished for a reliable, easy-to-use mass spectral detector. On the other hand, workers approaching this issue from the MS standpoint have always desired a direct liquid interface to the mass spectrometer which could handle samples in solution without requiring extensive sample workup. The solution to both of these problems has been the development of LC/MS. However, this is a case where the whole is certainly greater than the sum of its parts.

Both MS and HPLC (often shortened to LC) are separation techniques. In HPLC, components are separated based on their relative chemical and/or physical affinity for some type of stationary phase. This is measured as the time it takes the component to traverse the chromatographic column. This separation is represented by a 'chromatogram' which is a plot of detector response versus time. The separation in MS occurs on the basis of mass-to-charge ratio (m/z). The representation of this separation is a 'mass spectrum' which plots detector response versus m/z. (Note: it is very important always to think in terms of mass-to-charge ratio and not just mass when discussing mass spectra.)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The combination of two separation techniques can result in a technique which is not just twice as powerful but actually orders of magnitude more powerful than either technique alone. If the two techniques are completely orthogonal, in other words, they separate components based on completely different physical principles, then the 'peak capacity', i.e. the number of separate components the technique can theoretically resolve, is the *product* of the peak capacities of the two techniques. A very conservative estimate of the peak capacity of modern HPLC is 100 and that for benchtop MS is 500. If we then assume completely orthogonal separation mechanisms, the combined technique results in a peak capacity of 50 000! This is a peak capacity which is very improbable to impossible to achieve with one technique alone, no matter how many LC columns are coupled together or mass analyzers are linked.

In their landmark textbook, *Introduction to Modern Liquid Chromatography*, Snyder and Kirkland enumerate the attributes of the ideal detector for LC.[1] Those attributes are listed here along with comments on how the mass spectrometer meets them:

- Has high sensitivity and the same predictable response (LC/MS is routinely used for picomolar or better sensitivities with excellent reproducibility).
- Has response to all solutes, or has a predictable specificity (generally, if the solute can be ionized, it can be seen in the mass spectrometer).
- Has a wide range of linearity (true for all commercially available LC/MS interfaces).
- Is unaffected by changes in temperature and mobile phase flow (true for LC/MS within typical laboratory variations).
- Responds independently of the mobile phase (generally true, although mobile phase constituents may effect ionization efficiency).
- Does not contribute to extra column band broadening (modern LC/MS interfaces, properly set up and using correct flow rates, do not present a problem).
- Is reliable and convenient to use (LC/MS is in routine use in thousands of laboratories and its use is growing rapidly).
- Has a fast response (certainly true).
- Provides qualitative information on the detected peak (this is key advantage of MS – generation of spectra which provide extensive information about compound structure).
- Is nondestructive of the solute (this is not true, at least for benchtop instruments in routine use. There are nondestructive MS techniques: Siuzdak et al.[2] were able to analyze an intact virus via Fourier transform ion cyclotron resonance MS, recover the virus and use it successfully to infect a host organism).

MS meets most, if not all, of these requirements. The most significant features of MS are (1) the ability to produce ions which represent the intact molecule and can therefore give a direct and extremely accurate measure of the molecular weight of the analyte compound and (2) the ability to produce characteristic fragment ion spectra which can be interpreted to give at least partial, and often complete, structural information about the molecule.

The discussion of MS as a detection technique leads directly to the question of what sort of data are produced by an LC/MS system. Figure 1 presents a schematic answer to that question. Starting at point (a) we have a typical chromatographic peak recorded as an analog signal over time from a standard HPLC detector. A computer can digitize this signal (b) and further process that information into a report of peak areas and retention times (c). This process is familiar to chromatographers.

Mass spectral data are very different. When we connect a mass spectrometer up to the effluent from the LC column we do not obtain nice analog peaks coming off one after another. The mass spectrometer simply monitors a

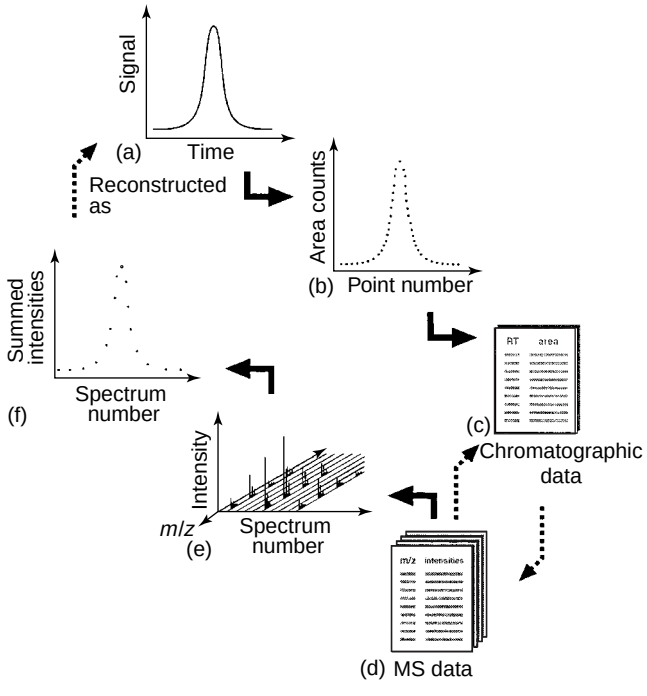

Figure 1 The data generation pathways in LC/MS. (Reprinted with permission from O.D. Sparkman and F.E. Klink, *Liquid Chromatography/Mass Spectrometry Fundamentals and Applications*, ACS Short Courses, American Chemical Society, Washington, DC, 1999. Copyright 1999 American Chemical Society.)

series of m/z values and records the intensity of the ion current at each value (Figure 1d). Once this series of values has been recorded, the instrument starts over and makes another set of recordings and so on until the operator stops the data collection. These recordings may be as slow as one per second or as many as 100 or more per second, depending on the type of instrument and how it is set up. This series of m/z vs intensity data sets can be plotted as spectra such as those shown in Figure 1 at point (e). Each spectrum represents a 'snapshot' in time of the ions passing out of the mass spectrometer at that moment. [The three-dimensional nature of the data in (e) is also a good illustration of the orthogonal nature of LC and MS.]

Those spectra in themselves are the reason why we use MS with LC since they are very valuable for qualitative analysis of the chromatographic components. Still, they do not look much like chromatography. However, by summing all of the intensities in each spectrum we reduce each spectrum to a single point. That single point can be plotted against the time at which the spectrum was collected (f) and by a connect-the-dots approach we have a chromatogram again (a). Now, the chromatographic information can be taken into a standard chromatographic data system, via the path (a)–(b)–(c) and quantitative analysis done just as we would with other HPLC detectors.

Creating chromatographic peaks in this manner is termed 'reconstruction' and the resulting chromatogram is called a reconstructed total ion current chromatogram (RTICC). We may not choose to look at all of the available ion current but only selected m/z values. If we can find an m/z value (i.e. an ion) which comes only from our analyte of interest, we can plot a chromatogram based only on that one ion. This technique is called selected ion monitoring (SIM) and is a very powerful selectivity tool in LC/MS. In fact, the mass spectral data set gives us great flexibility to reconstruct dozens of different types of chromatograms and to extract tremendous amounts of useful information, all from a single chromatographic run.

Clearly, the issue holding back development in LC/MS was not mere recognition of the natural synergy between LC techniques and MS but rather the development of interfacing strategies which could bring these advantages to the average scientist's laboratory bench. By comparison, the interfacing of a capillary gas chromatography (GC) effluent to MS was relatively straightforward. In gas chromatography/mass spectrometry (GC/MS) a high-temperature gas-phase material is introduced into a high-temperature vacuum environment. The analytes of interest are much higher in mass than the helium carrier gas and are readily separated from it by their relative momentum (see **Gas Chromatography/Mass Spectrometry**).

This is not so straightforward for LC/MS. An amusing illustration of this 'odd couple' relationship was published by Arpino[3] and appears in Figure 2. The HPLC analysis lives in a solution environment at elevated pressure whereas the mass spectrometer requires a high vacuum environment with ionized gas-phase analytes. Developing a 'meaningful relationship' between these two is therefore a considerable challenge.

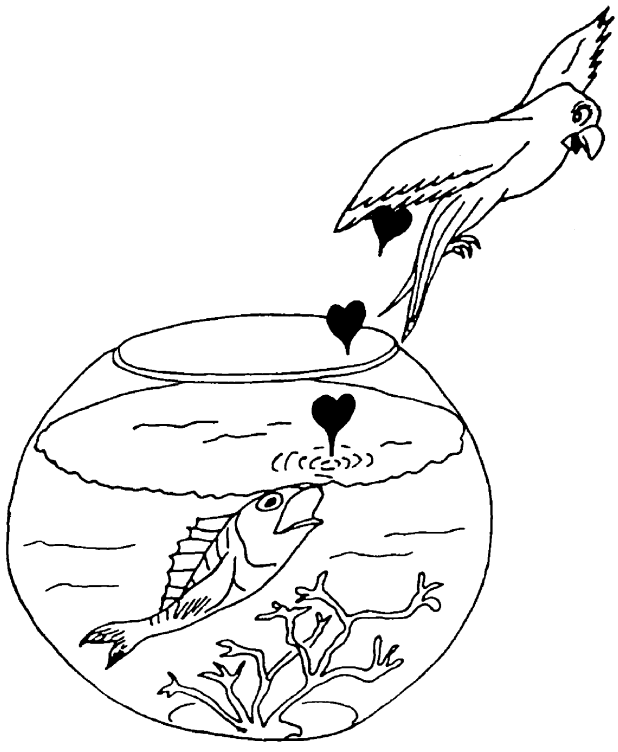

Figure 2 Arpino's illustration of the challenge of interfacing the HPLC 'fish' with the MS 'bird'. (Reprinted from P.J. Arpino, 'On-line Liquid Chromatography/Mass Spectrometry? An Odd Coupled', *Trends Anal. Chem.*, **1**, 154–158 (1982) with permission of Elsevier Science.)

The primary problem in interfacing LC and MS is the elimination of solvent and the preservation of sufficient amounts of analyte to give reasonable detection limits. A flowing system delivering $1\,mL\,min^{-1}$ acetonitrile becomes $533\,mL\,min^{-1}$ of gas when vaporized at standard temperature and pressure (STP). This solvent may be carrying micrograms per milliliter or less of the analyte which must also be put into the gas phase in order to enter the mass analyzer of the MS instrument. Hence the challenge is to eliminate literally liters of unwanted solvent vapor while selectively retaining infinitesimally smaller amounts of analyte. The problem is further compounded by the fact that many LC analytes are minimally volatile or involatile and may also be thermally labile.

Two general technical approaches have been taken to produce successful LC/MS interfaces: solvent elimination/gas-phase ionization and desorption/ionization.

1.1 Development of Solvent Elimination/Gas-phase Ionization Interfacing for Liquid Chromatography/Mass Spectrometry

The first approach to solvent elimination and gas-phase ionization was the obvious approach of simply plugging an LC effluent stream directly into the mass spectrometer ion source. This direct liquid introduction technique was proposed by Talroze et al.[4] in 1972 and demonstrated practically by Arpino et al.[5] in 1974. The technique was limited by the very low column flow rates required ($10–50\,\mu L\,min^{-1}$), which were impractical for HPLC applications at that time. The solvent stream was nebulized through a 4-μm orifice into a heated desolvation chamber and then into a chemical ionization (CI) source. Volatilized solvent acted as the ionization reagent. The direct inlet probe, as this interface was known, suffered from clogging of the orifice and short filament life in the CI source. Nevertheless, some basic work on pesticide analysis and biomolecules was accomplished.

In the same year that Arpino et al. publicized the direct inlet probe, Hornung et al.[6] published their work on atmospheric pressure ionization (API). This technique is very similar to the direct inlet probe except that nebulization and solvent-mediated CI take place in an atmospheric pressure chamber outside the mass spectrometer. Now known as APCI, this technique is one of the most popular LC/MS techniques in use today, particularly for high-sensitivity quantitative analysis. The technique is described in detail later in this article.

The moving belt interface of McFadden et al.[7] was a clever and direct approach to the issue of solvent elimination followed by analyte volatilization. Introduced in 1976, this interface was initially designed using a wire loop and later a belt which allowed greater surface area for LC effluent. The belt was moved into and out of the vacuum of the mass spectrometer by drive wheels, much like a conveyer belt. LC effluent from the end of the column or detector dripped on to the belt on the atmospheric end and then passed through a series of increasingly higher vacuum regions accompanied by mild heating which removed the residual solvent. In the final step, the belt, now with dry analyte on its surface, moved into the ion source of the mass spectrometer where the analyte was flash vaporized into the vacuum region. The belt was heated and cleaned during its return journey to the atmospheric side of the interface. Both the moving belt and direct liquid inlet could be used with a generic ion source which could accommodate other types of sample introduction, e.g. GC, when the HPLC interface as not in use.

The thermospray interface[8,9] was first described in 1983 by Vestal et al. In this interface, shown in Figure 3, the LC effluent is introduced through a heated probe into a reduced pressure area (5×10^{-15} Torr), forming a jet of droplets. The heat imparted allows the solvent to evaporate from these droplets and to be swept out through the vacuum system. Meanwhile, ions are formed either by solvent-mediated CI or by ion evaporation processes. In the CI mode, the solvent acts as the reagent gas and is ionized by either a stream of high-energy electrons from a filament or by a corona discharge electrode. One mode or the other is selected based on the mobile phase aqueous–organic composition. Aside from operation at reduced pressure, this is the same mechanism employed in APCI. The ion evaporation mode was used with both the filament and discharge electrodes off. It is analogous to ESI, which is described later.

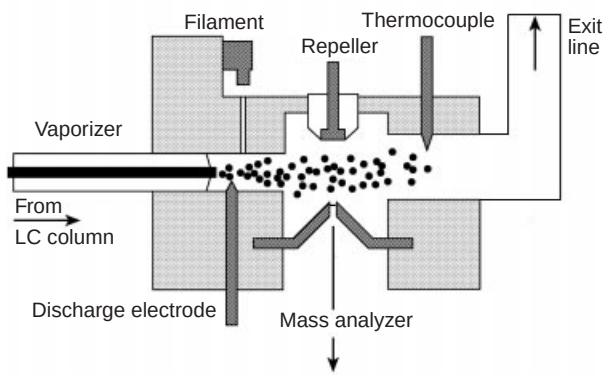

Figure 3 The thermospray LC/MS interface. Filament or discharge electrode facilitates ion formation through solvent-mediated CI. With filament and electrode off, ions are formed through ion evaporation. (Reprinted with permission from O.D. Sparkman and F.E. Klink, *Liquid Chromatography/Mass Spectrometry Fundamentals and Applications*, ACS Short Courses, American Chemical Society, Washington, DC, 1999. Copyright 1999 American Chemical Society.)

After the introduction of thermospray, the floodgates opened to new LC/MS interfacing techniques. The 1980s was by far the most significant decade in the history of the technique.

The PB [or monodisperse aerosol generation interface for chromatography (MAGIC)] interface was developed by Willoughby and Browner in 1984[10] and has been applied to both HPLC and GC. This interface is still commercially available and popular today and therefore will be described later. At this point, it is sufficient to say that the PB system eliminates solvent through nebulization and desolvation assisted by heat and vacuum. The analyte is then ionized in a standard EI source. The PB is thus the only LC/MS interface currently available from commercial sources which is capable of producing standard EI mass spectra.

1.2 Development of Desorption/Ionization Techniques for Liquid Chromatography/Mass Spectrometry

The second technical approach, desorption/ionization, involves direct ionization of analyte from a liquid or solid

matrix without initial solvent elimination or vaporization of the analyte. The desorption/ionization techniques have had their greatest impact in the analysis of biomolecules. Since the analyte is ionized directly from a liquid or solid matrix, these techniques are not limited to the analysis of volatile molecules. Proteins, nucleic acids and other biological macromolecules are easily ionized and brought into the mass spectrometer by desorption/ionization.

A desorption/ionization technique for the analysis of static liquid samples was introduced in 1981 and termed fast atom bombardment (FAB).[11] (The technique is also known as 'static FAB' to distinguish it from 'continuous-flow fast atom bombardment (CF/FAB)', which is discussed below.) FAB relies on introduction of approximately 1 μL of sample in an involatile matrix such as glycerol. The glycerol solution is introduced on a probe directly into the MS source. The probe geometry positions the sample where it can be struck by a beam of atoms, e.g. xenon. The impact of the atom beam on the sample sputters the analytes off of the surface layer which are in turn ionized through a variety of processes. Analyte ions are then focused and directed into the mass analyzer. The glycerol matrix remains as a liquid for several minutes in the vacuum, which allows continuous production of sample ions over this time.

In 1985–86, two developments appeared which allowed the use of FAB with flowing LC effluents. The first was 'frit FAB'[12] and the second was termed CF/FAB.[13] CF/FAB relies on the same principal as static FAB but rather than using a drop of sample–matrix solution, a flowing sample is employed. The sample is flowed continuously up through the sample probe using a solution that is approximately 5% glycerol rather than the 90% glycerol matrix used in static FAB. The aqueous–organic portion of the solution evaporates quickly, leaving the glycerol matrix and sample behind. As solution is pumped through the probe, the lower viscosity of the fresh, aqueous solution forces it to the top of the bubble where it is exposed to the atom beam and is ionized. The lower percentage of glycerol reduces background interference compared with static FAB and the continuous flow allows interfacing with LC. However, CF/FAB flow rates are of the order of 5 μL min^{-1}, necessitating the use of split-flow systems with standard HPLC or the employment of micro LC or capillary electrophoresis.

Both static FAB and CF/FAB have been employed in biological macromolecule analysis.[14] However, these techniques are being largely superseded in biotechnology applications by the desorption/ionization techniques, matrix-assisted laser desorption/ionization (MALDI) and ESI. MALDI[15] is a static technique in which samples are prepared in a specific matrix, placed on a target or plate and allowed to dry. This plate is then placed in the ion source of the mass spectrometer where an ultraviolet (UV) or infrared (IR) laser is fired at the sample causing a type of CI between the matrix and sample molecules. Choice of matrix is one of the critical operating parameters in MALDI. The routine use of MALDI is limited to static operation although some work is under way to develop a continuous-flow version of MALDI using a moving tape interface.[16]

Ironically, the process which today is generating the most attention in the interfacing of LC with MS is also the process which has the longest history in the literature. The electrospray process, which relies on nebulization of a solvent stream into an electrical field, was first described by Zeleny[17] in 1917. Its applicability as a mass spectral interfacing technique was shown by Dole et al.[18] in 1968. Dole et al. described the major advantage of the ESI technique for LC/MS, namely the ability to produce macromolecular ions. This opened the door to biological investigations using MS. ESI was introduced for bioapplications by Whitehouse et al. in 1984[19] and further refined for use with typical HPLC flow rates by Bruin et al.'s concept of pneumatically assisted ESI, or ionspray, in 1987.[20] Both MALDI and ESI techniques are highly applicable to large, involatile analyte molecules and will be discussed at length later in this article.

2 LIQUID CHROMATOGRAPHY CONSIDERATIONS FOR MASS SPECTROMETRY

The technique of HPLC is truly the analytical technique of the 20th century. Michel Tsvet's seminal paper first describing liquid–solid partition chromatography[21] appeared just after the turn of the 20th century. Martin and Synge's Nobel Prize-winning work in 1941[22] described much of the theoretical foundation for later HPLC developments. It was not, however, until the late 1960s that high-pressure pump hardware became available to the analytical community at large and it was from this point that HPLC use started a meteoric rise. Majors' development of microparticulate bonded phases in 1972[23] allowed the technique rightfully to take on the name 'high performance'. In less than 20 years, the use of HPLC grew from a curiosity to the most widely used analytical technique in the world. Its applications span all forms of chemical and biological research, biomedical applications, environmental analysis and nearly every industrial application of chromatography. While the analytical use of HPLC is the most widespread, there is a very large segment of the user population who take advantage of the nondestructive nature of the technique for preparative applications (see the entries in *Liquid Chromatography*).

Most present-day users of analytical HPLC employ 3–4.6 mm inner diameter (ID) columns using solvent flow rates of 0.5–2 mL min^{-1}. Growing interest is seen in 'fast' LC columns, i.e. shorter length columns but still of ~4 mm ID.[24] This has remained the case for over 20 years in spite of the development during that time of micro-HPLC. The development of the 'micro' technique began with smaller diameter columns of 1 or 2 mm ID up to the present day when nano-LC columns may be of <100 μm ID. In the last few years, the major HPLC manufacturers have begun to produce high-performance pumps to deliver solvents accurately at the lower flow rates required by these columns. However, to switch over to the micro-LC techniques requires replacing nearly all of the components of the HPLC system. Extra-column band-broadening effects from the internal volumes of tubing and injection systems which are not suited to small-diameter columns will compromise the results of analyses. Analytical results may, in some cases, be poorer with small diameter columns if these hardware issues are not taken into account.

HPLC's success is the very thing which holds back its rapid technological evolution in the average analytical laboratory: the financial investment required to replace outright many thousands of dollars worth of equipment is not justified by solvent or sample savings realized by use of micro-HPLC techniques.

LC/MS may change this picture. ESI/MS and APCI/MS are highly sensitive techniques which have no difficulty detecting the very small amounts of sample introduced through micro-HPLC columns. Even more important, micro-flow rates actually improve electrospray performance and sensitivity. The electrospray technique was originally developed for use with very low flow rates. Subsequent developments in ESI (such as ionspray described earlier) focused on adapting the technique to higher flow rates but this came at the cost of performance. With the development of practical micro-HPLC systems, users are rediscovering the outstanding performance of ESI as a low-flow technique. This and other aspects of low-flow-rate HPLC will be discussed later.

There are three basic LC hardware configurations for interfacing a flowing liquid effluent to an MS interface, as illustrated in Figure 4(a–c). Figure 4(a) shows the simplest configuration of flowing liquid interface. A simple pump, e.g. a syringe pump, drives liquid sample directly into the MS interface. This is referred to as continuous infusion or simply infusion. This technique is very useful in method development because it can be used to provide a continuous background of the analyte spectrum in the mass spectrometer. This allows tuning of the mass spectral conditions and flow rate without having to make multiple injections. If a pump is used which is capable of multisolvent proportioning, this technique

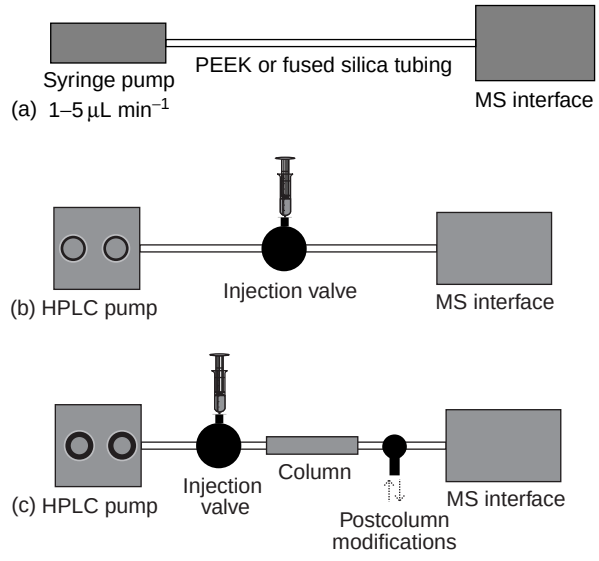

Figure 4 Three modes of interfacing LC output to MS interfaces: (a) direct infusion; (b) flow injection; and (c) complete HPLC/MS hardware.

can also be used to study the effects of various solvent systems on ionization efficiency. The new protein analysis technique called nanospray employs continuous infusion for analytical use and will be discussed later.

Figure 4(b) shows the next step up in complexity. A standard analytical HPLC pump plus an injection valve replace the continuous-flow pump used for infusion. Sample is now introduced by injection into the flowing solvent stream and hence this technique is known as flow injection. This technique is also very useful in method development because the effect of different injection volumes, solvents and concentrations on analyte ionization efficiency can be quickly evaluated. It is also ideal for observing the mass spectral effects of changing mobile phase solvents, additives, pH, etc. without worrying about their impact on a chromatographic separation. It may, in fact, be possible to use this technique in the final analytical method if column separation of the analytes is not required. For example, analytes which are clearly separated by mass-to-charge ratio (m/z) in the mass spectrum can be readily distinguished and quantified without further separation. Qualitative information can also be obtained in this case by the employment of MS/MS in which a single ion representing the intact analyte molecule is subjected to a second stage of MS where fragmentation occurs. This fragmentation gives a pattern characteristic of the structure of the molecule (see **Tandem Mass Spectrometry: Fundamentals and Instrumentation**). A standard HPLC autosampler may be employed in place of the injection valve for high throughput, automated analyses.

The final technique (Figure 4c) involves the use of a complete HPLC system including an analytical column for interface to the MS. For many closely related compounds, e.g. drugs and metabolites, it is not possible to obtain sufficient separation by MS alone. In these cases, an HPLC column separation is required. In other analyses, interferences present in the sample matrix may suppress ion formation of the analyte (ion suppression effects will be described in more detail at the end of this article). Again, an HPLC column separation of matrix components from analyte prior to the MS interface will solve this problem. The HPLC separation can sometimes be compromised because of the additional separation power of the MS. This is becoming increasingly clear in the area of sample preparation. In many recent applications using short columns or solid-phase extraction (SPE) cartridges, it is not so easy to distinguish the 'sample preparation' step from the 'chromatographic' step.[25]

In Figure 4(c), a hardware component has been added after the column and before the MS interface to allow postcolumn modifications to the solvent stream going into the mass spectrometer. The solvent system needed for the HPLC separation may be incompatible with the MS interface. Postcolumn modifications to the solvent stream are then employed to optimize the ionization efficiency.

3 MODERN METHODS OF INTERFACING LIQUID CHROMATOGRAPHY EFFLUENT WITH MASS SPECTROMETRY

In the Introduction, a brief history of the development of LC/MS interfacing techniques was presented. Three significant techniques have survived to the present day and represent most of the technical literature and all of the available commercial instrumentation. The PB interface and APCI rely on volatilization of solvent and analyte followed by gas phase ionization. ESI works on the principle of desorption/ionization and is therefore capable of handling large, involatile analyte molecules.

3.1 Particle Beam Interface

The only currently available commercial PB interface is the Waters Thermabeam™. The PB interface as originally designed required high flow rates of nebulizing gas and exhibited very poor sensitivity in comparison with other LC/MS techniques such as APCI and ESI. The introduction of the Thermabeam™ interface addressed both of these problems primarily through use of a heated nebulizing chamber. Sensitivities in the low-nanogram range are routinely seen with this instrument. Although APCI and ESI are still more sensitive techniques, PB has the advantage of producing true EI spectra. Research work has been done with a micro-PB interface[26] in an attempt to improve sensitivity. Sensitivities of <1 ng have been reported using LC effluent flow rates of $1-5\,\mu L\,min^{-1}$.

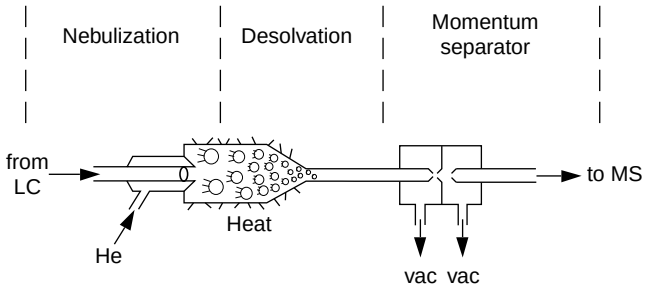

Figure 5 Illustration of the PB interface. (Reprinted with permission from J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, Philadelphia, 1997.)

Figure 5 is a schematic diagram of a PB system. The LC effluent is sprayed through a 5–10 μm ID orifice. This results in the formation of a fine liquid jet which forms a stream of uniformly sized droplets. A stream of high velocity gas, usually He, is introduced at 90° to the liquid jet and serves to disperse the liquid into a fine aerosol. The first step in the interfacing process is the elimination of mobile phase solvent which is vaporized at atmospheric pressure in a 'desolvation chamber'. Under ideal conditions, all of the mobile phase is vaporized, and the analytes remain as a semisolid or less volatile liquid droplet (colorfully referred to as 'dirty snowballs'). The solvent vapor and analyte molecules pass through a two-stage momentum separator to produce a PB of analyte. This PB then strikes a heated plate in the ion source to produce the gas-phase molecules. These molecules rebound off the plate to be ionized by a 70-eV electron beam in a conventional EI source. As with other EI techniques, the ions formed from intact molecules are typically odd-electron species, $M^{+\bullet}$ or $M^{-\bullet}$.

An EI source operates, by convention, at 70 eV. This level of ionizing energy ensures the maximum possible ionization efficiency. Figure 6 shows ion current produced by varying levels of ionization energy. In region A of the curve, the only ions formed are molecular ions. In region B, some fragmentation begins to occur owing to the higher energy imparted to the ions. Finally, in region C, the ion current has reached its maximum. In this region, significant amounts of fragment ions are formed from decomposition of molecular ions and it is these fragment ion spectra which provide structural information about the molecule. Operation at 70 eV ensures that spectra will be collected well into the C region for all molecules. At 70 eV, molecules of a given type will always yield ions with the same energy distribution and therefore the same fragmentation patterns and fragment relative abundances. These are the spectra which are

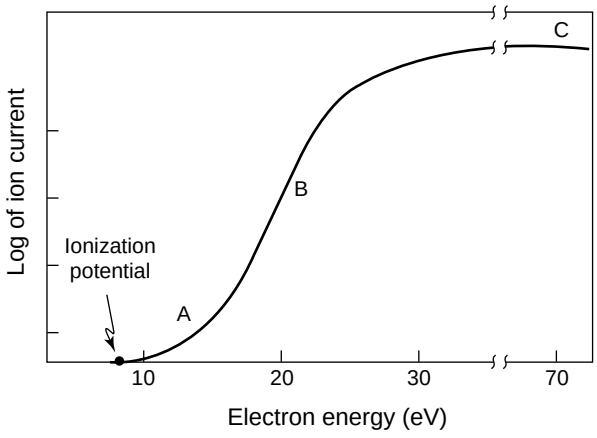

Figure 6 Appearance of molecular ions (region A) and fragment ions (regions B and C) with increasing ionization energy in EI sources. (Reprinted with permission from J.T. Watson, *Introduction to Mass Spectrometry*, 3rd edition, Lippincott-Raven, Philadelphia, 1997.)

considered 'standard' mass spectra and are available in a variety of commercial databases and search engines. PB is the only LC/MS interfacing technique which yields standard EI mass spectra. (See **Electron Ionization Mass Spectrometry**.)

CI sources have also been used with the PB interface, although no commercial products are currently available. CI will typically result in a far less complex spectrum than is seen with EI. Often these spectra contain only one ion, that being one which represents the intact molecule which provides information about the molecular weight of the molecule but no structural information. CI will be discussed in detail in section 3.2.1.

PB is ideal for those molecules which cannot survive the high-temperature environment of a gas chromatograph but are still volatile enough to go into the gas phase at moderately elevated temperatures. Because of the volatility requirement, PB is limited to analytes of molecular weight ≤ 1000 Da. Natural product molecules, steroids and carbamate pesticides are good examples. PB is also an excellent companion technique to photodiode array (PDA) detection. PDA data allow the analyst to determine the areas of maximum chromatographic peak purity and therefore the best portions of the chromatogram from which to obtain EI spectra. This allows for the collection of good, library-searchable spectra from less-than-ideal chromatography. Of course, a PDA detector also has this advantage for use with other LC/MS interfaces.

3.2 Atmospheric Pressure Ionization Interfaces

The most widely used LC/MS interfacing techniques are currently the API interfaces, namely the APCI interface and ESI. Figure 7 illustrates both of these variations of the API concept and serves to show that there are more similarities than differences in these two techniques. When these interfaces first appeared they were thought of as little more than 'molecular weight machines' because they produce only ions derived from the intact analyte molecule. The ionization process, unlike EI, does not impart sufficient energy to cause fragmentation and therefore no standard set of spectra can be generated. However, CAD does allow for fragmentation and these fragmentation patterns can be interpreted in qualitatively meaningful ways.

3.2.1 Atmospheric Pressure Chemical Ionization

The first LC/MS technique to gain widespread acceptance for routine analytical work was the APCI interface. APCI has extremely high sensitivity, is simple to use and is very rugged in high-throughput operation. When coupled with MS/MS, the technique provides excellent quantitative sensitivity and high specificity. Based on these performance characteristics, APCI is the method of choice for open access LC/MS systems. Open-access systems are specifically designed for analysts with little MS experience. These users need only set a couple of simple method parameters to process their samples and obtain usable mass spectral data.

CI is a reaction between a reagent ion and an analyte molecule in the gas phase. APCI uses the vaporized LC mobile phase (solvent) as the CI reagent. It is therefore not necessary in APCI to add a CI reagent from an external source as is required with GC/MS.

Referring to Figure 7, the APCI inlet is shown at the bottom left. The LC effluent flows in from the left side of the diagram through a capillary inlet needle. This needle is surrounded by a second needle carrying nebulizer gas resulting in a nebulized spray of LC effluent being introduced into the interface at atmospheric pressure. A heater at the outlet of the nebulizer assembly ensures vaporization of the LC solvent and analyte molecules. APCI interfaces are generally operated from 150 to as high as 400 °C depending on the volatility and thermal lability of the solvents and analytes. The temperature of the interface is an important method parameter since analyte volatility is essential to the APCI process. A potential difference of ~3–5 kV is set between the capillary nebulizer and the inlet to the mass spectrometer. The sign of this potential determines whether positive or negative ions will be drawn into the mass analyzer.

Just beyond the outlet of the capillary nebulizer assembly is a corona discharge needle. A high voltage applied to this needle leads to the formation of a corona of ionized atmospheric gas around the needle. Mobile phase vapor passing through this corona is ionized, e.g. $H_2O \longrightarrow H_3O^+$. Further down the path to the mass

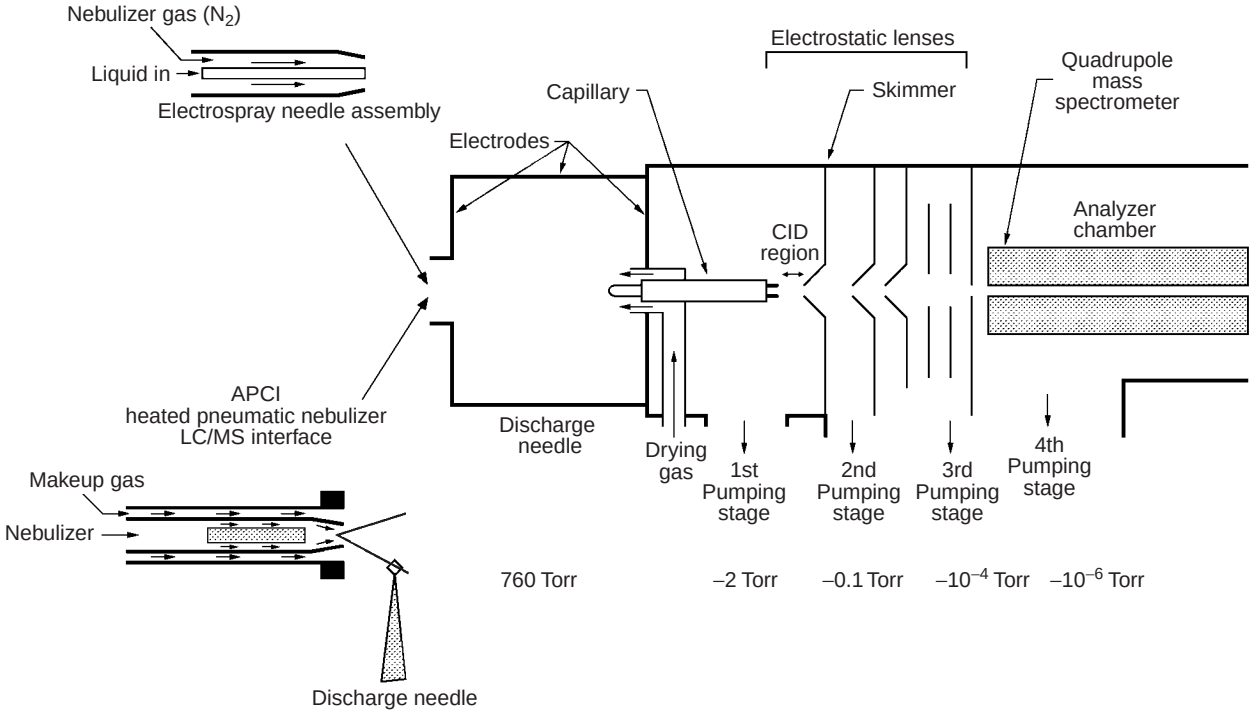

Figure 7 Schematic diagram of generic API interface showing both APCI configuration (bottom left) and electrospray (top left). The mass analyzer (far right) is shown here as a transmission quadrupole but may also be a quadrupole ion trap, time-of-flight (TOF) or magnetic analyzer. (Reprinted with permission from O.D. Sparkman and F.E. Klink, *Liquid Chromatography/Mass Spectrometry Fundamentals and Applications*, ACS Short Courses, American Chemical Society, Washington, DC, 1999. Copyright 1999 American Chemical Society.)

spectrometer inlet, gas-phase ion–molecule reactions take place which result in the formation of analyte ions. This process is also called solvent-mediated CI. For example, in the analysis of an amine RNH_2, the CI reaction may be as shown in Equation (1):

$$H_3O^+ + RNH_2 \longrightarrow H_2O + RNH_3^+ \qquad (1)$$

The ionized analyte is taken into the mass analyzer and the neutral water molecule is swept out through the vacuum system in the first pumped low-pressure region of the MS inlet.

The modes of ionization common in CI are represented in Scheme 1, in which M represents the analyte molecule.

The formation of even-electron ions by protonation, proton abstraction and adduct attachment is by far the most common APCI ion formation reaction scheme. It is important to note that in each case the resultant ion has a mass different from the analyte molecule. These ions are therefore not referred to as molecular ions but as ions which represent the intact molecule. Fragmentation of these ions is possible using CAD. CAD fragmentation may occur prior to entry into the mass analyzer or in a collision cell which is part of an MS/MS system. CAD will be discussed in more detail at the end of this section.

Protonation:

$$H_3O^+ + M \longrightarrow (M+H)^+ + H_2O$$

Proton abstraction:

$$HO^- + M \longrightarrow (M-H)^- + H_2O$$

Adduct attachment:

$$NH_4^+ + M \longrightarrow (M+NH_4)^+$$

$$CH_3COO^- + M \longrightarrow (M+CH_3COO)^-$$

Charge exchange:

$$C_6H_6^{+\bullet} + M \longrightarrow M^{+\bullet}$$

Electron capture:

$$e^- + M \longrightarrow M^{-\bullet}$$

Scheme 1

In order to obtain the best results from APCI, it is necessary to understand something of the mechanism of the gas-phase ion–molecule reactions. Figure 8(a) and (b) shows the common HPLC solvents and analyte functional groups and their relative ionization potentials.[27] Figure 8(a) shows positive ion operation. During the gas-phase ion–molecule reactions protons will, in effect, transfer downward in this chart, i.e. toward the formation of the weakest acid. Therefore, the molecule with the highest proton affinity will show the greatest ionization efficiency. If this molecule is a mobile phase component,

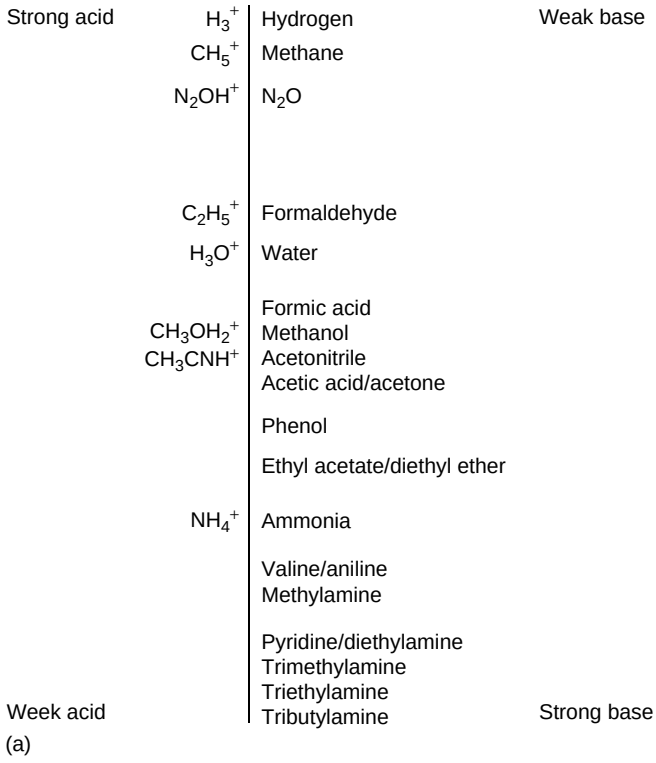

Figure 8 (a) Relative proton affinity of common HPLC reagents/analytes (right column) and their resultant cations (left column). (b) The corresponding chart to (a) but in this case for negative ion operation. (Adapted from Bruins.[27])

ionization of the analyte molecule will be completely suppressed owing to the much greater abundance of mobile phase compared with the analyte. On the other hand, if the molecule of higher proton affinity is a sample component, a better HPLC separation of this component from the analyte or a different sample preparation procedure will solve the problem. Figure 8(b) shows the corresponding chart for negative ion operation. In this case, the APCI reaction goes toward the weakest base.

APCI works well for most small molecule analytes, even those which contain no acidic or basic sites (e.g. alcohols, aldehydes, esters, ketones and hydrocarbons). The analyte must be volatile and proper setting of the capillary nebulizer heater temperature is important. It must be sufficient to volatilize the solvents and analytes but no so high as to decompose them thermally. APCI, along with PB, is essentially a small molecule interface. The upper limit of analyte mass in APCI is around 2000 Da. The lower limit is approximately 150–200 Da owing to low mass noise in the spectra. This low mass noise is a result of solvent ions taken into the mass analyzer along with the analyte.

APCI is extremely sensitive. The ionization efficiency is the best of all interfaces discussed in this article and can approach 100% of the analyte introduced into the interface, of which approximately 1% reaches the MS detector.[28] Routine sensitivities in the low-picogram range are observed and femtogram limits of detection are not uncommon.

3.2.2 Electrospray Interfacing

ESI is the second API technique, along with APCI. The hardware for ESI has much in common with APCI and, in fact, modern instrumentation is easily switched between these two ionization modes. In ESI, as in APCI, a capillary needle sprayer introduces sample into the atmospheric pressure interface (Figure 7). An electric potential is again applied across the capillary needle and the entrance to the low-pressure region of the mass analyzer. However, in ESI there is no corona discharge needle and temperatures are not as high as those used in APCI.

The first important consideration in understanding ESI is the nature of the ionization process. In all other interfacing techniques discussed in this article, ionization takes place *in the interface*. Hence all of these other techniques have the term 'ionization' in their name. Electrospray is different – ions must exist in solution for the electrospray process to work as a mass spectral interface.[29] Consequently, it is properly referred to as an electrospray *interface*, not electrospray ionization.

It is convenient to think of the ESI process as an electrochemical cell which operates in the gas phase.[30] The circuit is composed of the capillary needle and the

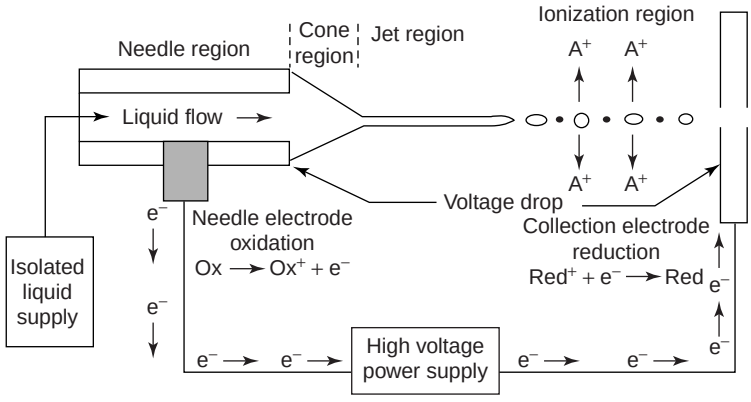

Figure 9 Schematic of an ESI operating in positive ion mode. (Reprinted with permission from Willoughby et al.[28] with permission of Global View Publishing.)

inlet to the mass spectrometer (across which the potential is applied) and is completed by the ionic LC effluent solution. Figure 9 shows an ESI producing positive ions. Negative counterions are depleted from solution in the oxidation process at the positively charged capillary sprayer. LC effluent droplets, now with an excess of positive ions, draw out of the tip of the capillary in a characteristic 'Taylor cone' formation and are pulled in a thin jet toward the negatively biased MS inlet. There are a number of competing theories about exactly how ions are evaporated from the surface of the liquid,[18,30–32] but each of these theories shares some common elements. It appears that as a droplet moves toward the MS inlet it breaks down into a series of successively smaller droplets as the neutral solvent molecules are evaporated. This process is aided by the charge repulsion of the like-charged ions clustered on the surface of the droplet. The entire process is called ion evaporation. Eventually, only partially solvated ions remain to be taken into the mass analyzer where the remaining vaporized solvent is swept out through the vacuum system of the instrument.

Since very low energy is applied in ESI, no fragmentation takes place. ESI typically produces even-electron protonated (+), deprotonated (−) or adducted (+ or −) species representing the intact analyte molecule (see section 3.2.1 for an explanation of these ion types). Formation of these ions is facilitated by control of the pH of the solution as it is introduced into the interface.[28,29] The chromatographic conditions, however, may preclude adjustment of the mobile phase pH to the required value. In this case, a postcolumn introduction of acid or base can be used to adjust pH prior to the ESI inlet.[33] Ion suppression resulting from acid–base ion–molecule reactions in the gas phase which were noted for APCI (Figures 8a and b) are also of concern in ESI. A discussion of interferences and suppression effects is included at the end of this article.

Electrospray has two key characteristics which make it ideal for the analysis of biological macromolecules: (1) nonvolatile analytes can be passed into the mass spectrometer using ESI and (2) multiply charged ions can be produced by ESI. These characteristics, taken together, allow very large biomolecules to be introduced to the mass spectrometer and permit them to be seen at reasonable m/z values well within the range of the mass analyzer. For example, with ESI it is possible to analyze a protein of 10 000 Da using a mass analyzer which has an upper m/z limit of 3000. The electrospray ion evaporation process facilitates putting this large, ionized molecule into the mass spectrometer. Second, if the molecule is capable of carrying, let us say, 10 charges, then the peak representing the ion will show up at $m/z = (10\,000\ \text{Da}/10\ \text{charges}) = 1000$.

Usually, however, biological macromolecules do not appear in the spectrum as a single peak but rather a cluster of peaks representing ions of different charge states. These clusters are characterized by a one-charge difference between adjacent peaks. The analyst can 'tune' the number of charges and the spectral positioning (in terms of m/z) of these peak clusters by altering the pH of the analyte solution. Deconvolution algorithms are used to interpret these multiple peak clusters and present the results on a mass (as opposed to an m/z) scale. The result is an artificial spectrum of the macromolecule as a singly charged ion. Because the deconvolution is based on many multiply charged ion peaks, a very accurate mass for the singly charged ion and, consequently, for the molecule can be determined.[29] Figure 10(a–c) shows a typical multiply charged ion cluster and the results of two different deconvolution approaches. Spectrum (a) represents the data obtained from the instrument. Each peak represents an ion of the intact rho protein (47 005 Da) and adjacent peaks differ by one charge. These ions carry charges in the range +70

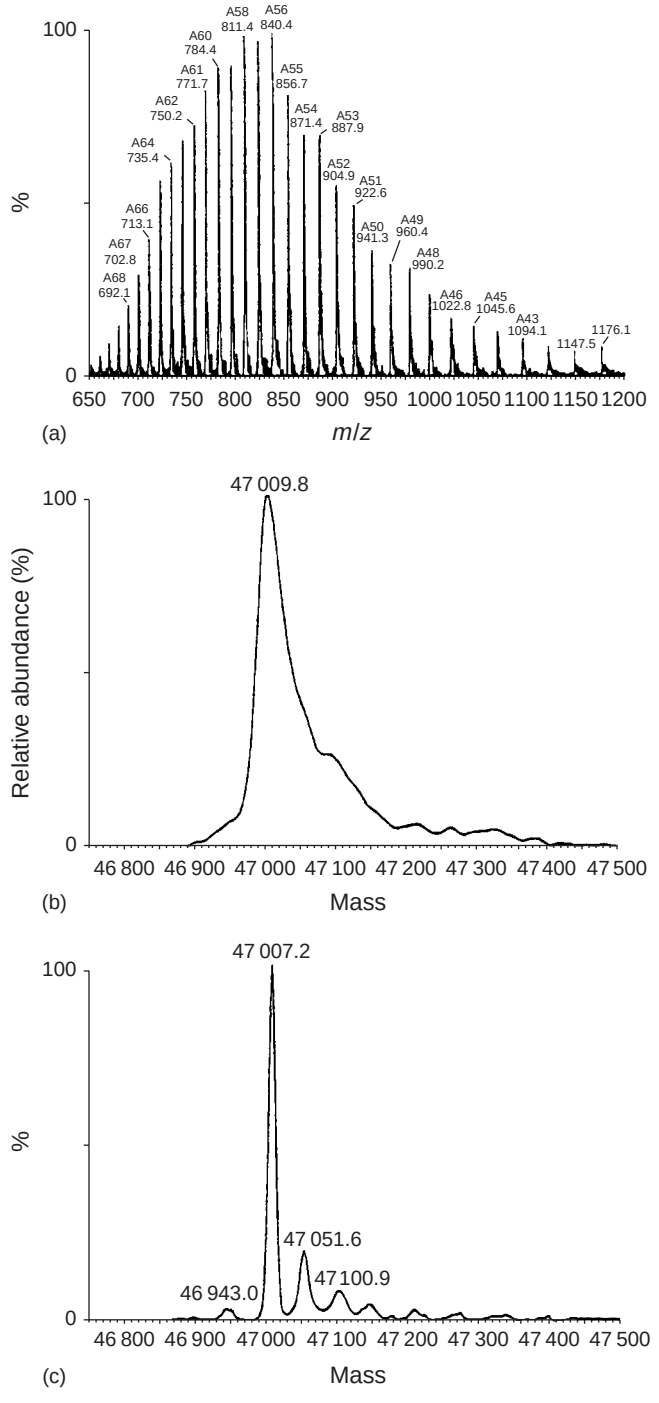

Figure 10 (a) ESI spectrum of Rho protein (47 005 Da) showing a cluster of multiple charge ions in the range of +40 to +70. (b) The same spectrum after deconvolution to mass scale. (c) The spectrum after maximum entropy processing. (Reprinted with permission from Gaskil.[29])

to +40. Spectrum (b) results from a simple deconvolution algorithm and spectrum (c) was generated using a more sophisticated 'maximum entropy' method which allows resolution of isotope peaks.[34]

Electrospray as used in LC/MS can be characterized as a concentration detector, meaning that the response is directly proportional to the concentration of analyte in the solution sprayed into the interface. The linear range of ESI is from 10^{-9} to 10^{-5} M. From 10^{-5} to 10^{-4} M the signal increases with approximately the square root of concentration and above 10^{-4} M the signal begins to decrease with increasing concentration.[30] Most detectors in HPLC are mass detectors, in which the response is proportional to the mass of analyte introduced. For mass detectors, a larger injection volume creates more signal. With an ESI concentration detector, an increased injection volume will have no effect on response. This is important information for method development where either concentration or dilution may be necessary to optimize the signal for a particular analyte.

The original descriptions of electrospray for MS[18] and for LC/MS[19] showed the technique to be limited to flow rates of the order of 1–10 μL min^{-1}. This technique used a simple capillary sprayer as illustrated in Figure 11(a). These low flow rates were not practical for routine chromatography until very recently. Ionspray[20] was developed in 1984 to overcome the low flow rate limitation. This modification is illustrated in Figure 11(b) and consists of a second needle concentric with the inlet capillary. A gas flow is introduced through this second needle and acts to nebulize the incoming liquid

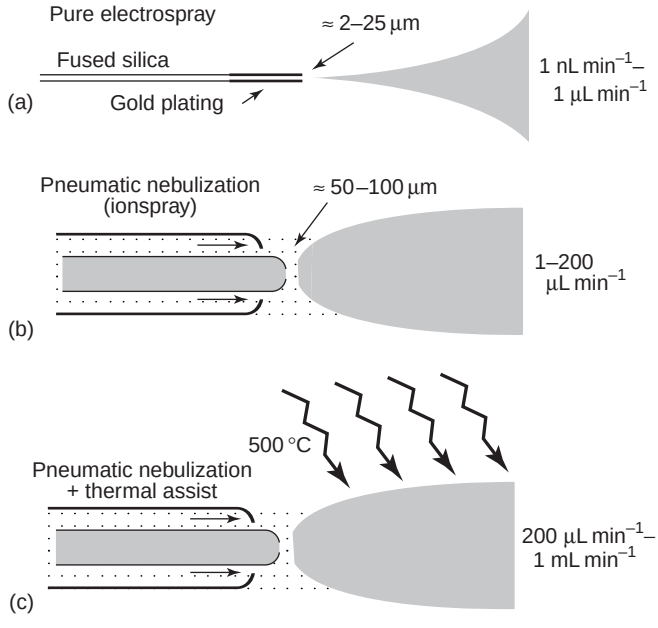

Figure 11 Various modes of operation of ESIs. (Reprinted with permission from O.D. Sparkman and F.E. Klink, *Liquid Chromatography/Mass Spectrometry Fundamentals and Applications*, ACS Short Courses, American Chemical Society, Washington, DC, 1999. Copyright 1999 American Chemical Society.)

flow. A further refinement of this approach was to add a heater (Figure 11c), resulting in a technique known as turbospray. Both turbospray and ionspray allow the use of ESI with typical analytical HPLC flow rates of up to $1\,mL\,min^{-1}$. Variations of these two nebulizer-assisted techniques are used in all current commercial ESI products.

With increasing interest in micro-flow rate HPLC techniques and the increased availability of HPLC pumps and accessories for micro-techniques, a corresponding surge in interest in low-flow rate ESI is occurring. Beginning in 1995,[35] Andrén et al. demonstrated use of a minaturized, nebulizer-assisted electrospray coupled with a micro-HPLC system operating at flow rates of $300-800\,nL\,min^{-1}$. This device, called micro-electrospray, is not a return to the original ESI design but a miniaturization of the ionspray concept. These researchers demonstrated optimum ESI sensitivity in the zeptomole and attomole range, improved performance with high-aqueous solvent systems and reduced contamination.

In parallel with Andrén et al.'s work, Wilm and Mann[36] developed the concept of nano-electrospray or, simply, nanospray. This interface uses backing gas pressure in place of a mechanical pump to drive solvent into the interface. Very small ID capillaries ($1-2\,\mu m$) are used and operation is carried out at very low flow rates, of the order of $20\,nL\,min^{-1}$. No nebulization is used and therefore this method can be thought of as a 'pure' electrospray method. Nanospray was originally developed as a continuous infusion method for protein analysis. By infusing sample continuously over a period of many hours, thousands of spectra are collected and averaged. This improves the signal-to-noise ratio and allows for very accurate m/z determination and high spectral resolution. For unknown protein samples, this results in excellent sensitivity and remarkably accurate exact mass determinations while consuming very little sample. More recently, this technique has been expanded to use nano-HPLC columns to achieve some degree of initial separation, although not with high chromatographic resolution.[37,38] Nanospray routinely exhibits femtomole and better sensitivities.

Commercial products for both micro- and nano-ESI are becoming widely available at the time of this writing.

The use of micro-diameter columns results in very narrow chromatographic peak widths and this in itself presents data acquisition problems for the mass spectrometer. Typical benchtop LC/MS systems employing transmission quadrupoles or quadrupole ion traps are limited to spectral data acquisition rates of 1–10 spectra per second. This is inadequate to obtain a sufficient number of spectral samples across the very narrow peaks typical of micro-HPLC. Consequently, commercial micro- and nano-ESI products are linked with time-of-flight (TOF) mass spectrometers. TOF instruments have spectral data acquisition rates of 100 spectra per second or higher. In addition, modern reflectron TOF instruments exhibit very high spectral resolution, excellent m/z accuracy and virtually unlimited m/z range.[39] (See **Time-of-flight Mass Spectrometry**.)

Electrospray (in its several forms) is ideally suited to analytes which can be readily ionized in solution, are marginally volatile or involatile and/or would thermally decompose at the higher operating temperatures of the other LC/MS interface types. Ion suppression effects (discussed below) must be understood and appreciated or poor performance could result. Electrospray generally works well with aqueous solvent systems and, since there is no concern about involatile analytes, this interface is the ideal companion to reversed-phase or aqueous size-exclusion chromatography. The sensitivities for modern electrospray systems approach that for APCI and ESI can be considered nearly on a par with APCI for ruggedness and ease-of-use. The most rapidly evolving area for electrospray is in biological macromolecule analysis (see **Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis; High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis** and **Mass Spectrometry of Nucleic Acids**).

4 COLLISIONALLY ACTIVATED DISSOCIATION IN LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY

The CI process in APCI and the ion evaporation process in ESI produce only ions of the intact analyte molecule. If this were the only capability of these two techniques, they would still be very useful in the determination of exact molecular weights and in the separation of analytes of different mass which could not be separated by the chromatographic process. However, through the process of CAD, it is possible to produce fragment ions and to apply mass spectral interpretation techniques to the fragment ion spectra to yield qualitative information about the analyte molecule. Collision-induced dissociation (CID) is a widely used synonym for CAD.

The CAD fragmentation process takes place after ions have been formed. This process is the basis for fragmentation in all forms of MS/MS regardless of the nature of the front-end interface to the mass analyzer (see **Tandem Mass Spectrometry: Fundamentals and Instrumentation**). In API interfaces another CAD option is available which is called 'up-front' or 'in-source' CAD.

In this instance CAD fragmentation takes place before the ions enter the mass analyzer ion optics. This region is labeled 'CID region' in Figure 7.

Note that this region is just before the skimmer cone, the first element of the ion optical system. Ions traversing this region of relatively high pressure (at least 10^{-4} Torr) are likely to collide with neutral gas molecules. If the collisions are sufficiently energetic, this increases the internal energy of the ions. To shed this excess energy, these ions will fragment. The velocity with which ions traverse the CAD region is controlled by the voltage applied to the skimmer. The voltage at which fragment ions begin to appear varies with analyte compound and the API system design.

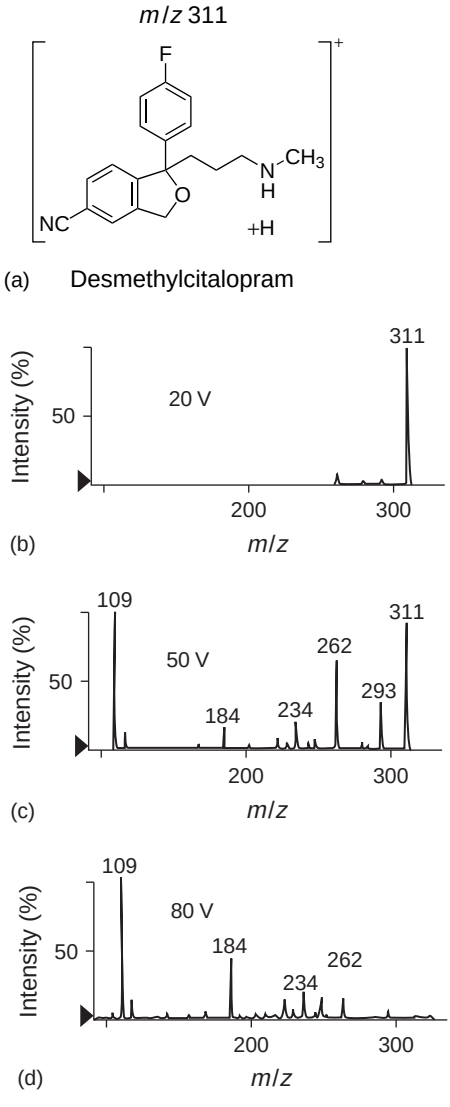

Figure 12 Illustration of up-front CAD of (a) desmethylcitalopram demonstrating the effect of increasing the voltage applied to the skimmer cone from (b) 20 to (c) 50 and (d) 80 V with ESI. (Reprinted with permission from Weinmann et al.[42])

Figure 12(a–d) illustrates up-front CAD fragmentation of the metabolite desmethylcitalopram at various voltages applied to the skimmer cone (see Figure 7). Referring to Figure 12(b), at 20 V there is no fragmentation evident. The single peak at m/z 311 is the M + H ion as shown in Figure 12(a). At 50 V (Figure 12c) there is now considerable production of fragment ions, as seen by the appearance of peaks at lower m/z values. If we were to examine actual ion abundances, not the relative intensity shown here, we would find that the absolute amount of the m/z 311 ion has dropped as the fragment ion abundances increased. Finally, at 80 V (Figure 12d) there is no M + H present and even the larger fragment ions have dropped in abundance relative to the m/z 109 fragment ion. This accelerating voltage is enough to decompose all desmethylcitalopram ions into very low-mass fragments.

The CAD processes in MS/MS and in up-front CAD are very similar. In up-front CAD, the gas molecules are largely nitrogen from the atmosphere or from the sweep gas used in some interfaces to aid solvent evaporation. In MS/MS, gases such as argon or helium are introduced to the collision cell to function as CAD gases. Up-front CAD and MS/MS can also produce very similar spectra. With MS/MS the analyst has the advantage of selecting only certain ions for fragmentation. Therefore, MS/MS can generate usable spectra even if the chromatographic separation has not resulted in a pure compound. Up-front CAD fragments all ions passing into the analyzer, which means that interpretable spectra can only be obtained if pure compounds are introduced to the API interface.

As described earlier, both APCI and ESI produce primarily even-electron ions, e.g. $[M + H]^+$. EI spectra, such as those from GC/MS and PB interfaces, produce odd-electron species, $M^{+\bullet}$ and $M^{-\bullet}$. Classical MS interpretation rules are based on the radical site-initiated fragmentation of these odd-electron ions. Even-electron ion fragmentation can therefore result in different fragments being produced than are seen for the same compound in EI. The 'even-electron rule' states that the neutral loss from an even-electron ion will tend to be the fragment with the lowest proton affinity and, as a consequence, the charge will be retained on the fragment with the highest proton affinity.[28] Rearrangements and multiple-bond cleavages are common in CAD spectra. (See **Electron Ionization Mass Spectrometry** for EI spectral interpretation and **Atmospheric Pressure Ionization Mass Spectrometry** for API spectral interpretation.)

Interpretation of CAD spectra can be further complicated in electrospray by the production of multiply charged ions. Fragments of multiply charged ions will carry fewer charges than the intact ion and therefore may appear at similar or even higher m/z values than the ion from which they originated. Data system software which not only deconvolutes multiply charged ion spectra but

also identifies fragment ions in these spectra is a necessity for biospectrometry with ESI/MS.

A unique use of both up-front CAD and MS/MS with ESI has been the work on sequencing of proteins and nucleic acids. Both of these biomolecules fragment in CAD in ways that can be interpreted to give at least a partial sequence of their component amino acids or bases. Specific software products are available to support CAD sequencing of biomolecules. (See **Mass Spectrometry in Structural Biology; Mass Spectrometry of Nucleic Acids** and **High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis**.)

EI spectra have long been used for exact qualitative analysis through the comparison of an unknown spectrum with libraries of spectra of identified compounds. Such libraries are available for drugs, metabolites, toxins, pesticides, environmental pollutants and their common GC derivatives. Because of variations in the HPLC environment from one analytical method to another and design variations from one interface manufacturer to another, it is not possible to create similar 'standard' libraries for CAD spectra using LC/MS and API interfaces. However, within the confines of a defined LC/MS method, users have been successful in creating libraries of CAD spectra and using these in routine identification of unknown spectra using both up-front CAD[40] and MS/MS.[41] One recent report described the creation of a 400-compound library of drugs and the use of this library to evaluate patient samples.[42] In all of these papers libraries were devised on the basis of two or more spectra from each compound taken at different applied voltages as shown in Figure 12(b–d). Excellent reproducibility of CAD spectra from one instrument to another instrument of the same model and manufacturer was also reported.

5 SAMPLE AND SOLVENT CONSIDERATIONS IN METHOD DEVELOPMENT

Components of the sample matrix or the chromatographic mobile phase may have an adverse effect on the results from an LC/MS analysis using API interfaces. Analysts must understand these potential adverse effects and the various means of controlling them. Although there is considerable overlap, it is convenient for this introductory discussion to place these effects into four categories: spectral interference, system compromise, adduct formation and ion suppression.

5.1 Spectral Interference

Ions which appear at the same or nearly the same m/z value as the component of interest will interfere with both qualitative and quantitative use of the spectrum. Ions of similar m/z value can generally be separated by the chromatographic process prior to entry into the mass spectrometer. Where this is not possible, a clear distinction can often be made between the components through use of selected reaction monitoring (SRM) with an MS/MS system. Finally, newer LC/MS systems employing TOF mass analyzers are generally able to distinguish much better than unit mass resolution and can therefore 'pull apart' spectral peaks which may show considerable overlap in lower resolution systems.

5.2 System Compromise

Many 'classical' mobile phase components used in HPLC may degrade the performance of the LC/MS system. Chromatographers approaching LC/MS for the first time must begin to switch their thinking from concepts related primarily to UV detectors and toward mass spectral considerations. Components of the mobile phase which are likely to precipitate in the API interface are generally buffer salts or ion-pair reagent salts. Phosphate buffers became the pH control system of choice in HPLC because they are usable down into the very low UV detection wavelengths. Obviously, this is not an issue for an MS system. In some cases, substitution with volatile buffers, e.g. ammonium acetate or formate, will be a reasonable approach. For very high sensitivity analysis in electrospray, both of these buffers have been shown to contribute to signal suppression.[43]

However, in most HPLC separations, buffering capacity is not required at all. Buffer salts may be eliminated in favor of the use of organic acids or ammonia solution to control pH.[44]

5.3 Adduct Formation

Adduction of another ion with the component of interest, shifting the m/z value at which it appears in the spectrum is common with API interfaces. Adduct ions such as sodium, potassium and ammonium may be readily picked up from the sample itself, from reagents used, or the container in which the sample is held (Na^+ from laboratory glassware is the most common example). Adduct formation in MS has often been used to improve sensitivity, especially for macromolecules. However, uncontrolled adduct formation is generally undesirable and requires specific sample preparation procedures to reduce or eliminate it. SPE methods on reversed-phase and ion-exchange sorbents are widely used[45] for the removal of salts and buffers which contribute to ion suppression effects and adduct formation. Naturally, many SPE and chromatographic procedures have the potential for integrated, automated operation with the LC/MS instrument.

5.4 Ion Suppression

Components of the sample or chromatographic mobile phase may suppress ionization of, or compete in the ionization process with, the analyte of interest. Ion suppression is the most critical of the four types of interference because it is often the most difficult to determine. For both qualitative and quantitative LC/MS work with API interfaces, understanding and eliminating ion suppression effects are essential.

The formation of ions in API interfaces through gas-phase ion–molecule reactions was described earlier. One of the more common types of ion suppression results from the presence of a large excess of a compound with greater proton affinity (+ ion operation) than the analyte of interest. In the vapor phase of the API interface this compound will be preferentially ionized. Figure 8(a–b) should be referred to for an understanding of these processes for both positive and negative ion operation.

Ion suppression effects due to mobile phase components will generally be easy to diagnose and eliminate during the methods development process. Use of the previously described flow injection technique (Figure 4b) allows the analyst a wide playing field for experimenting with the effects of various mobile phase components. Both standards and actual samples should be run during these experiments.

Components of the sample matrix, some of which do not show up in the mass spectrum in the m/z range being investigated, can also cause ion suppression. In biological samples, natural variations in endogenous compound concentrations from one sample to another lead to varying levels of ion suppression.[46] This, in turn, contributes to unacceptable variability in the signal response for the compounds of interest. The solution for these cases is found in better chromatographic separation of analytes from interferences[46] or improved sample preparation procedures.[47,48]

Another type of ion suppression is thought to occur when very strong ion pairs are formed which are not broken apart by the conditions in the API interface. The simplest solution is to eliminate the ion-pairing agent. This may shorten retention times, cause peak overlap and/or degrade peak shape but these issues are often not critical with LC/MS since usable spectral information may still be obtained from 'poor' chromatography. Ion-pairing agents which often cannot be eliminated are trifluoroacetic acid (TFA) and heptafluorobutyric acid (HFBA) in peptide/protein and other complex molecule analysis. TFA and HFBA are thought to improve peak shapes in reversed-phase chromatography by ion pairing with basic sites on the molecule, thereby eliminating mixed retention mechanisms. However, TFA ion pairs are not broken up in the mild conditions of ESI and this prevents ion evaporation of the unpaired analyte ion. The problem can be reduced by limiting the concentration of TFA to <0.1% or using 'TFA light', which is a reduced amount of TFA, perhaps 0.05%, and the same concentration of acetic acid. This serves to control pH and promote ionization, as required for electrospray, without forming an abundance of ion pairs.[25]

ABBREVIATIONS AND ACRONYMS

APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
CAD	Collisionally Activated Dissociation
CF/FAB	Continuous-flow Fast Atom Bombardment
CI	Chemical Ionization
CID	Collision-induced Dissociation
EI	Electron Ionization
ESI	Electrospray Interface
FAB	Fast Atom Bombardment
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HFBA	Heptafluorobutyric Acid
HPLC	High-performance Liquid Chromatography
ID	Inner Diameter
IR	Infrared
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
MAGIC	Monodisperse Aerosol Generation Interface for Chromatography
MALDI	Matrix-assisted Laser Desorption/ Ionization
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
PB	Particle Beam
PDA	Photodiode Array
RTICC	Reconstructed Total Ion Current Chromatogram
SIM	Selected Ion Monitoring
SPE	Solid-phase Extraction
SRM	Selected Reaction Monitoring
STP	Standard Temperature and Pressure
TFA	Trifluoroacetic Acid
TOF	Time-of-flight
UV	Ultraviolet

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological

Macromolecules • Mass Spectrometry in Structural Biology

Environment: Water and Waste **(Volume 3)**
Capillary Electrophoresis Coupled to Inductively Coupled Plasma-Mass Spectrometry for Elemental Speciation Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Food **(Volume 5)**
Liquid Chromatography in Food Analysis

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids • Sequencing Strategies and Tactics in DNA and RNA Analysis

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • High-performance Liquid Chromatographic Separations and Equipment in Peptide and Protein Analysis, Miniaturization of • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Reversed-phase High-performance Liquid Chromatography in Peptide and Protein Analysis

Pesticides **(Volume 7)**
High-performance Liquid Chromatography Methods in Pesticide Residue Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Gas and Liquid Chromatography, Column Selection for, in Drug Analysis • Mass Spectrometry in Pharmaceutical Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Biopolymer Chromatography • Capillary Electrophoresis • Microscale High-performance Liquid Chromatography and the Evolution of Capillary Electrochromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

REFERENCES

1. L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, 2nd edition, John Wiley and Sons, New York, 1979.
2. G. Siuzdak, B. Bothner, M. Yeager, C. Brugidou, C. Fauquet, K. Hoey, C. Chang, 'Mass Spectrometry and Viral Analysis', *Chem. Biol.*, **3**, 45–48 (1996).
3. P.J. Arpino, 'On-line Liquid Chromatography/Mass Spectrometry? An Odd Couple', *Trends Anal. Chem.*, **1**, 154–158 (1982).
4. V.L. Talroze, V.E. Skurat, I.G. Gorodetski, N.B. Zolotai, *Russ. J. Phys. Chem.*, **46**, 456 (1972).
5. P.J. Arpino, M.A. Baldwin, F.W. McLafferty, 'Liquid Chromatography/Mass Spectrometry Continuous Monitoring', *Biomed. Mass Spectrom.*, **1**, 80–82 (1974).
6. E.C. Horning, D.I. Carroll, I. Dzidic, K.D. Haegele, M.G. Horning, R.N. Stillwell, 'Atmospheric Pressure Ionization (API) Mass Spectrometry; Solvent-mediated Ionization of Samples Introduced in Solution and a Liquid Chromatograph Effluent Stream', *J. Chromatogr. Sci.*, **12**, 725–729 (1974).
7. W.H. McFadden, H.L. Schwartz, S. Evans, 'Direct Analysis of Liquid Chromatographic Effluents', *J. Chromatogr.*, **122**, 389–396 (1976).
8. C.R. Blakely, M.L. Vestal, 'Thermospray Interface for Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, **55**, 750–754 (1983).
9. M.L. Vestal, 'High Performance Liquid Chromatography–Mass Spectrometry', *Science*, **226**, 275–281 (1984).
10. R.C. Willoughby, R.F. Browner, 'Monodisperse Aerosol Generation Interface for Combined Liquid Chromatography with Mass Spectrometry', *Anal. Chem.*, **56**, 2626–2631 (1984).
11. M. Barber, R.S. Bordoli, R.D. Sedgewick, A.N. Tyler, 'F.A.B.', *J. Chem. Soc., Chem. Commun.*, 325–327 (1981).
12. Y. Ito, T. Takeuchi, D. Ishii, M. Goto, 'Direct Coupling of Micro High-performance Liquid Chromatography with Fast Atom Bombardment Mass Spectrometry', *J. Chromatogr.*, **346**, 166–181 (1985).
13. R.M. Caprioli, T. Fan, J.S. Cottrell, 'Continuous-flow Sample Probe for Fast Atom Bombardment Mass Spectrometry', *Anal. Chem.*, **58**(14), 2949–2953 (1986).
14. J. Chapman, *Protein and Peptide Analysis by Mass Spectrometry*, Humana Press, Totowa, NJ, 1996.
15. F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser Desorption/Ionization of Biopolymers', *Anal. Chem.*, **63**, 1191A–1203A (1991).
16. J. Preisler, F. Foret, B.L. Karger, 'On-line MALDI–TOF MS Using a Continuous Vacuum Deposition Interface', *Anal. Chem.*, **70**(24), 5278–5287 (1998).

17. J. Zeleny, 'Instability of Electrified Liquid Surfaces', *Phys. Rev.*, **10**, 1 (1917).
18. M. Dole, L.L. Mach, R.L. Hines, R.C. Mobley, L.P. Ferguson, M.B. Alice, 'Molecular Beams of Macroions', *J. Chem. Phys.*, **42**, 2240 (1968).
19. C.M. Whitehouse, J.B. Fenn, M. Yamashita, 'An Electrospray Ion Source for Mass Spectrometry of Fragile Organic Species', in *Proceedings of the 32nd Annual Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, May 27–June 1, 1984*, 188–189, American Society for Mass Spectrometry (ASMS), Sante Fe, New Mexico, 1984.
20. A.P. Bruins, T.R. Covey, J.D. Henion, 'Ionspray Interface for Combined Liquid Chromatography/Atmospheric Pressure Ionization Mass Spectrometry', *Anal. Chem.*, **59**, 2642–2646 (1987).
21. M. Tsvet, 'A New Category of Adsorption Phenome and Its Application to Biochemical Analysis', *Trav. Soc. Nat. Varisovic*, **14**, (1903).
22. A.J.P. Martin, R.L.M. Synge, 'A New Form of Chromatogram Employing Two Liquid Phases', *Biochem. J.*, **35**, 91 and 1358 (1941).
23. R.E. Majors, 'High Performance Liquid Chromatography on Small Particle Silica Gel', *Anal. Chem.*, **44**, 1722–1726 (1972).
24. R.E. Majors, 'Current Trends in HPLC Column Usage', *LC/GC*, **15**(11), 1008–1015 (1997).
25. F. Klink, 'New Sample Preparation Approaches to Biological Matrices for LC/MS', *LC/GC*, **17**(12), 1084–1093 (1999).
26. A. Capiello, G. Famiglini, F. Bruner, 'Determination of Acidic and Base/Neutral Pesticides in Water with a New Microliter Flow Rate LC/MS Particle Beam Interface', *Anal. Chem.*, **66**, 1416–1423 (1994).
27. A.P. Bruins, 'Mass Spectrometry with Ion Sources Operating at Atmospheric Pressure', *Mass Spectrom. Rev.*, **10**(1), 53–77 (1991).
28. R. Willoughby, E. Sheehan, S. Mitrovich, *A Global View of LC/MS*, Global View Publishing, Pittsburgh, PA, 1998.
29. S.J. Gaskil, 'Electrospray: Principles and Practice', *J. Mass Spectrom.*, **32**, 677–688 (1997).
30. M.G. Ikonomou, A.T. Blades, P. Kebarle, 'Electrospray–Ion Spray: A Comparison of Mechanisms and Performance', *Anal. Chem.*, **63**, 1989–1998 (1991).
31. J.V. Iribarne, B.A. Thomson, 'Evaporation of Small Ions from Charged Droplets', *J. Chem. Phys.*, **64**, 2287–2294 (1976).
32. J.B. Fenn, J. Rosell, C.K. Meng, 'In Electrospray Ionization, How Much Pull Does an Ion Need to Escape its Droplet Prison?', *J. Am. Soc. Mass Spectrom.*, **8**(11), 1147–1157 (1997).
33. R. Voyksner, 'Short Course on Liquid Chromatography–Mass Spectrometry: Art and Practice', presented at the 44th ASMS Conference on Mass Spectrometry and Allied Topics, Portland, OR, May 11–12, 1996.
34. A.G. Ferrige, M.J. Seddon, S. Jarvis, *Rapid Commun. Mass Spectrom.*, **5**, 374 (1991).
35. P.E. Andrén, M.R. Emmett, R.M. Capriolli, 'Micro Electrospray: Zeptomole and Attomole per Microliter Sensitivity for Peptides', *J. Mass Spectrom.*, **30**, 817 (1995).
36. M. Wilm, M. Mann, 'Analytical Properties of the Nanoelectrospray Ion Source', *Anal. Chem.*, **68**, 1–8 (1996).
37. H.M. Svendsen, J. Andersen, T. Covey, O. Vorm, 'New Approaches to On-line Nanoelectrospray and Nano-scale LC–MS', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, May 31–June 1, 1998*, ASMS, Santa Fe, New Mexico, 1998.
38. R.K. Blackburn, R.J. Anderegg, 'Characterization of Femtomole Levels of Proteins in Solution Using Rapid Proteolysis and Nanoelectrospray Ionization Mass Spectrometry', *J. Am. Soc. Mass. Spectrom.*, **8**, 483–494 (1997).
39. C. Henry, 'Electrospray Continues to Evolve', *Anal. Chem.*, **69**, 427A–432A (1997).
40. J. Keever, R.D. Voyksner, J. Hough, 'Development and Evaluation of Electrospray CID Mass Spectral Searching for Compound Identification', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, May 31–June 4, 1998*, 86, 1998, ASMS, Santa Fe, New Mexico, 1998.
41. J.L. Jospehs, M. Sanders, M. Bolgar, Y. Gervacio, G. DiDonato, M. Hail, 'The Role of LC/MS/MS Spectral Library Searching for Impurity Analysis Using Ion Trap and Triple Quadrupole Mass Spectrometers', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, May 31–June 4, 1998*, 579, 1998, ASMS, Santa Fe, New Mexico, 1998.
42. W. Weinmann, A. Wiedemann, B. Eppinger, M. Renz, 'Screening for Drugs in Serum by Electrospray Ionization/Collision-induced Dissociation and Library Searching', *J. Am. Soc. Mass Spectrom.*, **10**, 1028–1037 (1999).
43. D. Temesi, B. Law, 'The Effect of LC Eluent Composition on MS Responses Using Electrospray Ionization', *LC/GC*, **17**(7), 626–632 (1999).
44. O.D. Sparkman, F.E. Klink, *Liquid Chromatography/Mass Spectrometry Fundamentals and Applications*, American Chemical Society Short Courses, American Chemical Society, Washington, DC, 1999.
45. E.M. Thurman, M.S. Mills, *Solid Phase Extraction Principles and Practice*, John Wiley and Sons, New York, 1998.
46. B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, 'Matrix Effect in Quantitative LC/MS/MS Analyses of Biological Fluids: A Method for Determination of Finasteride in Human Plasma at Picogram Per Milliliter Concentrations', *Anal. Chem.*, **70**, 882–889 (1998).
47. R. Wieboldt, V.S. Ong, T.A. El-Shourbagy, J.M. McVey, L.B. Sapochak, 'A Comparison of Three Extraction Methods for Determination of HIV Protease Inhibitors by LC/MS', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics, Dallas, TX,*

June 13–17, 1999, ASMS, Santa Fe, New Mexico, 1999.

48. G. Hopfgartner, C. Husser, M. Zell, A. Maschka-Selig, B. Hess, G. Fischer, B. Lausecker, 'Strategies for Sample Preparation to Support Multi-component Analysis in Bioanalytics', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics, Dallas, TX, June 13–17, 1999*, ASMS, Santa Fe, New Mexico, 1999.

Literature of Mass Spectrometry

Stephen Down
HD Science Limited, Nottingham, UK

1 Introduction **1**
2 Primary Journals **1**
2.1 Printed Form 1
2.2 Internet Form 5
3 Secondary Publications **6**
3.1 Mass Spectrometry Publications 6
3.2 Other Publications 9
4 Miscellaneous Literature Sources **11**
5 Internet Resources **11**
6 Books **12**
6.1 Mass Spectrometry Data Compilations 13
6.2 Basic Mass Spectrometry and Teaching Texts 14
6.3 Applications of Mass Spectrometry 18
Acknowledgments **24**
Abbreviations and Acronyms **24**
Related Articles **25**
References **26**

The mass spectrometry (MS) literature is one of the major sources available for keeping up to date with new developments in the technique. There are different ways of reporting and presenting MS information, such as journals, abstracts and books, and several different means of distribution, including printed, diskette, CD-ROM and the Internet.

1 INTRODUCTION

MS has grown to be one of the most important and powerful analytical techniques available today and the value of the literature in the advancement of the technique cannot be underestimated. Although many major developments in MS have been reported at conferences and symposia, most are recorded for posterity in the printed literature. This article examines the various currently available literature sources and their distribution media, including primary journals, secondary (abstracts) journals and books, published in printed, diskette, CD-ROM and Internet formats. Other Internet-based literature resources are also discussed. A compilation of MS books with brief descriptions of their content is presented.

2 PRIMARY JOURNALS

A primary scientific journal is defined as one that publishes original research work, whereas a secondary journal contains original information from one or more primary journals that may be in the form of abstracts or other summaries. The number of primary journals related to chemistry is increasing regularly, with new journals appearing each year and very few journals being discontinued. This rise, combined with the increase in the numbers of papers published within the journals, is illustrated clearly in the average number of papers (excluding patents and books) abstracted by Chemical Abstracts over comparative 4-year periods. In 1983–86, the average annual number of abstracted papers was 379 078. In the period 1995–98 this average was 571 686.[1] For journals specializing in MS, the same pattern is visible, and these primary MS journals are a major resource for the mass spectrometrist, both in keeping pace with new developments and applications and, to a lesser extent, for retrospective research.

2.1 Printed Form

Since the first journal devoted to MS was published (*Journal of the Mass Spectrometry Society of Japan*, JMSSJ, Tokyo, Japan, 1953) several new MS primary journals have been launched, most recently *European Mass Spectrometry* (EMS, IM Publications, Chichester, UK) in 1995, and some have been amalgamated, but none have been discontinued. This popularity of the MS literature reflects the way that the MS technique has become a routine laboratory tool over the last two decades. It also mirrors the development and application of the newer MS techniques such as electrospray ionization and matrix-assisted laser desorption/ionization (MALDI), which have become relatively affordable to most laboratories and have opened up the biological and (bio)polymer areas of science to MS. This, in turn, means that we find many significant papers on MS in ''non-MS'' journals that deal primarily with biochemistry or biology, far more so than in the mid-1980s when electron ionization and chemical ionization were the most common MS techniques.

There are several journals dealing exclusively with MS, as detailed in Table 1. Most of these are fully referred or sponsored so that the quality is assured. However, there

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Primary mass spectrometry journals

Journal title	Publisher	First issue	Number of 1999 issues	ISSN
European Mass Spectrometry	IM Publications	1995	6	1356–1049
International Journal of Mass Spectrometry	Elsevier Science	1968	27	1387–3806
Journal of the American Society for Mass Spectrometry	American Society for Mass Spectrometry	1990	12	1044–0305
Journal of Mass Spectrometry	Wiley Interscience	1995	12	1076–5174
Journal of the Mass Spectrometry Society of Japan	Mass Spectrometry Society of Japan	1953	6	1340–8097
Mass Spectrometry Reviews	Wiley Interscience	1982	4	0277–7037
Rapid Communications in Mass Spectrometry	Wiley Interscience	1987	24	1939–2010

are different philosophies behind some journals, involving criteria such as speed of publication and content.

The first journal on MS was published in Japan by the MS Society of Japan and began in 1953. Named *Mass Spectroscopy*, it was retitled *Journal of the Mass Spectrometry Society of Japan* (JMSSJ) in 1992. It contains many excellent review papers and reports of primary research, and is available worldwide, but is intended largely for the home market because many papers are published in the Japanese language. However, the non-Japanese scientist may still extract much useful information because the titles, author abstracts, figures and tables, references, and some technical data in the body of the text are in English.

The *International Journal of Mass Spectrometry and Ion Physics* (Elsevier Science, Amsterdam, The Netherlands) was the second MS journal to be published, being launched in 1968. It was retitled *International Journal of Mass Spectrometry and Ion Processes* in 1993, then *International Journal of Mass Spectrometry* (IJMS) in 1998. It has always aimed to publish papers on the fundamental aspects of MS and ion processes, including new instrumentation and practical applications. This journal also produces regular special issues, some of the later ones covering MALDI MS, electrospray ionization and ion spectroscopy. Other special issues are regularly produced in honor of eminent scientists in the field.

Also first published in 1968 was *Organic Mass Spectrometry* (OMS, Heyden). Its sister publication *Biomedical Mass Spectrometry* (BMS, Heyden, London, UK) was launched in 1974 and both journals were acquired by John Wiley & Sons in 1982. BMS underwent two name changes to *Biomedical and Environmental Mass Spectrometry* (1985), then *Biological Mass Spectrometry* (1990), before OMS and BMS were merged in 1995 to form the *Journal of Mass Spectrometry* (JMS), which is published under the Wiley Interscience imprint. The main reason for the merger was the convergence of the content of articles in both journals, with the movement towards biological MS in the community, and the fall in the number of pure organic MS studies, so that the journals were beginning to appear similar in any case. The current scope of the journal is wide, encompassing all aspects of MS. As well as regular articles and accelerated communications, it includes ''Perspectives'' and ''Tutorial'' articles. Each issue also contains a current awareness database that lists citations to papers on MS that have appeared recently in the chemical literature.

As a direct response to criticisms from authors about the delay times from submission of articles to publication, a new journal was published in 1987: *Rapid Communications in Mass Spectrometry* (RCM, Wiley Interscience). Contributions may be submitted to the journal by the normal method, whereby the Editor will send them to independent referees for assessment. Alternatively, the novel concept of sponsorship referees may be employed. Here, the authors themselves arrange for an independent referee or sponsor from another institution to act as sole referee for the paper, which is then submitted to the journal. Papers approved by referees or sponsors are generally accepted within 1–2 days and published within 4–5 weeks. The publishers also emphasize that ''rapid'' does not necessarily mean brief: there is no formal limit on paper length. The quick turnaround of papers has proved extremely popular with authors, as reflected in the number of issues per annual volume. This has increased from the original 12 issues to 24 in 1999. The journal publishes papers on all aspects of gas-phase ions and the associated scientific disciplines. It also publishes occasional issues that have groups of papers from a single laboratory, to illustrate the contribution that that particular organization has made to the development of MS.

In 1990, the American Society for Mass Spectrometry (ASMS) launched its official publication *Journal of the American Society for Mass Spectrometry* (JASMS) in conjunction with Elsevier Science. This journal also claims rapid publication of articles, with an average time following acceptance of 4 months for full papers and 2 months for communications. JASMS publishes

comprehensive research papers on all aspects of MS. The publication schedule has increased from 6 to 12 issues per annum, again reflecting both the popularity of the journal and the growth of MS. JASMS is available on subscription as with regular journals but is also supplied free to all members of the ASMS.

The most recent new journal devoted to MS was launched in 1995 when IM Publications launched EMS. To quote Allan Maccoll, the Chairman of the original Managing Board, it was felt that ''there is a place for a journal of independent spirit. The unique organisation of the Journal will ensure, through its Managing Board, close contact between the Publisher and the authors and readers.'' Contrary to the implication in the journal title, EMS publishes papers from worldwide sources. It aims to publish papers covering all aspects of MS, with special emphasis on the fundamental processes underlying MS: ionization, fragmentation, ion/molecule reactions and collisions, spectroscopy of gaseous ions, instrumentation and computers in MS.

A further MS journal is *Mass Spectrometry Reviews* (MSR, John Wiley & Sons), which is published six times a year at present. According to the co-editors Dominic Desiderio (USA) and Nico Nibbering (The Netherlands) the goals of the journal include the active solicitation of high-quality reviews from leading researchers in the field, the provision of critiques of the research published in different sub-specialities of MS, and the provision of an appropriate suppository of the research accumulated in a particular area of research in MS. Up to six reviews are published in each issue. MSR also contains a regular section entitled ''Selected Reviews on Mass Spectrometric Topics'', which cites published literature references on particular areas of MS.

Apart from the core journals for MS, there exist a multitude of journals that include major papers dealing with MS developments and applications. A newcomer to MS may be surprised at the number and spread of different journals containing ''useful'' MS information. These are not confined to the analytical chemistry literature but are included in many different types of journals that cover chromatography, forensic chemistry, food science, biotechnology, biochemistry, toxicology, occupational hygiene, environmental chemistry, drug metabolism and pharmacokinetics, clinical chemistry and natural products chemistry.

For example, *Analytical Chemistry* (American Chemical Society, Washington, DC) is devoted to the broad area of analytical chemistry but contains many key papers on MS. It also contains one issue each year devoted to reviews in analytical chemistry, covering particular calendar periods. These special issues are called *Fundamentals* and *Applications*, the former appearing in even-numbered years and the latter in odd-numbered years. An article on MS appears in the *Fundamentals* issue. Similarly, *Journal of Chromatography A* and *Journal of Chromatography B* (Elsevier) include a large number of papers on gas chromatography/mass spectrometry (GC/MS), liquid chromatography/mass spectrometry (LC/MS) and electrophoresis coupled with MS (e.g. capillary electrophoresis/MS).

An indication of this diversity may be found by browsing the secondary journals *Mass Spectrometry Bulletin* (MSB, The Royal Society of Chemistry, Cambridge, UK) and *LC/MS: The Applications Database* (formerly *LC/MS Update*) (HD Science Ltd, Nottingham, UK) from 1990 onwards. This date coincides approximately with the explosion of LC/MS and soft ionization MS techniques, which expanded MS into the biological/biotechnological areas and into the bio-type journals.

MSB covers all aspects of MS in organic and analytical chemistry, physical chemistry and chemical physics, including theory, instrumentation and applications. A count of the papers in MSB selected from particular primary journals shows the level of MS content in these journals. Over the 5-year period 1994–98, 905 papers from *Analytical Chemistry* were selected for MSB. For the same period, 734 papers from *Journal of Chromatography A* and 462 from *Journal of Chromatography B* were included. This distribution of MS over the different subject areas is illustrated in Table 2, which shows the

Table 2 Distribution of MS papers across different disciplines from 1994 to 1998[a]

Journal title	Area of study	Number of papers
Analytical Chemistry	General analytical	905
Clinical Chemistry	Clinical	62
Drug Metabolism & Disposition	Drug metabolism and pharmacokinetics	147
Environmental Science & Technology	Environmental	144
Journal of Agricultural and Food Chemistry		361
Journal of Analytical Toxicology	Forensic and toxicology	174
Journal of Biological Chemistry	General biochemistry	360
Journal of Chromatography A	General chromatography	734
Journal of Chromatography B	General chromatography	462
Journal of Natural Products	Natural products extraction and characterization	609

[a] As selected for inclusion in MSB over that period.

Table 3 List of publications scanned regularly for items for inclusion in *LC/MS: The Applications Database*

Acc. Chem. Res.	*Chem. Phys. Lett.*	*Jpn. J. Toxicol. Environ. Health*
Acta Chem. Scand.	*Chem. Phys. Lipids*	*J. Agric. Food Chem.*
Agents Actions	*Chem. Res. Toxicol.*	*J. Am. Chem. Soc.*
Am. Ind. Hygiene Assoc. J.	*Chemom. Intell. Lab. Systems*	*J. Am. Oil Chem. Soc.*
Am. J. Clin. Pathol.	*Chemosphere*	*J. Am. Soc. Mass Spectrom.*
Am. J. Hematol.	*Chemtech*	*J. Anal. Appl. Pyrol.*
Am. J. Med.	*Chinese J. Anal. Chem. (Fenxi Huaxue)*	*J. Anal. At. Spectrom.*
Am. J. Physiol.	*Chinese J. Chromatogr. (Sepu)*	*J. Anal. Toxicol.*
Am. Lab.	*Chinese J. Pharm. Anal.*	*J. Antibiot.*
Amino Acids	*Chirality*	*J. Assoc. Off. Anal. Chem.*
Anal. Biochem.	*Chromatographia*	*J. Bacteriol.*
Anal. Chem.	*Clin. Biochem.*	*J. Biochem.*
Anal. Chim. Acta	*Clin. Chem.*	*J. Biol. Chem.*
Anal. Commun.	*Clin. Chim. Acta*	*J. Biotechnol.*
Anal. Lett.	*Clin. Exp. Immunol.*	*J. Carbohydrate Chem.*
Anal. Methods Instrum.	*Clin. Pharmacokin.*	*J. Cardiovasc. Pharmacol.*
Anal. Sci.	*Comp. Biochem. Physiol.*	*J. Cereal Sci.*
Analusis	*Comput. Chem.*	*J. Chem. Ecol.*
Analyst	*Current Biol.*	*J. Chem. Educ.*
Anesthesiology	*Current Opin. Biotechnol.*	*J. Chem. Inf. Comput. Sci.*
Angew. Chem. Int. Ed. Engl.	*Cytotechnol.*	*J. Chemometrics*
Ann. Clin. Biochem.	*Deutsche Lebensm. Rundschau*	*J. Chromatogr. A*
Ann. Clin. Lab. Sci.	*Drug Metab. Dispos.*	*J. Chromatogr. B*
Anticancer Res.	*Eisei Kagaku*	*J. Chromatogr. Sci.*
Appl. Environ. Microbiol.	*Endocrinol.*	*J. Clin. Invest.*
Appl. Radiat. Res.	*Environ. Health Perspect.*	*J. Clin. Pharmacol.*
Arch. Biochem. Biophys.	*Environ. Pollut.*	*J. Dairy Res.*
Arch. Environ. Contam. Toxicol.	*Environ. Sci. Technol.*	*J. Endocrinol.*
Arch. Toxicol.	*Eur. J. Biochem.*	*J. Ethnopharmacol.*
Arzneim. Forsch.	*Eur. J. Clin. Chem. Clin. Biochem.*	*J. Food Sci.*
Aust. J. Chem.	*Eur. J. Clin. Invest.*	*J. Forensic Sci.*
Biochem. Biophys. Res. Commun.	*Eur. Mass Spectrom.*	*J. High Res. Chromatogr.*
Biochem. J.	*Experientia*	*J. Immunol.*
Biochem. Mol. Biol. Int.	*FASEB J.*	*J. Inher. Metab. Dis.*
Biochem. Pharmacol.	*Fat Sci. Technol.*	*J. Ind. Microbiol.*
Biochem. Soc. Trans.	*FEBS Lett.*	*J. Inorg. Biochem.*
Biochemistry	*Fenxi Ceshi Xuebao*	*J. Labelled Compds. Radiopharm.*
Biochim. Biophys. Acta	*Food Addit. Contam.*	*J. Lipid Res.*
Bioconjugate Chem.	*Food Chem.*	*J. Liq. Chromatogr.*
Biomarkers	*Food Chem. Toxicol.*	*J. Mass Spectrom.*
Biomed. Chromatogr.	*Forensic Sci. Int.*	*J. Mass Spectrom. Soc. Jpn.*
Bioorg. Med. Chem.	*Forensic Sci. Rev.*	*J. Med. Chem.*
Bioorg. Med. Chem. Lett.	*Free Radical Biol. Med.*	*J. Microbiol. Methods*
Biosci. Biotech. Biochem.	*Free Radical Res. Commun.*	*J. Microcol. Sep.*
BioTechniques	*Fresenius' J. Anal. Chem.*	*J. Mol. Biol.*
Biotechnol. Bioeng.	*Genetic Comp. Endocrinol.*	*J. Nat. Prod.*
Br. J. Clin. Pharmacol.	*Genetic Eng. Biotechnol.*	*J. Neurochem.*
Br. J. Surgery	*GIT Fachz. Lab.*	*J. Occup. Med.*
Bull. Chem. Soc. Jpn.	*Helv. Chim. Acta*	*J. Org. Chem.*
Bull. Environ. Contam. Toxicol.	*Heterocycles*	*J. Pediatrics*
Bull. Soc. Chim. Belg.	*Inorg. Chem.*	*J. Peptide Res.*
Bull. Soc. Chim. Fr.	*Inorg. Chim. Acta*	*J. Pharm. Biomed. Anal.*
Bunseki Kagaku	*Int. Arch. Occup. Environ. Health*	*J. Pharm. Pharmacol.*
Can. J. Chem.	*Int. Chromatogr. Lab.*	*J. Pharm. Sci.*
Carbohydrate Chem.	*Int. J. Colorect. Dis.*	*J. Planar Chromatogr.*
Carbohydrate Res.	*Int. J. Environ. Anal. Chem.*	*J. Plant Growth Regul.*
Carcinogenesis	*Int. J. Mass Spectrom.*	*J. Sci. Food Agric.*
Cereal Chem.	*Int. J. Peptide Protein Res.*	*J. Soc. Cosmetic Chem.*
Chem. Br.	*Int. J. Pharmaceutics*	*J. Steroid Biochem. Mol. Biol.*
Chem. Commun.	*Int. J. Radiat. Biol.*	*Lab. Equip. Digest*
Chem. Eur. J.	*Int. Lab.*	*Lab. Microcomp.*
Chem. Pharm. Bull.	*Int. Spectrosc. Lab.*	*Lab. Practice*

Table 3 (*continued*)

Lab. Products Technol.	*Peptide Res.*	*Semin. Thromb. Hemostasis*
LC/GC Int.	*Peptides*	*Spectrochim. Acta*
Lebensmittelchem.	*Pest. Sci.*	*Spectroscopy*
Liebigs Ann./Recueil	*Pharmacol. Toxicol.*	*Spectrosc. Eur.*
Life Sci.	*Pharmazie*	*Spectrosc. Lett.*
Lipids	*Photochem. Photobiol.*	*Steroids*
Mass Spectrom. Rev.	*Phytochem. Anal.*	*Synth. Commun.*
Meat Sci.	*Phytochemistry*	*Talanta*
Metabolism	*Planta*	*Teratogen. Carcinogen. Mutagen.*
Microb. Pathogen.	*Planta Med.*	*Tetrahedron*
Microbios	*Plant Physiol.*	*Tetrahedron Lett.*
Mol. Microbiol.	*Proc. Natl. Acad. Sci. USA*	*Ther. Drug Monit.*
Nachr. Chem. Tech. Lab.	*Prostaglandins*	*Trends Anal. Chem.*
Nat. Prod. Rep.	*Prostaglandins, Leukotrienes Ess. Fatty Acids*	*Trends Biotechnol.*
Nat. Toxins	*Protein Sci.*	*Trends Food Sci. Technol.*
Nature	*Pure Appl. Chem.*	*VG Monographs*
Naturwissenschaften	*Radiation Res.*	*Xenobiotica*
Neurosci. Lett.	*Rapid Commun. Mass Spectrom.*	*Z. Lebensm. Unters. Forsch.*
Occup. Environ. Med.	*Rev. Sci. Instrum.*	*Z. Naturforsch.*
Paed. Res.	*Science*	
Pathology	*Sci. Justice*	

numbers of papers in journals representing different disciplines over the period 1994–98, as selected for MSB.

The list of publications from which papers have been selected for *LC/MS: The Applications Database* is reproduced in Table 3.

2.2 Internet Form

All of these MS journals remain popular in printed form, but now most of them are also accessible to various degrees as on-line versions via the Internet.

The bibliographic details of the Elsevier journals IJMS[2] and JASMS[3] can be searched at the Elsevier Web site. A free public service enables nonregistered users to search the tables of contents using titles, authors or keywords, and these are the only details displayed as search results. Subscribers to the printed versions may search the full papers on-line, with a free trial period. The Internet version of JASMS also displays ''fast track'' articles. These are fully peer-reviewed papers that have been accepted for publication, which are posted electronically as pre-copy-edited files within 2 weeks of acceptance. They are still considered to be ''in press'' by the publisher until the full articles appear on the printed page.

Two further general Web-based services offered by Elsevier are ContentsDirect,[4] and ContentsSearch.[5] ContentsDirect is a free e-mail alerting service for their books and journals. Tables of contents are delivered to individual e-mail accounts before publication, based on the user's predefined journal titles and their subject areas of interest for books. With ContentsSearch, the user can search the tables of contents on-line.

Internet versions of the journals JMS,[6] RCM[7] and MSR[8] are searchable through the Wiley Web site. Non-subscribers can display the bibliography for a particular issue but do not have access to the full articles. Fully registered users can interrogate the full journal issues using simple or advanced searches, the latter permitting up to three terms to be linked (such as title, author, affiliation, author abstract) and issue and date restrictions to be applied.

The current literature databases of JMS which appear at the back of each printed issue may also be downloaded for searching within Microsoft® Access in two forms: as the files alone or as a run-time version within Access.[9] The 1997 and 1998 version are currently available.

Wiley also offers the Web-based service Wiley Chemistry.[10] This offers links to several key Wiley products in chemistry. It is a different point of entry to search the book list or the Web-based journals and also provides access to news and content for Wiley journals spanning every subject area.

EMS is available as a full-text on-line journal on the IM Publications Web site.[11] Subscribers to the printed journal receive free and full access to the electronic version, and there is also the option of taking out an electronic-only subscription, at a discounted price. The on-line journal is searchable with the Adobe® Acrobat© Reader search software, which can be downloaded and used free of charge from the Adobe® Web site.[12] The electronic journal is a direct reproduction of the printed page, including text and figures. Titles, authors and abstracts of the papers can be searched and those users with an electronic subscription can access the

paper directly from the search results. Other users (nonsubscribers) can view the abstract only.

A further service offered by IM Publications is an alerting service for new papers in EMS. Called EMS Scan,[13] this provides registered users with listings of titles and authors of papers in each issue of EMS when the final abstracts are made available on the Web site, generally 3 weeks before printed copies are available. Users can then visit the Web site and view these abstracts. There is no charge for this option. A sample copy may be downloaded and searched for free as a trial.

There is also a limited Web-searchable version of JMSSJ.[14] Beginning with the 1998 volume, the contents of each annual volume to date, which comprise the title, authors, corresponding address and the journal citation, can be searched using the user's browser software. Similarly, each individual issue within a volume can be searched, from 1998 onwards, with the browser software over the same fields as above, with the addition of the author abstract. There are also plans to make a searchable keyword index available on the Web site.

Of the other non-MS journals that include papers on MS, many are also available on the Internet. Most publishers now have Internet versions or are actively working towards them. Subscribers to printed versions may be allowed free access or a specially reduced subscription to the Internet version. Some journals such as *Analytical Chemistry* also accept supplementary material for Web display. Apart from the usual figures and tables, this may also include animations and movie clips. The main advantage to using Internet versions is that they can be searched to locate papers of interest that can be downloaded for future reference. For retrospective searching over a series of issues, the time savings are significant. There are too many individual journal-containing Web sites to mention in this chapter and new ones are being created regularly. The reader is referred to the Web sites of the various scientific publishers for more information.

3 SECONDARY PUBLICATIONS

The previous section highlights the large number of journals that are of interest to MS practitioners, or to analysts wishing to find out about a particular MS method or an MS protocol for a particular analyte. These journals are excellent for keeping up to date, but further help is often needed to distil this information into manageable quantities, or for retrospective searching over a particular period. The on-line versions of primary journals go some way to help with this problem, although it means that the different journals must be searched separately.

Another solution is provided by secondary publications, which are generally abstracting publications. In general, these provide the titles, bibliography and abstracts of papers of interest, sometimes with added index terms. The abstracts may be direct copies of the original author abstracts, or, alternatively, are especially written for the particular abstracts publication. This type of publication is generally distributed in various formats, including printed, electronic and, more recently, on-line form. A review of the PC-based literature abstract sources for MS has recently been published.[15]

3.1 Mass Spectrometry Publications

The products that deal solely with MS rather than chemistry as a whole are *CA Selects* (Chemical Abstracts Service (CAS), Columbus, OH, USA), *CASurveyor*® (CAS), MSB, and a series derived from the Update Series: *LC/MS: The Applications Database* (formerly *LC/MS Update), MALDI: The Applications Database* (formerly *MALDI Update), Drugs: The Mass Spectrometry Applications Database* and *Pesticides: The Mass Spectrometry Applications Database* (HD Science Limited, Nottingham, UK).

3.1.1 CA Selects

CAS, a division of the American Chemical Society, produces a large abstracts database. About 8000 titles are monitored annually, including journals, patents, symposia, conference proceedings, dissertations, technical reports and new books in chemistry. From these, approximately 700 000 documents are selected for abstracting each year and the completed abstracts are placed in one of 80 sections. *CA Selects* is produced by searching through all of these 80 sections for newly added papers in one of 33 specially selected subject areas, such as *Mass Spectrometry*. Every 2 weeks, these are gathered together and printed in booklets. Each entry includes the title, authors, the address where the work was conducted, full journal citation and full CAS abstract. The CAS abstracts are a mixture of original author abstracts and those written by CAS staff. The CAS document analysts are guided in their abstracting and indexing by the contents of the original author abstracts. If an English-language author abstract accurately and clearly summarizes the principal findings, then this abstract will probably form the basis for the CAS abstract. It may even constitute the complete CAS abstract. CAS-written abstracts are intended to provide sufficient information on the chemical content of the document abstracted to allow the reader to determine whether it is necessary to consult the original publication for complete details. The MS detail contained in the abstract can range from highly detailed to sparse,

depending on whether MS is a major or less important part of the original paper. A typical abstract is shown below.

> *130: 104703a **Determination of olanzapine in human plasma and serum by liquid chromatography/tandem mass spectrometry**. Berna, M.; Shugert, R.; Mullen, J. (Dep. Drug Disposition, Eli Lilly and Co., Indianapolis, IN 46285 USA) *J. Mass Spectrom.* **1998**, 33(10) 1003–1008 (Eng), John Wiley & Sons, Ltd. A liq. chromatog./tandem mass spectrometric (LC/MS/MS) assay was developed for the quant. detn. of olanzapine (LY 170053, OLZ) in human plasma and serum. Bond Elut C_2 solid-phase extn. cartridges (single cartridge or 96-well format), in conjunction with a pos. pressure manifold, were used to ext. OLZ and its internal std., LY 170222, from the biol. matrix. Chromatog. resoln. of OLZ from endogenous plasma interferences and its metabolites was accomplished with a MetaChem monochrom HPLC (4.6×150 mm, d_p 5 μm). Detection was carried out with a Perkin-Elmer SCIEX API III Plus mass spectrometer using pos. ion atm. pressure chem. ionization and multiple reaction monitoring protocol. The linear dynamic range was from 250 pg mL^{-1} to 50 ng mL^{-1} of human plasma/serum using a 0.5 mL aliquot. The inter-day precision (relative std. deviation) and accuracy (relative error) in plasma ranged from 6.26 to 7.66% and from −3.54 to 7.52%, resp. The intra-day precision and accuracy in serum ranged from 3.46 to 8.76% and from −8.06 to 12.46%, resp. The assay is sensitive and selective, and will be used to support both human clin. and toxicol. analyses. Furthermore, using the 96-well solid-phase extn. format, sample prepn. can be easily automated.

Even though there are 26 issues each year within each *CA Selects* title, they are intended as browsing publications for current awareness, because there are no indices of any sort to help with retrospective searching.

3.1.2 CA Selects Plus™

As the title implies, *CA Selects Plus™* is a supplemented form of *CA Selects*. It is based on the same 33 subject areas, with the rapid inclusion of references from 1300 priority journals that CAS processes first, before they appear in CA itself. References to meetings, letters to the editor and other items that are not full papers and do not appear in CA are also added. Copies of original documents can be ordered from the CAS Document Detective Service.

3.1.3 CASurveyor®

This is another product derived from the full CA database, and distributed specifically on CD-ROM for Microsoft® Windows™ and Macintosh® systems. This series has 9 subject areas, called topics, one of which is entitled *Mass Spectrometry and Applications*. This topic includes papers on MS methodology and apparatus, experimental results obtained by spectroscopic techniques, ion cyclotron resonance/MS, ion kinetic energy studies, energetics, structural studies, reaction kinetics and mechanisms, and analytical applications. The references are included from 1300 priority journals that CAS processes immediately and completely, supplemented by partial coverage of many more journals. There is complete coverage of selected core journals, such as JMS.

CASurveyor® is updated monthly on a rolling basis, each new disk giving the last 36 issues. The information can be searched, displayed, downloaded and printed. The search software will allow a particular period to be searched, such as the latest update. Searches can be conducted as a text search, in which the fields are not specified, or the search term can be specified within a particular field, such as author or journal name. The displayed records include title, authors, author affiliation, citation and abstract. Structural diagrams and CAS Registry Numbers are also included, with a link from the Registry Number to a different screen showing the relevant substance record, which may include IUPAC name, common and trade names, molecular formula and structure.

3.1.4 Liquid Chromatography/Mass Spectrometry: The Applications Database, Matrix-assisted Laser Desorption/Ionization: The Applications Database, Drugs: The Mass Spectrometry Applications Database, Pesticides: The Mass Spectrometry Applications Database

These four products are discussed together, as they are all published by the same organization, with abstracts in the same style across all publications. Originally produced as printed publications and diskette-based products (*LC/MS Update; MALDI Update; GC/MS Update, Part A: Environmental* and *GC/MS Update, Part B: Biomedical, Clinical, Drugs*) the new generation are produced on CD-ROM only. Over 250 core journals are scanned regularly, and papers are located from other journals. Each item includes the full bibliography, including all author affiliations, abstract and keywords. The abstracts differ from those in other abstracting publications in that they have a practical bias. They are especially written for these databases from the MS point of view, to highlight the mass spectrometric content of the paper and include as much of the experimental set-ups for the liquid chromatography (LC), gas chromatography (GC) and MS systems as is reported in the original. This includes the type of column, temperature or solvent program, and all published mass spectrometer conditions.

* Reproduced by permission from the American Chemical Society, Vol. 130, 1999, © 1999 by the American Chemical Society.

The keyword index terms are also listed at the end of each document. A typical record, taken from *LC/MS: The Applications Database*, is shown below.

> *L98:2/3:426. **Identification of adenine nucleotide-containing metabolites of bisphosphonate drugs using ion-pair liquid chromatography-electrospray mass spectrometry**
>
> **Auriola, S. (*1), Frith, J. (3), Rogers, M.J. (4), Koivuniemi, A. (1), Monkkonen, J. (2)**
>
> (1) Dept. Pharm. Chem., (2) Dept. Pharmaceutics, Univ. Kuopio, Fin-70211 Kuopio, Finland, (3) Dept. Human Metab. Clin. Biochem., Univ. Sheffield, (4) Dept. Med. Therapeut., Univ. Aberdeen, UK.
>
> **REF: J. Chromatogr. B 704(1,2) 1997 187–195.**
>
> Neg.-ion electrospray (ESP) LC/MS/MS is used to analyze the adenine nucleotide-contg. mammalian metabolites of clodronate [CAS 10596-23-3], tiludronate [CAS 89987-06-4] and etidronate [CAS 2809-21-4]. Cultured murine J774 macrophage-like cells are extd. and the metabolites dissolved in 50 mM dimethylhexylamine formate (DMHF) (pH 5.0) for anal.
>
> **LC: Rheos 4000 pump, 20 µl Rheodyne 7725 inj., 12.5 cm × 3 mm i.d. (5 µm Purospher C18e), 23°, A: 10 mM DMHF (pH 5.0 with formic acid), B: 20 mM DMHF (pH 5.0)-acetonitrile (50:50, v/v), 4–20% B in 10 min, to 80% B in 30 min, 0.5 ml/min.**
>
> **MS: Finnigan MAT LCQ, ESP, neg. mode, −4.5 kV, capillary −39 V and 20°, lens 55 V, N_2 sheath gas, m/z 200–700, 200 ms/scan, CAD: collision energy 18%, m/z 160–700.**
>
> Mass spectra and chromatograms of the bisphosphonate adduct metabolites are presented. The adenosine phosphate metabolites of the 3 compds. are identified. Alendronate [CAS 66376-36-1] does not yield an ATP analogue.
>
> 18 Refs. 6 Figs. 0 Tables.
>
> **Keywords:** Clodronate; Drugs, metabolism; Electrospray/MS/MS, drugs; Etidronate; Ion trap, drugs; MS/MS, drugs; Tiludronate; Tissue, drugs.

All databases originate from 1991 and every CD-ROM issued contains the full database. Customers for the LC/MS and MALDI products receive one CD-ROM a year, plus three updates, and the Drugs and Pesticides databases are issued annually. The LC/MS version also contains infusion and flow injection MS papers that involve soft ionization techniques. The MALDI database also includes laser desorption/ionization (i.e. with no matrix). The Drugs and Pesticides databases contain GC/MS, LC/MS, fast atom bombardment (FAB) MS and nonhyphenated MS methods for detecting and analyzing these types. There is some overlap between the analyte-based products (Drugs and Pesticides) and the technique-based ones. For example, a paper on the LC/MS determination of a pesticide will appear in the LC/MS and Pesticides databases. However, there are many extra references in the Drugs and Pesticides databases, from work done by the other techniques.

The data are provided in two formats for searching with Adobe® Acrobat© and the Info Select® program (Micro Logic, Hackensack, NJ, USA). The Acrobat© reader program may be downloaded from the Adobe® Web site free of charge. Info Select® is a commercial personal information manager program that works in Microsoft® Windows™ format with hypertext type searching. Each complete item is treated as one field, so that any word in any item can be searched, without having to select which fields to search. The searching is interactive: as the search string is typed, the database is searched and the hits can be viewed at any time during input of the search term. The search item is highlighted in the retrieved records. The displayed hits can be written to a special subset created by the user, or printed off. Additionally, users may use the software to add their own comments and details to other files, or, if they copy the database files to the hard disk, they may annotate the abstracts themselves.

3.1.5 Mass Spectrometry Bulletin

Although the MSB does not contain abstracts, it is included here because it is a secondary publication dealing solely with MS and related physical processes. The journal was first published in 1966 and is issued 12 times a year, with currently about 500 items per issue. Items are selected from over 800 primary journals as well as books, conference proceedings, reports and patents, and are placed in one of eight sections: Instruments, Isotopic analysis and mass measurement, Chemical analysis, Organic chemistry, Atomic and molecular processes, Surface and solid state chemistry, Thermodynamics and reaction kinetics, and Miscellaneous. Each item lists the title, authors, full journal citation and then index terms from two separate indices. The Subject index is a fixed thesaurus of about 350 terms, corresponding to common MS techniques and concepts. Topics not covered here are included in the General Index, which is divided into nine sub-sections. These indices are listed at the end of the publication, with the author index. An additional two-part term at the end of each title gives an indication of the level of MS interest in the paper, from minor to major, with full, partial or no mass spectra. A record from MSB is shown below.

> *Tamoxifen metabolism in rat liver microsomes: identification of a dimeric metabolite derived from free radical

* Reproduced by permission from HD Science.

* Reproduced by permission from the Royal Society of Chemistry, Vol. 33, 1999.

intermediates by liquid chromatography/mass spectrometry (A2)

Jones, R.M., Yuan, Z.-X., Lim, C.K.

Source details: *Rapid Commun. Mass Spectrom.* 1999, **13**: 4211–215.
Document type: Journal
CODEN: RCMSEF
Language: English

Keywords: Biochemical; Drugs; Electrospray technique; LC/MS; Metabolites; Pharmacology; Rat; Structure confirmation; Tamoxifen; Biological fluids

Sections: 43; General organic chemistry, Drugs and pharmacology

MSB can also be accessed on-line via ChemWeb[16] and, for the UK higher education community, via BIDS (Bath Information and Data Services).[17] In ChemWeb, the whole backfile from 1989 to date is searchable, with over 80 000 items available over all of the divisions within the printed bulletin. All the fields and indices from the printed version can be searched, either globally or over selected fields or combinations of fields. Searches may also be limited to a particular Bulletin section. The search results are displayed as a list of titles and years of publication within MSB. Full items are viewed by selecting from this list one by one. Under the terms of the agreement with the publisher, up to 10% of one database update may be downloaded or copied, and up to 1000 records may be downloaded and held for searching within the user's organization.

3.2 Other Publications

There are numerous other products that contain abstracts of the scientific literature for various specialist markets, such as medicine, agriculture and physics. Many of these products contain some references that will be of interest to mass spectrometrists, but, for this article, dealing with the MS literature, only those products that are more specific towards MS are considered.

3.2.1 Analytical Abstracts

Although Analytical Abstracts (AA, The Royal Society of Chemistry, Cambridge, UK) is not specific to MS, it is the only other major product that produces abstracts specifically in analytical chemistry. The printed journal was established in 1954 and provides comprehensive coverage of new techniques and applications in all branches of analytical chemistry. Over 230 core journals are scanned for AA and papers from about 3000 different journals have been included. AA is updated monthly, with the addition of about 1200 items. There is a section on MS, which is generally fairly small, with most papers that include MS appearing in other sections such as Clinical and biochemical analysis, Pharmaceutical analysis and Environmental, agriculture and food. Each item contains the full bibliographic details, including the address of the corresponding (or first) author and the abstract. The abstracts concentrate on the analytical part of the original paper, even if this is not a major part of the work. The procedure is outlined and the apparatus described briefly. Analytical data such as recoveries, detection limits and coefficients of variation are included. Much detail is given on the chromatograph column and operating conditions in GC/MS and LC/MS systems, but very little information is provided on the mass spectrometer set-up. A typical abstract from AA is reproduced below.

*TI: Electrospray ionization mass spectrometry for the study of non-covalent complexes: an emerging technology.
AU: Pramanik, -BN; Bartner, -PL; Mirza, -UA; Liu, -Y-H; Ganguly, -AK
AD: Schering-Plough Res. Inst., Kenilworth, NJ 07033, USA
CP: USA
SO: J-Mass-Spectrom. Oct 1998; 33(10): 911–920
IS: 1076-5174
CO: JMSPFJ
PY: 1998
LA: English
PT: Journal
AB: A tutorial review is presented, of the utility of electrospray ionization MS for the determination of the molecular masses of non-covalent complexes with molecular masses in the range 19 000–34 000 Da. Typical examples are presented from work carried out by the authors which cover a full range of non-covalent interactions from weak to relatively strong binding energies, and these include a protein-ligand, inhibitor-protein-ligand, protein-protein and protein-metal complex interactions. The molecular masses of the complexes were determined with a mass accuracy of better than 0.01%, which is far superior to traditional methods based on SDS PAGE and GPC. The key to success in such studies depends on an understanding and manipulation of electrospray ionization source parameters and sample solution conditions. The following were found to be particularly important: (i) the orifice potential, which should be sufficient to promote ionization without dissociation of the complex; (ii) the acidity of the complex; and (iii) the use of an organic co-solvent is strictly prohibited since the interior pocket of the protein would be disrupted resulting in denaturation. The usefulness of the technique for biological problems leading to the discovery of new therapeutics and the rapid screening of potential drug candidates is also illustrated. (33 references).
IA: complexes-A: study of non-covalent, by electrospray MS, tutorial review

* Reproduced by permission from the Royal Society of Chemistry, Vol. 61, 1999.

IC: mass-spectrometry, -electrospray- (electrospray-MS) -C: in study of non-covalent complexes, tutorial review
SC: C-Spectroscopy-and-Radiochemical-Methods
SS: 40 000
CR: G001; F0
COP: Copyright: The Royal Society of Chemistry
AN: 6103C00058
UD: 6103

The printed version has one index, the Subject index, that is prepared with a controlled vocabulary in order to facilitate searching. Within this index, there are three categories of index term: analyte, matrix and concept.

AA is also available on CD-ROM,[18] using the SilverPlatter Information Retrieval System (SPIRS) software in DOS, Windows™ and Macintosh® formats and can be searched on the Internet, also using SPIRS. The database is covered from 1980 onwards, with quarterly updates adding about 3600 items each time. The total size is over 235 000 records. Only items from 1984 onwards contain abstracts. Those from 1980 to 1983 contain titles, bibliographic detail and index entries. In this case, the three categories of Subject index term can be separately searched. There is a basic index comprising words from nearly all the fields, which can be used to locate techniques or chemicals. Otherwise the separate fields can be specified, to locate, for example, a specific compound as analyte (analyte index) or a technique (in the concept index). In retrieved items, the search term is highlighted. Further search terms can be selected from displayed records to refine the search. Records may be sorted for viewing, printing or downloading. Searches can be saved for a later session and the saved search run on a later CD-ROM update. In the WebSPIRS version, AA can be accessed over the Internet.[19] Customizable user interfaces can be created by the user, search screens can be tailored and there are simple and advanced search options. The software enables cross-database searching to be implemented and searches can be set to run automatically at timed intervals or when a new update is added.

AA is also available via ChemWeb[16] and several on-line hosts, including Dialog, DataStar (Dialog and DataStar, Dialog Corporation, Mountain View, CA, USA), Orbit (Questel.Orbit, Nanterre, France), and STN (Columbus, OH, USA).

3.2.2 Chromatography Abstracts

A further abstracting journal in the analytical chemistry area is *Chromatography Abstracts* (The Royal Society of Chemistry), produced in printed form only. This was originally published by the Chromatographic Society and is now in its 42nd annual volume. It covers chromatography and separation science, divided into five sections: General and Miscellaneous Techniques, GC, LC, Electrophoresis and Thin-layer chromatography. The MS papers can occur in any of the five sections. The abstracts are specially written in a similar manner to those of AA, concentrating on the analytical and chromatographic details of the original publication, such as sample sizes, column type and column conditions. Although the mass spectrometric experimental details are not highlighted, more detail is sometimes given than in the original author abstract. A typical record is shown below.

*1222. **Gas chromatography-mass spectrometry method for the quantitation of 3-chloro-*p*-toluidine hydrochloride in birds using a deuterated surrogate.**

Hurlbut, D.B., Johnston, J.J., Daniel, S.R., Tawara, J., *J. Agric. Food Chem.*, Nov 1998, **46**(11), 4610–4615.

3-Chloro-*p*-toluidine hydrochloride (**I**) and its d_6-labelled analogue (internal standard: IS) were extracted from spiked pigeon breast or gastrointestinal tissue with 2N-NaOH saturated with NaCl and back-extracted with propan-2-ol and hexane, then isolated on a silica (1 g) SPE column. Analytes **I** and IS were eluted with 2 ml n-butyl acetate, subsequently diluted to 5 ml. Portions were analysed by GC on a fused-silica column (30 m × 0.25 mm i.d.) coated with DB-1 (0.25 µm) operated with temperature programming from 90 °C (held for 0.25 min) to 300 °C at 35 °C/min with He as carrier gas (1 ml/min) and EIMS detection operated in selected-ion monitoring mode at *m*/*z* 106, 140 and 141 and 112, 147 and 149, respectively, for **I** and IS. The calibration graph was linear from 0.025–2 µg/ml of **I**, with a detection limit of 0.03 µg/g in both tissues. Recoveries of **I** from fortified tissues were 56.4–92.4%, whereas surrogate corrected recoveries were 95–101%.

The Subject index leads to some papers via the index term ''mass spectrometry'', but many other papers are found under index terms for the analyte or group of analytes, so much browsing is required to locate all of the MS articles.

3.2.3 Current Contents

Current Contents (CC) is a bibliographic database published by the Institute for Scientific Information, Philadelphia, PA, USA. It is a more general product, covering the published research literature, using over 7000 journals and 2000 books for inclusion in one of seven editions. The three editions that contain material of interest to mass spectrometrists are Physical, Chemical and Earth Sciences, Life Sciences and Clinical Medicine, covering 957, 1374 and 1045 publications, respectively.

* Reproduced by permission from The Royal Society of Chemistry, Vol. 42, 1999.

The items include full bibliography, keywords and, except for the printed edition, author abstracts. No special efforts are made outside of the normal keyword creation to emphasize the MS content of the original publication. Five distribution formats are available for CC: printed, diskette, via File Transfer Protocol (FTP), CD-ROM and on the World Wide Web.

The diskette version is supplied in Windows™ or Macintosh® versions and the inclusion of abstracts is optional. The FTP option allows delivery of the same data and functionality as the diskette, the main advantage being speed of delivery. Multi-issue searching of the diskette and FTP versions is possible, with nine different search options, including journal, title, author and keyword. The search results may be printed and downloaded. The tables of contents of journals may also be browsed page by page.

CC on CD-ROM is issued weekly on a 52-week rolling basis, with an annual CD-ROM at the end of each year. This version is in Windows™ and DOS formats, and uses the Ovid® software (Ovid® Technologies, New York) for searching. There are 18 discrete search fields. The Web version, Current Contents Connect™, also allows access to a 1-year rolling file and has multi-query searching. All data, including abstracts, can be searched using simple search strategies of personalized search profiles. The records may be exported into other data management software. A novel aspect of the Web version is the inclusion of Web links to facilitate communication with the corresponding authors or publishers.

All five CC formats have facilities for ordering copies of the original document. Two complementary products from ISI are the CC Reference Edition and CC Proceedings. The former is a quarterly CD-ROM that accumulates to an annual archive disc and is useful in conjunction with the diskette and printed versions. CC Proceedings is a monthly diskette product containing complete information on conferences, symposia and workshops in the two editions: Engineering and Physical Sciences and Biomedical, Biological and Agricultural Sciences.

4 MISCELLANEOUS LITERATURE SOURCES

Other printed sources of MS information include trade journals or magazines that are free to individual subscribers who fulfill the qualifying criteria (which are not usually very rigorous). There are none covering MS alone, but several deal with spectroscopy in general, including *Spectroscopy* (Advanstar Communications, Cleveland, OH, USA), *Spectroscopy Europe* (Wiley-VCH, Weinheim, Germany, IM Publications, Chichester, UK) and *International Spectroscopy Laboratory* (a subset of *International Laboratory*, International Scientific Communication, Shelton, CT, USA). Other publications of this type that are directed towards chromatographers, such as *CAST* (*Chromatography and Separation Technology*, RoseWard Publishing, Guildford, UK) and the more general laboratory publication, such as *LabAnalytica International* (Globetech Publishing, Wilton, CT, USA), also occasionally contain items related to MS, although some are written as promotions of MS equipment by representatives of instrument manufacturers.

Many conference proceedings are also published separately, but circulation may be limited to the conference delegates. In these cases, it is arguable whether or not the proceedings should be considered as published. In addition, the contents often consist of posters or presented papers that have not been refereed before publication. The quality of some entries may be questionable and much of the work is still in progress. Some, but by no means all, of these conference papers may be published later as full articles in a regular journal (or as a special journal issue) or as a book, when the work is more complete and the submissions have been refereed. Only then do some scientists accept them as valid, trustworthy publications.

One notable exception to the perceived value of nonrefereed conference proceedings is the extended abstracts from the ASMS annual meeting, published in book form up to 1998 and on a CD-ROM from 1999. Despite the lack of refereeing, it is evident from citations in other published MS papers that there is more confidence attached to these submissions and they are often used to establish scientific credit.

A further source of MS information is the application notes published by the mass spectrometer manufacturer. These will contain much technical information, generally on novel aspects or applications of the instrumentation, and are usually available by ordering from the appropriate Web site.

5 INTERNET RESOURCES

The Internet is a vast resource available to scientists, and reference to its uses in other aspects of MS and analytical chemistry will be made in other parts of this encyclopedia. For mass spectrometrists, one of the main uses is to identify the many different sources of chemical literature that may be available in printed, diskette, CD-ROM, online and Internet-based formats. These may be produced by government organizations, commercial organizations and academia, and no attempt is made here to locate and identify them all. In general, these sources include the bibliography from the original citation and some also

contain chemical structures and author abstracts. They are intended for more general markets, such as chemistry or medicine, but remain useful for directing the user within that area to more specific papers on MS and related subjects.

Some of the initial problems encountered in using the Internet involved filtering the sites to find the useful ones, but these have been partly eradicated. However, a general search of ''mass spectrometry'' will give a large, almost unmanageable number of hits, with many useful ones far down the list. To alleviate this problem, some Web sites have been specifically designed to provide links to sites of most interest to mass spectrometrists. The first and currently one of the best of these was set up and maintained at Emory University by Kermit Murray, and is now operated by John Wiley & Sons under the title Base Peak,[20] with Murray as Editor. A full explanation of the original purpose, design and intended use of this site has been published.[21] It is organized under various headings, all listing live links to other sites: organizations (national societies and associations), journals, meetings, local MS discussion groups, listservers, Usenet groups, information/archives, education, employment and non-English language resources. There is also a section set up so that the user can search the World Wide Web by selecting one of the many popular Web search engines. Wiley have also added links to their own Web site to the journals and books pages.

Many of the sites, such as those of the national MS societies and groups, also have links to other MS resources, including publishers, so that an intricate and valuable subset of the World Wide Web devoted to MS is building up. Many sites now exist simply to provide links to other sites, by collecting together the addresses of journals that are published on the World Wide Web. Some subscription agencies, such as Ebsco On-line,[22] Blackwell's Electronic Journal Navigator,[23] Dawson/Faxon Information Quest[24] and SwetsNet,[25] provide links to the electronic journals on the Internet. Other Web sites, like ChemPort,[26] offer additional services, such as links from search results in several reference databases, to the corresponding full electronic text at publishers' sites. Some sites, many of which are maintained by particular academic departments, exist mainly to provide a service to their students, by collecting together information on sites of interest. Many of these sites are not restricted to the student but may also be accessed from outside the organization.

The ASMS is also making available the preliminary abstracts of its annual conference on the Web, at the ASMS Web site. The abstracts may be searched on-line.[27] Additionally, the required viewer program and the abstracts from any or all of the conference sessions can be downloaded for the PC[28] or the Macintosh®[29] and searched free of charge. The extended abstracts are distributed on CD-ROM.

A more specialized, but nevertheless useful, Internet resource comprises part of the Web site at www.lcms.com.[30] Listed here are the appendices from the book *A Global View of LC/MS*[31] (see entry under the book section). These provide sources for LC/MS information under various headings and are kept up to date, so providing dynamic appendices to the book. The literature resources include books, selected reviews, on-line services, journals, training and education.

The first Internet-only journal to be published, *The Internet Journal of Chemistry* (IJC),[32] was published in 1998. IJC is a purely electronic journal, with all submissions and publications done over the Internet. The statement of purpose includes the intention to use the special features of the Internet to represent images in three dimensions and in color, to publish dynamic processes as movies, to keep costs to a minimum and to give the subscriber the ability to customize journal output. It is probably only a matter of time before more journals, including MS journals, are published this way. The representation of peptide and protein processes in MS, for example, would benefit greatly by the use of three-dimensional imaging and color to represent conformational changes and fragmentations. The journal *Analytical Chemistry* now accepts supplementary material for display on the Internet and this concept is under consideration by the publishers of the main MS journals.

6 BOOKS

Books on all aspects of MS are included in this section, with the general guideline that they are included if published from 1990 onwards. This rule of thumb is relaxed for older books which are still relevant today, such as books on the principles of MS, rather than, for example, books on MS in the biosciences, in which there have been huge advances during the 1990s. This article attempts no critical review of any of the books, but rather an indication of the content of each. For details of new books published after March 1999, the reader is referred to the Web-based catalogs of the various publishers, and also to the Web sites at www.amazon.com[33] and www.bol.com,[34] which contain details of several million published books on all subjects.

The books listed here are segregated into three broad sections: Data compilations, Basic MS and teaching texts, and Applications. Within each section the books are sorted chronologically.

6.1 Mass Spectrometry Data Compilations

Mass Spectra and GC Data of Steroids, Androgens and Estrogens

H.L.J. Makin, J. Nolan, D.J.H. Trafford

In the SpecBook Series by Chemical Concepts, this collection contains almost 2500 newly measured mass spectra of androgens and estrogens and is the most comprehensive collection of mass spectral data currently available. Each spectrum is accompanied by its structure, systematic and trivial names, retention index, molecular formula and mass, CAS Registry Number and sample source.

1999, Wiley-VCH, ISBN 3-527-29644-1

Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and their Metabolites, Parts I–IV, 2nd edition

K. Pfleger, H.H. Maurer, A. Weber

This expanded second edition compiles GC data and the complete mass spectra of 6000 toxic compounds and their metabolites, with the addition of over 1500 new mass spectra in the newly released Volume IV. Tables of approximately 7500 potential poisons, giving empirical formulae, molecular weights, CAS Registry Numbers and pharmacological/toxicological categories make it possible to attempt identification of compounds whose spectra have not yet been recorded. Relevant search criteria (e.g. name, category, molecular weight) can be searched in separate lists. All spectra were measured in the authors' laboratory by a standardized procedure.

1999, Wiley-VCH, ISBN 3-527-29293-6

Instrumental Data for Drug Analysis, Volumes I–VII

T. Mills III, J.C. Roberson, W.H. Wall, K.L. Lothridge, W.D. McDougall, M.W. Gilbert (eds.)

In the Forensic and Police Science Series, this self-contained set contains information on over 1600 drugs, metabolites, drug-related compounds and pesticides. Data include chemical title, molecular formula and weight, synonyms and trade names, structures, Fourier transform infrared, nuclear magnetic resonance, ultraviolet, GC data and detailed mass spectral data. Volume VII contains the cumulative indices.

Vol. I, 1992, CRC Press, ISBN 0-8493-9521-6

Vol. II, 1992, CRC Press, ISBN 0-8493-9522-4

Vol. III, 1992, CRC Press, ISBN 0-8493-9523-2

Vol. IV, 1992, CRC Press, ISBN 0-8493-9524-0

Vol. V, 1992, CRC Press, ISBN 0-8493-9516-X

Vol. VI, 1996, CRC Press, ISBN 0-8493-8114-2

Vol. VII, 1996, CRC Press, ISBN 0-8493-8115-0

Identification of Essential Oil Components by Gas Chromatography/Mass Spectroscopy

R.P. Adams

This updated book contains the mass spectra and retention times of over 1200 of the most common compounds in essential oils derived from plants. The compounds are listed in order of their elution on a DB-5 column and Kovat's indices are included. The book is available separately or in combination with a diskette-based library for use with mass spectrometer data systems. The instrument formats available are Finnigan GCQ/Magnum & ITS 40 (.LBR), Finnigan ICIS (LIB), Finnigan ion trap (.LIB), Finnigan Incos (.LS), Fisons/VG Mass Lab (.IDB), Fisons/VG Lab Base (.IDB), HP JCAMP & Unix (.HPJ), HP Chemstation (.L), net CDG (.CDF), Palisade BenchTop/PBM (.SPC), Shimadzu QP5000 (.NAM), Teknivent Vector/2 (.SPC), Varian Saturn (.LBR) and Waters Millennium (.SPC).

1995, Allured Publishing, ISBN 0-931710-42-1

The Eight Peak Index of Mass Spectra, 4th edition

This printed index of mass spectral data contains the eight most abundant m/zs and intensities of over 81 000 mass spectra covering over 65 000 compounds. Molecular formula and weight, compound name, parent peak intensity and CAS Registry Number are also present. The complete data are sorted in three different ways, laid out in tables to allow easy identification of unknown compounds by (i) molecular weight subindexed on formula, (ii) molecular weight subindexed on m/z values, and (iii) m/z of the two most abundant ions.

1991, Royal Society of Chemistry, ISBN 0-85186-417-1

GC/MS of the Gibberellins and Related Compounds: Methodology and a Library of Spectra

P. Gaskin, J. MacMillan (eds.)

This book contains 951 electron impact mass spectra of gibberellins as trimethylsilyl, methyl ester-trimethylsilyl ether and trimethylsilyl ether-trimethylsilyl ester derivatives, to provide GC/MS reference data. Also included are chapters on methodology and on procedures for editing GC/MS data. A companion diskette version containing the line diagram mass spectra of all compounds is also available.

1991, Cantocks Enterprises Ltd., ISBN 0-86292-388-3

Important Peak Index of the Registry of Mass Spectral Data

F.W. McLafferty, D.B. Stauffer (eds.)

This is a derivative of the Wiley/NBS Registry of Mass Spectral Data (1989) and presents the same

information in different orders for different points of entry to the data. It selects from each spectrum only those peaks that are diagnostic for that compound. There are 136 141 different mass spectra of 115 917 compounds.

1991, Wiley Interscience, ISBN 0-471-55270-4

Wiley/NBS Registry of Mass Spectral Data

F.W. McLafferty, D.B. Stauffer (eds.)

This collection covers spectra for 112 272 compounds with structures for 89 903 compounds. The spectra are selected from 150 000 different spectra. The Registry features vertical column formatting for easy cross-comparison of mass spectra. Data are contributed from laboratories worldwide. The electronic version of this product has continued to be expanded and now contains 392 000 mass spectra.

1989, Wiley Interscience, ISBN 0-471-62886-7

Analytical Artifacts, GC, MS, HPLC, TLC and PC

B.S. Middleditch

This contains information on artifacts encountered in organic analysis. Emphasis is on impurities, by-products, contaminants and other artifacts and over 1100 entries are included. Each entry contains the common name, mass spectrum, GC data, CAS name and Registry Number, synonyms and a narrative discussion. Mass spectral data are indexed in a six-peak index. There are also formula, author and subject indices. An extensive bibliography contains complete literature citations.

1989 (reprinted 1991), Elsevier Science, ISBN 0-444-87158-6

Advances in Prostaglandin, Thromboxane and Leukotriene Research. Vol. 18. Mass Spectra of Prostaglandins and Related Products

C.R. Pace-Asciak

This title contains line diagram mass spectra of derivatized prostaglandins and related compounds.

1989, Raven Press, ISBN 0-88167-474-5

Electron Capture Negative Ion Mass Spectra of Environmental Contaminants and Related Compounds

R.A. Hites, E.A. Stemmler

This volume contains the electron capture negative ionization (ECNI) line diagram mass spectra of 361 compounds selected to include the major classes of environmental contaminants. Compound names, CAS Registry Numbers, molecular weights and structures are also given.

1988, John Wiley, ISBN 0-471-18709-7

The Bile Acids. Chemistry, Physiology and Metabolism. Vol. 4: Methods and Applications

K.D.R. Setchell, D. Kritchevsky, P.P. Nair (eds.)

One of the chapters on the MS of bile acids (Lawson and Setchell) gives 112 line diagram mass spectra of bile acid methyl ester–trimethylsilyl ether derivatives, and an appendix lists GC retention indices.

1988, Plenum Press, ISBN 0-306-42622-6

Pharmaceutical Mass Spectra

R.E. Ardrey, A.R. Allen, T.S. Bal, A.C. Moffat

This is a compilation of mass spectra of drugs, metabolites and other compounds of pharmaceutical interest. Over 1100 full spectra with ion lists are arranged in alphabetical order of compound name. Derivatives and thermal degradation products are indexed under the parent compound. It includes a molecular weight index, an eight peak index and an alphabetical index of compound names.

1985, Pharmaceutical Press, ISBN 0-85369-172-X

6.2 Basic Mass Spectrometry and Teaching Texts

The Encyclopedia of Mass Spectrometry

M.L. Gross, R. Caprioli (eds.)

This eight-volume reference work will provide comprehensive coverage of the full range of topics and techniques in the field of MS. The format will provide access to specific subjects, with volumes organized by broad topic. Volumes will include: Ionization Methods; Mass Analyzers and Associated Instrumentation; Hyphenated Methods; Elemental MS and Isotope Ratio MS; Chemistry and Physics of Gas-phase Ions; Applications of MS; Applications to Biochemistry and Medicine; Historical Perspective.

2001, Elsevier Science, ISBN 008 0438504

Encyclopedia of Spectroscopy and Spectrometry

J.C. Lindon, G.E. Tranter, J.L. Holmes (eds.)

This three-volume set provides comprehensive coverage of the state-of-the-art in spectroscopy and related areas. There are about 350 articles in total and MS is one of the eight major topics covered. Purchasers of the printed edition will also be able to register for on-line access to an electronic version.

1999, Academic Press, ISBN 0-12-226680-3

Understanding Mass Spectra. A Basic Approach

R.M. Smith, K.L. Busch

This textbook for higher level students provides a basic understanding of the interpretation of mass spectra and

can also serve as a self-help book. An overview of the theory and instrumentation is provided. Real-life problems are used as examples for the reader to solve, with answers and complete explanations at the end of the book.

1999, John Wiley, ISBN 0-471-29704-6

Mass Spectrometry, 2nd edition

J. Barker

This is the latest edition in the Analytical Chemistry by Open Learning (ACOL) Series. It provides an introduction to MS suitable for self-learning. The practical approach combines theoretical and operational aspects. Ion formation, instrumentation, interpretation of simple spectra and recent interfaces (electrospray, continuous-flow) are included.

1998, John Wiley, ISBN 0-471-96764-5

A Beginner's Guide to Mass Spectral Interpretation

T.A. Lee

This book is dedicated to the interpretation of mass spectra and excludes information on hardware and experimental conditions, which may be of limited use in the interpretation. It is aimed at under- and postgraduate level. Each chapter concentrates on a different class of compound, such as hydrocarbons, halocarbons, alcohols, ethers and phenols, aldehydes and ketones, N-containing compounds, thiols and thio ethers and heterocyclic compounds.

1998, John Wiley, ISBN 0-471-97628-8

GC/MS. A Practical User's Guide

M.C. McMaster, C. McMaster

This guide provides practical information on using, maintaining and troubleshooting a GC/MS system. It progresses through the stages of setting up a system, and performing an analysis, with practical tips. A basic coverage of the theory is provided to illustrate operational variables and their control.

1998, John Wiley, ISBN 0-471-24826-6

Liquid Chromatography/Mass Spectrometry. 2nd edition, Revised and Expanded

W. Niessen

This revision, part of the Chromatographic Science Series (no. 79), provides an updated and comprehensive account of the latest developments, principles and applications of LC/MS and related techniques. It covers all interfaces currently used, especially atmospheric pressure chemical ionization and electrospray, and includes discussions on the optimization of experimental and ionization conditions. There is an expanded applications section covering environmental, pharmaceutical, natural products/endogenous compounds, biochemical and miscellaneous applications.

1998, Marcel Dekker, ISBN 0-8247-1936-0

A Global View of LC/MS. How to Solve Your Most Challenging Analytical Problems

R. Willoughby, E. Sheehan, S. Mitrovich

Written in the style of a ''how-to'' book, this is a practical guide to learning about, acquiring equipment and making the correct choices in LC/MS. The four sections cover: Is LC/MS right for you?; Acquiring LC/MS capabilities; Solving problems with LC/MS; and Appendices (Sources for LC/MS information, Practical information for LC/MS, Interfaces – how do they work? Mass analyzers – how do they work? Pioneers in LC/MS). Structural approaches are given for interpretation, target analysis, quantification and method development.

1998, Global View Publishing, ISBN 0-9660813-0-7

Ionization Methods in Organic Mass Spectrometry

A.E. Ashcroft

This book concentrates on instrumental operation and gives step-by-step instructions on how to set up mass spectrometers using various ionization types. Atmospheric pressure chemical ionization, electrospray, electron ionization, chemical ionization, FAB, fast ion bombardment ionization (including continuous-flow), field desorption, field ionization, thermospray, and MALDI are included.

1997, Royal Society of Chemistry, ISBN 0-85404-570-8

Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications

R.B. Cole (ed.)

This is the only book currently available that deals solely with electrospray ionization. It presents a wide view of the current knowledge of the basic aspects of electrospray ionization MS. It is divided into four main sections, covering fundamental and mechanistic aspects of the electrospray process, coupling electrospray to various mass analyzers, interfacing electrospray to separation techniques, and applications to biochemistry.

1997, John Wiley, ISBN 0-471-14564-5

Spectral Interpretation of Organic Compounds

E. Pretsch, J.-T. Clerc

In the series Spectroscopic Techniques: An Interactive Course, this is a book/CD-ROM combined tutorial introducing the interpretation of ^{13}C- and ^{1}H-nuclear magnetic resonance, infrared, ultraviolet/visible spectroscopy and MS data. It is a hands-on tutorial based on the SpecTool® product.

1997, Wiley-VCH, ISBN 3-527-28826-0

Tandem Techniques

R.P.W. Scott

In the Separation Science Series, the functions and uses of instruments combining separation and identification techniques for analyzing mixtures are described. The interfaces are described in detail and comparative performance data is presented. The techniques include GC/MS and LC/MS.

1997, John Wiley, ISBN 0-471-96760-2

Spectrometric Identification of Organic Compounds, 6th edition

R.M. Silverstein, F.X. Webster

This is a self-teaching book on the identification of organic compounds from their mass, infrared and nuclear magnetic resonance spectra. It follows a problem-solving approach, with a large set of real data problems and an extensive set of reference charts and tables.

1997, John Wiley, ISBN 0-471-13457-0

Introduction to Mass Spectrometry, 3rd edition

J.T. Watson

This completely revised and updated edition provides a guide to the concept of MS and demonstrates its potential and limitations. Using real-life examples, the book presents 18 realistic cases of qualitative and quantitative applications of MS. Systematic descriptions of the various types of mass analyzers and other mass spectrometer components, and strategies for interpretation of data, are included. This edition has eight new chapters, including six devoted to various types of ionization, such as electrospray, MALDI and inductively coupled plasma.

1997, Raven Press, ISBN 0-39751-688-6

Mass Spectrometry and its Applications

E. De Hoffmann, J. Charette, V. Stroobant

This is an introductory book covering the basics of MS. It covers many important modern techniques, such as electrospray ionization and MALDI, as well as modern instrumentation. It includes applications of MS to biological molecules, analysis of complex mixtures, and analysis of samples that require derivatization, chromatography and/or tandem analysis.

1996, John Wiley, ISBN 0-471-96696-7 (hardcover), ISBN 0-471-96697-5 (paperback)

Mass Spectrometry for Chemists and Biochemists, 2nd edition

R.A.W. Johnstone, M.E. Rose

This is a teaching text aimed at undergraduate and postgraduate levels, requiring no previous knowledge of MS and adopting an instructional approach. This edition includes new methods and applications that have developed in the past 10 years. Methods combining MS with newer separation techniques and analysis of polar and large-mass compounds such as proteins are discussed.

1996, Cambridge University Press, ISBN 0-521-41466-0 (hardback), ISBN 0-521-42497-6 (paperback)

Gas Chromatography and Mass Spectrometry. A Practical Guide

F.G. Kitson, B.S. Larsen, C.N. McEwen

This book covers the GC separation conditions for many compound types provided as derivatized and underivatized compounds. There is extensive correlation of ion masses and neutral losses with possible structures, and examples of spectra are provided. Also included is basic information on instrumentation, ionization methods, quantification, operating tips and derivatization procedures. The various appendices include troubleshooting tips for GC and MS.

1996, Academic Press, ISBN 0-12-833385-3

Quadrupole Mass Spectrometry and its Applications

P.J. Dawson

This book was originally published in 1976 by Elsevier and is reprinted as an American Vacuum Society Classic. It provides a broad study of the development and use of quadrupoles and the actions of radiofrequency fields, from basic operating principles to their use in mass filters, monopoles, ion traps and time-of-flight mass spectrometers.

1995, American Institute of Physics, ISBN 1-56396-455-4

Ion Traps

P.K. Ghosh

In the International Series of Monographs on Physics, this is the first book to present the principles of operation of both Paul and Penning ion traps. It examines confinement in both traps, including ion cooling. Applications of the traps in atomic physics and in frequency standards, collision studies and the use of the Paul trap in analytical chemistry are covered.

1995, Clarendon Press, ISBN 0-19-853995-9

Practical Aspects of Ion Trap Mass Spectrometry. Volume I. Fundamentals of Ion Trap Mass Spectrometry

R.E. March, J.F.J. Todd (eds.)

This is an account of the development and theory of the quadrupole ion trap, its utilization as an ion storage device, a reactor for ion/molecule reactions and as a mass spectrometer. Four chapters cover the fundamental aspects, including theory, development, operating principles and the production of the first

commercial version. Four more chapters cover the environment within the ion trap, including collisional cooling, ion trajectory simulations, boundary excitation and ion/molecule reactions.

1995, CRC Press, ISBN 0-8493-4452-2

Practical Aspects of Ion Trap Mass Spectrometry. Volume II. Ion Trap Instrumentation

R.E. March, J.F.J. Todd (eds.)

This volume covers the research and applications of ion trap instruments, often in tandem with other instruments or components, such as external ion sources and lasers. Also covered are high-resolution MS and mass measurement accuracy, ion trap confinement of externally generated ions, including by electrospray, ion structure differentiation, lasers, and ion traps in the study of physics.

1995, CRC Press, ISBN 0-8493-8253-X

Experimental Mass Spectrometry

D.H. Russell (ed.)

This book is designed as a guide for non-specialists, providing a thorough discussion of practice in modern MS, at the time of publication. Areas include biological MS, negative ion chemistry, collision-induced dissociation in tandem/structural MS, fundamental aspects of Fourier transform/ion cyclotron resonance (FT/ICR) MS and aspects of protein structure.

1994, Plenum Press, ISBN 0-306-44457-7

Practical Organic Mass Spectrometry. A Guide for Chemical and Biochemical Analysis, 2nd edition

J.R. Chapman

A fully updated edition, it includes time-of-flight and Fourier transform instrumentation and has an expanded LC/MS section, a new section on SFC/MS, and a new chapter on ''in-solution'' ionization techniques and their use in LC/MS coupling. Other chapters cover chemical ionization and ion–molecule reactions, negative ion chemical ionization, tandem MS (MS/MS) and quantitative analysis.

1993, John Wiley, ISBN 0-471-95831-X

Spectroscopic Methods and Analyses. NMR, Mass Spectrometry, and Metalloprotein Techniques

C. Jones, B. Mulloy, A.H. Thomas (eds.)

In the Methods in Molecular Biology Series (No. 17), this book gives practical information on these three research tools, such as required sample sizes and purities, advantages and limitations of the techniques, and data interpretation. The MS section includes an introduction and sections on laser desorption/ionization of bio-organic molecules, plasma desorption time-of-flight mass spectrometry (TOF MS) of peptides and proteins, FAB MS of peptides and tandem techniques.

1993, Humana Press, ISBN 0-89603-215-9

Interpretation of Mass Spectra, 4th edition

F.W. McLafferty, F. Turecek

This is the latest edition of the classic book on mass spectral interpretation, which uses set problems throughout the text for the reader to practice and gain experience. This edition contains new chapters related to ion energetics involved in fragmentation.

1993, University Science Books, ISBN 0-935702-25-3

Laser Ionization Mass Spectrometry

A. Vertes, R. Gijbels, F. Adams (eds.)

This book takes a comprehensive look at the new hardware and investigative possibilities of this form of analysis. It links theory with applications and provides chemists using MS with a definitive guide to laser ionization MS up to the early 1990s.

1993, John Wiley, ISBN 0-471-53673-3

Liquid Chromatography–Mass Spectrometry: Techniques and Applications

A.L. Yergey, C.G. Edmonds, I.A.S. Lewis, M.L. Vestal

This is an older treatment of LC/MS but is still very useful as there are so few books dealing exclusively with the subject. It covers direct liquid introduction interfaces, mechanical transport devices, thermospray and particle beam interfaces. Special coverage is given to nucleic acid constituents, glucuronides, acylcarnitines, sulfates, amino acids, peptides and proteins.

1991, Plenum Press, ISBN 0-306-43186-6

Field Desorption Mass Spectrometry

L. Prokai

In the Practical Spectroscopy Series, this book covers all aspects of field desorption MS: principles, experimental techniques and methods, general practice and applications.

1990, Marcel Dekker, ISBN 0-8247-8303-4

Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry

K.L. Busch, G.L. Glish, S.A. McLuckey

The only book devoted to MS/MS, this deals with the basic concepts of the technique and reviews the various types of instrumentation used. The different types of reaction involved in MS/MS are also considered, along with MS/MS spectra interpretation. Applications

in environmental chemistry, natural products chemistry, industrial products, foods and flavours, forensic chemistry, petroleum chemistry and the pharmaceutical industry are covered.

1989, Wiley-VCH, ISBN 0-471-18699-6

Quadrupole Storage Mass Spectrometry

R.E. March, R. Hughes

This monograph provides an extensive discussion of the operation of the ion trap detector and its application in MS. It surveys results obtained with the ion trap detector in gaseous ion spectrometry and physics. Applications of ion trap MS to toxicological studies, metabolism studies, environmental problems, water analysis and forensic science are covered.

1989, John Wiley, ISBN 0-471-85794-7

Fourier Transforms in NMR, Optical, and Mass Spectrometry. A User's Handbook

A.G. Marshall, F.R. Verdun

A unified treatment of the three most popular types of Fourier transform spectroscopy is presented, with uniform notation. All mathematics is self-contained and requires only a simple knowledge of calculus. Following the general treatments, one chapter is devoted to FT/ICR MS.

1989, Elsevier Science, ISBN 0-444-87360-0

Basic Gas Chromatography/Mass Spectrometry. Principles and Techniques

F.W. Karasek, R.E. Clement

This book covers the basic principles of GC and MS alone and in a GC/MS system. The data generated and their use are covered in detail, with extended coverage of the use of the computer and mass spectral search techniques. Qualitative and quantitative aspects are covered. GC/MS-computer applications are presented.

1988 (reprint 1991) Elsevier Science, ISBN 0-444-42760-0

Techniques of Combined Gas Chromatography Mass Spectrometry: Applications in Organic Analysis

W.H. McFadden

A comprehensive book on the state-of-the-art of GC/MS, this book still contains much relevant material today.

1988, Krieger Publishing Co., ISBN 0-8946-4280-4

Tandem Mass Spectrometry

F.W. McLafferty (ed.)

This book contains a collection of topics dealing with MS/MS as practiced in the early 1980s.

1983, John Wiley, ISBN 0-471-86597-4

Quantitative Mass Spectrometry

B.J. Millard

Even though this book is over 20 years old, it is the only book devoted to quantitative MS and contains information that is still relevant today.

1978, Heyden Publishers, ISBN 0-8550-1156-4

Mass Spectrometry. Organic Chemical Applications

K. Biemann

This is Volume 1 of the ASMS series Classic Works in Mass Spectrometry, which consists of books of historical and practical interest that are out of print. Chapters on instrumentation, sample handling and operating techniques, interpretation of mass spectra, and isotopically labelled compounds are followed by several chapters on the mass spectra of different compound classes, such as fatty acids and alkaloids.

1962, McGraw-Hill, no ISBN, republished 1997 by the ASMS

Mass Spectrometry and its Application to Organic Chemistry

J. Beynon

This is the second book in the ASMS series Classic Works in Mass Spectrometry.

1960, Elsevier, no ISBN, republished 1999 by the ASMS

6.3 Applications of Mass Spectrometry

GC/MS in Clinical Chemistry

P. Gerhards, U. Bons, J. Sawazki, J. Szigan, A. Wertmann

The basic principles of GC/MS are explained, with reference to applications in industrial medicine and drug analysis. Sample preparation methods and quality assurance are also covered. Drug information on epidemiology and pharmacology is presented to aid in drug screening studies.

1999, Wiley-VCH, ISBN 3-527-29623-9

New Methods for the Study of Biomolecular Complexes

W. Ens, K.G. Standing, I.V. Chernushevich (eds.)

This volume is in the NATO Science Series: C Mathematical and Physical Sciences (Vol. 510) and covers the proceedings of the NATO Advanced Study Institute Workshop held in 1996 in Alberta, Canada, on the study and characterization of noncovalent biomolecular complexes by electrospray and MALDI MS. Mass spectra of large complexes (several hundred kilodaltons) have been obtained on sector, quadrupole, time-of-flight and ion cyclotron resonance instruments. The use of other techniques in combination with MS for studying supramolecular complexes is also discussed.

1998, Kluwer Academic Publishers, ISBN 0-7923-5003-0

Advances in Mass Spectrometry, Vol. 14

E.J. Karjalainen, A.E. Hesso, J.E. Jalonen, U.P. Karjalainen (eds.)

This is the latest volume in the series reporting the proceedings of the International MS Conferences, this one being held in Tampere, Finland, in 1997. The articles are divided between a book and a CD-ROM. The book contains 33 papers. The CD-ROM has 300 additional papers, a copy of the book for easy searching, the conference abstract book and a QuickTime film recording of the closing lecture by R.G. Cooks.

1998, Elsevier Science, ISBN 0-444-50054-5

Mass Spectrometry of Biological Materials, 2nd edition, Revised and Expanded

B.S. Larsen, C.N. McEwen (eds.)

This edition has been completely rewritten and enlarged. It reviews ionization methods and instrumentation, introduces new technology, covers the isolation and characterization of very low levels of biological molecules and supplies in-depth coverage of screening protein databases using mass spectrometric data.

1998, Marcel Dekker, ISBN 0-8247-0157-7

Inductively Coupled Plasma Mass Spectrometry. From A to Z

A. Montaser (ed.)

This book covers new instrumentation, theory and applications of inductively coupled plasma MS. Fundamental applications, sample introduction, radiofrequency generators, and comparisons with other plasma mass spectrometric systems are covered.

1998, John Wiley, ISBN 0-471-18620-1

Current Practice in Liquid Chromatography/Mass Spectrometry

W.M.A. Niessen, R.D. Voyksner (eds.)

A spin-off from the *Journal of Chromatography A*, Volume 794, this book contains invited papers on modern aspects of LC/MS, all of which have been peer-reviewed in the normal manner for journals. The sections cover pharmaceutical/biomedical applications, environmental applications, natural products analysis, electromigration techniques and fundamentals/instrumentation.

1998, Elsevier Science, ISBN 0-444-82938-5

Trace Determination of Pesticides and their Degradation Products in Water

D. Barcelo, M.-C. Hennion (eds.)

In the Techniques and Instrumentation in Analytical Chemistry Series (Vol. 19), this is a critical compilation of analytical methods for monitoring pesticides and their degradation products in water. The section on general chromatographic methods includes electrophoresis, GC/MS and LC/MS techniques.

1997, Elsevier Science, ISBN 0-444-81842-1

Secondary Ion Mass Spectrometry. SIMS X

A. Benninghoven, B. Hagenhoff, H.W. Werner (eds.)

This is the latest book in a series, containing the proceedings of the 10th International Conference on Secondary Ion MS (1995), providing an overview of current research. Over 200 papers are included.

1997, John Wiley, ISBN 0-471-95897-2

Selected Topics in Mass Spectrometry in the Biomolecular Sciences

R.M. Caprioli, A. Malorni, G. Sindona (eds.)

This volume is in the NATO Science Series: C Mathematical and Physical Sciences (Vol. 504) and covers the Proceedings of the NATO Advanced Study Institute held in Altavilla-Milicia, Italy in 1996. Many fundamental aspects of MS are presented, including the principles and applications of electrospray, ion spray, MALDI, GC/MS interfaces and tools for quantitative analysis. The major part of the book deals with peptides, proteins, oligonucleotides, DNA adducts, polysaccharides, lipids, and plant metabolites.

1997, Kluwer Academic Publishers, ISBN 0-7923-4849-4

Time-of-Flight Mass Spectrometry. Instrumentation and Application in Biological Research

R.J. Cotter

The basic principles of TOF MS are presented, with a strong emphasis on biological research. Techniques such as orthogonal extraction, post-source decay and delayed extraction and the various compatible ionization techniques are covered. Applications to peptides, proteins, oligonucleotides and human genome sequencing are discussed.

1997, American Chemical Society, ISBN 0-8412-3474-4

Stable Isotopes

H. Griffiths, D. Robinson, P. Van Gardingen

The growing range of applications of stable isotopes for probing biological and geochemical cycles is reviewed in an interdisciplinary approach.

1997, Bios Scientific Publishers, ISBN 1-8599-6135-5

Plasma Source Mass Spectrometry. Developments and Applications

S.D. Tanner, J.G. Holland (eds.)

A current perspective on elemental analysis by plasma source MS is presented. The contents include elemental

analysis in liquids and solids, on-line separation, ion optics, instrumental developments, analytical applications and the quantification of radionuclides.

1997, Royal Society of Chemistry, ISBN 0-85404-727-1

Handbook of Drug Analysis Applications in Forensic and Clinical Laboratories

R.H. Liu, D.E. Gadzala

This book covers preliminary test and chromatographic methods in forensic drug testing, reviews identification methods such as molecular spectrophotometry, nuclear magnetic resonance (NMR) and MS, and discusses the fundamental relationship between instrumentation and drug analysis. It also evaluates the characteristics and pretreatment approaches for common sample categories and provides proven laboratory procedures for the analysis of commonly abused drugs.

1997, American Chemical Society, ISBN 0-8412-3448-5

Techniques in Protein Chemistry

D.R. Marshak (ed. 1996, 1997)

J.W. Crabb (ed. 1994, 1995)

R.H. Angeletti (ed. 1993)

This benchtop reference guide for protein chemists features up-to-date advances in protein methodologies. Most volumes contain sections on the use of MS in protein and peptide analysis.

Vol. VIII, 1997, Academic Press, ISBN 0-12-473557-6 (hardback), ISBN 0-12-473558-4 (comb-bound)

Vol. VII, 1996, Academic Press, ISBN 0-12-473555-X

Vol. VI, 1995, Academic Press, ISBN 0-12-194712-2 (hardback), ISBN 0-12-194713-0 (paperback)

Vol. V, 1994, Academic Press, ISBN 0-12-194710-6 (hardback), ISBN 0-12-194711-4 (comb-bound)

Vol. IV, 1993, Academic Press, ISBN 0-12-058757-2 (hardback), ISBN 0-12-058758-0 (comb-bound)

Modern Isotope Ratio Mass Spectrometry

I.T. Platzner

This is the first book on this subject for many years and it includes a comprehensive element-by-element review of 94 elements. Applications from the fields of analytical chemistry, geochemistry, astronomy and biomedical science are included, with special reference to technical details and instrumentation.

1997, John Wiley, ISBN 0-471-97416-1

Large Ions. Their Vaporisation, Detection and Structural Analysis

T. Baer, C.Y. Ng, I. Powis (eds.)

In the Wiley Series in Ion Chemistry and Physics, this book covers the mass spectrometric analysis of large molecules by techniques that include Fourier transform MS and MALDI MS. Models for MALDI, the laser ejection of oligonucleotides, and collisional activation of large molecules are studied.

1996, John Wiley, ISBN 0-471-96239-2

Applications of LC/MS in Environmental Chemistry

D. Barcelo (ed.)

In the *Journal of Chromatography* library (Vol. 59), this is a critical evaluation of LC/MS in environmental chemistry, written by experienced practitioners as a practical reference guide. An introductory section on fundamental aspects and instrumentation is followed by applications using most interfaces (particle beam, thermospray, electrospray). Analytes include pesticides, detergents, dyes, polar metabolites, waste streams, organo-tins and marine toxins.

1996, Elsevier Science, ISBN 0-444-82067-1

Mass Spectrometry of Soils

T.W. Boutton, S. Yamasaki (eds.)

This single-source reference is the first interdisciplinary work of its kind on the subject. It provides detailed coverage of the applications of proven spectrometric techniques in soil science, and presents the analytical approaches required. It contains nearly 1200 bibliographic citations, figures, and equations.

1996, Marcel Dekker, ISBN 0-8247-9699-3

Mass Spectrometry in the Biological Sciences

A.L. Burlingame, S.A. Carr (eds.)

Leading practitioners describe in detail advanced methods used in the analysis of biomacromolecules of natural and recombinant origin. They cover real-world problems in protein biochemistry, immunology, glycobiology and in the biochemistry of human bacterial pathogens, lipids and nucleic acids. The MS methods include MALDI TOF MS, electrospray, tandem techniques and FT/ICR. Database searching is also covered. A series of 10 appendices list useful reference data for the MS of peptides and carbohydrates.

1996, Humana Press, ISBN 0-89603-340-6

Mass Spectrometry in Biomolecular Sciences

R.M. Caprioli, A. Malorni, G. Sindona (eds.)

This volume is in the NATO Science Series: C Mathematical and Physical Sciences (Vol. 475) and covers the Proceedings of the NATO Advanced Study Institute held in 1993 at Lacco Ameno, Ischia, Italy. It covers

technological and applied developments, and includes ionization techniques and ion analysis. The newer ionization methods are reviewed and major instrumentation involved in mass analysis is described. The on-line combination of GC, LC and supercritical fluid chromatography with MS is also included.

1996, Kluwer Academic Publishers, ISBN 0-7923-3946-0

Protein and Peptide Analysis by Mass Spectrometry

J.R. Chapman (ed.)

In the Methods in Molecular Biology Series (No. 61), this book contains a comprehensive collection of MS methods for the analysis of peptides and proteins. Experts have contributed detailed step-by-step protocols covering topics such as recombinant protein analysis, peptide sequencing and quantification, analysis of glycoproteins and glycopeptides, measurement of drug loading on proteins, study of protein secondary structure, noncovalent interactions and using MS with PAGE.

1996, Humana Press, ISBN 0-89603-345-7

Methods in Enzymology. Volume 270. High Resolution Separation and Analysis of Biological Macromolecules, Part A

B.L. Karger, W.S. Hancock (eds.)

With its companion Volume 271, they cover the fundamentals and applications of LC, electrophoresis and MS, with chapters written by experts in their fields. The MS chapters cover FAB MS, electrospray ionization MS, matrix-assisted laser desorption MS of proteins, and quadrupole ion trap MS.

1996, Academic Press, ISBN 0-12-182171-4

Methods in Enzymology. Volume 271. High Resolution Separation and Analysis of Biological Macromolecules, Part B

B.L. Karger, W.S. Hancock (eds.)

The MS chapters in this volume cover protein structure analysis, carbohydrate sequence analysis by electrospray ionization MS, reversed-phase peptide mapping of recombinant human glycoproteins using LC/electrospray ionization MS, peptide characterization, capillary electrophoresis/MS, trace analysis peptide mapping by high-performance displacement chromatography and measuring DNA adducts by GC-electron capture/MS.

1996, Academic Press, 0-12-182172-2

Stable Isotopes in Human Nutrition

F. Mellon, B. Sandstrom (eds.)

In the Food Science and Technology International Series, this is a handbook of methodology dealing with the use of stable isotopes in human studies of absorption and the metabolism of essential mineral and trace elements. Sections cover nutritional techniques, sample handling and preparation, quantification of stable isotopes, and future developments. Quantification methods include electron ionization, GC/MS, FAB MS, thermal ionization MS and inductively coupled plasma MS.

1996, Academic Press, ISBN 0-12-490540-4

Applications of Modern Mass Spectrometry in Plant Science Research

R.P. Newton, T.J. Walton (eds.)

The basic principles, techniques and capabilities of MS in the analysis of plant-derived compounds are covered. Ionization methods, MS/MS techniques, MALDI MS and electrospray MS are prominent. The types of compounds covered include proteins, DNA, RNA, lignins, phytohormones, cyclic nucleotides, flavonoid glycosides, fatty acid derivatives, tropane alkaloids and brassinosteroids. This book is in the series of the Proceedings of the Phytochemical Society of Europe.

1996, Clarendon Press, ISBN 0-19-854965-2

Biofluid Assay for Peptide-related and Other Drugs. Methodological Surveys in Bioanalysis of Drugs, Vol. 24

E. Reid, H.M. Hill, I.D. Wilson (eds.)

This book describes state-of-the-art technology in the bioanalysis of drugs. The Assay Compendium contains a summary of literature based on therapeutic type, emphasizing more sensitive assays and automation. Approaches described for peptide-related drugs include immuno-based methods and MS methods, including MALDI MS, atmospheric pressure chemical ionization LC/MS, electrospray LC/MS and MS/MS techniques.

1996, Royal Society of Chemistry, ISBN 0-85404-741-7

Mass Spectrometry for Biotechnology

G. Siuzdak

This covers the fundamental concepts of MS and how the instruments work, with descriptions of recent advancements. Also included are a glossary of basic terms, practical tips on using mass spectrometric techniques, and a simplified description of MS, including MALDI, electrospray and FAB.

1996, Academic Press, ISBN 0-12-647471-0

Protein Sequencing Protocols

B.J. Smith (ed.)

In the Methods in Molecular Biology Series (No. 64), this book provides practical step-by-step details of techniques utilized for protein sequence analysis. Techniques covered include peptide isolation and mapping, amino acid

analysis, chemical and mass spectrometric approaches to N- and C-terminal sequencing, and analysis of sequence information.

1996, Humana Press, ISBN 0-89603-353-8

Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry. ACS Symposium Series No. 619

A.P. Snyder (ed.)

This book introduces improved characterization of proteins and biological molecules. The chapters include specific applications of electrospray ionization MS for analyzing proteins, peptides, glycolipids, lipopolysaccharides, oligodeoxynucleotides, nucleic acids, drugs and marine toxins.

1996, American Chemical Society, ISBN 0-8412-3378-0

The Protein Protocols Handbook

J.M. Walker (ed.)

This handbook offers comprehensive protocols for the analysis of proteins and peptides. The condensed techniques provided cover all of the basic methods regularly employed: 145 concise, step-by-step protocols which involve mainly chromatography and electrophoresis but with some MALDI and GC/MS.

1996, Humana Press, ISBN 0-89603-338-4 (hardback), 0-89603-339-2 (comb-bound)

Inductively Coupled and Microwave Induced Plasma Sources for Mass Spectrometry

E.H. Evans, J.J. Giglio, T.M. Castillano, J.A. Caruso

One of the RSC Analytical Chemistry Monographs, this book presents practical applications of the theory of inductively coupled plasma sources in a tutorial style, emphasizing the interfacing of microwave induced and inductively coupled plasma sources with MS, and interfacing chromatographic techniques.

1995, Royal Society of Chemistry, ISBN 0-85404-560-0

New Methods in Peptide Mapping for the Characterization of Proteins

W.S. Hancock (ed.)

This is a volume in the Analytical Biotechnology Series, and is devoted to the characterization of recombinant DNA-derived proteins by peptide mapping. It describes new technological procedures, including capillary electrophoresis, analysis of glycopeptides and the use of electrospray and MALDI MS.

1995, CRC Press, ISBN 0-8493-7822-2

Steroid Analysis

H.L.J. Makin, D.B. Gower, D. Kirk (eds.)

This volume provides comprehensive and up-to-date information on the analysis of steroids, mainly in biological fluids. Chapters on the extraction, separation and quantification of steroids are followed by chapters describing the analysis of specific steroid groups, including mass spectrometric methods.

1995, Blackie Academic, ISBN 0-7514-0128-5

Chemical, Environmental and Biomedical Applications of Ion Trap Mass Spectrometry

R.E. March, J.F.J. Todd (eds.)

This is really the third volume in the ion trap series (see the two books in the section on Basic MS, by the same editors). This volume consists of 14 chapters in four parts: Fundamentals; Practical ion trap methodology; Applications involving small molecules; Environmental and Biomedical Applications.

1995, CRC Press, ISBN 0-8493-8251-3

Applied Pyrolysis Handbook

T.P. Wampler (ed.)

This comprehensive book on the instrumentation and analytical applications of pyrolysis includes sections on pyrolysis/GC, pyrolysis/GC/MS and pyrolysis/MS.

1995, Marcel Dekker, ISBN 0-8247-9446-X

Forensic Applications of Mass Spectrometry

J. Yinon

The techniques discussed include capillary GC/MS, thermospray LC/MS, MS/MS and pyrolysis. A variety of applications include analysis of body fluids and hair for drugs of abuse, drug testing in sports, analysis of accelerants in fire debris and protection of industrial products by isotopic signature.

1995, CRC Press, ISBN 0-8493-8252-1

Time-of-flight Mass Spectrometry. ACS Symposium Series No. 549

R.J. Cotter (ed.)

A comprehensive discussion of TOF MS examining the use of time-of-flight TOF instruments in tandem mode and their interfacing with continuous ionization techniques. It includes examples of applications to biological research.

1994, American Chemical Society, ISBN 0-8412-2771-3

Mass Spectrometry. Clinical and Biomedical Applications. Vol. 2

D.M. Desiderio (ed.)

This volume reviews novel instrumentation and techniques as applied to the analysis of diacylglyceryl-phospholipids, modifications to DNA, the characterization

of variant hemoglobins and urinary nucleosides. Microdialysis/MS is also featured. (See also Volume 1 below, 1992.)

1994, Plenum Press, ISBN 0-306-44455-0

Biological Mass Spectrometry. Present and Future

T. Matsuo, R.M. Caprioli, M.L. Gross, Y. Seyama (eds.)

This volume covers developments in the field of biological MS, including electrospray ionization, matrix-assisted laser desorption, LC/MS, ion trap MS and FT/ICR. It contains specific applications to proteins, drugs and natural products.

1994, John Wiley, ISBN 0-471-93896-3

Time-of-Flight Mass Spectrometry and its Applications

E.W. Schlag (ed.)

This book was previously published as a special issue of the journal *International Journal of Mass Spectrometry and Ion Processes* (1994) and gives a survey of current applications from many active groups in the field.

1994, Elsevier Science, ISBN 0-444-81875-8

Applications of Plasma Source Mass Spectrometry

G. Holland, A.N. Eaton (eds.)

This series of books represents the current state of plasma source MS in all its forms, and charts the latest developments in a wide range of applications. Special attention is given to inductively coupled plasma MS and glow discharge MS.

1993, Royal Society of Chemistry, ISBN 0-85186-465-1

1991, Royal Society of Chemistry, ISBN 0-85186-566-6

Mass Spectrometry for the Characterization of Microorganisms. ACS Symposium Series No. 541

C. Fenselau (ed.)

The book includes features on GC/MS for the clinical diagnosis of viral infection, and summarizes the status of detection of biological agents and of modifications in ribonucleic acids in microorganisms. It also describes electrospray, laser desorption and MS/MS for protein processing and modification in viruses.

1993, American Chemical Society, ISBN 0-8412-2737-3

Advances in Chemical Diagnosis and Treatment of Metabolic Disorders I

I. Matsumoto (ed.)

This book contains a cross-section of updated and revised papers (many GC/MS) from Symposia on Chemical Diagnosis of Metabolic Disorders held in Japan (1990, 1991).

1993, John Wiley, ISBN 0-471-93889-0

Mass Spectrometry of Lipids

R.C. Murphy

This is a fundamental overview of the general physical concepts of MS that details the unique ion chemistry of specific lipids, ranging from the simple unsaturated fatty acids to more complex glycosphingolipids and glycerophospholipids.

1993, Plenum Press, ISBN 0-306-44361-9

Modern Methods and Applications in Analysis of Explosives

J. Yinon, S. Zitrin

This book discusses developments in techniques and methods for the analysis of explosives, and reviews recent applications. It contains a list of explosive compounds based on a survey of specialist laboratories all over the world. The contents include explosive compounds and mixtures, chromatographic methods, mass spectrometric methods, analysis of explosive residues, environmental analysis of explosives and detection of hidden explosives.

1993, John Wiley, ISBN 0-471-93894-7 (hardback), 0-471-96562-6 (softback)

Mass Spectrometry. Clinical and Biomedical Applications. Vol. 1

D.M. Desiderio (ed.)

The first Volume offers a thorough exploration of significant advances in theoretical and experimental research and instrumentation. Topics include the principles and practice of electrospray, quantitative analysis of neuropeptides, in-born errors of metabolism and the use of carnitine and acylcarnitines in metabolic disease diagnosis and management. (See also Volume 2 above, 1994.)

1992, Plenum Press, ISBN 0-306-44261-2

Chemical Ionization Mass Spectrometry, 2nd edition

A.G. Harrison

This is a comprehensive review of chemical ionization MS applications in structural elucidation and quantitative analytical studies. Chemical ionization mass spectra of 13 classes of compounds are discussed in detail to illustrate the influence of functional groups on the spectra observed.

1992, CRC Press, ISBN 0-8493-4254-6

Mass Spectrometry in the Biological Sciences. A Tutorial

M.L. Gross (ed.)

This volume is in the NATO Science Series: C Mathematical and Physical Sciences (Vol. 353) and covers the Proceedings of the NATO Advanced Study Institute held in 1990 at Cetraro, Italy. It is divided into three

sections, dealing with instrumentation for mass analysis and detection, methods in MS and applications to biomolecules.
1991, Kluwer Academic Publishers, ISBN 0-7923-1539-1

Continuous-flow Fast Atom Bombardment Mass Spectrometry

R.M. Caprioli (ed.)

This book is based on the presentations made at a workshop on continuous-flow FAB held in Annapolis, Maryland and, although fairly old, is included because there are so few books dealing with LC/MS as the main subject. The papers presented have been expanded and edited to form a small volume detailing advances in a number of key areas.

1990, John Wiley, ISBN 0-471-92863-1

Analytical Microbiology Methods. Chromatography and Mass Spectrometry

A. Fox, S.I. Morgan, L. Larsson, G. Odham

This book covers a broad array of approaches for automated and rapid identification of microorganisms, characterization of microbial products and constituents and trace detection of microbial chemicals.
1990, Plenum Press, ISBN 0-306-43536-5

Lasers and Mass Spectrometry

D.M. Lubman (ed.)

In the Oxford Series in Optical and Imaging Sciences No. 1, each chapter deals with a particular application area and forms an entry point for the utilization of MS for graduate students and researchers.

1990, Oxford University Press, ISBN 0-19-505929-8

Methods in Enzymology, Vol. 193, Mass Spectrometry

J.A. McCloskey (ed.)

The first section covers a range of general techniques and topics of contemporary importance in the applications of MS to the biological sciences in general. The three remaining sections emphasize specific applications divided among three major classes of molecules: peptides and proteins, glycoconjugates, nucleotides and nucleosides. Each of these sections opens with a critical overview of the applications of MS to that area.

1990, Academic Press, ISBN 0-12-182094-7

Liquid Chromatography/Mass Spectrometry. Applications in Agricultural, Pharmaceutical and Environmental Chemistry. ACS Symposium Series No. 420

M.A. Brown (ed.)

This volume describes the LC/MS techniques and instrumentation available and being used for analyzing agrochemicals, drugs, their metabolites and environmental pollutants. The analysis of pesticides, herbicides (chlorinated, sulfonylurea), antibiotics, retinoic acid, target and nontarget pollutants, phenols and phenol conjugates and polar urinary metabolites are included.

1989, American Chemical Society, ISBN 0-8412-1740-8

Mass Spectrometry in Clinical Biochemistry

A.M. Lawson (ed.)

This book serves as a reference handbook of applications of MS to the practice of clinical biochemistry. Applications to the work of specialized laboratories are discussed, covering subjects such as endocrinology, psychiatry, pediatrics, immunology, nutrition and others that embrace clinical biochemistry.

1989, Walter de Gruyter, ISBN 311-007751-5

Mass Spectrometry

M.E. Rose (ed.)

This Specialist Periodical Report series is now discontinued. It provided full reviews on the MS literature published over 2-year periods.

Vol. 10, 1989, Royal Society of Chemistry (1986–1988) ISBN 0-85186-348-5
Vol. 9, 1987, Royal Society of Chemistry (1984–1986) ISBN 0-85186-338-8
Vol. 8, 1985, Royal Society of Chemistry (1982–1984) ISBN 0-85186-328-0
Vol. 7, 1983, Royal Society of Chemistry (1980–1982) ISBN 0-85186-318-3
Vol. 6, 1981, Royal Society of Chemistry (1978–1980) ISBN 0-85186-308-6

ACKNOWLEDGMENTS

The author gratefully acknowledges the invaluable assistance of the publishers of the featured printed and electronic products in the preparation of this article and for access to Web-based products.

ABBREVIATIONS AND ACRONYMS

AA	Analytical Abstracts
ASMS	American Society for Mass Spectrometry
BMS	Biomedical Mass Spectrometry
CAS	Chemical Abstracts Service
CC	Current Contents
ECNI	Electron Capture Negative Ionization
EMS	European Mass Spectrometry

FAB	Fast Atom Bombardment
FT/ICR	Fourier Transform/Ion Cyclotron Resonance
FTP	File Transfer Protocol
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
IJC	The Internet Journal of Chemistry
IJMS	International Journal of Mass Spectrometry
JASMS	Journal of the American Society for Mass Spectrometry
JMS	Journal of Mass Spectrometry
JMSSJ	Journal of the Mass Spectrometry Society of Japan
LC	Liquid Chromatography
LC/MS	Liquid Chromatography/Mass Spectrometry
MALDI	Matrix-assisted Laser Desorption/Ionization
MS	Mass Spectrometry
MSB	Mass Spectrometry Bulletin
MS/MS	Tandem Mass Spectrometry
MSR	Mass Spectrometry Reviews
NMR	Nuclear Magnetic Resonance
OMS	Organic Mass Spectrometry
RCM	Rapid Communications in Mass Spectrometry
SPIRS	SilverPlatter Information Retrieval System
TOF	Time-of-Flight
TOF MS	Time-of-Flight Mass Spectrometry

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Laser Mass Spectrometry in Trace Analysis

Environment: Water and Waste **(Volume 3)**
Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry • Volatile Organic Compounds in Groundwater, Probes for the Analysis of

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications • Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Capillary Electrophoresis/Mass Spectrometry in Peptide and Protein Analysis • High-performance Liquid Chromatography/Mass Spectrometry in Peptide and Protein Analysis • Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis

Pesticides **(Volume 7)**
Gas Chromatography/Mass Spectrometry Methods in Pesticide Analysis • High-performance Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Pulp and Paper **(Volume 10)**
Pyrolysis in the Pulp and Paper Industry

Atomic Spectroscopy **(Volume 11)**
Glow Discharge Optical Spectroscopy and Mass Spectrometry

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry

• Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Tandem Mass Spectrometry: Fundamentals and Instrumentation • Time-of-flight Mass Spectrometry

Radiochemical Methods **(Volume 14)**
Mass Spectrometry of Long-lived Radionuclides

General Articles **(Volume 15)**
Literature Searching Methodology

REFERENCES

1. Chemical Abstracts Publication Record 1907–1998, CAS Statistical Summary, March, 1999.
2. http://www.elsevier.nl/inca/publications/store/5/0/0/8/4/7/
3. http://www.elsevier.nl/inca/publications/store/5/0/5/7/2/7/
4. http://www.elsevier.nl/homepage/alert.htt?mode=direct&main=/homepage/about/contentsdirect
5. http://www.elsevier.nl/homepage/alert.htt?mode=contents&main=/homepage/about/estoc
6. http://www.interscience.wiley.com/jpages/1076–5174/
7. http://www.interscience.wiley.com/jpages/0951–4198/
8. http://www.interscience.wiley.com/jpages/0277–7037/
9. ftp://ftp.wiley.co.uk/pub/journals/jms/
10. http://wiley.co.uk/products/subject/chemistry/ or http://www.wiley.com/products/subject/chemistry/
11. http://www.impub.co.uk/electronic.html
12. http://www.adobe.com/prodindex/acrobat/adobepdf.html
13. http://www.impub.co.uk/scan.html
14. http://wwwsoc.nacsis.ac.jp/mass/jmssjtop.html
15. O.D. Sparkman, 'Review of Personal Computer Literature Abstracts Sources for Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **7**(3), 299–305 (1996).
16. http://www.chemweb.com/databases/RSC/
17. http://www.bids.ac.uk/rsc.html
18. Analytical Abstracts on CD-ROM, SilverPlatter Information, Norwood, MA, USA.
19. http://www.silverplatter.com/
20. http://base-peak.wiley.com/
21. K.K. Murray, 'Internet Resources for Mass Spectrometry', *J. Mass Spectrom.*, **34**(1), 1–9 (1999).
22. http://www.ebsco.com/home/bropage1.asp
23. http://www.blackwell.co.uk
24. http://www.informationquest.com
25. http://www.swetsnet.nl
26. http://chemport.org/html/
27. http://www.asms.org/abstract_srch.html
28. http://www.asms.org/asms1999.exe
29. http://www.asms.org//asms1999Mac.hqx
30. http://www.lcms.com
31. R. Willoughby, E. Sheehan, S. Mitrovich, *A Global View of LC/MS. How to Solve Your Most Challenging Analytical Problems*, Global View Publishing, Pittsburgh, PA, 1998.
32. http://www.ijc.com
33. http://www.amazon.com
34. http://www.bol.com

Quadrupole Ion Trap Mass Spectrometer

Raymond E. March
Trent University, Peterborough, Canada

1 Introduction 2
2 Structure of the Quadrupole Ion Trap Mass Spectrometer 2
3 Literature 5
4 Trapping Pseudo-potential Well 5
5 Structure of the Quadrupole Ion Trap 6
6 Theory of Quadrupole Ion Trap Operation 6
6.1 An Ion in a Quadrupole Field 7
6.2 The Mathieu Equation 7
6.3 Potentials on the Electrodes 8
6.4 Stretched Ion Trap 9
7 Regions of Ion Trajectory Stability 9
7.1 Stability Region 10
7.2 Secular Frequencies 10
8 Resonant Excitation 11
9 Calculations 12
9.1 q_z and Low-mass Cut-off 12
9.2 β_z 12
9.3 w_z 13
9.4 Mass Range 13
9.5 Mass Range Extension 13
9.6 Mass Resolution 13
9.7 $\overline{D}_z$ 13
10 Mass Spectrometric Operation of the Quadrupole Ion Trap 13
10.1 Scan Function 15
10.2 Collision-induced Dissociation 15
11 Tandem Mass Spectrometry 17
11.1 Tandem Mass Spectrometry 17
11.2 Scan Functions 18
11.3 Tandem Mass Spectrometry Determination of Eluting and Co-eluting Compounds 19
11.4 Tandem Mass Spectrometric Operation 20
12 Chemical Ionization and Ion/Molecule Reactions 21
13 Conclusions 23
Acknowledgments 23
Abbreviations and Acronyms 23
Related Articles 23
References 24

The quadrupole ion trap is constructed of three electrodes that, when held at appropriate potentials, cause the formation of a trapping pseudo-potential well so that charged particles, or gaseous ions, may be confined or stored for long periods of time. The two end-cap electrodes resemble saucers while the ring electrode resembles a napkin ring; all of the electrodes are of hyperbolic geometry. The ion trap itself functions as a mass spectrometer when the ion-confining conditions are modified such that ions are ejected mass-selectively from the trapping potential well. As ions of successive mass/charge ratios are ejected in turn from the ion trap, they impinge upon an external detector whereby ion signals are created in proportion to the ion number of each species; in this manner, a mass spectrum is generated. The QITMS (quadrupole ion trap mass spectrometer) is an extraordinary instrument in that it is physically small (the entire electrode assembly can be held in the palm of one's hand) compared with magnetic and electric sector instruments, it is relatively inexpensive, it is one of the most, if not the most, sensitive mass spectrometers and, since several mass-selective operations can be carried in succession, the ion trap can function as a tandem mass spectrometer. Tandem mass spectrometric operation is described as $(MS)^n$. With the QITMS, $(MS)^n$ is carried out in time in the same volume of space whereas $(MS)^n$ in sector instruments is carried out in space. With sector instruments, the maximum value of n is $n = 4$ yet with the ion trap, $(MS)^n$ where $n = 4–6$ can be carried out routinely and $n = 13$ has been achieved. The QITMS shares several similarities with the ion cyclotron resonance mass spectrometer yet the cost of the former is about one-tenth that of the latter. One striking difference between the QITMS and all other mass spectrometers is that the QITMS operates at a pressure of 10^{-3} Torr compared with $10^{-6}–10^{-9}$ Torr for other mass spectrometers.

The theory of ion confinement and ion trajectory manipulation in the QITMS has been explained relatively simply so far. Since the theory differs widely from those of sector instruments and ion cyclotron resonance mass spectrometry (ICR/MS), it will not be familiar to those who have not had the opportunity to examine ion motion in quadrupole fields. Optimum operation of the QITMS is effected by collisional focusing of the ion cloud to the center of the ion trap under the influence of helium buffer gas. Since the movement of ions confined in the ion trap is periodic, the trajectories of collisionally focused ions can be expanded by resonance excitation effected by the imposition of supplementary radio frequency (rf)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

potentials of low amplitude to the end-cap electrodes of the ion trap. This excitation operation permits isolation of selected ion species, by ejection of unwanted ion species, followed by selective ion/molecule reaction or by collision-induced dissociation (CID) with subsequent mass analysis of the product/fragment ions formed. Sample calculations are given of all of the relevant trapping parameters. Applications of the QITMS as a single stage mass spectrometer and as a tandem mass spectrometer are discussed. The operation of the QITMS for the identication of dioxins and furans co-eluting from a gas chromatograph is described. In addition, the application of chemical ionization (CI) for the identification of co-eluting polychlorinated biphenyl (PCB) congeners is discussed. The QITMS is an extraordinary instrument that is capable of great sensitivity, high mass range and high mass resolution. Since the QITMS is compatible with methods for generating ions externally, such as electrospray ionization (ESI), its continued growth in many areas of mass spectrometry (MS) is assured.

1 INTRODUCTION

The QITMS functions not only as a mass spectrometer of high sensitivity and high specificity but also as an ion store in which gaseous ions can be confined for periods of some hundreds of milliseconds; both functions are employed in normal ion trap operation. While other mass spectrometers operate at pressures $<10^{-6}$ Torr, the ion trap operates at a pressure of 1 mTorr of helium buffer gas. As a storage device, the ion trap acts as an ''electric field test tube'' for the confinement of gaseous ions, either positively charged or negatively charged, in the absence of solvent molecules. The confining capacity of the ion trap arises from the formation of a trapping potential well when appropriate potentials are applied to the electrodes of the ion trap. As an ion storage device, the ion trap permits the study of gas-phase ion chemistry and the dissociation of ions for elucidation of ion structures. When stages of mass selectivity are employed repeatedly, the ion trap functions as a tandem mass spectrometer.

The ion trap functions as a mass spectrometer when the field within the device is changed, so that the trajectories of simultaneously trapped ions of consecutive specific mass/charge ratio (m/z) become sequentially unstable, and ions leave the trapping field in order of mass/charge ratio. Upon ejection from the ion trap, ions strike a detector and provide an output signal. This relatively simple method of mass-selective operation of the ion trap has led to a revolution in MS. It is estimated that some 4500 ion trap instruments have been sold thus far at a total cost of about a quarter of a billion US dollars; the volume of these sales has led to a reorganization of the MS industry. The combination of a quadrupole ion trap interfaced with a gas chromatograph is now available commercially at a price which permits the acquisition of these instruments by most academic departments of chemistry; thus these instruments are now becoming accessible to relatively large numbers of students, both graduate and undergraduate.

With the advent of new methods by which ions can be formed in the gas phase from polar as well as covalent molecules and introduced subsequently into an ion trap, the range of applications of the quadrupole ion trap is now considerable. The coupling of liquid chromatography (LC) with ESI and with MS in the early 1980s, together with the rapid advancement in ion trap technology, have led to the development of new ion trap instruments for the analysis of nonvolatile, polar, and thermally labile compounds. In 1995, new ion trap instruments (Finnigan's LCQ™ and GCQ™, and Bruker-Franzen's ESQUIRE™) were introduced which employ external ion sources with injection of externally generated ions into the ion trap. The major focus for the application of these new instruments, using LC/ESI/MS (liquid chromatography/electrospray ionization mass spectrometry), has been the analysis of high-molecular-weight biopolymers such as proteins, peptides, and oligodeoxyribonucleotides.

2 STRUCTURE OF THE QUADRUPOLE ION TRAP MASS SPECTROMETER

The quadrupole ion trap consists essentially of three electrodes which are shown in open array in Figure 1. Two

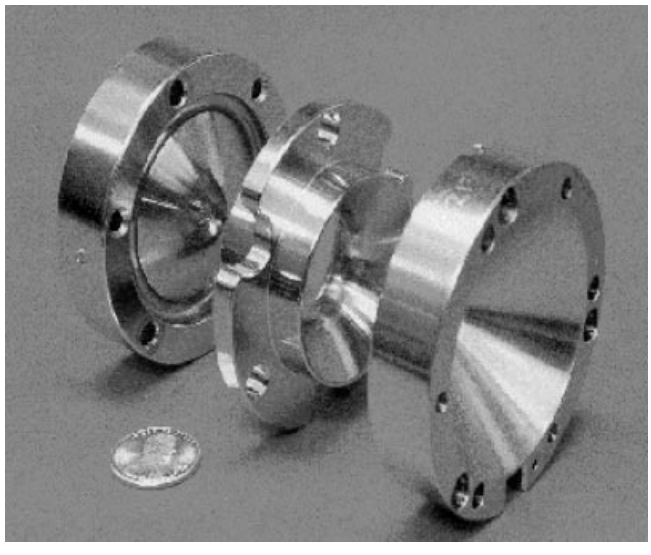

Figure 1 Three electrodes of the quadrupole ion trap shown in open array.

of the three electrodes are virtually identical and, while having hyperboloidal geometry, resemble small inverted saucers; these saucers are called end-cap electrodes and in Figure 1 are distinguishable by the number of holes in the center of each electrode. One end-cap electrode has a single small central aperture through which electrons and/or ions can be gated periodically while the other has several small apertures arranged centrally also and through which ions pass to a detector; in ion trap instruments with external ion sources, each end-cap electrode has a single perforation. The third electrode is also of hyperboloidal geometry (of two sheets, in this case, rather than one) and is called the ring electrode; it not only resembles a serviette or napkin holder but is of roughly the same size in that the radius in the central plane, r_0, of the ring electrode is approximately 1 cm. The ring electrode is positioned symmetrically between the two end-cap electrodes as shown in Figure 2; Figure 2(a) shows a photograph of an ion trap cut in half along the axis of cylindrical symmetry, while Figure 2(b) is a cross-section of an ideal ion trap showing the asymptotes and the dimensions r_0 and z_0, where $2z_0$ is the separation of the two end-cap electrodes measured along the axis of the ion trap.

The electrodes in Figure 2 are necessarily truncated for practical purposes but, in theory, they extend to infinity and meet the asymptotes shown in Figure 2. The asymptotes arise from the hyperboloidal geometries of the three electrodes and, in the ideal case, are inclined at an angle of $53^\circ\,34'$ to the cylindrical axis (the z-axis) of the electrode arrangement. The geometries of the electrodes have been chosen so as to produce a near ideal quadrupole field when an rf potential is applied to the ring electrode and the two end-cap electrodes are grounded. In a quadrupole field, the forces acting on an ion depend linearly upon the displacement, x, y, or z from the ion trap center.

In Figure 3 is shown a representation of the instantaneous saddle-shaped potential surface generated by the application of an rf potential of, say, 1000 V_{0-p} (zero-to-peak) to the ring electrode; the open circle represents an ion on this surface. Because of the cylindrical symmetry of the ion trap, the x- and y-components of the field are often combined to yield a single radial r-component using $x^2 + y^2 = r^2$. As the rf potential oscillates sinusoidally with a frequency of the order of 1 MHz, the field is periodically reversed; this situation can be envisaged as Figure 3 inverted with, say, $-1000\,V_{0-p}$ applied to the ring electrode and the end-cap electrodes grounded. The ion in Figure 3 rolls down the slope and approaches the ion trap center radially, that is, the ion is being focused towards the center; the ion may continue to roll downhill and moves axially away from the center, that is, the ion is being defocused. Upon reversal of the field in Figure 3, the ion is defocused radially and focused axially. When the focusing and defocusing forces are balanced by the imposition of appropriate conditions, the ion can remain confined on the surface as

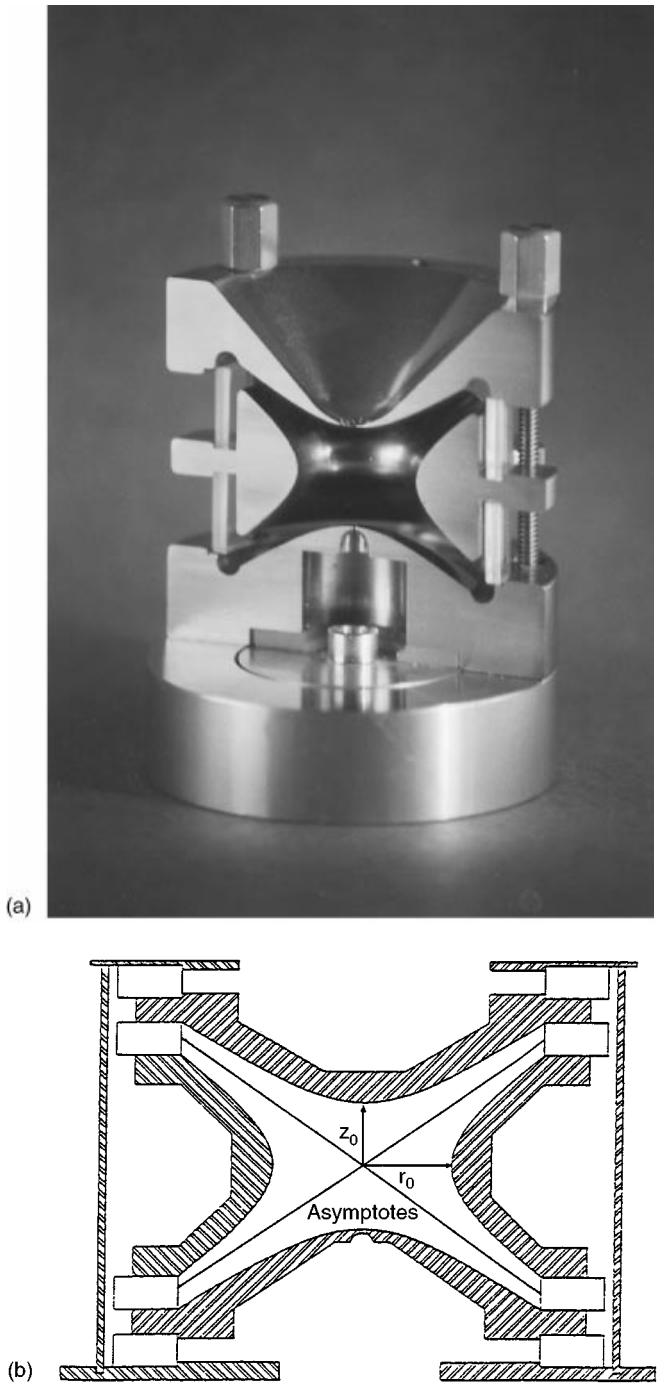

Figure 2 Quadrupole ion trap; (a) photograph of an ion trap cut in half along the axis of cylindrical symmetry; (b) schematic diagram of the three-dimensional ideal ion trap showing the asymptotes and the dimensions r_0 and z_0.

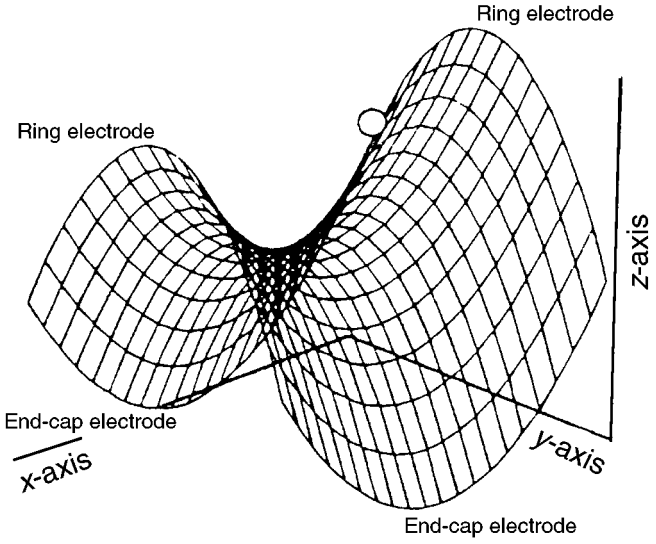

Figure 3 Instantaneous pure quadrupole field, or potential surface, for a quadrupole ion trap. Note the four poles of the surface and the similarity of the field shape to the trajectory in Figure 6. The open circle represents an ion on the potential surface.

though it were trapped in a pseudo-potential well. Thus, a trapping pseudo-potential well, parabolic in cross-section, can be created when an rf potential is applied to the ring electrode and the two end-cap electrodes are grounded.

The trapping pseudo-potential well is shown in Figure 4; the potential well in the axial direction is of depth $\overline{D}_z$ while in the radial direction the depth is $\overline{D}_r$. The value of $\overline{D}_z$ is about twice that of $\overline{D}_r$ so that the potential well resembles more a flower vase than it does a bowl. The depth of the trapping pseudo-potential well represents a measure of the kinetic energy that is required in order for an ion to escape the well; for an ion of m/z 200, under the conditions of Figure 3, the kinetic energy required is ca. 58 eV.

In the ion trap, the trajectory or path of an ion is characterized by two frequencies, or secular frequencies,

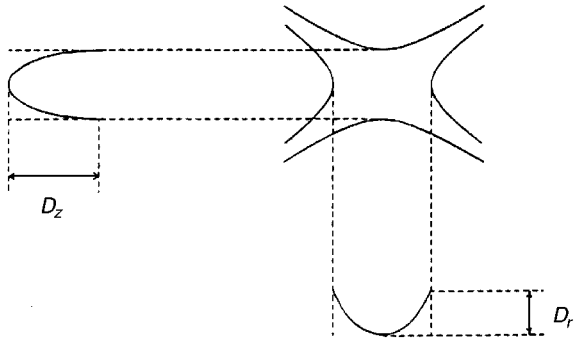

Figure 4 Representation of the parabolic trapping potential wells of depths D_z and D_r.

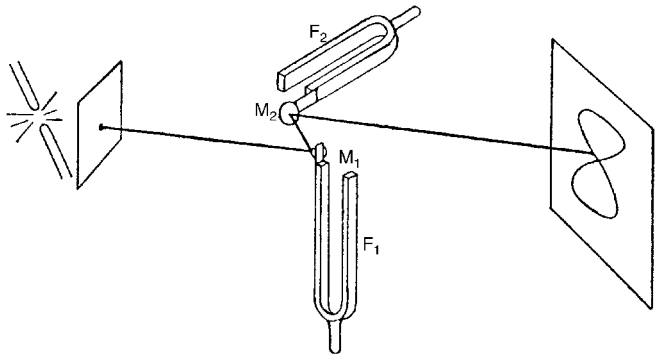

Figure 5 Apparatus for producing Lissajous figures.

that arise from the axial and radial motions of the ion induced by the trapping field. At a given instant of time, an ion is focused axially towards the center of the ion trap while it is simultaneously defocused radially so that it moves away from the ion trap center; some half a microsecond later the situation is reversed, and so on. The resultant trajectory from these axial and radial oscillations resembles a figure-of-eight; such a figure is known as a Lissajous figure. A Lissajous figure can be demonstrated readily with a light beam reflected in turn from each of two mirrors, M_1 and M_2 attached respectively to two tuning forks, F_1 and F_2, arranged orthogonally as shown in Figure 5. In the case of the ion trajectory, there is a high-frequency ripple superimposed on the Lissajous figure; the ripple is due to the rf potential that oscillates at 1 MHz.

In Figure 6(a) is shown the trajectory of a single ion in an ion trap; an enlargement of this trajectory is shown in Figure 6(b) where the Lissajous-type motion and the high-frequency ripple are seen clearly. The projection of the trajectory onto the $x-y$ plane forms a straight line as shown in Figure 6(b); this line shows that the ion motion is restricted to a plane.

Mass-selective ejection of ions from the potential well within the ion trap is accomplished by ramping in a linear fashion the amplitude of a rf potential applied to one of the ion trap electrodes; each ion species is ejected from the potential well at a specific rf amplitude and, because the initial amplitude and ramping rate are known, the mass/charge ratio can be determined for each ion species upon ejection. This relatively simple method for measuring the mass/charge ratios of confined ions upon ejection was developed by Stafford et al.[2] and is known as the "mass-selective axial instability mode"; this method made possible the commercialization of the quadrupole ion trap around 1984. A prerequisite of this method of mass-selective ion ejection is that ions be herded initially to the center of the ion trap under the action of momentum-dissipating collisions; helium atoms are used for this purpose.

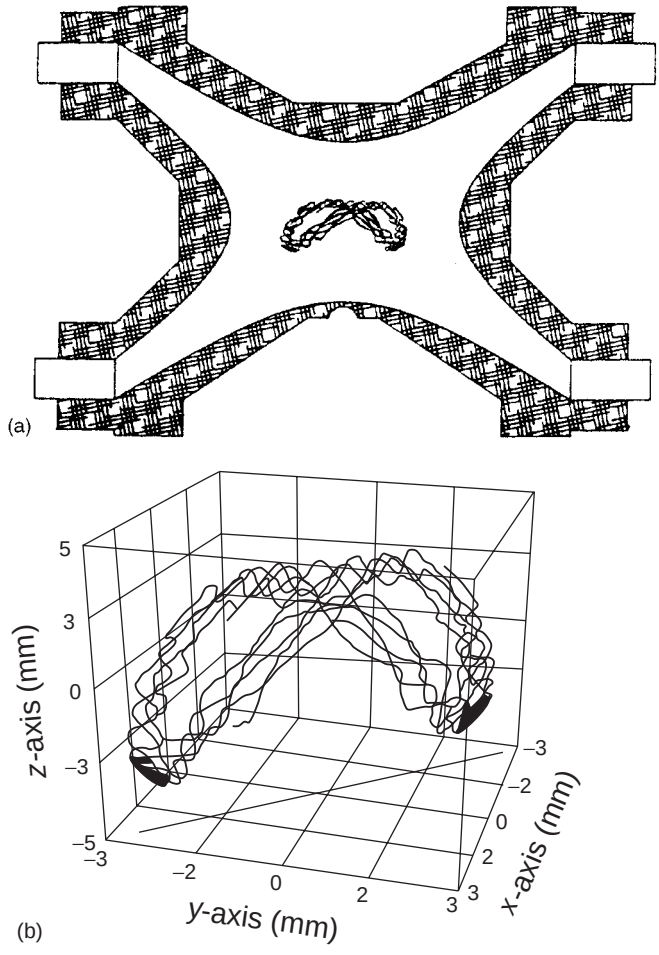

Figure 6 (a) Trajectory of a trapped ion of m/z 105 in a quadrupole ion trap. (b) Expanded view of the ion trajectory showing the magnitude of the trajectory. The projection of the trajectory onto the $x-y$ plane forms a straight line and illustrates planar motion in three-dimensional space. The trajectory develops a shape that resembles a flattened boomerang. Taken from Nappi et al.[1]

3 LITERATURE

The quadrupole ion trap and the related quadrupole mass filter were invented by Paul and Steinwedel[3] in 1960 and their pioneering work was recognized by the award of a 1989 Nobel Prize in Physics to Wolfgang Paul.[4] Yet the basis of the theory of operation of quadrupole devices was laid down about 140 years ago by Mathieu to whose work we shall return. In this presentation, references are given to those works that are regarded as landmarks in the field of ion trap MS and to those publications that have been used to illustrate specific modes of operation of the ion trap.

Detailed accounts of the early development of the quadrupole-type devices as mass spectrometers were published by Dawson and Whetten[5] and by Dawson;[6] the latter publication has now been reprinted as part of the series of "*American Vacuum Society Classics*" by the American Institute of Physics under ISBN 1563964554. A full treatment of ion trap theory can be found in the now standard but somewhat dated text by March, Hughes and Todd;[7] the historical account in this text by Todd has been expanded into a full-scale review.[8] Other reviews on specific topics have been contributed by Cooks, et al.[9,10] and a special collection of papers reporting upon recent developments has appeared also.[11]

The field of quadrupole ion trap MS was reviewed extensively for the 12th International Mass Spectrometry Conference,[12] held in Amsterdam in 1991. Since that time, three volumes have been published in the CRC Series on *Modern Mass Spectrometry*.[13] Volume 1 of this series covers the history of the quadrupole ion trap, nonlinear ion traps, ion activation, ion/molecule reactions and ion trajectory simulations; the reader is referred to Chapter 2 of Volume 1 for a detailed exposition of the mathematical basis of the operation of the ion trap. Volume 2 deals with enhancement of ion trap performance, ion trap confinement of externally generated ions, ion structure differentiation, ion photodissociation, lasers and the ion trap, and ion traps in the study of physics. Volume 3 includes a review of fundamentals in addition to extensive expositions on gas chromatography/ion trap tandem mass spectrometry (GC/MS/MS) and liquid chromatography/ion trap tandem mass spectrometry (LC/MS/MS); examples of applications in each of the three areas above are given. In all, the above three volumes contain 30 chapters that originated from 18 of the leading ion trap research laboratories. An introduction to the quadrupole ion trap written by this author in a tutorial form has appeared.[14]

4 TRAPPING PSEUDO-POTENTIAL WELL

In a quadrupole ion trap, ions are focused by collisions with helium buffer gas towards the center of the ion trap. This focusing process imposes a degree of order upon the ions much like that of a flock of sheep that has been shepherded together by sheepdogs. Ions of lowest mass/charge ratio have the smallest excursions from the center of the ion trap so that they reside closest to the ion trap center; since these ions have the least kinetic energy, they populate the lowest region in the potential well. The ions of next higher mass/charge ratio are arranged in the ion trap about the ions of lowest mass/charge ratio and populate the next higher region of the potential well, and so on. Under these conditions, trapped ion species are arranged about the ion trap center rather like the layers

of an onion with the ions of lowest mass/charge ratio residing in the center of the onion. Mass-selective ion ejection in order of increasing mass/charge ratio requires that ions of lowest mass/charge ratio be ejected first so that the mass-selective ion ejection process is akin to removing the layers of the onion, one by one, starting with the layer nearest the onion center.

When ions of lowest mass/charge ratio are extracted from the center of the onion, they experience space charge perturbation while passing through the layers of ions of higher mass/charge ratio, become spatially dispersed and, upon ejection, produce ion signals with poor mass resolution. To obviate this problem, ions of lowest mass/charge ratio are made to come into resonance with a fixed frequency applied across the end-cap electrodes. Ions come into resonance as the rf amplitude is ramped, that is, as the amplitude is increased. As the ions become excited resonantly in the axial direction, they emerge from the onion and experience a brief period free of space charge immediately prior to ejection. During ejection, the ions remain clustered together and, upon impacting on the detector, produce ion signals of higher mass resolution. This process of ion axial excitation prior to ion ejection is termed axial modulation.

Let us consider an analogy to this trapping, focusing, and ejection process. The trapping pseudo-potential well created within the electrode assembly of the quadrupole ion trap can be likened to a bowl or glass of parabolic cross-section; ion species are confined in layers in the bowl rather like an exotic drink of several liqueurs arranged carefully and tastefully in horizontal layers according to their density, as shown in Figure 7. The liqueur of greatest density represents the ions of lowest mass/charge ratio. The tilting of the bowl or the lowering of the side of the bowl corresponds to the ramping of the rf amplitude; the liqueur glass in Figure 7 corresponds to the detector. The removal of ions from near the bottom of the potential well is effected with the straw in Figure 7. The tilting of the bowl is analogous to ramping the rf amplitude while the use of a straw is analogous to axial modulation in which ions are brought successively into resonance with an applied AC (alternating current) frequency, as is discussed later. As the liqueurs are drawn up the straw in order of decreasing density, so the ions are ejected in order of increasing mass/charge ratio and impinge upon a detector. The signals from the detector create a mass spectrum of the contents of the ion trap.

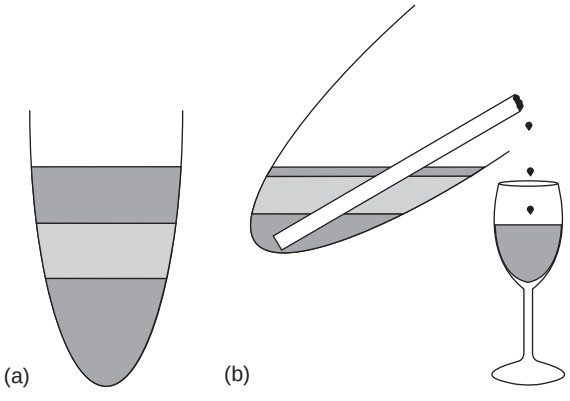

Figure 7 Schematic presentation of (a) a trapping parabolic potential well where the three liquids differing in density represent ions differing in mass/charge ratio; (b) the tilting of the well corresponds to ramping of the rf potential while the straw, with which ions are withdrawn in order of increasing mass/charge ratio, represents axial modulation.

5 STRUCTURE OF THE QUADRUPOLE ION TRAP

The quadrupole ion trap is composed of three electrodes with the ring electrode located symmetrically between two end-cap electrodes as shown in Figure 2. The shape of the ring electrode is given by Equation (1)

$$\frac{r^2}{r_0^2} - \frac{z^2}{z_0^2} = 1 \tag{1}$$

while the shapes of the end-cap electrodes are given by Equation (2)

$$\frac{r^2}{r_0^2} - \frac{z^2}{z_0^2} = -1 \tag{2}$$

For an ideal quadrupole field, the following identity is given (usually with more authority than truth)

$$r_0^2 = 2z_0^2 \tag{3}$$

so that, once the magnitude of r_0 is given, the sizes of all three electrodes and the spacing between the electrodes are fixed. However, it has been pointed out by Knight[15] that, contrary to Equation (3), the ratio of $z_0^2 : r_0^2$ is not necessarily restricted to 2. Regardless of the value of this ratio, the size of the ion trap, is determined largely by the magnitude of r_0 and, in the majority of commercial ion traps in use today, r_0 is either 1.00 cm or 0.707 cm.

6 THEORY OF QUADRUPOLE ION TRAP OPERATION

The motions of ions in quadrupole devices differs greatly from the straight lines and arcing curves of ions in field-free regions and in magnetic and electrostatic sectors, respectively, familiar to those conversant with sector mass spectrometers. The quadrupole ion trap and

the quadrupole mass filter or analyzer are described as *dynamic* instruments since ion trajectories in these instruments are influenced by a set of *time-dependent forces* (which render the trajectories mathematically more difficult to predict compared with sector instruments). Sector instruments are described as *static* devices in that the field is maintained at a constant value for transmission of an ion. In quadrupole instruments, a quadrupole field is established when a potential is applied to electrodes that have a hyperbolic geometric form. Let us examine the movement of charged particles in a quadrupole field by considering first the forces acting on a single ion within a quadrupole field.

6.1 An Ion in a Quadrupole Field

An ion, positively charged or negatively charged, in a quadrupole field experiences strong focusing in that the restoring force, which drives the ion back towards the center of the device, increases as the ion deviates from the center of the device. The motion of ions in a quadrupole field can be described mathematically by the solutions to the second-order linear differential equation described originally by Mathieu;[16] this equation is known as the Mathieu equation. From Mathieu's investigation of the mathematics of vibrating stretched skins, he was able to describe solutions in terms of regions of stability and instability; these solutions and the criteria for stability and instability have been used to describe the trajectories of ions confined in quadrupole devices and to define the limits to trajectory stability. In order to adopt the solutions to the Mathieu equation, we must verify that the equation of motion of an ion confined in a quadrupole device can be described by the Mathieu equation. The path that is followed here concerns the expression for a force (mass × acceleration) in Mathieu's equation, and comparison of that expression with one for the force on an ion in a quadrupole field. This comparison is laid out below in simple mathematical terms; thus it is possible to express the magnitudes and frequencies of the potentials applied to ion trap electrodes, the size of the ion trap, and the mass/charge ratio of ions confined therein in terms of Mathieu's dimensionless parameters, a_u and q_u. On this basis, we shall adopt the idea of stability regions in a_u, q_u space in order to discuss the trapping, and limits thereto, of gaseous ions in quadrupole devices.

6.2 The Mathieu Equation

The canonical or commonly accepted form of the Mathieu equation is, Equation (4):

$$\frac{\mathrm{d}^2u}{\mathrm{d}\xi^2} + (a_u - 2q_u \cos 2\xi)u = 0 \tag{4}$$

where u represents the coordinate axes x, y and z, ξ is a dimensionless parameter equal to $\Omega t/2$ such that Ω must be a frequency as t is time, and a_u and q_u are additional dimensionless parameters known as trapping or stability parameters. The introduction of Ω here is not entirely serendipitous since it will reappear as the radial frequency (in rad s^{-1}) of the rf potential applied to the ring electrode. Now it can be shown by substituting $\xi = \Omega t/2$ (from Equation 4) that

$$\frac{\mathrm{d}^2u}{\mathrm{d}t^2} = \frac{\Omega^2}{4}\frac{\mathrm{d}^2u}{\mathrm{d}\xi^2} \tag{5}$$

Substitution of Equation (5) into Equation (4), multiplying throughout by m and rearranging yields

$$m\frac{\mathrm{d}^2u}{\mathrm{d}t^2} = \frac{-m\Omega^2}{4}(a_u - 2q_u \cos \Omega t)u \tag{6}$$

We note that the left side of Equation (6) can represent the force on an ion, that is, mass times acceleration in each of the x, y and z directions.

Now the field in quadrupole devices is uncoupled so that the forces in the three coordinate directions may be determined separately. Let us then consider the force in the x-direction, F_x, experienced by an ion of mass m and charge e at any point within a quadrupole field, Equation (7)

$$F_x = ma = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -e\frac{\partial\phi}{\partial x} \tag{7}$$

where a is the acceleration of the ion, e is the electronic charge and ϕ is the potential at any point (x,y,z) within the field. Similar expressions for F_y and F_z can be obtained. It should be noted that Equation (7) relates the force on an ion to the field within the ion trap. The quadrupole potential ϕ can be expressed as

$$\phi = \frac{\phi_0}{r_0^2}(\lambda x^2 + \sigma y^2 + \gamma z^2) \tag{8}$$

where ϕ_0 is the applied electric potential (which we shall see later is an rf potential either alone or in combination with a direct current (DC) potential), λ, σ and γ are weighting constants for the x, y, and z coordinates, respectively, and r_0 is a constant which is defined separately depending on whether the quadrupole device is an ion trap or mass filter. It can be seen from Equation (8) that the potential increases quadratically with x, y, and z. In any electric field, it is essential that the Laplace condition, which requires that the second differential of the potential at a point be equal to zero, be satisfied; the Laplace condition ensures that the field in the x, y, and z directions is linear and does not change. When that is done, it is found that, Equation (9)

$$\lambda + \sigma + \gamma = 0 \tag{9}$$

For the ion trap, $\lambda = \sigma = 1$ and $\gamma = -2$, whereas for the quadrupole mass filter $\lambda = -\sigma = 1$ and $\gamma = 0$. Substituting the values $\lambda = \sigma = 1$ and $\gamma = -2$ into Equation (8), we obtain Equation (10) for the potential at any point within the quadrupole field in a quadrupole ion trap.

$$\phi_{x,y,z} = \frac{\phi_0}{r_0^2}(x^2 + y^2 - 2z^2) \tag{10}$$

This equation can be transformed into cylindrical coordinates by employing the standard transformations $x = r\cos\theta, y = r\sin\theta, z = z$. Thus Equation (10) becomes Equation (11)

$$\phi_{r,z} = \frac{\phi_0}{r_0^2}(r^2\cos^2\theta + r^2\sin^2\theta - 2z^2) \tag{11}$$

When we apply the trigonometric identity $\cos^2 + \sin^2 = 1$, we obtain Equation (12)

$$\phi_{r,z} = \frac{\phi_0}{r_0^2}(r^2 - 2z^2) \tag{12}$$

The applied electric potential, ϕ_0, (that is, applied to the ring electrode) is either an rf potential $V\cos\Omega t$ or a combination of a DC potential, U, of the form, Equation (13)

$$\phi_0 = U + V\cos\Omega t \tag{13}$$

where Ω is the angular frequency (in rad s^{-1}) of the rf field. Note that Ω is equal to $2\pi f$, where f is the frequency in hertz.

When the expression for ϕ_0 as given by Equation (13) and $\lambda = 1$ are substituted into Equation (8) and ϕ is differentiated with respect to x, Equation (14) is obtained for the potential gradient

$$\frac{\partial\phi}{\partial x} = \frac{2x}{r_0^2}(U + V\cos\Omega t) \tag{14}$$

Substitution of Equation (14) into Equation (7) yields an expression for the force on an ion, Equation (15)

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = \frac{-2e}{r_0^2}(U + V\cos\Omega t)x \tag{15}$$

We can now compare directly the terms on the right hand sides of Equations (6) and (15), recalling that u represents x, to obtain, Equation (16)

$$a_x = \frac{8eU}{mr_0^2\Omega^2}; \qquad q_x = \frac{-4eV}{mr_0^2\Omega^2} \tag{16}$$

When this derivation is repeated to obtain the force on an ion in the y-direction in a quadrupole mass filter, one finds that $q_x = -q_y$; this relationship is obtained since $\lambda = -\sigma = 1$. For the quadrupole ion trap, it is found that $q_x = q_y$; this equality is found since $\lambda = \sigma = 1$. It is suggested that it would be instructive for the reader to derive similarly the a_z and q_z trapping parameters for positively charged ions (Equation 17) which are used frequently in discussions of the stability diagram of the quadrupole ion trap; in this case, $\lambda = \sigma = 1$ and $\gamma = -2$. For negatively charged ions, the signs of the a_z and q_z trapping parameters in Equation (17) are reversed:

$$a_z = \frac{-8eU}{mr_0^2\Omega^2}; \qquad q_z = \frac{4eV}{mr_0^2\Omega^2} \tag{17}$$

Let us ignore a_z (which is proportional to U, a DC potential) for the present since most commercial ion trap instruments do not offer the flexibility of applying a DC potential to the electrodes; therefore, a_z is held equal to zero such that the most common mode of ion trap operation is said to correspond to operation on the q_z axis. The expression for q_z in Equation (17) contains the mass/charge ratio for a given ion, the size of the ion trap, r_0, the amplitude V of the rf potential and the radial frequency Ω; these parameters are all that we shall need in order to understand the various operations of the ion trap. The solutions to the Mathieu equation can be calculated in terms of a_z and q_z and these solutions can be interpreted in terms of trajectory stability (and instability) in each of the x, y (or r), and z directions. When the confinement conditions correspond simultaneously to trajectory stability in both r- and z-directions, a charged particle may be stored. The trajectories of a charged particle are characterized by the fundamental secular frequencies of ion motion in the radial, r, and axial, z, directions.

6.3 Potentials on the Electrodes

The potentials on the ring and end-cap electrodes may be verified in the following manner with reference to Figure 2 and Equation (12). Consider the intersection of the central radial plane with the surface of the ring electrode, such that $z = 0$ and $r = r_0$; the potential at the ring electrode is given by Equation (18)

$$\phi_{r_0,0} = \frac{\phi_0}{r_0^2}r_0^2 = \phi_0 \tag{18}$$

Now consider the intersection of the central axis of cylindrical symmetry with the surface of either end-cap electrode, such that $r = 0$ and $z = z_0$; recalling the identity of Equation (3) and the potential of Equation (12), the potential at each end-cap electrode is given by Equation (19)

$$\phi_{0,z_0} = \frac{\phi_0}{r_0^2}(-2z_0^2) = -\phi_0 \tag{19}$$

However, no commercial quadrupole ion trap is operated in this fashion; rather, the end-cap electrodes are held at ground potential (except for the imposition of oscillating potentials of low amplitude, hundreds of millivolts to a few volts). The net effect of applying ϕ_0 to the ring electrode and grounding the end-cap electrodes is to halve the mass range of the ion trap as a mass spectrometer.

In order to verify the potentials on the ring and end-cap electrodes in the commercial ion trap, an alternative equation to Equation (12) must be used; this equation is

$$\phi_{r.z} = \frac{\phi_0(r^2 - 2z^2)}{2r_0^2} + c \tag{20}$$

where c is a constant. The ring electrode potential (Equation (20) with $z = 0$, $r = r_0$) is given by Equation (21)

$$\phi_{r_0,0} = \frac{\phi_0 r_0^2}{2r_0^2} + c = \phi_0 \tag{21}$$

from which we obtain $c = \phi_0/2$. The potential at the end-cap electrodes ($r = 0$ and $z = z_0$) is given by Equation (22)

$$\phi_{0,z_0} = \frac{-2\phi_0 z_0^2}{2r_0^2} + \frac{\phi_0}{2} = 0 \tag{22}$$

Hence Equation (20) reduces to Equation (23)

$$\phi_{r,z} = \frac{\phi_0(r^2 - 2z^2)}{2r_0^2} + \frac{\phi_0}{2} \tag{23}$$

The constant term does not change the equations of motion derived from the partial differentials, but the potential along the asymptotes of the hyperbolas is changed. It should be noted that an ion at the center of a commercial ion trap experiences a potential of $\phi_0/2$ and "sees" a potential of $-\phi_0/2$ on the end-cap electrodes and a potential of $\phi_0/2$ on the ring electrode.

6.4 Stretched Ion Trap

As we discussed above, the electrodes of the ion trap are truncated in order to obtain a practical working instrument but this truncation introduces higher order multipole components to the potential as shown in Equation (24)

$$\phi_{r,z} = C_0^0 + C_1^0 z + C_2^0\left(\tfrac{1}{2}r^2 - z^2\right) + C_3^0 z\left(\tfrac{3}{2}r^2 - z^2\right) + C_4^0\left(\tfrac{3}{8}r^4 - 3r^2z^2 + z^4\right) + \cdots \tag{24}$$

The C_n^0 coefficients, where $n = 0, 1, 2, 3$, and 4, correspond to monopole, dipole, quadrupole, hexapole and octopole components, respectively, of the potential $\phi_{r,z}$. For the pure quadrupole ion trap, only the coefficients corresponding to $n = 0$ and $n = 2$ are nonzero.

In order to compensate for these higher order multipole components, the electrodes of most commercial instruments were assembled in such a way that the distance between the end-cap electrodes was increased or "stretched"; the value of z_0 was increased by 10.6%. However, there was no corresponding modification of the shapes of the electrodes which would be required in order to maintain a purely quadrupolar geometry.

The immediate consequences are that the asymptotes to the end-cap electrodes no longer coincide with those for the ring electrode. Furthermore, $r_0^2 \neq 2z_0^2$. In order to compensate, in part, for the stretching of the ion trap, the trapping parameters are now calculated using the actual values of z_0 and r_0, as follows, Equation (25)

$$a_r = \frac{8eU}{m(r_0^2 + 2z_0^2)\Omega^2}; \quad q_r = \frac{-4eV}{m(r_0^2 + 2z_0^2)\Omega^2} \tag{25}$$

and

$$a_z = \frac{-16eU}{m(r_0^2 + 2z_0^2)\Omega^2}; \quad q_z = \frac{8eV}{m(r_0^2 + 2z_0^2)\Omega^2} \tag{26}$$

When $r_0^2 = 2z_0^2$ (Equation 3) is substituted into Equation (26), we obtain the trapping parameters given in Equation (17). It should be noted that for the ion trap in the LCQ™ and GCQ™ instruments, $r_0 = 0.707$ cm and $z_0 = 0.785$ cm such that the geometry has been stretched by ca. 57%.

7 REGIONS OF ION TRAJECTORY STABILITY

The operation of the quadrupole ion trap is concerned with the criteria that govern the stability (or instability) of the trajectory of an ion in the ion trap, that is, the experimental conditions that determine whether an ion is stored within the device or whether it is ejected and whether it is excited resonantly or not. The experimental conditions for the manipulation of ion trajectories are represented mathematically by the solutions to Mathieu's equation.

The solutions to Mathieu's equation are of two types: (i) ion motion is periodic but unstable, and (ii) ion motion is periodic and stable. Solutions of type (i) are called Mathieu functions of integral order and form the boundaries of unstable regions on the stability diagram. The boundaries, which are referred to as characteristic curves or characteristic values, correspond to the values of a new trapping parameter, β_z, that is, $0, 1, 2, 3, \ldots$; β_z is a complex function of a_z and q_z to which we shall return. The boundaries represent, in practical terms, the point at which the trajectory of an ion becomes unbounded.

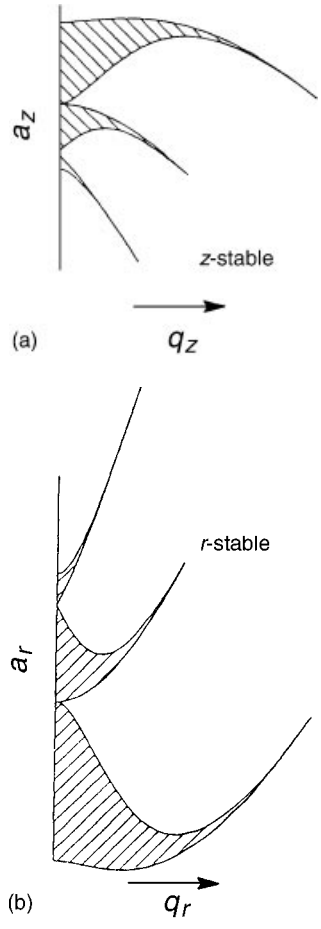

Figure 8 Several Mathieu stability regions for the three-dimensional quadrupole field. (a) diagrams for the z-direction of (a_z, q_z) space; (b) diagrams for the r-direction of (a_z, q_z) space.

Solutions of type (ii) determine the motion of ions in an ion trap. The stability regions corresponding to stable solutions of the Mathieu equation in the z-direction are shaded and labeled "z-stable" in Figure 8(a). The stability regions corresponding to stable solutions of the Mathieu equation in the r-direction are shaded and labeled "r-stable" in Figure 8(b); it can be seen that they are doubled in magnitude along the ordinate and inverted. It is seen from Equations (25) and (26) that $a_z = -2a_r$ and $q_z = -2q_r$, that is, the stability parameters for the r- and z-directions differ by a factor of -2.

7.1 Stability Region

Ions can be stored in the ion trap provided that their trajectories are stable in the r- and z-directions simultaneously; such trajectory stability is obtained in the region closest to the origin, that is, region A in Figure 9 which is plotted in a_u, q_u space, that is, where a_u is plotted against q_u. Regions A and B are referred to as *stability regions*; region A is of the greatest importance at this time

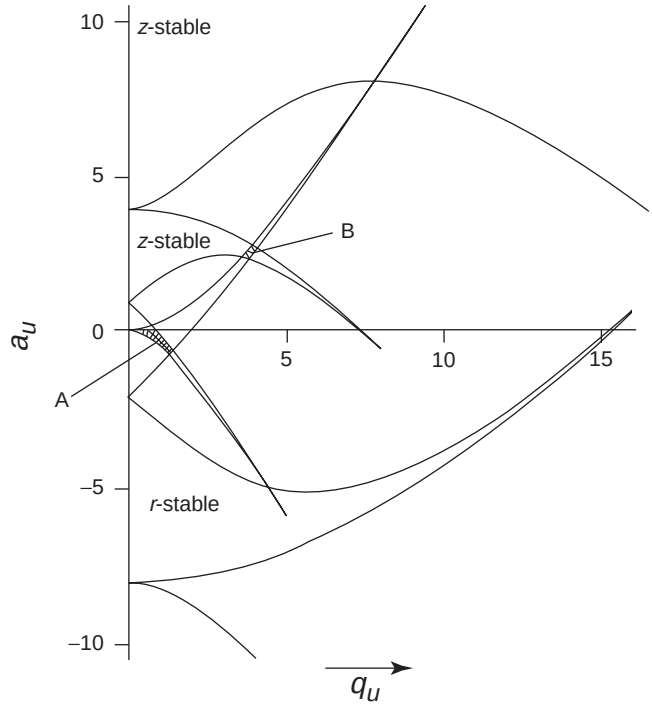

Figure 9 Mathieu stability diagram in (a_z, q_z) space for the quadrupole ion trap in both the r- and z-directions. Regions of simultaneous overlap are labeled A and B.

(region B remains to be explored) and is shown in greater detail in Figure 10. The coordinates of the stability region in Figure 10 are the Mathieu parameters a_z and q_z. Here, we plot a_z versus q_z rather than a_u versus q_u which is unnecessarily confusing because $u = r, z$. In Figure 10, the $\beta_z = 1$ stability boundary intersects with the q_z axis at $q_z = 0.908$; this working point is that of the ion of lowest mass/charge ratio (that is, low-mass cut-off (LMCO), as discussed below) which can be stored in the ion trap. The stability diagram for negatively charged ions is the mirror image about the q_z-axis of the stability diagram in Figure 10.

7.2 Secular Frequencies

A three-dimensional representation of an ion trajectory in the ion trap, as shown in Figure 6, has the general appearance of a Lissajous curve or figure-of-eight composed of two fundamental frequency components, radial and axial, respectively, $w_{r,0}$ and $w_{z,0}$ of the secular motion.[1] The description of "fundamental" infers that there exist other higher order (n) frequencies and that the entire family of frequencies is described by $w_{r,n}$ and $w_{z,n}$. These secular frequencies are given by Equations (27) and (28):

$$w_{u,n} = \left(n + \tfrac{1}{2}\beta_u\right)\Omega, \qquad 0 \le n < \infty \tag{27}$$

$$w_{u,n} = -\left(n + \tfrac{1}{2}\beta_u\right)\Omega, \qquad -\infty < n < 0 \tag{28}$$

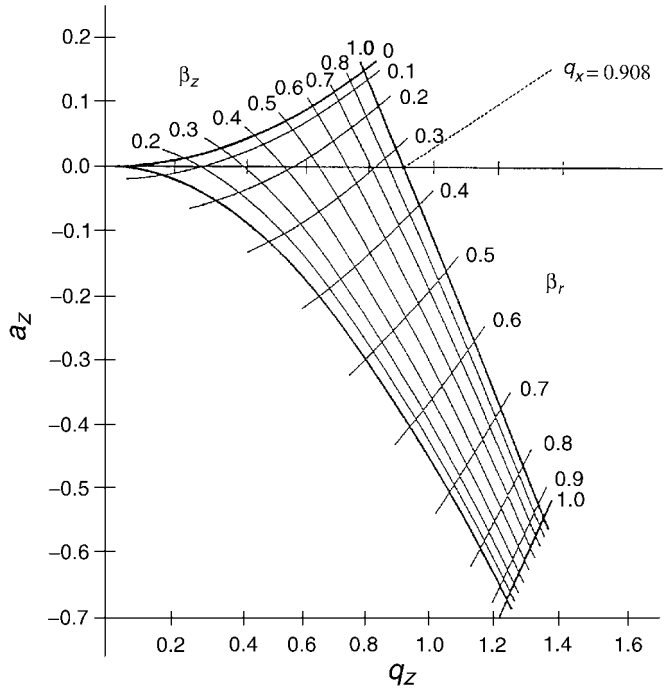

Figure 10 Stability diagram in (a_z, q_z) space for the region of simultaneous stability in both the r- and z-directions near the origin for the three-dimensional quadrupole ion trap; the iso-β_r and iso-β_z lines are shown in the diagram. The q_z-axis intersects the $\beta_z = 1$ boundary at $q_z = 0.908$, which corresponds to $q_{\max}$ in the mass-selective instability mode.

where

$$\beta_u \approx \sqrt{\left(a_u + \frac{q_u^2}{2}\right.} \tag{29}$$

for $q_r < 0.2$ and $q_z < 0.4$. It should be noted that while the fundamental axial secular frequency, $w_{z,0}$, is usually given in units of hertz in the literature and referred to simply as w_z, it should be given in radians per second. At this time, the higher-order frequencies are of little practical significance.

It should be noted further that the definition of β_z given in Equation (29) above is only an approximation, known as the Dehmelt approximation after Hans Dehmelt who shared the 1989 Nobel Prize in Physics along with Norman Ramsey and Wolfgang Paul; β_u is defined precisely by a continued fraction expression in terms of a_u and q_u, as shown in Equation (30)

$$\beta_u^2 = a_u + (q_u^2) \quad \left((\beta_u + 2)^2 - a_u - (q_u^2) \quad (\beta_u^2 + 4)^2 \right.$$
$$- a_u - (q_u^2) \,/\, \left((\beta_u + 6)^2 - a_u - \cdots\right)$$
$$+ (q_u^2) \quad \left((\beta_u - 2)^2 - a_u - (q_u^2) \quad (\beta_u - 4)^2 \right.$$
$$- a_u - (q_u^2) \,/\, \left((\beta_u - 6)^2 - a_u - \cdots\right) \tag{30}$$

The resemblance of the simulated ion trajectory shown in Figure 6 to a roller coaster ride is due to the motion of an ion on the potential surface shown in Figure 3. The oscillatory motion of the ion results from the undulations of the potential surface that can be envisaged as rotation of the potential surface. The simulation of the ion trajectory was carried out using the ITSIM (ion trajectory simulation program),[17] while the potential surface was generated[18] from Equation (24) by calculating $\phi_{r,z}$ for $C_2^0 = 1$ and all of the other coefficients equal to zero for increment steps of 1 mm in both radial and axial directions.

8 RESONANT EXCITATION

As the motion of ions confined in a quadrupole ion trap is characterized by two secular frequencies, axial and radial, ion motion can be excited upon resonant irradiation at either or both of these frequencies. Such irradiation can be effected by applying a small supplementary oscillating potential of some hundreds of millivolts across the end-cap electrodes, that is, in dipolar mode. Resonant excitation, using the axial secular frequencies of confined ions, has become a powerful technique in quadrupole ion trap MS due to the utilization of predetermined waveforms composed of specified frequencies or frequency ranges. Prior to resonant excitation of ion trajectories, ions are focused collisionally to the vicinity of the center of the ion trap under the influence of collisions with helium buffer gas atoms. This process is described as "ion cooling" in that ion kinetic energies are reduced to ca. 0.1 eV, corresponding to ca. 800 K as calculated from $3RT/2 = 0.1$ eV. Ion excursions from the center of the ion trap are less than 1 mm.

Resonant excitation of cooled ions, brought about by a supplementary potential oscillating at the axial secular frequency of a specific ion species and of amplitude but a few hundreds of millivolts, causes those ions to move away from the ion trap center in such a way that they experience a greater trapping field. This process of ion excitation is often referred to as "tickling". The ions are accelerated further by the trapping field so that they can achieve kinetic energies of tens of electron volts.

Resonant excitation is used in the following ways:

1. to remove unwanted ions during ionization so as to isolate a narrow range of mass/charge ratios; in this case, wavebands of frequencies are applied to

the end-cap electrodes to excite and eject many ion species simultaneously leaving a single ion species (or a small range of mass/charge ratios) isolated within the ion trap;

2. to increase ion kinetic energy so as to promote endothermic ion/molecule reactions;
3. to increase ion kinetic energy so as to deposit internal energy in ions through momentum-exchange collisions with helium atoms; in the limit, ions dissociate. This mode of resonant excitation is discussed in more detail in section 10.2 on CID;
4. to increase kinetic energy so as to move ions close to an end-cap electrode where an image current can be detected. This mode permits the nondestructive measurement and remeasurement of the mass/charge ratio of confined ions;
5. to increase kinetic energy so as to cause ions to escape from the trapping potential and be ejected. This mode can be used either to eject unwanted ions, as in ion isolation, or to eject ions mass selectively while the applied frequency is swept;
6. to eject ions while the amplitude V of the main rf potential is being ramped up. This mode, known as axial modulation, is used in conjunction with an rf ramp so that ions come into resonance with a fixed frequency of ca. 6 $V_{(p-p)}$ (peak-to-peak) just before their trajectories are made unstable. In this case, ions of low mass/charge ratio are removed from the perturbing influences of ions of higher mass/charge ratio and are detected with enhanced resolution.

In axial modulation, the resonant frequency is just less than half the main drive frequency Ω. Resonant excitation at lower frequencies has been used with great success to extend the normal mass range of the ion trap.

9 CALCULATIONS

On many occasions while working with a quadrupole ion trap it becomes necessary to calculate some of the ion trapping parameters such as q_z, LMCO value (see below), β_z, the secular frequency w_z and the potential well depth $\overline{D}_z$. In modern ion trap instruments, these calculations can be carried out using the accompanying software but it is instructive to examine the manner in which each of the following parameters is calculated.

Let us consider an ion of butylbenzene (m/z 134) in a normal stretched ion trap which has a ring electrode of radius $r_0 = 1.00\,\text{cm}$ and with $z_0 = 0.783\,\text{cm}$ (corresponding to an electrode spacing, $2z_0$, of 15.66 mm) and under the following conditions:

$$U = 0; \quad V = 757\,V_{(0-p)} \text{ at } 1.05\,\text{MHz}$$

$$\Omega = 2\pi f = 2\pi \times 1.05 \times 10^6\,\text{rad}\,\text{s}^{-1}$$

$$m = \frac{134\,\text{Da}}{\text{Avogadro's number}} = \frac{134 \times 10^{-3}\,\text{kg}\,\text{mol}^{-1}}{6.022 \times 10^{23}\,\text{mol}^{-1}}$$

9.1 q_z and Low-mass Cut-off

From Equation (26), we recall that

$$q_z = \frac{8eV}{m(r_0^2 + 2z_0^2)\Omega^2} \tag{26}$$

Thus

$$\begin{aligned} q_z = \; & 8(1.602 \times 10^{-19}\,\text{C})(757\,\text{kg}\,\text{m}^2\,\text{s}^{-2}\,\text{C}^{-1}) \\ & \times (6.022 \times 10^{23}\,\text{mol}^{-1}) \qquad (134 \times 10^{-3}\,\text{kg}\,\text{mol}^{-1}) \\ & \times ([1.000 + 1.226]10^{-4}\,\text{m}^2)(2\pi \times 1.05 \times 10^6\,\text{s}^{-1})^2 \\ = \; & 0.450 \end{aligned}$$

We have now calculated that m/z 134 has a q_z value of 0.450 under these conditions, but what is the LMCO value at q_z slightly less than 0.908? Since $m \times q_z =$ constant at constant V from Equation (26), the LMCO value can be calculated as:

$$(\text{LMCO})(0.908) = (m/z\ 134)(0.450)$$

Rearranging,

$$\begin{aligned} \text{LMCO} &= (m/z\ 134)(0.450)/(0.908) \\ &= m/z\ 66.4 \end{aligned}$$

That is, with a potential of $757\,V_{(0-p)}$ applied to the ring electrode, only those ions of $m/z > 66.4$ will be stored. The potential V to be applied to the ring electrode to effect a given LMCO is given as

$$\begin{aligned} V &= (\text{LMCO}) \times (757\,V_{(0-p)})/(m/z\ 66.4) \\ &= (11.40 \times \text{LMCO})V_{(0-p)} \end{aligned}$$

This calculation is particularly useful when an ion is to be fragmented and one wishes to know the low mass/charge limit for fragment ions stored, that is, the LMCO.

9.2 β_z

From Equation (29), we see that β_z is given approximately by $\sqrt{(q_z^2/2)}$ thus, when $q_z = 0.450$, $\beta_z = 0.318$. However, we have exceeded the limit of the approximation relating

q_z and β_z and so the calculated value of β_z is high by about 5%. For m/z 1340, where $q_z = 0.0450$ such that the above approximation is valid, the value of $\beta_z = 0.0318$.

9.3 ω_z

From Equation (27), the fundamental axial secular frequency, w_z, (or, more properly, $w_{z,0}$) is given by $\beta_z\Omega/2$ thus, when $\beta_z = 0.318$ and $\Omega = 2\pi \times 1.05 \times 10^6\,\mathrm{rad\,s^{-1}}$, $w_z = 1.049 \times 10^6\,\mathrm{rad\,s^{-1}}$ or, more conventionally, $w_z =$ 167 kHz; w_z is correspondingly high by about 5%. However, for m/z 1340, w_z is 16.7 kHz.

9.4 Mass Range

The upper limit of the mass range is given by the mass/charge ratio having a q_z value of, let us say, exactly 0.900 when the maximum rf amplitude is applied to the ring electrode. From Equation (26) it is seen that $m \times q_z/V =$ constant; this constant can be evaluated from the above expression for q_z as 0.0797. With $q_z = 0.900$ and $V = 7340\,V_{(0-p)}$, the mass range is found to be 650 Da.

9.5 Mass Range Extension

With $V = 7340\,V_{(0-p)}$ and ion ejection brought about by axial modulation at, say, $q_z = 0.900$, the upper mass limit of the ion trap is 650 Da; that is, for m/z 650, $q_z = 0.900$. However, under these trapping conditions, ions of m/z 1300 remain stored in the ion trap and have a q_z value of 0.450. If resonant excitation had been carried out at $q_z = 0.45$ with excitation at 167 kHz (see above), ions of m/z 1300 would be on the point of ejection at the maximum rf amplitude and the mass range of the ion trap would have been doubled to 1300 Da. Similarly, with resonant excitation at $q_z = 0.045$, the mass range would be extended to 13 000 Da. This method has been applied[19] with great success, using axial modulation at low q_z and a slightly lower drive frequency Ω, such that the mass/charge range was extended to 72 000 Da per charge.

9.6 Mass Resolution

The normal mass scan rate of Finnigan and Varian ion traps prior to 1995 was 5555 Da s^{-1} and over the normal mass range of 10–650 Da, peak width was maintained at ca. 0.5 Da. Mass resolution is defined as the ratio of the mass of an ion to the peak width at half height; hence mass resolution increases with mass. However this statement is somewhat misleading as can be seen from the following example; mass resolution for m/z 65 is ca. 130 while that for m/z 650 is 1300. The real performance of the instrument is gauged from the width of the ion signals produced upon ion ejection. Upon reducing the scan rate, it was observed[20] that the peak width was reduced so that for a given mass, mass resolution was increased. While in research instruments peak widths of some 20 μDa have been observed, the narrowest peak widths in commercial instruments are ca. 0.2 Da so that, for m/z 2000, a mass resolution approaching 10 000 is achieved.

9.7 $\overline{D}_z$

The essential importance of the potential well depth is that it determines both the minimum kinetic energy that an ion must acquire by resonant excitation in order to be ejected from the ion trap and the maximum kinetic energy that an externally generated ion may possess and still be trapped. The magnitude of the potential well in the z-direction may be estimated from the approximation that $\overline{D}_z \approx q_z V/8$ thus, for our example of m/z 134 at $q_z = 0.450$, $\overline{D}_z \approx 0.45 \times 757\,V/8 = 43\,V$.

10 MASS SPECTROMETRIC OPERATION OF THE QUADRUPOLE ION TRAP

A schematic diagram of the QITMS is shown in Figure 11. This figure does not show the end of the gas chromatograph capillary column from which elutes the sample of interest in the helium carrier gas. The gate control in Figure 11 permits the passage of a 200-μs burst of electrons from the filament for automatic gain control (AGC). The fundamental rf voltage is ramped rapidly to eject ions formed during the brief burst of electrons. Under AGC, the instrument calculates the duration of ionization appropriate to the stipulated ion count and the prevailing partial pressure of sample in the ion trap. All forms of resonant excitation are effected with the supplementary AC voltage applied to the end-cap electrodes as shown in Figure 11. When this supplementary AC voltage is used for axial modulation, ions are ejected in order of mass/charge ratio, impinge upon the electron multiplier detector, and produce a signal after preamplification. All

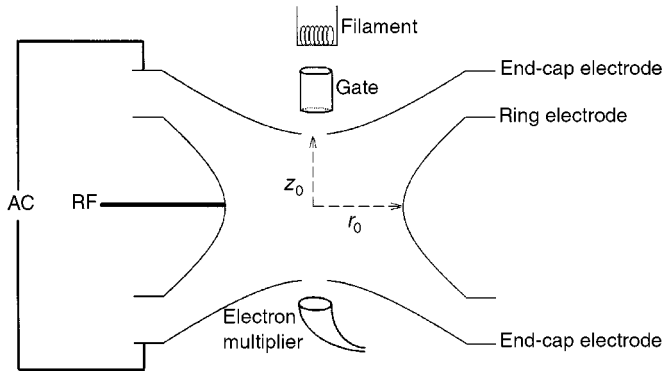

Figure 11 Schematic diagram of the QITMS.

of the voltages applied to the ion trap are controlled by the scan acquisition processor.

Dioxin (2,3,7,8-tetrachlorodibenzo-*p*-dioxin T_4CDD) is used for illustrating the use of the ion trap as a MS; dioxin is an excellent example of the application of ion trap MS to analytical chemistry. Compounds such as dioxin are analyzed from complex mixtures where they are present in trace amounts and, under such circumstances, gas chromatography (GC) is a common method for sample introduction; in this example, ca. 400 pg of dioxin dissolved in 2 μL of *n*-nonane is used. The pressure of helium in the ion trap is set at about 10^{-3} Torr to provide adequate cooling of ions formed; while helium can be used directly from a cylinder, the helium carrier gas flowing through a chromatograph provides an adequate background pressure in the ion trap.

Dioxin molecules elute directly into the ion trap where they are bombarded with electrons of 50–80 eV emitted from a heated filament and gated into the ion trap, as shown in Figure 11. The duration of ionization is determined by AGC; the ion number formed during an ionization burst is used to scale the ionization time in order to produce the required number of ions. The ions thus formed come immediately under the influence of the trapping potential within the ion trap. During ionization, the ring electrode is driven at an initial rf voltage V_0 and a fixed frequency ($f \approx 1$ MHz in many instruments) so that all ions in a given range of mass/charge ratio are trapped within the imposed quadrupole field. The value of the initial rf voltage V_0 imposes a LMCO, usually in the range m/z 20–50, so that ions of lower mass/charge ratio are not stored. No DC potential is applied between the ring and end-cap electrodes ($U = 0$) so that the confining field is purely oscillatory. During and after ionization, ions follow Lissajous-type trajectories and are subjected simultaneously to about 20 000 collisions per second with helium; those ions that are not lost from the ion trap become focused near the trap center. The rf amplitude is ramped during the analytical scan and as the axial secular frequency of ion motion comes into resonance with the axial modulation frequency, mass-selective ion ejection and mass analysis occur.

Each ion species confined within the ion trap is associated with a q_z value which is calculated according to Equation (26) and which lies on the q_z-axis on the stability diagram; ions of relatively high mass/charge ratio have q_z values near the origin while ions of lower mass/charge ratio have q_z values which extend towards the $\beta_z = 1$ stability boundary, as shown diagrammatically using stick-people of various sizes in Figure 12(a). At the intersection of the $\beta_z = 1$ stability boundary and the q_z-axis, where $q_z = 0.908$ (see Figure 10), the trajectories of trapped ions become unstable axially such that ions of mass/charge ratio less than the LMCO are not stored.

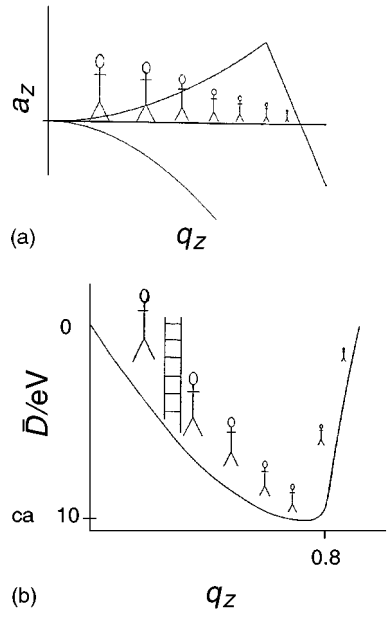

Figure 12 (a) Schematic representation of working points (that is, coordinates in a_z, q_z space) in the stability diagram for several species of ions stored concurrently. The arrangement of working points with respect to mass/charge ratio is depicted by figures which differ in size. (b) Ions are shown residing near the bottom of their respective axial potential wells of depth $\overline{D}_z$; the ladder represents the opportunity for resonant ejection of an ion species.

Once the ion cloud within the ion trap has been focused collisionally to the center of the ion trap over a period of some 1–30 ms, the amplitude of the rf potential is ramped. The ramping of the rf potential amplitude, which is described as an analytical ramp or analytical scan, causes the q_z values of all ion species to increase throughout the ramp. As the q_z value for each ion species reaches a value of 0.908, the ions are ejected axially through the end-cap electrodes. This method of ion ejection, which can occur only at a boundary of the stability diagram and is referred to as *mass-selective axial instability*, has been supplanted by the use of *axial modulation*. Originally, axial modulation was the name given to resonant ejection of ions at a frequency of 485 kHz and at a q_z value only slightly less than 0.908.

When the rf amplitude is ramped, ions come into resonance at 485 kHz as their q_z values approach 0.908; in this manner, ions are ejected axially in order of increasing mass/charge ratio. As ions are focused into a region at the center of the ion trap, they are arranged in concentric layers rather in the fashion of an onion. Resonance ejection affects first the ions of low mass/charge ratio residing in the innermost layer of the onion by removing them axially from ions of higher mass/charge ratio. In this manner, the influence of space charge perturbations induced by the ion of higher mass/charge ratio is sharply

reduced so that the ions of low mass/charge ratio are ejected free of space charge and with enhanced mass resolution. An additional advantage of resonant ejection is that it can be carried out at any frequency. Resonant ejection is depicted pictorially in Figure 12(b) where ions are shown residing near the bottom of their respective axial potential wells of depth $\overline{D}_z$; the ladder represents the opportunity for resonant ejection of an ion species at any frequency. For axial modulation, the ladder is positioned at a q_z value just a little less than 0.908. Resonantly ejected ions pass through holes in the end-cap electrodes so that only half of them impinge upon an electron multiplier located behind one of the end-cap electrodes; ion signals are created which produce a mass spectrum in order of increasing mass/charge ratio. The quadrupole ion trap functions as a mass spectrometer when operated in this manner.

A mass spectrum is obtained by running the scan function (see below) a number of times specified by the number of microscans; the signals which compose the mass spectrum are the result of averaging those obtained from each microscan. In the following example of obtaining a mass spectrum, one mass spectral file was generated each second. A mass spectrum of 2,3,7,8-T_4CDD obtained in this manner is shown in Figure 13. $M^{+\bullet}$ is of m/z 320 and the $[M + 2]^{+\bullet}$ ion of the molecular cluster which is shown at m/z 322 corresponds to the inclusion of one ^{37}Cl atom; the ion of m/z 259 is a primary fragment ion and is due to the loss of $COCl^{\bullet}$ from $[M + 2]^{+\bullet}$ and of $CO^{37}Cl^{\bullet}$ from $[M + 4]^{+\bullet}$. The peak at m/z 334 is $[M + 2]^{+\bullet}$ from a relatively small amount of ^{13}C-labeled $^{13}C_{12}$-2,3,7,8-T_4CDD, $M^{+\bullet}$ of m/z 332, added to the sample injected.

10.1 Scan Function

The sequence of events described above can be expressed succinctly in a scan function which shows the temporal variation of all of the potentials applied to the ion trap electrodes. A scan function is a visual representation of the sequence of program segments in the software that controls ion trap operation. The scan function for the mass spectrometric operation of the ion trap described above is shown in Figure 14.

10.2 Collision-induced Dissociation

CID of an isolated ion species in a quadrupole ion trap has become a powerful technique for both the determination of ion structures and the analytical determination of, for example, compounds of environmental interest such as dioxin. Before we explore each of these uses, let us consider the CID process, how ion dissociation is effected, the efficiency of the process, and the duration.

While CID is effected by resonant excitation of a selected ion species in the ion trap, the process is made more complex since the frequencies of ion motion change slightly as an ion moves away from the ion trap center; the values of the frequencies of ion motion are dependent also on space charge, that is, on the number and nature of other ions present in the ion trap. The variation in axial frequency with axial excursion, which is shown in

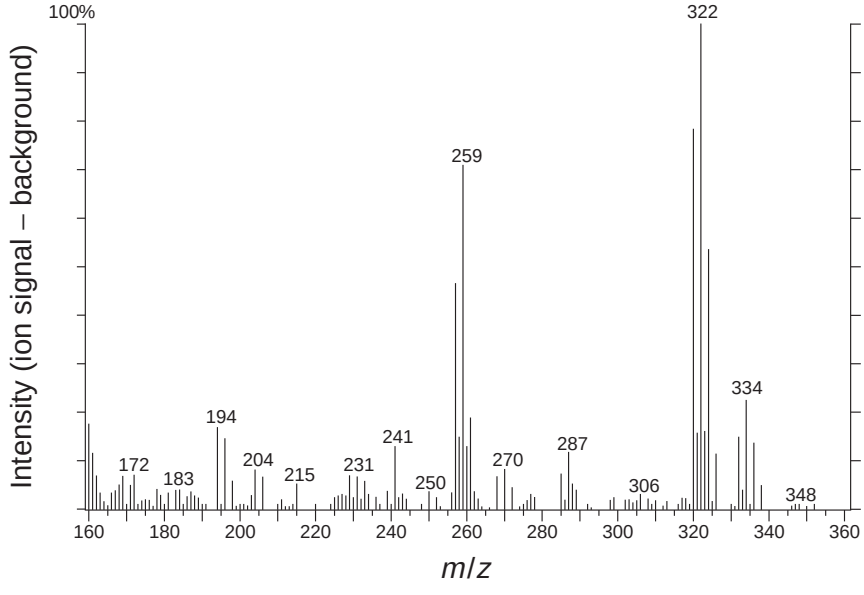

Figure 13 Mass spectrum of 2,3,7,8-T_4CDD. The $[M + 2]^{+\bullet}$ ion of the molecular cluster is shown at m/z 322, while that of a trace of $^{13}C_{12}$-2,3,7,8-T_4CDD is shown at m/z 334; the primary fragmentation process is the loss of $COCl^{\bullet}$ giving rise to m/z 259 and 270, respectively.

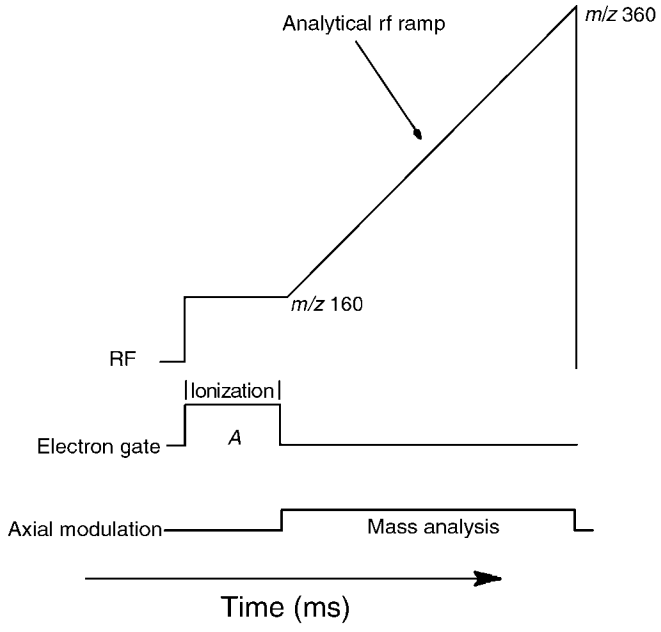

Figure 14 Scan function for obtaining an EI (electron ionization) mass spectrum. The scan function shows the ionization period, *A*, followed immediately by the analytical ramp with concurrent axial modulation. Note that the prescan for the AGC algorithm is not shown.

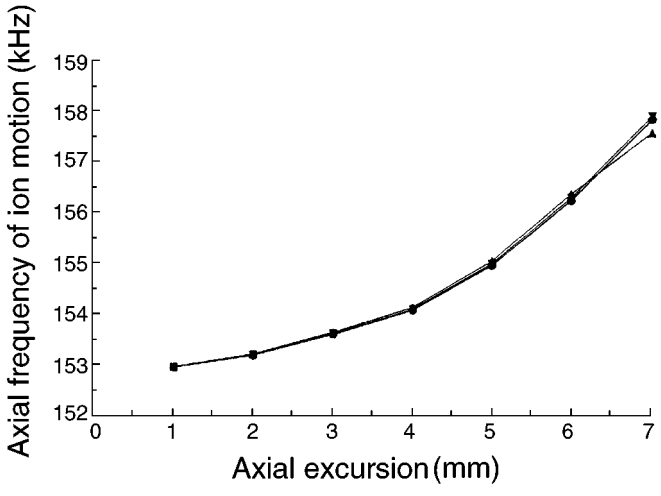

Figure 15 Variation of ion axial secular frequency as a function of the axial excursion from the center of the ion trap. Ion trajectories were calculated for three *m/z* 134 ions, 1 (●), 2(■) and 3( ), taken at random, with $q_z = 0.4$ and subjected to resonant excitation. While the three ions differed in their initial position and velocity, all were close to the center of the ion trap and had been cooled collisionally. When specific axial excursions had been reached, ion excitation was arrested and the trajectory of each ion was subjected to frequency analysis.

Figure 15 as obtained from analysis of simulated ion trajectories, arises from the superimposition of higher order field components as a result of the axial "stretching" of the ion trap.[21,22] The secular frequency variation necessitates the development of strategies for optimizing CID, particularly when the duration of irradiation must be minimized as in GC/MS (gas chromatography/mass spectrometry) analytical applications; these strategies include secular frequency modulation, single frequency irradiation with rf modulation, and multiple frequency irradiation.

Under the influence of the applied resonant excitation voltage, ions are moved away from the center to a region of higher potential whereupon they are accelerated (under the influence of the higher potential rather than the resonant excitation voltage alone) so that their kinetic energies are increased. Subsequent collisions with buffer gas atoms lead to enhancement of ion internal energy. Ions undergo rapid changes in kinetic energy on the microsecond timescale whereas changes in internal energy occur more slowly on the millisecond timescale. In analytical applications, the objective is usually to dissociate all of the isolated ions and to maximize the trapping of fragment ions produced; the achievement of this objective requires a balancing of ion kinetic energy uptake so that ion internal energy may be accumulated incrementally and rapidly but ejection of both isolated ions and fragment ions is prevented. Incremental loss of internal energy in collisions can occur with both buffer gas atoms and sample molecules while total loss of internal energy can occur upon charge exchange with sample molecules; yet these loss mechanisms merely slow down the overall accumulation of internal energy as there is no net loss of charge involved.

The role played by time is enormously important in CID. The total time during which ions are subjected to collisions and the total time during which ions are subjected to irradiation can be used as variable parameters to direct dissociation into a chosen fragmentation pathway. While the mode of irradiation is also influential in directing fragmentation, alternating periods of irradiation (with collisions) and collisional cooling can affect significantly the fragment ion abundances. During an irradiation period, the time-averaged ion kinetic energies increase along with internal energy: during a subsequent collisional cooling period in which the excitation voltage is removed, ion kinetic energies are quenched more readily than are ion internal energies. Thus ions become kinetically relaxed during a cooling period and sample once more the lower reaches of the trapping potential well, yet the internal energies gained are virtually unimpaired. The internally excited ions, kinetically cool and focused near the center of the ion trap, can be excited further during the next irradiation period. In this manner, several electron volts of internal energy can be deposited in ions to allow access to fragmentation pathways of high activation energy.

11 TANDEM MASS SPECTROMETRY

Tandem (latin: at length) mass spectrometry, (MS/MS) or $(MS)^n$, is the practice of carrying out one mass-selective operation after another, much as one rider is seated after the other on a tandem bicycle. The objective of the first mass-selective operation is to isolate an ion species designated as the *parent* ion, while that of the second operation is to determine the mass/charge ratios of the fragment, or *product*, ions formed by CID of the parent ions. MS/MS can be effected *in space*, by placing one mass spectrometer after another, or by carrying out successive mass-selective operations *in time* in a quadrupole ion trap. The ion trap offers two principal advantages when used in the MS/MS mode. First, the ion trap operates in a pulsed mode, compared with sector and triple stage quadrupole instruments that operate in a continuous mode, so that it can accumulate ions mass-selectively over time. In this way, a target ion number can be selected so as to ensure constant signal/noise ratio over a wide range of eluent concentrations. A disadvantage of ion accumulation is that the resultant product ion signal intensity can correspond to integration of eluent concentration over some 25 ms. Second, CID in the ion trap is wrought by some hundreds of collisions of mass-selected ions with helium buffer gas atoms. Under these conditions, the energy transferred in a single collision is seldom greater than that of a vibrational quantum such that the dissociation reaction channels of lowest energy of activation are accessed almost exclusively; this behavior is highly advantageous in analytical chemistry because the total charge is conserved within a single fragment ion species. However, alternating periods of excitation and collisional cooling can be used to allow access to dissociation reaction channels of higher energy of activation. Furthermore, it is possible to dissociate completely the accumulated mass-selected ions and to confine within the ion trap fragment ions arising from some 90% of the accumulated mass-selected ions in favorable cases.

To achieve CID with such high efficiency (defined as 100 times the ratio of the sum of product ion signal intensities to that of the parent ion) when the ion trap is interfaced with a gas chromatograph, it is imperative that the resonant excitation conditions be optimized with respect to the duration of the CID episode. Here, we use multifrequency irradiation (MFI) as it has been shown to be an effective method for CID because it exhibits high efficiency and requires a relatively short period of irradiation, ca. 10 ms.

11.1 Tandem Mass Spectrometry

Here, MS/MS is illustrated using an extension of the dioxin example used in the above discussion. This application,[23] while not illustrating the ultimate limit of sensitivity of the ion trap of some hundreds of

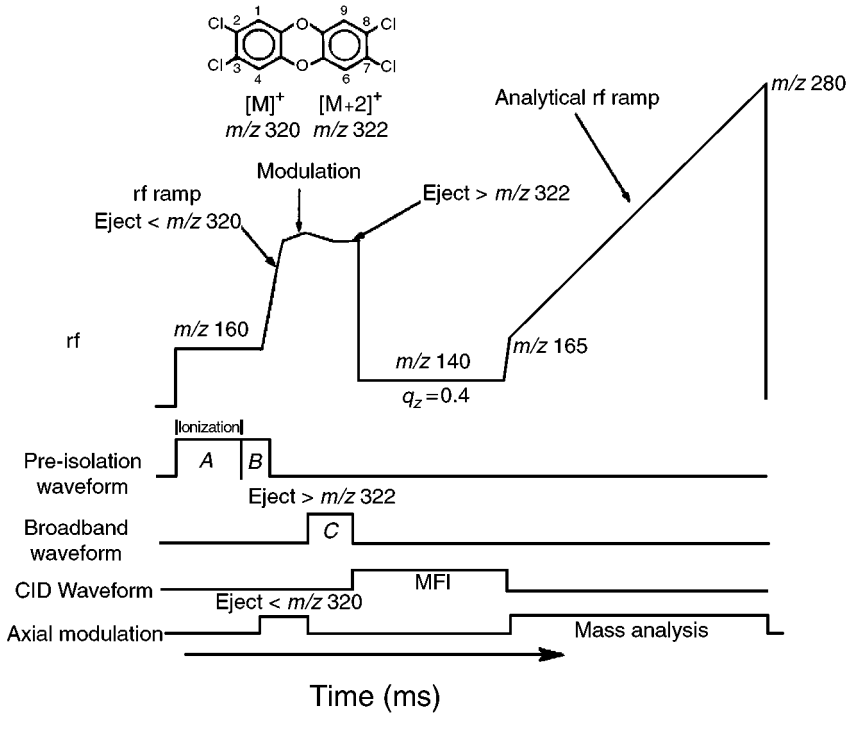

Figure 16 Scan function for MS/MS of dioxin T_4CDD; the prescan for AGC is not shown.

femtograms of dioxin, does illustrate the high sensitivity of the ion trap for the determination of co-eluting dioxin congeners. Polychlorodibenzo-*p*-dioxins (PCDDs) and polychlorodibenzofurans (PCDFs) are persistent organochlorine compounds in the environment; most of these chlorocongeners are obtained from municipal and industrial waste incinerators, automobile exhaust, and the manufacture of chlorophenol products. The compounds having the highest toxicity have been identified as those congeners having 2,3,7,8-tetrachloro-substitution. When GC is interfaced with MS, individual chlorocongeners can be detected at the hundreds of femtograms level. The high specificity or informing power obtainable with GC/MS/MS is achieved by observation of specific fragment ion signals, such as $[M - COCl^{\bullet}]^{+}$, from an isolated ion species $M^{+\bullet}$ formed from M which elutes within a specified retention-time window. The essential stages of MS/MS are portrayed in the scan function for dioxin shown in Figure 16. It is instructive to examine this scan function in advance of the following explanation of ionization, ion isolation, cooling period, CID, and the analytical ramp over a selected mass range so as to detect product ions.

We shall examine the MS/MS of $[M]^{+\bullet}$ and $[M + 2]^{+\bullet}$ for T_4CDF (tetrachlorodibenzofurans), T_4CDD and their internal (labeled) standards following elution from a GC column, where $[M + 2]^{+\bullet}$ is a molecular ion containing a single $^{37}Cl^{\bullet}$ atom.

11.2 Scan Functions

The scan function employed for the MS/MS determination of T_4CDDs is shown in Figure 16. The rf voltage is applied to the ring electrode with a drive frequency f of 1.05 MHz. The LMCO value, as determined by the amplitude of the rf potential, was set at m/z 160 during the ionization period (A) for all the scan functions so that the q_z value for $M^{+\bullet}$ (m/z 320) was 0.45. While the same scan function can be used for all T_4CDDs, as they do not co-elute, different scan functions must be used in sequence for the determination of a specific T_4CDD and its co-eluting labeled isotopomer.

A total ion number target of 35 000 counts was set for the AGC algorithm; with a filament emission current of 50 μA, the maximum ionization time employed was 20 ms. The supplementary alternating voltages applied to the end-cap electrodes in dipolar fashion are referred to as waveforms; these waveforms are employed for ion isolation, ion excitation, and axial modulation. A preisolation waveform was imposed during ionization (A) and prolonged after the cessation of ionization during period B. The pre-isolation waveform consisted of multiple frequencies covering the range 3.7–513.5 kHz with a 1 kHz notch corresponding to the secular frequency of the molecular ions to be isolated. For the example of a T_4CDD scan function shown in Figure 16, the notch is centered at 174.5 kHz in order to isolate both m/z

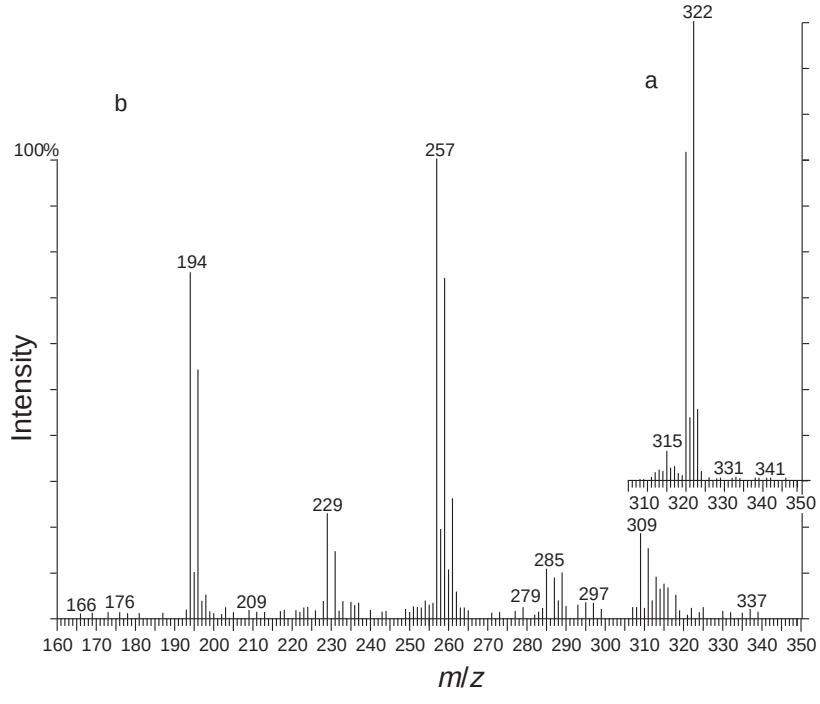

Figure 17 Mass spectra of T_4CDD: (a) mass spectrum showing the isolated $[M]^{+\bullet}$ and $[M + 2]^{+\bullet}$ ions of the molecular ion cluster; (b) product ion mass spectrum obtained by CID of the two isolated ion species shown in part (a). Note the complete disappearance of the isolated molecular ions.

320 and m/z 322, $[M]^{+\bullet}$ and $[M+2]^{+\bullet}$, respectively, having q_z values of 0.454 and 0.451, respectively. The amplitude of the pre-isolation waveform was $20\,V_{(0-p)}$ for all scan functions; the function of the pre-isolation waveform was to eject all ions except those within a mass range of ca. 10 Da about the selected ones. Fine isolation was achieved by ramping the rf amplitude until the LMCO was just less than m/z 320 at which point ions of lower mass/charge ratios were ejected; ejection of ions of higher mass/charge ratio was facilitated by concurrent application of axial modulation with an amplitude of $3\,V_{(0-p)}$. The rf amplitude was modulated moderately over a small amplitude range in order to avoid ejection of the selected ions. The ejection of ions with $m/z > 322$ (C) was achieved by applying a broad band waveform having an amplitude of $30\,V_{(0-p)}$ and lasting for 5 ms.

Once isolation of the selected ion species m/z 320 and m/z 322 for T_4CDD as illustrated in Figure 17(a) was completed, the rf amplitude was reduced to obtain a q_z value of 0.4 for m/z 322; for T_4CDD, a q_z value of 0.4 for m/z 322 corresponds to a LMCO value of m/z 142. The bandwidth of the waveform employed to carry out CID using MFI was composed of 13 frequency components spaced at intervals of 0.5 kHz so as to cover a 6 kHz band of frequencies. The total bandwidth is almost double the required bandwidth since the Toolkit® software locks the center of the MFI band to the axial secular frequency of the previously selected ion species; only the width of the MFI band can be varied. The MFI waveform amplitudes were $2.55\,V_{(0-p)}$ for T_4CDF and $2.45\,V_{(0-p)}$ for T_4CDD. The voltage amplitude of each component of the MFI waveform is equal to the amplitude of the waveform divided by the number of frequency components. The rf amplitude was modulated so as to cover completely the frequency range of the 13 frequency components of MFI.

Following CID by MFI for 10 ms, the analytical mass range of 165–350 Da was scanned for product ions from T_4CDF and T_4CDD so as to monitor all major product ions save that resulting from chlorine atom loss. A mass spectrum of the product ions from dioxin is shown in Figure 17(b): m/z 257 and m/z 194 are due to the losses of $COCl^{\bullet}$ or $2COCl^{\bullet}$, respectively, from $M^{+\bullet}$ (m/z 320). The analytical ramp was scanned at 5555 Da s^{-1}; axial modulation was carried out with an amplitude of $3\,V_{(0-p)}$ and at a frequency of 485 kHz. The electron multiplier was biased at a voltage of ca. 1800 V to provide an ion signal gain of 10^5.

Each scan function used had a duration of ca. 125 ms such that each acquisition point or mass spectral file was generated from four microscans; thus two mass spectra were accumulated each second. It should be noted that when a small mass range only is monitored during the analytical rf ramp, say for two fragment ions only, the duration of the scan function can be reduced to only 50 ms; in this case, the number of mass spectral files can be increased to five per second, each based on four microscans. The mass spectra obtained from the application of each scan function within an ion preparation file were merged together to form a single merged mass spectrum. The significance of this merging procedure is described below.

11.3 Tandem Mass Spectrometry Determination of Eluting and Co-eluting Compounds

The total ion chromatogram shown in Figure 18 was obtained by GC/MS/MS of 1 μL of solution containing 200 pg of each of a T_4CDF, a $^{13}C_{12}$-T_4CDF, a T_4CDD and a $^{37}Cl_4$-T_4CDD and 100 pg of each of two $^{13}C_{12}$-T_4CDDs. The peak indicated with an asterisk in Figure 18 is an impurity present in the sample. This chromatogram is described more properly as a merged total ion chromatogram since it is a display of the total ion count of each merged mass spectrum obtained as described above.

Let us consider the first three peaks of interest which were observed in this merged total ion chromatogram, that is, those peaks labeled 1, 2 and 3. Peak 1 is due to a T_4CDF and its co-eluting $^{13}C_{12}$-labeled isotopomer (a component of the solution injected); peak 2 is due solely to a $^{13}C_{12}$-T_4CDD while peak 3 is a composite peak of a native T_4CDD and its $^{13}C_{12}$- and $^{37}Cl_4$-labeled isotopomers.

For the T_4CDF of molecular weight is 304, the molecular ion $M^{+\bullet}$ of m/z 304 and the $[M+2]^{+\bullet}$ of m/z 306 were isolated simultaneously then dissociated collisionally using MFI in accordance with a specific scan function. The signal intensities of fragment ions formed by the loss of $COCl^{\bullet}$ or $CO^{37}Cl^{\bullet}$ to yield m/z

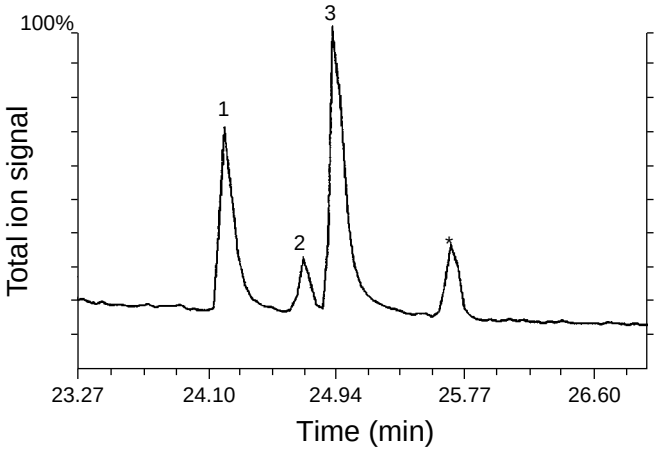

Figure 18 Merged total ion chromatogram of six dioxins and furans and their labeled isotopomers. Each point expresses the sum of all of the relevant fragment ion peaks.

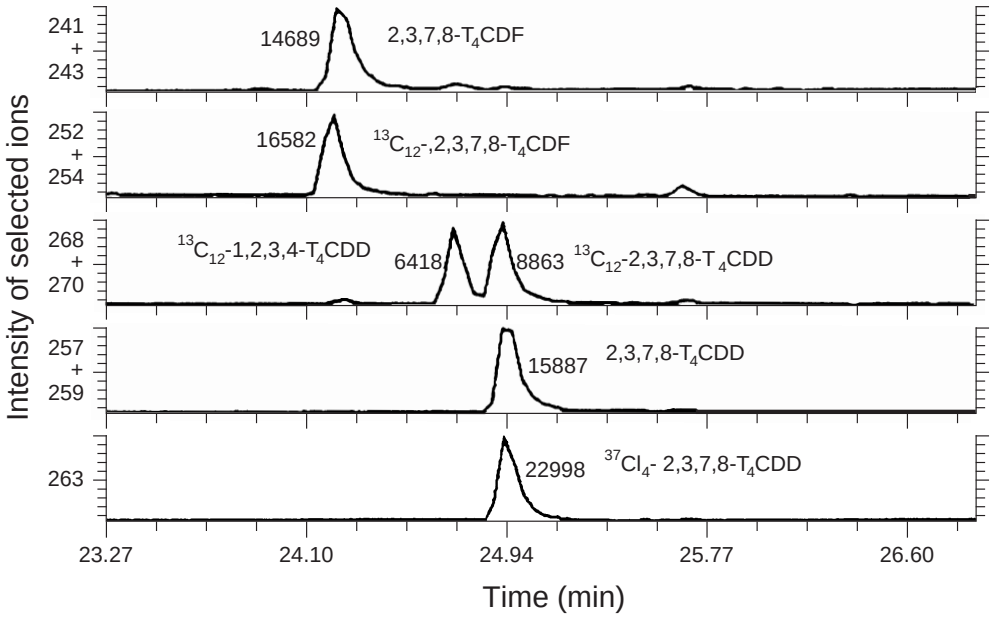

Figure 19 Selected ion chromatogram showing that the signal area of each of six compounds can be determined using GC/MS/MS.

241 and *m*/*z* 243 were recorded, summed and shown in Figure 19. For MS/MS of the co-eluting $^{13}C_{12}$-T_4CDF, the second scan function comes into play for the ionization and simultaneous isolation of both $M^{+\bullet}$ of *m*/*z* 316 and $[M + 2]^{+\bullet}$ of *m*/*z* 318; the isolated ion species were then subjected to CID with MFI. The signal intensities of fragment ions formed by the loss of $COCl^{\bullet}$ or $CO^{37}Cl^{\bullet}$ to yield *m*/*z* 252 and *m*/*z* 254 were recorded and summed as shown in Figure 19. Throughout the period when the native and labeled T_4CDFs are co-eluting, the first two scan functions are used alternately.

Now let us consider peak 2 of Figure 18. A third scan function is required for the MS/MS determination of the labeled $^{13}C_{12}$-T_4CDD to control the ionization, isolation, and CID of both the molecular ion $M^{+\bullet}$ of *m*/*z* 332 and the $[M + 2]^{+\bullet}$ of *m*/*z* 334. The signal intensities of fragment ions formed by the loss of $COCl^{\bullet}$ or $CO^{37}Cl^{\bullet}$ to yield *m*/*z* 268 and *m*/*z* 270 were summed and recorded in Figure 19.

Peak 3 of Figure 18 is composed of the fragment ion counts from three compounds. The scan function used in peak 2 was used again for the MS/MS determination of a T_4CDD together with two additional scan functions for the co-eluting $^{13}C_{12}$- and $^{37}Cl_4$-T_4CDD isotopomers. For T_4CDD of molecular weight 320, the molecular ions $M^{+\bullet}$, of *m*/*z* 320, and $[M + 2]^{+\bullet}$, of *m*/*z* 322, were isolated simultaneously then were dissociated as before; the resulting signal intensities of the fragment ions of *m*/*z* 257 and *m*/*z* 259, as shown in Figure 17, were summed and recorded in Figure 19. From the MS/MS of the $^{13}C_{12}$-T_4CDD, the signal intensities of the fragment ions of *m*/*z* 268 and *m*/*z* 270 were summed and recorded in Figure 19. For the MS/MS of $^{37}Cl_4$-T_4CDD of molecular weight 328, the molecular ion $M^{+\bullet}$ of *m*/*z* 328 was isolated then dissociated as before and the signal intensity of the fragment ions arising from the loss of $CO^{37}Cl^{\bullet}$ to yield *m*/*z* 263 was recorded in Figure 19.

The signal areas of each of the six separated compounds are shown in Figure 19 as selected ion chromatograms; the signal intensities of fragment ions due to loss of $COCl^{\bullet}$ or $CO^{37}Cl^{\bullet}$ are plotted here. In order to perform MS/MS on the total of six different tetrachlorinated dioxins and furans which constituted the first three peaks of Figure 18, a total of five different scan functions was required. In this example, the enormous power of MS/MS is evident in the determination of three co-eluting compounds.

In the above discussion, only a single fragmentation channel was considered, that of the loss of $COCl^{\bullet}$ or $CO^{37}Cl^{\bullet}$. While $COCl^{\bullet}$ loss is the major fragmentation channel, it is not the sole loss channel. For the dioxins, minor losses of $Cl^{\bullet}$ and $2COCl^{\bullet}$ are observed while for the furans, minor losses of $Cl^{\bullet}$, $COCl_2$ and $COCl_3^{\bullet}$ are observed. The duration of the analytical rf ramp was reduced by selecting a small mass range that excluded the fragment ions of low intensity.

11.4 Tandem Mass Spectrometric Operation

The above sequence of ion isolation and CID can be repeated many times in a process known as $(MS)^n$. While this process is not illustrated here, multiple stages of mass-selective operation are used frequently for ion structure determination where the quadrupole ion trap is used

in combination with external ionization sources, such as ESI. Here, multiply charged molecules having masses of thousands of daltons can be stored in an ion trap and subjected to $(MS)^n$ so as to follow the stepwise ion dissociation for the elucidation of ion structure.

12 CHEMICAL IONIZATION AND ION/MOLECULE REACTIONS

In the quadrupole ion trap, several types of reactions can and do occur simultaneously and spontaneously once EI of a compound has occurred. Ion/molecule reactions involving charge transfer, proton transfer, and clustering occur sequentially in a type of thermodynamic ''waterfall'' which results in the formation of stable even-electron ions of greater mass/charge ratio than that of the molecular ion. Proton transfer CI, involves the transfer to a neutral species of a proton from an ion which has been formed in an ion/molecule reaction. Parenthetically, CI can be effected by the transfer of other even-electron charged particles.

Let us consider the CI of the co-eluting PCB congeners 77 and 110 shown in Scheme 1.[24] Congener 77, which is highly toxic, is a nonortho compound because the molecule is not chlorine-substituted in the ortho positions vicinal to the phenyl–phenyl bond. On the other hand, congener 110 is a diortho compound and is much less toxic. Since the concentration of toxic congener 77 in environmental samples is often <1% of that of congener 110, quantitative analysis of congener 77 is a challenge.

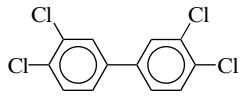

Congener 77

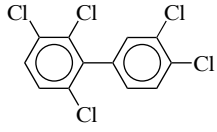

Congener 110

Scheme 1 PCB congeners 77(3,3′,4,4′,-tetrachlorobiphenyl) and 110(2,3,3′,4′,6-pentachlorobiphenyl).

The EI mass spectrum, averaged over the GC composite peak of the co-eluting congeners 77 and 110, is shown in Figure 20. The base peak group in the mass spectrum is the molecular ion cluster of the tetrachloro-congener 77, where m/z 292 is $[M_{77} + 2]^{+\bullet}$; this peak is superimposed on the fragment ion cluster owing to the loss of one chlorine atom from pentachloro congener 110 $[M_{110} + 2]^{+\bullet}$, m/z 326, where m/z 291 is $[M_{110} + 2 - Cl]^{+\bullet}$. The peak at m/z 220 is due to loss of $2Cl^{\bullet}$ from $[M_{77} + 2]^{+\bullet}$ and $3Cl^{\bullet}$ from $[M_{110} + 2]^{+\bullet}$, while the peak at m/z 256 is due principally to the loss of $2Cl^{\bullet}$ from $[M_{110} + 2]^{+\bullet}$. It is not possible to determine quantitatively the relative contributions of the two congeners to this mass spectrum. Now let us examine the CI of these congeners using methane as the CI reagent.

The ionization of methane yields the molecular ion that reacts with methane to form CH_5^+, Equation (31)

$$CH_4^{+\bullet} + CH_4 \longrightarrow CH_5^+ + CH_3^{\bullet} \qquad (31)$$

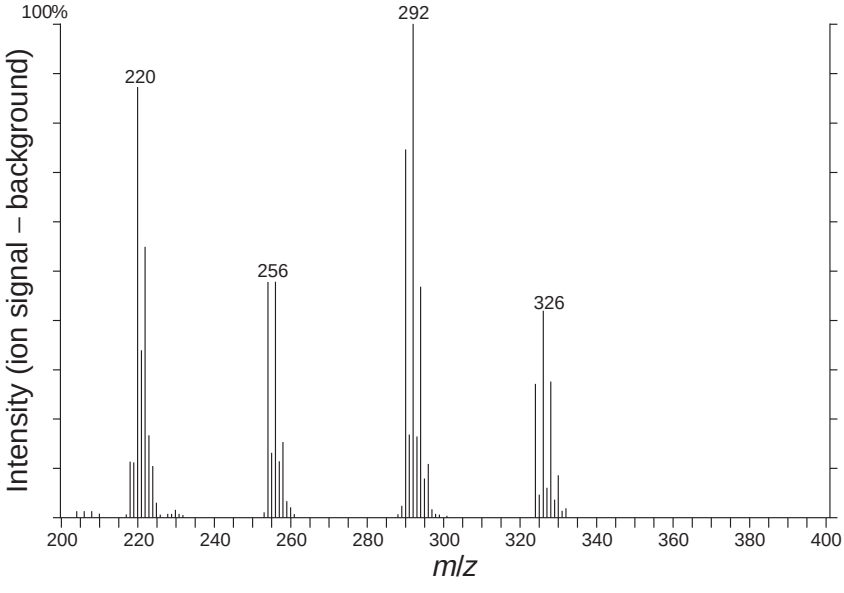

Figure 20 EI mass spectrum averaged over the GC composite peak of the co-eluting polychlorinated congeners 77 and 110.

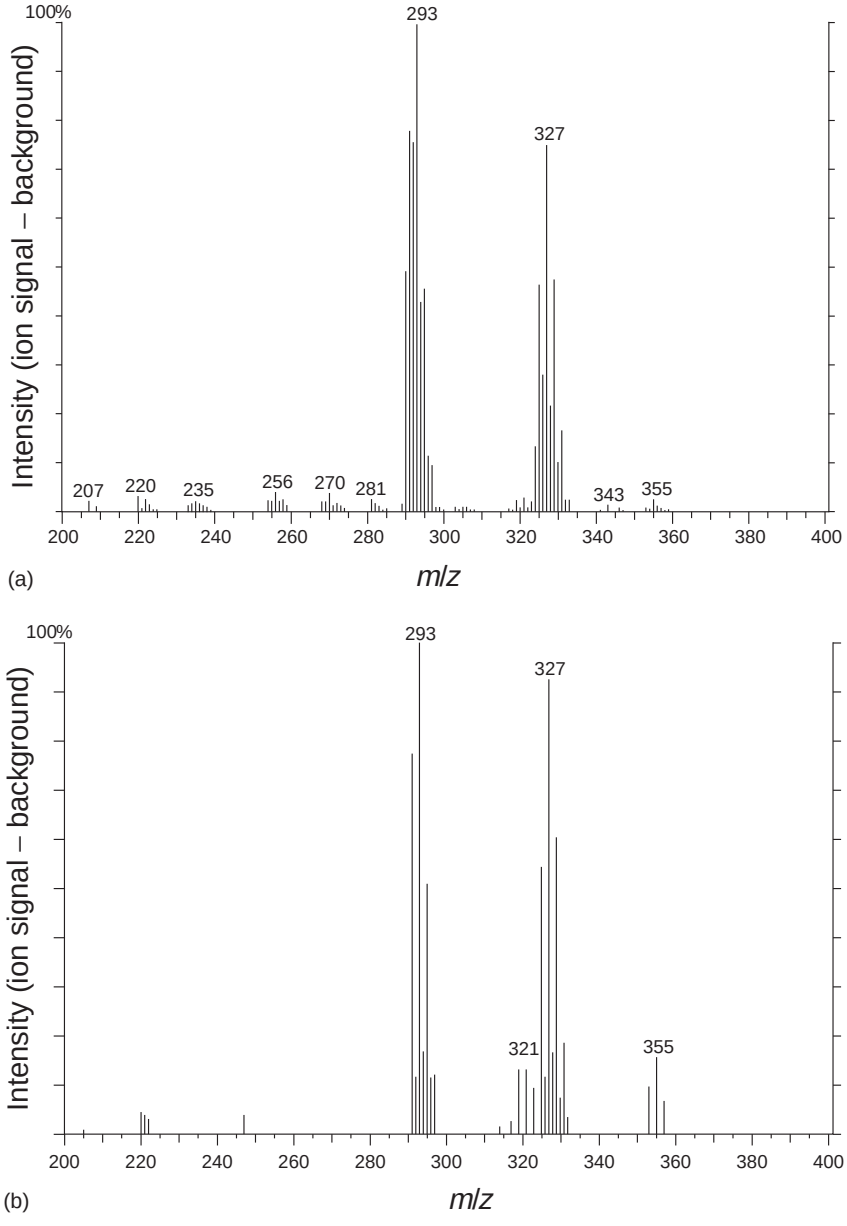

Figure 21 (a) CI mass spectrum averaged over the GC composite peak of the co-eluting PCB congeners 77 and 110; methane CI gas at a pressure and reaction time such that the ratio of signal intensities of *m*/*z* 29 and *m*/*z* 17 was 1 : 1. (b) Selected reagent ion CI mass spectrum averaged over the GC composite peak of the co-eluting PCB congeners 77 and 110 using *m*/*z* 29 from methane.

and the CH_3^+ ion which reacts with methane to form $C_2H_5^+$, Equation (32)

$$CH_3^+ + CH_4 \longrightarrow C_2H_5^+ + H_2 \qquad (32)$$

In practice, an LMCO of ca. *m*/*z* 12 is established by the rf amplitude then methane is introduced briefly into the ion trap and ionized simultaneously with the congeners. Owing to the preponderance of methane, CH_5^+ and $C_2H_5^+$ are formed rapidly during a short interval inserted into the scan function for ion/molecule reactions and can be isolated; during the isolation process, primary ions of methane and the congeners together with other secondary ions from methane are ejected and surplus methane is pumped away. Proton transfer then occurs during a reaction period of ca. 10 ms.

When methane is used as a CI reagent and the reaction time is varied such that the signal intensities of CH_5^+ and $C_2H_5^+$ are similar, CI of a mixture of congeners 77 and 110 yields a mixture of proton transfer CI (to form $[M + H]^+$) and charge transfer (to form $M^{+\bullet}$), as shown in Figure 21(a). However, when a single reagent ion

species, in this case $C_2H_5^+$ of m/z 29, is isolated for CI, proton transfer *only* is observed for both congeners, ($[M_{77} + 2 + H]^+$, m/z 293 and $[M_{110} + 2 + H]^+$, m/z 327), as shown in Figure 21(b). The isotopic ratios for each cluster of protonated congeners are in good agreement with those expected for pure CI: the contribution of charge exchange to the congener 77 molecular cluster is 3%. The mass spectrum is entirely free of fragment ions below $[M_{77} + H]^+$. The adduct formation channel can be suppressed by reducing reagent gas pressure.

Now that external ion sources can be used with an ion trap, CI reagent ions can be created externally and injected subsequently into the ion trap, isolated mass selectively and allowed to react therein with sample molecules.

13 CONCLUSIONS

The QITMS is a versatile instrument of high sensitivity and high specificity. The relatively low cost of commercial instrumentation has permitted a substantial growth in the practice of MS and a pronounced diminution of the average cost per mass spectrum. The theory of ion trap operation differs from those of other mass spectrometers and presents a challenge to the MS community.

ACKNOWLEDGMENTS

The author acknowledges gratefully the support of Trent University and the Natural Sciences and Engineering Research Council of Canada. The author is indebted to Denise McMaster, Amy Trang, Dr Mehran Sharifi, Glenn Legault, and Douglas Simmons who have assisted by reading and criticizing the text and by the preparation of figures.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
AGC	Automatic Gain Control
CI	Chemical Ionization
CID	Collision-induced Dissociation
DC	Direct Current
EI	Electron Ionization
ESI	Electrospray Ionization
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GC/MS/MS	Gas Chromatography/Ion Trap Tandem Mass Spectrometry
ICR/MS	Ion Cyclotron Resonance Mass Spectrometry
ITSIM	Ion Trajectory Simulation Program
LC	Liquid Chromatography
LC/ESI/MS	Liquid Chromatography/Electrospray Ionization Mass Spectrometry
LC/MS/MS	Liquid Chromatography/Ion Trap Tandem Mass Spectrometry
LMCO	Low-mass Cut-off
MFI	Multifrequency Irradiation
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
$(MS)^n$	Tandem Mass Spectrometric Operation
PCB	Polychlorinated Biphenyl
PCDD	Polychlorodibenzo-*p*-dioxin
PCDF	Polychlorodibenzofuran
QITMS	Quadrupole Ion Trap Mass Spectrometer
rf	Radio Frequency
T_4CDD	Tetrachlorodibenzo-*p*-dioxin
T_4CDF	Tetrachlorodibenzofuran

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Mass Spectrometry in Structural Biology

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Laser Mass Spectrometry in Trace Analysis

Environment: Water and Waste **(Volume 3)**
Dioxin-like Compounds, Screening Assays • Explosives Analysis in the Environment

Environment: Water and Waste cont'd **(Volume 4)**
Polychlorinated Biphenyls Analysis in Environmental Samples • Trace Organic Analysis by Gas Chromatography with Quadrupole Mass Spectrometry

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Pesticides **(Volume 7)**
Carbamate and Carbamoyloxime Insecticides: Single-class, Multiresidue Analysis of • High-performance

Liquid Chromatography/Mass Spectrometry Methods in Pesticide Analysis

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Mass Spectrometry, Low-resolution Electron Impact, for the Rapid Analysis of Petroleum Matrices

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Gas Chromatography **(Volume 12)**
Hyphenated Gas Chromatography • Instrumentation of Gas Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. M. Nappi, C. Weil, C.D. Cleven, L.A. Horn, H. Wollnik, R.G. Cooks, 'Visual Representations of Simulated Three-dimensional Ion Trajectories in an Ion Trap Mass Spectrometer', *Int. J. Mass Spectrom. Ion Processes*, **161**, 77–85 (1997).
2. G.C. Stafford Jr, P.E. Kelley, J.E.P. Syka, W.E. Reynolds, J.F.J. Todd, 'Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology', *Int. J. Mass Spectrom. Ion Processes*, **60**, 85–98 (1984).
3. W. Paul, H. Steinwedel, 'Apparatus for Separating Charged Particles of Different Specific Charges', German Patent 944,900,1956; US Patent 2,939,952, 1960.
4. W. Paul, 'Electromagnetic Traps for Charged and Neutral Particles', *Angew. Chem.*, **29**, 739–748 (1990).
5. P.H. Dawson, N.R. Whetten, 'Radio Frequency Quadrupole Mass Spectroscopy', *Adv. Electron. Electron Phys.*, **27**, 58–158 (1969).
6. P.H. Dawson, *Quadrupole Mass Spectrometry and Its Applications*, Elsevier, Amsterdam, 1976.
7. R.E. March, R.J. Hughes, J.F.J. Todd, *Quadrupole Storage Mass Spectrometry*, Wiley Interscience, New York, 1989.
8. J.F.J. Todd, 'Ion Trap Mass Spectrometer – Past, Present and Future(?)', *Mass Spectrom. Rev.*, **10**, 3–52 (1991).
9. B.D. Nourse, R.G. Cooks, 'Aspects of Recent Developments in Ion Trap Mass Spectrometry', *Anal. Chim. Acta*, **228**, 1–21 (1990).
10. R.G. Cooks, R.E. Kaiser Jr, 'Quadrupole Ion Trap Mass Spectrometry', *Acc. Chem. Res.*, **23**, 213–219 (1990).
11. G.L. Glish, S.A. McLuckey (eds.), 'Quadrupole Ion Traps', *Int. J. Mass Spectrom. Ion Processes*, **106** (1991).
12. R.E. March, 'Ion Trap Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **118/119**, 71–135 (1992).
13. R.E. March, J.F.J. Todd (eds.), *Practical Aspects of Ion Trap Mass Spectrometry*, Modern Mass Spectrometry Series, Vol. 1, Fundamentals of Ion Trap Mass Spectrometry, Vol. 2, Ion Trap Instrumentation, Vol. 3, Chemical, Biomedical, and Environmental Applications, CRC Press, Boca Raton, FL, 1995.
14. R.E. March, 'An Introduction to Quadrupole Ion Trap Mass Spectrometry', *J. Mass Spectrom.*, **32**, 351–369 (1997).
15. R.D. Knight, 'The General Form of the Quadrupole Ion Trap Potential', *Int. J. Mass Spectrom. Ion Phys.*, **51**, 127–131 (1983).
16. E. Mathieu, 'Mémoire sur le Mouvement Vibratoire d'une Membrane de Forme Elliptique', *J. Math. Pures Appl. (J. Liouville)*, **13**, 137 (1868).
17. H.-P. Reiser, R.K. Julian Jr, R.G. Cooks, 'A Versatile Method of Simulation of the Operation of Ion Trap Mass Spectrometers', *Int. J. Mass Spectrom. Ion Processes*, **121**, 49–63 (1992).
18. M. Splendore, Personal Communication.
19. R.E. Kaiser Jr, R.G. Cooks, G.C. Stafford Jr, J.E.P. Syka, P.E. Hemberger, 'Operation of a Quadrupole Ion Trap Mass Spectrometer to Achieve High Mass/Charge Ratios', *Int. J. Mass Spectrom. Ion Processes*, **106**, 79–115 (1991).
20. J.C. Schwartz, J.E.P. Syka, I. Jardine, 'High-resolution on a Quadrupole Ion Trap Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **2**, 198–204 (1991).
21. M. Splendore, M. Lausevic, Z. Lausevic, R.E. March, 'Resonant Excitation and/or Ejection of Ions Subjected to DC and RF Fields in a Commercial Quadrupole Ion Trap', *Rapid Commun. Mass Spectrom.*, **11**, 228–233 (1997).
22. M. Splendore, F.A. Londry, R.E. March, R.J.S. Morrison, P. Perrier, J. André, *Int. J. Mass Spectrom. Ion Processes*, **156**, 11–29 (1996).
23. M. Splendore, J.B. Plomley, R.E. March, R.S. Mercer, 'Tandem Mass Spectrometric Determination of Polychlorodibenzo-*p*-dioxins and Polychlorodibenzofurans in a Quadrupole Ion Trap Using Multi-frequency Resonant

Excitation', *Int. J. Mass Spectrom. Ion Processes*, **165/166**, 595–609 (1997).

24. M. Lausevic, J.B. Plomley, X. Jiang, R.E. March, C.D. Metcalfe, 'Analysis of Polychlorinated Biphenyls by Quadrupole Ion Trap Mass Spectrometry: Part 1 – Chemical Ionization Studies of Co-eluting Congeners 77 and 110', *Eur. Mass Spectrom.*, **1**, 149–159 (1995).

Secondary Ion Mass Spectrometry as Related to Surface Analysis

Scott D. Hanton and Paula A. Cornelio Clark
Air Products and Chemicals, Inc., Allentown, USA

1 Introduction **1**
2 Surface Analysis **2**
2.1 Dynamic Secondary Ion Mass Spectrometry 3
2.2 Static Secondary Ion Mass Spectrometry 5
3 Instrumentation **6**
3.1 Primary Ion Sources 6
3.2 Mass Spectrometers 8
3.3 Imaging 10
3.4 Charge Compensation 11
4 Sample Preparation **11**
4.1 Sample Cleanliness 11
4.2 Analysis of the Air/Solid Interface 12
4.3 Analysis of Solid/Solid Interfaces 12
4.4 Bulk Organic Molecule Measurements 12
5 Secondary Ion Formation **13**
5.1 Desorption 13
5.2 Energy Distribution 15
5.3 Ionization Mechanisms 15
6 Applications **16**
6.1 Dynamic Secondary Ion Mass-spectrometry Depth Profiling 16
6.2 Static Secondary Ion Mass-spectrometry Analysis of Poly(vinyl chloride) Sheets Using Time-of-flight Secondary Ion Mass Spectrometry 16
6.3 Static Secondary Ion Mass-spectrometry Analysis of Silicon Wafers Using Time-of-flight Secondary Ion Mass Spectrometry 17
6.4 Imaging 17
6.5 Bulk Polymer Molecular Weight Analysis 19
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **21**
References **21**

Many important processes such as catalysis, corrosion, and adhesion depend on the elemental and molecular composition of the surface or interfacial region. The focus of this article is specifically surface applications of secondary ion mass spectrometry (SIMS). Conceptually, SIMS is a very simple technique consisting of a primary ion bombardment step, an energy transfer step, a particle desorption step, and an ion detection step. SIMS experiments are carried out in either a dynamic or a static mode. The two modes are distinguished on the basis of ion dose. Dynamic SIMS uses high primary ion beam doses and is generally used for elemental depth profiling. Static SIMS uses low primary ion beam doses ($\leq 1 \times 10^{12}$ ions cm^{-2}) and is generally used for the chemical characterization of surfaces. The major advantages of SIMS include high sensitivity (parts per million detection limits for some elements), the ability to obtain molecular information, isotopic analysis, lateral characterization (imaging), and analysis of low-atomic-number elements such as H and Li. These characteristics combine to make SIMS a very powerful analytical tool. The key disadvantages of SIMS are that the instrumentation is relatively expensive and that the data can be complex. Related analytical techniques for surface analysis include X-ray photoelectron spectroscopy (XPS), Auger electron spectroscopy (AES), Scanning Auger Microscopy (SAM), and laser desorption mass spectrometry (LDMS).

1 INTRODUCTION

Many important processes, such as catalysis, corrosion, and adhesion, depend on the elemental and molecular composition of the surface or interfacial region. For example, the anticorrosive nature of 316L stainless steel is due to a 100 Å Cr-rich surface layer. Surface analysis can refer to a broad range of physical and spectroscopic techniques. Each of these techniques has advantages and disadvantages and it is significant to emphasize the power of a multitechnique approach in industrial problem solving. The focus of this article is specifically surface applications of SIMS. Conceptually, SIMS is a very simple technique consisting of a primary ion bombardment step, an energy transfer step, a particle desorption step, and an ion detection step. The major advantages of SIMS include high sensitivity (parts per million detection limits for some elements), the ability to obtain molecular information, isotopic analysis, lateral characterization (imaging), and analysis of low-atomic-number elements such as H and Li. These characteristics combine to make SIMS a very powerful analytical tool. Initially used as an ion microprobe technique, SIMS has played an important role in the semiconductor industry. Today, thanks to the pioneering efforts of Benninghoven,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © Copyright Property of Air Products & Chemicals, Allenton, USA

Hercules, Briggs, and many others, SIMS is finding wide application in industrial polymer analysis.

2 SURFACE ANALYSIS

We must begin by answering three fundamental questions: what is a surface?; how deep is the surface region?; and why are surfaces technologically important?

What is a surface? From first principles, we know that the four states of matter in the universe are solid, liquid, gas, and plasma (ionic species). None of these phases is infinite. The region of a sample where one phase ends and another begins is called the *surface phase* or *interfacial region*. The possible boundaries can be gas–liquid, gas–solid, liquid–liquid, liquid–solid, and solid–solid.[1–3]

How deep is the surface region? It takes work to bring a molecule from the interior of the material to the surface.[3–5] This work is the surface free energy. The work arises from the fact that the molecule experiences an imbalance in forces as it moves to the surface. Note that a molecule in the bulk experiences no such imbalance in forces. The imbalance in forces results in ''surface tension''.[3] Surface tension can be thought of as the horizontal pull arising from the work to raise a molecule out of the bulk to the surface.

In the 1800s, Laplace postulated that the depth of the ''unbalanced forces'', or the depth of the surface region, depends on short-range intermolecular forces.[5] With the exclusion of ionic species, the principal forces between molecules are van der Waals forces. These forces decrease with about the seventh power of the intermolecular distance! Hence the interactions between nearest neighbors are the most important. A molecule experiences essentially symmetrical forces once it is a few molecular diameters away from the surface.

Why are surfaces technologically important? Many important processes, such as catalysis, corrosion, and adhesion, depend on the elemental and molecular composition of the interfacial region. For example, the anticorrosive nature of 316L stainless steel is due to an enrichment of Cr at the surface. The Cr to Fe ratio at the surface is about 2 : 1 compared with the bulk ratio of about 0.20 : 1, as measured by AES. This Cr-enriched layer is only about 100 Å (or 0.01 μm) thick.

Surface analysis can refer to a broad range of physical and spectroscopic methods. A summary of some of the commonly employed surface analytical methods are provided in Tables 1 and 2. Each of these techniques has advantages and disadvantages. Each also has a sampling depth or information depth. The sampling depth will determine how much of the surface is measured. The focus of this article is specifically surface applications of SIMS in the analysis of surfaces.

Figure 1 shows a schematic diagram of a typical SIMS experiment.

Conceptually, SIMS is a very simple technique consisting of a primary ion bombardment step, an energy transfer step, a particle desorption step, and an ion detection step. The sample (usually a solid) is introduced into an ultrahigh vacuum (UHV) chamber at pressure

Table 1 Techniques for surface analysis – common acronyms

Acronym	Technique	Analytical information
–	Adsorption isotherms	Surface area and pore-size distributions
AES/SAM	Auger electron spectroscopy/scanning Auger microscopy	Elemental composition, depth profiling and mapping of metals and semiconductors
AFM	Atomic force microscopy	Surface topography
EDX	Energy-dispersed X-rays	Elemental composition
ESCA	Electron spectroscopy for chemical analysis[a]	See XPS
–	Ellipsometry	Layer thickness
FESEM	Field emission scanning electron microscopy	Morphology and elemental composition
ISS	Ion scattering spectroscopy	Atomic arrangement and composition of atoms within top atomic layer
MESIMS	Matrix-enhanced secondary ion mass spectrometry	Elemental and molecular composition
RBS	Rutherford backscattering spectrometry	Composition and depth distribution of elements
SEM	Secondary electron microscopy	Morphology
SIMS	Secondary ion mass spectrometry	Elemental and molecular composition
SNMS	Secondary neutral mass spectrometry	Elemental and molecular composition
STM	Scanning tunneling microscopy	Morphology
TEM	Transmission electron microscopy	Morphology and elemental composition
UPS	Ultraviolet photoelectron spectroscopy	Valence band structure
XPS	X-ray photoelectron spectroscopy[a]	Elemental composition, oxidation state of metals, catalysts, semiconductors, and polymers

[a] XPS and ESCA refer to the same technique.

Table 2 Techniques for surface analysis – technical summary[a]

Technique	Signal	Elements	Quantitative	Accuracy (%)	Detection limits	Depth	Lateral resolution
AES/SAM	Auger electrons	Li–U	With standards	±20	0.1–1 at %	2–10 nm	100 nm
ESCA/XPS	Photoelectrons	Li–U	With standards	±5–50	0.1–1 at %	5–10 nm	200 μm
RBS	He particles	Li–U	Without standards	±2	1–10 at % ($Z < 20$); 0.01–0.001 at % ($Z > 70$)	5–700 nm	200 μm
SEM/EDX	Secondary electrons/X-rays	Na–U	EDX with standards	±15	0.5 at %	1–5 μm	SEM 5 nm
			EDX without standards	±20–50			EDX 1 μm
SIMS (Magnetic sector)	Secondary ions	H–U	With standards	±5–50	10^{12}–10^{16} at cm^{-3}	5–30 nm	100 nm
SIMS (Quadrupole)	Secondary ions	H–U	With standards	±1–20	10^{14}–10^{16} at cm^{-3}	5–30 nm	5 μm
SIMS/(TOF)[b]	Secondary ions	H–U	With standards	±100	<1 ppm	0.1 monolayer	1 μm
SNMS	Secondary atoms	Li–U	With standards	±5	10–100 ppm	3 nm	5 mm

[a] Technical Summary from Charles Evans and Associates, Redwood City, CA, 1990.
[b] TOF, time-of-flight.

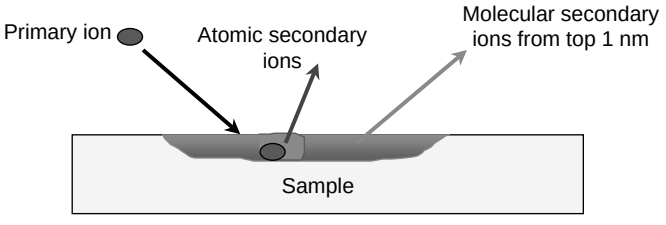

Figure 1 Schematic diagram of SIMS process.

$\leq 1 \times 10^{-9}$ Torr (1 Torr = 133.3 Pa). The sample is bombarded by a primary ion beam (typically Xe, Ar, Ga, or Cs ions) accelerated to 5–25 kV. The primary ion impacts the surface and penetrates about 100 Å into the sample. The impact of the primary ion results in an energy and momentum transfer process called the collision cascade. The energy and momentum transfer result in the desorption (sputtering) of neutral species, electrons, and secondary ions from the surface of the sample. The secondary ions are mass analyzed. The major advantages of SIMS include, high sensitivity (parts per million detection limits for some elements), the ability to obtain atomic and molecular information, isotopic analysis, lateral characterization (imaging), and analysis of low atomic number elements such as H and Li.[6,7]

SIMS experiments are carried out in either a dynamic or a static mode. In the early 1980s, the two modes were distinguished on the basis of primary ion beam current density. Today, however, it is more common to distinguish the two modes on the basis of ion dose. Dynamic SIMS uses high primary ion beam doses and is generally used for elemental depth profiling. Static SIMS uses low primary ion beam doses ($\leq 1 \times 10^{12}$ ions cm^{-2}). In static SIMS each sputtering event is thought to originate from a virgin surface and the SIMS spectrum is characteristic of the top few atomic layers.[6,8]

2.1 Dynamic Secondary Ion Mass Spectrometry

In its early applications, SIMS was referred to as an ion probe technique and found applications in elemental compositional analysis.[9] Detection limits in the parts per billion (ppb) range were achieved by using a high primary ion flux. The high primary ion flux also ensured the removal of a number of monolayers per second, making dynamic SIMS well suited for elemental depth profiling. Dynamic SIMS is an extremely valuable tool to the semiconductor industry.

Ideally, the number of ions sputtered from the sample should be linearly related to concentration, independent of the element, independent of the sample being analyzed, and independent of the primary bombarding species. Unfortunately, none of these conditions are met in SIMS. The linear dependence of signal is observed only for dilute systems, secondary ion yields vary greatly for different elements, secondary ion yields vary greatly for the same element in different matrices (matrix effects), and secondary ion yields depend upon the kinetic energy, size, and incident angle of the primary ion. Figure 2 illustrates the variation in secondary ion yields for various elements.[10,11]

Secondary ion yields are strongly influenced by the chemical and electrical properties of the surface.[9] Benninghoven showed that secondary ion yields for elements

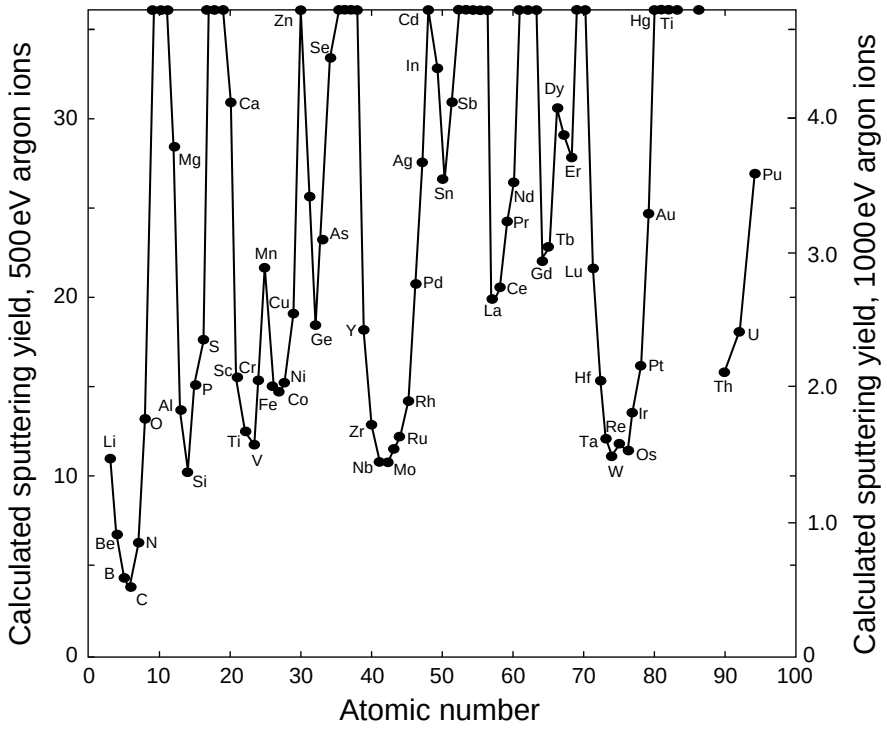

Figure 2 Variation in sputter yields for different elements. (Reprinted from M.P. Seah, *Thin Solid Films*, **81**, 279 (1981), Copyright 1981, with permission from Elsevier Science and *Practical Surface Analysis – Auger and XPS*, 1st edition, D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

sputtered from clean metal surfaces are significantly lower than those of the same elements sputtered from an oxidized surface.[12]

In SIMS, one often observes significant variations in secondary ion intensity for a given element in different matrices. Figure 3(a–c) illustrate this matrix effect.[9] Figure 3(a) shows the depth profile of Be in GaAs. Figure 3(b) shows a depth profile of Mg, and Figure 3(c) shows Be implanted through a Si_3N_4 layer on GaAs. Ideally, the distribution of the Mg and Be species should be continuous with the ion implant shape (as in Figure 3a). In the case of Be, however, we observe a significant decrease in the secondary ion intensity within the GaAs matrix. The Mg secondary ion intensity is not strongly affected by the GaAs matrix. Matrix effects are an issue in all SIMS experiments and can be a particular problem in the quantitative analysis of multiphase systems.

Despite these technical issues, SIMS can provide both a qualitative and a quantitative description of the sample in certain instances.[9] For example, quantitative analysis of homogeneous systems such as glasses and semiconductors is possible, provided that calibration standards are available. In these cases, the secondary ion signal, I_i, is proportional to the primary ion flux, I_p.[13] The secondary ion signal is related to concentration by Equation (1):

$$I_i = (AX_iY_j\varepsilon_{i,j}\Gamma_{i,j}\beta_{i,j}T_{i,j}D_i)I_p \qquad (1)$$

where A is the fractional abundance of the isotope sampled, X_i is the average concentration of the element in question within the sputtered atom escape depth (weighted by the relative probabilities of sputtered atom escape from the surface and deeper layers), Y_j is the sputtering yield of the sample (i.e. the total number of atoms of all species sputtered per incident primary ion), $\varepsilon_{i,j}$ is the fraction of element i that is sputtered as the species monitored, $\Gamma_{i,j}$ is a preferential sputtering term, $\beta_{i,j}$ is the ionization probability of the sputtered element, $T_{i,j}$ is the mass spectrometer transmission efficiency, and D_i is the efficiency of the mass spectrometer detector for different elements. Williams[13] gave a more detailed discussion of the terms in Equation (1).

Despite the fact that researchers have a good understanding of the qualitative aspects of the ionization and the charge-transfer processes in SIMS, a quantitative theory of SIMS still does not exist. Thus, for practical quantitative analysis, one must rely on standards which model the analytical sample. If this requirement is met, all the variables in Equation (1) are calibrated together in a single sensitivity factor, K_i, described by

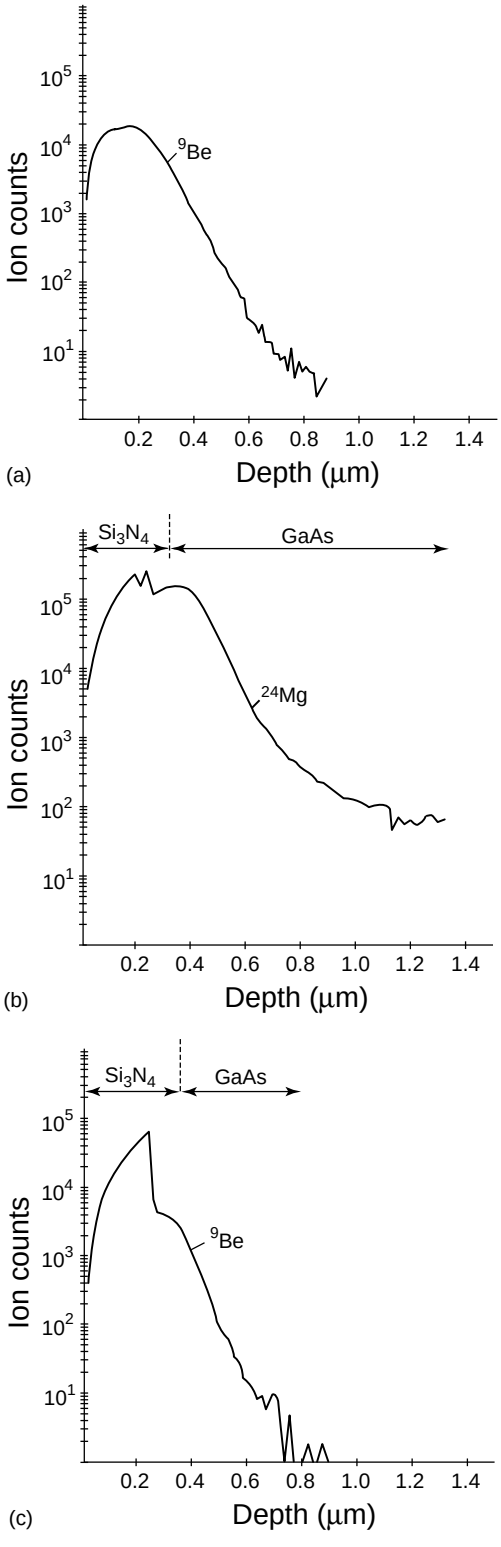

Figure 3 Dynamic SIMS depth profile of (a) Be implanted into GaAs, (b) Mg implanted into Si_3N_4, and (c) Be implanted into Si_3N_4. The Be profile shows a significant drop in the secondary ion intensity in the GaAs layer due to matrix effects. (Reproduced with the permission of Cambridge University Press.)

Equation (2):[13]

$$X_i = \frac{K_i I_i}{A} \tag{2}$$

The proportionality constant depends on the element of interest and the analytical conditions. The data are typically normalized to the secondary ion signal from a matrix element (I_m) to minimize the influence of analytical conditions (Equation 3):

$$X_i = \frac{K_{i/m} I_i}{A I_m} \tag{3}$$

The relative sensitivity factor, $K_{i/m,}$ is valid only when X is a dilute solute in the matrix. Generally, the accuracy of quantitative analyses carried out in this manner is ±10%. Owing to the matrix effects, it is important that the standards are produced from the same matrix as the samples to be analyzed. The most common standards are specific metal alloys developed as National Institute of Standards and Technology (NIST) standards, modified glass materials produced with known levels of impurities, and ion implantation standards.[14] Ion implantation standards have been particularly useful in the semiconductor industry for depth profiling of silicon wafers. Newbury[14] gives a more detailed discussion of the role of standards in SIMS.

2.2 Static Secondary Ion Mass Spectrometry

In the 1960s, Alfred Benninghoven et al. showed that it was possible to maintain the surface integrity during a SIMS experiment over long periods of time by decreasing the primary ion fluence.[6,15] In this way, one could preserve the molecular composition of the surface. As described earlier, primary particle bombardment results in (1) the creation of a perturbed region within the surface region (collision cascade) and (2) secondary emission (ions, neutrals, and electrons).[6] After prolonged primary ion bombardment, the surface will be covered by a large number of ''perturbation areas,'' σ, corresponding to individual ion impacts. The goal in static SIMS is to avoid secondary ion emission from these perturbed or prebombarded areas. This is typically achieved by limiting primary ion bombardment so that $P \ll 1$ (Equation 4):

$$P = \frac{\sum \sigma}{A} \tag{4}$$

where P is the probability for prebombardment, σ is the damage cross-section and A is the bombarded target area. The static SIMS limit is then achieved by limiting primary ion beam exposure to no more than 1×10^{12} ions cm^{-2}.

3 INSTRUMENTATION

The main components of a SIMS instrument include the vacuum system, sample introduction, sample manipulation, primary ion source and primary ion optics, secondary ion extraction and transfer optics, energy filter, mass spectrometer, detector, and data acquisition.[16] Modern SIMS instruments are designed to operate under UHV conditions ($\leq 1 \times 10^{-9}$ Torr). UHV conditions are required to prevent the loss of secondary ion signal from in-vacuum collisions. Turbomolecular pumps, ion pumps, and titanium sublimators are employed to achieve UHV. Modern SIMS instruments offer rapid sample introduction through a gate-valve-protected sample air lock; in this way, a sample can be quickly exchanged without adversely affecting the main chamber vacuum. Extensive discussions of vacuum components are readily available.[10] Sample manipulation devices are also available to tilt or rotate a sample, and to cool or heat the sample.[16]

Primary ion sources are either reactive or inert gas-phase sources (electron impact or plasma), surface ionization sources, or liquid metal field emission sources. The primary ion optics are used to extract ions from the source, filter out impurities, accelerate and focus the beam toward the sample, and raster the beam over the sample. The secondary ion extraction and transfer optics collect the secondary ions from a defined area inside the sputter region and transfer the ions to the mass filter. Energy filters are used to select ions from the broad energy distribution of the secondary ions sputtered from the surface and to eliminate scattered primary ions.

The mass spectrometer functions to separate the ions of differing mass. SIMS instruments usually employ either a quadrupole mass filter, magnetic sector field, or TOF mass spectrometer. The detector schemes may consist of either a channeltron, Faraday cup, CuBe secondary electron multiplier, or single- or dual-microchannel plate with either a fluorescent screen and camera or resistive anode encoder. Data acquisition in commercial SIMS instruments is controlled by a computer. Imaging SIMS instruments require dedicated data systems with large memories to store intensity distributions of several mass lines. Jede et al.[16] gives a more detailed discussion of instrument components.

3.1 Primary Ion Sources

In general, common primary ion sources include (a) the electron impact ionizer, (b) the plasma ion source, (c) the surface ionization source, and (d) the liquid metal field emission ion source.[16]

3.1.1 Electron Impact Ionizer

A schematic diagram of a gas-phase primary ion gun equipped with an extractor ion source is shown in Figure 4.[16] In this ion gun, an ionizer is placed inside a grounded, differentially pumped housing. A small, grounded aperture between the ionizer/extraction region and the lens arrangement gives a pressure step of three orders of magnitude. The extracted ions are focused by an Einzel lens to form a cross-over in the aperture plane. The tube lens images this cross-over to the sample surface. The typical operating parameters for an extractor type ion source are summarized in Table 3.[16]

3.1.2 Plasma Ion Sources

Plasma ion sources are characterized by the formation of a dense plasma from which ions are electrostatically extracted through a small orifice. The plasma is formed by a low-voltage arc between a cathode and an anode.[16] Because of the high density of charged particles, plasma ion sources are brighter than primary electron impact ionizer sources. With a brighter, higher primary ion flux to the sample, plasma sources have been used extensively in dynamic SIMS applications. The most common plasma

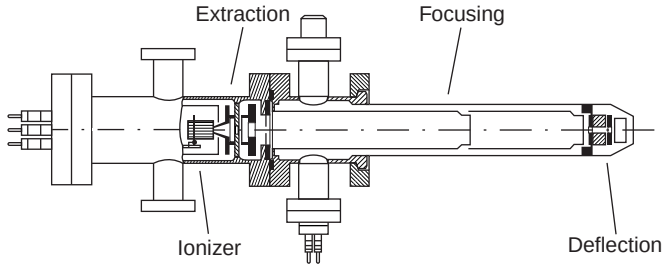

Figure 4 Schematic diagram of gas-phase primary ion gun equipped with an extractor ion source. (*Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

Table 3 Operating parameters for common sources

Parameter	Electron ionization	Plasma	Surface ionization	Liquid metal
Source (pressure)	Ar, Xe (10^{-7}–10^{-4} mbar)	Ar, Xe, Ne, O_2 (10^{-2} mbar)	Cesium	^{69}Ga
Acceleration voltage	0.1–5 keV	10 keV	10–20 keV	5–35 keV
Beam size at sample	0.1–2 mm	2–100 μm	≥200 nm	<100 nm
Current density	4 mA cm^{-2}	20 mA cm^{-2}	100 mA cm^{-2}	1000 mA cm^{-2}

ion sources are the hot-filament duoplasmatron source and the cold-cathode duoplasmatron source. In the former case, the cathode consists of a heated tungsten or tantalum filament. The hot filament is used with noble gas (e.g. Ar, Xe, and Ne). In dynamic SIMS applications, the cold-cathode source can be used with O_2 to enhance positive ion yield.

A schematic diagram of a cold-cathode duoplasmatron ion source is shown in Figure 5.[16,17] The plasma is constricted by a magnetic field generated by a coil. The primary ions are extracted by a grounding electrode and focused by electrostatic condenser and objective lenses in the optical column. Beam scanning is accomplished by a double-deflection system. The plasma drastically enhances charged particle density and, thus, the primary ion source brightness. This type of ion gun is commonly used on dynamic SIMS instruments. The typical operating parameters for a duoplasmatron ion source are summarized in Table 3.[16]

3.1.3 Surface Ionization Ion Sources

Surface ionization sources are also important primary ion sources for SIMS because (a) these sources have higher particle density (i.e. high brightness) as compared with duoplasmatron sources and (b) in the case of Cs^+ sources, cesium deposition on the surface can enhance negative ion yields.

A schematic diagram of the cesium surface ionization source is shown in Figure 6.[16,18] For this source, cesium

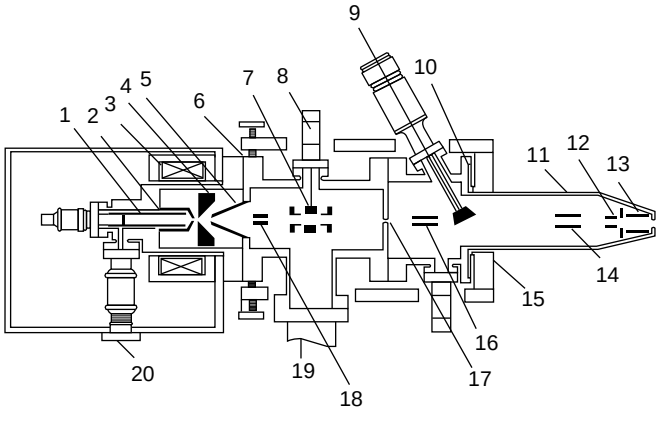

1 Cold cathode
2 Intermediate electrode
3 Magnetic field coil
4 Anode
5 Extractor
6 Source flange
7 Condenser lens
8 High-voltage feedthrough
9 Variable aperture
10 Source port aligner
11 2 degree bend
12 Beam scanning octupole
13 Objective lens
14 Beam scanning quadrupole
15 Mounting flange
16 Beam alignment and blanking quadrupole
17 Pressure step
18 Beam alignment quadrupole
19 Differential pumping port
20 Gas inlet port

Figure 5 Schematic diagram of a cold cathode duoplasmatron primary ion source. (Reproduced by permission of Physical Electronics Inc and *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

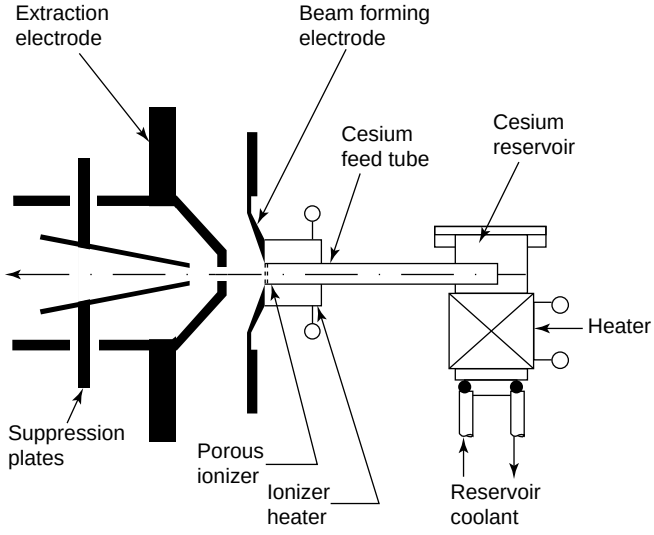

Figure 6 Schematic of a cesium surface ionization primary ion source. (Reprinted with permission from P. Williams, R.K. Lewis, C.A. Evans, Jr.[18] Copyright (1977) American Chemical Society and *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

vapor from a heated reservoir is led through a heated duct closed by a porous tungsten plug. The Cs^+ ions are vaporized by the outside of the plug and are accelerated by an electrical field between the extraction electrode and the suppression electrode. In comparison with duoplasmatron sources, Cs^+ sources have a higher primary ion flux and Cs^+ deposition on the surface dramatically increases the negative secondary ion yield. Cesium surface ionization sources are well suited for dynamic SIMS applications. The typical operating parameters for a Cs^+ source are summarized in Table 3.[16]

3.1.4 Liquid Metal Field Emission Ion Source

The advent of liquid metal field emission ion sources has brought improvements in microprobe and imaging SIMS.[16] A schematic of a liquid metal field emission ion source, or liquid metal ion gun (LMIG) is shown in Figure 7(a–d).[16]

The basic principle of operation is based on the field ion emission process where the surface potential barrier of a metal is modified by very high electric fields. When electrons tunnel into the solid, surface atoms are ionized and desorb in the electric field (field evaporation), as illustrated. The most common liquid metal source is isotopically enriched ^{69}Ga. The typical operating parameters for an LMIG source are summarized in Table 3.[16]

Note that, in time-of-flight secondary ion mass spectrometry (TOFSIMS) instruments the primary ion gun

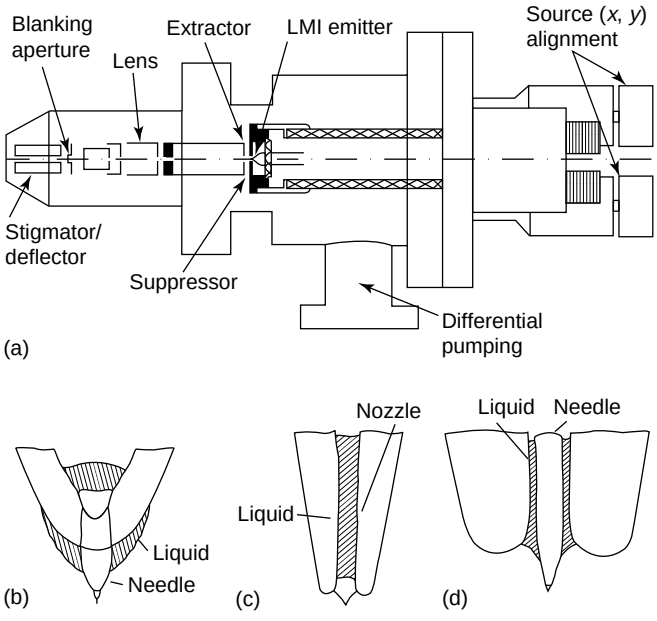

Figure 7 Schematic diagrams of (a) a LMIG and liquid metal ion sources: (b) needle/filament type, (c) nozzle type, and (d) needle nozzle type. (*Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

must be pulsed to produce a time-resolved ''package'' of primary ions. Pinkston et al.[19] gives an additional discussion of the pulsing schemes necessary for pulsed primary ion sources.

3.2 Mass Spectrometers

Three different types of mass spectrometers are commonly used in SIMS instruments: quadrupole mass filters, magnetic sector mass spectrometers, and TOF mass spectrometers.

3.2.1 Quadrupole Mass Filters

A schematic diagram of a quadrupole mass filter SIMS instrument is shown in Figure 8.[16]

Quadrupole mass filters work like tuneable, variable-band-pass filters. Only ions within a narrow mass region of <1 u are allowed to pass through the mass filter to the detector. Separation over a given mass range is achieved by electronically sweeping the position of the band-pass region.[20] The quadrupole mass filter consists of a set of four electrodes. The filtering action is obtained by the application of a combination of a time-independent (direct current, dc) potential and a time-dependent (radio frequency, rf) potential. By scanning both the dc voltage and the rf, while keeping their ratio constant, ions of different mass-to-charge ratio (m/z) are passed to

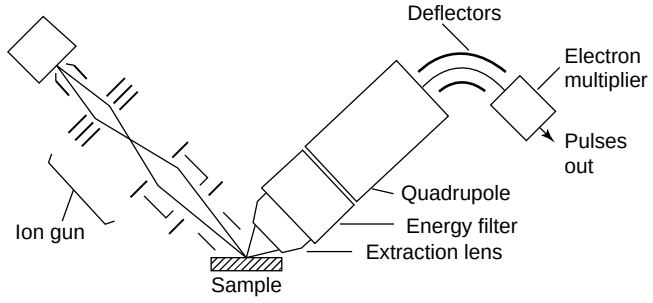

Figure 8 Schematic diagram of quadrupole mass filter SIMS instrument. (*Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

the detector.[21] If the dc/rf ratio is constant, then the operation of the mass filter is defined by a straight line which has a zero intercept. This line is known as the mass scan line. Mass separation is achieved by sweeping the voltage applied to the electrode structure such that an ion with a specific m/z is allowed to pass through the center axis to the detector.

3.2.2 Time-of-flight Mass Spectrometers

The basic principle of TOF mass separation is that ions separate according to their velocity in a field-free region. A basic time-of-flight mass spectrometry (TOFMS) system consists of an ion source, ion focusing optics, a field-free drift region and a time-sensitive detector. Since TOFMS instruments measure the flight time of ions, the start of the flight must be known. To produce and coordinate the start time, the primary ion gun is pulsed to produce a time-resolved ''package'' of primary ions which impinge on the sample. The secondary ions are extracted into the drift tube. Secondary ions are accelerated with a constant kinetic energy, V_0 (Equation 5), mass separate according to their velocity (lighter ions striking the detector first, followed by heavier ions) (Equation 6), and are detected after their full TOF (Equation 7):

$$V_0 = \frac{1}{2mv^2} \tag{5}$$

$$v = \left(\frac{2zeV_0}{m}\right)^{1/2} \tag{6}$$

$$t = \frac{L}{v} \tag{7}$$

where t is the ion flight time, L is the field-free flight path length, V_0 is the accelerating voltage, e is the unit electronic charge, and z is the number of charges on the ion. The arrival time is then converted to mass via calibration.

In TOFMS, the mass resolving power, R, is determined by the width of the ion peaks. Typically, full width at

half-maximum (fwhm) is used to define the peak width (Equation 8):

$$R = \frac{m}{\Delta m} = \frac{m}{(\mathrm{d}m/\mathrm{d}t)/\Delta t} = \frac{t}{2\Delta t} \qquad (8)$$

where m is the mass of the ion, Δm is the fwhm of the ion peak in mass units, t is the arrival time of the ion, and Δt is the fwhm of the ion peak in time.

Three factors limit mass resolution in TOFSIMS. The first factor is the pulse width of the primary ion source at the sample. The primary ion pulse width introduces variations in secondary ion start time. The second factor is time focusing aberrations of the analyzer. These aberrations contribute to line broadening. The third factor is the rise time and dead time of the detector. A typical mass resolution attainable in TOFSIMS is about 9000 for the $^{28}Si^{+}$ ion desorbed from a clean silicon wafer.

Some of the advantages of TOF mass analysis include very high mass range, high mass resolution, multiplex detection of secondary ions injected into the instrument (i.e. parallel detection) rather than detection of only those for which the analyzer is tuned, and a constant transmission function. In general, TOFSIMS instruments are commonly used for static SIMS applications. Owing to the pulsed nature of the experiment, TOFSIMS instruments are often not well suited for traditional dynamic SIMS depth profiling studies. The duty cycle of the pulsed ion gun is often not sufficient for dynamic SIMS. The dc ion beams used for dynamic SIMS are better suited to the continuous-source mass spectrometers (quadrupoles and magnetic sectors). However, some researchers use TOFSIMS to perform shallow-depth profiling.[22]

3.2.2.1 Reflectron Time-of-flight Secondary Ion Mass Spectrometry A schematic diagram of a reflectron TOFSIMS instrument is shown in Figure 9.[16,23,24] A reflectron is a mechanism to extend the actual flight time of the ions and increase the mass resolution. The secondary ions penetrate into the ion reflector (located at the end of the first drift space). In many reflectron instruments the sample is held at ground potential and the ions are extracted into the drift tube with the opposite sign potential. The ion reflector compensates for flight time dispersion. Faster ions penetrate deeper into and spend more time in the ion reflector field region than slower ions. Thus, a negative flight time dispersion from the initial kinetic energy is compensated for by a positive flight time dispersion in the reflector region. The secondary ions are then registered at the detector. Imaging using a reflectron TOFSIMS system is carried out by raster scanning the primary ion beam across the sample.

3.2.2.2 TRIFT® Time-of-flight Secondary Ion Mass Spectrometry Microscope A schematic of the TRIFT® TOFSIMS instrument is shown in Figure 10.[16,25,26] In this design, the sample is held at high voltage potential and the extraction lens is held at ground potential. In the TRIFT® instrument, an image of secondary ions is formed at the immersion lens and the transfer lens; the secondary ion image is extracted into the flight path region, which is defined by three 90° electrostatic sector analyzers (ESAs) with intermediate drift spaces. A spatially resolved intensity–time image is registered at the detector. This is converted to a spatially resolved intensity–mass image via calibration.

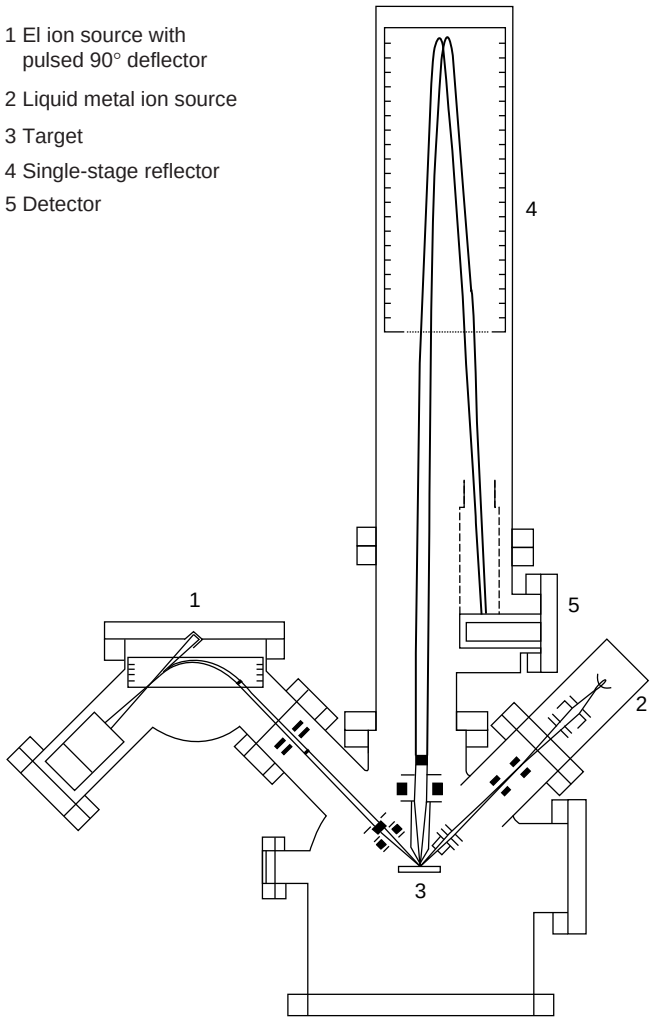

Figure 9 Schematic of reflectron TOFSIMS instrument. (Reprinted with permission from E. Niehuis, T. Heller, H. Feld, A. Benninghoven, Design and Performance of a Reflectron Based TOF Secondary Ion Mass Spectrometer with Electrodynamic Primary Ion Mass Separation, *J. Vac. Sci. Technol. A*, **5**, 1243–1246, Copyright 1987, American Vacuum Society and *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

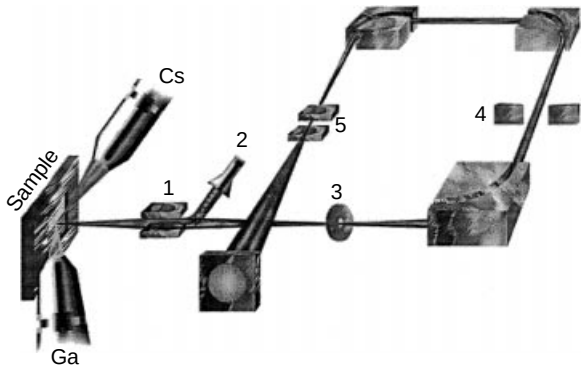

Figure 10 Schematic diagram of TRIFT® TOFSIMS instrument showing dual Cs and Ga primary ion sources, the high-mass blanker (1), the secondary electron detector (SED) (2), the angular filter (3), the energy slit (4), and the post-spectrometer blanking plates (5). (Reproduced by permission of Physical Electronics from 'Unique Features of the TRIFT® II Spectrometer and Their Analytical Benefits'.[25])

Some of the unique features of the TRIFT® instrument are illustrated in Figure 10.[25] A high-mass blanker (1) allows operation at primary ion repetition rates up to 40 kHz without ''wrap-around'' of high-mass ions into the low-mass region. A SED (2) after the extraction lens allows for static secondary electron detection imaging. An angular filter (3) operates at the cross-over of the secondary ion beam and is used to remove high-energy secondary ions with large angles of emission. An energy slit (energy filter, 4), located between the first ESA and second ESA, is used to filter out low-energy metastable ions. Post-ESA blanking plates (5), positioned at the cross-over in the secondary ion beam, are used to reject individual masses from the beam path before they strike the detector. The TRIFT® TOFSIMS system can be operated in microscope mode or in microprobe mode. In microscope mode, the sample is bombarded with a large-beam-size primary ion beam and the secondary ion image is registered at a position-sensitive detector. In microprobe mode, a small-probe-size primary ion beam is raster scanned across the sample and the secondary ions are registered at a multichannel detector. With the development of ^{69}Ga LMIG sources, superior imaging is now achieved with microprobe mode over microscope mode.

3.2.3 Magnetic Sector Field Secondary Ion Mass Spectrometry

A schematic diagram of a double-focusing magnetic sector field SIMS instrument is shown in Figure 11.[15] Magnetic sector mass spectrometers separate ions based on their flight through a magnetic field. The ions are accelerated into the magnetic field (Equation 9) and follow a curved path through the field (Equation 10). The magnetic field strength, B, is scanned

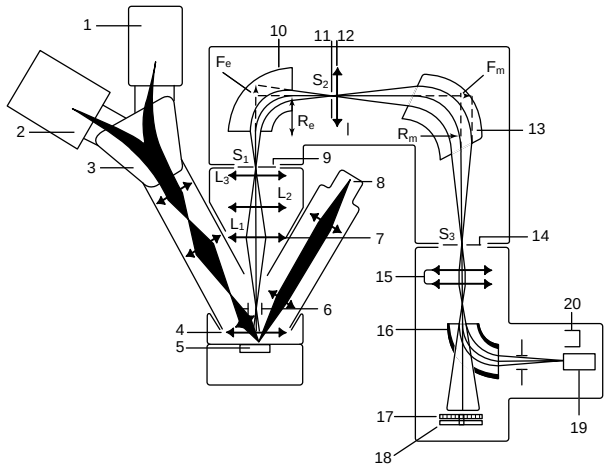

1 Cesium ion source
2 Duoplasmatron ion source
3 Primary beam mass filter
4 Immersion lens
5 Sample
6 Dynamic emittance matching
7 Transfer lens system
8 Liquid metal ion source
9 Entrance slit S_1
10 90° electrostatic sector
11 Energy slit S_2
12 Intermediate lens I
13 90° magnetic sector
14 Exit slit S_3
15 Projection lenses
16 Projection deflector
17 Channel plate
18 Fluorescent screen
19 Electron multiplier
20 Faraday cup

Figure 11 Schematic diagram of magnetic sector field SIMS instrument. (Reproduced by permission of CAMECA and *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, Vol. 2, eds. D. Briggs, M.P. Seah. Copyright John Wiley & Sons Limited. Reproduced with permission.)

to pass individual masses through the exit slits of the instrument (Equation 11).[19]

$$KE = \frac{1}{2mv^2} = zV \tag{9}$$

$$\frac{mv^2}{r} = Bzv \tag{10}$$

$$\frac{m}{z} = \frac{B^2r^2}{2V} \tag{11}$$

In Equations (9–11), KE is the kinetic energy of the ions injected into the magnetic field, m is the mass, z is the charge on the ion, and r is the radius of the curved ion flight path.

Some of the first SIMS instruments were magnetic sector field mass spectrometers.[15] These instruments were used in dynamic SIMS depth profiling and imaging. Magnetic sector instruments are more sensitive than quadrupole instruments and can be operated at very high mass resolution.

3.3 Imaging

SIMS imaging is an important component of many modern instruments. Secondary ion images are

constructed either by raster scanning the primary ion beam across the sample or by using the TRIFT® instrument (discussed above in section 3.2.2) in microscope mode with a position-sensitive detector. With the development of very small-probe-size LMIG sources, raster scanning is the most common technique to acquire images. To collect a raster-scanned image, the position of the primary ion beam is registered along with the mass of each secondary ion detected. From these data secondary ion maps or images can be plotted. These images can be for all of the ions detected (total ion image) or for individual secondary ions. The combination of mass data and lateral position data can be very powerful in analytical problem solving.

3.4 Charge Compensation

SIMS is commonly used to analyze a wide range of materials, including metals, catalysts, semiconductors, and polymers. Special experimental considerations must be given to SIMS of insulating materials. If a primary ion deposits a charge on the surface of a bulk insulator, the material will tend to build up an electrical charge. This static charge will prevent the desorption of any charged particles from the surface of the bulk insulator. Since most primary ion sources deliver positive ions, samples charge positively, so charging problems can be particularly troublesome in the negative SIMS mode. Electron bombardment from an electron flood gun is commonly used to dissipate the positive static charge on the surface.[15,17] In quadrupole-based SIMS instruments, the low-energy electron flood is simply operated simultaneously with the continuous primary ion beam. However, in TOFSIMS instruments, the electron gun must be operated in a pulsed mode and pulses are delivered between primary ion beam pulses. In the case of the TRIFT® TOFSIMS instrument (and also magnetic sector instruments), where the sample floats at high potential, a high-speed switch is used to drop the sample potential to ground during the electron beam pulse, to facilitate charge compensation. Vickerman[15] gives an extensive discussion of charging and charge compensation schemes.

4 SAMPLE PREPARATION

4.1 Sample Cleanliness

Owing to the surface sensitivity of SIMS and other surface analysis techniques, great care must be taken to preserve the original surface chemistry of the sample. Under static SIMS analysis conditions, only the top few monolayers of the sample are analyzed.[27] Hence contamination of a surface can be a significant problem. Surface contamination can be of two types: (a) adventitious carbon and (b) contamination due to improper sample handling and/or storage. The former does not present a major problem, but the latter can be very problematic.

Metals and highly reactive materials readily adsorb contaminants from the atmosphere. In fact, the surface of a ''clean'' metal placed in a UHV system at 10^{-7} Torr will become contaminated in only 1 s.[28] The degree of contamination depends on a number of factors, including the material's tendency to adsorb species either from the ambient atmosphere or in vacuum and the cleanliness of the vacuum chamber. Adventitious carbon contamination is commonly observed on metals and is considered to be saturated hydrocarbons.[29] Although surface techniques ''sample'' this adventitious carbon layer, the layer is thin enough that it does not completely attenuate the signal from the metal surface below.

Improper sample handling and storage can be problematic. Samples should not be handled by hand because the human fingerprint can readily contaminate the sample. Figure 12 shows a static SIMS spectrum of one of the author's thumbprints taken from a contaminated silicon wafer.

A substantial amount of contamination can attenuate the detection of secondary ions from the sample. Samples should be prepared with tools (scissors, tweezers, etc.) which are routinely cleaned with organic solvents (e.g. hexane, chloroform, methanol) to remove contamination. Improper sample storage can also lead to contamination. For example, samples stored in certain plastic bags can pick up processing and compounding aids such as *N*-stearyleurcamide or phthalate plasticizers. Figure 13 shows a TOFSIMS spectrum of a plastic bag, where

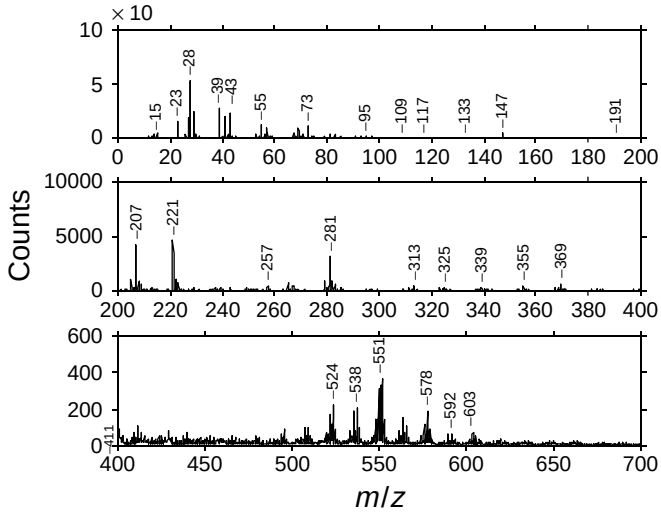

Figure 12 Static SIMS spectrum of one of the author's thumbprints on a silicon wafer.

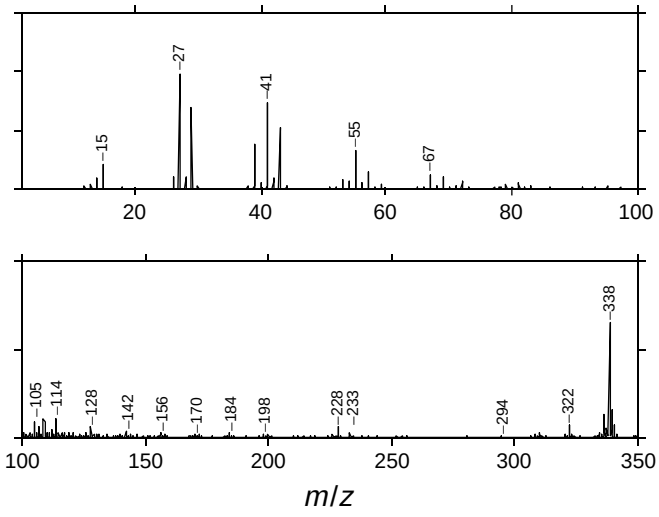

Figure 13 Static SIMS spectrum of a plastic bag.

the peak at m/z 338 is the $[M + H]^+$ molecular ion for *N*-stearyleurcamide.

Despite the potential sample cleanliness issues, static SIMS can be a very powerful tool for industrial problem solving. Proper sample handling and storage must be considered an integral part of good laboratory practice. An American Society for Testing and Materials (ASTM) guide to handling samples for surface analysis is available.[30]

4.2 Analysis of the Air/Solid Interface

SIMS is applicable to a broad range of materials. SIMS spectra can routinely be obtained from materials such as cured or coalesced polymer coatings, fabrics, silicon wafers, metals, catalysts, and powders. In many of these cases, solid samples are simply mounted on a sample holder and analyzed as received. Supported catalysts and powders can be pressed on to double-sided tape or pressed into cleaned In foil. Again, the tape and In foil should be characterized to ensure that the supporting media do not contribute to sample contamination. SIMS instruments can also be equipped with cryogenic sample introduction and cryogenic stages. In this way, SIMS can be applied to materials with measurable vapor pressures at room temperature.

4.3 Analysis of Solid/Solid Interfaces

Analysis of the solid/solid interface presents some unique challenges since the interface of interest may be well beyond the actual sampling depth of SIMS. Special sample preparation methods such as crater edge profiling, angle lapping, and ball cratering are well known in SAM.[10] The general principle here is to lap or cut a sample at a small tilt angle so as to expose the various layers for analysis by an analytical technique with good lateral resolution. Metallic multilayer structures have been prepared in this manner and characterized. Recently, Gardella et al. extended the crater edge profiling methods to polymers, with some success.[31] Several researchers have used traditional cryo-fracturing and cryo-ultramicrotomy specimen preparation methods. Cryo-ultramicrotomy is a tool commonly used in the preparation of ultrathin sections for TEM.

4.4 Bulk Organic Molecule Measurements

While SIMS is a powerful tool for traditional surface analysis, it is also an effective mass spectrometry tool for organic and oligomer molecular weight distribution analysis. A typical sample preparation method for a low-molecular-weight oligomer sample starts with a dilute solution of the analyte in an organic solvent. For example, a low-molecular-weight oligomer sample of polystyrene (PS) 2450 can be prepared for SIMS analysis by preparing a $1\,mg\,mL^{-1}$ solution in tetrahydrofuran (THF). The substrate for this experiment is a cleaned piece of silver foil. The silver foil can be cleaned by traditional acid oxidation or by lightly rubbing with a fine-grit emery paper.[6] The foil is cleaned by rinsing with solvent. While holding the roughened foil at about a 45° angle to the benchtop, a few microliters of the analyte solution are applied to the top of the substrate. The liquid drop is allowed to flow off the bottom of the substrate. This method delivers a very thin film of analyte to the silver substrate. An example SIMS spectrum of PS 2450 is shown in Figure 14.

An alternative method of sample preparation for liquid organic samples is to apply a small aliquot of the analyte

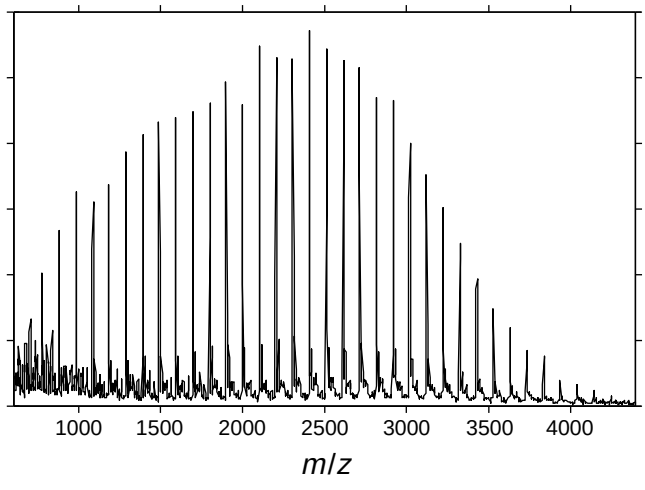

Figure 14 Static SIMS spectrum of a very thin layer of PS 2450 on silver foil.

directly to a piece of filter paper. The analyte will wet the surface of the filter paper and occupy the surface analyzed by SIMS.

A new method of analyzing low-molecular-weight oligomers and biological samples by SIMS is called MESIMS.[32–34] This method uses sample preparation techniques borrowed from matrix-assisted laser desorption/ionization (MALDI).[35,36] The following is a typical MESIMS sample preparation for poly(methyl methacrylate) (PMMA) 2900. A 5 mg mL^{-1} solution of PMMA 2900 is prepared in acetone. An aliquot of this solution is mixed 1 : 7 by volume with a 0.25 M solution of 2,5-dihydroxybenzoic acid. A few microliters of the resulting solution are applied to a clean aluminum substrate. Figure 15 shows the MESIMS spectrum acquired from this sample preparation.

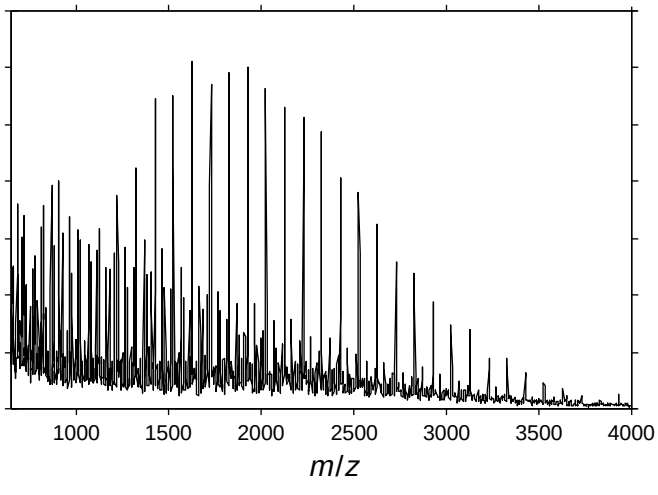

Figure 15 MESIMS mass spectrum of PMMA 2900 prepared with 2,5-dihydroxybenzoic acid as the matrix.

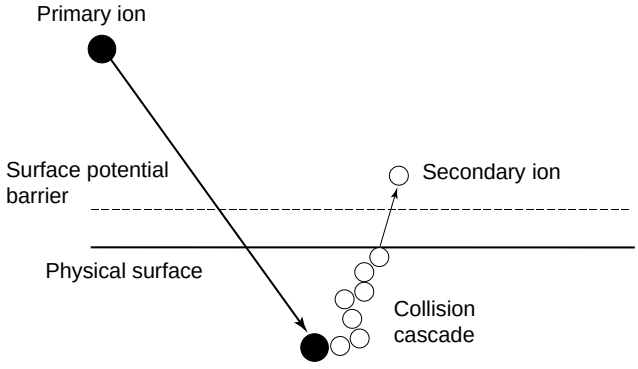

Figure 16 Schematic diagram of secondary ion ejection during a SIMS experiment, showing the impact and burial of the primary ion, the collision cascade, and the ejection of a secondary ion.

5 SECONDARY ION FORMATION

5.1 Desorption

As described previously, the SIMS experiment is initiated with the bombardment of the analyte with the primary ion. Figure 16 shows a schematic diagram of the impact of the primary ion on a sample.

The high-energy primary ion impacts the surface of the sample and begins to transfer energy to the sample. The highest energy density from the collision occurs in and around the primary ion impact area. The relatively high energy deposited in this region creates primarily atomic and small-molecule fragment ions. The energy from the primary ion impact is transferred via collisions between atoms and molecules surrounding the impact. This transfer process is called the collision cascade. The energy delivered during any given collision decreases in proportion to the distance from the original impact. Hence ions ejected further from the point of primary ion impact will have experienced less energetic collisions. These ions will tend to be the higher molecular mass fragments and intact molecular ions.

The details of the desorption process are not yet well understood. The basic concepts of the desorption models center around collision cascades that transfer energy first from the primary ion to the molecular structure of the analyte. These collisions continue between different components of the analyte, until a portion of the analyte receives sufficient energy to escape and is desorbed into the gas phase. Many different workers have contributed to these desorption models.[6,13,27,37–42]

There are two primary desorption models, knock-on sputtering and electronic excitation sputtering.[38] The knock-on sputtering model covers the ejection of target atoms that receive sufficient energy to desorb directly from the primary ion. The electronic excitation model was developed primarily to explain desorption from insulators. It incorporates the idea that excited-state lifetimes may be sufficiently long to allow the transfer of energy from electronic excitation to atomic motion. Both models cover three different regimes (Figure 17a–c) that are concerned with differing numbers of atoms involved in the collision cascade: single knock-on for individual atoms, linear cascade for a small region of moving atoms, and the spike for a large region of moving atoms.[38,42] Real systems are probably affected by forces discussed in both models.

Winograd et al.[41] have developed molecular dynamics models to help explain ion desorption. Their models use classical mechanics and examine only pair-wise forces

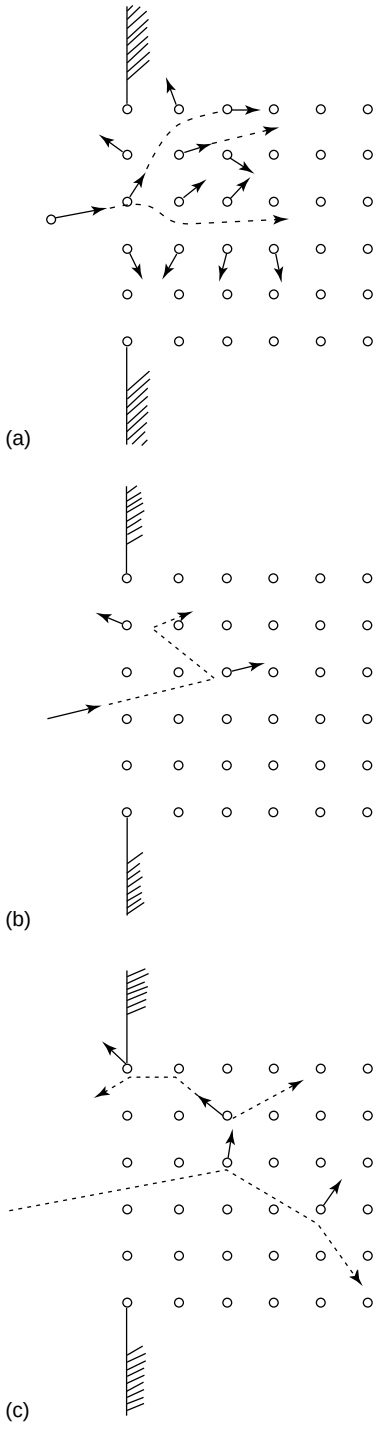

Figure 17 Schematic diagrams showing the three regimes of knock-on and electronic excitation sputtering: (a) single knock-on; (b) linear cascade; and (c) spike. (Reproduced by permission from Springer-Verlag.)

between neighboring atoms or atomic ions in the solid. Figure 18(a) and (b) show the results of a desorption model calculation of a thin organic film on a metal substrate from Garrison.[43]

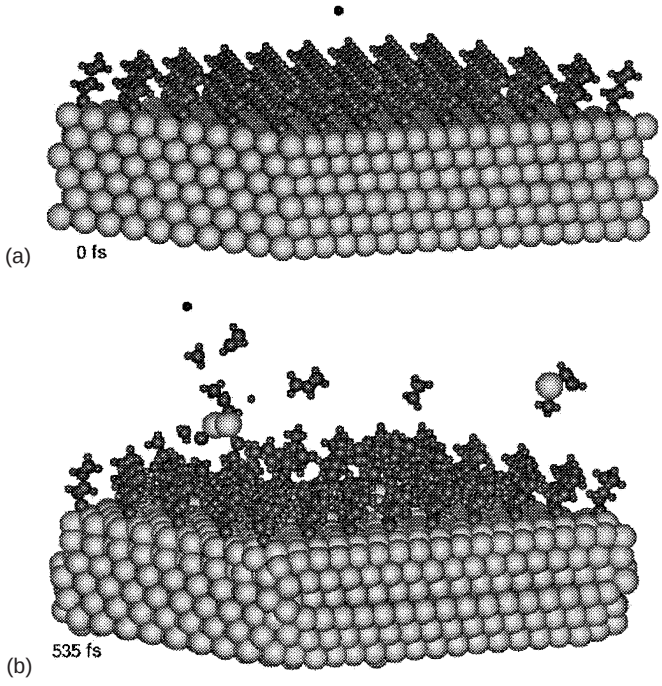

Figure 18 Results from a desorption model calculation of a thin organic film on a metal substrate. (Reproduced by permission of Pennsylvania State University.)

Species that escape the surface are counted as successfully desorbed. These molecular dynamics models tend to treat desorption independently from ionization. While the models must be simplified to allow reasonable computer calculation, they have provided an important insight into the desorption process and good agreement with some experiments.[44,45]

One of the key findings from the molecular dynamics models is that sputtered particles can recombine above the surface. This result has two important consequences: are multimers observed in the SIMS mass spectrum a reflection of the surface being analyzed, and can reactions be an important component of an ionization mechanism? If multimers can recombine with different species after sputtering, the utility of SIMS as a surface structural tool is diminished. For a series of copper clusters and copper oxides, it has been observed that these species were formed by recombination above the surface, but that they were formed primarily from nearest neighbors from the solid.[46] In this example, the structural integrity of the SIMS experiment was maintained. We must keep in mind that ions observed by SIMS may originate from collisions in the selvedge region rather than reflect true surface species. The selvedge region is the space directly above the impact of the primary ion into the sample that is neither solid nor vacuum. It is a relatively high-pressure environment in the transition from solid to vacuum containing both neutral and charged particles.

5.2 Energy Distribution

Once a particle has successfully escaped the surface and has desorbed, it can have a range of kinetic energies. The kinetic energy available to the sputtered particle depends on the energy of the collisions from the collision cascade. It also depends on the size of the particle, including the number of bonds broken to desorb it, any bonds reformed upon sputtering, and any internal degrees of freedom available to disperse energy.

Figure 19 shows a schematic diagram of typical kinetic energy distributions of sputtered atoms and clusters.[47] For typical metal atomic ions, M_1^+, we observe a distribution containing ions having from 0 V to more than 20 V of kinetic energy. There is still significant ion intensity in the high-energy tail of the distribution. For typical cluster ions, M_2^+ and M_3^+, we see fewer ions in the high-energy tail and more ions in the low-energy portion of the distribution. As the chemical complexity of the secondary ion increases, the distribution of kinetic energies tends to decrease.

For sputtered molecular ions we tend to observe kinetic energy distributions that peak at lower energy and have a significantly lower population in the high-energy tail. Since these differences are observed from sputtering by the same primary ion, this result is consistent with the collision cascade delivering a consistent range of energies back to the surface and available for desorption. The more complex molecular species require more energy to desorb, leaving less energy available for kinetic energy. The high-energy tail may disappear for molecular ions because of fragmentation. Those molecular ions with access to energy above 10 eV can readily fragment and are then lost to this kinetic energy distribution.[40]

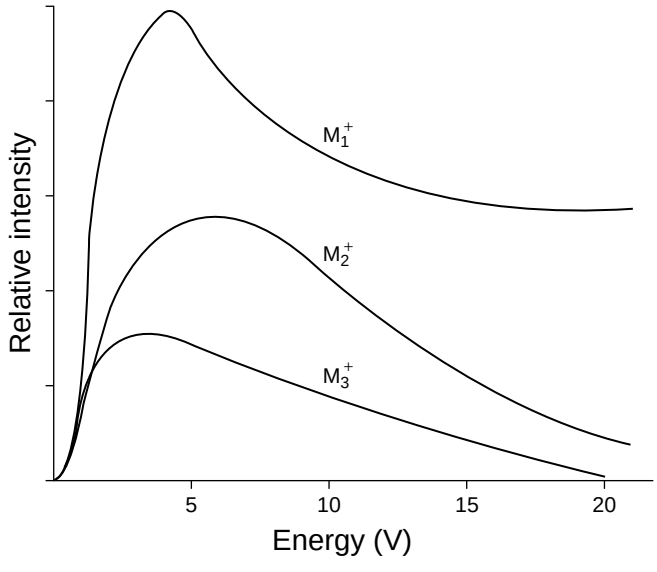

Figure 19 Schematic diagrams of typical kinetic energy distributions of sputtered atoms and clusters. (Reproduced by permission of Charles Evans and Associates.)

5.3 Ionization Mechanisms

The particles desorbed due to the impact of the primary ion can be neutral or charged. Charged particles can be readily introduced to a mass spectrometer and mass analyzed. The distribution of charged species to neutral species is completely dependent on the details of each experiment. For most organic samples, most sputtered particles are electrons, followed by neutral species and then secondary ions: secondary electrons $\gg$ neutrals $\gg$ secondary ions. The secondary electrons can be detected with an SED. The sputtered neutrals can be studied by SNMS, a complementary technique to SIMS. The precise mechanisms involved in ionization during SIMS experiments are not fully understood. Good descriptions of different models have been published by Leggett and Vickerman,[38] Cooks et al.,[39,40,48] and Benninghoven et al.[6] There appear to be at least four different ionization mechanisms in play during a SIMS experiment: bond breaking, preformed intact ions, cationization or anionization reactions, and unimolecular decomposition. Real samples probably have secondary ions produced by each of these mechanisms.

The high energy released in the impact region of the primary ion is sufficient to create a plethora of charged species, from electrons to small molecular ions. These ions are created by momentum transfer from the primary ion.[38] They are primarily atomic ions and are thought to be formed by bond-breaking interactions in the top layer of the sample surface. Once the ion escapes the surface, the relative energy of the ion state compared with the neutral state will determine if the ion is observed.[27]

In samples that have clear preformed ions, such as quaternary ammonium salts, charged species already exist in the solid prior to the primary ion impact. These ion species can be ''liberated'' by the primary ion impact and readily detected by SIMS. Preformed ions can produce significant SIMS signal intensity because no energy is consumed in a bond-breaking step to form a fragment ion or in an ionization step to form a parent ion. It is clear that preformed ions are important in ionic materials. Benninghoven et al. investigated the role of preformed ions in neutral materials and published studies of the SIMS analyses of peptides on various metal substrates.[6] In these experiments, they observed various $[M + H]^+$, $[M - H]^-$, and $[M + \text{metal}]^+$ secondary ions. The relative intensities of these ions correlated with the chemical activity of the underlying metal substrate. From these data, they concluded that protonated or deprotonated species existed prior to analysis on the metal substrate, ''... very high secondary ion yields occur only if the

considered ion is already preformed at and bonded to the surface''.[6]

Neutral species that are desorbed can undergo cationization or anionization reactions in the selvedge region. In the selvedge region there can be many low-energy collisions between desorbed neutrals and atomic ions. These collisions lead to ionization reactions resulting in an adduct between the ion and the neutral. The atomic ions in the selvedge region are created by the bond-breaking interactions described above. The molecular dynamics calculations of Winograd et al. and Garrison discussed in section 5.1 indicate that recombination reactions in the selvedge region is common in SIMS.

Whenever a multi-atomic species is produced in a SIMS experiment, it has the chance of fragmenting after desorption. This is true for both desorbed ions and neutrals. An ion produced by any of the preceding ionization mechanisms can fragment by unimolecular decay if it has sufficient energy. In addition, initially neutral desorbed species can also create ions by dissociative ion formation.[27] Dissociative ion formation can occur if the ion and neutral potential energy surfaces cross at an appropriate distance from the surface. These fragment ions contain significant chemical information about the surface of the sample. Static SIMS experiments appear to have components of both high-energy mass spectra, such as electron ionization or laser desorption with significant molecular fragmentation, and low-energy mass spectra, such as MALDI, with significant intact molecular ions.

SIMS experiments on real samples probably produce ions by all of these mechanisms. Many of the concepts of these mechanisms have been gathered into a single model.

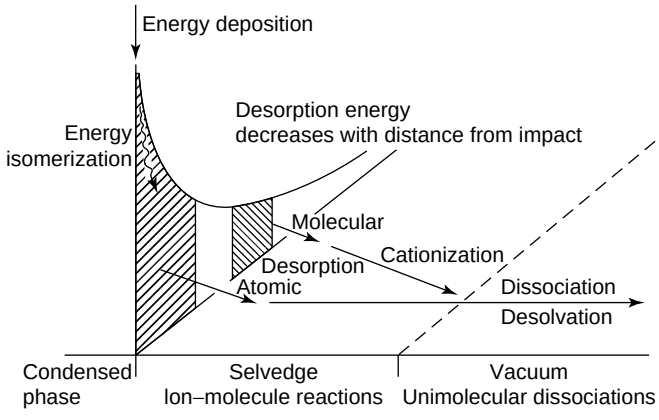

Figure 20 Schematic diagram of the Cooks and Busch desorption ionization model. (Reprinted from *Int. J. Mass Spectrom. Ion Phys.*, **53**, R.G. Cooks, K.L. Busch, Matrix Effects, Internal Energies and MS/MS Spectra of Molecular Ions Sputtered from Surfaces, 111–124, Copyright (1983) with permission from Elsevier Science.)

Figure 20 shows a schematic diagram of the desorption ionization model published by Cooks and Busch.[27,48]

This model separates the processes of desorption and ionization. The energy from the primary ion is transformed into thermal/vibrational motion. Species existing as charged particles in the solid can be desorbed immediately. Desorbed neutral species can be cationized (or anionized) through low-energy collisions in the selvedge region. Regardless of the ionization mechanism, ions can undergo unimolecular decay prior to detection in the mass spectrometer.

6 APPLICATIONS

6.1 Dynamic Secondary Ion Mass-spectrometry Depth Profiling

Ion implantation is a method used for doping the surface of a semiconductor with an electrically active species.[9] Ion implantation damages the sample and the dopant species may not be located on lattice sites. The sample is then annealed to restore crystallinity and achieve the desired electrical properties. SIMS is well suited to studying the implant distribution in these systems. A typical analysis is shown in Figure 21(a–d).[9] In this application, dynamic SIMS was used to examine the behavior of Mg implanted into Fe-doped InP as the sample was annealed.

In this example, Fe is present in the semi-insulating substrate and Mg is a p-type dopant. Figure 21(a) shows the typical Mg implant distribution and a uniform Fe doping level. Annealing at 600 °C (Figure 21b) has little effect on the Mg and only a slight redistribution of the Fe is observed. Annealing at 700 °C (Figure 21c) results in Mg diffusion into the InP and also back toward the surface. The profile also shows that Fe appears on the damage sites and is displaced from the bulk material by Mg diffusion. After annealing at 800 °C (Figure 21d), Mg has diffused 3.5 μm into the InP.

6.2 Static Secondary Ion Mass-spectrometry Analysis of Poly(vinyl chloride) Sheets Using Time-of-flight Secondary Ion Mass Spectrometry

TOFSIMS is used extensively to understand polymer additives.[49–51] TOFSIMS can also be very useful in solving problems. For example, reasons for poor adhesion of poly(vinyl acetate–ethylene) to a poly(vinyl chloride) (PVC) sheet were determined by surface analysis. Positive ion TOFSIMS spectra of two PVC sheets are shown in Figure 22(a) (problem sample) and Figure 22(b) (control sample).[51] Poor adhesion to the problem PVC sheet was attributed to the presence of polydimethylsiloxane (PDMS), as evidenced by the $C_5H_{15}Si_2O^+$ ion at 147 u,

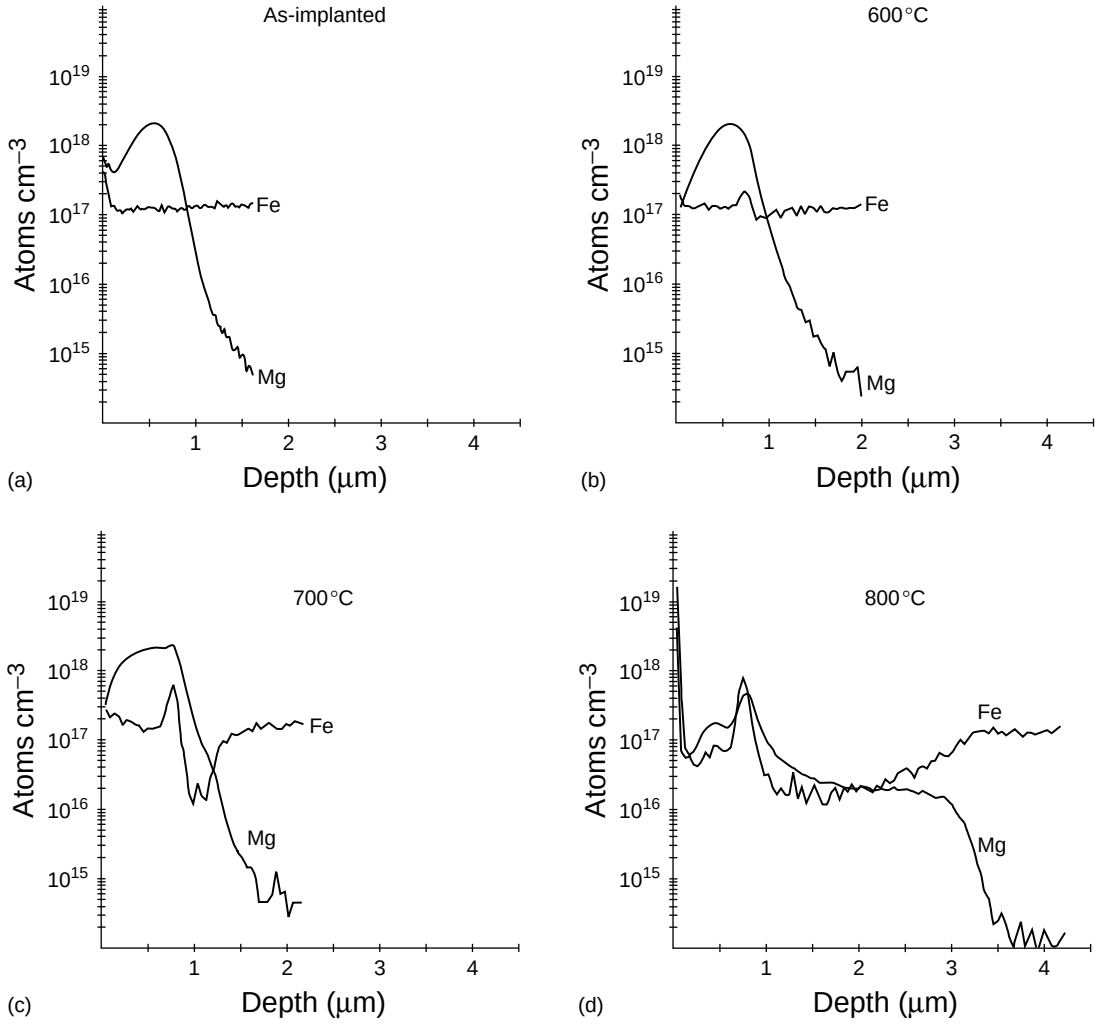

Figure 21 Dynamic SIMS depth profiles of Mg implanted into Fe-doped InP. The depth profiles illustrate the effects of annealing: (a) prior to annealing; (b) annealing at 600 °C; (c) annealing at 700 °C; (d) annealing at 800 °C. (Reprinted with the permission of Cambridge University Press.)

and phthalate plasticizer, as evidenced by the $C_8H_5O_3^+$ ion at 149 u.

PDMS is a common cause of adhesion problems in coating applications. PDMS has a very recognizable pattern of ion peaks.[49–52] Figure 23 shows a typical static SIMS spectrum containing PDMS.

6.3 Static Secondary Ion Mass-spectrometry Analysis of Silicon Wafers Using Time-of-flight Secondary Ion Mass Spectrometry

TOFSIMS is routinely used in the microelectronics industry. Figures 24 and 25 show TOFSIMS spectra from Fe implanted in a silicon wafer. Figure 24 shows the region at nominal mass 29 u. One advantage of TOFSIMS is the ability to obtain high-mass-resolution SIMS spectra. Figure 24 shows four distinct peaks at nominal mass 29 u. We observe $^{29}Si^+$, $^{28}SiH^+$, COH^+, and $C_2H_5^+$. The arrival time difference between the $^{29}Si^+$ ion and the $^{28}SiH^+$ ion is only 2 ns. This corresponds to fifteen 138 ps channels on the time-to-digital converter. Figure 25 shows the 56 u mass region, where $^{56}Fe^+$, $C_4H_8^+$, and Si_2^+ are all resolved.

6.4 Imaging

One of the more powerful capabilities of many commercial TOFSIMS instruments is to produce chemically sensitive ion images. The development of extremely small spot LMIG sources enables us to obtain secondary ion images with lateral resolution of the order of 1–5 μm length scale. In addition to obtaining high-resolution images, images for individual ions can be collected. The combination of image and mass information can be a powerful tool to solve problems. Figure 26(a–d) shows four

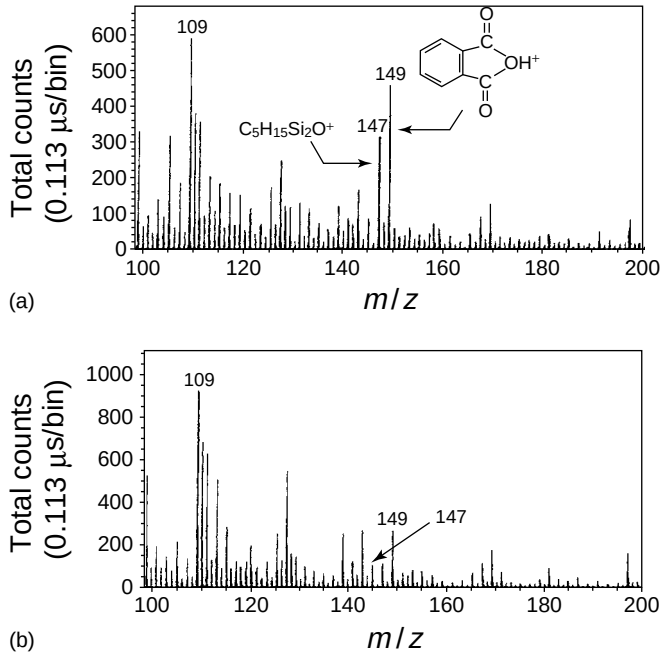

Figure 22 Positive ion TOFSIMS spectrum in the 100–200 u mass range: (a) the problem PVC sample; (b) the control PVC sample, obtained under static SIMS conditions.

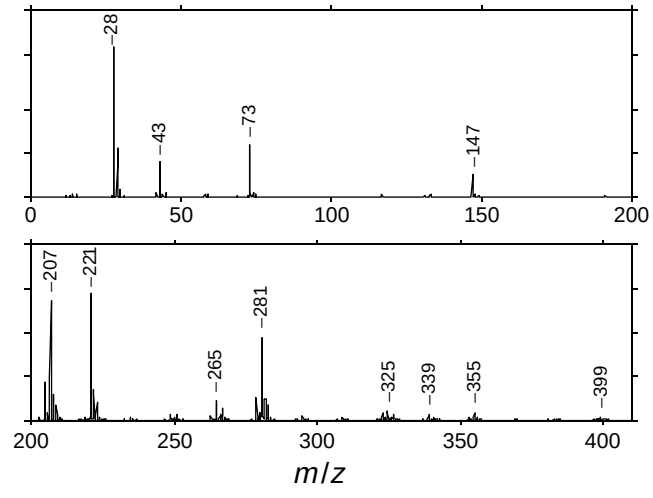

Figure 23 Static SIMS spectrum of PDMS low-molecular-weight species.

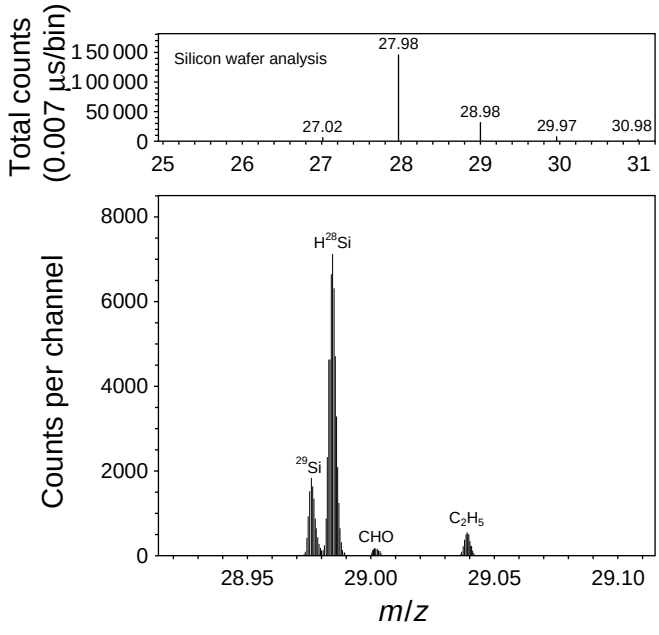

Figure 24 TOFSIMS spectrum from Fe implanted in a Si wafer at the nominal mass of 29 u, obtained under static SIMS conditions.

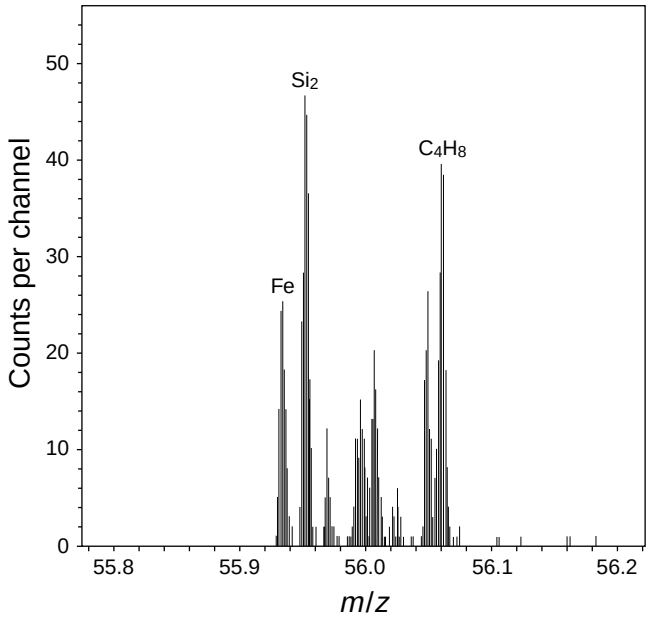

Figure 25 TOFSIMS spectrum from Fe implanted in a Si wafer at the nominal mass of 56 u, obtained under static SIMS conditions.

individual ion images from a microtomed cross-section of an automotive coating sample.[53]

In Figure 26(a–d), we see chemically sensitive ion images from the various layers of the coating. These images were collected on a TRIFT® TOFSIMS instrument with a $200 \times 200\,\mu m$ raster using a ^{69}Ga LMIG primary ion source. The Cl^- ion is indicative of the approximately 6 μm thick primer layer, a chlorinated adhesion promoter. The Al^+ ion is indicative of the approximately 20 μm thick basecoat. The NCO^- ion is indicative of the polyurethane base coat and approximately 37 μm thick clear coat. Finally, the Mg^+ ion is indicative of the substrate. The spots in the Mg^+ ion image indicate that the talc in the substrate is present as small (<5 μm) clusters.[53] These unique secondary ion

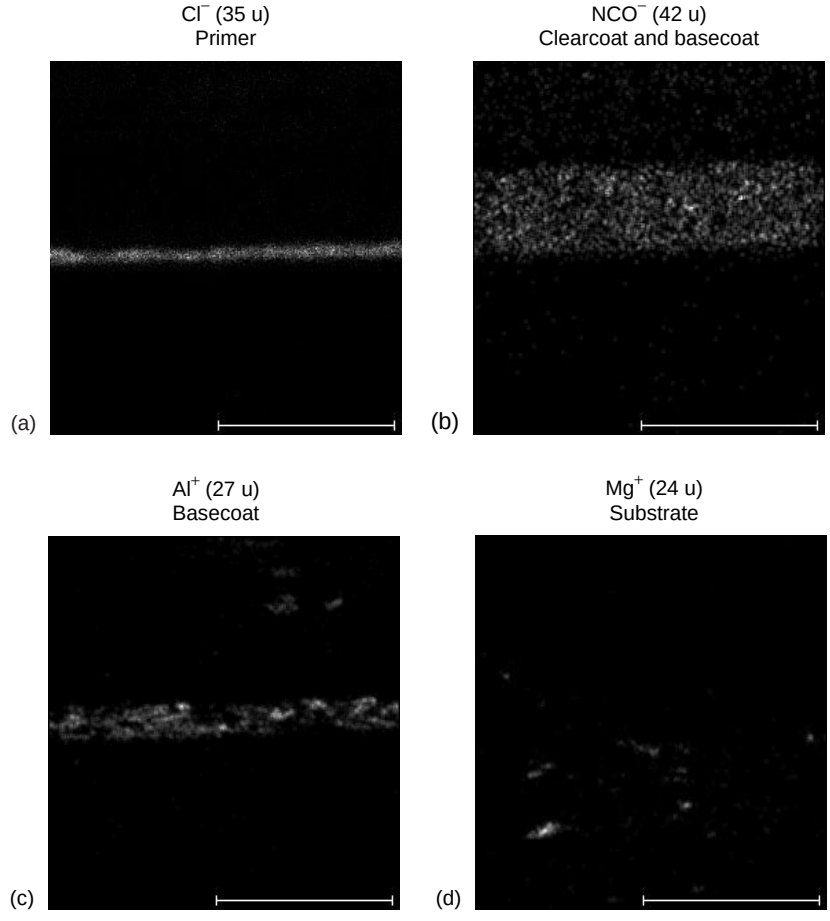

Figure 26 Static SIMS chemically specific images from a paint sample showing images for (a) Cl^-, (b) NCO^-, (c) Al^+ and (d) Mg^+. Bar = 100 µm. (Reproduced by permission of Physical Electronics.)

images enable static SIMS to identify and probe each layer of this multilayer sample.

6.5 Bulk Polymer Molecular Weight Analysis

SIMS can also be used to provide molecular weight and chemical structure information on low-molecular-weight oligomer samples. As described in section 4.4, we can use traditional monolayer sample preparation or matrix-enhanced sample preparation for low-molecular-weight samples and obtain mass spectra containing intact molecular ions of the oligomers. Figure 27 shows the SIMS spectrum of a sodium-cationized Surfynol® surfactant analyzed using MESIMS.

In the mass spectrum, we see a series of oligomer peaks spaced by 44 u. This Surfynol® surfactant is an ethoxylated alkyne.[54] Figure 28 shows the chemical structure of the Surfynol® surfactants.

The 44 u peaks are the different ethoxylated oligomer chain lengths. From these peak masses we can verify the mass of the alkyne backbone and that the surfactant is ethoxylated. From the distribution of the ion peaks we can calculate the average molecular weights of the material and calculate the average degree of ethoxylation.

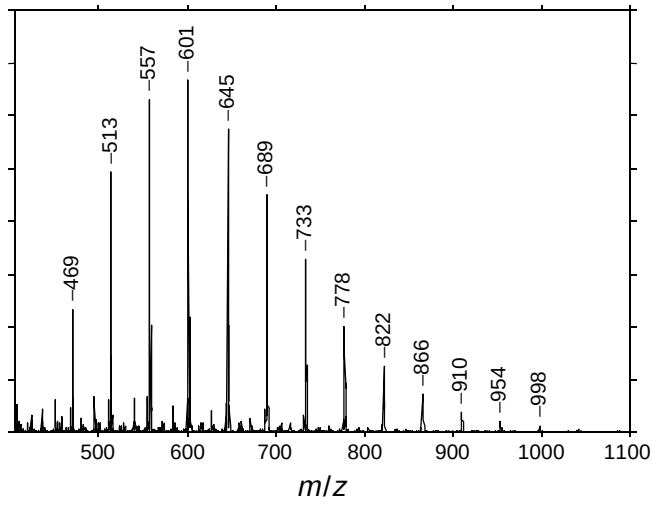

Figure 27 MESIMS spectrum of an ethoxylated Surfynol® surfactant.

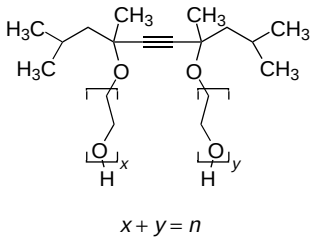

Figure 28 Structure of an ethoxylated Surfynol® surfactant; *n* represents the degree of ethoxylation.

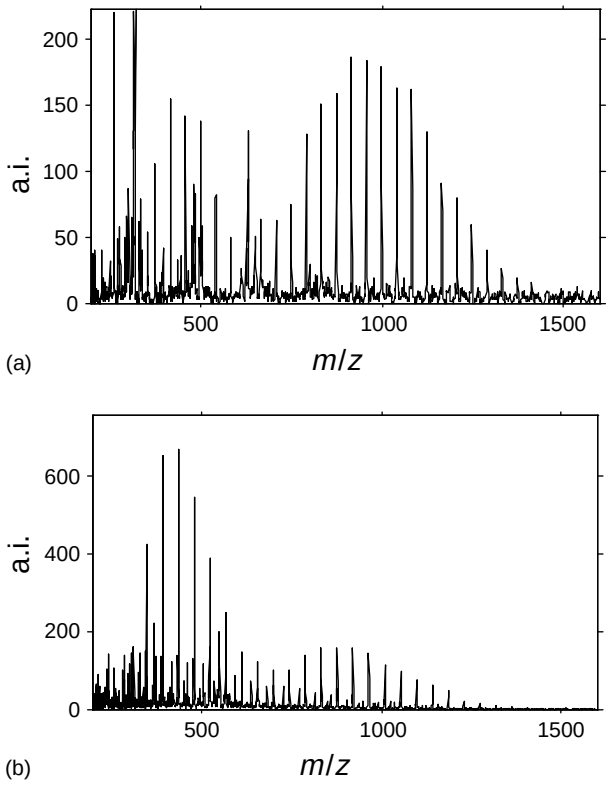

Figure 29 Mass spectra of a 1 : 1 molar blend of PEG 400 and PEG 1000, (a) by MALDI and (b) by MESIMS. The sample was prepared with methanol and 2,5-dihydroxybenzoic acid.

MESIMS can also be used in combination with MALDI to investigate the effects of surface segregation in blend analyses.[34,55] For example, we prepared 1 : 1 molar blends of poly(ethylene glycol) (PEG) 400 and PEG 1000. These very low-molecular-weight oligomers were chosen because we need component samples that can be detected and quantified by static SIMS. Figure 29(a) and (b) show two mass spectra of the 1 : 1 molar blend of PEG 400 and PEG 1000. The results of the MALDI and MESIMS analyses are given in Table 4.

Whereas we obtain similar measures of the average molecular weights, we observe a significant difference in the measured relative blend composition. Interestingly, under these conditions, neither technique measures a 1 : 1 molar blend composition. The MALDI analysis shows greater intensity for the higher molecular weight component. On the same sample, MESIMS analysis shows greater intensity for the lower molecular weight component. These experiments provide evidence that components of a sample can segregate during sample preparation. In this case, we observe evidence for segregation of the lower molecular weight oligomers at the surface of the sample.

Table 4 Comparison of results for a 1 : 1 molar blend of PEG 400 and PEG 1000

Technique	Component	M_n	M_w	Area
MALDI	PEG 400	485	510	5000
MALDI	PEG 1000	1000	1070	8800
MESIMS	PEG 400	475	490	4000
MESIMS	PEG 1000	950	970	3100

ACKNOWLEDGMENTS

We thank Air Products and Chemicals, Inc., for supporting this work and Dr Menas Vratsanos, Dr Bob Coraor, and Dr Tom Mebrahtu for helping prepare the article.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscopy
ASTM	American Society for Testing and Materials
EDX	Energy-dispersed X-rays
ESA	Electrostatic Sector Analyzer
ESCA	Electron Spectroscopy for Chemical Analysis
FESEM	Field Emission Scanning Electron Microscopy
fwhm	Full Width at Half-maximum
ISS	Ion Scattering Spectroscopy
LDMS	Laser Desorption Mass Spectrometry
LMIG	Liquid Metal Ion Gun
MALDI	Matrix-assisted Laser Desorption/ionization
MESIMS	Matrix-enhanced Secondary Ion Mass Spectrometry
NIST	National Institute of Standards and Technology
PDMS	Polydimethylsiloxane
PEG	Poly(ethylene glycol)
PMMA	Poly(methyl methacrylate)

PS	Polystyrene
PVC	Poly(vinyl chloride)
RBS	Rutherford Backscattering Spectrometry
SAM	Scanning Auger Microscopy
SED	Secondary Electron Detector
SEM	Secondary Electron Microscopy
SIMS	Secondary Ion Mass Spectrometry
SNMS	Secondary Neutral Mass Spectrometry
STM	Scanning Tunneling Microscopy
TEM	Transmission Electron Microscopy
THF	Tetrahydrofuran
TOF	Time-of-flight
TOFMS	Time-of-flight Mass Spectrometry
TOFSIMS	Time-of-flight Secondary Ion Mass Spectrometry
UHV	Ultrahigh Vacuum
UPS	Ultraviolet Photoelectron Spectroscopy
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Microscopy of Coatings

Peptides and Proteins **(Volume 7)**
Surface Plasmon Resonance Spectroscopy in Peptide and Protein Analysis

Polymers and Rubbers **(Volume 9)**
Surface Energetics of Polymers and Rubbers, Characterization of

Pulp and Paper **(Volume 10)**
X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Surfaces **(Volume 10)**
Surfaces: Introduction • Auger Electron Spectroscopy in Analysis of Surfaces • Differential Reflectance Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • Ellipsometry in Analysis of Surfaces and Thin Films • Infrared and Raman Spectroscopy in Analysis of Surfaces • Ion Scattering Spectroscopy in Analysis of Surfaces • Photoluminescence in Analysis of Surfaces and Interfaces • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces • Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

Electronic Absorption and Luminescence **(Volume 12)**
Surface Measurements using Absorption/Luminescence

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Inorganic Substances, Mass Spectrometric in the Analysis of • Time-of-flight Mass Spectrometry

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

General Articles **(Volume 15)**
Multivariate Image Analysis

REFERENCES

1. A.W. Czanderna, 'Introduction', in *Methods and Phenomena 1 – Methods of Surface Analysis*, ed. A.W. Czanderna, Elsevier, New York, 1, 1975.
2. P.W. Atkins, *Physical Chemistry*, W.H. Freeman, New York, 150, 1986.
3. A.W. Adamson, in *Physical Chemistry of Surfaces*, 4th edition, John Wiley & Sons, New York, 1 and 57, 1982.
4. J.J. Bickerman, *Physical Surfaces*, Academic Press, New York, 1, 1970.
5. P.S. Laplace, *Mecanique Celeste*, Suppl. to the 10th book, Imperiale, Paris, 1805, and *Laplace's Ouevres*, Little and Brown, Boston, Vol. IV, 1839.
6. A. Benninghoven, F.G. Ruedenauer, H.W. Werner, *Secondary Ion Mass Spectrometry*, John Wiley & Sons, New York, 1987.
7. A. Benninghoven, 'Some Aspects of Secondary Ion Mass Spectrometry of Organic Compounds', *Int. J. Mass Spectrom. Ion Phys.*, **53**, 85–99 (1983).
8. D.M. Hercules, S.H. Hercules, 'Analytical Chemistry of Surfaces – Part III: Ion Spectroscopy', *J. Chem. Educ.*, **61**(7), 593–599 (1984).
9. D.E. Sykes, 'Dynamic Secondary Ion Mass Spectrometry', in *Methods of Surface Analysis – Techniques and Applications*, ed. J.M. Walls, Cambridge University Press, New York, 216–262, 1989.
10. D. Briggs, M.P. Seah, *Practical Surface Analysis – Auger and X-ray Photoelectron Spectroscopy*, 1st edition, John Wiley & Sons, New York, 1983.
11. M.P. Seah, 'Pure Element Sputtering Yields Using 500–1000 eV Argon Ions', *Thin Solid Films*, **81**, 279–287 (1981).

12. A. Benninghoven, *Surf. Sci.*, **53**, 596–625 (1975).
13. P. Williams, 'Quantitative Analysis Using Sputtering Techniques: Secondary Ion and Sputtered Neutral Mass Spectrometry', in *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, eds. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 177–228, Vol. 2, 1992.
14. D.E. Newbury, 'The Role of Standards in Secondary Ion Mass Spectrometry,' in *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, eds. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 635–652, Vol. 2, 1992.
15. J.C. Vickerman, 'Static Secondary Ion Mass Spectrometry', in *Methods of Surface Analysis – Techniques and Applications*, ed. J.M. Walls, Cambridge University Press, New York, 169–215, 1989.
16. R. Jede, O. Ganschow, U. Kaiser, 'Instrumentation for SIMS', in *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, eds. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 19–104, Vol. 2, 1992.
17. R. Smith, J.M. Walls, 'Ion Erosion in Surface Analysis', in *Methods of Surface Analysis – Techniques and Applications*, ed. J.M. Walls, Cambridge University Press, New York, 20–56, 1989.
18. P. Williams, R.K. Lewis, C.A. Evans, Jr, 'Evaluations of a Cesium Primary Ion Source on a Ion Microprobe Mass Spectrometer', *Anal. Chem.*, **49**(9), 1399–1403 (1977).
19. J.D. Pinkston, M. Rabb, J.T. Watson, J. Allison, 'New Time-of-flight Mass Spectrometer for Improved Mass Resolution, Versatility, and Mass Spectrometry/Mass Spectrometry Studies', *Rev. Sci. Instrum.*, **57**(4), 583–592 (1986).
20. P.E. Miller, M. Bonner Denton, 'The Quadrupole Mass Filter: Basic Operating Concepts', *J. Chem. Educ.*, **63**(7), 617–622 (1986).
21. R. Davis, M. Frearson, *Mass Spectrometry*, John Wiley & Sons, Chichester, 1987.
22. P.K. Chu, B.W. Schueler, F. Reich, P.M. Lindley, 'Determination of Trace Metallic Impurities on 200-mm Silicon Wafers by Time-of-flight Secondary-ion Mass Spectrometry', *J. Vac. Sci. Technol. B*, **15**(6), 1908 (1997).
23. J. Schwieters, H.-G. Cramer, U. Juergens, E. Niehuis, J. Zehnpfenning, A. Benninghoven, 'High Mass Resolution Surface Imaging with a Time-of-flight Secondary Ion Mass Spectroscopy Scanning Microprobe', *J. Vac. Sci. Technol. A*, **9**, 2864–2871 (1991).
24. E. Niehuis, T. Heller, H. Feld, A. Benninghoven, 'Design and Performance of a Reflectron Based Time-of-flight Secondary Ion Mass Spectrometer with Electrodynamic Primary Ion Mass Separation', *J. Vac. Sci. Technol. A*, **5**(4), 1243–1246 (1987).
25. 'Unique Features of the TRIFT® II Spectrometer and Their Analytical Benefits', Physical Electronics, Eden Prairie, MN, 1996.
26. B. Schueler, 'Microscope Imaging by Time-of-flight Secondary Ion Mass Spectrometry,' *Microsc. Microanal. Microstruct.*, **3**, 119–139 (1992).
27. N.M. Reed, J.C. Vickerman, 'Static SIMS – Surface Analysis of Inorganic Materials', in *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, ed. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 303–366, Vol. 2, 1992.
28. D.M. Hercules, S.H. Hercules, 'Analytical Chemistry of Surfaces – Part I: General Aspects', *J. Chem. Educ.*, **61**(5), 402–409 (1984).
29. D. Briggs, M.P. Seah, *Practical Surface Analysis – Auger and X-ray Photoelectron Spectroscopy*, 2nd edition, John Wiley & Sons, New York, Vol. 1, 1992.
30. 'ASTM Standard Definitions, Guides, and Practices for Surface Analysis', reprinted under the auspices of the Education Committee of the American Vacuum Society, 1990, especially ASTM designation E 1078–85.
31. P.C. Schamberger, G.L. Jones, J.A. Gardella, P.J. McKeown, L.E. Davis, 'Visualization of a Buried Organic Interface by Imaging TOFSIMS and Scanning Auger Microprobe of an Ion Beam Crater Edge', *J. Vac. Sci. Technol A*, **14**(4), 2289–2302 (1996).
32. K.J. Wu, R.W. Odom, 'Matrix-enhanced Secondary Ion Mass Spectrometry: A Method for Molecular Analysis of Solid Surfaces', *Anal. Chem.*, **68**, 873–882 (1996).
33. A.J. Nicola, D.C. Muddiman, D.M. Hercules, 'Enhancement of Ion Intensity in Time-of-flight Secondary-ionization Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **7**, 467–472 (1996).
34. S.D. Hanton, P.A. Cornelio Clark, K.G. Owens, 'Investigations of MALDI Sample Preparation by TOFSIMS', *J. Am. Soc. Mass Spectrom.*, **10**, 104–111 (1999).
35. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**, 2299–2301 (1988).
36. K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshido, T. Yoshido, 'Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **2**, 151–153 (1988).
37. G. Betz, G.K. Wehner, in *Sputtering by Particle Bombardment II: Topics in Applied Physics*, ed. R. Behrisch, Springer-Verlag, Berlin, Vol. 52, 1983.
38. G.J. Leggett, J.C. Vickerman, 'An Empirical Model for Ion Formation from Polymer Surfaces During Analysis by Secondary Ion Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **122**, 281–319 (1992).
39. R.J. Day, S.E. Unger, R.G. Cooks, 'Molecular Secondary Ion Mass Spectrometry', *Anal. Chem.*, **52**(4), 557–566 (1980).
40. S.J. Pachuta, R.G. Cooks, 'Mechanisms in Molecular SIMS', *Chem. Rev.*, **87**(3), 647–669 (1987).
41. N. Winograd, D.E. Harrison, B.J. Garrison, 'Structure Sensitive Factors in Molecular Cluster Formation by Ion

Bombardment of Single Crystal Surfaces', *Surf. Sci.*, **78**, 467–477 (1978).
42. P. Sigmund, 'Sputtering by Ion Bombardment: Theoretical Concepts', in *Sputtering by Particle Bombardment I: Topics in Applied Physics*, ed. R. Behrisch, Springer-Verlag, Berlin, Vol. 47, 1981.
43. B.J. Garrison, Internet Homepage at Pennsylvania State University: www. galilei.chem.psu.edu/welcome2, 1998.
44. D.E. Harrison, W.L. Moore, H.T. Holcombe, 'Computer Simulation of Sputtering II', *Radiat. Effects*, **17**, 167–183 (1973).
45. N. Winograd, D.E. Harrison, B.J. Garrison, 'Angular Distributions of Ejected Particles from Ion-bombarded Clean and Reacted Single-crystal Surfaces', *Phys. Rev. Lett.*, **41**, 1120–1123 (1978).
46. N. Winograd, D.E. Harrison, B.J. Garrison, 'Atomic and Molecular Ejection from Ion-bombarded Reacted Single-crystal Surfaces – Oxygen on Copper (100)', *Phys. Rev. B*, **18**, 6000–6010 (1978).
47. R. Flemming, Charles Evans and Associates Internet Webpage: www.cea.com/cai/simstheo/distrib.htm, 1995.
48. R.G. Cooks, K.L. Busch, 'Matrix Effects, Internal Energies and MS/MS Spectra of Molecular Ions Sputtered from Surfaces', *Int. J. Mass Spectrom. Ion Phys.*, **53**, 111–124 (1983).
49. D. Briggs, A. Brown, J.C. Vickerman, *Handbook of Static Secondary Ion Mass Spectrometry*, John Wiley & Sons, New York, 1989.
50. D. Briggs, 'Static SIMS – Surface Analysis of Organic Materials', in *Practical Surface Analysis – Ion and Neutral Spectroscopy*, 2nd edition, ed. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 367–423, Vol. 2, 1992.
51. P.A. Cornelio Clark, S.A. Gardner, D. Horwat, 'Probing the Interfacial Properties of Poly(vinyl acetate–ethylene) Copolymer/Poly(vinyl chloride) Laminations by Time-of-flight Secondary Ion Mass Spectrometry', *J. Vac. Sci. Technol. A*, **13**(3), 1351–1358 (1995).
52. J.G. Newman, B.A. Carlson, R.S. Michael, J.F. Moulder, T.A. Hohlt (eds.), *Static SIMS Handbook of Polymer Analysis*, Perkin-Elmer Corp., Physical Electronics Division, Eden Prairie, MN, 1991.
53. A.M. Belu, J. McGuinness, 'TOF-SIMS Characterization of Multi Layer Paint Coatings', Application Note 9812, Physical Electronics, Eden Prairie, MN, 1998.
54. D.M. Parees, S.D. Hanton, P.A. Cornelio Clark, D.A. Willcox, 'Comparison of Mass Spectrometric Techniques for Generating Molecular Weight Information on a Class of Ethoxylated Oligomers', *J. Am. Soc. Mass Spectrom.*, **9**, 282–291 (1998).
55. S.D. Hanton, K.G. Owens, 'The Investigation of Polymer and Matrix Solubility by MALDI and MESIMS', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics in Orlando, FL*, American Society for Mass Spectrometry, Santa Fe, NM, 1185, 1998.

Tandem Mass Spectrometry: Fundamentals and Instrumentation

Edmond de Hoffmann
Catholic University of Louvain, Louvain-la-Neuve, Belgium

1 Introduction 1
1.1 What is Tandem Mass Spectrometry? 1
1.2 Tandem-in-space Mass Spectrometry 1
1.3 Tandem-in-time Mass Spectrometry 2
Nomenclature 2
2 Ion Dissociation 2
2.1 Stable and Metastable Ions 2
2.2 Collision Energy 2
3 Tandem Mass Spectrometry Scan Modes 4
3.1 Product Ion Scan 4
3.2 Precursor Ion Scan 4
3.3 Neutral Loss Scan 4
3.4 Selected Reaction Monitoring 4
4 In-space Instruments 5
4.1 Electric/Magnetic Sector Instruments 5
4.2 Multiple Sector Magnetic Instruments 6
4.3 Triple Quadrupole Instruments 7
4.4 Time-of-flight Mass Spectrometers 8
5 In-time Instruments 9
5.1 Ion Trap-based MS^n 9
5.2 MS^n in Ion Cyclotron Resonance Fourier Transform Mass Spectrometry Instruments 12
6 Hybrid Instruments 12
7 Analytical Methods Based on Tandem Mass Spectrometry 14
7.1 Structure Elucidation 14
7.2 Determination of Diastereoisomers 15
7.3 Methodology for Compound Class-selective Detection 16
7.4 Increase in Sensitivity by Tandem Mass Spectrometry 18
8 Fields of Application of Tandem Mass Spectrometry 19
9 Conclusions 20
Acknowledgments 20
Abbreviations and Acronyms 20
Related Articles 21
References 21

Tandem mass spectrometry (MS/MS) is a methodology which allows selection of ions of a given mass, fragmentation of these ions and detection of the fragments. It requires the use of dedicated instruments of different types. Some of them allow repetition of the ion selection step to fragment fragments further.

This methodology allows structural information on pure compounds or on components of mixtures to be obtained. It also allows highly selective and sensitive detection of target compounds or of compound classes, even from complex mixtures. It is widely used in environmental analysis, pharmaceutical research, biotechnology, forensic science and toxicology, clinical analysis, organic synthesis, process control and many other areas.

1 INTRODUCTION

1.1 What is Tandem Mass Spectrometry?

A mass analyzer is a device able to separate gas-phase ions according to their mass-to-charge ratio (m/z), allowing one to determine that mass. A detector allows one to count the ions of that mass. The combination of an ion source, a mass analyzer, a detector and a computer to handle the data constitute a mass spectrometer.

A tandem mass spectrometer is a device able, at least, to isolate ions of a selected mass, to let them fragment either spontaneously or by some activation process, and to yield a mass spectrum of the resulting fragments. This is called an MS/MS experiment. It is possible to increase the number of steps: select ions of a first mass, then select ions of a second mass from the fragments obtained, and finally analyze the fragments of these last selected ions. This is labeled as an MS/MS/MS or MS^3 experiment. The number of steps can be further increased to yield MS^n spectra.

Basically, a tandem mass spectrometer can be conceived in two ways, either performing MS/MS in space by coupling two physically distinct instruments, or in time by performing an appropriate sequence of events in an ion storage device.

1.2 Tandem-in-space Mass Spectrometry

In this approach several separate mass analyzers, most often two, are combined. They are combined in such a way that ions can be transferred from one mass analyzer

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

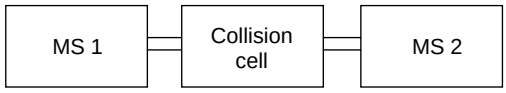

Figure 1 Two mass spectrometers combined in space. A collision cell enables the fragmentation of ions to be induced.

to the other in an instrument schematically represented in Figure 1.

In most cases, a collision cell is placed between the analyzers in order to cause the fragmentation of the ions. These instruments are named *tandem-in-space*, because the combined mass analyzers and collision cells are placed at different locations in space. Common instruments of this type have two mass analyzers, allowing MS/MS experiments to be performed. To obtain higher order MS^n spectra requires the combination of n analyzers, increasing the complexity and thus also the cost of the spectrometer.

1.3 Tandem-in-time Mass Spectrometry

The second approach to building a spectrometer with MS/MS capability is to use an ion storage device able to store only ions of selected masses by ejecting the unwanted ones, and as well being able to analyze the ions present in the storage device according to their masses. With such an instrument, the operator can, at least, select ions of a chosen mass, let them fragment either spontaneously or by an artificial excitation and determine the masses and abundances of the ions produced. Alternatively, after the fragmentation step, the operator can select ions of a new chosen mass, let them fragment and analyze the fragments. This process can, at the will of the operator, be repeated several times, in practice up to six generations of ions, thus producing MS^n spectra. In these instruments, all the procedures take place in one and the same device, so that MS/MS is achieved by successive operations in time. These are time-based instruments. There are two types: ion trap and Fourier transform ion cyclotron resonance (FTICR) instruments. MS^n experiments are performed in a single device. Figure 2 represents a schematic comparison of a simple MS/MS product ion scan performed by either a space-or a time-based instrument.

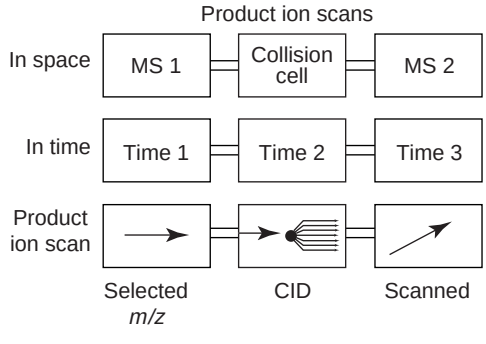

Figure 2 Comparison of a product ion scan performed by a space-based and a time-based instrument.

NOMENCLATURE

The main terms used in MS/MS are listed below.

Molecular ion: ion formed by the addition or the removal of one or several electrons to or from the sample molecule.
Adduct ion: ion formed through the interaction of two species and containing all the atoms of one of them plus one or several atom(s) of the other.
Ion of the molecular species or pseudomolecular ion: ion originating from the analyte molecule by the abstraction of a proton $(M-H)^-$ or by an hydride abstraction $(M-H)^+$, or by the formation of an adduct with another ion of the ionizing plasma. These ions allow the molecular weight to be deduced.
Precursor ion: (used to be called parent ion) any ion undergoing either a decomposition or a charge change.
Product ion: (used to be called daughter ion) ion resulting from the fragmentation of a selected precursor.
Selected reaction monitoring (SRM): monitoring in function of the time of one or several selected product ions from one or more selected precursor ions.
Neutral loss: fragment lost as a neutral species
Source spectrum: spectrum acquired with a single analyzer.

2 ION DISSOCIATION

2.1 Stable and Metastable Ions

Some ions have received a sufficient amount of energy to dissociate before leaving the source and are designated *source ions*. Ions that are stable in the source but have excess internal energy can dissociate spontaneously during flight in the analyzers to the detector; these are *metastable ions*. The other ''stable'' ions normally reach the detector without fragmentation. However, to increase the number of fragment ions, dissociation can be induced by activation, or energy transfer to the ion. The most common way to achieve this is by collision with an inert gas.

2.2 Collision Energy

The kinetic to internal energy transfers are governed by the laws concerning collisions of a mobile species, the ion, and an essentially static target, the collision gas.

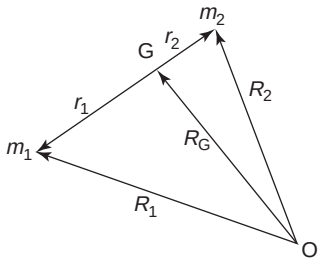

Figure 3 Vectorial demonstration of the link between the position vectors in the laboratory and in the center of gravity reference frames.

Consider an ion with mass m_1 hitting a target with mass m_2. If **R** represents the position vectors in the laboratory reference frame O, and **r** represents the position vectors in the reference frame linked to the center of gravity G, Figure 3 shows that Equations (1) and (2) follow:

$$R_1 = R_G + r_1 \tag{1}$$

and

$$R_2 = R_G + r_2 \tag{2}$$

Defining the center of gravity by Equation (3):

$$(m_1 + m_2)R_G = m_1R_1 + m_2R_2 \tag{3}$$

thus Equations (4) and (5) follow

$$r_1 = \frac{m_2(R_1 - R_2)}{m_1 + m_2} \tag{4}$$

and

$$r_2 = \frac{-m_1(R_1 - R_2)}{m_1 + m_2} \tag{5}$$

Differentiating with respect to t gives Equations (6) and (7):

$$u_1 = \frac{m_2\mathbf{g}}{m_1 + m_2} \tag{6}$$

and

$$u_2 = \frac{-m_1\mathbf{g}}{m_1 + m_2} \tag{7}$$

where $\mathbf{g} = v_1 - v_2$ and $m_1u_1 + m_2u_2 = 0$ and where u is the velocity in the G (center of gravity) frame and v the velocity in the O (laboratory) frame.

According to the law of momentum conservation, the kinetic energy of a rapid particle colliding with a static target ($v_2 = 0$) cannot be entirely converted into internal energy. The kinetic energy available for conversion, E_t, termed *relative kinetic energy*, is actually the kinetic energy in the center of mass reference frame. Indeed, in this G frame the momentum is equal to zero. We thus obtain Equation (8)

$$E_t = \frac{m_1u_1^2 + m_2u_2^2}{2} = \frac{\mu g^2}{2} \tag{8}$$

where $\mu = m_1m_2/(m_1 + m_2)$. Thus Equation (9) follows

$$E_t = E_c\frac{m_2}{m_1 + m_2} \tag{9}$$

where $E_c = m_1v_1^2/2$.

The kinetic energy is conserved in an *elastic* collision, Equation (10)

$$\frac{\mu g^2}{2} = \frac{\mu g'^2}{2} \tag{10}$$

In an *inelastic* collision, a part Q of the kinetic energy, at a maximum equal to E_t, is converted into internal energy, Equation (11):

$$\frac{\mu g^2}{2} = Q + \frac{\mu g'^2}{2} \tag{11}$$

For instance, a 100 Da ion with a kinetic energy of 10 eV colliding with argon (atomic mass 40 Da) has a maximum increase in its internal energy amounting to $10[40/(40 + 100)] = 2.86$ eV

Always keep in mind that 1 eV per ion is equivalent to around 100 kJ mol^{-1}.

This calculation supposes a totally elastic collision between spherical species.

A measure of the average amount of kinetic energy converted into internal energy in the collision process was obtained by Harrison and Lin.[1] They observed that the fragmentation of the *n*-butylbenzene molecular ion gives two fragments, $C_7H_7^+$ and $C_7H_8^{\bullet+}$ and that the ratio of their intensities increases with increasing amount of internal energy of the parent ion $C_{10}H_{14}^{\bullet+}$ measured by charge exchange. Nacson and Harrison,[2] using similar techniques, concluded that 87% of the available collision energy is converted into internal energy for collision of 60 eV *n*-octylbenzene with an N_2 target while the efficiency decreases to 33% for *n*-butylbenzene with an Ar target.

In practice, two collision regimes should be distinguished: low energy, in the range of 1–100 eV, as occurs in quadrupole or ion trap instruments, and high energy, several thousands of electron volts, as is common for magnetic instruments.

For theoretically comparable energy exchange, as calculated by Equation (9), different fragmentation patterns are observed for low and high collision energy. Generally, high-energy spectra give simpler, more clear-cut fragmentations, while low-energy spectra lead to spectra with more diverse fragmentation pathways, often including more rearrangements.

3 TANDEM MASS SPECTROMETRY SCAN MODES

3.1 Product Ion Scan

Time-based MS/MS instruments can produce fragment ions from one or several selected precursors, but can only do this over several ion generations, yielding MS^n spectra. On the other hand, instruments with several mass analyzers in space can be scanned in several ways. The most important ones for two analyzers are displayed in Figure 4.

In the first scan mode an ion of a chosen m/z is selected with the first spectrometer. This ion collides inside the collision cell and fragments. The reaction products are analyzed by the second mass spectrometer. This is a *fragment ion scan* or *product ion scan*. This method used to be called a *daughter scan*. Symbolism describing the various scan modes is represented in Figure 5. This fragment ion scan mode is represented by the *product ions scan* symbolism.

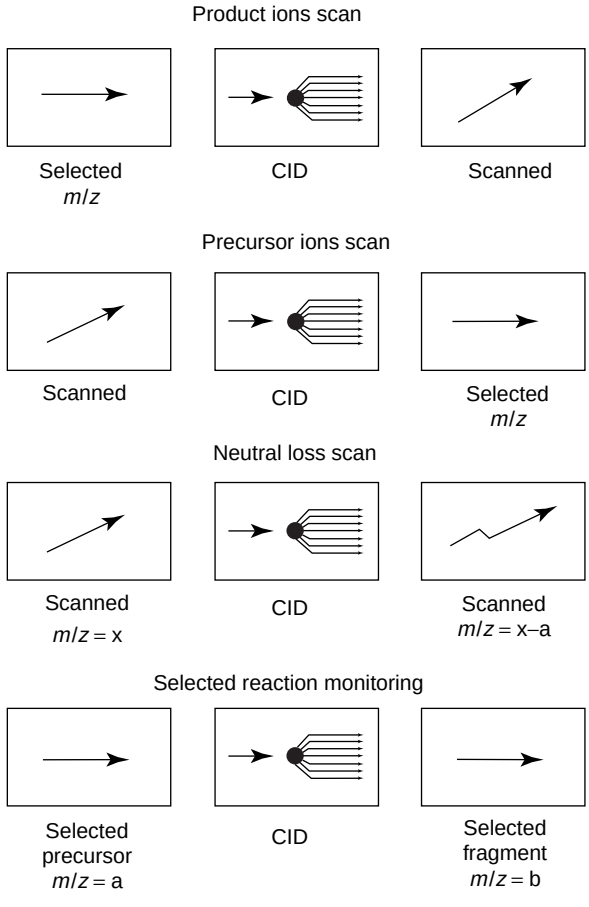

Figure 4 Different scan modes for a tandem in space mass spectrometer.

3.2 Precursor Ion Scan

In the second scan mode the second spectrometer (MS2) is focused on a selected ion, while the masses are scanned using the first spectrometer (MS1). All of the precursor ions that produce ions with the selected mass through reaction or fragmentation are thus detected. This method is called *precursor scan* because the precursor ions are identified. It used to be called *parent scan*. Using the symbolism of Figure 5, it is represented by the *precursor ions scan* symbolism.

3.3 Neutral Loss Scan

In the third common scan mode, both mass spectrometers are scanned together, but with a constant mass offset between the two. Thus, for a mass difference a, when an ion of mass m goes through the first mass spectrometer, detection occurs if this ion has produced a fragment ion of mass $(m - a)$ when it leaves the collision cell. This is a *neutral loss scan*, the neutral having the mass a. For example, in chemical ionization positive ions containing an hydroxy group very often lose a water molecule. Alcohols are thus detected by scanning a neutral loss of 18 mass units. On the other hand, a given mass increase can be detected if a reactive gas is introduced within the collision cell, but this is not frequently applied. The symbol for this scan mode is that of the *neutral loss scan* in Figure 5.

3.4 Selected Reaction Monitoring

For SRM, both the first and second analyzers are focused on selected masses. There is thus no scan. The method is analogous to selected ion monitoring in standard mass spectrometry (MS). But here the ions selected by the first mass analyzer are only detected if they produce a given fragment, by a selected reaction. The absence of scanning allows one to focus on the precursor and fragment ions over longer times, increasing the sensitivity as for selected ion monitoring, but this sensitivity is now associated with a high increase in selectivity. This is often used to detect a selected compound in an elution timeframe during chromatography. The high selectivity corresponds to a low probability of detecting an

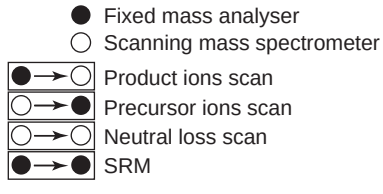

Figure 5 Symbolism proposed by Kondrat and Cooks for simple representation of various scan modes, reviewed by Schwartz et al.[3]

interfering compound. This reduces the background noise and eliminates most interferences. If required, it allows the time taken for the chromatography to be shortened. This scan mode is represented by the symbol for *SRM* in Figure 5.

4 IN-SPACE INSTRUMENTS

Many of these instruments are built especially to perform MS/MS. However, some instruments allow MS/MS experiments to be carried out with just small modifications in the instrument, most often electronic parts.

4.1 Electric/Magnetic Sector Instruments

Basic instruments in this category combine an electric sector (E) with a magnetic sector (B), in the order BE or EB. The ion paths in these sectors have respectively radii R_E and R_B.

If the geometry of the instrument is BE, a product ion scan can be performed in the following way. The field of the magnetic sector is set to allow ions of a given mass to pass. If the electric field is held at its normal value, that is the value used for standard scans, the ions of this selected mass will arrive at the detector. Fragmentation may occur between the magnetic and electric sectors, either because metastable ions spontaneously fragment, or because a collision cell has been inserted between the two sectors in the second field free region, as displayed in Figure 6. An important basic feature of fragmentations is that the velocity v_f of the fragments is almost the same as that of the precursor ions v_p, thus $v_f = v_p$. However, the mass has changed, $m_p \neq m_f$ and thus so also have the momentum and the kinetic energy. The value of the electric field E that will allow an ion to pass depends on its kinetic energy by the relation, Equation (12):

$$qE = \frac{mV^2}{R_E} \tag{12}$$

where R_E is the radius of the ion path in the electric sector and q the electric charge of the ion ($q = ze$, z is the number of elementary charges). If we consider now the values of E required to focus successively the precursor and the considered fragment, Equations (13) and (14) are obtained:

$$qE_p = \frac{m_p V^2}{R_E} \tag{13}$$

$$qE_f = \frac{m_f V^2}{R_E} \tag{14}$$

The ratio of these two electric field values is equal to the ratio of the masses, Equation (15):

$$\frac{E_f}{E_p} = \frac{m_f}{m_p} \tag{15}$$

E_f is measured, while m_p and E_p are known. Thus, m_f can be determined. Scanning E allows all the fragments of the selected precursor to be detected successively. Actually, what is measured is not the mass, but the kinetic energy of the fragment. This is why the method is called MIKE spectroscopy. If some kinetic energy is released during the fragmentation, for instance by the opening of a strained small cyclic structure or relaxation of steric interaction, this will result in a small dispersion of the kinetic energy, leading to a broadening of the peak while scanning E. This broadening allows the kinetic energy released to be measured quantitatively.

Another way to obtain fragment ion spectra from sector instruments of either BE or EB configuration is to use a B/E linked scan. Remember that after fragmentation the product ions have the velocity of the precursor, but differ by the mass. As the focusing condition in a magnetic sector is given by Equation (16)

$$qB = \frac{mV}{R_B} \tag{16}$$

where B is the magnetic field, R_B is the radius of the guide tube in the magnetic sector, q is the charge of the ion, m is its mass and V is its velocity. Applied to a precursor ion and one of its fragments, Equations (17) and (18) are obtained:

$$qB_p = \frac{m_p V}{R_B} \tag{17}$$

$$qB_f = \frac{m_f V}{R_B} \tag{18}$$

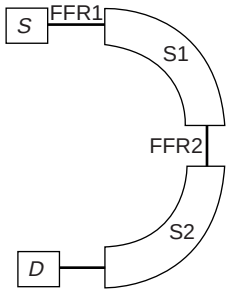

Figure 6 Schematic diagram of an electric/magnetic sector mass spectrometer. One of the two S1 or S2 sectors is an electric sector E, while the other is a magnetic one, B. The two field free regions FFR1 and FFR2 can be used to fragment precursor ions. If FFR2 is used, mass analyzed ion kinetic energy (MIKE) spectroscopy is performed and a BE configuration is needed. If FFR1 is used, the configuration may be either BE or EB. S is the source and D the detector.

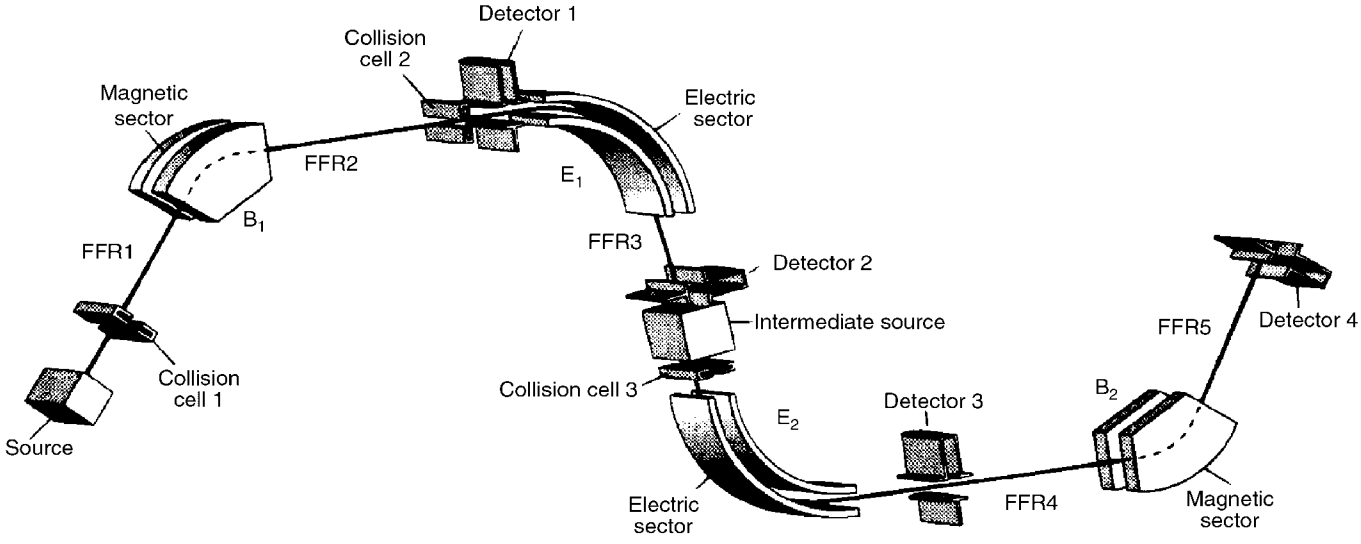

Figure 7 One example of a multiple sector magnetic mass spectrometer. This configuration is BEEB, with four detectors, four field free regions useful for MS/MS (FFR1 to FFR4). The first three have collision cells. Several intermediate detectors are also provided. (Reproduced from de Hoffmann.[4])

The ratio of these two equations gives Equation (19):

$$\frac{B_f}{B_p} = \frac{m_f}{m_p} \tag{19}$$

The corresponding focusing conditions for the electric sector have been deduced for MIKE spectroscopy and are expressed by Equation (20):

$$\frac{E_f}{E_p} = \frac{m_f}{m_p} \tag{20}$$

As they both equal the ratio of the fragment mass to the precursor mass, one can write Equations (21) and (22):

$$\frac{E_f}{E_p} = \frac{B_f}{B_p} \tag{21}$$

or

$$\frac{B_p}{E_p} = \frac{B_f}{E_f} \tag{22}$$

Thus, to select a precursor ion, the operator adjusts the values of B and E to those used for that ion in normal scan. The precursor ion is therefore focused in both sectors. Then, both sectors are scanned together in such a way that the ratio B/E remains constant and equal to B_p/E_p. The successive fragment ions are detected and the mass assigned from either the corresponding B or E value. The name B/E linked scan has been coined for this scan mode. Other scan modes such as precursor or neutral loss scans can also be performed by other linked scans of magnetic instruments, but are rarely used.

4.2 Multiple Sector Magnetic Instruments

Many combinations of more than two sector magnetic instruments are in use. Figure 7 displays an instrument with BEEB configuration. Many more types of experiment can be performed with such elaborate instruments. First, high resolution can be achieved in principle both for precursor selection and for product ion analysis. It should be noted however that sensitivity is lost when the resolution is increased. Because the collision process also results in the loss of the majority of the ions, more than one high resolution stage is almost never used, and one only rarely. Some very specific analyses, such as quantification of dioxine, require a high-resolution selection of the precursor ions.

To analyze ions with a sector analyzer, the ions must have about the same kinetic energy. If, for example, ions fragment in the field-free region after the first two BE sectors, in FFR3, the fragments will not have the same kinetic energy because, as described before, they will all have the velocity of their precursor but different masses. This means that the last two sectors must be scanned in the B/E linked scan mode. For different precursors, different calibrations of the two last sectors are needed.

An alternative way is to slow down the ions, make the collisions at low energy, i.e. a few electron volts, and then reaccelerate at several kiloelectron volts. This allows there to be product ions with about the same kinetic energy, but many ions are lost in the process. There is also still some dispersion in kinetic energy, which reduces the resolution that can be obtained.

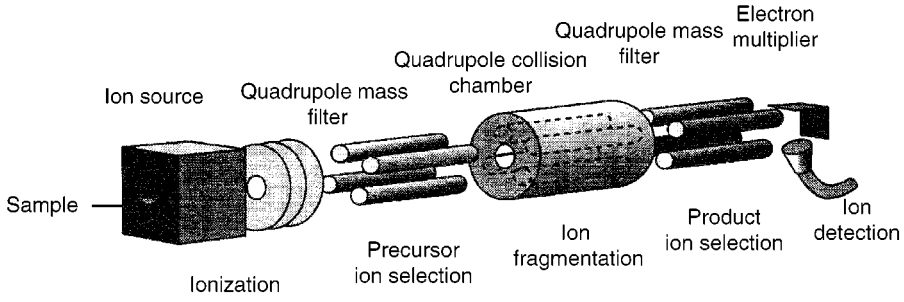

Figure 8 View of a triple quadrupole instrument. (Reproduced from de Hoffmann.[4])

4.3 Triple Quadrupole Instruments

A single quadrupole mass spectrometer cannot be used to perform true MS/MS experiments. However, with an electrospray source it is possible to cause the ions to fragment by adjusting the kinetic energy of the ions in a region where the gas density is still high enough to cause collisions. If under normal conditions, the source produces only ions of the molecular species, these collisions will produce fragments and thus produce a kind of MS/MS fragment ion spectra. However, if fragment ions are present or if a mixture is analyzed, all the fragments or all the ions of the different species will fragment together. This is, of course, avoided by using a true MS/MS instrument, which permits the selection of ions of a chosen mass. Instruments that include two quadrupole mass analyzers (Q) separated by an RF-only (radio frequency) quadrupole collision cell (q) are the most common tandem-in-space mass spectrometers available (Figure 8). They are symbolized QqQ, where Q stands for a quadrupole mass analyzer, while q refers to a quadrupole collision cell.

Quadrupoles have the advantage of being true mass filters. Magnetic sectors separate ions according to their momentum, electric sectors according to their kinetic energy, time-of-flight (TOF) instruments according to their m/z velocities, but quadrupoles according to their values. They are largely not sensitive to variations in kinetic energy and thus do not require special scans of the second analyzer or special calibration. The drawbacks are the impossibility of obtaining high resolution, a lower sensitivity in the higher mass region and the limitation of using only collisions at low energy. Besides their lower cost, their success is due to the great ease of their use.

To be used as a mass analyzer, both an RF and a DC (direct current) potential have to be applied together on the rods. If only a RF potential is applied, ions of all m/z values above a given low mass cutoff are allowed to pass. This low mass limit depends on the applied RF potential and is usually set at m/z 4. The advantage of using a RF-only quadrupole as a collision cell is the focusing property of this cell. When ions collide with a target and fragment,

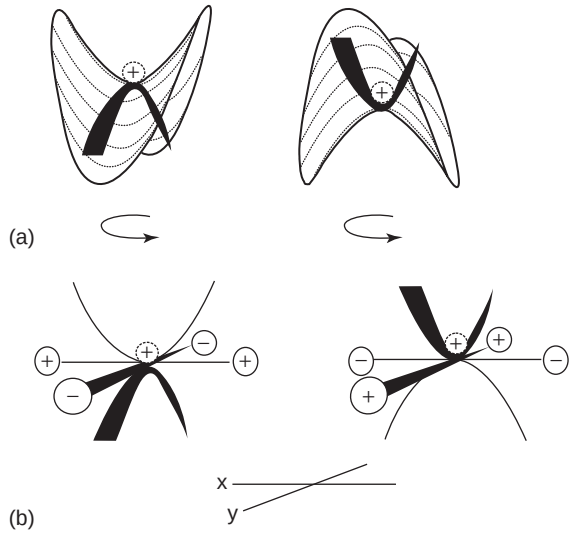

Figure 9 (a) Ball-on-a-saddle-point model of ion trapping. The ball will tend to roll down, but if the saddle is rotated, the ball will be constrained to roll uphill to the middle of the saddle. Similarly (b), a positive ion submitted to a quadrupolar field will move to a negative rod, and vice versa for negative ions. But if the polarity changes quickly, the ion will be brought back to the center of the rods. This is the principle of the focusing effect of RF-only quadrupole collision cells. (Reproduced from de Hoffmann.[4])

the fragment ions produced are scattered in all directions and a majority of ions are lost. The use of an RF-only quadrupole causes the ions to be refocused at the center of the quadrupole. This can be visualized by comparison with a ball on a saddle, as represented in Figure 9. A positive ion will travel toward the negative rod, but owing to the frequency of oscillation of the field, the polarity of the positive rods will quickly change to negative and vice versa. This change in polarity can be compared to a saddle on which a ball has been placed. The ball will roll down the slope, but if the saddle is quickly rotated by 90°, the rolling ball will face uphill and will roll back to the center of the saddle. If there are several balls and the trajectories change by collision between them, they will be driven back to the center by the effect of the fast

rotating saddle. This is what happens to the ions in the central quadrupole. Even if they undergo collisions with a neutral gas in the cell, the effect of the RF will be to bring them back to the center. This prevents the loss of ions by scattering after collision.

An offset voltage between the source and this RF-only quadrupole can be adjusted to allow the collision energy to be varied between zero and several hundred electron volts. This is low compared with magnetic instruments where the usual values are fixed somewhere in the range 2–10 keV. However, under normal operating conditions, a relatively large number of collisions can occur in the collision cell, causing the build-up of energy in the precursor ions. Conversion of precursor ions to fragments can usually occur to a much larger extent than possible with a magnetic sector instrument. Important advantages of triple quadrupole instruments are the relatively lower cost and ease of use. Once both quadrupole mass analyzers have been calibrated, switching between different scan modes and mass ranges can be done instantaneously. Furthermore, unit resolution on both quadrupole mass analyzers is obtained for all types of MS/MS experiment.

Instruments with more than two quadrupole mass analyzers have been constructed in research laboratories, but have never been used commercially. A review paper has been published on QqQqQ instruments.[5]

4.4 Time-of-flight Mass Spectrometers

When ions fragment, the fragments have almost the same velocity as their precursors and will thus not be distinguished by their TOF. However, if the instrument is equipped with a reflectron device, then MS/MS can be performed. This is best achieved by combining a linear with a reflectron spectrometer, as shown in Figure 10. When an ion fragments, its product ions have the same velocity as the precursor and thus the same flight time in the absence of any field. However, they will have different kinetic energies. If m_p is the mass of the precursor and m_f the fragment mass, their respective kinetic energies K_p and K_f are (as they have the same velocity v_{ix}) given by Equations (23) and (24):

$$K_p = \frac{m_p v_{ix}^2}{2} \tag{23}$$

and

$$K_f = \frac{m_f v_{ix}^2}{2} \tag{24}$$

thus $K_p/K_f = m_p/m_f$ or $K_f = K_p \frac{m_f}{m_p}$.

The penetration depth x in the reflectron is given by $x = K/qE$. Hence, for the precursor and the fragment,

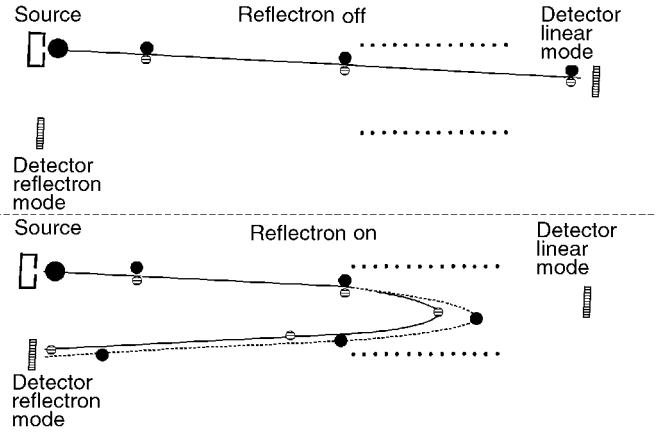

Figure 10 MS/MS with combined linear and reflectron TOF instrument. ● = Bunch of ions with one given mass leaving the source; during the flight, a fraction of these ions fragments; • = Survivor ions; ⊖ = Fragment ions. Those ions fragmenting between the source and the reflectron are called post-source decay (PSD) ions. (Reproduced from de Hoffmann et al.[6])

the Equations (25) and (26) are relevant:

$$x_p = \frac{K_p}{qE} \tag{25}$$

and

$$x_f = \frac{K_f}{qE} = \frac{K_p(m_f/m_p)}{qE} \tag{26}$$

Hence Equation (27)

$$x_f = x_p \frac{m_f}{m_p} \tag{27}$$

The respective flight times in the reflectron are given by Equations (28) and (29)

$$t_{rp} = \frac{4x_p}{v_{ix}} \tag{28}$$

and

$$t_{rf} = \frac{4x_f}{v_{ix}} = \frac{4x_p(m_f/m_p)}{v_{ix}} \tag{29}$$

Hence Equation (30)

$$t_{rp} = 4x_{rp} \frac{m_f}{m_p} \tag{30}$$

This demonstrates that the time spent in the reflectron will be shorter for the fragment than for the precursor, the ratio of the respective times being equal to the mass ratio. Thus, as shown in Figure 10, the TOF will be the same for the precursors and fragments in the linear mode, but will differ in the reflectron mode. When comparing the two spectra, ions observed in the reflectron spectrum but

not in the linear mode result from PSD fragmentations having occurred between the source and the reflectron device.

An alternative method involves detecting the neutrals with a linear detector.[7] This method requires the acceleration of the ions at several tens of kiloelectron volts, typically 30 keV, in order to have enough kinetic energy in the neutrals to be detectable. This is illustrated in Figure 11.

If the instrument is fitted with electrodes after the source on the ion path, a precursor ion can be selected to analyze its fragmentation by PSD. As shown in Figure 12, a potential is applied to eliminate the ions. Cutting off this potential for a short time, corresponding to the passage of the ions with a given mass allows these ions to be selected. The resolution on modern (1998) instruments is about 50. This gives a m/z 20 window at m/z 1000. The mass of the passing ions can be determined in the linear flight mode, while the fragments are observed in the reflectron mode.[8]

Cornish and Cotter[9] have developed a compact laser desorption tandem TOF mass spectrometer. The instrument incorporates two dual-stage reflectron analyzers

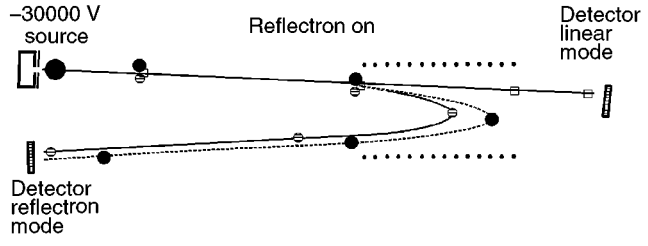

Figure 11 ● = Parent ions or survivor ions; ⊖ = Fragment ions; = Neutral fragments. The last are not deflected and reach the linear detector. They will be detected provided that the parent ions have been accelerated at a sufficient kinetic energy, typically 30 keV. (Reproduced from de Hoffmann et al.[6])

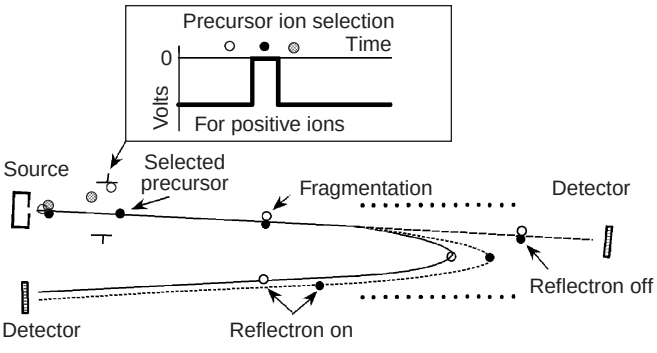

Figure 12 TOF spectrometer fitted with a deflection electrode for precursor ion selection. The instrument can be operated in either the linear or the reflectron mode. The selection resolution is about 50, which, for example, gives a m/z 20 window at mass m/z 1000. (Reproduced from de Hoffmann et al.[6])

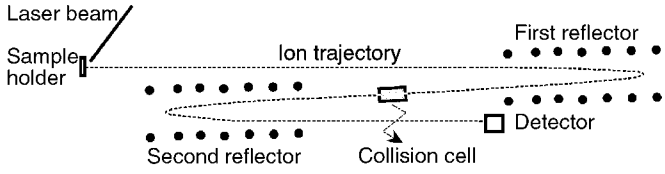

Figure 13 Schematic representation of a tandem mass spectrometer comprising two TOF with reflectron analyzers and, in between, a collision cell. (Reprinted with permission from R.J. Cotter,[11] *Time-of-flight MS: Instrumentation and Applications in Biological Research*, Copyright 1997, American Chemical Society.)

separated by a collision region for producing ions by collision-induced dissociation (CID) (Figure 13). The first reflectron provides energy focusing while the second provides both energy focusing and time dispersion of ions having different masses.

They introduced a curved field reflectron[10] to allow the recording of product ion spectra without stepping or scanning of the reflectron potential.

5 IN-TIME INSTRUMENTS

5.1 Ion Trap-based MS^n

Instruments with ion storage capabilities, i.e. ion traps or ICR (ion cyclotron resonance) cells, allow "in-time" multiple stage MS^n experiments to be performed with an appropriate time sequence of events, as represented for simple MS/MS experiments in Figure 2. Ion traps, like quadrupole instruments, rely on oscillating electric fields to operate, while ICR instruments makes use of a magnetic field to store the ions associated with oscillating electric fields to perform the various mass spectroscopic experiments.

A schematic representation of an ion trap is given in Figure 14.

Ion traps can be operated in various ways, by applying DC and RF voltages to the ring electrode or to the end caps. However, in commercial instruments, a ring RF voltage is applied to the ring electrode. Additional RF voltages are applied between the end caps for MS^n operations. These are the ion injection, ion expulsion, ion selection and ion excitation RF voltages.

To understand the MS/MS operation in an ion trap, let us first recall the typical stability diagram for an ion trap as represented in Figure 15. From Equation (31)

$$r_0^2 = 2z_0^2 \tag{31}$$

then Equations (32) and (33) of the coordinates of this stability diagram, a_z and q_z can be written for

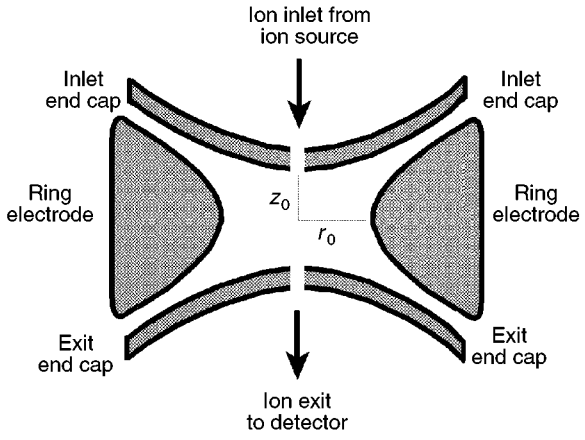

Figure 14 Schematic representation of an ion trap. It is composed of one ring electrode and two end cap electrodes. z_0 is the distance from the center to the end cap and r_0 is the radius of the ring electrode.

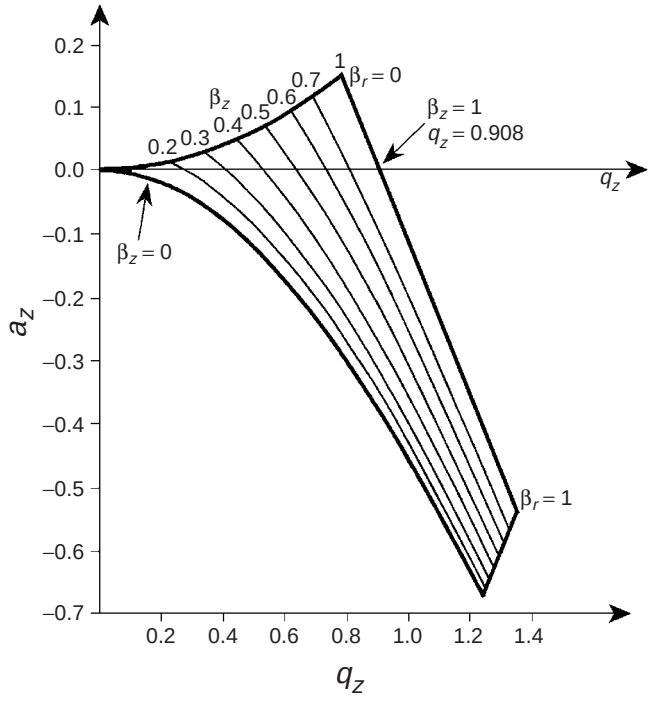

Figure 15 Typical stability diagram for a quadrupole ion trap. (Redrawn from March and Hughes[12] (Figure 3.62).)

monocharged ions respectively:

$$a_z = \frac{-8eU}{mr_0^2\Omega^2} \tag{32}$$

$$q_z = \frac{-4eV}{mr_0^2\Omega^2} \tag{33}$$

In these equations, e is the charge of a monocharged ion (1.6×10^{-19} C), U is the DC voltage applied to the ring electrode, r_0 is the radius of the trap in centimeters with a typical value of 1 cm, and Ω (impulse or angular velocity) is equal to $2\pi\nu$, where ν is the frequency of the RF potential, and V is the amplitude zero to peak in volts of this RF voltage. The common order of magnitude of the frequency is around 1 MHz and V varies between about 0 and 7000 V. An ion will only have a stable trajectory if its a_z and q_z values are inside the boundaries indicated in the diagram. These boundaries depend on a β stability parameter, whose values have to be between 0 and 1 in both the r and z directions.

Iso-β_z lines are also displayed in Figure 15. β_z can be calculated from a_z and q_z. For q_z lower than 0.4, it can be approximately calculated from Equation (34):

$$\beta_z = a_z + \frac{q_z^2}{2}^{1/2} \tag{34}$$

We will now consider, as is the case for most commercial ion traps, that $U = 0$, and thus $a_z = 0$. Thus, the ion trap operation line is on the q_z axis. In the expression of q_z, $-4e$ is a constant term for monocharged ions, and r_0 and Ω are constant for a given ion trap. Then q_z is proportional to V/m : for ions with a fixed mass m, q_z will be proportional to V. Respectively, at a fixed V value, ions with higher masses will be characterized by a lower q_z value.

It is important for MS/MS that ions will oscillate at a secular frequency f_z, linked to the ring electrode RF

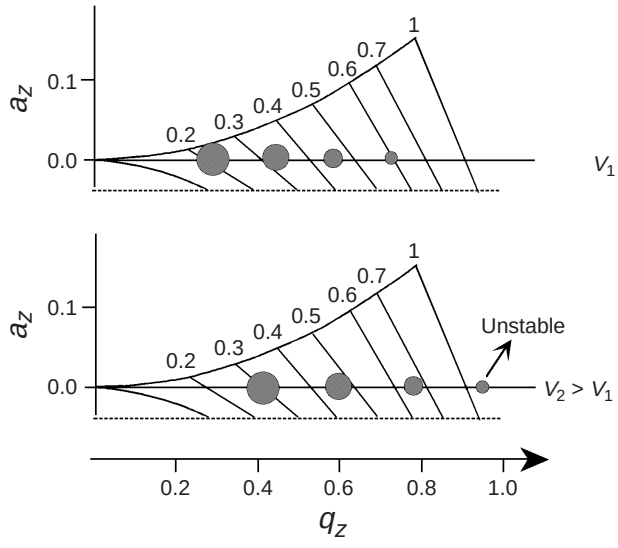

Figure 16 At a fixed value of the RF potential V applied to the ring electrode, heavier ions will have lower β_z values and thus lower secular frequencies. If V is increased, β_z values increase for all the ions, and thus also the secular frequencies. In the example given, the lightest ion now has a β_z value larger than one and is thus expelled from the trap.

frequency v by Equation (35):

$$f_z = \frac{\beta_z v}{2} \tag{35}$$

As the maximum value for β is 1, it follows that the maximum secular frequency f_z for an ion with a stable trajectory will be $v/2$.

Thus at a given V value, ions of mass m will oscillate at a frequency f_z. If an RF voltage of this same frequency f_z is applied between the end caps, it will be in resonance. Thus these ions will oscillate with a larger and larger amplitude, depending on the amplitude of the applied RF voltage and the time. If this amplitude is low, ions will progressively gain energy and dissociate by collisions with helium gas, which is always present in the trap. At higher amplitude, the ions will be quickly expelled from the trap in the z direction.

Figure 16 shows ions whose masses are represented by the volumes of the spheres. It shows that for a fixed value of the amplitude V of the RF potential applied to the ring electrode, ions with a lower mass have a larger β_z value, and thus a higher secular frequency f_z. If V is increased, the secular frequencies of all the ions increase.

By irradiation at all the frequencies except the one corresponding to ions of a selected mass, all the ions can be expelled from the trap except the selected ones. This is achieved by irradiation with an appropriate waveform, calculated by inverse Fourier transformation (Figure 17). This is the SWIFT method, selected waveform inverse Fourier transform. Alternatively, the waveform can be applied during ion injection from the source. Only ions of one selected mass are then stored; this is also used in selected ion monitoring.

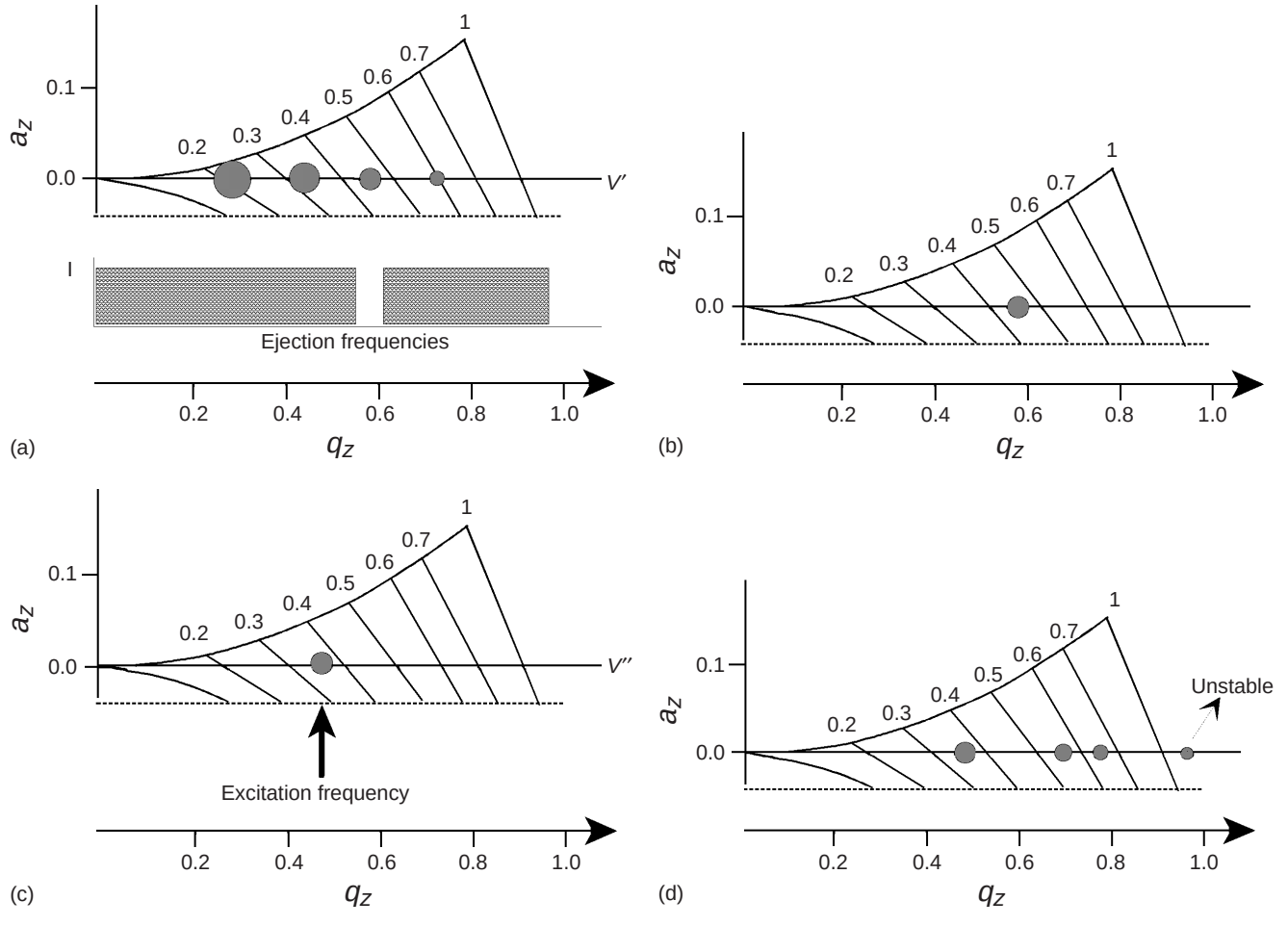

Figure 17 (a) Ions from the source are stored in the trap. These ions are irradiated by applying a waveform to the endcaps. (b) Only the ions whose secular frequency was absent in the waveform remain in the trap. (c) The ring RF potential V is changed to adjust the secular frequency of the selected ions to the excitation frequency applied to the endcaps. (d) Fragment ions are produced. The smallest will be expelled because it is not within the stability boundaries. Afterwards, the fragment ions can be analyzed, or the sequence of isolation/excitation can be repeated on one of the fragments; if this is repeated $(n-1)$ times, an MS^n experiment is performed.

After ions of a selected mass have been stored, fragmentation must be induced. To do this, it is possible to rely on multiple collisions of the ion with the helium damping gas. Fragmentation can be improved by causing the ions to oscillate at an increased amplitude, which can be performed as explained before by irradiating the ions at their secular frequency. Often, a fixed frequency is used for irradiation. Then the secular frequency of the selected ions must be adjusted by changing the V amplitude of the circular electrode RF potential. This process is illustrated in Figure 17. To be observed, fragment ions have to be inside the stability boundaries, i.e. $0 < \beta_z < 1$. Generally, fragments whose mass is less than about 25% of the mass of the precursor will be lost. This sequence of ion isolation, excitation, and analysis of the product ions can be repeated on one of the fragments. This can be done n times, allowing observation of successive generations of ions in a MS^n experiment. Typically for analytical purposes a maximum of five or six steps are used.

An advantage of ion traps in comparison with beam instruments is the high efficiency of ion transmission. When an ion is fragmented, the total ion current of the fragments can be 100% of the ion current of the fragmented precursor ion. In practical use, this transmission is at least 50%. For beam instruments, this figure is somewhere between 1 and 10% typically. The advantage of this efficiency increases exponentially with the number of fragmentation stages.

5.2 MS^n in Ion Cyclotron Resonance Fourier Transform Mass Spectrometry Instruments

In an ICR cell, ions exposed to an axial B magnetic field and having a charge ze have a cyclotron motion with frequency given by Equation (36):

$$\varpi_c = \frac{zeB}{m} \tag{36}$$

Thus, monocharged ions will have, for a given value of B depending on the instrument, a frequency inversely related to their mass. They can be detected by the current they induce in the wall of the cell, in a nondestructive way. This is a distinct feature of ICR, important in MS/MS experiments. By applying an AC (alternating current) electric field perpendicular to the axis, it is possible to excite ions, or to expel ions by discharge on the wall. This allows one to expel all the ions except those having a selected mass. These ions are then allowed to fragment over time. Sometimes a collision gas is introduced for a short time to induce fragmentation by collision, and excitation by irradiation at the cyclotron frequency is also used. However, sustained excitation at the resonance frequency results in larger cyclotron radii and fragment ions are then produced with nonzero magnetron radii. It is now common practice to use sustained off-resonance irradiation (SORI).[13,14] This results in ions being alternatively accelerated and decelerated, limiting the cyclotron radius. The fragments are then produced close to the center. Fragments are detected while they are formed in a nondestructive way. They can be detected repetitively, increasing sensitivity and resolution. The selection and fragmentation process can then be repeated, providing MS^n capability without reloading ions from the source, as the detection is not destructive. This contrasts with the situation in ion traps, where the detection of ions empties the trap which has then to be reloaded from the source.

However, the magnetic field has no focusing properties. Over time, there is an off-axis displacement of the center of the ion cyclotron orbit, known as magnetron radial expansion. Reaxialization can be performed using the focusing properties of a quadrupolar AC electric field (see Figure 9). Combined with collision damping, this allows the magnetron radial expansion to revert and greatly improves the resolution and the sensitivity, especially by reducing the loss of ions during MS/MS experiments. High resolution can be obtained on both precursor and fragment ions.

6 HYBRID INSTRUMENTS

Hybrid instruments are made by combining different types of mass analyzer to obtain a tandem mass spectrometer. Many hybrid instruments have been built. The most common ones result from coupling a magnetic instrument with a quadrupole, either in the BEqQ or the EBqQ configuration. Between the last sector and the RF-only quadrupole, ions must be decelerated from kiloelectron volts typical of magnetic instruments to tens of electron volts needed for quadrupole analysis.

If this is performed before CID fragmentation, all the ions leaving the magnetic instrument have the same kinetic energy, and deceleration occurs at the same potentials for all. CID then occurs in the RF-only quadrupole, at low energy, and fragments are analyzed by the last quadrupole.

Alternatively, CID can be effected in a collision cell at the focal point at the exit of the last sector. High-energy collisions occur. The ions are then slowed down before they enter the quadrupole stage. However, all the ions no longer have the same kinetic energy, because after collision, fragment ions have the same velocity

and thus lower kinetic energy than their precursors. The ions must therefore be slowed down by a difference of potential on the electrodes varied as a function of the mass.

Another instrument combines a magnetic or quadrupole mass spectrometer with an orthogonal TOF analyzer. As already mentioned, when ions fragment during their flight, with or without CID, precursor and fragments have the same velocity and will arrive together at the detector. If however they are accelerated perpendicularly to their flight in the TOF mass spectrometer, they will arrive at different times at the array detector and will thus be detected according to their mass. Figure 18 displays the scheme of such an instrument.

The combination of a quadrupole analyzer, a quadrupole collision cell and a reflectron TOF analyzer gained in importance in 1998. The reflectron TOF allows high-resolution spectra to be obtained, with typically a 10 000 resolution.

These instruments are used especially in the area of protein analysis. Short sequences of amino acids from an unknown protein can be compared using a database to find out to which protein they correspond. The efficiency of this search is greater if the mass is known with accuracy.

Another advantage of this high resolution is that it allows the charge state of ions to be determined. Indeed, electrospray ionization (ESI) produces multiply charged ions. From a statistical analysis, the charge state z can be determined and thus from the measurement of m/z, the mass m of the protein can be assigned.

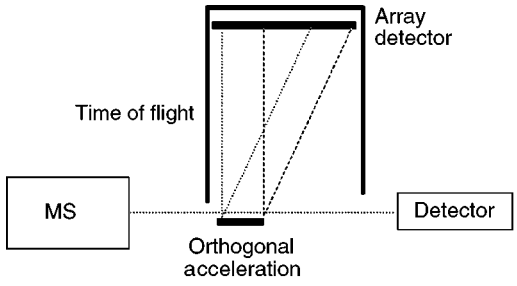

Figure 18 Principle of the combination of a mass spectrometer with an orthogonal TOF spectrometer. Ions arriving from the mass spectrometer are directed to the detector. When a pulse voltage is applied on the orthogonal acceleration repeller, the ions are analyzed by the TOF instrument. Faster ions will arrive farther along the array detector. (Reproduced from de Hoffmann.[4])

If a multiply charged ion is selected as a precursor in MS/MS, the assignment of the charge state of the fragments is not straightforward. However, if the precursor ion is selected with a sufficiently low resolution, the natural isotopes will be included. As they have an entire multiple Dalton difference, they will appear in the mass spectrum at m/z values at a distance of $1/z$ from each other. This is illustrated in Figure 19.[15]

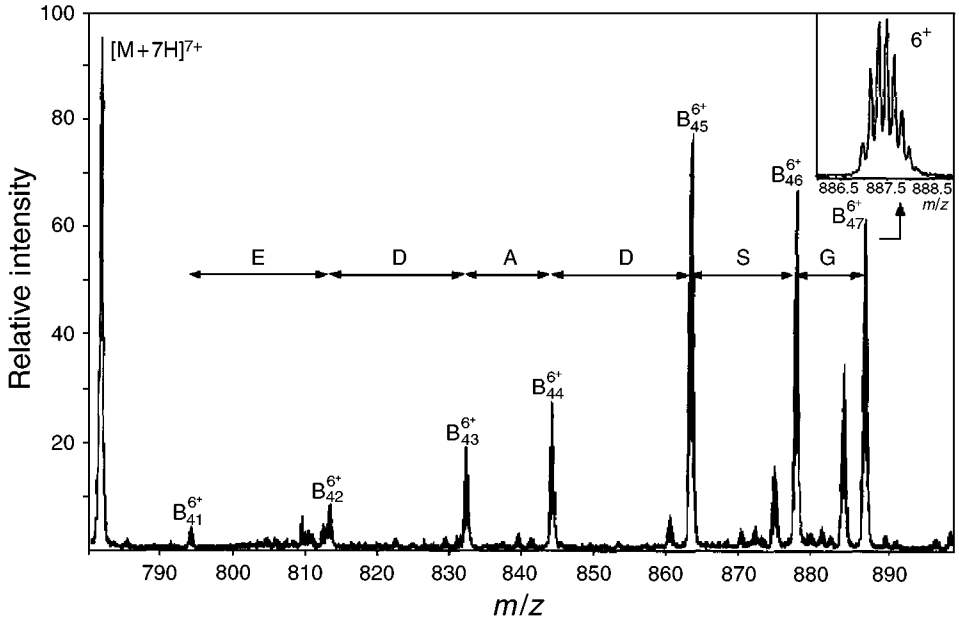

Figure 19 Product ion spectrum of the $[M + 7H]^{7+}$ ion from the peptide ALVRQGLAKVAYVYKPNNTHEQHLRKSEAQAKKEKLLNIWSEDNADSGQ. Fragment ions having lower charge number z may appear at higher m/z values than the precursor, which indeed occurs in the spectrum shown. The inset shows that, owing to the high resolution, the isotopic peaks appear at a distance of 1/6 mass units, and thus $1/z = 1/6$ or $z = 6$. (Reproduced with permission from Andersen et al.[15])

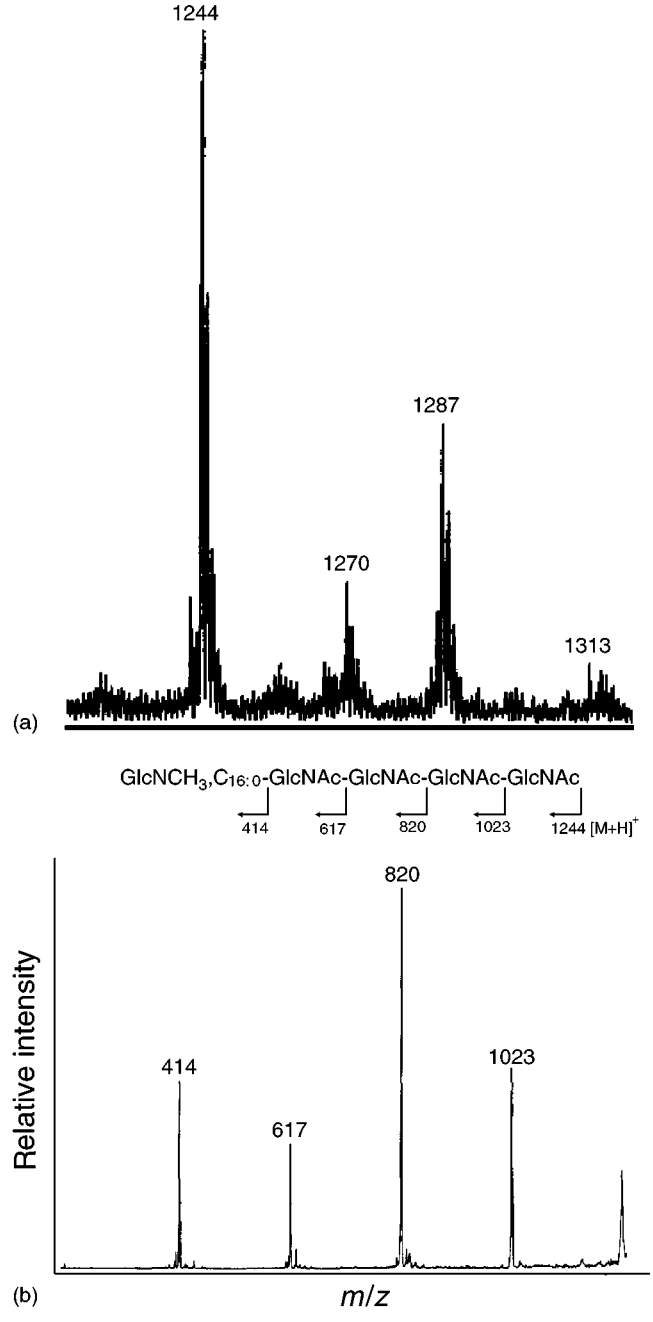

Figure 20 Structure elucidation by MS/MS of a nodulation factor. (a) m/z values observed by FAB (fast atom bombardment) from a chromatographic fraction of Nod factors. (b) the m/z 1244 ion is selected as precursor and fragmented by a B/E linked scan in a magnetic instrument. The clear-cut fragmentation at the glycosylic bonds allows straightforward sequence assignment as shown by the fragmentation scheme. (Reproduced from Mergaert et al.[16])

As the resolution is defined as $m/\delta m$, the required resolution is 887/(1/6), i.e. about 5400. TOF instruments with a reflector have resolutions in the 10 000–20 000 range.

7 ANALYTICAL METHODS BASED ON TANDEM MASS SPECTROMETRY

7.1 Structure Elucidation

An example of structure elucidation by MS/MS is displayed in Figure 20. The FAB ionization mass spectrum of an isolated fraction contains several nodulation factors appearing at different m/z values. One of them, m/z 1244, is selected and fragmented using a B/E linked scan on a magnetic instrument. The observed fragment ions suggest the structure shown, corresponding to features known for this class of compounds, i.e. an oligosaccharide bearing at one end an alkyl chain bound to an amino sugar.

The analogy to a chromatographic separation coupled to a mass spectrometer should be mentioned. In this case, the FAB mass spectrum is analogous to a chromatographic trace displaying compounds of the mixture. The B/E linked scan yields the mass spectrum of one of these compounds. While chromatography gives a separation requiring a long time, the separation according to the mass is almost instantaneous, providing a much shorter analysis time.

Another typical example of structure determination by MS/MS is the widely used sequence determination of peptides. As an example, an ESI spectrum of mellitin obtained by a hybrid magnetic orthogonal TOF instrument is displayed in Figure 21. The notation of the fragments is as shown in Scheme 1.

This example also illustrates the clear-cut fragments obtained from a doubly charged ion produced by ESI: doubly charged fragments, appearing at lower m/z values, belong to the b series, while monocharged ones, some of

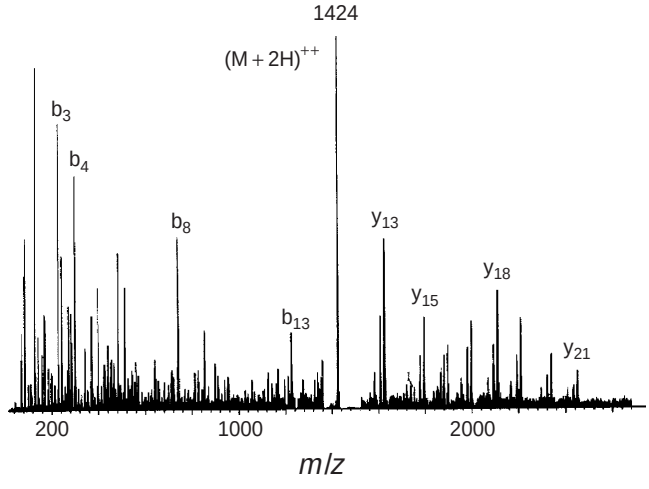

Figure 21 ESI tandem mass spectrum obtained with an orthogonal TOF mass spectrometer for mellitin at 400 eV in air with CID. $E_{com} = 7.8$ eV at 600 MHz time-to-digital converter. The doubly charged ion of the molecular species yields monocharged y fragments. Courtesy Micromass.

them appearing at m/z higher than the precursor ion, belong to the y series.

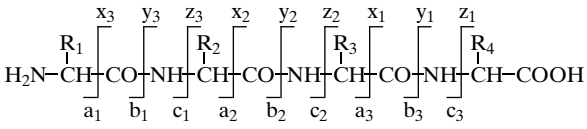

Scheme 1

MS/MS also allows the determination of isomers. As an example, the product spectra of the acylglycerol pseudomolecular ions contain a type of ion whose formal mass-based composition corresponds to a ketone obtained by the combination of two fatty acid chains with a carbonyl group, minus a proton. The ketone contains mainly the chains of the central fatty acid combined with one of the two external fatty acids, even if a ketone containing the two external fatty acids is always present with a much weaker intensity. The formation of these ions may be explained by an internal Claisen condensation followed by a fragmentation and a decarboxylation, as is shown in Figure 22. The expulsion of the ketone containing the hydrocarbon chains of the central fatty acid combined with one of the outermost fatty acids (Figure 22, pathway c) requires the formation of a neutral epoxide, which is faster than the neutral oxetane formation necessary for the expulsion of the ketone resulting from the condensation between the two outermost fatty acids (Figure 22, pathway f).

This method can be directly applied to acylglycerols present in a mixture, as is illustrated by the mass spectrometric analysis of natural cocoa butter.[17] This analysis, Figure 23, allowed the determination of the complete structure of the predominant acylglycerols in this cocoa (Table 1).

7.2 Determination of Diastereoisomers

Glycosylmonophosphopolyisoprenols[18,19] display abundant $(M-H)^-$ ions in the negative ion FAB mode.

When these ions are fragmented under low-energy collisions, an abundant fragment corresponding to the phospholipid anion is observed, together with fragments across the sugar ring. This fragmentation is very sensitive to the stereochemistry, as illustrated in Figure 24. Decaprenylphospho-D-ribose displays the decaprenylphosphate anion as almost the sole fragment,

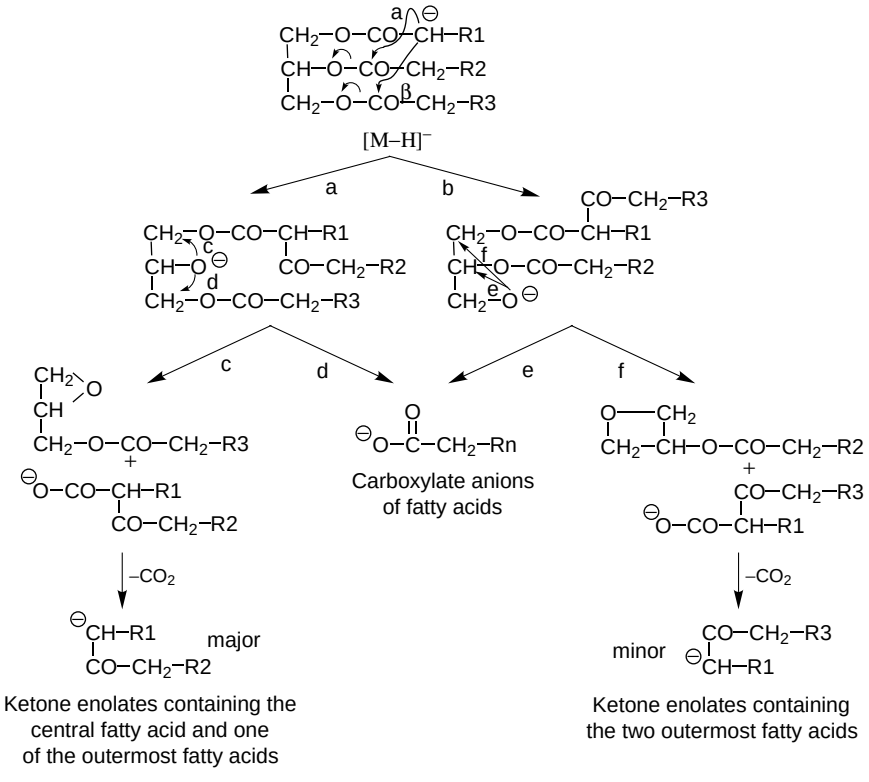

Figure 22 Claisen condensation mechanism for the formation of ketones, allowing the determination of positional isomers. (Reprinted by permission of Elsevier Science from V. Stroobant, R. Rozenberg, E.M. Bouabsa, E. Deffense, E. de Hoffmann, 'Fragmentation of Conjugate Bases of Esters Derived from Multifunctional Alcohols Including Triacylglycerols', *Journal of the American Society for Mass Spectrometry*, Vol. 6, 498–506, Copyright 1995 by American Society for Mass Spectrometry.)

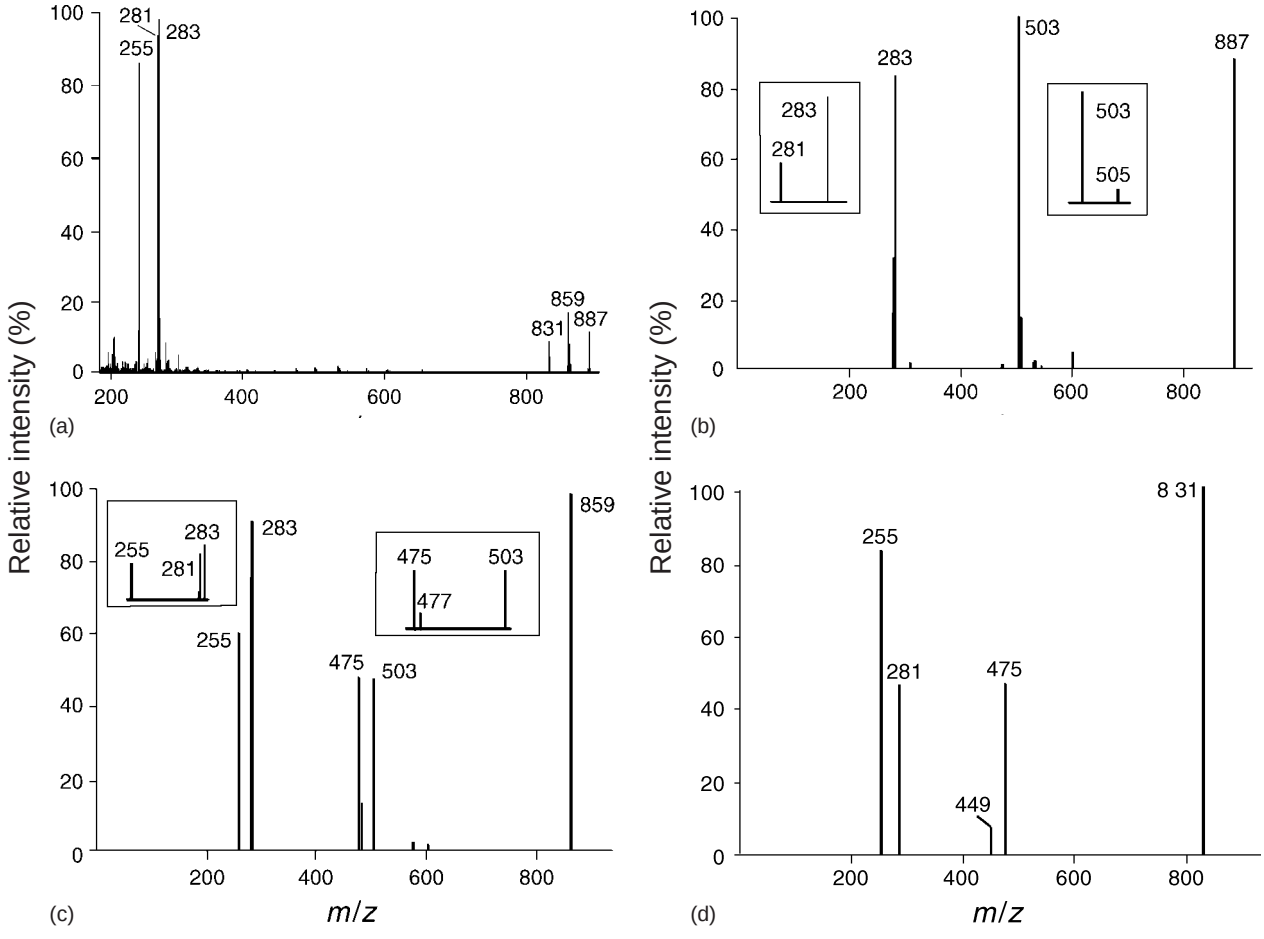

Figure 23 DCI (desorption chemical ionization) mass spectrum of natural cocoa butter (a) and DCI MS/MS CID traces of pseudomolecular negative ions of (b) *m/z* 887, (c) *m/z* 859 and (d) *m/z* 831. (Reprinted by permission of Elsevier Science from V. Stroobant, R. Rozenberg, E.M. Bouabsa, E. Deffense, E. de Hoffmann, 'Fragmentation of Conjugate Bases of Esters Derived from Multifunctional Alcohols Including Triacylglycerols', *Journal of the American Society for Mass Spectrometry*, Vol. 6, 498–506, Copyright 1995 by American Society for Mass Spectrometry.)

Table 1 Masses observed in the various fragment spectra of the three pseudomolecular ions present in the DCI spectrum of natural cocoa butter

$(M-H)^-$	R_nCOO^-	$R_nCOR'_n$	Deduced structure
831	255 (P), 281 (O)	475 (PO), *449*	POP
859	255 (P), 281 (O), 283 (S)	475 (PO), 503 (OS), *477*	POS
887	281 (O), 283 (S)	503 (SO), *505*	SOS

P = palmitic acid, O = oleic acid, S = stearic acid. Figures in italic indicate ions of low abundance.

while the analogous D-arabinose derivatives also produce fragments resulting from cleavage across the sugar ring.

From a study of several compounds, it appears that a 2-hydroxy group *trans* to the phosphate strongly favors the formation of the phospholipid anion. The ribose derivative differs from the arabinose derivative only by the *trans* configuration of the 2-hydroxy group with respect to the phosphate.

7.3 Methodology for Compound Class-selective Detection

A typical methodology to develop class-selective analysis of mixtures will now be illustrated. Figure 25 displays the CID product ions spectrum of octanoylcarnitine and a corresponding fragmentation scheme is shown in Figure 26. A typical feature of this product ion spectrum is the absence of isotopic masses. Indeed, the selected

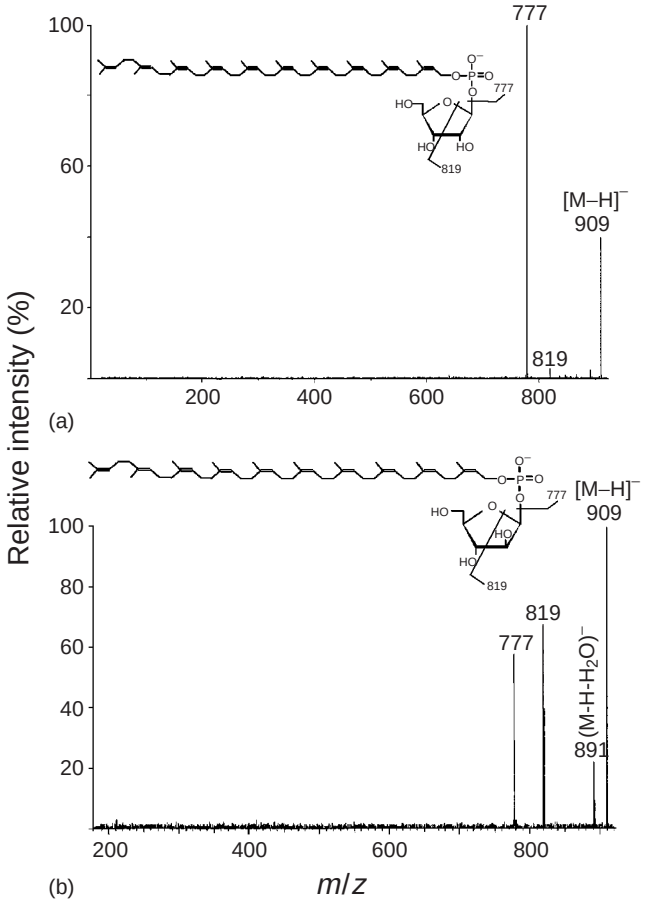

Figure 24 FAB CID at low collision energy of the $(M-H)^-$ anion from the decaprenylphosphates of β-D-ribose (a) and β-D-arabinose (b), respectively. The *trans* configuration of the 2-OH and phosphate groups in β-D-ribose favors the formation of the phospholipid fragment. (Reproduced modified from Wolucka and de Hoffmann,[19] with permission.)

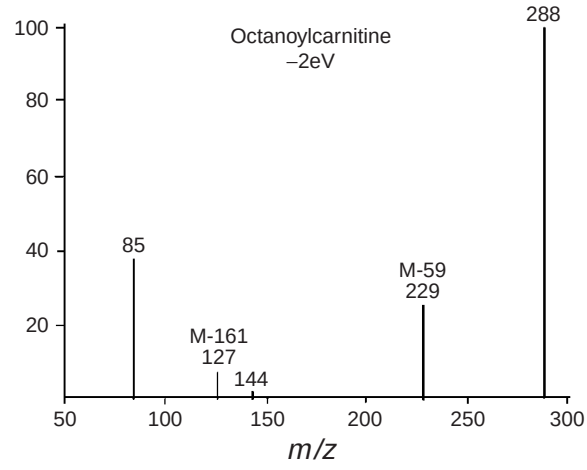

Figure 25 Product ions spectrum of octanoylcarnitine. For the nature of the fragments, see the fragmentation scheme in Figure 26.

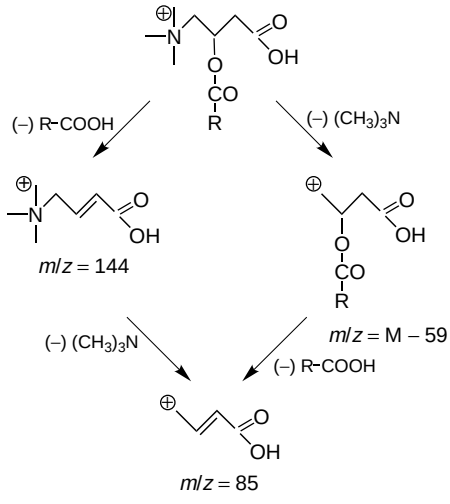

Figure 26 Fragmentation scheme of acylcarnitines, as deduced from the fragment spectra of octanoylcarnitine and confirmed by the spectra of other carnitine conjugates.

precursor ion (*m*/*z* 288) contains only the main isotopes ^{12}C, ^{1}H, ^{14}N and ^{16}O. At this low precursor kinetic energy (2 eV), few fragments are observed. From the interpretation of this spectrum, useful scan modes that detect carnitine conjugates selectively from different fatty acids can be found. This is important because the presence of carnitine conjugates in body fluids is diagnostic of several metabolic diseases and carnitine itself is used as a drug in some of these diseases.

Referring to the scheme in Figure 26, it can be seen that precursor scans of *m*/*z* either 144 or 85 are good candidates. Indeed, these fragments do not contain parts of the fatty acids and should thus not be sensitive to their nature. The fragment ions at *m*/*z* 229 (M-59) and 127 (M-161), acylium ion obtained after loss of the carnitine moiety, contain the fatty acid chain and thus are not good candidates for precursor ion scans that would detect all the carnitine conjugates. However, the neutrals lost in these fragmentations contain a part (59 Da) or the complete (161 Da) carnitine moiety. A neutral loss scan of either 59 Da or 161 Da should thus also allow the selective detection of all the carnitine conjugates.

Figure 27 displays the spectrum of both the FAB/MS spectrum of a biological sample containing carnitines and the precursor scan of *m*/*z* 85. If it is followed by a fragment ion scan of each detected conjugate, individual spectra for each compound are obtained. This example illustrates a general methodology for the selective detection of compounds or compound classes.

This can be combined with a chromatographic separation, allowing selective detection of one compound

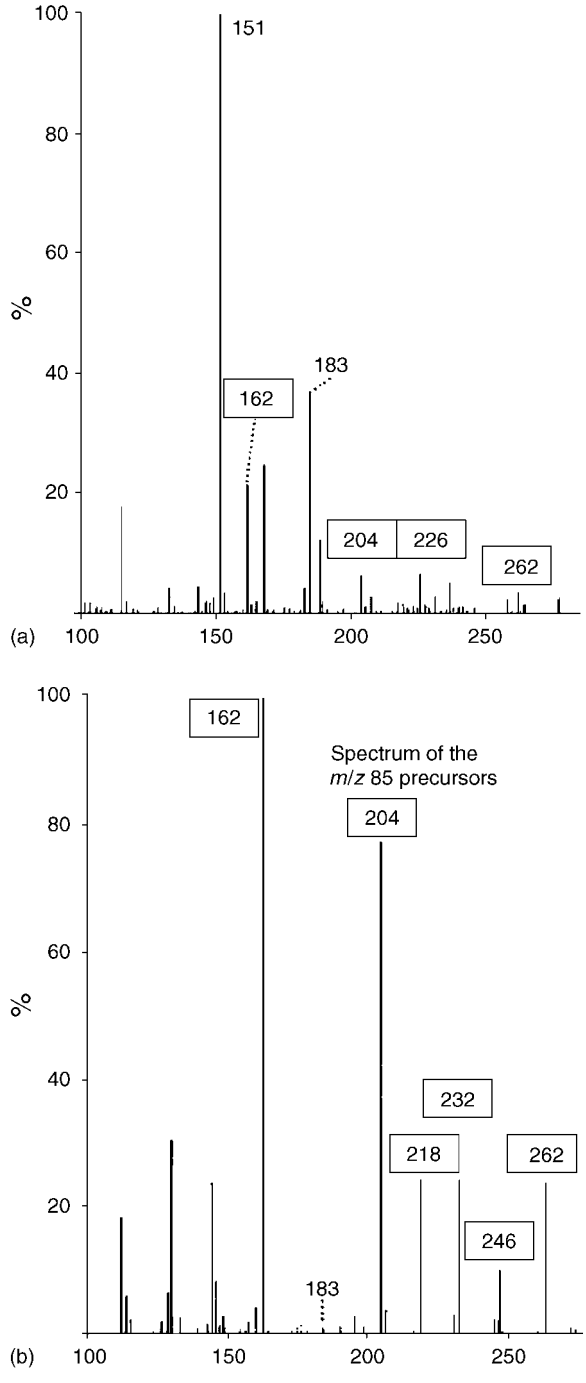

Figure 27 (a) FAB mass spectrum of a biological sample of acylcarnitines. (b) Same sample, but precursor scan of *m/z* 85. *m/z* 162: free carnitine, 204: acylcarnitine, 218 propionylcarnitine, etc. A comparison of these two spectra shows that the selectivity of the MS/MS scan strongly improves the signal-to-noise ratio (S/N). (Reproduced from de Hoffmann et al.[6])

in a complex mixture. This is largely used in the pharmaceutical industry to monitor and quantitate a selected compound in pharmacokinetic studies.

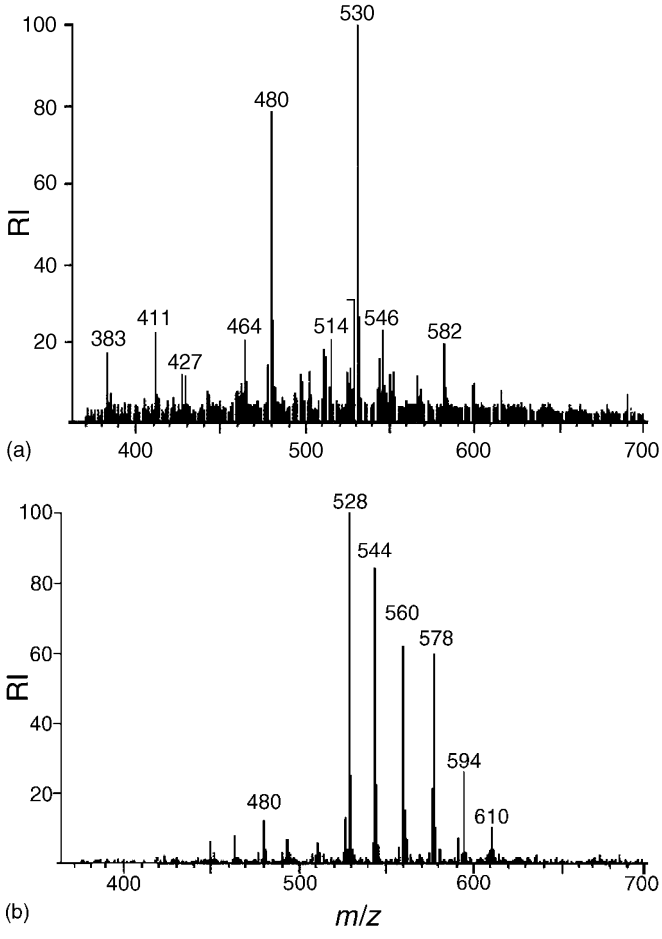

Figure 28 (a) FAB spectrum of a mixture of bile acids extracted from a physiological sample. (b) Selective detection of the sulfated bile acids by the neutral loss scan of 80 Da (SO_3). This spectrum displays compounds completely buried in the chemical background of the FAB spectrum. (Reproduced from Libert et al.[20])

7.4 Increase in Sensitivity by Tandem Mass Spectrometry

MS/MS also allows the S/N to be substantially improved in the detection of selected compounds or compound classes. This is illustrated by the spectra in Figure 28. The first spectrum is a FAB/MS spectrum (source spectrum) of a mixture of a physiological sample of bile acids.[20] Some of these bile acids are sulfated and for the diagnosis of biliary atresia it is important to detect them selectively. In the negative ion mode, sulfated bile acid loses SO_3, 80 Da, on CID fragmentation of the $[M-H]^-$ precursor anion. A scan of this neutral loss is displayed in the second spectrum. It shows that sulfated bile acids buried in the chemical background of the source spectrum are now clearly detected, such as those appearing at *m/z* 560, 578, 594 and 610. This is because selective detection does not apply to most of the components in the background, thus

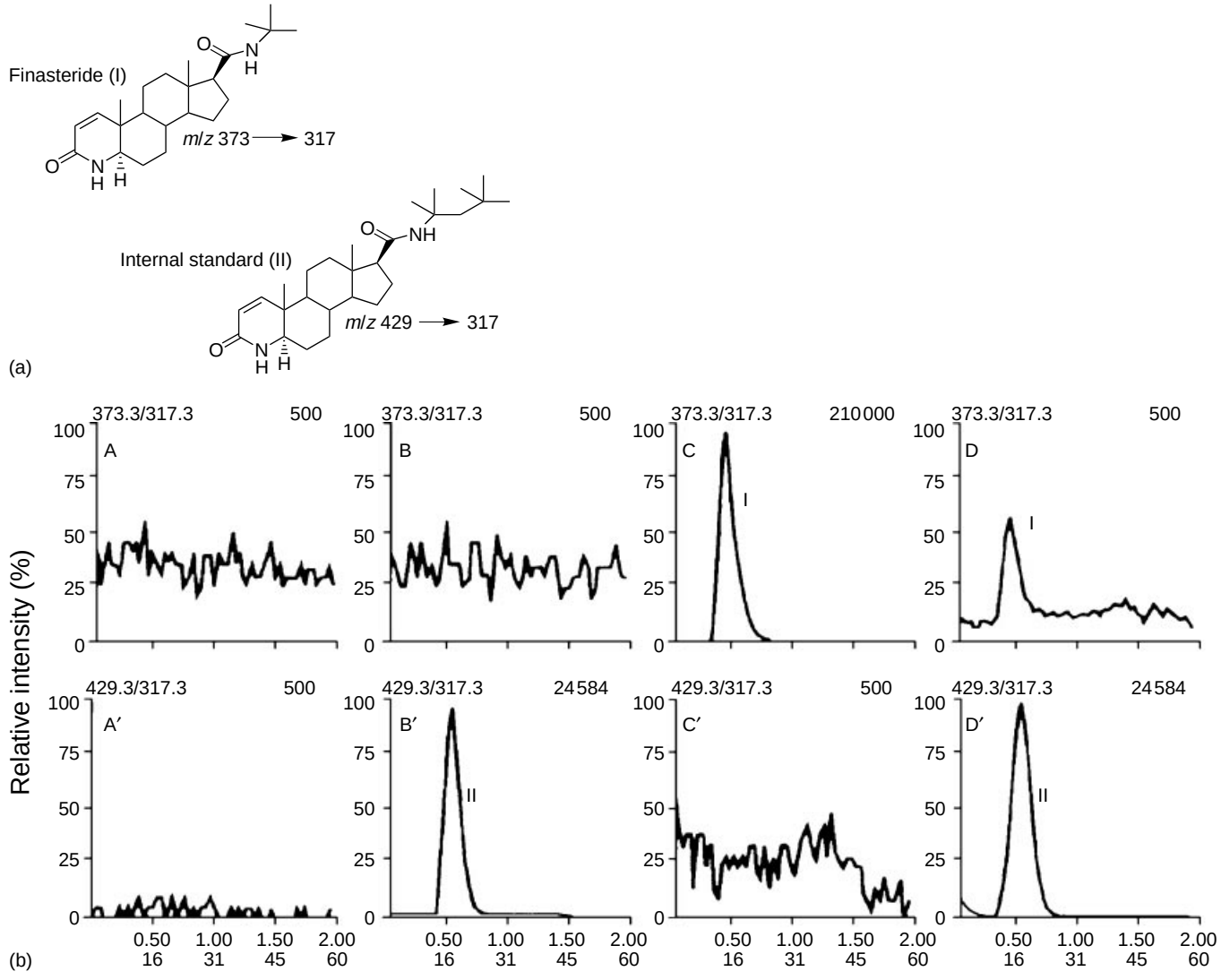

Figure 29 (a) Structure of finasteride and a derivative used as internal standard. The fragmentation used corresponds to the loss of the alkyl chain attached to the amide nitrogen, less a hydrogen atom. (b) The upper traces are relative to finasteride and the lower ones to the reference compound, monitored by the selected reaction indicated on the formulas. A, A′ : Both blank; B, B′ : blank and reference; C, C′ : control plasma spiked with 100 ng/mL and blank ; D, D′ : control plasma spiked with 0.05 ng mL^{-1} and 10 ng mL^{-1} of reference compound. (Reproduced from Matuszewski et al.[21] Reprinted with permission from B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng,[21] 'Matrix Effect in Quantitative LC/MS/MS Analyses of Biological Fluids: A Method for Determination of Finasteride in Human Plasma at Picogram Per Milliliter Concentrations', Copyright 1998, American Chemical Society.)

strongly reducing the background. As sensitivity depends on S/N, this example also shows that MS/MS can greatly improve the sensitivity towards selected compounds, especially when analyzing complex mixtures. It should be noted that this increase in sensitivity is observed even if actually more than 90% of the ions are lost in the fragmentation process.

The possibility of obtaining both high selectivity and high sensitivity by MS/MS means that it is often used in pharmacokinetics, which is the field that uses the largest number of tandem mass spectrometers. The example shown in Figure 29 illustrates that a good S/N can be obtained from a plasma sample spiked with 0.05 ng mL^{-1} (i.e. 50 ppb) of the drug, with a total chromatography time of only 2 min and elution after 0.5 min.

8 FIELDS OF APPLICATION OF TANDEM MASS SPECTROMETRY

Modern mass spectrometers can ionize and assign the m/z for a large variety of compounds, including

nonvolatile ones, such as proteins, synthetic polymers, oligosaccharides, and so on up to more than 100 000 Da. Easy gas chromatographic coupling is now extended to HPLC (high-performance liquid chromatography) and other separation techniques such as capillary electrophoresis.

MS/MS increases the capacity of the technique by allowing more efficient structure elucidation, selective detection, increase in sensitivity and faster mixture analysis. This makes MS/MS useful and it is now largely applied in almost all areas of organic and biochemical compounds analysis. Some areas, however, use more MS/MS than others. Pharmaceutical applications are the main domain of application of MS/MS, judged by the number of instruments used.

The development of combinatorial chemistry methods has created the need for faster structure determination or verification and selective compound class determination.

The parallel development of faster and more reliable in vitro tests to check metabolization has led to the tendency to check the rate of metabolization of new compounds earlier in the development process of a new drug. Here again, MS/MS is very useful. There are a few main metabolization routes leading to important expected phase II metabolites: glucuronides, sulfates, or glutathione conjugates. These can be easily detected by compound class-selective MS/MS methods. If a compound is quickly metabolized and eliminated, it is most probably not worthwile to try to develop it as a drug. The US Food and Drug Administration has issued a recommendation to the pharmaceutical industry to perform such in vitro tests.[22]

Furthermore, before approval of a drug, pharmacokinetics on the metabolites from human volunteers belonging to a representative panel of various ethnic origins have to be performed. Here again, MS/MS, coupled to HPLC, is the method of choice, as it allows fast sensitive analysis of target compounds.

Toxicology and clinical analysis also need highly selective and sensitive detection of target compounds, and MS/MS is developing quickly in these areas.

Proteomics has emerged as an important new application area in 1997 and 1998. More is known about the genomes by the sequencing of nucleic acids, and for some species it has been completely determined. Once the DNA sequence is known, the sequences of the corresponding proteins can be deduced. There is a great interest now in determining where those proteins, if expressed, are located and what is their role in the living cell. Because many proteins often act together in multistep synthesis, it is important to be able to determine which proteins are located in a delimited part of the cell. MS allows the molecular masses of proteins to be obtained quickly and MS/MS yields information about the sequence. As molecular masses combined with short sequences allow the identification of proteins in a database, established from the DNA sequence, MS/MS is again the method of choice for assigning the part of the genome from which a group of proteins originate.

Environmental pollution control uses MS widely, still mainly gas chromatography coupled to MS. HPLC/MS is becoming more frequently applied for nonvolatile compounds. MS/MS can be applied with cost affordable instrumentation and because it allows faster, more selective and sensitive detection, it is expanding fast in this field too.

9 CONCLUSIONS

MS/MS has many outstanding advantages as an analytical technique: high sensitivity into the femtomol range, rich structural information, high selectivity for the analysis of target compounds. As MS/MS can be applied to volatile or nonvolatile compounds, even with high molecular weights and can be coupled to separation techniques such as gas or liquid chromatography, its areas of application are very broad.

These advantages are increasing continuously with the development of instruments with still better performances and cheaper instrumentation. The future of MS/MS thus appears to be very bright.

ACKNOWLEDGMENTS

The author wishes to thank Alexandre Spôte for his constant and competent assistance, as well as all those who have contributed through their research work to the scientific content of this article.

ABBREVIATIONS AND ACRONYMS

AC	Alternating Current
CID	Collision-induced Dissociation
DC	Direct Current
DCI	Desorption Chemical Ionization
ESI	Electrospray Ionization
FAB	Fast Atom Bombardment
FTICR	Fourier Transform Ion Cyclotron Resonance
HPLC	High-performance Liquid Chromatography
ICR	Ion Cyclotron Resonance
MIKE	Mass Analyzed Ion Kinetic Energy
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass-to-charge Ratio

PSD	Post-source Decay
RF	Radio Frequency
S/N	Signal-to-noise Ratio
SORI	Sustained Off-resonance Irradiation
SRM	Selected Reaction Monitoring
SWIFT	Selected Waveform Inverse Fourier Transform
TOF	Time-of-flight

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry • Atmospheric Pressure Ionization Mass Spectrometry • Chemical Ionization Mass Spectrometry: Theory and Applications • Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry • Liquid Chromatography/Mass Spectrometry • Literature of Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Time-of-flight Mass Spectrometry

REFERENCES

1. A.G. Harrison, M.S. Lin, 'Energy Dependance of the Fragmentation of the *n*-butylbenzene Molecular Ion', *Int. J. Mass Spectrom. Ion Phys.*, **51**, 353–356 (1983).
2. S. Nacson, A.G. Harrison, 'Energy Transfer in Collisional Activation. Energy Dependance of the Fragmentation of *n*-alkylbenzene Molecular Ions', *Int. J. Mass Spectrom. Ion Processes*, **63**, 325–337 (1985).
3. J.C. Schwartz, A.P. Wade, C.G. Enke, R.G. Cooks, 'Systematic Delineation of Scan Modes in Multidimensional Mass Spectrometry', *Anal. Chem.*, **62**, 1809–1818 (1990).
4. E. de Hoffmann, 'Tandem Mass Spectrometry: a Primer', *J. Mass Spectrom.*, **31**, 129–137 (1996).
5. M.N. Eberlin, 'Triple-Stage Pentaquadrupole (QqQqQ) Mass Spectrometry and Ion/Molecule Reactions', *Mass Spectrom. Rev.*, **16**, 113–144 (1997).
6. E. de Hoffmann, J. Charette, V. Stroobant, *Mass Spectrometry: Principles and Applications*, John Wiley & Sons, Chichester, 1996.
7. K.G. Standing, in *The Analysis of Peptides and Proteins by Mass Spectrometry, Proceedings of the Fourth Texas Symposium on Mass Spectrometry*, ed. C.J. McNeal, John Wiley & Sons, New York, 267–278, 1988.
8. Dr P.C. Christiansen, Communication, Finnigan MAT User's Meeting, Chicago, May 1994.
9. T.J. Cornish, R.J. Cotter, 'Tandem Time-of-Flight Mass Spectrometer', *Anal. Chem.*, **65**, 1043–1047 (1993).
10. T.J. Cornish, R.J. Cotter, 'A Curved-field Reflectron for Improved Energy Focusing of Product-ions in Time-of-Flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 1037–1040 (1993).
11. R.J. Cotter, *Time of Flight Mass Spectrometry: Instrumentation and Applications in Biological Research*, American Chemical Society, Washington, DC, 1997.
12. R.E. March, R.J. Hughes, *Quadrupole Storage Mass Spectrometry*, Chemical Analysis, Wiley Interscience, Vol. 102, 1989.
13. J.W. Gauthier, T.R. Trautman, D.B. Jacobson, 'Sustained Off-Resonance Irradiation for Collision-activated Dissociation involving Fourier Transform Mass Spectrometry. Collision-activated Dissociation Technique that Emulates Infrared Multiphoton Dissociation', *Anal. Chim. Acta*, **246**, 211–225 (1991).
14. I.J. Amster, 'Fourier Transform Mass Spectrometry', *J. Mass Spectrom.*, **31**, 1325–1337 (1996).
15. J. Andersen, H. Molina, E. Moertz, T.N. Krogh, I. Chernuchevich, L. Taylor, O. Vorm, M. Mann, 'Quadrupole-TOF Hybrid Mass Spectrometers Bring Improvements to Protein Identification and MS/MS Analysis of Intact Proteins', The 46th Conference on Mass Spectrometry and Allied Topics, Orlando, FL, 978, 1998.
16. P. Mergaert, W. D'Haeze, D. Geelen, D. Promé, M. Van Montagu, R. Geremia, J.C. Promé, M. Holsters, 'Biosynthesis of *Azorhizobium caulinodans* Nod Factors', *J. Biol. Chem.*, **49**, 29217–29223 (1995).
17. V. Stroobant, R. Rozenberg, E.M. Bouabsa, E. Deffense, E. de Hoffmann, 'Fragmentation of Conjugate Bases of Esters Derived from Multifunctional Alcohols Including Triacylglycerols', *J. Am. Soc. Mass Spectrom.*, **6**, 498–506 (1995).

18. B. Wolucka, M.R. McNeil, E. de Hoffmann, T. Chojnacki, P.J. Brennan, 'Recognition of the Lipid Intermediate for Arabinogalactan/Arabinomannan Biosynthesis and Its Relation to the Mode of Action of Ethambutol on Mycobacteria', *J. Biol. Chem.*, **269**, 23328–23335 (1994).
19. B. Wolucka, E. de Hoffmann, 'The Presence of β-D-Ribosyl-1-monophosphodecaprenol in Mycobacteria', *J. Biol. Chem.*, **270**, 20151–20155 (1995).
20. R. Libert, D. Hermans, J.P. Draye, F. Van Hoof, E. Sokal, E. de Hoffmann, 'Bile Acids and Conjugates Identified in Metabolic Disorders by Fast Atom Bombardment and Tandem Mass Spectrometry', *Clin. Chem.*, **37**, 2102–2110 (1991).
21. B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, 'Matrix Effect in Quantitative LC/MS/MS Analyses of Biological Fluids: A Method for Determination of Finasteride in Human Plasma at Picogram Per Milliliter Concentrations', *Anal. Chem.*, **70**, 882–889 (1998).
22. 'Guidance for Industry. Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies in Vitro'. Department of Health and Human Services, US Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Rockville, MD, April 1997.

Time-of-flight Mass Spectrometry

Scot R. Weinberger
Ciphergen Biosystems, Inc., Palo Alto, USA

Stephen Davis and Alexander Makarov
H D Technologies, Manchester, UK

Steve Thompson
Scientific Analysis Instruments Ltd., Manchester, UK

Randy Purves
59-6247 Sundown Crescent, Gloucester, Canada

Randy M. Whittal
University of Alberta, Edmonton, Canada

1 Introduction **2**

2 A Historical Perspective of Time-of-flight Mass Spectrometry **4**

3 Time-of-flight Mass Spectrometry Theory and Fundamental Operation **8**

4 Time-of-flight Mass Spectrometry Interface Configurations, Ion Extraction Geometries, and Operational Schemes **10**
4.1 Interface Configurations 10
4.2 Ion Extraction Geometries: Parallel Ion Extraction 10
4.3 Orthogonal Ion Extraction 12

5 Factors Affecting Performance in Time-of-flight Mass Spectrometry **13**
5.1 Mass Accuracy 13
5.2 Resolution 15

6 Ion Focusing Means to Improve Resolution **18**
6.1 Time-lag Focusing 18
6.2 Post-source Pulse Focusing 20
6.3 Dynamic Pulse Focusing 20
6.4 The Ion Mirror or Reflectron 21
6.5 Higher Order Correction 22

7 Analytical Sensitivity **23**
7.1 Factors Affecting Analytical Sensitivity 24
7.2 Detectors for Time-of-flight Mass Spectrometry 24
7.3 Electron Work Function, Post-acceleration, and Secondary Ion Generation 25
7.4 Direct Measurement/Charge-based Detectors 27

8 Liquid Interfaces for Time-of-flight Mass Spectrometers **28**
8.1 Electrospray Ionization 28
8.2 Noncovalent Biological Complexes 29
8.3 High-performance Liquid Chromatography/Electrospray Ionization/Time-of-flight Mass Spectrometry 29
8.4 Capillary Electrophoresis/Capillary Electrochromatography Electrospray Ionization/Time-of-flight Mass Spectrometry 30
8.5 Matrix-assisted Laser Desorption-ionization 31

9 Time-of-flight Analyzers with Gas-phase Ion Sources **32**
9.1 Interfacing Continuous Ion Sources with the Time-of-flight Analyzer 32
9.2 Aspects of Gas Chromatography/Time-of-flight Mass Spectrometry 35

10 Inductively Coupled Plasma/Time-of-flight Mass Spectrometry **36**

11 Quadrupole Ion Trap/Time-of-flight Mass Spectrometry **37**

12 Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry **39**
12.1 Matrix-assisted Laser Desorption/Ionization Fundamentals 39
12.2 Matrix-assisted Laser Desorption/Ionization Applications 41

13 Surface-enhanced Laser Desorption/Ionization **43**
13.1 Surface-enhanced Laser Desorption/Ionization Versus Laser Desorption/Ionization and Matrix-assisted Laser Desorption/Ionization 43
13.2 Surface-enhanced Laser Desorption/Ionization Applications 45

14 Tandem Time-of-flight Mass Spectrometry **47**
14.1 In-source Decay 47
14.2 Post-source Decay 47
14.3 Collision-induced Dissociation/Tandem Mass Spectrometry Techniques 49

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

15 Time-of-flight Secondary Ion Mass Spectrometry **50**
15.1 Concept of a Maximum Dose 50
15.2 Instrumentation 51
15.3 Charge Neutralization 53
16 Future Advancements and Concluding Remarks **54**
Dedication **54**
Abbreviations and Acronyms **55**
Related Articles **56**
References **56**

This article reviews the developmental history, fundamental technology, and general applications of time-of-flight mass spectrometry (TOFMS). The fundamentals of instrument operation and components are discussed. Applications in terms of gas chromatography (GC), liquid chromatography (LC), capillary electrophoresis (CE), laser desorption, multiphoton desorption, plasma desorption (PD), matrix-assisted laser desorption, surface-enhanced laser desorption, inductively coupled plasma (ICP), and secondary ion mass spectrometry (SIMS) are presented. A review of tandem time-of-flight (TOF) techniques with a focus on ion trap TOF instrumentation is provided.

The purpose of this article is to provide the reader with a fundamental understanding of TOFMS principles and applications. Information is presented in a simple, straightforward, didactic fashion.

1 INTRODUCTION

During the past decade, the technology of TOFMS has demonstrated impressive development and analytical popularity. As is the case with all mass spectrometers, a TOFMS system consists of two major components: an ionization source (roughly the sample introduction means: also known as the interface or inlet system) and the mass analyzer (the mass measurement and ion detection portion). Present-day TOF mass analyzers are routinely interfaced with solid-state, gaseous, and liquid samples. Currently, there are some 15 different manufacturers of commercial TOFMS analyzers for use in surface investigation, life science studies, elemental analysis, as well as inorganic and organic chemical analysis (see Table 1).

Conceived in 1946,[1] the fundamental strengths of this mass analyzer when compared to other contemporary devices drove an initial phase of growth and acceptance during the late 1950s and early 1960s that resulted in a 1964 worldwide base of more than 200 units.[2] Indeed, these strengths were readily recognized by many of the TOF pioneers.[1–3] To begin with, the TOF analyzer is remarkably simple in its concept, operation, and design. Compared with quadrupole radiofrequency (RF)-mass filter, magnetic sector, and ion trap mass analyzers, the moderate mechanical alignment and machining requirements for the TOF analyzer mitigated the need for extreme alignment and fabrication tolerances, greatly simplifying the manufacturing and installation processes while concomitantly improving product reliability. Such inherent simplicity has resulted in the creation of several simple-to-use, inexpensive, small devices. Undoubtedly, this trend toward miniaturization and improved simplicity will continue in the years to come.

In terms of analytical performance, the TOF analyzer excels in a number of capabilities. Because scanning mass analysis is performed without the use of time-dependent changing magnetic or electric fields, the TOF analyzer demonstrates an extraordinarily fast scan speed and associated high scan-mode duty cycle. Mass analysis is generally accomplished by employing uniform, static electric fields and, as such, system response time is solely dependent upon ion introduction time, ion extraction rate, and the velocity of the heaviest ion of interest. Ions may be formed within the extraction region or in an outside area requiring about 1–2 ms to enter the extraction region. Modern TOF devices intended for use in continuous-flow gas- or liquid-phase analysis are capable of extracting ions at a rate of 10^4–10^5 cycles per second. Furthermore, the flight time for an ion of about 50 000 m/z is roughly 260 µs (15 keV energy, 2.0 m free flight). Accordingly, the duty cycle becomes ultimately limited by the TOF of the heaviest ions. The latter results in nominal duty cycles of 5–50%.[4] This high scan rate and associated duty cycle makes the TOF device the molecular weight (MW) detector of choice for many fast chromatographic techniques such as capillary GC, CE, micro and capillary LC, as well as capillary electrochromatography (CEC).

Another analytical strength of TOFMS is its mass range. Because mass assignment is solely dependent upon ion flight time, the mass range of the TOF instrument is theoretically unlimited. However, practical limitations of ionization and detection efficiency do play governing roles. Even so, singly charged organic polymers approaching 1.0 mDa in mass have been detected using MALDI.[5] Furthermore, the same technique has successfully detected singly charged ions of human immunoglobulin M (IgM) at ~980 kDa,[6] whereas TOFMS with ESI has been employed to detect multiply charged ions of single-stranded and double-stranded DNA in the 1.0–5.0 mDa range.[7–9]

Table 1 TOF instrument manufacturers in 1999

Company	TOF products	Location/URL
Analytica of Branford	API-TOFMS	Branford, CT, USA www.aob.com
Bergmann Messgerate Entwicklung KG	ESI, cluster, laser photon ionization-TOF	Murnau, Germany www.bme-bergmann.de
Bruker Daltonics	ESI, LDI, MALDI/TOFMS	Bremen, Germany www.bruker.de
Comstock	LDI, MALDI, EI/TOFMS	Oak Ridge, TN, USA www.comstockinc.com
Ciphergen Biosystems	LDI, MALDI, SELDI/TOF	Palo Alto, CA, USA www.ciphergen.com
GSG Analytical	LDI, MALDI/TOF	Karlsruhe, Germany www.gsg-analytical.com
HD Technologies	EI, HSI/TOF, accessories	Manchester, UK www.hdtech.demon.co.uk
Kore Technologies	TOFSIMS, accessories	Cambridge, UK www.kore.co.uk
Kratos/Shimadzu	LDI, MALDI/TOF	Manchester, UK www.kratos.co.uk
Leco	ICP/TOF, ESI/TOF	St. Joseph, MI, USA www.leco.com
Micromass	LDI, MALDI, ESI/TOF	Manchester, UK www.micromass.co.uk
PE Biosystems	LDI, MALDI, ESI/TOF	Foster City, CA, USA www.pebio.com
Physical Electronics	TOFSIMS	Eden Prairie, MN, USA www.phi.com
Scientific Analysis Instruments	LDI, MALDI/TOF	Manchester, UK www.saiman.demon.co.uk
Thermoanalysis	LDI, MALDI/TOF	Santa Fe, NM, USA www.maldi.com

EI, electron ionization; ESI, electrospray ionization; ESI/TOF, electrospray ionization/time-of-flight; MALDI/TOF, matrix-assisted laser desorption/ionization/time-of-flight; TOFSIMS, time-of-flight secondary ion mass spectrometry; LDI, laser desorption ionization; MALDI matrix-assisted laser desorption/ionization.

The high ion transmission properties of the TOFMS instrument, particularly linear TOF devices, make it perhaps the most sensitive scanning mass spectrometer. In a linear TOFMS instrument, overall ion transmission efficiency is primarily regulated by ion transverse velocity spread.[10] Other losses occur when ions collide with grid elements of ion optical components. Such effects are typically controlled by employing gridless, radial ion focusing means,[11] high ion accelerating voltages, and reasonably sized ion detectors. In this manner most of the ions formed are efficiently collected, accelerated, and transmitted down to a detection apparatus. Because no slits or energy-selective filters are employed, minimal ion loss due to unimolecular decay occurs, and products of unimolecular decay can be detected with high efficiency at the same time as their parent ions. This makes this device ideal for the analysis of ion species with very short lifetimes. As will be demonstrated later, subfemtomole levels of detection for even large, labile molecules such as bovine immunoglobulin G (IgG, ~148 kDa) is possible with reasonable signal-to-noise levels.

Without doubt, the coupling of the TOFMS analyzer with the soft ionization techniques of MALDI[12–15] and ESI[16] has enabled facile mass spectral analysis of complex biological samples and has been a primary driving force behind the recent growth and popularity of this approach. A good portion of this article will be dedicated to describing these techniques. However, equally important to this rapid adoption curve are recent technological advances in laser design; high-voltage electronics; high-speed digital and broad-bandwidth analog electronics; as well as high-speed, high-voltage pulsing technology. Because of the latter, MALDI and ESI TOFMS analyzers are now capable of providing low parts-per-million mass accuracy and a mass resolving power ($m/\mathrm{d}m$) in excess of 20 000, greatly enabling empirical studies and the ability to identify and characterize unknown biopolymers.

This article provides a brief historical review of TOFMS and then proceeds to describe the fundamental theory and components of TOF mass spectrometers. A number of different TOF geometries and ion interfaces are

discussed, along with exemplary applications. The goal is to provide new and future users of this technique with sufficient knowledge to allow for intelligent decisions regarding the appropriate use, limitations, and possible selection of TOF instrumentation.

Because the scope and required expertise to attain successfully the goals of this article are quite demanding, this work is the product of the combined efforts of a number of expert contributors. The section addressing gas chromatography/time-of-flight mass spectrometry (GC/TOFMS) is authored by Stephen Davis; the section covering inductively coupled plasma/time-of-flight mass spectrometry (ICP/TOFMS) is authored by Alexander Makarov; the section addressing liquid chromatography/time-of-flight mass spectrometry (LC/TOFMS) is authored by Randy Whittal; the section reviewing TOFSIMS is authored by Steven Thompson; the section describing QIT/TOFMS (quadrupole ion trap) is written by Randy Purvis; and all other sections were provided by Scot R. Weinberger.

2 A HISTORICAL PERSPECTIVE OF TIME-OF-FLIGHT MASS SPECTROMETRY

Stephens first suggested the concept of TOFMS in 1946.[1] In a brief correspondence, he envisaged a simplified mass spectrometer utilizing no magnets or slits, with a high duty cycle and a millisecond scan period, amenable for gas composition control, rapid gas analysis, as well as portable use. Although noting that such a device was being constructed, actual published data from this work was not seen until 1953.[17]

In 1948, Goudsmit described a virtual instrument that introduced and accelerated ions in a homogeneous magnetic field.[18] Under such conditions, he noted that these ions would assume helical trajectories with angular velocities independent of initial ion velocity and release direction. The latter minimizes the perturbing effects of varied initial ion position and energy upon mass resolving power and mass assignment accuracy. In this case, ion TOF was described to be directly proportional to ion mass and inversely proportional to magnetic field strength. This analyzer was ultimately constructed and described in 1951.[19] Although initially shown to be successful, this approach had a limited practical mass range, because heavier ions require increasingly larger helical path radii, mandating a large system footprint and excessive vacuum requirements. Consequently, this geometry has been effectively abandoned.

In 1948, Cameron and Eggers provided the first published data from a TOF mass spectrometer.[20] This device was a gas-phase TOF analyzer known as an ''Ion Velocitron''. The instrument was a constant-energy, linear TOF mass spectrometer in which gas-phase ions were created using an EI source and accelerated to constant energy by a three-element ion gun. After acceleration, ions traveled down a 317-cm long drift tube to strike ultimately a charge collector plate that functioned as an ion detector.

A constant-energy TOF device functions by accelerating all ions to the same level of kinetic energy. In this manner, ions will assume velocities as defined by Equation (1), where v is final ion velocity, e is ion charge, V is acceleration voltage, and m is ion mass:

$$v = \left(\frac{2eV}{m}\right)^{1/2} \tag{1}$$

Ion TOF is the product of free flight distance and $1/v$. Resolution for this approach is defined by Equation (2) where R is resolution (a unitless entity), T is the ion flight time and Δt is the peak width at half-height of the measured ion signal:

$$R = \frac{T}{2\Delta t} \tag{2}$$

Demonstrated resolution for the analysis of analytes ranging from 1 (H^+) to 200 (Hg^+) Da by the ion velocitron was of the order of 2–3. The authors accurately attributed the limited resolving power to nonmonoenergetic ion populations and ion multipath trajectories. Although providing limited mass resolving power, mass accuracy, and MW dynamic range, the device did incorporate a number of advanced technological approaches, such as differential vacuum pumping, a resistor high-voltage dividing network to establish ion optic potentials, a pulsed electrostatic ion deflector used for ion bunching, and a post-source focusing lens assembly used for beam collimating and steering.

In 1953, Wolff and Stephens published results from the instrument first described in 1946.[17] This device was the first to employ pulsed ion extraction (PIE). Furthermore, a 0.5-μs-wide pulsed EI source was used to generate ions within a multiple-lens ion optic assembly. For constant-energy TOF operation, the ion extraction pulse was applied for a period of time sufficiently long to allow complete extraction of all ions of interest. In another mode of operation – constant momentum or impulse TOF analysis – the ion extraction pulse was applied and then shut off prior to the time required for any ions to leave the extraction region. In this fashion, ions were accelerated to final velocities that were inversely proportional to their mass. Hence, all ions had constant momentum.

Final ion velocity in impulse TOFMS is described by Equation (3), in which E is the applied ion acceleration

field and T_p is the duration of the ion extraction pulse:

$$v = \frac{ET_p e}{m} \quad (3)$$

It can be shown that resolution in an impulse TOF analyzer is defined by Equation (4):

$$R = \frac{T}{\Delta t} \quad (4)$$

A quick comparison of Equations (2) and (4) would lead one to believe that improved mass resolving power would be obtained by employing the impulse TOF approach in lieu of the constant energy approach, should a constant degree of ion time spread be realized in each system. Indeed, Wolff and Stephens noted the apparent resolution advantage of impulse TOF and they spoke of improved resolution for heavier masses using this technique. However, they did realize that thermal velocity distributions creating nonmonoenergetic ion populations would also limit resolving power. Katzenstein and Friedland further substantiated this in 1955.[21] Mathematically, they demonstrated that thermal spread greatly minimized the resolving power of constant-impulse TOF measurements and that, under these conditions, constant-energy TOF was favored. They constructed a pulsed-ion-extraction, constant-energy TOF device that demonstrated a fivefold improvement in resolving power when compared to the constant-impulse device of Wolff and Stephens.

Later in 1955, Wiley and McLaren introduced a new dual-stage ion gun providing improved ion focusing.[3] In the same publication, they also introduced the landmark PIE technique of time-lag focusing (TLF), capable of correcting for initial ion spatial or initial ion energy spread, and providing previously unattainable mass resolving power. This technique is described at length in section 6.1. In their device, the primary limit to resolution was determined to be the spatial distribution of ions formed by their EI source. Employing TLF, unit mass resolution for ions as large as 300 Da was achieved for the first time. All modern TOFMS systems utilize some form of TLF during the ion extraction duty cycle.

Bendix Corporation (Detroit, MI) ultimately commercialized the TLF technology of Wiley and McLaren, as described by Wiley in 1956.[22] The Bendix TOF device had extended the resolvable mass range out to about 600 Da and demonstrated about 1 ppm sensitivity for gaseous analytes in ''the low mass range''. The system incorporated a magnetic electron multiplier detector that provided of the order of 10^8 gain.[2] This technology was ultimately sold to CVC Products (Rochester, NY).[23] Overall, the Bendix TOF system proved to be highly successful and was modified by a number of researchers, adapting it to a plurality of ion sources and a wide range of applications (see Table 2).

During the 1960s, TOF devices were also interfaced to a number of light-based ionization sources. In 1964, Isenor interfaced a TOF device with a Q-switched ruby laser ($\lambda = 694$ nm) to evaluate the effect of laser fluence upon the initial kinetic energy of ions desorbed from Zn, Pb, and Mg targets.[29] Lincoln (1965) used a high-intensity, xenon flash tube to desorb and ionize cellulose, inorganic compounds, and ions from meteor fragments prior to analysis in a Bendix Model 14-107 TOFMS system.[30] In 1966, Fenner and Daly used a short-pulse-duration Q-switched ruby laser to generate ions from thin metal foils prior to analysis in a 1.0-m TOFMS system.[31] Vastola and Pirone (1968) interfaced a frequency-doubled ruby laser ($\lambda = 347$ nm) to a Bendix 12-107 TOFMS system to analyze various polyaromatic hydrocarbons prepared in a pyrite matrix.[32]

In the mid-1970s, a significant advancement in the technology of laser-based ionization TOF instruments occurred in the creation of the laser microprobe mass

Table 2 Various sample interfaces developed for the Bendix TOF mass spectrometer

Interface	Application	Originator
Capillary gas inlet system	Pulmonary gas analysis and space cabin atmosphere analysis	Bendix Corporation[2]
Capillary GC	Analysis of gas chromatographic bands with <1-s peak widths	McFadden[2]
Flash pyrolysis interface	Analysis of poorly volatile liquids and solids	Friedman[24]
Heated molecular leak gas inlet	Gas analysis at multiple leak rates	Bendix Corporation[2]
Heated solid probe	Study of poorly volatile liquids and solids	Damoth[25]
Ion beam source	Gas-phase ion–molecule or neutral molecule interactions, and SIMS	Lehrle[26]
Ion molecule reaction chamber	Analysis of gas-phase ion–molecule reactions	Melton and Hamill[2]
Knudsen cell	Thermal ionization studies of poorly volatile liquids and solids up to 2500 °C	White[27]
Plasma furnace interface	Analysis of ions or neutrals from 11 000 K plasma	O'Halloran and Fluegge[28]
Residual gas analysis	Analysis of vacuum residuals in evacuated chambers	Bendix Corporation[2]
Van de Graaf accelerator	Study of ionizing millielectron-volt protons	Shuler and Stuber[2]

analyzer (LAMMA). Hillenkamp (1975) converted a laser microprobe microscope device into a laser desorption/ionization TOFMS ion source.[33] The system utilized a frequency-doubled, Q-switched ruby laser to desorb successfully and ionize a number of epoxy-doped metals. In 1978, Hillenkamp linked a frequency-doubled ruby laser, a frequency-quadrupled neodymium laser ($\lambda = 265\,\text{nm}$), and various nitrogen lasers ($\lambda = 337\,\text{nm}$) to a linear TOF system and analyzed alkali metal and divalent cation profiles within thin tissue slices of frog skeletal muscle.[34] Hillenkamp's efforts later became commercialized by Leybold–Heraeus (Koln, Germany) as the LAMMA 500,[35] a simple linear TOF analyzer, and as the LAMMA 1000, which utilized a reflectron TOF mass spectrometer.[36]

Significant growth in the use of TOFMS analyzers occurred shortly after the introduction of a number of soft ionization techniques that enabled the analysis of poorly volatile analytes, particularly biopolymers. Prior to the introduction of such techniques, molecules such as enzymes, amino acids, proteins, nucleic acids, oligonucleotides, oligosaccharides, and glycoconjugates remained refractory to mass spectral analysis when employing electron impact, thermal and flash pyrolysis, as well as early photoinduced ionization means. Biopolymers exhibit extensive intermolecular attractive forces in aqueous and solid states and, as such, possess extremely low vapor pressures. Under such circumstances, it is difficult to create gas-phase neutrals for subsequent ionization and mass spectrometry (MS) analysis. In this case, the simplest method for increasing biomolecule vapor pressure is to heat the sample. However, biopolymers are known to be exquisitely heat labile and, as such, decompose when subjected to elevated temperatures over an extended period of time.

Initial attempts to perform biomolecular MS analysis involved the use of chemical derivatization. In this manner, polar groups of biological compounds are modified to nonpolar substituents. These nonpolar groups do not produce the strong intermolecular interactions of their predecessors, and the modified compound exhibits significantly elevated vapor pressure. Such approaches have successfully linked gas chromatographic separation of some temperature-stable biocompounds, such as small lipids and carbohydrates, to MS analysis.

To provide effective MS analysis of thermally labile biocompounds, volatilization should be achieved with a minimal amount of decomposition. In this fashion, parent ions of the nascent biomolecule may be studied in a mass spectrum that is relatively free of artifact fragment ions. The latter allows for the analysis of complex biomolecular mixtures, where multiple compounds are simultaneously present, as well as the study of possible biopolymer modifications, where subtle mass differences between populations exist due to post-translational modifications or other biomediated changes.

Parent ion fragmentation typically occurs through unimolecular decay mechanisms when the heat of volatilization excites the molecule to unstable vibrational states that result in bond schism. The rate of decomposition is primarily determined by the time needed to deposit energy into the unstable vibrational mode and by the vibrational period. Consequently, if the rate of surface vaporization exceeds that of unimolecular decay, most desorbed ions should have sufficient time to cool, thus minimizing the yield of fragment products. However, volatilization is only part of the process. Ultimately, a volatized neutral molecule must be ionized for MS analysis.

The means by which volatilized neutral molecules are ionized also plays a role in unimolecular decay dynamics. With respect to biomolecules, ionization is typically achieved through the addition or deletion of a proton (H^+) and less frequently via the addition or deletion of an electron (e^-), or the addition of some metal ion. Varying degrees of ion fragmentation will result if the addition of the charging species creates excessive internal excitation. Such phenomena have been discussed in the case of chemical ionization (CI)[37] and MALDI.[38] In the final analysis, successful MS investigation of labile biopolymers must utilize a volatilization and ionization technique that does not facilitate excessive unimolecular decay. Such techniques have been loosely termed as being soft.

The first greatly successful soft ionization technique applied to the analysis of biopolymers was californium-252 PD, originated by Macfarlane et al.[39] In this approach, samples are prepared as solid-state crystals or films upon thin nickel foils. Desorption and ionization are achieved using the high-energy fission products of ^{252}Cf decay. Spontaneous fission of ^{252}Cf is asymmetric, producing two nuclei of uneven mass and energy, released in opposing directions. Typical products such as $^{142}Ba^{18+}$ and $^{106}Tc^{22+}$ emerge with respective kinetic energies of 79 and 104 meV. One of the pair of liberated fission products is directed to pass through the foil, whereas the other is directed toward a fragment detector, which is used to create the time zero mark for the TOF measurement. It was initially believed that the high-energy fragments that passed through the sample foil created a thermal spike that vaporized quantities of mobile impurity ions upon the foil surface (mostly H^+, Na^+, and H^-). These secondary ions would then react with the sample molecules in the region of the thermal spike to produce parent molecular ions. Later it was believed that the thermal spike created a desorbed molecular dimer, which in turn would split up into two parent molecular ions of opposite charge.[37]

Macfarlane et al.[39] mated this desorption/ionization source with a linear TOF mass spectrometer. Two different devices were noted. One device had an 8-m-long flight tube and was used for high-resolution experiments. The second device utilized a 1.5-m-long flight tube and was used for high-sensitivity measurements. PD TOF technology was ultimately commercialized by Bio-Ion AB (Uppsala, Sweden).

PD has been applied most successfully to the analysis of peptides and proteins.[37,40] Although identified as a soft ionization technique, evidence of unimolecular decay and prompt parent ion fragmentation was noted from the onset.[39] The potential for such prompt fragmentation was demonstrated to increase with increasing MW. The level of desorption energy deposited by ^{252}Cf fission products was eventually recognized as being too high for the analysis of large, labile proteins. The effective mass limit of this technique was found to be about 50 kDa.[41] Although widely employed from the late 1970s through to the 1980s, this technique was ultimately abandoned after the advent of MALDI and ESI.

In 1981, Chait and Standing interfaced a pulsed alkali metal ion gun to a simple linear TOF mass spectrometer known as Manitoba I.[42] The pulsed ion gun provided a desorption and ionization means for samples that were deposited as a thin layer of solid-state crystals or films upon a solid support target. When these pulses of primary ions struck the target, they liberated secondary ions. This ionization technique is known as SIMS and section 15 of this article is dedicated towards further discussion of this technique. The combination of SIMS with a TOF analyzer has been colloquially coined as TOFSIMS.

Chait and Standing reported the successful analysis of bio-ions with m/z up to 1355 Da. Later, they adapted a similar pulsed SIMS source to a reflectron TOF mass analyzer (Manitoba II).[10] This device demonstrated a mass resolving power $m/\mathrm{d}m$ of up to 10 000 and was used to detect successfully the parent molecular ions of bovine insulin: MW = 5733 Da. Using this device, they reported the detection of unimolecular decay products from ions generated during this process. Consequently, this ionization technique was not totally ''soft''.

TOFSIMS devices are used extensively today, but primarily for surface analysis. The use of TOFSIMS for biopolymer analysis has been greatly diminished since the advent of MALDI and ESI approaches.

In the late 1970s, another soft ionization technique known as multiphoton ionization (MPI) emerged. In this approach, samples are introduced as a gaseous molecular beam[43] or solid.[44,45] Direct ionization of the gaseous beam is accomplished while solid samples are typically desorbed using an infrared laser. An ultraviolet or visible laser is used to provide ionizing photons. Gas-phase neutrals absorb several photons, exciting them to an energy level equivalent to the total energy of the absorbed photons. In this manner, even visible light can lead to ultimate ionization. Initial photon absorption causes electronic transitions from the ground state (S_o) to an S_1 excitation state, and subsequent photon absorption results in further transition to an ionization phase where electron loss results in the creation of a gaseous cation. Should the wavelength of the ionizing laser match the absorbance maxima of the neutral's S_1 state, so-called resonance MPI occurs, allowing selective ionization of trace components within complex mixtures.[46]

When ionizing laser fluences of the order of 10^6 W cm^{-2} are employed, soft ionization is observed. However, fragmentation has been demonstrated for fluences of the order of about 10^8 W cm^{-2}. Such selective fragmentation has been used to provide structural information from detected parent ions. Tandem MS measurements have been demonstrated using a scanning reflectron approach similar to that employed during post-source decay (PSD) analysis of today's MALDI/TOF devices.[43]

Schlag et al. described a molecular beam MPI source coupled to a simple linear TOFMS device utilizing a 30-cm-long flight tube. This system provided unit mass resolution for compounds up to 260 m/z. Higher resolution was achieved for a reflectron device (3900 $m/\mathrm{d}m$ for benzene).[43] Zare described a linear TOF device for analyzing solid samples via MPI. Neutrals were desorbed using a pulsed CO_2 laser ($\lambda \sim 10.6\,\mu$m) and subsequently ionized via a frequency-quadrupled Nd : YAG ($\lambda = 265$ nm) laser. This device was used for quantitative analysis of phenylthiohydantoin-conjugated amino acids.[45] Zare's quantitative approach was ultimately patented (US Patent No. 4 988 879) and was shown to be successful in the analysis of as little as 5 fmol of protoporphyrin IX, 50 fmol of β-estradiol, and 200 fmol of adenine.[47] Lubman and Trembreull described a solid-phase MPI system that utilized a supersonic pulse of gas as an ion-cooling medium that greatly improved the mass resolving power of a linear multiphoton ionization/time-of-flight (MPI/TOF) system. Typical resolution of the order of about 250–300 for a number of amino acids, neuroleptic drug agents, water-soluble vitamins, and purine bases was demonstrated.[44]

Although MPI proved to be valuable in the study of many organic and small biological compounds, it was never applied successfully to routine studies of larger biopolymers. As such, plasma desorption mass spectrometry (PDMS) and ultimately the techniques of MALDI and ESI greatly restricted the use of this approach in the arena of life science research.

Undoubtedly, the soft ionization approaches of ESI and MALDI are responsible for the recent, exponential growth in the use of TOFMS technology. ESI, first described by Fenn et al.[48] was initially adapted to a

TOF device by Boyle and Whitehouse.[16] An improved means of ion cooling has enabled efficient coupling of these two techniques,[49] allowing for the availability of a number of commercial devices (see Table 1).

Two groups independently developed MALDI: the group of Tanaka and the group of Hillenkamp. In 1987, Tanaka reported the use of a colloidal matrix of cobalt powder, glycerol, ethanol, and acetone to analyze chicken egg-white lysozyme. Pseudomolecular ion clusters of 108 kDa $[(7M + H)^+]$ were detected successfully and measured with an accuracy of better than ±1% absolute error.[14,15] This was the first reported instance where ions in excess of about 50 kDa were detected. In 1988, Karas and Hillenkamp reported the first use of an organic matrix, nicotinic acid, for the detection of proteins in excess of 10 kDa.[12,13] Eventually, organic matrix MALDI was used to study proteins approaching 1 mDa in mass.[6] For the most part, present-day MALDI analyses employ organic matrices and the colloidal approach has been all but abandoned. MALDI has been successful not only for the analysis of proteins and peptides, but also in the analysis of small organics, large organic polymers, oligosaccharides, oligonucleotides, as well as lipid moieties. As with ESI/TOF, a number of commercial MALDI/TOF devices exist today (see Table 1). Because the impact of ESI and MALDI has greatly affected bioanalysis and TOFMS in general, this article dedicates separate sections (sections 8.1 and 12) to describe these two techniques.

3 TIME-OF-FLIGHT MASS SPECTROMETRY THEORY AND FUNDAMENTAL OPERATION

The TOF mass spectrometer employs uniform, static electric fields to accelerate ions to some final velocity v. Two fundamental approaches have been developed: constant-energy TOF[1] and constant-momentum or -impulse TOF.[20] The superiority of the constant-energy approach has eclipsed the impulse method, and all modern-day devices accelerate the ions formed to constant energy. Consequently, only theory regarding constant-energy TOF devices will be discussed.

Figure 1 depicts a single-stage ion optic, linear, constant-energy TOF mass spectrometer. It consists of two portions: an ion acceleration region containing an ion optic assembly; and an ion free flight region comprising an ion drift tube. Two different ions of charge z and mass M_1 and M_2, where the MW of M_1 is greater than that of M_2, are formed at the same time and at the same location toward the beginning of the ion optic assembly. This ion optic assembly consists of two elements: a repeller lens element and ground aperture. The repeller is raised to some potential V while the ground aperture is held at ground potential. If ions M_1 and M_2 are cations, the repeller plate is raised to some positive potential V. If ions M_1 and M_2 are anions, the repeller plate is set to some negative potential V. Only a single polarity of ions may be analyzed at any given time. In this fashion, an electric field is created between the repeller and ground aperture so that formed ions are accelerated along distance s, through the opening in the ground aperture, and out into the drift tube region of the mass spectrometer. The acceleration of ions M_1 and M_2 within the ion optic assembly and their final velocities can be described using fundamental equations of electrostatic theory, Newtonian mechanics, and rectilinear motion.

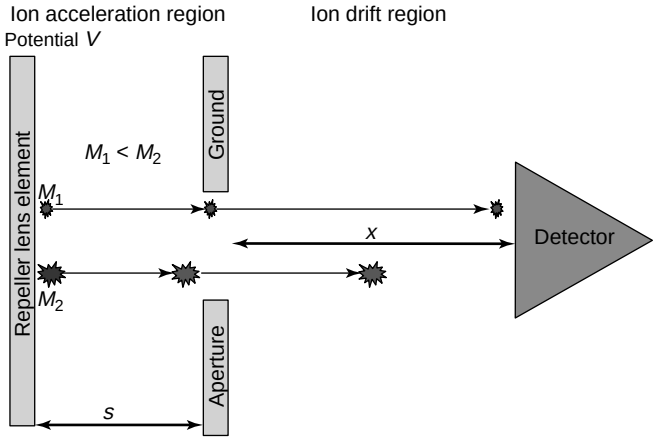

Figure 1 A single-stage ion optic, linear, constant-energy TOF mass spectrometer. Two ions M_1 and M_2, where M_2 is heavier than M_1, are accelerated to constant energy. Ion TOF is dependent upon m/z, acceleration potential V, acceleration distance s, and free flight distance x. See text for derivation.

Consider ion M_1 with charge z located above the repeller plate within the ion optic assembly. If the repeller plate is raised to a potential V, then the electric field strength (E) within the assembly is defined by Equation (5) and the ion acceleration force exerted by this field is defined by Equation (6):

$$E = \frac{V}{s} \tag{5}$$

$$F = Ez \tag{6}$$

From Newtonian mechanics, it is known that the force acting upon mass M_1 creating an acceleration of a within electric field E can be described by Equation (7):

$$F = M_1 a \tag{7}$$

As shown by Equation (8), combining and rearranging Equations (6) and (7) provides an expression describing

the acceleration of ion M_1 with charge z in field E:

$$a = \frac{Ez}{M_1} \tag{8}$$

From the simple equations of rectilinear motion, it is known that a particle with no initial velocity achieves a final velocity (v) after being accelerated at a rate of a through distance s as described by Equation (9):

$$v = (2as)^{1/2} \tag{9}$$

Substituting Equation (8) into Equation (9) yields Equation (10), which describes the final velocity of M_1 after leaving the ion optic assembly:

$$v = \left(\frac{2sEz}{M_1}\right)^{1/2} \tag{10}$$

Equation (10) describes v in terms of $\mathrm{d}s/\mathrm{d}t$. Consequently, the inverse of Equation (10), Equation (11), describes $1/v$ in terms of $\mathrm{d}t/\mathrm{d}s$:

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \left(\frac{M_1}{2sEz}\right)^{1/2} \tag{11}$$

In order to determine the TOF of ion M_1 in region s, one integrates $\mathrm{d}t$ with respect to s, yielding Equation (12):

$$t_s = \left(\frac{M_1}{2Esz}\right)^{1/2} 2s \tag{12}$$

The product of the free flight distance x and Equation (11) defines the TOF for ion M_1 in the free flight region. Consequently, the total TOF (t_t) for M_1 is given by the sum of the latter with Equation (12), which when simplified is described by Equation (13):

$$t_t = \left(\frac{M_1}{2Esz}\right)^{1/2} (2s + x) \tag{13}$$

From Equation (13) it can be seen that t_t is related to $(M_1)^{1/2}$. Accordingly, ion M_1 will reach the ion detector prior to ion M_2. Equation (14) is now derived to describe the M_1/z with respect to t_t, E, s, and x:

$$\frac{M_1}{z} = \frac{t_t^2 2Es}{(2s + x)^2} \tag{14}$$

From Equation (14), the M/z of any ion can be determined by measuring its total flight time once values for E, s, and x are established. For a given TOFMS system, the values of E, s, and x are intentionally held constant, reducing Equation (14) to Equation (15):

$$\frac{M_1}{z} = kt_t^2 \tag{15}$$

A more classical derivation of the TOF equation can be performed by examining the conservation of kinetic energy for an ion of mass M_1 and charge z, traveling with final velocity v as the consequence of being accelerated through a potential difference of voltage V. This derivation is appropriate for a system where free flight distance x far exceeds distance s and, as such, free flight time for ion M_1 is far greater than its acceleration period, making t_t and ion free flight time equivalent.

Newtonian mechanics describe the kinetic energy of ion M_1 traveling with a final velocity v as noted in Equation (16). Electrostatic theory equates the kinetic energy of the same particle with charge z accelerated through an electrical potential difference V as shown in Equation (17). Equations (16) and (17) are combined and rearranged in Equation (18) to describe ion M_1/z in terms of v^2 and V:

$$KE = \frac{mv^2}{2} \tag{16}$$

$$KE = Vz \tag{17}$$

$$\frac{M_1}{z} = \frac{2V}{v^2} \tag{18}$$

In this simplified model, v is taken to be the quotient of distance x and flight time t_t. Substituting the latter into Equation (18) yields Equation (19):

$$\frac{M_1}{z} = \frac{2Vt_t^2}{x^2} \tag{19}$$

For a given TOFMS system, the values of V and x are held constant, thus reducing Equation (19) to Equation (15).

From Equation (15), it can be seen that the m/z of any ion can be determined by measuring its flight time, once a value for the constant k is derived. Determining k for a TOFMS system is referred to as calibrating the mass spectrometer. Calibration is achieved by measuring the flight time for a number of well-characterized analytes and reconciling this time with established m/z values. For the most part, a least-squares-fit algorithm is applied to determine slope and intercept values for the linear relationship between m/z and the square of total flight time. A calibration expression similar to that of Equation (20) is subsequently used to determine the m/z for all unknown samples:

$$\frac{m}{z} = at_t^2 \pm b \tag{20}$$

Two calibration strategies are typically employed: internal standard and external standard calibration. An internal standard calibration approach spikes the sample of interest with at least one and usually a plurality of calibrants. In this fashion, the resultant spectrum from this measurement is calibrated using known m/z values for each calibrant, and a highly accurate m/z determination for the unknown is created. External

calibration experiments are performed when calibrants and unknowns are analyzed in independent experiments. The calibration function derived during the calibration experiment is applied to the unknown spectrum to provide a determination of m/z.

Because slight variations in V, s, and x will occur from measurement to measurement and from sample to sample, internal standard calibration experiments typically provide about 10–100-fold improvement in mass accuracy compared to external standard approaches. However, the addition of calibrants to an unknown sample matrix is often plagued with difficulties. In some instances, the ion formation potential of the added calibrants may exceed that of the unknowns, producing significant suppression in the formation of unknown ions, thus squelching the detection of the desired samples. In other cases, similar or identical m/z for unknown and calibrant species may result in the masking of unknown signals. The latter two are particularly troublesome for TOFMS approaches that do not employ an upstream chromatographic or sample separation scheme such as laser desorption/ionization/time-of-flight (LDI/TOF) and TOFSIMS experiments. Because of the difficulties of applying internal standard calibration approaches, external standard measurements are employed most routinely.

4 TIME-OF-FLIGHT MASS SPECTROMETRY INTERFACE CONFIGURATIONS, ION EXTRACTION GEOMETRIES, AND OPERATIONAL SCHEMES

4.1 Interface Configurations

Interface configurations for the TOF mass spectrometer are typically dependent upon sample inlet system requirements, sample ionization means, as well as sample extraction and mass analyzer geometries. Inlet system and ionization source requirements are often tightly coupled. For the most part, the state of the sample (gas, solid, liquid) and its means of introduction dictate inlet system and ionization specifics.

In many instances, the TOF device serves as a MW detector for some upstream chromatographic scheme such as GC, high-performance liquid chromatography (HPLC), CE, and CEC. In this fashion, a dynamic interface supporting continuous-flow sample introduction is utilized along with an appropriate ionization source. In the case of GC, modern devices couple a continuous-flow gas interface with EI, CI, metastable atom bombardment (MAB), or supersonic molecular beam (SMB) ionization sources. For LC, CE, and CEC, an atmospheric-pressure, liquid-phase interface is coupled with primarily ESI or CI sources.

Static sample introduction of solid-state and liquid analytes is typically achieved using a solid probe interface. The probe can be a simple, single-position device capable of rotation or it can be a more complex multiposition device capable of rotational, rectilinear, or two-dimensional (linear–linear, or circular–linear) movement. Such interfaces are typically employed with LDI, MALDI, surface-enhanced laser desorption/ionization (SELDI), PD, and SIMS sources.

4.2 Ion Extraction Geometries: Parallel Ion Extraction

Presently, TOF devices use one of two fundamental schemes of ion extraction: parallel ion extraction and orthogonal ion extraction. Static sample interface systems typically utilize parallel ion extraction. A parallel ion extraction LDI device is depicted in Figure 2. In this example, the energy inherent in a short-duration laser pulse creates ions in a well-defined region immediately above a solid probe. In this case, the solid probe is contiguous with the repeller lens of the ion optic assembly. In some instances, such as with two-dimensional sample stages, the solid probe is the repeller. In contrast to the simple ion optic assembly of Figure 1, the assembly of Figure 2 is a dual-stage design containing an additional plate known as an extractor. During ion acceleration, the extractor potential is established to be less than that of the repeller and greater than that of ground.

For this geometry, ion extraction can be performed in a continuous or pulsed manner. In the case of continuous ion extraction (CIE), ions are formed in the presence of an accelerating field created by the potential drop between repeller and extractor elements, tightly coupling ion formation with ion extraction. In PIE, ions form in an equipotential region created by holding the repeller and extractor lenses at the same voltage. Ions are subsequently extracted by the application of a high-voltage pulse, which either raises the repeller or lowers the extractor potential, creating an accelerating electrical field.

Extracted ions are further accelerated between the extractor and ground lens elements and ultimately enter the free flight region. In some devices, emitted ions enter a short free flight region located between the ion optic assembly and an ion deflector, where they begin to separate out in time and space as dictated by their m/z-dependent final velocities. The ion deflector of Figure 2 incorporates a time-based, pulsed, orthogonal, electrostatic acceleration field, which perturbs the trajectory of all ions that enter it when the pulse is applied. In this manner, ions of little interest can be deflected from striking the ion detector. As will be explained further, the latter improves detector sensitivity and signal baseline stability. The pulse is shut off to allow all other ions ultimately to strike the detector.

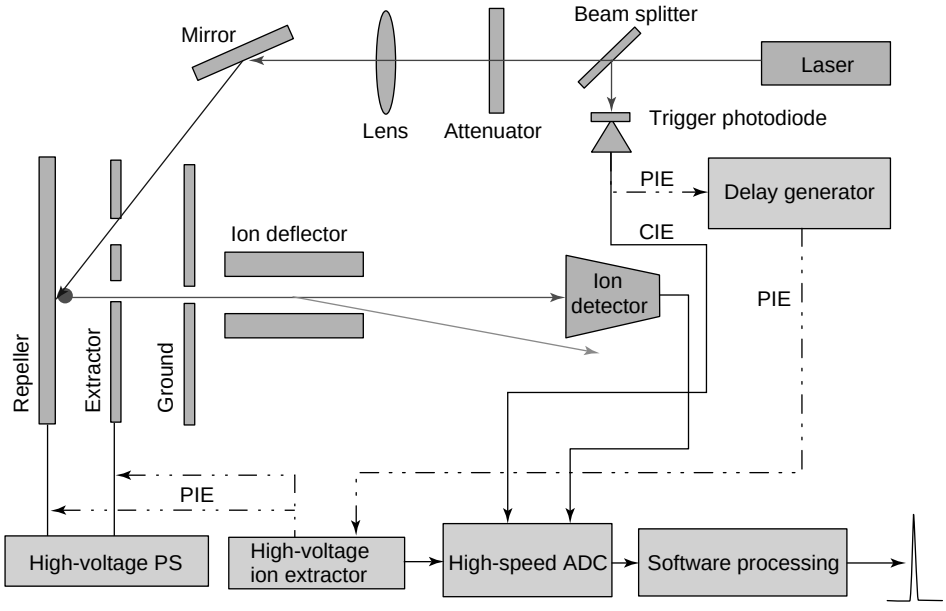

Figure 2 A parallel extraction, LDI, linear TOF mass spectrometer. Both constant extraction and PIE modes are illustrated; PS = power supply. See text for details.

Ions striking the detector are converted to electrons and the subsequent current creates an electrical potential that is digitized by a high-speed, time-array recording device that digitally captures the entire analog signal of each TOF cycle. Time-array recording devices include waveform (or transient) recorders, analog-to-digital converters (ADCs), and digital oscilloscopes. In many cases, digitization is provided by a high-speed ADC. The ADC transfers digital information to a data reduction software package, which provides visualization and higher order processing of the ion signal. Resolved ion populations are typically represented as a peak whose intensity and area are related to abundance, and flight time is related to m/z.

Flight time measurement is performed by the high-speed ADC. After receiving a start trigger, the ADC integrates detector output voltage at regular time intervals. Arrival of the ADC start signal is coordinated with the onset of ion extraction. For CIE systems, the start signal is tightly coupled to the point of ion creation. In Figure 2, an optical beam splitter is employed to direct a small portion of the laser beam to strike a trigger photodiode that serves as a lasing event detector. The lasing event signal (photodiode solid lines) provides the ADC start trigger for CIE analyses. For PIE approaches, the lasing event signal (interrupted line) provides a start trigger for a time-delay generator, which, after a pre-established time period, provides a start trigger to a pulsed high-voltage ion extractor. Coincident with the applied ion extraction pulse, the high-voltage ion extractor provides a start signal to the ADC.

After receipt of the start trigger, the ADC sorts integrated detector voltage values into discrete time-dependent bins. Each bin possesses a finite level of time resolution that is dependent upon the data acquisition rate (DAR) of the ADC. For the most part, ADCs with 500–2000 megasample s^{-1} acquisition rates are employed, providing a temporal resolution of 2000–500 ps per data point. ADC data storage continues for a scan period that is defined by the TOF of the heaviest ion of interest. It is not unusual to see scan periods of the order of 200 μs, requiring about 400 kbytes of memory for a 2000 megasample s^{-1} DAR.

Because the majority of ions formed are extracted in every scan and because an ADC takes data at every scan-period time bin, parallel extraction approaches can provide a complete mass spectrum after a single scan. However, it is common practice to sum a number of scans together to improve the overall signal-to-noise ratio (S/N) for the resultant spectrum. Furthermore, geographical heterogeneity of solid-state samples frequently requires the interrogation of different regions to provide complete ion profiles. Consequently, single-scan spectra are rarely compiled.

Scan rates of parallel extraction devices are basically limited by the ion source duty cycle and sample movement time. In the case of LDI, repetition rates are of the order of 5–20 lasing events per second and fine sample movement is achieved in about 300 ms. A quality spectrum is generally produced after the summing of 20–50 scans, resulting in an effective spectral acquisition time of about 1–30 s.

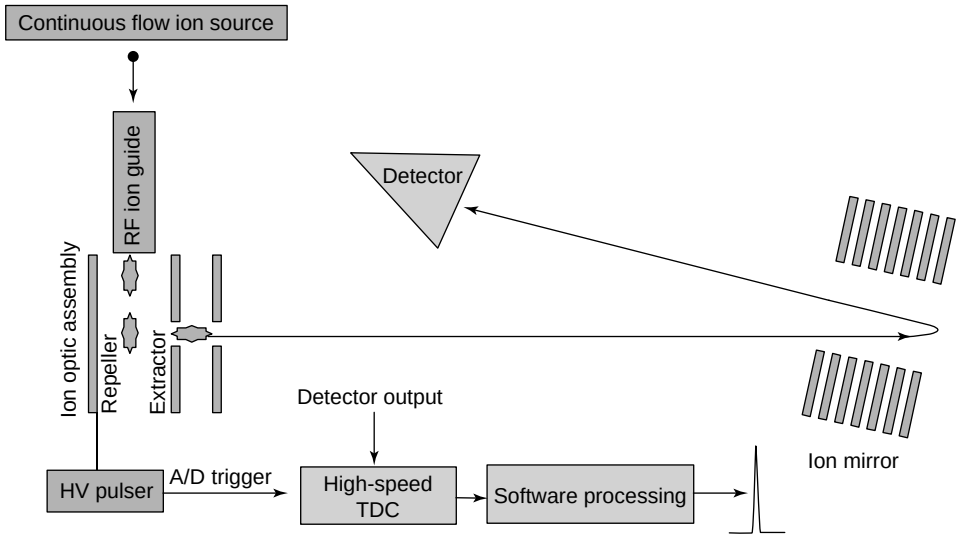

Figure 3 An orthogonal extraction TOF device with reflectron analyzer. See text for details.

4.3 Orthogonal Ion Extraction

Orthogonal extraction geometries are generally used with dynamic or continuous-flow interfaces. Historically, this geometry has been coupled with EI analysis of cold molecular beams as well as atmospheric-pressure plasma ion sources.[50] More recently these approaches have been coupled with ESI[16,51–55] and MALDI.[56–61] Figure 3 illustrates a typical orthogonal extraction TOF device coupled with a continuous-flow interface. Unlike the linear, parallel extraction geometry of Figure 2, this analyzer uses an ion mirror or reflectron to provide improved mass resolving power. The specifics of this process will be discussed later.

Ions enter the TOF analyzer by traveling through an ion guide assembly. For the most part, these ion guides are multipolar RF devices usually consisting of four or eight cylindrical rods. A four-pole array is known as a quadrupole assembly and an eight-pole configuration is referred to as an octopole assembly. A potential well is created within the center of this guide that functions to constrain ion radial trajectory, thus minimizing radial ion spread. Ions travel through the length of the guide by primarily relying upon their nascent velocities and secondarily assisted by accelerating electrical fields and pressure drops. As ions travel through the guide, they undergo a series of collisions with background gas. These collisions function to dampen the initial energy of these ions to a point that approximates to that of their thermal state, almost eliminating their radial velocity, while slowing their axial velocities down to the order of $25\,m\,s^{-1}$.[55] As will be explained later, the latter is essential to provide reasonable mass resolving power.

A dampened, collimated ion beam enters a field-free region of the ion optic defined between the repeller and extractor plates. When this region becomes filled with ions, the potential of the repeller plate is elevated through the use of a high-voltage pulser, and the ions are ejected out of the optic assembly and into the free flight region. Ions enter the ion mirror and then are turned about to strike a detector.

The ultimate rate at which ions are extracted is dependent upon a balance that is struck between the flight time of the heaviest ion and the time required for all ions to enter and not overfill the extraction region. Should the required flight time exceed that of the fill-up time of the extraction region, the instrument's duty cycle is effectively reduced and sensitivity suffers.[50] For the most part, ions are extracted at rates of 10^4–10^5 cycles per second. In many instances, these extraction rates significantly exceed the rate of ion introduction, dividing the introduced ion load into a number of low-ion-density TOF scans. Resultant TOF scans typically contain less than one ion per injection pulse per ion species. This is in stark contrast to parallel extraction, where many ions for a given species are introduced in every scan.

In most orthogonal extraction TOF spectrometers, a high-speed, time-interval-recording device such as a time-to-digital converter (TDC) measures ion TOF. Unlike time-array devices, a TDC is an individual pulse-counting mechanism that records the detector signal over a finite time span. The recording event is followed by a dead time period when the TDC remains refractory to any further pulse counting. Only a single impulse is stored in each time bin. Typical time bin resolution is of the order of hundreds of picoseconds, whereas dead time periods exist for a few nanoseconds. Considering the

latter, it becomes clear that true waveform reproduction is solely possible when only a single ion is present for each counting period. Furthermore, a single TOF scan does not produce a complete spectrum. Consequently, many thousand of scans must be acquired to provide a spectrum that accurately reflects the true ion population.

Triggering the TDC is generally accomplished by the high-voltage ion pulser, thus establishing the beginning of the TOF cycle at the point of ion extraction. Pulse counting occurs when detector output voltage exceeds a pre-established level of discrimination. This is followed by a dead period, after which the cycle repeats. Several ion counts are performed for each scan. Many scans are summed within the TDC prior to transfer to the system computer, where signal-reduction software ultimately provides visualization and higher order processing of the ion profiles.

For a given chromatographic or static analysis, total spectral acquisition time for this approach is dependent upon mass range and sensitivity requirements. With increasing mass range, some systems suffer duty cycle loss, requiring greater scan numbers. For trace analysis, large scan numbers are also needed. The rate of data transfer from the TDC to the data-reducing computer also plays a limiting role. For the most part, quality spectra are produced within a few seconds of time. In the extreme case of trace analysis and/or extended mass range, acquisition times may approach 30–60 s. Obviously, the latter becomes obviating in terms of high-resolution chromatographic techniques. Consequently, most continuous-flow, trace analysis approaches avoid upstream chromatography and resort to flow introduction of a cleaned up or homogeneous sample.

5 FACTORS AFFECTING PERFORMANCE IN TIME-OF-FLIGHT MASS SPECTROMETRY

5.1 Mass Accuracy

As is the case with many other analytical methods, fundamental performance in TOFMS is described in terms of system accuracy, resolving power, and sensitivity. TOFMS devices measure the fundamental mass-to-charge ratio (m/z) of an ion. Ion m/z is defined by Equation (21), where A_{mw} is analyte MW, Cs_{mw} is the total MW of added or deleted charging species, and cs is the total charge of all charging species:

$$\frac{m}{z} = \frac{A_{mw} \pm \sum Cs_{mw}}{\sum cs} \tag{21}$$

Once the nature of an ion's charge is established, analyte MW is easily determined. In most cases, cations are created by the addition of a proton (H^+), a metal (Na^+, K^+, Ag^+, etc.), or the loss of an electron. Anions usually arise by the loss of a proton or the addition of an electron. It is common to ignore electron mass when determining analyte MW.

The metric of mass accuracy is often discussed in terms of error with respect to analyte true mass. Error is usually reported as an absolute difference in m/z or MW (Da), or as a fractional error with respect to true MW (percent error or parts per million). To understand the sources of mass measurement error in TOFMS, it is best to re-examine the fundamental TOF expression as described by Equation (14) for the TOF arrangement of Figure 1.

Equation (14) describes the determined m/z of ions with charge z and total kinetic energy that is entirely dependent upon acceleration within the ion optic assembly. In this simplified case, all ions are assumed to form at the same time, at the same ion optic location, with no initial kinetic energy. In reality, this is rarely the case and most ion populations demonstrate some variance in initial position, time of formation, and possess some initial kinetic energy (U_o). Considering the latter, Equation (14) must be modified to account for U_o, yielding Equation (22), where E is ion acceleration field strength, s is the distance through which an ion is accelerated, x is ion free flight distance, and t_t is total ion flight time:

$$\frac{m}{z} = [(2Esz + U_o)(2s + x)^{-2}]t_t^2 \tag{22}$$

Rearranging Equation (5) in terms of applied voltage and substituting this into Equation (22) yields Equation (23), where V is the voltage applied to the repeller (see Figure 1):

$$\frac{m}{z} = [(2V + U_o)(2s + x)^{-2}]t_t^2 \tag{23}$$

In most TOF systems, design approaches are taken to minimize variation in V, s, and x, thus reducing Equation (23) to Equation (15), and enabling simple calibration of the system as noted by Equation (20). During actual practice, system warm-up time and ambient temperature conditions often result in high voltage instability and subtle shifts in ion acceleration and free flight distances. Furthermore, changes in sample introduction conditions, such as ion phase space and solid crystal size or film thickness, create notable alterations of ion acceleration distance. Likewise, variations in the position of ion formation create changes in both s and x, and variations in time of ion creation or onset of TOF measurement create ambiguities in t_t. All of these factors combine to perturb the consistency of V, s, x, and t_t, resulting in attendant mass assignment error.

Variation in U_o is another source of mass assignment error. If U_o increases, total kinetic energy increases

and the resultant TOF for a given ion will accordingly decrease, creating a lighter than actual MW determination. Should U_o decrease, the opposite occurs. Because of the random, combinatorial effects of changes in V, s, x, t_t, and U_o, internal standard measurements generally provide greater mass accuracy than external standard approaches.

A special consideration of U_o and mass assignment accuracy occurs if U_o is not constant for all m/z. This condition violates the assumptions for Equations (22) and (23) and thus the simple linear relationship of calibration (Equation 20) no longer applies. Such a condition has been reported for MALDI/TOF experiments with both CE and PIE. Zhou reported the existence of an energy deficit between low-molecular-weight ions and higher-molecular-weight ions as studied by MALDI with CE.[62] In this instance the U_o of some proteins and peptides in the range 1–12 kDa was found to decrease with increasing MW. Nelson reported a trend of increasing U_o with analyte MW while analyzing human IgM (MW $\sim$ 980 kDa) using MALDI/TOF with CE.[6] Unpublished results obtained by this author on the study of a number of proteins in the range 6–100 kDa using MALDI/TOF with CE and PIE agreed with the findings of Nelson. This apparent increase in U_o as a function of increasing MW may be explained by considering the initial desorption velocities (V_o) of MALDI-formed ions.[63]

Several studies of V_o for MALDI-formed ions have been performed.[62,64–69] Although discrepancies exist in terms of the absolute value of V_o, a common trend of decreasing V_o with increasing analyte MW is reported. Furthermore, average ion V_o has been demonstrated asymptotically to reach a minimum value at some ultimate m/z. Consequently, ion U_o as dictated by a constant V_o would increase with increasing MW as predicted by Equation (16).

Figure 4 depicts the effect of U_o upon mass calibration by plotting analyte MW against the square of ion flight time. Two major populations of ions are illustrated: $m/z < 200$ kDa and $m/z > 400$ kDa. A second-order-fit line is easily driven through all points. Clearly, the simple, linear relationship between m/z and the square of ion flight time does not apply. If a simple linear fit were employed to calibrate the system using the first several points of the 200-kDa group, the function depicted by the low-MW calibration line would result. Applying this low-MW calibration function to the measurement of ions in excess of 200 kDa would result in determined MWs that are less than their actual values. In a similar manner, if calibration was performed using the last several points of the >400-kDa group, a calibration function as depicted by the high-MW calibration line would result. If low-MW ions were analyzed using the high-MW calibration expression, measured MWs would again be less than their actual values.

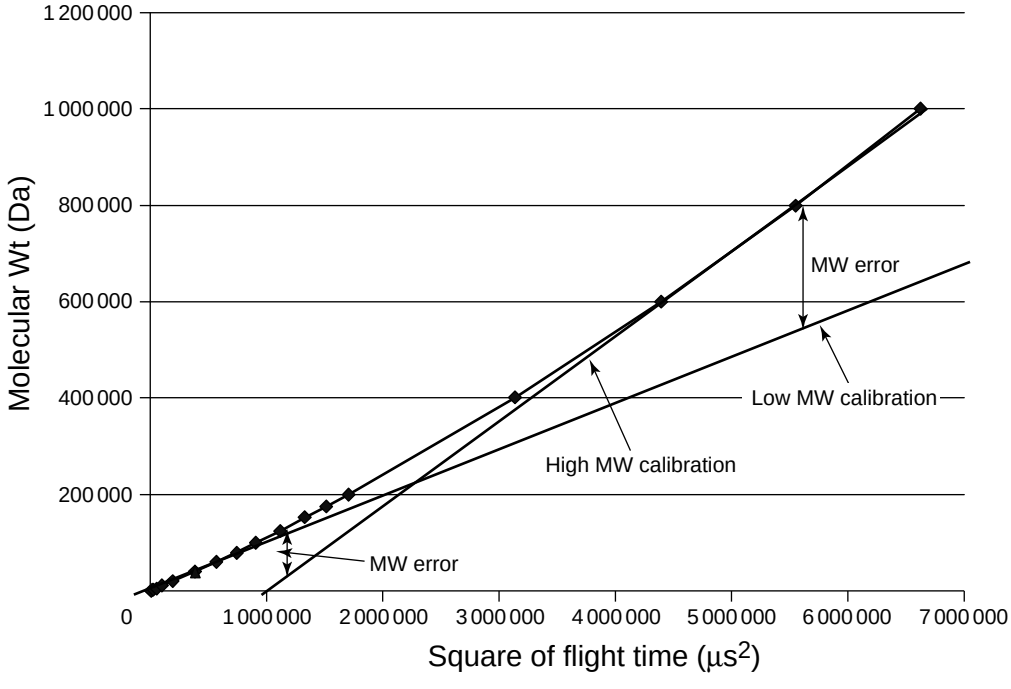

Figure 4 The effect of ion initial kinetic energy (U_o) on the linearity of the TOF calibration expression and mass determination accuracy is presented. In this plot, ion TOF was calculated based upon 30-kV acceleration, $V_o = 350\,\mathrm{m\,s^{-1}}$, and 1-m free flight. The trend is carried out to 1.2 mDa for demonstrative purposes.

As demonstrated, non-linearity of the TOF equation will result in additional mass assignment error whenever measurements are performed outside the m/z limits of the calibration experiment and extrapolated MWs are determined. This applies to both internal and external standard approaches. Although it is possible to correct for this nonlinearity by using higher order fits,[70] accurate MW determination for large ions remains dependent upon calibrating against a well-characterized standard in the high-MW domain. In reality this approach is problematic, because very few well-characterized calibrants with MWs in excess of 70 kDa are readily available today.

An instrumental means to address this problem was taken by Whittal et al.[63] This approach, known as functional wave TLF, employed an ion extraction pulse for which a time-dependent change in extraction field strength (E) was applied in a manner to compensate for U_o differences between light and heavy ions. At the beginning of the ion extraction cycle, E is held at its maximum value. As ions are accelerated out of the extraction region, the value of E decreases. In this manner, ion total acceleration will be dependent upon m/z and the integral of the wavefunction for E during the ion extraction period. Because lighter ions leave the extraction region before heavier ions, they receive a greater degree of acceleration when compared to their heavier counterparts. A functional wave pulse is selected to provide a mass-dependent difference in total ion acceleration that compensates for differences in U_o between light and heavy ion species, allowing for the use of simple, linear-fit calibration expressions. A 10–20-fold improvement for linear-fit, internal standard mass assignments for several proteins in the range 5.7–29 kDa was demonstrated for the functional wave system when compared to the conventional pulsed ion approach of TLF.

The overall affects of varying initial ion position, ion time of formation, and ion U_o upon mass accuracy is dependent upon ion extraction geometry. For the arguments presented above, parallel ion extraction geometry is assumed. In the case of modern orthogonal extraction, collisional cooling minimizes ion radial U_o disparities that would deleteriously affect total kinetic energy in the direction of TOF measurement. Furthermore, any change of initial ion axial velocity does not affect total ion flight time in the orthogonal direction. In this fashion, the effects of ion U_o are greatly removed from producing mass assignment error. Because ions are created downstream, often requiring 1–2 ms to enter an orthogonal extraction region, final ion TOF becomes independent of the original time or location of ion formation. However, within the ion extraction and free flight regions, mass accuracy in orthogonal TOF devices remains affected by variations in ion acceleration voltage, as well as variations in ion acceleration and free flight distances.

Dependant upon ion source and ion extraction geometries, today's state-of-the-art TOF devices provide typical mass accuracy of the order of 1–10 ppm for internal standard measurements and 20–200 ppm for external standard approaches. Because initial ion position, ion time of formation, and ion U_o are uncoupled from total ion flight time in orthogonal TOF, orthogonal devices typically provide superior external standard mass accuracy when compared to parallel extraction systems.

5.2 Resolution

Resolution is basically the ability to detect subtle differences in m/z among different ion populations. Several mathematical models and definitions of resolution have been suggested.[71] For the most part, TOF mass resolving power has been defined by Equation (24), where m is the determined mass of an ion, T is ion TOF, dm is the width (in Da) of the measured peak at half-magnitude, and dT is the half-height-measured peak width in time:

$$R = \frac{m}{\mathrm{d}m} = \frac{T}{2\,\mathrm{d}t} \tag{24}$$

Equation (24) may be derived by first rearranging Equation (15) in terms of M_1 to yield Equation (25). Taking the derivative of M_1 in Equation (25) with respect to t_t yields Equation (26). Dividing Equation (25) by Equation (26) yields Equation (24).

$$M_1 = zkt_t^2 \tag{25}$$

$$\mathrm{d}M_1 = zk2t_t\,\mathrm{d}t_t \tag{26}$$

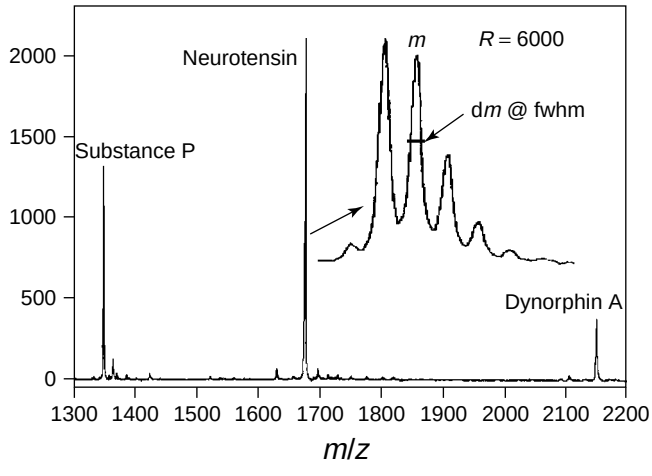

Figure 5 MALDI/TOF analysis of substance P, neurotensin, and dynorphin A. An expanded view of neurotensin is provided, demonstrating the fwhm and resultant resolution of about 6000 m/dm. (Figure adapted from data provided by Dr Robert Brown, Utah State University.)

Figure 5 depicts a MALDI/TOFMS analysis for a mixture of peptides. The neurotensin signal is expanded to reveal the individual isotopic populations. In this figure, the full width at half-maximum (fwhm) of the ^{13}C peak is indicated. The resolution of this signal was determined to be roughly 6000 m/dm.

The interdependency of mass accuracy and mass resolving power has been known for some time. In the 1960s, Campbell and Halliday showed that the standard deviation of mass measurement is related to signal resolution and the number of ions detected for a given species.[72] When resolution is extremely high, single-isotope ion populations are detected with good counting statistics and ion TOF is easily determined as the apex of a symmetrical peak. When resolution suffers, isotope discrimination becomes limited, and the introduction of peak asymmetry from convoluted isotopic envelopes as well as distortion due to poor ion counting statistics skew the true mass from the peak's apex. In order to compensate for these effects, ion TOF is generally estimated using some form of moment analysis. Gaussian or Lorentzian algorithms are often applied to determine the centroid of the measured signals. These centroid TOF values are subsequently used to determine m/z.

Factors affecting TOF resolution are associated with the initial conditions of ion formation and acceleration, free flight perturbations, as well as instrumental limitations. Figure 6 demonstrates the effect of various parameters of initial ion formation upon time spread for a homogeneous population of ions constantly extracted by a parallel electrical field. The upper panel describes a system in which all ions are introduced at the same location at the same time, with the same initial speed and equivalent energies. However, initial ion directions of travel are different. Consequently the initial velocities of these ions (V_o) with respect to the plane of detection are different, as indicated by V_1, V_2, and V_3. Ion optic acceleration and focusing of these ions drive all species toward the detection plane. Ion final velocity is the vector sum of ion V_o and imparted velocity due to electrostatic acceleration. Because of the latter, these three ions achieve different final velocities and strike the detector at different times where the TOF of $V_1 < V_3 < V_2$. In this particular case, mass resolving power is limited by the spread of V_o and cannot be improved by lengthening free flight. Resolution may only be enhanced by methods that compensate for differences in V_o by using differential acceleration or by those that provide different free flight distances for ions of different final velocities.

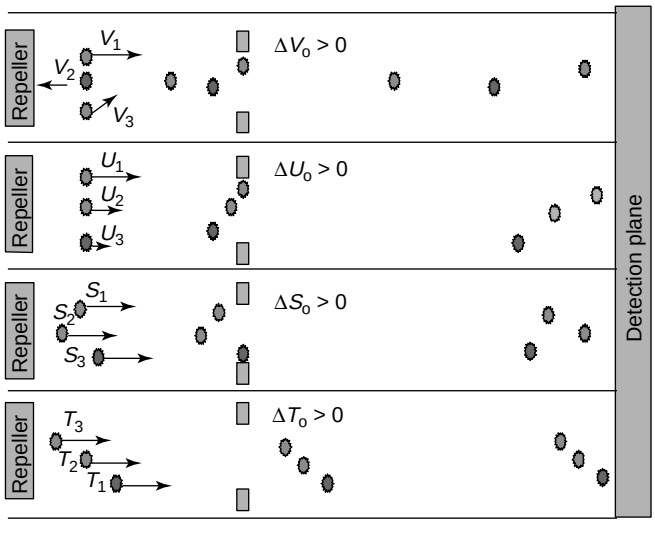

Figure 6 The effect of varying initial velocity, initial energy, initial position, and initial time of ion formation upon ion flight time and resolution. See text for further details. (The figure is recreated from the teachings of Professor Robert J. Cotter, The Johns Hopkins University.)

The second panel displays a system in which all ions are created at the same time, in the same location, with different speeds traveling in identical directions. In this manner, these ions have different initial energies U_o, created by differing V_o. All ions are identically accelerated to different final velocities, creating a resolution-limiting, ion arrival time spread at the detection plane. As in the first panel, resolution may only be enhanced by methods that compensate for δV_o and δU_o.

The third panel illustrates a condition where all ions are formed at the same time, with identical V_o and U_o at different locations. In this instance, these ions vary in final velocity (v) (and final energy) because the acceleration that each ion receives is different. As predicted by Equation (10), ions closer to the repeller are accelerated to greater v than those located away from the repeller. Although initial ion positions place S_3 as being the closest to the detector, followed by S_1 and S_2, differential acceleration of these ions creates final TOF profiles for which S_2 arrives at the detector first, followed by S_1 and then S_3. In this example, resolution is limited by the initial spatial distribution of all ions formed and cannot be corrected by lengthening the free flight. However, spatial spread may be corrected by applying an appropriate degree of differential acceleration so that all ions are temporally coincident upon the detection plane.

The lower panel depicts a scenario in which all ions are formed in the presence of an accelerating field at the same location, with equal U_o and V_o, but at different times. Here all ions are accelerated to identical final velocities, but a fixed time spread defined by the δT of ion formation will be maintained throughout the TOF period. In this case, lengthening of the free flight distance may enhance resolution.

Because it is customary to sum several TOF scans to create a given spectrum, there is an additive effect upon

the uncertainties in V_o, U_o, S_o, and T_o. All of these variances add in quadrature, and they must be minimized to preserve mass resolving power.

When ions leave the ion optic and enter the free flight region, collisions with background gas result in momentum transfer processes that ultimately alter ion final velocity. In this fashion a time-focused, homogeneous population of ions will exhibit time spread as a direct consequence of such collisions. This effect is illustrated in Figure 7. As analyzer region pressure decreases, resolution in the MALDI/TOF analysis of human insulin, equine cardiac cytochrome *c*, and BSA improves. Schuerch et al. reported similar results for the MALDI/TOF analysis of poly(ethylene glycol) 4000.[73] Generally speaking, acceptable TOF resolution is achieved for free flight pressures in the range 10^{-7}–10^{-8} Torr.

Instrumental limitations also play a governing role in mass resolving power. Any condition that results in instability of ion acceleration, such as variance in high-voltage power supply regulation or collapse of ion acceleration fields, creates accompanying shifts of ion flight time. Furthermore, irregularity in the onset of the TOF measurement cycle with respect to ion acceleration also produces ambiguities in TOF. Because a number of TOF scans are generally summed to create spectra, these factors randomly combine to widen measured peak widths and to limit resolution.

Aside from instrumental instabilities, fundamental instrument properties such as free flight distance, total ion acceleration energy, DAR, and detector system analog bandwidth also affect mass resolving power. Figure 8 demonstrates the effect of free flight distance and ion

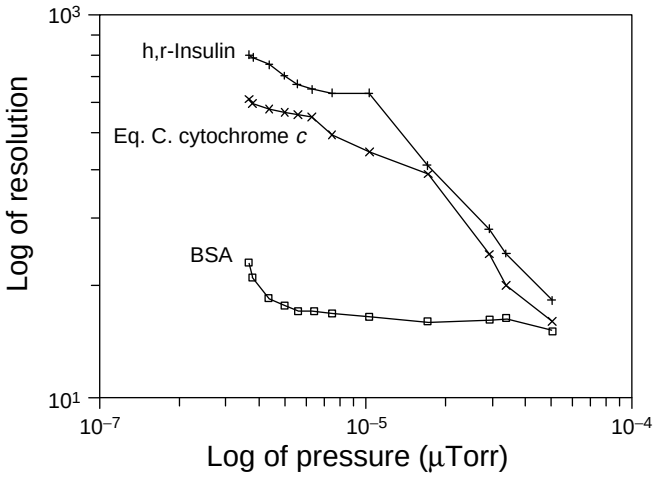

Figure 7 Resolution for the MALDI/TOF analysis of human insulin (h,r-Insulin), equine cardiac cytochrome *c* (Eq. C. Cytochrome *c*) and bovine serum albumin (BSA) is plotted as a function of free flight pressure. In all cases, resolution increases with decreasing background pressure. See text for further details. (Unpublished data from Scot R. Weinberger.)

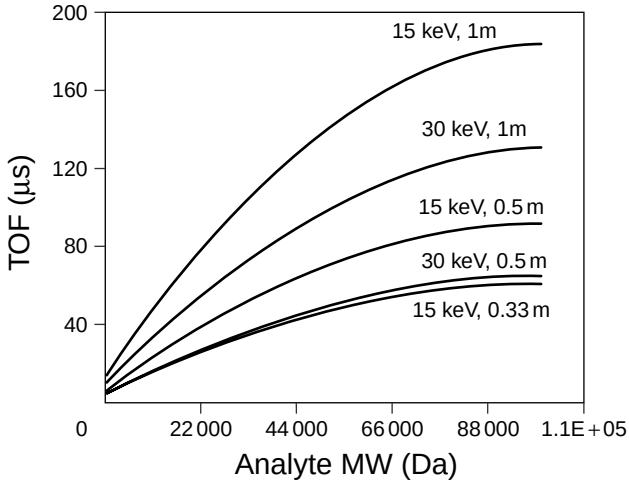

Figure 8 Analyte TOF vs MW for 15 and 30 keV of final energy and 1.0, 0.5, and 0.33 m free flight. As total energy increases and flight tube length decreases, $\delta t/\delta m$ decreases, placing a greater burden upon digitizing speed and analog bandwidth.

total kinetic energy upon the change in TOF (δt) with respect to the change in MW (δm). As free flight distance decreases and total kinetic energy increases, $\delta t/\delta m$ decreases and a greater DAR and broader analog bandwidth is required to preserve mass resolving power. It is noteworthy to discuss the effect of these factors on the resolution of a small peptide with MW ~ 1000 Da. In order to maintain unit mass resolution, a system utilizing 30 kV of acceleration and 0.33 m of free flight would require a 1-gHz detector bandwidth and must digitize at more than three gigasamples per second (330 ps time resolution). Similarly, a system using 15 kV acceleration and 1 m of free flight would require about a 300-mHz bandwidth and a one gigasample per second DAR.

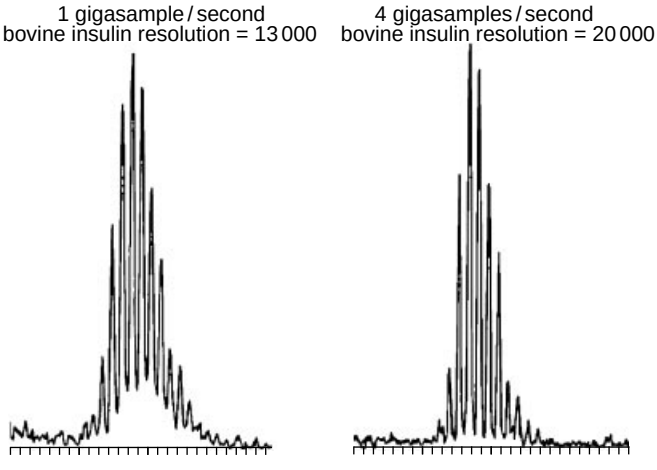

Figure 9 The effect of DAR on the resolution of bovine insulin. (Figure adapted from data courtesy of Bruker Daltonics, Billerica, MA, USA.)

Figure 9 further demonstrates the resolution effects of DAR. In this example, bovine insulin is analyzed using constant acceleration potential, free flight distance, and analog bandwidth but employing two different digitizing rates: one and four gigasamples per second. Ostensible differences in resolution are demonstrated between the one and four gigasample per second DARs.

From first principles it has been demonstrated that, for a given amount of initial energy spread, resolution improves with increasing ion acceleration potential.[3] Although this is certainly true, it is important to note the interactive impact of total kinetic energy, free flight distance, and DAR upon resolution. Figure 10 demonstrates the effect of total kinetic energy upon resolution for MALDI/TOF analysis with CE of human insulin, bradykinin, and the pentapeptide VPDPR (valine-proline-aspartic acid-proline-arginine). As acceleration potential is increased from 10 to 20 kV, the observed resolution for all three compounds improves. However, when acceleration potential is further increased, improved resolution is noted only for insulin. In this case, the additional acceleration has narrowed the bradykinin and VPDPR peak widths to a point where the system's analog bandwidth and DAR effectively limit them. As acceleration potential increases, TOF decreases but peak widths remain fixed at the data system limit. The end result is a reduction in mass resolving power. In this instance, increasing free flight distance would meliorate this effect.

As can be seen, optimizing the resolution often requires a compromise between system free flight distance, ion acceleration potential, system analog restrictions, and DAR. As will be demonstrated later, TOFMS detection sensitivity is related to ion total kinetic energy. As such, most useful TOF devices accelerate ions to at least 20 keV of total energy. Considering the latter, small, footprint devices should use broad-bandwidth, high-speed data acquisition systems to provide adequate mass resolving power and mass accuracy. Large, research-grade instruments combine 2–3-m-long free flight regions with high-energy acceleration and high-performance data acquisition systems to provide mass resolving power in excess of 10 000 m/dm and associated mass accuracy in the low parts-per-million range.

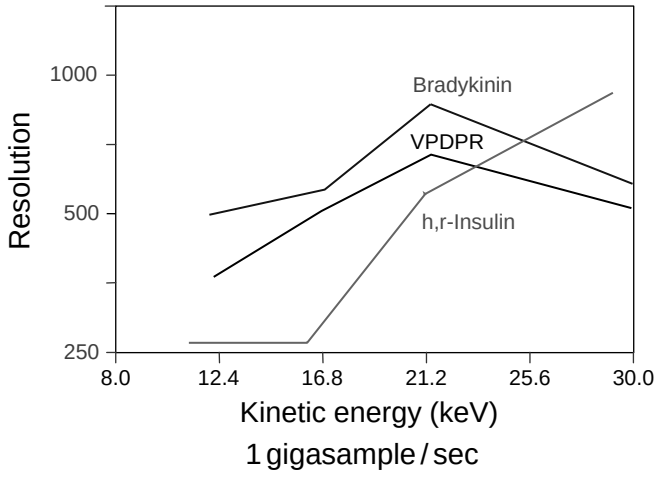

Figure 10 Resolution vs total kinetic energy for CIE MALDI/TOF analysis of human insulin (h,r-insulin), bradykinin, and pentapeptide VPDPR. Optimal total kinetic energy for resolution is dependent upon resultant peak width and system DAR. (1.0-m free flight, 500-mHz bandwidth). (Unpublished data from Scot R. Weinberger.)

6 ION FOCUSING MEANS TO IMPROVE RESOLUTION

The deleterious effects of energy, spatial, and time spread of ion creation upon mass resolving power became painfully clear to many early TOF pioneers. The resolving power of early devices was significantly limited by energy and spatial variations.[17,20,21] For many years, the TOF mass spectrometer had a reputation for providing poor resolution, greatly restricting its applicability and further fueling the growth of other MS approaches.[2] Eventually, several energy and spatial focusing techniques were developed, ultimately improving the mass resolving power in TOFMS. This section reviews these solutions.

6.1 Time-lag Focusing

Perhaps the most significant development in improving resolution in TOFMS came in 1955, when Wiley and McLaren introduced the concept of TLF.[3] Wiley and McLaren mathematically derived expressions describing ion flight time in a linear TOF device using a dual-stage ion optic assembly and an EI source. They separately considered the effects of spatial and energy aberrations upon mass resolving power. In terms of spatial variance, the optimal resolving power of a system exhibiting a maximum spatial spread of Δs was shown to be described by Equation (27), where M_s is the spatial-spread-limited maximum resolution, s_o is average initial ion position, and k_o is an instrumental constant:

$$M_s \approx 16k_o \left(\frac{s_o}{\Delta s}\right)^2 \tag{27}$$

In the case of no spatial spread, they noted the maximum resolving power (M_u) of a device with an initial energy spread of U_o and total kinetic energy of U_t to be described

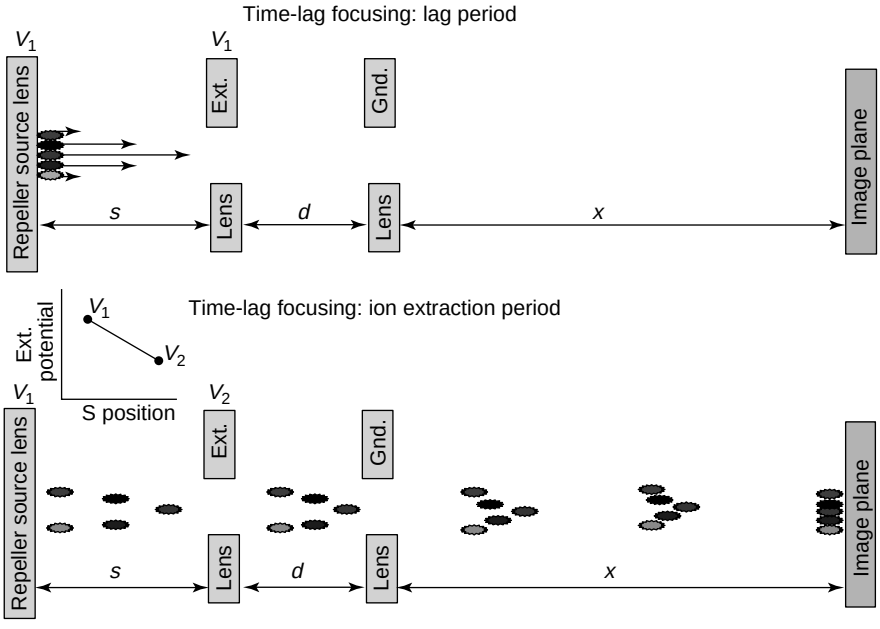

Figure 11 Steps involved in TLF. See text for details.

in terms of Equation (28). Overall mass resolving power was described as shown in Equation (29):

$$M_u = \frac{1}{4}\left(\frac{U_t}{U_o}\right)^{1/2} k_o \tag{28}$$

$$\frac{1}{R} = \frac{1}{M_s} + \frac{1}{M_u} \tag{29}$$

Wiley and McLaren conceived the concept of energy or spatial TLF to correct for variations in initial ion energy or position. Let us consider the case of energy TLF. PIE is employed and ion formation and acceleration become uncoupled (see Figure 11). In the first phase of TLF – the lag period – ions are created in a field-free region and are allowed to drift to final positions that are dependent upon their nascent velocities. Upon completion of the lag period, an ion extraction pulse is applied to either increase the potential of the repeller or decrease the potential of the extractor, thus creating an accelerating electrical field. At the time of extraction, ions have drifted to some final position that is correlated to their initial velocities. Ions closest to the repeller – the lagging group – are accelerated through a greater extraction (ext.) potential difference than those further away – the leading group. The parameters of lag time and pulse voltage are selected so that lagging ions catch up with leading ions at some image plane. In most cases, the image plane is the point of ion detection.

Energy TLF conditions are determined by taking the derivative of the total ion flight time (T) with respect to initial energy ($\mathrm{d}T/\mathrm{d}U_o$) and solving for the condition where $\mathrm{d}T/\mathrm{d}U_o$ is zero. The optimal lag for a given ion using a pulse voltage of V_p is described by Equation (30), where t_L is lag time, s is the length of ion extraction, m is ion MW, z is ion charge, U_o is ion initial kinetic energy and k_1 is an instrumental constant:[74]

$$t_L \propto s\left(\frac{m}{z}V_p\right)^{1/2} \frac{1}{k} + \left(\frac{U_o}{V_p}\right)^{1/2} \tag{30}$$

Space focusing conditions are similarly arrived at by taking the derivative of total ion flight time with respect to initial ion position (s) and solving for the condition where $\mathrm{d}T/\mathrm{d}s$ is zero. When this is done, Equation (31) is derived, where D is the point of ion focus, s_o is average initial ion position, k_o is the instrument constant of Equations (27) and (28), and d is the acceleration distance of the ion optic second stage (see Figure 11). As can be seen, this focus condition is the same for all ions and is independent of the total energy of the system.[3]

$$D = 2s_o k_o^{1/2} \frac{1-d}{(k+k^{1/2})s_o} \tag{31}$$

Wiley and McLaren noted that solutions for energy focusing violate conditions for space focusing, and therefore first-order TLF corrections for both spatial and energy perturbations are not possible. In this manner, TLF conditions are chosen to compensate for the factor that most greatly minimizes resolution. Because energy-focusing conditions are dependent upon analyte m/z, no single set of conditions provides enhanced resolution over the entire mass range. For a given lag time and instrument

constraints (ion optic subsequent acceleration and total free flight distance) the required pulse voltage for energy focusing is directly proportional to m/z. Conversely, under the same assumptions, the required lag time of a constant ion ejection pulse is related to the square root of m/z.

TLF has been applied to a variety of ionization schemes, including laser desorption,[75] MPI,[76] MALDI,[74,77–83] and orthogonal TOF geometries.[54–58] In the case of MALDI/TOF, TLF has been referred to as delayed extraction, delayed ion extraction, dynamic extraction, and space velocity correlation focusing. The addition of TLF to MALDI analyses has resulted in significant improvements in mass resolving power, mass accuracy, and analytical sensitivity. Uncoupling MALDI ion creation from ion acceleration has resulted in softer extraction of analyte ions from dense clouds of matrix ions and neutrals. In this manner, collisions between parent ions and matrix products are minimized, decreasing energy dispersion as well as fragmentation of analyte ions during the extraction process. The result is a more narrowed energy distribution and marked reduction of chemical noise.[81] An additional benefit of TLF and MALDI is the ability to obtain structural information from prompt fragmentation.[77] This technique, known as in-source decay (ISD) analysis, will be discussed further in section 14.2.

The addition of TLF to continuous-flow liquid- and gas-phase TOF devices has significantly improved TOF resolution, mass accuracy, and overall utility as a chromatographic detector. In this application, TLF corrects for a spatial distribution prior to analysis in an orthogonal, reflecting TOF analyzer.

6.2 Post-source Pulse Focusing

Although TLF provides enhanced resolution for a restricted mass range in linear systems suffering from U_o dispersion, under some conditions it may actually create defocusing of signals outside of the m/z range of enhancement. Because of the latter, several researchers suggested the use of post-source pulse focusing (PSPF) as a means of improving resolution.[84–89] In this approach, CIE is employed and ions are accelerated to enter a PSPF region. The PSPF region is either the last stage of an ion optic or a discrete assembly segregated from the source ion optic by a short free flight region. In a manner analogous to TLF, a discrete packet of ions enters the PSPF region with an inherent final velocity spread. At first, the PSPF region is acceleration field free. When focusing is desired, a high-voltage pulse is applied to create a post-source acceleration field. Differential acceleration proceeds with lagging ions receiving a greater degree of acceleration than their leading counterparts. The amplitude of the pulse potential is chosen so that lagging ions catch the leading ions at the system's detector.

The advantage of this approach, when compared to TLF, is that one can select an m/z range for focusing without defocusing other ions. When the focusing pulse is not applied, all other ions pass through the PSPF region without subsequent acceleration. However, because the focused ions have received a different degree of acceleration when compared to all other ions, they now have a different amount of kinetic energy. Consequently, these ions require a different calibration expression for m/z determination to that for unfocused ions. Higher order calibration expressions have been derived to determine the m/z of PSPF ions but these algorithms are not simultaneously applicable to unfocused ions.[88,89]

In the final analysis, PSPF effectively enhances resolution and mass accuracy for a limited m/z range in linear TOF devices. Because PSPF employs CIE, the advantages of uncoupling ion formation from acceleration are not realized. Furthermore, because PSPF acceleration occurs at an appreciable time after ion formation, a greater amount of metastable decay products are present at the time of differential acceleration, increasing the overall chemical noise. Because of these issues, the use of TLF over PSPF in linear geometries is generally favored.

6.3 Dynamic Pulse Focusing

Other attempts at extending the m/z focusing range to enhance mass accuracy and resolution have concentrated upon altering the nature of the ion extraction pulse. In classical TLF, a square-wave extraction pulse is applied to the energy-focusing region of an ion optic assembly. In this manner, a constant-focusing field is applied to ions of all m/z and some ions become focused but others are not. Dynamic pulse focusing takes advantage of the different residence times of ions of varying m/z within the energy-focusing ion optic region. These approaches apply a time-dependent extraction pulse in order to focus more appropriately the ions over a broader m/z range.

In 1974, Marable and Sanzone discussed the use of an impulse field focusing approach for which a time-dependent change in ion extraction field strength could be employed to extend the focusing range of an electron ionization/time-of-flight (EI/TOF) device out to $2600\,m/z$.[90] They later reduced this theory to practice in 1981.[91] In this device a two-step ion draw-out field was used. Although the mass range of resolution enhancement was increased, optimization was still shown to be m/z dependent.

Muga described a system that used a second-order time-dependent post-acceleration field in the second stage of a two-stage ion optic assembly. CIE acceleration was

performed in the first stage. This device showed unit mass resolution for EI/TOF analysis out to 4000 m/z.[92] As discussed previously, Whittal described the use of an exponentially decaying pulsed-ion extraction field to compensate for total energy disparities between low- and high-m/z ions. This system demonstrated improved mass accuracy when compared to square-wave TLF.[63]

Kovtoun introduced the concept of mass-correlated delayed extraction, where a time-varied correction field is added to a static PIE field so that every iso-mass group has its own history of energy correction while traveling through the acceleration region of an ion source.[93] This approach was applied to a MALDI/TOF device that demonstrated enhanced resolution for the simultaneous measurement of peptides in the range 900–4500 Da.[94]

Dynamic pulse focusing techniques are demonstrating significant promise to improve the ease of use of linear TLF devices, because the need to adjust TLF parameters as a function of m/z ultimately may be eliminated. These dynamic approaches impart an m/z-dependent final velocity and total kinetic energy profile. Consequently, system calibration must be performed using a higher order calibration expression that accounts for the change in total kinetic energy as a function of ion TOF. Indeed, this may be the preferred method of parallel ion extraction for future linear TOF platforms.

6.4 The Ion Mirror or Reflectron

In 1973, Mamyrin introduced the concept of the ion mirror or reflectron.[95] This device compensates for differences in U_o by making ions of different U_o travel proportionally different flight distances. Consider the ion-mirror-containing TOF system depicted in Figure 12. In this case, an ion source provides a space-focus object plane just outside of the ground aperture. At this point ions are spatially coherent but possess different final velocities because of varying initial U_o. All ions continue in free flight until they enter the uniform-retarding electrical fields of the mirror. In the reflectron, ion velocity slows and reaches zero at the point where the product of ion mirror potential and ion charge is equivalent to total ion energy. Ions are then turned about and accelerated out of the mirror to reach an exiting velocity that is the opposite of the entrance velocity.

In this case, an ion with additional velocity δ penetrates more deeply the reflectron field and travels a longer distance to the detection plane. This difference in total flight distance compensates for δ and isomass ions arrive at the detection plane in temporal coherence. Tang et al. provide a more thorough mathematical derivation of this system.[10]

Figure 13 depicts the two most common ion mirror geometries. The upper panel displays a coaxial reflectron

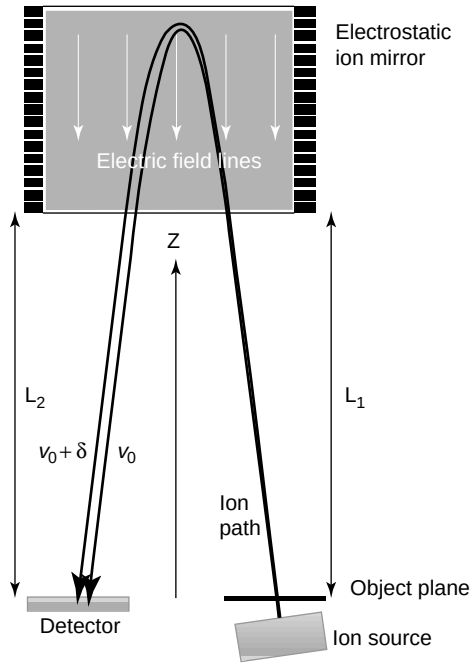

Figure 12 A single-stage electrostatic ion mirror or reflectron. See text for details. (Figure provided courtesy of Kenneth G. Standing and Werner Ens, University of Manitoba, Winnipeg, Manitoba, Canada.)

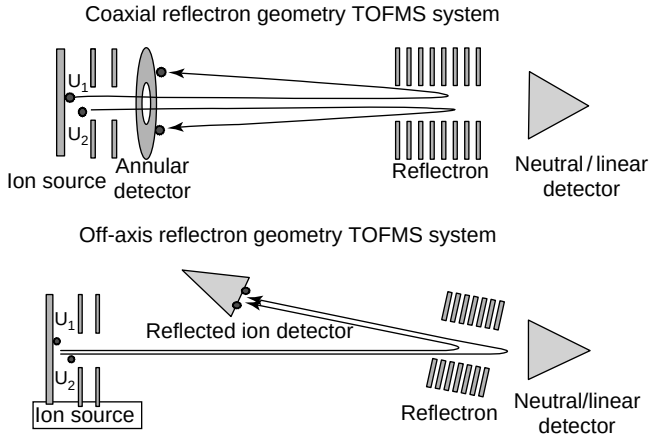

Figure 13 Illustration of coaxial and off-axis reflectron geometries. See text for details.

system. In this arrangement, two detectors are employed: a linear or neutral detector and a reflected-ion annular detector. Ions are emitted from the source through the opening of the annular detector. For linear mode operation, the reflectron is maintained at free flight potential and all ions strike the neutral/linear detector. When the reflectron is energized, reflected ions are directed toward the annular detector. The off-axis

geometry is shown in the lower panel. In this approach the same design may be employed for both detectors.

Over the years, reflectrons have been used in TOF analysis for ions formed from LDI,[96–98] MPI,[99] SIMS,[10,100,101] MALDI,[102] and ESI.[54] Although performing relatively well in focusing ions below 10 000 Da, most reflectrons do not improve resolution for heavier ions. This is primarily due to two factors: background gas collisions within the reflectron, and metastable decay of larger ions. As ion m/z increases, so does ion cross-sectional area, as well as the probability of collision with background gas. When large ions are decelerated to near-zero velocities in the reflectron, collisions with background gas create deleterious changes in ion position and velocity, manifesting as significant time spread at the detector.

Unimolecular decay of parent ions has been well established for a number of ionization schemes.[10,77,103] This decay results from bond schism that is a direct consequence of depositing excessive vibrational energy during ion desorption, desolvation, ionization, or extraction processes. The result is unimolecular or metastable decay, producing neutrals and product ions. During free flight, parent ions, product ions, and neutrals all have the same velocity. But because product ions and neutrals weigh less than parent ions, their total energy is less than that of their parent species. In this manner, the electrical field strength of the reflectron will not optimally focus all ions, and peaks exhibit significant fronting and asymmetry due to noncoherent early arrival of metastable products. This effect may be mitigated by the use of parabolic or curved-field reflectrons.[104,105] This concept will be discussed when considering tandem mass spectrometry (MS/MS) geometries, in section 14.

6.5 Higher Order Correction

As noted previously, TLF has demonstrated significant potential to improve the mass resolving power and accuracy for linear TOF devices by compensating for initial ion spatial, initial ion velocity, or initial ion formation time spread. First-order simultaneous correction of all three factors is not possible. In some TOF applications, such as MALDI, ions are ejected from a well-defined surface so the initial ion spatial spread is small. However, MALDI ions demonstrate significant initial velocity spread that is often exacerbated by collisions if ions are instantly accelerated by the application of a strong extraction field.[62] Although an ion mirror can provide first-order energy focusing, it cannot correct for time spread created during ion acceleration.[58] Consequently, the demonstrated resolution for systems combining MALDI, CIE and a reflectron is often less than that of a well-designed linear TLF device.

In the case of many continuous-flow interfaces, ions are formed with simultaneous initial velocity and spatial distributions. In this fashion, the simple application of either TLF or a reflection will not provide significant resolution enhancement. The desire to correct for both initial spatial and energy spreads has prompted many investigators to consider higher order correction approaches.

In 1974, Stein noted the possibility of both space and velocity focusing. Stein notes: ''If a conservative force field is space focusing to a given order with particles starting at initial points and arriving at a final point, this same force field becomes velocity focusing to the same order by starting particles at the final point, reflecting them at the initial points and collecting them at the point from which they started''. Stein provided several theoretical models that demonstrated spatial and velocity focusing through the use of equal-momentum acceleration, time-varying fields, two-dimensional ion motion, and a combination of conditions such as Wiley–McLaren TLF with the ion mirror of Mamyrin.[106]

In 1989, Cotter suggested a similar ''double-focusing'' system consisting of TLF and a reflectron. In addition to higher order correction, it was felt that the employment of delayed ion extraction would uncouple parent ion fragmentation from ion acceleration, thus providing structural information and eliminating the metastable ion-defocusing effects often seen in reflectron systems.[23]

Vestal et al. demonstrated significant resolution improvements in parallel-extraction MALDI by combining TLF with a reflectron. During the analysis of adrencorticotropic hormone clip 18–39 (monoisotopic m/z 2465.2), Vestal demonstrated a twofold improvement of resolution for the TLF/reflectron system when compared to a linear TLF configuration. The same system used to analyze bovine insulin (average $m/z = 5733.5$) successfully resolved the protein's isotopic envelope with a resultant resolution of 12 500. In this arrangement, TLF was used to create a first-order time focus near the initial point of free flight. Higher order energy focusing was provided by the reflectron.[74] In addition to reflectron focusing, the use of TLF in this double-focusing system also narrowed the initial ion velocity spread, because the collisions involved with immediate ion extraction were avoided. All of today's research-grade parallel-extraction MALDI/TOF devices utilize this double-focusing geometry.

TLF and reflectrons are combined in all modern orthogonal extraction devices. In this arrangement, TLF is used to create a spatial focus plane just outside of the accelerating ion optics (see Figure 14). Energy focusing is accomplished using an ion mirror. Detection and space-focus planes are arranged to be equidistant from the reflectron. In this manner, reflected ions arrive at the detection plane with both energy

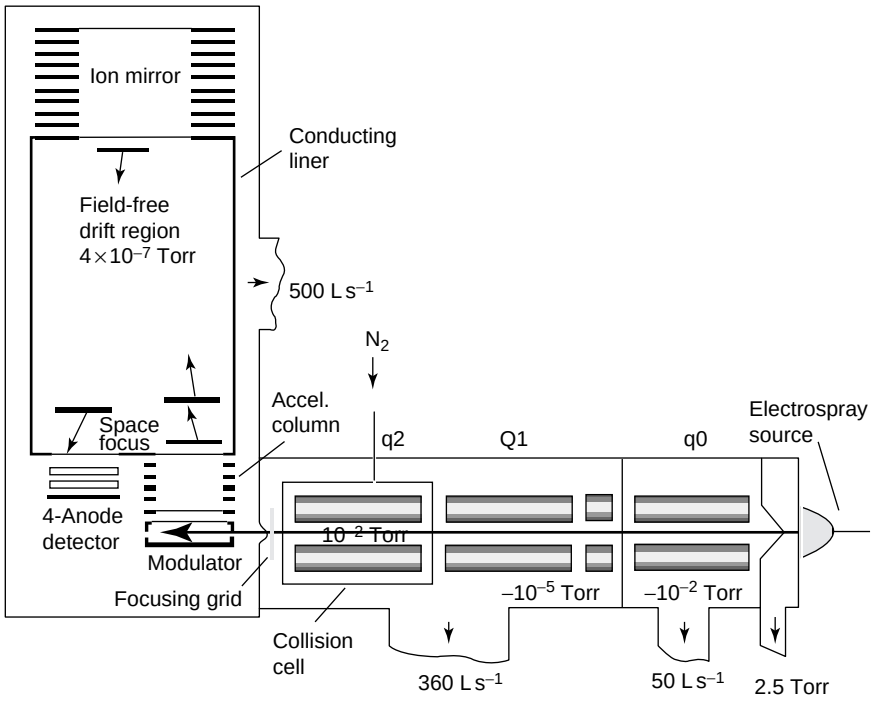

Figure 14 An orthogonal electrospray TOF device equipped with quadrupole ion guide providing collisional cooling (q0). Two other quadrupole regions are included for MS/MS purposes (Q1 and q2). TLF performed in the ion optic assembly corrects for initial spatial distribution, whereas the ion mirror corrects for velocity disparities. (Figure courtesy of Werner Ens and Kenneth G. Standing, University of Manitoba, Winnipeg, Manitoba, Canada.)

and spatial coherence. A similar device was used by Kurtchinksy to provide a deconvoluted mass spectrum of bovine insulin with isotopic resolution ($R = 10\,000$).[55] This geometry was later adapted to MALDI-generated ions.[58] Unlike the parallel-extraction double-focusing geometries suggested by Cotter and Vestal, the double-focusing orthogonal MALDI system does not require the adjustment of TLF parameters as a function of ion m/z. The latter is attributable to Wiley–McLaren space-focusing conditions that are independent of ion m/z. In this fashion, orthogonal MALDI ion acceleration represents an easier-to-use solution when compared to parallel devices.

7 ANALYTICAL SENSITIVITY

The metric of analytical sensitivity is often described in terms of some detection limit established at a minimal S/N. Mass spectral data are typically displayed as a two-dimensional plot of intensity vs m/z, where signal intensity is represented as total ion counts or some other arbitrary scale created to be proportional to total ion current for a given m/z domain. In either case, signal amplitude is taken to be the vertical displacement of the displayed peak as measured from the baseline.

Noise is often characterized in terms of its origin and frequency. In TOFMS, noise sources include electronic noise and chemical noise. For the most part, electronic noise such as thermal noise, environmental noise, shot noise, Johnson noise, and flicker noise, have been greatly reduced in today's TOF instruments. Undoubtedly, the greatest contribution to overall TOFMS noise is chemical noise. Chemical noise is created by the arrival of time-incoherent particles to the detector surface. These particles are energetic ions, neutrals, and, in the case of negative ion scans, electrons that arise from events of ion creation, collection, and acceleration. Processes such as metastable decay, fast ion and electron bombardment of ion optic elements, and poor ion focusing promote chemical noise creation.

Noise manifests as random signals of various frequencies. High-frequency noise is most troublesome, because many times its frequency closely matches that of the measured signals and, as such, adds constructively and destructively to the overall signal profile. Intermediate and low-frequency noise generally create annoying baseline displacement and drift, often complicating automated peak height measurements. Noise has typically been quantified in terms of maximum peak-to-peak (p–p) amplitude or root mean square (RMS) values for a given TOF period. A S/N is calculated by dividing the

total signal by the determined noise. Because signals are measured in terms of peak to baseline excursion, most investigators use the maximum p–p value near the time of signal detection for determining S/N. Overall S/N can be improved by summing a number of scans for a given measurement. Because true noise is random, summing n number of scans for a constant signal source improves the S/N by a factor of $n^{1/2}$.

Detection limits are established in terms of total sample mass or sample concentration. The mass detection limit of a system is generally defined as the minimum quantity of analyte that can be detected with a given S/N, whereas the concentration detection limit is taken to be the minimum concentration of sample that can be detected under the same circumstances. Most laboratories established these detection limits at S/N = 2–3. As will be explained later, overall analytical sensitivity is affected by a number of factors, including analyte m/z. However, generally, modern TOF devices demonstrate mass detection limits of the order of high attomole to low femtomole amounts of sample and concentration detection limits in the nanomolar scale.

7.1 Factors Affecting Analytical Sensitivity

Overall analytical sensitivity in TOFMS is a confounded product of sample ionization efficiency, ion collection efficiency, ion transmission efficiency, and ion detection efficiency. Sample ionization efficiency is dependent upon the ionization potential of the analyte as well as the efficiency of the ionization source. In some cases, ion suppression results from preferential ionization of other compounds within the sample matrix. These competing compounds could be other analytes present within a heterogeneous mixture or the necessary components of the sample matrix, such as buffers, ion-pairing agents, denaturants, chaotropic agents, or chromatographic mobile phases. Ionization source efficiency results from the combined effects of analyte desorption/desolvation and ionization. In terms of MALDI, this process is greatly driven by the geometry of the probed area, the applied laser fluence, and matrix efficiency. With respect to ESI, desolvation and ionization efficiency remain tightly coupled and are ultimately affected by flow rate. Improvements in this process have been made through source miniaturization and will be discussed later.[107]

After formation, ions need to be collected and accelerated into the free flight region. During this phase, ion collection efficiency becomes dependent upon source duty cycle and focusing parameters. Duty cycle has been discussed previously and will not be elaborated further here. Ion focusing should proceed in a manner that ultimately corrals all ions of a given m/z to a point of temporal coherence upon the detection plane. In some cases, it is necessary to trade off resolution for sensitivity, because ions with extreme radial position or velocities may only be collected at the expense of total TOF spread.

Ion transmission efficiency is dependent upon perturbations of free flight as well as overall radial focusing. If analyzer pressure is too high, frequent collisions with background gas ultimately create radial and axial ion displacement, causing some ions to miss the detector and temporal incoherence among detector incident ions. For a given pressure and free flight distance, the probability of background gas collision is dependent upon ion cross-sectional area. In this manner, ion transmission efficiency for a given charge would be inversely proportional to MW^2. Likewise, the presence of multiple charges often produces coulombic forces of repulsion within a given ion, causing molecular unfolding or extension that increases cross-sectional area. This is of particular concern for ESI/TOF measurements of large proteins.

Many ions formed have some inherent radial velocity that must be controlled to ensure detector incidence. This is generally accomplished through the use of radial focusing in the ion optics to create an appropriately sized ion image upon the detector surface. Another approach incorporates an ion particle guide to diminish free flight radial spread. An electrostatic ion guide used for this purpose is described by a number of groups.[37,71,81,108,109]

Ion detection efficiency is inherently linked to the probability of converting incident ions to electrons as well as the ability to provide subsequent electron gain. A number of factors come into play here, including fundamental properties of the detector and signal amplifiers. These are discussed in the next section.

7.2 Detectors for Time-of-flight Mass Spectrometry

Detectors for TOFMS are generally classified as electron multipliers or direct-measurement/charge-based detectors. Among the electron multiplier group, two distinctions are made: discrete dynode electron multipliers and continuous electron multipliers. Discrete dynode multipliers consist of a series of separate shaped plates or assemblies of plates connected together by a chain of resistors, usually of equal value (see Figure 15). In this manner, high voltage applied across the chain will create an equal voltage drop across each dynode. Potentials are established to drive secondary electrons down to the detector anode. This may be accomplished by raising the first dynode to some negative potential while maintaining the anode at virtual ground, or by grounding the first dynode and floating the anode to some positive potential. The latter is a preferred approach to circumvent the deleterious effects of accelerating metastable ion products.

An ion striking the first dynode will produce secondary electrons that are accelerated to the second dynode,

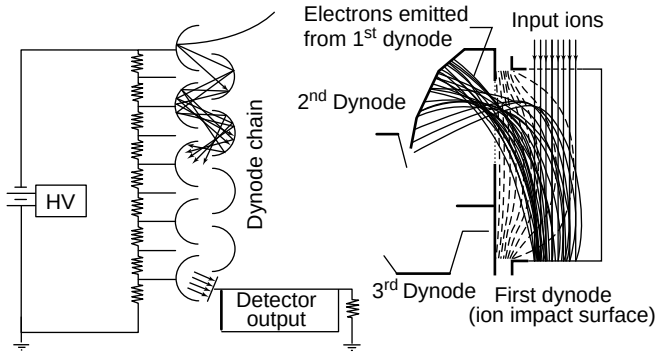

Figure 15 A discrete dynode electron multiplier detector. See text for details. (Figure courtesy of ETP Electron Multipliers, Ermington, Australia.)

and so on through the multiplier. Changing the applied voltage, usually over the range 1000–5000 V, varies the multiplier gain. Generally about 15–30 stages are used, with an interstage voltage drop of no more than 200 V. Interstage potential drops in excess of 200 V are likely to create field emission, thus elevating detector noise. Overall gain is usually of the order of about 10^8.

Continuous dynode electron multipliers consist of a single dynode surface. For TOFMS applications, microchannel plate (MCP) devices are most commonly used (see Figure 16). An MCP assembly is a lead silicate glass wafer with electrode etching containing a number of small channels. When suitably processed, these surfaces have secondary electron emission properties and are electrically resistive. About 60% of the MCP surface is comprised of these channels, and the remaining 40% provides structural integrity. A potential difference of about 800–1000 V is created across the MCP. Input ions strike the interior walls of a microchannel and release secondary electrons. An accelerating electrical field drives secondary electrons into the wall to create subsequent amplification. Output electrons are collected on the surface of the detector anode. Polarity considerations

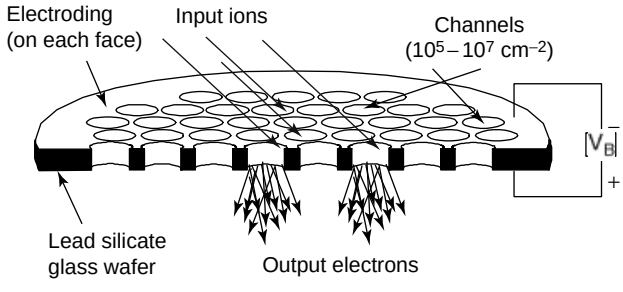

Figure 16 A MCP continuous dynode multiplier. See text for details. (Figure courtesy of Burle Electro-Optics, Lancaster, PA, USA; formerly Galileo Corporation.)

are similar to that of the electron multiplier, and varying the potential difference across the plate controls the gain. Maximum gain for this system is of the order of 10^4. Subsequent gain is achieved by stacking two or three of these MCPs in a chevron array. In this manner, the maximum gain ranges from 10^8 to 10^{12}.

As a general rule, MCP detectors have faster response times and broader analog bandwidths than discrete electron multipliers. However, compared to discrete electron multipliers, MCP multipliers have limited dynamic range. This is particularly true when a large ion load enters the first MCP of a stack, resulting in a saturating level of secondary electrons entering the second MCP. In this case, saturation is the direct result of charge depletion and acceleration field collapse. This effect creates a dead time period for which the second MCP remains refractory to electron gain and transmission. Because of the latter, many MCP-equipped TOF devices employ a means to minimize detector exposure to unwanted high ion currents. Ion deflection has been achieved through the use of electrostatic deflection lenses similar to that depicted in Figure 2.[20,21,110–112] Furthermore, pulsed application of a perturbing voltage to electrostatic ion guides has also demonstrated utility in deflecting unwanted ions.[81,109,113] Another means of minimizing detector saturation is achieved by decreasing the gain so that secondary electrons are not subsequently amplified.[25,114]

An extended dynamic range performance from MCP detectors has been achieved by linking the output of an MCP to a high-speed scintillator. In this manner, secondary electrons strike the scintillator surface to cause the emission of photons. Photon detection and subsequent gain are provided by a fast response photomultiplier tube. This arrangement avoids saturation, because there is no second MCP. Additionally, no anode is used, providing a facile means to float the output stage of the MCP without risk of deleterious anode discharge. This is particularly useful when applying post-acceleration fields for both positive and negative ions. A commercial embodiment of this device is available from Burle Electro-Optics Inc., (formerly Galileo Corporation), Lancaster, Pennsylvania, USA.

7.3 Electron Work Function, Post-acceleration, and Secondary Ion Generation

The first step of ion detection for all electron multiplier devices is ion to secondary electron conversion. Electron emission from conversion dynode surfaces is dependent upon the work function of that surface. Most dynode and MCP surfaces are coated with a low-work-function alloy or metal oxide so that a minimal amount of energy transfer can result in the emission of secondary electrons. Geno and Macfarlane described the probability of MCP

secondary electron emission as a function of the secondary electron coefficient, a term analogous to the electron work function.[115] The probability of secondary electron emission was defined as shown in Equation (32), where P is probability and γ is the secondary electron coefficient as described by Equation (33), where m is ion mass and v is ion final velocity:

$$P = 1 - e^{-\gamma} \tag{32}$$

$$\gamma = 2.58 \times 10^{-7} m[\exp(2.32 \times 10^{-4} v)] \tag{33}$$

For $\gamma = 5$, about 99% of all incident ions create secondary electron emissions. Geno and Macfarlane noted that, for a given velocity, γ increases with increasing MW, and for a given MW, γ increases with increasing velocity. Consequently, for a given amount of ion energy, the probability of secondary electron conversion and overall detection sensitivity is inversely proportional to ion MW. This relationship predicts that about 100% ion to secondary electron conversion efficiency is achieved for human insulin (5807.65 Da) at about 35 keV energy, whereas more than 65 keV of energy is needed for near-optimal conversion of bovine IgG (MW ~ 150 kDa).

The acceleration of ions to energies greater than 30 keV is difficult if one relies solely upon ion optic acceleration. Problems with electrical stability often become limiting here. Additional ion energy is often provided by the use of a post-acceleration region just prior to the detector's conversion surface. In this fashion, an additional 10–20 keV of energy may be supplied. For cations, a negative voltage post-acceleration field is employed, whereas a positive voltage post-acceleration field is used for anion detection. One shortcoming of this approach is the subsequent creation of time spread among free-flight metastable ion products.[80]

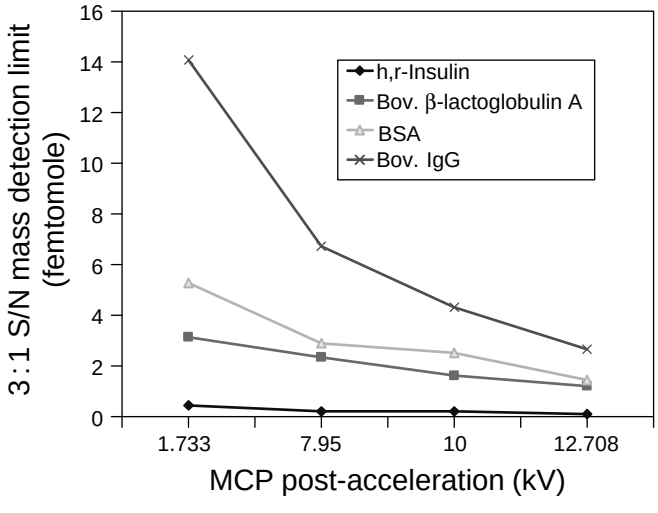

Figure 17 The 3:1 S/N mass detection limit for human insulin (MW = 5807.65 Da), bovine β-lactoglobulin A (MW = 18 363.3 Da), bovine serum albumin (BSA, MW = 66 430.2 Da), and bovine IgG (MW ~ 148 kDa). MALDI analysis was performed using a positive ion scan, 25-kV ion acceleration, and an MCP detector. (Unpublished data from Scot R. Weinberger.)

The overall effect of post-acceleration upon the 3 : 1 S/N mass detection limit for MALDI ions of various MWs is demonstrated in Figure 17. Here all ions were accelerated to 25 keV and then subsequently post-accelerated with up to an additional 12.7 keV. As can be seen, modest improvements of the mass detection limit for human insulin and bovine β-lactoglobulin A is demonstrated, but appreciable detection sensitivity enhancement is shown for BSA and bovine IgG. These results are consistent with the theory of Geno and Macfarlane.

The use of a secondary ion generator (SIG) has shown utility in improving the detection sensitivity for large ions.[110] Such elements are generally solid surfaces or moderate transmission grids that provide a means to generate secondary ions as a direct consequence of primary ion impact. The dynamics involved in this process have been studied by a number of researchers.[110,116–119] In terms of direct ion generation, it is generally believed that the impact of primary ions upon SIG surfaces results in the sputtering of surface contaminant ions such as alkali metals, H^+, H^-, and other adsorbed vacuum-system contaminants.[117] Furthermore, secondary electron emission from these same contaminants occurs in the presence of a negative potential.[119] I have discovered that a negatively biased copper mesh provides improved generation of secondary ions when compared to ground potential copper grids.

Because the MW of secondary ions is significantly less than that of the incident ions, subsequent post-acceleration of secondary products typically impart velocities that are greater than that of the original projectiles. In this fashion, the secondary ion conversion efficiency is greater than that for parent ions. In fact, even though the relative abundance of these secondary products is less than that of the parent ion, in many cases the amplitude of the measured secondary ion signal exceeds that of the high-molecular-weight parent ions. Figure 18 depicts a plot of the SIG signal normalized to the parent ion signal for a copper-mesh SIG/MCP detector system. As analyte MW increases, the ratio of the SIG-to-parent signal increases.

The disparate final velocities of post-accelerated secondary and primary ions complicate the use of the SIG signal to enhance the detection of high-molecular-weight ions. In many cases, subsequent post-acceleration of both populations results in attendant time spread and possible resolution of two discrete ion signals. I have demonstrated that these ion populations can be made temporally coherent to the detector by the use of a differential acceleration grid (DAG) positioned between

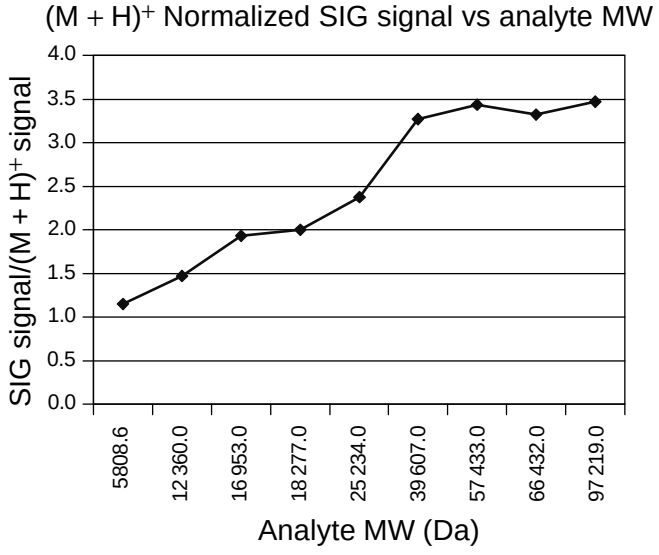

Figure 18 SIG signal as a function of analyte MW. The relative ratio of SIG ion signal normalized to parent ion signal is plotted against parent ion MW for a number of proteins ranging from 5800 to 97 000 Da. The normalized SIG signal increases with increasing projectile MW. Signals were generated during MALDI analysis using 25-kV acceleration potential, 10-kV post-acceleration potential, 60% transmission copper grid SIG, and MCP detection surface. (Unpublished data from Scot R. Weinberger.)

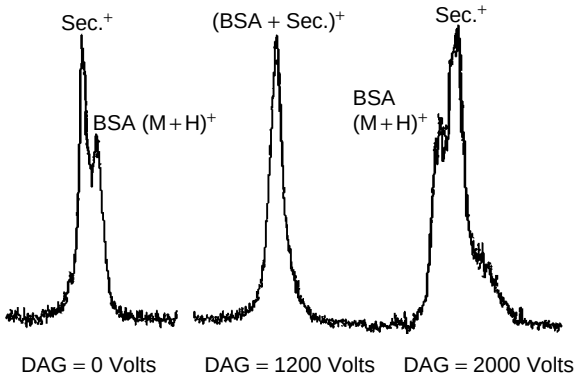

Figure 19 The effect of DAG potential upon BSA secondary and parent ion detector arrival times. See text for details. (Unpublished data from Scot R. Weinberger.)

the SIG and the detector surface. The DAG potential is controlled so that secondary ions are first softly accelerated, allowing primary ions to pass them. Upon moving through the DAG, secondary ions are accelerated to catch up with parent ions at the detector surface. Figure 19 demonstrates the effect of DAG potential upon secondary and parent ion profiles for MALDI-generated BSA cations. At zero DAG potential, secondary ions arrive at the detector before parent ions, and two discrete ion signals are noted. When the DAG is set to 1200 V, secondary ion arrival to the detector is slowed to be coincident with parent ion arrival. A single signal with improved resolution and S/N is achieved. Overfocusing is possible by setting the DAG to potentials greater than 1200 V. In this case, BSA parent ions precede the arrival of secondary ions. I have combined the above-noted SIG scheme with a 15-kV post-acceleration MCP detector assembly that has provided the ability to detect 700 amol of IgG (MW ~ 148 kDa) as well as the trimer [$(3M + H)^+$, MW ~ 444 kDa] and tetramer [$(4M + H)^+$, MW ~ 592 kDa] ions for 5 fmol of IgG (see Figure 20).

7.4 Direct Measurement/Charge-based Detectors

Because of the above-described molecular-weight-dependent loss of analytical sensitivity demonstrated by electron multipliers, many laboratories have been investigating the use of charged-based detectors to aid in the analysis of very large ions.[7–9,120,121] Bahr described a Faraday cup type of detector creating a 2:1 S/N for the detection of 18 000 elementary charges. The overall efficiency of this device was better than most electron multipliers for ions greater than 30 kDa.[120] Park and Callahan described an inductive detector that only used about 30% of all incident ions.[121] The remaining 70% of the ion load are available for subsequent experiments. The overall resolution demonstrated by this device probably makes it impractical for most TOF applications. Benner et al. described an inductive tube detector with a detection limit of about 150 elementary charges. This ultrasensitive device has been applied successfully to the measurement of megabase DNA.[7–9] To date, no commercial devices employ charge-based detector systems.

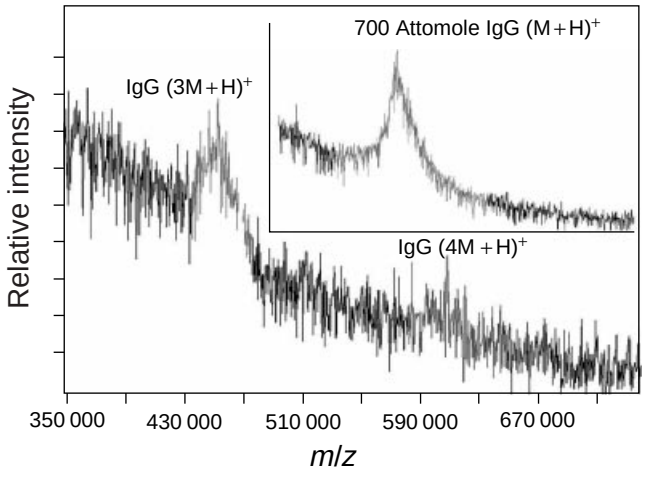

Figure 20 MALDI/TOF detection of 700 amol of IgG and trimer and tetramer IgG ions for a 5-fmol sample load. See text for details. (Unpublished data from Scot R. Weinberger.)

8 LIQUID INTERFACES FOR TIME-OF-FLIGHT MASS SPECTROMETERS

This section covers liquid interfaces coupled to TOFMS instruments, specifically ESI and the liquid separation methods coupled to it, including HPLC, CE, and CEC, as well as MALDI coupled with on-line liquid separation methods. ESI, being a solution-phase analysis method, is more readily coupled with liquid-phase separation techniques, hence the larger volume of literature on the topic. MALDI is more typically done using solid-phase matrices, such as 2,5-dihydroxybenzoic acid; coupling a liquid-phase separation to MALDI is thus inherently more difficult.

8.1 Electrospray Ionization

The challenge presented in performing ESI on a TOF instrument is in coupling a continuous ionization source to a mass spectrometer that is inherently pulsed. The TOF instrument requires precise measurement of the time between the start pulse and arrival of an ion packet at the detector; thus, ions cannot be introduced continuously into the analyzer region of the mass spectrometer but can only be introduced at specific intervals. The first widely known reported coupling of ESI to a TOF instrument was in 1992 by Boyle and Whitehouse.[122] They overcame the problem of continuous ion introduction by developing a TOF instrument where the ions are introduced to the spectrometer between the ion repeller and extractor lenses in a direction that is orthogonal to the field-free flight region (see Figure 21). The ions are then pulsed into the field-free flight region to allow for their velocity-dependent separation and focusing by a two-stage reflectron.[95] Simultaneously, the region between the repeller and extractor lenses, usually referred to as the ion storage region, is again allowed to fill ready for the next repeller/extractor pulse, giving a maximum duty cycle of approximately 25%. This instrument had a mass resolution of 1000 $m/\mathrm{d}m$.

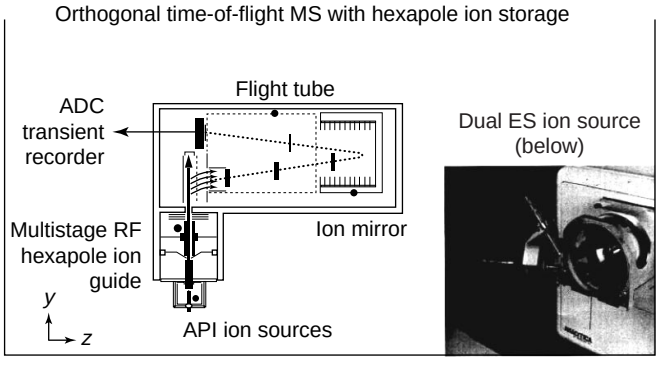

Figure 21 Orthogonal TOF device with hexapole ion storage system. (Figure courtesy of Analytica, Branford, CT, USA.)

Meanwhile, the independent development of orthogonal TOF with an ESI source had taken place at the Russian Academy of Sciences by Dodonov et al. in 1987.[51–53] This instrument was also equipped with a two-stage reflectron and had a mass resolution of 1000 $m/\mathrm{d}m$.

A recent review of orthogonal injection of ions into a TOF instrument was completed by Chernushevich et al. in July 1999.[4] They review the advantages of the orthogonal TOF configuration for continuous ion sources. Briefly, the TOF instrument can tolerate a relatively large spread in the velocity or spatial distribution of ions in the plane perpendicular to the spectrometer axis (y-axis) and typically there is a small velocity spread along the spectrometer axis (z-axis) such that high resolution can be achieved more readily.

An alternative method of coupling ESI to a TOF instrument was reported in 1993 by Lubman et al.[123] Briefly, instead of injecting ions orthogonally the authors injected ions into a QIT for temporary storage of the ions before analysis by the TOF instrument.

Since the first reports of coupling ESI to a TOF instrument, there have been several reports following the same basic configuration.[54,124–126] Dodonov et al. had improved their instrument by 1994 such that it had a mass resolution limited only by their data collection system to 2000 $m/\mathrm{d}m$.[124] That same year Dodonov and Laiko gave a theoretical treatment of the orthogonal TOF instrument equipped with a two-stage reflectron, complete with equations describing the expected resolution, spectral line shape, the effect of grid wires, and the effect of initial ion velocity.[125] Also that year, Verentchikov et al. reported the development of an ESI/TOF instrument based on the Russian design that had a mass resolution of 5000 $m/\mathrm{d}m$ for peptides.[54] However, the resolution decreased with mass for all of the early orthogonal TOF instruments.

A dramatic improvement in the performance of ESI orthogonal TOF instruments came with the addition of collisional cooling or focusing developed by Douglas and French for quadrupole mass spectrometers.[127] Collisional focusing is achieved through the use of an RF quadrupole or hexapole lens at elevated pressure, where collisions with surrounding gas reduce ion energies close to thermal values.[55] Without collisional focusing, ions introduced from atmospheric pressure into a vacuum are essentially a supersonic beam where all ions travel at approximately the same velocity. As the ions are pulsed out of the ion storage region, they continue to travel orthogonally (y-component of velocity) with the same velocity as they had when entering the high-vacuum region. To strike the detector, the correct ratio of velocities between ions traveling along the flight path

of the instrument (z-component of velocity) to the y-component of velocity must be maintained. Because ions pulsed from the source have a z-component of velocity that is inversely proportional to $(m/z)^{1/2}$, for all ions to strike the detector the y-component of velocity must also be inversely proportional to $(m/z)^{1/2}$. Collisional cooling reduces ion velocity to near-thermal values, after which they are electrostatically re-accelerated to give them the required energy to strike the detector, independent of mass.[4]

For the analysis of proteins, the vacuum pressure inside the high-vacuum region of the orthogonal TOF instrument is an important experimental parameter. Chernushevich et al. reported in 1996 that resolution of protein mass spectra is dependent upon the number of collisions between the protein ions and the residual background gas in the mass spectrometer, higher pressures giving more poorly resolved spectra.[128]

Recently, commercial versions of the orthogonal TOF instrument equipped with an ESI interface have become available. The current manufacturers include PE Biosystems (Framingham, MA), makers of the Mariner Biospectrometry Workstation with a maximum pulse repetition rate of 11 kHz, Micromass Inc. (Manchester, UK), makers of the LCT with a maximum pulse repetition rate of 20 kHz, Analytica of Branford Inc. (Branford, CT), makers of the Enterprise API-TOF/MS (the instrument in Figure 21) with a maximum pulse repetition rate of 20 kHz, and Bruker Daltonics (Bremen, Germany), makers of the Bio TOF. The Mariner can be purchased with a standard ESI interface or the PE Sciex (Concord, Canada) ESI interface (for higher throughput applications) or the Nano-link Spray Chamber (nanospray) for high sensitivity/limited sample applications. The Micromass ESI TOF can be purchased with a standard Z-spray ESI interface or an interface with up to four ESI sprayers, known as the MUX Technology interface. Potentially, four HPLC systems can be connected to one orthogonal TOF instrument or alternatively two nozzles may be used for internal standard introduction for exact mass measurements. The Enterprise API-TOF/MS can be purchased with a standard ESI interface ($1-1000\,\mu L\,min^{-1}$), an atmospheric pressure chemical ionization (APCI) interface ($100-2000\,\mu L\,min^{-1}$), a low-flow ESI interface ($50-1000\,nL\,min^{-1}$) or a dual ESI source. This source allows one to perform continuous internal calibrations for exact mass measurements without mixing for both HPLC or flow-injection analysis. The Bio TOF is intended for the study of noncovalent biological complexes and protein conformations. The nanoelectrospray source, developed by Wilm and Mann,[107] has been incorporated onto the orthogonal quadrupole TOF instrument for peptide sequencing.[129] Commercial versions of the quadrupole TOF instrument are available from Micromass (Manchester, UK), makers of the Q-TOF, and from PE-Sciex (Concord, Canada), makers of the Q-STAR.

8.2 Noncovalent Biological Complexes

The unlimited mass range of the TOF instrument makes it well suited for the study of large noncovalent complexes that would be beyond the mass range of most instruments. In 1994, Tang et al. reported the study of protein–protein noncovalent complexes on their recently constructed ESI orthogonal TOF instrument.[130] On the same instrument, Fitzgerald et al. have shown large noncovalent complexes with m/z beyond 4000.[131] Samples were introduced in 5 mM ammonium bicarbonate buffer at pH 7.5. The m/z distribution of the ions shifted to higher m/z ratios (lower charge states) at physiological pH and the protein hexamer was only observed with gentle declustering potentials (the potential difference between the atmospheric pressure plate and the first skimmer in the instrument); as the declustering potential was increased, the hexamer dissociated into pentamers, tetramers, and monomers. More recently, the same instrument was used to study protein–protein and protein–ligand equilibria.[132,133] Equilibrium was observed between the *Escherichia coli* citrate synthase dimer and hexamer that depends upon the concentration of nicotinamide adenine dinucleotide (reduced form) (NADH). Hexamer complexes with a mass of 287 322 U were observed at m/z up to 10 000. A protein–ligand–DNA complex was also studied on the same instrument for the tryptophan repressor.[134]

8.3 High-performance Liquid Chromatography/ Electrospray Ionization/Time-of-flight Mass Spectrometry

With the addition of the ESI continuous ion source to the TOF instrument, it was a natural transition to couple liquid separation methods to the TOF instrument through the ESI interface. As discussed previously, manufacturers offer ESI/TOF instruments capable of performing HPLC separations with or without HPLC control. A number of research groups have published reports combining HPLC separations with an orthogonal ESI/TOF instrument. Banks and Gulcicek developed an HPLC analysis method based upon separation on nonporous silica.[135] They achieved a complete separation of the peptides from a myoglobin tryptic digest in 3.5 min. Chromatographic peak widths were of the order of 0.5–3 s, necessitating data collection at a rate of 16 spectra per second. Here the TOF instrument is especially advantageous because a complete spectrum is acquired with each pulse. Of course, the pulse rate is actually much

higher than 16 Hz so each spectrum was actually a sum of 512 spectra. Banks also showed the gain in sensitivity achieved by using smaller internal diameter LC columns.[136] A 0.1-mm internal diameter column was 428 times more sensitive than a 2.1 mm internal diameter column packed with the same material. Nielen and Buijtenhuijs have used an orthogonal ESI/TOF instrument coupled to size-exclusion chromatography, gradient polymer elution chromatography, and LC at the critical point of adsorption for polymer analysis.[137] They point to the higher duty cycle and extended mass range of the orthogonal TOF over the quadrupole MS instrument as critical to the analysis of larger oligomers and the detection of minor impurities.

The excellent mass accuracy of the orthogonal TOF spectrometer makes it possible to achieve continuous exact mass measurements from LC-separated compounds. Hogenboom et al. showed that with the post-column addition of a single internal standard, the exact mass of several LC-separated pesticides could be determined.[138] With the use of a pesticide database, unambiguous identification of the pesticides was achieved in most cases, the only exceptions being isomeric compounds.

The quadrupole orthogonal TOFMS was recently investigated for its quantitative aspects as a high-performance liquid chromatography/tandem mass spectrometry (HPLC/MS/MS) instrument.[139] They found it to have an impressive dynamic range of over four orders of magnitude, with better sensitivity than an on-line fluorescence detector for the compound 3,4-ethylenedioxymethamphetamine.

8.4 Capillary Electrophoresis/Capillary Electrochromatography Electrospray Ionization/Time-of-flight Mass Spectrometry

The orthogonal TOF instrument is especially suited to liquid separation techniques that produce narrow chromatographic peaks (one to a few seconds in width). The TOF instrument is capable of collecting an entire mass spectrum with each pulse and modern data systems are capable of collecting several spectra per second, making the TOF more compatible with CE separations than the more widely used quadrupole mass spectrometer. CE interfaces for ESI/TOFMS was reviewed in 1997 by Banks.[140] Several approaches can be used to interface CE to the ESI/TOF instrument. In 1994, Fang et al. reported a direct coupling of CE to a commercial ESI interface without the use of a sheath flow buffer.[126] The tip of the capillary was inserted through the stainless-steel ESI nebulizer assembly and held in electrical contact using a gold wire. The ESI nebulizer assembly was held at ground potential, requiring the maintenance of the ESI counter-electrodes and ESI interface at negative potentials of 3–4 kV. A typical problem with CE coupled to ESI is the limited choice in CE running buffer. This is especially problematic with the sheathless CE interface design, because the CE running buffer must be compatible with ESI. Feng et al. used 50 mM acetic acid in 50 : 50 methanol–water.[126] It is more typical for capillary electrophoresis/electrospray ionization (CE/ESI) interfaces to use a sheath flow. In this configuration, the CE capillary is inserted through the ESI nebulizer assembly that delivers a liquid sheath flow (in addition to the gas sheath flow to assist in desolvation) that mixes with the CE column effluent. The sheath liquid is designed to allow efficient ESI so that CE buffers that are less compatible with ESI can be used. The sheath flow liquid is typically introduced at a few microliters per minute, whereas the CE effluent is typically introduced at a rate ten times less than the sheath liquid. Dilution of the CE effluent and thus the possible decrease in the sensitivity are natural consequences of the sheath flow design. However, this must be weighed against the decrease in CE separation efficiency that would otherwise result if the CE buffer were compromised.

Muddiman et al. have developed a mathematical correction method for separating analyte signal from background noise when reconstructing the electropherogram in capillary electrophoresis/electrospray ionization/time-of-flight mass spectrometry (CE/ESI/TOFMS).[141] Their method, known as sequential paired covariance, can provide better than a 100-fold increase in the S/N of the electropherogram over simple integration of the total ion current. They used a sheathless interface design and

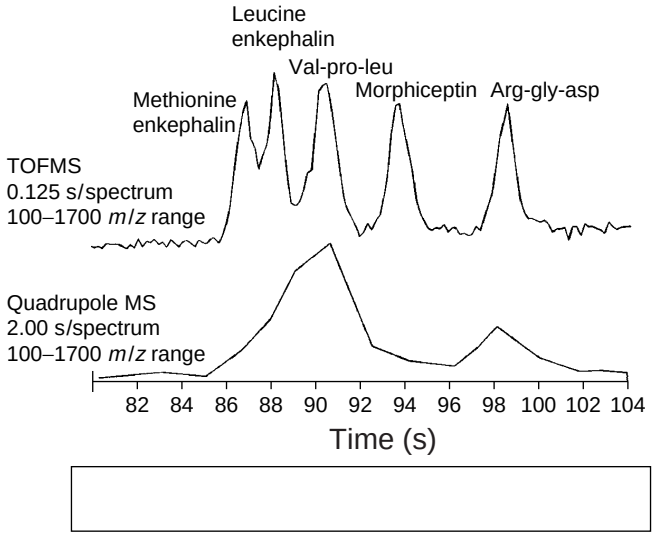

Figure 22 Comparison of the total ion currents from a peptide separation (amine-coated column) using TOF and quadrupole instruments. (Reprinted with permission.)

demonstrated the CE separation and detection of peptide standards.

Banks and Dresch have shown that CE-separated peptides and proteins that produce peaks 1–2 s in width are easily analyzed by an orthogonal TOF instrument, whereas peak detection suffered on their quadrupole mass spectrometer (see Figure 22).[142] Eight spectra per second were collected (each a sum of 1024 spectra) on the TOF instrument compared to a maximum of one spectrum every 2 s on the quadrupole. They used a sheath flow design and were able to analyze leucine enkephalin with a limit of detection of 8 fmol. Lazar et al. developed a CE/ESI interface for an orthogonal TOF instrument.[143] They used a sheath flow at higher flow rates and a sheathless design for flow rates less than 300 nL min^{-1} (microspray), maintaining ~75% separation efficiency when compared to ultraviolet detection. In 1998, Lazar et al. reported an improved version of this interface with a limit of detection of 1–10 fmol for CE/ESI/TOFMS.[144]

Cao and Moini have used CE/ESI interfaced to an orthogonal TOF instrument for the analysis of proteins followed by the analysis of the resulting tryptic peptides from these proteins for protein identification through database searching.[145,146] McComb et al. developed a sheathless CE/ESI interface using a gold-coated capillary for the separation of proteins and peptides.[147] Peak widths were typically 3–5 s for peptides with an S/N of 250–400 at femtomole sensitivity. Barnridge et al. developed chromium–gold-coated capillaries for use in a sheathless CE/ESI interface that had a lifetime of approximately 100 h.[148] Stable spray was observed from 30 nL to 1 μL min^{-1} with a limit of detection of 68 amol for MRF (methionine-arginine-phenylalanine)-enkephalin using the instrument and data system developed by Lazar et al.[149]

CEC combines CE with packed-column LC. Fast separations and narrow peak widths are typical with this technique. Choudhary et al. demonstrated the CEC separation of phenylthiohydantoin-amino acids using an orthogonal TOF instrument for detection through a sheath-flow ESI source.[150] They used a gradient elution of a packed capillary column and a CE instrument to apply high voltage along the length of the column. The separation of twelve phenylthiohydantoin-amino acids was completed in 10 min, with all components resolved in a 75 μm × 15 cm column.

8.5 Matrix-assisted Laser Desorption-ionization

Unlike the ESI sources, there is currently no commercially available source for coupling a liquid separation method directly to a MALDI/TOF instrument. A recent review of this area has been completed by Murray in 1997.[151] Off-line coupling of a liquid separation to MALDI does not require any special modification of the MALDI instrument and will not be discussed here. Two major on-line coupling methods are found in the literature. One method developed by Li et al. known as continuous-flow MALDI,[152] is similar to continuous-flow fast atom bombardment.[153] The other method developed by Murray et al., known as aerosol MALDI,[154] is an extension of aerosol interfaces previously developed for CI.[155]

In continuous-flow MALDI a liquid matrix, 3-nitrobenzyl alcohol, is mixed with methanol, ethylene glycol, and the analyte (in 0.1% trifluoroacetic acid) and introduced to the mass spectrometer through a 75-μm fused-silica capillary. The capillary was butted against a stainless-steel frit that was wrapped with filter paper to adsorb excess matrix and solvent. Desorption took place directly from the surface of the frit using an ultraviolet laser (Nd : YAG, 266 nm). This design was later changed to a fritless probe and coupled with an on-line HPLC separation.[156] Here the matrix was added post-column through a mixing tee and introduced through the flow probe to the mass spectrometer. In both reports, the MALDI probe was introduced to the spectrometer orthogonal to the instrument axis. The mass spectral resolution was unimpressive, typically being of the order of 10–20 m/dm. A later report from Whittal et al. explored the use of a flow probe introduced directly through the center of the repeller plate of the MALDI instrument.[157] In addition, the use of TLF or delayed extraction was also explored for improvement of the resolution. With TLF the resolution was improved to 2000 m/dm for peptides compared to ~200 m/dm using continuous extraction.

Aerosol MALDI is performed by directly spraying a matrix (4-nitroaniline)/analyte solution into the source region of a TOFMS. The aerosol particles are dried by passing the aerosol through a heated tube and then the dried particles are irradiated with an ultraviolet laser (Nd : YAG, 355 nm). Using a linear TOF instrument the aerosol MALDI apparatus has a mass resolution of ~6 m/dm for bovine insulin.[158] For combination with LC separation, the matrix solution was added post-column via a mixing tee and the mixed solution was introduced through the aerosol MALDI interface.[159] Improved mass resolution, to ~300 m/dm for peptides was demonstrated by Fei et al. for aerosol MALDI using a reflectron instrument and 2,5-dihydroxybenzoic acid as matrix.[160,161] Fei and Murray also demonstrated the coupling of aerosol MALDI to gel permeation chromatography for the separation and characterization of polyethylene glycol and polypropylene glycol.[162]

The on-line MALDI methods are far from routine and cannot compete with on-line CE or HPLC/ESI/TOFMS in terms of sensitivity, resolution, and ease of use at the

present time. Off-line separation combined with MALDI is a routine method of interfacing liquid separations to MALDI because no special interface is necessary.

9 TIME-OF-FLIGHT ANALYZERS WITH GAS-PHASE ION SOURCES

Compared to other forms of MS, the primary advantages of combining TOF detection with gas-phase analysis involve high-speed mass scanning, accurate mass assignment, and extended mass range. High-speed full-scan data acquisition facilitates analysis of transient processes and in particular opens up the possibilities for fast chromatography/MS. The possibility of accurate mass determination compared to quadrupole and QIT instruments extends the analytical power of empirical MW determinations. In the case of gas-phase sample introduction into the ion source, the higher mass range afforded by the TOF instrument is of no great benefit because the mass range is ultimately limited by sample volatility.

TOF with gas-phase ion sources has found broad application in molecular physics and chemistry due to its unique suitability for transient analysis. For example, a large amount of work has been done on laser ionization of gas-phase samples (the latter may be produced by laser ablation, laser desorption, arc, etc.),[46,99] photoemission EI,[163] CI (e.g. by metastable atoms),[164] and even gas-phase cationization of peptides.[165] Although this research allowed significant advance in these important fields, so far it has not resulted in a widely used analytical technique for gas-phase analysis.

From the other side, such a technique was born when TOF had been coupled to GC.[166] Since then the main area of analytical activity has become GC/TOFMS. Therefore, this section will concentrate mainly on various aspects of GC/TOFMS. Owing to the vast amount of library spectral information and method development that has been done using EI, GC/TOFMS is also primarily based around the use of EI ion sources. As both an illustration of an application where the TOF instrument is used to analyze transient processes and an illustration of a relatively new gas-phase ion source, bacteria identification using pyrolysis and MAB will also be discussed.[164,167]

9.1 Interfacing Continuous Ion Sources with the Time-of-flight Analyzer

9.1.1 Duty Cycle and Direct Current Ion Storage

The TOF analyzer has a high duty cycle when coupled to pulsed ion sources, because the source is not pulsed again until the ions from the preceding pulse have all arrived at the detector. Sample is therefore not wasted during the analysis period. To obtain high duty cycles with continuous ion beam sources some form of storage of the ions from the continuous ion beam is required during the analysis period. The most common approaches are:

- quasi-static ion storage (e.g. in the potential well of space charge),[3,95,168–171]
- ion storage in RF fields,[172,173]
- orthogonal acceleration.[174–178]

Wollnik et al.[168] have reused a method of creating the ions within a trapping electric field, which stores them prior to pulsing into the TOF mass analyzer.[3,95] This source has also been used in work by Enke et al.,[169,170] and is used in the Pegasus™ GC/TOFMS instrument (Leco Instruments, MI, USA).[171]

In a static-storage ion source, ions are created in an electrostatic potential well. For example, the potential well for positive ions can be created by two plates held at the same potential with an intermediate grid of a lower potential, or the attractive potential of the electron beam itself can also perform the function of the grid. Ions formed in the region between the two plates will have thermal energy and a potential energy determined by the potential at the position at which they were formed. Because the ions are formed in the potential well they will be trapped, and will oscillate to some extent between the two trapping plates. The trapping plate nearest the analyzer has a grid to allow transmission of ions out of the source once the voltages are pulsed to empty the source into the mass analyzer.

The TOF aberrations caused by the spatial spread of ions between the trapping plates can be corrected by a suitable arrangement of electric field slopes in the way described by Wiley and McLaren.[3] During an analysis cycle the stored ions are ejected from the source by pulsing the voltages on one or both of the trapping plates to form acceleration fields of the required slopes for ion TOF space focusing.

A disadvantage of this form of ion storage is that the oscillations of the ions between the trapping plates mean that there is also a velocity distribution as well as a spatial distribution of ions. This leads to further time aberrations, which reduce the mass resolution of the instrument. In particular, there is a significant effect of ''turn-around time'' where ions travelling away from the analyzer at the time of ejection from the ion source have to stop and accelerate back towards the analyzer.[3] Also, in GC/TOFMS applications ion loss rate from the potential well is adversely affected by the presence of carrier gas.[179] Another important disadvantage of this source is grid contamination within the source.

9.1.2 *Radiofrequency-storage Electron Ionization Sources*

Sample utilization efficiency could be improved also by using a QIT,[173] or a simpler cylindrical ion trap as the storage device.[179] In principle, this configuration also affords MS/MS experiments for molecular structure elucidation. Compromised mass resolution of the TOF analyzer, increased complexity of electronics, and low space-charge capacity of RF traps have impeded wide use of this type of source so far.

9.1.3 *Orthogonal Acceleration Time-of-flight Analyzers*

The most common method of coupling TOF with a continuous beam ion source is orthogonal acceleration. The use of orthogonal acceleration in TOF mass analyzers has meant that TOF can be coupled with any type of continuous ion source and still maintain a reasonably high duty cycle.[174–178] The use of orthogonal extraction helps to reduce the effects of ''turn-around time'' and initial velocity distribution in general, thereby improving the mass resolution of the instrument.

In orthogonal acceleration the continuous beam from the ion source is collimated and enters a field-free region between two plates at low energy (usually several tens of volts). Collimation of the beam reduces the component of energy orthogonal to the input direction of the beam. Ions are injected into the mass analyzer by pulsing voltages on one or both of the plates defining the field-free region to produce an acceleration field. As with the storage source, the magnitudes of the acceleration fields are chosen to provide TOF space focusing of the beam width. The direction of extraction and timing is orthogonal to the direction of input of the ion beam, i.e. in the direction where the energy spread has been minimized via collimation of the input ion beam.

The ion packets injected into the mass analyzer are accelerated usually to several kilovolts of energy. Orthogonal acceleration provides ion storage by virtue of slow filling of the extraction region due to low input energy and relatively fast analysis time after acceleration to several kilovolts. The duty cycle depends on the relative lengths of fill time to analysis time. Analysis time is dependent on the mass range to be analyzed and on the size of the instrument.

9.1.4 *Orthogonal Extraction Time-of-flight Analyzer with Electron Ionization Source*

The EI sources used in conjunction with orthogonal acceleration TOF analyzers tend be similar to EI sources used on quadrupole or magnetic sector instruments. The EI source is usually located in a differentially pumped region adjacent to the orthogonal acceleration module.

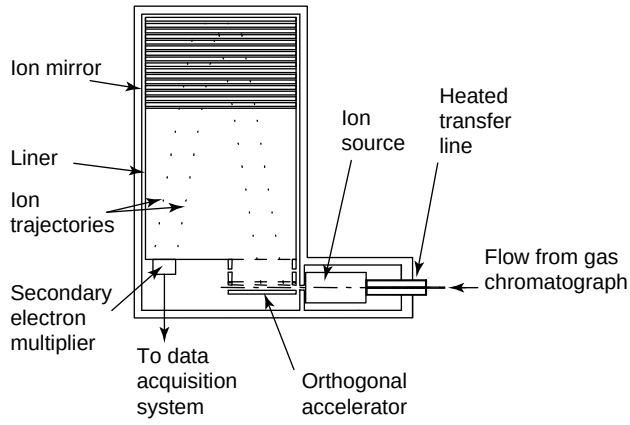

Figure 23 SprinT™ orthogonal acceleration GC/TOFMS device. See text for further details. (Figure courtesy of HD Technologies, Manchester, UK.)

Output from the GC column is normally directly inside the ionization volume of the EI source. Aside from the speed of analysis, TOF analyzers yield EI spectra identical to those found in the standard libraries. The TOF system could therefore use already existing protocols of gas chromatography/mass spectrometry (GC/MS) analysis.

Commercially available orthogonal acceleration GC/TOFMS instruments with conventional EI sources are the SprinT™[180,181] (Figure 23) (HD Technologies Ltd., Manchester, UK) and the GCT™[182] (Micromass, Manchester, UK). The GCT™ also offers CI sources and field ionization sources.[182]

9.1.5 *Orthogonal Extraction Time-of-flight Analyzer with Supersonic Molecular Beam*

A completely different type of GC/MS interface is the SMB interface.[183,184] Davis et al. have coupled an SMB interface with an orthogonal extraction TOF analyzer for GC/MS analyses (Figure 24).[185] In the SMB interface the GC column eluent is mixed with a make-up gas at atmospheric pressure and the mixture is expanded through a 150 μm heated nozzle into a low-pressure region. The make-up gas is helium or hydrogen. The sample molecules from the GC eluent are entrained in the supersonic expansion and the supersonic beam is skimmed at different stages of differential pumping prior to passing into the ionization region.

The interface consists of several differentially pumped stages to reduce the pressure from the atmosphere to the high vacuum in the TOF analyzer. The GC column eluent and make-up gas expand into a region pumped by a 28-$m^3\,h^{-1}$ rotary pump that reduces the pressure to around 0.8 Torr. The SMB is skimmed into a second stage that is pumped by a 70-$L\,s^{-1}$ turbo pump, which reduces the pressure to around 5 mTorr and further skimmed into

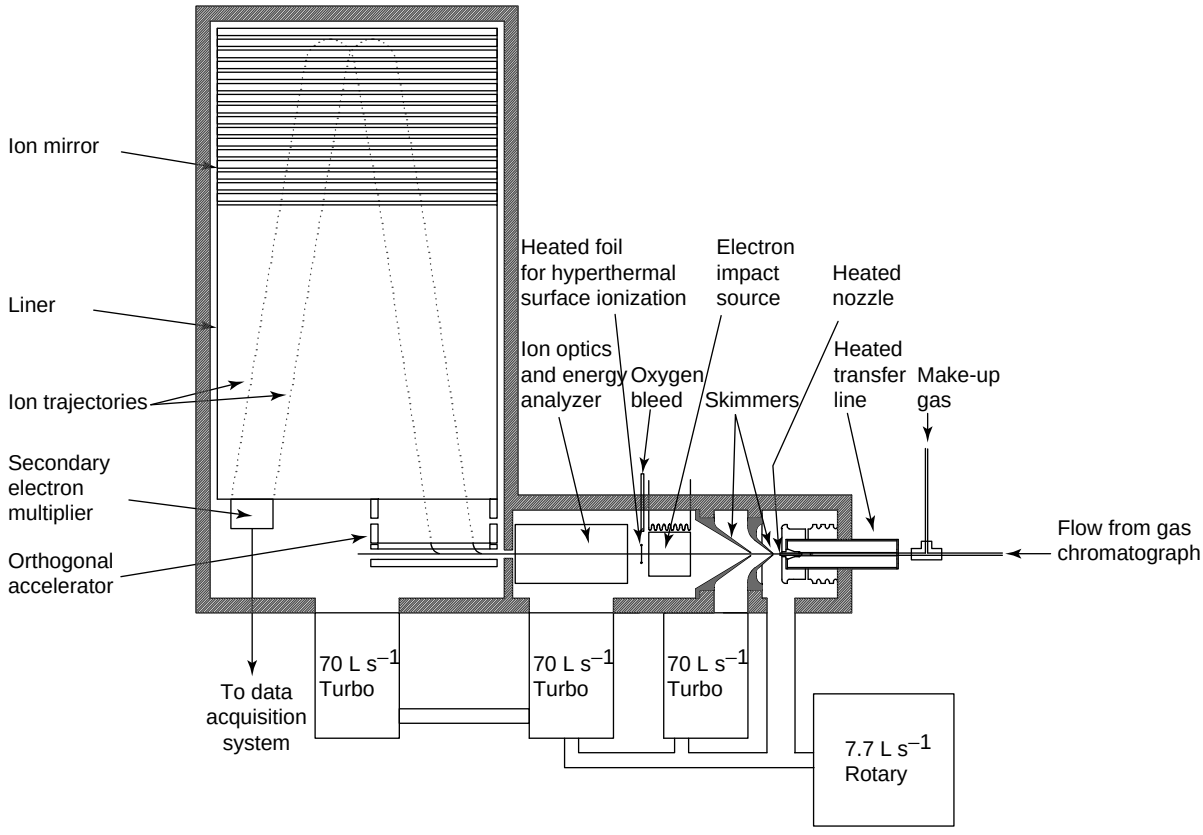

Figure 24 Layout of the HyperJet MS system, which illustrates the HyperJet sources and their orientation to the TOF extraction region and analyzer. Pressure in the "T" connecting the column is close to atmosphere. See text for further details. (Figure courtesy of HD Technologies, Manchester, UK.)

a region containing the ion source that is pumped by another 70-L s^{-1} turbo pump. The pressure in the ion source region is around 100 μTorr. The ion beam from the ion source passes through a conductivity restrictor into the TOF analyzer, which is pumped by a third turbo pump to a pressure in the low microtorr range. The turbo pumps all have a drag stage to maintain a sufficient compression ratio for pumping the light gases and so they can be backed by the same rotary pump from the first stage.

In the supersonic molecular beam/electron ionization (SMB/EI) source the supersonic beam of neutrals flows through the middle of a region where they are bombarded with ionizing electrons. The multitude of collisions in the supersonic expansion results in vibrational cooling of the molecules and the molecules have much lower internal energy prior to ionization compared with a more conventional EI source. The lower internal energy results in less fragmentation and molecular ion enhancement. The degree of molecular ion enhancement is compound dependent but is particularly evident in compounds like long-chain hydrocarbons where the molecular ion is the base peak in the mass spectrum.

Another feature of SMB/EI is the ability to perform background ion filtration. The total energy of the ions after acceleration, from a conventional EI source, is determined by the voltage in the ion source volume where the ions were formed. In the case of SMB/EI there is an additional component of energy due to the kinetic energy that the molecules possessed in the SMB at the time of ionization. Background molecules that were not entrained within the SMB will have only thermal kinetic energy at the time of ionization. This means that background ions can be filtered out using a simple energy filter and therefore any source memory effects are minimized.

Amirav et al. have also developed a surface ionization source for the SMB interface known as hyperthermal surface ionization (HSI), which has also been used for GC/MS.[183,184] Most HSI work has been done so far in the positive ion mode. The positive-ion HSI source involves collision of the SMB with a heated rhenium foil. The active surface is rhenium oxide, which is maintained by oxygen at a pressure around 10 μTorr in the vicinity of the foil. The hyperthermal energy of the molecules in the SMB allows the molecules to overcome the image

potential close to the surface. Ionization efficiency is a function of the difference between the ionization potential of the molecule and the work function of the surface. Rhenium oxide is an efficient surface for the formation of positive ions because it has a high work function. The ionization efficiency is dependent on the ionization potential of the molecules. Molecules with low ionization potentials will be ionized more efficiently. This leads to selectivity of the HSI source on the basis of ionization potential. For example, in the case of crude oil analysis, the aromatic components are ionized very efficiently but there is very little ionization of the aliphatic components.

The sensitivity of the HSI source can be very high. For compounds with low ionization potentials, such as anthracene and lidocaine, the ionization efficiency can be up to two orders of magnitude higher than EI. HSI mass spectra often show fragment ions, but the mass spectral fingerprints are generally different to those from EI.

9.1.6 Orthogonal Extraction Time-of-flight Analyzer with a Metastable Atom Bombardment Ion Source

In the MAB source, ionization is effected by a beam of excited neutral atoms.[164] The excited atoms are generally noble gases, although molecular gases such as nitrogen can also be used. The beam of neutrals is created using a Penning discharge and charge exchange, somewhat akin to the fast atom bombardment gun. The ion source is very similar to an EI source where the electron beam is replaced by the metastable atom beam.

The interesting thing about the MAB source is that the energy available for ionization and fragmentation is discrete, i.e. fixed to the excitation energy of the atom. This means that the MAB source has selective ionization and fragmentation capabilities through the use of different rare gases in the atom gun. If helium is used, there is substantial fragmentation similar to an EI mass spectrum. If xenon is used, the ionization is very soft and the spectra consist basically of molecular ions.

The combination of MAB with TOF has been used for bacteria identification using pyrolysis,[167] and has several advantages for this application. The complex spectra normally obtained using pyrolysis with EI are much simplified when the MAB source is used to produce molecular ions and eliminate the fragmentation from the complex mixture. The fast data acquisition of the TOF analyzer also allows the transient process to be characterized. The time profile of the spectra emerging during the pyrolysis process provide a characteristic fingerprint of the bacteria.

9.2 Aspects of Gas Chromatography/Time-of-flight Mass Spectrometry

9.2.1 Trade-offs of Analysis

Mass spectrometer designs are subject to various constraints to give the best compromise in operating parameters for their intended application. TOF instruments are no exception and compromises must be made in relation to parameters such as mass resolution, mass accuracy, sensitivity, and accuracy of quantitation. These parameters also have an impact on coupling with fast chromatography.

The trade-off between mass resolution and sensitivity is perhaps most familiar to those involved in MS. In general, higher mass resolution results in less sensitivity. In the case of orthogonal acceleration TOF instruments, higher mass resolution instruments require longer flight paths. Longer flight paths mean that the analysis time increases relative to the fill time of the push-out region and therefore the duty cycle (and hence sensitivity) of the instrument is reduced. Mass accuracy and quantitation are also intimately linked to these parameters.

For a given number of ions in a mass spectral peak, the higher the mass resolution, the better will be the mass accuracy of the determined peak centroid. With higher mass resolution the time width of the peak is narrower and therefore the time error for the centroid position will be smaller. However, the accuracy of the centroid of the peak is also a function of the ion statistics within the peak, i.e. the mass accuracy of the peak is limited by the number of ions in the peak. The number of ions in the peak is related to sensitivity, which is related to mass resolution. In other words, mass accuracy is not just a function of mass resolution. It is possible to get the same mass accuracy from a lower resolution instrument; this is subject of course to their being sufficient mass resolution to resolve isotopic peaks in the lower resolution case and the lower mass resolution instrument having sufficient transmission and duty cycle to provide correspondingly higher ion statistics. This also means that even high-resolution instruments will not be able to provide accurate mass at trace levels of signal.

If we now impose the additional constraint of chromatographic timescales there are yet more trade-offs that must be examined. In GC/MS a fixed amount of sample is injected and the compounds elute in a short time frame. Orthogonal TOF instruments have a very high underlying DAR, typically tens of kilohertz. At an acquisition rate of 40 kHz there are 40 000 individual mass spectra obtained per second. However, there are generally not enough ions in one of these mass spectra to be statistically meaningful, and in any case this is far too much data to store to disk at this rate. As a result, a number of individual shots

are added together before the resultant mass spectrum is saved to disk. For GC/MS mass ranges, mass spectra can be saved to disk at rates of up to several hundred per second. At an acquisition rate of 40 kHz, if 100 spectra per second are saved to disk then each mass spectrum will be the sum of 400 individual shots. For a given chromatographic peak the number of ions in the mass spectral peaks will thus be a function of the scan rate, where scan rate in this context refers to the number of spectra per second saved to disk.

This means that there is a compromise between scan rate and mass accuracy, because the ion statistics are poorer at higher scan rates. Limits of detection will also be a function of scan rate because S/N will be poorer at high scan rates for the same statistical reasons. To maintain levels of mass accuracy and S/N at high scan rates, the ions available need to arrive in a narrower time window, i.e. the GC peaks need to be narrower.

9.2.2 Quantitation

For signals well above the detection limit, quantitation on a GC/TOFMS is relatively independent of the scan rate. This is due to the very fast underlying DAR of several tens of kilohertz. The number of ions at any given mass from a GC peak will be the sum of ions from the individual spectra obtained during the elution time of the peak. This total sum will remain the same if it is partitioned into several sub-sums during the time of peak elution, i.e. if a higher scan rate is used. The higher scan rate simply adds more points to the histogram representing the integral under the GC peak. However, at very high scan rates the accuracy of quantitation may be reduced due to difficulties in peak detection within the stored mass spectra because of poor ion statistics. A higher scan rate, however, provides more points across the GC peak, thereby better defining the chromatography. The greater the number of points across the peak, the easier it will be to deconvolve closely eluting peaks.

9.2.3 Dynamic Range

The dynamic range of TOF instruments is limited mainly by its data acquisition system. In ion counting mode using a TDC not more than 0.5 ion per shot per mass could be acquired, which limits the counting rate by $1-2(\times 10^4)$ counts s^{-1} $mass^{-1}$. With background typically at levels of a few counts s^{-1} $mass^{-1}$, this limits the dynamic range by 10^3-10^4. This dynamic range tends to exceed that of microbore columns. However, it is not sufficient for standard narrow-bore fused-silica columns. The required extension of dynamic range may be achieved by using an ADC running alongside the TDC.[186–188] Alternatively, a multi-anode detector and multiple TDCs may be run in parallel.[189]

10 INDUCTIVELY COUPLED PLASMA/ TIME-OF-FLIGHT MASS SPECTROMETRY

It is probably true to say that the entry of TOF into the field of ICP MS has proved to be a formidable task due to the maturity of this field. Indeed, several TOF parameters were to be enhanced sometimes by orders of magnitude from typical values just to achieve performance of the already highly developed quadrupole and sector instruments.[190]

The progress in ICP/TOFMS instrumentation development is summarized in Table 3. Generally, ICP/TOFMS has approached average levels typical for quadrupole mass spectrometers (except that sensitivity continues to be inferior at higher masses). Presently, two commercial ICP/TOFMS systems are available: the Optimass 8000™ produced by GBC Scientific Pty. Ltd. (Australia),[188,191] and the Renaissance™ produced by Leco Incorporated (St. Joseph, MI, USA).[192–195] Both machines consist of a differentially pumped atmosphere-to-vacuum interface, focusing ion optics, pulsed ion accelerator, pulsed ion gate for removal of matrix ions, ion mirror, and a multidynode electron multiplier. In the Optimass 8000™, ions after leaving the focusing ion optics are accelerated orthogonally to the original direction of their motion. In the Renaissance™ ICP/TOFMS instrument ions are accelerated along the same direction.

The main driving force of ICP/TOFMS development is associated, as usual, with unique applications, possibly due to the speed and all-mass detection capabilities of the TOF analyzer.[196] For high-throughput multi-element analysis (>5–10 samples min^{-1}, with >10–20 elements monitored), ICP/TOFMS could provide better overall performance (sensitivity, isotope ratio accuracy, dynamic range) than sequential analysers. For laser-ablation atomization, ICP/TOFMS provides a unique capability of quasi-simultaneous (within <0.1 ms) multi-element monitoring. In terms of high-throughput multi-element monitoring, the success of ICP/TOFMS will depend on improvements of sample introduction techniques. GC,[197] LC, supercritical fluid chromatography, capillary zone electrophoresis, field-flow fractionation, counter-current chromatography, etc. have been interfaced to ICP/TOFMS with different degrees of success.[198] Development of these hyphenated techniques will strongly benefit from the parallel-analysis capability of the TOF analyzer. For laser ablation applications, the success of ICP/TOFMS will depend on improving the robustness of the atmosphere-to-vacuum interface in the presence of relatively high contamination caused by laser-sputtered samples.[196]

Table 3 Progress in ICP/TOFMS instrumentation

Main parameters	Typical values for ICPMS[a]	Associated difficulties specific to TOF	Solutions
Sensitivity	Better than 1 ng L^{-1} for $m/z > 100$ (in the absence of interfering species)	1. Background from scattered matrix ions (Ar, ArO, etc.) 2. Insufficient ion count rate due to lower duty cycle, and transmission comparing to traditional continuous-beam analysers 3. Electronic noise	a) Filtering of background using parallel-plate "smart gate"[191,199] or blanker[194] or filter grids near detector[200–202] b) Increasing transmission using axial scheme[192] or gridless design of orthogonal accelerator c) Thorough shielding
Quantitation (linearity and dynamic range)	10^7 with 0.99 correlation coefficient	1. Limited dynamic range of TDC (about 10^4) or ADC (about 100)	a) Combined TDC/TR[b] system with logarithmic amplifier[191] b) On-the-fly matching of TDC and TR data
Abundance sensitivity	Intensity of ions of mass M must diminish by $>10^6$ at $M \pm 1$	1. Characteristic for TOF high tailing of mass peak[39,200–202] 2. Ringing of secondary electron multiplier[39,200–202]	a) "Smart gating" b) Better impedance matching of SEM c) Higher resolution fwhm is provided by an ion mirror
Accuracy of isotope ratio	Mass discrimination should be around 1% per amu	1. High mass bias of conventional orthogonal TOF due to mass dependence of kinetic energies caused by jet expansion[195]	a) Higher energy of ions in orthogonal direction makes mass dependence negligible b) Alternatively, axial scheme[192] could be used
Robustness	>6 months SEM life, 3 months between major services while aspirating liquids	1. Oversaturation of SEM by intense ion packets 2. Contamination of grids	a) "Smart gating" or blanking of matrix peaks b) Gridless design of orthogonal accelerator

[a] ICPMS, inductively coupled plasma mass spectrometry.
[b] TR, transient recorder.

11 QUADRUPOLE ION TRAP/ TIME-OF-FLIGHT MASS SPECTROMETRY

The QIT accumulates ions, thereby amplifying the observed ion current, especially for weak ion beams.[203] In addition, the QIT offers other advantages for mass analysis, including MS/MS.[204] Typically, mass-selective detection is used for ions trapped in a QIT,[205] however an alternative approach involves a TOF mass analyzer for "simultaneous" detection of the ions, thereby enabling a much more rapid analysis.[206] In addition, for ions produced using a continuous ionization source, such as ESI, the QIT provides a convenient interface for converting the continuous beam of ions into discrete packets for TOFMS. Thus, the duty cycle is essentially limited only by the trapping efficiency of the QIT. Consequently, a QIT/TOFMS enables ions to be detected over a wide m/z range without scanning, thereby offering a means for rapid analysis, suitable for applications such as a detector for a fast separation technique.

Although QIT/TOFMS was described as early as 1979,[173] Lubman et al. were the first to explore the potential of this hybrid for analytical applications,[206–208] including the first description of electrospray ionization/quadrupole ion trap/time-of-flight mass spectrometry (ESI/QIT/TOFMS).[123] Figure 25 shows a (simplified) schematic of a QIT/TOFMS instrument used as a detector for an external ionization source. Ion optics (typically Einzel lenses) before and/or after the QIT are also used with this hybrid but have been omitted in this diagram for simplicity. The operation of the QIT/TOFMS instrument can be divided into two events: ion storage and ion extraction. During ion storage, both endcap electrodes

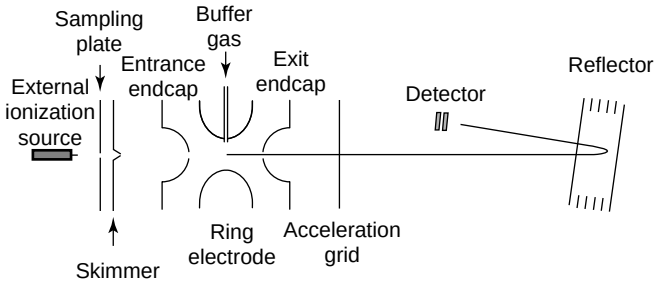

Figure 25 Simplified schematic of the main components of a QIT/TOFMS for an external ionization source. The solid line illustrates the path of an ion during extraction.

are typically held at ground potential while an RF voltage (typically a frequency of ~1 mHz and an amplitude of 200–3000 V_{p-p} are used) is applied to the ring electrode. A buffer gas (e.g. helium) is also added to the QIT to assist in cooling ions to the center of the QIT.[205] The buffer gas is especially important for reducing the kinetic energies of externally generated ions that pass into the QIT through one of the endcap electrodes.[203] Under the appropriate conditions,[209] ions will have stable trajectories within the QIT and will undergo collisional cooling and accumulate in the center of the QIT. After a predetermined trapping time, ions are extracted from the QIT, which now serves as part of the acceleration region of the TOFMS instrument. In the first ESI/QIT/TOFMS study,[123] the RF voltage was shut off, and after a small time to allow for the decay of the RF voltage the ions were extracted from the QIT by pulsing the exit endcap electrode. The extracted ions separated in the drift tube of a reflectron TOFMS instrument and were detected using dual MCPs.

The role of several parameters, including the RF voltage and the method of ion extraction, on the performance of QIT/TOFMS has been investigated.[209–214] Chambers et al. reported significant improvements in their QIT/TOFMS spectra by using a technique known as ''RF clamping''.[213,214] With this method, the RF voltage is turned off and immediately clamped to zero instead of being allowed to decay. A decaying RF voltage during ion extraction will cause broadening in the velocity distribution of the ions, thereby decreasing the observed mass resolution. Pulsing both endcap electrodes (bipolar extraction) instead of only the exit endcap electrode (unipolar extraction), to maintain a more uniform electric field in the extraction region, was also introduced.[212,213] This change resulted in significant improvements in the mass resolution of the QIT/TOFMS spectra and is now used routinely. Finally, the role of the QIT in this hybrid has been expanded by Lubman et al., who have recently described procedures for MS/MS.[215,216]

To date, the emphasis of the literature published on QIT/TOFMS has involved the use of this hybrid as a detector for fast separation techniques such as CE.[217] However, QIT/TOFMS has been used also to investigate unimolecular reactions in the gas phase.[218–220] Furthermore, although ESI has been most commonly used with QIT/TOFMS, the implementation of other ionization sources, both internal and external to the QIT, has been described.[207]

Several studies have used ESI/QIT/TOFMS to investigate small analytes; however, the analysis of higher MW compounds, such as proteins, has not been nearly as successful. Purves and Li showed that in their linear QIT/TOFMS instrument the quality of mass spectra for protein ions in excess of ~20 000 Da degraded rapidly.[209] The reason was attributed to the higher pressure (~10^{-3}–10^{-4} Torr) in the acceleration region due to the buffer gas. This result is not surprising because Chernushevich et al. have shown that the mass spectra of protein ions begin to degrade at pressures as low as 10^{-6} Torr in orthogonal ESI/TOF.[128] Consequently, a conflict in the analysis of larger MW species with this hybrid arises because the buffer gas plays a crucial role in trapping externally generated ions. For this hybrid to be useful for the analysis of protein ions, this fundamental problem must be addressed. Other methods of ion cooling need to be explored or, alternatively, the ion trap could be uncoupled from the extraction process (a method for the latter is described below).

Several variations of the original design of the QIT/TOFMS instrument have been reported. For example, Wang et al. developed a QIT that was orthogonal to a TOFMS instrument to allow uncoupling of the operation of the QIT and TOFMS.[221] In addition, some recent hybrids reported in the literature have significant changes compared with the Lubman design, but retain the common theme of trapping ions and subsequent analysis using TOFMS.[206] Three of these designs that are still in their early development stages are described below.

A segmented-ring, cylindrical ion trap TOFMS instrument was reported by Enke et al.[172] This design replaces the hyperbolic-shaped electrodes with planar endcaps, and two, thin, circular diaphragms further approximate the cylindrical ring electrode. This geometry produces a unidirectional linear field during ion extraction and thereby avoids the need to cool the ions to the center of the ion trap. The advantage of this technique is that a high mass spectral generation rate can be maintained (>1 kHz) compared with QIT/TOFMS (<100 Hz).

The linear ion trap (LIT) coupled to a TOFMS instrument was reported by Douglas et al.[222] This instrument uses an RF-only quadrupole that is coupled orthogonally to a TOFMS instrument. By using timed stopping potentials on the entrance and exit apertures of the quadrupole, an ion trap is created. Although results to date with LIT/TOFMS are not comparable with QIT/TOFMS, LIT/TOFMS offers some advantages over QIT/TOFMS, such as a higher ion capacity. Furthermore, the separation of the trapping hardware from the acceleration region of the TOFMS instrument in LIT/TOFMS may offer additional advantages in the analysis of larger MW analytes.

Finally, an entirely different type of ion trap that is operational at atmospheric pressure (760 Torr) has been developed for use with a TOFMS instrument.[223] This ion trap was developed based on the ion focusing principles of a cylindrical-geometry, high-field, asymmetric-waveform ion mobility spectrometer.[224] The mechanism by which an ion is trapped is based on changes in ion mobility at

high fields and is dependent on ion properties other than m/z. The use of this ion trap with TOFMS may offer advantages for analytical applications.

At the present time, there are no commercial versions of QIT/TOFMS devices.

12 MATRIX-ASSISTED LASER DESORPTION/IONIZATION TIME-OF-FLIGHT MASS SPECTROMETRY

Undoubtedly, the most widely applied laser desorption technique to date has been MALDI. Two groups independently developed MALDI: the group of Tanaka and the group of Hillenkamp.[12–15] Further historical details of these developments are found in section 2 of this article. A number of reviews have been written regarding this technique.[102,225–227] Consequently, this section will provide a brief overview of MALDI fundamentals and popular applications.

12.1 Matrix-assisted Laser Desorption/Ionization Fundamentals

MALDI is a desorption/ionization technique for poorly volatile compounds that utilizes the energy inherent in a moderately focused pulsed laser. Although the technique has been applied to liquid samples,[152,154,157–162] it has been far more successful in the analysis of solid-state crystals or thin films. Consequently, this section will only focus on solid-state MALDI analysis.

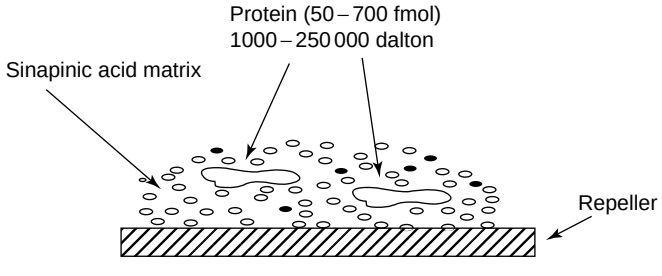

Figure 26 MALDI sample preparation and probe geometry. See text for details.

MALDI samples are prepared on the surface of a solid probe (see Figure 26). Generally speaking, analytes are mixed with a solution containing a laser energy-absorbing compound, known as the matrix. Sample and matrix material can be premixed prior to probe deposition or mixed directly on the probe surface.[228–230] In some instances, a seed layer of matrix is first applied followed by the addition of a sample–matrix mixture.[231,232] Sample–matrix co-crystallization is allowed to proceed by atmosphere evaporation,[12] or by rapid evaporation facilitated by the application of a vacuum.[233] Although the exact nature of desorption and ionization are still being investigated, it is clear that the generation of quality ion signal depends upon efficient sample incorporation into the matrix material. In this fashion, it is necessary to match the solubility characteristics of the matrix to the analyte. Additionally, it is necessary to match the absorbance maxima of the matrix to the spectral wavelength of the laser source. Most commercial MALDI devices use inexpensive nitrogen lasers ($\lambda = 337$ nm). Table 4 lists some of the commonly used nitrogen-laser-compatible matrices for bioanalysis, along with their application and solvent systems.

After crystallization, the probe is inserted into the TOFMS instrument where it becomes intimate with the repeller lens of the ion optic assembly. In some cases, such as with two-dimensional sample stages, the probe is the repeller. After a brief pump-down period, desorption and ionization are promoted by a short-duration pulse (typically <10 ns) of moderately focused laser light. Coarsely speaking, desorption and ionization are achieved by a thermal energy transfer process from the matrix to the analyte.

In contrast to LDI microprobe techniques that highly focus incident laser radiation to small spots ($<10\,\mu$m in diameter) and high irradiance (10^{12} W cm^{-2}),[23] MALDI typically employs spots of the order of several hundred microns in diameter with resultant power densities of about 10^6 W cm^{-2}.[102] Ultraviolet lasers used in MALDI include: nitrogen lasers; excimer lasers ($\lambda = 193$, 248, 308, and 351 nm); frequency-doubled,

Table 4 Commonly used MALDI matrices, their applications, and solvent systems

Matrix	Application	Solvent[a]
Sinapinic acid[230,234]	Proteins and large peptides	ACN, H_2O, MeOH, TFA, FA
α-Cyano-4-hydroxycinnamic acid (HCCA)[235]	Amino acids, peptides, small proteins	ACN, H_2O, MeOH, TFA
2,5-Dihydroxybenzoic acid (DHB)[236]	Small peptides and glycoproteins	MeOH, H_2O
Ferulic acid[237]	Large proteins	IPA, H_2O
2,6-Dihydroxyacetophenone[238]	Peptides, proteins, oligonucleotides	ACN, MeOH, DAHC, H_2O
3-Hydroxypicolinic acid[239]	Oligonucleotides	MeOH, H_2O

[a] ACN, acetonitrile; MeOH, methanol; TFA, trifluoroacetic acid; FA, formic acid; IPA, isopropyl alcohol; DAHC, diammonium hydrogen citrate.

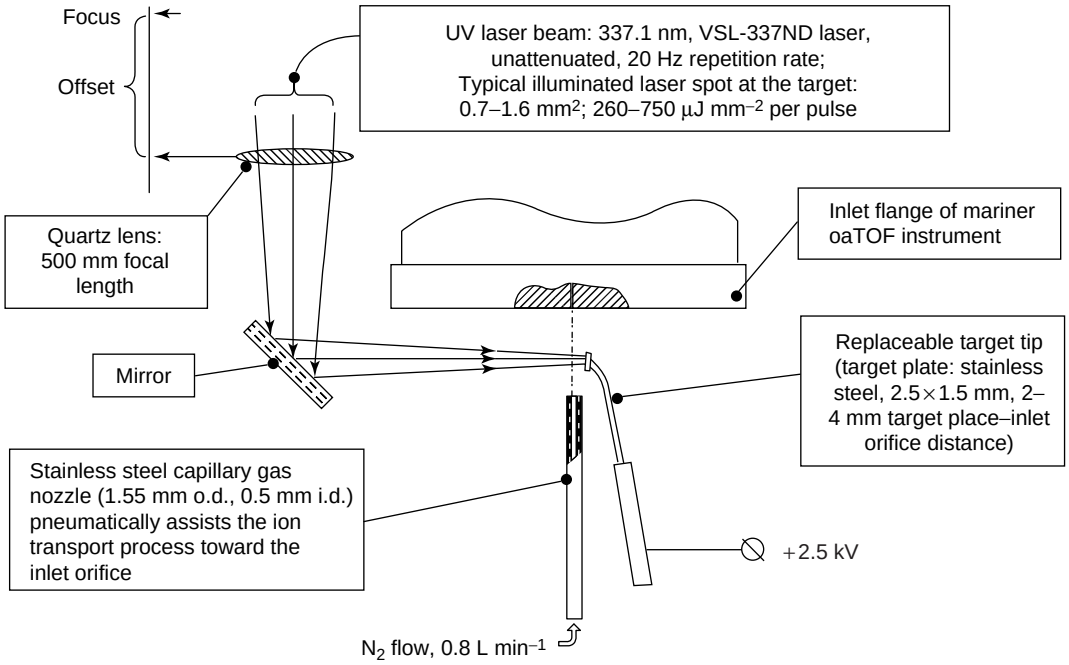

Figure 27 Atmospheric pressure MALDI interface. An atmosphere housed MALDI source is coupled with an orthogonal acceleration TOF device (Mariner oa-TOF, PE Biosystems, Framingham, Massachusetts, USA). See text for further details. (Figure courtesy of Professor Alma L. Burlingame, University of California at San Francisco.)

excimer, pumped-dye lasers ($\lambda = 220–300$ nm); and Q-switched, frequency-tripled and -quadrupled Nd : YAG lasers ($\lambda = 355$ and 266 nm, respectively).[102] Infrared lasers used in MALDI applications include: transverse electrode atmospheric (TEA)–CO_2 ($\lambda = 10.6$ μm); Er : YAG ($\lambda = 2.94$ μm);[156,240,241] and optical parametric oscillator lasers ($\lambda = 2.8–3.3$ μm).[242,243]

Originally, MALDI sources were coupled with linear, constant, parallel ion extraction TOF devices.[12] Eventually ion mirrors were added, improving the resolution for low-molecular-weight compounds.[15,102] The rediscovery of TLF and its subsequent addition to MALDI analysis resulted in significant improvements in mass resolving power and mass accuracy.[63,70,74,77–83,93,94,244] Higher order corrections combining TLF with reflectrons has extended the mass resolving power of this approach to about 20 000 m/dm, with low parts-per-million mass assignment error.

For all of the above-mentioned TOF geometries, MALDI ions are generated in vacuo and parallel ion extraction is utilized. Recently, a number of groups have experimented with orthogonal acceleration extraction coupled with the generation of MALDI ions at elevated pressures.[4,58–61] Initial attempts at orthogonal extraction of MALDI ions met with modest success. Difficulties with initial spatial and velocity distributions created spectra with poor resolution and sensitivity.[56,57] The addition of collisional cooling within a quadrupole ion guide greatly improved matters.[58] Eventually, a tandem orthogonal MALDI quadrupole–quadrupole TOF device was created, capable of high accuracy, high-resolution MS and MS/MS operation.[4] This device is described further in section 14 of this article.

Laiko and Burlingame introduced the concept of atmospheric pressure MALDI when they combined an external MALDI source with an orthogonal TOF analyzer (see Figure 27). They reported a reduced level of unimolecular decay, possibly due to collisional dampening with atmospheric gas. Superior analysis for multicomponent peptide mixtures was demonstrated compared to conventional MALDI.

Verentchikov et al. combined the MALDI, two-dimensional sample stage of a commercial device (Voyager, PE Biosystems, Framingham, MA, USA) with the same orthogonal TOF analyzer used by Laiko and Burlingame (see Figure 28). This device was used to analyze a number of proteins as well as DNA molecules.[59] Optimum ion stability was demonstrated when desorption was performed at 1 Torr of pressure. Detector post-acceleration technology was employed to enhance the sensitivity of the device for large proteins.[60] Isotope-envelope-limited resolution for enolase, a protein of 47 kDa MW, was demonstrated. The authors also noted partial resolution of BSA isoforms. Because of the latter, they now believed protein resolution to be limited by sample heterogeneity.[59]

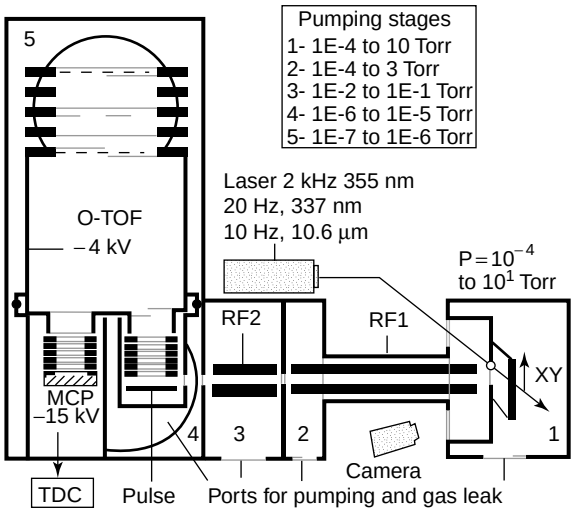

Figure 28 MALDI orthogonal acceleration TOF device (o-TOF) of Verentchikov et al. See text for details. (Figure courtesy of Dr Anatoli Verentchikov, PE Biosystems, Framingham, MA, USA.)

Orthogonal acceleration MALDI geometries employing collisional cooling have a number of advantages when compared to in vacuo, parallel extraction approaches. These include:

- Uncoupling the initial ion position and time of creation from the TOF measurement obviates concerns for precise sample positional and electrical clamping.
- Ions are extracted at rates 1000-fold greater than they are created. This allows single ion per m/z detection and the use of low-cost TDC devices.
- Desorption performed at elevated pressures permits the use of higher laser energies without deleteriously effecting signal fidelity. This circumvents the need to hunt and peck for good crystals, greatly facilitating analysis.
- The combination of the two preceding points allows for the creation of a pseudo-continuous introduction of ions. In this manner, MALDI ion current stability approaches that of GC and LC sources.
- Higher order TLF and reflectron focusing provides enhanced resolution at all m/z values without the need to adjust pulse voltage or time lag parameters.
- Calibration of ion TOF is independent of initial conditions, making the calibration extremely linear over several orders of m/z magnitude.

Because of these factors, orthogonal MALDI systems show great promise in simplifying and enabling the use of this analytical technique.

12.2 Matrix-assisted Laser Desorption/Ionization Applications

MALDI has been extended to the analysis of small molecules,[245–249] organic polymers,[5,14,15,250–256] proteins,[12–15,75,102,111,225,230,234,235,257] oligosaccharides,[258–265] and oligonucleotides.[9,59,244,266–284] The preponderance of matrix chemical noise has greatly limited the utility of MALDI to study compounds with $m/z < 1000$. In terms of organic polymer analysis, difficulties in ionizing many species coupled with sample preparation and instrumentation effects upon the determination of mean MW and polydispersity have also limited the acceptance of MALDI in this analytical arena.[252,253] Consequently, this review will focus on the applications of protein/peptide, oligonucleotide, and oligosaccharide analyses.

12.2.1 Protein and Peptide Analysis

Proteins and peptides were the first compounds successfully studied by MALDI.[12–15,102,225,230,252,253,257] As such, it was a natural step to link MALDI analysis to LC and gel-isolated proteins. LC-isolated proteins and peptides are generally collected as fractions and, in off-line fashion, mixed with matrix for MALDI analysis. Although in situ MALDI analysis of gel-isolated proteins has been developed,[285] this direct approach has demonstrated poorer resolution and mass accuracy when compared to off-line approaches. Consequently, most laboratories link MALDI analysis to gel-isolated proteins through the use of membrane blots or buffer extraction.[286–290]

Although providing a facile means to assess rapidly the purity or to analyze mixtures, very little information is actually afforded by an accurate MW assignment for a given protein population. Significant analytical power is unleashed when MALDI/TOF analysis is combined with enzymatic and/or chemical modification of target proteins, enabling the elucidation of structural components, post-translational modifications, and protein identification. During this process, a sequential or parallel application of protein modifications is often employed, followed by MS analysis. MS profiles are reconciled with known mechanisms of chemical and enzymatic activity, providing a priori knowledge of protein characteristics.

Peptide mapping of proteins has been facilitated by the use of various endopeptidases.[291,292] Subsequent peptide maps have been submitted for protein and cDNA database searches in attempts to provide possible protein identification.[293] Exopeptidases have been applied to peptides for the purpose of generating a ladder sequence.[294,295] In this fashion, enzyme/substrate ratio and digestion times are altered to create overlapping peptide populations with varying degrees of exopeptidase digestion. The resultant spectrum creates a ladder of

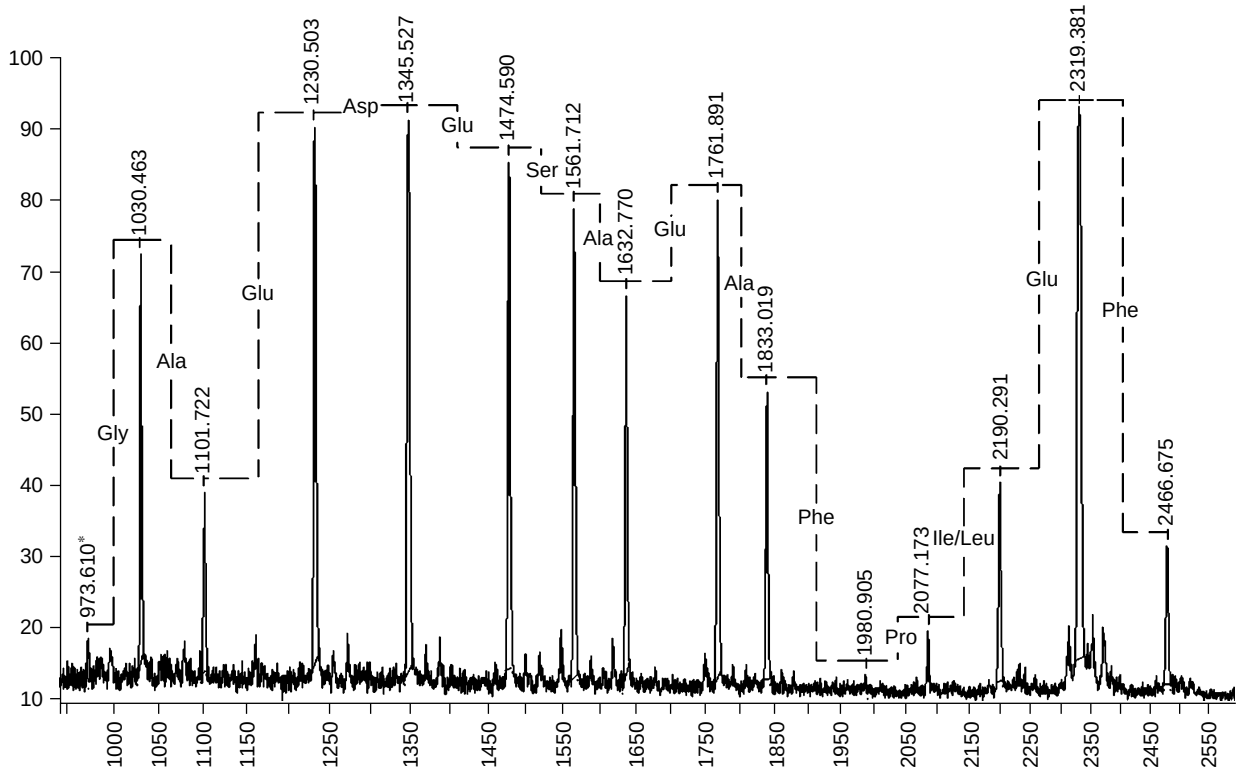

Figure 29 Peptide ladder sequence of adrenal corticotropic hormone clip 18–39 using carboxypeptidase *y*. (Figure courtesy of Thermo Bioanalysis, Inc., Santa Fe, NM, USA.)

signals where primary amino acid sequence is revealed by taking the mass difference between adjacent peaks (see Figure 29). Chemical methods of creating peptide ladder sequences have also been developed.[296,297] Protein post-translational modifications have been studied through the use of phosphorylases[227,298] and deglycosidases.[258,259]

12.2.2 Oligonucleotide Analysis

Early attempts at MALDI analysis of DNA and RNA using various protein matrices produced spectra of poor resolution and sensitivity.[276] Compared to proteins and peptides, the analysis of oligonucleotides was complicated by a strong affinity for alkali metals and significant unimolecular decay. Ion exchange approaches utilizing cation-exchange beads[244] or ammonium salts of organic acids[299] functioned well to eliminate superfluous sodium or potassium ions, simplifying oligonucleotide ion profiles to primarily parent molecular ions.

The primary source for oligonucleotide unimolecular decay has been attributed to gas-phase acid-catalyzed base loss among purines and cytidine.[276,281] Credence to this mechanism has been further provided by experiments in which 7-deaza-guanosine and 7-deaza-adenosine nucleosides demonstrated significant improvements in MALDI stability when compared to their normal analogs.[282] Ultimately, the discovery of milder matrices such as 3-hydroxypicolinic acid,[239] 2,4,6-trihydroxyacetophenone,[299] and 2,6-dihydroxyacetophenone facilitated MALDI analysis of oligonucleotides.[238] The latter, coupled with TLF MALDI/TOF geometries, has greatly extended the mass resolving power and mass accuracy for MALDI oligonucleotide applications.[244,277]

As was the case for protein analysis, accurate MW assignments of species within oligonucleotide mixtures only provide indications of purity or mixture status. Primary sequence information can be created when DNA or RNA molecules are enzymatically or chemically perturbed. Unlike proteins, the problem of sequence determination is facilitated by the fact that the smallest MW difference among all bases is about 9 Da (thymine and adenine). Primary sequence analysis of oligonucleotides has been performed using 3′- and 5′-exonucleases,[299,300] as well as primer-based dideoxynucleotide chain termination approaches.[283,284]

12.2.3 Oligosaccharide Applications

MALDI analysis of carbohydrates is complicated by the fact that these species have very low proton affinities. As such, ionization typically occurs by conjugation

with an alkali metal, usually sodium or potassium.[262] Often the analysis is aided by the addition of a small amount of alkali salt to preferentially drive the ionization process to a single species. Dilute sodium chloride or potassium chloride solutions (<1 mM) are typically used. For the most part, 2,5-dihydroxybenzoic acid has been used for this application.[262] Mohr et al. found the addition of 1-hydroxy isoquinoline to 2,5-dihydroxybenzoic acid improved resolution, S/N, and tolerance to buffer solutions.[261] The use of 2,4,6-trihydroxyacetophenone or 6-aza-thiothymine, along with negative ion scanning, proved effective for the analysis of acidic oligosaccharides and glycopeptides.[263]

Numerous papers report the analysis of complex oligosaccharides and glycoproteins by MALDI/TOF.[258–262,264,265] As was the case for proteins, peptides, and oligonucleotides, oligosaccharide characterization has been performed using a combination of enzymolysis and MS. Exoglycosidases have been used to elucidate primary sequence and conjugation sites for *n*-linked glycoforms.[258,259] PSD analysis has also been used to provide oligosaccharide structural information.[242,265]

13 SURFACE-ENHANCED LASER DESORPTION/IONIZATION

The advent of ESI[48,301,302] and MALDI[12,14] has extended the application of TOFMS to the study of proteins within complex biological systems. However, the nascent state of biological materials such as blood, sera, plasma, lymph, interstitial fluid, urine, exudates, whole cells, cell lysates, as well as cellular secretion products, typically precludes direct TOFMS analysis. These samples are usually complex, heterogeneous mixtures of many biomolecules, often in the presence of excess organic and inorganic salts. As such, significant sample preparation and purification steps need to be employed prior to MS investigation.

Classical methods of sample purification, such as LC (ion exchange, size exclusion, affinity, and reverse phase), membrane dialysis, centrifugation, immunoprecipitation, and electrophoresis, are labor intensive, often demanding a preponderance of sample and suffering from attendant analyte loss due to nonspecific binding and dilution effects. In this fashion, a direct, facile means for TOFMS study of proteins capable of detecting major and minor components is not easily realized. Furthermore, the desire to characterize proteins by the combined use of chemical/enzymatic modification followed by MS analysis often requires considerable off-line sample manipulation. Clearly, a simple platform allowing for serial and parallel alterations of the same sample would greatly facilitate the efficiency of this approach.

The need for an improved, simplified technique that offloaded complicated protein extraction, purification, and characterization requirements became quickly evident. In 1993, the concept of SELDI was suggested by Hutchens and Yip.[303] SELDI was described as a new strategy for MS analysis of macromolecules that provided an improved means for sample extraction and effective on-probe investigation of biopolymers when compared to conventional LDI or MALDI approaches. This landmark work incited enthusiastic interest in a number of researchers and sparked a cascade of studies in a plurality of laboratories.[268,273,286,288,304–319] Soon SELDI applications were described in terms of MS immunoassay,[312] Probe-Immobilized Affinity Chromatography/MS,[320] Probe Affinity MS,[307] and Solid Phase Probe Extraction.[319,321,322] Over the past 6 years, this technology has been applied successfully not only to the study of peptides and proteins,[242,268,274,286–288,303–307,309–316,320,323–333] but also to the analysis of oligonucleotides,[266,267,271] bacteria,[308] and small molecules.[334]

Recent advances in SELDI technology has led to the creation of a ProteinChip® Array System (Ciphergen Biosystems, Inc., Palo Alto, CA, USA) that has been applied successfully to a number of biological problems.[335–339] The ProteinChip® Array system is a SELDI-based, protein analysis platform that incorporates a straightforward sample preparation and detection approach, which de-emphasizes the need for sample preparation and MS expertise.

13.1 Surface-enhanced Laser Desorption/Ionization Versus Laser Desorption/Ionization and Matrix-assisted Laser Desorption/Ionization

Although distantly similar to LDI and MALDI, SELDI possesses a number of significant functional and practical distinctions from these other MS techniques. All three of these techniques rely upon the energy inherent in a focused laser beam to promote the creation of gaseous ions from mostly solid-state matter. Samples of interest are presented as crystals or thin films upon a sample support typically referred to as a probe. For both LDI and MALDI applications, the sample probe surface plays a passive role in the analytical scheme. The probe merely presents the sample to the mass spectrometer for analysis. In this fashion, raw samples must first be fractionated and desalted in order to produce a usable TOF signal. Furthermore, in the case of MALDI, large biopolymer analysis is only possible when analytes are co-crystallized with a solution of energy-absorbing compounds, typically referred to as matrix.

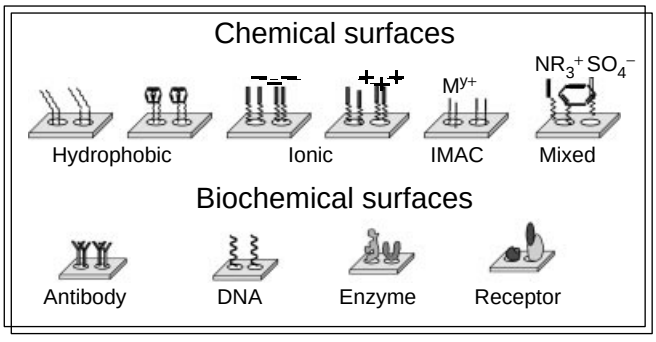

Figure 30 SELDI ProteinChip® Arrays. See text for details.

In contrast to LDI and MALDI, the sample probe used in SELDI plays an active role in the overall analytical scheme. As originally defined by Hutchens and Yip, SELDI consists of three subsets of technology: surface-enhanced affinity capture (SEAC), surface-enhanced neat desorption (SEND), and surface-enhanced photolabile attachment and release (SEPAR).[303,333,340] In SEAC, the probe surface plays an active role in the extraction, presentation, structural modification, and/or amplification of the sample. This is accomplished by creating an affinity interaction between the probe surface and the analytes of interest. Figure 30 depicts a number of SELDI ProteinChip® Array surfaces used in SEAC applications. Two fundamental surfaces are shown: chemical and biochemical surfaces. Chemical surface probes may be coated with reverse-phase, ion-exchange, immobilized metal affinity capture (IMAC) and mix-mode chromatographic media. In this case, analyte–stationary-phase dynamics proceed as predicted by the intermolecular attractive forces of hydrophobic, electrostatic, coordinate covalent bond, and Lewis acid–base interaction.

Biochemical surfaces contain covalently bound macromolecules such as antibodies, DNA, enzymes, receptors, ligands, and lectins. These surfaces facilitate the study of biomolecule–biomolecule interactions, as well as providing a means for specific sample extraction and enrichment, or on-chip enzymatic modification.

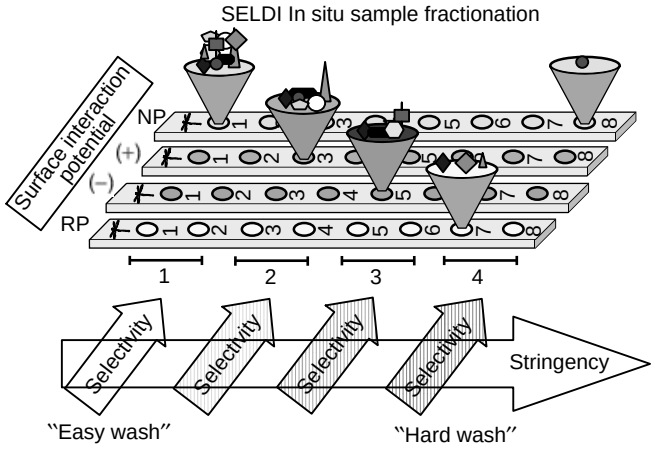

Figure 31 SELDI in situ sample clean-up. See text for details.

The process of SELDI sample fractionation for a number of chemical surfaces is demonstrated in Figure 31. A series of orthogonal SELDI surfaces are arranged in parallel array. Raw sample solutions are deposited upon each active spot. Each chip is washed with appropriate eluent in a gradient manner so that subsequent locations experience a greater degree of stringency, removing analytes with weak surface interaction potential and enriching those of strong surface affinity. When biochemical surfaces are employed, specific interactions are usually in play and a single surface is used with a gradient of wash conditions.

Independent of surface characteristics, SEAC functions as a solid-phase extraction technique that allows for sample clean-up and enrichment directly on the surface of the MS probe. This direct approach has a number of advantages when compared to MALDI or LDI. Because analyte is retained as dictated by surface affinity, retained compounds are driven to a state of physical–chemical homology. Under these conditions, the probability of analyte signal suppression is reduced when compared to MALDI analyses. Furthermore, in situ clean-up effectively diminishes sample loss by eradicating the effects of excessive dilution or nonspecific binding to sample purification vessels. Finally, because surface interaction potential is dependent upon physical–chemical properties such as hydropathicity, total charge, p*I*, phosphorylation, glycosylation, and primary composition, a priori knowledge of the analyte is provided when compounds are retained to a given surface under a specific set of wash conditions.

SEND is a process by which analytes may be desorbed and ionized without the need for the addition of matrix. This is accomplished by attachment of an energy-absorbing compound to the probe surface using the process of covalent modification or physisorption.[340] Compared to SEAC, this technology is still in its germinal stages and, with the possible exception of small molecule analysis, has demonstrated limited utility. SEPAR is a hybrid combination of SEAC and SEND, where the affinity capture device also functions as an energy-absorbing molecule promoting analyte desorption and ionization. Like SEND, this technology is also in its infancy.

After purification upon the SELDI surface, analyte detection is achieved by the use of a TOFMS system equipped with a laser desorption ion source. In the case of SEAC analysis, a matrix solution is added to promote laser-based ionization and desorption. SEND and SEPAR surfaces are analyzed directly.

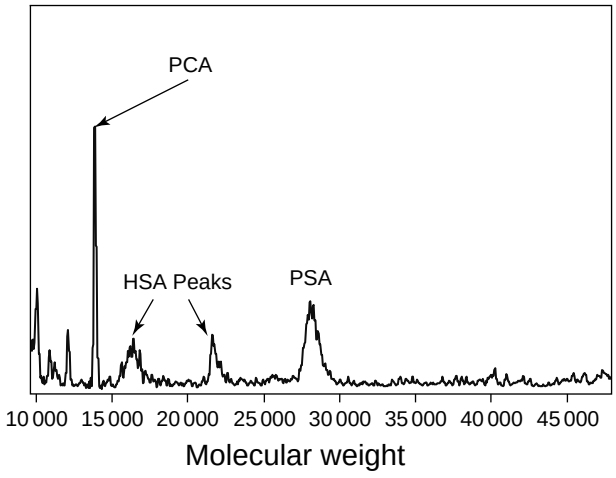

Figure 32 SEAC antibody isolation of PCA and PSA from human serum. Human serum albumin (HSA) pseudomolecular ions of $(M + 3H)^{3+}$ and $(M + 4H)^{4+}$ are also detected. (Unpublished data courtesy of Ciphergen Biosystems, Palo Alto, CA, USA.)

13.2 Surface-enhanced Laser Desorption/Ionization Applications

Because of the similarities between MALDI and SELDI, SELDI applications encompass all of the analytical possibilities for MALDI while providing improved performance in many areas. Unlike MALDI, the analysis of small molecules is facilitated by the use of SEND surfaces.[340] Because no matrix is added, SEND analyses have a significantly diminished level of chemical noise in the low-molecular-weight domain, facilitating the detection of analytes with $m/z < 1$ kDa. Recently, SEND-like applications on the surface of modified porous silicon have been demonstrated.[334]

SEAC applications have been applied to the extraction of analyte from complex biological fluids. Figure 32 displays a multiplex antibody extraction of human prostate cancer antigen (PCA) and prostate-specific antigen (PSA) from serum. Unlike other immunoassay techniques, SEAC allows for simultaneous analysis of different antigens, because detection is based upon m/z assignment and not some bulk property such as radioactivity, fluorescence, or luminescence. A copper IMAC chip has been used to extract histidine-rich amino acids from fetal calf serum at concentrations approaching 50 fmol μL^{-1} (see Figure 33). As demonstrated here, the modified SEAC surface does not perturb mass accuracy. Resolution for this analysis was of the order of 2500 m/dm.

A more general application of SEAC involves differential protein display for phenotypic investigation in the arenas of functional genomics and biomarker discovery. Biomarkers are biological compounds indicative of a particular physiological or pathological state. In many instances, these biomarkers are proteins or protein conjugates. During the initial discovery phase, protein profiles for control and experimental groups are compared in an attempt to elucidate statistically significant differences in protein expression. Various biological samples are analyzed using the previously described SELDI purification scheme, and the protein profiles detected for each group are contrasted. Reliable correlations between

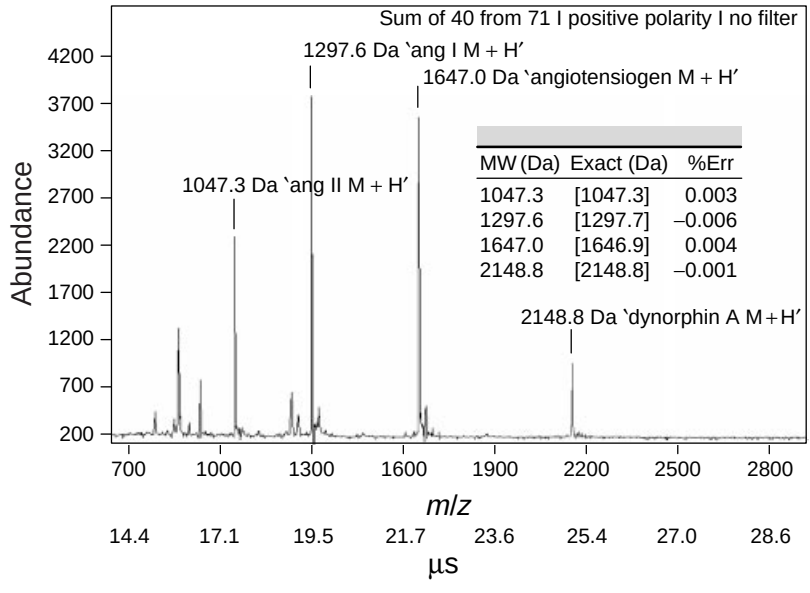

Figure 33 IMAC for Cu extraction of antiotensin I and II, angiotensinogen, and dynorphin from fetal calf serum. Fetal calf serum was spiked with 50 fmol μL^{-1} of each peptide. See text for further details. (Unpublished data courtesy of Ciphergen Biosystems, Palo Alto, CA, USA.)

disease state or biological activity are made with upward or downward expression of a single or multiple protein signals. These proteins become biomarker candidates and the process of characterization begins.

Biomarker candidates are characterized by on-chip application of various chemical or enzymatic strategies to fragment the protein into constituent peptides in order to identify post-translational modifications, provide primary sequence information, and ultimately elucidate protein identity. As a typical starting point, proteins are reduced chemically and then exposed to endoprotease digestion. The resultant peptides may also be treated with dephosphorylating, deglycosylating, or exoprotease enzymes. At each step along the way, MS detection is performed to monitor changes in the peptide m/z profile. Observed m/z changes are reconciled with known enzymatic or chemical modes of action to provide structural insight. Peptide profiles of biomarker digests are used to mine protein and cDNA databases for possible matches.

Figure 34 depicts an on-chip isolation of a possible prostate cancer biomarker. Ion signals have been

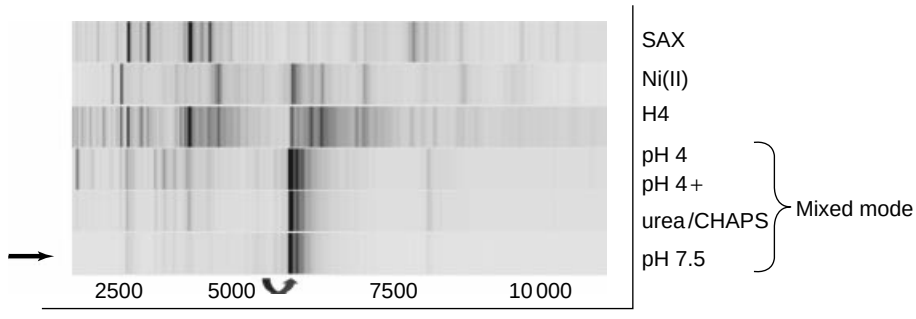

Figure 34 SELDI ProteinChip® Array isolation of possible prostate cancer biomarker. Enrichment to near purity was achieved using a mixed mode surface and high pH wash conditions. See text for further details. SAX, strong anion exchanger; H4, reverse phase surface; CHAPS, 3-([chloroamidopropyl]dimethylammonia)-1-propane sulfonate. (Unpublished data courtesy of Ciphergen Biosystems, Palo Alto, CA, USA. Samples courtesy of Dr George L. Wright, Virginia Prostate Cancer Center, University of Virginia School of Medicine.)

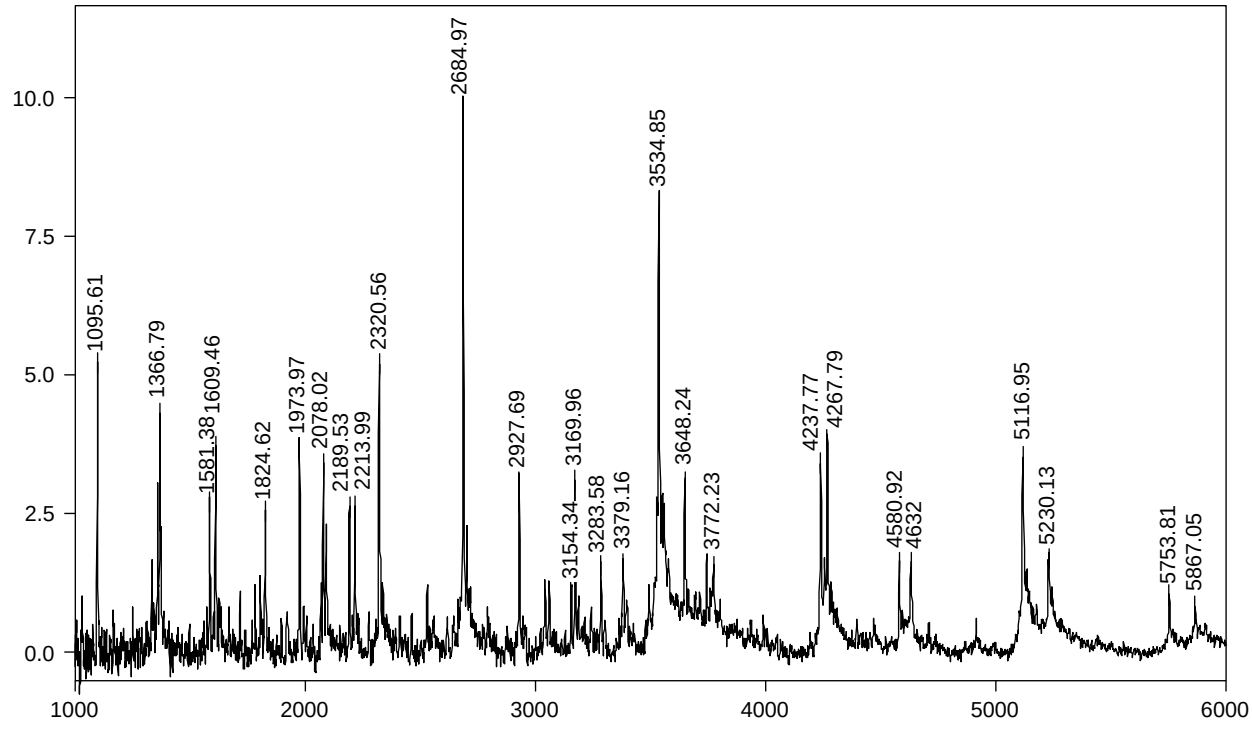

Figure 35 On-chip tryptic digest of enriched prostate cancer biomarker candidate. See text for details. (Unpublished data courtesy of Ciphergen Biosystems, Palo Alto, CA, USA.)

converted to a virtual gel display for viewing simplicity. Near-homogeneous purification of the biomarker candidate is achieved by washing the surface of a mixed-mode chip with a buffer solution at pH 7.5. Figure 35 demonstrates the result of an on-chip tryptic digest of the enriched candidate. Peptide signals were used to mine protein and cDNA databases. Combining these results with data generated by collisional-induced dissociation MS/MS analyses allowed for unambiguous identification of the marker as seminal basic protein. Validation studies for confirming the utility of seminal basic protein as a prostate cancer biomarker are currently under way.

14 TANDEM TIME-OF-FLIGHT MASS SPECTROMETRY

TOF analyzers have been combined in tandem fashion with a plurality of analyzers, including magnetic sector, electrostatic analyzer, ion trap, quadrupole, and TOF devices. A thorough review of this subject is provided by Cotter.[341] For the most part, TOF tandem arrangements have been created in order to provide structural information of biopolymers. Characterizations of post-translational modifications and primary sequence determination studies have been performed. A didactic review of MS amino acid sequencing of proteins is provided by Biemann and Payayannopoulos.[342]

Protein characterization, as performed by MS/MS, primarily requires initial enzymatic or chemical cleavage into peptides. MS-based fragmentation of peptides then proceeds through unimolecular decay or collision-induced mechanisms. Regardless of mechanism, peptide fragments are classified in terms of parent ion cleavage location and product ion charge retention. Medzihradszky and Burlingame provide a detailed discussion

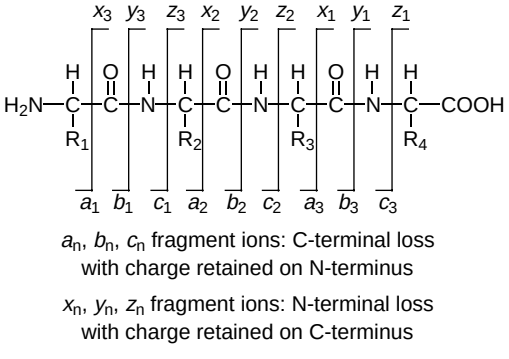

Figure 36 Peptide/protein MS-based fragmentation nomenclature. See text for details. (Figure courtesy of Dr Robert S. Brown, Utah State University, Logan, Utah, USA.)

of this nomenclature along with postulated reaction mechanisms.[343]

Figure 36 outlines the nomenclature involved in labeling peptide/protein MS-generated fragments. Two primary classes of ions exist: *x*,*y*,*z* and *a*,*b*,*c*. If ion cleavage results in charge retention on the N-terminus, the ion is of the *a*,*b*,*c* class. If charge retention is located on the C-terminus, the ion is of the *x*,*y*,*z* variety. Cleavage between the carbonyl group and an N-terminal α-carbon results in the creation of either *x* or *a* ions. Cleavage at the amide bond results in the creation of *b* or *y* ions, whereas cleavage between the amine and C-terminal α-carbon creates *z* or *c* ions.

This section further examines tandem time-of-flight mass spectrometry (TOFMS/MS) analysis in terms of ISD, PSD, and collision-induced dissociation (CID).

14.1 In-source Decay

Although not a true form of MS/MS, ISD analysis does provide structural information for purified compounds. ISD is a prompt decay process for which parent ions fragment into products while remaining in the ion optic region. In this manner, product ions and intact parent ions are accelerated to the same final kinetic energy, allowing direct analysis and accurate MW determination for all species when using a simple, linear TOF device. ISD in TOF applications has been reported for PDMS,[39] MPI,[46] and MALDI.[77,344–350] Currently, ISD is mostly used in conjunction with MALDI analysis, and, as such, subsequent discussion will focus upon this application.

Matrix-assisted laser desorption/ionization/in-source decay (MALDI/ISD) analysis is best performed using a TLF device. During the lag period, ions undergo prompt decay, and are subsequently accelerated out during the ion extraction phase. For the most part, *c* and *y* series ions are generated. As fragment MW increases, predominantly *c* ion fragments are seen.[38,344]

The problem in performing ISD analysis is that there is no means to select the precursor or parent ion. Consequently, product ions form from all species, and if the sample is impure, interpretation becomes complicated. Even so, ISD has shown reasonable utility in the sequencing of a number of purified peptides,[349] as well as identifying disulfide-linked peptides.[347] In contrast to other forms of MS/MS, ISD does demonstrate reasonable efficiency in the direct analysis of large peptides and proteins.[346,350] Figure 37 depicts a series of c-ion fragments generated during direct ISD analysis of BSA (MW = 66 430 Da).

14.2 Post-source Decay

Like ISD, PSD is a unimolecular decay process, but on a longer time frame. PSD occurs typically several

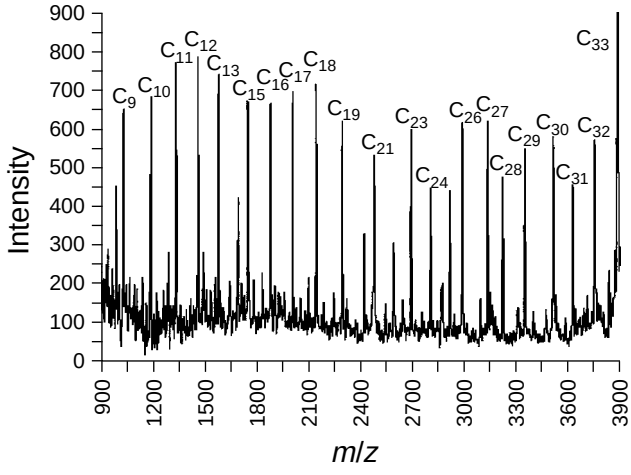

Figure 37 ISD spectrum of BSA. Predominantly *c* series ions are seen. (Figure courtesy of Dr Robert S. Brown, Utah State University, Logan, Utah, USA.)

microseconds after ion creation. In this fashion, fragmentation takes place during free flight, after ions have been accelerated out of the source region. PSD has been reported for SIMS,[10] MPI,[43] and MALDI/TOF MS.[103,104,265,348,351–366] Again, the focus of this section is on matrix-assisted laser desorption/ionization/post-source decay (MALDI/PSD) analysis.

MALDI/PSD analysis is typically performed using a reflectron-type TOFMS (see Figure 38). Precursor ion selection is accomplished using an electrostatic ion gate. Fragment ions enter the reflectron and get turned about according to their entrance energies. Because metastable decay occurs during free flight, product ions and neutrals have the same velocity as parent ions but less mass. Accordingly so, the total kinetic energy of these fragments can be considerably less than that of their parent ion. Although reflectrons function to compensate for variations in total kinetic energy, most linear field reflectors do not adequately focus ions over a broad energy range. Consequently, PSD analysis is generally performed by scanning the reflectron through a series of applied voltages. Because reflectron ion deceleration and acceleration are voltage dependent, a separate calibration function must be derived for each scan. Linking the results of each scan together in histogram fashion creates a composite MS/MS spectrum. This approach was first used by Schlag for MPI analysis[43] and was later applied to MALDI by Spengler.[363,364]

Compared with linear field reflectrons, the curved field reflectron of Cotter has a broader energy focusing range and can be used without scanning (see Figure 38).[105] This approach greatly simplifies PSD analysis, because a single spectrum provides an entire PSD/MS/MS profile. This device has been used for CID analysis,[104] peptide sequencing,[352] as well as generating fragment information on atomole levels of small peptides.[366] Recently, this curved-field reflectron PSD has been combined with carboxypeptidase activity to create what has been termed as nested PSD spectra (see Figure 39).[357] Intact peptides are simultaneously analyzed with carboxypeptidase products. In this manner, additional C-terminal fragments are created, improving the overall efficiency of fragment yield and sequence call.

PSD peptide fragments are generally of the *a*, *b*, and *y* variety. Additionally, side-chain fragments as well as fragments with the addition of water and losses of ammonia have been reported.[351,360,362,363] For the most part, efficient PSD analysis requires at least 500 fmol of peptide. The overall complexity of interpreting these spectra combined with the relatively low efficiency of this process has limited its general applicability. For the most part, collision-induced dissociation/tandem mass spectrometry (CID/MS/MS) approaches provide easier to interpret results with significantly lower sample requirements.

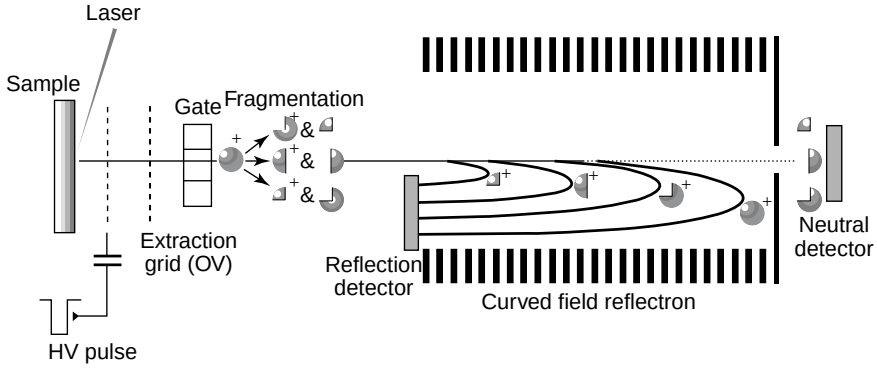

Figure 38 A curved-field reflectron post-source decay/tandem mass spectrometry (PSD/MS/MS) device. See text for details. (Figure courtesy of Kratos Analytical, Manchester, UK.)

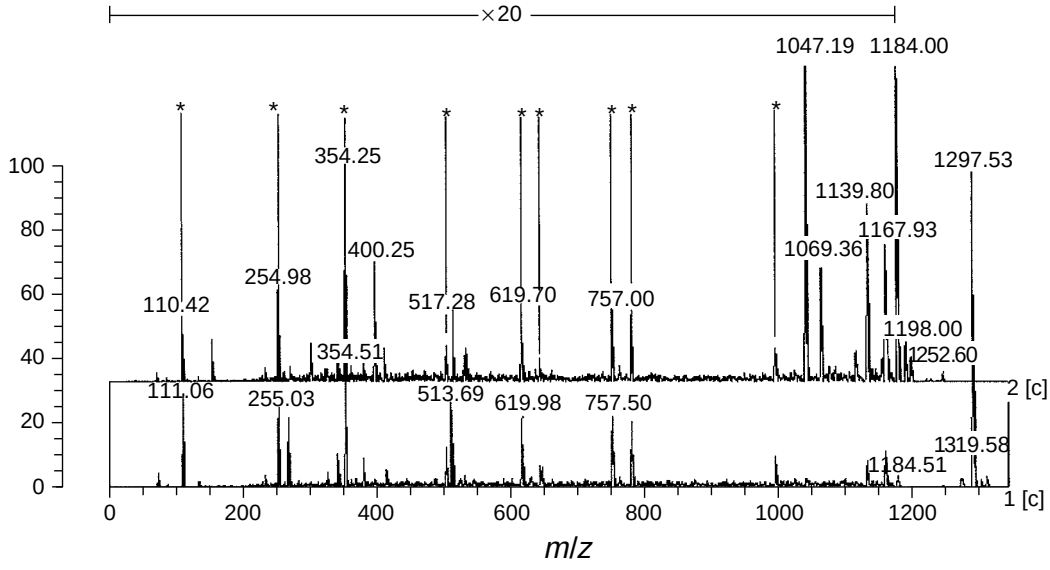

Figure 39 Nested PSD spectrum of human angiotensin I treated with carboxypeptidase *y*. The upper spectrum reveals additional C-terminal peaks due to elimination of the C-terminal amino acid leucine. (Figure courtesy of Kratos Analytical, Manchester, UK.)

14.3 Collision-induced Dissociation/Tandem Mass Spectrometry Techniques

Unlike ISD and PSD, CID is a bimolecular process where ions generally collide with collision gas molecules (see Figure 40). These collisions occur within a specialized region known as a collision cell. Collision cells have been bordered by electrostatic lenses,[353,356,367–369] formed within ion trap bodies,[370,371] and RF quadrupole arrangements.[4,55,58] Low-energy CID occurs with collision energies typically less than 200 eV. High-energy collisions usually involve collision energies of 1000–10 000 eV.[343]

CID fragmentation produces products that conserve mass and charge. Singly charged parent ions fragment into a neutral and product ion, whereas doubly charged parents create a single doubly charged product ion and neutral or two product ions. For the most part, *b* and *y* ions are formed. This process is generally more efficient than unimolecular decay mechanisms, often providing more complete fragmentation profiles and sequence information, even at femtomole levels of peptide. Parent ion selection is performed in MS_1 and product ion analysis is performed in MS_2. Tandem TOF approaches generally employ a TOF device as MS_2, taking full advantage of this system's scan speed and mass range.

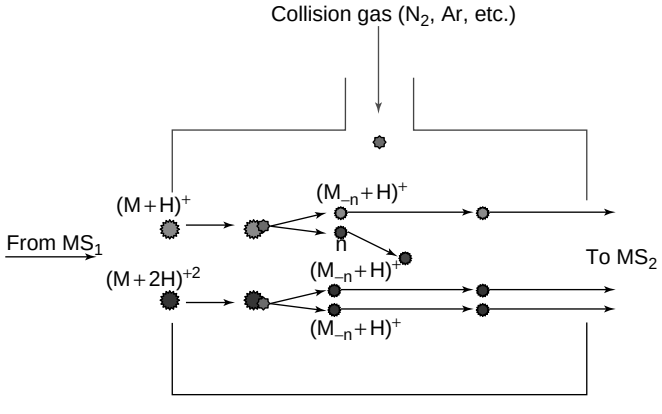

Figure 40 CID collision cell schematic. See text for details.

Three major tandem TOF geometries are currently showing promise for efficient CID/MS/MS operation: ion trap TOF, tandem TOF, and quadrupole–quadrupole TOF. Ion trap TOF has been applied to the analysis of ES I-[208,215,216] and MALDI-formed ions.[370,371] For the most part, tandem TOF analyzers have been used for MALDI analysis.[353,356,367–369] No commercial versions of ion trap TOF or tandem TOF devices exist today.

Quadrupole–quadrupole TOF devices have been applied to ESI[129,372–374] and MALDI applications.[375,376] Commercial versions of ESI quadrupole–quadrupole TOF are presently available from Micromass (Manchester, UK) and PE-Sciex (Concord, Canada). An LDI quadrupole–quadrupole TOF device developed by Krutchinsky, Loboda, Spicer, Dworschak, Ens, and Standing has recently demonstrated significant MS/MS efficiency for the MALDI analysis of human angiotensin I (see Figure 41). The overall signal quality for the MS/MS analysis of 4 fmol of angiotensin I was found to be similar to that for 4 pmol.

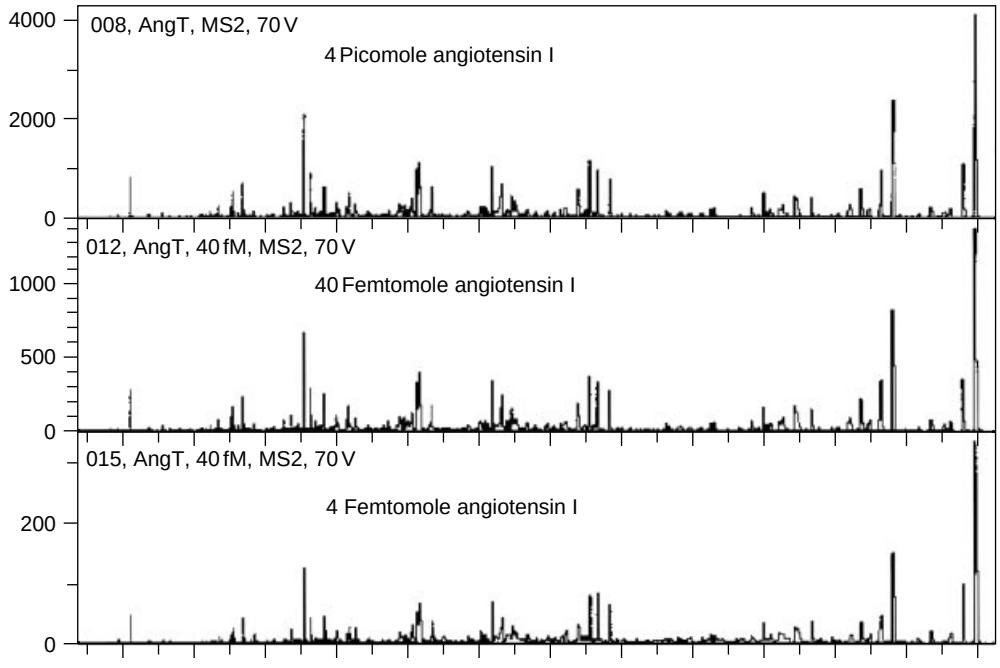

Figure 41 Analysis of angiotensin I by quadrupole–quadrupole TOFMS coupled with LDI and CID. See text for details. (Unpublished results from Scot R. Weinberger (Ciphergen Biosystems, Inc., Palo Alto, CA), Alexander Loboda, and Werner Ens (University of Manitoba, Canada).)

15 TIME-OF-FLIGHT SECONDARY ION MASS SPECTROMETRY

When a beam of energetic primary ions is directed onto and impinges upon a solid surface, atoms and molecules of the surface are released. This process is known as sputtering and it defines the fundamentally destructive nature of this method of analysis. A small fraction of the atoms and molecules released from the surface carry a charge. These ''Secondary Ions'', when mass analyzed, form the basis for the technique known as Secondary Ion Mass Spectrometry or SIMS for short.

The development of this technique has generated two separate application areas. In the first, use is made of the phenomenon known as sputtering. Here, the primary ion beam has two purposes: to erode the atomic surface in a controlled way and therefore investigate the atomic composition of the surface as a function of depth; and to maximize the efficiency of production of positively charged secondary ions by producing the primary ion beam from a chemically aggressive atomic species such as oxygen. Cesium primary ion beams similarly enhance the yield of negative secondary ions. This technique, which has become known as dynamic SIMS depth profiling, has established itself as the standard method for the determination of the concentration profile of semiconductor dopants.[377] It can deliver accurate concentrations from the 1% level down to parts per billion over a range of depths down to the micron level. However, the property of the technique that gives rise to this extraordinary atomic sensitivity has a side-effect: the chemically active primary ions become incorporated into and thereby alter the chemical nature of the surface. Although this phenomenon may be tolerable to the semiconductor analyst, it can be a major disadvantage to the surface chemist. Surface analysis thus represents the second area of the applications of SIMS referred to above.

15.1 Concept of a Maximum Dose

Consider a monolayer of surface. Each square centimeter consists of approximately 10^{15} molecules. In order that the technique is regarded as nondestructive to the surface as a whole, only about 1% of the surface molecules may be interrogated. This gives a maximum number of molecules available for analysis by SIMS of 10^{13}. If one incident primary ion removes a maximum of ten molecules then the maximum permitted primary ion dose is $\sim 10^{12}$ primary ions for one square centimeter of surface. Suppose that one cares to examine the surface by taking account of the spatial distribution of chemistries to a minimum feature size of $\sim 10\,\mu m$. Now the maximum dose has reduced to only 10^6 primary ions. At a continuous primary ion current of 1 nA this represents a maximum experimental analysis time per pixel of 160 μs. Clearly any type of scanning mass spectrometer is going to find

the task of collecting the full mass range in this short time impossible should more than a single m/z be analyzed at any time. However, with a TOF mass spectrometer the situation is different because the secondary ions are released during a short time period of $1 \sim 100$ ns when the primary ion beam is pulsed. There then follows a period of up to 100 µs during which the flight times of all the secondary ions are measured. Thus mass analysis of the complete spectrum from that pulse point occurs before the next primary pulse arrives. The duty cycle, i.e. the fraction of the time that the primary ion beam is playing on the sample, is typically only 10^{-4}. The mass spectral information from each image pixel is summed and stored within the data system. The efficiency of collection of mass spectral information is therefore at a maximum. This simple example illustrates the main application area of surface chemical imaging where TOFMS has made a fundamental impact. Scanning mass spectrometers (mainly quadrupoles) have been used satisfactorily to produce large-area (1–2 mm) chemical spectra of surfaces since the early 1980s but the requirements for spatial information have led to dramatic developments in TOF instrumentation for SIMS, which we shall now examine in detail.

15.2 Instrumentation

15.2.1 Primary Ion Probes

First let us look at the technologies now existing for producing finely focused primary ion probes and the way they have been developed for TOFSIMS. The brightest of all ion sources for laboratory use are the field emitters.[378] This family of ion sources merits some discussion because virtually all modern TOFSIMS instruments utilize this type of source to realize the highest performance for surface imaging. The types of field-emitting ion sources used commonly are known as liquid-metal ion sources because they use a supported liquid-metal film as the field-emitting electrode. To understand how this comes about, it is instructive to consider the evolution of the emitter as the potential applied to the source is gradually raised. When the potential on the emitter is zero the liquid metal film conforms to the underlying electrode structure, as shown in Figure 42(a). As the potential on the emitter is raised, at first no current is observed. Only as the potential on the emitter exceeds a value of several kilovolts is emission observed to commence suddenly. This onset voltage coincides with a spontaneous change in the geometry of the liquid film into a cone shape (Figure 42b). This is known after its discoverer as a "Taylor Cone" and it represents a geometry where surface tension and electrostatic forces are in equilibrium. The apex of the cone evolves to a point where its dimensions are atomic and here the cohesive forces acting between atoms in the liquid are reduced and the liquid is virtually torn asunder by the immense local electric field. At the same time electrons in the metal atoms can tunnel back into the liquid cone, leaving positive ions of metal atoms all formed within atomic dimensions (see Figure 42c). These ions form a diverging beam but, because they were formed in a region of a few nanometers, they can be brought back into focus using standard ion optical techniques to give a probe of ~50 nm in extent. For technological reasons

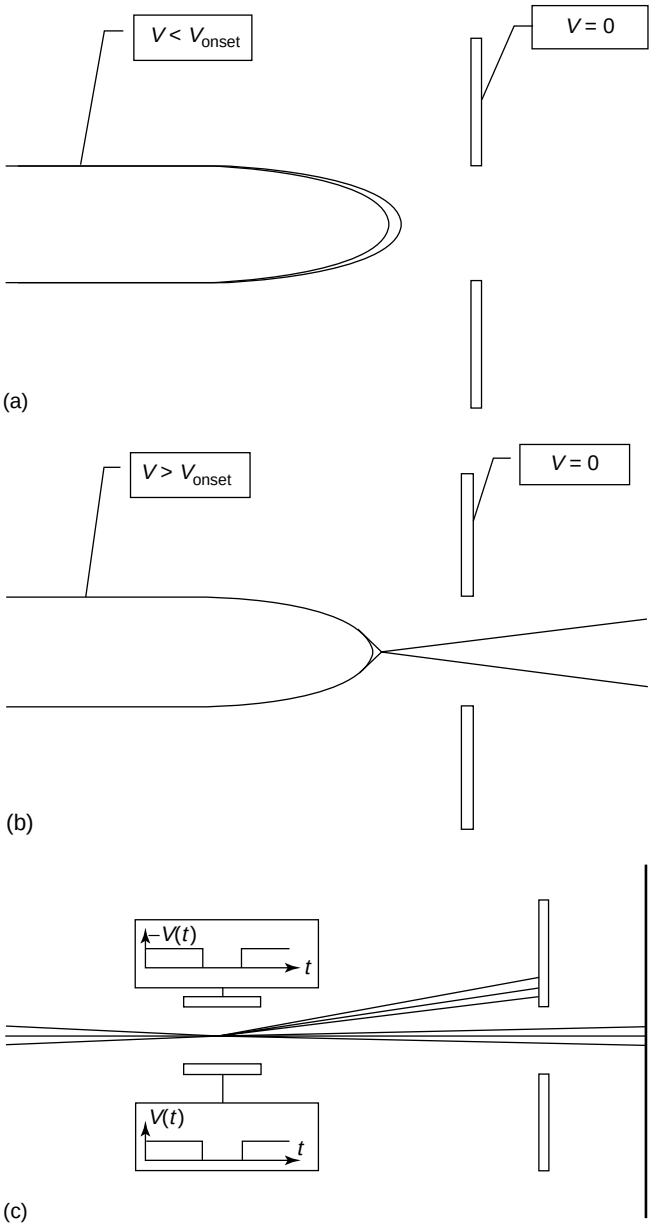

Figure 42 Evolution of the Taylor Cone as the applied voltage exceeds the onset voltage and the metal film suddenly forms a liquid cone and emits ions. See text for details. (Figure courtesy of Steven Thompson, Scientific Analysis Instruments Ltd, Manchester, UK.)

gallium metal is most commonly used in these sources but the move to more massive primary ion probes has rekindled interest in gold and bismuth.

Other primary ion sources used in TOFSIMS instruments are surface ionization sources for cesium and electron bombardment ion sources for inert gas ions. These also use conventional ion optical techniques to produce focused beams of ions. One ion optical technique that is particular to TOFSIMS concerns the production of short pulses of ions or ion packets. The design constraints on these ion packets are considerable: their spatial extent should not exceed that of a continuous beam, or spatial resolution will be lost; and their temporal extent has a major impact on the ultimate mass resolution available from the instrument because it directly influences the time period over which secondary ions are generated. Short pulses of ions are produced if the continuous beam enters a pair of electrostatic deflection plates energized with a potential of a few hundred volts such that the beam is deflected from the ion optical axis. If the potential on the plates is suddenly brought to zero and then returned to the original value, the beam will be brought back to the ion optical axis for a short time before being returned to the deflected state. If the duration of this voltage pulse is varied, then the length of the ion pulse can be controlled. Furthermore, if the ion beam forms a crossover within the deflection plates, this method of ion pulse blanking can be achieved without apparent movement of the ion probe. This so-called ''motionless blanking'' is illustrated in Figure 43.

Other schemes for the production of ion packets have been used successfully.[379] If the beam is directed into a 90° ion mirror then the beam will only emerge along its chosen trajectory when a suitable deflecting potential is suddenly applied to the mirror electrode. Again, varying the duration of the applied potential can control the pulse width. These techniques can produce ion pulse widths in the range 3–100 ns. To achieve the highest mass resolution available from the TOF analyser, these pulse widths are still too long. The only option then is to use a method of electrodynamic pulse compression or ''bunching''.[380]

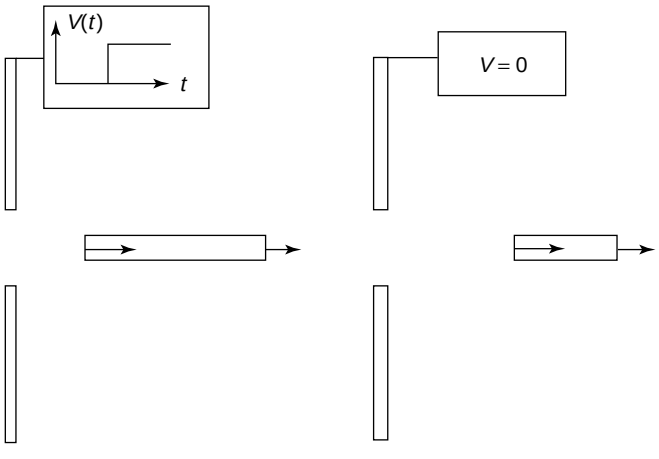

Figure 43 Motionless beam blanking by suddenly reducing the potentials on the plates to zero. See text for details. (Figure courtesy of Steven Thompson, Scientific Analysis Instruments Ltd, Manchester, UK.)

Bunching involves admitting a pulse of ions into a chamber consisting of two electrodes separated by a distance somewhat greater than the spatial extent, δz, of the pulse of ions while they are in flight. This is given from the product of the drift velocity of the ions, v, and the temporal pulse width, δt. At the instant the ions are between the two electrodes, the rearmost electrode suddenly has its potential raised by a few kilovolts, thus generating an electric field between the electrodes. The ions within the pulse gain an amount of energy from the field that depends upon their positions when the field was suddenly applied. So the ions at the rear of the pulse emerge from the chamber with a drift velocity somewhat greater than the ions at the front of the pulse. As the pulse of ions proceeds downstream toward the sample under analysis, it will become bunched up in time. The skill comes in ensuring that the pulse has its minimum width as it hits the sample. By techniques such as this, ion pulse widths as short as 800 ps can be achieved. The disadvantage of this method is that the pulse compression introduces a large chromatic spread into the beam. The chromatic aberration of the probe-forming lens limits the ultimate spatial resolution of the probe to a few microns.

When the ion source uses surface ionization of cesium as the primary ion, then mass filtering of the primary ion beam is not necessary. In general, however, it must be considered. Nonmass-filtered primary ion beams can lead to minor isotopes or impurities being admitted through the primary ion column. These will have drift velocities that differ from the primary species. Once an ion pulse packet is formed, dispersion in time will occur and pulse broadening will result, leading to a loss of resolution in the analyzer. This is a particular problem when using gallium.

Before moving on to consider the analyzer it is necessary to consider recent developments in the use of massive primary ions. As the primary ion impinges onto the sample surface it transfers its momentum into the lattice. Locally the lattice is in considerable turmoil but this diminishes as one moves away from the impact site. It is known that intact molecular ions are released several atomic diameters away from the impact site. As the impact site is approached and the extent of lattice vibration increases, then only molecular fragment ions are released. As the momentum transferred to the lattice is increased, by the use of high-mass primary ions, then the zone of formation of intact molecular

ions increases disproportionately. So the use of massive primary particles increases the relative yield of intact molecules from the surface. Originally, this led to the use of fission fragments from radioactive sources but recently the trend has been toward cluster sources from gold, Au_3^{2+}, and carbon, C_{60}^{+}.

15.2.2 Mass Spectrometers

The properties of TOF mass analysers, which collect the secondary ions, are in many ways more straightforward than the primary probes that deliver them. The first stage of the mass spectrometer invariably consists of an accelerating field to collect the secondary ions. Two electrodes, the first of which is the electrode onto which the sample is loaded, form this accelerating field. The second electrode will form an annulus about the ion optical axis such that the secondary ions will be swept through the central aperture and on to the next stage of the mass spectrometer. Frequently this second electrode will have further apertures disposed at an angle of 30–40° to the ion optical axis to permit entry of the primary ion probes and a charge-neutralizing electron beam. The secondary ions are released from the surface with a kinetic energy of several electron-volts and a range of angles about the ion optical axis, so even after acceleration they form a cone of diverging ion trajectories as they pass out of the accelerating region. Typically they are then collected and focused into a well-collimated beam by an electrostatic lens. The secondary ions can then be left to drift in field-free space until they are detected. This simple type of mass spectrometer has two disadvantages. First, the small energy spread of the secondary ions cause temporal dispersion of the secondary ion packet as it travels down the drift tube, leading to a broadening of the detected signal and a loss of mass resolution. Second, the length of the flight tube sets a limit on the drift length of the ions after acceleration; again, this affects the mass resolution available.

If the ion trajectories could be folded back on themselves by an ion mirror or ''reflectron'' placed at the opposite end of the flight tube from the acceleration region, then the effective pathlength of the secondary ions could be greatly increased.[95] This can be achieved by mounting a set of annular electrodes with suitable potentials in order to produce a constant retarding electric field on or near the ion optical axis. If the reflectron is mounted at a small angle to the ion optical axis, the ions leaving the reflectron and moving in the opposite direction to the ions entering will have diverging trajectories. This simplifies detection. The use of a reflectron has two benefits: it extends the effective pathlength of the secondary ions for a given flight tube and it provides a first-order correction for the ion energy spread. These benefits significantly improve the mass resolution available from the instrument.

Although the ion mirror reflectron is a popular device for the correction of ion energy spreads, other devices have been used. A remarkable instrument manufactured by the Physical Electronics Company (Eden Prairie, MN, USA) uses three 90° cylindrical analysers arranged in a way that not only provides a flat focal plane of time focus at the detector but also has spatial focusing properties. This means that the instrument can be used as a chemical imaging microscope where the spatial resolution is independent of the properties of the primary probe. The primary beam can therefore be set-up for the minimum pulse width to realize the maximum mass resolution from the analyser while still giving spatial resolution of the order of 10 μm.

15.3 Charge Neutralization[381]

Many of the samples of interest to the surface chemist are likely to be poor electrical conductors, and yet the sample has to form one of the electrodes of the extraction system supporting an accelerating electric field of several kilovolts per centimeter. Furthermore, many secrets of the surface chemistry can be unraveled by studying both the positive and the negative secondary ions released from the surface. The clue to balancing the surface charge during analysis can be determined from other electric techniques such as X-ray photoelectron spectroscopy. It is known that effective charge balancing can be achieved by generating a cloud of very-low-energy electrons in the near-surface region of the sample. In some cases this cloud is confined by an axial magnetic field. Any change in the surface potential will result in a shift of the photoelectron peak. TOFSIMS instruments employ a strong extraction field for efficient collection of the secondary ions, so it is virtually impossible to direct a suitably low electron beam at the sample when the extraction field is applied. The technique used is to turn off the extraction field after every pulse of the primary ion beam, while the secondary ions are dispersing in time within the spectrometer. Then a short pulse of electrons is directed at the sample, after which the extraction field is restored and the next cycle commences.

If there is one type of application, that may be described as routine for TOFSIMS then it concerns investigation of the surface chemistry of synthetic polymers.[382] There exists a considerable body of literature, that has been collected and collated into a series of handbooks. These discuss the interpretation of spectra in terms of the polymer properties, such as end groups, cross-linking and branching. The imaging capabilities of TOFSIMS are valuable when investigating fibers and composites.

Some of the most beautiful applications of TOF-SIMS have been demonstrated by the Physical Electronics Company (Eden Prairie, MN, USA) using the TRIFT II TOF imaging microscope.[383] These applications have explored the full benefits of the instrumentation. Before looking at some of these examples, it is important to emphasize the requirement for sample preparation. Often the surface of interest is an internal surface requiring that the sample be mounted and then sectioned by a microtome in order to preserve the chemistry. Biological samples can, in principle, be analyzed using these methods but in this case the microtoming may have to be accomplished at low temperature and the sample kept at low temperature during the analysis. Such low-temperature sample-manipulating devices are available presently on TOFSIMS instruments but it serves to emphasize that the sample preparation can often take much longer than the analysis.

Clearly such instruments can be used to solve special problems as well as for routine use; because of this it is instructive to look at two or three different examples that represent the state of the art rather than discuss standard methods. Surface contamination of silicon wafers can be particularly difficult to detect, partly because the very high purity of materials used in this industry demands a low detection limit and partly because the regions of a silicon wafer available for analysis by a destructive technique like SIMS are so restricted. When the impurity of interest is iron, nominal mass 56 Da, this mass interferes with the abundant molecule $(^{28}Si_2)^+$ and it all looks like a lost cause. Fortunately, $^{56}Fe^+$ has a significant mass defect of 7 mDa, whereas for $(^{28}Si_2)^+$ it is only 4 mDa; all organic interferences at that nominal mass have a positive defect.

Using the technique of electrodynamic pulse compression on the primary ion pulse to achieve a pulse duration of less than 1 ns, the spectrometer was able to resolve $^{56}Fe^+$ from $(^{28}Si_2)^+$, demonstrating a mass resolution, m/dm of ~5000. The area of the wafer sampled was only 40 μm × 40 μm. From the accumulated count rate and relative sensitivity measurements, it was possible to determine the surface concentration of iron to be at a level of less than 1×10^{10} atoms cm^{-2}.

A second example further demonstrates the versatility of a good TOFSIMS instrument as a chemical microscope. Drug delivery systems frequently contain the active ingredient within a polymeric coating so that the release time of the active ingredient can be delayed following ingestion. These pellets form small spheres of 500 μm diameter held together by the coating, which is ~25 μm thick. The pellet was mounted, cross-sectioned and finally introduced for analysis. Remarkably, the intact molecular ion of the active ingredient, Metoprolol (MW of M + H = 268 Da), was detected easily. Furthermore, the chemical image of the molecule could be seen surrounding a silicon core and extending uniformly out to the coating characterized by an ion map at 59 Da. This was later identified to be a fragment of ethylcellulose, the coating material. In this application, TOFSIMS has shown all the attributes for a true imaging chemical analyzer. The analysis proceeds directly on electrically insulating substrates, the charge balance being maintained by the use of a pulsed low-energy source of electrons. Detection of the intact molecule provides direct proof of the spatial distribution of specific organic species.

16 FUTURE ADVANCEMENTS AND CONCLUDING REMARKS

Most certainly, the recent trend of technical advances in TOFMS will continue in the years to come. TOF devices will become simpler to use and smaller in total size. Field-portable devices may ultimately be realized. TOF will continue to play an increased role in the realm of bioscience investigations. Improvements in sample preparation and SELDI technology will create even greater access to biological problems. SELDI-based bioassays may eventually find utility in research and clinical applications

In terms of TOFMS technical innovations, future TOF devices will most likely obviate the requirement to adjust TLF parameters as a function of m/z, while improving the accuracy of external standard measurements. The latter may be achieved through refinements in dynamic pulse focusing and/or orthogonal ion extraction. Improvements in sensitivity, especially for large ions, are also expected. Progress in CID/MS/MS technology, such as ion trap TOF, tandem TOF, and quadrupole–quadrupole TOF, will result in commercial devices compatible with ESI and MALDI ionization.

Indeed, the next millennium will see an increased role for TOFMS analysis. This technology will make significant contributions to the advancement of surface science, life science, elemental analysis, as well as inorganic and organic chemistry.

DEDICATION

This article is dedicated to my lovely wife Carla and my unborn child, Joseph Anthony or Leah Arabella. Their indelible patience and steadfast support were a continued source of inspiration.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
APCI	Atmospheric Pressure Chemical Ionization
BSA	Bovine Serum Albumin
CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CE/ESI	Capillary Electrophoresis/ Electrospray Ionization
CE/ESI/TOFMS	Capillary Electrophoresis/ Electrospray Ionization/ Time-of-flight Mass Spectrometry
CI	Chemical Ionization
CID	Collision-induced Dissociation
CID/MS/MS	Collision-induced Dissociation/ Tandem Mass Spectrometry
CIE	Continuous Ion Extraction
DAG	Differential Acceleration Grid
DAR	Data Acquisition Rate
EI	Electron Ionization
EI/TOF	Electron Ionization/Time-of-flight
ESI	Electrospray Ionization
ESI/QIT/TOFMS	Electrospray Ionization/ Quadrupole Ion Trap/ Time-of-flight Mass Spectrometry
ESI/TOF	Electrospray Ionization/ Time-of-flight
fwhm	Full Width at Half-maximum
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GC/TOFMS	Gas Chromatography/ Time-of-flight Mass Spectrometry
HPLC	High-performance Liquid Chromatography
HPLC/MS/MS	High-performance Liquid Chromatography/Tandem Mass Spectrometry
HSI	Hyperthermal Surface Ionization
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICP/TOFMS	Inductively Coupled Plasma/ Time-of-flight Mass Spectrometry
IMAC	Immobilized Metal Affinity Capture
ISD	In-source Decay
LAMMA	Laser Microprobe Mass Analyzer
LC	Liquid Chromatography
LC/TOFMS	Liquid Chromatography/ Time-of-flight Mass Spectrometry
LDI	Laser Desorption Ionization
LDI/TOF	Laser Desorption/Ionization/ Time-of-flight
LIT	Linear Ion Trap
MAB	Metastable Atom Bombardment
MALDI	Matrix-assisted Laser Desorption/ Ionization
MALDI/ISD	Matrix-assisted Laser Desorption/ Ionization/In-source Decay
MALDI/PSD	Matrix-assisted Laser Desorption/ Ionization/Post-source Decay
MALDI/TOF	Matrix-assisted Laser Desorption/ Ionization/Time-of-flight
MCP	Microchannel Plate
MPI	Multiphoton Ionization
MPI/TOF	Multiphoton Ionization/ Time-of-flight
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MW	Molecular Weight
NADH	Nicotinamide Adenine Dinucleotide (Reduced Form)
PCA	Prostate Cancer Antigen
PD	Plasma Desorption
PDMS	Plasma Desorption Mass Spectrometry
PIE	Pulsed Ion Extraction
PSA	Prostate-specific Antigen
PSD	Post-source Decay
PSD/MS/MS	Post-source Decay/Tandem Mass Spectrometry
PSPF	Post-source Pulse Focusing
QIT	Quadrupole Ion Trap
RF	Radiofrequency
RMS	Root Mean Square
SEAC	Surface-enhanced Affinity Capture
SELDI	Surface-enhanced Laser Desorption/Ionization
SEND	Surface-enhanced Neat Desorption
SEPAR	Surface-enhanced Photolabile Attachment and Release
SIG	Secondary Ion Generator
SIMS	Secondary Ion Mass Spectrometry
SMB	Supersonic Molecular Beam
SMB/EI	Supersonic Molecular Beam/ Electron Ionization
S/N	Signal-to-noise Ratio
TDC	Time-to-digital Converter
TLF	Time-lag Focusing

TOF	Time-of-flight
TOFMS	Time-of-flight Mass Spectrometry
TOFMS/MS	Tandem Time-of-flight Mass Spectrometry
TOFSIMS	Time-of-flight Secondary Ion Mass Spectrometry

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Biomolecules Analysis: Introduction • High-performance Liquid Chromatography of Biological Macromolecules • Mass Spectrometry in Structural Biology

Chemical Weapons Chemicals Analysis **(Volume 2)**
Gas Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention • Gas Chromatography/Mass Spectrometry in On-site Analysis of Chemicals Related to the Chemical Weapons Convention • Liquid Chromatography/Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Gas Chromatography and Mass Spectrometry in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Automotive Emissions Analysis with Spectroscopic Techniques

Environment: Water and Waste **(Volume 3)**
Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Field-portable Instrumentation **(Volume 5)**
Mobile Mass Spectrometry used for On-site/In situ Environmental Measurements

Forensic Science **(Volume 5)**
Mass Spectrometry for Forensic Applications

Nucleic Acids Structure and Mapping **(Volume 6)**
Mass Spectrometry of Nucleic Acids

Peptides and Proteins **(Volume 7)**
Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis

Pharmaceuticals and Drugs **(Volume 8)**
Mass Spectrometry in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Mass Spectrometry in Process Analysis

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Artificial Intelligence and Expert Systems in Mass Spectrometry • Gas Chromatography/Mass Spectrometry • High-resolution Mass Spectrometry and its Applications • Liquid Chromatography/Mass Spectrometry • Quadrupole Ion Trap Mass Spectrometer • Secondary Ion Mass Spectrometry as Related to Surface Analysis • Tandem Mass Spectrometry: Fundamentals and Instrumentation

REFERENCES

1. W.E. Stephens, 'A Pulsed Mass Spectrometer with Time Dispersion', *Phys. Rev.*, **69**, 691 (1946).
2. D.C. Damoth, 'Recent Advances in Time-of-flight Mass Spectrometry', in *Advances in Analytical Chemistry and Instrumentation*, ed. D.C.N. Reilley, John Wiley & Sons, New York, 371–410, 1964.
3. W.C. Wiley, I.H. McLaren, 'Time-of-flight Mass Spectrometer with Improved Resolution', *Rev. Sci. Instrum.*, **26**(12), 1150–1157 (1955).
4. I. Chernushevich, W. Ens, K.G. Standing, 'Orthogonal-injection TOFMS for Analyzing Biomolecules', *Anal. Chem.*, **71**, 452A–461A (1999).
5. D.C. Schriemer, L. Li, 'Detection of High Molecular Weight Narrow Polydisperse Polymers up to 1.5 Million Daltons by MALDI Mass Spectrometry', *Anal. Chem.*, **68**, 2721–2725 (1996).
6. R.W. Nelson, D. Dogruel, P. Williams, 'Mass Determination of Human Immunoglobulin IgM Using Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 627–631 (1994).
7. J.C. Schultz, C.A. Hack, W.H. Benner, 'Mass Determination of Megadalton-DNA Electrospray Ions Using Charge Detection Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **9**, 305–313 (1997).
8. S.D. Furstenau, W.H. Benner, 'Molecular Weight Determination of Megadalton DNA Electrospray Ions Using Charge Detection Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1528–1538 (1995).
9. J.C. Schultz, C.A. Hack, W.H. Benner, 'Polymerase Chain Reaction Products Analyzed by Charge Detection Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **13**, 15–20 (1999).
10. X. Tang et al., 'A Secondary Ion Time-of-flight Mass Spectrometer with an Ion Mirror', *Int. J. Mass Spectrom. Ion Processes*, **85**, 13–67 (1988).

11. L. Schmidt et al., 'Gridless Ion Acceleration Systems for Time-of-flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 782–786 (1993).
12. M. Karas, F. Hillenkamp, 'UV Laser Desorption of Ions above 10 000 Daltons', in *International Mass Spectrometry Conference*, Bordeaux, France, 12, 1988.
13. M. Karas, F. Hillenkamp, 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Anal. Chem.*, **60**(20), 2299–2301 (1988).
14. K. Tanaka et al., 'Protein and Polymer Analyses up to *m*/*z* 100 000 by Laser Ionization Time-of-flight Mass Spectrometry', in *Proceedings of the Second Japan–China Joint Symposium on Mass Spectrometry*, Osaka, Japan, 37, 1987.
15. K. Tanaka et al., 'Protein and Polymer Analyses up to *m*/*z* 100 000 by Laser Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **2**, 151–153 (1988).
16. J.G. Boyle, C.M. Whitehouse, 'Time-of-flight Mass Spectrometry with an Electrospray Ion Beam', *Anal. Chem.*, **64**, 2084–2089 (1992).
17. M.M. Wolff, W.E. Stephens, 'A Pulsed Mass Spectrometer with Time Dispersion', *Rev. Sci. Instrum.*, **24**(8), 616–617 (1953).
18. S.A. Goudsmit, 'A Time-of-flight Mass Spectrometer', *Phys. Rev.*, **74**, 622–623 (1948).
19. S.A. Goudsmit, E.E. Hayes, P.I. Richards, 'A Magnetic Field Time-of-flight Mass Spectrometer', *Phys. Rev.*, **84**(4), 824–829 (1951).
20. A.E. Cameron, J.D.F. Eggers, 'An Ion Velocitron', *Rev. Sci. Instrum.*, **19**(9), 605–607 (1948).
21. H.S. Katzenstein, S.S. Friedland, 'New Time-of-flight Mass Spectrometer', *Rev. Sci. Instrum.*, **26**(4), 324–327 (1955).
22. W.C. Wiley, 'Bendix Time-of-flight Mass Spectrometer', *Science*, **124**, 123–124 (1956).
23. R.J. Cotter, 'Time-of-flight Mass Spectrometry: An Increasing Role in the Life Sciences', *Biomed. Environ. Mass Spectrom.*, **18**, 513–532 (1988).
24. H.L. Friedman, 'Products of Flash Pyrolysis of Phenol–Formaldehyde by Time-of-flight Mass Spectrometry', in *Eleventh Annual ASTM Committee E-14 Conference on Mass Spectrometry*, San Francisco, CA, 47, 1963.
25. D.C. Damoth, 'Recent Improvements in the Time-of-flight Mass Spectrometer', in *Eighth Annual ASTM Committee E-14 Conference on Mass Spectrometry*, Atlantic City, NJ, 26, 1960.
26. R.W. Lehrle, J.C. Thomas, *J. Sci. Instrum.*, **39**, 458–463 (1962).
27. D.O. White et al., 'Applications of the Time-of-flight Mass Spectrometer to the Study of Inorganic Materials at Elevated Temperatures', *Adv. Mass Spectrom.*, **2**, 546–600 (1963).
28. G.J. O'Halloran, R.A. Fluegge, 'Plasma Analysis with the Bendix Time-of-flight Mass Spectrometer', in *Eleventh Annual ASTM Committee E-14 Conference on Mass Spectrometry*, San Francisco, CA, 39, 1963.
29. N.R. Isenor, 'High Energy Ions from a Q-Switched Laser', *Can. J. Phys.*, **42**, 1413–1416 (1964).
30. K.A. Lincoln, 'Flash-vaporization of Solid Materials for Mass Spectrometry by Intense Thermal Radiation', *Anal. Chem.*, **37**(4), 541–543 (1965).
31. N.C. Fenner, N.R. Daly, 'Laser used for Mass Analysis', *Rev. Sci. Instrum.*, **37**(8), 1068–1070 (1966).
32. F.J. Vastola, A. Pirone, 'Ionization of Organic Solids by Laser Irradiation', *Adv. Mass Spectrom.*, **4**, 107–111 (1968).
33. F. Hillenkamp, E. Unsold, 'Laser Microprobe Mass Analysis of Organic Materials', *Nature*, **256**, 119–120 (1975).
34. F. Hillenkamp et al., 'Mass Spectrometric Analysis of Laser Induced Microplasmas from Organic Samples', *Isr. J. Chem.*, **17**, 181–184 (1978).
35. R.J. Cotter, 'Laser Mass Spectrometry: An Overview of Techniques, Instruments, and Applications', *Anal. Chim. Acta*, **195**, 45–59 (1987).
36. P. Feigl, B. Schueler, F. Hilenkamp, 'A New Instrument for Bulk Microprobe Mass Analysis by Pulsed Laser Irradiation', *Int. J. Mass Spectrom. Ion Phys.*, **47**, 15–21 (1983).
37. R.D. Macfarlane, D.F. Torgerson, 'Californium-252 Plasma Desorption Mass Spectrometry', *Science*, **191**, 920–925 (1976).
38. R.S. Brown, J.J. Lennon, 'Factors that Influence the Observed Fast Fragmentation of Peptides in Matrix Assisted Laser Desorption', *J. Am. Soc. Mass Spectrom.*, **7**(3), 225–232 (1995).
39. R.D. Macfarlane, D.F. Torgerson, R.P. Skowronski, 'New Approach to the Mass Spectroscopy of Non-volatile Compounds', *Biochem. Biophys. Res. Commun.*, **60**(2), 616–621 (1974).
40. P. Roepstorff, K. Breddam, K. Klarskov, 'C-Terminal Sequence Determination of Peptides Degraded with Carboxypeptidases of Different Specificities and Analyzed by 252-Cf Plasma Desorption Mass Spectrometry', *Anal. Biochem.*, **180**, 28–37 (1989).
41. D.N. Nguyen, G.W. Becker, R.M. Riggin, 'Protein Mass Spectrometry: Applications to Analytical Biotechnology', *J. Chromatogr.*, **705**, 21–45 (1995).
42. B.T. Chait, K.G. Standing, 'A Time-of-flight Mass Spectrometer for Measurement of Secondary Ion Mass Spectra', *Int. J. Mass Spectrom. Ion Phys.*, **40**, 185–193 (1981).
43. E.W. Schlag, H.J. Neusser, 'Multiphoton Mass Spectrometry', *Acc. Chem. Res.*, **16**, 355–360 (1983).
44. D.M. Lubman, R. Tembreull, 'Pulsed Laser Desorption of Biological Molecules in Supersonic Beam Mass Spectrometry with Resonant Two-photon Ionization Detection', *Anal. Chem.*, **59**, 1082–1088 (1987).
45. R. Zare et al., 'Determination of Phenylthiohydantoin–Amino Acids by Two-step laser Desorption/

Multiphoton Ionization', *Anal. Chem.*, **59**, 909–912 (1987).
46. E.W. Schlag et al., 'High Resolution Laser Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **60**, 147–156 (1984).
47. R.N. Zare, F. Engelke, J.H. Hahn, 'Apparatus and Method for Laser Desorption of Molecules for Quantitation', *US Patent No. 4 988 879*, The Board of Trustees of the Leland Stanford Junior College, USA, 1991.
48. J.B. Fenn, M. Yamashita, *J. Chem. Phys.*, **88**, 4451 (1984).
49. K.G. Standing et al., 'Collisional Damping Interface for an Electrospray Ionization Time-of-flight Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **9**, 569–579 (1998).
50. M. Guilhaus, 'Spontaneous and Deflected Drift Trajectories in Orthogonal Acceleration Time-of-flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**, 588–595 (1994).
51. A.F. Dodonov et al., *Inventors Certificate No. 1 681 340A1*, Russian Academy of Sciences, USSR, 1987.
52. A.F. Dodonov et al., *International Patent Application PCT/SU89/00228*, PCT, World Intellectual Property Organization, International Bureau, Geneva, Switzerland, 1989.
53. A.F. Dodonov, I.V. Chernushevich, V.V. Laiko, 'Electrospray Ionization on a Reflecting Time-of-flight Mass Spectrometer', in *American Chemical Society Symposium Series 549*, American Chemical Society, Washington, DC, 108–123, 1994.
54. A.N. Verentchikov, W.E. Ens, K.G. Standing, 'Reflecting Time-of-flight Mass Spectrometer with an Electrospray Ion Source and Orthogonal Extraction', *Anal. Chem.*, **66**(1), 126–133 (1993).
55. A.N. Krutchinsky et al., 'Collisional Damping Interface for an Electrospray Ionization Time-of-flight Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **9**, 569–579 (1998).
56. Y. Pan, R.J. Cotter, 'Measurement of Initial Translational Energies of Peptide Ions in Laser Desorption/Ionization Mass Spectrometry', *Org. Mass Spectrom.*, **27**, 3–8 (1992).
57. V. Mlynski, M. Guilhaus, 'Matrix-assisted Laser/Desorption Ionization Time-of-flight Mass Spectrometer with Orthogonal Acceleration Geometry: Preliminary Results', *Rapid Commun. Mass Spectrom.*, **10**, 1524–1530 (1996).
58. A.N. Krutchinsky, A.V. Loboda, V.L. Spicer, D. Dworschak, W. Ens, K.G. Standing, 'Orthogonal Injection of Matrix-assisted Laser Desorption/Ionization Ions into a Time-of-flight Spectrometer through a Collisional Damping Interface', *Rapid Commun. Mass Spectrom.*, **12**, 508–518 (1998).
59. I. Smirnov et al., 'Orthogonal MALDI in the Analysis of Proteins and DNA', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 2329–2330, 1999.
60. A. Verentchikov, I. Smirnov, M.L. Vestal, 'Collisional Cooling and Ion Formation Processes in Orthogonal MALDI at Intermediate Gas Pressure', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 3052–3053, 1999.
61. V.V. Laiko, A.L. Burlingame, 'Atmospheric Pressure Matrix Assisted Laser Desorption Ionization Mass Spectrometry', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1196–1197, 1999.
62. J. Zhou et al., 'Kinetic Energy Measurements of Molecular Ions Ejected into an Electric Field by Matrix-assisted Laser Desorption', *Rapid Commun. Mass Spectrom.*, **6**, 671–678 (1992).
63. R.M. Whittal et al., 'Functional Wave Time-lag Focusing Matrix-assisted Laser Desorption/Ionization in a Linear Time-of-flight Mass Spectrometer: Improved Mass Accuracy', *Anal. Chem.*, **69**(11), 2147–2153 (1997).
64. D. Kirsch, B. Spengler, R. Kaufman, 'Initial Kinetic Energies of Peptide Ions Formed by MALDI', in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, CA, 670a, 1993.
65. W. Ens et al., 'Properties of Matrix-assisted Laser Desorption Measurements with a Time-to-digital Converter', *Rapid Commun. Mass Spectrom.*, **5**(3), 117–123 (1991).
66. R.C. Beavis, B.T. Chait, 'Velocity Distributions of Intact High Mass Polypeptide Molecule Ions Produced by Matrix Assisted Laser Desorption', *Chem. Phys. Lett.*, **181**(5), 479–484 (1991).
67. V. Bokelmann, B. Spengler, R. Kaufmann, 'Dynamical Parameters of Ion Ejection and Ion Formation in Matrix-assisted Laser Desorption/Ionization', *Eur. Mass Spectrom.*, **1**, 81–93 (1995).
68. K. Dreisewerd et al., 'Influence of Laser Intensity and Spot Size on the Desorption of Molecules and Ions in Matrix-assisted Laser Desorption/Ionization with a Uniform Beam Profile', *Int. J. Mass Spectrom. Ion Processes*, **141**, 127–148 (1995).
69. B. Spengler, V. Bokelmann, 'Angular and Time Resolved Intensity Distributions of Laser Desorbed Matrix Ions', *Nucl. Instrum. Methods Phys. Res., B*, **82**, 379–385 (1993).
70. P. Juhasz, M.L. Vestal, S.A. Martin, 'On the Initial Velocity of Ions Generated by Matrix-assisted Laser Desorption Ionization and Its Effect on the Calibration of Delayed Extraction Time-of-flight Mass Spectra', *J. Am. Soc. Mass Spectrom.*, **8**, 209–217 (1997).
71. D. Ioanoviciu, 'Ion-optical Solutions in Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 985–997 (1995).

72. A.J. Campbell, J.S. Halliday, 'Mass Measurement Accuracy and Mass Resolving Power', in *13th Annual Symposium on Mass Spectrometry*, Pergamon Press, New York, 129, 1962.
73. S. Schuerch et al., 'Enhanced Mass Resolution in Matrix-assisted Laser Desorption/Ionization Linear Time-of-flight Mass Spectrometry', *Biol. Mass Spectrom.*, **23**, 695–700 (1994).
74. M.L. Vestal, P. Juhasz, S.A. Martin, 'Delayed Extraction Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 1044–1050 (1995).
75. J.C. Tabet, R.J. Cotter, 'Laser Desorption Time-of-flight Mass Spectrometry of High Mass Molecules', *Anal. Chem.*, **56**, 1662–1667 (1984).
76. W. Martin, R. O'Malley, 'The Importance of the Time-lag Parameter in a Laser/Time-of-flight Mass Spectrometer', *Int. J. Mass Spectrom.*, **77**, 203–221 (1987).
77. J.J. Lennon, R.S. Brown, 'Pulsed Ion Extraction Combined with High Accelerating Potentials for Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics*, Chicago, IL, 501, 1994.
78. R.S. Brown, S. Hermsen, D. Christie, 'Evaluation of Pulsed Ion Extraction at High Accelerating Fields in a MALDI Time-of-flight Mass Spectrometer', in *35th Rocky Mountain Conference on Analytical Chemistry*, Denver, CO, 23, 1993.
79. S.M. Colby, T.B. King, J.P. Reilly, 'Improving the Resolution of Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry by Exploiting the Correlation between Ion Position and Velocity', *Rapid Commun. Mass Spectrom.*, **8**, 865–868 (1994).
80. T.B. King, S.M. Colby, J.P. Reilly, 'High Resolution MALDI-TOF Mass Spectra of Three Proteins Obtained Using Space–Velocity Correlation Focusing', *Mass Spectrom. Ion Processes*, **145**, L1–L7 (1995).
81. R.S. Brown, J.J. Lennon, 'Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-assisted Laser Desorption/Ionization Linear Time-of-flight Mass Spectrometer', *Anal. Chem.*, **67**, 1998–2003 (1995).
82. R.M. Whittal, L. Li, 'High Resolution Matrix-assisted Laser Desorption/Ionization in a Linear Time-of-flight Mass Spectrometer', *Anal. Chem.*, **67**, 1950–1954 (1995).
83. S.M. Colby, J.P. Reilly, 'Space–Velocity Correlation Focusing', *Anal. Chem.*, **68**, 1419–1428 (1996).
84. G.R. Kinsel, M.V. Johnston, 'Post Source Pulse Focusing: A Simple Method to Achieve Improved Resolution in a Time-of-flight Mass Spectrometer', *Int. J. Mass Spectrom. Ion Processes*, **91**, 157–176 (1989).
85. G.E. Yefchak, C.G. Enke, 'Models for Mass-independent Space and Energy Focusing in Time-of-flight Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **87**, 313–330 (1989).
86. G.R. Kinsel et al., 'Resolution Enhancement in Linear Time-of-flight Mass Spectrometry by Post-source Pulse Focusing', *Int. J. Mass Spectrom. Ion Processes*, **104**, 35–44 (1991).
87. G.R. Kinsel, J.M. Grundwuermer, J. Grotemeyer, 'High-resolution Mass Spectrometry of Large Molecules in a Linear Time-of-flight Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **4**, 2–10 (1992).
88. J.M. Grundwurmer et al., 'High-resolution Mass Spectrometry in a Linear Time-of-flight Mass Spectrometer', *Int. J. Mass Spectrom. Ion Processes*, **131**, 139–148 (1994).
89. D. Ioanoviciu, 'Linear Time-of-flight Mass Spectrometers: Post-source Pulse Focusing Conditions and Mass Scale', *J. Am. Soc. Mass Spectrom.*, **6**, 889–891 (1995).
90. N.L. Marable, G. Sanzone, 'High-resolution Time-of-flight Mass Spectrometry Theory of the Impulse-focused Time-of-flight Mass Spectrometer', *Int. J. Mass Spectrom. Ion Phys.*, **13**, 185–194 (1974).
91. J.A. Browder et al., 'High Resolution TOF Mass Spectrometry II. Experimental Confirmation of Impulse-field Focusing Theory', *Int. J. Mass Spectrom. Ion Phys.*, **37**, 99–108 (1981).
92. M.L. Muga, 'Velocity Compaction – Theory and Performance', *Anal. Instrum.*, **16**(1), 31–50 (1987).
93. S. Kovtun, R.J. Cotter, 'Analysis of Mass Correlated Delayed Extraction Operation Modes Applied to Typical MALDI Conditions', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, 1262, 1998.
94. S.V. Kovtoun, R.J. Cotter, 'The Mass-correlated Delayed Extraction: Feasibility and Resolution', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1605–1606, 1999.
95. B.A. Mamyrin, V.I. Karctaev, D.V. Shmink, V.A. Zagulin, 'The Mass Reflectron, a New Nonmagnetic Time-of-flight Mass Spectrometer with High Resolution', *Sov. Phys. JETP*, **37**(1), 45–48 (1973).
96. M. Yang, J.P. Reilly, 'A Reflectron Mass Spectrometer with UV Laser-induced Surface Ionization', *Int. J. Mass Spectrom. Ion Processes*, **75**, 209–219 (1987).
97. M. Yang, 'A New High-resolution TOF Mass Spectrometer', *Anal. Instrum.*, **16**(1), 133–150 (1987).
98. M. Yang, J.P. Reilly, 'A Reflectron Mass Spectrometer with UV Laser-induced Surface Ionization', *Int. J. Mass Spectrom. Ion Processes*, **75**, 209–219 (1987).
99. U. Boesel, R. Weinkauf, E.W. Schlag, 'Reflectron Time-of-flight Mass Spectrometry and Laser Excitation for the Analysis of Neutrals, Ionized Molecules and Secondary Fragments', *Int. J. Mass Spectrom. Ion Processes*, **112**, 121–166 (1992).
100. E. Niehuis, T. Heller, A. Benninghoven, 'Design and Performance of a Reflectron Based Time-of-flight Secondary Ion Mass Spectrometer with Electrodynamic Primary Ion Mass Separation', *J. Vac. Sci. Technol.*, **5**(4), 1243–1246 (1987).

101. E. Niehuis, T. Heller, U. Jurgens, 'Surface and Trace Analysis by High-resolution Time-of-flight Secondary Ion Mass Spectrometry', *J. Vac. Sci. Technol.*, **7**(3), 1823–1828 (1989).
102. F. Hillenkamp et al., 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers', *Anal. Chem.*, **63**(24), 1193A–1203A (1991).
103. R. Kaufman, B. Spengler, F. Lutzenkirchen, 'Mass Spectrometric Sequencing of Linear Peptides by Product-ion Analysis in a Reflectron Time-of-flight Mass Spectrometer Using Matrix-assisted Laser Desorption Ionization', *Rapid Commun. Mass Spectrom.*, **7**, 902–910 (1993).
104. T.J. Cornish, R.J. Cotter, 'A Curved Field Reflectron Time-of-flight Mass Spectrometer for the Simultaneous Focusing of Metastable Product Ions', *Rapid Commun. Mass Spectrom.*, **8**, 781–785 (1994).
105. T.J. Cornish, R.J. Cotter, 'A Curved-field Reflectron for Improved Energy Focusing of Product Ions in Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 1037–1040 (1993).
106. R. Stein, 'Space and Velocity Focusing in Time-of-flight Mass Spectrometers', *Int. J. Mass Spectrom. Ion Phys.*, **14**, 205–218 (1974).
107. M.S. Wilm, M. Mann, 'Analytical Properties of the Nanospray Ion Source', *Anal. Chem.*, **68**, 1–8 (1996).
108. B.J. Bolanos et al., 'Evaluation of Wire Guides in MALDI-TOF', in *Proceedings of the 42nd Annual ASMS Conference on Mass Spectrometry and Allied Topics*, Chicago, IL, 745, 1994.
109. C.D. Hanson, C.L. Just, 'Selective Background Suppression in MALDI-TOF Mass Spectrometry', *Anal. Chem.*, **66**(21), 3676–3680 (1994).
110. S.R. Weinberger et al., 'Laser Desorption Ionization Mass Monitor', *US Patent No. 5 382 793*, Hewlett-Packard, USA, 1995.
111. R.C. Beavis, B.T. Chait, 'Instrument and Method for the Laser Desorption of Ions in Mass Spectrometry', *US Patent No. 5 045 694*, Rockefeller University, USA, 1991.
112. T. Dresch, 'Properties of the Ideal Electrostatic Deflector in Time-of-flight Mass Spectrometry', in *46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, 1234, 1998.
113. B.J. Bolanos, D.A. Laude, 'High Voltage Deflectors for Mass Selection on a Wire Guide MALDI/TOF', in *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, GA, 131, 1995.
114. O. Vorm, P. Roepstorff, 'Detector Bias Gating for Improved Detector Response and Calibration in Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *J. Mass Spectrom.*, **31**, 351–356 (1996).
115. P.W. Geno, R.D. McFarlane, 'Secondary Electron Emission Induced by Impact of Low-velocity Molecular Ions on a Microchannel Plate', *Int. J. Mass Spectrom. Ion Processes*, **92**, 195–210 (1989).
116. B. Spengler et al., 'The Detection of Large Molecules in Matrix-assisted UV-laser Desorption', *Rapid Commun. Mass Spectrom.*, **4**(9), 301–305 (1990).
117. R. Kaufmann et al., 'Secondary-ion Generation from Large keV Molecular Primary Ions Incident on a Stainless-steel Dynode', *Rapid Commun. Mass Spectrom.*, **6**, 98–104 (1992).
118. R.S. Brown, N.L. Gilfrich, 'Ion Detection Limitations to Mass Resolution in Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 690–696 (1992).
119. W.D. Fraunberg et al., 'Collision-induced Electron Emission from Surfaces in Negative Ion Time-of-flight Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **133**, 211–219 (1994).
120. U. Bahr et al., 'A Charge Detector for Time-of-flight Mass Analysis of High Mass Ions Produced by Matrix-assisted Laser Desorption/Ionization(MALDI)', *Int. J. Mass Spectrom. Ion Processes*, **153**, 9–21 (1996).
121. M.A. Park, J.H. Callahan, 'An Inductive Detector for Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 317–322 (1994).
122. J.G. Boyle, C.M. Whitehouse, 'Time-of-flight Mass Spectrometry with an Electrospray Ion Beam', *Anal. Chem.*, **64**, 2084–2089 (1992).
123. S.M. Michael, B.M. Chien, D.M. Lubman, 'Detection of Electrospray Ionization Using a Quadrupole Ion Trap Storage Reflectron Time-of-flight Mass Spectrometer', *Anal. Chem.*, **65**, 2614–2620 (1993).
124. O.A. Migorodskaya et al., 'Electrospray Ionization Time-of-flight Mass Spectrometry in Protein Chemistry', *Anal. Chem.*, **66**, 99–107 (1994).
125. V.V. Laiko, A.F. Dodonov, 'Resolution and Spectral Line Shapes in the Reflecting Time-of-flight Mass Spectrometer with Orthogonally Injected Ions', *Rapid Commun. Mass Spectrom.*, **8**, 720–726 (1994).
126. L. Fang et al., 'On-line Time-of-flight Mass Spectrometric Analysis of Peptides Separated by Capillary Electrophoresis', *Anal. Chem.*, **66**, 126–133 (1994).
127. D.J. Douglass, J.B. French, 'Collisional Focussing Effects in Radiofrequency Quadrupoles', *J. Am. Soc. Mass Spectrom.*, **3**, 398–408 (1992).
128. I.V. Chernushevich et al., 'Effect of Ion–Molecule Collisions in the Vacuum Chamber of an Electrospray Time-of-flight Mass Spectrometer on Mass Spectra of Proteins', *J. Am. Soc. Mass Spectrom.*, **9**, 569–579 (1996).
129. A. Shevchenko et al., 'Rapid 'de Novo' Peptide Sequencing by a Combination of Nanoelectrospray Isotopic Labeling and a Quadrupole/Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **11**, 1015–1024 (1997).
130. X.J. Tang et al., 'Investigation of Protein–Protein Noncovalent Interactions in Soybean Agglutinin by Electrospray Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **8**, 750–754 (1994).

131. M.C. Fitzgerald et al., 'Probing the Oligomeric Structure of an Enzyme by Electrospray Ionization Time-of-flight Mass Spectrometry', *Proc. Natl. Acad. Sci. USA*, **93**, 6851–6856 (1996).
132. I.V. Chernushevich, W. Ens, K.G. Standing, 'Measurement of Noncovalent Complexes with High *m*/*z* by Electrospray Time-of-flight Mass Spectrometry', in *New Methods for the Study of Biomolecular Complexes*, ed. W. Ens, Kluwer Academic, Norwell, MA, 101–116, 1998.
133. A. Ayed et al., 'Quantitative Evaluation of Protein–Protein and Ligand–Protein Equilibria of a Large Allosteric Enzyme by Electrospray Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**, 339–344 (1998).
134. N. Potier et al., 'Study of the Non-covalent TRP Repressor: DNA Operator Complex by Electrospray Ionization Time-of-flight Mass Spectrometry', *Protein Sci.*, **7**, 1388–1395 (1998).
135. J.F. Banks, E.E. Gulcicek, 'Rapid Peptide Mapping by Reverse Phase Liquid Chromatography with Electrospray Ionization and Time-of-flight Mass Spectrometry', *Anal. Chem.*, **69**, 3973–3978 (1997).
136. J.F. Banks, 'Proteins Analyzed by Packed Capillary Liquid Chromatography with Electrospray Ionization and Time-of-flight Mass Spectrometry Detection', *J. Chromatogr.*, **786**, 67–73 (1997).
137. M.W.F. Nielen, F.A. Bujitenhuijs, 'Polymer Analysis by Liquid Chromatography Electrospray Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **71**, 1809–1814 (1999).
138. A.C. Hogenboom et al., 'Accurate Mass Determinations for the Confirmation and Identification of Organic Micro-contaminants in Surface Water Using On-line Solid-phase Extraction Liquid Chromatography Electrospray Orthogonal-acceleration Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **13**, 125–133 (1999).
139. K.M. Clauwaert et al., 'Investigation of the Quantitative Properties of the Quadrupole Orthogonal Acceleration Time-of-flight Mass Spectrometer with Electrospray Ionization Using 3,4-Methylenedioxymethanphetamine', *Rapid Commun. Mass Spectrom.*, **13**, 1540–1545 (1999).
140. J.F. Banks, 'Recent Advances in Capillary Electrophoresis Electrospray Mass Spectrometry', *Electrophoresis*, **18**, 2255–2266 (1997).
141. D.C. Muddiman et al., 'Application of Sequential Paired Covariance to Capillary Electrophoresis Electrospray Ionization Time-of-flight Mass Spectrometry – Unraveling the Signal from the Noise in the Electropherogram', *Anal. Chem.*, **67**, 4371–4375 (1995).
142. J.F. Banks, T. Dresch, 'Detection of Fast Capillary Electrophoresis Peptide and Protein Separations Using Electrospray Ionization with a Time-of-flight Mass Spectrometer', *Anal. Chem.*, **68**, 1480–1485 (1996).
143. I.M. Lazar et al., 'Evaluation of an Electrospray Interface for Capillary Electrophoresis Time-of-flight Mass Spectrometry', *J. Chromatogr. A*, **791**, 269–278 (1997).
144. I.M. Lazar et al., 'General Considerations for Optimizing a Capillary Electrophoresis Electrospray Ionization Time-of-flight Mass Spectrometry System', *J. Chromatogr. A*, **829**, 279–288 (1998).
145. P. Cao, M. Moini, 'Pressure-assisted and Pressure-programmed Capillary Electrophoresis Electrospray Ionization Time-of-flight Mass Spectrometry for the Analysis of Peptide Mixtures', *Electrophoresis*, **19**, 2200–2206 (1998).
146. P. Cao, M. Moini, 'Capillary Electrophoresis Electrospray Ionization High Mass Accuracy Time-of-flight Mass Spectrometry for Protein Identification Using Peptide Mapping', *Rapid Commun. Mass Spectrom.*, **12**, 864–870 (1998).
147. M.E. McComb et al., 'Sensitive High-resolution Analysis of Biological Molecules by Capillary Zone Electrophoresis Coupled with Reflecting Time-of-flight Mass Spectrometry', *J. Chromatogr. A*, **800**, 1–11 (1998).
148. D.R. Barnridge et al., 'Metallized Sheathless Electrospray Emitters for use in Capillary Electrophoresis Orthogonal Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **13**, 994–1002 (1999).
149. I.M. Lazar et al., 'High Speed TOF MS Detection for Capillary Electrophoresis', *Anal. Chem.*, **71**, 2578–2581 (1999).
150. G. Choudhary, C. Horvath, J.F. Banks, 'Capillary Electrochromatography of Biomolecules with On-line Electrospray Ionization and Time-of-flight Mass Spectrometry', *J. Chromatogr. A*, **828**, 469–480 (1998).
151. K.K. Murray, 'Coupling Matrix-assisted Laser Desorption/Ionization to Liquid Separations', *Mass Spectrom. Rev.*, **16**, 283–299 (1997).
152. L. Li, A.P.L. Wang, L.D. Coulson, 'Continuous-flow Matrix-assisted Laser Desorption Ionization Mass Spectrometry', *Anal. Chem.*, **65**, 493–495 (1993).
153. R.M. Caprioli, T. Fan, J.S. Cottrell, 'Continuous-flow Sample Probe for Fast Atom Bombardment Mass Spectrometry', *Anal. Chem.*, **58**, 2949–2954 (1986).
154. K.K. Murray, D.H. Russell, 'Liquid Sample Introduction for Matrix-assisted Laser Desorption/Ionization', *Anal. Chem.*, **65**, 2534–2537 (1993).
155. P. Arpino, M.A. Baldwin, F.W. McLafferty, 'Liquid Chromatography–Mass Spectrometry. II. Continuous Monitoring', *Biomed. Mass Spectrom.*, **1**, 80–82 (1974).
156. D.S. Nagra, L. Li, 'Liquid Chromatography Time-of-flight Mass Spectrometry with Continuous-flow Matrix-assisted Laser Desorption/Ionization', *J. Chromatogr. A*, **711**, 235–245 (1995).
157. R.M. Whittal, L.M. Russon, L. Li, 'Development of Liquid Chromatography Mass Spectrometry Using Continuous Flow Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *J. Chromatogr. A*, **794**, 367–375 (1998).

158. K.K. Murray, D.H. Russell, 'Aerosol Matrix-assisted Laser Desorption Ionization Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **5**, 1–9 (1994).
159. K.K. Murray et al., 'Aerosol Matrix-assisted Laser Desorption Ionization for Liquid Chromatography/Time-of-flight Mass Spectrometry', *Anal. Chem.*, **66**, 1601–1609 (1994).
160. X. Fei, G. Wei, K.K. Murray, 'Aerosol MALDI with a Reflectron Time-of-flight Mass Spectrometer', *Anal. Chem.*, **68**, 1143–1147 (1996).
161. L. He, G. Wei, K.K. Murray, 'Fragmentation of Vitamin B-12 in Aerosol Matrix-assisted Laser Desorption Ionization', *J. Am. Soc. Mass Spectrom.*, **8**, 140–141 (1997).
162. X. Fei, K.K. Murray, 'On-line Coupling of Gel Permeation Chromatography with MALDI Mass Spectrometry', *Anal. Chem.*, **68**, 3555–3560 (1996).
163. S.M. Colby, J.P. Reilly, 'Photoemission Electron Impact Ionization in Time-of-flight Mass Spectrometry: An Examination of Experimental Consequences', *Int. J. Mass Spectrom. Ion Processes*, **131**, 125–138 (1994).
164. P. Martin, O. Peraldi, A.L. Heureux, M.J. Bertrand, 'An Integrated Benchtop Mass Spectrometer for Pyrolytic Analysis of Complex Chemical and Biological Samples', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry*, Dallas, TX, 849–850, 1999.
165. B.H. Wang, K. Dreisewerd, U. Bahr, M. Karas, F. Hillenkamp, 'Gas-phase Cationization and Protonation of Neutrals Generated by Matrix-assisted Laser Desorption', *J. Am. Soc. Mass Spectrom.*, **4**, 393–398 (1993).
166. R.S. Gohlke, 'A Gas Chromatographic Time-of-flight Mass Analyzer', *Anal. Chem.*, **31**, 535 (1959).
167. D. Faubert, G.J. Paul, J. Giroux, M.J. Bertrand, 'Selective Fragmentation and Ionization of Organic Compounds Using an Energy-tunable Rare-gas Metastable Beam Source', *J. Mass Spectrom. Ion Processes*, **124**, 69–77 (1993).
168. R. Grix, U. Gruner, G. Li, H. Stroh, H. Wollnik, 'An Electron Impact Storage Ion Source for Time-of-flight Mass Spectrometers', *Int. J. Mass Spectrom. Ion Processes*, **93**, 323–330 (1989).
169. J.D. Pinkston, M. Rabb, J.T. Watson, J. Allison, 'New Time-of-flight Mass Spectrometer for Improved Mass Resolution, Versatility, and Mass Spectrometry/Mass Spectrometry Studies', *Rev. Sci. Instrum.*, **57**, 583 (1986).
170. D.J. Beussman, P.R. Vlasak, R.D. McLane, M.A. Seeterlin, C.G. Enke, 'Tandem Reflectron Time-of-flight Mass Spectrometer Utilizing Photodissociation', *Anal. Chem.*, **67**, 3952–3957 (1995).
171. V.B. Artaev, N. Brichford, J. Mitchell, R. Parry, 'Time-of-flight Mass Spectrometer as a Detector for High Speed Gas Chromatography', in *Proceedings of the 44th ASMS Conference On Mass Spectrometry and Allied Topics*, Portland, OR, 798, 1996.
172. Q. Ji, M.R. Davenport, C.G. Enke, J.F. Holland, 'A Segmented Ring Cylindrical Ion Trap Source for Time-of-flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **7**, 1009–1017 (1996).
173. R.M. Waldren, J.F.J. Todd, 'The Quadrupole Ion Store (QUISITOR) Part III. Studies on Phase-synchronized Ion Ejection: The Effects of Ejection Pulse Width and Detection Pulse Delay', *Int. J. Mass Spectrom. Ion Phys.*, **29**, 315–335 (1979).
174. G.J. O'Halloran, R.Q. Fluegge, *Report ASD-TDF 62-644*, Bendix Co. Res. Lab., Detroit, MI, 1964.
175. J.H.J. Dawson, M. Guilhaus, 'Orthogonal-acceleration Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **3**, 155 (1989).
176. J. Coles, M. Guilhaus, 'Orthogonal Acceleration – A New Direction for Time-of-flight Mass Spectrometry: Fast, Sensitive Mass Analysis for Continuous Ion Sources', *Trends Anal. Chem.*, **12**, 203 (1993).
177. A.F. Dodonov, I.V. Chernushevich, V.V. Laiko, 'Electrospray Ionization in a Reflecting Time-of-flight Mass Spectrometer', in *Time-of-flight Mass Spectrometry*, ed. R.J. Cotter, *ACS Symposium Series*, American Chemical Society, Washington, DC, 108–123, Vol. 549, 1994.
178. T. Bergman, T.P. Martin, H. Schaber, 'High Resolution Time-of-flight Mass Spectrometer', *Rev. Sci. Instrum.*, **60**, 792 (1989).
179. Q. Ji, M.R. Davenport, C.G. Enke, J.F. Holland, 'A Segmented Ring Cylindrical Ion Trap Source for Time-of-flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **7**, 1007–1017 (1996).
180. S.C. Davis, A.A. Makarov, J.D. Hughes, 'Ultrafast Gas Chromatography Using Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **13**, 237–241 (1999).
181. J.D. Hughes, A.A. Makarov, S.C. Davis, A.D. Hoffmann, 'Fast GC/MS of Complex Mixtures Using Benchtop TOF_MS', in *Proceedings of the 47th ASMS Conference On Mass Spectrometry and Allied Topics*, Dallas, TX, 1700–1701, 1999.
182. M. Green, R. Bateman, R. Carruthers, A. Gilbert, 'GC/MS Automated High Throughput Exact Mass Measurement Using a Time-of-flight Mass Spectrometer', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 715–716, 1999.
183. A. Amirav, 'Electron Impact and Hyperthermal Surface Ionization Mass Spectrometry in Supersonic Molecular Beams', *Org. Mass Spectrom.*, **26**, 1–17 (1991).
184. S. Danon, A. Amirav, 'Isotope Molecular and Surface Effects on Hyperthermal Surface-induced Dissociative Ionization', *Int. J. Mass Spectrom. Ion Processes*, **125**, 63–74 (1993).
185. S.C. Davis, A.A. Makarov, J.D. Hughes, 'Supersonic Molecular Beam–Hyperthermal Surface Ionization Coupled with Time-of-flight Mass Spectrometry Applied to Trace Level Detection of Polynuclear Aromatic Hydrocarbons in Drinking Water for Reduced Sample

Preparation and Analysis Time', *Rapid Commun. Mass Spectrom.*, **13**, 247–250 (1999).
186. J.F. Holland, B. Newcombe, R.E. Tecklenburg, M. Davenport, J. Allision, J.T. Watson, C.G. Enke, 'Design, Construction, and Evaluation of an Integrating Transient Recorder for Data Acquisition in Capillary Gas Chromatography/Time-of-flight Mass Spectrometry', *Rev. Sci. Instrum.*, **62**, 69–76 (1991).
187. J.F. Holland, C.G. Enke, M.R. Davenport, L.W. Janow, 'Integrating Transient Recorder for Time Array Detection in Time-of-flight Mass Spectrometry', *US Patent No. 5 367 162*, 1994.
188. S.C. Davis, D.R. Bandura, A.D. Hoffmann, A.A. Makarov, 'A Time-of-flight Mass Analysis Device', *World Patent No. WO9840 907*, 1998.
189. A.L. Rockwood, L.J. Davis, 'Multi-anode Time-to-digital Converter', *US Patent No. 5 777 326*, 1998.
190. A. Montaser, *Inductively Coupled Plasma Mass Spectrometry*, Wiley–VCH, Weinheim, 1998.
191. S.C. Davis, D.R. Bandura, 'Gate for Eliminating Charged Particles in TOF MS', *World Patent No. WO9 833 203*, 1998.
192. G. Li, G.M. Hieftje, *US Patent No. 5 614 711*, 1997.
193. D.P. Myers, K. Brushwyler, 'Benefits of Simultaneous ICP–TOF MS', in *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, 483, 1997.
194. D.P. Myers, K. Brushwyler, 'Optimizing Selective Ion Deflection in ICP TOF MS', in *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, 330, 1998.
195. F. Vanhacek, L. Moens, R. Dams, L. Allen, S. Georgits, 'Evaluation of the Isotope Ratio Performance of an Axial TOF ICP MS', *Anal. Chem.*, **71**, 3297–3303 (1999).
196. P.P. Mahoney, G. Li, G.M. Hieftje, 'Laser Ablation ICP MS with a TOF Mass Analyzer', *J. Anal. Spectrosc.*, **11**, 401 (1996).
197. S.M. Gallus, K.G. Heumann, 'Development of a Gas Chromatography Inductively Coupled Plasma Isotope Dilution Mass Spectrometry System for Accurate Determination of Volatile Element Species', *J. Anal. Atom. Spectrom.*, **11**, 887 (1996).
198. N.P. Vela, L.K. Olson, J.A. Caruso, 'Elemental Speciation with Plasma Spectrometry', *Anal. Chem.*, **65**, 585A (1993).
199. S.C. Davis, D.R. Bandura, in *World PCT*, World Intellectual Property Organization, International Bureau, Geneva, 1998.
200. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An ICP–TOF for Elemental Analysis: Part 1: Optimization and Characteristics', *J. Am. Soc. Mass Spectrom.*, **5**, 1008 (1994).
201. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An ICP–TOF for Elemental Analysis. Part 2: Direct Current Quadrupole Lens System for Improved Performance', *J. Am. Soc. Mass Spectrom.*, **6**, 400 (1995).
202. D.P. Myers, G. Li, P.P. Mahoney, G.M. Hieftje, 'An ICP–TOF for Elemental Analysis. Part 3: Analytical Performance', *J. Am. Soc. Mass Spectrom.*, **6**, 411–420 (1995).
203. S.A. McLuckey et al., 'Ion Trap Mass Spectrometry of Externally Generated Ions', *Anal. Chem.*, **66**(13), 689A–696A (1994).
204. J.N. Louris et al., 'Instrumentation and Energy Deposition in Quadrupole Ion-trap Mass Spectrometry', *Anal. Chem.*, **59**, 1677–1685 (1987).
205. J.G.C. Stafford et al., 'Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology', *Int. J. Mass Spectrom. Ion Processes*, **60**, 85–98 (1984).
206. S.M. Michael, M. Chien, D.M. Lubman, 'An Ion Trap Storage/Time-of-flight Mass Spectrometer', *Rev. Sci. Instrum.*, **63**, 4277–4284 (1992).
207. B.M. Chien, S.M. Michael, D.M. Lubman, 'The Design and Performance of an Ion Trap Storage–Reflectron Time-of-flight Mass Spectrometer', *Int. J. Mass Spectrom. Ion Processes*, **131**, 149–179 (1994).
208. M.G. Qian, D.M. Lubman, 'A Marriage Made in MS', *Anal. Chem.*, **67**(7), 234A–242A (1995).
209. R.W. Purves, L. Li, 'Development and Characterization of An Electrospray Ionization Ion Trap/Linear Time-of-flight Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **8**(10), 1085–1093 (1997).
210. L. He, D.M. Lubman, 'Simulation of External Ion Injection, Cooling, and Extraction Processes with SIMION 6.0 for the Ion Trap/Reflectron Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **11**, 1467–1477 (1997).
211. U. Wilhelm, C. Weickhardt, J. Grotemeyer, 'Ion Trajectory Calculations for a Quadrupole-Ion-trap–Reflectron–Time-of-flight Hybrid Instrument: Effects of the Initial RF-phase and the Trapping Time on an Ion Bunch Produced from a Molecular Beam', *Rapid Commun. Mass Spectrom.*, **10**, 473–477 (1996).
212. A. Mordehai, J. Karnicky, B. Limbek, S.E. Buttrill, 'A New LC Electrospray Ion Trap Time-of-flight Mass Spectrometer', in *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, GA, 1995.
213. D.M. Chambers, S.W. Thomas, L.I. Grace, B.D. Andresen, 'The QUISTOR as a Pulsed-extraction Source for Time-of-flight Mass Spectrometry of VOC's in Air', in *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics*, Atlanta, GA, 1995.
214. D.M. Chambers, L.I. Grace, B.D. Andresen, 'Development of an Ion Store/Time-of-flight Mass Spectrometer for the Analysis of Volatile Compounds in Air', *Anal. Chem.*, **69**(18), 3780–3790 (1997).
215. M.G. Qian, D.M. Lubman, 'Procedures for Tandem Mass Spectrometry on an Ion Trap Storage/Reflectron

Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **10**, 1911–1920 (1996).
216. L. He et al., 'Development of a Capillary High Performance Liquid Chromatography Tandem Mass Spectrometer Using SWIFT Technology in an Ion Trap/Reflectron Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **11**, 1739–1748 (1997).
217. J. Wu et al., 'On Line Analysis by Capillary Separations Interfaced to an Ion Trap Storage/Reflectron Time-of-flight Mass Spectrometer', *J. Chromatogr. A*, **794**, 377–389 (1998).
218. K.P. Aicher et al., 'Design and Set-up of an Ion Trap–Reflectron–Time-of-flight Mass Spectrometer', *Eur. Mass Spectrom.*, **1**, 331–340 (1995).
219. U. Wilhelm, K.P. Aicher, J. Grotemeyer, 'Ion Storage Combined with Reflectron Time-of-flight Mass Spectrometry: Ion Cloud Motions as a Result of Jet-cooled Molecules', *Int. J. Mass Spectrom. Ion Processes*, **152**, 111–120 (1996).
220. T.L. Grebner, H.J. Neusser, 'Dissociation Rates of Energy-selected Benzene Cations: Comparison of Experimental Results from Ion Cyclotron Resonance and Cylindrical Ion Trap Time-of-flight Mass Spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **185/186/187**, 517–532 (1999).
221. L. Wang et al., 'Photo-detachment Photoelectron Spectroscopy of Multiply Charged Anions Using Electrospray Ionization', *Rev. Sci. Instrum.*, **70**(4), 1957–1966 (1999).
222. J.M. Campbell, B.A. Collings, D.J. Douglass, 'A New Linear Ion Trap Time-of-flight System with Tandem Mass Spectrometry Capabilities', *Rapid Commun. Mass Spectrom.*, **12**, 1463–1474 (1998).
223. R. Guevremont, R.W. Purves, 'Atmospheric Pressure Ion Focusing in a High Field Asymmetric Waveform Ion Mobility Spectrometer', *Rev. Sci. Instrum.*, **70**, 1370–1383 (1999).
224. R. Guevremont, R.W. Purves, 'Ion Trapping at Atmospheric Pressure (760 Torr) and Room Temperature with a High-field Asymmetric Waveform Ion Mobility Spectrometer', *Int. J. Mass Spectrom.*, **193**, 45–56 (1999).
225. M. Karas, U. Bahr, 'Laser Desorption Ionization Mass Spectrometry of Large Biomolecules', *Trends Anal. Chem.*, **9**(10), 321–325 (1990).
226. A.L. Burlingame, R.K. Boyd, S.J. Gaskell, 'Mass Spectrometry', *Anal. Chem.*, **70**(16), 647r–716r (1998).
227. J.T. Stults, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS)', *Curr. Opin. Struct. Biol.*, **5**, 691–698 (1995).
228. S.J. Doktycz, P.J. Savickas, D.A. Krueger, 'Matrix/Sample Interactions in Ultraviolet Laser-desorption of Proteins', *Rapid Commun. Mass Spectrom.*, **5**(4), 145–148 (1991).
229. G. Bolbach et al., 'Influence of Target Preparation and Laser Irradiance in Matrix-assisted UV Laser Desorption of Proteins', *Analusis*, **21**, 383–387 (1993).
230. R.C. Beavis, B.T. Chait, 'Matrix-assisted Laser-desorption Mass Spectrometry Using 355 nm Radiation', *Rapid Commun. Mass Spectrom.*, **3**(12), 436–439 (1989).
231. S.R. Weinberger, K.O. Boernsen, 'Quantitative Aspects of Laser/Desorption Ionization Mass Monitoring for Peptide Studies', in *Sixth Symposium of the Protein Society*, San Diego, CA, 231, 1992.
232. O. Vorm, P. Roepstorff, M. Mann, 'Improved Resolution and Very High Sensitivity in MALDI TOF of Matrix Surfaces Made by Fast Evaporation', *Anal. Chem.*, **66**(19), 3281–3287 (1994).
233. S.R. Weinberger et al., 'A Comparison of Sample Preparation Techniques for MALDI-TOF MS', in *Proceedings of the 41st Annual ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, CA, 775a–b, 1993.
234. R.C. Beavis, B.T. Chait, 'Cinnamic Acid Derivatives as Matrices for Ultraviolet Laser Desorption Mass Spectrometry of Proteins', *Rapid Commun. Mass Spectrom.*, **3**(12), 432–435 (1989).
235. R.C. Beavis, B.T. Chait, 'Alpha-cyano-4-hydroxycinamic Acid as a Matrix for Matrix-assisted Laser Desorption Mass Spectrometry', *Org. Mass Spectrom.*, **27**, 156 (1992).
236. K. Strupart, M. Karas, F. Hillenkamp, 'Dihydroxybenzoic Acid: A New Matrix for Laser Desorption/Ionization Mass Spectrometry', *Int. J. Mass Spectrom. Ion. Processes*, **111**, 89 (1991).
237. K. Tang, S.L. Allman, C.H. Chen, 'Matrix Assisted Laser Desorption Ionization of Oligonucleotides with Various Matrices', *Rapid Commun. Mass Spectrom.*, **7**, 943 (1993).
238. U.P. Schlunegger, 'Studies on the Selection of New Matrices for Ultraviolet Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**, 1927–1933 (1996).
239. K.J. Wu, A. Steding, C.H. Becker,'Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry of Oligonucleotides Using 3-Hydroxypicolinic Acid as an Ultraviolet-sensitive Matrix', *Rapid Commun. Mass Spectrom.*, **7**, 142–146 (1993).
240. E. Nordhoff et al., 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Nucleic Acids with Wavelengths in the Ultraviolet and Infrared', *Rapid Commun. Mass Spectrom.*, **6**, 771–776 (1992).
241. A. Overberg et al., 'Matrix-assisted Infrared-laser (2.95 μm) Desorption/Ionization Mass Spectometry of Large Biomolecules', *Rapid Commun. Mass Spectrom.*, **4**(8), 293–296 (1990).
242. M. Sadeghi et al., 'Compact Tunable Cr:LiSAF Laser for Infrared Matrix-assisted Laser Desorption/Ionization', *Rapid Commun. Mass Spectrom.*, **11**, 393–397 (1997).

243. K.L. Caldwell, K.K. Murray, 'Mid-infrared Matrix Assisted Laser Desorption Ionization with a Water/Glycerol Matrix', *Appl. Surf. Sci.*, **127**(129), 242–247 (1998).
244. P. Juhasz et al., 'Applications of Delayed Extraction Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry to Oligonucleotide Analysis', *Anal. Chem.*, **68**(6), 941–946 (1996).
245. K.O. Boernsen, M. Schaer, H.M. Widmer, 'Matrix-assisted Laser Desorption Ionization Mass Spectrometry and Its Applications in Chemistry', *Chimia*, **44**, 412–416 (1990).
246. S.A. Pergantis, W.R. Cullen, G.K. Eigendorf, 'Evaluation of Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry for the Analysis of Organoarsenic Compounds of Environmental Interest', *Biol. Mass Spectrom.*, **23**, 749–755 (1994).
247. D.C. Muddiman et al., 'Quantitative Measurement of Cyclosporin A in Blood by Time-of-flight Mass Spectrometry', *Anal. Chem.*, **66**, 2362–2368 (1994).
248. M.W. Duncan, G. Matanovic, A. Cerpa-Poljak, 'Quantitative Analysis of Low Molecular Weight Compounds of Biological Interest by Matrix-assisted Laser Desorption Ionization', *Rapid Commun. Mass Spectrom.*, **7**, 1090–1094 (1993).
249. R.M. Jones, J.H. Lamb, C.K. Lim, 'Urinary Porphyrin Profiles by Laser Desorption/Ionization Time-of-flight Mass Spectrometry without the use of Classical Matrices', *Rapid Commun. Mass Spectrom.*, **9**, 921–923 (1995).
250. G. Montaudo et al., 'Characterization of Polymers by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry: Molecular Weight Estimates in Samples of Varying Polydispersity', *Rapid Commun. Mass Spectrom.*, **9**, 453–460 (1995).
251. G. Montaudo et al., 'Molecular Weight Distribution of Poly(dimethylsiloxane) by Combining Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry with Gel-permeation Chromatography Fractionation', *Rapid Commun. Mass Spectrom.*, **9**, 1158–1163 (1995).
252. D.C. Schriemer, L. Li, 'Mass Discrimination in the Analysis of Polydisperse Polymers by MALDI Time-of-flight Mass Spectrometry. 2. Instrumental Issues', *Anal. Chem.*, **69**(20), 4176–4183 (1997).
253. D.C. Schriemer, L. Li, 'Mass Discrimination in the Analysis of Polydisperse Polymers by MALDI Time-of-flight Mass Spectrometry. 1. Sample Preparation Issues', *Anal. Chem.*, **69**(20), 4165–4175 (1997).
254. P.O. Danis, D.E. Karr, F. Mayer, A. Holle, C.H. Watson, 'The Analysis of Water Soluble Polymers by, Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Org. Mass Spectrom.*, **27**, 843–846 (1992).
255. J.C. Blais et al., 'Matrix-assisted Laser Desorption Ionisation Time-of-flight Mass Spectrometry of Synthetic Polyesters', *Int. J. Mass Spectrom. Ion Processes*, **144**, 131–138 (1995).
256. U. Bahr, A. Deppe, M. Karas, F. Hillenkamp, U. Giessmann, 'Mass Spectrometry of Synthetic Polymers by UV-Matrix Assisted Laser Desorption/Ionization', *Anal. Chem.*, **64**, 2866–2869.
257. R.C. Beavis, B.T. Chait, 'High-accuracy Molecular Mass Determination of Proteins Using Matrix Assisted Laser Desorption Mass Spectrometry', *Anal. Chem.*, **62**(17), 1836–1840 (1990).
258. C.W. Sutton, J.A. O'Neill, J.S. Cotrell, 'Site-specific Characterization of Glycoprotein Carbohydrates by Exoglycosidase Digestion and Laser Desorption Mass Spectrometry', *Anal. Biochem.*, **218**, 34–46 (1994).
259. Y. Yang, R. Orlando, 'Simplifying the Exoglycosidase Digestion/MALDI-MS Procedures for Sequencing N-linked Carbohydrate Side Chains', *Anal. Chem.*, **68**(3), 570–572 (1996).
260. M.C. Huberty et al., 'Site-specific Carbohydrate Identification in Recombinant Proteins Using MALDI-TOF MS', *Anal. Chem.*, **65**(20), 2791–2800 (1993).
261. M. Mohr, K.O. Boernsen, H.M. Widmer, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry: Improved Matrix for Oligosaccharides', *Rapid Commun. Mass Spectrom.*, **9**, 809–814 (1995).
262. B. Stahl et al., 'Analysis of Neutral Oligosaccharides by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **63**(14), 1463–1466 (1991).
263. D.I. Papac, A. Wong, A.J.S. Jones, 'Analysis of Acidic Oligosaccharides and Glycopeptides by Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Anal. Chem.*, **68**, 3215–3223 (1996).
264. D.J. Harvey, P.M. Rudd, 'Examination of Complex Oligosaccharides by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry on Time-of-flight and Magnetic Sector Istruments', *Org. Mass Spectrom.*, **29**, 753–765 (1994).
265. B. Spengler et al., 'Structure Analysis of Branched Oligosaccharides Using Post-source Decay in Matrix-assisted Laser Desorption Ionization Mass Spectrometry', *J. Mass Spectrom.*, **30**, 782–787 (1995).
266. C. Jurinke et al., 'Recovery of Nucleic Acids from Immobilized Biotin–Streptavidin Complexes Using Ammonium Hydroxide and Applications in MALDI-TOF Mass Spectrometry', *Anal. Chem.*, **69**, 904–910 (1997).
267. K. Tang et al., 'Matrix-assisted Laser Desorption–Ionization Mass Spectrometry of Immobilized Duplex DNA Probes', *Nucleic Acids Res.*, **23**(16), 3126–3131 (1995).
268. J. Bai, Y.H. Liu, D.M. Lubman, D. Siemieniak, 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Restriction Enzyme-digested Plasmid DNA Using an Active Nafion Substrate', *Rapid Commun. Mass Spectrom.*, **8**, 687–691 (1994).

269. C.F. Bleczinski, C. Richert, 'Monitoring the Hybridization of the Components of Oligonucleotide Mixtures to Immobilized DNA via Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **12**(22), 1737–1743 (1998).
270. B.A. Bruenner, T.T. Yip, T.W. Hutchens, 'Quantitative Analysis of Oligonucleotides by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**(14), 1797–1801 (1996).
271. Y.H. Liu et al., 'Use of a Nitrocellulose Film Substrate in Matrix-assisted Laser Desorption/Ionization Mass Spectrometry for DNA Mapping and Screening', *Anal. Chem.*, **67**(19), 3482–3490 (1995).
272. Y.H. Liu et al., 'Rapid Screening of Genetic Polymorphisms Using Buccal Cell DNA with Detection by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**(9), 735–743 (1995).
273. E. Nordhoff et al., 'Rapid Identification of DNA-binding Proteins by Mass Spectrometry', *Nature Biotechnol.*, **17**(9), 884–888 (1999).
274. T.T. Yip et al., 'Prognostic Significance of DNA Flow Cytometric Analysis in Patients with Nasopharyngeal Carcinoma', *Cancer*, **83**(11), 2284–2292 (1998).
275. Y. Zhu et al., 'Improved Resolution in the Detection of Oligonucleotides up to 60-mers in Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry Using Pulsed-delayed Extraction with a Simple High Voltage Transistor Switch', *Rapid Commun. Mass Spectrom.*, **11**(9), 987–992 (1997).
276. M.C. Fitzgerald, L.M. Smith, 'Mass Spectrometry of Nucleic Acids: The Promise of Matrix-assisted Laser Desorption–Ionization (MALDI) Mass Spectrometry', *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 117–140 (1995).
277. Y. Dai et al., 'Accurate Mass Measurement of Oligonucleotides Using a Time-lag Focusing Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **10**, 1792–1796 (1996).
278. W. Tang et al., 'Positive Ion Formation in the Ultraviolet Matrix-assisted Laser Desorption/Ionization Analysis of Oligonucleotides by Using 2,5-Dihydroxybenzoic Acid', *J. Am. Soc. Mass Spectrom.*, **8**, 218–224 (1997).
279. G.J. Currie, J.R. Yates, 'Analysis of Oligodeoxynucleotides by Negative-ion Matrix-assisted Laser Desorption Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **4**, 955–963 (1993).
280. E. Nordhoff et al., 'Comparison of IR- and UV-matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Oligodeoxynucleotides', *Nucleic Acids Res.*, **22**(13), 2460–2465 (1994).
281. E. Nordhoff et al., 'Ion Stability of Nucleic Acids in Infrared Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Nucleic Acids Res.*, **21**(15), 3347–3357 (1993).
282. K. Schneider, B.T. Chait, 'Increased Stability of Nucleic Acids Containing 7-Deaza-guanosine and 7-Deaza-adenosine may Enable Rapid DNA Sequencing by Matrix-assisted Laser Desorption Mass Spectrometry', *Nucleic Acids Res.*, **23**(9), 1570–1575 (1995).
283. T. Shaler et al., 'Analysis of Enzymatic DNA Sequencing Reactions by Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **9**, 942–947 (1995).
284. C.H. Chen et al., 'MALDI for Fast DNA Analysis and Sequencing', *Lab. Robotics Automation*, **8**, 87–99 (1995).
285. R.R.O. Loo et al., 'Mass Spectrometry of Proteins Directly from Polyacrylamide Gels', *Anal. Chem.*, **68**(11), 1910–1917 (1996).
286. M.M. Vestling, C. Fenselau, 'Poly(vinylidene difluoride) Membranes as the Interface between Laser Desorption Mass Spectrometry, Gel Electrophoresis and In Situ Proteolysis', *Anal. Chem.*, **66**(4), 471–477 (1994).
287. M.M. Vestling, C. Fenselau, 'Surfaces for Interfacing Protein Gel Electrophoresis Directly with Mass Spectrometry', *Mass Spectrom. Rev.*, **14**, 169–178 (1995).
288. E.J. Zaluzec et al., 'Direct Matrix-assisted Laser Desorption Ionization Mass Spectrometric Analysis of Proteins Immobilized on Nylon-based Membranes', *J. Am. Soc. Mass Spectrom.*, **5**, 230–237 (1994).
289. J.A. Blackledge, A.J. Alexander, 'Investigation of Polyethylene Membranes as Potential Sample Support for Linking SDS-PAGE with MALDI-TOF MS for the Mass Measurement of Proteins', in *Techniques in Protein Chemistry VI*, Academic Press, New York, 13–19, 1995.
290. J.A. Blackledge, A.J. Alexander, 'Polyethylene Membrane as a Sample Support for Direct Matrix-assisted Laser Desorption/Ionization Mass Spectrometric Analysis of High Mass Proteins', *Anal. Chem.*, **67**(5), 843–848 (1995).
291. M. Schaer, K.O. Boernsen, E. Gassmann, 'Fast Protein Sequence Determination with Matrix-assisted Laser Desorption and Ionization Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **5**, 31–326 (1991).
292. E.J. Zaluzec, 'Matrix-assisted Laser Desorption Ionization Mass Spectrometry: Applications in Peptide and Protein Characterization', *Protein Express. Purif.*, **6**, 109–123 (1995).
293. J.R. Yates et al., 'Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Data Base', *Anal. Chem.*, **67**(8), 1426–1436 (1995).
294. M. Schaer et al., 'Monitoring of Carboxypeptidase Digestion by Matrix Assisted Laser Desorption Ionization Mass Spectrometry', *Chimia*, **45**, 123–126 (1991).
295. D.H. Patterson et al., 'C-Terminal Ladder Sequencing via Matrix-assisted Laser Desorption Mass Spectrometry Coupled with Carboxypeptidase Y Time-dependent

and Concentration-dependent Digestions', *Anal. Chem.*, **67**(21), 3971–3978 (1995).
296. M. Bartlet-Jones et al., 'Peptide Ladder Sequencing by Mass Spectrometry Using a Novel, Volatile Degradation Reagent', *Rapid Commun. Mass Spectrom.*, **8**, 737–742 (1994).
297. B.T. Chait et al., 'Protein Ladder Sequencing', *Science*, **262**, 89–92 (1993).
298. A.G. Craig et al., 'Monitoring Protein Kinase and Phosphatase Reactions with Matrix-assisted Laser Desorption/Ionization Mass Spectrometry and Capillary Zone Electrophoresis: Comparison of the Detection Efficiency of Peptide–Phosphopeptide Mixtures', *Biol. Mass Spectrom.*, **23**, 519–528 (1994).
299. U. Pieles et al., 'Matrix-assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry: A Powerful Tool for the Mass and Sequence Analysis of Natural and Modified Oligonucleotides', *Nucleic Acids Res.*, **21**(14), 3191–3196 (1993).
300. J.M. Schuette et al., 'Sequence Analysis of Phosphorthioate Oligonucleotides via Matrix Assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry', *J. Pharm. Biomed. Anal.*, **13**, 1195–1203 (1995).
301. M.J. Labowksy, J.B. Fenn, M. Yamashita, 'Method and Apparatus for the Mass Spectrometric Analysis of Solutions', in *US PTO*, Yale University, *US Patent #4531056*, 1985.
302. J.B. Fenn, M. Yamashita, C. Whitehouse, 'Process and Apparatus for Changing the Energy of Charged Particles Contained in a Gaseous Medium', in *US PTO*, Yale University, *US Patent #4542293*, 1985.
303. T.W. Hutchens, T.-T. Yip, 'New Desorption Strategies for the Mass Spectrometric Analysis of Macromolecules', *Rapid Commun. Mass Spectrom.*, **7**, 576–580 (1993).
304. C.I. Papac, J. Hoyes, K.B. Tomer, 'Direct Analysis of Affinity-bound Analytes by MALDI/TOF MS', *Anal. Chem.*, **66**(17), 2609–2613 (1994).
305. D.I. Papac, J. Hoyes, K.B. Tomer, 'Epitope Mapping of the Gastrin-releasing Peptide/Anti-bombesin Monoclonal Antibody Complex by Proteolysis followed by Matrix-assisted Laser Desorption Ionization Mass Spectrometry', *Protein Sci.*, **3**, 1485–1492 (1994).
306. J. Bai et al., 'Matrix-assisted Laser Desorption/Ionization Using an Active Perfluorosulfonated Ionomer Film Substrate', *Anal. Chem.*, **66**(20), 3423–3430 (1994).
307. A.H. Brockman, R. Orlando, 'New Immobilization Chemistry for Probe Affinity Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **10**(13), 1688–1692 (1996).
308. J. Bundy, C. Fenselau, 'Lectin-based Affinity Capture for MALDI-MS Analysis of Bacteria', *Anal. Chem.*, **71**(7), 1460–1463 (1999).
309. D. Dogruel, P. Williams, R.W. Nelson, 'Rapid Tryptic Mapping Using Enzymatically Active Mass Spectrometer Probe Tips', *Anal. Chem.*, **67**(23), 4343–4348 (1995).
310. J.R. Krone, R.W. Nelson, D. Dogruel, P. Williams, R. Granzow, 'Interfacing Mass Spectrometric Immunoassays with BIA', *Anal. Biochem.*, **244**(1), 124–132 (1997).
311. R.W. Nelson et al., 'Peptide Characterization Using Bioreactive Mass Spectrometer Probe Tips', *Rapid Commun. Mass Spectrom.*, **9**, 1380–1385 (1995).
312. R.W. Nelson et al., 'Mass Spectrometric Immunoassay', *Anal. Chem.*, **67**, 1153–1158 (1995).
313. R.W. Nelson, 'The use of Bioreactive Probes in Protein Characterization', *Mass Spectrom. Rev.*, **16**(6), 353–376 (1997).
314. R.W. Nelson, J.R. Krone, O. Jansson, 'Surface Plasmon Resonance Biomolecular Interaction Analysis Mass Spectrometry. 1. Chip-based Analysis', *Anal. Chem.*, **69**(21), 4363–4368 (1997).
315. R.W. Nelson, J.R. Krone, O. Jansson, 'Surface Plasmon Resonance Biomolecular Interaction Analysis Mass Spectrometry. 2. Fiber Optic-based Analysis', *Anal. Chem.*, **69**(21), 4369–4374 (1997).
316. R.W. Nelson et al., 'BIA/MS of Epitope-tagged Peptides Directly from *E. coli* Lysate: Multiplex Detection and Protein Identification at Low-femtomole to Sub-femtomole Levels', *Anal. Chem.*, **71**(14), 2858–2865 (1999).
317. X. Qian et al., 'Direct Analysis of the Products of Sequential Cleavages of Peptides and Proteins Affinity-bound to Immobilized Metal Ion Beads by Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Biochem.*, **274**(2), 174–180 (1999).
318. H. Wang, K. Tseng, S. Lebrilla, 'A General Method for Producing Bioaffinity MALDI Probes', *Anal. Chem.*, **71**(10), 2014–2020 (1999).
319. L. Zhang, R. Orlando, 'Solid-phase Extraction/MALDI-MS: Extended to Ion-pairing Surfaces for the On-target Cleanup of Protein Samples', *Anal. Chem.*, **71**(20), 4753–4757 (1999).
320. A.H. Brockman, R. Orlando, 'Probe-immobilized Affinity Chromatography/Mass Spectrometry', *Anal. Chem.*, **67**(24), 4581–4585 (1995).
321. M.E. Warren, A.H. Brockman, R. Orlando, 'On-probe Solid-phase Extraction/MALDI-MS Using Ion-pairing Interactions for the Cleanup of Peptides and Proteins', *Anal. Chem.*, **70**(18), 3757–3761 (1998).
322. L. Zhang, R. Orlando, 'Solid-phase Extraction/MALDI-MS: Extended Ion-pairing Surfaces for the On-target Cleanup of Protein Samples', *Anal. Chem.*, **71**(20), 4753–4757 (1999).
323. D.C. Schriemer, L. Liang, 'Combining Avidin–Biotin Chemistry with Matrix-assisted Laser Desorption/Ionization Mass Spectrometry', *Anal. Chem.*, **68**(19), 3382–3387 (1996).
324. K. Strupat et al., 'Matrix-assisted Laser Desorption Ionization Mass Spectrometry of Proteins Electroblotted after Polyacrylamide Gel Electrophoresis', *Anal. Chem.*, **66**, 464–470 (1994).

325. J.R. Krone et al., 'BIA/MS: Interfacing Biomolecular Interaction Analysis with Mass Spectrometry', *Anal. Biochem.*, **244**(1), 124–132 (1997).
326. S. Mouradian, C.M. Nelson, L.M. Smith, 'A Self-assembled Matrix Monolayer for UV-MALDI Mass Spectrometry', *J. Am. Chem. Soc.*, **118**(36), 8639–8645 (1996).
327. R.W. Nelson, J.R. Krone, 'Advances in Surface Plasmon Resonance Biomolecular Interaction Analysis Mass Spectrometry (BIA/MS)', *J. Mol. Recognit*, **12**(2), 77–93 (1999).
328. T.T. Yip, T.W. Hutchens, 'Immobilized Metal Ion Affinity Chromatography', *Methods Mol. Biol.*, **59**, 197–210 (1996).
329. T.T. Yip et al., 'Cryptic Antigenic Determinants on the Extracellular Pyruvate Dehydrogenase Complex/Mimeotope Found in Primary Biliary Cirrhosis. A Probe by Affinity Mass Spectrometry', *J. Biol. Chem.*, **271**(51), 32 825–32 833 (1996).
330. J.C. Blais et al., 'MALDI–TOFMS Identification of 'Odorant Binding Proteins' (OBP's) Electroblotted onto Poly(vinylidene difluoride) Membranes', *Rapid Commun. Mass Spectrom.*, **10**, 1–4 (1996).
331. A. Brockman, B. Dodd, R. Orlando, 'A Desalting Approach for MALDI-MS Using On-probe Hydrophobic Self-assembled Monolayers', *Anal. Chem.*, **69**(22), 4716–4720 (1997).
332. A. Brockman, N. Shah, R. Orlando, 'Optimization of a Hydrophobic Solid-phase Extraction Interface for Matrix-assisted Laser Desorption/Ionization', *J. Mass Spectrom.*, **33**(11), 1141–1147 (1998).
333. T.W. Hutchens, T.T. Yip, 'Method and Apparatus for Desorption and Ionization of Analytes', United States Patent and Trademark Office, Baylor College of Medicine, Houston, TX, 1998.
334. J. Wei, J.M. Buriak, G. Siuzdak, 'Desorption–Ionization Mass Spectrometry on Porous Silicon', *Nature*, **399**, 243–246 (1999).
335. B. Austen et al., 'The Role of Cholesterol in the Biosynthesis of Beta-amyloid', *Neuroreport*, **10**(8), 1699–1705 (1999).
336. S. Beck, 'Ciphergen's ProteinChip® Arrays: Miniaturization Continues on a Grand Scale', *Scientist*, **12**(21), 15 (1998).
337. V. Glaser, 'Protein Chips and Mass Spectrometry Bolster Proteomics Research', *Genet. Eng. News*, **18**(11), 1, 18, 37 (1998).
338. V. Glaser, V. Rosetta, 'Miniaturizing a Protein Lab', *Nature Biotechnol.*, **15**, 937–938 (1998).
339. E. Strauss, 'New Ways to Probe the Molecules of Life', *Science*, **282**(5393), 1406–1407 (1998).
340. T.W. Hutchens, T.T. Yip, 'Surface-enhanced Neat Desorption for Desorption and Detection of Analytes', in *US PTO*, Baylor College of Medicine, Houston, TX, *US Patent #5894063*, 1999.
341. R.J. Cotter, *Time-of-flight Mass Spectrometry: Instrumentation and Applications to Biological Research*, American Chemical Society, Washington, DC, 1997.
342. K. Biemann, I.A. Papayannopoulos, 'Amino Acid Sequencing of Proteins', *Acc. Chem. Res.*, **27**, 370–378 (1994).
343. K.F. Medzihradszky, A.L. Burlingame, 'The Advantages and Versatility of a High-energy Collision-induced Dissociation-based Strategy for Sequence and Structural Determination of Proteins', *Compan. Methods Enzymol.*, **6**, 284–300 (1994).
344. R.S. Brown, J.J. Lennon, 'Sequence Specific Fragmentation of Matrix-assisted Laser Desorbed Protein/Peptide Ions', *Anal. Chem.*, **67**(21), 3990–3999 (1995).
345. V. Katta, D.T. Chow, M.F. Rohde, 'Applications of In-source Fragmentation of Protein Ions for Direct Sequence Analysis by Delayed Extraction MALDI-TOF Mass Spectrometry', *Anal. Chem.*, **70**(20), 4410–4416 (1998).
346. J.J. Lennon, K.A. Walsh, 'Direct Sequence Analysis of Proteins by In-source Fragmentation During Delayed Ion Extraction', *Prot. Sci.*, **6**, 2446–2453 (1997).
347. S.D. Patterson, V. Katta, 'Prompt Fragmentation of Disulfide-linked Peptides During Matrix-assisted Laser Desorption Ionization Mass Spectrometry', *Anal. Chem.*, **66**(21), 3727–3732 (1994).
348. T. Pfeiffer, M. Drewello, A. Schierhorn, 'Using a Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometer for Combined In-source Decay/Post-source Decay Experiments', *J. Mass Spectrom.*, **34**, 644–650 (1999).
349. D.C. Reiber, R.S. Brown, S.R. Weinberger, J. Kenney, J. Bailey, 'Unknown Peptide Sequencing Using Matrix-assisted Laser Desorption/Ionization and In-source Decay', *Anal. Chem.*, **70**(6), 1214–1222 (1998).
350. D.C. Reiber, 'Identifying Proteins Using Matrix-assisted Laser Desorption/Ionization In-source Fragmentation Data Combined with Database Searching', *Anal. Chem.*, **70**(4), 673–683 (1998).
351. P. Chaurand, F. Luetzenkirchen, B. Spengler, 'Peptide and Protein Identification by Matrix-assisted Laser Desorption Ionization (MALDI) and MALDI–Post-source Decay Time-of-flight Mass Spectrometry', *J. Am. Soc. Mass Spectrom.*, **10**, 91–103 (1999).
352. M.M. Cordero et al., 'Sequencing Peptides without Scanning the Reflectron: Post Source Decay with a Curved-field Reflectron Time-of-flight Mass Spectrometer', *Rapid Commun. Mass Spectrom.*, **9**, 1356–1361 (1995).
353. T. Cornish, R. Cotter, 'A Compact Time-of-flight Mass Spectrometer for the Structural Analysis of Biological Molecules Using Laser Desorption', *Rapid Commun. Mass Spectrom.*, **6**, 242–248 (1992).

354. T.J. Cornish, R.J. Cotter, 'A Curved-field Reflectron for Improved Energy Focusing of Product Ions in Time-of-flight Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **7**, 1037–1040 (1993).
355. T.J. Cornish, R.J. Cotter, 'A Curved Field Reflectron Time-of-flight Mass Spectrometer for the Simultaneous Focusing of Metastable Product Ions', *Rapid Commun. Mass Spectrom.*, **8**, 781–785 (1994).
356. R.J. Cotter, T.J. Cornish, 'Tandem Time-of-flight Mass Spectrometer', United States Patent and Trademark Office, The Johns Hopkins University, *US Patent #5202563*, 1993.
357. R.J. Cotter, 'The New Time-of-flight Mass Spectrometer', *Anal. Chem.*, **71**, 445A–451A (1999).
358. P.R. Griffin et al., 'Direct Database Searching with MALDI/PSD Spectra of Peptides', *Rapid Commun. Mass Spectrom.*, **9**, 1546–1551 (1995).
359. R. Kaufman, B. Spengler, F. Lutzenkirchen, 'Mass Spectrometric Sequencing of Linear Peptides by Product-ion Analysis in a Reflectron Time-of-flight Mass Spectrometer Using Matrix-assisted Laser Desorption Ionization', *Rapid Commun. Mass Spectrom.*, **7**, 902–910 (1993).
360. R. Kaufman, D. Kirsch, B. Spengler, 'Sequencing Peptides in a Time-of-flight Mass Spectrometer: Evaluation of Post Source Decay Following Matrix Assisted Laser Desorption Ionization (MALDI)', *Int. J. Mass Spectrom. Ion Processes*, **131**, 355–385 (1994).
361. P.C. Liao, Z.H. Huang, J.A. Ison, 'Charge Remote Fragmentation of Peptides Following Attachment of a Fixed Positive Charge: A Matrix Assisted Laser Desorption/Ionization Post-source Decay Study', *J. Am. Soc. Mass Spectrom.*, **8**, 501–509 (1997).
362. J.C. Rouse, W. Yu, S.A. Martin, 'A Comparison of the Peptide Fragmentation Obtained from a Reflector Matrix-assisted Laser Desorption–Ionization Time-of-flight Mass Spectrometer and a Tandem Four Sector Mass Spectrometer', *J. Am. Soc. Mass Spectrom.*, **6**, 822–835 (1995).
363. B. Spengler, D. Kirsch, R. Kaufmann, 'Peptide Sequencing by Matrix-assisted Laser-desorption Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **6**, 105–108 (1992).
364. B. Spengler, D. Kirsch, R. Kaufmann, 'Metastable Decay of Peptides and Proteins in Matrix-assisted Laser Desorption Mass Spectrometry', *Rapid Commun. Mass Spectrom.*, **5**, 198–202 (1991).
365. W. Yu et al., 'Identification of the Facile Gas-phase Cleavage of the Asp-Pro and Asp-Xxx Peptide Bonds in Matrix-assisted Laser Desorption Time-of-flight Mass Spectrometry', *Anal. Chem.*, **65**(21), 3015–3023 (1993).
366. A.S. Woods et al., 'Peptide Analysis to the Atomole Level Using a Curved-field Reflectron MALDI-TOF Spectrometer', *J. Mass Spectrom. Soc. Jpn.*, **46**(1), 91–96 (1998).
367. T.J. Cornish, R.J. Cotter, 'Collision-induced Dissociation in a Tandem Time-of-flight Mass Spectrometer with Two Single-stage Reflectrons', *Org. Mass Spectrom.*, **28**, 1129–1134 (1993).
368. T.J. Cornish, R.J. Cotter, 'Tandem Time-of-flight Mass Spectrometer', *Anal. Chem.*, **65**(8), 1043–1047 (1993).
369. M.A. Baldwin et al., 'A Novel Tandem MALDI-TOF/TOF System: Peptide Sequencing', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1999.
370. H. Lee, D.M. Lubman, 'Sequence-specific Fragmentation Generated by Matrix-assisted Laser Desorption/Ionization in a Quadrupole Ion Trap/Reflectron Time-of-flight Device', *Anal. Chem.*, **67**(8), 1400–1408 (1995).
371. K. Tanaka, E. Kawatoh, Li Ding, A. Smith, S. Kamashiro, 'A MALDI–Quadrupole Ion Trap–TOF Mass Spectrometer', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1823–1824, 1999.
372. D. Maltby, L. Huang, M.A. Baldwin, K.F. Medzihradszky, A.L. Burlingame, 'MS/MS Experiments on a Hybrid Quadrupole–Orthogonal-TOF Mass Spectrometer', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 616–617, 1999.
373. M.E. McComb, A. Krutchinsky, W. Ens, K.G. Standing, H. Perreault, 'CE-ESI/TOF MS and CE-ESI MS/MS Analysis of Protein Digests', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1958–1960, 1999.
374. C.M. Lock, S. McKinnon, D. Figeys, W. Ens, 'Identification of Peptides by Micro-capillary HPLC–Nanospray Qq-TOF', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1954–1955, 1999.
375. A. Loboda, A. Krutchinsky, V. Spicer, W. Ens, K.G. Standing, 'Design and Performance of a MALDI–QqTOF Mass Spectrometer', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1956–1957, 1999.
376. A. Shevchenko, A. Loboda, W. Ens, K.G. Standing, 'Identification of Low Molecular Weight Gel Separated Proteins by MALDI MS and MALDI–QqTOF MS', in *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, 1961–1962, 1999.
377. M.G. Dowsett, 'SIMS Depth Profiling of Ultra Shallow Implants', *Proc. SIMS XI*, Wiley, New York, 1998.
378. R. Clampitt, D.K. Jefferies, 'Miniature Ion Sources for Analytical Instruments' *Nucl. Instrum. Methods*, **149**, 734–742 (1978).
379. E. Niehuis, T. Heller, U. Jurgens, A. Benninghoven, 'Surface and Trace Analysis by High Resolution Time-of-flight Secondary Mass Spectrometry', *J. Vac. Sci. Technol. A*, **7**, 1823–1828 (1989).

380. J. Schwieters, H.G. Cramer, T. Heller, U. Jurgens, E. Niehuis, J. Zehnpfenning, A. Benninghoven, 'High Mass Resolution Surface Imaging with a Time-of-flight Secondary Ion Mass Spectroscopy Scanning Microprobe', *J. Vac. Sci. Technol. A*, **9**, 2864–2871 (1991).

381. S.R. Bryan, F. Reich, 'Improvements in TOF/SIMS Analysis of Organic Materials Using an Indium Liquid-metal Ion Source', *Polym. Prepr. (Am Chem. Soc., Div. Polym. Chem.)*, **37**(1), 839–840 (1996).

382. D. Briggs, A. Brown, J.C. Vickerman, 'Handbook of Static Secondary Ion Mass Spectrometry', Wiley, New York, 1989.

383. M.A. Douglass, 'Time-of-flight Secondary Ion Mass Spectrometry', *Texas Instrum. Tech. J.*, **12**(5), 19–29 (1995).

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Cecil Dybowski
University of Delaware, Newark, USA

1 Introduction 1
2 Discovery 1
3 Applications in Chemical Analysis 1
4 Expansion of Applications 2
5 Applications to Solids 3
6 Current Status of Magnetic Resonance 3
Abbreviations and Acronyms 3
References 3

1 INTRODUCTION

Nuclear magnetic resonance (NMR) is a phenomenon of certain nuclear isotopes, in which they absorb energy from a radiofrequency (RF) electromagnetic field. To evince this phenomenon, a nucleus must possess spin angular momentum, which – by virtue of the fact that it is charged – results in a nuclear magnetic moment that will couple to a magnetic field by the Zeeman interaction. (See **Zeeman Interaction in Nuclear Magnetic Resonance** by Dybowski.)

Protons and neutrons, as well as electrons, possess spin. Atomic nuclei are composites of protons and neutrons. Because of the coupling of the angular momenta of the nuclear particles, an atomic nucleus will have a nonzero spin if (1) the mass number A is odd, or (2) the mass number A is even and the atomic number Z is zero.[1] Nuclei with an even mass number A and even atomic number Z will have a spin of zero in the nuclear ground state.

To create the energy states between which the transitions occur, the sample must be in a magnetic field, the strength of which determines the transition energy. Electron spin resonance (ESR) is a similar effect for electrons, as discussed by Giamello in **Electron Spin Resonance Spectroscopy**. The analytical uses of these techniques rely on the fact that the absorption frequency depends on the nature of the local environment in a manner easily correlated with chemical structure.

2 DISCOVERY

Magnetic resonance was independently discovered immediately after World War II by research teams led by Zavoisky,[2] Bloch,[3] and Purcell.[4] Zavoisky's experiment demonstrated the effect known today as ESR (or paramagnetic resonance), while the experiments of Bloch and Purcell demonstrated the NMR effect for the proton. In part, these experiments resulted from developments in RF and microwave technology necessary for detection systems such as radar. The experiments were similar, in that they involved the absorption of energy by nuclear and electron spins bathed in a magnetic field. The possibility of such resonant absorption had been implied by earlier experiments on atomic beams in nonhomogeneous magnetic fields.[5]

After these discoveries were announced, it was thought that the primary use of such magnetic resonance techniques would be the determination of fundamental constants of the nucleus or electron – the magnetogyric ratios – since the specific energy difference in a given magnetic field is determined by this property. However, it was soon discovered that NMR absorption depended not just on the magnetogyric ratio, but also slightly on the local environment.[6–8] For example, the ^{14}N resonance absorptions of NH_4NO_3 were found to occur at two different energies in the same magnetic field.

3 APPLICATIONS IN CHEMICAL ANALYSIS

Of course, a technique to sense nuclei in different chemical environments immediately attracted the attention of chemists, who have always sought convenient ways to analyze materials, and the history of NMR spectroscopy as a convenient analytical tool in chemistry began with the discovery of the chemical shift. (See **Chemical Shifts in Nuclear Magnetic Resonance** by van Bramer.) The observation of spin–spin couplings in simple molecules demonstrated that the resonant absorption was also sensitive to the presence of other nearby nuclei,[9–11] further enhancing the analytical utility of the technique. (See **Scalar Couplings in Nuclear Magnetic Resonance, General** by Karlsson.)

ESR spectroscopy, the outgrowth of Zavoisky's experiment, showed a similar sensitivity to the local environment of the electron spin through the dependence of the electron g factor on local environment and hyperfine couplings to nuclear spins.[12] Of course, to detect an ESR signal requires a material containing unpaired electron spins. Thus, the technique has become particularly valuable for analyses of materials containing radicals and unpaired electron spins.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Within a decade of the discovery of NMR, proton magnetic resonance had assumed a prominent place in chemical analysis, particularly of organic materials. Chemists used it routinely to determine the structure of newly synthesized materials, and it was found that simple correlations of structure with NMR parameters allowed a picture of the molecule(s) in a sample to be pieced together from the proton NMR spectrum.[13] The use of proton NMR spectroscopy in analysis of organic materials had a profound effect on the manner in which chemists did their work. By the 1960s, chemistry departments at universities and colleges had proton NMR instruments in routine use for analysis.

4 EXPANSION OF APPLICATIONS

During this period, the NMR properties of other nuclei were also being reported, pointing the way towards a more general use of NMR in the investigation of a broader range of materials. A particular development that has had wide application in chemical analysis came to fruition in the 1970s, with the ready availability of spectroscopy of nuclei such as ^{13}C, ^{31}P, and ^{29}Si. Today we take for granted that modern NMR spectrometers are built to allow routine analysis of many different nuclear species, as discussed in **Nuclear Magnetic Resonance Instrumentation** by Neiss. However, early commercial spectrometers were single-frequency instruments, usually devoted to the proton. The analysis of other nuclei was the bailiwick of only a few specialists who had created instruments for detection of specific nuclei, usually with considerable effort. The introduction of commercial spectrometers to address the spectroscopy of nuclei other than protons, particularly ^{13}C, reliably opened a wider use of NMR for analysis. (See **Chemical Shifts in Nuclear Magnetic Resonance** by van Bramer, and **Parameters, Calculation of Nuclear Magnetic Resonance** by Jameson, for example.)

For these other nuclear species (except tritium), the signal per nucleus is less intense than that of a proton (at fixed field), making it harder to detect. The often low natural abundance of many of the other isotopes having nuclear spins compounded this difficulty, and so analysis was tricky. The development of magnets of higher field and stability and of the Fourier transform technique to allow averaging to increase signal-to-noise ratio all addressed this problem, as discussed by Neiss in **Nuclear Magnetic Resonance Instrumentation**. By the mid-1970s it was likely that analysis of nuclei such as ^{13}C was an integral part of a routine analysis of some newly created material. Nowadays, analyses of many different nuclei are easily carried out with modern instrumentation.

The simplicity of correlations between NMR parameters and chemical structure is the principal feature that has made it such a useful technique.[14,15] (See **Parameters, Calculation of Nuclear Magnetic Resonance** by Jameson.) As interest in more and more complex materials has grown, the NMR spectra obtained have become increasingly more and more complex. As a result, techniques for spectral assignment became the focus of activity in NMR research.[16] The development of selective decoupling experiments led to studies of how one might use the inherent NMR properties to manage information available from NMR spectra. More versatile RF electronics allowed the development of pulse decoupling methods that used power more efficiently and provided better decoupling to simplify spectra. With the advent of sophisticated computer systems and more control of the spectrometer's properties, NMR spectroscopists developed experiments to manage the information, such as multidimensional NMR spectroscopy for correlation of NMR parameters,[17] as discussed in Byrd's article, **Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules**. The result was simplification of the interpretation of the often complex spectra. An understanding of the way these experiments worked resulted in the development of a plethora of techniques now in common use for spectral simplification and analysis, so that nowadays one may do a series of NMR experiments on a complex material to provide a detailed chemical structure.[16]

Articles by Eriksson **Nuclear Magnetic Resonance and Nucleic Acid Structures**, Yuan **Magnetic Resonance Angiography**, Menon **Magnetic Resonance Imaging, Functional**, Richards **Multinuclear Magnetic Resonance Spectroscopic Imaging**, Smith **Magnetic Resonance in Medicine, High Resolution Ex Vivo**, McCarthy **Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials** and Fan **Nuclear Magnetic Resonance in Analysis of Plant Soil Environments** discuss a variety of applications of magnetic resonance to medicine and biology, an important use of the technique. However, magnetic resonance has found many industrial applications, for example in the analysis of polymers (**Nuclear Magnetic Resonance, Imaging of Polymers** by Hafner, **Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers** by Mirau), coatings (Brandolini, **Nuclear Magnetic Resonance of Coating and Adhesive Systems**) and fossil fuels and hydrocarbons (Miknis, **Nuclear Magnetic Resonance Characterization of Petroleum**). It is also used in fields such as environmental monitoring (MacFall, **Nuclear Magnetic Resonance for Environmental Monitoring**) and drug

detection (Dawson, **Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs**).

5 APPLICATIONS TO SOLIDS

In the 1960s, 1970s and 1980s, NMR spectroscopists revisited the study of solid materials, where analysis of chemical structure had been limited by inherent NMR properties of solids, particularly dipole–dipole coupling and quadrupolar coupling. The development of techniques to decouple nuclei efficiently and manipulate the information in the NMR spectrum brought the analysis of solids to a further level of refinement.[19–23] One technique that has become routine uses rapid spinning about an axis inclined at the ''magic angle'' to the magnetic field to produce a spectrum of spins like ^{13}C that is qualitatively similar to a solution spectrum.

However, the development of techniques to emulate solution-state spectra was not the only development in solid-state NMR. The analysis of materials in the solid state gives more detailed pictures of the electronic properties. Today, for example, the measurement of properties such as: the coupling of a spin $I > 1/2$ to the electric-field gradient at the nuclear site through its electric quadrupole moment, or the tensor properties of the chemical shift, is used to identify a material. Various aspects of the application of NMR to solids and solid-like materials are discussed in the articles by Freude **Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance**, Canet **Relaxation in Nuclear Magnetic Resonance, General**, McGregor **Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton**, Gerald **Solid-state Nuclear Magnetic Resonance**, Neue **Nuclear Magnetic Resonance in Metals Analysis**.

6 CURRENT STATUS OF MAGNETIC RESONANCE

In the articles in this section, you will find explanations of how magnetic resonance is used in analysis throughout chemistry and allied sciences. Some articles focus on the esoterica of the technology of NMR spectroscopy, while others discuss applications of NMR spectroscopy to analysis of specific types of materials. We have tried to strike a balance between topics that emphasize the constant development of the technology of magnetic resonance and the applications of the technique that make it useful for the chemist interested in identification of materials. Inevitably, some topics will be overlooked or emphasized less; there are many other topics that could, no doubt, be included, but we hope the sampling in the encyclopedia will provide an overview of the many ways magnetic resonance is currently used.

One thing is clear in all these examples: the ability of magnetic resonance spectroscopy to provide analytical information on a wide variety of materials in a format that is convenient to interpret in terms of structure and chemical function. This is the central feature of the magnetic resonance spectroscopic technique – one that makes it a primary analytical technique in the arsenal of tools chemists use to define the systems they produce in virtually every venue in which a chemist is active – geology, biology, industrial production, or laboratory-bench research.

ABBREVIATIONS AND ACRONYMS

ESR	Electron Spin Resonance
NMR	Nuclear Magnetic Resonance
RF	Radiofrequency

REFERENCES

1. J.A. Pople, W.G. Schneider, H.J. Bernstein, *High-resolution Nuclear Magnetic Resonance*, McGraw-Hill, New York, 5, 1959.
2. E. Zavoisky, *J. Phys. USSR*, **9**(211), 245 (1945).
3. F. Bloch, W.W. Hansen, M. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
4. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
5. I.I. Rabi, S. Millman, P. Kusch, J.R. Zacharias, 'The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments', *Phys. Rev.*, **55**, 526–535 (1936).
6. W.G. Proctor, F.C. Yu, 'The Dependence of a Nuclear Magnetic Resonance Frequency Upon Chemical Compound', *Phys. Rev.*, **77**, 717 (1950).
7. W.C. Dickinson, 'Dependence of the ^{19}F Nuclear Resonance Position on Chemical Compound', *Phys. Rev.*, **77**, 736–737 (1950).
8. H.S. Gutowsky, C.J. Hoffman, 'Chemical Shifts in the Magnetic Resonance of ^{19}F', *Phys. Rev.*, **80**, 110–111 (1950).
9. N.F. Ramsey, E.M. Purcell, 'Interactions Between Nuclear Spins in Molecules', *Phys. Rev.*, **85**, 143–144 (1952).
10. E.L. Hahn, D.E. Maxwell, 'Spin Echo Measurements of Nuclear Spin Coupling in Molecules', *Phys. Rev.*, **88**, 1070–1084 (1952).

11. H.S. Gutowsky, D.W. McCall, C.P. Slichter, 'Nuclear Magnetic Resonance Multiplets in Liquids', *J. Chem. Phys.*, **21**, 279–292 (1953).
12. J.E. Wertz, J.R. Bolton, *Electron Spin Resonance: Elementary Theory and Practical Applications*, McGraw-Hill, New York, 1972.
13. R.M. Silverstein, G.C. Bassler, T.C. Morrill, *Spectrometric Identification of Organic Compounds*, John Wiley & Sons, New York, 1991.
14. E.D. Becker, *High Resolution NMR: Theory and Chemical Applications*, Academic Press, New York, 1980.
15. R.K. Harris, *Nuclear Magnetic Resonance Spectroscopy*, Pitman Publishing, London, 1983.
16. M.D. Bruch, *NMR Spectroscopy Techniques*, Marcel Dekker, Inc., New York, 1996.
17. A. Bax, *Two-dimensional Nuclear Magnetic Resonance in Liquids*, D. Riedel Publishing Company, Dordrecht, 1982.
18. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
19. C.P. Slichter, *Principles of Magnetic Resonance*, Springer-Verlag, Berlin, 1992.
20. C. Fyfe, *Solid State NMR for Chemists*, C.F.C. Press, Guelph, 1983.
21. B.C. Gerstein, C.R. Dybowski, *Transient Techniques in NMR of Solids*, Academic Press, New York, 1985.
22. U. Haeberlen, *High Resolution NMR in Solids: Selective Averaging*, Academic Press, New York, 1976.
23. M. Mehring, *Principles of High Resolution NMR in Solids*, Springer-Verlag, Berlin, 1982.

Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Thomas G. Neiss
DuPont Pharmaceuticals Company, Deepwater, New Jersey, USA

1 Introduction **1**
1.1 Development of Carbon-13 Nuclear Magnetic Resonance Spectroscopy 2
1.2 Applications Summary 3
2 Theory **4**
2.1 Physical Description of the Magnetic Resonance Phenomenon 4
2.2 Interactions Present in Carbon-13 Nuclear Magnetic Resonance Measurements 6
2.3 Heteronuclear Decoupling 7
2.4 Chemical Shifts in Carbon-13 Nuclear Magnetic Resonance Spectroscopy 9
2.5 Nuclear Overhauser Effects 17
2.6 Coherence Transfer in Carbon-13 Nuclear Magnetic Resonance Measurements 19
2.7 Nuclear Relaxation in Carbon-13 Nuclear Magnetic Resonance Spectroscopy 21
2.8 Multidimensional Carbon-13 Nuclear Magnetic Resonance Experiments 24
3 Current Perspective and Future Development **26**
Abbreviations and Acronyms **27**
Related Articles **27**
References **28**

Carbon-13 (^{13}C) nuclear magnetic resonance spectroscopy (NMR) is the measurement of the precession or resonance frequencies of the net magnetization for ^{13}C nuclei whose individual magnetic moments have been oriented in a strong magnetic field. Nuclei differing in their electronic shielding precess about the magnetic field at different Larmor or resonance frequencies. A high-power radio-frequency (rf) pulse is used to perturb the magnetization vectors from their equilibrium distribution, generating an observable transverse magnetization. Precession of this net magnetization vector about the static magnetic field induces a voltage into the NMR probe coil. Relaxation pathways promote the repartitioning of the individual magnetic moments to their equilibrium Boltzmann distributions and a dephasing of the individual magnetization vectors in the transverse plane. This signal is detected as a function of time through a phase-sensitive receiver, digitized and Fourier transformed (FT) into a frequency domain spectrum. The NMRs, referenced relative to the resonance frequency of a standard are shifted in a manner characteristic of hybridization of the atom, electronegativity of the substituents attached and the steric environment of the nucleus. These shifts typically follow a standard set of rules and thus spectra can be simulated either empirically through measurement of the additive effects of substituents or through ab initio and semi-empirical computational methods. Scalar couplings to directly attached hydrogen-1 (^{1}H) nuclei split resonance lines into $(n_H + 1)$ lines and require the use of broadband ^{1}H decoupling to remove this splitting for sensitivity enhancement. The presence of attached or nearest neighbor ^{1}H atoms also provides a means of enhancing the signal intensity of ^{13}C spectra and for selective observation of ^{13}C signals through polarization transfer. Multidimensional NMR experiments are available which permit the ^{13}C–^{13}C or ^{13}C–^{1}H correlations as well as a means of measuring $^{n}J_{CH}$ couplings.

Analysis of samples can be done without the need for internal standards or calibration standards since NMR signals are directly proportional to the moles of analyte present and there are no response factors or absorptivities to be determined for quantitative analysis. Although ^{13}C NMR spectroscopy is widely used and is a very powerful structural technique in organic and polymer chemistry, the technique suffers from an inherent lack of sensitivity due to a natural abundance of only 1.1 percent and a small magnetogyric ratio (γ_C). Long relaxation times present in small molecules undermine the quantitative accuracy of this method or require lengthy amounts of instrument time to acquire an NMR spectrum. Quantitative results can be obtained with the selection of appropriate experimental parameters often combined with the use of relaxation agents.

1 INTRODUCTION

Carbon-13 (^{13}C) NMR is the measurement of the precession or resonance frequencies of the net magnetization for ^{13}C nuclei whose individual magnetic moments have been oriented in a strong magnetic field. Nuclei differing in their electronic shielding precess about the magnetic field at different Larmor or resonance frequencies. A high-power rf pulse is used to perturb the magnetization vectors from

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

their equilibrium distribution, generating an observable transverse magnetization. Precession of this net magnetization vector about the static magnetic field induces a voltage into the NMR probe coil. Relaxation pathways promote the repartitioning of the individual magnetic moments to their equilibrium Boltzmann distributions and a dephasing of the individual magnetization vectors in transverse plane. This signal is detected as a function of time through a phase-sensitive receiver, digitized and FT into a frequency domain spectrum. The NMRs, referenced relative to the resonance frequency of a standard are shifted in a manner characteristic of hybridization of the atom, electronegativity of the substituents attached and the steric environment of the nucleus. These shifts typically follow a standard set of rules and thus spectra can be simulated either empirically through measurement of the additive effects of substituents or through ab initio and semi-empirical computational methods. Scalar couplings to directly attached hydrogen-1 (^{1}H) nuclei split resonance lines into $(n_H + 1)$ lines and require the use of broadband ^{1}H decoupling to remove this splitting for sensitivity enhancement. The presence of attached or nearest-neighbor ^{1}H atoms also provides a means of enhancing the signal intensity of ^{13}C spectra and for selective observation of ^{13}C signals through polarization transfer. Multidimensional NMR experiments are available which permit the ^{13}C–^{13}C or ^{13}C–^{1}H correlations as well as a means of measuring $^nJ_{CH}$ couplings.

Analysis of samples can be done without the need for internal standards or calibration standards since NMR signals are directly proportional to the moles of analyte present and there are no response factors or absorptivities to be determined for quantitative analysis. Although ^{13}C NMR spectroscopy is widely used and is a very powerful structural technique in organic and polymer chemistry, the technique suffers from an inherent lack of sensitivity due to a natural abundance of only 1.1 percent and a small magnetogyric ratio (γ_C). Long relaxation times present in small molecules undermine the quantitative accuracy of this method or require lengthy amounts of instrument time to acquire an NMR spectrum. Quantitative results can be obtained with the selection of appropriate experimental parameters often combined with the use of relaxation agents.

1.1 Development of Carbon-13 Nuclear Magnetic Resonance Spectroscopy

While the concept of nuclear magnetism had been discussed in the early part of this century,[1–2] the nearly simultaneous reports of Bloch[3] and Purcell[4] contained the first experiments to demonstrate the phenomenon of magnetic resonance. With improvements to the homogeneity of the external applied magnetic field, resolved ^{1}H-NMR spectra of simple molecules such as ethanol could be acquired, displaying the potential of NMR as a tool for the structural analysis of organic molecules.[5] However, it wasn't until 1957 that the first study of ^{13}C NMR spectra of organic compounds appeared in the scientific literature.[6] Development of this method suffered largely from the low sensitivity inherent in the NMR experiment, hindered by the low natural abundance of the ^{13}C isotope and its small magnetogyric ratio relative to ^{1}H nuclei (1 : 4). In addition, ^{13}C signals were split into multiplet structures due to attached protons, further reducing the signal-to-noise ratio (S/N) of the NMR signal. The introduction of broadband decoupling of ^{1}H nuclei,[7] the advent of pulse FT NMR experiments[8–11] and the development of higher magnetic field NMR spectrometers resulted in the emergence of ^{13}C-NMR as a routine analytical tool for the structural analysis of organic materials. Pulsed experiments also made viable the concept of signal averaging,[12] the accumulation and addition of multiple scans to improve sensitivity of the NMR experiment. This averaging has the effect of summing coherent resonance signals while the random noise found in an NMR spectrum does not, resulting in an increase in the S/N of a spectrum proportional to (number of scans)$^{1/2}$. Once all of these gains were realized, the acquisition of ^{13}C spectra became a routine event and the study of organic molecules using NMR rapidly grew to encompass most areas of chemistry and molecular biology.

Further refinement of NMR instrumentation, particularly in the area of superconducting magnets, extensive use of computers, and the development of multiple pulse NMR experiments has further expanded the utility of ^{13}C-NMR experiments. Introduction of polarization transfer methods such as the Insensitive Nucleus Enhancement via Polarization Transfer (INEPT)[13,14] and Distortionless Enhancement via Polarization Transfer (DEPT)[15,16] multiple pulse sequences improved the sensitivity of the ^{13}C experiment through the transfer of spin distribution information from attached protons with the potential to increase the ^{13}C signal intensity by γ_H/γ_C, which is very close to four. The DEPT sequence can also be used to determine the multiplicity through spectral editing based on modulation of the carbon signal by scalar couplings to attached protons. Use of these pulse methods has also been extended to multidimensional NMR experiments in which ^{13}C signals are correlated to other nuclei with methods such as heteronuclear correlation (HETCOR) for the correlation of ^{13}C signals to those NMRs of attached protons and incredible natural abundance double quantum transfer experiment (INADEQUATE), which is a homonuclear ^{13}C–^{13}C correlation experiment.

1.2 Applications Summary

Applications of ^{13}C-NMR spectroscopy, facilitated by the continued development of larger superconducting magnetic fields, encompass most organic chemical systems, from small molecules to synthetic and biological polymers with molecular weight ranges from a few to millions of daltons. Use of this technique is focused on the wealth of information contained in the NMR spectrum, revealing not only structural identity, but quantitative composition, conformational properties, molecular dynamics and reaction mechanisms. In many applications, ^{13}C-NMR is the only analytical tool capable of supplying quantitative structural information in a nondestructive manner.

NMR spectroscopy is a principal characterization tool for the organic chemist. A ^{13}C NMR spectrum serves as a fingerprint of molecular structure, with signals due to specific functional groups occurring in regions of the NMR spectrum established through decades of experimental study. Acquisition of a ^{13}C spectrum for a small molecule (MW $\cong$ 400 daltons) routinely requires less than an hour, in part due to higher magnetic fields and the proliferation of automated NMR spectrometers. From a simple one-dimensional (1-D) ^{13}C spectrum, one can obtain an approximate measure of sample purity, as well as an assay of the types of functional moieties present. With more sophisticated polarization transfer methods, the number of protons attached to each carbon can be determined. The measurement of scalar $^{3}J_{CH}$ couplings can provide information on molecular conformation of the molecules under study.[17,18] Multiple quantum methods such as ^{13}C–^{13}C homonuclear correlation spectroscopy enable the spectroscopist to determine the carbon connectivities and thus the molecular framework of an unknown structure.[19,20] Reaction mechanisms and dynamic processes can also be examined through ^{13}C NMR methods. The larger spectral dispersion is especially favorable for the study of exchange phenomena.[21,22] Line-shape analysis as a function of temperature can be used to determine the rate of a dynamic process such as conformer exchange and the thermodynamic stability of each isomer.[23–25] With appropriate care in sample temperature measurement, an analysis of the line-shape or kinetic rate constants using NMR yields the Arrhenius activation energy (E_A) and the free enthalpy of activation ($\Delta G^{\pm}$). Isotopic labeling is also a valuable tool for elucidating mechanisms by analyzing the placement of the isotope in the product formed in the reaction. Isotope labels used in ^{13}C-NMR experiments include ^{13}C, ^{15}N and ^{2}H.

NMR is often used to determine the structure and conformational properties of natural products such as terpenes,[26–28] steroids,[29,30] carbohydrates and polysaccharides,[31–34] nucleotides,[35–37] and proteins.[38–42] ^{13}C-NMR of natural products synthesized under controlled conditions with ^{13}C-labeled starting materials reveals the route by which these products are created biologically.[43–45] Since samples can be examined in a variety of solvents, over a wide range of temperature, and with numerous experimental methods, the structural parameters which can be obtained on biomolecules are extensive. For example, the ^{13}C chemical shift of the anomeric carbon in polysaccharides is not only sensitive to the type of adjacent monosaccharide, but also to which position on that residue it is attached and to its anomeric configuration, α or β. Selective nuclear Overhauser effect (NOE) experiments and HETCOR techniques contribute through-space distances and dihedral angles which are used in conjunction with molecular modeling techniques to determine the three-dimensional (3-D) structure of biological molecules.

Perhaps the most insightful studies using ^{13}C-NMR spectroscopy are in the study of polymer systems. NMR is the only analytical method capable of providing information on the monomer composition, conformation, configuration, monomer distribution and polymer dynamics in situ and nondestructively. Analysis ranges from synthetic polymers with a typically regular microstructure to heterogeneous biopolymers such as polysaccharides derived from higher-order plants. NMR spectroscopy has been used extensively to determine the compositional and configurational parameters in synthetic polymers.[46–49] ^{13}C chemical shifts are sensitive to parameters such as the tacticity of consecutive monomers and the distribution of different monomers in mixed polymer systems.[46] A classic example is in the spectrum of polypropylene. The methyl signal appears as a complex series of lines from changes in monomer sequence configurations up to four units away, resulting in shifts of the resonance and provide enough data to perform sequence analysis at the heptad level. Observed shifts are also sensitive to the structure of neighboring monomers in the polymer chain and, from these shifts, a statistical model can be derived for monomer sequence distributions in a polymer. ^{13}C relaxation studies are effective measures of the segmental mobility and functional group dynamics in polymer systems.[50,51]

Some examples can be found where ^{13}C-NMR has been used for quantitative purposes.[52–56] It is not considered a routine method for quantitative analysis due to relaxation effects, peak enhancement from NOE, offset effects, and the cost of NMR spectrometers. Often relaxation enhancement agents are necessary to shorten ^{13}C relaxation times and quench NOEs for quantitative purposes.[53,54,57] Of course this sort of sample contamination is not acceptable if limited sample is available for analysis by other methods. Many examples exist in the literature in which ^{13}C NMR is used semiquantitatively to determine the molar composition of synthetic

polymers, polysaccharides and other multicomponent systems where normal wet chemical or chromatographic methods may not be applicable. Internal standard methods can also be used to determine weight percent concentrations of analytes in solution if the solvent is transparent to the NMR spectrometer.

2 THEORY

^{13}C nuclei are spin 1/2. The physical basis of the NMR experiment discussed here is applicable to other nuclei with a spin quantum number $I = 1/2$, such as ^{1}H, ^{19}F, ^{31}P or ^{29}Si. Understanding the magnetic resonance phenomenon is a complex quantum-mechanical problem and no attempt is will be made here to offer a complete theoretical description, but a general physical chemical description is given to familiarize the reader with the basic principles of NMR. Numerous treatises are available which explain such theoretical formulations as density matrix theory, product operator formalisms or Bloch equations for the description of the magnetic resonance phenomenon.[58–61]

Important aspects of ^{13}C-NMR for the analytical chemist are the concept of chemical shift and the ability to determine quantitatively the constituent groups in a molecule. Chemical shift is the primary source of structural information, i.e. what type of functional groups are present in a sample and how are they connected to each other. Quantitative accuracy depends primarily on relaxation processes that return magnetization to equilibrium. These topics are presented in detail to provide the chemist with a practical understanding of their use and the limitation of NMR in quantitative analysis.

2.1 Physical Description of the Magnetic Resonance Phenomenon

Nuclei with either odd atomic number or mass possess the property of nuclear spin, denoted by the spin quantum number I. Isotopes such as $^{12}C_6$, which has both an even atomic mass and atomic number, do not have this property and are not NMR-active. This property of spin gives rise to a magnetic moment (μ_i) for the nucleus since the oscillation of an electrical charge generates a magnetic field. In the absence of any external magnetic field, μ_i has no preferential orientation, a net equilibrium magnetization (M_z) does not exist, and there is no observable NMR signal. Placing the nuclei in a strong magnetic field orients the individual moments either with the external field in a low-energy state (α) or against the applied field in a high-energy state (β). These magnetic moments precess about the external magnetic field at a frequency that depends on the size of the external field, known as the Larmor frequency, ν_0. The quantum numbers of energy levels that nuclear spins may occupy are defined by m as shown in Equation (1).

$$m = I, I-1, I-2, \ldots -I \tag{1}$$

^{13}C nuclei have $I = 1/2$, for which there are two energy levels for ^{13}C with an energy separation defined in Equation (2):

$$\Delta E = \frac{\gamma_C B_0 h}{2\pi} \tag{2}$$

where γ_C is the magnetogyric ratio for ^{13}C (6.7263×10^7 rad T^{-1} s^{-1}), B_0 is the external magnetic field strength, and h is Planck's constant (6.6256×10^{-34} J s). This is illustrated in Figure 1. For a 11.744 Tesla magnet with a ^{1}H resonance frequency of 500.1 MHz, the energy-level separation for ^{13}C is defined by Equation (3):

$$\nu_0 = \frac{\gamma_C B_0}{2\pi} \tag{3}$$

which corresponds to 125.7 MHz. This frequency is the rate of precession about the external field and that frequency corresponding to the energy necessary to cause a transition from one energy level to another. The populations of the α and β energy levels follow a Boltzmann distribution shown in Equation (4):

$$\frac{N_\alpha}{N_\beta} = e\frac{\Delta E}{kT} = e\frac{\gamma_C B_0 h}{kT2\pi} \tag{4}$$

in which k is the Boltzmann constant (1.38044×10^{-23} J K^{-1}) and T is the sample temperature in kelvins. At

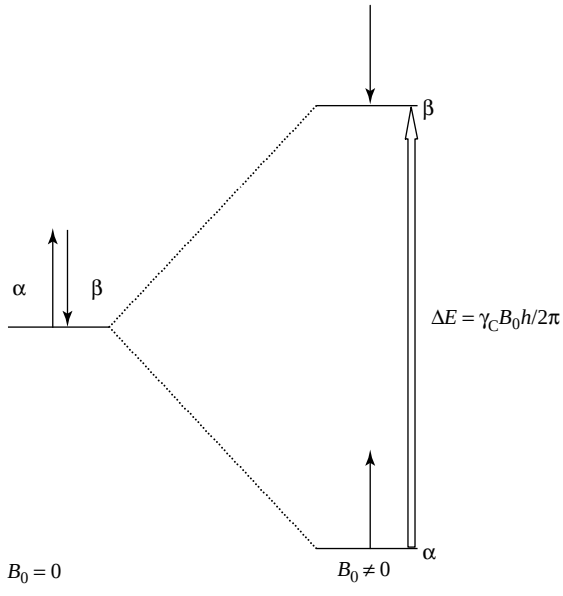

Figure 1 Energy-level diagram for nuclei with spin quantum number $I = 1/2$.

$T = 298\,\text{K}$ and $B_0 = 11.744\,T$, N_α/N_β is 1.00004, indicating the very small difference in equilibrium populations of the energy levels available for observation in the NMR experiment. This equation demonstrates that the population difference is dependent on B_0, being larger in a large magnetic field. Thus, the higher field increases the sensitivity of the NMR experiment.

The basic NMR experiment is shown in Figure 2. When a sample is exposed to a pulse of rf radiation at ν_0 with an intensity given by the field, B_1, the individual magnetization vectors begin to precess about this field as well as the external field B_0. This additional precession effectively tilts M_z away from B_0 through an angle θ, defined in Equation (5):

$$\theta = \gamma_C B_1 t_p \tag{5}$$

where t_p is the pulse duration time for B_1. Once in the transverse (x,y) plane, M_t precesses about B_0 at its individual precession frequency, ν_i, which is the sum of ν_0 and the chemical shift effect described in section 2.4. This precession, still about B_0, generates a voltage in the NMR receiver coil modulated at ν_i, and is detected as a function of time as shown in Figure 2(d). After digitization, the signal is analyzed with Fourier transformation to present the frequency components corresponding to the precession frequencies of individual ^{13}C species, shown in Figure 2(e).

During precession of M_t, numerous processes occur. Chemical shift effects cause nuclei of different electronic structure to precess at different rates, some faster and some slower than ν_0. Spin–spin relaxation processes cause the various components of M_t to dephase in the x,y plane, which gives rise to the finite width of the resonance line in the NMR spectrum. Spin-lattice relaxation also allows the individual μ_i to return to their equilibrium Boltzmann distributions. The resulting time-dependent signal is continually decaying due to relaxation and is termed the free induction decay (FID). Any scalar, dipolar or quadrupolar couplings also modulate the magnetization, observed as splittings or broadening of the resonances in the frequency spectrum. Once M_z has been re-established, the experiment can be repeated and the resulting signals added to increase sensitivity of the experiment.

Pulse NMR spectroscopy essentially detects the signals from all nuclei that resonate at or near the frequency of

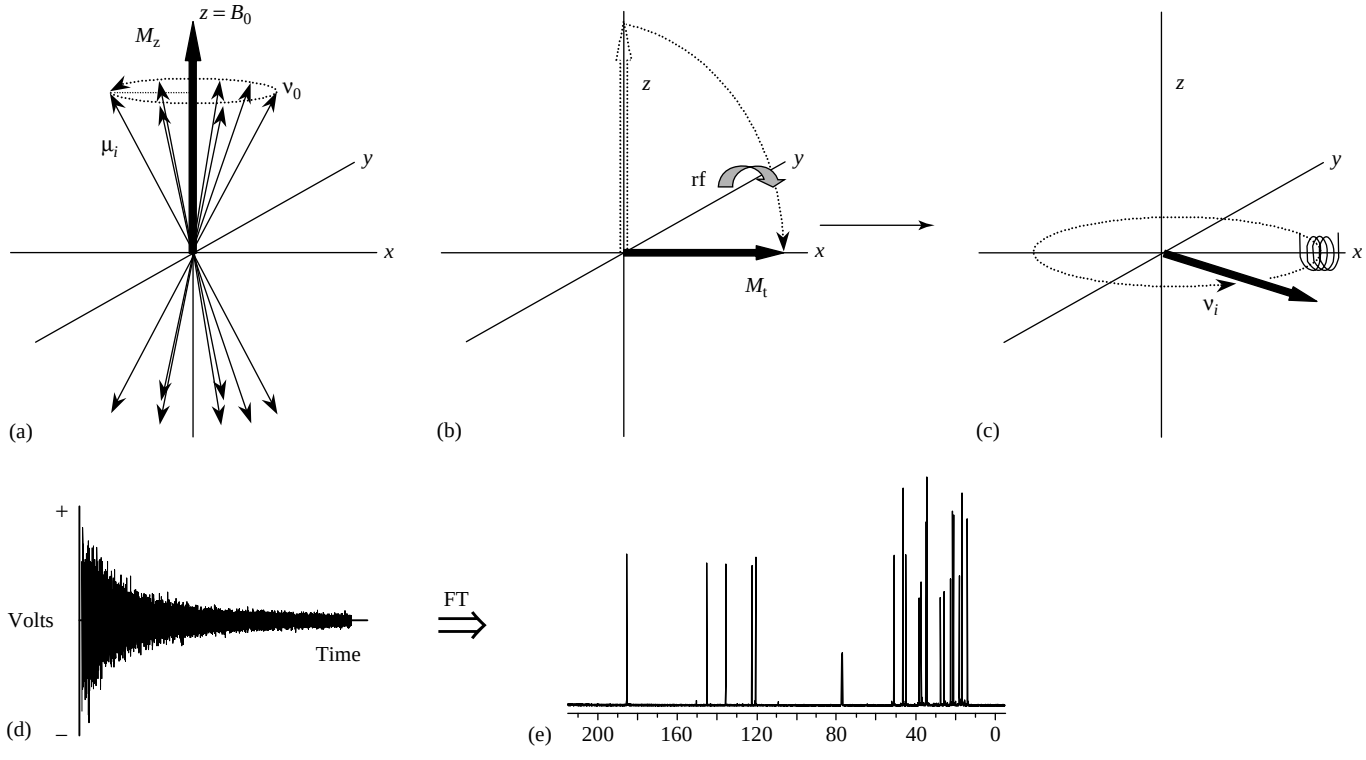

Figure 2 Diagram of the basic pulse NMR experiment. (a) Individual magnetic moments, μ_i, precess about the static magnetic field, B_0, at their Larmor frequency, ν_0. (b) An oscillating rf pulse, generates a field B_1, also at ν_0, perpendicular to B_0, tilting the net magnetic moment, M_z, into the transverse plane, to become observable magnetization M_x. (c) The observable magnetization vector M_x, then precesses about B_0 in the transverse plane (x,y plane) at its specific ν_i. (d) This oscillation induces a voltage in the NMR probe coil, which is measured as a function of time. (e) Fourier transformation of the resulting signal converts the time domain signal into its frequency components, yielding an NMR spectrum.

the high-power rf pulse. This high-power pulse has the effect of exciting a wide range of spectral frequencies determined by the relationship shown in Equation (6):

$$\text{Spectral width (SW)} = \frac{1}{2t_p} \tag{6}$$

where t_p is the time of the excitation pulse in seconds. A standard ''hard'' pulse is about 8.0 microseconds, corresponding to an effective excitation region of 62.5 kilohertz (kHz), sufficient to acquire a ^{13}C spectrum at all currently available magnetic field strengths.

The FID observed after a rf pulse is difficult to interpret, as a result of the complexity brought on by multiple frequency components which interfere with each other, creating an interferogram in time. FT resolves these components by converting the time-domain interferogram into a frequency spectrum using the mathematical operation shown in Equation (7):[60]

$$S(w) = \int_{-\infty}^{\infty} s(t)e^{-iwt}\,dt \tag{7}$$

where $S(w)$ is the frequency domain function and $s(t)$ is the time domain function. Current computer software uses the fast Fourier transform (FFT)[61] to complete this transformation rapidly to produce a frequency-domain spectrum with resonances at positions corresponding to their precession frequencies and whose intensities are proportional to the number of nuclei of each type in the sample.

2.2 Interactions Present in Carbon-13 Nuclear Magnetic Resonance Measurements

Magnetic resonance involves numerous interactions that must be described to obtain a complete picture of the information in an NMR spectrum. Several of these interactions will be discussed in further detail in following sections (chemical shift and NOE). The major interactions of interest are summarized in Table 1. The Zeeman interaction exists for all nuclei with a nonzero spin quantum number and is the interaction of the applied field, B_0 with μ_i, as discussed in section 2.1. The size of this interaction is linearly proportional to B_0, resulting in an increasing separation of energy levels with an increase in B_0.

Quadrupolar interactions occur when nuclei have a spin quantum number, $I > 1/2$; thus they do not affect ^{13}C spectra directly. The quadrupolar interaction arises from the interaction of a nuclear spin of $I > 1/2$ with an electric field gradient. The effect on a ^{13}C resonance is through interactions with other nuclei coupled to the electric field gradient through this interaction. The major determinants of this interaction's strength are the symmetry of the electronic structure around the nucleus, the correlation time of molecular motion (τ_C) and the size of the quadrupole coupling constant, shown in Equation (8).

Table 1 Interactions present in NMR

Interaction	Hamiltonian	Size (Hz)
Zeeman	$H_Z = -\gamma_C \hbar B_0 I_Z$	10^6–10^9
Quadrupolar	$H_Q = \overrightarrow{I} \cdot \hat{Q} \cdot \overrightarrow{S}$	0–10^9
Chemical shift	$H_{CS} = \gamma_C \hbar \overrightarrow{S} \cdot \hat{\sigma} \cdot \overrightarrow{H}$	0–10^5
Dipolar	$H_D = (\gamma_1 \gamma_C \hbar^2/r^3)\, \overrightarrow{I} \cdot \hat{D} \cdot \overrightarrow{S}$	0–10^5
Scalar	$H_{SC} = \overrightarrow{I} \cdot \hat{J} \cdot \overrightarrow{S}$	0–10^4

$$\chi = \frac{e^2 Qq}{h} \tag{8}$$

In solution ^{13}C-NMR, the dominant relaxation pathways are dipolar, but the fluctuating interaction of the quadrupole moment of quadrupolar nucleus with an electric field gradient can affect relaxation pathways of attached nuclei or those in close spatial proximity. In the case of ^{14}N, this can lead to broadening of ^{13}C spectral lines arising from directly attached nuclei while increasing the time for attached ^{13}C nuclei to return to their equilibrium spin distributions. Scalar couplings of quadrupolar nuclei with ^{13}C nuclei are also possible, typically observed for ^{2}H nuclei which have a small value for χ or a high degree of symmetry about the quadrupolar nucleus.

The chemical shift interaction is the most significant component of the NMR experiment, as it is the most diagnostic of the chemical environment around a nucleus, providing a spectral resonance for each chemically distinct nucleus in a molecule. It can best be described as the difference in Larmor frequencies of nuclei with differing electronic (chemical) environments. Effects of the electronic environment can shield or deshield the nucleus from B_0 relative to another resonance, in essence adding or subtracting an additional B_i magnetic field, so that the precession frequency of a chemically distinct nucleus, ν_i is not exactly the same as ν_0. This effect is defined by a magnetic shielding constant (σ). This constant has contributions from diamagnetic and paramagnetic shielding, magnetic anisotropy from surrounding nuclei, ring currents, conjugation, steric effects, electric fields, and intermolecular effects such as hydrogen bonding and solvation. Shifts from ν_0 are linear in B_0 and are measured relatively in parts-per-million (ppm). These shifts from ν_0 are on the order of several kilohertz (kHz) as compared to the 20 to 200 MHz range typical of ^{13}C Larmor frequencies at current magnetic field strengths. A more detailed discussion of the chemical shift can be found in section 2.4.

Scalar or spin–spin couplings result from interactions between adjacent magnetic dipoles that perturb the local magnetic fields around a nucleus. Since these dipoles may be aligned with or against B_0, they may increase or decrease the effective field at the nucleus being observed, causing a shift from ν_i. Since an adjacent dipole for ^{13}C is most likely ^{1}H nuclei, the states, α and β, have almost equal probabilities, this coupling splits the ^{13}C resonance into two lines of equal intensity centered on ν_i. Scalar couplings are indirect spin–spin couplings and occur through chemical bonds. A direct dipolar coupling occurring through space is also possible, but with rapid isotropic motion found in solution NMR, this coupling is averaged to zero. In solid-state NMR this direct coupling is a dominant effect and can be used as a means of efficient polarization transfer from abundant ^{1}H nuclei to dilute ^{13}C nuclei, enhancing the sensitivity of the ^{13}C NMR experiment. In the ^{13}C experiment, heteronuclear scalar couplings to protons are removed with broadband ^{1}H decoupling to remove the $^{1}H-^{13}C$ splittings. This simplifies the spectra and improves the sensitivity of the experiment. Scalar couplings are also used for polarization-transfer experiments to improve sensitivity or to modulate the ^{13}C resonances depending on the number of directly attached ^{1}H nuclei.

Dipolar couplings are direct spin–spin couplings that exist through space and not through chemical bonds as in the case of indirect scalar couplings. The local field around a ^{13}C nucleus created by a nearby ^{1}H dipole is shown in Equation (9):

$$B_{\mathrm{LOC}} = \pm \frac{\mu_i (3\cos^2\theta_{ij} - 1)}{r_{ij}^3} \tag{9}$$

where r_{ij} is the internuclear distance and θ_{ij} is the angle between r_{ij} and B_0. The dependence of B_{LOC} on θ_{ij} in a rapidly tumbling molecule results in a rapid modulation of the local field and provides an effective pathway for relaxation. The strong dependence on $1/r_{ij}^3$ means this is a localized phenomenon. Thus ^{13}C nuclei with directly attached ^{1}H nuclei will have more rapid and efficient relaxation pathways. Further discussion of dipolar interactions and relaxation will be presented in sections 2.5 and 2.7.

2.3 Heteronuclear Decoupling

Interaction of the numerous dipoles present in an organic molecule can complicate a ^{13}C-NMR spectrum and make structural identification a complex, if not impossible, task. Resonance splittings and line-width broadening due to scalar couplings decrease the sensitivity of the NMR experiment. For ^{13}C nuclei, directly attached protons are the dominant partner in scalar couplings, though longer-range couplings through multiple bonds also produce smaller couplings. The scalar coupling between two dipoles is designated $^{n}J_{ij}$, where n is the number of bonds connecting the two spins, and i and j are the identities of the scalar coupled nuclei. A typical carbon–proton one-bond scalar coupling, $^{1}J_{CH}$, is on the order of 125 to 150 Hz. Longer-range couplings diminish rapidly to several Hz for a $^{3}J_{CH}$ coupling. The number of lines, L, in a multipler from a ^{13}C resonance is given in Equation (10):

$$L = 2nI + 1 \tag{10}$$

where n is the number of attached nuclei and I is the spin quantum number of those spins. As an example, a methylene group has two hydrogens, both with a magnetic dipole and an $I = 1/2$, and would be split into three lines separated by $^{1}J_{CH}$. A fully proton coupled ^{13}C NMR spectrum is shown in Figure 3(a) for abietic acid, a tricyclic rosin acid isolated from pine tar. A methyne signal at 51 ppm is shown in the insert, with both the splitting and a significant broadening of the resonance due to longer-range couplings and dipolar couplings. The size of scalar couplings is dependent primarily on hybridization of the carbon atom (sp^3, sp^2 or sp) and the relative electronegativity of any substituents attached to the atom. While higher-order couplings may be positive or negative, $^{1}J_{CH}$ values are always positive.

While there is useful information in C,H scalar couplings, removing them is necessary to simplify ^{13}C spectra to permit interpretation of the underlying spectrum in terms of chemical groups. This process of heteronuclear decoupling yields a spectrum denoted as a $^{13}C\{^{1}H\}$ spectrum, where the nucleus in brackets is decoupled in the experiment.

There are numerous methods of spin decoupling. The most widely used is broadband decoupling. A second rf pulse, B_2, is applied to the sample at the ^{1}H Larmor frequency with sufficient power to cover the entire range of ^{1}H Larmor frequencies. The power required to do this increases with B_0. At 500 MHz, a decoupler field strength of 6 kHz would be required to decouple nuclei effectively over the entire range of carbon chemical shifts. To minimize the power required to decouple protons, methods have been developed using composite pulse decoupling (CPD) schemes which change the phase of the rf pulse, creating rapid spin transitions between α and β energy levels.[62,63] As long as the rate of transition exceeds the maximum scalar coupling found in a molecule, the ^{1}H spins are effectively decoupled from the ^{13}C spins. With this type of experiment, effective heteronuclear decoupling can be achieved at very low rf power levels, minimizing frequency offset problems and sample heating seen at higher power levels.

Experimentally, decoupling only needs to be done during the period in which the ^{13}C signal is acquired.

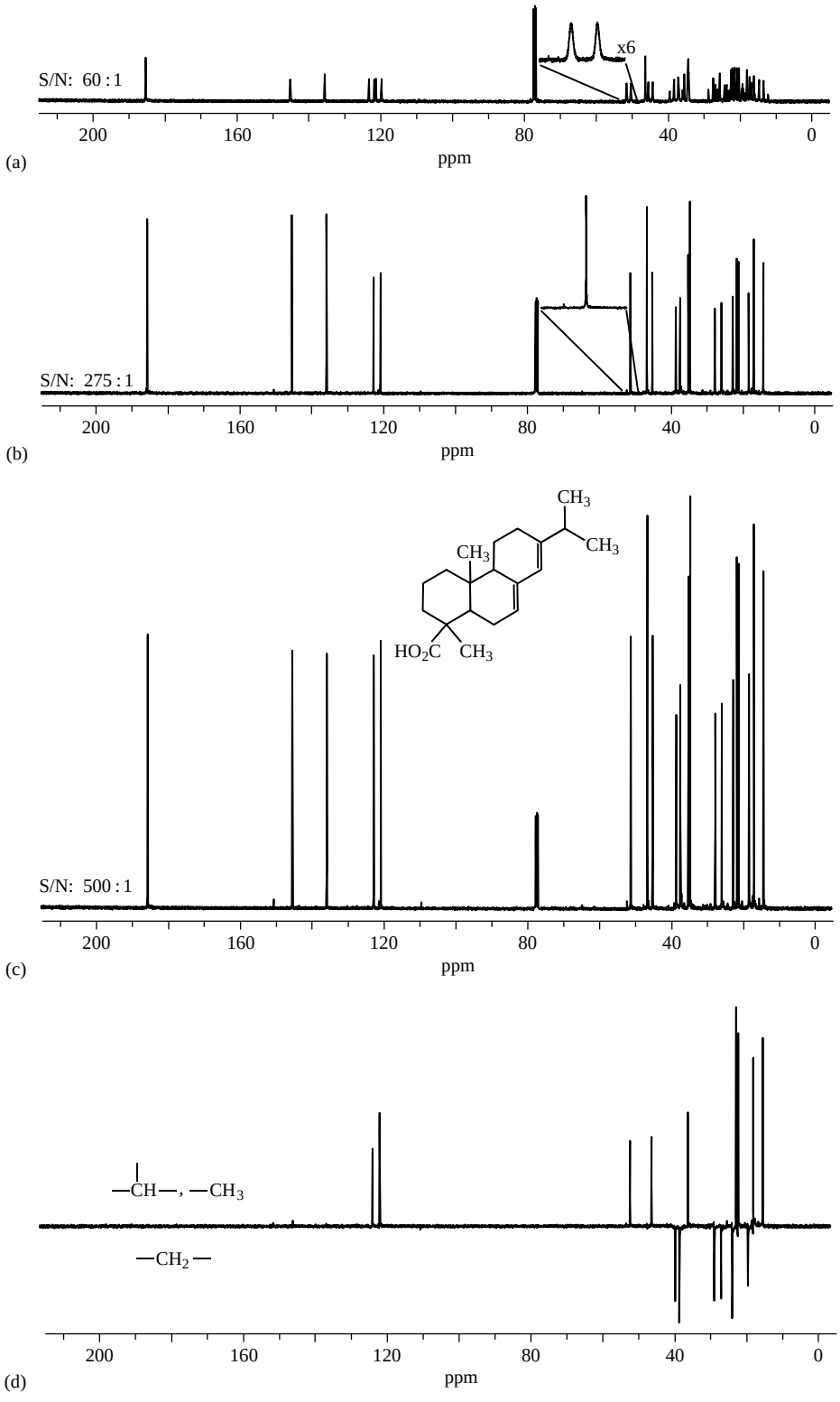

Figure 3 ^{13}C NMR spectra of abietic acid ($C_{20}H_{30}O_2$), a terpene resin acid acquired with: (a) single-pulse excitation and no ^{1}H decoupling; (b) single-pulse excitation with ^{1}H decoupling during acquisition only; (c) single-pulse excitation with ^{1}H decoupling during acquisition and relaxation delays; (d) DEPT experiment with ^{1}H decoupling during acquisition only. Expansions shown in (a) and (b) are of the same methyne signal at 51 ppm plotted on the same horizontal scale.

A $^{13}C\{^1H\}$ NMR spectrum of abietic acid is shown in Figure 3(b) using low-power CPD. Comparison to the fully ^1H coupled spectrum of Figure 3(a) demonstrates the enhancement in sensitivity and simplification of the NMR spectrum when decoupling is used. While the effects of scalar couplings are lost in decoupling experiments, there are other methods to retrieve the multiplicity of information they carry. In the inset of Figure 3(b), the methyne resonance of Figure 3(a) is shown in expansion. In addition to collapsing the doublet to a singlet, there is a very dramatic decrease in line width. This narrowing is the result of removing unresolved long-range scalar couplings of only a few hertz and residual dipolar couplings whose effects are to broaden the resonance line.

Enhancement of the ^{13}C NMR signal is also obtained if the ^1H B_2 field remains on during the relaxation or delay period between signal acquisitions. A comparison of Figure 3(c) with Figure 3(b) demonstrates the substantial gain in signal intensity when the B_2 field is left on during the relaxation delay after acquisition. This results from the NOE generated between the ^1H and ^{13}C nuclei. Section 2.5 explains this effect and its potential use in structural studies. Enhancement is dependent on the number of protons attached and the internuclear distance to the ^{13}C nucleus. Use of NOEs increases sensitivity, but often precludes quantitation of dissimilar functional groups because NOE enhancement rates differ from group to group.

Other methods of decoupling include selective decoupling (to establish connectivity between ^{13}C and ^1H nuclei), continuous wave (CW) decoupling, and off-resonance decoupling. In selective decoupling experiments, use of a low-power B_2 field whose frequency is matched to a specific ^1H resonance affects only ^{13}C resonances directly bonded or strongly dipolar coupled to the irradiated ^1H nucleus. These resonances will have NOE enhancement or decoupling, a fact that enables assignment of ^{13}C resonances from correlation with proton resonances. Repeated use of this kind of experiment allows one to assign the coupled resonances with a series of experiments. This technique is not used nowadays as often as it once was because multidimensional NMR methods have been developed that provide this same coupling information for all resonances in one experiment. CW decoupling is not often used in solution NMR as the power required is large on high-field spectrometers. This causes sample heating which can detune the NMR probe and cause broadening due to magnetic field inhomogeneities introduced through heating the sample. It is routinely used in solid-state NMR to remove the strong direct dipolar couplings in solid samples. In this case, heteronuclear dipolar couplings are very large, typically in the 20–80 kHz range for which low-power heteronuclear decoupling would be ineffective. Off-resonance decoupling is completed with the B_2 decoupling field moved away from the ^1H Larmor frequency, which retains the C,H couplings but scales them depending on the difference between their precession frequency and the applied B_2 field frequency. Off-resonance decoupling has been replaced by the more rapid and sensitive methods found in coherence-transfer techniques discussed in section 2.6.

2.4 Chemical Shifts in Carbon-13 Nuclear Magnetic Resonance Spectroscopy

The utility of the NMR experiment arises from the ability to distinguish one atom from another in the same molecule. This is the single most important capability of the technique, to study chemical structure at the atomic level. While the Zeeman interaction determines the average Larmor frequency of a nuclear type, i.e. ^{13}C versus ^1H, in a given magnetic field, local electronic shielding around a specific ^{13}C nucleus diminishes the strength of the B_0 field at a specific nucleus, thus shifting its Larmor frequency, ν_i. The chemical shift interaction, as shown in Table 1, contains a tensor element in its Hamiltonian. The chemical shift is an interaction determined by the shape and orientation of the molecular framework relative to B_0. In the extreme narrowing limit where random molecular reorientation is rapid, the chemical shift is determined by the average of the principal elements of this tensor – the isotropic value; this is the position of the resonance in a solution. On the other hand, solid-state measurements retain this quality of spatial chemical shift and require the use of high-speed sample spinning techniques to facilitate spectral observation of the isotropic-like spectrum.

Numerous factors can affect shielding, and thus the position of a resonance, including bonding, hybridization, functional groups, steric interactions and motion. The shift of the Larmor frequency of a specific nucleus from a defined reference position is the chemical shift, δ.

2.4.1 Introduction to Chemical-shift Concepts

The shielding of a nucleus by electrons is proportional to the strength of the applied magnetic field, B_0. This proportionality factor is the magnetic shielding constant, σ. The effective field at nucleus i can be defined in terms of σ and B_0, so that Equation (3) can be written as in Equation (11):[21]

$$\nu_i = \frac{\gamma_C}{2\pi} B_0(1 - \sigma_i) \tag{11}$$

where ν_i is the observed Larmor frequency for a nucleus and σ_i is the magnetic shielding constant for that nucleus. Each chemically different nucleus will have a unique

ν_i, dependent on the local electronic environment. This chemical shift is reported as a value relative to some reference standard. For ^{13}C spectra, the deuterated solvent used for magnetic field stabilization typically serves as an internal reference, having been set to define the chemical shift of tetramethylsilane (TMS) to be zero.

Chemical shifts are reported in parts-per-million, ppm, and denoted as δ. Historic convention established the units of ppm from the relationship in Equation (12):

$$\delta = \frac{\nu_i - \nu_{ref}}{\nu_{ref}} \times 10^6 \quad (12)$$

where ν_i is the resonance frequency of nucleus i, ν_{ref} is the resonance frequency of TMS or other known reference signal. As the difference $(\nu_i - \nu_{ref})$ is in hertz, the shift of Larmor frequencies is small and the ratio in Equation (12) is Hz/MHz or parts-per-million. Resonances to the left of the position (of higher Larmor frequency) in a spectrum have positive values of the chemical shift, and those to the right, negative values. Use of the ppm scale permits comparison of spectra acquired at different magnetic field strengths, since the fractional frequency shift should be independent of the magnetic field strength. For example, a resonance with $\nu_i = 7645$ Hz acquired at a B_0 field of 7.05 T ($\nu_1 = 75.45$ MHz), would have $\delta = 101.33$ ppm. On an instrument with $B_0 = 11.744$ T ($\nu_1 = 125.72$ MHz), the same resonance would have $\nu_i = 12\,739$ Hz and from Equation (12), $\delta = 101.33$ ppm. Resonances with positive chemical shifts are said to be deshielded relative to TMS (and in a nomenclature derived from experiments in which swept magnetic fields were used, downfield of TMS), and those with negative values relative to TMS are said to be more shielded (or in the older nomenclature, upfield of TMS). From Equation (11), σ_i values for deshielded resonances are smaller than those for more shielded nuclei. The terms upfield and downfield are also used in discussions of resonance positions relative to other signals in the NMR spectrum. The carboxyl resonance of acetic acid at 179 ppm is said to be downfield of the carboxylate resonance of methyl acetate at 171 ppm.

Resonance positions are sensitive to many external factors, and parameters such as temperature, pH, solvent and reference signal must routinely be cited when reporting chemical shifts. Both temperature and pH can affect the shift of resonance lines relative to a reference, particularly in aqueous samples, and should be controlled. Nuclei can also have different chemical shifts in different solvents through solvation effects and polarity of the solvent. An advantage of ^{13}C-NMR is the wide dispersion of chemical shift values, providing well-resolved NMR spectra. The standard range of chemical shifts is from -10 to 230 ppm, though there are materials that resonate outside this range such as transition-metal complexes, molecules with strongly electronegative substituents, paramagnetic centers and other structural features which distort the electron density around a ^{13}C nucleus.

2.4.2 *Factors Affecting Chemical Shift Positions*

Any structural feature which affects the electron density around a nucleus will affect σ, and as a result, δ. These include hybridization of the atom, charge density, substituent electronegativity, inductive effects, ring currents, steric interactions, unpaired electrons, hydrogen bonding, isotope effects, intermolecular interactions, solvent, temperature, pH and conformational properties of the molecule. Hybridization of the ^{13}C nucleus is primarily responsible for determining the chemical shift range for a specific nucleus. In general, sp^3 atoms resonate furthest upfield (at lowest frequency) in the -10 to 110 ppm region, sp^2 atoms in the 110 to 230 ppm region and sp atoms in the 80 to 120 ppm region of the NMR spectrum. Theoretical approaches to calculating σ_i are difficult and are not precise, due to the wide range of contributions to shielding and the complexity of completing such calculations even for relatively simple classes of molecules.

For the organic chemist, the most important contributions to the chemical shift are the electron distribution around the nucleus, characterized by σ_{dia} and σ_{para}, and neighboring-group effects including magnetic anisotropy, electric field effects, intermolecular interactions and ring-current effects in π-bonding groups. Any calculation of σ may be expanded to include terms for these effects, as shown in Equation (13):[22,64]

$$\sigma_i = \sigma_{dia} + \sigma_{para} + \sum \sigma_N + \sum \sigma_R + \sum \sigma_e + \sum \sigma_I \quad (13)$$

where σ_{dia} is the contribution from isotropic circulation of the spherically symmetrical electron cloud to shielding, σ_{para} is the nonspherically symmetric contribution, σ_N is the magnetic anisotropy contribution from neighboring atoms, σ_R represents ring current effects, σ_e are contributions from electric fields from polar groups in the molecule, and σ_I is a result of intermolecular contributions resulting from hydrogen bonding or solvation. The first three terms are applicable for all observed carbon resonances, while the other terms depend on the structure being studied and whether, for example, arene rings are present to contribute to σ_R or polar groups such as carbonyls are present to contribute intramolecular electric fields. Paramagnetic shielding is considered the dominant term for shielding in ^{13}C-NMR.

Isotropic circulation of the electrons surrounding a nucleus generates a spherically symmetrical field, creating a local magnetic field opposite in orientation to B_0. This field is proportional to the electron density around the nucleus and results in an upfield shift

of the resonance line as this local density increases. Diamagnetic shielding is inversely proportional to the distance between the nucleus and surrounding electrons, with s electrons contributing the dominant shielding term. Local contributions to this shielding from neighboring atoms are also included in calculations of the shielding parameter.[22,64] Numerically, σ_{dia} is a large term in the magnetic shielding, but is not considered to be the dominant factor in determining overall shielding and chemical shift of a ^{13}C nucleus.

Paramagnetic contributions arise from nonsymmetrical electron circulation around a nucleus as a result of chemical bonding and bond hybridization effects. Shielding of this nature can be described with Equation (14):[66]

$$\sigma_{para} = -\frac{e^2h^2}{m^2c^2}\Delta E^{-1} r_{2p}{}^{-3}[Q_{AA} + \sum Q_{AX}] \qquad (14)$$

where ΔE is the mean electron excitation energy, r_{2p} is the average distance between the nucleus and a $2p$ electron, and $[Q_{AA} + \quad Q_{AX}]$, the charge density bond order matrix. Contributions to paramagnetic shielding depend on accessibility of low-level excited states, the effective nuclear charge, number of electrons in p orbitals, and their distances from the nucleus. This latter contribution is considered the primary contribution to σ_{para}, producing the wide range of chemical shifts in ^{13}C spectra.

Neighboring anisotropic effects contribute to shielding of a nucleus due to the circulation of electrons on nearby atoms and of the currents generated through bonding electrons. This shielding term is defined in Equation (15):[22,67,68]

$$\sigma_N = \frac{1}{3R_{NB}^{-3}\Delta\chi_B(1 - 3\cos^2\theta_B)} \qquad (15)$$

where R_{NB} is the distance between the nucleus and dipole B, $\Delta\chi_B$ is the anisotropic term for the magnetic susceptibility of dipole B, and θ is the angle defined by the vector between the nucleus and dipole B and the symmetry axis of dipole B. A summation over all nuclei bound to the atom being observed and other nuclei in close proximity is done to account for the contribution of this term. From Equation (15) it is evident that this contribution to shielding depends only on the distance between the dipole and nucleus and the angle θ, while it is not dependent on the nature of the nuclei under observation.

Increasing electronegativity of a substituent produces a downfield shift of the α carbon signal due to the inductive nature of these functional groups. Attraction of electron density to the electronegative group reduces the electron density at the carbon nucleus, resulting in a deshielding of the nucleus. Unshared electron pairs also produce a downfield shift in a ^{13}C resonance, as their presence increases the contribution from σ_{para} to chemical shielding. Electron deficiencies found in positively charged species, also result in a downfield shift of ^{13}C resonances. Multiple bond conjugation effectively increases the shielding of the central nuclei in the delocalized bond environment, yielding an upfield shift of the affected resonances. Steric effects arise as a result of the close spatial proximity of atoms in a molecule. Repulsive forces from spatially close atoms may produce either upfield or downfield shifts, dependent on the molecular geometry. For example, γ–g-gauche effects in alkanes typically shift ^{13}C resonances upfield, but this effect is dependent upon spatial orientations and can be either shielding or deshielding.[69,70] Functional groups containing ionic charge also have electric fields that can distort the electron density and bond polarizability of neighboring carbon atoms, producing both up- and downfield resonance shifts.

Many factors influence the chemical shift of a nucleus. Rationalization of chemical shift positions based on a limited number of factors may result in erroneous attribution of a shift. Competing contributions between two or more structural features may make the interpretation of the spectrum in structural terms ambiguous. Molecular geometry, motion and exchange processes can also complicate spectra, making interpretation on first principles difficult. To understand an NMR spectrum fully, it is often beneficial to examine known samples of similar structure. In the following sections, representative examples of ^{13}C chemical shifts are presented, along with a survey of methods for calculating chemical shifts.

2.4.3 *Representative Examples of Carbon-13 Chemical Shifts*

An immense amount of literature exists about ^{13}C chemical shifts of organic molecules and an impossible task would be to summarize that literature completely. Numerous reviews and compendia of NMR spectra are available to provide more comprehensive tabulations.[21,22,46,71–79] Normal chemical shift ranges for organic functional groups are shown in Figure 4. The ranges presented are for carbon atoms α to the functional group or for the underlined carbon atom in the functional group in the left column. Other carbons in a molecule with this functional group will have resonances consistent with their structural identity, though shifted in a way that depends on the long-range shift effects of the functional group, typically up to four carbons away.

The ^{13}C NMR spectrum is, in the most simplified view, considered to have four regions where signals can be found and easily classified. The region from zero to 35 ppm is where most aliphatic carbons with only carbon

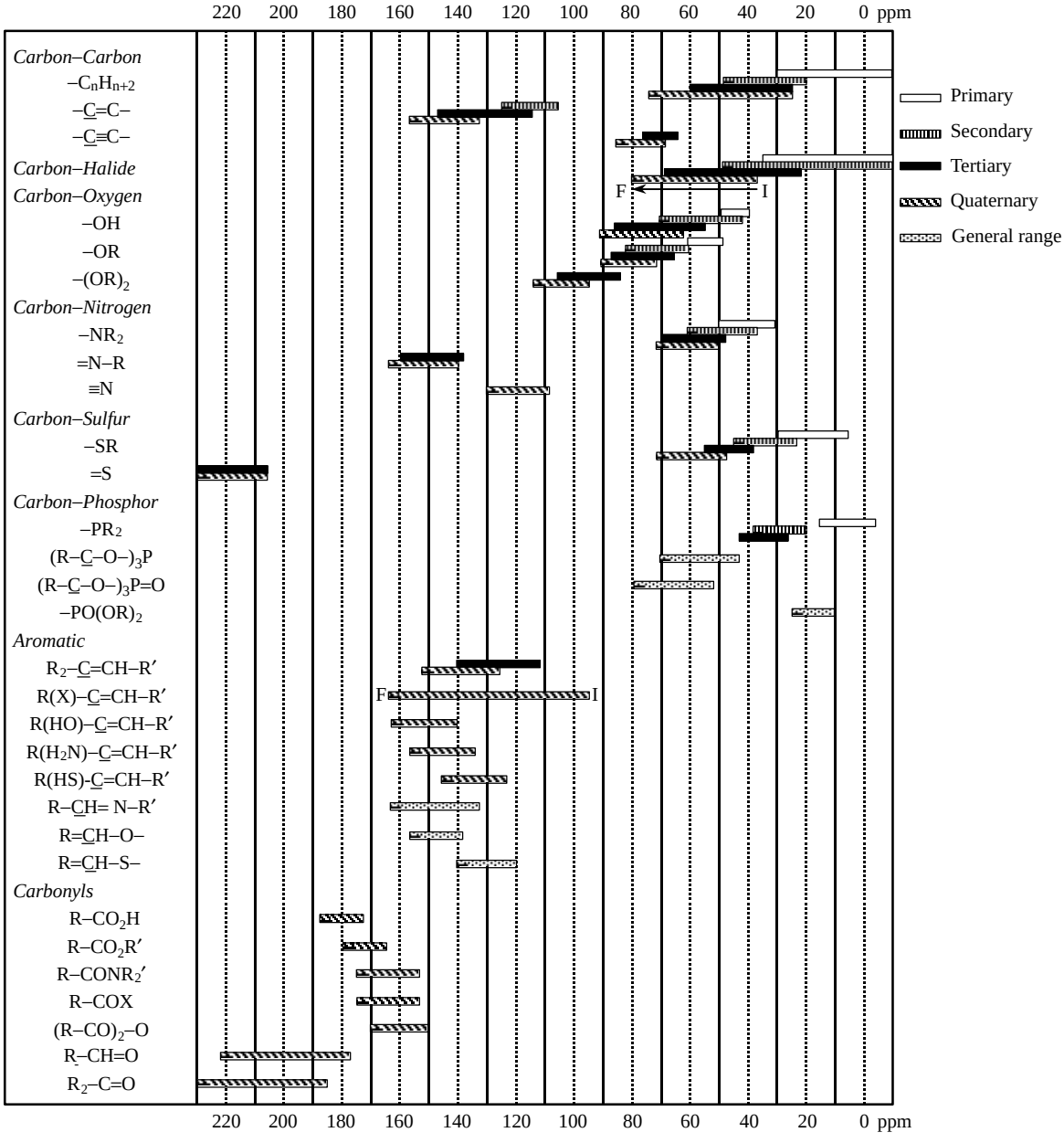

Figure 4 ^{13}C Chemical shift ranges for common organic functional groups.

or hydrogen substituents resonate. Functional groups on sp^3 carbons are found in the 40–110 ppm region, carbons with one or two oxygens resonating further downfield than carbons with nitrogens. Aromatic and olefinic sp^2 carbons appear in the 110–150 ppm region, the perturbation of *p* shell electrons producing a deshielding effect. Carbonyl resonances are found in the 150–220 ppm region of the spectrum, with aldehydes and ketones found shifted further downfield than carboxylic acids, esters and amides. Table 2 contains ^{13}C chemical shifts of a representative group of compounds from many areas of chemistry.

Some general observations can be discussed and a few representative spectra shown to illustrate the technique. Linear aliphatic groups contain distinct signals for a methyl at 14 ppm, a methylene α to the methyl at 23 ppm, β at 32 ppm and internal methylenes in the chain at 29 ppm. The spectrum of a mixture of C_8–C_{10} straight-chain carboxylic acids is shown in Figure 5(a), displaying a typical aliphatic pattern and a carboxylic acid resonance at 180.3 ppm. The peak at 34 ppm in the aliphatic region is the carbon α to the carbonyl group, owing to the downfield-shift contribution of the carbonyl. Other aliphatic compounds, i.e. amines, alcohols, halides,

Table 2 ^{13}C chemical shifts for representative organic structures[21,22,62,70,71]

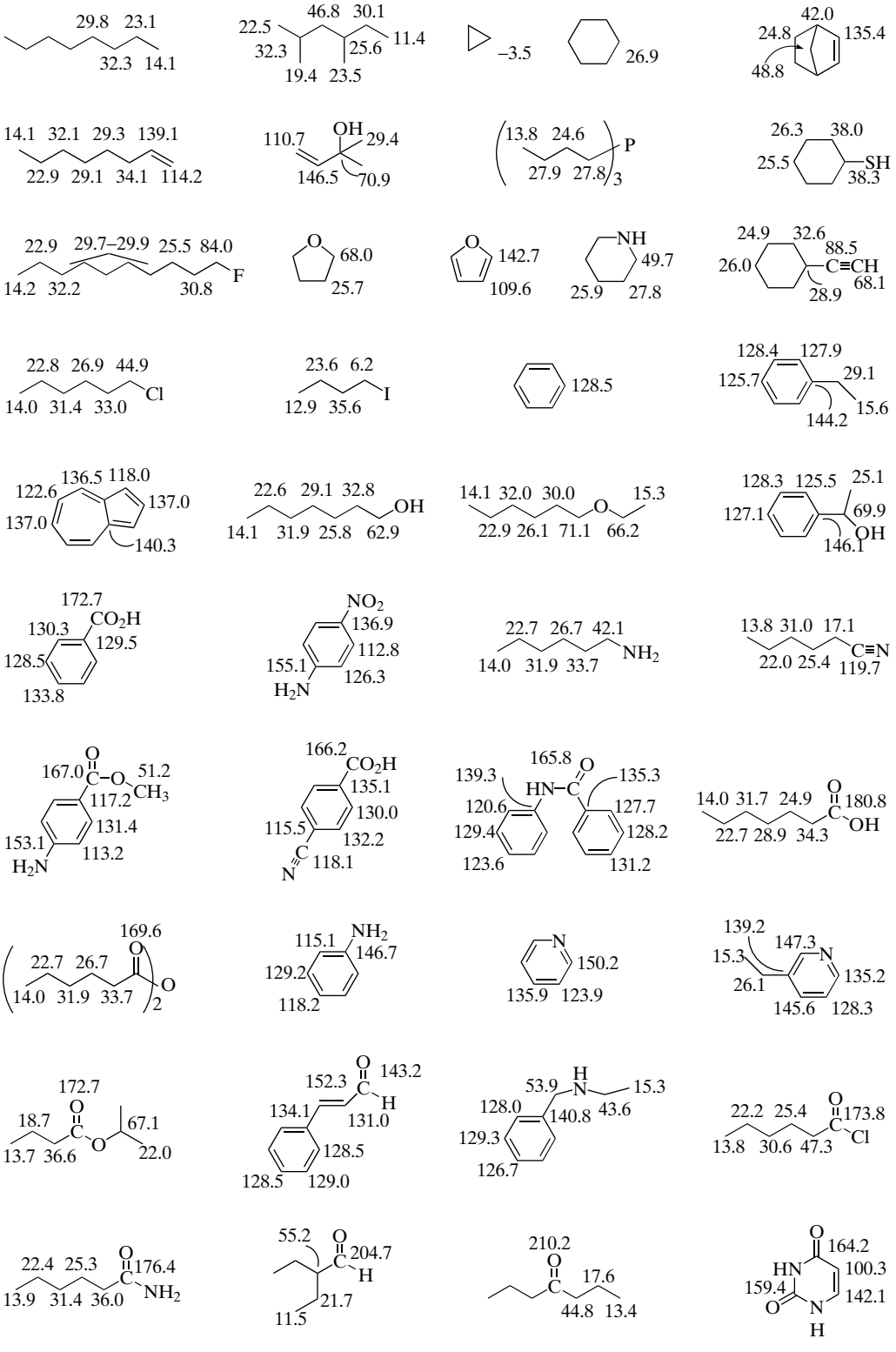

(*continued overleaf*)

Table 2 (*continued*)

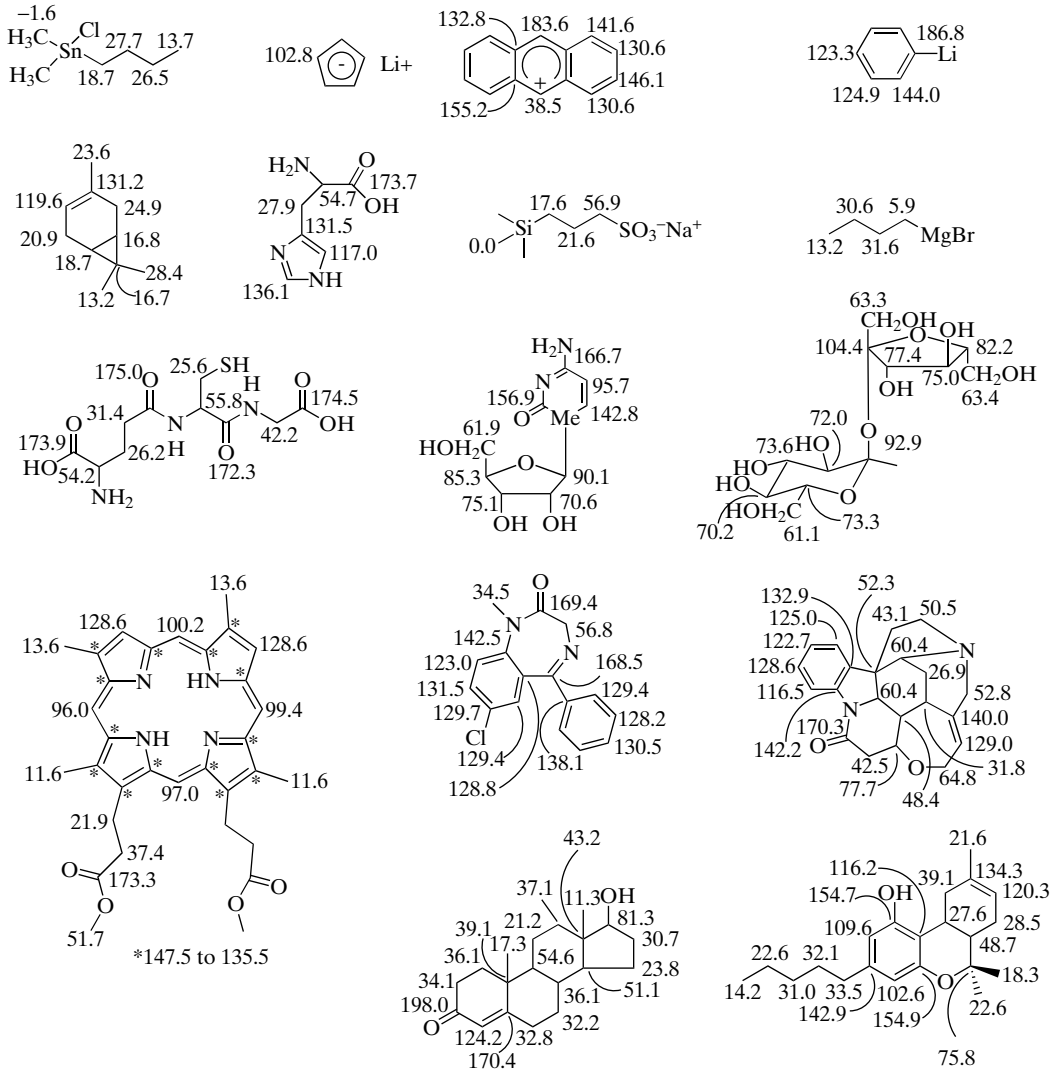

esters, ketones, etc. have similar aliphatic patterns, with the exception of the carbon α to the functional group. There is always some shifting due to longer-range effects, and any carbons that are part of a functional group such as an aldehyde carbonyl will be shifted.

The ^{13}C spectrum of a seaweed galactan is shown in Figure 5(b). The signals around 100 ppm arise from the acetal carbons of the pyranose rings which have two oxygen atoms bound to them. The region from 85 to 60 ppm contains signals from carbons with a single oxygen attached either as a secondary ether linkage or as a primary or secondary alcohol. The trend is for primary to be furthest upfield, with increasing substitution causing a downfield shift. The ^{13}C-NMR spectrum of a terpene mixture is shown in Figure 5(c), exhibiting a typically crowded aliphatic region from 10 to 50 ppm, olefinic and aromatic resonances from 110 to 150 ppm and various carbonyl signals in the 170–190 ppm region. Figure 5 also shows a division of the ^{13}C-NMR spectrum into four general regions, given below the spectral axis. For most organic compounds, the ^{13}C shifts are predictable, and a general conclusion about functionality can quickly be made by noting the presence of signals in these four regions. Inclusion of transition metals or the presence of charge can complicate the interpretation of spectra since these structural moieties can shift ^{13}C signals to regions where they would not typically be present or to regions well outside the normal 230 to −10 ppm spectral region.

2.4.4 Review of Methods for Calculating Chemical Shifts

Predicting or simulating ^{13}C-NMR spectra as an aid to the assignment of a spectrum or for developing a qualitative description of an organic structure can be approached

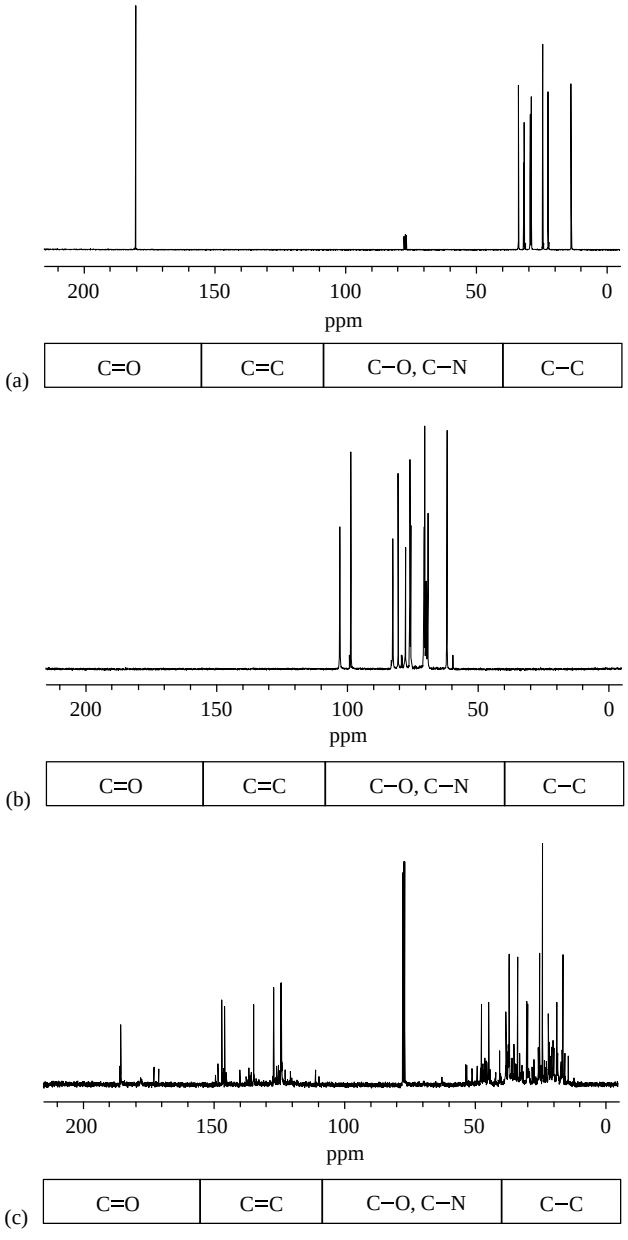

Figure 5 ^{13}C NMR spectra of (a) a C_8-C_{10} carboxylic acid mixture, (b) a seaweed galactan, and (c) a terpene resin mixture.

through several distinct methods. Prediction is typically done through empirical methods, i.e. developing additivity rules for classes of compounds or through semiempirical and ab initio calculation of the chemical shifts of a spin system by solving Schroedinger's equation for a molecule.

Development of additivity rules is an empirical exercise based on the examination of an appropriate number of model compounds through a regression analysis to determine substituent effects on ^{13}C chemical shifts. Numerous schemes exist (depending on the class of chemical compounds being examined) and current development in this area focuses on both new classes of compounds and computer methods for developing additivity relationships. Grant and Paul first pioneered these principles in their development of additivity rules for linear and branched hydrocarbons.[80,81] This work reported that the chemical shift of a carbon nucleus could be determined through the relationship (Equation 16):

$$\delta_{obs} = -2.3 + n_\alpha S_\alpha + n_\beta S_\beta + n_\gamma S_\gamma + n_\delta S_\delta + n_\varepsilon S_\varepsilon + C \quad (16)$$

where S_α is the increment in chemical shift due to the α substituent and n_α is the number of these substituents α to the carbon in question, C is an additional term related to the substitution of the nucleus observed and the most highly substituted α carbon, and -2.3 is the reference chemical shift of unsubstituted methane. The incremental shift parameters for the Grant–Paul equation are given in Table 3.

Further development of this relationship was reported by Linderman and Adams, in a paper in which they extended the relationship to that shown in Equation (17):[22,69,82]

$$\delta_C(k) = B_S + \sum_{M=2}^{4} D_M A_{SM} + \gamma_S N_\gamma + \delta_S N_\delta \quad (17)$$

where $\delta_C(k)$ is the chemical shift of the kth carbon, B_S, A_{SM}, γ_S and δ_S are constants from Table 4.

For these constants S is the number of carbons bonded to the nucleus whose shift is being calculated, D_M is the number of carbons bonded to k with M carbons attached, with N_γ and N_δ representing the number of carbons γ and δ to carbon k respectively. A comparison of results obtained using this method, the Grant–Paul equation, prediction from a spectral database software program and experimentally determined results for 3-methylpentane is shown in Figure 6. As seen in this figure, all three methods provide reasonable estimates for δ values of the example hydrocarbon. These rules were subsequently extended to substituted hydrocarbons using an equation of the

Table 3 Shift parameters for Grant–Paul additivity relationship[79]

Position	S	Number of carbons attached to:		C
		Observed carbon	α Carbon	
α	9.1	1	3	−1.1
β	9.4	1	4	−3.4
γ	−2.5	2	3	−2.5
δ	0.3	2	4	−7.2
ε	0.1	3	2	−3.7
		3	3	−9.5
		4	1	−1.5
		4	2	−8.4

Table 4 Shift parameters for the Linderman–Adams additivity relationship[80]

B_1	6.80	A_{12}	9.56	γ_1	−2.99	δ_1	0.49
		A_{13}	17.83				
		A_{14}	25.48				
B_2	15.34	A_{22}	9.75	γ_2	−2.69	δ_2	0.25
		A_{23}	16.70				
		A_{24}	21.43				
B_3	23.46	A_{32}	6.60	γ_3	−2.07		
		A_{33}	11.14				
		A_{34}	14.70				
B_4	27.77	A_{42}	2.26	γ_4	0.68		
		A_{43}	3.96				
		A_{44}	7.35				

same form as Equation (16), but with the coefficients shown in Table 5 for some common organic functional groups.[21,22,83–86]

Predictive models of this type have been extended to cyclic alkanes,[87–89] substituted cyclic alkanes,[90,91] alkenes and substituted alkenes[92–94] and substituted benzenes.[21,22,69,95–97] Table 5 contains the additivity constants for calculating the chemical shifts of substituted benzenes. A base shift of 128.5 ppm is used for all positions and the appropriate shift contributions are added or subtracted from this number for each substituent on the aromatic ring for each carbon in the aromatic ring. Comparison of calculated and experimental values for *p*-fluorotoluene are given in Figure 7. Methods of this type employing empirically derived relationships provide

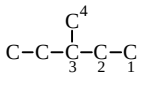

C_3 from Grant–Paul additivity rules:

$$\delta_{obs} = -2.3 + n_\alpha S_\alpha + n_\beta S_\beta + n_\gamma S_\gamma + n_\delta S_\delta + n_\varepsilon S_\varepsilon + C$$

$$\delta_{obs} = -2.3 + (2 \times 9.1) + (3 \times 9.4) + 0 + 0 + 0 + (-7.4)$$

$$\delta_{obs} = 36.7 \text{ ppm}$$

C_3 from Linderman–Adams additivity rules:

$$\delta_C(k) = B_S + \sum_{M=2}^{4} D_M A_{SM} + \gamma_S N_\gamma + \delta_S N_\delta$$

$$\delta_C(k) = 23.46 + (2 \times 6.6) + 0 + 0$$

$$\delta_C(k) = 36.7 \text{ ppm}$$

Carbon	δ (ppm)			
	1	2	3	4
Grant–Paul	11.5	29.7	36.7	19.5
Linderman–Adams	10.9	29.4	36.7	18.7
Database prediction	11.3	29.2	36.6	18.6
Experimental	11.4	29.3	36.4	18.8

Figure 6 Predicted chemical-shift values for 3-methylpentane using Grant–Paul[78,79] and Linderman–Adams[22,59,80] additivity formulas and constants from Tables 3 and 4.

Table 5 Additivity constants for calculating ^{13}C chemical shifts of substituted alkanes using Equation (16)[21,22,81–84]

Functional group	S_α	S_β	S_γ	S_δ
$-CH_n-$	9.1	9.4	−2.5	0.3
−C=C−	21.5	6.9	−2.1	0.4
−C≡C−	4.4	5.6	−3.4	−0.6
$-(C_6H_5)$	22.1	9.3	−2.6	0.3
−OH, −OR	49	10.1	−6	0.3
$-NH_2$, $-NR_2$	28.3	11.3	−5.1	0.3
$-CO_2H$	20.1	2.0	−2.8	0.0
$-CO_2R$	22.6	2.0	−2.8	0.0
−O(O=C)−R	54.5	6.5	−6.0	0.0
−NH(O=C)−R	28.0	6.8	−5.1	0.0
−CHO	29.9	−0.6	−2.7	0.0
−(C=O)−R	22.5	3.0	−3.0	0.0
−S−	10.6	11.4	−3.6	−0.5
−CN	3.1	2.4	−3.3	−0.5
$-NO_2$	61.6	4.4	−4.6	−0.9
−F	70.1	7.8	−6.8	0.0
−Cl	31.0	10.0	−5.1	−0.5
−Br	18.9	11.0	−3.8	−0.7
−I	−7.2	10.9	−1.5	−0.9

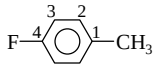

Carbon	δ (ppm)			
	1	2	3	4
Fluorine	−4.4	0.9	−14.4	35.1
Methyl	8.9	0.7	−0.1	−2.9
Base shift	128.5	128.5	128.5	128.5
Calculated shift	133.0	130.1	114.0	160.7
Experimental	133.3	130.5	115.0	161.4

Figure 7 Experimental and calculated chemical-shift values determined for *p*-fluorotoluene using the shift parameters shown in Table 6.[21,22,69,93–95]

reasonably accurate estimates of ^{13}C chemical shifts and are widely used as interpretative aids for structural analysis.

Progress in the areas of chemical shift prediction and additivity relationships has focused on the use of different computational methods to derive the correlations necessary for prediction within classes of compounds. Recent studies have employed computerized methods using multilinear regression (MLR) analysis or artificial neural networks (ANN) for the development of additivity relationships.[98–108] Creation of large chemical databases containing known chemical structures or fragments with correct chemical shift assignments has fostered development of computer software capable of finding structures to match an experimental spectrum, predict the structure of an unknown from a spectrum or construct an NMR spectrum from an input structure.[109–111] These methods are rapid and can be tailored to the chemistry of interest.

Table 6 ^{13}C shift parameters for functional groups on aromatic rings[21,22,69,91,93,94]

Functional group	$C_{attached}$	*o*	*m*	*p*
$-CH_n-$	8.9	0.7	−0.1	−2.9
−C=C−	9.5	−2.0	0.2	−0.5
−C≡C−	−5.8	3.9	0.1	0.4
$-(C_6H_5)$	13.1	−1.1	0.4	−1.2
−OH	26.9	−12.7	1.4	−7.3
−OR	31.4	−14.4	1.0	−7.7
$-NH_2$	18.0	−13.3	0.9	−9.8
$-NR_2$	19.9	−15.3	1.4	−12.2
$-CO_2H$	2.1	1.5	0.0	5.1
$-CO_2R$	1.3	−0.5	−0.5	3.5
−CHO	9.0	1.2	1.2	6.0
−(C=O)−R	7.6	−1.0	−1.0	3.5
−SH	2.3	1.1	1.1	−3.1
−SR	10.2	−1.8	0.4	−3.6
−CN	−15.4	3.6	0.6	3.9
$-NO_2$	20.0	−4.8	0.9	5.8
−F	35.1	−14.4	0.9	−4.4
−Cl	6.4	0.2	1.0	−1.9
−Br	−5.9	3.0	1.5	−1.5
−I	−32.3	9.9	2.6	−0.4

Commercial packages also ''learn'' to predict spectra or structures as the basis set of structures for prediction is expanded. Empirical methods are rapid and provide the most utility for analytical spectroscopy.

Ab initio and semiempirical methods are used to solve Schroedinger's equation for a molecule, a description of the energy states in the system. Chemical shifts are calculated after adding the perturbation effects of H_0, and related to the total energy of the molecule and nuclear magnetic moment mathematically. The electronic environment around a nucleus has an orientation in the molecular framework and is sensitive to the direction of the applied magnetic field, thus the chemical shift must be described as a tensor to account for all possible orientations of the molecule and the magnetic field. In liquids and gases with isotropic molecular motion, this directional dependence is averaged out and the observed chemical shift is the average of the principal tensor elements.

Treatment of magnetic shielding in a molecule has been done through numerous theoretical frameworks, ranging from Rayleigh–Schroedinger perturbation theory[112,113] to self-consistent field (SCT) theory.[114–117] An exact solution of Schroedinger's equation for typical organic molecules is not possible and approximate wave functions are used to obtain solutions. Linear combinations of atomic orbitals (LCAOs), are used to derive a basis set of molecular orbitals. The orientational dependence of the chemical shift of a carbon in a molecule in a magnetic field cannot be adequately described with normal wave functions because they cannot describe the gauge invariance of these functions. A larger basis set could be used, but this approach would not be viable for larger molecules. The introduction of gauge independent atomic orbital methods (GIAO), which use local gauge origins for LCAOs, made possible a comprehensive framework for calculation of chemical shifts.[118–122] Additional methods such as individual gauge for localized orbitals (IGLO) and localized orbitals, localized origin (LORG) are also available and are computationally faster methods.[123–128] However, development and applications of the GIAO abound in the literature, focusing on optimization of calculations and the application to progressively larger molecules.[128–137]

Semiempirical methods introduce approximations to the formalisms used in ab initio methods to simplify expressions for describing the energies and associated wave functions of complex molecular systems. Theoretical approximations for describing diamagnetism were first described by Pople and still referred to in current studies of ^{13}C chemical shifts.[138,139] Various approaches to semiempirical calculations exist based on bond polarization theory,[140] average excitation energy approximations,[141] valence bond theory,[142] and empirical parameterization from previous experimental results.[143–145]

While more complex and with the potential to relate chemical shifts to molecular structure at a basic level, these methods are more qualitative in their results, as a result of assumptions and approximations required to make computation manageable. Difficulties also arise in accounting for all the potential electronic and nuclear interactions that may exist in the molecular spin systems. Coupled with this problem are any potential internuclear perturbations from solvents or through interactions such as hydrogen bonding. While these methods are developing rapidly as computer calculation speeds increase and theoretical treatments are refined, this methodology is not generally of practical use for the practicing analytical chemist completing routine structural analysis. However, ab initio and semiempirical methods can provide insight into structural phenomena such as folding, binding, and the orientation of the chemical shift tensor with respect to the molecular framework, an insight into the electronic environment around a nucleus.

2.5 Nuclear Overhauser Effects

When heteronuclear dipolar couplings exist between two spins, their relaxation behavior is interrelated and changes in one spin reservoir will have an effect on the entire coupled system. In ^{13}C-NMR spectroscopy, the dominant dipolar couplings occur between ^{13}C nuclei and their directly attached ^{1}H spins, serving as the primary relaxation pathway for protonated ^{13}C spins. Use of heteronuclear decoupling to suppress scalar couplings

causes a saturation of the ^{1}H spin system, altering the energy-level populations for ^{13}C–^{1}H dipolar coupled systems. This perturbation and resultant enhancement of the ^{13}C-NMR signal is known as the NOE.[146–149]

2.5.1 *Origin of the Nuclear Overhauser Effect in Heteronuclear Spin Systems*

Dipolar coupling of nuclear spins alters the energy-level diagram shown in Figure 1 to that for a two-spin system shown in Figure 8. In this spin system, there are four single-quantum nuclear-spin transition probabilities with P_{13} and P_{24} corresponding to ^{1}H spin transitions and P_{12} and P_{34} denoting the corresponding ^{13}C spin transitions. Single-quantum transitions are those responsible for spin-lattice relaxation processes (see section 2.7). The transitions P_2 and P_0 represent double-quantum and zero-quantum transitions respectively, which cannot be directly excited with rf pulses or directly observed in the NMR spectrum. Through dipole–dipole relaxation processes, these two transitions do occur and are not affected by other relaxation processes that affect the single-quantum transitions.

Saturation of ^{1}H transitions with a continuous rf pulse, as found in ^{13}C-NMR experiments where ^{1}H spins are decoupled to remove scalar couplings, produces a nonequilibrium spin distribution. Continuous saturation produces the state where $P_{13} = P_{31}$ and $P_{24} = P_{42}$ so that these relaxation processes cannot re-establish thermal equilibrium populations in each of the four energy levels.

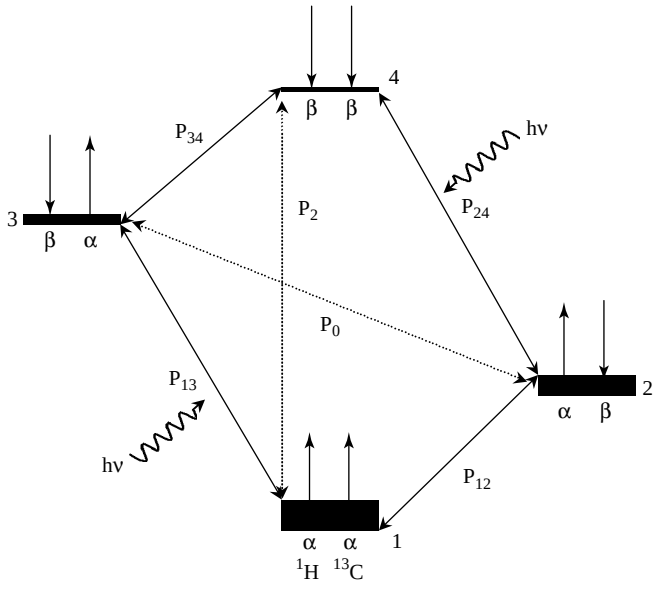

Figure 8 Energy-level diagram for a two-spin, A,X system with single quantum transitions shown in solid lines and zero (P_0) and double (P_2) quantum transitions represented with broken arrows.

The effect of this saturation is equalize the populations so that $N_1 = N_3$ and $N_2 = N_4$. At this point, the zero- and double-quantum dipolar relaxation processes, with P_2 and P_0 transition probabilities, begin to establish equilibrium, population distributions. Since N_1/N_4 is less than that found at equilibrium, repopulation of N_1 through the P_2 pathway proceeds, decreasing N_4 and increasing N_1. The net effect is to increase the population differences $N_3 - N_4$ and $N_1 - N_2$. ^{13}C signal intensities are directly related to these population differences, so that any increase in these differences will produce an increase in the ^{13}C signal intensity. However, the zero-quantum relaxation pathway, with transition probability P_0, is also active, attempting to re-establish the N_2 equilibrium population while reducing the N_3 population. This would decrease the $(N_1 - N_2)$ and $(N_3 - N_4)$ population differences and reduce the ^{13}C NMR signal intensity.

A competition exists between these two processes, one which would increase the ^{13}C signal intensity and one which would decrease the intensity. Molecular motion determines which process is more effective and, consequently, whether the NOE will produce a larger ^{13}C signal. For molecules with short molecular correlation times (τ_C), the P_2 transition probability prevails and the ^{13}C signal enhancement is positive. Since this is a double-quantum pathway, the dipolar modulation frequency required to initiate these transitions is the sum of Larmor frequencies of the nuclei involved ($\nu_H + \nu_C$) in the process. The zero-quantum transition has a much smaller frequency requirement ($\nu_H - \nu_C$) and is the favored transition when τ_C is longer, indicating slower molecular tumbling. Small organic molecules generally have a positive signal enhancement, while larger molecules, such as biopolymers, may have zero or negative NOEs. Since both P_2 and P_0 depend on both the Larmor frequencies and τ_C, NOEs are magnetic-field-dependent. Since τ_C is not field-dependent, increasing magnetic field strengths may produce smaller or negative NOE values when $(w_H + w_C)\tau_C \approx 1$.

The NOE enhancement factor (η) is dependent on the magnetogyric ratios of the nuclei involved in the dipolar coupled processes, as shown in Equation (18):

$$\eta_{(NOE)} = \frac{\gamma_H}{2\gamma_C} \tag{18}$$

which for ^{13}C{^{1}H} experiments is 1.988. For a ^{13}C spectral resonance with a full NOE, the signal intensity would then be (Equation 19):

$$I = I_0(1 + \eta_{(NOE)}) \tag{19}$$

where I is the NOE-enhanced intensity and I_0 is the equilibrium peak intensity. Full enhancement is realized when dipolar relaxation is the principal spin-lattice

relaxation mechanism for a molecule tumbling in the extreme-narrowing region of motion. In many cases, full enhancement is not obtained due to numerous effects such as multispin effects on NOE rates and transition probabilities.

Measurement of NOE values are carried out by several methods, the most common involving the acquisition of two spectra, one with the ^{1}H decoupler on during the delay between the end of spectral acquisition and the next observation pulse, and the second with the decoupler off during this period. The delay between acquisition should be longer than five times the longest spin-lattice relaxation time to allow the ^{13}C spin system to reach thermal equilibrium. The spectral intensities are ratioed to determine the NOE enhancement factor. Samples should be prepared in fully deuterated solvents and care should be taken to eliminate paramagnetic impurities such as O_2 from the solvent.

2.5.2 *Applications of Nuclear Overhauser Effect Measurements*

Heteronuclear NOEs are used to enhance the signal sensitivity for nuclei with a small γ and/or a low isotopic natural abundance, both categories include ^{13}C nuclei. Other nuclei also exhibit strong NOEs, such as ^{15}N, ^{29}Si and ^{31}P. Since some of these nuclei have negative magnetogyric ratios, their NOE enhancements may be negative and unfortunate selection of experimental conditions can actually destroy sensitivity or completely suppress resonances. As the NOE values obtained for different structural moieties and molecules of differing size may vary, the use of NOEs often precludes accurate quantitative analysis and transforms ^{13}C-NMR into a semi-quantitative or qualitative method. As seen in Figure 3(c), the spectrum of abietic acid acquired with steady-state NOE has a greater S/N than those acquired (a and b) where NOE was not active.

Selective NOE measurements are a valuable aid in structural elucidation and conformational analysis. Use of ^{13}C measurements is limited to nonprotonated atoms since ^{1}H$-^{13}$C NOEs are dominated by directly attached protons and no spatial information is likely to be obtained from protonated carbon resonances. Dipolar relaxation rates have a $1/r_{ij}^6$ dependence, making the NOE phenomenon a local one. Selective saturation of a single ^{1}H resonance will provide an enhancement of ^{13}C nuclei within effectively 2.5 Å of the proton under irradiation. This should limit enhancement to directly bonded spins or those in close spatial proximity. The former serves as an assignment aid,[150,151] while analyzing the NOEs of the latter can be used to determine molecular conformation or association between different molecules.[152–155] Less-than-complete enhancement can be used to determine internuclear distances.[156,157] Hydrogen bonding can be explored through saturation of proton resonances and searching for enhancements in the ^{13}C resonances of atoms which may participate in bonding with the protons.[158] NOE measurements have also found application in determining the chain microstructure in synthetic polymers.[159]

The low sensitivity inherent in ^{13}C-NMR limits the use of selective NOE measurements. Polarization transfer methods have been developed which provide proton multiplicities and increased sensitivity to replace NOE methods. ^{13}C NOE measurements have largely been usurped by a combination of ^{1}H detected heteronuclear (^{13}C$-^1$H) and homonuclear (^{1}H$-^1$H) 2-D (two-dimensional) measurements which are much more rapid and contain substantially more information for the study of molecular conformation and dynamics. The development of high-field superconducting NMR spectrometers has made acquiring and interpreting homonuclear ^{1}H$-^1$H NOE data less complicated and more informative than heteronuclear NOE methods. Multidimensional HETCOR experiments designed to provide carbon–proton bonding networks have replaced selective NOE experiments for both assignment confirmation and, through measurement of $^3J_{CH}$ couplings, a wealth of conformational data can be determined in one spectrum. These techniques are part of a class of pulse sequences called inverse methods since the NMR spectrum is detected as a ^{1}H signal to improve sensitivity. Pulsed polarization transfer methods can also be used for multiplicity determinations in ^{13}C spectra. Development of these techniques has limited the scope of applications for ^{13}C$-^1$H NOE measurements to instances where results of ^{1}H experiments are ambiguous.

2.6 Coherence Transfer in Carbon-13 Nuclear Magnetic Resonance Measurements

The multiplicities of ^{13}C-NMR signals are important structural parameters for completing spectral resonance assignments. Examination of a ^{1}H-coupled ^{13}C spectrum to determine the number of attached hydrogens can be complicated by overlapping resonance splitting patterns, as seen in Figure 4(a) and also suffers from an inherent insensitivity due to the loss of resonance intensity when sharp single lines are split into multiplet structures. Alternative methods for determining the multiplicity of ^{13}C lines rely on pulsed polarization transfer to increase ^{13}C signal sensitivity while the NMR signals are modulated as a result of $^1J_{CH}$ couplings to produce multiplicity information. Numerous polarization-transfer methods exist, though two nonselective methods that find the most utility in ^{13}C-NMR are INEPT,[13,14,35] and DEPT.[15,16]

Polarization transfer is somewhat analogous to the NOE experiment since both processes lead to an enhancement of ^{13}C-NMR signals, the former using scalar

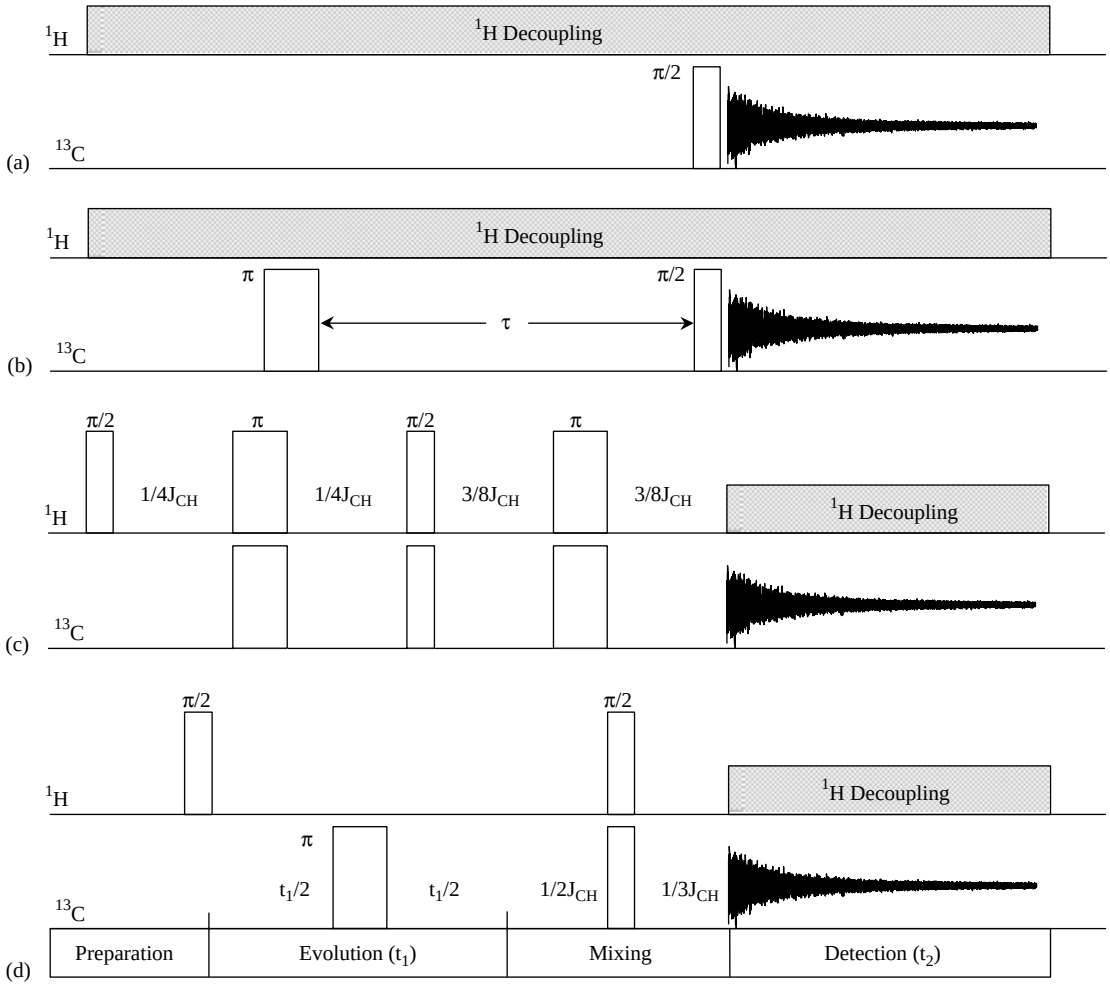

Figure 9 Pulse sequences used to acquire (a) a ^{13}C NMR spectrum with NOE and, (b) an inversion-recovery experiment for measurement of T_1, (c) a refocused INEPT spectrum with a positive phase for CH and CH_3 groups and a negative phase for CH_2 groups and (d) the sequence for acquiring a 2-D, ^{13}C–^{1}H HETCOR spectrum.

couplings through direct bonding pathways in contrast to the through-space dipolar couplings active in NOE. As a consequence, only ^{13}C nuclei with directly bonded protons will be observable in the spectrum obtained with INEPT, DEPT and other polarization-transfer schemes. Polarization transfer is completed using a series of pulses to effect transfer through spin reorientation, rather than saturation of ^{1}H energy levels with continuous irradiation, as is done in the NOE experiment. The pulse sequence required to effect transfer is shown in Figure 9(c) and compared to a standard 1-D ^{13}C{^{1}H} pulse sequence with NOE in Figure 9(a). Polarization-transfer sequences of this type are also the fundamental units of heteronuclear multidimensional NMR sequences discussed in section 2.8.

The basic premise of polarization transfer lies in the ability to change ^{13}C spin distributions (α and β) to that of directly attached ^{1}H spins at equilibrium. Observable ^{1}H magnetization is generated in the transverse plane with a nonselective pulse. Precession of the magnetization vectors occurs for a time period of $1/4J_{CH}$. This time delay serves to label the ^{1}H magnetization components by the orientation (α or β) of their coupled ^{13}C spin components. A refocusing/inversion pulse is then applied to both ^{1}H and ^{13}C spin systems, refocusing the ^{1}H chemical shifts through a spin echo and inverting the ^{13}C spin populations, effectively changing the identity and precession direction of the resulting magnetization vectors. After a second $1/4J_{CH}$ delay, the spin-labeled magnetization vectors are 180° out of phase with each other. A final ^{1}H 90° pulse returns the magnetization to the z axis. However, the initial energy-level populations have been changed so that level 2 in Figure 8 has been switched with that of level 4. The population differences for the allowed ^{13}C transitions, $1 \rightarrow 2$ and $3 \rightarrow 4$ are greater and ^{13}C signals are enhanced by the ratio defined

in Equation (20).

$$\eta_{(\mathrm{INEPT})} = \frac{\gamma_{\mathrm{H}}}{\gamma_{\mathrm{C}}} \tag{20}$$

This ratio is 3.98 for ^{13}C and ^{1}H nuclei. A refocusing period and heteronuclear decoupling of ^{1}H spins produces a sensitivity-enhanced ^{13}C spectrum in which methylenes are phase inverted relative to methyne and methyl signals, and quaternary signals are suppressed. Delays between successive scans can also be made shorter as ^{1}H relaxation now limits the pulse repetition rate.

Owing to the sensitivity of the refocused INEPT to the delay periods around the second refocusing pulses, the DEPT sequence is more commonly used in sensitivity enhancement and multiplicity determinations. Similar in nature to INEPT, this sequence is less sensitive to refocusing delays and can be used to generate subspectra of differing multiplicities. The second ^{1}H π pulse in the INEPT sequence is replaced with a pulse of varying flip angle Θ. Three spectra are acquired with Θ set to 45°, 90° and 135°. The spectrum with $\Theta = 90°$ has only methyne carbon signals, the $\Theta = 135°$ spectrum will have positive methyl and methyne peaks and negative phase methylene peaks. The three spectra are mathematically added and subtracted to generate the following subspectra:

DEPT90	CH subspectrum.
DEPT45 − DEPT135	CH_2 subspectrum.
DEPT45 + DEPT135 − (0.707 ∗ DEPT90)	CH_3 subspectrum.

A common practice is to acquire only the DEPT spectrum with $\Theta = 135°$, since it contains the same multiplicity information as the attached proton test (APT) experiment with higher sensitivity. Comparison of the DEPT135 with a normal ^{13}C spectrum will permit assignment of quaternary carbons, as they will be suppressed in all DEPT spectra. The DEPT135 spectrum of abietic acid is shown in Figure 3(d). Although the methyne and methyl signals are of the same phase, the intensity and chemical shift information is often sufficient to distinguish methyls (more intense, more upfield) from methynes.

2.7 Nuclear Relaxation in Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Magnetization is observed in the NMR experiment through application of an rf pulse. This pulse forces the net magnetization vector, M_z, to precess about this applied field, toward the observable transverse plane. At the end of this pulse, the magnetization vectors begin the relaxation process through the exchange of excess energy in the spin system with other spins and with the local surroundings to return to equilibrium. The process of energy exchange with other spins is called spin–spin relaxation, T_2. Loss of energy to the surrounding structural environment is called spin-lattice relaxation, T_1. T_2 relaxation causes a dephasing of spin components of M_z in the x,y plane. In contrast, T_1 relaxation is the mechanism through which the spin system returns to equilibrium.

Relaxation processes occur as a result of the interaction of the magnetization vectors with oscillating local fields in the molecule created by molecular motion of the proper frequency. As the rates of molecular motion, lattice vibrations, molecular tumbling, and reorientation, change, both T_1 and T_2 will change. For small molecules, T_1 can be long, seconds to minutes in length, while polymers typically exhibit relaxation times in the 0.1 to 1 second that grow longer, after passing through a minimum, with decreased molecular motion as temperature is lowered. Small molecule T_2s are typically close to their T_1 values in a primarily dipolar relaxation spin system, and decrease to a minimum value as molecular size increases and molecular motion slows. The rate of molecular motion is important, represented by the molecular correlation time, τ_C, as are the number of directly attached protons. In all systems, $T_2 \leq T_1$, since the magnetization cannot be at equilibrium if there is still a transverse component. The importance of relaxation in spectroscopic analysis is the fact that it limits the ability to acquire resolved NMR spectra in as short a time as possible. T_1 relaxation creates a lower limit on the time delay between successive scans in the pulse NMR experiment to obtain quantitative results. Delay times of 5 T_1s are necessary for spin systems to return to full equilibrium, substantially increasing signal acquisition times as T_1 values lengthen. T_2 limits the resolution in a spectrum by causing a line broadening as T_2 decreases.

Relaxation time experiments in solution have traditionally focused on the measurement of T_1 values, as these values are diagnostic of structure and dynamics. Since relaxation times largely reflect molecular motion, studies are usually undertaken to study the type and rate of motions and factors that affect these parameters, such as associative phenomena. The rate and strength of intermolecular associative complexes have an effect on isotropic motion and can be examined through T_1 measurements. Relaxation studies can be used to distinguish different modes of motion or segmental mobility in a molecule, reflected in different T_1 values for carbon atoms in the same molecule. For small rigid molecules, T_1 values can be used to distinguish the number of protons attached to a carbon, since for rigid molecules, correlation times determined at all carbon positions are the same, and differences in T_1 reflect only differences in the number of directly attached protons. Determination of

T_1 values is also necessary to establish quantitative conditions for analytical measurements. The longest T_1 value in a spectrum is used to set the delay period between scans in quantitative experiments, five times T_1, to allow every part of the spin system to return to equilibrium. T_1 measurements can be used to determine τ_C, an important parameter for conformational analysis using molecular modeling. The dynamics of a molecular system must be known to use NOE experiments effectively for distance determinations. T_2 relaxation times are of value when designing new NMR experiments, since spin dephasing limits the length of time a pulse sequence can be carried out before observable magnetization is gone and cannot be refocused.

2.7.1 Spin-lattice (T_1) Relaxation

Spin-lattice relaxation is most efficient at motional frequencies close to the Larmor frequency of the nucleus being observed and is thus a field-dependent parameter. Contributions to T_1 result from a series of mechanisms, as shown in Equation (21):[21,22,64]

$$\frac{1}{T_1} = \frac{1}{T_{1(\mathrm{DD})}} + \frac{1}{T_{1(\mathrm{SR})}} + \frac{1}{T_{1(\mathrm{SC})}} + \frac{1}{T_{1(\mathrm{CSA})}} + \frac{1}{T_{1(\mathrm{ES})}} \quad (21)$$

where $T_{1(\mathrm{DD})}$ is the component due to heteronuclear dipole–dipole interactions, $T_{1(\mathrm{SR})}$ is from spin rotation, $T_{1(\mathrm{SC})}$ from scalar couplings, $T_{1(\mathrm{CSA})}$ from chemical shift anisotropy (CSA), and $T_{1(\mathrm{ES})}$ from unpaired-electron-spin interactions. In the absence of unpaired electrons in the molecule or paramagnetic impurities in the sample, the predominant relaxation mechanism in ^{13}C NMR of organic molecules is through heteronuclear dipole–dipole interactions between carbon atoms and directly bonded protons or those in close spatial proximity. Molecular tumbling and C–H bond rotation reorient the spin dipoles, creating local magnetic fields that constantly fluctuate, providing a pathway for relaxation. Oscillating local fields induce nuclear spin transitions if there are molecular motions with frequency components near ν_0 of the observed nucleus. $T_{1(\mathrm{DD})}$ for a molecule in dilute solution with isotropic motion in the extreme narrowing limit $(w_H + w_C)\tau_C \ll 1$ can be simplified to Equation (22):[22]

$$\frac{1}{T_{1(\mathrm{DD})}} = N\left(\frac{h}{2\pi}\right)^2 \gamma_C^2 \gamma_H^2 (r_{CH}^{-6})\tau_C \quad (22)$$

where N is the number of directly bonded ^{1}H nuclei, r_{CH}^{-6} is the internuclear distance and τ_C is the molecular correlation time. While this equation does not show the field dependence of T_1, it is reduced from an equation containing spectral density functions which contain w_C and w_H terms[22] that approach unity with rapid isotropic motion. Using this equation, a methylene group with $\tau_C = 10^{-11}$ s, would have a $T_{1(\mathrm{DD})}$ of ~1.5 s neglecting neighbor and intermolecular effects. This component of spin-lattice relaxation is the basis of NOE measurements discussed in section 2.5 and the experimentally determined NOE factor can be used to determine the contribution of dipole–dipole interactions to the overall T_1. A full NOE indicates that dipole relaxation, and thus $T_{1(\mathrm{DD})}$, is the only contributor to observed relaxation behavior. Incomplete NOE reveals contributions from other relaxation mechanisms.

Molecules or portions of a molecule that have an angular velocity produce local magnetic fields due to the reorientation of the electrons contained in chemical bonds. This spin rotation contributes to relaxation as $T_{1(\mathrm{SR})}$ and is described in Equation (23):[22,23,71,160]

$$\frac{1}{T_{1(\mathrm{SR})}} = \left\{\frac{2kT}{3(h/2\pi)^2} I_m (C_\perp^2 + C_\parallel^2)\right\} \tau_{SR} \quad (23)$$

where I_m is the moment of inertia relative to axis m, $C_\perp$ and $C_\parallel$ are spin rotation constants perpendicular and parallel to the symmetry axis of the molecule or functional group, and τ_{SR} is the correlation time for spin rotation or angular momentum. In typical systems, perturbation of angular velocity occurs at rates making this mechanism inefficient in liquids. Contributions from spin rotation can be significant for unprotonated carbons with modest dipole relaxation components, or pendant alkyl groups with motional velocities on the order of ^{13}C Larmor frequencies. Since angular velocity increases with temperature, contributions to relaxation will also increase and an examination of the temperature dependence of T_1 will confirm the presence of $T_{1(\mathrm{SR})}$ mechanisms.

Coupling of ^{13}C nuclei to other atoms provides a potential mechanism for relaxation if that nucleus has a nonzero spin quantum number. If the bonded nucleus relaxes rapidly, a condition found with many common quadrupolar nuclei, oscillating local fields are created which provide a pathway for ^{13}C relaxation. The scalar relaxation contribution is defined in Equation (24):[22,69,161]

$$\frac{1}{T_{1(\mathrm{SC})}} = \frac{8\pi^2 J^2}{3} S(S+1)\, \frac{\tau_{SC}}{1 + (w_S - w_C)^2 \tau_{SC}^2} \quad (24)$$

where J is the scalar coupling between the carbon atom and nucleus S, S is the spin quantum number of nucleus S, τ_{SC} is the relaxation time of nucleus S, and w_S is the Larmor frequency of nucleus S. In typical ^{13}C spectra, this mechanism does not significantly contribute to T_1 relaxation, but can be the dominant factor in T_2. This mechanism is field-dependent, since its mathematical description contains the term $(w_S - w_C)$. As these terms are also in the denominator of Equation (24), the second term is small. Only in the event of a very large coupling

do scalar coupling mechanisms contribute to spin-lattice relaxation.

As discussed in section 2.4, the chemical shift of a carbon atom is described by a tensor which has some orientation within the molecular framework. The ^{13}C chemical shielding is thus dependent on the orientation of the molecule to H_0, as well as the size of H_0, and becomes a significant contributor to relaxation at larger magnetic fields. Rapid isotropic motion averages the chemical shift to an isotropic value, but the anisotropic components create an oscillating field due to the rapid reorientation of the chemical shift tensor with respect to H_0. This relaxation rate is described in Equation (25):[22,71,87]

$$\frac{1}{T_{1(\mathrm{CSA})}} = \frac{2}{15}\gamma_C^2 H_0^2(\sigma_\parallel - \sigma_\perp)^2\tau_C \qquad (25)$$

where $\sigma_\parallel$ and $\sigma_\perp$ are the components of the CSA which are oriented parallel and perpendicular to H_0. Contributions to relaxation from the CSA will depend on the size and shape of the anisotropy, as well as τ_C.

Unpaired electrons are an effective contributor to T_1 and T_2 relaxation. Electrons can interact with ^{13}C nuclei either through dipole–dipole couplings or through a scalar electron–nucleus coupling. These contributors to relaxation rates are defined in Equations (26) and (27):[22]

$$\frac{1}{T_{1(\mathrm{ES})}} = \frac{2S(S+1)\gamma_S^2\gamma_C^2}{15r_{CS}^6}\left(\frac{3\tau_{C1}}{1+w_e^2\tau_{C1}^2} + \frac{7\tau_{C2}}{1+w_e^2\tau_{C2}^2}\right) \qquad (26)$$

$$\frac{1}{T_{1(\mathrm{ES})}} = \frac{8\pi^2A^2}{3}S(S+1)\frac{\tau_{e2}}{1+(w_C-w_e)^2\tau_{e2}^2} \qquad (27)$$

where γ_S is the magnetogyric ratio for an electron, w_e is the resonance frequency of the electron, A is the hyperfine spin nuclear coupling constant, and τ_{C1} and τ_{C2} are correlation times consisting of contributions from τ_C, τ_{e1} and τ_{e2}, the electron relaxation times and τ_m, the lifetime of the complex contributing the unpaired electron.[22] The dipole mechanism in Equation (26) only requires some interaction between the unpaired electron and the molecule being observed, while the scalar mechanism is only active when the unpaired spin can be delocalized through bonding. Since electrons have very large magnetic moments, couplings between them and surrounding nuclei are greater than nuclear dipole couplings and the presence of unpaired electrons can dominate both spin-lattice and spin–spin relaxation rates. An advantage to the presence of unpaired spins is the shorter T_1 relaxation times found for ^{13}C spins, permitting shorter delay times between successive scans. Paramagnetic relaxation agents added to shorten T_1 s can also quench NOEs, making data more quantitative at the cost of sensitivity enhancement. Unpaired spins shorten T_2, effectively broadening spectral resonances.

T_1 values are measured using an inversion-recovery sequence.[162,163] In this experiment, shown in Figure 9(b), the net magnetization is inverted with a nonselective π pulse followed by a delay, τ, to allow relaxation to occur. This is followed by a $\pi/2$ pulse and acquisition of the ^{13}C NMR spectrum. A series of spectra are collected while changing the delay between pulses to cover a range from 0.1 T_1 to 5 T_1. A spectrum with a very short t will have all lines inverted in phase relative to one with a long t value. The integrated peak areas or intensities are plotted as a function of t to produce a plot like that shown in Figure 10. This curve is fitted to an exponential function of the type shown in Equation (28):

$$I = I_0(1 - Ae^{(-\tau/T_1)}) \qquad (28)$$

where I is the measured peak area or intensity, I_0 is the equilibrium magnetization value, A is a constant and τ is the delay between pulses. The data are fitted to this equation iteratively to obtain T_1. Other methods for measurement of T_1 include saturation-recovery,[164] and progressive-saturation.[165,166] These methods examine the recovery of magnetization, but through either saturation of the spins with a delay before observation or attenuation of the observed magnetization using a multiple-pulse sequence. The inversion-recovery method is the most commonly used sequence, but the progressive-saturation method is advantageous for measuring long T_1 values.

2.7.2 Spin–Spin (T_2) Relaxation

Spin–spin relaxation is a function of molecular motion and the inhomogeneity of the external magnetic field. In

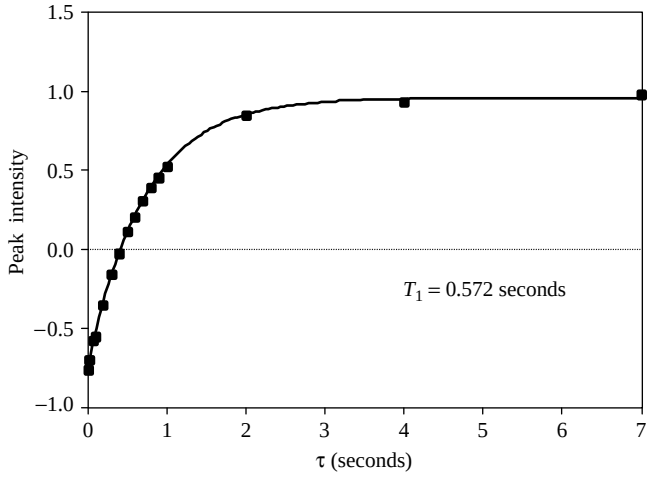

Figure 10 Plot of ^{13}C peak intensities as a function of t acquired using the pulse sequence shown in Figure 9 and fit to Equation (28) to determine T_1 using the inversion recovery method.

the latter case, molecules moving through the magnetic field experience slightly different H_0 fields and thus precess with slightly different ν_0 values. This is minimized, but not eliminated, by adjusting the magnetic field to make it homogeneous across the sample. As a result, spin coherence is lost and the magnetization dephases in the transverse plane. Unlike T_1, T_2 is a result of energy exchange between spins rather than a loss of energy to the molecular lattice. In the motional narrowing limit, $(w_H + w_C)\tau_C \ll 1$, $T_2 = T_1$ as in Equation (22). T_2 relaxation is also sensitive to the same mechanisms as discussed in the previous section, particularly when scalar couplings to quadrupolar nuclei or couplings to unpaired electrons are present.

Spectral line widths are determined by T_2 values and spectrometer magnetic-field inhomogeneities shown by Equations (29) and (30):[21,64]

$$\Delta\nu_{1/2} = \frac{1}{\pi T_2^*} \tag{29}$$

$$\frac{1}{T_2^*} = \frac{\gamma \Delta B_0}{2} + \frac{1}{T_2} \tag{30}$$

where $\Delta\nu_{1/2}$ is the full width at half height of the spectral resonance. T_2^* is the sum of magnetic-field inhomogeneity effects, $\gamma\Delta B_0$, and T_2 due to spin–spin interactions. For typical NMR spectrometers with small molecules in the extreme-narrowing region of molecular motion, the resonance line width is dominated by field inhomogeneities. In this case, spin-echo experiments are used to determine intrinsic T_2 values. Larger molecules and polymers, having slower molecular motions that shorten T_2, exhibit broadened lines resulting in $\Delta\nu_{1/2} \sim 1/\pi T_2$. In molecules with quadrupolar nuclei or unpaired spins, T_2 decreases and broadens spectral resonances.

Measurement of T_2 is carried out with the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence.[167,168] This experiment begins with a 90° pulse followed by a series of 180° refocusing pulses, spaced by an appropriate interval. During the delay, both chemical-shift and spin–spin interactions occur. The 180° pulse refocuses the chemical-shift process and, at an appropriate time, the magnetization has been refocused. Acquisition of an FID at this point will contain a normal spectrum, but the peaks will be attenuated due to T_2 relaxation effects because the dephasing due to T_2 is not refocused in this experiment. A series of spectra are collected, each with an increasing number of echo pulses and delays. Peak intensities will exponentially decay and can be fitted to Equation (31):[21]

$$I_t = I_0 e^{(-t/T_2)} \tag{31}$$

with the same connotations as mentioned for T_1 analysis.

2.8 Multidimensional Carbon-13 Nuclear Magnetic Resonance Experiments

With the development of sophisticated computers and electronics capable of acquiring and processing large amounts of spectral data, multidimensional NMR methods have become the mainstay of the chemist wishing to identify an unknown material, determine physical parameters such as scalar couplings or study molecular conformations and associative phenomena. First proposed in 1971,[169] the seminal work of Ernst et al.[170,171] made 2-D NMR a reality and heralded a new era in NMR spectroscopy. Multidimensional NMR methods contain information on spin systems, i.e. which nuclei are scalar coupled to each other through scalar or dipolar interactions or through chemical-exchange pathways. Since the initial development of these methods, 2-D and higher-order NMR experiments have become pervasive tools in studying the molecular structure of organic materials, especially in the area of proteins, where information about the 3-D structure can be obtained through NOE and scalar coupling experiments. As in all NMR methods though, direct observation of ^{13}C signals reduces the sensitivity of any experiment and most of the recent developments are for inverse or indirect detection experiments where the spectrum is acquired through the ^1H signal. Rather than observe the effect of C–H couplings on the carbon signals, it has become much easier and quicker to observe the same effect on the proton signals. Consequently, there are only a few 2-D experiments for which the ^{13}C signal is measured and these are discussed below. Several excellent reference volumes provide a detailed description of multidimensional NMR and its implementation for structural determination and the study of dynamic phenomena.[60,172,173]

2.8.1 Description of Multidimensional Nuclear Magnetic Resonance Spectroscopy

A standard chemical-shift experiment comprises two basic components, preparation and detection of the NMR signal. The preparation step in ^{13}C-NMR generates ^{13}C observable magnetization with an rf pulse while saturating the ^1H spins with heteronuclear decoupling. Detection involves turning on the receiver and observing the voltage induced in the sample probe coil. More sophisticated techniques, such as INEPT, may have a mixing period in which spins exchange information through scalar or dipolar couplings. The detection period where the NMR signal is observed as a function of time is called the acquisition period, or T_2. Multidimensional techniques have a second time period, T_1, which is known as the evolution time. This second evolution period is for spin systems to evolve under the influence of scalar couplings, dipolar oscillations creating cross-relaxation pathways,

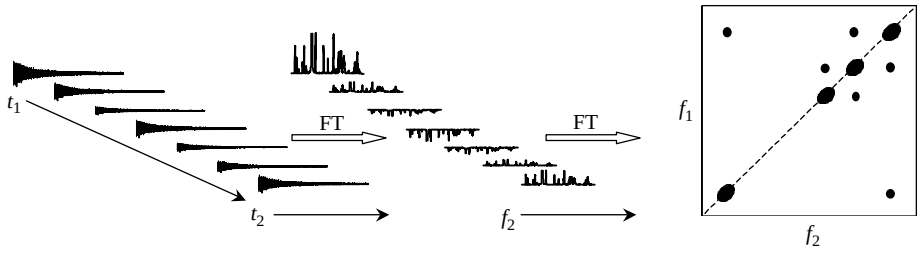

Figure 11 Schematic diagram of the 2-D NMR experiment. A series of NMR spectra (typically 256–512) are acquired, spaced by a constantly increasing delay in the pulse sequence. These spectra, collected in an array, are then Fourier transformed once in each dimension to yield a new array with frequency units in both dimensions.

chemical dynamics or exchange as a function of the T_1 time increment. A 2-D NMR spectrum is created by acquiring a series of 1-D FIDs (64 to 512) spaced by a constant increment of the evolution time, T_1. The result is an array of FIDs, each in the time variable T_2, spaced uniformly over the total T_1 period. A double FT, once along each axis of the array, produces a 2-D spectrum with dimensions in frequency units f_2 and f_1. This process is shown schematically in Figure 11. For 2-D spectra where ^{13}C is the detected nucleus, f_2 is the ^{13}C chemical-shift dimension and f_1 is either the chemical shift of some other nucleus, the heteronuclear coupling dimension or the double-quantum-frequency dimension. A pulse program for a $^{13}C-^{1}H$ HETCOR experiment is shown in Figure 9(d). The basic unit of many 2-D and higher-order NMR experiments is the polarization transfer sequence discussed earlier and shown in Figure 9(c). The appropriate selection of pulse sequences establishes the type of pathway or spin evolution that is to occur in T_1.

2.8.2 Two-dimensional Nuclear Magnetic Resonance

There are three primary 2-D NMR experiments for the study of ^{13}C nuclei, not including the inverse, ^{1}H detected analogs of each. $^{13}C-^{1}H$ scalar couplings and carbon multiplicities can be observed through the 2-D J-resolved technique, which has ^{13}C chemical shifts along one axis and $^{13}C-^{1}H$ $^{1}J_{CH}$ coupling multiplets in the second dimension. HETCOR experiments establish connectivities between scalar coupled carbon and hydrogen atoms in a molecule. Carbon–carbon connectivities can be established through 2-D INADEQUATE spectroscopy, a difficult but rewarding technique for study of the carbon skeleton in a molecule.

Determination of $^{1}J_{CH}$ scalar couplings is valuable in structural analysis for the assignment of multiplicities of carbon atoms, the relative s character in carbon hybridization and the electron-withdrawing nature of substituents. The J-resolved 2-D spectrum separates ^{13}C chemical shifts from $^{1}J_{CH}$ scalar couplings, while suppressing ^{1}H chemical shifts and homonuclear couplings.[171,174] This simple experiment uses a spin echo to refocus ^{13}C chemical-shift information, while allowing $^{13}C-^{1}H$ scalar couplings to modulate the echo in the evolution period. The result is a 2-D spectrum with ^{13}C chemical shifts on the f_2 axis and C–H coupling information in the f_1 dimension. From studying the f_1 slice at each ^{13}C resonance, the multiplicity and $^{1}J_{CH}$ can be determined for each signal. Analogous to acquiring a 1-D ^{13}C spectrum with the decoupler off during acquisition, this method has the advantage of removing the overlap found in 1-D spectra, such as the spectrum of abietic acid shown in Figure 3(a).

Assignment of spectral resonances is a crucial step in the process of structure determination. If assignments of ^{13}C or ^{1}H signals are known, then HETCOR is used to establish the assignment of the coupled nucleus.[175,176] The HETCOR experiment transfers information generated from one-bond scalar $^{1}J_{CH}$ couplings with a polarization-transfer sequence to observe ^{13}C resonances. The time evolution of the $^{1}J_{CH}$ coupling introduces a phase modulation of the detected ^{13}C signals and, when Fourier transformed, yields a ^{1}H dimension in the 2-D spectrum. The $^{13}C-^{1}H$ HETCOR spectrum of abietic acid is shown in Figure 12. The horizontal axis is the ^{13}C chemical-shift dimension and the vertical is the ^{1}H dimension. Bonding partners are seen as peaks where the ^{13}C and ^{1}H frequencies intersect. The solid lines in the spectrum trace several of the correlations, for example, the methyl carbon resonance at 14 ppm is seen to correlate with the proton signal at 0.8 ppm.

Structural analysis requires the unambiguous establishment of carbon–carbon connectivities, identification of functional-group arrangements and the presence of heteronuclei in the carbon skeleton. A difficult but rewarding technique for this purpose is the INADEQUATE.[177,178] In this experiment, $^{13}C-^{13}C$ connectivities are determined by allowing spins to precess, while evolving only under the influence of $^{1}J_{CC}$ couplings. The resulting spectrum, an example of which is shown in Figure 13, contains ^{13}C chemical shifts along one axis and the double-quantum frequency (the sum of the chemical shifts of the coupled spins) axis along the second dimension. If two

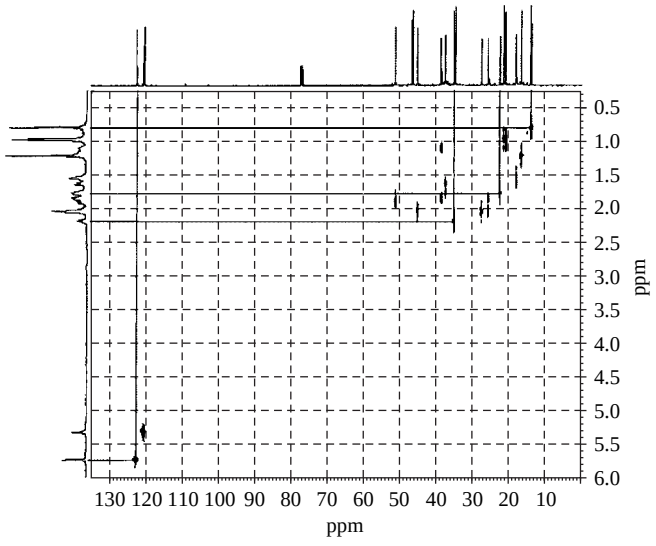

Figure 12 $^{13}C-^{1}H$ HETCOR 2-D spectrum of abietic acid.

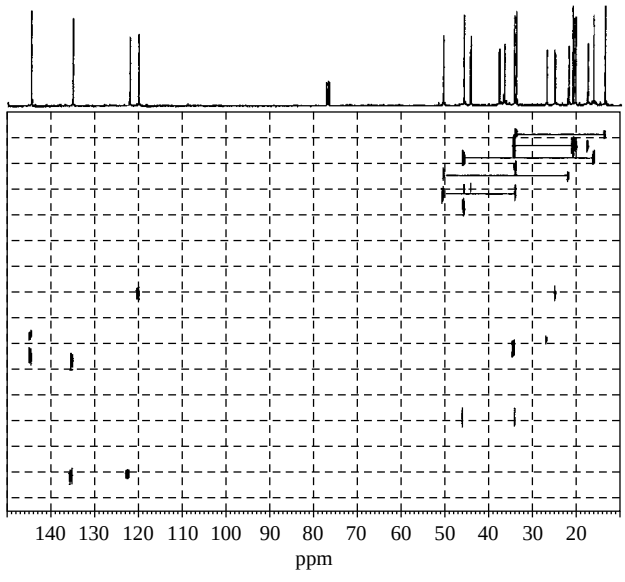

Figure 13 $^{13}C-^{13}C$ INADEQUATE 2-D correlation spectrum of abietic acid.

spins are coupled, their resonances, at ν_{Ca} and ν_{Cb}, each have a correlation peak in the f_1 dimension at $(\nu_{Ca} + \nu_{Cb})$, shown by solid lines in the spectrum of Figure 13 for several spin pairs. Each correlation peak is split into doublets spaced by $^{1}J_{CC}$. The carbon skeleton can be followed by "walking" through the connectivity pathways shown in the spectrum. Any breaks in the correlation path indicate either insertion of a heteroatom in the skeleton or the termination of a structure. Since this experiment suppresses the central ^{13}C resonance while the $^{13}C-^{13}C$ satellites are observed, signals which are only 0.55% of the central ^{13}C signal, it is a difficult experiment that requires a high-field NMR spectrometer and maximum sample concentrations. This technique is difficult and is often replaced by a combination of ^{1}H inverse experiments.

Development of ^{1}H multidimensional NMR methods has surpassed that of ^{13}C-NMR for many reasons. Sensitivity of the ^{1}H experiment is the first advantage, coupled with shorter relaxation times. Spectra of sub-microgram samples can be obtained in minutes and, with the application of field gradients, 2-D experiments can be acquired in under 30 minutes. NMR probes in the inverse configuration are used to acquire HETCOR spectra with ^{1}H rather than ^{13}C detection, with significant time savings.[179–182] In many studies of molecular structure, only a 1-D ^{13}C spectrum is acquired, with all other information derived through inverse experiments. Even the 2-D INADEQUATE for determining $^{13}C-^{13}C$ connectivities has a ^{1}H detected version, 2-D INEPT–INADEQUATE[183] that takes advantage of heteronuclear polarization transfer to transfer carbon skeleton information to attached protons. Of course, the drawback to these experiments is that quaternary carbons may be overlooked since polarization-transfer methods do not detect them.

3 CURRENT PERSPECTIVE AND FUTURE DEVELOPMENT

^{13}C-NMR spectroscopy is and will continue to be one of the most valuable tools for the analysis of organic molecules. It is pervasive throughout the chemical literature in most areas of chemistry and biochemistry and is routinely used for the structural analysis of organic molecules, the study of polymer structure and dynamics, analysis of molecular conformations in biopolymers, elucidation of reaction mechanisms or biosynthetic pathways. Applications have grown substantially in the last two decades and will continue to do so with the further refinement of NMR instrumentation and the development of polarization-transfer techniques for improved sensitivity and spectral-editing capabilities. NMR spectroscopy is a routine analytical tool for the student as well as the professional scientist and access to instrumentation is widespread.

The simulation and prediction of NMR spectra will continue to develop. There is still much to be understood about the chemical shift and how best to calculate meaningful tensor values. An understanding of the chemical shift and its orientation in the molecular framework are important parameters in both structural and conformational analysis. Along with the fundamental principles defining the chemical shift, significant developments in the area of molecular-modeling software have simplified, and will continue to simplify, the task of solving

Schroedinger's equation for molecular energies. Dramatic improvements in the speed of computer hardware for doing such calculations puts the challenge of doing ab initio calculations within the grasp of those with access to only desktop PC systems.

While suffering from low inherent sensitivity compared to ^{1}H, the spectral dispersion of ^{13}C-NMR provides significantly greater resolution of functional groups than does ^{1}H NMR. To circumvent low sensitivity, larger magnetic-field strengths are being made available at a rapid rate. Magnets with field strengths of 18 tesla are commercially available with ^{13}C resonance frequencies at 200 MHz. This increase in field strength continually enhances the utility of ^{13}C-NMR measurements, since the net observable magnetization scales with H_0. At these higher fields, it may be possible to learn more about the CSA of ^{13}C nuclei and the orientation of this anisotropy to the molecular framework. Larger magnetic fields can produce preferential orientation of molecules, reintroducing anisotropic motion and line broadening. Improved resolution may also permit the observation of conformational and sequential effects in polymers over a larger distance than currently available. Refinement of statistical models for reactivity and catalyst activity will result.

However, one cannot ignore the shortcomings of ^{13}C-NMR as a method for quantitative analysis. Even with great care in selection of parameters,[57,184–186] quenching of NOE and with an understanding of relaxation in the system of interest, one may achieve a relative precision rate of one percent, and worse at lower component levels. Under highly quantitative conditions, analysis times will be very long. In cases of very low component levels, the technique may not have adequate sensitivity to observe trace materials. While ^{13}C-NMR may never be the most precise analytical technique, it finds substantial use as a qualitative technique and will most likely remain a powerful identifier of functional groups in a sample, serving as a reference method for other analytical techniques. The cost of instrumentation, up to several million dollars for state-of-the-art 800 MHz systems, makes it prohibitive to place an NMR in plant operations. Using NMR to calibrate other methods, such as infrared and Raman spectroscopy, does however allow NMR to be of use in commercial operations.

ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Networks
APT	Attached Proton Test
CPD	Composite Pulse Decoupling
CPMG	Carr–Purcell–Meiboom–Gill
CSA	Chemical Shift Anisotropy
CW	Continuous Wave
DEPT	Distortionless Enhancement via Polarization Transfer
FFT	Fast Fourier Transform
FID	Free Induction Decay
FT	Fourier Transform
GIAO	Gauge Independent Atomic Orbital Methods
HETCOR	Heteronuclear Correlation
IGLO	Individual Gauge for Localized Orbitals
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
INEPT	Insensitive Nucleus Enhancement via Polarization Transfer
LCAO	Linear Combinations of Atomic Orbital
LORG	Localized Orbitals, Localized Origin
MLR	Multilinear Regression
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
rf	Radiofrequency
SCT	Self-consistent Field
S/N	Signal-to-noise Ratio
SW	Spectral Width
TMS	Tetramethylsilane
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Food **(Volume 5)**
Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Steel and Related Materials **(Volume 10)**
Nuclear Magnetic Resonance in Metals Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. W. Pauli, 'Recent Investigations of the Structure of Colloids', *Naturwissenschaften*, **12**, 741 (1924).
2. I.I. Rabi, S. Millman, P. Kusch, P. Zacharias, 'The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments', *Phys. Rev.*, **55**, 526 (1939).
3. F. Bloch, W.W. Hansen, M. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
4. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37 (1946).
5. J.T. Arnold, S.S. Dharmatti, M.E. Packard, 'Chemical Effects on Nuclear Induction Signals from Organic Compounds', *J. Chem. Phys.*, **19**, 507 (1951).
6. P.C. Lauterbur, 'C^{13} Nuclear Resonance Spectra', *J. Chem. Phys.*, **26**, 217 (1957).
7. R.R. Ernst, 'Nuclear Magnetic Double Resonance with an Incoherent Radio-frequency Field', *J. Chem. Phys.*, **45**, 3845 (1966).
8. F. Bloch, 'Nuclear Induction', *Phys. Rev.*, **70**, 460 (1946).
9. E.L. Hahn, 'Spin Echoes', *Phys. Rev.*, **80**, 580 (1950).
10. I.J. Lowe, R.E. Norberg, 'Free Induction Decay in Solids', *Phys. Rev.*, **107**, 46 (1957).
11. R.R. Ernst, W.A. Anderson, 'Application of Fourier Transform Spectroscopy to Magnetic Resonance', *Rev. Sci. Instr.*, **37**, 93 (1966).
12. R.R. Ernst, 'Sensitivity Enhancement in Magnetic Resonance', *Adv. Magn. Res.*, **2**, 1 (1966).
13. G.A. Morris, R. Freeman, 'Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer', *J. Am. Chem. Soc.*, **101**, 760–762 (1979).
14. D.P. Burum, R.R. Ernst, 'Net Polarization Transfer via a J-Ordered State for Signal Enhancement of Low-sensitivity Nuclei', *J. Magn. Reson.*, **39**, 163–168 (1980).
15. M.R. Bendall, D.M. Doddrell, D.T. Pegg, 'Editing of Carbon-13 NMR Spectra: 1. A Pulse Sequence for the Generation of Subspectra', *J. Am. Chem. Soc.*, **103**, 4603–4605 (1981).
16. D.M. Doddrell, D.T. Pegg, M.R. Bendall, 'Distortionless Enhancement of NMR Signals by Polarization Transfer', *J. Magn. Reson.*, **48**, 323–327 (1982).
17. A.E. Derome, *Modern NMR Techniques for Chemistry Research*, Pergamon, Oxford, 1991.
18. G. Bodenhausen, R. Freeman, R. Niedermeyer, D.L. Turner, 'Double Fourier Transformation in High-resolution NMR', *J. Magn. Reson.*, **26**, 133 (1977).
19. V.V. Krishnamurthy, J.G. Shi, G.A. Ollah, 'The Structure of Tetrahydrobinor-*S*(pentacyclo[8.4.0.02, 6.03, 8.09, 13]tetradecane) Based on Carbon–Carbon Connectivity', *J. Org. Chem.*, **50**, 3005 (1985).
20. A. Bax, R. Freeman, T.A. Frenkiel, 'An NMR Technique for Tracing Out the Carbon Skeleton of an Organic Molecule', *J. Am. Chem. Soc.*, **103**, 2102 (1981).
21. E. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy*, 3rd edition, VCH, Weinheim, 1989.
22. F.W. Wehrli, A.P. Marchand, S. Wehrli, *Interpretation of Carbon-13 NMR Spectra*, 2nd edition, John Wiley & Sons, New York, 1988.
23. R. Schrotooten, F. Borremans, M. Anteunis, 'Determination of the Conformational Preference of the Halogens in 4-halooxacylclohexanes', *Spectrochim. Acta*, **34B**, 297 (1978).
24. G. Binsch, H. Kessler, 'New Analytical Methods. 18. Kinetic and Mechanistic Evaluation of NMR Spectra', *Angew. Chem.*, **92**, 445 (1980).
25. R.A. Heintz, T.G. Neiss, K.H. Theopold, 'Unusually Large Isotope Effects in NMR Chemical Shifts of Paramagnetic Organometallic Compounds', *Angew. Chem.*, **106**, 2389 (1994).
26. P.J.P. Nathan, R.L. Santillan, P. Schmitt, H. Gunther, 'The Study of Longifolene by Two-dimensional NMR Spectroscopy', *Org. Magn. Reson.*, **22**, 450 (1984).
27. W. Offermann, 'Carbon-13 NMR of Camphene and α-Pimene: Revised Assignments', *Org. Magn. Reson.*, **20**, 203 (1982).
28. E. Wenkert, B.L. Buckwalter, 'Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Naturally Occurring

Substances. X. Pimaridenes', *J. Am. Chem. Soc.*, **94**, 4367 (1972).
29. J.F. Templeton, S. Majid, A. Marr, K. Marat, 'Zinc-Acetic Acid Reduction of the Steroid 4-En-3-one: Novel Conversion of the 4-En-3-one via a Vinyl Chloride', *J. Chem. Soc. Perkin Trans.*, **1**, 2581 (1990).
30. T. Yasuzawa, M. Yoshida, H. Sano, 'Structure of a Paxisterol, a Sterol from *Penicillium paxilli*', *J. Chem. Soc. Perkin Trans.*, **2**, 3145 (1990).
31. L. Poppe, H. van Halbeek, 'Proton-detected Measurements of Long-range Heteronuclear Coupling Constants: Application to a Trisaccharide', *J. Magn. Reson.*, **92**, 636 (1991).
32. J. Dabrowski, 'Application of Two-dimensional NMR Methods in the Structural Analysis of Oligosaccharides and Other Complex Carbohydrates', *Meth. Stereochem. Anal.*, **9**, 349 (1987).
33. H. Grasdalen, A.K. Andersen, B. Larsen, 'NMR Spectroscopy Studies of the Action Pattern of Tomato Pectinesterase: Generation of Block Structure in Pectin by a Multiple-attack Mechanism', *Carbohydr. Res.*, **289**, 105 (1996).
34. A.I. Usov, S.V. Yarotsky, N.D. Shashkov, 'Carbon-13 NMR Spectroscopy of Red Algal Galactans', *Biopolymers*, **19**, 977 (1980).
35. E.P. Nikonowicz, A. Pardi, 'Application of Four-dimensional Heteronuclear NMR to the Structure Determination of a Uniformly ^{13}C Labeled RNA', *J. Am. Chem. Soc.*, **114**, 1082 (1992).
36. I.W.J. Still, N. Plavac, D.M. McKinnon, M.S. Chauhan, 'Carbon-13 Nuclear Magnetic Resonance Spectra of *N*-, *O*-, and *S*-Methylated Uracil and Thiouracil Derivatives', *Can. J. Chem.*, **56**, 725 (1978).
37. M.J. Michnicka, J.W. Harper, G.C. King, 'Selective Isotopic Enrichment of Synthetic RNA: Application to the HIV-1 TAR Element', *Biochemistry*, **32**, 395 (1993).
38. M.H. Gelb, R.H. Abeles, 'Mechanism of Inactivation of Chymotrypsin by 3-Benzyl-6-chloro-2-pyrone', *Biochemistry*, **23**, 6596 (1984).
39. J.N.S. Evans, P.E. Fagerness, N.E. Mackenzie, A.I. Scott, 'Proton, Tritium and Carbon-13 NMR Studies on Porphobilinogen', *Magn. Reson. Chem.*, **23**, 939 (1985).
40. S. Spera, A. Bax, 'Empirical Correlation Between Protein Backbone Conformation and C_α and C_β ^{13}C Nuclear Magnetic Resonance Chemical Shifts', *J. Am. Chem. Soc.*, **113**, 5490 (1991).
41. J.N.S. Evans, *Biomolecular NMR Spectroscopy*, Oxford University Press, Oxford 1995, and references therein.
42. D.S. Wishart, B.D. Sykes, F.M. Richards, 'Relationship Between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure', *J. Mol. Biol.*, **222**, 311 (1991).
43. P.W. Jeffs, D. McWilliams, 'Carbon-13 Nuclear Magnetic Resonance Study of the Biosynthesis of Cycloheximide. Stereospecific Incorporation into the Glutarimide Ring', *J. Am. Chem. Soc.*, **103**, 6185 (1981).
44. N. Neuss, C.N. Nash, P.A. Lembke, J.B. Grutzner, 'Use of Carbon-13 Nuclear Magnetic Resonance (Cmr) Spectroscopy in Biosynthetic Studies. Incorporation of Carboxyl and Methyl Carbon-13 Labeled Acetates into Cephalosporin C', *J. Am. Cem. Soc.*, **93**, 2337 (1971).
45. W. Trowitzsch, K. Gerth, V. Wray, G. Hofle, 'On the Biosynthesis of Myxovirescin A1 by *Myxococcus virescens*', *J. Chem. Soc., Chem. Comm.*, **1174** (1983).
46. J.C. Randall, *Polymer Sequence Determination*, Academic Press, New York, 1977.
47. A.E. Tonekki, *NMR Spectroscopy and Polymer Microstructure: The Conformational Connection*, VCH, New York, 1989.
48. F.A. Bovey, *Chain Structure and Conformation of Macromolecules*, Academic Press, New York, 1982.
49. F. Laupretre, L. Monnerie, O. Vogl, '^{13}C NMR Spectroscopy of Head-to-Head Polystyrene: Molecular Motion in Solution', *Europ. Polym. J.*, **14**, 981 (1978).
50. A. Perico, A. Altomare, D. Catalino, M. Columbari, C.A. Veracini, 'Relaxation of Segment Orientation in Dilute Polymer Solution: Interpretation of Carbon-13 Spin-lattice Relaxation Time and Nuclear Overhauser Effect Experiments on Dilute Poly[(*R*,*S*)-3,7-dimethyl-1-octene] and poly[(*R*,*S*)-3-methyl-1-octene]', *Macromolecules*, **23**, 4912 (1990).
51. J.A. Ratto, P.T. Inglefield, R.A. Rutkowski, K.-L. Li, A.A. Jones, A.K. Roy, 'A Comparison of Spin Relaxation and Local Motion Between Symmetrically and Asymmetrically Ring-substituted Bisphenol Units in Dissolved Polycarbonates', *J. Polym. Sci., Polym. Phys. Ed.*, **25**, 119 (1987).
52. S. Gillet, J.-J. Delpeuch, P. Valentin, J.-C. Escalier, 'Optimum Conditions for Crude Oil and Petroleum Product Analysis by Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **52**, 813 (1980).
53. T.D. Alger, M. Solum, D.M. Grant, G.D. Silcox, R.J. Pugmire, 'Spin-lattice Relaxation Parameters in the Quantitative Determination of Condensed Aromatic Compounds by Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **53**, 2299 (1981).
54. T.J. Wenzel, M.E. Ashley, R.E. Sievers, 'Water-soluble Paramagnetic Relaxation Reagents for Carbon-13 Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **54**, 615 (1982).
55. E. Hofer, R. Keuper, H. Renken, 'Chiral Relaxation Reagents (CRR) in Carbon-13 NMR Spectroscopy', *Tetrahed. Lett.*, **25**, 1141 (1984).
56. A. Rapp, A. Markowetz, 'Use of NMR Spectroscopy in Wine Analysis', *Lebensmittelchem. Gerichtl. Chem.*, **43**, 73 (1989).
57. D.E. Cookson, B.E. Smith, 'Optimal Conditions for Obtaining Quantitative Carbon-13 NMR Data', *J. Magn. Reson.*, **57**, 355–368 (1984).
58. A. Abragam, *The Principles of Nuclear Magnetism*, Oxford University Press, New York, 1961.

59. C.P. Slichter, *Principles of Magnetic Resonance*, Harper, New York, 1963.
60. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, New York, 1987.
61. J.W. Cooley, J.W. Tukey, 'An Algorithm for the Machine Calculation of Complex Fourier Series', *Math. Comput.*, **19**, 297 (1965).
62. M.H. Levitt, R. Freeman, 'Composite Pulse Decoupling', *J. Magn. Reson.*, **43**, 502 (1981).
63. M.H. Levitt, R. Freeman, T. Frenkiel, 'Broadband Heteronuclear Decoupling', *J. Magn. Reson.*, **47**, 328 (1982).
64. H. Friebolin, *Basic One- and Two-dimensional NMR Spectroscopy*, 2nd edition, VCH, Weinheim, 1993.
65. A. Saika, C.P. Slichter, 'Calculation of Diamagnetic Shielding in Molecules', *J. Chem. Phys.*, **49**, 3122 (1968).
66. M. Karplus, J.A. Pople, 'Theory of Carbon NMR Chemical Shifts in Conjugated Molecules', *J. Chem. Phys.*, **38**, 2803 (1963).
67. H.M. McConnell, 'Theory of Nuclear Magnetic Shielding in Molecules. I. Long-range Dipolar Shielding of Protons', *J.Chem. Phys.*, **27**, 226 (1957).
68. J.A. Pople, 'The Theory of Chemical Shifts in Nuclear Magnetic Resonance. II. Interpretation of Proton Shifts', *Proc. Roy. Soc.*, **A239**, 550 (1957).
69. D.M. Grant, V.B. Cheney, 'Carbon-13 Magnetic Resonance. VII. Steric Perturbation of the Carbon-13 Chemical Shift', *J. Am. Chem. Soc.*, **89**, 5315 (1967).
70. D.K. Dalling, D.M. Grant, 'Carbon-13 Magnetic Resonance. XXI. Steric Interactions in the Methylcyclohexanes', *J. Am. Chem. Soc.*, **94**, 5318 (1972).
71. G.C. Levy, R.L. Lichter, G.L. Nelson, *Carbon-13 Nuclear Magnetic Resonance Spectroscopy*, John Wiley & Sons, New York, 1980.
72. W.W. Simons, *The Sadtler Guide to Carbon-13 NMR Spectra*, Sadtler, Philadelphia, 1983.
73. C.J. Pouchert, J. Behnke, *The Aldrich Library of ^{13}C and ^{1}H FT NMR Spectra*, Aldrich Chemical Co., Milwaukee, 1993.
74. Q.T. Pham, *Proton and Carbon NMR Spectra of Polymers*, Penton Press, London, 1991.
75. J.D. Memory, N.K. Wilson, *NMR of Aromatic Compounds*, Wiley, New York, 1982.
76. B.E. Mann, B.F. Taylor, *^{13}C MR Data for Organometallic Compounds*, Academic Press, New York, 1981.
77. E. Breitmaier, G. Haas, W. Voelter, *Atlas of Carbon-13 NMR Data*, Heyden, London, 1979.
78. L.F. Johnson, W.C. Jankowski, *Carbon-13 NMR Spectra: A Collection of Assigned, Coded, and Indexed Spectra*, R.E. Krieger Publ. Co., Huntington, 1978.
79. W. Bremser, B. Franke, H. Wagner, *Chemical Shift Ranges in Carbon-13 NMR Spectroscopy*, Verlag Chemie, Weinheim, 1982.
80. E.G. Paul, D.M. Grant, 'Additivity Relationships in Carbon-13 Chemical Shift Data for the Linear Alkanes', *J. Am. Chem. Soc.*, **85**, 1701 (1963).
81. D.M. Grant, E.G. Paul, 'Carbon-13 Magnetic Resonance. II. Chemical Shift Data for Alkanes', *J. Am. Chem. Soc.*, **86**, 2984 (1964).
82. L.P. Lindeman, J.Q. Adams, 'Carbon-13 Nuclear Magnetic Resonance Spectrometry. Chemical Shifts for Paraffins Through C_9', *Anal. Chem.*, **43**, 1245 (1971).
83. J.T. Clerc, E. Pretsch, S. Sternhill, *^{13}C-Kernresonanzspektroskopie*, Springer Verlag, Frankfurt, 1973.
84. T. Pehk, E. Lippmaa, 'Carbon-13 Chemical Shifts of Monosubstituted Cyclohexanes', *Org. Magn. Reson.*, **3**, 679 (1970).
85. J.B. Stothers, 'Carbon-13 Nuclear Magnetic Resonance Spectroscopy', *Q. Rev. Chem. Soc.*, **19**, 144 (1965).
86. H.N. Cheng, '^{13}C Spectral Simulation and Shift Prediction', *Trends Anal. Chem.*, **13**, 95 (1994).
87. D.K. Dalling, D.M. Grant, 'Carbon-13 Magnetic Resonance. XXI. Steric Interactions in the Methylcyclohexanes', *J. Am. Chem. Soc.*, **94**, 5318 (1972).
88. A.S. Perlin, H.J. Koch, 'Carbon-13 and Proton Chemical Shifts of Cyclohexanes in Relation to Interaction Energy', *Can. J. Chem.*, **48**, 2639 (1970).
89. K.M. Creceley, R.W. Creceley, J.H. Goldstein, 'Carbon-13 Nuclear Magnetic Resonance Spectra of Monosubstituted Cyclopropanes', *J. Phys. Chem.*, **74**, 2680 (1970).
90. H.J. Schneider, V. Hoppin, 'Stereochemical and Carbon-13 Magnetic Resonance Spectroscopic Investigations. VI. Equilibriums and Carbon-13 Magnetic Resonance Shifts of Cyclohexane Conformers', *Tetrahedr. Lett.*, **30**, 579 (1974).
91. O.A. Subbotin, N.M. Sergeyev, 'Pulsed Carbon-13 Fourier Transform Nuclear Magnetic Resonance Spectra of Monohalo-substituted Cyclohexanes at Low Temperatures', *J. Am. Chem. Soc.*, **97**, 1080 (1975).
92. G.E. Maciel, 'Carbon-13 Chemical Shifts of Vinyl Carbons', *J. Phys. Chem.*, **69**, 1947 (1965).
93. D.E. Dorman, M. Jautelat, J.D. Roberts, 'Carbon-13 Nuclear Magnetic Resonance Spectroscopy. Quantitative Correlation of the Carbon Chemical Shifts of Acyclic Alkenes', *J. Org. Chem.*, **36**, 2757 (1971).
94. G. Mujajima, K. Takahashi, K. Nishimoto, 'Carbon-13 Nuclear Magnetic Resonance Spectroscopy. IX. Monosubstituted Ethylenes', *Org. Magn. Reson.*, **6**, 413 (1974).
95. G.L. Nelson, E.A. Williams, 'Electronic Structure and Carbon-13 NMR', *Progr. Phys. Org. Chem.*, **12**, 229 (1976).
96. D.F. Ewing, 'Carbon-13 Substituent Effects in Monosubstituted Benzenes', *Org. Magn. Reson.*, **12**, 499 (1979).
97. G.L. Nelson, G.C. Levy, J.D. Cargioli, 'Solvent Effects in Carbon-13 Nuclear Magnetic Resonance. Electronic Perturbation of Aromatic Systems', *J. Am. Chem. Soc.*, **94**, 3089 (1972).
98. O. Ivanciuc, J.P. Rabine, D. Cabrol-Bass, '^{13}C NMR Chemical Shift Sum Prediction for Alkanes Using Neural Networks', *Comput. Chem.*, **21**, 437 (1997).
99. O. Ivanciuc, J.P. Rabine, D. Cabrol-Bass, A. Panaye, J.P. Doucet, '^{13}C NMR Chemical Shift Prediction of

the sp3 Carbon Atoms in the Position Relative to the Double Bond in Acyclic Alkenes', *J. Chem. Inf. Comput. Sci.*, **57**, 587 (1997).
100. O. Ivanciuc, J.P. Rabine, D. Cabrol-Bass, A. Panaye, J.P. Doucet, '^{13}C NMR Chemical Shift Prediction of the sp2 Carbon Atoms in Acyclic Alkenes Using Neural Networks', *J. Chem. Inf. Comput. Sci.*, **36**, 644 (1996).
101. D. Jahnels, U. Edlund, E. Johansson, S. Wold, 'A Multivariate Method for Carbon-13 NMR Chemical Shift Prediction Using Partial Least Squares Data Analysis', *J. Magn. Reson.*, **55**, 316 (1983).
102. Z. Li, Y. Huang, F. Hu, Q. Sheng, S. Peng, L. Mo, G. Chen, B. Yu, 'Neural Networks in Spectroscopy. Estimation and Prediction of Chemicals of ^{13}C NMR in Alkanes by Using Subgraphs', *Bopuxue Zazhi*, **14**, 507 (1997).
103. S.R. El Azzouzi, B.T. Fan, A. Panaye, J.P. Doucet, 'Application of Artificial Neural Networks to ^{13}C-NMR Shift Simulation of Methyl Substituted Monocycloalkanes and Polycyclohexanes', *Org. React.*, **31**, 3 (1997).
104. J.P. Doucet, A. Panaye, E. Feuilleaubois, P. Ladd, 'Neural Networks and Carbon-13 NMR Shift Prediction', *J. Chem. Inf. Comput. Sci.*, **33**, 320 (1993).
105. D.L. Clouser, P.C. Jurs, 'Simulation of the ^{13}C Nuclear Magnetic Resonance Spectra of Trisaccharides Using Multiple Linear Regression Analysis and Neural Networks', *Carbohydr. Res.*, **27**, 65 (1995).
106. H.N. Cheng, '^{13}C Spectral Simulation and Shift Prediction', *Trends Anal. Chem.*, **13**, 95 (1994).
107. L. Chen, W. Robien, 'OPSI: A Universal Method for Prediction of Carbon-13 NMR Spectra Based on Optimized Additivity Models', *Anal. Chem.*, **65**, 2282 (1993).
108. J.W. Ball, P.C. Jurs, 'Automated Selection of Regression Models Using Neural Networks for Carbon-13 NMR Spectral Prediction', *Anal. Chem.*, **65**, 505 (1993).
109. D.S. Wishart, M.S. Watson, R.F. Boyko, B.D. Sykes, 'Automated ^{1}H and ^{13}C Chemical Shift Prediction Using the BioMagResBank', *J. Biomol. NMR*, **10**, 329 (1997).
110. R.C. Schweitzer, G.W. Small, 'Automated Spectrum Simulation Methods for Carbon-13 Nuclear Magnetic Resonance Spectroscopy Based on Database Retrieval and Model-building Strategies', *J. Chem. Inf. Comput. Sci.*, **37**, 249 (1997).
111. R.C. Schweitzer, G.W. Small, 'Performance Enhancement of Vector-based Search Systems: Application to Carbon-13 Nuclear Magnetic Resonance Chemical Shift Prediction', *J. Chem. Inf. Comput. Sci.*, **36**, 46 (1996).
112. N.F. Ramsey, 'The Internal Diamagnetic Field Correction in Measurements of the Proton Magnetic Moment', *Phys. Rev.*, **77**, 567 (1950).
113. N.F. Ramsey, 'Chemical Effects in Nuclear Magnetic Resonance and Diamagnetic Susceptibility', *Phys. Rev.*, **86**, 243 (1952).
114. R. Grinter, 'Nuclear Spin Coupling', *Nucl. Magn. Reson.*, **1**, 51 (1972).
115. R.M. Stevens, R.M. Pitzer, W.N. Lipscomb, 'Perturbed Hartree–Fock Calculations. I. Magnetic Susceptibility and Shielding in the LiH Molecule', *J. Chem. Phys.*, **38**, 550 (1963).
116. R.M. Stevens, W.N. Lipscomb, 'Perturbed Hartree–Fock Calculations. V. Magnetic Properties of the BH Molecule', *J. Chem. Phys.*, **42**, 3666 (1965).
117. E.A. Laws, R.M. Stevens, W.N. Lipscomb, 'Theoretical Determination of the Magnetic Properties of the Diatomic AlH', *Chem. Phys. Lett.*, **4**, 159 (1969).
118. R. Ditchfield, 'Self-consistent Perturbation Theory of Diamagnetism. I. A Gage-invariant LCAO (Linear Combination of Atomic Orbitals) Method for NMR Chemical Shifts', *Mol. Phys.*, **27**, 789 (1974).
119. R. Ditchfield, 'Theoretical Studies of Magnetic Shielding in Water and Dimeric Water', *J. Chem. Phys.*, **65**, 3123 (1976).
120. F. Ribas Prado, C. Giessner-Prettre, J.-P. Dudley, A. Pullman, F. Young, J.F. Hinton, D. Harpool, 'Nuclear Magnetic Resonance Spectroscopy Applied to Lithium Ion Complexes by Small Ligands: Ab Initio and Experimental Study of Lithium-7 Chemical Shifts', *J. Magn. Reson.*, **37**, 431 (1980).
121. H. Fukui, K. Muira, H. Yamazaki, T. Nosaka, 'Calculation of NMR Chemical Shifts. V. The Gauge Invariant Coupled Hartree–Fock Calculation for Water, Oxonium (H_3O^+) and Hydroxide Ion', *J. Chem. Phys.*, **82**, 1410 (1983).
122. K. Wolinski, J.F. Hilton, P. Pulay, 'Efficient Implementation of the Gauge-independent Atomic Orbital Method for NMR Chemical Shift Calculations', *J. Am. Chem. Soc.*, **112**, 8251 (1990).
123. M. Schindler, W. Kutzelrigg, 'Theory of Magnetic Susceptibilities and NMR Chemical Shifts in Terms of Localized Quantities. II. Application to Some Simple Molecules', *J. Chem. Phys.*, **76**, 1919 (1982).
124. M. Schindler, W. Kutzelrigg, 'Theory of Magnetic Susceptibilities and NMR Chemical Shifts in Terms of Localized Quantities. III. Application to Hydrocarbons and Other Organic Molecules', *J. Am. Chem. Soc.*, **105**, 1360 (1983).
125. A.E. Hansen, T.D. Bowman, 'Localized Orbital/Localized Origin Method for Calculation and Analysis of NMR Shieldings. Applications to Carbon-13 Shielding Tensors', *J. Chem. Phys.*, **82**, 5035 (1985).
126. A.E. Hansen, T.D. Bowman, 'Calculation, Display and Analysis of the Nature of Nonsymmetric Nuclear Magnetic Resonance Shielding Tensors: Application to Three-membered Rings', *J. Chem. Phys.*, **91**, 3552 (1989).
127. J.D. Baker, M.C. Zerner, 'Calculation of Carbon-13 NMR Chemical Shift Using the Intermediate Neglect of Differential Overlap Model', *Int. J. Quant. Chem.*, **43**, 327 (1992).
128. A. Soderquist, J.C. Facelli, W.J. Horton, D.M. Grant, 'Low Temperature ^{13}C Magnetic Resonance. 9. Steric

Effects for Methyl Chemical Shift Tensors in Methylcyclohexane', *J. Am. Chem. Soc.*, **117**, 8441 (1995).
129. G.M. Bernard, G. Wu, R.E. Wasylishen, 'An Experimental and Theoretical Investigation of the Olefinic Carbon Chemical Shift Tensors in Trans-Stilbene and Pt(h_2-*trans*-stilbene)(PPh_3)$_2$', *J. Chem. Phys. A*, **102**, 3184 (1998).
130. A.M. Orendt, J.Z. Hu, Y.J. Jiang, J.C. Facelli, W. Wang, R.J. Pugmire, C. Ye, D.M. Grant, 'Solid-state ^{13}C NMR Measurements in Methoxynaphthalenes: Determination of the Substituent Chemical Shift Effects in the Principal Values', *J. Phys. Chem. A*, **101**, 9169 (1997).
131. G. Zheng, L. Wang, J. Hu, X. Zhang, L. Shen, C. Ye, G.A. Webb, 'Hydrogen Bonding Effects on the ^{13}C NMR Chemical Shift Tensors of Some Amino Acids in the Solid State', *Magn. Reson. Chem.*, **35**, 606 (1997).
132. T. Xu, D.H. Barich, P.D. Torres, J.B. Nicholas, J.F. Haw, 'Carbon-13 Chemical Shift Tensors for Acylium Ions: A Combined Solid-state NMR and Ab Initio Molecular Orbital Study', *J. Am. Chem. Soc.*, **119**, 396 (1997).
133. P.v.R. Schleyer, J. Gauss, M. Buehl, R. Greatrex, M.A. Fox, 'Even More Reliable NMR Chemical Shift Computations by the GIAO–MP2 Method', *J. Chem. Soc., Chem. Comm.*, **23**, 1766 (1993).
134. G. Zheng, J. Hu, X. Zhang, L. Shen, C. Ye, G.A. Webb, 'Quantum Chemical Calculation and Experimental Measurement of the ^{13}C Chemical Shift Tensors of Vanillin and 3,4-Dimethoxybenzaldehyde', *Chem. Phys. Lett.*, **266**, 533 (1997).
135. P.T. Czech, X.Q. Ye, R.F. Fenske, 'Application of the Penske-Hall Molecular Orbital Method to the Calculation of Carbon-13 NMR Shifts in Organometallic Compounds: Correlation Between Calculated sp Values and the Observed Chemical Shift', *Organometallics*, **9**, 2016 (1990).
136. J.G. Pearson, H. Le, L.K. Sanders, N. Godbout, R.H. Havlin, E. Oldfield, 'Predicting Chemical Shifts in Proteins: Structure Refinement of Valine Residues by Using Ab Initio and Empirical Geometry Optimization', *J. Am. Chem. Soc.*, **119**, 11941 (1997).
137. L. Olsson, D. Cremer, 'Sum-over-states Density Functional Perturbation Theory: Prediction of Reliable ^{13}C, ^{15}N and ^{17}O Nuclear Magnetic Resonance Chemical Shifts', *J. Chem. Phys.*, **105**, 8995 (1996).
138. J.A. Pople, 'Molecular-orbital Theory of Diamagnetism. I. An Approximate LCAO Scheme', *J. Chem. Phys.*, **37**, 53 (1962).
139. J.A. Pople, 'Molecular-orbital Theory of Diamagnetism. II. Calculation of Pascal Constants for Some Noncyclic Molecules', *J. Chem. Phys.*, **37**, 60 (1962).
140. U. Sternberg, W. Prieb, 'New Semi-empirical Approach for the Calculation of ^{13}C Chemical-shift Tensors', *J. Magn. Reson.*, **125**, 8 (1997).
141. J.A. Pople, 'The Theory of Carbon Chemical Shifts in NMR', *Mol. Phys.*, **7**, 301 (1964).
142. B.V. Cheney, D.M. Grant, 'Carbon-13 Magnetic Resonance. VIII. Theory of Carbon-13 Chemical Shifts Applied to Saturated Hydrocarbons', *J. Am. Chem. Soc.*, **89**, 5319 (1967).
143. H. Beierbeck, J.K. Saunders, J.W. ApSimon, 'The Semiempirical Derivation of Carbon-13 Nuclear Magnetic Resonance Chemical Shifts. Hydrocarbons, Alcohols, Amines, Ketones, and Olefins', *Can. J. Chem.*, **55**, 2813 (1977).
144. H. Beierbeck, J.K. Saunders, 'Analysis of Carbon-13 Nuclear Magnetic Resonance Chemical Shifts of Acyclic Hydrocarbons', *Can. J. Chem.*, **58**, 1258 (1980).
145. M. Gonzalez-Sierra, D.A. Bustos, E.A. Ruveda, A.C. Oliveri, M. Grasselli, 'Semiempirical Calculation of Carbon-13 Nuclear Magnetic Resonance Chemical Shifts of Acyclic Hydrocarbons. Application to the Stereochemical Analysis of Steroidal Side Chains', *Can. J. Chem.*, **66**, 77 (1988).
146. A.W. Overhauser, 'Paramagnetic Relaxation in Metals', *Phys. Rev.*, **89**, 689 (1953).
147. A.W. Overhauser, 'Polarization of Nuclei in Metals', *Phys. Rev.*, **92**, 411 (1953).
148. I. Solomon, 'Relaxation Processes in a System of Two Spins', *Phys. Rev.*, **99**, 559 (1955).
149. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, Weinheim, 1989.
150. K. Tori, J. Nishikawa, Y. Takeuchi, 'NMR Studies of Penicillins and Cephalosporins. Part 2. Carbon-13 NMR Spectra of Cephalosporins. Signal Assignments of Free Acids and Esters', *Tetrahedr. Lett.*, **22**, 2793 (1981).
151. A.J. Jones, D.M. Grant, K.F. Kuhlmann, 'Carbon-13 Magnetic Resonance. XIII. Assignment of Carbon-13 Resonances Using the Nuclear Overhauser Effect', *J. Am. Chem. Soc.*, **91**, 5013 (1969).
152. P. Sohar, J. Szabo, G. Stajer, G. Bernath, 'NMR Study of the Stereochemistry of Polycyclic Compounds', *Kem. Kozl.*, **73**, 283 (1991).
153. L.A.C. Pieters, A.J. Vlietinck, 'Application of Quantitative Proton and Carbon-13 NMR Spectroscopy in Drug Analysis', *J. Pharm. Biomed. Anal.*, **7**, 1405 (1989).
154. R. Neidlein, W. Kramer, V. Ulrich, 'Application of Selective ^{13}C(Proton) – NOE Difference Spectroscopy in Determination of the Configuration of Geometric Isomers', *Helv. Chim. Acta*, **69**, 898 (1986).
155. M.F. Aldersley, F.M. Dean, B.E. Mann, 'The Use of Carbon-13 (Proton Selective) Nuclear Overhauser Enhancement Experiments in the Determination of the Structure of a Highly Crowded Quinone–Diazoalkane Epoxide Adduct', *J. Chem. Soc., Chem. Comm.*, 107 (1983).
156. K.E. Kover, G. Batta, 'Theoretical and Practical Aspects of One- and Two-dimensional Heteronuclear Overhauser Experiments and Selective Carbon-13 T1-Determinations of Heteronuclear Distances', *Prog. Nucl. Magn. Reson.*, **19**, 223 (1987).

157. H.E. Gottlieb, 'Techniques for Signal Assignments in Carbon-13 NMR Spectroscopy of Naturally Occurring Aromatic Compounds', *Isr. J. Chem.*, **16**, 57 (1977).
158. M.A. Khaled, C.L. Watkins, 'Detection of Hydrogen Bonding in Peptide by $^{13}C[^{1}H]$ Nuclear Overhauser Effect', *J. Am. Chem. Soc.*, **105**, 3363 (1983).
159. Y.E. Shapiro, 'Analysis of Chain Microstructure by Proton and Carbon-13 NMR Spectroscopy', *Bull. Magn. Reson.*, **7**, 27 (1985).
160. C. Deverell, 'Nuclear Magnetic Relaxation by Spin-rotation Interaction. Determination of Spin-rotation Interaction Constants from Nuclear Magnetic Shielding Constants', *Mol. Phys.*, **18**, 319 (1970).
161. A. Carrington, A.D. McLachlan, *Introduction to Magnetic Resonance*, Harper and Row, New York, 1967.
162. R.L. Vold, J.S. Waugh, M.P. Klein, D.E. Phelps, 'Measurement of Spin Relaxation in Complex Systems', *J. Chem. Phys.*, **48**, 3831 (1968).
163. R. Freeman, H.D.W. Hill, 'High-resolution Studies of Nuclear Spin-lattice Relaxation', *J. Chem. Phys.*, **51**, 3140 (1969).
164. R. Freeman, H.D.W. Hill, 'Fourier Transform Study of NMR Spin-lattice Relaxation by 'Progressive Saturation' ', *J. Chem. Phys.*, **54**, 3367 (1971).
165. J.L. Markley, W.J. Horsley, M.P. Klein, 'Spin-lattice Relaxation Measurements in Slowly Relaxing Complex Spectra', *J. Chem. Phys.*, **55**, 3604 (1971).
166. G.C. McDonald, J.S. Leigh, Jr., 'New Method for Measuring Longitudinal Relaxation Times', *J. Magn. Reson.*, **9**, 358 (1973).
167. H.Y. Carr, E.M. Purcell, 'Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments', *Phys. Rev.*, **94**, 630 (1954).
168. S. Meiboom, D. Gill, 'Modified Spin-echo Method for Measuring Nuclear Relaxation Times', *Rev. Sci. Instrum.*, **29**, 688 (1958).
169. J. Jeener, Ampere International Summer School, Basko Polje, Yugoslavia, 1971.
170. L. Mueller, A. Kumar, R.R. Ernst, 'Two-dimensional Carbon-13 NMR Spectroscopy', *J. Chem. Phys.*, **63**, 5490 (1975).
171. W.P. Aue, E. Bartholdi, R.R. Ernst, 'Two-dimensional Spectroscopy. Application to Magnetic Resonance', *J. Chem. Phys.*, **64**, 2229 (1976).
172. W.R. Croasmun, R.M.K. Carlson, *Two-dimensional NMR Spectroscopy Applications for Chemists and Biochemists*, 2nd edition, VCH, New York, 1994.
173. G.E. Martin, A.S. Zektzer, *Two-dimensional NMR Methods for Establishing Molecular Connectivity. A Chemist's Guide to Experiment Selection, Performance, and Interpretation*, VCH, New York, 1988.
174. G. Bodenhausen, R. Freeman, R. Niedermeyer, D.L. Turner, 'Double Fourier Transformation in High-resolution NMR', *J. Mag. Reson.*, **26**, 133 (1977).
175. R. Freeman, G.A. Morris, 'Experimental Chemical Shift Correlation Maps in Nuclear Magnetic Resonance Spectroscopy', *J. Chem. Soc. Chem. Commun.*, **684** (1978).
176. A. Bax, G.A. Morris, 'An Improved Method for Heteronuclear Chemical Shift Correlation by Two-dimensional NMR', *J. Magn. Reson.*, **42**, 501 (1981).
177. J. Buddrus, H. Bauer, 'New Analytical Methods. Part 32. Determination of the Carbon Skeleton of Organic Compounds by Double Quantum Coherent Carbon-13 NMR Spectroscopy, the INADEQUATE Pulse Sequence', *Angew. Chem.*, **99**, 642 (1987).
178. A. Bax, R. Freeman, T.A. Frenkiel, 'An NMR Technique for Tracing Out the Carbon Skeleton of an Organic Molecule', *J. Am. Chem. Soc.*, **103**, 2102 (1981).
179. G. Bodenhausen, D.J. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185 (1980).
180. A. Bax, M.F. Summers, 'Proton and Carbon-13 Assignments from Sensitivity-enhanced Detection of Heteronuclear Multiple-bond Connectivity by 2D Multiple Quantum NMR', *J. Am. Chem. Soc.*, **108**, 2093 (1986).
181. L. Muller, 'Sensitivity Enhanced Detection of Weak Nuclei Using Heteronuclear Multiple Quantum Coherence', *J. Am. Chem. Soc.*, **101**, 4481 (1979).
182. R.E. Hurd, B.K. John, 'Gradient-enhanced Proton-detected Heteronuclear Multiple-quantum Coherence Spectroscopy', *J. Magn. Reson.*, **91**, 648 (1991).
183. J. Weigelt, G. Otting, '^{1}H-Detected INEPT–INADEQUATE at Natural ^{13}C Abundance', *J. Magn. Reson. Ser. A*, **113**, 128 (1995).
184. S. Gillett, J.-J. Delpeuch, 'Optimum Conditions for Nondestructive Quantitative Analysis by Carbon-13 NMR', *J. Magn. Reson.*, **38**, 433 (1980).
185. S.J. Opella, D.J. Nelson, O. Jardetzky, 'Dynamics of Nuclear Overhauser Enhancement in Proton Decoupled Carbon-13 Nuclear Magnetic Resonance', *J. Chem. Phys.*, **64**, 2533 (1976).
186. D. Canet, 'Systematic Errors Due to Improper Waiting Times in Heteronuclear Overhauser Effect Measurements by the Gated Decoupling Technique', *J. Magn. Reson.*, **23**, 361 (1976).

Chemical Shifts in Nuclear Magnetic Resonance

Scott E. Van Bramer
Widener University, Chester, USA

1 Introduction **1**
1.1 Definition of Chemical Shift 1
1.2 Chemical Shift Reference 1
1.3 Solvent Peaks 2
1.4 Theory 2
2 Carbon-13 Chemical Shift **4**
2.1 Correlation Table 4
2.2 Alkanes 4
2.3 Substituted Alkanes 6
2.4 Substituted Alkenes 8
2.5 Substituted Benzenes 8
2.6 Cyclic Compounds 9
3 Proton Chemical Shifts **10**
3.1 Correlation Table 10
3.2 Substituted Alkanes 11
3.3 Substituted Alkenes 14
3.4 Substituted Benzenes 15
3.5 Cyclic Compounds 16
3.6 Heteroatoms 16
4 Coupling Constants **16**
5 Computer Software **16**
6 Additional Information **17**
Abbreviations and Acronyms **17**
Related Articles **17**
References **17**

This article explains a variety of methods used to calculate ^{13}C and proton (^{1}H) chemical shifts for alkanes, alkenes, aromatic compounds, and cyclic compounds. Additivity rules, which account for the effect of different functional groups, are used to calculate chemical shifts for a wide range of compounds. Example calculations are provided for each rule, to clarify the use and for comparison with experimental results. The chemical shifts calculated with these rules are typically accurate to ±10 ppm for ^{13}C nuclear magnetic resonance (NMR) and ±0.5 ppm for ^{1}H-NMR.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

NMR spectra contain several types of information that are used for interpretation. This article explains a variety of methods used to calculate ^{13}C and proton (^{1}H) chemical shifts. These predicted chemical shifts are used to identify functional groups present in a compound, to distinguish possible unknown structures, and to identify peaks in a spectrum. The calculations used to predict chemical shifts are based upon additivity rules that account for how functional groups affect the chemical shifts of neighboring atoms. These calculated chemical shifts can be accurate to ±10 ppm for ^{13}C-NMR and ±0.5 ppm for ^{1}H-NMR. Each of the additivity rules presented here were developed for certain compound classes. The accuracies reported are to fit the set of compounds used to construct the rule. The greater the difference between the compound of interest and the compounds used to create the rules, the greater the uncertainty in the calculated chemical shift. Rules for many common compound classes and functional groups are provided, but interested readers should refer to the references for additional information.

1.1 Definition of Chemical Shift

The chemical shift, δ, in NMR spectroscopy is defined in Equation (1):

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{reference}}}{\nu_{\text{reference}}} \times 10^6 \text{ ppm} \quad (1)$$

The reference frequency, $\nu_{\text{reference}}$, is measured by adding a small amount of the reference compound to the sample. The signal frequency, ν_{sample}, is the Larmor frequency for the nucleus of interest. Equation (1) is frequently written using the spectrometer frequency, $\nu_{\text{spectrometer}}$, as shown in Equation (2):

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{reference}}}{\nu_{\text{spectrometer}}} \times 10^6 \text{ ppm} \quad (2)$$

The strength of the instrument's static magnetic field determines $\nu_{\text{spectrometer}}$. Because this is a fixed value, the chemical shift axis is readily calculated. In liquid ^{13}C and ^{1}H NMR spectra the difference between Equations (1) and (2) is insignificant.

1.2 Chemical Shift Reference

Tetramethylsilane (TMS) is used as the primary reference for ^{1}H- and ^{13}C-NMR. The spectrum is plotted with TMS, at 0 ppm, on the right. Increasing chemical shift, δ in parts per million, is plotted from right to left. Increasing chemical shift also corresponds to increasing frequency, decreasing magnetic field strength,

Table 1 Chemical shifts of common standards. (Reproduced by permission of Academic Press and G. Levy from G. Levy, J. Cargioli, *J. Mag. Reson.*, **6**, 143–144 (1972).)

Compound (solvent)	^{13}C chemical shift downfield from internal TMS (ppm)	
	Protio compound	Perdeutero compound
Cyclohexane	27.51	26.06
Acetone	30.43	29.22
Dimethylsulfoxide	40.48	39.56
Methylene chloride	54.02	53.61
Dioxane	67.40	
Chloroform	77.17	76.91
Carbon tetrachloride	95.99	
Benzene	128.53	127.96
Acetic acid (CO)	178.27	
CS_2	192.8	
CS_2 capillary (external TMS)	193.7	

or decreased shielding. Decreased shielding is frequently referred to as deshielding, so increasing chemical shift also corresponds to increased deshielding.

TMS is the primary chemical shift standard and should be used whenever possible. Samples in polar solvents and many older literature values are referenced to a secondary standard. This is most common for ^{13}C data in the older literature. Table 1, from Levy and Cargioli,[1] lists the chemical shifts of several other compounds that have been used as ^{13}C NMR standards.

1.3 Solvent Peaks

Carbon-13 NMR spectra are typically acquired by dissolving a sample in a suitable solvent. Table 2, from Breitmaier and Voelter,[2] gives the chemical shift, splitting patterns, and coupling constants for the most common NMR solvents. The chemical shift for many solvents is concentration-dependent and may change when a solute is added. As a result, the solvent is not an appropriate chemical shift reference for NMR.

1.4 Theory

NMR is such a powerful tool for structural identification because the chemical shift depends upon the electronic

Table 2 Carbon-13 NMR solvent peaks. (Reproduced by permission of W. Voelter from E. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy, Methods and Applications in Organic Chemistry and Biochemistry*, VCH, Weinheim, Germany, 1990.)

Compound	Multiplicity C–D splitting	Coupling J_{CD} (Hz)	Chemical shift ^{13}C (ppm)	Isotopic ^{1}H compound	
				^{13}C (ppm)	J_{CH} (Hz)
*d*2-Methylene chloride	Quintet	27	53.1	53.8	177.5
d-Chloroform	Triplet	32	77.0 ± 0.5	78.0 ± 0.5	210.5
d-Bromoform	Triplet	31.5	10.2	10.3	205
*d*3-Nitromethane	Septet	23.5	60.5	61.1	146.5
*d*3-Acetonitrile	Septet	32	1.3	1.7	117.4
	Not resolved	<1	118.2	118.2	<5
*d*4-Methanol	Septet	21.5	49.0	49.9	141
*d*6-Ethanol	Septet	19.5	15.8	16.9	127
	Quintet	22	55.4	56.3	140.5
*d*10-Diethylether	Septet	19	13.4	14.5	127
	Quintet	21	64.3	65.3	142
*d*8-1,4-Dioxane	Quintet	22	66.5	67.6	146
*d*8-Tetrahydrofuran	Quintet	20.5	25.2	26.2	132
	Quintet	22	67.4	68.2	137
*d*6-Acetone	Septet	20	29.8	30.7	126
	Not resolved	<1	206.5	206.7	7.5
*d*6-Dimethylsulfoxide	Septet	21	39.7	40.9	137
*d*4-Acetic acid	Septet	20	20.0	20.9	129
	Not resolved	<1	178.4	178.8	7.5
*d*7-Dimethylformamide	Septet	21	30.1	30.9	138
	Septet	21	35.2	36.0	138
	Triplet	30	167.7	167.9	188
*d*18-Hexamethyl phosphoramide	Septet	21	35.8	36.9	129.5
*d*12-Cyclohexane	Quintet	19	26.4	27.8	126
*d*6-Benzene	Triplet	24	128.0 ± 0.5	128.5 ± 0.5	161
*d*5-Pyridine	Triplet (3,5)	25	123.5	124.2	166
	Triplet (4)	24.5	135.5	136.2	163
	Triplet (2,6)	27.5	149.2	149.7	177

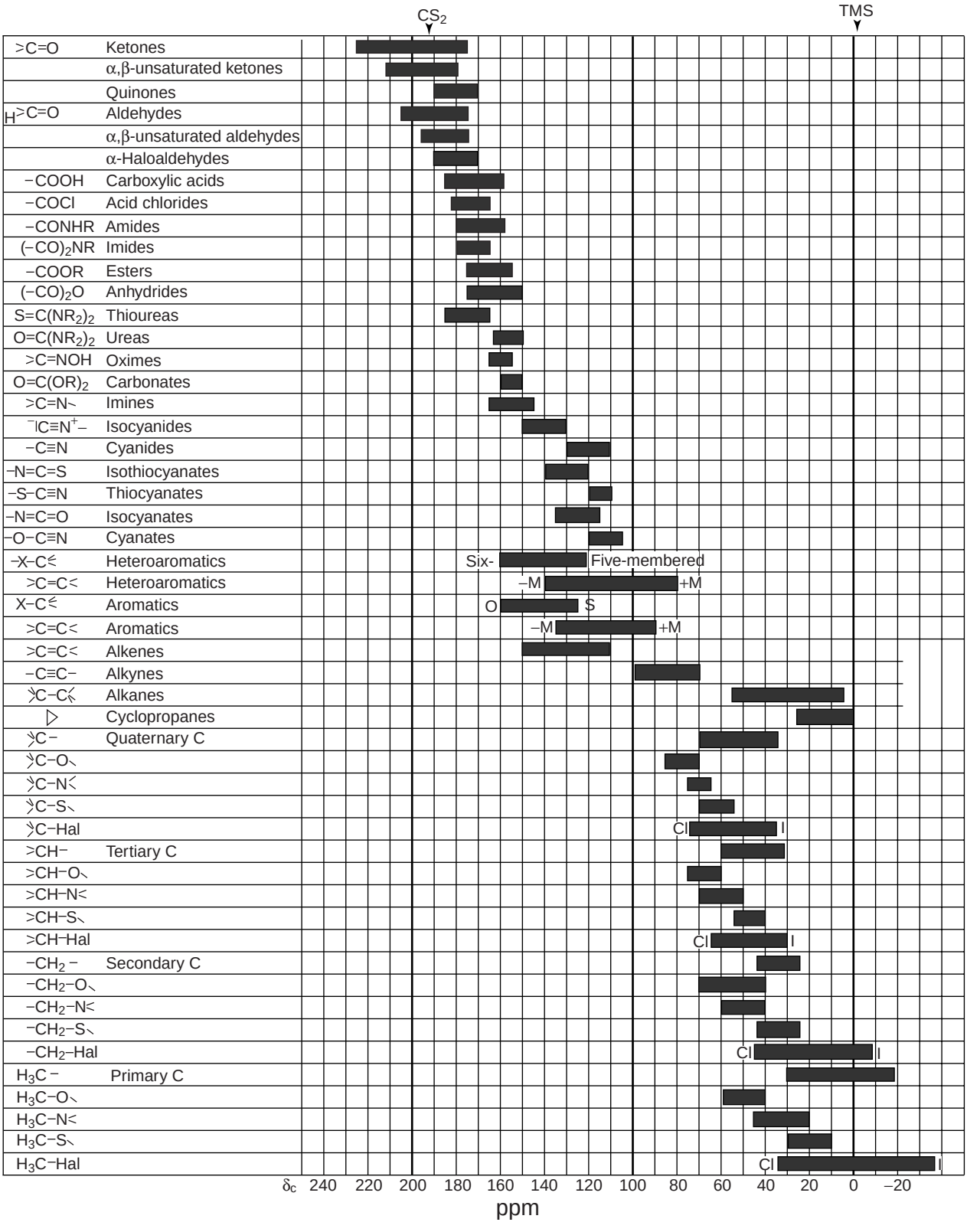

Figure 1 Carbon-13 chemical shift correlation. (Reproduced by Permission of Wiley–VCH and E. Breitmaier from E. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy, Methods and Applications in Organic Chemistry and Biochemistry*, VCH, Weinheim, Germany, 1990.)

environment surrounding each atom in a molecule. When a sample is placed in the spectrometer, the magnetic field causes the electron cloud surrounding a nucleus to circulate. This circulating current produces a magnetic field that opposes the magnetic field of the spectrometer. The result is a shielding of the nucleus from the applied magnetic field. Different chemical shifts are caused by different amounts of shielding. One example of this is caused by highly electronegative elements. When an electronegative element is nearby, it reduces the electron density surrounding a nucleus. This deshields the nucleus so that the NMR signal is observed with a large chemical shift. The chemical shift of methane is 0.13 ppm but the chemical shift of chloromethane is 3.05 ppm. Another example of this shielding is the ring current observed in aromatic compounds. This ring current is caused by the free circulation of electrons around the aromatic ring. Depending upon the geometry of the molecule, these induced currents may result in either shielding or deshielding of nuclei.

2 CARBON-13 CHEMICAL SHIFT

2.1 Correlation Table

Carbon-13 chemical shifts are predicted using additivity rules to account for the deshielding caused by various functional groups. Figure 1 is a correlation table, based upon data from Breitmaier and Voelter,[2] that shows the typical ^{13}C chemical shift range for observing various functional groups. This figure is a useful starting point, but additivity rules are needed for more precise estimates of the chemical shift.

2.2 Alkanes

2.2.1 Grant and Paul Rule for Linear Alkanes

Grant and Paul[3] developed Equation (3) for calculating the chemical shift of linear alkanes:

$$\delta = -2.6 + 9.12n_\alpha + 9.38n_\beta - 2.53n_\gamma + 0.37n_\delta + 0.17n_\varepsilon \quad (3)$$

In Equation (3) δ is the calculated ^{13}C chemical shift; -2.6 is the ^{13}C chemical shift for methane; n_α is the number of α-carbons, one bond removed from the carbon of interest; n_β is the number of β-carbons, two bonds removed from the carbon of interest; n_γ is the number of γ-carbons, three bonds removed from the carbon of interest; n_δ is the number of δ-carbons, four bonds removed from the carbon of interest; and n_ε is the number of ε-carbons, five bonds removed from the carbon of interest. This version of the Grant and Paul rule was developed using linear alkanes up to *n*-decane. The equation fits the 30 observed carbons in this data set with a standard deviation of ± 0.21 ppm.

Table 3 Calculated ^{13}C shifts for *n*-pentane using the Grant and Paul rule for linear alkanes (Equation 3)

Position	C1 (ppm)	C2 (ppm)	C3 (ppm)
Methane (δ)	-2.6	-2.6	-2.6
$\alpha(n_\alpha \times 9.12)$	1×9.12	2×9.12	2×9.12
$\beta(n_\beta \times 9.38)$	1×9.38	1×9.38	2×9.38
$\gamma(n_\gamma \times -2.53)$	$1 \times (-2.53)$	$1 \times (-2.53)$	$0 \times (-2.53)$
$\delta(n_\delta \times 0.37)$	1×0.37	0×0.37	0×0.37
$\varepsilon(n_\varepsilon \times 0.17)$	0×0.17	0×0.17	0×0.17
Calculated shift	13.74	22.49	34.40
Experimental shift[4]	14.2	22.8	34.8

In Table 3, the Grant and Paul rule for linear alkanes is used to calculate the ^{13}C chemical shift for *n*-pentane (109-66-0) **(1)**. The chemical shifts calculated from Equation (3) very closely match the experimental data given by Pretsch et al.[4]

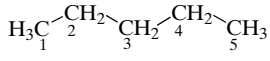

(1)

2.2.2 Grant and Paul Rule for Branched Alkanes

The version of the Grant and Paul rule given in Equation (3) was developed for linear alkanes. If it is used to calculate the chemical shift for carbons in branched alkanes, the error increases significantly. To compensate for this, Grant and Paul developed a second equation that includes correction terms for branched alkanes. This equation was developed by fitting both the linear alkanes used for Equation (3) and branched alkanes with up to six carbons. The Grant and Paul rule for branched alkanes[3] is given in Equation (4):

$$\delta = -2.6 + 9.09n_\alpha + 9.40n_\beta - 2.49n_\gamma + 0.31n_\delta + 0.11n_\varepsilon + \sum A \quad (4)$$

The coefficients and the variables in Equation (4) are almost identical to Equation (3). As with Equation (3) δ is the calculated ^{13}C chemical shift; -2.6 is the ^{13}C chemical shift for methane; n_α is the number of α-carbons, one bond removed from the carbon of interest; n_β is the number of β-carbons, two bonds removed from the carbon of interest; n_γ is the number of γ-carbons, three bonds removed from the carbon of interest; n_δ is the number of δ-carbons, four bonds removed from the carbon of interest; and n_ε is the number of ε-carbons, five bonds removed from the carbon of interest. Although the coefficients for Equation (4) are slightly different from those of Equation (3), the most significant difference in

Table 4 Correction terms for calculating chemical shifts of branched alkanes using the Grant and Paul rule for branched alkanes (Equation 4). (Reproduced with permission from D. Grant, E. Paul, *J. Am. Chem. Soc.*, **86**, 2984–2990 (1964). Copyright 1964 American Chemical Society.)

Correction term[a]	Coefficient (ppm)
1°(3°)	−1.12
1°(4°)	−3.37
2°(3°)	−2.50
2°(4°)	−7.23
3°(2°)	−3.65
3°(3°)	−9.47
4°(1°)	−1.50
4°(2°)	−8.36

[a] The first digit is for the carbon being calculated (1° – methyl, CH_3; 2° – methylene, CH_2; 3° – methine, CH; 4° – quaternary, C). The second digit is for the type of adjacent carbon. The correction term 2°(3°) is used for a secondary carbon that has a tertiary carbon attached.

Equation (4) is term A. The values for A, given in Table 4, are added for each attached carbon to compensate for branching. Grant and Paul found that the calculated chemical shift using Equation (4) and the corrective terms in Table 4 predicted the measured value for 53 observations of linear and branched alkanes in their data set with a standard deviation of ±0.3 ppm.

Table 5 shows the calculations for the ^{13}C chemical shifts in 2,2,4-trimethylhexane (16747-26-5) **(2)** using the Grant and Paul rule for branched alkanes. For comparison, the results of the Grant and Paul rule for linear alkanes and experimental data from Lindeman and Adams[5] are also shown.

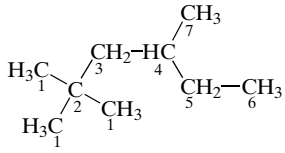

(2)

2.2.3 Lindeman and Adams Rule for Alkanes

A different method to calculate the chemical shift of alkanes was developed by Lindeman and Adams.[5] The Lindeman and Adams rule was developed using all paraffin isomers through C_8 and 24 of the possible C_9 isomers. Lindeman and Adams compared their results with the Grant and Paul rules. The Grant and Paul rule for linear alkanes gave a standard error of ±2.0 ppm with the data set of 59 compounds used by Lindeman and Adams. Using the modified Grant and Paul rule for branched alkanes reduced the standard error to ±1.4 ppm. The Lindeman and Adams rule gave a standard error of ±0.79 ppm. For primary and secondary carbon atoms, all of the rules work well. The Lindeman and Adams rule, however, works better for the tertiary and quaternary carbon atoms in this data set. The authors note that the Lindeman and Adams rule does not work well for highly branched compounds.

The Lindeman and Adams rule is given in Equation (5):

$$\delta_c(k) = B_S + \sum D_M A_{SM} + \gamma_S N_{k3} + \Delta_S N_{k4} \qquad (5)$$

The values for B_S, A_{SM}, γ_S and Δ_S are given in Table 6; $\delta_c(k)$ is the calculated chemical shift of carbon k; S is the number of carbons attached to carbon k; the value for B_S is given in Table 6 for primary, secondary, tertiary

Table 5 Calculated ^{13}C shifts for 2,2,4-trimethylhexane using the Grant and Paul rule for branched alkanes (Equation 4)

Position	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C5 (ppm)	C6 (ppm)	C7 (ppm)
Methane (δ)	−2.6	−2.6	−2.6	−2.6	−2.6	−2.6	−2.6
$\alpha(n_\alpha \times 9.09)$	1 × 9.09	4 × 9.09	2 × 9.09	3 × 9.09	2 × 9.09	1 × 9.09	1 × 9.09
$\beta(n_\beta \times 9.40)$	3 × 9.40	1 × 9.40	5 × 9.40	2 × 9.40	2 × 9.40	1 × 9.40	2 × 9.40
$\gamma(n_\gamma \times -2.49)$	1 × (−2.49)	2 × (−2.49)	1 × (−2.49)	3 × (−2.49)	1 × (−2.49)	2 × (−2.49)	2 × (−2.49)
$\delta(n_\delta \times 0.31)$	2 × 0.31	1 × 0.31	0 × 0.31	0 × 0.31	3 × 0.31	1 × 0.31	3 × 0.31
$\varepsilon(n_\varepsilon \times 0.11)$	1 × 0.11	0 × 0.11	0 × 0.11	0 × 0.11	0 × 0.11	3 × 0.11	0 × 0.11
Terms A	1 × (−3.37)	3 × (−1.50) 1 × (−8.36)	1 × (−2.50) 1 × (−7.23)	2 × (−3.65)	1 × (−2.50)		1 × (−1.12)
Calculated shift from Equation (4)	29.56	25.63	50.36	32.35	30.32	11.55	20.12
Calculated shift from Equation (3)	33.04	38.57	60.01	35.93	32.98	11.72	21.33
Experimental shift[5]	29.9	31.0	51.0	31.9	31.0	11.2	21.9

Table 6 Terms for calculating ^{13}C shifts using the Lindeman and Adams rule (Equation 5).[5] (Reproduced with permission from L. Lindeman, J. Adams, *Anal. Chem.*, **43**, 1245–1252 (1971). Copyright 1971 American Chemical Society.)

Type of carbon at position k	B_S value (ppm)	Type of adjacent carbon	A_{SM} value (ppm)	γ_S value (ppm)	Δ_S value (ppm)
Primary (CH_3)	6.80	Secondary (CH_2)	9.56	−2.99	0.49
		Tertiary (CH)	17.83		
		Quaternary (C)	25.48		
Secondary (CH_2)	15.34	Secondary (CH_2)	9.75	−2.69	0.25
		Tertiary (CH)	16.70		
		Quaternary (C)	21.43		
Tertiary (CH)	23.46	Secondary (CH_2)	6.60	−2.07	0
		Tertiary (CH)	11.14		
		Quaternary (C)	14.70		
Quaternary (C)	27.77	Secondary (CH_2)	2.26	0.68	0
		Tertiary (CH)	3.96		
		Quaternary (C)	7.35		

Table 7 Calculated ^{13}C shifts for *n*-pentane using the Lindeman and Adams rule (Equation 5)

Position	C1 (ppm)	C2 (ppm)	C3 (ppm)
B_S	6.80	15.34	15.34
$D_M \times A_{SM}$	1 × 9.56	1 × 9.75	2 × 9.75
$N_{k3} \times \gamma_S$	1 × (−2.99)	1 × (−2.69)	0 × (−2.69)
$N_{k4} \times \Delta_S$	1 × 0.49	0 × 0.25	0 × 0.25
Calculated shift using the Lindeman and Adams rule [Equation 5]	13.86	22.4	34.84
Calculated shift using the Grant and Paul rule [Equation 3]	13.74	22.49	34.40
Experimental shift[4]	14.2	22.8	34.8

or quaternary carbons; the next term in Equation (5) accounts for each of the carbons attached to k; M is the number of carbons attached to each of these carbons; A_{SM} is given in Table 6 for secondary, tertiary and quaternary carbons attached to each of the possible types of carbon at position k; D_M is the number of each type of carbon present in the structure; there is no correction term for the carbons separated from carbon k by two bonds; N_{k3} is the number of carbons three bonds away from k; γ_S is a constant for the carbons three bonds away from k; N_{k4} is the number of carbons four bonds away from k; and Δ_S is a constant for the carbons four bonds away from k.

Although the use of the Lindeman and Adams rule appears complicated, the layout of Table 6 simplifies its use. To demonstrate the application of the Lindeman and Adams rule to a linear alkane, calculations are given in Table 7 for *n*-pentane **(1)**. To demonstrate the application of the Lindeman and Adams rule to a branched alkane, calculations are given in Table 8 for 2,2,4-trimethylhexane **(2)**. The results from the Lindeman and Adams rule and results from the appropriate Grant and Paul rule compare favorably with the experimental data from Pretsch et al.[4] with the exception of C2 in 2,2,4-trimethylhexane. The Lindeman and Adams rule gives much better results for this highly branched carbon.

Table 9 gives ^{13}C chemical shifts for selected linear and branched alkanes from Grant and Paul.[3]

2.3 Substituted Alkanes

The Grant and Paul rules and the Lindeman and Adams rule are used to calculate the chemical shift of saturated alkanes. The addition of functional groups has a significant effect on the ^{13}C chemical shift of any neighboring carbon atoms. The change in chemical shift, $\Delta\delta$, caused by substitution is consistent and additive. There are two common methods to account for the effect on the chemical shift of the alkane backbone.

Table 8 Calculated ^{13}C shifts for 2,2,4-trimethylhexane using the Lindeman and Adams rule (Equation 5)

Position	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C5 (ppm)	C6 (ppm)	C7 (ppm)
B_S	6.80	27.77	15.34	23.46	15.34	6.80	6.80
$D_M \times A_{SM}$	1 × 25.46	1 × 2.26	1 × 16.70 1 × 21.43	2 × 6.60	1 × 16.70	1 × 9.56	1 × 17.83
$N_{k3} \times \gamma_S$	1 × (−2.99)	2 × 0.68	1 × (−2.69)	3 × (−2.07)	1 × (−2.69)	2 × (−2.99)	2 × (−2.99)
$N_{k4} \times \Delta_S$	2 × 0.49	1 × 0	0 × 0.25	0 × 0	3 × 0.25	1 × 0.49	3 × 0.49
Calculated shift using the Lindeman and Adams rule [Equation 5]	30.25	31.39	50.78	30.45	30.10	10.87	20.12
Calculated shift using the Grant and Paul rule [Equation 4]	29.56	25.63	50.36	32.35	30.32	11.55	20.12
Experimental shift[5]	29.9	31.0	51.0	31.9	31.0	11.2	21.9

Table 9 Carbon-13 chemical shifts for selected linear and branched alkanes. (Reproduced with permission from D. Grant, E. Paul, *J. Am. Chem. Soc.*, **86**, 2984–2990 (1964). Copyright 1964 American Chemical Society.)

Compound	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C5 (ppm)
Methane	−2.3				
Ethane	5.7				
Propane	15.4	15.9			
n-Butane	13.0	24.8			
2-Methylpropane	24.2	25.0			
n-Pentane	13.5	22.4	34.3		
2-Methylbutane	21.8	29.7	31.6	11.3	
2,2-Dimethylpropane	31.3	27.7			
n-Hexane	13.7	22.7	31.8		
2-Methylpentane	22.33	27.6	41.6	20.5	13.9
3-Methylpentane	11.1	29.1	36.5	18.4	
2,2-Dimethylbutane	28.7	30.2	36.5	8.5	
2,3-Dimethylbutane	19.2	33.9			
n-Heptane	13.8	22.8	32.2	29.3	

Original data are referenced to neat benzene. Values here were adjusted to TMS using the formula $\delta_{TMS} = 128.5 - \delta_{benzene}$.

Table 10 Substituent effects on ^{13}C chemical shifts ($\Delta\delta_C$) in 1-substituted pentanes. (Reproduced by permission of John Wiley & Sons Limited from R.J. Abraham, J. Fisher, P. Loftus, *Introduction to NMR Spectroscopy*, John Wiley & Sons, New York, 1995.)

n-Pentane	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C5 (ppm)
	13.7	22.6	34.5	22.6	13.7
Substituent	Substituent effect, $\Delta\delta$ (ppm)				
	α	β	γ	δ	ε
F	70.1	8.0	−6.7	0.1	0.0
Cl	30.6	10.0	−5.3	−0.5	−0.1
Br	19.3	10.1	−4.1	−0.7	0.0
I	−7.4	10.5	−2.1	−1.1	−0.1
CH_3	9.3	9.4	−2.5	0.4	0.2
NH_2	29.7	11.2	−5.0	0.1	0.0
OH	48.3	10.1	−6.0	0.3	0.2
CHO	31.4	0.7	−1.9	0.8	0.5
$COCH_3$	30.7	2.1	−1.2	1.4	1.2
COOH	20.7	2.3	−2.7	0.2	0.3
C≡N	3.7	3.2	−2.9	−0.4	−0.8
C≡CH	5.0	5.8	−3.0	0.4	–
$CH{=}CH_2$	20.3	6.2	−2.8	0.0	−0.1

2.3.1 Substituted n-*Pentane*

The change in the chemical shift of *n*-pentane caused by substitution at C1 is useful for calculating chemical shifts in substituted alkanes. Table 10 lists the change in the chemical shift for 13 different 1-substituted pentanes given by Abraham et al.[6] The values in Table 10 are added to the chemical shift of the alkane backbone using the shift for C1 for the α-carbon at the point of substitution, the shift for C2 for the β-carbon adjacent to the point of substitution, and continuing through the molecule. As shown in Table 10, the substituent effect drops off strongly after C3, the γ-carbon. 1-Hexanol (111-27-3) **(3)** is used in Table 11 as an example of how these values are used to calculate the chemical shift of a substituted alkane.

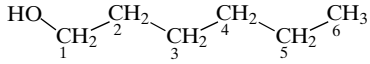

(3)

Table 11 Calculated ^{13}C shifts for 1-hexanol using substituent effects on 1-substituted pentanes

	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C5 (ppm)	C6 (ppm)
n-Hexane[3]	13.7	22.7	31.8	31.8	22.7	13.7
OH position	α	β	γ	δ	ε	–
OH ($\Delta\delta$)	48.3	10.1	−6.0	0.3	0.2	0
Calculated shift	62.0	32.8	25.8	32.1	22.9	13.7
Experimental shift[4]	62.1	32.9	25.8	31.9	23.0	14.2

Table 12 Substituent effects on ^{13}C shifts with terminal and internal substitution. (Reproduced by permission of John Wiley & Sons Limited from F.W. Wehrli, A.P. Marchand, S. Wehrli, *Interpretation of Carbon-13 NMR Spectra*, 2nd edition, John Wiley & Sons, New York, 1988.)

Substituent	α		β		γ (ppm)
	Normal (ppm)	Iso (ppm)	Normal (ppm)	Iso (ppm)	
CH_3	9	6	10	8	−2
COOH	21	16	3	2	−2
COO^-	25	20	5	3	−2
COOR	20	17	3	2	−2
COCl	33	28		2	
COR	30	24	1	1	−2
CHO	31		0		−2
Phenyl	23	17	9	7	−2
OH	48	41	10	8	−5
OR	58	51	8	5	−4
OCOR	51	45	6	5	−3
NH_2	29	24	11	10	−5
NH_3^+	26	24	8	6	−5
NHR	37	31	8	6	−4
NR_3	42		6		−3
NO_2	63	57	4	4	
CN	4	1	3	3	−3
SH	11	11	12	11	−4
SR	20		7		−3
F	68	63	9	6	−4
Cl	31	32	11	10	−4
Br	20	25	11	10	−3
I	−6	4	11	12	−1

Table 13 Calculated ^{13}C shifts for 1-bromobutane and 2-bromobutane using shifts for terminal and internal substitution from Wehrli et al.[7]

	1-Bromobutane				2-Bromobutane			
	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)
n-Butane[3]	13.0	24.8	24.8	13.0	13.0	24.8	24.8	13.0
Br position (*n*-)	α	β	γ		β	α	β	γ
Br substituent effect	20	11	−3		10	25	10	−3
Calculated shift	33.0	35.8	21.8	13.0	23.0	49.8	34.8	10.0
Experimental shift[8]	33.6	34.8	21.4	13.2	26.0	53.5	34.2	12.2

2.3.2 Terminal and Internal Substitution

Because the effect of internal substitution is slightly different to the effect of terminal substitution, Wehrli et al.[7] provide substituent effects for α, β, and γ carbons with either terminal or internal substitution. These values, given in Table 12, are based upon average chemical shifts for a number of alkanes, so they differ slightly from the *n*-pentane values. The ^{13}C chemical shifts for 1-bromobutane (109-65-9) **(4)** and 2-bromobutane (78-76-2) **(5)** are calculated in Table 13. The experimental data for these compounds are from Aldrich.[8]

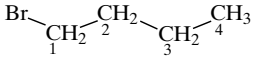

(4)

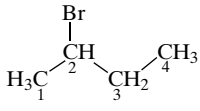

(5)

2.4 Substituted Alkenes

The chemical shift of ethylene, 123.3 ppm, is used as the basis for calculating the ^{13}C chemical shifts of alkene carbons using Equation (6):

$$\delta = 123.3\,\text{ppm} + \sum A_{ki} + \sum A'_{ki} + \sum \text{correction} \quad (6)$$

The effect of substitution on the chemical shift of ethylene is given in Table 14 with data from Wehrli et al.[7] The data used by Wehrli for acyclic alkenes are derived from Dorman et al.,[9] where the predicted chemical shift for acyclic alkenes has a standard error of ±0.9 ppm. The chemical shift calculated using Equation (6) and Table 14 is for the C* carbon. Effects for α, β and γ are with substitution on the same side of the double bond; effects for α', β' and γ' are with substitution across the double bond; and the correction factors account for geometry across the double bond. Table 15 demonstrates the use of these data to calculate the ^{13}C chemical shifts for 3-methyl-2-butene-1-ol (556-82-1) **(6)**. The experimental chemical shifts in Table 15 are from Aldrich.[8]

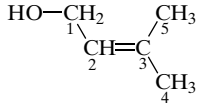

(6)

Table 14 Alkene chemical shifts, $\Delta\delta$, with α, β, and γ substitution at C* for use with Equation (6). (Reproduced by permission of John Wiley & Sons Limited from F.W. Wehrli, A.P. Marchand, S. Wehrli, *Interpretation of Carbon-13 NMR Spectra*, 2nd edition, John Wiley & Sons, New York, 1988.)

Substituent	Parameters A'_{ki} (ppm)			Parameters A_{ki} (ppm)		
	C- - - -C- - - -C- - - -C=			C*- - - -C- - - -C- - - -C		
	γ'	β'	α'	α	β	γ
C	1.5	−1.8	−7.9	10.6	7.2	−1.5
OH		−1			6	
OR		−1	−39	29	2	
OAc			−27	18		
$COCH_3$			6	15		
CHO			9	4		
COOR			7	6		
CN			15	−16		
Cl		2	−6	3	−1	
Br		2	−1	−8	∼0	
I			7	−38		
C_6H_5			−11	12		

Correction terms

Geometry	Correction (ppm)
α, α' (*trans*)	0
α, α' (*cis*)	−1.1
α, α	−4.8
α', α'	2.5
β, β	2.3
All other interactions	∼0

2.5 Substituted Benzenes

Data for the ^{13}C chemical shifts of monosubstituted benzene are given in Table 16. This listing from Silverstein and Webster[10] is derived from a much more extensive compilation by Ewing.[11] The values for the substituent effect, $\Delta\delta$, in Table 16 are used with Equation (7), where

Table 15 Calculated ^{13}C shifts for 3-methyl-2-butene-1-ol using alkene chemical shifts with α, β, and γ substitution and Equation (6)

	C2 (ppm)	C3 (ppm)
Ethene shift	123.3	123.3
OH position and $\Delta\delta$	β (6)	$\beta'(-1)$
C1 position and $\Delta\delta$	α (10.6)	$\alpha'(-7.9)$
C4 position and $\Delta\delta$	$\alpha'(-7.9)$	α (10.6)
C5 position and $\Delta\delta$	$\alpha'(-7.9)$	α (10.6)
Correction term		
C1 to C4	α, α' *trans* (0)	α, α' *trans* (0)
C1 to C5	α, α' *cis* (-1.1)	α, α' *cis* (-1.1)
C4 to C5	α', α' (2.5)	α, α (-4.8)
Calculated shift [Equation 6]	125.5	129.7
Experimental shift[8]	123.8	136.0

128.5 ppm is the ^{13}C chemical shift of benzene:

$$\delta = 128.5\,\text{ppm} + \Delta\delta \qquad (7)$$

The substituent effect is additive, but multiple substituents increase the uncertainty in the calculated chemical shift. Calculations for 3-hydroxybenzaldehyde (100-83-4) **(7)** and experimental data from Aldrich[8] are given in Table 17.

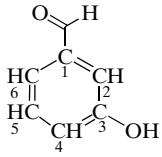

(7)

2.6 Cyclic Compounds

The ^{13}C chemical shifts have been determined for many cyclic compounds. Figure 2 gives the ^{13}C shifts for some of the most common cyclic compounds, from

Table 16 Carbon-13 shifts, $\Delta\delta$, for substituted benzenes. (Reproduced by permission of John Wiley & Sons Limited, Copyright 1998 from R. Silverstein, F. Webster, *Spectrometric Identification of Organic Compounds*, 6th edition, John Wiley & Sons, New York, 1998.)

Substituent	C1 (ppm)	C2 (*ortho*) (ppm)	C3 (*meta*) (ppm)	C4 (*para*) (ppm)	Substituent shift (ppm)
H	0.0	0.0	0.0	0.0	
CH_3	+9.3	+0.7	−0.1	−2.9	21.3
CH_2CH_3	+15.6	−0.5	0.0	−2.6	29.2 (CH_2)
					15.8 (CH_3)
$CH(CH_3)_2$	+20.1	−2.0	0.0	−2.5	34.4 (CH)
					24.1 (CH_3)
$C(CH_3)_3$	+22.2	−3.4	−0.4	−3.1	34.5 (C)
					31.4 (CH_3)
$CH{=}CH_2$	−9.1	−2.4	+0.2	−0.5	137.1 (CH)
					113.3 (CH_2)
C≡CH	−5.8	+6.9	+0.1	+0.4	84.0 (C)
					77.8 (CH)
C_6H_5	+12.1	−1.8	−0.1	−1.6	
CH_2OH	+13.3	−0.8	−0.6	−0.4	64.5
CH_2OCOCH_3	+7.7	0.0	0.0	0.0	20.7 (CH_3)
					66.1 (CH_2)
					170.5 (C=O)
OH	+26.6	−12.7	+1.6	−7.3	
OCH_3	+31.4	−14.4	+1.0	−7.7	54.1
OC_6H_5	+29.0	−9.4	+1.6	−5.3	
$OCOCH_3$	+22.4	−7.1	−0.4	−3.2	23.9 (CH_3)
					169.7 (C=O)
COH	+8.2	+1.2	+0.6	+5.8	192.0
$COCH_3$	+7.8	−0.4	−0.4	+2.8	24.6 (CH_3)
					195.7 (C=O)
COC_6H_5	+9.1	+1.5	−0.2	+3.8	196.4 (C=O)
$COCF_3$	−5.6	+1.8	+0.7	+6.7	
COOH	+2.9	+1.3	+0.4	+4.3	168.0
$COOCH_3$	+2.0	+1.2	−0.1	+4.8	51.0 (CH_3)
					166.8 (C=O)
COCl	+4.6	+2.9	+0.6	+7.0	168.5
C≡N	−16.0	+3.6	+0.6	+4.3	119.5

(*continued overleaf*)

Table 16 (*continued*)

Substituent	C1 (ppm)	C2 (*ortho*) (ppm)	C3 (*meta*) (ppm)	C4 (*para*) (ppm)	Substituent shift (ppm)
NH_2	+19.2	−12.4	+1.3	−9.5	
$N(CH_3)_2$	+22.4	−15.7	+0.8	−11.8	40.3
$NHCOCH_3$	+11.1	−9.9	+0.2	−5.6	
NO_2	+19.6	−5.3	+0.9	+6.0	
N=C=O	+5.7	−3.6	+1.2	−2.8	129.5
F	+35.1	−14.3	+0.9	−4.5	
Cl	+6.4	+0.2	+1.0	−2.0	
Br	−5.4	+3.4	+2.2	−1.0	
I	−32.2	+9.9	+2.6	−7.3	
CF_3	+2.6	−3.1	+0.4	+3.4	
SH	+2.3	+0.6	+0.2	−3.3	
SCH_3	+10.2	−1.8	+0.4	−3.6	15.9
SO_2NH_2	+15.3	−2.9	+0.4	+3.3	
$Si(CH_3)_3$	+13.4	+4.4	−1.1	−1.1	

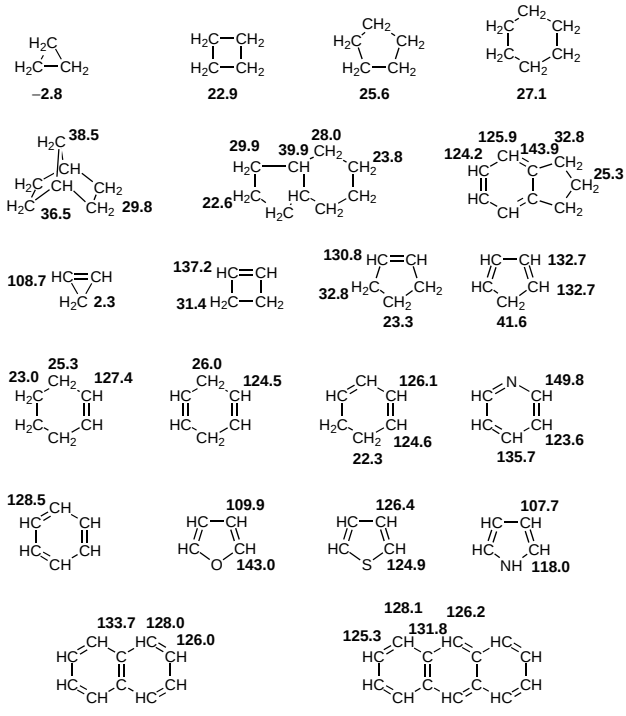

Figure 2 Carbon-13 shifts of cyclic compounds.

Pretsch et al.[4] These values may be used with the additivity values given for alkanes, alkenes, and benzene to calculate the chemical shifts for many substituted cyclic compounds. Figure 2 provides a useful starting point but for more accurate calculations with these compounds other resources should be consulted.

3 PROTON CHEMICAL SHIFTS

Calculations of proton (^{1}H) chemical shifts are usually less accurate than the calculations for ^{13}C chemical shifts.

Table 17 Calculated ^{13}C shifts for 3-hydroxybenzaldehyde using Equation (7)

Effect	C1 (ppm)	C2 (ppm)	C3 (ppm)	C4 (ppm)	C5 (ppm)	C6 (ppm)
Benzene	128.5	128.5	128.5	128.5	128.5	128.5
COH position	1	*Ortho*	*Meta*	*Para*	*Meta*	*Ortho*
COH shift, Δδ	8.2	1.2	0.6	5.8	0.6	1.2
OH position	*Meta*	*Ortho*	1	*Ortho*	*Meta*	*Para*
OH shift, Δδ	1.6	−12.7	26.6	−12.7	1.6	−7.3
Calculated shift [Equation 7]	138.3	117.0	155.7	121.6	130.7	122.4
Experimental shift[8]	137.6	114.8	158.1	121.0	129.9	121.8

Many of the rules given here result in a standard error of ±0.1 ppm for the set of compounds used to produce the rule. For compounds not included in the data set, errors of ±0.5 ppm are typical.

In ^{1}H-NMR, the substituent effects are frequently called Shoolery coefficients or deshielding values. Shoolery's original listing of ten values in a Varian technical bulletin has been modified by many subsequent authors who optimized the coefficients with different sets of compounds. The first two methods given below are fairly general and easy to apply, but are the least accurate. The other methods provide improved accuracy for a smaller range of proton types. For the most accurate results, find the method that most closely matches the proton of interest.

3.1 Correlation Table

Figure 3 is a correlation table for proton chemical shifts using data from Breitmaier.[12] This provides an overview of proton chemical shifts for various functional groups. The additivity rules provide shifts for additional functional groups and improved accuracy.

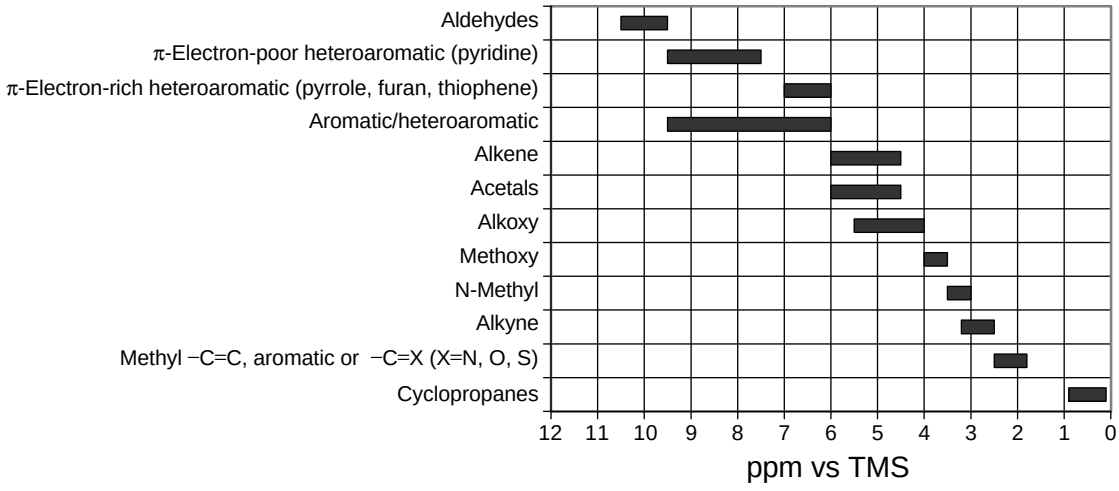

Figure 3 Proton chemical shift correlation. (Reproduced by permission of John Wiley & Sons Limited from E. Breitmaier, *Structure Elucidation by NMR in Organic Chemistry, A Practical Guide*, John Wiley & Sons, New York, 1995.)

Table 18 Friedrich brief proton chemical shifts. (Reproduced by permission of Division of Chemical Education, Inc. Copyright 1984 from E. Friedrich, K. Runkle, *J. Chem. Educ.*, **61**, 830–832 (1984).)

Attached atom	Constant (ppm)
H	0.5
C (sp^3)	0.9
C (sp^2 or sp)	1.6
O (all)	2.8
N (except $O_2N = 3.5$)	1.9
S (except $O_2S = 2.6$)	1.8
F	3.4
Cl, Br, I	2.5

Table 19 Calculated proton shifts for chloroethane using Friedrich brief proton chemical shifts

Substituent on C	H1 (ppm)	H2 (ppm)
Attached atom (shift)	H (0.5)	H (0.5)
	C (0.9)	H (0.5)
	Cl (2.5)	C (0.9)
Calculated shift	3.9	1.9
Experimental shift[4]	3.47	1.33

3.2 Substituted Alkanes

3.2.1 Friedich Brief Proton Chemical Shifts[13]

This method provides quick estimates for methyl (CH_3) and methylene (CH_2) protons. The appropriate value from Table 18 for each substituent on the carbon is added to calculate the proton chemical shift. For most common functional groups, the results are generally within ±0.5 ppm. Table 19 shows the proton chemical shift calculations for chloroethane (75-0-3).

Table 20 Jeffreys proton chemical shifts. (Reproduced by permission of Division of Chemical Education, Inc. Copyright 1979 from J.A.D. Jeffreys, *J. Chem. Educ.*, **56**, 806–809 (1979).)

Substituent	Attached	Deshielding value (ppm)
Elements attached to carbon	B, Si, Sn	−0.3
	H	0.2
	C, P	0.7
	N, S	1.7
	O, Cl, Br, I	2.7
	F	3.7
Double bonds	On carbon (i.e. H−C=C)	4.2
	On adjacent carbon (i.e. H−C−C=C)	0.8
Benzene ring (also pyridine and diazines)	Carbon in ring	6.0
	Carbon attached to ring	1.2
Triple bond	On carbon (i.e. H−C≡C)	2.1
	On adjacent carbon (i.e. H−C−C≡C)	0.8
Rings	Four-membered ring	0.7
	Four-membered ring	−0.7

3.2.2 Jeffreys Proton Chemical Shifts[14]

This is another quick method to calculate the chemical shift of protons attached to a carbon atom. It is a more general method than the Friedrich brief proton method because it works for protons on any type of carbon. The deshielding values for various carbon substituents are given in Table 20. These values are added together to calculate the predicted chemical shift

Table 21 Calculated proton shifts for 3-methyl-2-buten-1-ol using Jeffreys proton chemical shifts

Proton	H1 (ppm)	H2 (ppm)	H4 (ppm)	H5 (ppm)
Substituent attached to α-C	−C−C=C (0.8) O (2.7) H (0.2)	−C=C (4.2) C (0.7)	−C−C=C (0.8) H (0.2) H (0.2)	−C−C=C (0.8) H (0.2) H (0.2)
Calculated shift	3.7	4.9	1.2	1.2
Experimental shift[8]	4.1	5.4	1.7	1.7

Table 22 Friedrich detailed proton chemical shifts. (Reproduced by permission of Division of Chemical Education, Inc. Copyright 1984 from E. Friedrich, K. Runkle, *J. Chem. Educ.*, **61**, 830–832 (1984).)

Substituent	σ (ppm)	Substituent	σ (ppm)	Substituent	σ (ppm)
H_3C	0.68	R_3Sn	−0.45	S=C=N	2.62
R	0.58	HO	2.56	N_3	1.97
ZCH_2 (Z = Cl or Br)	0.91	RO	2.36	R(H)S	1.64
Cyclopropane	0.66	C_6H_5O	2.94	ArS	1.90
Oxirane	0.77	R(H)C(=O)O	3.01	RSS	1.72
F_3C	1.14	$C_6H_5C(=O)O$	3.27	$CH_3C(=O)S$	1.94
HF_2C	1.12	R(Ar)NHC(=O)O	3.16	$C_6H_5C(=O)S$	2.16
Cl_3C	1.55	$F_3CC(=O)O$	3.35	$R_2NC(=O)S$	1.97
Br_3C	1.92	RSO_2O	3.13	N≡C−S	2.04
R(H)C(=O)	1.50	$ArSO_2O$	3.06	R(Ar)S(=O)	1.74
R(H)OC(=O)	1.46	N≡C−O	3.57	$R(Ar)SO_2$	2.45
$C_6H_5C(=O)$	1.90	$R_2(H_2)N$	1.57	H_2NSO_2	2.29
$C_6H_5OC(=O)$	1.50	C_6H_5NH	2.04	$ClSO_2$	2.86
$R_2(H_2)NC(=O)$	1.47	R_3N^+	2.55	$R_2P(=O)$	0.98
$C_6H_5NHC(=O)$	1.45	R(H)C(=O)NH	2.27	$(RO)_2P(=O)$	1.09
ClC(=O)	1.84	$C_6H_5C(=O)NH$	2.43	$R_2P(=S)$	0.83
C_6H_5	1.83	ROC(=O)NH	2.25	H	0.34
$R_2(H_2)C=CR(H)$	1.32	RSC(=O)NH	2.55	F	3.30
R(H)C≡C	1.44	$ArSO_2NH$	1.98	Cl	2.53
N≡C	1.59	O_2N	3.36	Br	2.33
$(CH_3)_3Si$	0.03	O=C=N	2.36	I	2.19

for the proton of interest. Results are generally within ±0.3 ppm; however, aldehyde protons give poor results. Table 21 gives the calculated chemical shifts for 3-methyl-2-butene-1-ol **(6)**, with experimental data from Aldrich[8] for comparison.

3.2.3 Friedich Detailed Proton Chemical Shifts[13]

This method uses a more extensive table of deshielding values to calculate more accurately the proton chemical shift. This method is only designed for use with methyl (CH_3) or methylene (CH_2) protons, but it should produce results within ±0.1 ppm of the experimental value. The chemical shift is calculated using Equation (8):

$$\delta_{XCH_2Y} = 0.23\,\text{ppm} + \sigma_x + \sigma_y \qquad (8)$$

The deshielding values, σ_x and σ_y, for substituent X and Y, respectively, are given in Table 22. Table 23 illustrates the use of this rule to calculate the proton chemical shifts of chloroethane.

3.2.4 Bell α-Substituted Methylene Protons[15]

This method is specifically for α-substituted methylene protons. The deshielding values for this method were developed using a data set of 395 compounds containing two α functional groups. The method produces a standard error of ±0.14 ppm for this data set. For compounds with a single α-substituent, the error increased to ±0.18 ppm. This method is very similar to the Friedich detailed proton method but the deshielding values were only optimized for methylene protons. As a result, the deshielding values are slightly different. For this rule, proton chemical shifts are calculated using Equation (9) for compounds $X-CH_2-Y$:

$$\delta = 0.23\,\text{ppm} + \sum \alpha \qquad (9)$$

Table 23 Calculated proton shifts for chloroethane using Friedrich detailed proton chemical shifts (Equation 8)

Substituent	H1 (ppm)	H2 (ppm)
CH_2	0.23	0.23
δ_x (substituent, shift)	C (0.68)	C (0.68)
δ_y (substituent, shift)	Cl (2.53)	H (0.34)
Calculated shift	3.44	1.25
Experimental shift[4]	3.47	1.33

Substituents X and Y each have a deshielding value, α, given in Table 24. Based upon these values, the chemical shift for the protons on dichloromethane (75-09-2) is $0.23 + 2.48 + 2.48 = 5.19$. This compares to the experimental value of 5.33 given by Pretsch et al.[4]

3.2.5 Bell α- and β-Substituted Methylene Protons[15]

This method is based upon a data set containing 1007 compounds with methylene protons having one α-substituent and one or more β-substituents. The standard error using Equation (10) and the deshielding values in Table 25 with the 1007 compounds in this data set was ± 0.15 ppm.

$$\delta = 1.20\,\text{ppm} + \sum \alpha + \sum \beta \qquad (10)$$

The rule is applied to compounds w$-CH_2-$C$-$xyz, where w is the α-substituent and x, y, and z are β-substituents.

Table 24 Bell α-substituted methylene proton chemical shifts. (Reproduced by permission of John Wiley and Sons Limited from H. Bell, L. Berry, E. Madigan, *Org. Magn. Reson.*, **22**, 693–696 (1984).)

Group	α-Shift (ppm)	Group	α-Shift (ppm)
−F	3.15	$-NR_2$•HX	2.46[a]
−Cl	2.48	$-NR_3^+$	2.56[b]
−Br	2.29	−NRCOR′, −NHCOR	2.23
−OH	2.46	−NHCOAr	2.33
−OR	2.27	−COOR, −COOH	1.49
−OAr	2.89	−COOAr	1.74
−OCOR	2.98	−COR	1.58
−OCOAr	3.23	−COAr	2.08
−SR, −SH	1.63	$-CONR_2$, $-CONH_2$	1.39
−SAr	1.92	−CONHAr	1.59
$-NH_2$	1.69	−C=C	1.33
−NHR	1.60	−Ar	1.85
$-NR_2$	1.41	−C≡CR, −C≡CH	1.52
−NHAr	2.15	−C≡C−Ar	1.77
−NRAr	2.39	−C≡N	1.73
$-NH_2$•HX, −NHR•HX	2.31[a]		

[a] Amine salt shifts are for D_2O; in DMSO-d_6 or DMSO-d_6/$CDCl_3$, shifts are smaller by 0.2–0.3 ppm.
[b] This value is for D_2O solutions; in $CDCl_3$, the shift is larger by 0.3–0.5 ppm.

Using this method to calculate the chemical shift for 1,2-dichloroethane (107-06-2) gives the proton chemical shift of $1.20 + 2.04 + 0.36 = 3.62$. Pretsch et al.[4] give an experimental value of 3.27 ppm.

Table 25 Bell α- and β-substituted methylene proton chemical shifts. (Reproduced by permission of John Wiley & Sons Ltd from H. Bell, L. Berry, E. Madigan, *Org. Magn. Reson.*, **22**, 693–696 (1984).)

Group	α-Shift (ppm)	β-Shift (ppm)	Group	α-Shift (ppm)	β-Shift (ppm)
−F	2.76	0.24	−NRAr	1.97	0.31
−Cl	2.04	0.36	$-NH_2$•HX, −NHR•HX	1.77[b]	0.60[b]
−Br	1.87	0.52	$-NR_2$•HX	1.95[b]	0.60[b]
−OH	2.07	0.23	$-NR_3^+$	2.07[c]	0.70
−OR	1.91	0.28[a]	−NRCOR′, −NHCOR	1.79[d]	0.26
−OAr	2.48	0.58	−COOR, −COOH	1.02[e]	0.34
−OCOR	2.61	0.43	−COR	1.11	0.38
−OCOAr	2.91	0.53	−COAr	1.61	0.58
−SR, −SH	0.15	0.30	$-CONR_2$, $-CONH_2$	0.89[f]	0.40
−SAR	1.44	0.28	−C=C	0.78	0.16
$-NH_2$	1.27	0.25	−Ar	1.31	0.37
−NHR	1.23	0.25	−C≡CR, −C≡CH	0.98[g]	0.21
$-NR_2$	1.00	0.25	−C≡N	1.15	0.42
−NHAr	1.74	0.31	−R	0.00	0.01

[a] The second OR group in acetals and ketals has a β-shift of zero.
[b] Listed shifts are for D_2O; in DMSO or DMSO/$CDCl_3$, α-shifts are lowered by 0.2–0.3 ppm and β-shifts are 0.05–0.10 ppm lower.
[c] D_2O solvent. In $CDCl_3$ the value is 0.3–0.5 ppm higher.
[d] For α-NHCOAr, add 0.1 ppm.
[e] Shift is 1.27 ppm for α-COOAr.
[f] Add 0.2 ppm for α-CONHAr.
[g] Shift is 1.23 ppm for α-C≡CAr.

Table 26 Substituent effects on chemical shift of protons on sp^3 carbons. (Reproduced by permission of Division of Chemical Education, Inc. Copyright 1997 from P. Beauchamp, R. Marquez, *J. Chem. Educ.*, **74**, 1483–1485 (1997).)

Substituent	α (ppm)	β (ppm)	γ (ppm)
R–	0.0	0.0	0.0
$R_2C{=}CR-$	0.8	0.2	0.1
$RC{\equiv}C-$	0.9	0.3	0.1
Ar–	1.4	0.4	0.1
F–	3.2	0.5	0.2
Cl–	2.2	0.5	0.2
I–	2.0	0.9	0.1
HO–	2.3	0.3	0.1
R–	2.1	0.3	0.1
$R_2C{=}CRO-$	2.5	0.4	0.2
ArO–	2.8	0.5	0.3
RCO_2-	2.8	0.5	0.1
$ArCO_2-$	3.1	0.5	0.2
$ArSO_3-$	2.8	0.4	0.0
H_2N-	1.5	0.2	0.1
RCONH–	2.1	0.3	0.1
O_2N-	3.2	0.8	0.1
HS–	1.3	0.4	0.1
RS–	1.3	0.4	0.1
OHC–	1.1	0.4	0.1
RCO–	1.2	0.3	0.0
ArCO–	1.7	0.3	0.1
HO_2C-	1.1	0.3	0.1
RO_2C-	1.1	0.3	0.1
H_2NOC-	1.0	0.3	0.1
ClOC–	1.8	0.4	0.1
$N{\equiv}C-$	1.1	0.4	0.2
RSO–	1.6	0.5	0.3
RSO_2-	1.8	0.5	0.3

Table 27 Calculated proton shifts for 2-methyl-1-propanol using Beauchamp's substituent effects (Equation 11)

Type	H1 Methylene, CH_2 (ppm)	H2 Methine, CH (ppm)	H3 Methylene, CH_3 (ppm)
Shift	1.2	1.5	0.9
OH position	α	β	γ
OH shift	2.3	0.3	0.1
Calculated shift [Equation 11]	3.5	1.8	1.0
Experimental shift[8]	3.4	1.8	0.9

3.2.6 Beauchamp Methyl, Methylene and Methine Protons[16]

This is a more general method for protons on sp^3-hybridized carbons. The method accounts for one or more substituents located within three carbons and uses a different constant for each type of proton. The values used for this method were obtained using substituted methyl, ethyl, propyl, isopropyl, and *t*-butyl compounds. The deshielding values in Table 26 are used with the appropriate constant in Equation (11):

$$\begin{aligned}\text{Methyl } (CH_3) \text{ protons } \delta &= 0.9\,\text{ppm} + \sum\alpha \\ &\quad + \sum\beta + \sum\gamma \\ \text{Methylene } (CH_2) \text{ protons } \delta &= 1.2\,\text{ppm} + \sum\alpha \\ &\quad + \sum\beta + \sum\gamma \\ \text{Methine } (CH) \text{ protons } \delta &= 1.5\,\text{ppm} + \sum\alpha \\ &\quad + \sum\beta + \sum\gamma \qquad (11)\end{aligned}$$

A reliability of ±0.5 ppm should be expected when using Equation (11) and Table 26. Table 27 illustrates these calculations for 2-methyl-1-propanol (78-83-1) **(8)**, with experimental data from Aldrich.[8]

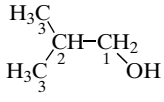

(8)

3.3 Substituted Alkenes

The calculation of proton chemical shifts for alkenes was studied by Pascual et al.[17] Using Equation (12) and the deshielding values in Table 28 with a data set of 1070 measured chemical shifts, substituent effects were calculated with a standard error of ±0.15 ppm.

$$\delta = 5.28\,\text{ppm} + \sum Z_i \qquad (12)$$

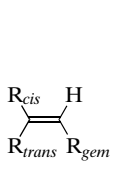

(9)

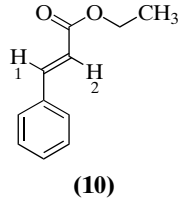

(10)

The deshielding effect differs for the substituents in the *gem*, *cis*, and *trans* positions of an alkene **(9)**. The proton chemical shift of ethylene is 5.28 ppm and Z_i is the substituent effect from Table 28 for the position from the proton of interest. Table 29 shows the calculations for the chemical shifts of the alkene protons in ethyl-*trans*-cinnamate (103-36-6) **(10)**, with experimental data from Aldrich.[8]

Table 28 Substituent effects on proton shifts of alkenes. (Reproduced by permission of Wiley–VCH Verlag GmbH from C. Pascual, J. Meier, W. Simon, *Helv. Chim. Acta*, **49**, 164–168 (1966).)

Substituent (R)	Z_i *gem* (ppm)	Z_i *cis* (ppm)	Z_i *trans* (ppm)
−H	0	0	0
Alkyl	0.44	−0.26	−0.29
Alkyl-ring	0.71	−0.33	−0.30
$-CH_2O$, $-CH_2I$	0.67	−0.02	−0.07
$-CH_2S$	0.53	−0.15	−0.15
$-CH_2Cl$, $-CH_2Br$	0.72	0.12	0.07
$-CH_2N$	0.66	−0.05	−0.23
−C≡C	0.50	0.35	0.10
−C≡N	0.23	0.78	0.58
−C=C	0.98	−0.04	−0.21
−C=C (conj.)[a]	1.26	0.08	−0.01
−C=O	1.10	1.13	0.81
−C=O (conj.)[a]	1.06	1.01	0.95
−COOH	1.00	1.35	0.74
−COOH (conj.)[a]	0.69	0.97	0.39
−COOR	0.84	1.15	0.56
−COOR (conj.)[a]	0.68	1.02	0.33
−COH	1.03	0.97	1.21
−CON	1.37	0.93	0.35
−COCl	1.10	1.41	0.99
−OR (R aliph.)	1.18	−1.06	−1.28
−OR (R conj.)[a]	1.14	−0.65	−1.05
−OCOR	2.09	−0.40	−0.67
−Ar	1.35	0.37	−1.10
−Cl	1.00	0.19	0.03
−Br	1.04	0.40	0.55
−NRR′ (R aliph.)	0.69	−1.19	−1.31
−NRR′ (R conj.)[a]	2.30	−0.73	−0.81
−SR	1.00	−0.24	−0.04
$-SO_2$	1.58	1.15	0.95

[a] The increment for R conj. is used when the substituent R or the double bond is conjugated.

Table 29 Calculated proton shifts for ethyl *trans*-cinnamate using Pascual's substituent effects on proton shifts of alkenes (Equation 12)

Position	H1 (ppm)	H2 (ppm)
Ethene	5.28	5.28
Gem substituent (Δδ)	−Ar (1.35)	−COOR (0.84)
Cis substituent (Δδ)	−COOR (1.15)	−Ar (0.37)
Trans substituent (Δδ)	−H (0)	−H(0)
Calculated shift	7.78	6.49
Experimental shift[8]	7.7	6.4

3.4 Substituted Benzenes

The substituent effects on the proton chemical shifts of benzene are given in Table 30.[6] These values are used with Equation (13):

$$\delta = 7.262\,\text{ppm} + \sum \Delta\delta_H \qquad (13)$$

Table 30 Proton substituent chemical shifts (Δδ) in benzenes (δ_H 7.262). (Reproduced by permission of John Wiley & Sons Limited from R.J. Abraham, J. Fisher, P. Loftus, *Introduction to NMR Spectroscopy*, John Wiley & Sons, New York, 1995.)

Substituent	$\Delta\delta_H$ (ppm)		
	Ortho	*Meta*	*Para*
NO_2	0.95	0.26	0.38
CO•OCH_3	0.71	0.11	0.21
CO•CH_3	0.62	0.14	0.21
CHO	0.56	0.22	0.29
CN	0.36	0.18	0.28
F	−0.29	−0.02	−0.23
Cl	0.03	−0.02	−0.09
Br	0.18	−0.08	−0.04
I	0.39	−0.21	0.00
OH	−0.56	−0.12	−0.45
OCH_3	−0.48	−0.09	−0.44
O•CO•CH_3	−0.25	0.03	−0.13
CH_3	−0.20	−0.12	−0.22
NH_2	−0.75	−0.25	−0.65
NMe_2	−0.66	−0.18	−0.67
CO•Ph	0.53	0.20	0.31
SO_2•Ph	0.69	0.26	0.27

Table 31 Calculated proton shifts for *p*-methylanisole (Equation 13)

Substituent	H1 (ppm)	H2 (ppm)
Benzene	7.262	7.262
OCH_3 (position, shift)	*Ortho*, −0.48	*Meta*, −0.09
CH_3 (position, shift)	*Meta*, −0.12	*Ortho*, −0.20
Calculated shift	6.66	6.97
Experimental shift[8]	6.8	7.1

According to Martin and Dailey,[18] the proton chemical shift for *ortho* disubstituted benzenes is not readily correlated, but for most other substituted benzenes the additivity rules provide accuracies of ±0.1 ppm. Table 31 shows the chemical shift calculations for *p*-methyl anisole (104-93-8) **(11)**, with experimental results from Aldrich.[8]

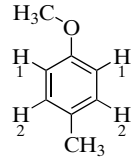

(11)

3.5 Cyclic Compounds

Proton chemical shifts given by Pretsch et al.[4] for a number of cyclic compounds are shown in Figure 4.

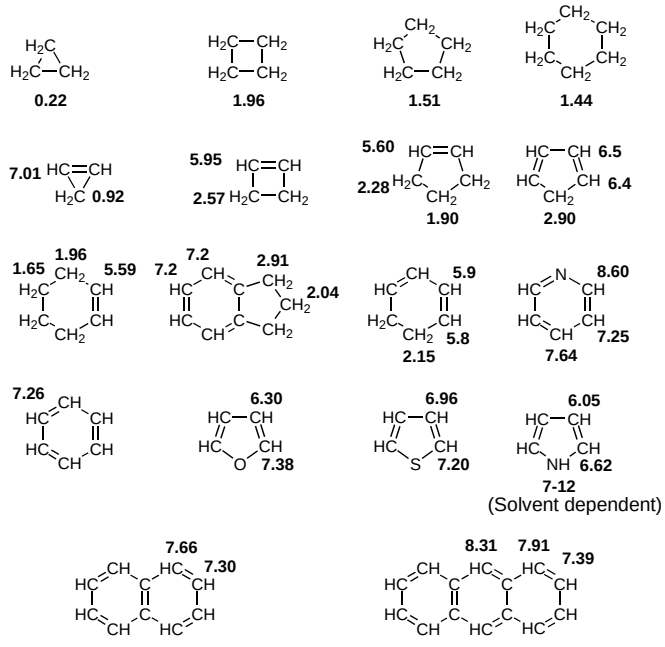

Figure 4 Proton shifts of cyclic compounds.

Although certainly not exhaustive, this provides proton chemical shifts for many common cyclic structures.

3.6 Heteroatoms

Figure 5 is a correlation chart from Silverstein and Webster[10] that shows the expected range for protons on heteroatoms. Many protons on heteroatoms are exchangeable on the NMR timescale and may not be observed in the spectrum. If these protons are observed, wide line widths are common and many chemical shifts show strong concentration, solvent, and pH effects. As a result, these chemical shifts are difficult to predict, and calculating the chemical shifts for protons on heteroatoms is not generally useful.

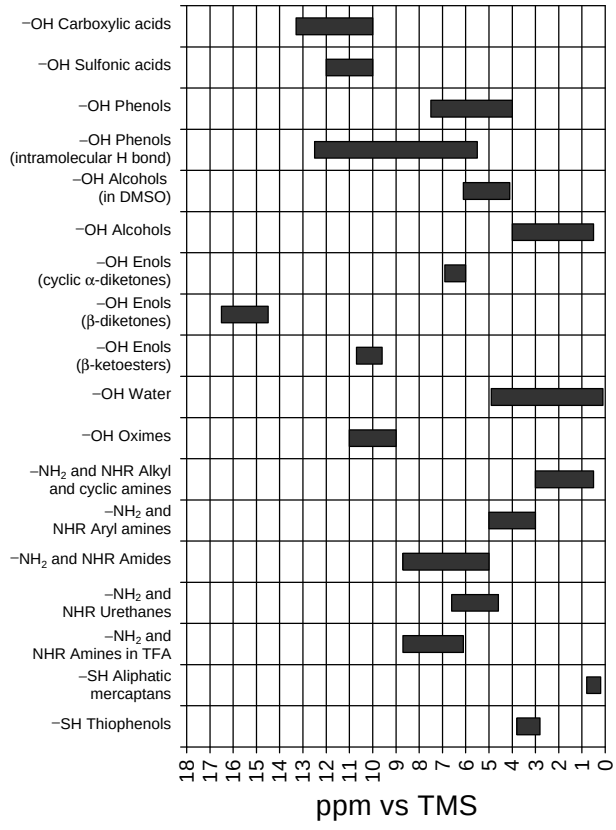

Figure 5 Chemical shifts of protons on heteroatoms. (Reproduced by permission of John Wiley & Sons Limited, Copyright 1998, from R. Silverstein, F. Webster, *Spectrometric Identification of Organic Compounds*, 6th edition, John Wiley & Sons, New York, 1998.)

4 COUPLING CONSTANTS

Although this article does not discuss coupling in NMR spectra, coupling constants are extremely useful in interpreting NMR spectra. Table 32 provides a listing of proton–proton and carbon–proton coupling constants from Pretsch et al.[4]

Table 32 Coupling constants[4]

Structure	Coupling (Hz)
CH_4	$^1J_{CH} = 125$
$CH_2{=}CH_2$	$^1J_{CH} = 156$
HC≡CH	$^1J_{CH} = 249$
H–C–C–H	$^3J_{HH} = 7$ (free rotating)
	$^3J_{HH} = 0–18$ (rigid)
H–C=C–H	$^2J_{HH} = 2.5$ (*gem*)
	$^3J_{HH} = 11.6$ (*cis*)
	$^3J_{HH} = 19.6$ (*trans*)
	Dependent upon electronegativity of substituent, *trans* > *cis*
Substituted benzene	$^3J_{HH} = 6–9$ (*ortho*)
	$^4J_{HH} = 1–3$ (*meta*)
	$^5J_{HH} = 0–1$ (*para*)

5 COMPUTER SOFTWARE

Although a detailed review of software is not within the scope of this article, there are software applications that perform many of the calculations discussed here. For routine use, these programs greatly simplify the calculation of chemical shifts. Most of these programs can

be divided into one of three distinct categories: chemical shift databases; calculated chemical shifts using additivity rules; and calculated chemical shifts using molecular modeling software.

Comparison with spectral databases is extremely useful for interpreting NMR spectra. The Aldrich NMR Library is available from Advanced Chemistry Development Inc., 133 Richmond Street West, Suite 1501, Toronto, Ontario, Canada M5H 2L3 and the Sadtler NMR Library is available from BioRad, Sadtler Group, 3316 Spring Garden Street, Philadelphia, PA 19104-2596, USA. Both databases are easy to use and contain thousands of experimental spectra. In addition to these two commercial packages, there are also several databases available on the Internet.

There are many software applications that calculate chemical shifts for a structure using databases and additivity rules. These programs include CNMR and HNMR from Advanced Chemistry Development Inc., 133 Richmond Street West, Suite 1501, Toronto, Ontario, Canada M5H 2L3; the C-13 module for ChemWindows from BioRad Sadtler Group, 3316 Spring Garden Street, Philadelphia, PA 19104-2596, USA; Win-SpecEdit from Bruker Instruments Inc., 19 Fortune Drive, Billerica, MA 01821, USA, tel. (508) 667-9580; and ChemDraw Ultra from CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140, USA.

It is also possible to calculate chemical shifts using molecular modeling software. These calculations are computationally very complex but offer the potential of calculating chemical shifts for novel compound classes that are not included in the current databases. These programs include an add-on for HyperChem from Hypercube, Inc., 1115 NW 4th Street, Gainesville, FL 32601, USA and Gaussian98 from Gaussian, Inc., Carnegie Office Park, Building 6, Suite 230, Carnegie, PA 15106, USA.

6 ADDITIONAL INFORMATION

Because the calculation of chemical shifts is so useful for the interpretation of NMR spectra, a considerable amount of information is available. Although this is not a thorough listing, some of the most useful texts are worth noting: the book by Silverstein and Webster[10] is an excellent starting point for information about ^{13}C and proton (^{1}H) chemical shifts; Pretsch et al.[4] provide an extensive set of tables for calculating chemical shifts of many different compound classes and data for many compounds; and Wehrli et al.[7] and Breitmaier and Voelter[2] provide detailed information about calculating ^{13}C chemical shifts.

ABBREVIATIONS AND ACRONYMS

NMR	Nuclear Magnetic Resonance
TMS	Tetramethylsilane

RELATED ARTICLES

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Scalar Couplings in Nuclear Magnetic Resonance, General • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules

REFERENCES

1. G. Levy, J. Cargioli, 'Carbon-13 Chemical Shifts on the TMS Scale', *J. Magn. Reson.*, **6**, 143–144 (1972).
2. E. Breitmaier, W. Voelter, *Carbon-13 NMR Spectroscopy, Methods and Applications in Organic Chemistry and Biochemistry*, VCH, Weinheim, Germany, 1990.
3. D. Grant, E. Paul, 'Carbon-13 Magnetic Resonance. II. Chemical Shift Data for the Alkanes', *J. Am. Chem. Soc.*, **86**, 2984–2990 (1964).
4. E. Pretsch, T. Clerc, J. Seibl, W. Simon, *Tables of Spectral Data for Structure Determination of Organic Compounds*, 2nd edition, Springer-Verlag, Berlin, 1989.
5. L. Lindeman, J. Adams, 'Carbon-13 Nuclear Magnetic Resonance Spectrometry Chemical Shifts for the Paraffins through C_9', *Anal. Chem.*, **43**(10), 1245–1252 (1971).
6. R.J. Abraham, J. Fisher, P. Loftus, *Introduction to NMR Spectroscopy*, John Wiley & Sons, New York, 1995.

7. F.W. Wehrli, A.P. Marchand, S. Wehrli, *Interpretation of Carbon-13 NMR Spectra*, 2nd edition, John Wiley & Sons, New York, 1988.
8. *Aldrich/ACD Library of FT NMR Spectra [CD-ROM]*, Advanced Chemistry Development, Inc., Toronto, Canada, 1998.
9. D. Dorman, M. Jautelat, J. Roberts, 'Carbon-13 Nuclear Magnetic Resonance Spectroscopy. Quantitative Correlations of the Carbon Chemical Shifts of Acyclic Alkenes', *J. Org. Chem.*, **36**, 2757–2766 (1971).
10. R. Silverstein, F. Webster, *Spectrometric Identification of Organic Compounds*, 6th edition, John Wiley & Sons, New York, 1998.
11. D. Ewing, '^{13}C Substituent Effects in Monosubstituted Benzenes', *Org. Magn. Reson.*, **12**(9), 499–524 (1979).
12. E. Breitmaier, *Structure Elucidation by NMR in Organic Chemistry, A Practical Guide*, John Wiley & Sons, New York, 1995.
13. E. Friedrich, K. Runkle, 'Empirical NMR Chemical Shift Correlations for Methyl and Methylene Protons', *J. Chem. Educ.*, **61**(9), 830–832 (1984).
14. J.A.D. Jeffreys, 'A Rapid Method for Estimating NMR Shifts for Protons Attached to Carbon', *J. Chem. Educ.*, **56**, 806–809 (1979).
15. H. Bell, L. Berry, E. Madigan, 'Additive Chemical Shift Parameters for Methylene Protons. Shoolery's Constants Reinvestigated', *Org. Magn. Reson.*, **22**(11), 693–696 (1984).
16. P. Beauchamp, R. Marquez, 'A General Approach for Calculating Proton Chemical Shifts for Methyl, Methylene, and Methine Protons When There Are One or More Substituents within Three Carbons', *J. Chem. Educ.*, **74**, 1483–1485 (1997).
17. C. Pascual, J. Meier, W. Simon, 'Reǵel zur Abschätzuńg der Chemischen Verschiebung von Protonen and Einer Doppelbinduńg', *Helv. Chim. Acta*, **49**(21), 164–168 (1966).
18. J. Martin, B. Dailey, 'Chemical Shifts in Polysubstituted Benzenes', *J. Chem. Phys.*, **39**(7), 1722–1728 (1963).

Electron Spin Resonance Spectroscopy

Mario Chiesa and Elio Giamello
Università di Torino, Turin, Italy

1 Introduction **2**
1.1 Paramagnetism, Unpaired Electrons and Electron Spin Resonance Spectroscopy 2
1.2 Origin of the Technique and Applications in Chemistry 2
1.3 Books and Chapters About Electron Spin Resonance Spectroscopy 3
2 Basic Principle of the Technique **3**
2.1 Electron Spin and the Zeeman Effect 3
2.2 Spin Resonance and Spin Relaxation 5
2.3 The Continuous Wave Electron Paramagnetic Resonance Spectrometer 6
3 Spin Resonance in Atomic and Molecular Systems **8**
3.1 The Electron Spin–Nuclear Spin (Hyperfine) Interaction 8
3.2 Organic Radicals in Solution 12
3.3 The g Tensor 13
3.4 Systems with $S > \frac{1}{2}$: the Fine Structure 16
4 Spectra of Polycrystalline Materials (Powder Spectra) **17**
4.1 Line Shape of Simple Powder Spectra 18
4.2 Powder Spectra of Some $S = \frac{1}{2}$ Transition Metal Ions 18
4.3 A Pragmatic Approach to Powder Spectra Interpretation 20
4.4 Solid-state Versus Solution Spectra 24
5 Advanced Electron Magnetic Resonance Techniques **25**
5.1 Electron Nuclear Double Resonance 25
5.2 Pulsed Techniques 25
6 Analytical Applications of Electron Paramagnetic Resonance **27**
6.1 Spin Trapping 27
6.2 Radiation Dosimetry 27
6.3 Electron Paramagnetic Resonance Dating 29
Acknowledgments **30**
Abbreviations and Acronyms **30**
Related Articles **30**
References **30**

Electron spin resonance (ESR) or Electron paramagnetic resonance (EPR) spectroscopy is an experimental technique for detecting and characterizing chemical systems bearing one or more unpaired electrons. The application of the technique, the background of which is similar to the more diffused nuclear magnetic resonance (NMR) spectroscopy, therefore concerns (i) organic and inorganic free radicals, (ii) trapped radicals in various matrices, including irradiated solids (radiation chemistry), frozen inert matrices and solid surfaces, (iii) transition metal ion compounds in classical inorganic systems or in biological systems, (iv) excited paramagnetic states (triplets) and several other systems of relevant scientific interest. The aim of this article is to provide an overall survey on the basic principles and on the various applications of the technique. This is done by describing, first, the physical basis of the electron resonance phenomenon (microwave absorption when the system is in the line of force of a strong external magnetic field) and, subsequently, the main types of interactions that the unpaired electron undergoes in the chemical system to which it belongs and which determines the features of the experimental spectra. These are essentially (i) the magnetic interaction of the electron with nuclei of nonzero nuclear spin (hyperfine interaction) which determines the multiline structure of the spectrum (hyperfine structure) and (ii) the interaction of the electron spin with the electron orbital angular momentum, occurring through the so-called spin–orbit coupling. This latter interaction causes the dependence of the resonance on the orientation of the radical in the external magnetic field. Particular emphasis will be given to the different types of spectra observed according to the physical state of the investigated sample (liquid solutions, single solid crystals or microcrystalline powdered solids). In the last case (powders), the essential methodology for understanding the often complex profiles of the experimental spectra is described. Advanced electron resonance techniques, essentially ENDOR (electron–nuclear double resonance) and time-resolved EPR will also be briefly mentioned.

The final part of the article is devoted to some analytical applications of the technique, including spin trapping (an essential tool for the detection and quantitative evaluation of reactive short-lived radicals), radiation dosimetry (with applications to medical problems and irradiated foodstuff analysis) and dating of geologically relevant systems.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

ESR or EPR is a powerful tool for investigating paramagnetic entities in the domains of chemistry, physics and biology. The limitation of the technique is evident from the previous definition, in that all diamagnetic systems, which represent the majority of the substances, are excluded from EPR investigations. However, this limitation is also an advantage since, for instance, reactive paramagnetic intermediates present in complex media or paramagnetic centers belonging to a complex chemical system (e.g. a paramagnetic ion in a protein) can be studied without any spectroscopic interference.

1.1 Paramagnetism, Unpaired Electrons and Electron Spin Resonance Spectroscopy

Before describing the origin and the concepts of ESR spectroscopy, parts of which are common to the more familiar and widely used NMR, it must be recalled that a paramagnetic body, when placed into the lines of force of a magnetic field, increases the density of the lines of force (called the magnetic flux density) within its own volume. In contrast, a diamagnetic substance decreases the magnetic flux density. Diamagnetism is present in all substances and is produced by electron circulation in atoms, ions and molecules having a closed shell of electrons. Paramagnetism is typical of substances having centers with one or more unpaired electrons and therefore bearing an intrinsic angular momentum – the spin – and a corresponding magnetic moment. These centers are essentially isolated from each other and, consequently, in the absence of magnetic cooperative effects causing ferromagnetism or antiferromagnetism.

In the presence of an external magnetic field, the magnetic moments on different centers tend to align with the field and hence with each other (this effect being contrasted by the randomizing effect of thermal energy), whereas in the absence of a field the unpaired electrons are aligned randomly.

Electron magnetic resonance occurs when an electromagnetic wave of suitable frequency (usually in the microwave region) interacts with the sample immersed in a magnetic field and causes a ''flipping over'' of the free electron spin vectors, along with the associated magnetic moment, so that a certain amount of electromagnetic energy is absorbed by the system. In quantum mechanical terms the description of the electron resonance (see section 2) can be reduced to the description of the behavior of a single unpaired electron undergoing, under the effect of an external magnetic field, the so-called Zeeman effect, i.e. an energy splitting between the two allowed spin states. Under such conditions an electron in the lower state is promoted to the upper state by absorption of a photon of suitable energy.

Two main features distinguish EPR (and the analogous NMR) from the ''classical'' types of spectroscopy employing electromagnetic waves to investigate matter. The first is that, in most molecular spectroscopic techniques, the electric-field component of the radiation interacts with some electric dipole moment in the molecules. In magnetic resonance spectroscopy, in contrast, it is the magnetic component of the radiation that interacts with the magnetic dipoles of paramagnetic centers. The second feature concerns the principle of the experimental set-up which, instead of being based on a source of electromagnetic radiation and a monochromator that selects the frequency to be sent to the sample, consists of a source of nearly monochromatic radiation coupled with a variable magnetic field. In a normal experiment with the standard continuous wave electron paramagnetic resonance (CWEPR), the sample is irradiated with a fixed frequency and a given interval of magnetic field is swept to look for the conditions (field and frequency) of the magnetic resonance. A CWEPR spectrum is thus a diagram where microwave absorption (or more frequently, for the reasons discussed in section 2, its first derivative) is reported as a function of the magnetic field swept in the experiment.

1.2 Origin of the Technique and Applications in Chemistry

The first observation of an EPR spectrum was made in Kazan (USSR) in 1945 by Zavoisky. The technique was, however, developed mostly at Oxford (UK) from both theoretical (Pryce, Abragam) and experimental (Bleaney) points of view, and also on the foundations provided by earlier work by Van Vleck on magnetic susceptibility. A considerable contribution to the exploitation of the EPR technique was due to the development of microwave systems, to be employed for radio detection and ranging (radar), during the Second World War. The first applications of EPR to chemistry were in the field of transition metal ions, and the observations on the spectra of organic radicals were performed later in 1952.

To classify the applications of EPR in chemistry, physics and biology it is convenient to have in mind a clear methodological distinction. EPR is employed according to two basic approaches. In the first, paramagnetic entities are introduced in diamagnetic, usually complex, systems to obtain information on the properties of the systems themselves. This technique is usually termed the ''spin-probe'' or ''spin-label'' method. In the former case, the probes, usually paramagnetic ions or stable nitroxide radicals, are added to complex systems such as biological

membranes to "explore" particular structural features or alterations of the properties caused by pathologies, interactions with drugs, etc. The spin label is also a stable radical like a nitroxide but it has a particular functional group suitable to bind the molecule permanently to particular sites of the system. For its high specificity the spin-probe–spin-label method will not be discussed further in this article.

In the second approach, EPR directly investigates paramagnetic centers present in various chemical and biological systems. Typical systems that are studied in EPR include

- Free radicals in chemistry, biochemistry and biology.
- Radicals and paramagnetic point defects trapped in solid matrix. EPR, in fact, is one of the leading techniques to study the effects of radiation in the solid state and, generally, for the study of defects in solid state physics and material science.
- Transition metal ion compounds in inorganic chemistry and bioinorganic chemistry.
- Systems with more than one unpaired electron (triplets, quadruplets, etc.). Some of these systems are stable ground states but, in most cases, they are excited short-lived states.

Several new specific applications of EPR have recently appeared which are aimed at solving particular problems in branches of science other than chemistry, physics and biology. This is the case, for instance, with dating techniques in archaeology and geology and with nuclear radiation dosimetry. These recent applications of EPR will be described in section 6.

The classical CWEPR is by far the most commonly used electron magnetic resonance technique. For this reason, this article mainly refers to this experimental approach. However, the constant evolution of the microwave physics and engineering has allowed the introduction of new experimental approaches in the field of electron magnetic resonance. The first and probably the most important one is the ENDOR, introduced in the late 1950s, which will be briefly described in section 5.1. More recently, pulsed techniques have been introduced in EPR and, after nearly 30 years of laboratory-built pulsed spectrometers and some years after the corresponding evolution of NMR, this kind of instrumental apparatus has become commercially available. The principal feature of pulsed EPR, which will be briefly described in section 5.2, consists of the substitution of the continuous microwave irradiation of the paramagnetic sample with suitable sequences of microwave pulses.

In conclusion to these introductory remarks, it should be pointed out that whether or not the reader has an immediate interest in the multitude of physical systems that can be investigated by EPR, this type of spectroscopy occupies a paramount position among the many physical techniques employed to investigate matter. In fact, there are hardly any other techniques that provide such a deep and detailed insight into the investigated systems. Furthermore, the quantum mechanics concepts that are needed for a nontrivial understanding of the principles and of the results of EPR have made this technique a stimulating intellectual challenge for many researchers in various domains of science.

1.3 Books and Chapters About Electron Spin Resonance Spectroscopy

The content of this article is basically a short introduction to the basic principles and applications of ESR. The reader wishing to increase their level of knowledge on this subject is referred to a series of modern general books[1–5] covering all the basic aspects of the technique. The particular domain of transition metal ions is well covered by several books,[6–9] as are the domains of spin labeling and biological applications of EPR.[10–12] Modern double-resonance and pulsed techniques, which are assuming increasing importance in the field of electron magnetic resonance, are illustrated in several books and review articles,[13–16] and an interesting report on new applications of EPR to fields such as dating and radiation dosimetry has also appeared recently.[17]

2 BASIC PRINCIPLE OF THE TECHNIQUE

2.1 Electron Spin and the Zeeman Effect

The fundamental properties of an electron are mass, charge and intrinsic angular momentum or spin, a vector denoted by the symbol $\boldsymbol{S}$. It is known from quantum mechanics that only the square of the intrinsic angular momentum ($\boldsymbol{S}^2$) and one component in a given direction (say z, hence S_z) can be measured with certainty. The intrinsic angular momentum of a given quantum system is defined by the spin quantum number S, which can assume integer or half-integer values only (0, $\frac{1}{2}$, 1, etc.). For an electron $S = \frac{1}{2}$. As a consequence, the allowed S_z values are two ($2S+1$ are the allowed values), i.e. $S_z = \frac{1}{2}\hbar$ (sometimes referred to as the $\uparrow$ or α state, quantum number $M_s = \frac{1}{2}$), $S_z = -\frac{1}{2}\hbar$ ($\downarrow$ or β, $M_s = -\frac{1}{2}$). The symbol $\hbar$ represents the unit of angular momentum, $\hbar = h/2\pi$, where h is Planck's constant, $h = 6.626 \times 10^{-34}$ J s. $\boldsymbol{S}^2$ in general is $S(S+1)\hbar$ and becomes, for an electron, $\boldsymbol{S}^2 = \frac{3}{4}\hbar^2$ with $S = \frac{\sqrt{3}}{2}\hbar$ (Figure 1).

The electron, by virtue of its spin, has associated with it a magnetic moment μ_s which is collinear and antiparallel

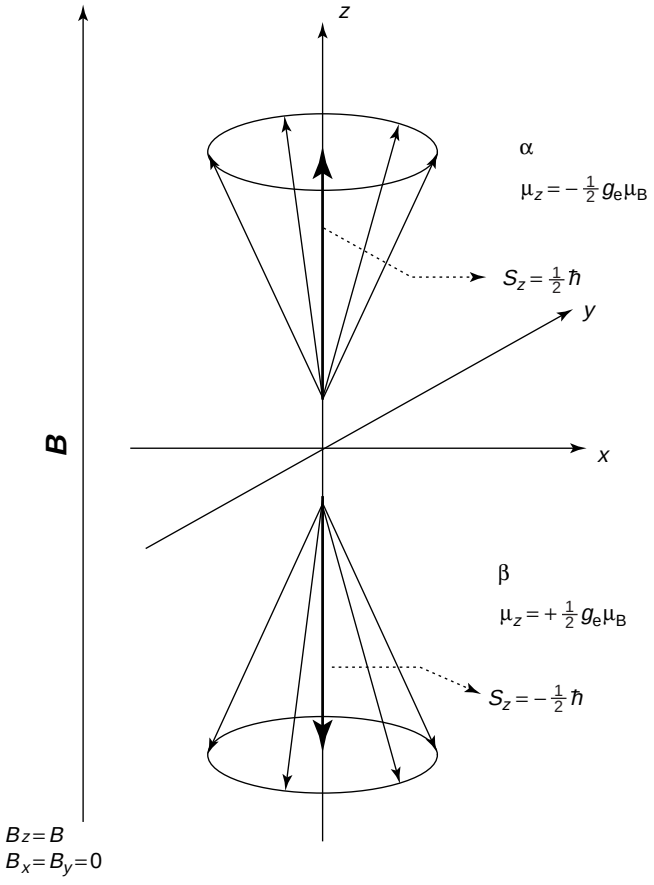

Figure 1 The two possible orientations of the spin of a free electron in a magnetic field oriented along the z-axis.

to the spin itself and given by Equation (1):

$$\mu_s = -g_e \mu_B \boldsymbol{S} \tag{1}$$

where g_e is the free electron g value ($g_e = 2.0023$), μ_B is the Bohr magneton ($\mu_B = eh/4\pi mc = 9.27 \times 10^{-24}\,\mathrm{J\,T^{-1}}$) with e and m being the electron charge and mass, respectively, and c the velocity of light. In the absence of a magnetic field the spin vectors $\boldsymbol{S}$ (and the corresponding magnetic moments) of an assembly of free electrons are randomly aligned, but in the presence of an external magnetic field the magnetic moments tend to align and define the axis of orientation of the S_z component. The situation is illustrated in Figure 1. Because of the complete uncertainty with respect to the S_x and S_y components, the vectors representing the spin of the electrons can lie on the surface of one of the two cones whose axes coincide with the z axis. In this way all electrons assume one of the two permitted S_z values ($+\frac{1}{2}\hbar$, $-\frac{1}{2}\hbar$). It is known from classical physics that an energy develops by interaction between a magnetic field and a magnetic dipole having a given magnetic moment μ which is given by a scalar product (Equation 2):

$$E = -\mu \boldsymbol{B} \tag{2}$$

[where $\boldsymbol{B}$ is the magnetic flux density in the matter measured in tesla (T) or gauss (G) ($1\,\mathrm{T} = 10^4\,\mathrm{G}$)] and which thus depends on the orientation of the magnetic dipole with the field ($E = \mu \boldsymbol{B} \cos\theta$). This means that for electrons in a magnetic field, the two permitted orientations of the spins (α and β) have different energies. In a field of 3000 G, for instance, the α state lies $0.3\,\mathrm{cm^{-1}}$ above the β state. This phenomenon is sometimes called the electronic Zeeman effect, or the first-order Zeeman effect. In quantum mechanics, the μ vector in Equation (2) is replaced by the corresponding operator (Equation 1) leading to the Hamiltonian in Equation (3), i.e. the energy operator:

$$H = g_e \mu_B \boldsymbol{B} \boldsymbol{S} \tag{3}$$

As $\boldsymbol{B}$ lies along the z-axis ($B_x = B_y = 0, B = B_z$), we have (Equation 4)

$$H = g_e \mu_B B \hat{S}_z \tag{4}$$

which is the simplest example of a spin Hamiltonian, i.e. the energy operator acting on the spin components of the whole electron wave function. $\hat{S}_z$ acts on the spin functions α and β according to

$$\hat{S}_z|\alpha\rangle = \tfrac{1}{2}|\alpha\rangle \quad \text{and} \quad \hat{S}_z|\beta\rangle = -\tfrac{1}{2}|\beta\rangle$$

The permitted energy levels of the system (i.e. the energies corresponding to the two allowed orientations of the spin) are therefore

$$E = \left(\pm\tfrac{1}{2}\right) g_e \mu_B B \tag{5}$$

The lower energy state occurs when the electron magnetic moment is lined up with the applied magnetic field ($M_s = -\frac{1}{2}$, β). A higher energy state occurs when μ_s lines up in opposition to the magnetic field ($M_s = +\frac{1}{2}, \alpha$). The energy difference between these two states is given by Equation (6):

$$\Delta E = g_e \mu_B B \tag{6}$$

At thermal equilibrium, under the influence of the external applied magnetic field, the spin population is split between the two levels according to the Maxwell–Boltzmann law (Equation 7):

$$\frac{n_\alpha}{n_\beta} = \mathrm{e}^{-\frac{\Delta E}{kT}} \tag{7}$$

where k is Boltzmann's constant, T the absolute temperature and n_α and n_β are the spin populations characterized by the M_s values $+\frac{1}{2}$ and $-\frac{1}{2}$ respectively. At 77 K, in a field of about 3000 G, n_1 and n_2 differ by less than 0.005.

2.2 Spin Resonance and Spin Relaxation

The ESR experiment is the observation of the energy required to reverse the direction of an electron spin in the presence of an external magnetic field. Electrons in the lower Zeeman state (β) can be promoted to the upper state (α) by irradiating the paramagnetic system with suitable electromagnetic radiation. To observe the spin transition the following condition (Equation 8) must be fulfilled:

$$h\nu = g_e \mu_B B \tag{8}$$

From Equation (8), it can easily be deduced that for a free electron in a 3000-G magnetic field the resonance frequency ν lies in the microwave region of the electromagnetic spectrum ($\sim$9 GHz). However, for reasons that will be explained in section 2.3, the actual EPR experiment is run by irradiating the sample with a fixed frequency and sweeping a suitable range of magnetic field. When the magnetic field reaches the value necessary to fulfill the resonance condition in Equation (8) (which is also called the resonant field), an absorption of energy is observed that results in a spectral line. The width of this line (reported in Figure 2 both as absorption and first derivative of absorption) essentially depends on the efficiency of the mechanism (relaxation mechanism) through which the excited system returns to the ground state.

By the absorption process, in fact, the populations of the two energy levels n_α and n_β tend to equalize. The odd electrons from the upper level give up an amount of energy corresponding to the $h\nu$ quantum to return to the lower level in order to satisfy the Maxwell–Boltzmann law. This energy may be dissipated in two different types of relaxation processes, spin–lattice and spin–spin relaxation, respectively. In the former mechanism, the magnetic energy absorbed by the paramagnetic system is transferred to vibrational degrees of freedom of the lattice. The motions of the lattice (phonons) give rise to fluctuating magnetic fields at the unpaired electron; if these fluctuations happen to have a component that oscillates at the transition frequency, then a transition can be induced. The width of the resonance line is inversely proportional to the lifetime of the excited state. This can be shown in terms of Heisenberg's uncertainty principle, which relates the line width to the lifetime of the excited state, by Equation (9):

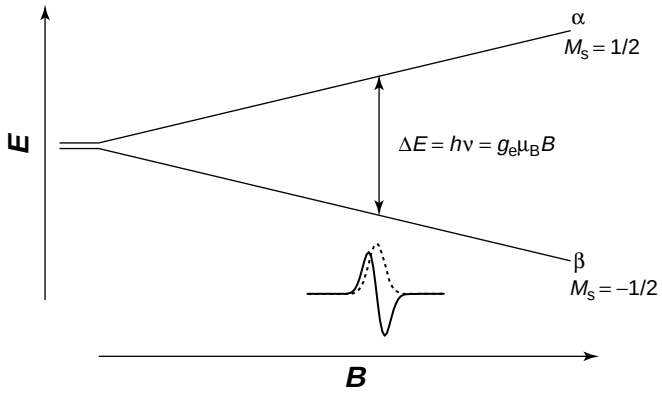

Figure 2 The electron Zeeman effect. The energy difference between the α and β spin states is proportional to the value of the applied magnetic field.

$$\Delta t \, \Delta E \approx \frac{h}{2\pi} \tag{9}$$

where Δt is the lifetime of the excited state and ΔE the effective width of the energy band involved. From Equation (9), it immediately follows that if Δt is small, ΔE will be large and the absorption will be observed over a range of energies (or magnetic field) corresponding to a large line width. Spin–lattice relaxation is characterized by an exponential decay of energy as a function of time. The exponential time constant is denoted T_1 or spin–lattice relaxation time. T_1 has been found to be sensitive in particular to

- Temperature. A decrease in temperature decreases the population of phonons and leads to a longer relaxation time and correspondingly narrower lines.
- The presence of nearby electronic excited states. Where these are close ($\sim$100 cm^{-1}) to the ground state, the relaxation time will be very short and the line width very large or even not observable.

The other relaxation process that may take place is the spin–spin relaxation process. This phenomenon is characterized by a time constant T_2 called the spin–spin relaxation time and involves an energy exchange between the electron spins without transfer of energy to the lattice. The width and the shape of the EPR lines are strictly related to the type and to the rate of the motion that a given paramagnetic center undergoes. An important application of EPR is the study of the motional behavior of particularly stable radicals in various media (e.g. natural and artificial membranes, liquid crystals) to obtain information on the whole system. The radical molecules employed in these kinds of studies are usually nitroxides (containing an $-NO^{\bullet}$ group) and are called spin probes or spin labels.

When a spin system is irradiated with a microwave power too high for the efficiency of the relaxation mechanism operating in the system itself, the populations of the two Zeeman levels tend to equalize and the intensity of the signal decreases, no longer being proportional to the number of spins present in the sample. This effect, known as "saturation", can be avoided by exposing the sample to suitable microwave power. In

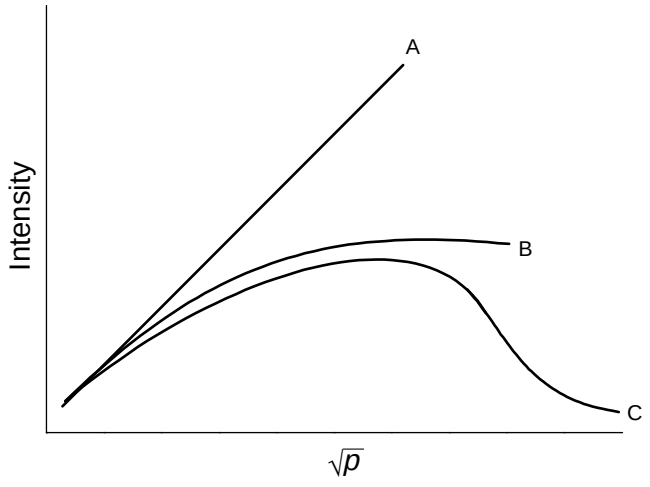

Figure 3 Different behaviors of the EPR spectral intensity as a function of microwave power: (A) no saturation; (B) homogeneous saturation; (C) heterogeneous saturation.

order to determine the microwave power that avoids saturation effects, the signal intensity is usually plotted as a function of the square root of the incident microwave power. Three examples of such a plot are shown in Figure 3.

The linear part of the plot indicates the power region where saturation effects are absent. Three different saturation trends can therefore be identified: (A) no saturation of the signal; (B) homogeneous saturation, typical of single EPR lines arising from a single EPR parameter and from samples exhibiting a single type of center or, at least, little heterogeneity of the centers; and (C) heterogeneous saturation, occurring when the EPR lines arise from the superimposition of several overlapping individual lines. The saturation of an EPR spectrum is sometimes employed in heterogeneous systems, or systems containing more than one type of paramagnetic species, in order to increase the intensity of one resonance with respect to a second partially overlapped signal.

When both spin–lattice and spin–spin relaxations contribute to the EPR line, a rough expression of the resonance line width is (Equation 10):

$$\Delta H \propto \frac{1}{T_1} + \frac{1}{T_2} \tag{10}$$

In general $T_1 > T_2$ and the line width depends mainly on spin–spin interactions: T_2 increases on decreasing the spin concentration, i.e. the spin–spin distances in the system.

There are two typical line shapes that are useful to describe the EPR lines, namely Gaussian and Lorentzian, whose analytical expressions are given by Equations (11) and (12), respectively:

$$y = a\mathrm{e}^{-bx^2} \tag{11}$$

$$y = \frac{a}{1 - bx^2} \tag{12}$$

A detailed description of the line shape in EPR is beyond the scope of this article. It is sufficient to note that the Gaussian line shape occurs in magnetic resonance from the statistical distribution of similar paramagnetic centers.

2.3 The Continuous Wave Electron Paramagnetic Resonance Spectrometer

A CWEPR spectrometer is designed to detect with high sensitivity the microwave absorption in a sample as a function of the external applied magnetic field. The EPR experiment is performed by maintaining a constant microwave frequency and sweeping the applied field until the resonance condition, given by Equation (8), is fulfilled, i.e. the incident radiation is absorbed.

The basic components of the spectrometer (Figure 4) are as follows:

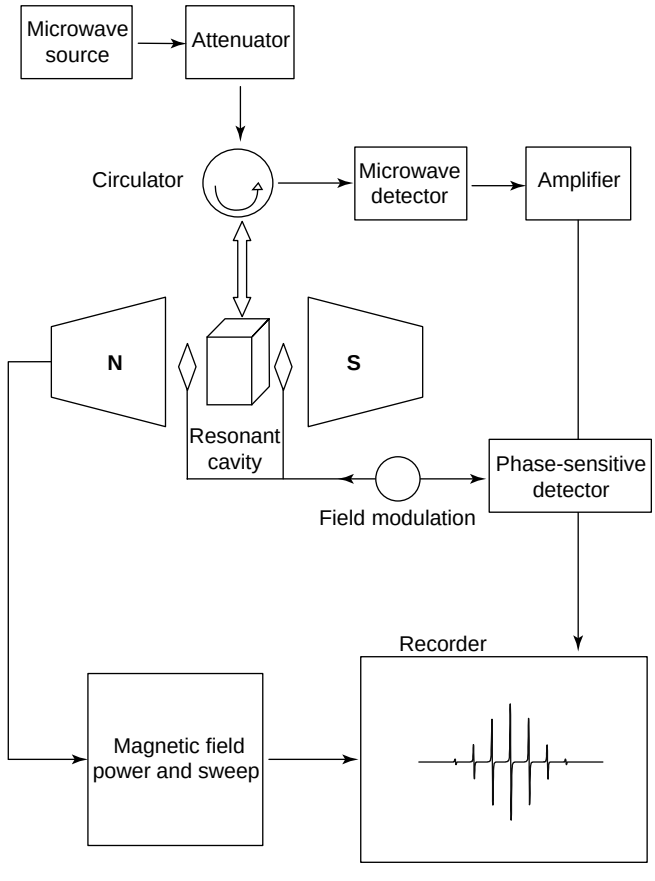

Figure 4 Scheme of a CWEPR spectrometer.

Table 1 Features (frequency, wavelength and typical magnetic field) of the various bands employed in ESR spectroscopy

Band	Typical EPR frequencies, ν (GHz)	Typical EPR field (mT)	λ (cm)
W	95	3400	0.3
Q	36	1300	0.80
K	23	820	1.5
X	9.5	340	3.0
S	3.0	110	10
L	1.5	54	20

- A microwave source (klystron or Gunn diode) supplying electromagnetic radiation at a constant frequency ν whose power is controlled by an attenuator between the microwave source and the sample. The frequencies most often employed are those corresponding to X (~9 GHz) and Q (~35 GHz) bands. In the second half of the 1990s, the high-frequency W band (96 GHz) became commercially available. Other frequencies, e.g. K (25 GHz) and S (3 GHz) band, are sometimes employed. The experimental features of these bands are reported in Table 1.
- A microwave guide system to direct the microwave radiation from the source to the resonant cavity.
- A cavity made from a highly conductive metal and having reflecting walls to accumulate power of the microwave radiation on to the sample. The internal dimensions of the cavity are similar to the wavelength of the microwaves.
- A powerful electromagnet capable of providing a homogeneous field that is controlled by a field probe. The rate and range of scanning are adjusted to provide the most suitable conditions for observation of the microwave absorption.
- A detector diode to measure the energy absorbed by the sample at resonance.
- A convenient amplifier system, a recorder, and an oscilloscope.

Superimposed on the main magnetic field, an oscillating field is obtained by applying an alternating current (typically 100 kHz) to a set of coils in the cavity walls. This modulating field converts the resonance to an alternating signal, which can be separated from random noise using a phase-sensitive detection system. This method of detection has a very high sensitivity and leads to an output whose amplitude is proportional to the slope A_1A_2, so that the resulting EPR signal now appears as the first-derivative curve of the absorption signal (Figure 5). The experimentalist selects the amplitude of the modulating field, which should be large enough to obtain a good signal-to-noise ratio but small enough to obtain a good first-derivative curve, i.e. to prevent distortions, which occur if the value of the amplitude approaches the resonance line width.

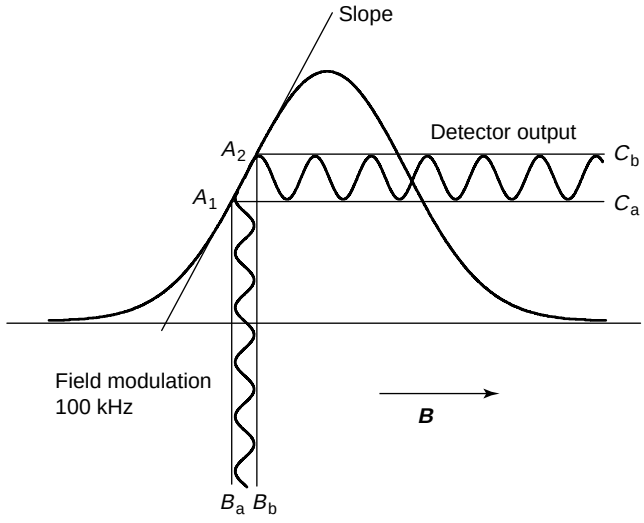

Figure 5 Effect of field modulation on the detector output current. The static magnetic field is modulated between B_a and B_b. The corresponding detector current varies between the limits C_a and C_b.

Paramagnetic compounds are usually studied by EPR either in solution or in the solid state. In the solid state one has to distinguish between single-crystal systems and polycrystalline systems. The latter case encompasses the true polycrystalline solids (powders) and the frozen solutions.

2.3.1 Fluid Solution

Samples of stable paramagnetic compounds in fluid solution are usually easy to prepare. The removal of oxygen (which is a paramagnetic $S = 1$ molecule) from the solvent is sometimes necessary in order to obtain narrower resonance lines. In certain cases (e.g. unstable radicals) it is possible to generate the radicals in situ in the EPR cavity. The usual sample tubes used for the EPR experiment are cylindrical tubes made of quartz glass, since Pyrex and other glasses contain small traces of paramagnetic impurities. It is also essential to expose only a thin sample to the microwave radiation, since all solvents absorb the microwaves to various degrees and if much absorption occurs it is not possible to obtain resonance. Good solvents are those that cause the minimum absorption effects and, in general, these are characterized by having low permittivity (dielectric constant). In the case of aqueous solutions (water has a high dielectric constant), special flat cells are usually employed, in which a liquid layer of about 0.1 mm is present. These undesirable effects of high dielectric constant solvents are strongly reduced in the solid state.

2.3.2 Single Crystal

Magnetically dilute single crystals of a paramagnetic compound can be obtained by growing a crystal of an isomorphous diamagnetic compound containing the paramagnetic species as impurity. Any of the standard crystal-growing methods may be used, the choice depending on the nature of the radical being studied. Alternatively, a single crystal containing a paramagnetic impurity may be produced by means of high-energy radiation of the bare crystal. As will be shown below, the EPR spectra of single crystals exhibit different features as a function of the orientation of the sample in the cavity (anisotropy of magnetic tensors).

2.3.3 Powders, Glasses and Frozen Solutions

The EPR spectra of polycrystalline materials exhibit different features from those of single crystals, since they are invariant to rotation of the sample in the cavity (see below). Polycrystalline materials can be either assemblies of very small single crystals randomly oriented in the space or microcrystals of a frozen solution. In fact, a convenient method for the study of many compounds is to observe their EPR spectra in solution and then to freeze the solution and observe the solid-state spectrum. The former spectra give the isotropic parameters for the various species, while the latter yield the anisotropic data. A particular case is glasses and amorphous materials which, from the point of view of EPR, display the same spectral features as a polycrystalline material.

Low-temperature EPR spectra can be recorded at 77 K using a special Dewar fitting in to the cavity. Variable-temperature EPR spectra can be recorded by using a flow of cooled, dried nitrogen in a liquid nitrogen cryostat for temperatures from about 120–130 K up to 450 K. A flow of cooled, dried helium in a liquid helium cryostat allows spectra to be recorded from 4 K. Both nitrogen and helium cryostats are fitted into the EPR microwave cavity.

The digital acquisition of spectra with on-line computers (which are included in modern spectrometers) allows an improvement of the signal-to-noise ratio by both spectra accumulation and spectral curve smoothing. Mathematical data handling can be also used to obtain second and third derivatives of the spectra, which are extremely useful for detecting overlapping signals. The measurement of spin concentration is done by double integration of the first-derivative EPR signal and comparison of the intensity of the integrated signal (which is proportional to the number of spins present in the system) with that of standard samples with a known number of spins.

3 SPIN RESONANCE IN ATOMIC AND MOLECULAR SYSTEMS

To understand how information about the nature and the structure of a real paramagnetic entity is obtained by means of EPR, we now have to consider how the unpaired electron interacts with its environment. The interactions of an unpaired electron with its surroundings (in addition to the Zeeman interaction with the external magnetic field) are essentially of magnetic or electric nature. We shall limit ourselves in this section to a purely phenomenological approach, considering the three main interactions potentially involving an unpaired electron. These are (i) the electron spin–nuclear spin interaction (hyperfine interaction), (ii) the coupling of the spin angular momentum with angular momenta of orbital origin and (iii) the interaction of the electron spin with other electron spins. An unpaired electron in a real chemical system does not necessarily undergo all these three types of physical interactions. It will be shown in the following, for instance, that the spectra of organic radicals in solution are mainly analyzed in terms of the electron spin–nuclear spin coupling. It will also be clarified in this section that the main problems in the analysis of experimental EPR spectra arise because of the anisotropic nature of all the above-mentioned magnetic interactions. This means that the extent of such interactions varies as a function of the orientation of the sample in the external magnetic field. The anisotropic effects show up for samples in the solid state, but are averaged (and, in some cases, vanish) for paramagnetic systems in low-viscosity solutions because of the rapid molecular motion.

3.1 The Electron Spin–Nuclear Spin (Hyperfine) Interaction

Most of the information which can be deduced from EPR spectra arises from the analysis of the hyperfine structure. This is particularly true for organic radicals which exhibit small g shifts (the other parameter characterizing EPR spectra and better observed in high-frequency Q or W bands rather than in the classical X-band) which, in any case, are not easy to correlate with the molecular structure.

A hyperfine structure is composed of a series of lines with different intensity and well-defined separation which arise from the interactions between electron and nuclear magnetic dipoles in the case of nuclei with nonzero nuclear spin. The case of a free radical with a single nucleus of spin $I = \frac{1}{2}$ will be discussed in some detail and the most important results will be extended to the case of interactions with n nuclei.

Several nuclei possess spins and associated magnetic moments. The magnetic moment μ_n associated with a

nucleus is collinear with the spin vector $\boldsymbol{I}$ according to Equation (13):

$$\mu_n = g_n \beta_n \boldsymbol{I} \tag{13}$$

where g_n is the nuclear g factor and β_n is the nuclear magneton ($\beta_n = 5.05 \times 10^{-27}\,\mathrm{J\,T^{-1}}$), which is the analog of the Bohr magneton for the proton mass. Nuclear and Bohr magnetons are correlated by a factor of 1838, i.e. the ratio between the proton and electron masses.

For a nucleus with nuclear spin $I = \frac{1}{2}$ the interaction of the nuclear magnetic moment with an external applied magnetic field can be described in terms of the following Hamiltonian (Equation 14):

$$H = -g_n \beta_n \boldsymbol{BI} \tag{14}$$

If $\boldsymbol{B}$ lies in the z direction ($B_x = B_y = 0$ and $B_z = B$), the interaction energy corresponds to (Equation 15)

$$H = -g_n \beta_n B \hat{I}_z \tag{15}$$

and the energy corresponding to the two allowed orientations of the nuclear spin (nuclear Zeeman effect) is given by Equation (16):

$$E = \left(\pm\tfrac{1}{2}\right) g_n \beta_n B \tag{16}$$

The total spin-Hamiltonian for the case under discussion is composed of three terms and can be written according to Equation (17):

$$H = g\mu_B B_z \hat{S}_z - g_n \beta_n B_z \hat{I}_z + a\boldsymbol{IS} \tag{17}$$

The first and second terms account for the electron–Zeeman (Equation 4) and nuclear–Zeeman (Equation 15) interactions respectively, which have been discussed above. The third term accounts for the electron spin–nuclear spin interaction (i.e. hyperfine interaction) and a is called the hyperfine splitting constant. This interaction originates from the fact that, with the electron and the nucleus being two magnetic dipoles situated in close proximity to one another, they undergo a reciprocal interaction.

Two types of electron spin–nuclear spin interactions must be considered depending on whether the electron interacting with the nucleus is in an s or in any other orbital of higher l (e.g. p, d and f orbitals). In the first case the interaction is called the isotropic hyperfine coupling or Fermi contact interaction and is related to the finite probability of the unpaired electron being at the nucleus. For odd electrons in p, d or f orbitals no Fermi contact occurs due to the nodes of the corresponding wave functions at the nucleus. The electron spin–nuclear spin interaction is in this case due to the nonspherical symmetry of the orbitals and can be described in terms of the classical dipolar interaction between magnetic moments. The interaction is anisotropic in that it depends on the orientation of the orbital with respect to the applied magnetic field.

3.1.1 Anisotropic Hyperfine Interaction

The interaction energy between two magnetic moments (μ_1 and μ_2) is classically given by Equation (18):

$$E = \frac{\mu_1 \mu_2}{r^3} - \frac{3(\mu_1 \boldsymbol{r})(\mu_2 \boldsymbol{r})}{r^5} \tag{18}$$

where $\boldsymbol{r}$ is the vector relating the two moments and r is the distance between them. The quantum mechanical analog of Equation (18) is obtained by replacing μ_1 and μ_2 by their corresponding operators (Equation 19):

$$H_{an} = -g_e \mu_B g_n \beta_n \left[\frac{\boldsymbol{IS}}{r^3} - \frac{3(\boldsymbol{Ir})(\boldsymbol{Sr})}{r^5}\right] \tag{19}$$

Since the electron is not localized at one position in space, Equation (19) must be averaged over the electron probability distribution function. H_{an} is averaged out to zero when the electron cloud is spherical (s orbital) and comes to a finite value in the case of axially symmetric orbitals (p, d, etc.).

3.1.2 Isotropic Hyperfine Interaction

We now turn to consider the situation when the electron is in an s-type orbital. The simplest example of isotropic hyperfine interaction is the one observed for the hydrogen atom. The classical dipolar interaction cannot account for the hyperfine splitting in the hydrogen atom (whose EPR spectrum is constituted by two lines separated by about 508 G) since the electron distribution in a 1s orbital is spherically symmetric. The dipolar interaction is thus averaged to zero.

As mentioned above, the isotropic hyperfine interaction is related to the unpaired electron density at the interacting nucleus which has a nonzero value for s-type orbitals. The spherical symmetry of s orbitals explains the isotropic nature of the contact interaction, which is also known as the Fermi contact interaction and is given by Equation (20):

$$a = a_{iso} = \frac{8\pi}{3} g_e g_n \mu_B \beta_n |\Psi_{(0)}|^2 \tag{20}$$

where $|\Psi_{(0)}|^2$ is the square of the value of the wave function of the unpaired electron evaluated in the volume of the nucleus.

Let us now tackle the example of an unpaired electron spin interacting with the proton ($I = \frac{1}{2}$) spin. The interaction occurs via isotropic Fermi contact only. The case of a hydrogen atom corresponds to the schemes proposed in this example which is, however, rather

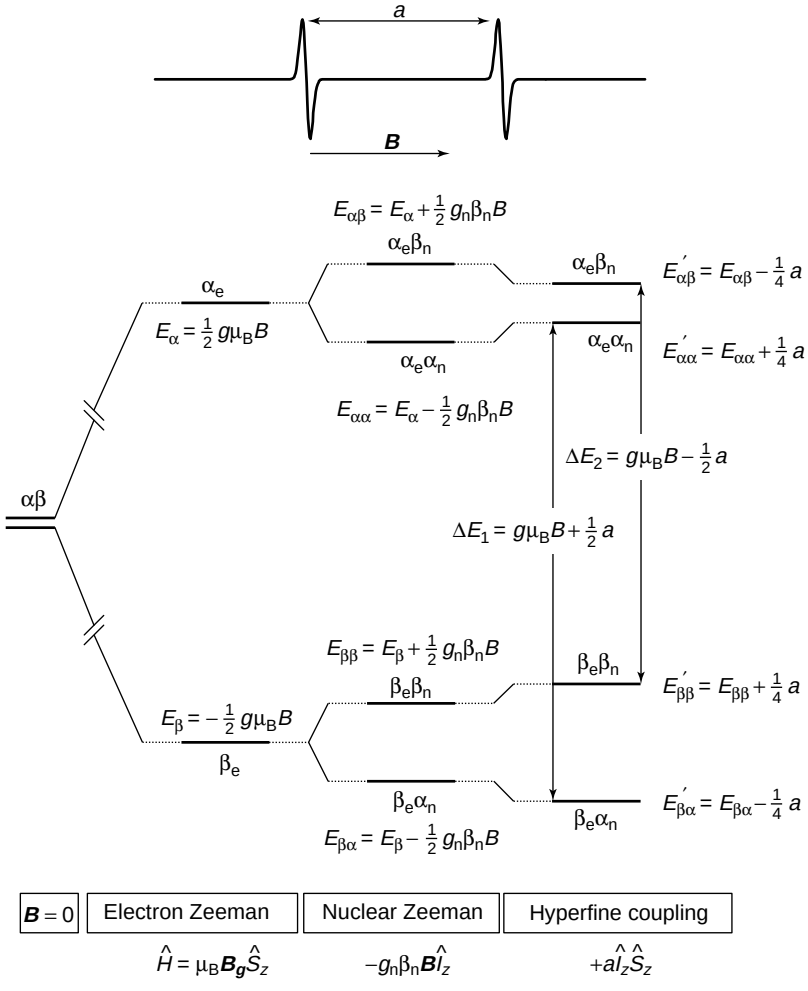

Figure 6 Energy levels scheme for the case of an unpaired electron ($S = \frac{1}{2}$) interacting with one nucleus having nuclear spin $I = \frac{1}{2}$. The scheme corresponds to the case of the hydrogen atom. The observed spin transitions are evidenced and the related EPR spectrum is shown.

general and includes also other chemical systems (see below). The nuclear spin may assume two possible orientations with $M_I = \pm\frac{1}{2}$. Hence the nuclear magnetic moment further splits each Zeeman level into two sublevels of different energy. The energy of the four electron–nuclear Zeeman levels is determined by the three terms of the Hamiltonian in Equation (17). In particular, the last term in this equation modifies the energies of the four levels as shown in Figure 6. The EPR selection rule allows transitions between β and α electron spin states, provided that the nuclear spin is kept constant. This is expressed by $\Delta M_s = \pm\frac{1}{2}$, $\Delta M_I = 0$. In the case of the system described here, the two allowed transitions (indicated in the figure) correspond to the following energy differences (Equations 21 and 22):

$$\Delta E_1 = h\nu_1 = g\mu_B B + \tfrac{1}{2}a \tag{21}$$

$$\Delta E_2 = h\nu_2 = g\mu_B B - \tfrac{1}{2}a \tag{22}$$

from which it can be easily seen that the nuclear Zeeman effect influences the energy of the four spin levels but not the energy difference of the allowed spin transition. The scheme in Figure 6 and the energy differences (Equations 21 and 22) have been conceived for practical reasons for a virtual CWEPR experiment at constant B and variable microwave frequency. In these terms the two allowed transitions are separated by the hyperfine splitting constant, expressed in energy units (Equation 23):

$$h\nu_1 - h\nu_2 = a \tag{23}$$

In a real CWEPR experiment (sweeping magnetic field at constant ν), the two transitions take place at two different B values (B_1 and B_2) separated by $a' = a/g\mu_B$ (in magnetic field units (Equation 6)). As already mentioned, the EPR spectrum of the hydrogen atom is composed of two lines split by approximately 508 G (Figure 6). This

value is the measure of the Fermi contact term at the nucleus.

The Fermi contact interaction is not the only effect which contributes to the isotropic hyperfine coupling. Since isotropic coupling is associated with unpaired electron spin density at the nucleus, then transition metal ions with unpaired electrons in d orbitals would not be expected to exhibit any isotropic coupling. However, appreciable isotropic coupling is always observed, e.g. vanadyl (VO^{2+}) complexes have $a_{iso}(^{51}V)$ values of up to nearly 120 G.

Another example is constituted by aromatic radicals such as $C_6H_6^-$. In this case the unpaired electron is located in a π molecular orbital formed from the overlap of carbon $2p_z$ orbitals. This molecular orbital has a node in the molecular plane containing the protons which give rise to the hyperfine structure. Also, in this case, the question is how the s orbitals of the hydrogen atoms can share a finite spin density to account for the observed hydrogen hyperfine structure.

In both cases a mechanism of spin polarization acts. This is described in terms of an exchange interaction between the unpaired electron and the paired s electrons. In the case of an odd electron in d orbitals, for example, spin polarization will occur between the unpaired d electron and the paired s electrons in the core ($1s^2$, $2s^2$, $3s^2$, etc.). The unpaired electron may be considered as attracting electrons of the same spin (Hund's rule) and repelling those of opposite spin. Hence the nucleus feels the influence of one of the s electrons (say β) more than the other (α) and so there is an apparent unpaired β-spin density at the nucleus as shown in Figure 7(a).

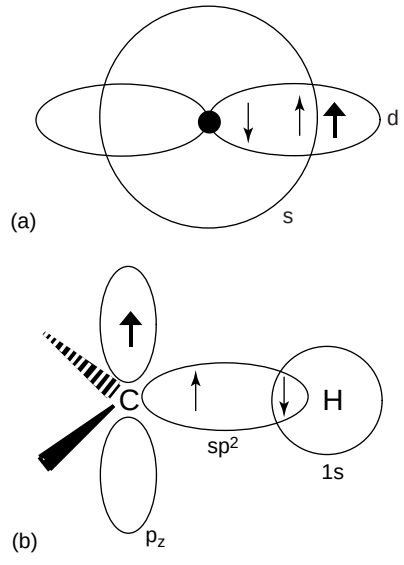

Figure 7 Schematic view of the spin polarization of s electrons by unpaired electrons in (a) d or (b) p orbitals (see text).

The case of the $C_6H_6^-$ is analogous to that described previously. We consider the isolated $-CH$ fragment with a π electron located in the carbon $2p_z$ orbital (Figure 7b). The π electron (with, say, α spin) causes the α spin in the C–H σ bond to be predominantly in its vicinity (as stated by Hund's rule), so that the β spin predominates in the vicinity of the proton. Therefore, although in the bond the spins are paired, the proton feels predominantly one spin, and there is a net interaction.

In the general case, both isotropic and anisotropic hyperfine interactions contribute to the experimental hyperfine tensor. The whole interaction is therefore dependent on orientation and must be expressed by a tensor. The effective spin-Hamiltonian for an $S = \frac{1}{2}$ system containing j nuclei with $I \neq 0$ thus becomes (Equation 24):

$$H = g\mu_B \boldsymbol{BS} + \sum_j \boldsymbol{IAS} + \sum_j g_n \beta_n \boldsymbol{BI} \qquad (24)$$

where $\boldsymbol{A}$ represents the total hyperfine tensor. The third term represents the nuclear Zeeman interaction which is included for the sake of completeness but, as already shown, does not influence the energy of the transitions.

The $\boldsymbol{A}$ tensor may be split into the isotropic and anisotropic parts as follows (Equation 25):

$$\boldsymbol{A} = \begin{matrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{matrix} = a_{iso} + \begin{matrix} T_1 & 0 & 0 \\ 0 & T_2 & 0 \\ 0 & 0 & T_3 \end{matrix} \qquad (25)$$

with $a_{iso} = (A_1 + A_2 + A_3)/3$. The anisotropic part of the $\boldsymbol{A}$ tensor corresponds to the dipolar interaction as expressed by the Hamiltonian in Equation (19). In a number of cases, the second term matrix of Equation (25) is a traceless tensor ($T_1 + T_2 + T_3 = 0$) and has the form ($-T$, $-T$, $+2T$). For this reason in the case of EPR spectra of paramagnetic species having both isotropic and anisotropic hyperfine coupling and undergoing rapid tumbling in a low-viscosity solution, the anisotropic term is averaged to zero and the observed hyperfine coupling corresponds to the isotropic part only (a_{iso}). Extremely useful information derived from the hyperfine structure of an EPR spectrum are the s and p characters of the orbital hosting the unpaired electron [or, in other words, the coefficient of the atomic wave functions involved in the semi-occupied molecular orbital (SOMO)]. These data can be deduced from Equations (26) and (27):

$$C_s^2 = \frac{a_{iso}}{A_0} \qquad (26)$$

$$C_p^2 = \frac{T}{B_0} \qquad (27)$$

where A_0 and B_0 are experimental or, more frequently, theoretical hyperfine coupling constants assuming pure

s and p orbitals for the element under consideration and $C_s^2 + C_p^2 + \cdots = 1$. For instance, the A_0 value for hydrogen is, in magnetic field units, 508 G.

3.2 Organic Radicals in Solution

As shown before for the EPR spectrum of the hydrogen atom, the effect of the hyperfine interaction with a single nucleus of $I = \frac{1}{2}$ is to split each electron energy level into two. In the general case, if the nuclear spin is I, then $2I + 1$ energy levels are present. Free organic radicals usually contain several nuclei having nuclear spin $I \neq 0$ (also called magnetic nuclei). The spectra of organic radicals in solution can be discussed in the same terms as those adopted for the H atom. This is because (1) the influence of the orbital momenta on the electron spin is negligible (see section 3.3) and the g value of organic radicals is close to the free electron value (g_e) and (2) the hyperfine interaction is reduced to the isotropic Fermi contact term since the anisotropic contribution is averaged to zero by molecular tumbling in solution. This gives rise to highly symmetric spectral structures.

The magnetic nuclei in an organic radical may be grouped into magnetically equivalent sets. Magnetically

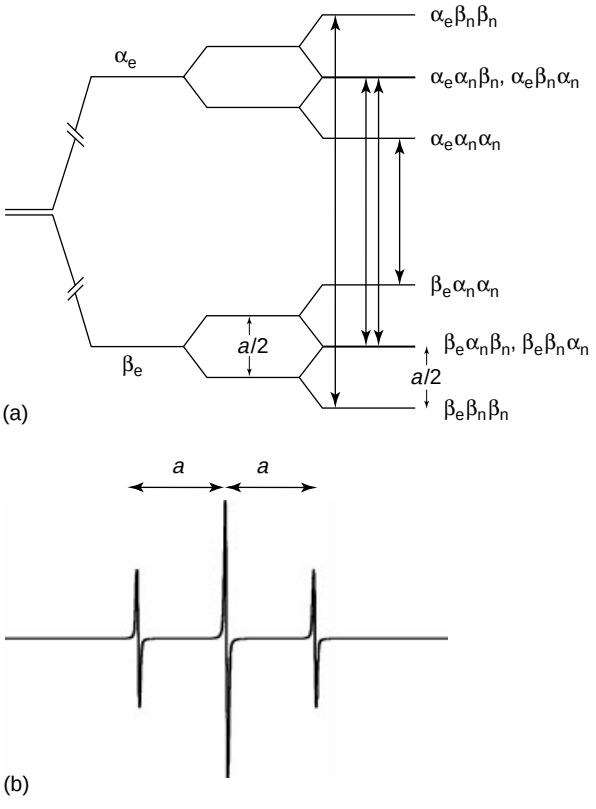

Figure 8 Energy levels scheme for the case of an unpaired electron ($S = \frac{1}{2}$) interacting with two equivalent H atoms (or two atoms having nuclear spin $I = \frac{1}{2}$). The observed spin transitions are evidenced and the related EPR spectrum is shown.

equivalent nuclei are nuclei which interact at the same extent with the unpaired electron. For example, the three H in the planar methyl CH_3 radical are a set of equivalent nuclei. EPR spectra of such systems may exhibit complex spectral patterns extremely rich in line components. In order to illustrate this situation, some examples are given, starting from the simple case of one unpaired electron interacting with two equivalent protons. The hyperfine energy levels may be obtained by successive splitting of levels as shown in Figure 8(a). This figure, in comparison with the realistic representation of the spin energy levels in the H atom (Figure 6), is obviously schematic. The interaction of the unpaired electron with the first nucleus causes the $a/2$ splitting of the two Zeeman electron levels. The interaction with the second nucleus causes each level to be split again by $a/2$ as magnetic equivalence implies that hyperfine splitting constants are identical. It is worth noting that in both $M_s = -\frac{1}{2}$ and $M_s = \frac{1}{2}$ (or α_e and β_e) groups, there is a twofold degeneracy of the levels with $M_I = 0$ associated with the two possible permutations of two nuclear spins giving rise to a zero net spin ($\alpha_n\beta_n$, $\beta_n\alpha_n$). This degeneracy is reflected in the 1 : 2 : 1 relative intensity of the allowed transitions and in the analogous 1 : 2 : 1 trend of the spectral intensities (Figure 8b). The three lines are separated by a, which is the value of the hyperfine coupling constant. In the general case for n equivalent nuclei with $I = \frac{1}{2}$, the degeneracies are just the coefficients in the binomial expansion of $(1 + x)^n$, which can readily be found from Pascal's triangle (Figure 9a). In order to depict this situation visually a collection of simulated spectra of radicals containing up to eight equivalent protons, all having the same coupling constant a, is presented in Figure 9(b).

In a large variety of cases one is dealing with radicals containing magnetically inequivalent protons. These in general exhibit different hyperfine splitting constants. The case of two inequivalent protons with hyperfine coupling constants $|a_1| > |a_2|$ is considered. A schematic energy-level diagram may be built up by considering firstly the $a_1/2$ hyperfine splitting due to the first proton and then each of these two levels to be split into two new levels separated by $a_2/2$. The scheme is shown, along with the corresponding EPR spectrum, in Figure 10(a) and (b).

The situation becomes much more complex when different sets of magnetically equivalent nuclei are present. In these cases the interpretation of the spectrum may not be straightforward, and isotopic labeling and computer simulation of the spectrum or double-resonance experiments, such as ENDOR, are sometimes necessary in order to assign successfully the large number of hyperfine lines. A beautiful example of such complexity is provided by the EPR spectrum of the radical cation [1.1.1.1]pagodane,[18] the structure of which is shown in Figure 11(a). The

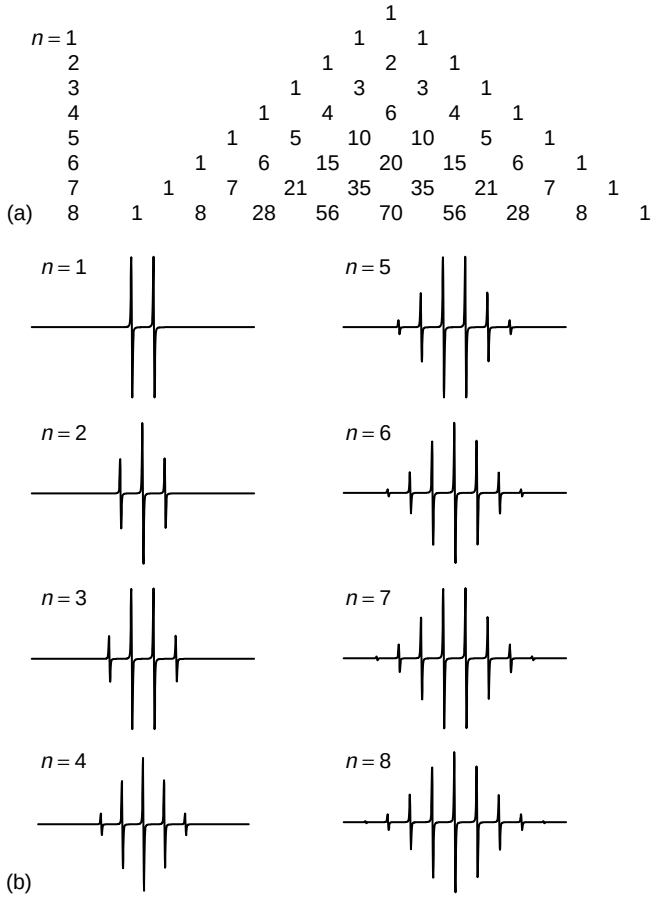

Figure 9 (a) Practical scheme to calculate the relative intensities of the EPR lines for organic radicals having a given number (n) of equivalent protons. The intensities are the coefficients of the binomial expansion $(1 + x)^n$. (b) Calculated spectra for organic radicals having from two to eight equivalent protons.

spectrum extends over 130 G (Figure 11b) and is composed of nine groups of lines, each group separated by 15.44 G, arising from the hyperfine interaction with eight equivalent protons. The splittings within the groups are ascribed to three sets of four equivalent protons, each with coupling constants of 1.17, 0.65 and 0.51 G. The total number of expected lines in this case is 1125.

3.3 The *g* Tensor

The magnetic moment of an electron due to its spin angular momentum given in Equation (1) is given by Equation (28):

$$\mu_s = -g_e \mu_B \boldsymbol{S} \tag{28}$$

where g_e is a dimensionless factor to which experiments ascribe the value 2.0023. If the magnitude of the magnetic moment associated with each unpaired electron in a paramagnetic system had this same value, then all

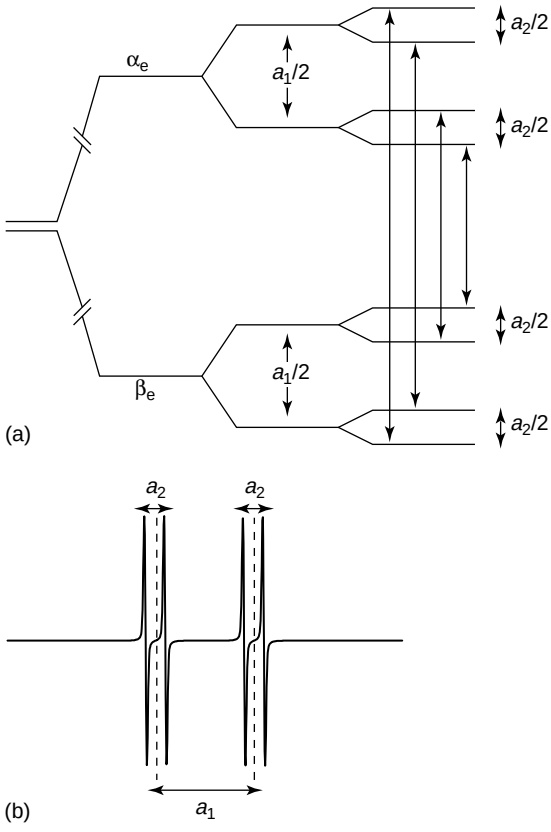

Figure 10 (a) Energy levels scheme for the case of an unpaired electron ($S = \frac{1}{2}$) interacting with two nonequivalent hydrogen atoms ($I = \frac{1}{2}$). The observed spin transitions are evidenced and the related EPR spectrum is shown in (b).

resonances would be expected to occur at the same field in a spectrometer operating at frequency ν according to the resonance condition (Equation 29):

$$B_{res} = \frac{h\nu}{g_e \mu_B} \tag{29}$$

However, in most experimental cases, the resonant field B_{exp} observed for a given signal determines a value for g (Equation 30):

$$g = \frac{h\nu}{B_{exp} \mu_B} \tag{30}$$

which is different from g_e and which is called the g factor. The difference between g and g_e is very small for free radicals but can be significant in the case of paramagnetic transition metal ions. To explain the origin of the g factor we have to consider the interaction between the unpaired electron and its environment. In particular, since the electron belongs to an orbital associated with the paramagnetic species, the interaction occurring between the spin $\boldsymbol{S}$ and the angular orbital momentum $\boldsymbol{L}$ must now be considered if the latter is different from zero.

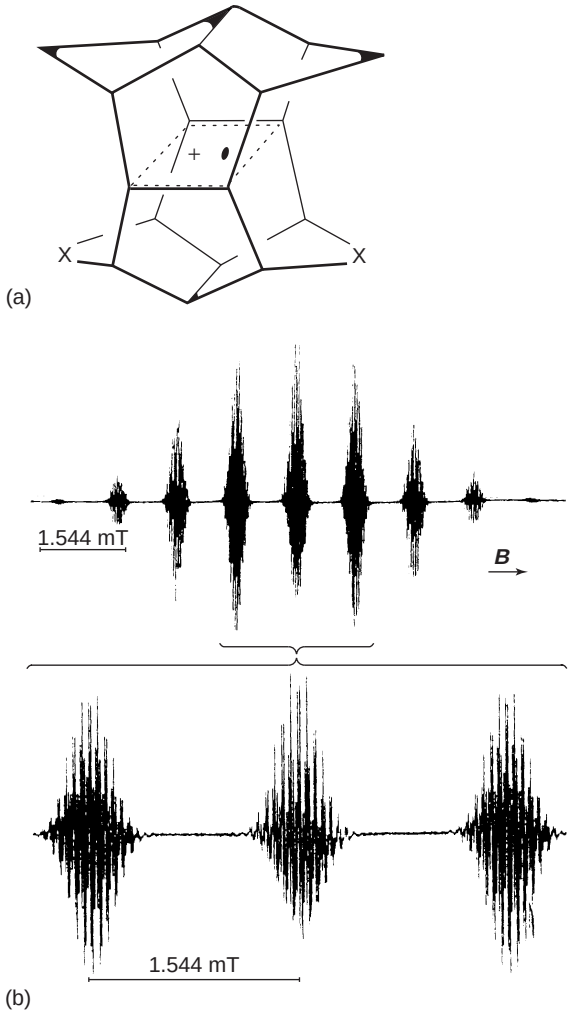

Figure 11 (a) A radical cation of the family of pagodane with $X = CH_2$ or $(CH_2)_2$ and (b) its EPR spectrum, showing the complete spectrum above and a magnification of the central part below. (Reproduced from Prinzbach et al.[18] with permission of VCH-Verlagsgesellschaft mbH.)

This interaction causes a deviation of the g value from 2.0023. Actually, for a series of reasons related to the electric fields present in molecules and solid systems, the orbital angular momentum $\boldsymbol{L}$ is quenched and its contribution to the total angular momentum should be zero in most cases and not only for S states (i.e. those states for which $\boldsymbol{L} = 0$). However, $\boldsymbol{L}$ is partially restored by a magnetic interaction that operates via an admixture of excited electronic states to the ground state. This interaction is called the spin–orbit coupling. The role of the partially restored $\boldsymbol{L}$ in the system containing the unpaired electron can be visualized by making the electron itself (which is orbiting a nucleus of charge Z_e) the origin of the coordinate system. In these terms the electron sees a positive charge moving about it and, as a result of this positive current, a magnetic moment originates at the electron position. This means that the electron spin magnetic moment interacts with a local magnetic field arising from its own orbital motion which, in the EPR experiment, is added to the external field. The true magnetic field experienced by the electron thus differs from the applied field.

The energy of the spin–orbit interaction is approximately given by the Hamiltonian (Equation 31)

$$H_{s-o} = \lambda \boldsymbol{LS} \tag{31}$$

where λ is the spin–orbit coupling constant, which depends strongly on the atomic number Z of the atom. The strength of the spin–orbit interaction increases with increasing Z. The value of λ can be estimated experimentally from the fine structure of atomic spectra. The whole spin Hamiltonian for a $S = \frac{1}{2}$ nondegenerate electronic ground state with zero nuclear moments exclusively ($\mu_n = 0$) can thus be written according to Equation (32):

$$H = g_e \mu_B \boldsymbol{BS} + \mu_B \boldsymbol{BL} + \lambda \boldsymbol{LS} \tag{32}$$

where the second term accounts for the energy of the interaction between the orbital angular momentum $\boldsymbol{L}$ and the external field (orbital Zeeman interaction). The previous Hamiltonian reduces to (Equation 33):

$$H = \mu_B \boldsymbol{BgS} \tag{33}$$

where g is now a symmetric (3×3) matrix (or a second-rank tensor), indicating that the Zeeman interaction is anisotropic. The described transformation of the scalar g_e value (2.0023) into a $\boldsymbol{g}$ matrix does not hold, of course, for $\boldsymbol{L} = 0$ systems (or s states such as the hydrogen atom). The first consequence of the tensorial nature of $\boldsymbol{g}$ is that the magnetic moment of an unpaired electron in a real chemical system is not, in general, exactly antiparallel to the effective spin, and its magnitude is not simply that of a free electron but depends on the orientation of the system in the magnetic field. This can be expressed by Equation (34):

$$\mu_s = -\mu_B \boldsymbol{gS} \tag{34}$$

which formally clarifies how the (3×3) $\boldsymbol{g}$ matrix acts on the (3×1) $\boldsymbol{S}$ vector. In any arbitrary coordinate system fixed in the crystal the $\boldsymbol{g}$ tensor has the general form

$$\begin{matrix} g'_{xx} & g'_{xy} & g'_{xz} \\ g'_{yx} & g'_{yy} & g'_{yz} \\ g'_{zx} & g'_{yz} & g'_{zz} \end{matrix} \tag{35}$$

and is almost always symmetrical ($g'_{zx} = g'_{yz}$). In experimental terms (for a sample constituted by a single crystal having all the paramagnetic centers similarly orientated

with respect to the crystal axes), the presence of a $\boldsymbol{g}$ matrix is the cause of the dependence of the resonant field on the orientation of the crystal axes in the magnetic field. This is illustrated in Figure 12(a–d), where the simulated spectra of a paramagnetic center (with $I = \frac{3}{2}$) in a single crystal are reported for four different orientations of the paramagnetic center axis in the field. The anisotropy of $\boldsymbol{g}$ is reflected in the different positions of the resonant fields for the various orientations and the anisotropy of A in the different hyperfine separations between the four spectral lines. The example shown in Figure 12(a–d) is an ideal case in fact, because in real systems the orientation of the paramagnetic species axes is, in principle, unknown.

The structure of the $\boldsymbol{g}$ tensor is therefore derived by running the EPR spectrum several times for various orientations of the crystal axes with respect to the magnetic field (Figure 13). The (3×3) $\boldsymbol{g}$ matrix in Equation (35) is thus built up and then can be diagonalized by a mathematical operation that allows one to obtain the principal values of the tensor (Equation 36):

$$\begin{matrix} g_{xx} & 0 & 0 \\ 0 & g_{yy} & 0 \\ 0 & 0 & g_{zz} \end{matrix} \tag{36}$$

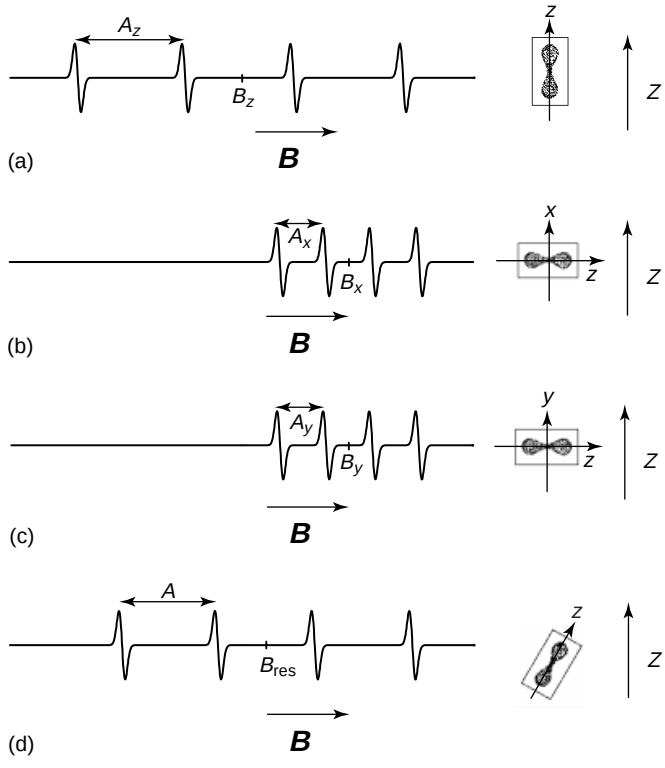

Figure 12 Simulated four-line EPR spectra of an $S = \frac{1}{2}, I = \frac{3}{2}$ system in a single crystal for different orientations of the crystal in the magnetic field. In the scheme the directions of the radical axes and of the crystal axes are arbitrarily assumed to be the same.

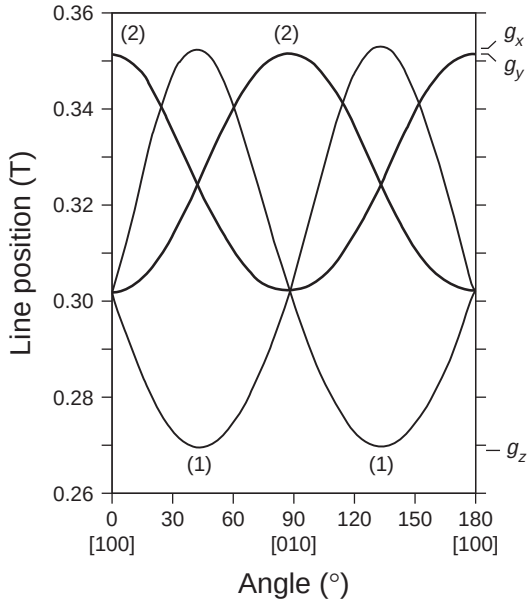

Figure 13 EPR resonant fields as a function of the orientation of a single crystal. The species has an axial $\boldsymbol{g}$ tensor. (Reproduced from J.A. Weil, J.R. Bolton and J.E. Wertz, *Electron Paramagnetic Resonance, Elementary Theory and Practical Applications*, Wiley-Interscience, New York, 92, 1994. Copyright © 1994, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

This operation is equivalent to a rotation of the Cartesian axes, after which the axes of the tensor (often, but not always, coinciding with molecular axes) and the laboratory axes coincide. In the first three cases in Figure 12(a–c) the direction of the external magnetic field coincides with the three principal axes of the paramagnetic system, two of which (Figure 12b and c) are equivalent and give rise to two coincident EPR spectra. The reported system is said to have an axial symmetry. The orientation in Figure 12(a) is called ''parallel'' and those in Figure 12(b) and (c) are called ''perpendicular''. The three values $g_{zz} = g_{\parallel}$, $g_{xx} = g_{yy} = g_{\perp}$ derived from the resonant fields $B_z, B_x = B_y (g_{ii} = h\nu/\mu_B B_i$, Equation 8) and corresponding to the principal orientations (Figure 12a–c) are the diagonal elements of the matrix in Equation (36). In the more general case of rhombic symmetry, the three diagonal elements of the tensor have distinct values. The principal g values of a given system can be evaluated theoretically, in terms of the deviation from g_e caused by the spin–orbit interaction, using perturbation theory and considering the second and third terms of Equation (32) as perturbations of the spin-Hamiltonian. In this way it can be shown that a given element of the $\boldsymbol{g}$ tensor can be expressed by Equation (37):

$$g_{ij} = g_e \delta_{ij} + 2\lambda \sum_{n=0} \frac{\langle 0|L_i|n\rangle\langle n|L_j|0\rangle}{E_n - E_0} \tag{37}$$

where δ_{ij} is the Kronecker delta ($\delta = 0$ for $i \neq j$ and $\delta = 1$ for $i = j$), L_i and L_j are the components of the orbital angular momentum operator, 0 represents the ground state and n the different excited states and E_0 and E_n are the energies of the ground state and the excited states, respectively. The sum of the diagonal elements of a tensor is called the trace and this quantity is invariant under any rotational transformation of the axis system. This means that for a sample in highly fluid solution, where it undergoes rapid and random tumbling, the position of the resonance is determined by the average of the diagonal elements of the $\boldsymbol{g}$ tensor. In this case the observed g value is therefore given by Equation (38):

$$g_{\text{av}} = \frac{g_{xx} + g_{yy} + g_{zz}}{3} \tag{38}$$

The absolute determination of the g values may, in principle, be carried out by independent and simultaneous measurements of B and ν, using a gauss meter and a frequency meter, respectively, according to Equation (30).

In practice, the g value can be also determined by comparing the field values at resonance for the sample investigated and that of a reference sample. As one can write (Equation 39)

$$h\nu = g_{\text{ref}}\mu_{\text{B}}B_{\text{ref}} = g\mu_{\text{B}}B \tag{39}$$

provided that ν is invariant during the whole experiment, g is given by Equation (40):

$$g = g_{\text{ref}}\frac{B_{\text{ref}}}{B} \tag{40}$$

The usual reference samples are diphenylpicrylhydrazyl (DPPH) ($g = 2.0036$), Varian Pitch ($g = 2.0028$) and Cr^{3+} in a MgO matrix ($g = 1.9797$). These reference samples give rise to narrow lines necessary for accurate determinations. The reference sample could be placed in one of the two compartments of a dual cavity or be stuck to the investigated sample quartz tube in the case of a single cavity.

Summarizing the concepts introduced in the present section, we have to recall that, measuring the position of the resonant field in an EPR spectrum, a g factor ($g \neq g_{\text{e}}$) is determined (Equation 30) which depends on the electronic structure of the paramagnetic species and which is somehow equivalent to the chemical shift in NMR. The shift of g from g_{e} is due to partial contributions of the orbital angular momentum which take place through the spin–orbit coupling, and can be calculated as a perturbation of the spin-Hamiltonian (Equation 37). For spectra in the solid state the anisotropy of the magnetic interaction between $\boldsymbol{B}$ and $\boldsymbol{S}$ (the latter ''modified'' by the spin–orbit coupling) shows up in terms of different values of g recorded for different orientations of the crystal in the magnetic field (Figures 12a–d and 13) and is formally expressed by a $\boldsymbol{g}$ tensor (Equation 35) whose elements range between three limiting values (g_{xx}, g_{yy}, g_{zz}) which are the principal values of the same tensor in diagonal form (Equation 36). The cases of isotropic, axial and rhombic symmetries have been described for which all the axes of g and the hyperfine tensor coincide. Cases with lower symmetry (not treated here for the sake of brevity) are the monoclinic and triclinic cases. In the monoclinic symmetry one axis of g and one of A are coincident, whereas in the triclinic symmetry there is complete noncoincidence of all axes. For spectra recorded in low-viscosity solutions a unique g factor is observed, owing to the rapid molecular tumbling which is the average of the three principal values (g_{av}). In Table 2 the g values of some $S = \frac{1}{2}$ paramagnetic species are reported.

3.4 Systems with $S > \frac{1}{2}$: the Fine Structure

The previous sections were devoted to the discussion of $S = \frac{1}{2}$ systems, i.e. of molecules or ions having one unpaired electron only. However, several paramagnetic systems have two, three or more unpaired electrons and consequently $S > \frac{1}{2}$. The relevant paramagnetic entities with $S > \frac{1}{2}$ are biradicals, molecules in excited triplet ($S = 1$) state, molecules in triplet, quadruplet ($S = \frac{3}{2}$) or quintuplet ($S = 2$) ground states, transition metal ions. The effect of the external magnetic field on the spin levels of an $S > \frac{1}{2}$ system is the splitting into $2S + 1$ Zeeman sublevels characterized by different M_{s} values. The Zeeman splitting increases with increasing B as in the $S = \frac{1}{2}$ case. The selection rule $\Delta M_{\text{s}} = \pm 1$ is still valid and produces $2S$ transitions. The position of the absorption

Table 2 Typical g values of some organic and inorganic radical species

Radical	g_{av}
$MoO(SCN)_5^{2-}$	1.935
$VO(acac)_2$	1.968
$(Cp)_2TiCl_2AlCl_2$	1.975
CO_2^-	2.0007
e^-	2.0023
CH_3	2.0026
$C_{14}H_{10}$ cation	2.0028
$C_{14}H_{10}$ anion	2.0029
R–O	2.0052
SO_2^-	2.0056
$\{(CH_3)_3C\}_2NO$	2.0063
R–S	2.0103
HO_2	2.014
Cl_2^-	2.030
$Cu(acac)_2$	2.13

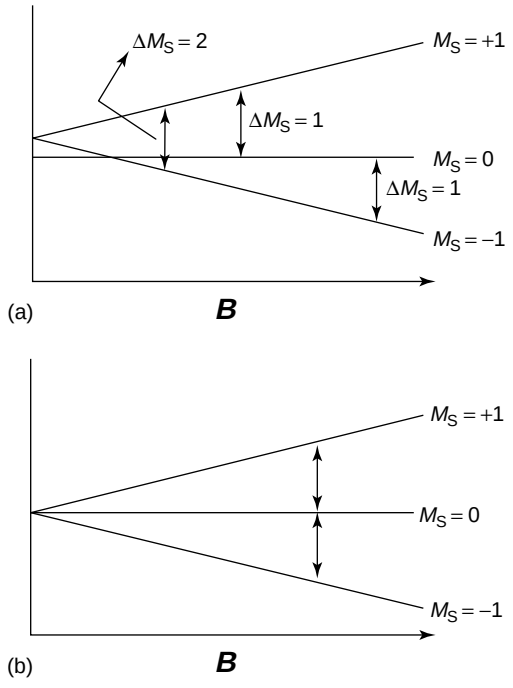

Figure 14 Energy levels for an $S = 1$, two electron spin system. In (a) there is a ZFS interaction and in (b) the ZFS interaction is zero.

lines, however, can vary markedly according to the extent of the so-called zero field splitting (ZFS) term. This term is a measure of the energy difference between the magnetic sublevels that occur in the absence of an external magnetic field and originates from the interaction between the magnetic moments of the various unpaired electrons. The interaction can be direct (for instance in the case of organic molecules in the triplet state) or can occur via spin–orbit coupling. The situation is illustrated in Figure 14(a) and (b) for the case of an $S = 1$ system. The separation between $M_s = \pm 1$ and $M_s = 0$ in the absence of a magnetic field is the ZFS. In terms of the spin-Hamiltonian the interaction between the two electron spins is expressed by Equation (41):

$$H = \boldsymbol{SDS} \tag{41}$$

where $\boldsymbol{D}$ is again a (3×3) traceless matrix analogous to the $\boldsymbol{A}$ tensor (Equation 25) defined for the anisotropic hyperfine coupling. Also in the case of the electron–electron interaction (or fine interaction) the extent of the coupling varies as a function of the orientation of the paramagnetic center in the external field. Owing to some mixing between the various magnetic states, the $\Delta M_s = \pm 2$ transition (in principle forbidden) becomes partially allowed and appears, usually with weak intensity, at low field (Figure 14a).

In particular cases of high symmetry of the environment, the zero field separation for the various

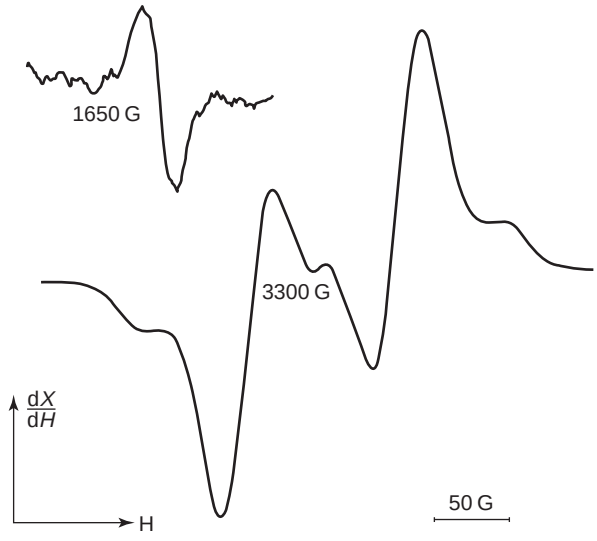

Figure 15 EPR spectrum of a system in the triplet state. The species is the diphenylamino radical dimer in rigid glass at 77 K. The line in the inset at 1650 G is the half-field $\Delta M_s = 2$ transition. (Reproduced with permission from Gordy.[4])

magnetic sublevels vanishes. The situation is reported in Figure 14(b) for $S = 1$. In this case the two allowed transitions occur at the same magnetic field value, resulting in a single resonance line. An experimental spectrum of a system in triplet state is reported in Figure 15.[4]

4 SPECTRA OF POLYCRYSTALLINE MATERIALS (POWDER SPECTRA)

In the previous section the role of anisotropy of both the electron–Zeeman and electron–nucleus hyperfine interactions was introduced and briefly discussed in the case of paramagnetic centers confined in a single crystal and therefore all having the same orientation. This was done because in the single crystal case (Figures 12a–d and 13) the main consequence of anisotropy (spectral features depending on sample orientation) clearly shows up. The very major fraction of the EPR spectra recorded in the solid state, however, does not come from single crystals but from polycrystalline materials or powders whose spectra are usually called ''powder'' spectra. As many chemical systems of interest are polycrystalline or cannot be easily prepared as single crystals, the practical interest in powder EPR spectra is very high. Powder spectra are recorded for various real chemical systems such as glasses, ceramics, catalysts, minerals, micro- and nanoporous materials and coordination compounds. Particularly important examples of powder spectra are obtained in the case of frozen solutions which are polycrystalline materials. The comparison of a solution

spectrum with the corresponding spectrum of the frozen solution is usually very fruitful for the determination of the properties of the paramagnetic compound.

4.1 Line Shape of Simple Powder Spectra

A polycrystalline material is constituted by a large number of microcrystals randomly oriented in space so that all possible orientations of one crystal in the external field are equally probable. Since the sample is solid (the paramagnetic centers do not have the typical motion of a molecule in solution and in many cases can be considered immobile), the anisotropy of the various magnetic interactions influences the spectra and the values are not averaged as in the solution case. However, the experimental spectra, owing to the presence of all possible orientations of the microcrystals, are invariant to rotation and are basically constituted by an envelope of the single crystal spectra corresponding to all possible crystal orientations. Let us consider first the simple case of an $S = \frac{1}{2}$ paramagnetic center with no hyperfine splitting. The EPR spectrum is spread over the entire resonant field range ($\Delta B = B_{\min} - B_{\max}$) determined by the values of the principal $\boldsymbol{g}$ components. The intensity of the powder spectrum, however, is not uniform over all the ΔB range as shown in Figure 16(a) and (b), in the case of an axial $\boldsymbol{g}$ tensor ($g_{\parallel} > g_{\perp}$) and in that of a rhombic tensor ($g_{xx} \neq g_{yy} \neq g_{zz}$), respectively. The nonuniformity of the intensity can be understood, in practical terms, taking into account (e.g. in the axial case) that, as $g_{xx} = g_{yy} = g_{\perp}$, the probability of having a microcrystal oriented in one of the two equivalent x and y directions is higher than the probability of an orientation along z. The relatively simple powder spectra lines reported in Figure 16(a) and (b) can become rather complex in the presence of hyperfine interactions. The whole shape of the hyperfine powder spectrum depends on various factors: the nuclear spin I of the magnetic nucleus (or nuclei) interacting with the unpaired electron; the isotopic abundance of the magnetic nucleus; the symmetry of the hyperfine tensor (axial, rhombic, etc.) and the values of the various elements of the tensor itself.

Figures 17(a–c) and 18(a–c) illustrate some simulated spectral profiles due to species with axial $\boldsymbol{g}$ tensors (Figure 17a–c) and rhombic $\boldsymbol{g}$ tensors (Figure 18a–c) having nuclei with a given I and different hyperfine constant values.

4.2 Powder Spectra of Some $S = \frac{1}{2}$ Transition Metal Ions

EPR spectroscopy has played a paramount role in the study of transition metal ions existing in inorganic molecular compounds, biologically important molecules such as metalloenzymes and solid materials. The interpretation

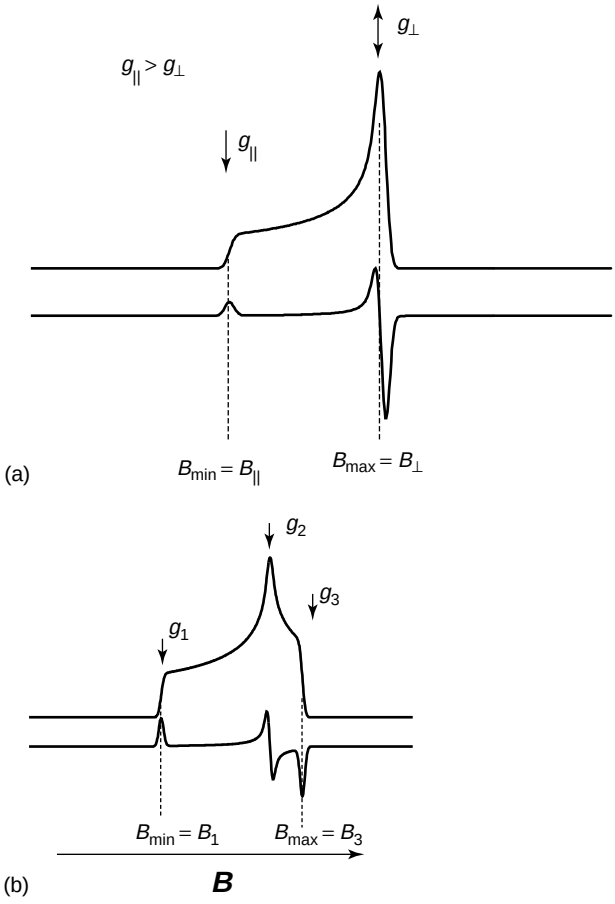

Figure 16 Calculated profiles (absorption and first derivative of absorption) of the powder spectra in the case of an axial (a) and a rhombic (b) $\boldsymbol{g}$ tensor in the absence of hyperfine interactions.

of the spectra for paramagnetic transition ions having $S > \frac{1}{2}$ can be extremely complex and will not be discussed here. The reader is therefore referred to relevant books that treat this subject.[6–9] We shall limit ourselves to the description of the features of some $S = \frac{1}{2}$ transition metal ions to illustrate with practical examples the role of the factors listed at the end of section 4.1 in determining the shape of the spectra. Let us consider the case of nd^1 ions such as Ti^{3+}, V^{4+}, Cr^{5+} and Mo^{5+}, all having a single electron in the d orbitals. We shall examine first the structure of the $\boldsymbol{g}$ tensor. These ions in coordination compounds or solid materials usually exhibit either octahedral or tetrahedral symmetry which splits the fivefold degenerate state of the free ion into two states, t_{2g} and e_g, which are respectively triply and doubly degenerate (Figure 19a and b) The degeneracy of these states is further lifted by other effects, leading to structural distortions.

Two common cases which are relevant to EPR spectroscopy are illustrated in Figure 19(a) and (b), namely the tetragonal distortions by compression of

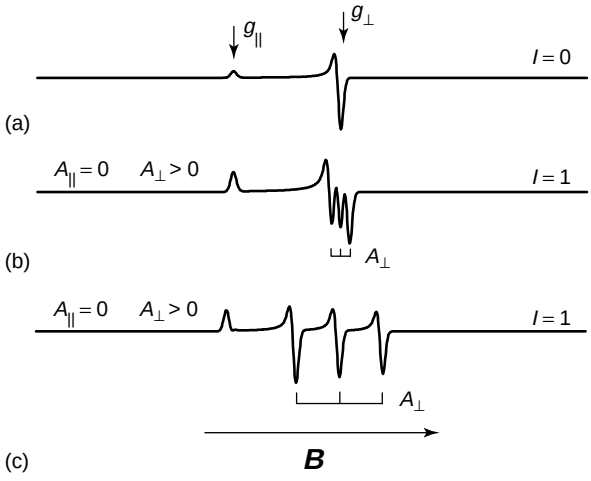

Figure 17 Calculated profiles of powder spectra of a species with axial $\boldsymbol{g}$ tensor: (a) no hyperfine interaction; (b) and (c) two cases with hyperfine structure due to one $I = 1$ nucleus. The perpendicular component only shows appreciable hyperfine coupling.

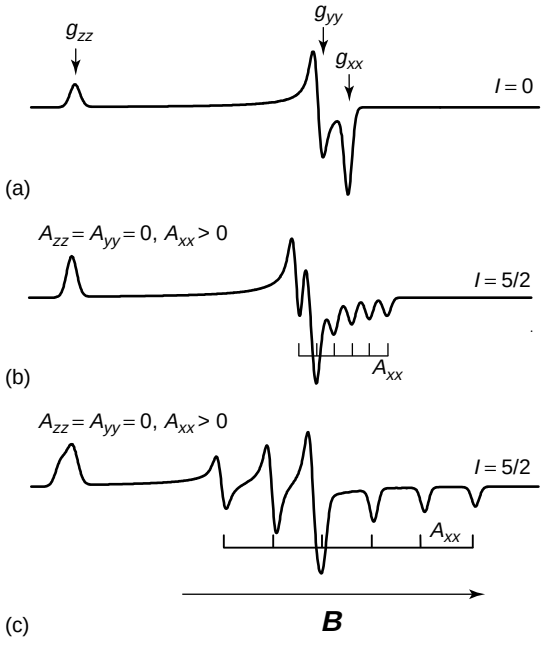

Figure 18 Calculated profiles of powder spectra of a species with rhombic $\boldsymbol{g}$ tensor: (a) no hyperfine interaction; (b) and (c) two cases with hyperfine structure due to one $I = \frac{5}{2}$ nucleus. The only x component shows appreciable coupling.

(a) tetrahedral and (b) octahedral structures. The latter, for example, is the typical structure of vanadyl (VO^{2+}) compounds. The $\boldsymbol{g}$ values in the two cases are evaluated on the basis of the perturbation theory according to Equation (37).

For tetragonally distorted tetrahedral crystal fields (Figure 19a) one has

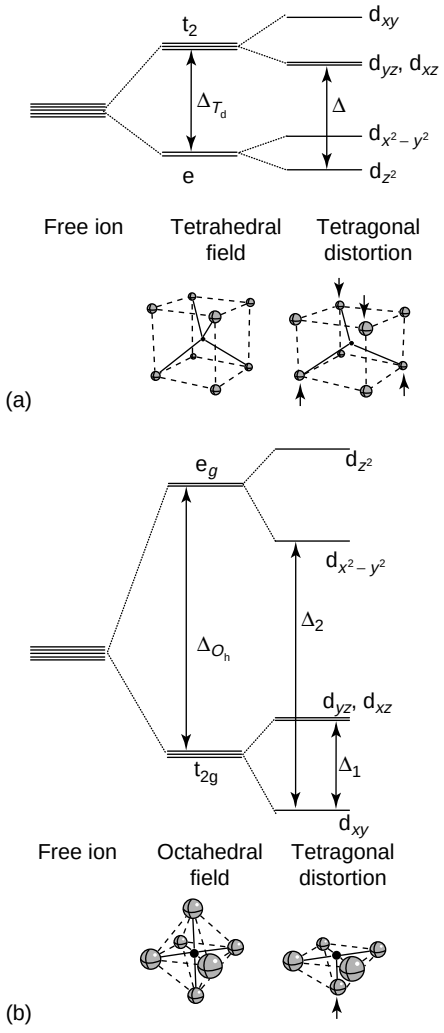

Figure 19 (a) d-Orbital energy levels for (from left to right) a free transition metal ion, the same ion in tetrahedral crystal field and undergoing a tetragonal distortion by compression along an axis. (b) d-Orbital energy levels for (from left to right) a free transition metal ion, the same ion in octahedral crystal field and undergoing a tetragonal distortion along the z-axis.

$$g_{zz} = g_{\parallel} = g_{\mathrm{e}} \tag{42}$$

$$g_{xx} = g_{yy} = g_{\perp} = g_{\mathrm{e}} - \frac{6\lambda}{\Delta} \tag{43}$$

whereas in the case of tetragonally distorted octahedral symmetry (Figure 19b) one has

$$g_{xx} = g_{yy} = g_{\perp} = g_{\mathrm{e}} - \frac{2\lambda}{\Delta_1} \tag{44}$$

$$g_{zz} = g_{\parallel} = g_{\mathrm{e}} - \frac{8\lambda}{\Delta_2} \tag{45}$$

Simulated spectra for these two cases (in the absence of hyperfine interactions) are reported in Figure 20, showing

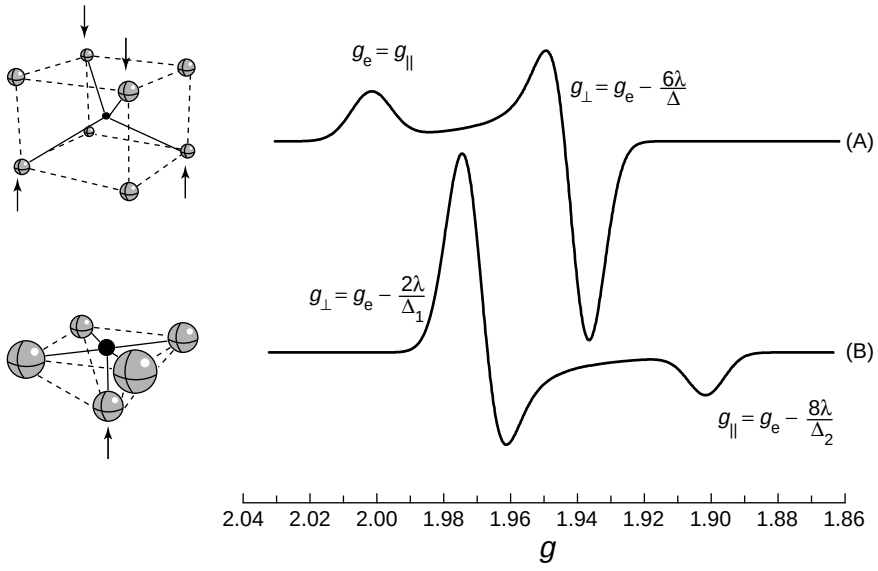

Figure 20 Schematic simulated EPR spectra of a $3d^1$ ion with no $I \neq 0$ nuclei in distorted (A) tetrahedral and (B) octahedral crystal fields.

that EPR provides information not only on the nature of the paramagnetic ions but also on their coordination and on structural parameters such as the various crystal field energy splittings Δ. The schematic spectra shown in Figure 20 are amenable to the case of Ti^{3+} which, owing to the low abundance of the $I = \frac{7}{2}$ ^{49}Ti nucleus, does not visually show any hyperfine structure or, at least, exhibits a very weak one. A real experimental spectrum of Ti^{3+} in tetragonal distorted octahedral coordination is shown in Figure 21, together with its computer simulation. Ti^{3+} ions are obtained by reducing Ti^{4+} ions embedded in the matrix of an ETS-10 zeolite (a molecular sieve of synthetic origin) where they are coordinated by six oxygen ions. The features of the spectrum, except for the line width, correspond to those in the schematic profile B in Figure 20. Traces of the hyperfine structure due to ^{49}Ti are visible at the left-hand side of the spectrum. Mo^{5+} and V^{4+} ions have the same d^1 electronic structure as Ti^{3+} and therefore, in a corresponding coordinative environment, also exhibit the same structure of the ***g*** tensor. The spectral profile under these conditions, however, is complicated by the presence of hyperfine structures due to magnetic nuclei. These are ^{95}Mo and ^{97}Mo $I = \frac{5}{2}$ (overall abundance 25.3%) in the former case and ^{51}V ($I = \frac{7}{2}$, abundance about 100%) in the latter. The spectra (and corresponding simulations) of Mo^{5+} and V^{4+} ions both stabilized at the surface of TiO_2 in two different heterogeneous catalysts are shown in Figures 22 and 23, respectively. The stick diagrams in the figures evidence the hyperfine structure which is (partially) resolved for the perpendicular component only in the case of Mo^{5+} (the parallel structure is buried in the broad $g_{||}$ line). The high abundance of ^{51}V determines the complex profile of the V^{4+} spectrum (Figure 23) where the two eight-line hyperfine structures ($I = \frac{7}{2}$, number of lines $2I + 1 = 8$) are widely overlapped. Furthermore, for second-order effects not discussed here, the spacing between the hyperfine lines is not exactly the same, and grows from low to high magnetic field. Comparison of Figures 21, 22 and 23 illustrates the complexity that can be attained by a powder EPR spectrum and outlines the need for spectral simulations to determine exactly the spin-Hamiltonian parameters (***g*** and ***A*** tensors).

4.3 A Pragmatic Approach to Powder Spectra Interpretation

In sections 4.1 and 4.2 the intrinsic complexity of the EPR powder spectra profile has been shown even in the case of model spectra (Figures 17a–c and 18a–c). The real, experimental spectra are usually complicated by additional effects such as the presence in the sample of various distinct species having different parameters, the presence of various nuclei with different nuclear spins in the same species and the broadening of lines due to spin–spin interaction or motional phenomena causing loss of resolution. For the above-mentioned reasons the evaluation of the spin-Hamiltonian parameters in the case of powder spectra is often performed by adopting additional techniques that may help in the assignments. The main additional tools which are employed are isotopic labeling, the multifrequency approach and the systematic use of computer simulation of experimental spectra. All

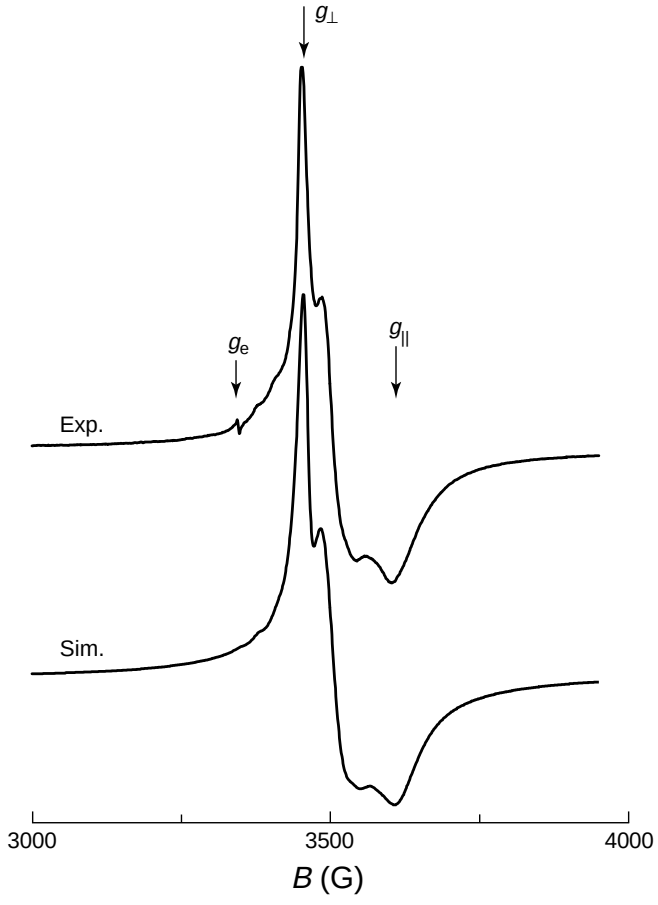

Figure 21 Experimental and simulated EPR spectra of Ti^{3+} ions ($3d^1$) in the octahedrally distorted crystal field of a molecular sieve.

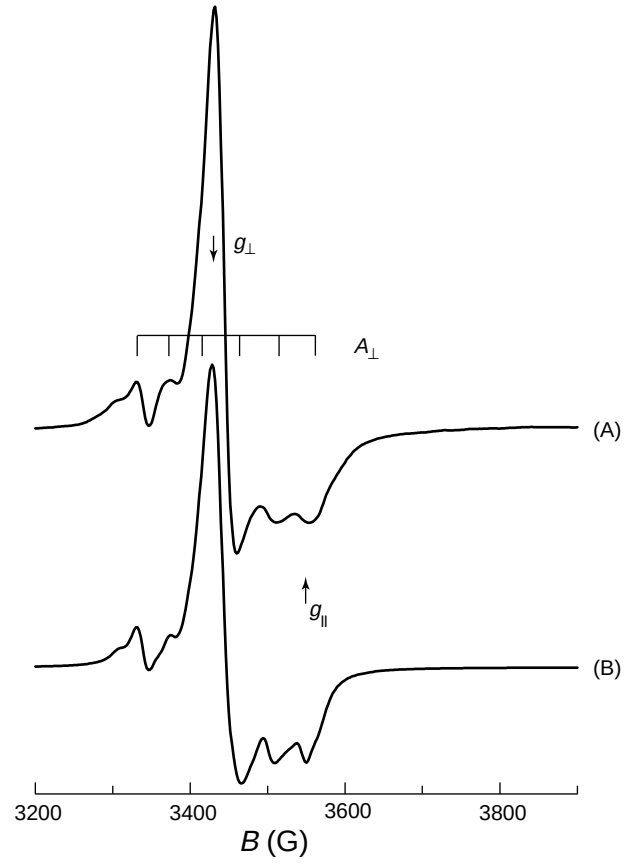

Figure 22 Experimental and simulated EPR spectra of Mo^{5+} ($3d^1$) ions at the surface of an MoO_3/TiO_2 catalyst. The stick diagram evidences the hyperfine structure due to the 25% abundant $I = \frac{5}{2}$ ^{95}Mo and ^{97}Mo nuclei.

these techniques will be briefly discussed in the following sections.

4.3.1 Isotopic Labeling

The isotopic labeling technique is employed in EPR for at least two distinct reasons. In the first case, isotopes having nuclear spin $I = 0$ are substituted in the paramagnetic center by other isotopes with $I \neq 0$. In this case it has to be mentioned, for the sake of clarity, that the isotopic labeling is not used to help in the assignment of a complex spectrum, as indicated at the beginning of this section, but to obtain a hyperfine structure and the related information (nature of the atoms interacting with the unpaired electron, extent of this interaction) otherwise absent if $I = 0$. A typical case is that of oxygen radical species obtained using reactants enriched with ^{17}O ($I = \frac{5}{2}$), since natural oxygen contains mainly ^{16}O with $I = 0$. In the second case the isotopic labelling is indeed used to elucidate the structure of complex signals by comparing spectra of species having different isotopic composition. An example of this technique is reported in Figure 24 where (upper spectrum) the complex X-band signal of powdered MgO_2 obtained by reaction of magnesium oxide with H_2O_2 is reported.[19] The signal is complex and is due to several oxygen-based paramagnetic species embedded in the solid matrix. However, two of the spectral lines (namely A and B, centered at $g = 2.0137$) were thought to be due to hyperfine interaction of an unpaired electron and a ^1H nucleus both belonging to an OH radical. The combined use of isotopic labeling, the multifrequency approach and computer simulation, has been determinant in performing this assignment.

The substitution of ^1H by ^2H (using D_2O_2 instead of H_2O_2 in MgO_2 synthesis) leads to the lower spectrum in Figure 24 where the A–B pair (57 G separation) transforms into a triplet of about 9 G separation (partially buried in the intense central line). This is because ^1H has $I = \frac{1}{2}$ and two lines ($2I + 1$) are expected for the interaction of the unpaired electron with one proton whereas, in the case of deuterium, $I = 1$ and a three-line structure therefore arises. The hyperfine constant (i.e. the separation between the hyperfine lines) is in

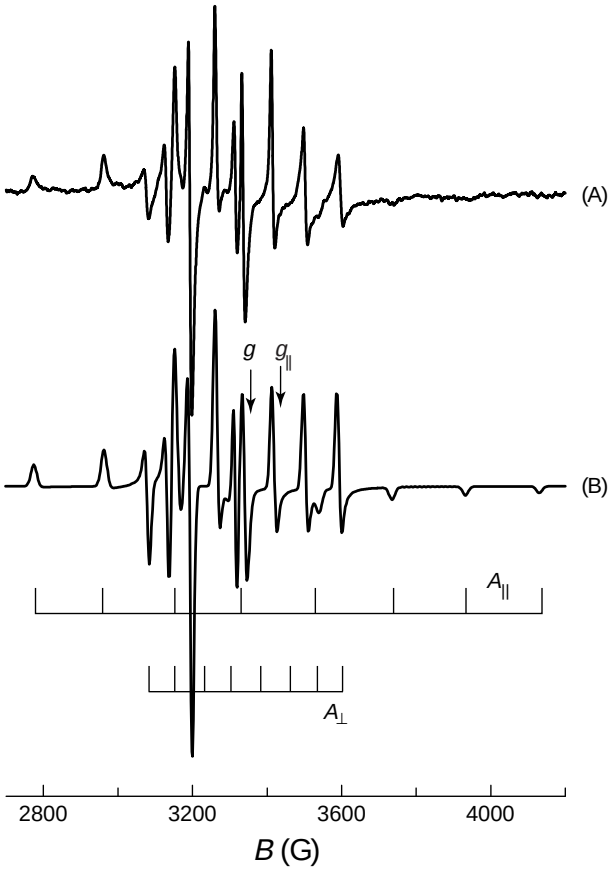

Figure 23 Experimental and simulated EPR spectra of V^{4+} ($3d^1$) ions at the surface of a V_2O_5/TiO_2 catalyst. The stick diagram evidences the hyperfine structure due to the 100% abundant $I = \frac{7}{2}$ ^{51}V nucleus.

the case of deuterium about 6.5 times smaller than that recorded for ^1H. This is due to the different values of the nuclear factor g_N (Equation 13) for the two nuclei whose ratio $g_n(H)/g_n(D)$ is indeed 6.51. The use of deuterium instead of hydrogen has therefore allowed the assignment of two of the lines in the complex spectral envelope of Figure 24 to an H hyperfine structure due to one H nucleus.

4.3.2 Multifrequency Approach

The multifrequency approach is an excellent tool for understanding spectra of powder solids in which anisotropies of the tensor give rise to considerable complexity. The principle of this approach is based on the fact that the magnetic field separation ΔB between two lines due to anisotropy of the $\boldsymbol{g}$ tensor varies linearly with the microwave frequency according to Equation (46):

$$\Delta B = \frac{h\nu}{\mu_B}\left(\frac{1}{g_1} - \frac{1}{g_2}\right. \qquad (46)$$

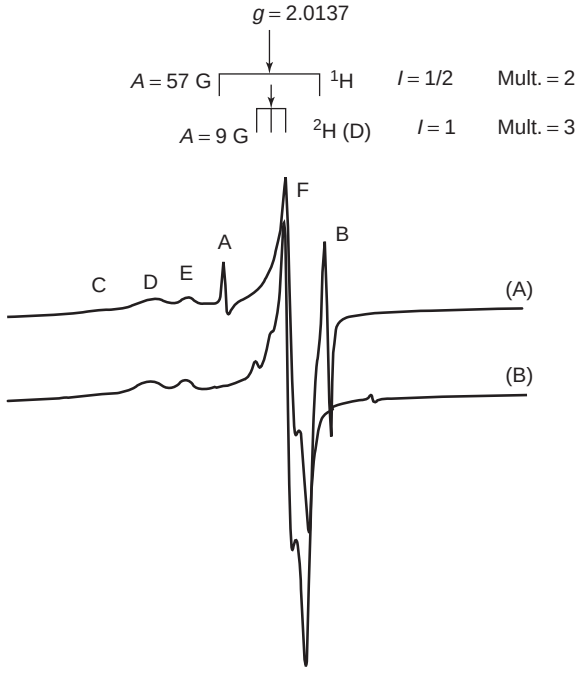

Figure 24 X-band ($\nu = 9.5$ GHz) powder spectra of various oxygen radicals trapped in a solid MgO_2 matrix obtained by contacting MgO with (A) hydrogen peroxide and (B) deuterium peroxide. The main lines due to the trapped hydroxyl (OH) or deuteroxyl (OD) radicals are evidenced by the stick diagram.

which is directly derived from Equation (29). The same does not apply to the separation of hyperfine lines arising from the interaction between the nuclear spin and the unpaired electron, which is unaffected by microwave frequency changes (see Equation 23). The use of a high-frequency mode, for instance, instead of a classical X-band mode, increases, for a given paramagnetic center, the magnetic field separation of the various g components, but does not influence the hyperfine line separation. This effect is very useful in the case of X-band spectra with overlapping hyperfine structures whose features can be better resolved by running a high-frequency spectrum. The most commonly used bands (alternative to the classical X-band at about 9.5 GHz) are the low-frequency S band (see Table 1), the Q-band at about 35 GHz and the W band (94 GHz), which has been commercially available since the mid-1990s. The Q-band spectrum of the system described in the previous section (OH radical and other oxygen radicals in MgO_2) is reported in Figure 25 together with its computer simulation.[19] The spectrum in Figure 25 confirms the existence of a doublet of hyperfine lines at $g = 2.0137$ separated by 57 G, but reveals another doublet ($g = 2.0038$, 25 G hyperfine separation) overlapped to the other spectral lines in the X-band spectrum of Figure 24, also belonging to the OH radical. The OH radical spectrum has a third broad and uncoupled feature at $g = 2.050$.

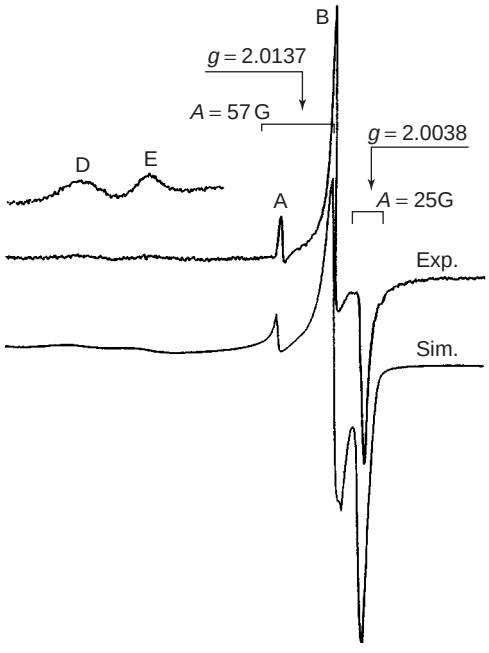

Figure 25 Experimental and simulated Q-band ($\nu = 35$ GHz) of the radical species trapped in the solid MgO_2 matrix corresponding to the spectrum in Figure 24 (A). (Reproduced from Giamello et al.[19] with permission of the American Chemical Society.)

4.3.3 Simulation of Experimental Powder Spectra

The set of spin-Hamiltonian parameters obtained from an experimental EPR spectrum (not only for powder spectra but also for single-crystal and solution spectra) may be confirmed, to avoid ambiguity, by computer simulation of the spectrum on the basis of a given set of data, usually including the various tensors included in the spin-Hamiltonian (***g***, ***A***, ***D***), the line shape and the line width. Computer programs based on various approaches (from first-principle calculations to crude approximations) are available which calculate a spectral profile on the basis of input data. The exactness of the fit with the experiment can be checked by visual comparison of the two spectral patterns or via best-fitting computer programs based on standard mathematical approaches such as Monte Carlo or simplex methods. Computer simulation of the experimental spectra reported in this section [oxygen-based radicals in MgO_2[19]] gave rise to the spectra reported in Figures 25 (Q-band spectrum) and 26(a) and (b) (X-band spectra with H and D hyperfine structure, respectively) The three different simulated profiles were obtained using exactly the same ***g*** tensor and using hyperfine values in the correct 6.5 : 1 ratio as would be expected on passing from H to D. The exactness of the fit for the three different experimental spectra definitely confirms the self-consistency of the set of spin-Hamiltonian parameters introduced in the simulation,

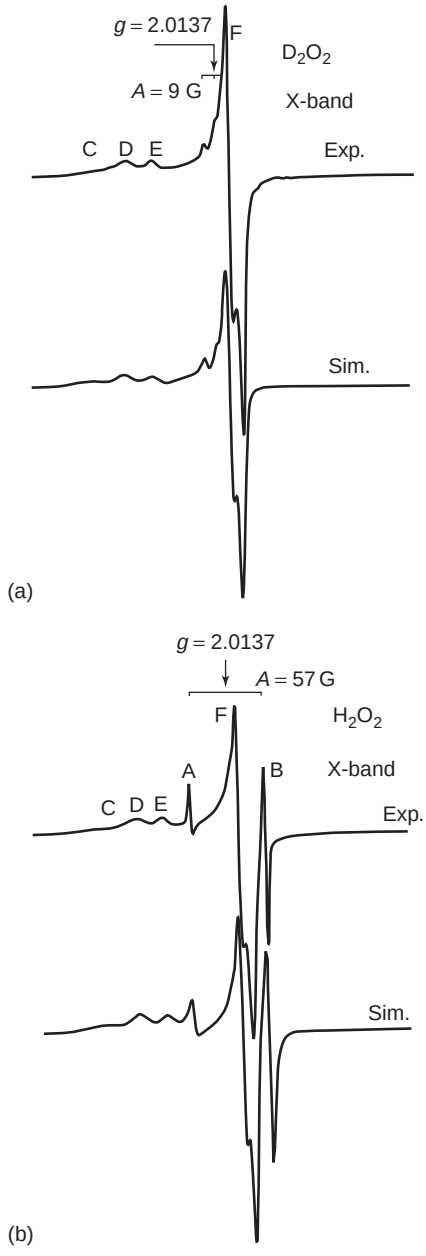

Figure 26 Computer simulation of the experimental spectra reported in Figure 24. The *g* values employed for simulation are the same as in Figure 25.

which are reported in Table 3. The total number of species identified in the MgO_2 matrix is three. Two are O^- and O_2^- radical ions and the third is the OH radical, the observation of which at room temperature is very uncommon owing to its high reactivity. The reported example should have shown that, in certain cases, a single experimental X-band spectrum is not sufficient to determine all the spin-Hamiltonian parameters of a given species, whereas a more complex approach based on complementary techniques allows, in some

Table 3 Spin-Hamiltonian parameters of the oxygen radical species observed in the H_2O_2/MgO system[19]

	Spin-Hamiltonian component			Abundance of species in simulated spectra		
	1	2	3	*a*	*b*	*c*
		Species I				
g	2.0038 ± 0.0003	2.0137 ± 0.0003	2.050 ± 0.005	0.20		0.16
$A_{(H)}$ (G)	25 ± 0.05	57.0 ± 0.05	0 ± 0.05			
		Species I′				
g	2.0038 ± 0.0003	2.0137 ± 0.0003	2.050 ± 0.005		0.13	
$A_{(D)}$ (G)	3.8 ± 0.02	8.8 ± 0.02	0 ± 0.05			
		Species II				
g	2.003 ± 0.001	2.008 ± 0.001	2.057 ± 0.001	0.64	0.62	0.60
		Species III				
g	2.002 ± 0.001	2.043 ± 0.001	2.043 ± 0.001	0.16	0.27	0.24

cases, an unambiguous assignment of the parameters and consequently a full description of the paramagnetic species.

4.4 Solid-state Versus Solution Spectra

The role of anisotropy of the magnetic tensors in causing complexity of the EPR spectra has been emphasized above. It is useful to summarize here the main practical consequences of anisotropy on the spectra in three limiting cases treated in this work, i.e. a solid single crystal, a polycrystalline solid and a low-viscosity solution. The complex cases of high-viscosity solutions or of liquid crystals are not considered here.

Single crystals: the anisotropy of $\boldsymbol{g}$ shows up in terms of different resonant fields for different orientation of the crystal in the external magnetic field. (Figure 12a–d). The anisotropy of $\boldsymbol{A}$ results in different hyperfine line separation, again according to the crystal orientation.

Polycrystalline solids. The anisotropy of $\boldsymbol{g}$ and $\boldsymbol{A}$ is observed in these spectra, which are (section 4.1) an envelope of all possible single-crystal spectra. The determination of all the principal values of the tensors is not straightforward and various methodological approaches (section 4.3) are adopted.

Solutions: molecular tumbling at a high rate has the effect of averaging the principal values of the tensors. In the case of $\boldsymbol{g}$ a g_{av} value is observed (Equation 38) whereas, in the case of the hyperfine tensor, the traceless anisotropic part is averaged to zero and only the Fermi contact term (a_{iso}) is detected. Figure 27(a–d) illustrate the above discussion, reporting the simulated spectra of a Cu(II) compound in terms of single-crystal spectrum (two orientations), powder spectrum and solution spectrum. Cu^{2+} is a $3d^9$ ($S = \frac{1}{2}$, $I = \frac{3}{2}$) ion and, in the cases of the very common elongated octahedral or square-pyramidal structures, an axial $\boldsymbol{g}$ tensor is expected, with (Equations 47 and 48)

$$g_{zz} = g_{\parallel} = g_e + \frac{8\lambda}{\Delta_1} \tag{47}$$

$$g_{xx} = g_{yy} = g_{\perp} = g_e + \frac{2\lambda}{\Delta_2} \tag{48}$$

where g_e and λ have the usual meanings and Δ_1 and Δ_2 are ligand field energy splittings between metal d orbitals. In such compounds $g_{\parallel} > g_{\perp}$ and, for the four-lines hyperfine structure $A_{\parallel} > A_{\perp}$.

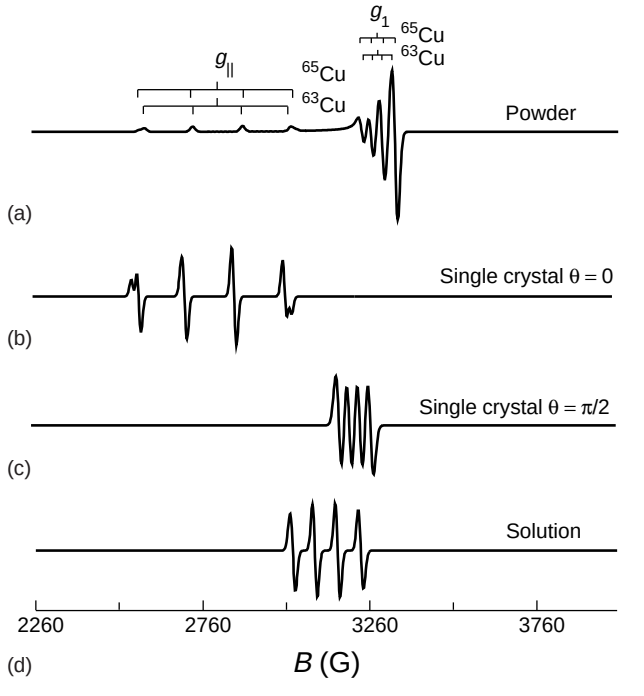

Figure 27 Simulated profiles of the EPR spectra of a Cu^{2+} compound: (a) powder spectrum; (b) and (c) single-crystal spectra (two limiting directions); (d) solution spectrum.

5 ADVANCED ELECTRON MAGNETIC RESONANCE TECHNIQUES

5.1 Electron Nuclear Double Resonance

ENDOR spectroscopy was applied for the first time in 1956 by Feher in solids. The ENDOR experiment is based on the partial saturation of one of the EPR hyperfine lines with an intense microwave field. A second irradiating field in the radiofrequency region induces nuclear spin transitions, thereby altering the populations of the energy levels, and this change is detected by observing changes in the polarization of the saturated EPR transition. Thus, in the ENDOR technique the NMR spectrum of a paramagnetic compound is measured indirectly via the electron spin. One of the principal advantages of the ENDOR technique over EPR is the higher resolution of the spectrum. The basic scheme of the ENDOR experiment is reported in Figure 28, where the EPR ($|\Delta M_S| = \pm 1$) and ENDOR ($|\Delta M_I| = \pm 1$) allowed transitions are shown. As a consequence of the different selection rules, the ENDOR spectrum shows a reduced number of lines and even in complex spin systems with many nuclei is usually still fairly simple.

The ENDOR spectrum of a paramagnetic compound in an external magnetic field can be described by means of the Hamiltonian in Equation (49):

$$H = \mu_B \boldsymbol{BgS} - g_n \beta_n \boldsymbol{BI} + \boldsymbol{IAS} + \boldsymbol{IQI} \qquad (49)$$

which accounts for the electron and nuclear Zeeman interaction, the hyperfine interaction and the nuclear quadrupolar interaction. The last term causes only small shifts in the resonance frequencies. However, in the ENDOR spectrum, the quadrupolar coupling for nuclei with $I \geq 1$ is directly observed as splittings of the lines, giving useful information on the nuclear charge distribution. Furthermore, in the ENDOR experiment, equivalent nuclei, whatever the nucleus is, contribute to the spectrum with a pair of lines symmetrically disposed around ν_n ($\nu_n = g_n \beta_n \boldsymbol{B}/h$), e.g. the Larmor frequency of the nucleus, and separated by the hyperfine splitting constant A if $\nu_n > A/2$. When $\nu_n < A/2$ the resonance lines will be symmetrically disposed around $A/2$ and separated by $2\nu_n$. In general, the ENDOR frequencies are given by Equation (50):

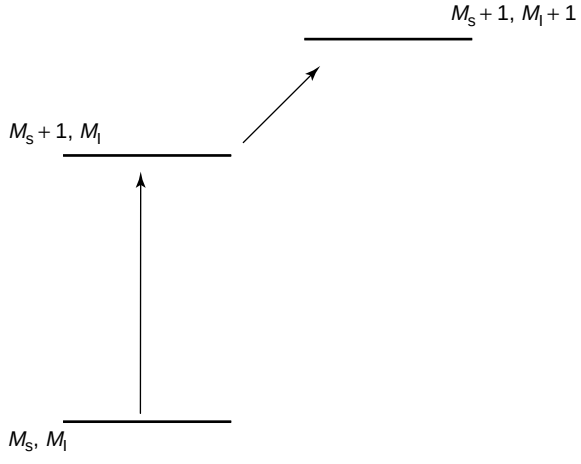

Figure 28 Basic scheme of the ENDOR experiment.

$$\nu_{ENDOR} = \nu_n \pm \frac{A}{2} \qquad (50)$$

It is noteworthy that nuclei with different magnetic moments, e.g. different Larmor frequencies, will appear in different regions of the spectrum. Since the nuclear g factors of most of the nuclei are known, the ENDOR spectrum allows the unambiguous identification of the nucleus under examination.

ENDOR spectroscopy thus offers a way around many of the common EPR problems such as poor resolution and inhomogeneously broadened spectra, the most relevant drawback being its lower sensitivity. Despite this disadvantage, ENDOR spectroscopy remains a powerful method for the analysis of unresolved EPR spectral lines, accurate measurement of hyperfine coupling constants, overlapping EPR signals due to the mixture of radicals or nonequivalent magnetic sites, measurement of the quadrupolar tensor.

5.2 Pulsed Techniques

So far our discussion has been restricted to CWEPR methods. We turn now to consider the field of time-domain EPR.[13–15]

In contrast to NMR spectroscopy, in which time-domain methodology has nowadays almost entirely superseded the conventional continuous wave (CW) methods, time-domain EPR has had a much shorter history, basically for technical reasons. The physics of the spins in both cases is essentially the same but the timescale characterizing the behavior of electron spins is much shorter than that for nuclear spins and this makes the technology needed for pulsed EPR experiments much more demanding. In recent years, however, pulsed EPR spectroscopy has undergone an extraordinary development and commercial instruments are now available.

In this section we shall limit ourselves to some introductory ideas about time-resolved EPR, introducing the basic features of pulsed EPR spectroscopy.

Let us consider a system made up by a number of electron spins $S = \frac{1}{2}$ in an external magnetic field B_0. As stated in section 2.1, electron spins will be distributed at thermal equilibrium between the two Zeeman levels

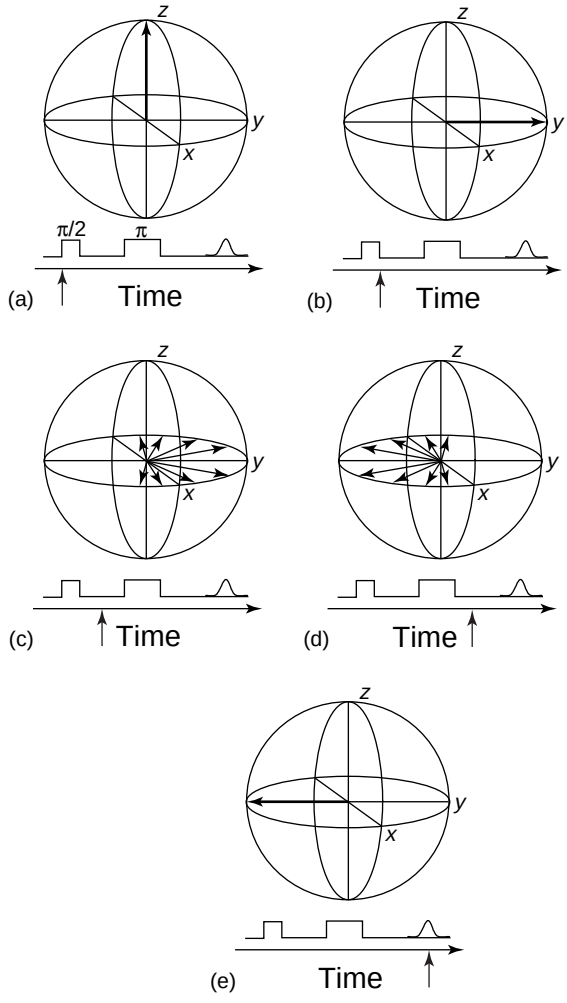

Figure 29 Evolution of the magnetization vector in a two-pulse electron spin-echo experiment. See text for explanation.

according to the Maxwell–Boltzmann law. This means that an excess of magnetic moments parallel to the applied field (β state, lower energy) will be present and a macroscopic magnetization M_z in the direction of the applied static field will be established (Figure 29a). In order to understand the time-dependent EPR experiment we start by describing the behavior of the spin magnetization vector $\boldsymbol{M}_z$ precessing about the static magnetic field direction (Figure 29b). Upon the application of a short intense microwave pulse, the magnetization vector is rotated through 90° in the rotating frame (i.e. a coordinate system rotating with the precessing electron spin magnetization vector about the static magnetic field direction) into a plane perpendicular to the field direction. $\boldsymbol{M}_z$ is made up of the contributions of all the spins present in the system; since these spins are subjected to different ''local fields'' once the pulse is turned off, the magnetization vectors of the individual spins lose their coherence and precess apart (Figure 29c). The oscillating signal, which decays owing to transverse relaxation or inhomogeneous line broadening, is called free induction decay (FID). The Fourier transform (FT) of the oscillating signal in the frequency domain gives a spectrum consisting of peaks at the magnetization precessing frequency, which is identical with the CWEPR spectrum. Hence, in principle, the Fourier transform electron paramagnetic resonance (FTEPR) method allows the EPR spectrum to be recorded within a few microseconds. However, experimental limitations due to the short electron spin relaxation times make this technique, in practice, unsuitable for the majority of the spin systems. Nevertheless, in some restricted areas such as free radicals chemistry, FTEPR spectroscopy has reached a stage of development sufficient to make it an important tool especially for the investigation of short-lived paramagnetic species.

If a second pulse of twice the duration of the first is then applied to the system after a time interval τ, the spin vectors are refocused back into phase at a time interval τ after the second pulse, producing the so-called Echo signal (Figure 29d). The evolution of the magnetization vector in the case of a two-pulse sequence $90°-\tau-180°-\tau$, which is the simplest one that can be used to generate a spin echo signal, can be followed in Figure 29(a–e). At time $t = 0$ the magnetization vector lies along the z-axis defined by the static magnetic field. The first $\pi/2$ pulse rotates the magnetization in the xy-plane after which the magnetization starts to decay for a time τ as the spin vectors spread out by virtue of their different precession frequencies. When a second pulse, π, is applied at a time τ after the first, all the spin vectors rotate at 180° in the xy-plane. Then they continue to move in the same direction, at the same rate as before, until at a time π after the second pulse they are all refocused and the magnetization vector is built up again, giving rise to the echo signal. Since the loss of coherence of the single spin vectors which constitute the magnetization vector $\boldsymbol{M}_z$ is related to the spin–spin and spin–lattice relaxations, EPR spectroscopy provides a direct way of measuring relaxation times which can only be indirectly estimated with the usual CWEPR methods. Pulsed techniques also allow data on the kinetics of short-lived radicals to be obtained and in some cases the analysis of the spin echo spectrum also provides useful information about hyperfine coupling.

The way in which echoes reflect hyperfine interaction will now be briefly described. Until now we have limited our attention to the case of electron spin only. Under certain circumstances, however, a lot of information about hyperfine coupling can be achieved via the echo experiment. A nuclear spin in the vicinity of the unpaired electron will experience a magnetic field that results from the external field and the dipolar field from the unpaired electron. The nuclear spin will precess about this

field, producing a modulation of the dipolar field at the electron due to the nuclear spin. The main consequence of this phenomenon, known as electron spin echo envelope modulation (ESEEM), is that a plot of echo amplitude versus delay time τ often displays periodic modulation due to hyperfine interactions of the unpaired electron with nearby nuclear spins.

6 ANALYTICAL APPLICATIONS OF ELECTRON PARAMAGNETIC RESONANCE

It should be clear from the previous sections that EPR is basically an experimental research technique, extremely useful for structural investigations of various systems, rather than being a routine analytical technique. Nevertheless some applications of the technique having an analytical character have been proposed in the recent past and are briefly described in this section.

6.1 Spin Trapping

Spin trapping is an indirect technique for detection, identification and quantitative determination of free radicals in solution.[20] The direct detection of highly reactive short-lived free radicals in fact is possible only for high concentrations of in situ-produced radicals generally obtained by irradiation or via a rapid-mixing flow system. The spin-trapping technique is based on the reaction of a free radical with the double bond of a diamagnetic compound to produce a less reactive radical. The diamagnetic compound is called the spin trap. The spin-trapping technique, very useful in general for studying short-lived reactive radicals, is essential for detecting radicals at very low concentration or with extremely short relaxation times (broad lines). Typical spin traps are nitroso compounds and nitrones. The well-known and widely employed 5,5-dimethylpyrroline *N*-oxide (DMPO) belongs to this latter class and reacts according to Scheme 1, which leads to the formation of a

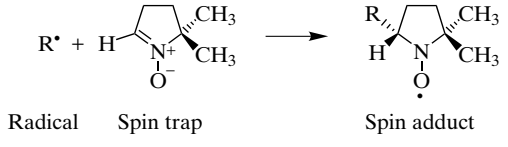

Scheme 1

cyclic nitroxide. The nitroxide spin adducts derived from nitroxones usually exhibit EPR spectra characterized by hyperfine interaction with nitrogen and with the β-hydrogen. The magnitude of the hyperfine coupling is dependent on the nature of the radical species. Therefore, the identification of the trapped radical is not straightforward and is based on the careful measure of the hyperfine constants and on the comparison of their values with those of reference compounds. In Figure 30(a) and (b) the spectra of the adducts of hydroxyl radicals ($OH^{\bullet}$) and CO_2^- radicals with DMPO are shown with the relative stick diagrams illustrating the origin of the two different sets of lines. Double integration of the first-derivative signals and comparison with suitable references allows quantitative estimation of the amount of radicals present in a given system (spin counting).

6.2 Radiation Dosimetry

The accurate measurement of absorbed radiation dose is critical to the proper and effective use of ionizing radiation.[17] A precise determination of absorbed dose is crucial in order to give accurate treatments in nuclear medicine, clinical radiology, food irradiation, the sterilization of medical devices, waste management and a wide variety of agricultural and industrial applications.

Ionization processes, along with excitations in atoms and molecules arising from the interaction between matter and ionizing radiation, are at the basis of the effects induced by radiation in biological tissues. A major task is to correlate the observed effects with the physical characteristics of the radiation field. This can be achieved by introducing a quantity known as absorbed dose (D) which represents the absorbed energy per unit of mass ($D = \mathrm{d}E/\mathrm{d}m$) and is measured in

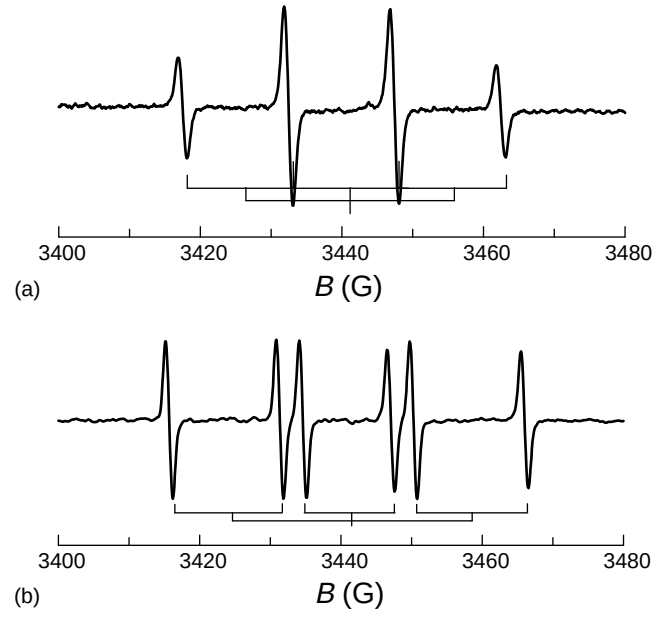

Figure 30 EPR solution spectra of DMPO spin adducts: (a) $R^{\bullet} = CO_2^-$; (b) $R^{\bullet} = OH^{\bullet}$.

gray, ($Gy = J\,kg^{-1}$). The absorbed dose can be directly evaluated by measuring the energy transferred to the matter (calorimetry) or indirectly, by measuring radiation effects such as fission track, thermoluminescence (TL), phosphorescence and optical absorption. Among the possible products originated by ionizing radiation are free radicals whose concentration can be determined by means of EPR spectroscopy. If these radicals are sufficiently stable and a linear correlation between the EPR signal intensity and the absorbed dose exists, then the absorbed dose can be measured. The intensity of an EPR signal originated by radiation is proportional to the radiation dose rate, D, and the irradiation time, t. The dose–response curve is simply described by Equation (51):

$$I(t) = kDt \tag{51}$$

where k is a constant. The correlation between EPR signal and absorbed dose is obtained by means of a calibration curve. EPR dosimetry can be applied to a wide variety of materials and a few examples will now be given.

Alanine has been studied as a dosimetric system since 1962.[21] Owing to the high thermal stability of the radicals generated by ionizing radiation, EPR/alanine dosimetry has been recognized as one of the most appropriate systems for monitoring the entire dose range from 10 to 10^4 Gy. The interaction between ionizing radiation and solid-state alanine gives rise to radical species, among which the stable radical $CH_3-\dot{C}H-COOH$, predominant at room temperature, exhibits a pronounced EPR spectral pattern. The signal intensity, obtained by double integration of the spectrum, is proportional to the concentration of unpaired spins (e.g. $CH_3-\dot{C}H-COOH$ radicals) and allows accurate and reliable absorbed dose measurement. EPR/alanine dosimetry has been studied in radiation processing plants, radiotherapy and an accelerator radiation environment. Another advantage of alanine is its chemical composition, which is close to that of tissue and makes this amino acid very suitable for neutron dosimetry.

Organic materials other then alanine can also be used as a tissue-equivalent dosimetric material. One example is represented by sugar. Sugar has been found to be sensitive to radiation down to 1 mGy. Sugar samples collected from houses have been used as an accident EPR dosimeter in the case of the Chernobyl reactor accident.[22]

One of the best biological dosimeters, however, is tooth enamel. Teeth consist of three main tissues: dentine, enamel and cement. Enamel contains 95–98% hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$] and a few percent of organic materials. Owing to their stability, radiation-induced CO_2^- radicals in hydroxyapatite have been extensively used for EPR dosimetry. EPR of tooth enamel has been applied to dosimetry of cancer patients under radiotherapy, to the cumulative X-ray dose used by dentists and to human dosimetry of radiation accidents. Other biological materials suitable for EPR dosimetry are shells, e.g. marine carbonates sensitive to radiation. Shell buttons have been used in atom-bomb radiation dosimetry.

Another field where EPR dosimetry is gaining more and more importance is the detection of irradiated foodstuffs. Irradiation is becoming an important method of food preservation and the first alternative process to the use of heat to kill microorganisms in food. Food irradiation can be used to reduce or eliminate food spoilage microorganisms, so extending the shelf life of food and improving its safety. Nevertheless, a precise determination of the absorbed dose is crucial in order to

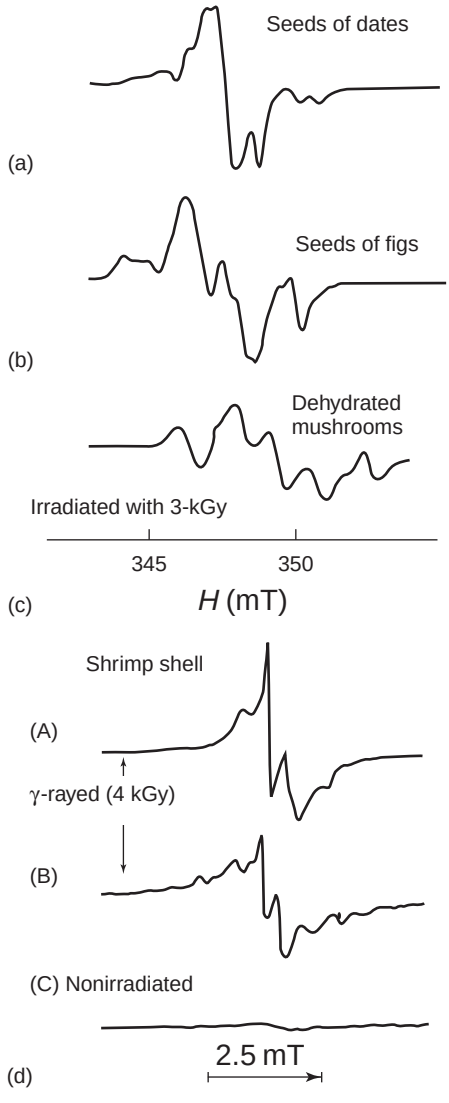

Figure 31 EPR spectra of some irradiated foodstuffs. (Figure reproduced from Ikeya[17] with permission of World Scientific.)

know whether this is within the safe upper dose endorsed by the World Health Organization (WHO) which has been set at 10 kGy. EPR dosimetry is a simple (no sample pretreatment is required) and nondestructive method for the detection of radiation-induced unpaired electrons in irradiated food. EPR spectra of some irradiated foods are shown in Figure 31(a–d).[17] CO_2^- radicals in bones and shells have been used for the dosimetry of irradiated poultry, fish and frog bones. Irradiated meat, vegetables and spices have also been investigated by means of EPR dosimetry.

6.3 Electron Paramagnetic Resonance Dating

Another intriguing application of EPR dosimetry is in the field of archaeological and geological dating.[17,23] The use of EPR as a dating technique is relatively recent. In recent years EPR dating has been applied to a number of different materials in various fields of earth science. Dating of carbonates, bones and tooth enamel in anthropological studies, and dating of burnt flints in archaeological applications have been carried out. EPR dating of geological fault movement, sun-bleached sediment and shock metamorphism has also been reported.

In EPR dating, the sample under investigation acts as a dosimeter, absorbing the doses from all radioactive sources in its immediate environment. Sources of radiation are cosmic rays and natural radioactive emitters such as the isotopes of the U- and Th-decay chains and ^{40}K. The interaction between ionizing radiation and matter gives rise to a number of paramagnetic species which show characteristic EPR spectra. The intensity of the EPR signal is proportional to the concentration of the paramagnetic species and hence to the strength of the dose rate and the time of irradiation (age). An EPR age is thus derived from Equation (52):

$$\text{age} = \frac{TD}{D} \tag{52}$$

where TD is the total dose of natural radiation received by the sample since its formation or its last zeroing, and D is the dose rate. It is worth pointing out that an EPR signal intensity does not indicate the radiation dose or the age unless the intensity is calibrated to the dose. The calibration can be set by the additive dose method, which consists in the exposure of the sample to increasing γ- or β-doses and in the extrapolation of the EPR signal intensity to zero (Figure 32). The dose rate (D) is derived from the chemical analysis of the radioactive elements of the sample and its immediate environment. In some cases the calibration of the EPR signal intensity can be obtained without radiation assessment. Several types of sample have annual growth lines, such as tree rings, which can be exploited to correlate the EPR signal intensities with the elapsed time in order to estimate the TD directly. This method is known as EPR dendrochronology.

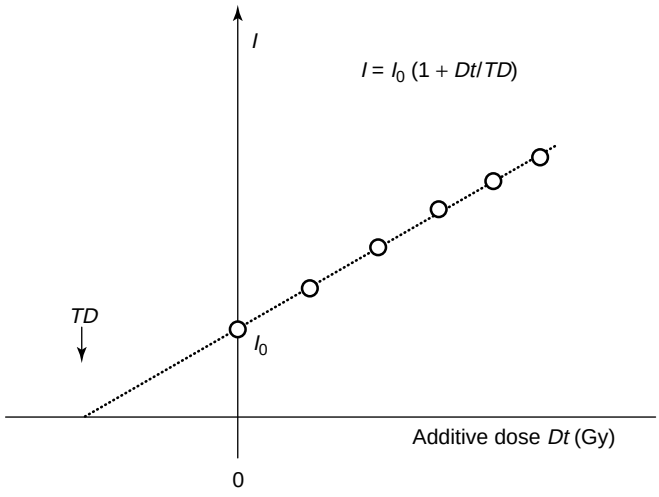

Figure 32 Example of additive dose method calibration curve. ED is determined by extrapolation of the curve.

The signal intensity of radiation-induced defects is related to the EPR age and is considered to be zero for a sample of zero age. The following zeroing processes can be utilized for dating:

- Crystallization: this allows the dating of carbonate such as speleothems, mollusc shells, corals, tooth enamel, gypsum and halite.
- Action of heat: defects produced by natural radiation are annealed by heating. This zeroing process allows the dating of volcanic minerals or materials fired by ancient man.
- Sunlight: shining light on materials leads to optical bleaching of the defects. This mechanism allows, in principle, the dating of sediments.
- Pressure: plastic deformation resets the EPR signal in quartz and ionic oxides.

The dating range of a specific EPR signal is mainly dependent on its radiation sensitivity and its thermal stability. An upper limit of billions of years has been proposed, based on the correlation between geological age and the E' center in quartz grains. The lower limit in EPR dating depends on the signal-to-noise ratio, the detection limit of the spectrometer and the presence of overlapping impurity signals. In certain cases the lower limit can be reduced to about 10 years ($TD = 10$ mGy).

A wide variety of materials and geological and archaeological events can, in principle, be dated by

EPR spectroscopy. The basic requirement is that the defects are stable for a sufficient length of time or a zeroing process, according to the events previously described, has occurred. Samples whose *TD*s have so far been determined by EPR spectroscopy include carbonates (stalacites, travertine, coral, shell, etc.), sulfates (anhydrite and gypsum), phosphates (hydroxyapatite, tooth, bone, etc.), silica and silicates (geological faults, volcanic rock, clay minerals) and organics (food, crops, leather, paper, alanine, mummies, blood).

EPR dating, owing to the large variation of materials and geological and archeological events to which this technique can be applied, has, in principle, one of the greatest potentials among dating methods. None the less, some problems have still not been conclusively solved. It is worth pointing out that an EPR result depends upon up to 15 independent factors and each can be influenced by various processes. EPR dating also lacks routine sample-preparation techniques, measurement conditions, signal definition, *TD* determination techniques and age-calculation programs.

ACKNOWLEDGMENTS

The authors thank Dr Eilsh O'Donoghue (Dublin Institute of Technology) for carefully revising the manuscript, and Dr Lorenzo Dall'Acqua and Dr Giors Gasco for technical assistance with graphical editing.

ABBREVIATIONS AND ACRONYMS

CW	Continuous Wave
CWEPR	Continuous Wave Electron Paramagnetic Resonance
DMPO	5,5-Dimethylpyrroline *N*-oxide
DPPH	Diphenylpicrylhydrazyl
ENDOR	Electron–Nuclear Double Resonance
EPR	Electron Paramagnetic Resonance
ESEEM	Electron Spin Echo Envelope Modulation
ESR	Electron Spin Resonance
FID	Free Induction Decay
FT	Fourier Transform
FTEPR	Fourier Transform Electron Paramagnetic Resonance
NMR	Nuclear Magnetic Resonance
radar	Radio Detection and Ranging
SOMO	Semi-occupied Molecular Orbital
TL	Thermoluminescence
WHO	World Health Organization
ZFS	Zero Field Splitting

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Multinuclear Magnetic Resonance Spectroscopic Imaging

Food **(Volume 5)**
Near-infrared Spectroscopy in Food Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Nuclear Magnetic Resonance of Geological Materials and Glasses • Solid-state Nuclear Magnetic Resonance

REFERENCES

1. J.A. Weil, J.R. Bolton, J.E. Wertz, *Electron Paramagnetic Resonance, Elementary Theory and Practical Applications*, Wiley-Interscience, New York, 1994.
2. N.M. Atherton, *Principles of Electron Paramagnetic Resonance*, Ellis Horwood, Chichester, 1993.
3. P.B. Ayscough, *Electron Spin Resonance in Chemistry*, Methuen, London, 1967.
4. W. Gordy, *Theory and Applications of Electron Spin Resonance*, Wiley, New York, 1980.
5. C.P. Poole, Jr, H.A. Farach, *The Theory of Magnetic Resonance*, Wiley-Interscience, New York, 1987.
6. A. Abragam, B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions*, Clarendon Press, Oxford, 1970.
7. J.R. Pilbrow, *Transition Ion Electron Paramagnetic Resonance*, Clarendon Press, Oxford, 1990.
8. F.E. Mabbs, D. Collison, *Electron Paramagnetic Resonance of d-Transition Metal Compounds, Studies in Inorganic Chemistry Series*, Vol. 6, Elsevier, Amsterdam, 1992.
9. A. Bencini, D. Gatteschi, *Electron Paramagnetic Resonance of Exchange-coupled Systems*, Springer, Berlin, 1990.
10. L.J. Berliner (ed.), *Spin Labelling – Theory and Applications*, Academic Press, New York, 1976.
11. L.J. Berliner (ed.), *Spin Labelling II – Theory and Applications*, Academic Press, New York, 1979.
12. L.J. Berliner, J. Reuben (eds.), *Biological Magnetic Resonance*, Plenum Press, New York, Vol. 1, 1978.
13. A. Schweiger, 'Pulsed Electron Spin Resonance Spectroscopy: Basic Principles, Techniques, and Examples of Applications', *Angew. Chem., Int. Ed. Engl.*, **30**, 265–292 (1991).

14. L. Kevan, R.N. Schwartz (eds.), *Time-domain Electron Spin Resonance*, Wiley, New York, 1979.
15. C.P. Keijzers, E.J. Reijerse, J. Schmidt (eds.), *Pulsed EPR – A New Field of Applications*, North-Holland, Amsterdam, 1989.
16. H. Kurreck, B. Kirste, W. Lubitz, *Electron Nuclear Double Resonance Spectroscopy of Radicals in Solution*, VCH, Weinheim, 1988.
17. M. Ikeya, *New Applications of Electron Spin Resonance, Dating, Dosimetry and Microscopy*, World Scientific, Singapore, 1993.
18. H. Prinzbach, B.A.R.C. Murty, W. Fessner, J. Mortensen, J. Heinze, G. Gescheidt, F. Gerson, 'One- and Two-electron Oxidations of Pagodanes and Bissecododecahedradienes: Unusually Persistent Radical Cations', *Angew. Chem., Int. Ed. Engl.*, **26**, 457–458 (1987).
19. E. Giamello, L. Calosso, B. Fubini, F. Geobaldo, 'Evidence of Stable Radicals and Other Oxygen Radical Species Generated by Interaction of Hydrogen Peroxide with Magnesium Oxide', *J. Phys. Chem.*, **97**, 5735–5740 (1993).
20. E.G. Janzen, D.L. Haire, 'Two Decades of Spin Trapping', in *Advances in Free Radical Chemistry*, ed. D.D. Tarmer, JAI Press, Greenwich, CT, Vol. 1, 253–295, 1990.
21. W.W. Bradshaw, D.G. Crawford, H.W.W. Spetzler, 'The Use of Alanine as a Solid Dosimeter', *Radiat. Res.*, **171**, 11–21 (1962).
22. T. Nakajima, T. Ohtsuki, I. Likhtariov, 'Tentative Dose Estimation in House at Pripyat-city on Chernobyl Accident with Sugar', *J. Nucl. Sci. Technol.*, **28**, 71–73 (1991).
23. R. Grün, 'Electron Spin Resonance (ESR) Dating', *Quat. Int.*, **1**, 65–109 (1989).

High-performance Liquid Chromatography Nuclear Magnetic Resonance

Harry C. Dorn
Virginia Polytechnic Institute and State University, Blacksburg, USA

1 Introduction 1
2 Theory and Operating Principles 2
3 Applications of High-performance Liquid Chromatography Nuclear Magnetic Resonance 4
3.1 Fuels 4
3.2 Polymers 4
3.3 Biological and Pharmaceuticals 7
4 Miscellaneous Applications 8
5 Perspective and Future Developments 10
Abbreviations and Acronyms 12
Related Articles 12
References 13

The direct coupling of high-performance liquid chromatography (HPLC) to nuclear magnetic resonance (NMR) has evolved as a powerful new tool (HPLC/NMR) for the analysis of mixtures. The technique was originally reported in the late 1970s, but was generally limited by NMR detector sensitivity constraints and the availability of routine high magnetic field dispersion instruments (11.8–17.6 T, 500–750 MHz ^{1}H-NMR). During the early 1980s, the HPLC/NMR technique was applied to problems involving the analysis of mixtures that were typically not sample limited, such as petroleum fuels. In the early 1990s, the advent of commercial HPLC/NMR equipment provided the impetus for a worldwide growth spurt of the HPLC/NMR technique. As outlined in this review, the HPLC/NMR technique has evolved as a powerful tool for applications involving polymers, pharmaceuticals, and biological mixtures.

An inherent advantage of HPLC/NMR is the ability to monitor the NMR chemical shift dimension. This dimension is particularly relevant for those applications involving structural isomer identification and has distinct advantages in comparison with the more commonly employed HPLC/MS (mass spectrometry) technique. Although sensitivity limitations of NMR as a detector are still a concern, the development of more complex hyphenated combinations (such as HPLC/NMR/MS) will undoubtedly continue, especially in the areas of pharmaceutical and drug development. Other technical and phenomenological considerations of the HPLC/NMR technique are presented in this review. Finally, a number of related variant techniques including capillary electrophoresis nuclear magnetic resonance (CE/NMR), capillary zone electrophoresis (CZE), capillary electrochromatography (CEC), supercritical fluid chromatography (SFC), and dynamic nuclear polarization (DNP) are also discussed.

1 INTRODUCTION

In 1978, Watanabe and Niki[1] reported the first results for direct coupling of HPLC to ^{1}H-NMR with stopped-flow detection. During the next two years, Bayer et al.,[2] Buddrus and Herzog,[3] and Haw et al.[4] extended the approach to continuous flow ^{1}H-NMR detection (HPLC/^{1}H-NMR). During the early 1980s, the HPLC/NMR technique continued to evolve and was applied to problems involving the analysis of mixtures that were typically not sample limited, such as petroleum fuels (see section 3.1). In the early 1990s, the major growth of HPLC/NMR as a viable analytical tool occurred after the commercial introduction by the NMR manufacturers Bruker and Varian. With the advent of commercial instruments, the number of laboratories worldwide with HPLC/NMR instrumentation increased by at least 10-fold during the 1990s to over 150. During this time, the availability of high-field NMR superconducting magnets (11.8–17.6 T, 500–750 MHz ^{1}H-NMR) has helped alleviate sensitivity constraints and improved spectral dispersion. The HPLC/NMR technique has evolved as a powerful tool for applications in pharmaceutical and related biological industries (see section 3.3). With these developments, it is not difficult to predict that HPLC/NMR will continue to evolve as a powerful analytical tool.[5–12]

One of the major advantages of NMR in comparison with other commonly employed chromatographic detectors is the high information content present in the NMR chemical shift dimension. The inherent advantage of monitoring the local electronic environment at each nonequivalent nuclear site (^{1}H) in a molecule has distinct advantages in comparison with other chromatographic detectors, such as ultraviolet (UV), refractive index (RI), and MS. This is a particularly relevant point for those applications involving isomer identification (especially when coupled with MS, i.e. HPLC/NMR/MS). The noninvasive nature of the NMR detector also has distinct advantages for the characterization of air-sensitive and/or light-sensitive labile compounds.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 THEORY AND OPERATING PRINCIPLES

Although standard high-resolution NMR normally utilizes samples spinning in cylindrical 5 mm tubes, a flow NMR detector represents different phenenomological and technical requirements. For example, the flow cell volume V_c is usually smaller than the 500–700 μL (5 mm tubes) typically employed in high-resolution static NMR studies. At present, commercial HPLC/^{1}H-NMR instruments employ detector volumes (V_c) of the order of 50–150 μL (see Figure 1). In addition, considerable progress has been made (notably by Olson et al.)[13] to extend this approach to nanoliter volumes.

Undoubtedly the major constraint in utilizing HPLC/NMR has been the sensitivity limitations of NMR as a detector. Although this factor has clearly limited development in the past, the advent of routine high-field NMR magnets (600–800 MHz, 14–19 T) has helped suppress this constraint. Furthermore, the recent development of cooled cryoprobes as NMR detectors promises further sensitivity gains by factors of 3–5. Figure 2 illustrates the present approximate detection limits for several detectors commonly employed in LC. It is apparent from Figure 2 that present commercial ^{1}H-NMR detectors have ultimate detection limits comparable with several common LC detectors, but fluorescence and MS still have a 10^2–10^4 sensitivity advantage. To provide a quantitative measure of the detection limits of HPLC/^{1}H-NMR, several factors must be considered. Nevertheless, for molecules with molecular weights of 200–400 Da, continuous flow (1–4 s observation times) HPLC/^{1}H-NMR detection limits of 5–2 μg are commonly achieved for 400–800 MHz ^{1}H-NMR detectors. For the case of stopped-flow ^{1}H-NMR measurements (several hours acquisition times) at these field strengths, a reasonable signal to noise ratio (S/N) can be achieved for samples of only 200–50 ng.

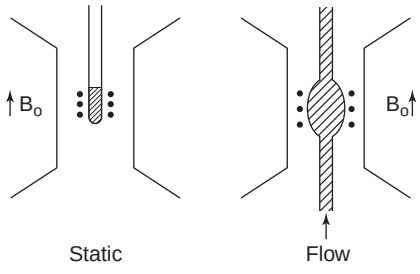

Figure 1 High-field NMR (static, spinning sample) and HPLC/NMR (flowing sample).

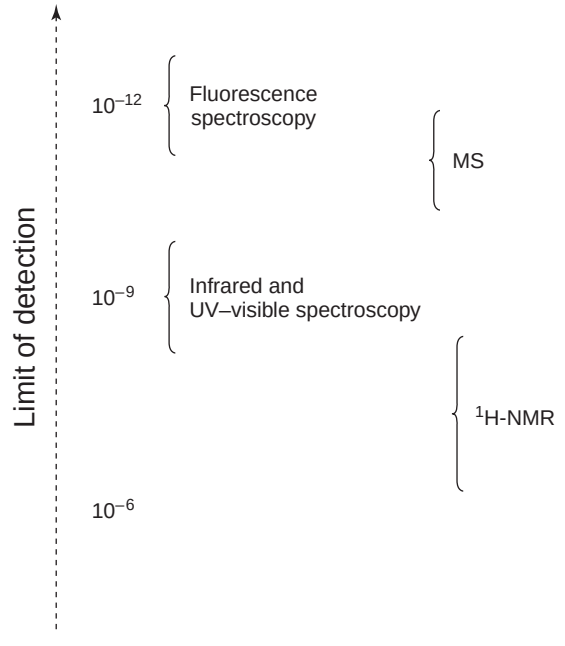

Figure 2 Limits of liquid chromatography (LC) detection (the ranges indicated are approximate).

A second problem originally anticipated for the LC/^{1}H-NMR technique was the limited choice of compatible chromatographic solvents (e.g. without hydrogen). This problem was originally alleviated by the use of extensive deuterated, chlorinated, and/or fluorinated solvents for normal-phase chromatography (e.g. freon 113, D_2O, or $DCCl_3$). In another approach, SFC coupled with ^{1}H-NMR has the advantage that certain common supercritical solvents (e.g. CO_2) do not contain hydrogen.[14–16] However, most state-of-the-art chromatographic separations utilize reversed-phase conditions and correspondingly large residual ^{1}H-NMR background signals are normally present in the reverse-phase deuterated solvents (D_2O, D_3CCN, or CD_3OD) usually employed. In these cases, a number of solvent-suppression sequences have been successfully developed to eliminate one or more remaining spectral solvent lines. Originally, Laude et al. utilized a 1–1 hard pulse sequence for solvent suppression (97 : 3, D_2O–acetonitrile) in a reversed-phase LC/^{1}H-NMR separation of the nucleosides uridine, cytidine, and adenosine.[17] Albert et al. have employed a binomial 1 : 3 : 3 : 1 sequence for solvent suppression of a 50 : 50 water–acetonitrile solvent system in a reversed-phase separation of several aromatic compounds (70 μg each).[18] More recently, very effective solvent suppression of one or more spectral lines has been reported by Smallcombe et al. utilizing a water elimination T_1 (spin–lattice relaxation time) (WET) sequence.[19] The very effective elimination of solvent peaks (as well as ^{13}C satellite lines is illustrated in Figure 3).

One phenomenological consideration of continuous flow HPLC/^{1}H-NMR and flow NMR is the essential requirement that the flowing bolus must reside in the

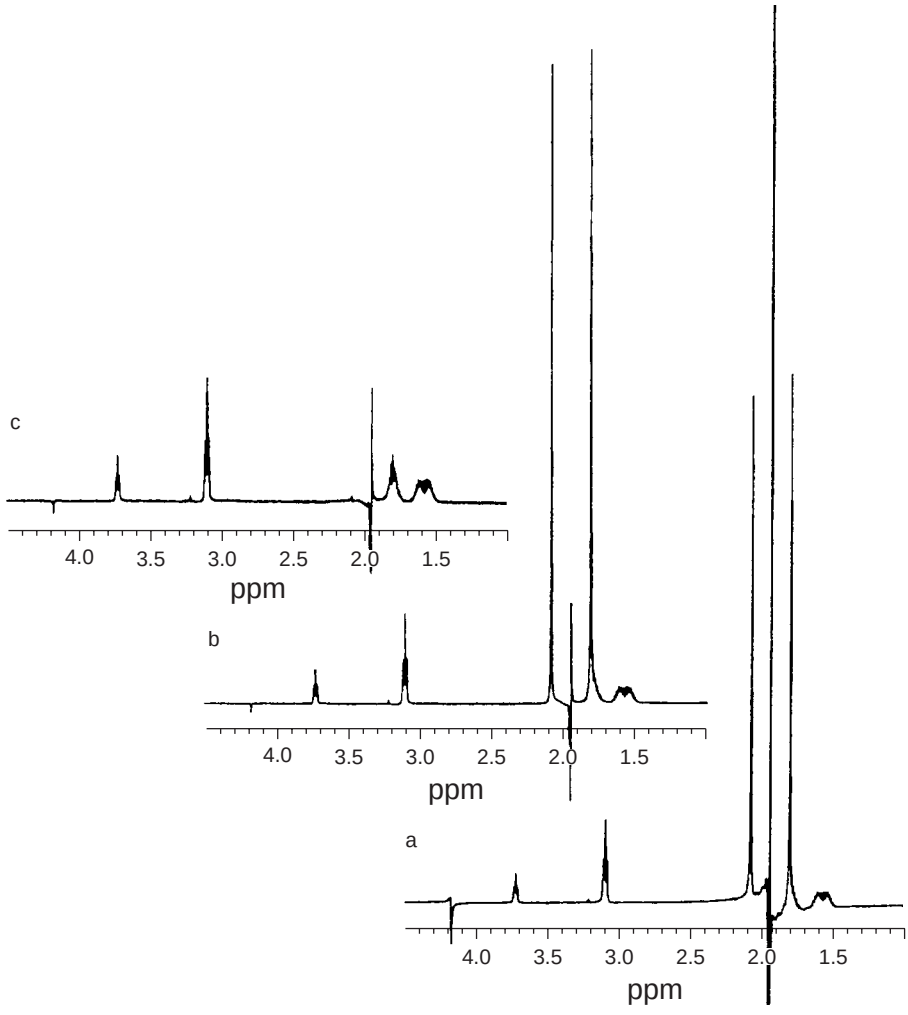

Figure 3 A comparison of WET and conventional presaturation solvent suppression on a sample of 10 mM arginine in 50 : 50 CH_3CN (0.1% trifluoroacetic acid (TFA))–D_2O flowing at 1.0 mL min^{-1}. Curve (a) used both transmitter and decoupler presaturation, each producing 40 Hz field strength 1.1 s square pulses, for the CH_3CN and monodeuterated water (HOD) resonances at 1.95 and 4.4 ppm, respectively. The residual center line of acetonitrile is plotted off scale. Curve (b) used WET without ^{13}C decoupling, whereas curve (c) used WET with ^{13}C decoupling. Curves (b and c) used 22.4 ms SEDUCE band-selective decoupling scheme pulses, the spin–lattice relaxation time (T_1) and B_1-insensitive (where B_1 is radiofrequency field) version of WET, a 10 ms *dz* delay following the final gradient pulse (G4), and a composite read pulse. Each spectrum is the result of four scans, and was processed using 1 Hz line broadening (no solvent-deconvolution software was used). The acquisition times for (b) and (c) were 2.0 s (no zero filling), whereas the presaturation spectrum (a) required a 1.0 s acquisition time and zero filling to allow the repetition rate for all three spectra to be kept constant (at 2.1 s). (Reprinted with permission, *J. Magn. Reson.*[19])

magnetic field (B_o) for a time (τ) that is long relative to the ^{1}H-NMR T_1s ($\geq$3–4 T_1s). This requirement ensures that the Boltzmann magnetization has sufficient time for equilibration in the magnetic field. For small molecules (100–300 Da) in solution the ^{1}H T_1s are usually of the order of several seconds, but are considerably shorter (<1 s) for larger molecules. Thus, this requirement is usually satisfied at low flow rates and/or the use of pre-equilibration volume in the magnetic field. One ideal solution that has not been commonly implemented is the placement of the LC column directly in the magnetic field.[20]

A second consequence of flow is the effective residence time of the bolus in the detector cell, t_c. The average residence time of the bolus in the detector cell, V_c can be expressed (Equation 1) as

$$t_c = \frac{V_c}{\bar{v}} \tag{1}$$

where the mean velocity $\bar{v}$ must be used. The mean velocity is defined because for other flow patterns (e.g. laminar) the velocity is a function of the distance from the detector cell walls. If one treats flow as a relaxation process, the lifetime broadening can be expressed as in

Equations (2) and (3):

$$\left(\frac{1}{T_1}\right)_{\text{flow}} = \left(\frac{1}{T_1}\right)_{\text{static}} + \frac{1}{t_c} \quad (2)$$

$$\left(\frac{1}{T_2}\right)_{\text{flow}} = \left(\frac{1}{T_2}\right)_{\text{static}} + \frac{1}{t_c} \quad (3)$$

It is apparent from Equation (3) that the observed flow NMR line width $\Delta\nu_{1/2} = (\pi T_2)^{-1}$ can exhibit a significant increase at high flow rates because of the limited residence time, t_c. In contrast, static samples only exhibit contributions from spin–spin relaxation processes (T_2) and the magnetic field inhomogeneity. For example, Fyfe et al.[21] reported ^{1}H-NMR line width data for flowing ethylbenzene using a 5 mm outside diameter flow tube. A linear plot (Equation 3) was obtained as a function of flow rate with line widths ranging from 2.2 Hz (static) to 8 Hz (flow, 120 mL min^{-1}). The bolus lifetime problem becomes more acute with the progression to nanoliter-scale volumes.[13] For example, at a flow rate of 10 mL min^{-1} and a cell volume V_c of about 80 μL, an average residence time t_c of about 0.5 s is readily calculated, which corresponds to an observed line width of several hertz. However, if a cell volume of only 800 nL is employed at the same flow rate, this corresponds to an NMR line width of several hundred hertz.

Finally, it is important to emphasize that the major focus of flow HPLC/NMR has been mainly limited to the ^{1}H nuclide. The described sensitivity limitations of the NMR detector have generally led to only HPLC/NMR studies of NMR high-sensitivity nuclides (e.g. ^{1}H and ^{19}F). However, the extension to other less favorable nuclides such as ^{13}C (section 4) is possible in future studies by indirect detection methods (via ^{1}H) and/or stopped-flow approaches. It is also important to note that HPLC/NMR is one application for flow NMR and that several other flow experiments have been reported, such as recycled flow.[12]

3 APPLICATIONS OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY NUCLEAR MAGNETIC RESONANCE

3.1 Fuels

One of the first applications of HPLC/^{1}H-NMR was the analysis of fuel samples from petroleum, shale, and coal samples.[22–26] As these studies were not hampered by the availability of the sample, larger-scale chromatographic columns were commonly employed. These early examples illustrate the power of the HPLC/NMR approach and the critical importance of the chemical shift dimension.

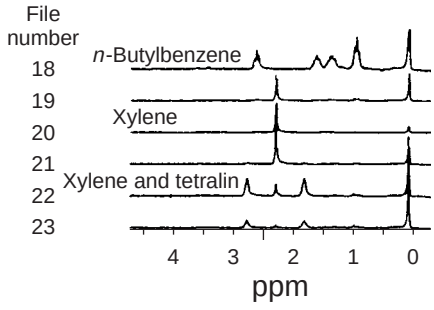

Figure 4 200 MHz LC/^{1}H-NMR profile of the monocyclic region of a model mixture consisting of *n*-butylbenzene, xylene, and tetralin. A Whatman M9 10.50 PAC column was used with the solvent system freon 113–$CDCl_3$ 97.5 : 2.5. Hexamethyldisiloxane (0.05%) was used as the reference. Each file represents 0.5 mL elution volume, 30 s per file. (Reprinted with permission from J.F. Haw, T. Glass, H.C. Dorn, *JEOL News*, **18A**, 6–9 (1982).)

For example, Figure 4 illustrates five progressive snapshots in the LC/^{1}H-NMR profile of the monocyclic aromatic region of a model fuel mixture. The progression from early-eluting compounds (*n*-butylbenzene, spectral file 18) to later-eluting compounds (*m*-xylene and tetralin) demonstrates resolution in the chromatographic sense. In spectral file 22 (Figure 4), two compounds (*m*-xylene and tetralin) that are not fully resolved chromatographically are easily identified along the chemical shift axis.

A practical example of LC/^{1}H-NMR utilizing a 200 MHz ^{1}H-NMR detector is presented in Figure 5(a). A sample consisting of 100 μL of a jet fuel was injected and the data acquisition required 20 min. The HPLC/^{1}H-NMR profile is clearly more informative than LC or static ^{1}H-NMR alone, providing a fingerprint for a fuel. The progression from linear alkanes to branched and cyclic hydrocarbons is readily apparent in files 2 to 7 (Figure 5a) for the aliphatic fraction of this fuel. In a similar fashion the monocyclic aromatic region clearly indicates a progression from short chain (3–4 alkyl carbons) monosubstituted aromatics in files 13 and 14 (Figure 5b) to di-, tri-, and tetra-methyl substituted aromatics in files 18–22 (Figure 5c). Methods for the determination of quantitative data from the HPLC/^{1}H-NMR profile of fuels using an internal standard have been reported.[27] Finally, properties such as density, freezing point, or aromaticity can be predicted from the HPLC/^{1}H-NMR profile.[28]

3.2 Polymers

One of the first series of detailed reports for analyzing soluble polymer mixtures was reported by Hatada and colleagues. During the 1990s, this group has utilized several chromatographic approaches,

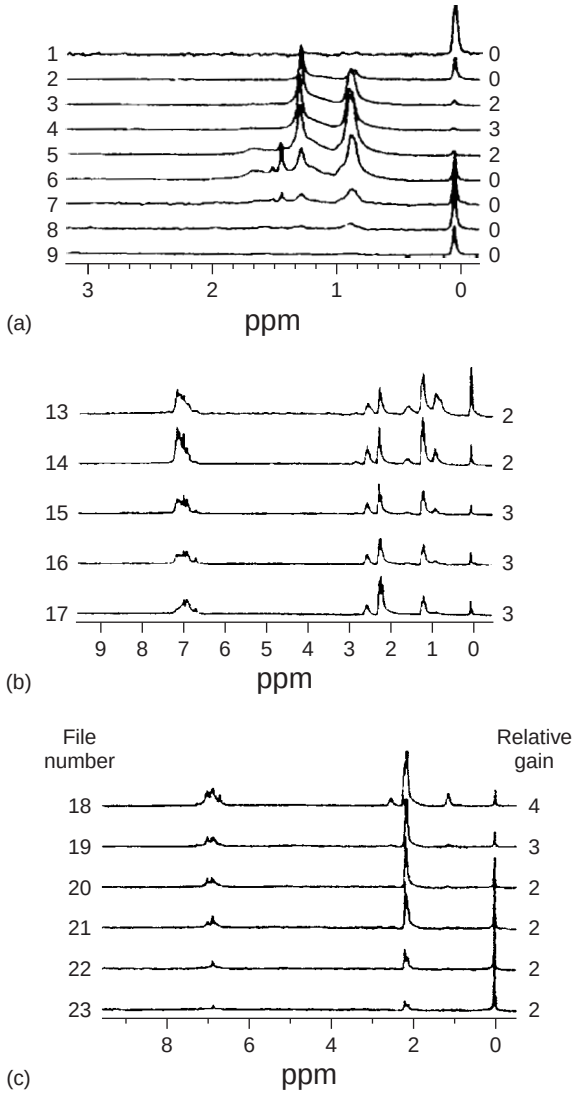

Figure 5 (a) 200 MHz LC/^{1}H-NMR profile of the aliphatic region of jet fuel sample. Each file represents 0.5 mL elution volume, 30 s per file. (b) 200 MHz LC/^{1}H-NMR profile of the monocyclic aromatic region of jet fuel sample. Each file represents 1.0 mL elution volume, 60 s per file. (c) 200 MHz LC/^{1}H-NMR profile of the monocyclic aromatic region of jet fuel sample. Each file represents 1.0 mL elution volume, 60 s per file. (A Whatman M9 10.50 PAC column was used with the solvent system freon 113–$CDCl_3$ 97.5 : 2.5. Hexamethyldisiloxane (0.05%) was used as the reference.) (Reprinted with permission from J.F. Haw, T. Glass, H.C. Dorn, *JEOL News*, **18A**, 6–9 (1982).)

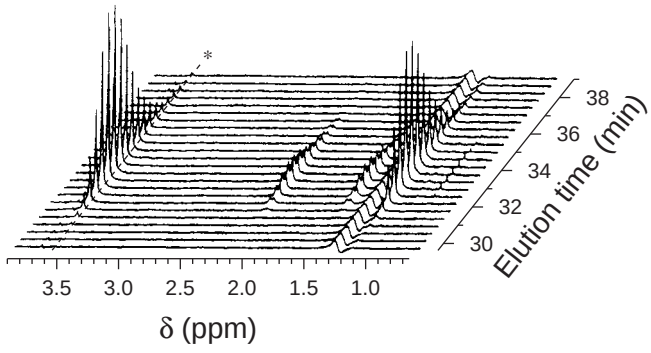

Figure 6 On-line GPC/NMR data of isotactic PMMA (M_n = 12 600). Cross-section at 3.60 ppm (∗), where the methoxy proton resonates, gives the ^{1}H-NMR-detected GPC curve (see Figure 7). (Reprinted with permission, *Polymer Journal*.[32])

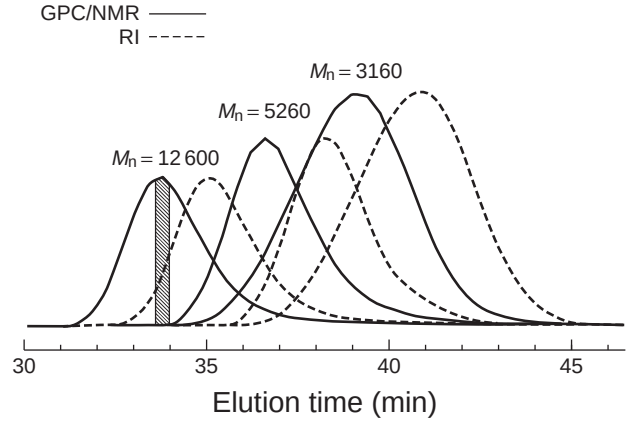

Figure 7 The ^{1}H-NMR-detected GPC curves of the isotactic PMMAs with M_n values of 12 600, 5260, and 3160. These curves were obtained by plotting the intensity of the methoxy proton signal (3.60 ppm) against elution time. The RI-detected GPC curves are also shown. (Reprinted with permission, *Polymer Journal*.[32])

including gel permeation chromatography (GPC) directly coupled to ^{1}H-NMR (GPC/^{1}H-NMR) to analyze several different soluble homopolymers, copolymers, and oligomers.[29–33] In one study, chloral oligomers were separated in both the chromatographic and ^{1}H chemical shift domain into dimer, trimer, tetramer, pentamer, hexamer, and heptamer fractions in the GPC/^{1}H-NMR profile.[29] In a GPC/^{1}H-NMR study of isotactic poly(methylmethylacrylates) (PMMAs) containing a single *t*-butyl end group, the molecular weights were directly determined from integration of the *t*-butyl and methoxy ^{1}H-NMR signals.[32] The on-line GPC/^{1}H-NMR profile for isotactic PMMA is shown in Figure 6, and the corresponding ^{1}H-NMR-detected GPC curves are shown in Figure 7. In this latter figure, the RI curves are similar to the NMR data, but displaced in time because of differences in the void volumes of the two detectors. A linear plot of log ($\overline{M}_n$) versus elution time was obtained from the ^{1}H-NMR data profile.

More recently, Hatada and colleagues[33] have reported a further study of the direct determination of molecular weight distribution of isotactic PMMAs ($\overline{M}_n = 3.27 \times 10^3 - 4.44 \times 10^4$) utilizing site-exclusion chromatography (SEC)/^{1}H-NMR at the higher frequency of 750 MHz. In this study, the relative ^{1}H-NMR intensity of the

end-groups was used as the internal standard (integrated relative to the repeating units) without recourse to the use of an internal standard. Also, the conditions were carefully optimized for quantitative measurements including the flow rate, pre-equilibration in the magnetic field, and other NMR operating parameters. The molecular weight distribution determined in this fashion was found to be accurate based on uniform isotactic PMMAs consisting of exactly 23, 40 and 80 repeating units. The results of this SEC/^{1}H-NMR study were compared with the usual SEC/RI approach. The greater accuracy of the SEC/^{1}H-NMR data that does not require a calibration curve was observed in this study, as were the advantages in S/N available at the higher operating frequency (750 MHz ^{1}H-NMR).

Pasch et al.[35] have utilized HPLC/^{1}H-NMR to analyze the tacticity of oligostyrenes. In this study, four separate isomers **(1–4)** in Scheme 1 of the trimer were

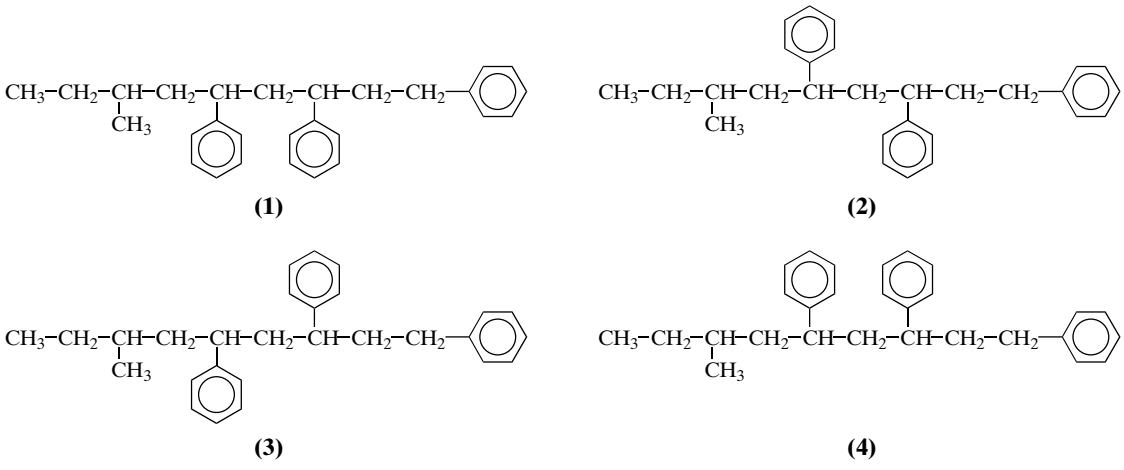

Scheme 1

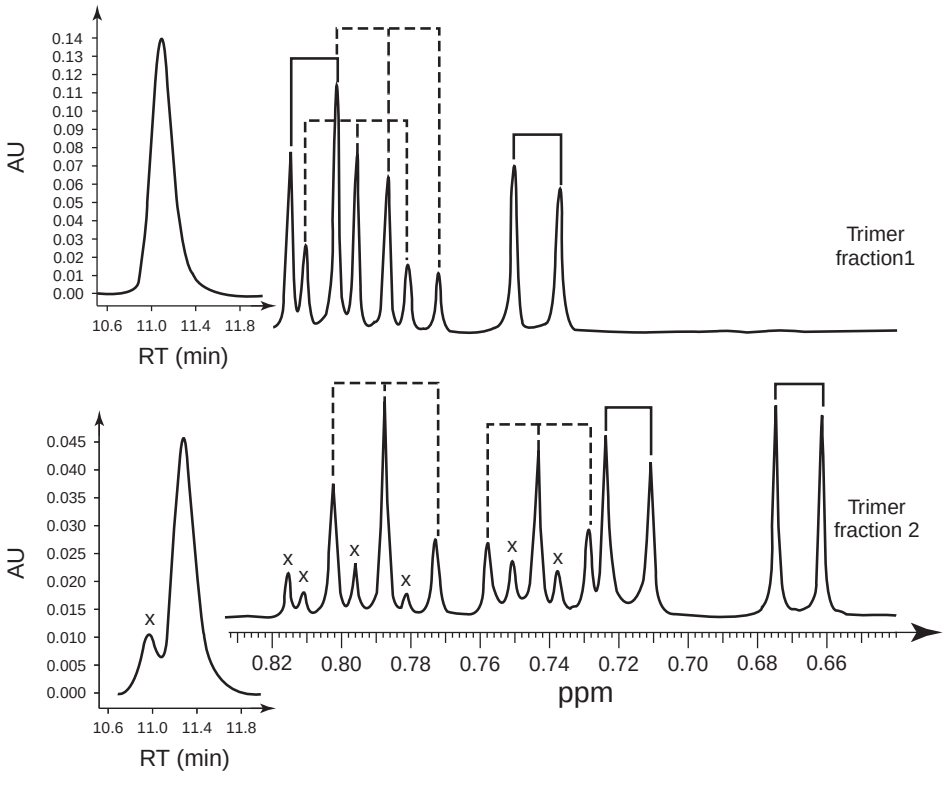

Figure 8 Chromatograms and ^{1}H methyl group regions of fractions 1 and 2 of the chromatographic separation of the trimer. (Reprinted with permission, *Polymer*.[34])

identified in two different chromatographic fractions, as outlined in Figure 8. The trimer fractions 1 and 2 each consist of two isomers with a small amount of fraction 1 eluting into fraction 2. The 500 MHz ^{1}H-NMR spectral methyl region for these fractions clearly identifies eight different methyl groups (four triplets and four doublets). However, one set of triplets overlaps (at about 0.784 ppm) in both fractions. A highlight of this study was the use of pure acetonitrile with no deuterium lock.

3.3 Biological and Pharmaceuticals

With the advent of commercial high magnetic field equipment during the early 1990s by Bruker and Varian, the utility of HPLC/NMR has rapidly evolved for many pharmaceutical applications. For example, a series of papers illuminate the utility of HPLC/^{1}H-NMR for the investigation of numerous urinary and bile metabolites.[35–42]

In one pioneering study, the advantages of monitoring both the ^{1}H and ^{19}F chemical shift dimension of the HPLC/NMR profile for the human urinary metabolites of fluorine containing the drug, flurbiprofen **(5)** was nicely demonstrated (Scheme 2).[40]

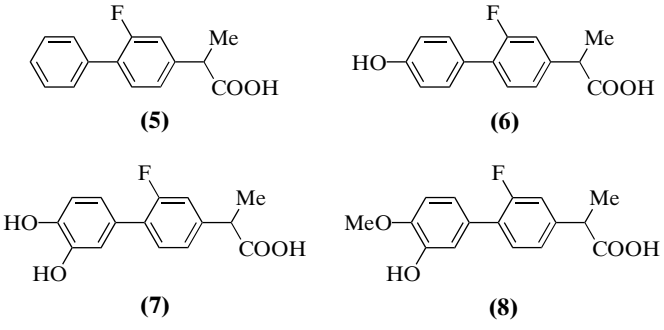

(5) **(6)** **(7)** **(8)**

Scheme 2

The lack of any significant endogenous fluorine-containing compounds yielded a HPLC/^{19}F-NMR profile of only four major fluorine-containing metabolites **(5–8)** in Scheme 2. The four fluorine-containing metabolites were then further characterized using the HPLC/^{1}H-NMR approach.

In another study,[41] phase II metabolites of the nonsteroidal anti-inflammatory drug tolfenamic acid **(9)** were clearly separated in the higher field chemical shift dimension (800 MHz ^{1}H-NMR) profile (Figure 9). This represents one of the first examples of direct identification of intact glucuronic acid conjugates of tolfenamic acid without β-glucuronidase treatment of the biological sample.

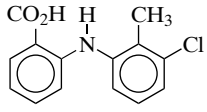

(9)

A 400 MHz and 600 MHz HPLC/^{1}H-NMR study of the isomers of 6,11-dihydro-11-oxodibenz[*b*,*e*]oxepin-2-acetic acid glucuronide (a nonsteroidal anti-inflammatory drug) has also been reported by this group.[42] This study illustrates the power of the ^{1}H-NMR chemical shift dimension to monitor subtle structural changes (represented by α and β anomers) which are involved in the acyl migration of isomers. Although the 2-, 3-, and 4-*O*-aryl transacylation products recovered from urine were chromatographically resolved, the α and β anomers were not completely resolved.

Figure 10 is the 600 MHz ^{1}H-NMR spectrum for the peak eluting at 11.68 min, and corresponds to the 3-*O*-acylglucuronide anomers **(12)** and **(13)**, Scheme 3. In this case, the α and β anomers were present in nearly equal amounts, although the doublet resonance H5′ signal (δ 4.30 ppm) is slightly suppressed relative to the β anomer H5′ signal at $\delta = 3.97$ ppm. The latter suppression is due to solvent irradiation of the HOD signal centered at $\delta = 4.35$ ppm. This paper nicely illustrates the practical utility of HPLC/^{1}H-NMR in the characterization of drug glucuronides in whole biofluids. This study also demonstrates that a 400 MHz ^{1}H-NMR detector was adequate for most glucuronides in this study, but the 600 MHz ^{1}H-NMR detector provided better sensitivity and spectral dispersion for one pair of isomeric glucuronides.

In a study by Ehlhardt et al.[43] the metabolites of the multidrug resistance modulator CY335979 were identified in a more complex matrix, namely bile. The rat bile metabolic pathway is outlined below.

The HPLC/MS profile clearly distinguished **(18)** and **(19)** ($m/z = 720$; Scheme 4) from **(20)** ($m/z = 736$) because of a parent peak 16 mass units greater in value. However, the HPLC/^{1}H-NMR (500 MHz) profile was required to distinguish metabolites **(18)** and **(19)**. As indicated in Figure 11, stopped-flow spectra HPLC/NMR spectra for peaks at 24.6 min and 24.8 min, corresponds to elution times for metabolite **(18)** and a mixture of **(18)** and **(19)**, respectively. This work clearly illustrates the utility of the HPLC/NMR technique to analyze metabolites in both liver microsome and bile matrices.

Finally, the rapid identification of naphthylisoquinoine alkaloids in crude plant extracts using HPLC/NMR and rotating frame Overhauser enhanced spectroscopy (ROESY) was recently reported by Bringmann et al.[44]

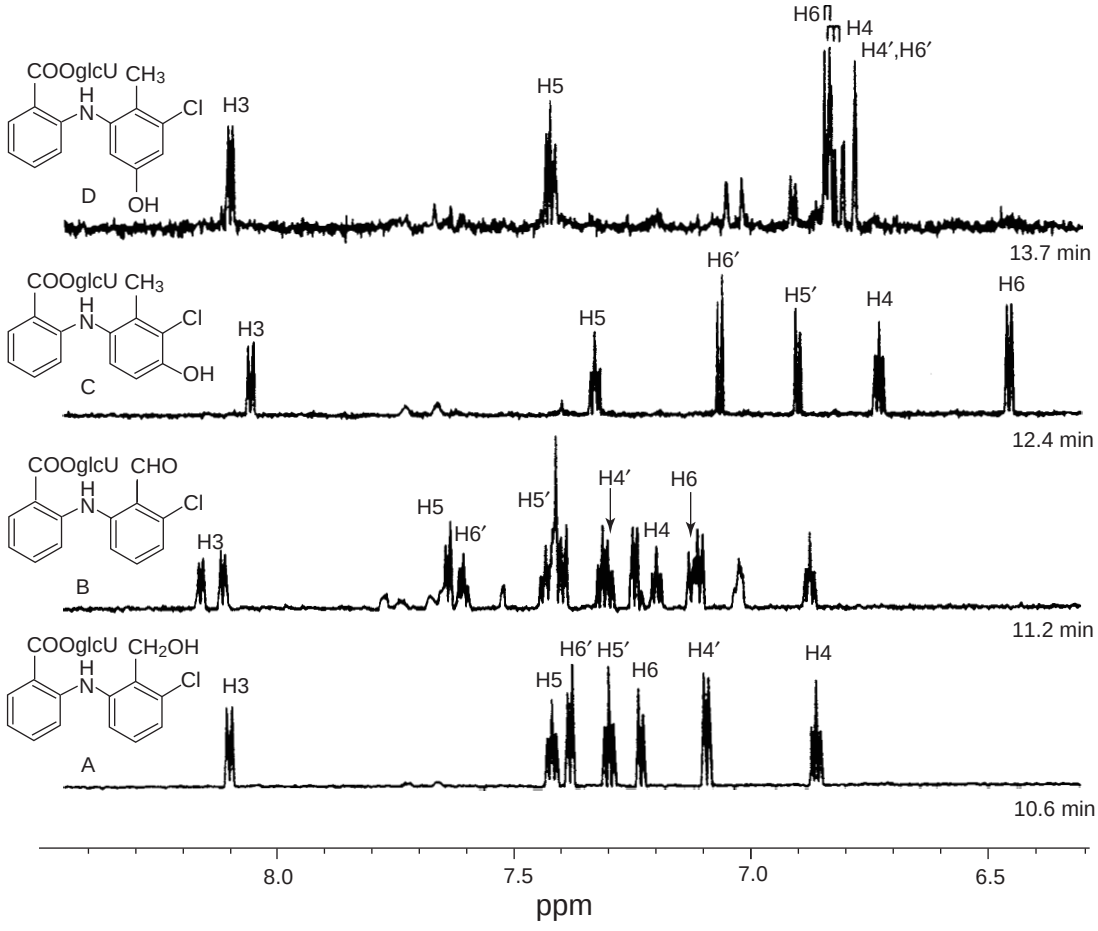

Figure 9 800 MHz one-dimensional ^{1}H-NMR spectra obtained from an HPLC/NMR experiment on the 60% methanol solid-phase extraction (SPE) fraction of a urine sample containing metabolites of tolfenamic acid (Structure 9). The aromatic spectral region (δ 6.3–8.5) is shown for clarity. All the metabolites seen were present in the form of glucuronic acid conjugates, identified on the basis of the resonance of the β-1′ proton at δ 5.8 (not shown in the figure). Additionally, the metabolic state of the CH_3 side chain was identified on the basis of the presence of a singlet at δ 2.3. H3–H6 and H3′–H6′ are the protons on the aromatic rings of tolfenamic acid and its metabolites. Spectrum A corresponds to metabolite IXa (shown), B corresponds to metabolites 9a and 9b (shown), C corresponds to metabolite 9c (shown), and D corresponds to metabolite 9d (shown). (Reprinted with permission, *Analytical Chemistry*.[41])

4 MISCELLANEOUS APPLICATIONS

One important extension of the HPLC/NMR technique from the present NMR detection volumes of 50–150 μL is the progression to nanoliter volumes (i.e. 40–200 nL). The development of nanoliter level flow cells and corresponding NMR microcoils has been hampered in the past by the loss in spectral resolution usually observed in preparing these devices. This is usually the result of differences in the magnetic susceptibility of the sample, glass support, and micro-cell. However, Olson et al.[6] have made considerable progress in suppressing the problems associated with magnetic susceptibility variations. In one of the first applications in a study by Albert and colleagues, capillary HPLC/^{1}H-NMR was used to determine the structure of retinyl acetate dimers[45] at nanoliter detection volumes (ca. 200 nL). Another possible application of nanoliter-volume NMR detectors is for separations based on CZE which allows rapid and efficient separation of charged analytes and biomolecules in small volumes. A diagram of a CE/NMR experimental layout is shown in Figure 12. Recently, Pusecker et al.[46] extended the capillary electrophoresis (CE) approach to CEC which is carried out in packed capillary columns. The separation of some hop bitter acids (humolones and isohumlones) was reported by this group using stopped-flow CZE/NMR.

Another variant is SFC and supercritical fluid extractions (SFE) with ^{1}H-NMR detection (SFC/^{1}H-NMR). This technique was originally reported in 1988 by

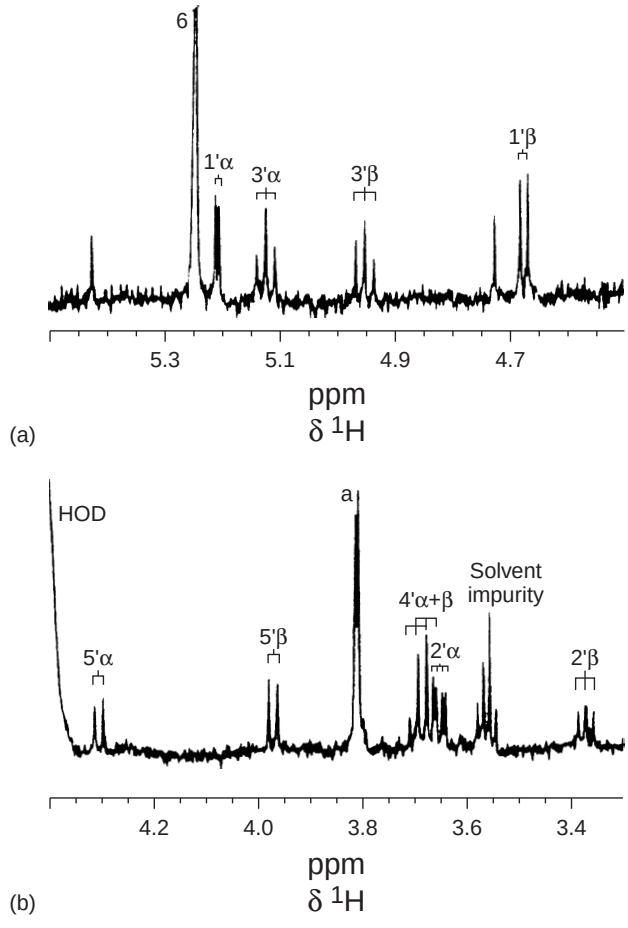

Figure 10 Partial 600 MHz ^{1}H-NMR spectrum (320 scans) of the mixture of α and β anomers of the 3-*O*-acylglucuronide (Structure 13). (Reprinted with permission, *Analytical Chemistry*.[42])

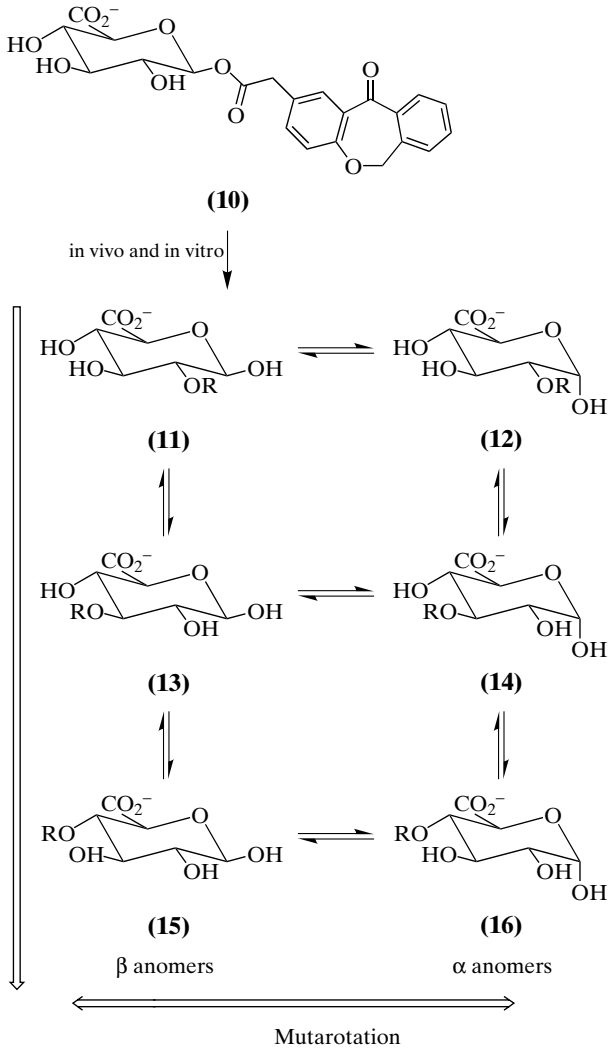

Scheme 3

Allen et al. for the analysis of fuel samples.[14] More recently, others have reported results for extraction of caffeine from roasted coffee (SFC/^{1}H-NMR) and the extraction of alkyl and aryl phthalates (Figure 13) utilizing SFE/^{1}H-NMR.[15,16] One major advantage in both (SFC and SFE) is the avoidance of background signals with the most common supercritical fluid CO_2.

Except for a few studies[40,47] utilizing HPLC/^{19}F-NMR, there have been relatively few reports employing other NMR active nuclides (such as ^{31}P, ^{15}N, and ^{13}C) for HPLC monitoring applications. Clearly, a limitation in the past has been the poor sensitivity of other important nuclides (^{13}C).

However, in the direct coupling of LC with ^{13}C, DNP detection (LC/^{13}C-DNP) has been employed to alleviate the sensitivity constraints of LC/^{13}C-NMR. Work by Stevenson and Dorn[48] has described directly coupled LC with continuous flow solid–liquid intermolecular transfer (SLIT) ^{13}C-DNP detection using a silica phase immobilized nitroxide (SPIN) sample (LC/^{13}C-DNP). Shown in Figure 14 is the LC/^{13}C-DNP profile of a single injected component of chloroform with carbon tetrachloride as the chromatographic solvent. The LC/^{13}C-DNP profile for the separation of chlorinated hydrocarbons is illustrated in Figure 15. Although this is a contrived example, it is noteworthy that in the corresponding flow ^{13}C-NMR (no microwave irradiation) experiment no signals were detected. With the improved SPIN samples used in this study, a ^{13}C-DNP enhancement of approximately 100 has been obtained in comparison with static thermal ^{13}C-NMR Boltzmann magnetization at 4.7 T. Clearly, improvements in higher field superconducting magnet design, instrumentation, and other polarization transfer experiments will provide continued evolution of improvements in NMR signal enhancement.

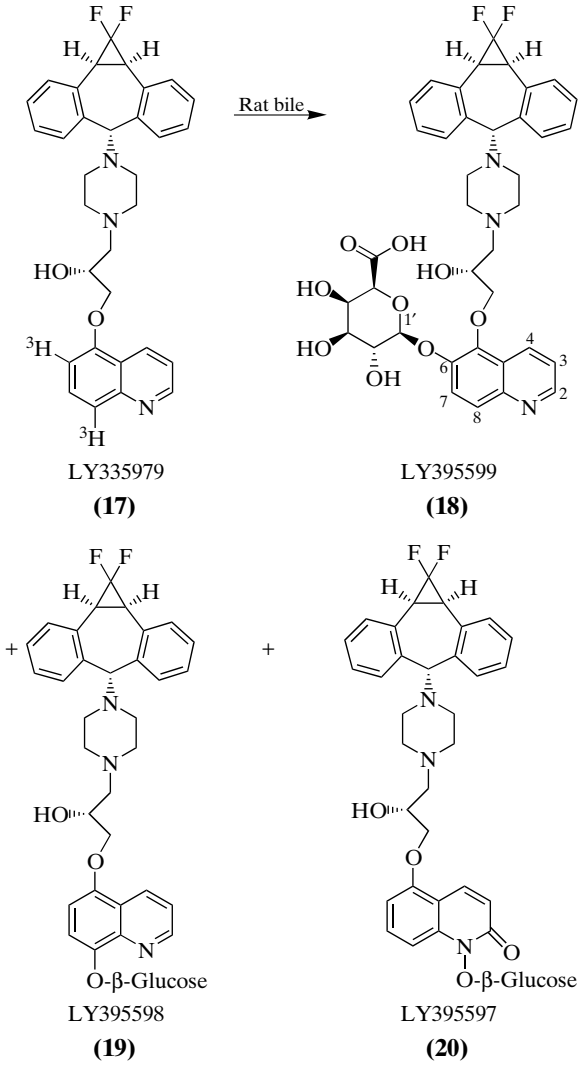

Scheme 4

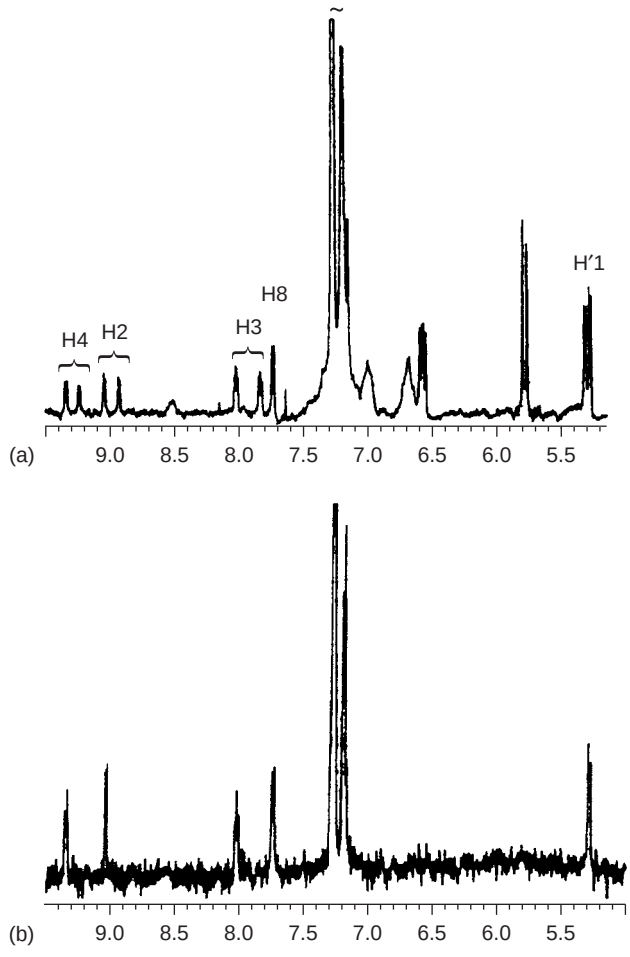

Figure 11 LC/NMR stop-flow ^{1}H-NMR spectrum. Expansion of the 5.0–9.5 ppm region of the stop-flow ^{1}H spectrum, (a) corresponding to the peak with a R_t (retention time) of 24.8 min in rat bile, and (b) analysis of the leading edge of that peak in a subsequent HPLC experiment 26 000 scans (11.8 hr) peak at 26.6 min. (Reprinted with permission, *Drug Metabolism and Disposition.*[43])

5 PERSPECTIVE AND FUTURE DEVELOPMENTS

It is reasonable to assume that the HPLC/NMR technique will continue to evolve as one of the dominant tools in high-resolution NMR. The commercial expansion, from a nonexistent commercial product in the early 1990s to a market today with multimillion dollar sales, further support this development. The commercial availability of high-resolution and high-field superconducting magnets suggests continued advances in the HPLC/NMR technique. Additional related techniques include direct injection of samples, nanoliter scale volumes, and improved cryoprobe NMR detectors, providing complimentary supporting approaches for the flow and HPLC/NMR techniques. One of the

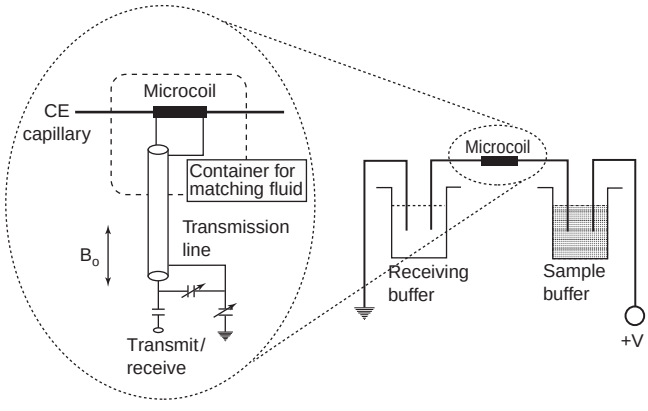

Figure 12 Experimental layout for CE/NMR showing the separation capillary and NMR probe. (Reprinted with permission, *Analytical Chemistry*.[13])

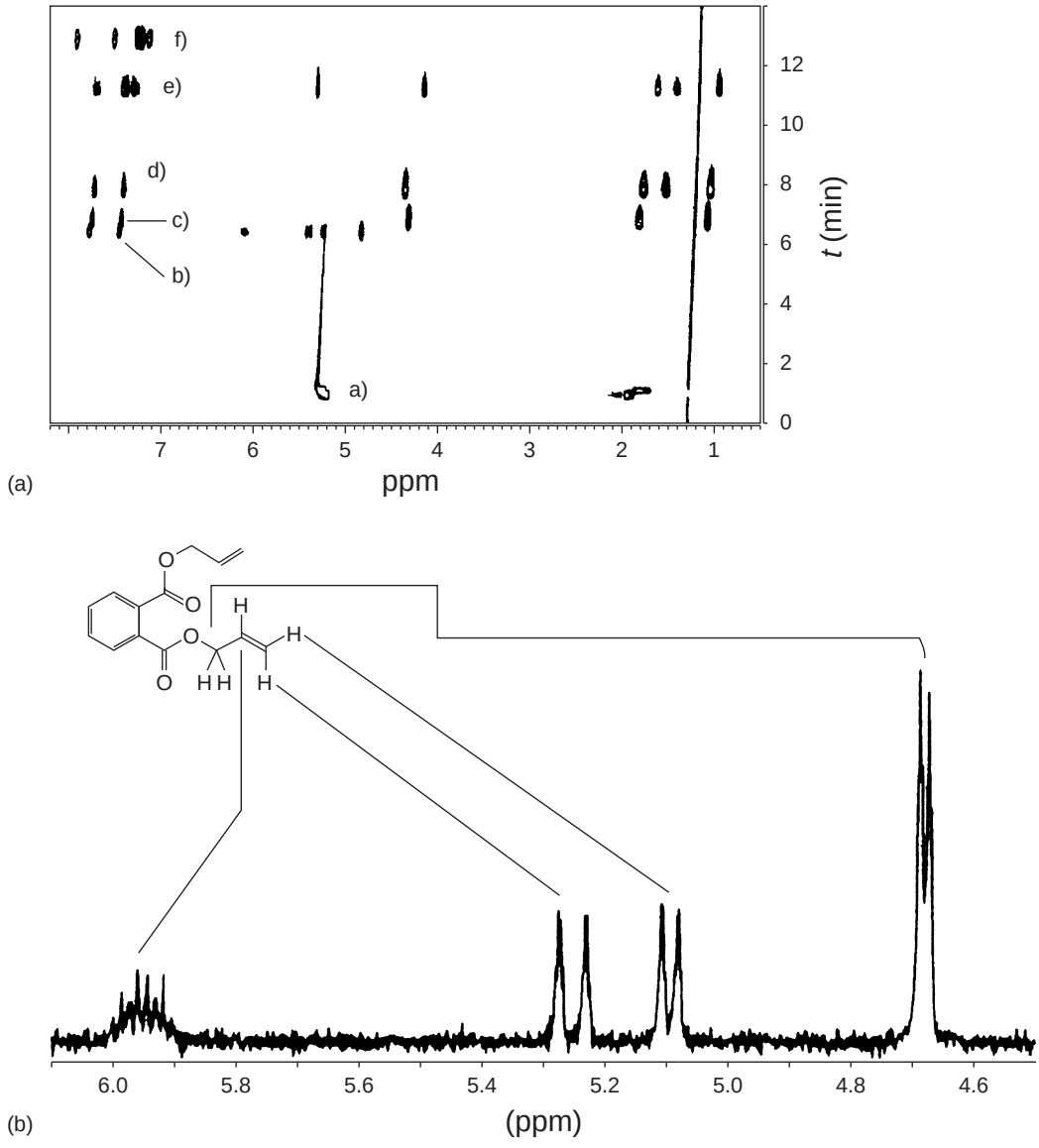

Figure 13 **(a)** ^{1}H-NMR chromatogram (contour plot, 400 MHz) of a SFC separation of five plastifiers: (a) CH_2Cl_2, (b) diallyl, (c) di-*n*-propyl, (d) di-*n*-butyl, (e) benzyl *n*-butyl, and (f) diphenylphthalate. **(b)** continuous-flow ^{1}H-NMR spectrum of diallylphthalate in supercritical CO_2; flow rate 2 mL min^{-1}, pressure approximately 136 bar; temperature 321 K. (Reprinted with permission, *Analytical Chemistry*.[15])

ultimate advantages of this technology is to reduce the labor-intensive preparation of individual sample tubes (and corresponding reference and lock material) for each separate sample. These conventional approaches also have the inherent problem of introducing new impurities. The HPLC/NMR technique is also readily amenable to monitoring photosensitive products or metabolites in a noninvasive fashion and can be readily coupled to other detectors (e.g. MS). As illustrated by the examples in this review, samples consisting of relatively complex matrices are readily analyzed by HPLC/NMR.

From the chromatographic perspective, the availability of commercial chiral chromatographic phases for separation of enantiomers suggests a future marriage of chiral chromatography with NMR detection as a new tool for pharmaceutical development of next-generation single-enantiomer drugs. The present limitation of HPLC/NMR to only the most sensitive NMR nuclides (^{1}H and ^{19}F) will undoubtedly expand to other nuclides with reasonable sensitivity (e.g. HPLC/^{31}P-NMR) in the near future. Even nuclides with relatively poor sensitivity are potentially amenable with stopped-flow and inverse detection (via ^{1}H) techniques.

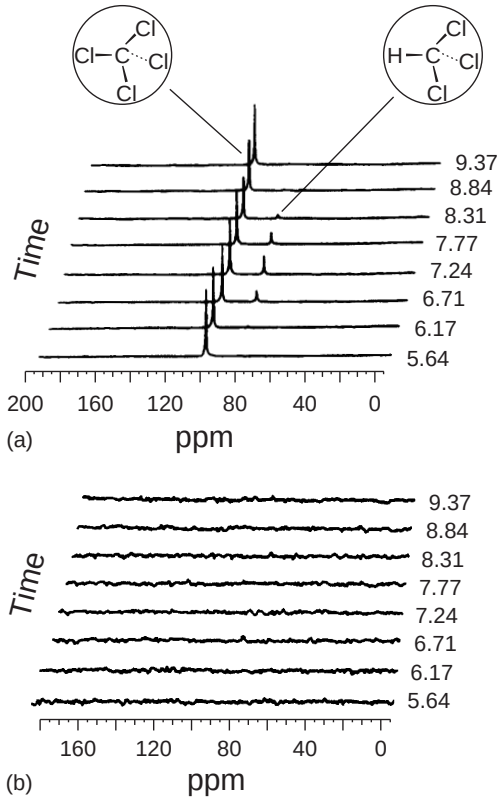

Figure 14 (a) LC/^{13}C-DNP profile: 500 µL of $HCCl_3$ injected, 3.5 mL min^{-1}, 10 scans per file, 31 s per file, and CCl_4 mobile phase. (b) LC/^{13}C-NMR profile (without DNP): 500 µL of $HCCl_3$ injected, 3.5 mL min^{-1}, 10 scans per file, and CCl_4 mobile phase. (Reprinted with permission, *Analytical Chemistry*.[48])

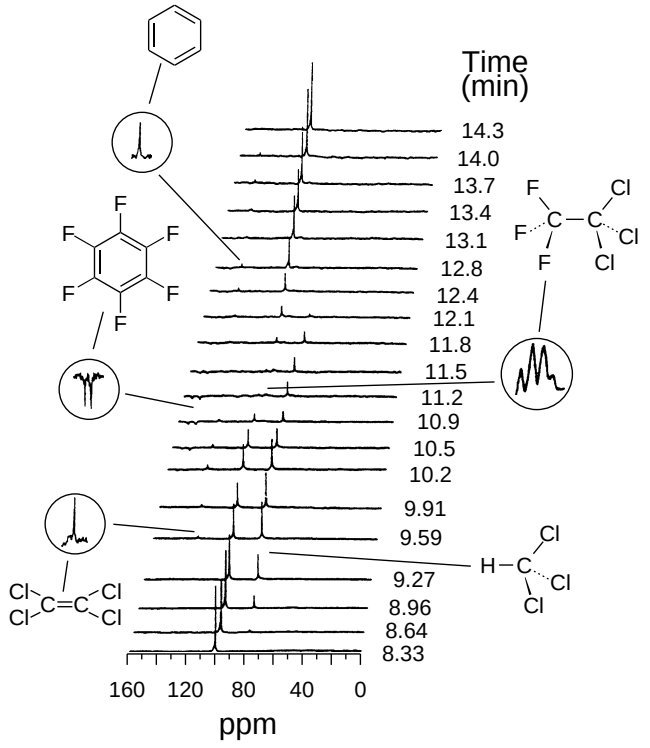

Figure 15 LC/^{13}C-DNP profile: 1000 µL chloroform, 1500 µL of tetrachloroethylene, 2500 µL of hexafluorobenzene, 1500 µL 1,1,1-trichlorotrifluoroethane, 3000 µL of benzene mixture, 2.5 mL min^{-1}, five scans per file, 19 s per file, and CCl_4 mobile phase. ^{19}F decoupling was not employed. ^{13}C-NMR signals for 1,1,1-trichlorotrifluoroethane ($^2J_{CF}$ quartet under heavy apodization, line broadening 11 Hz) and hexafluorobenzene ($^1J_{CF}$ doublet) are observed. (Reprinted with permission, *Analytical Chemistry*.[48])

ABBREVIATIONS AND ACRONYMS

CE	Capillary Electrophoresis
CEC	Capillary Electrochromatography
CE/NMR	Capillary Electrophoresis Nuclear Magnetic Resonance
CZE	Capillary Zone Electrophoresis
DNP	Dynamic Nuclear Polarization
GPC	Gel Permeation Chromatography
HOD	Monodeuterated Water
HPLC	High-performance Liquid Chromatography
LC	Liquid Chromatography
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
PMMA	Poly(methylmethylacrylate)
RI	Refractive Index
ROESY	Rotating Frame Overhauser Enhanced Spectroscopy
R_t	Retention Time
SEC	Site-exclusion Chromatography
SFC	Supercritical Fluid Chromatography
SFE	Supercritical Fluid Extractions
SLIT	Solid–Liquid Intermolecular Transfer
S/N	Signal to Noise Ratio
SPE	Solid-phase Extraction
SPIN	Silica Phase Immobilized Nitroxide
T_1	Spin–Lattice Relaxation Time
T_2	Spin–Spin Relaxation Time
TFA	Trifluoroacetic Acid
UV	Ultraviolet
WET	Water Elimination T_1 (Spin–Lattice Relaxation Time)

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance in Medicine, High Resolution Ex Vivo • Multinuclear Magnetic Resonance Spectroscopic Imaging

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules • Nuclear Magnetic Resonance of Biomolecules

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring • Solid-phase Microextraction in Environmental Analysis • Supercritical Fluid Extraction of Organics in Environmental Analysis

Food **(Volume 5)**
Near-infrared Spectroscopy in Food Analysis • Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Biological Matrices: Pesticides Content Sampling, Sample Preparation and Preservation

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Mass Spectrometry, High-resolution, (Homolog)-type Analysis of Petroleum and Synfuel Distillates and Refinery Streams • Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Combinatorial Chemistry Libraries, Analysis of • Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis • Solid-phase Extraction and Clean-up Procedures in Pharmaceutical Analysis

Process Instrumental Methods **(Volume 9)**
Chromatography in Process Analysis • Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Liquid Chromatography **(Volume 13)**
Liquid Chromatography: Introduction • Column Theory and Resolution in Liquid Chromatography • Supercritical Fluid Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • Liquid Chromatography/Mass Spectrometry

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Relaxation in Nuclear Magnetic Resonance, General

REFERENCES

1. N. Watanabe, E. Niki, 'Direct-coupling of FT-NMR to High Performance Liquid Chromatography', *Proc. Jpn. Acad. Ser. B.*, **54**, 194 (1978).
2. E. Bayer, K. Albert, M. Neider, E. Grom, T. Keller, 'On-line Coupling of High Performance Liquid Chromatography and Nuclear Magnetic Resonance', *J. Chromatography*, **186**, 497–507 (1979).
3. J. Buddrus, H. Herzog, 'Coupling of HPLC and NMR', *Org. Magn. Reson.*, **13**, 153 (1980).
4. J.F. Haw, T.E. Glass, D.W. Hausler, E. Motell, H.C. Dorn, 'Direct Coupling of a Liquid Chromatograph to a Continuous Flow Hydrogen Nuclear Magnetic Resonance Detector for Analysis of Petroleum and Synthetic Fuels', *Anal. Chem.*, **52**, 1135–1140 (1980).
5. H.C. Dorn, '^{1}H-NMR: A New Detector for Liquid Chromatography', *Anal. Chem.*, **56**, 747A–758A (1984).
6. D.A. Laude, Jr, C.L. Wilkins, 'Nuclear Magnetic Resonance Detection for the On-line Identification of Liquid Chromatography Eluents', *Trends Anal. Chem.*, **5**, 230 (1986).
7. K. Albert, E. Bayer, 'High-performance Liquid Chromatography–Nuclear Magnetic Resonance On-line Coupling', *Trends Anal. Chem.*, **7**(8), 288–292 (1988).
8. M. Spraul, M. Hofmann, J.C. Lindon, J.K. Nicholson, I.D. Wilson, 'Liquid Chromatography Coupled with High-field Proton Nuclear Magnetic Resonance Spectroscopy; Current Status and Future Prospects', *Anal. Proc.*, **30**, 390 (1993).
9. K. Albert, 'On-line Use of NMR Detection in Separation Chemistry', *J. Chromatogr. (A)*, **703**, 123–147 (1995).
10. S.A. Korhammer, A. Bernruether, 'Hyphenation of High-performance Liquid Chromatography (HPLC) and Other Chromatographic Techniques (SFC, GPC, GC, CE) with Nuclear Magnetic Resonance (NMR)', *Fresenius' J. Anal. Chem.*, **354**, 131–135 (1996).
11. J.C. Lindon, J.K. Nicholson, I.D. Wilson, 'Direct Coupling of Chromatographic Separations to NMR Spectroscopy', *Prog. Nucl. Magn. Reson. Spectrosc.*, **29**, 1–49 (1996).
12. H.C. Dorn, 'Flow NMR', in *Encyclopedia of NMR*, eds. D. Grant, K.K. Harris, John Wiley & Sons, 2026–2037, Vol. 3, 1996.
13. D.L. Olson, M.E. Lacey, J.V. Sweedler, 'The Nanoliter Niche', *Anal. Chem.*, 257A (1998).
14. L.A. Allen, T.E. Glass, H.C. Dorn, 'Direct Monitoring of Supercritical Fluids and Supercritical Chromatographic Separations by Proton Magnetic Resonance', *Anal. Chem.*, **60**, 390 (1988).
15. K. Albert, U. Braumann, L.H. Tseng, G. Nicholson, E. Bayer, M. Spraul, M. Hofmann, C. Dowle, M. Chippendale, 'On-line Coupling of Supercritical Fluid Chromatography and Proton High-field Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **66**, 3042–3046 (1994).

16. U. Braumann, H. Handel, K. Albert, R. Ecker, M. Spraul, 'On-line Monitoring of the Supercritical Fluid Extraction Process with Proton Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **67**, 930–935 (1995).
17. D.A. Laude, R.W.K. Lee, C.L. Wilkins, 'Reversed-phase High-performance Liquid Chromatography/Nuclear Magnetic Resonance Spectrometry Separations of Biomolecules with 1-1 Hard Pulse Solvent Suppression', *Anal. Chem.*, **57**, 1464–1469 (1985).
18. K. Albert, M. Kunst, E. Bayer, M. Spraul, W. Bermel, 'Reversed-phase High-performance Liquid Chromatography–Nuclear Magnetic Resonance On-line Coupling with Solvent Non-excitation, *J. Chromatogr.*, **463**, 355–363 (1989).
19. S.H. Smallcombe, S.L. Patt, P.A. Keifer, 'WET Solvent Suppression and Its Applications to LC NMR and High-resolution NMR Spectroscopy', *J. Magn. Reson. A*, **117**, 295–303 (1995).
20. D.A. Laude, C.L. Wilkins, 'Direct-linked Analytical Scale High-performance Liquid Chromatography/Nuclear Magnetic Resonance Spectrometry', *Anal. Chem.*, **56**, 2471–2475 (1984).
21. C.H. Fyfe, M. Cocivera, W.H. Damji, T.A. Hostetter, O. Sproat, J. O'Brien, 'Apparatus for the Measurement of Transient Species and Effects in Flowing Systems by High-resolution Nuclear Magnetic Resonance Spectroscopy', *J. Magn. Reson.*, **23**, 377–384 (1976).
22. J.F. Haw, T.E. Glass, H.C. Dorn, 'Continuous Flow High Field Nuclear Magnetic Resonance Detector for High-performance Liquid Chromatographic Analysis of Fuel Samples', *Anal. Chem.*, **53**, 2327–2332 (1981).
23. J.F. Haw, T.E. Glass, H.C. Dorn, 'Analysis of Coal Conversion Recycle Solvents by High-performance Liquid Chromatography with Nuclear Magnetic Resonance Detection', *Anal. Chem.*, **53**, 2332–2336 (1981).
24. J.F. Haw, T.E. Glass, H.C. Dorn, 'Liquid Chromatography/Proton Nuclear Magnetic Resonance Spectrometry Average Composition Analysis of Fuels', *Anal. Chem.*, **55**, 22–29 (1983).
25. R.N. Hazlett, H.C. Dorn, T.E. Glass, 'Analysis of Middle Distillate Fuels By Flow Liquid Chromatography/Proton Nuclear Magnetic Resonance', Magnetic Resonance. Introduction, *Advanced Topics and Applications to Fossil Energy*, 709–720 (1984).
26. R.N. Hazlett, H.C. Dorn, T.E. Glass, 'Analysis of Middle Distillate Fuels by Flow Liquid Chromatography/Proton Nuclear Magnetic Resonance', *NATO ASI Ser., Ser. C, 124 (Magn. Reson.)*, 709–720 (1984).
27. J.F. Haw, T.E. Glass, H.C. Dorn, 'Conditions for Quantitative Flow FT-^{1}H NMR Measurements Under Repetitive Pulse Conditions', *J. Magn. Reson.*, **49**, 22–31 (1982).
28. K. Allen Caswell, T.E. Glass, M. Swann, H.C. Dorn, 'Rapid Prediction of Various Physical Properties for Middle Distillate Fuels Utilizing Directly Coupled Liquid Chromatography/^{1}H Nuclear Magnetic Resonance', *Anal. Chem.*, **61**, 206 (1989).
29. K. Ute, M. Kashiyama, K. Oka, K. Hatada, 'On-line GPC/Analysis of the Mixture of Chloral Oligomers', *Makromol. Chem., Rapid Commun.*, **11**, 31–35 (1990).
30. K. Ute, K. Hatada, 'Analysis of Polymers and Oligomers by On-line GPC/NMR Using a 500 MHz Spectrometer as a Detector of GPC', *Anal. Sci.*, **7**, 1629–1632 (1991).
31. K. Hatada, K. Ute, T. Kitayama, T. Nishimura, M. Kashiyama, N. Fujimoto, *Polym. Bull. (Berlin)*, **23**, 549–554 (1990).
32. K. Hatada, K. Ute, M. Kashiyama, M. Imanari, 'Direct Determination of Molecular Weight and Its Distribution by the Absolute Calibration Method Using the On-line GPC/NMR', *Polym. J. (Tokyo)*, **22**, 218–222 (1990).
33. K. Ute, R. Nimi, S. Hongo, K. Hatada, 'Direct Determination of Molecular Weight Distribution by Size Exclusion Chromatography with 750 MHz ^{1}H NMR Detection (On-line SEC–NMR)', *Polymer J.*, **30**(5), 439–443 (1998).
34. H. Pasch, W. Hiller, R. Haner, 'Investigation of the Tacticity of Oligostyrenes by On-line h.p.l.c./^{1}H n.m.r.', *Polymer*, **39**, 1515 (1998).
35. M. Spraul, M. Hofmann, P. Dvortsak, J.K. Nicholson, I.D. Wilson, 'Liquid Chromatography Coupled with High-field Proton NMR for Profiling Human Urine for Endogeneous Compounds and Drug Metabolites', *J. Pharm. Biomed. Anal.*, **10**, 601–605 (1992).
36. M. Spraul, M. Hofmann, J.C. Lindon, J.K. Nicholson, I.D. Wilson, 'Liquid Chromatography Coupled with High-field Proton Nuclear Magnetic Resonance Spectroscopy: Current Status and Future Prospects', *Anal. Proc.*, **30**, 390–393 (1993).
37. S. Johnson, E.D. Morgan, I.D. Wilson, M. Spraul, M. Hofmann, 'Photo-isomerization of Azadirachtin Studied by High Performance Liquid Chromatography Coupled to High Field Proton NMR Spectroscopy', *J. Chem. Soc., Perkin Trans. I*, 1499 (1994).
38. I.D. Wilson, J.K. Nicholson, M. Hofmann, M. Spraul, J.C. Lindon, 'Investigation of the Human Metabolism of Antipyrine Using Coupled Liquid Chromatography and Nuclear Magnetic Resonances Spectroscopy of Urine', *J. Chromatogr.*, **617**, 324–328 (1993).
39. J.P. Shockcor, S.E. Unger, I.D. Wilson, P.J.D. Foxall, J.R. Nicholson, J.C. Lindon, 'Combined HPLC, NMR Spectroscopy, and Ion-trap Mass Spectrometry with Application to the Detection and Characterization of Xenobiotic and Endogenous Metabolites in Human Urine', *Anal. Chem.*, **68**, 4431–4451 (1996).
40. M. Spraul, M. Hofmann, I.D. Wilson, E. Lenz, J.K. Nicholson, J.C. Lindon, 'Coupling of HPLC with ^{19}F- and ^{1}H-NMR Spectroscopy to Investigate the Human Urinary Excretion of Flurbiprofen Metabolites', *J. Pharm. Biomed. Anal.*, **11**, 1009–1015 (1993).
41. U.G. Sidelmann, U. Braumann, M. Hofmann, M. Spraul, J.C. Lindon, J.K. Nicholson, S.H. Hansen, 'Directly Coupled 800 MHz HPLC–NMR Spectroscopy of Urine and Its Application to the Identification of the Major Phase

II Metabolites of Tolfenamic Acid', *Anal. Chem.*, **69**, 607–612 (1997).

42. E.M. Lenz, D. Greatbanks, I.D. Wilson, M. Spraul, M. Hofmann, J. Troke, J.C. Lindon, J.K. Nicholson, 'Direct Characterization of Drug Glucuronide Isomers in Human Urine by HPLC–NMR Spectroscopy: Application to the Positional Isomers of 6,11-Dihydro-11-oxodibenz[*b*,*e*]oxepin-2-acetic Acid Glucuronide', *Anal. Chem.*, **68**, 2832–2837 (1996).
43. W.J. Ehlhardt, J.M. Woodland, T.M. Baughman, M. Vandenbranden, S.A. Wrighton, J.S. Kroin, B.H. Norman, S.R. Maple, 'Liquid Chromatography/Nuclear Magnetic Resonance Spectroscopy and Liquid Chromatography/Mass Spectrometry Identification of Novel Metabolites of the MultiDrug Resistance Modulator LY335979 in Rat Bile and Human Liver Microsomal Incubations', *Drug Metab. Dispos.*, **26**(1), 42–51 (1998).
44. G. Bringmann, C. Gunther, J. Schluner, M. Ruckert, 'HPLC/NMR On-line Coupling Including the ROESY Technique: Direct Characterization of Naphthylisoquinoline Alkaloids in Crude Plant Extracts', *Anal. Chem.*, **70**, 2805–2811 (1998).
45. G. Schlotterbeck, L.-H. Tseng, H. Handel, U. Braumann, K. Albert, 'Direct On-line Coupling of Capillary HPLC with 1H NMR Spectroscopy for the Structural Determination of Retinyl Acetate Dimers: 2D NMR Spectroscopy in the Nanoliter Scale', *Anal. Chem.*, **69**, 1421–1425 (1997).
46. K. Pusecker, J. Schweitz, P. Gfrorer, L.H. Tseng, K. Albert, E. Bayer, 'On-line Coupling of Capillary Electrochromatography, Capillary Electrophoresis, and Capillary HPLC with Nuclear Magnetic Resonance Spectroscopy', *Anal. Chem.*, **70**, 3280–3285 (1998).
47. L.A. Allen, M.P. Spratt, T.E. Glass, H.C. Dorn, Directly Coupled LC/^{19}F FT NMR for the Analysis of Fluorine Containing Mixtures', *Anal. Chem.*, **60**, 675 (1988).
48. S. Stevenson, H.C. Dorn, '^{13}C Dynamic Nuclear Polarization: A Detector for Continuous-flow, On-line Chromatography', *Anal. Chem.*, **66**, 2993–2999 (1994).

Multidimensional Nuclear Magnetic Resonance of Biomolecules

David J. Craik
The University of Queensland, St Lucia, Australia

James Horne and Martin J. Scanlon
Monash University, Parkville, Australia

1 Introduction **1**
1.1 Physical Basis of Nuclear Magnetic Resonance Spectroscopy 1
1.2 Instrumentation 2
1.3 Nuclear Magnetic Resonance in Biological Systems 3
1.4 Nuclear Magnetic Resonance Parameters and Reference Standards 4
1.5 One-dimensional Nuclear Magnetic Resonance Spectroscopy 6
2 Two-dimensional Nuclear Magnetic Resonance **6**
2.1 Historical Basis of Two-dimensional Nuclear Magnetic Resonance 6
2.2 Description of a Generic Two-dimensional Experiment 7
2.3 Two-dimensional Nuclear Magnetic Resonance of Biomolecules 8
2.4 Limitations of Two-dimensional Nuclear Magnetic Resonance Methods 9
3 Three- and Four-dimensional Nuclear Magnetic Resonance **10**
3.1 More Dimensions 10
3.2 Homonuclear Methods 10
3.3 Heteronuclear Methods 10
3.4 Heteronuclear-edited Nuclear Magnetic Resonance 12
3.5 Triple-resonance Nuclear Magnetic Resonance 13
3.6 Limitations and Future Developments 17
Acknowledgments **18**
Abbreviations and Acronyms **18**
Related Articles **18**
References **19**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

Nuclear magnetic resonance (NMR) is associated with transitions, induced by radiofrequency (RF) irradiation, between energy levels of nuclei with nonzero spin quantum numbers in a magnetic field. Traditional ''one-dimensional'' (1-D) NMR spectra are presented as a plot of signal intensity versus applied frequency and provide information about the chemical environment of magnetically active nuclei. This is used to deduce information about the chemical structure and dynamics of molecules containing the magnetically active nuclei (most often protons). The frequencies of individual nuclei are referred to as their chemical shifts*. Multidimensional NMR encompasses a range of related techniques, based on the application of a variety of precisely timed RF pulses to the sample, which extends traditional 1-D NMR into two-, three-, or four-frequency dimensions. These additional frequency dimensions may be the same as the first frequency dimension, referred to as* homonuclear multidimensional NMR*, or different, referred to as* heteronuclear NMR*. In multidimensional NMR of biomolecules, the most common frequencies correspond to ^{1}H in the directly detected dimension and one or more of ^{1}H, ^{13}C, or ^{15}N in the additional dimensions. Peaks in such spectra are defined by their intensity and frequency coordinates in two, three, or four dimensions obtained by extrapolation back to the respective frequency axes, which are in turn determined by the nature of the applied RF pulse sequences. The intensity and chemical shifts of such peaks provide information about the chemical environment, molecular connectivity, or spatial proximity of the participating nuclei. Multidimensional NMR has major advantages over 1-D NMR: reduction of peak overlap and provision of information on connectivity (either through bonds or through space). Multidimensional NMR spectroscopy of biomolecules is usually done for samples in solution at millimolar concentration. Where experiments require the use of ^{13}C or ^{15}N frequencies, it is usually necessary to isotopically enrich the macromolecule with these nuclei to >90%. This is readily achieved with modern molecular biology techniques and does not present a major limitation. The biggest limitation relates to molecular size, with studies generally limited to macromolecules of less than approximately 35 kDa.*

1 INTRODUCTION

1.1 Physical Basis of Nuclear Magnetic Resonance Spectroscopy

NMR occurs when nuclei with nonzero spin quantum numbers (e.g. ^{1}H, ^{13}C, ^{15}N) are placed in a powerful magnetic field and subjected to RF irradiation. The derived signals provide valuable information about the

structure and dynamics of molecules containing the NMR-active nuclei, making NMR spectroscopy one of the most powerful techniques for examining molecules in solution. As the basic theory of NMR spectroscopy has been covered elsewhere in this publication (see **Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction**), the theoretical details are not repeated here, where the focus is on aspects unique to multidimensional NMR of biomolecules. It is sufficient to provide a brief reminder that the essential features of NMR spectroscopy are a powerful magnetic field, which results in the alignment of magnetically active nuclei into a discrete number of orientations (dependent upon the spin quantum number), and an RF field that is sufficient to disturb the alignment of the nuclei from their equilibrium positions. The signals detected during recovery (relaxation) of the nuclear magnetization back to its equilibrium value are manifest as oscillating voltages in an RF receiver coil and are subsequently amplified and digitized. The data are usually presented as a spectrum (a plot of intensity versus one or more frequency dimensions). In addition to the information provided in earlier articles of this publication, the detailed principles of 1-D and two-dimensional (2-D) NMR are covered in the monograph by Ernst et al.,[1] and the technical aspects of NMR relevant in biomolecular studies are addressed extensively by Wider.[2] Theoretical[3] and practical[4] aspects of macromolecular NMR have also been addressed in recent monographs. This article focuses on NMR spectroscopy of proteins. The applications of NMR to studies of other biomolecules have been addressed elsewhere.[5,6]

The general equation that governs the NMR phenomenon is Equation (1):

$$f = \gamma B_0 \tag{1}$$

where f is the observed resonance frequency, γ is the gyromagnetic ratio of the nucleus concerned, and B_0 is the applied magnetic field. The single most important factor that makes the NMR phenomenon useful in chemical and biological applications is that this basic frequency is moderated by the local electronic environment of the nucleus. Thus, the frequency of an individual nuclear spin in a molecule is given more correctly by Equation (2):

$$f = \gamma(1 - \sigma)B_0 \tag{2}$$

where σ is a shielding factor (of the order of a few parts per million) that reflects the local electronic environment of the individual nucleus. This shielding factor is the origin of the so-called chemical shift that discriminates between chemically inequivalent nuclei of a given type (e.g. protons). The separation of NMR signals via chemical shifts allows site-specific information to be obtained even in complex biomolecules.

1.2 Instrumentation

As protons are the most sensitive and most commonly detected nuclei, an NMR instrument is often specified in terms of its nominal proton frequency. In general, the sensitivity of NMR instruments increases with increasing frequency, and because of the stringent sensitivity requirements for biological samples, 500 MHz system (corresponding to a magnetic field of 11.7 T) tends to be the entry level of spectrometer for biological applications. The spectrometers currently used for biomolecular studies typically include 500-, 600-, 750-, and 800-MHz systems, with 900-MHz system on the immediate horizon. These systems utilize superconducting magnets, which are based on a solenoid of a suitable alloy (e.g. niobium/titanium or niobium/tin) immersed in a dewar of liquid helium. The extremely low temperature of the magnet itself (4.2 K) is well insulated from the sample chamber in the center of the magnet bore. The probe in which the sample is housed usually incorporates accurate temperature control over a range typically of 4–40 °C for biological samples. The RF coil in the probe is connected in turn to a preamplifier, receiver circuitry, analog-to-digital converter (ADC), and a computer for data collection.

Consideration of Equation (2) provides an insight into some of the RF requirements for NMR spectrometers. Because γ is substantially different for different nuclei, the absolute frequencies of different nuclei are quite different in a given magnetic field. For example, in a 500-MHz (11.7-T) spectrometer, the ^{1}H frequency is 500 MHz, but the ^{13}C frequency is only 125 MHz and the ^{15}N frequency is 50.6 MHz. This vast range of frequencies, combined with substantially different intrinsic sensitivities, means that it is impractical to simultaneously detect the signals from all nuclear types in a single NMR experiment. Instead, the RF probe and receiver electronics are selectively tuned for detection of a single nuclear type at any one time. The range of frequencies present for a particular nuclear type in a biomolecule depends on the range of different environments for that nucleus, and is reflected in the chemical shift ranges, typically 10 ppm for ^{1}H or 200 ppm for ^{13}C. These ranges correspond to approximately 5 and 25 kHz, respectively, in an 11.7-T magnetic field, and such frequency bandwidths are readily detected and digitized. Although only a single nuclear type (usually protons) is *detected* in a single experiment, most modern multidimensional NMR experiments involve the simultaneous *excitation* of multiple-frequency bands. Thus, NMR instruments have evolved to be multichannel

devices, with several frequency sources and with RF probes that typically contain two double-tuned coils for delivery of RF pulses at ^{1}H, ^{13}C, ^{15}N, and sometimes ^{2}H frequencies. In addition, the recent advent of methods based on pulsed-field gradients (PFGs) has added to the sophistication of NMR instruments and these methods are particularly important in the multinuclear experiments described in Section 3.3.

NMR was revolutionized in the late 1960s and early 1970s by the development of Fourier transform (FT) methods.[7] In this approach, a high-power RF pulse simultaneously excites a wide range of frequencies (~100 kHz) for a given nuclear type, giving a marked sensitivity advantage over earlier continuous wave (CW) methods, which essentially excite only one transition at a time. Figure 1 shows a schematic illustration of a simple FT experiment in which an RF pulse is followed by acquisition of a time-domain free induction decay (FID). Fourier transformation of the FID results in the conventional frequency-domain spectrum. As a result of the ubiquitous use of FT methods in modern biological NMR studies, spectrometers are typically equipped with sophisticated pulse-programming facilities so that RF pulses of accurately controlled duration, intensity, homogeneity, frequency, and phase can be programmed and applied to the sample via the RF probe. In multinuclear NMR experiments, the single pulse in Figure 1 is replaced by a series of pulses, as seen later.

So far we have discussed the magnet, RF and detection considerations, and FT methods. Another important aspect of modern NMR spectrometers for biological studies is an appropriately large data-handling system. Two-dimensional NMR data sets typically require several megabytes of disk storage, whereas multidimensional data sets may consume more than 40 MB of disk storage per experiment. For the determination of three-dimensional (3-D) structures of macromolecules, 6–12 2-D, 3D, or four-dimensional (4-D) data sets may be recorded for an individual sample. Data storage and processing requirements are thus significant.

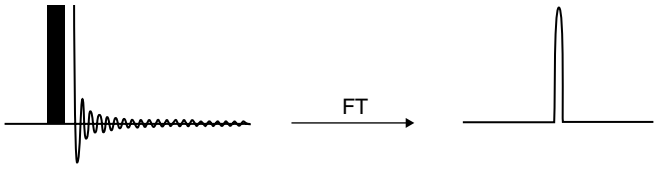

Figure 1 1-D pulsed FT-NMR experiment. The RF pulse tips the net magnetization, leading to an induced voltage or free induction decay (FID) in the receiver coil. Fourier transformation produces the usual frequency-domain representation of the spectrum.

1.3 Nuclear Magnetic Resonance in Biological Systems

Biological macromolecules contain only a very limited subset of the elements from the periodic table. In particular, the most abundant elements in proteins, nucleic acids, lipids, and carbohydrates include H, C, N, O, S, and P. All of these elements contain at least one isotope that is magnetically active and hence accessible via NMR. However, in some cases (^{17}O and ^{33}S), the properties of the nuclei make detection inconvenient or difficult, and in others (^{13}C and ^{15}N) the natural abundance, and hence sensitivity to detection, is very low. For detection of ^{13}C and ^{15}N it is usual to work with isotopically labeled proteins. Both ^{1}H and ^{31}P are essentially 100% abundant and are the most sensitive naturally occurring nuclei in biological macromolecules. The ^{1}H nucleus is, by far, the most widely applied and is the one detected in most modern biological NMR experiments.

Table 1 provides some frequency and sensitivity information on the nuclei that are commonly used in multidimensional NMR experiments of biological macromolecules. Other nuclei are often used to facilitate multidimensional NMR experiments but play a passive rather than an active role in the experiments. For example, ^{2}H is commonly used to dilute protons in large macromolecules to indirectly increase the sensitivity of various multidimensional experiments by reducing dipolar broadening. Many other nuclei have roles as selective probes. For example, ^{3}H and ^{19}F are used as isosteric replacements for protons because their low background levels in macromolecules allow their detection in the absence of interfering signals from other nuclei. However, most experiments involving these or other selective probe nuclei tend to involve simple 1-D or 2-D methods and are not described further here, where the emphasis is on multidimensional NMR. For multidimensional NMR applications, ^{1}H, ^{13}C, and ^{15}N are by far the most important nuclei.

In most intact biological systems, macromolecules are present only at very low concentrations, ranging from nano- to micromolar. Thus, in order to best mimic such systems, it is, in principle, desirable to conduct

Table 1 Properties of nuclei used in multidimensional NMR studies of macromolecules

Isotope	Spin	Frequency (MHz) at 11.7 T	Natural abundance (%)	Relative sensitivity[a]
^{1}H	1/2	500.0	99.98	1.000
^{13}C	1/2	125.7	1.11	0.016
^{15}N	1/2	50.7	0.37	0.001
^{31}P	1/2	202.4	100.00	0.066

[a] For equivalent numbers of nuclei (i.e. for 100% of the indicated isotope).

spectroscopic investigations at rather low concentrations. In general, sensitivity limitations preclude NMR studies at submicromolar concentrations for 1-D NMR experiments or submillimolar concentrations for 2-D or higher dimensional experiments. The additional complication of difficulty, in general, in obtaining gram quantities of biological macromolecules means that most studies are carried out at millimolar concentrations, with 1 mM having become almost the de facto standard for studies of protein solutions by multidimensional NMR methods. In the usual case where water is the solvent, the proton concentration (110 M) in water is typically up to 100 000 times higher than that of the solute (1 mM) in these solutions. One of the consequences of the use of FT methods and the need for digitization of the analog time domain NMR signal is that it is extremely difficult to adequately digitize those signals from macromolecular solutes in the presence of such a large water peak. Water-suppression methods are thus critical for multidimensional NMR studies for biological macromolecules.

The earliest and still very widely used approach for suppression of the water signal involves presaturation of the signal with a long RF pulse prior to collection of NMR data. The method is equally applicable to 1-D, 2-D, 3-D, or 4-D NMR experiments. Such presaturation methods require excellent shimming of the magnetic field so the water signal is as sharp as possible, thereby maximizing the efficiency of the applied radiation so that the minimal amount of RF power is used. Too much power for irradiation results in bleaching of adjacent signals and can result in a significant loss of information content, particularly for Hα proton signals, which occur with frequencies close to the water signal.

More recently, the advent of PFG pulses has revolutionized several aspects of multidimensional NMR experiments, including the procedures used for water suppression. The acronym given for a simple, but very effective and widely used, gradient-based method is WATERGATE.[8] The pulse sequence (shown in Figure 2) consists of two gradient pulses that sandwich a selective 180° pulse (or pulse train). The technique is based on the fact that all coherences dephased by the first gradient pulse are rephased by the second, provided that they have experienced the 180° flip. The selective 180° pulse is thus designed so that all coherences of interest are rotated by 180° but the signal at the water frequency is not rotated and therefore not rephased after the second gradient pulse. Water suppression of the order of 10^4–10^5 is achieved using this sequence, which can be incorporated into a variety of multidimensional NMR experiments.

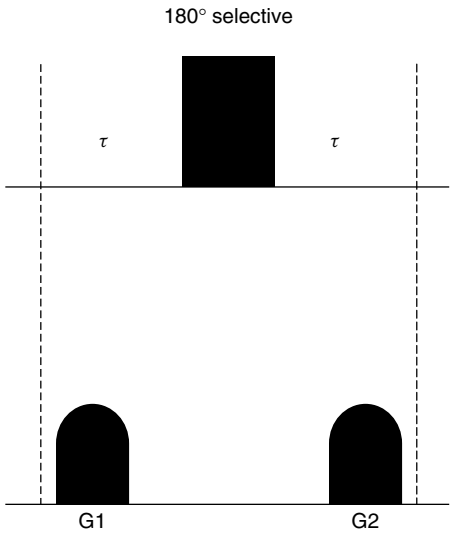

Figure 2 Schematic illustration of the WATERGATE pulse sequence.[8] Typical settings include a gradient pulse width of 1 ms, gradient amplitude 15–20 G cm^{-1}, and $\tau \sim 1$ ms. The selective 180° pulse often comprises a tailored excitation pulse train (e.g. a 3–9–19 sequence), which excites all frequencies outside the suppression band at the water frequency. This pulse sequence may be incorporated into a range of different 1-D and 2-D NMR experiments. G1 and G2 denote gradient pulses 1 and 2.

1.4 Nuclear Magnetic Resonance Parameters and Reference Standards

NMR is unique amongst the various forms of spectroscopy, in that there are so many parameters that can be measured, each providing information about different physical phenomena. Some of these parameters include chemical shifts, coupling constants, relaxation times, nuclear Overhauser effects (NOEs), peak intensities, and line widths. The theory behind many of these parameters is described in earlier articles of this publication, including **Chemical Shifts in Nuclear Magnetic Resonance**; **Parameters, Calculation of Nuclear Magnetic Resonance**; **Relaxation in Nuclear Magnetic Resonance, General**; **Scalar Couplings in Nuclear Magnetic Resonance, General**, and is hence not described in extensive detail here. However, because some aspects of the parameters have a central bearing on NMR of biomolecules, brief descriptions of the key elements are given below.

1.4.1 Chemical Shifts

Chemical shift defines the frequency at which an individual nucleus resonates and is perhaps the best known and most important NMR parameter. The most important factor determining chemical shifts is the electronic environment surrounding the nucleus and this depends on a number of factors, including local elements

of structure. Chemical shifts are expressed in units of parts per million and essentially indicate frequency offsets from a reference standard. Tetramethylsilane is the traditional standard of choice for organic applications of NMR but is not soluble in water, the medium used in most biological studies, so a water-soluble derivative, 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), is normally used for defining the zero point for ^{1}H- and ^{13}C-NMR shifts. A detailed discussion of the relative merits of various standards for these and other nuclei is given elsewhere.[9] Although a variety of standards have been proposed for heteronuclei such as ^{15}N or ^{31}P, a convenient solution to the referencing problem that is now widely adopted is to make use of the ''σ ratio'' method based on the relationship between frequency and γ defined by Equation (1) for a fixed magnetic field strength. The details of this method are described by Wishart et al.[9] Basically, the method involves the use of the following σ ratios for determining chemical shift zero points for heteronuclei: $^{13}C/^{1}H = 0.251449530$, $^{15}N/^{1}H = 0.101329118$. Measurement of the zero point for a single reference compound (DSS) for ^{1}H allows the zero points for other nuclei to be determined, thus saving both time and effort. The σ ratios (which reflect ratios of γ values in Equation 1) are universal and may be used on a variety of commercial spectrometers at any field strength.

The fact that the theory relating structural factors to chemical shifts is still not developed sufficiently to allow definitive predictions of chemical shifts in simple molecules, let alone macromolecules, does not diminish the usefulness of chemical shifts. From the measurements on large numbers of biomolecules, expected ranges of shifts for particular types of environments have been described for many nuclei, including ^{1}H, ^{13}C, and ^{15}N. The ''random coil'' shifts of ^{1}H nuclei in the 20 common amino acids that make up most protein structures are given in Table 2. This illustrates the general trend that the amide backbone shifts of proteins generally occur around 7–9 ppm, the α protons are in the range 4–5 ppm, and the signals from side-chain protons are typically in the range 0–4 ppm. In structured proteins, considerable variation in the shifts of individual protons from those given in Table 2 is expected but the random-coil values are nevertheless useful in identifying patterns of chemical shifts for different amino acids.

Table 2 ^{1}H chemical shifts for the 20 common amino acid residues[a] in random coil peptides

Residue	NH	Hα	Hβ	Others
Ala	8.24	4.32	1.39	–
Arg	8.23	4.34	1.89, 1.79	CγH$_2$ 1.70, 1.70
				CδH$_2$ 3.32, 3.32
				NH 7.17, 6.62
Asn	8.40	4.74	2.83, 2.75	NγH$_2$ 7.59, 6.91
Asp	8.34	4.64	2.84, 2.75	–
Cys	8.43	4.71	3.28, 2.96	–
Gln	8.32	4.34	2.13, 2.01	CγH$_2$ 2.38, 2.38
				NδH$_2$ 6.87, 7.59
Glu	8.42	4.35	2.09, 1.97	CγH$_2$ 2.31, 2.28
Gly	8.33	3.96	–	–
His	8.42	4.73	3.26, 3.20	2H 8.12
				4H 7.14
Ile	8.00	4.17	1.90	CγH$_2$ 1.48, 1.19
				CγH$_3$ 0.95
				CδH$_3$ 0.89
Leu	8.16	4.34	1.65, 1.65	Hγ 1.64
				CδH$_3$ 0.94, 0.90
Lys	8.29	4.32	1.85, 1.76	CγH$_2$ 1.45, 1.45
				CδH$_2$ 1.70, 1.70
				CεH$_2$ 3.02, 3.02
				NεH$_3^+$ 7.52
Met	8.28	4.48	2.15, 2.01	CγH$_2$ 2.64, 2.64
				CεH$_3$ 2.13
Phe	8.30	4.62	3.22, 2.99	2,6H 7.30
				3,5H 7.39
				4H 7.34
Pro		4.42	2.28, 2.02	CγH$_2$ 2.03, 2.03
				CδH$_2$ 3.68, 3.65
Ser	8.31	4.47	3.88, 3.88	–
Thr	8.15	4.35	4.22	CγH$_3$ 1.23
Trp	8.25	4.66	3.32, 3.19	2H 7.24
				4H 7.65
				5H 7.17
				6H 7.24
				7H 7.50
				NH 10.22
Tyr	8.12	4.55	3.13, 2.92	2,6H 7.15
				3,5H 6.86
Val	8.03	4.12	2.13	CγH$_3$ 0.97, 0.94

[a] The backbone shifts (Hα and NH) are from Wishart et al.[10] The remaining shifts are from Wüthrich.[11]

1.4.2 *Coupling Constants*

Coupling constants reflect the splitting of individual resonances that arises from through-bond (scalar) interactions between NMR-active nuclei. Among the applications to biological macromolecules, coupling constants are most valuable for providing dihedral angle constraints in structure calculations. This application owes its origin to the well-known Karplus relationship, originally derived in organic molecules, between three-bond coupling constants (^{3}J) and the intervening dihedral angle ϕ (Figure 3). In the case of proteins, the relevant coupling constant is $^{3}J_{\mathrm{HNH}\alpha}$, which defines the ϕ angle in the peptide backbone.[11]

1.4.3 *Relaxation Parameters T_1 and T_2*

The time constants T_1 (the spin–lattice relaxation time) and T_2 (the spin–spin relaxation time) define the rate of recovery of the nuclear magnetization following a perturbation from equilibrium: T_1 measures recovery

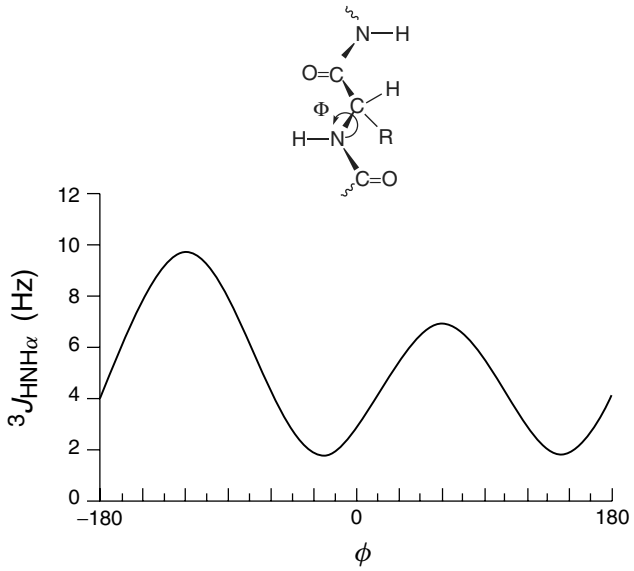

Figure 3 Illustration of the Karplus relationship between three-bond scalar coupling constants and the dihedral angle of the intervening bond. The relationship is indicated for the ϕ torsion angle of the peptide backbone, defined by $C'(i-1)-N(i)-C\alpha(i)-C'(i)$.

in the longitudinal direction (i.e. in the direction of the magnetic field) and T_2 defines the rate in the transverse direction. Both time constants are extremely sensitive to molecular motion and are widely used for the determination of correlation times and amplitudes of internal motions in proteins and nucleic acids. T_2 becomes smaller with increasing molecular size, and because it determines the rate of loss of nuclear magnetization it can become a limiting factor in NMR studies of larger molecules. It has an important bearing on the design of some of the complex multidimensional NMR pulse sequences described later in this article.

1.4.4 The Nuclear Overhauser Effect

The NOE was originally noticed as a change in the intensity of one signal in an NMR spectrum upon irradiation of other signals. The effect owes its origin to dipolar interactions between nuclei less than approximately 5 Å apart. Perturbation of the populations of one transition can cause changes in other related transitions, leading to intensity changes in 1-D spectra. In multidimensional NMR, NOEs are manifest as cross peaks in nuclear Overhauser effect spectroscopy (NOESY) spectra, with the intensity of the cross peak proportional to the size of the NOE. The value of NOEs is directly related to internuclear distances and so measurement of NOE cross-peak intensities can be used to derive distance restraints used in the calculation of 3-D structures of macromolecules.

1.5 One-dimensional Nuclear Magnetic Resonance Spectroscopy

The theory of multidimensional NMR, as it applies to macromolecular studies, is best introduced by building on the principles inherent in 1-D NMR spectroscopy, and this is conveniently done with an example. Figure 4(a) shows a typical 1-D spectrum of a medium-sized protein – in this case MiAMP1, a 70 amino acid protein from the seeds of macadamia nuts that exhibits antimicrobial activity.[12,13] The spectrum is shown to illustrate the strengths and weaknesses of 1-D NMR of biomolecules. The main advantage of such 1-D spectra is that they rapidly allow assessment of the folded state of biological macromolecules. This is usually judged by good dispersion of the amide resonances (7–10 ppm), shifts of Hα proton signals upfield (indicative of α helix) or downfield (indicative of β sheet) from random coil values, and upfield shifting of methyl resonances – usually an indication of a highly structured environment. These features are evident in the 1-D spectrum of Figure 4(a), but the spectrum also shows the main disadvantage of 1-D NMR of biomolecules – there is extensive overlap of the signals, and assignment of individual peaks in this spectrum is essentially impossible. The need for 2-D or higher dimensional spectra is thus clear.

2 TWO-DIMENSIONAL NUCLEAR MAGNETIC RESONANCE

2.1 Historical Basis of Two-dimensional Nuclear Magnetic Resonance

The original concept of 2-D NMR spectroscopy is widely attributed to a presentation by Jeener at a European summer school in 1971.[14] A formal demonstration of the principles of 2-D NMR was made in 1976.[15] The main advantage of 2-D over 1-D methods is the dispersion of signals into two frequency domains in a way that provides increased resolution as well as fundamental new information on connectivity between peaks. Figure 4(c) shows the relationship between 1-D and 2-D spectra, illustrated with a 2-D total correlation spectroscopy (TOCSY) spectrum of MiAMP1. The diagonal of this spectrum is essentially the 1-D spectrum shown in Figure 4(a), and the off-diagonal or cross peaks are the key features of 2-D NMR spectroscopy. As is probably clear from a brief inspection of the landscape view of the 2-D spectrum in Figure 4(b), it is far more convenient to view the 2-D spectrum in a contour map representation, as shown in Figure 4(c). In contour maps, the intensity of peaks (both diagonal and cross peaks) is represented by the number of contour levels.

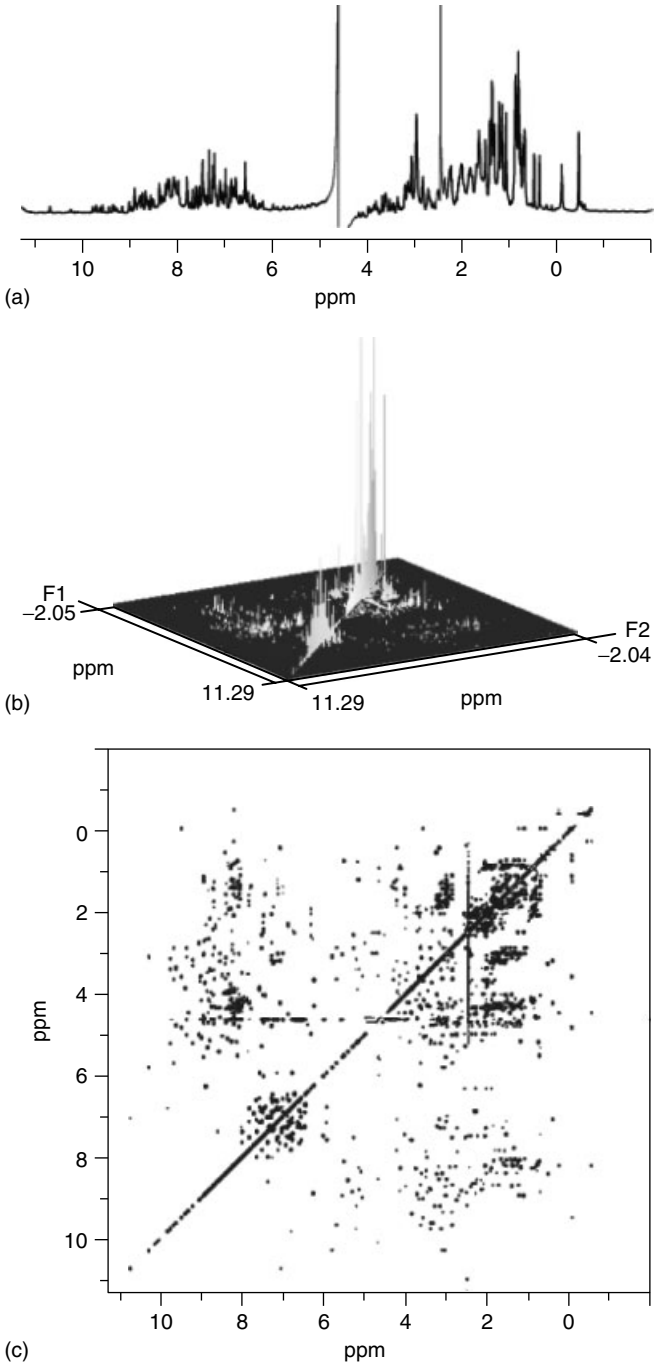

Figure 4 Illustrative 1-D and 2-D spectra of the antifungal protein MiAMP1.[12,13] (a) The 1-D spectrum recorded at 750 MHz. (b) The 2-D spectrum in landscape mode. The diagonal corresponds to the 1-D spectrum and the cross peaks indicate correlations between diagonal peaks. (c) A contour plot representation of the 2-D spectrum.

The 2-D NMR technique evolved rapidly from its inception, with the main limiting factor being instrument limitations. As the power of the new 2-D methods became clear, instrumentation rapidly improved to capitalize fully on the new methods. Instrumental improvements included the incorporation of multiple-frequency sources within the spectrometer console so that RF pulses at different frequencies could be applied as required. Pulse-programming hardware and software for accurately controlling the timing of these RF pulses was also critical, as was the general stability of the RF console to minimize spectral artifacts.

2.2 Description of a Generic Two-dimensional Experiment

The theory and applications of 2-D NMR spectroscopy to organic molecules are covered in another article in this publication (see **Two-dimensional Nuclear Magnetic Resonance of Small Molecules**) and elsewhere,[16,17] and because the basic principles are the same for small or large molecules we focus specifically on aspects relevant to 2-D, 3-D, and 4-D NMR of biomolecules.

The prototypical 2-D experiment involves a series of RF pulses (at least two) separated by an evolution delay that is incremented in successive steps of the 2-D experiment. In their most general form, 2-D NMR experiments may be thought of as consisting of four successive time periods comprising a preparation period; an evolution period, t_1, during which the spins are labeled according to their chemical shifts; a mixing period during which the spins are correlated with one another; and a detection period, t_2, where a FID is recorded. This experiment is repeated several times with linearly incremented values of t_1 to yield a data matrix $S(t_1, t_2)$. Fourier transformation in the t_2 dimension yields a set of n 1-D spectra in which the intensities of the resonances are sinusoidally modulated as a function of the t_1 duration. Subsequent Fourier transformation in the t_1 dimension yields the desired frequency-domain 2-D spectrum $S(\omega_1, \omega_2)$. In most homonuclear 2-D experiments, the diagonal corresponds to the 1-D spectrum, and the symmetrically placed cross peaks on either side of the diagonal indicate the existence of interactions between spins.

Different types of 2-D experiments are distinguished by the nature of the interactions that are probed during the mixing period. In a correlation spectroscopy (COSY) experiment, the cross peaks arise from through-bond scalar correlations, whereas in an NOE experiment they arise from through-space correlations. The basic pulse sequences for several common 2-D ^{1}H-NMR experiments applicable to biological macromolecules are given in Figure 5. It is clear from this figure that an important difference between the experiments lies in the nature of the mixing period, as is further discussed below.

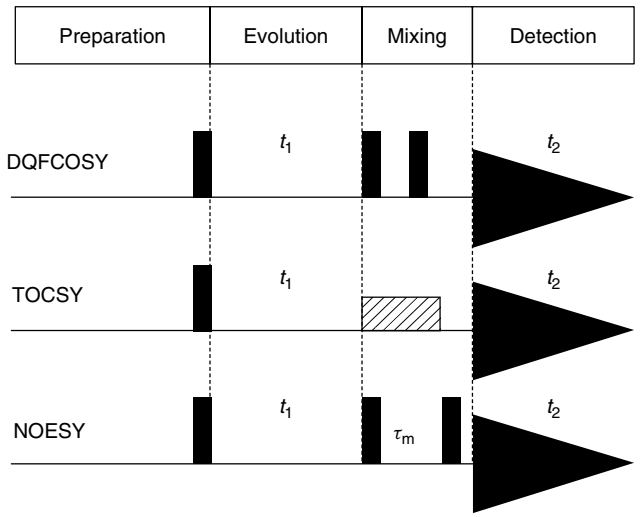

Figure 5 Basic pulse sequences of the double-quantum-filtered correlation spectroscopy (DQFCOSY), TOCSY, and NOESY experiments, illustrating that the nature of the mixing period determines the types of correlations observed in the experiment. Filled rectangles represent 90° pulses; the isotropic mixing (spin lock) pulse is represented by a cross-hatched rectangle; τ_m is the mixing time that allows development of the NOE; t_1 is the incremented delay; and t_2 is the detection period. (Adapted from King and Mackay[18].)

2.2.1 *Double-quantum-filtered Correlation Spectroscopy and Exclusive Correlation Spectroscopy*

In the double-quantum-filtered correlation spectroscopy (DQFCOSY) experiment,[19] the mixing time consists of two 90° pulses separated by a short delay to allow for the switching of RF phases between the second and third pulse in the sequence. The receiver phase and the phase of the final pulse are used to select the magnetization present as double-quantum coherence prior to the final read pulse in the experiment. The DQFCOSY experiment provides information on through-bond connectivities via scalar coupling, and is typically used to obtain spin system assignments and to measure $^3J_{HNH\alpha}$ coupling constants. This experiment has superseded the basic COSY as the experiment of choice for elucidating these couplings.

In the related exclusive correlation spectroscopy (ECOSY) experiment, phase cycling is used to select only the active couplings from the otherwise complex multiplet.[20] The main application of this experiment is to determine $^3J_{H\alpha-H\beta}$ couplings, which are used to obtain stereospecific assignments, and χ_1 restraints, which define the orientation of amino acid side chains.

2.2.2 *Total Correlation Spectroscopy*

The TOCSY (also referred to as *homonuclear Hartmann–Hahn spectroscopy*) experiment also yields scalar connectivities. In this experiment, the second pulse in the COSY experiment is replaced by a spin-locking pulse (termed the *isotropic mixing period*) that permits magnetization transfer from one proton to another at a rate determined by their J coupling.[21,22] The high spin-locking field strength (10 kHz) results in an efficient magnetization transfer during the mixing period, and the extent to which magnetization is propagated along a spin system depends on the duration of the spin-locking pulse (typically 30–120 ms). This experiment allows the mutual correlation of all protons within a spin system. This information is very valuable in the assignment of the spectra of proteins, for example, because different amino acids have different numbers of protons at different frequencies (Table 2). As a consequence, cross peaks in TOCSY spectra yield patterns characteristic of individual amino acid types.

2.2.3 *Nuclear Overhauser Effect Spectroscopy*

In the NOESY experiment, which is based on the NOE, two 90° pulses sandwich an extended mixing time (50–400 ms) during which the NOEs build up.[23,24] During the mixing time, magnetization of one nucleus can lead to variation in the magnetization of another, and the final read pulse detects the modulated signals. Protons mutually exchange their nuclear magnetization at a rate that is inversely proportional to the sixth power of the distance between them, and cross peaks are thus observed between protons that are at a maximum of 5–6 Å apart in space.

2.3 Two-dimensional Nuclear Magnetic Resonance of Biomolecules

Although there are myriad possible 2-D spectra, the assignment of peaks in NMR spectra of peptides and small proteins is usually accomplished with just a handful of experiments; in particular, TOCSY and NOESY spectra are the most important. As mentioned above, the TOCSY spectrum is useful for classifying spin systems into amino acid type, with typically the most useful region being the ''skewers'' emanating from individual NH shifts, as illustrated in Figure 6, again using MiAMP1 as an example. For each NH proton in the peptide, a series of cross peaks to the α, β, γ, etc., protons is observed and these patterns define the spin system as belonging to a particular type of amino acid.

In the early days of structure determination of proteins by NMR, the task of classifying peaks into amino acid type was done mainly using DQFCOSY spectra because limitations on hardware at that time made TOCSY spectra technically more difficult to record. Although these hardware limitations have been overcome in modern

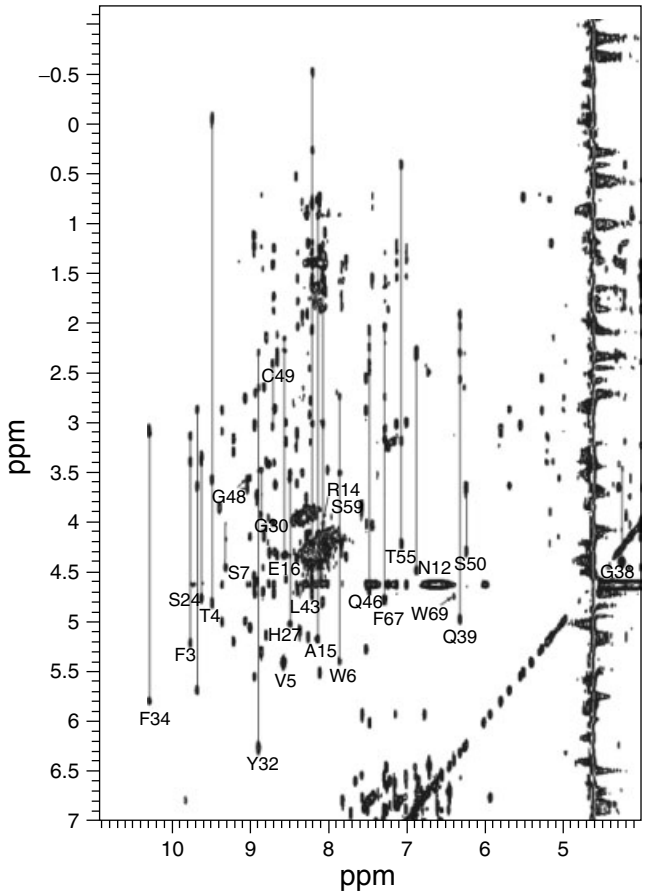

Figure 6 Region of the 2-D TOCSY spectrum of MiAMP1 showing selected amino acid assignments.[12]

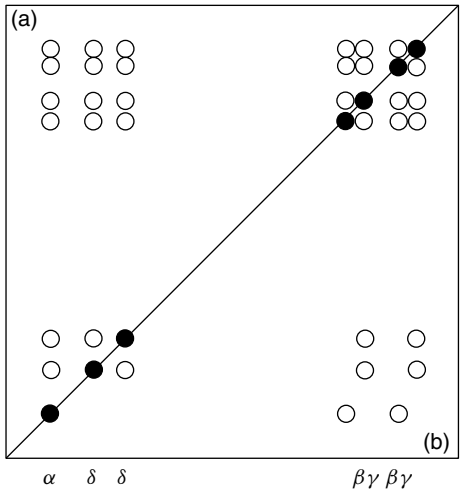

Figure 7 Comparison of TOCSY (a) and DQFCOSY (b) spectra showing connections between protons in a proline spin system. In the TOCSY spectrum, correlations to α, β, γ, and δ protons occur at the frequencies of each of these signals, making it impossible to distinguish between the α shifts and δ shifts if they are close. On other hand, in the DQFCOSY spectrum cross peaks are seen between α and β protons, between β and γ protons, and between γ and δ protons, but not between protons separated by more than three bonds.

instruments, and TOCSY spectra have principally taken over this role, DQFCOSY experiments are still useful in resolving ambiguities in TOCSY spectra, particularly for the spin systems of longer amino acids and for proline, as illustrated in Figure 7.

Once a reasonable number of amino acid spin systems have been assigned to amino acid type from the TOCSY and DQFCOSY spectra, the NOESY spectrum is inspected to locate adjacent amino acid spin systems. This is done principally by identifying a cross peak between the α proton of one residue (i) and the NH of the following residue $(i+1)$, often denoted as $d\alpha N(i, i+1)$, although support for the assignment is often also observed in $d\beta N(i, i+1)$ and $dNN(i, i+1)$ correlations. If the dipeptide pair of adjacent residues identified in this way is unique in the sequence, then the sequence-specific assignment of these residues is complete, and the assignment process continues in a similar manner for the remaining residues. In longer proteins unique stretches of three, four or more residues may need to be identified before a starting point for the sequence-specific assignment is established. The process is complicated by the fact that in a folded protein it is possible that any given α proton may be within 5 Å of more than one amide proton and thus have multiple $d\alpha N(i, j)$ NOEs. Thus, at the early stages of an assignment it is impossible to be certain whether a particular cross peak is a sequential or longer range cross peak, but as the assignment procedure progresses ambiguities get resolved. The assignment process is generally highly convergent in that once a series of correct assignments is made the number of choices for remaining cross peaks diminishes, in principle making their assignment easier.

Once all peaks in the 2-D NOESY spectrum have been assigned, their intensities are used to derive a series of interproton distance restraints. These are then used in a simulated annealing algorithm, together with dihedral angle restraints derived from DQFCOSY and ECOSY spectra to derive a family of 3-D structures consistent with the input restraints.

2.4 Limitations of Two-dimensional Nuclear Magnetic Resonance Methods

As indicated in the previous section, 2-D methods have the advantage of reducing overlap in 1-D spectra and providing additional information in terms of correlations and connectivities between peaks. However, the larger a biological macromolecule becomes, the more protons it contains and hence the higher the probability of overlap. In addition, line width increases with increasing molecular

weight, further magnifying problems of overlap. The net result from both of these factors is that for proteins containing more than approximately 100 amino acids it becomes exceedingly difficult and time consuming to assign spectra and determine 3-D structures from 2-D spectra. Higher dimensional spectra associated with ^{15}N and/or ^{13}C, ^{15}N-labeled molecules are needed.

3 THREE- AND FOUR-DIMENSIONAL NUCLEAR MAGNETIC RESONANCE

3.1 More Dimensions

Higher-dimensionality experiments can be regarded simply as linear combinations of successive 2-D experiments. For example, a 3-D pulse sequence can be constructed from two 2-D experiments as shown in Figure 8. In principle, it is possible to produce combinations of almost any 2-D experiments. This has led to an explosion in the number of techniques available, using both homonuclear and heteronuclear methods. In addition to extending the capacity of NMR for sequential assignment and structure determination as described above, these new techniques have provided an entirely different route to the assignment of protein spectra and have extended the range of NMR spectroscopy to proteins beyond the ~30-kDa size range.

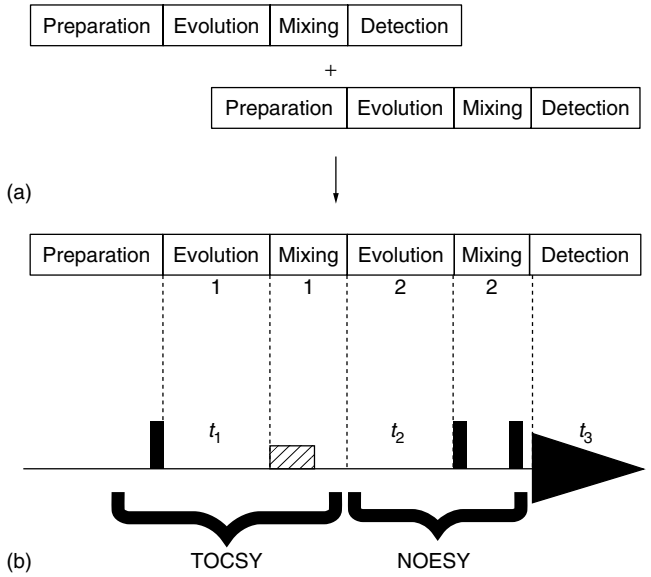

Figure 8 (a) Schematic representation of a generic 3-D pulse sequence that can be obtained by combination of two 2-D pulse sequences, exemplified in (b), where 2-D NOESY and TOCSY experiments have been combined to produce the 3-D NOESY/TOCSY.

3.2 Homonuclear Methods

Homonuclear 3-D NMR experiments were amongst the first to be developed.[25] The major advantage compared to their 2-D counterparts is that they offer the possibility to circumvent some of the problems associated with chemical shift degeneracy. Furthermore, by combining a TOCSY (or COSY), which provides scalar correlations, with a NOESY, which gives information on the distance between nuclei, it is possible, in principle at least, to obtain all the information necessary for complete assignment of the protein molecule.

There are a number of problems associated with homonuclear 3-D experiments. Although they offer the potential for improved resolution relative to the corresponding 2-D experiments, they present a number of practical difficulties. The first is in the acquisition of the data. These experiments require the incrementation of two successive delays to produce the 3-D data matrix. As a result of the relatively poor chemical shift dispersion observed for ^1H nuclei, it is necessary to record a large number of increments to obtain satisfactory digital resolution. Hence, the acquisition times for the experiments can become prohibitively long. Furthermore, because the 3-D experiment contains all the information from the two constituent 2-D experiments, there is a corresponding increase in the number of cross peaks observed, which increases the requirements for good digital resolution. Another potential problem lies in the interpretation of these experiments. The calculation of protein structures from NMR data is critically dependent on the ability to measure a large number of NOEs as accurately as possible. For homonuclear 3-D experiments, the peak volumes are the product of two transfer coefficients, which can complicate the process of extracting accurate distance information for structure calculation. Finally, as is the case with their 2-D counterparts, these experiments are based upon the observation of relatively small ^3J scalar couplings. Consequently, they only provide useful information for proteins <10 kDa. As a result, homonuclear 3-D methods are less widely used than double and triple resonance experiments that involve the use of uniformly ^{15}N and/or ^{13}C-labeled proteins.

3.3 Heteronuclear Methods

The development of recombinant DNA technology has made the production of isotopically labeled proteins relatively straightforward. Thus, it is possible to produce a variety of different protein samples with ^{15}N and/or ^{13}C labeling. Both of these nuclei have a spin quantum number $I = 1/2$, which makes them suitable for high-resolution NMR. A variety of labeling strategies are available, ranging from uniform enrichment of either

^{15}N only or both ^{15}N and ^{13}C to selective incorporation of isotope labels for particular amino acids. A major advantage of isotope labeling is that it allows the exploitation of the relatively large short-range scalar couplings that exist between both the heteronuclei and their directly attached protons (Scheme 1). These large couplings result in magnetization transfer between the heteronuclei being very much more efficient than for homonuclear experiments where $^{3}J \approx 3$–14 Hz.

The ability to transfer magnetization between heteronuclei is the essential foundation upon which many of the multidimensional methods for resonance assignment and structure determination are based. Although a full mechanistic description[3] of the myriad sophisticated pulse sequences available is beyond the scope of this article, the general principles of heteronuclear magnetization transfer can be outlined in the relatively simple example of an ^{1}H–^{15}N-heteronuclear single-quantum coherence (HSQC) experiment, which correlates the chemical shifts of the ^{15}N nuclei with those of their attached protons.[26] A typical pulse sequence for the HSQC experiment is shown in Figure 9(a). The first step of the experiment involves the transfer of magnetization from the ^{1}H nuclei to ^{15}N using insensitive nuclei enhancement by polarization transfer (INEPT).[27] There follows an incremented delay, during which the magnetization is labeled with the ^{15}N chemical shift. Finally, the magnetization is transferred back to the ^{1}H nuclei, using a reverse-INEPT sequence, and detected. The experiment has very high sensitivity, and Fourier transformation of the data produces a 2-D spectrum with a single in-phase cross peak for each H–N correlation. A typical ^{1}H–^{15}N HSQC spectrum is shown in Figure 9(b).

As is the case with homonuclear 2-D experiments, the HSQC experiment may be regarded as a building block that can be incorporated into more complex multidimensional pulse sequences. Similar heteronuclear magnetization transfer schemes have been described that correlate the chemical shifts of ^{1}H, ^{13}C, and ^{15}N nuclei to facilitate resonance assignment.

Heteronuclear experiments place greater demands upon the spectrometer hardware than homonuclear spectroscopy. Certain experiments require the generation of suitable pulses for excitation of ^{1}H, ^{15}N, and ^{13}C nuclei. For the purposes of the majority of these experiments, aliphatic ^{13}C nuclei are treated separately

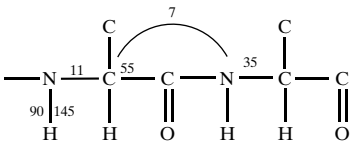

Scheme 1. Heteronuclear coupling constants (in hertz) found in proteins.

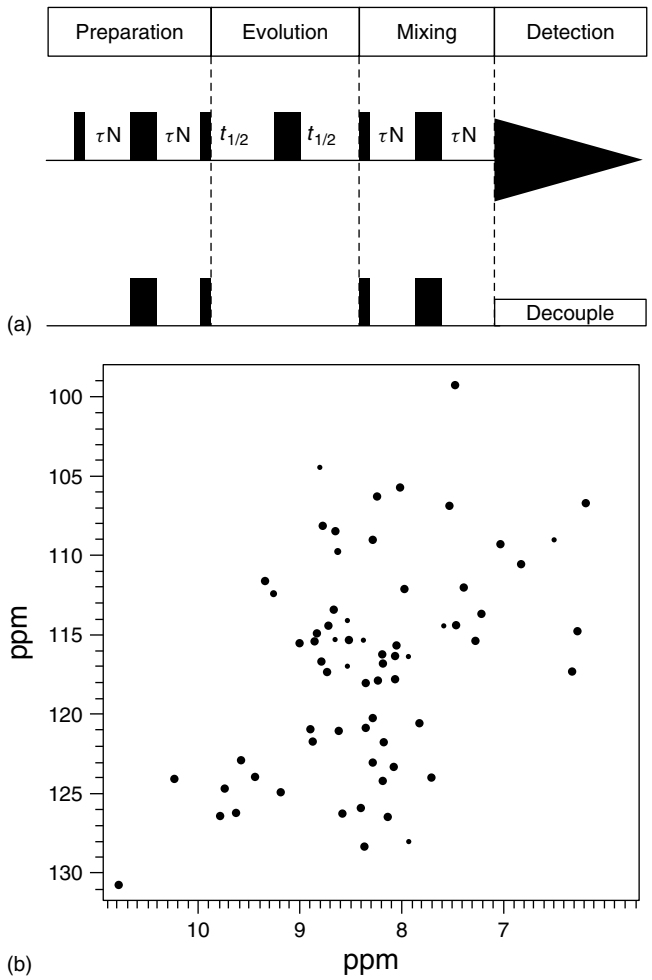

Figure 9 (a) Pulse sequence for the 2-D HSQC experiment. (b) The HSQC spectrum of MiAMP1.[12]

from carbonyl ^{13}C nuclei. Consequently, the spectrometer must possess at least three separate amplifiers (to excite the three separate nuclei), and it must be capable of synthesizing frequency-selective pulses in order to selectively excite the appropriate regions of the ^{13}C spectrum. Many heteronuclear experiments also use PFGs for the suppression (or selection) of coherences, which requires that the spectrometer have the capacity to produce inhomogeneous RF pulses. Furthermore, in cases where samples have been deuterated, a fourth frequency is introduced and, consequently, an additional amplifier is required. Key developments for multidimensional heteronuclear spectroscopy have included the production of higher magnetic fields, which improves resolution and has important advantages in terms of sensitivity, while improvements in probe technology, notably in the production of shielded-PFG probes, have seen a dramatic improvement in spectral quality.

3.4 Heteronuclear-edited Nuclear Magnetic Resonance

The simplest of the heteronuclear 3-D experiments involve the use of the heteronucleus to edit a standard homonuclear experiment. Typically, an HSQC experiment is combined with either a homonuclear NOESY or TOCSY. In general, these experiments are carried out using ^{15}N as the heteronucleus owing both to its favorable spectral properties and the relatively cheaper cost of producing uniformly ^{15}N-labeled proteins.

The ^{1}H–^{15}N total correlation spectroscopy/heteronuclear single-quantum coherence (TOCSY/HSQC) separates amide protons according to the ^{15}N chemical shifts of the attached nitrogen, as is the case with the HSQC spectrum shown in Figure 9(b).[28] In the third dimension there are cross peaks at the frequencies of all proton resonances that have scalar correlations to the amide proton. Thus the 3-D ^{1}H–^{15}N TOCSY/HSQC spectrum consists of ''towers'' of cross peaks built upon ^{1}H–^{15}N correlations observed in the HSQC experiment containing all of the scalar correlations observed in the amide proton region of a conventional 2-D ^{1}H–^{1}H TOCSY. These spectra are usually displayed as 2-D strips centered at the ^{15}N chemical shift that contain all of the TOCSY cross peaks for each amide proton, as shown schematically in Figure 10.

The related 3-D ^{1}H–^{15}N NOESY/HSQC spectrum also contains ''towers'' of cross peaks built upon ^{1}H–^{15}N correlations, but in this case the third dimension contains cross peaks at the frequency of all proton resonances that have dipolar interactions to the amide proton. Using a combination of the 3-D ^{1}H–^{15}N TOCSY/HSQC and 3-D ^{1}H–^{15}N NOESY/HSQC, it is possible to obtain sequential assignments using essentially the same strategy described above for homonuclear 2-D experiments. The advantage of increasing the dimensionality lies in the greater resolution obtained, as illustrated in Figure 10.

Similar experiments can be performed on suitably labeled samples using ^{13}C as the heteronucleus. Thus the 3-D ^{1}H–^{13}C NOESY/HSQC provides a method for resolving NOEs to protons based on the chemical shift of the attached carbon. In addition to the increased cost of incorporating ^{13}C labels compared to the incorporation

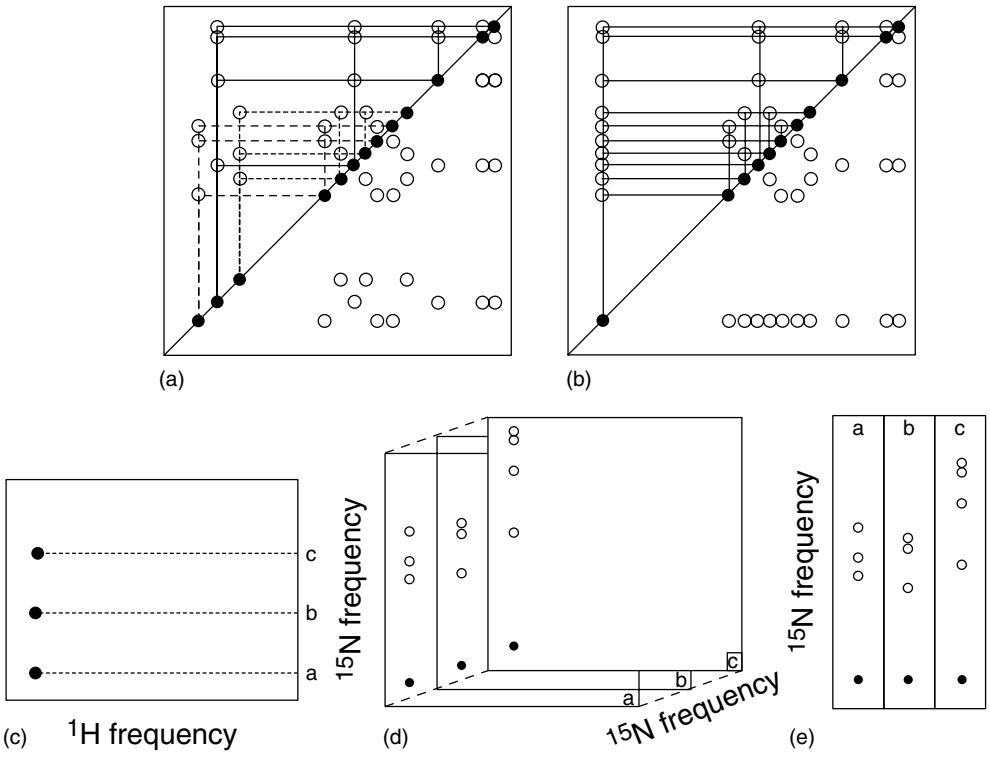

Figure 10 Schematic representation of the improved resolution afforded by 3-D heteronuclear-edited NMR. The homonuclear 2-D spectrum in (a) contains cross peaks between scalar coupled protons. For larger proteins, lack of signal dispersion often results in resonances becoming overlapped with one another, as depicted in (b). In this case it is no longer possible to distinguish between the three amide protons because they all resonate at the same frequency. However, in many cases the amide proton signals can be separated in an HSQC spectrum according to the resonance frequency of their attached nitrogen (a, b, or c), shown in (c). In this way, overlapping signals in a homonuclear 2-D TOCSY experiment can be resolved in the 3-D ^{15}N–^{1}H TOCSY/HSQC. Correlations to the side-chain protons are observed in the 3-D spectrum at the resonance frequency of the amide nitrogen (d). Such spectra are typically viewed as narrow 2-D strips taken at the ^{15}N amide nitrogen frequency (e).

of ^{15}N, there are several reasons why the 3-D ^{1}H–^{15}N NOESY/HSQC is a more widely used experiment. First, the transverse relaxation rates of ^{13}C nuclei can be relatively fast, which results in a significant loss of signal. Secondly, both ^{13}C and ^{1}H chemical shifts are closely correlated with secondary structure; thus, it is often the case that the regions of greatest overlap coincide in the two frequency dimensions, which limits the potential gains in resolution. Because there is no correlation between ^{15}N chemical shift and secondary structure, there is often a correspondingly greater increase in resolution.

For double-labeled proteins, it is also possible to record NOESY spectra that are edited with respect to both ^{13}C and ^{15}N. Such an experiment would be a 4-D ^{13}C–^{15}N HSQC/NOESY/HSQC. This experiment further separates dipolar correlations observed in the 3-D ^{1}H–^{15}N NOESY/HSQC on the basis of the ^{13}C chemical shifts of the carbon nucleus to which any carbon-bound protons are attached. Effectively, a series of 3-D ^{1}H–^{15}N NOESY/HSQC data sets are recorded at separate ^{13}C frequencies. This type of experiment is particularly useful for resolving NOEs between methyl groups in the protein core, where the dispersion is often very poor.

The obvious limitation of the heteronuclear-edited approach to obtaining resonance assignments is that the 3-D ^{1}H–^{15}N TOCSY/HSQC experiment is based on magnetization transfer via relatively small ^{3}J scalar couplings. Therefore, as is the case with homonuclear 2-D and 3-D experiments, this approach is only successful with proteins <10 kDa. Although 3-D and 4-D NOESY/HSQC experiments can be recorded successfully for very much larger proteins, analysis of these data is crucially dependent upon the ability to obtain sequence-specific assignment of the resonances.

3.5 Triple-resonance Nuclear Magnetic Resonance

An alternative approach to the sequential assignment problem involves utilizing the one- and two-bond couplings that exist between the heteronuclei and their directly attached protons. These heteronuclear couplings are large in comparison to homonuclear $^{3}J_{HH}$ couplings. As a result, magnetization transfer between the heteronuclei is very much more efficient than homonuclear scalar transfer. Many experiments that allow magnetization to be transferred between nuclei via scalar couplings have been devised. These multidimensional experiments correlate the chemical shifts of different nuclei, depending on the specific pulse sequence. One such pulse sequence, that of the HNCO experiment,[29] is illustrated in Figure 11. In terms of sensitivity, it is desirable to have the magnetization originate on a ^{1}H nucleus and to detect ^{1}H signals in

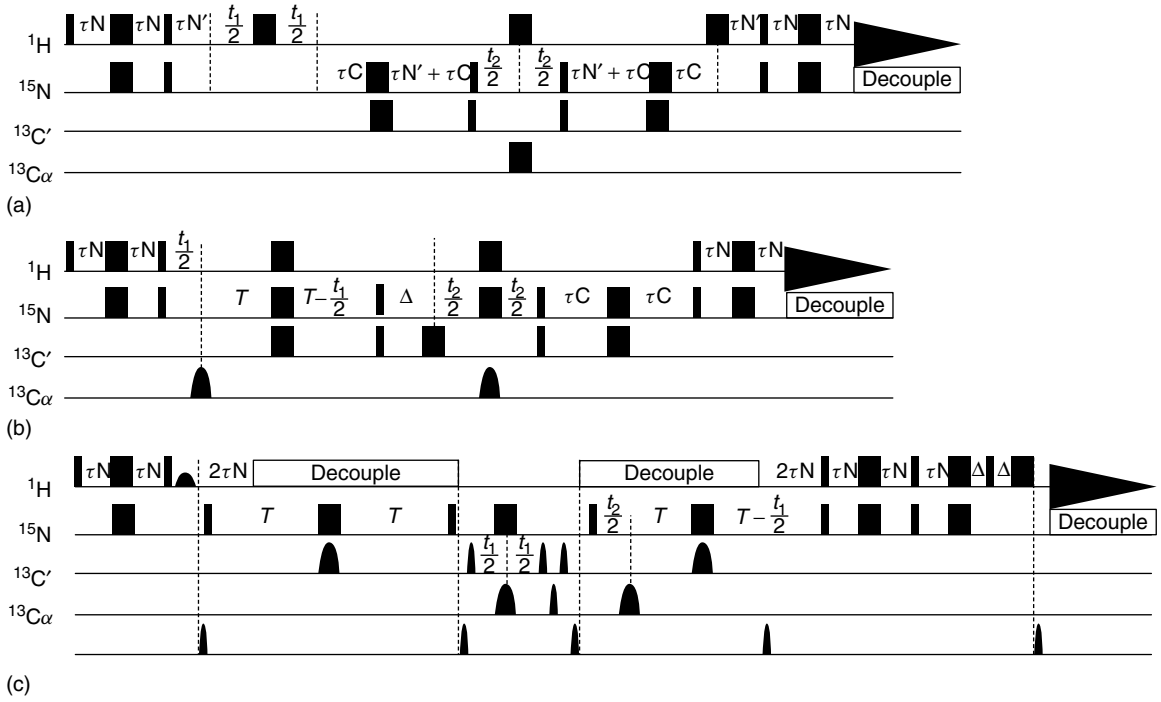

Figure 11 Three pulse sequences for the triple-resonance 3-D HNCO experiment. (a) A standard HNCO sequence. (b) A constant-time HNCO. (c) A constant-time HNCO sequence using PFGs incorporating minimal saturation of the water and sensitivity enhancement. In all cases, narrow and wide pulses correspond to 90° and 180° pulses, respectively, and nonrectangular blocks indicate selective pulses. Periods during which broadband decoupling is applied are indicated.

the direct dimension. In the case of HNCO, there is only one proton involved in the transfer pathway, and it is therefore necessary to adopt a so-called out-and-back approach. Thus ^{1}H magnetization is transferred to ^{15}N, then on to the ^{13}C′, and back to the ^{1}H via the ^{15}N. In general, triple-resonance experiments are undertaken on uniformly labeled samples, which makes it necessary for reasons of resolution and sensitivity to refocus the ^{13}C′–^{13}Cα interaction during the time that the ^{13}C′ coherence is evolving. Owing to the large chemical shift difference that exists between the ^{13}C′ and ^{13}Cα resonances, this is easily achieved through the application of appropriate selective 180° pulses at the ^{13}Cα resonance frequency.

By convention, triple-resonance experiments are named on the basis of the nuclei that are involved in the magnetization transfer pathway. Hence, the HNCO experiment correlates the amide proton $H^{N}(i)$, the amide nitrogen $N(i)$, and the carbonyl carbon of the preceding residue $C'(i-1)$. Nuclei that are involved in the transfer pathway but not frequency labeled are given in parentheses. Thus, the HN(CO)CA experiment correlates the amide proton $H^{N}(i)$, the amide nitrogen $N(i)$, and the Cα carbon of the preceding residue $C\alpha(i-1)$ via the transfer pathway $H^{N}(i) \Rightarrow N(i) \Rightarrow C'(i-1) \Rightarrow C\alpha(i-1) \Rightarrow C'(i-1) \Rightarrow N(i) \Rightarrow H^{N}(i)$, where the C′ is not frequency labeled. In the names of the pulse sequences, HN indicates that both the amide proton and amide nitrogen are frequency labeled, whereas CO, CA, CB, HA, and HB refer to C′, Cα, Cβ, Hα, and Hβ, respectively. A comprehensive description of all triple-resonance pulse sequences is beyond the scope of this article, although several reviews and monographs that describe many of the commonly used techniques are available.[3,29,30] A major advantage of these heteronuclear experiments is that they allow a number of interresidue correlations to be made that provide a direct and unambiguous route to obtaining sequence-specific assignments.

As the number of triple-resonance experiments has multiplied, so have the available strategies for obtaining sequence-specific assignments. The precise strategy employed can be tailored to the molecule being studied, but in general a complementary pair of experiments is recorded, one of which contains only interresidue correlations and the other contains both intra- and interresidue correlations. This general strategy is possible because, as shown in Scheme 1, the amide nitrogen $N(i)$ is coupled to the Cα carbons of both the same residue $C\alpha(i)$ and the preceding residue $C\alpha(i-1)$. Thus, when magnetization is transferred from N to Cα, two pathways that are modulated by the $^{1}J_{C\alpha N}$ and $^{2}J_{C\alpha N}$ coupling constants, respectively, are available. It is possible to select for the interresidue pathway alone by transferring

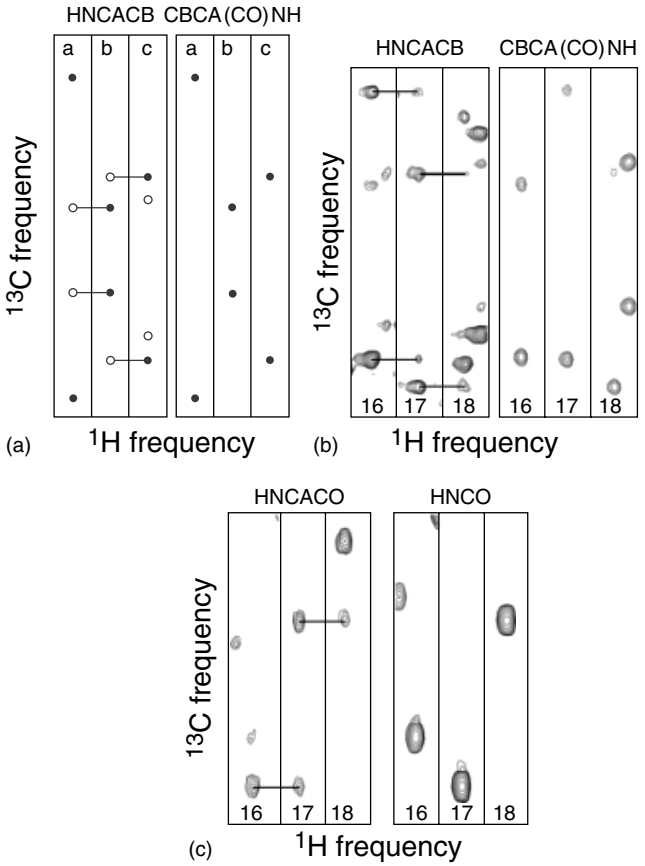

Figure 12 An illustration of the process of sequential assignment from triple-resonance data. (a) Schematic representation of the HNCACB experiment that correlates the amide proton and nitrogen with the intraresidue Cα and Cβ carbon resonances as well as those of the preceding residue. The complementary CBCA(CO)NH spectrum contains correlations to the preceding residue only. As is the case with heteronuclear-edited spectra, these data are typically viewed as narrow 2-D strips taken at the ^{15}N resonance frequency (a, b or c). Comparison of strips taken at the same resonance frequency in the two spectra enables sequential assignments to be obtained. (b) Strips from the HNCACB and CBCA(CO)NH experiments recorded at 600 MHz and 320 K for *Vibrio cholerae* DsbA. (c) Complementary strips from the HNCO and HN(CA)CO experiments on the same sample.

magnetization from $N(i)$ to $C\alpha(i-1)$ (and conversely from $C\alpha(i-1)$ to N) via the carbonyl carbon $C'(i-1)$.

Two such complementary experiments are the HNCA and the HN(CO)CA experiments. The HNCA experiment correlates the amide proton and ^{15}N resonance with the intraresidue ^{13}Cα via the $^{1}J_{C\alpha N}$ coupling constant and with the ^{13}Cα of the preceding residue via the $^{2}J_{C\alpha N}$ coupling constant, giving rise to a spectrum with two ^{13}Cα cross peaks for each ^{15}N resonance, as shown schematically in Figure 12(a). The HN(CO)CA experiment provides correlations exclusively to the ^{13}Cα

resonance of the preceding residue by relaying magnetization via the intervening ^{13}C carbonyl. In this case, only a single $^{13}C\alpha$ cross peak is observed for each ^{15}N resonance. With a combination of these two experiments, it is possible, at least in theory, to distinguish the intraresidue and sequential cross peaks and to obtain sequential assignments for each of the backbone nuclei. In practice, however, problems with overlapping or missing signals in the spectrum often prevent full assignments being obtained. It is common to begin by recording some of the more sensitive and straightforward out-and-back experiments described above, such as HNCO, HNCA, and HN(CO)CA. Beyond that point the given strategy is often determined by the sample under investigation. In practice, several triple-resonance spectra are usually required, the selection of which is determined by problems encountered during the assignment process.

An alternative approach involves using the HNCACB and the CBCA(CO)NH pair. In these experiments the amide proton and ^{15}N resonance are correlated with the $^{13}C\alpha$ and $^{13}C\beta$ spins of both the same and the preceding residue, as shown in Figure 12(b). Although the HNCACB experiment may contain up to four ^{13}C resonances in each ^{15}N strip, analysis is simplified because the $^{13}C\alpha$ and $^{13}C\beta$ cross peaks are of opposite sign. Furthermore, with knowledge of both $^{13}C\alpha$ and $^{13}C\beta$ assignments, it is possible, in many cases, to unequivocally assign the residue type, which simplifies the subsequent sequence-specific assignment.

Once the backbone resonances have been assigned, it is necessary to record experiments that correlate the side-chain nuclei. Two of the most common experiments for doing so are the HCCH-COSY and the HCCH-TOCSY. These experiments involve transfer of proton magnetization to a directly attached ^{13}C spin, followed by COSY- or TOCSY-type transfer along the carbon backbone. Finally, the magnetization is transferred back to the protons where it is detected. These experiments provide essentially the same information as homonuclear 2-D COSY/TOCSY experiments, but have the advantage that magnetization transfer is modulated by large one-bond couplings, making them useful for the assignment of side-chain resonances in proteins of up to 30 kDa. Furthermore, the addition of the third dimension correlating ^{13}C resonances to their directly attached protons dramatically improves resolution relative to the 2-D experiments. A schematic representation of the data obtained from a 3-D HCCH-TOCSY is presented in Figure 13.

Some of the strategies for obtaining sequence-specific assignments using a pair of triple-resonance experiments are depicted in Figure 14. These experiments have been developed to combat the problems of extensive overlap and large line width that prevent sequence-specific

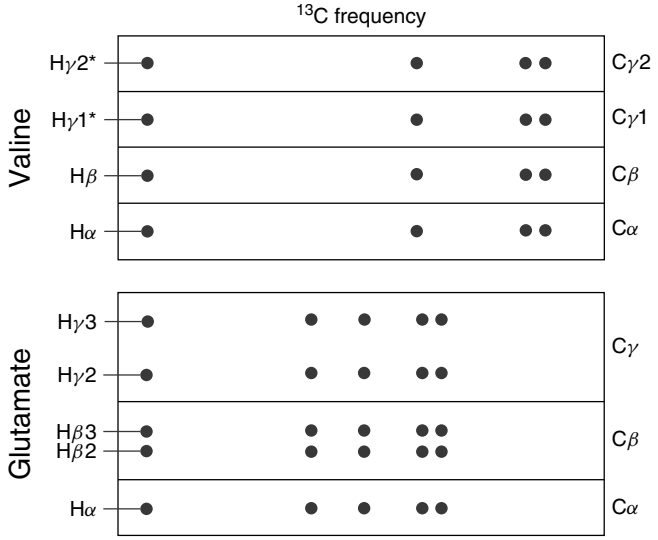

Figure 13 Schematic representation of the data observed in a 3-D HCCH-TOCSY spectrum. In this case 2-D strips are taken at the ^{13}C frequency of each carbon nucleus in the spin system. In principle, each strip contains cross peaks at all of the proton resonances for the given residue. In practice, some cross peaks may be missing or at the same frequency.

assignments being obtained by homonuclear methods. In general, the assignments are only a starting point in the process of macromolecular structure determination, which requires the unambiguous assignment of through-space NOE interactions. These are obtained using the 3-D NOESY/HSQC and 4-D HSQC/NOESY/HSQC experiments described above.

One of the major problems faced by triple-resonance experiments relates to transverse relaxation, T_2. Many triple resonance experiments consist of multiple steps wherein magnetization is transferred between different nuclei via a series of INEPT steps. The signal associated with each nucleus loses intensity over time as a result of a loss of phase coherence, at a rate determined by the T_2 value for that nucleus. In the design of triple-resonance experiments, therefore, it is important to minimize the amount of time spent on nuclei that have short T_2 values, such as $C\alpha$. As shown in Figure 9(a), the INEPT sequence that is used to transfer magnetization between nuclei involves a delay of $\frac{1}{2}J$ between certain pulses. Thus the duration of the experiment is determined, at least in part, by the magnitude of the appropriate coupling constant for the desired magnetization transfer pathway. An obvious way to increase the sensitivity of any experiment is to reduce the total duration of the pulse sequence, so that less magnetization is lost to T_2 relaxation before the signal is acquired. The first triple-resonance experiments, such as the HNCO pulse sequence shown in Figure 11(a), consisted of a magnetization transfer step, followed by a period for frequency labeling. More recently, it has been

Experiment		Couplings used	References
Three-dimensional			
HNCO	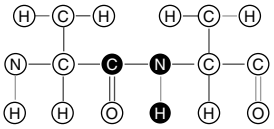	$^1J_{NH}$ $^1J_{NC'}$	(27, 29)
HN(CA)CO	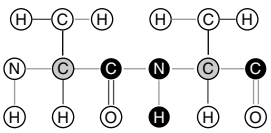	$^1J_{NH}$ $^1J_{NC\alpha}$ $^2J_{NC\alpha}$ $^1J_{C\alpha C'}$	(30)
HN(CO)CA	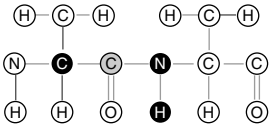	$^1J_{NH}$ $^1J_{NC'}$ $^1J_{C\alpha C'}$	(29, 31)
HNCA	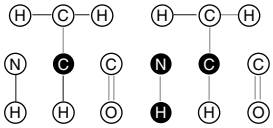	$^1J_{NH}$ $^1J_{NC\alpha}$ $^2J_{NC\alpha}$	(27, 29, 32)
CBCA(CO)NH	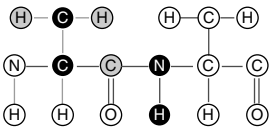	$^1J_{CH}$ $^1J_{C\alpha C\beta}$ $^1J_{C\alpha C'}$ $^1J_{NC'}$ $^1J_{NH}$	(33)
HNCACB	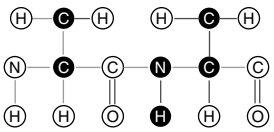	$^1J_{NH}$ $^1J_{NC\alpha}$ $^2J_{NC\alpha}$ $^1J_{C\alpha C\beta}$	(34)
Four-dimensional			
HCA(CO)NH	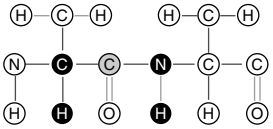	$^1J_{CH}$ $^1J_{C\alpha C'}$ $^1J_{NC'}$ $^1J_{NH}$	(35)
HCANH	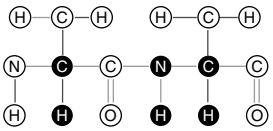	$^1J_{CH}$ $^1J_{NC\alpha}$ $^2J_{NC\alpha}$ $^1J_{NC'}$ $^1J_{NH}$	(36)

Figure 14 Selected triple-resonance experiments. Spectra contain correlations between the nuclei in black circles. Nuclei involved in the transfer pathway that are not detected in the experiment are shaded gray. The heteronuclear couplings that are employed to mediate the magnetization transfer are indicated.

shown that these steps can occur simultaneously, which results in much shorter recording times of the so-called constant-time experiments (Figure 11(b)). It is not always the case that modifications that lead to improvements in sensitivity result from a shortening in the duration of the pulse sequence. A more recent HNCO pulse sequence is shown in Figure 11(c). In addition to utilizing PFGs, this sequence employs minimal saturation of the solvent resonance and sensitivity enhancement.[31]

Most triple resonance pulse sequences are now acquired with the use of PFGs.[32,33] This involves the use of inhomogeneous RF pulses for either the suppression or selection of desired coherences. Prior to the use of PFGs, triple-resonance experiments relied on phase cycling to remove unwanted magnetization. This resulted in the experiments having long measuring times and relied crucially on the spectrometer being highly stable for the duration of the experiments. Using PFGs for

the suppression of unwanted coherences eliminates the need for much of the phase cycling so that in many cases it is now possible to employ two-step phase cycles for many triple-resonance experiments. This means that the measuring time for an experiment is now usually determined by considerations of signal-to-noise for any given sample and experiment and not by the requirements of the phase cycle. An additional advantage of PFGs is that they eliminate unwanted signals in a nonsubtractive manner and provide a straightforward mechanism for cleanly eliminating the solvent signal.

A major driving force in the development of triple-resonance NMR spectroscopy was the necessity to extend the size of macromolecules that were accessible to structure determination. However, larger molecules generally have shorter T_2 values, and hence give broader lines. There are two principal contributions to the line broadening observed for larger molecules. The first is due to the fact that for larger molecules a greater proportion of the sample is tumbling at the zero-quantum frequency where it can induce spin exchange and thereby contribute to dipole–dipole relaxation. Secondly, the chemical shift of a particular nucleus is dependent upon its orientation in the magnetic field. When a molecule tumbles rapidly in solution, the chemical shifts of all the different environments are averaged to give an ''isotropic'' chemical shift. The amount of variation in chemical shift with orientation is referred to as the *chemical shift anisotropy* (CSA). As the molecule tumbles more slowly, the oscillating magnetic field associated with the CSA can also induce transitions that lead to an increase in the rate of relaxation.

Several methods that attempt to minimize signal loss due to T_2 relaxation are available. Incorporation of deuterons into the protein is being widely used to extend the applicability of multidimensional NMR to larger systems. A deuterium nucleus has a much lower magnetogyric ratio than the proton and therefore it contributes far less to dipole–dipole relaxation. Even partial deuteration of a protein sample results in significant improvements in T_2. Different strategies for protein deuteration have been proposed, involving either partial or full deuteration, but the overall effect is a drastic reduction in the dipolar broadening of resonances, owing to the removal of a large number of protons. More recently, further sensitivity gains have been accomplished by the development of transverse relaxation-optimized spectroscopy (TROSY).[34] This approach utilizes the finding that at a particular field strength the CSA and dipole–dipole relaxation can be canceled out. This generates one component of the observable signal in the HSQC spectrum that essentially has very little T_2 relaxation. Using the combined approach of protein perdeuteration and TROSY techniques, spectra have been recorded on protein complexes in the range of 1 MDa.[35]

In addition to increasing the molecular-weight range of molecules that are amenable to structure determination using NMR, triple-resonance experiments allow the measurement of a number of parameters that may be used directly in the structure calculation and refinement. The most obvious are the chemical shifts of the heteronuclei that are determined. Chemical shifts have been used widely to identify elements of secondary structure in proteins.[36,37] More recently, ^{1}H and ^{13}C chemical shifts have been included as refinement terms in protein structure calculations.[38,39] In addition, the incorporation of heteronuclei allows the measurement of many more scalar couplings than that possible in homonuclear spectra. Heteronuclear coupling constants provide torsion angle information that may also be used in structure calculations.[40,41]

Both chemical shifts and coupling constants, like the NOE, contain information only about the local conformation of a macromolecule. Two additional parameters that may be obtained for isotopically labeled samples provide long-range information that can be used to calculate the relative orientation of separate domains even in cases where few short-range interactions exist. The first of these involves the measurement of residual dipolar couplings.[42] In cases where macromolecules tumble anisotropically, the dipolar interaction between pairs of spins, which is related to the angle between the vector joining the spins and the externally applied magnetic field, is not averaged to zero. This results in the observation of small splittings in the spectra of dipolar coupled nuclei. The size of the observed splittings is related to several factors, including the strength of the applied magnetic field and the degree of anisotropy. In practice, dipolar couplings are often measured under conditions that induce alignment (and hence anisotropy), such as solutions of magnetically aligned bicelles or filamentous bacteriophages. The relationship between the size of the splitting and the angular orientation of the dipole vector enables the relative orientations of structural elements within the protein to be established.[43] The second method relies on measurement of heteronuclear relaxation parameters T_1 and T_2. In molecules that tumble isotropically, the T_1/T_2 ratio is constant. However, molecules that tumble anisotropically display angular dependence of the T_1/T_2 ratio. Again, this allows the relative orientations of discrete domains in macromolecules to be determined.[44]

3.6 Limitations and Future Developments

Recent developments in multidimensional heteronuclear experiments, as well as improved strategies for isotope

labeling,[45] have extended the scope of macromolecular NMR to larger biomolecules, and structural studies on proteins with molecular weights of up to 30–40 kDa are becoming routine. Furthermore, additional information available from heteronuclear spectra has improved the precision and accuracy with which macromolecular structures can now be calculated from NMR data. There are also some examples of very much larger systems where structural information has been obtained.[35,46] In addition to the structural information, NMR spectroscopy continues to provide a unique method for the study of protein dynamics. There are now several examples where NMR has been used in conjunction with other techniques to provide clearer insights into motions over a range of timescales, and it is becoming increasingly clear that flexibility plays a key role in the function of many proteins.[47–49]

An emerging area is the application of NMR spectroscopy to the study of intrinsically unstructured proteins.[50] Typically, these systems have very poor signal dispersion, which makes the process of obtaining sequential assignments more challenging. In principle, the problem of overlap can be reduced by increasing the dimensionality of an experiment, but this comes at the cost of increased requirements on spectrometer recording time. Recently, a number of schemes have been developed to address this problem.[51,52] These projection–reconstruction methods allow for rapid acquisition of high-dimensionality experiments, which would otherwise have prohibitively long acquisition times. Recently, a six-dimensional experiment acquired using automated projection spectroscopy (APSY) that allows full backbone assignments from a single experiment was reported.[53] The peak lists generated in such experiments are very accurate, which makes the assignment process amenable to automated analysis.[54]

Collectively, these new tools extend the scope of NMR spectroscopy to a wider range of biomolecular systems, increase throughput, reliability and, ultimately, the quality of the data that can be produced. These advances have made NMR spectroscopy an invaluable tool for the study of protein structure and protein dynamics in a number of structural biology and structural genomics programs.

ACKNOWLEDGMENTS

The work in our laboratory is supported by grants from the Australian Research Council (ARC). D.J.C. is an ARC Professorial Fellow. We thank Ailsa McManus for providing the spectra of MiAMP1.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
APSY	Automated Projection Spectroscopy
COSY	Correlation Spectroscopy
CSA	Chemical Shift Anisotropy
CW	Continuous Wave
DQFCOSY	Double-quantum-filtered Correlation Spectroscopy
DSS	2,2-Dimethyl-2-silapentane-5-sulfonate
ECOSY	Exclusive Correlation Spectroscopy
FID	Free Induction Decay
FT	Fourier Transform
HSQC	Heteronuclear Single-quantum Coherence
INEPT	Insensitive Nuclei Enhancement by Polarization Transfer
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PFG	Pulsed-field Gradient
RF	Radiofrequency
TOCSY	Total Correlation Spectroscopy
TOCSY/HSQC	Total Correlation Spectroscopy/Heteronuclear Single-quantum Coherence
TROSY	Transverse Relaxation-optimized Spectroscopy
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
4-D	Four-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance in Medicine, High Resolution Ex Vivo • Magnetic Resonance, General Medical

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Two-dimensional Nuclear Magnetic Resonance of Small Molecules

REFERENCES

1. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford, 1987.
2. G. Wider, 'Technical Aspects of NMR Spectroscopy with Biological Macromolecules and Studies of Hydration in Solution', *Prog. NMR Spectrosc.*, **32**, 193–275 (1998).
3. J. Cavanagh, W.J. Fairbrother, A.G.I. Palmer, N.J. Skelton, *Protein NMR Spectroscopy, Principles and Practice*, Academic Press, London, 1996.
4. G.C.K. Roberts, *NMR of Macromolecules, a Practical Approach*, Oxford University Press, Oxford, 1993.
5. A. Pardi, 'Multidimensional Heteronuclear NMR Experiments for Structure Determination of Isotopically Labeled RNA', *Methods Enzymol.*, **261**, 350–380 (1995).
6. S. Campbell-Burk, S. Zhong, 'Biomolecular Applications of Heteronuclear Multidimensional NMR', *Curr. Opin. Biotechnol.*, **5**, 346–354 (1994).
7. R.R. Ernst, W.A. Anderson, 'Application of Fourier Transform Spectroscopy to Magnetic Resonance', *Rev. Sci. Instrum.*, **37**, 93–102 (1966).
8. M. Piotto, V. Saudek, V. Sklenar, 'Gradient-tailored Excitation for Single-quantum NMR Spectroscopy of Aqueous Solutions', *J. Biomol. NMR*, **2**, 661–665 (1992).
9. D.S. Wishart, C.G. Bigam, J. Yao, F. Abildgaard, H.J. Dyson, E. Oldfield, J.L. Markley, B.D. Sykes, '1H, 13C and 15N Chemical Shift Referencing in Biomolecular NMR', *J. Biomol. NMR*, **6**, 135–140 (1995).
10. D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, '1H, 13C and 15N Random Coil NMR Chemical Shifts of the Common Amino Acids. I. Investigations of Nearest-neighbor Effects', *J. Biomol. NMR*, **5**, 67–81 (1995).
11. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, 1986.
12. A.M. McManus, K.J. Nielsen, J.P. Marcus, S.J. Harrison, J.L. Green, J.M. Manners, D.J. Craik, 'MiAMP1, a Novel Protein from Macadamia Integrifolia Adopts a Greek Key Beta-barrel Fold Unique Amongst Plant Antimicrobial Proteins', *J. Mol. Biol.*, **293**, 629–638 (1999).
13. S.J. Harrison, A.M. McManus, J.P. Marcus, K.C. Goulter, J.L. Green, K.J. Nielsen, D.J. Craik, D.J. Maclean, J.M. Manners, 'Purification and Characterization of a Plant Antimicrobial Peptide Expressed in Escherichia Coli', *Protein Expr. Purif.*, **15**, 171–177 (1999).
14. J. Jeener, *Ampere International Summer School*, Basko Polje, Yugoslavia, 1971.
15. W.P. Aue, E. Bartholdi, R.R. Ernst, 'Two Dimensional Spectroscopy: Application to Nuclear Magnetic Resonance', *J. Chem. Phys.*, **64**, 2229–2246 (1976).
16. W.R. Croasman, R.M.K. Carlson, *Two Dimensional NMR Spectroscopy. Applications for Chemists and Biochemists*, 2nd edition, VCH Publishers, New York, 1994.
17. A. Bax, *Two Dimensional Nuclear Magnetic Resonance in Liquids*, Delft University Press, Holland, 1982.
18. G.F. King, J.P. Mackay, 'Protein Structure Determination Using NMR Spectroscopy', in *NMR in Drug Design*, ed. D.J. Craik, CRC Press, Boca Raton, 101–200, 1996.
19. M. Rance, O.W. Sorenson, G. Bodenhausen, G. Wagner, R.R. Ernst, K. Wüthrich, 'Improved Spectral Resolution in COSY 1H NMR Spectra of Proteins via Double Quantum Filtering', *Biochem. Biophys. Res. Commun.*, **117**, 479–485 (1983).
20. C. Greisinger, O.W. Sorenson, R.R. Ernst, 'Practical Aspects of the E.COSY Technique: Measurement of Scalar Spin-Spin Coupling Constants in Peptides', *J. Magn. Reson.*, **88**, 177–185 (1987).
21. A. Bax, D.G. Davis, 'MLEV-17 Based Two Dimensional Homonuclear Magnetization Transfer Spectroscopy', *J. Magn. Reson.*, **65**, 355–360 (1985).
22. L. Braunschweiler, R.R. Ernst, 'Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy', *J. Magn. Reson.*, **53**, 521–528 (1983).
23. J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, 'Investigation of Exchange Processes by Two Dimensional NMR Spectroscopy', *J. Chem. Phys.*, **71**, 4546–4553 (1979).
24. A. Kumar, R.R. Ernst, K. Wüthrich, 'A Two-dimensional Nuclear Overhauser Enhancement (2D NOE) Experiment for the Elucidation of Complete Proton-proton Cross-relaxation Networks in Biological Macromolecules', *Biochem. Biophys. Res. Commun.*, **95**, 1–6 (1980).
25. C. Griesinger, O.W. Sorenson, R.R. Ernst, 'Three Dimensional Fourier Transform Spectroscopy. Application to High Resolution NMR', *J. Magn. Reson.*, **84**, 14–63 (1989).

26. G. Bodenhausen, D.J. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185–188 (1980).
27. G.A. Morris, R. Freeman, 'Enhancement of Nuclear Magnetic Resonance Signals by Polarisation Transfer', *J. Am. Chem. Soc*, **101**, 760–762 (1979).
28. S.W. Fesik, E.R. Zuiderweg, 'Heteronuclear Three-dimensional NMR Spectroscopy of Isotopically Labelled Biological Macromolecules', *Quart. Rev. Biophys.*, **23**, 97–131 (1990).
29. L.E. Kay, M. Ikura, R. Tschudin, A. Bax, 'Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins', *J. Magn. Reson.*, **89**, 496–514 (1990).
30. A.S. Edison, F. Abildgaard, W.M. Westler, E.S. Mooberry, J.L. Markley, 'Practical Introduction to Theory and Implementation of Multinuclear, Multidimensional Nuclear Magnetic Resonance Experiments', *Methods Enzymol.*, **239**, 3–79 (1994).
31. L.E. Kay, G.Y. Xu, T. Yamazaki, 'Enhanced Sensitivity Triple Resonance Spectroscopy with Minimal H_2O Saturation', *J. Magn. Reson. Ser. A*, **109**, 129–133 (1994).
32. A. Bax, S.S. Pochapsky, 'Optimized Recording of Heteronuclear Multi-dimensional NMR Spectra using Pulse Field Gradients', *J. Magn. Reson.*, **99**, 638–643 (1992).
33. R.E. Hurd, 'Gradient-enhanced Spectroscopy', *J. Magn. Reson.*, **87**, 422–428 (1990).
34. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, 'Attenuated T2 Relaxation by Mutual Cancellation of Dipole-dipole Coupling and Chemical Shift Anisotropy indicates an Avenue to NMR Structures of very Large Biological Macromolecules in Solution', *Proc. Natl. Acad. Sci. U.S.A*, **94**, 12366–12371 (1997).
35. J. Fiaux, E.B. Bertelsen, A.L. Horwich, K. Wüthrich, 'NMR Analysis of a 900 K GroEL GroES Complex', *Nature*, **418**, 207–211 (2002).
36. D.S. Wishart, B.D. Sykes, 'The 13C Chemical-Shift Index: a Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-shift Data', *J. Biomol. NMR*, **4**, 171–180 (1994).
37. D.S. Wishart, B.D. Sykes, F.M. Richards, 'The Chemical Shift Index: a Fast and Simple Method for the Assignment of Protein Secondary Structure Through NMR Spectroscopy', *Biochemistry*, **31**, 1647–1651 (1992).
38. J. Kuszewski, J. Qin, A.M. Gronenborn, G.M. Clore, 'The Impact of Direct Refinement Against 13C Alpha and 13C Beta Chemical Shifts on Protein Structure Determination by NMR', *J. Magn. Reson. B*, **106**, 92–96 (1995).
39. J. Kuszewski, A.M. Gronenborn, G.M. Clore, 'The Impact of Direct Refinement Against Proton Chemical Shifts on Protein Structure Determination by NMR', *J. Magn. Reson. B*, **107**, 293–297 (1995).
40. J.S. Hu, A. Bax, 'Determination of Phi and Chi1 Angle in Proteins from 13C-13C Three Bond J Couplings Measured by Three-dimensional Heteronuclear NMR. How Planar is the Peptide Bond', *J. Am. Chem. Soc*, **119**, 6360–6368 (1997).
41. R. Konrat, D.R. Muhandiram, N.A. Farrow, L.E. Kay, 'Pulse Schemes for the Measurement of 3JC′C Gamma and 3JNC Gamma Scalar Couplings in 15N,13C Uniformly Labeled Proteins', *J Biomol. NMR*, **9**, 409–422 (1997).
42. C. Gayathra, A.A. Bothner-By, P.C.M. van Zijl, C. MacLean, 'Dipolar Magnetic Field Effects in NMR Spectroscopy', *Chem. Phys. Lett.*, **87**, 192–196 (1982).
43. N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, 'Use of Dipolar 1H-15N and 1H-13C Couplings in the Structure Determination of Magnetically Oriented Macromolecules in Solution', *Nat. Struct. Biol.*, **4**, 732–738 (1997).
44. N. Tjandra, D.S. Garrett, A.M. Gronenborn, A. Bax, G.M. Clore, 'Defining Long Range Order in NMR Structure Determination from the Dependence of Heteronuclear Relaxation Times on Rotational Diffusion Anisotropy', *Nat. Struct. Biol.*, **4**, 443–449 (1997).
45. R. Xu, B. Ayers, D. Cowburn, T.W. Muir, 'Chemical Ligation of Folded Recombinant Proteins: Segmental Isotopic Labeling of Domains for NMR Studies', *Proc. Natl. Acad. Sci. U.S.A*, **96**, 388–393 (1999).
46. V. Tugarinov, W.Y. Choy, V.Y. Orekhov, L.E. Kay, 'Solution NMR-derived Global Fold of a Monomeric 82-kDa Enzyme', *Proc. Natl. Acad. Sci. U.S.A*, **102**, 622–627 (2005).
47. A. Mittermaier, L.E. Kay, 'New Tools Provide New Insights in NMR Studies of Protein Dynamics', *Science*, **312**, 224–228 (2006).
48. K.A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M.A. Wilson, G.A. Petsko, M. Karplus, C.G. Hubner, D. Kern, 'Intrinsic Motions Along an Enzymatic Reaction Trajectory', *Nature*, **450**, 838–844 (2007).
49. K. Henzler-Wildman, D. Kern, 'Dynamic Personalities of Proteins', *Nature*, **450**, 964–972 (2007).
50. K. Sugase, H.J. Dyson, P.E. Wright, 'Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein', *Nature*, **447**, 1021–1025 (2007).
51. M. Mishkovsky, E. Kupce, L. Frydman, 'Ultrafast-based Projection-reconstruction Three-dimensional Nuclear Magnetic Resonance Spectroscopy', *J. Chem. Phys.*, **127**, 034507 (2007).
52. T. Szyperski, H.S. Atreya, 'Principles and Applications of GFT Projection NMR Spectroscopy', *Magn. Reson. Chem.*, **44**, S51–S60 (2006).

53. F. Fiorito, S. Hiller, G. Wider, K. Wüthrich, 'Automated Resonance Assignment of Proteins: 6D APSY-NMR', *J. Biomol. NMR*, **35**, 27–37 (2006).

54. S. Hiller, F. Fiorito, K. Wüthrich, G. Wider, 'Automated Projection Spectroscopy (APSY)', *Proc. Natl. Acad. Sci. U S A*, **102**, 10876–10881 (2005).

Nuclear Magnetic Resonance Instrumentation

Shi Bai
University of Delaware, Newark, DE, USA

1 Introduction **1**
2 Brief History of Nuclear Magnetic Resonance Instrumentation **4**
3 Current State of the Art in Nuclear Magnetic Resonance Instrumentation **5**
4 Description of Nuclear Magnetic Resonance Spectrometer **6**
4.1 Superconducting Magnet 8
4.2 Signal Generation and Radio-frequency Amplifiers 9
4.3 Detection and Radio-frequency Receiver 10
4.4 The Spectrometer Field-frequency Lock System 11
4.5 The Shim System 12
4.6 The Pulsed-field-gradient Amplifier 13
4.7 The Computer and Software 13
4.8 The Nuclear Magnetic Resonance Probe 15
4.9 Automation and Sample Changers 17
5 Practical Aspects of Nuclear Magnetic Resonance Spectrometer Operation **17**
5.1 Safety Precautions 17
5.2 Sample Preparation 18
5.3 Data Acquisition 19
5.4 Data Processing and Reduction 20
6 Future Perspectives in the Development of Nuclear Magnetic Resonance Spectrometers **21**
Acknowledgments **22**
Abbreviations and Acronyms **23**
Related Articles **23**
References **24**

The nuclear magnetic resonance (NMR) spectrometer consists of a superconducting magnet, a sample probe, a radio-frequency (RF) transmitter, a receiver, and a computer for instrument control and data processing. The superconducting magnet provides an ultrastable magnetic field in the central volume of the magnet. To maintain the current, the solenoid that produces the field is immersed in liquid helium to maintain the superconducting state. Surrounding the magnet volume is a set of vacuum jackets and a liquid nitrogen–filled volume to isolate the low-temperature regions from room temperature. The magnet has a bore to house the sample probe and a room-temperature shim (RTS) coil assembly to reduce inhomogeneity of the magnetic field across the active sample volume. The sample to be analyzed is usually dissolved in a deuterated solvent and is contained in a specially ground glass tube that is inserted into the volume of the probe coil. RF pulses, from one or a set of high-power transmitters, are applied to the sample through the RF coil that surrounds the active sample area to perturb the sample. The RF pulses are controlled to ensure precise frequency, phase, duration, and amplitude of the RF voltage. The response of the sample is gathered by recording the induced voltage across the RF coil. The response (the signal) is digitized as a so-called free induction decay (FID) with a high-performance RF receiver including a preamplifier, an amplifier, and an analog-to-digital converter (ADC). By using Fourier transformation (FT), the time-domain FID is converted into the frequency-domain spectrum, often known as the NMR spectrum.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

1 INTRODUCTION

NMR spectroscopy has a wide range of applications in chemistry, pharmaceuticals, proteomics, structural biology, structural genomics, metabolomics, and materials science and engineering, where it is used to identify structural features of the molecules in a sample. It is a powerful analytical tool for quantitative and qualitative characterization of these molecules. Because of the wide range of problems that analysis by NMR can address, NMR spectrometers are found in industrial, academic, and government laboratories around the world, wherever identification and quantification of the components of a sample are important.

The NMR signal arises from interactions between the magnetic moments of nuclear spins and the applied external magnetic field (*see* **Zeeman Interaction in Nuclear Magnetic Resonance**). The NMR spectrometer is used to detect these NMR signals, the interpretation of which yields important structural and dynamic information. The traditional information obtained from the NMR spectrum is the various interactions of nuclear spins with environment, such as chemical shifts, scalar coupling, and relaxation rates. For certain nuclei with nuclear spin $I > 1/2$ the magnitude of the quadrupolar coupling can also give information on the local electronic state (*see* **Quadrupole Couplings in Nuclear Magnetic Resonance, General**).

When a collection of NMR-active nuclear spins (which we consider here to have $I = 1/2$) is placed in a strong static magnetic field B_0, the magnetic moments, μ, associated with the nuclear spins precess about the direction of the magnetic field. The precession frequency is known as the *Larmor frequency*, ω_0, of the nucleus. If the magnetic moment is not directed along the magnetic field, it has two components, the projection along the field and the projection along the plane perpendicular to the field. The precession about the field causes this latter component to be time dependent.

After a sample has remained in a field for a time sufficiently long to establish equilibrium, (*see* **Relaxation in Nuclear Magnetic Resonance, General**) the collection of spins shows the following two qualities: (i) there is a finite magnetization aligned with the magnetic field and (ii) there is no net magnetization perpendicular to the field. The first of these phenomena is understood by the fact that there are two states with projection along the field, either parallel (α) or antiparallel (β). The population of α nuclear spins is slightly greater than that of β spins because of a slightly lower energy associated with α nuclear spins. The second is a result of the fact that the precession of each of the individual moments occurs in such a way that the various spin moments have incoherent phase. The Larmor frequency, ω_0, which is the nuclear precession frequency in the applied external magnetic field expressed in radians per second, is proportional to the strength of the external field:

$$\omega_0 = \gamma B_0 \tag{1}$$

where γ is the magnetogyric ratio of the nucleus being observed, and B_0 is the magnetic-field strength in tesla. When the Larmor frequency is expressed in units of cycles per second (Hz), we have $\upsilon_0 = \gamma B_0/2\pi$. For a magnet with a field of 14.09 T, the nominal Larmor frequencies of ^{1}H and ^{13}C are 600.0 and 150.9 MHz, respectively. For this reason, besides quoting the magnetic-field strength in tesla, it is common parlance to specify the field strength of an NMR spectrometer in terms of the ^{1}H Larmor frequency. It also happens that the Larmor frequency is directly proportional to the energy difference between the two states:

$$\Delta E = \hbar\omega_0 \tag{2}$$

where $\hbar$ is Planck's constant divided by 2π. Thus, a measure of the Larmor frequency is a measure of the energy spacing between the two stationary states of a spin. For example, for a 600 MHz magnet at 300 K, the energy difference, ΔE, is 3.97×10^{-25} J for ^{1}H nuclei. This small energy difference results in a very small population difference between the two levels at room temperature. The induced voltages (the NMR signal) are therefore quite small, even at the highest achievable magnetic fields. This fact makes NMR spectroscopy a relatively insensitive technique, when compared to other spectroscopic methods such as infrared or optical spectroscopy. Because of this issue, the majority of developments in NMR instrumentation focus on various means to increase the sensitivity of NMR measurements.

Since the Boltzmann factor that determines the magnitude of the NMR signal is proportional to the externally applied field B_0, higher magnetic-field strength leads to higher sensitivity in the NMR experiment. The additional benefit of higher field is that, at higher frequencies, the natural noise in the detection circuits (including the NMR probe, receiver, and amplifiers) is lower, so that the signal-to-noise ratio (S/N) increases faster than linearly with the magnetic field. In recent decades, we have seen a continual development of higher and higher magnetic fields for use in NMR spectroscopy.

A second practical reason for the use of higher magnetic fields follows directly from quantum mechanical consequences of the direct dependence of the magnitudes of various couplings on magnetic field. The two couplings that are the ''bread and butter'' of NMR are the chemical shift and the scalar, or J, coupling (*see* **Chemical Shifts in Nuclear Magnetic Resonance**; **Scalar Couplings in Nuclear Magnetic Resonance, General**). What makes NMR such an important technique is that nuclei in different local environments have precession frequencies (and hence energy spacings) slightly different from the nominal Larmor frequency. The measurement of all of these frequencies, reported as chemical shifts, specifies the local structure of the molecule in which the spins are embedded. The chemical shift range expressed in hertz is proportional to the applied magnetic-field strength, B_0. Thus, at higher fields, two resonances are separated by a larger frequency difference than they would be at lower field. On the other hand, the spin–spin coupling is independent of field. Thus, at higher field, the chemical shift difference between two resonances becomes relatively larger than the J coupling. The form of the NMR spectrum depends on the relative size of the J coupling and chemical shift difference between the two resonances. In particular, when the J coupling is much smaller than the chemical shift difference, the form of the spectrum becomes quite simple. Thus, evaluating a spectrum at high field guarantees a simpler spectrum than one acquired at lower field.

For spins other than 1/2 there is an additional effect, the quadrupolar coupling, to electric field gradients that can affect the spectrum. This interaction is independent of field strength, so carrying out experiments at higher magnetic field provides, in some cases, a simplification of the NMR spectrum. Therefore, there are several reasons to develop higher magnetic fields for NMR spectroscopy or to use the highest possible magnetic field for studies.

As mentioned above, the equilibrium condition is such that the sample has a macroscopic magnetization that lies along the magnetic field, with no macroscopic magnetization perpendicular to the magnetic field. A common way to measure the precession frequencies is to place magnetic moments perpendicular to the main magnetic field so that they precess about its direction. This strategy employing the transverse magnetization allows one to record magnetic fields in the sample that are many orders of magnitude smaller than the main static fields. One may create this transverse magnetization by applying appropriate bursts of RF energy to the sample, after which the sample acquires a macroscopic magnetization perpendicular to the main field direction for a finite amount of time. The RF burst (called a *pulse*) must have a carrier frequency close to the frequency of precession of the nuclear spins; it must be reasonably strong; and it must produce a coherent time-dependent magnetic field, B_1. An RF current through a solenoidal coil (or coils in some geometries) produces a B_1 field along the axis of the solenoid; if the current is time dependent, the magnitude of the field is also time dependent. The application of such a field (oriented perpendicular to the main field) to a system initially at equilibrium in the main field results in the creation of magnetization perpendicular to the main field. If the time of excitation is sufficiently short (i.e. the excitation is a short pulse), the transverse magnetization at the end of the RF pulse will be coherent and will subsequently precess about the main field at a frequency characteristic of its particular environment, as determined by interactions such as the chemical shift and J coupling. The RF pulse may produce various effects, depending on the amplitude, the RF phase, and the carrier frequency, so these parameters must be precisely controlled to make the greatest use of a modern NMR spectrometer.

Let us introduce a bit of nomenclature. If the magnetization is allowed to rotate 90° from its equilibrium position along the external magnetic field (or z axis) during the pulse, the RF excitation is referred to as a 90° pulse. If the magnetization is allowed to rotate 180° during the RF excitation, the pulse is said to be a 180° pulse, and so on. In addition, the relative phase of the RF excitation is important. Relative to some coordinate system, the B_1 field may fall along the x axis or the y axis, or any other axis in the transverse plane. Generally, the x and y directions (and sometimes the $-x$ and $-y$ directions) are the most important, and so the pulse is called an *x pulse* or a *y pulse*, with symbols 90°_x and 90°_y, respectively. The phase of the pulse indicates the Cartesian axis around which the magnetization is rotated and, hence, where the magnetization ends after the 90° pulse. An x pulse puts the magnetization initially parallel to the y axis, whereas a y pulse puts the magnetization along the $-x$ axis. It is important to realize that the sense of rotation is arbitrary; by convention a right-hand Cartesian axis system is chosen; however, the rules can be reformulated for left-hand rotations, with equivalent results.

Following a 90° RF pulse, the net magnetization, being perpendicular to the field, precesses at the Larmor frequency around the direction of external magnetic field. This motion of the magnetization induces a time-dependent voltage in the probe coil, which can be detected and digitized as a function of time, which is the FID. FT of the time-domain FID produces a frequency-domain NMR spectrum that contains all the frequency components of the magnetization. From the frequency spectrum, the analyst can determine the presence or absence of certain types of nuclei and certain types of chemical environment. Aside from this simplest NMR experiment, the control of pulses allows one to perturb the magnetization's evolution by a variety of other experimental protocols to provide selective information on couplings and kinetic constants.

In an NMR spectrometer designed for detection of nuclei in liquid samples, there are usually several RF systems operating simultaneously, generally at different frequencies. These are termed *channels* of the spectrometer. In addition to the channel employed for detection of the desired nucleus (the *observe channel*), one channel is usually dedicated to the detection of the deuterium signal from the deuterated solvent (e.g. $CDCl_3$) in which the sample is dissolved. This channel is often referred to as the *lock channel* because the corresponding signal is used for the purpose of controlling the long-term stability of the magnetic field. The number of additional RF channels depends on the configuration of the spectrometer, and that in turn is dictated by the kinds of experiments being done on the spectrometer. For most routine analytical NMR spectrometers, besides the lock channel, there are two observe RF channels, one for high-frequency ^1H detection and another for multinuclear detection usually at lower frequencies, with the capability of providing any frequency over a wide range. This channel is often labeled the *broadband channel*. In addition, many experiments require decoupling of some nuclei (often protons), so the high-frequency channel may also be known as the *decoupling channel*. For high-end biomolecular NMR research spectrometers, there are at least three observe channels (^1H, ^{13}C, and ^{15}N) because of the need to carry out highly sophisticated multiple-frequency experiments. Modern instruments usually have equivalent broadband transmitters for each channel, while the amplifiers dedicated to different nuclei may be different (e.g. ^1H amplifier has a unique high-frequency configuration, with the low-frequency broadband amplifiers used for the remaining channels being the same).

It should be obvious from the short discussion above that the NMR spectrometer requires the following: (i) a strong, stable, and highly homogeneous magnet, (ii) a

probe circuit that houses the sample and has a coil, often a solenoid, to which RF voltage is applied to produce the perturbing field, (iii) a set of RF amplifiers to generate the RF voltages (pulses) applied to the coil, (iv) an RF receiver capable of detecting the tiny voltages generated by the precession of the nuclear magnetization, (v) a system to perform the detection (which usually involves some demodulation, amplification at a lower frequency, the necessary devices to transform the analog voltages into digital data, and some means of storage), and (vi) a number of computers with special software to control all aspects of spectrometer RF pulsing, data acquisition, and processing.

After a brief discussion of the history of, and the current state of the art in, NMR instrumentation, each part of the NMR spectrometer is described in more detail. Since our laboratory is mostly equipped with Bruker BioSpin instrumentation, the Bruker NMR spectrometer is used as an example to illustrate the concepts and architecture of the NMR instrumentation in the following sections. Similar descriptions could be made of laboratories employing other NMR spectrometers (e.g. Varian, JEOL, or Tecmag), although the details of analysis, processing, and setup may be slightly different. A general procedure to operate a solution NMR spectrometer and to acquire an NMR spectrum is given. Finally, the future perspective in the development of NMR instrumentation is discussed.

2 BRIEF HISTORY OF NUCLEAR MAGNETIC RESONANCE INSTRUMENTATION

Since the first detection of NMR signals in bulk materials independently by Bloch et al.[1] and Purcell et al.[2] in 1946, NMR instrumentation has evolved rapidly and continuously for more than 60 years. Early NMR experiments were carried out with permanent magnets or electromagnets. In those days, the magnetic field was slowly swept over a region with an RF field constantly impinging on the sample. When the resonance condition of Equation (1) was encountered, the response was observed as a change in the quality (Q) factor of the coil. Thus, early spectrometers (known as *continuous-wave (CW) spectrometers*) produced a plot of the Q factor of the coil as a function of the field, the Q factor of the coil being a measure of the absorption of energy by the sample. In principle, one could sweep the frequency while holding the field constant, and that was indeed done by some groups. It was, however, more difficult to sweep the frequency of the excitation field, so the majority of spectrometers at that time were of the field-swept variety.

The first commercial NMR spectrometer with a permanent magnet running at the ^{1}H resonance frequency of 30 MHz was pioneered by Varian in 1952. It was a field-swept spectrometer. Already, the trend toward building magnets with higher fields was evident. By the early 1960s, electromagnets with field strengths up to 100 MHz were commercially available. In general, these spectrometers could detect only the resonance of a single nuclear species, usually the proton.

In the mid-1960s, the first superconducting magnet was introduced by Varian.[3] The superconducting solenoid magnet, much more stable than electromagnets, provided magnetic-field strengths unattainable with electromagnets. This resulted in an overall increase in sensitivity. In addition, because at higher fields one could still attain the same magnetic-field homogeneity, the resolution of the NMR measurements was greatly enhanced by the use of the superconducting magnets.

At about the same time that superconducting magnets became available, the demonstration[4] of the Fourier transform relationship between the FID and the frequency spectrum and the development of fast Fourier transform (FFT) paved a new path for NMR spectroscopy. The combination of the detection of the time response and the superconducting magnet added another dimension to the NMR experiment. Because of the ultrastability of superconducting magnets and the exact repeatability of the time response, signal averaging (which had been done only occasionally in swept-mode systems) became commonplace. One year after the demonstration of the technique, the first FT NMR with a superconducting magnet was marketed.[5] The FT experiment is generally repeated in a few tens of seconds, rather than the many minutes it took for acquisition of a single CW spectrum, and allowed a dramatic increase in the S/N within a given time. This added the real possibility of detecting nuclei like ^{13}C with routine spectroscopy, and soon multinuclear spectroscopy became a part of the analytical toolbox.

Of equal importance has been the fact that excitation sequences were developed to parse or factor the information in a spectrum. For example, the utilization of RF-pulse sequences has resulted in multidimensional NMR techniques, from which rich structural information is easily extracted.[6,7] These multidimensional NMR experiments have become part of routine experimental protocols and are now offered as a standard component of any NMR spectrometer software package.

Since its introduction, the superconducting magnet has been one of the driving forces behind modern NMR spectroscopy. The magnetic-field strength attainable by the highest-field commercial superconducting magnet has increased roughly linearly from a Varian 200 MHz spectrometer in 1966 to a Bruker BioSpin 950 MHz spectrometer in 2006.[3]

Developments in probe technology also greatly expanded the utility of NMR over the years. In early

studies, the probe circuit was tuned for a single nucleus (usually ^{1}H) and the response of the ^{1}H was monitored with all the interactions that naturally affect the proton being seen in the structure of the spectrum. Over the years, probes have evolved from a single-channel type to those that contain multiple RF channels to accommodate the excitation and detection of several nuclei simultaneously. Early multichannel probes had to be tuned manually, which is a time-consuming operation. Modern probe design and computer control allow fully automatic tuning and matching (ATM) of several different simultaneous circuits. This automatic operation has produced an open-access laboratory environment, where the experimenter does not have to delve into the eccentricities of RF electronics to obtain usable NMR spectra. Recently, the development of cryogenically cooled probes has led to a leap in sensitivity owing to noise reduction in the probe electronics.

In the wake of the rapid development of digital electronics, extensive digital control of the spectrometer is common nowadays. This also simplifies the adjustments that were necessarily made by operators with older analog (or partially analog) spectrometers. Needless to say, the rapid developments in computer technology quickly become incorporated into NMR spectrometers.

3 CURRENT STATE OF THE ART IN NUCLEAR MAGNETIC RESONANCE INSTRUMENTATION

A typical twenty-first-century state-of-the-art NMR spectrometer is shown in Figure 1. An actively shielded superconducting solenoid produces a 600 MHz magnetic field (A). The magnet is mounted on antivibration posts (B) to isolate low-frequency (<25 Hz) building floor vibrations that might affect the NMR spectroscopy. Floor vibrations can be easily observed as satellite peaks appearing around the main resonance peaks in a high-resolution NMR spectrum. With reasonable magnetic-field homogeneity, the intensities of floor vibration peaks can be reduced to less than 0.55% of the main peak intensity.

The RF generators, amplifiers, receivers, lock channel, pulsed-field-gradient (PFG) amplifiers, and other spectrometer control units are mounted inside the spectrometer console (E). In a basic configuration, there are three RF channels including one for ^{1}H, a second for other (sometimes called X) nuclei (i.e. ^{13}C, ^{15}N, $^{31}P\ldots$), and the third for the deuterium lock that regulates the magnetic field. The RF pulses are created by a computer-controlled pulse program. This computer program controls precisely the frequencies and the timing

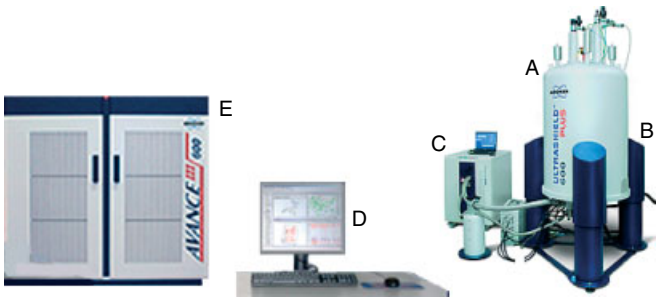

Figure 1 Modern NMR spectrometer: A, actively shielded superconducting magnet; B, antivibration posts; C, cryogenic cooling unit; D, primary computer; E, spectrometer console.

of the RF pulses. Frequency generation and timing of RF pulses are carried out digitally to produce pure frequencies. The digital pulse instructions are converted into analog waveforms that are brought to the appropriate power level by the RF amplifiers, which are linear over a wide range (<120 dB) and capable of producing several hundred watts of RF power. Modern receivers are completely digital and use oversampling technology to eliminate artifacts related to the digitization process.

The ^{2}H lock channel is also digitally controlled by its own microprocessor to compensate for any field drift and to maintain field homogeneity through adjustment of the RTS coils. The primary computer (D) provides a graphic user interface (GUI) that allows the spectrometer operator to set up a pulse program, control experimental progress, and process the acquired data.

The type of probe in a particular system depends on the type of research for which the spectrometer is used or the role that the NMR spectrometer serves. In Figure 1, the cryogenically cooled probe (mounted inside the magnet) is controlled by a cryocooling unit (C). Other multiple probe configurations are available commercially. The most popular configuration for routine use in analytical solution NMR is a 5-mm inverse broadband probe with PFG capability. Another useful probe for routine service is the so-called quad nuclei probe, which is pretuned for four different nuclei, with ease of switching among the channels by digital commands. These probes are readily used for acquiring data in one-dimensional (1-D), two-dimensional (2-D), and three-dimensional (3-D) NMR experiments without the necessity of changing cabling or retuning the probe.

The architecture of modern NMR spectrometers is modular, which makes expansion of the spectrometer functionality relatively simple. The addition of a sample changer or other automation accessories and the associated control software makes a routine open-access spectrometer that can be easily operated by users having only had a brief basic training. An additional RF amplifier and transmitter (e.g. ^{15}N) can be added

to the basic spectrometer to extend the capability for analyzing biological samples. In addition, many types of cryogenically cooled probes are available for the enhanced sensitivity required in experiments such as the detection of natural-abundance ^{15}N, the ^{13}C 2-D INADEQUATE (incredible natural-abundance double quantum transfer experiment), the ^{13}C measurement of polymeric materials at high temperature, and triple-resonance (^{1}H, ^{13}C, ^{15}N) experiments that are essential to analysis of biological macromolecular samples. When equipped with a high-power amplifier, a solid-state probe, and a magic-angle spinning (MAS) controller, the modern NMR spectrometer can readily analyze solid materials.

NMR spectroscopy has also become an integral part of many hyphenated analytical techniques such as liquid chromatography/nuclear magnetic resonance (LC-NMR) (*see* **High-performance Liquid Chromatography Nuclear Magnetic Resonance**). These techniques find a great deal of use in the pharmaceutical and chemical industries, where effective separation and structural elucidations may be carried out in a single apparatus. The mixture separated by liquid chromatography (LC) is routed to the NMR spectrometer where the individual pure components of the mixture are injected individually into a flow NMR probe to be analyzed.

Further developments even combine LC, NMR, and mass spectrometry into a liquid chromatography/nuclear magnetic resonance/mass spectrometry (LC-NMR/MS) system. Such systems are valuable because it is often desirable to have multiple means of analyzing some samples. A commercially available integrated system allows the use of an NMR spectrometer and a time-of-flight (TOF) mass spectrometer to give both analyses of a sample. This technology is an ideal platform for metabolomic studies. The use of solid phase extraction (SPE) provides an efficient interface between LC and NMR when demanded for special flow probes. The liquid chromatography-mass spectrometry (LC-MS) provides the exact mass of the analyzed sample and access to the sum formula. The combination of these three powerful techniques often gives the answer to a question in a matter of minutes.

The ubiquity of nuclear magnetic resonance spectroscopy, especially its wide applications to analysis of chemicals in the laboratory or in the factory, has created an industry catering to the needs of NMR spectroscopy. Companies specialize in the production of NMR solvents, or in specifically isotopically enriched materials for use in studies that focus on a particular site (such as in studies of enzymes). Some companies specialize in consumables such as NMR tubes. Finally, the development of NMR software for specific applications is yet another niche market that is serviced by a number of companies. Readers are referred to the website http://www.ebyte.it/library/NmrMriCompanies.html for a constantly updated directory of NMR and magnetic resonance imaging (MRI) companies.

4 DESCRIPTION OF NUCLEAR MAGNETIC RESONANCE SPECTROMETER

A modern NMR spectrometer consists of a superconducting magnet, including a cryoshim coil assembly to compensate for the inhomogeneity of magnetic field, an RTS coil assembly to adjust the magnetic homogeneity variations between the different samples, a sample probe, transmitter channels including a lock channel, an RF receiver system, and a computer. In addition, there may be a subsidiary slave acquisition computer (AC) for controlling the RF transmitters, which is addressed by the primary computer. A schematic diagram of an NMR spectrometer is shown in Figure 2. The primary computer (A) serves as an interface between the spectrometer operator and all the components of the spectrometer. The operator's instructions are interpreted by the primary computer and are transferred to an AC through an Ethernet cable (B). From the AC, all incoming information is routed to appropriate devices through a data bus or through RS-232 or RS-485 serial lines. The information returned to the AC, such as variable temperature-control output, lock signal, shimming responses, and other control signals are transferred back to the primary computer through the Ethernet cable (B). The time-critical information such as pulse timing and frequency control are sent back and forth from the AC to the appropriate control boards through a high-speed data bus. Signals that are not time critical (shim, lock, and variable temperature control) are transferred through an RS-232 or RS-485 cable. Most recently, the RS-485, which is faster than RS-232 and can be used to control several units simultaneously, has been used for communication to the preamplifiers and signal-generation units.

All NMR experiments start with a sequence of instructions for RF pulsing, delays, and data acquisition. The pulse sequence, entered by the operator at the primary computer, is transferred to the AC as the instructions for the synthesizer to produce RF signals with a specific frequency near the Larmor frequencies of the nucleus being studied. All RF events are synchronized to the same clock. The digital RF signal instructions are converted into analog commands (voltages) that can drive the RF amplifiers. The output voltage sequence of the RF amplifiers (ranging from milliwatts for decoupling to hundreds of watts for excitation) is applied across the probe coil to produce the RF magnetic field that excites the sample.

At appropriate times, the probe coil is also used as a detector of the NMR signal. The resulting induced

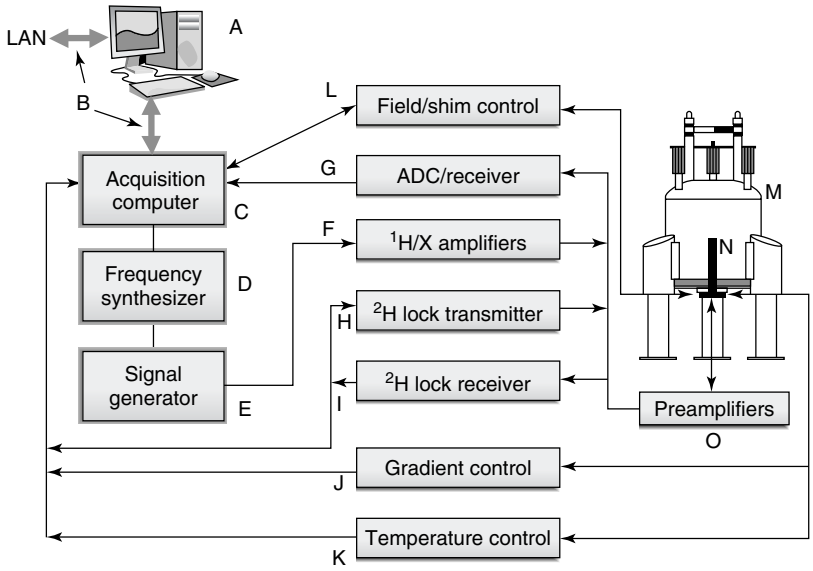

Figure 2 Schematic diagram of an NMR spectrometer: A, primary computer; B, Ethernet connections; C, acquisition computer; D, frequency synthesizer; E, signal generator; F, RF amplifiers; G, RF receivers and analog-to-digital convert; H, ^{2}H lock transmitter; I, ^{2}H lock receiver; J, gradient control and amplifier; K, temperature-control unit; L, room-temperature field and shim control and power supply; M, superconducting solenoid magnet; N, NMR probe; O, preamplifiers.

voltage is amplified by an RF preamplifier. The amplified signal is fed to the receiver section, which includes a reference mixer to reduce the frequency of NMR signal from the RF region to the audio frequency region. At this point, the signal is often amplified with an audio amplifier. The analog voltage is then digitized. Prior to digitization, a digital noise reduction filter is often used to remove aliased signals and noise. The digitized signal is then stored in the principal computer. When the experiment is repeated, the data are coadded and stored in the principal computer.

If the spectrometer is equipped with a gradient accessory for destruction of undesired magnetization remaining between repeated scans, gradient-pulsing control is carried out by the same computer, so that RF and gradient control are synchronized. As shown in Figure 2, the lock channel (F and H) can be considered as an independent ^{2}H spectrometer parallel to the *observe* and *decoupling* channels.

Subsequent to data collection, the digitized signal (the FID) is transferred to a frequency-domain spectrum using the FFT algorithm by the primary computer. At this point, digital manipulation of the FID is possible, for example, with various apodization procedures. An example of FFT is given in Figure 3, where the time-domain FID is on the left and the frequency-domain spectrum is on the right.

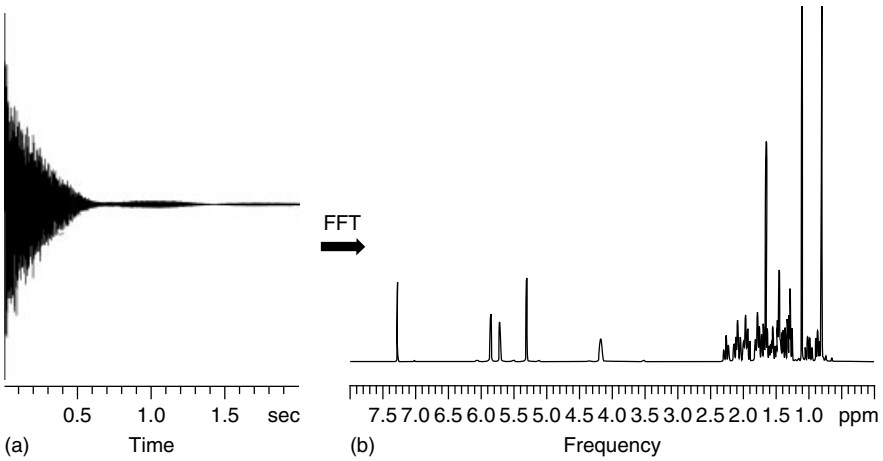

Figure 3 ^{1}H NMR spectra of gramicidine-S in DMSO-d6: (a), the FID in time domain and (b), spectrum in frequency domain. The Fourier transform is used to convert the time-domain FID to frequency-domain spectrum.

The exponential decay of the FID signal results from nuclear relaxation processes.

4.1 Superconducting Magnet

An ultrastable strong magnet field in the active sample volume is provided by a solenoid made of alloys generally containing niobium, which are superconducting at the temperature of liquid helium. A schematic diagram of a superconducting magnet is given in Figure 4. The intense static magnetic field is the result of the constant current in a coil of superconducting niobium-alloy wire that comprises the magnet (F). The wire must be immersed in a bath of liquid helium (<6 K) (D) to keep it in superconducting state. In the superconducting state, the electric current running through the magnet coil is said to be *persistent*, in that there are no losses of current due to resistive effects. Besides the main coil, a set of cryoshim coils also in the superconducting state is provided to create small fields to offset inhomogeneities in the main field. Once set, the currents in all the superconducting coils remain constant and are not generally changed by the user. To reduce the boil-off of liquid helium, it is encased in a vacuum jacket packed with materials that limit radiative losses. This setup is surrounded by a further jacket (E) filled with liquid nitrogen (77 K) and another vacuum jacket (C) to isolate it as much as possible from the room-temperature environment. The central bore of the magnet, which houses the sample probe, is at room temperature.

At the time of installation of the NMR spectrometer, the jackets are pumped down to $\sim 1 \times 10^{-5}$ Pa to remove as much air and other ''contaminants'' (if present) as possible before the cryogenic liquids are added. The helium jacket is generally precooled with liquid nitrogen at this stage, with the liquid nitrogen being removed before the liquid helium is added. After the cool down, the magnet is charged with electric current by connecting it to a special power supply. For an actively shielded 600 MHz magnet, a 134-A current is slowly introduced into the magnet coil until the magnetic-field strength reaches 14.09 T. Following this procedure, the cryoshim coils are also given appropriate currents. The optimal set of currents is determined by observation of the line shape of a water sample. Generally, one should be able to achieve a line width of less than 0.5 ppm with a reasonable line shape using only the cryoshim system. Once the magnet field strength is reached and the cryoshim currents have been optimized, a set of heaters that has kept some parts of the wire nonsuperconducting to allow connection to the power supply is carefully turned off to produce a self-contained superconducting magnet with persistent currents. The final step is the complete removal of the power supply from the magnet. The magnet is not connected to any source of power during operation and will continue to be persistent almost indefinitely, as long as it is kept at temperatures low enough to maintain the wires in the superconducting state.

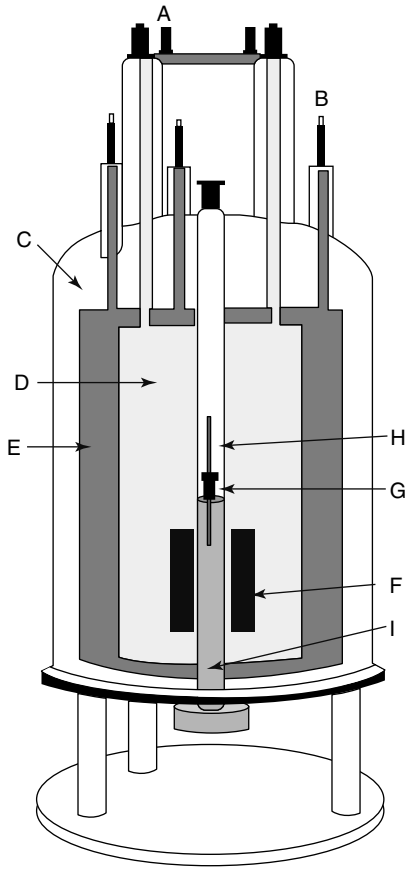

Figure 4 Schematic diagram of a superconducting solenoid magnet: A, liquid helium refill port; B, liquid nitrogen refill port; C, outer vacuum chamber; D, liquid helium bath; E, liquid nitrogen bath; F, superconducting solenoidal magnet; G, sample spinner; H, sample tube; I, NMR probe.

Any coil produces a continuous field, both inside the coil and outside. The components of the field outside the coil affect the space around the magnet and could potentially present safety hazards for those working around the magnet. Magnets produced most recently are *actively shielded*. In actively shielded magnets, the field components outside the magnet housing are compensated with additional coils and special windings to minimize the intrusion of the magnetic field into that space. Active shielding allows much more effective use of laboratory space.

Magnets in the range of 300–500 MHz (proton frequency) are currently very popular in the analytical NMR laboratory. The use of magnets in the range of 600–750 MHz is becoming increasingly accessible as the technology becomes more stable and the costs of these

systems decrease. As of 2007, a 950 MHz magnet has been successfully tested in a manufacturing facility and will be relocated to a customer's laboratory. Thus, one should expect the trend toward the use of ever-higher magnetic fields to continue for some time.

As we have mentioned, magnets are generally (if somewhat colloquially) specified by the proton resonance frequency and the diameter of the room-temperature bore. Thus, a 500/SB MHz magnet means that the proton resonance frequency is 500 MHz and the room-temperature bore diameter is 54 mm (''standard bore''). There are also wide-bore (WB) systems, and these are similarly expressed as 500/WB, typically meaning a bore of 89 mm.

To maintain the magnet in the superconducting state, a weekly refill of the liquid nitrogen reservoir and a periodic refilling of the liquid helium reservoir (approximately every 2–5 months, depending on the magnet design) are required.

4.2 Signal Generation and Radio-frequency Amplifiers

In a modern NMR spectrometer, a set of RF transmitters produces the voltage that creates the perturbing time-dependent magnetic field. Usually, an RF transmitter consists of an RF source, a phase modulator, and an amplifier to generate the voltage of the desired amplitude, frequency, and phase. The phase modulator determines the phases of the pulses during the experiment, such that one may demand an x, a y, or any other pulse of RF magnetic field. The amplifier provides the RF voltage that impinges on the coil. It must be capable of providing outputs of a few watts (for a solution-state NMR spectrometer) to a few hundreds of watts or even kilowatts (for a solid-state NMR spectrometer).

The RF system provides the means to produce the voltages that excite the nuclei in the sample. The RF waveforms are generally produced at low power with various gating devices that allow the passage of RF current when an appropriate voltage is applied. The RF signal from a digital synthesizer passes through this stage, and from the output comes the appropriately modulated RF signal. The modulators in this stage are all connected to the AC that controls timing, frequency, shapes, phase, and amplitudes of the RF pulses according to the pulse program loaded into it from the primary computer. In modern NMR experiments, there are often excitations at several different frequencies for the various nuclei in a sample, so there are several channels, each similar to the others in function but working at a different frequency. When multiple RF channels are involved, all channels must be able to operate independently, as well as being synchronized with each other through their control by the AC. In a modern NMR spectrometer, the timing controllers are able to control more than 60 events with a timing resolution of 12 ns. The modulated output drives the high-power amplifiers that produce voltages corresponding to radiation up to several kilowatts across the sample coil.

All NMR experiments start from the operator's instructions, a computer-programmed pulse sequence. In a pulse program, the frequencies, timing, phase, and amplitude of pulses in the channels are specified by the experimenter. This is usually done at the primary computer. Execution of the pulse program presents a series of digital instructions to the NMR spectrometer interface (usually through the intermediacy of the AC), containing a DAC that produces the appropriate analog waveform from the digital instructions of the pulse program. All ''run-time'' decisions, such as conditional loops and phase and frequency shifts, are made by the AC to control counters, gates, and electronic phase shifters in the RF signal-generation unit. The signal-generation unit creates the analog RF pulses and delays that drive the sample to make a response. A typical signal-generation unit has a frequency range of 3–1100 MHz, a frequency resolution less than 0.005 Hz, a switching time less than 300 ns, and phase resolution less than 0.006°. An additional function of a signal-generation unit is to produce reference signals, gating, and the dwell clock for the receivers.

The output of the signal-generation unit is passed to high-power RF amplifiers for further amplification. The output of the amplifiers is passed directly to the sample coil, where it perturbs the sample in a manner determined by the pulse program written by the experimenter. The high-power amplifier has a wide linear range (120 dB or more) so that it can accurately reproduce the RF signal passed to it. The RF output of the amplifier ranges from a few hundred watts for a solution-state spectrometer to a kilowatt for a solid-state spectrometer.

The pulse has a limited region of frequencies around the carrier frequency that it affects. The bandwidth of the excitation by an RF pulse is determined by the RF field strength, B_1, which also determines the 90° pulse width. The relationship between pulse width and the RF field strength is straightforward:

$$\frac{\gamma B_1}{2\pi} = \frac{1}{PW_{360}} = \frac{1}{4PW_{90}}\text{Hz} \tag{3}$$

where PW_{360} and PW_{90} are the pulse widths for 360° and 90°, respectively. For example, a 90° pulse width of 10 μs corresponds to a B_1 field strength of 25 kHz, a typical value for pulse excitation on a modern solution NMR spectrometer. This pulse width is generally acceptable to excite the full range of possible nuclear frequencies for nuclei such as ^{1}H and ^{13}C. This type of pulse is sometimes

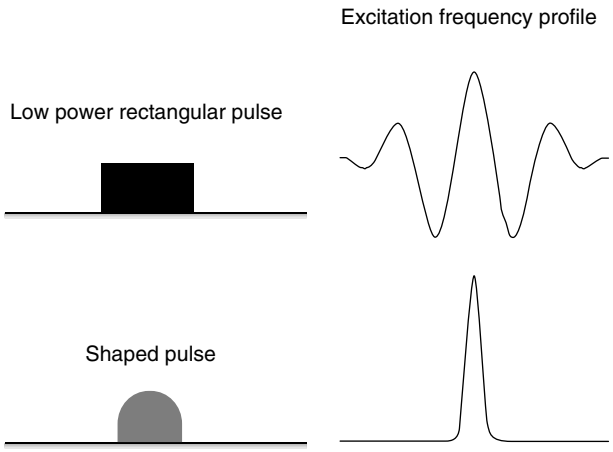

Figure 5 Schematic excitation profiles of a low-power rectangular pulse and a shaped pulse.

called a *hard pulse*. For some nuclei, a uniform excitation is not always possible using the conventional solution NMR hardware because of the large range of frequencies the excitation must span. As mentioned above, pulses also have phase properties that determine along which transverse axis, x, y, $-x$, or $-y$, the magnetization lies after the RF pulse. This pulse phase is defined relative to an internal reference signal shared with all transmitter channels and receivers.

Pulse shapes also affect the excitation bandwidth. The common hard pulse described above is rectangular, with a finite width, constant amplitude, and phase. Sometimes it is necessary to excite a narrow region of the spectrum for selective transfer experiments, solvent suppression, and reduction of the dimensionality of multiple-dimension pulse sequences. In principle, the simple selective pulse can be a long, weak rectangular pulse. However, as shown in Figure 5, the excitation profile of a rectangular pulse often involves extended side lobes in the frequency excitation profile, which often create a sinc-like oscillation in the resulting spectrum. Also shown in Figure 5, the selective pulse, sometimes called a *soft pulse*, is mostly a shaped pulse, which is often used to smooth out the side lobes in the excitation profile caused by the rectangular pulse shape. The most commonly used shaped pulses have a profile of a ramp, a sine, a trapezoid, or a Gaussian. The selective shaped pulses, characterized by their duration, frequency profile, and phase behavior, are typically 1–100 ms in length, three orders of magnitude longer than the hard pulses.

4.3 Detection and Radio-frequency Receiver

Each RF channel is equipped with high-performance preamplifiers and RF receivers to detect and amplify the signal. Subsequent to excitation, the detected sample response is an analog voltage across the coil. The signal is ultimately converted to a digital form by a highly sophisticated ADC. The weak NMR signal (approximately microvolts) goes through preamplification while it is still a radio-frequency signal. This preamplification occurs in a unit separated from the console and located near the magnet to avoid loss in transmission through the cable. With an actively shielded magnet, the spectrometer console can be located much closer to the magnet than with a conventional magnet. In this case, the preamplifiers may be housed inside the spectrometer console to maximize the usage of laboratory space. The preamplified signal is usually then transformed into a lower-frequency signal by demodulation against a reference signal. After that, the lower-frequency signal is amplified again (often to filter high-frequency noise) to the level of several volts and finally digitized for storage and further processing.

A receiver must be able to amplify the signal over all frequencies within the spectral width (SW) linearly; otherwise, the amplitudes in the spectra cannot be used for quantitative results. In addition, the receiver must retain all phase information of the input signal. A large dynamic range is desirable for a receiver so that small signals in the spectrum are detected with appropriate ratios so that they can be quantitated. The receivers in modern spectrometers have variable gain under the control of the experimenter. It is very important to set the receiver gain properly. If the signal is overamplified (i.e. the gain is set very high so that receiver output is greater than 5 V), the digitizer is saturated by the largest signals, resulting in some signals seeming to have the same amplitude. A ''clipped'' FID is the result of saturating the digitizer. If the gain is too small, only a part of the dynamic range of a digitizer is used and poor reproduction of the signal results.

The chemical shift ranges (kilohertz) are much smaller than Larmor frequencies (typically tens or hundreds of megahertz). For example, the frequency range of a proton spectrum at 600 MHz is between 7 and 8 kHz, and is about 20 kHz for a ^{13}C spectrum obtained on the same spectrometer. Often, one is only interested in chemical shift differences, instead of the absolute Larmor frequency. Therefore, it is a common approach to demodulate the signal to a lower frequency, which is equivalent to subtracting a reference frequency from the detected signal frequency before digitization. The reference frequency signal is usually chosen to be the carrier frequency of the exciting pulse. In addition, the RF signal is frequently referenced against a waveform with a phase reference differing by 90°. This gives the so-called in-phase and out-of-phase components of the signal, which are simultaneously digitized and stored together. This data detection protocol, referred to as *quadrature*

detection, is widely used. It has the advantage that the Fourier transform of the quadrature signal distinguishes between positive and negative frequencies (relative to the reference).

The digitizer works on the ''sample and hold'' principle, in which the signal is sampled repetitively with a specified interval. The time between samples is called the *dwell time* (*DW*). The minimum DW one may use is dictated by the Nyquist theorem, which states that the digitization rate must be at least twice as large as the highest frequency one wishes to observe. If the Nyquist theorem is violated during digitization, spectral aliasing or folding may occur, in which peaks that have frequencies outside of the spectral window are detected as peaks at frequencies within the spectral window, often with a weaker intensity or a phase shift. To improve the performance of digitization, one may use an analog or digital filter to remove signals at frequencies outside of the spectral window. Digital filtering has replaced analog filtering done with passive analog electronic circuits. Employing oversampling, where many more points than required by the Nyquist theorem are collected, has helped the performance of digital filter programs and is common in most modern spectrometers.

In quadrature detection, the NMR signal consists of two time series related by the phase between the two reference signals. Because there are twice as many data in a quadrature set, the S/N in a spectrum collected in the quadrature mode is increased by a factor of $\sqrt{2}$ over that of a spectrum that comes from a single time series. In an experiment using quadrature detection, the RF-pulse frequency is at the center of the spectrum to assure a uniform excitation across the spectrum. Since it is almost impossible to make the two channels in the quadrature detection exactly identical, there are sometimes quadrature-detection-related artifacts that appear in the Fourier spectrum. Phase cycling of the RF excitation and the detection in a repeated experiment is often required to eliminate such artifacts.

4.4 The Spectrometer Field-frequency Lock System

The magnetic field and the frequency must both remain constant to within at least one part in 10^{10} during the course of an NMR experiment. Some experiments such as ^{13}C acquisition, 2-D, and 3-D experiments may take several hours to several days. Magnetic fields of a superconducting magnet drift more than this (one part in 10^{10}) over a period of time, which leads to a loss of resolution in the NMR spectrum. To overcome the field drift, a field-frequency lock system is widely used in modern solution NMR spectrometers. The frequency of the deuterium (^2H) signal of the deuterated solvent of the sample is constantly monitored and reports on the magnetic-field drift. The value of the deuterium frequency is then used to ''correct'' for the field drift by applying a small additional field parallel to the main field using an RTS coil. This feedback system, which can be considered a parallel spectrometer, is often referred to as *the lock channel*. The lock channel contains a dedicated transmitter, receiver, and compensation circuit.

The choice of the ^2H signal to be that of the deuterated solvent (D_2O, $CDCl_3$, CD_3CN, DMSO, etc.) is convenient since the material is present in the sample and does not generally interfere with the spectroscopy of the commonly observed nuclei. If the nucleus of detection is ^2H, then one usually uses the NMR spectrum of some other nucleus (such as ^{19}F or ^1H) as the basis for the field-frequency lock system.

The lock system works in the following way. The absorption component of the ^2H NMR signal of the deuterated solvent is generally displayed by the computer. The dispersion component is used to form the compensation voltage that regulates the field, as shown in Figure 6. When the magnetic-field shifts, there is a change in the deuterium resonance frequency. The error signal that occurs depends on the magnitude of the frequency shift and on the direction of drift. This error signal controls a feedback circuit that adjusts the current in the B_0 shim coil in the RTS assembly to compensate for the field drift. In most cases, the deuterium NMR signal is detected by the CW method, because the CW method is simple and provides sufficient sensitivity for monitoring the frequency change of a single and strong resonance.

Since the lock channel is virtually a separate ^2H NMR spectrometer, its operation relies on careful settings of the ^2H transmitter power, the ^2H receiver gain, and the receiver phase. The highest possible lock transmitter

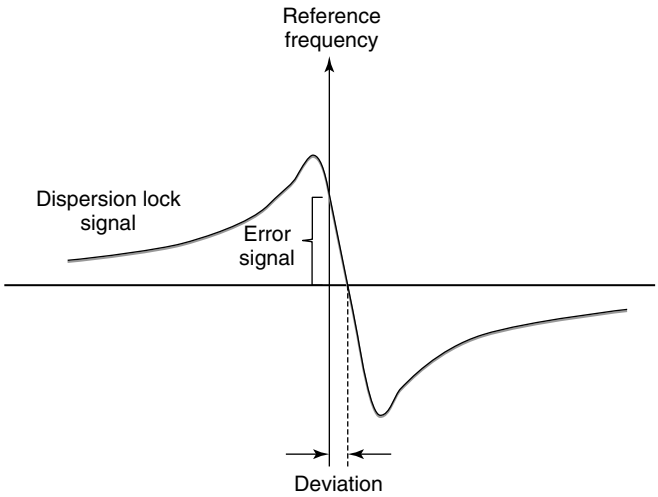

Figure 6 The sign and magnitude of error signal from deuterated solvent is used in a frequency-lock feedback circuit to compensate the field drift.

power (without saturation) ensures the highest sensitivity in the lock signal. Saturation of the ^2H through use of too high a RF power will limit the ability to detect a frequency change. The amplifier gain controls the magnitude of the lock signal. If the gain is too low, the error signal is low and the amount of signal is not sufficient. On the other hand, excessively high amplification introduces unnecessary noise that may cause the lock to suffer random jumps. The error signal also depends on the proper setting of the lock phase.

4.5 The Shim System

The typical solution-station NMR resonance line widths in solution range from a few tenths of hertz to a few hertz. To maintain a magnetic field over an active volume of about 0.3–0.7 cm^3 and over a period of hours is extremely demanding. For example, in order to resolve fine structure with a splitting of the order of 0.5 Hz in a 600 MHz spectrometer, a relative field homogeneity $\Delta B/B_0$ of better than 10^{-9} is required. Inhomogeneous magnetic fields across the sample volume cause various distortions of the line shape, which leads to poor resolution and loss of sensitivity. Besides the careful design and construction of the superconducting solenoid, the magnetic field's homogeneity is achieved by adjusting currents by two methods: cryoshimming and room-temperature shimming. The shim coils are a set of coils wound in various configurations that carry electric currents to generate small magnetic-field gradients to compensate or cancel out small gradients in the magnetic field of the large solenoid. The cryoshims are superconducting coils that are adjusted when the magnet is energized. The cryoshims remove gross inhomogeneities in the magnetic field.

It is important to remember at this juncture that the static field of the typical superconducting magnet and cryoshim system is already extremely homogeneous, often with inhomogeneities of no more than 10^{-8} over the volume of the sample. That high quality notwithstanding, to obtain the maximum performance required for high-resolution NMR, further optimization of the field homogeneity is needed. This second procedure is carried out with the RTS system. The set of room-temperature coils is mounted in the magnet bore surrounding the sample probe. The current in these coils is adjusted for each sample to compensate any remaining field gradients, specifically including gradients caused by differences from sample to sample.

The process of preparing the system to measure an NMR spectrum requires one to carry out what is called *shimming*, adjustment of the currents through the RTS coils to optimize the magnetic-field homogeneity across the sample volume, which allows one to maximize the sensitivity and resolution. Some people enjoy this iterative process, and some people find it an uncomfortable part of the setup to carry out the NMR experiment; it must be done with care to achieve a good NMR spectrum. To monitor homogeneity improvement while shimming, one uses the increase in the deuterium lock signal, or perhaps the time extent of an FID, to indicate approach to the optimal conditions. In addition, one sometimes examines the line shape of some resonance to ensure a combination of narrowness of the resonance band with a symmetric line shape. As discussed below, modern instruments are often equipped with triple-axis pulse field gradients, and shimming can be performed automatically by first mapping the magnetic-filed profile in the sample and then compensating for the field inhomogeneity with adjustments of the appropriate RTS coils.

There are two types of shim coils, axial and transverse, which are also known as *spinning* and *nonspinning* shims, respectively, to indicate the condition under which each is adjusted. Each shim coil is designed to produce a gradient profile near the center of the coil that varies as a specific mathematical function of the coordinates. The basis set functions corresponding to the individual coil geometries in almost all spectrometers are the spherical harmonic functions, the solutions of Laplace's equation.[8] The coils are therefore labeled by the Cartesian representation of the appropriate spherical harmonic function. Table 1 lists 17 of the RTSs and their corresponding labels. In some spectrometers, there may be as many as 64 RTSs that can be adjusted. The spinning shims have orientations that produce a gradient along the axis of sample rotation. They are best adjusted when the sample is spinning. The nonspinning shims are adjusted without sample rotation because they produce gradients perpendicular to the direction of the main field. Rotation would average these out, and one wishes to compensate them with the gradient coils, so the adjustment is done without spinning.

One of the most exciting developments in the NMR instrumentation in recent years is the gradient-shimming method,[9,10] which completely changes the conventional method of magnet shimming described above. Gradient shimming utilizes a gradient-echo sequence. The result of this process is a resonance phase map, which directly correlates to the field inhomogeneity across the sample. This profile is obtained experimentally and is compared with a reference phase map previously recorded on the basis of the known electric current settings of each individual shim coil, so that all required changes of current in each shim coil can be calculated. The gradient-shimming method was first applied to NMR spectroscopy of biomolecular samples that had a strong solvent (90% H_2O) signal. The strong solvent signal is used to map the inhomogeneity across the active sample volume in three dimensions. The gradient-shimming method is now

Table 1 Common room-temperature shims and their Cartesian representations

Shim	Order	Orientation		Function
Z^1	1	Axial	Spinning	z
Z^2	2	Axial	Spinning	$2z^2 - (x^2 - y^2)$
Z^3	3	Axial	Spinning	$z[2z^2 - 3(x^2 + y^2)]$
Z^4	4	Axial	Spinning	$8z^2[z^2 - 3(x^2 + y^2)] + 3(x^2 + y^2)^2$
Z^5	5	Axial	Spinning	$48z^3[z^2 - 5(x^2 + y^2)] + 90z(x^2 + y^2)^2$
X	1	Transverse	Nonspinning	x
Y	1	Transverse	Nonspinning	y
XZ	2	Transverse	Nonspinning	zx
YZ	2	Transverse	Nonspinning	zy
XY	2	Transverse	Nonspinning	$x[4z^2 - (x^2 + y^2)]$
X^2-Y^2	2	Transverse	Nonspinning	$y[4z^2 - (x^2 + y^2)]$
Z^2X	3	Transverse	Nonspinning	$x^2 - y^2$
Z^2Y	3	Transverse	Nonspinning	xy
ZXY	3	Transverse	Nonspinning	$z(x^2 - y^2)$
$Z(X^2-Y^2)$	3	Transverse	Nonspinning	xyz
X^3	3	Transverse	Nonspinning	$x(x^2 - 3y^2)$
Y^3	3	Transverse	Nonspinning	$y(3x^2 - y^2)$

available for samples in organic solvents where the ^{2}H resonance of a deuterated solvent is observed through the lock channel for inhomogeneity mapping. The accessory required for ^{2}H gradient shimming is now a standard part of a commercial spectrometer. An example of gradient shimming is shown in Figure 7, where spectrum A is taken with the current in all RTSs set to zero and spectrum B is obtained after a two-minute ^{2}H gradient shimming with only one iteration. The high efficiency of shimming demonstrated here is almost impossible to achieve by the manual iterative technique, even for experienced NMR spectroscopists.

4.6 The Pulsed-field-gradient Amplifier

The use of the PFG method has become very popular because of its applications in a wide variety of experiments, including coherence selection, solvent suppression, gradient shimming, and its use as a spoil pulse to dephase magnetization. The technique works by transiently applying a magnetic-field gradient to the sample through a gradient coil in the probe, along the z axis (single-axis gradient) or along x, y, and z axes (three-axis gradients). The gradient amplifier is operated independently from, but is always in synchrony with, the RF channels. Usually a gradient-control unit generates the gradient pulse on the basis of commands from the pulse program. Since the gradient pulse involves application of a high direct current across the gradient coil, residual eddy currents are unavoidable. Most commercial systems have self-screening to minimize the influence of these eddy currents. The shape of the gradient can be linear or rectangular, but other shapes can be programmed as well. For reduction of eddy currents, the gradient pulse shape may be a sinusoidal waveform. Like an RF pulse, a gradient pulse is specified by duration, a shape, an amplitude, and a polarity.

4.7 The Computer and Software

There are at least two computers and a number of microprocessors in a spectrometer. The primary computer provides the GUI to the spectroscopist. The computer controls spectrometer action, processes and displays, and data, and performs other tasks necessary to obtain a spectrum. The primary computer usually is a Linux workstation running Enterprise WS4 or WS5, or a PC running Windows Vista or XP. A typical workstation configuration includes 1 GB of internal memory, a 200 GB hard drive, and a 19'' (or larger) flat panel monitor, a DVD/CD-R drive, and two network cards for spectrometer and local area network (LAN) connections. With rapid developments in computer information technology, the specific characteristics of the primary computer change almost on a monthly basis. Using the GUI of the primary computer, the NMR spectroscopist selects an experiment from a list of experimental protocols, changes the parameters if necessary, and commands the computer to start the experiment. The spectroscopist's instructions are sent from the primary computer to an AC located in the console through an Ethernet connection. The AC works as an interface between the transmitter, receiver, lock channel, gradient amplifier, probe, and the primary computer. This computer does not perform onboard data manipulation and only transmits data between various spectrometer components. The communication to timing-critical units such as receiver, transmitter, digitizer, and lock channel is carried out through a high-speed virtual memory expansion (VME) data bus. The Ethernet link

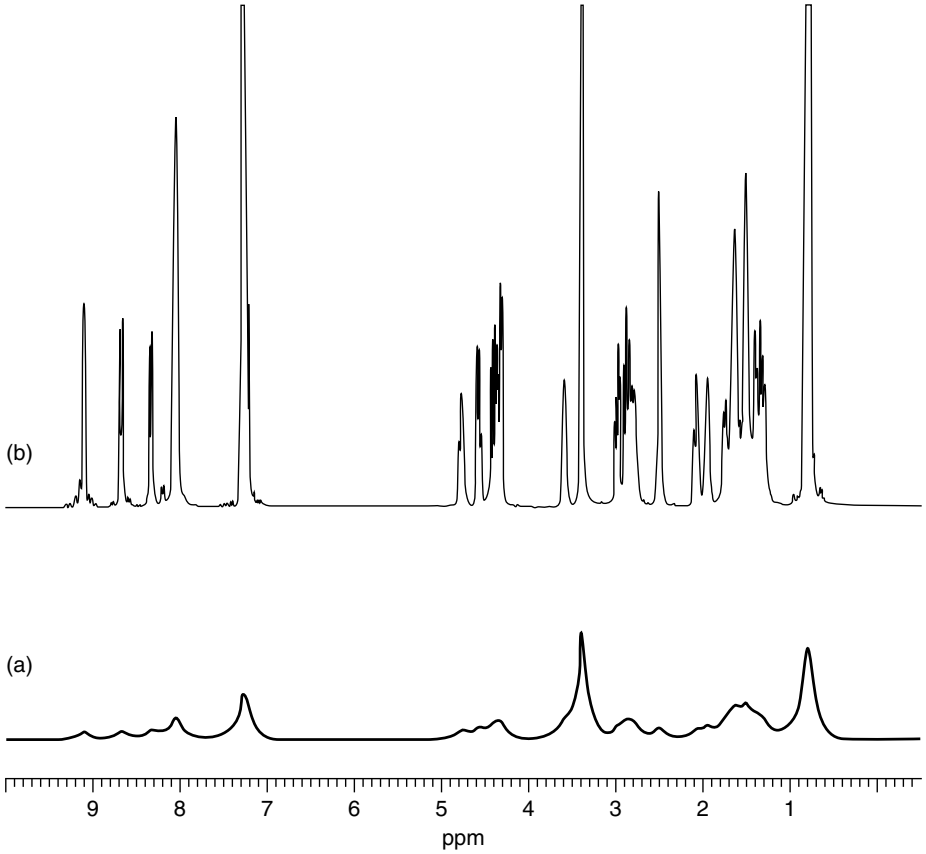

Figure 7 The gradient-shimming method: ^{1}H spectrum of gramicidine-S in DMSO-d6. The bottom trace (a) was obtained by setting all on-axis shims to zero. After a deuterated gradient shimming of z1–z5 for a few minutes, a high-resolution spectrum (b) was obtained.

to the primary computer provides an effective connection by which the commands entered at the operator's desk are transmitted to the AC and hence to various spectrometer components. Non-timing-critical units such as the variable temperature control, the pneumatic unit for probe control, and the automatic sample changer are connected to the spectrometer computer through an RS-232 or RS-485 bus.

Acquired and digitized data are transferred to the primary computer for storage and further processing. Data are saved on a hard drive and may be transferred to a file server through an Ethernet connection. The time-domain data are generally transformed into a frequency-domain spectrum by NMR software resident in the primary computer. A local printer or a network printer can be used to plot the spectrum obtained through this process.

NMR software has several major functions: spectrometer control, data acquisition, and data processing. Most NMR software code is written in the C and C++ programming languages, but may include recent Java and Tcl/Tk add-ons on a Linux- or a Windows-based PC computer. The software packages have an extensive microprogramming language for operators to develop automated operations, to control the spectrometer, and to process the data. The NMR software package is usually accessed through the graphical interface. NMR software usually presents a large display of the current NMR data in either the time or the frequency domain. The program subroutines can be accessed through drop-down menus with many levels of commands that handle tasks from data acquisition to processing. Most commands can also be entered through a command line. There are also control icons for quick access to many routinely used procedures. Interactive operations, such as spectral display, phasing, peak picking, and integration are done by convenient mouse operations.

Automation of operations is an important convenient feature of recently released NMR software. This is especially useful for laboratories where minimally trained users submit samples. The user places the NMR sample on a sample changer carousel and selects the NMR experiments to be carried out through the GUI of the spectrometer's primary computer. The sample is automatically moved, at the appropriate time, from the carousel to the NMR probe by the sample changer, and

the experiment preparation and execution are carried out by the NMR software with basically no intervention from the user. After data acquisition and reduction, a notification of the experiment, usually including the processed NMR spectrum as a portable document format (PDF) file, is sent to the user through an email. The system administrator may choose to archive acquired data to a network data server through a network file system (NFS) protocol or other Internet file-transfer protocols.

A database is used to manage all pulses and other experimental parameters, including durations and power levels for RF pulses for each probe. These parameters can be easily accessed when a new experimental protocol is to be parameterized. As a part of the software package, there are many modules to perform specific functions, such as nuclear relaxation analysis, analysis of shaped pulses, spectral simulation, and experimental simulation, as well as graphic display of the pulse program. A molecular drawing program is even included in some NMR packages for users to document molecular structures for each spectrum. Recently released software also provides web access, which allows users to check the status of data acquisition and to control the spectrometer through the Internet. In addition, there are frequently subroutines for calculation of theoretical spectra or fitting the spectra to some form. NMR software continues to be a fast-developing component of most modern NMR spectrometers, a result of the rapid advances in computer technology and programming.

4.8 The Nuclear Magnetic Resonance Probe

The NMR probe, housing the sample for analysis, transmits RF radiation to the sample and detects the weak NMR response. Most solution NMR probes are mounted in the magnet bore from the bottom of the magnet. For insertion of the sample, a spinner loaded with the NMR sample tube is placed into the probe from the top of the magnet bore. A pneumatic device supplies a cushion of air or nitrogen gas that buoys the spinner/sample combination, allows it to be slowly positioned into the probe as the pressure is reduced, and finally ejects the spinner when the experiment is finished. The sample tube, surrounded by the RF coils, is usually spun at 15–20 Hz to reduce residual magnetic-field inhomogeneities. A simplified schematic diagram of the NMR probe and sample spinner is given in Figure 8.

In general, the sample temperature in the NMR probe can be controlled from -100 to $+150\,^{\circ}\mathrm{C}$ to within $\pm 0.1\,^{\circ}\mathrm{C}$. If the desired sample temperature is higher than room temperature, heated air or nitrogen gas is passed over the sample tube. Cold nitrogen gas generated by a heater immersed in a liquid nitrogen dewar is used as a thermally controlled bath for the sample if the desired sample temperature is below room temperature. Recently, various gas refrigeration devices have become available commercially for cooling nitrogen gas before it is introduced into the probe for temperature regulation at or slightly below room temperature.

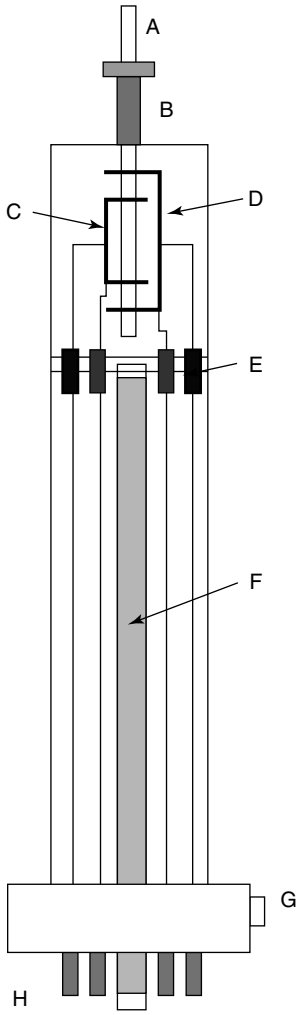

Figure 8 Schematic diagram of NMR probe: A, sample tube; B, sample spinner; C, inner RF coil; D, outer RF coil; E, variable capacitors for probe LC circuit frequency tuning; F, variable temperature dewar and heater; G, RF connectors for transmitting RF pulses and receiving NMR signals; H, knobs for probe tuning.

A simplified probe (LC) circuit, which consists of an RF coil having inductance L and two variable capacitors, is given in Figure 9. Optimal performance of the NMR spectrometer depends critically on detection sensitivity, which is achieved by adjusting the tuning capacitor C_{T} and matching capacitor C_{M} such that the resonance frequency of the probe LC circuit is the nuclear Larmor frequency and the impedance of the circuit is matched to 50 Ω.

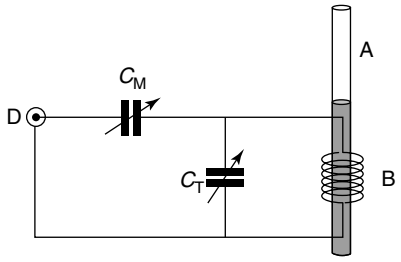

Figure 9 Simplified probe LC circuit: A, NMR sample tube; B, RF coil; C_T, variable capacitor for tuning resonance frequency; C_M, variable capacitor for matching the impedance of the LC circuit; D, BNC coaxial connector to the transmitter and receiver.

In general, insertion of a sample changes the tuning of the probe circuit. Samples containing a high-dielectric material (such as water) may particularly cause a change in probe tuning. Samples that are electrically conducting cause significant changes in both matching and tuning. Therefore, it is imperative to tune and match the probe for each sample. As shown in Figure 8, conventional probe tuning and matching knobs that control the capacitor settings are located at the bottom of the probe. To tune the probe circuit at the appropriate frequency, the response of an RF channel to an imposed RF voltage is compared to the response of a purely resistive 50 Ω load. The RF voltage is generally swept through a range of frequencies as it originates from a sweep generator. Adjustment of the two capacitors, usually by iterative changes, produces a situation in which the RF response of the circuit appears identical to that of the purely resistive load. To aid in this tuning step, the reflected voltage from the circuit is usually displayed through a GUI at the primary computer display. Sometimes, the response is sent to an array of light-emitting diodes (LED), which report the reflected voltage as a string of lit diodes. Whatever the device for detection, the object of the tuning step is to minimize reflected RF voltage from the probe RF circuit.

Tuning the probe circuit requires a set of skills of the user. It can be a time-consuming and frustrating process for one who has not experienced the process. A recent development in the NMR probe technology is ATM by a device that carries out this procedure automatically under software control. The advantage of ATM is especially obvious in an open-access NMR laboratory, where many users are untrained or minimally trained in tuning RF circuits.

Various high-resolution probes for study of solution-state samples are commercially available. Each type of probe meets some requirements for specific applications in chemistry, biochemistry, and material sciences. However, many features are common to all probes, which makes it possible to classify probes into well-defined groups according to the particular values of certain parameters. For example, commercial probes are determined by the sample tube diameter. Currently, one may purchase probes that accept tubes with diameters of 1.7, 2.5, 5, 8, 10, and 20 mm. The choice of diameter depends on the application. Probes with a diameter of less than 3 mm are called *microsample probes* and are used in situations where the amount of sample is limited. The high efficiency of a microsample probe is a result of a high filling factor due to the limited volume of the sample tube. The most commonly used NMR probes accept 5-mm-diameter NMR tubes. Probes for larger tubes can be employed for special applications, where the sensitivity of the measurement is limited.

Besides the sample tube diameter, solution NMR probes can be classified as *standard* or *inverse* based on the configuration for signal detection and for decoupling. A typical standard high-resolution probe has at least two RF coils. The inner coil is for the X-nucleus detection, while the outer coil is mainly used for ^{1}H decoupling. The outer coil is sometimes used for observation of a nuclear signal as well, but this almost always involves a lower sensitivity. The outer (or ^{1}H) coil is usually doubly tuned to allow it to be used to detect the ^{2}H resonance for use in the field-frequency lock system. In some probes, the X channel is tuned only at a fixed frequency (e.g. ^{13}C). However, there are so-called broadband probes in which the frequency may be tuned over a range to allow detection of a variety of nuclei. A typical configuration for a broadband probe might allow detection of nuclei that have frequencies in the range from ^{109}Ag to ^{31}P. The standard probe is generally optimized to observe an X nucleus, normally ^{13}C, while allowing decoupling of protons. Other standard probes may be tuned for multiple resonances in the X channel, for example by throwing a switch. One such probe is the so-called quad nuclei probe that has an inner coil tuned to predefined nuclei, usually ^{31}P, ^{13}C, and ^{15}N, or ^{19}F, ^{31}P, and ^{13}C. In such probes, the outer coil is doubly tuned to ^{1}H and ^{2}H.

An inverse probe reverses the uses of the inner and outer coils. The inner coil is tuned to the ^{1}H resonance frequency and the outer coil is used for the X-nucleus frequency (^{13}C, ^{15}N, ^{31}P, etc.). Such probes are often used when one wishes to detect the ^{1}H resonance while applying decoupling power to the X nucleus. Because of the high sensitivity of ^{1}H detection, inverse probes have been widely used for indirect-detection 2-D experiments such as heteronuclear multiple quantum correlation (HMQC) and heteronuclear multiple band correlation (HMBC). Inverse probes are also optimized for 2-D ^{1}H–^{1}H correlation experiments such as the correlated spectroscopy (COSY) experiment and the nuclear Overhauser effect spectroscopy (NOESY) experiment. Most triple-resonance probes (that can be used for detection or decoupling of three different nuclei in an experiment) are of the inverse design, with an inner coil

tuned to ^{1}H resonance frequency and an outer coil doubly tuned for simultaneous decoupling two nuclei, often ^{13}C and ^{15}N.

Another aspect of probe technology is the incorporation of coils that allow the generation of PFGs. For most of multiple-dimensional experiments it is necessary to apply a gradient to the main field for a short time, which is accommodated by these coils. The gradient coils are actively shielded, with stray magnetic fields screened by a second coil located around the principal gradient coil. Some probes contain a single z-gradient coil that provides a linear gradient in the direction of the main magnetic field. For more complex experiments, one may also purchase probes that contain an assembly of gradient coils that independently generate linear gradients in each of the three orthogonal directions x, y, and z to allow a succession of different gradients that are required in some sophisticated experiments.

For special experimental situations such as LC-NMR, the probe has to be designed to accommodate the needs of the other procedure. In particular, in LC-NMR the sample flows through a capillary tube, the active volume of which is only microliters. The inverse mode is often used for LC probe to provide enhanced sensitivity for ^{1}H observation. Inverse operation is necessitated because of the small amount of sample. Another kind of probe used widely in industrial laboratories is a high-throughput flow probe, similar in concept to the probe used in LC-NMR. In these probes, a capillary tube is used as a sample tube with a volume of a few hundred microliters.

One of the most exiting developments in probe technology in recent years is the cryogenically cooled probe. Such a probe was proposed decades ago[11,12] but it has only recently become commercially available. The cryogenically cooled probe offers a dramatic increase in the S/N by reducing the operating temperature of the NMR coil assembly and preamplifiers to 25–30 and 77 K, respectively, which reduces the inherent noise in the RF circuits. The probe coils, a few millimeters away from the sample tube, which is usually near room temperature, are cooled with cold helium gas that is controlled by an automatic closed-cycle cooling system. The cryogenically cooled probe can be configured as a standard probe where the inner coil is for X detection or as an inverse probe. In the standard configuration, the cooled probe can be used for collecting natural-abundance ^{15}N spectra and even for acquiring ^{13}C spectra of high-molecular-weight polymers where the sample is maintained at high temperature. The ability to insulate the region of the RF coils from a very different temperature at the sample (which is only millimeters away) is one of the triumphs of the design of NMR technology. Triple-resonance cryogenic probes operated in the inverse mode have been extensively used for analysis of proteins, as well as for other biomolecular analyses. It appears that, the high cost and demanding requirements for laboratory infrastructure notwithstanding, cryogenically cooled probes are becoming mainstays of industrial, academic, and government NMR laboratories.

4.9 Automation and Sample Changers

In the modern NMR laboratory, more and more work is performed automatically. While all of the operations of the spectroscopy have been under computer control for many years, it is only recently that the initial step of selecting and inserting a sample has become a fully automated part of the process of acquiring an NMR spectrum. The sample changer is, at its heart, a robot that selects a specific sample by command of the computer and places it in the NMR spectrometer; at the end of an acquisition, it reverses this process by ejecting the sample. The choice of a particular sample changer depends on the number of samples to be analyzed. A small sample changer that can handle from 8 to 24 samples is typically mounted on top of the magnet and samples are analyzed sequentially. A larger sample changer (handling up to 120 samples) is often floor-mounted and uses pneumatic or magnet-mounted arms to move the sample between the magnet and sample carousel. NMR software controls the queuing system and determines when a sample is analyzed on the basis of queuing rules preprogrammed by the system administrator.

5 PRACTICAL ASPECTS OF NUCLEAR MAGNETIC RESONANCE SPECTROMETER OPERATION

To obtain a high-quality spectrum, the sample must be prepared carefully. The appropriate NMR probe and experiments must be chosen for the specific application. The standard experimental procedures must be followed to assure reproducibility and accuracy of measurements. As with any experimental procedure, safety precautions must be observed. In this section, practical details of procedures and precautions are discussed.

5.1 Safety Precautions

Extreme caution must be taken while working around magnets, particularly when magnetic materials such as gas cylinders and tools are used in the vicinity of the magnets. Persons with medical metal implants must consult a physician before working around magnets. Credit cards and other magnetic recording materials must be kept away from the magnets.

5.2 Sample Preparation

Sample preparation is the first step, and a critical one, if one is to obtain a high-quality NMR spectrum. Usually the sample is dissolved in a deuterated solvent that serves at the same time as the origin of the lock signal. A list of common NMR solvents is given in Table 2, along with chemical shift references and other physical properties. Nonviscous solvents provide a sharp NMR resonance and are always the preferred choice for use in preparing a sample. If a solvent is viscous but necessary to the chemistry, NMR spectra may be acquired at an elevated temperature to increase the mobility of the solute and solvent molecules to improve the resolution. It is preferable to avoid the overlap of NMR signals from the residual protonated solvent molecule and the sample signal, if possible. If temperature is a variable in the NMR measurement, boiling point and melting point of the solvent must be taken into consideration. If a large amount of the solvent is to be used for NMR analysis, then cost may also be a factor in the solvent selection. Deuterated chloroform is, by far, the most widely used solvent for studies involving organic molecules. Deuterated acetone, because of its low viscosity and very sharp resonance, is often used as the standard sample for establishing homogeneity of the magnetic field. Because of its ability to dissolve many organic materials, deuterated dimethylsulfoxide (DMSO) has also become popular as a solvent for NMR studies. In studies of biological macromolecular systems, detection of amide and imido protons is critical to structure determination with NMR. A mixture of 90% H_2O with 10% D_2O is often used as the solvent in these studies to limit deuterium exchange from D_2O to backbone amide NH protons in peptides and proteins or to imido protons of DNA and RNA base pairs. Because H_2O gives a large signal in the proton spectrum of these solutions, solvent-suppression techniques to reduce the intensity of the water signal have been developed.

For obtaining a ^1H NMR spectrum, the sample concentration should be kept at moderate or low levels to reduce spectral artifacts, such as spinning sidebands. For a ^{13}C analysis in which the natural abundance limits the signal strength, a high sample concentration is preferable.

Table 2 Common NMR solvents[13]

Solvent (formula)	^1H chemical shift (ppm) (multiplicity)	^{13}C chemical shift (ppm) (multiplicity)	Melting point (°C)	Boiling point (°C)
Acetic acid-d_4 (CD_3COOD)	11.65(1) 2.04(5)	178.99(1) 20.0(7)	16.7	118
Acetone-d_6 (CD_3COCD_3)	2.05(5)	206.68(1) 29.9(7)	−94	56.5
Acetonitrile-d_3 (CD_3CN)	1.94(5)	118.69(1) 1.39(7)	−45	81.6
Benzene-d_6 (C_6D_6)	7.16(1)	128.39(3)	5.5	80.1
Chloroform-d_1 ($CDCl_3$)	7.24(1)	77.23(3)	−63.5	61.62
Deuterium oxide-d_2 (D_2O)	4.80(1)		3.81	101.42
Dichloromethane-d_2 (CD_2Cl_2)	5.32(3)	54.0(5)	−95	39.75
N, *N*-Dimethyl formamide-d_7 ($(CD_3)_2NCDO$)	8.03(1) 2.92(5) 2.75(5)	163.15(3) 34.89(7) 29.76(7)	−61	153
Dimethylsulfoxide-d_6 (CD_3SOCD_3)	2.50(5)	39.51(7)	18.45	189
Methanol-d_4 (CD_3OD)	4.78(1) 3.31(5)	49.15(7)	−97.8	64.7
Pyridine-d_5 (C_5D_5N)	8.74(1) 7.58(1) 7.22(1)	150.35(3) 135.91(3) 123.87(3)	−42	115 −116
Tetrahydrofuran-d_8 (C_4D_8O)	3.58(1) 1.73(1)	67.57(5) 25.37(5)	−108.5	66
Toluene-d_8 ($C_6D_5CD_3$)	7.09(m) 7.00(1) 6.98(5) 2.09(5)	137.86(1) 129.24(3) 128.33(3) 125.49(3) 20.4(7)	−95	110.6
Trifluoroacetic acid-d_1 (CF_3COOD)	11.50(1)	164.2(4) 116.6(4)	−15.4	72.4
Trifluoroethanol-d_3 (CF_3CD_2OD)	5.02(1) 3.88(4 × 3)	126.3(4) 61.5(4 × 5)	−43.3	75

There are several features of solution preparation that are necessary to ensure a high-quality spectrum. Insoluble particles suspended in the solution to be analyzed lead to a situation in which the NMR resonances are often poorly resolved. Generally, careful filtration of the solution is recommended. Gases, especially oxygen, dissolved in the solution provide additional nuclear spin relaxation pathways for the spins in the analyte, with the net effect that the resonances may be broader than one might expect. The result is that the resonances are less well resolved. Degassing the sample by the freeze-pump-thaw method can generally reduce the amount of dissolved gas in a sample. It is especially important to carry out this removal of dissolved oxygen if the experiment involves measuring the intrinsic relaxation rate for the analyte. As an internal standard of the chemical shift, a trace amount of tetramethylsilane (TMS, 0.0 ppm) is added to a sample as an internal chemical shift reference. Sometimes, the residual proton signal of a deuterated solvent may also be used as a secondary chemical shift reference, rather than adding TMS. The ^{1}H and ^{13}C chemical shifts with respect to TMS of some common NMR solvent are listed in Table 2.

5.3 Data Acquisition

After a sample is loaded into a probe, acquisition parameters, including observe and decoupling frequencies, are loaded into the data directory of the current experiment by a command to the NMR-control software. If the sample is to be examined at a temperature other than room temperature, the temperature controller is adjusted and a sufficient time for temperature equilibration must be allowed. It is important to note that the temperature that is reported by the hardware is not the temperature at the sample, and the temperature must be determined by some previous calibration of the system.

The probe must be tuned to the frequencies specified by the acquisition parameters. In addition to sample properties, temperature is a variable that may affect the tuning, so it is important to allow the probe to equilibrate before carrying out tuning. The tuning and matching capacitors of the probe LC circuit are adjusted iteratively until the circuit is resonant at the appropriate frequency and its impedance is matched to that of the spectrometer's RF output.

To average out inhomogeneities across the active volume, the sample is spun at 15–20 Hz. To begin, the ^{2}H signal of the deuterated NMR solvent is located by adjusting magnetic-field offset while keeping the deuterium frequency constant. Once the lock signal is found, the magnetic field is locked with the solvent deuterium signal on resonance. After the frequency lock has been established, the magnetic-field homogeneity in the active sample volume is optimized by shimming. Shimming is carried out iteratively. The procedure begins by adjusting the spinning shims iteratively until no improvement in the lock level is possible. The sample spinning is stopped, and the nonspinning shims are adjusted iteratively until the optimal condition is achieved. Although these two steps sound straightforward, the iteration among a large number of shims takes some time, unless performed automatically. With proper room-temperature shimming, the ^{1}H line width of typical small organic molecules such as $CHCl_3$ can be improved from a few hundred hertz in a cryoshimmed magnet to a few tenths of a hertz. For a routine operation in a properly maintained spectrometer, shimming generally only involves adjusting the first- and/or the second-order shims when the sample is changed. However, for a new probe or spectrometer, shimming to a proper line resolution and line shape may take hours or even days.

Before beginning the acquisition of data, one must choose the appropriate pulse program and enter the appropriate acquisition parameters, including pulse widths, related power levels, SW, number of data points to be acquired (time domain, TD), DW, and number of scans (NS) to be coadded. In some cases, one may obtain the appropriate set of acquisition parameters from files saved on the computer, particularly for routine ^{1}H and ^{13}C experiments. Calibration of pulse widths and amplitudes must be done routinely and properly documented.

Here we discuss some of the parameters that must be adjusted before beginning an experiment with quadrature detection. In quadrature detection, the transmitter frequency should be placed at the center of the SW such that only a frequency range of $\pm$SW/2 needs to be digitized. The data-sampling rate, dictated by the Nyquist criterion, must be larger than or equal to SW (i.e. $2 \times (SW/2)$). By this requirement, the relation of the sweep width to the data-sampling time interval, or DW, is as follows:

$$DW = \frac{1}{SW} \tag{4}$$

The total acquisition time (AT) is simply the product of the dwell time and half of the time-domain data points (TD/2), which provides the following relation:

$$AT = DW\frac{TD}{2} = \frac{1}{SW}\frac{TD}{2} \tag{5}$$

The digital resolution (DR) of the spectrum is defined as the ratio of total SW in hertz to the total number of points in the spectrum. If zero filling is not applied to the time-domain data set before FT, the number of points in the frequency-domain spectrum is one half of TD or

TD/2. The DR can then be expressed as

$$DR = \frac{SW}{(TD/2)} = \frac{1}{AT} \tag{6}$$

As shown in Equation (6), digital resolution is simply the reciprocal of the total acquisition time. DR should always be smaller than the natural line width of the NMR resonances of the sample to assure the ability to resolve fine structure. For a typical ^{1}H spectrum, the resonance line width ranges from a few tenths of a hertz to a few hertz, and the AT must be set to a few seconds or longer. Prior to data acquisition, the receiver gain should be adjusted to prevent the signal from being too large or too small for the detection; it is particularly important that the gain be not so big as to produce signals that saturate the ADC converter, a situation that introduces artifacts into the spectrum. The acquisition is generally terminated when the number of FIDs requested by the operator have been added together, given the symbol NS or NA. Most NMR-control programs allow the spectrometer operator to terminate acquisition for any reason at any time before NS is reached.

5.4 Data Processing and Reduction

Upon the completion of data acquisition, the FID is generally apodized by multiplication with a function defined by the operator. This action is usually done to improve either the S/N or the resolution. To ensure that zero-frequency artifacts are minimized at this point, the average of the data over some region is subtracted from each point to ensure that the oscillation is about the zero of the data. After manipulation of the data by these procedures, the data are ready to analyze. The FFT is used to convert the time-domain FID to a frequency-domain spectrum. Subsequent to this procedure, the NMR spectrum may require a phase correction, which can be carried out with a subroutine. In addition, the spectrum may be further treated to ''flatten the baseline''. Here a standard protocol is used to determine the low-frequency broad components of the spectrum, often by fitting certain parts of the spectrum and subtracting, in a point-by-point fashion, to ''remove'' these components from the spectrum.

An important part of presenting a spectrum is ensuring that the data are properly referenced. This referencing is frequently done by choosing the peak representative of a particular chemical shift (0.0 ppm for TMS) and issuing a command to the computer to set all frequencies relative to its frequency, the reference position. If the TMS signal is absent from the spectrum, one sometimes uses the residual ^{1}H signals of the deuterated solvent as a secondary chemical shift reference.

At this point, one usually wishes to identify all of the resonances in the spectrum by specifying their positions relative to the reference. This can be done manually, but there are computer programs that can also perform this operation. In the semiautomatic mode, the spectrometer operator sets the minimum and maximum limits for the program to accept as a peak. The process of identifying the resonances in this manner is called *peak picking*. Most NMR software automatically labels the chemical shift of each peak found between the limits defined by user with its chemical shift.

An important quality is the relative amount of signal corresponding to each peak. This is found by integration of the peak. The spectrometer operator may use a software function to select spectral regions for automatic calculation of the peak integrals. These are displayed on the graphical presentation of the spectrum and are frequently listed in a table generated by the software.

At this point, the only remaining procedure is to make a hard-copy version of the spectrum. Most NMR software provides a plotting facility to do this. Depending on the needs of the spectroscopist, one may plot different regions of the spectrum to demonstrate some particularly important aspect of the data. The format of the plot depends on the kind of experiment. For multidimensional experiments, the plotting may be quite complicated. For example, for 2-D experiments, the ''forest'' style plot that was prevalent in early versions of NMR software has now been replaced almost completely by the contour plot.

The operating procedures discussed above are summarized in Figure 10. All steps can be automated in an

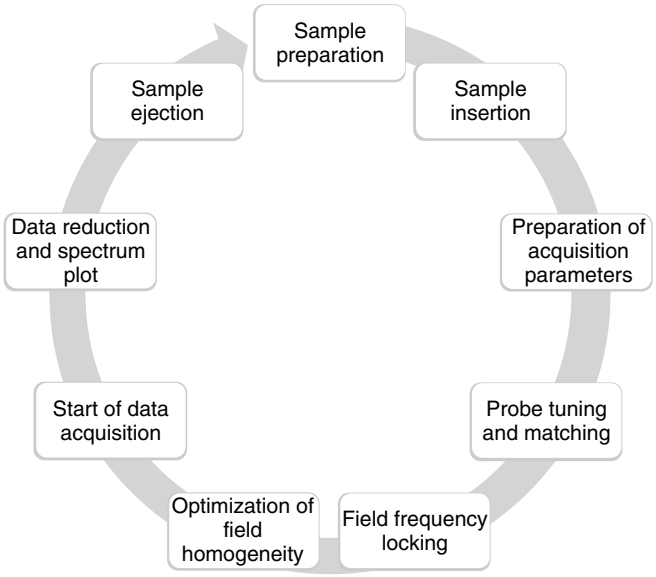

Figure 10 A typical flow chart for a routine NMR operation.

open-access NMR laboratory. It is worthwhile to mention that the automation procedures are suitable to a direct-injection autosampler system. Samples are extracted from vials or well plates with a needle that can be injected into a capillary tube. The capillary tube allows flow into the NMR-active region, where the sample is analyzed. The typical probe used in this flow injection autosampler system is an ATM capillary-flow probe.

6 FUTURE PERSPECTIVES IN THE DEVELOPMENT OF NUCLEAR MAGNETIC RESONANCE SPECTROMETERS

As the NMR spectrometer is one of the critical tools used in industrial, academic, and government laboratories, each development that improves the NMR spectrometer has made a significant impact on the chemical, pharmaceutical, and biotechnology industries. There is no doubt that this trend will continue. Newly introduced methods and techniques have been quickly applied to all areas of NMR instrumentation, whether they be developments affecting the magnet, the probe, the console, the computer hardware, or the NMR software. In general, the developments have focused on enhancement of sensitivity and resolution, automation of the spectrometer operation, reduction of space required for the NMR system, enhancing the reliability of components, mating NMR with other technologies to provide synergy, and ease of operation by less-experienced users.

The driving force to produce ever-higher magnetic field is the enhancement of sensitivity and resolution, especially for the analysis of biological macromolecular samples in solution or in solids. In the solid state, quadrupolar coupling, which broadens the NMR line of nuclei possessing a spin greater than 1/2 (two-thirds of the elements in the periodic table), can be made nearly first order in sufficient high fields. Thus, there is a tremendous emphasis in both solid- and solution-state NMR to provide reliable magnets that have higher fields, such as the recently introduced 950 MHz spectrometer. This trend will not slow down in the future.

The introduction of actively shielded magnets has led to less need for laboratory space. For example, the 5-G line of an actively shielded Bruker BioSpin 800 MHz magnet is about 1.5 m, which is comparable to the stray field footprint of an unshielded 300 MHz magnet. This means a great reduction in the space required for NMR spectrometers. It is certain that there will be continual pressure to replace older magnets with modern shielded magnets.

The conservation of cryogenic liquids (nitrogen and helium) is another driving force in the development of magnet technology, particularly as these liquids become harder to obtain and therefore more expensive. Several commercial systems exist for this purpose. For example, using the excess cooling power of a cooling unit for a cryogenically cooled probe, nitrogen gas can be condensed and introduced back into the nitrogen chamber of the magnet for reuse. With this technique, it is possible to eliminate or reduce the number of liquid nitrogen refills, reduce cryogen costs, and increase the flexibility for performing long-term experiments by eliminating interruptions for nitrogen refills.

Considering that helium is a nonrenewable resource, it is anticipated, not too far in the future, that similar techniques will be employed to cool down helium gas from the boil-off to help maintain the temperature of the superconducting solenoid. Even now, it has become economically feasible and socially important to collect helium boil-off gas and return the gas, rather than allow it to escape into space.

A great potential saving may accrue if the new technologies in high-temperature superconducting materials can be used to replace the current superconducting materials that require cooling to liquid helium temperatures with materials that only require cooling to liquid nitrogen temperatures. There are still many technological hurdles to overcome before this becomes reality, and there might be some technological reasons that it cannot be done, but if such a magnet could be developed, it would have a tremendous impact on the way in which NMR spectroscopy is done.

With the rapid development of digital electronics, much of the spectrometer control and operation have been consolidated onto a single computer motherboard. For example, identical peripheral component interconnect (PCI) cards perform individual functionalities, such as timing, frequency, and gradient controlling, depending on a preloaded firmware. The communication backbone of the NMR spectrometer now consists of two data paths, a LAN-based approach (Ethernet) for non-time-critical operations (diagnostics, firmware upgrades, and configuration parameters) and an ultrafast low-voltage differential signaling (LVDS) bus for time-critical (real-time) applications (pulse timing, loop, and other real-time decisions). The resolutions of frequency and timing control have been greatly improved with LVDS communications. Special ADC architectures have been developed in which the dynamic range of the digitizer is expanded from the previous 16 to 22 bits. In addition, receiver bandwidth has been expanded to as much as 5 MHz. Both of these developments benefit areas such as polymer analysis, solid-state applications, and impurity detection in mixtures.

With wide-bandwidth receivers, a recent technique of parallel detection of multiple nuclei[14] allows simultaneous recording of multiple-dimension NMR spectra for more than one nucleus. This method reduces the total time required for data collection per sample. A further development by Varian of a receiver with a single 80 MHz ADC to digitize the NMR signal directly at the intermediate frequency provides spectra quickly with fewer steps: these spectra have flat baselines and are free of quadrature detection–related artifacts. This new capability allows single-scan operation with the associated higher throughput when sensitivity is not an issue. It is anticipated that more and more ultrafast ADC will become available and be used in an NMR spectrometer.

To obtain information on the absolute concentration of individual components of a mixture with ^{1}H NMR spectroscopy requires the addition of an external reference compound that may overlap some of the resonance of the material to be analyzed. A new technique[15] employs an artificial electronically synthesized reference signal, which can be tuned in terms of chemical shift displacement and amplitude to suit the particular constraints of the experiments. This effectively replaces the external reference and avoids contaminating the sample.

The most active area in NMR instrumentation development in recent years has been probe development. In particular, cryogenically cooled probes have begun to be utilized in a wide variety of applications, not only in complex biological systems but also in small molecule structure elucidation in organic and inorganic chemistry. The principal advantage of a cryogenically cooled probe is the tremendous increase in S/N, which makes possible a substantial reduction in total time required for data collection, in some cases by up to a factor of 16 when compared with a conventional probe. The cryogenically cooled probes also allow the detection of ^{1}H and ^{13}C NMR spectra of extremely small ($\sim$30 μL) samples often found in studies in drug research in the pharmaceutical industry. For the same reason, cryogenically cooled probes have a great potential in the NMR microimaging probes.

In solid-state NMR, similar probe developments have provided new arenas for experimentalists. The development of the ultrafast MAS probes that spin up to 70 kHz has been offered improved spectral resolutions for ^{1}H and has eliminated the need of high-power decoupling. In biosolid NMR applications, RF heating in the sample chamber due to electric fields (E-fields) of the RF pulses severely affects the lifetime of the biosolid sample under MAS conditions.[16,17] New coil designs from Bruker and Varian reduce or eliminate RF heating caused by the E-field.

NMR software packages have been continuously updated to be consistent with advances in hardware. Most NMR software is designed for operation under Windows and Linux operating systems. The software generally has a straightforward interface that takes advantage of widespread PC standards that are commonly used in word processing, graphics, and presentation programs. Recent versions of NMR software keep track of NMR experimental parameters for each probe by retrieving information stored in a microchip in the probe. With these new versions, the RF power is monitored under software control to ensure that power does not exceed the limit for each probe; if this happens, the software terminates the experiment to prevent damage to the probe.

Another innovation is web-based access to the spectrometer through a new generation of software. With this software, one may monitor the status of an experiment or even start and terminate an experiment from a remote computer. With further development of this sort of software control, an operator will soon be able to access and control the NMR spectrometer through a cellular phone or other mobile communication device. A recent software development is the integration of accounting programs into the software to track usage for administrative purposes.

The developments in software, from remote control to automation, have changed the concept of spectrometer operation. For example, the tedious work associated with magnetic-field shimming has been greatly reduced with the utilization the gradient-echo technique. Using a new algorithm available in the most recently released software by Bruker, it takes less than 1 h for a spectrometer operator to produce a homogeneous magnetic field that meets the resolution and line-shape requirements, started from a newly installed magnet. Without this spectrometer automation, the same process may have taken a few days for an experienced console engineer to accomplish when this was done manually.

One should expect that integrated analytical systems involving hyphenated techniques that involve NMR will continue to grow in importance. The combinations of LC with NMR and of LC-MS with NMR are now popular options for an analytical chemist. Integrated software that interprets NMR spectra and correlates the structural features from various analytical techniques is the next logical point for development as these hyphenated techniques become more common in the analytical laboratory.

ACKNOWLEDGMENTS

Cecil Dybowski has helped by reading the initial manuscript and providing suggestions. Mark Chaykowski of Bruker BioSpin and Daniel Barabino of Varian, Inc. have provided NMR pictures and copyrighted materials that have greatly helped in the development of this article.

ABBREVIATIONS AND ACRONYMS

AC	Acquisition Computer
ADC	Analog-to-Digital Converter
AT	Acquisition Time
ATM	Automatic Tuning and Matching
BNC	Bayonet-Neill-Concelman Connector
COSY	Correlated Spectroscopy
CW	Continuous Wave
DAC	Digital-to-Analog Converter
DR	Digital Resolution
DW	Dwell Time
DMSO	Dimethylsulfoxide
E-Field	Electric Field
FFT	Fast Fourier Transform
FID	Free Induction Decay
FT	Fourier Transform
GUI	Graphic User Interface
HMBC	Heteronuclear Multiple Band Correlation
HMQC	Heteronuclear Multiple Quantum Correlation
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
LAN	Local Area Network
LC	Liquid Chromatography
LC-MS	Liquid Chromatography/Mass Spectrometry
LC-NMR	Liquid Chromatography/Nuclear Magnetic Resonance
LC-NMR/MS	Liquid Chromatography/Nuclear Magnetic Resonance/ Mass Spectrometry
LED	Light-emitting Diode
LVDS	Low-voltage Differential Signaling
MAS	Magic-angle Spinning
MRI	Magnetic Resonance Image
MS	Mass Spectrometry
NFS	Network File System
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NS	Number of Scans
PC	Personal Computer
PCI	Peripheral Component Interconnect
PDF	Portable Document Format
PFG	Pulsed Field Gradient
Q	Quality
RF	Radio Frequency
RTS	Room-temperature Shim
S/N	Signal-to-noise ratio
SPE	Solid Phase Extraction
SW	Spectral Width
TD	Time Domain
TMS	Tetramethylsilane
TOF	Time-of-flight
VME	Virtual Memory Extension
WB	Wide bore
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Biomedical Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Glycoprotein Analysis: Using Nuclear Magnetic Resonance

Verification of Chemicals Related to the Chemical Weapons Convention **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Coating Analysis **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Environmental Analysis of Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring

Food Analysis Techniques **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Hydrocarbons Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Process Analysis **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Steel and Related Materials **(Volume 10)**
Nuclear Magnetic Resonance in Metals Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. F. Bloch, W.W. Hansen, M. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
2. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
3. J.W. Emsley, J. Feeney, 'Forty Years of Progress in Nuclear Magnetic Resonance Spectroscopy', *Prog. in NMR Spectrosc.*, **50**, 179–198 (2007).
4. I.J. Lowe, R.E. Norberg, 'Free-Induction Decays in Solids', *Phys. Rev.*, **107**, 46–61 (1957).
5. E.D. Becker, 'A Brief History of Nuclear Magnetic Resonance', *Anal. Chem.*, **65**, A A 295–302 (1993).
6. L. Mueller, A. Kumar, R.R. Ernst, 'Two-dimensional Carbon-13 NMR spectroscopy', *J. Chem. Phys.*, **63**, 5490–5491 (1975).
7. W.P. Aue, E. Baryholdi, R.R. Ernst, 'Two-dimensional Spectroscopy. Application to Nuclear Magnetic Resonance', *J. Chem. Phys.*, **64**, 2229–2246 (1976).
8. G.N. Chmurny, D.I. Hoult, 'The Ancient and Honorable Art of Shimming', *Concepts Magn. Reson.*, **2**, 131–149 (1990).
9. P.C.M. Van Zijl, S. Sukumar, M. Johnson, P. Webb, R.E. Hurd, 'Optimized shimming for high-resolution NMR using three-dimensional image-based filed mapping', *J. Magn. Reson. A*, **111**, 203–207 (1994).
10. S. Sukumar, M.O. Johnson, R.E. Hurd, P.C.M. Van Zijl, 'Automated shimming for deuterated solvents using filed mapping', *J. Magn. Reson.*, **125**, 159–162 (1997).
11. D.I. Hoult, R.R. Richards, 'Signal-to-noise ratio of nuclear magnetic resonance experiments', *J. Magn. Reson.*, **24**, 71–85 (1976).
12. P. Styles, N.F. Soffe, C.A. Scott, D.A. Cragg, D.J. White, P.C.J. White, 'A high-resolution NMR probe in which the coil and preamplifier are cooled with liquid helium', *J. Magn. Reson.*, **60**, 397–404 (1984).
13. S. Budavari, M.J. O'Neil, A. Smith, P.E. Heckelman, *The Merck Index, an Encyclopedia of Chemicals, Drugs, and Biologicals*, Eleventh Edition, Merck Co., Inc., Rahway, NJ, 1989.
14. E. Kupce, R. Freeman, B.K. John, 'Parallel acquisition of two-dimensional NMR spectra of several nuclear species', *J. Am. Chem. Soc.*, **128**, 9606–9607 (2006).
15. S. Akota, L. Barantin, M. Trieweiler, 'Concentration measurement by proton NMR using the ERTIC method', *Anal. Chem.*, **71**, 2554–2557 (1999).
16. J.B. d'Espinose de Lacailleriea, B. Jarry, O. Pascui, D. Reichert, 'Cooking the sample: Radiofrequency induced heating during solid-state NMR experiments', *Solid State NMR*, **28**, 225–232 (2005).
17. A. Krahn, U. Priller, L. Emsley, F. Engelke, 'Resonator with reduced sample heating and increased homogeneity for solid-state NMR', *J. Magn. Reson.*, **191**, 78–92 (2008).

Nuclear Magnetic Resonance of Geological Materials and Glasses

Grainne Moran and Russell F. Howe
University of New South Wales, Sydney, Australia

1 Introduction **2**
1.1 Overview of Nuclear Magnetic Resonance Applications 2
1.2 Scope 2
2 Experimental Issues **3**
2.1 Overview of Solid-state Nuclear Magnetic Resonance 3
2.2 Static (Wide-line Nuclear Magnetic Resonance) of Solids 6
2.3 Magic-angle Spinning for $I = \frac{1}{2}$ Nuclei 7
2.4 Cross-polarization 8
2.5 Dipolar Decoupling 9
2.6 Chemical Shift Correlation Experiments 9
2.7 Double-resonance Experiments 10
2.8 Sample Spinning Experiments with Quadrupolar Nuclei 11
2.9 Other Techniques for Quadrupolar Nuclei 12
2.10 Nuclear Magnetic Resonance Experiments at High Temperature 13
2.11 Quantitative Solid-state Nuclear Magnetic Resonance 13
2.12 Nuclear Magnetic Resonance Relaxation Experiments 14
2.13 Magnetic Resonance Imaging 15
3 Applications **16**
3.1 Speciation and Structure Determination 16
3.2 Nuclear Magnetic Resonance Relaxation Applications for Mobile Ions in Glasses 32
3.3 Nuclear Magnetic Resonance Imaging 34
4 Future Trends **36**
Acknowledgments **37**
Abbreviations and Acronyms **37**
Related Articles **37**
References **38**

This article concerns the application of nuclear magnetic resonance (NMR) spectroscopic techniques to the analysis of geological materials and glasses. It includes inorganic minerals and glasses, but does not cover soils and clay minerals to any extent. The first part presents an overview of solid-state NMR experiments applicable to inorganic solids. The range of nuclei accessible by NMR and the particular requirements for obtaining good spectra of more difficult nuclei, including quadrupolar nuclei, are discussed. The experimental conditions under which solid-state NMR can provide quantitative analytical measurements are also considered.

NMR spectroscopy provides element-specific speciation and structural information including coordination environment and bond distances. It is particularly valuable for the analysis of amorphous and partially crystalline materials not amenable to structure determination by X-ray diffraction. NMR can detect and quantify materials containing multiple phases and can be used for in situ monitoring of phase transitions. It provides qualitative and quantitative speciation information for framework nuclei in glasses and is also capable of characterizing the dynamic behavior of mobile ions and solvent through relaxation, diffusion and chemical exchange experiments. Thus it has become an important technique for the study of hydrous minerals and gels and for ion-conducting glasses.

The sensitivity of NMR varies considerably, depending on the nucleus observed, its abundance and its chemical environment. NMR sensitivities and detection limits cannot compete with elemental analytical techniques such as X-ray fluorescence. In addition, quantitation in solids requires careful control of experimental parameters and careful calibration to avoid artifacts. However, the selectivity of NMR as an analytical technique is unsurpassed and the development of specialist pulse sequences to measure specific local interactions in solids makes it particularly well suited to the analysis of geological materials and glasses.

NMR imaging enables structural information to be resolved in three dimensions, although resolution in solid-state imaging is limited to tens of micrometers at best owing to the broader line widths and lower signal-to-noise ratios compared with liquids. However, capabilities in this field are being extended all the time and the application of NMR imaging techniques in materials analysis is expanding steadily. An alternative to conventional imaging is NMR force microscopy, which shows potential for surface mapping at higher resolutions.

The applications section is organized on the basis of the materials being analyzed. Within each section, NMR techniques are classified according to the observed nuclei. The emphasis is on more recent applications and on developments leading to enhanced resolution and sensitivity. Silicates and aluminosilicates represent

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the largest class of applications, followed by phosphate, borate and other oxide glasses. Sol–gel materials and the reactions leading to gel formation are also discussed, as are hydrous materials more generally. The other important areas of application reviewed are minerals, ceramics, high-temperature melts and ion-conducting glasses. Finally, the applications of NMR imaging in this field, although sparse at present, are potentially very great. This is illustrated by two major areas of application to date, namely the imaging of bone minerals and cement and concrete.

1 INTRODUCTION

1.1 Overview of Nuclear Magnetic Resonance Applications

NMR spectroscopy measures the transitions between nuclear spin energy levels in the presence of an external magnetic field. The energies of the transitions are sensitive to the local chemical environment of the nuclei. NMR spectroscopy can be applied to any nucleus having a spin $I > 0$, although in practice some nuclei have low abundances and/or low sensitivities, making them difficult to observe, especially in solid samples. Wide-line spectra on static solids often consist of unresolved peaks or powder patterns whose line shape is sensitive to chemical shift anisotropy, dipole–dipole interactions and other interactions, moderated by molecular motion. High-resolution spectra of solids, achieved using a combination of magic-angle spinning (MAS) and specialist pulse sequences, provide compositional and structural information on the sample and in favorable cases on internuclear distances, while signal integration provides quantitation. In the solid state, the resolution is lower than in liquids, but NMR is still capable of resolving different species according to coordination number and ligand type, as exemplified by Si speciation in silicates, for which ^{29}Si NMR is now the primary method.

NMR is particularly valuable in the analysis of amorphous and partially crystalline materials not amenable to structure determination by X-ray diffraction. NMR can detect and quantify materials containing multiple phases and can be used for in situ monitoring of phase transitions. It provides qualitative and quantitative speciation information for framework nuclei in glasses and is also capable of characterizing the dynamic behavior of mobile ions and solvents through relaxation, diffusion and chemical exchange experiments. Hence it has become an important technique for the study of hydrous minerals and gels and for ion-conducting glasses.

The sensitivity of NMR varies considerably, depending on the nucleus observed, its abundance and its chemical environment. NMR sensitivities and detection limits cannot compete with elemental analytical techniques such as X-ray fluorescence. In solids, between 10^{17} and 10^{20} nuclei are required, depending on the sensitivity of the nucleus, the temperature and the applied field. In addition, quantitation in solids requires rigorous control of experimental parameters and careful calibration to avoid artifacts. Since standard reference materials are rarely available for the sorts of materials being considered here, a combination of internal and external standards is needed for absolute concentrations to be obtained. However, the selectivity of NMR as an analytical technique is unsurpassed and the development of specialist pulse sequences to measure specific local interactions in solids makes it particularly well suited to the analysis of geological materials and glasses.

NMR spectroscopy can also be compared with vibrational spectroscopy (infrared and Raman) as a tool for structural determination in disordered materials. Vibrational spectroscopy is still an important complementary technique for the characterization of glasses. NMR spectra can be broadened considerably owing to paramagnetic components in the sample which do not affect the resolution of infrared or Raman spectra. Vibrational techniques are also easier to implement under conditions of high temperature and pressure.

NMR imaging enables structural information to be resolved in three dimensions, although resolution in solid-state imaging is limited to tens of micrometers at best, owing to the broader line widths and lower signal-to-noise ratios compared with liquids. However, capabilities in this field are being extended all the time and the application of NMR imaging techniques in materials analysis is expanding steadily. An alternative to conventional imaging is NMR force microscopy, which shows potential for surface mapping at higher resolutions.

The versatility of NMR and its ability to provide a unique combination of structural, dynamic and quantitative information are its main advantages in the analysis of geological materials and glasses.

1.2 Scope

This article concerns the application of NMR techniques to the analysis of geological materials and inorganic glasses. It has been necessary to be selective in the choice of applications, given the very broad scope of the field from the point of view of both the NMR techniques and the samples involved. The greatest emphasis has been placed on glasses and gels. We have included inorganic minerals but have not covered soils and clay minerals to any extent. The broad field of synthetic zeolites and mesoporous materials has been excluded, except to illustrate particular experimental points. Some solution-state NMR applications are

Table 1 NMR reviews and monographs

Section	Title	Refs.
2	Recent Advances in Experimental Solid-state NMR Methodology for Half-integer Spin Quadrupolar Nuclei	5
3.1.9	NMR at High Temperature	33
3.1	Solid-state NMR Studies of Molecular Sieve Catalysis	137
3.1	NMR Techniques and Applications in Geochemistry and Soil Chemistry	138
3.1	High-resolution Fluorine-19 Magnetic Resonance of Solids	139
3.1, 3.2, 3.3	Structural Characterization of Noncrystalline Solids and Glasses Using Solid-state NMR	1
3.1	Solid-state NMR II: Inorganic Matter	9
3.1	Solid-state NMR for Chemists	140
3.1	Solid-state NMR – Special Issue on Glasses	141
3.1	Characterizing Porous Media with NMR Methods	142
3.1	Solid-state NMR Investigations on the Nature of Hydrogen Bonds	143
3.1	NMR and NQR Studies of Boron in Vitreous and Crystalline Borates	144
3.1	Structure, Properties and Application of Oxynitride Glasses	145
3.1	Solid-state NMR as a Tool of Structure and Dynamics in Solid-state Chemistry and Materials Science – Recent Progress and Challenges	50
3.1.5, 3.1.6	Application of NMR Spectrometry in the Study of Alkoxysilane Polycondensation	146
3.1	The Dynamics of H_2O in Minerals	147
3.1	Fluorine-19 MAS NMR	84
3.1	Structure and Chemical Modification in Oxide Glasses	51
3.1	NMR Investigation of the Structures of Phosphate and Phosphate-containing Glasses – a Review	80
3.1, 3.2	Structure and Dynamics in Glassy and Molten Silicates	113
3.1	High Resolution NMR of Solids	148
3.1, 3.2, 3.3	NMR Tomography Diffusometry Relaxometry	41
3.1	MAS NMR Spectroscopy of Minerals and Glasses	2
3.1, 3.2	NMR Spectroscopy and Dynamic Processes in Mineralogy and Geochemistry	4
3.1.9	Structure, Dynamics and Properties of Silicate Melts	3
3.3	Stray Field MRI of Solids	45, 46
3.3	3-D Magnetic Resonance Microscopy of Materials	149
3.3	Magic Echoes and NMR Imaging of Solids	150
3.3	Solid-state NMR – Special Issue on MRI of Materials	151
3.3	Solid-state NMR as an Analytical Technique Offering Spatially Resolved Chemical Information	152
3.3	NMR Imaging of Materials	44
3.3	Principles of NMR Microscopy	43

3-D, three-dimensional; MRI, magnetic resonance imaging.

also described, for sol–gel reaction systems and high-temperature liquids and melts. The literature reviewed is primarily from 1993 onwards. A list of reviews covering NMR applications to glasses and minerals, along with relevant monographs, is given in Table 1, to provide access to earlier literature. In particular, the 1992 review by Eckert[1] provides a comprehensive discussion of the literature on glasses up to that point, while NMR applications in mineralogy are reviewed in three monographs.[2–4]

The applications section is organized on the basis of the materials being analyzed. Within each section, NMR techniques are classified according to the observed nuclei. The emphasis is on more recent applications and on developments leading to enhanced resolution and sensitivity. Silicates and aluminosilicates represent the largest class of applications, followed by phosphate, borate and other oxide glasses. Sol–gel materials and the reactions leading to gel formation are also discussed, as are hydrous materials more generally. The other important areas of application reviewed are minerals, ceramics, high-temperature melts and ion-conducting glasses. Finally, the applications of NMR imaging in this field, although sparse at present, are potentially very great. This is illustrated by two major areas of application to date, namely the imaging of bone minerals and cement and concrete.

2 EXPERIMENTAL ISSUES

2.1 Overview of Solid-state Nuclear Magnetic Resonance

NMR spectroscopy as a structural tool first found wide application when applied to molecules in the liquid phase. The technique involves measuring transitions between nuclear spin energy levels, and these transitions are in turn sensitive to the chemical environment in which the nuclei

are found. In the first forms of NMR spectroscopy to be developed, only liquid samples gave lines narrow enough to reveal directly chemical information. The Hamiltonian operator describing interactions of nuclear spins with an external field and with each other in the liquid state can be written according to Equation (1):

$$H = H_Z + H_\sigma + H_J \quad (1)$$

where H_Z describes the nuclear Zeeman interaction, H_σ is the magnetic shielding term and H_J is the indirect spin–spin coupling term.

The nuclear Zeeman interaction between the external magnetic field and the magnetic moment arising from the nuclear spin depends on the field strength and on the magnitude of the nuclear spin magnetic moment. There exist 117 stable naturally occurring isotopes with nonzero nuclear spins; the nuclear spin magnetic moments range in magnitude from 0.111 (^{187}Os) to 4.837 (^{1}H) in units of the nuclear Bohr magneton (the radioactive isotope ^{3}H has a larger moment than that of the proton). Tables 2 and 3 list the properties of nuclei commonly studied in glasses and geological materials; these are divided into spin $= \frac{1}{2}$ and spin $> \frac{1}{2}$ because of the importance of quadrupolar interactions discussed further below for nuclei with spin $> \frac{1}{2}$. Tables 2 and 3 also list the frequencies at which NMR transitions are observed for different nuclei at a magnetic field strength of 7 T, a typical field strength used for solid-state NMR studies.

Table 2 NMR properties of selected $I = \frac{1}{2}$ nuclei

Nucleus	Natural abundance (%)	Sensitivity (relative to ^{1}H) at constant field	NMR frequency at 7.05 T (MHz)	Approx. chemical shift range (ppm)
^{1}H	99.98	1.00	300.00	10
^{13}C	1.108	1.76×10^{-4}	75.432	300
^{15}N	0.37	3.85×10^{-6}	30.398	1000
^{19}F	100	0.83	282.231	800
^{29}Si	4.7	3.69×10^{-4}	59.595	250
^{31}P	100	6.63×10^{-2}	121.442	2000
^{77}Se	7.58	5.525×10^{-4}	57.203	2400
^{109}Ag	48.18	4.86×10^{-5}	13.956	1000
^{113}Cd	12.26	1.33×10^{-3}	66.548	800
^{205}Tl	70.5	0.13	173.124	7000
^{207}Pb	22.6	2.07×10^{-3}	62.765	5000

Table 3 NMR properties of selected quadrupolar nuclei

Nucleus	Spin	Natural abundance (%)	Sensitivity (relative to ^{1}H) at constant field	NMR frequency at 7.05 T (MHz)	Approx. chemical shift range (ppm)	Quadrupole moment (10^{-28} m^2)
^{2}H	1	0.015	1.45×10^{-6}	46.051		2.8×10^{-3}
^{6}Li	1	7.42	6.31×10^{-4}	44.146	12	-8×10^{-4}
^{7}Li	$\frac{3}{2}$	92.58	0.27	116.590	12	-4×10^{-2}
^{9}Be	$\frac{3}{2}$	100	1.39×10^{-2}	42.160	25	5×10^{-2}
^{11}B	$\frac{3}{2}$	80.42	0.13	96.251	100	4.1×10^{-2}
^{14}N	1	99.63	1.01×10^{-3}	21.671	120	1×10^{-2}
^{17}O	$\frac{5}{2}$	3.7×10^{-2}	1.08×10^{-5}	40.670	600	-2.6×10^{-2}
^{23}Na	$\frac{3}{2}$	100	9.25×10^{-2}	79.353	110	0.10
^{25}Mg	$\frac{5}{2}$	10.13	2.71×10^{-4}	18.358	50	0.22
^{27}Al	$\frac{5}{2}$	100	0.21	78.172	300	0.15
^{47}Ti	$\frac{5}{2}$	7.28	1.52×10^{-4}	16.910		0.29
^{49}Ti	$\frac{7}{2}$	5.51	2.07×10^{-4}	16.914		0.24
^{51}V	$\frac{7}{2}$	99.76	0.38	78.864	1300	-5×10^{-2}
^{95}Mo	$\frac{5}{2}$	15.72	5.07×10^{-4}	19.544	4000	0.011

The sensitivity of a particular nucleus in the NMR experiment depends on the population difference between the spin states being excited, which in turn is a function of both the magnetic field strength and the magnetic moment of the nucleus. The larger the magnetic moment and the higher the magnetic field, the stronger the signal becomes. A third factor which may also be experimentally adjusted is the abundance of the particular isotope being examined. For example, ^{17}O is an isotope for which the NMR spectrum can give a great deal of structural information about oxide materials. With a natural abundance of 0.037%, however, the sensitivity is normally too low to be useful, but isotopic enrichment is frequently used to raise the sensitivity to observable levels (see examples in section 3). Tables 2 and 3 also include a relative sensitivity factor for each nucleus, which expresses the sensitivity relative to that of the proton, assuming that the isotope is present at natural abundance.

The magnetic shielding interaction H_σ, more frequently referred to as the chemical shift, provides unique information about the chemical environment of the nucleus. It is due to the fact that electron density around the nucleus perturbs the magnetic field to a small extent, shifting the NMR frequency. Chemical shifts are usually reported in terms of a frequency shift from a reference compound divided by the resonance frequency of the reference compound, thus cancelling the effect of the nuclear Zeeman term. In the liquid phase, the chemical shift is averaged by rapid reorientation of molecules on the NMR timescale to give the isotropic value. The range of chemical shifts observed generally increases with the atomic number of the element concerned, e.g. from about 10 ppm for ^{1}H to 200 ppm for ^{13}C and 5000 ppm for ^{207}Pb.

The indirect spin–spin coupling interaction between nuclear spins (H_J) is the result of coupling between nuclei and electrons in chemical bonds intervening between the atoms concerned. This interaction is extremely weak, typically a few hertz, but is routinely used as a powerful structural tool in high resolution liquid-phase NMR of complex molecules. The weakness of the interaction means that heteronuclear spin–spin coupling for molecules in the liquid phase is readily removed by low-power decoupling.

In solids, there are up to three additional interactions present which are not evident in liquid-phase spectra, and which cause dramatic increases in line width. First, the shielding interaction (chemical shift) is anisotropic, i.e. it depends on the orientation of the nuclear environment in the magnetic field. A low-symmetry nuclear environment (and it is the crystallographic site symmetry that is important rather than the molecular symmetry) will have its shielding described by three different principal components of the shielding or chemical shift tensor. If the environment has axial symmetry, two of the principal components will be equal (corresponding to the chemical shift when the field is perpendicular to the symmetry axis), and different from the third (the so-called parallel component). Only in the case of cubic symmetry will the chemical shift appear to be isotropic.

The consequence of shielding anisotropy for single-crystal samples is that the NMR frequency will vary with orientation of the crystal in the magnetic field. For polycrystalline powders or glasses, which contain randomly oriented nuclear environments, the NMR spectrum will be broadened over the frequency range corresponding to the chemical shift anisotropy. Figure 1(a–c) shows schematically the form of the so-called powder pattern expected for cubic, axial and nonaxial symmetry. A narrow line is obtained only for cubic symmetry. For the axial and nonaxial cases, the principal components of the shielding tensor can be deduced from the turning points in the spectrum.

The shielding anisotropy can yield useful information about the structure of the system being studied. In general, however, the spectral broadening caused by shielding anisotropy is perceived as degrading resolution, and MAS techniques are applied to remove it, as described below.

The second interaction present in solids which averages to zero in liquid-phase spectra is dipolar spin–spin coupling. The dipolar coupling interaction is the classical through-space interaction between two magnetic dipoles, which may both be the same nucleus (homonuclear dipolar coupling) or different (heteronuclear dipolar coupling). The magnitude of the interaction depends on the magnetic moments of the two nuclear spins, the distance between them and the orientation of the vector joining the spins in the magnetic field. The orientation dependence of the dipolar coupling is similar to that of the shielding anisotropy, i.e. dipolar coupling can also be

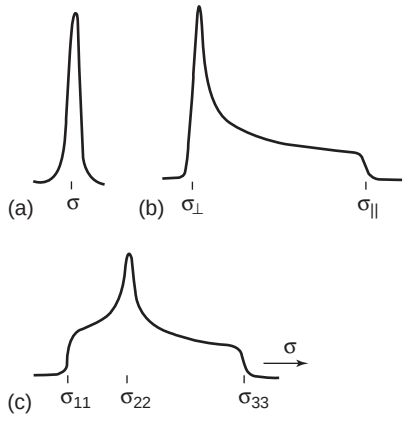

Figure 1 Powder patterns produced by shielding (chemical shift) anisotropy for sites with (a) cubic symmetry, (b) axial symmetry and (c) lower symmetry.

described by a second-rank tensor with three principal components. The spectra of powder or glassy samples in which dipolar coupling is present may contain broad poorly resolved lines. In samples which contain high concentrations of nuclei with large magnetic moments (such as ^{1}H or ^{19}F), the broadening caused by dipolar coupling may be as large as tens or hundreds of kilohertz, which will totally obscure any chemical shift information in the spectrum. Detection of such broad signals is also experimentally difficult. Developing efficient strategies for eliminating homonuclear dipolar coupling has been a major achievement of solid-state NMR in recent years. In other cases in which only heteronuclear dipolar coupling is present, high-power decoupling methods can be used to eliminate the interaction.

A third interaction present in the spectra of solids containing nuclei with spins $I > \frac{1}{2}$ is the nuclear quadrupole interaction. Nuclei with $I > \frac{1}{2}$ have a nonspherically symmetrical nuclear charge distribution producing a quadrupole moment. The interaction of this quadrupole moment with electric field gradients caused by asymmetric electron distributions in molecules or lattice sites perturbs the nuclear spin energy levels. For single-crystal samples, the quadrupole interaction causes a splitting of the single resonance into multiple components which will vary with orientation (depending on the symmetry of the electric field gradient). The magnitude of the interaction depends on the size of the quadrupole moment and the magnitude of the electric field gradient. If the interaction is small relative to the nuclear Zeeman term, it may be accurately described by first-order perturbation theory. In this treatment, the central $(\frac{1}{2} \leftrightarrow -\frac{1}{2})$ transition for nuclei with nonintegral nuclear spins is unaffected by the quadrupole interaction. This transition will therefore still be observed in spectra of polycrystalline powders or glasses, whereas the satellite transitions, which are strongly orientation dependent, may be broadened beyond detection. Nuclei with integral spins will give broad powder patterns for all transitions, although the principle components of the quadrupole coupling tensor may be deduced from the turning points. Figure 2(a, b) illustrates hypothetical powder patterns for $I = 1$ and $\frac{5}{2}$ nuclei, assuming a first-order quadrupole perturbation.

When the quadrupole interaction becomes larger relative to the nuclear Zeeman term, a second-order perturbation treatment of the interaction must be used. In this case, even the central transition for nuclei with nonintegral spin becomes anisotropic, and in polycrystalline or glassy samples gives a broad and complex line shape (Figure 2c). Note also that the centre of gravity of the spectrum is no longer at the isotropic chemical shift ν_0. MAS (Figure 2d) does not completely remove the second-order quadrupolar broadening. More generally, spectra

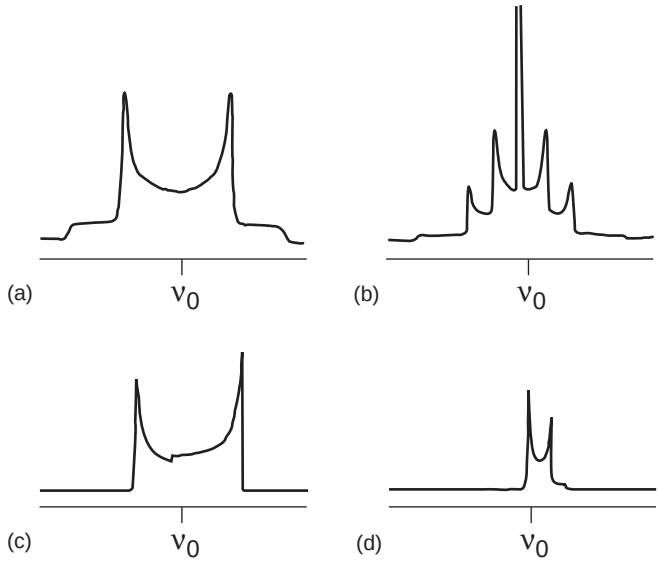

Figure 2 Schematic powder patterns for quadrupolar nuclei. (a) $I = 1$, first-order (static); (b) $I = \frac{5}{2}$, first-order (static); (c) central transition of $I = \frac{5}{2}$, second-order, axial symmetry (static); (d) MAS.

of quadrupolar nuclei in powders or glasses will have line shapes determined by both quadrupole interactions and chemical shift anisotropy, and computer simulation methods must be used to extract reliable parameters.

In summary, the Hamiltonian describing the interactions of nuclear spins in a solid sample contains five terms, four of which are anisotropic (Equation 2):

$$H = H_Z + H_\sigma + H_J + H_D + H_Q \qquad (2)$$

The indirect coupling term is normally (but not always) too small to be observed in solid-state NMR spectra. The remaining three terms all contribute to line width and loss of resolution in solid-state spectra; obtaining chemically informative isotropic chemical shift information from a solid sample therefore requires steps to be taken to remove the anisotropic interactions (so-called high-resolution solid-state NMR spectroscopy). On the other hand, the anisotropic interactions contain a great deal of additional structural information. A second development in solid-state NMR spectroscopy has been the development of techniques which allow that additional information to be extracted. Both approaches are illustrated below.

2.2 Static (Wide-line Nuclear Magnetic Resonance) of Solids

Historically, NMR spectra of solids were first detected using continuous wave (CW) techniques; either the detector response was measured as the radiofrequency (RF)

or the magnetic field was swept through resonance under slow passage conditions. The sensitivity is inherently low, however, and instruments capable of performing this experiment are no longer commercially available.

The application of pulse NMR techniques dramatically improved the sensitivity and versatility of NMR spectroscopy applied to liquid-phase samples with inherently narrow lines. The broad lines typical of solids are, however, much more difficult to measure accurately using pulse techniques. In a conventional one-pulse experiment, where a single pulse is followed by measurement of the free induction decay (FID), a significant fraction of the FID from a broad solid spectrum will be lost in the receiver deadtime immediately following the pulse. Fourier transformation of such an FID will give an inaccurate depiction of the spectrum. The deadtime problem can be overcome by Fourier transforming the decay of a spin echo; a number of different echo-generating pulse sequences have been used for this purpose.[5] Figure 3(a) and (b) shows, for example, static ^{27}Al NMR spectra of α-alumina measured by single-pulse and spin echo acquisition. The single-pulse experiment allows the turning points expected in the powder pattern for an $I = \frac{5}{2}$ nucleus to be correctly recorded, but does not correctly display the intervening intensity. The spectrum acquired by Fourier transformation of the spin echo decay gives a much better (although still not completely accurate) depiction of the true line shape.

No single-pulse or echo measurement will allow correct line shapes to be recorded for lines broader than about 200 kHz. This limitation arises from the maximum excitation bandwidth achievable with a conventional high-power pulse. Excitation by composite pulses or shaped pulses may achieve some small improvement in bandwidth.[6] The alternative approach for broad lines is to measure the spectrum conventionally or using echoes at a series of different frequencies or fields (i.e. to combine the advantages of the old CW technique with those of the pulse experiment). For example, Bastow and Smith[7] obtained a broad-line ^{91}Zr spectrum of ZrO_2 by measuring the spin echo amplitude at many different excitation frequencies to map out the line shape. This is, however, a tedious experiment requiring retuning of the probe between each point. Alternatively, the magnetic field may be swept. Poplett and Smith[8] recently described an instrument containing a superconducting magnet which could be swept over a field range of 1 T; this allowed ^{27}Al spectra with line widths beyond 3 MHz to be accurately measured.

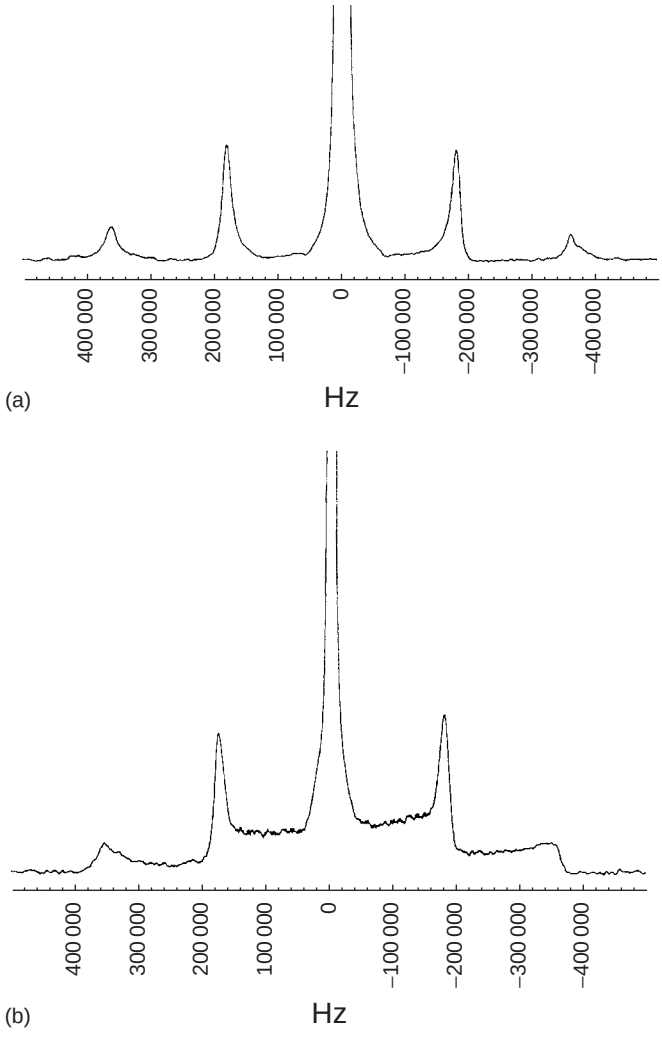

Figure 3 Static ^{27}Al NMR spectra of α-alumina obtained using (a) single pulse and (b) spin-echo acquisition. [Reproduced with permission from Smith and Van Eck.[5]]

2.3 Magic-angle Spinning for $I = \frac{1}{2}$ Nuclei

The terms in the dipolar coupling Hamiltonian and the anisotropic component of the shielding Hamiltonian both contain the geometric factor $3\cos^2\theta - 1$, where θ describes the orientation with respect to the magnetic field. In the liquid phase, rapid isotropic molecular tumbling averages this factor to zero, so that dipolar coupling and shielding (chemical shift) anisotropy vanish. A similar averaging can in principle be achieved by rotating a solid sample rapidly about an axis making an angle of 54.7° to the magnetic field (the angle at which $3\cos^2\theta - 1 = 0$). In the case of dipolar coupling, the rotation frequency must exceed the magnitude of the coupling. Since this may be hundreds of kilohertz, and practically feasible sample rotation frequencies are limited to <20 kHz, the MAS technique is not generally used for the removal of dipolar broadening. Broadening due to chemical shift anisotropy, however, may often be small enough in magnitude to qualify for this technique. Figure 4 shows, for example, the

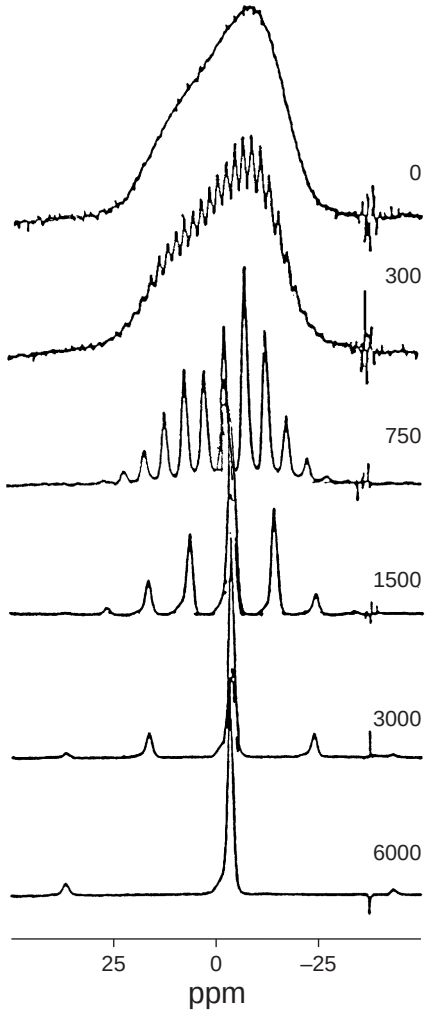

Figure 4 ^{31}P MAS NMR spectra of $WO_3 \cdot P_2O_5$ at different spinning speeds. [Reproduced with permission from Blumich.[9]]

^{31}P spectrum of solid $WO_3 \cdot P_2O_5$ measured at different MAS spinning speeds.[9] The static spectrum shows a typical powder pattern for an axially symmetric chemical shift tensor, with a total anisotropy of ca. 40 ppm or 6000 Hz. When the sample is spun at the magic angle at a frequency of 6000 Hz, the spectrum collapses to a single line at the isotropic value of the chemical shift. At lower spinning speeds, spinning side bands are seen which trace out approximately the intensity profile of the powder pattern, and which are separated by the spinning frequency.

Following the first introduction of a commercial MAS probe in 1978, the design and performance of such probes has steadily improved and routine spinning of solid samples at frequencies of 18 kHz or higher under computer control is now feasible. This achieves removal of chemical shift anisotropy from spectra of lower atomic number elements although, as noted above, heavier elements such as ^{207}Pb may show too large an anisotropy to be fully removed. Where MAS does not totally remove chemical shift anisotropy, the intensity distribution of the spinning sidebands can be analyzed to deduce the principal components of the chemical shift tensor.[10] Pulse sequences are also available for suppressing the spinning side bands. The effects of MAS on spectra of quadrupolar nuclei are discussed in section 2.8.

2.4 Cross-polarization

The technique of cross-polarization (CP) is usually coupled with MAS in measuring the spectra of magnetically dilute nuclei, to enhance sensitivity. This is achieved in two ways: by transferring magnetization from a more abundant nucleus to the dilute nucleus of interest, and by shortening the spin–lattice relaxation time allowing faster pulse repetition and thereby achieving a better signal-to-noise ratio in a given measurement period. Detailed accounts of CP and the pulse sequences used are given elsewhere.[11,12] The key to successful CP is strong dipolar coupling between the abundant nucleus and the nucleus being measured. The abundant spin is first spin locked in the rotating frame, where it oscillates at a frequency controlled by the RF field strength. The dilute spin is then rotated in the rotating frame at a frequency matching that of the abundant spin, allowing spin flip transitions to occur and magnetization of the dilute spin to grow. The RF excitation of the dilute spin is then turned off, and its FID measured. The RF excitation of the abundant spin remains on during the FID acquisition, to decouple heteronuclear dipolar interactions between the two spins. The delay between successive pulses and FID acquisitions is now determined by the spin–lattice relaxation time of the abundant nucleus, rather than that of the dilute nucleus.

The efficiency of CP is determined by the extent of dipolar coupling and the rate of spin diffusion of the abundant nucleus, which causes loss of the spin-locked signal. The balance achieved between these two competing processes will determine the extent to which enhancement of signal intensity occurs. Optimum CP requires the abundant nucleus to have a relatively long spin diffusion rate (characterized by $T_{1\rho}$, the relaxation time in the rotating frame) and a relatively short spin–lattice relaxation time.

CP is usually combined with MAS, which can cause some difficulties. As noted above, MAS will tend to average the dipolar coupling interaction to zero. If it were to do this completely, the efficiency of CP would fall to zero. In practice, the MAS spinning rate is rarely fast enough to eliminate dipolar coupling; nevertheless, it can greatly reduce the efficiency of CP. There is therefore an incentive to avoid high-speed MAS when carrying out CP.

Alternatively, a recently introduced variable-amplitude cross-polarization (VACP) experiment appears to offer promise for efficient CP at high MAS frequencies.[13]

CP has been widely used in ^{13}C NMR studies of organic compounds, where abundant protons transfer magnetization to dilute ^{13}C. The technique can, however, be used in principle with any dilute spin system. For example, in ^{29}Si NMR of silica gels and glasses, ^{1}H CP will selectively enhance only the signals due to Si atoms which are close to protons (i.e. SiOH groups).

2.5 Dipolar Decoupling

Removal of the heteronuclear dipolar coupling contribution to line broadening in the NMR spectra of solids is similar in principle to the decoupling process in liquid-phase NMR: the second nucleus is irradiated at its NMR frequency with sufficient power to saturate the resonance, reducing the magnetic moment of the second nucleus effectively to zero. Because the dipolar coupling is much larger in magnitude than the indirect (scalar) coupling found in the liquid phase, the power levels needed to achieve effective heteronuclear decoupling in the solid state are considerably higher than those in liquid samples. Effective decoupling requires the continuous application of an RF field to the second nucleus during acquisition of the FID from the nucleus of interest. This can entail transmitter power levels of 1 kW or higher, which created some experimental difficulties during the early development of solid-state NMR, but which is now routinely available. Such high-power decoupling also eliminates the indirect heteronuclear coupling interaction although, as noted above, this is normally too weak to have a significant impact on line widths.

High-power decoupling cannot be used to eliminate the homonuclear dipolar coupling interaction which dominates the NMR spectra of solids containing magnetically abundant nuclei in close proximity (particularly ^{1}H and ^{19}F). This interaction is generally too strong to be removed by MAS; although the angular dependence of the dipolar coupling interaction is identical with that of the chemical shift anisotropy, spinning speeds needed to eliminate dipolar broadening cannot normally be realized. One chemical approach to the problem is to achieve magnetic dilution by selective isotopic labeling. For example, substitution of ^{1}H by ^{2}H in selective sites allows measurement of quadrupolar line shapes in the proton decoupled ^{2}H NMR spectrum, which give information about molecular order and dynamics. Examples of such ^{2}H NMR studies are described below. Generally however, such chemical modification may cause structural changes to the system of interest, and therefore cannot compete with physical decoupling methods.

Physical decoupling of the homonuclear dipolar coupling interaction can be achieved with multiple pulse sequences. These consist of a cycle of RF pulses which is applied repetitively to the spin system. Qualitatively, such a sequence can be visualized as rotating the spin magnetic moments in spin space rather than laboratory space, achieving a similar averaging of the dipolar coupling interaction to zero to that found in liquid samples. For example, the so-called Waugh–Huber–Haeberlin (WAHUHA) four pulse sequence applies successive 90° pulses about the x, $-x$, y and $-y$ axes in the rotating frame, with suitable delays between. This can be visualized as forcing the spins to spend equal times oriented along the x, y and z axes in the rotating frame, which produces the same average effect as alignment along the MAS axis.

A more correct and quantitative description of multiple pulse averaging requires the use of density operator theory to solve the time-dependent Schrödinger equation, which will not be undertaken here. Increasing the number of pulses and varying their phases can achieve averaging of the dipolar coupling interaction to higher order and compensate for magnetic field inhomogeneities and finite pulse lengths. Since the entire pulse sequence must be applied between the acquisition of each point in the FID, the need to generate precisely timed sequences of extremely short pulses places formidable demands on pulse generation and pulse timing hardware. The technique has nevertheless now become readily available on commercial instruments. The combination of multiple pulse averaging of homonuclear dipolar coupling and MAS to remove chemical shift anisotropy has become known as combined rotation and multiple pulse spectroscopy (CRAMPS); this method for achieving high-resolution solid-state spectra is now widely used for ^{1}H- and ^{19}F-containing solids.[14]

2.6 Chemical Shift Correlation Experiments

Correlation spectroscopy is a two-dimensional (2-D) NMR experiment which produces correlation diagrams describing relationships between spectral features. Such 2-D correlations have been used for many years in liquid-phase NMR to determine molecular structure; more recently the techniques have also been applied to solids. Conventional one-dimensional (1-D) NMR involves conversion of the time domain signal following an RF pulse (the FID) into a frequency domain spectrum by Fourier transformation. For many pulse sequences, however, the FID may also depend on a second time variable (e.g. the delay between two successive pulses), which is known as the evolution period. If this second time variable is designated t_1 and the acquisition time of the FID (the detection period) is t_2, then the FID is a

function of both t_1 and t_2. If the experiment is carried out to measure the FID at many different t_1 values, the resulting data set will be a 2-D array in which the FID intensity depends on t_1 and t_2. Fourier transformation in both dimensions will then yield a 2-D spectrum which can be plotted in three dimensions as intensity versus the two frequency components f_1 and f_2. The significance of f_1 and f_2 will depend on the particular pulse sequences employed.

The interaction between nuclei on two sites with different chemical environments may be through chemical bonds (scalar or *J*-coupling) or through space (dipolar coupling). Although the scalar coupling interaction is too weak to be normally resolved in conventional 1-D spectra of solids, the interaction can nevertheless be used to establish correlations between nuclei which are linked by chemical bonds. For example, the so-called COSY (correlated spectroscopy) experiment involves measuring the FID following two successive 90° pulses separated by the evolution time t_1. During the evolution period, magnetization is transferred between spins which are scalar coupled, modulating the subsequent FID. The 2-D spectrum obtained by double Fourier transformation is a plot of intensity versus chemical shift in both dimensions. Figure 5 shows, for example, the ^{29}Si COSY spectrum of ^{29}Si-enriched zeolite ZSM-39 reported by Fyfe et al. in 1989.[15] The peaks along the diagonal axis correspond to the normal 1-D spectrum of this material, which contains three distinct silicon sites. The cross peaks arise from scalar coupling between ^{29}Si nuclei on different sites. In this case, sites 1 and 2 are coupled, as are sites 2 and 3, establishing the connectivity through oxygen between the sites. Because of the low natural abundance of ^{29}Si (Table 2), initial COSY experiments used isotopically enriched samples, although this was later shown not to be absolutely necessary. Signal enhancement was also achieved by exploiting CP from protons of organic template molecules within the zeolite.

An alternative correlation experiment which does not require isotopic enrichment is INADEQUATE (incredible natural abundance double quantum transfer experiment). This experiment was first introduced to establish ^{13}C–^{13}C connectivities in organic compounds at natural abundance levels, but it can also be applied to ^{29}Si in solids. The result of an INADEQUATE experiment is a 2-D plot in which only cross peaks are present at positions corresponding to the chemical shifts of nuclei which are scalar coupled to each other. Compared with COSY, the INADEQUATE experiment has the considerable advantage that there are no intense peaks on the diagonal, making it easier to observe connectivities between closely spaced resonances. However, implementing the optimum pulse sequence for an INADEQUATE experiment requires knowledge (or at least a reasonable estimate) of the scalar coupling constant, which is not resolved in the 1-D spectrum. Fyfe et al. determined the Si–O–Si *J* coupling constant in zeolites to be about 10 Hz (from observed splittings in COSY experiments with isotopically enriched samples).

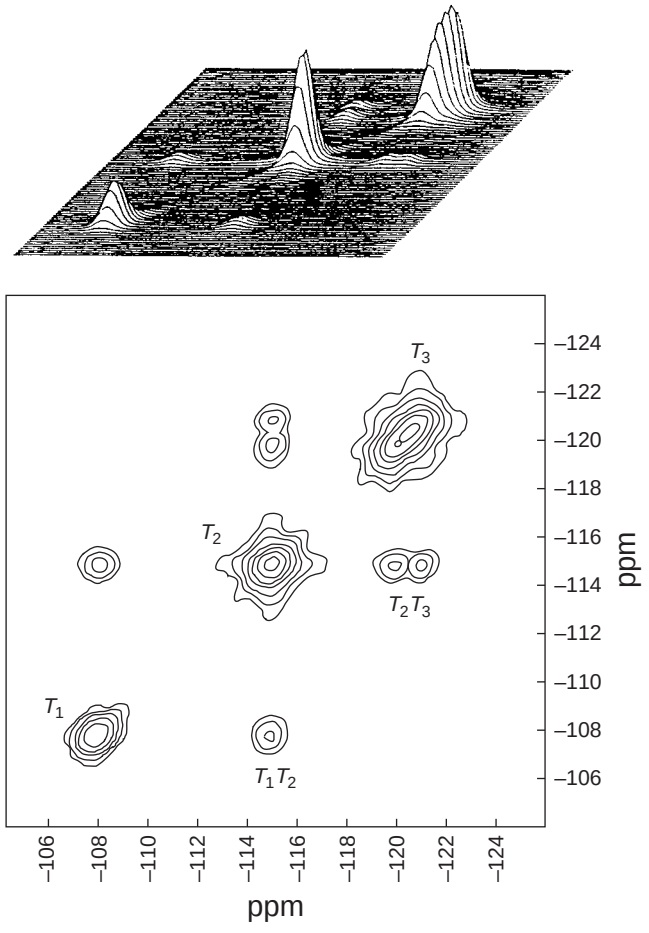

Figure 5 ^{29}Si COSY spectrum of ZSM-39. [Reproduced with permission from Fyfe et al.[15]]

2.7 Double-resonance Experiments

Information about connectivities or proximities between different nuclei in solid samples can be obtained from a variety of different double-resonance experiments. These experiments rely on the heteronuclear dipolar coupling between spins to establish their proximity in a sample. For example, the SEDOR (spin echo double resonance) experiment determines the dipolar coupling constant for a static sample by applying a 180° pulse to the second nucleus during a spin echo pulse sequence applied to the first nucleus. This has the effect of inverting the sign of the dipolar coupling between the two nuclei, reducing the echo amplitude. From a plot of spin echo intensity

versus echo time or pulse position, the dipolar coupling constant can be obtained. From the dipolar coupling constant, the internuclear distance is calculated.[16] The REDOR (rotational echo double resonance) experiment is obtained by combining SEDOR with MAS. MAS greatly enhances sensitivity and resolution; a sequence of rotor synchronized 180° pulses is applied to the second nucleus during a spin echo sequence, and the echo amplitude is attenuated by the dipolar coupling. The dipolar coupling constant (and hence the internuclear distance) is obtained by measuring the echo amplitude as a function of the number of rotor periods or of the position of the 180° pulses.[17]

Both SEDOR and REDOR are difference experiments, where the echo intensity decrease is a measure of the dipolar coupling. TEDOR (transferred echo double resonance), on the other hand, measures an intensity increase. The TEDOR experiment is essentially comprised of two REDOR pulse sequences applied to both spin systems in succession. The echo intensity for the second spin is measured as a function of the number of MAS rotor cycles or as a function of the position of the 180° pulses.[18]

The determination of internuclear distances from such 1-D double resonance experiments is readily achieved if the sample contains isolated pairs of spins. In the more usual situation where a nucleus may be coupled to a number of neighboring nuclei at different distances, interpretation of TEDOR or REDOR curves becomes more difficult, and it may become necessary to assume a particular structural model in order to extract distance information.

A qualitative double resonance experiment applicable to quadrupolar nuclei is TRAPDOR (triple resonance adiabatic passage double resonance). This still involves modulation of the heteronuclear dipolar coupling to attenuate an echo signal, as in TEDOR, but uses continuous irradiation of the second (quadrupolar) spin during the spin echo sequence applied to the first nucleus (which may or may not be quadrupolar). A TRAPDOR experiment allows the qualitative determination of which sites in a multicomponent spectrum are dipolar coupled to a quadrupolar nucleus, but no distance information can be obtained.[19]

2-D heteronuclear correlation experiments can be used to provide connectivity information in a similar manner to the homonuclear COSY experiment described above. In the so-called HETCOR (heteronuclear correlation spectroscopy) experiment, magnetization is transferred between the two spin systems during the evolution period by either CP or TEDOR. The resulting 2-D spectrum shows cross peaks for nuclei which are dipolar coupled.[20]

2.8 Sample Spinning Experiments with Quadrupolar Nuclei

As noted above, the central ($\frac{1}{2} \leftrightarrow -\frac{1}{2}$) transition in the spectra of nonintegral quadrupolar nuclei is independent of sample orientation to first order, so that this contribution to the spectrum is not broadened in polycrystalline or glassy samples. The second-order quadrupolar interaction is anisotropic, however, and the angular dependence involves two different Legendre polynomials such that there is no one orientation about which the interaction can be averaged to zero. Conventional MAS will remove the $3\cos^2\theta - 1$ component of the second-order quadrupolar broadening, and narrow but not totally remove the other component. The spinning speed needed to achieve this depends on the magnitude of the quadrupolar coupling and the magnetic field at which the spectrum is measured (since the second-order broadening depends on the relative magnitudes of the quadrupolar and Zeeman interactions). For an $I = \frac{5}{2}$ nucleus such as ^{27}Al, for example, the maximum quadrupolar coupling interaction which can be narrowed by MAS is given by Equation (3):

$$C_q = 8.82(\nu_R \nu_0)^{1/2} \tag{3}$$

Thus, in a 7-T magnet, spinning at 10 kHz will narrow ^{27}Al lines with a quadrupolar coupling constant of up to 7.78 MHz. In a 14-T magnet, on the other hand, spinning at 18 kHz will reduce the line width for coupling of up to 14.7 MHz. There is therefore an incentive to record MAS spectra of quadrupolar nuclei with nonintegral spins in as high a magnetic field as possible, with the fastest spinning speed possible, in order to narrow the central transition as much as possible; 18 kHz MAS speeds are now routinely possible, and speeds as high as 35 kHz have been reported.[21]

A second consequence of the second-order quadrupolar interaction is that the center of gravity of the narrowed central transition in the MAS spectrum will not correspond to the isotropic chemical shift. Only in the limit of very high magnetic field where the second-order quadrupolar interaction vanishes will the peak position become exactly the isotropic chemical shift. For example, the apparent chemical shift in the ^{27}Al MAS spectrum of kaolinite varies from 36 ppm at 3.5 T to −8 ppm at 14 T.[5] This provides a further incentive to measure spectra of quadrupolar nuclei at as high a field as possible if accurate chemical shifts are to be obtained, and to measure spectra at two different magnetic fields to check for the presence of any second-order shift.

Complete removal of the second-order quadrupolar interaction would allow in principle the accurate determination of isotropic chemical shifts and the resolution of small chemical shift differences. Spinning about a single axis cannot achieve this, but spinning about two axes

simultaneously can. The so-called double angle rotation (DOR) technique achieves this by enclosing a spinning rotor within a second spinning rotor such that the axis of rotation of the inner rotor describes a well defined trajectory as a function of time. The second-order quadrupolar frequency of the central transition then experiences a double modulation in the laboratory frame instead of the single modulation imposed by conventional MAS. The two angles at which the rotors spin are chosen to null the two different angular components of the second order quadrupolar interaction (e.g. 54.74° and 30.56° or 70.15°).[22] There are severe experimental difficulties in achieving stable spinning of such a double rotor, and the effectiveness of the technique is restricted by the relatively slow spinning speed that can be achieved in the outer rotor (which produces extensive spinning side band patterns). Nevertheless, the technique has been very successful in resolving chemical shift information in a number of cases.[23]

An alternative to DOR which avoids the mechanical complexity and expense of a double rotation MAS probe is so-called dynamic angle spinning (DAS). This is a 2-D technique in which the orientation of the sample rotation axis is different for every evolution period.[24] The resonance frequency of a spin when rotating at one orientation is correlated with its frequency when rotating at a second angle. A mathematical justification of how such a correlation can be used to eliminate second-order quadrupolar broadening of the central transition for nonintegral spins was given by Grandinetti.[25] The two angles are chosen to satisfy simultaneously Equations (4) and (5):

$$P_2(\cos\theta_1) = -kP_2(\cos\theta_2) \tag{4}$$

$$P_4(\cos\theta_1) = -kP_4(\cos\theta_2) \tag{5}$$

where P_2 and P_4 are the second- and fourth-order Legendre polynomials (responsible for the angular dependence of the first- and second-order quadrupolar interaction, respectively), and k is a scaling factor. There is a continuous set of solutions for these equations, each with a different scaling factor. For these solutions, the second-order powder pattern at one angle is an exact mirror image of that at the other angle. If the time at the first orientation is t_1, an echo will form at time $t_2 = kt_1$ which is isotropic. By collecting the intensity of the echo maxima as a function of t_1 and Fourier transforming the resulting interferogram, an isotropic DAS spectrum is obtained. Alternatively, the complete echo is acquired as a function of t_1 and the resulting 2-D spectrum is subjected to a so-called shearing transform to display the isotropic spectrum.

The DAS experiment requires a probe in which the rotation angle can be switched quickly and reproducibly (e.g. in <10 ms). The technique, being 2-D, involves much longer acquisition times than simple MAS or DOR, and it cannot be applied to systems having short T_1 values (which will show a substantial loss of intensity during the time taken to reorient the rotation axis). Nevertheless, it has been widely applied to nuclei such as ^{17}O.

2.9 Other Techniques for Quadrupolar Nuclei

Quadrupole nutation NMR is a 2-D experiment introduced by Kundla et al.[26] to separate quadrupolar interactions from other line broadening interactions. The spin system is allowed to evolve in time for a period t_1 in which both the RF and quadrupolar interactions are present, then for a period t_2 in which only the quadrupolar interaction is present. An FID is acquired during t_2 as a function of the RF pulse length t_1. Double Fourier transformation gives a 2-D spectrum in which the F_2 projection is the normal powder pattern showing second-order quadrupolar broadening and chemical shift anisotropy, and the F_1 projection is the so-called nutation spectrum which contains contributions only from quadrupolar interactions and is independent of chemical shift. The technique thus allows discrimination between species with similar chemical shifts but different quadrupolar interactions. The nutation spectrum can, however, be a complex pattern which is difficult to interpret, and computer simulation is often used to assist.[27]

A new experiment which was first reported in 1995[28] shows great promise for solid-state NMR studies of quadrupolar nuclei. The 2-D multiple quantum magic-angle spinning (MQMAS) experiment enhances resolution of the spectra of half integral quadrupolar nuclei. The experiment correlates the $(\frac{1}{2} \leftrightarrow -\frac{1}{2})$ single quantum transition to the $(m \leftrightarrow -m)$ multiple quantum transition. The quadrupole frequencies of both transitions are correlated; thus the anisotropic parts of the quadrupole interaction are refocused, forming an echo. In one form of the experiment, the multiple quantum transition is excited by a single high-power RF pulse, and the multiple quantum coherence allowed to evolve for a period t_1 before a second pulse is applied, which converts the multiple quantum coherence into a single quantum coherence which is observed during t_2. The resulting 2-D spectrum obtained by measuring the FID acquired in t_2 as a function of t_1 can be processed to yield both the isotropic chemical shifts and the second order quadrupole induced shifts.

Figure 6 shows, for example, the triple quantum ^{17}O 2-D MQMAS spectrum of albite glass. This material contains Si–O–Si and Si–O–Al bridging oxide ions. In the normal 1-D MAS experiment, the signals from these two species overlap and cannot be separated. In the 2-D MQMAS experiment, however, the 2-D plot shows clear

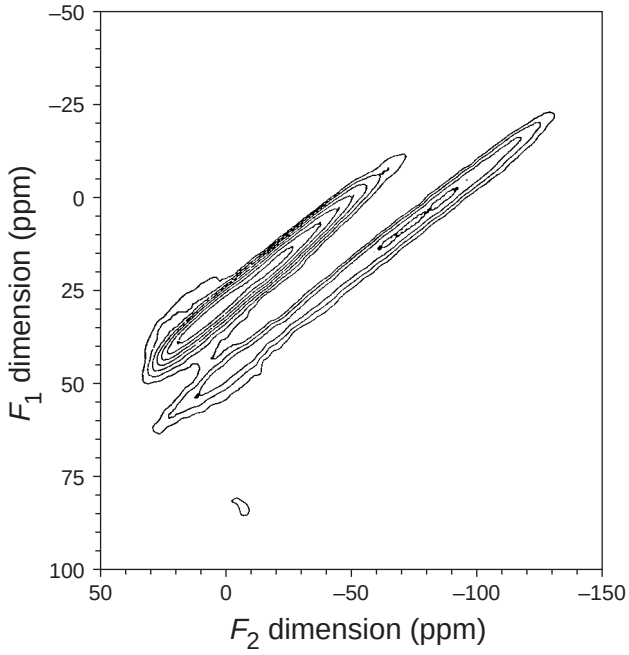

Figure 6 Triple quantum ^{17}O 2-D MQMAS spectrum of albite glass (^{17}O enriched). [Reproduced with permission from Dirken et al.[29]]

resolution of the two types of oxygen, and the quadrupole coupling parameters for both can be obtained.[29]

Many different schemes and strategies for acquiring MQMAS spectra and processing the data obtained have been described,[5] and the technique will undoubtedly find growing applications in the area of glasses and geological materials. Comparisons of MQMAS with the earlier techniques (DOR and DAS) have been reported.[30–32]

2.10 Nuclear Magnetic Resonance Experiments at High Temperature

High-temperature NMR measurements present particular challenges. The technical details are not presented here but are reviewed elsewhere.[3,33] Static (wideline) NMR experiments allow temperatures up to about 1700 °C to be reached, while MAS above 600 °C is possible with a heated gas flow, but is still nonroutine. Laser heating allows even higher temperatures to be achieved (up to 2500 °C), although this approach presents other challenges such as controlling the homogeneity of heating, temperature calibration and the possibility of inducing changes in the sample or loss of sample during heating.

The sensitivity of the NMR experiment is reduced at higher temperatures owing to the less favorable Boltzmann population of the excited states, but this can be partly offset by reduced line widths in both liquids and solids due to increased mobility.[33]

2.11 Quantitative Solid-state Nuclear Magnetic Resonance

In many NMR experiments the signal intensity does not correlate quantitatively with the number of nuclei in the sample. This is especially the case in heterogeneous solids, although under carefully controlled conditions it is possible for signal integration to be quantitative. There are a number of contributing factors. First, for broad line signals in solids not all nuclei may be uniformly excited. Small flip-angle pulses can be used to ensure uniform response across the signal.[5] In samples with a wide range of relaxation times, some nuclei may relax too quickly (or too slowly) to be observed under the experimental conditions. In single-pulse wideline experiments there will be significant signal loss during the dead-time of the receiver. Even with MAS, if line widths are large as they are in glasses, this will remain a problem and differential signal loss may be obtained across the spectrum. Variable delays before signal acquisition can be used to detect the problem and in some cases to extrapolate back to 100% signal intensities, although this will not compensate for signals which have completely relaxed before detection can begin. Spin-echo techniques can be used to avoid the dead-time problem. Recycle delay times must also be long enough to avoid saturation of nuclei with long spin–lattice relaxation times. This may ultimately limit sensitivity by increasing acquisition times.

Internal standards have been used to calibrate changes in signal intensity in single-pulse experiments. A recent example is the use of AlN as the internal standard to calibrate the detected Al signals in as-synthesized versus calcined MCM-41[34] (Figure 7).

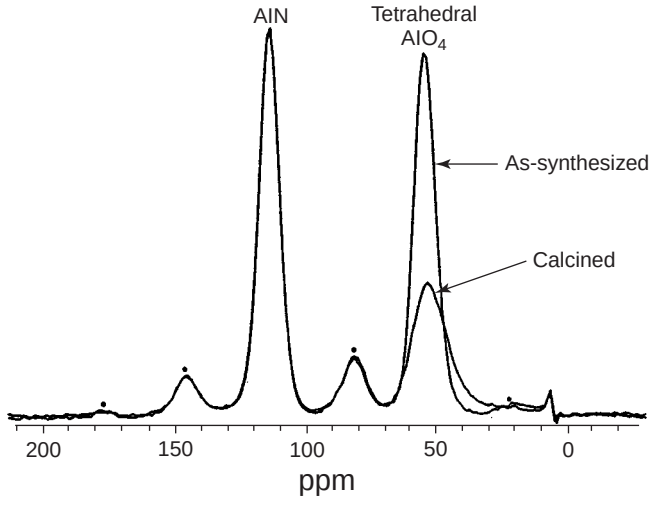

Figure 7 ^{27}Al NMR signals in as-synthesized versus calcined MCM-41a acquired using AlN as an internal standard. [Reproduced with permission from Janicke et al.[34]]

The determination of crystalline silica in iron oxide pigments[35] provides a good example of the experimental considerations required for calibration in ^{29}Si NMR. Single-pulse excitation with MAS spinning was used. Because of the long T_1 relaxation times for the silica, a reduced pulse angle of 10° with a 20-min relaxation delay between acquisitions was required. A blind test of silica standards was used for comparison, with X-ray diffraction analyses as a test of the protocol and tetrakis(trimethylsilyl)silane was employed as the spin-counting standard.

The use of CP to transfer magnetization from abundant to dilute nuclei leads to enhanced sensitivity but also to quantitation problems because the CP efficiency varies with the strength of the heteronuclear dipolar coupling. Combined with variations in proton relaxation times in the rotating frame ($T_{1\rho}$) and added difficulties at high MAS speeds, quantitative transfer of excitation is difficult. Detailed characterization of contact times and relaxation properties must be carried out and CP is often not used if quantitative analysis is required. A linear ramp of CP amplitudes in combination with high-speed MAS has been shown to improve quantitation in ^{1}H–^{13}C cross-polarization/magic-angle spinning (CPMAS) experiments.[36]

For quadrupolar nuclei of half-integer spin, there are additional complications, since it may be only the central transition that can be accurately observed. However, techniques for the correction of the center band to obtain the total signal intensity have been investigated. The current state-of-the art for the quantitation of half-integer quadrupolar nuclei has been recently reviewed by Smith and van Eck.[5] Finally, even when all experimental precautions have been taken, there remains the question of whether deconvolution of overlapping signals in order to obtain quantitative speciation in amorphous samples is valid. The deconvolution issue has been critically reviewed for ^{29}Si Q^n speciation in silicates[37] but the concerns raised are likely to be just as significant for speciation of other nuclei. Factor analysis has also been evaluated for separating the spectral components in ^{113}Cd MAS spectra consisting of overlapping isotropic peaks and is proposed as superior to least-squares fitting methods.[38]

The other issues of interest to analytical chemists are those of sensitivity and detection limit. As a point of comparison, it is interesting to compare the detection levels reported recently for silicone species by ^{29}Si distortionless enhancement by polarization transfer (DEPT) NMR *in solution* with detection limits for elemental methods of analysis such as inductively coupled plasma atomic emission spectrometry (ICPAES) or inductively coupled plasma mass spectrometry (ICPMS). On a standard commercial 17.61 T instrument, using an unenriched sample of hexamethyldisiloxane, the ^{29}Si detection limit was 150 ng mL^{-1}, requiring an acquisition time of 52 h.[39] For solids it is more difficult to generalize on the question of detection limits in NMR experiments, except that they will be substantially larger than those in solution. The sensitivity of an NMR experiment depends primarily on the nucleus (its abundance, gyromagentic ratio and quadrupole moment) and on the applied magnetic field and temperature. However, it is also affected by symmetry and relaxation effects in the sample. The concentrations of species which can be detected for a particular nucleus in the solid state depend therefore on both instrumental and sample matrix considerations, but between 10^{17} and 10^{20} spins are generally required. It should also be noted that capabilities of NMR for the detection of low concentrations are continually being improved.

2.12 Nuclear Magnetic Resonance Relaxation Experiments

2.12.1 Spin–Lattice Relaxation

When the nuclear spins in a sample are excited with an RF pulse, a magnetization M_z develops along the direction of the magnetic field. This magnetization relaxes exponentially back to the equilibrium value M_0. This process is called longitudinal or spin–lattice relaxation. It involves loss of excitation energy to the surrounding lattice and is characterized, at least in mobile liquids, by a single relaxation time T_1, the spin–lattice relaxation time. Spin–lattice relaxation can occur by a number of different mechanisms, but all require fluctuations in the local environment of the nucleus at frequencies comparable to the Larmor frequency, w_0. Thus spin–lattice relaxation is sensitive to atomic and molecular motion and is used as a probe of mobility in both liquids and solids. Spin-$\frac{1}{2}$ nuclei relax predominantly by dipolar interactions, whereas in quadrupolar nuclei ($I > \frac{1}{2}$) interactions with fluctuating electric field gradients contribute to relaxation. Detailed discussion of relaxation mechanisms is given in many NMR texts and monographs and is not repeated here.[40,41] In simple systems, the relationship between the relaxation rate $1/T_1$ and the correlation time for molecular motion τ_c is given by Equation (6):

$$\frac{1}{T_1} = (\text{constant}) \times \frac{\tau_c}{1 + w_0^2\tau_c^2} \tag{6}$$

In solids, where mobility is low and $\tau_c \gg 1/w_0$, T_1 decreases with increasing temperature. A minimum in T_1 is reached when $\tau_c \approx 1/w_0$ and relaxation becomes slower again as mobility increases further. In the motional narrowing region, where $\tau_c \gg 1/w_0$, T_1 shows

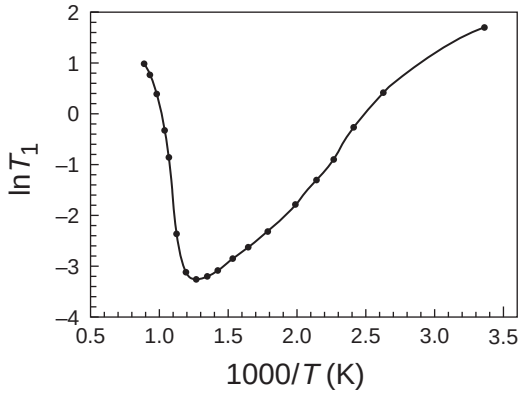

Figure 8 Plot of $\ln T_1$ versus inverse temperature for ^{7}Li in Li_2SO_4. [Reproduced with permission from Xu and Stebbins.[42]]

the reverse temperature dependence. This is illustrated by results for ^{7}Li spin–lattice relaxation in Li_2SO_4 (Figure 8).[42]

If the temperature dependence of τ_c follows Arrhenius-type behavior with an activation energy E_a, then T_1 data can be used to extract the activation energy. This is often assumed as a model for interpreting T_1 results, with activation energies derived from plots of $\ln T_1$ versus inverse temperature both above and below the T_1 minimum. Not surprisingly, in many samples, and especially in glasses, deviations from this simple model are observed since assumptions of single correlation times are unlikely to be valid. For the Li_2SO_4 example shown in Figure 8, although the individual ^{7}Li relaxation times were well fitted by single exponential fits, the Arrhenius plot was asymmetric, indicating different apparent activation energies in the fast and slow motion regions.[42] This effect is commonly observed in ion-conducting materials.

Apart from its utility as a probe of local motion, a knowledge of relaxation behavior is important for measuring NMR spectra. A recycle delay of $5T_1$ is needed between pulse sequences to avoid saturating the signals of slowly relaxing nuclei. On the other hand, if relaxation is too fast, this contributes to line broadening. Spin–lattice relaxation times are generally measured by the inversion–recovery pulse sequence $180°-\tau-90°$. It can be difficult to excite broad signals uniformly in solids with a 180° pulse and an alternative method, saturation recovery, $(90°)_n-t_1-90°$, can be used.[40] Relaxation time can also be used as the basis for contrast in NMR imaging.[41]

2.12.2 Spin–Spin Relaxation

When the spins in the sample are subjected to a 90° pulse, the z magnetization is rotated into the xy plane. The decay of this magnetization in the xy plane is known as transverse or spin–spin relaxation. It corresponds to an exchange of energy among nuclei with slightly different resonance frequencies. The RF signal in the xy plane gradually loses phase coherence and decays to zero and this is seen directly in the decay of the FID. The spin–spin relaxation time directly affects the line width of the NMR signal and estimates of T_2 can be obtained from the line width at half-height. T_2 is generally measured using modifications of the Carr–Purcell pulse sequence.[40] In liquids, under motional narrowing conditions, $T_1 = T_2$. However, in solids T_2 will be much shorter than T_1. Complex relaxation behavior is also observed for liquids in confined-pore materials.[41]

2.12.3 Spin–Lattice Relaxation in the Rotating Frame

Under conditions of CP, the abundant spin is locked in the rotating frame. The rate at which the magnetization in the rotating frame decays is characterized by a relaxation time $T_{1\rho}$, and is sensitive to fluctuating local fields in the kilohertz range. In CP experiments, $T_{1\rho}$ for the abundant spins can be determined from plots of signal intensity versus contact time. An initial increase in intensity is observed determined by the CP time constant, with a decay in signal intensity at longer contact times due to rotating frame relaxation. The efficiency of CP in solids thus depends on the relative rates of these two processes. In ^{29}Si–^{1}H CP in silicates or aluminosilicates, for example, the proton concentrations may be low, making T_{SiH} relatively long and comparable to the proton $T_{1\rho}$. This affects the maximum ^{29}Si intensity which can be achieved and hence limits the sensitivity of the experiment. In heterogeneous systems, a single $T_{1\rho}$ may not always be found and this can provide evidence for the existence of multiple phases since the relaxation is not averaged by spin diffusion.

2.13 Magnetic Resonance Imaging

MRI, like NMR spectroscopy itself, is not a single technique but covers a wide range of capabilities from qualitative imaging of fluids based on relaxation contrast to parameter mapping, where quantitative 3-D maps of specific parameters such as relaxation time or spin density are obtained.[41,43] MRI has become routinely available outside the medical field in recent years. Materials analysis, particularly studies of fluids in porous materials and fluid diffusion into bulk materials, has been a major area of interest.[44] While the majority of MRI is still of mobile ^{1}H nuclei, imaging of other nuclei and of solids is expanding rapidly, offering even more powerful methods for the NMR analysis of geological and glassy materials.

The basic imaging experiment relies on a linear magnetic field gradient to introduce position-dependent frequency shifts enabling signals from different parts of the sample to be distinguished. Gradients can then be varied in the x, y and z directions to produce 3-D images. Spin–lattice relaxation (T_1) maps are commonly used for water and other mobile fluids. The spatial resolution is limited in practice to a few micrometers in liquids and an order of magnitude larger in solids, because of line width, signal-to-noise ratio and diffusion effects.

Contrast techniques for solid-state imaging enable a much wider range of inorganic materials to be investigated. Imaging in solids is inherently more difficult owing to dipole–dipole and chemical shift interactions leading to large line widths and low signal-to-noise ratios. The lower sensitivity ultimately limits the resolution which can be obtained. This problem can be reduced by a range of strategies including the use of larger magnetic field gradients and by line-narrowing techniques. Multiple-pulse methods used to obtain high resolution in conventional NMR spectra of solids have been incorporated into imaging from its early development. A large range of line-narrowing methods is now in use, including MAS and magic angle in the rotating frame and multiple pulse techniques such as CRAMPS.[44]

Conventional MRI techniques are eventually limited by the rate of spin–spin relaxation though continuing developments in probe technology as well as pulse sequences are extending its capabilities to systems with shorter T_2 values. The complementary approach of stray field imaging (STRAFI) extends the applications of imaging in solids even further.[45–47] STRAFI makes use of the very large gradients in the fringe field of superconducting magnets. The RF pulse excites a narrow slice of the sample orthogonal to the gradient direction. The sample is then moved through the gradient and a series of slices is measured using echo detection, giving a 1-D profile. Rotation of the sample can be used to give 2-D or 3-D images. However 3-D imaging using STRAFI is still technically difficult since it requires precise translation and rotation of the sample, with reconstruction of the image from individual slices by back-projection. STRAFI is successful for signals over a very wide range of T_2 values including very broad signals. Apart from ^{1}H, its application to a number of other nuclei of importance in inorganic materials has been demonstrated,[45–47] including ^{19}F, ^{11}B, ^{23}Na, ^{27}Al and ^{7}Li.

Finally, another development of importance to the imaging inorganic solids by NMR should be mentioned. This is NMR force microscopy,[48,49] which has the potential to achieve higher spatial resolution in solids than conventional imaging.

3 APPLICATIONS

3.1 Speciation and Structure Determination

Tables 4–10 summarize the applications of NMR to speciation and structure determination in glasses, gels, ceramics and minerals. The examples in the tables are organized according to the nucleus studied. In the following sections we discuss selected applications to the different classes of materials.

3.1.1 Silicate and Aluminosilicate Glasses

^{29}Si NMR has been extensively applied to these systems for many years, and the subject has been comprehensively reviewed on several occasions.[1,50–52] The isotropic chemical shift for a ^{29}Si atom depends on both its chemical environment and on the local geometry; ^{29}Si spectra can thus provide information about both aspects of glass structure.

For tetrahedrally coordinated silicon, the chemical shift depends on the number of bridging oxygen atoms attached to the silicon. The standard terminology used is to identify the silicon sites as Q_1, Q_2, Q_3 or Q_4, where the subscript refers to the number of bridging oxygens connected to other tetrahedral sites. For Q_1, Q_2 and Q_3 silicons, the remaining coordination sites are occupied by hydroxyl groups. Approximate chemical shift ranges for each type of silicon are as follows (relative to tetramethylsilane):[52]

Q_4	$Si(OSi)_4$	−(100–116) ppm
Q_3	$Si(OSi)_3OH$	−(85–100) ppm
Q_2	$Si(OSi)_2(OH)_2$	−(80–105) ppm
Q_1	$Si(OSi)(OH)_3$	−(70–85) ppm

A second influence on the ^{29}Si chemical shift is the substitution of silicon by other tetrahedral elements, particularly aluminum.The chemical shift becomes less negative with increasing replacement of adjacent silicons by aluminum, thus:

Q_4 (0Al)	$Si(OSi)_4$	−(100–116) ppm
Q_4 (1Al)	$Si(OSi)_3OAl$	−(96–106) ppm
Q_4 (2Al)	$Si(OSi)_2(OAl)_2$	−(92–102) ppm
Q_4 (3Al)	$Si(OSi)(OAl)_3$	−(88–96) ppm
Q_4 (4Al)	$Si(OAl)_4$	−(80–90) ppm

Speciation of silicon in a silicate or aluminosilicate requires resolution of the different components in the spectrum. In the case of glasses, the lines are relatively broad, and curve-fitting techniques must be employed. Mahler and Sebald[53] have given a critique of the assumptions used in line shape fitting. Most authors have assumed a Gaussian line shape function, but Mahler and Sebald showed by means of selected examples that it

Table 4 Applications of ^{29}Si NMR for speciation and structure determination

Sample or experiment type	Example	NMR technique	Analytical information	Ref.	Other related studies
Inorganic composites	Si in SiO_2	MAS		86	153
Hybrid inorganic–organic materials	TEOS–organic polymer composites	MAS	Silicon coordination	154	155–157
	Polysilsesquioxane materials	MAS	Silicon coordination and connectivity	158	159–163
	Polysilsesquioxane–transition metal composites	MAS	Silicon coordination and connectivity	164	165
	Alkoxysilane–TEOS composites	MAS, CP, 2-D HETCOR	Silicon coordination	166	
	Silicon oxycarbide ceramics	MAS, CP	Silicon coordination and connectivity	167	168–170
Aluminosilicates	Ceramics	MAS	Si, Al ordering	98	171, 172
	Minerals	MAS, CP	Si, Al ordering	101	173
	Glasses and gels	MAS, CP	Si, Al ordering	55	174–177
Magnesium silicates	Crystallization of ceramics	MAS	Si environment	178	179, 180
Titanium silicates	Mixed oxides	MAS	Si environment, evidence for Si–O–Ti bonds	90	174, 181, 182
Zirconium silicates	Microporous frameworks	MAS	Si environment	183	
	Polymer to ceramic conversion	MAS	Si coordination	184	87, 185
	Mixed oxides	MAS, CP	Si environment	174	186
Silicon nitrides	Glasses	MAS	Evidence for Si–N bonds	175	145, 187
Silicon phosphates	Ceramics	MAS	Si environment, evidence for 6-coordinate Si	188	
Minerals	Characterization	MAS, CP	Phase identification, Si, Al ordering	100	173, 189–191
	Phase transitions	MAS	Phase identification, Si, Al distribution	192	193–195
	Modification of clays	MAS	Si, Al distribution	196	101, 197–201
Chemical shift–structure correlations	Oxides, nitrides and carbides	MAS	Structure	56	
	Glasses	MAS	Structure	53	
Thin films	Silica	MAS	Structure	202	203, 204

TEOS, tetraethoxysilane.

may not be possible to obtain a single unique solution to the problem of fitting a ^{29}Si NMR spectrum with several components. Not withstanding this cautionary conclusion, curve fitting is still used to obtain at least semiquantitative distributions of silicon species. For example, Fayon et al.[54] examined a series of lead silicate glasses of varying lead contents. In this case the observed ^{29}Si spectra contain up to five components which are assigned to silicon atoms bonded through oxygen to four, three, two, one or no neighboring silicon atoms. Figure 9 shows the distribution obtained.

At high lead contents, the dominant silicon species are those bridged to two, one or no Si atoms, corresponding to $Si_3O_{10}^{8-}$, $Si_2O_7^{4-}$ and SiO_4^{4-} anions respectively. Comparison with ^{207}Pb NMR spectra for the same series of samples leads to the conclusion that as the lead content is increased the silicate network is destroyed and replaced by a polymeric lead oxide chain structure.

The sensitivity of the ^{29}Si chemical shift to the number of aluminum next-nearest neighbors in aluminosilicate glasses has been used recently by Lee and Stebbins[55] to investigate the degree of aluminum avoidance in such materials. The often-cited Loewenstein's rule postulates that the Si–O–Al linkage is more favorable than the alternative Si–O–Si and Al–O–Al combination. In crystalline aluminosilicates, Loewenstein's rule seems to be well established, but in the case of glasses the extent to which it is followed is still an open question. This is largely due to the difficulties of resolving the signals due to different $Si(OAl)_n$ species in glass samples. The approach taken by Lee and Stebbins was to fit the broad asymmetric single peaks with up to five Gaussian peaks corresponding to the five possible Q_4 Si species with 0–4 Al next-nearest neighbors. A speciation model was used in which the relative proportion of Al–O–Si, Si–O–Si and Al–O–Al linkages were calculated as a function of glass composition, assuming the quasi-chemical approximation (which assumes the number of each species depends on their relative energy differences). Spectra were then fitted with different possible values of the aluminum

Table 5 Applications of ^{27}Al NMR for speciation and structure determination

Sample type	Example	NMR technique	Analytical information	Ref.	Other related studies
Glasses	Aluminophosphate	MAS	Al coordination	105	205–207
	Aluminophosphate	TRAPDOR double resonance	Evidence for AlOH	208	
	Aluminosilicate	Nutation	Quadrupole parameters, Al coordination	62	
	Aluminosilicate	Satellite transition spectroscopy	Quadrupole parameters, Al coordination	64	
	Aluminoborate	REDOR	Al–O–B connectivity	78	
	Cordierite	MAS, nutation, satellite transition spectroscopy	Quadrupole parameters, Al coordination	177	180
	SiAlONs	MAS	Al coordination	175	145, 172
Gels	Alumina	MAS	Al coordination	209	210–215
	Aluminosilicate	MAS	Al coordination	216	174, 217
	Aluminosilicate	MQMAS, nutation	Quadrupole parameters, Al coordination	88	
	Aluminophosphate	MQMAS	Al coordination	218	
	Other alumina–oxide systems	MAS	Al coordination	174	219
Ceramics	Mullite	MAS	Al coordination	193	220
	Mullite	Satellite transition spectroscopy	Quadrupole parameters, Al coordination	221	
	Mullite	MQMAS	Quadrupole parameters, Al coordination	222	
	Melilite and ALON ceramics	MAS	Al coordination	223	187, 224, 225
	Synroc	MAS	Al coordination	226	
	Hollandite	MAS	Al coordination	227	228
Minerals	Cement phases	MAS	Quantitative analysis	229	230
	Other minerals	MAS	Al coordination	231	173, 191, 232
	Andalusite and kyanite	MAS, MQMAS	Al coordination	21	
Clays	Various	MAS	Al coordination and quantification	233	190, 196, 198, 200, 201, 231, 234–238
	Layered double hydroxides	2-D	Al coordination	239	

Table 6 Applications of ^{17}O NMR for speciation and structure determination

Sample type	Example	NMR technique	Analytical information	Ref.	Other related studies
Gels	Aluminosilicate	MAS	Local structure	240	
	Lanthana	MAS, satellite transition spectroscopy	Local structure, quadrupole parameters	106	241
	Titanosilicates	MAS	Local structure	90	242, 243
	Siloxane–oxide hybrid materials	MAS	Local structure	87	244
Ceramics	Nanophase MgO	MAS	Local structure	245	
	Mixed metal oxides	MAS	Local structure, chemical shift-structure correlations	246	
Minerals	Stilbite	Static, MAS, DAS, MQMAS	Local structure	102	

avoidance parameter Q ($Q = 0$ represents a totally random distribution of Si and Al, taking no account of Loewenstein's rule, and $Q = 1$ perfect compliance with Loewenstein's rule).

For calcium aluminosilicate glasses, the observed spectra could be fitted over the complete compositional range with values of Q between 0.8 and 0.875. For sodium aluminosilicate glasses, on the other hand, values of Q closer to 1.0 (0.93–0.99) were obtained. In both types of glass there is clearly a very substantial degree of aluminum ordering present, representing the structure of the liquid at the glass transition temperature.

Table 7 Applications of ^{31}P NMR for speciation and structure determination

Sample type	Example	NMR technique	Analytical information	Ref.	Other related studies
Glasses	Phosphate glasses	Static, MAS	Chemical shift tensor components, local structure	81	207, 208, 247
Ceramics	Phosphate ceramics	MAS	Local structure, phase identification	248	79, 205, 249, 250
	Phosphate ceramics	2-D	P connectivity	251	82, 105, 247
	NASICON	MAS	Local structure	252	253
Gels	Silicophosphate	MAS	Local structure	254	
	Aluminum polyphosphate thermoreversible	MAS	Local structure	255	
	Hybrid materials	MAS	Local structure	158	
Minerals	Viseite	MAS	Phase identification	191	

Table 8 NMR Applications of other framework elements for speciation and structure determination

Nucleus	Sample type	Example	NMR technique	Analytical information	Ref.	Other related studies
^{9}Be	Ceramics	Beryllium silicates, germanates	MAS	Local structure	256	
^{11}B	Glasses	Borosilicates	Satellite transition spectroscopy	Speciation, quadrupole parameters	77	
		Borosilicates	Nutation	Speciation, quadrupole parameters	257	
		Boroaluminates	CPMAS, REDOR	B–O–Al connectivity	78	
		Lithium triborates	Static, MAS	Coordination, local structure	258	
		Boronsilicon oxycarbides	MAS	Local structure	259	260
	Gels	Hybrid borosilicates	MAS	Local structure	89	
	Ceramics	Indialite	MAS	Local structure	261	
		Mixed oxides	MAS	Local structure	262	
^{25}Mg	Minerals	Various	Static spin echo	Coordination number	263	264
	Minerals	Various	MAS	Local structure	265	266–271
^{47,49}Ti	Ceramics	Various	Static spin echo	Local structure	272	273, 274
^{51}V	Gels	Mixed oxides	MAS	Local structure	275	276, 277

Table 9 Applications of ^{19}F NMR for speciation and structure determination

Sample type	Example	NMR technique	Analytical information	Ref.	Other related studies
Glasses	Fluorophosphates	MAS	F coordination	123	278
	Aluminum fluoride	MAS	F coordination and structure	279	
	Boron oxyfluoride	MAS	F coordination and structure	280	281
	Transition metal fluoride	MAS	F coordination and structure, quantification	85	282–284
Ceramics	Fluorapatite	MAS	F quantification	285	
Clays	Various	MAS	F coordination	286	

The low resolution in ^{29}Si NMR spectra of glasses is due to the presence of a distribution of Si–O–Si bond angles. Theoretical calculations indicate a spread of as much as 40° around a mean of ca. 145°.[53] Correlations of chemical shift with bond angle are well established in the crystalline silicate and aluminosilicate literature, and supported by quantum mechanical calculations.[56] The possibility of obtaining experimental evidence for Si–O–Si bond angle distributions from ^{29}Si NMR spectra might therefore be considered. In practice, however,

Table 10 NMR spectra of mobile and exchangeable ions: speciation and structure determination

Nucleus	Sample type	Example	NMR technique	Analytical information	Ref.	Other related studies
^{23}Na	Glasses	Sodium silicate	Spin echo decay	Distribution of Na–Na distances	71	
		Sodium borate and germanate	MAS	Local structure	287	105
	Ceramics	NASICON	MAS	Local structure	288	
	Minerals	Zeolites	MAS	Cation-exchange sites	289	290
		Zeolites	MQMAS	Cation-exchange sites	103	183, 291
		Clays	MAS and DOR	Cation-exchange sites	292	104, 293
^{7}Li	Glasses	Lithium borate	Static, MAS	Li coordination	258	
	Gels	Ormolyte hybrid gels	MAS	Li mobility	294	
	Ceramics	Nanocrystalline anatase	MAS	Li sites	295	
		Various	Static	Quadrupole parameters, Li sites	296	
		Lithium aluminate	Spin–lattice relaxation times	Impurity content	108	
	Clays	Montmorillonite	MAS	Ion-exchange sites	297	298
^{6}Li	Glasses	Lithium silicate	MAS	Li coordination	70	
	Ceramics	Nanocrystalline anatase	MAS	Li sites	295	
^{133}Cs	Clays	Various	MAS	Ion-exchange sites	299	104, 293
^{1}H	Glasses	Hydrous silicates	MAS, CRAMPS, 2-D	Proton speciation	119	
^{109}Ag	Glasses	Silver tellurite	MAS	Ag coordination and mobility	300	301, 302
^{113}Cd	Clays	Various	MAS	Cation-binding sites	104	
^{209}Pb	Glasses	Lead silicates	MAS	Local structure	54	303
		Lead silicates	MAS	Local structure and bonding state	304	
	Ceramics	Titanate/zirconates	MAS	Local structure	107	305
		Lead silicates	Static, MAS	Chemical shift–structure correlation	306	

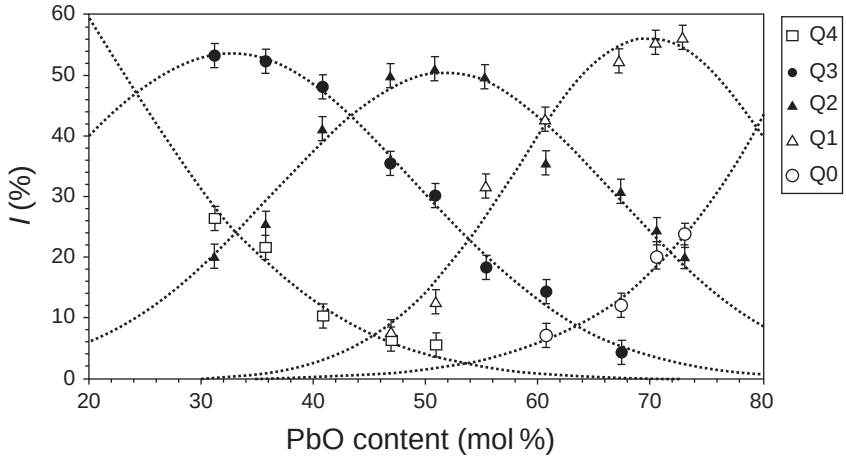

Figure 9 Distribution of Q_n silicon sites as a function of PbO content in lead silicate glasses. [Reproduced with permission from Fayon et al.[54]]

the uncertainties associated with curve fitting and the difficulty of separating structural from chemical influences on chemical shift preclude any reliable conclusions being drawn. As discussed below, ^{17}O NMR, which observes the bridging oxygen atoms directly, may offer more hope of studying bond angle distributions.

^{29}Si NMR spectroscopy can easily distinguish five- and six-coordinate Si in glasses containing these unusual species. Octahedrally coordinated Si has chemical shifts between −180 and −221 ppm,[57] while pentacoordinated Si has been detected in small concentrations in certain cesium and sodium silicate glasses at a chemical shift of about −150 ppm.[58]

As noted in section 2.2.6 (and discussed further below), 2-D correlation ^{29}Si NMR spectroscopy has proved to be a powerful tool for studying connectivity of Si atoms

in crystalline silicates and aluminosilicates, particularly zeolites. Knight et al.[59] first described a ^{29}Si COSY experiment on silicate glasses which had been isotopically enriched with ^{29}Si. They showed the first direct evidence that Q_3 and Q_4 Si sites in sodium silicate glasses were connected; the intensities of the cross peaks in the 2-D COSY spectra arising from ^{29}Si–^{29}Si dipolar coupling decreased as the level of ^{29}Si enrichment was lowered.

More recently, an alternative approach for measuring Si connectivity has been described by Glock et al.[60] This uses the 2-D double quantum NMR experiment first described by Fieke et al.[61] In this experiment, only dipolar coupled ^{29}Si nuclei are detected (isotopic enrichment was used to enhance sensitivity). Peaks due to double quantum coherences between chemically identical spins appear on the diagonal of the 2-D spectrum, and off-diagonal peaks are due to double quantum coherences between chemically different sites. For a ^{29}Si-enriched sodium silicate glass, Glock et al. identified three different Q_3 Si sites at about −89, −92 and −95 ppm, assigned to Q_3 silicon atoms connected to $3Q_3$ silicons, $2Q_3$ and $1Q_4$ silicons and $1Q_3$ and $2Q_4$ silicons respectively. Likewise, the three different Q_4 sites resolved at about −102, −105 and −108 ppm were assigned to Q_4 silicon atoms connected to different numbers of Q_3 and Q_4 silicons. The power of this technique to identify connectivities at a level of resolution impossible in conventional one-pulse MAS NMR spectra offers many possibilities for the study of silicate glasses (the authors have also used this method to study phosphate glasses, as described below).

^{27}Al NMR spectroscopy has provided many insights into the structure and speciation of aluminosilicate glasses. The effects of quadrupolar broadening were described in section 2.8. In glasses, the additional broadening caused by structural disorder makes analysis of conventional one-pulse MAS ^{27}Al NMR spectra difficult. The presence of four-, five- and six-coordinate aluminum (chemical shifts at −10 to 15, 35 to 40 and 50 to 80 ppm, respectively, in aluminosilicates) can often be qualitatively determined by curve fitting asymmetric line shapes. Several of the newer NMR techniques for enhancing ^{27}Al resolution and extracting structural information from quadrupole parameters have been successfully applied to glass systems.

For example, Dirken et al.[62] used the off-resonance nutation technique to determine the quadrupole parameters for ^{27}Al in alkali metal aluminosilicate glasses. The conventional MAS spectra of these samples show a single broad peak at the chemical shift expected for tetrahedrally coordinated aluminum, with a line width which decreases with increasing size of the alkali metal cation. The contributions to the line width from a distribution of chemical shifts and from second-order quadrupolar broadening cannot be easily separated, however (unless spectra are recorded at several different field strengths). As described in section 2.9, nutation is a 2-D experiment in which the f_1 dimension contains the nutation spectrum, which is a sensitive function of the quadrupole parameters only. By comparing the observed nutation spectra with those computer simulated for different possible quadrupole parameters, the mean coupling constants and asymmetry parameters could be obtained.

The mean coupling constants in this series of glasses were found to correlate linearly with the polarizing power of the alkali metal cation, suggesting that the quadrupole interaction is dominated by the perturbation of electron density around Al by a nearby alkali metal ion. The values of the quadrupole parameters obtained from the nutation spectra were then used to calculate the conventional MAS line shapes and deduce the isotropic chemical shift and chemical shift dispersion. The isotropic chemical shifts were found to decrease with increasing size of the cation, which may indicate changes in the Al–O–Si bond angle,[63] but the chemical shift dispersion (reflecting the degree of structural disorder) remained approximately constant. The authors concluded that ^{27}Al NMR of these systems is much more structurally informative than ^{29}Si NMR.

An alternative approach which has been applied with some success to ^{27}Al in disordered systems is satellite transition spectroscopy. As noted above, the central ($\frac{1}{2} \leftrightarrow -\frac{1}{2}$) transition for a nonintegral quadrupolar nucleus is broadened owing to second-order quadrupolar interactions and to chemical shift dispersion. The satellite transitions (e.g. $\frac{3}{2} \leftrightarrow \frac{1}{2}$) show higher resolution, however, which can allow the isotropic chemical shift and quadrupole coupling constants to be obtained with a greater degree of confidence. For example, Figure 10(a–f) compares the conventional (central transition) MAS spectra of several potassium aluminosilicate glasses with differing potassium contents with the corresponding satellite transition spectra.[64] The signals due to tetrahedral and octahedral Al are clearly much better resolved in the satellite transition spectra, and can be readily quantified. Furthermore, in two of the samples there is partial resolution of two different tetrahedral sites with slightly different isotropic chemical shifts. The quadrupole coupling constants calculated from the spinning side band patterns decreased on adding potassium.

Direct structural information about silicate and aluminosilicate glasses is obtained from observation of ^{17}O NMR spectra. The Si, Al ordering in such glasses can be investigated since the ^{17}O quadrupolar coupling constants are significantly different for Si–O–Si, Si–O–Al and Al–O–Al linkages. An example from Dirken et al.[29] was shown in Figure 6. In this particular example of an albite glass, the conventional single quantum ^{17}O MAS spectrum shows two overlapping signals which cannot

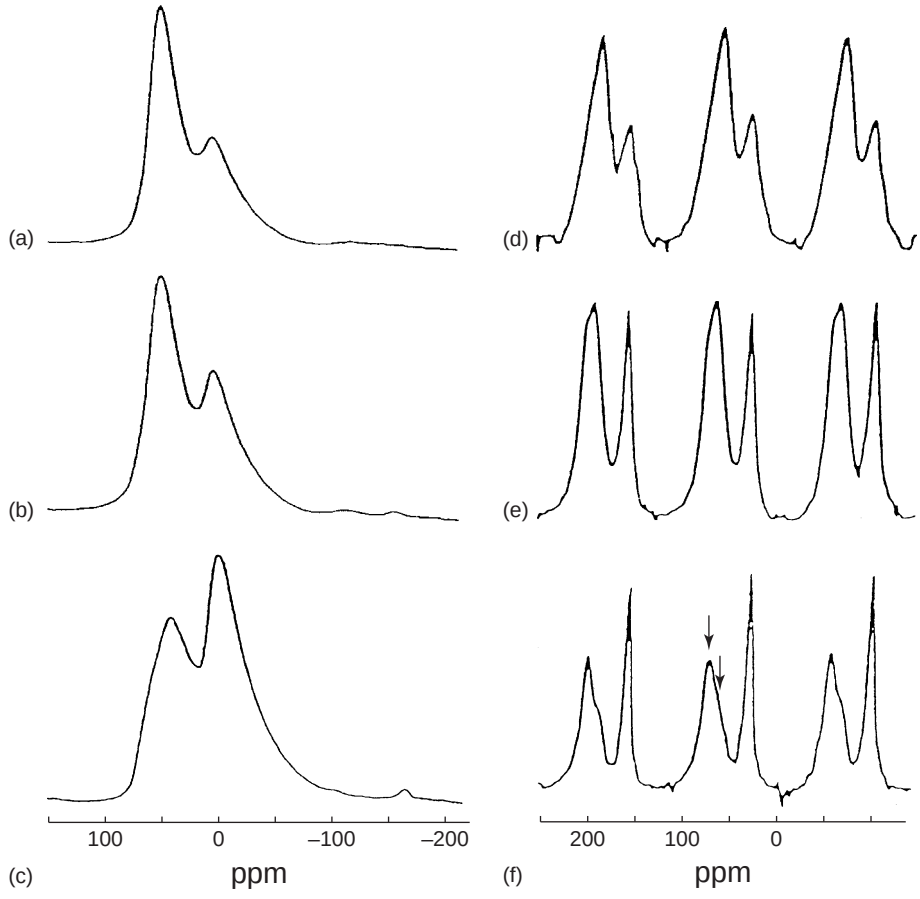

Figure 10 (a–c) Central and (d–f) satellite transition ^{27}Al spectra of potassium aluminosilicate glasses with different K : Al ratios. (a), (d), K : Al = 0.25; (b), (e), K : Al = 0.15; (c), (f), K : Al = 0.0. Arrows mark different tetrahedral sites. [Reproduced with permission from Mundus and Mueller-Warmuth.[64]]

be quantified. The triple quantum spectrum (Figure 6) clearly separates two signals which can be assigned to Si–O–Si and Si–O–Al bridging oxygens; the single quantum spectrum can then be fitted with the sum of these two contributing signals to obtain a reliable estimate of the relative numbers of the two types of bridging oxygen.

Related studies of oxygen speciation have been reported by Stebbins et al.[65–67] For example, in mixed barium–calcium silicate glasses the MQMAS experiment separates bridging (Si–O–Si) from terminal (Si–O) oxygen atoms. The terminal Si–O oxygen projections on the isotropic chemical shift axis for the mixed glasses could be fitted assuming a random distribution of Ba and Ca on the three cation sites adjoining each terminal oxygen, i.e. there is no evidence for cation ordering in these particular materials.[65] In a calcium aluminosilicate glass, the ^{17}O MQMAS experiment finds direct evidence for Al–O–Al linkages in violation of Loewenstein's rule,[66] whereas in hydrous silicate glasses at least four different oxygen species have been detected (Si–O–Si, Si–O, Si–OH and H_2O).[67]

NMR spectroscopy has also been applied to study the cations in silicate and aluminosilicate glasses and also the framework elements. Table 10 lists some of the NMR-visible nuclei which have been studied in this way. In the case of protons, Schaller and Sebald[68] applied the CRAMPS technique to observe OH and H_2O species in hydrous silicate glasses. The line widths of the two signals at about 12 and 4 ppm were unchanged, however, between the CRAMPS experiment, in which dipolar decoupling is applied, and the conventional MAS experiment, indicating that the line width is mainly due to chemical shift dispersion. The presence of dipolar coupling between the OH and H_2O sites was shown by the observation of cross peaks in a 2-D spin exchange spectroscopy (EXSY) experiment. Schaller and Sebald also showed that the spinning side band patterns associated with the two kinds of proton sites could be separated by carrying out a 2-D CRAMPS/MAS correlation experiment in which the f_2 axis is the MAS spectrum and the f_1 axis is the CRAMPS spectrum. Figure 11 shows the resulting 2-D spectrum; the spinning side band pattern for each of the two sites can

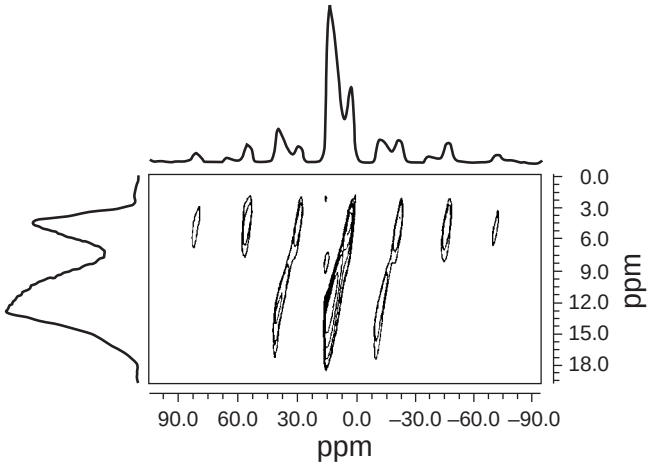

Figure 11 ^{1}H CRAMPS/MAS correlation spectrum of a hydrous sodium silicate glass. [Reproduced with permission from Schaller and Sebald.[68]]

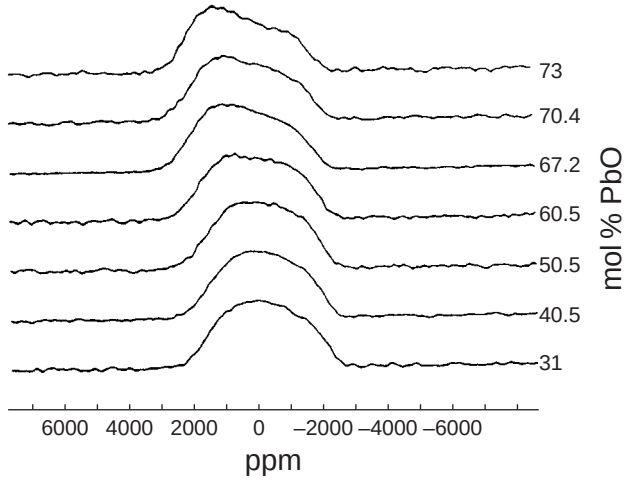

Figure 12 ^{207}Pb static spectra of lead silicate glasses. [Reproduced with permission from Fayon et al.[54]]

then in principle be analyzed to obtain dipolar coupling constants.

The local environment and spatial distribution of alkali metal ions in silicate and aluminosilicate glasses has been studied by ^{23}Na, ^{7}Li and ^{6}Li NMR. For ^{23}Na and ^{7}Li the line widths in conventional MAS NMR spectra are dominated by quadrupolar effects, and direct chemical information is difficult to obtain. Nevertheless, Ali et al.[69] were able to show for a series of mixed sodium–lithium silicate glasses that the apparent ^{23}Na and ^{7}Li chemical shifts and line widths varied continuously across the composition range, concluding that the two cations were randomly mixed with no ordering occurring. In the case of ^{6}Li, the quadrupole moment is sufficiently small that the observed center of gravity in a MAS spectrum is very close to the isotropic chemical shift, i.e. this nucleus behaves effectively as though it were spin $= \frac{1}{2}$. ^{6}Li chemical shifts in a range of different crystalline environments have been established; there is an approximately linear correlation between ^{6}Li chemical shift and coordination number (e.g. from 1.5 ppm relative to aqueous LiCl for three-coordinate Li to −1.5 ppm for eight-coordinate Li).[70] In a lithium silicate glass, the ^{6}Li spectrum showed a broad peak indicating a distribution of coordination numbers between 4 and 6.

Gee and Eckert[71] used ^{23}Na spin echo decay spectroscopy to measure ^{23}Na–^{23}Na dipole–dipole couplings in sodium silicate glasses. Spin echo decay spectroscopy measures the intensity of a spin echo as a function of the evolution time; in these systems the mechanism responsible for spin echo decay is assumed to be dipolar coupling. The observed decay can be fitted to obtain a ^{23}Na–^{23}Na second moment, which contains the dipolar coupling. The resulting second moments, when plotted against the sodium content of the glass, best fitted a model of homogeneous distribution of sodium cations rather than a random distribution.

The double resonance techniques of REDOR and SEDOR have been applied to investigate the ^{23}Na/^{7}Li or ^{23}Na/^{6}Li distributions in mixed sodium–lithium silicate glasses,[72–74] but in these materials the data could be best fitted by assuming a random distribution of the two cations.

The investigation of lead silicate glasses by ^{29}Si NMR was described above. The ^{207}Pb signals from these materials are broad and poorly defined, reflecting the large chemical shift range shown by ^{207}Pb. Nevertheless, as shown in Figure 12, there is a definite evolution in line shape and position as the lead content of the glasses changes.

Up to 50 mol% PbO the isotropic chemical shift remains constant, at a value consistent with chains of distorted PbO_n pyramids; at higher loadings the increasing chemical shift was interpreted in terms of formation of Pb–O–Pb units.

3.1.2 Germanate and Borate Glasses

Germanium dioxide shows many structural similarities to silica, and can form similar glassy structures. Hussin et al.[75] recently reported an ^{17}O NMR investigation of a GeO_2 glass to determine the Ge–O–Ge bond angle distribution. Samples were enriched in ^{17}O to enhance sensitivity, and spectra were collected in static and MAS modes. The spectra (both static and MAS) of the glass differed from those of the crystalline quartz-like GeO_2 only in line width. The glass line shapes were fitted with a model assuming a Gaussian distribution of Ge–O–Ge bond angles; it was concluded that the distribution is fairly

narrow, around a mean of 130° with a half-width of no more than 4°.

In the case of borate or borosilicate glasses, ^{11}B NMR can provide useful structural information. The relative concentrations of three- and four-coordinate boron can easily be determined from static and MAS spectra, and the DAS technique has been applied to enhance resolution.[76] As in the case of aluminum, resolution enhancement can also be achieved using ^{11}B satellite transition spectroscopy. Jäger et al.[77] examined a series of sodium borosilicate glasses with this technique, and identified two different BO_4 sites and three different BO_3 sites.

Double resonance $^{11}B-^{27}Al$ and $^{27}Al-^{11}B$ experiments on sodium aluminoborate glasses have been reported by Van Wuellen et al.[78] The single-pulse ^{11}B MAS spectra of these glasses show two overlapping signals due to three- and four-coordinate boron. Application of ^{27}Al CP enhances only the three-coordinate boron signal, showing unambiguously the connectivity of BO_3 and AlO_4 sites. Likewise, a ^{11}B (^{27}Al) REDOR experiment showed that there is a much larger dipolar coupling between ^{27}Al and the three-coordinate boron than the four-coordinate boron.

In borosilicate glasses, the assumption is generally made that three-coordinate boron does not mix with the silica network. Wang and Stebbins[79] have recently shown by triple quantum magic angle spinning (3QMAS) ^{17}O NMR that this assumption is not correct. Figure 13 shows the triple quantum spectrum of a ^{17}O enriched borosilicate glass containing 40 mol% B_2O_3.

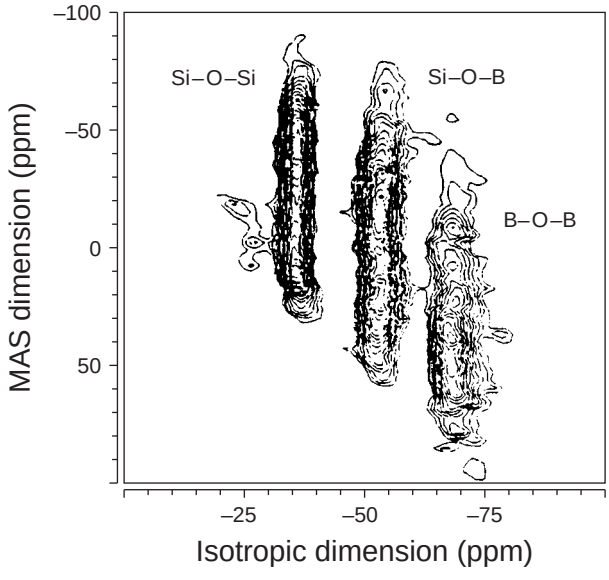

Figure 13 2-D MQMAS ^{17}O spectrum of borosilicate glass with 40 mol% B_2O_3. [Reproduced with permission from Wang and Stebbins.[79]]

The enhanced resolution of the MQMAS experiment allows the presence of three different bridging oxygen species to be resolved. In particular, there is a large contribution to the spectrum from Si–O–B species, implying that three-coordinate boron is incorporated into the silica network.

3.1.3 Phosphate Glasses

Kirkpatrick and Brow[80] reviewed in 1995 the use of NMR to characterize phosphate and phosphate-containing glasses. ^{31}P, as a spin $= \frac{1}{2}$ nucleus, gives MAS spectra in which the isotropic chemical shift is a sensitive function of the coordination environment and geometry. For example, the degree of polymerization of tetrahedral phosphate units is easily determined from the chemical shift, and the effects of other elements on the phosphate polymerization and the formation of P–O–X units has been widely studied. Losso and Sternberg[81] examined the ^{31}P chemical shift tensor components and the distribution of values existing in glasses using static and MAS measurements.

The connectivity between different phosphate units in sodium, calcium phosphate glasses has been investigated by 2-D exchange NMR.[82] The 1-D MAS spectra of these materials show two signals due to phosphorus connected through oxygen to one and two neighboring phosphorus, respectively (so-called Q_1 and Q_2). In the 2-D spectra, cross peaks are observed due to dipolar coupling between Q_1 and Q_2 sites, showing that these sites are connected.

Connectivity between phosphorus and aluminum or sodium in phosphorus-containing aluminosilicate glasses has been investigated by Schaller et al.[83] using the double resonance TRAPDOR technique. The ^{31}P MAS spectra of these samples show up to five different signals. The TRAPDOR measurements allow these to be assigned to phosphorus sites linked to different numbers of Na and/or Al cations.

3.1.4 Fluoride-containing Glasses

^{19}F as a highly abundant spin $= \frac{1}{2}$ nucleus can provide much useful information about fluoride-containing glasses. The chemical shift range for fluorine-containing compounds covers more than 200 ppm.[84] As an example, Figure 14 shows the ^{19}F MAS spectrum of a PZG ($PbF_2-ZnF_2-GaF_3$) glass reported by Bureau et al.[85] The spectrum is broad because of dipolar coupling not removed by MAS, and contains spinning side bands. Nevertheless, the spectrum could be fitted as the sum of the three components shown, at around 10, 65 and 156 ppm (relative to C_6F_6).

From comparisons with fluoride compounds of known structure, the authors assigned the three signals to free fluoride ions, fluoride ions in MF_6 octahedra

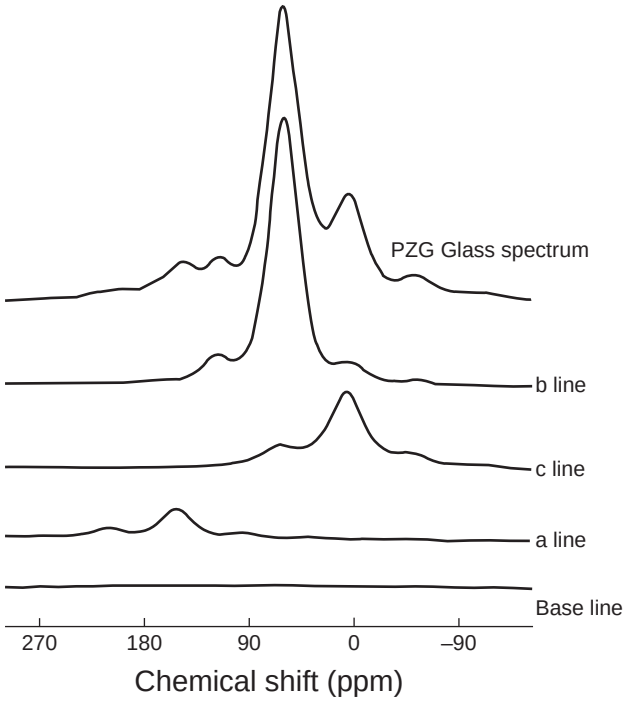

Figure 14 ^{19}F NMR spectrum of a PZG glass. [Reproduced with permission from Bureau et al.[85]]

and fluoride ions bridging between MF_6 octahedra, respectively, and they were able to follow the relative concentrations of these species as a function of glass composition.

3.1.5 Gels

The speciation of silicon in gels and gel precursors by ^{29}Si NMR follows very closely that in glasses: the chemical shift is a sensitive function of the silicon coordination and local structure. For example, Pauthe et al.[86] prepared silicon nanocrystallites embedded in a silica matrix by a sol–gel route using triethoxysilane as the precursor. ^{29}Si NMR spectra were measured after firing the gels at temperatures from 100 to 1000 °C, and showed the gradual loss of monomeric $HSiO_3$ units and the appearance of the Q_3 and Q_4 signals characteristic of bulk silica gel. At high temperatures, a signal assigned to crystalline silicon at −81 ppm appeared.

^{29}Si NMR has also been used in the characterization of hybrid gel materials prepared from polydimethylsiloxane precursors. In this case, retention of the linear dimethylsiloxane backbone was confirmed from the spectrum.[87]

^{27}Al NMR is likewise used to characterize precursors and intermediate phases in the formation of alumina or composite alumina-containing gels. Figure 15 shows, for example, ^{27}Al MAS NMR spectra of aluminosilicate gels

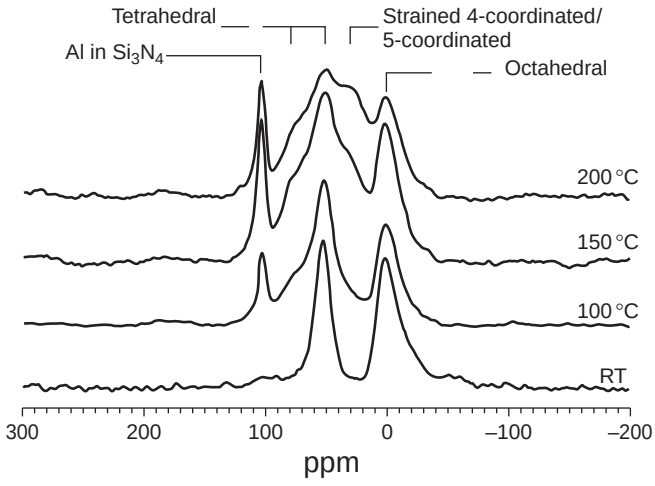

Figure 15 ^{27}Al MAS spectra of aluminosilicate gels after heating for 2 h at each of the temperatures indicated. [Reproduced with permission from Peeters and Kentgens.[88]]

as a function of firing temperature, reported by Peeters and Kentgens.[88]

The initially formed gel shows two well resolved peaks associated with tetrahedral and octahedral aluminum, respectively, but after firing the spectrum broadens in a complex manner. In order to enhance the resolution, Peeters and Kentgens performed 3QMAS measurements on the same samples. Figure 16 shows the 3QMAS spectrum of the gel heated to 200 °C, in which the signals of the three different aluminum species can be clearly resolved and quantified.

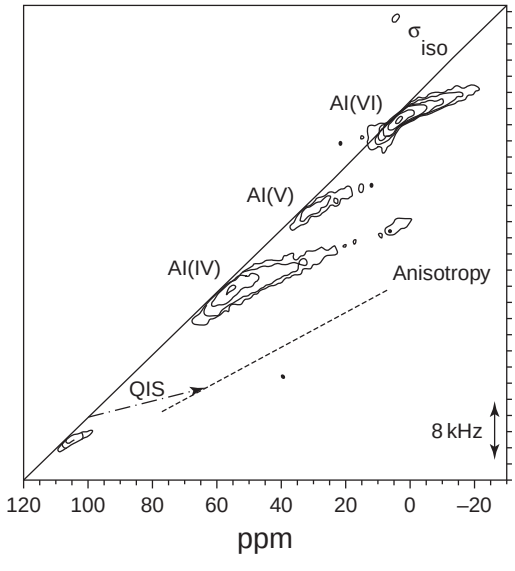

Figure 16 ^{27}Al 3QMAS spectrum of aluminosilicate gel fired to 200 °C. The deviations from the diagonal axis reflect a distribution of quadrupole parameters. [Reproduced with permission from Peeters and Kentgens.[88]]

Table 11 Sol–gel precursors and reactions

Nucleus	Sample	NMR techniques	Analytical information	Ref.	Other related studies
^{29}Si	Silicate solutions	PGSE[a] self-diffusion measurements	Self-diffusion and chemical-exchange studies	307	308
^{23}Na, ^{27}Al, ^{29}Si	Albite ($NaAlSi_3O_8$)	^{29}Si, ^{27}Al and ^{23}Na NMR, ^{29}Si {^{1}H} CPMAS	Gel synthesis and heat treatment of an albite glass	309	
^{207}Pb	$PbTiO_3$ films and powders	^{27}Pb NMR	Sol–gel processing of $PbTiO_3$	305	
^{1}H	Water-glass silica solutions	^{1}H NMR and T_2	Gel transformations in silicas	94	
^{17}O	Sol–gel-derived La_2O_3	^{17}O MAS; satellite transition NMR	Study of La_2O_3 formation	106	
^{17}O, ^{29}Si	Hybrid zirconia–silica glasses	^{17}O, ^{29}Si solution-state NMR of reaction intermediates	Optimization of synthesis	310	311, 90
^{17}O, ^{29}Si, ^{13}C	Hybrid polymer–silica materials	^{17}O, ^{29}Si solution-state and MAS NMR; ^{13}C MAS	Synthesis, structure and dynamics of hybrid materials	244	312, 159, 313
^{17}O	TEOS, TMOS reaction mixtures	^{17}O solution-state NMR	The fate of water in TEOS and TMOS hydrolysis	97	
^{13}C, ^{17}O, ^{29}Si, ^{27}Al	TEOS and aluminum *sec*-butoxide reactions	^{13}C, ^{29}Si, ^{17}O and ^{27}Al solution and solid-state NMR	Aluminosilicate sol–gel synthesis	240	
^{17}O, ^{29}Si	Alkoxycarbide gels and precursors	Solution-state NMR; ^{29}Si–^{1}H HETCOR MAS NMR	Reactions of alkoxysilanes containing Si–H bonds	314	
^{1}H, ^{13}C, ^{29}Si, ^{31}P	TMOS–phosphorus alkoxide sol–gels	^{1}H, (CSi)–^{13}C–^{29}Si and ^{31}P liquid- and solid-state NMR	Sol–gel conversion	254	
^{29}Si	TEOS reaction mixtures	Solution-state ^{29}Si NMR	Kinetics of sol–gel formation	315	316
^{29}Si	MTES reaction products	Solution-state ^{29}Si NMR (DEFT)	Polymerization reactions in MTES	96	95, 146

[a] PGSE, pulsed gradient spin echo; MTES, methyltriethoxysilane.

The quadrupole coupling parameters and distribution were estimated from comparisons of the observed spectra with calculated spectra, and the values obtained were corroborated by off-resonance nutation measurements.

A similar approach was taken by Soraru et al.,[89] who used ^{11}B 3QMAS NMR to resolve different boron sites in organically modified borosilicate gels. In particular, a high concentration of B–O–Si bonds was detected by this technique.

^{17}O NMR is a powerful technique for characterizing both gels and glasses. For example, ^{17}O MAS NMR was used by Dirken et al.[90] to distinguish between Ti–O–Si and Ti–O–Ti bonds in titanosilicate gels.

3.1.6 Gel Precursors

Oxide glasses and hybrid organic–inorganic glassy materials can be synthesized by low-temperature hydrolysis and condensation of soluble precursors, usually alkoxides. Thus silicates are produced from alkoxysilanes, principally TEOS and tetramethoxysilane (TMOS). The initially formed hydrated gels can be further processed by dehydration and thermal treatment to form dense glasses. Not surprisingly, NMR spectroscopy finds an important role in the characterization of the resulting materials as discussed in section 3.1.5. There is a vast literature covering this area, a detailed review of which is not attempted here.[91] Therefore, this section will cover the reactions leading to the formation of gels and the applications of NMR in characterizing the intermediates (see Table 11).

Multinuclear magnetic resonance spectroscopy has become an important technique in sol–gel chemistry where the hydrolysis and condensation reactions leading to gel formation can be monitored. High-resolution solution-state spectroscopic techniques are applicable prior to gel formation and provide structural, kinetic and mechanistic data. ^{29}Si NMR is used to monitor the Si species formed during reactions of alkoxysilanes. Figure 17 shows the ^{29}Si chemical shift assignments for silicon sites in TEOS sol–gel systems.[92]

The ^{29}Si NMR spectrum of sol–gel systems can be observed past the gelation point, under solution-state conditions, indicating that the mobility of the oligomers remains.[93] The transition from sol to gel has been monitored in situ by ^{1}H spin–spin relaxation measurements.[94] Gelation of water-glass silicate solutions was accompanied by a decrease in T_2 of the water protons to a minimum value. Subsequent aging of the gels was associated with a

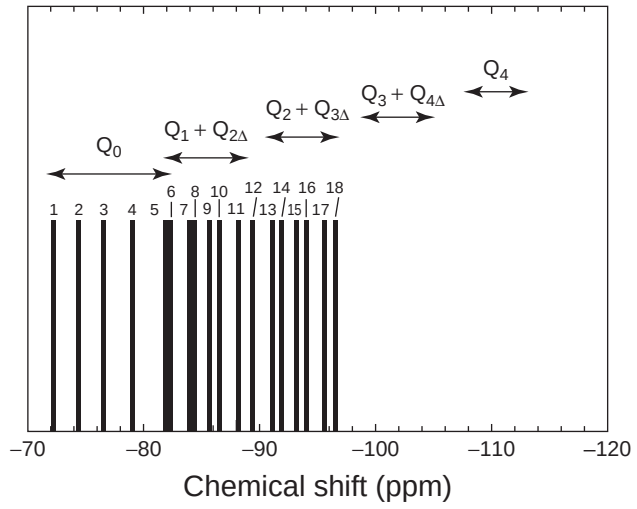

Figure 17 Chemical shift assignments for ^{29}Si species in TEOS hydrolysis and condensation. Peaks are marked by numbers corresponding to the following Si sites: 1, Q_0^4; 2, Q_0^3; 3, Q_0^2; 4, Q_0^1; 5, Q_0^0; 6, Q_1^3; 7, $Q_{2\Delta}^2$; 8, Q_1^2; 9, $Q_{2\Delta}^1$; 10, Q_1^1; 11, $Q_{2\Delta}^0$; 12, Q_0^1; 13, Q_{2x}^2; 14, Q_2^2; 15, Q_{2x}^1; 16, Q_2^1; 17, Q_{2x}^1; 18, Q_2^0. Δ and x subscripts designate sites in three- and four-membered rings, respectively. [Reproduced with permission Sefcik and McCormick.[92]]

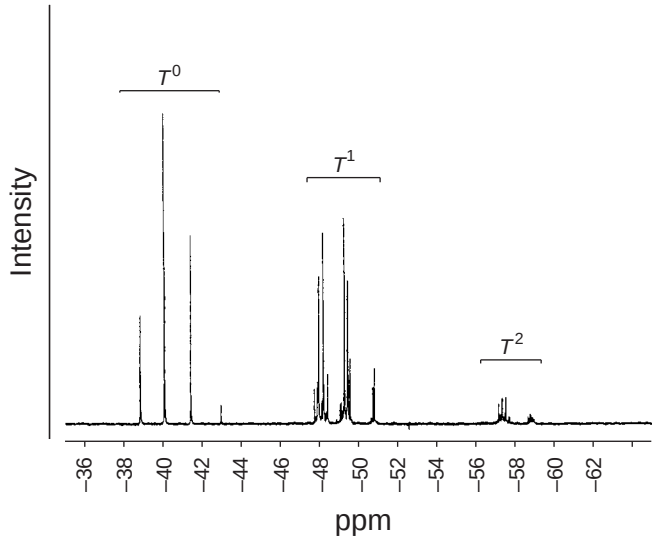

Figure 18 ^{29}Si NMR {DEPT} of an MTES reaction mixture at 296 K and 59.59 MHz, with ^{1}H broadband decoupling (MTES–C_2H_5OH–H_2O with H_2O : Si ratio $w = 2$; pH = 2.4). [Reproduced with permission from Brunet.[96]]

gradual increase in T_2 again, as the pore structure developed. Relaxation changes thus provide a sensitive probe of the gel transition, allowing in situ studies which are difficult by other techniques. Polarization transfer methods [insensitive nuclei enhanced by polarization (INEPT) and DEPT] have been employed to overcome two of the difficulties with ^{29}Si NMR, the relative insensitivity of the ^{29}Si nucleus and its slow relaxation.[95,96] DEPT experiments also give improved quantitation and allow the assignment of Si sites according to the number of attached hydroxy groups. Figure 18 shows the ^{29}Si DEPT spectrum of a MTES reaction mixture at pH 2.4, in which the various silicon monomers and oligomers are resolved. The T^n notation refers to the number of siloxane bonds so that T^0 are monomers, T^1 end groups of chains and T^2 central sites in chains or rings.

^{17}O enrichment is easily achieved in sol–gel systems and ^{17}O NMR has found increasing application in this field both to follow the fate of water during the early stages of reaction[97] and using MAS spectroscopy to detect different oxygen coordination modes in gels and oxide products.

3.1.7 Minerals

The speciation of silicon in silicate minerals by ^{29}Si NMR follows that in other materials; the higher resolution achieved in crystalline mineral phases means that Q_4 and Q_3 species with different next-nearest neighbors can be more reliably identified and quantified than in silicate glasses, for example.[98] Kohn et al. have identified up to 10 distinct Si sites in various synthetic leucite analogs.[99] The use of 2-D chemical shift correlation spectroscopy to establish connectivity between different silicon sites in crystalline synthetic zeolites by Fyfe et al.[15] was referred to in section 2.6. ^{29}Si MAS and CPMAS have been used to characterize the phases present in microcrystalline opals.[100] ^{1}H CP was used to show the presence of large concentrations of internal SiOH groups in these materials. ^{29}Si NMR has been used also to follow the process of dealumination of kaolinite clays with hydrochloric acid treatment.[101] In this case, the conversion of Si–O–Al species to Si–O–Si species with reaction time could be quantitatively monitored and a mechanistic model proposed.

The problem of quantifying and speciating aluminum in crystalline minerals, particularly in terms of eliminating quadrupolar broadening, has been addressed by several groups recently. Alemany et al. have undertaken very fast ^{27}Al MAS measurements on kyanite and andalusite at three different magnetic field strengths (9.4, 11.65 and 18.8 T), and compared the spectra obtained with the corresponding 3QMAS spectra.[21] Figure 19 shows 1-D MAS and 2-D 3QMAS spectra of kyanite, measured at 9.4 T and a MAS rate of 22 kHz.

Kyanite contains four different octahedral Al sites, and in the 1-D MAS spectrum these cannot be separated, regardless of the magnetic field strength used. The three signals clearly resolved in the 3QMAS spectrum have an intensity ratio of 2 : 1 : 1, indicating that two of the Al sites have the same apparent isotropic chemical

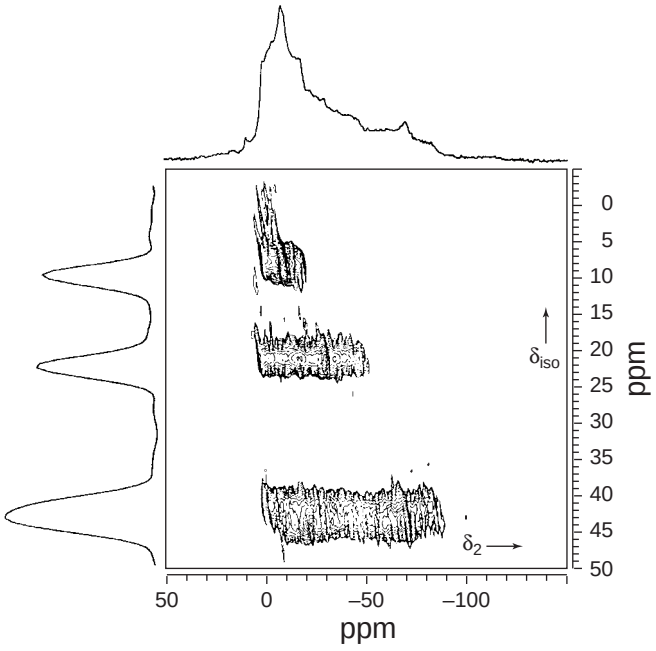

Figure 19 ^{27}Al MAS and 3QMAS spectra of kyanite. The MAS spectrum is shown at the top and the summed 3Q signals on the left. [Reproduced with permission from Alemany et al.[21]]

shift at this field. From measurements of the MAS and 3QMAS spectra at three different magnetic fields, however, Alemany et al. were able to calculate accurately the quadrupole coupling constants and true isotropic chemical shifts for the four sites.

A useful comparison of the different techniques for obtaining ^{17}O NMR spectra [static, MAS, variable-angle spinning (VAS), DAS and MQMAS] of the zeolite mineral stilbite has been reported by Xu and Stebbins.[102] Figure 20(a–d) summarizes the spectra obtained.

The static spectrum clearly shows two overlapping quadrupolar powder patterns, with a total width of ca. 400 ppm at 9.4 T. This is narrowed to ca. 80 ppm in the MAS spectrum, but there is insufficient resolution to determine unambiguously the parameters of the two overlapping signals. Partial resolution is achieved in the DAS 2-D spectrum; the two oxygen sites Si–O–Si and Si–O–Al are separated by ca. 20 ppm along the isotropic dimension. The improved resolution is due to the removal of anisotropic second-order quadrupolar broadening by the DAS experiment, and the remaining line widths of ca. 20 ppm are due at least in part to a distribution of isotropic chemical shifts in the sample. The distribution of quadrupole parameters for the two sites was investigated by simulating VAS slices across the 2-D DAS spectrum. In the 3QMAS spectrum, chemical shift anisotropy and dipolar interactions are eliminated to first order, and the two oxygen sites are now clearly resolved with line widths of $<$10 ppm in the isotropic dimension. The quadrupole parameters obtained by simulating the different spectra (static, DAS and 3QMAS) agree moderately well; Xu and Stebbins argued that the 3QMAS-derived quadrupole data are the most reliable.

The 3QMAS technique has also been applied to ^{23}Na NMR of sodium cations in various zeolites by Hunger et al.[103] In these systems, the presence of Na^+ in one or more cation-exchange sites is evident from the occurrence of partially resolved shoulders on the 1-D MAS spectra, but the distinction between a second-order quadrupolar powder pattern and multiple overlapping isotropic chemical shifts cannot easily be made. The 3QMAS experiment allows this separation to be done; Hunger et al. reported isotropic chemical shifts and quadrupole coupling constants for Na^+ in the different exchange sites identified.

^{113}Cd NMR has been used by Sullivan et al.[104] to investigate cation binding in montmorillonite clays. In particular, they undertook double resonance SEDOR experiments to investigate the coupling between ^{113}Cd and ^{27}Al in order to study cation sites in the clay. In the case of hydrated clays, no coupling was observed, owing to rapid motion of hydrated cations within the clay layers. For freeze-dried samples, however, significant SEDOR intensity reduction was observed, indicating strong dipolar coupling between ^{113}Cd and ^{27}Al cations in the octahedral sheets of the montmorillonite structure. Since the distance between such aluminum ions and the interlayer space of the clay is $>$ 5 Å (the largest distance over which dipolar coupling is operative), it was concluded that in the dried clay the Cd cations move to pseudo-hexagonal holes in the tetrahedral sheets.

3.1.8 Ceramics

NMR has found many applications in studying the formation and crystallization of ceramic phases prepared from sol–gel and glass precursors. For example, ^{23}Na, ^{27}Al and ^{31}P NMR have been used by Abrahams et al. to characterize bioactive ceramics prepared from $CaO–Na_2O–P_2O_5–Al_2O_3$ glasses.[105] ^{23}Na MAS spectra from such materials are broad and poorly resolved, but suggest the possibility of two different Na^+ sites. By applying 3QMAS NMR, these two species could be clearly separated, and their quadrupole coupling parameters determined. The reliability of the analysis was checked by successfully simulating the single-quantum MAS spectra with the parameters obtained. The corresponding ^{31}P MAS spectra showed two principal bands assigned to Q_1 and Q_2 phosphate units. From a Herzfeld–Berger analysis[10] of the spinning side band intensities in these spectra, the chemical shift tensor components of these two species were obtained, confirming the assignment.

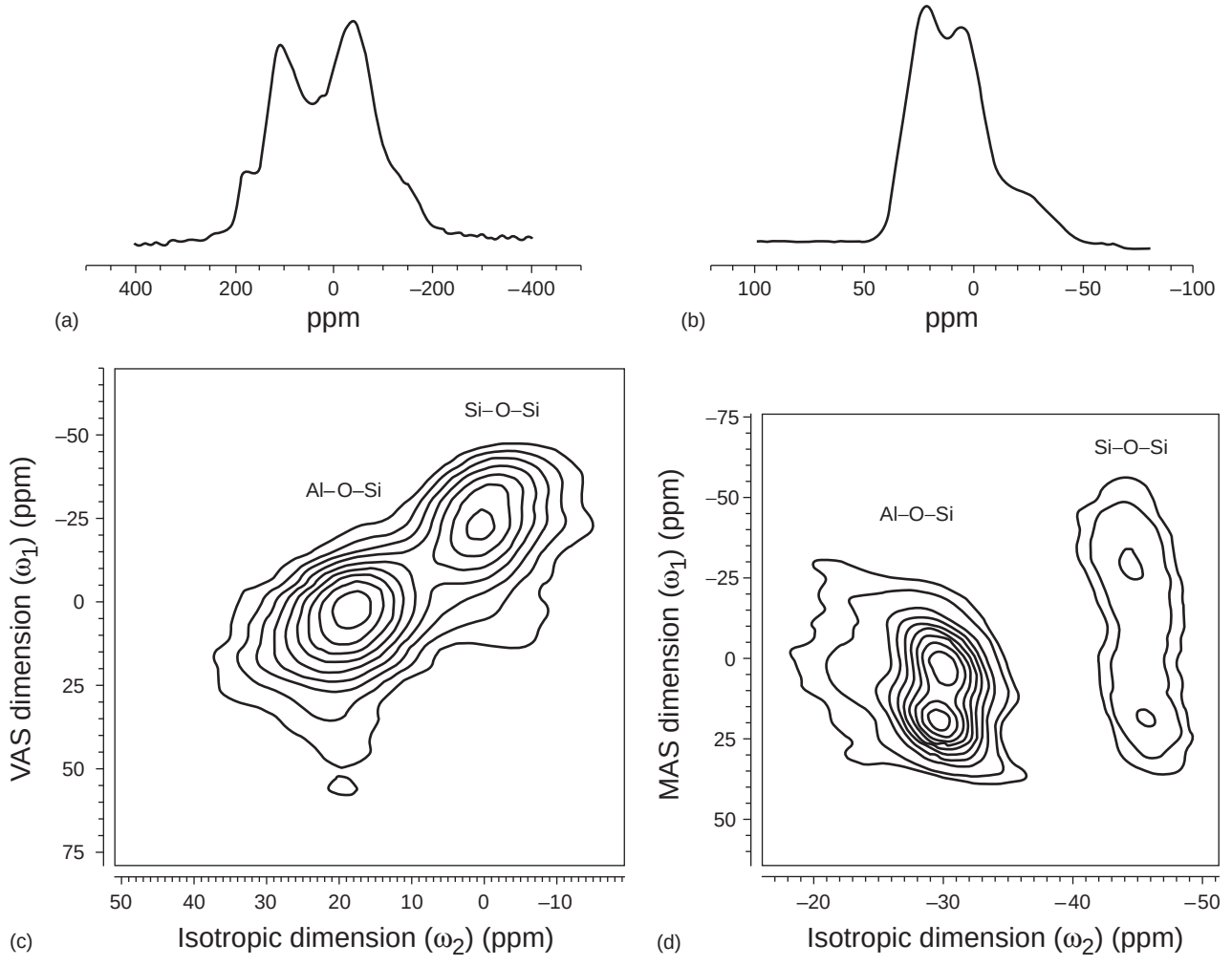

Figure 20 Comparison of (a) static, (b) MAS, (c) DAS and (d) 3QMAS ^{17}O spectra of ^{17}O-enriched stilbite. [Reproduced with permission from Xu and Stebbins.[102]]

Changes in the speciation of oxygen during the formation of La_2O_3 from various precursors was followed by Ali et al.[106] using ^{17}O MAS NMR on enriched materials. The formation of the oxide was shown to occur via an intermediate LaO(OH) species not previously observed in other oxide gels.

Brieger et al.[107] used ^{207}Pb NMR to follow the pyrolysis and crystallization of lead zirconate titanate ceramics prepared by sol–gel methods. The parent gels show broad, unresolved ^{207}Pb MAS spectra; the absence of resolved spinning side bands was attributed to the superposition of signals from lead species in a distribution of different environments. Upon crystallization, the signals narrow and spinning side bands are resolved. The line widths of the resulting signals were strongly influenced, however, by the particular preparation method used, indicating that there can be wide variations in local order within ceramics showing identical crystallinity by X-ray diffraction.

The use of ^{6}Li and ^{7}Li MAS NMR to aid in the nondestructive analysis of lithium aluminate ceramic phases has been reported by Stewart et al.[108] ^{6}Li was found to give a single, narrow resonance with no spinning side bands, suggesting that a single lithium site is occupied, and the intensity of the signal was found to be directly proportional to the concentration of lithium in the sample, provided that enough time was allowed for spin–lattice relaxation. ^{7}Li gave a much more intense signal than ^{6}Li, with spinning side bands resulting from the larger quadrupolar interactions expected for this nucleus. The static spectrum was also measured to obtain the quadrupole coupling constant.

3.1.9 High-temperature Glasses, Liquids and Melts

The use of NMR to investigate the structure and dynamics of oxide melts is aimed at understanding both geological and synthetic glassy materials. Magmas and

lava cannot be studied directly by NMR but laboratory experiments can provide insights into the properties of these molten systems through model studies of oxide melts and supercooled liquids over an increasingly wide range of temperatures and pressures. Spectroscopic analysis of glasses formed by quenching of melts at different rates and of the melts themselves provides fundamental information on the structure of the liquid phase. Investigation of glass structure and dynamics at the glass transition temperature (T_g) and above provides information on slower exchange processes which are completely averaged out in the melt. This area has been reviewed in detail in 1991 and 1995[3,33] and NMR spectroscopy has also been evaluated in comparison with vibrational spectroscopy for the analysis and characterization of melts.[109]

The onset of molecular motion above T_g results in exchange of Si between coordination sites with complete averaging and narrow line widths in the melt. Figure 21(a–f) shows the ^{29}Si NMR spectrum of $K_2Si_4O_9$ as a function of temperature from 697 to 997 °C. The increasing rate of chemical exchange between Q^3 and Q^4 sites was determined by simulation of the peak shapes.

Similar line narrowing is observed for nonframework nuclei such as Li, Na and Mg. Static ^{25}Mg spectra for a magnesium aluminosilicate were only observed above 1000 °C and showed significant line narrowing on further heating from 1200 to 1440 °C.[110] High-temperature 2-D exchange spectra are sensitive to processes on a timescale of seconds. Such experiments have been carried out on liquids just above T_g, enabling exchange to be detected at frequencies less than the peak width. This is illustrated by the spectra for $K_2Si_4O_9$ above T_g at 555 °C (Figure 22a and b), which show off-diagonal peaks at longer mixing times, indicative of substantial exchange between Q^3 and Q^4 silicon sites.

Fast exchange in high-temperature melts means that structural details of Si coordination in the liquid are averaged out, as seen above. There are two strategies to recover this information. One is to quench the melt at different rates, thus freezing in changes in liquid structure at a fictive temperature T_f. This strategy has been used to identify Si(V) and Si(VI) coordination sites which are found to increase in importance at higher temperatures in the melt. Similarly, in aluminosilicates, Al(V) and Al(VI) coordination has been observed in glasses quenched at high T_f and high pressure.[111] Oxygen site exchange accompanies averaging of Si, Al or B sites in oxide melts and hence ^{17}O NMR has also been used to identify changes in bridging and nonbridging oxygen sites in quenched glasses.[112] In cases where the liquid can be supercooled, exchange is slowed and coordination information can be obtained in the liquid state.[113]

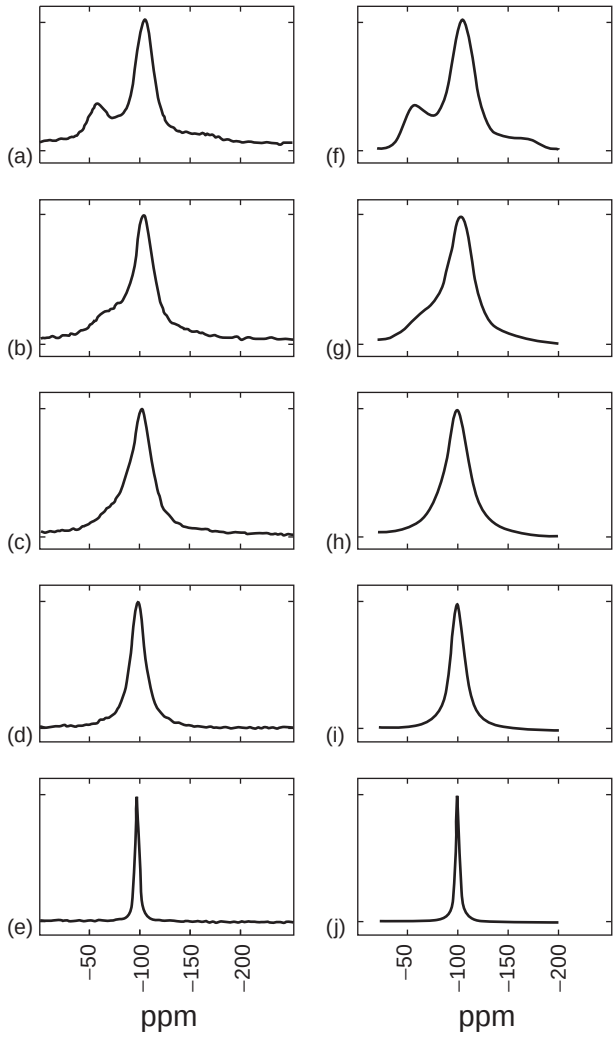

Figure 21 ^{29}Si NMR spectra (a–e) of $K_2Si_4O_9$ as a function of temperature from 697 to 997 °C and simulations (f–j). Temperatures were (a) 697, (b) 774, (c) 800, (d) 847 and (e) 997 °C. Exchange rates for the simulations were (f) 2000, (g) 10 000, (h) 25 000, (i) 50 000 and (j) 500 000 Hz. [Reproduced with permission from Stebbins et al.[3]]

Relaxation measurements are sensitive to motion on a timescale of nanoseconds. On heating a silicate glass above T_g, the temperature dependence of the spin–lattice relaxation time, T_1, shows a discontinuity associated with the onset of long-range motion in the glass.[114] Non-framework nuclei do not show the same effect. Apparent activation energies can be extracted from Arrhenius plots of relaxation data both above and below the T_1 minimum. However, in both glasses and melts, these values are generally not the same, with a higher E_a found on the high-temperature side of the minimum. This has been attributed to the dominance of high-energy diffusive motions in the relaxation mechanism at high temperature.[3]

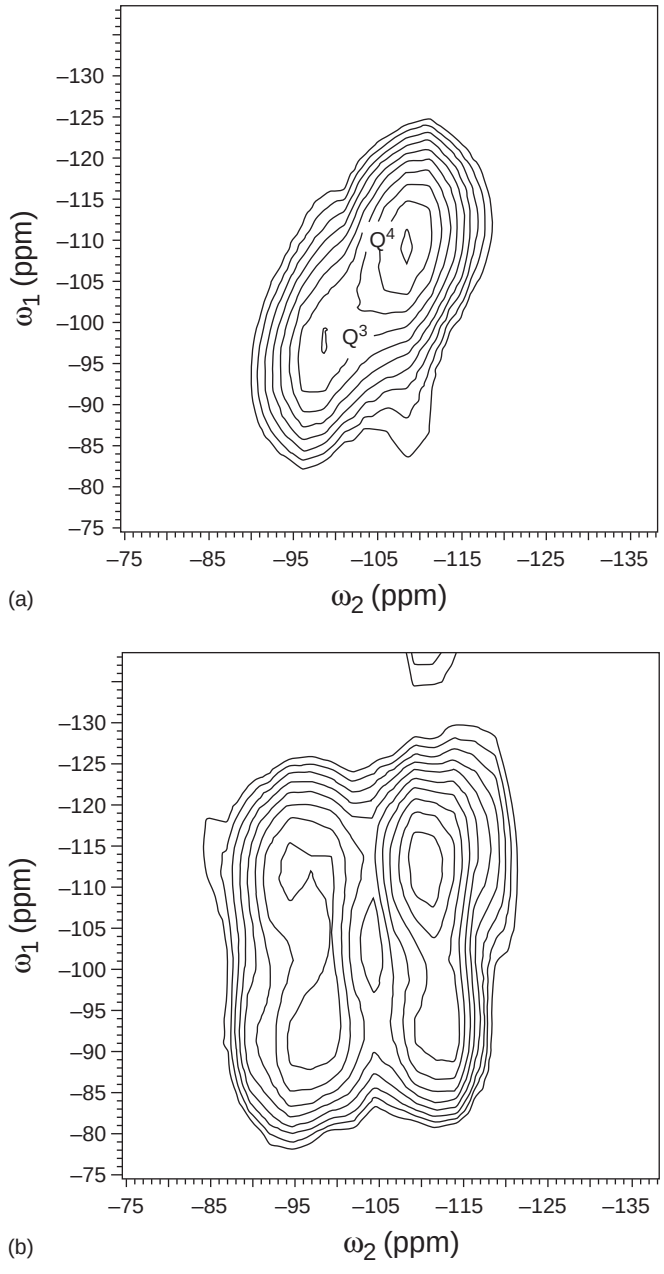

Figure 22 2-D MAS exchange spectra for $K_2Si_4O_9$ at 555 °C. Mixing time: (a) 0.5 s; (b) 4.0 s. [Reproduced with permission from Stebbins et al.[3]]

^{11}B NMR, along with ^{29}Si and ^{27}Al NMR, has been used to monitor structural changes in borate, borosilicate and boroaluminosilicate liquids. High-resolution experiments using 3QMAS spectroscopy have recently been carried out. This technique allowed BO_3 ring sites to be distinguished from nonring sites in a fast-cooled $(Na_2O)_{0.5}(B_2O_3)_{0.95}$ glass sample.[115]

Most NMR studies of melts concern oxides, but some nonoxide glasses have also been analyzed. ^{31}P and ^{77}Se NMR have been used to measure multiple site exchange in glass-forming melts of phosphorus and selenium.[116]

Table 12 provides a summary of the recent literature on oxide melts and high-temperature glasses.

3.1.10 Hydrated Minerals and Glasses

In this section, the application of ^1H and ^2H NMR spectroscopy to the analysis of glasses and minerals is considered. The NMR characterization of water and other fluids trapped in porous materials is a significant area of application which is not covered here. Imaging is considered separately in section 3.3. This section covers NMR analysis of water and hydroxy groups in hydrated minerals and glasses from a structural point of view, as summarized in Table 13.

^1H wideline NMR spectra of hydrous silicates can show Pake patterns typical of immobilized protons. Mobile water and the Si–OH protons with which it is in fast exchange show narrow, motionally averaged line shapes. Superposition of both types of line shape is common in heterogeneous materials. Static ^1H line shape studies have proved useful in determining the orientation of water in hydrated layers in sheet silicates and to investigate mobility of water in hydrated silicates more generally.[117]

In ^1H MAS spectra, spinning speeds high enough to remove the homonuclear dipole–dipole couplings would need to be of the order of 30–40 kHz. MAS spectra at lower spinning speeds therefore contain a series of spinning sidebands. They nevertheless have sufficient resolution to distinguish water and silanol OH groups in hydrated silicates. The ^1H MAS spectrum of a hyalite mineral consisted of three partly overlapping signals at 7.1, 5.6 and 3.9 ppm, which were assigned to isolated water molecules, water clusters and two types of silanol groups.[118] Multiple pulse techniques such as CRAMPS enable sidebands to be suppressed and provide higher resolution, but in amorphous glasses, chemical shift dispersion may still make it difficult to resolve different OH species. 2-D CRAMPS MAS correlation spectroscopy enables the individual MAS spectra of the OH sites to be separated, permitting analysis of the spinning sideband patterns. Figure 23(a) and (b) shows MAS slices from a CRAMPS MAS correlation spectrum of an $Na_2O \cdot 4SiO_2 \cdot 0.7H_2O$ glass, showing the separated spectra for H_2O at 4.0 ppm and OH at 14.0 ppm.[119]

^1H relaxation measurements (T_1 and T_2) are used to investigate water mobility in hydrated glasses, minerals and gels. In heterogeneous systems, where water in different phases is not in fast exchange, multiple relaxation times can be resolved. This has been used to good effect in studies of the curing of cement, where relaxation times were used to distinguish three separate phases.[120]

Table 12 High-temperature glasses, liquids and melts

Nucleus	Sample	NMR techniques	Analytical information	Ref.	Other related studies
^{6}Li, ^{7}Li, ^{29}Si	Lithium disilicate glasses quenched at high temp. and pressure	MAS and nutation NMR	Nucleation and growth in lithium disilicate	317	
^{29}Si	$K_2Si_4O_9$	2-D MAS spectra, to 583 °C	Si exchange and formation of Si(V) species	318	
^{27}Al, ^{23}Na, ^{29}Si, ^{133}Cs	Analcime	Room temp. to 550 °C and ^{27}Al, T_1	Dehydration of Analcimes at high temp.	289	
^{23}Na	$Na_2Si_3O_7$, $Na_3AlSi_6O_{15}$, $NaAlSi_3O_8$	Chemical shifts and T_1 data, to 1200 °C	Neighboring sites and disorder around Na	319	320
^{23}Na	Germanates and borates	Chemical shifts to 1200 °C	Correlation of Na–O bond lengths and δ_{iso}(^{23}Na)	321	
^{25}Mg	Mg silicate liquids	Chemical shifts and T_1, 1000–1470 °C	Mg coordination and motion in the melt	110	264, 263
^{7}Li, ^{23}Na, ^{29}Si	Li_2SiO_3, Na_2SiO_3	MAS and relaxation data to 20 °C below melting	Cation dynamics and premelting	322	
^{23}Na	$NaAlO_2$–SiO_2 melts	^{23}Na spectra to 1600 °C	Na coordination	323	324
^{27}Al	CaO–Al_2O_3 liquids	High-resolution NMR at 2500 K	Structure and dynamics	325	
^{11}B, ^{23}Na	Borate and silicate liquids	Chemical shifts and T_1 data, to 1200 °C	Dynamics of cation exchange	326	
^{29}Si	Sodium silicate glass and liquid (60 mol% SiO_2)	MAS NMR to 600 °C	Silicate species exchange	327	
^{25}Mg	Forsterite, Mg_2SiO_4	NMR to 1400 °C	Magnesium site exchange	328	
^{29}Si, ^{11}B	Alkali metal borosilicate liquid	High-temp. MAS NMR	Microscopic dynamics and viscous flow	329	330, 331, 115
^{17}O	$K_2Si_4O_9$ and $Na_2Si_4O_9$ quenched glasses	^{17}O static and MAS spectra	Nonbridging and bridging oxygens at Si(V) and Si(VI) sites	112	332

^{2}H has spin $I = 1$ and has a much lower gyromagnetic ratio than ^{1}H. Its spectra and relaxation properties are therefore dominated by quadrupolar rather than dipole–dipole interactions. ^{2}H NMR of isotope-enriched samples therefore finds some useful applications in investigating hydration sites in glasses and minerals. ^{2}H static NMR line shapes give typical powder patterns for immobilized molecules and partially averaged quadrupolar splittings in the presence of molecular motion. D_2O in clay minerals shows residual quadrupolar splittings of between 100 Hz and 12 kHz depending on the water content.[121] ^{2}H relaxation is dominated by quadrupolar interactions and relaxation studies provide a useful probe of water mobility in deuterated materials.[121,122]

3.2 Nuclear Magnetic Resonance Relaxation Applications for Mobile Ions in Glasses

The mobility of nonframework ions in silicates and more generally of both cations and anions ions in ion-conducting glasses has predominantly been investigated using NMR relaxation measurements. Li^+ is the conducting species in a large range of ion-conducting glasses and hence the NMR relaxation properties of its two NMR-active nuclei (^{6}Li and ^{7}Li) are extensively used to characterize the structure and dynamics of these materials.

^{7}Li is the more abundant isotope and has spin $I = \frac{3}{2}$. Peak broadening due to quadrupolar interactions and disorder in glasses, along with a narrow chemical shift dispersion, means that structural information is not easy to obtain.[123] On the other hand, ^{7}Li relaxation measurements are extremely useful in studying Li^+ dynamics. ^{7}Li quadrupolar interactions and chemical shift anisotropy are both relatively small in solids, so that relaxation is dominated by homonuclear dipole–dipole interactions. Coupling to paramagnetic centers is important in naturally occurring minerals.[70] ^{6}Li and ^{7}Li have complementary NMR properties so that a combination of ^{6}Li chemical shift and ^{7}Li relaxation measurements can provide a more detailed insight into structure and dynamics in Li^+ conducting glasses. Figure 24 shows the resolution of Li coordination sites in the MAS spectrum of ^{6}Li-enriched lithium orthosilicate.

^{6}Li 2-D exchange spectroscopy in lithium orthosilicate enabled exchange rates and activation energies for exchange between different Li sites to be determined.[42] Close agreement was found between the activation energy for LiO_3 to LiO_4 exchange with that for conductivity.

Table 13 ^{1}H and ^{2}H NMR applications in hydrated minerals and glasses

NMR nucleus	Sample	NMR techniques	Analytical information	Ref.	Other related studies
^{1}H, ^{17}O, ^{23}Na, ^{27}Al, ^{29}Si	Silica, albite and $Na_2Si_4O_9$ glasses	Multinuclear NMR; quantitative ^{1}H NMR	Water in hydrous glasses	117	333, 334
^{1}H, ^{23}Na, ^{29}Si	Silicate electrode glasses; kaolinite	$^{23}Na-^{1}H$ and $^{27}Al-^{1}H$ REDOR NMR	Hydrated layers in electrode glasses	335	336, 337, 338
^{1}H	Aluminate minerals	^{1}H T_1	Hydration of aluminate minerals	339	
^{1}H, ^{29}Si	Hyalite	^{1}H and ^{29}Si MAS $^{29}Si\{^{1}H\}$ CPMAS	Si coordination; water and silanol sites	118	
^{1}H	Clinopyroxene, enstatite, forsterite	^{1}H MAS	H_2O in minerals	340	341, 342
^{2}H, ^{7}Li	Layer silicates; Li-saponite; montmorillonite	^{2}H and ^{7}Li NMR, ^{2}H T_1	Dynamics of intercalated water	343	
^{1}H, ^{31}P	Zinc ultraphosphate glasses	$^{31}P\{^{1}H\}$ CPMAS; ^{1}H MAS; ^{1}H T_1 and T_2	Local structure, water and hydroxy groups	344	345
^{1}H, ^{31}P	Phosphoellenbergerite and holtedahlite	H and ^{31}P MAS, $^{31}P\{^{1}H\}$ CPMAS	OH sites in magnesium phosphates	346	
^{1}H, ^{29}Si	Kaolinite/HCl(aq.) reactions	^{1}H CRAMPS and MAS, ^{29}Si NMR	Si and OH sites in the dealumination of Kaolinite	101	
^{1}H, ^{29}Si, ^{43}Ca	Calcium silicate hydrates	^{1}H CRAMPS, ^{29}Si, ^{43}Ca	Calcium silicate hydrates in cement and concrete	347	348
^{1}H	$Na_2O \cdot 4SiO_2 \cdot 0.7H_2O$	MAS, CRAMPS and 2-D CRAMPS/MAS	Structure of hydrous silicates	119	
^{2}H	Clay minerals	^{2}H variable-temp. T_1	Interaction of water with clay minerals	121	
^{2}H	Epoxy–glass composites	^{2}H NMR	Water mobility in epoxy–glass composites	122	

The relaxation behavior of framework nuclei, particularly ^{31}P in phosphate glasses, has also been shown to give useful insights into ion conduction mechanisms. Spin–lattice relaxation of ^{31}P in the rotating frame ($T_{1\rho}$) proved the most suitable method of investigating the effect of Li^+ dynamics on the glassy framework in $LiPO_3$.[124] Since rotating frame relaxation is sensitive to motions in the kilohertz range, this enabled a full temperature dependence study of the relaxation to be carried out, whilst maintaining the temperature below T_g. The derived activation energy was in good agreement with that obtained for lithium ion diffusion from conductivity measurements.

Mixed-alkali glasses show a nonlinear dependence of thermal and electrical conductivity properties on cation composition. Cation mobility is also reduced compared with single-ion glasses. Numerous NMR studies have sought to understand the origins of these effects in terms of cation site distributions or the presence of distinct cation environments. It has proved difficult to directly observe distinct sites in NMR spectra of these mixed-alkali systems due to poor chemical shift resolution in 1-D spectra of ions such as $^{7}Li^+$ and $^{23}Na^+$. Recently, a detailed comparison was made between the speciation information obtained from 1-D ^{23}Na and ^{31}P MAS spectra.[125] Inclusion of DAS enhanced the spectral resolution but not sufficiently to allow the detection of inequivalent Na sites. However, the 2-D technique of $^{23}Na-^{31}P$ DAS/CPMAS HETCOR enabled ^{23}Na nuclei associated with Q_1 and Q_2 phosphate sites to be separated and the spectra of these fractions showed splittings attributed to distinct Na environments.

The question of cation clustering in mixed alkali metal silicate and chalcogenide glasses has been addressed by $^{23}Na^{6}Li$ and $^{23}Na^{7}Li$ SEDOR spectroscopy.[72,74,126] Generally, there is no evidence for cation clustering in these systems, although if clustering occurs, the NMR experiments are consistent with preferential interactions of like cations.

Lithium-doped chalcogenide glasses are generally more conducting than their oxide counterparts and there is a considerable literature on lithium ion mobility in thioborates, thiogermanates and thiosilicates, as well as oxides and silicates, as summarized in Table 14. Other important nuclei for investigation of cation-conducting glasses are ^{23}Na and ^{109}Ag, and ^{85}Rb, ^{133}Cs and ^{205}Tl have also found a few applications.

Some ^{109}Ag NMR studies have been carried out on silver borate, phosphate and thiogermanate glasses. Ag has a low gyromagnetic ratio and slow relaxation times but motional line narrowing in Ag^+ conducting materials

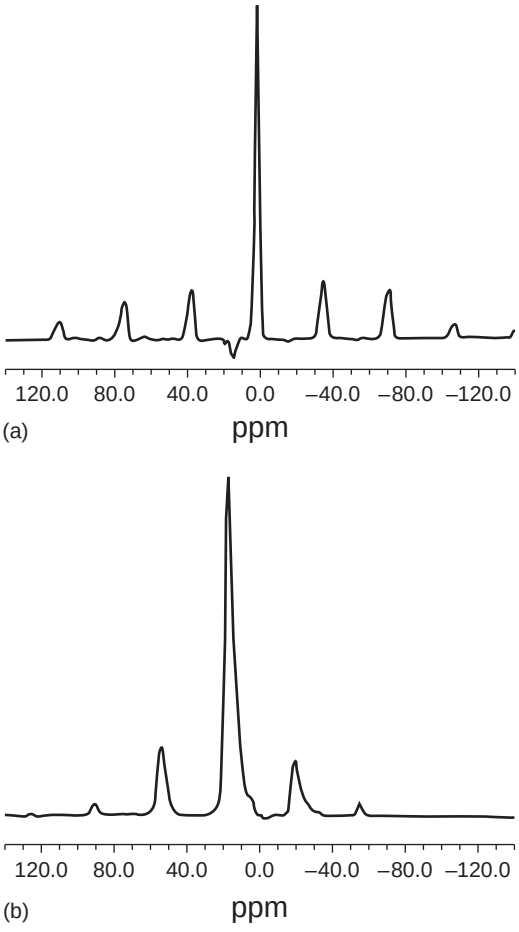

Figure 23 Slices through the F_1 (CRAMPS) dimension of the ^{1}H CRAMPS/MAS correlation spectrum of an $Na_2O \cdot 4SiO_2 \cdot 0.7H_2O$ glass showing "MAS" spectra from (a) the H_2O resonance at 4.0 ppm and (b) the OH resonance at 14.0 ppm. [Reproduced with permission from Schaller and Sebald.[119]]

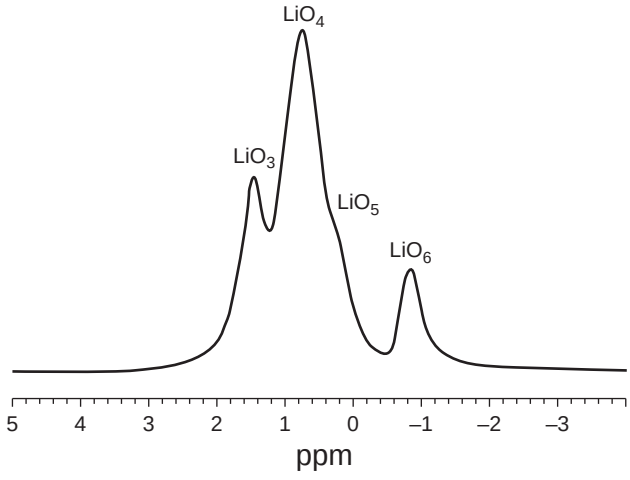

Figure 24 ^{6}Li MAS spectrum of ^{6}Li-enriched lithium orthosilicate, Li_4SiO_4. [Reproduced with permission from Xu and Stebbins.[70]]

results in readily accessible NMR spectra. ^{109}Ag has a broad chemical shift range of about 1000 ppm, making it sensitive to coordination environment. For example, silver iodide doping of silver phosphate glasses causes dramatic increases in ionic conductivity.[127] The ^{109}Ag chemical shifts in this system are strongly dependent on temperature and glass composition, the composition effect attributed to the difference in coordination between oxide and iodide.

In silver iodomolybdate glasses of the type $AgI–Ag_2MoO_4$, ^{109}Ag NMR gave evidence for two distinct populations of mobile Ag^+ ions.[128]

ZrF_4-based glasses are the most important class of anion conducting materials and have been extensively investigated by NMR spectroscopy. At low temperatures, ^{19}F relaxation is caused by low-frequency excitations of disordered modes in the glass. It has also been suggested that spin diffusion to paramagnetic impurities contributes to ^{19}F T_1 relaxation in this range. Diffusive ionic motions begin above about 300 K and this is reflected in narrowed line widths and a decrease in T_1 due to modulation of ^{19}F dipole–dipole interactions between 300 K and T_g. This is illustrated in Figure 25, which shows the temperature dependence of ^{19}F relaxation rate (T_1^{-1}) as a function of inverse temperature for a fluorozincate glass having a T_g of 520 K.[129]

3.3 Nuclear Magnetic Resonance Imaging

NMR imaging (NMRI or MRI) is becoming an increasingly important approach to the analysis of heterogeneous inorganic materials. Key reviews are listed in Table 1 and applications from the recent literature of relevance to inorganic oxides and minerals are summarized here (Table 15). Imaging of liquids in porous materials is an important area which is covered in more detail elsewhere.

Applications concerning the hydration and dehydration of oxide materials are well represented by studies on cement and concrete. Figure 26 shows 1-D STRAFI ^{1}H images of the penetration of water into two cylinders of Portland cement previously hydrated with D_2O.[130] H_2O was added to the tops of the cylinders and the STRAFI profiles after 30 min show the retardation of water uptake in sample A, whereas B is fully penetrated.

Relaxation time mapping using single-point imaging (SPI) techniques have also been applied to imaging of protons in concrete, enabling nuclei with fast relaxation times to be imaged[131] (Table 15). Imaging of protons in hydrous minerals has been demonstrated for the first time, in samples of stilbite, agate and enhydros from geothermal fields. Again this required techniques for

Table 14 Spin–lattice relaxation and ion mobility in glasses

NMR nucleus	Sample	NMR techniques	Analytical information	Ref.	Other related studies
^{23}Na, ^{7}Li	Disilicate and trisilicate glasses	Variable-temp. T_1	Ionic conduction mechanisms and mixed cation effects	349	350
^{7}Li, ^{31}P	Lithium silicophosphate glasses	Variable-temp. T_1 from 300 K to T_g	Diffusive motion of Li^+ ions	351	352, 353
^{7}Li, ^{31}P	0.35 LiF · 0.65 $LiPO_3$	T_1, $T_{1\rho}$	^{31}P NMR relaxation in glassy lithium fluorophosphates	124	278
^{6}Li, ^{7}Li	Lithium aluminate ceramics	T_1	Mobility and relaxation determinations of lithium	108	
^{7}Li	Lithium silicophosphate glasses	Variable-temp. T_1, T_2	Lithium dynamics in thioborates	354	355
^{7}Li, ^{19}F	ZrF_4-based fluoride glasses	^{7}Li and ^{19}F NMR	Ion conductivity and diffusion in ZrF_4-based fluoride glasses	284	
^{7}Li	Lithium lanthanum titanates	Variable-temp. (150–900 K) ^{7}Li T_1	A distribution of activation energies for the local and long-range ionic motion	356	357
^{7}Li, ^{11}B	Lithium triborate glass	^{11}B and ^{7}Li static and MAS NMR; ^{7}Li T_1	Dynamics in sol–gel versus melt-quenched lithium triborate glass	258	
^{7}Li	TeO_2–$LiO_{0.5}$ · LiX (X = F, Cl) glasses	Variable-temp. T_1	Li^+ ion mobility from NMR and impedance spectroscopy	358	
^{23}Na	NASICON superionic conductors	2-D nutation NMR	Ion dynamics	359	
^{7}Li, ^{6}Li	Natural lepidolites	^{6}Li MAS; ^{7}Li T_1	Lithium coordination number and relaxation	70	360, 361
^{7}Li, ^{11}B	$xLi_2S \cdot (1-x)B_2S_3$	T_1	Ionic motions in glassy fast ionic conductors	362	363
^{7}Li, ^{23}Na, ^{29}Si	(Li, Na) disilicate glasses	^{29}Si, ^{23}Na and ^{7}Li MAS NMR; ^{7}Li T_1	Mixed alkali metal effect and lithium mobility	69	
^{7}Li, ^{11}B	Lithium thioborate glasses	^{7}Li and ^{11}B NMR and T_1; PGSE diffusion	Boron sites in lithium thioborate glasses	364	365, 366, 367
^{7}Li, ^{23}Na	Mixed alkali metal thiogermanate glasses	^{23}Na{^{7}Li} (SEDOR)	Alkali metal ion distribution in mixed alkali metal chalcogenide glasses	126	72
^{31}P	Nd^{3+} phosphate glasses	Variable-temp. line widths and T_1	Correlation of NMR and AC susceptibility data	368	
^{19}F, ^{27}Al, ^{23}Na	Cryolite (Na_3AlF_6), chiolite ($Na_5Al_3F_{14}$)	Variable-temp. line shape and T_1	Diffusion and the dynamics of displacive phase transitions in cryolite and chiolite	369	
^{93}Nb, ^{107}Ag, ^{109}Ag	Ag_9GaSe_6, Ag_7NbS_6	Variable-temp. T_1	Mobility of silver ions in silver ion conductors	370	
^{109}Ag	$(AgI)_{1-x}(Ag_2MoO_4)_x$	T_1	Structure and cation dynamics	302	

imaging nuclei with short T_2 values; a spatial resolution of 0.31 mm was reported.[132] Drying of sodium silicate films was monitored using STRAFI. Spatially resolved T_2 images allowed the drying of the films to be investigated as a function of temperature.[133]

Solid-state imaging has not yet found widespread applications but, as mentioned in section 2.13, the range of nuclei which can potentially be detected include several of importance in inorganic oxides and minerals. ^{11}B imaging of solids such as borax having boron at sites of low symmetry has been demonstrated using STRAFI.[134] It is also clear that techniques already developed for imaging of bone minerals, for example, should be directly applicable to a much wider range of minerals and glasses. The difference in the ^{1}H–^{31}P CP efficiency in protonated compared with unprotonated phosphate ions was used to selectively image brushite and hydroxyapatite minerals.[135] Continuing developments in this field demonstrate the value of this approach in phosphate imaging,[136] although extension to a wider range of materials has not yet occurred.

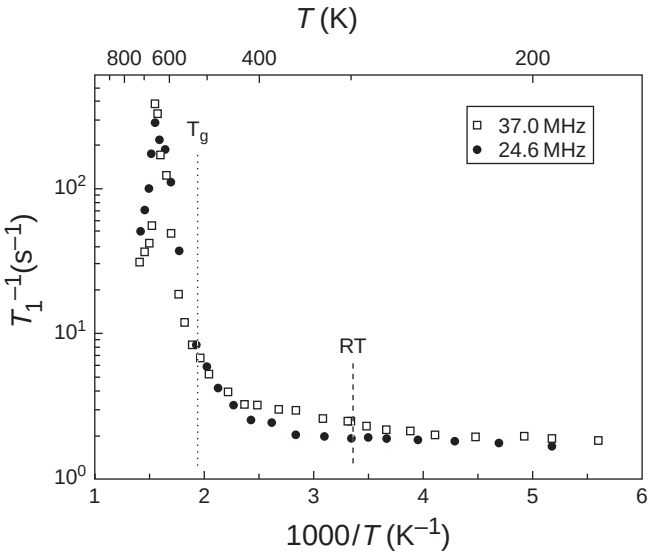

Figure 25 ^{19}F relaxation rate (T_1^{-1}) as a function of inverse temperature for a fluorozincate glass measured at 24.6 and 37 MHz. [Reproduced with permission from Auler et al.[129]]

4 FUTURE TRENDS

NMR spectroscopy, after 55 years of development, is now a mature technique which is routinely providing a great deal of structural, dynamic and quantitative information about glasses and geological materials. Commercial instrumentation is widely available, and the MAS experiment with a wide range of spin $= \frac{1}{2}$ nuclei is easily implemented by the nonspecialist. MAS probes are available with large volumes for high sensitivity, or with high speeds for high resolution, and the ever increasing magnetic field strengths available in superconducting magnets enhance both sensitivity and resolution.

In terms of spectroscopy, further advances in probe design for high-temperature in situ measurements will encourage wider use of NMR to study phase transitions, melts and high-temperature cation mobilities. The most striking advances in the past 5 years have come, however, in the observation of quadrupolar nuclei. The ability to extract not only chemical shift information but also quadrupole parameters makes such nuclei particularly powerful probes of local structure in solids. The techniques of DOR, DAS and MQMAS are all at a relatively early stage in their development, and many further developments in the design of both hardware and pulse sequences can be anticipated. The routine availability of these techniques to materials scientists will stimulate progress also in the theoretical understanding of structure and dynamics in noncrystalline materials.

One area of application which seems likely to make a large future impact is that of NMR imaging. It is clear from the recent literature that solid-state imaging techniques developed for bone minerals, for example, should also

Table 15 NMR imaging

NMR nucleus	Sample	NMR techniques	Analytical information	Ref.	Other related studies
^{31}P	Synthetic calcium phosphates and bone mineral	Differential CP solid-state imaging	Imaging of bone mineral	371	
^{31}P	Hydroxyapatite and bone	Solid state 3-D ^{31}P imaging	Bone mineral density	372	135
^{1}H, ^{31}P	Bone, tooth and calcium phosphates	3-D solid-state imaging	^{1}H and ^{31}P solid-state imaging of bone minerals	136	
^{1}H	White Portland cement	1-D SPI	Moisture profiles in Portland cement	373	374, 375
^{1}H, ^{27}Al	Portland cement	MAS NMR; ^{1}H STRAFI	Pozzolanic binders in Portland cement	130	376
^{1}H	Concrete	2-D imaging	Concrete freeze–thaw studied by MRI	377	378
^{1}H	Stilbite, agate and enhydros from geothermal fields	3-D imaging	NMR imaging of ^{1}H in hydrous minerals	132	
^{1}H	Epoxy–glass composites	2-D imaging	The interface of epoxy resin–glass fiber reinforced composites	379	
^{1}H	Sodium silicate films	STRAFI imaging	Drying of sodium silicate films	133	
^{1}H	Ethanol/water in aluminum oxide spheres	3-D solid-state imaging	Structure of porous catalyst spheres	380	
^{7}Li	LiF in a glass vial	STRAFI imaging	STRAFI imaging of quadrupolar nuclei of half-integer spin	47	
^{11}B	Borax; sodium fluoroborate	STRAFI imaging	NMR imaging of ^{11}B quadrupolar solids	134	
^{1}H	Mortar and concrete	SPI	Relaxation time mapping of short T_2^* nuclei	131	

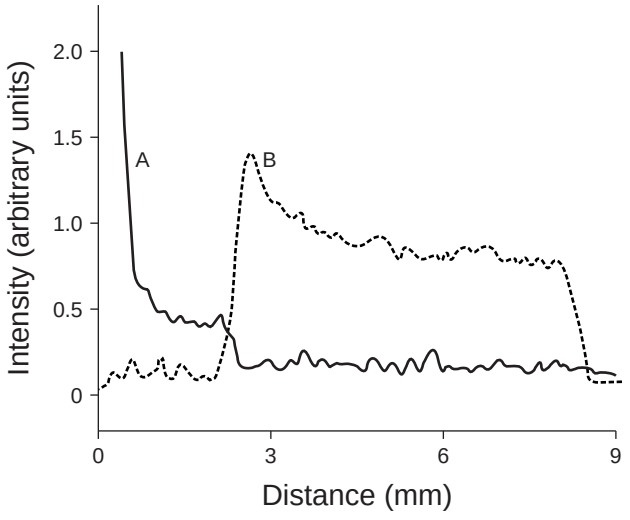

Figure 26 1-D STRAFI ^{1}H images of the penetration of water into two cylinders, A and B, of Portland cement. A was treated with siloxane to a depth of 3 mm and B was untreated. [Reproduced with permission from Hewlett et al.[130]]

be applicable to a much wider range of heterogeneous materials.

ACKNOWLEDGMENTS

We thank Yuni Krisnandi and Jixiu Jiang for their assistance in compiling references for this review.

ABBREVIATIONS AND ACRONYMS

COSY	Correlated Spectroscopy
CP	Cross-polarization
CPMAS	Cross-polarization/Magic-angle Spinning
CRAMPS	Combined Rotation and Multiple Pulse Spectroscopy
CW	Continuous Wave
DAS	Dynamic Angle Spinning
DEPT	Distortionless Enhancement by Polarization Transfer
DOR	Double Angle Rotation
EXSY	Two-dimensional Spin Exchange Spectroscopy
FID	Free Induction Decay
HETCOR	Heteronuclear Correlation Spectroscopy
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
INEPT	Insensitive Nuclei Enhanced by Polarization
MAS	Magic-angle Spinning
MQMAS	Multiple Quantum Magic-angle Spinning
MRI	Magnetic Resonance Imaging
MTES	Methyltriethoxysilane
NMR	Nuclear Magnetic Resonance
PGSE	Pulsed Gradient Spin Echo
REDOR	Rotational Echo Double Resonance
RF	Radiofrequency
SEDOR	Spin Echo Double Resonance
SPI	Single-point Imaging
STRAFI	Stray Field Imaging
TEDOR	Transferred Echo Double Resonance
TEOS	Tetraethoxysilane
TMOS	Tetramethoxysilane
TRAPDOR	Triple Resonance Adiabatic Passage Double Resonance
VACP	Variable-amplitude Cross-polarization
VAS	Variable-angle Spinning
WAHUHA	Waugh–Huber–Haeberlin
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
3QMAS	Triple Quantum Magic Angle Spinning

RELATED ARTICLES

Chemical Weapons Chemicals Analysis **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring

Food **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. H. Eckert, 'Structural Characterization of Noncrystalline Solids and Glasses Using Solid State NMR', *Prog. Nucl. Magn. Reson. Spectrosc.*, **24**(3), 159–293 (1992).
2. R.J. Kirkpatrick, 'MAS NMR Spectroscopy of Minerals and Glasses', *Rev. Mineral.*, **18**, 341–403 (1988).
3. J.F. Stebbins, P.F. McMillan, D.B. Dingwell, *Rev. Mineral.*, **32**, 191–246 (1995).
4. J.F. Stebbins, 'NMR Spectroscopy and Dynamic Processes in Mineralogy and Geochemistry', *Rev. Mineral.*, **18**, 405–429 (1988).
5. M.E. Smith, E.R.H. van Eck, 'Recent Advances in Experimental Solid State NMR Methodology for Half-integer Spin Quadrupolar Nuclei', *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**(2), 159–201 (1999).
6. C. Dybowski, G. Neue, 'Wide Line NMR', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 5042–5047, 1996.
7. T.J. Bastow, M.E. Smith, '^{91}Zr NMR of ZrO_2', *Solid State Nucl. Magn. Reson.*, **3**, 17–22 (1994).
8. I.J.F. Poplett, M.E. Smith, 'Field Swept Broad Line NMR Spectroscopy', *Solid State Nucl. Magn. Reson.*, **11**(2), 211–214 (1998).
9. B. Blumich (ed.), 'Solid-state NMR II: Inorganic Matter', in *NMR Basic Principles and Progress*, Springer, Heidelberg, 31, 1994.
10. J. Herzfeld, A.E. Berger, 'Sideband Analysis in MAS NMR of Solids', *J. Chem. Phys.*, **73**, 6021–6030 (1980).
11. D.P. Burum, 'Cross Polarization in Solids', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 1535–1542, 1996.
12. F. Engelke, 'Cross Polarization in Rotating Solids: Spin-1/2 Nuclei', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 1529–1535, 1996.
13. O.B. Peersen, X. Wu, I. Kustanovich, S.O. Smith, 'Variable Amplitude Cross Polarization', *J. Magn. Reson. Ser. A*, **104**, 334–339 (1993).
14. B.C. Gerstein, 'CRAMPS', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 1501–1509, 1996.
15. C.A. Fyfe, H. Gies, Y. Feng, '2 Dimensional 29Si NMR of ZSM-39', *J. Am. Chem. Soc.*, **111**, 7702–7707 (1989).
16. E.R.H. Van Eck, W.S. Veeman, 'Determination of Average ^{27}Al:^{31}P Distance in Aluminophosphate Molecular Sieves with SEDOR NMR', *J. Magn. Reson. Solids*, **1**, 1–6 (1992).
17. T. Gullion, J. Schaeffer, 'Detection of Weak Heteronuclear Dipolar Couplings by REDOR NMR', *Adv. Magn. Reson.*, **13**, 57–70 (1989).
18. E.R.H. Van Eck, W.S. Veeman, 'Spin Density Description of REDOR and TEDOR Solid State NMR', *Solid State Nucl. Magn. Reson.*, **2**, 307–315 (1993).
19. C.P. Grey, W.S. Veeman, A.J. Vega, 'Rotational Echo Triple Resonance Solid State Nuclear Magnetic Resonance: A Probe of ^{13}C:^{14}N Distances', *J. Chem. Phys.*, **98**, 7711–7715 (1993).
20. E.R.H. Van Eck, W.S. Veeman, 'Solid State 2D-heteronuclear ^{27}Al:^{31}P Correlation NMR Spectroscopy of Aluminophosphate VPI-5', *J. Am. Chem. Soc.*, **115**, 1168–1169 (1993).
21. L.B. Alemany, S. Steurnagel, J.P. Amoureux, R.L. Callender, A.R. Barron, 'Very Fast MAS and MQMAS NMR Studies of the Spectroscopically Challenging Minerals Kyanite and Andalusite on 400, 500 and 800 MHz Spectrometers', *Solid State Nucl. Magn. Reson.*, **14**(1), 1–18 (1999).
22. A. Samoson, E. Lipmaa, A. Pines, 'High-resolution Solid State NMR. Averaging of Second Order Effects by Means of a Double Rotor', *Mol. Phys.*, **65**, 1013–1018 (1988).
23. E.W. Wooten, K.T. Mueller, A. Pines, 'New Angles in NMR Sample Spinning', *Acc. Chem. Res.*, **25**, 209–215 (1992).
24. I. Farnan, P.J. Grandinetti, B.J.H. Baltisberger, J.F. Stebbins, U. Werner, M.A. Eastman, A. Pines, 'Quantification of the Disorder in Network-modified Silicate Glasses', *Nature (London)*, **358**, 31–35 (1992).
25. P.J. Grandinetti, 'Dynamic Angle Spinning NMR', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 1768–1776, 1996.
26. E. Kundla, A. Samoson, E. Lipmaa, 'Nutation Spectroscopy of Quadrupolar Nuclei', *Chem. Phys. Lett.*, **83**, 229–232 (1981).

27. B.C. Gerstein, 'Quadrupole Nutation Spectroscopy', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 3360–3364, 1996.
28. A. Medek, J.S. Harwood, L. Frydman, 'MQ MAS NMR: A New Method for the Study of Quadrupolar Nuclei in Solids', *J. Am. Chem. Soc.*, **117**, 12 779–12 787 (1995).
29. P.J. Dirken, S.C. Kohn, M.E. Smith, E.R.H. van Eck, 'Complete Resolution of Si–O–Si and Si–O–Al Fragments in an Aluminosilicate Glass by ^{17}O Multiple Quantum MAS NMR Spectroscopy', *Chem. Phys. Lett.*, **266**, 568–574 (1997).
30. S.H. Wang, Z. Xu, J.H. Baltisberger, L.M. Bull, J.F. Stebbins, A. Pines, 'Multiple-quantum Magic-angle Spinning and Dynamic-angle Spinning NMR Spectroscopy of Quadrupolar Nuclei', *Solid State Nucl. Magn. Reson.*, **8**(1), 1–16 (1997).
31. T. Anupold, A. Reinhold, P. Sarv, A. Samoson, 'A Comparison of DOR and MQMAS Spectra', *Solid State Nucl. Magn. Reson.*, **13**(1), 87–91 (1998).
32. A.P.M. Kentgens, 'A Practical Guide to Solid-state NMR of Half-integer Quadrupolar Nuclei with Some Applications to Disordered Systems', *Geoderma*, **80**(3–4), 271–306 (1997).
33. J.F. Stebbins, 'Nuclear Magnetic Resonance at High Temperature', *Chem. Rev.*, **91**, 1353–1373 (1991).
34. M.T. Janicke, C.C. Landry, S.C. Christiansen, S. Birtalan, G.D. Stucky, B.F. Chmelka, 'Low-silica MCM-41 Composites and Mesoporous Solids', *Chem. Mater.*, **11**(5), 1342–1351 (1999).
35. G. Kowalczyk, J.E. Roberts, 'Solid State ^{29}Si NMR Determination of Crystalline Silica in Natural Iron Oxide Pigments', *Anal. Chem. Acta*, **286**(1), 25–35 (1994).
36. G. Metz, M. Ziliox, S.O. Smith, 'Towards Quantitative CP/MAS NMR', *Solid State Nucl. Magn. Reson.*, **7**(3), 155–160 (1996).
37. J. Mahler, A. Sebald, 'Deconvolution of ^{29}Si Magic-angle Spinning Nuclear Magnetic Resonance Spectra of Silicate Glasses Revisited – Some Critical Comments', *Solid State Nucl. Magn. Reson.*, **5**(1), 63–78 (1995).
38. J.M. Koons, P.D. Ellis, 'Applicability of Factor Analysis in Solid State NMR', *Anal. Chem.*, **67**(23), 4309–4315 (1995).
39. C.T.G. Knight, S.D. Kinrade, 'Silicon-29 Nuclear Magnetic Resonance Spectroscopy Detection Limits', *Anal. Chem.*, **71**(1), 265–267 (1999).
40. M.D. Bruch (ed.), *NMR Spectroscopy Techniques*, 2nd edition, Practical Spectroscopy Series, Marcel Dekker, New York, Vol. 21, 1996.
41. R. Kimmich, *NMR Tomography Diffusometry Relaxometry*, Springer, Berlin, 1997.
42. Z. Xu, J.F. Stebbins, 'Cation Dynamics and Diffusion in Lithium Orthosilicate – Two-dimensional Lithium-6 NMR', *Science*, **270**(5240), 1332–1334 (1995).
43. P.T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Clarendon Press, Oxford, 1991.
44. J.B. Miller, 'NMR Imaging of Materials', *Prog. Nucl. Magn. Reson. Spectrosc.*, **33**(4), 273–308 (1998).
45. P.J. McDonald, B. Newling, 'Stray Field Magnetic Resonance Imaging', *Rep. Prog. Phys.*, **61**(11), 1441–1493 (1998).
46. P.J. McDonald, 'Stray Field Magnetic Resonance Imaging', *Prog. Nucl. Magn. Reson. Spectrosc.*, **30**(1–2), 69–99 (1997).
47. P. Bodart, T. Nunes, E.W. Randall, 'Stray-field Imaging of Quadrupolar Nuclei of Half Integer Spin in Solids', *Solid State Nucl. Magn. Reson.*, **8**(4), 257–263 (1997).
48. O. Zuger, S.T. Hoen, C.S. Yannoni, D. Rugar, 'Three-dimensional Imaging with a Nuclear Magnetic Resonance Force Microscope', *J. Appl. Phys.*, **79**(4), 1881–1884 (1996).
49. D. Rugar, O. Zuger, S. Hoen, C.S. Yannoni, H.M. Vieth, R.D. Kendrick, 'Force Detection of Nuclear Magnetic Resonance', *Science*, **264**(5165), 1560–1563 (1994).
50. H. Eckert, 'Solid State NMR as a Tool of Structure and Dynamics in Solid State Chemistry and Materials Science – Recent Progress and Challenges', *Curr. Opin. Solid State Mater. Sci.*, **1**(4), 465–476 (1996).
51. J.W. Zwanziger, 'Structure and Chemical Modification in Oxide Glasses', *Int. Rev. Phys. Chem.*, **17**(1), 65–90 (1998).
52. R.J. Kirkpatrick, 'MAS NMR Spectroscopy of Minerals and Glasses', *Rev. Mineral.*, **18**, 341–403 (1988).
53. J. Mahler, A. Sebald, 'Deconvolution of ^{29}Si MAS NMR Spectra of Silicate Glasses Revisited – Some Critical Comments', *Solid State Nucl. Magn. Reson.*, **5**(1), 63–78 (1995).
54. F. Fayon, C. Bessada, D. Massiot, I. Farnan, J.P. Coutures, '^{29}Si and ^{207}Pb NMR Study of Local Order in Lead Silicate Glasses', *J. Non-cryst. Solids*, **234**, 403–408 (1998).
55. S.K. Lee, J.F. Stebbins, 'The Degree of Aluminum Avoidance in Aluminosilicate Glasses', *Am. Mineral.*, **84**(5–6), 937–945 (1999).
56. R. Wolff, H. Jancke, R. Radeglia, 'On Reasons of ^{29}Si NMR Chemical Shift/Structure Relations for Silicon Oxides, Nitrides, and Carbides – an Individual-gauge-for-localized-orbitals Study', *Solid State Nucl. Magn. Reson.*, **9**(2–4), 177–189 (1997).
57. A.D. Grimmer, F. von Lampe, M. Magi, 'Solid State High Resolution ^{29}Si MAS NMR of Silicates with Sixfold Coordinated Silicon', *Chem. Phys. Lett.*, **132**, 549–553 (1986).
58. J.F. Stebbins, P. McMillan, 'Compositional and Temperature Effects on Five Coordinated Silicon in Ambient Pressure Silicate Glasses', *J. Non-cryst. Solids*, **160**, 116–125 (1993).
59. C.T.G. Knight, R.J. Kirkpatrick, E. Oldfield, 'The Connectivity of Silicon Sites in Silicate Glasses as Determined by Two Dimensional ^{29}Si NMR Spectroscopy', *J. Non-cryst. Solids*, **116**(1), 140–144 (1990).

60. K. Glock, O. Hirsch, P. Rehak, B. Thomas, C. Jaeger, 'Novel Opportunities for Studying the Short and Medium Range Order of Glasses by MAS NMR, ^{29}Si Double Quantum NMR and IR Spectroscopies', *J. Non-cryst. Solids*, **232–234**, 113–118 (1998).
61. M. Fieke, R. Graf, I. Schnell, C. Jaeger, H.W. Spiess, 'Structure of Crystalline Phosphates from ^{31}P Double Quantum NMR Spectroscopy', *J. Am. Chem. Soc.*, **118**, 9631–9634 (1996).
62. P.J. Dirken, G.H. Nactegaal, A.P.M. Kentgens, 'Off Resonance Nutation NMR Study of Framework Aluminosilicate Glasses with Li, Na, K, Rb or Cs as Charge Balancing Cations', *Solid State Nucl. Magn. Reson.*, **5**(1), 189–200 (1995).
63. E. Lipmaa, A. Samoson, M. Magi, 'High Resolution ^{27}Al NMR of Aluminosilicates', *J. Am. Chem. Soc.*, **108**, 1730–1735 (1986).
64. C. Mundus, W. Mueller-Warmuth, '^{27}Al MAS NMR Satellite Transition Spectroscopy of Glasses in the System $K_2O-Al_2O_3-SiO_2$', *Solid State Nucl. Magn. Reson.*, **5**(1), 79–88 (1995).
65. J.F. Stebbins, J.V. Oglesby, Z. Xu, 'Disorder Among Network-modifier Cations in Silicate Glasses – New Constraints from Triple-quantum ^{17}O NMR', *Am. Mineral.*, **82**(11–12), 1116–1124 (1997).
66. J.F. Stebbins, S.K. Lee, J.V. Oglesby, 'Al–O–Al Oxygen Sites in Crystalline Aluminates and Aluminosilicate Glasses: High-resolution Oxygen-17 NMR Results', *Am. Mineral.*, **84**(5–6), 983–986 (1999).
67. Z. Xu, H. Maekawa, J.V. Oglesby, J.F. Stebbins, 'Oxygen Speciation in Hydrous Silicate Glasses – an Oxygen-17 NMR Study', *J. Am. Chem. Soc.*, **120**(38), 9894–9901 (1998).
68. T. Schaller, A. Sebald, 'One and Two Dimensional ^{1}H MAS Experiments on Hydrous Silicate Glasses', *Solid State Nucl. Magn. Reson.*, **5**(1), 89–102 (1995).
69. F. Ali, A.V. Chadwick, G.N. Greaves, M.C. Jermy, K.L. Ngai, M.E. Smith, 'Examination of the Mixed-alkali Effect in (Li, Na) Disilicate Glasses by Nuclear Magnetic Resonance and Conductivity Measurements', *Solid State Nucl. Magn. Reson.*, **5**(1), 133–143 (1995).
70. Z. Xu, J.F. Stebbins, '^{6}Li Nuclear Magnetic Resonance Chemical Shifts, Coordination Number and Relaxation in Crystalline and Glassy Silicates', *Solid State Nucl. Magn. Reson.*, **5**(1), 103–112 (1995).
71. B. Gee, H. Eckert, '^{23}Na Nuclear Magnetic Resonance Spin Echo Decay Spectroscopy of Sodium Silicate Glasses and Crystalline Model Compounds', *Solid State Nucl. Magn. Reson.*, **5**(1), 113–122 (1995).
72. B. Gee, H. Eckert, 'Cation Distribution in Mixed-alkali Silicate Glasses – NMR Studies by ^{23}Na(^{7}Li) and ^{23}Na(^{6}Li) Spin Echo Double Resonance', *J. Phys. Chem.*, **100**(9), 3705–3712 (1996).
73. L. Vanwullen, B. Gee, L. Zuchner, M. Bertmer, H. Eckert, 'Connectivities and Cation Distributions in Oxide Glasses – New Results From Solid State NMR', *Ber. BunsenGes. Phys. Chem.*, **100**(9), 1539–1549 (1996).
74. B. Gee, H. Eckert, 'Cation Ordering Scenarios in Mixed Alkali Silicate Glasses – Experimental Constraints from ^{23}Na(^{6}Li) Spin Echo Double Resonance NMR', *Ber. BunsenGes. Phys. Chem.*, **100**(9), 1610–1616 (1996).
75. R. Hussin, R. Dupree, D. Holland, 'The Ge–O–Ge Bond Angle Distribution in GeO_2', *J. Non-cryst. Solids*, **246**, 159–168 (1999).
76. R.E. Youngman, J.W. Zwanziger, 'Multiple Boron Sites in Borate Glass Detected with Dynamic Angle Spinning Nuclear Magnetic Resonance', *J. Non-cryst. Solids*, **168**(3), 293–297 (1994).
77. C. Jäger, K. Herzog, B. Thomas, M. Feike, G. Kunath-Fandrei, 'Detection of Multiple Boron Sites in Glasses by ^{11}B Satellite Transition NMR Spectroscopy', *Solid State Nucl. Magn. Reson.*, **5**(1), 51–62 (1995).
78. L. Van Wuellen, L. Zuechner, W. Mueller-Warmuth, H. Eckert, '^{11}B(^{27}Al) and ^{27}Al(^{11}B) Double Resonance Experiments on a Glassy Sodium Aluminoborate', *Solid State Nucl. Magn. Reson.*, **6**(1), 203–212 (1996).
79. S.H. Wang, J.F. Stebbins, 'On the Structure of Borosilicate Glasses – a Triple-quantum Magic-angle Spinning ^{17}O Nuclear Magnetic Resonance Study', *J. Non-cryst. Solids*, **231**(3), 286–290 (1998).
80. R.J. Kirkpatrick, R.K. Brow, 'Nuclear Magnetic Resonance Investigation of the Structures of Phosphate and Phosphate-containing Glasses – a Review', *Solid State Nucl. Magn. Reson.*, **5**(1), 9–21 (1995).
81. P. Losso, U. Sternberg, 'Distribution of Principal Values of the ^{31}P Chemical Shift Tensor in Phosphate Glasses', *Solid State Nucl. Magn. Reson.*, **13**(1), 113–118 (1998).
82. P. Hartmann, J. Vogel, C. Jager, '2D Exchange NMR Investigations of Phosphate Glass Ceramic', *Ber. BunsenGes. Phys. Chem.*, **100**(9), 1658–1660 (1996).
83. T. Schaller, C. Rong, M.J. Toplis, H. Cho, 'TRAPDOR NMR Investigations of Phosphorus Bearing Aluminosilicate Glasses', *J. Non-cryst. Solids*, **248**, 19–27 (1999).
84. J.M. Miller, 'Fluorine-19 Magic-angle Spinning NMR', *Prog. Nucl. Magn. Reson. Spectrosc.*, **28**(3–4), 255–281 (1996).
85. B. Bureau, G. Silly, J.Y. Buzare, J. Emery, C. Legein, C. Jacoboni, 'Investigation of Fluorine Octahedron Connectivities in Transition Metal Fluoride Glasses by Solid-state ^{19}F Magic-angle-spinning Nuclear Magnetic Resonance Spectroscopy', *J. Phys. Condens. Matter*, **9**(31), 6719–6736 (1997).
86. M. Pauthe, E. Bernstein, J. Dumas, L. Saviot, A. Pradel, M. Ribes, 'Preparation and Characterization of Si Nanocrystallites Embedded in a Silica Matrix', *J. Mater. Chem.*, **9**(1), 187–191 (1999).
87. C. Guermeur, J. Lambard, J.F. Gerard, C. Sanchez, 'Hybrid Polydimethylsiloxane–Zirconium Oxo Nanocomposites. Part 1. Characterization of the Matrix and the Siloxane–Zirconium Oxo Interface', *J. Mater. Chem.*, **9**(3), 769–778 (1999).

88. M.P.J. Peeters, A.P.M. Kentgens, 'A ^{27}Al MAS, MQMAS and Off-resonance Nutation NMR Study of Aluminum Containing Silica-based Sol–Gel Materials', *Solid State Nucl. Magn. Reson.*, **9**(2–4), 203–217 (1997).
89. G.D. Soraru, N. Dallabona, C. Gervais, F. Babonneau, 'Organically Modified SiO_2–B_2O_3 Gels Displaying a High Content of Borosiloxane (=B–O–Si) Bonds', *Chem. Mater.*, **11**(4), 910–919 (1999).
90. P.J. Dirken, M.E. Smith, H.J. Whitfield, '^{17}O and ^{29}Si Solid State NMR Study of Atomic Scale Structure in Sol–Gel-prepared TiO_2–SiO_2 Materials', *J. Phys. Chem.*, **99**(1), 395–401 (1995).
91. C.J. Brinker, G.W. Scherer, *Sol–Gel Science: The Physics and Chemistry of Sol–Gel Processing*, Academic Press, Boston, MA, 1990.
92. J. Sefcik, A.V. McCormick, 'Kinetic and Thermodynamic Issues in the Early Stages of Sol–Gel Processes Using Silicon Alkoxides', *Catal. Today*, **35**(3), 205–223 (1997).
93. F. Devreux, J.P. Boilot, F. Chaput, A. Lecomte, 'Sol–Gel Condensation of Rapidly Hydrolyzed Silicon Alkoxides: A Joint ^{29}Si NMR and Small-angle X-ray Scattering Study', *Phys. Rev. A*, **41**(12), 6901–6909 (1990).
94. W.H. Dokter, H.F. Vangarderen, T.P.M. Beelen, J.W. Dehaan, L.J.M. Vandeven, R.A. Vansanten, 'Gel Transformations in Silicas – a Combined NMR and SAXS Study', *Colloids Surf. A*, **72**, 165–171 (1993).
95. F. Brunet, P. Lux, J. Virlet, 'New NMR Tools for the Characterization of Hybrid Systems', *New J. Chem.*, **18**(10), 1059–1064 (1994).
96. F. Brunet, 'Polymerization Reactions in Methyltriethoxysilane Studied Through ^{29}Si NMR with Polarization Transfer', *J. Non-cryst. Solids*, **231**(1–2), 58–77 (1998).
97. M.P.J. Peeters, T.N.M. Bernards, M.J. Van Bommel, '^{17}O NMR of Sol–Gel Processes of TEOS and TMOS', *J. Sol–Gel Sci. Technol.*, **13**(1–3), 71–74 (1998).
98. M.T. Dove, 'The Use of ^{29}Si MAS/NMR and Monte Carlo Methods in the Study of Al/Si Ordering in Silicates', *Geoderma*, **80**(3–4), 353–368 (1997).
99. S.C. Kohn, C.M.B. Henderson, R. Dupree, 'NMR Studies of the Leucite Analogues', *Phys. Chem. Miner.*, **21**, 176–90 (1994).
100. H. Graetsch, H. Gies, I. Topalovic, 'NMR, XRD and IR Study on Microcrystalline Opals', *Phys. Chem. Miner.*, **21**(3), 166–175 (1994).
101. J.J. Fitzgerald, A.I. Hamza, C.E. Bronnimann, S.F. Dec, 'Studies of the Solid/Solution Interfacial Dealumination of Kaolinite in HCl(aq) using Solid-state ^{1}H CRAMPS and SP/MAS ^{29}Si NMR Spectroscopy', *J. Am. Chem. Soc.*, **119**(30), 7105–7113 (1997).
102. Z. Xu, J.F. Stebbins, 'Oxygen Sites in the Zeolite Stilbite – a Comparison of Static, MAS, VAS, DAS and Triple Quantum MAS NMR Techniques', *Solid State Nucl. Magn. Reson.*, **11**(3–4), 243–251 (1998).
103. M. Hunger, P. Sarv, A. Samoson, 'Two Dimensional Triple Quantum ^{23}Na MAS NMR Spectroscopy of Sodium Cations in Dehydrated Zeolites', *Solid State Nucl. Magn. Reson.*, **9**(1), 115–120 (1997).
104. D.J. Sullivan, J.S. Shore, J.A. Rice, 'Assessment of Cation Binding to Clay Minerals Using Solid-state NMR', *Clays Clay Miner.*, **46**(3), 349–354 (1998).
105. I. Abrahams, K. Franks, G.E. Hawkes, G. Philippou, J. Knowles, P. Bodart, T. Nunes, '^{23}Na, ^{27}Al and ^{31}P NMR and X-ray Powder Diffraction Study of Na/Ca/Al Phosphate Glasses and Ceramics', *J. Mater. Chem.*, **7**(8), 1573–1580 (1997).
106. F. Ali, M.E. Smith, S. Steuernagel, H.J. Whitfield, '^{17}O Solid-state NMR Examination of La_2O_3 Formation', *J. Mater. Chem.*, **6**(2), 261–264 (1996).
107. J. Brieger, R. Merkle, H. Bertagnolli, K. Muller, 'Investigation of the Pyrolysis and Crystallization of Lead Zirconate Titanate Ceramics Prepared via the Sol Gel Process', *Ber. BunsenGes. Phys. Chem.*, **102**(10), 1376–1386 (1998).
108. F.F. Stewart, J.F. Stebbins, E.S. Peterson, I. Farnan, S.O. Dunham, E. Adams, P.W. Jennings, 'Mobility and Relaxation Determinations of Lithium in Lithium Aluminate Ceramics Using Solid-state NMR Spectroscopy', *Chem. Mater.*, **7**(2), 363–367 (1995).
109. P.F. McMillan, G.H. Wolf, *Rev. Mineral.*, **32**, 247–315 (1995).
110. A.M. George, J.F. Stebbins, 'Structure and Dynamics of Magnesium in Silicate Melts – a High-temperature ^{25}Mg NMR Study', *Am. Mineral.*, **83**(9–10), 1022–1029 (1998).
111. J. Yarger, J. Diefenbacher, B.T. Poe, G.H. Wolf, P.F. McMillan, 'High Coordinate Si and Al in Sodium Aluminosilicate Glasses Quenched from High Pressure', *Trans. Am. Geophys. Union*, **75**, 713–718 (1994).
112. X.Y. Xue, J.F. Stebbins, M. Kanzaki, 'Correlations Between ^{17}O NMR Parameters and Local Structure Around Oxygen in High-pressure Silicates – Implications for the Structure of Silicate Melts at High Pressure', *Am. Mineral.*, **79**(1–2), 31–42 (1994).
113. I. Farnan, 'Structure and Dynamics in Glassy and Molten Silicates', *Curr. Opin. Solid State Mater. Sci.*, **3**(4), 371–377 (1998).
114. I. Farnan, J.F. Stebbins, 'High-temperature ^{29}Si NMR Investigation of Solid and Molten Silicates', *J. Am. Chem. Soc.*, **112**(1), 32–39 (1990).
115. S. Sen, Z. Xu, J.F. Stebbins, 'Temperature Dependent Structural Changes in Borate, Borosilicate and Boroaluminate Liquids–High-resolution ^{11}B, ^{29}Si and ^{27}Al NMR Studies', *J. Non-cryst. Solids*, **226**(1–2), 29–40 (1998).
116. R. Maxwell, H. Eckert, 'Molten-state Kinetics in Glass-forming Systems – a High-temperature NMR Study of the System Phosphorus–Selenium', *J. Phys. Chem.*, **99**(13), 4768–4778 (1995).
117. H. Maekawa, T. Saito, T. Yokokawa, 'Water in Silicate Glass – ^{17}O NMR of Hydrous Silica, Albite, and

$Na_2Si_4O_9$ Glasses', *J. Phys. Chem. B*, **102**(39), 7523–7529 (1998).
118. H. Graetsch, I. Topalovicdierdorf, 'MAS NMR Spectra of Hyalite from Gran Canaria', *Chem. Erde-Geochem.*, **56**(4), 387–391 (1996).
119. T. Schaller, A. Sebald, 'One- and Two-dimensional ^{1}H Magic-angle Spinning Experiments on Hydrous Silicate Glasses', *Solid State Nucl. Magn. Reson.*, **5**(1), 89–102 (1995).
120. L.J. Schreiner, J.C. McTavish, J.C. Miljkovic, M.M. Pintar, R. Blinc, G. Lahajnar, D. Lasic, L.W. Reeves, 'NMR Line Shape Spin–Lattice Relaxation Correlation Study of Portland Cement Hydration', *J. Am. Ceram. Soc.*, **68**, 10–16 (1985).
121. C.A. Weiss, W.V. Gerasimowicz, 'Interaction of Water with Clay Minerals as Studied by ^{2}H Nuclear Magnetic Resonance Spectroscopy', *Geochim. Cosmochim. Acta*, **60**(2), 265–275 (1996).
122. J. Klotz, W. Brostow, M. Hess, W.S. Veeman, 'Epoxy and Glass Composites in Water Studied With ^{2}H NMR', *Polym. Eng. Sci.*, **36**(8), 1129–1133 (1996).
123. S.V. Dvinskikh, I.V. Murin, A.F. Privalov, A.A. Pronkin, E. Rossler, H.M. Vieth, 'Microscopic Structure of the Glassy Ionic Conductor $X \cdot LiF + (1 - X) \cdot LiPO_3$ from NMR Data', *J. Non-cryst. Solids*, **240**(1–3), 79–90 (1998).
124. S. Berger, J. Roos, A.Y. Zavidonov, D. Brinkmann, '^{31}P NMR Relaxation in Glassy $0.35\ LiF \cdot 0.65LiPO_3$', *Solid State Ionics*, **112**(1–2), 87–93 (1998).
125. R.M. Wenslow, K.T. Mueller, 'Cation Sites in Mixed-alkali Phosphate Glasses', *J. Non-cryst. Solids*, **231**(1–2), 78–88 (1998).
126. B. Gee, H. Eckert, A. Pradel, G. Taillades, M. Ribes, 'Alkali Ion Distribution in Mixed-alkali Chalcogenide Glasses – ^{23}Na(^{7}Li) Spin Echo Double Resonance NMR Studies of the System', *J. Non-cryst. Solids*, **215**(1), 32–40 (1997).
127. K.K. Olsen, J.W. Zwanziger, 'Multi-nuclear and Multi-dimensional Nuclear Magnetic Resonance Investigation of Silver Iodide Silver Phosphate Fast Ion Conducting Glasses', *Solid State Nucl. Magn. Reson.*, **5**(1), 123–132 (1995).
128. P. Mustarelli, C. Tomasi, A. Magistris, M. Cutroni, 'Ion Dynamics and Devitrification in $0.75AgI–0.25Ag_2MoO_4$ Fast Ion Conducting Glass – an XRD, DSC and ^{109}Ag NMR Study', *J. Non-cryst. Solids*, **234**, 532–538 (1998).
129. T. Auler, J.P. Donoso, P.L. Frare, C.J. Magon, Y. Messaddeq, A.A.S.T. Delben, 'Nuclear Magnetic Relaxation of ^{19}F in Fluoride Glasses', *J. Non-cryst. Solids*, **247**, 92–97 (1999).
130. P. Hewlett, G. Hunter, R. Jones, 'Bridging the Gaps', *Chem. Br.*, **35**(1), 40–43 (1999).
131. S.D. Beyea, B.J. Balcom, P.J. Prado, A.R. Cross, C.B. Kennedy, R.L. Armstrong, T.W. Bremner, 'Relaxation Time Mapping of Short T_2^* Nuclei with Single-point Imaging (SPI) Methods', *J. Magn. Reson.*, **135**(1), 156–164 (1998).
132. Y. Nakashima, S. Nakashima, D. Gross, K. Weiss, K. Yamauchi, 'NMR Imaging of ^{1}H in Hydrous Minerals', *Geothermics*, **27**(1), 43–53 (1998).
133. P.D.M. Hughes, P.J. McDonald, N.P. Rhodes, J.W. Rockliffe, E.G. Smith, J. Wills, 'A Stray Field Magnetic Resonance Imaging Study of the Drying of Sodium Silicate Films', *J. Colloid Interface Sci.*, **177**(1), 208–213 (1996).
134. E.W. Randall, D.G. Gillies, 'Multinuclear NMR Imaging of ^{11}B Quadrupolar Solids with the Fringe-field Method', *J. Magn. Reson. Ser. A*, **121**(2), 217–220 (1996).
135. J.R. Moore, L. Garrido, J.L. Ackerman, 'Solid State Phosphorus-31 Magnetic Resonance Imaging of Bone Mineral', *Magn. Reson. Med.*, **33**(3), 293–299 (1995).
136. Y.T. Wu, D.A. Chesler, M.J. Glimcher, L. Garrido, J.X. Wang, H.J. Jiang, J.L. Ackerman, 'Multinuclear Solid-state Three-dimensional MRI of Bone and Synthetic Calcium Phosphates', *Proc. Nat. Acad. Sci. USA*, **96**(4), 1574–1578 (1999).
137. J. Klinowski, 'Solid-state NMR Studies of Molecular Sieve Catalysis', *Chem. Rev.*, **91**, 1459–1479 (1991).
138. M.A. Wilson, *NMR Techniques and Applications in Geochemistry and Soil Chemistry*, Pergamon Press, Oxford, 1987.
139. R.K. Harris, P. Jackson, 'High-resolution Fluorine-19 Magnetic Resonance of Solids', *Chem. Rev.*, **91**, 1427–1440 (1991).
140. C.A. Fyfe, *Solid State NMR for Chemists*, C.F.C. Press, Guelph, 1983.
141. H. Eckert, 'Special Issue on Glasses – Preface', *Solid State Nucl. Magn. Reson.*, **5**(1), 1–2 (1995).
142. A.T. Watson, C.T.P. Chang, 'Characterizing Porous Media with NMR Methods', *Prog. Nucl. Magn. Reson. Spectrosc.*, **31**(4), 343–386 (1997).
143. E. Brunner, U. Sternberg, 'Solid-state NMR Investigations on the Nature of Hydrogen Bonds', *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**(1), 21–57 (1998).
144. P.J. Bray, 'NMR and NQR Studies of Boron in Vitreous and Crystalline Borates', *Inorg. Chim. Acta*, **289**(1–2), 158–173 (1999).
145. S. Sakka, 'Structure, Properties and Application of Oxynitride Glasses', *J. Non-cryst. Solids*, **181**(3), 215–224 (1995).
146. J. Brus, P. Kotlik, 'Application of NMR Spectrometry in the Study of Alkoxysilane Polycondensation', *Chem. Listy*, **90**(5), 316–323 (1996).
147. B. Winkler, 'The Dynamics of H_2O in Minerals', *Phys. Chem. Miner.*, **23**(4–5), 310–318 (1996).
148. L.W. Jelinski, M.T. Melchior, 'High Resolution NMR of Solids', in *NMR Spectroscopy Techniques*, ed. M.D. Bruch, Marcel Dekker, New York, 417–485, 1996.
149. R.E. Botto, G.D. Cody, S.L. Dieckman, D.C. French, N. Gopalsami, P. Rizo, 'Three-dimensional Magnetic

Resonance Microscopy of Materials', *Solid State Nucl. Magn. Reson.*, **6**(4), 389–402 (1996).
150. S. Hafner, D.E. Demco, R. Kimmich, 'Magic Echoes and NMR Imaging of Solids', *Solid State Nucl. Magn. Reson.*, **6**(4), 275–293 (1996).
151. R.E. Botto, 'Special Issue on Magnetic Resonance Imaging of Materials – Preface', *Solid State Nucl. Magn. Reson.*, **6**(4), R6 (1996).
152. G. Neue, U. Skibbe, C. Dybowski, 'Solid-state NMR as an Analytical Technique Offering Spatially Resolved Chemical Information', *Fresenius' J. Anal. Chem.*, **349**(1–3), 112–116 (1994).
153. M. Casu, F.C. Marincola, A. Lai, A. Musinu, G. Piccaluga, 'A ^{29}Si MAS and ^{1}H NMR Investigation of Fe_2O_3–SiO_2 Nanocomposites', *J. Non-cryst. Solids*, **234**, 329–334 (1998).
154. W.B. Chen, H.Q. Feng, C.H. Ye, 'Synthesis and Characterization of Styrene–Butadiene Block Copolymer/Silicate Hybrid Materials Via the Sol–Gel Process I – Four-arm Star Styrene–Butadiene Block Copolymer/Silica Hybrids', *Polym. J.*, **29**(12), 992–996 (1997).
155. W.B. Chen, H.Q. Feng, D.Y. He, C.H. Ye, 'High Resolution Solid-state NMR and DSC Study of Poly(ethylene glycol) Silicate Hybrid Materials Via Sol–Gel Process', *J. Appl. Polym. Sci.*, **67**(1), 139–147 (1998).
156. K. Haraguchi, Y. Usami, Y. Ono, 'The Preparation and Characterization of Hybrid Materials Composed of Phenolic Resin and Silica', *J. Mater. Sci.*, **33**(13), 3337–3344 (1998).
157. Q. Deng, W. Jarrett, R.B. Moore, K.A. Mauritz, 'Novel Nafion(R)/Ormosil Hybrids Via In Situ Sol–Gel Reactions. 2. Probe of Ormosil Phase Nanostructure by ^{29}Si Solid State NMR Spectroscopy', *J. Sol–Gel Sci. Technol.*, **7**(3), 177–190 (1996).
158. J.P. Bezombes, C. Chuit, R.J.P. Corriu, C. Reye, 'Preparation and Characterization of New Organic–Inorganic Hybrid Materials Incorporating Phosphorus Centres', *J. Mater. Chem.*, **8**(8), 1749–1759 (1998).
159. F. Babonneau, 'NMR Characterization of Hybrid Siloxane–Oxide Materials', *New J. Chem.*, **18**(10), 1065–1071 (1994).
160. P.J. Barrie, S.W. Carr, D.L. Ou, A.C. Sullivan, 'Formation and Characterization of Novel Organodisilicate Glasses', *Chem. Mater.*, **7**(2), 265–270 (1995).
161. B. Boury, R.J.P. Corriu, R. Nunez, 'Hybrid Xerogels from Dendrimers and Arborols', *Chem. Mater.*, **10**(7), 1795–1804 (1998).
162. R. Corriu, 'A New Trend in Metal-alkoxide Chemistry – the Elaboration of Monophasic Organic–Inorganic Hybrid Materials', *Polyhedron*, **17**(5–6), 925–934 (1998).
163. P. Lacan, C. Guizard, L. Cot, 'Chemical and Rheological Investigations of the Sol–Gel Transition in Organically-modified Siloxanes', *J. Sol–Gel Sci. Technol.*, **4**(2), 151–162 (1995).
164. B. Alonso, J. Maquet, B. Viana, C. Sanchez, 'Hybrid Organic–Inorganic Polydimethylsiloxane–Vanadium–Oxo Materials Crosslinked at the Molecular Level', *New J. Chem.*, **22**(9), 935–939 (1998).
165. G. Cerveau, R.J.P. Corriu, C. Lepeytre, 'Synthesis of Monophasic Hybrid Organic–Inorganic Solids', *Chem. Mater.*, **9**(11), 2561–2566 (1997).
166. M.P.J. Peeters, W.J.J. Wakelkamp, A.P.M. Kentgens, 'A ^{29}Si Solid-state Magic Angle Spinning Nuclear Magnetic Resonance Study of TEOS-based Hybrid Materials', *J. Non-cryst. Solids*, **189**(1–2), 77–89 (1995).
167. Q. Liu, W. Shi, F. Babonneau, L.V. Interrante, 'Synthesis of Polycarbosilane/Siloxane Hybrid Polymers and their Pyrolytic Conversion to Silicon Oxycarbide Ceramics', *Chem. Mater.*, **9**(11), 2434–2441 (1997).
168. L. Bois, J. Maquet, F. Babonneau, H. Mutin, D. Bahloul, 'Structural Characterization of Sol–Gel Derived Oxycarbide Glasses. 1. Study of the Pyrolysis Process', *Chem. Mater.*, **6**(6), 796–802 (1994).
169. L. Bois, J. Maquet, F. Babonneau, D. Bahloul, 'Structural Characterization of Sol–Gel Derived Oxycarbide Glasses. 2. Study of the Thermal Stability of the Silicon Oxycarbide Phase', *Chem. Mater.*, **7**(5), 975–981 (1995).
170. R. Kalfat, F. Babonneau, N. Gharbi, H. Zarrouk, '^{29}Si MAS NMR Investigation of the Pyrolysis Process of Cross-linked Polysiloxanes Prepared from Polymethylhydrosiloxane', *J. Mater. Chem.*, **6**(10), 1673–1678 (1996).
171. A. Nordmann, Y.B. Cheng, T.J. Bastow, A.J. Hill, 'Structural Characterization of Lithium Aluminosilicate Glass and Glass Ceramics Derived From Spodumene Mineral', *J. Phys. Condens. Matter*, **7**(16), 3115–3128 (1995).
172. A. Nordmann, Y.B. Cheng, M.E. Smith, 'Role of Nitrides in Oxynitride Glasses and Glass-ceramics – an NMR Investigation', *Chem. Mater.*, **8**(10), 2516–2522 (1996).
173. S.C. Kohn, C.M.B. Henderson, R. Dupree, 'Si–Al Ordering in Leucite Group Minerals and Ion-exchanged Analogs – an MAS NMR Study', *Am. Mineral.*, **82**(11–12), 1133–1140 (1997).
174. J.M. Miller, L.J. Lakshmi, 'Spectroscopic Characterization of Sol–Gel-derived Mixed Oxides', *J. Phys. Chem. B*, **102**(34), 6465–6470 (1998).
175. P.F. McMillan, R.K. Sato, B.T. Poe, 'Structural Characterization of Si–Al–O–N Glasses', *J. Non-cryst. Solids*, **224**(3), 267–276 (1998).
176. P.F. McMillan, A. Grzechnik, H. Chotalla, 'Structural Characterization of SiO_2–$CsAlO_2$ and SiO_2–$RbAlO_2$ Glasses', *J. Non-cryst. Solids*, **226**(3), 239–248 (1998).
177. W. Storek, R. Mueller, G. Kunath-Fandrei, '^{27}Al and ^{29}Si MAS NMR Investigations of Cordierite Glass, μ- and α-Cordierite', *Solid State Nucl. Magn. Reson.*, **9**(2), 227–239 (1997).

178. T. Ban, Y. Ohta, Y. Takahashi, 'Low-temperature Crystallization of Forsterite and Orthoenstatite', *J. Am. Ceram. Soc.*, **82**(1), 22–26 (1999).
179. J.B.D. Delacaillerie, M. Kermarec, O. Clause, '^{29}Si NMR Observation of an Amorphous Magnesium Silicate Formed During Impregnation of Silica with Mg(II) in Aqueous Solution', *J. Phys. Chem.*, **99**(47), 17 273–17 281 (1995).
180. K. Okada, H. Kawashima, S. Hayashi, M. Sugai, K.J.D. Mackenzie, 'Crystallization Behavior and Cordierite Formation in Rapidly Quenched $MgAl_2O_4$–SiO_2 Glasses of Various Chemical Compositions', *J. Mater. Res.*, **13**(5), 1351–1357 (1998).
181. S.A. Khainakov, A.I. Bortun, L.N. Bortun, A. Clearfield, C. Trobajo, J.R. Garcia, 'Hydrothermal Synthesis and Characterization of Alkali Metal Titanium Silicates', *J. Mater. Chem.*, **9**(1), 269–272 (1999).
182. M. Feike, K. Meisegresch, Q. Chen, G.H. Frischat, 'NMR and FTIR Characterization of Sol–Gel Derived Ternary Oxide Glasses in the System BaO–TiO_2–SiO_2', *Z. Naturforsch., Teil A*, **50**(9), 837–844 (1995).
183. Z. Lin, J. Rocha, P. Ferreira, A. Thursfield, J.R. Agger, M.W. Anderson, 'Synthesis and Structural Characterization of Microporous Framework Zirconium Silicates', *J. Phys. Chem. B*, **103**(6), 957–963 (1999).
184. S. Dire, R. Campostrini, R. Ceccato, 'Pyrolysis Chemistry of Sol–Gel-derived Poly(dimethylsiloxane)–Zirconia Nanocomposites – Influence of Zirconium on Polymer-to-ceramic Conversion', *Chem. Mater.*, **10**(1), 268–278 (1998).
185. M. Jansen, E. Guenther, 'Oxide Gels and Ceramics Prepared by a Nonhydrolytic Sol–Gel Process', *Chem. Mater.*, **7**(11), 2110–2114 (1995).
186. M. Andrianainarivelo, R. Corriu, D. Leclercq, P.H. Mutin, A. Vioux, 'Mixed Oxides SiO_2–ZrO_2 and SiO_2–TiO_2 by a Non-hydrolytic Sol–Gel Route', *J. Mater. Chem.*, **6**(10), 1665–1671 (1996).
187. K.S. Chee, Y.B. Cheng, M.E. Smith, 'NMR Investigation of the Structure of Aluminum-containing Nitrogen Melilite (M′(ss))', *Chem. Mater.*, **7**(5), 982–988 (1995).
188. Z.C. Cao, B.I. Lee, W.D. Samuels, L.Q. Wang, G.J. Exarhos, 'Sol–Gel Synthesis of Phosphate Ceramic Composites II', *J. Mater. Res.*, **13**(6), 1553–1559 (1998).
189. A. Filippidis, A. Godelitsas, D. Charistos, C. Tsipis, 'Natrolite/Thomsonite Intergrowths from Basaltic Rocks of Deccan Traps (Western India)', *Neues Jahrb. Mineral. Monatsh.*, **4**, 161–170 (1996).
190. H.J. Jakobsen, N.C. Nielsen, H. Lindgreen, 'Sequences of Charged Sheets in Rectorite', *Am. Mineral.*, **80**(3–4), 247–252 (1995).
191. Y. Kim, R.J. Kirkpatrick, 'Application of MAS NMR Spectroscopy to Poorly Crystalline Minerals – Viseite', *Mineral. Mag.*, **60**(403), 957–962 (1996).
192. D.W. Fasshauer, N.D. Chatterjee, B. Marler, 'Synthesis, Structure, Thermodynamic Properties, and Stability Relations of K-Cymrite', *Phys. Chem. Miner.*, **24**(6), 455–462 (1997).
193. K.J.D. Mackenzie, R.H. Meinhold, I.W.M. Brown, G.V. White, 'The Formation of Mullite from Kaolinite Under Various Reaction Atmospheres', *J. Eur. Ceram. Soc.*, **16**(2), 115–119 (1996).
194. S. Chandrasekhar, 'Influence of Metakaolinization Temperature on the Formation of Zeolite 4A from Kaolin', *Clay Miner.*, **31**(2), 253–261 (1996).
195. J. Cuadros, J. Linares, 'Experimental Kinetic Study of the Smectite-to-illite Transformation', *Geochim. Cosmochim. Acta*, **60**(3), 439–453 (1996).
196. W.E. Dubbin, T.B. Goh, 'Solid-state ^{27}Al and ^{29}Si NMR Analysis of Hydroxy-Cr and -Al Interlayered Montmorillonite', *Clay Miner.*, **32**(3), 485–492 (1997).
197. W.P. Gates, J.W. Stucki, R.J. Kirkpatrick, 'Structural Properties of Reduced Upton Montmorillonite', *Phys. Chem. Miner.*, **23**(8), 535–541 (1996).
198. J.T. Kloprogge, J. Breukelaar, A.E. Wilson, J.W. Geus, J.B.H. Jansen, 'Solid-state Nuclear Magnetic Resonance Spectroscopy on Synthetic Ammonium/Aluminum-saponites', *Clays Clay Miner.*, **42**(4), 416–420 (1994).
199. K. Okada, A. Shimai, T. Takei, S. Hayashi, A. Yasumori, K.J.D. Mackenzie, 'Preparation of Microporous Silica from Metakaolinite by Selective Leaching Method', *Microporous Mesoporous Mater.*, **21**(4–6), 289–296 (1998).
200. G.E. Roch, M.E. Smith, S.R. Drachman, 'Solid State NMR Characterization of the Thermal Transformation of an Illite-rich Clay', *Clays Clay Miner.*, **46**(6), 694–704 (1998).
201. I. Tkac, P. Komadel, D. Muller, 'Acid-treated Montmorillonites – a Study by ^{29}Si and ^{27}Al MAS NMR', *Clay Miner.*, **29**(1), 11–19 (1994).
202. S. Maekawa, K. Okude, T. Ohishi, 'Synthesis, Properties, and Molecular Structure Analysis of SiO_2 Thin Films Prepared by Sol–Gel Method', *Electron. Commun. Jpn. Part 2: Electron. (Engl. Transl. Denshi Tsushin Gakkai Ronbunshi)*, **77**(5), 86–92 (1994).
203. M. Mayrand, J.E. Quinson, V. Roisne, H. Guyon, 'Development of Heteropolysiloxane Coatings on Electrogalvanized Steel', *J. Sol–Gel Sci. Technol.*, **13**(1–3), 779–782 (1998).
204. M. Mayrand, J.F. Quinson, A. Roche, V. Roisne, H. Guyon, 'Heteropolysiloxane Coatings on Electrogalvanized Steel – Elaboration and Characterization', *J. Sol–Gel Sci. Technol.*, **12**(1), 49–57 (1998).
205. P. Hartmann, J. Vogel, B. Schnabel, 'NMR Study of Phosphate Glasses and Glass Ceramic Structures', *J. Non-cryst. Solids*, **176**(2–3), 157–163 (1994).
206. L. Kerns, M.C. Weinberg, S. Myers, R. Assink, 'Al Coordination in Sol–Gel and Conventional Calcium Aluminate Glasses', *J. Non-cryst. Solids*, **234**, 86–92 (1998).
207. M.G.W. Lockyer, D. Holland, A.P. Howes, R. Dupree, 'MAS NMR Study of the Structure of Some PbO–

$Al_2O_3-P_2O_5$ Glasses', *Solid State Nucl. Magn. Reson.*, **5**(1), 23–34 (1995).

208. M.A. Harmer, A.J. Vega, 'Nuclear Magnetic Resonance Study of a High-surface-area Aluminum Phosphate Glass and its Thermally Reversible Sol–Gel Precursor', *Solid State Nucl. Magn. Reson.*, **5**(1), 35–49 (1995).
209. R. Mezei, T. Karancsi, J. Rohonczy, K. Sinko, 'Chemical Processes Involved in the Sol–Gel Preparation of an Aluminum Oxohydroxide Gel from Aluminum Nitrate in Organic Medium', *J. Mater. Chem.*, **8**(9), 2095–2101 (1998).
210. S. Acosta, R.J.P. Corriu, D. Leclercq, P. Lefevre, P.H. Mutin, A. Vioux, 'Preparation of Alumina Gels by a Non-hydrolytic Sol–Gel Processing Method', *J. Non-cryst. Solids*, **170**(3), 234–242 (1994).
211. D. Coster, J.J. Fripiat, 'Memory Effects in Gel Solid Transformations – Coordinately Unsaturated Al Sites in Nanosized Aluminas', *Chem. Mater.*, **5**(9), 1204–1210 (1993).
212. U. Janosovits, G. Ziegler, U. Scharf, A. Wokaun, 'Structural Characterization of Intermediate Species During Synthesis of Al_2O_3-aerogels', *J. Non-cryst. Solids*, **210**(1), 1–13 (1997).
213. S. Rezgui, B.C. Gates, 'Sol Gel Synthesis of Alumina in the Presence of Acetic Acid – Distinguishing Gels and Gelatinous Precipitates by NMR Spectroscopy', *Chem. Mater.*, **6**(12), 2386–2389 (1994).
214. S. Rezgui, B.C. Gates, S.L. Burkett, M.E. Davis, 'Chemistry of Sol Gel Synthesis of Aluminum Oxides with In Situ Water Formation – Control of the Morphology and Texture', *Chem. Mater.*, **6**(12), 2390–2397 (1994).
215. F. Touati, N. Gharbi, H. Zarrouk, 'Synthesis of New Hybrid Organic–Inorganic Alumina Gels by the Sol–Gel Method', *J. Sol–Gel Sci. Technol.*, **8**(1–3), 595–598 (1997).
216. L. Laby, L.C. Klein, A. Turniansky, D. Avnir, 'Synthesis of Alumina and Silica-containing Alumina Xerogel Hosts', *J. Sol–Gel Sci. Technol.*, **10**(2), 177–184 (1997).
217. M. May, M. Asomoza, T. Lopez, R. Gomez, 'Precursor Aluminum Effect in the Synthesis of Sol–Gel Si–Al Catalysts – FTIR and NMR Characterization', *Chem. Mater.*, **9**(11), 2395–2399 (1997).
218. R. Iwamoto, C. Fernandez, J.P. Amoureux, J. Grimblot, 'Characterization of Mo–P–Alumina Sol–Gel Catalysts By Solid-state ^{31}P and ^{27}Al Magic-angle Spinning Nuclear Magnetic Resonance and Two-dimensional ^{27}Al Multiple-quantum Magic-angle Spinning Nuclear Magnetic Resonance', *J. Phys. Chem. B*, **102**(22), 4342–4349 (1998).
219. G. Pachecomalagon, A. Garciaborquez, D. Coster, A. Sklyarov, S. Petit, J.J. Fripiat, '$TiO_2-Al_2O_3$ Nanocomposites', *J. Mater. Res.*, **10**(5), 1264–1269 (1995).
220. C. Gerardin, S. Sundaresan, J. Benziger, A. Navrotsky, 'Structural Investigation and Energetics of Mullite Formation from Sol–Gel Precursors', *Chem. Mater.*, **6**(2), 160–170 (1994).
221. G. Kunathfandrei, P. Rehak, S. Steuernagel, H. Schneider, C. Jager, 'Quantitative Structural Analysis of Mullite by ^{27}Al Nuclear Magnetic Resonance Satellite Transition Spectroscopy', *Solid State Nucl. Magn. Reson.*, **3**(5), 241–248 (1994).
222. P.R. Bodart, J. Parmentier, R.K. Harris, D.P. Thompson, 'Aluminum Environments in Mullite and an Amorphous Sol–Gel Precursor Examined by ^{27}Al Triple-quantum MAS NMR', *J. Phys. Chem. Solids*, **60**(2), 223–228 (1999).
223. J.Y. Kim, M.A. Sriram, P.H. McMichael, P.N. Kumta, B.L. Phillips, S.H. Risbud, 'New Molecular Precursors from the Reaction of Hydrazine and Aluminum Alkoxide for the Synthesis of Powders in the Al–O–N System', *J. Phys. Chem. B*, **101**(24), 4689–4696 (1997).
224. J.J. Fitzgerald, S.D. Kohl, G. Piedra, S.F. Dec, G.E. Maciel, 'Observation of 4-coordinate Aluminum Oxynitride (AlO_4-XNx) Environments in Alon Solids by MAS ^{27}Al NMR at 14-T', *Chem. Mater.*, **6**(11), 1915–1917 (1994).
225. K.S. Chee, Y.B. Cheng, M.E. Smith, 'The Solubility of Aluminum in Rare Earth Nitrogen Melilite Phases', *J. Eur. Ceram. Soc.*, **15**(12), 1213–1220 (1995).
226. J.S. Hartman, E.R. Vance, 'A ^{27}Al MAS NMR Study of Synroc Crystallization from Alkoxide Precursors', *J. Mater. Res.*, **9**(7), 1714–1720 (1994).
227. S. Kohn, M. Jansen, 'Preparation of Whisker Penetrated Ceramics by Internal Precipitation of Hollandite Within a Tielite Matrix', *J. Mater. Chem.*, **8**(7), 1593–1599 (1998).
228. S. Kohn, M. Jansen, 'Precipitation of Hollandite Whiskers within a Tielite Matrix via the Sol–Gel Process', *Ber. BunsenGes. Phys. Chem.*, **100**(9), 1450–1452 (1996).
229. J. Skibsted, H.J. Jakobsen, C. Hall, 'Quantitative Aspects of ^{27}Al MAS NMR of Calcium Aluminoferrites', *Adv. Cement Based Mater.*, **7**(2), 57–59 (1998).
230. J. Skibsted, H.J. Jakobsen, C. Hall, 'Direct Observation of Aluminum Guest Ions in the Silicate Phases of Cement Minerals by ^{27}Al MAS NMR Spectroscopy', *J. Chem. Soc., Faraday Trans.*, **90**(14), 2095–2098 (1994).
231. P. Ildefonse, R.J. Kirkpatrick, B. Montez, G. Calas, A.M. Flank, P. Lagarde, '^{27}Al MAS NMR and Aluminum X-ray Absorption Near Edge Structure Study of Imogolite and Allophanes', *Clays Clay Miner.*, **42**(3), 276–287 (1994).
232. P.J. Dirken, A.P.M. Kentgens, G.H. Nachtegaal, A.M. J. Vandereerden, J.B.H. Jansen, 'Solid-state MAS NMR Study of Pentameric Aluminosilicate Groups with 180° Intertetrahedral Al–O–Si Angles in Zunyite and Harkerite', *Am. Mineral.*, **80**(1–2), 39–45 (1995).
233. P.A. Schroeder, R.J. Pruett, 'Fe Ordering in Kaolinite – Insights from ^{29}Si and ^{27}Al MAS NMR Spectroscopy', *Am. Mineral.*, **81**(1–2), 26–38 (1996).

234. K. Hackbarth, M. Fechtelkord, J.C. Buhl, 'Crystallization Kinetics of $Na_{7.4}(AlSiO_4)_6(Co_3)_{0.7} \cdot 4H_2O$, an Intermediate Phase between Cancrinite and Sodalite, Grown under Low Temperature Hydrothermal Conditions', *React. Kinet. Catal. Lett.*, **65**(1), 33–39 (1998).
235. K.B. Brandt, R.A. Kydd, 'Characterization of Synthetic Microporous Pillared Beidellites of High Thermal Stability', *Chem. Mater.*, **9**(2), 567–572 (1997).
236. C.N. Rhodes, D.R. Brown, 'Autotransformation and Ageing of Acid-treated Montmorillonite Catalysts – a Solid-state ^{27}Al NMR Study', *J. Chem. Soc., Faraday Trans.*, **91**(6), 1031–1035 (1995).
237. R.H. Newman, C.W. Childs, G.J. Churchman, 'Aluminum Coordination and Structural Disorder in Halloysite and Kaolinite by ^{27}Al NMR Spectroscopy', *Clay Miner.*, **29**(3), 305–312 (1994).
238. P.J. Sanchezsoto, I. Sobrados, J. Sanz, J.L. Perezrodriguez, '^{29}Si and ^{27}Al Magic-angle Spinning Nuclear Magnetic Resonance Study of the Thermal Transformations of Pyrophyllite', *J. Am. Ceram. Soc.*, **76**(12), 3024–3028 (1993).
239. S. Carlino, M.J. Hudson, 'Thermal Intercalation of Layered Double Hydroxides – Capric Acid into an Mg–Al LDH', *J. Mater. Chem.*, **5**(9), 1433–1442 (1995).
240. G.A. Pozarnsky, A.V. McCormick, 'Multinuclear NMR Study of Aluminosilicate Sol–Gel Synthesis Using the Prehydrolysis Method', *J. Non-cryst. Solids*, **190**(3), 212–225 (1995).
241. F. Ali, A.V. Chadwick, M.E. Smith, 'EXAFS Analysis of the Structural Evolution of Gel-formed La_2O_3', *J. Mater. Chem.*, **7**(2), 285–291 (1997).
242. J. Blanchard, C. Bonhomme, J. Maquet, C. Sanchez, 'Characterization of Sol–Gel Derived Titanium Oxopolymers – First Evidence of Ti–OH Groups Through ^{1}H–^{17}O CP NMR Experiments', *J. Mater. Chem.*, **8**(4), 985–989 (1998).
243. J.K. Walters, J.S. Rigden, P.J. Dirken, M.E. Smith, W.S. Howells, R.J. Newport, 'An Atomic-scale Study of the Role of Titanium in TiO_2–SiO_2 Sol–Gel Materials', *Chem. Phys. Lett.*, **264**(5), 539–544 (1997).
244. B. Lebeau, J. Maquet, C. Sanchez, F. Beaume, F. Laupretre, 'Structural and Dynamical Studies of Hybrid Siloxane–Silica Materials', *J. Mater. Chem.*, **7**(6), 989–995 (1997).
245. A.V. Chadwick, I.J.F. Poplett, D.T.S. Maitland, M.E. Smith, 'Oxygen Speciation in Nanophase MgO From Solid-state ^{17}O NMR', *Chem. Mater.*, **10**(3), 864–870 (1998).
246. T.J. Bastow, P.J. Dirken, M.E. Smith, H.J. Whitfield, 'Factors Controlling the ^{17}O NMR Chemical Shift in Ionic Mixed Metal Oxides', *J. Phys. Chem.*, **100**(47), 18 539–18 545 (1996).
247. I. Abrahams, G.E. Hawkes, J. Knowles, 'Phosphorus Speciation in Sodium–Calcium–Phosphate Ceramics', *J. Chem. Soc., Dalton Trans.*, **9**, 1483–1484 (1997).
248. L.Q. Wang, W.D. Samuels, G.J. Exarhos, B.I. Lee, Z.C. Cao, '^{31}P and ^{29}Si NMR Study of Sol–Gel-synthesized Phosphate Ceramics', *J. Mater. Chem.*, **8**(1), 165–169 (1998).
249. H.L. Ren, Y. Yue, C.H. Ye, L.P. Guo, J.H. Lei, 'NMR Study of Crystallization in $MgO–CaO–SiO_2–P_2O_5$ Glass-ceramics', *Chem. Phys. Lett.*, **292**(3), 317–322 (1998).
250. S. Arsalane, M. Ziyad, 'Structural and Spectroscopic Properties of $M_1Th_2(PO_4)_3$ ($M = Cu^+, Ag^+, Na^+, K^+$)', *Mater. Res. Bull.*, **31**(12), 1567–1572 (1996).
251. P. Hartmann, C. Jana, J. Vogel, C. Jager, '^{31}P MAS and 2D Exchange NMR of Crystalline Silicon Phosphates', *Chem. Phys. Lett.*, **258**(1–2), 107–112 (1996).
252. S. Licoccia, M.L. Di Vona, E. Traversa, L. Montanaro, 'NMR Study of Sol–Gel Processed NASICON', *J. Eur. Ceram. Soc.*, **19**(6–7), 925–929 (1999).
253. N. Gasmi, N. Gharbi, H. Zarrouk, P. Barboux, R. Morineau, J. Livage, 'Comparison of Different Synthesis Methods for Nasicon Ceramics', *J. Sol–Gel Sci. Technol.*, **4**(3), 231–237 (1995).
254. J.C. Schrotter, A. Cardenas, M. Smaihi, N. Hovnanian, 'Silicon and Phosphorus Alkoxide Mixture – Sol–Gel Study by Spectroscopic Techniques', *J. Sol–Gel Sci. Technol.*, **4**(3), 195–204 (1995).
255. E.C.D. Lima, J.M.M. Neto, F.Y. Fujiwara, F. Galembeck, 'Aluminum Polyphosphate Thermoreversible Gels – a Study by ^{31}P and ^{27}Al NMR Spectroscopy', *J. Colloid Interface Sci.*, **176**(2), 388–396 (1995).
256. S.E. Dann, M.T. Weller, 'Correlations Between ^{9}Be Magic-angle Spinning Nuclear Magnetic Resonance Spectra and the Geometry of Beryllium Containing Framework Structures', *Solid State Nucl. Magn. Reson.*, **10**(1–2), 89–94 (1997).
257. A.J. Woo, D.Y. Han, K.H. Yang, '^{11}B Nutation NMR Study of Powdered Borosilicates', *Bull. Korean Chem. Soc.*, **19**(5), 519–524 (1998).
258. P. Mustarelli, E. Quartarone, F. Benevelli, 'A ^{11}B and ^{7}Li MAS/NMR Study of Sol–Gel Lithium Triborate Glass Subjected to Thermal Densification', *Mater. Res. Bull.*, **32**(6), 679–687 (1997).
259. A.M. Wootton, M.H. Lewis, 'Sol–Gel Synthesis and Structure of Borosilicon Oxycarbide Glasses', *J. Sol–Gel Sci. Technol.*, **13**(1–3), 1001–1004 (1998).
260. A.M. Wootton, M. Rappensberger, M.H. Lewis, S. Kitchin, A.P. Howes, R. Dupree, 'Structural Properties of Multi-component Silicon Oxycarbide Glasses Derived from Metal Alkoxide Precursors', *J. Non-cryst. Solids*, **204**(3), 217–227 (1996).
261. T. Sei, K. Eto, T. Tsuchiya, 'The Role of Boron in Low-temperature Synthesis of Indialite (α-$Mg_2Al_4Si_5O_{18}$) by Sol–Gel Process', *J. Mater. Sci.*, **32**(11), 3013–3019 (1997).
262. M.A. Aramendia, V. Borau, C. Jimenez, J.M. Marinas, A. Porras, F.J. Urbano, 'Synthesis and Characterization

of $MgO-B_2O_3$ Mixed Oxides Prepared by Coprecipitation; Selective Dehydrogenation of Propan-2-ol', *J. Mater. Chem.*, **9**(3), 819–825 (1999).
263. P.S. Fiske, J.F. Stebbins, I. Farnan, 'Bonding and Dynamical Phenomena in MgO – a High Temperature ^{17}O and ^{25}Mg NMR Study', *Phys. Chem. Miner.*, **20**(8), 587–593 (1994).
264. P.S. Fiske, J.F. Stebbins, 'The Structural of Mg in Silicate Liquids – a High-temperature ^{25}Mg, ^{23}Na, and ^{29}Si NMR Study', *Am. Mineral.*, **79**(9–10), 848–861 (1994).
265. K.J.D. Mackenzie, R.H. Meinhold, '^{25}Mg Nuclear Magnetic Resonance Spectroscopy of Minerals and Related Inorganics – a Survey Study', *Am. Mineral.*, **79**(3–4), 250–260 (1994).
266. K.J.D. Mackenzie, R.H. Meinhold, 'Thermal Decomposition of Brucite, $Mg(OH)_2$ – a ^{25}Mg MAS NMR Study', *Thermochim. Acta*, **230**, 339–343 (1993).
267. K.J.D. Mackenzie, R.H. Meinhold, B.L. Sherriff, Z. Xu, '^{27}Al and ^{25}Mg Solid-state Magic-angle Spinning Nuclear Magnetic Resonance Study of Hydrotalcite and its Thermal Decomposition Sequence', *J. Mater. Chem.*, **3**(12), 1263–1269 (1993).
268. K.J.D. Mackenzie, R.H. Meinhold, 'Role of Additives in the Sintering of Silicon Nitride – a ^{29}Si, ^{27}Al, ^{25}Mg and ^{89}Y MAS NMR and X-ray Diffraction Study', *J. Mater. Chem.*, **4**(10), 1595–1602 (1994).
269. K.J.D. Mackenzie, R.H. Meinhold, 'The Thermal Reactions of Talc Studied by ^{29}Si and ^{25}Mg MAS NMR', *Thermochim. Acta*, **244**, 195–203 (1994).
270. K.J.D. Mackenzie, R.H. Meinhold, 'The Thermal Reactions of Synthetic Hectorite Studied by ^{29}Si, ^{25}Mg, ^{7}Li Magic Angle Spinning Nuclear Magnetic Resonance', *Thermochim. Acta*, **232**(1), 85–94 (1994).
271. K.J.D. Mackenzie, R.H. Meinhold, 'Thermal Reactions of Chrysotile Revisited – a ^{29}S and ^{25}Mg MAS NMR Study', *Am. Mineral.*, **79**(1–2), 43–50 (1994).
272. T.J. Bastow, M.A. Gibson, C.T. Forwood, '^{47}Ti, ^{49}Ti NMR – Hyperfine Interactions in Oxides and Metals', *Solid State Nucl. Magn. Reson.*, **12**(4), 201–209 (1998).
273. O. Kanert, H. Schulz, J. Albers, 'Nuclear Magnetic Resonance Study of the Cubic-to-tetragonal Phase Transition in $BaTiO_3$', *Solid State Commun.*, **91**(6), 465–469 (1994).
274. A. Labouriau, W.L. Earl, 'Titanium Solid-state NMR in Anatase, Brookite and Rutile', *Chem. Phys. Lett.*, **270**(3–4), 278–284 (1997).
275. D.C.M. Dutoit, M. Schneider, P. Fabrizioli, A. Baiker, 'Vanadia–Silica Low-temperature Aerogels – Influence of Aging and Vanadia Loading on Structural and Chemical Properties', *Chem. Mater.*, **8**(3), 734–743 (1996).
276. K.T. Lee, G. Pozarnsky, O. Zarembowitch, A. McCormick, 'V-51 NMR of Homogeneous Multicomponent Vanadium Oxide Solutions', *Chem. Eng. J.*, **64**(2), 215–223 (1996).
277. U. Scharf, M. Schneider, A. Baiker, A. Wokaun, 'Vanadia–Titania Aerogels. 2. Spectroscopic Investigation of the Structural Properties', *J. Catal.*, **149**(2), 344–355 (1994).
278. S. Berger, J. Roos, D. Brinkmann, B.V.R. Chowdari, '$^{6,7}Li$ and ^{19}F NMR Studies of $X.LiF + (1-X)LiPO_3$ Glasses', *Solid State Ionics*, **86–8**(1), 475–479 (1996).
279. T. Yano, J. Mizuno, S. Shibata, M. Yamane, S. Inoue, Y. Onoda, 'NMR Study on Glass Structure of Chlorine-doped AlF_3-based Glasses with Various Glass-forming Abilities', *J. Non-cryst. Solids*, **213**, 345–352 (1997).
280. C. Boussardpledel, M. Lefloch, G. Fonteneau, J. Lucas, S. Sinbandhit, J. Shao, C.A. Angell, J. Emery, J.Y. Buzare, 'The Structure of a Boron Oxyfluoride Glass, an Inorganic Cross-linked Chain Polymer', *J. Non-cryst. Solids*, **209**(3), 247–256 (1997).
281. M. Lefloch, C. Boussardpledel, G. Fonteneau, J. Lucas, S. Sinbandhit, J. Emery, J.Y. Buzare, 'Gas Nucleation in a Boron Oxyfluoride Glass by ^{11}B and ^{19}F NMR', *J. Phys. Condens. Matter*, **10**(24), 5475–5483 (1998).
282. R.W.A. Franco, C.C. Tambelli, C.J. Magon, J.P. Donoso, Y. Messaddeq, S.J.L. Ribeiro, M. Poulain, 'Differential Scanning Calorimetry, X-ray Diffraction and ^{19}F Nuclear Magnetic Resonance Investigations of the Crystallization of InF_3-based Glasses', *J. Chem. Phys.*, **109**(6), 2432–2436 (1998).
283. J.M. Bobe, J.M. Reau, J. Senegas, M. Poulain, 'F^- Ion Conductivity and Diffusion Properties in ZrF_4-based Fluoride Glasses With Various NaF Concentrations $(0 \cdot X(NaF) \cdot 0.45)$', *Solid State Ionics*, **82**(1–2), 39–52 (1995).
284. J.M. Bobe, J.M. Reau, J. Senegas, M. Poulain, 'Ion Conductivity and Diffusion in ZrF_4-based Fluoride Glasses Containing LiF $(0 \cdot X(LiF) \cdot 0.60)$', *J. Non-cryst. Solids*, **209**(1–2), 122–136 (1997).
285. C. Jana, M. Braun, '^{19}F NMR Spectroscopy of Glass Ceramics Containing Fluorapatites', *Biomaterials*, **17**(21), 2065–2069 (1996).
286. A. Labouriau, Y.W. Kim, S. Chipera, D.L. Bish, W.L. Earl, 'A ^{19}F Nuclear Magnetic Resonance Study of Natural Clays', *Clays Clay Miner.*, **43**(6), 697–704 (1995).
287. A.M. George, S. Sen, J.F. Stebbins, 'Na-23 Chemical Shifts and Local Structures in Crystalline, Glassy and Molten Sodium Borates and Germanates', *Solid State Nucl. Magn. Reson.*, **10**(1), 9–17 (1995).
288. M.L. Di Vona, S. Licoccia, L. Montanaro, E. Traversa, 'Sol–Gel Synthesis of NASICON: 1D and 2D NMR Investigation', *Chem. Mater.*, **11**(5), 1336–1341 (1999).
289. Y. Kim, R.J. Kirkpatrick, 'High-temperature Multinuclear NMR Investigation of Analcime', *Am. Mineral.*, **83**(3–4), 339–347 (1998).
290. J.C. Buhl, J. Lons, 'Synthesis and Crystal Structure of Nitrate Enclathrated Sodalite Na-8', *J. Alloys Compd.*, **235**(1), 41–47 (1996).

291. Z. Lin, J. Rocha, P. Brandao, A. Ferreira, A.P. Esculcas, J.D.P. Dejesus, A. Philippou, M.W. Anderson, 'Synthesis and Structural Characterization of Microporous Umbite, Penkvilksite, and Other Titanosilicates', *J. Phys. Chem. B*, **101**(36), 7114–7120 (1997).
292. J.J. Liang, B.L. Sherriff, 'Lead Exchange into Zeolite and Clay Minerals – a ^{29}Si, ^{27}Al, ^{23}Na Solid-state NMR Study', *Geochim. Cosmochim. Acta*, **57**(16), 3885–3894 (1993).
293. Y. Kim, R.J. Kirkpatrick, '^{23}Na and ^{133}Cs NMR Study of Cation Adsorption on Mineral Surfaces – Local Environments, Dynamics, and Effects of Mixed Cations', *Geochim. Cosmochim. Acta*, **61**(24), 5199–5208 (1997).
294. K. Dahmouche, P.H. De Souza, T.J. Bonagamba, H. Paneppucci, P. Judeinstein, S.H. Pulcinelli, C.V. Santilli, 'Investigation of New Ion Conducting Ormolytes Silica–Polypropyleneglycol', *J. Sol–Gel Sci. Technol.*, **13**(1–3), 909–913 (1998).
295. V. Luca, T.L. Hanley, N.K. Roberts, R.F. Howe, 'NMR and X-ray Absorption Study of Lithium Incorporation in Micro and Nanocrystalline Anatase', *Chem. Mater.*, **11**(8), 2089–2102 (1999).
296. K.M. Nairn, M. Forsyth, M. Greville, D.R. Macfarlane, M.E. Smith, 'Solid State NMR Characterization of Lithium Conducting Ceramics', *Solid State Ionics*, **86–88**(2), 1397–1402 (1996).
297. M.D. Alba, R. Alvero, A.I. Becerro, M.A. Castro, J.M. Trillo, 'Chemical Behavior of Lithium Ions in Reexpanded Li-montmorillonites', *J. Phys. Chem. B*, **102**(12), 2207–2213 (1998).
298. R. Alvero, M.D. Alba, M.A. Castro, J.M. Trillo, 'Reversible Migration of Lithium in Montmorillonites', *J. Phys. Chem.*, **98**(32), 7848–7853 (1994).
299. Y. Kim, R.J. Kirkpatrick, R.T. Cygan, '^{133}Cs NMR Study of Cesium on the Surfaces of Kaolinite and Illite', *Geochim. Cosmochimi. Acta*, **60**(21), 4059–4074 (1996).
300. J. Sanz, P. Herrero, R. Rojas, J.M. Rojo, S. Rossignol, J.M. Reau, B. Tanguy, '^{109}Ag NMR Investigation of Tellurite Glasses Containing AgI or the $AgI_{0.75}TlI_{0.25}$ Eutectic Mixture', *Solid State Ionics*, **82**(3–4), 129–134 (1995).
301. M. Hosono, J. Kawamura, H. Itoigawa, N. Kuwata, T. Kamiyama, Y. Nakamura, 'Structure and Ionic Conductivity of Rapidly Quenched $AgI–Ag_2WO_4$ Superionic Conductor Glasses', *J. Non-cryst. Solids*, **244**(1), 81–88 (1999).
302. P. Mustarelli, C. Tomasi, E. Quartarone, A. Magistris, M. Cutroni, A. Mandanici, 'Structure and Cation Dynamics in the System $AgI–Ag_2MoO_4$ – a ^{109}Ag NMR Study', *Phys. Rev. B Condens. Matter*, **58**(14), 9054–9061 (1998).
303. F. Fayon, C. Landron, K. Sakurai, C. Bessada, D. Massiot, 'Pb^{2+} Environment in Lead Silicate Glasses Probed by Pb-L-III Edge XAFS and ^{207}Pb NMR', *J. Non-cryst. Solids*, **243**(1), 39–44 (1999).
304. F. Fayon, C. Bessada, A. Douy, D. Massiot, 'Chemical Bonding of Lead in Glasses Through Isotropic vs Anisotropic Correlation: PASS Shifted Echo', *J. Magn. Reson.*, **137**(1), 116–121 (1999).
305. L. Lan, A. Montenero, G. Gnappi, E. Dradi, '$PbTiO_3$ Sol–Gel Process Studied By ^{207}Pb-NMR, IR, DTA and XRD', *J. Mater. Sci.*, **30**(12), 3137–3141 (1995).
306. F. Fayon, I. Farnan, C. Bessada, J. Coutures, D. Massiot, J.P. Coutures, 'Empirical Correlations Between ^{207}Pb NMR Chemical Shifts and Structure in Solids', *J. Am. Chem. Soc.*, **119**(29), 6837–6843 (1997).
307. E.K.F. Bahlmann, R.K. Harris, K. Metcalfe, J.W. Rockliffe, E.G. Smith, 'Silicon-29 NMR Self-diffusion and Chemical-exchange Studies of Concentrated Sodium Silicate Solutions', *J. Chem. Soc., Faraday Trans.*, **93**(1), 93–98 (1997).
308. L.V. Ng, A.V. McCormick, 'Self-diffusion Coefficients of Sol–Gel Intermediates', *AIChE J.*, **40**(7), 1193–1202 (1994).
309. C.E. Schmelz, J.F. Stebbins, 'Gel Synthesis of an Albite ($NaAlSi_3O_8$) Glass – an NMR Analysis', *Geochim. Cosmochim. Acta*, **57**(16), 3949–3960 (1993).
310. L. Delattre, M. Roy, F. Babonneau, 'Design of Homogeneous Hybrid Materials Through a Careful Control of the Synthetic Procedure', *J. Sol–Gel Sci. Technol.*, **8**(1–3), 567–570 (1997).
311. L. Delattre, F. Babonneau, '^{17}O Solution NMR Characterization of the Preparation of Sol–Gel Derived SiO_2/TiO_2 and SiO_2/ZrO_2 Glasses', *Chem. Mater.*, **9**(11), 2385–2394 (1997).
312. T. Iwamoto, K. Morita, J.D. Mackenzie, 'Liquid State ^{29}Si NMR Study on the Sol Gel Reaction Mechanisms of Ormosils', *J. Non-cryst. Solids*, **159**(1–2), 65–72 (1993).
313. S. Goizet, J.C. Schrotter, M. Smaihi, A. Deratani, 'Sol–Gel Polyimide–Silica Composite Films – Correlations Between the Microstructure and the Synthesis Parameters', *New J. Chem.*, **21**(4), 461–468 (1997).
314. V. Gualandris, F. Babonneau, M.T. Janicke, B.F. Chmelka, 'NMR Studies on Hydrolysis and Condensation Reactions of Alkoxysilanes Containing Si–H Bonds', *J. Sol–Gel Sci. Technol.*, **13**(1–3), 75–80 (1998).
315. C.A. Fyfe, P.P. Aroca, 'Quantitative Kinetic Analysis by High-resolution ^{29}Si NMR Spectroscopy of the Initial Stages in the Sol–Gel Formation of Silica Gel from Tetraethoxysilane', *Chem. Mater.*, **7**(10), 1800–1806 (1995).
316. Y. Sugahara, T. Inoue, K. Kuroda, '^{29}Si NMR Study on Co-hydrolysis Processes in $Si(OEt)_4–RSi(OEt)_3–EtOH–Water–HCl$ Systems (R = Me, Ph) – Effect of R Groups', *J. Mater. Chem.*, **7**(1), 53–59 (1997).
317. D.J.M. Burkhard, G. Nachtegaal, 'Nucleation and Growth Behavior of Lithium Disilicate at Pressures to 0.5 GPa Investigated by NMR Spectroscopy', *J. Non-cryst. Solids*, **209**(3), 299–304 (1997).

318. I. Farnan, J.F. Stebbins, 'The Nature of the Glass Transition in a Silica-rich Oxide Melt', *Science*, **265**(5176), 1206–1209 (1994).
319. A.M. George, J.F. Stebbins, 'Dynamics of Na in Sodium Aluminosilicate Glasses and Liquids', *Phys. Chem. Miner.*, **23**(8), 526–534 (1996).
320. A.M. George, J.F. Stebbins, 'High-temperature ^{23}Na MAS NMR Data for Albite – Comparison to Chemical-shift Models', *Am. Mineral.*, **80**(9–10), 878–884 (1995).
321. A.M. George, S. Sen, J.F. Stebbins, '^{23}Na Chemical Shifts and Local Structure in Crystalline, Glassy, and Molten Sodium Borates and Germanates', *Solid State Nucl. Magn. Reson.*, **10**(1–2), 9–17 (1997).
322. A.M. George, P. Richet, J.F. Stebbins, 'Cation Dynamics and Premelting in Lithium Metasilicate (Li_2SiO_3) and Sodium Metasilicate (Na_2SiO_3): A High-temperature NMR Study', *Am. Mineral.*, **83**(11–12, Part 1), 1277–1284 (1998).
323. H. Maekawa, T. Nakao, S. Shimokawa, T. Yokokawa, 'Coordination of Sodium Ions in $NaAlO_2$–SiO_2 Melts – a High Temperature ^{23}Na NMR Study', *Phys. Chem. Miner.*, **24**(1), 53–65 (1997).
324. X.Y. Xue, J.F. Stebbins, '^{23}Na NMR Chemical Shifts and Local Na Coordination Environments in Silicate Crystals, Melts and Glasses', *Phys. Chem. Miner.*, **20**(5), 297–307 (1993).
325. B.T. Poe, P.F. McMillan, B. Cote, D. Massiot, J.P. Coutures, 'Structure and Dynamics in Calcium Aluminate Liquids – High-temperature ^{27}Al NMR and Raman Spectroscopy', *J. Am. Ceram. Soc.*, **77**(7), 1832–1838 (1994).
326. J.F. Stebbins, S. Sen, A.M. George, 'High Temperature Nuclear Magnetic Resonance Studies of Oxide Melts', *J. Non-cryst. Solids*, **193**, 298–305 (1995).
327. J.F. Stebbins, S. Sen, I. Farnan, 'Silicate Species Exchange, Viscosity, and Crystallization In a Low-silica Melt – In Situ High-temperature MAS NMR Spectroscopy', *Am. Mineral.*, **80**(7–8), 861–864 (1995).
328. J.F. Stebbins, 'Magnesium Site Exchange in Forsterite – a Direct Measurement By High-temperature ^{25}Mg NMR Spectroscopy', *Am. Mineral.*, **81**(11–12), 1315–1320 (1996).
329. J.F. Stebbins, S. Sen, 'Microscopic Dynamics and Viscous Flow in a Borosilicate Glass-forming Liquid', *J. Non-cryst. Solids*, **224**(1), 80–85 (1998).
330. J.F. Stebbins, S.E. Ellsworth, 'Temperature Effects on Structure and Dynamics in Borate and Borosilicate Liquids – High-resolution and High-temperature NMR Results', *J. Am. Ceram. Soc.*, **79**(9), 2247–2256 (1996).
331. J.F. Stebbins, 'Cation Sites in Mixed-alkali Oxide Glasses – Correlations of NMR Chemical Shift Data with Site Size and Bond Distance', *Solid State Ionics*, **112**(1–2), 137–141 (1998).
332. D.R. Spearing, J.F. Stebbins, 'Oxygen Displacement Through the Ferroelectric Phase Transition of Barium Titanate – High-temperature ^{17}O NMR', *J. Am. Ceram. Soc.*, **77**(12), 3263–3266 (1994).
333. H. Koller, G. Engelhardt, J. Felsche, 'Variable Temperature ^{1}H, ^{23}Na, and ^{29}Si MAS NMR Studies on Sodium Silicate Hydrates of Composition $Na_2O \cdot SiO_2 \cdot nH_2O$ ($n = 9, 6, 5$) – Local Structure in Crystals, Melts, Supercooled Melts, and Glasses', *Z. Anorg. Allg. Chem.*, **621**(2), 301–310 (1995).
334. S.J. Gaffey, 'H_2O and OH in Echinoid Calcite – a Spectroscopic Study', *Am. Mineral.*, **80**(9–10), 947–959 (1995).
335. K. Herzog, B. Thomas, D. Sprenger, C. Jager, 'REDOR NMR – Approaching Structural Elucidation of Hydrated Layers in Silicate Electrode Glasses', *J. Non-cryst. Solids*, **190**(3), 296–300 (1995).
336. K. Herzog, K. Scholz, B. Thomas, 'Structure of Hydrated Layers on Silicate Electrode Glasses', *Solid State Nucl. Magn. Reson.*, **3**(1), 1–15 (1994).
337. K. Herzog, K. Scholz, B. Thomas, 'Investigation of the Insertion of Water in Hydrated Layers of Silicate Electrode Glasses', *Fresenius' J. Anal. Chem.*, **349**(1–3), 223–224 (1994).
338. A. Burneau, J. Lepage, G. Maurice, 'Porous Silica–Water Interactions. 1. Structural and Dimensional Changes Induced by Water Adsorption', *J. Non-cryst. Solids*, **217**(1), 1–10 (1997).
339. A.B. Kudryavtsev, T.V. Kouznetsova, W. Linert, G. Hunter, 'A Study of the Hydration of Aluminate Minerals Based on the Measurements of the Mean and the Variance of the Proton Magnetic Resonance Relaxation Rate', *Chem. Phys.*, **215**(3), 419–427 (1997).
340. S.C. Kohn, 'Solubility of H_2O in Nominally Anhydrous Mantle Minerals Using ^{1}H MAS NMR', *Am. Mineral.*, **81**(11–12), 1523–1526 (1996).
341. H. Cho, G.R. Rossman, 'Single-crystal NMR Studies of Low-concentration Hydrous Species in Minerals – Grossular Garnet', *Am. Mineral.*, **78**(11–12), 1149–1164 (1993).
342. B. Charoy, P. Dedonato, O. Barres, C. Pintocoelho, 'Channel Occupancy in an Alkali-poor Beryl from Serra Branca (Goias, Brazil) – Spectroscopic Characterization', *Am. Mineral.*, **81**(3–4), 395–403 (1996).
343. S. Ishimaru, R. Ikeda, 'NMR Studies on the Dynamics of Intercalated Water in Li-saponite', *Z. Naturforsch., Teil A*, **52**(12), 863–866 (1997).
344. C. Mercier, L. Montagne, H. Sfihi, G. Palavit, J.C. Boivin, A.P. Legrand, 'Local Structure of Zinc Ultraphosphate Glasses Containing Large Amount of Hydroxyl Groups – ^{31}P and ^{1}H Solid State Nuclear Magnetic Resonance Investigation', *J. Non-cryst. Solids*, **224**(2), 163–172 (1998).
345. S. Dick, G. Grossmann, G. Ohms, T. Zeiske, 'Aluminum Phosphates with Non-centrosymmetric Layer- and Framework-structures of Topologically Related Motifs – 1 – $KAl_2(PO_4)_2(OH) \cdot 4H_2O$', *Z. Naturforsch., Teil B*, **52**(12), 1439–1446 (1997).

346. F. Brunet, T. Schaller, 'Protons in the Magnesium Phosphates Phosphoellenbergerite and Holtedahlite – an IR and NMR Study', *Am. Mineral.*, **81**(3–4), 385–394 (1996).
347. H. Zanni, R. Rassembertolo, S. Masse, L. Fernandez, P. Nieto, B. Bresson, 'A Spectroscopic NMR Investigation of the Calcium Silicate Hydrates Present in Cement and Concrete', *Magn. Reson. Imaging*, **14**(7–8), 827–831 (1996).
348. S. Thomas, K. Meisegresch, W. Mullerwarmuth, I. Odler, 'MAS NMR Studies of Partially Carbonated Portland Cement and Tricalcium Silicate Pastes', *J. Am. Ceram. Soc.*, **76**(8), 1998–2004 (1993).
349. S. Sen, A.M. George, J.F. Stebbins, 'Ionic Conduction and Mixed Cation Effect in Silicate Glasses and Liquids – ^{23}Na and ^{7}Li NMR Spin–Lattice Relaxation and a Multiple-barrier Model of Percolation', *J. Non-cryst. Solids*, **197**(1), 53–64 (1996).
350. M.J. Kang, M.S. Kim, S.K. Oh, S.J. Chung, K.S. Hong, M.S. Sim, '^{7}Li NMR Studies in Mixed Alkali Silicate Glasses', *J. Korean Phys. Soc.*, **26**(4), 361–368 (1993).
351. V. Blache, J. Forster, H. Jain, O. Kanert, R. Kuchler, K.L. Ngai, 'Ionic Motion in Lithium Silicophosphate Glasses by Nuclear Spin Relaxation and Electrical Conductivity', *Solid State Ionics*, **115**(Special Issue SI), 723–731 (1998).
352. B. Ruffle, S. Beaufils, J. Gallier, 'Low-frequency Motions in an Alkali Phosphate Glass Studied by ^{7}Li and ^{31}P NMR', *Chem. Phys.*, **195**(1–3), 339–350 (1995).
353. J. Dieckhofer, O. Kanert, R. Kuchler, A. Volmari, H. Jain, 'Composition Dependence of Low-frequency Excitations in Lithium Silicophosphate Glasses By Nuclear Magnetic Resonance and Electrical Conductivity', *Phys. Rev. B Condens. Matter*, **55**(22), 14 836–14 846 (1997).
354. R. Bertermann, W. Muller-Warmuth, C. Jansen, F. Hiltmann, B. Krebs, 'NMR Studies of the Lithium Dynamics in Two Thioborate Superionic Conductors: $Li_9B_{19}S_{33}$ and $Li_{4-2x}Sr_{2+x}B_{10}S_{19}$ (x approximate to 0.27)', *Solid State Ionics*, **117**(3–4), 245–255 (1999).
355. P. Vinatier, M. Menetrier, A. Levasseur, 'Aggregation of the Doping Salt in B_2S_3–Li_2S–LiI Glasses, Effect on the Dynamical Properties', *Solid State Ionics*, **116**(1–2), 35–45 (1999).
356. O. Bohnke, J. Emery, A. Veron, J.L. Fourquet, J.Y. Buzare, P. Florian, D. Massiot, 'A Distribution of Activation Energies for the Local and Long-range Ionic Motion is Consistent with the Disordered Structure of the Perovskite $Li_{3x}La_{2/3-x}TiO_3$', *Solid State Ionics*, **109**(1–2), 25–34 (1998).
357. T. Matsuo, M. Shibasaki, T. Katsumata, Y. Onoda, 'Nuclear Magnetic Resonance of ^{7}Li in $Li_2B_4O_7$ Single Crystal and Glass', *Jpn. J. Appl. Phys., Part 1*, **33**(7A), 3913–3917 (1994).
358. J.M. Reau, S. Rossignol, B. Tanguy, M.A. Paris, J.M. Rojo, J. Sanz, 'Li^+ Ion Mobility in TeO_2–$LiO_{0.5}$–LiX (X = F, Cl) Glasses Determined by ^{7}Li NMR and Impedance Spectroscopy', *Solid State Ionics*, **80**(3–4), 283–290 (1995).
359. N. Nakamura, 'Nuclear Quadrupole Interaction and Ionic Dynamics in Sodium Superionic Conductors', *Z. Naturforsch., Teil A*, **49**(1–2), 337–344 (1994).
360. J.F. Stebbins, Z. Xu, D. Vollath, 'Cation Exchange Rates and Mobility in Aluminum-doped Lithium Orthosilicate – High-resolution Lithium-6 NMR Results', *Solid State Ionics*, **78**(1–2), L1–L8 (1995).
361. S. Sen, J.F. Stebbins, 'Phase Separation, Clustering, and Fractal Characteristics in Glass – a Magic-angle-spinning NMR Spin–Lattice Relaxation Study', *Phys. Rev. B Condens. Matter*, **50**(2), 822–830 (1994).
362. K.L. Ngai, 'Further Tests of the Coupling Model of Ionic Motions in Glassy Fast Ionic Conductors – Conductivity Relaxation and ^{7}Li and ^{11}B NMR in $X.Li_2S + (1-X)B_2S_3$', *Solid State Ionics*, **105**(1–4), 225–230 (1998).
363. I. Svare, F. Borsa, D.R. Torgeson, S.W. Martin, 'Connection Between NMR Relaxation and Electrical Conductivity from Distributions of Activation Energies of Ionic Motion in Some Fast-ion Conductors', *J. Non-cryst. Solids*, **172**(2), 1300–1305 (1994).
364. K.H. Kim, D.R. Torgeson, F. Borsa, J.P. Cho, S.W. Martin, I. Svare, G. Majer, 'Evidence of Complex Ionic Motion in $X.Li_2S + (1-X)B_2S_3$ Glassy Fast Ionic Conductors From ^{7}Li and ^{11}B NMR and Ionic Conductivity Measurements', *J. Non-cryst. Solids*, **211**(1–2), 112–125 (1997).
365. K.H. Kim, D.R. Torgeson, F. Borsa, S.W. Martin, 'Two Separate Li^+ Ionic Motions Observed by ^{7}Li and ^{11}B NMR in $X.Li_2S + (1-X).B_2S_3$ Glassy Fast Ionic Conductors', *Solid State Ionics*, **90**(1–4), 29–38 (1996).
366. M. Grune, W. Mullerwarmuth, P. Zumhebel, B. Krebs, 'Complex Lithium Dynamics in the Novel Thioborate $Li_5B_7S_{13}$ Revealed by NMR Relaxation and Line Shape Studies', *Solid State Ionics*, **78**(3–4), 305–313 (1995).
367. M. Grune, W. Mullerwarmuth, 'The Importance of Paramagnetic Impurities to the Nuclear Magnetic Resonance Relaxation of Ion-conducting Glasses', *Solid State Nucl. Magn. Reson.*, **5**(1), 145–150 (1995).
368. J.M. Keartland, I.P. Goudemond, M.J.R. Hoch, H.B. Senin, P.J. Ford, G.A. Saunders, 'Magnetic Properties of Nd^{3+} Phosphate Glasses', *Physica B*, **194**(1), 397–398 (1994).
369. D.R. Spearing, J.F. Stebbins, I. Farnan, 'Diffusion and the Dynamics of Displacive Phase Transitions in Cryolite (Na_3AlF_6) and Chiolite ($Na_5Al_3F_{14}$) – Multi-nuclear NMR Studies', *Phys. Chem. Miner.*, **21**(6), 373–386 (1994).
370. M. Tansho, H. Wada, M. Ishii, Y. Onoda, 'Mobility of Silver Ions in Silver Ion Conductor Ag_7NbS_6 Studied by Ag and Nb NMR', *J. Phys. Chem. B*, **102**(26), 5047–5049 (1998).

371. C. Ramanathan, J.L. Ackerman, 'ADRF Differential Cross Polarization Spectroscopy of Synthetic Calcium Phosphates and Bone Mineral', *J. Magn. Reson.*, **127**(1), 26–35 (1997).
372. Y. Wu, J.L. Ackerman, D.A. Chesler, J. Li, R.M. Neer, J. Wang, M.J. Glimcher, 'Evaluation of Bone Mineral Density Using Three-dimensional Solid State Phosphorus-31 NMR Projection Imaging', *Calcified Tissue Int.*, **62**(6), 512–518 (1998).
373. S.D. Beyea, B.J. Balcom, T.W. Bremner, P.J. Prado, A.R. Cross, R.L. Armstrong, P.E. Grattan-Bellew, 'The Influence of Shrinkage-cracking on the Drying Behavior of White Portland Cement Using Single-point Imaging (SPI)', *Solid State Nucl. Magn. Reson.*, **13**(1–2), 93–100 (1998).
374. M. Bogdan, B.J. Balcom, T.W. Bremner, R.L. Armstrong, 'Single-point Imaging of Partially Dried, Hydrated White Portland Cement', *J. Magn. Reson., Ser. A*, **116**(2), 266–269 (1995).
375. J. Kaufmann, W. Studer, J. Link, K. Schenker, 'Study of Water Suction of Concrete with Magnetic Resonance Imaging Methods', *Mag. Concrete Res.*, **49**(180), 157–165 (1997).
376. T. Nunes, E.W. Randall, A.A. Samoilenko, P. Bodart, G. Feio, 'The Hardening of Portland Cement Studied by ^{1}H NMR Stray-field Imaging', *J. Phys. D Appl. Phys.*, **29**(3), 805–808 (1996).
377. P.J. Prado, B.J. Balcom, S.D. Beyea, T.W. Bremner, R.L. Armstrong, P.E. Grattanbellew, 'Concrete Freeze/Thaw as Studied by Magnetic Resonance Imaging', *Cement Concrete Res.*, **28**(2), 261–270 (1998).
378. P.J. Prado, B.J. Balcom, S.D. Beyea, T.W. Bremner, R.L. Armstrong, R. Pishe, P.E. Grattenbellew, 'Spatially Resolved Relaxometry and Pore Size Distribution by Single-point MRI Methods – Porous Media Calorimetry', *J. Phys. D Appl. Phys.*, **31**(16), 2040–2050 (1998).
379. C.D. Arvanitopoulos, J.L. Koenig, 'An NMR Imaging Study of the Interface of Epoxy Resin–Glass Fiber Reinforced Composites', *J. Adhes.*, **53**(1–2), 15 ff. (1995).
380. J. Timonen, L. Alvila, P. Hirva, T.T. Pakkanen, D. Gross, V. Lehmann, 'NMR Imaging of Aluminum Oxide Catalyst Spheres', *Appl. Catal. A*, **129**(1), 117–123 (1995).

Parameters, Calculation of Nuclear Magnetic Resonance

Cynthia J. Jameson
University of Illinois at Chicago, USA

1 Introduction **2**
1.1 Absolute Shielding Tensor and Nuclear Magnetic Resonance Chemical Shift 2
1.2 Indirect Spin–Spin Coupling Tensor 3
1.3 Electric Field Gradient Tensor and Nuclear Quadrupole Coupling Constant 3
1.4 Nuclear Magnetic Resonance Parameters in Gases, Liquids, and Solids 3

2 General Theoretical Methods **4**
2.1 Multiple Perturbation Theory 4
2.2 Gauge Origin Problem in Calculations of Chemical Shift 7
2.3 Difficulties of Describing Triplet States in Calculations of Spin–Spin Coupling 8
2.4 Ab Initio Methods 9
2.5 Density Functional Methods 11
2.6 Relativistic Calculations 13

3 Calculations of Nuclear Magnetic Resonance Chemical Shifts **14**
3.1 Comparison of Various Computational Methods Using the Same Set of Test Molecules 14
3.2 Comparison of Carbon Chemical Shift Tensor Components with Calculations 16
3.3 Other First Row Nuclei 17
3.4 Second Row Nuclei 18
3.5 Other Heavy Nuclei 18
3.6 Transition Metal Nuclei 19

4 Calculations of Spin–Spin Coupling Constants **19**
4.1 One-bond Coupling Constants 19
4.2 The Two-bond Coupling Constant 22
4.3 Coupling Over Three Bonds 22
4.4 Relativistic Effects 22

5 Calculations of Electric Field Gradients **22**
5.1 Electric Field Gradient Tensor Versus Electronic Structure in the Solid State 22
5.2 Calculations of Electric Field Gradients at Nuclei in Isolated Small Molecules 23
5.3 Simulations of Nuclear Quadrupole Coupling in Associated Liquids 23
5.4 Relation Between Chemical Shift and Electric Field Gradient Tensors 23

6 Influence of Intramolecular Geometry and Environment on Nuclear Magnetic Resonance Parameters **23**
6.1 Nuclear Magnetic Resonance Parameter Dependence on Local Geometry: Bond Lengths, Bond Angles, Torsion Angles 24
6.2 Intermolecular Effects 24

7 Future Developments **25**

Abbreviations and Acronyms **26**

Related Articles **26**

References **27**

The fundamental parameters that reproduce a nuclear magnetic resonance (NMR) spectrum in gases, liquids, and solids are the four tensor quantities: the NMR chemical shift of the nucleus, the indirect nuclear spin–spin coupling, the nuclear electric quadrupole coupling, and the direct dipolar coupling tensors. The first three are intimately related to the local electronic structure at the nucleus and the chemical bonds connecting the nuclei. On the other hand, the direct nuclear spin dipole–dipole interaction depends directly and entirely on the third power of the inverse of the direct through-space distance between two nuclei, whether bonded or otherwise. In gases and in liquids where free tumbling of the molecules bearing the nuclear spins leads to isotropic averaging of these quantities, only the isotropic average values, the average of three components along the principal axes of the chemical shift and the indirect spin–spin coupling tensors determine the observed frequencies in the NMR spectrum. In the solid state, restricted motion permits the tensors to manifest all the components, whether the sample is a polycrystalline powder, an amorphous solid, or a single crystal. Theoretical calculations of the NMR chemical shift, indirect spin–spin coupling, and nuclear quadrupole coupling parameters, using quantum mechanical methods, permit the prediction of NMR spectra and provide the physical basis for the relationship between the parameters and molecular electronic structure, which may include local electronic structure (electronic distribution in the immediate vicinity of the nucleus and neighboring bonds), local molecular geometry, bond connectivities, stereochemical structure, as well as subtle

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

effects of the chemical environment, such as contributions from remote parts of the molecule, tertiary and secondary structure, crystal packing, solvent effects, and isotopic substitution.

1 INTRODUCTION

The fundamental parameters that reproduce a NMR spectrum in gases, liquids, and solids are: the NMR chemical shift of the nucleus, the indirect nuclear spin–spin coupling, and the nuclear quadrupole coupling. All three quantities are tensors whose directional properties are intimately related to the local electronic structure at the nucleus. In gases and in liquids where free tumbling of the molecules bearing the nuclear spin leads to isotropic averaging of these quantities, only a single number determines the frequencies in the NMR spectrum: the isotropic average value, the average of three components along the principal axes of the tensor. In the solid state, restricted motion permits the tensors to manifest all the components, whether the sample is a polycrystalline powder, an amorphous solid, or a single crystal.

There is a fourth parameter which is just as important, the direct nuclear spin dipole–dipole interaction, which depends directly and entirely on the third power of the inverse of the distance between two nuclei, whether bonded or otherwise, and irrespective of the electronic structure. It is a very important parameter in the solid state because it depends on structure, and for protons in a large molecule in solution, it provides the nuclear Overhauser effect (NOE) at intensities which makes it possible to elicit the geometrical structure of the nonfloppy parts of the molecule. This direct dipole–dipole interaction parameter is not considered in the theoretical calculations described here because of the trivial mathematical relationship between the parameter and the direct through-space distance.

1.1 Absolute Shielding Tensor and Nuclear Magnetic Resonance Chemical Shift

The NMR spectrum provides the chemical shift δ relative to a chosen reference substance in a chosen medium. The definition of the chemical shift, usually expressed in ppm, is given in Equation (1)

$$\delta \equiv \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}} \tag{1}$$

where ν is the resonance frequency for the nucleus of interest in the sample and ν_{ref} is the resonance frequency for the reference. The resonance frequency is determined by a fundamental molecular electronic property called the nuclear magnetic shielding, $\boldsymbol{\sigma}$, which is defined by the Hamiltonian for the energy of a single nucleus N possessing a nuclear magnetic moment $\boldsymbol{\mu}_{\text{N}}$, in an external magnetic field $\mathbf{B}$, Equation (2) (where Z stands for Zeeman and CS stands for Chemical Shift)

$$\mathcal{H}_{\text{Z+CS}} = -\boldsymbol{\mu}_{\text{N}}(1 - \boldsymbol{\sigma})\mathbf{B} \tag{2}$$

The magnetic field experienced by the nucleus at its site is different from the applied magnetic field $\mathbf{B}$ because of the small field $\mathbf{B}_{\text{local}}$ arising from the circulations of the electrons induced by the external magnetic field, Equation (3)

$$(\mathbf{B}_{\text{local}})_\alpha = (1 - \boldsymbol{\sigma})_{\alpha\beta}\mathbf{B}_\beta, \qquad \alpha, \beta = x, y, z \tag{3}$$

Thus, Equation (4)

$$\mathcal{H}_{\text{CS}} = +\boldsymbol{\mu}_{\text{N}}\boldsymbol{\sigma}\mathbf{B} \tag{4}$$

The term ''magnetic shielding'' implies that the magnetic dipole of a nucleus at that site would be shielded from the full effect of the external field by the influence of the induced electronic motions. For free atoms $\boldsymbol{\sigma}$ is always positive because this circulation generates a shielding field which opposes the applied field. In a molecule the presence of other nuclei hinders this circulation to an extent that depends on the electronic distribution and may even lead to a negative $\boldsymbol{\sigma}$. Depending on the symmetry of the electronic distribution at the nuclear site, some of the components $\boldsymbol{\sigma}_{\alpha\beta}$ may be zero or identical. For example, for a linear molecule there are only two unique components, $\boldsymbol{\sigma}_{zz}$ and $\boldsymbol{\sigma}_{xx}$, where z is along the molecular axis; these components are designated as $\boldsymbol{\sigma}_{\parallel}$ and $\boldsymbol{\sigma}_{\perp}$, respectively.

Theoretical calculations of the nuclear magnetic shielding provide the entire shielding tensor $\boldsymbol{\sigma}$ on an absolute basis, i.e. with respect to a bare nucleus. The chemical shift δ expresses a difference in nuclear magnetic shielding, Equation (5)

$$\delta \equiv \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}} = \frac{\boldsymbol{\sigma}_{\text{ref}} - \boldsymbol{\sigma}}{1 - \boldsymbol{\sigma}_{\text{ref}}} \tag{5}$$

Usually, though not always, $\boldsymbol{\sigma}_{\text{ref}}$ can be neglected relative to 1.0, so sometimes it is sufficient to use Equation (6),

$$\delta \approx (\boldsymbol{\sigma}_{\text{ref}} - \boldsymbol{\sigma}) \tag{6}$$

A *negative chemical shift* means that the nucleus located at site A sees a more shielded (smaller) magnetic field than does the nucleus in the reference substance, so that the *applied field has to be made higher* in order to achieve resonance with the nuclear spin energy separation at site A.

We see from Equation (2) that the mathematical terms in the total energy of a molecule that are bilinear in the external homogeneous magnetic field $\mathbf{B}$ and the nuclear magnetic moment $\boldsymbol{\mu}$ determine the nuclear magnetic

shielding $\boldsymbol{\sigma}$ for a nucleus in a molecule. The theoretical calculation of NMR chemical shifts from first principles therefore consists of collecting all such bilinear terms in the energy of a molecule in the presence of both an external magnetic field and a nuclear magnetic moment located at the observed site to obtain the absolute shielding tensor quantities $\boldsymbol{\sigma}$. A separate calculation is required for the reference molecule. The NMR chemical shift tensor can then be calculated from differences between $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}_{\mathrm{ref}}$.

1.2 Indirect Spin–Spin Coupling Tensor

Usually, more than one nuclear spin is present in the observed molecule. The interaction of nuclear spins N and N′ is composed of a direct through-space dipolar coupling (coupling of the bare nuclear magnetic dipole moments) and an indirect interaction by way of the electrons. The Hamiltonian for this interaction energy is, Equation (7)

$$_{\mathbf{D+J}} = \boldsymbol{\mu}_{\mathrm{N}}(\mathbf{D}+\mathbf{J})\boldsymbol{\mu}_{\mathrm{N}'} \tag{7}$$

The direct dipolar coupling tensor **D** is symmetric with the principal components summing to zero (a traceless tensor), and depends entirely on the distance vector between N and N′. In an oriented system both **D** and **J** (the spin–spin coupling) contribute to the observed spectrum. In a rapidly tumbling molecule in solution, only the isotropic average of **J** survives ($\mathbf{J}_{\mathrm{iso}} = (1/3)[\mathbf{J}_{xx}+\mathbf{J}_{yy}+\mathbf{J}_{zz}]$); the anisotropic part averages to zero. A *positive* **J** results from an interaction which minimizes the energy when the two nuclear spins are *antiparallel*. Theoretical calculations of the **J** tensor from first principles consists of collecting all such bilinear terms in the energy of a molecule, as shown in Equation (7).

1.3 Electric Field Gradient Tensor and Nuclear Quadrupole Coupling Constant

All nuclei with spin $I > 1/2$ have an ellipsoidal distribution of charge and an electric quadrupole moment eQ, where e is the magnitude of the charge of an electron. Q is positive if the nucleus is prolate (cigar-like), negative if oblate (pancake-like). Q is an intrinsic property of the nucleus. Energy is minimized by appropriate alignment of an electric quadrupole in an electric field gradient. At a nuclear site in a molecule, there is an electric field gradient when there is an asymmetry in the charge distribution due to the electrons and other nuclei. This electric field gradient is represented by $e\mathbf{q}$. The energy of a nuclear quadrupole is quantized according to its orientation in the electric field gradient, even in the absence of an external magnetic field. The electrostatic energy of interaction between the electric quadrupole moment and the electric field gradient is expressed in terms of the nuclear quadrupole coupling constant ($e^2Qq_{zz}/\hbar$).

The magnetic dipole moment of a quadrupolar nucleus is along the axis of symmetry of the nuclear charge distribution. Thus, when a quadrupolar nucleus is placed in a magnetic field so that the nuclear magnetic dipole tends to align with the external magnetic field, the interaction of the electric quadrupole with the internal electric field gradient at the nuclear site in the molecule affects the nuclear magnetic energy levels. The tensor coupling between the nuclear spin and the electric field gradient $e\mathbf{q}$ at the nucleus is described by the Hamiltonian, Equation (8)

$$_{Q} = \mathbf{I}_{\mathrm{N}}\frac{eQ}{2I(2I-1)}e\mathbf{q}\mathbf{I}_{\mathrm{N}'} \tag{8}$$

Like **D**, the electric field gradient tensor is traceless: the isotropic average of energy terms involving **q** is zero. Thus, in liquids or gases the positions of the lines in the NMR spectrum are not affected by the nuclear quadrupole coupling. In solids the nuclear quadrupole coupling can dominate the NMR spectrum and measurements of the nuclear quadrupole coupling tensor in single crystals or powders provide the electric field gradient tensors.

1.4 Nuclear Magnetic Resonance Parameters in Gases, Liquids, and Solids

In gases and liquids, isotropic averaging caused by the rapid tumbling of molecules leads to observations of only the isotropic part of $\boldsymbol{\sigma}$ and **J**, which are given by one third the sum of the principal components of these tensors. At the same time, **D** and **q** being traceless means that this sum is zero. Thus, in liquids or gases the positions of the lines in the NMR spectrum are not affected by either the direct dipolar coupling or the nuclear quadrupole coupling. To a good approximation, neither the chemical shift nor the spin–spin coupling **J** is dependent on the strength of the magnetic field. Actually, one has to be quite specific in defining the environment of the nucleus in both sample and reference because the NMR chemical shift is very sensitive to these. For example, the chemical shift of a ^{13}C nucleus in molecule A relative to the usual reference, tetramethylsilane (TMS) is $\delta_{\mathrm{A}} = \sigma(^{13}\mathrm{C}, \text{in TMS, in } CDCl_3 \text{ solution}, x_{\mathrm{A}}, x_{\mathrm{TMS}}, x_{CDCl_3}, 300\,\mathrm{K}) - \sigma(^{13}\mathrm{C}, \text{in A, in } CDCl_3 \text{ solution}, x_{\mathrm{A}}, x_{\mathrm{TMS}}, x_{CDCl_3}, 300\,\mathrm{K})$.

It is important to specify completely all the variables (e.g. mole fractions x_{A}, etc.) that determine the observed chemical shift because the nuclear magnetic shielding is so sensitive to factors of molecular structure and environment. There is an intrinsic mass and temperature dependence of the chemical shift, the spin–spin coupling, and the nuclear quadrupole coupling because all three are functions of the electron distribution, which in

turn is a function of the nuclear positions. As the internuclear separations are weighted according to the vibrational functions, the thermal average values of these NMR parameters are dependent on the vibrational and rotational state populations. Furthermore, all three NMR parameters are dependent on the medium since each one is affected by the electronic environment, and the electron distribution is affected by intermolecular interactions and external electric fields. For protons the medium effects are generally small, whereas they can be quite large for other nuclei.

In oriented molecules, such as in liquid crystal solutions, polycrystalline powders, single crystals, or amorphous powders, the tensor nature of the three NMR parameters manifest themselves in the spectrum. In principle, one can measure both the anisotropy and asymmetry of the **J** tensor in rigid solids. However, the anisotropy of **J** transforms similarly to the direct dipolar coupling, thus the two interactions cannot be easily separated via experiment. The anisotropy in **J** is predicted to become more important for coupling constants involving heavier nuclei, whereas **D** depends only on the internuclear distance. Rapid magic-angle spinning (MAS) can be used to obtain high-resolution spectra of solids by removing the effects of the anisotropic terms, which in general have a $P_2(\cos\theta)$ dependence. The angle for which $(3\cos^2\theta - 1)$ equals zero is the magic-angle 54.74°. The terms that give rise to the NMR spectrum of quadrupolar nuclei include in addition, a $P_4(\cos\theta)$ dependence. Various techniques have been used to determine these three NMR tensors individually by experiment, including the orientations of their principal axis systems. In solids, the spinning sidebands observed in slow MAS NMR spectra arising from tightly **J**-coupled spin pairs contain valuable information about NMR parameters such as the orientation of chemical shift tensors and the sign of **J**. Multidimensional NMR spectra in solids permit the separate determination of the isotropic chemical shifts and the anisotropic line shapes that contain chemical shift tensor and quadrupole coupling information for each site.

Thus, we have seen that the parameters of an NMR spectrum are related to fundamental molecular electronic properties: the chemical shift is related to nuclear magnetic shielding $\boldsymbol{\sigma}$ and the nuclear quadrupole coupling is related to the electric field gradient tensor $e\mathbf{q}$. The indirect spin–spin coupling **J** is itself a molecular electronic property. Therefore, the general approach to the theoretical calculations of these NMR parameters is through a quantum mechanical calculation of molecular electronic properties in the isolated molecule. Any medium effects that have to be included, when they are large enough, require, in addition, ensemble averages for a gas, liquid, or solution. The calculation of the electric field gradient is simplest, since this is a property of the unperturbed electronic state of the molecule. Since $\boldsymbol{\sigma}$ and **J** are electronic properties associated with the presence of magnetic fields and fields generated by nuclear magnetic moments, their calculation requires the general approaches that apply to multiple perturbations. Furthermore, since the probe nucleus senses electronic environments in the immediate vicinity of the nucleus, high-level calculations that take into account electron correlation have to be used for all three parameters to achieve accuracy, and relativistic corrections are sometimes necessary. Density functional methods have been very successful and can compete favorably with ab initio calculations.

2 GENERAL THEORETICAL METHODS

The mechanisms by which a nuclear magnetic moment interacts with the molecular field and with external magnetic or electric fields in the ground vibronic state were originally articulated in fundamental work by Ramsey.[1–3] For a unified approach to molecular electronic properties which explicitly shows where the contributing terms arise and thereby also permits the relationships between electronic properties to be perceived, consult the articles by Michelot.[4,5] The complete molecular Hamiltonian in the presence of external magnetic and electric fields, including all relevant interaction terms involving nuclear magnetic moments (such as interaction between the nuclear magnetic moment and the field induced at the nucleus by the molecular motion, as well as those related to the interaction of the magnetic moment induced by this molecular motion with an external magnetic field), treats electrons and nuclei as Dirac particles. Relativistic effects are included from the beginning and effects due to the finite dimensions of nuclei are also taken into account, so that the nuclear quadrupole coupling is a natural outcome.[4] Using this Hamiltonian with relativistic corrections for a free molecule in a nondegenerate electronic state, a second-order calculation in degenerate perturbation theory leads to the explicit expressions for the contributing terms to nuclear magnetic shielding $\boldsymbol{\sigma}$, indirect spin–spin coupling **J**, nuclear electric quadrupole coupling, and all other molecular electronic properties.[5]

2.1 Multiple Perturbation Theory

All the terms in the molecular Hamiltonian given by Michelot[4] may be treated as perturbations added to a zeroth-order part (the kinetic energy of the electrons together with the total coulomb potential energy of all the electrons and nuclei, assumed to have already been solved). These include terms bilinear in $\boldsymbol{\mu}_N$ and **B**. In first order, these will lead to energy terms that are of the form

given by Equation (4), providing the formal expression for the so-called diamagnetic part of the nuclear magnetic shielding $\boldsymbol{\sigma}$. In second order, the terms in the molecular Hamiltonian that are linear in $\boldsymbol{\mu}_N$ together with those linear in $\mathbf{B}$ will lead to energy terms that are also of the form given by Equation (4), providing the formal expression for the so-called paramagnetic part of the nuclear magnetic shielding $\boldsymbol{\sigma}$. Michelot's expression[5] derived for nuclear magnetic shielding $\boldsymbol{\sigma}$ reduces to that given by Ramsey[1,2] if the origin of the molecular frame is placed at the center of the nucleus of interest, the orientation being that of the Eckart frame. Formally, the shielding term for nucleus N is given by Equation (9):

$$\begin{aligned}\sigma_{\alpha\beta}(\mathrm{N}) = \frac{\mu_0 e^2}{8\pi m}\langle 0|\sum_k \{(\boldsymbol{r}_{k\mathrm{N}}\boldsymbol{r}_{k0})r_{k\mathrm{N}}{}^{-3}\delta_{\alpha\beta} \\ - (r^{\alpha}_{k\mathrm{N}}r^{\beta}_{k0})r_{k\mathrm{N}}{}^{-3}\}|0\rangle - \frac{\mu_0 e^2}{8\pi m^2}\sum_{n\neq 0}({}^1E_n - E_0)^{-1} \\ \times \quad \langle 0|\sum_k r_{k\mathrm{N}}{}^{-3}L^{\alpha}_{k\mathrm{N}}|n\rangle\langle n|\sum_j L^{\beta}_{j0}|0\rangle \\ + \langle 0|\sum_j L^{\beta}_{j0}|n\rangle\langle n|\sum_k r_{k\mathrm{N}}{}^{-3}L^{\alpha}_{k\mathrm{N}}|0\rangle \end{aligned} \tag{9}$$

The first index $\alpha(=x, y, z)$ is associated with the nuclear magnetic moment and the second index $\beta(=x, y, z)$ is associated with the external magnetic field. L^{β}_{j0} is the β component of the orbital angular momentum operator for the jth electron with respect to the chosen origin (so-called gauge origin) and $\boldsymbol{r}_{k0}$ is the distance vector between the kth electron and the origin. $L^{\alpha}_{k\mathrm{N}}$ is the α component of the orbital angular momentum operator for the kth electron with respect to the nucleus N as origin. $\boldsymbol{r}_{k\mathrm{N}}$ is the distance vector between the kth electron and the nucleus N. m and e are the mass and charge of the electron, μ_0 is the magnetic permeability of a vacuum. E stands for the energy at states 0 and n (0 is the lowest state and n is an index that runs through all the states of the molecule). In Equation (9) the second term in the expression for $\sigma_{\alpha\beta}(\mathrm{N})$, the paramagnetic term, is written in the so-called sum over states (SOS) form, that is, in terms of the sum $\sum_n$ over all excited states designated by the symbol $|n\rangle$. In the symbol $\langle 0|\sum_k r_{k\mathrm{N}}{}^{-3}L^{\alpha}_{k\mathrm{N}}|n\rangle$ the operators for the angular momentum and the distance vector for the electrons are integrated over the ground state $\langle 0|$ and the excited state $|n\rangle$.

In the same way, the terms in the molecular Hamiltonian that are bilinear in $\boldsymbol{\mu}_N$ and $\boldsymbol{\mu}_{N'}$ lead to the energy terms that are already of the form given by Equation (7) give rise to the first order part of the indirect spin–spin coupling $\mathbf{J}$, usually denoted by $\mathbf{J}^{(1a)}$. The terms linear in $\boldsymbol{\mu}_N$ in the molecular Hamiltonian in the nonrelativistic limit are three, labeled orbital, spin dipolar (SD), and Fermi contact (FC). In second order, products of these lead to various contributions to the spin–spin coupling $\mathbf{J}$. The product of orbital terms leads to $\mathbf{J}^{(1b)}$; $\mathbf{J}^{(1a)}$ is sometimes called the diamagnetic orbital (OD) contribution or $\mathbf{J}^{(OD)}$, and $\mathbf{J}^{(1b)}$ the paramagnetic orbital (OP) contribution or $\mathbf{J}^{(OP)}$ because of the analogy with the diamagnetic and paramagnetic parts of the shielding tensor. The product of SD terms leads to $\mathbf{J}^{(2)}$ or $\mathbf{J}^{(SD)}$, and the product of FC terms leads to $\mathbf{J}^{(3)}$ or $\mathbf{J}^{(FC)}$. By symmetry, there is only one nonvanishing cross-term, resulting from the product of the SD and the contact terms, referred to as $\mathbf{J}^{(4)}$. $\mathbf{J}^{(FC)}$ is purely scalar (isotropic), whereas the others are anisotropic. The motional average of $\mathbf{J}^{(4)}$ is zero, thus, all but $\mathbf{J}^{(4)}$ contribute to the observed isotropic average spin–spin coupling for a rapidly tumbling molecule in solution. All terms contribute to the observed NMR spectrum in solids. The formal expressions for spin–spin coupling in the nonrelativistic limit are shown below in terms of the spin ($\mathbf{S}_k$) and orbital ($L^{\alpha}_{k\mathrm{N}}$) angular momentum of the (kth) electron, Equations (10) to (14):

$$\begin{aligned}\mathbf{J}^{(1a)}_{\alpha\beta} = \frac{2m}{h}\mu_B^2\left(\frac{\mu_0}{4\pi}\right)^2\gamma_N\gamma_{N'}\langle 0|\sum_k r_{k\mathrm{N}}{}^{-3}r_{k\mathrm{N}'}{}^{-3} \\ \times [(\boldsymbol{r}_{k\mathrm{N}}\boldsymbol{r}_{k\mathrm{N}'})\delta_{\alpha\beta} - r^{\alpha}_{k\mathrm{N}}r^{\beta}_{k\mathrm{N}'}]|0\rangle \end{aligned} \tag{10}$$

$$\begin{aligned}\mathbf{J}^{(1b)}_{\alpha\beta} = \frac{1}{h}(2\mu_B)^2\left(\frac{\mu_0}{4\pi}\right)^2\gamma_N\gamma_{N'}\sum_n({}^1E_n - E_0)^{-1} \\ \times \quad \langle 0|\sum_k r_{k\mathrm{N}}{}^{-3}L^{\alpha}_{k\mathrm{N}}|n\rangle\langle n|\sum_j r_{j\mathrm{N}'}{}^{-3}L^{\beta}_{j\mathrm{N}'}|0\rangle + \mathrm{cc} \end{aligned} \tag{11}$$

$$\begin{aligned}\mathbf{J}^{(2)}_{\alpha\beta} = -\frac{1}{h}(2\mu_B\hbar)^2\left(\frac{\mu_0}{4\pi}\right)^2\gamma_N\gamma_{N'}\sum_n({}^3E_n - E_0)^{-1} \\ \times \quad \langle 0|\sum_k 3r_{k\mathrm{N}}{}^{-5}(\mathbf{S}_k\mathbf{r}_{k\mathrm{N}})r^{\alpha}{}_{k\mathrm{N}} - r_{k\mathrm{N}}{}^{-3}\mathbf{S}^{\alpha}_k|n\rangle \\ \times \langle n|\sum_j 3r_{j\mathrm{N}'}{}^{-5}(\mathbf{S}_k\mathbf{r}_{j\mathrm{N}'})r^{\beta}_{j\mathrm{N}'} - r_{j\mathrm{N}'}{}^{-3}\mathbf{S}^{\beta}_j|0\rangle + \mathrm{cc} \end{aligned} \tag{12}$$

$$\begin{aligned}\mathbf{J}^{(3)}_{\alpha\beta} = -\frac{1}{h}\left(\frac{16\pi\mu_B\hbar}{3}\right)^2\left(\frac{\mu_0}{4\pi}\right)^2\gamma_N\gamma_{N'} \\ \times \sum_n({}^3\mathrm{E}_n - \mathrm{E}_0)^{-1} \quad \langle 0|\sum_k \delta(r_{k\mathrm{N}})\mathbf{S}^{\alpha}_k|n\rangle \\ \times \langle n|\sum_j \delta(r_{j\mathrm{N}'})\mathbf{S}^{\beta}_j|0\rangle + \mathrm{cc} \end{aligned} \tag{13}$$

$$\mathbf{J}^{(4)}_{\alpha\beta} = -\frac{2}{h}\frac{16\pi\boldsymbol{\mu}_B\hbar}{3}2\boldsymbol{\mu}_B\hbar \quad \frac{\boldsymbol{\mu}_0}{4\pi} \quad {}^2\gamma_N\gamma_{N'} \times \sum_n ({}^3E_n - E_0)^{-1} \quad \langle 0| \sum_k \delta(r_{kN})\mathbf{S}^\alpha_k|n\rangle \times \langle n| \sum_j 3r_{jN'}{}^{-5}(\mathbf{S}_k\boldsymbol{r}_{jN'})r^\beta_{jN'} - r_{jN'}{}^{-3}\mathbf{S}^\beta_j|0\rangle + \text{cc} \tag{14}$$

where cc indicates the conjugate term in which the two operators are switched. $\boldsymbol{\mu}_B$ is the Bohr magneton, γ_N is the magnetogyric ratio for the nucleus N. $\delta(\boldsymbol{r}_{kN})$ is the Dirac delta function which picks out the value at $\boldsymbol{r}_{kN} = 0$ in any integration over the coordinates of the kth electron.

The coupling contributions $\mathbf{J}^{(1a)}$ or $\mathbf{J}^{(OD)}$ and $\mathbf{J}^{(1b)}$ or $\mathbf{J}^{(OP)}$ can be thought of as arising through paramagnetic and diamagnetic currents induced in the molecular electronic distribution by the nuclear magnetic moment of one of the nuclei, coupling to the magnetic moment of the other nucleus. The coupling contribution $\mathbf{J}^{(3)}$ or $\mathbf{J}^{(FC)}$ can be considered as arising from the transmission of spin information from nuclear spin to electron spin. Owing to the finite density of the electron at the nucleus, this information is passed on through the spin interaction between electrons in the molecule and transmitted at the other end via electron spin density at the other nucleus. The dipole–dipole interaction between the nuclear and electron spins lead to the coupling contribution $\mathbf{J}^{(2)}$ or $\mathbf{J}^{(SD)}$.

The expressions shown here in Equations (9–14) are cast in the form of sums over excited states, as they were originally cast in the Ramsey formulation. However, practical calculations are not actually carried out in this form for several reasons. Multiple perturbation theory is more conveniently carried out by using directly the first-order perturbed wavefunction or the first-order density matrix. In other words, for the multiple perturbation of the external magnetic field and the nuclear magnetic moment, the nuclear magnetic shielding may be calculated by first calculating the first-order density matrix of the molecule in the external magnetic field alone, using the operator $\sum_j L^\beta_{j0}$, and with this the integrals that account for the second perturbation imposed by the nuclear magnetic moment are then evaluated, using the operator $\sum_k r_{kN}{}^{-3}L^\alpha_{kN}$. Or, independently, one could first find the first-order density matrix of the molecule in the presence of the nuclear magnetic moment alone, and with this, evaluate the integrals that account for the second perturbation imposed by the external magnetic field. A physical interpretation is that the nuclear magnetic shielding arises from the interaction of the magnetic moment of the nucleus with the magnetic field due to the current density induced by the external magnet or, equivalently, from the interaction of the current density induced by the nuclear magnetic moment with the external magnetic field. The most common approach is to construct first the electronic response induced by the external magnet and then study its interaction with various nuclei in the molecule. Note that written in the form of Equation (9), the diamagnetic part of the shielding is very easy to calculate, since it requires an average over the unperturbed electronic ground state function only. On the other hand, the second-order term, the paramagnetic contribution, requires a knowledge of how the presence of the external magnetic field changes the electronic wavefunction of the molecule and the integration requires that this knowledge be especially accurate in the immediate vicinity of the nucleus of interest.

The calculation of the electric field gradient tensor does not require a perturbation treatment since this is one of those electronic properties that can be calculated as an average over the electronic ground-state wavefunction. The zz component of the electric field gradient tensor is given by Equation (15)

$$q_{zz} = \sum_j e_j(3z_j{}^2 - r_j{}^2)r_j^{-5} \tag{15}$$

where e_j is the charge of the jth particle (the electrons, other nuclei, external charges) in the system and the z axis is in the nuclear-fixed coordinate system, that is along the nuclear axis of spin. The spin axis of the nucleus is allowed to rotate with respect to the laboratory frame of reference and the nuclear wavefunction will be a product of the intrinsic $\Psi_{\text{intrinsic}}$ and orientation $\Psi_{I,M}$ functions for spin angular momentum described by quantum numbers I and M, Equation (16)

$$\Psi_{\text{total}} = \Psi_{I,M}\Psi_{\text{intrinsic}}\Phi(\text{electrons, other nuclei, external charges}) \tag{16}$$

For a nucleus in a molecule oriented in the laboratory framework, the components of the field gradient tensor are q_{XX}, q_{YY}, q_{ZZ}. The principal field gradient tensor component q_{zz} is related to the laboratory values through the direction cosines between the axes, as follows, Equation (17):

$$q_{zz} = (C_{Xz})^2q_{XX} + (C_{Yz})^2q_{YY} + (C_{Zz})^2q_{ZZ} \tag{17}$$

In the absence of a magnetic field, the energy of a quadrupolar nucleus in the electric field gradient will be obtained by averaging the electric quadrupole charge over the wavefunction $\Psi_{\text{intrinsic}}$, averaging the squares of the direction cosines over $\Psi_{I,M}$, and averaging $q_{ZZ} = \sum_j e_j(3Z_j{}^2 - r_j{}^2)r_j^{-5}$ over the wavefunction Φ (electrons, other nuclei, external charges) expressed in the laboratory frame. The average of the direction cosines over $\Psi_{I,M}$

leads to an energy expression that is proportional to $[3M^2 - I(I+1)]$ in the absence of a magnetic field.

In the presence of a magnetic field, the magnetic dipole moment of a quadrupolar nucleus, which lies along the axis of symmetry of the cigar-like or pancake-like nuclear charge distribution, interacts with the magnetic field. Thus, when a single quadrupolar nucleus in a molecule is placed in a magnetic field, the interaction of the electric quadrupole with the internal electric field gradient at the nuclear site in the molecule leads to a series of $2I$ resonance lines. Thus, a spin $I = 1$ nucleus in an axially symmetric electric field gradient gives a pair of lines separated by $[e^2 q_{ZZ} Q](3/2)$ or more explicitly, by $[e^2 q_{zz} Q](3/2)(3\cos^2\theta - 1)/2$, where θ is the angle that the principal symmetry axis of the electric field gradient makes with the external magnetic field. The pair of lines is centered at a frequency that provides the shielding tensor. Both the shielding tensor and the electric field gradient tensor in the *XYZ* (i.e. the laboratory-fixed) coordinate system can be obtained from an oriented molecule in the solid state. Equation (17) permits the determination of the tensor in the *xyz* (i.e. the molecule-fixed) coordinate frame system. Since the electric field gradient is a traceless tensor, the isotropic average is zero. Thus, in the liquid phase the positions of the lines in the NMR spectrum are not affected by the nuclear quadrupole coupling, although information about the latter can still be obtained from quadrupolar relaxation times.

Theoretical calculations of the electric field gradient tensor in *the molecular frame* of an isolated molecule involves evaluating the quantum mechanical average of the operator $q_{zz} = \sum_j e_j(3z_j^2 - r_j^2)r_j^{-5}$ over the ground state electronic wavefunction for the molecule, where j runs over all electrons and the origin is set at the nucleus in question. To this electronic contribution must be added the nuclear contribution, by evaluating a similar algebraic expression in which e_j are the charges of the other nuclei and z_j and r_j are their positions in the molecular framework with the origin at the nucleus in question. For molecules in a liquid, electric field gradient contributions from neighbors have to be included, which may require a quantum mechanical average or an approximate sum over fixed partial charges.

2.2 Gauge Origin Problem in Calculations of Chemical Shift

In deriving the expressions shown here, the external magnetic field **B** itself does not appear in the Hamiltonian. What appears instead are the magnetic vector potentials associated with the magnetic fields, Equation (18):

$$\mathbf{B} = \nabla \mathbf{A} \tag{18}$$

where ∇ is the gradient vector. That is, $B_z = (\partial A_y/\partial x) - (\partial A_x/\partial y)$, for one component. While **B** is determined uniquely if **A** is given, unfortunately there is an ambiguity because there is no unique **A** that produces a given **B**. Any transformation that takes a particular **A** into another functional form that also reproduces the same **B** upon applying Equation (18) is called a gauge transformation. A mere translation of the origin of the coordinate system can do this, therefore the set of problems associated with this ambiguity is called the gauge origin problem. Physically, there should be no problem at all, since an arbitrary choice of coordinate system should not affect an observable property. Similarly, theoretically there should be no problem at all; any physical quantities resulting from any calculations involving **A** or **B** or physical quantities related to them must be gauge invariant, provided the calculations are done exactly. In fact, calculations are not usually done exactly when one uses an incomplete set of basis functions in which to do the calculations. It has been shown that if the Hartree–Fock equations are solved exactly (which is only possible in the limit of a complete basis) the total current density is gauge independent, as is the nuclear magnetic shielding $\boldsymbol{\sigma}$, while the two parts which are usually called the ''diamagnetic'' and ''paramagnetic'' contributions in Equation (18) are not individually gauge invariant. In practice, calculations are not carried out in the Hartree–Fock limit so the results of such calculations are not gauge invariant. When a single origin is chosen common to all electrons in the molecule in the definition of $\sum_j L_{j0}^{\beta}$ (and $\boldsymbol{r}_{j0}$), the method is the so-called ''common origin'' coupled Hartree–Fock (CHF) method.

Consider an isolated atom. The external magnetic field induces a current density. The current density vector is orthogonal to the magnetic field vector **B** and to the position vector $\boldsymbol{r}_j$. For a magnetic field in the z direction the current density vectors lie in planes parallel to the xy plane, following the tangents of concentric circles. Here, the natural choice of origin is the position of the nucleus; this leads to a vanishing paramagnetic current density. The current density is entirely the diamagnetic part and corresponds to a local field that opposes the external field **B**. Moving the origin off-center to other than the position of the nucleus would make the two parts more difficult to evaluate, but the sum should still be the same as before, so there is no reason to adopt an alternative origin. In molecules, however, there is no choice of origin that will make the paramagnetic part vanish. Changing the location of the origin, in the definition of $\boldsymbol{r}_{j0}$ and $\sum_j L_{j0}^{\beta}$ in Equation (18) leads to differing amounts of positive and negative terms. The worst choice gives very large not quite canceling terms. Clearly, the inner shell electrons in a molecule behave like they do in the free atom, so that it makes sense to choose the nucleus as the origin when

calculating integrals over orbitals centered on that atom. However, that same origin would be a bad choice for orbitals centered on another atom, whereas choosing the nucleus of that atom as origin would present a favorable atom-like calculation for its own inner electrons. Thus, it becomes clear that in order to avoid calculating large positive and negative terms which imperfectly cancel in a single origin method, some method of using distributed origins would present a practical advantage in computing nuclear magnetic shielding for nuclei in molecules.

The theoretical calculations of nuclear magnetic shielding did not become generally practical even for very small molecules until (1) various ways of using distributed origins were devised, and (2) efficient algorithms for evaluating second-order properties were developed. The various schemes for using distributed origins are known by the acronyms LORG (localized orbital/local origin),[6] IGLO (individual gauge for localized orbitals),[7] GIAOs (gauge including atomic orbitals),[8] and IGAIM (individual gauges for atoms in molecules).[9] The success of distributed origins comes from the avoidance of calculating large imperfectly canceling contributions. In the first two methods, gauge factors are applied to localized molecular orbitals instead of every atomic orbital. The LORG and IGLO methods introduce an approximation in the form of the closure relation and LORG uses commutation rules and identities. Both have been very successful, although there is the problem of lack of uniqueness in the localization method used. The GIAO method uses gauge factors on every atomic orbital. Although this method of distributed origins had been introduced much earlier than all the others, it was not until the efficient implementation by Peter Pulay using the analytic gradients approach that it became widely successful. The convergence of calculated $\boldsymbol{\sigma}$ values with increasing quality of basis set employed appears to be faster with the GIAO method. GIAOs (sometimes called London orbitals) constitute a physically motivated, compact basis set for magnetic calculations. The field-dependent exponential factor in the London orbital depends on the origin of the coordinate system. A displacement of the origin changes the phase factor of an orbital centered on a nucleus by a factor which is independent of the electronic coordinates. Thus, the calculated properties such as shielding remain unaffected and methods based on the use of such orbitals are gauge invariant. The most important property of the GIAO method is not this formal translational invariance but that the GIAO (the atomic orbital multiplied by the gauge factor) itself represents to first order the eigenfunctions of a one-electron system which has been perturbed by an external magnetic field. GIAOs thus incorporate the bulk of the effect of the magnetic field at the basis function level. The IGAIM approach amounts to constructing the induced current density distribution of a molecule from its constituent atoms, following the highly successful atoms-in-molecules concepts of R.F.W. Bader. It differs from LORG, IGLO, and GIAO in that the gauge origins are determined by properties of the charge density in real space rather than by the behavior exhibited by the basis functions in the Hilbert space of the molecular wavefunction. All these distributed origin methods (GIAO, IGLO, LORG, IGAIM) and any single common origin method should lead to identical results at the Hartree–Fock level in the limit of a complete set of basis functions. The differences lie in the rate of convergence as the number of basis functions are increased. The various distributed-origins methods converge toward the Hartree–Fock limit faster than using a single origin. Common origin calculations require much larger basis sets to provide nearly origin-independent results comparable to the results from distributed-origin methods. An alternative method of doing calculations with a single origin is to cast the diamagnetic term in the same (a SOS) form as the paramagnetic term.[10] This makes the rate of convergence of the calculations of the two parts equally slow and the calculations equally difficult. This (Geertsen's method) has the virtue of being origin-independent (i.e. gives the same answer for any choice of common origin) at any basis set size.

2.3 Difficulties of Describing Triplet States in Calculations of Spin–Spin Coupling

There are no gauge problems in spin–spin coupling calculations; as seen in Equations (10–14) only operators with their origin at the nucleus ($\boldsymbol{r}_{k\mathrm{N}}$ and $\boldsymbol{r}_{k\mathrm{N}}{}^{-3}L^{\alpha}_{k\mathrm{N}}$) appear. The calculations of spin–spin coupling have their own associated difficulties. As can be seen in Equations (10–14), the nature of some of the indirect spin–spin coupling mechanisms requires calculations with uncoupled spin states. Thus, the spin-unrestricted approaches that are normally applied to open shell systems have to be used. When there is a nonsinglet ground state with lower energy than the restricted Hartree–Fock singlet ground state, the calculations of the $\mathbf{J}^{(\mathrm{SD})}$ and $\mathbf{J}^{(\mathrm{FC})}$ terms require higher order calculations than CHF. Furthermore, the usually (not always) dominant FC contribution in Equation (13) requires that the spin densities are highly accurate at the location of the nucleus, and this is not easily achieved when the basis functions used are the standard gaussian form, having no cusp at the nucleus. Relativistic effects influence spin–spin couplings much earlier (at lower atomic numbers) than other properties, owing to the strong dependence of $\mathbf{J}$ on the electronic structure at the position of the nucleus and its immediate vicinity. In fact, Equations (10–14) are valid only for the point nucleus in the nonrelativistic limit. For heavy nuclei it is necessary

to start out with the relativistic treatment described in section 2.6, since the nonrelativistic theory may lead to unrealistic calculated values.[11]

2.4 Ab Initio Methods

In ab initio calculations of NMR parameters there are several things that have to be considered: (1) the level of theory that is used (without, with some, or with substantial electron correlation, with or without relativistic corrections), (2) the number of basis functions, (3) the desired degree of averaging over molecular configurations (with or without rovibrational averaging, with or without medium effects), (4) the availability of tensor data, and (5) the availability of absolute shielding test data in the case of chemical shifts. In the case of spin–spin coupling, there is very limited information beyond that of the isotropic average values obtained in solution. Only in extremely rare cases is there anisotropy information, so that the only viable additional tests are those of isotopic effects on spin–spin coupling. In the case of ^{13}C chemical shifts the amount of detailed tensor information from single crystals and polycrystalline powders is so rich that the level of theory, and size of basis sets needed, and the accuracy of geometrical structure data required to achieve agreement with experiment has been established (by D.M. Grant and associates) for a large variety of carbon site types. The level of theory used in ab initio calculations of NMR parameters had to improve continuously with the challenges posed by attempts to match experimental results for specific small molecules, as we shall see below (section 3.1). Depending on the need for accuracy, depending on the nucleus and the nature of the nuclear site, the appropriate level of calculation can be done. Various levels of theory have been used with various numbers of basis functions. The need to establish good basis functions still hinders accurate calculations for transition metal nuclei. Furthermore, in some cases there is a lack of a good test molecule in the gas phase for which absolute shielding has been established. For the purpose of distinguishing between two structures or even two chemical identities, calculations of $\boldsymbol{\sigma}$ using uncorrelated wavefunctions constructed with smallish basis sets have sometimes been employed. This is a very dangerous approach, which has been criticized. Ideally, a calculation to estimate the importance of electron correlation to the nuclear shielding should be done. An estimate of the magnitude of the effects of averaging over molecular configurations is needed, in order to determine whether it is sufficient to do a calculation for a single molecule at a fixed geometry in a vacuum. An estimate of the magnitude of relativistic effects on shielding is needed when next neighbors are halogen atoms, or when the nucleus is a heavy nucleus. The involvement of multiple bonds or presence of lone pairs at the nucleus of interest is usually an indication that correlated methods have to be used.

An alternative to the CHF approach is the use of polarization propagator or the equation of motion (EOM) methods. With these latter two methods the level of calculation equivalent to the CHF level is the random phase approximation (RPA). For the same basis set, using the same origin choice, calculations using CHF and RPA should provide the same results. In terms of perturbation theory, the RPA is the consistent first-order approximation to the polarization propagator or the EOM. The CHF, RPA, and the first-order polarization propagator method represent *the same approximation* for frequency-independent properties such as the NMR properties $\boldsymbol{\sigma}$ and **J**. With respect to the extent of inclusion of electron correlation, CHF and RPA provide the first level of calculations.

In the finite field method, the NMR parameter (for example, **J**) is obtained by differentiating the energy in the presence of the nuclear magnetic moments (or in the presence of the nuclear moment and the external magnetic field for $\boldsymbol{\sigma}$) with respect to the nuclear moments (or with respect to the nuclear moment and the external field, for $\boldsymbol{\sigma}$). The addition of a finite field to the total energy expression is a simple extension of existing computer codes for electronic structure calculations and is one of the standard methods for calculating higher order molecular electronic properties such as the nonlinear polarizabilities, for example. Thus, the finite field method is easily used, without additional theoretical development, to study the effects of electron correlation on properties. The drawback is that a finite field calculation has to be carried out for each tensor component of the property. Thus, while the purely isotropic FC term of **J** is easily done with finite field methods at various levels such as various many body perturbation theory (nth order term) (MBPT(n)) and various coupled cluster (CC) methods, the tensor types of mechanisms given in Equations (10) to (12) and (14) require several calculations to yield the various xx, xy, xz, yy, zz, components. Direct methods such as the polarization propagator or EOM method on the other hand, use analytic expressions that provide all components of the tensor with one calculation. In an MBPT(n) calculation, or alternatively the Møller–Plesset nth order term (MPn) perturbation series, all energy contributions less than or equal to order n in perturbation theory are included. CC methods on the other hand, in addition to being consistent to a particular order in perturbation theory include certain classes of energy contributions summed to infinite order. The same kind of infinite summations are also included in polarization propagator methods (RPA, SOPPA, etc.). Thus, SOPPA

(second-order polarization propagator approximation) is not equivalent to the MBPT(2) (second-order many body perturbation theory) or MP2 (Møller–Plesset second-order term) level of approximation. The coupled cluster reference state including singles and doubles excitations (CCSD) has been introduced in post Hartree–Fock shielding calculations by Gauss, CC has been used in finite field calculations of $\mathbf{J}^{(FC)}$ by Bartlett et al. and has been used in the polarization propagator method as an extension of SOPPA for the calculations of all the mechanisms of **J**.

For chemical shift calculations the CHF level, sometimes also referred to as the self-consistent field (SCF) level since the calculations use the SCF wavefunctions, appears to give quite good results for CH_4 and other saturated ^{13}C sites, and even olefinic sites. The major problems that have been discovered for carbon are the C≡O, −C≡N, and >C=O environments, for which sites electron correlation is extremely important. The SCF level has been found to be insufficient even for calculations using very large basis sets to reproduce the anisotropy of the ^{31}P shielding in the molecule PH_3.[12] Apparently, electron correlation is important for ^{31}P calculations in molecules where a lone pair is on the phosphorus, even when only hydrogen atoms are attached to it. In these cases, some level of post-Hartree–Fock theory is necessary in order to obtain meaningful results.

The next level of theory for shielding provides electron correlation effects at the level of MBPT(2) or MP2 perturbation theory. MP2 calculations may be carried out with a conventional common origin or may be used with any of the distributed origin approaches. An alternative approach to second-order electron correlation effects called SOPPA belongs to the family of propagator techniques, but is different from and not equivalent to MP2. Many quantum mechanical software packages provide MP2 level of wavefunctions. These are supposed to take care of the dynamic correlation effects and hence to improve results for closed shell systems where CHF already gives good results. When the degree of electron correlation contribution to the shielding is small, this level is usually sufficient to provide useful comparisons with experiment. When the difference between SCF and MP2 results is not small, then it may be necessary to go up to MP4 level, with the corrections at MP2, MP3, and MP4 often alternating in sign. Despite its successes, the inherent problem with this approach is the slow convergence of the perturbation series for those systems with strong correlation effects, that is for those cases where MBPT(2) or MP2 level is no longer adequate.

Multiconfiguration self-consistent field wavefunctions (MCSCF) may sometimes be necessary for calculations of shielding, when the electronic ground state cannot be adequately described by a single Slater determinant. A multireference calculation can properly account for strong correlation effects if closed shell SCF is too poor as an initial approximation (for example, in O_3, NSF, SO_2, N_2O_3 molecules). CHF methods completely fail for such systems. MCSCF wavefunctions, generated using complete active spaces (CAS) and restricted active spaces (RAS), may be used with any of the distributed origin methods, spawning such methods as MCSCF/GIAO[13] and MCIGLO (multiconfiguration individual gauge for localized orbitals)[14]. The MCIGLO formulation was presented formally by van Wullen and Kutzelnigg,[14] and applied to appropriate cases such as carbenes and dinitrogen oxides.[14] The dinitrogen oxides $(NO)_2$, $(NO)(NO_2)$, and $(NO_2)_2$ have strong correlation effects which affect the shielding tensors. The ^{15}N and ^{17}O shieldings have been measured in these molecules and the effects of correlation are particularly interesting in that they are of both signs. In the nitroso nitrogen, the electron correlation effect is to enhance deshielding, while in the nitro nitrogen the correlation effects are to increase shielding. Thus, in $(NO)(NO_2)$ the correlation effects are large and opposite in sign for the two types of N. Correlation contributions to the isotropic ^{15}N shielding range are −558 ppm in $(NO)_2$, +61 ppm in $(NO_2)_2$ and −63 and +116 ppm in $(NO)(NO_2)$.

If different electronic configurations dominate the wavefunction at different geometries, the calculation of the shielding surface also requires a computational method based upon a multiconfiguration wavefunction.[14] The MCSCF approaches are hampered by the same sort of problem because rather large active spaces are needed to obtain satisfactorily converged results. While static correlation effects on shielding arising from near degeneracies are efficiently treated by the MCSCF methods described above, many body perturbation theory (MBPT, also known as Møller–Plesset perturbation theory) has been used to treat dynamical correlation effects.[15–17]

One of the most successful approaches for the treatment of electron correlation is provided by CC theory. While ultimately based on a single determinant reference function, the exponential parametrization of the wavefunction ensures an efficient treatment of electron correlation. In particular, dynamic correlation effects are accounted for with nearly quantitative accuracy at a fraction of the cost needed to obtain similar precision with MCSCF approaches. Among the various schemes suggested in the literature, the CCSD approximation[18] in which single and double excitations are considered in the cluster operator, has proven specially useful in calculations of other molecular properties. CC approaches can be considered as infinite-order generalizations of the MBPT series. The implementation of GIAOs for the CCSD approach has been carried out by Gauss and Stanton,[19]

and further augmented by a perturbative correction for connected triple excitations coupled cluster singles and doubles with some triple excitations (CCSD(T)).[20] The principal advantage of the GIAO method is the ease with which high-level treatments of electron correlation may be handled by straightforward application of analytic derivative theory.

MCSCF/GIAO calculations for triple-bonded systems, in particular HC≡N, HN≡C, MeC≡N, and MeN≡C[21] show that the electron correlation effects are large for the triple-bonded nuclei, especially the component perpendicular to the triple bond axis, and largest for the terminal nucleus. For example, the electron correlation contribution to $\sigma_{\perp}^{13}$C shielding is +47 to +54 ppm in the $-N\equiv{}^{13}C$ nuclear sites, and to $\sigma_{\perp}$ nitrogen shielding is +87 to +79 ppm in the $-C\equiv N$ sites.[21] These MCSCF/GIAO results do not compare as well with experiments as do the calculations by Gauss using the CCSD method.[19,20] The ultimate level of theory would be full configuration interaction (FCI), but this is only possible for very small systems and is rarely used.[22]

The same general methods for multiple perturbations are used for calculating spin–spin couplings, with the difference that there are no gauge origin problems in spin–spin coupling calculations. All but the $\mathbf{J}^{(OD)} + \mathbf{J}^{(OP)}$ mechanisms mix triplet states with the unperturbed electronic singlet ground state. Thus, the unrestricted Hartree–Fock (UHF) method is sometimes used to generate the unperturbed electronic ground state, even for closed shell molecular systems such as CH_4. The extent to which electron correlation needs to be included depends on the system, just as in shielding calculations. The same molecules that are found to be pathological cases in shielding calculations also pose problems in spin–spin coupling calculations. The uncorrelated finite field SCF calculations, which are the same approximation as CHF and RPA calculations, in most cases give good results for $\boldsymbol{\sigma}$ while this is not often true for $\mathbf{J}$. The $\mathbf{J}^{(OD)}$ term in spin–spin coupling, like the diamagnetic part of $\boldsymbol{\sigma}$, is generally easy to calculate; good results can be obtained with low-level electronic correlation included and using moderately sized basis sets. The sum over all excited states in $\mathbf{J}^{(OP)}$ extends over singlet excited states since most of the common ground states that chemists are interested in are singlet states. However, in the $\mathbf{J}^{(SD)}$, $\mathbf{J}^{(FC)}$, and $\mathbf{J}^{(SDFC)}$ expressions, the excited states have different spin multiplicity from the ground state. Because there is a state of triplet symmetry either very close to or sometimes below the restricted Hartree–Fock singlet ground state for many molecules, these mechanisms for spin–spin coupling are very poorly described at the CHF level of approximation. This problem affects spin–spin coupling calculations, but not calculations of $\boldsymbol{\sigma}$. A resurgence of interest in accurate ab initio calculations of coupling constants was initiated by Oddershede et al.[23–25] using polarization propagator methods. The SOPPA has been found to yield reliable coupling constants in some instances.[23,26] When higher accuracy is required, CCSD reference states (limited to single and double excitations) have been used within the polarization propagator method.[26,27] Midway between first (RPA) and second-order is HRPA (higher random phase approximation), which has been used by Galasso in the EOM approach to calculate one- and two-bond coupling constants $^{1 \text{ or } 2}\mathbf{J}$(CC) in ring systems.[28] MCSCF wavefunctions are appropriate in cases where the SCF calculations predict unrealistic coupling constants,[29–31] such as in (pathological) molecules involving multiply bonded nitrogens.[30] The method is MCLR (multiconfiguration linear response).[29] MCSCF functions have been used for the hydrides of group IV (C, Si, Ge, Sn).[31] Finite field methods have been used by Bartlett et al. to calculate the FC mechanism of spin–spin coupling using various levels including electron correlation, up to CCSD.[32] MBPT has also been used.[33]

2.5 Density Functional Methods

One method of including electron correlation effects is through density functional theory (DFT). DFT methods have become widely used. DFT methods are based on a theorem which states that for a scalar potential $V(\mathbf{r})$ the ground state N-electron density uniquely determines the potential that gives rise to it. The total electronic energy is a unique functional of the density $\rho(\mathbf{r})$.[34] Although constructing an accurate approximation to the kinetic exchange correlation functional $G[\rho(\mathbf{r})]$ is a formidable task, it need only be done once because the form of G is independent of the form of $V(\mathbf{r})$. Approximations are required because the functional is not known exactly but these approximations are getting better and better. Developments in exchange correlation functionals have made DFT methods viable alternatives to those of conventional quantum mechanical calculations. DFT combines the promise of accurate results (that is, more accurate than Hartree–Fock level quantum calculations) with cheaper computation (because it scales up to more electrons less steeply than conventional methods that include some electron correlation). Several approximate functionals of the electron density are in common use and are relatively successful in prediction of molecular structure, and are known to yield geometries and energies of at least MP2 quality. The difficulties of calculating magnetic response properties using DFT arise in two major ways. The first is intrinsic to all DFT methods, because only approximate functionals are available and they are deficient in various ways. NMR parameters show up these deficiencies most glaringly because of their

extreme sensitivity to the electron distribution in the immediate vicinity of the nucleus. Second, is that in the presence of a vector potential (when magnetic fields or magnetic moments are present) the functionals of both the current density and the electron density are needed, thus a current density functional theory (CDFT) is the appropriate theory. A current-dependent DFT has been derived by Vignale et al.[35–37] Handy et al. have used this approach.[38] On the other hand, more commonly, a generalization of the Kohn–Sham density functional theory (KSDFT) has been used to obtain magnetic responses using only the functional of the electron density; the current density part of the calculation is not included. This is by far the most commonly used calculation method. It remained to be shown by Grayce and Harris[39,40] that when the magnetic field is produced by a constant applied field and a single nuclear magnetic dipole, the current density is a functional of the electron density. Furthermore, they showed that in the linear response regime, the current density functional depends on the *zero field electron density*. As a consequence, magnetic responses in the linear regime are solely functionals of the electron density in the absence of a magnetic field. Grayce Harris call this the magnetic field density functional theory (BDFT).[39,40] The problem is that in all DFT approaches, only *approximate* functionals are yet available, and the magnetic response DFT approaches, whether CDFT or KSDFT or BDFT, all suffer from this same difficulty.

There are several independent formulations of DFT of shielding. A large number of DFT calculations have been carried out by Malkin, Salahub et al. without including the effects of the current density, using a local density approximation (LDA) in a SOS method and IGLO method of local origins for shielding tensor calculations.[41,42]

In the CDFT of Vignale et al., in addition to the usual exchange correlation functionals of the density that are used for solving electronic structure problems in the absence of a magnetic field, the effects of a current density are included.[35–37] Van Wüllen has derived the coupled perturbation equations for calculating nuclear magnetic shielding tensors using CDFT in both the IGLO and the GIAO method of introducing local gauge origins.[43] Lee, Handy, and Colwell derived equations within the Kohn–Sham formulation of DFT for calculations of nuclear magnetic shielding tensors with GIAO basis functions using CDFT, including the use of a local exchange correlation functional which depends on both the electron density and the paramagnetic current density.[38] To put the various DFT formalisms and calculations in context they applied their working expressions to the systems HF, N_2, CO, F_2, and H_2O. By doing computations using conventional atomic basis functions versus GIAOs basis functions, using various local functionals of the density in popular use, such as the exchange functional of Becke with the correlation functional of Becke–Lee–Yang–Parr (BLYP) and others, with and without including the current-dependent functional proposed by Vignale, Rasolt, and Geldart,[37] with or without the ad hoc correction of Malkin et al.[41] Lee et al.[38] provide some very useful comparisons. They of course found what is already well known, that using GIAOs leads to better results than using conventional (gauge-less) atomic basis functions employing CHF or any DFT method. They found that including the current-dependent functional proposed by Vignale, Rasolt, and Geldart[37] gives only small corrections. They also established that DFT and CDFT methods exhibit general difficulty in describing multiply bonded systems such as N_2 and CO. An important observation is that calculations using local functionals of the density give severely deficient eigenvalues. To overcome this, a more accurate functional must be developed. Since the Malkin correction is to modify the energy denominators, this has the effect of shifting the incorrect eigenvalues already noted above. Indeed, direct comparisons by Lee et al. using various functionals with and without the ad hoc Malkin correction lead to a significant improvement in the CO case. They also found that, unlike in the HF molecule, exchange terms are significant in the CO molecule and the current-density terms are no longer negligibly small. The general conclusions are that the use of local density functionals is a major deficiency and overwhelms the small current density corrections.[38] The best results for CO in the Lee et al. formulation of CDFT/GIAO appear to come from the hybrid B3LYP functional[44,45] combined with the Malkin correction.[41] Thus, in spite of what appears to be a lack of solid theoretical foundation, the ad hoc Malkin correction gives very promising results.

There are several other implementations of DFT in shielding calculations, all of which use only current-independent exchange correlation functionals: the IGLO-based DFT of Malkin and Salahub already mentioned and the GIAO-based DFT calculations introduced by Schreckenbach and Ziegler,[46,47] Pulay et al.[48] and Cheeseman, Trucks, Keith, and Frisch.[49]

In practice, the calculations in the Ziegler DFT/GIAO implementation employ Slater-type orbitals as atomic basis functions (unlike most computations which use Gaussian-type basis functions).[46,47] They have also used the frozen core approximation in some systems.[50] Pulay et al. have developed a DFT/GIAO based on their program system which uses analytic derivative theory (TX90).[48] They derive their DFT/GIAO equations in the density matrix formulation used originally.[8] Compared to the Hartree–Fock case, the only new quantity is the first-order exchange correlation term. In the Pulay implementation, these terms are evaluated by the same Becke

numerical integration scheme they use for the exchange matrix elements themselves. Although in principle this method is identical to that of Schreckenbach and Ziegler, there are major differences in implementation. First they use Gaussian basis functions rather than Slater basis functions. Second, the Pulay implementation does not use fitting functions for the Coulomb and exchange functions. Rather they calculate the Coulomb term exactly, using traditional two-electron integral evaluation, and calculate exchange correlation terms by the numerical integration technique of Becke. Independently, Cheeseman et al. presented their formulation of the DFT/GIAO method and also DFT/CSGT (Continuous set of gauge transformations) (or IGAIM).[49] The latter makes use of the highly successful IGAIM[9] and the more general CSGT[51] method of Keith and Bader (wherein the current is determined through the definition of a CSGT, a separate gauge origin to calculate the current $\mathbf{J}^{(1)}(\boldsymbol{r})$ at each point $\boldsymbol{r}$ in real space). They found the IGAIM results essentially identical to the CSGT. Cheeseman et al. compared the results from the DFT methods with CHF/GIAO and CHF/CSGT calculations of isotropic nuclear magnetic shielding of ^{13}C in a large number of molecules.[49]

How well does DFT predict absolute shielding? DFT accounts for correlation effects implicitly in the exchange correlation functionals used and thus might be expected to give superior results in comparison to CHF calculations for a given GIAO set of basis functions, in those molecules such as CO, N_2, NNO, HCN where electron correlation effects on shielding are important. Absolute shieldings obtained using the gradient-corrected functionals are consistently better than CHF in these molecules,[49] although the improvement is small in some cases. Using a basis set which is sufficient to predict accurate shifts using GIAO/MP2 theory, various DFT functionals consistently predict chemical shifts that are too deshielded compared with experiment. The absolute shielding results are too deshielded by 10–20 ppm for ^{13}C, by 6–40 ppm for ^{15}N, and by 30–40 ppm for ^{17}O in the selected molecules where the absolute shielding results were known. Anisotropies are even worse. The isotropic chemical *shifts* correlate better with experiment, with somewhat smaller average standard deviation in the DFT results in comparison with experiment compared to CHF using GIAOs. The test of any theoretical method for calculating $\boldsymbol{\sigma}$ is the comparison of the results with the benchmark CCSD(T) calculations of Gauss for a set of small molecules at the same fixed geometries. The successful wide applicability of the DFT method for calculating $\boldsymbol{\sigma}$ lies in its applications to molecules with large numbers of electrons, where the accurate CCSD(T) calculations are not feasible and even MP2 level calculations are prohibitively expensive and impractical.

The DFT method has severe limitations for the calculations of spin–spin coupling which are connected to the inability of the presently available exchange correlation functionals (LDA and general gradient approximation) to produce the highly accurate spin densities required to describe properly the FC term for molecules containing atoms lying at the right of the periodic table and containing lone pairs.[52] The $\mathbf{J}^{(OD)}$ term is the easiest to calculate, a straightforward numerical integration in the DFT method because this contribution depends only on the unperturbed ground state density. The calculations of the $\mathbf{J}^{(OP)}$ and $\mathbf{J}^{(FC)}$ terms, just as in the ab initio calculations of $\mathbf{J}$, require the spin-unrestricted approach which is normally applied to open shell systems. The $\mathbf{J}^{(SD)}$ contribution is the most time-consuming and is usually neglected in DFT calculations of $\mathbf{J}$ because it is usually smaller than the error in the $\mathbf{J}^{(FC)}$ calculation by this method. Improved results will require a better exchange correlation functional to describe the spin polarization more precisely. A parameterized functional trained to reproduce a set of gas phase isotropic hyperfine coupling constants in nonsinglet radicals would be a good start.

2.6 Relativistic Calculations

It is well known that relativistic effects are very important in the study of heavy elements. Instead of the Schrödinger equation one has to solve a many electron generalization of the four-component Dirac equation. Fully relativistic calculations are very time-consuming even at the SCF level. Other approaches are based on perturbation theory, starting from a nonrelativistic calculation as zeroth-order approximation. A direct perturbation theory formalism has been proposed by Kutzelnigg in 1990 and recently he and his co-workers have laid out the formalism for stationary direct perturbation theory which is conceptually strictly equivalent to a theory in terms of four-component spinors, but operationally on a two-component or even one-component level, and yet avoiding the singularities that arise in using other approaches.[53] The formulation is a relativistic Hartree–Fock theory for closed shell states describable by a single Slater determinant. The eagerly awaited next step is yet in progress, application of multiple perturbation theory formalism to treat relativistic effects on molecular properties such as shielding. Meanwhile, present relativistic calculations of nuclear magnetic shielding are still very approximate.

The relativistic effects on shielding may be considered in three parts. One is the direct effect of the relativistic contraction of s and p inner shells and the relativistic SCF expansion of d and f shells on the diamagnetic contribution. The contraction of the s and p shells leads to larger values of $\langle r_{kN}{}^{-3}\rangle$ and $\langle r_{kN}{}^{-1}\rangle$. The essentially constant relativistic effect on the diamagnetic contribution

coming from the core electrons can be a large correction but hardly changes as the atom is compared from one molecule to the next. This is of little concern in taking differences, the chemical shifts. Another relativistic effect is the indirect effect of this s and p contraction and d and f expansion within the Hartree–Fock scheme, an effect that varies from one molecule to the next. Both of these relativistic effects are "spin-free" in nature. Third is the effect of introducing the spin–orbit terms. To calculate the additional contributions to the nuclear magnetic shielding associated with spin–orbit interactions, the latter has been approached approximately by a nonrelativistic treatment with the spin–orbit operator added on as a perturbation. These corrections can be large for nuclei whose immediate neighbors are atoms that have large values of spin–orbit coupling constants, e.g. the heavier halogens. Both the spin-free relativistic term and the spin–orbit terms can be important and they can couple with each other, as they do in ^{199}Hg shielding in mercury halides.[54] One way of approximately including relativistic effects is to use relativistic effective core potentials. Nakatsuji et al. have introduced this and other additional approximations.[55] The so-called normal halogen dependence of chemical shifts (increasing shielding upon substitution of neighboring atoms by Cl, Br, I, in that order) which had previously been attributed to relativistic effects,[56] has been accounted for entirely by the spin–orbit contributions centered on the halogen atoms in approximate calculations of shielding of ^{1}H, ^{13}C, ^{29}Si, ^{71}Ga, and ^{115}In nuclei in the respective halides.[55,57,58]

These approximate calculations have established the importance of the spin–orbit terms for shielding of nuclei having Cl, Br and I neighbors. However, calculations of heavy atom shielding are not yet on a sound footing at the level of relativistic theory used here. Furthermore, the number of electrons involved is very large and basis sets used are far from saturated so that it is not yet possible to have good nonrelativistic baseline values against which it may be possible to judge quantitatively that the relativistic approximations used are bridging the gap between nonrelativistic calculations and experiment. When a robust method is used which is consistent in the level of inverse powers of the speed of light used for each term, as is promised by the Kutzelnigg approach,[53,59] for example, coupled with a serious basis set study, this problem will be more rigorously addressed. The second problem which is just as important is the lack of experimental absolute shielding data in the gas phase for such heavy nuclei. This leads to comparisons of theoretical values with solution data where solvent effects may even bring some doubt as to the actual chemical species being observed. The problem is particularly severe with the calculations involving bare anions of In, for example, rather than neutral species.

The spin–spin coupling is itself a purely relativistic phenomenon. However, starting from a relativistic Hamiltonian such as the Dirac–Coulomb–Breit Hamiltonian, and neglecting the small components of the four component functions, the expressions for the spin–spin coupling tensor in the nonrelativistic limit can be derived.[5] Relativistic formulations of **J** have been derived. Aucar and Oddershede have formulated a fully relativistic ab initio theory of the spin–spin coupling in its most general form within the polarization propagator approach.[60] They neglect the Breit interaction to derive the formulas which look very much like the nonrelativistic expression for spin–spin coupling in the propagator approach, except that the elements involve integrals of the full four component wavefunctions. The large component of the relativistic wavefunction is the nonrelativistic wavefunction. One relativistic expression replaces the nonrelativistic $\mathbf{J}^{(OP)}$, $\mathbf{J}^{(SD)}$, and $\mathbf{J}^{(FC)}$ terms. Ramsey's nonrelativistic expressions are obtained in the limit that the speed of light goes to infinity. Kutzelnigg has also developed a relativistic theory of spin–spin coupling.[59] Earlier, Pyykkö derived a relativistic analog to Ramsey's theory, using a relativistic nuclear Zeeman hyperfine Hamiltonian as a perturbation.[11] Implementation of Pyykkö's theory has been limited to one-bond couplings.[61]

3 CALCULATIONS OF NUCLEAR MAGNETIC RESONANCE CHEMICAL SHIFTS

3.1 Comparison of Various Computational Methods Using the Same Set of Test Molecules

Which method would be best to use for calculating NMR chemical shifts? The answer depends on the question being asked. Is the chemical shift to be used to discriminate between two or more proposed chemical structures? Is the goal to verify a particular structure? Is it to determine if the molecule is fluxional or not, if forming a complex or not, on the basis of the NMR chemical shifts? Is it to assign the multitude of peaks observed in a crystalline sample? Sometimes we just want to understand what it is about the electronic structure that gives rise to an observed chemical shift or its temperature dependence. Depending on the accuracy that is required to answer the question being asked, a particular method and level of calculation and a particular size of basis set may be sufficient. It is not always necessary to use the most accurate method and the largest basis set. But first, we will compare methods across the board, using several benchmark molecules, in order to see the level of accuracy that may be expected.

Table 1 Comparison between absolute shielding calculations, all using GIAOs and experiment[a]

Method	^{13}C in CH_4	^{13}C in CO	^{13}C in HCN	^{15}N in NH_3	^{15}N in N_2	^{15}N in HCN
DFT/BLYP[b]	184.33	−15.35			−84.82	
DFT/BLYP[c]	187.5	−17.7		259.2	−87.9	
DFT/BLYP[d]	187.80	−12.27	71.74	259.42	−80.55	−43.47
DFT/BLYP[e]	191.2	−9.3	91.5	262.0	−72.9	8.4
SCF[f]	194.8	−25.5	70.9	262.3	−112.4	−50.7
MBPT(2)[f]	201.0	10.6	87.6	276.5	−41.6	−0.3
MBPT(3)[f]	198.8	−4.2	80.0	270.1	−72.2	−26.2
MBPT(4)[f]	198.6	4.1	84.3	269.9	−60.1	−14.9
MCSCF	198.2[g]	8.22[g]	86.76[h]		−52.2[g]	2.63[h]
CCSD[i]	198.7	0.8	84.1	269.7	−63.9	−16.7
CCSD(T)[i]	198.9	5.6	86.3	270.7	−58.1	−13.6
Expt. σ_0^j	195.1 ± 1.2	1.0 ± 0.9	82.1 ± 0.9	264.5 ± 0.2	−61.6 ± 0.2	−20.4 ± 0.2
Expt. σ_e^k	198.7 ± 1.2	3.2 ± 0.9	84 ± 1	273.3 ± 0.2	−59.6 ± 0.2	−15 ± 1

[a] σ_e values are the ones that should be compared with the calculations. All shielding values are in ppm.
[b] Lee et al.[38]
[c] Cheeseman et al.[49]
[d] Rauhut et al.[48]
[e] Schreckenbach and Ziegler.[46]
[f] Gauss,[17]
[g] Ruud et al.[13]
[h] Barszczewicz et al.[21]
[i] Gauss and Stanton.[20]
[j] These are absolute shielding values σ_0 which are isotropic averages in the gas at the zero-pressure limit. They correspond to the thermal average for an isolated molecule. The error bars are associated with the determination of the absolute scale based on spin rotation constants for specific molecules (the CO molecule for ^{13}C, NH_3 for ^{15}N).
[k] The estimates of the vibrational corrections have been subtracted from σ_0 to obtain the value σ_e (the value for a rigid isolated molecule at its equilibrium molecular geometry) with which calculations are to be compared. See Jameson[62] for the references for experimental data and vibrational corrections.

We present some comparisons of absolute isotropic shieldings calculated using GIAOs in Tables 1 and 2. The calculations are for a fixed geometry, and therefore should be compared with the value for the equilibrium geometry of the molecule, σ_e, whereas the room temperature average value in the limit of zero density, $\sigma_0(300\,K)$, is for a rotating vibrating molecule. Electron correlation effects on the individual components of the tensor differ and so are partly washed out in the isotropic average value (which is one-third the sum of the principal components of the tensor). Nevertheless, these comparisons are revealing. Here we included the various levels of many body perturbation expansion MBPT(2), MBPT(3), MBPT(4) used by Gauss,[17] the coupled cluster approach for the singles and doubles approximation (CCSD)[19] and augmented by a correction for triple excitations (CCSD(T))[20] by Gauss and Stanton, a FCI calculation for BH molecule by Gauss and Ruud,[22] and the MCSCF results using GIAOs from Ruud et al.[13,21] We also include the results from DFT using the exchange functional of Becke with the correlation functional of Lee–Yang–Parr DFT/BLYP from calculations using the different implementations of Lee et al.[38] Cheeseman et al.[49] Pulay et al.[48] and Schreckenbach and Ziegler.[46]

Examination of Tables 1 and 2 permit the following conclusions. For the hydrides HF, H_2O, NH_3 and CH_4, electron correlation effects described by triple excitations are small, amounting to less than 1 ppm for the nonhydrogen nuclei. The effects for proton shieldings are not shown in the tables, but they are even smaller, of the order of 0.1 ppm. MBPT(4), CCSD and MCSCF all provide an adequate treatment of electron correlation effects for these simple systems. Furthermore, the agreement with experiment is very good. It has been suggested that these calculations are good enough to be able to say that the ^{17}O shielding scale (which has an error bar of ±17.2 ppm) may actually be closer to the less shielded edge of the error limits. It has been found in systematic studies of a large number of ^{13}C chemical shifts that MBPT(2)-level results are much closer to experiment than CCSD results.[64] It appears that MBPT(2) benefits from a fortuitous but consistent error cancellation, while CCSD (which is theoretically more complete and is in principle a more reliable approach) does not. Triplet excitation effects are considerably more important for the multiply bonded systems CO, N_2 and HCN. The magnitude of the triplet excitation corrections (2–6 ppm) for these systems leads to calculated results that are closer to experiment. For the F_2 molecule, inclusion of triple excitation corrections

Table 2 Comparison between absolute shielding calculations, all using GIAOs and experiment[a]

Method	^{17}O in H_2O	^{17}O in CO	^{19}F in HF	^{19}F in F_2	^{1}H in BH	^{11}B in BH
DFT/BLYP[b]	317.86	−77.14	405.05	−271.70		
DFT/BLYP[c]	324.8	−80.7				
DFT/BLYP[d]	326.37	−73.60	410.87	−277.14		
DFT/BLYP[e]	331.5	−68.4	412.5	−282.7		
SCF[f]	328.1	−87.7	413.6	−167.9	24.21	−261.25
MBPT(2)[f]	346.1	−46.5	424.2	−170.0	24.12	−220.67
MBPT(3)[f]	336.7	−68.3	417.8	−176.9	24.14	−201.92
MBPT(4)[f]	337.5	−52.0	418.7	−174.0	24.22	−184.18
MCSCF[g]	335.3	−38.92	419.6	−136.6		−174.83[h]
CCSD[i]	336.9	−56.0	418.1	−171.1	24.74	−166.64
CCSD(T)[i]	337.9	−52.9	418.6	−186.5	24.62	−170.46
FCI[j]					24.60	−170.08
Expt. σ_0^k	344.0 ± 17.2	-42.3 ± 17.2	410.0 ± 6	-232.8 ± 6		
Expt. σ_e^l	357.6 ± 17	-36.7 ± 17	419.7 ± 6	-192.8 ± 6		

[a] σ_e values are the ones that should be compared with the calculations. All shielding values are in ppm.
[b] Lee et al.[38]
[c] Cheeseman et al.[49]
[d] Rauhut et al.[48]
[e] Schreckenbach and Ziegler.[46]
[f] Gauss.[17]
[g] Ruud et al.[13]
[h] van Wüllen.[63]
[i] Gauss and Stanton.[20]
[j] Gauss and Ruud[22]
[k] These are absolute shielding values σ_0 which are isotropic averages in the gas at the zero-pressure limit. They correspond to the thermal average for an isolated molecule. The error bars are associated with the determination of the absolute scale based on spin rotation constants for specific molecules (the CO molecule for ^{13}C, NH_3 for ^{15}N).
[l] The estimates of the vibrational corrections have been subtracted from σ_0 to obtain the value σ_e (the value for a rigid isolated molecule at its equilibrium molecular geometry) with which calculations are to be compared. See Jameson[62] for the references for experimental data and vibrational corrections.

leads to a change of about 15 ppm and brings the calculated value closer to experiment. Results for F_2 at lower levels of calculation do not provide satisfactory agreement with experiment. Except for the F_2 molecule, GIAO/MCSCF calculations using very large active spaces (only those are shown in Tables 1 and 2) provide results comparable to CCSD. It has been found, and is obvious in Tables 1 and 2, that the DFT method consistently overestimates the paramagnetic term leading to too much deshielding for these benchmark molecules. The SCF value is good enough for CH_4, NH_3, and HF molecules to agree reasonably with the thermal average value (since the neglect of electron correlation effects in these and most molecules is compensated for by the neglect of rovibrational averaging), whereas this level of theory is clearly inadequate for the multiply bonded CO, HCN, and N_2, and also for H_2O and F_2.

The benchmark test molecules shown in Tables 1 and 2, except for CH_4, are specifically chosen as examples that presented problems of electron correlation, especially in ^{15}N, ^{17}O, and ^{19}F shielding. Observe in Tables 1 and 2 the slow convergence in some molecules, faster in others, of the series SCF, MBPT(2), MBPT(3), MBPT(4). Observe also the consistent improvement over SCF afforded by the approximate exchange correlation functionals used in DFT calculations especially for CO, N_2, HCN. Observe also how close CCSD(T) results come to the FCI (in BH molecule). More typical of the applications of calculated NMR chemical shifts to analysis of mixtures are calculations of ^{13}C chemical shifts. Table 3 demonstrates the importance of electron correlation to ^{13}C chemical shifts in the comparison with chemical shifts in the gas phase at the low density limit. It can be seen that the second-order electron correlation generally brings the calculations close enough to experiment to be useful for analysis.

3.2 Comparison of Carbon Chemical Shift Tensor Components with Calculations

A more stringent test of the calculations has to do with reproducing the elements of the shielding tensor, not just the isotropic average that is obtained in solution or in a MAS experiment in the solid state. In a single crystal study of a sugar, for example, there are a large number of peaks which have to be assigned in order to verify the structure. Complete assignment of ^{13}C shielding tensors in the entire molecule from single crystal studies has

Table 3 Calculated ^{13}C chemical shifts, in ppm relative to $^{13}CH_4$, and experimental values in the gas phase

A	$\sigma(CH_4) - \sigma(A)$, Calculated		Expt.[b]
	SCF[a]	MBPT(2)[a]	
CH_3CH_3	11.7	13.5	14.2
$H_2C{=}CH_2$	135.8	130.3	130.6
$HC{\equiv}CH$	81.8	78.2	77.9
CH_3F	71.6	79.7	78.3
CH_3OH	52.0	59.3	58.5
CH_3NH_2	31.9	36.6	36.8
CH_3CHO	33.5	38.7	37.9
$(CH_3)_2CO$	32.2	37.0	37.1
CH_3CN	4.8	7.9	7.4
CO	224.9	190.4	194.1
CO_2	147.9	138.0	136.3
H_2CO	205.0	194.8	
CH_3CHO	211.3	200.3	201.8
$(CH_3)_2CO$	218.8	207.3	208.2
HCN	127.5	114.2	113.0
CH_3CN	135.1	125.4	121.3
$CH_2{=}C{=}CH_2$	240.0	227.5	224.4
$CH_2{=}C{=}CH_2$	81.7	80.6	79.9
CF_4	116.4	137.1	130.6
C_6H_6	140.6	137.5	137.9

[a] Gauss.[16]
[b] Jameson and Jameson.[65]

been developed to the highest level by Grant et al. The multiple axis sample reorientation mechanism developed in this group permits the study of crystals containing 50–100 magnetically different nuclei per unit cell. In a polycrystalline solid with a very large number of distinct ^{13}C chemical sites, it is possible, using multidimensional NMR techniques to obtain the individual shielding tensor elements for each isotropic peak in the MAS NMR spectrum. To assign all these, ab initio calculations of shielding tensor elements are indispensable.

How well do calculations predict the tensor elements? It is important to be able to do these calculations in a relatively routine manner (one cannot use CCSD level of calculations) so that fitting to the observed spectra can be done expeditiously. Otherwise, theoretical calculations would not be practically useful for analysis. The group of D.M. Grant has carried out the largest number of such analyses.[66] Single crystal NMR experiments produce a complete description of the shielding tensor with six independent components specifying the tensor in a fixed crystallographic coordinate system (the so-called icosahedral tensor representation). Figure 1 shows the degree of success of SCF level calculations using a modest size basis set.[67] The high level of agreement between calculated and experimental tensors for ^{13}C is such that only the structural parameters (bond distances and angles) limit the level of agreement. This means that ab initio calculations and measurements together can be used to address certain fine details of solid-state structure, surpassing the accuracy of X-ray data.[68] This is possible because the shielding tensor is exquisitely sensitive to bond distances.

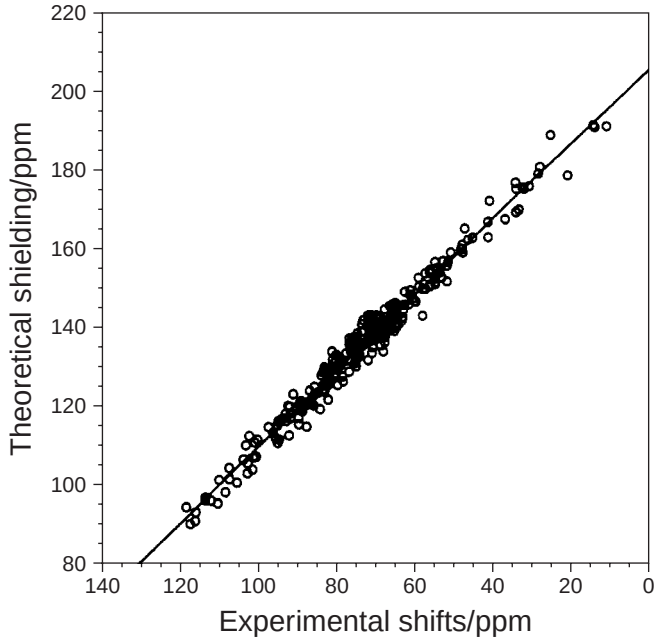

Figure 1 Comparison of computed full ^{13}C shielding tensors versus experimental values of icosahedral shift from TMS obtained from single crystal measurements. Both are expressed in the icosahedral representation that includes the principal axis system orientation. (Reproduced by permission from F. Liu, C.G. Phung, D.W. Alderman, D.M. Grant, *J. Am. Chem. Soc.*, **118**, 10629–10634 (1996). Copyright (1996) American Chemical Society.[67])

It is quite important to be able to predict theoretically the individual tensor components of the building blocks of proteins, in order to establish that it is possible to use NMR chemical shifts in the determination of protein structure. In Figure 2 is the demonstration of the degree of success of SCF level calculations using a modest basis set for the tensor elements of ^{13}C in a single crystal of threonine.[69]

3.3 Other First Row Nuclei

A systematic study of B, N, O, and F shielding using the IGLO method of distributed origins provides a measure of the initial successes of theoretical calculations for these nuclei in systems of known structure.[7] ^{11}B NMR chemical shift calculations have been used for the analyses of new boron compounds, which are particularly useful when more than one structure can fit the electron diffraction data. Shielding calculations for N, O, and F in most molecules do require a theoretical treatment including electron correlation. Ab initio calculations for

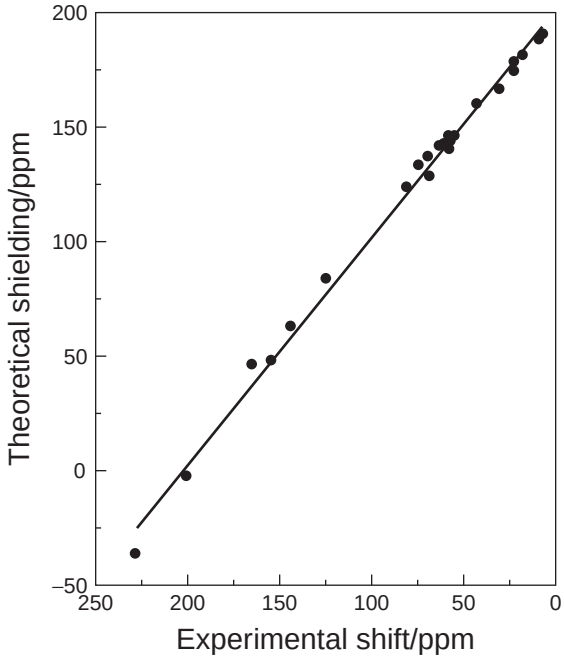

Figure 2 Comparison of computed full ^{13}C shielding tensors versus experimental values obtained from single crystal measurements for threonine. Both are expressed in the icosahedral representation that includes the principal axis system orientation. (Reproduced by permission from A.C. de Dios, D.D. Laws, E. Oldfield, *J. Am. Chem. Soc.*, **116**, 7784–7786 (1984). Copyright (1994) American Chemical Society.[69])

^{17}O are of sufficiently high quality to indicate that the absolute shielding for ^{17}O in CO (used to define the experimental absolute shielding scale) is very likely at the lower edge of the reported error bars. DFT and ab initio methods have been used for ^{17}O shielding calculations in carbonyl complexes of transition metals (Ti, Zr, Hf, Fe, Rh, Cr, Mo, and W) by Kaupp et al.[70,71] and also by Ziegler et al.[72] In using ^{17}O experiments to validate the calculations, the anisotropy of the tensor $(\sigma_{\parallel} - \sigma_{\perp})$ contains none of the uncertainty of the absolute shielding scale for ^{17}O and gives a good measure of the predictive success of ^{17}O calculations. Using this criterion, the DFT results agree with one of the crystalline sites in the symmetrical carbonyls $Cr(CO)_6$, and $Mo(CO)_6$, and differ by at least 10–20 ppm from the solid-state experiments in $W(CO)_6$.

3.4 Second Row Nuclei

Systematic studies of the shielding of nuclei ^{29}Si, ^{31}P, ^{33}S, and ^{35}Cl using IGLO distributed origins in a wide variety of molecules have been compared against chemical shifts, leading to reasonably good straight lines.[7] DFT calculations of ^{35}Cl shielding in XCl_4, where X = C, Si, Ge, Sn, and Ti gave good agreement with experimental chemical shifts observed in solution,[73] however, the calculations were not validated against any of the available absolute shielding, such as in ClF ($\sigma = -516 \pm 23$ ppm) or HCl ($\sigma = 952$ ppm). GIAO calculations including second-order MP2 level electron correlation for ^{31}P has been used to estimate the infinite-order results, and these agree very well with the absolute shieldings that are known for molecules ranging from PN to P_4 a range of about 900 ppm.[74]

3.5 Other Heavy Nuclei

Heavy nuclei present some general problems for calculations of NMR shielding. First, the larger number of electrons requires larger numbers of basis functions and including electron correlation becomes very expensive. Second, relativistic effects could be important. Third, in some cases there are few if any gas-phase data that can be used to test the absolute shieldings from the calculations; in many cases, only solution-phase experimental data is available for comparison with calculations (e.g. ^{69}Ga, ^{115}In, ^{75}As, ^{121}Sb, ^{129}Xe). ^{77}Se is an exception to this third difficulty and can be used to explore the range of problems associated with heavy nuclei in general.

Correlation effects must be included for a quantitative description of ^{77}Se chemical shifts in those same bonding situations where ^{17}O shielding has been found to require correction for correlation effects. The additional complication of the large number of electrons therefore makes the ^{77}Se calculations more challenging. The results of ab initio calculations are very good. At the SCF level, for example, various calculations for the isotropic shielding of H_2Se lead to 2167.6, 2170, and 2171 ppm, which are very close to each other and reasonably close to the value calculated at the CCSD level (2213 ppm) and the experimental value: 2101 ± 64 ppm on the absolute shielding scale[75] (without the relativistic corrections for the diamagnetic shielding of the free atom). The electron correlation effects are only 2% of the total shielding in H_2Se and O=C=Se; they are 7% of the total shielding in Se=C=Se. This is very encouraging. Table 4 shows only the highest level ab initio calculations compared with DFT calculations and experiment. Keeping in mind that the rovibrational corrections are about −60 ppm, (that is, σ_e may be converted into experimental σ_0 by adding −60 ppm) the CCSD values are within 3–5% of the experimental values. The DFT results are less shielded than the CCSD values by 100–200 ppm.

The situation for ^{125}Te is comparable to that in ^{77}Se. ^{125}Te in TeF_6 gas has an absolute shielding of 2570 ± 130 ppm in the zero density limit, if the nonrelativistic diamagnetic shielding of the free atom is used. The DFT nonrelativistic calculations give 2260 ppm,[78] which is 200–300 ppm less shielded. This is in the same direction

Table 4 Calculated ^{77}Se shielding compared with experimental absolute shielding values in the gas phase (in ppm)

Molecule	$H_2C{=}Se$	Me_2Se	H_2Se	$O{=}C{=}Se$	$Se{=}C{=}Se$
CCSD σ_e^a	−741	1877.5	2213	2345	1596
DFT σ_e^b		1668	2093	2270	1441
Expt. σ_0^c	-900 ± 200	1756 ± 64	2101 ± 64	2348 ± 60	1610 ± 80

[a] Bühl et al.[76]
[b] Schreckenbach et al.[50]
[c] See Jameson[77] for the original sources of the experimental data.

as the difference between DFT and CCSD in the ^{77}Se case. The ability of DFT calculations to reproduce the full range of ^{125}Te chemical shifts in all types of chemical bonding situation is very encouraging.[78]

3.6 Transition Metal Nuclei

For transition and post-transition metal nuclei there are the usual problems associated with heavy nuclei, plus the lack of absolute shielding information. With one exception, there are no gas-phase data to establish absolute shielding scales. Despite these problems some DFT calculations have been done for ^{103}Rh, ^{91}Zr, ^{57}Fe- and ^{59}Co by Bühl and others. A relativistic formulation in which spin orbit contributions are neglected, resulting in the so-called scalar relativistic method, is being explored within the DFT/GIAO method.[79] In calculations for other transition and post-transition metal nuclei for which the absolute shielding is not known, only shielding differences, i.e. chemical shifts, have been used to compare with experiment. The agreement is not yet at the level that is achievable routinely with ^{13}C shielding calculations. The range of transition metal shifts is usually very large and these exhibit useful diagnostic variations with ligand types. The theoretical calculations have yet to catch up with experiment. Witness, for example, the one case where gas-phase data are available: $CdMe_2$ molecule. Beam measurements show neat $CdMe_2$ liquid being deshielded by 1746 ppm from the free Cd atom, which has an absolute shielding of 4813 ppm. Thus $\sigma = 3067$ ppm for neat $CdMe_2$ liquid at room temperature. The gas is found to be 62.1 ppm unusually less shielded than the neat liquid,[80] so that $\sigma = 3005$ ppm for gaseous $CdMe_2$ at 97 °C. This is to be compared with 3504.5 ppm (too shielded by 500 ppm) from GIAO Cd shielding calculations on an isolated molecule of $CdMe_2$ with a spread of 105 ppm depending on basis set used, neglecting relativistic corrections or electron correlation.[81] Other calculations give more shielded values than this, as much as 900 ppm more shielded than 3005 ppm. Since the chemical shift range of Cd is about 900 ppm, the level of accuracy needs considerable improvement.

4 CALCULATIONS OF SPIN–SPIN COUPLING CONSTANTS

The various mechanisms in the nonrelativistic limit, given by Equations (10–14) are $\mathbf{J}^{(OD)}$, $\mathbf{J}^{(OP)}$, $\mathbf{J}^{(FC)}$, $\mathbf{J}^{(SD)}$, and the cross-term $\mathbf{J}^{(SDFC)}$ which has no isotropic part. The OD term, $\mathbf{J}^{(OD)}$, is the only term that is not expressed as a SOS in the Ramsey formulation, rather it is an average value of an operator containing two nuclear spins. $\mathbf{J}^{(OD)}$ is not usually small and can be rather large for ^2J(HH). A systematic study of this term shows that it is not very sensitive to basis set choice (double zeta with polarizations functions are sufficient) and to inclusion of electronic correlation (SCF average values will do); it is particularly important for ^nJ(HH), independent of n.[82] The sign of the contribution is negative for two-bond HH coupling. From a systematic study of the OP mechanism using DFT, $\mathbf{J}^{(OP)}$ appears to be significant for most couplings although not dominant, and is particularly important for couplings involving a nucleus with lone pairs.[83] The sign of the contribution (reduced so as to not include the nuclear gamma values) can be positive or negative; $\mathbf{J}^{(OP)}$ is negative and is the largest contribution for CO and N_2 molecules, for example. The SD term $\mathbf{J}^{(SD)}$ is the most time-consuming to calculate and so is sometimes neglected; it is not small when multiple bonds are involved between the coupled nuclei. For example, for N_2 molecule it is comparable to and partly cancels the $\mathbf{J}^{(FC)}$ term.[25,27,84] Electron correlation is very important for multiple bonded systems and must be accounted for to obtain reliable results: results at the RPA level have the wrong sign and magnitudes for coupling in both CO and N_2 molecules. The sign and magnitude of the FC, $\mathbf{J}^{(FC)}$, term varies across the periodic table. Where no multiple bonds are involved, this mechanism usually provides the largest contribution to one-bond coupling constants. Electron correlation is very important for this mechanism and unrealistic values may result from calculations at the RPA level.

4.1 One-bond Coupling Constants

Table 5 shows the various contributions to the one-bond couplings in HF, HCl, CO, and N_2. The uncorrelated

Table 5 One-bond spin–spin coupling constants (Hz)

Molecule	Method	Ref.	$\mathbf{J}^{(FC)}$	$\mathbf{J}^{(OP)}$	$\mathbf{J}^{(OD)}$	$\mathbf{J}^{(SD)}$	**J**	**J**, Expt.
HF	RPA	85	467.3	119.3	−0.1	−12.4	654.1	
	SOPPA	85	338.3	195.7	−0.1	−1.0	532.9	
	MBPT	86	390.71	195.14	1.69	−17.47	570.01	
	CCSD/PPA[a]	85	329.4	195.7	−0.1	−0.6	524.4	
	DFT	87	198.1	198.0	0.1		396.2	
	EOM/CCSD	27	338.2	176.2	0.0	−1.0	513.4	529 ± 23
$H^{35}Cl$	RPA	27	16.78	13.70	0.00	−0.45	30.03	
	MBPT	86	12.52	12.02	0.00	−0.08	24.45	
	EOM/CCSD	27	22.04	12.65	0.00	0.34	35.03	37.7
$^{13}C^{17}O$	RPA	25	−8.1	12.2	0.1	−9.3	−5.1	
	SOPPA	25	7.3	14.8	0.0	−4.0	18.1	
	MCLR	84	6.69	13.66	0.09	−4.33	16.11	
	DFT	87	13.4	12.4	0.1		25.9	
	EOM/CCSD	27	7.0	13.0	0.1	−4.6	15.5	16.4 ± 0.1
$^{14}N^{15}N$	RPA	25	−7.65	0.50	0.0	−8.13	−15.26	
	SOPPA	25	0.45	3.25	0.0	−1.55	2.18	
	MCLR	84	−0.23	2.83	0.02	−1.85	0.77	
	DFT	87	2.0	2.7	0.0		4.7	
	EOM/CCSD	27	0.3	2.8	0.02	−1.7	1.4	1.8 ± 0.6

[a] PPA, polarization propagator approximation.

results (RPA) are shown to be inadequate. Various methods of including electron correlation are (1) through the polarization propagator approach to SOPPA and CCSD level, (2) MBPT, (3) MCLR theory, (4) DFT, and (5) EOM/CCSD. The DFT calculations suffer from the inadequacy of the approximate exchange correlation functionals available. These functionals may be good enough to reproduce binding energies, but offer less accurate descriptions of the electron spin distributions where they are needed in calculations of **J**, especially the FC mechanism. MCLR theory uses an MCSCF reference state and is capable of describing electronic systems with large static correlation effects.

How well do calculations predict the simple one-bond ^1J(CH)? This may be observed in Table 6, where it is seen that the isotropic value is entirely dominated by the FC mechanism and is easily reproduced by calculations that include correlation, including DFT. Correlation effects can be substantial. For example, the uncorrelated calculation of the term FC for ^1J(CH) in HCCH molecule leads to 449.3 Hz, whereas SOPPA which includes correlation up to second order gives

Table 6 One-bond CH coupling constants, ^1J(CH) (Hz)

Molecule	Method	Ref.	$\mathbf{J}^{(FC)a}$	**J**	**J**, Expt.
CH_4	CCSD/PPA	26	122.12	123.87	120.78 ± 0.05[b]
C_2H_4	EOM/CCSD	88	145.63	147.66	156.4
HCCH	SOPPA	24	246.5	246.5	248.7
CH_3F	EOM/CCSD	88	136.44	137.89	149.1
CH_3CN	EOM/CCSD	88	123.29	125.69	136.0
CH_3NH_2	EOM/CCSD	88	120.88[c]	123.29[c]	132.5
Cyclopropene	MCSCF	89	163.6	164.7	167
Cyclopropene	MCSCF	89	212.7	213.1	226
Cubane C_8H_8	EOM/HRPA	90	145.62	146.45	154.5
CH_4	DFT	87	122.0	123.9	120.78 ± 0.05[b]
C_2H_4	DFT	42		153.9	156.4
CH_3CH_3	DFT	42		123.87	124.9
HCCH	DFT	42		250.8	248.7

[a] The FC contribution only.
[b] This value, corresponding to the equilibrium structure, is obtained after the experimental value is corrected for rovibrational averaging. All other experimental values are uncorrected thermal averages at room temperature.
[c] Average values.

Table 7 One-bond CF coupling constants, ^1J(CF) (Hz)

Molecule	Method	Ref.	Calculated	Expt.
CH_3F	EOM/CCSD	88	−172.37	−157.5 ± 0.2
CH_3F	DFT	52	−268.12	−157.5
CH_2F_2	DFT	52	−343.11	−234.8
CF_4	DFT	52	−379.37	−259.2
CH_3CF_3	DFT	52	−379.06	−271
CHF_3	DFT	52	−390.72	−274.3
$CHCl_2F$	DFT	52	−388.50	−293.8
CF_3Cl	DFT	52	−415.33	−299
F_2CO	DFT	52	−426.22	−308.4
$CFCl_3$	DFT	52	−448.83	−337
FC(O)H	DFT	52	−455.01	−369
F_2CSe	DFT	52	−510.87	−408

246.5 Hz, which agrees quite well with the experimental value of 248.7 Hz.

On the other hand, calculations are less successful with the one-bond ^1J(CF). DFT is found to underestimate the FC contribution to the one-bond coupling because of the inability of such exchange correlation functionals to produce the accurate spin densities required for this calculation. In the presence of polarizable lone pairs the correlation problem is more severe, and the available functionals do not describe the spin densities well enough in the case of ^1J(CF) which are predicted to be about 110 Hz away from experimental values, as shown in Table 7.

There are many interesting trends observed in coupling constants, in signs, magnitudes, dependence on substituents, stereochemistry, position of coupled nuclei in the periodic table, and so on.[91] Many of these trends have been very useful in analysis of spectra and yet a sound theoretical basis for few of such trends has been established.

The anisotropy of the tensors calculated with and without electron correlation are shown in Table 8 for HF and HCl.[86] First of all, note that the FC mechanism is purely isotropic and the cross-term $\mathbf{J}^{(SDFC)}$ which has no isotropic part, is responsible for a large part of the total anisotropy of the tensor. The anisotropy of the orbital mechanisms are opposite in sign and partly canceling. The contribution to the anisotropy from the SD mechanism is small. Any anisotropy observed in the **J** tensor in oriented molecules has to come from the mechanisms other than the FC term. However, because of the very large contribution from the cross-term $\mathbf{J}^{(SDFC)}$ to the anisotropy (78% in HF and 93% in HCl), the magnitude of the measured anisotropy unfortunately conveys very little information about the magnitude of the contributions of mechanisms other than the FC term to the isotropic average observed in solution. The effect of electron correlation on the individual components of $\mathbf{J}^{(OD)}$ is small. (It is well known that the effect of correlation on the isotropic average of $\mathbf{J}^{(OD)}$ is small and that it is not very sensitive to basis set choice.) The orbital mechanisms have opposite contributions to the anisotropy, $\mathbf{J}_{\parallel}^{(OD)}$ and $\mathbf{J}_{\perp}^{(OP)}$ are similar in sign (positive) and magnitude (large), and so are $\mathbf{J}_{\parallel}^{(OP)}$ and $\mathbf{J}_{\perp}^{(OD)}$ (negative and smaller). The effect of electron correlation on the cross-term $\mathbf{J}^{(SDFC)}$ is about 5%. If this is typical, uncorrelated calculations should permit estimation of the **J** anisotropy that may be expected in oriented systems. There are only a few measurements of the anisotropy of the **J** tensor because the observable quantity in solids is the $(\mathbf{D} + \mathbf{J})$ tensor, and the direct dipolar coupling tensor **D** overwhelms the sum. The anisotropy of the **J** tensor has been determined in a few favorable cases, such as $^1J(^{31}P{-}X)$, where $X = {}^{199}Hg$, ^{95}Pt, ^{115}In, in Wasylishen's laboratory.[92] A typical such measurement in a single crystal of a mercury phosphine complex shows the experimental technique for arriving at $J_{\parallel} = 11800$ Hz, $J_{\perp} = 6400$ Hz, and the isotropic value is 8200 Hz. While the isotropic value is very likely to be dominated by the FC mechanism, the

Table 8 Contributions to the calculated anisotropy of the one-bond coupling, $\Delta\mathbf{J} = (\mathbf{J}_{\parallel} - \mathbf{J}_{\perp})$ (Hz)[a]

		$\mathbf{J}^{(FC)}$	$\mathbf{J}^{(OD)}$	$\mathbf{J}^{(OP)}$	$\mathbf{J}^{(SD)}$	$\mathbf{J}^{(SDFC)}$	Total
HF,	$\mathbf{J}_{\parallel}$	453.44	143.50	−11.34	−71.96	−392.22	121.41
SCF	$\mathbf{J}_{\perp}$	453.44	−69.43	297.82	0.44	196.11	878.38
	$\Delta\mathbf{J}$	0	212.93	−309.16	−72.4	−588.33	−756.97
HF,	$\mathbf{J}_{\parallel}$	390.71	143.12	−8.92	−58.78	−373.40	92.73
MBPT(2)	$\mathbf{J}_{\perp}$	390.71	−69.03	297.17	3.18	186.63	808.65
	$\Delta\mathbf{J}$	0	212.15	−306.09	−61.96	−560.03	−715.92
HCl,	$\mathbf{J}_{\parallel}$	25.17	13.19	−2.90	−2.33	−42.26	−9.14
SCF	$\mathbf{J}_{\perp}$	25.17	−6.60	18.95	0.71	21.13	59.36
	$\Delta\mathbf{J}$	0	19.79	−21.85	−3.04	−63.39	−68.50
HCl,	$\mathbf{J}_{\parallel}$	12.52	13.19	−2.52	−1.84	−41.38	−20.03
MBPT(2)	$\mathbf{J}_{\perp}$	12.52	−6.59	19.29	0.80	20.68	46.70
	$\Delta\mathbf{J}$	0	19.78	−21.81	−2.64	−62.06	−66.73

[a] Fukui et al.[86]

anisotropy $\Delta\mathbf{J} = 5400$ Hz comes entirely from the non-FC mechanisms.[92]

4.2 The Two-bond Coupling Constant

The geminal coupling constant ^2J(HH) turns out to be very difficult to predict. As is the case for all ^nJ(HH), the $\mathbf{J}^{(OD)}$ term is important. So also is the $\mathbf{J}^{(OP)}$ term, but it has the opposite sign to the $\mathbf{J}^{(OD)}$ term. For the series CH_4, SiH_4, GeH_4, SnH_4, the orbital mechanism $J^{(OD)}$ and $J^{(OP)}$ terms have opposite signs and they very nearly cancel in CH_4. The magnitude of the $\mathbf{J}^{(FC)}$ term varies from large negative to large positive. There is poor agreement of the total calculated value with experiment.[31] The experimental variation of ^2J(HH) with the nature of the intervening atom is not predicted quantitatively, although the trend of algebraically increasing from C to Sn is reproduced at every level of correlation treatment.[31,52] In the series CH_4, NH_3, OH_2, ^2J(HH) has the sign of the $\mathbf{J}^{(FC)}$ term, but is by no means dominated by it. Here too, the experimental variation of ^2J(HH) with the position of the intervening atom in the periodic table is not predicted quantitatively, although the trend of algebraically increasing from C to N to O is reproduced at every level of correlation treatment.[27]

4.3 Coupling Over Three Bonds

From a practical viewpoint, one of the very early major successes of theoretical calculations of spin–spin couplings is the prediction of the torsion angle dependence of ^3J(HCCH), known as the Karplus equation. The very simple valence bond calculation[93] using a small four-atom fragment (HCCH) led to an unequivocal prediction which permitted a practical determination of structure strictly from the observed isotropic value of the coupling constant. It was found later that the dihedral angle dependence of the three-bond coupling is general and Karplus-type equations have been used to describe many types of three-bond coupling pathways, for example ^3J(X–Y–C–H), where X represents other nuclei such as ^{31}P or ^{13}C or ^{15}N, and three-bond coupling paths such as PtCCC, PWNN, PCPSe, etc. Used with caution, experimental ^3J values and a Karplus equation make a reasonable conformational probe. The original Karplus equation is written in the form of Equation (19)

$$^3J(\mathrm{HCCH}) = C_0 + C_1\cos\phi + C_2\cos(2\phi) \qquad (19)$$

with $C_0 = 8.02$, $C_1 = -1.2$, and $C_2 = 7.0$ Hz as the empirical parameters, although other forms have also been used. The coefficients in the above equation have been calculated by various methods using ethane as the model. The $\mathbf{J}^{(FC)}$ contribution is the largest and $\mathbf{J}^{(SD)}$ the smallest. One such calculation, with second-order correlation for all contributions except the FC contribution (which was done with third-order correlation), leads to $C_0 = 4.66$, $C_1 = 0.39$, and $C_2 = 5.78$ Hz.[94] In the general case, the ^3J value also depends on the bond angles between any two adjacent bonds out of the three, and there are substituent effects.

4.4 Relativistic Effects

Why are relativistic effects important for spin–spin couplings of heavy nuclei? Relativistic effects are particularly important for electronic properties which depend on the electronic wavefunctions very near nuclei where electrons move fast. Relativistic effects on the electronic structure of atoms and molecules consist of a contraction of s and p shells, the spin–orbit splitting of the non-s shells, and the relativistic SCF expansion of d and f shells. The contraction of the s and p shells leads to larger spin densities at the nuclei (FC term) and also larger values of $\langle r_{k\mathrm{N}}{}^{-3}\rangle$ (other mechanisms). An a posteriori correction of the nonrelativistic values of these by a multiplicative factor $B(n, Z)$, depending on the principal quantum number n and the nuclear charge Z, was suggested by Breit in 1930[95] and this multiplicative factor has been used by Pyykkö and others to impose a simple relativistic correction on the values of J calculated using the nonrelativistic formulas.[96] This factor, $B(n, Z)$, is 1.348 for the $n = 5$ shell of Sn and is 2.592 for the $n = 6$ shell of Pb.[96] That is, the nonrelativistic calculations underestimate the value of J(SnH) by a factor of 1.348. When both nuclei involved in the coupling are heavy, the product of two such factors is substantial.

5 CALCULATIONS OF ELECTRIC FIELD GRADIENTS

5.1 Electric Field Gradient Tensor Versus Electronic Structure in the Solid State

The electric field gradient tensor is intimately related to the local molecular structure. In crystalline silicates, for example, the measured ^{17}O nuclear quadrupole coupling constant serves as a probe of oxygen coordination number and geometry. Using experimental correlations between structure (Si–O–Si bond angles, for example) in crystalline silicates and the measured ^{17}O quadrupolar coupling constants, the Si–O–Si bond angle distribution in silicate glasses can be deduced, bridging and nonbridging oxygens can be distinguished.[97] Electric field gradient tensors of deuterium nuclei in hydrogen bonded positions, such as the amide or carboxy hydrogen in peptides, give deuteron/proton bond directions with an accuracy rivaled only by neutron diffraction, since it has been established that the unique eigenvector of a deuteron quadrupole

coupling tensor is approximately parallel to the bond direction of the deuteron.[98]

5.2 Calculations of Electric Field Gradients at Nuclei in Isolated Small Molecules

The most reliable method of obtaining the intrinsic electric quadrupole moment of a nucleus is by very high quality ab initio calculations of the electric field gradient tensor in selected small molecules in which the nuclear quadrupole couplings of the nucleus have been measured accurately via microwave spectroscopy. The value of eQ is obtained as a fit parameter.[99] Once calibrated, this eQ value can be used to deduce from experiment the electric field gradient tensor for any other nuclear site.

Accurate theoretical calculations of electric field gradients for small molecules pose no special problems; requirements of basis set quality and appropriate level of electron correlation depend on the molecule, just as for calculations of shielding, but less demanding in that only the ground electronic wavefunction is required. Just as for shielding calculations, the r^{-3} factor in the electric field gradient requires wavefunctions that are accurate in the immediate vicinity of the nucleus.

5.3 Simulations of Nuclear Quadrupole Coupling in Associated Liquids

The presence of neighboring molecules influences the electric field gradient at a nuclear site, by directly providing additional charge distributions outside of the molecule and also by distorting the electronic distribution of the molecule of interest. An extreme case is a liquid in which hydrogen bonding or complex formation is present. One approach to the calculation is to consider the liquid as having a distribution of clusters of all sizes, monomers, dimers, n-mers where n is truncated at some value when the contribution to the average value is sufficiently small. Molecular geometries of each n-mer are optimized and the electric field gradients are calculated at each nuclear site in the n-mer. Molecular partition functions are calculated for each n-mer, and from thermodynamic calculations the distributions of the n-mers are obtained. The average electric field gradient for each cluster is weighted with the cluster distribution to obtain the electric field gradient values in the liquid phase. The ^{14}N, ^{17}O, and ^{2}H of the carbonyl and *cis* and *trans* amides have been calculated in liquid formamide by this method, for comparison with the experimental values of NMR quadrupolar relaxation time as a function of temperature.[100] Cyclic hexamers are found to be the dominant species at room temperature, consistent with structural data from neutron diffraction, low-frequency Raman, and far-infrared spectra. This method of calculation has been applied to liquid HCN, in which the calculated values for the isolated monomer, dimer, and trimer successfully predict the values known independently from pulsed Fourier transform microwave experiments on the van der Waals complexes.[101] Theoretical calculations such as these, combined with measurements of the nuclear quadrupole coupling constants as a function of temperature, can provide a useful general probe of electronic changes accompanying hydrogen bonding, cluster formation, solvation, phase condensation, and other phenomena in condensed media.

5.4 Relation Between Chemical Shift and Electric Field Gradient Tensors

NMR measurements in single crystals permit the independent determination of the principal axis systems of the electric field gradient tensor and the shielding tensor.[102] Even in the powder it may be possible to find the relative orientation of these two axis systems by referring to the known axis system for the dipolar coupling. The two axis systems are not necessarily coincident. Theoretical calculations of both the electric field gradient and the shielding tensors at the same nuclear site provide descriptions of the electronic distribution and chemical bonding which can be checked directly against experiment. They provide respectively, a measure of the bond direction and the strength of the hydrogen bond for the deuterium nucleus, for example. In materials that exhibit a distribution of nuclear sites, such as glasses or polymers, multidimensional solid-state NMR techniques permit the determination of the anisotropic chemical shift as a function of the isotropic chemical shift or of the electric field gradient as a function of the isotropic chemical shift. From such measurements, the anisotropic chemical shift of ^{29}Si and the electric field gradient of ^{17}O nuclei, for example, can both be used to characterize a silicate glass or other complex materials, providing complementary information.[103,104] Thus, these two tensors provide local electronic information even in complex materials. With the assistance of theoretical calculations such multidimensional solid-state NMR experiments can provide answers to questions about the microscopic structure of solids, on the extent of order/disorder in cation environments, random distributions or amorphous/crystalline domains, short-range and long-range order, and so on.

6 INFLUENCE OF INTRAMOLECULAR GEOMETRY AND ENVIRONMENT ON NUCLEAR MAGNETIC RESONANCE PARAMETERS

The temperature dependence, mass dependence (isotope effects), and site sensitivity (dependence on secondary/tertiary structure of proteins, for example) of

chemical shifts, spin–spin couplings, and electric field gradients provide additional information about structure, dynamics, and environment of a molecule or a particular part of a molecule.

6.1 Nuclear Magnetic Resonance Parameter Dependence on Local Geometry: Bond Lengths, Bond Angles, Torsion Angles

The insight into structure and environment provided by the NMR chemical shift is obtained by a combination of theoretical calculations and experiments.[105] The NMR chemical shift discriminates between the various alanine residues in the same protein molecule, between two nuclear sites identical in every way except that one has ^{18}O in a neighboring bond rather than ^{16}O (isotope shift studies), between a ^{15}N (and ^{13}C) in a dynamically averaged rather than rigid headgroup at an oriented membrane interface, for example. In most cases, theoretical calculations using innovative model fragment systems are required to interpret the relation between the structure and the chemical shift. The discrimination is afforded by the extreme sensitivity of the shielding tensor to the local geometry: the bond lengths, bond angles, and torsion angles. The mathematical surface describing the shielding tensor as a function of these geometrical parameters is called a shielding surface. Vibrational averaging over the shielding surface, weighted by the probabilities of finding the molecule at the geometries described by these parameters (the vibrational wavefunction provides these probabilities) gives average shielding values that are different for different isotopomers, gives average shielding values that are different for particular (ϕ, ψ) angles that characterize particular alanine residues in a protein. Thus, isotope effects on chemical shifts can be predicted; the distinguishing chemical shifts of different alanine residues in a protein can be associated with specific local conformations, leading to structure determination. The application of quantum mechanical calculations of shielding surfaces to the structural characterization of proteins was introduced by de Dios, Pearson, and Oldfield.[106] This approach has led to the possibility of secondary and tertiary protein structure determination from NMR chemical shifts in solution using ^{13}C alone.[107] The method is extremely powerful when combined with complementary information obtained from geometry sensitivity of other NMR parameters such as spin–spin coupling and ^{1}H chemical shifts.[108] The use of theoretical calculations of NMR shielding surfaces to elucidate structure and dynamics finds application in the gas phase, in catalysis, as well as in biomolecular systems.[105,109,110]

The dependence of the spin–spin coupling, the electric field gradient and other molecular electronic properties on bond lengths and bond angles, and the observations that are the experimental manifestation of this, are similar to that discussed above for shielding. The general theoretical basis for isotope effects and temperature dependence of these properties is the same. With the assumption of the Born–Oppenheimer separation of nuclear motion from electronic motion, the mathematical surface describing the dependence of **J** (or other) on the geometrical parameters (such as bond lengths) exists, just as does the potential energy surface of the molecule. Averaging on the property surface is according to the probability of finding the molecule at various geometries, which in turn is determined by the vibrational wavefunctions corresponding to the potential energy surface.[111] Raynes et al. have provided very good case studies of these effects on coupling constants in polyatomic molecules, including details of the theoretical surfaces, dynamic averaging, and experimental measurements of the temperature dependence of spin–spin coupling in the various isotopomers.[112–115]

6.2 Intermolecular Effects

NMR shielding is extremely sensitive to intermolecular effects. This sensitivity is manifested by the very large gas-to-liquid shifts (4.4 ppm for ^{1}H in H_2O, 19.5 ppm for ^{15}N in NH_3, 77 ppm for ^{31}P in P_4, 120 ppm for ^{77}Se in H_2Se, around 200 ppm for ^{129}Xe in xenon), by the aromatic solvent-induced proton chemical shifts, and by the very large average chemical shifts observed for Xe in various media such as zeolites and polymers.[105,110] Theoretical calculations of these intermolecular effects are sometimes carried out by approximating the medium as a continuum and considering the molecule in a cavity within this medium of fixed dielectric constant. Such an option is routine in many quantum mechanical software packages. Another approach is to consider the intermolecular effects in terms of electrical polarization effects of fixed partial electrical charges centered on surrounding atoms in a crystalline system. A more complete treatment is to calculate the nuclear shielding in the molecule together with the solvent at various internuclear separations and orientations and then to average such shieldings over the appropriately weighted configurations at each temperature.[116,117] The effects of hydrogen bonding on the ^{1}H shielding tensor in ice have been reproduced by calculations using 17 H_2O molecules arranged in the experimental ice configuration, for example, emphasizing the importance of long-range effects.[118] While the short range effects on shielding of ^{13}C, ^{15}N, ^{17}O, and ^{1}H nuclei in each amino acid residue protein (the geometry dependence and hydrogen bonding) can be calculated using a model fragment, the long-range effects of neighboring residues may be considered in the same way as intermolecular effects from solvent molecules. It

has been found that such long-range electrostatic effects have an important role to play in interpreting ^{1}H, ^{15}N, and ^{17}O chemical shift inequivalencies in proteins, and neighbor anisotropy has an important role in the case of ^{1}H. Molecular dynamics (MD) has been used to produce a large number of instantaneous configurations for which the component magnitudes and directions of the ^{15}N shielding tensor can be calculated in a simple model system (three *N*-methylacetamide complexes) constructed from a gramicidin channel in a fully hydrated phospholipid bilayer.[119] The MD method of sampling reveals fluctuations in the tensor properties which may be used to investigate the types of NMR spectra produced during such motional averaging for a protein observed in a fluid bilayer environment.

A clear case of intermolecular effect on chemical shifts is found in the xenon atom. The use of ^{129}Xe chemical shifts in studies of various electronic environments (zeolites, polymers, clays, coals, biological systems) depends on theoretical calculations of the intermolecular effects on ^{129}Xe shielding in the xenon atom. The dispersion of the Xe signal in these various media is very useful as a diagnostic tool in the analysis of structure of the medium, the distribution of Xe atoms within, and the rate of exchange (diffusion) of Xe from one cage (or channel or domain) to another, as well as from within the medium to the bulk phase. The ability to reproduce the temperature dependence of the intensities and the individual chemical shifts in the ^{129}Xe NMR spectrum of the various Xe_1, Xe_2, $Xe_3, \ldots, Xe_8$ signals observed for xenon trapped in variety of *A* zeolites, (Na*A*, $Ca_xNa_{12-2x}A$, K*A*, Ag*A*) using a combination of quantum mechanical calculations and statistical mechanical averaging,[117,120–123] permits the interpretation of the NMR observations in various other zeolites wherein fast xenon exchange leads to one signal that contains the average over all distributions.[124] The Xe chemical shifts are very large (several hundred ppm) and discriminate between amorphous and crystalline regions in polymers and coals; however, the theoretical prediction of the Xe NMR spectrum in these more complex materials lags behind the experiments. The same approach as used in the crystalline zeolites should work in these more complex systems, provided that an appropriate model system can be introduced and tested in each case.

Theoretical approaches for intermolecular effects on **J** would be the same as for chemical shifts, although the observed effects are generally smaller. Intermolecular effects on electric field gradients at a nucleus can be substantial. In van der Waals dimers and higher *n*-mers, the nuclear quadrupole coupling constant can be sufficiently different from that found in the monomer and can be used to deduce the structure of the clusters formed. Calculations of the electric field gradient at a nucleus in a cluster treated as a supermolecule generally have to be carried out as a function of geometry, and it is usually necessary for internal coordinates to be permitted to vary with intermolecular separation. Furthermore, consistent with the strength of intermolecular forces, cluster vibrations can be very anharmonic, so averaging will have to be done accordingly. Farrar and Weinhold have carried out averaging of the electric field gradient at the ^{14}N and ^{2}H nuclei of HCN in $(HCN)_2$ and $(HCN)_3$ and $(HCN)_n$ up to $n = 6$.[101] The interpretation of the nuclear quadrupole couplings in liquid HCN as a function of temperature also requires that the distribution of the dominant clusters be calculated by statistical mechanics. The nonpairwise additive cooperativity effects in the hydrogen bonding that are comparable in magnitude to that of dimer formation in this system cannot be neglected in the calculation of the NMR parameters. For the same reason, the NMR parameters of liquid water or ice cannot be deduced from calculations on $(H_2O)_2$ alone.

Many interesting materials analyzed by NMR spectroscopy are made up of covalent networks. Calculations of NMR parameters in extended networks is not yet tractable. One approach to calculations in such systems is to consider models consisting of truncated clusters. For example, ^{29}Si shielding, ^{17}O shielding and electric field gradients, and ^{27}Al electric field gradients in silicates and aluminosilicates are of great interest in the NMR analyzes of zeolite structure. Zeolites are aluminosilicates with crystalline structure which are constructed from SiO_4 tetrahedra joined together by sharing edges and vertices. Al may replace Si in the lattice, with the charge balanced by extra-framework cations. Thus, the crystal is formed by a network of Al/Si atoms connected by bridging O atoms. Some clusters that have been used in calculations of NMR parameters of zeolites include fragments of the sodalite cage terminated by H atoms replacing the truncated Al or Si atoms in the extended network, or by replacing the truncated bridging oxygen atoms by H atoms.[125,126] Another possible approach is the use of an embedded small fragment within the much larger cluster. The small fragment which consists of the nucleus of interest and its immediate neighbors, is treated ab initio quantum mechanically at an appropriate level, the rest of the atoms in the much larger truncated cluster carved out of the network being treated at a lower level (semiempirically or by molecular mechanics). Such hybrid approaches appear promising for extended covalent networks and polymers.

7 FUTURE DEVELOPMENTS

Theoretical calculations of shielding and electric field gradient tensors for isolated molecules are well in hand,

except where relativistic effects are important. Relativistic calculations have yet to be carried out in a theoretically rigorous manner, rather than using a patchwork of spin-free and spin–orbit contributions. Heavy nuclei, particularly transition metal nuclei, which are important components of technologically important solid materials, cannot be treated accurately until the problem of relativistic effects is solved. The extreme sensitivity of shielding and electric field gradient tensors to intermolecular effects and local geometries (imposed by longer range order and distributions) presents a distinct advantage as well as difficulties. Applications of calculations of shielding and electric field gradients to interpretations and analyses of complex systems require construction of appropriate useful models that can be tested in simpler systems and extended to complex ones. By separating the various short-range (geometrical, hydrogen bonding) and long-range intermolecular contributions to the chemical shift, a powerful tool in analyzing structure, dynamics, and environment can be realized. The approach used in proteins can be extended to other polymers and even mixtures of polymers. Theoretical calculations of spin–spin coupling constants are more difficult, even in isolated small molecules, but are slowly becoming more tractable. These have sensitivity to local geometries but are less susceptible to intermolecular effects and long-range contributions. The use of density functional techniques offers hope that larger systems can be handled, but the difficulties with the presently available approximate exchange correlation functionals naturally limit the accuracy of the results that can be obtained. When more accurate functionals become available, this will clearly be the method of choice for very large systems.

ABBREVIATIONS AND ACRONYMS

BDFT	Magnetic Field Density Functional Theory
BLYP	Becke–Lee–Yang–Parr
CAS	Complete Active Spaces
CC	Coupled Cluster
CCSD	Coupled Cluster Singles and Doubles
CCSD(T)	Coupled Cluster Singles and Doubles with Some Triple Excitations
CDFT	Current Density Functional Theory
CHF	Coupled Hartree–Fock
CSGT	Continuous Set of Gauge Transformations
DFT	Density Functional Theory
EOM	Equation of Motion
FC	Fermi Contact
FCI	Full Configuration Interaction
GIAO	Gauge Including Atomic Orbital
HRPA	Higher Random Phase Approximation
IGAIM	Individual Gauges for Atoms in Molecules
IGLO	Individual Gauge for localized Orbitals
KSDFT	Kohn–Sham Density Functional Theory
LDA	Local Density Approximation
LORG	Localized Orbital/Local Origin
MAS	Magic-angle Spinning
MBPT	Many Body Perturbation Theory
MBPT(2)	Second-order Many Body Perturbation Theory
MBPT(n)	Many Body Perturbation Theory (nth order term)
MCIGLO	Multiconfiguration Individual Gauge for Localized Orbitals
MCLR	Multiconfiguration Linear Response
MCSCF	Multiconfiguration Self-consistent Field
MD	Molecular Dynamics
MP2	Møller–Plesset Second-order Term
MPn	Møller–Plesset nth order term
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
OD	Diamagnetic Orbital
OP	Paramagnetic Orbital
PPA	Polarization Propagator Approximation
RAS	Restricted Active Spaces
RPA	Random Phase Approximation
SCF	Self-consistent Field
SD	Spin Dipolar
SOPPA	Second-order Polarization Propagator Approximation
SOS	Sum Over States
TMS	Tetramethylsilane
UHF	Unrestricted Hartree–Fock

RELATED ARTICLES

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Quadrupole Couplings in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General

REFERENCES

1. N.F. Ramsey, 'Magnetic Shielding of Nuclei in Molecules', *Phys. Rev.*, **78**(6), 699–703 (1950).
2. N.F. Ramsey, 'Chemical Effects in Nuclear Magnetic Resonance and in Diamagnetic Susceptibility', *Phys. Rev.*, **86**(2), 243–246 (1952).
3. N.F. Ramsey, 'Electron Coupled Interactions Between Nuclear Spins in Molecules', *Phys. Rev.*, **91**(2), 303–307 (1953).
4. F. Michelot, 'Nuclear Hyperfine Interactions in Nonlinear Semi-rigid Molecules I. The Molecular Hamiltonian', *Mol. Phys.*, **45**(5), 949–970 (1982).
5. F. Michelot, 'Nuclear Hyperfine Interactions in Nonlinear Semi-rigid Molecules II. The Effective Hamiltonian for a Non-degenerate Electronic State', *Mol. Phys.*, **45**(5), 971–1001 (1982).
6. A.E. Hansen, T.D. Bouman, 'Localized Orbital/Local Origin Method for Calculation and Analysis of NMR Shieldings. Applications to ^{13}C Shielding Tensors', *J. Chem. Phys.*, **82**(11), 5035–5047 (1985).
7. W. Kutzelnigg, U. Fleischer, M. Schindler, 'Ab Initio Calculation and Interpretation of Chemical Shifts in NMR by Means of the IGLO Method', *NMR Basic Principles and Prog.*, **23**, 165–262 (1990).
8. K. Wolinski, J.F. Hinton, P. Pulay, 'Efficient Implementation of the Gauge-independent Atomic Orbital Method for NMR Chemical Shift Calculations', *J. Am. Chem. Soc.*, **112**(23), 8251–8260 (1990).
9. T.A. Keith, R.F.W. Bader, 'Calculation of Magnetic Response Properties Using Atoms in Molecules', *Chem. Phys. Lett.*, **194**(1–2), 1–8 (1992).
10. J. Geertsen, 'A Solution of the Gauge-origin Problem for the Magnetic Shielding Constant', *Chem. Phys. Lett.*, **179**(5–6), 479–482 (1991).
11. P. Pyykkö, 'Relativistic Theory of Nuclear Spin–Spin Coupling in Molecules', *Chem. Phys.*, **22**, 289–296 (1977).
12. K. Wolinski, C.L. Hsu, J.F. Hinton, P. Pulay, 'Hartree–Fock and 2nd Order Møller–Plesset Perturbation Theory Calculations of the ^{31}P NMR Shielding Tensor in PH_3', *J. Chem. Phys.*, **99**(10), 7819–7824 (1993).
13. K. Ruud, T. Helgaker, R. Kobayashi, P. Jørgensen, K.L. Bak, H.J.A. Jensen, 'Multiconfigurational Self Consistent Field Calculations of Nuclear Shieldings Using London Atomic Orbitals', *J. Chem. Phys.*, **100**(11), 8178–8185 (1994).
14. C. van Wüllen, W. Kutzelnigg, 'Calculation of NMR Shieldings and Magnetic Susceptibilities Using Multiconfiguration Hartree–Fock Wavefunctions and Local Gauge Origins', *J. Chem. Phys.*, **104**(6), 2330–2340 (1996).
15. J. Gauss, 'Calculation of NMR Chemical Shifts at 2nd Order Many Body Perturbation Theory Using Gauge-including Atomic Orbitals', *Chem. Phys. Lett.*, **191**(6), 614–620 (1992).
16. J. Gauss, 'Effects of Electron Correlation in the Calculation of Nuclear Magnetic Resonance Chemical Shifts', *J. Chem. Phys.*, **99**(5), 3629–3643 (1993).
17. J. Gauss, 'GIAO-MBPT(3) and GIAO-SDQ-MBPT(4) Calculations of NMR Shielding Constants', *Chem. Phys. Lett.*, **229**(3), 198–203 (1994).
18. G.D. Purvis, R.J. Bartlett, 'A Full Coupled Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples', *J. Chem. Phys.*, **76**(4), 1910–1918 (1982).
19. J. Gauss, J.F. Stanton, 'Coupled Cluster Calculations of NMR Chemical Shifts', *J. Chem. Phys.*, **103**(9), 3561–3577 (1995).
20. J. Gauss, J.F. Stanton, 'Perturbative Treatment of Triple Excitations in Coupled Cluster Calculations of Nuclear Magnetic Shielding Constants', *J. Chem. Phys.*, **104**(7), 2574–2583 (1996).
21. A. Barszczewicz, T. Helgaker, M. Jaszunski, P. Jørgensen, K. Ruud, 'NMR Shielding Tensors and Indirect Spin–Spin Coupling Tensors in HCN, HNC, CH_3CN and CH_3NC Molecules', *J. Magn. Reson., Ser. A*, **114**(2), 212–218 (1995).
22. J. Gauss, K. Ruud, 'On the Convergence of MBPT and CC Nuclear Shielding Constants of BH Toward the Full CI Limit', *Int. J. Quantum Chem.*, **S29**, 437–442 (1995).
23. J. Geertsen, J. Oddershede, 'Second Order Polarization Propagator Calculations of Indirect Nuclear Spin–Spin Coupling Tensors in the Water Molecule', *Chem. Phys.*, **90**, 301–311 (1984).
24. J. Geertsen, J. Oddershede, 'Higher RPA and Second Order Polarization Propagator Calculations of Coupling Constants in Acetylene', *Chem. Phys.*, **104**, 67–72 (1986).
25. J. Geertsen, J. Oddershede, G.E. Scuseria, 'Spin–Spin Coupling Constants of CO and N_2', *J. Chem. Phys.*, **87**(4), 2138–2142 (1987).
26. J. Geertsen, J. Oddershede, W.T. Raynes, G.E. Scuseria, 'Nuclear Spin–Spin Coupling in the Methane Isotopomers', *J. Magn. Reson.*, **93**, 458–471 (1991).
27. S.A. Perera, H. Sekino, R.J. Bartlett, 'Coupled Cluster Calculations of Indirect Nuclear Coupling Constants: The Importance of Non-Fermi Contact Contributions', *J. Chem. Phys.*, **101**(3), 2186–2191 (1994).
28. V. Galasso, 'Ab Initio Study of the **J**(CC) Indirect Nuclear Spin–Spin Coupling Constants in Cyclobutane and Related Systems', *Chem. Phys.*, **117**, 415–420 (1987).
29. O. Vahtras, H. Agren, P. Jørgensen, H.J.Aa. Jensen, S.B. Padkjær, T. Helgaker, 'Indirect Nuclear Spin–Spin Coupling Constants from Multiconfigurational Linear Response Theory', *J. Chem. Phys.*, **96**(8), 6120–6125 (1992).
30. A. Barszczewicz, T. Helgaker, M. Jaszunski, P. Jørgensen, K. Ruud, 'Multiconfigurational Self Consistent Field Calculations of Nuclear Magnetic Resonance Indirect Spin–Spin Coupling Constants', *J. Chem. Phys.*, **101**(8), 6822–6828 (1994).

31. S. Kirpekar, H.J.Aa. Jensen, J. Oddershede, 'Correlated Calculations of Indirect Nuclear Spin–Spin Coupling Constants for XH_4 (X = Si, Ge, and Sn)', *Chem. Phys.*, **188**, 171–181 (1994).
32. H. Sekino, R.J. Bartlett, 'Nuclear Spin–Spin Coupling Constants Evaluated Using Many Body Methods', *J. Chem. Phys.*, **85**(7), 3945–3949 (1986).
33. H. Fukui, K. Miura, H. Matsuda, 'Calculation of the Nuclear Spin–Spin Coupling Constants. VI. Many Body Perturbation Theoretical Calculation of Electron Correlation Effect', *J. Chem. Phys.*, **94**(1), 533–536 (1991).
34. R.G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, 1989.
35. G. Vignale, M. Rasolt, 'Density Functional Theory in Strong Magnetic Fields', *Phys. Rev. Lett.*, **59**(20), 2360–2363 (1987).
36. G. Vignale, M. Rasolt, 'Current- and Spin-density Functional Theory for Inhomogeneous Electronic Systems in Strong Magnetic Fields', *Phys. Rev. B*, **37**(18), 10685–10696 (1988).
37. G. Vignale, M. Rasolt, D.J.W. Geldart, 'Magnetic Fields and Density Functional Theory', *Adv. Quantum Chem.*, **21**, 235–253 (1990).
38. A.M. Lee, N.C. Handy, S.M. Colwell, 'The Density Functional Calculation of Nuclear Shielding Constants Using London Atomic Orbitals', *J. Chem. Phys.*, **103**(23), 10095–10109 (1995).
39. C.J. Grayce, R.A. Harris, 'Magnetic Field Density Functional Theory', *Phys. Rev. A*, **50**(4), 3089–3095 (1994).
40. C.J. Grayce, R.A. Harris, 'Two Dipole Magnetic Field Density Functional Theory', *J. Phys. Chem.*, **99**(9), 2724–2726 (1995).
41. V.G. Malkin, O.L. Malkina, M.E. Casida, D.R. Salahub, 'NMR Shielding Tensors Calculated with a Sum Over States Density Functional Perturbation Theory', *J. Am. Chem. Soc.*, **116**(13), 5898–5908 (1994).
42. V.G. Malkin, O.L. Malkina, L.A. Eriksson, D.R. Salahub, in 'Modern Density Functional Theory: A Tool for Chemistry', *Theoretical and Computational Chemistry*, eds. P. Politzer, J.M. Seminario, Elsevier, Amsterdam, 273, Vol. 2, 1995.
43. C. van Wüllen, 'Density Functional Calculation of NMR Chemical Shifts', *J. Chem. Phys.*, **102**(7), 2806–2811 (1995).
44. A.D. Becke, 'Density Functional Thermochemistry. III. The Role of Exact Exchange', *J. Chem. Phys.*, **98**(7), 5648–5652 (1993).
45. C. Lee, W. Yang, R.G. Parr, 'Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density', *Phys. Rev. B*, **37**(2), 785–789 (1988).
46. G. Schreckenbach, T. Ziegler, 'Calculation of NMR Shielding Tensors Using Gauge-including Atomic Orbitals and Modern Density Functional Theory', *J. Phys. Chem.*, **99**(2), 606–611 (1995).
47. G. Schreckenbach, R.M. Dickson, Y. Ruiz-Morales, T. Ziegler, in *Density Functional Theory in Chemistry*, eds. B. Laird, R. Ross, T. Ziegler, American Chemical Society, Washington, DC, 1996.
48. G. Rauhut, S. Puyear, K. Wolinski, P. Pulay, 'Comparison of NMR Shieldings Calculated from Hartree–Fock and Density Functional Wavefunctions Using Gauge-including Atomic Orbitals', *J. Phys. Chem.*, **100**(15), 6310–6316 (1996).
49. J.R. Cheeseman, G.W. Trucks, T.A. Keith, M.J. Frisch, 'A Comparison of Models for Calculating Nuclear Magnetic Resonance Shielding Tensors', *J. Chem. Phys.*, **104**(14), 5497–5509 (1996).
50. G. Schreckenbach, Y. Ruiz-Morales, T. Ziegler, 'The Calculation of ^{77}Se Chemical Shifts Using Gauge-including Atomic Orbitals and Density Functional Theory', *J. Chem. Phys.*, **104**(21), 8605–8612 (1996).
51. T.A. Keith, R.F.W. Bader, 'Calculation of Magnetic Response Properties Using a Continuous Set of Gauge Transformations', *Chem. Phys. Lett.*, **210**(1–3), 223–231 (1993).
52. O.L. Malkina, D.R. Salahub, V.G. Malkin, 'Nuclear Magnetic Resonance Spin–Spin Coupling Constants from Density Functional Theory: Problems and Results', *J. Chem. Phys.*, **105**(19), 8793–8800 (1996).
53. W. Kutzelnigg, E. Ottschofski, R. Franke, 'Relativistic Hartree–Fock by Means of Stationary Direct Perturbation Theory. 1. General Theory', *J. Chem. Phys.*, **102**(4), 1740–1751 (1995).
54. H. Nakatsuji, M. Hada, H. Kaneko, C.C. Ballard, 'Relativistic Study of Nuclear Magnetic Shielding Constants – Mercury Dihalides', *Chem. Phys. Lett.*, **255**(1–3), 195–202 (1996).
55. H. Nakatsuji, T. Nakajima, M. Hada, H. Takashima, S. Tanaka, 'Spin–Orbit Effect on the Magnetic Shielding Constants Using the Ab Initio UHF Method – Silicon Tetrahalides', *Chem. Phys. Lett.*, **247**(4–6), 418–424 (1995).
56. C.J. Jameson, J. Mason, 'The Chemical Shift', in *Multinuclear NMR*, ed. J. Mason, Plenum, New York, 51–88, 1987.
57. H. Nakatsuji, H. Takashima, M. Hada, 'Spin–Orbit Effect on the Magnetic Shielding Constant Using the Ab Initio UHF Method', *Chem. Phys. Lett.*, **233**(1–2), 95–101 (1995).
58. H. Takashima, M. Hada, H. Nakatsuji, 'Spin–Orbit effect on the Magnetic Shielding Constant Using the Ab Initio UHF Method – Ga and In Tetrahalides', *Chem. Phys. Lett.*, **235**(1–2), 13–16 (1995).
59. W. Kutzelnigg, 'Origin and Meaning of the Fermi Contact Interaction', *Theor. Chim. Acta*, **73**, 173–200 (1988).
60. G.A. Aucar, J. Oddershede, 'Relativistic Theory for Indirect Nuclear Spin–Spin Couplings within the

Polarization Propagator Approach', *Int. J. Quantum Chem.*, **47**, 425–435 (1993).
61. P. Pyykkö, J.R. Wiesenfeld, 'Relativistically Parameterized Extended Hückel Calculations IV. Nuclear Spin–Spin Coupling Tensors for Main Group Elements', *Mol. Phys.*, **43**(3), 557–580 (1981).
62. C.J. Jameson, 'Theoretical and Physical Aspects of Nuclear Shielding', in *Nuclear Magnetic Resonance*, ed. G.A. Webb, Royal Society of Chemistry, London, 469–487, Vol. 26, 1997.
63. C. van Wüllen, 'Magnetic Properties of the BH Molecule – a CASSCF Study', *Theor. Chim Acta*, **87**(1–2), 89–95 (1993).
64. J. Gauss, 'Accurate Calculation of NMR Chemical Shifts', *Ber. Bunsenges. Phys. Chem.*, **99**(8), 1001–1008 (1995).
65. A.K. Jameson, C.J. Jameson, 'Gas Phase ^{13}C Chemical Shifts in the Zero-pressure Limit. Refinements to the Absolute Shielding Scale for ^{13}C', *Chem. Phys. Lett.*, **134**(5), 461–466 (1987).
66. D.M. Grant, J.C. Facelli, D.W. Alderman, M.H. Sherwood, 'Carbon-13 Chemical Shielding Tensors in Sugars: Sucrose and Methyl-(-D-Glucopyranoside)', in *Nuclear Magnetic Shielding and Molecular Structure*, ed., J.A. Tossell, Kluwer Academic, Dordrecht, 367–384, 1993.
67. F. Liu, C.G. Phung, D.W. Alderman, D.M. Grant, 'Carbon-13 Chemical Shift Tensors in Methyl Glycosides, Comparing Diffraction and Optimized Structures with Single-Crystal NMR', *J. Am. Chem. Soc.*, **118**(43), 10629–10634 (1996).
68. J.C. Facelli, D.M. Grant, 'Determination of Molecular Symmetry in Crystalline Naphthalene Using Solid State NMR', *Nature*, **365**, 325–327 (1993).
69. A.C. de Dios, D.D. Laws, E. Oldfield, 'Predicting ^{13}C NMR Chemical Shielding Tensors in Zwitterionic L-Threonine and L-Tyrosine via Quantum Chemistry', *J. Am. Chem. Soc.*, **116**(17), 7784–7786 (1994).
70. M. Kaupp, O.L. Malkina, V.G. Malkin, 'The Calculation of ^{17}O Chemical Shielding in Transition Metal Oxo Complexes. I. Comparison of DFT and Ab Initio Approaches, and Mechanisms of Relativity-induced Shielding', *J. Chem. Phys.*, **106**(22), 9201–9212 (1997).
71. M. Kaupp, 'Analysis of ^{13}C and ^{17}O Chemical Shift Tensors and ELF View of Bonding in $Fe_2(CO)_9$ and $Rh_6(CO)_{16}$', *Chem. Ber.*, **129**, 527–533 (1996).
72. Y. Ruiz-Morales, G. Schreckenbach, T. Ziegler, 'Theoretical Study of ^{13}C and ^{17}O NMR Shielding Tensors in Transition Metal Carbonyl Based on Density Functional Theory and Gauge Including Atomic Orbitals', *J. Phys. Chem.*, **100**(9), 3359–3367 (1996).
73. M.A. Fedotov, O.L. Malkina, V.G. Malkin, '$^{35/37}Cl$ NMR Chemical Shifts and Nuclear Quadrupole Couplings for Some Small Chlorine Compounds: Experimental and Theoretical Study', *Chem. Phys. Lett.*, **258**, 330–335 (1996).
74. D.B. Chesnut, E.F.C. Byrd, 'The Inclusion of Correlation in the Calculation of Phosphorus NMR Chemical Shieldings', *Heteroatom Chem.*, **7**(5), 307–312 (1996).
75. C.J. Jameson, A.K. Jameson, 'Concurrent ^{19}F and ^{77}Se or ^{19}F and ^{125}Te NMR T_1 Measurements for Determination of ^{77}Se and ^{125}Te Absolute Shielding Scales', *Chem. Phys. Lett.*, **135**(3), 254–259 (1987).
76. M. Bühl, J. Gauss, J.F. Stanton, 'Accurate Computations of ^{77}Se NMR Chemical Shifts Using the GIAO-CCSD Method', *Chem. Phys. Lett.*, **241**(3), 248–252 (1995).
77. C.J. Jameson, 'Theoretical and Physical Aspects of Nuclear Shielding', in *Nuclear Magnetic Resonance*, ed. G.A. Webb, Royal Society of Chemistry, London, 39–82, Vol. 25, 1996.
78. Y. Ruiz-Morales, G. Schreckenbach, T. Ziegler, 'Calculation of ^{125}Te Chemical Shifts Using Gauge Including Atomic Orbitals and Density Functional Theory', *J. Phys. Chem.*, **101**(22), 4121–4127 (1997).
79. G. Schreckenbach, T. Ziegler, 'Calculation of NMR Shielding Tensors Based on Density Functional Theory and a Scalar Relativistic Pauli-type Hamiltonian. The Application to Transition Metal Complexes', *Int. J. Quantum Chem.*, **61**, 899–918 (1997).
80. P.D. Ellis, J.D. Odom, A.S. Lipton, Q. Chen, J.M. Gulick, 'Gas Phase Measurements and Ab Initio Calculations of ^{77}Se and ^{113}Cd Chemical Shifts', in *Nuclear Magnetic Shielding and Molecular Structure*, ed. J.A. Tossell, Kluwer Academic, Dordrecht, 539–555, 1993.
81. T. Higashioji, M. Hada, M. Sugimoto, H. Nakatsuji, 'Basis Set Dependence of Magnetic Shielding Constant Calculated by the Hartree–Fock/Finite Perturbation Method', *Chem. Phys.*, **203**, 159–175 (1996).
82. G.E. Scuseria, 'A Systematic Study of the Ab Initio Diamagnetic Spin–Orbital Contribution to Calculated Spin–Spin Coupling Constants', *Chem. Phys.*, **107**, 417–427 (1986).
83. R.M. Dickson, T. Ziegler, 'NMR Spin–Spin Coupling Constants from Density Functional Theory with Slater-type Basis Functions', *J. Phys. Chem.*, **100**(13), 5286–5290 (1996).
84. O. Vahtras, H. Agren, P. Jørgensen, T. Helgaker, H.J.Aa. Jensen, 'The Nuclear Spin–Spin Coupling in N_2 and CO', *Chem. Phys. Lett.*, **209**(3), 201–206 (1993).
85. J. Geertsen, J. Oddershede, G.E. Scuseria, 'Calculation of Spectra and Spin–Spin Coupling Constants Using a Coupled Cluster Polarization Propagator Method', *Int. J. Quantum Chem. Symp.*, **21**, 475–485 (1987).
86. H. Fukui, K. Miura, H. Matsuda, T. Baba, 'Calculation of Nuclear Spin–Spin Couplings. VII. Electron Correlation Effects on the Five Coupling Mechanisms', *J. Chem. Phys.*, **97**(4), 2299–2304 (1992).
87. V.G. Malkin, O.L. Malkina, D.R. Salahub, 'Calculations of Spin–Spin Coupling Constants Using Density Functional Theory', *Chem. Phys. Lett.*, **221**(1–2), 91–99 (1994).

88. S.A. Perera, M. Nooijen, R.J. Bartlett, 'Electron Correlation Effects on the Theoretical Calculation of Nuclear Magnetic Resonance Spin–Spin Coupling Constants', *J. Chem. Phys.*, **104**(9), 3290–3305 (1996).
89. A. Barszczewicz, M. Jaszunski, K. Kamienska-Trela, T. Helgaker, P. Jørgensen, O. Vahtras, 'Ab Initio Study of the NMR Shielding Constants and Spin–Spin Coupling Constants in Cyclopropene', *Theor. Chim. Acta*, **87**, 19–28 (1993).
90. V. Galasso, 'Theoretical Study of Spectroscopic Properties of Cubane', *Chem. Phys.*, **184**, 107–114 (1994).
91. C.J. Jameson, J. Mason, 'Spin–Spin Coupling' in *Multinuclear NMR*, ed., J. Mason, Plenum, New York, 89–132, 1987.
92. M.D. Lumsden, K. Eichele, R.E. Wasylishen, T.S. Cameron, J.F. Britten, 'Determination of a ^{199}Hg–^{31}P Indirect Spin–Spin Coupling Tensor via Single-crystal Phosphorus NMR Spectroscopy', *J. Am. Chem. Soc.*, **116**(24), 11129–11136 (1994).
93. M. Karplus, 'Contact Electron Spin Coupling of Nuclear Magnetic Moments', *J. Chem. Phys.*, **30**(1), 11–15 (1959).
94. H. Fukui, H. Inomata, T. Baba, K. Miura, H. Matsuda, 'Calculation of Nuclear Spin–Spin Couplings. VIII. Vicinal Proton–Proton Coupling Constants in Ethane', *J. Chem. Phys.*, **103**(15), 6597–6600 (1995).
95. G. Breit, 'Possible Effects of Nuclear Spin on X-Ray Terms', *Phys. Rev.*, **35**(12), 1447–1451 (1930).
96. P. Pyykkö, E. Pajanne, M. Inokuti, 'Hydrogen Like Relativistic Corrections for Electric and Magnetic Hyperfine Integrals', *Int. J. Quantum Chem.*, **7**, 785–806 (1973).
97. K.E. Vermillion, P. Florian, P.J. Grandinetti, 'Relationships Between Bridging Oxygen ^{17}O Quadrupolar Coupling Parameters and Structure in Alkali Silicates', *J. Chem. Phys.*, **108**(17), 7274–7285 (1998).
98. C. Müller, W. Schajor, H. Zimmermann, U. Haeberlen, 'Deuteron Chemical Shift and EFG Tensor in α-Glycine', *J. Magn. Reson.*, **56**, 235–246 (1984).
99. P.L. Cummins, G.B. Bacskay, N.S. Hush, 'The Prediction of Nuclear Quadrupole Moments from Ab Initio Quantum Chemical Studies on Small Molecules. II. The Electric Field Gradients at the ^{17}O, ^{35}Cl, and ^{2}H Nuclei in CO, NO^+, OH^-, H_2O, CH_2O, HCl, LiCl, and FCl', *J. Chem. Phys.*, **87**(1), 416–423 (1987).
100. R. Ludwig, F. Weinhold, T.C. Farrar, 'Experimental and Theoretical Studies of Hydrogen Bonding in Neat, Liquid Formamide', *J. Chem. Phys.*, **102**(13), 5118–5125 (1995).
101. B.F. King, T.C. Farrar, F. Weinhold, 'Quadrupole Coupling Constants in Linear $(HCN)_n$ clusters: Theoretical and Experimental Evidence for Cooperativity Effects in C–H···N Hydrogen Bonding', *J. Chem. Phys.*, **103**(1), 348–352 (1995).
102. W.P. Power, R.E. Wasylishen, 'NMR Studies of Isolated Spin Pairs in the Solid State', in *Annual Reports on NMR Spectroscopy*, ed. G.A. Webb, Academic Press, London, 1–84, Vol. 23, 1991.
103. P. Zhang, C. Dunlap, P. Florian, P.J. Grandinetti, I. Farnan, J.F. Stebbins, 'Silicon Site Distributions in an Alkali Silicate Glass Derived by Two-dimensional ^{29}Si NMR', *J. Non-cryst. Solids*, **204**, 294–300 (1996).
104. I. Farnan, P.J. Grandinetti, J.H. Baltisberger, J.F. Stebbins, U. Werner, M.A. Eastman, A. Pines, 'Quantification of the Disorder in Network-modified Silicate Glasses', *Nature*, **358**, 31–35 (1992).
105. A.C. de Dios, C.J. Jameson, 'The NMR Chemical Shift: Insight into Structure and Environment', in *Annual Reports on NMR Spectroscopy*, ed. G.A. Webb, Academic Press, London, 1–69, Vol. 29, 1994.
106. A.C. de Dios, J.G. Pearson, E. Oldfield, 'Secondary and Tertiary Structural Effects on Protein NMR Chemical Shifts: An Ab Initio Approach', *Science*, **260**, 1491–1496 (1993).
107. A.C. de Dios, E. Oldfield, 'Ab Initio Study of the Effects of Torsion Angles on Carbon-13 Nuclear Magnetic Resonance Chemical Shielding in *N*-Formyl-L-Alanine, *N*-Formyl-L-Valine Amide, and Some Simple Model Compounds: Applications to Protein NMR Spectroscopy', *J. Am. Chem. Soc.*, **116**(12), 5307–5314 (1994).
108. H.B. Le, J.G. Pearson, A.C. de Dios, E. Oldfield, 'Protein Structure Refinement and Prediction via NMR Chemical Shifts and Quantum Chemistry', *J. Am. Chem. Soc.*, **117**(13), 3800–3807 (1995).
109. C.J. Jameson, A.C. de Dios, 'The Nuclear Shielding Surface: The Shielding as a Function of Molecular Geometry and Intermolecular Separation', in *Nuclear Magnetic Shielding and Molecular Structure*, ed. J.A. Tossell, Kluwer Academic, Dordrecht, 95–116, 1993.
110. C.J. Jameson, 'Understanding NMR Chemical Shifts', *Ann. Rev. Phys. Chem.*, **47**, 135–169 (1996).
111. C.J. Jameson, 'Rovibrational Averaging of Molecular Electronic Properties', in *Theoretical Models of Chemical Bonding*, Part 3. Molecular Spectroscopy, Electronic Structure, and Intramolecular Interactions, ed. Z.B. Maksic, Springer-Verlag, Berlin, 457–519, 1991.
112. W.T. Raynes, 'Theory of Vibrational Effects on Properties of Methane and its Isotopomers. The Polarizability, Magnetizability, Carbon-13 Shielding and the One- and Two-bond Spin–Spin Coupling Constants', *Mol. Phys.*, **63**(4), 719–729 (1988).
113. J. Geertsen, J. Oddershede, W.T. Raynes, T.L. Marvin, 'Accurate Ab Initio Carbon–Proton and Proton–Proton Spin–Spin Coupling Surfaces for the Methane Molecule', *Mol. Phys.*, **82**(1), 29–50 (1994).
114. W.T. Raynes, J. Geertsen, J. Oddershede, 'Nuclear Spin–Spin Coupling and Nuclear Motion', *Int. J. Quantum Chem.*, **52**(1), 153–163 (1994).
115. B. Bennett, W.T. Raynes, C.W. Anderson, 'Temperature Dependence of J(CH) and J(CD) in $^{13}CH_4$ and

some of its Isotopomers', *Spectrochim. Acta A*, **45**(8), 821–827 (1989).

116. A.C. de Dios, C.J. Jameson, 'The ^{129}Xe Nuclear Shielding Surfaces for Xe Interacting with Linear Molecules CO, N_2 and CO_2', *J. Chem. Phys.*, **107**(11), 4253–4270 (1997).
117. C.J. Jameson, H.M. Lim, A.K. Jameson, 'The Role of Polarization of Xe by Di- and Monovalent Cations in ^{129}Xe NMR Studies in Zeolite A', *Solid State Nucl. Magn. Reson.*, **9**, 277–301 (1997).
118. J.F. Hinton, P. Guthrie, P. Pulay, K. Wolinski, 'Ab Initio Quantum Mechanical Calculation of the Chemical Shift Anisotropy of the Hydrogen Atom in the $(H_2O)_{17}$ Cluster', *J. Am. Chem. Soc.*, **114**(5), 1604–1605 (1992).
119. T.B. Woolf, V.G. Malkin, O.L. Malkina, D.R. Salahub, B. Roux, 'The Backbone ^{15}N Chemical Shift Tensor of the Gramicidin Channel. A Molecular Dynamics and Density Functional Study', *Chem. Phys. Lett.*, **239**, 186–194 (1995).
120. C.J. Jameson, H.-M. Lim, 'Ab Initio Studies of the Nuclear Magnetic Resonance Chemical Shifts of a Rare Gas Atom in a Zeolite', *J. Chem. Phys.*, **103**(10), 3885–3894 (1995).
121. C.J. Jameson, A.K. Jameson, B.I. Baello, H.M. Lim, 'Grand Canonical Monte Carlo Simulations of Xenon in Zeolite NaA, I. Distribution and ^{129}Xe Chemical Shifts', *J. Chem. Phys.*, **100**(8), 5965–5976 (1994).
122. C.J. Jameson, A.K. Jameson, R.E. Gerald II, H.-M Lim, 'Xe_n Clusters in the Alpha Cages of Zeolite KA', *J. Chem. Phys.*, **103**(20), 8811–8820 (1995).
123. C.J. Jameson, H.M. Lim, 'Distribution and ^{129}Xe Chemical Shifts of Xe_n Clusters in the Alpha Cages of Zeolite AgA', *J. Chem. Phys.*, **107**(11), 4373–4383 (1997).
124. C.J. Jameson, A.K. Jameson, R.E. Gerald II, H.M. Lim, 'Anisotropic Xe Chemical Shifts in Zeolites. The Role of Intra- and Inter-crystallite Diffusion', *J. Phys. Chem.*, **101**(42), 8418–8437 (1997).
125. J. Sauer, J.R. Hill, 'The Acidity of Surface Silanol Groups. A Theoretical Estimate Based on Ab Initio Calculations on a Model Surface', *Chem. Phys. Lett.*, **218**, 333–337 (1994).
126. F. Haase, J. Sauer, 'Interaction of Methanol with Brønsted Acid Sites of Zeolite Catalysts: An Ab Initio Study', *J. Am. Chem. Soc.*, **117**(13), 3780–3789 (1995).

Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance

Dieter Freude
Universität Leipzig, Leipzig, Germany

1	**Introduction**	1
2	**Basic Theory**	2
3	**Excitation**	5
4	**Line Shape of Static and Magic-angle Spinning Nuclear Magnetic Resonance Spectra**	7
5	**Double-rotation and Dynamic-angle Spinning**	10
6	**Multiple-quantum Magic-angle Spinning Technique**	12
7	**Nutation Technique**	15
8	**Echo Techniques**	16
9	**^{2}H Nuclear Magnetic Resonance Spectroscopy**	17
10	**Survey of Nuclear Magnetic Resonance Parameters for Selected Compounds**	25
	Acknowledgments	25
	Abbreviations and Acronyms	25
	Related Articles	26
	References	26

Solid-state nuclear magnetic resonance (NMR) spectroscopy is mostly applied to ^1H or ^{13}C nuclei with the nuclear spin $I = \frac{1}{2}$, but about 100 of 130 NMR isotopes have $I > \frac{1}{2}$, and the electric quadrupole interaction strongly broadens the NMR signal in the solid-state powder spectra. The perturbing effect of the electric quadrupole interaction is reduced at the higher magnetic fields which are available at present. In addition, approaches of the solid-state NMR traditionally used in the study of spin-$\frac{1}{2}$ nuclei have been adapted for use with quadrupolar nuclei, and some techniques, e.g. double-rotation (DOR) and multiple-quantum transition in combination with fast spinning [multiple-quantum magic-angle spinning (MQMAS)], were recently developed for quadrupolar nuclei with half-integer spins.

This article describes the basic theory, the line shape for first- and second-order quadrupole broadened spectra with and without magic-angle spinning (MAS) of the powder, the most important experimental techniques for the study of quadrupolar nuclei with half-integer spins such as DOR, dynamic-angle spinning (DAS), MQMAS, echo and nutation techniques, and some recent developments in deuterium NMR. Electric field gradient and chemical shift data for the most commonly studied quadrupolar nuclei with half-integer spins, ^{27}Al, ^{23}Na, and ^{17}O, and a few references to recent solid-state NMR studies of some other quadrupolar nuclei are given.

1 INTRODUCTION

The applicability of the NMR technique continues to expand in physics, chemistry, materials science, geology, biology, and medicine for either spectroscopic studies or imaging purposes. The majority of the recent applications can be found in the field of analytical chemistry. In 1998, *Current Contents (Physical, Chemical and Earth Sciences)* referred to about 8500 NMR studies, including 1800 studies of solids. Solid-state NMR, for which some special techniques must be used in order to obtain highly resolved spectra, is mostly applied to ^{1}H or ^{13}C nuclei with spin $I = \frac{1}{2}$. About 100 of 130 NMR isotopes have $I > \frac{1}{2}$, and the electric quadrupole interaction strongly broadens the NMR signal in the solid-state powder spectra. However, more than 400 references to solid-state NMR studies of quadrupolar nuclei in *Current Contents* in 1998 demonstrate that this problem could be overcome.

The demand for the characterization of inorganic materials and some recently developed experimental techniques have led to growing interest in high-resolution solid-state NMR spectroscopy of quadrupolar nuclei in powder samples. In addition, the perturbing effect of the electric quadrupole interaction is reduced at higher magnetic fields, which are now commercially available up to 21 T also for solid-state NMR spectrometers. In addition, approaches of the solid-state NMR traditionally used in the study of spin-$\frac{1}{2}$ nuclei have been adapted for use with quadrupolar nuclei, e.g. echo methods, sample spinning about the magic angle (MAS), multiple-dimensional spectroscopy, double resonance and adiabatic transfer. Some techniques, e.g. DOR and multiple-quantum transition in combination with fast spinning (MQMAS), were developed for quadrupolar nuclei with half-integer spins.

This article describes the basic theory, the line shape for first- and second-order quadrupole broadened spectra with and without MAS of the powder, the most important experimental techniques for the study of quadrupolar nuclei with half-integer spins such as DOR, DAS, MQMAS, echo and nutation techniques, and finally some recent developments of the deuterium NMR. Since single-crystal studies are not considered in this article,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

we refer to recent papers by Vosegaard et al.,[1,2] which describe the up-to-date experimental techniques, and the review by Kanert and Mehring, 'Static Quadrupole Effects in Disordered Cubic Solids'.[3] The classical papers of Volkoff et al., 'Nuclear Electric Quadrupole Interaction in Single Crystals',[4,5] of Das and Saha, 'Electric Quadrupole Interaction and Spin Echoes in Crystals',[6] and the classical review by Cohen and Reif, 'Quadrupole Effects in NMR Studies of Solids',[7] should also be mentioned.

Some other reviews on solid-state NMR of quadrupolar nuclei include mainly powder studies. They are, in chronological order, Spiess, 'Deuteron NMR – A New Tool for Studying Chain Mobility and Orientation in Polymers',[8] Freude and Haase, 'Quadrupole Effects in Solid-state NMR',[9] Jäger, 'Satellite Transition Spectroscopy of Quadrupolar Nuclei',[10] Hoatson and Vold, '^{2}H-NMR Spectroscopy of Solids and Liquid Crystals',[11] Chmelka and Zwanziger, 'Solid-state NMR Line Narrowing Methods for Quadrupolar Nuclei – Double Rotation and Dynamic-angle Spinning',[12] Chandrakumar, 'Spin-1 NMR',[13] Ulrich and Grage, 'H-2 NMR (Solid-state NMR of Polymers)',[14] and Smith and van Eck, 'Recent Advances in Experimental Solid-state NMR Methodology for Half-integer Spin Quadrupolar Nuclei'.[15]

For the spin–lattice relaxation of quadrupole spins, which is not considered in the present article, we refer to Chapter 7 in Freude and Haase.[9] A survey of nuclear quadrupole frequency data published before the end of 1982 is given by Chihara and Nakamura in *Landolt-Börnstein*, Vol. 20.[16] Values of the chemical shift of quadrupolar nuclei in solids can be found in books such as *Multinuclear NMR* edited by Mason.[17] For ^{2}H NMR we refer to other publications.[11,13,14,18] In section 10 we present as a continuation of Freude and Haase[9] some electric field gradient and chemical shift data for the most commonly studied quadrupolar nuclei with half-integer spins ^{27}Al, ^{23}Na and ^{17}O. A few references to recent solid-state NMR studies of some other quadrupolar nuclei are also given.

2 BASIC THEORY

At least three Cartesian coordinate systems were used for the description of the NMR interactions: the external magnetic field in the z-direction and the radiofrequency (RF) coil in the x-direction determine the laboratory axis system (LAB) with the coordinates (x, y, z). The interaction representation is based on the coordinates $x_\mathrm{i} = x\cos wt$, $y_\mathrm{i} = \pm y\cos wt$, $z_\mathrm{i} = z$. This system rotates around the direction of the external magnetic field with the applied RF $w = 2\pi\nu$. The microscopic properties of the complex under study such as the dipolar interactions, the anisotropy of the chemical shift or the electric field gradient are described in the principal axis system (PAS). The principal axes (X, Y, Z) are related to the structure of the complex, e.g. for the dipolar interaction the Z-direction is parallel to the internuclear vector.

The various interactions of the nuclear spin $\boldsymbol{I}$ can be described by the corresponding Hamiltonians. In the following we use I for the spin number and $\boldsymbol{I}$ for the spin angular momentum vector and also for the corresponding vector operator (Hamiltonian). The interaction of a nuclear spin with an external magnetic field $\boldsymbol{B}$ gives the Hamiltonian (Equation 1)

$$= \boldsymbol{IZB} \tag{1}$$

$\boldsymbol{Z} = -\gamma\hbar\boldsymbol{1}$ includes the unity matrix $\boldsymbol{1}$, the gyromagnetic ratio γ of the nuclear spins and Planck's constant, $h = 2\pi\hbar$.[19] For the case of a static external magnetic field B_0 pointing in the z-direction and the application of an RF field $B_x(t) = 2B_\mathrm{rf}\cos(wt)$ in the x-direction, we have for the external interactions Equation (2):

$$_0 + \quad _\mathrm{rf} = \hbar w_\mathrm{L} I_z + 2\hbar w_\mathrm{rf}\cos(wt) I_x \tag{2}$$

where $w_\mathrm{L} = 2\pi\nu_\mathrm{L} = -\gamma B_0$ denotes the Larmor frequency and the nutation frequency w_rf is defined as $w_\mathrm{rf} = -\gamma B_\mathrm{rf}$. The transformation from the laboratory frame to the rotating frame gives, by neglecting the part which oscillates with twice the RF, Equation (3):

$$_{0,\mathrm{i}} + \quad _{\mathrm{rf,i}} = \hbar\Delta w I_z + \hbar w_\mathrm{rf} I_x \tag{3}$$

where $\Delta w = w_\mathrm{L} - w$ denotes the resonance offset.

In addition to the external interactions there exist internal interactions of a nuclear spin, which can be efficiently expressed in the notation of irreducible tensor operators.[20] The definition and the most important feature (coordinate rotation) of an irreducible tensor operator of rank k with $(2k + 1)$ components is given by Equation (4):

$$P_R T_q^{(k)} P_R^{-1} = \sum_{q'=-k}^{+k} T_{q'}^{(k)} D_{q'q}^{(k)}(R) \tag{4}$$

where $D_{q'q}^{(k)}(R)$ denote the matrix elements of the irreducible representation of the group $\mathrm{O}^+(3)$ of the ordinary three-dimensional rotations. The transformation operator P_R is defined as $P_R = \exp\{\mathrm{i}\Omega\boldsymbol{e}_\mathrm{r}\boldsymbol{J}\}$, where $\boldsymbol{J}$ is the total angular momentum operator, $\boldsymbol{e}_\mathrm{r}$ is the unit vector pointing in the direction of rotation and Ω represents the angle of rotation. The rotation transforms the eigenfunctions, $|Im\rangle$, of the angular momentum operator

I_z as Equation (5):

$$P_R|Im\rangle = \sum_{m'} |Im'\rangle D^{(k)}_{m'm}(R) \tag{5}$$

The matrix elements $D^{(k)}_{q'q}(R)$ (Wigner matrices) can be written as Equation (6):

$$D^{(k)}_{m'm}(\alpha, \beta, \gamma) = \exp\{i\alpha m'\} d^{(k)}_{m'm}(\beta) \exp\{i\gamma m\} \tag{6}$$

with the Euler angles α, β and γ. The elements $d^{(k)}_{m'm}(\beta)$ of the reduced Wigner matrices depend on the angle β only. They are given in Table 1 for rank 2; values for rank 4 can be found in Zheng et al.[21] As presented in Figure 1, a positive rotation to a frame (x, y, z) about the Euler angles includes the rotation α about the original z axis, the rotation β about the obtained y' axis and the rotation γ about the final z'' axis [cf. Rose[22]].

Table 1 Reduced Wigner matrix $d^{(k)}_{qq'}(\beta)$ for rank 2[a]

q'/q	-2	-1	0	1	2
-2	a^4	$-\sqrt{2}a^2c$	$\sqrt{\frac{3}{2}}c^2$	$-\sqrt{2}b^2c$	b^4
-1	$\sqrt{2}a^2c$	$a^2(2d-1)$	$-\sqrt{3}cd$	$b^2(2d+1)$	$-\sqrt{2}b^2c$
0	$\sqrt{\frac{3}{2}}c^2$	$\sqrt{3}cd$	$\frac{1}{2}(3d^2-1)$	$-\sqrt{3}cd$	$\sqrt{\frac{3}{2}}c^2$
1	$\sqrt{2}b^2c$	$b^2(2d+1)$	$\sqrt{3}cd$	$a^2(2d-1)$	$-\sqrt{2}a^2c$
2	b^4	$\sqrt{2}b^2c$	$\sqrt{\frac{3}{2}}c^2$	$\sqrt{2}a^2c$	a^4

[a] The following abbreviations are used: $a = \cos\frac{\beta}{2}$, $b = \sin\frac{\beta}{2}$, $c = \sqrt{2}ab = \frac{1}{\sqrt{2}}\sin\beta$ and $d = a^2 - b^2 = \cos\beta$.

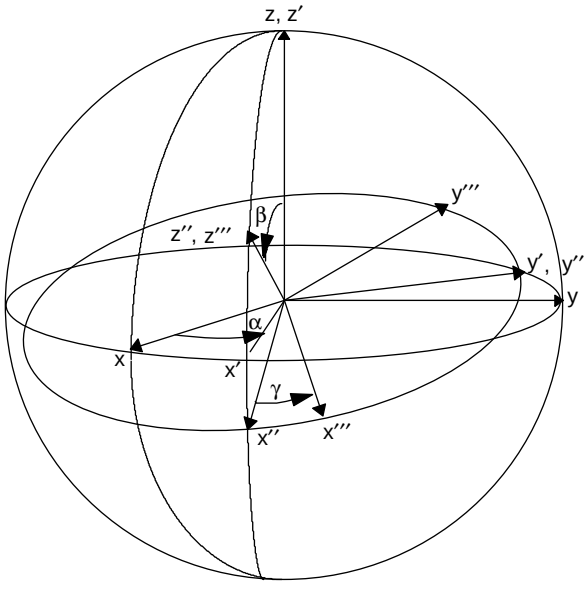

Figure 1 Euler angles.

The commutator shows another important feature of an irreducible tensor operator (Equations 7–9):[20]

$$[I_z, T^{(k)}_q] = qT^{(k)}_q \tag{7}$$

and

$$[I_{\pm1}, T^{(k)}_q] = \mp\sqrt{\tfrac{1}{2}\{k(k+1) - q(q\pm1)\}}\, T^{(k)}_{q\pm1} \tag{8}$$

with

$$I_{\pm1} = \mp\tfrac{1}{\sqrt{2}}(I_x \pm iI_y) \tag{9}$$

The tensor product $\boldsymbol{U}^{(\boldsymbol{K})}_{\boldsymbol{Q}}$ of two operators $T^{(k)}_q$ and $V^{(k)}_q$ is defined according to Equation (10):[20]

$$\boldsymbol{U}^{(\boldsymbol{K})}_{\boldsymbol{Q}} = \sum_{qq'} T^{(k)}_q V^{(k')}_{q'} \langle kk'qq'|kk'KQ\rangle \tag{10}$$

where $\langle kk'qq'|kk'KQ\rangle$ are Clebsch–Gordan coefficients.[20] The scalar product of the operators $T^{(k)}_q$ and $V^{(k)}_q$ is simpler (Equation 11):

$$T^{(k)}_q \cdot V^{(k)}_q = \sum_{q=-k}^{+k} (-1)^q T^{(k)}_q V^{(k)}_{-q} \tag{11}$$

Fortunately, most internal interactions in the NMR can be written in this form.[23] In the following the operators $T^{(k)}_q$ and $V^{(k)}_q$ act on two noninteracting systems, the nuclear spin coordinates and the spatial coordinates (lattice parameters), respectively.

In the strong external magnetic field it holds for the internal interactions by neglecting the nonsecular part of the corresponding Hamiltonian that (Equation 12)

$$[\ _0, \ _{\text{internal}}] = 0 \tag{12}$$

Comparison of Equation (12) with Equation (7) shows that only tensor elements with $q = 0$ contribute to the secular part of the internal Hamiltonians in a strong external magnetic field, if the system is described in the laboratory frame (LAB). However, the microscopic properties of the system are described in the PAS and a rotation of the coordinates from the PAS to the LAB by means of the Wigner matrix elements $D^{(k)}_{q'0}(R)$ must be performed. For example, the element $q = 0$ of the tensor $V^{(k)}$ is obtained by Equation (13):

$$V^{(k)}_0(\text{LAB}) = \sum_{q'=-k}^{k} V^{(k)}_{q'}(\text{PAS}) D^{(k)}_{q'0}(R) \tag{13}$$

Operators of rank 0 are invariant with respect to rotations. For the chemical shift we have (Equations 14–16)

$$T^{(0)}_0 = \tfrac{-1}{\sqrt{3}} I_z B_0 \tag{14}$$

$$V^{(0)}_0 = \sqrt{3}\sigma_{\text{iso}} \tag{15}$$

and

$$\begin{matrix}\text{isotropic}\\ \text{chemical shift}\end{matrix} = -\gamma\hbar\sigma_{\text{iso}} I_z B_0 \tag{16}$$

where σ_{iso} is the isotropic part of the shielding tensor (Equation 17):

$$\sigma_{\text{iso}} = \frac{\sigma_{XX} + \sigma_{YY} + \sigma_{ZZ}}{3} \tag{17}$$

Rank 0 operators do not contribute to dipolar interactions or quadrupolar interaction in first order. The contribution of the rank 1 operators can be neglected for all considered interactions. Therefore, anisotropy of the chemical shift, dipolar interactions and quadrupolar interactions in first order can be described by rank 2 operators in the form of Equation (18):

$$= C \sum_{q=-2}^{+2} (-1)^q T_q^{(2)} V_{-q}^{(2)} \tag{18}$$

Several contributions as described in Table 2 can be superimposed. The elements of the shielding tensor σ (trace $3\sigma_{\text{iso}}$) and of the traceless tensor of the electric field gradient $\boldsymbol{V}$ [which must not confused with the operator $V^{(k)}$] are given in the PAS. Parameters of the anisotropy are $\delta = \sigma_{ZZ} - \sigma_{\text{iso}}$ for the chemical shift and $V_{ZZ} = eq$ for the electric field gradient, where e denotes the elementary charge and the value q alone has no physical meaning in SI units. Q is the quadrupolar moment and eQ is called the electric quadrupolar moment. The asymmetry parameter η is in the range $0 \le \eta \le 1$. With the convention $|V_{ZZ}| \ge |V_{YY}| \ge |V_{XX}|$ or $|\sigma_{ZZ} - \sigma_{\text{iso}}| \ge |\sigma_{YY} - \sigma_{\text{iso}}| \ge |\sigma_{XX} - \sigma_{\text{iso}}|$, we obtain (Equations 19 and 20)

$$\eta = \frac{V_{XX} - V_{YY}}{V_{ZZ}} \tag{19}$$

or

$$\eta = \frac{\sigma_{XX} - \sigma_{YY}}{\sigma_{ZZ} - \sigma_{\text{iso}}} \tag{20}$$

Equation (21):

$$I_{\pm 1} = \mp\tfrac{1}{\sqrt{2}}(I_x \pm \mathrm{i} I_y) \tag{21}$$

and Equation (22):

$$F(\alpha, \beta, \eta) = \left(\frac{3\cos^2\beta - 1}{2} + \frac{\eta}{2}\sin^2\beta\cos 2\alpha \right. \tag{22}$$

are used in Table 2. The last row in Table 2 gives components which were transformed from the PAS into the LAB by the Euler angles α and β using Equation (13) for rank 2. With Equation (18) we obtain for the secular part of the Hamiltonian in the LAB for quadrupolar interaction in the first order (Equation 23)

$$\begin{aligned} \text{Q} = {} & \frac{eQV_{ZZ}}{4I(2I-1)}[3I_z^2 - I(I+1)] \\ & \times \left(\frac{3\cos^2\beta - 1}{2} + \frac{\eta}{2}\sin^2\beta\cos 2\alpha \right. \end{aligned} \tag{23}$$

The quadrupolar coupling constant C_{qcc} is commonly defined according to Equation (24):

$$C_{\text{qcc}} = \frac{e^2 qQ}{h} \tag{24}$$

Table 2 Some contributions to the Hamiltonian[a]

Parameter	Chemical shift	Dipolar interaction between $\boldsymbol{I_i}$ and $\boldsymbol{I_k}$	Quadrupole interaction
C	$\hbar\gamma$	$-2\gamma_i\gamma_k\hbar^2\frac{\mu_0}{4\pi}$	$\frac{eQ}{2I(2I-1)}$
$T_0^{(2)}$	$\sqrt{\frac{2}{3}}I_z B_0$	$\sqrt{\frac{1}{6}}(3I_{zi}I_{zk} - \boldsymbol{I}_i\boldsymbol{I}_k)$	$\sqrt{\frac{1}{6}}[3I_z^2 - I(I+1)]$
$T_{\pm 1}^{(2)}$	$\sqrt{\frac{1}{2}}I_{\pm 1}B_0$	$\sqrt{\frac{1}{2}}(I_{\pm 1i}I_{zk} + I_{zi}I_{\pm 1k})$	$\sqrt{\frac{1}{2}}(I_{\pm 1}I_z + I_{-z}I_{\pm 1})$
$T_{\pm 2}^{(2)}$	0	$I_{\pm 1i}I_{\pm 1k}$	$I_{\pm 1}^2$
$V_0^{(2)\text{PAS}}$	$\sqrt{\frac{3}{2}}(\sigma_{ZZ} - \sigma_{\text{iso}}) = \sqrt{\frac{3}{2}}\delta$	$\sqrt{\frac{3}{2}}r_{ik}^{-3}$	$\sqrt{\frac{3}{2}}V_{ZZ} = \sqrt{\frac{3}{2}}eq$
$V_{\pm 1}^{(2)\text{PAS}}$	0	0	0
$V_{\pm 2}^{(2)\text{PAS}}$	$\frac{1}{2}(\sigma_{XX} - \sigma_{YY}) = \frac{1}{2}\eta\delta$	0	$\frac{1}{2}(V_{XX} - V_{YY}) = \frac{1}{2}\eta eq$
$V_0^{(2)\text{LAB}}$	$\sqrt{\frac{3}{2}}\delta F(\alpha, \beta, \eta)$	$\sqrt{\frac{3}{2}}\frac{1}{r_{ik}^3}\left(\frac{3\cos^2\theta_{ik} - 1}{2}\right.$	$\sqrt{\frac{3}{2}}eqF(\alpha, \beta, \eta)$

[a] The dipolar interaction is the homonuclear one. For heteronuclear dipolar interactions $T_0^{(2)}$ must be substituted by $\sqrt{1/6}I_{zi}I_{zk}$.

However, for the quadrupolar frequency, ν_Q or w_Q, different definitions exist in the literature. We use the values according to Equation (25):

$$\nu_Q = \frac{3e^2qQ}{2I(2I-1)h} = \frac{3C_{qcc}}{2I(2I-1)} \tag{25}$$

or Equation (26):

$$w_Q = \frac{3e^2qQ}{2I(2I-1)\hbar} \tag{26}$$

which were introduced for half-integer spin nuclei in the field of nuclear quadrupolar resonance (NQR) by Das and Hahn,[24] and established by Abragam[19] also for NMR. For different definitions in the case of half-integer spin nuclei, see Freude and Haase[9] (p. 33) and some papers following Frydman and Harwood.[25] For integer spins even less notational consistency exists, cf. Hoatson and Vold[11] (pp. 5 and 8). Equation (24) is still valid, but the definition of the quadrupolar frequency after Equation (26) is seldom used for spin-1 nuclei.

By substituting $V_{ZZ} = eq$ and using the angular dependent quadrupole frequency (Equation 27)

$$\nu'_Q = \nu_Q \left(\frac{3\cos^2\beta - 1}{2} + \frac{\eta}{2}\sin^2\beta\cos 2\alpha \right. \tag{27}$$

one can write (Equation 28)

$$\mathcal{H}_Q^{\text{first-order}} = \frac{h\nu'_Q}{6}[3I_z^2 - I(I+1)] \tag{28}$$

Equation (28) represents the first-order contribution of the quadrupolar interaction in the strong external magnetic field. From the second-order contribution the secular part with respect to I_z is given by Equation (29):

$$\mathcal{H}_Q^{\text{second-order}} = \frac{h\nu_Q^2}{9\nu_L}\left\{ 2I_Z \quad 2I_Z^2 - I(I+1) + \frac{1}{4} \quad V_{-1}V_1 \right.$$
$$\left. + I_Z \quad I_Z^2 - I(I+1) + \frac{1}{2} \quad V_{-2}V_2 \right\} \tag{29}$$

The components $V_{p,p'}$ in Equation (29) correspond to the LAB. They can be obtained from the components in the PAS by means of Equation (13).

The homonuclear dipolar interaction of a pair (i, k) of spins with the distance r_{ik} and an angle θ between the internuclear vector and the direction of the external magnetic field is described by the Hamiltonian (Equation 30)

$$\mathcal{H}_D = \frac{\mu_0}{4\pi}\gamma_i\gamma_k\hbar^2 \left(\frac{3\cos^2\theta_{ik} - 1}{2r_{ik}^3} \quad (3I_{z,i}I_{z,k} - \boldsymbol{I}_i\boldsymbol{I}_k) \right. \tag{30}$$

and it holds that $\gamma_i = \gamma_k$. For heteronuclear dipolar interactions we have $\gamma_i \neq \gamma_k$ and $(3I_{z,i}I_{z,k} - \boldsymbol{I}_i\boldsymbol{I}_k)$ in Equation (30) must be substituted by $(I_{z,i}I_{z,k})$.

The Hamiltonian of the chemical shift is the sum of the isotropic and the anisotropic contributions (Equation 31):

$$\mathcal{H}_{CS} = \gamma\hbar B_0 I_z \quad \sigma_{iso} + \delta \left(\frac{3\cos^2\beta - 1}{2} + \frac{\eta}{2}\sin^2\beta\cos 2\alpha \right. \tag{31}$$

3 EXCITATION

The energy of a spin interaction can be described by the size of the corresponding Hamiltonian $\| \mathcal{H} \|$. For NMR the usual sequence is $\| \mathcal{H}_0 \| \gg \| \mathcal{H}_{rf} \| \gg \| \mathcal{H}_Q \|$, $\| \mathcal{H}_D \|$, $\| \mathcal{H}_{CSA} \|$. Expressed in frequencies: the Larmor frequency ($\| \mathcal{H}_0 \|$), cf. Equation (2), is of the order of magnitude of 10–1000 MHz, the nutation frequency ($\| \mathcal{H}_{rf} \|$), cf. Equation (2), is of the order of magnitude of 1–400 kHz and the internal interactions are smaller than the RF field strength. However, the latter is sometimes not correct for quadrupolar nuclei in solids. We can assume the relation $\| \mathcal{H}_0 \| \gg \| \mathcal{H}_Q \| \gg \| \mathcal{H}_D \|$, $\| \mathcal{H}_{CSA} \|$ and have to distinguish the well-defined cases $\| \mathcal{H}_{rf} \| \gg \| \mathcal{H}_Q \|$ (*) or $\| \mathcal{H}_{rf} \| \ll \| \mathcal{H}_Q \|$ (**) and the ill-defined intermediate case (***). A so-called hard pulse can perform a nonselective excitation of the whole quadrupolar broadened spectrum, if the RF field strength meets (*). The soft pulse (**) excitation is limited to any single transition ($m = -I, -I+1, \ldots, I-1$) in a single crystal or to the central transition ($m = -\frac{1}{2}$) for powdered materials. Here the transition m denotes the transition between the magnetic quantum numbers $m \longleftrightarrow m+1$.

The excitation of quadrupolar nuclei with half-integer spins is discussed in detail by Freude and Haase.[9] For a short description of the problem of excitation of any NMR transition, we start from the basic equation for the frequency spectrum of a rectangular pulse with the duration τ and the carrier frequency ν_0 with $\delta\nu = \nu - \nu_0$ (Equation 32):

$$f(\delta\nu) = \frac{1}{\tau}\int_{-\tau/2}^{+\tau/2} \cos(2\pi\delta\nu t)\,\mathrm{d}t = \frac{\sin(\pi\delta\nu\tau)}{\pi\delta\nu\tau} \tag{32}$$

It has its maximum $f(\delta\nu) = 1$ for $\nu = \nu_0$ and the first nodes in the frequency spectrum occur at $\delta\nu = \pm 1/\tau$. The spectral energy density E is proportional to the square of the RF field strength which includes the offset dependence of Equation (32). If we define the usable bandwidth of excitation $\delta\nu_{1/2}$ in analogy to electronics by Equation (33):

$$E\left(\nu = \pm\delta\nu_{1/2}\right) = \tfrac{1}{2}E(\nu = \nu_0) \tag{33}$$

or Equation (34):

$$f^2(\nu = \pm\delta\nu_{1/2}) = \tfrac{1}{2}f^2(\nu = \nu_0) \tag{34}$$

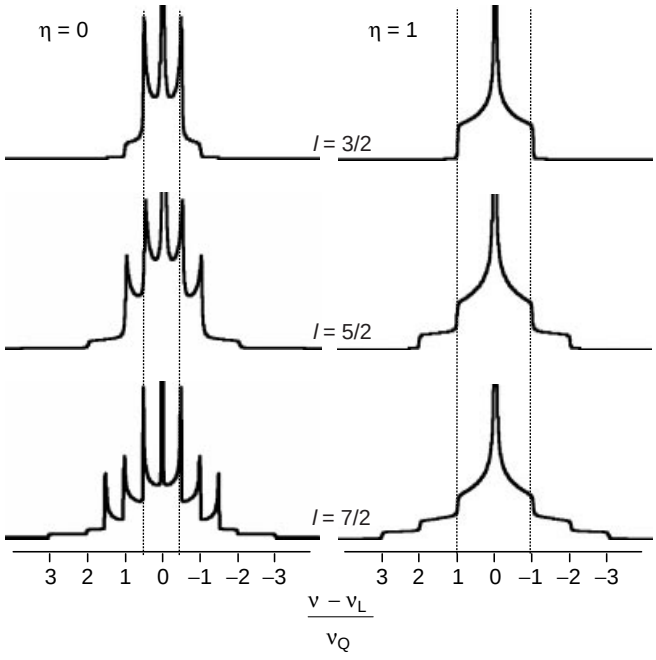

Figure 2 Examples of the static line shape (without MAS) of first-order quadrupolar broadened powder spectra for nonselective excitation. The central line is not fully represented (cut off). The theoretical line shapes are slightly Gaussian broadened.

the bandwidth of excitation $\delta\nu_{1/2} \approx 0.886/\tau$ follows from Equation (32). It should be noted that also the quality factor of the probe, $Q = \nu/\delta\nu_{\text{probe}}$, limits the bandwidth of excitation independently of the applied RF field strength or pulse duration. A superposition of the free induction decay (FID) of the NMR signals (liquid sample excited by a very short pulse) for some equidistant values of the resonance offset (without retuning the probe) shows easily the bandwidth $\delta\nu_{\text{probe}}$ of the probe. The spectral range of resonances for first-order quadrupolar shifts, which should be excited, amounts to $\nu_Q(2I-1)$ or $3C_{\text{qcc}}/2I$ (cf. Figure 2). The superposition of both limitations gives the conditions in Equation (35):

$$\frac{1}{\tau} > 1.13\nu_Q(2I-1) \approx \frac{1.4C_{\text{qcc}}}{2I} \tag{35}$$

and Equation (36):

$$\delta\nu_{\text{probe}} > \nu_Q(2I-1) = \frac{3C_{\text{qcc}}}{2I} \tag{36}$$

for a ''nonselective'' excitation of the spectrum. For increasing pulse width τ or decreasing bandwidth $\delta\nu_{\text{probe}}$ of the probe, a ''partly selective'', i.e. nonuniform, excitation of the transitions or a ''selective'' excitation of one transition only takes place.

If the single pulse is substituted by a composite pulse, a designed broadband or narrowband excitation can be achieved.[26]

The intensity of the FID $G(t=0)$ after the pulse with the RF field strength w_{rf} and the duration τ is, for nonselective excitation of all transitions $m \leftrightarrow m+1$, given by Equation (37):[9]

$$G_{m,m+1}^{\text{nonselective}}(0) = \frac{3[I(I+1) - m(m+1)]}{2I(I+1)(2I+1)} \sin w_{\text{rf}}\tau \tag{37}$$

Equation (37) gives also the relative intensities of all transitions, e.g. 12/30, 9/35 and 4/21 for the central lines in the case of nonselective excitation of the $I = \frac{3}{2}, \frac{5}{2}$ and $\frac{7}{2}$ nuclei, respectively. The selective excitation of a single transition can be described by Equation (38):[9]

$$G_{m,m+1}^{\text{selective}}(0) = \frac{3\sqrt{I(I+1) - m(m+1)}}{2I(I+1)(2I+1)} \times \sin \sqrt{I(I+1) - m(m+1)}\, w_{\text{rf}}\tau \tag{38}$$

Comparison with Equation (37) reveals that the maximum observed intensity is reduced by $\sqrt{I(I+1) - m(m+1)}$, but the effective nutation frequency is enhanced by the same value. For the central transition, $m = -\frac{1}{2}$, we obtain Equation (39):

$$\nu_{\text{rf}}^{\text{eff}} = (I + \tfrac{1}{2})\nu_{\text{rf}} \tag{39}$$

Thus, for the selective excitation of the central transition, the optimum pulse duration is equal to the duration of a nonselective $\pi/2$ pulse divided by $I + \frac{1}{2}$. The RF power should be as low as possible. Second-order quadrupolar broadening, the chemical shift anisotropy and dipole broadening give the lower limit for the RF power which should be applied, in order to excite the full central transition in powder spectra. Then, only a small portion of the satellite powder spectra is also excited. The partly selective excitation of more than one transition can be treated only numerically.[9]

The problem of ill-defined excitation in cases of strong quadrupolar interaction can be overcome by adiabatic frequency sweeps. Adiabatic passages are well known in NMR.[19] Kentgens[27] established in 1991 the quantitative excitation of a half-integer spin system by a frequency-stepped adiabatic half-passage (FSAHP) on the basis of an approach of Sindorf and Bartuska[28] and introduced in 1999 double-frequency sweeps (DFS) in static, MAS and MQMAS NMR experiments on the basis of an appropriate time-dependent amplitude modulation to the carrier frequency, which was applied before by Fu et al.[29] to integer spin systems in single crystals.

In the FSAHP, the spin system is far off-resonance at the beginning of the irradiation. The frequency is then

stepped through the region of resonances slowly enough that the density operator can follow the Hamiltonian. Switching off the RF power at the resonance position of the central transition creates a single-quantum coherence like a $\pi/2$ pulse applied to a spin-$\frac{1}{2}$ system. A full passage would be comparable to a nonselective π pulse. A frequency sweep that adiabatically inverts the satellite transitions, in order to transfer magnetization to the central levels and increase the intensity of the corresponding signal, was introduced by Haase and Conradi.[30] The combination of this population enhancement with cross-polarization (CP) from ^{27}Al to ^{17}O was demonstrated.[31] A DFS can be generated by an amplitude modulation of the RF which causes two sidebands that are swept from a start frequency to a final frequency during the pulsing.[32] The authors claim that the DFS technique enhances the signal intensity also in MQMAS experiments.

The exclusive detection of the $\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$ transitions, which is based on a central-transition signal enhancement by means of an adiabatic RF sweep was introduced by McDowell et al.[33] in order to simplify the spectra and increase the sensitivity. Another approach to resolution enhancement by population transfer between Zeeman levels of the quadrupolar spin system with selective pulses and adiabatic frequency sweeps was first applied to ^{63}Cu in superconductors by Haase et al.[34]

Another NMR technique that uses adiabatic passage in combination with rotational echo and double resonance was introduced by Gullion.[35] He combined the principles of rotational-echo double resonance (REDOR),[36] with the transfer-of-populations double-resonance (TRAPDOR) developed by Grey et al.[37] The so-called REAPDOR NMR technique allows first, like TRAPDOR, the indirect detection of signals which are too broad to be directly observable for the single-resonance observation of the quadrupolar nucleus and second, like REDOR, the measurement of distances between spin pairs.[35]

CP excites a spin system by polarization transfer from another spin system. CP was introduced by Hartmann and Hahn[38] and is described in detail in the textbook by Slichter.[39] S. Vega[40] first applied static CP to half-integer nuclei and A.J. Vega[41,42] considered the spin dynamics of cross-polarization magic-angle spinning (CPMAS) NMR as a function of the ratio $\nu_{rf}^2/\nu_Q\nu_{rot}$. Applications until 1992 are reviewed in Chapter 6.2 of Freude and Haase.[9] Recent discussions about the spin-locking efficiency of the CPMAS NMR of half-integer quadrupolar nuclei were presented by Hayashi,[43,44] Ding and McDowell[45,46] and Sun et al.[47] The combination of CP with other NMR techniques will be considered in sections 5–7.

4 LINE SHAPE OF STATIC AND MAGIC-ANGLE SPINNING NUCLEAR MAGNETIC RESONANCE SPECTRA

The following considerations concern the angular dependent quadrupolar shift. Assuming the resonance offset to be zero, the quadrupolar shift should be given as $\nu - \nu_L$. For simplicity, we omit $-\nu_L$, the subtraction of the Larmor frequency. The conventions $\nu_{m,m+1}$ and $\nu_{m,-m}$ for single-quantum transitions and symmetric transitions, respectively, are used here in agreement with the majority of the literature. For the central transition $\nu_{-1/2,+1/2}$ we have $m = -\frac{1}{2}$. Then, the first-order quadrupolar shift becomes, with ν'_Q from Equation (27), Equation (40):

$$\nu_{m,m+1} = \nu'_Q(m + \tfrac{1}{2}) \tag{40}$$

for single-quantum transitions, which means zero for the central transition $m = -\frac{1}{2}$. The first-order quadrupolar contribution is zero for all symmetric transitions $m \leftrightarrow -m$. See Equation (28) and note that $I_z \equiv m$ is a number in this case.

The second-order quadrupolar shift can be obtained by means of Equation (29) as Equation (41):

$$\nu_{m,m+1} = -\frac{\nu_Q^2}{18\nu_L}\{[24m(m+1) - 4I(I+1) + 9]V_1V_{-1} + [6m(m+1) - 2I(I+1) + 3]V_2V_{-2}\} \tag{41}$$

or Equation (42):

$$\nu_{m,-m} = -\frac{m\nu_Q^2}{18\nu_L}\{[4I(I+1) - 8m^2 - 1]V_1V_{-1} + [2I(I+1) - 2m^2 - 1]V_2V_{-2}\} \tag{42}$$

for single or symmetric quantum transitions, respectively. The components V_j are given in the LAB and can be described, with the Wigner matrices of rank 2, as functions of corresponding values in the PAS, which are given in Table 2. The V_jV_{-j} terms in Equations (41) and (42), therefore, can be written also as Wigner matrices for which the rank goes up to 4. Amoureux[48] gave a corresponding equation with the coefficients for the transformation of Equation (42) from the LAB into the PAS.

The second-order quadrupolar shift of the central transition can also be obtained with $m = -\frac{1}{2}$ from Equations (41) and (42) (Equation 43):

$$\nu_{-1/2,+1/2} = -\frac{\nu_Q^2}{6\nu_L}\left[I(I+1) - \frac{3}{4}\right] \times (A\cos^4\beta + B\cos^2\beta + C) \tag{43}$$

Table 3 Functions A, B and C for the second-order static and MAS line shapes of the central transition (cf. Equation 43)

Function	Static	MAS
A	$-\frac{27}{8}-\frac{9}{4}\eta\cos 2\alpha-\frac{3}{8}\eta^2\cos^2 2\alpha$	$+\frac{21}{16}-\frac{7}{8}\eta\cos 2\alpha-\frac{7}{48}\eta^2\cos^2 2\alpha$
B	$+\frac{15}{4}-\frac{1}{2}\eta^2+2\eta\cos 2\alpha+\frac{3}{4}\eta^2\cos^2 2\alpha$	$-\frac{9}{8}-\frac{1}{12}\eta^2+\eta\cos 2\alpha-\frac{7}{24}\eta^2\cos^2 2\alpha$
C	$-\frac{3}{8}+\frac{1}{3}\eta^2+\frac{1}{4}\eta\cos 2\alpha-\frac{3}{8}\eta^2\cos^2 2\alpha$	$+\frac{5}{16}-\frac{1}{8}\eta\cos 2\alpha+\frac{7}{48}\eta^2\cos^2 2\alpha$

The components V_pV_{-p} are substituted now by functions depending on the Euler angles α and β and the asymmetry parameter η. The functions *A, B, C* (Table 3, static) were first used by Narita et al.[49] in order to calculate the second-order powder pattern of the central transition for a static (nonrotating) sample. Müller[50] showed that in the case of very fast MAS of the samples Equation (43) holds with other functions *A, B, C* (Table 3, MAS).

In order to deduce quadrupole parameters from experimentally obtained spectra, it is usual to fit the experimental spectrum with calculated spectra using a utility program of the spectrometer software. Figures 2 and 3 show spectra, which were obtained with the program WINFIT from Bruker. If the spinning speed is not large compared with the second-order shift, the average Hamiltonian can be used for numerical calculations.

Equation (43) allows an analytical determination of the center of gravity $\nu_{cg} \equiv \nu_{iso}$ for the powder pattern, which is called the isotropic quadrupolar shift of the signal of the central transition. This shift is not influenced by a rotation of the sample. We denote it as $\nu_{iso\ Q}$ (Equation 44):

$$\nu_{iso\ Q} = -\frac{\nu_Q^2}{30\nu_L}\ I(I+1)-\frac{3}{4}\ \left(1+\frac{\eta^2}{3}\right. \qquad (44)$$

An important parameter of the powder line shape $f(\nu)$ is the second moment defined by Equation (45):

$$M_2 = \int (\nu-\nu_{iso})^2 f(\nu)\,d\nu \qquad (45)$$

with $\int f(\nu)\,d\nu = 1$. The dimension Hz^2 of M_2 changes to s^{-2} or T^2 if the line shape is given as $f(w)$ or $f(B)$, respectively. The square root of M_2 characterizes the second-order broadening of the signal. Equations (46–48) give the second moments as a function of ν_{iso} for the static and for the MAS spectrum and in addition the narrowing factor, which can be achieved by the application of MAS to the second-order quadrupolar broadening of the central transition:

$$M_2^{static} = \nu_{iso}^2\frac{23}{7} \qquad (46)$$

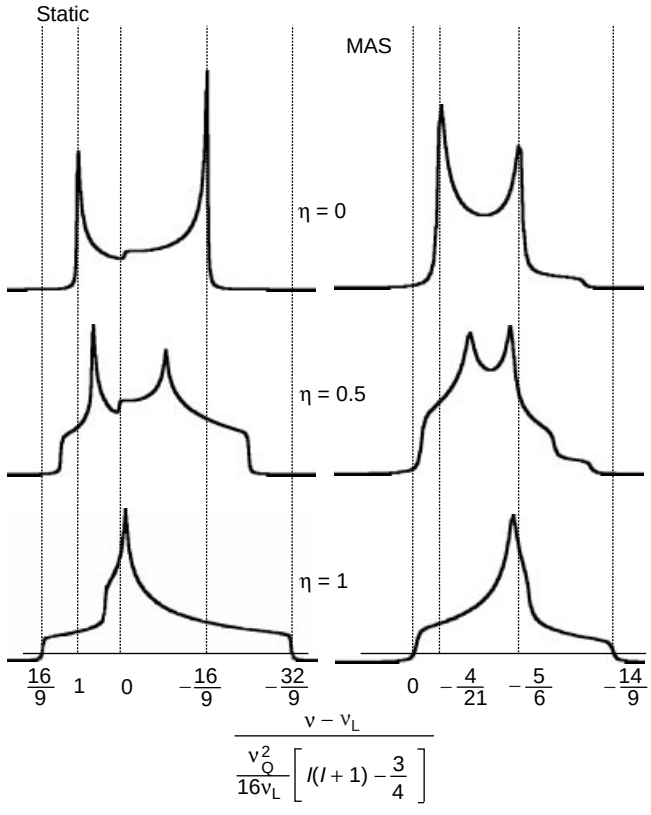

Figure 3 Examples of the line shapes, without and with MAS, of second-order quadrupolar broadened powder spectra of the central transition. The theoretical line shapes are slightly Gaussian broadened.

$$M_2^{MAS} = \nu_{iso}^2\frac{1}{4} \qquad (47)$$

$$\frac{M_2^{static}}{M_2^{MAS}} = \sqrt{\frac{92}{7}} \approx 3.6 \qquad (48)$$

Equations (46–48) in combination with Equation (44) allow the determination of $\nu_Q\sqrt{1+\eta^2/3}$ from the second moment of the line shape, if the latter is exclusively broadened by second-order quadrupolar interaction. However, other contributions to the second moment of the static line shape should also be considered. The

anisotropy of the chemical shift gives Equation (49):

$$M_2^{\mathrm{csa}} = \frac{4}{9}\frac{(\Delta\sigma\nu_{\mathrm{L}})^2}{5}\left(1+\frac{\eta^2}{3}\right. \tag{49}$$

where $\Delta\sigma = \sigma_{ZZ} - (\sigma_{XX} - \sigma_{YY})/2$ denotes the total anisotropy and η is the asymmetry parameter (cf. Equations 20 and 21) and ν_{iso} in Equation (45) must be taken as the isotropic value of the chemical shift.

If the dipolar interaction is small compared with the quadrupolar interaction, the spin flipping between different transitions is prohibited, and the second moment due to the dipolar interaction is modified.[9,19] We will give here only the equations for dominating dipolar interaction. The dipolar second moment of a spin system consisting of N resonant spins of type I (homonuclear interaction) and M nonresonant spins of type S (heteronuclear interaction) can be determined by the dimensionless Equation (50):

$$\frac{M_2/\mathrm{Hz}^2}{(\gamma_{\mathrm{I}}/2\pi)^2} = \frac{M_2/\mathrm{s}^{-2}}{\gamma_{\mathrm{I}}^2} = M_2/\mathrm{T}^2 = \frac{C_{\mathrm{I}}}{r_{\mathrm{I}}^6/\mathrm{m}^6} + \frac{C_{\mathrm{S}}}{r_{\mathrm{S}}^6/\mathrm{m}^6} \tag{50}$$

with (Equations 51–54)

$$C_{\mathrm{I}} = \frac{3}{5}I(I+1)\left(\frac{\mu_0}{4\pi}\right)^2\gamma_{\mathrm{I}}^2\hbar^2 \tag{51}$$

$$C_{\mathrm{S}} = \frac{4}{15}S(S+1)\left(\frac{\mu_0}{4\pi}\right)^2\gamma_{\mathrm{S}}^2\hbar^2 \tag{52}$$

$$\frac{1}{r_{\mathrm{I}}^6} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k\neq i}^{N}\frac{1}{r_{ik}^6} \tag{53}$$

$$\frac{1}{r_{\mathrm{S}}^6} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{M}\frac{1}{r_{ik}^6} \tag{54}$$

The value C_{S} of a nonresonant nucleus can be easily calculated from the value of the same resonant nucleus C_{I} by $C_{\mathrm{S}} = C_{\mathrm{I}} \times 4/9$. If in Equation (50) the unit of the magnetic flux density T is substituted by 10^{-4} T (this corresponds to the old cgs unit gauss) and the unit m is substituted by the old unit Å, then the values C_{I} can be taken from Table 4.

Another dimensionless equation is very helpful in order to correlate the second moment of a line, which is broadened by any interaction, to the line width, which is commonly described by the full width at half-maximum (fwhm $\equiv \delta\nu_{1/2}$). Under the assumption of a gaussian line shape we obtain, for $M_2/\mathrm{s}^{-2} = \gamma_{\mathrm{I}}^2 M_2/\mathrm{T}^2 = (\gamma_{\mathrm{I}}/2\pi)^2 M_2/\mathrm{Hz}^2$, cf. Equation (50), with T_2 as the transversal relaxation time and the line width $\delta\nu$ given in hertz, Equation (55):

$$M_2/\mathrm{s}^{-2} = \frac{2}{(T_2/\mathrm{s})^2} = (\delta\nu_{1/2}/\mathrm{Hz})^2\frac{\pi^2}{\ln 4} \approx 7.12(\delta\nu_{1/2}/\mathrm{Hz})^2 \tag{55}$$

Table 4 Values of C_{I} in the equation

$$M_2/10^{-8}\mathrm{T}^2 = \left(\frac{C_{\mathrm{I}}}{r_{\mathrm{I}}^6/\mathrm{Å}^6} + \frac{C_{\mathrm{S}}}{r_{\mathrm{S}}^6/\mathrm{Å}^6}\right.$$ [a]

^{1}H	358.1	^{19}F	316.9	^{63}Cu	125.8	^{121}Sb	239.2
^{2}H	22.50	^{23}Na	125.3	^{65}Cu	144.3	^{127}I	167.2
^{7}Li	270.4	^{27}Al	283.6	^{75}As	52.52	^{133}Cs	129.4
^{9}Be	35.36	^{29}Si	14.13	^{77}Se	13.02	^{195}Pt	16.55
^{10}B	66.16	^{31}P	58.68	^{79}Br	112.4	^{199}Hg	11.38
^{11}B	184.3	^{35}Cl	17.19	^{81}Br	130.6	^{201}Hg	7.798
^{13}C	22.64	^{37}Cl	11.91	^{87}Rb	191.7	^{203}Tl	116.9
^{14}N	4.983	^{51}V	519.3	^{93}Nb	706.0	^{205}Tl	119.2
^{15}N	3.677	^{55}Mn	254.1	^{117}Sn	45.45	^{207}Pb	15.67
^{17}O	76.81	^{59}Co	419.3	^{119}Sn	49.74	^{209}Bi	305.1

[a] Values of C_{S} are given by $C_{\mathrm{S}} = C_{\mathrm{I}} \times 4/9$. The distances r_{I} and r_{S} denote in a simple model the internuclear distances of one resonant spin to another resonant spin and to a nonresonant spin, respectively. For the general case, see the definitions of r_{I} and r_{S} in Equations (53) and (54).

Now we go back to Equation (41) and consider the MAS spectra of the satellite transitions in addition to the central transition. Samoson[51] has shown that the second-order quadrupolar shift of the $m \leftrightarrow m+1$ transition can be split into an isotropic part describing the center of gravity of the quadrupolar shift and an angular-dependent part (Equation 56):

$$\nu_{m,m+1} = -\frac{\nu_{\mathrm{Q}}^2}{30\nu_{\mathrm{L}}}[I(I+1) - 9m(m+1) - 3]\left(1+\frac{\eta^2}{3}\right. - \left(\frac{\nu_{\mathrm{Q}}^2}{30\nu_{\mathrm{L}}}\right. I(I+1) - \frac{17}{3}m(m+1) - \frac{13}{6} \times \left(A\cos^4\beta + B\cos^2\beta + C\right) \tag{56}$$

with (Equations 57–59)

$$A = \frac{105}{16} - \frac{35}{8}\eta\cos 2\alpha + \frac{35}{48}\eta^2\cos^2 2\alpha \tag{57}$$

$$B = -\frac{45}{8} + \frac{5}{12}\eta^2 + 5\eta\cos 2\alpha - \frac{35}{24}\eta^2\cos^2 2\alpha \tag{58}$$

$$C = +\frac{9}{16} + \frac{1}{3}\eta^2 - \frac{5}{8}\eta\cos 2\alpha + \frac{35}{48}\eta^2\cos^2 2\alpha \tag{59}$$

For the central transition, $m = -\frac{1}{2}$, Equation (56) corresponds to Equation (43) and the isotropic, upper part in Equation (56) corresponds to Equation (44). The true value of the isotropic chemical shift can be determined from the experimentally obtained isotropic shift (center of gravity of the signal), if the isotropic quadrupolar shift can be determined as shown below.[10,51,52] For nonselective excitation, or at least partially selective excitation, MAS sidebands can be observed which are just outside of the spectral range of the static spectrum of the central transition. Their main intensity results from the $\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$ transitions, whereas the center band

results mainly from the $+\frac{1}{2} \leftrightarrow -\frac{1}{2}$ transition. The average resonance position of two equal-order sidebands can be obtained experimentally as the center of gravity of the corresponding sidebands. Thus, the difference $\Delta = \nu_{\text{iso Q}\pm1/2,\pm3/2} - \nu_{\text{iso Q}-1/2,+1/2}$ of the average resonance position of two first satellite sidebands to the center of gravity of the center band (central transition) can be measured and compared with the corresponding difference derived from Equation (56) (Equation 60):

$$\Delta = -\frac{\nu_{\text{Q}}^2}{30\nu_{\text{L}}} 9 \left(1 + \frac{\eta^2}{3}\right. \tag{60}$$

Finally, the combination of Equations (56) and (60) gives the isotropic quadrupolar shift of the center band, which for $I = \frac{5}{2}$ is equal to $\nu_{\text{iso Q}-1/2,+1/2} = \Delta 8/9$.

This procedure for the determination of the quadrupolar shift or other quadrupolar parameters was introduced by Samoson.[51] It is convenient for $I = \frac{5}{2}$ nuclei, because the linewidth of their $\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$ satellites is decreased by a factor of 0.3 with respect to the central line. Jäger used this sideband analysis for the study of various inorganic compounds[10] and denoted it SATRAS (satellite-transition spectroscopy).

Now we go back to Equation (42) and use the representation of Amoureux[48] for the shift of a symmetric transition in the case of very fast sample rotation around the magic angle $\theta = \arccos 3^{-1/2} \approx 54.74°$. Then, the contributions from the rank 2 components disappear and we obtain[53] Equation (61):

$$\begin{aligned}
\nu_{p/2,-p/2} &= \nu_{\text{iso Q}(p)} + \nu_{\text{aniso Q}(p)} \\
&= \frac{p\nu_{\text{Q}}^2(3+\eta^2)}{90\nu_{\text{L}}} \left\{ I(I+1) - \frac{3}{4}p^2 \right\} \\
&\quad - \frac{p\nu_{\text{Q}}^2}{12960\nu_{\text{L}}} \Bigg\} \quad (18+\eta^2) d_{0,0}^{(4)} \\
&\quad + \sqrt{360}\,\eta\, d_{2,0}^{(4)} \cos 2\alpha + \sqrt{70}\,\eta^2\, d_{4,0}^{(4)} \cos 4\alpha \\
&\quad \times \left\{ 36I(I+1) - 17p^2 - 10 \right\} \left\{ \left(-\frac{9}{28} \right. \right. \\
&\quad \left. \left. \times (35\cos^4\theta - 30\cos^2\theta + 3) \right\} \right.
\end{aligned} \tag{61}$$

where p denotes the quantum level pQ and symmetric coherences with the notation $p/2 \leftrightarrow -p/2$ instead of $m \leftrightarrow -m$ are considered. The Euler angles (α, β) describe the spinner axis with respect to the PAS. The elements d of the reduced Wigner matrices are related to β by Equations (62–64):

$$d_{0,0}^{(4)} = \frac{1}{8}(35\cos^4\beta - 30\cos^2\beta + 3) \tag{62}$$

$$d_{2,0}^{(4)} = \frac{\sqrt{10}}{8}(7\cos^2\beta - 1)\sin^2\beta \tag{63}$$

$$d_{4,0}^{(4)} = \frac{\sqrt{70}}{16}\sin^4\beta \tag{64}$$

Equation (61) still contains a function of the rotor angle θ in the last brackets, in order to give one basic equation for MQMAS and DOR as well: if we insert the magic angle $\theta = \arccos\sqrt{1/3} \approx 54.74°$, then the value of the last brackets is equal to one and Equation (61) can be used as the basic equation for MQMAS NMR (cf. section 6). If we consider a DOR and the second rotor angle is fixed to $\arccos\sqrt{(6 \pm \sqrt{96/5})/14} \approx 30.56°$ or $70.12°$, then the value of the last brackets in Equation (61) is equal to zero and the total anisotropic part of Equation (61) disappears. For the central transition, $p = -1$, only the isotropic quadrupolar shift remains, which can be directly observed by DOR NMR. The isotropic contribution in Equation (61) is identical with Equation (44) and with the isotropic part of Equation (56), if the central transition is considered ($p = -1$ and $m = -\frac{1}{2}$).

5 DOUBLE-ROTATION AND DYNAMIC-ANGLE SPINNING

Samoson et al.[54] succeeded in 1988 in building a double-rotor probe. The outer rotor, inclined by $\beta_2 = 54.74°$ with respect to the external field, rotated at about 400 Hz, and had a diameter of 20 mm. The sample in the inner rotor, 5 mm diameter, rotated at about 2 kHz, and the angle between both axes of rotation was $\beta_1 = 30.56°$.[54]

A torque-free rotation can be achieved only for a certain ratio of $k = w_1/w_2$. The two contributions to the inner-rotor angular velocity are w_1 and the angular velocity w_2 imposed by the outer rotor. The latter can be decomposed into one component $w_2\cos\beta_1$ along the z-axis of the inner rotor and another component $w_2\sin\beta_1$ perpendicular to this axis in y-direction, so that y-direction, axis z of the inner rotor, and axis Z of the outer rotor are in one plane. Thus, $w_x = 0$, $w_y = w_2\sin\beta_1$, $w_z = w_1 + w_2\cos\beta_1$, and the resulting angular momentum $\boldsymbol{L}$ must be in the Z-direction in order to be torque-free[55] (cf. Figure 4) (Equation 65):

$$\tan\beta_1 = \frac{J_y w_2 \sin\beta_1}{J_z(w_1 + w_2\cos\beta_1)} \tag{65}$$

For $\beta_1 = \arccos\sqrt{(6+\sqrt{96/5})/14} \approx 30.56°$, we obtain Equation (66):

$$\begin{aligned}
\frac{w_1}{w_2} \equiv \frac{\nu_{\text{inner}}}{\nu_{\text{outer}}} = k &= \overline{\frac{6+\sqrt{96/5}}{14}} \left(\frac{J_y}{J_z} - 1 \right. \\
&\approx 0.86 \left(\frac{J_y}{J_z} - 1 \right.
\end{aligned} \tag{66}$$

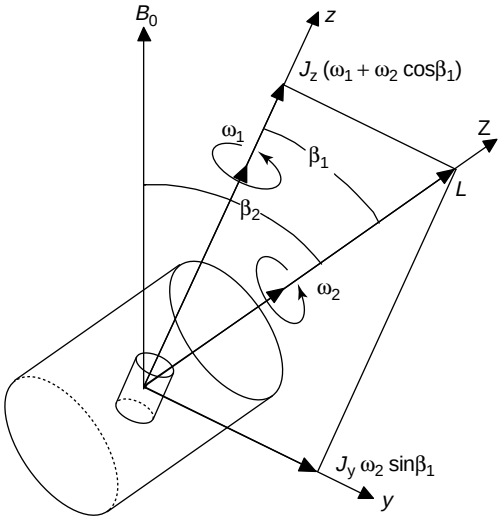

Figure 4 Moments of inertia for DOR after Wu et al.[55] By adjusting the ratio of moments, J_y/J_x, the addition of the two vector components of the angular momentum along the inner-rotor axis z and the y-axis can be made to point along the outer-rotor axis Z, so that spinning the outer rotor will not affect the orientation of the total angular momentum of the inner rotor.

and the usual design of the inner rotor gives values of J_y/J_z so that $5 < k < 6$. Wu et al.[55] explained that the spinning system is in a stable state when the torque-free condition, which is given by Equation (66), is slightly violated and the frequencies obey Equation (67):

$$\nu_{\text{inner}} = k\nu_{\text{outer}} + \Delta\nu_{\text{inner}} \tag{67}$$

where the excess of the inner frequency $\Delta\nu_{\text{inner}}$ is about 1 kHz.

The application of rotation-synchronized pulses has been shown to be effective in eliminating half of the DOR sidebands.[56] The total suppression of sidebands can be achieved by the application of additional pulses[57] and by application of the magic-angle turning technique.[58] A theoretical consideration of the DOR sidebands can be found in the literature.[59–61]

DOR NMR gives accurate values of the isotropic shift, δ_{iso}, of the NMR of a quadrupolar nucleus. Two shift effects are superimposed: the isotropic quadrupolar shift, which is described in Equation (44), and the isotropic value of the chemical shift, which was introduced in Equation (16). We substitute σ_{iso}, the isotropic part of the shielding tensor, by the value of the chemical shift $\delta_{\text{CS iso}} = \sigma_{\text{ref}} - \sigma_{\text{iso}}$ (with respect to a given isotropic chemical shift of a reference compound). Then we substitute the second-order quadrupolar shift $\nu_{\text{iso Q}}$ in Equation (44) by the dimensionless value $\delta_{\text{Q iso}} = (\nu_{\text{iso Q}})/\nu_{\text{L}}$. In this notation we obtain for the observed isotropic shift in the DOR experiment Equation (68):

$$\delta_{\text{DOR}} = \delta_{\text{CS iso}} + \delta_{\text{Q iso}} = \delta_{\text{CS iso}} - \frac{\nu_{\text{Q}}^2}{30\nu_{\text{L}}^2}\ I(I+1) - \frac{3}{4}\ \left(1 + \frac{\eta^2}{3}\right. \tag{68}$$

For common use the shifts can be expressed in parts per million (ppm) by multiplying the right-hand side of Equation (68) by 10^6. Equation (68) reflects the fact that the value of the isotropic chemical shift cannot be obtained from *one* DOR experiment if the quadrupolar shift is unknown. However, two DOR experiments at different external field strengths, which means at different Larmor frequencies, give the value of the chemical shift $\delta_{\text{CS iso}}$ and also the quadrupolar parameter $\nu_{\text{Q}}\sqrt{1+\eta^2/3}$.

The DOR technique has been applied, first, to the ^{23}Na resonance in sodium oxalate[54,56,62] and later to $NaNO_2$[56,63] to the ^{17}O resonance in diopside[64] and silicates[65] and to the ^{27}Al resonance of YAG.[63] The majority of the DOR NMR studies was devoted to zeolites of faujasite type[66–75] and to zeolite-like materials such as aluminophosphates,[47,66,76–86] zincosilicates[87] or sodalite.[88,89] Other materials studied included aluminas[90] and aluminum borate[91] (^{27}Al), gallium phosphate[92] (^{71}Ga) and boron nitride[93] (^{11}B). CP effects have been studied.[47,79] Amoureux[94] proposed a combination of DOR and variable-angle spinning (VAS) in which the quadrupolar shift of DOR NMR can be deduced with a simulation of the second-dimension spectra obtained with stepped values of the angle of the outer rotor with respect to the external magnetic field.

The rotation frequencies of the inner and outer rotor at present do not exceed $\nu_{\text{inner}} = 12$ kHz and $\nu_{\text{outer}} = 2$ kHz, respectively. A pneumatic unit, which is controlled by a computer, simplifies the experimental set-up and makes it safer. However, compared with the MAS technique, a more complicated set-up and stronger wear of the rotors must still be accepted for DOR experiments.

A DAS experiment was first performed independently by Chmelka et al.[64] and Llor and Virlet.[95] The DAS technique uses a time-dependent angle of the rotor axis, which toggles in the simple experiment between 37.38° and 79.19° for two equal periods of time. During the hopping time, necessary for switching the angle (>30 ms), the evolution of the magnetization can be reduced by means of a selective $\pi/2$ pulse pair applied at a time t_1 after the preparation pulse. The signal consists of a second-order quadrupolar echo with a maximum at the time t_1 after the last pulse. The data acquisition of the FID starting at the echo maximum gives the data in the t_2 domain. The two-dimensional (2-D) spectrum is obtained after a 2-D Fourier transformation with respect to t_1 and t_2. An advanced concept introduces a fourth $\pi/2$ pulse in order to retain two pathways from the t_1 evolution yielding a

pure absorption-phase spectrum and introduces a third angle of rotation of 54.74°, in order to correlate in two dimensions the DAS spectra with MAS spectra.[96,97] A sideband pattern emerges if the evolution period of the 2-D DAS experiment is divided into unequal parts by RF pulses or spinner reorientation, as shown by Grandinetti et al.[98] Fyfe et al.[99] found that the switch to the magic angle after the CP step is useful for obtaining high-resolution spectra of slowly relaxing spin species. A 2-D heteronuclear correlation DAS NMR experiment with CP from ^{23}Na to ^{31}P was applied by Jarvie et al.[100] in order to find connectivities between the nuclei. Wenslow and Mueller[101] demonstrated the increased information of this so-called DAS/CP/MAS NMR experiment for the study of cation sites in glasses. Medek et al.[102] introduced the three-dimensional dynamic-angle correlation spectroscopy (DACSY), which is based on the acquisition of DAS NMR signals as a function of different initial and final spinning axes.

Two drastic limitations for the application of the DAS technique compared with the DOR technique should be noted. First, the spin–lattice relaxation time T_1, which is often short for quadrupolar nuclei, has to be larger than the time period necessary for flipping the rotor axis. Second, the spin exchange due to dipolar interactions, which cannot be eliminated during the relatively long flipping time of the rotor axis, has to be sufficiently small.

6 MULTIPLE-QUANTUM MAGIC-ANGLE SPINNING TECHNIQUE

In 1995, Frydman and Harwood[25] proved the feasibility of a 2-D NMR experiment that makes use of invisible multiple-quantum transitions combined with MAS to remove the anisotropy of the quadrupolar interaction. Symmetric $p/2 \leftrightarrow -p/2$ coherences with the quantum level pQ were selected, since the corresponding powder resonances are devoid of first-order quadrupolar effects.

The phase development $\varphi(t)$ of the single or multiple-quantum coherence can be written as Equation (69):

$$\begin{aligned}\frac{\varphi(t)}{2\pi t} &= \Delta\nu p + \nu_{p/2,-p/2} \\ &= \Delta\nu p + \frac{p\nu_Q^2(3+\eta^2)[4I(I+1)-3p^2]}{360\nu_L} \\ &\quad - \frac{p\nu_Q^2\{36I(I+1)-17p^2-10\}}{12960\nu_L}\{(18+\eta^2)d_{0,0}^{(4)} \\ &\quad + \sqrt{360}\,\eta d_{2,0}^{(4)}\cos 2\alpha + \sqrt{70}\,\eta^2 d_{4,0}^{(4)}\cos 4\alpha\} \qquad (69)\end{aligned}$$

with $\nu_{p/2,-p/2}$ from Equation (61), which is reduced under MAS conditions. The contributions from the chemical shift and from the resonance offset are included in $\Delta\nu = \sigma_{iso}\nu_L - \nu_{offset}$. Equation (69) shows that by going from the multiple-quantum level pQ to the -1Q level of observation, the sign of the phase development can be inverted. Thus, the influence of the anisotropy of the second-order quadrupolar interaction is averaged out, if the times t_1 and t_2 spent on the quantum levels pQ and -1Q, respectively, fulfill the condition in Equation (70):

$$t_2 = p\frac{36I(I+1)-17p^2-10}{36I(I+1)-27}t_1 = R(I,p)t_1 \qquad (70)$$

This relation describes the appearance of the isotropic echo, and thus gives the slope $R(I,p)$ of the anisotropic axis in the 2-D spectrum, which is obtained after the 2-D Fourier transformation with respect to t_2 and t_1.

Several pulse sequences were used in order to excite a multiple-quantum coherence and convert it to the observable single-quantum coherence after the evolution time. The pulse sequence, which is shown in Figure 5, is mostly applied now. The highest possible RF power should be used for the first and second pulse, which excites the triple- (or quintuple-) quantum coherence and converts it back to zero-quantum coherence, respectively. The third pulse (z-filter) should be much weaker, in order to excite the central transition only. It converts nonobservable zero-quantum coherence and populations ($p = 0$) into observable single-quantum coherences ($p = -1$).

The symmetric 3Q coherence pathway in Figure 5 is $(0 \rightarrow \pm 3 \rightarrow 0 \rightarrow -1)$. The symmetric quintuple-quantum

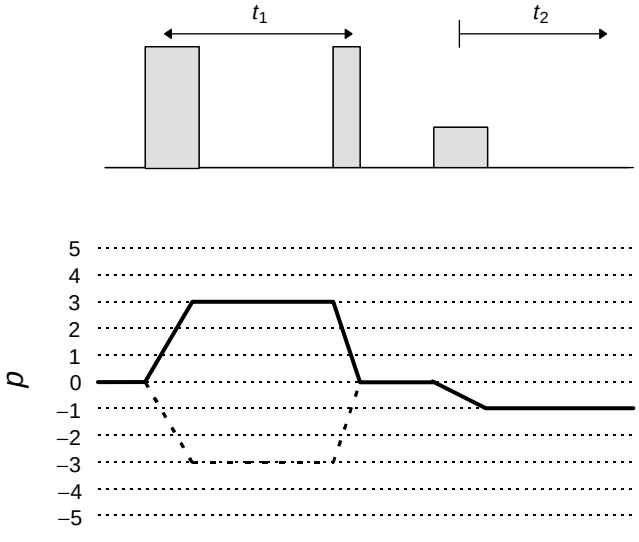

Figure 5 Pulse sequence and coherence transfer pathway for the echo (upper way) and antiecho (lower way) consisting of two strong pulses and a weak z-filter pulse. The triple-quantum MAS pathway is shown on the lines for the nuclear spin $I = \frac{5}{2}$.

(5Q) pathway is $(0 \rightarrow \pm 5 \rightarrow 0 \rightarrow -1)$. Two-pulse sequences (without z-filter) can be described by the notation $(0 \rightarrow \pm p \rightarrow -1)$. Although only one sign of p (cf. Equation 70) yields an echo which can be observed, it is needed to acquire both $\pm p$ coherence transfer pathways when trying to produce 2-D MQMAS spectra that are free from dispersive distortions.[103] Equally strong signals from both pathways must be collected, in order to obtain pure adsorptive spectra. This is more difficult if a two-pulse sequence is used, which includes the different transitions $+p \rightarrow -1$ and $-p \rightarrow -1$ for the last step in the echo and antiecho coherence transfer pathway, respectively. However, both pathways are completely symmetrical when considering the coherence jumps in the z-filter pulse sequence.[104,105]

The flip angles of the first and second pulses should be optimized with respect to the best signal amplitude. In principle, this can be done by numerical calculations,[106] e.g. by the program PULSAR.[107] The result depends on the nuclear spin I and its Larmor frequency ν_L, the chosen coherence pQ and the ratio between the nutation frequency $\nu_{rf} = \gamma B_{rf}/2\pi$ and the quadrupolar frequency ν_Q. One conclusion is that the MQMAS NMR spectra, which are obtained by strong pulses, do not present quantitative results with respect to the concentration of species, since the multiple-quantum transition efficiency is different for nuclei with different quadrupolar frequency ν_Q.

The calculations need a full set of NMR parameters including the quadrupolar parameters and do not take into consideration second-order and spinning effects. Therefore, the optimization of the pulse duration should be performed at the beginning of the NMR experiment by variation of the pulse duration; and only the initial flip angles for this procedure should be taken from the numerical calculations. Nonselective flip angles of about 180° and 50° for the first and second pulse, respectively, can be used for any I and pQ, if the Larmor frequency is about 100 MHz and the ratio between the quadrupolar frequency ν_Q and the nonselective nutation frequency is $\nu_Q/\nu_{rf} \approx 10$. More precise flip angles are given in the literature[106,108] as a function of I and pQ. (Note that definition of the quadrupolar frequency in the literature[25,106,108] is different to the definition of Abragam;[19] cf. Equation (25), which is used here: $\nu_Q^{Frydman} = \nu_Q^{Abragam}/6$.)

An alternative approach to the generation of multiple-quantum transitions by strong pulses is to use the adiabatic transfer by the interchange between the eigenstates 3Q and 1Q undergoing MAS.[42] Wu et al.[109] introduced a triple-quantum magic-angle spinning (3QMAS) experiment by rotation-induced adiabatic coherence transfer (RIACT) for $I = \frac{3}{2}$ nuclei and applied it to the study of sodium sites in inorganic salts. They claimed that this method gives quantitative MQMAS NMR spectra, since the adiabatic coherence transfer is independent of the strength of the quadrupolar interaction. However, theoretical calculations[108] showed that, compared with the nonadiabatic excitation, the disturbing off-resonance effects are more extensive for RIACT. Lim and Grey[110] performed numerical simulations of the RIACT experiment as a function of asymmetry parameter and pulse spacings and discussed the effect of very fast sample spinning. Mildner et al.[111] demonstrated by means of ^{17}O studies of silicates that RIACT can be applied to $I = \frac{5}{2}$ spins at the moderate RF field strength of $\nu_{rf} \approx 30$ kHz.

Ding and McDowell[112,113] claimed that a shaped first excitation pulse would also achieve quantitative results in nonadiabatic 3QMAS NMR experiments and that properly shaped pulses can also reduce the requirements on the RF power and sample spinning speeds. The superior excitation of triple-quantum coherences by composite pulse schemes was also shown by Marinelli et al.[114]

Kentgens and Verhagen[32] showed that DFSs are very efficient for the conversion of triple- to single-quantum coherence in MQMAS spectroscopy. Goldbourt et al.[115] used instead fast radiofrequency amplitude modulation (FAM) in the version FAM and FAM1 for $I = \frac{3}{2}$ and $I = \frac{5}{2}$ nuclei, respectively, and obtained a substantial intensity and resolution enhancement of the signal over the commonly used MQMAS pulse schemes. A very strong RF field, no resonance offset and a high spinning rate, if possible 30 kHz as shown by Rocha,[116] are still the requirements for most MQMAS experiments.

For both the nonadiabatic and adiabatic MQMAS experiments described above, the selection of desired multiple-quantum coherences is done by phase cycling.[111,117] Methods of hypercomplex or time-proportional phase increment (TPPI) 2-D Fourier transformation [cf. Ernst et al.[118]] are used for the acquisition of real and imaginary part with respect to the t_1/ν_1 dimension. Fyfe et al.[119] used pulsed field gradients instead of phase cycling. The advantage is the simplification of the MQMAS pulse program, which allows the implementation of correlation experiments. However, the insertion of the field gradient coil reduces the maximum achievable RF power.

A shearing transform of the experimentally obtained 2-D spectrum, which aligns anisotropic ridges parallel to the ν_2 axis in such a way that the isotropic dimension can be displayed along the ν_1 axis, is described below. Brown et al.[104,120,121] introduced an experiment with an additional pulse called split-t_1 MQMAS that is pure adsorptive and includes the shearing ratio into the pulse program so that anisotropic ridges

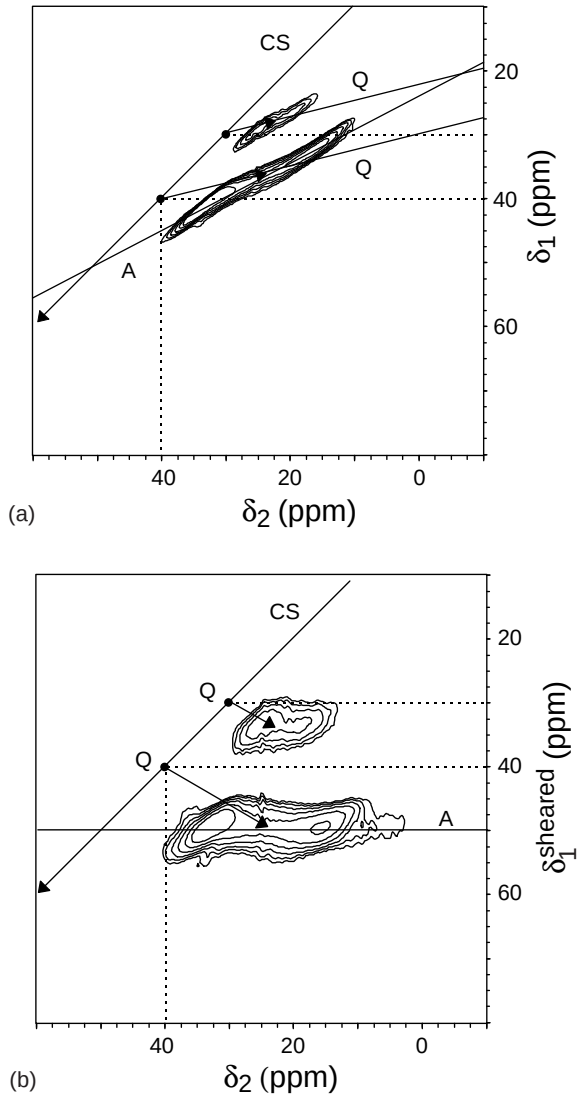

Figure 6 2-D ^{17}O 3QMAS NMR spectra of the zeolite Na-ZSM-5 (a) after the 2-D Fourier transformation and (b) after the following shearing transformation. The two signals were assigned to SiOSi and SiOAl fragments with concentrations of 80 and 20%, $\delta = 40$ and 30 ppm, $C_{qcc} = 5.3$ and 3.5 MHz, $\eta = 0.12$ and 0.29, respectively.[53] (The figure was provided by Ulf Pingel.)

are parallel to the ν_2 axis after the 2-D Fourier transformation.

Figure 6(a) and (b) shows the 2-D ^{17}O 3QMAS NMR spectra of the zeolite Na-ZSM-5 before and after the shearing procedure. The two signals in the spectra are due to Si−O−Si and Si−O−Al bonds.[53] The spectra demonstrate how the projection on the ν_1 axis is changed by the shearing. The analytical treatment starts from the isotropic part of Equation (69). Without resonance offset, the total shift is a superposition of the isotropic part of the chemical shift, $\sigma_{iso}\nu_L$, and the isotropic part of the quadrupolar shift, $\nu_{iso\ Q}$ (Equation 71):

$$\begin{aligned}\nu_{iso} &= \sigma_{iso}\nu_L p + \nu_{iso\ Q} \\ &= \sigma_{iso}\nu_L p + \frac{p\nu_Q^2(3+\eta^2)[4I(I+1)-3p^2]}{360\nu_L} \end{aligned} \quad (71)$$

Equation (71) can be transformed into relative units (ppm) by multiplication by $10^6/(-p\nu_L)$. The value $-p\nu_L$ describes the apparent Larmor frequency and includes the real Larmor frequency for $p = -1$. For the ν_1 axis (before shearing) and for the ν_2 axis, we have Equations (72) and (73), respectively:

$$\begin{aligned}\delta_1 &= \delta_{CS\ iso} + \delta_{Q\ iso} \\ &= \delta_{CS\ iso} - \frac{\nu_Q^2(3+\eta^2)[4I(I+1)-3p^2]}{360\nu_L^2} \end{aligned} \quad (72)$$

$$\begin{aligned}\delta_2 &= \delta_{CS\ iso} + \delta_{Q\ iso} \\ &= \delta_{CS\ iso} - \frac{\nu_Q^2(3+\eta^2)[4I(I+1)-3]}{360\nu_L^2} \end{aligned} \quad (73)$$

Equation (73) describes the shift δ_2 of the center of gravity of the anisotropic ridge, which is not influenced by the shearing transformation. The coordinate $\delta_1^{shearing}$ after the transformation is given by[108] Equation (74):

$$\begin{aligned}\delta_1^{shearing} &= \delta_{CS\ iso} + \delta_{Q\ iso}^{shearing} \\ &= \delta_{CS\ iso} + \frac{\nu_Q^2(3+\eta^2)[4I(I+1)-3]}{612\nu_L^2} \end{aligned} \quad (74)$$

It can be seen that after shearing the quadrupolar shift becomes always positive and independent of p. Comparison of Equation (74) with Equation (73) gives the slope of two special lines in the sheared 2-D spectrum. One line with slope 1, which is denoted CS in Figure 6, describes the variable isotropic chemical shift in the case of zero quadrupolar shift. Only one line CS exists and crosses the coordinates (0, 0). Lines Q with slope −10/17 describe variable quadrupolar shifts at one value $\delta_{CS\ iso}$, which is given by the crossing of line Q with line CS. The shift difference between this crossing point and the center of gravity of a signal gives the corresponding quadrupolar shift $\delta_{Q\ iso}^{shearing}$. One-dimensional anisotropic slices A are obtained from the 2-D spectrum. These slices can be fit by a line-shape simulation, which gives values for ν_Q and η, whereas $\delta_{Q\ iso}$ is a measure of $\nu_Q\sqrt{1+\eta/3}$. The quadrupolar parameter obtained from the 2-D spectrum can then be used for a simulation of the single-quantum MAS NMR spectrum. This procedure yields correct information about the relative concentration of the various species giving rise to the corresponding signals. The program REGULAR, which was developed by Delevoye,[122] takes full advantage of the information included in the 2-D MQMAS NMR spectrum and thereby

quantifies the distribution of nucleus surroundings also in badly crystallized and amorphous materials. A 2-D result is obtained which represents quadrupolar constants versus isotropic chemical shifts with correct relative intensities. Scaling and labeling of the isotropic axis (ν_1 axis after shearing) was considered in detail by Man.[123]

The residual line width of the MQMAS NMR spectra is a matter of current research. Wu and Wasylishen[124] showed that the heteronuclear dipolar interaction between two nuclei does not cause residual dipolar couplings in MQMAS NMR experiments.

The spinning sidebands in the multiple-quantum dimension are a common feature of dipolar-coupled spin $I = \frac{1}{2}$ and half-integer quadrupolar systems. Both cases were considered by Friedrich et al.[125] The case of half-integer spins is the subject of various studies.[126–129] Spinning sidebands arise even in the absence of explicit time modulations of any spin interactions, due to orientation and time dependence characterizing the multiple- to single-quantum conversion process, as shown by Marinelli and Frydman.[126] The sidebands cause signal loss and misinterpretation of spectra. Rotor synchronization, which means setting the evolution dwell time equal to the rotation period, is useful for suppressing these spinning sidebands.[130] However, it limits the spectral width of the multiple-quantum dimension to the rotation frequency. This drawback increases on going from triple- to quintuple-quantum transitions.

Several combinations of MQMAS with other NMR techniques have been successfully applied. Vosegaard et al.[131] added the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence to the MQMAS pulse program, in order to improve the sensitivity by splitting the line shape in the ν_2 dimension into manifolds of spin-echo sidebands. High-power ^{1}H decoupling improves the resolution in both dimensions.[132,133] Lacassagne et al.[134] demonstrated improved resolution for strong heteronuclear coupling by applying composite decoupling schemes during the evolution and acquisition time. Wang et al.[135] provided chemical shift anisotropy data by multiple-quantum variable-angle spinning (MQVAS). CP provides information about connectivities.[136–139] REDOR[139–141] and heteronuclear chemical shift correlation (HETCOR)[138,142] were also applied in addition to MQMAS. The most difficult combination, DOR in addition to MQMAS, was presented by Samoson et al.[143,144] The advantage of multiple-quantum double-rotation (MQDOR) compared with MQMAS is that the need for simulations is avoided.[143] However, although in principle most informative, experimental difficulties and low multi-quantum conversion efficiency make the applicability of the MQDOR technique currently relatively limited.[144]

MQMAS NMR studies have been carried out on all half-integer quadrupolar spins from $I = \frac{3}{2}$ to $I = \frac{9}{2}$ [^{93}Nb[106]]. The nucleus ^{59}Co ($I = \frac{7}{2}$) in nitrogen compounds was studied to show the interplay between the chemical shift anisotropy and the quadrupolar interaction. However, the majority of the studies were focused on the two nuclei ^{27}Al ($I = \frac{5}{2}$) and ^{23}Na ($I = \frac{3}{2}$). Two examples are the study of the aluminum coordination in $SrAl_{12}O_{19}$ by Jansen et al.[145] and the study of sodium cations in dehydrated zeolites by Hunger et al.[146] Zeolites, zeolite-like materials and glasses[147,148] have been investigated in many studies.

7 NUTATION TECHNIQUE

Nutation NMR spectroscopy represents a 2-D experiment: a strong RF field is irradiated during the evolution period t_1, and the FID is observed during the detection period t_2. In the rotating frame, the spins *nutate* (precess) around the strong RF field with specific nutation frequencies w_1; here the subscript 1 denotes the frequency axis w_1 in the 2-D spectrum corresponding to the Fourier transform with respect to t_1 and should not be confused with $w_{rf} = \gamma B_{rf}$ which is a constant for the 2-D experiment. The original one-dimensional nutation experiment of Torrey[149] uses only one point of the FID.

Samoson and Lippmaa[150,151] introduced 2-D nutation NMR studies of half-integer quadrupolar nuclei. A corresponding one-dimensional experiment was performed by Fenzke et al.[152] For quadrupolar nuclei, the effective nutation frequencies w_1 depend on the strength of the quadrupolar interaction, as has been shown in section 3. If $w_Q \ll w_{rf}$, then the transverse magnetization responds to the RF pulse like spin-$\frac{1}{2}$ nuclei, thus $w_1 = w_{rf}$. If $w_Q \gg 10w_{rf}$, the central transition can be treated as a two-level system and an effective nutation frequency is expected, that is, increased by a factor $I + \frac{1}{2}$, i.e. $w_1 = (I + \frac{1}{2})w_{rf}$ (cf. Equation 39). For the intermediate case, $0.1 < w_Q/w_{rf} < 100$, the nutation spectra are complicated because of the influence of the partly excited outer satellite transitions.

Nutation spectra can be calculated using numerical procedures or by means of an analytical description which is the more complicated the larger the spin quantum number. The powder average, which produces complicated nutation spectra, is necessary, in order to deduce quadrupolar parameters from the spectra and to compare the experimental spectrum with a set of calculated spectra. An analytical approach was found for $I = \frac{3}{2}$,[153–156] for $I = \frac{5}{2}$,[157,158] and for $I = \frac{7}{2}$[159,160] yielding $(I + \frac{1}{2})^2$ different nutation frequencies of the central transition. Each frequency gives rise to a powder

pattern, and $2I$ of the $(I + \frac{1}{2})^2$ frequencies are dominant. Calculated line shapes of nutation spectra can be found in papers by Kentgens et al.[161,162] and in a review by Freude and Haase.[9] The latter reviews the quadrupolar nutation literature from 1983 to 1992 (about 40 papers), including experimental aspects such as inhomogeneity of the RF field, resonance offset, influence of MAS, spin–lattice relaxation, spin–spin relaxation in the rotating frame and some advanced concepts in nutation spectroscopy. Since 1993, another 15 papers were published, among them recent work about nutation MAS NMR by Ding and McDowell.[46,163]

The most advanced concept, the off-resonance nutation NMR of half-integer nuclei, was introduced by Kentgens.[164] The large resonance offset, $\nu_{\text{offset}} \gg \nu_{\text{rf}}$, during the irradiation causes an evolution of the spin system in the effective field, $\boldsymbol{B}_{\text{eff}} = \boldsymbol{B}_{\text{offset}} + \boldsymbol{B}_{\text{rf}}$, whereas the signal is as usually observed in resonance. The off-resonance technique extends the range of applicability at given $\boldsymbol{B}_0$ and $\boldsymbol{B}_{\text{rf}}$ to larger values of C_{qcc}. The acquisition of several spectra with different ν_{offset} improves the reliability of the fitting process, which is essential for obtaining information from nutation spectra. A loss of signal intensity due to the uninformative zero-frequency peak, which is typical for off-resonance irradiation, can be avoided by a soft selective pulse or by using an FSAHP (cf. section 3), for the preparation of the central transition magnetization only. A recent review including the perspectives of the nutation NMR of half-integer quadrupolar nuclei was given by Kentgens.[162] He proposed a combination of MQMAS, off-resonance nutation and echo techniques for sites with very large C_{qcc}.

8 ECHO TECHNIQUES

Echo techniques provide two data sets in the NMR time domain. First, the echo decay dependence on the time t_2 starting from zero at the maximum of the echo gives similar information to the initial FID after the preparation pulse, but the loss of signal during the ring down of the probe and the recovery of the receiver immediately after the pulse is overcome in the echo technique. Second, by varying the pulse distance t_1 between the preparation pulse and the refocusing second pulse, the resulting envelope of the echo decay gives additional information about the spin system. The refocusing effect of a second pulse after a preceding $\pi/2$ pulse for an interaction that is proportional to I_z is called the Hahn echo. The original experiment by Hahn[165] was performed with identical phase incoherent pulses in an inhomogeneous external magnetic field. The term Hahn echo is now in use for the spin-echo (after $\pi/2{-}t_1{-}\pi{-}t_1$) of spin-$\frac{1}{2}$ nuclei or of quadrupolar nuclei with half-integer spins, if the central transition is selectively excited.

However, the quadrupolar interaction strongly influences the formation of an echo: homonuclear dipolar interactions which mainly cause the decay of the spin-echo amplitude become less effective since spin-flipping between different transitions is forbidden. Owing to the nature of the quadrupolar coupling, which is to first order proportional to I_z^2, refocusing is not complete. Also, the limited range of excitation for very broad lines of powder samples causes complicated spin-echo behavior.

Echo studies on quadrupolar nuclei with half-integer spin (for integer spin, see section 9) were reviewed by Freude and Haase,[9] beginning with the pioneering work of Das and Saha[6] in 1955 up to the publications in 1993, e.g. Haase and Oldfield.[166] Several cases, which can be mainly divided into nonselective and selective excitation, were considered, as follows.[9]

Nonselective and hard pulses without dipole interaction cause several echoes at times $k \times t_1$ with $k = \frac{1}{2}, 1, \frac{3}{2}$ and $k = \frac{1}{2}, 1, \frac{3}{2}, 2, 3$ for $I = \frac{3}{2}$ and $I = \frac{5}{2}$, respectively. Some of these are forbidden echoes: $k = \frac{1}{2}, \frac{3}{2}$ and $k = \frac{3}{2}, 3$ for $I = \frac{3}{2}$ and $I = \frac{5}{2}$, respectively. On going from hard to soft pulses (cf. section 3), the influence of the quadrupolar interaction during the pulse has to be considered and the selection rule is changed. Then also the forbidden echoes can be observed. The consideration of dipole interaction for first-order quadrupolar echoes can be divided into two parts. First, the heteronuclear dipole interaction, which is proportional to $I_z S_z$, can generally be removed by the application of a second π-pulse. However, strong coupling among the S spins, and also a short spin–lattice relaxation time of the S spins, can destroy the formation of the I spin echo. Second, if the homonuclear dipolar interaction is large compared with the quadrupolar interaction, $\| \omega_{\text{D}} \| \gg \| \omega_{\text{Q}} \|$, its influence is essentially the same as for spin-$\frac{1}{2}$ nuclei. If the quadrupolar interaction is large compared with the dipolar interaction, $\| \omega_{\text{Q}} \| \gg \| \omega_{\text{D}} \|$, spin exchange due to homonuclear dipole interaction between adjacent levels can be considered as being suppressed completely and the echo decay can be investigated by selective excitation of each transition, separately.

$\| \omega_{\text{Q}} \| \gg \| \omega_{\text{rf}} \| \gg \| \omega_{\text{D}} \|$ and $\| \omega_{\text{Q}} \| \gg \| \omega_{\text{rf}} \| \approx \| \omega_{\text{D}} \|$ denote the cases of selective excitation of a single transition without and with dipole interaction, respectively. For powder we can mainly excite the central transition, and for single crystals we can tune to any transition and describe it by the reduced spin-$\frac{1}{2}$ formalism.[19] Without the dipole interaction an echo is observed at the time t_1 after the π pulse in full analogy with spin-$\frac{1}{2}$ nuclei subjected to an inhomogeneous interaction. Since the calculation of the spin-echo amplitude in the presence of dipolar interactions can be complicated,

the spin-echo decay has been discussed by Haase and Oldfield[166] in terms of the second moment spin-echo decay (cf. section 5.2.5 in Freude and Haase[9]).

Echo techniques in combination with coherence selection (multiple-quantum filter) reduce the overlap of signals of quadrupolar nuclei in anisotropic soft matter, as demonstrated by Furo and Halle.[167,168] Recent progress in this field was presented by Eliav et al.[169–172] The use of echoes for the editing of ^{27}Al MAS NMR spectra of zeolite catalysts was shown by Schmitt et al.[173] The connection between all NMR interactions (except the homonuclear dipole interaction) and the optimum experimental conditions for the avoidance of spectral distortions was described by Dumazy et al.[174]

The leading authority in the field of quadrupolar echoes is Pascal Man, who has published more than 10 papers on this topic during the last decade. Most of them are discussed in recent publications.[175–177]

The CPMG[178,179] pulse sequence $\pi/2_x$, $(t_1, \pi_y, t_1)_n$ can be considered as an advanced concept of the Hahn echo. Larsen et al. reconsidered the quadrupolar version of CPMG (QCPMG) without MAS[180] and with MAS.[181] For MQMAS NMR and ^{2}H NMR, see sections 6 and 9, respectively. More than one order of magnitude can be gained in sensitivity by the application of this QCPMG technique.

9 ^{2}H NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

^{2}H is the most important NMR isotope with an integer nuclear spin $I = 1$, rather seldom ^{6}Li and ^{14}N NMR studies can be found in the literature. About 10% of the solid-state NMR studies of quadrupolar nuclei concern deuteron magnetic resonance spectroscopy, which is a powerful technique for investigating molecular structure and dynamics. This small section does not correspond to the importance of the deuterium NMR, especially for the study of dynamics. The textbook by Schmidt-Rohr and Spiess, *Multidimensional Solid-state NMR and Polymers*[18] and the reviews by Hoatson and Vold, '^{2}H-NMR Spectroscopy of Solids and Liquid Crystals',[11] and Ulrich and Grage, '^{2}H NMR (Solid-state NMR of Polymers)',[14] demonstrate the wealth of modern applications of solid-state ^{2}H NMR.

The experimental techniques for nuclei with the spin $I = 1$ are different to the techniques for nuclei having half-integer spins mainly for one reason, that no central transition exists. There are only two transitions for the magnetic quantum number m: $-1 \leftrightarrow 0$ and $0 \leftrightarrow 1$. They cause a line splitting that is symmetric with respect to the Larmor frequency if the quadrupolar interaction is considered in first-order perturbation theory with respect to the Zeeman interaction (cf. Equation 40). From Equation (41) it follows that the second-order shifts for both transitions are identical. Thus, the splitting between the two lines (difference frequency), which is the experimentally observed parameter in studies of single crystals, is not influenced by the second-order quadrupolar interaction.

Equations (40) and (41) can be used for nuclei with the spin $I = 1$, but the quadrupolar frequency should be substituted by the quadrupolar coupling constant using $\nu_Q = 3C_{qcc}/2$. The definition of the quadrupolar frequency for half-integer spins in Equations (25) and (26) is rather unusual for integer spins. The use of the quadrupolar coupling constant C_{qcc} after Equation (24) is commonly accepted, but for integer spins no notational consistency exists with respect to the quadrupolar frequency ν_Q.

The asymmetry parameter η (cf. Equations 19 and 20), is relatively small for the majority of the ^{2}H NMR studies due to a rotation symmetry of the hydrogen bonding. The theoretical powder line shape of nuclei with $I = 1$ and $\eta = 0$ is equal to the line shape of isolated pairs of nuclei with the spin $I = \frac{1}{2}$. Therefore, the typical ^{2}H NMR pattern of the so-called static powder (no sample rotation and no internal rotation) is often referred to as the Pake powder pattern. Pake in 1948 studied the ^{1}H NMR spectra of gypsum and discussed the angular dependence of the doublets in the spectra of a single crystal and the powder pattern in the spectrum of a powder sample as well.[182]

The splitting of lines of single crystals or the full width of powder patterns reach values up to 300 kHz. From $\delta\nu_{1/2} \approx 0.886/\tau$ (cf. Equations 32–34), it follows that a duration $\tau \leq 2\,\mu$s for the $\pi/2$ pulse is necessary, in order to obtain a quite uniform excitation of the whole spectrum. In addition to the need for high RF power, an echo pulse sequence must be applied in order to overcome the dead time problem of the receiver. A simple pulse sequence consisting of two $\pi/2$ pulses with a pulse duration of 2 μs and a pulse separation of 20 μs with phase cycling (the phase of the first pulse is cycled with respect to the synthesizer, and in addition the phase of the second pulse is cycled with respect to the first pulse) is the standard for solid-state ^{2}H NMR experiments. The pulse separation must be adjusted to one rotation period if MAS is applied. Hoatson discussed the broadband composite excitation sequences[183] and the Hadamard pulse excitation[184,185] for detecting quadrupolar order in ^{2}H NMR. Noise spectroscopy provides broad-bandwidth excitation with minimal applied RF field strengths. The state-of-the-art for applications to the ^{2}H solid-state NMR was described by Yang et al.[186,187] More detailed consideration of the quadrupolar echo in ^{2}H NMR can be found in the review by Hoatson and Vold,[11] Chapter 2.2, and the references

therein. (The NMR echo for nuclei with half-integer spin is discussed in section 8 of this article.) The excitation with symmetric quadrupolar-shaped (SQUASH)[188] and time-asymmetric-shaped (QUASH) pulses[189] improves the excitation bandwidth.

Deuterium labels are introduced synthetically or by a postsynthesis treatment of the material. The labeled hydrogen segment can be investigated in terms of its specific structural and dynamic properties. The natural abundance of ^{2}H is only 150 ppm. The relative sensitivity of naturally abundant ^{2}H with respect to the ^{1}H NMR sensitivity at constant external field is only 1.5 ppm. Nevertheless, natural abundance high-resolution solid-state ^{2}H NMR investigations can be performed if MAS, high-power ^{1}H decoupling and ^{1}H–^{2}H CP are used.[190,191]

^{1}H–^{2}H CP effects in single- and double-quantum (DQ) transitions of the ^{2}H NMR were theoretically and experimentally investigated by Vega et al.[192] in 1980. There are two reasons why CP is rarely used in the solid-state ^{2}H NMR: the maximum increase in the ^{2}H NMR intensity due to the CP from ^{1}H spins is given by the gyromagnetic ratio $\gamma_H/\gamma_D \approx 6.5$ for infinitely diluted ^{2}H nuclei. This value decreases if the concentration of ^{2}H nuclei increases with the deuteron enrichment. However, the increasing level of deuteration increases the ^{2}H NMR intensity more effectively. The second point is that the high-power decoupling of the protons close to the deuterons causes some problems. However, ^{1}H–^{2}H CP is invaluable for the investigation of the interfaces of the deuterated segments, e.g. in polymer blends. Ba et al.[193] proposed a quadrupolar echo double resonance (QEDOR) and a solid echo double resonance (SOLEDOR) NMR experiment in order to provide information about deuteron–proton distances. A new theoretical model that describes CPMAS experiments for abundant $I = \frac{1}{2}$ spins coupled to a single spin $S = 1$ was presented by Marks et al.[194] Gan and Robyr[195] described a new 2-D CPMAS experiment with a three-pulse sequence for the determination of the relative orientation between ^{2}H quadrupolar coupling tensors.

The MAS technique for powder materials offers a more sensitive technique (compared with the quadrupolar echo technique) and the opportunity for an accurate measurement of the chemical shift.[196] A simple variation of the processing procedure of the FID, which is obtained in a one-pulse experiment, gives a 2-D spectrum with quadrupolar pattern in one dimension and the isotropic chemical shift in the other dimension.[197,198] The 2-D data set in the time domain can be constructed from the one-dimensional set, which should be acquired with a dwell time τ_{dwell} and a rotation frequency ν_{rot} in the order of magnitude of 1 μs and 5 kHz, respectively. $N = (2\nu_{rot}\tau_{dwell})^{-1}$ points with the time distance of $2\tau_{dwell}$ starting at the top of each rotational echo define the t_2 domain, and the increments of t_1 as the rotation period $1/\nu_{rot}$ give the t_1 domain. Spaniol et al.[199] recently used a 2-D ^{2}H MAS NMR experiment for the separation of paramagnetic and quadrupolar interactions.

Quadrupolar effects do not shift (or broaden in the case of powder) the signals arising from DQ transitions. Therefore, 2-D ^{2}H DQ NMR spectra provide information about the chemical shift of the deuterons.[200] Advantages of the MAS NMR technique compared with the DQ NMR technique are the higher sensitivity of the single-quantum compared with the DQ transition and the higher sensitivity of the MAS technique compared with the static echo technique. Multiple-quantum effects in combination with MAS are less often used for ^{2}H nuclei[201–203] compared with nuclei with half-integer spin (cf. section 6). Chandrakumar et al.[203] use a four-pulse experiment (QUADSHIFT) with a t_1 interval in the middle (which is incremented by the rotation period) in order to obtain chemical shift values and quadrupolar parameters from DQ coherences under MAS conditions [see also Chandrakumar[13]]. ^{2}H MAS NMR imaging experiments were described by Blümich.[185]

Deuterium NMR spectroscopy of single-crystal systems is an active field of research. The original approach of Volkoff et al.[4,5] is still in use, in order to obtain the quadrupolar parameters from the angular dependence of the line splitting, and new aspects were found and applied by Haeberlen et al.[204,205] The work up until 1994 is reviewed by Chandrakumar[13] and recent results are also available[204,205] The ordering behavior at the phase transition in single crystals of betaine phosphate and betaine phosphite has also been studied by one-dimensional and 2-D ^{2}H NMR.[206]

The majority of ^{2}H NMR studies concern molecular dynamics. Many polymer systems are studied, because ^{2}H NMR is ideally suited to explore molecular motions in polymers.[14] There is a wide range of timescales for the correlation times that are accessible by various methods: 10^2–10^{-4} s (2-D exchange spectroscopy), 10^2–10^{-5} s (selective inversion), 10^{-4}–10^{-8} s (line-shape analysis), 10^2–10^{-4} s (T_2 anisotropy), 10^{-8}–10^{-12} s (T_{1Z} anisotropy) and 10–10^{-12} s (T_{1Q} anisotropy).[11] For exchange spectroscopy we refer to the textbook by Schmidt-Rohr and Spiess.[18] The introduction of multi-dimensional exchange experiments by Spiess et al. gave rise to strong developments in this field. Kubo et al.[207] recently combined deuterium selective-excitation exchange spectroscopy with the Hadamard method for the measurement of reorientation rates. The technique of selective inversion is based on the selective excitation or the so-called hole burning that has been well known in NMR from the very beginning.[208] Recent applications to ^{2}H NMR have been reviewed by Hoatson and Vold.[11]

Table 5 Quadrupolar coupling constant $C_{qcc} = e^2qQ/h$, the asymmetry parameter η and the isotropic value of the chemical shift δ (referred to 1.0 M $AlCl_3 \cdot 6H_2O$) for the ^{27}Al NMR of powder compounds at ambient temperature[a]

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Refs.
Aluminates					
$CaO \cdot 6Al_2O_3$	AlO_4	2.0	$0_{assumed}$	65	261
	AlO_5	6.7	$0_{assumed}$	27.5	262
	AlO_6-I	1.5	$0_{assumed}$	9	261
	AlO_6-II	<1	$0_{assumed}$	16	261
$CaO \cdot 2Al_2O_3$	AlO_4-I	6.7	0.8	78	261
	AlO_4-II	13	0.1	~60	261
$CaO \cdot Al_2O_3$	AlO_4-I	2.7	0.85	80	261
	AlO_4-II	2.7	0.85	83	261
$4CaO \cdot 3Al_2O_3$	AlO_4	2.4	0.95	80	261
$12CaO \cdot 7Al_2O_3$	AlO_4-I	3.7	0.9	79	261
	AlO_4-II	11	0.2	85	263
$3CaO \cdot Al_2O_3$	AlO_4	9.7	0.3	85	263
	AlO_4-I	8.69	0.32	79.5	264
	AlO_4-II	9.3	0.54	78.25	264
$4CaO \cdot 3Al_2O_3 \cdot 3H_2O$	AlO_4-I	1.8	0.5	78	261
	AlO_4-II	5.4	0.45	79	261
$KAlO_2$	AlO_4	1.1	0.7	76	263
$KAlO_2 \cdot 0.5H_2O$	AlO_4	5.6	0.0	77	263
$KAlO_2 \cdot H_2O$	AlO_4	6.5	0.6	83	263
$KAlO_2 \cdot 1.5H_2O$	AlO_4	5.0	0.25	81	263
$5BaO \cdot Al_2O_3$	AlO_4	2.3	0.8	80	263
$BaO \cdot Al_2O_3$	AlO_4	2.4	0.4	78	263
$\alpha BaO \cdot Al_2O_3 \cdot 2H_2O$	AlO_4-I	3.4	0.5	81	263
	AlO_4-II	5.1	0.9	80	263
α-$LiAlO_2$	AlO_6	2.8	0.05	16	265
β-$LiAlO_2$	AlO_4	1.8	0.55	82	265
	AlO_4	1.86	0.56	83.0	266
γ-$LiAlO_2$	AlO_4	3.2	0.7	81	265
β-$NaAlO_2$	AlO_4	1.4	0.5	80	263
α-Al_2O_3	AlO_6	2.40	0.01	18.8	267
γ-Al_2O_3	AlO_4	>5.0	~0	>67	268, 269
	AlO_6	4.5	~0	>9	268, 269
γ-Al_2O_3 (rehydrated)	AlO_4	5.1	–	71.5	270
	AlO_5	5.1	–	44.0	270
	AlO_6	3.55	–	10.0	270
κ-Al_2O_3	AlO_4	7.6	0.3	81.5	271
	AlO_6-(1)	5.0	–	ca. 13	271
	AlO_6-(4)	8.5	–	18	271
χ-Al_2O_3	AlO_4	5.0	0.3	71.5	272
	AlO_5	2.7	0.3	38.5	272
	AlO_6	4.5	0.3	11.5	272
$Cd_8(AlO_2)_{12}S_2$ (sodalite)	AlO_4	2.00	<0.1	80.4	273
$Ca_8(AlO_2)_{12}S_2$ (sodalite)	AlO_4	3.55	<0.1	79.1	273
$Cd_8(AlO_2)_{12}Se_2$ (sodalite)	AlO_4	3.95	<0.1	78.7	273
$Cd_8(AlO_2)_{12}(SO_4)_2$ (sodalite)	AlO_4	3.24	<0.1	79.1	273
$Sr_8(AlO_2)_{12}S_2$ (sodalite)	AlO_4	4.65	<0.1	76.9	273
$Sr_8(AlO_2)_{12}Se_2$ (sodalite)	AlO_4	5.10	<0.1	76.6	273
$Sr_8(AlO_2)_{12}(CrO_4)_2$ (sodalite)	AlO_4	6.75	<0.1	75.5	273
$SrAl_{12}O_{19}$	AlO_4	3.45	0.1	68.0	145
	AlO_5	2.1	0.7	18.0	145
	AlO_6-(1)	0.6	1	17.1	145
	AlO_6-(2)	1.3	1	9.6	145
	AlO_6-(3)	4.9	0.63	21.7	145
Alumoborates					
$2SrO \cdot Al_2O_3 \cdot B_2O_3$	AlO_4	4.3	0.65	83.5	91
$2CaO \cdot Al_2O_3 \cdot B_2O_3$	AlO_4	6.25	0.45	79.5	91

(*continued overleaf*)

Table 5 (*continued*)

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Refs.
$2Li_2O{\cdot}Al_2O_3{\cdot}B_2O_3$	AlO_4	6.0	0.45	76	91
$3Li_2O{\cdot}Al_2O_3{\cdot}2B_2O_3$	AlO_4	6.7	0.83	70	91
$9Al_2O_3{\cdot}2B_2O_3$	AlO_4	6.8	0.1	53	91
	AlO_5	4.8	0.3	31	91
	AlO_6	6.2	0.4	10.5	91
Aluminumphosphates					
$AlPO_4$ (quartz)	AlO_4	4.2	0.35	44.8	274
$AlPO_4$ (tridimite)	AlO_4	0.75	0.95	39.8	274
$AlPO_4$ (cristobalite)	AlO_4	1.2	0.75	42.5	274
$Al_2PO_4(OH)_3$ (augelite)	AlO_5	5.5	0.78	30	275
	AlO_6	4.7	0.2	−3	275
$Al_2PO_4(OH)_3{\cdot}H_2O$ (senegalite)	AlO_5	2.87	$0_{assumed}$	30	275
	AlO_6	4.09	$0_{assumed}$	1.7	275
$KAlP_2O_7$	AlO_6	1.2	0.25	−16	276
$AlPO_4$-5 (molecular sieve)	AlO_4	2.3	0.95	40.4	274
$AlPO_4$-8 (dehydrated sieve)	AlO_4-(1)	3.9	$0.5_{assumed}$	40.1	84
	AlO_4-(2)	3.6	$0.5_{assumed}$	40.6	84
	AlO_4-(3)	3.6	$0.5_{assumed}$	47.0	84
	AlO_4-(4)	3.0	$0.5_{assumed}$	42.9	84
	AlO_4-(5)	3.4	$0.5_{assumed}$	42.6	84
$AlPO_4$-14 (molecular sieve)	AlO_4-(3)	1.75/1.74	0.70/0.63	43.2/42.9	83/277
	AlO_4-(2)	4.15/4.08	0.82/0.82	44.0/43.5	83/277
	AlO_5-(1)	5.66/5.58	0.89/0.97	27.2/27.1	83/277
	AlO_6-(4)	2.60/2.57	0.68/0.70	−0.9/−1.3	83/277
$AlPO_4$-21 (molecular sieve)	AlO_4	3.7	0.15	47.3	78
	AlO_5-I	5.9	0.68	14.6	78
	AlO_5-II	7.4	0.52	15.7	78
$AlPO_4$-25 (molecular sieve)	AlO_4-I	1.9	$0.67_{assumed}$	40.8	78
	AlO_4-II	0.8	$0.67_{assumed}$	39.5	78
$AlPO_4$-25 (dehydrated molecular sieve)	AlO_4-I	2.3	$0.67_{assumed}$	39.2	78
	AlO_4-II	1.1	$0.67_{assumed}$	37.5	78
Alumosilicates					
Al_2SiO_5 (sillimanite)	AlO_4	6.77	0.53	64.5	278
	AlO_6	8.93	0.46	4.0	278
Al_2SiO_5 (andalusite)	AlO_5	5.73	0.7	35	279
	AlO_6	15.5	0.0	10	279
Al_2SiO_5 (kyanite)	AlO_6-(1)	10.1	0.27	13.0	280
	AlO_6-(2)	3.8	0.85	4.0	280
	AlO_6-(3)	6.4	0.70	5.7	280
	AlO_6-(4)	9.2	0.38	5.9	280
Mullite ($\sim 3Al_2O_3{\cdot}1SiO_2$)	AlO_6	7.3	0	6.3	281
	AlO_4(T)	7.3	0	68	281
	AlO_4(T′)	6	0.5	53	281
	AlO_4(T*)	4	0.5	45	281
Pennine	AlO_4	2.8	–	72	278
	AlO_6	1.4	–	10	278
$KAlSi_2O_6$ (leucite)	T1-T3	2.1–2.3	$0_{assumed}$	61–69	282
$CaAl_2Si_2O_8$ (anorthite)	T1-T6	2.7–8.2	0.45–0.70	61–66	282
$KAl_2\{(OH,F)_2/AlSi_3O_{10}\}$	AlO_4	2.1	–	72	278
(muscovite)	AlO_6	2.2	–	5	278
$CaAl_2\{(OH)_2/Al_2Si_2O_{10}\}$	AlO_4	4.2	–	76	278
(margarite)	AlO_6	6.3	–	11	278
Xantophyllite	AlO_4	2.8	–	76	278
	AlO_6	2.0	–	11	278
$Na_8Al_2Be_2Si_8O_{24}Cl_2$ (tugtupide)	AlO_4	1.36	0.08	63.4	218

Table 5 (*continued*)

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Refs.
$NaAlSi_3O_8$ (low albite)	AlO_4	3.29	0.62	62.7	283
$KAlSi_3O_8$ (microcline)	AlO_4	3.22	0.21	58.5	283
$(Mg, Fe)Al_3SiBO_9$ (grandidierite)	AlO_5	8.7	0.95	41.0	284
	AlO_6-(1)	3.5	0.5	9.0	284
	AlO_6-(2)	8.6	0.95	11.0	284
Natrolite	AlO_4	1.67	0.50	64	278
LiCl-sodalite	AlO_4	0.98	0.59	71.9	217
LiBr-sodalite	AlO_4	0.71	0.61	70.9	217
NaCl-sodalite	AlO_4	0.94	0.32	62.9	217
NaBr-sodalite	AlO_4	0.81	0.29	61.8	217
NaI-sodalite	AlO_4	0.57	0.34	60.4	217
Na-A, hydrated zeolite	AlO_4	1.1	0.75	59.2	278
Na-Y, hydrated zeolite	AlO_4	2.0	0.5	62.8	278
Na-Y, dehydrated zeolite	AlO_4	5.5	0.3	ca. 60	285
H-Y, dehydrated zeolite	AlO_3	13.1	0.75	105 ± 20	285
	AlO_3	13.5	0.4	–	286
	AlO_3	15.3	0.4	60	287
	AlO_4	6.0	0.7	–	286
H-Y, dealuminated, dehydr.	AlO_3	13.7	0.5	–	286
	AlO_4	7.0	0.7	–	286
USY	AlO_4	2.8	–	60.0	270
	AlO_5	4.1	–	34.5	270
	AlO_6	2.9	–	4.0	270
HMOR, dehydrated zeolite	AlO_3	15.0	0.35	–	286
	AlO_4	6.8	0.7	–	286
Na-ZSM-5, dehydrated zeolite	AlO_4	4.7	0.5	ca. 60	285
H-ZSM-5, dehydrated zeolite	AlO_3	16.0/15.5	0.1/0.5	82 ± 20/–	285/286
	AlO_4	7.3	0.7	–	286
H, Al-MCM-41 (as synthesized)	AlO_4	2.3	$0_{assumed}$	52.6	288
Fluoroaluminates					
$H_3AlF_6 \cdot 6H_2O$	AlF_6	0.3	0.0	−2.8	289
$K_2AlF_5 \cdot H_2O$	AlF_6	12	0.0	0	289
$Rb_2AlF_5 \cdot H_2O$	AlF_6	13	0.0	0	289
$CsAlF_5 \cdot H_2O$	AlF_6	7.5	0.15	−10	289
NH_4AlF_4	AlF_6	10	0.1	−6	289
$KAlF_4$	AlF_6	12	0.0	−5	289
$RbAlF_4$	AlF_6	13	0.1	−4	289
$Al_2SiO_4F_2$ (topas)	AlF_6	1.7	0.4	0.3	289
Others					
$Al(acac)_3$	AlO_6	3.03	0.15	0.0	290
$Al(trop)_3$		4.43	0.08	36.6	290
$Al(TMHD)_3$	AlO_6	3.23	0.10	1.5	290
$AlCl_3 \cdot 3Al(OH)_3 \cdot 6H_2O$	AlO_6	6.9	0.4	7	291
$AlCl_3 \cdot 4Al(OH)_3 \cdot 7H_2O$	AlO_6	5.7	0.7	3	291
$AlCl_3 \cdot OPCl_3$	$AlCl_3O$	6.0	0.15	88	292
$Al_2Ge_2O_7$	AlO_5	8.8	0.4	36	293
$AlLaGe_2O_7$	AlO_5	7.2	0.37	36	293
$Al_2(MoO_4)_3$	AlO_6-(1)	1.12	0.65	−12.4	272
	AlO_6-(2)	0.88	0.95	−13.4	272
	AlO_6-(3)	1.21	1.0	−10.3	272
	AlO_6-(1)	0.78	0.8	−11.1	272
$Al_2(OH)_2(H_2O)_8(SO_4)_2 \cdot 2H_2O$	AlO_6	4.6	0.4	3	291
$Al_2(OH)_4SO_4 \cdot 7H_2O$ (aluminite)	AlO_6-I	10.1	0.1	6.9	291
	AlO_6-II	11.6	0.15	6.4	291
$KAl(SO_4)_2 \cdot 12H_2O$	AlO_6	0.400	0.00	−4.1	266
$NH_4Al(SO_4)_2 \cdot 12H_2O$	AlO_6	0.456	0.00	−0.4	266

[a] The data published in 1983–92 were compiled by Dirk Müller.

Table 6 Quadrupolar coupling constant $C_{qcc} = e^2qQ/h$, the asymmetry parameter η and the isotropic value of the chemical shift δ (refered to 1.0 M NaCl) for the ^{23}Na NMR of powdered substances at ambient temperature[a]

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Ref.
Sodium–nitrogen compounds					
$NaNO_3$		0.337	0.00	−8.0	266
$NaNO_2$		1.09	0.11	−0.1	266
NaN_3		0.297	0.12	−3.8	266
Aluminosilicates					
NaX (Si/Al = 1.0)	I	1.1	0.5	5.2*	294
	I′	5.8	0.0	−12.8*	294
	II	5.0	0.0	−8.8*	294
	III′(1,2)	2.2	0.7	−10.8*	294
	III′(3)	1.2	0.9	−22.8*	294
NaX (Si/Al = 1.23)	I	0.0	0.0	1.2*	74
	I′	5.2	0.0	−11.8*	74
	II	4.6	0.0	−7.8*	74
	III′(1,2)	2.6	0.7	−5.8*	74
	III′(3)	1.6	0.9	−21.8*	74
NaY (Si/Al = 2.5)	I	1.2	$0_{assumed}$	2.2*	146
	I′	4.8	0.0	3.2*	74
	II	3.9	0.0	−4.8*	74
EMT (Si/Al = 3.7)	I	1.0	$0_{assumed}$	0.7*	146
	I′ + II	4.1	0.3	0.2*	146
NaMOR (Si/Al = 7.1)	12-ring	2.0	$0_{assumed}$	−6.8*	146
	Sidepockets	3.1	$0_{assumed}$	−16.8*	146
NaZSM-5 (Si/Al = 18)		2.0	$0_{assumed}$	−10.8*	146
$NaAlSi_3O_8$ (albite)		2.69	0.25	−7.1	295
$Na_8Al_2Be_2Si_8O_{24}Cl_2$ (tugtupide)		1.41	0.44	7.7	218
Amphibole HSMC	M(4)	3.9	0.49	9.3	296
	A	2.9	0.26	5.5	296
NaCl-sodalite, dehydrated		0–0.5	–	−8.8	217
		∼0	$0.67_{assumed}$	6.3	71
NaCl-sodalite, hydrated		≤0.1	–	–	297
NaBr-sodalite, dehydrated		0.72	0.12	−9.9	217
		1	$0.67_{assumed}$	8.5	71
NaBr-sodalite, hydrated		0.6–0.8	–	–	297
NaI-sodalite, dehydrated		1.73	0.06	−20.6	217
		1.9	$0.67_{assumed}$	9.3	71
NaI-sodalite, hydrated		1.5–1.8	0	–	297
Na-hydroxosodalite		2.00	0.10	3.2*	298
Na-nitride sodalite		1.00	0.18	0.4*	298
Silicates					
$Na_2H_2SiO_4{\cdot}8H_2O$		1.11	0.72	3.8*	298
$Na_2H_2SiO_4{\cdot}5H_2O$	I	1.35	0.45	5.7*	298
	II	2.01	0.70	0.0*	298
$Na_2H_2SiO_4{\cdot}4H_2O$	I	1.80	0.75	9.0*	298
	II	2.83	0.17	9.5*	298
Others					
NaOH		3.5	0.00	19.4*	298
$Na_3P_3O_9$	I	1.57	0.55	1.6*	298
	II	2.20	0.70	−7.6*	298
$Na_2(OOCCH)_2{\cdot}H_2O$	I	1.34	0.80	−2.2*	298
	II	0.80	0.75	0.9*	298
$Na_2C_2O_4$		2.50	0.74	17.9	266
$CH_3COONa{\cdot}3H_2O$		0.779	0.38	2.1	266
Na_2SO_4		2.60	0.58	−1.3*	298

[a] An asterisk denotes values of the chemical shift which were originally referred to solid NaCl and transformed by δ(1 M NaCl) = δ(solid NaCl) + 7.2 ppm. Zeolite samples listed in this table were dehydrated.

Table 7 Quadrupolar coupling constant $C_{qcc} = e^2qQ/h$, the asymmetry parameter η and the isotropic value of the chemical shift δ (referred to H_2O) for the ^{17}O NMR of powder compounds at ambient temperature[a]

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Ref.
Aluminum oxides and hydroxides					
α-Al_2O_3 (corundum)	OAl_4	2.17	0.55	75	299
AlO(OH) (boehmite)	OAl_4	1.20	0.1	70	299
	OAl_4	1.15	0.13	70.0	266
	Al_2OH	5.0	0.5	40	299
$Al(OH)_3$ (bayerite)	Al_2OH	6.0	0.3	40	299
γ-Al_2O_3	OAl_4	1.8	$0_{assumed}$	73	299
η-Al_2O_3	OAl_4	1.6	$0_{assumed}$	73	299
δ-Al_2O_3	OAl_4	1.6	$0_{assumed}$	72	299
θ-Al_2O_3	OAl_4	1.2	$0_{assumed}$	72	299
	OAl_3	4.0	0.6	79	299
$Al_{13}O_{40}$ cluster	AlOAl	1.2	0	50	300
Microporous materials					
$AlPO_4$-5	AlOP	5.7	0.0	63	301
$AlPO_4$-11	AlOP	5.7	0.0	64	301
$AlPO_4$-17	AlOP	5.6	0.1	67	301
Ga-sodalite	SiOGa	4.0	0.3	29	301
	SiOSi	5.1	0.0	52	301
NaBa-Ga-sodalite	SiOGa	4.0	0.3	29	301
	SiOSi	5.1	0.0	52	301
Ga-X	SiOGa	4.0	0.3	28	301
	SiOSi	5.0	0.0	50	301
Na-Y	SiOAl	3.1	0.2	31	302
	SiOSi	4.6	0.2	46	302
Ba, Na-Y	SiOAl	3.4	0.2	45	302
	SiOSi	5.1	0.1	57	302
NH_4-Y	SiOAl	3.2	0.2	31	302
	SiOSi	5.0	0.1	47	302
Na-Y, dealuminated	SiOSi	5.2	0.1	47	302
Sil-Y, dehydrated	SiOSi O1	5.1	0.3	42.3	85
	SiOSi O4	5.28	0.2	34.8	85
	SiOSi O3	5.14	0.1	47.3	85
	SiOSi O2	5.39	0.2	37.2	85
Na-A, dehydrated	SiOAl	3.2	0.2	33	302
Na-A, hydrated	SiOAl O1	3.4	0	43.6	86
	SiOAl O3	3.4	0.25	40.5	86
	SiOAl O2	3.4	0	31.0	86
Na-LSX, hydrated	SiOAl O1	3.2	0.4	50.3	86
	SiOAl O3	3.4	0.3	45.0	86
	SiOAl O2	3.3	0.3	41.7	86
	SiOAl O4	3.6	0.15	36.9	86
Na-ZSM-5, hydrated	SiOSi	5.3	0.12	40.0	53
	SiOAl	3.5	0.29	30.0	53
Stilbite	SiOSi	5.1	0.18	43	303
	SiOAl	3.5	0.28	33	303
Silicates					
SiO_2 (low cristobalite)	SiOSi	5.3	0.0	46	302
SiO_2 (cristobalite)	SiOSi	5.3	0.125	40	304
SiO_2 (amorphous)	SiOSi	5.8	0.0	50	305
	SiOH	4.0	0.3	20	305
SiO_2 (stishovite)	SiOSi	6.5	0.125	109	306
SiO_2 (coesite)	SiOSi O5	5.16	0.292	58	307
	SiOSi O2	5.43	0.166	41	307
	SiOSi O3	5.45	0.168	57	307
	SiOSi O4	5.52	0.169	53	307
	SiOSi O1	6.05	0.000	29	307

(*continued overleaf*)

Table 7 (*continued*)

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Ref.
Mg_2SiO_4 (forsterite)	SiOMg-I	2.35	0.2	61	308
	SiOMg-II	2.35	1.0	62	308
	SiOMg-III	2.70	0.3	47	308
	SiOMg-I	2.8	$0_{assumed}$	64	309
	SiOMg-II	3.3	$0_{assumed}$	72	309
	SiOMg-III	3.0	$0_{assumed}$	49	309
$MgSiO_3$ (clinoenstatite)	SiOMg-I	3.2	0.0	60	310
	SiOMg-II	3.2	0.0	42	310
	SiOSi	5.1	0.3	62	310
	SiO-?	2.9–5.2	$0_{assumed}$	57–70	309
$CaMgSi_2O_6$ (diopside)	SiOCa	2.7	0.0	84	310
	SiOMg	2.7	0.1	63	310
	SiOSi	4.4	0.3	69	310
	SiOCa	2.8	$0_{assumed}$	86	309
	SiOMg	2.7	$0_{assumed}$	64	309
	SiOSi	4.5	$0_{assumed}$	69	309
$Mg_3Si_4O_{10}(OH)_2$ (talc)	SiOMg	3.2	0.0	40	305
	SiOSi	5.8	0.0	50	305
	MgOH	7.3	0.0	0	305
$Li_2Si_2O_5$	br O1	5.6	0.55	108	311
	br O2	4.05	0.05	35	311
	nb O3	2.45	0.1	38	311
α-$Na_2Si_2O_5$	br O1	5.74	0.2	52	311
	br O2	4.67	0.3	74	311
	nb O3	2.4	0.2	36	311
$K_2Si_2O_5$	br O1	5.1	0.1	114	311
	br O2	4.7	0.2	69	311
	nb O3	2.1	0.5	72	311
$Rb_2Si_2O_5$	br O1	4.4	0.1	124	311
	br O2	4.7	0.5	59	311
	nb O3	1.9	0.5	93	311
$K_2Si_4O_9$ (wadeite)	br O1	4.45	0.35	62.5	306
	SiOSi O2	4.9	0.2	97	306
$BaSiO_3$	br O	3.7	0.4	87	310
α-$SrSiO_3$	br O	4.1	0.4	80	310
α-$CaSiO_3$	br O	3.8	0.2	75	310
$CaSiO_3$ (wollastonite)	SiO-?	2.3–4.7	$0_{assumed}$	115–167	309
Ca_2SiO_4 (larnite)	SiO-?	2.5–2.8	$0_{assumed}$	122–134	309
Na-ilerite (RUB-18)	SiOSi	5.1	0	42.6	312
	SiOH	3.1	0	61.2	312
Aluminosilicate glasses					
YAS	br O	3.1	–	54	313
	nb O			143	313
	nb O			210	313
LAS	br O	3.1	–	58	313
	nb O			178	313
$NaAlSi_3O_8$	SiOSi	5.1	0.15	49	314
	SiOAl	3.2	0.05	33	314
$NaAlSi_3O_8$	SiOSi	5.2	0.2	40	315
	SiOAl	3.8	0.2	25	315
Analcime	H_2O	7.6	0	18	316
Silicate glasses					
$Na_2Si_4O_9$	SiOSi	5.2	0.22	51	315
	nb O	2.7	0.25	40	315
	H_2O	6.0	0.7	20	315
$Na_2Si_2O_5$	br O	4.9	0.1	69	311
	nb O	2.35	0.2	37	311

Table 7 (*continued*)

Compound	Site	C_{qcc} (MHz)	η	δ (ppm)	Ref.
$Na_2Si_3O_7$	br O	5.0	0	60	306
	nb O	2.3	0	39	306
$Li_2Si_2O_5$	br O	5.0	0.15	68	311
	nb O	2.55	0.2	42	311
$K_2Si_2O_5$	br O	4.7	0.25	60	311
	nb O	2.5	0.45	84	311
$K_2Si_4O_9$	br O	4.9	0	52	306
	nb O	2.3	0	76	306
$Cs_2Si_2O_5$	br O	4.55	0.3	68	311
	nb O	3.1	0.55	145	311
Ba Si glass	br O	4.0	0.3	78	317
Ba Ca Si glass	br O	4.1	0.3	68	317
Ca Si glass	br O	4.7	0.3	59	317
	br O	4.6	0.0	66	318
	nb O	2.1	0.2	110	318
Na_2O-GeO_2 crystals and glasses					
GeO_2 (quartz)	GeO_4	7.3	0.48	70	319
GeO_2 (rutile)	GeO_6	7.5	0.10	160	319
Na_2GeO_3 (crystal)	GeO_4	5.2	0.5	70	319
	NBO	2.5	0.5	47	
$Na_2O{\cdot}9GeO_2$	GeO_4	7.0	0.5	165	319
	GeO_6				
$2\ Na_2O{\cdot}9GeO_2$	GeO_4	6.0	0.5	80	319
Others					
TiO_2 (rutile)	TiOTi	1.5	0.87	596.5	249
$HfGeO_4$	HfOGe	5.2	0.65	185	320
$Mg(OH)_2$ (brucite)	MgOH	6.8	0	20	321
$Mg(OH)_2$ polycristalline	MgOH	6.8	0.0	25	305
$Mg(OH)_x(OCH_3)_{2-x}$	MgOH	7.25	0	−25	321
Hydroxyapatite $Ca_5(P^{17}O_4)_3(OH)$		4.0/4.1	0.0/0.1	108/115	322
$CaHP^{17}O_4{\cdot}2H_2O$		4.2/4.3	0/0	98/96	322
KH_2PO_4		5.2	0.55	92	322
$NH_4H_2PO_4$		5.1	0.55	93	322
$Ba(ClO_3)_2{\cdot}H_2{}^{17}O$	H_2O	6.8	1.00	22	322
$Ca(OH)_2$		6.5	0.00	62	322
$CaCO_3$		6.97	1	204	323

[a] The data published after 1989 were compiled by Ulf Pingel. A site nb O and br O denotes nonbridging and bridging oxygen, respectively.

If the quadrupolar echo decay $(\pi/2, t_1, \pi/2, t_1, t_2)$ is monitored as a 2-D data set, the doubly Fourier transformed 2-D spectrum can be influenced by a slow motion. Schleicher et al.[209] applied this technique for the observation of dynamics in alanine. Müller et al.[210] used a multiple-pulse quadrupolar echo sequence, in order to obtain information about the chemical exchange. For other examples, including the influence of dynamics on the longitudinal relaxation time in the Zeeman reservoir T_{1Z} or the relaxation time in the quadrupolar reservoir T_{1Q}, see Vold et al.[11,211]

Larsen et al.[212] applied the CPMG experiment to deuterons (^{2}H QCPMG) and thereby enhanced the sensitivity by about one order of magnitude and the dynamic range by two orders of magnitude.

Dynamic experiments were mainly performed with nonrotating samples. However, dynamic ^{2}H MAS NMR is now in progress. Weintraub and Vega[213] studied dimethyl sulfone by MAS and off-MAS techniques. Kristensen et al. investigated multiaxial dynamics[214] and effects of restricted rotational diffusion.[215]

10 SURVEY OF NUCLEAR MAGNETIC RESONANCE PARAMETERS FOR SELECTED COMPOUNDS

Collections of electric field gradient and chemical shift data for the most commonly studied quadrupolar nuclei with half-integer spins ^{27}Al, ^{23}Na and ^{17}O are presented in Tables 5, 6 and 7, respectively. For

solid-state NMR studies of other quadrupolar nuclei we give only a few references to recent publications: lithium-6,[216] lithium-7,[74,217] beryllium-9,[218] boron-11,[219–221] nitrogen-14,[222] magnesium-25,[223] sulfur-33,[224,225] chlorine-35/37,[226] potassium-39,[227,228] calcium-43,[229–231] scandium-45,[232] titanium-47/49,[233,234] vanadium-51,[235,236] chromium-53,[237] cobalt-59,[236,238,239] copper-65,[240] zinc-67,[241,242] gallium-69/71,[243–245] bromine-81,[246] rubidium-85/87,[247,248] zirconium-91,[249] niobium-93,[33,249] molybdenum-95,[250] indium-115,[251] antimony-121,[252] iodine-127,[253,254] cesium-133,[255,256] barium-137,[257,258] lanthanum-139[73,259] and lutetium-175.[260]

ACKNOWLEDGMENTS

I am grateful to Jean-Paul Amoureux, Horst Ernst, Jürgen Haase, Michael Hunger, Christian Jäger, Dieter Michel, Toralf Mildner, Dirk Müller, Ulf Pingel, Dagmar Prager, Ago Samoson, Mark E. Smith, Daniel Prochnow, and Ingo Wolf for contributions and advice. This work was supported by the Deutsche Forschungsgemeinschaft, SFB 294, project Fr 902/9, and the Graduiertenkolleg Physikalische Chemie der Grenzflächen.

ABBREVIATIONS AND ACRONYMS

CP	Cross-polarization
CPMAS	Cross-polarization Magic-angle Spinning
CPMG	Carr–Purcell–Meiboom–Gill
DACSY	Dynamic-angle Correlation Spectroscopy
DAS	Dynamic-angle Spinning
DFS	Double-frequency Sweeps
DOR	Double-rotation
DQ	Double-quantum
FAM	Fast Radiofrequency Amplitude Modulation
FID	Free Induction Decay
FSAHP	Frequency-stepped Adiabatic Half-passage
fwhm	Full Width at Half-maximum
HETCOR	Heteronuclear Chemical Shift Correlation
LAB	Laboratory Axis System
MAS	Magic-angle Spinning
MQDOR	Multiple-quantum Double-rotation
MQMAS	Multiple-quantum Magic-angle Spinning
MQVAS	Multiple-quantum Variable-angle Spinning
NMR	Nuclear Magnetic Resonance
NQR	Nuclear Quadrupolar Resonance
PAS	Principal Axis System
QEDOR	Quadrupolar Echo Double Resonance
REDOR	Rotational-echo Double Resonance
RF	Radiofrequency
RIACT	Rotation-induced Adiabatic Coherence Transfer
SATRAS	Satellite-transition Spectroscopy
SOLEDOR	Solid Echo Double Resonance
TPPI	Time-proportional Phase Increment
TRAPDOR	Transfer-of-populations Double-resonance
VAS	Variable-angle Spinning
2-D	Two-dimensional
3QMAS	Triple-quantum Magic-angle Spinning

RELATED ARTICLES

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General

REFERENCES

1. T. Vosegaard, E. Hald, V. Langer, H. Skov, P. Daugaard, H. Bildsoe, H. Jakobsen, 'Improved Hardware and Software for Single-crystal NMR Spectroscopy', *J. Magn. Reson.*, **135**, 126–132 (1998).
2. T. Vosegaard, E. Hald, P. Daugaard, H.J. Jakobsen, 'A Two-axis Goniometer for Sensitivity Enhancement in Single-crystal NMR Spectroscopy', *Rev. Sci. Instrum.*, **70**, 1771–1779 (1999).
3. O. Kanert, M. Mehring, 'Static Quadrupole Effects in Disordered Cubic Solids', *NMR Basic Principles Prog.*, **3**, 1–81 (1971).

4. B.V.G. Volkoff, H.E. Petch, D.W.I. Smellie, 'Nuclear Electric Quadrupole Interaction in Single Crystals', *Can. J. Phys.*, **30**, 270–289 (1952).
5. B.G.M. Volkoff, 'Second Order Nuclear Quadrupole Effects in Single Crystals', *Can. J. Phys.*, **31**, 820–836 (1953).
6. T.P. Das, A.K. Saha, 'Electric Quadrupole Interaction and Spin Echoes in Crystals', *Phys. Rev.*, **98**, 516–524 (1955).
7. M.H. Cohen, F. Reif, 'Quadrupole Effects in Nuclear Magnetic Resonance Studies of Solids', *Solid State Phys.*, **5**, 321–438 (1957).
8. H.W. Spiess, 'Deuteron NMR – A New Tool for Studying Chain Mobility and Orientation in Polymers', *Adv. Polym. Sci.*, **66**, 23–58 (1985).
9. D. Freude, J. Haase, 'Quadrupole Effects in Solid-state NMR', *NMR Basic Principles Prog.*, **29**, 1–90 (1993).
10. C. Jäger, 'Satellite Transition Spectroscopy of Quadrupolar Nuclei', *NMR Basic Principles Prog.*, **31**, 133–170 (1994).
11. G.L. Hoatson, R.L. Vold, '^{2}H-NMR Spectroscopy of Solids and Liquid Crystals', *NMR Basic Principles Prog.*, **32**, 1–67 (1994).
12. B.F. Chmelka, J.W. Zwanziger, 'Solid-state NMR Line Narrowing Methods for Quadrupole Nuclei – Double Rotation and Dynamic-angle Spinning', *NMR Basic Principles Prog.*, **33**, 79–124 (1994).
13. N. Chandrakumar, 'Spin-1 NMR', *NMR Basic Principles Prog.*, **34**, 1–117 (1996).
14. A.S. Ulrich, S.L. Grage, 'H-2 NMR (Solid-state NMR of Polymers)', *Stud. Phys. Theor. Chem.*, **84**, 190–211 (1998).
15. M.E. Smith, E.R.H. van Eck, 'Recent Advances in Experimental Solid State NMR Methodology for Half-integer Spin Quadrupole Nuclei [Review]', *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 159–201 (1999).
16. H. Chihara, N. Nakamura, *Nuclear Quadrupole Resonance Spectroscopy Data, Landold-Börnstein, New Series, Group III*, Springer, Berlin, Vol. 20, 1988.
17. J. Mason, *Multinuclear NMR*, Plenum Press, New York, 1989.
18. K. Schmidt-Rohr, H.W. Spiess, *Multidimensional Solid-state NMR and Polymers*, Academic Press, London, 1994.
19. A. Abragam, *Principles of Nuclear Magnetism*, Oxford University Press, Oxford, 1961.
20. M. Weissbluth, *Atoms and Molecules*, Academic Press, New York, London, 1978.
21. Z.W. Zheng, Z.H. Gan, N.K. Sethi, D.W. Alderman, D.M. Grant, 'An Efficient Simulation of Variable-angle Spinning Lineshapes for the Quadrupole Nuclei with Half-integer Spin', *J. Magn. Reson.*, **95**, 509–522 (1991).
22. M.E. Rose, *Elementary Theory of Angular Momentum*, John Wiley, New York, 1967.
23. U. Haeberlen, *High Resolution NMR in Solids, Selective Averaging*, Academic Press, New York, 1976.
24. T.P. Das, E.L. Hahn, 'Nuclear Quadrupole Resonance Spectroscopy', *Solid State Phys., Suppl.*, **1**, 1–223 (1958).
25. L. Frydman, J.S. Harwood, 'Isotropic Spectra of Half-integer Quadrupole Spins from Bidimensional Magic-angle Spinning NMR', *J. Am. Chem. Soc.*, **117**, 5367–5368 (1995).
26. S. Wimperis, 'Broadband and Narrowband Composite Excitation Sequences', *J. Magn. Reson.*, **86**, 46–59 (1990).
27. A.P.M. Kentgens, 'Quantitative Excitation of Half-integer Quadrupole Nuclei by a Frequency-stepped Adiabatic Half-passage', *J. Magn. Reson.*, **95**, 619–625 (1991).
28. D.W. Sindorf, V.J. Bartuska, 'Wide-line NMR Spectroscopy in Solids Using Variable Frequency Pulses', *J. Magn. Reson.*, **85**, 581–585 (1989).
29. R.Q. Fu, V.L. Ermakov, G. Bodenhausen, 'Divergent Double Chirp Pulses for Refocusing Quadrupole Interactions', *Solid State NMR*, **7**, 1–10 (1996).
30. J. Haase, M.S. Conradi, 'Sensitivity Enhancement for NMR of the Central Transition of Quadrupole Nuclei', *Chem. Phys. Lett.*, **209**, 287–291 (1993).
31. J. Haase, E. Oldfield, 'Aluminum to Oxygen Cross-polarization in Alpha-Al_2O_3 (Corundum)', *Solid State NMR*, **3**, 171–175 (1994).
32. A.P.M. Kentgens, R. Verhagen, 'Advantages of Double Frequency Sweeps in Static, MAS and MQMAS NMR of Spin $I = \frac{3}{2}$ Nuclei', *Chem. Phys. Lett.*, **300**, 435–443 (1999).
33. A.F. McDowell, M.S. Conradi, J. Haase, 'First-satellite Spectroscopy, a New Method for Quadrupole Spins', *J. Magn. Reson. A*, **119**, 211–218 (1996).
34. J. Haase, N.J. Curro, R. Stern, C.P. Slichter, 'New Double Resonance Technique for Quadrupole Nuclei', *Mol. Phys.*, **95**, 891–896 (1998).
35. T. Gullion, 'Measurement of Dipolar Interactions Between Spin-$\frac{1}{2}$ and Quadrupole Nuclei by Rotational-echo, Adiabatic-passage, Double-resonance NMR', *Chem. Phys. Lett.*, **246**, 25–330 (1995).
36. T. Gullion, 'Introduction to Rotational-echo, Double Resonance NMR', *Concepts in Magn. Reson.*, **10**, 277–289 (1998).
37. C.P. Grey, W.S. Veeman, A.J. Vega, 'Referenz', *J. Chem. Phys.*, **98**, 7711 (1993).
38. S.R. Hartmann, E.L. Hahn, 'Nuclear Double Resonance in the Rotating Frame', *Phys. Rev.*, **128**, 2042–2053 (1962).
39. C.P. Slichter, *Principles of Magnetic Resonance*, Springer, Berlin, 1990.
40. S. Vega, 'Multiple-quantum Cross-polarization NMR on Spin Systems with $I = \frac{1}{2}$ and $S = \frac{3}{2}$ in Solids', *Phys. Rev. A*, **23**, 3152–3173 (1981).
41. A.J. Vega, 'CP/MAS of Quadrupole $S = \frac{3}{2}$ Nuclei', *Solid State NMR*, **1**, 17–32 (1992).
42. A.J. Vega, 'MAS NMR Spin Locking of Half-integer Quadrupole Nuclei', *J. Magn. Reson.*, **96**, 50–68 (1992).

43. S. Hayashi, K. Hayamizu., 'Line Shapes in CP/MAS NMR Spectra of Half-integer Quadrupole Nuclei', *Chem. Phys. Lett.*, **203**, 319–324 and **205**, 597 (1993).
44. S. Hayashi, 'MAS NMR of Half-integer Quadrupole Nuclei – Effect of Spin-locking Efficiency on Powder Lineshapes', *Solid State NMR*, **3**, 93–101 (1994).
45. S.W. Ding, C.A. McDowell, 'Theoretical Calculations of the CPMAS Spectral Lineshapes of Half-integer Quadrupole Systems', *J. Magn. Reson. A*, **114**, 80–87 (1995).
46. S.W. Ding, C.A. McDowell, 'Application of Floquet Theory to Quadrupole Nuclear Spin-$\frac{5}{2}$ Nuclei in Solids Undergoing Sample Rotation', *Mol. Phys.*, **95**, 841–848 (1998).
47. W. Sun, J.T. Stephen, L.D. Potter, Y. Wu, 'Rotation-induced Resonance and Second-order Quadrupole Effects on Spin Locking of Half-integer Quadrupole Nuclei', *J. Magn. Reson. A*, **116**, 181–188 (1995).
48. J.P. Amoureux, 'High-resolution Solid-state NMR for Spin $\frac{3}{2}$ and $\frac{9}{2}$: The Multi-quantum Transitions Method', *Solid State NMR*, **2**, 83–88 (1993).
49. K. Narita, J.J. Umeda, H. Kusumoto, 'NMR Powder Patterns of the Second-order Nuclear-Quadrupole Interaction on the Central Line in Solids with Asymmetric Field Gradient', *J. Chem. Phys.*, **44**, 2719–2723 (1966).
50. D. Müller, 'Zur Bestimmung Chemischer Verschiebungen der NMR-Frequenzen bei Quadrupolkernen aus den MAS NMR-Spektren', *Ann. Phys.*, **39**, 451–460 (1982).
51. A. Samoson, 'Satellite Transition High-resolution NMR of Quadrupole Nuclei in Powders', *Chem. Phys. Lett.*, **119**, 29–32 (1985).
52. H.J. Jakobsen, J. Skibsted, H. Bildsoe, N.C. Nielsen, 'Magic-angle Spinning NMR Spectra of Satellite Transitions for Quadrupole Nuclei in Solids', *J. Magn. Reson.*, **85**, 173–180 (1989).
53. J.-P. Amoureux, F. Bauer, H. Ernst, C. Fernandez, D. Freude, D. Michel, U.-T. Pingel, 'O-17 Multiple-quantum and ^{1}H MAS NMR Studies of Zeolite ZSM-5', *Chem. Phys. Lett.*, **285**, 10–14 (1998).
54. A. Samoson, E. Lippmaa, A. Pines, 'High Resolution Solid State NMR Averaging of Second-order Effects by Means of a Double-rotor', *Mol. Phys.*, **65**, 1013–1018 (1988).
55. Y. Wu, B.Q. Sun, A. Pines, A. Samoson, E. Lippmaa, 'NMR Experiments with a New Double Rotor', *J. Magn. Reson.*, **89**, 297–309 (1990).
56. A. Samoson, E. Lippmaa, 'Synchronized Double-rotation NMR Spectroscopy', *J. Magn. Reson.*, **84**, 410–416 (1989).
57. A. Samoson, J. Tegenfeldt, 'Suppression of DOR Sidebands', *J. Magn. Reson. A*, **110**, 238–244 (1994).
58. D. Kuwahara, T. Nakai, 'Spinning-sideband Suppression in DOR NMR Spectra Using a Magic-angle-turning Technique', *Chem. Phys. Lett.*, **260**, 249–252 (1996).
59. B.Q. Sun, J.H. Baltisberger, Y. Wu, A. Samoson, A. Pines, 'Sidebands in Dynamic Angle Spinning (DAS) and Double Rotation (DOR) NMR', *Solid State NMR*, **1**, 267–295 (1992).
60. E. Cochon, J.P. Amoureux, 'Sideband Analysis in DOR NMR Spectra. 1. 'Infinite' Inner-rotor Speed', *Solid State NMR*, **2**, 205–222 (1993).
61. J.P. Amoureux, E. Cochon, 'Sideband Analysis in DOR NMR Spectra. 2. Real Finite Inner-rotor Speed', *Solid State NMR*, **2**, 223–234 (1993).
62. A. Samoson, A. Pines, 'Double Rotor for Solid-state NMR', *Rev. Sci. Instrum.*, **60**, 3239–3241 (1989).
63. D. Müller, M. Smith, L. Wunsch, 'NMR at More Than One Magic Angle', Bruker Report, Bruker, Karlsruhe, 31–32, 1990.
64. B.F. Chmelka, K.T. Mueller, A. Pines, J. Stebbins, Y. Wu, J.W. Zwanziger, 'Oxygen-17 NMR in Solids by Dynamic-angle Spinning and Double Rotation', *Nature*, **339**, 42–43 (1989).
65. K.T. Mueller, Y. Wu, B.F. Chmelka, J. Stebbins, A. Pines, 'High-resolution Oxygen-17 NMR of Solid Silicates', *J.Am. Chem. Soc.*, **113**, 32–38 (1991).
66. R. Jelinek, B.F. Chmelka, Y. Wu, M.E. Davis, J.G. Ulan, R. Gronsky, A. Pines, 'Adsorption Effects in Aluminophosphate Molecular Sieves Studied by Al-27 Double-rotation NMR', *Catal. Lett.*, **15**, 65–73 (1992).
67. G.J. Ray, A. Samoson, 'Double Rotation and Variable Field Al-27 NMR Study of Dealuminated Y-Zeolites', *Zeolites*, **13**, 410–413 (1993).
68. M. Hunger, G. Engelhardt, H. Koller, J. Weitkamp, 'Characterization of Sodium Cations in Dehydrated Faujasites and Zeolite EMT by Na-23 DOR, 2D-Nutation, and MAS NMR', *Solid State NMR*, **2**, 111–120 (1993).
69. H.A.M. Verhulst, W.J.J. Welters, G. Vorbeck, L.J.M. Vandeven, V.H.J. Debeer, R.A. Vansanten, J.W. Dehaan, 'New Assignment of the Signals in Na-23 DOR NMR to Sodium Sites in Dehydrated NaY Zeolite', *J. Phys. Chem.*, **98**, 7056–7062 (1994).
70. G.W. Haddix, M. Narayana, S.C. Tang, Y. Wu, 'Double-rotation NMR, Magic Angle Spinning NMR, and X-ray Diffraction Study of the Structure of Aluminum Molybdate', *J. Phys. Chem.*, **97**, 4624–4627 (1993).
71. R. Jelinek, S. Özkar, G.A. Ozin., 'Extraframework Sodium Cation Sites in Sodium Zeolite Y Probed by 23-Na Double-rotation NMR', *J. Am. Chem. Soc.*, **114**, 4907–4908 (1992).
72. R. Jelinek, S. Ozkar, G.A. Ozin, 'Intrazeolite Topotaxy – Na-23 Double-rotation NMR Study of Transition-metal Hexacarbonyls and Oxides Encapsulated in Sodium Zeolite-Y', *J. Phys. Chem.*, **96**, 5949–5953 (1992).
73. M. Hunger, G. Engelhardt, J. Weitkamp, 'Solid-state Na-23, La-139, Al-27 and Si-29 NMR Spectroscopic Investigations of Cation Location and Migration in

Zeolites LaNaY', *Microporous Mater.*, **3**, 497–510 (1995).
74. M. Feuerstein, M. Hunger, G. Engelhardt, J. Amoureux, 'Characterisation of Sodium Cations in Dehydrated Zeolite NaX by Na-23 NMR Spectroscopy', *Solid State NMR*, **7**, 95–103 (1996).
75. H. Koller, E. Meijer, R. Vansanten, 'Al-27 Quadrupole Interaction in Zeolites Loaded with Probe Molecules – A Quantum-chemical Study of Trends in Electric Field Gradients and Chemical Bonds in Clusters', *Solid State NMR*, **9**, 165–175 (1997).
76. Y. Wu, B. Chmelka, A. Pines, M.E. Davis, P.J. Grobet, P.A. Jacobs, 'High-resolution Al-27 NMR Spectroscopy for the Aluminophosphate Molecular Sieve VPI-5', *Nature*, **346**, 550–552 (1990).
77. P.J. Grobet, A. Samoson, H. Geerts, J.A. Martens, P.A. Jacobs, 'Crystallographic Site Discrimination in Aluminophosphate Molecular Sieves by Solid-state NMR – Influence of the Field-dependent Isotropic 2nd-order Quadrupole Shift', *J. Phys. Chem.*, **95**, 9620–9622 (1991).
78. R. Jelinek, B.F. Chmelka, Y. Wu, P.J. Grandinetti, A. Pines, P.J. Barrie, J. Klinowski, 'Study of the Aluminophosphates $AlPO_4$-21 and $AlPO_4$-25 by ^{27}Al Double-rotation NMR', *J. Am. Chem. Soc.*, **113**, 4097–4101 (1991).
79. Y. Wu, D. Lewis, J.S. Frye, A.R. Palmer, R.A. Wind, 'Cross-polarization Double-rotation NMR', *J. Magn. Reson.*, **100**, 425–430 (1992).
80. J. Jänchen, M.P.J. Peeters, J.W. Dehaan, L.J.M. Vandeven, J.H.C. Vanhooff, I. Girnus, U. Lohse, 'Adsorption Calorimetric Measurements and Al-27 DOR NMR Studies on the Molecular Sieve $AlPO_4$-18', *J. Phys. Chem.*, **97**, 12042–12046 (1993).
81. G.J. Kennedy, J.B. Higgins, C.F. Ridenour, H.X. Li, M.E. Davis, 'Multifield MAS and DOR NMR Studies of a Hydrated Aluminophosphate Molecular Sieve – $AlPO_4$-H_2', *Solid State NMR*, **4**, 173–178 (1995).
82. M. Kovalakova, P. Grobet, 'Double Rotation Al-27 NMR in Molecular Sieves', *Appl. Magn. Reson.*, **10**, 447–476 (1996).
83. C.A. Fyfe, H.M.Z. Altenschildesche, K.C. Wong-Moon, H. Grondey, J.M. Chezeau, '1D and 2D Solid State NMR Investigations of the Framework Structure of As-synthesized $AlPO_4$-14', *Solid State NMR*, **9**, 97–106 (1997).
84. M. Kovalakova, P. Grobet, 'The Al-27 DOR NMR Characterization of the Molecular Sieve $AlPO_4$-8', *Solid State NMR*, **9**, 107–113 (1997).
85. L. Bull, A. Cheetham, T. Anupold, A. Reinhold, A. Samoson, J. Sauer, B. Bussemer, Y. Lee, S. Gann, J. Shore, A. Pines, R. Dupree, 'A High-resolution O-17 NMR Study of Siliceous Zeolite Faujasite', *J. Am. Chem. Soc.*, **120**, 3510–3511 (1998).
86. U.T. Pingel, J.P. Amoureux, T. Anupold, F. Bauer, H. Ernst, C. Fernandez, D. Freude, A. Samoson, 'High-field O-17 NMR Studies of the SiOAl Bond in Solids', *Chem. Phys. Lett.*, **194**, 345–350 (1998).
87. C. Rohrig, I. Dierdorf, H. Gies, 'X-ray Powder Diffraction and NMR-spectroscopic Investigations on a Porous Zincosilicate Related to the Zeolite VPI-7 (VSV)', *J. Phys. Chem. Solids.*, **56**, 1369–1376 (1995).
88. R. Jelinek, B.F. Chmelka, A. Stein, G.A. Ozin, 'Multinuclear Magic-angle Spinning and Double-rotation NMR Study of the Synthesis and Assembly of a Sodalite Semiconductor Supralattice', *J. Phys. Chem.*, **96**, 6744–6752 (1992).
89. G. Engelhardt, P. Sieger, J. Felsche, 'Multinuclear Solid State NMR of Host–Guest Systems with TO_2 (T = Si, Al) Host-frameworks – A Case Study on Sodalites', *Anal. Chim. Acta*, **283**, 967–985 (1993).
90. T.J. Bastow, J.S. Hall, M.E. Smith, S. Steuernagel, 'Characterisation of Hydrated Aluminas by MAS and DOR Al-27 NMR', *Mater. Lett.*, **18**, 197–200 (1994).
91. D. Massiot, D. Müller, T. Hubert, M. Schneider, A.P.M. Kentgens, B. Cote, J.P. Coutures, W. Gessner, 'Double Rotation and Magic-angle Spinning Nuclear Magnetic Resonance Study of Al-27: Reexamination of the Aluminum Borate $9Al_2O_3 \cdot 2B_2O_3$', *Solid State NMR*, **5**, 175–180 (1995).
92. F. Taulelle, A. Samoson, T. Loiseau, G. Ferey, 'ULM-18, a Fluorinated Gallium Phosphate with Perforated Layers: XRD and NMR, Structure Determination, and HF Localization in a D4R', *J. Phys. Chem. B*, **102**, 8588–8598 (1998).
93. G. Jeschke, 'Spin Locking of $I = \frac{3}{2}$ Nuclei in Static and Spinning Samples – A Description by Abstract Spins and Floquet Formalism', *J. Chem. Phys.*, **108**, 907–917 (1998).
94. J.P. Amoureux, 'Variable-angle Double Rotation Technique: A New Two-dimensional High-resolution Technique for Quadrupole Nuclei', *Solid State NMR*, **4**, 229–239 (1995).
95. A. Llor, J. Virlet, 'Towards High-resolution NMR of More Nuclei in Solids: Sample Spinning with Time-dependent Spinner Axis Angle', *Chem. Phys. Lett.*, **152**, 248–253 (1988).
96. K.T. Mueller, E.W. Wooten, A. Pines, 'Pure-absorption-phase Dynamic-angle Spinning', *J. Magn. Reson.*, **92**, 620–627 (1991).
97. P.J. Grandinetti, J.H. Baltisberger, A. Llor, Y.K. Lee, U. Werner, M.A. Eastman, A. Pines, 'Pure-absorption-mode Lineshapes and Sensitivity in 2-Dimensional Dynamic-angle Spinning NMR', *J. Magn. Reson. A*, **103**, 72–81 (1993).
98. P.J. Grandinetti, Y.K. Lee, J.H. Baltisberger, B.Q. Sun, A. Pines, 'Sideband Patterns in Dynamic-angle-spinning NMR', *J. Magn. Reson. A*, **102**, 195–204 (1993).
99. C.A. Fyfe, K.C. Wongmoon, H. Grondey, K.T. Mueller, 'Increased Efficiency of Coherence Transfer in Cross-polarization NMR from Quadrupole

Nuclei – Dynamic-angle Cross-polarization', *J. Phys. Chem.*, **98**, 2139–2142 (1994).
100. T.P. Jarvie, R.M. Wenslow, K.T. Mueller, 'High-resolution Solid-state Heteronuclear Correlation NMR for Quadrupolar and Spin-$\frac{1}{2}$ Nuclei', *J. Am. Chem. Soc.*, **117**, 570–571 (1995).
101. R.M. Wenslow, K.T. Mueller, 'Cation Sites in Mixed-alkali Phosphate Glasses', *J. Non-Cryst. Solids*, **231**, 78–88 (1998).
102. A. Medek, J.R. Sachleben, P. Beverwyk, L. Frydman, 'Multi-rank Nuclear Magnetic Resonance Studies of Half-integer Quadrupole Nuclei in Solids by Three-dimensional Dynamic-angle Correlation Spectroscopy', *J. Chem. Phys.*, **104**, 5374–5382 (1996).
103. C. Fernandez, J.P. Amoureux, 'Triple-quantum MAS-NMR of Quadrupole Nuclei', *Solid State NMR*, **5**, 315–321 (1996).
104. S.P. Brown, S. Wimperis, 'Two-dimensional Multiple-quantum MAS NMR of Quadrupole Nuclei: A Comparison of Methods', *J. Magn. Reson.*, **128**, 42–61 (1997).
105. J.P. Amoureux, C. Fernandez, S. Steuernagel, 'Z Filtering in MQMAS NMR', *J. Magn. Reson. A*, **123**, 116–118 (1996).
106. J.P. Amoureux, C. Fernandez, L. Frydman, 'Optimized Multiple-quantum Magic-angle Spinning NMR Experiments on Half-integer Quadrupoles', *Chem. Phys. Lett.*, **259**, 347–355 (1996).
107. J.P. Amoureux, C. Fernandez, Y. Dumazy, 'A Useful Tool for the Elaboration of New Solid-state NMR Experiments: PULSAR', *J. Chim. Phys.*, **92**, 1939–1942 (1995).
108. J.P. Amoureux, C. Fernandez, 'Triple, Quintuple and Higher Orders Multiple Quantum MAS NMR of Quadrupole Nuclei', *Solid State NMR*, **10**, 211–224 (1998).
109. G. Wu, D. Rovnyak, R.G. Griffin, 'Quantitative Multiple-quantum Magic-angle-spinning NMR Spectroscopy of Quadrupole Nuclei in Solids', *J. Am. Chem. Soc.*, **118**, 9326–9332 (1996).
110. K.H. Lim, C.P. Grey, 'Analysis of the Anisotropic Dimension in the RIACT (II) Multiple Quantum MAS NMR Experiment for $I = \frac{3}{2}$ Nuclei', *Solid State NMR*, **13**, 101–112 (1998).
111. T. Mildner, M.E. Smith, R. Dupree, 'Rotationally Induced Triple Quantum Coherence Excitation in MAS NMR Spectroscopy of $I = \frac{5}{2}$ Spins', *Chem. Phys. Lett.*, **301**, 389–394 (1999).
112. S. Ding, C.A. McDowell, 'Shaped Pulse Excitation in Multi-quantum Magic-angle Spinning Spectroscopy of Half-integer Quadrupole Spin Systems', *Chem. Phys. Lett.*, **270**, 81–86 (1997).
113. S.W. Ding, C.A. McDowell, 'Multiple-quantum MAS NMR Spectroscopy of Spin-$\frac{3}{2}$ Quadrupole Spin Systems Using Shaped Pulses', *J. Magn. Reson.*, **135**, 61–69 (1998).
114. L. Marinelli, A. Medek, L. Frydman, 'Composite Pulse Excitation Schemes for MQMAS NMR of Half-integer Quadrupole Spins', *J. Magn. Reson.*, **132**, 88–95 (1998).
115. A. Goldbourt, P.K. Madhu, L. Frydman, S. Vega, 'Fast RF Amplitude Modulation in MQMAS NMR Spectroscopy', in *Abstracts of the 41st Rocky Mountain Conference*, 97–98, 1999.
116. J. Rocha, 'Direct Observation of Highly Distorted Hexacoordinated Aluminum in Andalusite by Very Fast Al-27 MAS NMR', *Chem. Commun.*, 2489 (1998).
117. D. Massiot, B. Touzo, D. Trumeau, J.P. Coutures, J. Virlet, P. Florian, P.J. Grandinetti, 'Two-dimensional Magic-angle Spinning Isotropic Reconstruction Sequences for Quadrupole Nuclei', *Solid State NMR*, **6**, 73–83 (1996).
118. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, London, 1987.
119. C.A. Fyfe, J. Skibsted, H. Grondey, H. Meyer zu Altenschildesche, 'Pulsed Field Gradient Multiple-quantum MAS NMR Spectroscopy of Half-integer Spin Quadrupolar Nuclei', *Chem. Phys. Lett.*, **281**, 44–48 (1997).
120. S.P. Brown, S.J. Heyes, S. Wimperis, 'Two-dimensional MAS Multiple-quantum NMR of Quadrupole Nuclei, Removal of Inhomogeneous Second-order Broadening', *J. Magn. Reson. A*, **119**, 280–284 (1996).
121. S.P. Brown, S.E. Ashbrook, S. Wimperis, 'Al-27 Multiple-quantum Magic Angle Spinning NMR Study of the Thermal Transformation Between the Microporous Aluminum Methylphosphonates AlMePO-beta and AlMePO-alpha', *J. Phys. Chem. B*, **103**, 812–817 (1999).
122. L. Delevoye, 'Quantification des Spectres MQMAS et Application a l'Etude des Materiaux Microporeux', Thesis, Lille, 1988.
123. P.P. Man, 'Scaling and Labeling the High-resolution Isotropic Axis of Two-dimensional Multiple-quantum Magic-angle-spinning Spectra of Half-integer Quadrupole Spins', *Phys. Rev. B*, **58**, 2764 (1998).
124. G. Wu, R.E. Wasylishen, 'The Influence of the Heteronuclear Dipolar Interaction on NMR Spectra of Quadrupole Nuclei', *Mol. Phys.*, **95**, 1177–1183 (1998).
125. U. Friedrich, I. Schnell, S.P. Brown, A. Lupulescu, D.E. Demco, H.W. Spiess, 'Spinning-sideband Patterns in Multiple-quantum Magic-angle Spinning NMR Spectroscopy', *Mol. Phys.*, **95**, 1209–1227 (1998).
126. L. Marinelli, L. Frydman, 'On the Origin of Spinning Side Bands in MQMAS NMR Experiments', *Chem. Phys. Lett.*, **275**, 188–198 (1997).
127. T. Charpentier, C. Fermon, J. Virlet, 'Efficient Time Propagation Technique for MAS NMR Simulation – Application to Quadrupole Nuclei', *J. Magn. Reson.*, **132**, 181–190 (1998).
128. T. Charpentier, C. Fermon, J. Virlet, 'Numerical and Theoretical Analysis of Multiquantum Magic-angle

Spinning Experiments', *J. Chem. Phys.*, **109**, 3116–3130 (1998).
129. J.P. Amoureux, M. Pruski, D.P. Lang, C. Fernandez, 'The Effect of RF Power and Spinning Speed on MQMAS NMR', *J. Magn. Reson.*, **131**, 170–175 (1998).
130. D. Massiot, 'Sensitivity and Lineshape Improvements of MQMAS by Rotor-synchronized Data Acquisition', *J. Magn. Reson. A*, **122**, 240–244 (1996).
131. T. Vosegaard, F.H. Larsen, H.J. Jakobsen, P.D. Ellis, N.C. Nielsen, 'Sensitivity-enhanced Multiple-quantum MAS NMR of Half-integer Quadrupole Nuclei', *J. Am. Chem. Soc.*, **119**, 9055–9056 (1997).
132. G. Wu, D. Rovnyank, B. Sun, R.G. Griffin, 'High-resolution Multiple Quantum MAS NMR Spectroscopy of Half-integer Quadrupole Nuclei', *Chem. Phys. Lett.*, **249**, 210–217 (1996).
133. M. Hanaya, R.K. Harris, 'Effect of ^{1}H-decoupling in Two-dimensional Multiple-quantum MAS NMR Spectroscopy of Na-23 in a Hydrous Layered Silicate', *Solid State NMR*, **8**, 147–151 (1997).
134. V. Lacassagne, P. Florian, V. Montoullout, C. Gervais, F. Babonneau, D. Massiot, 'Resolution Enhancement in Solid-state MQMAS Experiments Achieved by Composite Decoupling', *Magn. Reson. Chem.*, **36**, 956–959 (1998).
135. S.H. Wang, Z. Xu, J.H. Baltisberger, L.M. Bull, J.F. Stebbins, A. Pines, 'Multiple-quantum Magic-angle Spinning and Dynamic-angle Spinning NMR Spectroscopy of Quadrupolar Nuclei', *Solid State NMR*, **8**, 1–16 (1997).
136. M. Pruski, D.P. Lang, C. Fernandez, J.P. Amoureux, 'Multiple-quantum Magic-angle Spinning NMR with Cross-polarization: Spectral Editing of High-resolution Spectra of Quadrupole Nuclei', *Solid State NMR*, **7**, 327–331 (1997).
137. C. Fernandez, L. Delevoye, J.P. Amoureux, '^{27}Al{^{1}H} Cross Polarization Triple Quantum Magic Angle Spinning NMR', *J. Am. Chem. Soc.*, **119**, 6858–6862 (1997).
138. S.H. Wang, S.M. De Paul, L.M. Bull, 'High-resolution Heteronuclear Correlation Between Quadrupole and Spin-$\frac{1}{2}$ Nuclei Using Multiple-quantum Magic-angle Spinning', *J. Magn. Reson.*, **125**, 364–368 (1997).
139. C. Magnenet, D. Massiot, I. Klur, J.P. Coutures, 'Characterization of TiO_2 pigments by Al-27 NMR', *J. Chim. Phys.*, **95**, 310–316 (1998).
140. C. Fernandez, D.P. Lang, J.P. Amoureux, M. Pruski, 'Measurement of Heteronuclear Dipolar Interactions Between Quadrupole and Spin-$\frac{1}{2}$ Nuclei in Solids by Multiple-quantum REDOR NMR', *J. Am. Chem. Soc.*, **120**, 2672–2673 (1998).
141. M. Pruski, C. Fernandez, D.P. Lang, J.P. Amoureux, 'Measurement of Interatomic Connectivities in Molecular Sieves Using MQMAS-based Methods', *Catal. Today*, **10**, 401–409 (1999).
142. S. Steuernagel, 'A New Sequence for the MQMAS/HETCOR Experiment', *Solid State NMR*, **11**, 197–201 (1998).
143. A. Samoson, 'Two-dimensional Isotropic NMR of Quadrupole Nuclei in Solids', *J. Magn. Reson.*, **121**, 209–211 (1996).
144. T. Anupold, A. Reinhold, P. Sarv, A. Samoson, 'A Comparison of Double Rotation and Multi-quantum Magic Angle Spinning Spectra', *Solid State NMR*, **13**, 87–91 (1998).
145. S.R. Jansen, H.T. Hintzen, R. Metselaar, J.W. de Haan, L.J.M. van de Ven, A.P.M. Kentgens, G.H. Nachtegaal, 'Multiple Quantum Al-27 Magic-angle-spinning Nuclear Magnetic Resonance Spectroscopic Study of $SrAl_{12}O_{19}$: Identification of an Al-27 Resonance from a Well-defined AlO_5 Site', *J. Phys. Chem.*, **102**, 5969–5976 (1998).
146. M. Hunger, P. Sarv, A. Samoson, 'Two-dimensional Triple-quantum Na-23 MAS NMR Spectroscopy of Sodium Cations in Dehydrated Zeolites', *Solid State NMR*, **9**, 115–120 (1997).
147. C. Jäger, P. Hartmann, G. Kunath-Fandrei, O. Hirsch, P. Rehak, J. Vogel, M. Feike, H.W. Spiess, K. Herzog, B. Thomas, 'Novel Opportunities of Structural Investigations of Glasses by 2D NMR', *Ber. Bunsenges. Phys. Chem.*, **100**, 1560–1562 (1996).
148. L. Zuchner, J.C.C. Chan, W. Müller-Warmuth, H. Eckert, 'Short-range Order and Site Connectivities in Sodium Aluminoborate Glasses: I. Quantification of Local Environments by High-resolution B-11, Na-23, and Al-27 Solid-state NMR', *J. Phys. Chem. B*, **102**, 4495 (1998).
149. H.C. Torrey, 'Transient Nutations in NMR', *Phys. Rev.*, **76**, 1059–1068 (1949).
150. A. Samoson, E. Lippmaa, 'Excitation Phenomena and Line Intensities in High-resolution NMR Spectra of Half-integer Quadrupole Nuclei', *Phys. Rev. B*, **28**, 6567 (1983).
151. A. Samoson, E. Lippmaa, 'Central Transition NMR Excitation Spectra of Half-integer Quadrupole Nuclei', *Chem. Phys. Lett.*, **100**, 205–208 (1983).
152. D. Fenzke, D. Freude, T. Fröhlich, J. Haase, 'NMR Intensity Measurements of Half-integer Quadrupole Nuclei', *Chem. Phys. Lett.*, **111**, 171–175 (1984).
153. A. Wokaun, R.R. Ernst, 'Selective Excitation and Detection in Multilevel Spin Systems: Application of Single Transition Operators', *J. Chem. Phys.*, **67**, 1752 (1977).
154. L. Pandey, S. Towta, D.G. Hughes, 'NMR Pulse Response and Measurement of the Quadrupole of $I = \frac{3}{2}$ Nuclei', *J. Chem. Phys.*, **85**, 6923–6927 (1986).
155. P.P. Man, 'Investigation of the Central Line of Mn-55 in $KMnO_4$ by a Two-dimensional NMR Method', *J. Magn. Reson.*, **67**, 78–90 (1986).
156. R. Janssen, W.S. Veeman, 'Quadrupole Nutation NMR in Solids', *J. Chem. Soc., Faraday Trans.*, **184**, 3747–3759 (1988).

157. J.A.M.v.d. Mijden, R. Janssen, W.S. Veeman, 'Analytical Description of the $I = \frac{5}{2}$ Quadrupole Nutation Experiment', *Mol. Phys.*, **69**, 53–64 (1990).
158. P.P. Man, 'Analytical Expression for the Spin-$\frac{5}{2}$ Line Intensities', *Mol. Phys.*, **78**, 307–318 (1993).
159. P.P. Man, P. Tougne, 'Exact Expression for the Spin-$\frac{7}{2}$ Line Intensities: Application to Solid State Co-59(III) NMR', *Mol. Phys.*, **83**, 997–1009 (1994).
160. S.Z. Ageev, B.C. Sanctuary, 'Analytical Solutions for Spin-$\frac{7}{2}$ Line Intensities in Solid State NMR', *Mol. Phys.*, **84**, 835–844 (1995).
161. A.P.M. Kentgens, J.J.M. Lemmens, F.M.M. Geurts, W.S. Veeman, 'Two-dimensional Solid-state Nutation NMR of Half-integer Quadrupole Nuclei', *J. Magn. Reson.*, **71**, 62–74 (1987).
162. A.P.M. Kentgens, 'Off-resonance Nutation NMR Spectroscopy of Half-integer Quadrupole Nuclei', *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**, 141–164 (1998).
163. S.W. Ding, C.A. McDowell, 'Nutation Magic-angle Spinning Spectra of Half-integer Quadrupole Spin Systems', *J. Magn. Reson. A*, **112**, 36–42 (1995).
164. A.P.M. Kentgens, 'Off-resonance Nutation NMR Spectroscopy of Half-integer Quadrupole Nuclei', *J. Magn. Reson. A*, **114**, 302–309 (1993).
165. E.L. Hahn, 'Spin Echoes', *Phys. Rev.*, **80**, 580–594 (1950).
166. J. Haase, E. Oldfield, 'Spin-echo Behavior of Non-integral Spin Quadrupole Nuclei in Inorganic Solids', *J. Magn. Reson. A*, **101**, 30–40 (1993).
167. I. Furo, B. Halle, 'Multiple Quantum NMR Spectroscopy on $I > 1$ Nuclei in Anisotropic Systems', *Mol. Phys.*, **76**, 1169–1197 (1992).
168. I. Furo, B. Halle, '2-D Quadrupole-echo Spectroscopy with Coherence Selection and Optimized Pulse Angle', *J. Magn. Reson.*, **98**, 388–407 (1992).
169. U. Eliav, G. Navon, 'A Study of Dipolar Interactions and Dynamic Processes of Water Molecules in Tendon by H-1 and H-2 Homonuclear and Heteronuclear Multiple-quantum-filtered NMR Spectroscopy', *J. Magn. Reson.*, **137**, 295–310 (1999).
170. U. Eliav, G. Navon, 'Corrections and Additions To ''Quadrupole-echo Techniques in Multiple-quantum-filtered NMR Spectroscopy of Heterogeneous Systems'', JMR 1995', *J. Magn. Reson.*, **128**, 98 (1997).
171. U. Eliav, G. Navon, 'Quadrupole-echo Techniques in Multiple-quantum-filtered NMR Spectroscopy of Heterogeneous Systems', *J. Magn. Reson. A*, **115**, 241–253 (1995).
172. U. Eliav, T. Kushnir, T. Knubovets, Y. Itzchak, G. Navon, 'The Effect of Magnetic Field Inhomogeneity on the Transverse Relaxation of Quadrupole Nuclei Measured by Multiple Quantum Filtered NMR', *J. Magn. Reson.*, **128**, 82–86 (1997).
173. K.D. Schmitt, J. Haase, E. Oldfield, 'Spectral Editing: A Quantitative Application of Spin-echo-nuclear Magnetic Resonance Spectroscopy to the Study of Zeolite Catalysts', *Zeolites*, **14**, 89 (1994).
174. Y. Dumazy, J.P. Amoureux, C. Fernandez, 'Theoretical and Experimental Study of Quadrupole Echoes in Solid State NMR', *Mol. Phys.*, **90**, 959–970 (1997).
175. P. Man, 'Second-order Quadrupole Effects on Hahn Echoes in Fast-rotating Solids at the Magic Angle [Review]', *Phys. Rev. B*, **55**, 8406–8424 (1997).
176. P. Man, 'Numerical Analysis of Solomon Echo Amplitudes in Static Solids', *J. Chem. Phys.*, **105**, 3908–3919 (1997).
177. S.Z. Ageev, P.P. Man, J. Fraissard, B.C. Sanctuary, 'Determination of Quadrupole Spin Coupling for Spin-$\frac{7}{2}$ Using Two Pulse Sequences', *Mol. Phys.*, **91**, 75–80 (1997).
178. H.Y. Carr, E.M. Purcell, 'Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments', *Phys. Rev.*, **80**, 630–638 (1954).
179. S. Meiboom, D. Gill, 'Modified Spin-echo Method for Measuring Nuclear Relaxation Times', *Rev. Sci. Instrum.*, **29**, 688–691 (1958).
180. F.H. Larsen, H.J. Jakobsen, P.D. Ellis, N.C. Nielsen, 'Sensivity-enhanced Quadrupole-echo NMR of Half-integer Quadrupole Nuclei. Magnitudes and Relative Orientation of Chemical Shielding and Quadrupole Coupling Tensors', *J. Phys. Chem.*, **101**, 8597–8606 (1997).
181. F.H. Larsen, H.J. Jakobsen, P.D. Ellis, N.C. Nielsen, 'High-field QCPMG-MAS NMR of Half-integer Quadrupolar Nuclei with Large Quadrupole Couplings', *Mol. Phys.*, **95**, 1185–1195 (1998).
182. G.E. Pake, 'Nuclear Resonance Adsorption in Hydrated Crystals: Fine Structure of the Proton Line', *J. Chem. Phys.*, **16**, 327–336 (1948).
183. G. Hoatson, 'Broadband Composite Excitation Sequences for Creating Quadrupole Order in H-2 NMR', *J. Magn. Reson.*, **94**, 152–159 (1991).
184. G.L. Hoatson, B. Blümich, 'Deuteron Hadamard NMR of Solids and Liquid Crystals', *J. Magn. Reson.*, **95**, 446–451 (1991).
185. B. Blümich, 'Wideline ^{2}H-NMR Spectroscopy and Imaging of Solids', *Z. Naturforsch., Teil A*, **49**, 19–26 (1994).
186. D.-K. Yang, J.E. Atkins, C.C. Lester, D.B. Zax, 'New Developments in NMR Using Noise Spectroscopy', *Mol. Phys.*, **95**, 747–757 (1998).
187. D.-K. Yang, D.B. Zax, 'Bandwidth Extension in Noise Spectroscopy', *J. Magn. Reson.*, **135**, 267–270 (1998).
188. S. Habot, D.L. Lu, N.J. Tagg, G.R. Gall, D.J. Siminovitch, 'Time Symmetry – An Application to Shaped Pulse Excitation of Spin-1 Systems', *Solid State NMR*, **10**, 137–150 (1998).
189. M.A. McCoy, W.S. Warren, 'Pulse Shaping in Quadrupole-echo NMR', *J. Magn. Reson.*, **98**, 24–35 (1992).
190. A.E. Aliev, K.D.M. Harris, D.C. Apperley, 'Natural Abundance High-resolution Solid State H-2 NMR Spectroscopy', *Chem. Phys. Lett.*, **226**, 193–198 (1994).

191. R. Poupko, Z. Olender, D. Reichert, Z. Luz, 'Deuterium MAS NMR in Natural Abundance', *J. Magn. Reson. A*, **106**, 113–115 (1994).
192. S. Vega, T.W. Shattuck, A. Pines, 'Double-quantum Cross-polarization NMR in Solids', *Phys. Rev. A*, **22**, 638–661 (1980).
193. Y. Ba, C.I. Ratcliffe, J.A. Ripmeester, 'Quadrupole Echo Double Resonance (QEDOR) and Solid Echo Double Resonance (SOLEDOR) NMR', *Chem. Phys. Lett.*, **299**, 201–206 (1999).
194. D. Marks, N. Zumbulyadis, S. Vega, 'Deuterium Cross-polarization Magic-angle Spinning', *J. Magn. Reson. A*, **122**, 16–36 (1996).
195. Z. Gan, P. Robyr, 'Deuterium Polarization Transfer in Rotating Solids and Its Application in Structural Investigation', *Mol. Phys.*, **95**, 1143–1152 (1998).
196. J.H. Kristensen, H. Bildsoe, H.J. Jakobsen, N.C. Nielsen, 'Deuterium Quadrupole Couplings from Least-squares Computer Simulations of ^{2}H MAS NMR Spectra', *J. Magn. Reson.*, **92**, 443–453 (1991).
197. B. Blümich, P. Blümler, J. Jansen, 'Presentation of Sideband Envelopes by Two-dimensional One Pulse (TOP) Spectroscopy', *Solid State NMR*, **1**, 111–113 (1992).
198. H. Ernst, D. Freude, I. Wolf, 'Multinuclear Solid-state NMR Studies of Brönsted Sites in Zeolites', *Chem. Phys. Lett.*, **212**, 588–596 (1993).
199. T.P. Spaniol, A. Kubo, T. Terao, 'Two-dimensional Deuterium MAS NMR of Paramagnetic Compounds – Separation of Paramagnetic and Quadrupole Interactions', *J. Chem. Phys.*, **106**, 5393–5405 (1997).
200. S. Vega, A. Pines, 'Operator Formalism for Double Quantum NMR', *J. Chem. Phys.*, **66**, 5624–5644 (1977).
201. R. Eckman, L. Mueller, A. Pines, 'Deuterium Double-quantum NMR with MAS', *Chem. Phys. Lett.*, **74**, 376–378 (1980).
202. Y. Yang, M. Schuster, B. Blümich, H.W. Spiess, 'Dynamic Magic-angle Spinning NMR Spectroscopy: Exchange Induced Sidebands', *Chem. Phys. Lett.*, **139**, 239–243 (1987).
203. N. Chandrakumar, G.v. Fircks, H. Günther, 'The 2-D Quadshift Experiment. Separation of Deuterium Chemical Shifts and Quadrupole Couplings by Two-dimensional Solid-state MAS NMR Spectroscopy', *Magn. Reson. Chem.*, **32**, 433–435 (1994).
204. A. Detken, H. Zimmermann, U. Haeberlen, Z. Luz, 'Dynamic Hydrogen Disorder in Solid Tropolone. A Single-crystal NMR Study of Hydroxyl Deuterons', *J. Magn. Reson.*, **126**, 95–102 (1997).
205. P. Speier, H. Zimmermann, U. Haeberlen, Z. Luz, 'Dynamic Disorder in 2,3-Dimethylnaphthalene: A Single Crystal Deuteron Nuclear Magnetic Resonance Study', *Mol. Phys.*, **95**, 1153–1167 (1998).
206. P. Freude, D. Michel, J. Totz, K.A, 'Ordering Behavior at the Antiferrodistortive Phase Transition in Betaine Phosphate and Betaine Phosphite', *Ferroelectrics*, **208**, 93–103 (1998).
207. A. Kubo, A. Yogo, F. Imashiro, T. Terao, 'Deuterium NMR Study of the Glassy Crystal Pentachlorotoluene – Hadamard Quadrupole-order Exchange NMR', *J. Phys. Chem.*, **100**, 15 933–15 941 (1996).
208. N. Bloembergen, E.M. Purcell, R.V. Pound, 'Relaxation Effects in NMR Absorption', *Phys. Rev.*, **102**, 679–712 (1948).
209. A. Schleicher, K. Müller, G. Kothe, '2-D NMR Relaxation Spectroscopy of Solids', *J. Chem. Phys.*, **92**, 6432 (1990).
210. K. Müller, R. Poupko, Z. Luz, '2-D Presentation of the Quadrupole-echo Modulation in Chemically Exchanging Systems of Spin $I = 1$', *J. Magn. Reson.*, **93**, 291–298 (1991).
211. R.R. Vold, R.L. Vold, 'Deuterium Relaxation in Molecular Solids', *Adv. Magn. Opt. Reson.*, **16**, 85–171 (1991).
212. F.H. Larsen, H.J. Jakobsen, P.D. Ellis, N.C. Nielsen, 'Molecular Dynamics from H-2 Quadrupole Carr–Purcell–Meiboom–Gill Solid-state NMR Spectroscopy', *Chem. Phys. Lett.*, **292**, 467–473 (1998).
213. O. Weintraub, S. Vega, 'Dynamic H-2 NMR of Rotating Solids', *Solid State NMR*, **4**, 341–351 (1995).
214. J.H. Kristensen, G.L. Hoatson, R.L. Vold, 'Investigation of Multiaxial Molecular Dynamics by H-2 MAS NMR Spectroscopy', *Solid State NMR*, **13**, 1–37 (1998).
215. J.H. Kristensen, G.L. Hoatson, R.L. Vold, 'Effect of Restricted Rotational Diffusion on H-2 MAS NMR Spectra', *J. Chem. Phys.*, **110**, 4533–4553 (1999).
216. Y.J. Lee, F. Wang, C.P. Grey, 'Li-6 and Li-7 MAS NMR Studies of Lithium Manganate Cathode Materials', *J. Am. Chem. Soc.*, **120**, 12601–12613 (1998).
217. N.C. Nielsen, H. Bildsoe, H.J. Jakobsen, P. Norby, 'Li-7, Na-23, and Al-27 Quadrupole Interactions in Some Aluminosilicate Sodalites from MAS NMR Spectra of Satellite Transitions', *Zeolites*, **11**, 622–633 (1991).
218. J. Skibsted, P. Norby, H. Bildsoe, H.J. Jakobsen, 'Line Shapes and Widths of MAS Sidebands for Al-27 Satellite Transitions – Multinuclear MAS NMR of Tugtupite $Na_8Al_2Be_2Si_8O_{24}Cl_2$', *Solid State NMR*, **5**, 239–255 (1995).
219. P.J. Bray, S.J. Gravina, P.E. Stallworth, S.P. Szu, J. Zhong, 'NMR Studies of the Structure of Glasses', *Exp. Tech. Phys.*, **36**, 397–413 (1988).
220. W. Müller-Warmuth, 'The Structure of Glasses Studied by MAS-NMR Spectroscopy of Quadrupole Nuclei', *Z. Naturforsch., Teil A*, **51**, 585–590 (1996).
221. S.J. Hwang, C. Fernandez, J.P. Amoureux, J. Cho, S.W. Martin, M. Pruski, 'Quantitative Study of the Short Range Order in B_2O_3 and B_2S_3 by MAS and Two-dimensional Triple-quantum MAS ^{11}B NMR', *Solid State NMR*, **8**, 109–121 (1997).
222. T.J. Bastow, D. Massiot, J.P. Coutures, 'N-14 NMR in AlN and BN', *Solid State Nucl. Magn. Reson.*, **10**, 241–245 (1998).

223. K.J.D. Mackenzie, R.H. Meinhold, 'MAS NMR Study of Pentacoordinated Magnesium in Grandidierite', *Am. Mineral.*, **82**, 479–482 (1997).
224. T.J. Bastow, S.N. Stuart, 'NMR Study of the Zinc Chalcogenides (ZnX, X = 0, S, Se, Te)', *Phys. Status Solidi B*, **145**, 719–728 (1988).
225. W.A. Daunch, P.L. Rinaldi, 'Natural-abundance Solid-state S-33 NMR with High-speed Magic-angle Spinning', *J. Magn. Reson. A*, **123**, 219–221 (1996).
226. M.A. Fedotov, O.L. Malkina, V.G. Malkin, 'Cl-35/37 NMR Chemical Shifts and Nuclear Quadrupole Couplings for Some Small Chlorine Compounds – Experimental and Theoretical Study of Chemical Shifts and Nuclear Quadrupole Couplings for Some Small Chlorine Compounds', *Chem. Phys. Lett.*, **258**, 330–335 (1996).
227. T.J. Bastow, S.L. Segel, K.R. Jeffrey, 'Antiferroelectric Transition in KOH(D) – A H-2 and K-39 NMR Study', *Solid State Commun.*, **78**, 565–568 (1991).
228. R. Blinc, J. Dolinsek, T. Apih, W. Schranz, A. Fuith, H. Warhanek, '2D K-39 NMR Study of the Phase Transition in KSCN', *Solid State Commun.*, **93**, 609–611 (1995).
229. P.V. Bellot, A. Trokiner, Y. Zhdanov, A. Yakubovskii, 'Ca-43 NMR in Solid State', *J. Chim. Phys.*, **95**, 280–288 (1998).
230. R. Dupree, A.P. Howes, S.C. Kohn, 'Natural Abundance Solid State Ca-43 NMR', *Chem. Phys. Lett.*, **276**, 399–404 (1997).
231. R.G. Bryant, S. Ganapathy, S.D. Kennedy, 'High-resolution Ca-43 NMR in Solids', *J. Magn. Reson.*, **72**, 376–378 (1987).
232. A.R. Thompson, E. Oldfield, 'Solid-state Scandium-45, Yttrium-89, and Lanthanum-139 NMR Spectroscopy', *J. Chem. Soc., Chem. Commun.*, 27–29 (1987).
233. A. Labouriau, W.L. Earl, 'Titanium Solid-state NMR in Anatase, Brookite and Rutile', *Chem. Phys. Lett.*, **270**, 278–284 (1997).
234. S.F. Dec, M.F. Davis, G.E. Maciel, C.E. Bronnimann, J.J. Fitzgerald, S.S. Han, 'Solid-state Multinuclear NMR Studies of Ferroelectric, Piezoelectric, and Related Materials', *Inorg. Chem.*, **32**, 955–959 (1993).
235. J. Skibsted, N.C. Nielsen, H. Bildsoe, H.J. Jakobsen, 'Magnitudes and Relative Orientation of V-51 Quadrupole Coupling and Anisotropic Shielding Tensors in Metavanadates and KV_3O_8 from V-51 MAS NMR Spectra – Na-23', *J. Am. Chem. Soc.*, **115**, 7351–7362 (1993).
236. S. Hayashi, 'V-51 and CO-59 Off-MAS NMR Spectra – Determination of Quadrupole Coupling, Chemical Shift Anisotropy and Their Relative Orientation', *Magn. Reson. Chem.*, **34**, 791–798 (1996).
237. G.N. Abelyashev, V.N. Berzhanskiii, S.N. Polulyah, N.A. Sergeev, 'Pecularities of the Formation of a Two Pulse Echo in Spin-$\frac{3}{2}$ Systems', *Fiz. Tverd. Tela*, **34**, 676–678 (1992) (in Russian).
238. F.H. Larsen, H.J. Jakobsen, P.D. Ellis, N.C. Nielsen, 'QCPMG-MAS NMR of Half-integer Quadrupole Nuclei', *J. Magn. Reson.*, **131**, 144–147 (1998).
239. T. Charpentier, J. Virlet, 'Triple Quantum MQMAS Spectroscopy of Co-59 ($I = \frac{7}{2}$) in $Na_3Co(NO_2)_6$ and *trans*-$Co[(en_2)(NO_2))_2]NO_3$ Interplay Between the Quadrupole Coupling and Anisotropic Shielding Tensors', *Solid State NMR*, **12**, 227–242 (1998).
240. S.L. Hu, J.A. Reimer, A.T. Bell, 'Cu-65 NMR Spectroscopy of Cu-exchanged ZSM-5 Catalysts', *J. Phys. Chem. B*, **101**, 1869–1871 (1997).
241. G. Wu, 'Zinc-67 NMR Spectroscopy of Solids', *Chem. Phys. Lett.*, **298**, 375–380 (1998).
242. T.J. Bastow, 'Zn-67 NMR in Zinc Metal', *J. Phys. Condens. Matter*, **8**, 11309–11315 (1996).
243. D. Massiot, V. Montouillout, F. Fayon, F.P.C. Bessada, 'QPASS – Towards Higher Resolution in NMR of Half Integer Quadrupole Nuclei with High Quadrupole Couplings', *J. Chim. Phys.*, **95**, 270–279 (1998).
244. T. Vosegaard, D. Massiot, N. Gautier, H.J. Jakobsen, 'Ga-71 Chemical Shielding and Quadrupole Coupling Tensors of the Garnet $Y_3Ga_5O_{12}$ from Single-crystal Ga-71 NMR', *Inorg. Chem.*, **36**, 2446–2450 (1997).
245. R. Klik, V. Bosacek, L. Kubelkova, D. Freude, D. Michel, 'Coordination State of Gallium in MFI Structures Prepared by Direct Synthesis and by Postsynthetic Modification of Boralites', *Zeolites*, **19**, 343–348 (1997).
246. R. Jelinek, A. Stein, G.A. Ozin, 'Br-81 and Cl-35 NMR Investigation of Sodium, Silver, Halo Sodalite Semiconductor Supralattices', *J. Am. Chem. Soc.*, **115**, 2390–2396 (1993).
247. T. Vosegaard, J. Skibsted, H. Bildsoe, H.J. Jakobsen, 'Combined Effect of Second-order Quadrupole Coupling and Chemical Shielding Anisotropy on the Central Transition in MAS NMR of Quadrupole Nuclei. Rb-87 MAS NMR', *J. Phys. Chem.*, **99**, 10731–10735 (1995).
248. T. Vosegaard, I.P. Byriel, H.J. Jakobsen, T. Vosegaard, I.P. Byriel, H.J. Jakobsen, 'Characterization of the Two Rubidium Sites in Rb_2CrO_4 by Rb-87 Single-crystal NMR', *J. Phys. Chem. B*, **101**, 8955–8958 (1997).
249. T.J. Bastow, 'Materials Characterisation by Nuclear Quadrupole Interaction', *Z. Naturforsch., Teil A*, **49**, 320–328 (1994).
250. V.M. Mastikhin, O.B. Lapina, R.I. Maximovskaya, 'Mo-95 Solid-state NMR Spectra of Molybdates', *Chem. Phys. Lett.*, **148**, 413–416 (1988).
251. O.H. Han, H.K.C. Timken, E. Oldfield, 'Solid-state MAS NMR Spectroscopic Study of Group III–V Semiconductors', *J. Chem. Phys.*, **89**, 6046–6052 (1988).
252. T. Ueda, N. Nakamura, 'Sb-121 NMR and SCF-MS-X-alpha Studies of Quadrupole Interaction and the Electronic Structure of Mixed-valence Compound, Cs_2SbC_{16}', *Z. Naturforsch., Teil A*, **51**, 672–676 (1996).
253. N. Lee, B.C. Sanctuary, T.K. Halstead, 'I-127 NMR Study of Quadrupole Echoes in KI', *J. Magn. Reson.*, **98**, 534–555 (1992).

254. G. Hinze, R. Bohmer, B. Zalar, R. Blinc, 'The Orientational Glass $(KI)_{(0.56)}(NH_4I)_{(0.44)}$ studied by I-127 NMR', *J. Phys. Condens. Matter*, **9**, 117–125 (1997).
255. S. Kroeker, K. Eichele, R. Wasylishen, J. Britten, 'Cesium-133 NMR Study of CsCD(SCN)(3) – Relative Orientation of the Chemical Shift and Electric Field Gradient Tensors', *J. Phys. Chem. B*, **101**, 3727–3733 (1997).
256. J. Skibsted, T. Vosegaard, H. Bildsoe, H.J. Jakobsen, 'Cs-133 Chemical Shielding Anisotropies and Quadrupole Couplings from MAS NMR of Cesium Salts', *J. Phys. Chem.*, **100**, 14 872–14 881 (1996).
257. K. Kumagai, Y.V. Piskunov, A.Y. Yakubovskii, S.V. Verkhovskii, A.P. Gerashenko, Y.I. Zhdanov, K.N. Mikhalev, I.V. Sagaradze, E.Y. Medvedev, S.M. Cheshnitskii, R.N. Pletnev, 'Local Distortions in the BiO_6-octahedra Sublattice $BaPb_XBi_{1-X}O_3$ as seen by Ba-137 NMR/NQR', *Physica C*, **247**, 209–220 (1997).
258. T.J. Bastow, 'An NMR Study of Ba-137 and Ti-47, 49 in Ferroelectric $BaTiO_3$', *J. Phys. Condens. Matter*, **1**, 4985–4991 (1989).
259. T.J. Bastow, 'La-139 Nuclear Magnetic Resonance Characterisation of La_2O_3 and $La_{1-x}Sr_xMO_3$ Where M = Cr, Mn or Co', *Solid State NMR*, **3**, 17–22 (1994).
260. P.J. Chu, B.C. Gerstein, H.D. Tang, R.N. Shelton, 'Studies of the Electronic Phase Transition in $Lu_5Ir_4Si_{10}$ and $Lu_5Rh_4Si_{10}$ by Lu-175 NMR', *Phys. Rev. B*, **37**, 1796–1806 (1988).
261. D. Müller, W. Gessner, A. Samoson, E. Lippmaa, G. Scheler, 'Solid-state Al-27 NMR Studies on Polycrystalline Aluminates of the Systems $CaO{\cdot}Al_2O_3$', *Polyhedron*, **5**, 779 (1986).
262. W.S. Veeman, 'Quadrupole Nutation NMR in Solids', *Z. Naturforsch., Teil A*, **47**, 353–360 (1992).
263. D. Müller, W. Gessner, A. Samoson, E. Lippmaa, G. Scheler, 'Solid-state Aluminum-27 Nuclear Magnetic Resonance Chemical Shift and Quadrupole Coupling Data for Condensed AlO_4 Tetrahedra', *J. Chem. Soc., Dalton Trans*, 1277–1281 (1986).
264. J. Skibsted, H. Bildsoe, H.J. Jakobsen, 'High-speed Spinning Versus High Magnetic Field in MAS NMR of Quadrupole Nuclei. Al-27 MAS NMR of 3Ca Al_2O_3', *J. Magn. Reson.*, **92**, 669–676 (1991).
265. D. Müller, W. Gessner, G. Scheler, 'Chemical Shift and Quadrupole Coupling of the Al-27 NMR Spectra of $LiAlO_2$ Polymorphs', *Polyhedron*, **2**, 1195–1200 (1984).
266. J. Skibsted, N.C. Nielsen, H. Bildsöe, H.J. Jacobsen, 'Satellite Transitions in MAS NMR Spectra of Quadrupolar Nuclei', *J. Magn. Reson.*, **95**, 88–117 (1991).
267. T. Vosegaard, H.J. Jakobsen, 'Al-27 Chemical Shielding Anisotropy', *J. Magn. Reson.*, **128**, 135–137 (1997).
268. H. Kraus, R. Prins, A.P.M. Kentgens, 'A Al-27 MQMAS and Off-resonance Nutation NMR Investigation of Mo-P/gamma-Al_2O_3 Hydrotreating Catalyst Precursors', *J. Phys. Chem.*, **100**, 16 336–164 345 (1996).
269. H. Kraus, M. Muller, R. Prins, A.P.M. Kentgens, 'Comments on the Al-27 NMR Visibility of Aluminas', *J. Phys. Chem. B*, **102**, 3862–3865 (1998).
270. D. Coster, A.L. Blumenfeld, J.J. Fripiat, 'Lewis Acid Sites and Surface Aluminum in Aluminas and Zeolites: A High-resolution NMR Study', *J. Phys. Chem.*, **98**, 6201–6211 (1994).
271. B. Ollivier, R. Retoux, P. Lacorre, D. Massiot, G. Ferey, 'Crystal Structure of kappa-Alumina – An X-ray Powder Diffraction, TEM and NMR Study', *J. Mater. Chem.*, **7**, 1049–1056 (1997).
272. G. Kunathfandrei, T.J. Bastow, J.S. Hall, C. Jäger, M.E. Smith, 'Quantification of Aluminum Coordinations in Amorphous Aluminas by Combined Central and Satellite Transition Magic Angle Spinning NMR Spectroscopy', *J. Phys. Chem.*, **99**, 15 138–15 141 (1995).
273. M.T. Weller, M.E. Brenchley, D.C. Apperley, N.A. Davies, 'Correlations Between Al-27 MAS NMR Spectra and the Coordination Geometry of Framework Aluminates', *Solid State NMR*, **3**, 103–106 (1994).
274. D. Müller, E. Jahn, G. Ladwig, U. Haubenreisser, 'High-resolution Solid-state 27-Al and 31P NMR: Correlation Between Chemical Shift and Mean Al–O–P angle in $AlPO_4$ Polymorphs', *Chem. Phys. Lett.*, **109**, 332–336 (1984).
275. W.F. Bleam, S.F. Dec, J.S. Frye, 'Al-27 Solid-state NMR Study of Five-coordinate Aluminum in Augelite and Senegalite', *Phys. Chem. Miner.*, **16**, 817–820 (1989).
276. D. Müller, I. Grunze, E. Hallas, G. Ladwig, 'Hochfeld-Al-27-NMR-Untersuchungen zur Aluminiumkoordination in kristallinen Aluminiumpohosphaten', *Z. Anorg. Allg. Chem.*, **500**, 80–88 (1983).
277. C. Fernandez, J.P. Amoureux, J.M. Chezeau, L. Delmotte, H. Kessler, 'Al-27 MAS NMR Characterization of $AlPO_4$-14. Enhanced Resolution and Information by MQMAS', *Microporous Mater.*, **6**, 331–340 (1996).
278. E. Lippmaa, A. Samoson, M. Mägi, 'High-resolution Al-27 NMR of Aluminosilicates', *J. Am. Chem. Soc.*, **108**, 1730–1735 (1986).
279. L.B. Alemany, D. Massiot, B.L. Sheriff, M.E. Smith, F. Taulelle, 'Observation and Accurate Quantification of Al-27 MAS NMR Spectra of Some Al_2SiO_5 Polymorphs Containing Sites with Large Quadrupole Interactions', *Chem. Phys. Lett.*, **177**, 301–306 (1991).
280. L.B. Alemany, S. Steuernagel, J.P. Amoureux, R.L. Callender, A.R. Barron, 'Very Fast MAS and MQMAS NMR Studies of the Spectroscopically Challenging Minerals Kyanite and Andalusite on 400, 500, 800 MHz Spectrometers', *Solid State NMR*, **14**, 1–18 (1999).
281. G. Kunathfandrei, P. Rehak, S. Steuernagel, H. Schneider, C. Jäger, 'Quantitative Structural Analysis of Mullite by Al-27 Nuclear Magnetic Resonance Satellite Transition Spectroscopy', *Solid State NMR*, **3**, 241–248 (1994).
282. J.H. Baltisberger, Z. Xu, J.F. Stebbins, S.H. Wang, A. Pines, 'Triple-quantum Two-dimensional Al-27 MAS

NMR Spectroscopic Study of Aluminosilicate and Aluminate Crystals and Glasses', *J. Am. Chem. Soc.*, **118**, 7209–7214 (1996).
283. R.J. Kirkpatrick, R.A. Kinsey, K.A. Smith, D.M. Henderson, E. Oldfield, 'High Resolution Solid-state Sodium-23, Aluminum-27, and Silicon-29 Nuclear Magnetic Resonance Spectroscopic Reconnaissance of Alkali and Plagioclase Feldspars', *Am. Mineral.*, **70**, 106–123 (1985).
284. M.E. Smith, S. Steuernagel, 'A Multinuclear Magnetic Resonance Examination of the Mineral Grandidierite', *Solid State NMR*, **1**, 175–183 (1992).
285. D. Freude, H. Ernst, I. Wolf, 'Solid-state NMR Studies of Acid Sites in Zeolites', *Solid State NMR*, **3**, 271–286 (1994).
286. M. Hunger, T. Horvath, 'Multi-nuclear Solid-state NMR Study of the Local Structure of SiOHAl Groups and Their Interaction with Probe-molecules in Dehydrated Faujasite, Mordenite and Zeolite ZSM-5', *Ber. Bunsenges. Phys. Chem.*, **99**, 1316–1320 (1995).
287. C.P. Grey, A.J. Vega, 'Determination of the Quadrupole of the Invisible Aluminum Spins in Zeolite HY with H-1/Al-27 TRAPDOR NMR', *J. Am. Chem. Soc.*, **117**, 8232–8242 (1995).
288. H. Kosslick, G. Lischke, G. Walther, W. Storek, A. Martin, R. Fricke, 'Physico-chemical and Catalytic Properties of Al-, Ga- and Fe-substituted Mesoporous Materials Related to MCM-41', *Microporous Mater.*, **9**, 13–33 (1997).
289. D. Müller, U. Bentrup, 'Al-27 NMR Studies on Alkali Fluoroaluminates', *Z. Anorg. Allg. Chem.*, **575**, 17–25 (1989).
290. R.W. Schurko, R.E. Wasylishen, H. Foerster, 'Characterization of Anisotropic Aluminum Magnetic Shielding Tensors. Distorted Octahedral Complexes and Linear Molecules', *J. Phys. Chem. A*, **102**, 9750–9760 (1998).
291. D. Müller, personal communication, 1992.
292. R.W. Schurko, R.E. Wasylishen, A.D. Phillips, 'A Definitive Example of Aluminum-27 Chemical Shielding Anisotropy', *J. Magn. Reson.*, **133**, 388–394 (1998).
293. D. Massiot, A. Kahn-Harari, D. Michel, D. Müller, F. Taulelle, 'Aluminum-27 MAS NMR of $Al_2Ge_2O_7$ and $LaAlGe_2O_7$: Two Pentacoordinated Aluminum Environments', *Magn. Reson. Chem.*, **28**, 82–88 (1990).
294. M. Feuerstein, G. Engelhardt, P.L. McDaniel, J.E. MacDougall, T.R. Gaffney, 'Solid-state Nuclear Magnetic Resonance Investigation of Cation Siting in LiNaLSX Zeolites', *Microporous Mater.*, **16**, 27–35 (1998).
295. A.M. George, J.F. Stebbins, 'High-temperature Na-23 MAS NMR Data for Albite – Comparison to Chemical-shift Models', *Am. Mineral.*, **80**, 878–884 (1995).
296. L. Delevoye, S.X. Liu, M.D. Welch, C. Fernandez, J.P. Amoureux, J. Klinowski, 'Triple-quantum Al-27 and Na-23 MAS NMR Study of Amphiboles', *J. Chem. Soc., Faraday Trans.*, **93**, 2591–2595 (1997).
297. R. Janssen, G.A.H. Tijink, W.S. Veeman, T.L.M. Maesen, J.F.v. Lent, 'High-temperature NMR Study of Zeolite Na-A: Detection of a Phase Transition', *J. Phys. Chem.*, **93**, 899–904 (1989).
298. G. Engelhardt, H. Koller, 'A Simple Procedure for the Determination of the Quadrupole Interaction Parameters and Isotropic Chemical Shifts from MAS NMR Spectra of Half-integer Spins in Solids', *Magn. Reson. Chem.*, **29**, 941–945 (1991).
299. T.H. Walter, E. Oldfield, 'Magic-angle Spinning Oxygen-17 NMR of Aluminum Oxides and Hydroxides', *J. Phys. Chem.*, **93**, 6744–6751 (1989).
300. A.R. Thompson, A.C. Kunwar, H.S. Gutowsky, E. Oldfield, 'Oxygen-17 and Aluminum-27 NMR Spectroscopic Investigations of Aluminum(III) Hydrolysis Products', *J. Chem. Soc., Dalton Trans.*, 2317–2322 (1987).
301. H.K.C. Timken, N. James, G.L. Turner, S.L. Lambert, L.B. Welsh, E. Oldfield, 'Solid-state Oxygen-17 NMR Spectroscopic Studies of Zeolites and Related Systems. Part 2', *J. Am. Chem. Soc.*, **108**, 7236–7241 (1986).
302. H.K.C. Timken, G.L. Turner, J.P. Gilson, L.B. Welsh, E. Oldfield, 'Solid-state Oxygen-17 NMR Spectroscopic Studies of Zeolites and Related Systems. Part 1', *J. Am. Chem. Soc.*, **108**, 7231–7235 (1986).
303. Z. Xu, J.F. Stebbins, 'Oxygen Sites in the Zeolite Stilbite: A Comparison of Static, MAS, VAS, DAS and Triple Quantum MAS NMR Techniques', *Solid State NMR*, **11**, 243–251 (1998).
304. D.R. Spearing, I. Farnan, J.F. Stebbins, 'Dynamics of the Alpha–Beta Phase Transitions in Quartz and Cristobalite as Observed by In Situ High Temperature Si-29 NMR and O-17 NMR', *Phys. Chem. Miner.*, **19**, 307–321 (1992).
305. T.H. Walter, G.L. Turner, E. Oldfield, 'Oxygen-17 Cross-polarization NMR Spectroscopy of Inorganic Solids', *J. Magn. Reson.*, **76**, 106–120 (1988).
306. X. Xue, J.F. Stebbins, M. Kanzaki, 'Correlations Between O-17 NMR Parameters and Local Structure Around Oxygen in High-pressure Silicates: Implications for the Structure of Silicate Melts at High Pressure', *Am. Mineral.*, **79**, 31–42 (1994).
307. P.J. Grandinetti, J.H. Baltisberger, I. Farnan, J.F. Stebbins, U. Werner, A. Pines, 'Solid-state O-17 Magic-angle and Dynamic-angle Spinning NMR Study of the SiO_2 Polymorph Coesite', *J. Phys. Chem.*, **99**, 12 341–12 348 (1995).
308. S. Schramm, E. Oldfield, 'High-resolution Oxygen-17 NMR of Solids', *J. Am. Chem. Soc.*, **106**, 2502–2506 (1984).
309. K.T. Mueller, J.H. Baltisberger, E.W. Woote, A. Pines., 'Isotropic Chemical Shifts and Quadrupole Parameters for Oxygen-17 Using Dynamic-angle Spinning NMR', *J. Phys. Chem.*, **96**, 7001–7004 (1992).
310. H.K.C. Timken, S.E. Schramm, R.J. Kirkpatrick, E. Oldfield, 'Solid-state Oxygen-17 NMR Spectroscopic Studies of Alkaline Earth Metasilicates', *J. Phys. Chem.*, **91**, 1054–1058 (1987).

311. H. Maekawa, P. Florian, D. Massiot, H. Kiyono, M. Nakamura, 'Effect of Alkali Metal Oxide on O-17 NMR Parameters and Si−O−Si Angles of Alkali Metal Disilicate', *J. Phys. Chem.*, **100**, 5525–5532 (1996).
312. U. Pingel, 'O-17 NMR of Porous Solids', Thesis, Universität Leipzig, 1999.
313. T. Schaller, J.F. Stebbins, 'The Structural Role of Lanthanum and Yttrium in Aluminosilicate Glasses: A Al-27 and O-17 MAS NMR Study', *J. Phys. Chem. B*, **102**, 10 690–10 697 (1998).
314. P.J. Dirken, S.C. Kohn, M.E. Smith, E.R.H. van Eck, 'Complete Resolution of Si−O−Si and Si−O−Al Fragments in an Aluminosilicate Glass by O-17 Multiple Quantum Magic Angle Spinning NMR Spectroscopy', *Chem. Phys. Lett.*, **266**, 568–574 (1997).
315. H. Maekawa, T. Saito, T. Yokokawa, 'Water in Silicate Glass: O-17 NMR of Hydrous Silica, Albite, and $Na_2Si_4O_9$ Glasses', *J. Phys. Chem. B*, **102**, 7523–7529 (1998).
316. Z. Xu, H. Maekawa, J.V. Oglesby, J.F. Stebbins, 'Oxygen Speciation in Hydrous Silicate Glasses: An Oxygen-17 NMR Study', *J. Am. Chem. Soc.*, **120**, 9894 (1998).
317. J.F. Stebbins, J.V. Oglesby, Z. Xu, 'Disorder Among Network-modifier Cations in Silicate Glasses – New Constraints from Triple-quantum O-17 NMR', *Am. Mineral.*, **82**, 1116–1124 (1997).
318. X. Cong, R.J. Kirkpatrick, 'O-17 MAS NMR Investigation of the Structure of Calcium Silicate Hydrate Gel', *J. Am. Ceram. Soc.*, **79**, 1585–1592 (1996).
319. R. Hussin, D. Holland, R. Dupree, 'Does Six-coordinate Germanium Exist in Na_2O-GeO_2 glasses – O-17 NMR Measurements', *J. Non-cryst. Solids*, **234**, 440–445 (1998).
320. T.J. Bastow, M.E. Smith, H.J. Whitfield, 'O-17 NMR Investigation of Hafnia and Ternary Hafnium Oxides', *J. Mater. Chem.*, **6**, 1951–1955 (1996).
321. E.R.H. Van Eck, M.E. Smith, 'Orientation of the Quadrupole and Dipole Tensors of Hydroxyl Groups by O-17 Quadrupole Separated Local Field NMR', *J. Chem. Phys.*, **108**, 5904–5912 (1998).
322. G. Wu, D. Rovnyak, P.C. Huang, R.G. Griffin, 'High-resolution Oxygen-17 NMR Spectroscopy of Solids by Multiple-quantum Magic-angle-spinning', *Chem. Phys. Lett.*, **277**, 79–83 (1997).
323. M.E. Smith, S. Steuernagel, H.J. Whitfield, 'O-17 Magic-angle Spinning Nuclear Magnetic Resonance of $CaCO_3$', *Solid State NMR*, **4**, 313–316 (1995).

Quadrupole Couplings in Nuclear Magnetic Resonance, General

Pascal P. Man
Université Pierre et Marie Curie, Paris, France

1 Introduction **1**
2 Frequency-domain Response **2**
2.1 Quadrupole Interaction 2
2.2 Energy Levels and Spectrum 3
3 Time-domain Response **9**
3.1 One-pulse Sequence: Two-dimensional Nutation Experiment 10
3.2 Composite-pulse Sequences: Spurious Signal Suppression 18
3.3 Two-pulse Sequences 19
3.4 Solomon Echo Sequences 19
3.5 Hahn Echo Sequences 24
4 Perspective and Future Developments **37**
List of Symbols **37**
Abbreviations and Acronyms **38**
Related Articles **38**
References **38**

Nuclear magnetic resonance (NMR) spectroscopy is continually finding new applications. It enables the local symmetry to be probed at the atomic scale using the nuclear spins I *of the compound under investigation. The nuclear spin is either a half-integer (or odd) number or an integer (or even) number. The nuclei in the periodic table can be divided into two parts – spin-$\frac{1}{2}$ nuclei and spin larger than $\frac{1}{2}$ nuclei. The spin larger than $\frac{1}{2}$ nuclei are called quadrupole nuclei because they possess an electric quadrupole moment which interacts with the electric-field gradient (EFG) generated by its surroundings. By extension, their spins are called quadrupole spins. Spin-$\frac{1}{2}$ nuclei are not sensitive to the EFG. Of the nuclei that possess a spin, 6% have integer quadrupole spins and 66% have half-integer quadrupole spins.*

*This article focuses on the half-integer quadrupole spins (*I $= \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$, *and* $\frac{9}{2}$*) in single crystals and in powder compounds. Most of these spins are observable. As they are multi-energy-level systems (the number of energy levels is* $2\mathrm{I}+1$*), multiple quantum (MQ) transitions occur during excitation of the spin system by a radiofrequency (RF) pulse sequence. As a result, quantum mechanical concepts are needed for an understanding of the spin dynamics and for interpretation of the results. In particular, the choice of pulse sequence and the experimental conditions, such as pulse duration, pulse strength, and phase cycling in the pulse sequence, depend on the strength of the EFG surrounding the nuclear spin.*

1 INTRODUCTION

Most nuclei in the periodic table have a spin I larger than $\frac{1}{2}$. These spins are called quadrupole spins and they are sensitive to the EFG generated by their surroundings. The coupling of the nuclear electric quadrupole moment eQ (a property of the nucleus) with an EFG (a property of the sample) is called the quadrupole interaction.[1–3] Nuclei with quadrupole spins are extensively used to probe static and dynamic phenomena accompanying reversible phase transitions in solids.[4–6] This article focuses on half-integer quadrupole spins ($I = \frac{3}{2}$, $\frac{5}{2}$, $\frac{7}{2}$, and $\frac{9}{2}$) in NMR experiments; that is, the interaction of the spin system with the Zeeman field $\mathbf{B}_0$ (the strong static magnetic field) is much larger than the quadrupole interaction.[7] Under these conditions, we consider the first two expansion terms of this interaction, namely the first- and second-order quadrupole interactions.

There are two ways to investigate the effects of an interaction in NMR:

- Studying the frequency-domain response of the spin system, which deals with transition frequencies between two energy levels and therefore the line positions, line shifts, and line shapes in the spectrum.
- Predicting the time-domain response of the spin system, namely the signal intensities, positions and amplitudes of echoes by using the density operator formalism.

Two-dimensional (2-D) NMR experiments are based on the time-domain response of the spin system to specific sequences, designed in such a way that the desired information is clearly shown in 2-D spectra.[8–10]

The main effects of the quadrupole interaction in the frequency domain are gathered in section 2 but are not developed because excellent review articles are available in the *Encyclopedia of Nuclear Magnetic Resonance*.[11,12] Expressions for the second-order quadrupole shift of the central line under static conditions, and for variable- and magic-angle spinning (MAS) are provided, together with the procedure for simulating powder patterns. As the spectra are the Fourier transforms of time-domain signals, which depend on the experimental parameters such as the strength of $\mathbf{B}_0$, that of the RF field $\mathbf{B}_1$, the pulse durations, the pulse separation, or the relative

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

phase of the RF pulses, understanding of the effects of these parameters enables the undistorted line shapes to be obtained.

Section 3 deals with the time-domain response of the spin system excited by one- and two-pulse sequences that include Solomon echoes[1,13] and Hahn echoes.[14] Section 3.1, which focuses on one-pulse and 2-D nutation experiments, gives the meanings of the elements of the density matrix at the end of an RF pulse. The properties of these elements are crucial for an understanding of echoes in NMR. Nonselective (or hard) pulse excitation, and selective excitation of the spin system are presented to establish the experimental conditions for quantification of the spin population. For this purpose, the nature of a line (only the central line or the full spectrum) must be determined by a one-dimensional (1-D) nutation experiment. Section 3.2 presents a composite-pulse sequence that cancels spurious signals generated by the NMR probe when low-gyromagnetic-ratio nuclei are studied in a Zeeman field $\mathbf{B}_0$. Quantification of the spin population with this composite-pulse sequence is not possible. The numerical procedures for calculating the line intensity from single-pulse and composite-pulse sequences are given.

Section 3.3 presents a method for obtaining the two quadrupole parameters (the quadrupole coupling constant and the asymmetry parameter) from a featureless central-transition line shape by using a two-pulse sequence with a short pulse separation. This method is an extension of the 1-D nutation experiment, which is based on the fact that the variation of the line intensity versus a pulse duration depends on the ratio of the quadrupole coupling w_Q to the strength w_{RF} of the RF field $\mathbf{B}_1$. Furthermore, knowledge of the two quadrupole parameters allows the true isotropic chemical shift of a line to be determined. The latter is related to bond angle in solids.[15]

The broad powder pattern is distorted by the dead time of the receiver following an RF pulse. The usual way to overcome this problem is to apply a spin echo sequence. However, there is a subtle but important difference between Solomon echoes and Hahn echoes. Extensive examples are provided for supporting predictions. For simplicity mainly spins $I = \frac{3}{2}$ and $\frac{5}{2}$ are discussed because they represent the most-studied nuclei. Section 3.4 discusses Solomon echoes with respect to the excitation conditions, namely nonselective (or hard-pulse) excitation, and soft-pulse excitation. These echoes are satellite-transition signals only, because the delay separating the two RF pulses is much shorter than the duration of the free-induction decay (FID) of the central transition.[16] A numerical procedure is proposed to calculate the amplitudes of Solomon echoes, allowing the conditions for their observation to be optimized. Quantification of the spin population is not possible with satellite-transition signals.

For Hahn echoes, which are discussed in section 3.5, the pulse separation is larger than the duration of the central-transition FID; both central-transition and satellite-transition echoes are observed.[17,18] As a result of the spin–spin relaxation, Hahn echo amplitudes are much smaller than Solomon echo amplitudes. The various excitation conditions are also reviewed to deduce the conditions for quantifying the spin population. Hahn echo sequences also cancel spurious signals. A numerical procedure is proposed to predict the echo amplitudes. Section 3.5.7 discusses Hahn echoes in rapidly rotating samples. Finally, a 2-D multiple quantum/magic-angle spinning (MQ/MAS) experiment is reviewed in section 3.5.8. Emphases are on the various conventions for scaling and labeling the high-resolution axis in a 2-D sheared spectrum.

2 FREQUENCY-DOMAIN RESPONSE

2.1 Quadrupole Interaction

In its principal-axis system (PAS), the EFG tensor $\mathbf{V}$ is given by Equation (1):

$$\mathbf{V} = \begin{matrix} V_{XX} & 0 & 0 \\ 0 & V_{YY} & 0 \\ 0 & 0 & V_{ZZ} \end{matrix} \tag{1}$$

As the Laplace equation is valid, $V_{XX} + V_{YY} + V_{ZZ} = 0$, two new parameters are used: the largest component eq (Equation 2) and the asymmetry parameter η (Equation 3),

$$eq = V_{ZZ} \tag{2}$$

$$\eta = \frac{V_{XX} - V_{YY}}{V_{ZZ}} \tag{3}$$

with $0 \leq \eta \leq 1$. If necessary, the axes of the EFG tensor are relabeled to satisfy Equations (2) and (3). The definition of η is not unique.[11]

In NMR, the quadrupole interaction is treated as a weak perturbation of the Zeeman interaction (Equation 4),

$$H_Z = -w_L I_z \tag{4}$$

the coupling of the nuclear spin I with the Zeeman field $\mathbf{B}_0$. The Larmor frequency (Equation 5) is

$$w_L = \gamma \mathbf{B}_0 \tag{5}$$

which is a positive number if the gyromagnetic ratio γ is positive. The standard approach mainly considers the

first two perturbation terms of the quadrupole interaction. These are the first-order quadrupole interaction (Equation 6),

$$H_{\mathrm{Q}}^{(1)} = N_{\mathrm{Q}} \tfrac{1}{6}\sqrt{6}\{3I_{\mathrm{z}}^2 - I(I+1)\}V_{2,0} \tag{6}$$

and the second-order quadrupole interaction (Equations 7 and 8),

$$H_{\mathrm{Q}}^{(2)} = -\frac{N_{\mathrm{Q}}^2}{w_{\mathrm{L}}}\left(\tfrac{1}{2}V_{2,-1}V_{2,1}\{4I(I+1) - 8I_{\mathrm{z}}^2 - 1\} + \tfrac{1}{2}V_{2,-2}V_{2,2}\{2I(I+1) - 2I_{\mathrm{z}}^2 - 1\}\right) I_{\mathrm{z}} \tag{7}$$

$$N_{\mathrm{Q}} = \frac{eq}{2I(2I-1)\hbar} \tag{8}$$

The terms $V_{2,k}$ are the components of the EFG expressed as a spherical tensor of rank 2. The interaction $H_{\mathrm{Q}}^{(1)}$ in Equation (6) is an *even* function of I_{z}. However, in Equation (7), $H_{\mathrm{Q}}^{(2)}$ is an *odd* function of I_{z}. That is, Equation (7) can be rewritten as Equation (9):

$$H_{\mathrm{Q}}^{(2)} = A_{\mathrm{Q}}^{(2)} I_{\mathrm{z}}^3 + B_{\mathrm{Q}}^{(2)} I_{\mathrm{z}} \tag{9}$$

The interaction $H_{\mathrm{Q}}^{(1)}$ is independent of w_{L}, whereas $H_{\mathrm{Q}}^{(2)}$ is inversely proportional to it. This means that the effect of $H_{\mathrm{Q}}^{(2)}$ decreases when the strength of $\mathbf{B}_0$ increases.

Under static conditions $H_{\mathrm{Q}}^{(1)}$ can be expressed as Equations (10–12):

$$H_{\mathrm{Q}}^{(1)} = \tfrac{1}{3}w_{\mathrm{Q}}\{3I_{\mathrm{z}}^2 - I(I+1)\} \tag{10}$$

$$w_{\mathrm{Q}} = \tfrac{3}{4}\Omega_{\mathrm{Q}}(3\cos^2\beta - 1 + \eta\sin^2\beta\cos 2\alpha) \tag{11}$$

$$\Omega_{\mathrm{Q}} = eQN_{\mathrm{Q}} = \frac{e^2qQ}{2I(2I-1)\hbar} \tag{12}$$

The definition of the quadrupole coupling w_{Q} in Equation (10) is not unique. It is half that used by Abragam.[1] Fortunately, the quadrupole coupling constant $e^2qQ/\hbar$ is universally accepted. The angles α and β are the first two Euler angles of $\mathbf{B}_0$ in the PAS of the EFG tensor. However, β and α are the polar angles of $\mathbf{B}_0$ in this system. The definition of the Euler angles is also not unique.[11] Various conventions about signs of phases and frequencies have appeared in the NMR literature.[19]

In a powder sample consisting of many small single crystals oriented randomly with respect to $\mathbf{B}_0$, the values of α and $\cos\beta$ are distributed with equal probability. The factor $\cos\beta$ is involved instead of β because the elementary solid angle is $\sin\beta\,\mathrm{d}\beta\,\mathrm{d}\alpha = \mathrm{d}(\cos\beta)\,\mathrm{d}\alpha$. The histogram $P(w_{\mathrm{Q}})$ is obtained by counting the frequency lying between w_{Q} and $w_{\mathrm{Q}} + \Delta w_{\mathrm{Q}}$ for any combination of α and $\cos\beta$.

Combining Equations (11) and (12) gives Equation (13):

$$y = w_{\mathrm{Q}}\frac{8I(2I-1)\hbar}{3e^2qQ} = 3\cos^2\beta - 1 + \eta\sin^2\beta\cos 2\alpha \tag{13}$$

which varies between $-(1+\eta)$ and 2, and reaches $\{-2, 2\}$ when $\eta = 1$. The interval $\{-2, 2\}$ is divided into n parts. For a given η, the values of y are computed for every combination of $\alpha = 2u\pi/M$ and $\cos\beta = r/M$($u, r = 0, \ldots, M$). The y values falling into the n intervals are counted. We chose $n = 41$ and $M = 300$. This procedure was proposed by Narita and colleagues.[20] It allows us to simulate the powder pattern (i.e. the equivalents of the spectrum of a powder sample) $P(y)$ of a satellite transition. Therefore, comparing the shape and the spreading of the experimental powder pattern with the simulated one provides the asymmetry parameter and the quadrupole coupling constant, respectively.

The spectral line intensity[21] and the echo amplitude (which are dealt with in section 3) do not depend on the sign of w_{Q} or y. So $P(y < 0)$ is co-added to $P(y > 0)$, reducing the numbers of values of $P(y)$ to $\frac{1}{2}(1+n)$. Other procedures giving the same result but with less computation time are available.[22,23]

2.2 Energy Levels and Spectrum

Figure 1 shows the splitting and shift of the energy levels, as well as those of the absorption lines (simply lines in the remainder of this article) in the spectra for a spin $I = \frac{3}{2}$ system in a single crystal, when considering $H_{\mathrm{Q}}^{(1)}$ and $H_{\mathrm{Q}}^{(2)}$.

Without $\mathbf{B}_0$, the four energy levels of a free spin-I system are degenerate (state a). When the spin system is introduced into $\mathbf{B}_0$, the Zeeman interaction splits these energy levels equally (state b) with $|I, m\rangle = |\frac{3}{2}, \frac{3}{2}\rangle, |\frac{3}{2}, \frac{1}{2}\rangle, |\frac{3}{2}, -\frac{1}{2}\rangle$, and $|\frac{3}{2}, -\frac{3}{2}\rangle$, or simply $|m\rangle = |\frac{3}{2}\rangle, |\frac{1}{2}\rangle, |-\frac{1}{2}\rangle$, and $|-\frac{3}{2}\rangle$. The difference between two consecutive energy levels defines w_{L}, giving a single line in the spectrum, shown as b′.

The first-order quadrupole interaction generates small shifts of the energy levels. As $H_{\mathrm{Q}}^{(1)}$ is an even function of I_{z}, the energy-level shifts are an even function of the magnetic number m. Therefore two energy levels having the same $|m|$ are shifted by the same quantity: $\langle m|H_{\mathrm{Q}}^{(1)}|m\rangle = \pm w_{\mathrm{Q}}$ for a spin $I = \frac{3}{2}$ (state c), depending on $|m|$. The transition $(-\frac{1}{2} \leftrightarrow \frac{1}{2})$ is called the central transition; the other two transitions $(-\frac{3}{2} \leftrightarrow -\frac{1}{2})$ and $(\frac{1}{2} \leftrightarrow \frac{3}{2})$ are called the satellite transitions. As the two energy levels of the central transition are shifted by the same quantity, the line in the spectrum, called the central line, remains at w_{L}. However, the satellite lines are shifted by $\pm 2w_{\mathrm{Q}}$ from the central line (shown as c′).

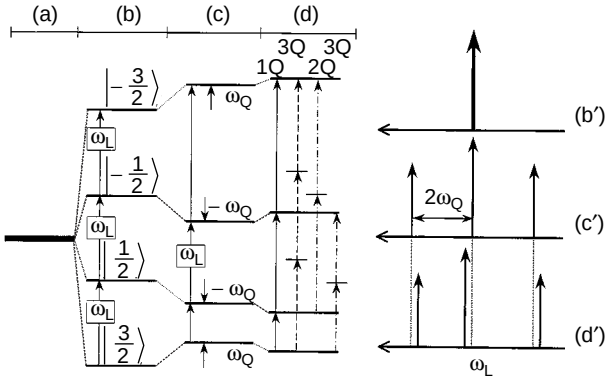

Figure 1 The four energy levels $|\frac{3}{2}\rangle$, $|\frac{1}{2}\rangle$, $|-\frac{1}{2}\rangle$, and $|-\frac{3}{2}\rangle$ for a spin $I = \frac{3}{2}$ system in a single crystal in (state a) zero Zeeman field $\mathbf{B}_0$, (b) in the presence of $\mathbf{B}_0$, (c) including the first-order quadrupole interaction $H_Q^{(1)}$, and (d) including the second-order quadrupole interaction $H_Q^{(2)}$. Their corresponding spectra are represented as arrows in (b′), (c′), and (d′), respectively. The figure provides definitions of the quadrupole coupling w_Q (c′) and of the Larmor frequency w_L (b). The transitions (d) between two energy levels occur during excitation of the spin system by an RF pulse: one-quantum (1Q) transitions are represented by solid arrows, 2Q transitions by dot-dashed arrows, and 3Q transitions by dashed arrows; for clarity, −1Q, −2Q, and −3Q transitions are not shown.

The effect of $H_Q^{(2)}$ is to shift the energy levels further (state d). As a result, an additional shift with respect to w_L, called the second-order quadrupole shift, occurs for each absorption line. However, this second-order quadrupole shift of a line is much smaller than the $\pm 2w_Q$ shift due to $H_Q^{(1)}$. As $H_Q^{(2)}$ is an odd function of I_z, the energy-level shifts are an odd function of m. Therefore, two energy levels having the same $|m|$ are shifted by the same quantity but in the opposite direction. Even the central line is shifted from w_L (shown as d′). Therefore the observed position of the central line is not w_L. The direction of the line shift depends on the orientation of $\mathbf{B}_0$ in the PAS of the EFG tensor.

To detect the presence of these split energy levels, a time-dependent perturbation has to be applied.[24] Continuous wave (CW) NMR applies an RF magnetic field $\mathbf{B}_1$ of *very weak* amplitude $w_{RF}(=\gamma B_1)$, whose Hamiltonian is (Equation 14)

$$H_{\text{pert}} \propto -w_{RF} I_x \cos w_c t \tag{14}$$

where w_c is the carrier or the spectrometer frequency. The RF field $\mathbf{B}_1$ is applied perpendicularly to $\mathbf{B}_0$. Therefore, the sample is surrounded with a bath of RF photons of energy w_c.[25] Excitation of the spin system and detection of its response occur simultaneously by using two RF coils. During the excitation of the spin system by $\mathbf{B}_1$, single quantum (SQ) transitions between two consecutive energy levels as well as MQ transitions between two nonconsecutive energy levels occur.[26] The probability per second that H_{pert} induces a transition between two states $|m\rangle$ and $|m'\rangle$ is proportional to $|\langle m'|I_x|m\rangle|^2$, which vanishes unless $m' = m \pm 1$. The condition $\Delta m = \pm 1$ is the selection rule for magnetic dipole transitions. Transverse magnetizations, detected by an RF coil, are related to SQ transitions. The expression $\zeta^2(m+1, m)$ is used to denote this probability (Equation 15):

$$\begin{aligned}\zeta^2(m+1, m) &\equiv |\langle m+1|I_x|m\rangle|^2 = |\langle m|I_x|m+1\rangle|^2 \\ &= I(I+1) - m(m+1)\end{aligned} \tag{15}$$

In the linear regime where the response of the spin system is proportional to the excitation, the spectral line intensity (the area of the absorption line) is proportional to ζ^2. Table 1 gives the values of ζ^2 as well as those of the relative spectral line intensity

$$\frac{\zeta^2}{\sum_{m=-I}^{I-1} \zeta^2}$$

for various half-integer spins. The effects of MQ transitions are observed in the nonlinear regime by using a stronger $\mathbf{B}_1$, so that the spin system is forced to absorb several RF photons simultaneously.[9,27]

In contrast, in pulsed NMR the process of these transitions is investigated only after the end of a strong

Table 1 Relative spectral line intensities ζ^2 ζ^2 of all the SQ transitions ($m \leftrightarrow m+1$) of half-integer spins I, with $\zeta^2(m+1, m) = I(I+1) - m(m+1)$

Spins I	Transitions $m \leftrightarrow m+1$	ζ^2	ζ^2 ζ^2
$\frac{1}{2}$	$-\frac{1}{2} \leftrightarrow \frac{1}{2}$	1	1
$\frac{3}{2}$	$\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$	3	3/10
	$-\frac{1}{2} \leftrightarrow \frac{1}{2}$	4	4/10
$\frac{5}{2}$	$\pm\frac{3}{2} \leftrightarrow \pm\frac{5}{2}$	5	5/35
	$\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$	8	8/35
	$-\frac{1}{2} \leftrightarrow \frac{1}{2}$	9	9/35
$\frac{7}{2}$	$\pm\frac{5}{2} \leftrightarrow \pm\frac{7}{2}$	7	7/84
	$\pm\frac{3}{2} \leftrightarrow \pm\frac{5}{2}$	12	12/84
	$\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$	15	15/84
	$-\frac{1}{2} \leftrightarrow \frac{1}{2}$	16	16/84
$\frac{9}{2}$	$\pm\frac{7}{2} \leftrightarrow \pm\frac{9}{2}$	9	9/165
	$\pm\frac{5}{2} \leftrightarrow \pm\frac{7}{2}$	16	16/165
	$\pm\frac{3}{2} \leftrightarrow \pm\frac{5}{2}$	21	21/165
	$\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$	24	24/165
	$-\frac{1}{2} \leftrightarrow \frac{1}{2}$	25	25/165

RF pulse by detecting a FID. Excitation and detection are performed with the same RF coil. Also, SQ and MQ transitions occur during the RF pulse, as in CW NMR. As an RF coil can only detect an FID generated by SQ transitions, a second RF pulse is required to probe the effects of MQ transitions.

When MQ transitions between $|m'\rangle$ and $|m\rangle$ are detected directly as in CW NMR, the absorption line would be located at about $(m' - m)w_L$ as the energy-level diagram suggests (state d in Figure 1). In fact this is not the case – the transition between two nonconsecutive energy levels is a multiple-photon process, as shown for a spin $I = \frac{3}{2}$. If $H_Q^{(1)}$ is the dominant interaction (state c), a 2Q transition is a two-photon process and the frequency of the photon is $w_L \pm w_Q$, depending on the transitions.[27,28] The 3Q transition is a three-photon process. The photon frequency of the central and 3Q transitions is w_L.[27] If $H_Q^{(2)}$ is present, their photon frequency is close to w_L. From a theoretical point of view, to avoid the effects of the Zeeman interaction on the transition frequency $w_{w,q}$ between two energy levels $|w\rangle$ and $|q\rangle$, we define $w_{w,q}$ in the rotating frame of the frequency carrier or, in short, the rotating frame. For the moment, we do not deal with the offset term $H_\Delta = -(w_L - w_c)I_z$. In other words, the Zeeman interaction is not considered in the definition of $w_{w,q}$ (Equation 16):

$$w_{w,q} = \langle w|(H_Q^{(1)} + H_Q^{(2)})|w\rangle - \langle q|(H_Q^{(1)} + H_Q^{(2)})|q\rangle = w_{w,q}^{(1)} + w_{w,q}^{(2)} \quad (16)$$

where $w_{w,q}^{(1)}$ and $w_{w,q}^{(2)}$ are, respectively, the first- and second-order quadrupole shifts of the $(w - q)$ quantum absorption line with respect to w_L.

In the simplest pulsed NMR experiment that uses a single RF pulse (Figure 2a), two durations are involved – the duration of the RF pulse t_1 and that of the signal acquisition τ_2. Furthermore we define the duration of the FID by T_{FID}. An FID in the time domain $F(t_1, \tau_2)$ and its spectrum in the frequency domain $S(t_1, w_2)$ are related by the Fourier transform of Equation (17):

$$F(t_1, \tau_2) = \int_{-\infty}^{+\infty} S(t_1, w_2) \exp(-i\tau_2 w_2)\, dw_2 \quad (17)$$

The amplitude of the FID at the end of the RF pulse is given by Equation (18),

$$F(t_1, \tau_2 = 0) = \int_{-\infty}^{+\infty} S(t_1, w_2)\, dw_2 \quad (18)$$

That is, the FID amplitude at the end of the RF pulse is related to the spectral line intensity. In the rotating frame (Figure 2b), the effect of an X pulse is the nutation of the magnetization vector **M** around $\mathbf{B}_1$ located along the x axis. The projection of **M** or the FID is detected along the y axis. However, the FID amplitude, also denoted by $\langle I_y(t_1)\rangle$, is related to the density operator $\rho(t_1)$ as Equation (19),

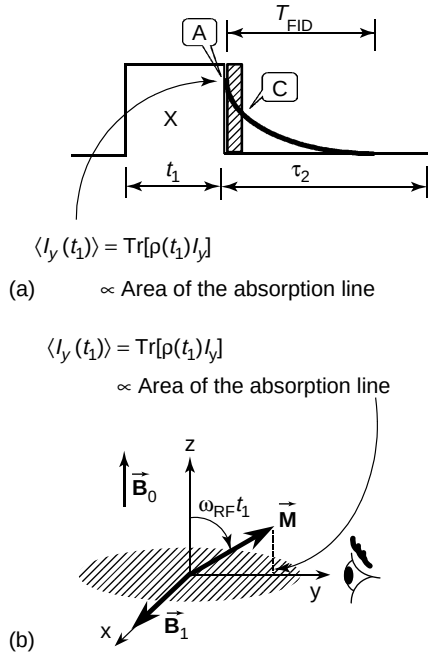

Figure 2 Relationship between the density operator $\rho(t_1)$, the spin operator I_y, and the area of an absorption line (spectral line intensity) or the amplitude of an FID $\langle I_y(t_1)\rangle$. (a) One-pulse experiment: the FID amplitude (point A) following the RF pulse is lost in the dead-time (shaded area) of the receiver; the acquisition of the FID starts from point C; t_1 is the pulse duration, τ_2 the acquisition period, and T_{FID} the duration of the FID. (b) Nutation of the magnetization vector **M** around the RF magnetic field $\mathbf{B}_1$ (along the x axis) in the rotating frame of the frequency carrier during the excitation of the spin system by an X pulse; $w_{RF}t_1$ is the pulse flip angle. The receiver, represented by an eye, detects the projection of **M** along the y axis. $\mathbf{B}_0$ is along the z axis.

$$\langle I_y(t_1)\rangle = F(t_1, \tau_2 = 0) = \mathrm{Tr}\{\rho(t_1)I_y\} \quad (19)$$

where Tr means trace. In practice, the FID is recorded with a quadrature detector. The complex spectral line intensity and the density operator are related by Equation (20):

$$\langle I_+(t_1)\rangle = \langle I_x(t_1) + iI_y(t_1)\rangle = \mathrm{Tr}\{\rho(t_1)I_+\} \quad (20)$$

Therefore, knowledge of the density operator allows us to predict the complex spectral line intensity and the dynamics of the spin system as usual (see section 3).

2.2.1 *Effect of the First-order Quadrupole Interaction on the Spectrum*

Two examples are used to illustrate the effects of $H_Q^{(1)}$ on the spectrum. The spectrum of ^{27}Al ($I = \frac{5}{2}$) from a single

crystal of α-Al_2O_3 consists of five lines (Figure 3a); the central line is in the middle of four satellite lines,[29] which means that the dominant interaction is $H_Q^{(1)}$. Figure 3(b) is the spectrum of two single crystals. As the central lines of the two crystals remain at w_L, they co-add, giving an intense line. In contrast, the two sets of four satellite lines are separated due to different orientations of $\mathbf{B}_0$ in the PAS of the EFG tensor for each crystal.

When the sample is a powder, there is a uniform distribution of the orientations of $\mathbf{B}_0$ in each crystallite.[30] If $H_Q^{(1)}$ is the dominant interaction, the contributions of all crystallites co-add, giving a sharp central line and a broad powder pattern of satellite lines. Figure 4 (spectrum a) shows the case of ^{23}Na ($I = \frac{3}{2}$) in a powder of $NaNO_3$.[31] The powder pattern can be simulated by using the numerical procedure described in section 2.1.

In a MAS experiment, the powder is packed in a rotor whose spinning axis relative to $\mathbf{B}_0$ is at the magic angle $\theta_m = 54°44'$. In the fast condition, that is if the rotor spinning rate is much larger than the line width, $H_Q^{(1)}$ will be canceled during free precession of the spin system.[18] Otherwise rotational echoes appear in the FID and spinning sidebands in the spectrum. Spectrum (c) in Figure 4 presents the ^{23}Na ($I = \frac{3}{2}$) MAS spectrum in powdered $NaNO_3$. The shape and the spread of these spinning sidebands enable determination of the asymmetry parameter η and the quadrupole coupling constant, respectively.

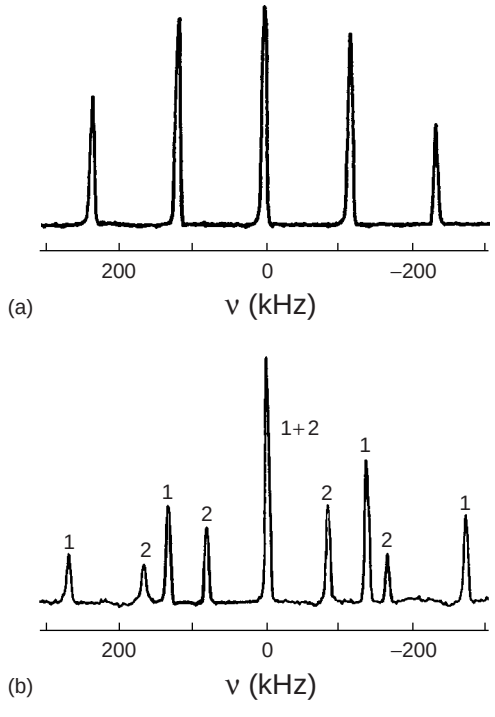

Figure 3 Effect of $H_Q^{(1)}$ on the one-pulse spectrum of ^{27}Al ($I = \frac{5}{2}$) in (a) a single crystal and (b) two crystals (denoted by 1 and 2) of α-Al_2O_3: the central lines co-add; the satellite lines are resolved.

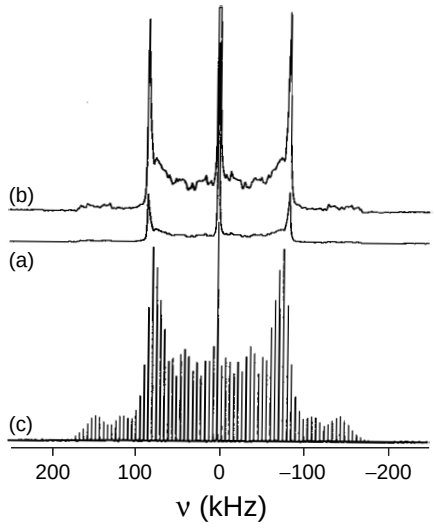

Figure 4 Effect of $H_Q^{(1)}$ on the spectrum of spin $I = \frac{3}{2}$ in a powdered sample. ^{23}Na ($I = \frac{3}{2}$) spectra of $NaNO_3$ recorded by using (a) static and (c) MAS conditions. The central line is cut off at (b) $\frac{1}{4}$ and (c) $\frac{1}{16}$ of its total height. (Reproduced by permission of Academic Press from J. Skibsted, N.C. Nielsen, H. Bildsøe, H.J. Jakobsen, *J. Magn. Reson.*, **95**, 88–117 (1991).)

2.2.2 *Effect of the Second-order Quadrupole Interaction on the Spectrum*

When $H_Q^{(2)}$ becomes important, the central line is also shifted. For a powder, the central line is also broadened but to a lesser extent than the satellite lines, which spread over the megahertz range. As a result the powder pattern of the central line is generally observed, whereas that of the satellite lines is not. The shape of the central-transition powder pattern depends on the asymmetry parameter η, as shown by the spectra for three ^{139}La ($I = \frac{7}{2}$) salts in Figure 5(a–c).[32] Under static conditions, the second-order quadrupole shift of the central line with respect to w_L is given by Equations (21–24):[11]

$$w_{-1/2,1/2}^{(2)\text{static}} = -\frac{1}{6w_L}\left(\frac{3e^2qQ}{2I(2I-1)\hbar}\right)^2 \left\{I(I+1) - \frac{3}{4}\right\} \times \{A(\alpha,\eta)\cos^4\beta + B(\alpha,\eta)\cos^2\beta + C(\alpha,\eta)\} \quad (21)$$

$$A(\alpha,\eta) = -\tfrac{27}{8} + \tfrac{9}{4}\eta\cos 2\alpha - \tfrac{3}{8}(\eta\cos 2\alpha)^2 \quad (22)$$

$$B(\alpha,\eta) = \tfrac{30}{8} - \tfrac{1}{2}\eta^2 - 2\eta\cos 2\alpha + \tfrac{3}{4}(\eta\cos 2\alpha)^2 \quad (23)$$

$$C(\alpha,\eta) = -\tfrac{3}{8} + \tfrac{1}{3}\eta^2 - \tfrac{1}{4}\eta\cos 2\alpha - \tfrac{3}{8}(\eta\cos 2\alpha)^2 \quad (24)$$

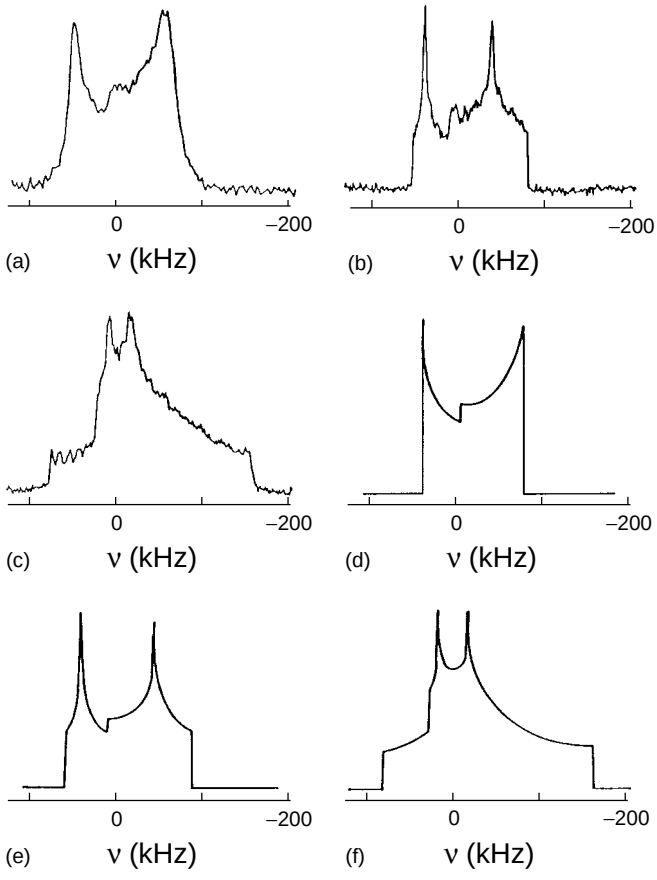

Figure 5 Effect of $H_Q^{(2)}$ on the static central line of a powder sample. ^{139}La ($I = \frac{7}{2}$) spectra of (a) $La(OH)_3$ with $\eta = 0.05$, (b) $LaCl_3$ ca. $6H_2O$ with $\eta = 0.40$, and (c) $La(NO_3)_3 \cdot 6H_2O$ with $\eta = 0.80$. The corresponding simulated spectra are shown in (d), (e), and (f), respectively. These spectra are the Fourier transform of half of the Hahn echo. (Reproduced by permission of The Royal Society of Chemistry from B. Herreros, P.P. Man, J.-M. Manoli, J. Fraissard, *J. Chem. Soc., Chem. Commun.*, 464–466 (1992).)

This shift is inversely proportional to w_L (see Equation 21). The central-transition powder pattern can be simulated by the angular part of Equations (21–24) using the procedure described in section 2.1. The limits are $\{-\frac{2}{3}(1+\eta), \frac{1}{24}(3+\eta)^2\}$ which become $\{-\frac{4}{3}, \frac{2}{3}\}$ when $\eta = 1$. Figure 5(d–f) are the simulated spectra.

Figure 6 shows the powder patterns of ^{59}Co ($I = \frac{7}{2}$) in $Na_3Co(NO_2)_6$ whose line width decreases when the strength of $\mathbf{B}_0$ increases.[33] However, above 11 T the dominant interaction becomes the chemical shift anisotropy whose contribution increases with the strength of $\mathbf{B}_0$.

When the rotor spinning axis is at the angle $\theta_r (\neq \theta_m)$ relative to $\mathbf{B}_0$, the experiment is called variable-angle spinning (VAS).[34] In the fast condition, the second-order quadrupole shift of the central line with respect to w_L is

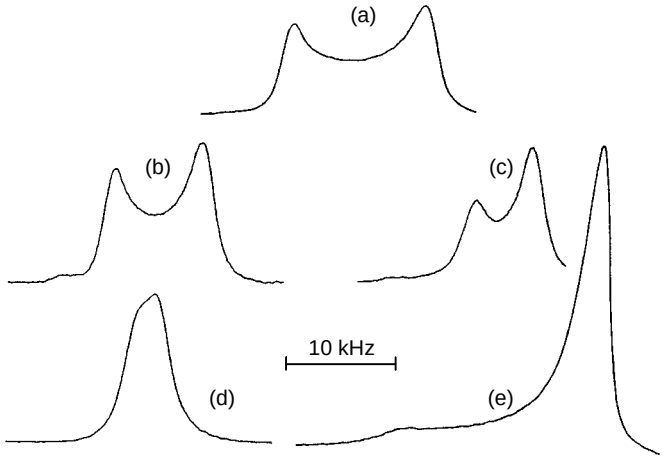

Figure 6 Effect of the strength of $\mathbf{B}_0$ on the central-transition powder pattern characterized by $H_Q^{(2)}$ on ^{59}Co ($I = \frac{7}{2}$) spectra of polycrystalline sodium hexanitrocobaltate(III) $Na_3Co(NO_2)_6$ in different $\mathbf{B}_0$: (a) 4.7, (b) 5.9, (c) 7.0, (d) 9.4, and (e) 11.7 T. (Reproduced by permission of The National Research Council of Canada from D.R. Eaton, R.J. Buist, B.G. Sayer, *Can. J. Chem.*, **65**, 1332–1335 (1987).)

defined by Equations (25–28):[18]

$$w_{-1/2,1/2}^{(2)\text{fast VAS}} = \frac{\Omega_Q^2}{w_L}\{I(I+1) - \tfrac{3}{4}\} \times \tfrac{1}{2}B_{0,0}(\eta) + 4\{B_{2,0}(\eta)d_{0,0}^{(2)}(\beta_1) + 2B_{2,2}(\eta) \times d_{2,0}^{(2)}(\beta_1)\cos 2\alpha_1\}P_2(\cos\theta_r) + 9\{B_{4,0}(\eta) \times d_{0,0}^{(4)}(\beta_1) + 2B_{4,2}(\eta)d_{2,0}^{(4)}(\beta_1)\cos 2\alpha_1 + 2B_{4,4}(\eta)d_{4,0}^{(4)}(\beta_1)\cos 4\alpha_1\}P_4(\cos\theta_r) \quad (25)$$

$$B_{0,0}(\eta) = -\tfrac{3}{5}\left(\tfrac{1}{3}\eta^2 + 1\right), \quad B_{2,0}(\eta) = \tfrac{3}{14}\left(\tfrac{1}{3}\eta^2 - 1\right) \quad (26)$$

$$B_{2,\pm2}(\eta) = \frac{3}{7\sqrt{6}}\eta, \quad B_{4,0}(\eta) = \tfrac{9}{70}\left(\tfrac{1}{18}\eta^2 + 1\right) \quad (27)$$

$$B_{4,\pm2}(\eta) = \frac{3}{14\sqrt{10}}\eta, \quad B_{4,\pm4}(\eta) = \frac{1}{4\sqrt{70}}\eta^2 \quad (28)$$

where $P_2(\cos\theta_r)$ and $P_4(\cos\theta_r)$ are Legendre polynomials (Equations 29 and 30):

$$P_2(\cos\theta_r) = \tfrac{1}{2}(3\cos^2\theta_r - 1) \quad (29)$$

$$P_4(\cos\theta_r) = \tfrac{1}{8}(35\cos^4\theta_r - 30\cos^2\theta_r + 3) \quad (30)$$

The Euler angles α_1 and β_1 are those of the rotor in the PAS of the EFG tensor. In Equation (25), we can replace Ω_Q^2/w_L by Equation (31).

$$\frac{\Omega_Q^2}{w_L} = \frac{1}{6w_L}\left(\frac{3e^2qQ}{2I(2I-1)\hbar}\right)^2\frac{2}{3} \quad (31)$$

The VAS central-transition powder pattern can be simulated by using Equations (25–31) and the procedure described in section 2.1 with the same limits as in static experiment, that is, $\{-\frac{4}{3}, \frac{2}{3}\}$. The VAS central-transition powder pattern for $\theta_r = 0$ is identical to the static spectrum.

For MAS experiments in the fast condition, we only need to specify $\theta_r = \theta_m$, that is, $P_2(\cos\theta_r) = 0$ in Equation (25) to obtain the expression of the second-order quadrupole shift of the central line with respect to w_L, which is equivalent to the classical relationship (Equations 32–35).

$$w_{-1/2,1/2}^{(2)\text{fast MAS}} = -\frac{1}{6w_L}\left(\frac{3e^2qQ}{2I(2I-1)\hbar}\right)^2 \left\{I(I+1) - \tfrac{3}{4}\right\} \times \{D(\alpha_1,\eta)\cos^4\beta_1 + E(\alpha_1,\eta)\cos^2\beta_1 + F(\alpha_1,\eta)\} \quad (32)$$

$$D(\alpha_1,\eta) = \tfrac{21}{16} - \tfrac{7}{8}\eta\cos 2\alpha_1 + \tfrac{7}{48}(\eta\cos 2\alpha_1)^2 \quad (33)$$

$$E(\alpha_1,\eta) = -\tfrac{9}{8} + \tfrac{1}{12}\eta^2 + \eta\cos 2\alpha_1 - \tfrac{7}{24}(\eta\cos 2\alpha_1)^2 \quad (34)$$

$$F(\alpha_1,\eta) = \tfrac{5}{16} - \tfrac{1}{8}\eta\cos 2\alpha_1 + \tfrac{7}{48}(\eta\cos 2\alpha_1)^2 \quad (35)$$

In rapid conditions, the MAS central-transition powder pattern can be simulated by using the angular part of Equations (32–35), and the procedure described in section 2.1 with the limits $\{-\frac{1}{14}(1-\eta)^2, -\frac{1}{2}(1+\frac{1}{6}\eta^2)\}$. In practice, the same limits are used as in static conditions in order to compare the line widths. The central-transition powder pattern is narrowed by a factor ranging from three to four with respect to that obtained in static condition, depending on the value of η.[11]

For all experiments (static or rotating sample), the second-order quadrupole shift of the center of gravity of the central-transition powder pattern due to $H_Q^{(2)}$ (Equation 36) is[18]

$$w_{-1/2,1/2}^{(2)\text{iso}} = -\frac{3\Omega_Q^2}{10w_L}\left\{I(I+1) - \tfrac{3}{4}\right\}\left(1 + \tfrac{1}{3}\eta^2\right) \quad (36)$$

with respect to w_L. Experimentally, the center of gravity is the position that divides the integrated area of the powder pattern into two equal parts. The observed chemical shift of the center of gravity of the central-transition powder pattern $\delta_{G2}^{(obs)}$ comes from two contributions,[18] namely the isotropic chemical shift $\delta_{CS}^{(iso)}$ and $w_{-1/2,1/2}^{(2)iso}$ such that (Equation 37)

$$\delta_{G2}^{(obs)} = \delta_{CS}^{(iso)} + \frac{1}{w_c} w_{-1/2,1/2}^{(2)iso} \quad (37)$$

As $w_{-1/2,1/2}^{(2)iso} \leq 0$, the result $\delta_{CS}^{(iso)} \geq \delta_{G2}^{(obs)}$ is always true.

Figure 7 shows the MAS spectra of four- and six-coordinated ^{27}Al ($I = \frac{5}{2}$) in $Y_3Al_5O_{12}$.[35] The center of gravity of the six-coordinated Al powder pattern, whose spectrum is typical of $H_Q^{(1)}$, remains at its isotropic chemical shift position whatever the strength of $\mathbf{B}_0$. In contrast, that of the four-coordinated Al powder pattern shifts towards its isotropic chemical shift when the strength of $\mathbf{B}_0$ increases because the dominant interaction is $H_Q^{(2)}$.

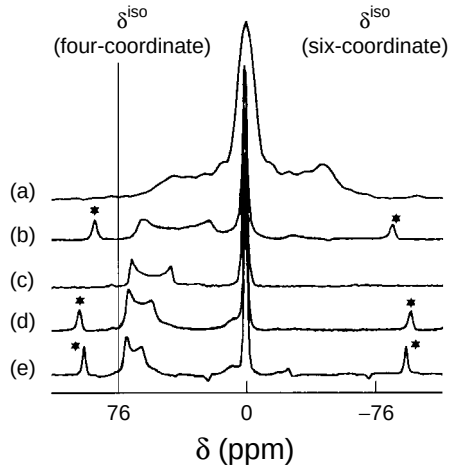

Figure 7 Effect of the strength of $\mathbf{B}_0$ on the shift and the shape of the central line characterized by $H_Q^{(1)}$ and $H_Q^{(2)}$. Spectra of four- and six-coordinate ^{27}Al ($I = \frac{5}{2}$) in $Y_3Al_5O_{12}$ crystalline powder acquired with varying $\mathbf{B}_0$: (a) 4.7, (b) 7.0, (c) 9.4, (d) 11.7, and (e) 14.1 T. Asterisked lines are spinning sidebands. The isotropic chemical shifts δ^{iso} of the four- and six-coordinate species are 76 and 0 ppm, respectively. (Reproduced by permission of Gordon and Breach Publishers, Harwood Academic, Yverdon Switzerland, from D. Massiot, B. Cote, F. Taulelle, J.-P. Coutures, *Application of NMR Spectroscopy to Cement Science*, eds. P. Colombet, A.R. Grimmer, 153–169 (1994).)

2.2.3 *Procedures used to Obtain the Spectrum*

Depending on the quadrupole coupling constant and the spin, the powder pattern is obtained by one of four procedures:

1. the Fourier transform of the FID following a single RF pulse such as those of ^{27}Al ($I = \frac{5}{2}$) in $Y_3Al_5O_{12}$ (Figure 7);
2. the Fourier transform of the second half of the echo obtained with the two-pulse sequences such as those of ^{139}La ($I = \frac{7}{2}$) in lanthanum salts (Figure 5);
3. the magnitude representation of the Fourier transform of the full echo;[36]
4. the summation of all the spectra obtained as in (2) by incrementing the carrier frequency by steps of 100 kHz, an example being the spectrum of ^{139}La ($I = \frac{7}{2}$) in $LaCoO_3$ (Figure 8).[37]

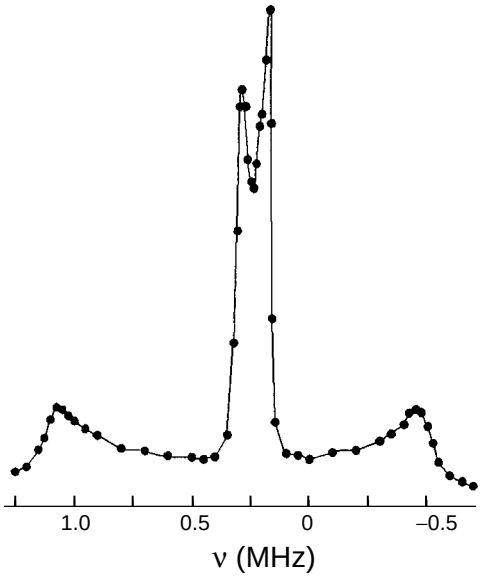

Figure 8 Use of the Hahn echo sequence for acquiring a very broad powder pattern (>200 kHz): ^{139}La ($I = \frac{7}{2}$) powder pattern of $LaCoO_3$ showing the central ($-\frac{1}{2} \leftrightarrow \frac{1}{2}$) transition (located between 0.5 and 0 MHz) and the ($\frac{1}{2} \leftrightarrow \frac{3}{2}$) and ($-\frac{3}{2} \leftrightarrow -\frac{1}{2}$) satellite transitions (located between 1.25 and −0.75 MHz). Each filled circle of the powder pattern is acquired with the Hahn echo sequence. (Reprinted from T.J. Bastow, '^{139}La NMR Characterization of La_2O_3 and La_{1-x} Sr_xMO_3 where M = Cr, Mn or Co', *Solid State NMR*, **3**, 17–22 Copyright (1994) with permission from Elsevier Science.)

3 TIME-DOMAIN RESPONSE

In limiting the discussion topics to the effects of interactions on the spectrum, the internal interactions H_{int} are restricted to secular Hamiltonians, that is, to those that commute with the Zeeman interaction. An example is the isotropic chemical shift shown in Equation (38):

$$H_{\text{CS}} = -w_{\text{c}}\delta_{\text{CS}}^{(\text{iso})}I_{\text{z}} \tag{38}$$

As the pulse duration is short (some microseconds), the rotor containing the sample appears to be static during excitation of the spin system by the RF pulses in MAS experiments. As a result, the internal interactions remain time independent. Only $H_{\text{Q}}^{(1)}$ is considered because this interaction can be much stronger than the amplitude w_{RF} of the RF pulse; thus the offset or shift interactions and $H_{\text{Q}}^{(2)}$ can be neglected. Homonuclear $H_{\text{D(II)}}$ and heteronuclear $H_{\text{D(IS)}}$ magnetic dipole–dipole interactions are also neglected during strong RF pulses. The assumptions during the RF pulses are the same for both static and MAS conditions, and both one- and two-pulse sequences (Figures 2 and 9).

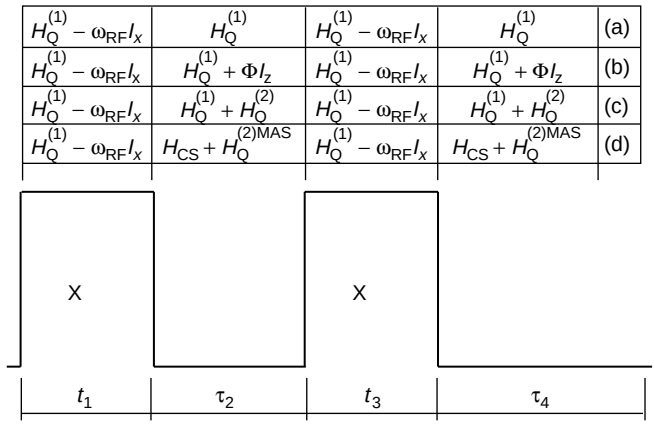

Figure 9 Interactions involved in the four periods of the two-pulse sequences discussed in this paper: the RF pulse of phase $X(-w_{\text{RF}}I_x)$, the first-order quadrupole interaction in static condition $H_{\text{Q}}^{(1)}$; the second-order quadrupole interaction in static condition $H_{\text{Q}}^{(2)}$ or in fast MAS condition $H_{\text{Q}}^{(2)\text{MAS}} \equiv H_{\text{Q}}^{(2)\text{fast MAS}}$; the isotropic chemical shift H_{CS}; and the secular part of the heteronuclear magnetic dipole-dipole interaction ΦI_z. (a) Two-pulse sequence with a short delay τ_2 or Solomon echo sequence in static condition; (b) and (c) Hahn echo sequence under static conditions; (d) Hahn echo sequence under rapid MAS conditions. The term t_1 is the first-pulse duration and t_3 is the second-pulse duration; τ_2 is the pulse separation or the experimental evolution period and τ_4 describes the acquisition period.

During free precession of the spin system, that is in the absence of RF pulses, interactions neglected during the RF pulses cannot be ignored. The simplification depends on the sample, the nuclei, and the experimental conditions (static or MAS conditions, and the strength of $\mathbf{B}_0$). The second-order quadrupole interaction can be neglected with strong $\mathbf{B}_0$ or with light nuclei. In MAS experiments, the chemical shift anisotropy, $H_{\text{Q}}^{(1)}$, as well as $H_{\text{D(IS)}}$, are canceled under rapid conditions, whereas $H_{\text{D(II)}}$ is dramatically reduced.

Throughout this article Hamiltonians H, the carrier frequency w_{c}, the Larmor frequency w_{L}, the RF pulse amplitude w_{RF}, and the line positions have been defined in angular frequency units. Only the frequency offsets are expressed in frequency units. Disregarding relaxation phenomena, the dynamics of a spin-I system submitted to two-pulse sequences is described by the density operator $\rho(t_1, \tau_2, t_3, \tau_4)$ expressed in the rotating frame. The duration t_1 is that of the first RF pulse, τ_2 is the pulse separation or experimental evolution period, t_3 is the second-pulse duration, and τ_4 the acquisition period (Figure 9). The Hamiltonian of an X pulse is given by Equation (39):

$$H_{\text{RF}} = -w_{\text{RF}}I_x \tag{39}$$

The spin dynamics from the Boltzmann equilibrium to the acquisition period are followed by using the density operators. In the high-temperature approximation, the initial state is described by the Boltzmann density operator $\rho(0) = I_z$. At the end of the first RF pulse the density operator is defined by Equation (40):

$$\rho(t_1) = \exp\{-\mathrm{i}(H_{\mathrm{int}} + H_{\mathrm{RF}})t_1\}\rho(0)\exp\{\mathrm{i}(H_{\mathrm{int}} + H_{\mathrm{RF}})t_1\} \tag{40}$$

Although the matrix representation of the secular Hamiltonian H_{int}, expressed in the eigenstates $|m\rangle$ of the Zeeman interaction H_z, is diagonal, this is not the case for H_{RF}. The matrix representation of $\exp\{-\mathrm{i}(H_{\mathrm{int}} + H_{\mathrm{RF}})t_1\}$ is the matrix whose elements are the exponential of each matrix element of $(H_{\mathrm{int}} + H_{\mathrm{RF}})$ if the latter is diagonal. Therefore, the first step is to diagonalize the matrix of $(H_{\mathrm{int}} + H_{\mathrm{RF}})$. For simplicity the matrix is denoted by the same symbol as the Hamiltonian. The eigenvalue matrix Ω, the eigenvector matrix T, and $(H_{\mathrm{int}} + H_{\mathrm{RF}})$ are related (Equation 41) as

$$\Omega = T(H_{\mathrm{int}} + H_{\mathrm{RF}})T^{\dagger} \tag{41}$$

where $T^{\dagger}$ is the complex conjugate of the transposed matrix of T. The calculation of Ω and T has been done analytically (Equation 42) for

$$H = H_{\mathrm{Q}}^{(1)} + H_{\mathrm{RF}} \tag{42}$$

in three cases: $I = \frac{3}{2}$,[38–41] $\frac{5}{2}$,[42,43] and $\frac{7}{2}$.[44–46] Otherwise we can always use a numerical procedure. The density matrix of Equation (40) becomes (Equation 43)

$$\rho(t_1) = T\exp(-\mathrm{i}\Omega t_1)T^{\dagger}\rho(0)T\exp(\mathrm{i}\Omega t_1)T^{\dagger} \tag{43}$$

After the first RF pulse, the density matrix is given by Equation (44),

$$\rho(t_1, \tau_2) = \exp(-\mathrm{i}H_{\mathrm{int}}\tau_2)\rho(t_1)\exp(\mathrm{i}H_{\mathrm{int}}\tau_2) \tag{44}$$

The effect of the second RF pulse is described by Equation (45),

$$\rho(t_1, \tau_2, t_3) = T\exp(-\mathrm{i}\Omega t_3)T^{\dagger}\rho(t_1, \tau_2)T\exp(\mathrm{i}\Omega t_3)T^{\dagger} \tag{45}$$

Finally, the density matrix during the acquisition period in a two-pulse sequence is

$$\rho(t_1, \tau_2, t_3, \tau_4) = \exp(-\mathrm{i}H_{\mathrm{int}}\tau_4)\rho(t_1, \tau_2, t_3)\exp(\mathrm{i}H_{\mathrm{int}}\tau_4) \tag{46}$$

Calculation of $\rho(t_1, \tau_2, t_3, \tau_4)$ requires 16 matrix multiplications. It would take a lot of time to perform them analytically to find out the position of the echo and the expression of the echo amplitude. Fortunately, careful analysis of $\rho(t_1, \tau_2, t_3, \tau_4)$ shows that only 12 matrix multiplications are needed (see sections 3.4.3 and 3.5.5).

3.1 One-pulse Sequence: Two-dimensional Nutation Experiment

In a one-pulse experiment, the Hamiltonian during the RF pulse is Equation (42), and that of the acquisition period is $H_{\mathrm{int}} = H_{\mathrm{Q}}^{(1)}$. First we describe the density matrix $\rho(t_1)$ at the end of the RF pulse. Then we discuss the simplest 2-D method, called nutation.

3.1.1 Spin Operators for $\mathrm{I} > \frac{1}{2}$

A matrix element of a spin operator O is denoted by $\langle r|O|c\rangle$ where the magnetic numbers r and c are half-integer numbers associated with the row and column of $\langle r|O|c\rangle$ in the matrix. We use the spin $I = \frac{5}{2}$ to describe the physical meanings of the density matrix elements at the end of the RF pulse. The extension to other spins is easy. A generalization of the three spin operators I_x, I_y, and I_z used to describe the dynamics of spin $I = \frac{1}{2}$ system are introduced for spins $I > \frac{1}{2}$: $I_x^{w,q}$, $I_y^{w,q}$, and $I_z^{w,q}$. Their matrix representations in the eigenstates of I_z are given by Equation (47).[8]

$$I_x^{w,q} = \begin{array}{c|ccc} & |w\rangle & & |q\rangle \\ \hline \langle w| & & & \frac{1}{2} \\ & & & \\ \langle q| & \frac{1}{2} & & \end{array}$$

$$I_y^{w,q} = \begin{array}{c|ccc} & |w\rangle & & |q\rangle \\ \hline \langle w| & & & -\frac{\mathrm{i}}{2} \\ & & & \\ \langle q| & \frac{\mathrm{i}}{2} & & \end{array}$$

$$I_z^{w,q} = \begin{array}{c|ccc} & |w\rangle & & |q\rangle \\ \hline \langle w| & \frac{1}{2} & & \\ & & & \\ \langle q| & & & -\frac{1}{2} \end{array} \tag{47}$$

These $(2I+1)\times(2I+1)$ matrices have only two nonzero elements. From Equation (47) we deduce that if we permute the two magnetic numbers w and q, we obtain Equation (48):

$$I_x^{w,q} = I_x^{q,w}, \quad I_y^{w,q} = -I_y^{q,w}, \quad I_z^{w,q} = -I_z^{q,w} \tag{48}$$

The operators I_x, I_y, and I_z are related to those of Equation (47) as Equations (49–52):[40]

$$I_x = \sum_{w,q} \sqrt{C(w,q)} I_x^{w,q} \tag{49}$$

$$I_y = \sum_{w,q} \sqrt{C(w,q)} I_y^{w,q} \tag{50}$$

$$I_z = \sum_{w,q} C(w,q) I_z^{w,q} \tag{51}$$

$$C(w,q) = I(I+1) - wq \tag{52}$$

In particular $C(m, m+1) = \zeta^2$, with ζ being defined by Equation (15). The operator I_x is related to the X pulse (Equation 39) and the operator I_y is related to the time-domain signal (Equations 19 and 20; Figure 2). For example, $I = \frac{5}{2}$ gives Equations (53–55):

$$I_x = \sqrt{5} I_x^{\frac{5}{2},\frac{3}{2}} + \sqrt{8} I_x^{\frac{3}{2},\frac{1}{2}} + 3 I_x^{\frac{1}{2},-\frac{1}{2}} + \sqrt{8} I_x^{-\frac{1}{2},-\frac{3}{2}} + \sqrt{5} I_x^{-\frac{3}{2},-\frac{5}{2}} \tag{53}$$

$$I_y = \sqrt{5} I_y^{\frac{5}{2},\frac{3}{2}} + \sqrt{8} I_y^{\frac{3}{2},\frac{1}{2}} + 3 I_y^{\frac{1}{2},-\frac{1}{2}} + \sqrt{8} I_y^{-\frac{1}{2},-\frac{3}{2}} + \sqrt{5} I_y^{-\frac{3}{2},-\frac{5}{2}} \tag{54}$$

$$I_z = 5 I_z^{\frac{5}{2},\frac{3}{2}} + 8 I_z^{\frac{3}{2},\frac{1}{2}} + 9 I_z^{\frac{1}{2},-\frac{1}{2}} + 8 I_z^{-\frac{1}{2},-\frac{3}{2}} + 5 I_z^{-\frac{3}{2},-\frac{5}{2}} \tag{55}$$

The matrix forms are given by Equations (56–58):

$$I_x = \begin{array}{c|cccccc} & \frac{5}{2} & \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{3}{2} & -\frac{5}{2} \\ \frac{5}{2} & 0 & \frac{1}{2}\sqrt{5} & 0 & 0 & 0 & 0 \\ \frac{3}{2} & \frac{1}{2}\sqrt{5} & 0 & \frac{1}{2}\sqrt{8} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2}\sqrt{8} & 0 & \frac{1}{2}\sqrt{9} & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{1}{2}\sqrt{9} & 0 & \frac{1}{2}\sqrt{8} & 0 \\ -\frac{3}{2} & 0 & 0 & 0 & \frac{1}{2}\sqrt{8} & 0 & \frac{1}{2}\sqrt{5} \\ -\frac{5}{2} & 0 & 0 & 0 & 0 & \frac{1}{2}\sqrt{5} & 0 \end{array} \tag{56}$$

$$I_y = \begin{array}{c|cccccc} & \frac{5}{2} & \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{3}{2} & -\frac{5}{2} \\ \frac{5}{2} & 0 & \frac{-\mathrm{i}}{2}\sqrt{5} & 0 & 0 & 0 & 0 \\ \frac{3}{2} & \frac{\mathrm{i}}{2}\sqrt{5} & 0 & \frac{-\mathrm{i}}{2}\sqrt{8} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{\mathrm{i}}{2}\sqrt{8} & 0 & \frac{-\mathrm{i}}{2}\sqrt{9} & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{\mathrm{i}}{2}\sqrt{9} & 0 & \frac{-\mathrm{i}}{2}\sqrt{8} & 0 \\ -\frac{3}{2} & 0 & 0 & 0 & \frac{\mathrm{i}}{2}\sqrt{8} & 0 & \frac{-\mathrm{i}}{2}\sqrt{5} \\ -\frac{5}{2} & 0 & 0 & 0 & 0 & \frac{\mathrm{i}}{2}\sqrt{5} & 0 \end{array} \tag{57}$$

$$I_z = \begin{array}{c|cccccc} & \frac{5}{2} & \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{3}{2} & -\frac{5}{2} \\ \frac{5}{2} & \frac{5}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{3}{2} & 0 & \frac{3}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 \\ -\frac{3}{2} & 0 & 0 & 0 & 0 & -\frac{3}{2} & 0 \\ -\frac{5}{2} & 0 & 0 & 0 & 0 & 0 & -\frac{5}{2} \end{array} \tag{58}$$

The two matrices of Equations (56) and (57) can also be obtained from Equation (59):

$$I_x = \frac{1}{2}(I_+ + I_-); \quad I_y = \frac{1}{2\mathrm{i}}(I_+ - I_-) \tag{59}$$

where I_+ and I_- are the raising and the lowering operators.

3.1.2 Meanings of the Matrix Elements of $\rho(t_1)$

A density matrix element $\langle w|\rho(t_1)|w\rangle$ in the main diagonal of $\rho(t_1)$ is defined by

$$\langle w|\rho(t_1)|w\rangle \equiv \langle I_z^{w,-w}(t_1)\rangle = \mathrm{Tr}\{\rho(t_1) I_z^{w,-w}\} \tag{60}$$

In other words, the main diagonal matrix element $\langle w|\rho(t_1)|w\rangle$ is the spin population of the energy level $|w\rangle$. The density matrix element $\langle a|\rho(t_1)|b\rangle$ below the main diagonal is defined by Equation (61):

$$\langle a|\rho(t_1)|b\rangle \equiv \langle I_+^{b,a}(t_1)\rangle = \mathrm{Tr}\{\rho(t_1) I_x^{b,a}\} + \mathrm{i}\mathrm{Tr}\{\rho(t_1) I_y^{b,a}\} \tag{61}$$

with the magnetic numbers a and b satisfying the condition $a < b$. The density matrix element $\langle w|\rho(t_1)|q\rangle$ above the main diagonal is defined by Equation (62):

$$\langle w|\rho(t_1)|q\rangle \equiv \langle I_-^{w,q}(t_1)\rangle = \mathrm{Tr}\{\rho(t_1) I_x^{w,q}\} - \mathrm{i}\mathrm{Tr}\{\rho(t_1) I_y^{w,q}\} \tag{62}$$

with $w > q$. The density matrix elements $\langle I_+^{b,a}(t_1)\rangle$ and $\langle I_-^{w,q}(t_1)\rangle$ are the spectral intensities of off-resonance lines because they are complex quantities. More generally, the density matrix element $\langle r|\rho(t_1)|c\rangle$ is called the p-quantum (pQ) coherence where $p = r - c$ is the coherence order. The spin populations of the energy levels are zero-quantum coherences. The matrix element $\langle w|\rho(t_1)|-w\rangle$ in the second main diagonal (Equation 63),

$$\langle w|\rho(t_1)|-w\rangle \equiv -\mathrm{i}\langle I_y^{w,-w}(t_1)\rangle = -\mathrm{i}\mathrm{Tr}\{\rho(t_1)I_y^{w,-w}\} \quad (63)$$

is an imaginary quantity. As the dispersion spectral line intensity $\langle I_x^{w,-w}(t_1)\rangle = 0$, $\langle w|\rho(t_1)|-w\rangle$ is related to the spectral intensity of the on-resonance absorption line. The SQ or $\pm$1Q coherences induce an RF voltage in the RF coil. These oscillating voltages are observed as an FID that follows the RF pulse. In contrast, MQ coherences do not induce an RF voltage. As a result, no signal is detected.

Figure 10 gives the density matrix $\rho(t_1)$ at the end of an X pulse for a spin $I = \frac{5}{2}$. The first-diagonal elements above the main diagonal of the density matrix are called 1Q coherences, the second-diagonal elements above the main diagonal are called 2Q coherences, and so on. The coherences below the main diagonal elements have a negative value of p. The RF coil only detects $\pm$1Q coherences. As the pQ coherences are the complex conjugates of the $-p$Q coherences, we consider only the $-$1Q coherences during the acquisition period.

The RF pulse dephased by an angle φ_1 relative to an X pulse is an important ingredient of MQ/MAS methodology[47] (section 3.5.8). A density matrix element of a φ_1 pulse is related to that of an X pulse by

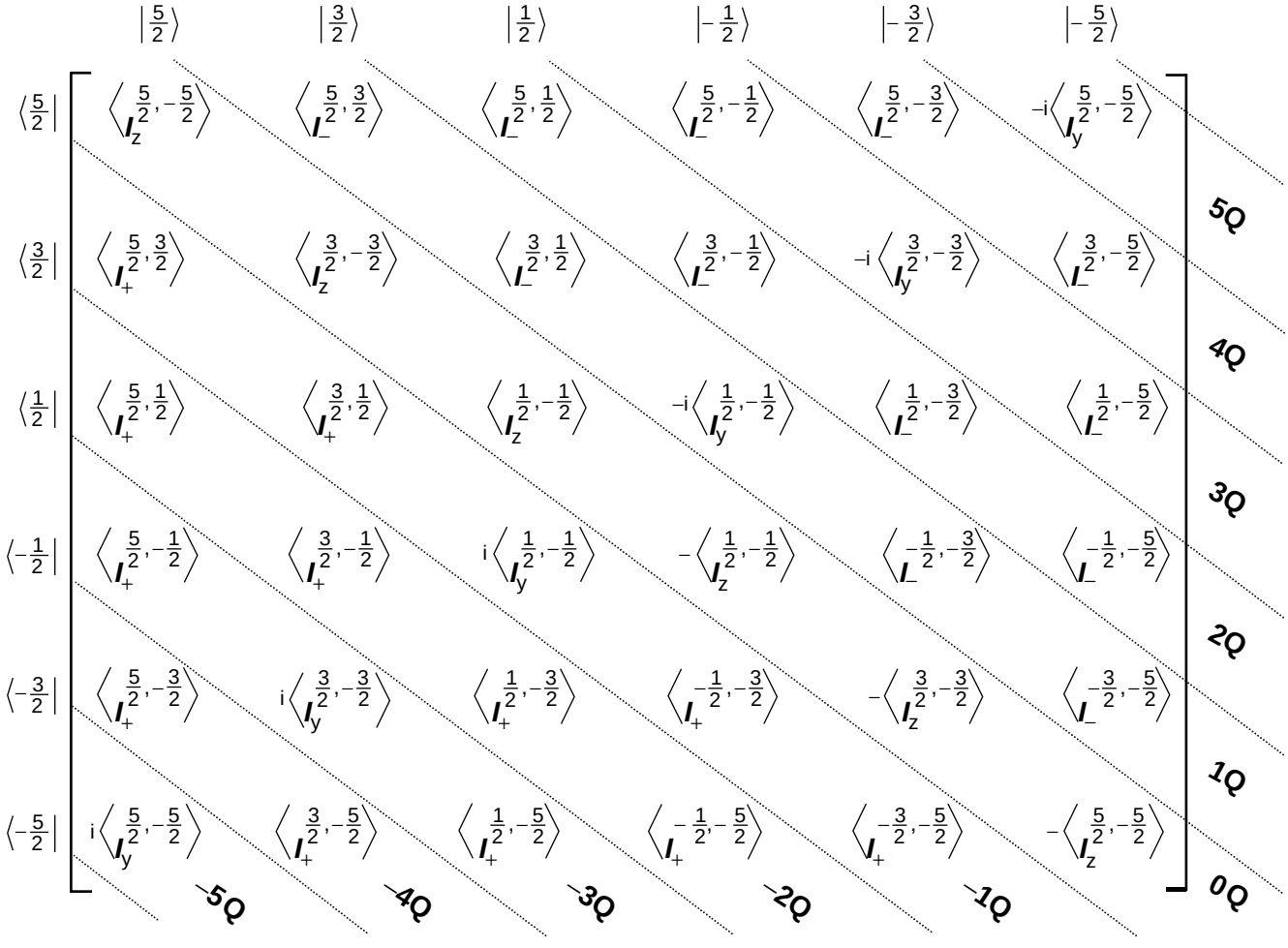

Figure 10 Definition of pQ coherences with the spin $I = \frac{5}{2}$ density matrix at the end of an X pulse, the coherence order p being an integer ranging from -5 to 5. The main diagonal contains real values, whereas the second main diagonal contains pure imaginary values. The other matrix elements are complex values. This matrix is a hermitian one. Matrix elements in the same parallel to the main diagonal have the same coherence order.

Equation (64):[48,49]

$$\langle r|\rho_{\varphi 1}(t_1)|c\rangle = \langle r|\rho(t_1)|c\rangle \exp\{-\mathrm{i}(r-c)\varphi_1\} = \langle r|\rho(t_1)|c\rangle \exp(-\mathrm{i}p\varphi_1) \quad (64)$$

In other words, when an RF pulse is dephased by an angle φ_1, a pQ coherence is dephased by $\exp(-\mathrm{i}p\varphi_1)$, whereas the zero-quantum coherences or spin populations are not affected.[8,25]

3.1.3 Spectral Line Intensity and Excitation Condition

The RF pulse amplitude w_{RF} is usually higher than internal interactions, except the quadrupole interaction. Therefore at least $H_Q^{(1)}$ must be taken into account during the RF pulse to predict the spin dynamics. Two extreme cases appear:

1. When $w_Q \ll w_{RF}$ all the transitions are irradiated uniformly, and the excitation is known as nonselective or hard-pulse excitation because $H_Q^{(1)}$ need not be considered during the RF pulse.
2. When $w_Q \gg w_{RF}$ two consecutive energy levels are generally irradiated, and the excitation is said to be selective because the multilevel system is reduced to a two-energy-level system whose absorption line is on resonance and $H_Q^{(1)}$ is taken into account during the RF pulse.

For nonselective excitation, the relative spectral line intensity $I_{NS}^{m+1,m}$ of the transition $(m+1 \leftrightarrow m)$ is (Equation 65)[50,51]

$$I_{NS}^{m+1,m} = \frac{\zeta^2}{\sum_{m=-I}^{I-1} \zeta^2} \sin w_{RF} t_1 \quad (65)$$

For selective excitation, the relative spectral line intensity $I_S^{m+1,m}$ of the transition $(m+1 \leftrightarrow m)$ is (Equation 66)[50,51]

$$I_S^{m+1,m} = \frac{\zeta}{\sum_{m=-I}^{I-1} \zeta^2} \sin \zeta w_{RF} t_1 \quad (66)$$

Comparison of Equations (65) and (66) shows that for selective excitation the pulse duration that maximizes the relative spectral line intensity of the on-resonance transition $(m+1 \leftrightarrow m)$ is ζ times as short as that for the nonselective excitation. Furthermore, there is also a loss of relative spectral line intensity by the same factor ζ.[52] If the pulse duration is sufficiently short, Equation (67),

$$\sin \zeta w_{RF} t_1 \approx \zeta w_{RF} t_1 \quad (67)$$

then Equations (65) and (66) become identical:

$$I_{NS}^{m+1,m} \approx I_S^{m+1,m} \approx \frac{\zeta^2}{\sum_{m=-I}^{I-1} \zeta^2} w_{RF} t_1 \quad (68)$$

Although Equation (68) has been deduced from two relationships with limited validity, it becomes valid for both $w_Q \ll w_{RF}$ and $w_Q \gg w_{RF}$ provided that the pulse duration is sufficiently short. Consequently, by continuity, Equation (68) remains valid for any ratio of w_Q/w_{RF}. In other words, there is a linear region in which the relative spectral line intensity is independent of w_Q, if the transition is excited on resonance with a short pulse duration.

Analytical expressions of the spectral line intensity for spins $I = \frac{3}{2}$,[21,53–56] $\frac{5}{2}$,[42,43] and $\frac{7}{2}$[44–46] are available for any w_Q/w_{RF} ratio. Taking an internal interaction such as $H_Q^{(1)}$ during the RF pulse is called soft-pulse excitation, which includes both nonselective and selective excitations. For spin $I = \frac{9}{2}$, a numerical procedure must be applied to solve Equations (41) and (43). The SQ spectral line intensity is not given by $\langle I_y^{w+1,w}(t_1)\rangle = \mathrm{Tr}\{\rho(t_1) I_y^{w+1,w}\}$, as shown in Figure 11 for a spin $I = \frac{5}{2}$, but by Equation (69):

$$\langle I_y^{w+1,w}(t_1)\rangle \zeta(w+1,w) = \zeta(w+1,w)\mathrm{Tr}\{\rho(t_1) I_y^{w+1,w}\} \quad (69)$$

This is because SQ coherences are detected directly by an RF coil, in contrast to MQ coherences. As the quadrupole coupling w_Q depends on the orientations of $\mathbf{B}_0$ with respect to the crystal, a computer is needed to average the response of the spin system over all of these orientations in order to predict the experimental FID amplitude for a powder sample.[57]

As the NMR absorption line is proportional to the number of spins, the usual way to quantify the number of spins is to compare it with a sample containing a known quantity of spins. Unfortunately, a delay of a few microseconds is necessary before acquisition of the FID. This delay, which depends on the design of the probe head, is called the dead-time of the receiver (Figure 2a). If the latter is much shorter than T_{FID}, the loss of spectral line intensity is negligible and the quantification of the number of spins is reliable. This is the case in high-resolution liquid-state NMR. In solid-state NMR, the absorption lines are broad, that is, T_{FID} is short. The dead-time of the receiver becomes a significant part of

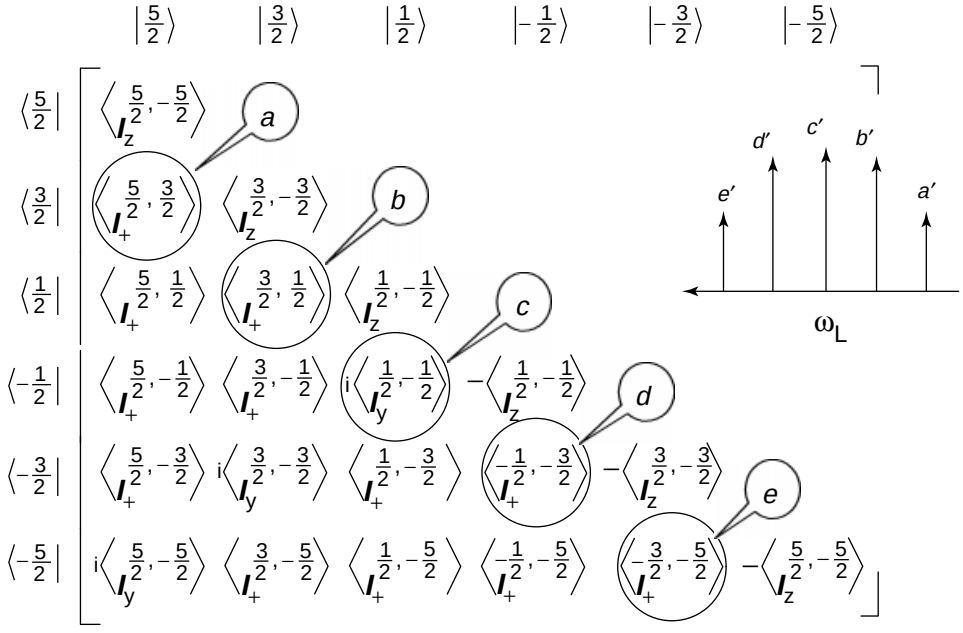

Figure 11 Directly observable coherences of spin $I = \frac{5}{2}$ at the end of an RF pulse. Only ± 1Q coherences, which are complex conjugated, are detected by an RF coil. By convention only -1Q coherences (circled) are considered. These -1Q coherences a–e are related to the area of the absorption lines ($a' \equiv \frac{3}{2} \leftrightarrow \frac{5}{2}$), ($b' \equiv \frac{1}{2} \leftrightarrow \frac{3}{2}$), ($c' \equiv -\frac{1}{2} \leftrightarrow \frac{1}{2}$), ($d' \equiv -\frac{3}{2} \leftrightarrow -\frac{1}{2}$), and ($e' \equiv -\frac{5}{2} \leftrightarrow -\frac{3}{2}$), respectively. The values of ζ (see Table 1) have not been considered here.

the FID and the quantification of the number of spins may not be reliable.

3.1.4 *One-dimensional and Two-dimensional Nutation Experiments*

The spectral line intensity is a sum of $(I + \frac{1}{2})^2$ sine functions[58] whose angular frequencies are the nutation frequencies of the spin system around the RF field $\mathbf{B}_1$. Study of the spectral line intensity versus the pulse duration is called the 1-D nutation experiment. Fourier transform of the spectral line intensity with respect to the pulse duration leads to the first 2-D NMR experiment for half-integer quadrupole spins, and is called the 2-D nutation experiment.[59] Figure 12 presents the principle of this method. The experiment consists in acquiring a series of FIDs following a single RF excitation pulse for increasing pulse duration. The F_2 dimension of the 2-D spectrum is the usual chemical shift axis, whereas the F_1 dimension is the axis of the nutation frequencies of the spin system around $\mathbf{B}_1$. The implementation is simple, but the interpretation is not obvious because the 2-D spectrum associated with a single crystallographic site may consist of several peaks along the F_1 dimension; the positions of these peaks are related to the nutation frequencies. Furthermore, only the SQ coherences generated by the RF pulse are detected by the RF coil; the MQ coherences are not. Rotary-echo and off-resonance nutations[55] have also been introduced. Simulation of experimental 2-D spectra is necessary to extract the two quadrupole parameters, namely the quadrupole coupling constant $e^2qQ/\hbar$ and the asymmetry parameter η.

3.1.4.1 Spin $\mathrm{I} = \frac{3}{2}$ Figure 13 shows the ^{23}Na ($I = \frac{3}{2}$) stacked plotted MAS spectra of a mixture of powdered NaCl and $NaNO_2$ for increasing pulse duration.[60] The powder pattern of Na in $NaNO_2$ is typical of a central line affected by $H_{\mathrm{Q}}^{(2)}$, in agreement with the noncubic crystallographic structure of $NaNO_2$. However, the powder pattern of Na in NaCl has a featureless line shape. As NaCl has a cubic crystallographic structure, the EFG around the sodium site must be very small, if any. All the three absorption lines (one central and two satellite lines) are not resolved in NaCl. It is well known that the NMR spectral line intensities are related to the number of spins in the sample; the ratio of two lines should give the spin population ratio. Figure 13 clearly shows that the spectral line intensity ratio varies with the pulse duration, giving any spin population ratio. This odd behavior is in agreement with the analytical result of Figure 14(a), showing the central-line intensity of a spin $I = \frac{3}{2}$ system versus the pulse duration for several $w_{\mathrm{Q}}/w_{\mathrm{RF}}$ ratios.[39] Only when the pulse duration is sufficiently short is the central-line intensity independent of the $w_{\mathrm{Q}}/w_{\mathrm{RF}}$ ratio. However, for satellite transitions (Figure 14b),[60] there is no linear region where the satellite-line intensity is independent of the $w_{\mathrm{Q}}/w_{\mathrm{RF}}$ ratio. This is because the

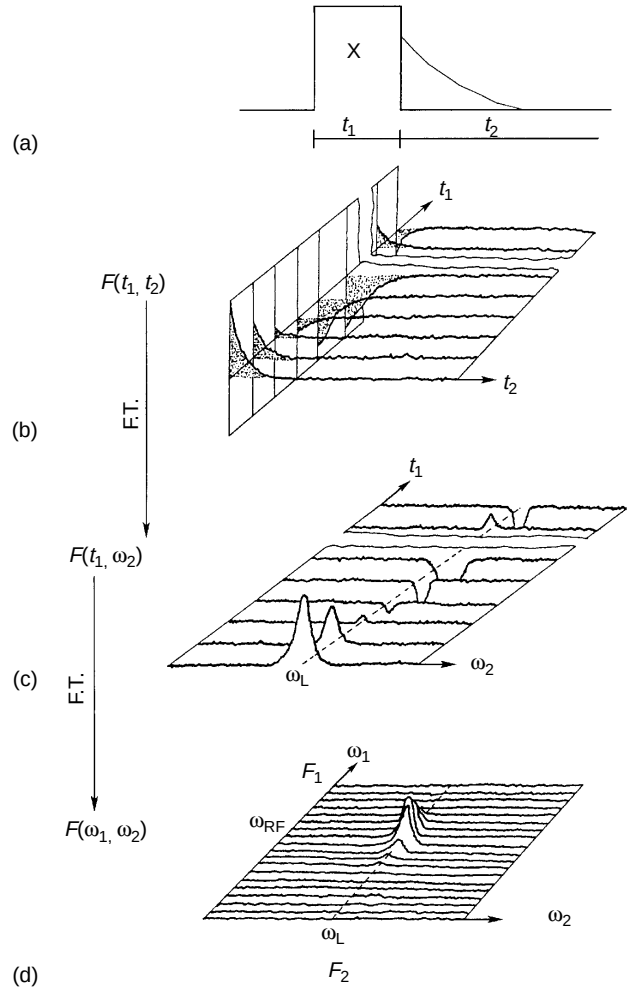

Figure 12 Nutation experiment. (a) The two time-domain parameters involved in 2-D experiments – the pulse duration t_1 describes the evolution period whereas t_2 describes the acquisition period. (b) $F(t_1, t_2)$ is a series of FID acquired with increasing pulse duration t_1. (c) $F(t_1, w_2)$, the Fourier transform of $F(t_1, t_2)$ with respect to t_2, is a series of spectra with increasing pulse duration t_1. (d) $F(w_1, w_2)$, the Fourier transform of $F(t_1, w_2)$ with respect to t_1, is the 2-D nutation spectrum. F_2 is the usual chemical shift axis and F_1 is the nutation frequency axis. The peak is located at w_{RF} in the F_1 dimension and w_L in the F_2 dimension for a spin $I = \frac{1}{2}$ system.

satellite lines become increasingly off resonance when the w_Q/w_{RF} ratio increases. The effective RF field that the spin system is submitted to differs with $\mathbf{B}_1$.

For a short pulse duration the spectral line intensity ratio does not depend on w_Q. However, precautions must be taken to ensure that the spectral line intensity ratio is that of the spin populations. For example in Figure 15,[59,60] the molar ratio of $NaNO_2$ to NaCl is 6.50, whereas the ^{23}Na experimental spectral line intensity ratio is 2.50. These two ratios are very different. As the central line of the spin $I = \frac{3}{2}$ system represents 4/10 of the

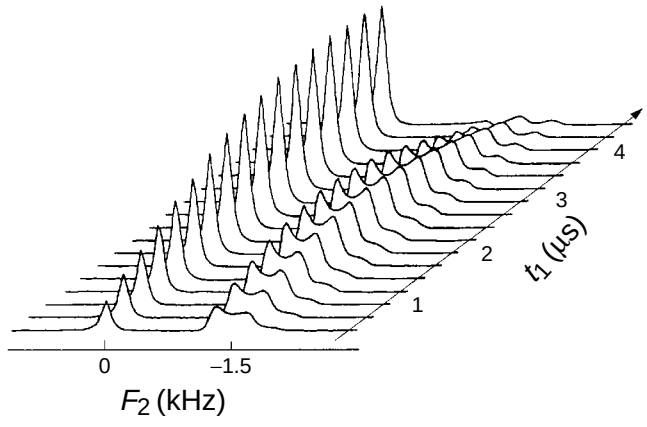

Figure 13 Effect of pulse duration on spectral line intensities: stacked plot of one-pulse ^{23}Na ($I = \frac{3}{2}$) MAS spectra in a mixture of powdered NaCl (line at 0 kHz) and $NaNO_2$ (line at −1.5 kHz) versus the pulse duration t_1, with a recycle delay $D_0 = 60$ s.

full spectrum (see Table 1), multiplying the experimental spectral line intensity ratio by 10/4 indeed gives the spin population ratio.

The main problem in quantifying the spin population is not only the application of a short pulse duration but also the identification of the nature of a featureless line shape (a central line or the full spectrum). One way to proceed with this is to study its behavior with respect to the pulse duration and compare it with the simulation. If the quadrupole coupling constant is relatively small, all the spectral line intensities should be detected.

Figure 16 shows the ^{23}Na 2-D nutation MAS spectrum of the same mixture of NaCl and $NaNO_2$.[61] Peak B is the 2-D line of $NaNO_2$, whereas peaks A and A′ are those of NaCl. The latter two peaks are located at w_{RF} and $2w_{RF}$, respectively, along the nutation axis. Peak B is located at $2w_{RF}$ because $\zeta = 2$ for $I = \frac{3}{2}$ (Table 1). The positions of peaks A and B are in agreement with the strength of the EFG around the sodium nuclei in the two compounds. The EFG around sodium nuclei in NaCl is nearly zero; the nutation frequency is that of a free spin not submitted to the quadrupole interaction. The nutation frequency of sodium nuclei in $NaNO_2$ is $2w_{RF}$ because the quadrupole coupling constant and therefore the quadrupole coupling is much larger than w_{RF}. Peak A′ occurs because the recycle delay D_0 in this experiment is too short compared to the spin–lattice relaxation time T_1 of sodium nuclei. A fuller explanation is given in section 3.1.4.3. The ^{23}Na 2-D nutation experiment has been applied to zeolite NaA.[62]

3.1.4.2 Spin $I = \frac{5}{2}$ The spin $I = \frac{5}{2}$ is not developed here because numerous papers deal with aluminum nuclei ^{27}Al ($I = \frac{5}{2}$) in zeolites.[50,63–66] The effects of pulse duration on the central-line shape are shown in

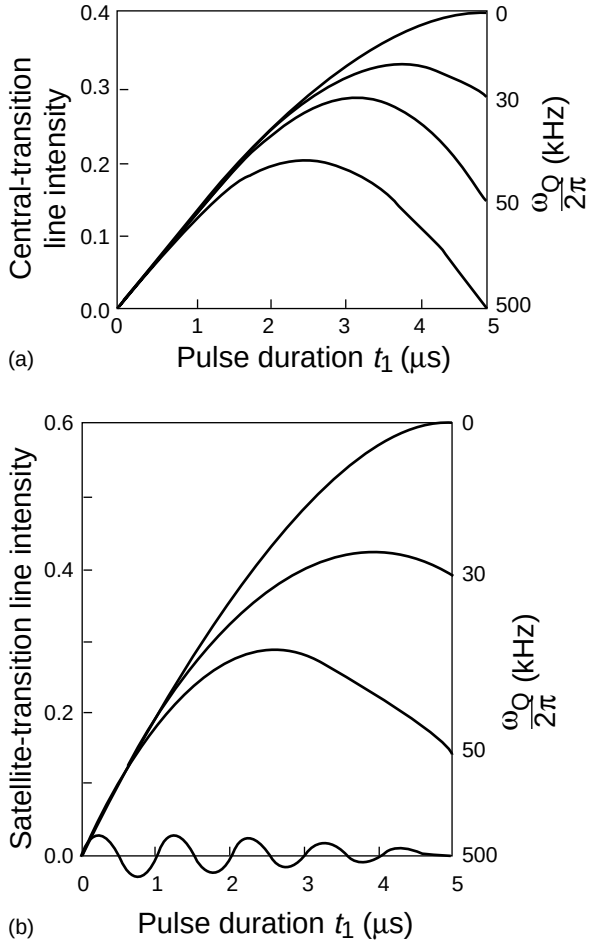

Figure 14 Calculated one-pulse spectral line intensities for a spin $I = \frac{3}{2}$ in a single crystal and submitted to $H_Q^{(1)}$ during the RF pulse versus the pulse duration t_1 for several values of the quadrupole coupling w_Q ($w_Q/(2\pi) = 0$, 30, 50, and 500 kHz), with $w_{RF}/(2\pi) = 50$ kHz: (a) the central transition ($-\frac{1}{2} \leftrightarrow \frac{1}{2}$); (b) the two satellite transitions ($\frac{1}{2} \leftrightarrow \frac{3}{2}$) and ($-\frac{3}{2} \leftrightarrow -\frac{1}{2}$).

Figure 17 which is the MAS central line of ^{55}Mn ($I = \frac{5}{2}$) in $KMnO_4$.[59] Only short pulse durations (<3 μs) yield undistorted MAS powder patterns in agreement with frequency-domain results.

3.1.4.3 Spin $I = \frac{7}{2}$ The FID amplitude in the steady state $\langle I_y(t_1)\rangle$ is related to that in the Boltzmann state $\langle I_z(t_1 = 0)\rangle$, the recycle delay D_0, and the spin–lattice relaxation time T_1 (Equation 70):[8,67]

$$\langle I_y(t_1)\rangle = \langle I_z(t_1 = 0)\rangle \frac{\{1 - \exp(-D_0/T_1)\} \sin w_{RF} t_1}{1 - \exp(-D_0/T_1) \cos w_{RF} t_1} \quad (70)$$

Figure 18 represents the graphs of $\langle I_y(t_1)\rangle : \langle I_z(t_1 = 0)\rangle$ versus the pulse flip angle $w_{RF}t_1$, ranging from 0 to π, for various ratios of D_0/T_1. The maximums of each curve as well as the associated pulse flip angle $w_{RF}t_1$ decrease

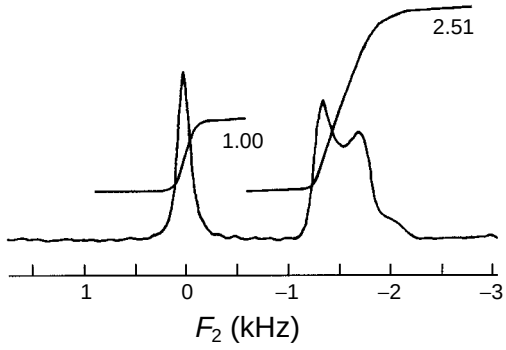

Figure 15 Quantification of spin populations with a short pulse duration $t_1 = 0.25$ μs: one-pulse ^{23}Na ($I = \frac{3}{2}$) MAS spectrum and the integrated area in a mixture of powdered NaCl (line at 0 kHz) and $NaNO_2$ (line at −1.5 kHz) with $D_0 = 60$ s. The molar ratio $NaNO_2$/NaCl is 6.50.

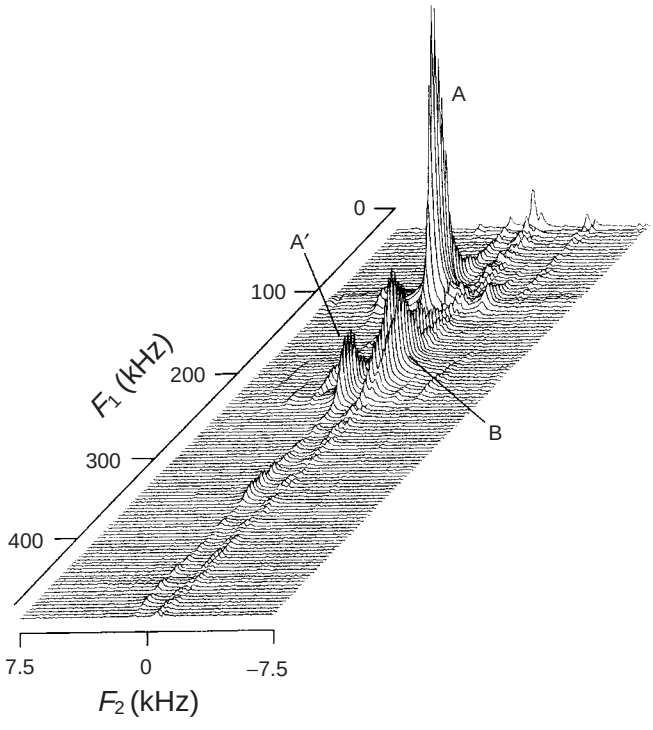

Figure 16 ^{23}Na ($I = \frac{3}{2}$) 2-D nutation MAS spectrum (magnitude mode) in a mixture of powdered NaCl (peaks A and A′) and $NaNO_2$ (peak B) with $D_0 = 5$ s.

when the D_0/T_1 ratio decreases. Only the duration of the π pulse, which yields a zero signal, does not depend on the D_0/T_1 ratio. Experimentally, the π-pulse duration is determined first with an aqueous solution, then the $\pi/2$-pulse duration is deduced by taking half the π-pulse duration.

Figure 19(a) represents, in magnitude, the Fourier transform of Equation (70) with respect to the pulse duration, for various D_0/T_1 ratios. It shows peaks

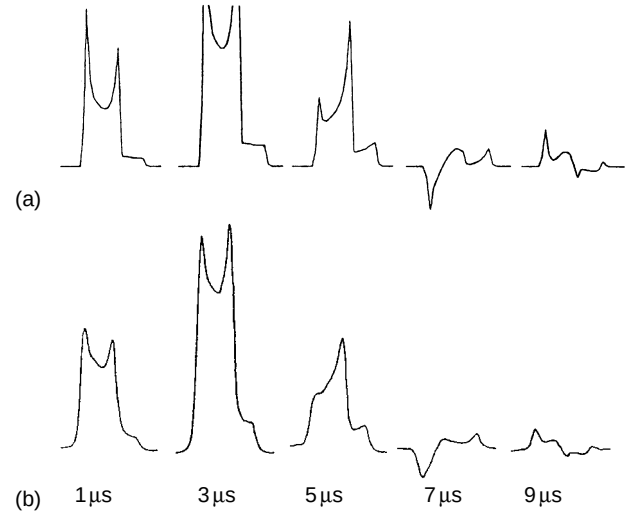

Figure 17 Effect of pulse duration on the central-line shape: (a) simulated and (b) experimental one-pulse MAS spectra of ^{55}Mn ($I = \frac{5}{2}$) in powdered $KMnO_4$ for several pulse durations $t_1 = 1$, 3, 5, 7, and 9 μs. (Reproduced by permission of The American Physical Society from A. Samoson, E. Lippmaa, *Phys. Rev. B*, **28**, 6567–6570 (1983).)

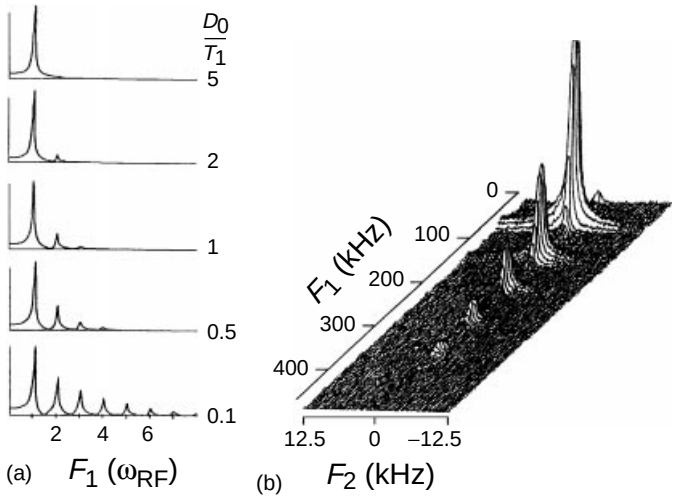

Figure 19 Effect of the recycle delay D_0 on the 2-D nutation spectrum. (a) Fourier transform of Equation (70) or curves in Figure 18 for several D_0/T_1 ratios (5, 2, 1, 0.5, and 0.1). (b) The ^{133}Cs ($I = \frac{7}{2}$) 2-D nutation spectrum (magnitude mode) in powdered CsCl, with $D_0 = 30$ s.

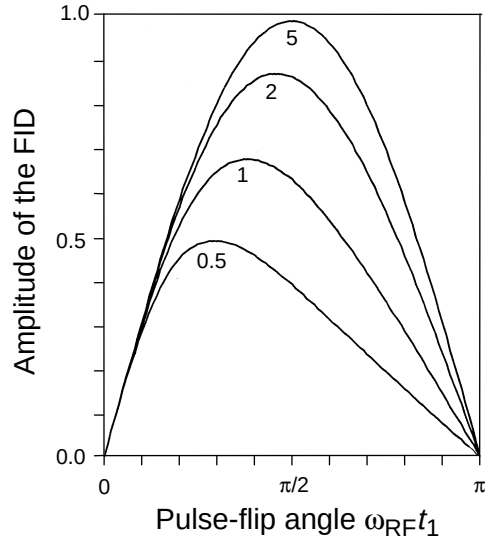

Figure 18 Effect of the recycle delay D_0 on the FID amplitude or the area of the corresponding absorption line: calculated FID amplitude in an aqueous solution versus the pulse flip angle $w_{RF}t_1$ for various D_0/T_1 ratios (5, 2, 1, and 0.5), where T_1 is the spin–lattice relaxation time.

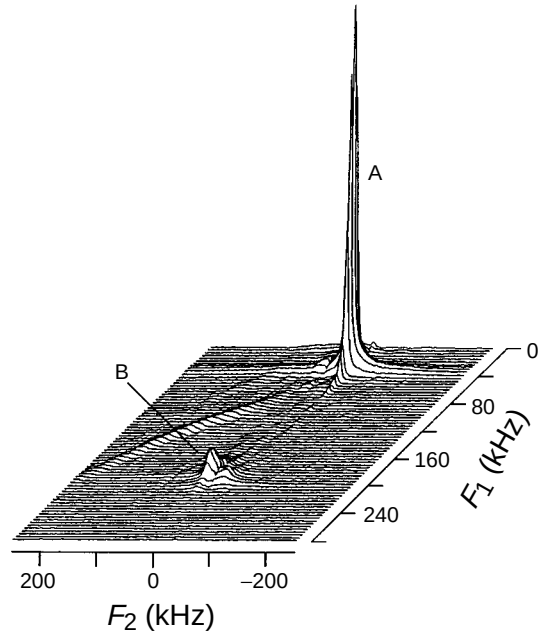

Figure 20 ^{93}Nb ($I = \frac{9}{2}$) 2-D nutation spectrum (magnitude mode) in a mixture of powdered NbO_2F (peak A) and $LiNbO_3$ (peak B) under static conditions with $D_0 = 0.1$ s.

at manifolds of w_{RF}. The number of peaks increases when the D_0/T_1 ratio decreases. We have checked this spin–lattice relaxation effect with cesium nuclei ^{133}Cs ($I = \frac{7}{2}$) in a powdered CsCl (Figure 19b). By comparing the two graphs of Figure 19, the spin–lattice relaxation time of cesium is estimated to be about 300 s.

3.1.4.4 Spin $I = \frac{9}{2}$ The largest stable nuclear spin for NMR is $I = \frac{9}{2}$. The nucleus most studied is niobium ^{93}Nb. Figure 20 shows the 2-D nutation spectrum of ^{93}Nb in a mixture of powdered $LiNbO_3$ and NbO_2F.[68] The crystallographic structure of NbO_2F is cubic, whereas that of $LiNbO_3$ is rhombohedral. Peak A corresponds to

NbO_2F and peak B to $LiNbO_3$. In the nutation frequency axis, peak A is located at w_{RF}, whereas peak B is at $5w_{RF}$ ($\zeta = 5$ for the central transition of a spin $I = \frac{9}{2}$). Peak A is much more intense than peak B because the latter represents only one $\frac{25}{165}$ part of the number of spins.

3.2 Composite-pulse Sequences: Spurious Signal Suppression

Spurious signals in pulsed NMR are often observed.[69] They disturb the signals in such a way that the interpretation of the spectra becomes difficult. We will focus on two kinds of spurious signals:

1. acoustic ringing from the NMR probe during the study of low gyromagnetic ratio nuclei in strong $\mathbf{B}_0$;
2. piezoelectric signals that occur when ferroelectric materials are studied in single crystal form.

This section discusses the first kind of spurious signal, postponing the second case to section 3.5.3. Figure 21(a) shows the ^{131}Xe ($I = \frac{3}{2}$) one-pulse spectra of xenon gas physisorbed in NaY zeolite, for increasing pulse durations.[70] The baselines of these spectra are so

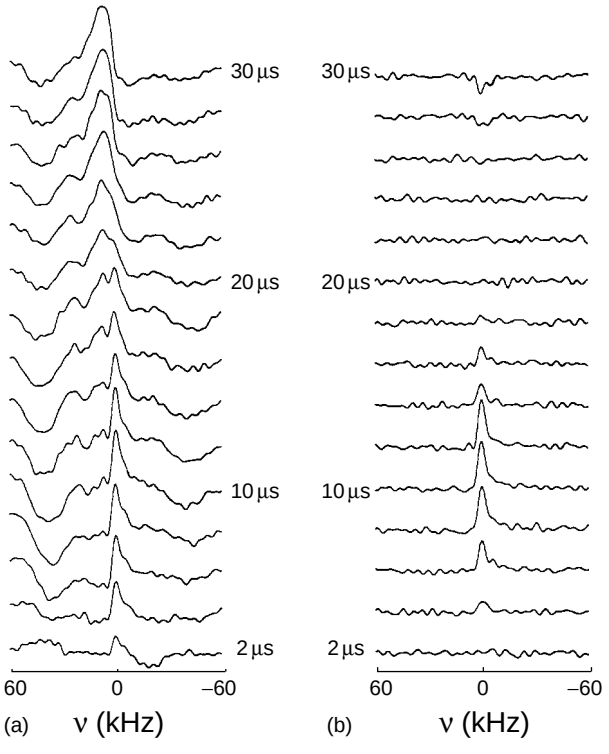

Figure 21 Suppression of the ringing signals from the NMR probe head by using a composite-pulse sequence (Figure 22 and sequence 1): ^{131}Xe ($I = \frac{3}{2}$) spectra of Xe gas physisorbed in NaY zeolite, acquired with (a) one-pulse sequence and (b) composite-pulse sequence for increasing pulse duration t_1 ranging from 2 to 30 µs by steps of 2 µs.

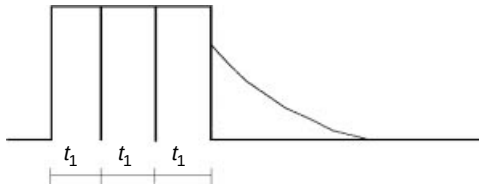

Figure 22 Composite-pulse sequence with phases given in sequence 1 cancels the ringing signals from the NMR probe head.

distorted that the ^{131}Xe signals are difficult to recognize. A composite-pulse sequence consisting of three successive RF pulses of the same duration t_1 (Figure 22) was applied with the following phases:[71]

first RF pulse:	$-X$	$-X$	$+X$	$+X$	
second RF pulse:	$+X$	$+X$	$+X$	$+X$	(sequence 1)
third RF pulse:	$+X$	$-X$	$-X$	$+X$	
receiver:	$+x$	$-x$	$+x$	$-x$	

This sequence, called sequence 1, consists of X and $-X$ pulses. Figure 21(b) presents ^{131}Xe ($I = \frac{3}{2}$) spectra acquired with the composite-pulse sequence.[70] These spectra have flattened baselines, allowing the ^{131}Xe signal to be recognized. The signal has a featureless line shape which does not allow the conclusion that the ^{131}Xe nuclei are submitted to quadrupole interaction. One response to this question is to study the behavior of the spectral line intensity versus the pulse duration t_1.

The Hamiltonian during an X pulse (Equation 71) is

$$H^{(X)} = H_Q^{(1)} - w_{RF} I_x \tag{71}$$

The Hamiltonian during a $-X$ pulse (Equation 72) is

$$H^{(-X)} = H_Q^{(1)} + w_{RF} I_x \tag{72}$$

The two Hamiltonians $H^{(X)}$ and $H^{(-X)}$ have the same eigenvalue matrix Ω, but different eigenvector matrices, denoted by T and U, respectively. The central-line intensity is given by $\zeta(\frac{1}{2}, -\frac{1}{2})\mathrm{Tr}\ \{\rho_1(t_1) - \rho_2(t_1) + \rho_3(t_1) - \rho_4(t_1)\} I_y^{\frac{1}{2},-\frac{1}{2}}$, where the signs of the four density operators $\rho_j(t_1)$ are those of the receiver phase given in sequence 1, with (Equation 73)

$$\rho_1(t_1) = T \exp(-2i\Omega t_1) T^\dagger U \exp(-i\Omega t_1) U^\dagger \rho(0) U \times \exp(i\Omega t_1) U^\dagger T \exp(2i\Omega t_1) T^\dagger \tag{73}$$

for the first acquisition of the FID. For the second FID acquisition Equation (74) applies:

$$\rho_2(t_1) = U \exp(-i\Omega t_1) U^\dagger T \exp(-i\Omega t_1) T^\dagger U \times \exp(-i\Omega t_1) U^\dagger \rho(0) U \exp(i\Omega t_1) U^\dagger T \times \exp(i\Omega t_1) T^\dagger U \exp(i\Omega t_1) U^\dagger \tag{74}$$

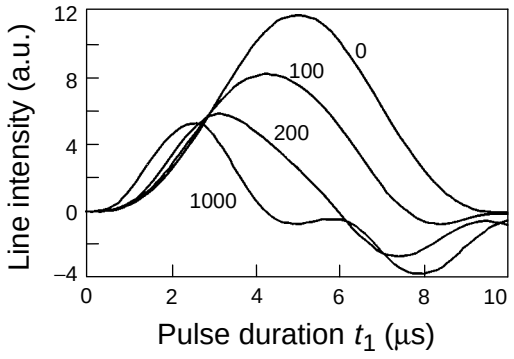

Figure 23 Calculated central-line intensities for spin $I = \frac{3}{2}$ in a powder sample, excited by the composite-pulse sequence of Figure 22, versus the pulse duration t_1 for several values of the quadrupole coupling constant $e^2qQ/\hbar$ (0, 100, 200, and 1000 kHz), with $\eta = 0$ and $w_{RF}/(2\pi) = 50$ kHz.

Equation (75) is for the third acquisition:

$$\rho_3(t_1) = U \exp(-i\Omega t_1) U^\dagger T \exp(-2i\Omega t_1) T^\dagger \rho(0) T \times \exp(2i\Omega t_1) T^\dagger U \exp(i\Omega t_1) U^\dagger \quad (75)$$

For the last acquisition Equation (76) is used:

$$\rho_4(t_1) = T \exp(-3i\Omega t_1) T^\dagger \rho(0) T \exp(3i\Omega t_1) T^\dagger \quad (76)$$

Figure 23 shows the graphs of the central-line intensities of a spin $I = \frac{3}{2}$ versus the pulse duration t_1 by taking into account $H_Q^{(1)}$ during the excitation of the spin system by the composite-pulse sequence.[70] Several quadrupole coupling constants have been used, which means that the sample is supposed to be a powder. The graph for $e^2qQ/\hbar = 0$ is bell shaped, and that for $e^2qQ/\hbar = 1$ MHz has two bell shapes of opposite signs. As with the one-pulse sequence applied to spin $I = \frac{3}{2}$ (Figure 14), the pulse duration that maximizes the spectral line intensity in the selective excitation is half that corresponding to the non-selective excitation condition. In contrast to the one-pulse sequence, there is no linear regime where the response of the spin system is proportional to the pulse duration, even for short pulse durations. Quantitative results based on spin population are not possible with this composite-pulse sequence. Furthermore, broad absorption lines are lost in the dead-time of the receiver, as in the one-pulse experiment. Spin-echo experiments are required in this case.

3.3 Two-pulse Sequences

Fitting the experimental spectral line intensity versus the pulse duration of the one-pulse sequence allows us to determine the two quadrupole parameters. However, this sequence depends only on two independent experimental parameters, the RF pulse amplitude w_{RF} and the pulse duration t_1. As w_{RF} has a limited range of variation, to increase the number of independent experimental parameters we can apply a two-pulse sequence (Figure 9a). Keeping w_{RF} at its maximum value provided by the high-power amplifier so that our assumptions remain valid (Figure 9), we have three time-domain parameters: the durations t_1 and t_3 of the two RF pulses, and the pulse separation τ_2. The graph of the spectral line intensity versus one of these three parameters, the other two being constant, allows us to extract the two quadrupole parameters. Fitting two sets of experimental spectral line intensities, obtained with different values of the two constants, should provide us with the same quadrupole parameters. In practice, we apply both one- and two-pulse sequences. As the delay required to fit the one-pulse spectral line intensities is short, we use the two quadrupole parameters provided by the one-pulse experiment as initial values to fit the two-pulse sequence data.

The internal interaction throughout the two-pulse experiment is $H_Q^{(1)}$. The analytical expressions of the central-line intensity after the second RF pulse for spins $I = \frac{3}{2}$,[21,72,73] $\frac{5}{2}$,[74] and $\frac{7}{2}$[75] in a single crystal are available. Fortunately they can be calculated numerically. As the spin–lattice relaxation is not taken into account, the pulse separation τ_2 must be short. It is of the order of magnitude of the dead-time of the receiver (typically 10–20 μs). All the density matrix elements at the end of the first RF pulse contribute to the spectral line intensity at the end of the second RF pulse. Equations (43–45) are used to calculate the central-line intensities, which are given by $\zeta(-\frac{1}{2}, \frac{1}{2})\mathrm{Tr}\ \rho(t_1, \tau_2, t_3) I_y^{\frac{1}{2}-\frac{1}{2}}$. For powder samples the procedure described in section 2.1 should be used to average the quadrupole coupling w_Q.

Figure 24 presents the ^{23}Na ($I = \frac{3}{2}$) spectra of $NaNO_3$ acquired with the two-pulse sequence for increasing second-pulse duration.[72] The Na central line has featureless line shape, irrelevant for line-shape analysis for the extraction of quadrupole parameters. Figure 25 shows the fitting of the experimental central-line intensities versus the second-pulse duration.[72] The result of the fitting for the quadrupole coupling constant is quite good.

Two-pulse sequences without pulse separation, that is, $\tau_2 = 0$, such as the spin-lock sequence[48,76] or the rotary echo sequence,[77] can also be applied to determine the quadrupole parameters from a featureless line shape.

3.4 Solomon Echo Sequences

The FIDs that have short durations T_{FID} are lost in the dead-time of the receiver following an RF pulse. This is the case of the FIDs of the satellite transitions whose powder pattern spreads far out from the central line. As the satellite-transition powder pattern provides us with the quadrupole parameters, we should apply the

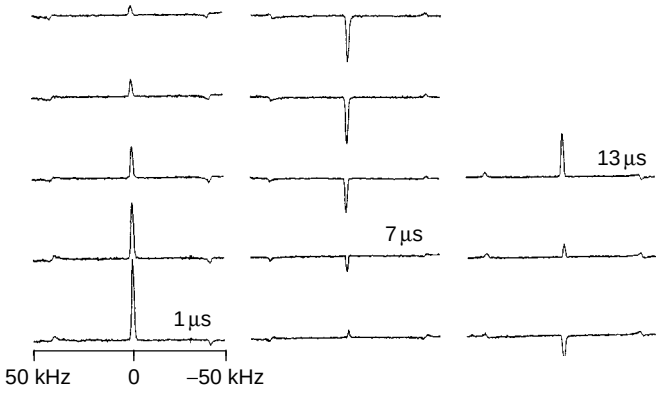

Figure 24 Two-pulse ^{23}Na ($I = \frac{3}{2}$) spectra in a powdered $NaNO_3$ for increasing second-pulse duration t_3 from 1 to 13 μs, in static condition with a short pulse separation $\tau_2 = 10\,\mu s$ and $D_0 = 60\,s$. (Reproduced from P.P. Man, 'Determination of the Quadrupolar Coupling Constant in Powdered Samples with a Two In-phase RF Pulse Sequence in Solid-state NMR', *Chem. Phys. Lett.*, **168**, 227–232, Copyright (1990) with permission from Elsevier Science.)

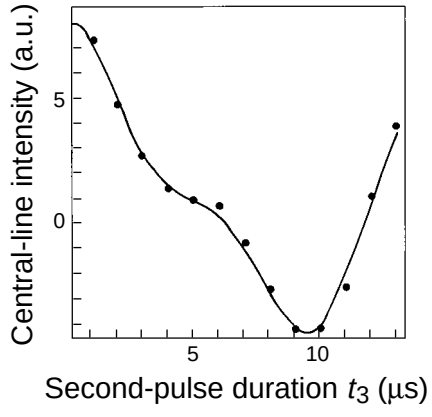

Figure 25 Experimental (filled circles) and calculated (solid line) central-line intensities of ^{23}Na ($I = \frac{3}{2}$) in a powdered $NaNO_3$ corresponding to Figure 24. The following parameters were used in the calculation: $t_1 = 4\,\mu s$; $\tau_2 = 10\,\mu s$; $t_3 = 1–13\,\mu s$; $e^2qQ/\hbar = 336\,kHz$; $\eta = 0$; $w_{RF}/(2\pi) = 46\,kHz$ (Reprinted from P.P. Man, 'Determination of the Quadrupolar Coupling Constant in Powdered Samples with a Two In-phase RF Pulse Sequence in Solid-state NMR', *Chem. Phys. Lett.*, **168**, 227–232, Copyright (1990) with permission from Elsevier Science.)

Solomon echo sequences, two-pulse sequences separated by a delay τ_2, to recover this powder pattern. These echo sequences allow us to detect the satellite transitions as echoes located far from the second RF pulse, therefore they are not affected by the dead-time of the receiver following the second RF pulse.

The positions of the Solomon echoes and the contribution of coherences generated by the first RF pulse have been already determined for the four half-integer quadrupole spins.[16] The interaction considered throughout the experiment is still $H_Q^{(1)}$ (see Figure 9a). The pulse separation τ_2 must be larger than the dead-time of the receiver but much smaller than the FID duration of the central transition T_{FID} so that the Solomon echoes can be detected. There is no Solomon echo for the central transition. The Solomon echoes are satellite-transition signals. However these echoes are superimposed on the central-transition FID following the second RF pulse. In the literature, echoes observed in these conditions for $I = \frac{3}{2}$ are called quadrupole echoes. In this paper, the term "Solomon echoes" includes those observed for $I = \frac{3}{2}$.

3.4.1 *Spin* I = $\frac{3}{2}$

Numerous studies have been done on spin $I = \frac{3}{2}$ systems[78,79] but the predictions are valid only in the hard-pulse or nonselective excitation condition. That is, $H_Q^{(1)}$ has been neglected during the excitation of the spin system by the RF pulses.

3.4.1.1 Hard-pulse Excitation Figure 26(a) is the oscillogram of the one-pulse experiment from ^{81}Br

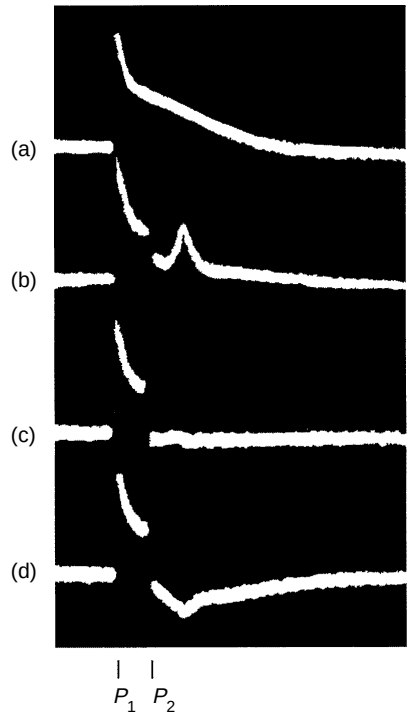

Figure 26 Oscillogram of Solomon echo from ^{81}Br ($I = \frac{3}{2}$) in KBr single crystal. The separation of the two RF pulses P_1 and P_2 is $\tau_2 = 180\,\mu s$. (a) One X pulse with $\pi/2$ pulse flip angle, i.e. $\{X\}_{\pi/2}$; (b) the sequence $\{X\}_{\pi/2}$—τ_2—$\{Y\}_{\pi/2}$; (c) the sequence $\{X\}_{\pi/2}$—τ_2—$\{X\}_{\pi/2}$; (d) the sequence $\{X\}_{\pi/2}$—τ_2—$\{X\}_{\pi/3}$. (Reproduced by permission of Pergamon Press from G. Bonera, M. Galimberti, *Solid State Commun.*, **4**, 589–591 (1966).)

($I = \frac{3}{2}$) in KBr single crystal;[78] two FIDs are clearly observed: the long one is that of the central transition, whereas the shorter one is for the two satellite transitions. Figure 26(b–d) shows the response of the spin system to three different experimental conditions of the Solomon echo sequence. The FID following the first RF pulse is that of the satellite transitions. This observation supports the condition that the pulse separation must be shorter than T_{FID}, the duration of the central-transition FID. The observed echo, which is very sharp, is the satellite-transition signal and is located at $\tau_4 = \tau_2$. The central-transition FID following the second RF pulse is not observed in these oscillograms due to the experimental conditions, which cause confusion with the Hahn echoes also located at $\tau_4 = \tau_2$ and obtained with a pulse separation $\tau_2 \geq T_{\mathrm{FID}}$ (see section 3.5).

For two in-phase RF pulses, the echo amplitude $E_{XX}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = \tau_2)$ versus the second-pulse flip angle β_3 is the expression $\frac{1}{2}(A + A')$ of Bonera and Galimberti (Equation 77):[78]

$$E_{XX}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = \tau_2) = \tfrac{3}{2}\cos^2\tfrac{1}{2}\beta_3 \sin^4\tfrac{1}{2}\beta_3 - \tfrac{3}{2}\sin^2\tfrac{1}{2}\beta_3 \cos^4\tfrac{1}{2}\beta_3 = -\tfrac{9}{8}\sin^2\beta_3 \cos\beta_3 \quad (77)$$

For two RF pulses in quadrature phase, the echo amplitude $E_{XY}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = \tau_2)$ versus the second-pulse flip angle is (Equation 78)[78]

$$E_{XY}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = \tau_2) = \tfrac{9}{8}\sin^2\beta_3 \quad (78)$$

The thin solid line in Figure 27(a) represents the opposite of Equation (77), whereas the thin dashed line in Figure 27(b) represents the opposite of Equation (78). This complication is due to the phases of the RF pulses used, namely $\{-X\}_\tau_2_\{-X\}$ for Figure 27(a) and $\{-X\}_\tau_2_\{Y\}$ for Figure 27(b). They provide a simple explanation for Figure 26(b–d). For two in-phase RF pulses, the Solomon echo is a maximum when the second-pulse flip angle is $\pi/3$ (Figure 26d and the thin solid line in Figure 27a) and zero when the second-pulse flip angle is $\pi/2$ (Figure 26c and the thin solid line in Figure 27a). However, for two RF pulses in quadrature phase, the Solomon echo is a maximum when the second-pulse flip angle is $\pi/2$ (Figure 26b and thin dashed line in Figure 27b).

3.4.1.2 Soft-pulse Excitation The limitation of the hard pulse or nonselective excitation is removed by taking into account $H_{\mathrm{Q}}^{(1)}$ during the RF pulses. A single echo is predicted at $\tau_4 = \tau_2$,[16,80] in agreement with the results of the literature. In fact this echo represents the refocusing

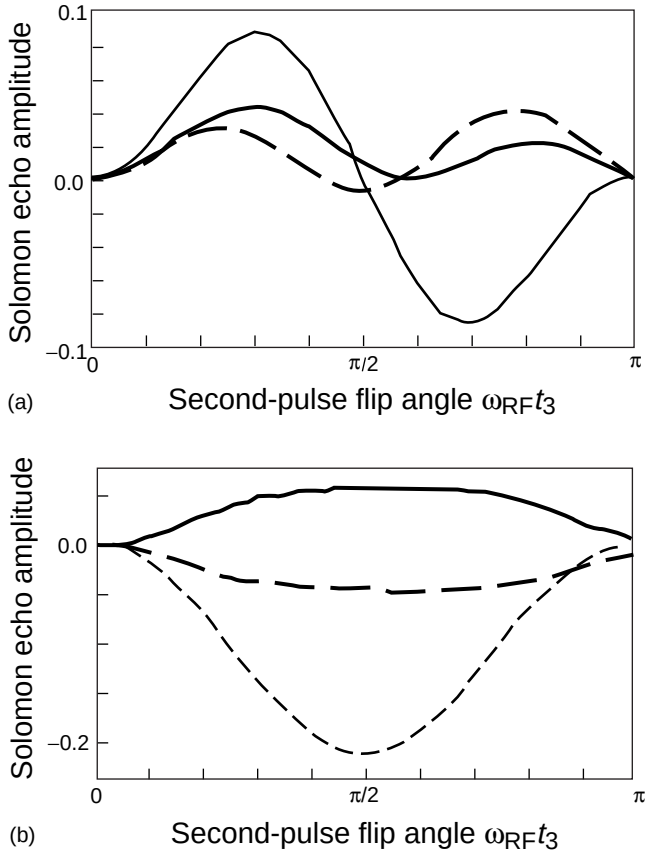

Figure 27 The $\tau_4 = \tau_2$ Solomon echo amplitudes of a spin $I = \frac{3}{2}$ versus the second-pulse flip angle $w_{\mathrm{RF}}t_3$. These echo amplitudes are associated with (a) two in-phase $(-X, -X)$ RF pulses and (b) two RF pulses in quadrature phase: their x components for the two $w_{\mathrm{Q}}/w_{\mathrm{RF}}$ ratios (thin dashed line for $w_{\mathrm{Q}}/w_{\mathrm{RF}} = 0.001$ and thick dashed line for $w_{\mathrm{Q}}/w_{\mathrm{RF}} = 1$) and their y components (thin solid line for $w_{\mathrm{Q}}/w_{\mathrm{RF}} = 0.001$ and thick solid line for $w_{\mathrm{Q}}/w_{\mathrm{RF}} = 1$). The first-pulse flip angle $w_{\mathrm{RF}}t_1$ is $\pi/2$ and $3\pi/10$ for the two $w_{\mathrm{Q}}/w_{\mathrm{RF}}$ ratios 0.001 and 1, respectively. (Reproduced by permission of American Institute of Physics from P.P. Man, *J. Chem. Phys.*, **106**, 3908–3919 (1997).)

of satellite transitions $I_{-}^{\frac{3}{2},\frac{1}{2}}(t_1)$ and $I_{+}^{-\frac{1}{2},-\frac{3}{2}}(t_1)$

and ±2Q coherences $I_{-}^{\frac{3}{2},-\frac{1}{2}}(t_1)$ and $I_{+}^{\frac{1}{2},-\frac{3}{2}}(t_1)$

generated by the first RF pulse. The contributions of ±2Q coherences to the Solomon echo were not predicted previously, because the initial state of the spin system is $\rho(0) = I_x$, whose matrix contains ±1Q coherences only.[78] Phase cycling the first RF pulse and the receiver allows us to study these two contributions independently.

Figure 27 presents the $\tau_4 = \tau_2$ Solomon echo amplitudes of a spin $I = \frac{3}{2}$ versus the second-pulse flip angle $w_{\mathrm{RF}}t_3$.[16] These echo amplitudes are associated with two in-phase RF pulses (Figure 27a) and two RF pulses

in quadrature phase (Figure 27b). The echo amplitudes decrease when the w_Q/w_{RF} ratio increases; they can be positive or negative depending on $w_{RF}t_3$ and the w_Q/w_{RF} ratio. The graphs for $w_Q/w_{RF} = 0.001$ are in agreement with Equations (77) and (78), apart from the sign as already explained above. In fact the curves in Figure 27 are plotted by using the numerical procedure described in section 3.4.3.

3.4.2 *Spin* $I = \frac{5}{2}$

The spin $I = \frac{5}{2}$ system was first studied by Solomon[13] who predicted four echoes but observed six experimentally for ^{127}I in KI. This is because he did not consider $H_Q^{(1)}$ during the RF pulses in his calculations. They are located at $\frac{3}{2}\tau$, 2τ, 2τ, $\frac{5}{2}\tau$, 3τ, and 4τ; the echo positions were referenced to the first RF pulse. The $\frac{3}{2}\tau$, 2τ, and 3τ echoes were called allowed echoes, whereas $\frac{5}{2}\tau$ and 4τ echoes were called forbidden echoes.

3.4.2.1 Hard-pulse Excitation The echo amplitudes had been determined for the allowed echoes, which are located at $\tau_4 = \frac{1}{2}\tau_2$, τ_2, and $2\tau_2$ in our notation. For two in-phase RF pulses, the echo amplitudes of the outer-satellite ($\frac{5}{2} \leftrightarrow \frac{3}{2}$) transition versus the second-pulse flip angle β_3 are given by Equations (79) and (80):[81]

$$E_{XX}^{\frac{5}{2},\frac{3}{2}}\left(\tau_4 = \tfrac{1}{2}\tau_2\right) = 20 \quad \cos^3\tfrac{1}{2}\beta_3 \sin^2\tfrac{1}{2}\beta_3 \quad \cos^5\tfrac{1}{2}\beta_3 - 4\cos^3\tfrac{1}{2}\beta_3 \sin^2\tfrac{1}{2}\beta_3 \quad - \cos^2\tfrac{1}{2}\beta_3 \times \sin^3\tfrac{1}{2}\beta_3 \quad \sin^5\tfrac{1}{2}\beta_3 - 4\cos^2\tfrac{1}{2}\beta_3 \times \sin^3\tfrac{1}{2}\beta_3 \tag{79}$$

$$E_{XX}^{\frac{5}{2},\frac{3}{2}}(\tau_4 = \tau_2) = 25 \quad \cos^2\tfrac{1}{2}\beta_3 \sin^8\tfrac{1}{2}\beta_3 - \cos^8\tfrac{1}{2}\beta_3 \times \sin^2\tfrac{1}{2}\beta_3 \tag{80}$$

Those of the inner-satellite ($\frac{3}{2} \leftrightarrow \frac{1}{2}$) transition versus the second-pulse flip angle β_3 are (Equations 81 and 82)[81]

$$E_{XX}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = \tau_2) = 64\left\{\cos^2\tfrac{1}{2}\beta_3 \sin^4\tfrac{1}{2}\beta_3 \quad \tfrac{3}{2}\cos^2\tfrac{1}{2}\beta_3 - \sin^2\tfrac{1}{2}\beta_3 \quad ^2 - \cos^4\tfrac{1}{2}\beta_3 \sin^2\tfrac{1}{2}\beta_3 \times \quad \cos^2\tfrac{1}{2}\beta_3 - \tfrac{3}{2}\sin^2\tfrac{1}{2}\beta_3 \quad ^2\right\} \tag{81}$$

$$E_{XX}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = 2\tau_2) = E_{XX}^{\frac{5}{2},\frac{3}{2}}\left(\tau_4 = \tfrac{1}{2}\tau_2\right) \tag{82}$$

Equations (79) and (82), and the sum of Equations (80) and (81) are equivalent to the terms $E_{1/2}$, E_2, and E_1 of Solomon, respectively.[13]

For two RF pulses in quadrature phase, the echo amplitudes of the outer-satellite ($\frac{5}{2} \leftrightarrow \frac{3}{2}$) transition versus the second-pulse flip angle β_3 are given by Equations (83) and (84):[81]

$$E_{XY}^{\frac{5}{2},\frac{3}{2}}\left(\tau_4 = \tfrac{1}{2}\tau_2\right) = -20 \quad \cos^3\tfrac{1}{2}\beta_3 \sin^2\tfrac{1}{2}\beta_3 \quad \cos^5\tfrac{1}{2}\beta_3 - 4\cos^3\tfrac{1}{2}\beta_3 \sin^2\tfrac{1}{2}\beta_3 \quad + \cos^2\tfrac{1}{2}\beta_3 \times \sin^3\tfrac{1}{2}\beta_3 \quad \sin^5\tfrac{1}{2}\beta_3 - 4\cos^2\tfrac{1}{2}\beta_3 \times \sin^3\tfrac{1}{2}\beta_3 \tag{83}$$

$$E_{XY}^{\frac{5}{2},\frac{3}{2}}(\tau_4 = \tau_2) = 25 \quad \cos^2\tfrac{1}{2}\beta_3 \sin^8\tfrac{1}{2}\beta_3 + \cos^8\tfrac{1}{2}\beta_3 \times \sin^2\tfrac{1}{2}\beta_3 \tag{84}$$

Those of the inner-satellite ($\frac{3}{2} \leftrightarrow \frac{1}{2}$) transition versus the second-pulse flip angle β_3 are (Equations 85 and 86)[81]

$$E_{XY}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = \tau_2) = 64\left\{\cos^2\tfrac{1}{2}\beta_3 \sin^4\tfrac{1}{2}\beta_3 \quad \tfrac{3}{2}\cos^2\tfrac{1}{2}\beta_3 - \sin^2\tfrac{1}{2}\beta_3 \quad ^2 + \cos^4\tfrac{1}{2}\beta_3 \sin^2\tfrac{1}{2}\beta_3 \times \quad \cos^2\tfrac{1}{2}\beta_3 - \tfrac{3}{2}\sin^2\tfrac{1}{2}\beta_3 \quad ^2\right\} \tag{85}$$

$$E_{XY}^{\frac{3}{2},\frac{1}{2}}(\tau_4 = 2\tau_2) = E_{XY}^{\frac{5}{2},\frac{3}{2}}\left(\tau_4 = \tfrac{1}{2}\tau_2\right) \tag{86}$$

3.4.2.2 Soft-pulse Excitation For any w_Q/w_{RF} ratio, six echoes have been predicted because the initial condition of the spin system is $\rho(0) = I_z$ instead of $\rho(0) = I_x$ as in the hard-pulse condition.[16,82] Three echoes, located at $\tau_4 = \tau_2$, $2\tau_2$, and $3\tau_2$, are inner-satellite transition ($\frac{1}{2} \leftrightarrow \frac{3}{2}$) signals:

- the $\tau_4 = \tau_2$ echo represents the refocusing of
  $$I_-^{\frac{3}{2},\frac{1}{2}}(t_1) \quad , \quad I_+^{-\frac{1}{2},-\frac{3}{2}}(t_1) \quad , \quad I_-^{\frac{3}{2},-\frac{1}{2}}(t_1) \quad , \text{and} \quad I_+^{\frac{1}{2},-\frac{3}{2}}(t_1)$$
  generated by the first RF pulse;
- the $\tau_4 = 2\tau_2$ echo represents the refocusing of
  $$I_-^{\frac{5}{2},\frac{3}{2}}(t_1) \quad , \quad I_+^{-\frac{3}{2},-\frac{5}{2}}(t_1) \quad , \quad I_-^{\frac{5}{2},-\frac{3}{2}}(t_1) \quad , \text{and} \quad I_+^{\frac{3}{2},-\frac{5}{2}}(t_1)$$

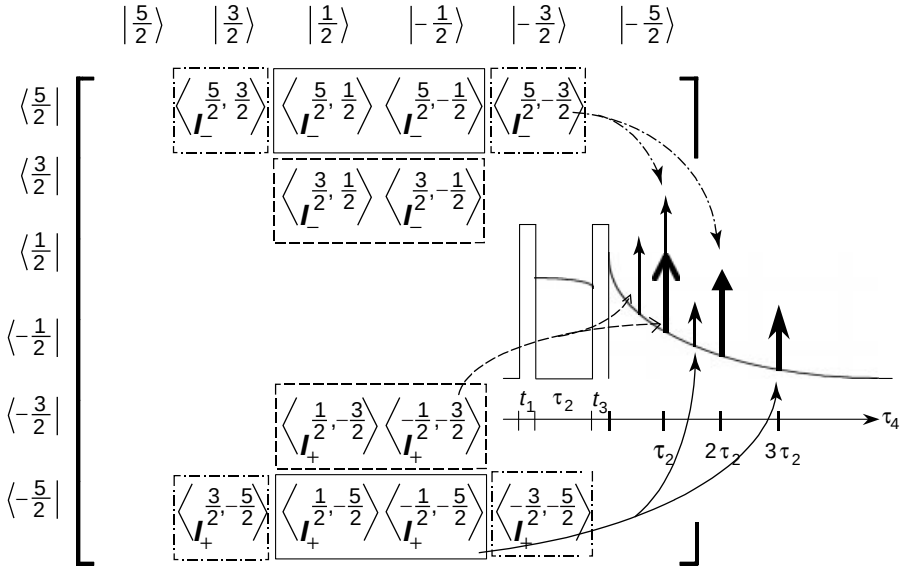

Figure 28 Coherences in the spin $I = \frac{5}{2}$ density matrix developed at the end of the first RF pulse and refocused as Solomon echoes by the second RF pulse. Echoes are represented by thick arrows for the inner-satellite transitions ($\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$) and thin arrows for the outer-satellite transitions ($\pm\frac{3}{2} \leftrightarrow \pm\frac{5}{2}$). Solid-line-framed coherences contribute to the $\tau_4 = 3\tau_2/2$ and $3\tau_2$ echoes. Dashed-line-framed coherences contribute to the $\tau_4 = \tau_2/2$ and τ_2 echoes. Dot-dashed-line-framed coherences contribute to the $\tau_4 = \tau_2$ and $2\tau_2$ echoes. Solomon echoes are superimposed on the central-transition FID following the second RF pulse. The pulse separation is much shorter than the duration T_{FID} of this FID.

- the $\tau_4 = 3\tau_2$ echo represents the refocusing of

$$I_-^{\frac{5}{2},\frac{1}{2}}(t_1) \quad , \quad I_+^{-\frac{1}{2},-\frac{5}{2}}(t_1) \quad , \quad I_-^{\frac{5}{2},-\frac{1}{2}}(t_1) \quad , \text{and}$$
$$I_+^{\frac{1}{2},-\frac{5}{2}}(t_1)$$

The other three echoes, located at $\tau_4 = \frac{1}{2}\tau_2$, τ_2, and $\frac{3}{2}\tau_2$, are outer-satellite transition ($\frac{3}{2} \leftrightarrow \frac{5}{2}$) signals; the coherences involved for each echo are given in Figure 28. In fact the 12 coherences are involved for both inner- and outer-satellite transition signals. The two forbidden echoes (the $\tau_4 = \frac{3}{2}\tau_2$ and $3\tau_2$ echoes) *solely* represent the refocusing of MQ ($\pm$2Q and $\pm$3Q) coherences generated by the first RF pulse, whereas the four allowed echoes (the $\tau_4 = \frac{1}{2}\tau_2$, τ_2, and $2\tau_2$ echoes) represent *both* the refocusing of SQ ($\pm$1Q) and MQ ($\pm$2Q and $\pm$4Q) coherences generated by the first RF pulse. The main drawback of Solomon echoes for spins $I = \frac{5}{2}$, $\frac{7}{2}$, and $\frac{9}{2}$ is that the Fourier transform of the time-domain signal generates distorted spectra. Direct analysis of the echoes in the time domain is necessary.

Figure 29 is the stacked plot of Solomon echoes from ^{127}I ($I = \frac{5}{2}$) in powdered RbI for increasing second-pulse duration.[70] Indeed six echoes are observed, superimposed on the FID following the second RF pulse. In contrast, Figure 30 is the stacked plot of Solomon echoes from ^{127}I ($I = \frac{5}{2}$) in the same sample, acquired

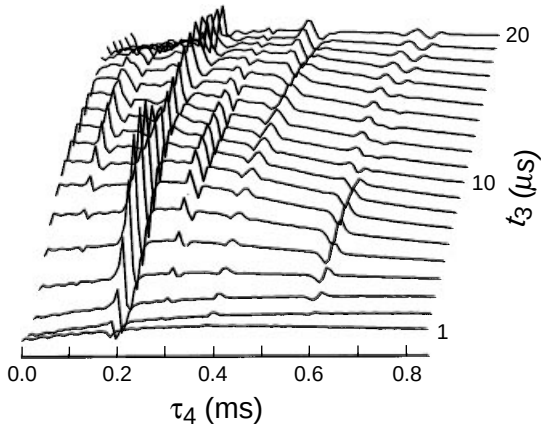

Figure 29 Stacked plot of Solomon echoes from ^{127}I ($I = \frac{5}{2}$) in powdered RbI for increasing second-pulse duration t_3 ranging from 1 to 20 μs, obtained with the $\{X\}_\tau_2_\{X\}$ sequence and a pulse separation $\tau_2 = 0.2$ ms. Echoes in the acquisition period are located at $\tau_4 = 0.1$, 0.2, 0.3, 0.4, and 0.6 ms.

with the following sequence:[70]

first RF pulse:	$X\ -X\quad Y\ -Y$	
second RF pulse:	$-X\ -X\ -X\ -X$	(sequence 2)
receiver:	$-y\ -y\quad y\quad y$	

Clearly the FID following the second RF pulse has been canceled by this sequence. However, the $\tau_4 = 2\tau_2$ echo is missing.

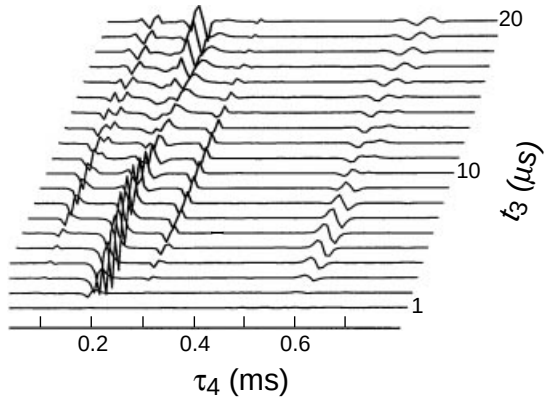

Figure 30 Cancellation of the FID following the second RF pulse in the Solomon echo sequence: stacked plot of Solomon echoes from ^{127}I ($I = \frac{5}{2}$) in powdered RbI for increasing second-pulse duration t_3 ranging from 1 to 20 μs, obtained with the phases given by sequence 2 and a pulse separation $\tau_2 = 200\,\mu s$. Echoes in the acquisition period are located at $\tau_4 = 0.1$, 0.2, 0.3, and 0.6 ms. The $\tau_4 = 0.4$ ms echo is missing.

3.4.3 Numerical Procedure

For simplicity, we propose a numerical procedure for calculating the Solomon echo amplitude versus the second-pulse duration to optimize the experimental conditions for two specific echoes of a spin $I = \frac{5}{2}$ system. The method is applicable to other echoes and to other half-integer quadrupole spins, because all the echo positions and the contribution of the coherences generated by the first RF pulse have been already defined.[16]

Figure 31 presents the numerical procedure for calculating the complex amplitude of the inner-satellite transition Solomon echo located at $\tau_4 = 2\tau_2$ and that of the outer-satellite transition located at $\tau_4 = \tau_2$ of a spin $I = \frac{5}{2}$. Equation (43), which describes the spin dynamics during the first RF pulse, is written in the matrix form. The initial condition for the first RF pulse is the Boltzmann density matrix $\rho(0) = I_z$, which has nonzero real values only along the main diagonal. We extract the four complex spectral line intensities (a_{12}, a_{15}, a_{62}, and a_{65}) and put them into another matrix whose elements are zeros. This matrix becomes the initial condition for the second RF pulse. The complex echo amplitudes $\zeta(\frac{1}{2}, \frac{3}{2})s_{32}$ and $\zeta(\frac{3}{2}, \frac{5}{2})s_{21}$ of the final matrix are that of the inner-satellite transition located at $\tau_4 = 2\tau_2$ and that of the outer-satellite transition located at $\tau_4 = \tau_2$, respectively. These two elements, s_{32} and s_{21}, are -1Q coherences detected by an RF coil. The other elements of this final matrix are meaningless from the point of view of Solomon echoes.

3.5 Hahn Echo Sequences

As shown in the Solomon echo sequence, taking only $H_Q^{(1)}$ into account throughout the experiment does not allow us to predict an echo for the central transition. As the spin–spin relaxation times are not introduced into the density matrix, the central transition, whose frequency is that of the frequency carrier, remains on resonance throughout the sequence. This suggests that the pulse separation must be much shorter than the central-transition FID duration T_{FID}, so that the magnetization of the central transition has no time to dephase during the pulse separation. However, by definition, the satellite transitions are off resonance and have time to dephase during the pulse separation and rephase during the acquisition period.

$$\begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} & a_{14} & \boxed{a_{15}} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & \boxed{a_{62}} & a_{63} & a_{64} & \boxed{a_{65}} & a_{66} \end{pmatrix} = \mathrm{Texp}(-\mathrm{i}t_1\Omega)T^{\dagger} \begin{pmatrix} \frac{5}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{5}{2} \end{pmatrix} \mathrm{Texp}(\mathrm{i}t_1\Omega)T^{\dagger}$$

$$\begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ \textcircled{s_{21}} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & \textcircled{s_{32}} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{pmatrix} = \mathrm{Texp}(-\mathrm{i}t_3\Omega)T^{\dagger} \begin{pmatrix} 0 & a_{12} & 0 & 0 & a_{15} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{62} & 0 & 0 & a_{65} & 0 \end{pmatrix} \mathrm{Texp}(\mathrm{i}t_3\Omega)T^{\dagger}$$

Figure 31 Numerical procedure for calculating the complex amplitude (s_{32}) of the inner-satellite transition ($\frac{1}{2} \leftrightarrow \frac{3}{2}$) Solomon echo located at $\tau_4 = 2\tau_2$ and that (s_{21}) of the outer-satellite transition ($\frac{3}{2} \leftrightarrow \frac{5}{2}$) Solomon echo located at $\tau_4 = \tau_2$ of a spin $I = \frac{5}{2}$ under static conditions. The initial condition for the first RF pulse is the Boltzmann density matrix containing nonzero but real diagonal elements. The initial condition for the second RF pulse is the matrix containing only four nonzero but complex elements (a_{12}, a_{15}, a_{62}, and a_{65}) provided by the density matrix of the first RF pulse. T and Ω are the eigenvector and eigenvalue matrices of the interactions involved during the pulses, namely the RF pulse and $H_Q^{(1)}$. The symbol † means transposed and complex conjugated. The two subscripts of the elements of a and s have their usual meanings, that is, the row and column of the matrix element.

To predict an echo for the central transition for pulse separation longer than T_{FID}, we include, during the free precession of the spin system, the secular part of the heteronuclear magnetic dipole–dipole interaction (Equations 87 and 88):[17]

$$H_{\mathrm{D(IS)}} = \Phi I_z \tag{87}$$

$$\Phi = \frac{\mu_0}{4\pi} \frac{\gamma_I \gamma_S \hbar S_z}{2(r_{I-S})^3}(1 - 3\cos^2\theta_{I-S}) \tag{88}$$

This is instead of the inhomogeneity of $\mathbf{B}_0$,[78,79] because the experiment is performed with a superconducting magnetic field whose homogeneity is very good for solid-state NMR. The interactions considered in this section are given in Figure 9(b). Only $H_{\mathrm{Q}}^{(1)}$ is considered during the RF pulses, an assumption that requires a strong RF pulse amplitude.

3.5.1 Hard-pulse Excitation

The $\tau_4 = \tau_2$ Hahn echo amplitude for the hard-pulse or nonselective excitation ($w_{\mathrm{Q}} \ll w_{\mathrm{RF}}$) was calculated by Mehring and Kanert.[83] Another calculation was performed by Suemitsu and Nakajo.[84] In these papers, two Y pulses were applied and the first RF pulse was a $\pi/2$ pulse so that the initial condition is $\rho(0) = I_x$. Their conventions differ with ours but the results concerning two in-phase RF pulses do not depend on this phase. A bell-like-shape echo associated with the transition $(m+1 \leftrightarrow m)$ is along the x axis and its amplitude is (Equations 89 and 90)

$$E_{\mathrm{NS}}(\tau_4 = \tau_2) = C_{\mathrm{NS}}(m, I) \sin w_{\mathrm{RF}} t_1 \tag{89}$$

$$C_{\mathrm{NS}}(m, I) = -\frac{3}{2}\frac{\zeta^2}{I(I+1)(2I+1)} \times \{d^{(I)}_{m,m+1}(w_{\mathrm{RF}} t_3)\}^2 \tag{90}$$

where $d^{(I)}_{m,m+1}(w_{\mathrm{RF}} t_3)$ is the reduced Wigner rotation matrix of order I. The two subscripts m and $m+1$ locate the matrix element. Equation (90) shows that the echo has a negative amplitude. As $C_{\mathrm{NS}}(m, I)$ is an even function of $w_{\mathrm{RF}} t_3$, alternating the phase of the second RF pulse does not change the echo amplitude $E_{\mathrm{NS}}(\tau_4 = \tau_2)$. For short duration of the two RF pulses, Equation (89) becomes Equation (91):[51]

$$E_{\mathrm{NS}}(\tau_4 = \tau_2) = -\frac{3}{2}\frac{\zeta^2}{I(I+1)(2I+1)} w_{\mathrm{RF}} t_1 \left\{\frac{1}{2}\zeta w_{\mathrm{RF}} t_3\right\}^2 \tag{91}$$

In fact the Hahn echo amplitude is the product of two functions:[17] the first function, called the excitation function, depends on the first-pulse duration; the second function, called the transfer or conversion function, depends on the second-pulse duration.

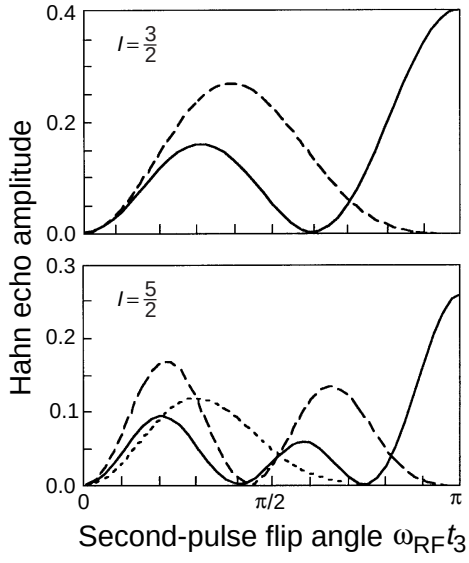

Figure 32 Graphs of the transfer function of the $\tau_4 = \tau_2$ Hahn echo amplitudes versus the second-pulse flip angle $w_{\mathrm{RF}} t_3$ for the spins $I = \frac{3}{2}$ and $\frac{5}{2}$ in the hard-pulse or non-selective excitation condition ($w_{\mathrm{Q}} \ll w_{\mathrm{RF}}$): solid line for the central-transition ($-\frac{1}{2} \leftrightarrow \frac{1}{2}$) echo; dashed line for the satellite-transition ($\pm\frac{1}{2} \leftrightarrow \pm\frac{3}{2}$) echoes; dotted line for the satellite-transition ($\pm\frac{3}{2} \leftrightarrow \pm\frac{5}{2}$) echoes. They are obtained with the numerical procedure described in section 3.5.5 or Figure 31. (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **52**, 9418–9426 (1995).)

Figure 32 shows the graphs of the transfer function of the $\tau_4 = \tau_2$ Hahn echo amplitudes versus the second-pulse flip angle for the spins $I = \frac{3}{2}$ and $\frac{5}{2}$,[17,83] obtained by using the numerical procedure described in section 3.5.5. Consider the case of $I = \frac{3}{2}$. The central-transition echo has two maximum and two zero values. The satellite-transition echoes have one maximum and two zero values. All the echoes are observed when the second-pulse flip angle is smaller than $\pi/2$. For larger flip angles, mainly the central-transition echo is observed.

Figure 33 shows the Hahn echoes of ^{37}Cl ($I = \frac{3}{2}$) in a single crystal of NaCl. The echoes of the central transition as well as those of the satellite transitions are observed because the shapes of the echoes vary with the second-pulse duration.[85] These experimental data are in agreement with the predictions (see Figure 32). Despite the cubic structure of NaCl, defects in the structure generate small electric field gradients, whose quadrupole couplings are much smaller than the RF pulse amplitude.

In contrast, Figure 34 shows the Hahn echoes of ^{23}Na ($I = \frac{3}{2}$) in the same single crystal of NaCl.[85] The behavior of these echoes differs from those of chlorine. All the echoes have the same shape. Figure 34

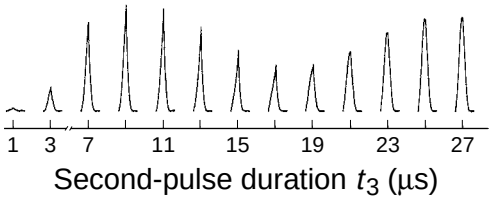

Figure 33 Experimental ^{37}Cl ($I = \frac{3}{2}$) $\tau_4 = \tau_2$ Hahn echo (in the time domain) of a single crystal of NaCl for $t_3 = 1$ to 27 μs by steps of 2 μs. These echoes are acquired with the $\{X\}_\tau_2_\{\pm X\}$ sequence; $\tau_2 = 1$ ms, $D_0 = 100$ s, and $T_{\text{FID}} = 1$ ms. The echo for $t_3 = 5$ μs is missing.

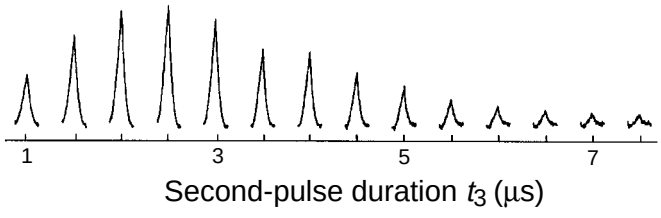

Figure 34 Experimental ^{23}Na ($I = \frac{3}{2}$) $\tau_4 = \tau_2$ Hahn echo (in the time domain) of a single crystal of NaCl for $t_3 = 1$ to 7.5 μs by steps of 0.5 μs. These echoes are acquired with the $\{X\}_\tau_2_\{\pm X\}$ sequence; $\tau_2 = 400$ μs, $D_0 = 100$ s, and $T_{\text{FID}} = 400$ μs.

is similar to that observed by Flett and Richard[86] and looks very similar to that of the satellite-transition echoes (Figure 32, dashed line in the upper figure). This discrepancy is due to the strength of the homonuclear magnetic dipole–dipole couplings with respect to that of the heteronuclear case. As the gyromagnetic ratio of the sodium nucleus is three times larger than for the chlorine nucleus, the homonuclear magnetic dipole–dipole coupling between sodium is much stronger than the heteronuclear magnetic dipole–dipole couplings between sodium and the two isotopes of chlorine. In other words, during free precession the dominant interactions are not those of Figure 9(b) for sodium. However, the heteronuclear magnetic dipole–dipole coupling of chlorine is much stronger than the homonuclear magnetic dipole–dipole coupling. Therefore chlorine is submitted to the interactions of Figure 9(b).

3.5.2 Selective Excitation

For selective excitation ($w_{\text{Q}} \gg w_{\text{RF}}$) the maximum of the echo amplitude is given by Equations (92) and (93):[51]

$$E_s(\tau_4 = \tau_2) = C_s(m, I) \sin \zeta w_{\text{RF}} t_1 \tag{92}$$

$$C_s(m, I) = -\frac{3}{2} \frac{\zeta}{I(I+1)(2I+1)} \sin^2 \frac{1}{2} \zeta w_{\text{RF}} t_3 \tag{93}$$

For a spin $I = \frac{1}{2}$ system, $\zeta = 1$ (see Table 1), and Equation (92) reduces to Equation (94):

$$E_s(\tau_4 = \tau_2) = -\sin w_{\text{RF}} t_1 \sin^2 \tfrac{1}{2} w_{\text{RF}} t_3 \tag{94}$$

The subscript s in Equation (94) is meaningless in the present case of $I = \frac{1}{2}$, because the latter has only two energy levels. Figure 35 is the oscillogram of proton ^{1}H ($I = \frac{1}{2}$) in glycerine showing an FID following the first RF pulse, an FID following the second RF pulse and the Hahn echo.[14] The latter two signals will be separated if the pulse separation is larger than $2T_{\text{FID}}$. The Hahn echo is due to the inhomogeneity of conventional iron $\mathbf{B}_0$. With modern spectrometers, alternating the phase of the second RF pulse will cancel the FID following the second RF pulse. As a result, the pulse separation can be reduced from $2T_{\text{FID}}$ to T_{FID} – the full echo is still observed but with larger amplitude than that observed when the pulse separation is $2T_{\text{FID}}$. These experimental conditions remain valid for half-integer quadrupole spins. Figure 36

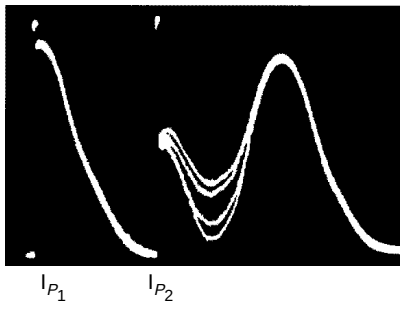

Figure 35 Proton ^{1}H ($I = \frac{1}{2}$) oscillogram in glycerine showing random interference of the FID following the second RF pulse with the $\tau_4 = \tau_2$ Hahn echo for several exposures. The two RF pulses P_1 and P_2 are phase incoherent relative to each other. (Reproduced by permission of The American Physical Society from E.L. Hahn, *Phys. Rev.*, **80**, 580–594 (1950).)

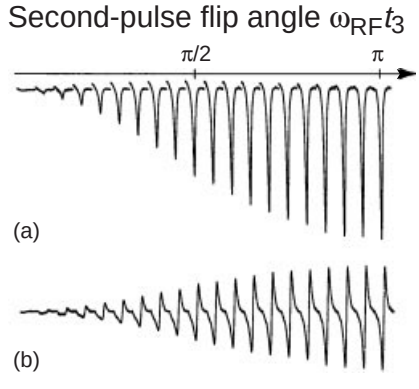

Figure 36 (a) Imaginary-part and (b) real-part echoes (in the time domain) of ^{23}Na in an aqueous solution of NaCl versus the second-pulse flip angle $w_{\text{RF}} t_3$. The echoes are acquired with the $\{X\}_{\pi/2}_\tau_2_\{X\}$ sequence.

shows the imaginary and real parts of the ^{23}Na echo in an aqueous solution of NaCl versus the second-pulse flip angle.[70] The homogeneity of the superconducting $\mathbf{B}_0$ was reduced by using the z shim. The echo amplitude increases with the second-pulse flip angle, in agreement with Equation (94).

Figure 37 is the ^{87}Rb ($I = \frac{3}{2}$) Hahn echo in powdered pyrochlore $RbNb_2O_5F$.[85] In this compound the EFG around the rubidium atoms is very strong. The variation of the echo amplitude versus the second-pulse duration is described by Equation (92).

As $C_s(m, I)$ is also an even function of $w_{RF}t_3$, alternating the phase of the second RF pulse does not change the echo amplitude $E_s(\tau_4 = \tau_2)$. As alternating the phase of the second RF pulse does not change the echo amplitude in both nonselective and selective excitations, it must be true for any excitation condition. Indeed analytical results on spin $I = \frac{3}{2}$,[85,87] $\frac{5}{2}$,[88,89] and $\frac{7}{2}$[45] support this observation.

If both pulse durations are short Equation (95) holds:

$$E_s(\tau_4 = \tau_2) = E_{NS}(\tau_4 = \tau_2) = -\frac{3}{2I(I+1)(2I+1)} w_{RF}t_1 \left(\tfrac{1}{2}\zeta^2 w_{RF}t_3\right)^2 \quad (95)$$

Therefore the echo amplitude is independent of w_Q but depends *linearly* on the first-pulse duration and *quadratically* on the second-pulse duration.[51] This excitation condition is useful for spin population determination in single crystals as well as in powders.

3.5.3 *Soft-pulse Excitation for Spin* I $= \frac{3}{2}$*: Piezoelectric Signal Suppression*

The secular part of the heteronuclear magnetic dipole–dipole interaction during free precession of the spin system and the first-order quadrupole interaction throughout the experiment are now considered (Figure 9b). For a spin $I = \frac{3}{2}$ system, three Hahn echoes are predicted and observed: two central-transition echoes located at $\tau_4 = \tau_2$ and $\tau_4 = 3\tau_2$, and the satellite-transition echoes located at $\tau_4 = \tau_2$.[17,88] The $\tau_4 = \tau_2$ Hahn echoes are the refocusing of 1Q coherences generated by the first RF pulse, namely

$$-\mathrm{i}\ I_y^{\frac{1}{2},-\frac{1}{2}}(t_1) \quad , \quad I_-^{\frac{3}{2},\frac{1}{2}}(t_1) \quad , \text{and} \quad I_-^{-\frac{3}{2},-\frac{1}{2}}(t_1)$$

The $\tau_4 = 3\tau_2$ echo is the refocusing of the 3Q coherence generated by the first RF pulse,

$$-\mathrm{i}\ I_y^{\frac{3}{2},-\frac{3}{2}}(t_1)$$

Figure 38 is the oscillogram of the ^{53}Cr ($I = \frac{3}{2}$) echoes observed in $CdCr_2Se_4$,[90] where the positions of the two echoes are referenced to the first RF pulse, that is, 2τ and 4τ instead of τ_2 and $3\tau_2$.

Figure 39 presents the graphs of the transfer function of the central transition versus the second-pulse flip angle for several w_Q/w_{RF} ratios.[17] They are obtained by using the procedure described in section 3.5.5. The graph for $w_Q/w_{RF} = 0$ is in agreement with that of Mehring and Kanert[83] or Equation (90). The graph for $w_Q/w_{RF} = 20$ is in agreement with Equation (93).

The piezoelectric signals are observed when a ferroelectric material[91] is studied in single crystal form. In the past several methods have been used for attenuating these piezoelectric signals. The simplest but least interesting method is to grind the single crystal into powder so that the centers of gravity of the positive and negative charges are merged. Alternatively, the single crystal can

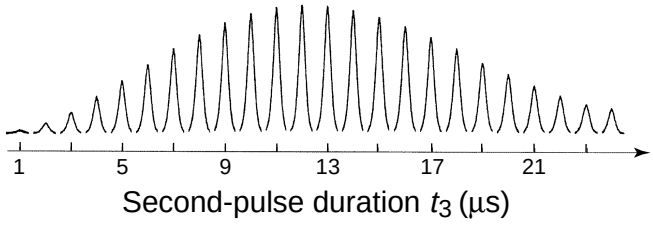

Figure 37 Experimental ^{87}Rb ($I = \frac{3}{2}$) $\tau_4 = \tau_2$ Hahn echo (in the time domain) of powdered pyrochlore $RbNb_2O_5F$ for $t_3 = 1$ to 24 μs by steps of 1 μs. These echoes are acquired with the $\{X\}_\tau_2_\{\pm X\}$ sequence; $t_1 = 6$ μs, $\tau_2 = 300$ μs, $D_0 = 1$ s, and $T_{FID} = 300$ μs. The RF pulse amplitude is $w_{RF}/(2\pi) = 21$ kHz, corresponding to a $\pi/2$ pulse duration of 12 μs.

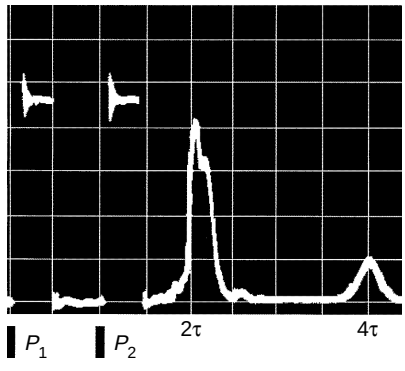

Figure 38 Oscillogram echo signals at times 2τ and 4τ of ^{53}Cr ($I = \frac{3}{2}$) in $CdCr_2Se_4$. The pulse separation is $\tau = 20$ μs. The horizontal scale is 10 μs per division. The positions of the echoes are referenced to the first RF pulse P_1. In fact they correspond to $\tau_4 = \tau_2$ and $3\tau_2$ in our notation where echo positions are referenced to the second RF pulse P_2. (Reprinted from G.N. Abelyashev, V.N. Berzhanskij, N.A. Sergeev, Y.V. Fedotov, 'Multiquantum Effects and NMR in Magnetically Ordered Substances', *Phys. Lett. A.*, **133**, 263–265, Copyright (1988) with permission from Elsevier Science.)

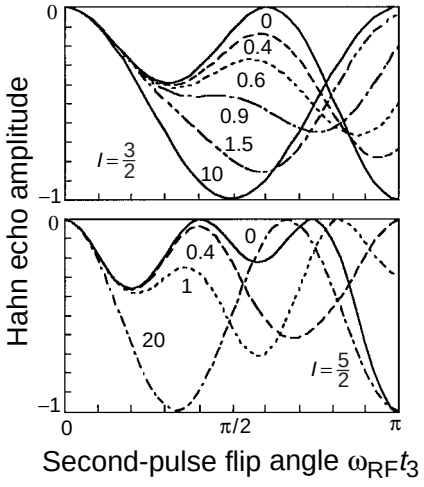

Figure 39 Graphs (for spins $I = \frac{3}{2}$ and $\frac{5}{2}$) of the transfer function of the central-transition Hahn echo amplitude versus the second-pulse flip angle $w_{RF}t_3$ for several w_Q/w_{RF} ratios (0, 0.4, 0.6, 0.9, 1, 1.5, 10, and 20). (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **52**, 9418–9426 (1995).)

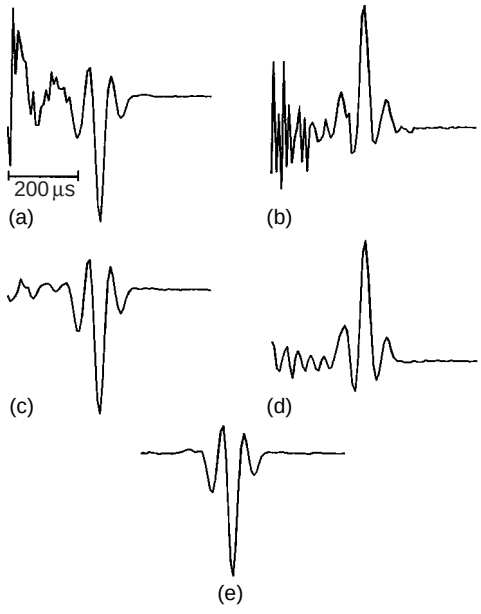

Figure 40 Suppression of the piezoelectric signals from a single crystal by using Hahn echo sequences. ^{7}Li ($I = \frac{3}{2}$) Hahn echoes (in the time domain) of a single crystal of $LiNbO_3$ acquired with the following sequences: (a) $\{X\}_\tau_2_\{X\}_$acquisition(y); (b) $\{X\}_\tau_2_\{Y\}_$acquisition(y); (c) $\{X\}_\tau_2_\{\pm X\}_$acquisition(y); (d) $\{X\}_\tau_2_\{\pm Y\}_$acquisition(y); (e) is sequence (c) followed by $\{X\}_\tau_2_\{\pm Y\}_$acquisition($-y$); $\pm y$ is the receiver phase; $D_0 = 100\,\text{s}$ and $\tau_2 = 300\,\mu\text{s}$. (Reprinted from P.P. Man, 'Study of a Spin-$\frac{3}{2}$ System by a Quadrupolar-echo Sequence: Suppression of Spurious Signals', *Solid State NMR*, **1**, 149–158, Copyright (1992), with permission from Elsevier Science.)

be put into oil, CCl_4 or paraffin to attenuate the piezoelectric signals. Shielding the sample with an extra coil inside the RF coil is also effective.[91]

Hahn echo sequences were applied to cancel these spurious signals. Five sequences were tested for the proposals. Figure 40 presents the efficiency of these sequences.[92] Clearly sequence (e) gives the best result. Sequence (a), which consists of two RF pulses of the same phase, and sequence (c), which is sequence (a) where the phase of the second RF pulse alternates, provide negative echoes. Sequence (b), which consists of two RF pulses in quadrature phase, and sequence (d), which is sequence (b) where the phase of the second RF pulse alternates, provide positive echoes. Sequence (e) is sequence (c) followed by sequence (d) whose receiver phase has opposite phase to make the echo negative so that the echoes co-add.[93]

first RF pulse:	$+X$	$+X$	$+X$	$+X$	
second RF pulse:	$+X$	$-X$	$+Y$	$-Y$	(sequence e)
receiver:	$+x$	$+x$	$-x$	$-x$	

The efficiency of sequence (e) to cancel the piezoelectric signals does not depend on the two pulse durations.[94] In fact this sequence also cancels the spurious signals generated by the NMR probe head. We have used this sequence to observe ^{131}Xe ($I = \frac{3}{2}$) physisorbed in NaY zeolite (Figure 41).[94]

3.5.4 Soft-pulse Excitation for Spin $I = \frac{5}{2}$

Figure 42 shows the coherences developed at the end of the first RF pulse and refocused as Hahn echoes by the second RF pulse.[17] Six echoes have been predicted: three as the central-transition echoes located at $\tau_4 = \tau_2$, $3\tau_2$, and $5\tau_2$; two as the inner-satellite transition echoes located at $\tau_4 = \tau_2$ and $3\tau_2$; and one echo as the outer-satellite transition echo located at $\tau_4 = \tau_2$. These echoes are the refocusing of

$$-\text{i}\ \ I_y^{\frac{1}{2},-\frac{1}{2}}(t_1)\ \ ,\ \ I_-^{\frac{3}{2},\frac{1}{2}}(t_1)\ \ ,\ \ I_-^{\frac{5}{2},\frac{3}{2}}(t_1)\ \ ,$$

$$-\text{i}\ \ I_y^{\frac{3}{2},-\frac{3}{2}}(t_1)\ \ ,\ \ I_-^{\frac{5}{2},-\frac{1}{2}}(t_1)\ \ ,\text{and}\ -\text{i}\ \ I_y^{\frac{5}{2},-\frac{5}{2}}(t_1)$$

respectively. In practice, only the $\tau_4 = \tau_2$ Hahn echoes are observed.

Figure 43 presents the one-pulse spectra of ^{27}Al ($I = \frac{5}{2}$) in a polycrystalline sample of $KAl(SO_4)_2 \cdot 12H_2O$;[89] the satellite-transition powder patterns are distorted by the dead-time of the receiver. However, Figure 44 presents the Hahn echo spectra of ^{27}Al $I = \frac{5}{2}$ in the same sample;[89] the satellite-transition powder patterns are reconstructed only when the second-pulse duration is short. For longer second-pulse durations, the powder patterns differ from what is expected. In this compound, an aluminum atom is hexacoordinated to six water

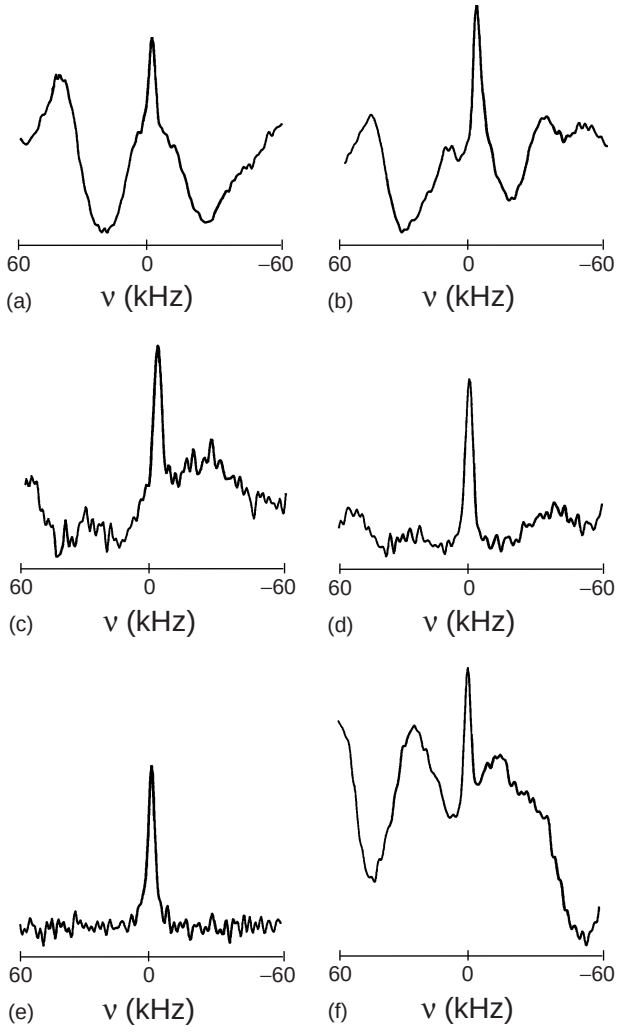

Figure 41 Suppression of the ringing signals from the NMR probe head by using Hahn echo sequences: ^{131}Xe ($I = \frac{3}{2}$) spectra of Xe gas physisorbed in NaY zeolite, acquired with one-pulse sequence (f) and Hahn echo sequences (a–e): sequences used for acquiring (a–e) are identical to those of Figure 40. (Reproduced by permission of Gordon and Breach, Harwood Academic, Yverdon Switzerland, from M.A. Hepp, P.P. Man, J. Fraissard, *Application of NMR Spectroscopy to Cement Science*, eds. P. Colombet, A.R. Grimmer, 455–466 (1994).)

molecules. The heteronuclear magnetic dipole–dipole interaction of aluminum with 12 protons is much larger than the homonuclear magnetic dipole–dipole interaction of aluminum with its surrounding aluminum atoms. In other words, the Hamiltonians (Figure 9b) required for the formation of Hahn echoes are verified for aluminum atoms in this compound.

3.5.5 Numerical Procedure

The Hahn echo amplitude is the product of two functions:[17] the first function, called the excitation function, depends on the first-pulse duration t_1; the second function, called the transfer or conversion function, depends on the second-pulse duration t_3. The positions of the central- and satellite-transition Hahn echoes and the coherences generated at the end of the first RF pulse and involved in the Hahn echo amplitude have been established for the four half-integer quadrupole spins.[17]

Figure 45 presents the numerical procedure for calculating the $\tau_4 = \tau_2$ Hahn echo complex amplitude (s_{43}) of the central transition in a spin $I = \frac{5}{2}$ system. The first part of this procedure concerning the first RF pulse is identical to that proposed for Solomon echoes (Figure 31). However, only one coherence or complex spectral line intensity,

$$a_{34} = -\mathrm{i} \quad I_{\mathrm{y}}^{\frac{1}{2},-\frac{1}{2}}(t_1)$$

contributes to the initial density matrix for the second RF pulse to calculate the transfer function, which is defined by $\zeta\left(\frac{1}{2}, -\frac{1}{2}\right)\mathrm{Im}(s_{43})/\mathrm{Im}(a_{34})$, where Im means the imaginary part. The other elements of the matrix s are meaningless. Figure 39 shows the graphs of the transfer function for spin $I = \frac{3}{2}$ and $\frac{5}{2}$, obtained with the numerical procedure; one important feature of these transfer functions is that they are always positive or negative, whereas Solomon echoes change sign with the second-pulse duration. For short pulse duration t_3, the transfer function is independent of the quadrupole coupling w_{Q} and varies quadratically with t_3 (see Equation 95).

3.5.6 Any Spin I *in a Static Sample*

The interactions considered during free precession define the nature and position of the echoes, whereas the interactions considered during the RF pulses affect the echo amplitude. However, the powder patterns are affected by the two pulse durations.

In particular, when $H_{\mathrm{Q}}^{(2)}$ is taken into account during free precession (Figure 9c), the secular term of the heteronuclear magnetic dipole–dipole interaction need not be considered because $H_{\mathrm{Q}}^{(2)}$ is an odd function of I_z (Equation 7). In other words, $H_{\mathrm{Q}}^{(2)}$ contains a term I_z (Equation 9) that allows the dephasing and the refocusing of the central-transition coherence. Dumazy and colleagues[95] have shown, in the case of $I = \frac{3}{2}$, that only the $\tau_4 = \tau_2$ Hahn echoes are observable.

So far homonuclear magnetic dipole–dipole interaction has been neglected. In fact this interaction, present in a static sample, generates other echoes,[96] which is why the samples used to illustrate theoretical results are so specific: chlorine in NaCl, rubidium in

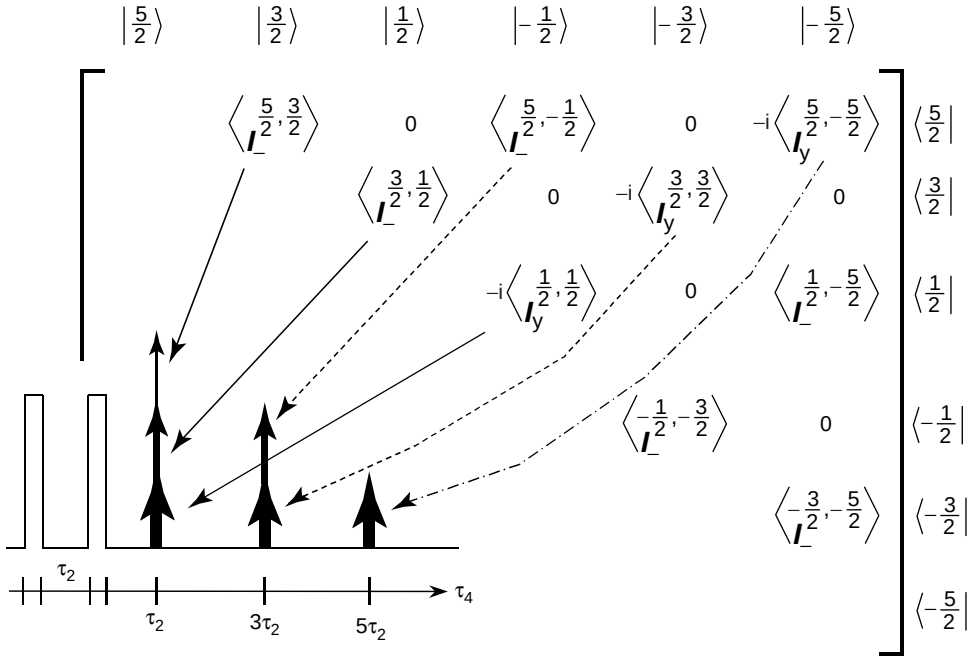

Figure 42 Coherences of the spin $I = \frac{5}{2}$ density matrix developed at the end of the first RF pulse and refocused as Hahn echoes by the second RF pulse. Echoes are represented by thick arrows for the central transition ($-\frac{1}{2} \leftrightarrow \frac{1}{2}$), medium arrows for the inner-satellite transition ($\frac{1}{2} \leftrightarrow \frac{3}{2}$), and a thin arrow for the outer-satellite transition ($\frac{3}{2} \leftrightarrow \frac{5}{2}$). Coherences and corresponding echoes located at $\tau_4 = \tau_2$, $3\tau_2$, and $5\tau_2$ are connected by arrows (solid line for 1Q, dashed line for 3Q, and dot-dashed line for 5Q). (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **52**, 9418–9426 (1995).)

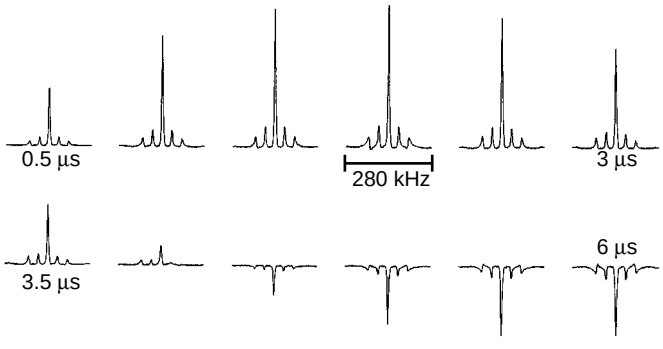

Figure 43 Effect of the dead-time of the receiver on the satellite-transition powder pattern: one-pulse spectra of ^{27}Al ($I = \frac{5}{2}$) in a polycrystalline sample of $KAl(SO_4)_2 \cdot 12H_2O$. The pulse duration t_1 increases from 0.5 to 6 µs by steps of 0.5 µs. (Reprinted from P.P. Man, E. Duprey, J. Fraissard, P. Tougne, J.-B. d'Espinose, 'Spin-$\frac{5}{2}$ Hahn Echoes in Solids', *Solid State NMR*, **5**, 181–188, Copyright (1995), with permission from Elsevier Science.)

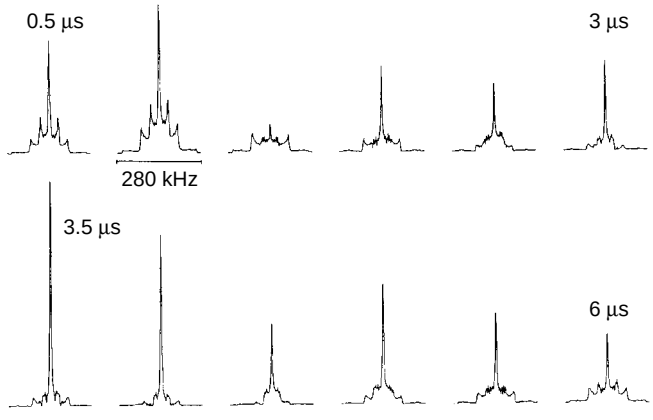

Figure 44 Effect of the second-pulse duration on the spectral line intensity and the shape of the spectrum, Fourier transform of half of the $\tau_4 = \tau_2$ Hahn echo: the $\tau_4 = \tau_2$ Hahn echo spectra of ^{27}Al ($I = \frac{5}{2}$) in a polycrystalline sample of $KAl(SO_4)_2 \cdot 12H_2O$, acquired with the sequence $\{X\}_\tau_2_\{\pm X\}$, $\tau_2 = 160$ µs and $D_0 = 1$ s; $T_{FID} = 200$ µs. The second-pulse duration t_3 increases from 0.5 to 6 µs by steps of 0.5 µs. (Reprinted from P.P. Man, E. Duprey, J. Fraissard, P. Tougne, J.-B. d'Espinose, 'Spin-$\frac{5}{2}$ Hahn Echoes in Solids', *Solid State NMR*, **5**, 181–188, Copyright (1995), with permission from Elsevier Science.)

the pyrochlore $RbNb_2O_5F$, or aluminum in $KAl(SO_4)_2 \cdot 12H_2O$.

3.5.7 Any Spin I *in a Fast-rotating Sample*

In this and the following section, the sample rotates at the magic angle θ_m under rapid conditions during the two-pulse experiment. The following interactions (Equations 96 and 97) are taken into account to predict the spin dynamics (Figure 9d):

$$H_{int} = H_{CS} + H_Q^{(2)\text{fast MAS}} \tag{96}$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & \boxed{a_{34}} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{pmatrix} = \mathrm{Texp}(-\mathrm{i}t_1\Omega)T^{\dagger} \begin{pmatrix} \frac{5}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{5}{2} \end{pmatrix} \mathrm{Texp}(\mathrm{i}t_1\Omega)T^{\dagger}$$

$$\begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & \circled{s_{43}} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{pmatrix} = \mathrm{Texp}(-\mathrm{i}t_3\Omega)T^{\dagger} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{34} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathrm{Texp}(\mathrm{i}t_3\Omega)T^{\dagger}$$

Figure 45 Numerical procedure for calculating the complex amplitude s_{43} of the central-transition $\tau_4 = \tau_2$ Hahn echo of a spin $I = \frac{5}{2}$, valid for static condition whatever the pulse durations or fast MAS condition with short pulse durations. The initial condition for the first RF pulse is the Boltzmann density matrix containing nonzero but real diagonal elements. The initial condition for the second RF pulse is the matrix containing only one nonzero but complex element, a_{34}, provided by the density matrix of the first RF pulse. The other parameters are identical to those of Figure 31.

during the free precession of the spin system, because $H_{\mathrm{Q}}^{(1)} = 0$ in the fast condition; and

$$H = H_{\mathrm{Q}}^{(1)} + H_{\mathrm{RF}} \tag{97}$$

during the RF pulses. The static Hamiltonian is considered during the RF pulses, which means that the pulse durations must be short enough so that the rotor appears static during the excitation of the spin system.

3.5.7.1 p-*Quantum Line Shift* In theoretical study, the isotropic chemical shift $\delta_{\mathrm{CS}}^{(\mathrm{iso})}$ of a line located at w_{L} is referenced relative to w_{c}. The contribution of H_{CS} (Equation 38) to the absorption line position with respect to w_{c} is defined by Equation (98):

$$\begin{aligned} w_{r,c}^{(\mathrm{iso})} &= \langle r|H_{\mathrm{CS}}|r\rangle - \langle c|H_{\mathrm{CS}}|c\rangle = -(r-c)w_{\mathrm{c}}\delta_{\mathrm{CS}}^{(\mathrm{iso})} \\ &= -(r-c)(w_{\mathrm{L}} - w_{\mathrm{c}}) \end{aligned} \tag{98}$$

The analytical expression of $H_{\mathrm{Q}}^{(2)\text{fast MAS}}$ is unknown but is defined by the second-order quadrupole shift $w_{r,c}^{(2)\text{fast MAS}}$ of the $(r-c)$-quantum absorption line relative to w_{L} (Equation 99):

$$w_{r,c}^{(2)\text{fast MAS}} = \langle r|H_{\mathrm{Q}}^{(2)\text{fast MAS}}|r\rangle - \langle c|H_{\mathrm{Q}}^{(2)\text{fast MAS}}|c\rangle \tag{99}$$

We confine ourselves to investigate the on-resonance symmetrical coherences. On-resonance means "in the second main diagonal of the density matrix". Symmetrical means "connecting $|r\rangle$ with $|-r\rangle$", that is, $c = -r$. In this case, the coherence order is $p = 2r$ and[18]

$$\begin{aligned} w_{p/2,-p/2}^{(2)\text{fast MAS}} = -\frac{\Omega_{\mathrm{Q}}^2}{w_{\mathrm{L}}} \quad & \tfrac{1}{2}C_0(I,p)B_{00}(\eta) + \tfrac{1}{2}C_2(I,p) \\ & \times \{B_{40}(\eta)d_{0,0}^{(4)}(\beta_1) + 2B_{42}(\eta)d_{2,0}^{(4)}(\beta_1)\cos 2\alpha_1 \\ & + 2B_{44}(\eta)d_{4,0}^{(4)}(\beta_1)\cos 4\alpha_1\}P_4(\cos\theta_{\mathrm{m}}) \\ &= w_{p/2,-p/2}^{(2)\mathrm{iso}}(I,p) - k(I,p)\xi_{-1/2,1/2} \\ &= \lambda(I,p)w_{-1/2,1/2}^{(2)\mathrm{iso}} - k(I,p)\xi_{-1/2,1/2} \end{aligned} \tag{100}$$

The parameters $C_0(I,p)$, $C_2(I,p)$, $\lambda(I,p)$, and $k(I,p)$, used in Equation (100) and given values in Table 2, are defined in Equations (101–104):

$$C_0(I,p) = p\left\{I(I+1) - \tfrac{3}{4}p^2\right\} \tag{101}$$

$$C_2(I,p) = p\left\{18I(I+1) - \tfrac{17}{2}p^2 - 5\right\} \tag{102}$$

$$\lambda(I,p) = \frac{C_0(I,p)}{C_0(I,-1)} \tag{103}$$

$$k(I,p) = -\frac{C_2(I,p)}{C_2(I,-1)} \tag{104}$$

The other parameters in Equation (100) have been defined previously.[18] The function $\xi_{-1/2,1/2}$ depends on orientation parameters. The second-order quadrupole shift of the center of gravity of the central-transition

Table 2 Parameters depending on the spin I and the coherence order p. $C_0(I,p)$ is defined by Equation (101) and $C_2(I,p)$ by Equation (102). The parameter $\lambda(I,p)$, given by Equation (103), is the ratio of the second-order quadrupole shift of the center of gravity of a spectrum generated by a pQ coherence to that of the -1Q coherence. $k(I,p)$, defined by Equation (104), defines the echo position

I	p	$C_0(I,p)$	$C_2(I,p)$	$\lambda(I,p)$	$k(I,p)$	$k(I,p)-p$
$\frac{3}{2}$	-1	-3	-54	1	-1	0
	-3	9	42	-3	7/9	34/9
$\frac{5}{2}$	-1	-8	-144	1	-1	0
	3	6	228	$-3/4$	19/12	$-17/12$
	-5	50	300	$-25/4$	25/12	85/12

powder pattern ($p = -1$) $w^{(2)\mathrm{iso}}_{-1/2,1/2}$ is given by Equation (36). If η cannot be determined by line-shape analysis, the parameter $C_{\mathrm{Q}\eta}$ relating $e^2qQ/\hbar$ and η (Equation 105),

$$C_{\mathrm{Q}\eta} = \frac{e^2qQ}{\hbar}\sqrt{\tfrac{1}{3}\eta^2 + 1} = I(2I-1)\frac{w_{\mathrm{L}}}{2\pi}\sqrt{\frac{40}{3\left\{I(I+1) - \frac{3}{4}\right\}}\left(-\frac{w^{(2)\mathrm{iso}}_{-1/2,1/2}}{w_0}\right)} \tag{105}$$

derived from Equations (12) and (36), is used for characterizing a material. This is especially the case for materials having featureless NMR line shapes in the F_2 dimension, such as glasses. As H_{CS} is also considered during free precession of the spin system, the total shift $w^{\text{fast MAS}}_{p/2,-p/2}$ of the pQ absorption line is given by Equation (106),

$$w^{\text{fast MAS}}_{p/2,-p/2} = \lambda(I,p)w^{(2)\mathrm{iso}}_{-1/2,1/2} - k(I,p)\xi_{-1/2,1/2} - pw_{\mathrm{c}}\delta^{(\mathrm{iso})}_{\mathrm{CS}} \tag{106}$$

which is the sum of Equations (98) and (100).

3.5.7.2 Phase Modulation of the Echo and Antiecho Amplitudes The amplitude of the central-transition FID during the acquisition period τ_4 is described by Equation (107):[18]

$$\begin{aligned} -\tfrac{1}{2}\ \rho^{\mathrm{s}}(t_1,\tau_2,t_3,\tau_4)\ \tfrac{1}{2}\ &= \exp(-\mathrm{i}\tau_4 w^{\text{fast MAS}}_{-1/2,1/2}) \\ &\times \sum_{\substack{p=-2I \\ p\ \text{odd}}}^{2I} -\tfrac{1}{2}\ \rho^{\mathrm{s}}_p(t_1,t_3)\ \tfrac{1}{2}\ \exp(-\mathrm{i}\tau_2 w^{\text{fast MAS}}_{p/2,-p/2}) \\ = \sum_{\substack{p=-2I \\ p\ \text{odd}}}^{2I} -\tfrac{1}{2}\ \rho^{\mathrm{s}}_p(t_1,t_3)\ \tfrac{1}{2}\ &\exp(-\mathrm{i}\{\tau_4 + \lambda(I,p)\tau_2\}w^{(2)\mathrm{iso}}_{-1/2,1/2}) \\ &\times \exp(-\mathrm{i}\{\tau_4 - k(I,p)\tau_2\}\xi_{-1/2,1/2}) \\ &\times \exp(-\mathrm{i}\{\tau_4 - p\tau_2\}w_{\mathrm{c}}\delta^{(\mathrm{iso})}_{\mathrm{CS}}) \end{aligned} \tag{107}$$

Echoes appear at the positions τ_4 defined by the three conditions of Equation (108):

$$\tau_4 + \lambda(I,p)\tau_2 = 0, \quad \tau_4 - k(I,p)\tau_2 = 0, \quad \tau_4 - p\tau_2 = 0 \tag{108}$$

By convention the refocusing of coherence occurring after the second RF pulse is an echo, whereas that occurring before the second RF pulse is an antiecho. They are central-transition signals. As $\lambda(I,p)$ and $k(I,p)$ are odd functions of p, the positions of the echoes and antiechoes are symmetrical with respect to $\tau_4 = 0$.

For a given I and p, the amplitudes of the echo and antiecho located at $\tau_4 = k(I,p)\tau_2$ are phase modulated as in Equations (109) and (110):

$$\begin{aligned} -\tfrac{1}{2}\ \rho^{\mathrm{s}}(t_1,\tau_2,t_3,\tau_4 = k(I,p)\tau_2)\ \tfrac{1}{2}\ &= -\tfrac{1}{2}\ \rho^{\mathrm{s}}_p(t_1,t_3)\ \tfrac{1}{2} \\ &\times \exp\{-\mathrm{i}\tau_2 w_{\mathrm{F1}}(I,p)\} \end{aligned} \tag{109}$$

$$\begin{aligned} w_{\mathrm{F1}}(I,p) &= \{k(I,p) + \lambda(I,p)\}w^{(2)\mathrm{iso}}_{-1/2,1/2} \\ &+ \{k(I,p) - p\}w_{\mathrm{c}}\delta^{(\mathrm{iso})}_{\mathrm{CS}} \end{aligned} \tag{110}$$

The term $w_{\mathrm{F1}}(I,p)$ depends on two isotropic shift values, $w^{(2)\mathrm{iso}}_{-1/2,1/2}$ and $\delta^{(\mathrm{iso})}_{\mathrm{CS}}$, and is therefore independent of the nature of the sample (single crystal or powder). As a result, $w_{\mathrm{F1}}(I,p)$ is not broadened by orientation-dependent parameters when the sample is polycrystalline. Equation (110) plays a crucial role in MQ/MAS methodology. The density matrix $\rho^{\mathrm{s}}_p(t_1,t_3)$ in Equations (107) and (109) is defined by[17]

$$\rho^{\mathrm{s}}_p(t_1,t_3) = T\exp(-\mathrm{i}\Omega t_3)T^{\dagger}\rho^{\mathrm{s}}_p(t_1)T\exp(\mathrm{i}\Omega t_3)T^{\dagger} \tag{111}$$

Equation (111) differs from Equation (43) by the initial density operator, $\rho^{\mathrm{s}}_p(t_1)$ instead of $\rho(0)$. In fact Equation (111) has been used within the two numerical procedures described in sections 3.4.3 and 3.5.5. Similarly, the amplitudes of echo and antiecho located at $\tau_4 = -\lambda(I,p)\tau_2$ are phase modulated by

$$\begin{aligned} w_{\mathrm{F1}}(I,p) &= -\{k(I,p) + \lambda(I,p)\}\xi_{-1/2,1/2} \\ &- \{p + \lambda(I,p)\}\delta^{(\mathrm{iso})}_{\mathrm{CS}}w_{\mathrm{c}} \end{aligned} \tag{112}$$

Those of the echo and antiecho located at $\tau_4 = p\tau_2$ are phase modulated by

$$w_{\mathrm{F1}}(I,p) = \{p + \lambda(I,p)\}w^{(2)\mathrm{iso}}_{-1/2,1/2} + \{p - k(I,p)\}\xi_{-1/2,1/2} \tag{113}$$

Equations (112) and (113) depend on $\xi_{-1/2,1/2}$, which is a function of orientation parameters. In a powder sample, this dependence may destroy the echoes and antiechoes.

For the $\tau_4 = \tau_2$ Hahn echo, which is the refocusing of the 1Q coherence developed at the end of the first RF pulse, we have from Table 2: $p = 1$, $k(I,1) = 1$, and $\lambda(I,1) = -1$, that is, $w_{\mathrm{F1}}(I,1) = 0$ for Equations (110), (112) and (113). The amplitudes of the $\tau_4 = \tau_2$ Hahn echo (1Q echo) and the $\tau_4 = -\tau_2$ Hahn antiecho (-1Q antiecho) are not phase modulated. The MQ/MAS method (see section 3.5.8), if it is applied to the $\tau_4 = \tau_2$ Hahn echo, will not split peaks in the F_1 dimension.

Figure 46 presents the echoes for the spin $I = \frac{3}{2}$ system.[49] The echo located at $\tau_4 = \frac{7}{9}\tau_2$, which is the refocusing of the -3Q coherence developed at the end of the first RF pulse, is called -3Q echo and is close to the 1Q echo. The latter has an amplitude much

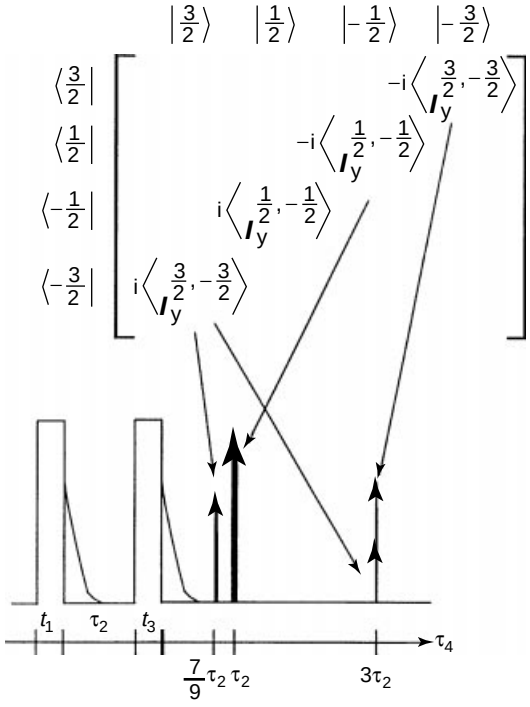

Figure 46 On-resonance coherences in the second main diagonal of the spin $I = \frac{3}{2}$ density matrix developed at the end of the first RF pulse and refocused as Hahn echoes by the second RF pulse in rapid MAS conditions: the thick arrow at $\tau_4 = \tau_2$ is the 1Q Hahn echo, the medium arrow at $\tau_4 = \frac{7}{9}\tau_2$ is the echo observed in the 3Q MAS experiment, and the thin arrows at $\tau_4 = 3\tau_2$ are echoes not often observed. On-resonance coherences and corresponding echoes are connected by arrows. Antiechoes are not shown. (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **55**, 8406–8424 (1997).)

larger than that of the $-3Q$ echo. As the 3Q and $-3Q$ coherences at the end of the first RF pulse have opposite amplitudes, Figure 47 only shows the graph of the imaginary part,

$$I_y^{\frac{3}{2},-\frac{3}{2}}(t_1) \quad = \mathrm{Tr} \quad \rho(t_1)\, I_y^{\frac{3}{2},-\frac{3}{2}}$$

of the 3Q coherence for a spin $I = \frac{3}{2}$, for three values of quadrupole coupling.[49] Their maxima are obtained with longer pulse duration than that used for $-1Q$ coherence. As far as the pulse durations are short enough, the transfer function can be calculated by using the numerical procedure described in section 3.5.5.

Phase cycling the RF pulse and the receiver is required to cancel the 1Q echo and the $-1Q$ antiecho without affecting the $-3Q$ echo and the 3Q antiecho of an $I = \frac{3}{2}$ system. The other method is to apply a pulsed magnetic field gradient to cancel the unwanted echo and antiecho.[97]

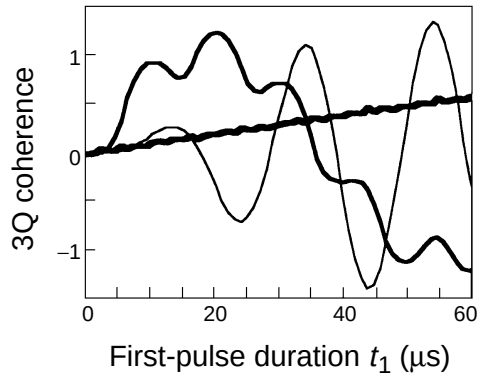

Figure 47 Graph of the excitation function (or the imaginary part) of the 3Q on-resonance coherence for a spin $I = \frac{3}{2}$, generated by the first RF pulse in a Hahn echo sequence consisting of two $-X$ pulses, versus the first-pulse duration t_1 for $w_{RF}/(2\pi) = 50\,\mathrm{kHz}$ and three values of $w_Q/(2\pi)$: thin line for 10 kHz, medium line for 50 kHz, and thick line for 200 kHz. (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **55**, 8406–8424 (1997).)

3.5.8 Multiple Quantum/Magic-angle Spinning Under Rapid Conditions

Frydman and colleagues[47,98] introduced the MQ/MAS methodology, which is based on 2-D methods. The MQ/MAS methodology takes advantage of the properties of the MQ coherences to generate high-resolution isotropic spectra along the F_1 dimension, giving the number of different crystallographic sites in a compound when the conventional 1-D MAS experiment provides overlapping absorption lines. Among all the coherences generated by the first RF pulse, the specific MQ coherences are detected selectively by phase cycling the RF pulses and the receiver in the Hahn echo sequence. A sheared 2-D MQ/MAS spectrum represents the correlation of a specific MQ coherence in the F_1 dimension with an SQ coherence in the F_2 dimension. The 2-D spectrum associated with a single site consists of a single peak, an MQ-filtered central-transition peak (simply called the peak in the remainder of this article). Provided that the sheared 2-D MQ/MAS spectrum is labeled in chemical shift units and the zero ppm positions in both axes are defined, knowledge of the two observed chemical shifts of the center of gravity $\delta_{G1}^{(obs)}$ and $\delta_{G2}^{(obs)}$ of a peak in the two dimensions (F_1 and F_2) enables determination of the true isotropic chemical shift $\delta_{CS}^{(iso)}$ of an absorption line. The asymmetry parameter η may be determined by the line shape of the cross-section of a peak parallel to F_2.

The rotor spinning axis is at the magic angle throughout the experiment. Thus a standard MAS probehead suffices for the experiment, but the NMR spectrometer must be equipped with a digital dephaser of the RF

pulses or a pulsed magnetic field gradient probe. The optimum experimental conditions for MQ/MAS experiments are strong RF pulses, small offset, and high rotor spinning rate.[99] Quantitative results on spin populations are not obvious, but are possible under particular conditions. Conventional 1-D MAS spectra obtained with short pulse durations allow determination of the spin populations if the absorption lines are not overlapping (section 3.1.4.1). If not, the MQ/MAS methodology can be used to determine the number of peaks or crystallographic sites, their true isotropic chemical shifts, and the quadrupole parameters. Thanks to these data, the 1-D MAS spectra can be simulated to obtain the spin population of each site.[100]

3.5.8.1 Selective Detection of the ±3Q Coherences To cancel the 1Q echo and −1Q antiecho in the two-pulse sequence, the standard way is to cycle the RF pulse phases and that of the receiver. We have already applied a mathematical approach.[49] Here we apply a diagramatic approach. Figure 48(a–c) shows the positions of the on-resonance coherences (±1Q and ±3Q) at the end of the first RF pulse when the RF field $\mathbf{B}_1$ is positioned at 0°, 120°, and 240° in the rotating frame. The ±3Q coherences remain along the *y* axis of the receiver, in agreement with Equation (64). Cycling according to these three cases co-adds the ±3Q coherences and cancels the ±1Q coherences. However the zero-quantum coherences are not affected by this RF pulse cycling and are also co-added; they will contribute to the FID following the second RF pulse. To cancel the zero-quantum coherences, $\mathbf{B}_1$ is positioned at 90°, 180°, and 300° (Figure 48d–f). The

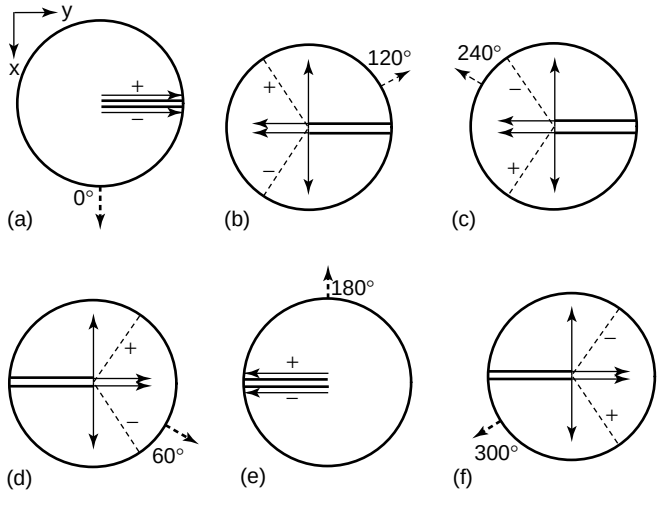

Figure 48 Evolution of the ±1Q on-resonance coherences (dashed lines, + for 1Q and − for −1Q) with their projections (arrows) along the *x* and *y* axis of the rotating frame and that of the ±3Q on-resonance coherences (thick lines) for the six positions of the RF field $\mathbf{B}_1$ (dashed arrows) at the end of the first RF pulse in the two-pulse MQ/MAS experiment: (a) 0°, (b) 120°, (c) 240°, (d) 60°, (e) 180°, and (f) 300°.

±1Q coherences are canceled but ±3Q coherences are along the $-y$ axis of the receiver. To co-add the contributions of the ±3Q coherences to those of Figure 48(a–c) the receiver phase must be set to $-y$. In this case the contributions of the zero-quantum coherences are also canceled. As a result, these six phase-cyclings of the first RF pulses and the receiver allow selective detection of the ±3Q coherences generated by the first RF pulse.[101]

3.5.8.2 Experimental and Data Processing From now on, standard notation for 2-D NMR is used: t_1 (the pulse separation) is the experimental evolution period, and t_2 is the acquisition period. First, the echo signal is optimized by systematic variation of each pulse duration.[99] Then, signal acquisition in the 2-D experiment occurs as follows: for each increment Δt_1 of the experimental evolution period, the time-domain signals appearing after the second RF pulse are acquired in the simultaneous mode. To obtain a pure 2-D absorption spectrum, the hypercomplex method[102] for generating quadrature detection in the t_1 domain is applied during the signal acquisition. Sometimes the increment of the experimental evolution period is synchronized with the rotor spinning rate.[103]

Double Fourier transform of the experimental data from $F(t_1, t_2)$ to $F(w_1, w_2)$ via $F(t_1, w_2)$ yields a pure 2-D absorption spectrum but tilted. However a shearing transformation of $F(t_1, w_2)$ yields a powder pattern along the F_2 dimension, and a high-resolution featureless line shape in the F_1 dimension. The sheared 2-D spectrum becomes a 2-D isotropic/anisotropic correlation spectrum. More details on the shearing transformation can be found in literature.[49,98,102]

3.5.8.3 Labeling the F_1 Axis in the Frequency Unit The spectral width (SW) in the F_1 dimension (SW1) of a 2-D MQ/MAS spectrum is equal to the inverse of the increment of the evolution period. Unfortunately, two definitions of the evolution period appear in the literature, giving two conventions for scaling in frequency units the SW1. The first convention (Cz) considers the experimental evolution period as the evolution period;[98,104] therefore $\mathrm{SW1_{Cz}} = 1/(\Delta t_1)$. The second convention (Ck) considers the position of the echo relative to the first RF pulse, equal to $(1 + k)$ times the experimental evolution period, as the evolution period;[102,105] therefore $\mathrm{SW1_{Ck}} = 1/\{(1 + k)\Delta t_1\}$. This article uses the Cz convention, because both conventions have been compared previously.[49]

If we synchronize the increment of the experimental evolution period Δt_1 with the inverse of the rotor spinning rate ν_{rotor}, that is if $\Delta t_1 = 1/\nu_{\mathrm{rotor}}$, then $\mathrm{SW1_{Cz}} = \nu_{\mathrm{rotor}}$. Spinning sidebands along the F_1 dimension, located outside the $\mathrm{SW1_{Cz}}$, will be folded back onto the centerband, increasing the spectral line intensity and improving the line shape of the centerband.[103] However,

Table 3 Relationships of various parameters of chemical shift (CS), offset, shift, and center of gravity (CG) expressed with respect to the carrier frequency w_c, the coherence order p of the echo located at kt_1, and λ (Equation 103) used in this paper and those of Amoureux and Fernandez[106] for sheared 2-D MQ/MAS spectra. ($w_{F1}(I,p) = (k+\lambda)w^{(2)iso}_{-1/2,1/2} + (k-p)\delta^{(iso)}_{CS}w_c$; $k \equiv R$ in Amoureux and Fernandez's notation.)

Parameter	This paper	Amoureux and Fernandez[106]
$\delta^{(obs)}_{G2}$:[a]	$\delta^{(iso)}_{CS} + \frac{w^{(2)iso}_{-1/2,1/2}}{w_c}$	$\delta^{(iso)}_{CS} + \frac{w^{(2)iso}_{-1/2,1/2}}{w_c}$
$\delta^{(obs)}_{G1z}$:[b]	$\frac{w_{F1}(I,p)}{w_c} = (k-p)\delta^{(iso)}_{CS} + (k+\lambda)\frac{w^{(2)iso}_{-1/2,1/2}}{w_c}$	$\frac{w_{F1}(I,p)}{(k-p)w_c} = \delta^{(iso)}_{CS} - \frac{10}{17}\frac{w^{(2)iso}_{-1/2,1/2}}{w_c}$
$\delta^{(iso)}_{CS}$:[c]	$\frac{10}{27}\delta^{(obs)}_{G2} - \frac{1}{p+\lambda}\delta^{(obs)}_{G1z}$	$\frac{10}{27}\delta^{(obs)}_{G2} + \frac{17}{27}\delta^{(obs)}_{G1z}$
$\frac{w^{(2)iso}_{-1/2,1/2}}{w_c}$:[d]	$\frac{1}{\lambda+p}\{\delta^{(obs)}_{G1z} - (k-p)\delta^{(obs)}_{G2}\}$	$\frac{17}{27}(\delta^{(obs)}_{G2} - \delta^{(obs)}_{G1z})$
Ω_{F2}:[e]	Experimental	Experimental
δ_{F2}:[f]	$\frac{\Omega_{F2}}{w_c}$	$\frac{\Omega_{F2}}{w_c}$
Ω_{F1z}:[g]	$(k-p)\Omega_{F2}$	$(k-p)\Omega_{F2}$
δ_{F1z}:[h]	$\frac{\Omega_{F1z}}{w_c} = (k-p)\delta_{F2}$	$\frac{\Omega_{F1z}}{(k-p)w_c} = \delta_{F2}$

[a] Observed chemical shift of the center of gravity of a peak in the F_2 dimension.
[b] Observed chemical shift of the center of gravity of a peak in the F_1 dimension.
[c] Isotropic chemical shift.
[d] Second-order quadrupole shift of the central line expressed as chemical shift.
[e] Offset of the frequency carrier with respect to the aqueous solution in the F_2 dimension.
[f] chemical shift of the frequency carrier with respect to the aqueous solution in the F_2 dimension.
[g] Offset of the frequency carrier with respect to the aqueous solution in the F_1 dimension.
[h] chemical shift of the frequency carrier with respect to the aqueous solution in the F_1 dimension.

if ν_{rotor} is not high enough and the spectrum consists of several centerbands, one or several of them can be located outside the $SW1_{Cz}$ but still appear within the $SW1_{Cz}$ as a result of folding. Consequently, the positions of the center of gravity of these folded centerbands will be incorrect.

3.5.8.4 Labeling the F_1 Axis in Chemical Shift Units In theoretical study, the isotropic chemical shifts of a peak along the F_1 and F_2 dimensions of a 2-D spectrum are referenced to w_c, which is located at the center of each of the two SWs. However, the second-order quadrupole shift is referenced to w_L. If the acquisition of the data starts at the position $t_2 = kt_1$ of the echo, the position $w_{F1}(I,p)$ of a peak relative to w_c along the F_1 dimension is given by Equation (110). Therefore, the observed chemical shift of its center of gravity $\delta^{(obs)}_{G1}$ with respect to w_c is equal to its position divided by w_c (Equation 114):

$$\delta^{(obs)}_{G1} = \frac{w_{F1}(I,p)}{w_c} \tag{114}$$

Amoureux and Fernandez[106] apply Equation (115):

$$\delta^{(obs)}_{G1} = \frac{w_{F1}(I,p)}{(k-p)w_c} \tag{115}$$

that is, the carrier frequency in the F_1 dimension is $(k-p)w_c$ instead of w_c. In some cases $(k-p)w_c$ takes a negative value. Table 3 shows the main differences between these two points of view.

As the aim of the shearing transformation is to shift data acquisition from the end of the second RF pulse to the echo position after the end of the 2-D signal acquisition,[49] Equations (110), (112) and (113) remain valid for a sheared 2-D spectrum. Along the F_1 dimension, the observed chemical shift of the center of gravity of a peak relative to w_c (Equation 114) for the Cz convention is given by Equation (116):

$$\delta^{(iso)}_{G1z} = (k-p)\delta^{(obs)}_{CS} + (k+\lambda)\frac{w^{(2)iso}_{-1/2,1/2}}{w_c} \tag{116}$$

As the observed chemical shift of the center of gravity of the central-transition powder pattern relative to w_c along the F_2 dimension $\delta^{(obs)}_{G2}$ is given by Equation (37), the latter and Equation (116) allow deduction of the true isotropic chemical shift of a peak relative to w_c (Equation 117):

$$\delta^{(iso)}_{CS} = \frac{(k+\lambda)\delta^{(obs)}_{G2} - \delta^{(obs)}_{G1z}}{p+\lambda} = \frac{10}{27}\delta^{(obs)}_{G2} - \frac{1}{p+\lambda}\delta^{(obs)}_{G1z} \tag{117}$$

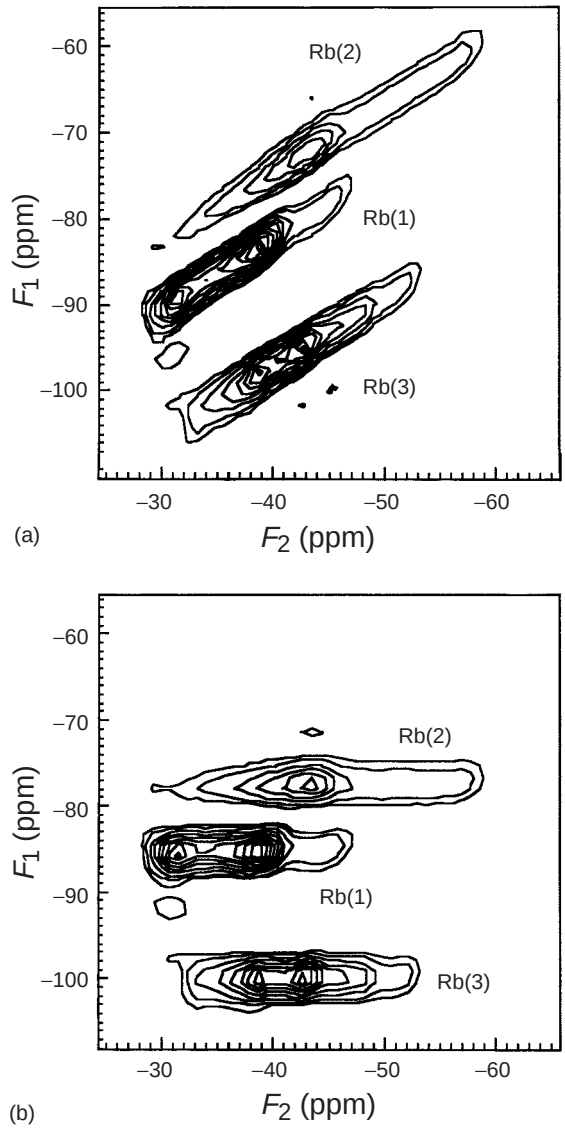

Figure 49 Contour plot of the (a) unsheared and (b) sheared 2-D 3Q-MAS spectra of ^{87}Rb ($I = \frac{3}{2}$) in $RbNO_3$ powder, obtained at 98.2 MHz. The three Rb sites, Rb(1), Rb(2), and Rb(3) are observed. The axis of the F_1 dimension is labeled using the Cz convention. The F_2 dimension is the usual chemical shift axis. (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **58**, 2764–2782 (1998).)

From a practical point of view, in a 1-D spectrum or the F_2 dimension of a 2-D spectrum, the chemical shift of an absorption line is referenced experimentally to an external aqueous solution with $\delta_{CS}^{(iso)} = 0$ ppm, by definition. For both dimensions of a 2-D MQ/MAS spectrum, knowing the frequency offsets of w_c relative to the aqueous solution, Ω_{F1} in the F_1 dimension and Ω_{F2} in the F_2 dimension, allows us to express the observed chemical shifts of the center of gravity of any peak ($\delta_{G1}^{(obs)}$ in the F_1 dimension and $\delta_{G2}^{(obs)}$ in the F_2 dimension) relative to the aqueous solution. Unfortunately, Ω_{F1} differs from Ω_{F2}. For the Cz convention, this difference is given by Equation (118):

$$\Omega_{F1z} = (k - p)\Omega_{F2} \tag{118}$$

Now we know the position (Equation 112) of a peak relative to w_c and the offset (Equation 118) of w_c relative to the aqueous solution in the F_1 dimension. Therefore we know the position of the peak relative to the aqueous solution, which is what we are looking for. As the offsets in both dimensions are known, Equations (37), (116) and (117) remain valid even when the observed chemical shifts of the center of gravity of a peak along the two axes are referenced to an aqueous solution instead of w_c.

We can also deduce $w_{-1/2,1/2}^{(2)iso}$ from Equations (37) and (116):

$$\frac{w_{-1/2,1/2}^{(2)iso}}{w_c} = \frac{1}{\lambda + p}\{\delta_{G1z}^{(obs)} - (k - p)\delta_{G2}^{(obs)}\} \tag{119}$$

However, Equation (119) does not allow us to determine the quadrupole coupling constant and the asymmetry

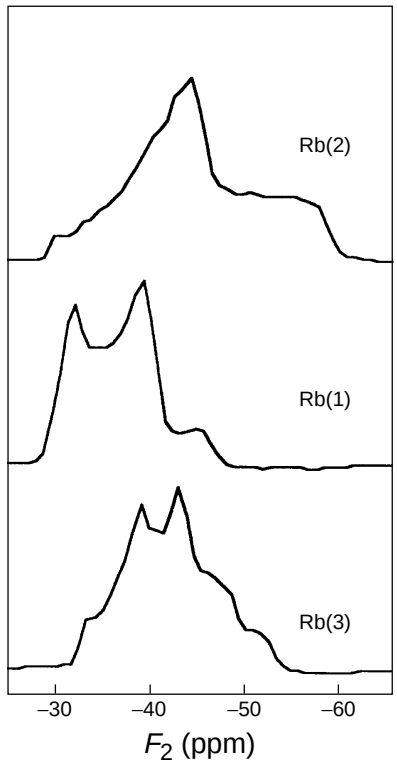

Figure 50 The three cross-section spectra (parallel to the F_2 dimension) of the sheared 2-D 3Q-MAS spectrum of ^{87}Rb ($I = \frac{3}{2}$) in $RbNO_3$ powder of Figure 49(b). (Reproduced by permission of The American Physical Society from P.P. Man, *Phys. Rev. B*, **58**, 2764–2782 (1998).)

Table 4 The carrier frequency, offset Ω_{F1} of the frequency carrier with respect to the aqueous solution, SW1, and the chemical shift range $\Delta\delta_{CS}$ in the F_1 dimension are related to the carrier frequency w_c and the offset Ω_{F2} of the frequency carrier relative to the aqueous solution in the F_2 dimension, the increment Δt_1 of the experimental evolution period, and the coherence order p of the echo located at kt_1 relative to the second RF pulse for the four conventions appearing in the literature. The Cz convention considers the experimental evolution period t_1 as the evolution period, whereas the Ck convention considers the position $(1 + k)t_1$ of the echo relative to the first RF pulse as the evolution period

Convention	Carrier frequency	Ω_{F1}	SW1	$\Delta\delta_{CS}$
Cz, Medek et al.[98]	w_c	$(k-p)\Omega_{F2}$	$\frac{1}{\Delta t_1}$	$\frac{1}{w_c \Delta t_1}$
Ck, Massiot[102]	w_c	$\frac{k-p}{1+k}\Omega_{F2}$	$\frac{1}{(1+k)\Delta t_1}$	$\frac{1}{w_c(1+k)\Delta t_1}$
Cz, Amoureux and Fernandez[106]	$(k-p)w_c$	$(k-p)\Omega_{F2}$	$\frac{1}{\Delta t_1}$	$\frac{1}{(k-p)w_c \Delta t_1}$
Ck, Amoureux and Fernandez[106]	$(k-p)w_c$	$\frac{k-p}{1+k}\Omega_{F2}$	$\frac{1}{(1+k)\Delta t_1}$	$\frac{1}{(k-p)w_c(1+k)\Delta t_1}$

parameter η independently. The latter may be obtained by fitting the line shape of the cross-section (parallel to the F_2 dimension) of the peak.

Figure 49 shows the contour plots of the unsheared and sheared 2-D 3Q-MAS spectra of ^{87}Rb ($I = \frac{3}{2}$) in $RbNO_3$ powder, acquired at 98.2 MHz.[49] The F_1 axis is labeled according to the Cz convention. The three Rb sites Rb(1), Rb(2), and Rb(3) are observed. The line widths of these sheared spectra (Figure 49b) in the F_1 dimension are smaller than those in the F_2 dimension, given the high-resolution spectra along the F_1 dimension. Figure 50 presents the cross-section spectra (parallel to the F_2 dimension in Figure 49b) of the three sites.[49] Lineshape analysis provides us with the asymmetry parameter of each rubidium site.

For convenience, Table 4 gathers the relevant relationships for the two conventions, Cz and Ck, and the two points of view (the present paper using w_c as the carrier frequency and Amoureux and colleagues using $(k-p)w_c$ as the carrier frequency in the F_1 dimension).

4 PERSPECTIVE AND FUTURE DEVELOPMENTS

In solid-state NMR, the quadrupole coupling constant and the asymmetry parameter of half-integer quadrupole spins in a powder sample are determined by the central-transition powder pattern if the second-order quadrupole interaction is the main interaction. This powder pattern is generally observed with the one-pulse sequence; however, this is not always the case for the satellite transition. This article has presented both 1-D and 2-D methods based on Hahn echoes to determine these quadrupole parameters.

When the first-order quadrupole interaction becomes the main interaction, the central transition has a featureless line shape. The two parameters should be determined by the satellite-transition powder pattern if it is to be observed properly. Alternatively, some of the methods proposed in this article should be applied to determine these parameters from a featureless central-transition line shape, such as the two-pulse sequence with a short pulse separation.

As the strength of the Zeeman field $\mathbf{B}_0$ increases year after year, the effects of the second-order quadrupole interaction, which is inversely proportional to $\mathbf{B}_0$, decrease. However, the first-order quadrupole interaction, which does not depend on $\mathbf{B}_0$, is likely to become the dominant interaction in the near future. If the satellite-transition powder pattern cannot be detected properly, Solomon echo sequences should be applied. Unfortunately, these echoes are not yet well characterized. Progress in understanding these echoes is required.

LIST OF SYMBOLS

$\mathbf{B}_0$	Zeeman field
$\mathbf{B}_1$	Radiofrequency field
$\delta_{CS}^{(iso)}$	Isotropic chemical shift
$\delta_{G1}^{(obs)}$	Observed chemical shift of the center of gravity of the central-transition powder pattern in the F_1 dimension
$\delta_{G2}^{(obs)}$	Observed chemical shift of the center of gravity of the central-transition powder pattern in the F_2 dimension
D_0	Recycle delay of an NMR experiment
$e^2qQ/\hbar$	Quadrupole coupling constant
$F(t_1, \tau_2 = 0)$	Amplitude of the FID at the end of an RF pulse
F_1	First dimension of a 2-D spectrum
F_2	Second dimension of a 2-D spectrum
η	Asymmetry parameter

H_{CS}	Isotropic chemical shift interaction
H_{Δ}	Offset interaction
$H_Q^{(1)}$	First-order quadrupole interaction
$H_Q^{(2)}$	Second-order quadrupole interaction
$H_Q^{(2)\text{fast MAS}}$	Second-order quadrupole interaction of a sample rotating at the magic angle under rapid conditions
H_Z	Zeeman interaction
I	Nuclear spin
$\langle I_y(t_1)\rangle$	Amplitude of an FID at the end of an RF pulse
θ_m	Magic angle ($= 54°44'$)
t_1	Duration of the first RF pulse, or the evolution period in a 2-D experiment
t_2	Acquisition period in a 2-D experiment
t_3	Duration of the second RF pulse in a two-pulse sequence
τ_2	Acquisition period in a one-pulse sequence, or separation between the two RF pulses in a two-pulse sequence
τ_4	Acquisition period in a two-pulse sequence
T_1	Spin–lattice relaxation time of a nuclear spin
T_{FID}	Duration of FID
$w_{-1/2,1/2}^{(2)\text{iso}}$	Second-order quadrupole shift of the center of gravity of the central-transition powder pattern
w_L	Larmor frequency
w_{RF}	Amplitude of the RF pulse
w_Q	Quadrupole coupling
w_c	Carrier frequency
$\zeta^2(m+1, m)$	Probability per second that an RF field induces a transition between $\|m\rangle$ and $\|m+1\rangle$

ABBREVIATIONS AND ACRONYMS

CW	Continuous Wave
EFG	Electric-field Gradient
FID	Free-induction Decay
MAS	Magic-angle Spinning
MQ	Multiple Quantum
MQ/MAS	Multiple Quantum/Magic-angle Spinning
NMR	Nuclear Magnetic Resonance
PAS	Principal-axis System
RF	Radiofrequency
SQ	Single Quantum
SW	Spectral Width
VAS	Variable-angle Spinning
1-D	One-dimensional
2-D	Two-dimensional

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Multinuclear Magnetic Resonance Spectroscopic Imaging

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Particle Size Analysis **(Volume 6)**
Surface Area and Pore Size Distributions

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Steel and Related Materials **(Volume 10)**
Nuclear Magnetic Resonance in Metals Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance of Geological Materials and Glasses • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, 1961.
2. R.V. Pound, 'Nuclear Electric Quadrupole Interactions in Crystals', *Phys. Rev.*, **79**, 685–702 (1950).
3. J.L. Dye, A.S. Ellaboudy, J. Kim, 'Solid State NMR of Quadrupolar Nuclei', in *Modern NMR Techniques and Their Application in Chemistry, Pract. Spectr. Ser.*, eds. A.I. Popov, K. Hallenga, Marcel Dekker, New York, 217–323, Vol. 11, 1991.

4. M.H. Cohen, F. Reif, 'Quadrupole Effects in Nuclear Magnetic Resonance Studies of Solids', in *Solid State Physics*, eds. F. Seitz, D. Turnbull, Academic Press, New York, 321–438, Vol. 5, 1957.
5. F. Borsa, A. Rigamonti, 'Comparison of NMR and NQR Studies of Phase Transitions in Disordered and Ordered Crystals', in *Topics in Current Physics, Structural Phase Transitions II*, eds. K.A. Müller, H. Thomas, Springer-Verlag, Berlin, 83–183,Vol. 45, 1991.
6. A. Rigamonti, 'NMR/NQR Studies of Structural Phase Transitions', *Adv. Phys.*, **33**, 115–191 (1984).
7. D. Freude, J. Haase, 'Quadrupole Effects in Solid-state Nuclear Magnetic Resonance', in *NMR Basic Principles and Progress*, eds. P. Diehl, E. Fluck, H. Günter, R. Kosfeld, J. Seelig, Springer-Verlag, Berlin, 1–90, Vol. 29, 1993.
8. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1990.
9. R. Freeman, *Spin Choreography, Basic Steps in High Resolution NMR*, Spektrum Academic, Oxford, 1997.
10. K. Schmidt-Rohr, H.W. Spiess, *Multidimensional Solid-state NMR and Polymers*, Academic Press, San Diego, 1994.
11. P.P. Man, 'Quadrupole Interaction', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, Wiley, Chichester, 3838–3848, 1996.
12. A.J. Vega, 'Quadrupolar Nuclei in Solids', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, Wiley, Chichester, 3888–3898, 1996.
13. I. Solomon, 'Multiple Echoes in Solids', *Phys. Rev.*, **110**, 61–65 (1958).
14. E.L. Hahn, 'Spin Echoes', *Phys. Rev.*, **80**, 580–594 (1950).
15. E. Lippmaa, A. Samoson, M. Mägi, 'High-resolution ^{27}Al NMR of Aluminosilicates', *J. Am. Chem. Soc.*, **108**, 1730–1735 (1986).
16. P.P. Man, 'Numerical Analysis of Solomon Echo Amplitudes in Static Solids', *J. Chem. Phys.*, **106**, 3908–3919 (1997).
17. P.P. Man, 'Numerical Analysis of Hahn Echoes in Solids', *Phys. Rev. B*, **52**, 9418–9426 (1995).
18. P.P. Man, 'Second-order Quadrupole Effects on Hahn Echoes in Fast-rotating Solids at the Magic Angle', *Phys. Rev. B*, **55**, 8406–8424 (1997).
19. M.H. Levitt, 'The Signs of Frequencies and Phases in NMR', *J. Magn. Reson.*, **126**, 164–182 (1997).
20. K. Narita, J.J. Umeda, H. Kusumoto, 'NMR Powder Patterns of the Second-order Nuclear Quadrupole Interaction in Solids with Asymmetric Field Gradient', *J. Chem. Phys.*, **44**, 2719–2723 (1966).
21. P.P. Man, 'Measurement of Quadrupolar Coupling with a Two-pulse Sequence in Solid-state NMR', *Mol. Phys.*, **69**, 337–346 (1990).
22. D.W. Alderman, M.S. Solum, D.M. Grant, 'Methods for Analyzing Spectroscopic Line Shapes. NMR Solid Powder Patterns', *J. Chem. Phys.*, **84**, 3717–3725 (1986).
23. M. Bak, N.C. Nielsen, 'REPULSION, a Novel Approach to Efficient Powder Averaging in Solid-state NMR', *J. Magn. Reson.*, **125**, 132–139 (1997).
24. C.P. Slichter, *Principles of Magnetic Resonance*, Springer-Verlag, Berlin, 1990.
25. M. Goldman, *Quantum Description of High-resolution NMR in Liquids*, Clarendon Press, Oxford, 1988.
26. G. Bodenhausen, 'Multiple Quantum NMR', *Progr. Nucl. Magn. Reson. Spectrosc.*, **14**, 137–173 (1981).
27. G. Lindblom, H. Wennerström, B. Lindman, 'Multiple Quantum Transitions for Spin-$\frac{3}{2}$ Nuclei in the NMR Spectra of Lyotropic Liquid Crystals', *J. Magn. Reson.*, **23**, 177–179 (1976).
28. P.A. Spencer, D.G. Hughes, 'Double-quantum NMR of ^{23}Na in $NaNO_3$ by Double Resonance', *J. Magn. Reson.*, **38**, 1–8 (1980).
29. P.P. Man, R. Couty, J. Fraissard, 'Determination of the Line Intensities of ^{27}Al in Al_2O_3 by Solid-state NMR', *J. Magn. Reson.*, **86**, 613–617 (1990).
30. P.C. Taylor, J.B. Baugher, H.M. Kriz, 'Magnetic Resonance Spectra in Polycrystalline Solids', *Chem. Rev.*, **75**, 203–240 (1975).
31. J. Skibsted, N.C. Nielsen, H. Bildsøe, H.J. Jakobsen, 'Satellite Transition in MAS NMR Spectra of Quadrupolar Nuclei', *J. Magn. Reson.*, **95**, 88–117 (1991).
32. B. Herreros, P.P. Man, J.-M. Manoli, J. Fraissard, '^{139}La Solid-state NMR Investigation of Lanthanum-exchanged Y Zeolites', *J. Chem. Soc., Chem. Commun.*, 464–466 (1992).
33. D.R. Eaton, R.J. Buist, B.G. Sayer, 'The ^{59}Co Nuclear Magnetic Resonance Spectra of Polycrystalline Complexes', *Can. J. Chem.*, **65**, 1332–1335 (1987).
34. Z. Zheng, Z. Gan, N.K. Sethi, D.W. Alderman, D.M. Grant, 'An Efficient Simulation of VAS Lineshapes for the Quadrupolar Nuclei with Half-integer Spin', *J. Magn. Reson.*, **95**, 509–522 (1991).
35. D. Massiot, B. Cote, F. Taulelle, J.-P. Coutures, '^{27}Al MAS NMR of Crystalline and Amorphous Materials', in *Application of NMR Spectroscopy to Cement Science*, eds. P. Colombet, A.R. Grimmer, Gordon and Breach, Yverdon, Switzerland, 153–169, 1994.
36. D. Massiot, I. Farnan, N. Gauthier, D. Trumeau, A. Trokiner, J.-P. Coutures, 'Ga-71 and Ga-69 NMR Study of Beta-Ga_2O_3: Resolution of Four- and Six-fold Coordinated Ga Sites in Static Conditions', *Solid State NMR*, **4**, 241–248 (1995).
37. T.J. Bastow, '^{139}La NMR Characterization of La_2O_3 and $La_{1-x}Sr_xMO_3$ where M = Cr, Mn or Co', *Solid State NMR*, **3**, 17–22 (1994).
38. S. Vega, Y. Naor, 'Triple Quantum NMR on Spin Systems with $I = 3/2$ in Solids', *J. Chem. Phys.*, **75**, 75–86 (1981).

39. P.P. Man, 'Investigation of the Central Line of ^{55}Mn in $KMnO_4$ by a Two-dimensional NMR Method', *J. Magn. Reson.*, **67**, 78–90 (1986).
40. A. Wokaun, R.R. Ernst, 'Selective Excitation and Detection in Multilevel Spin Systems: Application of Single Transition Operators', *J. Chem. Phys.*, **67**, 1752–1758 (1977).
41. S. Ding, C.A. McDowell, 'Spin-locking Spectral Lineshapes of the Central Transition of Half-integer Quadrupole Systems Under MAS', *J. Mol. Struct.*, **355**, 135–142 (1995).
42. J.A.M. van der Mijden, R. Janssen, W.S. Veeman, 'Analytical Description of the $I = 5/2$ Quadrupole Nutation Experiment', *Mol. Phys.*, **69**, 53–64 (1990).
43. P.P. Man, 'Analytical Expression for the Spin-$\frac{5}{2}$ Line Intensities', *Mol. Phys.*, **78**, 307–318 (1993).
44. P.P. Man, P. Tougne, 'Exact Expression for the Spin-$\frac{7}{2}$ Line Intensities: Application to Solid State ^{59}Co(III) NMR', *Mol. Phys.*, **83**, 997–1009 (1994).
45. P. Kempgens, J. Hirschinger, P. Granger, J. Rosé, 'Spin-$\frac{7}{2}$ Nutation and Hahn-echo Amplitudes in Model Compounds and Application to the Tetrahedral Cluster $Co_4(CO)_{12}$', *Solid State NMR*, **10**, 95–103 (1997).
46. S.Z. Ageev, B.C. Sanctuary, 'Analytical Solutions for Spin-$\frac{7}{2}$ Line Intensities in Solid State NMR', *Mol. Phys.*, **84**, 835–844 (1995).
47. L. Frydman, J.S. Harwood, 'Isotropic Spectra of Half-integer Quadrupolar Spins from Bidimensional MAS NMR', *J. Am. Chem. Soc.*, **117**, 5367–5368 (1995).
48. S.Z. Ageev, P.P. Man, B.C. Sanctuary, 'Detection of Double and Four Quantum Coherences for Spin-$\frac{7}{2}$ Excited by Spin Lock Pulse Sequences', *J. Magn. Reson.*, **128**, 12–20 (1997).
49. P.P. Man, 'Scaling and Labeling the High-resolution Isotropic Axis of Two-dimensional Multiple Quantum Magic-angle-spinning Spectra of Half-integer Quadrupole Spins', *Phys. Rev. B*, **58**, 2764–2782 (1998).
50. P.P. Man, J. Klinowski, A. Trokiner, H. Zanni, P. Papon, 'Selective and Non-selective NMR Excitation of Quadrupolar Nuclei in the Solid State', *Chem. Phys. Lett.*, **151**, 143–150 (1988).
51. P.P. Man, 'Excitation Conditions for Quantitative Determination of Quadrupolar Spins with One Pulse or Spin-echo Sequences in Solid-state NMR', *Appl. Magn. Reson.*, **4**, 65–87 (1993).
52. V.H. Schmidt, 'Pulse Response in the Presence of Quadrupolar Splitting, in Pulsed Magnetic and Optical Resonance', in *Proceedings of the Ampere International Summer School 1971*, ed. R. Blinc, J. Stefan Institut, Ljubljana, 75–83, 1972.
53. L. Pandey, D.G. Hughes, 'RF Pulse Response of a System of $I = 3/2$ Nuclei in the Presence of Quadrupole Splitting', *J. Mol. Struct.*, **111**, 91–96 (1983).
54. L. Pandey, S. Towta, D.G. Hughes, 'NMR Pulse Response and Measurement of the Quadrupole Coupling Constant of $I = 3/2$ Nuclei', *J. Chem. Phys.*, **85**, 6923–6927 (1986).
55. A.P.M. Kentgens, 'Off-resonance Nutation NMR Spectroscopy of Half-integer Quadrupolar Nuclei', *Progr. Nucl. Magn. Reson. Spectrosc.*, **32**, 141–164 (1998).
56. H.W. Shin, S.H. Choh, T.H. Yeom, K.S. Hong, D.Y. Han, '1-D Nutation NMR of ^{7}Li in $LiTaO_3$ and $LiNbO_3$ Single Crystals and Powders', *J. Korean Phys. Soc.*, **32**, S662–S664 (1998).
57. A.P.M. Kentgens, J.J.M. Lemmens, F.M.M. Geurts, W.S. Veeman, 'Two-dimensional Solid-state Nutation NMR of Half-integer Quadrupolar Nuclei', *J. Magn. Reson.*, **71**, 62–74 (1987).
58. A. Samoson, E. Lippmaa, '2-D NMR Nutation Spectroscopy in Solids', *J. Magn. Reson.*, **79**, 255–268 (1988).
59. A. Samoson, E. Lippmaa, 'Excitation Phenomena and Line Intensities in High-resolution NMR Powder Spectra of Half-integer Quadrupolar Nuclei', *Phys. Rev. B*, **28**, 6567–6570 (1983).
60. P.P. Man, 'Measurement of Quadrupolar Spin Population by Solid-state NMR', *J. Magn. Reson.*, **77**, 148–154 (1988).
61. P.P. Man, 'The Two-dimensional Fourier Transform Analysis of Spin-lattice Relaxation Time', in *XXIII Congress Ampere on Magnetic Resonance*, eds. B. Maraviglia, F.D. Lucas, R. Campanella, Instituto Superiore di Sanita, Rome, 574–575, 1986.
62. G.A.H. Tijink, R. Janssen, W.S. Veeman, 'Investigation of the Hydration of Zeolite NaA by Two-dimensional ^{23}Na Nutation NMR', *J. Am. Chem. Soc.*, **109**, 7301–7304 (1987).
63. A. Samoson, E. Lippmaa, G. Engelhardt, U. Lohse, G. Jerschkewitz, 'Quantitative High-resolution Aluminium-27 NMR. Tetrahedral Nonframework Aluminium in Hydrothermally Treated Zeolites', *Chem. Phys. Lett.*, **134**, 589–592 (1987).
64. J. Klinowski, 'Nuclear Magnetic Resonance Studies of Zeolites', *Progr. Nucl. Magn. Reson. Spectrosc.*, **16**, 237–309 (1984).
65. J.M. Thomas, J. Klinowski, 'The Structure of Aluminosilicate and Related Catalysts by High-resolution Solid-state NMR Spectroscopy', *Adv. Catal.*, **33**, 199–374 (1985).
66. G. Engelhardt, D. Michel, *High Resolution Solid-state NMR of Silicates and Zeolites*, Wiley, Chichester, 1987.
67. R. Freeman, H.D.W. Hill, 'Phase and Intensity Anomalies in Fourier Transform NMR', *J. Magn. Reson.*, **4**, 366–383 (1971).
68. P.P. Man, H. Theveneau, P. Papon, 'Investigation of the Central Line of ^{93}Nb by a Two-dimensional NMR Method', *J. Magn. Reson.*, **64**, 271–277 (1985).
69. I.P. Gerothanassis, 'Methods of Avoiding the Effects of Acoustic Ringing in Pulsed Fourier Transform NMR

Spectroscopy', *Progr. Nucl. Magn. Reson. Spectrosc.*, **19**, 267–329 (1987).
70. P.P. Man, unpublished results (1998).
71. S. Zhang, X. Wu, M. Mehring, 'Elimination of the Ringing Effects in Multiple-pulse Sequences', *Chem. Phys. Lett.*, **173**, 481–484 (1990).
72. P.P. Man, 'Determination of the Quadrupolar Coupling Constant in Powdered Samples with a Two In-phase RF Pulse Sequence in Solid-state NMR', *Chem. Phys. Lett.*, **168**, 227–232 (1990).
73. M.A. Hepp, P.P. Man, A. Trokiner, H. Zanni, J. Fraissard, 'Determination of the Electric Field Gradient in $RbCaF_3$ Near the Phase Transition', *Solid State Commun.*, **84**, 869–873 (1992).
74. P.P. Man, 'Determination of Spin-$\frac{5}{2}$ Quadrupolar Coupling with Two-pulse Sequences', *Solid State NMR*, **2**, 165–180 (1993).
75. S.Z. Ageev, P.P. Man, J. Fraissard, B.C. Sanctuary, 'Determination of Quadrupolar Spin Coupling for Spin-$\frac{7}{2}$ Using Two-pulse Sequences', *Mol. Phys.*, **91**, 75–80 (1997).
76. P.P. Man, 'Double-quantum Coherence Detection in a Spin-$\frac{3}{2}$ System Excited by Two Consecutive RF Pulses', *Mol. Phys.*, **76**, 1119–1129 (1992).
77. P.P. Man, 'Measurement of the Quadrupolar Coupling with Two Pulses of Opposite Phase', *J. Magn. Reson.*, **94**, 258–267 (1991).
78. G. Bonera, M. Galimberti, 'Phase-dependence of Quadrupolar Echoes in Solids', *Solid State Commun.*, **4**, 589–591 (1966).
79. J. Butterworth, 'Spin Echoes in Solids', *Proc. Phys. Soc., London*, **86**, 297–304 (1965).
80. P.P. Man, 'Study of a Spin-$\frac{3}{2}$ System by a Spin-echo Sequence', *Mol. Phys.*, **72**, 321–331 (1991).
81. I.D. Weisman, L.H. Bennett, 'Quadrupolar Echoes in Solids', *Phys. Rev.*, **181**, 1341–1350 (1969).
82. P.P. Man, 'Solomon Echoes Extended to Soft Pulse Excitation', *Z. Naturforsch., A*, **49**, 89–96 (1994).
83. M. Mehring, O. Kanert, 'Line Shape Analysis of Spin Echo Signals in Cubic Solids', *Z. Naturforsch., A*, **24**, 768–774 (1969).
84. M. Suemitsu, N. Nakajo, 'Charged Point Defects in GaAs Crystals Evaluated by NMR Spin Echo', *J. Appl. Phys.*, **66**, 3178–3186 (1989).
85. P.P. Man, 'Importance of the Magnetic Dipolar Coupling on the Quadrupolar Spin-echo Amplitude', *J. Magn. Reson.*, **100**, 157–165 (1992).
86. A.M. Flett, J.C.S. Richards, '^{23}Na Spin Echoes in NaCl Single Crystals', *Proc. Phys. Soc.*, **86**, 171–179 (1965).
87. P.P. Man, 'Study of a Spin-$\frac{3}{2}$ System by a Spin-echo Sequence (Part 2)', *J. Chim. Phys. Phys.-Chim. Biol.*, **89**, 335–358 (1992).
88. P.P. Man, 'Mathematical Analysis of Electric-quadrupole and Heteronuclear Magnetic-dipole Interactions on Spin-$\frac{5}{2}$ Hahn Echo Amplitudes in Solids', *J. Magn. Reson. A*, **114**, 59–69 (1995).
89. P.P. Man, E. Duprey, J. Fraissard, P. Tougne, J.-B. d'Espinose, 'Spin-$\frac{5}{2}$ Echoes in Solids', *Solid State NMR*, **5**, 181–188 (1995).
90. G.N. Abelyashev, V.N. Berzhanskij, N.A. Sergeev, Y.V. Fedotov, 'Multiquantum Effects and NMR in Magnetically Ordered Substances', *Phys. Lett. A*, **133**, 263–265 (1988).
91. D.G. Hughes, L. Pandey, 'Spurious Signals Caused by the Piezoelectric Ringing of $NaNO_2$ in Pulsed NMR', *J. Magn. Reson.*, **56**, 428–442 (1984).
92. P.P. Man, 'Study of a Spin-$\frac{3}{2}$ System by a Quadrupolar-echo Sequence: Suppression of Spurious Signals', *Solid State NMR*, **1**, 149–158 (1992).
93. A.C. Kunwar, G.L. Turner, E. Oldfield, 'Solid-state Spin-echo Fourier Transform NMR of ^{39}K and ^{67}Zn Salts at High-field', *J. Magn. Reson.*, **69**, 124–127 (1986).
94. M.A. Hepp, P.P. Man, J. Fraissard, 'Excitation Conditions for Quantitative Determination of Quadrupolar Spins with a Spin-echo Sequence; Cancellation of Spurious Signals', in *Application of NMR Spectroscopy to Cement Science*, eds. P. Colombet, A.R. Grimmer, Gordon and Breach, Yverdon, Switzerland, 455–466, 1994.
95. Y. Dumazy, J.-P. Amoureux, C. Fernandez, 'Theoretical and Experimental Study of Quadrupolar Echoes in Solid State NMR', *Mol. Phys.*, **90**, 959–970 (1997).
96. R. Kimmich, *NMR: Tomography, Diffusometry, Relaxometry*, Springer-Verlag, Berlin, 1997.
97. C.A. Fyfe, J. Skibsted, H. Grondey, H.M.z. Altenschildesche, 'Pulsed Field Gradient MQ MAS NMR Spectroscopy of Half-integer Spin Quadrupolar Nuclei', *Chem. Phys. Lett.*, **281**, 44–48 (1997).
98. A. Medek, J.S. Harwood, L. Frydman, 'Multiple Quantum MAS NMR: a New Method for the Study of Quadrupolar Nuclei in Solids', *J. Am. Chem. Soc.*, **117**, 12 779–12 787 (1995).
99. J.-P. Amoureux, C. Fernandez, L. Frydman, 'Optimized MQ MAS NMR Experiments on Half-integer Quadrupoles', *Chem. Phys. Lett.*, **259**, 347–355 (1996).
100. J.-P. Amoureux, 'High-resolution Solid-state NMR for Spin-$\frac{3}{2}$ and $\frac{9}{2}$: the Multi-quantum Transition Method', *Solid State NMR*, **2**, 83–88 (1993).
101. A. Wokaun, R.R. Ernst, 'Selective Detection of Multiple Quantum Transitions in NMR by 2D Spectroscopy', *Chem. Phys. Lett.*, **52**, 407–412 (1977).
102. D. Massiot, B. Touzo, D. Trumeau, J.P. Coutures, J. Virlet, P. Florian, P.J. Grandinetti, 'Two-dimensional MAS Isotropic Reconstruction Sequences for Quadrupolar Nuclei', *Solid State NMR*, **6**, 73–83 (1996).
103. D. Massiot, 'Sensitivity and Lineshape Improvements of MQ/MAS by Rotor-synchronized Data Acquisition', *J. Magn. Reson. A*, **122**, 240–244 (1996).

104. M. Hanaya, R.K. Harris, 'Optimization of 2D MQ/MAS NMR Experiments for $I = 3/2$ Nuclei on a Moderate-field Spectrometer', *J. Phys. Chem. A*, **101**, 6903–6910 (1997).
105. S.H. Wang, Z. Xu, J.H. Baltisberger, L.M. Bull, J.F. Stebbins, A. Pines, 'MQ MAS and DAS NMR Spectroscopy of Quadrupolar Nuclei', *Solid State NMR*, **8**, 1–16 (1997).
106. J.-P. Amoureux, C. Fernandez, 'Triple, Quintuple and Higher Order Multiple Quantum MAS NMR of Quadrupolar Nuclei', *Solid State NMR*, **10**, 211–223 (1998).

Relaxation in Nuclear Magnetic Resonance, General

Daniel Canet and Pierre Mutzenhardt
Université Henri Poincaré, Nancy, France

1 Introduction 1

2 Nuclear Magnetic Resonance Relaxation Parameters and Their Determination 2
- 2.1 Longitudinal Relaxation 2
- 2.2 Transverse Relaxation 5
- 2.3 The Rotating Frame Relaxation Time $T_{1\rho}$ 8
- 2.4 Cross-relaxation Rates – Nuclear Overhauser Enhancement Factors 8
- 2.5 Cross-correlation Rates 13

3 Relaxation Mechanisms 14
- 3.1 Interaction with Local Fluctuating Magnetic Fields 14
- 3.2 Overview of the Various Interactions Responsible for Relaxation Phenomena 16
- 3.3 Relaxation by Chemical Shift Anisotropy 16
- 3.4 Scalar Relaxation of the Second Kind 17
- 3.5 Reduced Spectral Densities 18
- 3.6 Relaxation in Coupled Spin Systems – Redfield Equations 19
- 3.7 Longitudinal Relaxation in a Weakly Coupled Two-spin-1/2 System 19
- 3.8 The Nuclear Overhauser Effect 20
- 3.9 The Chemical Shift Anisotropy–Dipolar Interference Term 21
- 3.10 Quadrupolar Relaxation 21
- 3.11 Paramagnetic Relaxation 22

4 Influence of Spin Relaxation on Quantitative Measurements 22
- 4.1 Sensitivity Improvement and Repetition Rate 22
- 4.2 Polarization Transfer in Heteronuclear Systems 23
- 4.3 Overlapping Patterns and Spectral Analysis 23
- 4.4 Partial Decoupling by Quadrupolar Nuclei 24

Abbreviations and Acronyms 25

Related Articles 26

References 26

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Relaxation phenomena in nuclear magnetic resonance (NMR) govern the recovery of nuclear magnetization after it has been taken from its equilibrium state (i.e. along the static magnetic field direction). Various interactions, leading to local fluctuating magnetic fields, are responsible for the timescale of these phenomena. By themselves, these interactions yield valuable dynamic and structural information. Conversely, the influence of relaxation on the quantitative aspect of NMR spectra has to be assessed prior to any experiment aiming at using NMR as an analytical tool. This implies an evaluation not only of the relaxation parameters but also of indirect relaxation processes such as the nuclear Overhauser effect (NOE).

1 INTRODUCTION

Relaxation in NMR is a generic term for phenomena which bring the nuclear magnetization back to its equilibrium configuration after it has been perturbed. An ensemble of nuclear spins, in the presence of a static magnetic field B_0, leads to a macroscopic magnetization aligned with B_0 (this direction is usually called longitudinal, while quantities relating to a plane perpendicular to B_0 are usually called transverse). Obtaining an NMR spectrum implies taking the nuclear magnetization into a transverse configuration. This is one possible perturbation but, as will be seen later, many others exist. It can be stated at the outset, however, that longitudinal and transverse magnetizations behave independently, the former recovering to the equilibrium magnetization according to a longitudinal relaxation time T_1 (also called spin–lattice relaxation time), the latter disappearing according to a transverse relaxation time T_2 (also called spin–spin relaxation time). However, in the case of common liquids, T_1 and T_2 are usually equal and lie in the range 100 ms–30 s. Conversely, in solids, T_2 may be short – some tens of microseconds – whereas T_1 may become very long – several hundred seconds. Although the consideration of only two relaxation times is a simplistic view, it can help in understanding the basics of the complicated aspects of spin relaxation. Even if one is simply interested in the NMR spectrum itself, which, by contrast to optical spectroscopies, is known to be quantitative (since it reflects in principle the nuclear magnetization, and thus the spin concentration), one has to account for relaxation phenomena in order to use NMR as an analytical tool. Historically, as reported by Slichter in one of the best known books on NMR,[1] the very first attempt to detect a nuclear resonance was

hampered by the choice of a solid material whose T_1 was exceedingly long. In a general way and regardless of the detection mode of the NMR signal, the peak line width at half-height $\Delta\nu$ is controlled by the transverse relaxation time T_2 (also termed spin–spin relaxation time) according to $\Delta\nu = 1/\pi T_2$, which indicates that short T_2s will be responsible for a lack of resolution and may require some numerical procedures in order to separate overlapping peaks. In the early days of NMR, when spectra were obtained by continuous wave (CW) methods, it was mandatory to avoid peak saturation (so as to obtain meaningful results); saturation was controlled by the level of the applied radiofrequency (RF) field (the alternating magnetic field needed for inducing transitions) and by both T_1 and T_2. Nowadays, NMR experiments are performed by RF pulse methods which yield a so-called free induction decay (FID) or interferogram. As Fourier transformation of the FID leads to the conventional spectrum, this form of spectroscopy is usually referred to as Fourier transform nuclear magnetic resonance (FTNMR).[2] One of its major interests lies in the ability to improve efficiently the signal-to-noise ratio (S/N). Indeed, it is well known that NMR is characterized by poor sensitivity compared with most spectroscopies, and it is often necessary to repeat the experiment and to add coherently the relevant FIDs. Again, this accumulation procedure is governed by T_1 and some care has to be exercised in order to obtain quantitative results. Some subtler effects, such as cross-relaxation, which manifests itself through the NOE, must be considered when thoroughly analyzing an NMR spectrum. This will be discussed later in the case of multispin systems.

It must be stressed that relaxation parameters are interesting in themselves because they arise from the various interactions to which nuclear spins are subjected, the most common being the classical dipolar interaction (between the magnetic dipoles associated with nuclear spins). In order to be active, these interactions must be time dependent, e.g. via molecular motions. Therefore, relaxation parameters encompass not only structural information (through the interactions themselves) but also dynamic information (through this time modulation). This latter information can be about (i) overall motions, translational or rotational, governed essentially by the molecular volume and the viscosity of the medium in which the molecules under investigation are embedded, (ii) local motions such as internal rotations around C–C bonds, or segmental motions in molecules or aggregates of appreciable size and (iii) exchange between two different molecules or between two distinct sites within the same molecule. Before considering in detail the various relaxation mechanisms, we shall first discuss experimental methods which yield the classical relaxation parameters with accuracy and reliability.

2 NUCLEAR MAGNETIC RESONANCE RELAXATION PARAMETERS AND THEIR DETERMINATION

2.1 Longitudinal Relaxation

Any experimental method aiming at the measurement of a dynamic parameter invariably starts by a perturbation which moves the system out of its equilibrium state; this is followed by an evolution period whose duration is generally denoted by τ, which is such that the dynamic parameters of interest manifest themselves. The ultimate stage consists of reading the actual state of the system. Generally, this measurement is performed for different durations of the evolution period so as to evaluate the dynamic parameters properly. Regarding longitudinal relaxation, it seems obvious that the first two stages (initial perturbation and evolution) should involve only longitudinal magnetization. However, the read stage necessarily implies transverse components which are the only ones giving rise to a detectable signal. In other words, longitudinal magnetization, after having evolved, must be converted into transverse magnetization. The efficiency of the whole process will be optimal, thus providing the greatest accuracy for the determination of the considered relaxation parameter, provided that the initial perturbation is maximum. For longitudinal relaxation, this corresponds to a *complete inversion* by means of a 180° (or π) RF pulse, in such a way that magnetization *remains longitudinal* even during the evolution period. Subsequent to the evolution period, the ongoing value of the longitudinal magnetization is read with a 90° (or $\pi/2$) pulse which converts it into transverse magnetization, directly detectable. This is the well-known *inversion–recovery*[3] experiment, which is shown schematically in Figure 1(a) and (b).

At this point, it may be instructive to recall briefly how an RF pulse works. Let us denote by ν_r the transmitter frequency and by ν_0 the resonance frequency corresponding to a given peak in the NMR spectrum. It is convenient to define a frame rotating at a frequency ν_r around the z-axis which is the direction of the static magnetic field B_0. In this rotating frame, it can be shown that the RF field B_1 acts as if it is stationary, say along x. Furthermore, if the condition $(\gamma B_1/2\pi) \gg |\nu_r - \nu_0|$ is fulfilled (the gyromagnetic ratio γ is a constant characteristic of the considered isotope), the sole magnetic field which is able to act on nuclear magnetization is B_1, so that a precessional motion around B_1 occurs at an angular velocity γB_1 (in the same way as transverse magnetization precesses at an angular velocity γB_0 – the so-called Larmor frequency – in the presence of a magnetic field B_0). Thus starting from magnetization along z, an RF field applied for a time τ_{90}, such that

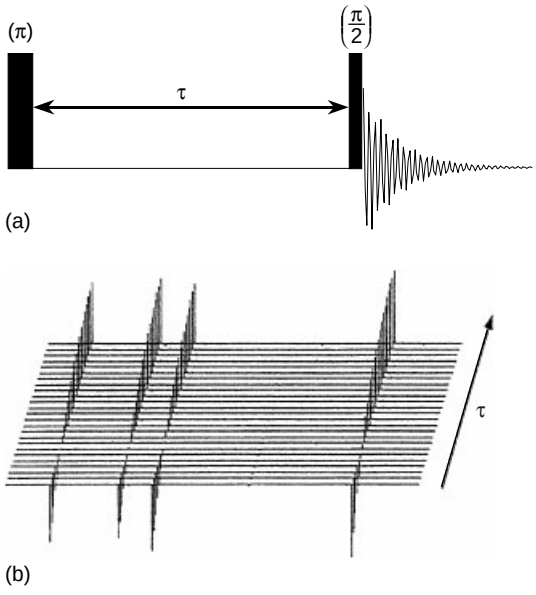

Figure 1 (a) Scheme of an inversion–recovery experiment yielding the longitudinal relaxation time [inversion is achieved by means of the (π) RF pulse, represented by a filled vertical rectangle]. FIDs (represented by a damped sine function) resulting from the ($\pi/2$) read pulse are subjected to Fourier transformation and lead to a series of spectra corresponding to the different τ values (evolution period). Spectra are generally displayed with a shift between two consecutive values of τ. The analysis of the amplitude evaluation of each peak from $-M_0$ to M_0 provides an accurate evolution of T_1. (b) The example concerns carbon-13 T_1 of *trans*-crotonaldehyde, the values of which from left to right are 20.5, 19.8, 23.3 and 19.3 s.

$\gamma B_1 \tau_{90} = \pi/2$, takes the magnetization on to the y-axis, whereas if it is applied for twice this duration it produces a complete inversion.

In a first approach, we shall limit ourselves to the hypothesis where nuclear magnetization conforms to Bloch equations. Concerning the longitudinal component, one has Equation (1):

$$\frac{\mathrm{d}M_z}{\mathrm{d}t} = -\frac{M_z - M_0}{T_1} \tag{1}$$

where M_0 is the equilibrium magnetization. This first-order differential equation can easily be solved to yield Equation (2):

$$M_z(\tau) = M_0 + [M_z(0) - M_0] \exp\left(\frac{-\tau}{T_1}\right. \tag{2}$$

We can verify that the right-hand side of Equation (2) is equal to $M_z(0)$ for $\tau = 0$ and to M_0 for $\tau \to \infty$. For the ideal case of a perfect inverting (180°) pulse, $M_z(0) = -M_0$ and Equation (2) reduces to Equation (3):

$$M_z(\tau) = M_0 \ \ 1 - 2\exp\left(\frac{-\tau}{T_1}\right. \tag{3}$$

or, in a logarithmic form, Equation (4):

$$\ln \ \frac{M_0 - M_z(\tau)}{2M_0} \ = -\frac{\tau}{T_1} \tag{4}$$

This latter form is especially convenient for a quick determination of T_1 which is the inverse of the slope of the linear representation of $\ln\left\{[M_0 - M_z(\tau)]/2M_0\right\}$ vs τ (Figure 2a and b). M_0 can be measured by means of a single read-pulse or for a time τ of the order of $5T_1$ (which insures a return to equilibrium of more than 99% of the magnetization).

It turns out that, in the case of an imperfect inverting pulse, the factor of 2 in Equation (3) must be substituted by an unknown factor ($K < 2$);[4] it is therefore recommended to turn to a nonlinear fit (based on the criterion of least-squares) (Equation 5):

$$M_z(\tau) = M_0 \ \ 1 - K\exp\left(\frac{-\tau}{T_1}\right. \tag{5}$$

where the three quantities M_0, K and T_1 have to be refined, starting for example from values deduced from

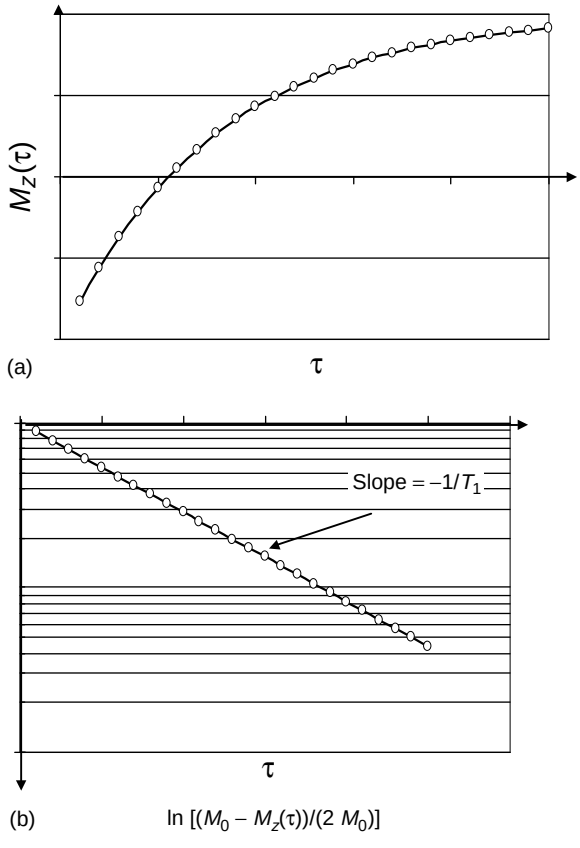

Figure 2 (a) Nonlinear fit of the data corresponding to an inversion–recovery experiment. (b) Another possible data analysis involving a semilogarithmic plot which leads directly to the longitudinal relaxation time.

Equation (4). The quality of inversion can be greatly improved by relying on the concept of *composite pulses*, which consists of replacing a single pulse by a *cluster* whose goal is to correct in a self-consistent way and, at least to first order, the various pulse imperfections. In the present context, an imperfection of the read-pulse does not matter; however, a composite pulse can be used to improve the efficiency of the initial inverting pulse, either if its duration has been misadjusted or if, because of spatial inhomogeneity of the RF field, the flip angle is not exactly 180° in all regions of the sample. Let us suppose that the inverting pulse is applied along the x-axis of the rotating frame, self-compensation can be seen to be effective if we resort to the cluster $(\pi/2)_x(\pi)_y(\pi/2)_x$,[5] where $(\pi/2)_x$ may be considered as a pulse of flip angle $90° - \varepsilon$ (whose duration is underestimated by the experimenter or due to B_1 inhomogeneity). We shall nevertheless assume that the $(\pi)_y$ pulse (acting along the y-axis of the rotating frame; this amounts to change by 90° the transmitter phase) is devoid of imperfections; this constitutes the "first-order approximation". From the magnetization motion sketched in Figure 3(a) and (b), we can recognize that the imperfection represented by ε has been removed.

For this type of composite pulse, neither imperfections of the central $(\pi)_y$ pulse nor possible off-resonance effects (which occur if the condition $(\gamma B_1/2\pi) \gg |\nu_r - \nu_0|$ is not fully satisfied) have been considered. It has been demonstrated empirically that both problems are attenuated by choosing a $(4\pi/3)_y$ pulse in place of the $(\pi)_y$ pulse, so that an efficient composite inverting pulse can be schematized as $(\pi/2)_x(4\pi/3)_y(\pi/2)_x$.[6] Finally, it must be stressed that the T_1 measurement, as described above, is meaningful only under the condition that the system returns to equilibrium between two experiments with different τ values, or between two consecutive scans if accumulation is necessary for improving the S/N (this requires a waiting time of the order of $5T_1$). Variants of the inversion–recovery method have been proposed for which this latter condition is lifted, thus avoiding prohibitively long measuring times.[7] The evolution of the longitudinal magnetization is simply given by Equation (5) as in the case of an imperfect inversion pulse. K depends

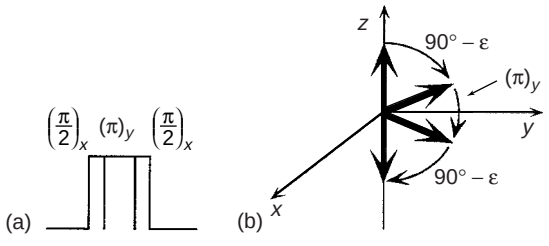

Figure 3 Compensations of pulse imperfections by means of an inverting "composite pulse".

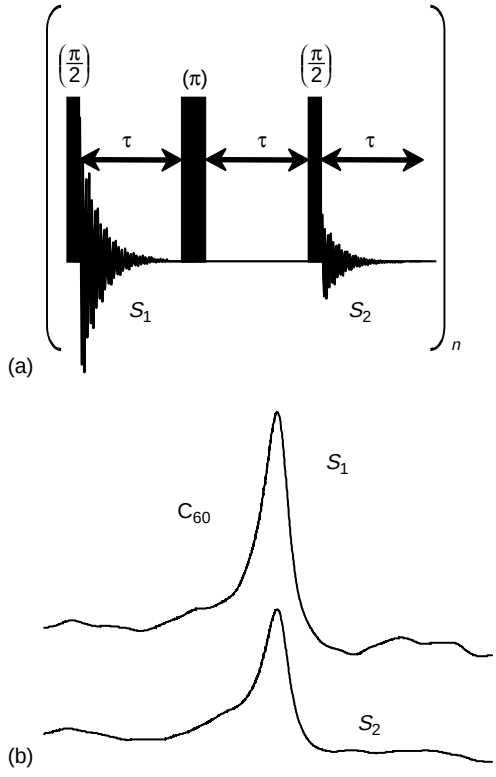

Figure 4 The SUFIR experiment. This experiment can be seen as a comparison of a reference spectrum (S_1) and of a partially relaxed spectrum (S_2). It can be used in extreme conditions: the two spectra correspond to S_1 and S_2 of a dilute species (1 mg of C_{60} fullerene dissolved in 1 mL of carbon disulfide) possessing an unusually long relaxation time ($T_1 = 153$ s; τ was set at a value of the same order of magnitude).

here on the ratio T/T_1 (where T is the repetition time) and becomes an adjustable parameter. It is even possible to go further by using the experiment sketched in Figure 4(a) and (b), denoted super fast inversion–recovery (SUFIR).[8]

The interval τ includes the acquisition of two different FIDs S_1 and S_2, which yields T_1 through Equation (6):

$$T_1 = \frac{\tau}{1 - \ln(S_2/S_1)} \tag{6}$$

The sequence is repeated n times so as to reach the desired S/N.

An interesting alternative to the inversion–recovery experiment (Figure 5) is the saturation–recovery experiment, which proceeds by an initial saturation of the nuclear magnetization (total cancellation by application of long scrambling RF pulses). In that case, the initial condition is $M_z(0) = 0$ instead of $-M_0$ and the evolution is governed by Equation (7):

$$M_z(\tau) = M_0 \quad 1 - \exp\left(\frac{-\tau}{T_1}\right. \tag{7}$$

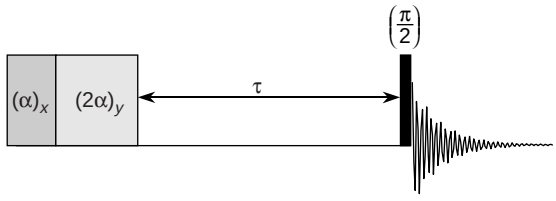

Figure 5 The saturation–recovery experiment. Saturation is achieved by the α pulses whose length lies in the millisecond range.

Although the dynamic range is reduced by a factor of two, there are two distinct advantages as neither a waiting time nor an adjustment of the π pulse is needed. It should be emphasized that this method becomes valuable when magnetization cannot be inverted in the case of a very short T_2 (because of relaxation taking place during the pulse) or in the case of the so-called "radiation damping"[9] (this occurs when the NMR signal is strong and induces a current in the detection coil capable of producing an alternating magnetic field which counteracts the magnetization motion, thus tending to bring it back to $+z$; this phenomenon is becoming very common with high-field sensitive spectrometers and concerns highly concentrated species, e.g. the solvent).

2.2 Transverse Relaxation

As stated before, the initial perturbation should be maximum with respect to the equilibrium state. Since we are dealing here with transverse magnetization, this maximum perturbation is obviously a 90° pulse. However, it can be immediately noticed that signals collected after a simple read-pulse decay exponentially according to a time constant T_2^* which differs from the genuine T_2 by a contribution due to the static induction B_0 inhomogeneity (Equation 8):

$$\frac{1}{T_2^*} = \frac{1}{T_2} + (B_0 \text{ inhomogeneity}) \tag{8}$$

Experimental methods must therefore be devised for removing this contribution. To reach this goal, an ingenious procedure was proposed by Hahn[10] in the early 1950s, soon after the first NMR experiments. It consists of the so-called echo sequence which involves a 180° pulse in the middle of an interval of duration 2τ. To understand how this sequence works, let us look at the events occurring after the application of the 90° pulse. Whenever nuclear magnetization is taken to the transverse plane, it precesses at the frequency given by Equation (9):

$$\nu_0 = \frac{\gamma B_0}{2\pi}(1 - \sigma) \tag{9}$$

where σ is the shielding coefficient, which defines the chemical shift, that is, the differentiation of various nuclei of the same isotopic species as a function of the electronic distribution by which they are surrounded (it is this feature which makes NMR such a powerful technique for identifying nuclei, e.g. protons, belonging to different chemical environments). A coil whose axis is in the transverse plane is therefore capable of detecting a signal [through an induced electromotive force (EMF)] oscillating at the frequency ν_0. Now, if this signal is detected with respect to a reference frequency precisely equal to ν_r (the frequency of the rotating frame, already invoked for describing the effects of an RF pulse, corresponding to the experimental arrangement used in most spectrometers), precession actually occurs at a low frequency $|\nu_0 - \nu_r|$. Of course, the signal is attenuated by T_2^* and one ends up with a damped sine or cosine function of the form $M_0 \cos[2\pi(\nu_0 - \nu_r)t] \exp(-t/T_2^*)$, assuming a (crude) exponential decay by the B_0 inhomogeneities, a situation generally encountered in the liquid phase.

Let us now suppose that the initial 90° RF pulse acts along the x-axis of the rotating frame taking the magnetization to the y-axis and that after a time τ we apply a 180° pulse along the y-axis as shown in Figure 6; this has the virtue of reversing any chemical shift effect so that after an additional time τ, the magnetization again lies along the y-axis and a signal reappears in the form of an echo.

This pulse sequence, $(\pi/2)_x$–τ–$(\pi)_y$–τ–acquisition, is the basic Hahn sequence, which in principle yields the true T_2 since any precession effect is removed (in particular, precession due to B_0 inhomogeneities), leaving a transverse magnetization attenuated according to the transverse relaxation time. By reference to the

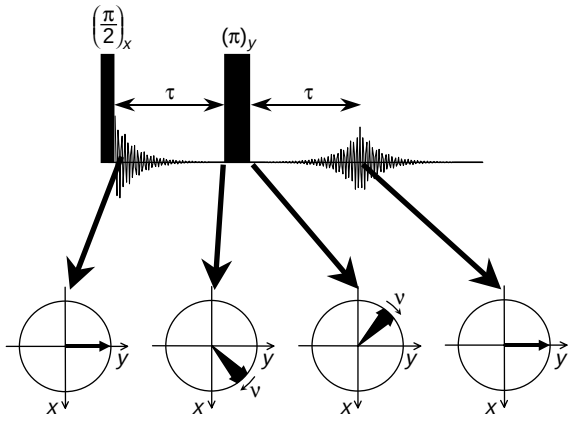

Figure 6 The basic Hahn sequence for the measurement of the transverse relaxation time T_2. Any precession motion characterized by the frequency ν in the rotating frame is refocused. This precession may arise either from chemical shift or from B_0 inhomogeneity (dashed area).

Bloch equations relative to transverse magnetization (Equation 10):

$$\frac{\mathrm{d}}{\mathrm{d}t}M_{x,y} = -\frac{M_{x,y}}{T_2} \tag{10}$$

one can recognize that the Fourier transform of the half-echo leads to a signal of amplitude $M_0 \exp(-2\tau/T_2)$. For a set of τ values, it therefore appears possible to extract an accurate value of T_2 for each line in the spectrum. Unfortunately, this analysis does not take into account translational diffusion phenomena, which will be now considered.

We shall assume that the B_0 induction is not perfectly homogeneous; for simplicity and without loss of generality, we shall make the hypothesis that it varies linearly across the whole sample in the X direction of the laboratory frame so that the induction sensed by a molecule at abscissa X is of the form (Equation 11):

$$B(X) = B_0 + g_0 X \tag{11}$$

where g_0 is a *uniform gradient* of the static field. Let us recall that such a gradient can be purposely created with the aim of producing NMR images or measuring the translational diffusion coefficient.[11] This gradient can also result from a nonexistent or incomplete compensation of the genuine inhomogeneity of B_0 (imperfect *shimming*; shims are additional coils used for correcting the main field B_0 inhomogeneity). In any event, the precession frequency depends on the location of the considered molecule via the spatial dependence of B_0. We have just seen that any precession effect is refocused by an echo sequence whose goal is precisely to get rid of B_0 inhomogeneity. However, this feature is impaired if, during the refocusing process, molecules translate from a location X to a location X' for which the precession frequency is different from that at location X. This should produce a further signal attenuation or an additional ''defocusing'' which arises from self-diffusion phenomena. The latter are characterized by a self-diffusion coefficient D and result in Equation (12):[11,12]

$$M_{x,y}(2\tau) = M_0 \exp\left(\frac{-2\tau}{T_2} \quad \exp \quad -\left(\frac{D\gamma^2 g_0^2}{3} \quad (2\tau^3)\right)\right) \tag{12}$$

It can be recognized that the compensation of all precession effects is still accounted for in Equation (12). It can also be stressed that translational diffusion in the presence of a gradient generates a decay depending on τ^3, whereas transverse relaxation produces a decay depending on τ. The idea of Carr and Purcell[13] was to minimize the effects of diffusion in a T_2 experiment. This goal is achieved by the train of 180° pulses in Figure 7 and stems from the fact that the differential equation must be solved for each interval following a 180° pulse since, at that point, the magnetization sign is suddenly modified; consequently, new boundary conditions prevail.

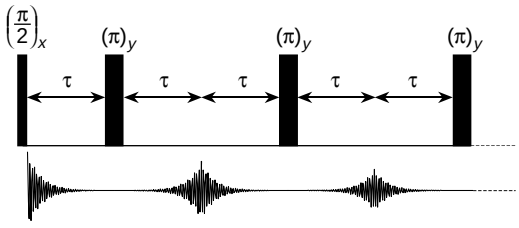

Figure 7 The CPMG (Carr–Purcell–Meiboom–Gill) pulse sequence. An echo is formed half way between two consecutive π pulses. The echo amplitude (or the Fourier transform of the half-echo) provides an evaluation of T_2 less affected by translational diffusion than in the simple Hahn sequence. The phase change of π pulses with respect to the initial $\pi/2$ pulse cancels the effect of (π) pulse imperfections.

Returning to the analysis of the Hahn sequence and extending it to the present experiment, we obtain the amplitude of the nth echo (Equation 13):

$$M_{x,y}(2n\tau) = M_0 \exp\left(\frac{-2n\tau}{T_2} \quad \exp \quad -\left(\frac{D\gamma^2 g_0^2}{3} \quad (2n\tau^3)\right)\right) \tag{13}$$

We note that for an evolution whose duration is identical with that of a Hahn sequence, the argument of the exponential relevant to diffusion has been divided by n^2. Therefore, the remedy to make negligible translational diffusion effects consists simply in increasing n, which amounts to bringing the π pulses closer to each other.

The trick introduced by Meiboom and Gill[14] is to *dephase* all π pulses in the Carr–Purcell train by an angle of 90° with respect to the initial $\pi/2$ pulse. As shown in Figure 8, without this phase change, imperfections of the π pulses are cumulative, whereas with the 90° phase change, self-compensation occurs for all echoes of even number.

The CPMG experiment can be handled in two ways. First, if the spectrum involves only one resonance (or if line widths do not allow for the separation of several resonances), a *single experiment* can be run with acquisition of the amplitude of each echo along the pulse train (for sensitivity enhancement, accumulations can be carried out). This experiment is especially valuable for determining the relative proportions of two species which differ by their transverse relaxation time, for instance the two types of water (free and bound) in various materials (wood is a good example) or water and lipids in foodstuffs (Figure 9). For this type of measurement, a ''low-resolution'' spectrometer (without any shim system) is sufficient.

Second, if the spectrum involves several well-resolved resonances, one can proceed by Fourier transformation of the half-echo (possibly with signal accumulation). As many experiments as necessary are performed by varying

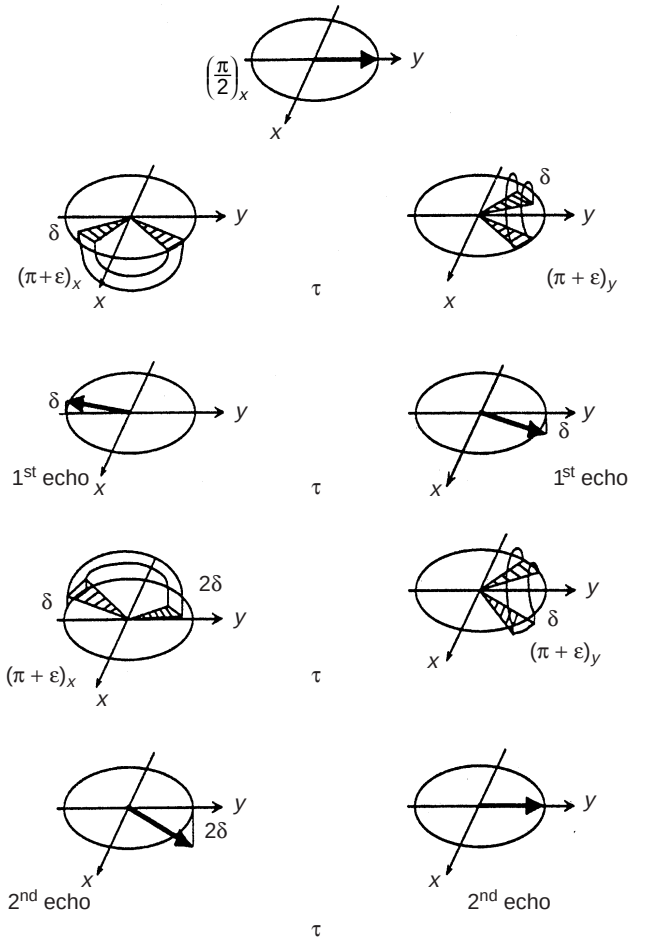

Figure 8 π pulse imperfections $(\pi + \varepsilon)$ are figured out by a flip angle in excess by δ. Left: errors cumulate in the simple Carr–Purcell sequence when π pulses have the same phase as the initial $\pi/2$ pulse. Right: self-compensation at each even-numbered echo by simply shifting the phase of all π pulses by 90° (CPMG sequence).

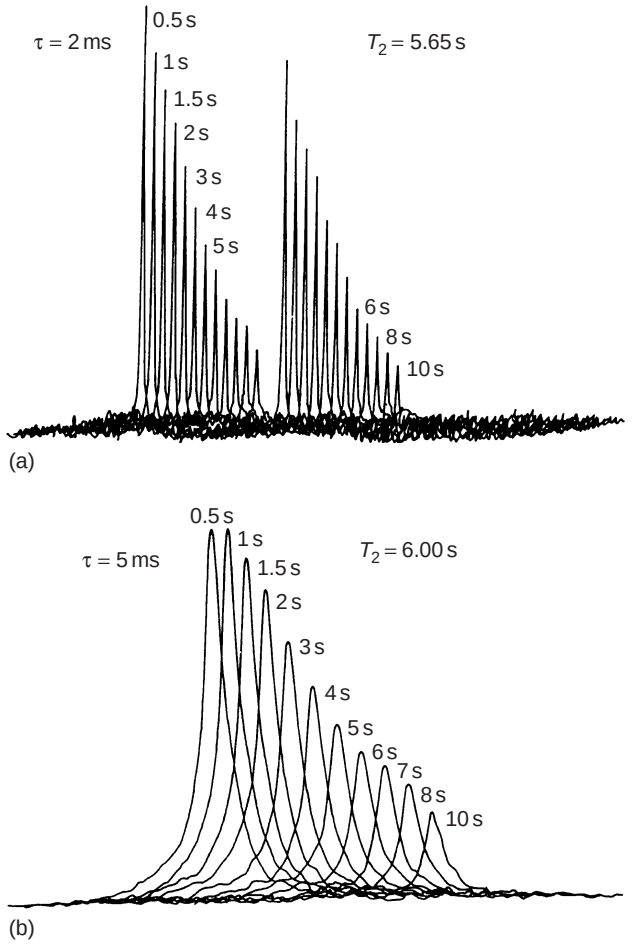

Figure 10 (a) Measurement of the transverse relaxation time in the carbon-13 doublet of formic acid (no proton decoupling) according to the sequence in Figure 7 with $\tau = 2$ ms and for successive values of n (the value of $2n\tau$ is given above each spectrum). (b) Same as (a) except that proton decoupling was applied only during signal acquisition and $\tau = 5$ ms.

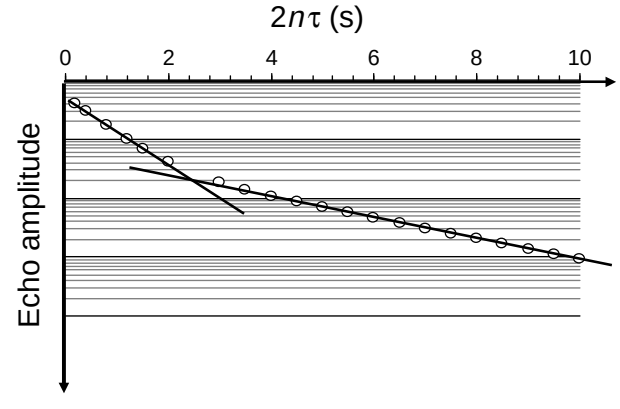

Figure 9 Biexponential evolution (in fact presented as a semilogarithmic plot) of the amplitude of echoes resulting from a CPMG pulse train in the case of a sample which involves two species differing by their transverse relaxation time.

n so that the successive values of $2n\tau$ (see Equation 13) lead to T_2 with the required accuracy (Figure 10a and b).

In any event, a time of ca. $5T_1$ must elapse between two consecutive experiments in order to allow for a complete return of the system to thermal equilibrium. In a general way, the spectroscopist must be aware of the difficulties associated with the determination of transverse relaxation times. In homonuclear coupled spin systems, the magnetization decay is affected by "J-modulation".[15] It can be recalled that, in addition to chemical shifts, high-resolution NMR spectra (of liquid sample) are characterized by a fine structure arising from the so-called J coupling (or indirect coupling or scalar coupling) which occurs via the spins of bonding electrons. It turns out that a 180° pulse does not refocus J coupling precession in a homonuclear system, hence the echo amplitude varies as a function of τ according to the J

value, making the analysis more difficult. Conversely, in the case of a heteronuclear system with observation of the rare nucleus (e.g. carbon-13), heteronuclear J coupling (e.g. with protons) is effectively refocused. On the other hand, as discussed later, a common practice in carbon-13 spectroscopy is to remove completely the effects of heteronuclear J couplings by strongly irradiating the proton transitions (proton broad-band decoupling).[6] However, it is strongly recommended that proton decoupling be avoided during the evolution period. This is because decoupling cannot be infinitely perfect and thus induces an extra line broadening which behaves like a transverse relaxation mechanism; it also induces some heating of the solution and sample vibrations. It therefore leads to apparent T_2 much shorter than expected. Of course, proton decoupling can be switched on for signal acquisition in order to deal with a simpler spectrum. Rotation and vibration of the sample should be avoided to prevent irreversible defocusing, which would preclude accurate measurement and variable-temperature experiments should be carried out under gas flow control.[16]

2.3 The Rotating Frame Relaxation Time $T_{1\rho}$[17]

Most of the drawbacks mentioned above can be circumvented by a simple experiment, although somewhat instrumentally demanding. The sequence is depicted in Figure 11. First, a standard $(\pi/2)_x$ flips the nuclear magnetization toward the y-axis of the rotating frame; thereafter an RF field B_1 is applied along that direction for a duration τ. This RF field must be sufficiently strong to avoid off-resonance effects but not too strong to prevent probe deterioration. During the τ interval, magnetization should nutate around B_1; since both are collinear, magnetization is stationary along the y-axis of the rotating frame. It is said to be locked, hence the terminology of ''spin-lock'' associated with this experiment.

Any modification of the magnetization thus arises from relaxation phenomena. The transverse magnetization

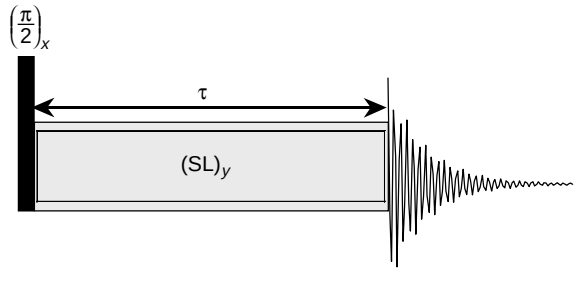

Figure 11 Principle of a spin-lock experiment leading to the determination of the relaxation time in the rotating frame ($T_{1\rho}$). $(SL)_y$ stands for the spin-lock period which corresponds to the application of an RF field along the y-axis of the rotating frame.

spin-locked along B_1 must end up at its thermal equilibrium value, that is, zero. The corresponding evolution is exponential with a time constant denoted by $T_{1\rho}$ (relaxation time in the rotating frame), very close to (if not identical with) T_2. In practice, the signal is measured (and subsequently Fourier transformed) for a set of τ values, in successive experiments, and obeys Equation (14):

$$S(\tau) = S_0 \exp\left(\frac{-\tau}{T_{1\rho}}\right) \tag{14}$$

where S_0 is the signal amplitude immediately after the $(\pi/2)_x$ pulse. An immediate advantage of the method is the absence of precession during the evolution period and thus, among other things, the absence of J modulation.

2.4 Cross-relaxation Rates – Nuclear Overhauser Enhancement Factors

Experimental methods outlined in the preceding section concern systems which obey Bloch equations and which consequently exhibit monoexponential evolutions according to a time constant T_1 for longitudinal magnetization and T_2 for transverse magnetization. Although in some instances Bloch equations are either strictly valid or represent a good approximation, there exist situations where biexponentiality shows up obviously and it would be a shame not to exploit this feature. Such a situation occurs whenever two spins A and X interact by a time-dependent mechanism. Such a process is termed *cross-relaxation* and means that any modification of the A magnetization induces a modification of X magnetization which adds up to the specific evolution of X magnetization; the symmetrical process, between X and A, of course, also holds. More formally, this coupling can be expressed via the well-known Solomon equations,[18] which are written as Equations (15) and (16) for the two longitudinal magnetizations I_z^A and I_z^X (as already mentioned, any coupling between longitudinal and transverse magnetizations is impossible):

$$\frac{dI_z^A}{dt} = -R_1^A(I_z^A - I_{eq}^A) - \sigma(I_z^X - I_{eq}^X) \tag{15}$$

$$\frac{dI_z^X}{dt} = -R_1^X(I_z^X - I_{eq}^X) - \sigma(I_z^A - I_{eq}^A) \tag{16}$$

where R_1^A represents the specific longitudinal relaxation rate of spin A ($R_1^A = 1/T_1^A$), σ is the *cross-relaxation* rate, which reflects the coupling between the two magnetizations alluded to above, and I_{eq}^A is the equilibrium magnetization. It can be seen that the Solomon equations are in fact Bloch equations to which the cross-relaxation rate σ has been appended. σ may have two origins:

1. The dipolar interaction between the two nuclei A and X, modulated by molecular motions. This contribution is very interesting with regard to the information that it contains and will be treated more explicitly in section 3.7. For the moment, it is important to know that it is proportional to $(r_{AX})^{-6}$, where r_{AX} is the internuclear distance between A and X; this will therefore tell us if the two spins are close to each other (this is, of course, essential information in terms of molecular structure or intermolecular proximity).
2. A chemical exchange process[19] such as spins move from site A to site X and conversely from site X to site A. If we define as τ the residence time in each site (this implies identical concentrations for A and X, thus in accord precisely with the situation considered for the two spins A and X), then $\sigma = -k = -1/\tau$. Moreover, the exchange rate k must be added to each specific rate R_1^A and R_1^X.

We shall be concerned here with the determination of the three dynamic parameters R_1^A, R_1^X and σ involved in Equations (15) and (16), deferring to section 3 their interpretation at a molecular level. A first approach consists of examining the whole evolution of I_z^A and I_z^X and, by a nonlinear analysis, extracting the three considered parameters. Because we are dealing with two simultaneous differential equations, this evolution is biexponential in nature (for both I_z^A and I_z^X). After solving the relevant equations, we obtain Equations (17) and (18):

$$\frac{I_{eq}^A - I_z^A(t)}{2I_{eq}^A} = a_1 \exp(\lambda_1 t) + a_2 \exp(\lambda_2 t) \tag{17}$$

$$\frac{I_{eq}^X - I_z^X(t)}{2I_{eq}^X} = x_1 \exp(\lambda_1 t) + x_2 \exp(\lambda_2 t) \tag{18}$$

where λ_1 and λ_2 are the roots of the characteristic equation associated with the differential Equations (17) and (18) (Equation 19):

$$\lambda_{1,2} = \frac{-R_+ \pm X}{2} \tag{19}$$

where $R_\pm = R_1^A \pm R_1^X$ and $X = \sqrt{R_-^2 + 4\sigma^2}$. The coefficients a_1, a_2, x_1 and x_2 depend not only on relaxation parameters but also on the initial conditions $I_z^A(0)$ and $I_z^X(0)$ (Equations 20 and 21):

$$a_{1,2} = \left\{ \frac{I_{eq}^A - I_z^A(0)}{2I_{eq}^A} \right\} \left(\frac{1}{2} \mp \frac{R_-}{2X} \quad \mp \quad \frac{I_{eq}^X - I_z^X(0)}{2I_{eq}^X} \right\} \times \left(\frac{I_{eq}^X}{I_{eq}^A} \quad \frac{\sigma}{X} \right. \tag{20}$$

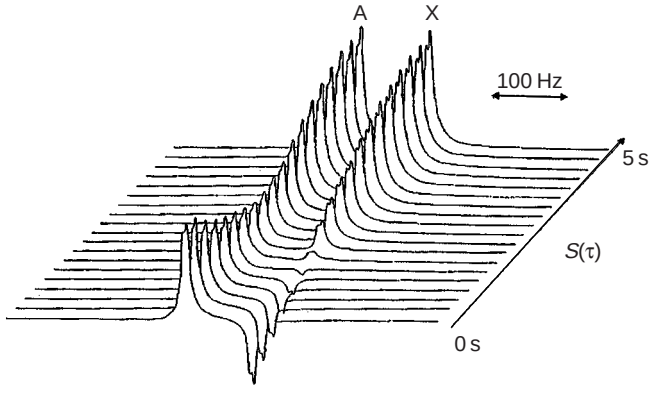

Figure 12 Evolution of two nuclear magnetizations obeying Equations (17) and (18) after the selective inversion of one of the magnetizations. A biexponential behavior is observed for both the magnetizations.

$$x_{1,2} = \left\{ \frac{I_{eq}^X - I_z^X(0)}{2I_{eq}^X} \right\} \left(\frac{1}{2} \pm \frac{R_-}{2X} \quad \mp \quad \frac{I_{eq}^A - I_z^A(0)}{2I_{eq}^A} \right\} \times \left(\frac{I_{eq}^A}{I_{eq}^X} \quad \frac{\sigma}{X} \right. \tag{21}$$

The above relations are quite general in the sense they can accommodate any kind of initial conditions: non-selective inversion, selective inversion of one of the magnetizations (Figure 12) or any intermediate situation.

However, the complexity of Equations (19–21) has prompted the design of more direct methods based on the initial behavior after an appropriate perturbation. Again, we start from Equations (15) and (16) and we assume that t is short enough to ensure that a first-order expansion is adequate (Equations 22 and 23):

$$I_z^A(t) = I_z^A(0) + t \left(\frac{dI_z^A}{dt} \right)_{t=0} \tag{22}$$

$$I_z^X(t) = I_z^X(0) + t \left(\frac{dI_z^X}{dt} \right)_{t=0} \tag{23}$$

It is convenient, from a conceptual and practical point of view, to express Equations (15) and (16) in the form of the initial slopes of the reduced quantities already introduced in Equations (17) and (18) (Equations 24 and 25):

$$S_A = \frac{d}{dt} \left(\frac{I_{eq}^A - I_z^A}{2I_{eq}^A} \right)_{t=0} = -R_1^A \left(\frac{I_{eq}^A - I_z^A(0)}{2I_{eq}^A} \right. - \sigma \left(\frac{I_{eq}^X - I_z^X}{2I_{eq}^A} \right. \tag{24}$$

$$S_X = \frac{d}{dt}\left(\frac{I_{eq}^X - I_z^X}{2I_{eq}^X}\right)_{t=0} = -R_1^X\left(\frac{I_{eq}^X - I_z^X(0)}{2I_{eq}^X}\right) - \sigma\left(\frac{I_{eq}^A - I_z^A}{2I_{eq}^A}\right) \quad (25)$$

These initial slopes are readily evaluated since we can always measure the signal corresponding to I_{eq}^A (or I_{eq}^X) and since the relevant instrumental factor is identical with that which prevails in the measurement of I_{eq}^A (or I_{eq}^X). Anyhow, initial conditions can be devised for determining separately one of the three relaxation parameters R_1^A, R_1^X or σ. The simplest experiment consists of selectively inverting one of the two magnetizations. Consider first the selective inversion of A magnetization for which the following initial conditions hold: $I_z^A(0) = -I_{eq}^A$ and $I_z^X(0) = I_{eq}^X$. This yields for the initial slopes Equations (26) and (27):

$$S_A \text{ (A selectively inverted)} = -R_1^A \quad (26)$$

$$S_X \text{ (A selectively inverted)} = -\sigma \quad (27)$$

The complementary experiment (selective inversion of X magnetization) leads to Equations (28) and (29):

$$S_A \text{ (X selectively inverted)} = -\sigma \quad (28)$$

$$S_X \text{ (X selectively inverted)} = -R_1^X \quad (29)$$

Another popular way of measuring the cross-relaxation rate σ relies upon the saturation (by continuous irradiation) of one of the two nuclei, say X. Referring again to Equation (15) with $I_z^X \equiv 0$ (saturation of spin X) and with $dI_z^A/dt = 0$, which implies that a stationary state has been reached prior to the measurement (this means that the irradiation of spin X has been applied for a time sufficiently long with respect to relaxation times), we obtain a new value for the longitudinal magnetization of spin A which will be denoted I_{stat}^A (Equation 30):

$$I_{stat}^A = I_{eq}^A\ 1 + \left(\frac{\gamma_X}{\gamma_A}\right)\left(\frac{\sigma}{R_1^A}\right) \quad (30)$$

Hence, provided that I_{eq}^A is known and that R_1^A has been determined by means of an independent experiment, I_{stat}^A provides the cross-relaxation rate σ. This stems from the NOE[18,20] (Overhauser was the first scientist to recognize that, by a related method, electron spin polarization could be transferred to nuclear spins). This effect is usually quantified by the so-called NOE factor η (Equation 31):

$$\eta = \frac{I_{stat}^A - I_{eq}^A}{I_{eq}^A} = \left(\frac{\gamma_X}{\gamma_A}\right)\left(\frac{\sigma}{R_1^A}\right) \quad (31)$$

In practice, I_{stat}^A and I_{eq}^A must be determined under identical instrumental conditions. The decoupler channel of the spectrometer is generally used for saturating the spin X. Homonuclear and heteronuclear systems have to

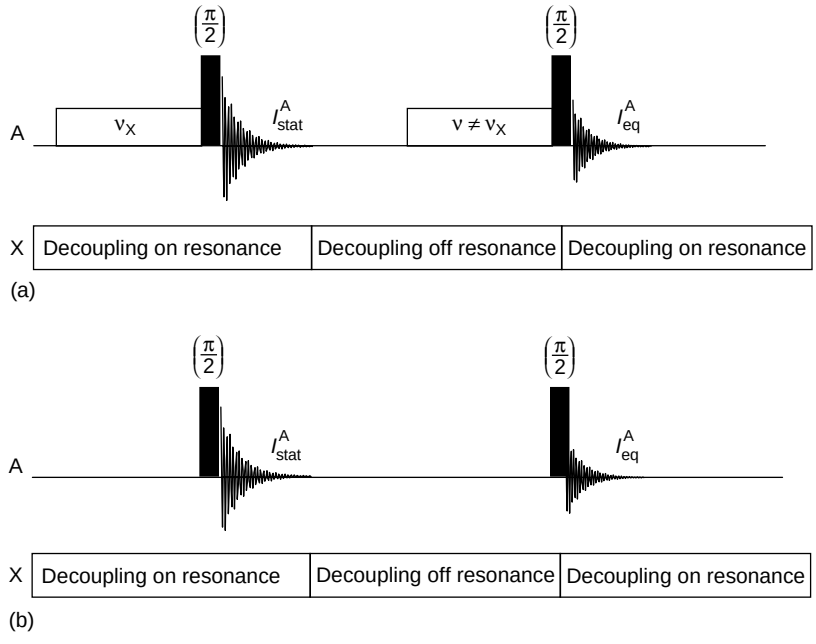

Figure 13 (a) Determination of the NOE factor in the homonuclear case. The frequency of the selective irradiation is moved at a value different from ν_X for the measurement of I_{eq}^A. (b) Determination of the NOE factor in the heteronuclear case. The broad-band decoupler is applied a long time prior to the measurement of I_{stat}^A. For the measurement of I_{eq}^A and for the sake of maintaining identical conditions, the decoupling frequency is moved far apart from its normal value and the modulation scheme is switched off (Dec. off).

be distinguished. In the former case, a weak (selective) secondary RF field centered on the X resonance is used for measuring I^A_{stat}, whereas I^A_{eq} is obtained by the same experiment with the irradiation frequency shifted to a zone devoid of resonances; the results of these two experiments are stored in different memory blocks (Figure 13a and b). In the heteronuclear case, broad-band X decoupling (which amounts to saturating X resonances) is used to measure I^A_{stat}. Since instrumental conditions as close as possible should be used for the measurement of I^A_{eq}, the decoupling frequency is moved to a value significantly different (as in the homonuclear case) and the modulation scheme[6] (which insured effective irradiation on a proper frequency band) is simply switched off so that no residual NOE can occur. The normal decoupling scheme is reset for signal acquisition (so as to obtain a decoupled spectrum); it should be mentioned that the NOE does not affect transverse magnetization. Again I^A_{stat} and I^A_{eq} are stored in different memory blocks for comparison purposes. The spectra in Figure 14 illustrate this procedure.

It is perfectly clear that, although the above experiments appear especially gratifying, they should be repeated for each pair of (A,X) nuclei (with the requirement of tedious experimental adjustments since selective pulses must be employed). Heteronuclear Overhauser effect measurements constitute an exception since they concern X nuclei in their globality. In fact,

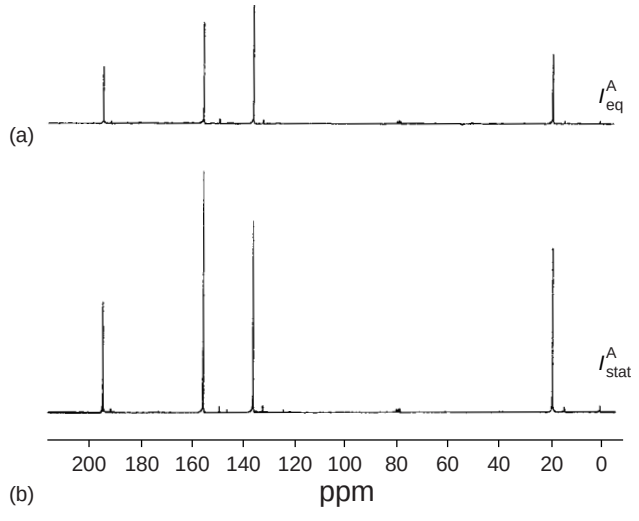

Figure 14 Determination of the heteronuclear Overhauser effect for the *trans*-crotonaldehyde molecule (experiment of Figure 13b). Carbon-13 is observed while proton is irradiated only during the signal acquisition (a) or continuously (b). The comparison of line intensities in the two spectra yields the NOE factor η for each carbon. Note: the ppm (parts per million) scale provides the position of a given resonance with respect to the one of a reference, independently of the magnetic field value B_0 value. It is given by $\delta_{ppm} = (\nu_0 - \nu_{ref})_{Hz}/(\nu_{measurement})_{MHz}$.

the experimental procedures can be simplified and automated through the two-dimensional techniques[21] briefly described below.

This two-dimensional method provides in a single experiment all cross-relaxation rates. The acronym NOESY (nuclear Overhauser effect spectroscopy)[22] originates from the one-dimensional experiments described above and is mostly reserved to the study of homonuclear cross-relaxation. In a general way, the analysis of a signal $S(t_1, t_2)$ (depending on two time variables) by means of a double Fourier transformation leads to a two-dimensional map in the frequency domains (ν_1, ν_2), whose cross-peaks (off-diagonal peaks for which $\nu_1 = \nu'_1$ and $\nu_2 = \nu'_2$) contain generally a correlation information regarding two nuclei resonating at frequencies ν'_1 and ν'_2. Concerning the NOESY sequence depicted in Figure 15, this correlation arises from cross-relaxation and indicates either a dipolar interaction (and thus a spatial proximity of the two considered nuclei) or chemical exchange between the two relevant sites. Let us emphasize again that the advantage of this two-dimensional technique stems from the global view of all dipolar or exchange correlations existing within the molecular system under investigation.

The preparation period allows for the return to equilibrium of nuclear magnetization. The evolution period provides a labeling according to the resonance frequency of each nucleus by repeating the experiment for incremented values of t_1. It will be assumed that transverse components during the mixing time t_m do not affect the final result (actually this can be achieved by several means, including phase cycling procedures). This amounts to considering, prior to the mixing period, exclusively longitudinal components which are off equilibrium by an amount depending on the chemical shift of each nucleus. It is therefore conceivable that the measured signal contains effects of specific and cross-relaxations, encoded in the t_1 dimension according to the chemical shift. This is illustrated by Figure 16 in which the diagonal peaks depend on specific longitudinal relaxation rates whereas cross-peaks (symmetrically located with

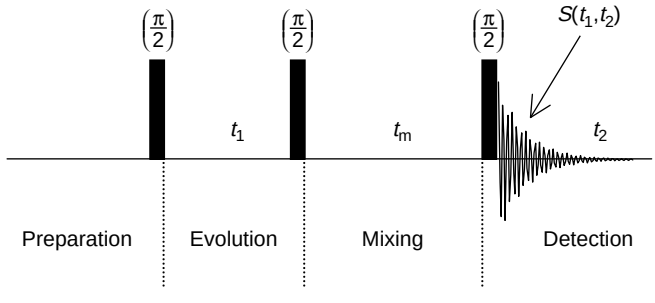

Figure 15 Basic scheme of the NOESY sequence which provides essentially homonuclear cross-relaxation rates from a double Fourier transformation of the signal $S(t_1, t_2)$.

respect to the diagonal) depend solely on the cross-relaxation rate between the two relevant nuclei. An example concerning biomolecular NMR is presented in Figure 17.

A similar experiment exists for measuring heteronuclear cross-relaxation rates, i.e. HOESY (heteronuclear Overhauser effect spectroscopy).[23,24] Of course, it employs RF pulses which must act on both nuclei. This experiment has mostly been used for studying intermolecular proximity.

We turn now to the determination of *transverse* cross-relaxation rates, by contrast with the NOESY experiment which is devoted to *longitudinal* cross-relaxation rates. It will be shown that the latter may be positive or negative, depending on the molecular mobility, and thus may go to zero; this situation occurs at certain

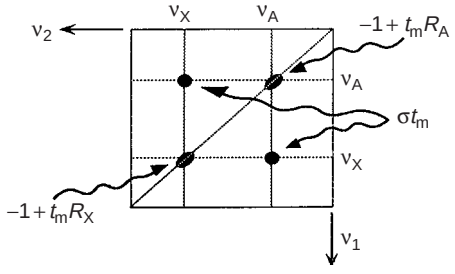

Figure 16 Meaning of the diagonal and cross-peaks in a NOESY two-dimensional diagram.

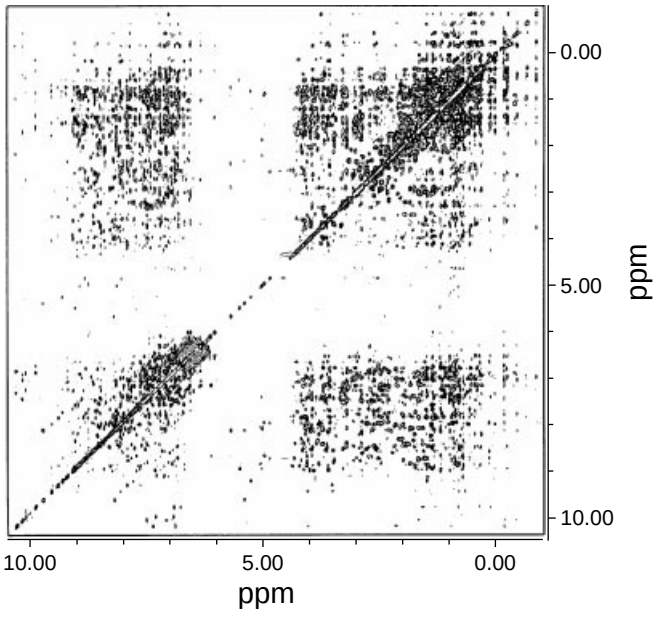

Figure 17 The NOESY spectrum of lysozyme at a concentration of 2 mM in H_2O whose signal has been suppressed by a special procedure. Note the wealth of information that can be extracted from such a spectrum.

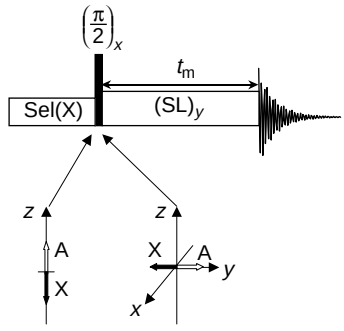

Figure 18 One-dimensional version of a ROESY experiment. Sel(X) stands for a selective pulse which takes the X magnetization toward $-z$. During the application of a spin-locking (SL) RF field along y and for a time t_m, transfers between the two magnetizations occur by transverse cross-relaxation.

measuring frequencies for slowly tumbling molecules such as proteins or nucleic acids. In contrast, transverse cross-relaxation rates are always positive. In that sense, a ROESY (rotating-frame Overhauser effect spectroscopy) experiment[25,26] which yields transverse cross-relaxation rates can be thought as complementary to a NOESY experiment. For theoretical reasons explained later, transverse cross-relaxation rates do not actually affect a conventional relaxation experiment using spin echoes. Rather, they become visible in a spin-lock experiment (see section 2.3). To maximize their effects, the two magnetizations of interest should be taken along the locking RF field in opposite directions. Hence, a possible design for a one-dimensional experiment is as follows: a selective pulse (i.e. where intensity has been adjusted to act only on the X nucleus) takes the X magnetization toward $-z$; it is followed by a nonselective pulse which brings both magnetizations along the spin-locking field which are then allowed to cross-talk during the mixing time t_m (Figure 18).

During the mixing time t_m, magnetizations evolve according to the relaxation rates in the rotating frame $R_{1\rho}^A$ and $R_{1\rho}^X (R = 1/T$, where R is a relaxation rate corresponding to a relaxation time), and also according to the transverse cross-relaxation rate denoted by σ_ρ (or $\sigma_\perp$). This is homologous to the relaxation parameters which govern the evolution of longitudinal magnetizations (R_1^A, R_1^X and σ). We can therefore expect a similar behavior, but with relaxation rates of different values under conditions of slow molecular mobility. The two-dimensional counterpart can be deduced from the sequence in Figure 18 by substituting in $[\mathrm{Sel(X)}(\pi/2)_x]$ a non-selective $(\pi/2)$ pulse followed by an evolution period as in the NOESY sequence in Figure 15. This yields qualitatively the same type of result as a NOESY experiment: a cross-peak reveals a nonzero transverse cross-relaxation rate.

2.5 Cross-correlation Rates

J splittings, when they exist, impose the definition of new spin quantities. These quantities also evolve according to relaxation phenomena and may interfere (by relaxation) with the usual magnetization components. This latter interference stems precisely from cross-correlation rates, i.e. relaxation parameters which involve two different mechanisms, e.g. dipolar interaction and the so-called chemical shift anisotropy (CSA) mechanism[27,28] (see section 3.3). In order to illustrate what is meant by ''new spin quantities'', let us consider the doublet (of splitting J) of one nucleus in a J-coupled two spin-1/2 system. Let us assume for simplicity that this spin is ''on resonance'' (that is, its resonance frequency ν_0 is identical with the reference frequency ν_r, which is also the frequency of the rotating frame), so that in the rotating frame, the resonance frequency is zero and the two lines of the doublet will just precess at frequency $+J/2$ and $-J/2$, respectively. As illustrated in Figure 19(a) and (b), the spin state at time t can be decomposed into an ''in-phase'' doublet (for the components along y) and an ''antiphase'' doublet (for the components along x).

Using the notations of the preceding section, the in-phase doublet is obviously represented by I_y^A whereas it can be shown that the antiphase doublet is represented by the product operator[29] $2I_x^A I_z^X$. It turns out that the two quantities I_y^A and $2I_x^A I_z^X$ can be coupled by a CSA–dipolar cross-correlation rate and, as a consequence, the two lines within the doublet possess different transverse relaxation rates.[28,30] This is a simple example of spin relaxation by cross-correlation. In many other situations, cross-correlation effects can be visible.[31–33] This will be discussed further in section 3, especially the so-called longitudinal-spin order represented by $2I_z^A I_z^X$. This latter spin state can be obtained by applying, to the previously discussed antiphase doublet ($2I_x^A I_z^X$), a $(\pi/2)_y$ selective pulse which acts exclusively on the A spin. It therefore transforms I_x^A into I_z^A and ($2I_x^A I_z^X$) into ($2I_z^A I_z^X$). On the other hand, because longitudinal magnetizations are coupled to the longitudinal order by CSA–dipolar cross-relaxation, the latter can be created in the course of a standard T_1 measurement and may alter the corresponding experimental results, as illustrated in Figure 20. This feature has also become especially

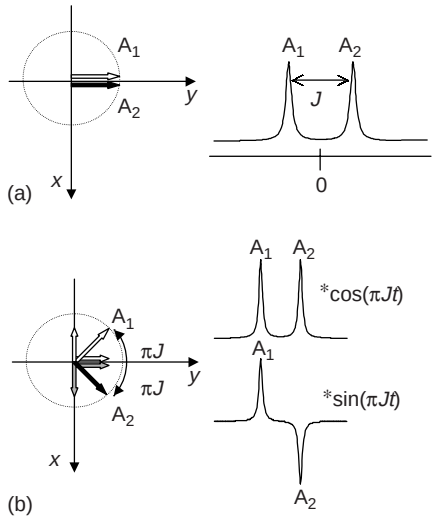

Figure 19 Spin states of a doublet: (a) after a $(\pi/2)x$ pulse and (b) some time t later, and its decomposition into an in-phase and an antiphase doublet.

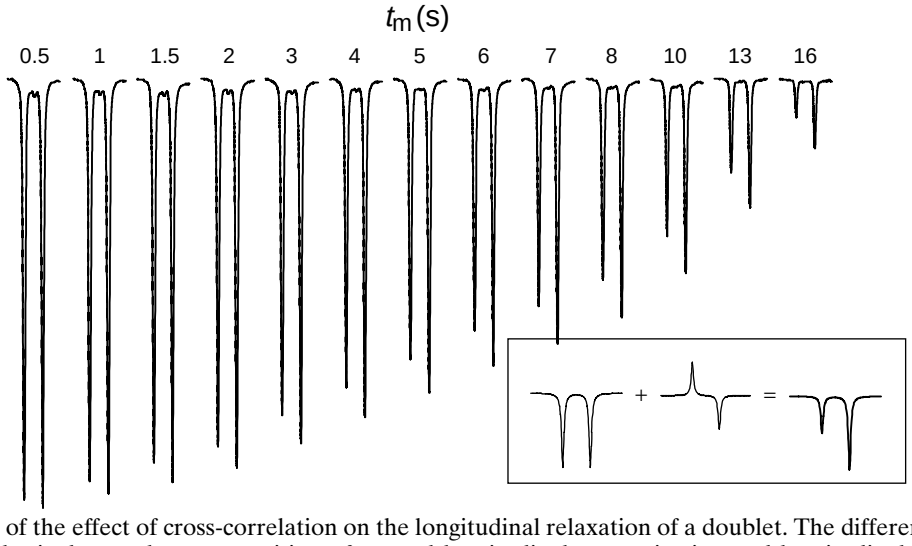

Figure 20 Example of the effect of cross-correlation on the longitudinal relaxation of a doublet. The differential evolution of each line within the doublet is due to the superposition of normal longitudinal magnetization and longitudinal spin order created by cross-correlation in a standard inversion–recovery experiment (the successive traces result from a read pulse, applied subsequently to the ''mixing'' period, which transforms the usual longitudinal magnetization into an in-phase doublet and the longitudinal spin order into an antiphase doublet).

apparent in measurements performed on proteins with spectrometers equipped with high magnetic field B_0.[34]

3 RELAXATION MECHANISMS

The only way to perturb a spin system from its equilibrium configuration or to take it back to its equilibrium state is to induce transitions. This, of course, can be done by application of an RF field or through magnetic fields governing relaxation phenomena. Consider a spin-1/2 system, that is, a system involving only two energy levels. As already stated, for inducing a transition between these two states, the experimenter can apply an RF field at the Larmor frequency ν_0 such that we have Equation (32) (see Figure 21):

$$h\nu_0 = \Delta E = \frac{\gamma h B_0}{2\pi} \tag{32}$$

where $\nu_0 = \gamma B_0/2\pi$ and the shielding coefficient has been omitted.

On the other hand, within a sample, an elementary nuclear magnetic moment μ is subjected to a local magnetic field $\boldsymbol{b}(t)$ originating from the various interactions to which this magnetic moment is subjected. Owing to molecular motions, these local fields are time dependent and consequently may be able to induce transitions. They must, however, mimic the action of an RF field and therefore fulfill the following conditions: (i) present some degree of coherence and (ii) be active at the frequency of the considered transition.

3.1 Interaction with Local Fluctuating Magnetic Fields

$\boldsymbol{b}(t)$ is, for instance, the magnetic field created by another spin (nuclear spin or the spin of an unpaired electron); in that case, it is proportional to $1/r^3$ (where r is the distance between the two spins). Its time dependence arises from the orientation of $\boldsymbol{r}$ and/or from the distance fluctuation. However, with the approach considered in this section, we disregard the origin of $\boldsymbol{b}(t)$ and we rest only on its general properties which arise from the random nature of molecular motions:

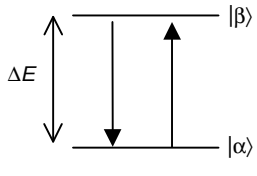

Figure 21 The two energy levels of a spin-1/2 system corresponding to the spin states $|\alpha\rangle$ and $|\beta\rangle$. The two possible transitions, represented by up and down arrows, must satisfy the Bohr relation $h\nu_0 = \Delta E$, where ν_0 is the resonance frequency.

1. the three components of a local magnetic field are not correlated (Equation 33):

$$\overline{b_x(t)b_y(t)} = \overline{b_x(t)b_z(t)} = \overline{b_y(t)b_z(t)} = 0 \tag{33}$$

   (the bar represents an ensemble average);
2. the mean value of each component is zero (Equation 34):

$$\overline{b_x(t)} = \overline{b_y(t)} = \overline{b_z(t)} = 0 \tag{34}$$

3. the mean values of their squares are identical (Equation 35):

$$\overline{b_x^2(t)} = \overline{b_y^2(t)} = \overline{b_z^2(t)} = \overline{b^2} \tag{35}$$

   (in an isotropic medium).

To understand the last two points, one can think of a quantity whose modulus is nonzero but which can take opposite values with the same probability.

The coherence of a random field can be evaluated by its correlation function, e.g. Equation (36):

$$\overline{b_x(t)b_x(0)} \neq 0 \tag{36}$$

which reflects its memory properties. Indeed, if the two quantities are totally uncorrelated (independent), one has Equation (37):

$$\overline{b_x(t)b_x(0)} = \overline{b_x(t)}\,\overline{b_x(0)} = 0 \tag{37}$$

Conversely, an RF field is totally correlated because it is represented by a sine (or cosine) function and, as a consequence, its value at any time t can be predicted from its value at time zero. The efficiency of a random field at a given frequency $\nu = w/(2\pi)$ can be appreciated by the Fourier transform of the above correlation function (Equation 38):

$$_{\mathrm{rf}}(w) = \int_0^\infty \overline{b_x(t)b_x(0)} \exp(-iwt)\,\mathrm{d}t \tag{38}$$

These quantities are called spectral densities and are involved in any relaxation parameter. Very often $\overline{b_x(t)b_x(0)}$ can be written as $\overline{b^2}\mathrm{e}^{-t/\tau_c}$, where τ_c is a *correlation time* associated with molecular reorientation (it can be thought as the time required for rotation by 1 rad; in liquids τ_c lies between 10^{-9} and 10^{-12} s). The above Fourier transform leads to a Lorentzian function of the type $\tau_c/(1 + w^2\tau_c^2)$. When the condition $w^2\tau_c^2 \ll 1$ is fulfilled (and this is the case for fast motions since the NMR measurement frequency is $<10^9$ Hz), the spectral density becomes independent of frequency. In such a case, one says that ''extreme narrowing'' conditions prevail.

We now intend to derive the Bloch equations in order to express T_1 and T_2 according to spectral densities at appropriate frequencies. The starting point

is the evolution equation of an elementary magnetic moment $\boldsymbol{\mu}$ subjected to a random field $\mathbf{b}$ (Equation 39):

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\mu} = \gamma\boldsymbol{\mu} \wedge \mathbf{b} \tag{39}$$

One has, e.g. for μ_z (Equation 40):

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_z = \gamma(\mu_{x'} b_{y'} - \mu_{y'} b_{x'}) \tag{40}$$

The prime indicates that one has switched to the so-called rotating frame in order to remove any precession effect at the angular velocity $w_0 = 2\pi\nu_0$, ν_0 being the resonance frequency (Equations 41 and 42):

$$b_{x'} = b_x \cos(w_0 t) - b_y \sin(w_0 t) \tag{41}$$

$$b_{y'} = b_x \sin(w_0 t) + b_y \cos(w_0 t) \tag{42}$$

One can always write (Equation 43):

$$\mu_z(t) = \mu_z(0) + \int_0^t \frac{\mathrm{d}\mu_z}{\mathrm{d}t'}\,\mathrm{d}t' \tag{43}$$

and proceed by time-dependent perturbation methods[35] (the relevant calculations will be avoided). It can be remembered that the quantity of interest is the z component of nuclear magnetization (Equation 44):

$$M_z(t) = \overline{\mu_z(t)} \tag{44}$$

After some algebra (and some physico-mathematical tricks), one arrives at Equation (45):

$$\frac{\mathrm{d}}{\mathrm{d}t}M_z = -(M_z - M_0)\gamma^2 \int_0^\infty \overline{b_x(t)b_x(0)} + \overline{b_y(t)b_y(0)} \times \mathrm{e}^{-iw_0 t}\,\mathrm{d}t \tag{45}$$

where M_0 is the equilibrium magnetization. This leads to the familiar Bloch equation pertaining to longitudinal magnetization (Equation 46):

$$\frac{\mathrm{d}}{\mathrm{d}t}M_z = -\left(\frac{1}{T_1}\right. \quad (M_z - M_0) \tag{46}$$

with (Equation 47) (rf standing for ''random fields''):

$$\left(\frac{1}{T_1}\right._{\mathrm{rf}} = (R_1)_{\mathrm{rf}} = 2\gamma^2 \;{}_{\mathrm{rf}}(w_0) \tag{47}$$

where R_1 is the inverse of the longitudinal relaxation time and is called the longitudinal relaxation rate. Similar calculations applied to transverse components lead to Equation (48):

$$\frac{\mathrm{d}}{\mathrm{d}t}M_{x,y} = -\left(\frac{1}{T_2}\right. \quad M_{x,y} \tag{48}$$

with (Equation 49):

$$\left(\frac{1}{T_2}\right._{\mathrm{rf}} = (R_2)_{\mathrm{rf}} = \gamma^2[\;{}_{\mathrm{rf}}(w_0) + \;{}_{\mathrm{rf}}(0)] \tag{49}$$

Notice the presence in R_2 of a spectral density *at zero frequency* arising from $\int_0^\infty \overline{b_z(t)b_z(0)}\,\mathrm{d}t$ (which evidently does not require a switch to the rotating frame).

Likewise (Equation 50):

$$\left(\frac{1}{T_{1\rho}}\right._{\mathrm{rf}} = \gamma^2[\;{}_{\mathrm{rf}}(w_0) + \;{}_{\mathrm{rf}}(w_1)] \tag{50}$$

with $w_1 = \gamma B_1 (B_1 = \text{spin-lock field})$.

It can be interesting, within this simple model, to vizualise the evolution of T_1 and T_2 as a function of ν_0 and of τ_c. For this goal we shall assume the simplest form for ${}_{\mathrm{rf}}(w_0)$, that is (Equation 51):

$${}_{\mathrm{rf}}(w_0) \propto \overline{b^2}\frac{\tau_c}{1 + w_0^2\tau_c^2} \tag{51}$$

This yields the curves for T_1 and T_2 given in Figures 22 and 23.

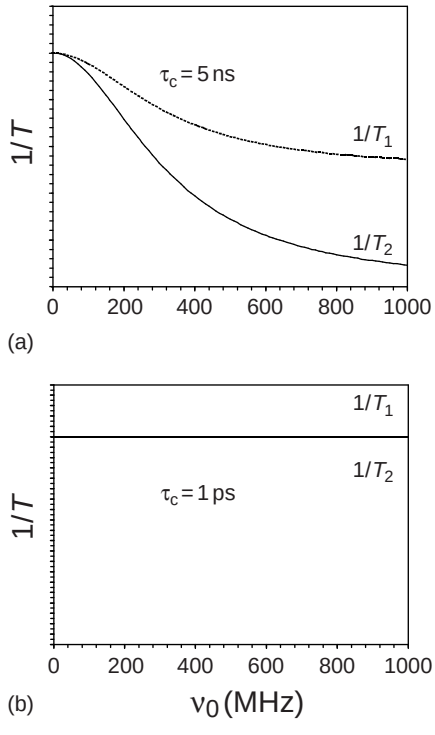

Figure 22 Evolution of the longitudinal and transverse relaxation times, T_1 and T_2, respectively, assuming that the considered spin is subjected to random fields whose correlation function is proportional to $\mathrm{e}^{(-t/\tau_c)}$, τ_c being the correlation time with the magnitude field B_0 (or, equivalently, with the measurement of resonance frequency, ν_0). For $\tau_c = 1\,\mathrm{ps}$, extreme narrowing conditions are seen to be fulfilled.

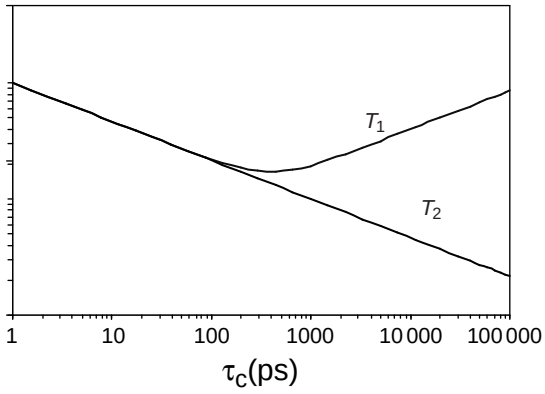

Figure 23 Same evolution as in Figure 22 but for a fixed frequency $\nu_0 = 400$ MHz, and as a function of the correlation time. Notice the continuous decease of T_2 as τ_c increases, while T_1 goes through a minimum.

3.2 Overview of the Various Interactions Responsible for Relaxation Phenomena

3.2.1 Dipolar Interaction

This mechanism concerns the direct interaction between the magnetic moments associated with nuclear or electron spins (see Figure 24). In the case of an unpaired electron, this mechanism is called pseudo-contact interaction.

The time modulation may have two origins: (i) reorientation of $\boldsymbol{r}$ due to molecular tumbling (intramolecular) or (ii) modulation of the distance $\boldsymbol{r}$ (intermolecular).

3.2.2 Chemical Shift Anisotropy

This originates from the tensorial nature of the shielding effect (responsible for chemical shift phenomena); its time dependence is due to the reorientation (with respect to the laboratory frame when the Z-axis coincides with B_0 of the tensor principal axes under molecular tumbling and is intramolecular by nature (Figure 25). These principal axes constitute a unique molecular frame depending on the electronic distribution and the considered interaction.

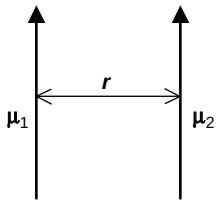

Figure 24 Representation used to figure out the dipolar interaction between two magnetic moments $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, separated by a vector $\boldsymbol{r}$.

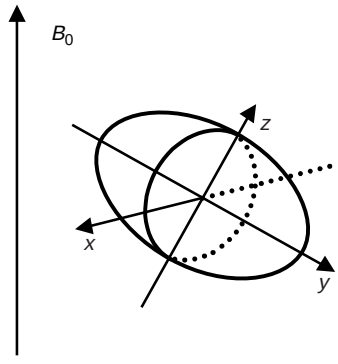

Figure 25 Representation used to figure out the influence of molecular tumbling (with respect to the fixed direction of B_0) via an interaction tensor whose xyz constitutes the principal axis system (PAS); (x, y, z) defines a molecular frame which undergoes molecular motions.

3.2.3 Quadrupolar Mechanism

The interaction between the quadrupole moment of nucleus of spin $>1/2$ with the electric field gradient (EFG) is tensorial and the time modulation arises again from the orientation of the PAS with respect to the B_0 direction, as for the CSA mechanism (Figure 25).

3.2.4 Spin Rotation

This is associated with the coupling between the spin angular momentum and the angular momentum describing the rotation of the electronic cloud. It is effective for small molecule or rotating methyl groups. Contrary to the other mechanisms, it produces a decrease in relaxation times when the temperature increases.

3.2.5 Scalar Relaxation of the First Kind

This concerns exchange phenomena and will not be considered further.

3.2.6 Scalar Relaxation of the Second Kind

This mechanism occurs when a slowly relaxing spin interacts with a fast relaxing spin (e.g. quadrupolar nucleus; unpaired electron, in that case it will be termed "contact interaction") to which it is J coupled (hyperfine coupling in the case of an unpaired electron).

3.3 Relaxation by Chemical Shift Anisotropy

Let B_0 be the static magnetic field of the NMR experiment. The actual field sensed by the nucleus has for components (in the laboratory frame X, Y, Z) (Equations 52–54):

$$b_X = \sigma_{XZ} B_0 \tag{52}$$

$$b_Y = \sigma_{YZ} B_0 \tag{53}$$

$$b_Z = (1 - \sigma_{ZZ}) B_0 \tag{54}$$

where σ_{XZ}, σ_{YZ} and σ_{ZZ} are elements of the shielding tensor expressed in the laboratory frame. It can be recalled that these quantities are very small (of the order of 10^{-6}) and reflect the actual field sensed by the nucleus under investigation. Of course, b_Z remains the major overwhelming component. As these elements fluctuate in time owing to molecular tumbling, the local field (random field) does so also. This mechanism is therefore active as far as spin relaxation is concerned. However, rather than dealing with spectral densities involving random fields, it would be more informative to deal with spectral densities involving elements of the shielding tensor. One therefore defines a new spectral density (Equation 55):

$$_{\mathrm{CSA}}(w) = \int_0^{\infty} \overline{\sigma_{ZZ}(t)\sigma_{ZZ}(0)} \exp(-iwt)\,\mathrm{d}t \qquad (55)$$

It can be shown that (in an isotropic medium) (Equation 56):

$$\begin{aligned} &\int_0^{\infty} \overline{\sigma_{XZ}(t)\sigma_{XZ}(0)} \exp(-iwt)\,\mathrm{d}t \\ &= \int_0^{\infty} \overline{\sigma_{YZ}(t)\sigma_{YZ}(0)} \exp(-iwt)\,\mathrm{d}t \\ &= \left(\tfrac{3}{4}\right)\ _{\mathrm{CSA}}(w) \end{aligned} \qquad (56)$$

Now, from Equation (45), we can remember that (Equation 57):

$$\left(\frac{1}{T_1}\right. = \gamma^2 \int_0^{\infty} \left(\overline{b_X(t)b_X(0)} + \overline{b_Y(t)b_Y(0)}\right) \mathrm{e}^{-iw_0t}\,\mathrm{d}t \qquad (57)$$

which leads to Equation (58):

$$(R_1)_{\mathrm{CSA}} = \left(\tfrac{3}{2}\right)\gamma^2 B_0^2\ _{\mathrm{CSA}}(w_0) \qquad (58)$$

Likewise (Equation 59):

$$(R_2)_{\mathrm{CSA}} = \gamma^2 B_0^2\ \left(\tfrac{3}{4}\right)\ _{\mathrm{CSA}}(w_0) + \ _{\mathrm{CSA}}(0) \qquad (59)$$

As immediate consequences, we can state that (i) R_1 and R_2 vary according to the square of the static magnetic field, and this contribution can therefore be sorted out by experiments performed at different magnetic fields, and (ii) under extreme narrowing conditions (Equation 60):

$$\frac{(R_2)_{\mathrm{CSA}}}{(R_1)_{\mathrm{CSA}}} = \frac{7}{6} \qquad (60)$$

In order to extract still more information from the CSA contribution to relaxation times, the next step is to switch to a molecular frame (x,y,z) where the shielding tensor is diagonal (x,y,z is called the PAS). The relevant calculations involve irreducible tensors, Wigner matrices and other spherical harmonics.[35] They yield for an isotropic reorientation (the molecule being supposed to behave as a sphere) (Equations 61 and 62):

$$(R_1)_{\mathrm{CSA}} = \frac{1}{15}(\gamma B_0)^2(\Delta\sigma)^2\left(1 + \frac{\eta_{\mathrm{CSA}}^2}{3}\right.\ \tilde{J}(w_0) \qquad (61)$$

$$\begin{aligned} (R_2)_{\mathrm{CSA}} = \frac{1}{15}(\gamma B_0)^2(\Delta\sigma)^2&\left(1 + \frac{\eta_{\mathrm{CSA}}^2}{3}\right. \\ &\times\ \frac{2\tilde{J}(0)}{3} + \frac{\tilde{J}(w_0)}{2} \end{aligned} \qquad (62)$$

where $\Delta\sigma = \sigma_{zz} - (\sigma_{xx} + \sigma_{yy})/2$ is the shielding anisotropy, $\eta_{\mathrm{CSA}} = (3/2)(\sigma_{xx} - \sigma_{yy})/\Delta\sigma$ is the asymmetry parameter and the reduced spectral density, $\tilde{J}(w)$, discussed in section 3.5, contains only dynamic information. In the case of an axially symmetry shielding tensor and for any kind of molecular tumbling, one can use the above equations with $\eta_{\mathrm{CSA}} = 0$ and evaluate the reduced spectral densities according to the actual molecular motions.

3.4 Scalar Relaxation of the Second Kind

The usual example is a spin-1/2 nucleus (A), J coupled to a fast relaxing quadrupolar nucleus (X) with relaxation times T_1^{X} and T_2^{X} (and spin number I_{X}). The relevant interaction is of the form in Equation (63) (J_{AX} in hertz):

$$J_{\mathrm{AX}}(I_x^{\mathrm{A}}I_x^{\mathrm{X}} + I_y^{\mathrm{A}}I_y^{\mathrm{X}} + I_z^{\mathrm{A}}I_z^{\mathrm{X}}) \qquad (63)$$

Because of fast X relaxation (which usually prevents the observation of J splittings; see section 4.4), $J_{\mathrm{AX}}I_x^{\mathrm{X}}$ can be considered as a random field γb_x acting on A. Consequently, correlation functions of the type $\overline{I_x^{\mathrm{X}}(t)I_x^{\mathrm{X}}(0)}$ have to be evaluated (in this case T_2^{X} plays the role of a correlation time). After some calculations,[35] one arrives at Equations (64) and (65):

$$(R_1^{\mathrm{A}})_{\mathrm{SC}} = \frac{8\pi^2 J_{\mathrm{AX}}^2}{3} I_{\mathrm{X}}(I_{\mathrm{X}}+1)\frac{T_2^{\mathrm{X}}}{1 + (w_{\mathrm{X}} - w_{\mathrm{A}})^2(T_2^{\mathrm{X}})^2} \qquad (64)$$

$$\begin{aligned} (R_2^{\mathrm{A}})_{\mathrm{SC}} = \frac{4\pi^2 J_{\mathrm{AX}}^2}{3} &I_{\mathrm{X}}(I_{\mathrm{X}}+1) \\ &\times\ \frac{T_2^{\mathrm{X}}}{1 + (w_{\mathrm{X}} - w_{\mathrm{A}})^2(T_2^{\mathrm{X}})^2} + T_1^{\mathrm{X}} \end{aligned} \qquad (65)$$

It can be seen that (i) owing to the importance of $(w_{\mathrm{X}} - w_{\mathrm{A}})^2(T_2^{\mathrm{X}})^2$, $(R_1^{\mathrm{A}})_{\mathrm{SC}}$ is generally negligible unless w_{X} is close to w_{A} (this may no longer be true if X is an unpaired electron), and (ii) for the same reason, $(R_2^{\mathrm{A}})_{\mathrm{SC}} \approx 4\pi^2 J_{\mathrm{AX}}^2 T_1^{\mathrm{X}} I_{\mathrm{X}}(I_{\mathrm{X}}+1)/3$ and is able to provide the value of the J_{AX} coupling constant.

This mechanism is also related to the "contact interaction" between an unpaired electron and a nuclear spin (in that case the hyperfine coupling replaces the J coupling).

3.5 Reduced Spectral Densities

In order to handle the correlation function involving the component of a cartesian tensor expressed in the laboratory frame, e.g. $\overline{\sigma_{ZZ}(t)\sigma_{ZZ}(0)}$, it is recommended to switch to a molecular frame. The objective of such an operation is to separate structural properties (molecular frame) from dynamic properties (contained in the function relating the laboratory frame to the molecular frame). This task is relatively easy when the considered interaction can be represented in the molecular frame by a unique direction (sometimes called the relaxation vector, e.g. the vector joining the two magnetic moments when dealing with dipolar interactions). This is the case (i) when molecular reorientation is isotropic (the considered molecule can be considered as a sphere) and all directions are equivalent, or (ii) if the tensor (expressed in the molecular frame) is axially symmetric. The relaxation vector corresponds in that case to the symmetry axis.

In such a situation, the spectral densities encountered so far and denoted are proportional to a reduced spectral density denoted by $\tilde{J}(w)$ (Equation 66):

$$\tilde{J}(w) = 4\pi \int_{-\infty}^{+\infty} \overline{Y_2^0[\Theta(t)]Y_2^0[\Theta(0)]} \exp(-iwt)\,\mathrm{d}t \quad (66)$$

where Θ is the angle between the relaxation vector and the preferential direction of the magnetic resonance experiment (i.e. the direction of B_0) and where the spherical harmonics Y_2^0 has the form (Equation 67):

$$Y_2^0(\Theta) = \sqrt{\frac{5}{16\pi}}(3\cos^2\Theta - 1) \quad (67)$$

In the case of an isotropic reorientation described by a simple correlation time τ_c one has (Equation 68):

$$\tilde{J}(w) = \frac{2\tau_c}{1 + w^2\tau_c^2} \quad (68)$$

Conversely, if the reorientation of a rigid molecule is anisotropic, three correlation times, associated with the PAS of the rotation diffusion tensor, are required. They lead to complicated (unwieldy) equations.[35]

In many situations (e.g. in large molecules), molecular reorientation can be viewed as the combination of an overall (slow) isotropic molecular reorientation (represented by a correlation time τ_s) and (fast) internal motions (represented by a correlation time τ_f) and it is recommended to turn to the widely used model-free approach,[36] equivalent to the two-step model[37] (Equation 69):

$$\tilde{J}(w) = \frac{2\tau_f}{1 + w^2\tau_f^2}(1 - S^2) + S^2\frac{2\tau_s}{1 + w^2\tau_s^2} \quad (69)$$

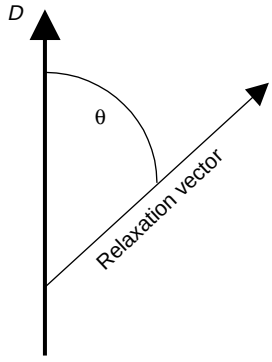

Figure 26 Definition of the angle θ involved in the calculation of the order parameter. D is a local director.

S is a generalized order parameter (similar to the quantity describing partial orientation in liquid crystals) which quantifies the motional restriction of the relaxation vector with respect to a local director D (e.g. the molecular long axis in the case of an alkyl chain). The order parameter is calculated as follows (Equation 70):

$$S = \tfrac{1}{2}\langle 3\cos^2\theta - 1\rangle \quad (70)$$

The brackets denote an average (see Figure 26 for the definition of θ). As a consequence, $-0.5 \leq S \leq 1$ and in the absence of any order $S = 0$ since $\langle\cos^2\theta\rangle = 1/3$. Measurements at different magnetic fields (different frequencies) yield in principle all dynamic parameters: τ_f, S and τ_s (see Figure 27).

It can be noted that (i) $\tilde{J}(0)$ (involved in R_2) is essentially equal to $2S^2\tau_s$ and (ii) at high frequencies, $\tilde{J}(w_0/(2\pi))$ is dominated by fast motions. Slow motions are probed by measurements performed at (very) low fields according to a technique known as relaxometry.[38]

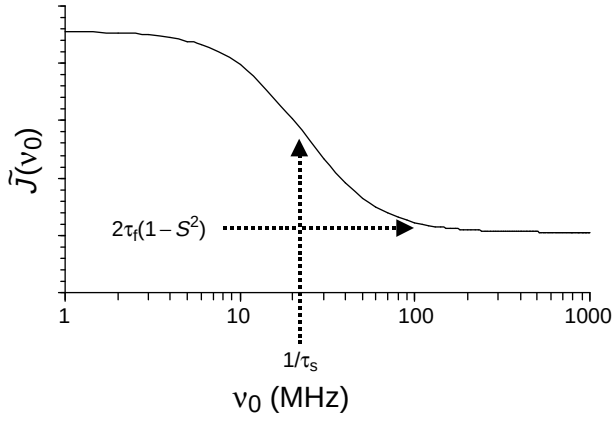

Figure 27 Evolution of the reduced spectral density as a function of the measurement frequency (or equivalently as a function of the magnetic field) when the model-free approach is invoked.

3.6 Relaxation in Coupled Spin Systems – Redfield Equations

If the spin system cannot be described by two states (e.g. $|\alpha\rangle$ and $|\beta\rangle$, as is the case for a single spin-1/2 system), macroscopic magnetization components are no longer sufficient (i.e. the classical description of spin dynamics). One must therefore turn to quantum mechanics. This is true for a system involving two spins-1/2 which possesses four eigenstates (for a weakly coupled spin system they are described by $|\alpha,\alpha\rangle$, $|\alpha,\beta\rangle$, $|\beta,\alpha\rangle$ and $|\beta,\beta\rangle$), and for a quadrupolar nucleus (three states for a spin-1, four states for a spin-3/2 and so on). The appropriate tool is the *density matrix*, which combines the description of the quantum system along with the statistical average over all the systems within the sample. The details of this approach, which require a considerable amount of mathematics, will be avoided. Only the essential features and properties of the density matrix will be outlined. Let $|a\rangle$ and $|a'\rangle$ be two spin states. The element $\sigma_{aa'}$ of the density matrix is defined according to Equation (71):

$$\sigma_{aa'} = \overline{c_a c_{a'}^*} \tag{71}$$

where c_a and $c_{a'}$ are the probability of finding a spin system in the states $|a\rangle$ and $|a'\rangle$, respectively. As usual, the bar denotes an ensemble average, i.e. an average over all systems constituting the sample. It follows that the diagonal element σ_{aa} corresponds to the population of the relevant energy level, P_a (Equation 72):

$$P_a = \sigma_{aa} \tag{72}$$

which, at equilibrium is proportional to $\exp(-E_a/kT)$. Diagonal elements are therefore related to longitudinal magnetization. The off-diagonal element $\sigma_{aa'}$ corresponds to a *coherence*. At equilibrium, there is no phase relationship between $|a\rangle$ and $|a'\rangle$, meaning that $\sigma_{aa'}$ is zero. A nonzero value of $\sigma_{aa'}$ indicates some degree of coherence between $|a\rangle$ and $|a'\rangle$, and thus the presence of some transverse magnetization associated with the transition $|a\rangle \rightarrow |a'\rangle$ (there is, of course, full coherence if $|a\rangle = |a'\rangle$). The density matrix evolves according to Redfield equations[39] (Equation 73):

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_{aa'} = i(w_{a'} - w_a)\sigma_{aa'} + \sum_{b,b'} R_{aa'bb'}\sigma_{bb'} \tag{73}$$

The first term, depending on $(w_{a'} - w_a)$, is related to precession ($w_{a'}$ and w_a represent the energy values of eigenstates $|a\rangle$ and $|a'\rangle$ expressed in radians per second) and the second term is related to relaxation ($\mathbf{R}$ = relaxation matrix). It is nonzero only if (Equation 74):

$$|w_{a'} - w_a| \approx |w_{b'} - w_b| \tag{74}$$

Some immediate consequences of the Redfield equations can be formulated as follows:

Longitudinal relaxation (Equation 75):

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_{aa} = \sum_b R_{aabb}\sigma_{bb} \tag{75}$$

which can be recast as Equation (76):

$$\frac{\mathrm{d}}{\mathrm{d}t}P_a = \sum_b W_{ab}P_b \tag{76}$$

with $W_{ab} = R_{aabb}$, which is the probability of a transition induced by relaxation. A diagonal element is coupled to other diagonal elements. Therefore, longitudinal magnetization and transverse magnetization are not connected by relaxation phenomena.

Transverse relaxation: consider an isolated line in the NMR spectrum, corresponding to the transition $|a\rangle \rightarrow |a'\rangle$. The relevant element of the density matrix evolves according to Equation (77):

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_{aa'} = [i(w_{a'} - w_a) + R_{aa'aa'}]\sigma_{aa'} \tag{77}$$

which would correspond (if B_0 inhomogeneity is negligible) to the line width at half-height equal to $R_{aa'aa'}/\pi$.

3.7 Longitudinal Relaxation in a Weakly Coupled Two-spin-1/2 System

From the diagram displayed in Figure 28, it is apparent that (Equations 78 and 79):

$$\langle I_z^{\mathrm{A}}\rangle \propto (P_1 - P_3) + (P_2 - P_4) \tag{78}$$

$$\langle I_z^{\mathrm{X}}\rangle \propto (P_1 - P_2) + (P_3 - P_4) \tag{79}$$

$\mathrm{d}\langle I_z^{\mathrm{A}}\rangle/\mathrm{d}t$ and $\mathrm{d}\langle I_z^{\mathrm{X}}\rangle/\mathrm{d}t$ will therefore involve all transition probabilities including zero-, one- and double-quantum transitions, at the respective frequencies $(w_{\mathrm{A}} - w_{\mathrm{X}})$, w_{A}, w_{X} and $(w_{\mathrm{A}} + w_{\mathrm{X}})$. It can be recalled that one-quantum

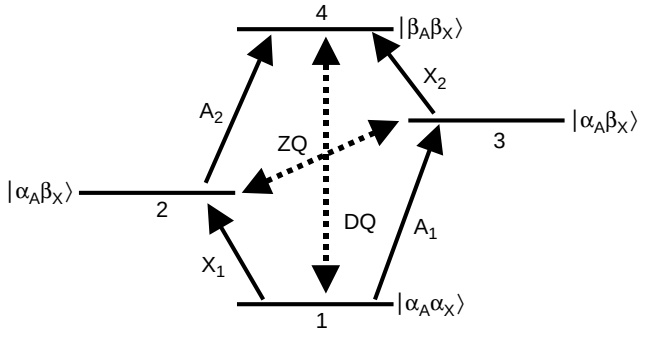

Figure 28 Energy level of a two-spin system. Full arrows, normally observed one-quantum (1Q) transitions; dashed arrows, zero-quantum (ZQ) and double-quantum (DQ) transitions.

transitions are those normally observed, which involve, for a given spin, the change of α into β (or β into α). Double-quantum transitions involve the same change for both spins, and zero-quantum transitions are such that for one spin α is changed into β and for the other spin β is changed into α.

Complete calculations lead to the Solomon equations given in section 2.4 (Equations 80 and 81):

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle I_z^{\mathrm{A}}\rangle = -R_1^{\mathrm{A}}(\langle I_z^{\mathrm{A}}\rangle - I_{\mathrm{eq}}^{\mathrm{A}}) - \sigma(\langle I_z^{\mathrm{X}}\rangle - I_{\mathrm{eq}}^{\mathrm{X}}) \qquad (80)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle I_z^{\mathrm{X}}\rangle = -R_1^{\mathrm{X}}(\langle I_z^{\mathrm{X}}\rangle - I_{\mathrm{eq}}^{\mathrm{X}}) - \sigma(\langle I_z^{\mathrm{A}}\rangle - I_{\mathrm{eq}}^{\mathrm{A}}) \qquad (81)$$

where R_1^{A} and R_1^{X} are specific relaxation rates depending in an additive way on the AX dipolar interaction, $(R_1^{\mathrm{A}})_{\mathrm{d}}$ or $(R_1^{\mathrm{X}})_{\mathrm{d}}$, and on other possible mechanisms. As already discussed in section 2.4, σ is a coupling term between A and X longitudinal magnetizations, also called cross-relaxation rate, depending solely on the AX dipolar interaction. From the above theory, relaxation parameters can be expressed as Equation (82):

$$(R_1^{\mathrm{A}})_{\mathrm{d}} = K_{\mathrm{AX}}[6\tilde{J}(w_{\mathrm{A}}+w_{\mathrm{X}}) + 3\tilde{J}(w_{\mathrm{A}}) + \tilde{J}(w_{\mathrm{A}}-w_{\mathrm{X}})] + (R_1^{\mathrm{A}})_{\mathrm{others}} \qquad (82)$$

(↑ DQ; ↑ 1Q; ↑ ZQ)

and (Equation 83):

$$\sigma = K_{\mathrm{AX}}[6\tilde{J}(w_{\mathrm{A}}+w_{\mathrm{X}}) - \tilde{J}(w_{\mathrm{A}}-w_{\mathrm{X}})] \qquad (83)$$

(↑ DQ; ↑ ZQ)

where DQ, 1Q, ZQ stand for double-, one- and zero-quantum coherences, respectively, and (Equation 84):

$$K_{\mathrm{AX}} = \left(\frac{1}{20}\right)\left(\frac{\mu_0}{4\pi}\right)^2\left(\frac{\gamma_{\mathrm{A}}\gamma_{\mathrm{X}}\hbar}{r_{\mathrm{AX}}^3}\right)^2 \qquad (84)$$

Conversely, the transverse relaxation rate originating from dipolar contribution is expressed as Equation (85), given here for the sake of comparison:

$$(R_2^{\mathrm{A}})_{\mathrm{d}} = K_{\mathrm{AX}}\left[3\tilde{J}(w_{\mathrm{A}}+w_{\mathrm{X}}) + \frac{3\tilde{J}(w_{\mathrm{A}})}{2} + 3\tilde{J}(w_{\mathrm{X}}) + \frac{\tilde{J}(w_{\mathrm{A}}-w_{\mathrm{X}})}{2} + 2\tilde{J}(0)\right] \qquad (85)$$

From the Solomon equations, we can immediately derive two important properties:

1. In a general way, longitudinal magnetization exhibits a biexponential behavior:
   (a) unless X is continuously irradiated, in which case (Equation 86):
   $$\frac{\mathrm{d}}{\mathrm{d}t}\langle I_z^{\mathrm{A}}\rangle = -R_1^{\mathrm{A}}(\langle I_z^{\mathrm{A}}\rangle - I_{\mathrm{stat}}^{\mathrm{A}}) \qquad (86)$$
   with (Equation 87):
   $$I_{\mathrm{stat}}^{\mathrm{A}} = I_{\mathrm{eq}}^{\mathrm{A}}\left(1 + \frac{\sigma}{R_1^{\mathrm{A}}}\frac{\gamma_{\mathrm{X}}}{\gamma_{\mathrm{A}}}\right. \qquad (87)$$
   (b) or unless A and X constitute a homonuclear system ($w_{\mathrm{a}} = w_{\mathrm{x}} = w_0$), subjected to nonselective pulses (like spins), in which case R_1^{A} and σ have to be added, leading to a monoexponential evolution with (Equation 88):
   $$(R_1^{\mathrm{like}}) = K[12\tilde{J}(2w_0) + 3\tilde{J}(w_0)] + (R_1)_{\mathrm{others}} \qquad (88)$$
2. If extreme narrowing conditions prevail (Equations 89–91):
   $$(R_1^{\mathrm{A}})_{\mathrm{d}} = K_{\mathrm{AX}}(20\tau_{\mathrm{c}}) \qquad (89)$$
   $$\sigma = K_{\mathrm{AX}}(10\tau_{\mathrm{c}}) = \frac{(R_1^{\mathrm{A}})_{\mathrm{d}}}{2} \qquad (90)$$
   $$(R_1^{\mathrm{like}})_{\mathrm{d}} = K(30\tau_{\mathrm{c}}) = \frac{3(R_1^{\mathrm{A}})_{\mathrm{d}}}{2}\left(\frac{3}{2}\ \text{effect}\right. \qquad (91)$$
   Under these conditions, it can be seen that $(R_1^{\mathrm{A}})_{\mathrm{d}} = (R_2^{\mathrm{A}})_{\mathrm{d}}$.

3.8 The Nuclear Overhauser Effect

In a two-spin system involving dipolar interaction, the quantity of interest is obviously the cross-relaxation rate because of its unique dependence on the internuclear distance and on the reorientation of the internuclear vector. Referring to Equation (83), we can always write (Equation 92):

$$\sigma \propto \frac{1}{r^6} \times (\tau_{\mathrm{c}}^{\mathrm{eff}}) \qquad (92)$$

(↑ geometry; ↑ dynamic properties)

If $\tau_{\mathrm{c}}^{\mathrm{eff}}$ can be estimated from another measurement, it yields the internuclear distance r whereas it provides dynamic information if r is known. The most popular way of measuring σ is the so-called NOE factor (see section 2.4) (Equation 93):

$$\eta = \frac{I_{\mathrm{stat}}^{\mathrm{A}} - I_{\mathrm{eq}}^{\mathrm{A}}}{I_{\mathrm{eq}}^{\mathrm{A}}} = \frac{\gamma_{\mathrm{X}}}{\gamma_{\mathrm{A}}}\frac{\sigma}{R_1^{\mathrm{A}}} \qquad (93)$$

R_1^A being determined from an independent measurement. If R_1^A is dominated by the dipolar mechanism and under extreme narrowing conditions (Equation 94):

$$\eta_{max} = \frac{\gamma_X}{\gamma_A}\frac{1}{2} = \begin{array}{l} 2 \text{ for the } {}^{13}C-{}^{1}H \text{ system} \\ \quad ({}^{1}H \text{ irradiated}) \\ 0.5 \text{ for a homonuclear case} \end{array} \quad (94)$$

Because of the difference in resonance frequencies, transverse magnetizations are not affected by cross-relaxation (see Equation 77). In the presence of a spin-lock field, this difference in resonance frequencies is removed and cross-relaxation reappears. This is the basis of the homonuclear ROESY experiment, which depends on cross-relaxation rates of the form in Equation (95):[40]

$$\sigma_\rho (= \sigma_\perp) = K_{AX}[3\tilde{J}(w_0) + 2\tilde{J}(w_1)] \quad (95)$$

with $w_1 = \gamma B_1$ much smaller than $w_0 = \gamma B_0$, while it can be recalled that the conventional cross-relaxation rate for a homonuclear system can be expressed as Equation (96):

$$\sigma = (\sigma_\parallel) = K_{AX}[6\tilde{J}(2w_0) - \tilde{J}(0)] \quad (96)$$

This involves a term affected by the minus sign, hence $\sigma_\parallel$ can go to zero. As a consequence, the cross-relaxation rate $\sigma_\parallel$ is prone to yield strong (negative) responses whenever $\tilde{J}(0)$ is overwhelming in the case of slow motions. In unfavorable circumstances (correlation time vs measurement frequency), $6\tilde{J}(2w_0)$ may become comparable to $\tilde{J}(0)$ and the responses will be negligibly small. In that case $\sigma_\perp$ should be preferred (Figure 29).

3.9 The Chemical Shift Anisotropy–Dipolar Interference Term

So far, for describing longitudinal relaxation in a two spin-1/2 system, two quantities have been invoked, namely $\langle I_z^A \rangle$ and $\langle I_z^X \rangle$. As this system involves four energy levels, four quantities would normally be required. A third (trivial) quantity is the identity. The fourth quantity is the so-called longitudinal order represented by the ''operator product'' $2I_z^A I_z^X$ already introduced in section 2.5. Longitudinal order is zero at thermal equilibrium and can be created through cross-correlation between the dipolar interaction and the CSA mechanism (interference term) in such a way that the Solomon equations must be extended[33,41] (Equations 97–99):

$$\frac{d}{dt}\langle I_z^A \rangle = -R_1^A(\langle I_z^A \rangle - I_{eq}^A) - \sigma_{AX}(\langle I_z^X \rangle - I_{eq}^X) - \sigma_{CSA(A),d}\langle 2I_z^A I_z^X \rangle \quad (97)$$

$$\frac{d}{dt}\langle I_z^X \rangle = -R_1^X(\langle I_z^X \rangle - I_{eq}^X) - \sigma_{AX}(\langle I_z^A \rangle - I_{eq}^A) - \sigma_{CSA(X),d}\langle 2I_z^A I_z^X \rangle \quad (98)$$

$$\frac{d}{dt}\langle 2I_z^A I_z^X \rangle = -R_1^{A,X}\langle 2I_z^A I_z^X \rangle - \sigma_{CSA(A),d}(\langle I_z^A \rangle - I_{eq}^A) - \sigma_{CSA(X),d}(\langle I_z^X \rangle - I_{eq}^X) \quad (99)$$

where $R_1^{A,X}$ is the specific relaxation rate of longitudinal spin order and $\sigma_{CSA(A),d}$ is the interference term, expressed according to Equation (100):

$$\sigma_{CSA(A),d} = -\left(\frac{1}{5}\quad \sqrt{\frac{2}{3}}(\gamma_A B_0)(\Delta\sigma_A)\quad \frac{\mu_0}{4\pi}\right. \times \left(\frac{\gamma_A \gamma_X \hbar}{r_{AX}^3}\quad \tilde{J}(w)\right. \quad (100)$$

It enhances the CSA ($\Delta\sigma_A$ stands here for the shielding anisotropy of nucleus A; see section 3.3) mechanism and has been detected in numerous instances at high field. For instance, in proteins, it leads to the determination of carbon-13, nitrogen-15 and even proton chemical shift anisotropies.[42] Analogous equations exist for transverse relaxation; they involve an interference term which, in the extreme narrowing, is equal to 7/6 $\sigma_{CSA(A),d}$.[43]

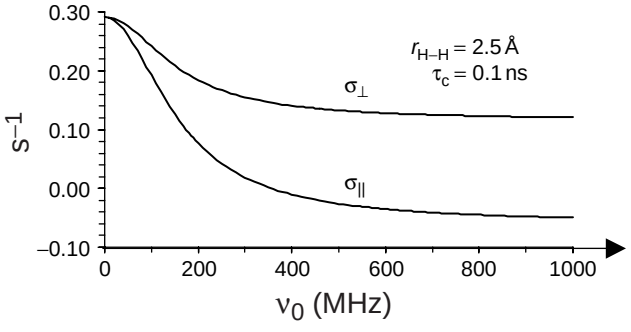

Figure 29 Comparison of the evolution of parallel and perpendicular homonuclear cross-relaxation rates as a function of the measurement frequency ν_0. Calculations have been performed for a pair of protons distant by 2.5 Å and a simple spectral density with $2\tau_c/(1 + w^2\tau_c^2)$ with $\tau_c = 0.1$ ns.

3.10 Quadrupolar Relaxation

Because the spin system (i.e. the quadrupolar nucleus) possesses more than two energy levels, a quantum mechanical treatment is required and, in general, the evolution of longitudinal and transverse magnetizations is not exponential. Additional quantities (in the same way as the longitudinal spin order for dipolar relaxation) may be necessary so as to describe completely the evolution provided that the NMR spectrum involves quadrupolar splittings (anisotropic medium).

If the environment of the considered nucleus is not of spherical symmetry, the quadrupolar mechanism is overwhelming. Some useful equations are given below.

1. In extreme narrowing conditions, τ_c^{eff} describing the reorientation of the EFG tensor z-axis, longitudinal and transverse relaxations are purely exponential with (Equation 101):

$$(R_1)_q = (R_2)_q = \left(\frac{3\pi^2}{10} \quad \left\{\frac{2I+3}{I^2(2I-1)}\right\} \times \chi^2(1+\eta_q^2)\tau_c^{\text{eff}} \right. \tag{101}$$

where I = spin number, $\chi = eQV_{zz}/h$ = quadrupole coupling constant with e = electron charge, Q = quadrupolar momentum, V_{zz} = either the largest element of the EFG tensor (which is supposed to be expressed in its PAS) or the element corresponding to the symmetry axis if the EFG tensor is axially symmetrical and $\eta_q = (V_{xx} - V_{yy})/V_{zz}$ (with $|V_{zz}| \geq |V_{yy}| \geq |V_{xx}|$) is the asymmetry parameter.

2. In the general case, the following approximations may be considered[44] (Equations 102 and 103):

$$(R_1)_q \approx \frac{2I+3}{5} \left[8 \ {}^q(2w_0) + 2 \ {}^q(w_0)\right] \tag{102}$$

$$(R_2)_q \approx \frac{2I+3}{5} \left[2 \ {}^q(2w_0) + 5 \ {}^q(w_0) + 3 \ {}^q(0)\right] \tag{103}$$

where the measurement frequency $\nu_0 = w_0/(2\pi)$, and Equation (104):

$${}^q(w) = \left(\frac{3\pi^2}{40} \quad \frac{1}{I^2(2I-1)} \quad \chi^2 \left(\frac{1+\eta_q^2}{3} \quad \tilde{J}(w)\right.\right. \tag{104}$$

These expressions are strictly valid for a spin-1 (^{2}H, ^{14}N).

3.11 Paramagnetic Relaxation[45]

A spin-1/2 nucleus close to an unpaired electron is subjected to (i) the dipolar interaction (dipolar or pseudo-contact mechanism) or (ii) the hyperfine interaction (scalar relaxation of the second kind or contact mechanism). The fast relaxing nature of the electron spin must be taken into account in the effective correlation times (and also possible exchange phenomena) (Equations 105 and 106):

$$\frac{1}{\tau_{c_1}} = \frac{1}{T_{1e}} + \frac{1}{\tau_R} + \frac{1}{\tau_e} + \frac{1}{\tau_M} \tag{105}$$

$$\frac{1}{\tau_{c_2}} = \frac{1}{T_{2e}} + \frac{1}{\tau_R} + \frac{1}{\tau_e} + \frac{1}{\tau_M} \tag{106}$$

where T_{1e} and T_{2e} are electronic relaxation times, τ_R is the rotational correlation time (isotropic motion assumed) and τ_e and τ_M are electron and nucleus residence times in the case of exchange.

Denoting the electron resonance frequency by w_e and realizing that $w_e \gg w_0$, one obtains Equations (107) and (108) for the dipolar contribution:

$$(R_1)_d = K_d \left(\frac{6\tau_{c_1}}{1+w_0^2\tau_{c_1}^2} + \frac{14\tau_{c_2}}{1+w_e^2\tau_{c_2}^2}\right. \tag{107}$$

$$(R_2)_d = K_d \left(4\tau_{c_1} + \frac{3\tau_{c_1}}{1+w_0^2\tau_{c_1}^2} + \frac{13\tau_{c_2}}{1+w_e^2\tau_{c_2}^2}\right. \tag{108}$$

For K_d, see K_{AX} defined for the dipolar interaction, where A here represents the nucleus and X the electron.

For the scalar relaxation contribution of both kinds (first and second) with (Equations 109–111):

$$\frac{1}{\tau_{e_1}} = \frac{1}{\tau_{c_1}} - \frac{1}{\tau_R} \tag{109}$$

$$\frac{1}{\tau_{e_2}} = \frac{1}{\tau_{c_2}} - \frac{1}{\tau_R} \tag{110}$$

$$K_{SC} = \frac{4\pi^2 A^2 S(S+1)}{3} \tag{111}$$

where A = hyperfine coupling in hertz and S = spin number of the electronic system, one obtains Equations (112) and (113):

$$(R_1)_{SC} = K_{SC}\frac{2\tau_{e_2}}{1+w_e^2\tau_{e_2}^2} \tag{112}$$

$$(R_2)_{SC} = K_{SC}\left(\tau_{e_1} + \frac{\tau_{e_2}}{1+w_e^2\tau_{e_2}^2}\right. \tag{113}$$

4 INFLUENCE OF SPIN RELAXATION ON QUANTITATIVE MEASUREMENTS

It is clear from the above considerations that spin relaxation plays a central role in any NMR experiment. A few illustrative examples, dealing with quantitative measurements, are provided in this section. Many more examples can be found in two review articles.[46,47]

4.1 Sensitivity Improvement and Repetition Rate

As already mentioned in the Introduction, NMR is very often impaired by poor sensitivity and this requires some remedy. In particular, it is well known that, after co-adding n scans, the S/N improves by $\sqrt{n}$. The only issue concerns the time needed for performing n scans. Roughly, a CW NMR spectrum is recorded in several minutes whereas an FID drops to an undetectable level in less than a few seconds (or milliseconds). This is why pulse NMR (or, in other words, FTNMR) has gained much popularity in the context of signal accumulation. The

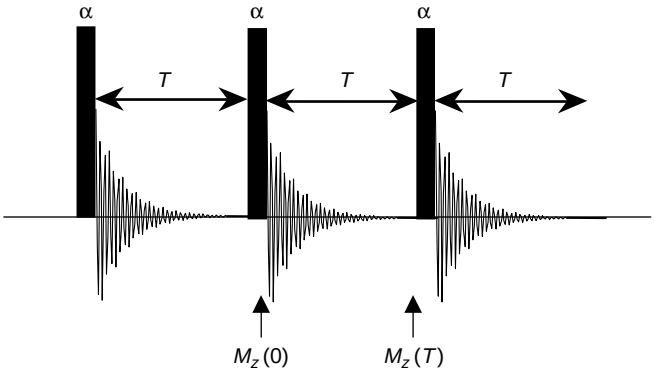

Figure 30 The standard NMR experiment, involving the coaddition of FIDs. They result from RF pulses (of flip angle α) with a repetition interval of duration T.

standard pulse NMR experiment is shown schematically in Figure 30.

From Equation (31) and assuming a stationary state [$M_z(0) = \cos\alpha M_z(T)$, where α is the flip angle of the RF pulse – the angle by which the nuclear magnetization is moved from the z-axis], it is easily shown that the measured quantity can be expressed as Equation (114):

$$M_0 \sin\alpha \frac{1 - \exp(-T/T_1)}{1 - \cos\alpha \exp(-T/T_1)} \tag{114}$$

Of course, ideally one should measure $M_0 \sin\alpha$, which would require T of the order of $5T_1$ so that $\exp(-T/T_1) = 0.01$. A compromise in terms of S/N is usually adopted through an optimum flip angle, α_{opt}, also called the Ernst angle (Equation 115):[2]

$$\cos\alpha_{opt} = \exp\left(\frac{-T}{T_1}\right. \tag{115}$$

for any arbitrary T value, but most of the time much shorter than $5T_1$. This entails again a bias as far as quantitative results are wanted (see Equation 114). Whenever T_1 is long, it must be reduced so that the condition $T \approx T_1$ can be fulfilled. Since paramagnetic relaxation (section 3.11) is very efficient, the usual trick consists in adding a very small amount of a paramagnetic substance, e.g. $Cu(SO_4)$ or $Cr(acac)_3$ [tris(acetato)chromium] according to their ability to be dissolved in the solution under investigation. Sometimes, dissolved oxygen present in the solution may prove sufficient.

4.2 Polarization Transfer in Heteronuclear Systems

NMR spectroscopy of heteronuclei is generally performed under proton broad-band decoupling with the aim of simultaneously simplifying the spectrum and increasing the sensitivity since all lines within a multiplet (due to J coupling with protons) coalesce into a single line. This is especially interesting for carbon-13 NMR at natural abundance level; in that case, the existence of two ^{13}C isotopes within the same molecule can be disregarded and splittings due to the carbon–carbon coupling are not expected. Each carbon in the molecule thus appears as a single line, provided of course that no nuclear spins other than protons (e.g. ^{31}P, ^{19}F) are present. However, as already mentioned, broad-band decoupling implies the saturation of proton transitions (so as to mix proton spin states and thus suppress J coupling effects), which leads to the NOE. As seen in section 3.8, the signal is enhanced by a factor equal to $1 + (\gamma_H/\gamma_C)(\sigma/R_1^C)$, where γ_C and γ_H are the carbon and proton gyromagnetic ratios, respectively, R_1^C is the carbon longitudinal relaxation rate and σ the cross-relaxation rate equal at most to $R_1^C/2$ (when extreme narrowing conditions prevail and when carbon relaxation is essentially of dipolar origin). Obviously, each carbon-13 in a given molecule has its own relaxation properties so that NOE enhancements are differentiated. For instance, a carbon bound to a proton is likely to benefit from a full NOE (in a small molecule) whereas the relaxation of a carbonyl is probably dominated by the CSA mechanism. Therefore, the carbonyl peak intensity is generally much less than that of an aliphatic carbon. Clearly, the quantitative aspect of carbon-13 spectroscopy is lost owing to the NOE. This undesirable feature combines with the loss of intensity arising from a repetition rate that is too high with respect to the longitudinal relaxation rate. Concerning NOE, the only remedy is to prevent its buildup by switching on the proton decoupler only during signal acquisition according to the scheme in Figure 31(a) and (b), where an example of a quantitative ^{13}C spectrum is shown.

Furthermore, the repetition rate must be lengthened because switching the proton decoupler on and off entails a perturbation of the proton magnetization. As a consequence, complete return to thermal equilibrium depends not only on carbon relaxation but also on proton relaxation. Overall, it is recommended to settle the repetition rate at 10 times the carbon-13 T_1 (according to the Solomon equations) and not to five times (as would be the case according to the Bloch equations). Finally, it can be mentioned that the continuous application of proton decoupling can be devastating in the case of nitrogen-15 spectroscopy; because ^{15}N and proton gyromagnetic ratios have opposite signs, this may lead to the disappearance of the ^{15}N signal in unfavorable conditions of cross-relaxation rate vs specific relaxation rate.

4.3 Overlapping Patterns and Spectral Analysis

We shall be concerned here with signals which overlap in the frequency domain owing to their line width.

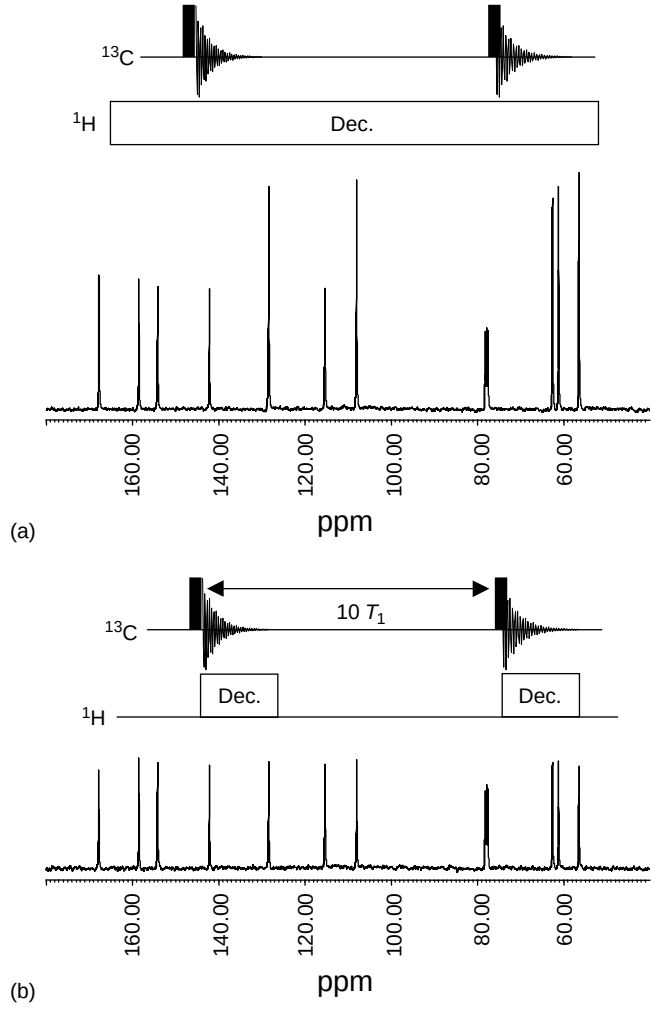

Figure 31 (a) A carbon-13 spectrum obtained under standard conditions. Except the multiplet at 78 ppm (deuterated solvent), each peak corresponds to one carbon in the same molecule. Notice the differences in peak heights. (b) The same spectrum obtained according to the sequence in the inset. All peak heights are almost identical; the remaining small differences are due to their line width (slight variations of the effective T_2) while peak intensities are effectively identical.

If we disregard the line broadening by B_0 homogeneity, this can occur because of large transverse relaxation rates impairing the direct measurement of line intensities. In that case a mathematical analysis must be carried out so as to separate in one way or another overlapping signals and thus pick up their individual characteristics. For practical reasons and to avoid artifacts associated with Fourier transformation, this is better performed in the time domain, that is, directly on the FID. We shall not present the numerous sophisticated methods, based on linear prediction or entropy maximization,[48] but rather rely on simple nonlinear least squares,[49] and consider the following target function

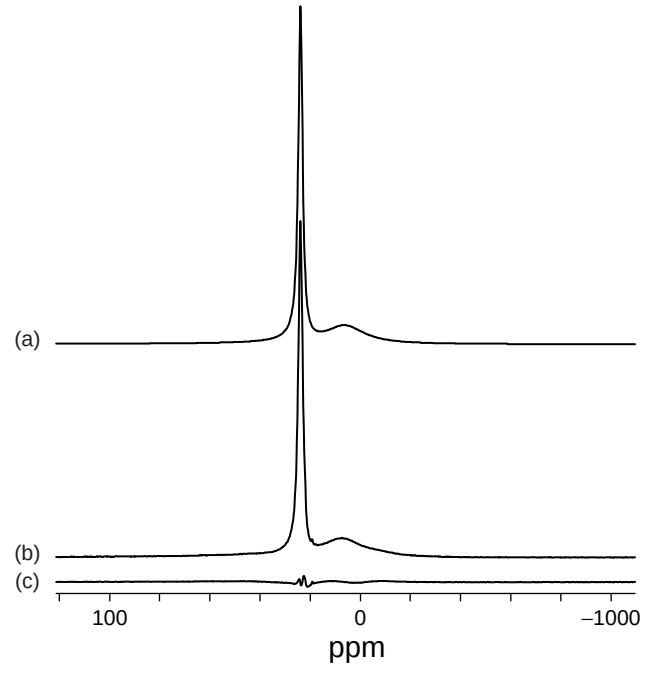

Figure 32 The successive traces represent the frequency domain of the ^{14}N spectrum involving two overlapping peaks whose characteristics could not have been obtained directly: (a) the reconstructed spectrum (after Fourier transformation); (b) the experimental spectrum; (c) the difference (b) − (a).

(Equation 116):

$$F = \sum_{j=1}^{N} \left(y_j - \sum_{k=1}^{K} A_k \exp[i(2\pi\nu k t_j + \varphi_k)] \exp(b_k t_j) \right)^2 \tag{116}$$

where y_j is the complex amplitude of the jth data point, acquired at time t_j, and the subscript k is associated with one of the K signals constituting the FID, assumed to lead to a Lorentzian line after Fourier transformation, ν_k is its frequency, φ_k a phase factor (depending on instrumental conditions), $b_k = 1/T_{2k}^*$ the inverse of its effective transverse relaxation time (including possibly B_0 inhomogeneity), and the quantity of interest, A_k, is the signal amplitude in the time domain which corresponds to the peak area in the frequency domain. These quantities are obtained by an iterative calculation aiming at the minimization of F and using as initial parameters those deduced from the Fourier transform of the FID. This procedure is exemplified by spectra (corresponding to the experimental and fitted FIDs) shown in Figure 32.

4.4 Partial Decoupling by Quadrupolar Nuclei

As reported in section 3.10, relaxation of nuclei of spin >1/2 is dominated by the so-called quadrupolar

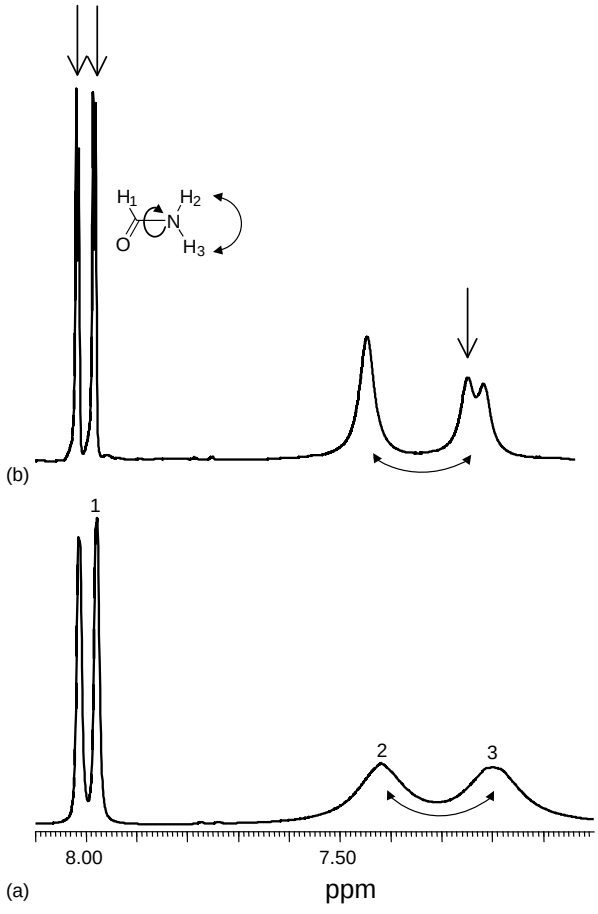

Figure 33 (a) The normal proton spectrum of formamide (with peaks broadened by $^{14}N-^{1}H$ couplings). (b) The same spectrum obtained under ^{14}N irradiation, which removes the effect of $^{14}N-^{1}H$ couplings. Notice that splittings due to $^{1}H-^{1}H$ couplings become visible. However, the remaining broadening is due to chemical exchange between the two amide protons resulting from the hindered rotation of the CN bond.

mechanism, provided that the considered nucleus is not in a symmetrical surrounding. It can be recalled that this mechanism is generally highly efficient leading to remarkably short T_1 and T_2, and hence to broad lines. In this respect, relaxation produces fast switching between spin states and acts as a decoupler of the quadrupolar nuclei, with respect to the other nuclei (of spin-1/2). It is for this reason that one seldom observes J splittings from nuclei such as ^{35}Cl, ^{37}Cl, ^{79}Br, ^{81}Br or ^{14}N, and that the only effect of a nearby quadrupolar nucleus is the so-called ''scalar relaxation of the second kind'' (see section 3.4). Now, if the surrounding of the quadrupolar nucleus becomes more or less symmetrical, the quadrupolar mechanism is no longer overwhelming and becomes comparable to other contributions. In other words, T_1 and T_2 have much longer values. As a consequence, the

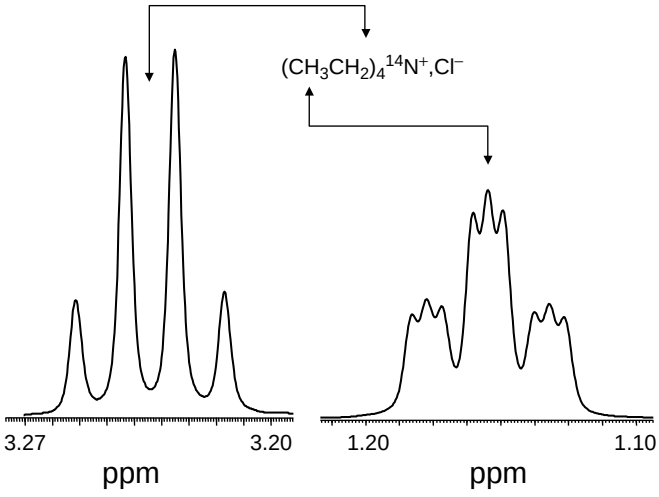

Figure 34 A further example of visible J couplings between ^{1}H and ^{14}N whenever the latter nucleus is in a quasi-symmetrical environment. Each line of the CH_3 triplet (the triplet is due to $^{1}H-^{1}H$ J coupling with the nearby CH_2) is further split into three lines of equal intensity due to the J coupling with the ^{14}N nucleus (which is a spin-1 nucleus, hence possessing three spin states, 1, 0 and -1).

decoupling effect tends to disappear and J splittings due to the quadrupolar nucleus are retrieved. These features, which can make the determination of line intensities difficult, are illustrated in Figures 33 and 34.

ABBREVIATIONS AND ACRONYMS

CPMG	Carr–Purcell–Meiboom–Gill
CSA	Chemical Shift Anisotropy
CW	Continuous Wave
EFG	Electric Field Gradient
EMF	Electromotive Force
FID	Free Induction Decay
FTNMR	Fourier Transform Nuclear Magnetic Resonance
HOESY	Heteronuclear Overhauser Effect Spectroscopy
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PAS	Principal Axis System
RF	Radiofrequency
ROESY	Rotating-frame Overhauser Effect Spectroscopy
S/N	Signal-to-noise
SUFIR	Super Fast Inversion–Recovery

RELATED ARTICLES

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. C.P. Slichter, *Principles of Magnetic Resonance*, Harper and Row, New York, 8, 1963.
2. R.R. Ernst, 'Sensitivity Enhancement in Magnetic Resonance', *Adv. Magn. Reson.*, **2**, 1–135 (1966).
3. R.L. Vold, J.S. Waugh, M.P. Klein, D.E. Phelps, *J. Chem. Phys.*, **48**, 3831–3837 (1968).
4. M. Sass, D. Ziessow, 'Error Analysis for Optimized Inversion–Recovery Spin–Lattice Relaxation Measurement', *J. Magn. Reson.*, **25**(2), 263–276 (1977).
5. R. Freeman, S.P. Kempsell, M.H. Levitt, 'Radiofrequency Pulse Sequences Which Compensate Their Imperfection', *J. Magn. Reson.*, **38**(3), 453–479 (1980).
6. M.H. Levitt, R. Freeman, T. Frenkiel, 'Broadband Decoupling in High Resolution Nuclear Magnetic Resonance Spectroscopy', *Adv. Magn. Reson.*, **11**, 47–110 (1983).
7. D. Canet, G.C. Levy, I.R. Peat, 'Time Saving in ^{13}C Spin–Lattice Relaxation by Inversion–Recovery', *J. Magn. Reson.*, **18**(1), 199–204 (1975).
8. D. Canet, J. Brondeau, K. Elbayed, 'Superfast T_1 Determination by Inversion–Recovery', *J. Magn. Reson.*, **77**(3), 483–486 (1988).
9. A. Abragam, *The Principle of Nuclear Magnetism*, Clarendon Press, Oxford, 73–74, 1961.
10. E. Hahn, 'Spin Echoes', *Phys. Rev.*, **80**(4), 580–594 (1950).
11. P.J. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Clarendon Press, Oxford, Chapters 1 and 6, 1991.
12. H.C. Torrey, 'Bloch Equations with Diffusion Terms', *Phys. Rev.*, **104**(3), 563–566 (1956).
13. H.Y. Carr, E.M. Purcell, 'Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments', *Phys. Rev.*, **94**(3), 630–638 (1954).
14. S. Meiboom, D. Gill, 'Modified Spin-echo Method for Measuring Nuclear Relaxation Times', *Rev. Sci. Instrum.*, **29**, 688–691 (1958).
15. E.L. Hahn, D.E. Maxwell, 'Chemical Shift and Field Independent Frequency Modulation of the Spin Echo Envelope', *Phys. Rev.*, **84**, 1246–1247 (1951).
16. B. Tiffon, B. Ancian, J.-E. Dubois, 'Accurate Measurement of Long Carbon 13 Spin–Spin Relaxation Times by the Spin-echo Fourier Transform Method with Carbon Disulfide as Example', *Anal. Chim. Acta*, **124**, 415–419 (1981).
17. D.C. Look, I.J. Lowe, 'Nuclear Magnetic Dipole–Dipole Relaxation Along the Static and Rotating Magnetic Field: Application to Gypsum', *J. Chem. Phys.*, **44**(8), 2995–3000 (1966).
18. J.H. Noggle, R.E. Schirmer, *The Nuclear Overhauser Effect. Chemical Applications*, Academic Press, New York, Chapter 1, 1971.
19. J.J. Delpuech, 'Timescales in NMR: Nuclear Site Exchange and Dynamics NMR', in *Dynamics of Solutions and Fluids Mixtures by NMR*, ed. J.J. Delpuech, Wiley, Chichester, 73–172, 1995.
20. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York, 1989.
21. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principle of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, Chapter 6, 1987.
22. J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, 'Investigation of Exchange Processes by Two Dimensional NMR Spectroscopy', *J. Chem. Phys.*, **71**(11), 4546–4553 (1979).
23. P.L. Rinaldi, 'Heteronuclear 2D-NOE Spectroscopy', *J. Am. Chem. Soc.*, **105**(15), 5167–5168 (1983).
24. C. Yu, G.C. Levy, 'Solvent and Intramolecular Proton Dipolar Relaxation of the Three Phosphates of ATP: a Heteronuclear 2D NOE Study', *J. Am. Chem. Soc.*, **105**(16), 6994–6996 (1983).
25. A.A. Bothner-By, R.L. Stephens, J.-M. Lee, C.D. Warren, R.W. Jeanloz, 'Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame', *J. Am. Chem. Soc.*, **106**(3), 811–813 (1984).
26. A. Bax, D.G. Davis, 'Practical Aspects of Two-dimensional Transverse NOE Spectroscopy', *J. Magn. Reson.*, **63**(1), 207–213 (1985).
27. E.L. Mac Kor, C. Mac Lean, 'Relaxation Processes in Systems of Two Non-identical Spins', *Prog. Nucl. Magn. Reson. Spectrosc.*, **3**, 129–157 (1967).
28. M. Goldman, 'Interference Effects in the Relaxation of a Pair of Unlike Spin-1/2 Nuclei', *J. Magn. Reson.*, **60**(3), 437–452 (1984).
29. O.W. Sorensen, G.W. Eich, M.H. Levitt, G. Bodenhausen, R.R. Ernst, 'Product Operator Formalism for Description of NMR Pulse Experiments', *Prog. Nucl. Magn. Reson. Spectrosc.*, **16**(2), 163–192 (1983).

30. M. Guéron, J.L. Leroy, R.H. Gaiffey, 'Proton Nuclear Magnetic Relaxation of ^{15}N-labelled Nucleic Acids via Dipolar Coupling and Chemical Shift Anisotropy', *J. Am. Chem. Soc.*, **105**(25), 7262–7266 (1983).
31. L.G. Werbelow, D.M. Grant, 'Intramolecular Dipolar Relaxation in Multispin Systems', *Adv. Magn. Reson.*, **9**, 190–299 (1977).
32. R.L. Vold, R.R. Vold, 'Nuclear Magnetic Relaxation in Coupled Spin Systems', *Prog. Nucl. Magn. Reson. Spectrosc.*, **12**, 79–133 (1978).
33. D. Canet, 'Construction, Evolution and Detection of Magnetization Modes Designed for Treating Longitudinal Relaxation of Weakly Coupled Spin 1/2 Systems with Magnetic Equivalence', *Prog. Nucl. Magn. Reson. Spectrosc.*, **21**(3), 237–291 (1989).
34. N. Tjandra, A. Szabo, A. Bax, 'Protein Backbone Dynamics and ^{15}N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects', *J. Am. Chem. Soc.*, **118**(29), 6986–6991 (1996).
35. D. Canet, *Nuclear Magnetic Resonance, Concepts and Methods*, Wiley, Chichester, Chapter 4, 1996.
36. G. Lipari, A. Szabo, 'Model-free Approach to the Interpretation of Nuclear Magnetic Relaxation in Macromolecules 1. Theory and Range of Validity', *J. Am. Chem. Soc.*, **104**(17), 4546–4559 (1982).
37. H. Wennerström, B. Lindman, O. Söderman, T. Drakenbers, J.B. Rosenholm, '^{13}C Magnetic Relaxation in Micellar Solutions. Influence of Aggregate Motion on T_1', *J. Am. Chem. Soc.*, **101**(73), 6860–6864 (1979).
38. R. Kimmich, *NMR, Tomography, Diffusometry, Relaxometry*, Springer, Berlin, Chapters 15 and 16, 1997.
39. A.G. Redfield, 'The Theory of Relaxation Processes', *Adv. Magn. Reson.*, **1**, 1–32 (1965).
40. T.E. Bull, *Prog. Nucl. Magn. Reson. Spectrosc.*, **24**, 377–410 (1992).
41. L.G. Werbelow, 'Relaxation-induced Transfer of Nuclear Spin Polarization as a Probe of Molecular Structure and Dynamics in Mobile Phase', in *Nuclear Magnetic Resonance Probes of Molecular Dynamics*, ed. R. Tycko, Kluwer Academic Publishers, Dordrecht, 223–263, 1994.
42. N. Tjandra, A. Bax, 'Solution NMR Measurement of Amide Proton Chemical Shift Anisotropy in ^{15}N-enriched Proteins. Correlation with Hydrogen Bond Length', *J. Am. Chem. Soc.*, **119**(34), 8076–8082 (1997).
43. M. Goldman, 'Interference Effects in the Relaxation of a Pair of Unlike Spin 1/2 Nuclei', *J. Magn. Reson.*, **60**, 437–452 (1984).
44. B. Halle, H. Wennerström, 'Nearly Exponential Quadrupolar Relaxation. A Perturbation Treatment', *J. Magn. Reson.*, **44**(1), 89–100 (1989).
45. L. Banci, I. Bertini, C. Luchinat, *Nuclear and Electron Relaxation*, VCH, Weinheim, Chapter 5, 1991.
46. J. Kowalewski, 'Nuclear Spin Relaxation in Diamagnetic Fluids. Part 1: General Aspects and Inorganic Applications', *Annu. Rep. NMR Spectrosc.*, **22**, 306–312 (1989).
47. J. Kowalewski, 'Nuclear Spin Relaxation in Diamagnetic Fluids. Part 2: Organic Systems and Solutions of Macromolecules and Aggregates', *Annu. Rep. NMR Spectrosc.*, **23**, 289 (1991).
48. R. de Beer, D. van Ormondt, 'Analysis of NMR Data Using Time Domain Fitting Procedures', *NMR Basic Principles Prog.*, **26**, 201–248 (1992).
49. F. Montigny, J. Brondeau, D. Canet, 'Analysis of Time-domain NMR Data by Standard Non-linear Least Squares', *Chem. Phys. Lett.*, **170**(2,3), 175–180 (1990).

Scalar Couplings in Nuclear Magnetic Resonance, General

Jo Woong Lee
Seoul National University, Seoul, Korea

1 Introduction 1
2 Theory of Scalar Spin Couplings 2
3 Correlation with Molecular Properties 5
3.1 Coupling between Directly Bonded Nuclei 5
3.2 Geminal and Vicinal Coupling 6
3.3 Long-range Coupling 7
3.4 Coupling through π Electrons 8
3.5 Through-space Coupling 9
4 Measurement of Scalar Spin Coupling Constants 9
Abbreviations and Acronyms 13
Related Articles 14
References 14

Indirect spin coupling refers to the magnetic interaction operative between a pair of magnetic nuclei in a molecule that is mediated by bonding electrons. The anisotropic part of this interaction averages out to zero when the internuclear vector connecting the two nuclei randomly changes its orientation in the isotropic fluid state as does the direct magnetic dipolar interaction between the same pair of nuclei which operates through space. The isotropic part, which is usually called the scalar spin coupling, survives such random modulations of the internuclear vector in the fluid state to give rise to splitting of the nuclear magnetic resonance (NMR) lines which would otherwise represent a mere collection of individual absorption lines corresponding to the chemical shift of each constituent spin. Adequate analysis of the multiplet structure of the NMR spectrum arising from a given set of nuclear spins can provide knowledge of the scalar coupling between each pair of nuclear spins which in turn yields a wealth of information on molecular structure as well as molecular processes occurring in the system.

1 INTRODUCTION

Direct magnetic dipole–dipole interaction between two magnetic nuclei can be shown to average out to zero in the isotropic fluid phase when the internuclear vector connecting these two nuclei randomly changes its spatial direction due to tumbling motions of the molecule on which they are embedded. Despite random molecular tumbling, however, it is usually found that the interaction between two magnetic nuclei is not always vanishing as manifested by the spectral splittings it causes. Hence this nonvanishing type of interaction can be thought to occur indirectly through mediation of electrons surrounding the two interacting nuclei, which we call nowadays indirect spin interaction (or coupling). This interaction was first observed by Proctor and Yu[1] as magnetic-field-independent splitting due to the coupling between ^{121}Sb and ^{19}F of the SbF_6^- anion in an aqueous solution of $NaSbF_6$. Shortly after this Gutowsky and McCall[2] reported on the spin coupling between ^{31}P and ^{19}F in $POCl_2F$, $POClF_2$, and CH_3OPF_2 which was followed by Hahn and Maxwell[3] with their report on H–H coupling in $CHCl_2CHO$. Based on their observations Gutowsky and McCall and Hahn and Maxwell independently proposed that the Hamiltonian $H_{NN'}$ describing this field-independent interaction in isotropic fluids should take the form of a scalar product as shown in Equation (1):

$$h^{-1}H_{NN'} = J_{NN'}\mathbf{I}_N \cdot \mathbf{I}_{N'} \tag{1}$$

where $\mathbf{I}_N$ and $\mathbf{I}_{N'}$ are respectively, the spin angular momentum operators for the nuclear spins N and N' and $J_{NN'}$ is the isotropic spin–spin coupling constant for the given pair of spins. [When there are several coupling pairs of magnetic nuclei, it is understood that Equation (1) is pairwise added.]

However, as Ramsey[4] pointed out, when the internuclear vector connecting the coupled nuclei is not randomly tumbling as it does in an isotropic fluid, the indirect spin interaction can depend on the spatial orientation of the internuclear vector with respect to the applied static magnetic field. The corresponding Hamiltonian expressed in the form of Equation (1) then should be modified to Equation (2):

$$h^{-1}H_{NN'} = \mathbf{I}_N \cdot \mathbf{J}_{NN'} \cdot \mathbf{I}_{N'} \tag{2}$$

where $\mathbf{J}_{NN'}$ is a second-rank tensor called the indirect spin coupling tensor.

The indirect spin coupling tensor may be written like any other tensor of rank 2 as a sum of two parts, the symmetric and the antisymmetric part,[5] as shown in Equation (3)

$$\mathbf{J}_{NN'} = \mathbf{J}^{\text{sym}}_{NN'} + \mathbf{J}^{\text{anti}}_{NN'} \tag{3}$$

where the symmetric part $\mathbf{J}^{\text{sym}}_{NN'}$ is defined by Equation (4):

$$\mathbf{J}^{\text{sym}}_{NN'} = \tfrac{1}{2}(\mathbf{J}_{NN'} + \mathbf{J}^{\text{T}}_{NN'}) \tag{4}$$

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

while the antisymmetric part $\mathbf{J}^{\text{anti}}_{NN'}$ may be expressed by Equation (5):

$$\mathbf{J}^{\text{anti}}_{NN'} = \tfrac{1}{2}(\mathbf{J}_{NN'} - \mathbf{J}^{\text{T}}_{NN'}) \tag{5}$$

with $\mathbf{J}^{\text{T}}_{NN'}$ denoting the transpose of $\mathbf{J}_{NN'}$. This tensor is antisymmetric with respect to exchange of row and column, whence its diagonal elements are all zero. For convenience the symmetric part $\mathbf{J}^{\text{sym}}_{NN'}$ is also divided into the isotropic part $\mathbf{J}^{\text{iso}}_{NN'}$ and the traceless part $\mathbf{J}^{\text{sym(0)}}_{NN'}$ which are defined, respectively, by Equations (6) and (7):

$$\mathbf{J}^{\text{iso}}_{NN'} = \left(\tfrac{1}{3}\text{Tr}\mathbf{J}_{NN'}\right)\mathbf{1} \tag{6}$$

$$\mathbf{J}^{\text{sym(0)}}_{NN'} = \tfrac{1}{2}\left(\mathbf{J}_{NN'} + \mathbf{J}^{\text{T}}_{NN'}\right) - \mathbf{J}^{\text{iso}}_{NN'} \tag{7}$$

where $\text{Tr}\mathbf{J}_{NN'}$ is the sum of three diagonal elements of the tensor $\mathbf{J}_{NN'}$ while $\mathbf{1}$ denotes the unit tensor. To the first-order approximation we may concern ourselves only with the symmetric part of $\mathbf{J}_{NN'}$ tensor because the contribution from the antisymmetric part becomes appreciable only in second- or higher-order calculation. Thus one may replace the full tensor $\mathbf{J}_{NN'}$ by its truncated form $\mathbf{J}^{\text{iso}}_{NN'} + \mathbf{J}^{\text{sym(0)}}_{NN'}$, which is symmetric and can be diagonalized, in our ensuing discussion.

Let us denote the three principal components of $\mathbf{J}_{NN'}$ by J_{xx}, J_{yy}, and J_{zz} (dropping the subscript NN' for brevity), with the convention shown in Equation (8):

$$|J_{zz}| \geq |J_{xx}| \geq |J_{yy}| \tag{8}$$

In order to describe the orientation dependence of NMR parameters it is convenient to define the anisotropy of that particular parameter. In the case of the indirect spin–spin coupling tensor we usually define its anisotropy by Equation (9):

$$\Delta J \equiv J_{zz} - \tfrac{1}{2}(J_{xx} + J_{yy}) \tag{9}$$

One of the easily tractable cases is when the two coupling nuclei are located on an axis of cylindrical symmetry, for which one may write Equation (10):

$$J_{zz} \equiv J_{\parallel} \quad \text{and} \quad J_{xx} = J_{yy} \equiv J_{\perp} \tag{10}$$

Therefore, in the presence of axial symmetry Equation (9) takes the form shown in Equation (11):

$$\Delta J = J_{\parallel} - J_{\perp} \tag{11}$$

In this case the Hamiltonian (Equation 2) may be expressed by Equation (12):

$$h^{-1}H_{NN'} = J\mathbf{I}_N \cdot \mathbf{I}_{N'} + \mathbf{I}_N \cdot \mathbf{J}' \cdot \mathbf{I}_{N'} \tag{12}$$

where the scalar spin coupling constant J means $\frac{1}{3}\text{Tr}(\mathbf{J}_{NN'})$ and the traceless tensor $\mathbf{J}'$ is given by Equation (13):

$$\mathbf{J}' = -\Delta\frac{J}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \tag{13}$$

In fluid phases the anisotropic part in the Hamiltonian (Equation 12) involving the tensor $\mathbf{J}'$ averages out to zero due to random tumbling of the principal-axes frame of this tensor, and only the isotropic part of $\mathbf{J}$ survives and affects the appearance of NMR spectra for the given spin system. The term ''scalar coupling'' invariably refers to the isotropic part of indirect spin coupling because it can be expressed in the form of a scalar product of spin angular momentum vectors of the two coupling nuclei. Based on the Hamiltonian (Equation 12) one may derive an expression explaining the spectral splitting caused by the indirect spin coupling which depends on the angle between the applied static field and internuclear axis. Thus, in principle one can obtain information on J and ΔJ by observing the spectral splittings due to indirect spin coupling in single crystals or powder samples as a function of this angle and in this way these parameters have been measured and reported for some molecules.[5] However, the anisotropic part of $\mathbf{J}$ transforms exactly like direct magnetic dipole–dipole interaction and, therefore, it will be difficult to disentangle the effect of the former from that of the latter and due to this accurate data on ΔJ are rare. Over the last 45 years most information on indirect spin coupling has been obtained mainly from the isotropic part and correlated with molecular structure in the liquid state. Therefore, in this article attention will be paid only to the isotropic part of $\mathbf{J}$.

2 THEORY OF SCALAR SPIN COUPLINGS

Shortly after the discovery of indirect nuclear spin–spin coupling, a suitable quantum mechanical interpretation for the mechanism of this interaction was provided for the first time by Ramsey and Purcell,[6] and later by Ramsey[4] in greater detail. For heavier nuclei the relativistic effect will undoubtedly be important,[7] but we restrict ourselves here to Ramsey's nonrelativistic description of indirect spin couplings. Although formal mathematical expressions for scalar spin coupling constants may be derived on the basis of perturbation theory, actual evaluation of these parameters from the derived expressions is a formidable task because we have little knowledge of the exact wavefunctions of the electronically excited states. Nevertheless, we can draw some qualitatively correct conclusions without performing complicated quantum mechanical calculations with the help

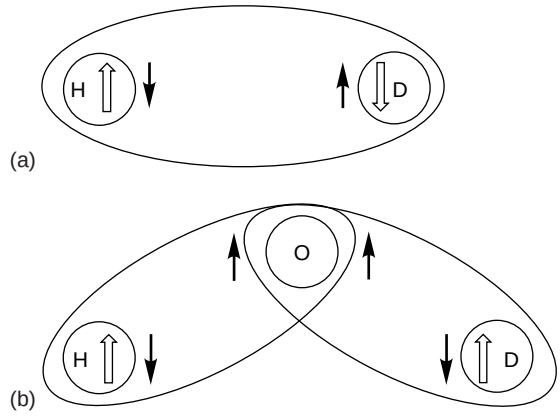

Figure 1 Dirac vector model for (a) directly bonded coupling and (b) geminal coupling between H and D in (a) HD and (b) HDO.

of a few assumptions and semiempirical results, which we will describe in the ensuing discussion.

Ramsey has proposed that Fermi contact interactions between the nuclei and electrons provide the major source of contribution to indirect spin coupling. When an electron occupies an s-orbital, it has nonvanishing density at the nuclear site. This causes the so-called contact interaction which was first studied by Fermi. How the indirect spin coupling is brought forth by the contact interaction may be understood in the following manner. The contact interaction favors the antiparallel orientation of an electron spin at a nuclear spin if the magnetogyric ratio of the nucleus is positive; in other words, the electron and nuclear magnetic moments are stable when parallel. Also, we note that electrons occupying the same orbital tend to have their spins aligned antiparallel (Pauli interaction) while those in the different orbitals favor the parallel spins (Hund-type interaction). Based on this simple model we are able to provide some physical pictures of scalar coupling transmitted through chemical bonds. First, consider a diatomic molecule, say, HD. This system of two directly bonded nuclei shown in Figure 1(a) becomes stabilized by having two nuclear spins antiparallel. In this case we say J is positive. (In general, J is not always positive for a pair of coupled nuclei.) The same logic may be applied to the case where there is an intervening atom between the two coupled nuclei to predict that J will be negative (Figure 1b) and smaller in magnitude compared with the case of directly bonded nuclei because correlations between electrons occupying different orbitals are expected to be less than those between electrons occupying the same orbital. This contact mechanism is important for (H, H) coupling because the electron densities about electrons are well represented by 1s atomic orbitals and may be extended further to the case of a larger number of bonds and we may anticipate that the sign of J alternates while its magnitude decreases rapidly as the number of intervening bonds increases. This simple model is based on the assumption that J coupling is transmitted through bonding σ electrons and the nucleus–electron interaction is dominated by the contact interaction. Although this model, frequently referred to as the Dirac vector model, is an oversimplification of the actual situation and there are invariably contradicting exceptions, it provides some qualitative explanation of general trends in variation of sign and magnitude of J. For example, for saturated hydrocarbons, (H, H) coupling is rarely observed when two protons are separated from each other by more than four or five bonds, but the attenuation of the magnitude of J is known to be less steep than the Dirac model would suggest.

To distinguish between number of intervening bonds we adopt the convention that we add a superscript prefix to the symbol J. Thus coupling between directly bonded nuclei is denoted by ${}^1J$ and coupling through two and three bonds, often known as geminal and vicinal coupling, is indicated by ${}^2J$ and ${}^3J$, respectively. The interproton coupling constant for the H_2 molecule is known to be ${}^1J_{HH} = 284\,Hz$ while the geminal (H, H) coupling constant for the CH_2 group in saturated hydrocarbons ranges from -9 to $-21\,Hz$, which is over an order of magnitude smaller than ${}^1J_{HH}$. Clearly, the Dirac model fits well in this case. However, when the intervening carbon atom is sp^2 hybridized, it is known that ${}^2J_{HH}$ is normally positive. This indicates that coupling through more than one bond may involve delocalization of electrons in contradiction to the Dirac vector model. Vicinal (H, H) coupling constants are of the same order of magnitude as geminal coupling constants but their signs are always positive. Frequently, we encounter the situation $|{}^2J_{HH}| < |{}^3J_{HH}|$, which means the rationale behind indirect spin coupling is not as simple as the Dirac model may suggest.

The quantum mechanical perturbation treatment leads to Equation (14) for the contribution to J from the contact mechanism:

$$J^c_{NN'} = -\frac{8\mu_0^2\mu_B^2\hbar}{9\pi}\gamma_N\gamma_{N'}\sum_{n(\neq 0)}\sum_k\sum_j \frac{1}{E_n - E_0} \times \langle 0|\delta(\mathbf{r}_{kN})\mathbf{S}_k|n\rangle \cdot \langle n|\delta(\mathbf{r}_{jN'})\mathbf{S}_j|0\rangle \qquad (14)$$

where $|0\rangle$ and $|n\rangle$ are respectively, the ground spin singlet state and the excited spin triplet state with E_0 and E_n denoting the corresponding energy eigenvalues. Other notations appearing in Equation (14) have the following meanings: μ_0 is the magnetic permeability of a vacuum, μ_B is the Bohr magneton, γ_N is the magnetogyric ratio of nucleus N, $\boldsymbol{S}_j$ is the spin angular momentum operator for

electron j, $\boldsymbol{r}_{jN}$ is the position vector of electron j relative to nucleus N, and $\delta(\boldsymbol{r}_{jN})$ is the Dirac delta function of $\boldsymbol{r}_{jN}$.

Although less dominant than the contact mechanism, the direct magnetic dipolar interaction of the nucleus with surrounding electrons can also be shown to make a contribution to J. That is, the magnetic moment of one of the coupled nuclei directly interacts with that of an electron spatially close to (but not in contact with) it which then correlates with another electron whose magnetic moment directly interacts with that of the other nucleus. This type of contribution can be shown to have the form shown in Equation (15):

$$J^{d}_{NN'} = -\frac{\mu_0^2\mu_B^2\hbar}{24\pi^3}\gamma_N\gamma_{N'}\sum_{n(\neq 0)}\sum_{k}\sum_{j}\frac{1}{E_n - E_0} \times \langle 0|3(\mathbf{S}_k\cdot\mathbf{r}_{kN})\mathbf{r}_{kN}r_{kN}^{-5} - \mathbf{S}_k r_{kN}^{-3}|n\rangle \times \langle n|3(\mathbf{S}_j\cdot\mathbf{r}_{jN'})\mathbf{r}_{jN'}r_{jN'}^{-5} - \mathbf{S}_j r_{jN'}^{-3}|0\rangle \quad (15)$$

Besides the contributions via interaction between nuclear spin and electron spin there can be another type of contribution to J as Gutowski et al.[8] have shown. That is, the electronic orbital motions induced by the magnetic moment of one nucleus may interact with the magnetic moment of the other nucleus in such a way that interaction between these two nuclei might not be averaged to zero by random molecular motions. The formal expression for this type of contribution obtained by the perturbation treatment is given by Equations (16) and (17):

$$J^{o}_{NN'} = \frac{\mu_0^2 e^2\hbar}{48\pi^3 m}\gamma_N\gamma_{N'}\sum_{k}\langle 0|\mathbf{r}_{kN}\cdot\mathbf{r}_{kN'}r_{kN}^{-3}r_{kN'}^{-3}|0\rangle - \frac{\mu_0^2\mu_B^2\hbar}{6\pi^3}\gamma_N\gamma_{N'}\sum_{n(\neq 0)}\sum_{k}\sum_{j}\frac{1}{E_n - E_0} \times \langle 0|\mathbf{m}_{kN}r_{kN}^{-3}|n\rangle\cdot\langle n|\mathbf{m}_{jN'}r_{jN'}^{-3}|0\rangle \quad (16)$$

where

$$\mathbf{m}_{kN} = \frac{e}{2m}\,\mathbf{r}_{kN}\times\left(\frac{\hbar}{i}\,\nabla_k\right) \quad (17)$$

and m, e, and ∇_k, respectively, stand for mass and charge of an electron and the gradient operator acting on the coordinates of electron k. Note that $J^{o}_{NN'}$ in Equation (16) consists of two parts: one- and two-electron contributions. Physically, the one-electron parts arise because nuclear N induces a local electron perturbation, producing a magnetic field which then acts across space on the second nucleus. The two-electron parts, on the other hand, arise because one nucleus induces a perturbation on its own atom, which in turn affects the distribution of a second electron on the other atom, and so on the second nucleus.

As mentioned earlier, exact calculation of the contributions to the scalar coupling constant from various sources on the basis of Equations (14), (15) and (16) is difficult unless the exact wavefunctions for the excited states are known. Thus, in earlier studies the so-called AEE (average excitation energy) approximation was adopted to alleviate this difficulty, in which the energy difference $E_n - E_0$ appearing in the above expressions for contributions to J is replaced by an AEE ΔE regardless of n.

Ramsey has considered only the contact term to derive Equation (18) giving J^{c}_{HD} for a HD molecule on the basis of AEE and electronic ground state function reported by James and Coolidge[9]

$$J^{c}_{HD} = \frac{2\mu_0^2\mu_B^2\hbar}{3\pi}\gamma_H\gamma_D\frac{1}{\Delta E}S_H(0)^2S_D(0)^2 = \frac{758.9}{\Delta E}\,(\text{Hz}) \quad (18)$$

where $S_H(0)^2S_D(0)^2$ is the electron density on the coupled proton and deuteron and ΔE the singlet–triplet excitation energy expressed in electron-volt units. A reasonable value of $\Delta E \approx 19.0$ eV gives a coupling constant of 40 Hz, which is close to the experimental value of 42.94 Hz. This indicates that the contribution from the contact term dominates those from other terms in the case of $^1J_{HD}$. A few years later, McConnell[10] suggested a simpler computational scheme on the basis of molecular orbital (MO) theory while Karplus et al.[11,12] proposed another scheme based on the valence bond (VB) method. In both approaches the AEE approximation was adopted. Within the frame of AEE approximation both the MO method and the VB method predict that the contribution from the contact mechanism is proportional to $\gamma_N\gamma_{N'}S_N(0)^2S_{N'}(0)^2$ where $S_N(0)^2$ denotes the density at the nucleus N of the valence s-electrons.

Coupling constants involving heavier nuclei are usually much larger than J_{HH}. For example, it was reported that $^1J(^{199}\text{Hg}, ^{31}\text{P}) = 9.66$ kHz and $^3J(^{205}\text{Tl}, ^{13}\text{C}) \approx 1$ kHz. This is mainly due to the fact that for heavier nuclei the valence s-electron density is much higher than for protons. But other mechanisms are also known to make substantial contributions in the case of heavier nuclei. To date, however, most extensively investigated and best understood from a theoretical point of view are the couplings between nuclei comprising organic molecules, and we will concern ourselves mainly with these nuclei.

The AEE approximation greatly simplifies the calculational procedure but is not always justified. Later, the calculations were carried out without invoking this approximation by Pople and others,[13,14] mainly using the MO expressions arising from Rayleigh–Schrödinger's "sum over states" method, finite perturbation theory, and self-consistent perturbation theory. A general VB approach developed by Karplus et al. was also successfully applied to the interpretation of the dependence of (H, H) vicinal couplings on the dihedral angle and of geminal coupling on the bond angle between the coupled

protons. However, mathematical difficulties associated with the VB method have made it less readily applicable than the MO approaches for calculation of spin–spin coupling constants. Until Barfield[15] proposed a new method based on the transition spin-density formulation, all the VB approaches had resorted to the AEE approximation. Theoretical calculations of coupling constants based on the various methods published thus far have all yielded results that agree qualitatively with experimental values both in sign and magnitude and can be helpful for understanding the nature of bonding and stereochemistry in relevant molecules. According to these calculations the contact term generally seems to be dominant for ^1J(C, H) and ^1J(C, C) coupling. However, orbital and dipolar contributions become significant for ^1J coupling involving multiple-bonded carbon atoms, fluorine, and other heavier atoms. We will not duplicate these data here and interested readers are referred to other NMR literature.[16–20]

In view of expressions for J derived thus far (see Equations 14, 15, and 16) it is obvious that J is proportional to the product of magnetogyric ratios of a coupled pair of nuclei. This is because all the interactions contributing to the indirect spin coupling are of magnetic origin and involve products of nuclear magnetic moments. This enables us to evaluate the coupling constants between magnetically equivalent nuclei. (See section 4 for the meaning of magnetic equivalence.) The couplings between magnetically equivalent nuclei cannot be directly observed because they cannot affect the NMR spectra. For example, for the H_2 molecule we cannot observe splitting of the proton absorption due to scalar spin coupling between two protons because they are magnetically equivalent. However, substitution of one of the two protons by a deuteron will result in splitting of the proton absorption line, from which we can measure $|J_{HD}|$. The observed value of $|J_{HD}|$ is known to be 42.94 Hz. Then, by making use of Equation (19):

$$\frac{J_{HH}}{J_{HD}} = \frac{\gamma_H}{\gamma_D} = 6.514 \tag{19}$$

we have $|J_{HH}| = 279.7$ Hz.

Table 1 Coupling constants $^1J_{XH}$ for group IV elements

Nuclei (X)	Molecule	$^1J_{XH}$ (Hz)	$K(10^{20}\,\mathrm{N\,A^{-2}\,m^{-3}})$
C	CH_4	+125	+4.1
Si	SiH_4	−202	+8.5
Ge	GeH_4	(−)98	(+)23.3
Sn	SnH_4	(−)1931(^{119}Sn)	+43.0

Adapted from Table 8–13 in Harris.[16]

Because of the fact that J is proportional to the product of magnetogyric ratios of the coupled pair of nuclei comparison between coupling constants of different elements will be meaningful only when the effect of magnetogyric ratios is eliminated. For this purpose one may define reduced coupling constants, usually denoted by K, using Equation (20):

$$K_{NN'} = \frac{4\pi^2 J_{NN'}}{h\gamma_N\gamma_{N'}} \tag{20}$$

Since the dependence on magnetogyric ratios has been eliminated, K shows better correlation with the variation of atomic number in a group.[16] For example, as shown in Table 1, the data for $^1J_{XH}$ in group IV elements (C, Si, Ge, Sn) show no trend in variation, even changing irregularly in sign. However, those for $^1K_{XH}$ for the same elements show a smooth trend, increasing with the atomic number (Z_X) of X. In fact, there has been an attempt to establish a universal relationship between $^1K_{XY}$ in isostructural molecules containing isoelectronic valence shells for the relevant atoms. Considering that the contact term is dominating and the relative coupling constants depend on the square of atomic s-electron charge densities at each nucleus, Reeves et al.[21–23] have proposed the relationship shown in Equation (21) for a given nucleus Y:

$$\sqrt{{}^1K_{XY}} = aZ_X + b \tag{21}$$

Equation (21) is known to be obeyed poorly if one of the coupling nuclei is fluorine or a heavy metal.

3 CORRELATION WITH MOLECULAR PROPERTIES

It is known that there are several factors influencing the sign and magnitude of the coupling constant and in this section we will discuss them in conjunction with molecular structure.

3.1 Coupling between Directly Bonded Nuclei

Ramsey's result (Equation 18) suggests that the nature of bonding can be understood by means of spin–spin coupling when the contact term is of greatest importance. Thus, it is not surprising that the relation shown in Equation (22) was proposed for $^1J(^{15}\mathrm{N}-^{13}\mathrm{C})$[24]

$$|{}^1J(^{15}\mathrm{N}-^{13}\mathrm{C})| \propto \%S_C\%S_N \tag{22}$$

where $\%S_C$ and $\%S_N$ are the percent s-character of the atomic orbitals forming the σ bond between the ^{15}N and ^{13}C nuclei. Finite perturbation theory/intermediate

neglect of differential overlap calculations developed by Pople et al.[14] show that the proportionality constant in Equation (22) should be 0.0106 and the sign of the coupling constant should be negative. (However, the same calculation shows that the contact term is not as dominant as in the case of $^1J_{HH}$.) Similar relations were also proposed. For example, for $^{13}C-^1H$ and $^{13}C-^{13}C$ coupling Equations (23)[25] and (24)[26,27] were proposed, respectively:

$$^1J(^{13}C-^1H) = 5.00\%S_{CH} \quad (23)$$

$$^1J(^{13}C_1-^{13}C_2) = 0.073\%S_{C_1}\%S_{C_2} - 17 \quad (24)$$

For a highly strained ring system Newton et al.[28–30] proposed, instead of Equation (23), Equation (25):

$$^1J(^{13}C_1-^{13}C_2) = 0.0621\%S_{C_1}\%S_{C_2} - 10.2 \quad (25)$$

Equation (23) indicates that hybridization of the carbon atom is important and is helpful for estimating the s-character of the C–H bond. For example, for the sp^3 bond as in methane and ethane $\%S$ is $\frac{1}{4}$, i.e. 25%, whence $^1J_{CH} = 125$ Hz, and for sp^2 and sp, as in ethylene and acetylene, $\%S$ are $\frac{1}{3}$ and $\frac{1}{2}$, respectively, with the corresponding one-bond coupling constant being 167 and 250 Hz. Measured coupling constants for ethane, ethylene, and acetylene are known to be 124.9, 156.4, and 248.7 Hz, respectively. Although the relation does not yield quantitatively satisfactory results, its use for qualitative estimation is thus justified. Introduction of an electronegative substituent usually increases $^1J_{CH}$ by increasing the s-character of the C–H bond and this effect is usually additive, but multiple substitution by highly electronegative groups induces some deviation from the trend of additivity, for which we need the introduction of some extra terms to account for pairwise interactions between substituents.

3.2 Geminal and Vicinal Coupling

Of particular interest are $^2J_{H-H}$ and $^3J_{H-H}$ because of their dependence on bond angles. Experiments show that in hydrocarbons geminal $^2J_{H-H}$ increases algebraically with increasing H–C–H bond angle, which was also confirmed by theoretical calculations. In sp^3 carbon, as in methane derivatives, $^2J_{H-H}$ is negative while it takes a positive value for sp^2 carbon (−12.4 Hz for CH_4 and +2.5 Hz for $CH_2{=}CH_2$). The $^2J_{H-H}$ values are known to be influenced by both α and β substituents.

Substitution of an electronegative group into the methylene group in question (α substitution) induces a positive change in the coupling constant. For example $^2J_{HH}$ is −10.8 Hz for CH_3Cl and −7.5 Hz for CH_2Cl_2 while it is −12.4 Hz in CH_4. In polysubstituted methanes the effect of substituents is, to a first approximation, additive as in the case of $^1J_{CH}$ for the same molecule. Nonbonding electron pairs that can be seen in the substituents containing nitrogen and oxygen atoms usually induce a positive change in $^2J_{HH}$, which is illustrated by a large positive geminal (H, H) coupling constant in formaldehyde (+41 Hz). Steric orientation of these nonbonding electron pairs relative to the orientation of C–H bonds is also known to influence the coupling constant in the case of sp^3 hybridized methylene groups. In contrast to the case of α substitution, an electronegative β substituent is known to result in a negative change in the coupling constant, while an electropositive β substituent produces opposite results. For instance, the $^2J_{HH}$ in $CHF{=}CH_2$ and $LiCH{=}CH_2$ are respectively, found to be −3.2 and +7.1 whereas that for $CH_2{=}CH_2$ is +2.5 Hz.

Thus far, we have been concerned mainly with transmission of J coupling through σ electrons; however, neighboring π electrons are also known to have a considerable influence on the magnitude of the geminal coupling constant. They are known to induce a negative change. [Note that methane has $^2J = -12.4$ Hz while N≡C–CH_2–C≡N (malononitrile) has $^2J = -20.4$ Hz.] As in the case of nonbonding electron pairs, the π bond effect is known to be a function of the angle φ between the π orbital and the C–H bond. The largest effect is observed when the neighboring π orbital and the plane of the methylene group are parallel.

The (H, H) vicinal couplings in ethane and ethylene derivatives have attracted much attention from chemists because of their importance in studying the molecular structure of these and related compounds and extensive data for these coupling constants are now accumulated and available. Both theoretical and experimental results show that the vicinal (H, H) coupling constants in H–C–C–H molecular fragments are invariably positive, in agreement with the simple Dirac vector model. In fact, the signs of other coupling constants are often determined relative to that of $^3J_{HH}$ which is always assumed to be positive. The vicinal coupling constants between H and other nuclei are also normally positive although $^3J_{FF}$ values are known to be negative.

One of the most conspicuous features in the context of $^3J_{HH}$ is its dependence on the H–C–C–H dihedral angle φ. This dependence, first theoretically predicted by Karplus[31] on the basis of VB theory and now called the Karplus equation, is usually expressed as in Equation (26):

$$^3J_{H-H} = A + B\cos\varphi + C\cos 2\varphi \quad (26)$$

where φ is the dihedral angle for this system (Figure 2). A typical set of values for the parameters A, B, and C for hydrocarbons are 7, −1, and 5 Hz, respectively, but FTP-INDO calculations give a somewhat different

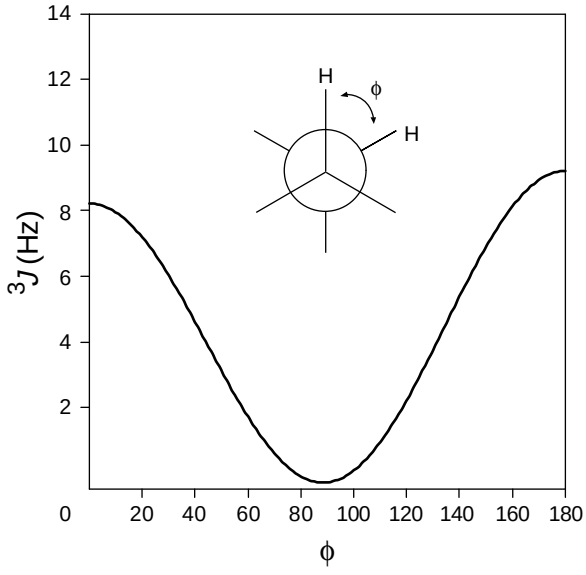

Figure 2 Dihedral angle in hydrocarbon chain and angular dependence of Karplus relation (calculated setting A, B, and C equal to 7, −1, and 5 Hz, respectively).

set of values for ethane ($A = 8.436$, $B = -2.835$, and $C = 7.523$ Hz respectively). The validity of the Karplus equation shown in Equation (26) is justified by many experimental examples. For instance, in ethylene $^3J_{HH}^{trans}$ (=19.1 Hz, $\varphi = \pi$) is greater than $^3J_{HH}^{cis}$ (=11.6 Hz, $\varphi = 0$) and $^3J_{HH}^{axial\text{-}axial}$ ($\varphi \approx \pi$) is greater than $^3J_{HH}^{axial\text{-}equatorial}$ ($\varphi \approx \pi/3$) in cyclohexane derivatives. Theory has shown that this dihedral angle dependence comes from electron spin correlation between electrons in different bonds and the most important exchange integral deciding this correlation is that between vicinal orbitals. This causes the electrons to be stable when their spins are antiparallel, thus leading to a positive vicinal coupling constant. The Karplus equation means that overlap of two vicinal orbitals is optimal when the C−H bonds are mutually cis and trans and this overlap is slightly greater for trans bonds than for cis bonds. If the molecule of interest has rapidly changing fluxional structure, only values averaged over various possible conformations are observed. As an example, we may consider a molecule of the type CH_3CHXY for which the observed $^3J_{HH}$ is $\frac{2}{3}J_g + \frac{1}{3}J_t$ where J_g and J_t are the gauche and trans coupling constants, respectively (Figure 3).

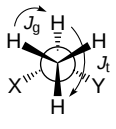

Figure 3 Gauche and trans coupling in the CH_3CHXY molecule.

The $^3J_{HH}$ values in both the saturated and unsaturated H−C−C−H fragment are also affected by the presence of substituents. One tendency is that the substitution of an electronegative atom causes a decrease in the average coupling. For quantitative purposes it is usually attempted to correlate $^3J_{HH}$ with the electronegativity change $\Delta E = E(X) - E(H)$ caused by the replacement of a hydrogen atom by a group X. Thus, for saturated and unsaturated H−C−C−H fragments the following empirical relations shown in Equations (27) and (28) have been found to be valid:[18]

$$^3J_{HH} = 9.41 - 0.80\Delta E \quad (\text{for } H_3C{-}CH_2X) \qquad (27)$$

and

$$^3J_{HH}^{trans} = 19.0 - 3.3\Delta E \quad (\text{for } H_2C{=}CHX) \qquad (28a)$$

$$^3J_{HH}^{cis} = 11.7 - 4.7\Delta E \quad (\text{for } H_2C{=}CHX) \qquad (28b)$$

The substituent effect is also known to depend on dihedral angle φ and the effect is greatest (that is, ^3J has the smallest value) when the substituent X is trans to one of the coupled protons.

The Karplus-type equations have also been proposed for a pair of nuclei other than the (H, H) pair such as (^{13}C, ^{1}H), (^{13}C, ^{13}C), and (^{31}P, ^{13}C), which have reasonable validity provided that closely similar compounds are treated, multiple-bonded atoms are excluded and the coupled nuclei do not possess lone pairs. For example, for hydrocarbons Equation (29) has been proposed:[32]

$$^3J(^{13}C, H) = 3.81 - 0.90\cos\varphi + 3.83\cos 2\varphi \qquad (29)$$

Other factors influencing ^3J are H−C−C valence angles and C−C bond length. Increase of either of the two factors or both of them is known to decrease ^3J. Although the Karplus equation is very useful for estimating dihedral angles in compounds of unknown conformation, this relation should not be indiscretely relied on because it is only valid in the absence of electronegative substituents and of distortion of tetrahedral angles at carbon. Only when values of A, B, and C are determined from model compounds closely related to the unknown can the Karplus relation be useful in a qualitative sense. Even so, it is not usually recommended to use this relation for quantitative prediction of dihedral angles.

3.3 Long-range Coupling

Couplings between two spins which are four or more bonds away from each other are usually classified as long-range couplings. The magnitude of (H, H) coupling constants in this class ranges from 0.1 to 3.0 Hz, much smaller than geminal and vicinal coupling constants which usually have values of 5–20 Hz. Nowadays, it is not difficult to measure 0.2 or smaller splittings

due to long-range couplings which are important for obtaining a wealth of information regarding structure and/or conformation of molecules.

In acyclic hydrocarbons $^4J_{HH}$ or higher coupling constants are normally too small to be measurable, but in saturated cyclic hydrocarbons these coupling constants are often found to be of considerable magnitude when two CH bonds involving coupling protons lie on a plane of zigzag carbon chain as shown in **(1)** and **(2)**. One such example can be seen in coupling over four bonds between two equatorial protons in six-membered saturated rings. In other words, long-range couplings (mediated by σ electrons) between two protons are often highly stereospecific.

(1) **(2)**

In order to understand how such a stereospecific long-range coupling arises let us consider the arrangement shown in **(1)** which is often called the ''W'' or ''M'' arrangement. Stereospecificity of long-range couplings indicates that direct spin correlation between electrons in vicinal orbitals, which we will henceforth refer to as vicinal orbital interaction, plays an important role. There may be several coupling pathways for the W arrangement. First, the C–H bonding σ orbital interacts with the neighboring geminal C–C bond orbital which then undergoes vicinal interaction with the C–H orbital at the other end, which makes a positive contribution to $^4J_{HH}$. Another coupling pathway consisting of two consecutive vicinal couplings between three C–H orbitals may also be possible, which gives a negative contribution to $^4J_{HH}$. The total coupling constant may therefore be positive or negative depending on which pathway dominates. However, considering that vicinal orbital interaction is maximum when the interacting orbitals are mutually trans, the first coupling pathway will probably weigh more in the W arrangement than the second, giving a positive sign of $^4J_{HH}$. When there are several zigzag pathways like the W arrangement discussed above, the magnitude of $^4J_{HH}$ rapidly increases as the number of pathways increases. Several such examples are shown in **(3–8)**. In saturated systems $^5J_{HH}$ is less frequently observed and if the coplanarity of the bonds involved is lost, both $^4J_{HH}$ and $^5J_{HH}$ rapidly decrease. However, if π electrons are involved in the coupling mechanism, even $^5J_{HH}$ can be appreciably large in magnitude as we will see shortly.

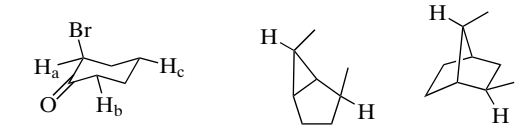

$J_{ab} \sim J_{ac} \sim J_{bc} \sim 1.1$ Hz **(3)**

$J = 0.9$ Hz **(4)**

$J = 3$–4 Hz **(5)**

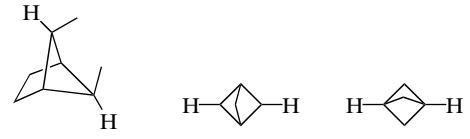

$J = 6.7$–8.1 Hz **(6)**

$J = 10$ Hz **(7)**

$J = 18$ Hz **(8)**

3.4 Coupling through π Electrons

In unsaturated compounds not only σ electrons but π electrons as well seem to be involved in the J coupling mechanism. For these compounds the observed coupling constants may be considered as the sum of the contributions from the σ and π mechanism, that is, $J^\sigma + J^\pi$. At first, this may seem puzzling because the π-electron distributions have nodes at the nuclei and, therefore, cannot contribute to J directly through contact terms. McConnell[10] has pointed out that π electrons can contribute to J through the σ–π exchange mechanism as in the case of hyperfine coupling between unpaired π electrons and nuclei in the electron spin resonance (ESR) of free radicals. Hence, the contribution to J from π electrons may be expressed in terms of the hyperfine coupling constant a as shown in Equations (30–32):[18]

$$\text{HC=CH (ethylenic)} : {}^3J^\pi \propto a(=\text{CH})\, a(=\text{CH}) \quad (30)$$

$$\text{HC=C–CH (allylic)} : {}^4J^\pi \propto a(=\text{CH})\, a(=\text{C–CH}) \quad (31)$$

$$\text{HC–C=C–CH (homoallylic)} : {}^5J^\pi \propto a(=\text{C–CH})\, a(=\text{C–CH}) \quad (32)$$

The π electron on ethylenic carbon induces negative electron spin density on the proton whence $a(=\text{CH})$ is negative. This enables us to predict from Equation (30) that $^3J^\pi$ will be positive. On the other hand, the π electron on the allylic carbon favors parallel alignment of the s-orbital electron in the H atom, making $a(=\text{C–CH})$ positive. It is known from ESR measurement that $a(=\text{C–H}) \approx -6.5 \times 10^7$ Hz and that $a(=\text{C–CH}) \approx 15 \cos^2 \varphi \times 10^7$ Hz, where φ is the angle by which the C–H bond is tilted away from the axis of the p-orbital on allylic carbon. Thus, one may expect that $^3J^\pi$ and $^5J^\pi$ would be positive while $^4J^\pi$ should be negative. Furthermore, we see that J is not affected by orientation about a double bond. Thus, cisoid allylic $^4J^\pi$ should equal transoid allylic $^4J^\pi$. Acetylenic carbon π electron also favors antiparallel alignment of the s-electron in the H atom as in the case of ethylenic coupling.

Since π-electron systems are often delocalized, spin correlation may occur for electrons at atoms separated by many chemical bonds. In other words π electrons provide a mechanism for long-range couplings. For example, $^5J^{\pi}$ in butatriene is known to be as large as ± 8.95 Hz for both cis and trans forms. In 1,3-butadiene $^4J_{HH}^{cis}(\approx -0.83\,Hz)$ is nearly equal to $^4J_{HH}^{trans}(\approx -0.86\,Hz)$ and $^5J_{HH}^{cis,cis}(\approx +0.69\,Hz)$ is very close to $^5J_{HH}^{cis,trans}(\approx +0.60\,Hz)$ which, together with the signs of these coupling constants, indicates that the π contribution is dominant. However, $^5J_{HH}^{trans,trans}(\approx +1.30\,Hz)$ has an appreciable contribution (almost 50%) from the σ mechanism due to the zigzag conformation. This should be compared with ethylenic vicinal coupling for which J^{π} is known to constitute approximately only 10% of the total coupling constant. One dramatic example that demonstrates the long-range nature of π-electron coupling is $CH_3C{\equiv}C{-}C{\equiv}C{-}C{\equiv}CH_2OH$, where the coupling constant between the methyl proton at one end of the molecule and the hydroxyl proton at the other end is known to be $+0.4$ Hz which comes entirely from the π mechanism.

3.5 Through-space Coupling

In some molecules anomalously large coupling constants are observed between two nuclei separated many bonds away but spatially in close proximity to each other. This type of coupling is thought to arise from nonbonding or van der Waals interaction of electron orbitals of two nuclei and is termed ''through-space'' coupling. It should be noted that there is some danger of confusing this term with the direct magnetic dipole–dipole interaction between two nuclei because the latter is also a through-space interaction.

Through-space coupling has been reported for spatially proximate pairs of nuclei of various types, including $^{19}F{-}^{19}F$, $^1H{-}^{19}F$, $^1H{-}^1H$, $^{13}C{-}^{19}F$, etc. In earlier days there has been some skepticism and dispute over the origin and existence of this type of coupling but numerous examples provide convincing evidence in favor of it.[33]

A semiempirical estimation of through-space (H, H) coupling is often made on the basis of Equation (33):

$$J_{AB}^{\text{through-space}} = \text{constant}\; S_A(0)^2 S_B(0)^2 S_{sAsB}^2 \qquad (33)$$

where S_{sAsB} is the overlap integral between the valence s-orbitals of atoms A and B and the constant is experimentally determined. However, Equation (33) does not seem to be suitable for estimating the through-space coupling constants involving ^{19}F for which the overlap of lone-pair p-orbitals is expected to make appreciable contributions.

Experimental studies have shown that through-space FF coupling changes very steeply with distance between the coupled fluorine nuclei. In an interesting study[33,34]

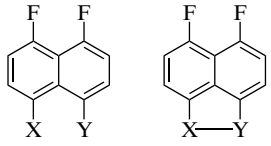

(9) **(10)**

that involves derivatives of 1,8-difluoronaphthalene **(9)** and 5,6-difluoroacenaphthalene **(10)** with various peri substituents serving to perturb the F–F distance in predictable ways it has been found that experimentally measured J_{FF} (in Hz) can be correlated with the F–F distance, d_{FF} (in nm), through Equation (34):

$$J_{FF} = (9.65 \times 10^5)\exp(-3.78 d_{FF}) \qquad (34)$$

Such an excellent correlation also supports the existence of the through-space mechanism. Thus, in general, it will be more appropriate to write down J as a sum of through-bond and through-space components, as shown in Equation (35), although it is not easy to dissect the observed coupling constant into each component:

$$J = J(\text{through-bond}) + J(\text{through-space}) \qquad (35)$$

4 MEASUREMENT OF SCALAR SPIN COUPLING CONSTANTS

A group of nuclear spins mutually coupled through scalar couplings comprises a spin system whose NMR spectrum shows a characteristic pattern of spectral splitting depending on the relative magnitude of chemical shifts and spin–spin coupling constants of involved spins. By observing and analyzing such a splitting pattern one can, in principle, extract information regarding the chemical shifts and spin–spin coupling constants for the given spin system. In practice, however, it is by no means trivial to extract all the relevant coupling parameters from overly crowded and overlapped spectra and to this end numerous ideas were contrived and applied in both one-dimensional and two-dimensional NMR spectroscopy. Both homonuclear and heteronuclear coupling constants for various spin systems have been successfully measured and correlated with electronic environments in many different kinds of molecules and the results are now tabulated in many standard NMR references. One can normally look up these tabulated data to gain information about the electronic structure of the molecules of interest with the aid of some semiempirical knowledge as explained in the previous section.

To be more specific, let us consider a spin system consisting of two coupled spins of $I = \frac{1}{2}$ in the liquid state. For this spin system the spin Hamiltonian, expressed in

frequency units, may be written as shown in Equation (36)

$$h^{-1}H = \nu_1 I_{1z} + \nu_2 I_{2z} + J\mathbf{I}_1 \cdot \mathbf{I_2} \tag{36}$$

where ν_1 and ν_2 are the resonance frequencies of the two spins 1 and 2 (with respect to the observing radiofrequency (RF) carrier frequency) and J is the scalar coupling constant.

By diagonalizing the Hamiltonian matrix formed by the basis set of product functions $|\alpha\alpha\rangle$, $|\alpha\beta\rangle$, $|\beta\alpha\rangle$, and $|\beta\beta\rangle$ we can find the eigenfunctions and eigenvalues for this two-spin system and the corresponding relative absorption line intensities can be found by evaluating $|\langle j|F^+|i\rangle|^2$ for every possible transition between two eigenstates i and j where $F^+ \equiv (I_{1x} + iI_{1y}) + (I_{2x} + iI_{2y})$. The resulting theoretical spectrum is shown in Figure 4. We see from Figure 4 that the appearance of the NMR spectrum for this spin system varies widely depending on the magnitude of $|J|$ relative to the chemical shift difference $|\nu_1 - \nu_2|$. In the limit $|J|/|\nu_1 - \nu_2| \ll 1$ each resonance line for two spins appears as a doublet centered at the respective chemical shift frequency with equal intensity and interval $|J|$. In this case the system is said to display a *first-order* spectrum and be *weakly* coupled. A weakly coupled two-spin system is conventionally denoted as AX. Why the name *first-order*? In order to understand the reason we rewrite the scalar coupling term in Equation (36) as Equation (37):

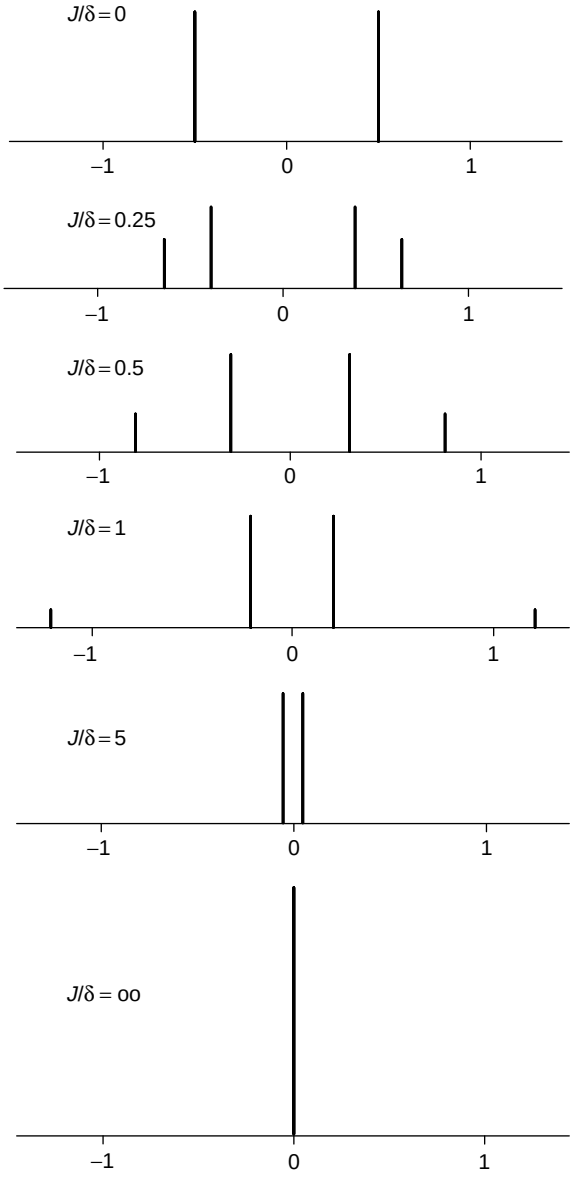

Figure 4 Stick plot for theoretical spectrum for AB spin system for various values of J/δ ($\delta \equiv |\nu_A - \nu_B|$).

$$J\mathbf{I}_1 \cdot \mathbf{I_2} = J[I_{1z}I_{2z} + \tfrac{1}{2}(I_1^+ I_2^- + I_1^- I_2^+)] \tag{37}$$

where I^+ and I^- stand, as explained in the standard textbooks on quantum mechanics, for the shifting operators for nuclear spin angular momentum $\mathbf{I}$. The first term on the right-hand side of Equation (37) gives the first-order correction to the unperturbed system consisting of two noncoupled $I = \frac{1}{2}$ spins while the second term makes the second- or higher order contributions only. In other words, if we approximate the whole scalar coupling term $J\mathbf{I}_1 \cdot \mathbf{I}_2$ by the first term $JI_{1z}I_{2z}$ only, we would obtain the spectrum corresponding to the limit $|J|/|\nu_1 - \nu_2| \ll 1$.

If $|J|$ is comparable with $|\nu_1 - \nu_2|$ in magnitude, the outer line in a doublet is always weaker than its inner line and an AB-type spectrum will result. In this case the second term in Equation (37) makes a significant contribution and the two spins are said to be *strongly* coupled. That is, when we describe a strongly coupled spin system, we have to keep the entire term in Equation (37).

Strongly coupled or not, the separation between the inner line and the outer one in a doublet is equal to $|J|$ in frequency units. Unlike the coupling constant, however, in the case of an AB-type spectrum the chemical shift difference must not be read off as the separation between the respective midpoints of two doublets as is usually done in the case of an AX spectrum. Instead, it must be evaluated from the fact that Δ, the separation between the outer line of one spin and the inner line of the other, and J are related through the relation shown in Equation (38):

$$\Delta = \sqrt{(\nu_1 - \nu_2)^2 + J^2} \tag{38}$$

However, in the AB and/or AX system the sign of $\nu_1 - \nu_2$ and J cannot be determined from the observed spectrum alone because its NMR spectrum is symmetrical with respect to the change of signs of these parameters. To determine their signs we need an extra source of information.

When the chemical shifts of the two spins 1 and 2 are exactly equal, we see from Figure 4 that splitting

disappears. It is generally true even for a multinuclear system that, if all the magnetic nuclei present have the same chemical shift, scalar couplings are inoperative in causing splitting.

In a heteronuclear two-spin system where two nuclei are of different kinds, as in the case of ^{1}H and ^{13}C, the chemical shift difference is usually very large compared with $|J|$ and the system can safely be regarded as an AX system. In a homonuclear spin system there is no such guarantee. Even in this case, however, we can speak of an AX system in the presence of strong static field B_0 because the chemical shift difference, expressed in frequency units, increases with increasing field strength and at very high field it can dominate $|J|$ which is influenced very little by B_0.

Another interesting spin system relevant to the scalar spin coupling is an A_2B system which consists of two magnetically equivalent spins denoted by A and a third spin, denoted by B, strongly coupled to them. (Two spins are said to be *magnetically* equivalent if they have the same chemical shifts *and* their coupling constants with all other magnetic nuclei present in the spin system are equal.) The spin Hamiltonian for this system may be written as shown in Equation (39):

$$h^{-1}H = \nu_A I_{A1z} + \nu_A I_{A2z} + J_{A1A2}\mathbf{I_{A1}} \cdot \mathbf{I_{A2}} + J_{AB}(\mathbf{I}_{A1} \cdot \mathbf{I}_B + \mathbf{I}_{A2} \cdot \mathbf{I}_B) \quad (39)$$

where we have noted that, since the two spins A_1 and A_2 are magnetically equivalent, $\nu_{A1} = \nu_{A2} \equiv \nu_A$ and $J_{A1B} = J_{A2B} \equiv J_{AB}$. The corresponding Hamiltonian matrix can be factored to four 1×1 and two 2×2 matrices if we choose a basis set complying with the fact that H is invariant with respect to the exchange of A_1 and A_2 and commutes with the operator $I_{A1z} + I_{A2z} + I_{Bz}$. Theoretically simulated spectra for this spin system are shown in Figure 5 for a set of $|\nu_A - \nu_B|/|J_{AB}|$ values. From this spectrum $\nu_A - \nu_B$ and $|J_{AB}|$ can be determined from the observed line positions but the sign of J_{AB} still remains unknown. Note in this case, too, that the scalar spin coupling between two magnetically equivalent nuclei A_1 and A_2 has no bearing on the appearance of spectra, whence no information on the J coupling between magnetically equivalent spins can be obtained. It is generally true that the scalar couplings among *magnetically* equivalent nuclei cannot result in any spectral change for larger spin systems, either, because the sum of magnetically equivalent spin operators $_k\mathbf{I}_k$ commutes with the sum of the scalar coupling terms among these spins $_{k>l}J_{kl}\mathbf{I}_k \cdot \mathbf{I}_l$.

The next simplest and frequently encountered spin system is an ABX system which consists of two strongly coupled nuclei and a third nucleus which is coupled weakly to the former. For this system the spin Hamiltonian may be written as shown in Equation (40):

$$h^{-1}H = \nu_A I_{Az} + \nu_B I_{Bz} + \nu_X I_{Xz} + J_{AB}\mathbf{I_A} \cdot \mathbf{I_B} + J_{AX} I_{Az} I_{Xz} + J_{BX} I_{Bz} I_{Xz} \quad (40)$$

If we form the energy matrix by employing the basis set of product functions $|\alpha_A\alpha_B\alpha_X\rangle$, $|\alpha_A\alpha_B\beta_X\rangle$, $|\beta_A\alpha_B\alpha_X\rangle$, etc., the matrix elements between two different spin states of X will be vanishing. Therefore, we may treat the AB part of the ABX spectrum as a superposition of two AB spectra centered at different positions,

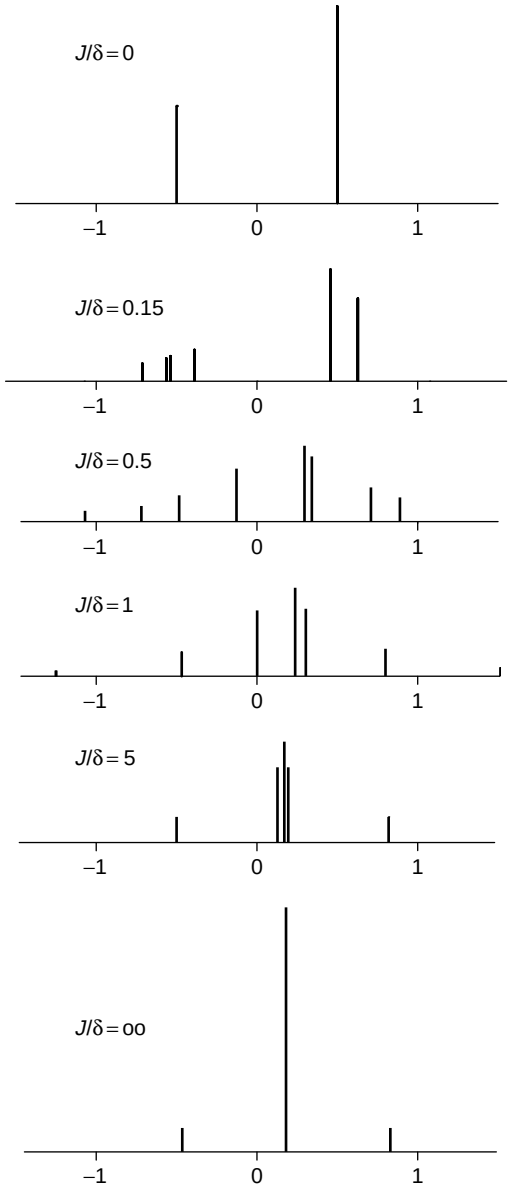

Figure 5 Stick plot for theoretical spectrum for A_2B spin system for various values of J/δ ($\delta \equiv |\nu_A - \nu_B|$).

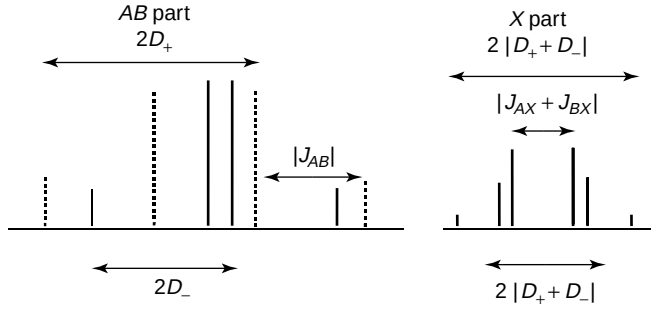

Figure 6 Structure of ABX-type spectrum where $D_{\pm}$ are defined by $2D_{\pm} = \sqrt{\left\{(\nu_A - \nu_B) \pm \frac{1}{2}(J_{AX} - J_{BX})\right\}^2 + J_{AB}^2}$.

one at $\frac{1}{2}[(\nu_A + \nu_B) + \frac{1}{2}(J_{AX} + J_{BX})]$ and the other at $\frac{1}{2}[(\nu_A + \nu_B) - \frac{1}{2}(J_{AX} + J_{BX})]$ as shown in Figure 6. The X region may be shown to consist of four lines originating from transitions due to X spin plus two weak combination lines. Interpretation of an ABX-type spectrum normally starts with identifying two sets of AB quartets, which will be denoted by $(ab)_+$ and $(ab)_-$, respectively. Once this is done, we can easily extract $|J_{AB}|$. However, as in the case of an AB spin system, only the information on the absolute value of J_{AB} can be obtained since the AB region of the spectrum does not change upon changing the sign of J_{AB}. Also, by measuring the distance between the center of two quartets $(ab)_+$ and $(ab)_-$ we know $\frac{1}{2}|J_{AX} + J_{BX}|$ which is also equal to half of the distance between the two strongest lines in the X region. The simultaneous change of sign of two coupling constants J_{AX} and J_{BX} leaves $|J_{AX} + J_{BX}|$, hence the entire spectrum itself, unchanged, which means we have no way of knowing whether $J_{AX} + J_{BX}$ is positive or negative. However, changing the sign of only one of these two coupling constants results in a change of spectral appearance, so we can gain information on their relative signs. It is by no means easy to interpret visually the ABX spectrum to extract the information on ν_A, ν_B, ν_X, J_{AB}, J_{AX}, and J_{BX}, although the routine procedure for this is described in many standard textbooks on NMR.

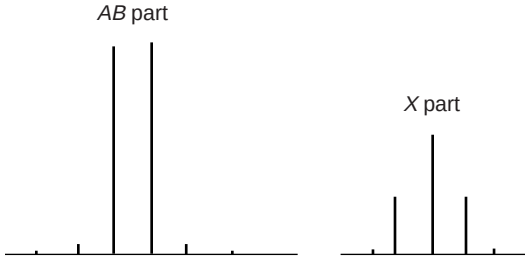

Figure 7 Stick plot for theoretical ABX spectrum when $\nu_A = \nu_B$.

Depending on the signs and magnitudes of J_{AX} and J_{BX}, the spectral feature varies so widely that sometimes one is easily misled to believe that it arises from a different spin system other than ABX. One such example may be seen in the case when $J_{AX} = 0, J_{BX} \neq 0$ with A and B strongly coupled. In the first-order interpretation one may expect that the X region consists merely of a doublet arising from coupling with B spin. In the actual ABX spectra the X spin signal usually shows additional splittings due to the presence of *A to which it is not actually coupled*. Musher and Corey[35] referred to this ghostly coupling as *virtual coupling*. However, this concept was introduced merely for convenience of understanding the spectra in the first-order approximation and should not be regarded as real. Another baffling example of ABX spectra is the case where $\nu_A \approx \nu_B$. As ν_B approaches ν_A, the spectral appearance changes drastically and when $\nu_A = \nu_B$ the AB region appears to be simply a doublet while the X region is a 1 : 2 : 1 triplet, as shown in Figure 7. Although there can be other weak lines as shown in Figure 7, they are normally lost in noise. This *deceptively simple* spectrum may mistakenly be interpreted as that of an A_2X spin system with only one J_{AX} coupling constant present if one does not realize that this is a special case of ABX (actually $AA'X$, since the chemical shifts of A and B are equal. In this case the two spins A and B are not *magnetically equivalent* although their chemical shifts are the same, hence being written as AA' instead of A_2). The apparent coupling constants which can be read off from splittings seen in the doublet or triplet are actually an average of two coupling constants J_{AX} and J_{BX}. For example, the proton resonance spectrum of 2,5-dichloronitrobenzene taken at 60 MHz[36] actually consists of a doublet and a triplet which one may be tempted to interpret as arising from an AX_2 spin system with $J_{HH}(\text{meta}) = J_{HH}(\text{para}) = 1.6$ Hz. However, it is found that, if we take the spectrum of the same sample at, say, 200 MHz, a typical ABX-type spectrum results, giving $J_{HH}(\text{meta}) \approx 3$ Hz and $J_{HH}(\text{para}) \approx 0$ Hz. Such *deceptively simple* spectra are widespread and not limited to ABX systems. We can easily beat these deceptively simple spectra by recording them on a spectrometer operating at higher fields because the splitting pattern of a true A_2X will not depend on the strength of the applied magnetic field.

The easiest case for a three-spin system is to interpret the AMX-type spectra. In this case the NMR absorption line for a spin, say A, is split first by M, interspaced by J_{AM}, and then further by the second coupling partner X, interspaced by J_{AX}, all with equal intensities. In particular, if all the spins involved have $I = \frac{1}{2}$ and the two coupling constants J_{AM} and J_{AX} are equal, that is, in the case of the AX_2 spin system, two of the split lines will overlap, giving the intensity ratio 1 : 2 : 1. This trend may be extended

to a multispin system. For example, in the case of the AX_3 spin system the A spectrum will consist of four lines interspaced by J_{AX} with the intensity ratio 1 : 3 : 3 : 1. Most generally, in the case of AX_n we will have the A spectrum consisting of $n+1$ lines with the intensity ratio of $\binom{n}{0} : \binom{n}{1} : \binom{n}{2} : \ldots : \binom{n}{n-1} : \binom{n}{n}$.

A most general ABC-type spectrum for a strongly coupled three-spin system cannot be treated exactly, but with the aid of a computer it can readily be interpreted numerically, yielding information about the magnitudes of three coupling constants J_{AB}, J_{AC}, and J_{AB}. For a higher spin system it is in general impossible to treat the problem analytically except for a few cases. However, numerical solutions even for the complicated multispin systems are available and have been implemented on modern NMR spectrometers. If all the lines for a given spin system are observed with good resolution, it is not difficult to extract information on the chemical shift of each constituent spin as well as the internuclear spin coupling constants, except for relative signs of these parameters, by analyzing the observed spectra as described above. In reality, however, for large molecules with many magnetic nuclei the one-dimensional NMR spectra are usually so crowded with many lines arising from nuclei of different chemical shifts as well as scalar spin couplings between them (and sometimes with considerable numbers of lines overlapped and weaker lines buried under the stronger ones) over the observed range of frequency that it is by no means easy to interpret and assign the spectral lines following the standard procedure described above. Moreover, in some cases it may not be clear which spin is scalar-coupled to which. To help ease such difficulties, in one-dimensional NMR one may employ double resonance techniques such as spin decoupling, spin tickling, internuclear double resonance (INDOR), pseudo-INDOR, selective population transfer, etc. Among these, spin decoupling is the most frequently used technique for reducing the spectral complexities in which one irradiates a specific nucleus with the second RF field at its resonance frequency. Upon irradiation the multiplet structures arising from coupling with the irradiated nucleus collapse, which results in considerable reduction of the complexity of the observed spectra. However, in this case the information on the coupling constant between the spin of interest and that irradiated is also lost. The procedure of irradiating the spins one by one employed in one-dimensional NMR is usually very cumbersome (and sometimes not possible), in which case the two-dimensional techniques such as correlation spectroscopy (COSY), two-dimensional J-resolved spectroscopy, etc. are more advantageous.

When the couplings between two different kinds of nuclei, such as ^{1}H and ^{13}C (that is, heteronuclear couplings), become a nuisance, one may employ the broadband decoupling technique in which the entire chemical shift range for one kind of coupling spin is irradiated with sufficient RF power. In this case the heteronuclear coupling information is completely lost, but homonuclear couplings can survive to give visible splitting as in the case of incredible natural abundance double quantum transfer experiment (INADEQUATE) spectra of ^{13}C–^{13}C coupling obtained under broadband ^{1}H decoupling. When the natural abundance of a coupling partner is so low that splitting caused by scalar spin coupling is not readily observed, one may employ such intensity enhancement techniques as insensitive nuclei enhanced by polarization transfer (INEPT). These and other techniques (β-COSY, E-COSY, etc.) may also be used to find the relative signs of various scalar coupling constants.

For some reasons (chemical exchange, quadrupolar relaxation, etc.) the scalar spin coupling expressed in the form of Equation (1) can be modulated randomly in time. In the case of chemical exchange one of the coupled spins may jump over different sites making the coupling constant J modulated. This kind of modulation can cause the scalar relaxation of the first kind. On the other hand, when one of the two coupled nuclei has its own strong relaxation mechanism such as quadrupolar relaxation, the scalar coupling is responsible for the scalar relaxation of the second kind. When one of the coupled spins, say S, relaxes fast (that is, its spin–lattice relaxation time is short compared with $1/J$), we can observe only a single line for the other spin, say I, instead of the expected multiplet. In this case we cannot measure the coupling constant J between two spins I and S. However, even in this case by making use of the relaxation data for two coupled spins we can estimate J.[37]

The author feels greatly indebted to all the authors of the original papers and/or literature cited in the present article.

ABBREVIATIONS AND ACRONYMS

AEE	Average Excitation Energy
COSY	Correlation Spectroscopy
ESR	Electron Spin Resonance
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
INDOR	Internuclear Double Resonance
INEPT	Insensitive Nuclei Enhanced By Polarization Transfer
MO	Molecular Orbital

NMR	Nuclear Magnetic Resonance
RF	Radiofrequency
VB	Valence Bond

RELATED ARTICLES

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Parameters, Calculation of Nuclear Magnetic Resonance • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. W.G. Proctor, F.C. Yu, 'On the Nuclear Magnetic Moments of Several Stable Isotopes', *Phys. Rev.*, **81**, 20–30 (1951).
2. H.S. Gutowsky, D.W. McCall, 'Nuclear Magnetic Resonance Fine Structure in Liquids', *Phys. Rev.*, **82**, 748–749 (1951).
3. E.L. Hahn, D.E. Maxwell, 'Chemical Shift and Field Independent Frequency Modulation of the Spin Echo Envelope', *Phys. Rev.*, **84**, 1246–1247 (1951).
4. N.F. Ramsey, 'Electron Coupled Interactions between Nuclear Spins in Molecules', *Phys. Rev.*, **91**, 303–307 (1953).
5. R.E. Wasylishen, 'Dipolar and Indirect Coupling Tensors in Solids', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 1685–1695, Vol. 3, 1996.
6. N.F. Ramsey, E.M. Purcell, 'Interactions between Nuclear Spins in Molecules', *Phys. Rev.*, **85**, 143–144 (1952).
7. P. Pyykkö, 'Relativistic Theory of Nuclear Spin–Spin Coupling in Molecules', *Chem. Phys.*, **22**, 289–296 (1977).
8. H.S. Gutowsky, D.W. McCall, C.P. Slichter, 'Nuclear Magnetic Resonance Multiplets in Liquids', *J. Chem. Phys.*, **21**, 279–292 (1953).
9. H.M. James, A.S. Coolidge, 'The Ground State of the Hydrogen Molecule', *J. Chem. Phys.*, **1**, 825–835 (1933).
10. H.M. McConnell, 'Molecular Orbital Approximation to Electron Coupled Interaction between Nuclear Spins', *J. Chem. Phys.*, **24**, 460–467 (1956).
11. M. Karplus, D.H. Anderson, T.C. Farrar, H.S. Gutowsky, 'Valence-bond Interpretation of Electron Coupled Proton–Proton Magnetic Interactions Measured via Deuterium Substitution', *J. Chem. Phys.*, **27**, 597–598 (1957).
12. M. Karplus, D.H. Anderson, 'Valence-bond Interpretation of Electron-coupled Nuclear Spin Interactions; Application to Methane', *J. Chem. Phys.*, **30**, 6–10 (1959).
13. J.A. Pople, D.P. Santry, 'Molecular Orbital Theory of Spin Coupling Constants', *Mol. Phys.*, **8**, 1–18 (1964).
14. J.A. Pople, J.W. McIver, N.S. Ostlund, 'Self-consistent Perturbation Theory. I. Finite Perturbation Methods', *J. Chem. Phys.*, **49**, 2960–2964 (1968).
15. M. Barfield, 'Extension of the Valence-bond Description of Nuclear Spin–Spin Coupling', *J. Chem. Phys.*, **46**, 811–812 (1967); 'Valence-bond Studies of Contact Nuclear Spin–Spin Coupling. I. A Truncated Matrix Sum Method', *J. Chem. Phys.*, **48**, 4458–4462 (1968).
16. R.K. Harris, *Nuclear Magnetic Resonance Spectroscopy*, Longman/Wiley, New York, 1986.
17. F.A. Bovey, L. Jelinski, P.A. Mirau, *Nuclear Magnetic Resonance Spectroscopy*, Academic Press, San Diego, 1988.
18. H. Günther, *NMR Spectroscopy*, 2nd edition, John Wiley & Sons, Chichester, 1995.
19. L.M. Jackman, S. Sternhell, *Applications of Nuclear Magnetic Resonance Spectroscopy in Organic Chemistry*, 2nd edition, Pergamon Press, Oxford, 1969.
20. I. Ando, G.A. Webb, *Theory of NMR Parameters*, Academic Press, London, 1983.
21. L.W. Reeves, 'Absolute Correlation of Nuclear Spin–Spin Coupling Constants with Atomic Number. II. Couplings J_{X-C-H} and J_{X-H}', *J. Chem. Phys.*, **40**, 2128–2131 (1964); 'Absolute Correlation of Nuclear Spin–Spin Coupling Constants with Atomic Number. III. J_{X-H} in Isoelectronic Series and the Lighter Elements', *J. Chem. Phys.*, **40**, 2132–2134 (1964).
22. P.T. Inglefield, L.W. Reeves, 'Correlation of Nuclear Spind–Spin Coupling Constants with Atomic Number. V. Deviation Parameters for Substituted Methanes and Silanes', *J. Chem. Phys.*, **40**, 2424–2425 (1964); 'Correlation of Nuclear Spin–Spin Coupling Constants with Atomic Number. VI. Exceptions to the Rule', *J. Chem. Phys.*, **40**, 2425–2426 (1968).
23. J. Feeny, R. Haque, L.W. Reeves, C.P. Yue, 'Nuclear Spin–Spin Coupling Constants; Equilibrium and Kinetic Studies for Fluoroberyllate Complexes in Solution', *Can. J. Chem.*, **46**, 1389–1398 (1968).
24. J.M. Schulman, T. Venanzi, 'The Relationship between $^{15}N^{13}C$ Coupling Constants and Hybridization', *J. Am. Chem. Soc.*, **98**, 6739–6741 (1976).
25. N. Müller, D.E. Pritchard, 'C^{13} Splittings in Proton Magnetic Resonance Spectra. I. Hydrocarbons', *J. Chem. Phys.*, **31**, 768–771 (1959).

26. R.E. Carhart, J.D. Roberts, 'Nuclear Magnetic Resonance Spectroscopy Analysis of the ^{1}H and ^{13}C Spectra of $Br^{13}CH_2$ $^{13}CH_2Br$', *Org. Magn. Res.*, **3**, 139–141 (1971).
27. F.H. Weigert, J.D. Roberts, 'Nuclear Magnetic Resonance Spectroscopy. Carbon–Carbon Coupling', *J. Am. Chem. Soc.*, **94**, 6021–6025 (1972).
28. M.D. Newton, J.N. Shulman, 'Theoretical Studies of Bicyclobutane', *J. Am. Chem. Soc.*, **94**, 767–773 (1972).
29. M.D. Newton, J.M. Shulman, M.M. Manus, 'Theoretical Studies of Benzene and Its Valence Isomers', *J. Am. Chem. Soc.*, **96**, 17–23 (1974).
30. J.M. Shulman, M.D. Newton, 'Contributions to the Nuclear Spin–Spin Coupling Constants of Directly Bonded Carbons', *J. Am. Chem. Soc.*, **96**, 6295–6297 (1974).
31. M. Karplus, 'Contact Electron-spin Coupling of Nuclear Magnetic Moments', *J. Chem. Phys.*, **30**, 11–15 (1959).
32. R. Aydin, J.-P. Loux, H. Günther, 'Karplus Curve for $^{3}J(^{13}C, ^{1}H)$ in Hydrocarbons', *Angew. Chem., Int. Ed. Engl.*, **21**, 449 (1982).
33. F.B. Mallory, 'Coupling through Space in Organic Chemistry', in *Encyclopedia of Nuclear Magnetic Resonance*, eds. D.M. Grant, R.K. Harris, John Wiley & Sons, Chichester, 1491–1500, Vol. 3, 1996.
34. F.B. Mallory, C.W. Mallory, M.-C. Fedarko, 'Substituent Effects on Through-space $^{19}F–^{19}F$ Coupling in the 1,8-Difluoronaphthalene System', *J. Am. Chem. Soc.*, **96**, 3536–3542 (1974).
35. J.I. Musher, E.J. Corey, 'Virtual Long-range Spin–Spin Couplings in NMR. The Linear 3-Spin System and Qualitative Implications of Higher Systems', *Tetrahedron*, **18**, 791–809 (1962).
36. E.D. Becker, *High Resolution NMR*, Academic Press, New York, 162, 1969.
37. T.C. Farrar, E.D. Becker, *Pulse and Fourier Transform NMR*, Academic Press, New York, 60–63, 1971.

Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton

Gang Wu
Queen's University, Kingston, Canada

1 Introduction **1**
2 Basic Principles and Techniques **2**
2.1 Nuclear Spin Hamiltonians 2
2.2 Experimental Solid-state Nuclear Magnetic Resonance Techniques 7
3 Solid-state Nuclear Magnetic Resonance of Main Group Elements **9**
3.1 Nitrogen-15 9
3.2 Silicon-29 10
3.3 Phosphorus-31 10
3.4 Selenium-77 11
4 Solid-state Metal Nuclear Magnetic Resonance **12**
4.1 Cadmium-113 and Mercury-199 12
4.2 Tin-119 13
4.3 Platinum-195 and Lead-207 14
4.4 Low-γ Nuclei: Iron-57, Yttrium-89, Silver-109 and Tungsten-183 14
5 Second-order Quadrupolar Effects on Spin-$\frac{1}{2}$ Spectra **15**
6 Concluding Remarks **16**
Acknowledgments **16**
Abbreviations and Acronyms **16**
Related Articles **16**
References **16**

Solid-state nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical technique with a wide range of applications in chemistry, biochemistry and materials science. Solid-state NMR is amenable to studies of molecular systems that are not suitable either for liquid-state NMR because of insolubility or for single-crystal diffraction techniques because of poor crystallinity. Therefore, solid-state NMR provides a natural connection between liquid-state NMR and single-crystal diffraction techniques. Furthermore, solid-state NMR is the best way of studying the anisotropic nature of nuclear magnetic properties, thus potentially yielding more complete information about molecular structure and chemical bonding. This article provides an overview of the fundamental principles of solid-state NMR with selected examples of chemical applications. Emphasis is placed on the fundamental information and practical aspects of solid-state multinuclear NMR experiments for the following spin-$\frac{1}{2}$ nuclei: ^{15}N, ^{29}Si, ^{31}P, ^{77}Se, ^{113}Cd, ^{199}Hg, ^{117}Sn, ^{195}Pt, ^{207}Pb, ^{57}Fe, ^{89}Y, ^{109}Ag and ^{183}W. A brief introduction to the second-order quadrupolar effect on spin-$\frac{1}{2}$ NMR spectra is also provided.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Solid-state NMR spectroscopy is a powerful analytical technique with a wide range of applications in chemistry, biochemistry and materials science. Over the past 20 years, the advances in both NMR methodology and instrumentation have made equally important impacts on the development of the field. Today, it is generally accepted that solid-state NMR is applicable to a great variety of chemical and biological systems. One unique feature of solid-state NMR is that the method provides a natural connection between liquid-state NMR and diffraction techniques. The rapid expansion in chemical applications of solid-state NMR has been brought about for several reasons. First, the technological improvement of NMR spectrometers in the past 20 years has made solid-state NMR experiments routine. Second, there is potentially more information in solid-state NMR spectra. Third, NMR parameters such as the chemical shift tensor and the dipolar coupling constant are influenced only by the short-range environment around the nucleus of interest. Therefore, NMR is amenable to studies of chemical systems that cannot be obtained in crystalline form, e.g. semicrystalline and amorphous polymers. For these systems, single-crystal diffraction techniques are difficult to apply. Finally, many chemical systems such as cross-linked polymers and ceramics are insoluble in common solvents and hence are not suitable for solution NMR study.

In this article, we describe the basic principles of solid-state NMR spectroscopy and some specific applications in chemistry. We focus on NMR studies of spin-$\frac{1}{2}$ nuclei other than ^{13}C and ^{1}H. This article is not intended to be a literature review, which is neither necessary nor possible within the limited space. Instead, the article will provide readers with an overview of the fundamental principles of solid-state NMR and some selected examples of applications. Emphasis will be placed on the basic information about solid-state multinuclear NMR experiments and discussions about specific applications will be kept brief. For more detailed information about solid-state NMR, readers are referred to several excellent monographs.[1–10]

Table 1 Nuclear properties and reference samples for selected spin-$\frac{1}{2}$ nuclei

Isotope	Natural abundance (%)	Magnetogyric ratio	NMR frequency at 11.75 T MHz	Reference (solution)	Set-up sample (solid)
^{1}H	99.985	26.7519	500.130000	TMS	–
^{15}N	0.37	−2.712	50.697	$MeNO_2$	$^{15}NH_4NO_3$ −359 ppm
^{29}Si	4.7	−5.3188	99.362	TMS	Q_8M_8 11.5 ppm (M band)
^{31}P	100	10.841	202.456	85% H_3PO_4	$NH_4H_2PO_4$ 0.81 ppm
^{57}Fe	2.19	0.8661	16.194	$Fe(CO)_5$	–
^{77}Se	7.58	5.12	95.382	Me_2Se	$(NH_4)_2SeO_4$ 1040.2 ppm
^{89}Y	100	−1.3155	24.591	1 M aq. YCl_3	$Y(NO_3)_3{\cdot}6H_2O$ −53.2 ppm
^{109}Ag	48.18	−1.250	23.274	$AgNO_3$(aq.)	Silver lactate 345.9, 320.2 ppm
^{113}Cd	12.26	−5.9550	110.995	0.1 M $Cd(ClO_4)_2$	$Cd(NO_3)_2{\cdot}4H_2O$ −100 ppm
^{119}Sn	8.58	−10.021	186.502	5% Me_4Sn in CH_2Cl_2	$(C_6H_{11})_4Sn$ −97.35 ppm
^{183}W	14.40	1.120	20.814	Na_2WO_4 (aq.)	$W(CO)_6$ −3470 ppm
^{195}Pt	33.8	5.768	107.100	K_2PtCl_6	$(NH_4)_2PtCl_6$ 240 ppm
^{199}Hg	16.84	4.8154	89.577	Me_2Hg	$[Hg(DMSO)_6](O_3SCF_3)_2$ −2313 ppm $NEt_4NaHg(CN)_4$ −434 ppm
^{207}Pb	22.6	5.540	104.630	(a) 85% Me_4Pb in toluene (b) 1.0 M $Pb(NO_3)_2$ (aq.) $\delta = -2961$ ppm	$(p\text{-tolyl})_4Pb$ −148.8 ppm

Q_8M_8, cubic octamer silicic acid trimethylsilyl ester; TMS, tetramethylsilane.

Research in multinuclear solid-state NMR spectroscopy is also periodically reviewed in the *Specialist Periodical Reports on NMR* published by the Royal Society of Chemistry, *Annual Reports on NMR Spectroscopy, NMR: Basic Principles and Progress, Progress in NMR Spectroscopy, Magnetic Resonance Review* and *Advances in Magnetic Resonance*. Another source of general information about NMR is the recently published *Encyclopedia of Nuclear Magnetic Resonance*.[11] The large eight volumes cover almost all aspects of NMR and contain many excellent articles about solid-state NMR applications and historical accounts of technique development. Finally, the review article by Sebald,[12] covering the literature up to 1993, also contains useful technical details and chemical applications of solid-state NMR of less common spin-$\frac{1}{2}$ nuclei.

Table 1 gives nuclear properties and reference samples for selected spin-$\frac{1}{2}$ nuclei.

2 BASIC PRINCIPLES AND TECHNIQUES

2.1 Nuclear Spin Hamiltonians

NMR spectroscopy, like all other spectroscopic techniques, is concerned with the study of a particular set of energy levels via the interaction between electromagnetic radiation and matter. More precisely, NMR spectroscopy deals with energy levels involving nuclear spins in the presence of a strong, static magnetic field. The electromagnetic radiation employed in NMR lies in the range of radiofrequency (RF), hence NMR spectroscopy is historically known as RF spectroscopy.

Generally, one can treat the nuclear spin system as an isolated system. That is, it is only necessary to consider a reduced-spin Hamiltonian rather than a total Hamiltonian describing the entire molecular system. Interactions of the nuclear spin system under observation with all other degrees of freedom including all possible time-dependent random interactions between the nuclear spins and environment are ascribed to the ''lattice''. Interactions between nuclear spins and the lattice are important in understanding various NMR relaxation phenomena, but this article is not concerned with NMR relaxation phenomena.

For nuclear spin systems containing only spin-$\frac{1}{2}$ nuclei, the following four nuclear spin interactions have to be considered for diamagnetic solids: (i) the nuclear Zeeman interaction, (ii) the chemical shielding interaction, (iii) the direct magnetic dipole–dipole interaction and (iv) the indirect electron-mediated spin–spin interaction or the J interaction.

2.1.1 The Nuclear Zeeman Interaction

The basic and strongest interaction in NMR is the interaction between a nuclear magnetic dipole moment, $\boldsymbol{\mu}$, and an external magnetic field, $\boldsymbol{B}_0$, namely the nuclear Zeeman interaction. A nuclear magnetic dipole moment is the intrinsic property of a nucleus and is related to the total nuclear spin angular momentum, $\hbar\boldsymbol{I}$, by

Equation (1):

$$\boldsymbol{\mu} = \gamma \left(\frac{h}{2\pi} \right) \boldsymbol{I} \tag{1}$$

where γ is the magnetogyric ratio and h is Planck's constant. The magnitude of the nuclear spin angular momentum is given by $h[I(I+1)]^{\frac{1}{2}}$, where I is either an integer or half-integer. If the direction of the external magnetic field is defined as the z direction, the nuclear Zeeman Hamiltonian can be written as Equation (2):

$$h^{-1}H_z = -\boldsymbol{\mu} \cdot \boldsymbol{B}_0 = -\left(\frac{\gamma}{2\pi} \right) \boldsymbol{I} \cdot \boldsymbol{B}_0 = -\left(\frac{\gamma}{2\pi} \right) B_0 m_Z \tag{2}$$

where m_Z is the z component of the nuclear spin angular momentum ($m_Z = I, I-1, \ldots, -I$). For a nucleus with $I = \frac{1}{2}$, m_Z takes values of $+\frac{1}{2}$ and $-\frac{1}{2}$, leading to a two-level system. The two spin states are usually denoted $|\alpha\rangle (m_Z = +\frac{1}{2})$ and $|\beta\rangle (m_Z = -\frac{1}{2})$, or as the ''up'' and ''down'' spin states, respectively. The energy separation between the two nuclear spin states is given by Equation (3):

$$\Delta E = \gamma \left(\frac{h}{2\pi} \right) B_0 \tag{3}$$

In the presence of electromagnetic radiation, transitions between the two nuclear spin states may be induced if the quanta of applied radiation satisfy the condition in Equation (4):

$$\nu_0 = \left(\frac{\gamma}{2\pi} \right) B_0 \tag{4}$$

where ν_0 is known as the Larmor frequency in units of s^{-1} (hertz or Hz). The magnitude of the nuclear Zeeman interaction is usually expressed by the Larmor frequency. It is clear from Equation (4) that the nuclear Zeeman interaction depends on the magnetogyric ratio of the nuclear species under observation and the strength of the applied magnetic field, $\boldsymbol{B}_0$. For protons, typical nuclear Zeeman interactions in the magnetic field of modern NMR spectrometers ($B_0 = 2.35$–17.62 T) are in the range 100–750 MHz. The nuclear Zeeman interaction is extremely weak compared with interactions studied by other spectroscopic techniques. For example, the nuclear Zeeman interaction energy of protons in a magnetic field of 14.10 T corresponds to a thermal energy of 28.8 mK.

In a macroscopic sample, one deals with a large number of nuclear spins (10^{18}–10^{23}). At thermal equilibrium the population of nuclear spins at $|\alpha\rangle$ and $|\beta\rangle$ states is given by the Boltzmann distribution (Equation 5):

$$\frac{n_\beta}{n_\alpha} = \exp\left(\frac{-\Delta E}{kT} \right) = \exp\left(\frac{-\gamma(h/2\pi)B_0}{kT} \right) \tag{5}$$

Since the nuclear Zeeman interaction is extremely weak, it results in an extremely small population difference between $|\alpha\rangle$ and $|\beta\rangle$ states. For instance, $n_\beta/n_\alpha = 0.99996$ for ^{1}H at 300 K in a magnetic field of 14.10 T. This makes NMR spectroscopy a rather *insensitive* spectroscopic technique in the sense that a relatively large number of nuclear spins are required in order to make NMR signals detectable. However, the extremely weak nuclear spin interactions make it possible for atomic nuclei to become a microscopic probe extremely sensitive to subtle variations in molecular and electronic structures.

It is clear from Equation (2) that the nuclear Zeeman interaction is of no explicit interest for chemists, since all nuclei of a given isotope would absorb RF radiations at the same frequency. If the nuclear Zeeman interaction were the only nuclear spin interaction present, NMR would be of no use to chemists. However, the nuclear Zeeman interaction is analogous to the carrier wave in radio broadcasting, without which all beautiful ''music'' arising from other subtle nuclear spin interactions could not be enjoyed.

2.1.2 The Chemical Shielding Interaction

In the above discussion of nuclear Zeeman interactions, one in fact deals with a bare nucleus; however, chemists are most interested in molecules. This implies that one must consider nuclei surrounded by electrons. Therefore, interactions among nuclei, electrons and an external magnetic field become important. It is well known that electron circulating motions induced by an external magnetic field will generate a secondary magnetic field at the nucleus. Thus, a nucleus will generally experience a magnetic field slightly different from the applied field. This phenomenon is known as chemical shielding. The effective magnetic field at the nucleus can be written as Equation (6):

$$B_{\text{eff}} = (1 - \sigma)B_0 \tag{6}$$

where σ is a small, dimensionless quantity ($\sigma \ll 1$) called the chemical shielding constant.

It can be readily appreciated that, if the electronic environment at the nucleus deviates from spherical symmetry, the chemical shielding at the nucleus will depend on the orientation of the molecule with respect to the external magnetic field. In general, the chemical shielding property can be described by a second-rank tensor and the chemical shielding Hamiltonian can be written as Equation (7):

$$\begin{aligned} h^{-1}H_{\text{CS}} &= \boldsymbol{\mu} \cdot \sigma \cdot \boldsymbol{B}_0 \\ &= \left(\frac{\gamma}{2\pi} \right) (I_x, I_y, I_z) \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ B_0 \end{pmatrix} \\ &= \left(\frac{\gamma}{2\pi} \right) (I_x, I_y, I_z) \begin{pmatrix} \sigma_{xz} \\ \sigma_{yz} \\ \sigma_{zz} \end{pmatrix} B_0 \end{aligned} \tag{7}$$

It is clear from Equation (7) that the chemical shielding introduces an additional magnetic field, $(\sigma_{xz}, \sigma_{yz}, \sigma_{zz})B_0$. Since $\sigma \ll 1$, the effective field that a nucleus feels is still approximately along the $\boldsymbol{B}_0$ direction (i.e. the z direction). This is known as the secular approximation or truncation.[3] Under this approximation, the chemical shielding Hamiltonian is rewritten as Equation (8):

$$h^{-1}H_{\mathrm{CS}} = \frac{\gamma}{2\pi}\ \sigma_{zz}B_0I_z \tag{8}$$

Usually σ is a general 3×3 matrix that can be separated into a symmetric and an antisymmetric matrix; however, the antisymmetric part does not contribute to σ_{zz} and therefore has no influence on NMR spectra under the secular approximation. The symmetric chemical shielding tensor can always be diagonalized and the three diagonal elements are called principal components of a chemical shielding tensor, denoted σ_{11}, σ_{22} and σ_{33} ($\sigma_{11} < \sigma_{22} < \sigma_{33}$). The transformation to fulfill such a diagonalization defines the orientation of a chemical shielding tensor with respect to the crystal axis system.

If the direction of the external magnetic field is described by directional cosines $(\cos X, \cos Y, \cos Z)$ in the principal axis system of a chemical shielding tensor, the observed NMR frequency can be written as Equation (9):

$$\nu = \nu_0[1 - (\sigma_{11}\cos^2 X + \sigma_{22}\cos^2 Y + \sigma_{33}\cos^2 Z)] \tag{9}$$

Often one has a sample consisting of many small crystallites in the form of a powder. Therefore, the orientation of the crystallites is random with respect to the external magnetic field. This is equivalent to saying that the orientation of the external magnetic field is random in the principal axis system of the chemical shielding tensor for all crystallites. This leads to a distribution of NMR frequencies and the resultant NMR line shapes are called NMR powder line shapes or powder patterns. Typical powder patterns due to anisotropic chemical shielding are shown in Figure 1(a–c).

The magnitude of the chemical shielding interaction is much smaller than that of the nuclear Zeeman interaction. However, it is its weakness that makes the chemical shielding interaction the most important NMR interaction for chemists. Atomic nuclei are spatially localized and are extremely sensitive to the electronic environment around them. The energy difference between different NMR signals could be of the order of $10^{-8}\,\mathrm{J\,mol^{-1}}$! It is this sensitivity, or perhaps resolution might be a better term, that leads to the many versatile chemical applications of NMR spectroscopy.

It should be mentioned that, in actual NMR experiments, one always measures NMR signals with respect to

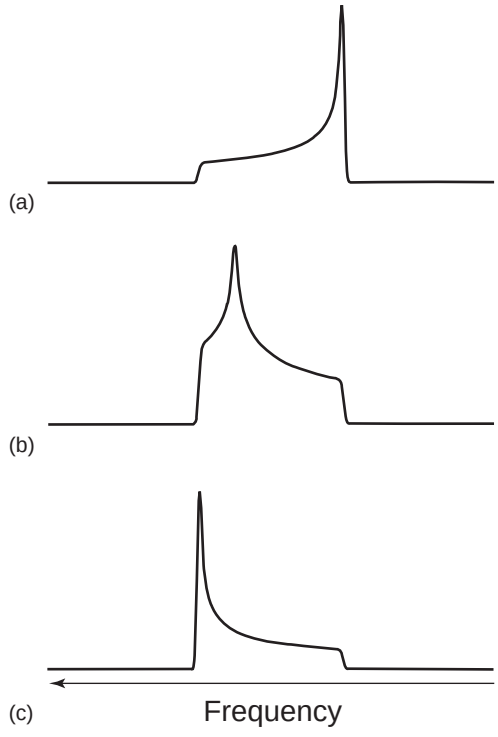

Figure 1 Typical NMR powder line shapes arising from the chemical shift anisotropy (CSA). $\kappa =$ (a) -1.0, (b) 0.5 and (c) $+1.0$.

that of a reference. The difference between NMR signals due to the sample of interest and that of the reference is called the chemical shift, δ, and expressed in units of parts per million (ppm). The relationship between the observed chemical shift and the chemical shielding is given by Equation (10):

$$\delta_{\mathrm{sample}} = \frac{10^6(w_{\mathrm{sample}} - w_{\mathrm{ref}})}{w_{\mathrm{ref}}} \approx 10^6(\sigma_{\mathrm{ref}} - \sigma_{\mathrm{sample}}) \tag{10}$$

Therefore, a chemical shielding tensor is also related to a chemical shift tensor. Once the absolute shielding for the NMR signal in a reference sample has been established, all chemical shift data for the same nuclear species can be easily converted to absolute chemical shielding.

The three principal components of a chemical shift tensor are sufficient to describe the anisotropic chemical shift; for convenience, three other quantities, known as the isotropic value, the span (Ω) and the skew (κ) of a chemical shift tensor, are also used for describing chemical shift tensors.[13] Their definitions are given in Equations (11–14):

$$\sigma_{\mathrm{iso}} = \frac{(\sigma_{11} + \sigma_{22} + \sigma_{33})}{3} \tag{11}$$

$$\delta_{\mathrm{iso}} = \frac{(\delta_{11} + \delta_{22} + \delta_{33})}{3} \tag{12}$$

$$\Omega = \sigma_{33} - \sigma_{11} = \delta_{11} - \delta_{33} \quad (13)$$

$$\kappa = \frac{3(\sigma_{iso} - \sigma_{22})}{\Omega} = \frac{3(\delta_{22} - \delta_{iso})}{\Omega} \quad (14)$$

The meanings of these quantities are clear in terms of the appearance of a powder pattern arising from anisotropic chemical shifts. The isotropic value defines the center of the powder pattern, the span describes the breadth of the powder pattern and the skew represents the asymmetry of the powder pattern. These conventions will be used throughout this article.

One of the advantages of solid-state NMR over solution NMR is the ability to measure the complete chemical shift tensor rather than its average (isotropic) value alone. Ideally, one prefers to study NMR signals of a single-crystal sample at different orientations with respect to the external magnetic field. From the variation of NMR signals as a function of the sample orientation, one can determine the orientation of a chemical shift tensor in the molecular frame of reference. If single crystals are unavailable, one can still obtain information on the three principal components of a chemical shift tensor from studying powder samples; but orientation information is unavailable. In some favorable situations, however, one can employ a technique known as dipolar NMR to powder samples and obtain information on both the magnitude and orientation of chemical shift tensors.[14,15] Chemical shift tensors reported in the literature have been compiled by Duncan.[16]

2.1.3 The Direct Magnetic Dipole–Dipole Interaction

In the presence of a strong external magnetic field, the magnetic dipole moment associated with a nuclear spin angular momentum can be visualised as a tiny bar magnet, hence producing a magnetic field around it. Any nucleus that resides in the vicinity of this tiny bar magnet will experience its magnetic field, in addition to the much larger applied magnetic field.

The interaction between two tiny bar magnets can be described by a magnetic dipole–dipole interaction. The magnetic dipole–dipole interaction Hamiltonian can be readily derived from the classical expression and it has the following form (Equation 15):[1]

$$H_{DD} = (A + B + C + D + E + F)R_{DD} \quad (15)$$

where the various terms are defined in Equations (16–22):

$$A = -I_{1z}I_{2z}(3\cos^2\theta - 1) \quad (16)$$

$$B = \tfrac{1}{4}(I_{1+}I_{2-} + I_{1-}I_{2+})(3\cos^2\theta - 1) \quad (17)$$

$$C = \tfrac{3}{2}(I_{1z}I_{2+} + I_{1+}I_{2z})\sin\theta\cos\theta\exp(-i\phi) \quad (18)$$

$$D = -\tfrac{3}{2}(I_{1z}I_{2-} + I_{1-}I_{2z})\sin\theta\cos\theta\exp(i\phi) \quad (19)$$

$$E = -\tfrac{3}{4}I_{1+}I_{2+}\sin^2\theta\exp(-2i\phi) \quad (20)$$

$$F = -\tfrac{3}{4}I_{1-}I_{2-}\sin^2\theta\exp(2i\phi) \quad (21)$$

$$R_{DD} = \frac{\mu_0}{4\pi}\gamma_1\gamma_2\left(\frac{h}{4\pi^2}\right)\left(\frac{1}{r^3}\right) \quad (22)$$

The terms A–F are known as the "dipolar alphabet" and R_{DD} is the direct dipolar coupling constant. The angles θ and ϕ are defined in Figure 2(a). In the presence of a strong magnetic field along the z direction, the dipolar alphabet is truncated, so that only the A and B terms remain for homonuclear spin pairs (Equation 23):

$$h^{-1}H_{DD} = (A + B)R_{DD} = -\frac{3\cos^2\theta - 1}{2}(3I_{1z}I_{2z} - \boldsymbol{I}_1\boldsymbol{I}_2)R_{DD} \quad (23)$$

It is worth mentioning, however, that the above truncation is valid only if the nuclear Zeeman interaction is much larger than any other interactions involving the nuclear spins. In the presence of other strong interactions, such as a quadrupole interaction, all other terms from C to F may be important.

Since the direct magnetic dipole–dipole interaction is axially symmetric and traceless, there is only one

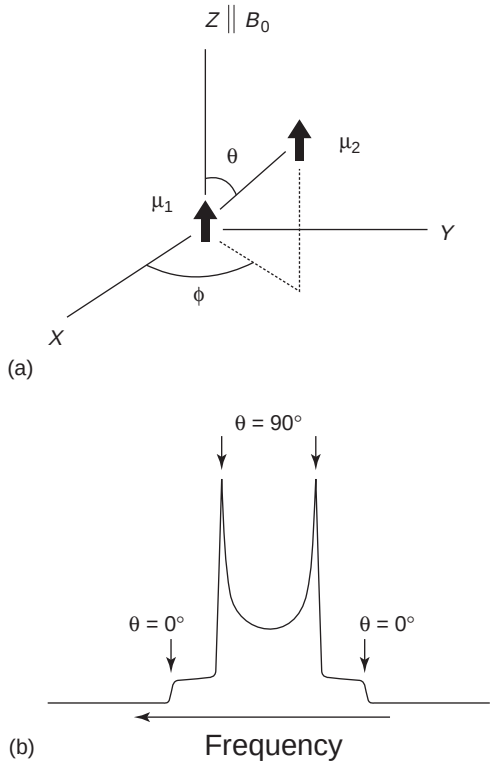

Figure 2 (a) Coordinate system describing two dipolar-coupled spin-$\frac{1}{2}$ nuclei in a strong magnetic field and (b) the resultant NMR powder line shape.

independent variable within the three principal components, i.e. $R_{11} = R_{22} = 3R_{DD}/2$ and $R_{33} = -3R_{DD}$. It is obvious that the unique axis of the principal axis system for the direct magnetic dipole–dipole interaction is along the internuclear vector. Similarly to the anisotropic chemical shielding interaction, the direct magnetic dipole–dipole interaction is also dependent on the crystallite orientation with respect to the external magnetic field. Therefore, for a powder sample, a distribution of NMR frequencies or a powder pattern will result. A typical powder pattern arising from a dipolar interaction between two identical spin-$\frac{1}{2}$ nuclei is displayed in Figure 2(b).

It is readily seen that a direct magnetic dipole–dipole interaction depends on the magnetic moments of two interacting nuclear spins and their spatial separation. The magnitude of a dipolar interaction is expected to be large for nuclear species with a high magnetogyric ratio and also for nuclei with a close spatial separation. By studying a direct magnetic dipole–dipole interaction, valuable structural information can be deduced. For example, one of the pioneering experiments in the early stages of NMR spectroscopy was performed by Pake.[17] He measured the dipolar interaction between each pair of hydrogen nuclei in the water molecules of hydration in gypsum crystals ($CaSO_4 \cdot 2H_2O$), from which a ^1H–^1H separation of 1.58 Å was obtained. As will be shown later, the dipolar interaction is the major source of information about molecular structure from solid-state NMR studies.

2.1.4 *The Indirect Electron-mediated Spin–Spin Interaction*

Another type of nuclear spin interaction in NMR is that between nuclear spins transmitted via intervening electrons, namely the J interaction. In contrast to the direct magnetic dipole–dipole interaction, the J interaction is indirect, since it involves nucleus–electron interactions. The J interaction is more complicated than the three aforementioned nuclear spin interactions. In general, there are three mechanisms for an indirect spin–spin interaction: (i) the interaction between a nuclear spin and electron orbital angular momenta; (ii) the magnetic dipole–dipole interaction between a nuclear spin and electron spins; and (iii) the Fermi contact interaction which arises from the fact that the s electrons have a nonzero probability of being found at the nucleus. Moreover, since the Hamiltonians describing the last two interactions do not commute, they do not have a common set of eigenfunctions. As a result, state mixing occurs between their corresponding eigenstates, leading to a cross-term that contributes to the indirect spin–spin interaction. More detailed discussion on J interactions can be found in the literature.[18]

The indirect spin–spin interaction Hamiltonian can be written as Equation (24):

$$h^{-1}H_J = I_1 \cdot J \cdot I_2 \tag{24}$$

Again, the presence of a strong magnetic field truncates the total Hamiltonian and the secular Hamiltonian becomes Equation (25):

$$h^{-1}H_J = J_{iso}\boldsymbol{I}_1 \cdot \boldsymbol{I}_2 + \frac{J_{zz} - J_{iso}}{2}(3I_{1z}I_{2z} - \boldsymbol{I}_1 \cdot \boldsymbol{I}_2) + \frac{J_{xy} - J_{yx}}{2}(I_{1x}I_{2y} - I_{1y}I_{2x}) \tag{25}$$

where J_{iso} is the isotropic value of the J tensor in frequency units. Often J is assumed to be a symmetric tensor. Under these conditions, the last term in Equation (25) vanishes. Regardless, this latter term does not contribute to the NMR line position in first order. A more detailed discussion concerning the antisymmetric part of a J tensor can be found in the literature.[2] If a J tensor is further assumed to be axially symmetric with the unique axis being coincident with the internuclear vector, one has Equation (26):

$$\begin{aligned} J_{zz} &= \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} J_\perp & 0 & 0 \\ 0 & J_\perp & 0 \\ 0 & 0 & J_\parallel \end{pmatrix} \\ &\quad \times \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}_{zz} \\ &= J_\perp \sin^2\theta + J_\parallel \cos^2\theta \end{aligned} \tag{26}$$

Substituting Equation (26) into Equation (25), one obtains the spin Hamiltonian due to a symmetric J tensor (Equation 27):

$$h^{-1}H_J = J_{iso}\boldsymbol{I}_1 \cdot \boldsymbol{I}_2 + \frac{\Delta J}{3}\frac{3\cos^2\theta - 1}{2}(3I_{1z}I_{2z} - \boldsymbol{I}_1 \cdot \boldsymbol{I}_2) \tag{27}$$

where (Equation 28)

$$\Delta J = J_\parallel - J_\perp \tag{28}$$

It is clear from Equations (23) and (27) that the spatial and spin dependence in the direct magnetic dipole–dipole and the J interaction Hamiltonians are similar. Therefore, it is usually difficult to separate a J tensor from a dipolar interaction tensor, that is, the experimental observable will be a combination of these two interactions. The experimental observable will be a simple quantity known as the effective dipolar coupling constant, R_{eff} (Equation 29):

$$R_{eff} = R_{DD} - \frac{\Delta J}{3} \tag{29}$$

Although numerous chemical applications have been made of the isotropic values of J interactions, much less is known about the anisotropy in J interactions. In fact, as long as the electronic environment at the nucleus deviates from spherical symmetry, the J interaction must be anisotropic. This can be readily appreciated when considering the similarity between the chemical shielding and one of the mechanisms for J interactions, the spin–orbital mechanism. In the case of chemical shielding, it is the external magnetic field causing unquenching of the electron orbital angular momentum, whereas in the case of J interaction, it is the field generated by the nuclear magnetic dipole moment that introduces unquenching of the electron orbital angular momentum. Since the electronic environment is not spherically symmetric, the degree of such unquenching will depend on the orientation of the nuclear magnetic moment in the molecule under study. Note that the orientation of the nuclear magnetic moment for a spin-$\frac{1}{2}$ nucleus is always forced to align along the external magnetic field. Hence the magnitude of a J interaction will depend on how the molecule orients itself with respect to the external magnetic field.

In general, it is difficult to measure J tensors experimentally. To date, the most definite J-tensor measurements have been based on single-crystal NMR studies.[19–21] There are also several cases where J tensors have been determined from solid-state NMR studies of powder samples.[22–30] It should be pointed out that the $^1J(^{31}P,^{31}P)$ tensors in tetraalkylphosphine disulfides,[31] the often quoted examples for J anisotropy, are proved to be nearly isotropic.[32] Finally, information about J tensors is also available from NMR studies of solute molecules dissolved in liquid crystalline solvents. Several J tensors have been measured by this approach, e.g. $J(^{19}F,^{13}C)$[33] and $J(^{199}Hg,^{13}C)$.[34]

2.2 Experimental Solid-state Nuclear Magnetic Resonance Techniques

In the solid state, the absence of rapid molecular tumbling makes the observation of a high-resolution NMR spectrum very difficult. Usually it is helpful to make a distinction between an abundant spin system and a rare spin system, since the NMR techniques involved in the two situations are different. Abundant spins are most often ^1H and ^{19}F nuclei and rare spins are magnetically dilute nuclei such as ^{15}N and ^{29}Si. In this article, we shall consider only solid-state NMR techniques for studying dilute spin-$\frac{1}{2}$ systems.

2.2.1 High-power Abundant Spin Decoupling and Magic-angle Spinning

One major problem encountered in obtaining high-resolution NMR spectra for solids arises from the anisotropic nature of all nuclear spin interactions. Among the anisotropic interactions present in magnetically dilute spin systems, the heteronuclear dipole–dipole interaction between dilute and abundant spins usually gives rise to broad, featureless NMR line shapes, whereas other interactions may result in NMR line shapes with some specific features. Therefore, it is desirable to eliminate the heteronuclear dipolar interaction between dilute and abundant spins. This can be achieved by applying strong RF radiation at the resonance frequency for abundant spins. The strong on-resonance RF radiation rapidly changes the spin state of the abundant spins and leads to spin decoupling. Under such conditions, the J interaction between abundant and dilute spins is also eliminated by spin decoupling. The resultant decoupled NMR spectra of the dilute spins will be determined solely by the chemical shielding. Although useful information can be extracted from such a decoupled NMR spectrum, spectral overlap usually limits its usage to nuclear systems consisting of only a small number of dilute spins.

In order to eliminate further the remaining anisotropic chemical shielding interaction, another technique known as magic-angle spinning (MAS)[35–37] must be employed. The MAS technique involves high-speed sample rotation about an axis inclined 54°44′ with respect to the external magnetic field. The rapid mechanical specimen rotation acts more-or-less like the molecular tumbling in isotropic liquids and results in solid-state NMR spectra consisting of sharp peaks. Typical sample spinning frequencies in modern NMR probes are below 35 kHz. If the sample rotation frequency is smaller than the frequency range that results from the anisotropic chemical shielding interaction, the central (isotropic) NMR peak will be flanked by a set of peaks with the separation between neighboring peaks being equal to the specimen rotation frequency. These peaks are known as the spinning sidebands. Only when the rotation frequency is much greater than the anisotropic interaction (in frequency units) will all the spinning sidebands have negligible intensities. Unfortunately, the CSA for magnetically dilute spins, especially for heavy elements, is usually very large, which leads to the presence of significant spinning sidebands.

As an example, Figure 3(a–d) shows several solid-state ^{31}P NMR spectra of $NH_4H_2PO_4$ under different conditions. It can be seen that without applying the abundant spin (proton) decoupling, the ^{31}P NMR spectrum shows a broad peak, which is due to both the anisotropic ^{31}P chemical shift and the heteronuclear dipolar coupling to protons (Figure 3a). When the proton decoupling is applied, the ^{31}P NMR spectrum becomes sharper and exhibits a typical powder pattern arising from the anisotropic chemical shift (Figure 3b). After combining MAS and high-power ^1H decoupling, the ^{31}P NMR

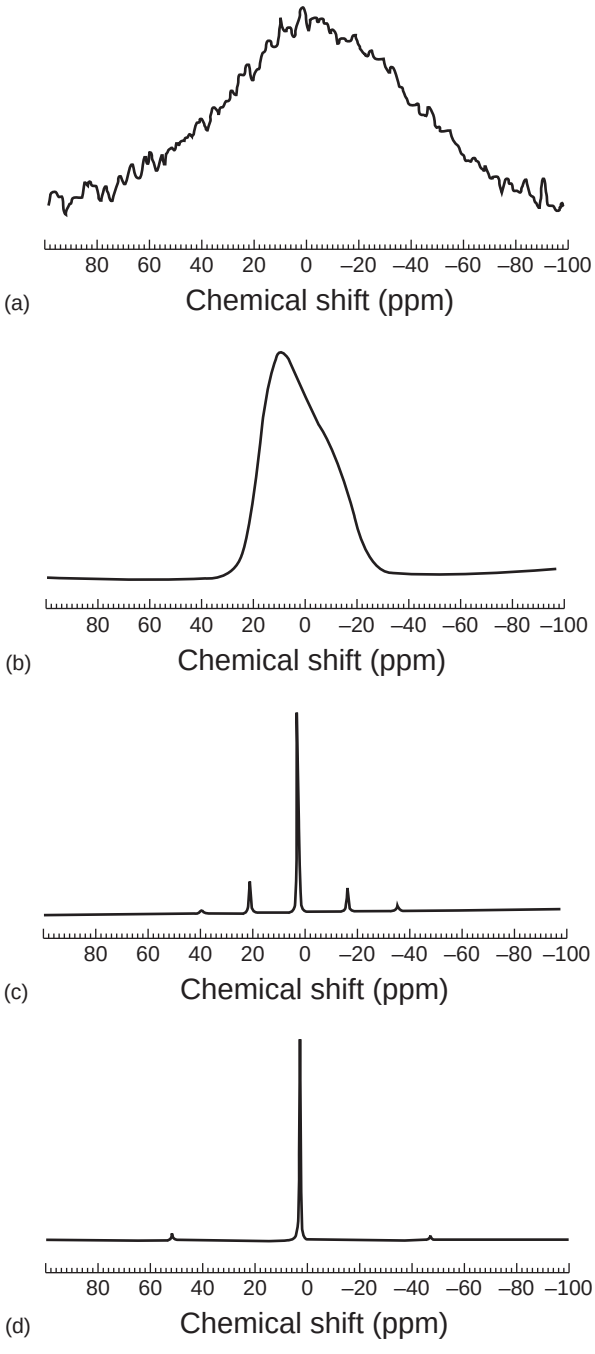

Figure 3 Solid-state ^{31}P NMR spectra of $NH_4H_2PO_4$: (a) without ^1H decoupling; (b) with high-power ^1H decoupling; (c) high-power ^1H decoupling and MAS at 1543 Hz; (d) high-power ^1H decoupling and MAS at 4024 Hz.

spectra bear significantly higher resolution (Figure 3c and d).

2.2.2 *Cross-polarization*

The concept of cross-polarization (CP) was first proposed by Hartmann and Hahn.[38] They demonstrated that spin polarization transfer can occur between two dipolar coupled spin systems, I and S, if the two systems are set into ''thermal contact''. The condition for such a transfer is now termed the Hartmann–Hahn match condition: $\gamma_I B_{1I} = \gamma_S B_{1S}$, where B_{1I} and B_{1S} are the rotating frame RF field strength for I and S, respectively. However, the CP experiment of Hartmann and Hahn did not find widespread applications in chemistry until a direct version of CP was introduced by Pines et al.,[39,40] who demonstrated that NMR signals of dilute spins can be enhanced by transferring magnetization from the abundant spins. This direct version of CP has become a standard technique in obtaining solid-state NMR spectra of dilute spins. The advantage of using CP is not only to enhance the NMR signals of dilute spins, but also to make the repetition time of the NMR experiment depend on the spin–lattice relaxation time of the abundant spins, which is often orders of magnitude shorter than that of the dilute spins. For instance, the ^{13}C spin–lattice relaxation time in a crystalline alkane $C_{60}H_{122}$ is more than 1660 s.[41] To study directly ^{13}C nuclei in this sample, one must wait for over 2 h between two subsequent ^{13}C 90° pulses in order for the ^{13}C spin system to return to thermal equilibrium! On the other hand, the ^1H spin–lattice relaxation time for this sample is of the order of seconds. Hence the repetition time for the CP experiment is only a few seconds. It is these two factors that make solid-state NMR studies of dilute spins feasible. The combination of CP, MAS and high-power abundant spin decoupling, often denoted cross-polarization/magic-angle spinning (CPMAS), was first demonstrated by Schaefer and Stejskal in 1976.[42] Since then CPMAS has become a powerful analytical technique routinely used by chemists in studies of a wide range of chemical systems.

2.2.3 *Single-crystal Nuclear Magnetic Resonance*

As mentioned earlier, single-crystal NMR is the ultimate method for yielding complete information about the orientation of a chemical shift tensor or any spin interaction tensor in the molecular frame of reference. Unfortunately, single-crystal NMR experiments are unfamiliar to chemists. The difficulties of such experiments are twofold. First, it is often difficult to obtain a single-crystal sample that is large enough for NMR measurement (e.g. sides of 1 mm or greater). Second, single-crystal NMR experiments are time-consuming. This is because one must carry out the same measurement for many different sample orientations. In addition, the spin–lattice relaxation times are usually very long in single crystals, adding another twist to the difficult experiments. Details concerning single-crystal NMR spectral analysis can be found in the literature.[43]

It should be noted that, as the single-crystal NMR probe development continues,[44] single-crystal NMR

studies are expected to become more feasible. These types of studies are of fundamental importance since they yield complete information (both the magnitude and orientation) about nuclear spin interaction tensors, providing benchmark data for testing ab initio quantum chemical calculations.

2.2.4 *Dipolar Recoupling: Molecular Structure Determination*

One of the most important solid-state NMR advances in the last decade is the development of NMR techniques that can measure geometric distances between two atomic nuclei. Since the essence of such techniques is to reintroduce dipolar interactions to NMR spectra under MAS conditions, they are also known as dipolar *recoupling* techniques. Recent developments in this area have been reviewed by Griffin et al.[45,46] and Garbow and Gullion.[47] However, reliable distance measurements have been limited to isolated spin pairs; it is natural to expect new methodologies for dealing with multiple-spin systems. Interestingly enough, complementary techniques have also appeared recently which, rather than measuring internuclear distances, provide information about the torsion angle between internuclear vectors.[48–51] Of course, with various torsion-angle constraints known, it is also possible to derive molecular structure.

3 SOLID-STATE NUCLEAR MAGNETIC RESONANCE OF MAIN GROUP ELEMENTS

3.1 Nitrogen-15

There are two NMR-active stable nitrogen isotopes: ^{14}N (spin $= 1$) and ^{15}N (spin $= \frac{1}{2}$). The natural abundance of the ^{15}N isotope is only 0.37%, which makes it an insensitive NMR nucleus at natural abundance. The magnetogyric ratio of ^{15}N is also quite low: $\gamma = -2.712 \times 10^7\,\mathrm{rad\,T^{-1}\,s^{-1}}$ (the ^{15}N NMR frequency is usually the lowest frequency to which most commercial MAS probes can be tuned). Consequently, solid-state ^{15}N NMR at natural abundance has been uncommon. Since ^{15}N isotopically enriched material is readily available now, most of the solid-state ^{15}N NMR studies deal with ^{15}N-labeled compounds. In fact, researchers in the field of protein structure determination using solution NMR spectroscopy usually utilize uniformly ^{15}N-labeled proteins which can be readily obtained by expression in bacteria grown on a medium with $^{15}NH_4Cl$. A similar trend is emerging in the area of solid-state ^{15}N NMR studies of proteins and polypeptides.[52]

The nitrogen chemical shift covers a range of approximately 1000 ppm. The primary external standard for ^{15}N chemical shift referencing is the neat $MeNO_2$. Another common ^{15}N chemical shift reference is liquid ammonia at 20 °C, whose ^{15}N NMR signal appears at −380.4 ppm with respect to that of $MeNO_2$. The only advantage of using liquid NH_3 as a reference is that its nitrogen nucleus is more shielded than most of the nitrogen nuclei in chemical compounds. As a result, nearly all the ^{15}N chemical shifts are positive on the liquid NH_3 scale. Nevertheless, one should be careful about the actual ^{15}N chemical shift scale used in the published reports. The absolute nitrogen shielding constant in neat $MeNO_2$ and liquid ammonia at 300 K is −135.8 and 244.6 ppm, respectively.[53] In ^{15}N CPMAS experiments, the set-up sample as well as the secondary reference is a solid sample of $^{15}NH_4NO_3$, which gives rise to an NMR signal at −359 ppm with respect to that of $MeNO_2$, or 23.8 ppm with respect to that of liquid ammonia.

Although the isotropic ^{15}N chemical shifts are known for hundreds of compounds over the past 50 years, much less experimental information is available on ^{15}N chemical shift tensors. The first ^{15}N chemical shift tensor was reported by Gibby et al. in 1972.[54] However, not until recently have considerable efforts been devoted to the systematic measurement of ^{15}N chemical shift tensors. To date, ^{15}N chemical shift tensors have been characterized in many important functional groups.[16]

As mentioned earlier, solid-state NMR studies of single crystals, powder and oriented samples can yield information about chemical shift tensors. The first ^{15}N chemical shift tensor for the peptide bond was fully characterized by single-crystal NMR.[55] Using a dipolar NMR approach, Wasylishen et al. have measured ^{15}N chemical shift tensors in the oxime, imine, azo, and nitroso fragments for powder samples.[56] Using oriented samples, Cross et al. studied the conformation of ^{15}N-labeled gramicidin A in an oriented phosphatidylcholine bilayer environment.[57,58]

The largest ^{15}N CSA values are those found in nitroso compounds and their complexes with transition metals. For example, the ^{15}N NMR spectrum of *p*-nitroso-*N*,*N*-dimethylaniline exhibits a remarkably large CSA: $\Omega = 1500$ ppm.[59] Groombridge et al.[60] also observed large ^{15}N CSAs in Co(NO)(TPP). Recently, Oldfield et al.[61,62] carried out extensive investigations on a series of nitrosoarene compounds and their complexes to heme proteins. They found that the orientation of the ^{15}N chemical shift tensor in the metal complexes is similar to that of the free ligand, but the anisotropy changes drastically. Figure 4(a) and (b) shows the orientations of the two ^{15}N chemical shift tensors in Fe(TPP)(Ph^{15}NO)-(1-methylimidazole) and

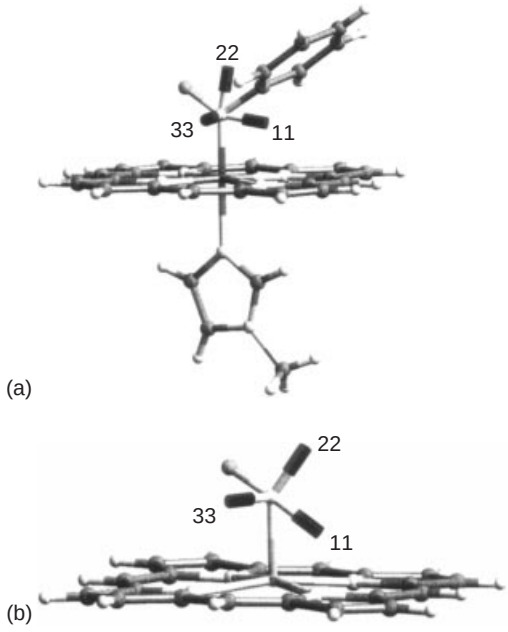

Figure 4 Orientations of the ^{15}N chemical shift tensors in (a) Fe(TPP)(Ph^{15}NO)(1-methylimidazole) and (b) Co(OEP)(^{15}NO). TPP = tetraphenylporphyrin, OEP = octaethylporphyrin. (Reprinted with permission from Godbout et al.[62] Copyright 1998 American Chemical Society.)

Co(OEP)(^{15}NO) according to the density functional theory (DFT) shielding calculations.

In the areas of organic and inorganic chemistry, there are also many useful applications of solid-state ^{15}N NMR. A more comprehensive account for ^{15}N NMR applications can be found in the literature.[63]

3.2 Silicon-29

The natural abundance of ^{29}Si is 4.7%. The ^{29}Si chemical shift reference is TMS. The ^{29}Si chemical shift range is small: approximately 150 ppm. Most of the ^{29}Si chemical shift values are negative and typical ^{29}Si CSAs are less than 300 ppm. The set-up sample for ^{29}Si CPMAS experiments is the Q_8M_8 $[Si_8O_{20}][Si(CH_3)_3]_8$. The most important applications of solid-state ^{29}Si NMR are in various areas of silicate science including zeolites and silicate glasses and in organosilicon chemistry. Reviews of solid-state ^{29}Si NMR can be found in the literature.[64–66]

Two-dimensional (2-D) ^{29}Si MAS NMR techniques such as correlation spectroscopy (COSY) and incredible natural abundance double quantum transfer experiment (INADEQUATE) have been used to establish the lattice connectivity in zeolites and silicate glasses. Figure 5 shows the 2-D solid-state ^{29}Si COSY spectrum of high-silica zeolite ZSM-12.[67] The observation of seven ^{29}Si NMR peaks with approximately equal intensities indicates that the asymmetric unit consists of seven crystallographically

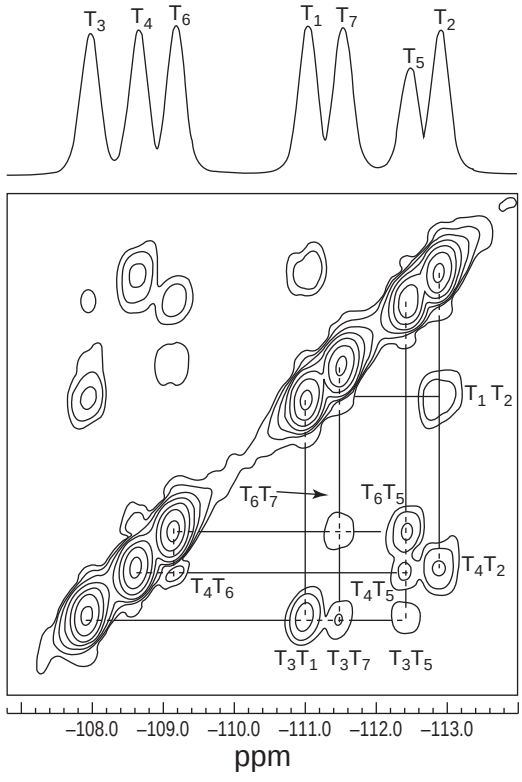

Figure 5 Contour plot of the 2-D ^{29}Si COSY spectrum of ZSM-12 with a projection in the F_2 dimension. (Reprinted with permission from Fyfe et al.[67] Copyright 1990 American Chemical Society.)

distinct Si sites with equal occupancies. From the 2-D spectrum, information about the $^{29}Si-O-^{29}Si$ framework connectivity can be unambiguously assigned from the cross peaks that connect different ^{29}Si NMR signals.

3.3 Phosphorus-31

Phosphorus-31 is among the most sensitive NMR nuclei because of its favorable NMR properties: spin = $\frac{1}{2}$, natural abundance = 100% and $\gamma = 10.841 \times 10^7$ rad T^{-1} s^{-1}. In contrast to ^{15}N, the ^{31}P NMR frequency is usually the high-frequency end of commercial MAS probes. The ^{31}P chemical shift standard is 85% aqueous H_3PO_4, which has an absolute shielding constant of $\sigma = 328.35$ ppm.[68] The set-up sample for ^{1}H-to-^{31}P CP experiments is solid ammonium dihydrogenphosphate, $NH_4H_2PO_4$. This compound is also used as a secondary reference for ^{31}P chemical shifts. The ^{31}P NMR signal of solid $NH_4H_2PO_4$ appears at 0.81 ppm with respect to that of 85% aqueous H_3PO_4. Typical ^{31}P chemical shifts for organophosphorus compounds are within a range of 500 ppm. However, extreme deshielding has also been observed for the ^{31}P nuclei in bridging phosphinidene complexes, $\delta = 900-1400$ ppm,[69] and in terminal phosphido

complexes, $\delta = 1000-1300$ ppm.[70,71] On an absolute shielding scale, the ^{31}P NMR signals of these complexes are approximately 600–1100 ppm more *deshielded* than that of a bare ^{31}P nucleus!

The utility of solid-state ^{31}P NMR in studying catalytic compounds was first recognized in the early 1980s.[72,73] Now solid-state ^{31}P NMR is a routine technique in inorganic chemistry and organometallic chemistry for identifying molecular structure, bonding and polymorphism in conjunction with the X-ray diffraction technique. Applications of ^{31}P CPMAS NMR to organophosphorus compounds and metal–phosphine complexes have been reviewed.[74] More advanced techniques including various 2-D NMR techniques such as *J*-resolved,[75] COSY[76] and dipolar recoupling[77] have also been applied to solid metal–phosphine complexes and phosphates.

Phosphorus chemical shift tensors have been measured in a large number of compounds.[16] Several examples containing terminal phosphorus atoms are given in Table 2. These complexes exhibit the largest ^{31}P CSAs observed to date. As seen in Figure 6(b), the spinning sidebands cover a frequency range of 300 kHz at 7.45 T. It is noted from Table 2 that the most shielded components, δ_{33}, for all the compounds listed are similar whereas the least shielded components, δ_{11}, differ by more than 2000 ppm.

Despite a large body of experimental data, a complete understanding of ^{31}P chemical shift tensors, especially their relationship to molecular structure and bonding, is still under development. It is noted that several single-crystal ^{31}P NMR studies of metal–phosphorus complexes have appeared in the literature.[82–84] These studies yield precise information about both the magnitude and orientation of the phosphorus chemical shift tensors, which can be used as benchmarks for testing ab initio chemical shielding calculations.

Table 2 Phosphorus chemical shift tensors in compounds containing a P–X triple bond

Compound	δ_{iso}	δ_{11}	δ_{22}	δ_{33}	Ω	Ref.
P≡N	275	734	734	−642	1376	78
P≡P	586	1200	1200	−641	1841	79
P≡CH	−74	209	209	−639	848	79
P≡C[$C_6H_2(C(CH_3)_3)_3$]	31	229	140	−274	503	80
[P≡N[$C_6H_2(C(CH_3)_3)_3$]$^+$ [$AlCl_4$]$^-$	77	308	196	−273	581	81
P≡Mo[N(tBu)C_6H_5]$_3$	1217	1987	1987	−324	2311	79
P≡Mo[N(tBu)(3,5-$C_6H_3Me_3$)]$_3$	1208	1978	1978	−330	2308	79
P≡Mo[NN_3]	1328	2125	2125	−267	2392	79
P≡W[NN_3]	1059	1728	1728	−280	2008	79
P≡Mo[NH_3]$_3$	1463	2228	2228	−67	2295	79

All chemical shifts are in ppm. $NN_3 = [Me_3SiNCH_2CH_2)_3N]^{3-}$.

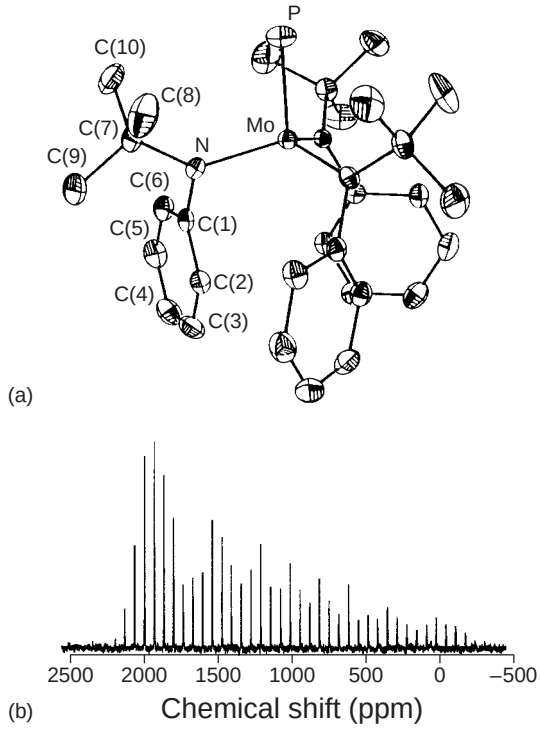

Figure 6 (a) Molecular structure and (b) ^{31}P MAS spectrum of Mo(P)[N(tBu)Ph]$_3$.

Solid-state ^{31}P NMR has also been a rich source for new spectral phenomena. The most important example is the discovery and development of the rotational resonance (R^2) methodology. The R^2 phenomenon was first observed by Andrew et al. in the ^{31}P MAS spectra of solid PCl_5.[85] They noticed that significant line broadening occurs when the sample spinning frequency is matched with the isotropic chemical shift difference between the signals arising from $[PCl_4]^+$ and $[PCl_6]^-$. However, this interesting observation had gone unnoticed until almost 25 years later when Raleigh et al.[86,87] demonstrated that it is possible to utilize R^2 to measure the distance between a pair of homonuclear spins in polycrystalline solids. Recent developments in this area have been reviewed.[45–47] Other examples of new phenomena related to solid-state ^{31}P NMR include the observations of *J* tensors[20,21] and *J*-recoupling phenomena.[88–90]

3.4 Selenium-77

The selenium element has only one naturally occurring spin-$\frac{1}{2}$ nucleus, ^{77}Se (7.58%). Solid-state ^{77}Se NMR is not very difficult with modern NMR instrumentation. The ^{77}Se chemical shift reference is $Se(CH_3)_2$ at 23 °C, $\delta = 0$ ppm. The ^{77}Se chemical shift spans a range of 1600 ppm. Ammonium selenate, $(NH_4)_2SeO_4$, is often

used as a set-up sample for ^{77}Se CPMAS experiments and also as a secondary reference sample, $\delta = 1040.2$ ppm.[91] The largest ^{77}Se CSA so far observed from solid-state NMR experiments is that found in $(NH_4)_2WSe_4$, $\Omega = 1500$ ppm.[92] Several large ^{77}Se CSAs of similar magnitude were also reported, e.g. $\Omega = 1105$ ppm in SeO_2Cl^-[91] and $\Omega = 1261$ ppm in $[Me_4N]_2[Hg(Se_4)_2]$.[93] It is also noted that, from a liquid-crystal ^{77}Se NMR study of carbon diselenide, a remarkably large CSA was reported, $\Omega = 2210$ ppm.[94] Although a large body of solution ^{77}Se NMR data is available in the literature, solid-state ^{77}Se NMR studies have been scarce, most being concerned with inorganic compounds and materials.[95] Since Se compounds are also important in organic chemistry, solid-state ^{77}Se NMR is expected to be useful in the study of organoselenium compounds.

4 SOLID-STATE METAL NUCLEAR MAGNETIC RESONANCE

4.1 Cadmium-113 and Mercury-199

Cadmium has two stable spin-$\frac{1}{2}$ isotopes, ^{111}Cd (12.75%) and ^{113}Cd (12.26%). Since the ^{113}Cd nucleus exhibits a slightly larger γ value than ^{111}Cd, the former has a higher NMR sensitivity. Hence ^{113}Cd is the nucleus of choice for studying cadmium. The ^{113}Cd chemical shift reference is the ^{113}Cd NMR signal of 0.1 M aqueous solution of $Cd(ClO_4)_2$. The set-up sample for ^{113}Cd CPMAS experiments is $Cd(NO_3)_2 \cdot 4H_2O$, which is also used as a secondary reference sample, $\delta = -100$ ppm.[96] The ^{113}Cd chemical shift range is approximately 900 ppm and the largest ^{113}Cd CSA so far reported is that found in the complex of 18-crown-6 and $CdCl_2$, $\Omega = 1100$ ppm.[97,98] Because of its reasonably high natural abundance and favorable NMR frequency, solid-state NMR of ^{113}Cd is straightforward.

The initial impetus of ^{113}Cd NMR studies includes the possibility of using this nucleus as a surrogate nuclear probe to calcium- and zinc-containing proteins.[99,100] In addition, since divalent group 12 metal ions are diamagnetic and have a closed-shell d^{10} electronic structure, neither electron spin resonance (ESR) nor optical spectroscopy is applicable. In this respect, ^{113}Cd NMR is a unique technique capable of providing useful information about the metal binding to proteins. To date ^{113}Cd NMR has been well established as an effective probe to Cd-substituted metalloproteins. Applications of ^{113}Cd NMR have been summarized by Summers[101] and more recently by Coleman.[102]

Several general trends concerning the ^{113}Cd chemical shift have been observed: (a) oxygen ligands generate the most shielded cadmium environment; (b) sulfur ligands are the least shielded; and (c) nitrogen and halide ligands provide a shielding environment between the oxygen and sulfur limits. For example, in a study of cadmium thiocyanate complexes, Eichele and Wasylishen found an almost additive relationship between the ^{113}Cd chemical shift and the number of S-bonded thiocyanate ligands.[103] As shown in Figure 7, replacement of an N-bonded thiocyanate ligand by an S-bonded ligand causes an average ^{113}Cd NMR signal shift of 38 ppm in the direction of deshielding.

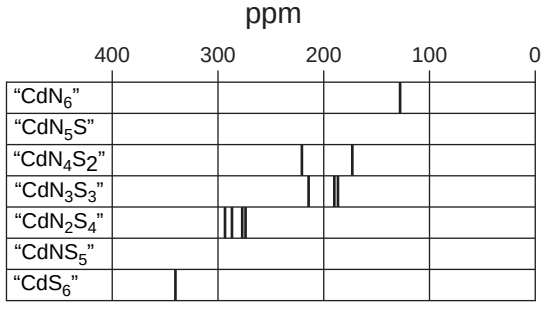

Figure 7 Dependence of the isotropic ^{113}Cd chemical shift on the coordination mode of the thiocyanate ligands. (Reprinted with permission from Eichele and Wasylishen.[103] Copyright 1994 American Chemical Society.)

In order to understand fully ^{113}Cd chemical shifts, considerable efforts have been devoted to the measurement of ^{113}Cd chemical shift tensors. In the last two decades, a large number of ^{113}Cd chemical shift tensors have been characterized using both MAS and single-crystal NMR techniques.[16] A good example illustrating why it is important to study the full chemical shift tensor rather than the isotropic chemical shift alone was given by Jakobsen et al.[104] in the study of the *meso*-TPP of cadmium (Cd-TPP) and its pyridyl adduct (Py-Cd-TPP). The isotropic ^{113}Cd chemical shift difference between the two compounds is only 33 ppm, making it difficult to distinguish between the two binding modes based on the observed isotropic chemical shifts alone. However, as seen in Figure 8, the two ^{113}Cd chemical shift tensors are remarkably different: $\Omega = 341$ and 105 ppm for Cd-TPP and Py-Cd-TPP, respectively. It is clear that the pyridyl ligation causes opposite shielding changes in the individual tensor components, which essentially cancel the change in the isotropic chemical shift. There are many examples of this kind that emphasize the rich information from studying chemical shift tensors.

Unlike ^{113}Cd NMR, solid-state ^{199}Hg NMR is less common. This is mainly because Hg compounds often exhibit very large CSAs, especially in linear Hg compounds. For example, liquid-crystal NMR studies showed that the ^{199}Hg CSA in Me_2Hg is $\Omega = 7475 \pm 80$ ppm.[105] This value corresponds to a frequency range of 536 kHz at 9.4 T. The ^{199}Hg chemical shift spans a range of 5000 ppm.

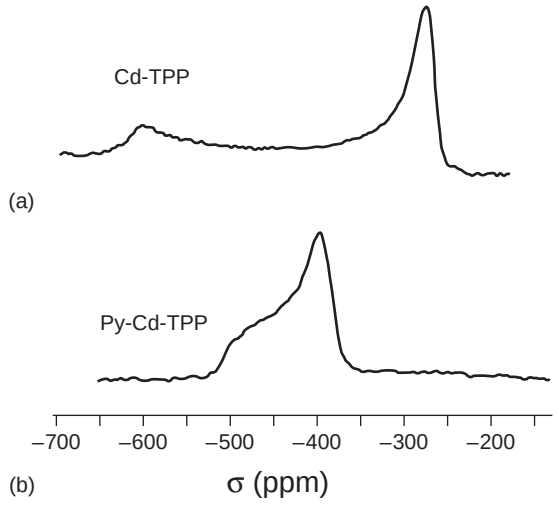

Figure 8 Experimental ^{113}Cd CPMAS spectra of (a) Cd-TPP and (b) Py-Cd-TPP. (Reprinted with permission from Jakobsen et al.[104] Copyright 1982 American Chemical Society.)

The commonly accepted ^{199}Hg chemical shift reference is neat dimethylmercury, $\delta = 0$ ppm (***Caution!*** this compound is highly toxic). As a result of the recent tragic death of a chemist from dimethylmercury poisoning,[106] attention has been drawn to using either a safer standard or an alternative approach for ^{199}Hg NMR referencing. The suggested alternative approach is to first measure the ^{1}H NMR frequency at 0 ppm, and to calculate the ^{199}Hg frequency at 0 ppm using the conversion factor of 0.17 910 323.[107] In the solid-state ^{199}Hg NMR community, researchers always use a *solid* sample as a set-up sample as well as a secondary reference. Although solid Hg compounds are far less hazardous than neat dimethylmercury, one must take precautions in handling any Hg compound.

The originally proposed set-up sample for ^{199}Hg CPMAS experiments was $Hg(CH_3COO)_2$, $\delta = -2497$ ppm.[108] However, this compound is far from being an ideal set-up sample since it has a very large value of ^{199}Hg CSA ($\Omega = 1826$ ppm) and a long ^{1}H spin–lattice relaxation time ($T_1 = 25$ s). These disadvantages render the set-up procedure time-consuming. Two new compounds have recently been proposed as set-up samples for ^{199}Hg CPMAS experiments: $[Hg(DMSO)_6](O_3SCF_3)_2$, $\delta = -2313$ ppm,[109] and $NEt_4Na[Hg(CN)_4]$, $\delta = -434$ ppm.[110] Using these new set-up samples, the set-up of ^{199}Hg CPMAS experiments becomes fairly easy.

Only a handful of ^{199}Hg chemical shift tensors have been reported in the literature.[16] Several Hg–thiolate complexes, which can serve as models for $[Hg(SCys)_n]$ centers in the bacterial mercury resistance proteins, have been studied by solid-state ^{199}Hg NMR.[111] An enormous ^{199}Hg CSA was observed for the linear Hg–thiolate complex, $Hg(S\text{-}2,4,6\text{-}iPr_3C_6H_2)_2$, $\Omega = 4479$ ppm. A similarly large ^{199}Hg chemical shift tensor was also reported for $Hg_2(NO_3)_2 \cdot 2H_2O$, which contains an Hg–Hg bond, $\Omega = 3203$ ppm.[112]

Since Hg compounds often have large CSAs, it is desirable to perform CPMAS experiments with the highest spinning frequency possible in order to eliminate the spinning sidebands. However, as mentioned earlier, when the sample spinning frequency reaches about 10 kHz or higher, the CP efficiency often decreases drastically. An approach known as variable-amplitude cross-polarization (VACP) was demonstrated to be effective in circumventing this problem.[113] In general, it is still difficult to obtain solid-state ^{199}Hg NMR spectra. However, recent investigations suggested that ^{199}Hg NMR is a more sensitive probe than ^{113}Cd NMR and can be applied to metalloproteins and protein–DNA complexes.[114] As NMR instrumentation continues to improve, it is expected that ^{199}Hg NMR will become increasingly accessible in biological solids.

4.2 Tin-119

Tin has three naturally occurring stable spin-$\frac{1}{2}$ isotopes: ^{115}Sn (0.35%), ^{117}Sn (7.61%) and ^{119}Sn (8.58%). Among these, ^{119}Sn exhibits the highest NMR sensitivity. The ^{119}Sn chemical shift standard is 5% $SnMe_4$ in CH_2Cl_2. The absolute ^{119}Sn chemical shielding scale was recently established, which yields the absolute chemical shielding constant for $SnMe_4$, $\sigma = 2180 \pm 200$ ppm.[115] This piece of information is necessary if one wishes to make direct comparisons between experimental ^{119}Sn chemical shifts and ab initio shielding calculations. Typical ^{119}Sn chemical shift values appear in a range of 800 ppm. The set-up sample for ^{119}Sn CPMAS experiments is tetracyclohexyltin, $(C_6H_{11})_4Sn$, which is also used as a secondary reference sample, $\delta = -97.35$ ppm.[108]

The first ^{119}Sn CPMAS NMR study was reported in 1978.[116] Since then, solid-state ^{119}Sn NMR has been applied to a large number of systems, including both inorganic materials and organotin compounds. A useful summary of solid-state ^{119}Sn NMR applications in organotin chemistry was given by Sebald.[117] Typical ^{119}Sn CSAs are below 1000 ppm. The largest ^{119}Sn CSA for diamagnetic tin-containing compounds is that for bis[bis(trimethylsilyl)methyl]tin (Figure 9a), $\Omega = 1500$ ppm. This compound also exhibits the most deshielded isotropic ^{119}Sn chemical shift, $\delta_{iso} = 700$ ppm[118] (see Figure 9b).

Potentially, solid-state ^{119}Sn NMR can be extended to the study of biological systems. For example, solution ^{119}Sn NMR has been proved useful in the studies of protoporphyrin–hemoprotein interactions[119] and of

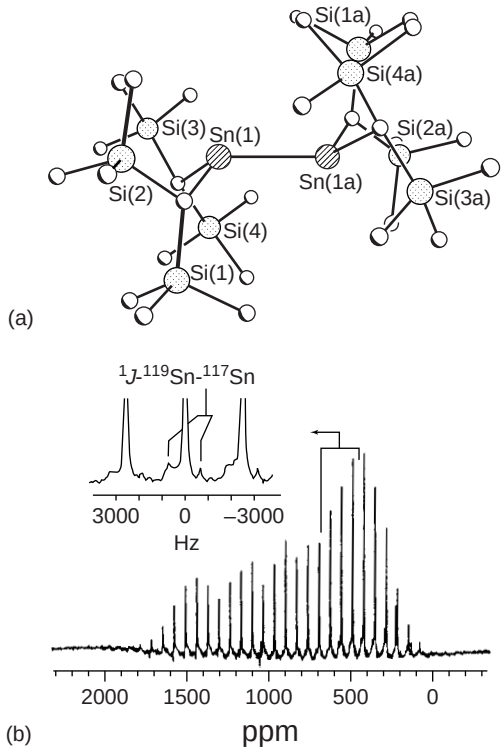

Figure 9 (a) Molecular structure and (b) ^{119}Sn CPMAS spectrum of bis[bis(trimethylsily)methyl]tin. The inset shows the presence of $^1J(^{117}\text{Sn}, ^{119}\text{Sn})$. (Reprinted with permission from Zilm et al.[118] Copyright 1987 American Chemical Society.)

new tin derivatives of antibiotics.[120] Solid-state ^{119}Sn NMR is expected to be useful in these areas.

4.3 Platinum-195 and Lead-207

The stable spin-$\frac{1}{2}$ nuclei of platinum and lead are ^{195}Pt (33.8%) and ^{207}Pb (22.6%). These two nuclei have very similar γ values so their NMR frequencies are also close, e.g. $\nu(^{195}\text{Pt}) = 107.100\,\text{MHz}$ and $\nu(^{207}\text{Pb}) = 104.630\,\text{MHz}$ at 11.75 T. The ^{195}Pt chemical shift standard is aqueous K_2PtCl_6. For ^{207}Pb NMR, the standard sample is 85% Me_4Pb in toluene. An external secondary reference sample of 1.0 M aqueous $Pb(NO_3)_2$ is often used in practice for ^{207}Pb chemical shift referencing, $\delta = -2961.2\,\text{ppm}$.

The chemical shift range for both ^{195}Pt and ^{207}Pb is very large, approximately 16 000 ppm. Such a large chemical shift range also suggests that the ^{195}Pt and ^{207}Pb CSAs must be large. In fact, a common feature of the solid-state ^{195}Pt and ^{207}Pb NMR spectra is the presence of a large number of spinning sidebands. The largest ^{195}Pt CSA was that found in K_2PtCl_4, $\Omega = 10.5 \times 10^3$ ppm, based on single-crystal ^{195}Pt NMR studies.[121,122] In this case, MAS would not be very useful in eliminating the ^{195}Pt CSA.

For ^{195}Pt CPMAS experiments, the originally recommended set-up sample was $K_2Pt(OH)_6$.[123] However, it often takes at least 100 scans to see a reliable signal, making this sample not completely satisfactory. A better set-up sample, $(NH_4)_2PtCl_6$, was suggested by Hayashi and Hayamizu.[124] Using this sample, a ^{195}Pt CPMAS NMR signal can be observed after one pulse, $\delta = 240\,\text{ppm}$ at 300 K. The set-up sample for ^{207}Pb CPMAS experiments is $(p\text{-tolyl})_4Pb$, $\delta = -148.8\,\text{ppm}$.[125,126]

It is well known that ^{195}Pt and ^{207}Pb NMR chemical shifts depend critically on the sample temperature at which the measurement is carried out. A recent study indicates that the solid-state ^{207}Pb NMR spectra of $Pb(NO_3)_2$ are highly sensitive to the variations of the sample temperature.[127] This study also reveals a localized heating effect caused by rapid sample spinning. The temperature-dependent phenomenon of solid $Pb(NO_3)_2$ suggests that this compound could be used for temperature calibration inside NMR probes.[128] Dybowski et al. have also studied the ^{207}Pb chemical shift tensors in a variety of lead salts and the novel Pb–Pb chemical bonding in lead oxides.[129–131]

A potentially important but unexplored research area is solid-state ^{195}Pt and ^{207}Pb NMR of biological systems. Solution ^{195}Pt NMR has been shown useful in studying interactions between antitumor drugs and DNA molecules.[132] Similarly, solution ^{207}Pb NMR has been demonstrated as a new probe to calcium-binding proteins.[133] It might be possible to apply solid-state ^{207}Pb NMR to this type of metalloproteins, in analogy with the successful applications of solid-state ^{113}Cd NMR. Since ^{207}Pb exhibits a chemical shift range approximately 20 times that of ^{113}Cd (or three times of that of ^{199}Hg), ^{207}Pb NMR should be more sensitive to the chemical environment at the metal active sites. Of course, the large ^{207}Pb CSAs may also make solid-state NMR measurement difficult.

4.4 Low-γ Nuclei: Iron-57, Yttrium-89, Silver-109 and Tungsten-183

In addition to the common main group elements and the aforementioned metal nuclei, there is another class of spin-$\frac{1}{2}$ metal nuclei that are distinct from an NMR viewpoint. These metal nuclei, including ^{57}Fe (2.19%), ^{89}Y (100%), ^{103}Rh (100%), $^{107/109}$Ag (51.82 and 48.18%), ^{183}W (14.40%) and ^{187}Os (1.64%), have very low γ values and are sometimes known as the "Cinderella nuclei".[134] Although in recent years these low-γ nuclei have become increasingly accessible by solution NMR spectroscopy,[135] solid-state NMR studies for this class of nuclei are still rare. The major problems with solid-state NMR experiments for low-γ nuclei include (a) exceedingly long spin–lattice relaxation times,

(b) large CSAs, (c) low intrinsic sensitivity associated with the low NMR frequencies and (d) acoustic ring effects of the NMR probe.

Despite all these difficulties, direct observations of ^{57}Fe, ^{89}Y and ^{183}W NMR signals in solids have been demonstrated by Oldfield et al.[136–138] and Dupree and Smith.[139] In the solid-state ^{57}Fe NMR studies, isotopic enrichment to 94.5% was employed to enhance the NMR signals. A large ^{183}W CSA was reported for the "Keggin" compound, $H_3[P(W_{12}O_{40})]\cdot nH_2O$: $\delta_{11} = 409$, $\delta_{22} = 148$ and $\delta_{33} = -1079$ ppm.[137] Solid-state ^{89}Y NMR has also been used to study high-T_c superconductors $YBa_2Cu_3O_{6+x}$.[140] For $^{107/109}$Ag, no direct observation of solid-state ^{109}Ag NMR was made in any silver compound.

One seemingly obvious solution to some of the difficulties encountered in direct NMR observation of low-γ nuclei is to transfer spin polarization from the abundant spins, namely the CP approach. However, the utility of CP under the Hartmann–Hahn condition to low-γ dilute spins is faced with some practical problems. Merwin and Sebald have demonstrated CPMAS experiments for ^{89}Y, ^{109}Ag and ^{183}W nuclei.[141–143] They found that it is difficult to directly optimize the CP efficiency. Therefore, they used an alternative approach in practice. The recommended procedure is (1) to set the proton transmitter power to a relatively low level (90° pulse width = 10 μs), then (2) to use an oscilloscope to calibrate the power level at the observation channel according to the Hartmann–Hahn condition and finally (3) to use a solid sample to fine-tune the Hartmann–Hahn match.

A common feature of the CPMAS experiments involving low-γ nuclei is that relatively long mixing times are required to achieve an appreciable degree of magnetization transfer. This is due to the fact that the dipolar interaction between ^{1}H and low-γ nuclei is always fairly weak as a result of the low γ values of the metal nuclei and the long metal–proton separations. Under such circumstances, CPMAS works only for those systems with long ^{1}H spin–lattice relaxation times in the rotating frame ($T_{1\rho}$). For systems with short ^{1}H $T_{1\rho}$ values, the ^{1}H magnetization would decay to zero before any substantial transfer occurs. Therefore, the drawback of the CPMAS experiments for low-γ nuclei is that it is difficult to predict whether or not one could observe any CP signals. For this reason, it is highly recommended that one measures ^{1}H $T_{1\rho}$ for the system in question before attempting CPMAS experiments. Another potentially important problem is that the CP efficiency decreases with spinning speed. As mentioned earlier, new CP techniques such as RAMP-CP or VACP should also be useful to low-γ spin-$\frac{1}{2}$ nuclei. Perhaps the only advantage associated with low-γ metal nuclei is that since the dipolar interactions between protons and low-γ metal nuclei are weak, it is often not so critical to have very high RF power for ^{1}H decoupling.

5 SECOND-ORDER QUADRUPOLAR EFFECTS ON SPIN-$\frac{1}{2}$ SPECTRA

In the above discussion, we have assumed that MAS can completely average the heteronuclear dipolar interaction between two dilute spins. This assumption is, however, only valid for cases where the two interacting spins are both spin-$\frac{1}{2}$ nuclei. If the spin-$\frac{1}{2}$ nucleus under study is dipolar coupled to a quadrupolar nucleus (spin number $> \frac{1}{2}$), the dipolar interaction between the two spins cannot be completely removed by MAS. This arises from the fact that the quadrupolar spins are no longer quantized along the direction of the external magnetic field due to the presence of the quadrupolar interaction. Consequently, the so-called "residual dipolar couplings" are often observed in the MAS spectra of spin-$\frac{1}{2}$ nuclei. The quadrupole perturbation on magnetic dipolar interactions was first noticed in solid-state ^{19}F and ^{1}H NMR spectra of stationary samples,[144,145] and later observed in ^{13}C MAS NMR spectra.[146,147] However, most of the recent cases are concerned with MAS spectra of spin-$\frac{1}{2}$ nuclei other than ^{13}C and ^{1}H. Therefore, it is worth discussing this subject further. In the discussion that follows, we present an example from the recent literature to illustrate the effect. More complete coverage of the literature can be found in the review article by Harris and Olivieri.[148]

Figure 10(a–d) shows the isotropic region of the ^{31}P MAS spectra of all-*trans*-$Ru(PEt_3)_2(CO)_2(C{\equiv}CPh)_2$.[149] The spectrum exhibits an intense central peak and small satellite peaks. The intense peak is due to the ^{31}P nucleus bound to the NMR-inactive (zero-spin) Ru isotopes, whereas the satellite peaks arise from the J-coupling between ^{31}P and $^{99/101}$Ru nuclei. Since both ^{99}Ru (12.7%) and ^{101}Ru (17.1%) are spin-$\frac{5}{2}$, one might expect the ^{31}P CPMAS spectrum of a molecule containing an Ru–P spin pair to consist of six equally spaced peaks, arising from the six different spin states, $m_Z = \frac{5}{2}, \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$ and $-\frac{5}{2}$. However, as clearly seen from Figure 10(a–d), the separations between the satellite peaks are "squeezed" at the high-frequency end and "stretched" at the opposite direction. These uneven separations between the satellite peaks are due to the different degrees of mixing between ^1J(Ru,P) and the "residual dipolar coupling". Analysis of this type of NMR spectra can yield information about $^1J(^{99/101}Ru,^{31}P)$, $^{99/101}$Ru quadrupole couplings and $^{99/101}$Ru–^{31}P dipolar couplings. Since direct NMR observation of quadrupolar nuclei is often difficult, the second-order quadrupole effect on spin-$\frac{1}{2}$ spectra is an

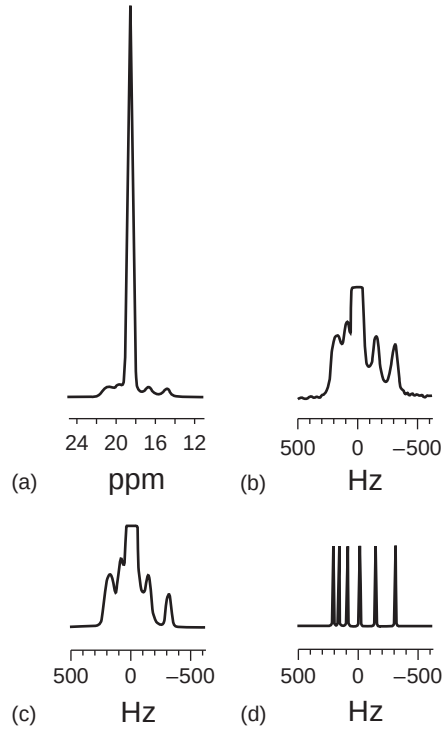

Figure 10 (a) Isotropic region of the ^{31}P CPMAS spectrum of all-*trans*-$Ru(PEt_3)_2(CO)_2(C{\equiv}CPh)_2$ obtained at 81.0 MHz. (b) Expansion of the satellite peaks in (a). (c) Calculated spectrum. (d) Schematic representation of the ^{99}Ru satellites. (Reprinted with permission from Eichele et al.[149] Copyright 1993 American Chemical Society.)

effective way of studying quadrupolar nuclei indirectly. Since the ''residual dipolar couplings'' are proportional to the quadrupolar coupling constant at the quadrupolar nucleus and inversely proportional to the external magnetic field strength, this phenomenon is important in cases where either large quadrupolar interactions are present or the NMR experiments are performed at low magnetic fields.

6 CONCLUDING REMARKS

In this article, we have described some basic principles and applications of solid-state NMR for spin-$\frac{1}{2}$ nuclei other than ^1H and ^{13}C. Emphasis has been placed on the fundamental and practical aspects of solid-state NMR for a group of selected spin-$\frac{1}{2}$ nuclei. The field of solid-state multinuclear NMR spectroscopy is not only diverse, but also rapidly expanding. It is therefore not possible within the limited space to provide a complete account. It is hoped that this article can provide readers with a brief overview of this branch of scientific research and a starting point to the extensive solid-state NMR literature.

ACKNOWLEDGMENTS

Our research in solid-state NMR is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

ABBREVIATIONS AND ACRONYMS

COSY	Correlation Spectroscopy
CP	Cross-polarization
CPMAS	Cross-polarization/Magic-angle Spinning
CSA	Chemical Shift Anisotropy
DFT	Density Functional Theory
ESR	Electron Spin Resonance
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
MAS	Magic-angle Spinning
NMR	Nuclear Magnetic Resonance
OEP	Octaethylporphyrin
Q_8M_8	Cubic Octamer Silicic Acid Trimethylsilyl Ester
R^2	Rotational Resonance
RF	Radiofrequency
TMS	Tetramethylsilane
TPP	Tetraphenylporphyrin
VACP	Variable-amplitude Cross-polarization
2-D	Two-dimensional

RELATED ARTICLES

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Parameters, Calculation of Nuclear Magnetic Resonance • Solid-state Nuclear Magnetic Resonance • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, 1961.
2. U. Haeberlen, *High-resolution NMR in Solids: Selective Averaging*, Academic Press, New York, 1976.

3. M. Mehring, *Principles of High-resolution NMR in Solids*, Springer, Berlin, 1983.
4. C.A. Fyfe, *Solid State NMR for Chemists*, CFC Press, Guelph, Ontario, 1983.
5. B.C. Gerstein, C.R. Dybowski, *Transient Techniques in NMR of Solids: An Introduction to Theory and Practice*, Academic Press, Orlando, FL, 1985.
6. C.P. Slichter, *Principles of Magnetic Resonance*, 3rd edition, Springer, Berlin, 1990.
7. J. Mason (ed.), *Multinuclear NMR*, Plenum Press, New York, 1987.
8. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, New York, 1991.
9. E.O. Stejskal, J.D. Memory, *High-resolution NMR in the Solid State: Fundamentals of CP/MAS*, Oxford University Press, New York, 1994.
10. K. Schmidt-Rohr, H.W. Spiess, *Multidimensional Solid-State NMR and Polymers*, Academic Press, San Diego, CA, 1994.
11. D.M. Grant, R.K. Harris (eds.), *Encyclopedia of Nuclear Magnetic Resonance*, John Wiley & Sons, Chichester, 1996.
12. A. Sebald, 'MAS and CP/MAS NMR of Less Common Spin-$\frac{1}{2}$ Nuclei', *NMR Basic Principles Prog.*, **31**, 91–131 (1994).
13. J. Mason, 'Conventions for the Reporting of Nuclear Magnetic Shielding (or Shift) Tensors Suggested by Participants in the NATO ARW on NMR Shielding Constants at the University of Maryland, College Park, July, 1992', *Solid State Nucl. Magn. Reson.*, **2**, 285–288 (1992).
14. D.L. VanderHart, H.S. Gutowsky, 'Rigid-lattice NMR Moments and Line Shapes with Chemical-shift Anisotropy', *J. Chem. Phys.*, **49**, 261–271 (1968).
15. K.W. Zilm, D.M. Grant, 'Carbon-13 Dipolar Spectroscopy of Small Organic Molecules in Argon Matrices', *J. Am. Chem. Soc.*, **103**, 2913–2922 (1981).
16. T.M. Duncan, *Chemical Shift Tensors*, 2nd edition, Farragut Press, Madison, WI, 1997.
17. G.E. Pake, 'Nuclear Resonance Absorption in Hydrate Crystals: Fine Structure of the Proton Line', *J. Chem. Phys.*, **16**, 327–336 (1948).
18. C.J. Jameson, 'Spin–Spin Coupling', in *Multinuclear NMR*, ed. J. Mason, Plenum Press, New York, 89–131, 1987.
19. A. Nolle, 'Direct and Indirect Dipole–Dipole Coupling Between ^{111}Cd, ^{113}Cd and ^{125}Te in Solid CdTe', *Z. Phys. B*, **34**, 175–182 (1979).
20. M.D. Lumsden, K. Eichele, R.E. Wasylishen, T.S. Cameron, J.F. Britten, 'Determination of a ^{199}Hg–^{31}P Indirect Spin–Spin Coupling Tensor via Single-crystal Phosphorus NMR Spectroscopy', *J. Am. Chem. Soc.*, **116**, 11129–11136 (1994).
21. M.D. Lumsden, R.E. Wasylishen, J.F. Britten, 'Anisotropy in the ^{199}Hg–^{31}P Indirect Spin–Spin Coupling Tensor of a 1:2 Mercury–Phosphine Complex. A Phosphorus Single-crystal NMR Study', *J. Phys. Chem.*, **99**, 16602–16608 (1995).
22. U. Haubenreisser, U. Sternberg, A.-R. Grimmer, 'High-field ^{31}P NMR Investigations of the Chemical Shielding and Indirect Dipolar Coupling of Polycrystalline Fluorophosphates', *Mol. Phys.*, **60**, 151–163 (1987).
23. G.H. Penner, W.P. Power, R.E. Wasylishen, 'Anisotropy of Indirect Spin–Spin Coupling Constants from Nuclear Magnetic Resonance Powder Patterns of Rigid Solids. ^1J(P-31, Hg-199) in $[HgP(o\text{-tolyl})_3(NO_3)_2]_2$', *Can. J. Chem.*, **66**, 1821–1823 (1988).
24. W.P. Power, M.D. Lumsden, R.E. Wasylishen, 'Anisotropic Indirect Spin–Spin Coupling in the Solid State. Experimental Evidence of Noncontact Contributions to $^1J(^{31}P,^{199}Hg)$', *J. Am. Chem. Soc.*, **113**, 8257–8262 (1991).
25. R.E. Wasylishen, K.C. Wright, K. Eichele, T.S. Cameron, 'Characterization of the $J(^{115}In,^{31}P)$ Tensor for a 1:1 Adduct of Indium Tribromide and a Triarylphosphine', *Inorg. Chem.*, **33**, 407–408 (1994).
26. W.P. Power, R.E. Wasylishen, 'Anisotropies of the ^{31}P Chemical Shift and ^{31}P–^{195}Pt Indirect Spin–Spin Coupling in Platinum(II) Phosphines', *Inorg. Chem.*, **31**, 2176–2183 (1992).
27. G. Wu, R.E. Wasylishen, 'Anisotropic Nuclear Spin Interactions in $K_2Hg(CN)_4$: A Multinuclear Solid State NMR Study', *J. Phys. Chem.*, **97**, 7863–7869 (1993).
28. P. Kempgens, J. Hirschinger, K. Elbayed, J. Raya, P. Granger, J. Rosé, 'Multinuclear NMR Study of $HFeCo_3(CO)_9[P(OCH_3)_3]_3$ in the Solid State and in Solution', *J. Phys. Chem.*, **100**, 2045–2052 (1996).
29. J.C. Cherryman, R.K. Harris, 'New Experiments on Solid Organotin Fluorides', *J. Magn. Reson.*, **128**, 21–29 (1997).
30. C. Marichal, A. Sebald, 'Anisotropy of $^2J(^{119}Sn,^{117}Sn)$ Determined by Off-magic-angle Spinning ^{119}Sn NMR', *Chem. Phys. Lett.*, **286**, 298–304 (1998).
31. P.N. Tutunjian, J.S. Waugh, 'Electron-coupled Spin–Spin Coupling Tensor for ^{31}P in Tetraethyldiphosphine Disulfide', *J. Chem. Phys.*, **76**, 1223–1226 (1982).
32. K. Eichele, G. Wu, R.E. Wasylishen, J.F. Britten, 'Phosphorus-31 NMR Studies of Solid Tetraethyldiphosphine Disulfide. A Reinvestigation of the ^{31}P,^{31}P Spin–Spin Coupling Tensor', *J. Phys. Chem.*, **99**, 1030–1037 (1995).
33. J. Jokisaari, Y. Hiltunen, J. Lounila, 'Methyl Fluoride-^{13}C in Nematic Liquid Crystals: Anisotropy of the Indirect ^{13}C–^{19}F Spin–Spin Coupling and of the ^{1}H,^{13}C, and ^{19}F Chemical Shieldings', *J. Chem. Phys.*, **85**, 3198–3210 (1986).
34. A. Pulkkinen, J. Jokisaari, Y. Hiltunen, 'Dimethylmercury in Mixtures of Thermotropic Nematogens', *Liq. Cryst.*, **3**, 737–744 (1988).
35. E.R. Andrew, A. Bradbury, R.G. Eades, 'Nuclear Magnetic Resonance Spectra from a Crystal Rotated at High Speed', *Nature*, **182**, 1659 (1958).

36. E.R. Andrew, A. Bradbury, R.G. Eades, 'Removal of Dipolar Broadening of Nuclear Magnetic Resonance Spectra of Solids by Specimen Rotation', *Nature*, **183**, 1802–1803 (1959).
37. I.J. Lowe, 'Free Induction Decays of Rotating Solids', *Phys. Rev. Lett.*, **2**, 285–287 (1959).
38. S.R. Hartmann, E.L. Hahn, 'Nuclear Double Resonance in the Rotating Frame', *Phys. Rev.*, **128**, 2042–2053 (1962).
39. A. Pines, G.M. Gibby, J.S. Waugh, 'Proton-enhanced Nuclear Induction Spectroscopy. ^{13}C Chemical Shielding Anisotropy in Some Organic Solids', *Chem. Phys. Lett.*, **15**, 373–376 (1972).
40. A. Pines, G.M. Gibby, J.S. Waugh, 'Proton-enhanced Nuclear Induction Spectroscopy. A Method for High Resolution NMR of Dilute Spins in Solids', *J. Chem. Phys.*, **56**, 1776–1777 (1972).
41. K.E. Russell, G. Wu, S. Blake, R.D. Heyding, '^{13}C NMR and X-ray Diffraction Studies of the Morphology of Alkanes and Linear Polyethylenes', *Polymer*, **33**, 951–957 (1992).
42. J. Schaefer, E.O. Stejskal, 'Carbon-13 Nuclear Magnetic Resonance of Polymers at the Magic Angle', *J. Am. Chem. Soc.*, **98**, 1031–1032 (1976).
43. M.A. Kennedy, P.D. Ellis, *Concepts Magn. Reson.*, **1**, 35–47, 109–129 (1989).
44. T. Vosegaard, V. Langer, P. Daugaard, E. Hald, H. Bildsøe, H.J. Jakobsen, 'A New Goniometer Design for Single-crystal Nuclear Magnetic Resonance Spectroscopy', *Rev. Sci. Instrum.*, **67**, 2130–2133 (1996).
45. J.M. Griffiths, R.G. Griffin, 'Nuclear Magnetic Resonance Methods for Measuring Dipolar Couplings in Rotating Solids', *Anal. Chim. Acta*, **283**, 1081–1101 (1993).
46. A.E. Bennett, R.G. Griffin, S. Vega, 'Recoupling of Homo- and Heteronuclear Dipolar Interactions in Rotating Solids', *NMR Basic Principles Prog.*, **33**, 1–77 (1994).
47. J.R. Garbow, T. Gullion, 'Measurement of Internuclear Distances in Biological Solids by Magic-angle Spinning ^{13}C NMR', in *Carbon-13 NMR Spectroscopy of Biological Systems*, ed. N. Beckmann, Academic Press, San Diego, CA, 65–115, 1995.
48. X. Feng, Y.K. Lee, D. Sandström, M. Edén, H. Maisel, A. Sebald, M.H. Levitt, 'Direct Determination of a Molecular Torsional Angle by Solid-state NMR', *Chem. Phys. Lett.*, **257**, 314–320 (1996).
49. K. Schmidt-Rohr, 'Torsion Angle Determination in Solid ^{13}C-labeled Amino Acids and Peptides by Separated-local-field Double-quantum NMR', *J. Am. Chem. Soc.*, **118**, 7601–7603 (1996).
50. D.P. Weliky, R. Tycko, 'Determination of Peptide Conformations by Two-dimensional Magic Angle Spinning NMR Exchange Spectroscopy with Rotor Synchronization', *J. Am. Chem. Soc.*, **118**, 8487–8488 (1996).
51. M. Hong, J.D. Gross, R.G. Griffin, 'Site-resolved Determination of Peptide Torsion Angle ϕ from the Relative Orientations of Backbone N–H and C–H Bonds by Solid-state NMR', *J. Phys. Chem. B*, **101**, 5869–5874 (1997).
52. A. Shoji, S. Ando, S. Kuroki, I. Ando, G.A. Webb, 'Structural Studies of Peptides and Polypeptides in the Solid State by Nitrogen-15 NMR', *Annu. Rep. NMR Spectrosc.*, **26**, 55–98 (1993).
53. C.J. Jameson, A.K. Jameson, D. Oppusunggu, S. Wille, P.M. Burrell, J. Mason, '^{15}N Nuclear Magnetic Shielding Scale from Gas Phase Studies', *J. Chem. Phys.*, **74**, 81–88 (1981).
54. M.G. Gibby, R.G. Griffin, A. Pines, J.S. Waugh, 'High Resolution NMR of ^{15}N in Solids', *Chem. Phys. Lett.*, **17**, 80–81 (1972).
55. G.S. Harbison, L.W. Jelinski, R.E. Stark, D.A. Torchia, J. Herzfeld, R.G. Griffin, '^{15}N Chemical Shift and ^{15}N–^{13}C Dipolar Tensors for the Peptide Bond in [1-^{13}C]Glycyl[^{15}N]Glycine Hydrochloride Monohydrate', *J. Magn. Reson.*, **60**, 79–82 (1984).
56. R.E. Wasylishen, R.D. Curtis, K. Eichele, M.D. Lumsden, G.H. Penner, W.P. Power, G. Wu, 'Applications of Dipolar NMR Spectroscopy in Characterizing Nitrogen and Phosphorus Shielding Tensors', in *Nuclear Magnetic Shieldings and Molecular Structure*, ed. J.A. Tossell, Kluwer, Dordrecht, 297–314, 1993.
57. P.V. LoGrasso, L.K. Nicholson, T.A. Cross, 'N–H Bond Length Determinations and Implications for the Gramicidin Channel Conformation and Dynamics from ^{15}N–^{1}H Dipolar Interactions', *J. Am. Chem. Soc.*, **111**, 1910–1912 (1989).
58. R.R. Ketchem, W. Hu, T.A. Cross, 'High Resolution Conformation of Gramicidin-A in a Lipid Bilayer by Solid State NMR', *Science*, **261**, 1457–1460 (1993).
59. M.D. Lumsden, G. Wu, R.E. Wasylishen, R.D. Curtis, 'Solid State ^{15}N NMR Studies of the Nitroso Group in the Nitrosobenzene Dimer and *p*-Nitroso-*N*,*N*-dimethylaniline', *J. Am. Chem. Soc.*, **115**, 2825–2832 (1993).
60. C.J. Groombridge, L.F. Larkworthy, J. Mason, 'Swinging of the Bent Nitrosyl Ligand in [Co(^{15}NO)(TPP)]: A Solid State Motion Detected by ^{15}N CPMAS NMR Spectroscopy', *Inorg. Chem.*, **32**, 379–380 (1993).
61. R. Salzmann, M. Wojdelski, M. McMahon, R.H. Havlin, E. Oldfield, 'A Solid-state Nitrogen-15 Nuclear Magnetic Resonance Spectroscopic and Quantum Chemical Investigation of Nitrosoarene–Metal Interactions in Model Systems and in Heme Proteins', *J. Am. Chem. Soc.*, **120**, 1349–1356 (1998).
62. N. Godbout, L.K. Sanders, R. Salzmann, R.H. Havlin, M. Wojdelski, E. Oldfield, 'Solid-state NMR, Mössbauer, Crystallographic, and Density Functional Theory Investigation of Fe–O_2 and Fe–O_2 Analog Metalloporphyrins and Metalloproteins', *J. Am. Chem. Soc.*, **121**, 3829–3844 (1999).

63. M. Witanowski, L. Stefaniak, G.A. Webb, 'Nitrogen NMR Spectroscopy', *Annu. Rep. NMR Spectrosc.*, **25**, 1–480 (1993).
64. G. Engelhardt, D. Michel, *High-resolution Solid-state NMR of Silicates and Zeolites*, Wiley, Chichester, 1987.
65. C.A. Fyfe, Y. Feng, H. Grondey, G.T. Kokotailo, H. Gies, 'One- and Two-dimensional High-resolution Solid-state NMR Studies of Zeolite Lattice Structures', *Chem. Rev.*, **91**, 1525–1543 (1991).
66. J. Klinowski, 'Applications of Solid-state NMR for the Study of Molecular Sieves', *Anal. Chim. Acta*, **283**, 929–965 (1993).
67. C.A. Fyfe, Y. Feng, H. Gies, H. Grondey, G.T. Kokotailo, 'Natural-abundance Two-dimensional Solid-state ^{29}Si NMR Investigations of Three-dimensional Lattice Connectivities in Zeolite Structures', *J. Am. Chem. Soc.*, **112**, 3264–3270 (1990).
68. C.J. Jameson, A. De Dios, A.K. Jameson, 'Absolute Shielding Scale for ^{31}P from Gas-phase NMR Studies', *Chem. Phys. Lett.*, **167**, 575–583 (1990).
69. G. Huttner, 'Zur Ursache der Extremen ^{31}P-NMR-Verschiebungen von Phosphinidenkomplexen', *J. Organomet. Chem.*, **308**, C11–C13 (1986).
70. C.E. Laplaza, W.M. Davis, C.C. Cummins, 'A Molybdenum–Phosphorus Triple Bond: Synthesis, Structure, and Reactivity of the Terminal Phosphido (P^{3-}) Complex [Mo(P)$(NRAr)_3$]', *Angew. Chem., Int. Ed. Engl.*, **34**, 2042–2044 (1995).
71. N.C. Zanetti, R.R. Schrock, W.M. Davis, 'Monomeric Molybdenum and Tungsten Complexes That Contain a Metal–Phosphorus Triple Bond', *Angew. Chem., Int. Ed. Engl.*, **34**, 2044–2046 (1995).
72. J.W. Diesveld, E.M. Menger, H.T. Edzes, W.S. Veeman, 'High Resolution Solid State Phosphorus-31 Nuclear Magnetic Resonance of Some Triphenylphosphine Transition-metal Complexes', *J. Am. Chem. Soc.*, **102**, 7935–7936 (1980).
73. G.E. Maciel, D.J. O'Donnell, R. Greavs, '^{31}P NMR Comparisons of the Crystalline and Solution States of Rhodium(I) Diphosphine Catalysts', *Adv. Chem. Ser.*, **196**, 389–408 (1982).
74. J.A. Davies, S. Dutremez, 'Solid State ^{31}P NMR Spectroscopic Studies of Tertiary Phosphines and their Complexes', *Coord. Chem. Rev.*, **114**, 61–103 (1992).
75. G. Wu, R.E. Wasylishen, 'Homonuclear ^{31}P *J*-resolved 2D Spectra of Rhodium(I) Phosphine Complexes in the Solid State', *Inorg. Chem.*, **31**, 145–148 (1992).
76. G. Wu, R.E. Wasylishen, 'Application of Two-dimensional ^{31}P CP/MAS NMR Techniques for Studying Metal Phosphine Complexes in the Solid State', *Organometellics*, **11**, 3242–3248 (1992).
77. M. Feike, R. Graf, I. Schnell, C. Jäger, H.W. Spiess, 'Structure of Crystalline Phosphates from ^{31}P Double-quantum NMR Spectroscopy', *J. Am. Chem. Soc.*, **118**, 9631–9634 (1996).
78. J. Raymonda, W. Klemperer, 'Molecular Beam Electric Resonance Spectrum of $^{31}P^{14}N$', *J. Chem. Phys.*, **55**, 232–233 (1971).
79. G. Wu, D. Rovnyak, M.J.A. Johnson, N.C. Zanetti, D.G. Musaev, K. Morokuma, R.R. Schrock, R.G. Griffin, C.C. Cummins, 'Unusual ^{31}P Chemical Shielding Tensors in Terminal Phosphido Complexes Containing a Phosphorus–Metal Triple Bond', *J. Am. Chem. Soc.*, **118**, 10654–10655 (1996).
80. J.C. Duchamp, M. Pakulski, A.H. Cowley, K.W. Zilm, 'Nature of the Carbon–Phosphorus Double Bond and the Carbon–Phosphorus Triple Bond as Studied by Solid State NMR', *J. Am. Chem. Soc.*, **112**, 6803–6809 (1990).
81. R.D. Curtis, M.J. Schriver, R.E. Wasylishen, 'Characterization of a Novel Triply Bonded Iminophosphenium Cation by Nitrogen-15 and Phosphorus-31 Solid State NMR Spectroscopy', *J. Am. Chem. Soc.*, **113**, 1493–1498 (1991).
82. A. Naito, D.L. Sastry, C.A. McDowell, 'Determination of ^{31}P Chemical Shielding Tensors from NMR Studies on Single Crystals of Chlorotris(triphenylphosphine) Rhodium(I)', *Chem. Phys. Lett.*, **115**, 19–23 (1985).
83. K. Eichele, R.E. Wasylishen, J.F. Corrigan, N.J. Taylor, A.J. Carty, 'Phosphorus-31 Chemical Shift Tensors of Phosphinidene Ligands in Ruthenium Carbonyl Cluster Compounds: A ^{31}P Single-crystal and CPMAS-NMR Study', *J. Am. Chem. Soc.*, **117**, 6961–6969 (1995).
84. K. Eichele, R.E. Wasylishen, K. Maitra, J.H. Nelson, J.F. Britten, 'Single-crystal ^{31}P NMR and X-ray Diffraction Study of a Molybdenum Phosphine Complex: (5-Methyldibenzophosphole)pentacarbonylmolybdenum(0)', *Inorg. Chem.*, **36**, 3539–3544 (1997).
85. E.R. Andrew, A. Bradbury, R.G. Eades, V.T. Wynn, 'Nuclear Cross-relaxation Induced by Specimen Rotation', *Phys. Lett.*, **4**, 99–100 (1963).
86. D.P. Raleigh, M.H. Levitt, R.G. Griffin, 'Rotational Resonance in Solid State NMR', *Chem. Phys. Lett.*, **146**, 71–76 (1988).
87. D.P. Raleigh, F. Creuzet, S.K. Das Gupta, M.H. Levitt, R.G. Griffin, 'Measurement of Internuclear Distances in Polycrystalline Solids: Rotational Enhanced Transfer of Nuclear Magnetization', *J. Am. Chem. Soc.*, **111**, 4502–4503 (1989).
88. R. Challoner, T. Nakai, C.A. McDowell, 'J Coupling in Chemically Equivalent Spin Pairs as Studied by Solid-state Nuclear Magnetic Resonance', *J. Chem. Phys.*, **94**, 7038–7045 (1991).
89. K. Eichele, G. Wu, R.E. Wasylishen, 'Unusual AB Spectra in High Resolution Magic-angle-spinning NMR Studies of Solids', *J. Magn. Reson., Ser. A*, **101**, 157–161 (1993).
90. G. Wu, R.E. Wasylishen, '*J*-Recoupling Patterns Arising from Two Chemically Equivalent Nuclear Spins in

Magic-angle Spinning Spectra of Solids', *J. Chem. Phys.*, **98**, 6138–6149 (1993).
91. M.J. Collins, C.I. Ratcliffe, J.A. Ripmeester, 'CP/MAS ^{77}Se NMR in Solids. Chemical Shift Tensors and Isotropic Shifts', *J. Magn. Reson.*, **68**, 172–179 (1986).
92. G. Wu, R.E. Wasylishen, unpublished data.
93. P.J. Barrie, R.J.H. Clark, R. Withnall, D.-Y. Chung, K.-W. Kim, M.G. Kanatzidis, '^{77}Se Solid State NMR Studies of $[M(Se_4)_2]^{2-}$ Anions (M = Zn, Cd, Hg)', *Inorg. Chem.*, **33**, 1212–1216 (1994).
94. J. Jokisaari, P. Lazzeretti, P. Pyykkö, 'NMR Chemical Shifts of CX and XCY Molecules (X,Y = O, S, Se, Te): A Comparison of Coupled Hartree–Fock, Semi-empirical REX and Experimental Results', *Chem. Phys.*, **123**, 339–350 (1988).
95. H. Duddeck, 'Selenium-77 Nuclear Magnetic Resonance Spectroscopy', *Prog. Nucl. Magn. Reson. Spectrosc.*, **27**, 1–323 (1995).
96. P.G. Mennitt, M.P. Shatlock, V.J. Bartuska, G.E. Maciel, '^{113}Cd NMR Studies of Solid Cadmium(II) Complexes', *J. Phys. Chem.*, **85**, 2087–2091 (1981).
97. M.A. Kennedy, P.D. Ellis, H.J. Jakobsen, '^{113}Cd Shielding Tensors of Cadmium Compounds. 6. Single Crystal NMR Study of the $CdCl_2 \cdot$ 18-crown-6-complex', *Inorg. Chem.*, **29**, 550–552 (1990).
98. M.F. Richardson, 'A Note on Symmetry and the Single-crystal NMR Shift Tensor of the $CdCl_2 \cdot$ 18-crown-6-complex', *Inorg. Chem.*, **29**, 3620–3621 (1990).
99. I.M. Armitage, R.T. Pajer, A.J.M.S. Uiterkamp, J.F. Chlebowski, J.E. Coleman, 'Cadmium-113 Fourier Transform Nuclear Magnetic Resonance of Cadmium(II) Carbonic Anhydrases and Cadmium(II) Alkaline Phosphatase', *J. Am. Chem. Soc.*, **98**, 5710–5712 (1976).
100. P.D. Ellis, 'Cadmium-113 Magnetic Resonance Spectroscopy', *Science*, **221**, 1141–1146 (1983).
101. M.F. Summers, '^{113}Cd NMR Spectroscopy of Coordination Compounds and Proteins', *Coord. Chem. Rev.*, **86**, 43–134 (1988).
102. J.E. Coleman, 'Cadmium-113 Nuclear Magnetic Resonance Applied to Metalloproteins', *Methods Enzymol.*, **227**, 16–43 (1993).
103. K. Eichele, R.E. Wasylishen, 'High Resolution ^{113}Cd MAS NMR Investigation of Structure and Bonding in Cadmium Thiocyanate Coordination Compounds. Distance Dependence of Cadmium–Nitrogen Indirect Spin–Spin Coupling Constants', *Inorg. Chem.*, **33**, 2766–2773 (1994).
104. H.J. Jakobsen, P.D. Ellis, R.R. Inners, C.F. Jensen, 'Multinuclear, Multifield, and Multiphase Nuclear Magnetic Resonance Study of Cadmium *meso*-Tetraphenylporphyrin and Its Pyridine Adduct', *J. Am. Chem. Soc.*, **104**, 7442–7452 (1982).
105. J.D. Kennedy, W. McFarlane, 'Studies of ^{199}Hg Nuclear Shielding Anisotropies and their Relation to Isotropic Chemical Shifts', *J. Chem. Soc., Faraday Trans. 2*, **72**, 1653–1662 (1976).
106. M.B. Blayney, J.S. Winn, D.W. Nierenberg, 'Handling Dimethylmercury', *Chem. Eng. News*, **75** (May 12), 7 (1997).
107. D. Live, 'Hazardous References', *Chem. Eng. News*, **75** (July 14), 7 (1997).
108. R.K. Harris, A. Sebald, 'Experimental Methodology for High-resolution Solid-state NMR of Heavy-metal Spin-$\frac{1}{2}$ Nuclei', *Magn. Reson. Chem.*, **25**, 1058–1062 (1987).
109. J.M. Hook, P.A.W. Dean, L.C.M. van Gorkom, 'Solid State Mercury-199 NMR of Hexakis(dimethyl sulfoxide)mercury(II) Trifluoromethanesulfonate: A New Standard for Mercury-199 CP/MAS Experiments', *Magn. Reson. Reson.*, **33**, 77–79 (1995).
110. K. Eichele, S. Kroeker, G. Wu, R.E. Wasylishen, 'Set-up Samples for ^{199}Hg Cross-polarization Magic-angle Spinning Nuclear Magnetic Resonance Spectroscopy', *Solid State Nucl. Magn. Reson.*, **4**, 295–300 (1995).
111. R.A. Santos, E.S. Gruff, S.A. Koch, G.S. Harbison, 'Solid-State ^{199}Hg and ^{113}Cd NMR Studies of Mercury–and Cadmium–Thiolate Complexes: Spectroscopic Models for $[Hg(SCys)_n]$ Centers in the Bacterial Mercury Resistance Proteins', *J. Am. Chem. Soc.*, **113**, 469–475 (1991).
112. R.A. Santos, G.S. Harbison, '^{199}Hg Chemical Shift Tensors of the Mercurous Ion in Mercurous Nitrate Dihydrate', *J. Am. Chem. Soc.*, **116**, 3075–3077 (1994).
113. M. Han, O.B. Peersen, J.W. Bryson, T.V. O'Halloran, S.O. Smith, 'Enhanced Cross Polarization in Magic Angle Spinning NMR of Metal Complexes', *Inorg. Chem.*, **34**, 1187–1192 (1995).
114. L.M. Utschig, J.W. Bryson, T.V. O'Halloran, 'Mercury-199 NMR of the Metal Receptor Site in MerR and Its Protein–DNA Complex', *Science*, **268**, 380–385. (1995).
115. A. Laaksonen, R.E. Wasylishen, 'An Absolute Chemical Shielding Scale for Tin from NMR Relaxation Studies and Molecular Dynamics Simulation', *J. Am. Chem. Soc.*, **117**, 392–400 (1995).
116. E.T. Lippmaa, M.A. Alla, T.J. Pehk, G. Engelhardt, 'Solid State High Resolution NMR Spectroscopy of Spin-$\frac{1}{2}$ Nuclei (^{13}C, ^{29}Si, ^{119}Sn) in Organic Compounds', *J. Am. Chem. Soc.*, **100**, 1929–1931 (1978).
117. A. Sebald, 'Solid State NMR Applications in Organotin and Organolead Chemistry', in *Advanced Applications of NMR to Organometallic Chemistry*, eds. M. Gielen, R. Willem, B. Wrackmeyer, Wiley, Chichester, Chapter 5, 1996.
118. K.W. Zilm, G.A. Lawless, R.M. Merrill, J.M. Millar, G.G. Webb, 'Nature of the Tin–Tin Double Bond As Studied by Solid State and Solution Nuclear Magnetic Resonance', *J. Am. Chem. Soc.*, **109**, 7236–7238 (1987).
119. R.S. Deeb, D.H. Peyton, '^{1}H and ^{119}Sn Magnetic Resonance Study of the Sn^{IV} Protoporphyrin IX Complex of Equine Myoglobin. Structure of the Porphyrin-binding Pocket', *J. Biol. Chem.*, **266**, 3728–3733 (1991).

120. J.M. Martins, F. Kayser, P. Verheyden, M. Gielen, R. Willem, M. Biesemans, 'Optimal Detection of Weak $^{n}J(^{1}H-^{119}Sn)$ Couplings by Gradient-enhanced 1D and 2D Heteronuclear Multiple-quantum Correlation Spectroscopy. Application to a Novel Tin Derivative of Erythromycin A', *J. Magn. Reson.*, **124**, 218–222 (1997).
121. H.J. Keller, H.H. Rupp, 'Anisotropie der ^{195}Pt-KMR-Verschiebungen Kristalliner Platin(II)-Verbindungen', *Z. Naturforsch., Teil A*, **26**, 785–786 (1971).
122. S.W. Sparks, P.D. Ellis, '^{195}Pt Shielding Tensors in Potassium Hexachloroplatinate(IV) and Potassium Tetrachloroplatinate(II)', *J. Am. Chem. Soc.*, **108**, 3215–3218 (1986).
123. R.K. Harris, P. Reams, K.J. Packer, 'High Resolution Solid State Platinum-195 Nuclear Magnetic Resonance', *J. Chem. Soc., Dalton Trans.*, 1015–1020 (1986).
124. S. Hayashi, K. Hayamizu, 'High Resolution Solid State ^{195}Pt Nuclear Magnetic Resonance in Hexachloroplatinates(IV)', *Magn. Reson. Chem.*, **30**, 658–661 (1992).
125. J.R. Ascenso, R.K. Harris, P. Granger, 'High Resolution ^{207}Pb NMR of Solid Organolead Compounds', *J. Organomet. Chem.*, **301**, C23–C26 (1986).
126. A. Sebald, R.K. Harris, '^{207}Pb CP MAS NMR Study of Hexaorganyldiplumbanes', *Organometallics*, **9**, 2096–2100 (1990).
127. L.C.M. van Gorkom, J.M. Hook, M.B. Logan, J.V. Hanna, R.E. Wasylishen, 'Solid State Lead-207 NMR of Lead(II) Nitrate: Localized Heating Effects at High Magic Angle Spinning Speeds', *Magn. Reson. Chem.*, **33**, 791–795 (1995).
128. G. Neue, C. Dybowski, 'Determining Temperature in a Magic-angle-spinning Probe Using the Temperature-dependence of the Isotropic Chemical Shift of Lead Nitrate', *Solid State Nucl. Magn. Reson.*, **7**, 333–336 (1997).
129. G. Neue, C. Dybowski, M.L. Smith, M.A. Hepp, D.L. Perry, 'Determination of Pb207(2+) Chemical Shift Tensors from Precise Lineshape Analysis', *Solid State Nucl. Magn. Reson.*, **6**, 241–250 (1996).
130. C. Dybowski, M.L. Smith, M.A. Hepp, E.J. Gaffney, G. Neue, D.L. Perry, 'Pb-207 NMR Chemical Shift Tensors of the Lead(II) Halides and the Lead(II) Hydroxyhalides', *Appl. Spectrosc.*, **52**, 426–429 (1998).
131. S.P. Gabuda, S.G. Kozlova, V.V. Terskikh, C. Dybowski, G. Neue, D.L. Perry, '^{207}Pb NMR Study of Novel Pb–Pb Chemical Bonding in Lead Monoxides, α-PbO and β-PbO', *Chem. Phys. Lett.*, **305**, 353–358 (1999).
132. P. Amo-Ochoa, V.M. Gonzales, J.M. Perez, J.R. Masaguer, C. Alonso, C. Navarro-Ranninger, 'Cytotoxicity, DNA Binding, and Reactivity against Nucleosides of Platinum(II) and (IV) Spermine Compounds', *J. Inorg. Biochem.*, **64**, 287–299 (1996).
133. J.M. Aramini, T. Hiraoki, M. Yazawa, T. Yuan, M. Zhang, H.J. Vogel, 'Lead-207 NMR: A Novel Probe for the Study of Calcium-Binding Proteins', *J. Biol. Inorg. Chem.*, **1**, 39–48 (1996).
134. B.E. Mann, 'The Cinderella Nuclei', *Annu. Rep. NMR Spectrosc.*, **23**, 141–207 (1991).
135. W. von Philipsborn, 'Probing Organometallic Structure and Reactivity by Transition Metal NMR Spectroscopy', *Chem. Soc. Rev.*, **28**, 95–105 (1999).
136. A.R. Thompson, E. Oldfield, 'Solid-state Scandium-45, Yttrium-89, and Lanthanum-139 Nuclear Magnetic Resonance Spectroscopy', *J. Chem. Soc., Chem. Commun.*, 27–29 (1987).
137. C.T.G. Knight, G.L. Turner, R.J. Kirkpatrick, E. Oldfield, 'Solid-state Tungsten-183 Nuclear Magnetic Resonance Spectroscopy', *J. Am. Chem. Soc.*, **108**, 7426–7427 (1986).
138. J. Chung, H.C. Lee, E. Oldfield, 'Iron-57 Nuclear Magnetic Resonance Spectroscopic Study of Alkyl Isocyanide Myoglobins and a Comparison of the ^{57}Fe Chemical-Shift Anisotropies of Alkyl Isocyanide Myoglobins, Carbonmonoxymyoglobin, Ferrocytochrome *c*, and $[^{57}Fe(bipy)_3]X_2 \cdot 5H_2O$ (X = Cl, Br, I)', *J. Magn. Reson.*, **90**, 148–157 (1990).
139. R. Dupree, M.E. Smith, 'Structural Influence on High Resolution Yttrium-89 NMR Spectra of Solids', *Chem. Phys. Lett.*, **148**, 41–44 (1988).
140. R.E. Walstedt, W.W. Warren, Jr, 'Nuclear Resonance Properties of $YBa_2Cu_3O_{6+x}$ Superconductors', *Science*, **248**, 1082–1087 (1990).
141. L.H. Merwin, A. Sebald, 'The First ^{89}Y CP-MAS Spectra', *J. Magn. Reson.*, **88**, 167–171 (1990).
142. L.H. Merwin, A. Sebald, 'The First Examples of ^{109}Ag CP MAS Spectroscopy', *J. Magn. Reson.*, **97**, 628–631 (1992).
143. L.H. Merwin, A. Sebald, 'Cross Polarization to Low-γ Nuclei: The First ^{183}W CP/MAS Spectra', *Solid State Nucl. Magn. Reson.*, **1**, 45–47 (1992).
144. P.A. Casabella, 'Nuclear Magnetic-dipole Coupling in Solid BF_3', *J. Chem. Phys.*, **41**, 3793–3798 (1964).
145. D.L. VanderHart, H.S. Gutowsky, T.C. Farrar, 'Dipole–Dipole Interactions of a Spin-$\frac{1}{2}$ Nucleus with a Quadrupole-coupled Nucleus', *J. Am. Chem. Soc.*, **89**, 5056–5057 (1967).
146. S.J. Opella, M.H. Frey, T.A. Cross, 'Detection of Individual Carbon Resonances in Solid Proteins', *J. Am. Chem. Soc.*, **101**, 5856–5857 (1979).
147. C.J. Groombridge, R.K. Harris, K.J. Packer, B.J. Say, S.F. Tanner, 'High-resolution ^{13}C NMR Spectra of Solids Nitrogen-containing Compounds', *J. Chem. Soc., Chem. Commun.*, 174–175 (1980).
148. R.K. Harris, A.C. Olivieri, 'Quadrupolar Effects Transferred to Spin-$\frac{1}{2}$ Magic-angle Spinning Spectra of Solids', *Progr. Nucl. Magn. Reson. Spectrosc.*, **24**, 435–456 (1992).
149. K. Eichele, R.E. Wasylishen, J.F. Corrigan, S. Doherty, Y. Sun, A.J. Carty, 'The First Observation of ^{1}J(Ru,P) in ^{31}P CP/MAS NMR Spectra of Solid Ruthenium Compounds', *Inorg. Chem.*, **32**, 121–123 (1993).

Solid-state Nuclear Magnetic Resonance

Rex E. Gerald II
Argonne National Laboratory, Argonne, USA

1 Introduction **2**
1.1 Spatial Dispersion of Atomic Magnetic Moments in Magnetic Fields 2
1.2 Nuclear Spin Flips in Molecular Beams 3
1.3 Early Attempts at Recording Nuclear Magnetic Resonance in Condensed Matter 4
1.4 First Successful Report of Nuclear Magnetic Resonance Phenomenon 4
1.5 Original Nuclear Magnetic Resonance Experiments in the Solid State 6
2 Basic Principles and Theory **7**
2.1 Nuclear Spins and Moments 7
2.2 Actions of Nuclear Spins in Magnetic Fields 8
2.3 Isotropic Nuclear Spin Interactions 9
2.4 Anisotropic Nuclear Spin Interactions 9
2.5 Averaging Spatial Anisotropy 14
2.6 Relaxation of Nuclear Spins 16
3 Solid-state Nuclear Magnetic Resonance Instrumentation **18**
3.1 Magnet System 18
3.2 Spectrometer Console 18
3.3 Nuclear Magnetic Resonance Detectors 19
4 Applications of Nuclear Magnetic Resonance to Solids **21**
4.1 Wideline Nuclear Magnetic Resonance Spectroscopy of Polymers 21
4.2 Mobile Guests in Solid Host Material 22
4.3 Metals 24
5 Perspective and Future Developments **25**
Acknowledgments **26**
Abbreviations and Acronyms **26**
Related Articles **27**
References **27**

The first experimental verification of nuclear magnetic resonance (NMR) in the solid state by Edward M. Purcell, Henry C. Torrey, and Robert V. Pound occurred on Saturday afternoon, December 15, 1945. Since then, new developments have transformed this form of spectroscopy from an interesting physical curiosity at its inception to one of the most widely used analytical tools for physical, chemical, biological, and materials research and applications. The capabilities of NMR include identification of molecules in all phases of matter by their atomic connectivities; probing the temperature-dependent conformation of molecules from cyclohexane to complex 50 kilo-Dalton proteins; molecular dynamics investigations from the picosecond timescale for rotation of small molecules to the second timescale for atoms diffusing throughout the three-dimensional nanoscopic channels of zeolites; elucidation of chemical reaction mechanisms; the determination of organization, reorientation, and diffusion of molecular solids; the identification of phase changes and morphology of solids; the measurement of bond lengths and interatomic distances between nonbonded atoms; the assessment of pore size in nanoporous solids as well as identification of adsorbates and their diffusivities; the estimation of physicochemical parameters of molecular systems such as equilibrium and kinetic rate constants, enthalpies and entropies of a reaction, and activation energies; the capability to follow a metabolic pathway taking place inside living cells; the quantification of the alignment of molecules in partially ordered systems; the determination of nanoscale domain sizes in heterogeneous materials; and the noninvasive imaging of all phases and forms of matter under various temperature, pressure, and mechanical conditions.

The scope of this article is very selective, covering some of the most widely used techniques and more unusual areas of application of solid-state NMR. These topics are covered at different levels of detail, and some are just mentioned by reference to published work. This article is not intended to present a comprehensive review of all state-of-the-art NMR techniques; for these the reader should consult references included in the text. Instead, this article provides a historical background and the fundamental concepts of magnetic resonance that preceded solid-state NMR, a general overview of the basic principles with emphasis on the visualization of anisotropic spin interactions, a review of the most widely used methods for investigating solids by NMR, and practical information to provide a modicum of guidance for application of the methods. There are several redundancies in terminology so that different sections can be read independently.

The references are an important component of the article and should be consulted for more complete and often original accounts of the theory and techniques for the application of NMR to solids. A modest attempt was made to include original and important references, but more

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

important, many were included to provide examples of chemical and physical insights arrived at in whole or in part by application of NMR techniques to the investigation of the problem. For figures and tables, if no reference or acknowledgment is provided, the spectrum or photograph in the figure or the table was previously unpublished.

The article is directed at a general audience of chemical and physical analytical scientists with an interest in NMR. In addition, NMR spectroscopists may find some interesting details of studies on metals and magnetic materials.

1 INTRODUCTION

In its present form, NMR spectroscopy is the application of highly evolved methods for manipulating nuclear spins, the purpose of which is to elicit a quantitative understanding of an atom's local environment and dynamics. NMR is rooted in quantum mechanics. A 1936 report by C.J. Gorter on the attempt to observe the NMR effect was founded on the quantum mechanical concept of energy transitions in two-level systems.[1–3] Gorter's experimental approach to record the first NMR response was based on a thermodynamic measurement. He understood the underlying theory and his estimation of the magnitude of the NMR effect was reasonable. Unfortunately, his attempt failed because the compounds he chose to investigate do not have an efficient pathway to dissipate energy from the system of spins to a calorimeter for measurement. Without an efficient energy pathway, any nuclear spin system will saturate when it is subjected to resonant energy. Without the absorption or emission of energy, the NMR effect cannot be detected.

At the time Gorter was performing these experiments, Isador I. Rabi was developing an interest in the spatial separation of atoms in atomic beams by the attractive and repulsive action of an inhomogeneous magnetic field on atomic magnetic moments that were both electronic and nuclear in origin. After setting up a laboratory of his own, Rabi demonstrated a mechanical method for flipping nuclear spins. The approach was rooted in the spatial quantization of spin angular momentum; however, Rabi used his keen kinematic sense for the behavior of classical angular momentum to develop the idea. He used his insight to envision the first (mechanical) manipulation of the orientation of a nuclear spin. Meanwhile, Gorter continued to report unsuccessful attempts to measure NMR transitions in compounds in the solid state. Following a short visit by Gorter to Rabi's laboratory, the Rabi group demonstrated the first successful NMR effect in a molecular beam experiment. Gorter was an influential factor. Several years later, a student of Rabi's, along with two collaborators, reported the detection of the NMR phenomenon in a solid compound by means of a radiofrequency apparatus. The following sections describe in greater detail the events that lead to the discovery of NMR in condensed matter.

1.1 Spatial Dispersion of Atomic Magnetic Moments in Magnetic Fields

The study of gas phase atomic magnetic moments in inhomogeneous magnetic fields provided the fundamental concepts for the manipulation of nuclear spins. Experiments that probed atomic and nuclear magnetic moments in molecular beams formed the basis for the first successful NMR experiments. These experiments were conducted in the laboratory of I.I. Rabi at Columbia University in the 1930s. The events that initiated Rabi's interest in molecular beams began earlier, in 1927, when he was awarded a Barnard Fellowship (September 22, 1927–June 8, 1928). Upon hearing of his award, Rabi applied for a leave of absence from his teaching duties in physics at the College of the City of New York. His request was denied. Shortly thereafter, he submitted a letter of resignation to the College and left for Europe to advance his understanding of the emerging radical new theories in physics, collectively called the new quantum mechanics. It had been eight years since he graduated with a PhD degree in chemistry from Cornell University, and he had grown restless in his teaching position. Unannounced, Rabi visited Erwin Schrödinger in Zurich, Arnold Sommerfeld in Munich, and Niels Bohr in Copenhagen. Bohr arranged for Rabi and a fellow student to go to Hamburg and work with Wolfgang Pauli.

Otto Stern, also in Hamburg at that time, was preparing a series of papers detailing studies on isolated atoms using his molecular beam method. In December 1927, Rabi suggested to Stern that an atomic beam could be directed into a well-defined homogeneous magnetic field at an oblique angle to separate atomic spin states. This technique could replace the traditional Stern–Gerlach method that relied on atomic deflection in an ill-defined inhomogeneous magnetic field. According to Rabi's idea, the spins would be separated spatially according to their orientation (spin-up or spin-down), just as white light separates into its components as it passes through a homogeneous glass prism. Stern thought Rabi had an unusual but reasonable idea, since an accurate calibration of an inhomogeneous field was difficult to achieve, and this shortcoming limited the accuracy of the determination of the fundamental magnetic properties of atoms. The keys to the Rabi method are the homogeneous field and the oblique angle trajectory. The advantage of the inhomogeneous magnetic field, used by Stern, is that the force is larger on one pole of a magnetic moment, and the spin moves in the direction of the magnetic field

gradient, or in the opposite direction, according to its orientation. A homogeneous field is easier to calibrate, but if the beam of spins were directed in the usual way, head-on, into a homogeneous field, then no splitting of the beam would occur because the magnetic force on the north and south pole of each magnetic moment would be equal and opposite. Rabi overcame this problem by aiming the beam into the field at an oblique angle, which refracted the beam. At Stern's invitation, Rabi set out to demonstrate his idea experimentally by using a beam of potassium atoms, and this later proved to be successful. Rabi's method for deflecting molecular beams was the first use of the homogeneous magnetic field (the Rabi field) that he later used for his famous resonance experiments.

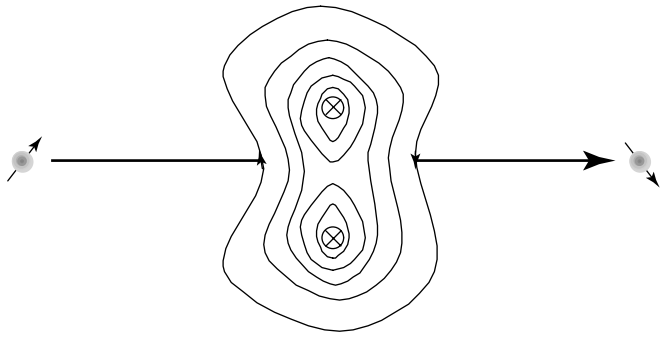

Figure 1 Transition or T-field devised by Rabi for manipulating the spatial orientation of nuclear magnetic moments. Magnetic field lines are generated from two parallel wires carrying direct current into the page. The molecular beam passes from left to right between the wires, where the field direction changes by 180°.

1.2 Nuclear Spin Flips in Molecular Beams

The experiments that separated atoms in space by attracting and repelling their electronic magnet moments led Otto Stern and his coworkers to consider how they might flip the spins in a beamlet that contained atomic spins of only one orientation; Rabi also wondered how he might flip the much weaker nuclear spins in a beamlet of polarized nuclear spins. While in Europe, Rabi was elated to receive an offer from Columbia University, his alma mater, for a lectureship in the physics department. He promptly accepted and within two years had set up his own molecular beam laboratory. For a period of about five years Rabi and his students made many measurements in his ''Rabi field'' to determine the magnetic properties of nuclei. He used the refocusing and zero-moment methods that he and his students developed.[2,4–11]

The transition field or T-field method, also conceived by Rabi, provided the avenue for measuring the sign and hence the sense of precession of a nuclear magnetic moment in an external magnetic field.[3,12–17] Two parallel wires carrying equal direct currents in the same direction created the transition field. A point of uniform magnetic field resulted between the two wires. On either side, away from both wires, the field was equal in intensity, but exactly opposite in direction, as illustrated in Figure 1. A nuclear magnetic moment of selected orientation was directed between the parallel wires and subjected to an abrupt change in the direction of the magnetic field. A transition or reorientation of the spin would take place in the process. The success of the T-field method was the impetus for the subsequent development of the resonance method; it was the drive to determine the sign of the nuclear magnetic moment that led Rabi to propose and use the T-field to manipulate the spatial orientation of a spin.

Prior to Rabi's early molecular beam work at Columbia University, Otto Stern in Hamburg had set out to test the idea of space quantization. Stern attempted to change the orientation of an atomic spin (of a potassium atom) prepared in an inhomogeneous magnetic field by passing it into a second, similar magnetic field, but one oriented in the opposite direction. Stern determined that on slow passage through two sequential magnetic fields of opposing senses (a transition field), the atomic spin maintained the same orientation. However, if the atomic spin was passed through the field inversion rapidly, then the spin did invert its orientation. Rabi later expounded on this effect in his 1936 paper titled ''On the Process of Space Quantization''.[12] A particle traveling through such a T-field would be subjected to a spatially abrupt change in the field direction from one sense to the opposite sense. This abrupt field inversion (or spatial gyration of the field as viewed by the moving particle) would couple to the precessing magnetic moment as it moved through the field. Rapidly moving a particle through the field inversion was equivalent to affecting a spin flip by applying a rotating field to a particle fixed in space. From the perspective of the particle, the effect was the same: the orientation of the nuclear spin was changed. This is the essence of NMR as it is applied today – the reorientation of spins in a controlled manner – but it was originally accomplished in a mechanical fashion and without the use of radiofrequency fields.

In the T-field method, the velocity of the particle and the spatial dimension over which the magnetic field inverted direction were used to calculate the rate of gyration that, in effect, the field would have from the perspective of a precessing magnetic moment in the nominal homogeneous magnetic field. If the frequency of precession of a nuclear magnetic moment was close to the apparent frequency of motion of the gyrating magnetic field, then there would be a strong coupling between the two and the gyrating field would change the orientation

of the nuclear spin. Rabi explained this in his 1937 paper titled ''Space Quantization in a Gyrating Magnetic Field'' in terms of a probability for flipping the spin.[18] A close match between the precession (or Larmor) frequency and the frequency of the gyrating magnetic field resulted in a high probability for nuclear spin reorientation, while a mismatch between the frequencies resulted in a concomitantly lower transition probability. The type of mechanical spin flip performed by Rabi's group, in retrospect, represented the first application of a π (180°) pulse to a nuclear magnetic moment, and it was accomplished without the application of a radiofrequency pulse, as would commonly be done today. The term ''resonance'', adopted later, was a suitable term to describe this interaction because the procedure for reorienting a magnetic moment required that the gyrating magnetic field move in a fashion coincident or resonant with the precessing magnetic moment. The physical intuitions, concepts, and theoretical underpinnings of manipulating the orientation of nuclear spins in magnetic fields were established. The stage was set for the experimental observation of NMR in the condensed phase.

1.3 Early Attempts at Recording Nuclear Magnetic Resonance in Condensed Matter

The first published attempt to measure an NMR response in condensed matter was reported by C.J. Gorter in 1936.[1] This unsuccessful experimental attempt was made on the ^{7}Li isotope of lithium in solid lithium fluoride (LiF) and on the ^{1}H isotope of hydrogen in the solid double salt potassium alum ($KAl(SO_4)_2 \cdot 12H_2O$). The experiments were performed at 20 K to take advantage of the larger Boltzmann magnetization available at low temperatures.[19] Gorter viewed the nuclear spin from the perspective of quantum mechanics. Thus, a nucleus with spin 1/2 will have two different orientations in a static magnetic field. The separation between the two energy levels is proportional to the strength of the magnetic field, and transitions from one energy level to the other occur in the presence of a transverse oscillating magnetic field. The apparatus that Gorter used to search for the NMR effect consisted of a variable field magnet (0–1.2 T) of poor homogeneity (~1% over the sample volume) by modern NMR standards, an oscillator that produced a 20-MHz radiofrequency magnetic field of several gauss, and a calorimeter.

Gorter's idea was to induce nuclear spin transitions with the radiofrequency field; following excitation, the spins would then relax back to a thermal distribution of orientations, and the energy given off by the spins to the surroundings (the lattice) would cause the temperature of the calorimeter to rise. The spins, in effect, would act as a conduit for transferring radiofrequency energy from the oscillator to the calorimeter. (A direct transfer of energy from the oscillator to the calorimeter occurred at a constant rate, and was taken into account.)

Gorter surmised that his experimental attempt to measure the NMR effect failed because the rate of energy transfer from the nuclear spins (in the excited state) to the lattice was too slow. That is, the spin-lattice relaxation rate was so small that the spin system saturated, and sufficient energy could not be transferred from the oscillator to the calorimeter, via the nuclear spins, in a short enough time to record a temperature change above the background level. Gorter also pointed out later that the experimental failure could be accounted for by viewing the results as a nuclear response with an inhomogeneous linewidth on the order of 10^6 Hz.[19] Presumably, the broad distribution of resonances and the slow modulation of the magnetic field (that was used to bring the nuclear spins into resonance) were not optimized for a measurable rate of energy transfer from the oscillator, via the nuclear spins, to the calorimeter. Thus, prior to Rabi's conceptual and later actual manipulation of spins, described by classic torque action, radiofrequency magnetic fields were applied to nuclear spins to affect incoherent transitions.

In September 1937, Gorter visited the physics department at Columbia University. Rabi was aware of his work in condensed matter physics, and after they discussed Rabi's mechanical concepts of the coupling of a gyrating magnetic field with a precessing nuclear magnetic moment, Gorter asked why Rabi's group was not using a radiofrequency magnetic field to reorient the nuclear spins. Rabi began to sense that he had a competitor. Hence, in addition to the continuing productive activity by Rabi's group with the refocusing effective zero-moment, and T-field methods for studying nuclear magnetic moments, Gorter's visit provided the stimulus for Rabi to develop the experimental apparatus required to attempt the resonance method.

1.4 First Successful Report of Nuclear Magnetic Resonance Phenomenon

Isador Rabi, Polykarp Kusch, Jerrold Zacharias, Jerome Kellogg, and Sidney Millman modified a molecular beam apparatus to accommodate a homogeneous magnetic field and radiofrequency coil for an attempt at observing the resonance effect in a molecular beam. They used a beam of lithium chloride (LiCl) molecules, and the resonance condition was set for ^{7}Li nuclei in a magnetic field of 0.21 T. In January 1938, they conducted the first successful experiment that demonstrated the resonance of a nuclear magnetic moment with a radiofrequency magnetic field. Their apparatus for this experiment consisted of three in-line magnets contained in a high-vacuum chamber. While

traversing the first inhomogeneous field, the beam of LiCl molecules was separated into four beamlets: two spin-up and two spin-down. (A ^{7}Li nucleus has a spin angular momentum quantum number, I, of 3/2.) The theory of quantum mechanics states that a particle with $I = 3/2$ can only have $2I + 1 = 4$ discrete orientations in space (space quantization). The discrete orientations are with respect to the axis of quantization, which is determined by the direction of a force field (in this case, the magnetic field). The orientation of a ^{7}Li nucleus in a magnetic field is given by its magnetic quantum number. Each beamlet contained LiCl molecules with the same magnetic quantum number for the ^{7}Li nucleus: +3/2 (spin-up), +1/2 (spin-up), −1/2 (spin-down), −3/2 (spin-down). The beamlets then passed through a hairpin-type radiofrequency coil that was located in the center of a second, homogeneous field. As long as the coil was not energized, the nuclear magnetic moments maintained their orientation on passing through this homogeneous field. The beamlets then passed through the third field, which was inhomogeneous, like the first, but oriented in the opposite direction, so that the beam would be refocused onto a detector. Provided the radiofrequency coil was not energized, the molecular beam diverged in the first field, traversed the second field unchanged, converged in the third field, and arrived at the detector intact. Under these conditions, the detector recorded a maximum beam signal. The radiofrequency coil was then energized at a constant frequency of 3.518 MHz, and the strength of the static, homogeneous magnetic field surrounding the coil was adjusted so that the frequency of precession of the nuclear magnetic moment of the ^{7}Li nuclei in the LiCl molecules would match the frequency of the oscillating magnetic field. This was considered to be the condition for mutual resonance between the precession of the magnetic moment and the rotation or gyration of the radiofrequency magnetic field. The condition of resonance caused a torque to act on the nuclear spin angular momentum, resulting in a reorientation or flip of the nuclear magnetic moment. Upon passing through the first (inhomogeneous) magnetic field, the spin-up nuclear magnetic moments were diverted to one side, while the spin-down moments were diverted to the opposite side. The third magnetic field, which was also inhomogeneous, turned each beam back towards the center. However, when each beamlet traveled through the homogeneous field with the radiofrequency magnetic field in operation at the resonance frequency of the ^{7}Li spins, spin-up nuclear magnetic moments traveling to the left suddenly flipped and became spin-down moments on the same course. On entering the refocusing (inverted) inhomogeneous magnetic field, this beam was further deflected to the left and not refocused. The analogous process occurred for the spin-down nuclear magnetic moments, and they too diverged and were not refocused. The detector, under the resonance condition, showed a dramatic reduction in beam signal output, as shown in Figure 2.

This success represented the first manipulation of nuclear spins in a homogeneous magnetic field by a rotating magnetic field that was generated by a radiofrequency oscillator.[20] It demonstrated (from the perspective of the nucleus) the fundamental concept of applying a torque to reorient a nuclear magnetic moment using a radiofrequency magnetic field. This accomplishment stands as a very significant step toward NMR experiments on condensed matter. The elements of a static homogeneous magnetic field and a radiofrequency magnetic field provided the avenue to leave the arena of rarefied matter and the spatial selection of unity polarization and enter the arena of condensed matter where nuclear polarization was determined by a Boltzmann distribution of spin orientations. Rabi's research efforts would continue in the gas phase, but the stage was set for NMR of solids.

A second published report by Gorter and Broer in 1942 indicated a subsequent unsuccessful attempt to record ^{7}Li NMR of lithium chloride (LiCl) and ^{19}F NMR of potassium fluoride (KF). The failed attempts were attributed to insufficient development of an initial Boltzmann nuclear polarization in a reasonable time or to a saturation of the nuclear magnetization during the experiment. That is, the nuclear spin-lattice relaxation rates for the pure substances at low temperature may have been exceedingly small.[19,21,22] Relaxation of nuclear magnetic moments in external magnetic fields was a key question at the time when the first NMR experiments in the condensed matter were attempted. Gorter discovered paramagnetic relaxation and was keenly aware of the

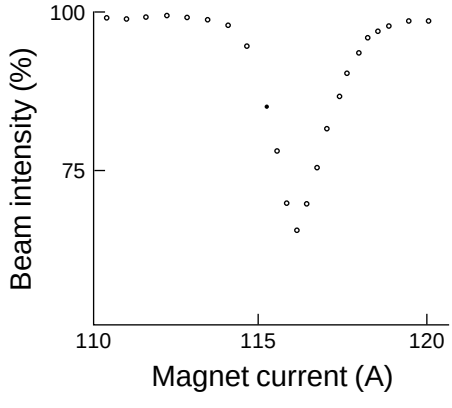

Figure 2 The decrease in beam intensity metered by the detector demonstrated the resonance condition for the ^{7}Li nuclei in a beam composed of isolated LiCl molecules. [Adapted from I.I. Rabi, J.R. Zacharias, S. Millman, P. Kusch,[20] 'A New Method of Measuring Nuclear Magnetic Moment', *Phys. Rev.*, **53**, 318 (1938).]

Figure 3 Professor C.J. Gorter, who made many attempts to detect NMR phenomena in the solid state, published several theoretical and experimental articles on NMR a decade before the first successful confirmation of NMR in condensed matter. [Adapted from C. Gorter,[25] 'Bad Luck in Attempts to Make Scientific Discoveries', *Physics Today*, **20**(1), 76–81 (1967).]

potential problems that small relaxation rates for nuclear magnetic moments could cause for the observation of the NMR phenomenon.[1,23,24] In consideration of his broad understanding of low-temperature physics and of paramagnetism in particular, and his many attempts to detect the NMR effect, Gorter argued that simple bad luck played a role in his misfortunes in the initial observation of NMR in condensed matter.[25] The preponderance of evidence indicates that Gorter was searching for NMR phenomena in condensed matter for about 14 years prior to the first successful reports. It is unfortunate that his contributions to NMR are not widely recognized. A photograph of Professor Gorter is shown in Figure 3.

1.5 Original Nuclear Magnetic Resonance Experiments in the Solid State

The first successful NMR experiments in a solid material, namely paraffin wax (a mixture of solid alkanes, $C_{23}-C_{29}$, with a melting point in the range 50–60 °C), was reported by Edward Purcell, Henry Torrey, and Robert Pound just after the end of World War II in 1945.[26] Independently, F. Bloch, W. Hansen, and M. Packard reported the first successful NMR experiments in a liquid, namely water.[27] Both groups recorded NMRs for the ^{1}H isotope of hydrogen. It is interesting to note that the suggestion by Edward Purcell that the resonant absorption of radiofrequency energy might be detected by nuclear magnetic moments when placed in an external magnetic field came about because of the confluence of scientists at the Massachusetts Institute of Technology Radiation Laboratory for war-related research on radar, the tedious task of documenting the radar work, and the lack of access to previously published reports of unsuccessful attempts by C.J. Gorter in German-occupied Holland.

Prior to working at the Radiation Laboratory, Henry Torrey was a graduate student of I.I. Rabi. He measured the sign of the nuclear magnetic moment of ^{23}Na and ^{39}K using the transition field method.[14] Rabi had won the Nobel Prize one year earlier (physics, 1944) for his molecular beam resonance method that was used to study nuclear magnetic moments. He and his students were renowned in Europe and the United States. Therefore, Torrey was an appropriate person for Purcell to approach with such an idea.

The experiment was conducted at nearby Harvard University because the magnet that would produce the first observed NMR in condensed matter was located there, and its owner, J. Curry Street, offered use of it to the trio. Robert Pound recounted that the detection of resonance absorption by nuclear spins was initially missed because the iron core of the magnet saturated unexpectedly and a magnetic field calibration error resulted.[28]

The detector that was used was not a Helmholtz or solenoid coil, although these types of coil are almost exclusively used in modern NMR probes for small solid and liquid samples. Instead, having spent the previous five years on microwave technology, Purcell and co-workers constructed a modified microwave cavity resonator that was designed to operate at the macrowavelength of 10 m.[26,29] Operation at this radiofrequency wavelength was dictated by the predicted resonance frequency for

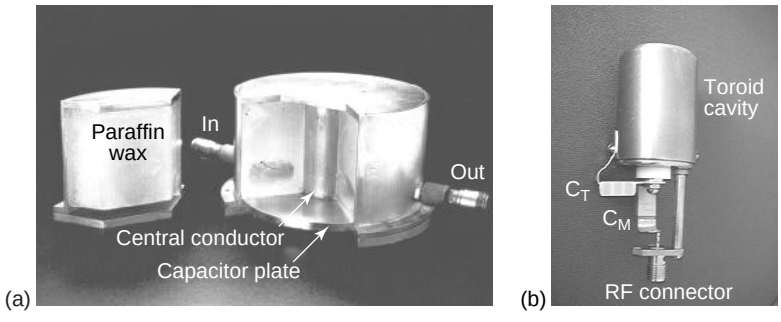

Figure 4 (a) Capacitively shortened, mica-dielectric coaxial cavity resonator filled with paraffin wax. (b) Toroid cavity detector used by the author. [(a) Adapted from R.V. Pound.[28]]

the nuclei of protons in paraffin wax when immersed in a magnetic field of 7 kG, such that the resonance absorption would be at a frequency of approximately 30 MHz. Their capacitively shortened, mica-dielectric coaxial cavity resonator is shown in Figure 4. It is similar to the so-called toroid cavity detector used by the author.[30,31]

Coaxial or toroid cavity NMR detector devices are quite simple in construction and consist of a metal can, usually made of copper, with a central conductor of copper wire. The cavity was used as a device to contain the sample and couple the resonant energy to the nuclear spins. Coupling energy into and out of the resonator is achieved by several methods. A set of two magnetic coupling loops (one for input, the other for output) can be inserted in the wall of the cavity, as Purcell and co-workers did. Another approach couples the radiofrequency energy into the cavity via a loading capacitor located at one end of the resonator; detection of the nuclear induction proceeds through the same port. The NMR effect in a two-pound sample of paraffin wax was demonstrated by the observation of a decrease in the output of a magnetic coupling loop, located opposite the loop used to energize the cavity, when at the specified magnetic field strength the proton nuclear spins absorbed energy. Thus, the energy detected in the cavity was attenuated because it was absorbed by the nuclear spins.

The proton NMR experiments conducted on paraffin wax demonstrated that the nuclear spins absorbed radiofrequency energy that was present in the cavity. The condition necessary for energy absorption by the sample was that the magnetic field intensity split the energy levels occupied by the nuclear magnetic moments of the hydrogen nuclei to an energy separation equal in value to the energy that was resonant in the cavity. However, this NMR experiment could be performed in a different manner. Instead of detecting the absorption of energy, the resonant emission of energy could be detected from the nuclear spins following an impulse of coherent excitation using radiofrequency energy. This approach, applied with a variety of different inductors, including Helmholtz and solenoid coils, single loops, Litz coils, and cavities, is the most common way that NMR experiments are performed today.

The resonator used by Purcell, Torrey, and Pound produced radiofrequency magnetic flux in concentric circles. This circular field geometry is very different from the unidirectional fields produced in the central volume of Helmholtz and solenoid coils. Coherent movement (flow) of nuclear spins in such a cavity produces NMR spectra with unusual properties, as shown recently by Gerald et al.[32] It is not uncommon that revisiting old approaches with modern equipment can lead to interesting and novel ideas and applications.

2 BASIC PRINCIPLES AND THEORY

NMR spectroscopy provides direct energy measurements of the interactions between electromagnetic fields and nuclei with spin. All nuclei with spin have a magnetic dipole moment; those with spin quantum number greater than the minimum of 1/2 also have an electric quadrupole moment. The magnetic and electric fields can be applied externally or come from the electronic structure surrounding the nuclei; they can also be static or dynamic. It is usually the case that nuclei in molecules at room temperature are simultaneously subjected to a combination of external and internal magnetic and electric fields that are both static and dynamic. The study of NMR spectroscopy and nuclear spin relaxation is founded on the quantum mechanics of spin interactions. Semiclassical and phenomenological treatments of spin interactions are widely used and provide a practical starting point for understanding NMR in solids.[33–36]

2.1 Nuclear Spins and Moments

Of the nuclei in the periodic table, most are said to be ''NMR active''. In a semiclassical treatment, these nuclei are viewed as charged structures that have an ''intrinsic'' or ''spin'' angular momentum given by $\hbar\vec{I}$, where $\hbar$ is Plank's constant and has units of angular momentum, and $\vec{I}$ is the dimensionless operator for angular momentum. The nuclear spin quantum number is a dimensionless, scalar quantity that is always positive and of magnitude $n/2$, where $n = 0, 1, 2, \ldots$; it is directly proportional to the maximum projection (z-component) of the total spin angular momentum onto the axis of spatial quantization. Several commonly investigated nuclei, including ^{1}H, ^{13}C, and ^{19}F, have the smallest spin quantum number, 1/2. An isomer of the hafnium nuclides, ^{177m2}Hf, has one of the largest nuclear spin quantum numbers, 37/2. The angular momentum of the nucleus is thought to be associated with the movement of constituent nuclear charge and has a sign, + or −. The orbiting charged particles that constitute the nucleus generate electric currents that, in turn, produce a nuclear magnetic moment, $\vec{\mu}$, given by $\vec{\mu} = \gamma\hbar\vec{I}$, where γ is the nuclear magnetogyric ratio.

Tables exist that give γ for NMR-active nuclei.[37–39] The nuclear magnetic moment also produces a dipolar magnetic field. Thus, a nucleus that possesses angular momentum, designated by a polar vector, has a magnetic moment, also designated by a polar vector; the two polar vectors are parallel when γ is positive and antiparallel when γ is negative. The sum of the z-components of the nuclear magnetic moments in an ensemble of spins constitutes the nuclear spin magnetization of the sample. Nuclei in which the constituent charge is not spherically distributed in space also possess an electric

quadrupole moment (nuclear electric dipole moments are theoretically not allowed). This electric moment produces an associated electric field that is different from the field produced by a point charge. Also, the orientation of the electric quadrupole moment is rigidly tethered to the orientation of the magnetic dipole moment. An electric field gradient and a homogeneous magnetic field can simultaneously interact with the nucleus and compete to align the nuclear angular momentum in the respective field directions.

2.2 Actions of Nuclear Spins in Magnetic Fields

The spin of an atom or nucleus is equivalent to the spin of a toy top. The spin in both cases imparts an angular momentum. The angular momentum, like any momentum, resists change. Gravity, acting on a slightly tilted spinning top, imparts a downward force on the angular momentum. The result is a torque that causes the top to precess about the direction of the gravitational field. The spin angular momentum of a nucleus and the magnetic moment of that nucleus are rigidly tethered together. A magnetic field acting on the tilted magnetic moment (tilted due to spatial quantization), in analogy to the gravitational field acting on the toy top causes the nuclear spin to precess about the direction of the magnetic field. Thus, like the top, the nuclear spin simultaneously spins about its own axis and precesses about the magnetic field direction. The spin rate determines the magnitude of the angular momentum, while the sense or sign of spin rotation determines the direction of the angular momentum. In turn, the magnitude and sign of the angular momentum determine the rate and sense (or the sign) of precession. The intrinsic nature of the spin angular momentum means that the rate of rotation of the nuclear spin is constant. And, therefore, the precession frequency (called the Larmor frequency) is also constant. For a toy top, in contrast, frictional forces slow the rate of rotation and thus the rate of precession.

A nuclear spin, precessing in a static magnetic field, can be manipulated by the application of an additional magnetic field that rotates synchronously with the precessional motion. The concerted effect of the rotational magnetic field, $\mathbf{B}_1$, and the static magnetic field, $\mathbf{B}^{ext}$, is analyzed in a frame of reference that rotates in step with the precessional motion of the nuclear spin, which is due only to $\mathbf{B}^{ext}$. In such a rotating frame, $\mathbf{B}^{ext}$ vanishes. If the frame of reference rotates at a rate different from the rate of precession of the spins then a pseudo (or offset) magnetic field $\Delta\mathbf{B}^{ext}$ results. The magnitude of $\Delta\mathbf{B}^{ext}$ would be commensurate with the difference between the different rates of rotation. Spin precession in synchronous and asynchronous rotating frames is illustrated in Figure 5.

An ensemble of identical spins will precess about $\mathbf{B}^{ext}$ at the same rate; however, each will have an instantaneous

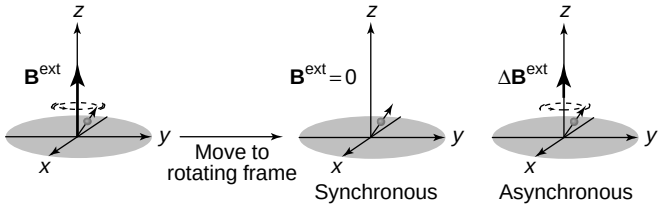

Figure 5 Transformation from the laboratory frame to the rotating frame. In the laboratory frame, spin precession is fast and proportional to $\mathbf{B}^{ext}$. In the synchronous frame, the spin precession is halted and $\mathbf{B}^{ext}$ appears to be zero. In the asynchronous frame, spin precession occurs, but at a much reduced rate, proportional to the magnitude of $\Delta\mathbf{B}^{ext}$.

random position. That is, their phases are not correlated. The sum of all projections of nuclear magnetic moments in thermal equilibrium on the axis defined by the static field direction is the Boltzmann nuclear magnetization, $\mathbf{M}$.[33] The rotating frame is useful for following the trajectory of the nuclear magnetization as it is manipulated by transient magnetic fields and then allowed to recover to equilibrium, subject only to $\mathbf{B}^{ext}$ and internal fields. The action of $\mathbf{B}_1$ (in the synchronous or on-resonance rotating frame) on $\mathbf{M}$ is illustrated for two transient pulses in Figure 6.

The application of a $\pi/2$ pulse rotates or nutates $\mathbf{M}$ by 90°; similarly, a π pulse rotates $\mathbf{M}$ by 180°, that is, $\mathbf{M}$ is completely inverted. In each case, $\mathbf{M}$ nutates only during the application of $\mathbf{B}_1$. The action of $\mathbf{B}_1^{eff}$, the vector sum

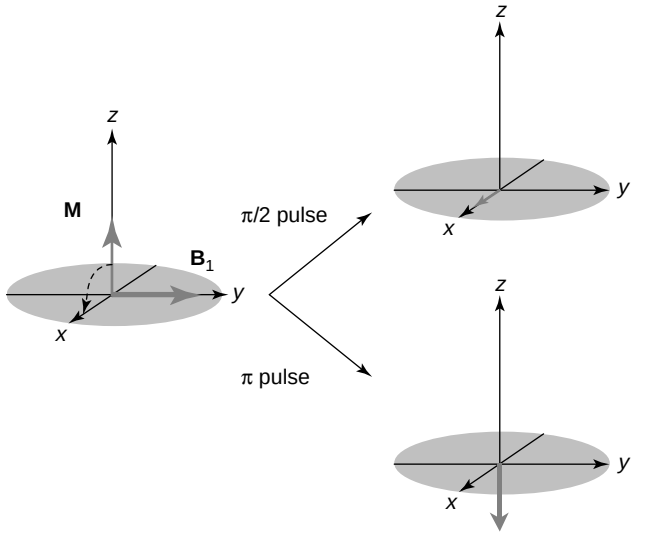

Figure 6 Magnetization, $\mathbf{M}$, is nutated through 90° and 180° following the application of a $\pi/2$ and π pulse of synchronous (on resonance) radiofrequency energy, respectively. $\mathbf{B}^{ext}$ is directed parallel to the z coordinate, but is not shown because it vanishes when the rotating coordinate system coincides with the precessional motion of the spins. $\mathbf{B}^{ext}$ is typically 1000–10 000 times larger than $\mathbf{B}_1$ in the fixed laboratory frame.

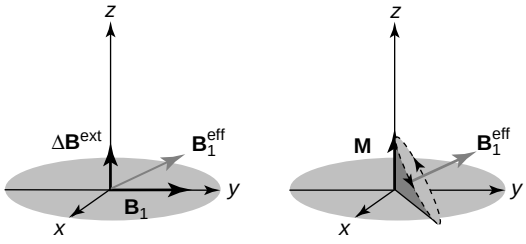

Figure 7 Nuclear spins that are subjected to an asynchronous (off-resonance) magnetic field experience an effective radiofrequency field, $\mathbf{B}_1^{\text{eff}}$, where $\mathbf{B}_1^{\text{eff}} = \mathbf{B}_1 + \Delta\mathbf{B}^{\text{ext}}$. A large difference between the nuclear precession and rotating frame frequencies causes $\mathbf{M}$ to nutate, about $\mathbf{B}_1^{\text{eff}}$, on the surface of a cone rather than on a plane.

of $\mathbf{B}_1$ and $\Delta\mathbf{B}^{\text{ext}}$ (in the asynchronous or off-resonance rotating frame) on $\mathbf{M}$ is illustrated in Figure 7. It is usually the case that $\mathbf{B}_1 \gg \Delta\mathbf{B}^{\text{ext}}$, so that the magnetization nutates in a plane (very nearly) perpendicular to $\mathbf{B}_1$. Following the application of $\mathbf{B}_1$, $\mathbf{M}$ is free to precess about $\mathbf{B}^{\text{ext}}$ in the laboratory frame and induces a voltage in a nearby receiver coil. The signal, a free induction decay (FID), is transformed from the time domain to the frequency domain (NMR spectrum) by the fast Fourier transform (FFT) algorithm.

2.3 Isotropic Nuclear Spin Interactions

The coupling between a nuclear magnetic moment and an externally applied homogeneous magnetic field that defines the laboratory z-coordinate, $\mathbf{B}^{\text{ext}} = (0, 0, B_0)$, is an isotropic interaction. The Zeeman Hamiltonian describes it:[40] $H_z = -\boldsymbol{\mu} \bullet \mathbf{B}^{\text{ext}}$. Only the dot-product term between the angular momentum component I_z and the nonzero magnetic field component survives, so that $H_z = -\gamma\hbar B_0 I_z$. The eigenvalues of the Zeeman Hamiltonian are the energies associated with the transitions between the energy levels that are occupied by the nuclear spin when it is placed in the magnetic field: $E_{\text{m}} = -\gamma\hbar B_0 m$. (The operation of I_z on a state of a spin with intrinsic angular momentum quantum number I results in an eigenvalue m, called the nuclear magnetic quantum number, that ranges from $-I$ to $+I$ in unit increments.) These energy levels are thought to correspond to discrete angular orientations of the spin angular momentum vector with respect to the direction of the external magnetic field (this correspondence is termed space quantization). Transitions are allowed between adjacent Zeeman energy levels ($\Delta m = 1$) such that the frequency of electromagnetic radiation needed to effect a transition is $\Delta E = \gamma\hbar B_0 = \hbar w_0$, where w_0 is called the Larmor frequency. An NMR-active nucleus placed in a magnetic field experiences a force that arises from the interaction between the field and the component of the magnetic moment perpendicular to the field. This force manifests itself as a torque, because the magnetic moment originates from a spin angular momentum, and a cross product describes the effect of the magnetic force on an angular momentum. The resultant motion is a precession of the magnetic moment about the magnetic field direction at the Larmor frequency. The frequency of the precession depends on the strength of the nuclear magnetic moment and the magnitude of the magnetic field, as given above.

To affect nuclear Zeeman transitions, the magnetic field component of circularly polarized electromagnetic radiation applied at the Larmor frequency, $\mathbf{B}_1(t)$, must have a component that is perpendicular to the direction of the external magnetic field. The circularly polarized magnetic field must also have the same sense of rotation as the nuclear precession. The Hamiltonian that describes the time-dependent interaction of the nuclear magnetic moment with $\mathbf{B}_1(t)$ is $H_{B_1} = -\boldsymbol{\mu} \bullet \mathbf{B}_1(t)$. Using the expression for $\mathbf{B}_1(t)$ at a frequency w near or at the Larmor frequency and an initial phase ϕ, that is, $\mathbf{B}_1(t) = B_1[\cos(wt + \phi)\hat{\imath} + \sin(wt + \phi)\hat{\jmath}]$, the expression for the Hamiltonian for the time-dependent magnetic field interactions is $H_{B_1} = \hbar w_1[\cos(wt + \phi)I_x + \sin(wt + \phi)I_y]$, where $w_1 = \gamma B_1$, and the operators I_x and I_y are the transverse components of the spin angular momentum operator. These operators can affect transitions of the nuclear spins between adjacent spin states.

2.4 Anisotropic Nuclear Spin Interactions

In NMR, nearly all observable properties are directly related to or described by a second-rank tensor, namely: diamagnetic and paramagnetic susceptibility, scalar (through-bond) coupling, dipolar (through-space) coupling, electric field gradient, and (chemical) shielding. Visualizing the spatial anisotropy of a second-rank tensor facilitates an understanding of many spectral features that are characteristic of NMRs in the solid state. Furthermore, a pictorial representation of an anisotropic spin interaction results in an automatic comparison of an experimental observable to the molecular architecture. The relationship is particularly clear for envisioning, for example, the variation of proton shielding as a function of hydrogen bond strength or the carbon shielding in strained ring systems.[41–43]

The electronic shielding of nuclei is a consequence of the interaction of the external magnetic field with the electrons of an atom or molecule.[44] Hansen and Bouman depict a shielding surface as a representation for the traceless symmetric constituent of the shielding tensor.[43,45] Resonance surfaces, which represent the NMR frequency directly, are similar to shielding surfaces. The purpose of this section is to explain resonance surfaces and to demonstrate their usefulness for understanding the anisotropic

nature of the interactions of spins with their environment. Also, this will provide an avenue to understand the technique of magic angle spinning (MAS), which is ubiquitous in NMR studies of solids.[36,46,47]

A Cartesian tensor of rank two is defined in every rectangular coordinate system by nine components, T_{kl}, which transform under a rotation of the axis system into a new set, T'_{ij}, according to Equation (1):

$$T'_{ij} = \sum_{k=1}^{3}\sum_{l=1}^{3} R_{ik}R_{jl}T_{kl}, \quad (i, j = 1, 2, 3) \tag{1}$$

where R_{ik} and R_{jl} are the direction cosines contained in the rotation matrix $[\mathbf{R}_{ij}]$.[48,49] In a Cartesian coordinate system, the subscripts i, j, k, and l range from 1 to 3 and are assigned to the three Cartesian axes, as follows: $1 = x$, $2 = y$, and $3 = z$. The first subscript of a direction cosine refers to the new or rotated axis ($x'_1 = x_{\text{new}}$, $x'_2 = y_{\text{new}}$, $x'_3 = z_{\text{new}}$), and the second subscript refers to the old or original axis ($x_1 = x_{\text{old}}$, $x_2 = y_{\text{old}}$, $x_3 = z_{\text{old}}$). The angle θ_{ij} follows the same convention (e.g. θ_{23} is the angle between the new y-axis, x'_2, and the old z-axis, x_3). The two sets of Cartesian axes and the corresponding angles that connect them are illustrated in Figure 8.

Rather than discuss the nonsymmetric second-rank tensor T_{kl} as an abstract mathematical construct, here we will pictorialize the resulting analytical expressions that describe an observable directly affected by a particular second-rank tensor, namely the electronic shielding tensor.[50–52] The description that follows is general and requires only a slight modification to incorporate the tensor of interest with its associated units of measurement. From this point forward it will be convenient to write the subscripts for tensor and vector components with the labels x, y, and z explicitly corresponding to the three Cartesian axes.

Each element of the shielding tensor $[\sigma_{ij}]$, Equation (2),

$$[\sigma_{ij}] = \begin{matrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{matrix} \tag{2}$$

determines, in part, a single component of the induced magnetic field vector in response to one of the components of the external magnetic field vector, $\mathbf{B}^{\text{ext}}$. The induced magnetic field is the response of a system of moving charged particles to a (typically large) static externally applied magnetic field. For example, the element σ_{xy} determines the magnitude of the x-component of the induced magnetic field vector, B_x^{ind}, caused by the y-component of the external magnetic field, B_y^{ext} (see Figure 9). This has been expressed succinctly as follows, Equation (3):

$$B_i^{\text{ind}} = \sigma_{ij}B_j^{\text{ext}} \quad (i, j = x, y, z) \tag{3}$$

The general (nonsymmetric) second-rank Cartesian tensor (nine independent elements) can be written as the sum of a symmetric (s) and an antisymmetric (a) tensor, Equation (4):

$$[\sigma_{ij}] = [\sigma_{ij}]^{\text{s}} + [\sigma_{ij}]^{\text{a}} \tag{4}$$

The symmetric tensor can also be decomposed into a sum of two tensors, the isotropic (i) and the traceless symmetric tensor (ts), Equation (5):

$$[\sigma_{ij}]^{\text{s}} = [\sigma_{ij}]^{\text{i}} + [\sigma_{ij}]^{\text{ts}} \tag{5}$$

Thus, in some arbitrarily chosen Cartesian coordinate system (with axes labeled x, y, and z) we can write the

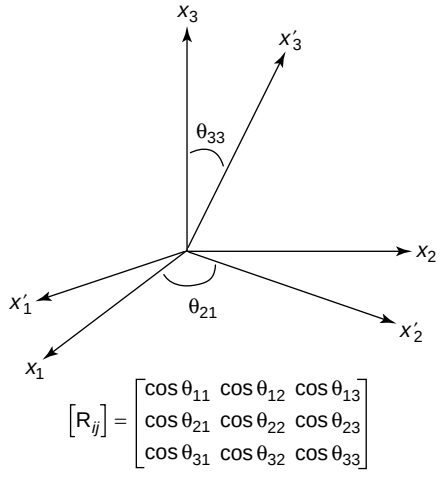

Figure 8 Matrix $[\mathbf{R}_{ij}]$ is used to rotate a second-rank Cartesian tensor (e.g. σ_{ij}) from its original representation in the old (unprimed) Cartesian coordinate system to a new (primed) Cartesian coordinate system.

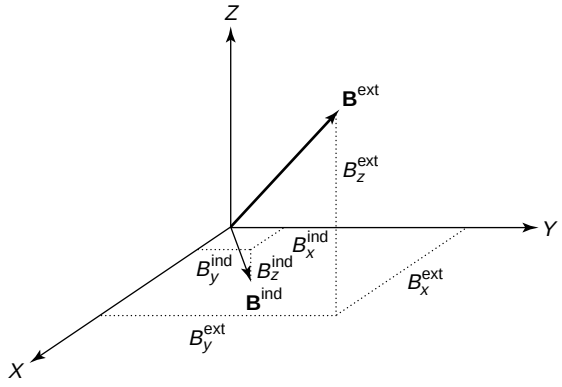

Figure 9 Electronic shielding tensor is defined in a Cartesian coordinate system. Each component of the induced magnetic field is given by $B_i^{\text{ind}} = \sigma_{ij}B_j^{\text{ext}}$ (i, $j = x$, y, z), where σ_{ij} is an element of the tensor matrix $[\sigma_{ij}]$ and B_j^{ext} is a component of the external magnetic field.

following tensor decomposition, Equation (6):

$$
\begin{aligned}
[\sigma_{ij}] = {} & \begin{bmatrix} \frac{\sigma_{xx}+\sigma_{yy}+\sigma_{zz}}{3} & 0 & 0 \\ 0 & \frac{\sigma_{xx}+\sigma_{yy}+\sigma_{zz}}{3} & 0 \\ 0 & 0 & \frac{\sigma_{xx}+\sigma_{yy}+\sigma_{zz}}{3} \end{bmatrix}_{\text{i}} \\
& + \begin{bmatrix} \sigma_{xx} - \left(\frac{\sigma_{xx}+\sigma_{yy}+\sigma_{zz}}{3} \right) & \frac{\sigma_{xy}+\sigma_{yx}}{2} & \frac{\sigma_{xz}+\sigma_{zx}}{2} \\ \frac{\sigma_{xy}+\sigma_{yx}}{2} & \sigma_{yy} - \left(\frac{\sigma_{xx}+\sigma_{yy}+\sigma_{zz}}{3} \right) & \frac{\sigma_{yz}+\sigma_{zy}}{2} \\ \frac{\sigma_{xz}+\sigma_{zx}}{2} & \frac{\sigma_{yz}+\sigma_{zy}}{2} & \sigma_{zz} - \left(\frac{\sigma_{xx}+\sigma_{yy}+\sigma_{zz}}{3} \right) \end{bmatrix}_{\text{ts}} \\
& + \begin{bmatrix} 0 & \frac{\sigma_{xy}-\sigma_{yx}}{2} & \frac{\sigma_{xz}-\sigma_{zx}}{2} \\ \frac{-(\sigma_{xy}-\sigma_{yx})}{2} & 0 & \frac{\sigma_{yz}-\sigma_{zy}}{2} \\ \frac{-(\sigma_{xz}-\sigma_{zx})}{2} & \frac{-(\sigma_{yz}-\sigma_{zy})}{2} & 0 \end{bmatrix}_{\text{a}}
\end{aligned} \tag{6}
$$

The tensor decomposition can be verified by adding corresponding elements from each of the constituents of the shielding tensor. For example, the shielding tensor element in the first column and second row is given by Equation (7),

$$\sigma_{yx} = 0 + \frac{\sigma_{xy}+\sigma_{yx}}{2} + \frac{-(\sigma_{xy}-\sigma_{yx})}{2} \tag{7}$$

The traceless symmetric constituent of the shielding tensor can be diagonalized by a similarity transformation. The process of the tensor (matrix) diagonalization is identical to determining the traceless symmetric tensor elements in a Cartesian coordinate system coincident with the so-called principal axis system. The principal axis system is a peculiar Cartesian coordinate system wherein the traceless symmetric constituent of the shielding tensor is described by three elements, the so-called principal elements. These elements are labeled σ_{11}, σ_{22}, and σ_{33}, with the convention that $\sigma_{11} \leq \sigma_{22} \leq \sigma_{33}$, and are located along the diagonal. The particular rotation matrix (determined from a similarity transformation) required to transform an absolute shielding tensor (described in some arbitrary Cartesian coordinate system) to the absolute shielding tensor in the principal axis system can likewise be applied to the isotropic constituent and the antisymmetric constituent of the shielding tensor. Three new elements are required to describe the antisymmetric constituent of the shielding tensor in the principal axis coordinate system. The antisymmetric principal elements, as they might be called, are labeled σ_{12}, σ_{13}, and σ_{23}.

Thus far we have accounted for six elements of a nine-element tensor. Since we have specified the principal axes as the coordinate system in which the tensor is described, the Euler angles that transform an arbitrary Cartesian coordinate system into the principal axis coordinate system account for the three missing elements. It will be useful to the reader of solid state NMR articles to show the complete description of the absolute shielding tensor in the principal axis system and some commonly used definitions that incorporate the principal elements. The complete description of the absolute shielding tensor described in the principal axis system can be written as follows, Equation (8):

$$
\begin{aligned}
[\sigma_{ij}]_{\text{pas}} = {} & \begin{bmatrix} \frac{\sigma_{11}+\sigma_{22}+\sigma_{33}}{3} & 0 & 0 \\ 0 & \frac{\sigma_{11}+\sigma_{22}+\sigma_{33}}{3} & 0 \\ 0 & 0 & \frac{\sigma_{11}+\sigma_{22}+\sigma_{33}}{3} \end{bmatrix}_{\text{pas}} \\
& + \begin{bmatrix} \sigma_{11} - \left(\frac{\sigma_{11}+\sigma_{22}+\sigma_{33}}{3} \right) & 0 & 0 \\ 0 & \sigma_{22} - \left(\frac{\sigma_{11}+\sigma_{22}+\sigma_{33}}{3} \right) & 0 \\ 0 & 0 & \sigma_{33} - \left(\frac{\sigma_{11}+\sigma_{22}+\sigma_{33}}{3} \right) \end{bmatrix}_{\text{pas}} \\
& + \begin{bmatrix} 0 & \sigma_{12} & \sigma_{13} \\ -\sigma_{12} & 0 & \sigma_{23} \\ -\sigma_{13} & -\sigma_{23} & 0 \end{bmatrix}_{\text{pas}}
\end{aligned} \tag{8}
$$

The average or isotropic shielding, σ^{ave}, is very nearly equal to that which would be measured at a nuclear site in an isotropically tumbling molecule (because second-order terms are not included), Equation (9):

$$\sigma^{\text{ave}} = \tfrac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \tag{9}$$

A partial measure of the spatial anisotropy of the electronic shielding arising only from the traceless symmetric constituent of the shielding tensor is given by $\Delta\sigma$, which is (loosely) called the anisotropy and is defined as follows, Equation (10):

$$\Delta\sigma = \sigma_{33} - \tfrac{1}{2}(\sigma_{11} + \sigma_{22}) \tag{10}$$

Nonzero antisymmetric principal elements also account for spatial anisotropy of the shielding tensor and are not included in $\Delta\sigma$.

The difference between the largest shielding element, σ_{33}, and the average shielding, σ^{ave}, is defined as delta, δ, and is equivalent to two-thirds of $\Delta\sigma$, Equation (11):

$$\delta = \sigma_{33} - \sigma^{\text{ave}} = \tfrac{2}{3}\Delta\sigma \tag{11}$$

Lastly, the axial asymmetry, η, which ranges from 0 (axially symmetric) to 1 (axially asymmetric or axially nonsymmetric) is defined by Equation (12),

$$\eta = \frac{\sigma_{22} - \sigma_{11}}{\delta} \tag{12}$$

The axial asymmetry is also a partial measure of the spatial anisotropy of the electronic shielding arising only from the traceless symmetric constituent of the shielding tensor.

Now that we have the general (nonsymmetric) shielding tensor broken down into its three constituent tensors, we can ask: *What effect does each of the constituent tensors have on the resonance frequency of a nuclear magnetic moment?* To answer this question we will find it convenient to write the induced magnetic field vector in terms of its spherical polar components. The induced magnetic field vector defined in Cartesian components, $\mathbf{B}^{\text{ind}}_{\text{Cartesian}} = B^{\text{ind}}_x\mathbf{i} + B^{\text{ind}}_y\mathbf{j} + B^{\text{ind}}_z\mathbf{k}$, can be expressed in spherical polar components, $\mathbf{B}^{\text{ind}}_{\text{Polar}} = B^{\text{ind}}_r\mathbf{e}_r + B^{\text{ind}}_\theta\mathbf{e}_\theta + B^{\text{ind}}_\phi\mathbf{e}_\phi$, by the following transformation, Equation (13):

$$\mathbf{B}^{\text{ind}}_{\text{Polar}} = \mathbf{P}\cdot\mathbf{B}^{\text{ind}}_{\text{Cartesian}} \tag{13}$$

where $\mathbf{P}$ is the transformation matrix for the conversion of Cartesian to spherical polar coordinates. Explicitly, we have the matrix Equation (14):

$$\begin{matrix} B^{\text{ind}}_r\mathbf{e}_r \\ B^{\text{ind}}_\theta\mathbf{e}_\theta \\ B^{\text{ind}}_\phi\mathbf{e}_\phi \end{matrix} = \begin{matrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\sin\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{matrix} \begin{matrix} B^{\text{ind}}_x\mathbf{i} \\ B^{\text{ind}}_y\mathbf{j} \\ B^{\text{ind}}_z\mathbf{k} \end{matrix} \tag{14}$$

For an applied static magnetic field $\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$ (in the direction $\mathbf{e}_r$) the induced magnetic field is given by $\mathbf{B}^{\text{ind}}_{\text{Cartesian}} = -[\sigma_{ij}]\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$ in Cartesian components and $\mathbf{B}^{\text{ind}}_{\text{Polar}} = -\mathbf{P}[\sigma_{ij}]\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$ in spherical polar components. It is from this last equation that the effects of $[\sigma_{ij}]^{\text{i}}$, $[\sigma_{ij}]^{\text{ts}}$, and $[\sigma_{ij}]^{\text{a}}$ on the nuclear magnetic moment precession frequency and the nuclear relaxation rates can best be described. By expressing $\mathbf{B}^{\text{ind}}_{\text{Cartesian}}$ in spherical polar coordinates and allowing $\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$ to define the radial direction, $|\mathbf{B}^{\text{ext}}_{\text{Cartesian}}|\mathbf{e}_r = \mathbf{B}^{\text{ext}}_{\text{Cartesian}}$, we will obtain components of $\mathbf{B}^{\text{ind}}_{\text{Polar}}$ which are parallel ($\mathbf{B}^{\text{ind}}_{\parallel}$) or perpendicular ($\mathbf{B}^{\text{ind}}_{\perp}$) to $\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$. The parallel component will increase or decrease the effective magnetic field directly and therefore affect the resonance frequency to first order. The perpendicular components will also affect the resonance frequency, but only to second order, and therefore to a much lesser degree (see Figure 10).

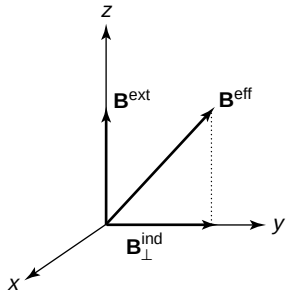

Figure 10 The resonance frequency of a nuclear spin caused by $\mathbf{B}^{\text{ind}}_{\perp}$ is directly proportional to the net, effective magnetic field, $\mathbf{B}^{\text{eff}}$, where $\text{B}^{\text{eff}} = \sqrt{\text{B}^{\text{ext2}} + \text{B}^{\text{ind2}}_{\perp}}$. When $\mathbf{B}^{\text{eff}} \gg \mathbf{B}^{\text{ind}}_{\perp}$, the nuclear resonance frequency is affected by the transverse field, $B^{\text{ind}}_{\perp}$, very weakly.

The interaction of molecular electrons, undergoing simple (diamagnetic) circular motion, with $\mathbf{B}^{\text{ext}}$ is described by the isotropic constituent of the shielding tensor and results in an induced magnetic field that is (always) antiparallel to the external magnetic field.[44] For any orientation of the external magnetic field in the Cartesian coordinate system in which the shielding tensor is defined, the perpendicular components of the induced magnetic field are always zero. The induced magnetic field is given by ${}^{\text{i}}\mathbf{B}^{\text{ind}}_{\text{Polar}} = -\mathbf{P}\cdot[\sigma_{ij}]^{\text{i}}\cdot\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$, which results in the following expression, Equation (15):

$${}^{\text{i}}B^{\text{ind}}_r(\theta,\phi) = -\sigma^{\text{i}}B^{\text{ext}}_{\text{Cartesian}} \tag{15}$$

where, Equation (16),

$$\sigma^{\text{i}} = \frac{\sigma_{xx} + \sigma_{yy} + \sigma_{zz}}{3} \tag{16}$$

The traceless symmetric constituent of the shielding tensor gives rise to an induced magnetic field with both parallel and perpendicular components relative to the external magnetic field. To first order in the resonance frequency of a nuclear spin, the spatial anisotropy of the shielding tensor is revealed by the parallel component. Additional (second order) spatial anisotropy of the shielding tensor arises from the transverse components. The equation ${}^{\text{ts}}\mathbf{B}^{\text{ind}}_{\text{Polar}} = -\mathbf{P}\cdot[\sigma_{ij}]^{\text{ts}}\cdot\mathbf{B}^{\text{ext}}_{\text{Cartesian}}$ results in expressions for both radial and transverse components of an induced magnetic field. The radial component affects the resonance frequency to first order and is given by Equation (17),

$$\begin{aligned} {}^{\text{ts}}B^{\text{ind}}_r(\theta,\phi) = \Bigg\{ &-\sin^2\theta\left\{(\sigma_{xx}-\sigma^{i})\cos^2\phi\right. \\ &\left. + (\sigma_{yy}-\sigma^{\text{i}})\sin^2\phi + \frac{\sigma_{xy}+\sigma_{yx}}{2}\sin 2\phi\right\} \\ &- \sin 2\theta\left\{\frac{\sigma_{xz}+\sigma_{zx}}{2}\cos\phi + \frac{\sigma_{yz}+\sigma_{zy}}{2}\sin\phi\right\} \\ &\left. - (\sigma_{zz}-\sigma^{\text{i}})\cos^2\theta\right\} B^{\text{ext}}_{\text{Cartesian}} \end{aligned} \tag{17}$$

Although the antisymmetric constituent of the shielding tensor manifests itself very weakly in the NMR resonance frequency, it can play a significant role in the longitudinal nuclear spin relaxation rate.[42] The antisymmetric constituent of the shielding tensor gives rise to an induced magnetic field consisting only of perpendicular components relative to the external magnetic field.[53–56]

Thus far we have determined the contributions from the isotropic and the traceless symmetric constituents of the

shielding tensor $[\sigma_{ij}]^{\mathrm{i}}$ and $[\sigma_{ij}]^{\mathrm{ts}}$ to the parallel component of the induced magnetic field, viz. Equation (18):

$$B_r^{\mathrm{ind}}(\theta,\phi) = {}^{\mathrm{i}}B_r^{\mathrm{ind}}(\theta,\phi) + {}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta,\phi) \tag{18}$$

The induced magnetic field affects several observables in an NMR experiment. We now turn to a quantitative examination of these effects.

The resonance frequency of a bare nucleus with spin 1/2 and placed in an external magnetic field is given by Equation (19),

$$\nu^{\mathrm{bare}} = \gamma B^{\mathrm{ext}} \tag{19}$$

where γ is the magnetogyric ratio, in units of $\mathrm{Hz\,G^{-1}}$ (or radians $\mathrm{T^{-1}}$). The presence of the moving electrons in the vicinity of the nucleus results in an induced magnetic field, which adds (vectorially) to the external magnetic field. Thus, in general, the effective magnetic field at the site of the nucleus is $\mathbf{B}^{\mathrm{eff}} = \mathbf{B}^{\mathrm{ext}} + \mathbf{B}^{\mathrm{ind}}$, and the observed resonance frequency of the nuclear magnetic moment becomes (Equation 20)

$$\nu^{\mathrm{obs}} = \gamma B^{\mathrm{eff}} \tag{20}$$

where $B^{\mathrm{eff}} = \sqrt{(B^{\mathrm{ext}} + B_r^{\mathrm{ind}}(\theta,\phi))^2}$ to first order.

The isotropic constituent of the shielding tensor gives rise to an induced magnetic field that is always antiparallel to the external magnetic field. Hence, the effective magnetic field at the nuclear site is always less than B^{ext}. The effective magnetic field is given by $B^{\mathrm{eff}} = B^{\mathrm{ext}} + {}^{\mathrm{i}}B_r^{\mathrm{ind}}(\theta,\phi)$, or simply $B^{\mathrm{eff}} = B^{\mathrm{ext}} - \sigma^{\mathrm{i}}B^{\mathrm{ext}}$, and the observed NMR resonance frequency is ${}^{\mathrm{i}}\nu^{\mathrm{obs}} = \gamma B^{\mathrm{ext}} - \gamma\sigma^{\mathrm{i}}B^{\mathrm{ext}}$. Clearly, for all combinations of θ and ϕ ($0^\circ \le \theta \le 180^\circ$, $0^\circ \le \phi \le 360^\circ$) the observed NMR resonance frequency is constant. We envisage the observed resonance frequency of the nuclear spin for a particular orientation of $\mathbf{B}^{\mathrm{ext}}_{\mathrm{Polar}}$ (given by the polar angles θ and ϕ) as a point on a closed surface. We plot the absolute value of the difference between the observed resonance frequency, ${}^{\mathrm{i}}\nu^{\mathrm{obs}}$, and the resonance frequency of the bare nucleus, ν^{bare} (i.e. ${}^{\mathrm{i}}\nu^{\mathrm{rel}} = {}^{\mathrm{i}}\nu^{\mathrm{obs}} - \nu^{\mathrm{bare}}$, where ${}^{\mathrm{i}}\nu^{\mathrm{rel}}$ is the relative frequency and $\nu^{\mathrm{bare}} = \gamma B^{\mathrm{ext}}$). The inherent resonance frequency of the nucleus, ν^{bare}, is a constant for a given magnetic field and, therefore, provides a frequency with which all frequency shifts (owing to the three constituents of the shielding tensor) can be compared.

A direct analogy to the above is made by entering the rotating frame at the precession frequency of the bare nucleus ν^{bare}, or Larmor frequency. The points that form the resonance surface $[{}^{\mathrm{i}}\nu^{\mathrm{rel}}(\theta,\phi),\theta,\phi]$, are rendered blue for ${}^{\mathrm{i}}\nu^{\mathrm{rel}} \ge 0$ (with $\sigma^{\mathrm{i}} \le 0$), an impossibility, and red for ${}^{\mathrm{i}}\nu^{\mathrm{rel}} \le 0$ (with $\sigma^{\mathrm{i}} \ge 0$). Figure 11 is a resonance surface depicting the effect on the resonance frequency of a nuclear spin due to the isotropic constituent of the

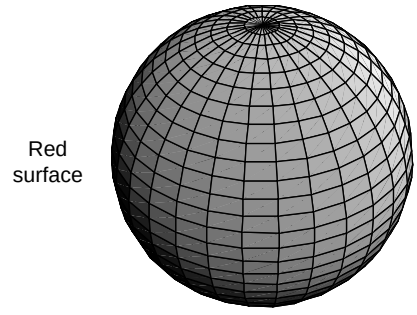

Figure 11 Resonance surface for the radial component of the induced magnetic field due to the isotropic constituent of the shielding tensor. The intersection of the applied static field with this surface (designated red) causes the nuclear spin to resonate at a lower frequency compared to the resonance frequency of the bare nucleus.

following shielding tensor, defined in the principal axis system, Equation (21):

$$[\sigma_{ij}^{\eta=0}] = \begin{matrix} 20.0 & 50.0 & 0.0 \\ -50.0 & 20.0 & 0.0 \\ 0.0 & 0.0 & 100.0 \end{matrix} \tag{21}$$

where the tensor elements are given in parts per million (ppm). For example, $\sigma_{xx} = 20.0\,\mathrm{ppm} = (20.0/1\,000\,000.0)$. (Although this shielding tensor is antisymmetric, the antisymmetry does not affect the resonance surface because of the isotropic constituent tensor.) The resonance surface depicted in Figure 11 and all subsequent resonance surfaces were calculated and rendered using Mathematica.[57] The symbolic language program AXIOM is also a useful mathematical tool for the investigation of second and higher rank tensors and is preferred by the author.[58] The AXIOM system is marketed and distributed by the Numerical Algorithms Group, Ltd., Oxford, UK, and Numerical Algorithms Group, Inc., 1400 Opus Place, Suite 200, Downers Grove, Illinois, USA (Phone: (630) 971–2337, Fax: (630) 971–2706). The AXIOM commands can be obtained from the author, R.E. Gerald II, Argonne National Laboratory, 9700 S. Cass Ave., Argonne, Illinois 60439, USA.

The effect of the radial component of the traceless symmetric constituent of the shielding tensor on the resonance frequency of a nuclear spin can be viewed as a three-dimensional surface. The collection of points $[{}^{\mathrm{ts}}\nu^{\mathrm{rel}}(\theta,\phi),\theta,\phi]$, where ${}^{\mathrm{ts}}\nu^{\mathrm{rel}} = \gamma^{\mathrm{ts}}B^{\mathrm{eff}} - \gamma B^{\mathrm{ext}}$ (i.e. ${}^{\mathrm{ts}}\nu^{\mathrm{rel}} = \gamma\sqrt{(B^{\mathrm{ext}} + {}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta,\phi))^2} - \gamma B^{\mathrm{ext}}$) results in a hyperboloid of two sheets. Again, with respect to the resonance frequency of the bare nucleus in the external static magnetic field (ν^{bare}), those surface points designated by red represent frequencies that are lower, and those surface points designated by blue represent frequencies that are higher. Equivalently, when $\mathbf{B}^{\mathrm{ext}}$ intersects the hyperboloid at a point on the red surface the nucleus is

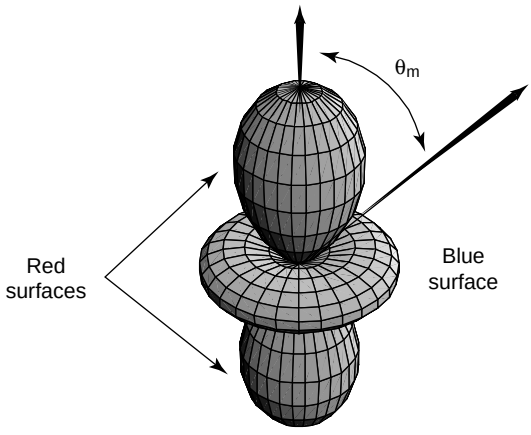

Figure 12 Resonance surface for the radial component of the induced magnetic field due to the traceless symmetric constituent of the shielding tensor. The surface is fixed to the molecular framework and results from the interaction of $\mathbf{B}^{ext}$ (the applied static magnetic field) with the orbiting molecular electrons. The intersection of $\mathbf{B}^{ext}$ with a red or blue surface causes the nuclear spin to resonate at a lower or higher frequency, respectively, compared to the resonance frequency of the bare nucleus. This surface corresponds to an axially symmetric shielding tensor. The intersection of the red and blue surfaces forms a conical node. The angle between the principal shielding axis (the direction of greatest electronic shielding) and the surface of the conical node is the magic angle, θ_m.

shielded (the magnetic field is attenuated) and therefore resonates at a lower frequency compared to the bare nucleus. Similarly, when $\mathbf{B}^{ext}$ intersects the hyperboloid at a point on the blue surface the nucleus is deshielded (the magnetic field is enhanced) and therefore resonates at a higher frequency compared to the bare nucleus. Figure 12 is a resonance surface depicting only the first-order effect of the traceless symmetric constituent of the same shielding tensor defined above, $[\sigma_{ij}^{\eta=0}]$, on the resonance frequency of a nuclear spin. Note the axial symmetry of the traceless symmetric resonance surface representing the axially symmetric ($\eta = 0$) tensor. Part of the spatial anisotropy of the electronic shielding tensor is due to the transverse induced magnetic fields that arise from the traceless symmetric constituent. However, the effect on the nuclear resonance frequency is very small and is not included in this analysis.

Figure 13 is also a resonance surface depicting only the first-order effect of the traceless symmetric constituent of the shielding tensor on the resonance frequency of a nuclear spin. However, in Figure 13 the shielding tensor is axially asymmetric ($\eta = 1/5$), and is defined in the principal axis system by Equation (22),

$$[\sigma_{ij}^{\eta=1/5}] = \begin{matrix} 20.0 & 50.0 & 0.0 \\ -50.0 & 30.0 & 0.0 \\ 0.0 & 0.0 & 100.0 \end{matrix} \qquad (22)$$

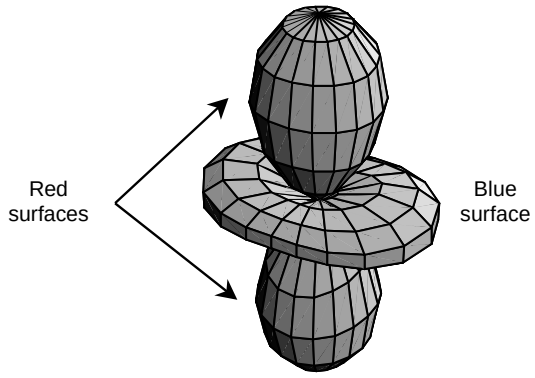

Figure 13 Shielding surface similar to Figure 12, except that this surface corresponds to an axially asymmetric shielding tensor.

The antisymmetric constituent of the shielding tensor does not have a first-order effect on the resonance frequency of a nuclear spin. However, the magnitudes of transverse induced magnetic fields due to nonzero antisymmetric principal elements can be similar to the magnitude of the radial component of the induced magnetic field from the principal elements of a shielding tensor. In addition, part of the spatial anisotropy of electronic shielding is attributed to the transverse induced magnetic field when the antisymmetric principal elements are not zero. (The same can be said of the transverse magnetic fields determined by nonzero principal elements.) The small effects of induced transverse magnetic fields on the nuclear resonance frequency are difficult to measure. However, if the transverse magnetic fields are time dependent, then the antisymmetric principal elements can be estimated from measurements of spin-lattice relaxation rate constants (vide infra). From the perspective of the nucleus, a fluctuating induced transverse magnetic field is similar to $\mathbf{B}_1$ fields generated by radiofrequency coils.

2.5 Averaging Spatial Anisotropy

The technique of magic angle sample spinning (MASS) is widely used to average the spatial anisotropy of nuclear spin interactions. A sample placed in a magic angle spinning (MAS) NMR probe is inclined at an angle of 54.74° with respect to $\mathbf{B}^{ext}$ and continuously rotated at speeds up to 30 000 Hz. Under these conditions, the sample is analyzed by a multitude of high-resolution solid-state NMR experiments.[36] Complete averaging of any anisotropic spin interaction that is mathematically described by a second-rank tensor occurs when the sample rotation frequency is several times larger than the frequency spread of the NMR spectrum for the static sample. A comparison of the achievable maximum spinning speeds with static NMR linewidths for a variety of solid materials

shows that the MASS technique is frequently suitable for averaging the spatial anisotropy due to electronic (chemical) shielding. Samples with a high density of proton or fluorine nuclei (e.g. hydrocarbons, fluorocarbons, polymers) exhibit broad static NMR linewidths due to magnetic dipole (through space) coupling between nuclei, and such interactions are difficult to average by MASS. For those cases, a separate technique called high-power dipolar decoupling is used to reduce NMR linewidths. Details of the spatial averaging of second-rank tensors by MASS will be discussed in the context of electronic shielding, but they are equally applicable to other magnetic and electric spin interactions that are spatially anisotropic.

The isotropic constituent of the shielding tensor results in an induced magnetic field that is antiparallel to the external magnetic field, as discussed previously. For any orientation of the external magnetic field in the Cartesian coordinate system in which the shielding tensor is defined, the perpendicular components of the induced magnetic field are always zero. Since the induced magnetic field due to the isotropic constituent of the shielding tensor is independent of θ and ϕ, rotation (of any type) has no effect. Therefore, the induced magnetic field component under MASS is given, as before, by (Equation 23)

$$\langle {}^{\mathrm{i}}B_r^{\mathrm{ind}}(\theta_{\mathrm{m}},\phi)\rangle^{\mathrm{MASS}} = -\sigma^{\mathrm{i}}B^{\mathrm{ext}}_{\mathrm{Cartesian}} \tag{23}$$

where the angle brackets indicate an average of the angle ϕ over the range $(0^\circ \le \phi \le 360^\circ)$, and the superscript MASS signifies that the external magnetic field and the z-axis of the Cartesian coordinate system in which the shielding tensor is described always maintain an angle equal to θ_{m} (vide supra).

As discussed previously, the traceless symmetric constituent of the shielding tensor gives rise to both parallel and perpendicular components relative to the external magnetic field. The so-called anisotropy and axial asymmetry of the shielding tensor are revealed by the radial (i.e. parallel) component. Additional anisotropy of the electronic shielding arises from the two transverse components. The equation for the time-dependent spherical polar component of the induced magnetic field due to the traceless symmetric constituent of the shielding tensor is obtained by replacing the angle ϕ with the time-dependent angle $2\pi\nu t$ in the corresponding equation derived previously. The equation that follows describes ${}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta_{\mathrm{m}},\phi)$, but is cast into a form (using the trigonometric identity $A\cos(\phi)+B\sin(\phi)=\sqrt{A^2+B^2}\sin(\phi+\zeta)$, where $\tan(\zeta)=A/B$) that will reduce easily under MASS to reveal its effect on the resonance frequency of a nuclear spin, Equation (24):

$$\begin{aligned}{}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta,\phi) = {} & -\sin^2\theta\left\{(\sigma_{xx}-\sigma^{\mathrm{i}})\cos^2\phi+(\sigma_{yy}-\sigma^{\mathrm{i}})\right.\\ & \times\sin^2\phi+\left(\frac{\sigma_{xy}+\sigma_{yx}}{2}\right.\left.\left.\sin 2\phi\right\}-\sin 2\theta\right.\\ & \times\sqrt{\left(\frac{\sigma_{xz}+\sigma_{zx}}{2}\right)^2+\left(\frac{\sigma_{yz}+\sigma_{zy}}{2}\right)^2}\\ & \times\sin(\phi+\zeta)\left.-(\sigma_{zz}-\sigma^{\mathrm{i}})\cos^2\theta\right\}B^{\mathrm{ext}}_{\mathrm{Cartesian}}\end{aligned} \tag{24}$$

where the expression for ζ is (Equation 25)

$$\tan\zeta=\frac{\sigma_{xz}+\sigma_{zx}}{\sigma_{yz}+\sigma_{zy}} \tag{25}$$

If we consider the effects of electronic shielding on the resonance frequency of a nuclear spin, under MASS, then the previous equation reduces to Equation (26):

$$\begin{aligned}\langle {}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta_{\mathrm{m}},\phi)\rangle^{\mathrm{MASS}} = {} & \left\{-\sin^2\theta_{\mathrm{m}}\left\{\frac{\sigma_{xx}-\sigma^{\mathrm{i}}}{2}+\frac{\sigma_{yy}-\sigma^{\mathrm{i}}}{2}\right\}\right.\\ & \left.-(\sigma_{zz}-\sigma^{\mathrm{i}})\cos^2\theta_{\mathrm{m}}\right\}B^{\mathrm{ext}}_{\mathrm{Cartesian}}\end{aligned} \tag{26}$$

The radial component of the induced magnetic field due to the traceless symmetric constituent of the shielding tensor is directly proportional to the observed resonance frequency and can be transformed into the following result by noting that $(\sigma_{xx}-\sigma^{\mathrm{i}})+(\sigma_{yy}-\sigma^{\mathrm{i}})=-(\sigma_{zz}-\sigma^{\mathrm{i}})$, Equation (27):

$$\begin{aligned}\langle {}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta_{\mathrm{m}},\phi)\rangle^{\mathrm{MASS}} = {} & \left\{-(\sigma_{zz}-\sigma^{\mathrm{i}})\frac{3\cos^2\theta_m-1}{2}\right\}\\ & \times B^{\mathrm{ext}}_{\mathrm{Cartesian}}\end{aligned} \tag{27}$$

Thus, the effect of ${}^{\mathrm{ts}}B_r^{\mathrm{ind}}(\theta_m,\phi)$ on the resonance frequency of a nuclear spin vanishes when the z-axis of the Cartesian coordinate system in which the shielding tensor is described rotates about an axis tilted from $\mathbf{B}^{\mathrm{ext}}_{\mathrm{Cartesian}}$ by the magic angle, as illustrated in Figure 14. At the start of the rotation cycle, consider that the Cartesian laboratory frame and the Cartesian coordinate system in which the tensor is defined are coincident. It was shown previously that a second-rank tensor could be described in any arbitrary coordinate system by application of the rotation matrix $[\mathbf{R}_{ij}]$. During a rotation cycle, the laboratory frame is fixed and the z-coordinate of the tensor frame traces out the surface of the magic angle cone in Figure 14. The x and y coordinates of the tensor frame twist in this process such that their projections onto the laboratory x–y plane

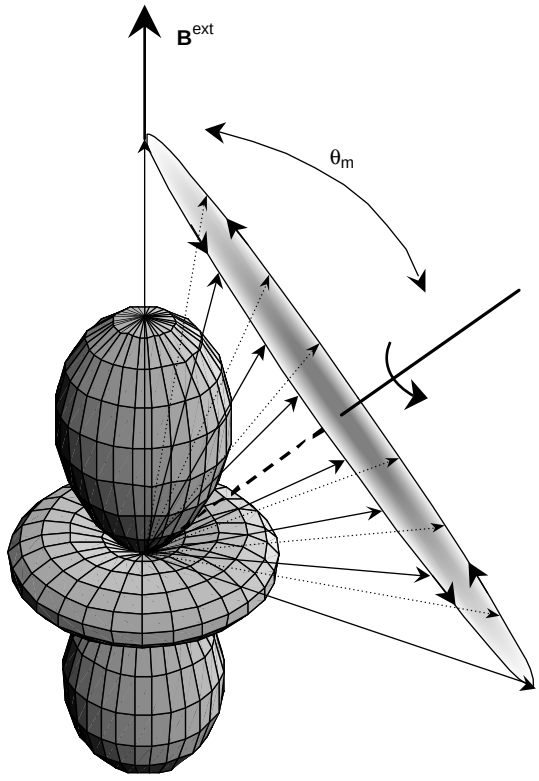

Figure 14 Illustration of the path taken by the shielding tensor in the laboratory frame for the MASS experiment. The resonance surface shown is rotated about an axis that is tilted from $\mathbf{B}^{ext}$ by the magic angle. The average of the NMR frequencies for a set of equally (angularly) spaced orientations of the tensor on the magic angle cone is zero. Thus, under MAS a nuclear spin resonates at a frequency determined (to first order) only by the isotropic constituent of the shielding tensor (see text).

move in circular motions. That is, the projection of the x-coordinate of the tensor frame, for example, makes an angle ϕ with the x-axis of the laboratory frame, and ϕ varies from 0° to 360° during each rotation cycle. Therefore, the chemical shift interaction is averaged over ϕ in the MASS experiment, and $\langle {}^{ts}B_r^{ind}(\theta_m, \phi)\rangle^{MASS} = 0$ for $\theta_m = 54.74°$.

Thus far we have determined the contributions from the two constituents of the shielding tensor $[\sigma_{ij}]^i$ and $[\sigma_{ij}]^{ts}$ to the parallel component of the induced magnetic field under MASS, viz. Equation (28):

$$\langle B_r^{ind}(\theta, \phi)\rangle^{MASS} = \langle {}^{i}B_r^{ind}(\theta, \phi)\rangle^{MASS} + \langle {}^{ts}B_r^{ind}(\theta, \phi)\rangle^{MASS} \tag{28}$$

Under MASS, the isotropic constituent of the shielding tensor gives rise to an induced magnetic field that is always antiparallel to the external magnetic field. The effective magnetic field at the nuclear site is not modulated in magnitude or direction by rotating the Cartesian coordinate system in which the shielding tensor is defined. The observed NMR resonance frequency is constant and given by ${}^{i}\nu^{obs} = \gamma B^{ext} - \gamma\sigma^{i}B^{ext}$.

The average of ${}^{ts}B_r^{ind}(\theta_m, \phi)$ was previously shown to vanish under MASS. However, unlike ${}^{i}B_r^{ind}(\theta_m, \phi)$, which is not modulated under MASS, ${}^{ts}B_r^{ind}(\theta_m, \phi)$ is modulated under MASS, even though the radial component produces a null average effect on the resonance frequency of the nuclear spin. This resonance frequency modulation manifests itself in spinning side bands in MASS experiments. Figure 15 shows the effect of spinning a sample at the magic angle at different spinning frequencies.[59] An analysis of the spinning side band intensities determines the principal elements of the chemical shift tensor, which are directly related by a constant to the principal elements of the chemical shielding tensor.

The Hamiltonian that describes the through-space dipole–dipole interaction can be visualized in part by a second-rank tensor. In this case, the tensor is symmetric, that is, it has no antisymmetric component. The isotropic constituent of the dipolar interaction tensor is zero and, therefore, no resonance shift is observed under isotropic tumbling. The traceless symmetric constituent of the tensor is axially symmetric. Thus, the resulting resonance surface is always similar to the surface shown in Figure 12.

The quadrupolar Hamiltonian describes the interaction of the nuclear electric quadrupole moment with an electric field gradient. This interaction is described by a second-rank tensor as well. However, because of spatial symmetry, the quadrupolar interaction is described by a symmetric tensor and, therefore, results in a resonance surface with a spatial anisotropy similar to the surfaces in Figures 12 and 13. The isotropic shift is zero, except for cases where second-order effects, due to very large interactions, are significant. Engelhardt and Michel have provided a useful theoretical description of the quadrupole interaction and its effects on NMR spectra of quadrupolar nuclei.[60]

2.6 Relaxation of Nuclear Spins

The description of the relaxation of nuclear spins is described in detail by Bloembergen.[61,62] The previous analysis of electronic shielding can be expanded to include nuclear spin relaxation. Transverse and longitudinal fluctuating magnetic fields, regardless of origin, are the prerequisite for nuclear spin relaxation.[63] Because the contribution to the induced magnetic field from the isotropic constituent of the shielding tensor is independent of θ and ϕ, there is no magnetic field modulation due to molecular rotation at the site of the nucleus. Hence, the isotropic constituent of the shielding tensor plays no role in the relaxation of the nuclear spin.

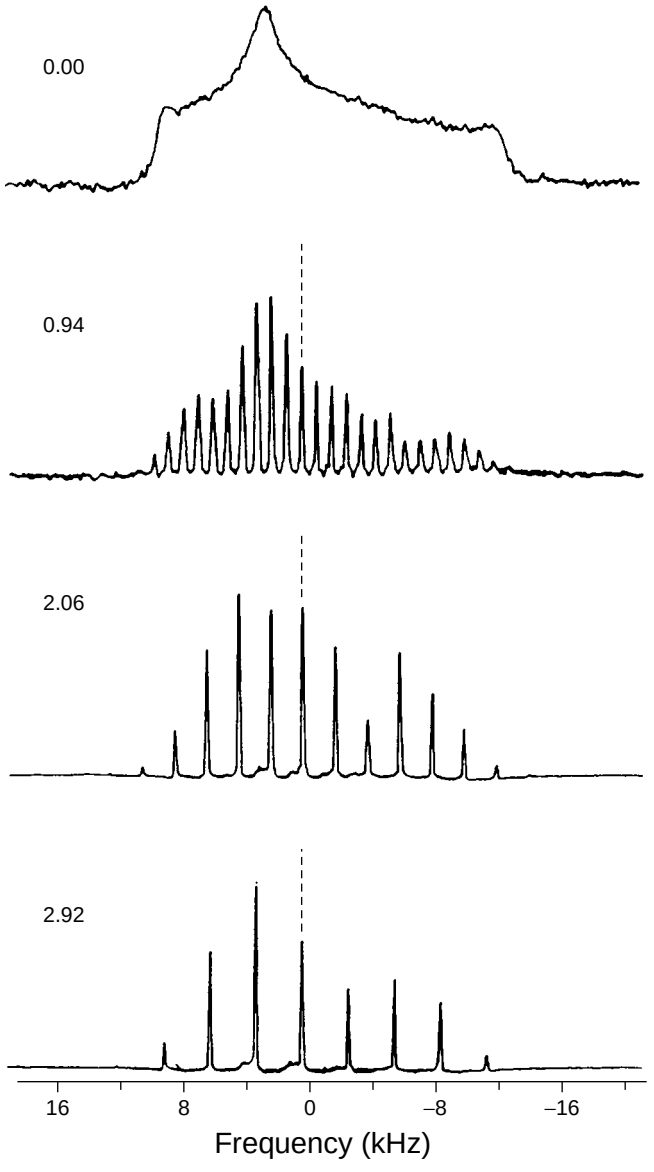

Figure 15 MASS NMR spectra of ^{31}P nuclei in solid dipalmitoyl phosphatidylcholine recorded at different rotation frequencies. The rotation frequencies are indicated (in kHz). The static NMR spectrum (called a powder pattern) is characteristic of an axially asymmetric shielding tensor, from which the principal elements of the chemical shift tensor can be obtained. The principal elements can be obtained with greater accuracy from analyses of the spinning side band intensities. The position of the isotropic chemical shift is indicated by vertical dashed lines and does not change with rotation frequency. The series of spectra confirms that there is only one chemically distinct phosphorous species. Dipolar decoupling eliminated the dipolar interactions between protons and phosphorous nuclei.

However, molecular rotation does cause a time-dependent parallel component of the induced magnetic field that is due to the traceless symmetric and antisymmetric constituents of the shielding tensor. The rate at which an isochromatic nuclear spin ensemble loses phase coherence depends on the period of the molecular rotation, the axis of rotation, the rate of reorientation, and the anisotropy of the electronic shielding of the nuclear spins. For isotropic motion, the contribution to the transverse relaxation rate, T_2^{-1}, from the symmetric constituent is given by Equation (29),

$$\frac{1}{T_2}^{\mathrm{s}} = \gamma^2 (B^{\mathrm{ext}})^2 \ \frac{1}{15} S^2 \left(\frac{4}{3}\tau + \frac{\tau}{1 + w^2\tau^2} \right) \tag{29}$$

where $S^2 = (\Delta\sigma)^2(1 + (1/3)\eta^2)w/2\pi$ is the Larmor frequency (in Hz), and τ is the correlation time for the motion of the molecule; the other parameters have been described above.

The perpendicular components of the induced magnetic field also become time dependent upon molecular rotation; however, the second-order effect on the transverse relaxation rate is negligible. The time-dependent perpendicular components of the induced magnetic field do, however, affect the longitudinal relaxation rate to first order. The resulting longitudinal relaxation rate depends on the magnitude of the Fourier components of the changing transverse magnetic field near the nuclear spin Larmor frequency. For isotropic motion, the contribution to the longitudinal relaxation rate, T_1^{-1}, from the symmetric constituent is given by Equation (30),

$$\frac{1}{T_1}^{\mathrm{s}} = \gamma^2 (B^{\mathrm{ext}})^2 \ \frac{2}{15} S^2 \frac{\tau}{1 + w^2\tau^2} \tag{30}$$

Because the antisymmetric constituent of the shielding tensor gives rise to an induced magnetic field, which is always perpendicular to the external static magnetic field, the second-order effect on the phase memory is very small. For isotropic molecular motion, the contribution to the transverse relaxation rate from the antisymmetric constituent is given by Equation (31),

$$\frac{1}{T_2}^{\mathrm{a}} = \gamma^2 (B^{\mathrm{ext}})^2 \ \frac{1}{3} A^2 \frac{\tau}{1 + w^2\tau^2} \tag{31}$$

where $A^2 = (\sigma_{12}^{\mathrm{a}})^2 + (\sigma_{13}^{\mathrm{a}})^2 + (\sigma_{23}^{\mathrm{a}})^2$.

The effect of the antisymmetric constituent of the shielding tensor on the longitudinal relaxation rate, however, is first order. For isotropic motion the contribution to the longitudinal relaxation rate from the antisymmetric constituent is given by Equation (32),

$$\frac{1}{T_1}^{\mathrm{a}} = \gamma^2 (B^{\mathrm{ext}})^2 \ \frac{2}{3} A^2 \frac{\tau}{1 + w^2\tau^2} \tag{32}$$

Anet and co-workers have reported experimental evidence for a system in which $T_2 > T_1$. This unusual inversion of relaxation time constants is a direct consequence of a relatively large antisymmetric constituent of the shielding tensor in the system they studied.[42]

3 SOLID-STATE NUCLEAR MAGNETIC RESONANCE INSTRUMENTATION

A modern NMR spectrometer consists of a superconducting magnet that generates a stable and homogeneous field; a radiofrequency console that provides control of frequency, duration, amplitude, and phase of pulsed radiofrequency energy output and input; a digital console that allows computer control of the radiofrequency console, and digitization of heterodyned NMR signals; and a computer that interfaces to the digital console for retrieval, processing, and storage of NMR data. Solid-state NMR spectrometers were historically distinguished from their solution-state counterparts by a larger magnet bore, higher power amplifiers, and a faster digitizer. Historically, a magnet bore diameter of 89 mm was desirable because it provided more room for large NMR probe heads of experimental design. The requirement for extra space in the magnetic field stemmed from the large high-voltage capacitors with high power handling capability that are required to generate large radiofrequency fields throughout the volume occupied by the sample. Short 90° pulses (1–2 μs) are required to record wideline and high-resolution solid-state NMR spectra for carbon or other low γ nuclei (the so-called X nuclei). In addition, 300–400 W of pulsed radiofrequency power with less than 5% drop over a 100-ms period is needed for proton decoupling. A faster digitizer ($\geq 1 \times 10^6$ samples s^{-1}) is also necessary for recording spectral windows that are at least ten times larger than the spectral windows for solution-state studies. Modern solid-state NMR spectrometers are often based on solution-state machines, such that if a solutions spectrometer were purchased today, the solid-state capabilities could be added easily in the laboratory at a later time.

3.1 Magnet System

In the late 1940s and early 1950s, the magnetic fields available in the laboratory did not produce a magnetic flux that was uniform in space or constant over time. These magnetic fields were also quite modest when compared to magnetic field strengths generated by superconducting magnets that are common today. The early requirement for spatial uniformity of the magnetic field over the sample volume was technologically very demanding.[64] A variation of less than 0.0000001% is necessary to resolve coarse differences in the isotropic chemical shifts of protons in organic materials. The variation in magnetic field homogeneity available then was perhaps a few tenths of a percent, and it was considered nearly impossible to achieve the required homogeneity. In addition, the magnitude of the magnetic field over the sample volume was required to remain within the maximum variation in field homogeneity over the period of time that the experiment was conducted. For solid-state NMR applications, the requirements for homogeneity and field stability are not as demanding because resonance linewidths have a minimum dispersion of approximately 0.5 ppm, which cannot be removed by MASS. A magnetic field strength of 7 T is suitable for many solid-state NMR experiments. Higher fields (>11 T) are desirable for studies of quadrupolar nuclei because the spectra are simpler to interpret.

It was evident early on that the homogeneity of the magnetic field was of critical importance in resolving resonances that were different because of small variations in the electronic environment around the nucleus. For solid-state NMR, the spatial uniformity of the magnetic field over the sample volume should be moderately high, with a variation of approximately 0.000001% (0.01 ppm). This requirement is easily achieved without applying external field corrections in modern magnets. A temporal magnetic field variation of less than 0.02 ppm h^{-1} is typical for modern superconducting magnets and is suitable for many solid-state NMR experiments conducted over several days without field drift adjustments.

3.2 Spectrometer Console

A spectrometer console for solid-state NMR applications consists of a pulse sequence generator, high-power radiofrequency amplifiers, a fast switching device for routing high-power pulses to the probe and low-power NMR signals from the probe, and a receiver unit for amplifying, filtering, and digitizing the NMR signals. A simplified scheme for a solid-state NMR spectrometer apparatus is shown in Figure 16. The low-power CW oscillator (A) generates four quadrature radiofrequency outputs. Each output is gated on or off by the pulse

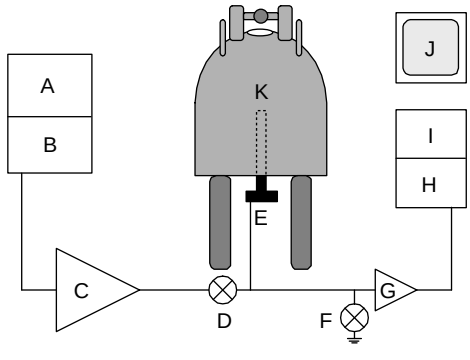

Figure 16 Simple scheme of a solid-state NMR spectrometer apparatus. (A) Low power radiofrequency continuous wave (CW) oscillator; (B) pulse programmer; (C) high-power amplifier; (D) pass-through crossed diodes; (E) probe; (F) grounded crossed diodes; (G) preamplifier; (H) receiver; (I) digitizer; (J) computer; (K) magnet.

programmer (B) according to a program called a pulse sequence. The high-power amplifier (C) increases the peak-to-peak voltage ($V_{p\text{-}p}$) of the radiofrequency pulse by a factor of about 200. The output of the amplifier ($\sim$200–300 $V_{p\text{-}p}$) is directed to a three point junction via pass-through crossed diodes (D). The crossed diodes are made by connecting ten fast-switching diodes in parallel but alternating (crossed) polarity. At the junction, the radiofrequency pulses are directed to the probe (E) because the grounded crossed diodes (F) cause a high impedance at the junction as seen by the preamplifier (G). Effective routing of the high-power radiofrequency pulses to the probe and not the preamplifier requires that the length of each cable between the junction and the amplifier, probe, and grounded crossed diodes be 1/4 of the wavelength of the radiofrequency wave in the cable.[65–68] The NMR signal (1–20 $\mu V_{p\text{-}p}$) that proceeds from the excitation pulse follows the low impedance path to the preamplifier where it is amplified and passed to the receiver (H).[69,70] The receiver heterodynes, filters, and amplifies the NMR signal before sending it to the digitizer (I). The digitizer samples the CW signal in quadrature at equally spaced intervals. The sampling rate is twice the frequency span that fully contains the NMR spectrum. The computer (J) is interfaced to the pulse programmer and the digitizer. It is used to upload pulse sequences to the pulse programmer and download data from the digitizer. The computer also processes and stores the NMR data.

3.3 Nuclear Magnetic Resonance Detectors

The selection of a probe for solid-state studies begins with a consideration of the systems that will be investigated (e.g. single crystals or powders, rigid or semi-rigid materials, gas/solid mixtures, etc.) and the experiments that will be performed (e.g. static or spinning, multinuclear observation, dipolar decoupled, etc.). Static, single-coil probes are the easiest to build and least expensive. For most solid-state NMR experiments, however, sample spinning will be desirable and required for recording high-resolution NMR spectra of powders. At a minimum, a solid-state spectrometer should be equipped with one spinning probe that can also be used in nonspinning mode for wideline experiments. The materials that are used to fabricate the stator, housing, rotor, and rotor caps of a MAS probe can be selected depending on the nuclei that will be studied. Probes for detecting NMR signals in solids require hard ceramic and plastic materials that can withstand high temperatures, and large centrifugal forces from spinning samples. These materials can produce background signals that can interfere with the samples under study. There are several combinations of ceramic and plastic materials that are used to build MAS probes. The materials are selected to minimize background signals from different nuclei, allow the highest spinning speeds, or give the largest temperature range for variable temperature studies. A combination of materials can be specified based on the type of studies that will be performed. Standard combinations are available for general-purpose probes. Table 1 includes materials that are currently used to fabricate stators, rotors, and caps for solid-state MAS NMR probes.[71]

Static probes do not have rapidly rotating parts. This type of probe includes wideline, single crystal, and several imaging probes. The wideline probe is used to record broad NMR spectra. It features a very fast recovery time following a radiofrequency pulse for nuclear spin excitation.[72,73] It is also designed to withstand 1–2 kW of pulsed (1–100 μs) and 100–400 W of continuous (1–100 ms) radiofrequency power. Probes for studying single crystals use a miniature goniometer to rotate the crystal within the NMR detector coil about an axis that is perpendicular to $\mathbf{B}^{ext}$.[74] Solid-state imaging probes are similar to conventional imaging probes but are built to withstand higher radiofrequency power and are equipped with powerful (1000 G cm^{-1}) pulsed magnetic field gradients.[75]

Table 1 Materials for fabrication of solid-state MASS NMR probes

Material	Major elements	Minor elements	Maximum temperature for caps (or rotors) (°C)	Range of spin rates for caps (or rotors) (kHz)
Silicon nitride	98% Si_3N_4	2% Y_2O_3	(1400)	(8–20)
Sapphire	99.95% Al_2O_3		(1400)	(5–9)
Zirconia	94% ZrO_2	3% Mg, Hf, 0.3% Si	(650)	(7–14)
Macor	Al, Si, O, B, K	2% F	250	6–13
			(250)	(5–8)
CFR-PEEK	H, C, O		200	<5
Vespel	H, C, O	N	240	7–22
PEK	H, C, O		150	<5
Aurum	H, C, O	N	80	7–20
Kel-f	F, C, Cl		70	4–13

CFR-PEEK, carbon fiber-reinforced polyetheretherketone; PEK, polyetherketone.

Spinning probes are designed to rotate a sample contained in a rotor rapidly in one of three ways: at one angle (the magic angle, 54.74°), at several different angles sequentially, or at two angles simultaneously. A typical standard MAS probe is designed for a specific magnetic field strength and is capable of tuning to nuclei on the observe channel (or X channel) from ^{103}Rh to ^{31}P. The decoupler channel of such a probe is usually tunable over a much narrower frequency range that includes ^{1}H and ^{19}F nuclei. Broadband MAS probes provide the best performance at higher frequencies. If a considerable amount of work will be done at only one frequency, then a probe optimized for that frequency should be considered. The size of the rotor is also an important specification. For general use, a 5- or 7-mm-diameter rotor is common. However, if a substantial number of studies will require samples of limited quantity (a few milligrams), then a MAS probe with a 2.5-mm diameter rotor should be considered to maximize the filling factor (the sample-to-coil volume ratio) and, therefore, the sensitivity.[76] The maximum spin rate of a probe is an important specification, but it should not be considered independently of rotor size and material. The rotors with the highest speed specifications are usually the smallest. If the amount of sample is not limited and the fastest available spinning speed not crucial for most experiments, then the larger slower rotors are more desirable. The temperature range for a standard spinning probe is −120 to +160 °C. The range can be extended, −180 to +250 °C, by a suitable change in design and probe material. A high-temperature probe is capable of reaching 650 °C, while a low-temperature probe can spin at −267 °C. A photograph of a MAS NMR probe head is shown in Figure 17. The outer diameter of the rotor is 7 mm and it is approximately 18-mm long. The probe is designed to operate continuously at a rotation frequency of 4 kHz.

Variable angle spinning (VAS) probes are similar in design to MAS probes. However, the spinning axis can be manually adjusted from 0° to 90° instead of a narrow range (±10°) near 54.74°. The angle is held fixed in a VAS probe throughout an experiment. The stator of switched angle spinning (SAS) probes can be computer manipulated via a rapid servo (in less than 50 ms) between two angles, usually 30° and 54.74°. In this probe, the coil follows the rotor, giving the optimal filling factor. The leads that connect to the movable coil require periodic replacement. The coil in a dynamic angle spinning (DAS) probe is stationary, and the spinning axis of the internal rotor can change to any orientation (360°) with respect to $\mathbf{B}^{ext}$. The maximum flip time is 20 ms, with a position accuracy of 0.4°.[71] The advantages of the DAS probe for experiments that require rapid changes between different spinning axes are high tuning stability (because the coil is fixed) and longer lifetime, in terms of number of flips. The primary disadvantage is that the minimum coil diameter is limited to the dimension of the rotor length, resulting in reduced filling factors and greater requirements for radiofrequency power to generate strong $\mathbf{B}_1$ fields. Double rotation (DOR) probes simultaneously spin two rotors, one within the other, to achieve high-resolution NMR spectra for quadrupolar nuclei. Stable rotation requires that a ratio of 1 : 5 be established for the spin rate of outer to inner rotors. The inner rotor contains the sample (100-μL volume), rotates at 6 kHz, and is tilted 30.12° with respect to the axis of the outer rotor. The outer rotor contains the spinning inner rotor, rotates at 1.2 kHz, and is tilted 54.74° with respect to $\mathbf{B}^{ext}$.

Figure 17 MASS NMR probe head manufactured by Doty Scientific, Inc. The rotor is shown suspended outside the stator. During operation, a ceramic top covers the rotor and is secured to the stator by two small knurl nuts. The stator assembly is inclined at the magic angle with respect to $\mathbf{B}^{ext}$. The probe head is mounted inside a protective metal cylinder called the probe body, which also serves as a radiofrequency shield.

Commercial spinning probes are fabricated from different materials, depending on the application, to eliminate or reduce specific background signals associated with the stator, housing, rotor, and rotor caps. For example, if a series of studies requires ^{29}Si NMR, a rotor constructed from sapphire would produce less silicon background than one made of alumina. In addition, the stator assembly supports the radiofrequency coil, and the material it is made from will give broad background signals. A zirconia stator (95% ZrO_2, 3.3% MgO, 1.5% HfO_2, 0.06% SiO_2, 0.03% Al) is common for general-purpose spinning probes and minimizes background signals from aluminum and silicon. An alumina stator (99.9% Al_2O_3, 0.2% SiO_2) has a low dielectric loss and therefore is more suitable for applications such as CRAMPS (combined rotation and multiple-pulse spectroscopy) that require series of pulses with large field strengths. The higher thermal conductivity of alumina makes it the best choice for low-temperature studies. A silicon nitride stator has superior electrical properties (lowest dielectric loss and permittivity), making it the

best choice for most applications, except those involving silicon and nitrogen. Even so, given the long T_1 time for silicon and the low natural abundance of nitrogen, this probe may still be suitable for some experiments involving these nuclei. The housing material is further removed from the detector coil, but it can still contribute to broad background signals. Its main purpose is to provide temperature control.

Probe materials are also selected for different operating temperatures. Macor (Al, Si, O, B, K, F) is used in probes that will detect carbon and proton signals with a maximum operating temperature below 160 °C. A Vespel (H, C, O, N) housing is useful for temperatures up to 300 °C, but it will provide a significant carbon background signal. The plastic Aurum (H, C, O, N) can be used to replace Vespel. A probe housing made of zirconia increases the temperature range up to 650 °C but is also more costly. Kel-f (*p*-chlorotrifluoroethylene) is a standard housing material that contains F, Cl, and C; however, the carbon signal is very broad because of C–F dipolar couplings. For fluorine studies, PEK (polyetherketone; H, C, O) can replace Kel-f and offers a larger operating temperature range (−180 to +225 °C). Glass dewars are commonly used for housings in probes for liquids but are rarely used in solid-state NMR probes because of the substantial background signal from Si, Al, Na, K, and other components of glass.

While the probe stator and housing can provide unwanted background signals that are usually broad, the rotor and rotor end caps can produce sharp background signals because these components, like the sample, are rotated rapidly and anisotropic spin interactions are averaged. In some cases a narrow background signal from a rotor can be used as an external reference for chemical shift and quantitation. In other cases, the signal from the rotor interferes with the desired spectral data and a different rotor material must be selected. Rotors are made from silicon nitride (Si_3N_3), sapphire (Al_2O_3), zirconia (ZrO_2), Macor (Al–Si glass), and some other materials. Silicon nitride rotors are the most expensive, about twice the cost of zirconia rotors, and are used for high spinning speeds. Sapphire rotors are about the same cost as zirconia rotors, are made of single crystals of Al_2O_3, and are translucent. They are useful for low-temperature studies, experiments that require the sample to be exposed to visible radiation, and NMR investigations of zirconium samples. Zirconia rotors are the most common, have high impact strength, can spin at speeds up to about 14 kHz, and are durable. Macor rotors are the least expensive, about 25% less than zirconia rotors, and are made of a glass material. These rotors are suitable for wet or air-sensitive samples. Samples of compressed gases (up to 50 atm) with solids or air-sensitive materials have been sealed in glass tube ampoules, placed inside rotors, and effectively spun.[77] The low-density plastic CFR-PEEK (carbon-fiber-reinforced polyetheretherketone) is used for DOR rotors.

It is also important to consider the material that is used to cap the rotor. Rotors are usually made in either of two forms: a hollow cylindrical sleeve or a cylindrical closed-end tube. The hollow cylindrical sleeve requires a turbine cap to drive the rotor and a simple plug cap to seal the opposite end. For higher spinning speed probes, the plug cap can be replaced by a second turbine cap. The closed-end tube rotor only requires a plug cap if the sealed end is machined into a turbine; otherwise, a turbine cap is used. The plug or turbine caps are made from plastic materials, e.g. Kel-f, Vespel, and Aurum. Kel-f produces F, Cl, and C background signals and is useful over a limited temperature range (−20 to +70 °C). Vespel is a harder material with a larger temperature range (−270 to +80 °C) than Kel-f. Vespel can withstand higher spinning speeds compared to other cap materials; however, it produces C and H background signals. Aurum is similar to and can replace Vespel for fabricating caps that have a temperature range of −30 to +80 °C. Macor or other hard ceramic materials are also used; machined grooves in these caps accommodate soft O-ring seals. The Viton O-rings in these caps can operate in the range −150 to +250 °C but produce C and H background signals. Usually the plug or turbine cap limits the temperature of the experiment.

4 APPLICATIONS OF NUCLEAR MAGNETIC RESONANCE TO SOLIDS

There are numerous applications of NMR to solid materials in the physical, chemical, and biological literature. Three very different applications are presented below. They include examples in polymers, zeolites, and metals.

4.1 Wideline Nuclear Magnetic Resonance Spectroscopy of Polymers

The physical properties of solids and semisolids are partially determined by the temperature-dependent mobility of molecules. One-dimensional wideline NMR spectroscopic and relaxation data are diagnostic of different types of isotropic and anisotropic hindered molecular motions. Examples of molecular motions in solids that have been investigated by NMR include methyl group rotations, small-angle librations, 180° phenylene ring flips, helical rotations, and translations. In solids that are not dominated by an extended network of nuclear dipolar interactions, the temperature-dependent lineshape of a chemical shielding powder pattern can provide convincing evidence for well-defined anisotropic molecular

rotations. NMR spin-relaxation experiments can probe the motion of molecules or segments of molecules over the timescale 10^{-10} to 1 s. Polymers cover a range of materials with physical properties of solids and semisolids. Investigations of polymers by one-dimensional wideline NMR spectroscopy and nuclear spin relaxation experiments illustrate the specificity of information that can be obtained from a collection of easily implemented standard experiments. McBrierty and Packer have given details of the general application of solid-state NMR methods to polymer research.[78] More advanced two-dimensional NMR experiments provide specific details about molecular motions such as the exchange between crystalline and amorphous polymer phases and are discussed in detail by Schmidt-Rohr and Spiess.[79]

A simple one-pulse NMR experiment on polymer samples with a high concentration of magnetic nuclei with $I = 1/2$, e.g. ^{1}H, ^{19}F, and ^{13}C, usually results in a very broad featureless spectrum. The broad line is the result of a complex extended network of nuclear spins that interact with each other via the magnetic dipole–dipole interaction. The linewidth and spin relaxation parameters T_1, T_2, and $T_{1\rho}$ can be measured as a function of temperature. These data provide information about molecular motion and phase transitions in solids.[80–82] For the case of dilute spins coupled to a network of abundant spins, as is found in many organic compounds, the ^{13}C NMR wideline spectrum can reveal chemical shift anisotropies, provided that the effect of the network of ^{1}H nuclei can be removed. This can be accomplished by acquiring the ^{13}C NMR signal while simultaneously irradiating the broad ^{1}H resonance with a radiofrequency magnetic field that is several times larger in magnitude than the ^{1}H linewidth. For solids with only a few distinguishable ^{13}C nuclei, the powder patterns for the different ^{13}C sites will show a temperature dependence that can partially, and sometimes fully, characterize the type of molecular motion (rotation, libration, reorientation) that takes place in the solid.

The ^{19}F NMR spectra of PTFE (poly(tetrafluoroethylene)), at different temperatures, recorded under conditions of homonuclear dipolar decoupling, provide an interesting example of two superposed chemical shielding powder patterns, as shown in Figure 18.[83] The broad asymmetric powder pattern recorded at 259 °C was attributed to the ^{19}F nuclei in crystallites imbedded in amorphous polymer. At high temperatures, the ^{19}F nuclei in the amorphous phase produce a narrow resonance because the mobility of these nuclei averages their anisotropic shielding tensors. Thus, the fractions of the polymer sample in the amorphous and crystalline phases were determined by decomposition of the observed spectra into a sharp symmetric resonance and a broad asymmetric powder pattern. At very low

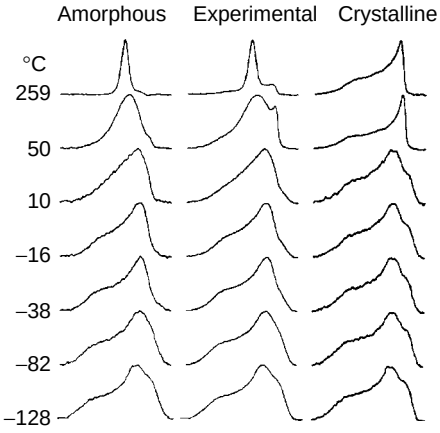

Figure 18 Solid-state ^{19}F-NMR spectra of a mixture of the amorphous and crystalline phases in PTFE at different temperatures. The experimental spectra were decomposed into chemical shielding power patterns representative of both polymer phases. An analysis of the temperature dependence of the crystalline phase component spectra determined that near room temperature the molecules rotate about the polymer chain axis. The polymer sample was 68% crystalline. (Adapted from Figure 7 in A.J. Vega, A.D. English, 'Multiple-Pulse Nuclear Magnetic Resonance of Solid Polymers. Polymer Motions in Crystalline and Amorphous Poly(tetrafluoroethylene)', *Macromolecules*, **13**, 1635–1647 (1980).)

temperatures, the experimental spectra were decomposed into two identical powder patterns. This indicated that the chemical shielding tensors and, therefore, the chemical environments of the ^{19}F nuclei in the amorphous and crystalline phases were the same. More important, the temperature dependence of the transformation from an axially symmetric (at 28 °C) to an axially nonsymmetric (at 14 °C) tensor powder pattern for the crystalline component spectra can be fitted to a model of rotational diffusion about the polymer chain axis. Thus, the axially nonsymmetric chemical shielding tensor for the ^{19}F nuclei in the crystalline phase was made to appear axially symmetric by rotation of the ^{19}F nuclei about the chain axis. The homonuclear decoupling scheme in this study could not be continuous because the ^{19}F nuclei required simultaneous detection. A multiple-pulse decoupling sequence was used effectively to decouple the ^{19}F nuclei from each other via a series of discrete pulses between which the data points for the spectra were acquired.

4.2 Mobile Guests in Solid Host Material

Many interesting heterogeneous systems (e.g. gas/solid, and liquid/solid) are only partially composed of solids at ambient temperature and pressure. They are included here because the analysis of such systems requires understanding of interactions in the solid state and the

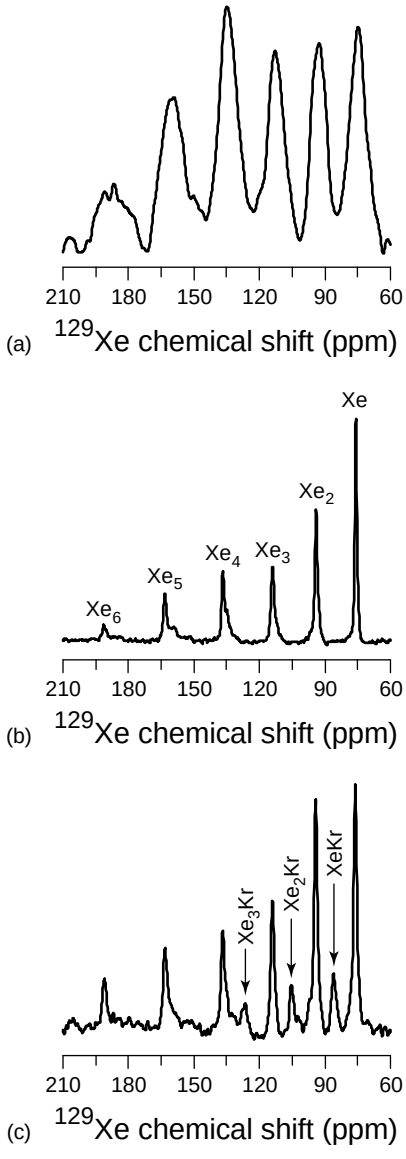

Figure 19 ^{129}Xe-NMR spectra of xenon atoms occulded in zeolite NaA. (a) Static wideline spectrum. (b) MASS spectrum recorded with 2 kHz spin rate. (c) MASS spectrum of coadsorbed xenon and krypton. (Adapted from Figure 1 in A.K. Jameson, C.J. Jameson, A.C. de Dios, E. Oldfield, R.E. Gerald II, G.T. Turner, '^{129}Xe Magic-angle Spinning Spectra of Xenon in Zeolite NaA: Direct Observation of Mixed Clusters of Co-adsorbed Species', *Solid State Nucl. Magn. Reson.*, **4**, 1–12 (1995).)

appropriate experimental techniques to average them. Xenon atom clusters trapped in the three-dimensional array of alpha cages in zeolite NaA provide a lucid example of how the local anisotropic cage environment determined by the host can influence the NMR spectrum of the guests. Solid-state NMR spectroscopy of zeolite hosts will not be discussed here. Engelhardt and Michel give a detailed description of solid-state NMR investigations of zeolite hosts.[60] The wideline (static) ^{129}Xe NMR spectrum of xenon atoms occluded in zeolite NaA is shown in Figure 19(a).[77] Distinct broad peaks have been observed for xenon atom clusters ranging in size from 1 to 8 atoms per alpha cage.[84] The combination of slow intercavity xenon atom exchange and large frequency shifts between the different ^{129}Xe peaks makes it possible to resolve the different Xe_n peaks. However, the peaks do not have a Lorentzian lineshape; the approximately 10-ppm peak widths are not consistent with the measured spin–spin relaxation time constants of 15–40 ms and selective inversion of narrow regions of a single peak is possible.[85] Therefore, the peaks could be a superposition of narrow Lorentzian resonances that have a range of positions covered by the width of the peak. The xenon atoms move rapidly within the alpha cages, even at low temperatures, so that the ^{129}Xe nuclei in a cluster exhibit a resonance frequency that represents an average over the possible cluster configurations. The set of configurations that each cluster Xe_n may have is determined by the multidimensional potential energy surface in configuration space defined by n interacting xenon atoms in the presence of the zeolite cage. For every configuration, each of the n xenon atoms in the Xe_n cluster has an associated shielding tensor. The resonance frequency of the ^{129}Xe nuclear spin of each atom in a configuration at any instant is determined by the orientation of its shielding tensor with respect to $\mathbf{B}^{ext}$. As the configurations of the clusters interchange, the shielding tensor and its orientation for each xenon atom in the configuration changes, resulting in a different resonance frequency for each spin. The occurrence of each configuration of n atoms in a Xe_n cluster is weighted by a Boltzmann factor and so too are the associated resonance frequencies for each member spin.

For individual clusters of xenon atoms occluded in zeolite NaA at room temperature, the rate of interconversion of configurations, including permutations of the xenon atoms among the n particle coordinates for a single Xe_n cluster configuration, is fast compared to the nuclear precession coherence lifetime of the ^{129}Xe spins. In other words, during the characteristic coherence dephasing time T_2 each xenon atom samples every particle position of every configuration, in the configuration space of the cluster, at least once. Therefore, an identical average resonance frequency is observed for each of the n xenon atoms in a Xe_n cluster. Furthermore, the average resonance frequency of a Xe_n cluster is dependent on the orientation of the noncubically symmetric alpha cage in $\mathbf{B}^{ext}$. For a powder sample of Xe/NaA, all orientations of the alpha cage are equally probable and each Xe_n cluster exhibits an inhomogeneous chemical shielding dispersion. Since the anisotropy of the alpha cage reveals the configuration

weighted anisotropic xenon chemical shielding tensor, the MASS technique is suitable for reducing the width of the Xe_n cluster peaks. Figure 19(b) shows the effect of MASS at 2 kHz on the same Xe/NaA sample used for the static spectrum (Figure 19a). The enhanced resolution afforded by MASS makes it possible to investigate binary mixtures of xenon and krypton atoms in NaA, as shown in Figure 19(c).

4.3 Metals

Nearly all metals have an isotope with a nonzero nuclear spin quantum number. Therefore, in principle, it is possible to record NMR spectra for these nuclei. In fact, a considerable amount of research on NMR properties of metals was conducted beginning in the 1950s. A monograph by Winter covers the fundamental properties of metals that influence their NMR spectra and provides several experimental results.[86] This section broadly summarizes several of the important factors that determine the resonance frequency and spin-lattice relaxation time of the NMR signal.

NMR in metals was described in 1949 as producing narrow resonances comparable to those observed in liquids.[87,88] While molecules in the liquid state tumble isotropically and in so doing, average out anisotropic spin interactions, atoms in a metal are fixed in the solid crystalline lattice and maintain a fixed orientation when placed in an external magnetic field. However, the conduction electrons rapidly move throughout the crystallite and produce an average fluctuating electronic environment at the nuclei. An estimate of the correlation time for a conduction electron moving from one unit cell to another is approximately 10^{-16} s; thus, the NMR nuclei with precessional periods of 10^{-7} to 10^{-9} s respond to an average electronic environment. That is, the conduction electrons average out the electronic shielding anisotropies that would cause inhomogeneous line broadening, and they move fast enough so that they do not shorten the lifetime of a nuclear spin in an excited state. This explains the narrow lines and lack of inhomogeneous broadening in NMR studies of metals.

Because the conduction electrons near the top of the Fermi distribution are not always spin paired, residual net magnetization from the magnetic moments of the electrons can exist at the nuclei. Knight used Fermi-Dirac statistics to predict the average net magnetization from the conduction electrons and showed that the nuclei would experience an additional field KB^{ext}, where K is the Knight shift and is usually positive. A measurement of the Knight shift can determine the number of conduction electrons per nucleus (i.e. metal atom) and their electronic wavefunctions. Hence, the Knight shift provides a measure of the coupling constant for the conduction electron/nucleus interaction.

Unpaired conduction electrons near the Fermi surface also contribute to magnetic field fluctuations at the nucleus. These fluctuating fields determine the spin-lattice relaxation time constants for the metal nuclei.[89] The conduction electron relaxation effect for metal nuclei is weak because the frequency of magnetic field fluctuations is seven orders of magnitude higher than the nuclear Larmor frequency for typical laboratory fields. Figure 20(a) shows an NMR image of two forms of lithium metal in a battery. Both resonances exhibit a Knight shift near 260 ppm. The dendritic lithium was electrochemically deposited on the central carbon conductor. The two forms of lithium metal differ in Knight shift by 18 ppm. The ^{7}Li Knight shift for an AlLi alloy is

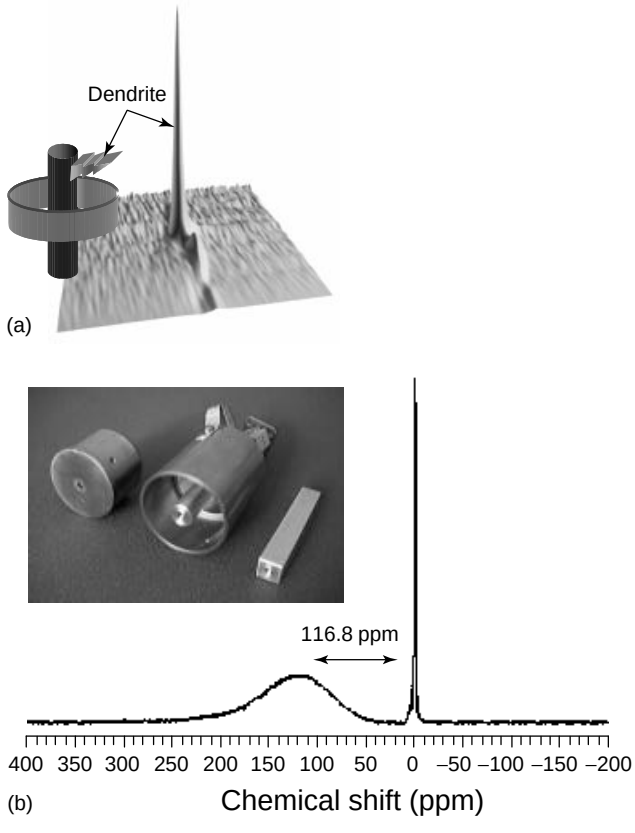

Figure 20 (a) NMR image of two forms of lithium metal in a battery (dendritic and counter electrode). Both resonances exhibit a Knight shift near 260 ppm, but differ from each other by 18 ppm. The dendritic lithium was electrochemically deposited on the central carbon conductor and has the weaker Knight shift. (b) The ^{7}Li Knight shift for an AlLi alloy made by Reynolds Aluminum and used for jet airplane wings. The central conductor shown inside the cavity was the sample. A capillary tube containing 1 molal aqueous LiCl was placed next to the central conductor and used as a chemical shift reference. It also served to quantify the amount of Li that was detected in the surface layer of the sample.

shown in Figure 20(b). The sharp resonance is an external chemical shift reference of 1 molal LiCl (0 ppm).[90–92]

Although some of the first NMR experiments were conducted on magnetic materials, many NMR spectroscopists dismiss magnetic materials as too problematic for NMR analysis.[93] In magnetic materials, the radiofrequency $\mathbf{B}_1$ field experienced by the nucleus with an unpaired electron is enhanced by as much as 10^6. The $\mathbf{B}_1$ field produced by the radiofrequency coil couples with the magnetization created by the unpaired electron. The result is an effective $\mathbf{B}_1$ field at the nucleus that can be dominated by the magnitude of the electronic magnetization. Thus, a weak radiofrequency $\mathbf{B}_1$ field strongly couples to an electronic magnetic moment and the two fields add together and effect NMR transitions. Hence, the magnetic field experienced by the nucleus is largely the result of the electronic magnetic moment, which is made time-dependent by coupling to the radiofrequency $\mathbf{B}_1$ field. Figure 21(a) shows the ^{57}Fe NMR power saturation curve for the iron nuclei in a surface layer of a large iron rod that also functions as the NMR detector element in a toroid cavity probe.[94–99] Figure 21(b) shows a representative ^{57}Fe NMR spectrum of the rod. The NMR spectra were recorded at room temperature in a hyperfine field of 33.5 T.[100]

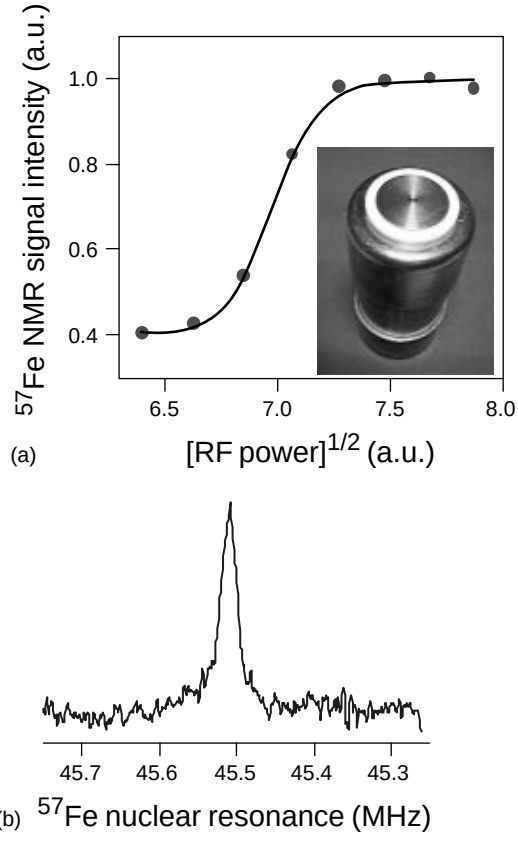

Figure 21 (a) ^{57}Fe-NMR power saturation curve for the iron nuclei in a surface layer of a large (15/16-in. diameter, 3-in. long) iron rod that also serves as the central conductor in a toroid cavity probe. (b) Representative ^{57}Fe-NMR spectrum of a thin (~40 μm) layer at the surface of the rod. The rod was demagnetized. All experiments conducted at room temperature in 33.5 tesla internal field.

5 PERSPECTIVE AND FUTURE DEVELOPMENTS

While NMR enjoys a high stature among spectroscopic methods because of the detailed information it can provide about composition, structure, mobility, quantity, and so on of molecules in all states of matter, its Achilles' heel is lack of sensitivity. Sensitivity enhancement techniques that have been developed usually have restrictions preventing general use; nonetheless, they are critical to various applications. Some of these methods include higher magnetic field, hyperpolarized gases, and optical methods.[101–106] The development of hardware with enhanced performance continues for conventional probes by employing tailor-made materials. Cold probes offer additional sensitivity because they operate at cryogenic temperatures where the Johnson noise is reduced. Enhancements to the spectrometer console include more sensitivity from the preamplifier and receiver, greater expandability for additional transmitter as well as receiver channels, smaller footprints for self-shielding magnets, and portable NMR units.

Several applications of NMR methods for the analysis of a material resemble the approach taken by analytical chemists. The material is first separated into chemical components. Second, these components are identified. Finally, the relative amounts of the components are determined. Similarly, in multidimensional NMR experiments, the first step is to separate the resonance signals based on knowledge of the spin interactions in the material. Second, the NMR spectrum is recorded for specific spin interactions and used to identify a component. Finally, this spectrum can be analyzed quantitatively because there is a one-to-one correspondence between the signal intensity and the amount of the component of interest.

The analytical chemist is interested in determining the composition of an unknown substance in terms of molecular composition and quantity of each component. An elemental analysis provides the quantity of the elemental constituents of the unknown substance. This analysis usually requires that a portion of the sample be destroyed in the process. Mixtures require separation before individual analysis and identification of the constituents. The sensitivity of techniques generally available to the analytical chemist ranges from picograms of identified substance per gram of unknown to a fractional percentage of the same. Techniques exist for determining the molecular and

atomic structure. Morphology, including size and shape of particles and defects, twinning, and larger features of the bulk solid can also be important properties of the material that in part determine their function and efficacy in a process. The chemical state or environment (e.g. oxidation state) is an important property that is necessary to characterize a material. Sample issues that arise in an analysis include uniformity, quantity, stability to temperature, pressure, and so on, and physical properties. For example, a high vapor pressure at room temperature may preclude analysis of a material in ultrahigh vacuum by electron microscopy at room temperature. Many techniques are invasive and result in complete destruction of the sample. This can present a severe limitation if the sample is available only in limited quantities.

Existing techniques can provide complementary and often comparable information for solid-state NMR analyses. Some examples are X-ray diffraction of powders and single crystals, neutron scattering, synchrotron radiation scattering, electron microscopy, and rheology techniques such as dielectric and mechanical relaxation measurements. Unlike some of these techniques, analysis of solids by NMR is noninvasive and leaves the sample available for subsequent analysis by invasive methods. While analysis by solid-state NMR does require a minimum sample weight of tens of milligrams, compared to a few picograms for analytical techniques such as atomic absorption spectroscopy, for cases where sample quantity in not severely limited, NMR can provide information that characterizes the sample in ways other than elemental composition. NMR allows the analyst to probe the sample selectively for information. For example, ^{1}H NMR could provide detailed site information and distribution of the residual protons in a sample that was perdeuterated.

Selectivity and separation of spectral information is increasing through the growing repertoire of multidimensional methods that provide added spectral resolution by expanding the NMR information into additional dimensions. These dimensions can be associated with molecular translation, chemical bonding, or coordination number. Many times these additional dimensions provide information that is not accessible by any other technique, such as structure–function correlations. It is important to recognize that one of the principal advantages of NMR is that it not only provides an electromagnetic spectrum (over a narrow range of energies) of a system (as in other forms of spectroscopy), but that it also permits correlations between different parts of the spectrum. The correlations can establish incontrovertible proof of atomic connectivity through chemical bonds and proximity through space. In addition, correlations between specific parts of molecules and well-defined rotational or librational motions can be ascertained. In short, it is the extensive selection, separation, and correlation of spectral signatures that separates the NMR method from its competitors.

As in most analyses, however, a combination of complementary techniques proves most useful for complete resolution of a problem. As an example, the determination of the structure of a crystal by X-ray diffraction provides the number, type, and position of atoms and ions in the unit cell, but not their electronic connectivity nor mobility, which might be inferred. The application of NMR methods in conjunction with the X-ray data can provide convincing details of atomic connectivity, methyl group rotation, proton bonding (including hydrogen bonding), and trapped liquids. NMR offers the advantage of measuring the weak interactions between nuclei that allow molecular features to be correlated. Analyses of these correlations form a picture of atomic connectivity, molecular motion and function, and molecular arrangements from the atomic to the microscale.

ACKNOWLEDGMENTS

Professor Cynthia J. Jameson of the University of Illinois in Chicago and Professor Ulrich Haeberlen of the Max-Planck-Institut für Medizinische Forschung in Heidelberg are gratefully acknowledged for allowing me a sojourn in Germany during which I learned about second rank tensors. Jerome W. Rathke (group leader, CMT-ANL) is greatfully acknowledged for fostering a unique research environment that made it possible for me to pursue this article. Editorial assistance and helpful comments from Jane Andrew (IPD-ANL) are gratefully acknowledged. The US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, supported this work under Contract W-31-109-Eng-38.

ABBREVIATIONS AND ACRONYMS

CFR-PEEK	Carbon Fiber-reinforced Polyetheretherketone
CRAMPS	Combined Rotation and Multiple-pulse Spectroscopy
CW	Continuous Wave
DAS	Dynamic Angle Spinning
DOR	Double Rotation
FFT	Fast Fourier Transform
FID	Free Induction Decay
MAS	Magic Angle Spinning
MASS	Magic Angle Sample Spinning
NMR	Nuclear Magnetic Resonance
PEK	Polyetherketone

PTFE	Poly(tetrafluoroethylene)
SAS	Switched Angle Spinning
VAS	Variable Angle Spinning

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Magnetic Resonance Angiography • Magnetic Resonance Imaging, Functional • Magnetic Resonance in Medicine, High Resolution Ex Vivo • Magnetic Resonance, General Medical • Multinuclear Magnetic Resonance Spectroscopic Imaging

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Chemical Weapons Chemicals Analysis **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Coatings **(Volume 2)**
Nuclear Magnetic Resonance of Coating and Adhesive Systems

Environment: Water and Waste **(Volume 4)**
Nuclear Magnetic Resonance for Environmental Monitoring • Organic Analysis in Environmental Samples by Electrochemical Methods

Food **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials • Water Determination in Food

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Polymers and Rubbers **(Volume 9)**
Nuclear Magnetic Resonance, Imaging of Polymers • Nuclear Magnetic Resonance, Solid State in Analysis of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Nuclear Magnetic Resonance and Magnetic Resonance Imaging for Process Analysis

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis • Metal Analysis, Sampling and Sample Preparation in • Nuclear Magnetic Resonance in Metals Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance • High-performance Liquid Chromatography Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Nuclear Magnetic Resonance of Geological Materials and Glasses • Parameters, Calculation of Nuclear Magnetic Resonance • Quadrupolar Nuclei in Solid-state Nuclear Magnetic Resonance • Quadrupole Couplings in Nuclear Magnetic Resonance, General • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solid-state Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-dimensional Nuclear Magnetic Resonance of Small Molecules • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules • Zeeman Interaction in Nuclear Magnetic Resonance

REFERENCES

1. C.J. Gorter, 'Negative Result of an Attempt to Detect Nuclear Magnetic Spins', *Physica (Utrecht)*, **3**, 995–998 (1936).
2. G. Breit, I.I. Rabi, 'Measurement of Nuclear Spin', *Phys. Rev.*, **38**, 2082–2083 (1931).
3. E. Majorana, 'Atomi Orientati in Campo Magnetico Variable', *Il Nuovo. Cimento.*, **9**, 43–50 (1932).
4. W. Cohen, 'The Nuclear Spin of Caesium', *Phys. Rev.*, **46**, 713–717 (1934).
5. S. Millman, 'On the Nuclear Spins and Magnetic Moments of the Principal Isotopes of Potassium', *Phys. Rev.*, **47**, 739–746 (1935).
6. I.I. Rabi, J.M.B. Kellogg, J. Zacharias, 'The Magnetic Moment of the Proton', *Phys. Rev.*, **46**, 157–163 (1934).
7. I.I. Rabi, J.M.B. Kellogg, J. Zacharias, 'The Magnetic Moment of the Deuton', *Phys. Rev.*, **46**, 163–165 (1934).

8. I.I. Rabi, S. Millman, P. Kusch, J.R. Zacharias, 'The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments', *Phys. Rev.*, **55**, 526–535 (1937).
9. S. Millman, M. Fox, 'Nuclear Spins and Magnetic Moments of Rb^{85} and Rb^{87}', *Phys. Rev.*, **50**, 220–225 (1936).
10. J.H. Manley, 'The Nuclear Spin and Magnetic Moment of Potassium (41)', *Phys. Rev.*, **49**, 921–924 (1936).
11. J.H. Manley, S. Millman, 'The Nuclear Spin and Magnetic Moment of Li^6', *Phys. Rev.*, **51**, 19–21 (1937).
12. I.I. Rabi, 'On the Process of Space Quantization', *Phys. Rev.*, **49**, 324–328 (1936).
13. J. Kellogg, I. Rabi, J. Zacharias, 'The Gyromagnetic Properties of the Hydrogens', *Phys. Rev.*, **50**, 472–481 (1936).
14. H. Torrey, 'The Sign of the Magnetic Moment of the K^{39} Nucleus', *Phys. Rev.*, **51**, 501–508 (1937).
15. T.E. Phipps, O. Stern, 'Uber die Einstellung der Richtungsquantelung', *Z. Physik*, **73**, 185–191 (1931).
16. R. Frisch, E. Segrè, 'Uber die Einstellung der Richtungsquantelung II', *Z. Physik*, **80**, 610–616 (1933).
17. P. Güttinger, 'Das Verhalten von Atomen im Magnetischen Drehfeld', *Z. Physik*, **73**, 169–184 (1931).
18. I.I. Rabi, 'Space Quantization in a Gyrating Magnetic Field', *Phys. Rev.*, **51**, 652–654 (1937).
19. C.J. Gorter, L.J.F. Broer, 'Negative Result of an Attempt to Observe Nuclear Magnetic Resonance in Solids', *Physica (Utrecht)*, **9**, 591–596 (1942).
20. I.I. Rabi, J.R. Zacharias, S. Millman, P. Kusch, 'A New Method of Measuring Nuclear Magnetic Moment', *Phys. Rev.*, **53**, 318 (1938).
21. N. Bloembergen, 'On the Interaction of Nuclear Spins in a Crystalline Lattice', *Physica*, **15**, 386–426 (1949).
22. F. Bruin, F.M. Schimmel, 'Théorie du Détecteur Autodyne pour la Résonance Paramagnétique', *Physica*, **21**, 867–876 (1955).
23. C.J. Gorter, 'Paramagnetic Relaxation', *Physica (Utrecht)*, **3**, 503–514 (1936).
24. C.J. Gorter, F. Brons, 'Magnetic Inhibition of Susceptibilities at Radio Frequencies', *Physica (Utrecht)*, **4**(7), 579–584 (1937).
25. C. Gorter, 'Bad Luck in Attempts To Make Scientific Discoveries', *Physics Today*, **20**(1), 76–81 (1967).
26. E.M. Purcell, H.C. Torrey, R.V. Pound, 'Resonance Absorption by Nuclear Magnetic Moments in a Solid', *Phys. Rev.*, **69**, 37–38 (1946).
27. F. Bloch, W.W. Hansen, M. Packard, 'Nuclear Induction', *Phys. Rev.*, **69**, 127 (1946).
28. R.V. Pound, 'From Radar to Nuclear Magnetic Resonance', *Rev. Mod. Phys.*, **71**, S54–S58 (1999).
29. R.W.P. King, H.R. Mimno, A.H. Wing, *Transmission Lines Antennas and Wave Guides*, Dover Publications, New York, 1965.
30. K. Woelk, J.W. Rathke, R.J. Klingler, 'The Toroid Cavity NMR Detector', *J. Magn. Reson., Ser. A*, **109**, 137–146 (1994).
31. R.E. Gerald, II, R.J. Klingler, J.W. Rathke, G. Sandí, K. Woelk, '*In Situ* Imaging of Charge Carriers in an Electrochemical Cell', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, New York, 111–119, 1998.
32. K. Woelk, R.E. Gerald, P. Trautner, B.L.J. Zwank, J. Bargon, R.J. Klingler, J.W. Rathke, 'Flow and Rheological Studies Using Toroid-Cavity NMR', *Program and Abstracts*, 40th Rocky Mountain Conference on Analytical Chemistry, Denver, CO, July 25–August 1, 1998, 157–158, 1998.
33. T.C. Farrar, E.D. Becker, *Pulse Fourier Transform NMR. Introduction to Theory and Methods*, Academic Press, New York, 1971.
34. D.D. Traficante, 'NMR Concepts', in *NMR Spectroscopy Techniques*, ed. M.D. Bruch, Marcel Dekker, New York, 1–60, 1996.
35. E.R. Andrew, 'Nuclear Magnetic Resonance', Cambridge University Press, Glasgow, 1955.
36. E.O. Stejskal, J.D. Memory, *High Resolution NMR in the Solid State. Fundamentals of CP/MAS*, Oxford University Press, New York, 1994.
37. D.R. Lide (ed.), 'CRC Handbook of Chemistry and Physics', 80th edition, CRC, Boca Raton, 1999.
38. J. Emsley, *The Elements*, 2nd edition, Oxford University Press, Oxford, UK, 1991.
39. P. Raghavan, 'Table of Nuclear Moments', *Atomic Data Nucl. Data Tables*, **42**(2), 189–291 (1989).
40. A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon, Oxford, 1961.
41. R.E. Gerald, II, T. Bernhard, U. Haeberlen, J. Rendell, S. Opella, 'Chemical Shift and Electric Field Gradient Tensors for the Amide and Carbonyl Hydrogens in the Model Peptide *N*-Acetyl-D, L-Valine. Single-Crystal Deuterium NMR Study', *J. Am. Chem. Soc.*, **115**, 777–782 (1993).
42. F.A.L. Anet, D.J. O'Leary, C.G. Wade, R.D. Johnson, 'NMR Relaxation by the Antisymmetric Component of the Shielding Tensor: A Longer Transverse than Longitudinal Relaxation Time', *J. Chem. Phys.*, **171**(5,6), 401–405 (1990).
43. A.E. Hansen, T.D. Bouman, 'Calculation, Display, and Analysis of the Nature of Nonsymmetric Nuclear Magnetic Resonance Shielding Tensors: Applications to Three-membered Rings', *J. Chem. Phys.*, **91**, 3552–3560 (1989).
44. N.F. Ramsey, 'Magnetic Shielding of Nuclei in Molecules', *Phys. Rev.*, **78**, 699–703 (1950).
45. A.E. Hansen, K.V. Mikkelsen, K.L. Bak, 'Pictorial Representations of Molecular Response Tensors; Nuclear Shielding, Magnetizability, and Polarizability', *Magn. Reson. Rev.*, **17**(2), 133–162 (1997).

46. M. Mehring, *Principles of High Resolution NMR in Solids*, Springer-Verlag, Würzburg, 1983.
47. E.W. Wooten, K.T. Mueller, A. Pines, 'New Angles in Nuclear Magnetic Resonance Sample Spinning', *Acc. Chem. Res.*, **25**, 209–215 (1992).
48. J.F. Nye, *Physical Properties of Crystals*, Clarendon, Oxford, 1957.
49. J.F. Nye, *Physical Properties of Crystals, their Representation by Tensors and Matrices*, Oxford University Press, Oxford, 1985.
50. F.A.L. Anet, D.J. O'Leary, 'The Shielding Tensor, Part I: Understanding its Symmetry Properties', *Concepts Magn. Reson.*, **3**, 193–214 (1991).
51. J.C. Facelli, D.M. Grant, 'Molecular Structure and Carbon-13 Chemical Shielding Tensors Obtained from Nuclear Magnetic Resonance', *Topics Stereochem.*, **19**, 1–61 (1989).
52. J.A. Tossell (ed.), 'Nuclear Magnetic Shieldings and Molecular Structure: Proceedings of the NATO Advanced Research Workshop on the Calculation of NMR Shielding Constants and their use in the Determination of the Geometric and Electronic Structures of Molecules and Solids. College Park, Maryland, USA, July 20–24, 1992', Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.
53. S. Ding, C. Ye, 'Exact Solution (of the) Chemical-shift Interaction', *Chem. Phys. Lett.*, **170**(2,3), 277–282 (1990).
54. R.F. Schneider, 'Asymmetry in Magnetic Second-rank Tensor Quantities', *J. Chem. Phys.*, **48**(11), 4905–4909 (1968).
55. A.D. Buckingham, S.M. Malm, 'Asymmetry in the Nuclear Magnetic Shielding Tensor', *Mol. Phys.*, **22**(6), 1127–1130 (1971).
56. S. Ding, C. Ye, 'Corrections to the Cross-polarization Dynamics in Solids: The Effect of Nonsecular Terms of Heteronuclear Dipolar Interaction on Cross-relaxation', *Solid State Nucl. Magn. Reson.*, **1**, 321–328 (1992).
57. Wolfram Research, Inc., *Mathematica*, Version 2.1, Wolfram Research, Champaign, IL, 1992.
58. R.D. Jenks, R.S. Sutor, *AXIOM: The Scientific Computation System*, Springer-Verlag, New York, 1992.
59. J. Herzfeld, A. Berger, 'Sideband Intensities in NMR Spectra of Sample Spinning at the Magic Angle', *J. Chem. Phys.*, **73**, 6021–6030 (1980).
60. G. Engelhardt, D. Michel, *High-resolution Solid-state NMR of Silicates and Zeolites*, John Wiley & Sons, Chichester, 1987.
61. N. Bloembergen, 'On the Interaction of Nuclear Spins in a Crystalline Lattice', *Physica*, **15**(3–4), 386–426 (1949).
62. N. Bloembergen, *Nuclear Magnetic Relaxation*, W. A. Benjamin, New York, 1961.
63. I. Ando, G.A. Webb, *Theory of NMR Parameters*, Academic Press, London, 1983.
64. R. Freeman, 'Pioneers of High-resolution NMR', *Concepts Magn. Reson.*, **11**(2), 61–70 (1999).
65. S.-I. Cho, N.S. Sullivan, 'Parameterized Description of NMR Signal Reception Systems, Part I: Impedance of Transmission Line and Probe', *Concepts Magn. Reson.*, **4**, 227–243 (1992).
66. S.-I. Cho, N.S. Sullivan, 'Parameterized Description of NMR Signal Reception Systems Part II: S/N of NMR Signal Reception Systems', *Concepts Magn. Reson.*, **4**, 293–306 (1992).
67. D.D. Traficante, 'Impedance: What Is It and Why It Must Be Matched', *Concepts Magn. Reson.*, **1**, 73–92 (1989).
68. I.J. Lowe, C.E. Tarr, 'A Fast Recovery Probe and Receiver for Pulsed Nuclear Magnetic Resonance Spectroscopy', *J. Phys. E*, **1**, 320–323 (1968).
69. O. Mueller, W.A. Edelstein, 'Low-noise Preamplifier Design for NMR', *RF Design*, 39–41 (1987).
70. M.G. Richards, A.R. Andrews, C.P. Lusher, J. Schratter, 'Cryogenic GaAs FET Amplifiers and Their Use in NMR Detection', *Rev. Sci. Instrum.*, **57**, 404–409 (1986).
71. F.D. Doty, 'Doty Scientific Catalog', Doty Scientific, Inc., 700 Clemson Road, Columbia, SC 29229, USA (Phone: (803) 788-6497, Fax: (803) 736-5495).
72. G.C. Chingas, 'Overcoupling NMR Probes to Improve Transient Response', *J. Magn. Reson.*, **54**, 153–157 (1983).
73. D.I. Hoult, 'Fast Recovery, High Sensitivity NMR Probe and Preamplifier for Low Frequencies', *Rev. Sci. Instrum.*, **50**, 193–200 (1979).
74. T. Vosegaard, E. Hald, V. Langer, H.J. Skov, P. Daugaard, H. Bildsøe, H.J. Jakobsen, 'Improved Hardware and Software for Single-Crystal NMR Spectroscopy', *J. Magn. Reson.*, **135**, 126–132 (1998).
75. W.E. Maas, L.H. Merwin, D.G. Cory, 'Materials Imaging with Examples from Solid Rocket Propellants', in *Spatially Resolved Magnetic Resonance*, eds. P. Blümler, B. Blümich, R. Botto, E. Fukushima, Wiley-VCH, New York, 141–161, 1998.
76. D.I. Hoult, R.E. Richards, 'The Signal-to-noise Ratio of the Nuclear Magnetic Resonance Experiment', *J. Magn. Reson.*, **24**, 71–85 (1976).
77. A.K. Jameson, C.J. Jameson, A.C. de Dios, E. Oldfield, R.E. Gerald, II, G.T. Turner, '^{129}Xe Magic Angle Spinning Spectra of Xenon in Zeolite NaA: Direct Observation of Mixed Clusters of Co-adsorbed Species', *Solid State Nucl. Magn. Reson.*, **4**, 1–12 (1995).
78. V.J. McBrierty, K.J. Packer, *Nuclear Magnetic Resonance in Solid Polymers*, Cambridge University Press, Cambridge, 1993.
79. K. Schmidt-Rohr, H.W. Spiess, *Multidimensional Solid-state NMR and Polymers*, Academic Press, San Diego, 1994.
80. A.G. Redfield, 'A Guide to Relaxation Theory', in *NMR, Basic Principles and Progress: Introductory Essays*, eds. P. Diehl, E. Fluck, R. Kosfeld, Springer-Verlag, Heidelberg, Vol. 13, 1976.

81. W.P. Slichter, 'NMR Studies of Solid Polymers', in *NMR, Basic Principles and Progress: Natural and Synthetic High Polymers*, eds. P. Diehl, E. Fluck, R. Kosfeld, Springer-Verlag, Heidelberg, Vol. 4, 1971.
82. T.M. Connor, 'Magnetic Relaxation in Polymers. The Rotating Frame Method', in *NMR, Basic Principles and Progress: Natural and Synthetic High Polymers*, eds. P. Diehl, E. Fluck, R. Kosfeld, Springer-Verlag, Heidelberg, Vol. 4, 1971.
83. A.J. Vega, A.D. English, 'Multiple-pulse Nuclear Magnetic Resonance of Solid Polymers. Polymer Motions in Crystalline and Amorphous Poly(tetrafluoroethylene)', *Macromolecules*, **13**, 1635–1647 (1980).
84. C.J. Jameson, A.K. Jameson, R.E. Gerald, II, A.C. de Dios, 'Nuclear Magnetic Resonance Studies of Xenon Clusters in Zeolite NaA', *J. Chem. Phys.*, **96**(3), 1676–1689 (1992).
85. R.E. Gerald, II, *A Single Crystal Study of Deuterons in a Dipeptide and Intercavity Dynamics of Xenon in Zeolite A*, PhD Thesis, University of Illinois, Chicago, 1994.
86. J. Winter, *Magnetic Resonance in Metals*, Oxford University Press, Oxford, 1971.
87. W.D. Knight, 'Nuclear Magnetic Resonance Shift in Metals', *Phys. Rev.*, **76**, 1259–1260 (1949).
88. G.C. Carter, L.H. Bennett, P.J. Kahan, 'Metallic Shifts in NMR, Part I', Pergamon, Elmsford, NY, 62–100, 1977.
89. J. Korringa, 'Nuclear Magnetic Relaxation and Resonance Line Shift in Metals', *Physica*, **XVI**, 601–609 (1950).
90. R.E. Gerald, II, L. Nuñez, R.J. Klingler, J.W. Rathke, '^{7}Li NMR Imaging at the Surface of an Aerospace Alloy', *The Chemist*, **76**(2), 45–50 (1999).
91. M. Winter, J.O. Besenhard, J.H. Albering, J. Yang, M. Wachtler, 'Lithium Storage Alloys as Anode Materials in Lithium Ion Batteries', *Prog. Batteries Battery Mater.*, **17**, 208–213 (1998).
92. N.J. Poulis, 'Nuclear Magnetic Relaxation in Metallic Lithium and Aluminum', *Physica*, **XVI**, 373–376 (1950).
93. H. Figiel, 'NMR of Magnetic Materials', *Magn. Reson. Rev.*, **16**, 101–134 (1991).
94. G.B. Benedek, J. Armstrong, 'Pressure and Temperature Dependence of the Fe^{57} Nuclear Magnetic Resonance Frequency in Ferromagnetic Iron', *J. Appl. Phys. Suppl.*, **32**, 106–110 (1961).
95. C. Robert, J.-M. Winter, 'Résonance Magnétique. Observation de la Résonance Nucléaire du Fer 57 dans le Fer Métallique Naturel en l'Absence de Champ Extérieur', *Comp. Rend. Acad. Sci.*, **250**, 3831–3833 (1960).
96. A.C. Gossard, A.M. Portis, 'Observation of Nuclear Resonance in a Ferromagnet', *Phys. Rev. Lett.*, **3**, 164–166 (1959).
97. A.C. Gossard, A.M. Portis, W.J. Sandle, 'Nuclear Magnetic Resonance in Ferromagnetic Fe^{57}', *J. Phys. Chem. Solids*, **17**, 341–342 (1961).
98. A.M. Portis, A.C. Gossard, 'Nuclear Resonance in Ferromagnetic Cobalt', *J. Appl. Phys. Suppl.*, **31**, 205S–213S (1960).
99. W. Marshall, 'Orientation of Nuclei in Ferromagnets', *Phys. Rev.*, **110**, 1280–1285 (1958).
100. D.J.E. Ingram, *Spectroscopy at Radio and Microwave Frequencies*, 2nd edition, Butterworths, London, 1967.
101. E. MacNamara, G. Fisher, J. Smith, C.V. Rice, S.-J. Hwang, D. Raftery, 'Cross Polarization and Cross Relaxation from Laser Polarized Xenon to Surface Species', *J. Phys. Chem. B*, **103**, 1158–1160 (1999).
102. E. Brunner, 'Enhancement of Surface and Biological Magnetic Resonance using Laser-polarized Noble Gases', *Concept Magn. Reson.*, **11**(5), 313–335 (1999).
103. E. Brunner, M. Haake, L. Kaiser, A. Pines, J.A. Reimer, 'Gas Flow MRI using Circulating Laser-polarized ^{129}Xe', *J. Magn. Reson.*, **138**, 155–159 (1999).
104. R. Tycko, 'Optical Pumping in Indium Phosphide: ^{31}P NMR Measurements and Potential for Signal Enhancement in Biological Solid State NMR', *Solid State Nucl. Magn. Reson.*, **11**, 1–9 (1998).
105. R. Seydoux, A. Pines, M. Haake, J.A. Reimer, 'NMR with a Continuously Circulating Flow of Laser-polarized ^{129}Xe', *J. Phys. Chem. B.*, **103**, 4629–4637 (1999).
106. M.S. Rosen, T.E. Chupp, K.P. Coulter, R.C. Welsh, S.D. Swanson, 'Polarized ^{129}Xe Optical Pumping/Spin Exchange and Delivery System for Magnetic Resonance Spectroscopy and Imaging Studies', *Rev. Sci. Instrum.*, **70**(2), 1546–1552 (1999).

Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton

Mike McGregor
University of Rhode Island, Kingston, USA

1 Introduction 1
1.1 Chemical Shift 2
1.2 Spin–Spin Coupling 2
1.3 Nuclear Overhauser Effect 2
1.4 Analytical Considerations for Spin-$\frac{1}{2}$ Nuclei 3
2 The More Commonly Observed Nuclei 4
2.1 Nitrogen-15 4
2.2 Fluorine-19 5
2.3 Silicon-29 6
2.4 Phosphorus-31 7
3 The Less Commonly Observed Nuclei 8
3.1 Hydrogen-3 8
3.2 Iron-57 9
3.3 Selenium-77 9
3.4 Yttrium-89 10
3.5 Rhodium-103 10
3.6 Silver-107 and Silver-109 10
3.7 Cadmium-111 and Cadmium-113 11
3.8 Tin-115, Tin-117 and Tin-119 11
3.9 Tellurium-123 and Tellurium-125 12
3.10 Xenon-129 13
3.11 Thulium-169 13
3.12 Ytterbium-171 13
3.13 Tungsten-183 13
3.14 Osmium-187 14
3.15 Platinum-195 14
3.16 Mercury-199 14
3.17 Thallium-203 and Thallium-205 15
3.18 Lead-207 16
Abbreviations and Acronyms 16
Related Articles 16
References 17

There are 22 spin-$\frac{1}{2}$ nuclei other than hydrogen and carbon, some with more than one spin-$\frac{1}{2}$ isotope. Four of these, ^{15}N, ^{19}F, ^{29}Si and ^{31}P, have been of major importance in chemistry and nuclear magnetic resonance (NMR) and are the subject of an extensive literature. Many spin-$\frac{1}{2}$ nuclides have had more specialized uses, such as ^{3}H, ^{77}Se, ^{109}Ag, ^{113}Cd, ^{119}Sn, ^{125}Te, ^{129}Xe, ^{195}Pt, ^{205}Tl, and ^{207}Pb. A few have been of limited use, generally due to poor sensitivity: ^{57}Fe, ^{89}Y, ^{103}Rh, ^{169}Tm, ^{171}Yb, ^{183}W, ^{187}Os and ^{199}Hg.

Spin-$\frac{1}{2}$ nuclei generally produce sharp spectral lines, which may allow a wealth of fine structure from spin–spin coupling to be observed. This coupling, along with chemical shift information and dipolar coupling information (nuclear Overhauser effect (NOE)) are used in a variety of ways to obtain structural information and solve chemical problems.

After a brief discussion of analytical techniques, the history, experimental practices, physical properties, relaxation properties, chemical shift referencing, applications and literature references for each nuclide are presented.

1 INTRODUCTION

There are 22 spin-$\frac{1}{2}$ nuclei other than hydrogen and carbon, some with more than one spin-$\frac{1}{2}$ isotope. The group includes metals, nonmetals, metalloids and the noble gas ^{129}Xe. They have two things in common, a spin quantum number of $\frac{1}{2}$ and a literature that reflects 50 years of intense scrutiny by determined physicists, chemists and spectroscopists.

Four of these nuclei, ^{15}N, ^{19}F, ^{29}Si and ^{31}P, have been of major importance in chemistry and NMR and are the subject of an extensive literature. Many spin-$\frac{1}{2}$ nuclides have had more specialized uses, such as ^{3}H, ^{77}Se, ^{109}Ag, ^{113}Cd, ^{119}Sn, ^{125}Te, ^{129}Xe, ^{195}Pt, ^{205}Tl, and ^{207}Pb. A few have been of limited use, generally due to poor sensitivity, including ^{57}Fe, ^{89}Y, ^{103}Rh, ^{169}Tm, ^{171}Yb, ^{183}W, ^{187}Os and ^{199}Hg.

The spin-$\frac{1}{2}$ nuclei range in NMR sensitivity over seven orders of magnitude. Although several of the most sensitive nuclei could be observed by older instruments using the continuous-wave technique, most of the less sensitive nuclei had to await the development of powerful superconducting magnets and the introduction of the pulsed Fourier transform (FT) technique. The sensitivity improvement meant that NMR could be extended to nearly all elements in the periodic table. The resulting explosion of data and publications, combined with the early work, made possible the groundbreaking book "NMR and the Periodic Table".[1] This 1978 text serves as the model for most subsequent multinuclear NMR literature. Outstanding texts describing this field include "The Multinuclear Approach to NMR Spectroscopy"[2] (1983), "NMR of the Newly Accessible Nuclei"[3] (1983) and "Multinuclear NMR"[4] (1987). More recently, the "Encyclopedia of NMR"[5] (1995) contains several excellent reviews of individual nuclei. Biannual reviews

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

of the NMR literature have appeared in *Analytical Chemistry*, providing a convenient starting point for learning about particular applications.

What distinguishes spin-$\frac{1}{2}$ nuclei spectroscopically is the relatively long lifetime of the excited state. This long lifetime leads to well-defined, sharp spectral lines, which allow a wealth of fine structure to be resolved. The linewidth (measured in hertz, at the half-height of a peak) is inversely related to the lifetime of the excited state. High resolution analysis conditions (described below) are required to properly observe such sharp lines.

The basic tools of the NMR spectroscopist are chemical shift, spin–spin coupling and the NOE.

1.1 Chemical Shift

The nucleus experiences the effects of small magnetic fields in its immediate environment. Circulating electrons in chemical bonds produce a magnetic field at the nucleus that may either oppose or reinforce the much larger external field. When the local magnetic field opposes the external field, the nucleus is shielded. A shielded nucleus experiences a lower effective magnetic field strength and resonates at a lower frequency.

As a result, each type of nucleus in a molecule may have a slightly different resonant frequency. This frequency difference is called the chemical shift, because it arises from the chemical environment of the nucleus. The NMR spectrometer accurately measures the resonant frequency of all nuclei being observed in the sample in units of hertz (Hz), or cycles per second. Chemical shift is defined as the difference between the resonant frequency of a nucleus in one type of chemical environment and that of a reference nucleus, divided by the spectrometer observe frequency. Because the spectrometer observe frequency is usually in the megahertz range, whereas the chemical shift is usually in the hertz or kilohertz range, this ratio is expressed in parts per million (ppm). Note that chemical shift is independent of the external magnetic field strength, although the resonant frequency of any given nucleus depends directly on the external field strength. The chemical shift is a sensitive measure of the electronic environment of the nucleus. Much can be deduced about the chemical structure, nature of the bonding and electronic distribution of a nucleus from its chemical shift. An excellent discussion of chemical shift can be found in Abraham et al.[6]

1.2 Spin–Spin Coupling

The small magnetic field of the nucleus exerts an effect on neighboring nuclei called spin–spin coupling. The effect is transmitted through bonds via bonding electrons. A nucleus that experiences no coupling produces a single resonance line. For spin-$\frac{1}{2}$ nuclei, this line will be split into two lines by coupling with a neighboring nucleus. The stronger the coupling, the farther these two lines will be spread. Such coupling is known as J coupling. The coupling constant J is measured in hertz. Coupling constants can be negative or positive, although the appearance of the coupled lines in the spectra is independent of the sign. In a series of compounds where J-coupling decreases, for example, positive values may decrease to zero and then become negative. The absolute values of such a series will appear to decrease and then increase. Only the absolute value of J is measurable from the spectrum, but relative signs of couplings may be obtained. Because J represents the strength of the interaction between two nuclei, it is independent of the external magnetic field strength. The J-coupling pattern for spin-$\frac{1}{2}$ nuclei contains $(n+1)$ lines, where n equals the number of neighboring nuclei. Smaller couplings may come from more distant nuclei, sometimes several bonds away. Coupling patterns may be considerably more complicated due to an effect known as strong coupling, if the frequency difference between the coupled nuclei is not large compared to J. The magnitude of the coupling constant can give information about the stereochemical relationship between coupled nuclei. Three-bond couplings (denoted ^{3}J) depend on the dihedral angle between the coupled nuclei. This dependence is known as the Karplus relationship and it has been used (and abused) in stereochemical analysis for forty years. Most, but not all, nuclei have been shown to exhibit this relationship. Briefly, coupling is at a minimum when this angle is near 90° and a maximum when the angle is near 180°. Intermediate angles produce intermediate coupling constants. Spin–spin coupling is a powerful tool both for establishing connectivity through bonds between atoms in a molecule and for establishing stereochemical relationships.

1.3 Nuclear Overhauser Effect

The NOE is a complex phenomenon that has been discussed in a recent definitive text.[7] The effect may be used to provide information on the spatial relationships of atoms in a molecule or to enhance the response of insensitive nuclei. The effect occurs only between nuclei that share a dipolar coupling, and dipolar relaxation makes a major contribution to the overall spin–lattice relaxation. The most common applications of the NOE are for the sensitivity enhancement of insensitive nuclei and for determining spatial relationships among the atoms in a molecule.

As a general rule the chemical shift range and the size of coupling constants increase as the atom gets larger. Thus the tritium nucleus, ^{3}H, being small, has only a 20-ppm chemical shift range. Medium size nuclei such

as nitrogen and phosphorus have larger chemical shift ranges of hundreds of ppm. Large ''heavy'' nuclei such as platinum and thallium have chemical shift ranges of thousands of ppm. Coupling constants follow a similar trend.

1.4 Analytical Considerations for Spin-$\frac{1}{2}$ Nuclei[8]

NMR spectroscopy is generally used for the identification and proof of structure of chemical compounds and quantitative determination of sample components. NMR also allows the study of a wide array of molecular details, including stereochemistry, molecular motion, intermolecular interactions, chemical equilibria, and many other applications. The technique is nondestructive, and highly selective, since the frequencies at which different nuclei are observed are extremely well dispersed throughout the radiofrequency spectrum. About 50 μM is required for efficient analysis. Signal averaging can reduce this quantity to 50 nM or less, depending on the complexity of the analysis. The less sensitive nuclei will require considerably higher concentrations. Samples may be gas, liquid or solid. Solid samples are usually dissolved, but both powdered and single-crystal samples may also be analyzed. Analysis time can be as little as five minutes, but dilute samples, insensitive nuclei or complex analyses usually require several hours, up to a realistic maximum of 48 hours. The precision of quantitative analyses is routinely ±5%; under rigorous conditions it can be ±0.1%.

There are a few important limitations of the technique. Low sensitivity is inherent to NMR. Even lower sensitivity will be encountered when there is a limited amount of sample, an insensitive nucleus is being observed, or a complex analysis is being performed. Pure compounds are usually required. Mixtures can often be analyzed, provided that each component of the mixture produces a unique NMR signal. Finally, background interferences become a limitation with very dilute samples.

The intensity of the NMR signal is directly proportional to the number of nuclei in the probe, making quantitative analysis straightforward. Like any analytical technique, quantification has more stringent requirements and demands closer controls than qualitative analysis. Both sample-dependent and instrument-dependent factors must be considered. Due to the low sensitivity of the technique, signal averaging is almost always required. This entails its own requirements.

A robust quantitative method must take into account the longitudinal relaxation time, T_1. The relaxation time is a measure of the time required for nuclei in the sample to return to their equilibrium distribution between ground state and excited state after an excitation pulse. It is important to note that, because the energy difference between ground state and excited state is so small, the normal kinetic energy of a room temperature sample is enough to keep the nuclei divided almost 50/50 between ground state and excited state. This is the main distinction between NMR spectroscopy and other spectroscopic methods such as ultraviolet spectroscopy where all species lie in the ground state until excited. In fact, the NMR signal arises from the small excess of nuclei in the ground state (the Boltzmann excess). Depending on the nuclide, the magnetic field strength and the temperature, the Boltzmann excess is of the order of one part in 10^5. In effect, this means that only 1/10 000 of the sample gives a signal. This is the principal reason for the insensitivity of NMR, and it is why signal averaging is almost always required. After an excitation pulse, the nuclei require time to re-establish their equilibrium distribution by losing the excitation energy to their surroundings. This exponential process is called spin-lattice relaxation. This process is characterized by a relaxation time constant, T_1.

Table 1 shows the percent recovery of magnetization after an excitation pulse as a function of T_1. As seen in Table 1, it requires five T_1 periods for the nuclei in the sample to recover 99.3% of their equilibrium distribution. However, if two 90° excitation pulses are separated by only one T_1 period, the second will generate only 63.2% as much signal as the first. To avoid this clearly nonquantitative response, a standard quantitative technique uses a waiting period or recycle time of five times the longest T_1 value among the nuclei being measured. Clearly, some information about the T_1 values of the nuclei in the sample is required. T_1 information can come from literature values, estimation, or preferably measurement. Depending on sample concentration, the T_1 values are easily measured by a variety of techniques. It is sometimes possible to add relaxation reagents to a sample to decrease T_1 by about an order of magnitude and shorten analysis time.

If decoupling is involved, the NOE may affect the intensity of different signals in a nonuniform way. The NOE can be suppressed by the addition of paramagnetic relaxation reagents or by a simple technique called inverse-gated decoupling. Inverse-gated decoupling uses

Table 1 The percent recovery of equilibrium magnetization after a pulse

Number of T_1 periods	Percent recovery of magnetization (%)
$1T_1$	63.2
$2T_1$	86.5
$3T_1$	95.0
$4T_1$	98.2
$5T_1$	99.3

a waiting time of five T_1 periods with the decoupler off and then acquires signal with the decoupler on.

Standardization of analytical results is accomplished in the same way as other instrumental techniques. In addition to the requirement for high purity, NMR standards must produce a signal that can be distinguished from that of the sample and impurities. Standards may be internal, dissolved in the same solution as the analyte, or external, usually contained in an axially concentric tube. A calibration curve may be prepared with a series of concentrations of an external standard. The concentration of an analyte may be determined by adding a known amount of a pure standard to a measured volume of a solution with unknown analyte concentration. The number of equivalent nuclei represented by each sample peak and the standard peak must be taken into account.

The analytical conditions for high resolution spectra include two considerations: lineshape and acquisition time. Briefly, lineshape must be optimized in order to obtain the best resolution and signal-to-noise ratio. This is accomplished by adjusting the magnetic field homogeneity in a process known as shimming. The resolution of the observation (in hertz) is equal to the inverse of the acquisition time (in seconds). In other words, the longer the acquisition time, the higher the resolution. Long acquisition times are associated with lower signal-to-noise ratios.

2 THE MORE COMMONLY OBSERVED NUCLEI

2.1 Nitrogen-15

Nitrogen-15 has been a very useful nuclide, but its extremely low sensitivity made it difficult to observe. Modern high field instruments and special pulse techniques were required before ^{15}N observation became practical. The low sensitivity arises from a low magnetogyric ratio (1/10 that of ^{1}H) and low natural abundance (0.37%). An additional complication comes from the negative magnetogyric ratio, which leads to negative NOE effects. The negative NOE can lead to the complete loss of some signals when proton decoupling is employed.

The huge volume of literature can be accessed through a book,[9] many periodic reports,[10] and reviews.[11,12]

2.1.1 Experimental Practices

The low sensitivity of ^{15}N is a constant concern for the analyst. The normal techniques of using the largest possible sample and extensive signal averaging do not adequately address this problem. In addition, relaxation times may be rather long, increasing the time required for acquisition. In cases where proton decoupling is used, it is possible to lose the nitrogen signal entirely via negative NOE effects. These effects can be suppressed by addition of paramagnetic compounds (not recommended for basic nitrogens) or, preferably, by inverse gated decoupling. Under favorable conditions, the NOE may lead to a large sensitivity increase. There is a relatively simple technique that can greatly improve the sensitivity, although it requires that a proton be either bound to the nitrogen or at least bound to another atom (such as carbon) that is attached to the nitrogen. The technique is called polarization transfer. It is achieved by the use of a pulse sequence known as insensitive nuclei enhanced by polarization transfer (INEPT). This can provide up to a 10- to 20-fold increase in sensitivity for nitrogens with directly attached protons, and less for nitrogens with long-range coupled protons. In cases where such protons exist this is clearly the method of choice.

2.1.2 Physical Properties

The physical properties of ^{15}N are listed in Table 2. Assuming that the low sensitivity problem can be overcome, ^{15}N can be an excellent nucleus for studying chemical environments. There is a large chemical shift range of 1350 ppm. The currently accepted chemical shift reference is neat liquid nitromethane, set at 0.0 ppm. Ammonia and ammonium salts have also been used, but these suffer from large pH and medium effects.

2.1.3 Chemical Shift

Both nitrogen chemistry and nitrogen chemical shift are strongly influenced by the presence of a lone pair of electrons. This makes ^{15}N useful for studying protonation effects, tautomerism and coordination effects. Protonation effects are particularly strong where a π bonding system is involved. For example, the chemical shift of pyridine changes by 100 ppm upon protonation. Chemical shift ranges for nitrogen in aliphatic compounds are in the -200 to -400 range and are somewhat similar to ^{13}C, although larger. The incorporation of an electronegative element on the nitrogen increases the chemical shift.

2.1.4 Coupling

Couplings to other nuclei are small, typically about half the size of the analogous carbon coupling. In the absence of lone pair contributions, coupling is linearly related to the percentage ''s'' character of the N–H bond. Thus, directly bonded protons show coupling constants of about -75 Hz, -90 Hz and -135 Hz, respectively, for sp^3, sp^2 and sp hybridized nitrogen. Such directly bonded protons may be subject to chemical exchange, so the couplings may be lost. Coupling constants may be used to study lone

Table 2 Nuclear parameters of the spin-$\frac{1}{2}$ nuclei

Nuclide	Larmor frequency (MHz) at 7.05 Tesla	Natural abundance (%)	Relative sensitivity[a] vs. proton	Natural sensitivity[b] vs. proton	Chemical shift range (ppm)	Reference compound
^{3}H	320	$<10^{-16}$	1.21	0	20	TMS[c]
^{15}N	30.4	0.365	1.04×10^{-3}	3.85×10^{-6}	1350	CH_3NO_2
^{19}F	282.2	100	0.83	0.83	550	$CFCl_3$
^{29}Si	59.6	4.7	7.84×10^{-3}	3.69×10^{-4}	2000	$Si(CH_3)_4$
^{31}P	121.4	100	6.63×10^{-2}	6.63×10^{-2}	2000	85% H_3PO_4
^{57}Fe	9.7	2.19	6.67×10^{-5}	1.46×10^{-6}	4000	$Fe(CO)_5$
^{77}Se	57.3	7.6	4.77×10^{-2}	3.6×10^{-3}	3300	$Se(CH_3)_2$
^{89}Y	14.7	100	1.07×10^{-2}	1.07×10^{-2}	500	$Y(ClO_4)_3$
^{103}Rh	9.56	100	2.8×10^{-3}	2.8×10^{-3}	10 000	Rh metal
^{107}Ag	12.14	51.8	3.1×10^{-3}	1.6×10^{-3}	270	$Ag_{(aq)}^{+}$
^{109}Ag	13.96	48.2	4.4×10^{-3}	2.1×10^{-3}	270	$Ag_{(aq)}^{+}$
^{111}Cd	63.6	12.75	1.10×10^{-1}	1.40×10^{-2}	800	$Cd(CH_3)_2$
^{113}Cd	66.6	12.26	1.09×10^{-2}	1.34×10^{-3}	800	$Cd(CH_3)_2$
^{115}Sn	98.2	0.35	3.49×10^{-2}	1.22×10^{-4}	5000	$Sn(CH_3)_4$
^{117}Sn	106.1	7.4	4.65×10^{-2}	3.44×10^{-3}	5000	$Sn(CH_3)_4$
^{119}Sn	111.9	8.58	5.17×10^{-2}	4.44×10^{-3}	5000	$Sn(CH_3)_4$
^{125}Te	94.65	6.99	3.16×10^{-2}	2.21×10^{-3}	4000	$Te(CH_3)_2$
^{129}Xe	82.97	26.4	2.12×10^{-2}	5.60×10^{-3}	7500	$XeOF_4$
^{169}Tm	24.82	100	5.66×10^{-4}	5.66×10^{-4}	–	–
^{171}Yb	52.84	14.27	5.47×10^{-3}	7.81×10^{-4}	–	–
^{183}W	14.0	14.28	7.28×10^{-5}	1.04×10^{-5}	6200	Na_2WO_4(aq)
^{187}Os	6.85	1.64	1.22×10^{-5}	2.00×10^{-7}	–	OsO_4
^{195}Pt	64.2	33.8	9.94×10^{-3}	3.36×10^{-3}	13 000	$Na_2Pt(Cl)_6$
^{199}Hg	53.7	16.84	5.67×10^{-3}	9.54×10^{-4}	3000	$Hg(CH_3)_2$
^{203}Tl	171.5	29.5	1.87×10^{-1}	5.51×10^{-2}	7000	$TlNO_3$
^{205}Tl	173.1	70.5	1.93×10^{-1}	1.36×10^{-1}	7000	$TlNO_3$
^{207}Pb	62.8	22.6	3.30×10^{-3}	2.07×10^{-3}	15 000	$Pb(CH_3)_4$

[a] For an equal number of nuclei.
[b] Product of relative sensitivity and natural abundance.
[c] TMS, tetramethylsilane.

pair effects, protonation, isomerization, stereochemistry and coordination. Couplings to carbon are again small. The range is about 0 Hz to −20 Hz. Long range coupling constants are also small, but three-bond coupling values have been used in Karplus relationships, for example in peptides.

2.1.5 Relaxation

Relaxation of ^{15}N is qualitatively similar to that of ^{13}C, although nitrogen's smaller magnetogyric ratio leads to less efficient dipole–dipole relaxation. In general, this means longer relaxation times. Relaxation times (T_1) range from about 15 s for amide nitrogens to 450 s in nitrobenzene, where the lack of attached protons greatly reduces dipolar relaxation.

2.2 Fluorine-19

Fluorine was one of the first nuclei to be observed by NMR. After a long history of significant applications, it continues to be one of the most useful. Due to its relatively high sensitivity it was possible to make observations on the early continuous-wave instruments. The early work is summarized in Dungan and van Wazer.[13] Chemical shift values in the older literature must be used cautiously as both reference material and sign conventions have changed. A large chemical shift range and large coupling constants help make the fluorine nucleus a sensitive probe of chemical environments.

Chemical shift compilations have been published.[14] Coupling constants have been reviewed.[15] An excellent overview of fluorine NMR has been published.[16]

2.2.1 Experimental Practices

Fluorine is easily observed, preferably with a probe optimized for fluorine detection, but also by simply retuning the proton probe. On some older instruments, the proton transmitter may have to be retuned or modified with an accessory to enable it to produce the required fluorine frequency. The large spectral width of possible

fluorine signals may require large data file sizes for high resolution, high sampling rates for the analog-to-digital converter and a wide frequency range for the excitation amplifier. These considerations become more important at higher field strengths. A broad background signal from fluorine-containing materials in the probe may be encountered. It is centered at about −120 ppm and is not usually seen unless very low concentration samples are observed.

2.2.2 Physical Properties

The physical properties of ^{19}F are listed in Table 2. The large magnetogyric ratio and high natural abundance make ^{19}F one of the most sensitive nuclei: 83% as sensitive as the proton. Fluorine has a wide chemical shift range of about 900 ppm, although most organic fluorides occur in a range of about 300 ppm. In much of the early literature trifluoroacetic acid (−76.6 ppm vs. $CFCl_3$) was used as an external chemical shift reference, but it has been replaced by external $CFCl_3$, which is assigned as 0 ppm. Perfluorobenzene, C_6F_6 (−163 ppm), has also been used as an external reference as well as a convenient, commercially available sensitivity standard.

2.2.3 Chemical Shifts

When fluorine is attached to carbon, the chemical shift will be from near zero to −275 ppm. Aromatic fluorides give chemical shifts between −90 and −150 ppm. Olefinic fluorides have a similar range, between −80 and −200 ppm. Approximate ranges for aliphatic fluorides are −200 to −300 ppm for monofluorides, −75 to −125 ppm for difluorides ($>CF_2$) and −50 to −100 ppm for the $-CF_3$ group. Positive chemical shifts occur when fluorine is attached to more electronegative elements than carbon, such as oxygen or the halogens. The chemical shift of F_2, for example, is 430 ppm.

2.2.4 Coupling

One of the outstanding characteristics of fluorine is the fact that it couples so well to neighboring nuclei, leading to large coupling constants and unusually long-range coupling. This makes fluorine a sensitive probe of the local electronic environment. Coupling may also be transmitted through nonbonding electrons, producing rather unusual, so-called ''through space'' couplings, thus affecting atoms that are some distance away.

Geminal coupling (denoted $^2J_{HF}$, indicating J-coupling through 2 bonds, between hydrogen and fluorine) to protons is large, often 5 to 10 times larger than the corresponding proton–proton couplings. Geminal couplings are in the range of 150 Hz to 300 Hz for aliphatic fluorides, but much smaller for olefinic fluorides, about 10 Hz to 80 Hz. Vicinal couplings ($^3J_{HF}$) are frequently in the 10 Hz to 20 Hz range and may show a strong angular dependence. Karplus relationships are not useful as the normal transmission of the coupling effect, which is through sigma bonds, is complicated by transmission through nonbonding electrons. Long-range couplings may be quite extensive, e.g. 20 Hz for $^5J_{HF}$ in aromatic systems.

Couplings to carbon range from 160–400 Hz for directly bonded fluorines to 20–45 Hz for $^2J_{CF}$. $^3J_{CF}$ and longer couplings are easily seen in a ^1H-decoupled carbon spectrum.

Fluorine couples with all the other commonly occurring nuclides such as ^{31}P and ^{15}N. Couplings are large, as much as 1000 Hz to ^{31}P and about 100–200 Hz to ^{15}N. Longer-range coupling is common. Coupling to metals is also known, and may be quite large, up to thousands of hertz. Coupling through nonbonding electrons may provide some very long-range interactions.

2.2.5 Relaxation

Relaxation of ^{19}F resembles proton relaxation. However, the fluorine nucleus is capable of relaxing by additional mechanisms, and this tends to make fluorine relaxation times somewhat shorter than the analogous proton relaxation times. NOE effects can be observed, making the ^{19}F nucleus a useful probe of interatomic distances.

2.3 Silicon-29

The study of ^{29}Si began before the nucleus could be observed directly. Proton spectra of silicon-containing compounds exhibit ''satellite'' peaks very similar to ^{13}C satellite peaks, so Si–H coupling constants could be measured. Silicon itself is of medium sensitivity, and isotopically dilute at 4.7% natural abundance. As silicon polymers and other organosilicon compounds became more and more useful, an increase in the popularity of ^{29}Si-NMR was natural. There is a large body of literature covering ^{29}Si, including a general review,[17] and a chemical shift compilation.[18]

2.3.1 Experimental Practices

The analyst observing ^{29}Si has several important considerations to take into account: background signals, negative NOE and long relaxation times. Because the NMR probe usually contains glass and other ceramics, there is a large, broad background ^{29}Si peak at about −110 ppm to contend with. In many cases the peaks of interest have a higher chemical shift and the background signal is simply ignored. The judicious choice of observation window can minimize the appearance of the glass peak. It is also possible to record a ''blank'' spectrum (everything but the

sample, tube included) that can be subtracted from the spectrum of the sample. This of course doubles the analysis time. An elegant solution to this problem is to use one of the polarization transfer experiments such as distortionless enhancement by polarization transfer (DEPT). In this case, only silicon atoms that are coupled to protons will be observed. The glass background disappears. This technique is limited to proton-coupled silicons, provided that proton relaxation times are sufficiently long. It is not a good method for quantitative work.

Because ^{29}Si has a negative magnetogyric ratio, the possibility of negative NOE enhancements exists. In a few cases, this can lead to the complete cancellation of the ^{29}Si signal. In any case, this effect, which follows upon proton decoupling, leads to very nonquantitative responses. The negative NOE effects can be suppressed by the addition of paramagnetic materials such as Cr^{3+} ion, which has the additional advantage of decreasing silicon relaxation times. However, the sample will be irretrievably contaminated, line widths will increase and there is the possibility of sample components reacting with the chromium. A cleaner way of suppressing unwanted NOE effects is to use the inverse-gated decoupling technique. The decoupler is applied only during the acquisition time. The decoupler is then turned off for a relatively long period (about 5 times the acquisition time or five times T_1).

A major impediment to ^{29}Si analyses is the long relaxation time of the ^{29}Si nucleus. Since signal averaging is required, the acquisition repetition rate is limited by how fast the silicon nuclei relax. The only way to avoid long acquisitions is to add paramagnetic materials as mentioned above. The addition of about 0.05 M Cr^{3+} will reduce ^{29}Si relaxation times by 50% to 90%, allowing faster recovery and more rapid pulsing. Quantitative work, therefore, usually relies on the combination of inverse-gated decoupling and addition of paramagnetics.

2.3.2 Physical Properties

The physical properties of ^{29}Si are summarized in Table 2. With a natural abundance of 4.7% and a relatively large magnetogyric ratio, ^{29}Si is moderately sensitive, rather like ^{13}C. The chemical shift ranges over about 400 ppm. Chemical shifts are referenced to tetramethylsilane (TMS), which also serves as reference for ^{1}H and ^{13}C.

2.3.3 Chemical Shifts

Although the total range of reported ^{29}Si chemical shifts is about 750 ppm, most organosilicon compounds fall in the range of 40 ppm to −220 ppm. Chemical shift compilations are available.[17] A major application of ^{29}Si is the identification and structural characterization of siloxane polymers, for which extensive chemical shift data are available.[17] Chemical shift data as well as additivity rules are available for polymeric silanes.[18]

2.3.4 Coupling

Due to the negative gyromagnetic ratio, most ^{29}Si couplings have a negative value. Data summaries and theoretical aspects of Si−H coupling have been published.[18] The values for $^{1}J_{SiH}$ are large, from about −100 Hz to −350 Hz. $^{2}J_{SiH}$ couplings are much smaller, in the range of 0 Hz to 10 Hz. Silicon–carbon couplings range from −50 Hz to −100 Hz for $^{1}J_{SiC}$ down to 0 Hz to 20 Hz for $^{2}J_{SiC}$. Silicon–fluorine couplings are positive and in the range of 100 Hz to 400 Hz.[15] Silicon–phosphorus couplings are found to have approximately the range of 20 Hz to 200 Hz.

2.3.5 Relaxation

It is important for the analyst to remember that ^{29}Si relaxation times are long. A summary of relaxation times has been published.[18] Typical values for most silicon atoms are 10–100 s. Unless paramagnetic relaxation reagents are employed, pulse repetition rates must be kept slow and analysis times correspondingly long. In addition, dipolar relaxation produces negative NOE effects as noted previously.

2.4 Phosphorus-31

Since the early days of NMR ^{31}P has been one of the most useful and most-studied nuclei. With 100% natural abundance and a high magnetogyric ratio, ^{31}P has a good sensitivity and is easy to observe. Most often observed in a broadband probe, phosphorus compounds generally produce clean, well-dispersed spectra with an abundance of coupling data. The importance of phosphorus in biological compounds has created intense interest in ^{31}P-NMR spectroscopy, resulting in an enormous literature on the subject. The data are most easily accessed through the *Handbook of ^{31}P NMR Data*,[19] which contains chemical shift and coupling constant data for a large collection of all types of phosphorus compounds. Applications of ^{31}P-NMR to organic chemistry, stereochemistry and structural characterization are presented in two dedicated monographs.[20,21] A recent, brief, overview of the field is available.[22]

2.4.1 Experimental Practices

The observation of ^{31}P is straightforward. It is usually accomplished in a broadband probe, with or without proton decoupling. The wide chemical shift range of more than 1000 ppm is accommodated with large data file

sizes. A typical ^{31}P spectrum is acquired in the same way as a proton spectrum, with a wide spectral width and a tip angle less than 90°, with a repetition time of about one or two seconds.

2.4.2 Physical Properties

The outstanding physical characteristic of ^{31}P is its sensitivity: only one order of magnitude less than the proton. This is a result of 100% natural abundance and a high magnetogyric ratio. ^{31}P chemical shifts have a range of over 2000 ppm. 85% Phosphoric acid is used as an external chemical shift reference, at 0.0 ppm. Alternative chemical shift references are in use that avoid the problem of the broadness of the phosphoric acid line.[22] Small differences between reported chemical shifts in the literature are common, due to matrix effects.

2.4.3 Chemical Shifts

Phosphorus chemistry is complex, with an enormous number of compounds having a wide range of coordination numbers and oxidation states. Most of the common organic phosphorus compounds give signals between 150 and −100 ppm.

2.4.4 Coupling

The phosphorus atom tends to have large couplings, easily visible in the proton or carbon spectra of organophosphate compounds. Extensive compilations of coupling constants are available.[20–22] Directly bonded ($^1J_{PX}$) couplings are exceptionally large, from 100 Hz to 1000 Hz for P−H coupling, −50 Hz to 500 Hz for P−C coupling and 500 Hz to 1500 Hz for P−F coupling. Geminal couplings ($^2J_{PX}$) are smaller: 0 Hz to 200 Hz for P−H coupling, 0 Hz to 100 Hz for P−C coupling and 0 Hz to 500 Hz for P−F coupling. Vicinal couplings ($^3J_{PH}$) are sometimes larger than the $^2J_{PH}$ values. The $^3J_{PH}$ coupling constants depend on the dihedral angle between the coupling atoms. These couplings conform to the normal Karplus relationship trends. Minimum values are found at angles near 90° and maximum values at angles near 180°. $^4J_{PH}$ values range from 0 Hz to 20 Hz. Very long-range couplings covering up to eight bonds are known. The values for such couplings are small.

2.4.5 Relaxation

Relaxation times for phosphorus compounds resemble those of protons, generally being in the 1 s to 20 s range. Organophosphate compounds usually have relaxation times in the 1 s to 10 s range. Matrix effects strongly affect these values. Compilations of relaxation times for phosphorus compounds are available.[19,21]

3 THE LESS COMMONLY OBSERVED NUCLEI

3.1 Hydrogen-3

Tritium is a rare isotope of hydrogen that has excellent NMR properties. It is the most sensitive nucleus: 21% more sensitive than ^1H. The principal obstacle to its use is that it is radioactive, a β-emitter with a half-life of 12.5 years. Early work was hampered by low sensitivity instruments, which required large sample sizes. Modern instrumentation allows the analyst to observe extremely small samples of tritium, minimizing the hazards of the radioactivity. A definitive compilation of tritium NMR methods, data and applications was published in 1985.[23] A 1995 review of more recent work is also available.[24]

3.1.1 Experimental Practices

The best source for experimental techniques in tritium NMR is '*Handbook of Tritium NMR Spectroscopy and Applications*'.[23] A great deal of work has been directed to developing safe techniques for handling tritium-containing samples. Tritium is a "soft" β-emitter, meaning that the particle emitted has very little energy. Such particles cannot penetrate even a thin surface of plastic, so effective shielding is easily accomplished. Samples are usually contained in commercially available doubly enclosed tubes, which can safely contain the sample even if a tube breaks. Modern instrumentation and techniques allow analysis of low concentration samples, minimizing the hazards. However, the amount of tritium necessary for analysis is 10^7 times that required for liquid scintillation analysis, a difference that illustrates both the sensitivity of scintillation counting and the general insensitivity of the NMR technique. Instrumentally, tritium observation closely resembles proton observation. The resonant frequency is 6.66% higher than proton, within the range of most ordinary proton observe channels. Conditions that are optimized for proton spectroscopy are also optimum for tritium observation.

3.1.2 Physical Properties

With the highest magnetogyric ratio of any nucleus, tritium is the most sensitive: 21% better than the proton. Having almost no natural abundance, the tritium must be synthetically introduced into the molecule. Its chemical shift range is very similar to the proton, about 20 ppm. Chemical shift referencing is accomplished with tritiated TMS. In practice, normal TMS may be used, because the tritium/proton Larmor frequency ratio is known very accurately. Measuring the proton frequency of the reference material and multiplying by the Larmor frequency ratio gives the frequency of the tritium reference. This is referred to as "ghost" referencing.[24]

3.1.3 Chemical Shift

All of the hydrogen isotopes have nearly identical nuclear shielding. This means that all three isotopes have nearly identical chemical shift scales and that the vast literature on proton chemical shifts (see **Chemical Shifts in Nuclear Magnetic Resonance**) can be applied to deuterium and tritium as well. Isotope effects are usually less than a few hundredths of a ppm. Replacement of a proton with a triton gives very clean spectra, with all coupling information preserved. Tritium may be introduced into a molecule in several ways, including catalytic tritiation. The products produced may differ from those expected, as the mechanism of the reaction is complex. This provides a clever method for mechanistic studies involving hydrogenation.

3.1.4 Coupling

Tritium coupling constants are nearly identical to proton coupling constants, being exactly 1.067 times larger. This similarity allows the use of existing literature on proton coupling constants in tritium NMR. A particularly interesting application allows the observation of couplings between protons with degenerate chemical shifts. For example, a CH_3 singlet will become a doublet if one proton is replaced with a triton, revealing the T–H coupling. By calculation, the previously hidden H–H coupling constant can be determined.

3.1.5 Relaxation

Relaxation processes for ^3H are the same as for ^1H, in contrast to ^2H, which relaxes by the quadrupolar mechanism. The extensive literature data on relaxation times for proton may be extended directly to tritium. In proton-decoupled spectra the prospect of an NOE must be taken into account for quantitative work. Qualitatively, the NOE can be used to study the proximity of a tritium atom to other protons in the molecule.

3.2 Iron-57

Iron has a spin-$\frac{1}{2}$ isotope, ^{57}Fe. Observation of this nuclide is extremely difficult because of its low natural abundance and very low magnetogyric ratio. There is very little literature available.[25] Only the Fe^0 and Fe^{2+} oxidation states are diamagnetic.

3.2.1 Experimental Practices

The low magnetogyric ratio produces a Larmor frequency that is below the design limit of a typical broadband probe. Extra effort is required to observe this nuclide.

3.2.2 Chemical Shift, Coupling and Relaxation

Literature reports are very sparse.[25] Most compounds are paramagnetic. Coupling has been observed between iron and carbon at about 5 Hz. Coupling to phosphorus is about 25 Hz.

3.3 Selenium-77

Selenium has been a useful NMR nucleus, with a relatively large body of data accumulated. A useful overall review[26] provides a good starting point for the literature and there are several data compilations.[27,28] Organoselenium chemistry is a large field, with many stereochemical applications including stereoselective synthesis and analysis of chiral compounds. Chiral derivatization reagents containing selenium have been used to detect chemical shift differences in diastereomeric adducts up to seven bonds away from the chiral site.

3.3.1 Experimental Practices

The observation of ^{77}Se is straightforward. The sensitivity is slightly better than ^{13}C, so acquisitions can be fairly rapid. Because the NOE is not usually an important factor, decoupling is simply done without any special precautions. ^{77}Se chemical shifts have a strong temperature dependence. Close temperature control is required. This temperature sensitivity has been exploited in dynamic NMR experiments, where organoselenium derivatives show coalescence temperatures about 50 °C higher than in proton or carbon spectra. This brings the study of lower energy processes into the range of variable temperature probes.

3.3.2 Physical Properties

The physical properties of ^{77}Se are listed in Table 2. The relatively high observe frequency of 57.3 MHz puts ^{77}Se in the middle of the range of most multinuclear probes and the sensitivity is relatively good. The chemical shift spans a range of 3300 ppm. Chemical shifts are referenced to $Se(CH_3)_2$ in $CDCl_3$ at 0.0 ppm. An alternative reference compound is $(C_6H_5)_2Se_2$ in $CDCl_3$ with a chemical shift of 463 ppm. This compound has the advantage of being an easily handled solid with no objectionable odor.

3.3.3 Chemical Shift

Chemical shifts reflect both oxidation state and the electronegativity of substituents. Se^{4+} compounds are deshielded versus Se^{2+} compounds. For example, decreasing electronegativity of the substituents explains the chemical shift trend in the series of $SeCl_2$ (1758 ppm),

$SeBr_2$ (1474 ppm) and SeI_2 (814 ppm).[28] Oxidation state is reflected by the chemical shifts of $Se(CH_3)_2$ (0 ppm) and $SeO(CH_3)_2$ (812 ppm). Electropositive substituents produce shielded chemical shift values, as in H_2Se (−226 ppm).[28]

3.3.4 Coupling

Coupling in selenium compounds has been reviewed.[27,28] Couplings are rather small. Directly bonded protons couple with selenium in the range of 40 Hz to 60 Hz. Coupling with directly bonded carbon gives values in the −40 Hz to −100 Hz range. Fluorine has one-bond coupling constants in the range of a few hundred hertz up to 1500 Hz. Longer range couplings in each case are about an order of magnitude less.

3.3.5 Relaxation

Relaxation time constants are in the range of 1 s to 30 s. There is not too much contribution from dipolar relaxation, so NOE effects are not too important. Pulse repetition rates can reflect the rather short T_1 values.

3.4 Yttrium-89

^{89}Y is 100% naturally abundant but suffers from a low gyromagnetic ratio. Long relaxation times make the nuclide difficult to observe. A review of available data has been published.[25]

3.4.1 Experimental Practices

Lack of sensitivity is the main concern of the analyst. The long relaxation times are particularly difficult to overcome. The NOE may be used to enhance sensitivity if coupled protons are available.

3.4.2 Physical Properties

As mentioned, even though 100% abundant, ^{89}Y has a low magnetogyric ratio.

3.4.3 Chemical Shift

Very few ^{89}Y chemical shifts have been reported.

3.4.4 Coupling

Some coupling data have been obtained from satellite spectra. Coupling to carbon was measured as 12 Hz.[4]

3.4.5 Relaxation

Relaxation times for ^{89}Y are very long, ranging from 1 min to 4 min.[4] The nucleus relaxes by a dipolar mechanism, leading to a potentially very large NOE. The maximum NOE is a factor of −10.

3.5 Rhodium-103

Although ^{103}Rh is 100% naturally abundant, it has one of the lowest magnetogyric ratios of all the spin-$\frac{1}{2}$ nuclei. The sparse literature on rhodium chemical shifts and coupling constants has been reviewed.[25] Rhodium NMR is not expected to make a significant contribution to rhodium chemistry.

3.5.1 Experimental Practices

The direct observation of rhodium presents a serious challenge to the analyst. Due to the low magnetogyric ratio, the observe frequency is often below the lower limit of tuning for a broadband probe. Most of the early work has been accomplished using double resonance techniques.

3.5.2 Physical Properties

^{103}Rh is 100% abundant, but extremely insensitive due to a low magnetogyric ratio. The chemical shift range of rhodium compounds is about 8000 ppm, referenced to rhodium metal.

3.5.3 Chemical Shift

Rhodium chemical shifts are temperature dependent. The chemical shift literature has been reviewed.[25]

3.5.4 Coupling

Rhodium coupling is similar to the other transition metals. Coupling constants are somewhat smaller. Directly bonded coupling constants to proton are a few tens of hertz, to carbon about 30 Hz to 80 Hz, to fluorine about 60 Hz and to phosphorus about 80 Hz to 150 Hz.

3.5.5 Relaxation

Relaxation times for rhodium are similar to the other transition metals.

3.6 Silver-107 and Silver-109

Silver has two spin-$\frac{1}{2}$ isotopes, ^{107}Ag and ^{109}Ag of nearly 50% natural abundance each. Most literature studies employ ^{109}Ag because it is about 40% more sensitive. Silver ranks near the bottom of the group in sensitivity and does not have an extensive literature. Chemical shift scales for the two nuclei are nearly identical, while coupling constants of ^{109}Ag

are expected to be about 15% larger than those for ^{107}Ag. Chemical shifts and coupling constants have been reviewed.[25]

3.6.1 Experimental Practices

The low frequency of the silver resonance is at the extreme lower range of most broadband probes. The principal difficulty in silver observation is low sensitivity. Relaxation times tend to be rather long, limiting the pulse repetition rate. These two impediments can be overcome only by long analysis times.

3.6.2 Physical Properties

The outstanding physical property of these two nuclides is their low magnetogyric ratio, leading to very low sensitivity. The chemical shift range is about 270 ppm, referenced to the aqueous silver ion, $Ag_{(aq)}{}^{+}$.

3.6.3 Chemical Shift

There is no extensive literature on chemical shifts for silver. Studies are often complicated by the fact that several species with complex equilibria may exist, making the technique even more difficult. Many of the chemical shifts are also highly concentration dependent.

3.6.4 Coupling

Few examples of silver coupling are known. Two-bond coupling to proton is about 10 Hz. Coupling to fluorine and phosphorus has been reported. The phosphorus–silver coupling constants were in the range of 500 Hz.

3.6.5 Relaxation

Relaxation times for the silver nucleus are long: about 50 s for aqueous silver ion and up to several minutes for silver salts in acetonitrile.

3.7 Cadmium-111 and Cadmium-113

Cadmium has been an interesting, if complicated, nucleus for NMR studies. There are two spin-$\frac{1}{2}$ isotopes, ^{111}Cd and ^{113}Cd. Both have about the same natural abundance and adequate sensitivity. Chemical shifts and coupling constants are essentially the same for both and have been reviewed.[25] Cadmium complexes are rather labile and the situation may be complicated in aqueous solution. Only one resonance may appear even though several species such as dimers, trimers, etc. may be present. Cadmium has a useful application as a surrogate for zinc and calcium anions in biomolecular NMR studies.

3.7.1 Experimental Practices

There are no special observational difficulties associated with cadmium. The negative magnetogyric ratio can lead to the expected negative NOE if proton decoupling is required. The possible loss of signal can be avoided using the techniques outlined in the ^{15}N and ^{29}Si sections. The sometimes long relaxation times of some cadmium compounds limit pulse repetition rates.

3.7.2 Physical Properties

With a mid-range Larmor frequency and about 12% natural abundance, either isotope is relatively sensitive and easy to observe. The chemical shift range is nearly 1200 ppm. Chemical shifts are referenced to $Cd(CH_3)_2$ at 0 ppm.

3.7.3 Chemical Shift

Most alkyl-substituted cadmium compounds have chemical shifts in the range of 0 ppm to −400 ppm. Inorganic cadmium salts are more likely to have chemical shifts in the range of 0 ppm to 600 ppm. Strong solvent dependence and complex equilibria are always possible.

3.7.4 Coupling

Directly bonded cadmium–carbon coupling constants are in the 500-Hz range. Protons in alkyl cadmium compounds are long-range coupled at about 50 Hz. Directly bonded phosphorus couplings to cadmium are about 1300 Hz.

3.7.5 Relaxation

Relaxation times for cadmium are long, often from 30 s to 60 s. Symmetrical compounds such as $Cd(CH_3)_2$ have the longest relaxation times, in this case 60 s. Alkyl cadmium compounds with longer chains have relaxation times between 1 s and 10 s. In the case of proton decoupling, the potential for negative NOE is present.

3.8 Tin-115, Tin-117 and Tin-119

There are three spin-$\frac{1}{2}$ isotopes of tin: ^{115}Sn, ^{117}Sn and ^{119}Sn. The most sensitive is ^{119}Sn, closely followed by ^{117}Sn. ^{115}Sn is about twenty times less sensitive. Most of the literature studies concern ^{119}Sn. Chemical shift scales are substantially identical for all three nuclides. Coupling constants show small differences, nearly negligible in the case of the two most sensitive nuclides. ^{119}Sn is easily observed, and makes an excellent tool for the study of tin chemistry. Literature reviews and chemical shift compilations are available.[26,27]

3.8.1 Experimental Practices

Observation of ^{119}Sn is straightforward, except for the negative magnetogyric ratio. The nuclide is sensitive, and relaxation times are not too long. Frequently, complex reactions with many intermediates and various products can be observed and characterized directly in the solution reaction medium. Close temperature control is required. Because ^{119}Sn has a negative magnetogyric ratio, the possibility of negative NOE enhancements exists. In a few cases, this can lead to the complete cancellation of the ^{119}Sn signal. In any case, this effect, which follows upon proton decoupling, leads to very nonquantitative responses. The negative NOE effects can be suppressed by the addition of paramagnetic materials such as Cr^{3+} ion, which has the additional advantage of decreasing tin relaxation times. However, the sample will be irretrievably contaminated, line widths will increase and there is the possibility of sample components reacting with the chromium. A cleaner way of suppressing unwanted NOE effects is to use the inverse-gated decoupling technique. The decoupler is applied only during the acquisition time. The decoupler is then turned off for a relatively long period (about 5 times the acquisition time). Finally, the technique of polarization transfer may be employed in the pulse sequences DEPT or INEPT, provided that there are some protons coupled to the tin.

3.8.2 Physical Properties

The moderately high magnetogyric ratio and natural abundance of ^{119}Sn give it moderate sensitivity. The Larmor frequency of 111.9 MHz is in the middle range of a typical broadband probe. ^{119}Sn features a large chemical shift range of 5500 ppm. Chemical shifts are referenced to $Sn(CH_3)_4$.

3.8.3 Chemical Shift

The chemical shift ranges from 3000 ppm to −2500 ppm. Chemical shifts are very temperature dependent and may be highly solvent dependent. In general, an increase in coordination number means an increase in shielding which results in lower chemical shifts.

3.8.4 Coupling

Most ^{119}Sn coupling constants are large and negative. One-bond couplings to proton are in the range of −1500 Hz to −2000 Hz. Coupling to carbon is in the range of −300 Hz to −500 Hz. Coupling to fluorine is about +1500 Hz. Longer range couplings are known. Many tin couplings have been obtained from satellite peaks in proton and carbon spectra.

3.8.5 Relaxation

Tin relaxes in a manner similar to the other spin-$\frac{1}{2}$ nuclei. Relaxation times are usually a few seconds. If dipolar relaxation with nearby protons is possible, the analyst must take into account the possibility of negative NOE effects.

3.9 Tellurium-123 and Tellurium-125

There are two spin-$\frac{1}{2}$ isotopes of tellurium. ^{125}Te is so clearly the preferable of the two that nearly all of the literature reports are studies of this isotope. The chemical shifts and coupling behavior of the two isotopes are not significantly different. Chemical shift and coupling data have been reviewed.[28–30]

3.9.1 Experimental Practices

The observation of ^{125}Te is straightforward. The sensitivity is more than twelve times better than ^{13}C, so acquisitions can be fairly rapid. Because the NOE is not usually an important factor, decoupling is simply done without any special precautions. ^{125}Te chemical shifts have a strong temperature dependence. Close temperature control is required. As with ^{77}Se, this temperature dependence has been exploited in dynamic NMR studies, where the coalescence temperatures of organotellurium compounds are even higher than the corresponding organoselenium compounds. This extends the range of dynamic processes that can be studied to lower energies. The higher coalescence temperatures for organotellurium compounds are more likely to be in the temperature range of variable temperature broadband probes.

3.9.2 Physical Properties

Tellurium-125 physical properties are listed in Table 2. With a relatively high magnetogyric ratio and 6.99% natural abundance, ^{125}Te is fairly sensitive. The chemical shift range of tellurium compounds is about 5000 ppm. Chemical shift compilations have been presented.[28–31] Chemical shift referencing is to $Te(CH_3)_2$ at 0 ppm or Ph_2Te_2 at 422 ppm.

3.9.3 Chemical Shift

Chemical shift trends for ^{125}Te are similar to those for ^{77}Se. Electronegative substituents produce high chemical shifts as in TeF_6 at 545 ppm and electropositive substituents produce low chemical shifts such as $Te(CH_3)_4$ at −67 ppm.

3.9.4 Coupling

Data on ^{125}Te coupling constants are not too abundant in the literature. Several reviews and compilations are

available.[29,30] A convenient rule of thumb is that ^{125}Te coupling constants will be about two to three times larger than the corresponding ^{77}Se coupling constants. Some representative coupling constants are: directly bonded proton Te–H, about −50 Hz; directly bonded carbon Te–C, −50 Hz to −100 Hz; and directly bonded fluorine, Te–F, about 2500 Hz to 3800 Hz.

3.9.5 Relaxation

Relaxation time constants are in the range of 1 s to 30 s. There is not too much contribution from dipolar relaxation, so NOE contributions are not too important. Pulse repetition rates can reflect the rather short T_1 values.

3.10 Xenon-129

Xenon is unique among the noble gases for its NMR properties as well as its chemistry. As the only spin-$\frac{1}{2}$ nucleus among the noble gases, ^{129}Xe is abundant, sensitive, and easy to observe. Much of what is known about xenon chemistry comes from ^{129}Xe-NMR studies. Xenon is a heavy nuclide with a large chemical shift range and large coupling constants. ^{129}Xe-NMR can be used as a sensitive probe of bonding and geometry in a wide variety of xenon compounds and complexes. The chemical shift and coupling constant literature has been reviewed.[32–34]

3.10.1 Experimental Practices

The Larmor frequency of ^{129}Xe lies in the middle range of a broadband probe. Sensitivity is good. Relaxation times are short, so pulse repetition rates can be rapid and analysis time short. The large chemical shift range requires large data sets for high resolution. Many less-stable xenon complexes require low temperature for stabilization. Special solvent conditions may also be required.

3.10.2 Physical Properties

The physical properties of ^{129}Xe are listed in Table 2. Features are a large magnetogyric ratio and medium natural abundance. The chemical shift range spans over 7000 ppm. Chemical shifts are referenced to $XeOF_4$ at 0 ppm.

3.10.3 Chemical Shift

Xenon chemical shifts depend on the oxidation state of Xe: Xe^0 compounds give shifts between −5460 ppm and −5331 ppm, Xe^{2+} compounds between −3769 ppm and −574 ppm, Xe^{4+} compounds between −662 ppm and 595 ppm, Xe^{6+} compounds between −211 ppm and 700 ppm, and Xe^{7+} compounds 2077 ppm and higher.[31] Temperature and other matrix effects are large.

3.10.4 Coupling

Many coupling constants between xenon and a wide variety of other nuclei have been reported. Xenon–fluorine couplings are particularly well studied. These coupling constants can be correlated with oxidation number: Xe^{2+} has Xe–F couplings from 4800 Hz to 7600 Hz, Xe^{4+} from 2400 Hz to 3800 Hz, and Xe^{6+} from 95 Hz to 1400 Hz.[31] Large temperature and solvent effects upon coupling constants have been observed.

3.10.5 Relaxation

Relaxation times for xenon compounds are short, usually less than one second.

3.11 Thulium-169

This nuclide is almost paramagnetic and there is little useful literature. There have been several solid-state studies published.[4]

3.12 Ytterbium-171

Very little literature exists for this nuclide. Only the Yb^{2+} oxidation state is diamagnetic, and compounds are not too stable. Coupling to fluorine has been reported as 120 Hz from satellite spectra.[4]

3.13 Tungsten-183

Tungsten has a spin-$\frac{1}{2}$ isotope, ^{183}W, but it has one of the lowest magnetogyric ratios of the group. This is somewhat compensated for by a moderate natural abundance. The lack of sensitivity has made this nuclide less interesting, as evidenced by the lack of literature. The available literature has been reviewed.[25]

3.13.1 Experimental Practices

Very poor sensitivity is the main concern. The Larmor frequency is at the extreme low end of the typical broadband probe.

3.13.2 Physical Properties

In spite of the low magnetogyric ratio, a chemical shift range of 6000 ppm has been observed for ^{183}W. Chemical shift referencing is to WF_6 at 0 ppm.

3.13.3 Chemical Shift

Chemical shifts depend on the oxidation state of tungsten. W^{6+} compounds have chemical shifts from 0 ppm

to 3000 ppm. Lower valent tungsten compounds have chemical shifts from −1500 ppm to −3500 ppm.

3.13.4 Coupling

Coupling constants for directly bonded protons are about 30 Hz to 80 Hz; for directly bonded carbon, 120 Hz to 190 Hz; for directly bonded fluorine, 10 Hz to 80 Hz; for directly bonded phosphorus, 200 Hz to 500 Hz.

3.13.5 Relaxation

Relaxation times for typical tungsten compounds are in the 1 s to 10 s range.

3.14 Osmium-187

Osmium has received very little attention due to low natural abundance and a very low magnetogyric ratio, the lowest of the spin-$\frac{1}{2}$ nuclei. The observe frequency is below the limit of a typical broadband probe.

3.14.1 Experimental Practices

Observation is quite difficult due to the very low Larmor frequency.

3.14.2 Physical Properties

Osmium is characterized by very low natural abundance and very low magnetogyric ratio.

3.14.3 Chemical Shift, Coupling and Relaxation Times

Coupling constants have been reported from satellite spectra. Coupling to protons is in the range of 16 Hz to 38 Hz. Coupling to carbon is from 80 Hz to 121 Hz.[4] Relaxation times for ^{187}Os are in the 1–25 s range.[25]

3.15 Platinum-195

Platinum is an easily observed nucleus of moderate sensitivity. Much early work was done by observation of the ^{195}Pt satellites in proton spectra. Along with the modern work, this has revealed a nucleus with an extremely wide chemical shift range and very large coupling constants. As expected, such a nucleus is quite sensitive to electronic effects as well as temperature and other matrix effects. Chemical shift and coupling constant data have been reviewed.[25]

3.15.1 Experimental Practices

Observation of ^{195}Pt is straightforward. Careful attention must be paid to matrix effects such as solvent, concentration, temperature and pH. The short relaxation times of platinum compounds usually allow rapid pulsing.

3.15.2 Physical Properties

The physical properties of ^{195}Pt are listed in Table 2. The Larmor frequency is in the middle range, and the natural abundance is high enough for moderately high sensitivity. The chemical shift range of 13 000 ppm is one of the largest of the spin-$\frac{1}{2}$ nuclei. $Pt(Cl)_6^{2-}$ is the reference compound, even though temperature and concentration effects are large. Coupling constants are large, often thousands of hertz. Relaxation times are rather rapid.

3.15.3 Chemical Shift

The large chemical shift range of ^{195}Pt is spanned by a series of halogen compounds $Pt(X)_6$: for X = F, 11 800 ppm, for X = Cl, 4521 ppm, for X = Br, 2651 ppm, and for X = I, −1528 ppm.[25] Chemical shifts are highly matrix dependent as previously mentioned. Generally, large chemical shift changes occur as ligands are changed, but changes in the ligands themselves have little effect on ^{195}Pt chemical shift. For example, platinum alkyl complexes have similar chemical shifts, even with different alkyl groups.

3.15.4 Coupling

Platinum coupling constants are large. One-bond coupling to proton is from 100 Hz to 1000 Hz. Two-bond coupling is from 40 Hz to 90 Hz. Three-bond proton coupling is from 15 Hz to 50 Hz. Platinum–carbon one-bond coupling ranges from 600 Hz to 2000 Hz. Couplings to most of the other spin-$\frac{1}{2}$ nuclei have been reviewed.[25]

3.15.5 Relaxation

Relaxation times of ^{195}Pt are short. Typical T_1 values are from 0.3 s to 1.3 s. Linewidths up to 20–50 Hz are frequently encountered.

3.16 Mercury-199

Mercury has a spin-$\frac{1}{2}$ isotope, ^{199}Hg. The advent of modern instrumentation made it possible to observe this nucleus rather easily. Chemical shift and coupling constant data have been reviewed.[25]

3.16.1 Experimental Practices

Mercury is observed in the mid-range of most broadband probes. The sensitivity is adequate and no special techniques are required for detection at the millimolar level.

3.16.2 Physical Properties

With nearly 17% natural abundance and a moderately high observe frequency, ^{199}Hg is in the middle range of

sensitivity. The chemical shift range is from about 0 ppm to −2500 ppm, referenced to external dimethylmercury, $Hg(CH_3)_2$. Although dimethylmercury is the accepted standard for chemical shift and all published values are referenced to it, the material represents an extreme health hazard. Contact can be fatal. It is easily absorbed through the skin and even through latex gloves. There is absolutely no place for this compound in the NMR laboratory. Alternative reference materials, including mercury salts, such as $Hg(ClO_4)_2$ or $Hg(NO_3)_2$, may be used, although the chemical shifts are concentration dependent.

3.16.3 Chemical Shift

Mercury compounds have a wide chemical shift range. Most organo-mercurials have chemical shifts from 0 ppm to −1000 ppm. A few compounds with electropositive substituents have positive chemical shifts such as $Hg(SiH_3)_2$ at 196 ppm. Most mercury salts have chemical shifts in the range of −1000 ppm to −3000 ppm.

3.16.4 Coupling

Mercury couples to proton in the 100–300 Hz range for two-bond coupling. Directly bonded carbon couplings are about 600–1800 Hz. Directly bonded mercury–phosphorus couplings are about 3700–5000 Hz. Two-bond couplings between mercury and fluorine have been measured to be 1200–2000 Hz.

3.16.5 Relaxation

Relaxation in mercury compounds is similar to other spin-$\frac{1}{2}$ nuclei, ranging from a few tenths of a second to a few seconds.

3.17 Thallium-203 and Thallium-205

By virtue of their sensitivity, ^{203}Tl and ^{205}Tl were among the first nuclei observed. The more abundant nuclide, ^{205}Tl, is almost 14% as sensitive as proton, making it the fourth most sensitive of the spin-$\frac{1}{2}$ nuclei. Thallium is very well suited for NMR use, having an extremely large chemical shift range and extremely large coupling constants. Both the chemical shifts and coupling constants change dramatically when the thallium environment changes. Temperature, solvent, and other matrix changes, as well as substituent changes, produce some of the largest chemical shift and coupling constant changes of any of the spin-$\frac{1}{2}$ nuclei. This makes thallium one of the most sensitive known probes of the electronic environment. A wealth of literature was comprehensively reviewed in 1988.[36] A recent overview of the field is also available.[37] Thallium has been used as a surrogate for potassium and sodium in biological studies, because it is much more likely to show subtle effects in its environment.

3.17.1 Experimental Practices

As both thallium isotopes are relatively sensitive, observation of NMR signals is rapid and easily accomplished. Thallium concentrations in the millimolar and micromolar range are sufficient for obtaining spectra. Relaxation is rapid enough so that acquisition can proceed quickly. The very properties that make thallium so sensitive to environmental effects also cause some experimental difficulties. The temperature must be very closely controlled. Other matrix effects must also be carefully watched. The large chemical shift range must be accommodated by large data file sizes and large observe window widths. Although no special precautions must be taken, good laboratory practice is essential due to the toxicity of thallium compounds.

3.17.2 Physical Properties

The similarity of the NMR properties of the two nuclides is reflected in the similarity of the observe frequencies: 173.1 MHz for ^{205}Tl and 171.5 MHz for ^{203}Tl. The chief reason for the relative sensitivity between the two nuclides is their natural abundance: 70.5% for ^{205}Tl and 29.5% for ^{203}Tl. The total chemical shift range for thallium is about 7000 ppm. Thallium is commonly found in two oxidation states: Tl^+, with a chemical shift range of 3000 ppm, and Tl^{3+} with a chemical shift range of about 7000 ppm. Thallium chemical shifts are extremely solvent dependent, and the temperature dependence is approximately 5 ppm per kelvin. The temperature dependence of thallium chemical shifts is so strong that chemical shift can be used to measure the temperature of the sample in the probe quite accurately. Chemical shifts are referenced to $Tl(NO_3)$.

3.17.3 Chemical Shift

Thallium chemical shifts are strongly influenced by temperature, solvent and other matrix effects and the nature of the attached substituents. The substituent effects are extremely large. For example, the change in chemical shift from $Tl(CH_3)_3$ to $Tl(CH_2CH_3)_3$ is 130 ppm.

3.17.4 Coupling

As seen with thallium chemical shifts, thallium coupling constants are large. Directly bonded protons show coupling constants of about 6000 Hz, directly bonded carbons about 10 000 Hz or larger and directly bonded phosphorus about 3000 Hz. As expected from these

large values, long-range couplings are also quite large. Even five- or six-bond couplings to proton are in the range of 100 Hz, which is comparable to directly bonded carbon–hydrogen coupling. These extremely long-range effects make thallium an excellent probe of electronic environmental effects for mechanistic studies, solvation and ion-pair studies and the study of intermolecular effects.

3.17.5 Relaxation

The relaxation behavior of both nuclides is the same and, as expected, extremely sensitive to environmental effects. Typical relaxation times are in the range of a few seconds. Thallium relaxation times are very sensitive to the presence of paramagnetic species in solution, which can drastically reduce the relaxation time to the millisecond range.

3.18 Lead-207

The NMR behavior of ^{207}Pb resembles that of ^{119}Sn, but on a larger scale. The atom is larger, and both the chemical shift range and coupling constant range are correspondingly larger. With a moderately high positive magnetogyric ratio and high natural abundance, sensitivity is good. The chemical shifts of ^{29}Si, ^{119}Sn and ^{207}Pb are related by linear equations. The literature on chemical shift and coupling constants has been reviewed.[26,27]

3.18.1 Experimental Practices

With modern instrumentation, the observation of ^{207}Pb is routine. Pulse repetition rates can be fairly rapid. A wide spectral width and large data sets are required for high resolution.

3.18.2 Physical Properties

The physical properties of ^{207}Pb are listed in Table 2. The Larmor frequency is easily found in the middle range of broadband probes. With 22.6% natural abundance, ^{207}Pb is rather sensitive. Lead spans an extremely large chemical shift range, from 10 000 to −6000 ppm. The large chemical shift range makes ^{207}Pb very sensitive to temperature, solvent and structural environment. Chemical shifts are referenced to tetramethyllead, $Pb(CH_3)_4$.

3.18.3 Chemical Shift

Lead chemical shifts are similar, and most are linearly related to ^{119}Sn chemical shifts. Electropositive substituents produce highly shielded chemical shifts, and in the normal fashion, electronegative substituents produce highly deshielded chemical shifts. Most four-coordinate lead compounds resonate in the range of 500 ppm to −800 ppm. Aromatic shieldings in the plumbocenes give chemical shifts in the −5000 ppm area.

3.18.4 Coupling

Lead couplings are large, and for the most part are linearly related to ^{119}Sn coupling constants. Directly bonded protons give couplings of about 2000 Hz. Coupling to carbon ranges from −1000 Hz to 250 Hz. Coupling to other metals produces extremely large coupling constants. Coupling to ^{119}Sn may be as large as −3600 Hz and coupling in some platinum compounds may be larger than 18 000 Hz.

3.18.5 Relaxation

Lead relaxation is very similar to the other spin-$\frac{1}{2}$ nuclei. In asymmetric compounds, linewidths of ^{207}Pb spectra may be smaller at lower magnetic field strengths due to chemical shift anisotropy.

ABBREVIATIONS AND ACRONYMS

DEPT	Distortionless Enhancement by Polarization Transfer
FT	Fourier Transform
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
TMS	Tetramethylsilane

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Multinuclear Magnetic Resonance Spectroscopic Imaging

Chemical Weapons Chemicals Analysis **(Volume 2)**
Nuclear Magnetic Resonance Spectroscopy in Analysis of Chemicals Related to the Chemical Weapons Convention

Food **(Volume 5)**
Nuclear Magnetic Resonance in Analysis of Plant Soil Environments • Nuclear Magnetic Resonance in the Analysis of Foodstuffs and Plant Materials

Forensic Science **(Volume 5)**
Nuclear Magnetic Resonance Spectroscopy for the Detection and Quantification of Abused Drugs

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Nuclear Magnetic Resonance Characterization of Petroleum

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Steel and Related Materials **(Volume 10)**
Nuclear Magnetic Resonance in Metals Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

REFERENCES

1. R.K. Harris, B.E. Mann (eds.), *NMR and the Periodic Table*, Academic Press, London, 1978.
2. J.B. Lambert, F.G. Riddell (eds.), *The Multinuclear Approach to NMR Spectroscopy*, Reidel, Dordrecht, 1983.
3. P. Laszlo (ed.), *NMR of Newly Accessible Nuclei*, Academic Press, New York, 1983.
4. J. Mason (ed.), *Multinuclear NMR*, Plenum Press, New York, 1987.
5. D.M. Grant, R.K. Harris (ed.), *The Encyclopedia of NMR*, Wiley, New York, 1995.
6. R.J. Abraham, J. Fisher, P. Loftus, *Introduction to NMR Spectroscopy*, Wiley, New York, 1988.
7. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York, 1989.
8. M. McGregor, 'Nuclear Magnetic Resonance Spectroscopy', in *Handbook of Instrumental Techniques for Analytical Chemistry*, ed. F. Settle, Prentice Hall, New York, Chapter 17, 1997.
9. G.C. Levy, R.L. Lichter, ^{15}N *NMR Spectroscopy*, Wiley-Interscience, New York, 1979.
10. M. Witanowski, L. Stefaniak, G.A. Webb, *Annu. Rep. NMR Spectrosc.*, **25**, 1 (1993).
11. J. Mason, 'Nitrogen NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 3222–3250, 1995.
12. J.Mason, 'Nitrogen NMR', in *Multinuclear NMR*, ed. J. Mason, Plenum Press, New York, Chap. 12, 207–244, 1987.
13. C.H. Dungan, J.R. van Wazer, *Compilation of Reported F-19 NMR Chemical Shifts*, Wiley-Interscience, New York, 1970.
14. J.W. Emsley, L. Phillips, 'Fluorine Chemical Shifts', *Prog. NMR Spectrosc.*, **7**, 1 (1971).
15. J.W. Emsley, L. Phillips, V. Wray, 'Fluorine Coupling Constants', *Prog. NMR Spectrosc.*, **10**, 83 (1976).
16. W.S. Brey, M.L. Brey, 'Fluorine-19 NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 2063–2071, 1995.
17. H.C. Marsmann, 'Silicon-29 NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 4386–4397, 1995.
18. H.C. Marsmann, 'Silicon-29 NMR', *NMR Basic Principles Prog.*, **17**, 65 (1981).
19. J.C. Tebby (ed.), *Handbook of Phosphorus-31 Nuclear Magnetic Resonance Data*, CRC Press, Boca Raton, FL, 1991.
20. J.G. Verkade, L.D. Quin (eds.), *Phosphorus-31 NMR Spectroscopy in Stereochemical Analysis*, VCH, Deerfield Beach, FL, 1987.
21. L.D. Quin, J. Verkade (eds.), *Phosphorus-31 NMR Special Properties in Compound Characterization and Structural Analysis*, VCH, Deerfield Beach, FL, 1994.
22. K. Karaghiosoff, 'Phosphorus-31 NMR' in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 3612–3618, 1995.
23. E.A. Evans, D.C. Warrell, J.A. Elvidge, J.R. Jones, *Handbook of Tritium NMR Spectroscopy and Applications*, Wiley, Chichester, 1–249, 1985.
24. M.G. Kubinec, P.G. Williams, 'Tritium NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 4819–4830, 1995.
25. R.G. Kidd, R.J. Goodfellow, 'The Transition Metals', in *Multinuclear NMR*, ed. J. Mason, Plenum Press, New York, 258–260, 1987.
26. B. Wrackmeyer, *Annu. Rep. NMR Spectrosc.*, **16**, 73 (1985).
27. B. Wrackmeyer, 'Germanium, Tin and Lead NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 2206–2214, 1995.
28. H. Duddeck, 'Sulfur, Selenium and Tellurium NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 4623–4635, 1995.
29. C. Rodger, N. Sheppard, H.C.E. McFarlane, W. McFarlane, in *NMR and the Periodic Table*, eds. R.K. Harris, B.E. Mann, Academic Press, London, 383, 1978.
30. H.C.E. McFarlane, W. McFarlane, in *Multinuclear NMR*, ed. J. Mason, Plenum Press, New York, 417, 1987.
31. N.P. Luthra, J.D. Odom, in *The Chemistry of Organic Selenium and Tellurium Compounds*, eds. S. Patai, Z. Rappoport, Wiley, Chichester, 189, Vol. 1, 1986.
32. G. Schrobilgen, 'Noble Gas Elements', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 3251–3261, 1995.

33. G. Schrobilgen, in *NMR and the Periodic Table*, eds. R.K. Harris, B.E. Mann, Academic Press, London, Chapter 14, 439–454, 1978.
34. B. Wrackmeyer, K. Horchler, *Annu. Rep. NMR Spectrosc.*, **22**, 249 (1989).
35. C.J. Jameson, in *Multinuclear NMR*, ed. J. Mason, Plenum Press, New York, Chapter 18, 463–477, 1987.
36. J.F. Hinton, K.R. Metz, R.W. Briggs, *Prog. Nucl. Magn. Reson. Spectrosc.*, **20**(5), 423 (1988).
37. J.F. Hinton, 'Thallium NMR', in *The Encyclopedia of NMR*, eds. D.M. Grant, R.K. Harris, Wiley, New York, 4702–4710, 1995.
38. J.D. Kennedy, W. McFarlane, in *Multinuclear NMR*, ed. J. Mason, Plenum Press, New York, 305–333, 1987.

Two-dimensional Nuclear Magnetic Resonance of Small Molecules

Henrik Pedersen
H. Lundbeck A/S, Copenhagen-Valby, Denmark

1 Introduction 1
2 The Second Dimension in Nuclear Magnetic Resonance 2
2.1 The Birth of a Two-dimensional Spectrum 2
2.2 Choice of Basic Parameters 2
2.3 Processing 3
2.4 Artifacts 3
2.5 Sensitivity Considerations 3
2.6 Solvent Suppression 3
3 Instrumentation 3
3.1 Magnets 3
3.2 Probe Heads 4
3.3 Pulsed Field Gradients 4
3.4 Console and Computer 5
4 Homonuclear Correlation Experiments 5
4.1 The Correlation Spectroscopy Experiment and Variants 5
4.2 The Total Correlation Spectroscopy Experiment 6
4.3 The Nuclear Overhauser Enhancement Spectroscopy Experiment 6
4.4 The Rotating Frame Overhauser Enhancement Spectroscopy Experiment 8
4.5 Incredible Natural Abundance Double Quantum Transfer Experiment 8
5 Homonuclear *J*-Resolved Spectroscopy 9
6 Heteronuclear Correlation Experiments 9
6.1 Heteronuclear Correlation with ^{13}C Detection 9
6.2 Heteronuclear Experiments with ^{1}H Detection 9
7 Pseudo-three-dimensional Experiments 12
8 Two-dimensional Nuclear Magnetic Resonance on Spin-$\frac{1}{2}$ Nuclei Other than ^{1}H and ^{13}C 14
9 Two-dimensional Nuclear Magnetic Resonance on Resin-bound Compounds 14
Abbreviations and Acronyms 15
Related Articles 16
References 16

The article deals with practical aspects in obtaining two-dimensional (2-D) nuclear magnetic resonance (NMR) spectra of small molecules [molecular weight (MW) < 1000]. This includes the choice of parameters for acquisition and processing and the required and optional instrument hardware. Examples of 2-D experiments are given with emphases on robustness and ease of use. The pulse sequences are given in schematic form together with suggestions for basic parameters and the phase cycle. The experiment examples include homonuclear and heteronuclear experiments with direct and indirect detection. Also the use of pulsed field gradients (PFGs) is exemplified. 2-D NMR on spin-$\frac{1}{2}$ nuclei other than ^{1}H and ^{13}C is exemplified by ^{19}F and ^{15}N experiments. Experiments on gel-phase samples under magic angle conditions are mentioned briefly.

1 INTRODUCTION

Since the first 2-D NMR experiment was suggested by Jeener[1] in 1971, an enormous number of experiments have been designed and published. In the last 10 years the major development of new pulse sequences has been associated with biomolecular applications, but some of the sequences may also be very useful for structural elucidation of small molecules.

The 2-D NMR techniques have been pushed forward by the continuing developments in spectrometer hardware, including higher magnetic fields, improved probe heads, increased stability and performance of the electronics and the host computers. The incorporation of a PFG in one or three dimensions has given access to shorter acquisition times and spectra with less artifacts. Also, the development of mathematical calculation tools has facilitated the development of new 2-D NMR experiments. For example, the product operator formalism developed by Sørensen et al.[2] makes it possible to calculate the fate of the magnetization at any point of a pulse sequence, and also helps in the design of new sequences.

Given the enormous number of possible 2-D pulse sequences, it is important for the chemist or spectroscopist, who wants to use 2-D NMR in structure elucidation or for spectral assignment, to be familiar with a relatively small number of experiments which can be employed on a more or less routine basis. The actual choice will be dependent on the available hardware and

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the degree of desirable user interaction. A pulse sequence which is clamed to give better resolution and/or higher sensitivity may be very sensitive to improper setting of the pulse widths or B_1 inhomogeneity of the probe head. Also, relaxation phenomena may be a limitation for sophisticated long-lasting sequences.

The term ''small molecules'' here refers to organic molecules of MW $< 1000\,\mathrm{g\,mol^{-1}}$. This will include many synthetic organic molecules, including drugs, but also oligosaccharides, peptides and nucleotides may fall in this category.

In this article, a basic knowledge of one-dimensional (1-D) NMR is presumed. Readers who want in-depth knowledge of the rather complex theory of 2-D NMR are referred to several excellent textbooks.[3–5] A very comprehensive overview with practical suggestions for many 2-D NMR experiments is given in the book by Braun et al.[6]

Programming of pulse sequences from literature sources is not a trivial task, but the instrument vendors normally supply an extensive library of pulse programs, from which the user may choose his or her favorites.

This article will focus on relatively few and simple experiments, which alone or in combination can give appropriate structural information. This may inspire the user to put together a portfolio of experiments depending on the needs and degree of instrument sophistication available.

2 THE SECOND DIMENSION IN NUCLEAR MAGNETIC RESONANCE

2.1 The Birth of a Two-dimensional Spectrum

A 2-D NMR spectrum consists of a series of 1-D NMR spectra arranged in a rectangular matrix and Fourier transformed in both dimensions. The individual 1-D spectra are acquired with the same pulse sequence, and the only difference is the duration of a delay in the sequence. A schematic representation of the building blocks in a 2-D experiment is given in Figure 1.

During the preparation period, the spins of the observed nuclei are allowed to relax back to the z-axis (along the external field) and then receive the first pulse or pulses that serve to bring magnetization of the nuclei into the xy-plane. The variable (incremented time) follows as the evolution period in which chemical shifts and couplings evolve to give the wanted correlation types. Observable transverse magnetization is created during the mixing period by pulses and delays, and finally the signal is acquired during the detection period (t_2) as a free induction decay (FID). The effect of incrementing the evolution time through a series of spectra is that the signal intensity and phase vary. When these spectra are arranged as rows in a matrix, the resulting columns of the matrix define a new set of FIDs. Subsequent Fourier transformation of this set of FIDs gives the 2-D spectrum.

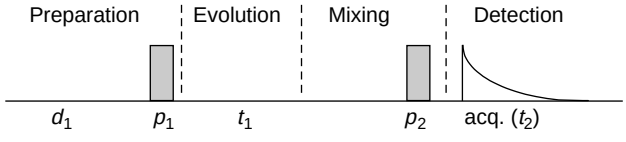

Figure 1 Schematic representation of a 2-D pulse sequence.

Depending on the experiment and the acquisition and/or processing method, the resulting spectrum may be phase sensitive with Lorentzian line shapes in both dimensions or it may be a magnitude spectrum. The latter takes up much less disk space and may be acquired in half the time, but the phase-sensitive spectrum gives better resolution. A magnitude spectrum may always be calculated from a phase-sensitive dataset, but not vice versa. In practice, phase-sensitive 2-D data are acquired by incrementing phases for each t_1 value by one of two methods, the States–Haberkorn–Ruben method[7] and the TPPI (time proportional phase increment) method.[8] The latter method relies on 90° step incrementation of the first 90° pulse and is used in the examples in this article.

2.2 Choice of Basic Parameters

The sweep width and spectrometer offset for a 2-D experiment are chosen from the corresponding 1-D spectra. If digital filters are available, the sweep width in the first dimension (referred to as F_2) can be set to focus on a minor part of the spectrum. If only analog filters are available, folding of signals from outside the sweep width around the spectral edges may occur. In the second dimension (referred to as F_1), cross peaks that lie outside the chosen sweep width in F_1 will always be folded.

In order to save hard disk space and to reduce processing time, it is an advantage to keep the resulting data matrix as small as possible. A $4\,\mathrm{K} \times 4\,\mathrm{K}$ matrix is the upper practical limit with state-of-the-art computers and software. For acquisition, the time requirement relies heavily on the number of individual serial files in the F_1 time domain. In most cases this can be kept as low as 128–256. If a greater number of serial files are collected, the signal-to-noise ratio (S/N) will decrease because of relaxation during the prolonged evolution delay in the last FIDs. If very detailed information is required (e.g. coupling constants), it can be advantageous to restrict the sweep area to the correlation cross peaks of interest, rather than to increase the digital resolution for the whole spectrum.

If vendor-supplied pulse programs are used, suitable choices for other parameters are usually proposed by

the vendor. A knowledge of 90° pulse lengths at the appropriate transmitter or decoupler power levels is, of course, necessary prior to setting up a 2-D experiment.

2.3 Processing

Prior to Fourier transformation, window functions are normally multiplied with the FID. A wide variety of functions are available, and the choice may have a great impact on the resulting spectrum. A popular window function is a shifted or unshifted sine-bell or squared sine-bell function.

The spectral resolution in F_1 can be enhanced by zero filling, i.e. transforming with a size greater than the number of increments. This will always lead to the formation of wiggles around the signals, but the level of wiggles may well be acceptable with two or four times zero filling. The newer generations of processing software offer the possibility of linear prediction in F_1.[9] The idea is that when the FID is cut even after a relatively short time in F_1, it contains all the needed information, and the further development of the FID may be calculated from the existing data. This means that linear prediction can be used as a kind of intelligent zero filling. If the S/N is not too low, linear prediction may be used routinely to double the number of data points, thus giving increased resolution or a shorter total experiment time. In fact, all the spectral examples shown in this article have been treated in this way.

2.4 Artifacts

Several types of unwanted peaks may occur in the resulting 2-D spectrum. Errors in the phase cycling scheme may give rise to unwanted coherences, and poor calibration of the pulse widths may give artifacts, especially when the full phase cycle is not obeyed. Wrong pulse lengths may in turn be a result of poor tuning and matching of the probe head.

Another phenomenon that is often encountered is the so-called t_1 noise, which appears as ridges in the vertical direction of the spectrum. The causes of this are all kinds of instrument instabilities, e.g. temperature variations, magnetic field disturbances and pulse instabilities.

Cosmetic improvements may be obtained by subtraction of rows that only contain noise peaks. This is particularly useful for magnitude spectra and spectra where all peaks of interest have the same sign.

2-D spectra of a symmetrical nature such as correlation spectroscopy (COSY), nuclear Overhauser enhancement spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) may be symmetrized about the diagonal. The algorithm compares points which are mirror images around the diagonal and exchange the greatest value with the smaller value. The major drawback with the method is that although t_1 noise ridges will disappear, they will leave some false cross peaks.

2.5 Sensitivity Considerations

It is impossible to give general data for sensitivity and detection limits for 2-D NMR experiments, since they depend entirely on the performance of the actual instrument. For 1–2 mM solutions of small molecules, homonuclear correlations and inverse detected proton–carbon correlation experiments should be obtainable overnight or over the weekend on instruments with a basic frequency of 200 MHz or more. A way of testing the chance of success is to acquire a single serial file, process it and judge the S/N. With some experience the possibility of obtaining 2-D spectra may be judged from the 1-D spectra. In the examples shown in later sections, the sample concentrations and total experiment times are given.

If only very small sample amounts (micrograms) are available, the instrument vendors offer probe heads which give improved sensitivity for small sample volumes per unit of available sample.

2.6 Solvent Suppression

Several factors limit the dynamic range of an NMR spectrometer. Among these are the limited resolution of the available analog-to-digital converters (ADCs) and the performance of the amplification stages. Therefore, it is essential to reduce the signals arising from dominating solvent molecules. Selective irradiation of the solvent proton peaks with a weak radiofrequency (RF) field during the relaxation delay is fairly efficient, and in the case of water it may be invoked in all ^{1}H detected experiments. If more solvent peaks are to be suppressed, an irradiation cycle where the irradiation frequency is moved from one peak to another with short intervals can be applied.

In heteronuclear correlation (HETCOR) experiments, water is the only solvent which is easily dealt with. For biomolecular applications, more sophisticated water suppression sequences may be employed, e.g. WATERGATE (water suppression by gradient tailored excitation) and variants[10,11] and WET (water suppression enhanced through T_1 effects),[12] which also requires the use of shaped pulses.

3 INSTRUMENTATION

3.1 Magnets

All types of magnets can in principle be used for 2-D NMR, but in practice only superconducting magnets are

useful, because they combine high field and resolution with good magnetic field stability. Even very small changes in the magnetic field during a 2-D acquisition will give rise to major line artifacts (t_1 noise) in the F_1 dimension. The magnetic field stability is greatly influenced by magnetic disturbances in the vicinity of the magnet in the form of moving metallic objects. This can be partly compensated by the new generation of digital lock systems. Otherwise it may be necessary to perform demanding experiments during the night or other quiet periods.

As in 1-D NMR, the resolution obeys a linear relationship with the field strength whereas the sensitivity in theory increases with the square of the field strength. In practice, other hardware limitations (probe heads, transmitters, etc.) cause the exponent of the relationship for the sensitivity to be around 1.5. Although high field strengths give the best resolution and sensitivity, there are also some drawbacks. With increasing field, the time required for shimming (optimization of the field homogeneity) will rise and so will the need for tuning and matching of the probe head for the individual samples. In practice, 500 MHz can be considered the upper limit for easy operation when working with samples of different nature in different solvents. However, the problem with tuning and matching may be circumvented by the use of auto-tune probes, which have recently been introduced by the vendors.

3.2 Probe Heads

The part of an NMR spectrometer which has the greatest impact on instrument performance at given magnetic field is the probe head, and new designs have drastically improved the sensitivity of modern probe heads.

A wide variety of probe designs are available, and most of them allow for pulsing on two different nuclei at the same time (double resonance probes) or three different nuclei at the same time (triple resonance probes). The probes fall into two major classes: proton sensitivity optimized (often called inverse probes) and X-nucleus sensitivity optimized (often called normal probes).

A double resonance probe will normally contain two RF coils. The inner coil, which is closest to the sample, will afford the highest sensitivity because of the better coupling between RF coil and sample. The gain in sensitivity for a given nucleus on going from the outer to the inner coil is roughly a factor of two. While the highest possible proton sensitivity is of major importance in biomolecular NMR, small molecule applications will often, if enough sample is available, benefit from probes which are easiest to work with. Here the normal probes benefit from easier shimming and less sensitivity to tuning and matching of the coils. If sample spinning is used, the use of inverse probes will often give the so-called Q-modulations, which show up as antiphase spinning sidebands. These artifacts arise from poor symmetry of the sample tube and poor concentricity of the spinner turbine and sample with the inner coil, and, unlike in-phase spinning sidebands, they cannot be removed by shimming. However, most experiments will work on superconducting magnets without sample spinning.

To maintain sufficient chemical shift stability, temperature control of the probe head should always be applied. The target temperature should be set at least 5 °C above the inlet temperature of the heating gas. This is especially important when decoupling sequences are employed, since these may give rise to substantial heating of the sample.

3.3 Pulsed Field Gradients

Although PFGs are the basis for magnetic resonance imaging and volume-selective NMR spectroscopy, the technique is relatively new in general solution-phase NMR. They do, however, represent a major step forward in reducing experiment time and spectral artifacts.

A PFG pulse in the z-axis direction will defocus all magnetization except for that aligned with the z-axis. The z-magnetization may then be brought back into the xy-plane by a 90° pulse, and is thus purified from unwanted magnetization. This is often referred to as purging gradients.

By using a set of two or more matched PFGs, the magnetization in the xy-plane may first be defocused, then manipulated by pulses and finally a PFG can refocus the magnetization of interest without bringing back magnetization arising from unwanted coherence. Hence the great advantage in using matched PFGs is that it is so efficient at selecting only wanted coherence that the need for phase cycling is circumvented, with the consequence that very short experiment times may be allowed, provided that the S/N is sufficient. Especially in indirectly detected heteronuclear experiments, the unwanted magnetization from protons attached to ^{12}C is very efficiently suppressed compared with other methods. When PFGs are used for coherence selection in HETCOR, the ratios between the individual gradients are dependent on the X-nuclei involved.

Weak PFGs in the z-axis direction may be generated from the z-shim coil, but these are only sufficient for homonuclear experiments. To achieve the full benefit of PFGs, a dedicated probe head with a gradient coil and a gradient amplifier is needed. Although triple axis PFG accessories, which can give gradient pulses of all spatial orientations, are available, small molecule work will hardly benefit from more than single-axis gradient equipment.

3.4 Console and Computer

A prerequisite for doing 2-D NMR experiments is that the instrument can work in the pulse Fourier transform mode. Modern instruments are fairly flexible and may be configured for the wanted experiment by computer control. Older instruments are often less flexible and may require some rewiring, if for instance pulsing with the proton decoupler and simultaneous proton observation are required.

2-D NMR has been greatly facilitated by the development of fast computers working under UNIX or another modern operating system, which allows for fast matrix processing and display. In the case where the spectrometer is equipped with an old, slow computer, raw data may be transferred to a modern computer and processed, plotted and displayed from there with an appropriate software package.

4 HOMONUCLEAR CORRELATION EXPERIMENTS

4.1 The Correlation Spectroscopy Experiment and Variants

The COSY experiment[13] is one of the most popular 2-D experiments in NMR, and it exists in numerous variations. The resulting 2-D NMR spectrum (Figure 2) of the antipsychotic drug sertindole (CAS 106516-24-9)[14] shows the 1-D spectrum on the diagonal and *J*-coupling correlation as off-diagonal cross peaks. The pulse sequence and phase cycle with optional PFGs are given in Figure 3. The resulting spectrum is in magnitude mode. The second 90° pulse may be changed to 45°, and at some expense of sensitivity this will narrow the diagonal, and thus ease the interpretation of tightly coupled spin systems. With the PFG version a total acquisition time of less than 5 min may be achieved, provided that enough sample is available.

If enhanced spectral resolution is required, double quantum filtered correlation spectroscopy (COSYDQ) (Figure 4)[15] may be the experiment of choice, since this may be performed in the phase-sensitive mode without giving dispersive diagonal signals. The double quantum filter serves to remove noncoupled signals (singlets) from the diagonal. Especially the PFG version[16] of this experiment (Figure 5) is effective in reducing singlet solvent lines. This is shown by the PFG/COSYDQ spectrum of sucrose (CAS 57-50-1) in 90% water (Figure 6).

In the case where the size and the sign of proton–proton coupling constants are wanted for complicated spin systems, a variant called E.COSY (Exclusive Correlation Spectroscopy)[17] gives simplified cross peaks, where a given multiplet component correlates selectively with only two components of its *J*-coupled proton, and not with all of its components as is the case in a normal COSY experiment. Since different spin system order requires different phase cycles, the experiment is hardly useful for routine work. In less demanding cases a COSY experiment where the second pulse is set to 45° instead of 90° may give a similar result (COSY-45).

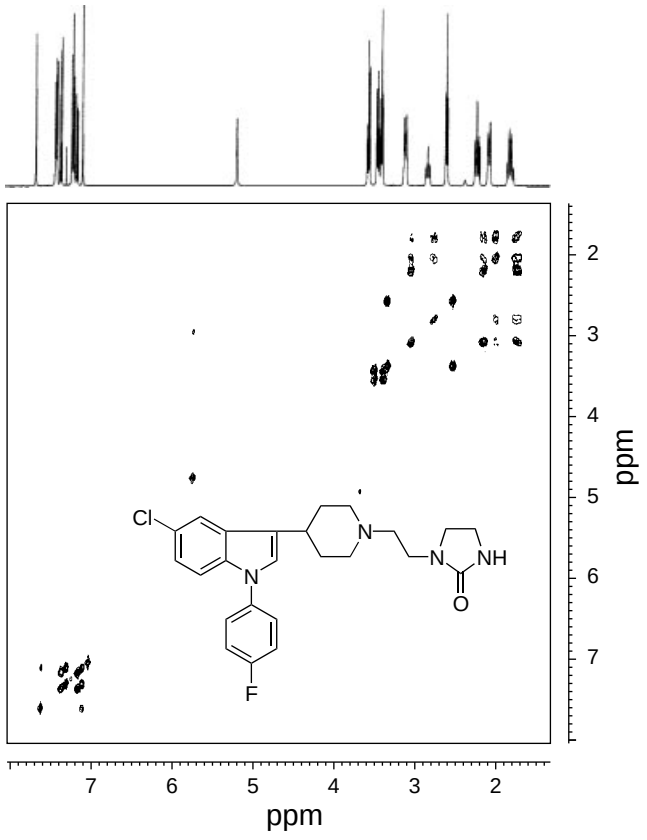

Figure 2 PFG/COSY of 30 mM (10 mg) sertindole in chloroform-*d* at 500 MHz with 128 increments of 1024 points and one scan per increment. Total experiment time was 3 min 45 s and d_1 was 1 s.

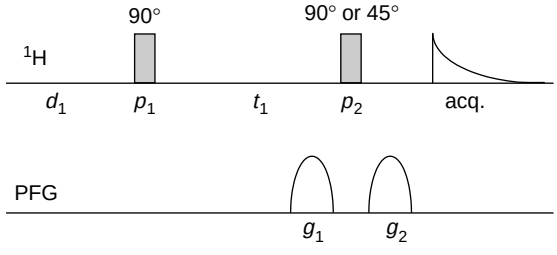

Figure 3 The pulse sequence for COSY with optional PFGs. Without PFGs the phase cycle is: p_1: $(x)_4$, $(y)_4$, $(-x)_4$, $(-y)_4$; p_2: $x,y,-x,-y$; acq.: $x, -x,x,-x, -y,y,-y,y, -x,x,-x,x, y,-y,y,-y$ with a minimum cycle of four. With PFGs this reduces to: p_1: x,x; p_2: $x,x,-x,-x$; acq.: $x,-x$. The gradient ratio is 1 : 1.

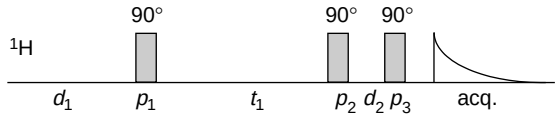

Figure 4 The pulse sequence for COSYDQ. Delay d_2 is short (typically 3 μs). The phase cycle is: p_1: $(y)_4$, $(x)_4$; p_2: $(x)_4$, $(y)_4$; p_3: $y,-x,-y-x$, $-x,-y,x,y$; acq.: $x,-x,-y,y$, $-y,-x,y,x$. The phase of p_1 is incremented for TPPI.

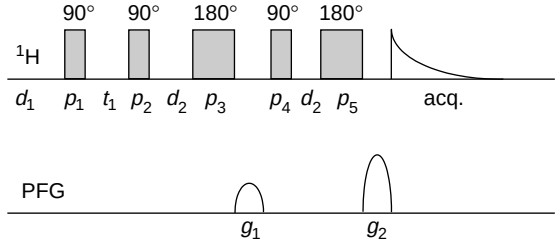

Figure 5 The pulse sequence PFG/COSYDQ. Delay d_2 is gradient pulse length plus ring downtime. The phase cycle is: p_1: $x,-x$; p_2, p_3, p_4, p_5; x; acq.: $x,-x$. The phase of p_1 is incremented for TPPI.

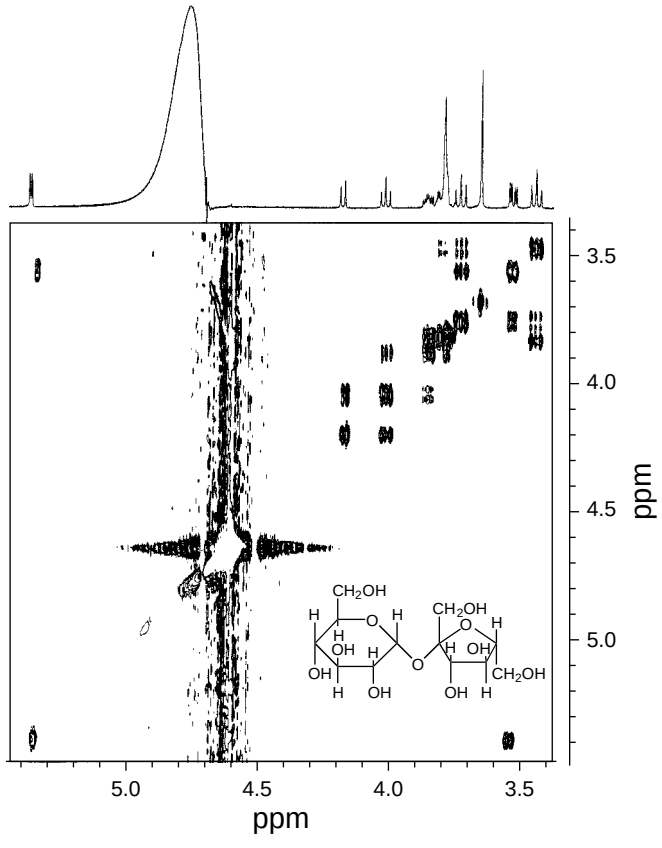

Figure 6 The PFG/COSYDQ spectrum of 2 mM sucrose in H_2O-D_2O (90:10) at 500 MHz with 256 increments of 1024 points and 32 scans per increment. The delays were $d_1 = 1.5$ s and $d_2 = 17$ ms. The total experiment time was 5 h 9 min.

4.2 The Total Correlation Spectroscopy Experiment

The TOCSY or HOHAHA (homonuclear Hartmann–Hahn)[16,18] is probably the most popular member of the family of so-called spin-lock experiments. After the initial 90° pulse, a strong field is applied which serves to lock the magnetization and avoid evolution of individual spins. The effect is that protons in a common coupling network will all correlate if the duration of the spin-lock field is sufficiently long. A COSY-like spectrum may, on the other hand, be obtained if the spin-lock time is short enough. This may give an advantage since the TOCSY cross peaks are in phase. A spin lock of 10–20 ms will give COSY-like spectra and 40–100 ms will correlate more spin systems in a coupling network. The spin lock may be generated in several different ways, but a very common way is to use the composite pulse train MLEV (Malcom Levitts)[18] surrounded by trim pulses.

The TOCSY sequence with optional PFGs is shown in Figure 7. The details of the MLEV cycle are not given because of its complexity. On older instruments the hardware must be configured in the inverse mode and the power of the preparation pulse must be lowered to that of the spin lock.

In the PFG/TOCSY of sertindole (Figure 8) several extra cross peaks are seen compared with the COSY spectrum in Figure 2.

4.3 The Nuclear Overhauser Enhancement Spectroscopy Experiment

The NOESY[7] experiment (Figure 9) correlates protons which are close in space through dipolar coupling. Whereas NOESY is an experiment of major importance in determining the structure of large molecules by distance constraints, the situation is less clear cut for small molecules. This is because the cross-peak sign and intensity are related to the tumbling rate of the molecules in solution and therefore to MW. Generally,

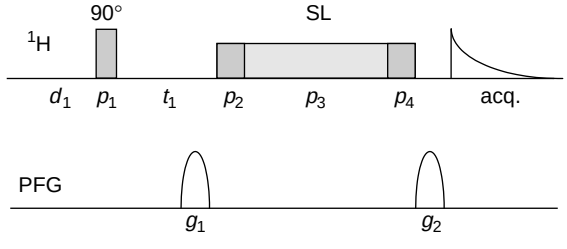

Figure 7 The pulse sequence for TOCSY with optional PFGs. The nongradient version may be acquired phase sensitive with incrementation of p_1 for TPPI using a minimum number of scans of four. In the PFG version the number of scans may be one per increment. The gradient ratio is 1:1. Alternatively, a gradient ratio of $1:-1$ may be used. In this case no trim pulses are needed, thus giving a higher S/N.

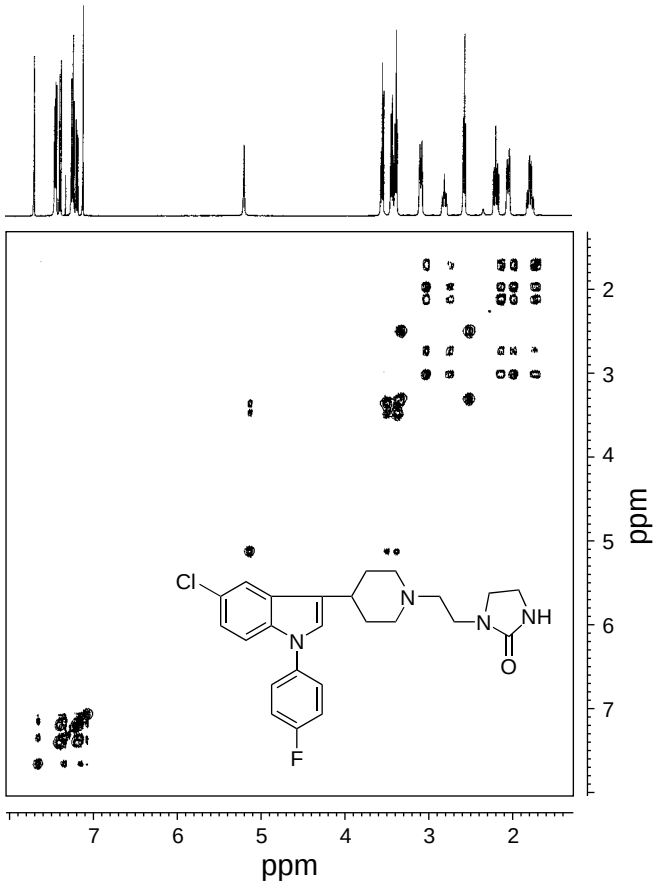

Figure 8 The TOCSY spectrum of 30 mM (10 mg) sertindole in chloroform-*d* at 500 MHz with 128 increment of 1024 points and one scan per increment. The delay d_1 was 1 s and the spin-lock time including two trim pulses of 2.5 ms was 75 ms.

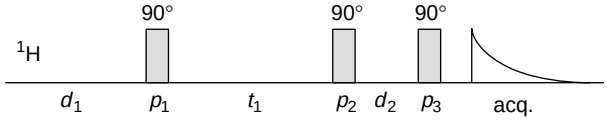

Figure 9 The pulse sequence for NOESY. The phase cycle is: p_1: $x,-x$; p_2: $(x)_8,(-x)_8$; p_3: $xx,-x,-x,y,y,-y,-y$; acq.: $x,-x,-x,x,y,-y,-y,y,-x,x,x,-x,-y,y,y,-y$. The minimum number of scans is eight. The first pulse is incremented for TPPI.

small molecules are in the positive nuclear Overhauser enhancement (NOE) regime, but depending on the molecule, the solvent and the field strength, the NOE effect may be close to zero. It is therefore often preferable to perform 1-D NOE experiments instead, either as an alternative or as a test for the chance of success in the 2-D case. The NOESY should always be performed in the phase-sensitive mode in order to distinguish the NOESY peaks from residual antiphase COSY peaks. The COSY peaks may be reduced by random variation of the delay d_2. The sign of NOESY cross peaks will be opposite of that of the diagonal (for small molecules). An important parameter is the mixing time (d_2). For small molecules it should be set to the approximate longitudinal relaxation time T_1 or lower.

Exchange phenomena between loosely bound protons or slowly exchanging rotamers may give intense cross peaks that are in phase with the diagonal and the NOESY experiment is therefore valuable in the study of such phenomena. In this case the experiment is sometimes called EXSY (exchange spectroscopy). In Figure 10 is shown part of the 2-D NOESY spectrum of the potential drug Lu 28–179 (CAS 147817-50-3)[19] as the hydrochloric acid salt. Small broad peaks in the reference spectrum, which may be mistaken for impurities, turn out to be a minor rotamer of the compound by showing strong exchange peaks with opposite sign to those from dipolar couplings. In the full spectrum exchange peaks between the water

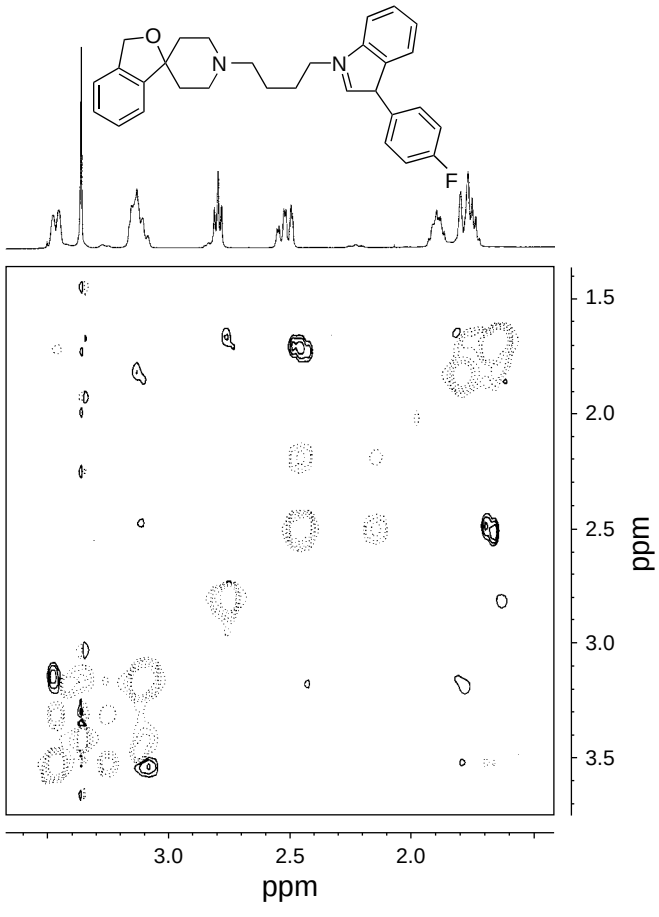

Figure 10 Aliphatic part of the 500-MHz NOESY spectrum of Lu 28–179 (40 mM in dimethyl-d_6 sulfoxide). The filled lines are NOE cross peaks and the dotted lines are exchange peaks and diagonal. 256 Increments of 1024 points and 16 scans were acquired. Delay d_1 was 2 s and the mixing time (d_2) was randomly varied about 0.6 s. The total experiment time was 5 h 40 min.

signal and the hydrogen chloride proton can also be seen.

4.4 The Rotating Frame Overhauser Enhancement Spectroscopy Experiment

In cases where the NOESY experiment fails, because the molecular tumbling rate gives rise to approximately zero NOEs, the ROESY (rotating frame Overhauser enhancement spectroscopy) experiment[20] may be an attractive alternative, since all sizes of molecules will behave as falling in the positive NOE regimen. ROESY is another spin-lock experiment like TOCSY (section 4.2), and basically the pulse sequence is the same as in Figure 7, but the power of the spin-lock field is lower, and is usually generated by two continuous-wave decoupling pulses with opposite phases. A good choice of spin-lock time in ROESY is 250 ms.

Because of the resemblance with the TOCSY experiment, artifacts arising from TOCSY peaks may be expected. These will have opposite phase to the ROESY peaks.

4.5 Incredible Natural Abundance Double Quantum Transfer Experiment

The 2-D INADEQUATE (Incredible Natural Abundance Double Quantum Transfer Experiment)[21] is at the borderline between homonuclear and heteronuclear experiments: it correlates nuclei of the same kind (^{13}C), but ^{1}H decoupling is crucial. It is certainly not a routine experiment, but is mentioned here because of the interest that has been shown in the literature in attempts to improve it. The experiment gives the ultimate structural information on organic molecules, since the whole carbon skeleton can be traced out. On the other hand, the sensitivity is extremely low. This is because the natural abundance of the NMR-active ^{13}C is only 1%. The likelihood that two ^{13}C atoms are neighbors in a molecule is therefore only 1 : 10 000, and the available signal is divided into doublets due to J_{CC} splitting in the original sequences. Furthermore, it must be considered that the gyromagnetic ratio of the ^{13}C nuclei is only 1/64 of that of the ^{1}H nuclei. The basic pulse sequence is shown in Figure 11(a) together with one of the latest improvements[22] in Figure 11(b). The 2-D INADEQUATE spectrum of menthol (CAS 1490-04-06) obtained with the pulse sequence in Figure 11(b) is shown in Figure 12. The correlation signals between the individual carbon atoms lie symmetrically about the diagonal on a horizontal line. Where the original sequence gives double quantum doublets, the newer sequence collects all signals in a single peak and therefore improves the S/N. A sample will give a reasonable 2-D INADEQUATE spectrum only if a 1-D single scan ^{13}C spectrum obtained with a 90° pulse has an S/N of at least 25 : 1. The delay d_2 must be adjusted to 1/4 J_{CC}, where J_{CC} is the ^{13}C–^{13}C one-bond coupling constant. Here a compromise must be made if a wide range of coupling constants is involved.

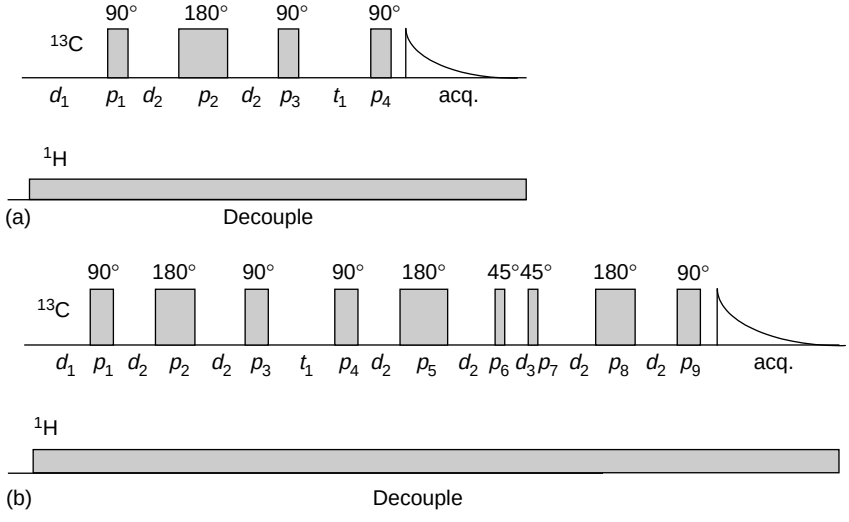

Figure 11 (a) Conventional 2-D INADEQUATE sequence. Delay d_2 is $1/4J_{CC}$. The first two pulses are varied in 45° steps. These phases are p_1 and p_3: 0, 135, 90, 45, 180, 315, 270, 225, 90, 225, 180, 135, 270, 45, 0, 315, 135, 270, 225, 180, 315, 90, 45, 0, 45, 180, 135, 90, 225, 0, 315, 270; p_2: 0, 135, 90, 45, 180, 315, 270, 225, 90, 225, 180, 135, 270, 45, 0, 315, 135, 270, 225, 180, 315, 90, 45, 0, 45, 180, 135, 90, 225, 0, 315, 270, 180, 315, 270, 225, 0, 135, 90, 45, 270, 45, 0, 315, 90, 225, 180, 135, 315, 90, 45, 0, 135, 270, 225, 180, 225, 0, 315, 270, 45, 180, 135, 90. The following are in 90° steps: p_4: $x,y,-x,-y$; acq.: $(x)_8$, $(-x)_8$, $(y)_8$, $(-y)_8$. (b) INADEQUATE sequence that collects all magnetization in one double quantum peak. The delay d_4 is short (4 µs). The phases for p_1-p_3 and the receiver are the same as for those in (a). For the other pulses the phases are: p_4: $-x,-y,x,y$; p_5, p_7, p_8: $x,y,-x,-y$; p_6: $-y,x,y,-x$; p_9: $y,-x,-y,x$.

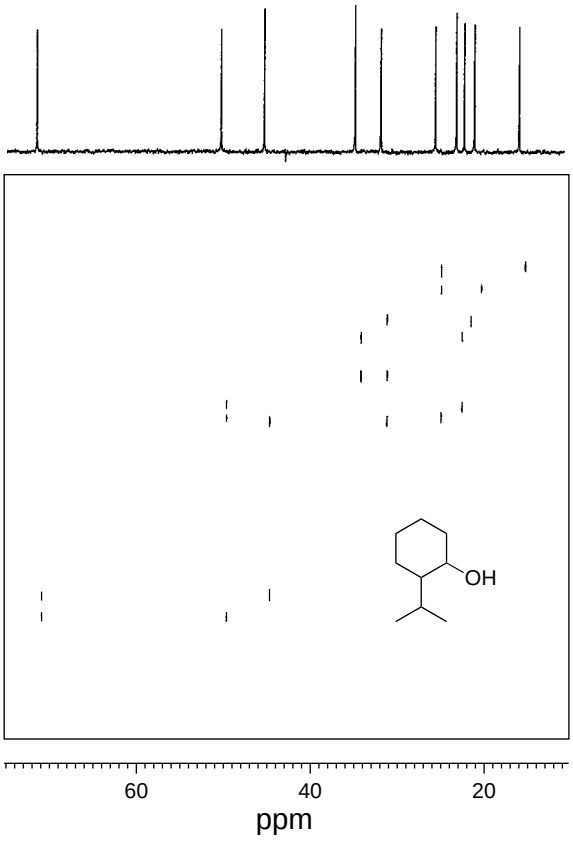

Figure 12 2-D INADEQUATE of 3 M menthol in chloroform-*d* obtained with the sequence in Figure 11(b) at 125 MHz (500-MHz instrument). 32 Increments of 2048 points with 256 scans per increment ware acquired. Delay d_1 was 3 s, d_2 was 7.14 ms and d_3 was 4 μs. Total experiment time was 7 h 31 min.

5 HOMONUCLEAR *J*-RESOLVED SPECTROSCOPY

With the JRES (*J*-resolved) experiment[13] (Figure 13), multiplet structures in crowded proton spectra may be revealed. Along the F_2 axis is the 1-D proton spectrum, while the F_1 axis show the splitting of the individual multiplets. If wanted, the spectrum may be tilted as a part of the processing, so that the multiplets appear to be perpendicular to the chemical shifts. Since the sweep width in the F_1 dimension only need be as large as the broadest multiplet, the experiment may be conducted with relatively few increments. For strongly coupled spin systems, cross peaks which are off the horizontal center line and between the wanted correlation peaks may be seen. The proximity and intensity of these peaks are dependent on the spin system order. Figure 14 shows the JRES spectrum of sertindole, which also shows some higher spin order cross peaks.

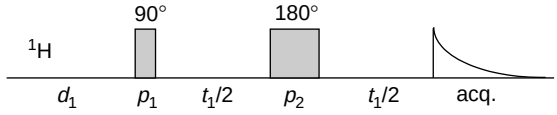

Figure 13 Pulse sequence for JRES. The phase cycle is: p_1: $(x)_4$, $(y)_4$, $(-x)_4$, $(-y)_4$; p_2: $x,-x,y,-y$, $y,-y,-x,x$, $y,-y,-x,x$, $-x,x,-y,y$; acq.: $x,x,-x,-x$, $y,y,-y,-y$. Minimum number of scans is four.

6 HETERONUCLEAR CORRELATION EXPERIMENTS

6.1 Heteronuclear Correlation with ^{13}C Detection

Direct ^{13}C detection was the original way of performing ^{13}C–^{1}H correlation experiments. These experiments may be performed on all NMR instruments capable of doing 2-D NMR and creating pulses with the ^{1}H-decoupler channel. Since the experiments rely on the acquisition of a series of ^{13}C spectra in the F_2 dimension, the sensitivity is limited. However, on instruments where the more sensitive indirect detection method requires rewiring of the spectrometer, direct detection can still be the method of choice for routine applications. The spectra of the two experiments mentioned are in magnitude mode.

One-bond correlation is obtained with the HETCOR experiment,[23] for which the pulse program is given in Figure 15. As an example, the HETCOR spectrum of setindole is shown in Figure 16. The delays d_2 and d_3 are set to $1/2J_{CH}$ and $1/3J_{CH}$, respectively.

In order to obtain two- or three-bond ^{13}C–^{1}H correlation, the delays d_2 and d_3 can be adjusted to correspond to the long-range coupling constant, but in this case the sensitivity is low owing to relaxation during the lengthy sequence. The method of choice is instead the COLOC (correlation through long-range couplings) experiment.[24] From the pulse sequence in Figure 17 it can be seen that the duration is constant with increasing t_1 values because p_2 and p_4 are moved within the sequence with increasing t_1. The evolution of long-range couplings takes place during d_3. Still, the sensitivity is considerably lower than for the one-bond experiment, as can be seen from the COLOC spectrum of sertindole in Figure 18.

6.2 Heteronuclear Experiments with ^{1}H Detection

Indirect or inverse detection has had a major impact on 2-D NMR, especially in biomolecular applications, where the sensitivity of the direct methods is much too low. The gain in sensitivity in indirect detection, which

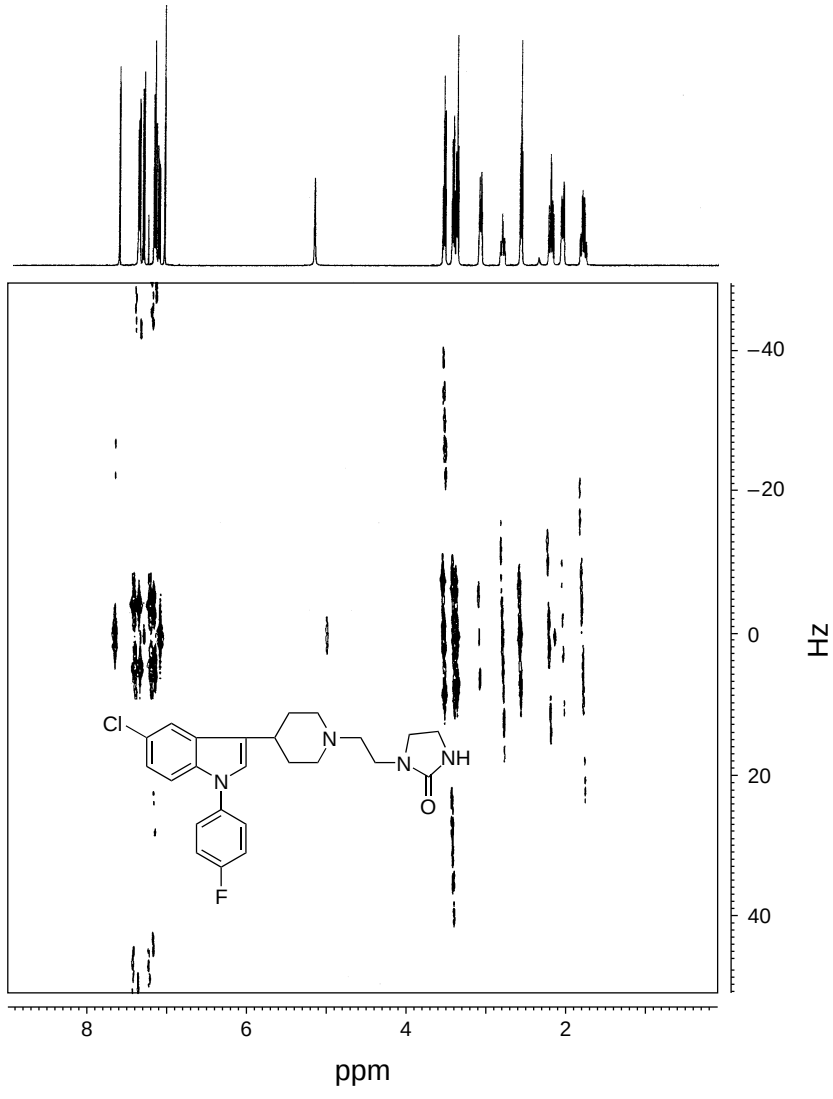

Figure 14 JRES spectrum of 30 mM sertindole in chloroform-*d* at 500 MHz with 64 increments of 1024 points and four scans per increment. Delay d_1 was 1.5 s and the sweep width in F_1 was 100 Hz. Total experiment time was 10 min.

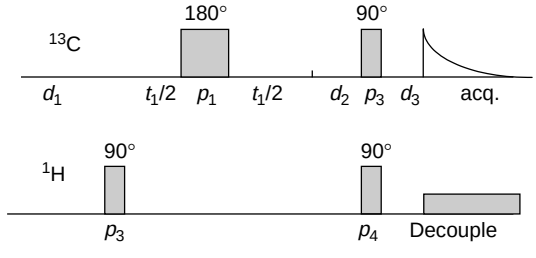

Figure 15 Pulse sequence for HETCOR. Delay d_2 must be adjusted to $1/2J_{CH}$ and d_3 to $1/3J_{CH}$. The phase cycle is: p_1: $(x)_4$, $(-x)_4$; p_2: $(x)_8$, $(y)_8$, $(-x)_8$, $(-y)_8$; p_3: x; p_4: $x,-x,y,-y$; acq.: $(x,-x,y,-y)_2$, $(y,-y,-x,x)_2$, $(-x,x,-y,y)_2$; $(-y,y,x,-x)_2$. Minimum number of scans is four.

is a factor of 16,[25] comes from the difference in sensitivity for ^{1}H and ^{13}C, where the higher gyromagnetic ratio for ^{1}H makes this the nucleus of choice for detection. If an X-nucleus decoupler is available, composite pulse decoupling, e.g. GARP (globally optimized alternating phase rectangular pulse),[26] can be done on carbon during acquisition. The most popular experiments for one-bond correlation are variants of HMQC (heteronuclear multiple quantum coherence)[25] and HSQC (heteronuclear single quantum coherence).[27,28] In general, HSQC has the advantage that ^{1}H–^{1}H couplings do not evolve in F_1, thus giving better resolved spectra. The HSQC pulse sequence is given in Figure 19 and a phase-sensitive PFG version is shown in Figure 20. The delay d_2 is adjusted to 1/4 J_{CH}. In order to acquire this experiment in the phase-sensitive mode, the so-called echo/antiecho method must be used instead of TPPI. If the processing software does not have the needed

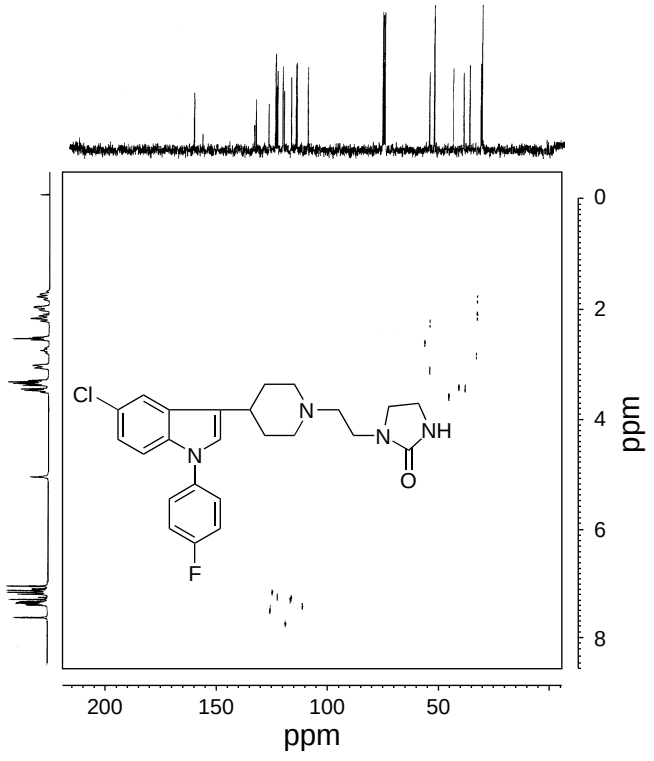

Figure 16 HETCOR spectrum of 0.15 M sertindole in chloroform-*d* on a 250-MHz instrument with 256 increments of 2048 points and 16 scans per increment. Delay d_1 was 1 s, d_2 was 3.4 ms and d_3 was 2.3 ms. Total experiment time was 1 h 24 min.

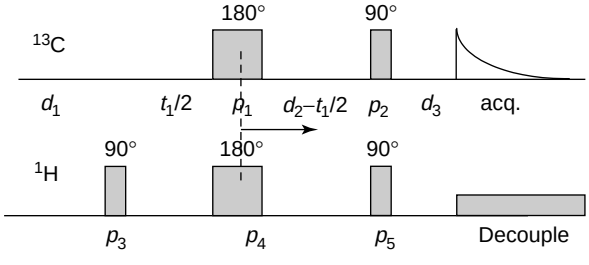

Figure 17 The COLOC sequence. Delay d_2 must be larger than the product of the time increment and number of increments. Delay d_3 is defined as in Figure 15, but for long-range couplings. The arrow indicates that the two 180° pulses are moved forward with increasing t_1, while the whole sequence time remains constant. The phase cycle is: p_1 and p_4: x; p_2: $x,y,-x,-y$, $y,-x,-y,x$, $-x,-y,x,y$, $-y,x,y,-y$, $y,-x,-y,x$, $-x,-y,x,y$, $-y,x,y,-y$, $x,y,-x,-y$, $-x,-y,x,y$, $-y,x,y,-x$, $x,y,-x,-y$, $y,-x,-y,x$, $-y,x,y,-x$, $x,y,-x,-y$, $y,-x,-y,x$, $-x,-y,x,y$; p_3: $(x)_{16}$, $(y)_{16}$, $(-x)_{16}$, $(-y)_{16}$; p_5: x,y,x,y, x,y,x,y, $-x,-y,-x,-y$, $-x,-y,-x,-y$; acq.: $x,x,-x,-x$, $y,y,-y,-y$. Minimum number of scans is 16.

processing feature, the experiment may be used for obtaining a magnitude spectrum. Figure 21 shows the 2-D spectrum of sertindole acquired with the sequence in Figure 20.

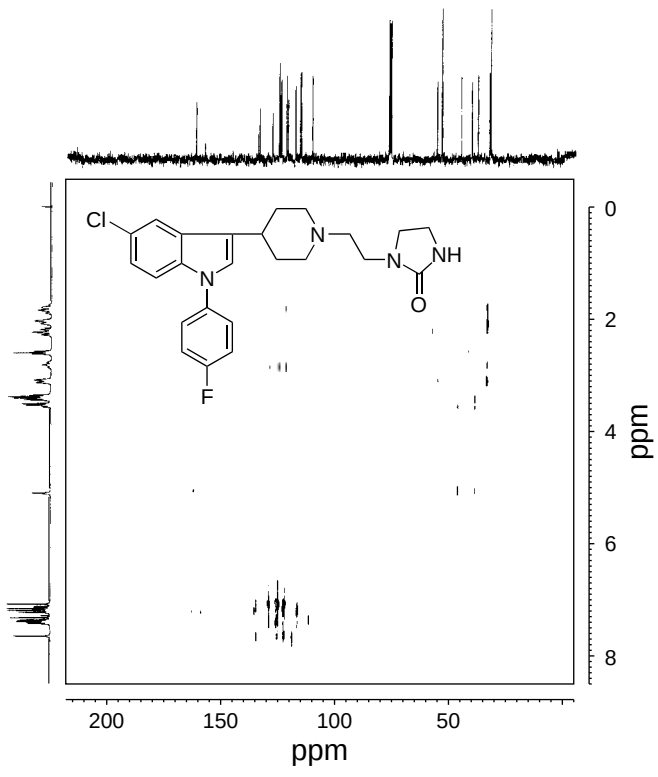

Figure 18 COLOC spectrum of 0.15 M sertindole in chloroform-*d* on a 250-MHz instrument with 128 increments of 2048 points and 256 scans per increment. Delay d_1 was 1 s, d_2 was 80 ms and d_3 was 40 ms. Total experiment time was 12 h.

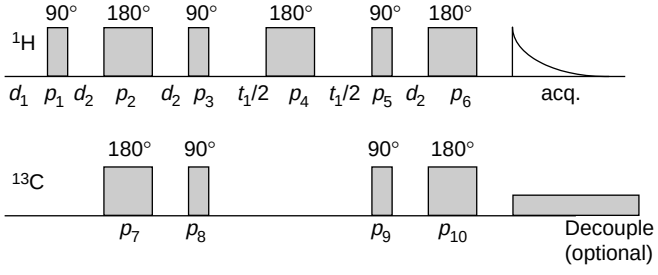

Figure 19 Pulse sequence for phase-sensitive HSQC. The delay d_2 is $1/4J_{CH}$. The phase cycle is: p_1, p_2, and p_6: x; p_3 and p_5: y; p_4: $x,x,-x,-x$; p_7 and p_{10}: $(x)_8$, $(-x)_8$; p_8: $x,-x$: p_9: $(x)_4$, $(-x)_4$; acq.: $x,-x,-x,x$, $-x,x,-x,x$. The phase of p_8 is incremented for TPPI. Minimum number of scans is four.

For long-range ^{1}H–^{13}C correlation, the experiment of choice is HMBC (heteronuclear multiple bond correlation),[29] which also exists in a PFG version. The pulse sequence with optional PFGs is shown in Figure 22 and the HMBC spectrum of sertindole is shown in Figure 23. Note that many more correlation peaks are seen here than in the spectrum in Figure 18 obtained with the COLOC sequence. Since one-bond

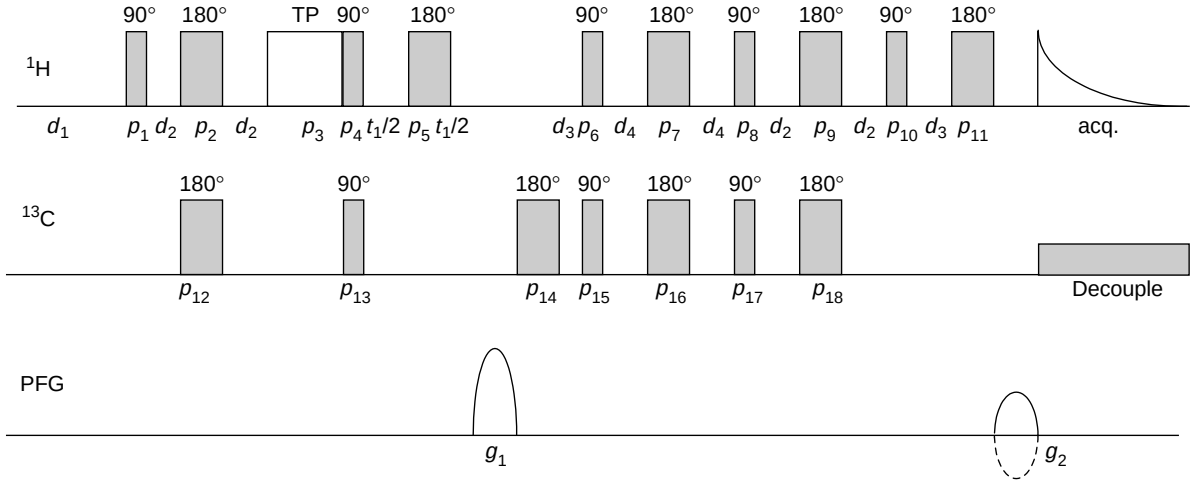

Figure 20 Pulse sequence for phase-sensitive PFG/HSQC. Delay d_2 is $1/4J_{CH}$, d_4 is $1/8J_{CH}$, d_3 is the effective gradient length, and p_3 is a high power trim pulse. The phase cycle is: p_1, p_2, p_3, p_6, p_7, p_9, p_{10}, p_{11}, p_{12}, p_{16} and p_{18}: x; p_4 and p_8: y; p_5, p_{14} and p_{15}: $x,x,-x,-x$; p_{13}: $x,-x$; p_{17}: $y,y,-y,-y$; acq.: $x,-x,-x,x$. The gradient ratio is 80 : 20, and the sign of g_2 is alternated for echo/antiecho detection.

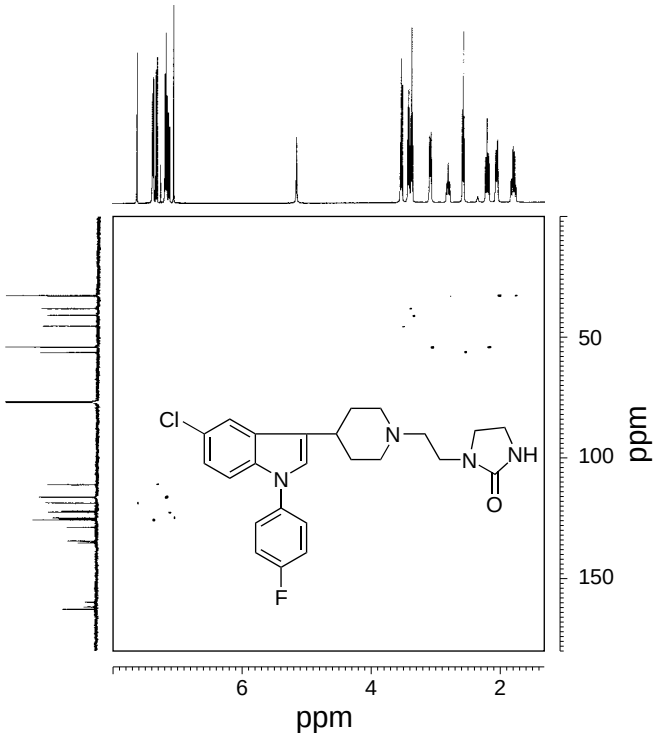

Figure 21 PFG/HSQC spectrum of 30 mM (10 mg) sertindole in chloroform-*d* at 500 MHz with 128 increments of 1024 points and two scans per increment. Delay d_1 was 1 s, d_2 was 1.9 ms, d_3 was 2.1 ms, d_4 was 1 ms and the high-power trim pulse was 2 ms. Total experiment time was 6 min.

couplings also evolve to some extent in the HMBC experiment, it is advantageous to avoid carbon decoupling. In this way, single and multiple bonds can easily be distinguished.

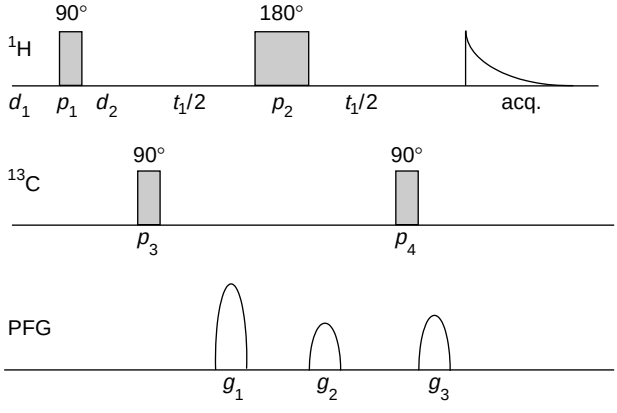

Figure 22 The pulse sequence for HMBC with optional PFGs. Delay d_2 ($1/2J_{CH}$) is optimized for long-range couplings. The phase cycle without PFGs is: p_1 and p_2: x; p_3: $x,y,-x,-y$; p_4: $(x)_4$, $(-x)_4$; acq.: $x,-y,-x,y$, $-x,y,x,-y$. Minimum number of scans is 16. With PFGs the cycle is: p_1: x; p_2: $x,x,-x,-x$; p_3: $x,-x$; p_4: $(x)_4$, $(-x)_4$; acq.: $x,-x,x,-x$, $-x,x,-x,x$. The gradient ratios are 50 : 30 : 40. Minimum number of scans is two.

7 PSEUDO-THREE-DIMENSIONAL EXPERIMENTS

The term pseudo-three-dimensional (3-D) covers a range of pulse sequences developed for 3-D work which have been modified. The main feature of 3-D experiments is the combination of two 2-D experiments in which the two evolution periods are incremented independently. If one of the evolution periods is omitted, a 2-D spectrum will be the result. An example is phase-sensitive 2-D HMQC/TOCSY[30] shown in Figure 24, which may easily be performed. The main advantage of this experiment is

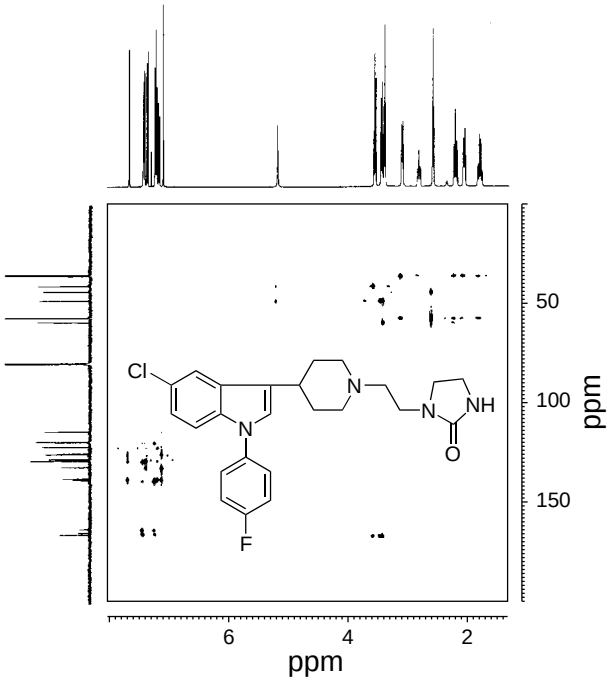

Figure 23 The PFG/HMBC spectrum of 30 mM sertindole in chloroform-*d* at 500 MHz with 128 increments of 1024 points and two scans per increment. Delay d_1 was 0.9 s and d_2 was 0.05 s. Total experiment time was 6 min.

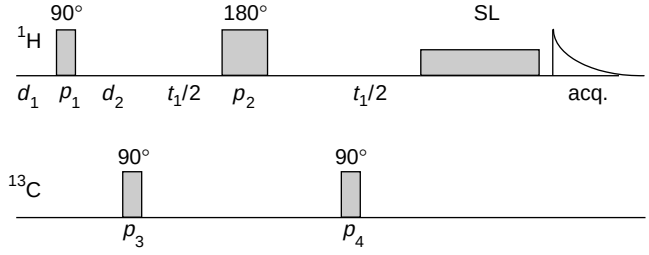

Figure 24 The pulse sequence for HMQC/TOCSY. Delay d_2 is $(1/2J_{CH})$ and the spin lock is adjusted for the wanted correlation type. For TPPI p_3 is incremented. Minimum number of scans is four.

that ^{1}H–^{1}H correlation in very crowded spectral areas may be spread out in the ^{13}C dimension, so that all cross peaks in a given row with one ^{13}C atom correspond to protons in a coupling network. The spectrum in Figure 25 was acquired without carbon decoupling. An important feature of this experiment type, when acquired without carbon decoupling, is that a ^{1}H–^{1}H correlation between protons with identical chemical shifts may be observed. For example, in Figure 25, a TOCSY correlation between symmetrically positioned protons in the piperidine ring of sertindole may be observed.

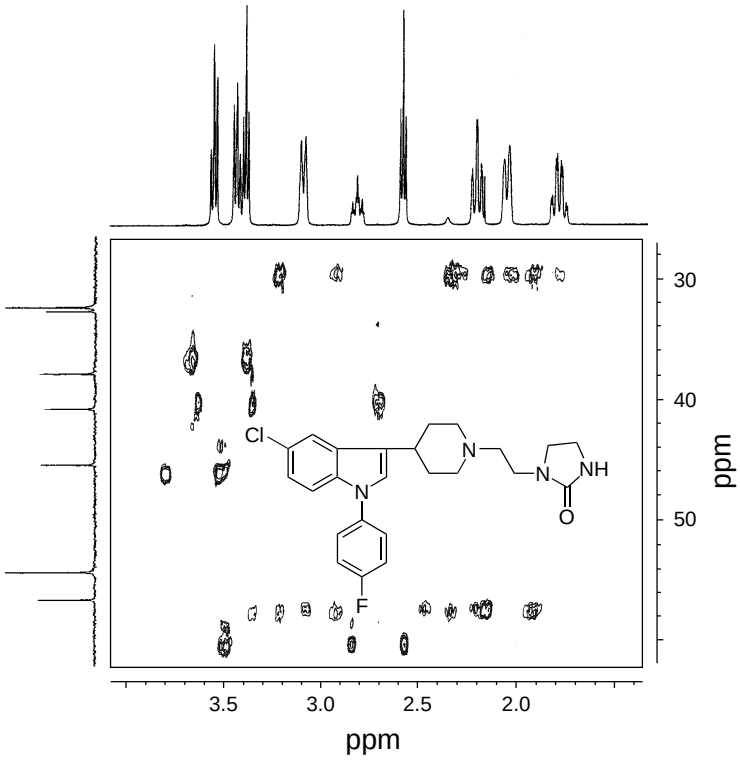

Figure 25 The phase-sensitive HMQC/TOCSY of 30 mM sertindole in chloroform-*d* at 500 MHz with 128 increments of 1024 points and eight scans per increment. Delay d_1 was 1 s, d_2 was 3.7 ms and the spin-lock time including two 2.5-ms trim pulses was 70 ms. Total experiment time was 23 min.

Other variants including gradient versions combine an indirect $^{1}H-^{13}C$ one-bond correlation with NOESY,[31] COSY,[32] and ROESY.[33]

8 TWO-DIMENSIONAL NUCLEAR MAGNETIC RESONANCE ON SPIN-$\frac{1}{2}$ NUCLEI OTHER THAN ^{1}H AND ^{13}C

In principle, all 2-D NMR experiments may involve any spin-$\frac{1}{2}$ nucleus if the abundance in the sample is sufficient and a probe head tunable for that nucleus is available. For small organic molecules, the most important are ^{15}N, ^{19}F, ^{29}Si and ^{31}P. From sensitivity considerations, both homonuclear and HETCOR are possible with ^{19}F, ^{29}Si and ^{31}P, whereas the low natural abundance and low gyromagnetic ratio of ^{15}N allow only for indirect-detected HETCOR. An example of this is the $^{1}H-^{15}N$ HMBC spectrum of sertindole in Figure 26. The spectrum was acquired with the pulse sequence in Figure 22, with a gradient ratio of 70 : 30 : 50.

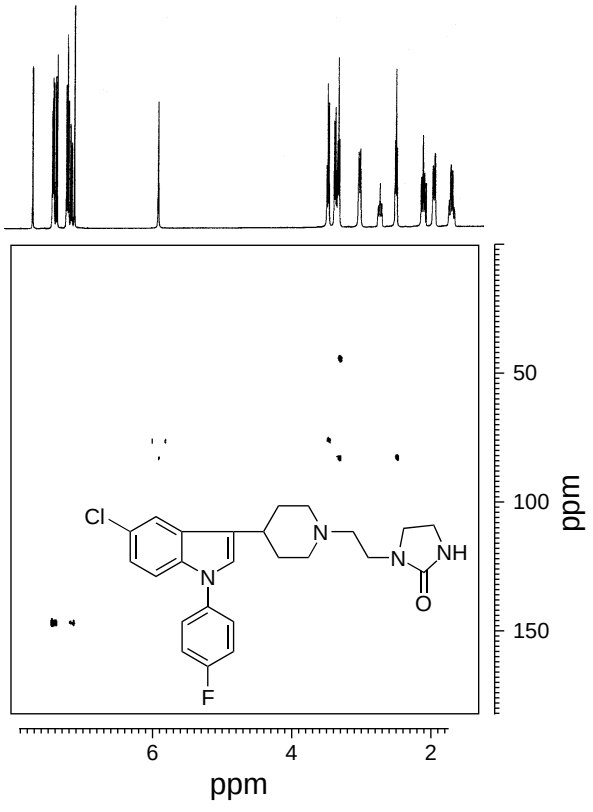

Figure 26 The $^{1}H-^{15}N$ PFG/HMBC of 150 mM sertindole in chloroform-*d* at 500 MHz with 128 increments of 1024 points and 128 scans per increment. Delay d_1 was 1 s, d_2 was 0.2 s. The gradient ratio was in this case 70 : 30 : 50. Total experiment time was 6 h 33 min.

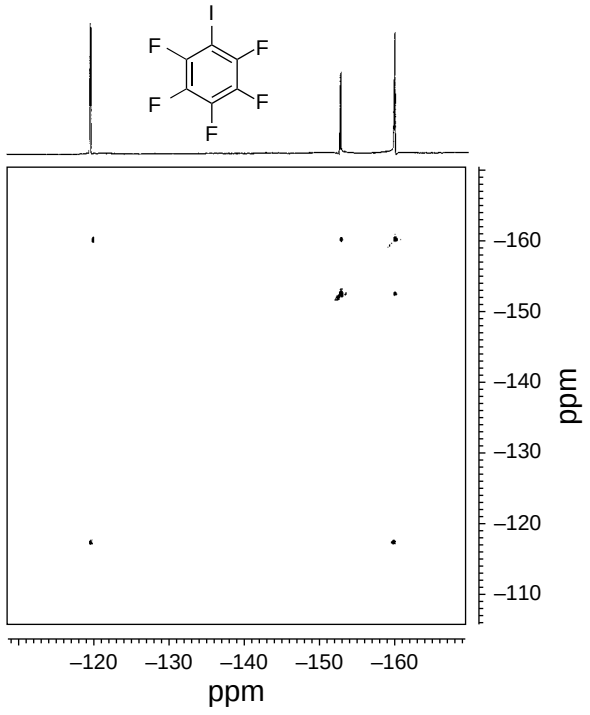

Figure 27 The $^{19}F-^{19}F$ COSY spectrum of 0.7 M iodopentafluorobenzene in chloroform-*d* obtained without PFGs with the sequence in Figure 3 at 235.3 MHz on a 250-MHz system. 128 Increments with 1024 points and four scans per increment. Delay d_1 was 2 s. Total experiment time was 18 min.

Since the Larmor frequency of ^{19}F is close to that of ^{1}H, the ^{1}H coil can in many cases be tuned and matched for ^{19}F with a reasonable result. This is how the spectra in Figures 27 and 28 of iodopentafluorobenzene were obtained. The $^{19}F-^{19}F$ COSY spectrum in Figure 27 was acquired with the pulse sequence in Figure 4 and the $^{19}F-^{13}C$ HMBC spectrum was acquired with the sequence in Figure 22, both without PFGs.

9 TWO-DIMENSIONAL NUCLEAR MAGNETIC RESONANCE ON RESIN-BOUND COMPOUNDS

The new era of combinatorial chemistry based on solid-phase synthesis has offered a new challenge to analytical methods, including NMR spectroscopy. Obtaining proton observed spectra of a compound bound to a resin swollen in an appropriate deuterated solvent requires magic angle spinning (MAS) in order to obtain a reasonable line width and resolution. NMR instrument vendors have introduced special probe heads with a deuterium lock channel suitable for this purpose, of which the most sophisticated have two channels and are equipped with a one-axis gradient coil. In principle, all the experiments

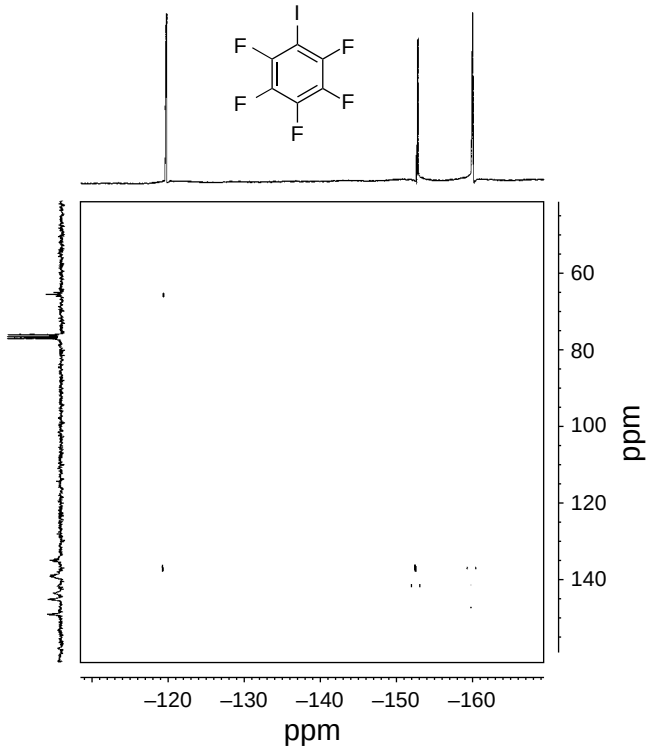

Figure 28 The $^{19}F-^{13}C$ HMBC spectrum of 0.7 M iodopentafluorobenzene in chloroform-*d* obtained without PFGs with the sequence in Figure 22 on a 250-MHz system. 64 Increments with 1024 points and 32 scans per increment. Delay d_1 was 2 s, d_2 was 25 ms. Total experiment time was 18 min.

described in the earlier sections may be conducted on resin-bound molecules, but there are some limitations. Some line broadening will occur even with fast spinning, and this is related to transverse relaxation processes, which are dependent on the mobility of the attached molecules in the swollen phase.

Thus, samples giving broad lines in the 1-D proton spectrum will give relatively poor S/N in the 2-D spectra if the pulse sequence is lengthy or if many t_1 increments are acquired. On the other hand, the time consumed by the 2-D pulse sequence will serve as an efficient filter for signals from the polymer backbone, which will have the lowest mobility.

Since the resolution observed in the 1-D proton spectra of this kind of sample is often so low that coupling patterns cannot be identified, the structural information that is available from 2-D spectra is of major importance. The use of the F_2 projection of the JRES experiment[34] and an E.COSY variant[35] has been described for revealing coupling constants in the F_1 dimension for samples with broadened signals. If, however, the size of the coupling constants is of less importance, more standard-like experiments may be used for obtaining structural information. Even though the effective sample volume may be as low as 25 μL, a resin-bound sample with a loading of 0.3 mmol g^{-1} giving broad lines in the 1-D spectrum may give gradient-selected COSY, HSQC and HMBC spectra in total acquisition times of less than 6 min.[36]

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
COLOC	Correlation Through Long-range Couplings
COSY	Correlation Spectroscopy
COSYDQ	Double Quantum Filtered Correlation Spectroscopy
E.COSY	Exclusive Correlation Spectroscopy
EXSY	Exchange Spectroscopy
FID	Free Induction Decay
GARP	Globally Optimized Alternating Phase Rectangular Pulse
HETCOR	Heteronuclear Correlation
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Coherence
HOHAHA	Homonuclear Hartmann–Hahn
HSQC	Heteronuclear Single Quantum Coherence
INADEQUATE	Incredible Natural Abundance Double Quantum Transfer Experiment
JRES	*J*-Resolved
MAS	Magic Angle Spinning
MLEV	Malcom Levitt
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Enhancement
NOESY	Nuclear Overhauser Enhancement Spectroscopy
PFG	Pulsed Field Gradient
RF	Radiofrequency
ROESY	Rotating Frame Overhauser Enhancement Spectroscopy
S/N	Signal-to-noise Ratio
TOCSY	Total Correlation Spectroscopy
TPPI	Time Proportional Phase Increment
WATERGATE	Water Suppression by Gradient Tailored Excitation
WET	Water Suppression Enhanced Through T_1 Effects
1-D	One-dimensional

2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Nuclear Magnetic Resonance of Biomolecules

Carbohydrate Analysis **(Volume 1)**
Glycoprotein Analysis: Using Nuclear Magnetic Resonance

Nucleic Acids Structure and Mapping **(Volume 6)**
Nuclear Magnetic Resonance and Nucleic Acid Structures

Pharmaceuticals and Drugs **(Volume 8)**
Nuclear Magnetic Resonance Spectroscopy in Pharmaceutical Analysis

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy: Introduction • Carbon-13 Nuclear Magnetic Resonance Spectroscopy • Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Nuclear Magnetic Resonance Instrumentation • Relaxation in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General • Solution Nuclear Magnetic Resonance: Spin-1/2 Nuclei Other than Carbon and Proton • Two-, Three- and Four-dimensional Nuclear Magnetic Resonance of Biomolecules

REFERENCES

1. J. Jeener, Unpublished Results, Ampere International Summer School, Basko Polje, Yugoslavia, 1971.
2. O.W. Sørensen, G.W. Eich, M.H. Levitt, G. Bodenhausen, R.R. Ernst, 'Product Operator Formalism for the Description of NMR Pulse Experiments', *Prog. Nucl. Magn. Reson. Spectrosc.*, **16**, 163–192 (1983).
3. R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One or Two Dimensions*, Clarendon Press, Oxford, 1987.
4. W.R. Croasmun, R.M.K. Carlson, *Two-dimensional NMR Spectroscopy, Applications for Chemists and Biochemists*, VCH, New York, 1994.
5. J.W. Hennel, J. Klinowski, *Fundamentals of Nuclear Magnetic Resonance*, Longman Scientific and Technical, Harlow, 1993.
6. S. Braun, H.-O. Kalinowski, S. Berger, *150 and More Basic NMR Experiments*, Wiley-VCH, Weinheim, 1998.
7. D.J. States, R.A. Haberkorn, D.J. Reuben, 'A Two-dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in Four Quadrants', *J. Magn. Reson.*, **48**, 286–292 (1982).
8. G. Drobny, A. Pines, S. Sinton, D. Weitekamp, D. Wemmer, 'Fourier Transform Multiple Quantum Nuclear Magnetic Resonance', *Faraday Symp. Chem. Soc.*, **13**, 49–55 (1979).
9. Y. Zeng, J. Tang, C.A. Bush, J.R. Norris, 'Enhanced Spectral Resolution in 2-D NMR Signal Analysis Using Linear Prediction Extrapolation and Apodization', *J. Magn. Reson.*, **83**, 473–483 (1989).
10. M. Piotto, V. Saudek, V. Sklenar, 'Gradient-tailored Excitation for Single-quantum NMR Spectroscopy of Aqueous Solutions', *J. Biol. NMR*, **2**, 661–665 (1992).
11. T.-L. Hwang, A.J. Shaka, 'Water Suppression that Works. Excitation Sculpting Using Arbitrary Waveforms and Pulsed Field Gradients', *J. Magn. Reson., Ser. A*, **112**, 275–279 (1995).
12. S.H. Smallcombe, S.L. Patt, P. Keifer, 'WET Solvent Suppression and Its Application to LC NMR and High-resolution NMR Spectroscopy', *J. Magn. Reson., Ser. A*, **117**, 295–303 (1995).
13. W.P. Aue, E. Bartholdi, R.R. Ernst, 'Two-dimensional Spectroscopy. Application to Nuclear Magnetic Resonance', *J. Chem. Phys.*, **64**, 2229–2246 (1976).
14. J. Perregaard, J. Arnt, K.P. Bøgesø, J. Hyttel, C. Sánchez, 'Noncataleptogenic, Centrally Acting Dopamine D-2 and Serotonin 5-HT_2 Antagonists Within a Series of 3-Substituted 1-(4-Fluorophenyl)-1*H*-indoles', *J. Med. Chem.*, **35**, 1092–1101 (1992).
15. U. Piantini, O.W. Sørensen, R.R. Ernst, 'Multiple Quantum Filters for Elucidating NMR Coupling Networks', *J. Am. Chem. Soc.*, **104**, 6800–6801 (1982).
16. R.E. Hurd, 'Gradient-enhanced Spectroscopy', *J. Magn. Reson.*, **87**, 422–428 (1990).
17. C. Griesinger, O.W. Sorensen, R.R. Ernst, 'Practical Aspects of the E.COSY Technique. Measurement of Scalar Spin–Spin Coupling Constants in Peptides', *J. Magn. Reson.*, **75**, 474–492 (1987).
18. A. Bax, D.G. Davis, 'MLEV-17-Based Two-dimensional Homonuclear Magnetization Transfer Spectroscopy', *J. Magn Reson.*, **65**, 355–360 (1985).
19. J. Perregaard, E.K. Moltzen, E. Meir, C. Sánchez, 'σ Ligands with Subnanomolar Affinity and Preference for the σ_2 Binding Site. 1. 3-(*w*-Aminoalkyl)-1*H*-indoles', *J. Med. Chem.*, **38**, 1998–2008 (1995).
20. A. Bax, D.G. Davis, 'Practical Aspects of Two-dimensional Transverse NOE Spectroscopy', *J. Magn. Reson.*, **63**, 207–213 (1985).

21. A. Bax, R. Freeman, T.A. Frenkiel, 'An NMR Technique for Tracing Out the Carbon Skeleton of an Organic Molecule', *J. Am. Chem. Soc.*, **103**, 2102–2104 (1981).
22. N.C. Nielsen, H. Thøgersen, O.W. Sørensen, 'A Systematic Strategy for Design of Optimum Coherent Experiments Applied to Efficient Interconversion of Double- and Single-quantum Coherences in Nuclear Magnetic Resonance', *J. Chem. Phys.*, **105**, 3962–3968 (1996).
23. A. Bax, G.A. Morris, 'An Improved Method for Heteronuclear Chemical Shift Correlation by Two-dimensional NMR', *J. Magn. Reson.*, **42**, 501–505 (1981).
24. H. Kessler, C. Griesinger, J. Zarbock, H.R. Loosli, 'Assignment of Carbonyl Carbons and Sequence Analyses in Peptides by Heteronuclear Shift Correlation via Small Coupling Constants with Broadband Decoupling in t_1 (COLOC)', *J. Magn. Reson.*, **57**, 331–336 (1984).
25. A. Bax, S. Subramanian, 'Sensitivity-enhanced Two-dimensional Heteronuclear Shift Correlation NMR Spectroscopy', *J. Magn. Reson.*, **67**, 565–569 (1986).
26. A.J. Shaka, P.B. Barker, R. Freeman, 'Computer-optimized Decoupling Scheme for Wideband Applications and Low-level Operation', *J. Magn. Reson.*, **64**, 547–552 (1985).
27. G. Bodenhausen, D.J. Ruben, 'Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy', *Chem. Phys. Lett.*, **69**, 185–189 (1980).
28. J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S.J. Glaser, O.W. Sørensen, C. Griesinger, 'A General Enhancement Scheme in Heteronuclear Multidimensional NMR Employing Pulsed Field Gradients', *J. Biomol. NMR*, **4**, 301–306 (1994).
29. M.F. Summers, L.G. Marzilli, A. Bax, 'Complete Proton and Carbon-13 Assignments of Coenzyme B12 through Use of New Two-dimensional NMR Experiments', *J. Am. Chem. Soc.*, **108**, 4285–4294 (1986).
30. D.G. Davis, 'Enhanced Correlations in ^{1}H-Detected Heteronuclear Two-dimensional Spectroscopy via Spin-locked ^{1}H Magnetization Transfer', *J. Magn. Reson.*, **84**, 417–424 (1989).
31. L.W. Castle, M.D. Johnston, Jr, C.L. Camoutsis, R.N. Castle, 'The HMQC–NOESY Experiment Without Decoupling During Acquisition for Detection of NOE between Pseudo-equivalent Protons', *J. Heterocycl. Chem.*, **29**, 1869–1871 (1992).
32. P.H. Bolton, 'Heteronuclear Relay Transfer Spectroscopy with Proton Detection', *J. Magn Reson.*, **62**, 143–146 (1985).
33. J. Kawabata, E. Fukushi, J. Mizutani, '2-D ^{13}C-coupled HMQC–ROESY: a Probe for NOEs between Equivalent Protons', *J. Am. Chem. Soc.*, **114**, 1115–1117 (1992).
34. M.J. Shapiro, J. Chin, R.E. Marti, M.A. Jarosinski, 'Enhanced Resolution in MAS NMR for Combinatorial Chemistry', *Tetrahedron Lett.*, **38**, 1333–1336 (1997).
35. A. Meissner, P. Bloch, E. Humpfer, M. Spraul, O.W. Sørensen, 'Reduction of Inhomogeneous Line Broadening in Two-dimensional High Resolution MAS NMR Spectra of Molecules Attached to Swollen Resins in Solid-phase Synthesis', *J. Am. Chem. Soc.*, **119**, 1787–1788 (1997).
36. T. Ruhland, K. Andersen, H. Pedersen, 'Selenium-linking Strategy for Traceless Solid-phase Synthesis: Direct Loading, Aliphatic C–H Bond Formation Upon Cleavage and Reaction Monitoring by Gradient MAS NMR Spectroscopy', *J. Org. Chem.*, **63**, 9204–9211 (1998).

Zeeman Interaction in Nuclear Magnetic Resonance

Cecil Dybowski
University of Delaware, Newark, USA

1 Introduction 1
2 Nuclear Spins 1
3 Magnetic Moments 2
4 The Zeeman Interaction, Historically 3
5 The Energies of Nuclear Spins in a Magnetic Field 4
6 Vector and Tensor Descriptions in Nuclear Magnetic Resonance Spectroscopy 4
7 Coupling of Spins Modeled as Zeeman Interactions 5
7.1 Shielding 5
7.2 The Dipole–Dipole Interaction 7
7.3 The J Coupling Between Spins 7
8 Conclusions 8
Abbreviations and Acronyms 8
Related Articles 8
References 8

The Zeeman interaction is the physical phenomenon underlying the coupling of magnetic moments to magnetic fields. Its effects are detected in the features of spectra of samples taken while the sample is immersed in a magnetic field. Specifically, the features of nuclear magnetic resonance (NMR) spectra depend strongly on the coupling of nuclear spins to various sources of magnetic field.

1 INTRODUCTION

Magnetic resonance is a staple of analysis in the modern chemistry laboratory, particularly for the analysis of organic materials. More generally, it is applied to a range of chemical and physical systems, both organic and inorganic, having nuclei with magnetic moments.[1] It has become a major tool in medical diagnoses through the development of imaging techniques that provide three-dimensional pictures of tissues. The application to the study of a wide variety of materials is a testament to the versatility and continued development of NMR techniques.

These NMR techniques rely on the singular physical fact that, in a magnetic field, the intrinsic nuclear angular-momentum (spin) states lie at different energies because of the presence of the magnetic field – the Zeeman interaction – first discovered by Pieter Zeeman in a different situation and explained by Lorentz (see Condon and Shortely[2]). The concept is widely applied to understand interactions with magnetic fields, one of which is the NMR experiment.

2 NUCLEAR SPINS

Nuclei are small islands carrying specific amounts of positive charge and mass, as exemplified in Table 1. They are quantum objects. In NMR studies, one generally focuses on the properties of the nucleus in its ground state, as addressed through measurement of the magnetic properties related to the intrinsic angular momentum.

Angular momentum, as a classical property, is a vector describing circularity of motion.[3] Consider a single particle of mass, m, in motion about a point. Its position relative to the point being $\boldsymbol{r}$ and its linear momentum being $\boldsymbol{p}$, the instantaneous orbital angular momentum, $\boldsymbol{l}$, is given by the simple formula given in Equation (1):

$$\boldsymbol{l} = \boldsymbol{r} \times \boldsymbol{p} = m\boldsymbol{r} \times \boldsymbol{v} \tag{1}$$

where the second equality indicates that angular momentum depends on mass and instantaneous velocity, $\boldsymbol{v}$. For a system of many particles, the total angular momentum is a vector sum of contributions from all particles. In many situations, one focuses on motion about some axis passing through the center of mass of the system.

The manner in which angular momentum changes with time is determined by the torque, τ, to which the system is subject, and is given by the formula in Equation (2):

$$\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}\boldsymbol{t}} = \tau \tag{2}$$

For a system subject to no torque, the angular momentum is time independent, with its magnitude and orientation fixed. The description of a vector such as the angular momentum requires one to specify the projections of the vector along the three axes of a Cartesian coordinate system, such as shown in Figure 1, or the equivalent.

One often thinks of angular momentum in terms of the motion of a particle tethered in some way, such as an electron in an orbit. However, there are many systems with angular momentum whose equations of motion are of the same form as those of the tethered particle. In

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Properties of representative nuclei

Nuclear isotope	Mass (g mol^{-1})[a]	Charge (in units of proton charge)	Spin quantum number, I	Magnetogyric ratio $\times 10^7$ (radians s^{-1} T^{-1})[b]
^{1}H	1.0078250	1	1/2	26.7519
^{2}H	2.0141018	2	1	4.1066
^{7}Li	7.0160030	3	3/2	10.3975
^{13}C	13.0033548	6	1/2	6.7283
^{15}N	15.00010897	7	1/2	2.712
^{19}F	18.99840325	9	1/2	25.181
^{23}Na	22.9897697	11	3/2	7.08013
^{27}Al	26.9815413	13	5/2	6.9760
^{29}Si	29.9737717	14	1/2	5.3188

[a] International Union of Pure and Applied Chemistry, *Quantities, Units and Symbols in Physical Chemistry*, Blackwell Scientific Publications, Oxford, 1998.
[b] R.K. Harris, *Nuclear Magnetic Resonance Spectroscopy*, Pitman, London, 1983.

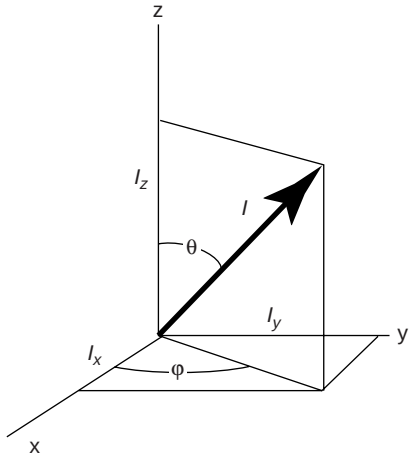

Figure 1 Representation of the angular momentum vector in a Cartesian coordinate system.

magnetic resonance, we focus particularly on the intrinsic angular momentum of nuclei and electrons.[4]

The particles are so small that one must consider them as quantum objects, for which one seeks the possible stationary states of a system, the equivalent of determining the time-independent properties in a classical system. For a system of particles in motion, there are states having constant magnitude of the angular momentum.

Unlike classical systems, not all variables of a quantum system may be simultaneously known with infinite precision. In particular, one is allowed by the uncertainty principle to measure only the magnitude and one component of the angular momentum precisely, with the result that there exists some uncertainty about the angular placement of the angular momentum vector. For a nucleus, the square of the magnitude of the intrinsic angular momentum, $\boldsymbol{I}$, is a fundamental property of the state of the nucleus, determined by the spin quantum number, I.[5] Its value is $I(I+1)\hbar^2$, where $\hbar$ is a fundamental constant (Planck's constant divided by 2π) equal to $1.0545726 \times 10^{-34}$ J s. I will always be a non-negative half or whole integer, as indicated in Table 1 for representative nuclear isotopes. In the quantum mechanical description of this kind of system, the state is also characterized by the projection along any one of the axes, usually chosen to be the z-axis of the coordinate system in which the system is described.[5] The value of the projection of the angular momentum along this axis is restricted to certain values $M\hbar$, where $-I \leq M \leq +I$. In isotropic space, these states of different M have the same energy. There are $2I+1$ different states at this ground nuclear energy level.

3 MAGNETIC MOMENTS

A charge of magnitude, Q, moving with velocity, $\boldsymbol{v}$, experiences a force, $\boldsymbol{F}_{\text{magnetic}}$, when interacting with a field of magnetic induction, $\boldsymbol{B}$, whose source is other remote moving charges (Equation 3):[4]

$$\boldsymbol{F}_{\text{magnetic}} = Q\boldsymbol{v} \times \boldsymbol{B} \tag{3}$$

This force causes the charge to circulate about the direction of the magnetic induction. The torque on such a system describes this change of direction of the flow of charge (Equation 4):

$$\tau = Q\boldsymbol{r} \times \boldsymbol{v} \times \boldsymbol{B} \tag{4}$$

For a system of charges Q_i, the total force is given by a sum over the distribution of charges (Equation 5):

$$\boldsymbol{F}_{\text{magnetic}} = \sum_i Q_i(\boldsymbol{r_i})\boldsymbol{v_i}(\boldsymbol{r_i}) \times \boldsymbol{B}(\boldsymbol{r_i}) \tag{5}$$

Analogously, the torque about the origin is given by Equation (6):

$$\tau = \sum_i Q_i \boldsymbol{r_i} \times \boldsymbol{v_i}(\boldsymbol{r_i}) \times \boldsymbol{B}(\boldsymbol{r_i}) \tag{6}$$

One may write analogous equations for a continuous distribution of charge.

If the magnetic induction is uniform across a distribution of constant total charge such as a nucleus, there is no net magnetic force. However, the magnetic torque on such a distribution is not zero. For example, for a distribution of discrete charges, it is (Equation 7):

$$\begin{aligned} \tau = & \sum_i \frac{Q_i}{m_i} \boldsymbol{r_i} \times \boldsymbol{p_i} \times \boldsymbol{B} \\ = & \sum_i \frac{Q_i}{m_i} \boldsymbol{l_i} \times \boldsymbol{B} \end{aligned} \tag{7}$$

One defines the magnetic moment of the distribution, $\boldsymbol{m}$, by the term in brackets. The magnetic moment is a function only of the variables of the distribution. The classical equation of motion of a magnetic moment in a magnetic field of induction $\boldsymbol{B}$ is then (Equation 8):

$$\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}\boldsymbol{t}} = \tau = \boldsymbol{m} \times \boldsymbol{B} \tag{8}$$

The magnetic moment and the total angular momentum are proportional:

$$\boldsymbol{m} = \gamma \boldsymbol{l} \tag{9}$$

where γ is the magnetogyric (also called the gyromagnetic) ratio of the system, which gives the classical equation of motion of the magnetic moment in a field of induction $\boldsymbol{B}$ (Equation 10):

$$\frac{\mathrm{d}\boldsymbol{m}}{\mathrm{d}\boldsymbol{t}} = \gamma \boldsymbol{m} \times \boldsymbol{B} \tag{10}$$

The existence of a torque on the magnetic moment in a magnetic field is an indication that the potential energy of the system depends on the relative orientation of the angular momentum vector and the magnetic field. One may show that the energy contribution from this effect is of the form (Equation 11):[4]

$$E_{\text{magnetic}} = -\boldsymbol{m} \cdot \boldsymbol{B} \tag{11}$$

This is the Zeeman energy of a magnetic moment in a magnetic field. In a quantum mechanical system, the states of constant energy are the states of constant angular momentum projection along the direction of the magnetic induction. While these equations are derived classically for a small current loop,[2] similar equations apply for the Hamiltonian operators of quantum systems such as nuclear magnets.

4 THE ZEEMAN INTERACTION, HISTORICALLY

Although the Zeeman interaction is a commonly used concept in NMR, its first observation and explanation were the result of observing atomic spectra in the presence of a magnetic field. Certain spectral lines that are single in the absence of a field appear as narrowly spaced multiplets when spectroscopy is carried out in the presence of a magnetic field. The presence of an energy term as given in Equation (11) explains these results quantitatively.

Consider one of the transitions of the sodium atom (Figure 2). The $^2S_{1/2} \rightarrow {}^2P_{1/2}$ transition is well known to occur at a transition energy of 16 961 cm^{-1}.[6] (This is one component of the famous sodium D line). For a system of atoms in a magnetic field irradiated with unpolarized light, one detects four transitions in the region of this nominal energy, where only one is detected outside the magnetic field. The spacing of these lines is proportional to magnetic field strength. One may understand this effect by the presence of additional energies. Each of the two terms, $^2S_{1/2}$ and $^2P_{1/2}$, has a total angular momentum of 1/2. Each, according to quantum mechanics, consists of two states, corresponding to different values of M_J, the quantum number for the angular momentum component along the field. These states differ in energy in the presence of the magnetic field, whereas they are of the same energy in the absence of the field. In the case of one substate, the Zeeman interaction slightly raises the energy relative to the nominal energy, and the other energy is slightly lowered. The result is that transitions between states occur at slightly different energies owing to the Zeeman interaction.

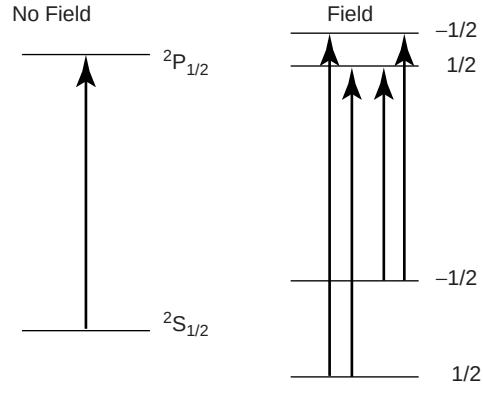

Figure 2 The splitting of one component of the sodium D line by the Zeeman coupling to a magnetic field.

5 THE ENERGIES OF NUCLEAR SPINS IN A MAGNETIC FIELD

Nuclei are charged particles described by the intrinsic angular momentum (spin). In a magnetic field, the states of different angular momentum projection along the field have energies that depend on the projection quantum number, the induction strength and the magnetogyric ratio, according to Equation (12):

$$E_M = -M\gamma\hbar B \tag{12}$$

As a result of the Zeeman interaction's dependence on the quantum number M, there is a series of energy levels ($2I + 1$ levels for a nucleus of spin quantum number, I), each of slightly different energy in the presence of a magnetic field. The spacing between them depends linearly on the magnetic induction, as shown in Figure 3 for spin 1/2. For a spin 1, there will be three states, and for a spin 3/2 there will be four states, etc.

Transitions between the levels by energy absorption from an electromagnetic field cause the nuclei to change between these states, as indicated by the bold arrow in Figure 3. In the typical experiment these are single-quantum transitions, $\Delta M = \pm 1$. The energy difference between the two states determines the type of electromagnetic field to produce the transitions, as given in Equation (13):

$$\Delta E = \hbar\gamma B = \hbar w_{\mathrm{L}} \tag{13}$$

The energy difference depends on the magnetic field, and also on the size of the magnetogyric ratio. The product of the magnetogyric ratio and the magnetic field is an angular frequency, w_{L}, the Larmor frequency. For typical magnitudes of magnetic fields achievable in the laboratory, the energy differences are quite small, and nuclear Larmor frequencies are in the radiofrequency region of the electromagnetic spectrum. Thus, NMR spectroscopy is thought of as a radiofrequency experiment, although it may occur at other frequencies depending on the magnetic field strength. Because the energy spacing is so small, the population difference between the two states is quite small and the signals are quite weak compared with many other forms of spectroscopy.

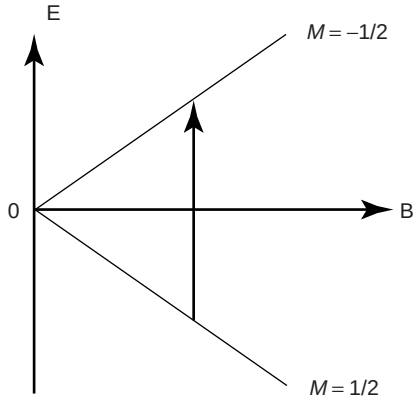

Figure 3 The energies, E, of the two substrates of a spin 1/2 in a magnetic field of induction, B.

From the sizes of the magnetogyric ratios in Table 1, it can be seen that the resonance of each nucleus occurs at a characteristic frequency quite different from those of other nuclei in a field of magnetic induction B, so that one may interrogate only one (or at most a limited number of different) nuclear species in an experiment. For example, in a field in which the proton resonance frequency is about 100 MHz, that of ^{13}C occurs near 25 MHz.

6 VECTOR AND TENSOR DESCRIPTIONS IN NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Magnetic fields are vectors, as are the magnetic moments of nuclei. The description of the energies of coupling between fields and moments, among moments and with other physical fields are couched as quantities proportional to products of the components of these vectors. The most general way to express these relationships is through tensors that specify the dependence of energies on the various components of vectors (see, for example, Jeffreys[7]). For example, the energy of the direct coupling of a magnetic moment to a magnetic field through the Zeeman interaction Equation (11) was stated to be proportional to the dot product of the vectors. One may think of this product as having the form (Equation 14):

$$E_{\mathrm{magnetic}} = -\boldsymbol{m}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\boldsymbol{B} \tag{14}$$

where the interposed matrix describes how the coupling of each component of $\boldsymbol{m}$ to each component of $\boldsymbol{B}$ contributes to the energy. In this case, only components of the two vectors along the same coordinate axis couple to produce energy, as indicated by zeros in the elements that would indicate coupling between components along different axes.

For more general cases, the energy of interaction may depend on the products of components along different axes. Thus, bilinear interaction energies (or, in quantum systems, the Hamiltonian operators bilinear in the operators of the system) may be written in the more

general form (Equation 15):

$$E_{\text{interaction}} = \boldsymbol{X} \cdot \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \cdot \boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{T} \cdot \boldsymbol{Y} \tag{15}$$

where $\boldsymbol{T}$ is a matrix of coefficients that describes the coupling between components of the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$.

An important feature of tensor interactions is the relation of the form of the information in one reference frame to that in another. For example, interactions are naturally expressed in terms of molecular coordinates defined by the features of the molecule and its surroundings, whereas the experiment is defined by a coordinate system specified by the magnetic field and the direction of a coil that emits and detects radiofrequency radiation. The relation of these two descriptions of the same interaction is determined by the relation of the expression of the tensor in the first frame to that in the latter.

The use of symmetry can often simplify the task of relating tensor descriptions from one frame to another. A general second-rank Cartesian tensor can be expressed as the sum of three tensors, each of which has certain symmetry properties (Equation 16):

$$\boldsymbol{T} = \boldsymbol{T_0} + \boldsymbol{T_1} + \boldsymbol{T_2} \tag{16}$$

The elements of the matrix of $\boldsymbol{T_0}$ have the following property (Equation 17):

$$t_{0,ij} = T_0 \delta_{ij} \tag{17}$$

T_0 is related to the trace of the tensor, given in terms of the diagonal elements of the tensor $\boldsymbol{T}$ (Equation 18):

$$T_0 = \frac{1}{3} \sum_i t_{ii} \tag{18}$$

The tensor $\boldsymbol{T_1}$ is antisymmetric about the diagonal (Equation 19):

$$t_{1,ij} = -t_{1,ji} \tag{19}$$

The tensor $\boldsymbol{T_2}$ is symmetric (Equation 20):

$$t_{2,ij} = t_{2,ji} \tag{20}$$

and has a trace of zero (Equation 21):

$$\sum_i t_{2,ii} = 0 \tag{21}$$

The elements of these two tensors may be expressed in terms of the elements of the general tensor, $\boldsymbol{T}$ (Equations 22 and 23):

$$t_{1,ij} = \tfrac{1}{2}(t_{ij} - t_{ji}) \tag{22}$$

and

$$t_{2,ij} = \tfrac{1}{2}(t_{ij} + t_{ij}) - T_0 \delta_{ij} \tag{23}$$

An examination of Equation (15) indicates that the right-hand side may be considered the dot product of two vectors. If one defines $\boldsymbol{Z}$ in the following manner (Equation 24):

$$\boldsymbol{Z} = \boldsymbol{T} \cdot \boldsymbol{Y} \tag{24}$$

the overall interaction is of the form of the dot product of two vectors. Specifically, if one vector is the magnetic moment, the resulting interaction is of the form of a Zeeman interaction if one considers the other vector to be an ''effective magnetic induction.'' In many cases, interactions spins experience can be put in this form. In some cases, the interaction is said to result from a ''local'' or ''effective'' field.

7 COUPLING OF SPINS MODELED AS ZEEMAN INTERACTIONS

Most spin interactions in NMR spectroscopy are represented by second-rank tensors of the type given above. In one convenient way, the interactions may be viewed as Zeeman-like couplings to ''effective fields.'' As examples, we quote these for the most common interactions.

7.1 Shielding

The interaction that makes NMR spectroscopy particularly useful for analysis of chemical structure is electronic shielding of the nucleus from the field. Generally, the effect is measured as the chemical shift of the resonance relative to some reference species.

The origin of shielding is the coupling of the nuclear spin to the orbital angular momentum of the electrons in the vicinity of the nucleus.[8,9] The simplest physical model is that of an unfettered charged particle, such as an electron, in a magnetic field circulating about the field direction. Such a circulating charge produces a field of its own, which contributes to the field at the site of a nucleus in the nearby region. The NMR experiment detects these variations in field as differences in effective nuclear Larmor frequencies at the various sites.

In quantum mechanical developments, the shielding is given as a second-order perturbation of the main Zeeman interaction by the direct interaction of the nuclear spin with the orbital angular momentum of the electrons in the molecular orbitals.[7] The resulting energy shift of a particular nuclear spin state is given by integrals over

the electronic state. The shielding field does not have to be collinear with the polarizing field, $\boldsymbol{B_0}$, and the energy corresponding to this interaction is of the form (Equation 25):

$$E_{\text{shielding}} = \boldsymbol{m} \cdot \boldsymbol{\sigma} \cdot \boldsymbol{B_0} = \boldsymbol{m} \cdot \boldsymbol{B}_{\text{cs}} \tag{25}$$

where $\boldsymbol{\sigma}$ is the shielding tensor, and $\boldsymbol{B}_{\text{cs}}$ is the chemical-shielding field. Expressions for the chemical shielding in terms of the molecular orbitals are given in texts on NMR. (see, for example, Slichter[10]). The representation of the shielding tensor in some Cartesian coordinates (x, y, z) can be divided into a sum of three tensors, as indicated above (Equations 26–29):

$$\boldsymbol{\sigma} = \boldsymbol{\sigma_0} + \boldsymbol{\sigma_1} + \boldsymbol{\sigma_2} \tag{26}$$

where

$$\boldsymbol{\sigma_0} = \begin{pmatrix} \sigma_{\text{iso}} & 0 & 0 \\ 0 & \sigma_{\text{iso}} & 0 \\ 0 & 0 & \sigma_{\text{iso}} \end{pmatrix} \tag{27}$$

and

$$\boldsymbol{\sigma_1} = \frac{1}{2} \begin{pmatrix} 0 & (\sigma_{xy} - \sigma_{yx}) & (\sigma_{xz} - \sigma_{zx}) \\ (\sigma_{yx} - \sigma_{xy}) & 0 & (\sigma_{yz} - \sigma_{zy}) \\ (\sigma_{zx} - \sigma_{xz}) & (\sigma_{zy} - \sigma_{yz}) & 0 \end{pmatrix} \tag{28}$$

and

$$\boldsymbol{\sigma_2} = \frac{1}{2} \begin{pmatrix} 2(\sigma_{xx} - \sigma_{\text{iso}}) & (\sigma_{xy} + \sigma_{yx}) & (\sigma_{xz} + \sigma_{zx}) \\ (\sigma_{yx} + \sigma_{xy}) & 2(\sigma_{yy} - \sigma_{\text{iso}}) & (\sigma_{yz} + \sigma_{zy}) \\ (\sigma_{zx} + \sigma_{xz}) & (\sigma_{zy} + \sigma_{yz}) & 2(\sigma_{zz} - \sigma_{\text{iso}}) \end{pmatrix} \tag{29}$$

Some physical qualities expressible as tensors have inherent symmetries requiring specific relationships among tensor elements. Although it is not required by the physics of NMR spectroscopy, the chemical-shielding tensor is often found to be symmetric.[11] Thus, $\boldsymbol{\sigma_1}$ is often practically taken to be the null matrix. Using this simplification (Equation 30),

$$\boldsymbol{B_{cs}} = \boldsymbol{\sigma_0} \cdot \boldsymbol{B_0} + \boldsymbol{\sigma_2} \cdot \boldsymbol{B_0} \tag{30}$$

The chemical shielding tensor in a particular molecular frame [the principal-axis system (PAS)] has the form (Equations 31 and 32):

$$\boldsymbol{\sigma_0}^{\text{PAS}} = \sigma_{\text{iso}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{31}$$

and

$$\boldsymbol{\sigma_2}^{\text{PAS}} = \delta \begin{pmatrix} -\frac{1}{2}(1+\eta) & 0 & 0 \\ 0 & -\frac{1}{2}(1-\eta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{32}$$

where the isotropic shielding, σ_{iso}, the anisotropy δ, and the asymmetry η are given in terms of the principal components, σ_{XX}, σ_{YY}, and σ_{ZZ} in the PAS (Equations 33–35):

$$\sigma_{\text{iso}} = \tfrac{1}{3}(\sigma_{XX} + \sigma_{YY} + \sigma_{ZZ}) \tag{33}$$

$$\delta = \tfrac{2}{3}\sigma_{ZZ} - \tfrac{1}{3}(\sigma_{XX} + \sigma_{YY}) \tag{34}$$

$$\eta = \frac{\sigma_{YY} - \sigma_{XX}}{\delta} \tag{35}$$

In any other frame of reference obtained by rotation of the coordinate axes, the tensor is given by the following relation to the tensor in the PAS (Equation 36):

$$\begin{aligned} \sigma &= \Re \cdot \sigma_0^{\text{PAS}} \cdot \Re^{-1} + \Re \cdot \sigma_2^{\text{PAS}} \cdot \Re^{-1} \\ &= \sigma_0^{\text{PAS}} + \Re \cdot \sigma_2^{\text{PAS}} \cdot \Re^{-1} \end{aligned} \tag{36}$$

where $\Re$ is a transformation matrix that connects the two reference frames.

The frame of interest in NMR spectroscopy is one in which one of the axes is along the direction of the magnetic induction, usually the z-axis since it is the component of the magnetic moment along this axis that, in the first approximation, determines the resonance condition. In high magnetic fields (which is usually satisfied by the fields achievable in modern magnets), only the zz component of the tensor in this frame is measurable with NMR. This component depends on the rotation angles that define $\Re$; these are the Euler angles, α, β and γ, which describe the relation of the PAS frame to the magnetic field frame.[6] In terms of these angles, one obtains an explicit form for this component of the shielding tensor (Equation 37):

$$\sigma_{zz} = \sigma_{\text{iso}} + \delta \left(\frac{3\cos^2\beta - 1}{2} + \frac{1}{2}\eta \sin^2\beta \cos 2\gamma \right) \tag{37}$$

The total effective field at a particular nuclear site is given by the sum of the chemical-shielding field and the strong external polarizing field, B_0 (Equation 38):

$$B = (1 - \sigma_{zz})B_0 \tag{38}$$

Frequently, one examines the NMR spectrum of a solution of molecules. The thermal motions of the molecules in the solution produce a modulation of the orientation of the PAS relative to the magnetic field frame. When this motion is fast and isotropic, the average field felt by the nuclei may be calculated as the average over all possible values of the Euler angles. In this case, the effective field becomes (Equation 39):

$$B = (1 - \sigma_{\text{iso}})B_0 \tag{39}$$

an approximation used in the analysis of most NMR spectra of isotropic solutions.

For solid samples, the chemical shielding field depends on the orientation of the PAS to the magnetic field axes. For samples in which several orientations of the PAS relative to the field direction are simultaneously present, the NMR spectrum will not generally show a single sharp line for a particular nuclear site, as is seen in solution. Thus, for single-crystal samples, the NMR spectrum reflects the number of geometrically unique sites in a unit cell. If the sample is a powder, then there will exist a range of values of the chemical shielding for a single site, and the spectrum for a nucleus will be a band whose range and relative shape are determined by the principal values of the shielding tensor.[12]

7.2 The Dipole–Dipole Interaction

Nuclear spins (or, for that matter, electronic spins) are sources of magnetic field that can affect the NMR properties of other nearby spins. This coupling of nearby spins occurs by the direct dipole–dipole interaction. For example, Figure 4 shows schematically the geometry of two magnetic moments that happen, in this case, to be polarized in the same direction. The moment, $\boldsymbol{m_1}$, acts like a bar magnet, producing a magnetic field at the site of $\boldsymbol{m_2}$, with which the latter interacts by a Zeeman interaction. The energy of this interaction is (Equation 40):

$$E_{\mathrm{dd}} = -\boldsymbol{m_2} \cdot \boldsymbol{B_{dd,1}} \tag{40}$$

The form of the field due to a dipole is found from the classical model of a steady-state current distribution.[2] Such a distribution is described by its magnetic moment, $\boldsymbol{m}$. At a point, $\boldsymbol{R}$, remote from this distribution, the vector potential is given by Equation (41):

$$\boldsymbol{A}(\boldsymbol{R}) = \frac{\boldsymbol{m} \times \boldsymbol{R}}{|\boldsymbol{R}|^3} \tag{41}$$

From this form, one finds the magnetic induction, $\boldsymbol{B(R)}$, as the curl of this vector potential. The resulting dipolar field of magnetic moment, $\boldsymbol{m_1}$, is (Equation 42):

$$\boldsymbol{B_{dd,1}}(\boldsymbol{R}) = 3 \quad \frac{(\boldsymbol{R}\boldsymbol{m_1})}{|\boldsymbol{R}|^5} \quad \boldsymbol{R} - \frac{1}{|\boldsymbol{R}|^3}\boldsymbol{m_1} \tag{42}$$

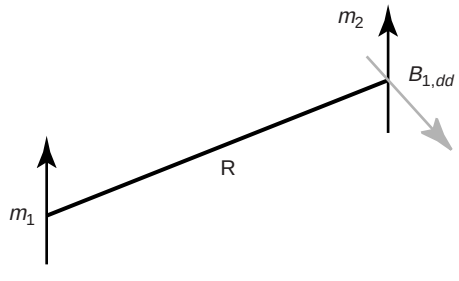

Figure 4 The geometry of the dipole–dipole coupling of two magnetic moments, showing the effective dipolar field.

This dipolar field has components along the dipole, $\boldsymbol{m_1}$, and along the vector, $\boldsymbol{R}$, connecting the position of the dipole with the point at which the induction is measured. In many materials, the field at a point is determined by a number of magnetic moments located at a variety of points, $\{\boldsymbol{R_i}\}$. In this case, the dipolar field is a sum over all such sources of magnetic induction (Equation 43):

$$\boldsymbol{B_{dd}}(\boldsymbol{R}) = \sum_i \left(\frac{3(\boldsymbol{R} - \boldsymbol{R_i})\boldsymbol{m_i}}{|\boldsymbol{R} - \boldsymbol{R_i}|^5} \quad (\boldsymbol{R} - \boldsymbol{R_i}) - \frac{1}{|\boldsymbol{R} - \boldsymbol{R_i}|^3}\boldsymbol{m_i} \right. \tag{43}$$

The interaction may also be couched in terms of a tensor product of the type (Equation 44):

$$E_{\mathrm{dd}} = \boldsymbol{m_1} \cdot \boldsymbol{D} \cdot \boldsymbol{m_2} \tag{44}$$

where the tensor $\boldsymbol{D}$ is the dipole–dipole tensor, whose elements may be determined from, for example, Equation (42).

Like the chemical-shielding tensor, the dipolar tensor may be expressed with respect to various reference frames, and the form of the interaction in one frame is related to that in another by a transformation described by a tensor $\Re$ that depends on the Euler angles. One may show that the isotropic part of the dipolar tensor is zero, a fact that has the profound effect that, for a molecule undergoing rapid isotropic tumbling, the time-averaged dipolar interaction does not affect the energy of the states in first order. In the NMR spectroscopy of molecules in nonviscous solutions, the dipolar coupling does not have to be considered in calculating observed shifts or splittings. (It does affect the NMR spectrum in second order and is often a major consideration in specifying line widths and relaxation times.) For solids, the dipolar coupling among the spins of a sample can be the major effect in an NMR spectrum, particularly for spin 1/2. Because of the large number of spins in a material, the spectrum must be interpreted in terms of a dipolar field of many spins, rather than that of a single spin. The development of techniques to overcome the effects of the dipolar field in a solid to reveal smaller interactions is a major triumph of the study of the physics of spin systems.[11]

7.3 The J Coupling Between Spins

In addition to the direct coupling of nuclei via the dipolar coupling and the coupling to the orbital motion of the electrons that gives rise to the chemical shielding, nuclei also experience an electron-mediated coupling, known as the J coupling or the indirect coupling, that may also be treated as a Zeeman coupling.[13,14] The origin of this effect is a slight interaction between a nuclear spin and

an electron that slightly modifies the state of the electron. This perturbation is then "felt" by the second nucleus with which the electron is in contact. The net effect is that the energy states of the spins depend on their relative spin states (Equation 45):

$$E_J = \boldsymbol{m_1} \cdot \boldsymbol{J} \cdot \boldsymbol{m_2} \tag{45}$$

The elements of the tensor $\boldsymbol{J}$ are integrals of functions of the electronic coordinates over all participating molecular orbitals.[15] Once again, one sees that, by treating the product of the $\boldsymbol{J}$ tensor with one of the magnetic moments as an effective magnetic field, this interaction is also of the form of a Zeeman interaction with a "local field."

The indirect coupling's tensor qualities can be seen, in some cases, in the spectra of solids.[16] However, in NMR spectra of isotropic solutions, one detects only the isotropic average indirect coupling between spins. Geometric effects are seen in the size of the isotropic coupling constants measured in solution because motion about bonds may be restricted. Thus, the values of three-bond couplings are determined by internal conformations through devices such as the Karplus relation.[2]

The J coupling presupposes the existence of a difference between nuclei due to differences in chemical shielding. I have tried to describe the effects of the various interactions in a classical manner as due to effective fields from the various interactions. Such a description works well if the sizes of the interaction energies are quite different. However, when the magnitudes of interactions become comparable, then one must consider the quantum nature of the interferences between the two interactions beyond the simple effective-field approach. The results for the case one is most likely to encounter, the J coupling and the chemical-shielding difference, have been worked out in most texts on NMR, and the effects are described in detail.[11]

8 CONCLUSIONS

The interaction of spins with magnetic fields produces splittings of energy levels, which are detected by energy absorption. One such coupling is the Zeeman interaction with an applied field of magnetic induction, $\boldsymbol{B}$. Other interactions of spins may be viewed as Zeeman interactions with "effective" fields. Specifically, the coupling to the orbital angular momentum of the electrons leads to the shielding by an effective field related to the electronic state. The direct dipolar coupling of a magnetic moment with the field produced by other magnetic moments can also be considered to result from a Zeeman interaction. The J coupling is represented by an effective electron-mediated field produced by interaction of electrons with a nucleus at another site. While the classical results of Zeeman-like models hold in many cases, where spins are strongly coupled quantum mechanical models of the interactions of several spins may not easily be pictured in terms of an effective field. Nevertheless, the concept of an effective Zeeman field often provides insight into the physics underlying the observation of multiple resonance lines in a spectrum.

ABBREVIATIONS AND ACRONYMS

NMR	Nuclear Magnetic Resonance
PAS	Principal-axis System

RELATED ARTICLES

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy **(Volume 13)**
Chemical Shifts in Nuclear Magnetic Resonance

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy cont'd **(Volume 14)**
Quadrupole Couplings in Nuclear Magnetic Resonance, General • Scalar Couplings in Nuclear Magnetic Resonance, General

REFERENCES

1. R.K. Harris, *Nuclear Magnetic Resonance Spectroscopy*, Pitman, London, 1983.
2. E.U. Condon, G.H. Shortley, *The Theory of Atomic Spectra*, Cambridge University Press, Cambridge, 4, 1985.
3. C. Kittel, W.D. Knight, M.A. Ruderman, *Mechanics*, McGraw-Hill, New York, 175–185, 1965.
4. E.M. Purcell, *Electricity and Magnetism*, McGraw-Hill, New York, 1965.
5. P.J. Hore, *Nuclear Magnetic Resonance*, Oxford University Press, Oxford, 1995.
6. G. Herzberg, *Atomic Spectra and Atomic Structure*, Dover, New York, 107, 1944.
7. H. Jeffreys, *Cartesian Tensors*, Cambridge University Press, Cambridge, 1974.
8. N.F. Ramsey, 'Magnetic Shielding of Nuclei in Molecules', *Phys. Rev.*, **78**, 699–703 (1950).
9. N.F. Ramsey, 'Chemical Effects in Nuclear Magnetic Resonance and in Diamagnetic Susceptibility', *Phys. Rev.*, **86**, 243–246 (1952).
10. C.P. Slichter, *Principles of Magnetic Resonance*, 3rd edition, Springer-Verlag, Heidelberg, 92–108, 1992.
11. U. Haeberlen, *High Resolution NMR in Solids: Selective Averaging*, Academic Press, New York, 31–35, 1976.

12. B.C. Gerstein, C. Dybowski, *Transient Techniques in NMR of Solids*, Academic Press, New York, 1985.
13. H.S. Gutowsky, D.W. McCall, C.P. Slichter, 'Nuclear Magnetic Resonance Multiplets in Liquids', *J. Chem. Phys.*, **21**, 279–292 (1953).
14. N.F. Ramsey, 'Electron-coupled Interactions between Nuclear Spins in Molecules', *Phys. Rev.*, **91**, 303–307 (1953).
15. J.A. Pople, W.G. Schneider, H.J. Bernstein, *High Resolution Nuclear Magnetic Resonance*, McGraw-Hill, New York, 1959.
16. G. Wu, R. Wasylishen, 'Anisotropic Nuclear Spin Interactions in $K_2Hg(CN)_4$: A Multinuclear Solid-State NMR Study', *J. Phys. Chem.*, **97**, 7863–7869 (1993).

Chemical Analysis by Nuclear Methods: Introduction

Zeev B. Alfassi
Ben Gurion University, Beer Sheva, Israel

Chemical analysis by nuclear methods is usually an elemental analysis, i.e. it determines the contents of the various elements in the analyzed sample but cannot tell in what chemical form (compounds, valence states) they are. The analysis is based on either a reaction of the analysed element with nuclear projectiles (neutrons or accelerated charged particles, e.g. electrons, protons, heavy ions or γ-photons) (see the following articles within this section: **Instrumental Neutron Activation Analysis; Instrumental Neutron Activation Analysis: Gamma Lines Table; Charged Particle Activation Analysis; Cyclic Activation Analysis; Particle-induced γ-Ray Emission; PIXE (Particle-induced X-ray Emission); Radiochemical Neutron Activation Analysis; Nuclear Reaction Analysis** and **Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons**) or on the scattering of accelerated charged particles. The reaction can be written in the same form as a chemical reaction:

$$\text{Target} + \text{Projectile} \longrightarrow \text{Light product} + \text{Heavy product}$$

or in the more concise form of writings of nuclear physics:

$$\text{Target (Projectile, Light product) Heavy product}$$

Thus, for example, the first production of artificial radionuclide by Joliot and Curie was done by the reaction of α-particles with aluminum metal:

$$^{27}\text{Al} + \alpha \longrightarrow {}^{30}\text{P} + \text{n} \quad (\alpha \text{ is the } {}^4\text{He nucleus})$$

or in the physics notation: $^{27}\text{Al}(\alpha, \text{n})^{30}\text{P}$.

The light product is similar to the projectile and can be a neutron, a small charged particle or a photon. The basis of the nuclear methods for chemical analysis via nuclear reaction (nuclear activation) is the measurement of the amount of light or heavy product formed in a known flux of projectiles for a known length of time (irradiation time). The amount produced is proportional to the number of target atoms, and hence measurement of the amount of the products yields the amount of target atoms. The amount of products formed is too small to be measured chemically (except in very rare cases), and the only way to measure them is by the nuclear physics method of pulse counting. The light product can be measured due to its energy if it is a photon or a charged particle of high kinetic energy, or due to its reaction if it is a neutron. The common denominator of all these processes is that they must be done a very short time after formation of the product, otherwise the light product will lose either its energy or its identity by reaction with the surrounding media. Consequently, measurement of the light product must be done during bombardment of the target with projectiles. This kind of measurement is called prompt activation analysis.

A similar situation exists when the analysis is done by scattering of the projectile, and the measurement is made simultaneously with the irradiation. In this case qualitative analysis is based not on the identity of the products but rather on the kinematics of the scattering, i.e. the energy and direction of the scattered particles. Quantitative measurement is done by counting the number of scattered particles. In the case where the heavy product is radioactive, i.e. it undergoes a spontaneous nuclear transformation, its identity and amount can be measured by its radioactivity. Because the radioactivity can be measured after the end of the interaction between the target and the beam of projectiles, this method of analysis is called delayed activation analysis. Because the heavy product must be radioactive, not all elements can be measured with each projectile. However, by careful choice of the bombarding projectile, almost every element can be determined.

Because many elements are activated simultaneously, the radioactivity measurement must yield both the identity of the radioactive nuclides and their amounts. The term that is sometimes used – radioisotopes – is wrong because when saying isotopes we should say of what element, see **Elemental Analysis by Isotope Dilution**. Nuclide is the nuclear physics term that is parallel to ''atom'' in chemistry. It consists of a nucleus (composed of neutrons and protons) and electrons around it. Because the identity of radionuclides (a short term for radioactive nuclides) in a mixture usually cannot be determined from β^-- or β^+-emission (unless there are only very few in the mixture and they differ strongly in their half-lives or energy of the β-particles), only radionuclides that also emit γ-rays can be used in activation analysis. The radionuclides used for identification are called indicator radionuclides. Only few pure β^--emitting indicator radionuclides, such as ^{32}P, can be used in activation analysis. In this case the phosphorus should be separated from the other radionuclides in the activated sample or the other radionuclides are left to decay before the measurement. The chemical analysis by nuclear activation in which chemical separation precedes the radioactivity measurement is named **Radiochemical Neutron Activation Analysis**, whereas analysis by nuclear activation with direct measurement of the nuclear activity is named Instrumental, see **Instrumental Neutron Activation Analysis** and **Instrumental Neutron Activation Analysis: Gamma**

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Lines Table. Owing to the method of measurement of pulse counting, very small amounts of radionuclides can be assessed quantitatively. This and the fact that most natural nuclides are nonradioactive are the basis for the use of radionuclides as tracers in chemical analysis.

Scattering of charged particles is measured both in the forward direction (by **Elastic Recoil Detection Analysis**) and in the backward direction (by **Rutherford Backscattering Spectroscopy**). In the first years of studies mainly the scattering of light ions was studied, but later the scattering of heavier ions became more frequent. The widespread availability of 2–3-MeV accelerators made their use more popular. However, in forward scattering they are restricted to the analyses of H and D. Heavier elements can be analyzed with the beams by backward scattering or by forward scattering of heavier ions. In the recent Ion Beam Analysis Conference at Lisbon (1997) there were two papers on He-induced elastic recoil detection analysis compared to 26 papers on heavy-ion elastic recoil detection analysis. However, for routine analysis the situation is different owing to the availability of the low-energy He accelerators.

Activation by neutrons is carried out mainly in the delayed mode, where very low detection limits can be obtained. Prompt activation is used mainly for in situ and on-line analysis and for elements that cannot be determined by the delayed mode due to the product being nonradioactive or having too short or too long a half-life. An example is the determination of hydrogen and nitrogen atoms by PGNAA (prompt gamma neutron activation analysis), mainly in food samples, see **Prompt γ-Neutron Activation Analysis**.

The slowing down of high-kinetic-energy charged particles passing through matter allows not only the determination of the total concentrations of some elements but also their distribution as a function of distance from the surface, i.e. depth profiles of the concentrations of the various elements. The slowing down can change the energy of the emitted charged particle, e.g. in the reaction $^{10}B(n, \alpha)^{7}Li$. The energy of the α-particle crossing the surface of the sample to reach the detector indicates the distance the α-particle travels in the sample via the energy loss from the original energy of the α-particle.

The slowing down of an impinged high-energy projectile can reduce its energy to the range where a resonance reaction can occur. Thus, by changing the original energy of the projectile, the depth profile of the concentration can be measured.

All nuclear methods, except those using radiotracers, use either nuclear reactors or ion accelerators. These are quite expensive machines that are not found in common analytical laboratories but rather in special laboratories. Many of these methods involve the cooperation of chemists and nuclear physicists.

Activation Analysis of Large Samples

Peter Bode
Delft University of Technology, Delft, The Netherlands

1 Introduction **1**
2 Large Sample Activation Analysis **2**
2.1 Large Sample Neutron Activation Analysis 2
2.2 Large Sample Prompt γ-Activation Analysis 3
2.3 Large Sample Photon Activation Analysis 4
3 Measurement Equation of Large Sample Neutron Activation Analysis **4**
4 Instrumentation **5**
4.1 Neutron Sources for Large Sample Activation Analysis 5
4.2 Irradiation Facility 6
4.3 Sample Containers for Irradiation 7
4.4 Counting Facility 8
5 Calibration **10**
5.1 Absolute Method 10
5.2 The Comparator Method 11
5.3 Single Comparator Method 11
5.4 k_0-Based Method for Standardization 11
5.5 Internal Monostandard Method 12
5.6 Neutron Fluence Rate Monitoring 12
5.7 γ-Ray Self-Attenuation 13
5.8 Extreme Inhomogeneities 13
6 Quality Control **14**
6.1 Quality Control in Large Sample Analysis 15
7 Sensitivity **17**
7.1 Natural Background 18
8 Applications **18**
8.1 Materials Difficult to Homogenize: Geological Samples, Ores, and Waste 18
8.2 Materials That May be Contaminated During Homogenization: High-Purity Materials 19
8.3 Materials Difficult to Subsample: Nutritional Studies 19
8.4 Valuable Material of Irregular Shape 19
8.5 Other Applications 19
References **19**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

There is a trend in many analytical techniques towards the use of smaller sizes of the test portion, and sometimes submilligram amounts are being used. The limitation to the size of the test portion can pose the analyst with problems when the amount of material collected is large. Subsampling and preparation of a representative test portion may be difficult if homogenization is impossible or extremely expensive, or if homogenization introduces contamination. An alternative approach has been introduced in the 1990s by the development of large sample neutron activation analysis (LSNAA), later followed by large sample photon activation analysis. These techniques are capable of direct analysis of samples with masses of hundreds of grams to several kilograms. Though the principles and physics of large sample activation analysis are thoroughly understood, the method is still not as versatile or applicable as, for example, normal small sample neutron activation analysis. In this article, the physics of LSNAA is described, including methods of calibration. Examples are given of irradiation and counting facilities and the special aspects of quality control are discussed. Several examples are given of applications of large sample analysis, e.g., for studies related to electronic waste, complete archaeological and cultural artifacts, high-purity materials, and materials of irregular shape.

1 INTRODUCTION

All the routine multielemental analysis techniques (inductively coupled plasma atomic emission spectrometry (ICP-AES), inductively coupled plasma mass spectrometry (ICP-MS), and instrumental neutron activation analysis (INAA), etc.) employ rather small test portions of material, varying from a few milligrams to a few grams of solids or in the range of a few milliliters in the case of liquids[1] (see Table 1). There is even a tendency to go for smaller test portions, such as in solid-state atomic absorption spectrometry (AAS), laser-ablation ICP, and total reflection X-ray fluorescence (XRF) spectrometry. In XRF, the use of quantities larger than required to prepare the target is anyhow meaningless, as the derived information is from the surface layers, representing a few milligrams only.

The limitation to the size of the test portion can pose problems to the analyst when the amount of material collected is large. This is often the case since soils, rocks, plant material, etc. can be more easily and

Table 1 Sizes of the samples and analytical portions handled in several multielement analysis techniques[1]

Analysis technique		Solid material mass used or prepared to test portion	Volume used as test portion
Atomic absorption spectroscopy (AAS)	Gas furnace	Typically 1–2 g dissolved; maximum approximately 10 g	10–20 μL
AAS	Flame		1–2 mL
Inductively coupled plasma spectroscopy (ICP)		Typically 1–2 g dissolved; maximum approximately 10 g	Approximately 500 μL
X-ray fluorescence spectroscopy (XRF)		10 g	
Instrumental neutron activation analysis (INAA)		Typically approximately up to 500 mg; in some cases up to 30 g	1–50 mL

representatively sampled at quantities in the order of hundreds of grams to kilograms than at quantities less than 1 g. A sample is denoted as ''representative'' when ''it can be expected to exhibit the average properties of the material, environment or population it was taken from''.[2] Representativeness is a priori preserved when (i) the sampling is performed according to specific, certified norms or when (ii) a truly homogeneous material is sampled. Homogeneity is defined as ''the degree to which a property or substance is randomly distributed throughout the material''.[2]

Homogeneity depends on the size of the units under consideration. A mixture of minerals may be inhomogeneous at the molecular or atomic level but homogenous at the particle level. In chemical analysis this unit is obviously correlated with its effect in the conduct of the analysis (e.g. differences in solubility) or in its interpretation. Thus, both from practical and sampling considerations often more material is collected and presented for analysis than can be handled.

Irrespective of the analysis technique selected, attention has to be paid to representative subsampling to obtain a relevant final analytical portion from the originally collected material. When restricting the discussion to the analysis of solid materials, this subsampling may imply sample size reduction techniques and other processing such as sieving, crushing, milling, or blending. Problems of the following types arise:

1. Homogenization is impossible, or extremely expensive, because of material properties. Examples are electronic circuits on printed boards, household waste, and scrap from recycled electronics, automobiles, and plastics. A solution to this problem is to sort the material and to perform individual homogenizations and, subsequently, analyze, thereby increasing the total project costs.
2. The homogenization step results in contamination of the sample. Often the contamination due to crushing or milling is not controllable on every sample type. When processing a large series of samples, careful interim cleaning may get less attention since it is time consuming and therefore expensive.

Testing the degree of homogeneity is a common practice in the preparation of reference materials, but for routine operations, such a procedure, the requirement of analysis and statistical evaluation of at least five or more test portions of each sample, would raise the cost of the analysis considerably.

Considerations from the above indicate that direct analysis of the voluminous solid sample as collected might have advantages, both analytical and economical.

Detection limits in trace-element studies are based on the signal-to-noise ratio. An additional feature of analysis of large test portions is that the detection limit for trace elements may be decreased considerably in case of high-purity materials. This was demonstrated, for example, by Verheijke[3] in the assessment of impurities in (5 in. diameter) silicon wafers to be used in the electronic industry.

2 LARGE SAMPLE ACTIVATION ANALYSIS

2.1 Large Sample Neutron Activation Analysis

In analytical terms, a ''large sample'' can be anything exceeding the regular size of a test portion in the process to determine the components of the material. The regular mass of a test portion in neutron activation analysis (NAA) varies from a few milligrams to 1 g. As already indicated, instrumental neutron activation analysis (INAA) has all the potentials to analyze, with adequate accuracy, test portions[4] in the kilogram range.[5–8] Both the incoming radiation for activation (neutrons) and the outgoing radiation to be measured (γ-rays) have sufficiently high penetrating power to facilitate NAA of portions of samples weighing kilograms. A ''large sample'' in NAA is defined as a test portion in which neutron and γ-ray self-attenuation cannot be neglected in view of the required degree of accuracy.

A few phenomena need more attention in large sample neutron activation analysis (LSNAA) than in normal NAA (handling test portions varying from micrograms to a maximum of 0.5 g), where these phenomena usually have only insignificant impact to the degree of accuracy of the results.[9] In large test portions, e.g., of kilogram size, neutron absorption and scattering result in substantial self-shielding, causing depression of the neutron flux at the center of the sample compared to the periphery. Neutron self-thermalization may cause substantial changes in the neutron spectrum throughout the sample if the sample material also contains, for example, hydrogen.

Similarly, the γ-radiation of the activation products deep inside in the sample will be more strongly absorbed and scattered before leaving the sample than the radiation resulting from, e.g., the surface of the sample; moreover the absorption and scattering increase rapidly at lower γ-ray energies. This effect is denoted as *γ-ray self-attenuation*. In addition, a sample of, say, 1 kg cannot be considered anymore as a more-or-less ''point source'' during counting at normal sample–detector distances of, e.g., 10–30 cm, resulting in a corresponding different response of the detector for the γ-radiation. In contrast to conventional INAA with small samples, analysis of test portions larger than a few hundred milligrams requires correction for these neutron self-shielding and γ-ray self-absorption effects, either via calibration or by modeling.

Trace-element determinations in large test portions have been carried out for decades in areas such as well logging,[10] on-line conveyor belt industrial analyzers,[11] and in vivo studies of, e.g., Ca in bones and Cd in kidneys.[12] In all these studies, the NAA technique is applied since the strong penetration of the neutrons as well as the resulting emitted γ-radiation (either prompt during the nuclear reaction or delayed from the activation products) allows bulk analysis rather than surface analysis, the ''limiting'' factor in XRF spectrometry. For cases such as coal and ore analysis, these bulk analyses have been applied in industrial settings and for field-mineral explorations using 14 MeV neutron generators[13] and isotopic neutron sources.[14] The analyses are mainly focused on raw material analysis and for product control of one or a few major constituents. These procedures are, through calibration, customized for the problems they have been developed for and cannot be translated into a routinely applicable method for the analysis of a large variety of sample types. The same applies to the use of NAA principles in well-logging devices, in which the entire surrounding rock may be considered as a large sample.

The advantage of reactor-based INAA in the analysis of large samples lies with the higher available neutron fluxes and activation cross sections, all leading to better sensitivities for trace elements than can be obtained with 14 MeV neutron generators and isotopic neutron sources and providing the opportunity for full multielement analysis.

LSNAA has evolved over the years[15–20] toward a capability for analysis of various samples types which otherwise would have been difficult to analyze. Both off-line (LSNAA) and on-line (LS-prompt γ-NAA) activation methods are used. The emphasis in this article is on (off-line) LSNAA; prompt γ-LSNAA is discussed briefly below.

2.2 Large Sample Prompt γ-Activation Analysis

Large sample prompt γ-neutron activation analysis (PGNAA) is being applied for many years in well logging, industrial (conveyor) belt analyzers, etc. using isotopic neutron sources such as ^{252}Cf or ^{241}Am(Be). The advantage of isotopic source-based PGNAA compared to normal NAA lies in the fact that the test portion may be analyzed locally rather than having to be taken to the laboratory and on-line information is obtained. Since the output of the sources is rather low, large samples are needed anyhow to obtain a measurable signal, usually from the main components in the material of interest. Industrial analyzers are commercially available for the on-line analysis of cement,[21] the determination of the sulfur content on coal,[22] for the detection of explosives in airline cargo,[23] etc.

Reactor-based large sample PGNAA, i.e. using an external neutron beam, was demonstrated by Sueki et al.[24] for a pottery sample of 15 cm diameter, 10 cm width, and 0.5 cm wall thicknesses. The neutron beam dimensions were approximately 2 cm $\times$ 2 cm.

Similar to ''normal'' LSNAA, in large sample PGNAA also the problems of neutron attenuation and γ-ray self-shielding have to be solved. In the example quoted above, the internal monostandard was used (see Section 5.3). Also, other intact archaeological objects were analyzed by this method, such as bronzes.[25] Moreover, neutron beams from reactors are relatively limited in dimensions (on the order of 5 cm $\times$ 3 cm), which sets also a limit to the size of the object activated. This limitation can be overcome by repositioning the sample in the beam.

An advantage of large sample PGNAA over normal LSNAA is that no special facilities have to be constructed in the reactor, and that the sample contains hardly any induced radioactivity, which is of importance when dealing with, for example, archaeological or cultural artifacts. The PGNAA setup can be standard,[24] but care has to be taken that the large object does not ''transform'' into a very intense source of prompt γ-radiation with associated radiation dose hazards for the researchers.

Other methods for standardization have been proposed too, and mostly are based on a priori available information on the (gross) composition of the object; e.g. using Monte Carlo simulations[26] or neutron transport codes[27] (''fixed point iteration method''). Degenaar developed a method in which no a priori information is used and the neutron self-shielding is estimated on basis of the attenuation and scattering of the neutron beam, measured outside the sample.[28]

2.3 Large Sample Photon Activation Analysis

Photon activation analysis has the potential to analyze very large samples for reasons similar to NAA: large penetration power of the incident bremsstrahlung photons (typically in the order of several tens of million electronvolts), and similar to NAA, large penetration of the γ-radiation from the induced radioactivity. There is also some similarity to prompt γ-NAA with respect to the size of the object that can be exposed at a time; here also, the sample can be ''moved'' through the beam to attain a homogeneous activation, or the beam can be scanned over the sample. It introduces an additional complication if the integral sample is counted after exposure: the different activated parts have different decay times but their signals are registered simultaneously. Alternatively, one may choose to limit the sample size.

At the Bundesanstalt fur Materialprufung (BAM) in Berlin, Germany, large sample photon activation analysis has been applied[29] using the 30 MeV linear accelerator for studies involving samples with sizes in the order of 8 cm height and approximately 2 cm thickness with masses of up to 100–200 g. The measurements were done using a twin detector set up; i.e. the sample was ''sandwiched'' between two side-looking semiconductor detectors.

One of the advantages of photon activation analysis over NAA is that the corrections for self-attenuation of the incoming bremsstrahlung photons are relatively easy to be applied on the basis of fluence rate monitors positioned before and after the sample.[29] Moreover, given the high energy of the photons, this attenuation is mostly relevant for thick targets with high average atomic number.

Large sample photon activation analysis has many interesting aspects and advantages compared to LSNAA, including its capability to detect elements such as C, N, and O, as well as Tl, Bi, and Pb. However, the number of photon activation analysis laboratories worldwide is very small and most of the large sample activation analysis studies are done with neutrons. For these reasons, this type of large sample analysis is not further elaborated upon in this article, and the reader is directed to the available literature.[29]

3 MEASUREMENT EQUATION OF LARGE SAMPLE NEUTRON ACTIVATION ANALYSIS

The basic measurement equations of NAA by which the mass of the unknown element is calculated directly demonstrates the fact that the technique does not set a priori constraints to the mass of the sample analyzed:

$$m_{\mathrm{unk}} = m_{\mathrm{std}} \frac{(A_{0,\mathrm{x}})}{(A_{\mathrm{s}})} R_\theta R_\phi R_{\mathrm{En}} R_\sigma R_{\mathrm{nss}} R_\varepsilon R_{\gamma\mathrm{ss}} R_{\mathrm{inh}} \quad (1)$$

$$A_0 = \Phi_{\mathrm{th}} \sigma_{\mathrm{eff}} \frac{N_{\mathrm{Av}} \theta m}{M} (1 - \mathrm{e}^{-\lambda t_{\mathrm{ir}}}) \mathrm{e}^{-\lambda t_{\mathrm{d}}} \frac{(1 - \mathrm{e}^{-\lambda t_{\mathrm{m}}})}{\lambda} \gamma \varepsilon \quad (2)$$

in which the subscripts ''unk'' and ''std'' refer to unknown and standard, respectively, and

A_0 = the area of the relevant peak in the γ-ray spectrum, corrected for differences in decay and measurement time between the unknown (x) and the standard (s);
R_θ = ratio of isotopic abundance of the element of interest in test portion and standard (often = 1);
R_ϕ = ratio of thermal neutron fluence rates in test portion and standard;
R_{En} = ratio of neutron energy distribution in test portion and standard;
R_σ = ratio of effective activation cross sections for the test portion and standard at the different neutron energy spectra;
R_{nss} = ratio of the neutron self-shielding in test portion and standard;
R_ε = ratio of the photopeak efficiency for the test portion and standard;
$R_{\gamma\mathrm{ss}}$ = ratio of the γ-ray self-attenuation in test portion and standard; and
R_{inh} = ratio of the effect of extreme inhomogeneities in test portion and standard.

Also,
Φ_{th} is the thermal neutron fluence rate ($\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$),
σ_{eff} is the effective absorption cross section (cm^2),
N_{Av} is the Avogadro's number (mol^{-1}),
θ is the isotopic abundance,
m is the mass of the irradiated element (g),
M is the atomic mass number ($\mathrm{g\,mol}^{-1}$),
λ is the decay constant of the radioisotope formed (s^{-1}),
t_{ir} is the irradiation duration (s),
t_{d} is the decay time (s),
t_{m} is the (live time) measuring time (s),
γ is the abundance in the nuclear decay of the γ-ray measured, and
ε is the full energy photopeak efficiency of the detector for the energy of the γ-ray measured.

Many of the correction terms, R_i, can often be neglected in normal sample analysis but some of them like R_{nss}, $R_{\gamma ss}$, R_{ε}, R_{En}, and R_{inh} become significant in large sample analysis. As such, algorithms in large sample NAA differ from normal NAA by the calculation/estimation of

- the neutron self-shielding and/or neutron fluence rate profile inside the test portion,
- the γ-ray self-attenuation,
- the volumetric photopeak source efficiency of the detector, and
- the impact of extreme inhomogeneity effects.

There are many approaches for these calculations, varying from pure theoretical modeling,[9] Monte Carlo modeling,[20] and modeling using a priori available information about the test portion composition[7] to pragmatic empirical estimations of the correction factors.[17] Modeling may even be avoided when, e.g., for routine applications a representative well-characterized (large sample) standard – even a reference material – is available.[18] These standardization methods are further discussed below.

4 INSTRUMENTATION

4.1 Neutron Sources for Large Sample Activation Analysis

The type and strength of the neutron source and energy characteristics play an important role in any type of NAA including LSNAA, as the radioactivity produced is directly proportional to the neutron flux (φ) and energy-dependent neutron absorption cross section (σ). The neutron source should provide a sufficiently high neutron fluence rate so as to keep the product of neutron fluence rate and large test portion mass almost equal to that in small test portion NAA. This criterion indicates that for test portions with masses in the order of 2 kg a neutron fluence rate of approximately $5 \times 10^{12} \times 0.2/2000 = 5 \times 10^{8}\,cm^{-2}\,s^{-1}$ would result in an adequate induced radioactivity during the irradiation time, similar to that applied in conventional NAA in which a 200 mg test portion is processed. Fluence rates on the order of 10^{8}–$10^{10}\,cm^{-2}\,s^{-1}$ are found at an extended distance from the core of small and medium-sized reactors,[15] in beam tubes, and in thermal columns (TCs).[5,19,20] However, low fluence rates can also be realized – or even may be preferred – by lowering the reactor power because of fuel economy considerations.[30] Table 2 also provides an indication of typical neutron sources available to provide the required neutron fluence rate.

The advantage of reactor TCs above, e.g., poolside facilities, is that the longitudinal neutron flux gradient (i.e. horizontally away from the reactor core) over the sample is much less steep because of the multiple neutron scattering in the graphite inside the TC, as can also be derived from the differences in thermal neutron diffusion length in carbon and water, viz., 64.2 and 2.76 cm, respectively. Rotating the sample along its vertical axis compensates partly for these gradients. However, in some materials the neutron self-attenuation combined with the neutron flux gradient may result in situations in which the center of the test portion is hardly activated compared to the periphery. In such cases, the measured γ-rays originate mainly from the periphery and the analysis result will apply merely to the outermost layers of the sample rather than reflecting the bulk composition. In such facilities, the approach is to set limits to the dimensions of the test portion. Moreover, neutron self-moderation will occur owing to the less thermalized neutron spectrum in poolside facilities, resulting in difficulties in the calculation of the element amounts.

Table 2 Typical neutron fluence rates for activation analysis of test portions of different mass along with neutron sources

Sample mass (g)	n-flux ($cm^{-2}\,s^{-1}$)	Neutron source
1–10	2.10^{12}–2.10^{11}	Reactor
10–100	2.10^{11}–2.10^{10}	Reactor
100–1000	2.10^{10}–2.10^{9}	Reactor
>1 kg	$<2.10^{9}$	Reactor, Isotopic n-source n-generator

(Mass × flux) = constant = 200 mg × $10^{13}\,cm^{-2}\,s^{-1}$

Neutron fluence rates $<10^{10}\,cm^{-2}\,s^{-1}$ may also be attainable with isotopic neutron sources and high-intensity neutron generators, and in (reactor-based) external neutron beams. ^{252}Cf is probably the most attractive isotopic neutron source from the point of view of the neutron spectrum shape and easy thermalization aspects. However, the short half-life (2.64 years) may be seen as an economical disadvantage. Other isotopic neutron sources have a relatively hard neutron spectrum, resulting in relatively low thermal neutron fluence rate equivalents. The applicability of such neutron sources may, therefore, be limited to the determination of the major components in a sample.

Neutron generators (3–14 MeV) have their own scope of applications.[31,32] One of the problems in using neutron generators and isotopic neutron sources is that the neutron fluence is anisotropic and therefore the neutron flux seen by a large sample is not the same in all parts of the sample. This can be overcome to some extent by rotating the test portion during irradiation. Given the fact that the thermal neutron fluence rate in, e.g., a D-D

(deuterium ions on a deuterium target) generator is of the order of 10^4 lower than in research reactors, an increase in sample mass of at least 10^4 (i.e. from 200 mg to 2 kg) would be needed to compensate for this low thermal fluence rate.

An advantage of large (kilogram size) sample NAA with D-D generators over reactors is that the sample may be quickly removed from its irradiation position upon shutting down the accelerator, facilitating the measurement of radionuclides with short half-lives. For samples in the order of tens of grams, pneumatic facilities may be designed. Larger sample masses may also be considered for delayed neutron counting procedures to reach lower minimum detectable amounts, although this implies that a larger delayed neutron counter is needed too.

However, several technological obstacles exist both with isotopic neutron sources as with neuron generators, such as the large void needed within the moderator of the device. Moreover, since these moderators are often based on hydrogenous materials such as polyethylene or paraffin, steep flux gradients may occur over the sample, similar to those outlined earlier for poolside reactor facilities. The large sample approach may also be considered for irradiation with the sub-fast neutrons, although here an increase in sample mass from, e.g., 200 mg to 200 g might be sufficient for reaching the sensitivity required. However, an additional problem is that large sample masses will increase the effect of neutron self-moderation.

Most of the external neutron beams from nuclear reactor, with neutron fluence rates of 10^6–$10^8\,cm^{-2}\,s^{-1}$, are suitable for PGNAA. An advantage associated with PGNAA is the flexibility in choosing the mass and shape of the test portion. However, one should be careful in increasing the test portion mass, as it might adversely affect the measurements because the background is sample dependent in PGNAA.[33] It becomes severe particularly in the cases where the hydrogen or boron fraction in the large test portion is high, as it results in an extremely intense source of prompt γ-radiation, which will affect the results.[34] External neutron beams of isotopic neutron sources have usually neutron fluence rates of $<10^7\,cm^{-2}\,s^{-1}$ and can in principle be used for PGNAA, though the energy definition becomes a tedious problem. Steep neutron gradients over the sample occur in neutron beam activation analysis too.

4.2 Irradiation Facility

Transferring the test portions to the irradiation position of the neutron source is usually done with pneumatic/hydraulic transfer facilities and/or manually or automatic loading facilities. Most of the pneumatic facilities are designed to handle volumes up to 5–50 mL, which are placed in a sample carrier known as ''rabbit''. Use of rabbit systems places constraints on the sample shape so as to maintain the defined geometry. In principle, it is possible to transfer test portions up to 1 kg or more through such facilities – such big systems already exist for transferring documents in offices and banks; however, it is yet to be explored and examined whether large rabbits can be obtained with the required specifications (quality of the rabbit materials, purity, and radiation/mechanical resistance) for application in reactors.

In some reactors TCs are available for accommodating a large sample irradiation facility (see Figures 1 and 2). Samples are placed in the irradiation position of the TC by the mechanical movement of a tray that houses the sample in a defined position.[5,19,20] In Table 3, suggestions for facilities for irradiation of test portions of different masses are indicated.

There are various design aspects to be taken into account for irradiation facilities:

1. A large-volume facility near the core of a nuclear reactor creates a void in the reactor's reflector, whereas loading and unloading may cause unwanted fluctuations in the core's reactivity. Moreover, a high amount of[41] Ar will be produced from activation of the air in the container.
2. The thermal neutron fluence rate gradient in the water reflector of a light-water-moderated reactor is quite steep, typically by a factor of 3 per each 3 cm, which is due to the neutron diffusion length (2.84 cm) in water. Such a strong gradient would also create an unwanted strong flux variation over the large test portion to be activated. This may be corrected for by rotating the sample during counting,[5] by mixing the sample after irradiation,[17] or by the use of *in situ* relative efficiency method.[19] Mixing, however, eliminates information about inhomogeneities. The problem of heterogeneity may also be addressed by dividing the large test portion into many smaller fractions, to be processed individually followed later on by combining of the results.
3. Large hydrogen mass fractions may result in neutron spectrum changes over the test portion volume due to self-thermalization. This phenomenon is difficult to correct for mathematically, and may be an additional reason to consider an irradiation facility with well-thermalized neutron spectrum, for instance, to be realized in a TC.
4. Large sample activation facilities at isotopic neutron sources must be designed in such a way that adequate shielding is ensured against the prompt γ-rays, which will be several orders of magnitude higher than with normal small samples.

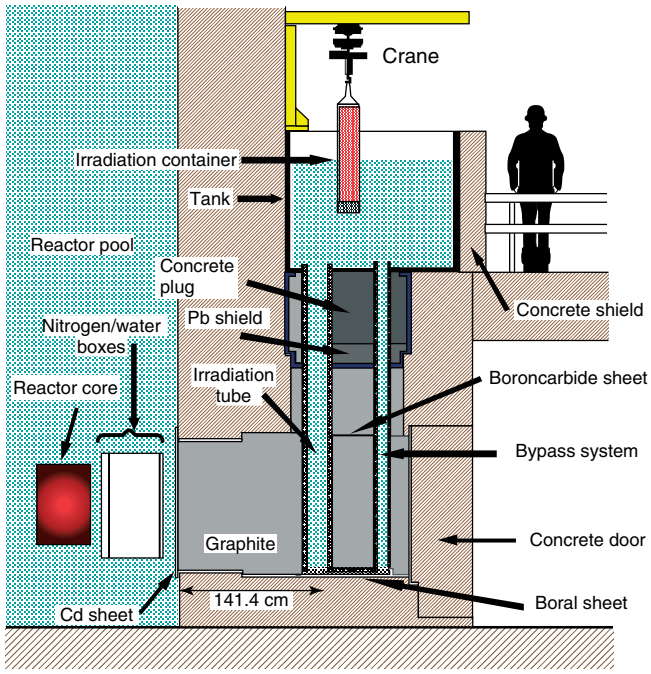

(a)

(b)

Figure 1 Schematic drawing (a) and picture of model (b) of the large sample neutron activation analysis facility in the thermal column of the Hoger Onderwijs Reactor in Delft.

A few of the large sample NAA reactor facilities that are currently operational are given in Table 4.[35] Large sample PGNAA facilities have been realized in Hungary and Japan. Isotope neutron source-based large sample PGNAA facilities are in use in some places.[36]

4.3 Sample Containers for Irradiation

The large sample container itself may be of any shape and type. A wide-neck bottle is easy to fill when coarse material has to be analyzed (Figure 3).[5] A container of inexpensive plastic may be preferred, as the impurities in the plastic itself (blank contribution) may be neglected at a given sample size (see Table 5). If the contribution from the sample holder is substantial, the irradiated large samples may also be easily transferred after irradiation into nonirradiated containers and possible small losses during transfer can be neglected in view of the large mass of the test portion. As such, a Marinelli beaker geometry or multisample container (Figure 4)[37] may

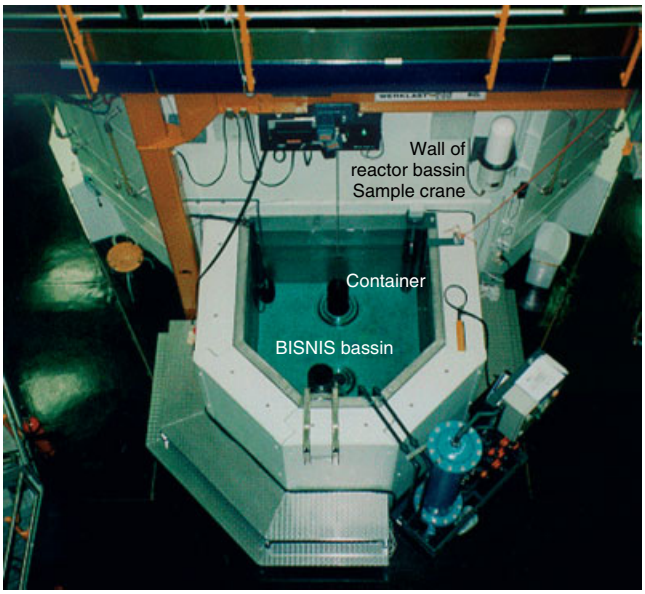

Figure 2 Top view of the large sample activation analysis facility in Delft.

have certain advantages, depending on the application (see below).

4.4 Counting Facility

Very large Ge detectors are available (crystal sizes up 400–800 cm^3, comparable to ''relative efficiencies'' of 100–200%). Such big detectors are an additional tool to maintain adequate sensitivity in NAA. Side-looking detectors (''horizontal dipstick'') have the advantage that cylindrical samples, positioned perpendicular to the detector axis, can easily be rotated around the sample axis to reduce geometrical effects. Vertical dipstick detectors have the advantage to measure large samples in the Marinelli beaker geometry. Well-type Ge detectors can handle test portion volumes of up to approximately 8 mL, and thus have their own niche in large sample NAA, especially as an addition to enhance sensitivity for test portions in the grams range.

In general, it is preferable to count the large sample by placing it at a certain distance from the detector end cap to minimize complications in the efficiency calculations, in particular in the coincidence summing corrections. The distance between the test portion and detector is guided by the sample activity; the higher the activity, the farther the sample to be placed from the detector. Automatic sample changing can be realized irrespective of the sample size. Sample changers are already commercially available for containers with volumes up to 1 L, although they have been designed for the Marinelli beaker measurement.

Table 3 Irradiation facilities needed for test portions of different masses

	Test portion mass (g)	Type of facility
1	1–10	Existing pneumatic facility
2	10–100	Existing/special pneumatic facility
3	100–1000	Existing manual loaded or special facility such as the thermal column of a reactor
4	>1000	Often a new special facility

Table 4 Details of some of the irradiation facilities used for LSNAA[37]

Institute	Nation	Reactor type	Test portion mass	Facility (TC = thermal column)	Neutron fluence rate ($cm^{-2}\,s^{-1}$)
Dalhousie University	Canada	SLOWPOKE	30 g	Rabbit system	2.5×10^{11}
International Centre for Environmental and Nuclear Sciences University of the West Indies	Jamaica	SLOWPOKE	30 g	Rabbit system	2.5×10^{11}
Atominstitut	Wien, Austria	TRIGA	5 g	Fast and normal Rabbit system	2×10^{12}
FRG-II	Munich, Germany	TRIGA	1 kg	Manual loading	6×10^{9}
Delft University of Technology, Reactor Institute Delft	Netherlands	Swimming pool reactor	50 kg	Manual loading, TC	3×10^{8}
BARC, Mumbai	India	Swimming pool reactor, Apsara	1–4 kg	Manual loading, TC	2×10^{8}
Demokritos	Greece	Swimming pool reactor	2 kg	Manual loading, TC	5×10^{6}
Institute of Nuclear Physics	Kazakhstan	Pool type reactor	10 × 100 mL	Manual loading, core	Low power operation

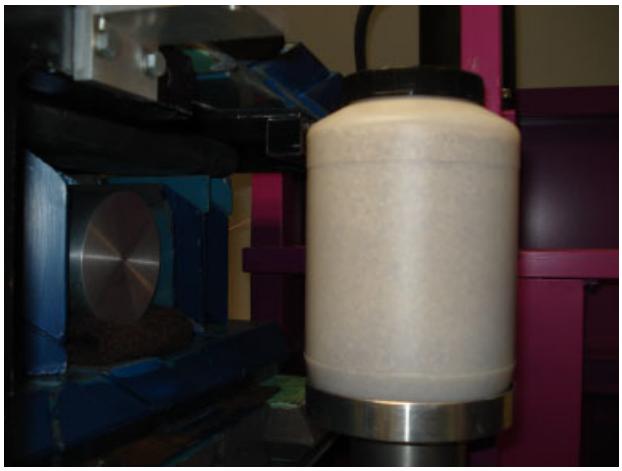

Figure 3 Large sample container (diameter 12 cm, height 20 cm) in front of side-looking Ge detector (cryostat diameter 9 cm).

Table 5 Typical masses of test portions and required bottles, indicating the reduction of the ratio of masses in the case of large sample NAA

Bottle volume (mL)	Bottle mass	Sample mass	Sample mass/ bottle mass
0.5	0.2 g	0.2–0.4 g	1–2
50	10	30–60	3–6
100	15	60–120	4–8
250	25	150–300	6–12
500	37	300–500	8–15
1000	100	600–1000	6–10

Adequate shielding of the stored samples remains, of course, a prerequisite.

Different detector calibration approaches are needed – taking into account voluminous photo peak efficiency, γ-ray self-attenuation, and coincidence summing correction if Marinelli beaker geometries or multisample containers are applied. In the multisample container setup, as the geometry for each container is the same, one can start with one container and measure up to 20 samples according to their decay.

The spectrometer may have to be equipped with a separate device to allow the determination of the effective γ-ray linear attenuation coefficients using a multi-γ-ray pencil beam of an external source (e.g. ^{152}Eu). An example of such a setup is given in Figure 5.

It may be necessary to collimate the detector in case the localization of the inhomogeneity is the subject of interest in large sample analysis (see Figure 6). Such a setup will allow, in principle, for emission tomography of the activated sample.[38]

More advanced spectrometer systems may be designed, in which two or more Ge detectors surround the large sample to create nearly 4π geometry of the detectors.[29]

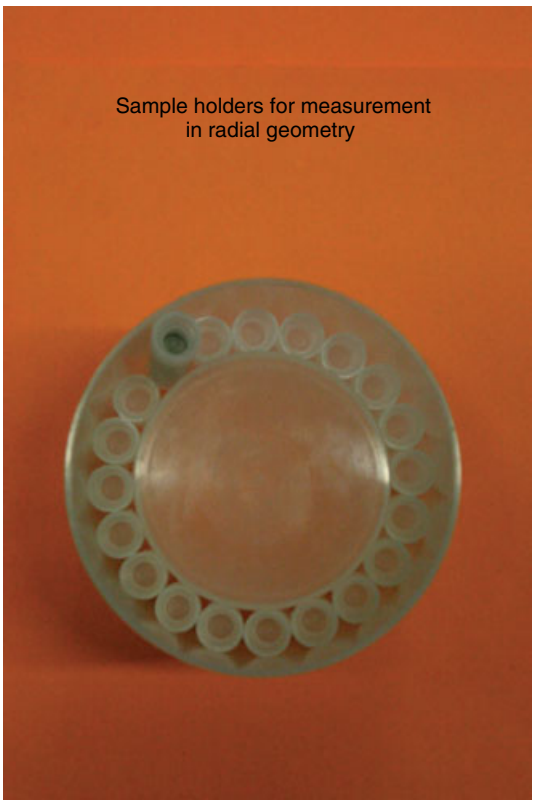

Figure 4 A multisample container.[37]

The individual spectra can be added later to create a composite spectrum with better statistics.

The γ-ray spectrometer should be equipped with dedicated high-count-rate electronics if the induced radioactivity to be measured would result in a count rate $>20\,000\,s^{-1}$. The prerequisites are a transit–reset preamplifier and pulse processing electronics (analog or digital), allowing for on-line dead-time compensation, e.g., based on the loss-free counting principle.[39] The latter is relevant only if count rates vary significantly during the counting time, as may be the case in counting radionuclides with very short half-lives (in the range of seconds to several minutes).

It is interesting to take note that there are a few applications where counting could be carried out using well-type scintillation detectors. In the cases where major components only have to be analyzed and the neutron sources are isotope based, then the resulting activity could be low and the γ-ray spectra might be relatively simple, which could be measured using high-efficiency scintillation detectors. Large sample NAA with isotopic sources and the use of scintillation detectors may be considered as an additional opportunity to further enhance the sensitivity. One may even construct a simple 4π detector by putting two well-type detectors against each other.

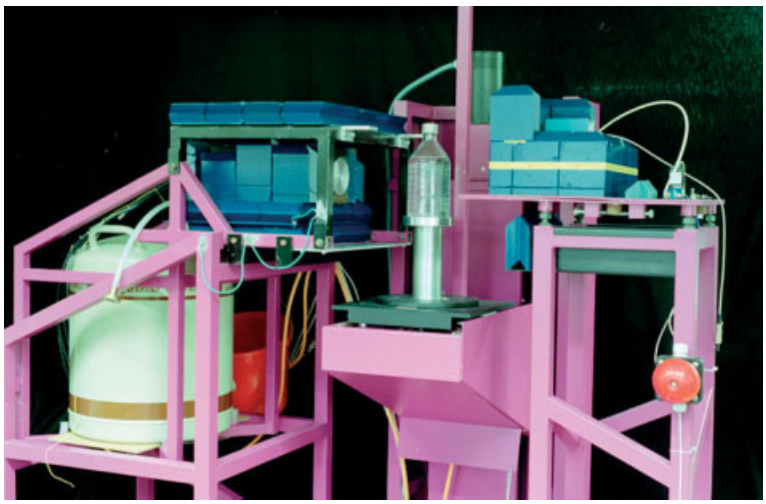

Figure 5 Large sample γ-ray spectrometer with (from left to right) the shielded side-looking Ge detector, the sample on the rotating turntable, and the lead shield with the ^{152}Eu source for γ-ray transmission. The source can be pneumatically moved in front of a point-source collimator.[1]

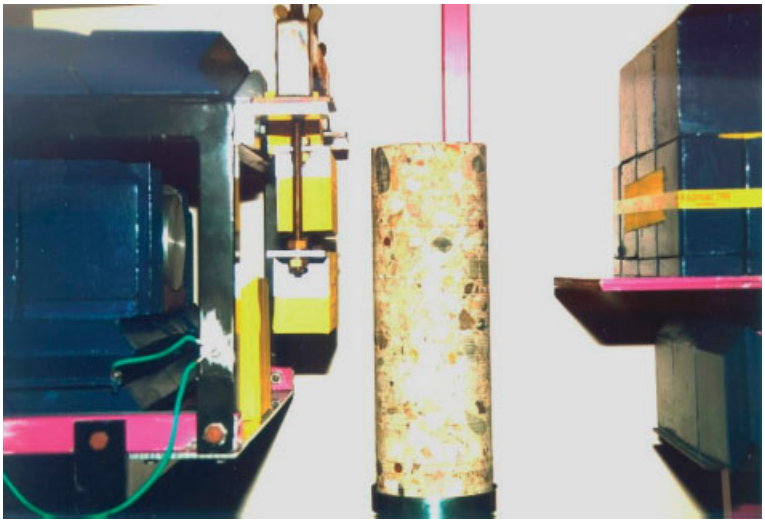

Figure 6 Close-up view of the γ-ray spectrometer showing (from left to right) the shielded Ge detector, a slit collimator, a (simulated) large sample, and the lead shield for the transmission source.[1]

5 CALIBRATION

Determination of the elemental masses in large sample NAA may be done as in small sample NAA, via

- the absolute method,
- the comparator method, or
- the internal standard method.

5.1 Absolute Method

The absolute method for standardization in NAA is based on using known values for neutron fluence rate and activation cross sections, derived either from previously performed measurements or from reactor physics estimations (neutron fluence rate) and from the literature data (cross sections); the same is the case with θ, N_{Av}, M, σ_{eff}, γ, and λ. For many (n,γ) reactions and radionuclides, the parameters σ_{eff}, γ, and λ are not precisely known, while in some cases θ also is not known accurately. Since the various parameters are often achieved via independent methods, their individual imprecisions will add up in the calculation of the elemental amounts, leading to large systematic errors. This method is best applicable if the composition of the sample matrix is well established in advanced, as is the case, e.g., when dealing with pure materials. It is well known that even in normal NAA this approach may not yield highly accurate data but still the results may be adequate

for the intended purpose. Still, additional estimates are needed for correcting the neutron fluence rate gradient and the γ-ray self-absorption. The first may be done using simplified models or using neutron transport codes, and the second after simple transmission measurements or using tabulated linear attenuation coefficients in case the sample is well defined with respect to its composition.

5.2 The Comparator Method

The test portion is irradiated together with a calibration sample containing a known amount of the element(s) of interest. The calibration sample is measured under (preferably) the same conditions as the sample (sample-to-detector distance, sample size, and if possible composition). From a comparison of the net peak areas in the two measured spectra, the mass(es) of the element(s) of interest can be calculated (see above, Equations (1) and (2)):

$$mass_{\mathrm{unk}} = mass_{\mathrm{std}} \frac{(A_{0,\mathrm{x}})}{(A_{\mathrm{s}})} R_\theta R_\phi R_{\mathrm{En}} R_\sigma R_{\mathrm{nss}} R_\varepsilon R_{\gamma\mathrm{ss}} R_{\mathrm{inh}} \qquad (3)$$

The relative standardization based on element standards is not immediately suitable for laboratories aiming at the full multielement powers of INAA. It is virtually impossible to produce a multielement standard containing known amounts of all 70 detectable elements with sufficient accuracy in a volume closely matching the size and the shape of the samples. For this reason, some laboratories prefer to use (certified) reference materials as a multielement standard. However, if dealing with large samples (gram to kilogram size), the use of (certified) reference materials is not practical (major differences may exist in neutron spectrum characteristics in the real sample and in the standard), economical (reference materials are expensive and large sample analysis would imply a high consumption rate), or ethical (certified reference materials are produced for method validation purposes, not necessarily for calibration).

Multielement INAA based on the relative standardization method is feasible when performed according to the principles of the single comparator method. Assuming that all relevant experimental conditions are stable over time, standards for all elements are co-irradiated, each in turn with the chosen single comparator element. Once the sensitivity for all elements relative to the comparator element has been determined (expressed as the so-called k-factor, see Section 5.3), only the comparator element has to be used in routine measurements instead of individual standards for each element.

5.3 Single Comparator Method

Originally, the single comparator method for multielement INAA was based on the ratio of proportionality factors of the element of interest and of the comparator element after correcting for saturation, decay, counting, and sample weights. Girardi et al.[40] defined the k-factor as

$$k = \frac{M_{\mathrm{t}} \gamma_{\mathrm{c}} \varepsilon_{\mathrm{c}} \theta_{\mathrm{c}} \sigma_{\mathrm{eff,c}}}{M_{\mathrm{c}} \gamma_{\mathrm{t}} \varepsilon_{\mathrm{t}} \theta_{\mathrm{t}} \sigma_{\mathrm{eff,t}}} \qquad (4)$$

in which the subscripts "t" and "c" refer to the element of interest in the sample and comparator, respectively. Mass fractions can then be calculated from these k-factors; for an element determined via a directly produced radionuclide, the mass fraction ρ follows from

$$\rho = \frac{(\mathrm{A/SDCw})_{\mathrm{t}}}{(\mathrm{A/SDCw})_{\mathrm{c}}} \cdot k \qquad (5)$$

where $S = (1 - \mathrm{e}^{-\lambda . t\mathrm{ir}})$

These experimentally determined k-factors are often more accurate than those calculated on basis of literature data as in the absolute standardization method. However, the k-factors are valid only for a specific detector, a specific counting geometry, and the irradiation facility and remain valid only as long as the neutron flux parameters of the irradiation facility remain stable.

The main problem of the single comparator method in LSNAA is that differences in neutron exposure, γ-ray attenuation, and volumetric counting efficiency between comparators and samples all have to be accounted for. In normal NAA, most of these differences can be neglected, also because the comparators are co-irradiated with the samples. But in LSNAA this may be practically impossible; the irradiation facility may not be spacious enough, and substantial difference may exist in the neutron exposure and flux gradients. Only if the neutron flux spectrum is well known – as in TCs – and the neutron fluence rate gradient can be established in each individual test portion, the comparator method provides an opportunity for applicability. In that case, even the k_0-based method for standardization may be applied.

The comparator method is, however, very well usable for large sample analysis of, e.g., large liquid samples such as water or oil since standard samples with element spikes into a similar matrix can easily be prepared.

5.4 k_0-Based Method for Standardization

At the Institute for Nuclear Sciences in Ghent, Belgium, an attempt has been made to define k-factors that should be independent of neutron flux parameters as well as spectrometer characteristics.

The expression for the activation reaction rate can written as

$$R = \Phi_{th}\sigma_0 + \Phi_{epi} I_0(\alpha) \tag{6}$$

The ratio f of the thermal neutron flux and the epithermal neutron flux is $f = \Phi_{th}/\Phi_{epi}$ and the ratio of the resonance integral and the thermal activation cross section can be expressed as $Q_0(\alpha) = I_0(\alpha)/\sigma_0$; thus the effective cross section is

$$\sigma_{eff} = \sigma_0 \left(1 + \frac{Q_0(\alpha)}{f}\right) \tag{7}$$

The k_0-factor is now defined as

$$k_0 = \frac{1}{k}\frac{1 + (Q_{0,c}(\alpha)/f)}{1 + (Q_{0,t}(\alpha)/f)}\frac{\varepsilon_c}{\varepsilon_t} = \frac{M_c\theta_t\sigma_{0,t}\gamma_t}{M_t\theta_c\sigma_{0,c}\gamma_c} \tag{8}$$

and the mass fraction, again for an element determined via a directly produced radionuclide, is found from

$$\rho = \frac{1 + (Q_{0,c}(\alpha)/f)}{1 + (Q_{0,t}(\alpha)/f)}\frac{\varepsilon_c}{\varepsilon_t}\frac{(\text{A/SDCw})_t}{(\text{A/SDCw})_c}\frac{1}{k_0} \tag{9}$$

The k_0-factor has thus become a purely nuclear parameter for the thermal neutron spectrum. In the k_0 convention, Au is proposed as the comparator element. The neutron flux parameters f and α no longer cancel out in concentration calculations and must be measured in each irradiation facility, preferably even for each irradiation and sample.[41] The k_0-factors are used in Delft for the analysis of very large samples.

5.5 Internal Monostandard Method

In the internal monostandard method, one of the radionuclides produced during activation of the test portion is used as monostandard. The rationale behind this is that the effect of neutron spectrum perturbation is the same for this parent element of this radionuclide as well as for all other elements in the sample; as such there is an implicit assumption that the test portion is ''macroscopically'' homogeneous.

In the case of internal monostandard method using TC irradiations followed by γ-spectrometric method, the ratio of mass (m) of an element (x) in the test portion (t) to mass of the internal comparator element (c) in the sample is given by Equation (10):

$$\frac{m_t}{m_c} = \frac{\left((\text{S.D.C}) \cdot (f + Q_o(\alpha))\right)_c}{\left((\text{S.D.C}) \cdot (f + Q_o(\alpha))\right)_t} \cdot \frac{A_{At}}{A_{Ac}} \cdot \frac{\left(\varepsilon_\gamma\right)_t}{\left(\varepsilon_\gamma\right)_c} \cdot \frac{1}{k_{0,c}(x)} \tag{10}$$

Here, $k_{0,c}(x)$ represents the relative sensitivity of element x with respect to y and is calculated from the $k_{0,Au}$ factors in the literature.

The internal monostandard method either results in elemental mass ratios (element of interest vs. monostandard element) and thus may serve for comparative studies, or, in the case of, e.g., materials of high purity and known stoichiometry, directly into mass fractions of the elements of interest if (one of) the major component(s) is used as the monostandard element.

5.6 Neutron Fluence Rate Monitoring

Neutron fluence rate monitoring is needed for the k_0 method of standardization and may be done noninvasively with the flux monitors outside the sample, as well as by inserting flux monitors inside the sample. The first approach may use, e.g., the neutron depression outside the large sample to estimate the neutron flux distribution;[42] the second approach provides direct information on the flux distribution.[43,44] The first approach is applied in Delft for large sample irradiations in the reactor's TC.[5] During the irradiation, the sample is surrounded by four flux monitors at any desired height around the sample. Since the unperturbed neutron flux gradient in the TC can be derived from the irradiation of a pure graphite sample, the neutron flux depression outside the sample can be estimated (Figure 7). This forms the basis for estimating the effective neutron diffusion length and neutron diffusion coefficient. Finally, the overall correction factor is calculated – which reflects how the large sample compares to a small sample – using the volume efficiency of the Ge

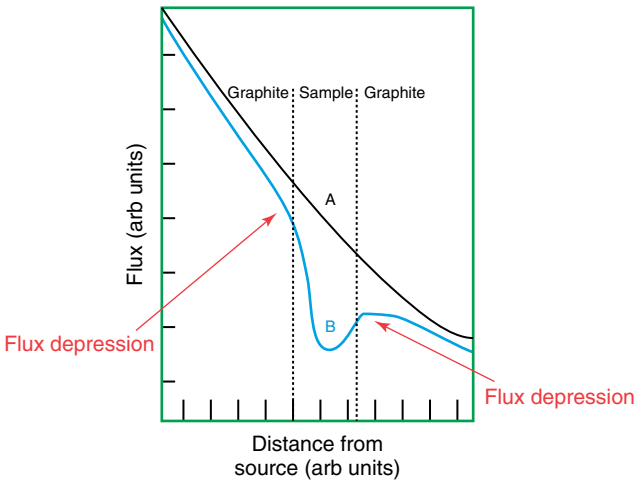

Figure 7 Neutron flux depression outside a large sample placed inside the graphite-filled thermal column. The curve A indicates the unperturbed flux gradient; curve B schematically shows the flux gradient if a sample absorbing and scattering thermal neutrons is placed in the irradiation position.

detector, the neutron diffusion length and coefficient, and the effective linear attenuation coefficients.

Changes in the neutron spectrum due to self-thermalization are much more difficult to deal with, as these are not easy to monitor. The extent of this effect depends, of course, on the neutron spectrum shape and the fraction of epithermal and fast neutrons compared to the thermal neutrons. In TC facilities, the ratio of thermal over nonthermal neutrons may be much larger than a factor of 1000, eliminating the significance of neutron self-thermalization. The user of the irradiation facility should be familiar with this phenomenon, and a priori information must always be collected about the sample composition so as to estimate the extent of these effects and to decide if empirical correction factors can be applied or if additional in situ monitoring is needed (i.e. invasive, by inserting suitable monitors inside the sample).

5.7 γ-Ray Self-Attenuation

The γ-ray self attenuation correction is relatively easy to establish once the effective linear attenuation coefficients are available, either by measurement or by calculation from the approximate (or well-known) elemental composition. A multi-γ-ray emitting source, with γ-ray energies distributed over the entire range of interest (such as ^{152}Eu, ^{182}Ta) can be used for this. A nearly pencil beam geometry can be created by locating this source behind a pinhole collimator, and the γ-ray transmission can be measured at several heights along the sample. This forms the basis for the estimation of the effective linear γ-ray attenuation coefficients.

The volumetric photopeak efficiency can be determined by Monte Carlo modeling, but this requires precise information about the inner geometry of the cryostat and detector configuration (including the dead-layer thickness). Empirical curves may be determined using standard sources in water, as the self-attenuation can simply be subtracted from the measured efficiency. In situ relative detection efficiency in a voluminous sample was determined using the multi-γ-emitters produced in the sample, and was adequate to calculate mass ratios with respect to the comparator using Equation (5).[45]

Pragmatic approaches have been suggested in which the large sample, after activation, is repacked into many small-diameter containers that are placed in a cylindrical holder surrounding the detector. If the detector crystal is perfectly symmetrically mounted inside the cryostat, the detection efficiency for each of the positions around the detector is the same, which simplifies the calculations. Besides, the dimensions of the small containers can be chosen such that γ-ray attenuation effects may be neglected.

It should be noted that the γ-ray spectrum due to the natural radioactivity of the sample material has also to be measured in large sample analysis, prior to the activation. These "sample background" peaks in the γ-ray spectrum should be separately treated later on in the neutron and γ-ray self-attenuation corrections.

5.8 Extreme Inhomogeneities

Combination of correction algorithms for neutron and γ-ray self-attenuation as well as for the volumetric photopeak efficiency yields an "overall correction factor", which reflects the difference between the actual detector response for a given γ-ray energy and the theoretical detector response if the sample were a massless point source located in the large sample's center, without any neutron and γ-ray attenuation.[9]

In these corrections, it is assumed implicitly that both trace elements and major (matrix) elements in the sample are homogeneously distributed on a macroscopic scale. If this condition is not met, there is a high probability that owing to the neglect of inhomogeneities the concentrations determined are not correct. How large these deviations may be as a result of such neglect has been studied via computer simulations. Inhomogeneities may influence the results of the irradiation as well as of the measurement; therefore, both have been treated separately. Inhomogeneous matrix composition has been modeled by composing a sample of cylinders with strongly differing neutron or γ-ray attenuation properties. Inhomogeneity for trace elements has been simulated by modeling extreme distributions of a trace element with either neutron or γ-ray attenuation properties, strongly differing from those of the main composition of the sample.[20,46] Both inhomogeneities in matrix composition (e.g., layered structures) and trace-element inhomogeneities (e.g., "hot spots") were taken into account.

Obviously, the results of these simulations demonstrated that false concentrations may be obtained if inhomogeneities are not accounted for in the interpretation step of large sample INAA. The smallest errors may occur for matrix inhomogeneities; the most pronounced effects can be expected when the trace element of interest is distributed either on the outside or on the axis of the cylindrical sample.

In these simulations, materials or elements were selected with neutron and γ-ray attenuation properties that were strongly different from the rest of the sample; as such they may be considered as "extremities", and the consequences on the inaccuracy of the results rather indicate "worst case" conditions.

5.8.1 *Determination of Inhomogeneities*

The presence of extreme inhomogeneities in large samples may be considered a nuisance. On the other hand, large sample analysis is a unique tool for determining these inhomogeneities without destroying the test portion. To this end, sample scanning using a collimated detector has been introduced, and it is assumed that the sample consists of volume elements that individually are considered homogeneous.

The set of spectra constituting one scan is statistically evaluated to determine whether fluctuations over the scan of the count rates of γ-rays of a given energy are only due to Poisson counting statistics or are also due to inhomogeneities in the sample. If inhomogeneities have only a layered structure in the direction of the cylinder axis of the sample, the sample can be analyzed layer by layer, and for the most extreme cases the analysis can be performed for each voxel separately. An example of a collimated detector system for large sample scanning is shown in Figure 6.

Baas et al.[38,47] developed a method for the detection of local inhomogeneities. The count rate in each channel of each segmented measurement can therefore be compared with the average count rate in each channel after summing all individual measurements. Such a comparison, taking into account uncertainties, is made analogous to the ζ-score principle. Values of $|\zeta| > 2$ or > 3 (depending on the analyst's fitness-for-purpose criterion) indicate a local inhomogeneity at the respective γ-ray energy in a particular voxel. This approach is visualized in Figure 8. A bottle of approximately 25 cm length and 8 cm diameter filled with soil was irradiated in the large sample facility in Delft. The induced radioactivity was measured with a 96% Ge detector, collimated with a 10-cm-thick Pb collimator with a 2-cm split opening. The figure shows the z' scores for each channel number (γ-ray energy) along the height of the sample. From the zoomed details it can clearly be seen that strongly deviating ζ-scores occur at, e.g., 439 and 1115 keV, indicating an inhomogeneity for zinc. In addition, the histogram of all ζ-scores in this figure provides also an insight into the presence of this inhomogeneity.[47]

6 QUALITY CONTROL

The high degree of accuracy in normal activation analysis results from decades of experience in the development of certified reference materials. Many sources of error and the quantification of their impact are known.[48,49] Methods commonly referred to as *quality control practices* have been developed to inspect the occurrence of errors during the analysis, whereas implementation of quality assurance contributes to minimizing and even avoiding the occurrence of errors. The known sources of error in normal activation analysis may occur in large sample analysis too. Some of them – such as γ-ray self-attenuation and neutron/photon fluence rate, or neutron spectrum gradients – have much larger effects. Extreme inhomogeneities are an additional phenomenon in large sample analysis,[46] with an impact on the degree of accuracy.

Quality control in normal activation analysis includes the simultaneous analysis of well-characterized quality

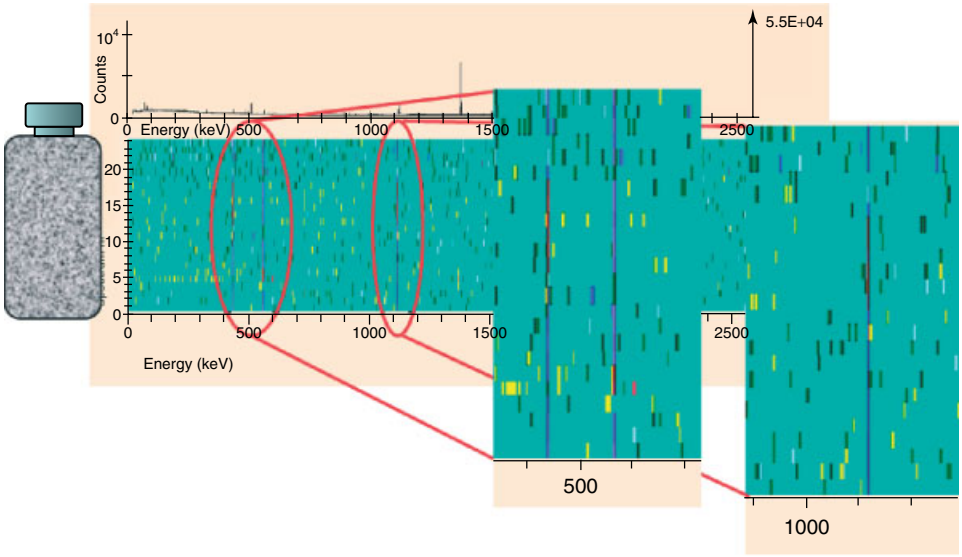

Figure 8 Energy- and position-correlated ζ-scores (see text) of measured radioactivity, indicating location of inhomogeneities in a sample of 20 cm height and 10 cm diameter.

Table 6 Opportunities of quality control measures traditionally applied in normal activation analysis for samples of larger sizes

	Quality control samples	Blanks	Duplicates	γ-Ray intensity ratios and multiple radio nuclides
1 g	Y	Y	Y	Y
10 g	N	Y/l.r.	Y/n.r.	Y
100 g	N	l.r.	n.r.	Y
1 kg	N	l.r.	n.r.	Y

Y = yes, application possible; N = no, not possible; l.r. = less relevant (see text); n.r. = not relevant.

control samples, blanks, and sometimes duplicates. In addition, inspection of the intensity ratios of γ-ray peaks of a given nuclide and/or the quantified results based on different radionuclides formed from a given element also provide a unique opportunity to inspect for errors. The applicability of these quality control approaches for samples of increasing mass is given in Table 6.

It is clear that basic problems emerge when extending the traditional approaches to samples with weights of more than a few grams. Firstly, well-characterized control samples of the size of large samples (several grams to kilograms) are either very expensive to use or not available at all. Secondly, large sample analysis may be required because of the heterogeneity of the object, which cannot be simulated by a control sample even if it were available. Thirdly, duplicates – assuming identical composition in mass fraction and in degree of homogeneity – may probably not be available when larger sample masses are needed. The problem related to the blank – impurities in the sample container and/or contamination – has, on the contrary, a smaller impact on the final result since the increase in the ratio of sample mass to container mass may result in negligible contribution of the blank (Table 5).

New strategies have to be developed to control the analytical quality in large sample analysis. One of the opportunities is to continue with the use of performance indicators, derived from the actual sample analyzed. In fact, this is not different from most quality control procedures in, e.g., manufacturing and production, in which the quality of a final product depends on predefined specifications being met, such as dimensions, tolerances, mass, color, or operation characteristics. The inspection of γ-ray intensity ratios and the use of different radionuclides of one element are already examples of such a form of quality control in activation analysis. This approach can further be extended to other sample/material characteristics on the basis of physical sample properties such as γ-ray self-attenuation and neutron attenuation parameters, as well as via the degree of inhomogeneity (Table 3).

6.1 Quality Control in Large Sample Analysis

Some materials may be difficult to be processed to such homogeneity that representative subsamples can be taken at the <1 g level. For such materials, it may be advantageous to analyze much larger quantities without homogenization and to assume that the inhomogeneities are randomly distributed throughout the sample, so that the entire quantity can be considered as homogeneous. However, this assumption has some limitations. Overwater and Bode demonstrated the impact of extreme inhomogeneities on the correction mechanisms for the attenuation of γ-ray attenuation and neutron self-shielding.[46] Inhomogeneities with strong γ-ray absorbing properties have stronger effects on the degree of accuracy than those with strong neutron absorbing properties. It is therefore relevant to inspect for the presence of such extreme inhomogeneities in order to decide on the value of the finally calculated mass fractions. Two opportunities to inspect such inhomogeneities are given here.

The effective linear γ-ray attenuation coefficient is usually determined by measuring the transmission of γ-rays of different energies emitted by an external source with known emission rates.[9,45] The values of the effective linear attenuation coefficients may be estimated using the tabulated values for the elements. Typically, for example, biological and geological material values of $\sim 0.15\,\mathrm{cm}^{-1} < \mu < \sim 0.60\,\mathrm{cm}^{-1}$ at ~ 100 keV; $\sim 0.12\,\mathrm{cm}^{-1} < \mu < \sim 0.25\,\mathrm{cm}^{-1}$ at ~ 300 keV; and $\sim 0.05\,\mathrm{cm}^{-1} < \mu < \sim 0.15\,\mathrm{cm}^{-1}$ at ~ 1000 keV can be found. A ''bandwidth'' of the linear attenuation coefficient can thus be determined at different γ-ray energies for different types of materials (e.g. environmental, geological) (Figure 9). This can assist in inspecting whether the experimentally determined attenuation coefficients have realistic values. Moreover, if scanned measurements are carried out, an indication of local (layer-type) inhomogeneities with strong γ-ray absorbing properties may already be obtained. An example of this approach is given in Figure 10. The transmission of the γ-rays of a ^{152}Eu source was measured at different heights along a $\sim 1-$m long, $\sim 15-$cm diameter soil drill core sample prior to neutron activation. The effective linear attenuation coefficients for each of the γ-ray energies were fitted with a polynomial so as to estimate the effective linear γ-ray attenuation at other γ-ray energies. The attenuation coefficients at 100, 300, and 1000 keV all fall within the expected bandwidths. There are no indications in this example for layer inhomogeneities with strongly differing γ-ray absorbing properties.

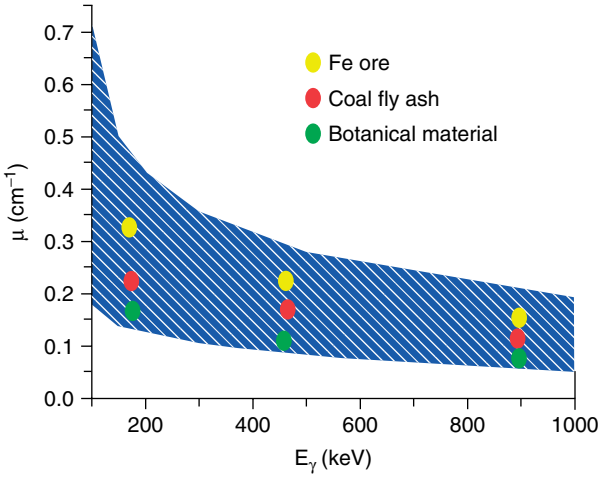

Figure 9 Range of linear γ-ray attenuation coefficients for typical materials.[50]

Additionally, the methods developed by Baas et al. (described in the preceding text) can be used for the detection of local inhomogeneities. It can now be decided on a case-to-case basis whether such inhomogeneities have any unwanted impact on the final analysis result.

The correction for neutron self-shielding in LSNAA may be made using information derived from the neutron fluence rate depression at positions in the irradiation facility just outside the sample. Overwater and Hoogenboom[42] developed this approach to estimate the thermal neutron diffusion length L_s and the thermal neutron diffusion coefficient D_s, which subsequently were used to reconstruct the neutron fluence rate profile inside the large sample. Both L_s and D_s are physical element properties and, similar to the effective linear γ-ray attenuation coefficient, boundaries can be estimated for the values of these two parameters in real materials. Though strongly correlated ($L_s^2 = D_s/\Sigma_a$, with Σ_a the macroscopic absorption cross section (cm^{-1})), typical values are $\sim 1\,cm < L_s < \sim 20\,cm$ and $\sim 0.2\,cm < D_s < \sim 3\,cm$. An example is given in Figure 4. A ~1-m long, ~15-cm diameter water basin sediment drill core was analyzed in the frame of a pollution research project. Zinc foils were used as neutron flux monitors. The monitors were positioned just outside the sample container to monitor neutron fluence rate depression by comparison with the neutron fluence rates as monitored in a separate irradiation with a solid graphite sample, thus simulating the unperturbed neutron flux. The calculated values of L_s and D_s at different heights are plotted within the bandwidths for these values (Figure 11). Also, the average neutron fluence rates as a function of sample height is given. The fluence rates can be fitted with a cosine function, reflecting the flux distribution within the reactor's TC.

Assuring the quality of the results requires insight, monitoring, and control of the sources of error. Quality control procedures as traditionally applied in chemical analysis are not fully applicable in large sample analysis.[50] One of the advantages of activation analysis is that some of the measured sample parameters dealing with γ-ray and neutron attenuation can only vary in ranges set by well-known values of elemental constants. These parameters can be much more easily determined in large sample analysis than with samples in the milligram range, thereby offering an outlook for direct verification of the quality of the related correction algorithms.

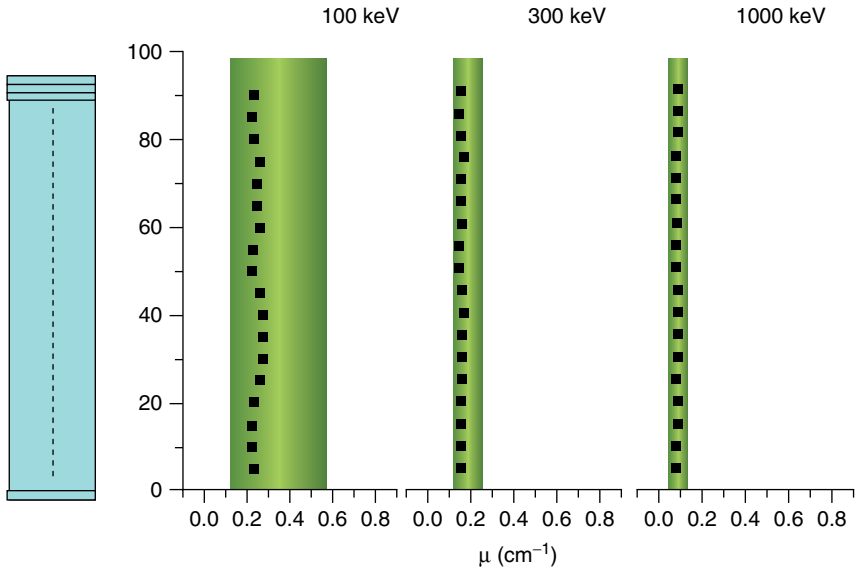

Figure 10 Measured linear attenuation coefficients at 100, 300 and 1000 keV of a 15 cm thick soil drill core.

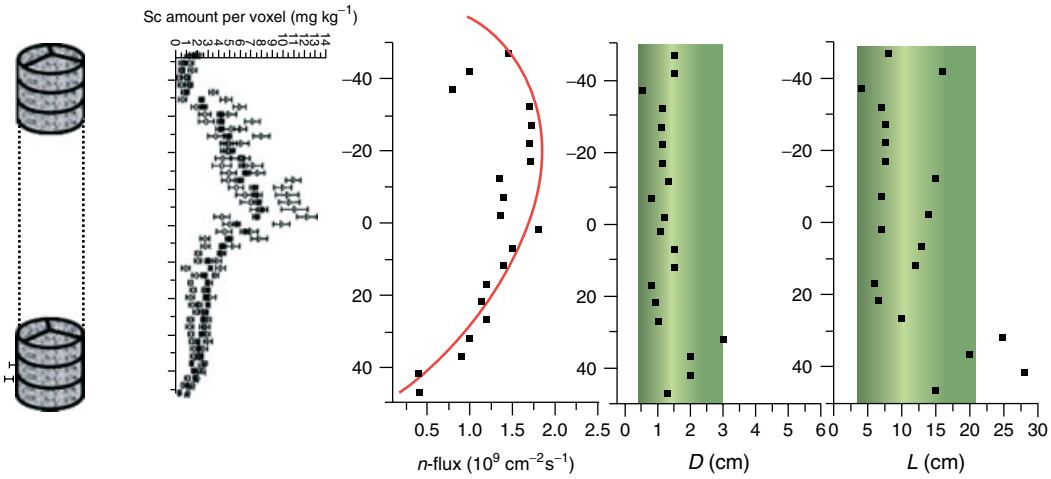

Figure 11 Results of collimated large sample analysis of a stack of soil samples, total of 1 m length, 12 cm diameter, showing (left to right) Sc amount per 120° voxel, neutron flux gradient variations in neutron diffusion coefficient and neutron diffusion length; the band widths indicate the theoretical boundaries of these parameters.[50]

7 SENSITIVITY

Sensitivity is defined as the gradient of the response curve: i.e., the change in instrument response that corresponds to a change in analyte concentration. This definition translates in INAA into the net peak area as a function of the analyte mass. Larger peak areas at a given sample mass can be obtained by the following:

- higher neutron fluxes and longer irradiation times
- more efficient detectors
- larger sample masses.

Higher neutron fluxes and longer irradiation times are often not easily attainable, as the first is limited by reactor design, whereas longer irradiation times only have a positive effect on the sensitivity for radionuclides with very long half-lives.

Absolute photopeak efficiencies of detectors of different sizes and for different geometries are given in Table 7, and a comparison of the absolute efficiencies of a regular 17% coaxial and one of the largest well-type detectors reported in the literature[51,52] is given in Figure 12. It should be noted that the use of Compton suppression shields does not result in an increase in sensitivity – as often erroneously suggested; after all, the signal resulting from the induced radioactivity does not increase and sometimes even decreases due to summing-out effects or larger sample-detector distances. Compton suppression systems find their advantage in a decrease in uncertainty of measurement due to decrease of the background under a peak.

Large sample masses can compensate for low neutron fluxes. As the limiting factor in NAA is merely the maximum acceptable induced radioactivity upon

Table 7 Absolute photopeak efficiencies at some photon energies for 3 detector types and different geometries[51,52]

Absolute photopeak efficiency (%)					
		1332 keV	750 keV	500 keV	250 keV
20%	0 cm	1.5	2.5	3.2	5
	5 cm	0.22	0.35	0.5	1
100%	0 cm	7.5	10	12	16
	5 cm	1	1.4	1.7	2.4
125 cm^3	Well	5	9	13	25

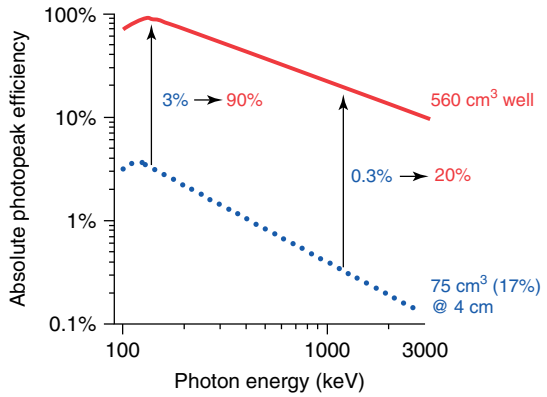

Figure 12 Schematic absolute full energy photopeak efficiency curves for a regular coaxial Ge detector (17% relative efficiency) and a very large well-type detector.

counting, a comparison has been made with normal INAA in which a hypothetical 200-mg sample is irradiated for a given time at a neutron flux of 10^{13} cm^{-2} s^{-1}. Assuming a certain neutron fluence rate and a cylindrical sample of similar length and diameter, and an average density of 0.5 g cm^{-3}, a first estimate can be made of the minimum

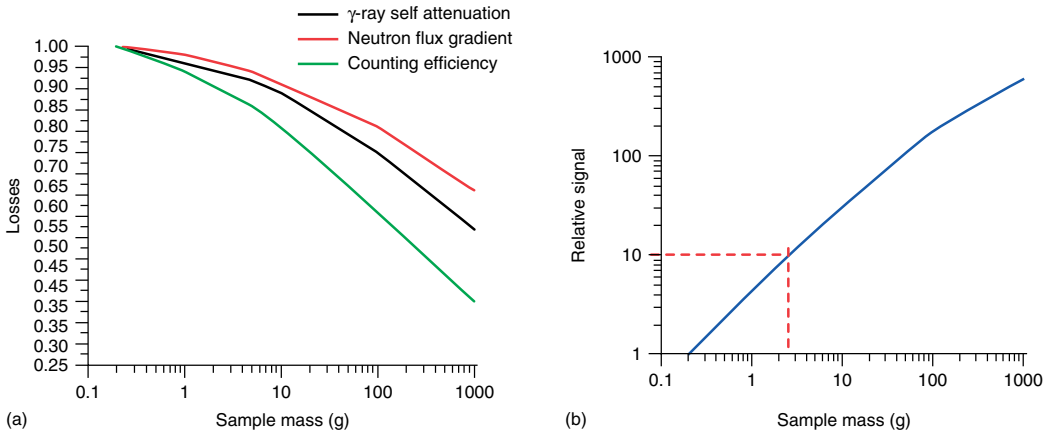

Figure 13 (a) Losses due to neutron and γ-ray self-shielding and enlarged source-to-detector distance, compared to an analysis of a 200-mg powdered siliceous sample and (b) net gain in signal at increased sample masses.

Table 8 Indication of minimum sample mass at a given neutron fluence rate to attain similar induced radioactivity as a 200-mg sample, irradiated at 10^{13} cm^{-2} s^{-1}

Neutron fluence rate (cm^{-2} s^{-1})	Sample size (L, D) (cm)	Sample mass (g)
2.10^{12}–2.10^{11}	1–2	1–10
2.10^{11}–2.10^{10}	2–4	10–100
2.10^{10}–2.10^{9}	4–7	100–1000
$<2.10^{9}$	>7	>1 kg

mass needed to reach the same sensitivity as for the small sample at a high neutron fluence rate (Table 8).[35]

However, these results must be corrected for the losses due to neutron self-shielding and γ-ray self-attenuation, and the fact that the center of voluminous samples, simply for physical reasons, is always positioned further away from the detector than in case of a small sample. In Figure 13 these effects have been combined, and it can be derived that at a given neutron fluence rate a net gain in signal of a factor of 10 can be obtained by increasing the sample mass by a factor of 15 (e.g., from 200 mg to 3 g).

7.1 Natural Background

A large sample NAA procedure should start with measurement of the natural radioactivity of the sample, as the corresponding peaks in the γ-ray spectrum should be separately treated later on in the neutron and γ-ray self-attenuation corrections.

8 APPLICATIONS

LSNAA has evolved over the years towards a capability for various samples types that otherwise would have been difficult to analyze. Accordingly, the developments faced many challenges, each one different from the other, and in this process its horizon of application to various fields has increased: e.g., nutritional studies, geology, biology, archaeology, waste analysis, high-purity material characterization, precious samples, and liquid samples, from all walks of life. A few of them are given in the following text.

8.1 Materials Difficult to Homogenize: Geological Samples, Ores, and Waste

LSNAA is highly suitable for the analysis of heterogeneous geological material such as rocks, coal (determination of quality), ores, and mineral concentrates. Conveyor belt monitoring of elemental concentrations by PGNAA in coal and cement raw materials has lead to an increased efficiency of coal-fired power plants and cement factories. It appears that this technique has the required potential to trigger industrial processes and gives reliable results.

Waste material in many cases is considered to contain hazardous substances whose behavior could result in their entry into biosphere through the atmosphere or ground water. Therefore, appropriate classification of the waste material is required in order to ensure safe disposal or further treatment and recycling. Construction material, domestic and electronic waste, as well as contaminated sediment and compost material are considered to be highly heterogeneous and therefore elaborate sampling procedures are required if representative sampling of these materials for analysis is needed. LSNAA has been effectively used to analyze large samples of soil,[53] electronic waste,[54,55] and other materials. These are the materials in which both subsampling and homogenization, steps not required in LSNAA, are very difficult and cost intensive. A typical example is the elemental composition analysis of waste from an incineration plant. This type of

sample cannot be easily homogenized and analyzed by other analytical techniques. However, a 1-kg portion of this waste was analyzed by using the LSNAA without homogenization.

8.2 Materials That May be Contaminated During Homogenization: High-Purity Materials

LSNAA is an extremely useful technique to analyze the metals and alloys for impurities. It is being used to analyze various finished products of alloys (Zircaloy 2, Zircaloy 4, SS-316M (stainless steel), and 1S aluminum) that are used in reactor technology[56] and for impurities in high technological materials such as silicon and superalloys.[3] The biggest advantage is that information is obtained on the entire specimen, and because of the absence of subsampling, contamination can be minimized.

8.3 Materials Difficult to Subsample: Nutritional Studies

LSNAA is extremely useful in the determination of major, minor, and trace elements in foodstuff, as large samples can be analyzed without resorting to subsampling. In fact, comparison of the results obtained from subsamples of varying mass gave an indication that a sample of 1 kg of wheat is more representative than small samples in the range of 40–1000 mg.[57] It is also feasible to determine trace elements in liquid diets, e.g., juices and milk, by LSNAA.

8.4 Valuable Material of Irregular Shape

Subsampling of archaeological and cultural heritage objects is generally prohibited, as these objects have to be preserved intact. LSNAA has the capability for nondestructive bulk analysis of the whole object. In comparison, other established nondestructive analytical methods, such as XRF or analytical techniques based on charged particle irradiation (PIXE, particle induced X-ray emission, IBA, ion beam analysis), can only analyze superficial layers of the sample and provide limited information over the whole volume of the object of interest.[58,59]

The particular advantage of INAA being noninvasive and a true multielemental technique is combined in LSNAA with the ability to analyze bulky objects as a whole, without any visual damage to the valuable cultural heritage objects. Art historians, conservators, and museum staff do not generally allow damaging such valuable objects by removing a portion for analytical purposes. However, sometimes only the elemental composition can decisively distinguish whether an object is different from what it appears to be from visual inspection.

8.5 Other Applications

Direct analysis of large samples provides a unique opportunity for (the validation of) (sub)sampling studies.

REFERENCES

1. P. Bode, 'Instrumental and Organizational Aspects of a Neutron Activation Analysis Laboratory', Ph.D. thesis, Delft University of Technology, 1996, ISBN 90-73861-42-X.
2. B. Kratochvil, J.K. Taylor, 'Sampling for Chemical Analysis', *Anal. Chem.*, **54**, 924A (1981).
3. M.I. Verheijke, H.J.J. Jaspers, J.M.G. Hanssen, 'Neutron Activation of Very Pure Silicon Wafers', *J. Radioanal. Nucl. Chem.*, **131**, 197 (1989).
4. IUPAC, 'Terminology in Soil Sampling', *Pure Appl. Chem.*, **77**, 825 (2005).
5. P. Bode, R.M.W. Overwater, 'Trace Element Determination in Very Large Samples: A New Challenge for Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **167**, 169 (1993).
6. R.M.W. Overwater, P. Bode, J.J.M. de Goeij, 'Large-sample Neutron Activation Analysis: Present Status and Prospects', *J. Radioanal. Nucl. Chem.*, **216**, 5 (1997).
7. R. Gwozdz, F. Grass, 'Activation Analysis of Large Samples', *J. Radioanal. Nucl. Chem.*, **244**, 523 (2000).
8. X. Lin, R. Henkelmann, 'Instrumental Neutron Activation Analysis of Large Samples: A Pilot Experiment', *J. Radioanal. Nucl. Chem.*, **251**, 197 (2002).
9. R.M.W. Overwater, 'The Physics of Big Sample Instrumental Neutron Activation Analysis', PhD thesis, Delft University of Technology, 1994.
10. C.G. Clayton, A.M. Hassan, M.R. Wormald, 'Multi-element Analysis of Coal During Borehole Logging by Measurement of Prompt γ-Rays from Thermal Neutron Capture', *Int. J. Appl. Radiat. Isot.*, **34**, 83 (1983).
11. G. Vourvopoulos, 'Industrial on Line Bulk Analysis Using Nuclear Techniques', *Nucl. Instrum. Methods*, **B56**, 917 (1991).
12. D.M. Franklin, R. Amstrong, D.R. Chettle, M.C. Scott, 'An Improved in vivo Neutron Activation System for Measuring Kidney Cadmium', *Phys. Med. Biol.*, **35**, 1397 (1990).
13. C.S. Lim, B.D. Sowerby, 'On Line Bulk Elemental Analysis in the Resource Industries Using Neutron-gamma Techniques', *J. Radioanal. Nucl. Chem.*, **264**, 15 (2005).
14. R. Doczi, M.A. Ali, M. Fayez-Hassan, J. Csikai, 'Determination of Hydrogen Content in Bulk Samples Using Neutron Activation Analysis', *Appl. Radiat. Isot.*, **63**, 137 (2005).

15. P.A. Beeley, R.G. Garrett, 'Neutron Activation Analysis of Multiple Large Geological Samples by in-pool Irradiation Using a Slowpoke-2 Reactor', *J. Radioanal. Nucl. Chem.*, **167**, 177 (1993).

16. P. Bode, R.M.W. Overwater, J.J.M. de Goeij, 'Large Sample Neutron Activation Analysis: Status and Prospects', *J. Radioanal. Nucl. Chem.*, **216**, 5 (1997).

17. R. Gwozdz, H.J. Hansen, K.L. Rasmussen, H. Kunzendorf, 'Instrumental Neutron Activation Analysis of Samples with Masses from Micrograms to Hectograms', *J. Radioanal. Nucl. Chem.*, **167**, 161 (1993).

18. J.I. Kim, H. Staerk, J. Fiedler, 'A Method of Long-time Irradiation of a Voluminous Liquid Sample in a Reactor Neutron Flux for Activation Analysis of Water', *Nucl. Instrum. Methods*, **177**, 557 (1980).

19. A.G.C. Nair, R. Acharya, K. Sudarshan, S. Gangotra, A.V.R. Reddy, S.B. Manohar, A. Goswami, 'Development of an Internal Monostandard INAA Method Based on in-situ Detection Efficiency for Analysis of Large and Non-standard Geometry Samples', *Anal. Chem.*, **75**, 4868 (2003).

20. F. Tzika, I. Stamatelatos, J. Kalef-Ezra, 'Neutron Activation Analysis of Large Volume Samples: the Influence of Inhomogeneity', *J. Radioanal. Nucl. Chem.*, **271**, 233 (2007).

21. M. Borsaru, R.J. Holmes, P.J. Mathew, 'Bulk Analysis Using Nuclear Techniques', *Int. J. Appl. Radiat. Isot.*, **34**, 397 (1983).

22. M.R. Wormald, C.G. Clayton, 'In-situ Analysis of Coal by Measurement of Neutron-induced Prompt Gamma-rays', *Int. J. Appl. Radiat. Isot.*, **34**, 71–82 (1983).

23. A.T. Farsoni, S.A. Mireshgi, 'Design and Evaluation of a TNA Explosive-detection System to Screen Carry-on luggage', *J. Radioanal. Nucl. Chem.*, **248**, 695 (2001).

24. K. Sueiki, K. Kobayashi, W. Sato, H. Nakahara, T. Tomizawa, 'Non Destructive Determination of Major Elements in Large Sample by Prompt Gamma Ray Neutron Activation Analysis', *Anal. Chem.*, **68**, 2203 (1996).

25. K. Sueki, Y. Oura, W. Sato, H. Nakahara, T. Tomizawa, 'Analysis of Archaelogical Samples by the Internal Monostandard Method of PGAA', *J. Radioanal. Nucl. Chem.*, **234**, 27 (1998).

26. I.H. Degenaar, M. Blaauw, 'The Neutron Energy Distribution to Use in Monte Carlo Modeling of Neutron Capture in Thermal Neutron Beams', *Nucl. Instrum. Methods*, **B 207**, 131 (2003).

27. J.H. Holloway, H. Akkurt, 'The Fixed Point Formulation for Large Sample PGNAA', *Nucl. Instrum. Methods*, **A522**, 529 (2004).

28. I.H. Degenaar, 'Towards a Methodology for Large Sample Prompt Gamma Neutron Activation Analysis', Ph.D. thesis, Delft University of Technology, 2004, ISBN 90-407-2509-8.

29. W. Gorner, S. Alber, A. Berger, O. Haase, G. Manse, C. Segebade, 'BEAMGAA A Chance for High Precision Analysis of Big Samples', *J. Radioanal. Nucl. Chem.*, **263**, 791 (2005).

30. R. Gwodz, 'Instrumental Neutron Activation Analysis of Samples with Volumes from 2 to 350 mL', *J. Radioanal. Nucl. Chem.*, **271**, 751 (2007).

31. S.S. Nargolwalla, E.P. Przybylowicz, *Activation Analysis with Neutron Generators*, John Wiley & Sons, New York, 1973.

32. K.G. Broadhead, D.E. Shanks, 'The Application of 2.8 MeV (D,D) Neutrons to Activation Analysis', *Int. J. Appl. Radiat. Isot.*, **18**, 279 (1967).

33. C.J. Evans, S.J.S. Ryde, D.A. Hancock, F. Al-Agel, 'Monte Carlo Simulation of Prompt Gamma Neutron Activation Analysis Using MCNP Code', *Int. J. Appl. Radiat. Isot.*, **49**, 541 (1998).

34. S.A. Natto, D.G. Lewis, S.J.S. Ryde, 'Benchmarking the MCNP Code for Monte Carlo Modeling of an in vivo Neutron Activation Analysis System', *Int. J. Appl. Radiat. Isot.*, **49**, 545 (1998).

35. 'Large Sample NAA Using Low Flux Irradiation Facilities', Report of a Technical Meeting, IAEA, Vienna, Austria, 7–11 November 2005.

36. R. Khelifi, Z. Idiri, P. Bode, A. Amokrane, 'Flux Calculation in LSNAA Using an Am-Be Source', *J. Radioanal. Nucl. Chem.*, **274**, 639 (2007).

37. R. Gwozdz, F. Grass, 'Application of a Radial Geometry for Counting Samples of Mass 10g, to 300g with a Ge-detector', *Environmental Radiochemical Analysis*, Special Publication No. 234, (G.W.A. Newton, Editor), Royal Society of Chemistry, Cambridge, UK, 289 1999.

38. H.W. Baas, M. Blaauw, P. Bode, J.J.M. de Goeij, 'Collimated Scanning Toward 3D-INAA of Inhomogeneous Large Samples', *Fresenius J. Anal. Chem.*, **363**, 753 (1999).

39. G.P. Westphal, 'On the Performance of Loss-Free Counting – A Method for Real-time Compensation of Dead-Time and Pileup Losses in Nuclear Pulse Spectroscopy', *Nucl. Instrum. Methods*, **163**, 189 (1979).

40. F. Girardi, G. Guzzi, J. Pauly, 'Reactor Neutron Activation Analysis by the Single Comparator Method', *Anal. Chem.*, **37**, 1085 (1965).

41. A. Simonits, F. de Corte, J. Hoste, 'Single Comparator Methods in Reactor Neutron Activation Analysis', *J. Radioanal. Chem.*, **24**, 31 (1975).

42. R.M.W. Overwater, J.E. Hoogenboom, 'Accounting for the Thermal Neutron Flux Depression in Voluminous Samples for Instrumental Neutron Activation Analysis', *Nucl. Sci. Eng.*, **117**, 141 (1994).

43. N.S. Shakir, R.E. Jervis, 'Correction Factors Required for Quantitative Large Volume Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **248**, 61 (2001).

44. K. Sudarshan, R. Tripathi, A.G.C. Nair, R. Acharya, A.V.R. Reddy, A. Goswami, 'A Simple Method for Correcting the Neutron Self-shielding Effect of Matrix and Improving the Analytical Response in PGNAA', *Anal. Chim. Acta*, **549**, 205 (2005).

45. R.M.W. Overwater, P. Bode, J.J.M. de Goeij, 'Gamma-ray Spectroscopy of Voluminous Sources: Corrections for Source Geometry and Self-attenuation', *Nucl. Instrum. Methods*, **A324**, 209 (1993).

46. R.M.W. Overwater, P. Bode, 'Computer Simulations of the Effects of Inhomogeneities on the Accuracy of Large Sample INAA', *Int. J. Appl. Radiat. Isot.*, **49**, 967 (1998).

47. H.W. Baas, 'Neutron Activation Analysis of Inhomogeneous Large Samples – An Explorative Study', Ph.D. Dissertation, Delft University of Technology, 2004.

48. P. Bode, E.A. De Nadai Fernandes, R.R. Greenberg, 'Metrology for Chemical Measurements: Purism, Pragmatism and the Position of INAA', *J. Radioanal. Nucl. Chem.*, **245**, 109 (2000).

49. R.M. Lindstrom, R. Zeisler, R.R. Greenberg, 'Accuracy and Uncertainty in Radioactivity Measurement for INAA', *J. Radioanal. Nucl. Chem.*, **271**, 311 (2007).

50. P. Bode, 'Quality Control in Large Sample Analysis', *J. Radioanal. Nucl. Chem.*, **271**, 333 (2007).

51. P. Bode, 'Detectors and Detection Limits in INAA. I. General Theoretical Relationships Between Detector Specifications and Detection Limits', *J. Radioanal. Nucl. Chem.*, **222**, 117 (1997).

52. P. Bode, 'Detectors and Detection Limits in INAA. II. Calculated and Observed Improvements in Detection Limits with Large Ge-detectors, Well-type Ge-detectors and Anti-Compton Spectrometers', *J. Radioanal. Nucl. Chem.*, **222**, 127 (1997).

53. M. Blaauw, H.W. Baas, M. Donze, 'Height-resolved Large-sample INAA of a 1 m long, 13 cm diameter Ditch-bottom Sample', *Nucl. Instrum. Methods*, **A505**, 512 (2003).

54. R.M.W. Overwater, P. Bode, J.J.M. de Goeij, J.E. Hoogenboom, 'Feasibility of Elemental Analysis of Kilogram-Size Samples by Instrumental Neutron Activation Analysis', *Anal. Chem.*, **68**, 341 (1996).

55. C. Segebade, P. Bode, W. Gorner, 'The Problem of Large Samples: An Activation Analysis Study of Electronic Waste Material', *J. Radioanal. Nucl. Chem.*, **271**, 261 (2007).

56. R. Acharya, A.G.C. Nair, A.V.R. Reddy, A. Goswami, 'Standard-less Analysis of Zircaloy Clad Samples by an Instrumental Neutron Activation Method', *J. Nucl. Mater.*, **326**, 80 (2004).

57. R. Acharya, A.G.C. Nair, K. Sudarshan, A.V.R. Reddy, A. Goswami, 'Development and Applications of the k_0-based Internal Mono Standard INAA Method', *Appl. Radiat. Isot.*, **65**, 164 (2007).

58. K. Janssens, R. van Grieken (eds.), *Comprehensive Analytical Chemistry XLII: Non-destructive microanalysis of cultural heritage materials*, Elsevier, 2005.

59. IAEA STI/DOC/010/416, 'Nuclear Analytical Techniques in Archaeological Investigations', Vienna, 2003.

Activity Determination and Localization of Radioactive Point Source in a Large Medium

Zeev B. Alfassi

Ben Gurion University, Beer Sheva, Israel Università degli Studi and INFN of Milano, Segrate

1 Introduction **1**
2 γ-Emitting Source **1**
2.1 Two Collinear Detectors 1
2.2 Four Detectors 5
3 Neutron Sources **10**
3.1 Two Neutron Detectors 11
3.2 More than Two Detectors 11
References **14**

The activity and localization of a small-sized radioactive source in a large medium was studied for various cases by the simultaneous use of several detectors. The γ-emitting source and the neutron-emitting source were both studied. This also led to an α-emitting source, which cannot be detected from outside of the medium unless the neutrons from their (α, n) reaction with low Z elements are detected. A special actual case of commercial lung counter with four semiplanar HPGe detectors is described in detail. For large boxes, NaI(Tl) detectors were used for γ measurements measuring only the full-energy peaks. For neutrons, only thermal neutrons were measured with ^{3}He detectors. To increase the fraction of the thermal neutrons and to protect the measurer, the sample was surrounded by a fence made of 5-cm polypropylene.

1 INTRODUCTION

Many measurements of radioactivity involve counting of γ rays or neutrons emitted by bulky samples, which are referred to as *a box*. While it is simple experimentally in the case of a point source in a known position, to transform the measurements of count-rate to activity (disintegration per unit time) and hence to the number of radioactive atoms, the problem can be the inaccuracy in the case of a small source in an unknown position in a bulky medium. The inaccuracy is because of two factors: (i) the self-absorption and scattering of the photons in the sample and (ii) the unknown solid angle in which the detector sees the point source. Thus, in the case of a radioactive small source in an unknown position in a large box, we need to find its position in order to transform the count-rate to activity (*see* **Instrumental Neutron Activation Analysis**).

The measurements are different if the source is emitting γ rays or neutrons. In the case of the γ-emitting source, measurements can be done only for the photons that do not interact with the media surrounding the source, except for the media of the detector, by measuring the full-energy peak. In the case of neutron emission, the measurement of the spectrum of the emission is more difficult. In addition, most of the emitted neutrons are fast neutrons, while most of the detectors are more sensitive to thermal neutrons; hence, for better statistics, the neutrons are measured after their interaction with the surrounding material.

The localization of the source can be done only by measurement with more than one detector. The simple case is when it is known that the source in on a given line. In this case, there are only two unknowns: the activity and the location on the line; and two detectors on parallel faces of the box, collinear with the source, are sufficient. However, in the general case, there are four unknowns: activity and the three coordinates in space; then, at least four detectors are needed.

2 γ-EMITTING SOURCE

2.1 Two Collinear Detectors

2.1.1 Nonabsorbing Media

Two factors cause the decrease in efficiency with the increase in distance from the detector: (i) the isotropic propagation of γ rays, which leads to a smaller solid angle in which the detector sees the source with the increasing distance, i.e. the 1/distance^2 law and (ii) the absorption of γ rays inside the sample, i.e. the $\exp(-\mu \cdot \text{distance})$ law. In order to separate these factors and dealing with only the first factor, a relative nonabsorbing media should be used. This can be done either with a low dense sample or, in the case of a low value of μ, the linear absorption coefficient, with high-energy γ rays.[1]

The mathematical treatment uses the assumption of the ''virtual point detector'' model for the voluminous γ detectors.[2–9] The bulky detector is treated as a virtual point detector on the symmetry axis of the detector at a distance d from the cap. This means that the count-rate in the absence of an absorbing/scattering medium due to a radioactive source at a distance y from the detector cap $C(y)$ is given by the equation $C(y) = \frac{C(0) \cdot d^2}{(y+d)^2}$, where d is

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

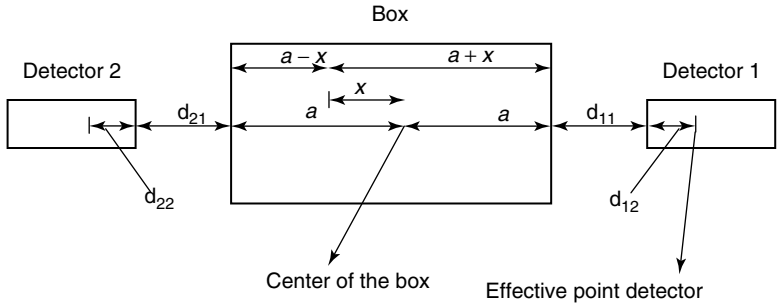

Figure 1 The dual detectors measuring system.

the distance from the detector cap to the virtual detector point, y is the distance from the source to the detector cap, and $C(0)$ is the count-rate measured on the detector cap.

The localization of the radioactive point source is done by two detectors on parallel faces of the box, collinear with the source, as shown in Figure 1. In Figure 1, d_{j1} is the distance between the inner face of the box and the detector j cap and d_{j2} is the distance between the detector cap and its effective virtual point detector; thus, if $d_j = d_{j1} + d_{j2}$, we get the equation for the count-rate because of a source at a distance y from the inner face of the box.

$$C_j(y) = \frac{C_j(0) \cdot {d_j}^2}{(y + d_j)^2} \quad j = 1, 2 \tag{1}$$

For a point source at the center of a box with a linear dimension $2a$ (on the line of centers), this equation leads to

$$C_j(\text{center}) = \frac{C_j(0) \cdot {d_j}^2}{(a + d_j)^2} \quad j = 1, 2 \tag{2}$$

For a point source located at a distance x from the center, the distances from the box faces are $a + x$ and $a - x$, and hence Equation (2) yields

$$C_1(y) = \frac{C_1(0) \cdot {d_1}^2}{(a + x + d_1)^2}; \quad C_2(y) = \frac{C_2(0) \cdot {d_2}^2}{(a - x + d_2)^2} \tag{3}$$

The normalized count-rate (normalized to the count-rate at the center of the box) $N_1(x)$ and $N_2(x)$ are

$$N_1(x) = \frac{(a + d_1)^2}{(a + x + d_1)^2}; \quad N_2(x) = \frac{(a + d_2)^2}{(a - x + d_2)^2} \tag{4}$$

Since both detectors are quite similar, we can accurately approximate $d_1 = d_2$. Equation (4) shows that, if we take the square roots of the reciprocal count-rate, the denominators are the same, while in the numerators, x appears once with a plus sign and once with a minus sign, which means that adding them cancels the terms with x:

$$\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} = 2 \tag{5}$$

This sum has the units $s^{0.5} \cdot \text{count}^{-0.5}$ and hence the square of its reciprocal value has the units counts per second, which are the units of count-rate. The definition of a new mean according to this equation, the harmonic mean of the square roots, leads to a mean that is independent of the position of the source.

$$M_{\text{SH}} = \left(\frac{2}{\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}}} \right)^2 \Rightarrow M_{\text{SH}} = \frac{4}{\left(\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} \right)^2}$$

$$M_{\text{SH}} = \frac{4N_1N_2}{(\sqrt{N_1} + \sqrt{N_2})^2} \tag{6}$$

Equation (5) indicates that M_{SH}, defined for the normalized counts, should be equal to 1, independent of x. It was found experimentally[1] that the deviation of M_{SH} from 1 is always below 10% and, in most cases, not more than 2–3%.

This constancy of M_{SH} indicates the method of measuring the activity of a point source accurately within a box without having to know the position of the source. As we do not know the counts with the source at the center of the box, M_{SH} is not calculated with the normalized count-rates, but with the experimental ones. In this way, M_{SH} is independent of the position of the source, but dependent proportionally on the activity of the source. The box is calibrated for its efficiency by a standard point source with known activity, which is denoted by Act(Std), positioned anywhere on the line of the centers. Let us assign its mean square harmonic M_{SH} by M_{SH}(Std). The activity of any unknown source Act(u) of the same energy, positioned

somewhere (unknown) on the line of centers, is given by

$$\mathrm{Act}(u) = \frac{M_{\mathrm{SH}}(u)}{M_{\mathrm{SH}}(\mathrm{Std})} \cdot \mathrm{Act}(\mathrm{Std}) \tag{7}$$

M_{SH} is calculated from the count-rates of the two detectors (Equation 6). As the unknown source cannot be removed for the measurement of the standard source, the counts due to the standard source for each detector are calculated by the subtraction of two counts; one with the unknown source alone and the other with both of the two, the unknown source and the standard one.

If the exact place of the source in the box is required, the position of the source can be found from the ratio of the count-rates of the two detectors. From Equation (5):

$$R = \sqrt{\frac{N_1}{N_2}} = \frac{[(d+a)+x]}{[(d+a)-x]} \tag{8}$$

and hence

$$x = (d+a) \cdot \frac{1-R}{1+R} \tag{9}$$

where x is the distance from the center of the box toward detector 2. Care should be taken to include d, besides the distance of the virtual point detector from the detector cap (which depends on the γ-ray energy) and also the thickness of the box and the distance of the outer face of the box from the detector cap.

2.1.1.1 Calculated Error for Longitudinal Sources The calculation till now assumes that the source is a very small one, close to a point. In the following, we calculate the error induced by a longitudinal source of length $2b$ lying on the line of centers with its midpoint coinciding with the midpoint of the line of centers (Figure 2).

The activity of an infinitesimal piece of the source with length dx is $A \cdot dx/2b$, where A is the total activity of the source. The normalized count-rates of the longitudinal source–detector (normalized to the whole activity being concentrated in the center) is given by the following integral

$$N = \frac{\dfrac{A}{2b}\displaystyle\int_0^{2b} \mathrm{d}x/(d+a-b+x)^2}{A/d+a}$$

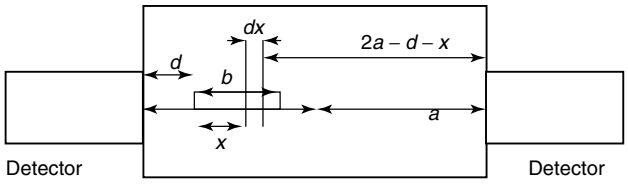

Figure 2 Longitudinal measuring geometry.

By integration, the normalized count-rates are

$$N = \frac{(d+a)^2}{(d+a)^2 - b^2} = \frac{1}{1-[b/(d+a)]^2} \tag{10}$$

If $b/(d+a) \ll 1$, N can be approximated as $N = 2b/(d+a)$; this is the fractional error because the source is not a point source. As long as $b/(d+a) < 0.3$, the error due to a longitudinal source (Equation 10) is less than 10%.

2.1.2 *Absorbing Media – Parallel γ Lines*

2.1.2.1 Activity Calculation The use of collimated detectors to measure only parallel photons, as, for example, in a medical double-headed γ camera, eliminates the decrease in the solid angle in which the detector sees the source with the increase in the source–detector distance and leaves the effect of the source–detector distance to the sole effect of absorption/scattering.[10] In this case, the geometric mean of the count-rate of the two detectors can be used to calculate the activity of the unknown source:

$$C_1 = C_{10} \cdot \mathrm{e}^{-\mu(a-x)}; \quad C_2 = C_{20} \cdot \mathrm{e}^{-\mu(a+x)} \tag{11}$$

From Equation (11), the geometric mean of the two count-rates measured by the two detectors is given by

$$\begin{aligned} M_{\mathrm{G}}(x) &= \sqrt{C_1(x) \cdot C_2(x)} = \sqrt{C_{10} \cdot C_{20} \cdot \mathrm{e}^{-2\mu \cdot a}} \\ &= \mathrm{e}^{-\mu \cdot a}\sqrt{C_{10} \cdot C_{20}} \end{aligned} \tag{12}$$

Equation (12) shows that the geometric mean of the two count-rates is independent of the position of the point source but depends linearly on the activity of the source. Thus, the geometric mean can be used to measure the activity of an unknown source using a calibrated source positioned on one of the faces of the box between the detector and the box:

$$\mathrm{Act}(u) = \frac{M_{\mathrm{G}}(u)}{M_{\mathrm{G}}(\mathrm{Std})} \cdot \mathrm{Act}(\mathrm{Std}) \tag{13}$$

2.1.2.2 The Error due to a Longitudinal Source The mathematical approach described previously assumed that the active source is very thin. In the following, the error is induced if the source has a dimension in the plane of the detectors. Let us assume that the whole activity is spread over a length b (Figure 2). Then, the count-rates due to infinitesimal length dx, assuming there is no absorption by the medium, are $\frac{C_{01}}{b}\mathrm{d}x$ and $\frac{C_{02}}{b}\mathrm{d}x$. The

count-rates due to the absorption effects are (Figure 3):

$$C_1 = \frac{C_{01}}{b}\int_0^b \mathrm{e}^{-\mu(x+d)}\,\mathrm{d}x = \frac{C_{01}}{\mu b}\mathrm{e}^{-\mu d}\cdot(1-\mathrm{e}^{-\mu b})$$

$$C_2 = \frac{C_{02}}{b}\int_0^b \mathrm{e}^{-\mu[2a-(x+d)]}\,\mathrm{d}x = \frac{C_{02}}{\mu b}\cdot\mathrm{e}^{-\mu 2a}\cdot\mathrm{e}^{\mu d}(\mathrm{e}^{\mu b}-1)$$

where $2a$ is thickness of the box, d is the distance of one side of the source from one of the detectors, and x is the distance of the infinitesimal length $\mathrm{d}x$ from the same side of the source. The geometric mean is

$$M_{\mathrm{G}} = \sqrt{C_1\cdot C_2} = \mathrm{e}^{-\mu a}\sqrt{C_{01}\cdot C_{02}}\frac{\sqrt{(1-\mathrm{e}^{-\mu b})(\mathrm{e}^{\mu b}-1)}}{\mu b} \tag{14}$$

The correction factor K due to the dimension b perpendicular to the detector's plane is given by

$$K = \frac{\sqrt{(1-\mathrm{e}^{-\mu b})(\mathrm{e}^{\mu b}-1)}}{\mu b} \Rightarrow K = \frac{\mathrm{e}^{\mu b/2}(1-\mathrm{e}^{-\mu b})}{\mu b} \tag{15}$$

As long as $\mu b < 1.5$, the error is <10% ($K < 1.10$). For example, for $^{99\mathrm{m}}$Tc in water $\mu = 0.152\,\mathrm{cm}^{-1}$, then for an object with size <10 cm, the error is <10%.

2.1.2.3 Source Localization The previous mathematical model yields the activity of the unknown source in the box without knowing where the activity is located. However, the location of the source can also be calculated using Equation (11). Dividing the two, Equation (11) yields

$$R(x) = \frac{C_1(x)}{C_2(x)} = \left[\frac{C_{10}}{C_{20}}\right]\cdot\mathrm{e}^{\mu\cdot 2x} \tag{16}$$

$$x = \frac{\ln[R(x)/(C_{10}/C_{20})]}{2\mu} \tag{17}$$

If the two detectors are equal, or are made to give equal responses by adding a factor to one of them, then $C_{10}/C_{20} = 1$.

If the material in the box is known, Equation (15) can be used for calculating x from the known μ. However, in case the content of the box is unknown, μ can also be calculated from the measurement of the external calibrated standard on the face of the box by the two detectors, as the only decrease in the counts is due to the absorption.

$$\mu = \frac{\ln[R(a)]}{2a} \tag{18}$$

Here, $R(a)$ is the ratio of the two count-rates due to the external source alone.

2.1.3 The General Case

The general case is where both, the solid angle in which the detector sees the source and the absorption/scattering, are operational.

2.1.3.1 Activity Calculation In the previous sections, it was shown that, when only the solid angle is important, the activity can be calculated from the square root harmonic average, while when only absorption is operational the geometric average is the right method to calculate the activity. Since these two effects are independent, it can be expected that, in the general case, the right answer is a weighted product of these two averages. The weighing must ensure that the units of the new average are the same as count-rate[10]:

$$M_{\mathrm{GSH}} = {M_{\mathrm{SH}}}^{\alpha}\cdot {M_{\mathrm{G}}}^{1-\alpha} \tag{19}$$

It was found that α depends on the contents of the box[10] and α can be found from the measurement of the count-rates at several points of the box. However, for an unknown box, this means opening the box or reproducing a similar box for calculating α. The method to measure the radioactivity in the box without the need to open it is by previous localization of the radioactive point source, as discussed in the next paragraph.

2.1.3.2 Localization of the Point Source It was shown[11] that the count-rate of a detector positioned on the face of a box as a function of the distance of a point source, positioned inside the box filled with polypropylene, from the detector is close to exponential dependence, but is not very accurate. The exponential dependence on the distance from the box's face is improved if, instead of taking the count-rate of one detector, the ratio of the count-rates of two detectors positioned on parallel faces is taken. Another possibility to improve the exponential dependence is to position the detector at a distance from the face of the box. However, increasing the distance between the detector and the box leads to worse statistics because of a smaller count-rate. In addition, the use of the ratio of the count-rates of two detectors is preferable as the equation of the count-rate $C(x)$ versus the distance of the source from the box face $-x$,

$$C(x) = C(0)\cdot\mathrm{e}^{-\mu\cdot x} \tag{20}$$

has two parameters, $C(0)$ and μ, which require at least two calibration experiments. Performing two calibration measurements is difficult if we do not want to open the box. On the other hand, if the two detectors are equal, or made to be equal, by adding a factor to one of them,

the two parameters of the equation of the ratio of the count-rates

$$R(x) = R(0) \cdot e^{-\mu \cdot x} \tag{21}$$

can be shown to be correlated, leading to only one parameter.

If the length of the box is a, then $R(0)$ and $R(a)$ are reciprocal numbers, as changing from 0 to a means replacing the detectors. Substituting $R(a) = 1/R(0)$ yields

$$R(0) = e^{\mu \cdot a/2} \tag{22}$$

Hence

$$x = -\frac{a}{2} \cdot \left(1 - \frac{\ln[R(x)]}{\ln[R(0)]}\right) \tag{23}$$

The activity of the source cannot be calculated from the ratio of the count-rates as $R(x)$ is independent of the source activity. For the determination of the source activity, the measurement of the count-rate of one detector must be used, together with the calculated x from Equation (23). Since in the case of the detectors close to the box, the dependence of the count-rate on the distance x is quite complex, it is preferable to use measurements at a distance between the detectors and the box. As mentioned before, there are two unknowns, $C(0)$ and μ, but μ for one detector measurement can also be deduced from the measurements of the ratio of the count-rates at the same distance detector box. It was shown earlier that μ for the expression $R(x)$ can be found from $R(0)$ and the box length. The μ for one detector count-rate was proven to be one-half of the μ for the ratio of the count-rates.[11]

2.2 Four Detectors

Old lung counters were made of two large-sized phoswich detectors. In order to increase the resolution of modern lung counters, semi-planar Ge detectors were used. Owing to the small size of the available Ge detectors, 3 or 4 detectors were typically used to cover the area of the two lungs previously covered by two phoswich detectors.[12,13] A commercially available set of four detectors is the Lung-Counter composed of four semiplanar HPGe detectors that were positioned two on each lung on the chest.

2.2.1 Lung Detector

The basic assumption generally made when calibrating a lung counter is that the deposition of the radioactive aerosols in lungs is homogenous; however, it has been reported that the distribution is a function of the aerosol size and the breathing rate and changes with time.[14] In some cases, the contamination can be present in the form of a single ''hot'' particle. The count-rate of an external detector is strongly dependent on the position of the ''hot particle''; the possible errors in activity determination due to deviations from homogeneity were evaluated to be factors of 26, 3, and 1.8 for photon energies of 17, 60, and 1000 keV respectively,[15] and factors of up to 20 were measured for natural uranium.[16] The extra information due to more than one detector allows the localization of a point source in the lungs.

2.2.1.1 The Mathematical Method The mathematical method used is applicable not only for localization of radioactive sources but also for any technique using several measurements to identify an unknown.[17] The most well known analysis using this mathematical method is the identification of an analyte from its electron-impact mass spectrum (EIMS),[18] but the method can be also used for source identification.[19–21] It is based on a calibration with a radioactive point source positioned consecutively in various locations in a phantom, which is the actual bulky material. For every positioning, the four detectors measure the pulse height spectra, and the count-rate for each γ full-energy peak is then calculated. The position of an unknown radioactive point source can be found by comparing its count-rates measured by the four detectors with similar data obtained for each of the various locations during the calibration. The mathematical problem is to find which of the sets of the standards (which, in our case, contain the results of the four detectors) best fits the four count-rates of the unknown.

The values of the count-rates of the four detectors are set as the components of a vector and all the vectors are normalized to unity length.[22] Each normalized vector can be seen as a single point on a hypersphere with a unity radius in an n-dimensional hyperspace, where n is the number of the vector components (the number of the detectors). If two vectors (points on the n-dimensional hypersphere) are identical in all the components, there will be a perfect ''match'' that represents the same point in the hyperspace. However, because of instrumental instability and the statistical nature of the measurements, the point of an unknown coincides exactly with one of the points in the library of standards only rarely. Actually, even repeating the same measurement at the same point rarely gives exactly the same count-rates. The problem of finding the best match for the vector of an unknown in the library of the vectors of standards can be solved in principle by two methods (two matching factors). The first method can be described as the ''nearest-neighbor'' technique, while the second can be defined as the ''minimal contact angle'', although, in many cases, it is called the *dot product method* after the mathematical operation used to calculate the angle.

The first method calculates the distances between the head of the vector of the unknown and the heads of each of the vectors of the standards (the library), looking for the minimum distance, i.e. the minimum of

$$\Delta_j = \sum_{i=1}^{n} (u_i - s_{i,j})^2 \tag{24}$$

where u_i is the ith component (the detector i) of the vector of the count-rates of the unknown and $s_{i,j}$ represents the ith component of the jth vector in the library of standards. The summation is performed over all the components of the vector (all the detectors). It should be emphasized that all the vectors must be normalized (to be of unity length) in order to compensate for different activities of the standard and the unknown:

$$u_i = \frac{\mathbf{U}_i}{\sum_{k=1}^{n} \mathbf{U}_k}; \quad s_{i,j} = \frac{\mathbf{S}_{i,j}}{\sum_{k=1}^{n} \mathbf{S}_{k,j}} \tag{25}$$

where **U** and **S** are the original un-normalized vectors.

In the second method, the angles between the unknown vector and each of the standard vectors are calculated. The cosine of the angle between two vectors can be calculated from the values of their components by using the scalar ''dot product'':

$$\cos\theta_j = \frac{\underline{s}_j \cdot \underline{u}}{|\underline{s}_j| \cdot |\underline{u}|} = \frac{\sum_{i=1}^{n} s_{i,j} \cdot u_i}{\sqrt{\sum_{i=1}^{n} s_{i,j}^2 \cdot \sum_{i=1}^{n} u_i^2}} \tag{26}$$

The summation is done over all the components of the two vectors (which are of the same dimension). Actually, the calculation of the angle does not require the normalization of the vectors and can be performed on the original data:

$$\cos\theta_j = \frac{\underline{\mathbf{S}}_j \cdot \underline{\mathbf{U}}}{|\underline{\mathbf{S}}_j| \cdot |\underline{\mathbf{U}}|} = \frac{\sum_{i=1}^{n} \mathbf{S}_{i,j} \cdot \mathbf{U}_i}{\sqrt{\sum_{i=1}^{n} \mathbf{S}_{i,j}^2 \cdot \sum_{i=1}^{n} \mathbf{U}_i^2}} \tag{27}$$

However, it is preferable to use the normalized vectors at least for all the standards. Even if the vectors are not normalized explicitly, normalization is still performed in the process of calculating the lengths of the vectors in the denominator of Equation (27), which includes division by the lengths of the vectors. It is better to calculate the lengths only once in the normalization process rather than to calculate them for each unknown.

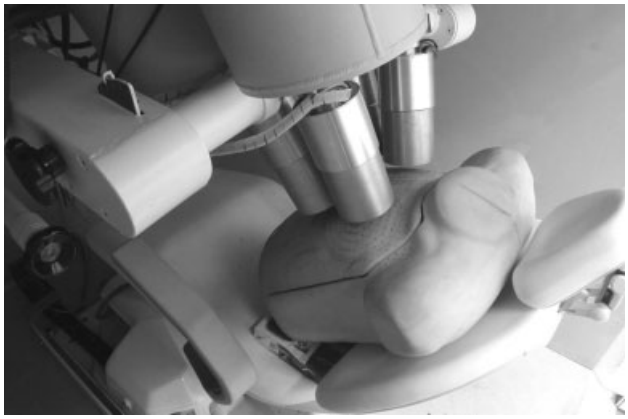

Figure 3 The realistic phantom used in the experiments.

Lengths of normalized vectors are unities, which simplifies Equation (26):

$$\cos\theta_j = \frac{\underline{s}_j \cdot \underline{u}}{|\underline{s}_j| \cdot |\underline{u}|} = \sum_{i=1}^{n} s_{i,j} \cdot u_i \tag{28}$$

The Experimental System The measurements were performed by the NRC-Negev lung counter system, which consists of four semiplanar HPGe detectors measuring a uranium source inside a phantom, as can be seen in Figure 3. The information from each detector was analyzed separately by a multichannel analyzer. The uranium source was a sealed cylinder with a radius of 1.5 mm and length of 5 mm, which contained natural uranium (99.27% ^{238}U and 0.72% ^{235}U) with an activity of 1.85 kBq. The main γ lines were 92 keV because of the ^{238}U chain (due to ^{234}Th) and 186 keV mainly because of ^{235}U. The quantitative information from each measurement was the count-rates of the 186 and 92 keV peaks (the areas under the peaks divided by the counting time). Other γ peaks were also observed; however, their counts were smaller, implying larger statistical errors, and they were not used in the analysis.

The detectors were calibrated using a realistic phantom designed by Lawrence Livermore National Laboratory. The phantom lungs used were made of tissue-equivalent plastic material and had 28 cylindrical holes, where point sources of natural uranium could be placed, as can be seen in Figure 4. All holes were normally filled with cylinders made of the same tissue-equivalent plastic as the lungs, and a small portion of the cylindrical filling was replaced by the source only at the checked position, so that the absorbing properties of the lungs remained practically unchanged and fully simulated the human lungs. The 56 points for placing the sources at different locations (two points in each hole) were distributed all over the lungs. Twenty-eight points were located on the upper surface of the lungs and 28 on the lower surface.

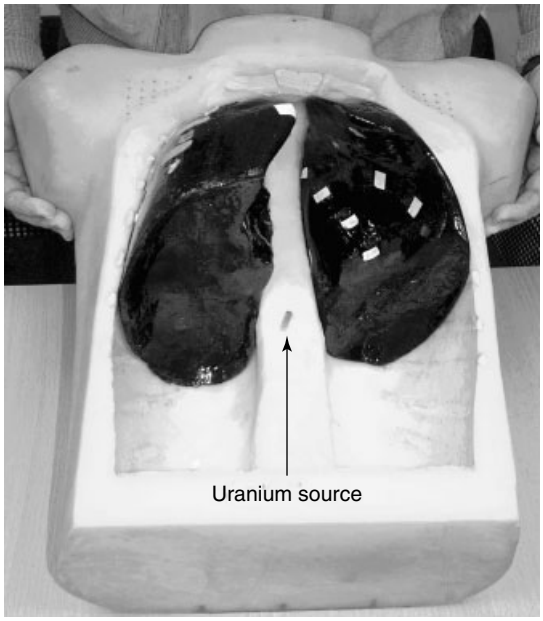

Figure 4 The four lung counters placed over the phantom.

The matching factors (Δ_j and $\cos\theta_j$) of the counts for the source placed at an unknown position with each of the 56 different points of the library data base were calculated. On the basis of the minimum value of Δ_j, the matching factor, or a maximum of $\cos\theta_j$, the other matching factor, a ''predicted point'' was determined and compared with the actual point location. If there was a match between them, it was called a *success* or just a *hit*.

The difference between the calculated and the actual activity (the error in calculation of the activity) was calculated as follows. After determination of the ''guessed'' point location, the guessed (calculated) source activity was calculated from the count-rates of the source at the unknown position, and the efficiency factor for the guessed point was determined. The difference between this calculated activity for each single measurement and the known source activity was the error for the specific measurement.

There was a one-to-one correlation between $\cos\theta$ (θ stands for the angle between the two vectors) and the distance between the points (heads of the vectors) in the hyperspace, and it was found, in the case of localization of a radioactive point source in the lung phantom, that Equations (3) and (5) actually gave the same results.[19]

The Use of Weighted Results Stein[23] found that the identification in EIMS gets better if, instead of raw ion intensities, the values (ion intensity · ion mass)$^{1/2}$ are taken as the components of the vectors. This choice of components of the vectors increases the weights of low-intensity peaks produced by ions of big masses.

In our case of the lung phantom, it was also found that the number of ''hits'' increased slightly if raw values of the count-rates were multiplied by weighing factors. The various weighing factors studied are given in Table 1. For calculation of the least distance, it was found that the best results were obtained for weights taken as the reciprocal of the mean of the normalized count-rates of the unknown and the library vectors. The raw vectors are first normalized, and the weights are calculated from the normalized vectors. In a mathematical notation, this means that, instead of searching for the library vector $\mathbf{L}$ that leads to the minimal values of $\sqrt{\sum_{i=1}^{n}(\mathbf{L}_i - \mathbf{U}_i)^2}$ for the various vectors of the library (where $\mathbf{U}$ is the vector of the unknown), we are looking for the vector $\mathbf{L}$ that yields a minimal value of $\sum_{i=1}^{n}(\mathbf{U}_i - \mathbf{L}_i)^2/(\mathbf{U}_i + \mathbf{L}_i)$. Because we are looking for the minimal distance, this assigns bigger weights to the measurements of larger counts with smaller statistical errors. When these weights are used, the fraction of the ''hits'' (where ''hit'' means that the method detects the correct location of the source) increases from 78 to 82%.[19] Although there were only 82% of ''hits'', the mean error in the calculated activity was only 10.4% because, in the ''nonhit'' cases, the resulted position was

Table 1 The various test criteria

1	2	3	4	5
$\cos^2\theta = \dfrac{\left(\sum \mathbf{L}_i \cdot \mathbf{U}_i\right)^2}{\sum \mathbf{L}_i^2 \cdot \sum \mathbf{U}_i^2}$	$\lvert r\rvert = \dfrac{\sum(\mathbf{U}_i - \overline{\mathbf{U}})\,(\mathbf{L}_i - \overline{\mathbf{L}})}{\sqrt{\sum(\mathbf{U}_i - \overline{\mathbf{U}})^2 \sum(\mathbf{L}_i - \overline{\mathbf{L}})^2}}$	$\sum_{i=1}^{n} \mathbf{L}_i \cdot \lvert\mathbf{U}_i - \mathbf{L}_i\rvert$	$\sum_{i=1}^{n} \mathbf{L}_i \cdot (\mathbf{U}_i - \mathbf{L}_i)^2$	$\sum_{i=1}^{n} \left\lvert \dfrac{\mathbf{U}_i}{\mathbf{L}_i} - 1 \right\rvert$
6	7	8	9	10
$\sum_{i=1}^{n} \left(\dfrac{\mathbf{U}_i}{\mathbf{L}_i} - 1\right)^2$	$\sum_{i=1}^{n} \dfrac{\lvert\mathbf{U}_i - \mathbf{L}_i\rvert}{\mathbf{U}_i}$	$\sum_{i=1}^{n} \dfrac{\lvert\mathbf{U}_i - \mathbf{L}_i\rvert}{\sqrt{\mathbf{U}_i \cdot \mathbf{L}_i}}$	$\sqrt{\sum_{i=1}^{n}(\mathbf{L}_i - \mathbf{U}_i)^2}$	$\sum_{i=1}^{n} \lvert\mathbf{L}_i - \mathbf{U}_i\rvert$
11	12	13	14	15
$\sum_{i=1}^{n} \dfrac{(\mathbf{U}_i - \mathbf{L}_i)^2}{(\mathbf{U}_i + \mathbf{L}_i)}$	$\sum_{i=1}^{n} \left[\dfrac{\mathbf{U}_i - \mathbf{L}_i}{\mathbf{U}_i + \mathbf{L}_i}\right]^2$	$\sum_{i=1}^{n} \dfrac{(\mathbf{U}_i - \mathbf{L}_i)^2}{\mathbf{U}_i \cdot \mathbf{L}_i}$	$\sum_{i=1}^{n} \dfrac{(\mathbf{U}_i - \mathbf{L}_i)^2}{\sqrt{\mathbf{U}_i \cdot \mathbf{L}_i}}$	$\sum_{i=1}^{n} \dfrac{\lvert\mathbf{U}_i - \mathbf{L}_i\rvert}{\mathbf{U}_i + \mathbf{L}_i}$

Table 2 The results obtained when either the 186 keV or the 92 keV peaks were used separately and when both peaks were used simultaneously, using the library of 224 measurements

Criterion number	92 keV		186 keV		92 + 186 keV	
	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)
1	78	17.6	81	10.9	90	5.5
2	67	23.8	71	16.8	73	9.5
3	64	19.2	72	20.4	74	13.0
4	70	21.0	77	16.1	79	10.7
5	79	14.6	81	11.4	88	6.0
6	81	14.0	80	11.0	89	5.5
7	82	11.3	81	11.4	89	5.5
8	82	11.3	80	12.3	90	5.4
9	78	17.6	81	10.9	90	5.5
10	80	14.5	82	10.4	88	6.3
11	82	11.2	84	9.8	94	4.1
12	82	13.9	81	10.9	90	5.7
13	82	13.9	81	10.9	90	5.7
14	82	11.3	82	9.5	94	3.8
15	83	11.3	80	12.3	90	5.4

quite close to the real one, which led to a relatively small error in the calculated activity.

Homogenous versus Point-wise Distribution Application of this method to a homogenously distributed source in the lungs of the phantom results in the best "fit" for one of the points in the center. The error in the activity is quite large in this case, almost 40%. It seems that this method can work only when applied to inhomogenously distributed radioactive contaminants. However, this problem can be solved by adding another vector to the set of the standards.[24] This vector is the response of the four detectors to a phantom with homogenously distributed radionuclides, which can be defined as a "virtual point". Another possibility for a vector representing the homogenously distributed lungs is to take the arithmetic mean of the 56 calibrating points for each one of the four detectors. With this extended set of standards, the "guessed point" for homogenously distributed lungs is always found in this virtual vector, and the error in the calculated activity is below 5%. Thus, one can actually distinguish between a homogenous distribution and a point location.

Simultaneous use of two γ lines In the case of uranium contamination, the count-rates of two relatively strong γ lines were measured by the four detectors, namely, the lines at 92 and 186 keV. The 92-keV line is due to a daughter in the ^{238}U chain, while the 186 keV line originates mainly from ^{235}U. These γ lines, when used separately, provide approximately the same fraction of "hits" and the same mean error in the activity, but a closer observation shows that the cases of "nonhits" are not the same for both the lines. This is reasonable because the main cause for a "miss" is the statistical nature of the count-rates. It was found[20] that analyzing the two energy peaks together could provide better results. Eight-component vectors were composed. Out of these, four components were the 92 keV count-rates measured by the four detectors, while the other four components were count-rates measured by the same detectors for the 186 keV line. The normalization of the vectors involved all the eight components. It can be seen in Table 2 that the fraction of "hits" increased from 82%, when each of the energies was analyzed separately, to 94%, when the two energies were analyzed together using an eight-dimensional vector. The mean error in the activity decreased from 10.4 to 3.8%.

Table 2 gives the percentage of "hits" and the error in the calculated activity for four detectors, both for using only the count-rates at either the 186 or the 92 keV peak (vectors of four dimensions) and the simultaneous count-rates at the 92 and the 186 keV peaks (eight-dimensional vectors). In the case of Table 2, the library was built from all the 224 experiments, and each standard vector was the average of four repeated experiments. The unknown vector was one of the 224 vectors, leading to the same vectors being used as part of the library, although in an average with three other experiments and as the unknown. In Table 3, the same results are given when the average of only two repeating experiments (112 measurements) was used to calculate the library, while

Table 3 Average results obtained when either the 186 or the 92 keV peaks were used separately and when both peaks where used simultaneously, using all six library combinations of 112 measurements

Criterion number	92 keV		186 keV		92 + 186 keV	
	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)	Number of "hits" from 224 measurements (%)	Average error in activity calculation (%)
1	64 ± 2	28.1	69 ± 2	17.8	74 ± 2	12.4
2	54 ± 4	51.6	58 ± 4	34.5	68 ± 3	16.3
3	51 ± 4	43.4	59 ± 5	30.6	60 ± 4	19.8
4	57 ± 4	36.4	63 ± 3	26.8	64 ± 3	18.2
5	64 ± 3	26.9	69 ± 2	16.4	76 ± 2	12.5
6	64 ± 2	28.9	68 ± 2	17.7	74 ± 2	13.3
7	65 ± 5	25.7	69 ± 2	16.6	75 ± 5	12.6
8	65 ± 4	27.2	69 ± 2	16.3	77 ± 2	11.4
9	64 ± 2	28.1	69 ± 2	17.8	74 ± 2	12.4
10	63 ± 1	28.4	68 ± 2	21.5	76 ± 3	11.3
11	68 ± 2	23.4	71 ± 2	16.4	79 ± 3	10.5
12	66 ± 3	26.6	69 ± 2	16.9	76 ± 2	12.0
13	66 ± 2	26.6	68 ± 2	17.3	75 ± 2	12.2
14	68 ± 2	23.2	71 ± 2	16.0	79 ± 3	10.7
15	65 ± 4	27.3	69 ± 2	16.3	77 ± 2	11.6

the other 112 measurements were treated as unknowns. There are six possibilities (C_4^2) to choose the two repeating experiments used for the library; the results in Table 3 are the average of these six possibilities.

However, using the two γ lines presents a problem when the isotopic composition of the uranium in the calibration standards is different from that of the unknown, because the two lines are produced by different isotopes. It was shown[21] that, if the isotopic compositions were different, errors in localizing a source and determining its activity were larger with the eight-dimensional vector than with the four-dimensional one. This study[21] shows that, in such a case, the method of simultaneous analysis of two γ energy peaks leading to eight-dimensional vectors can still be used to improve the accuracy, but the eight-dimensional vectors should be normalized in an unusual way, which can be called *double half-normalization*. In this method, the four count-rates of each γ energy are normalized separately. The components of the normalized vector c_j are

$$c_j = \frac{C_{92,j}}{\sum_{k=1}^{4} C_{92,k}}; \quad \text{for } j = 1\text{–}4$$

$$c_j = \frac{C_{186,j-4}}{\sum_{k=1}^{4} C_{185,k}}; \quad \text{for } j = 5\text{–}8$$

In the expressions above, the measured count-rates are $C_{92,j}$ and $C_{186,j}$, where $j = 1$–4 represent the four different detectors. Pelled[21] found that this complex normalization (the lengths of the vectors are actually 2) decreases the fraction of "hits" in the case of the same isotopic composition only from 94 to 93%, which is still much higher than the success rate when a single γ line in used. When the isotopic compositions are different, this larger dimensionality increases the fraction of "hits" almost to the same degree as in the case of the same isotopic composition. However, when the isotopic compositions are extremely different, the statistics of one of the energy lines is expected to be poor; therefore, it is better to use only the most abundant γ line.

Detectors at the Back Another way to increase the dimensionality of the vectors is to use more detectors. In practice, the four detectors cover most of the area of the lungs, and, physically, there is no place for additional ones. However, they can be mounted on the back of the phantom. Counts by detectors mounted on the back are usually lower due to stronger absorption of the photons. Hence, longer counting times are required to get good statistics. Count-rates for the normal counting time with the detectors positioned at the back of the phantom were studied,[24] thus simulating an array of eight detectors. The results can be represented by eight-dimensional vectors or 16-dimensional vectors if both the γ energies (92 and 186 keV) are used. The percentage of "hits" exceeded 99%, suggesting that placing lung detectors behind the phantom, by mounting them in a chair support, is worthwhile. There is actually no need for eight detectors as only two additional detectors at the back increase the fraction of "hits" to about 99%.

Table 4 The percentage of "hits" and the error in the activity calculation with four detectors for 56 points (224 measurements)

Combination number	Number of detectors	Detectors	92 + 186 keV		186 keV	
			Number of "hits" (%)	Error in activity (%)	Number of "hits" (%)	Error in activity (%)
1	**4**	**A, B, C, D**	**94**	**3.8**	**83**	**9.4**
2	4	BA, BB, BC, BD	93	4.7	82	10.1
3	4	A, B, BA, BB	98	1.9	95	2.4
4	4	A, D, BA, BD	98	2.1	92	5.9
5	4	B, C, BB, BC	99	1.8	95	2.9
6	4	A, B, BC, BD	98	1.3	97	1.4
7	4	C, D, BA, BB	96	2.0	95	2.2
8	4	C, D, BC, BD	96	1.5	96	1.6
9	4	A, B, D, BA	99	1.6	96	2.2
10	4	A, B, D, BB	99	1.6	97	1.8
11	4	A, B, D, BC	99	1.4	98	1.6
12	4	A, B, D, BD	99	1.7	95	2.1
13	4	A, B, C, BA	98	1.8	92	3.2
14	4	A, B, C, BB	97	2.1	96	2.1
15	4	A, B, C, BC	99	1.6	97	1.7
16	4	A, B, C, BD	99	1.5	95	1.8
17	4	A, C, D, BA	98	2.0	93	3.4
18	4	A, C, D, BB	99	1.7	96	2.3
19	4	A, C, D, BC	99	1.5	97	1.8
20	4	A, C, D, BD	98	1.9	94	2.5
21	4	B, C, D, BA	98	1.7	94	2.4
22	4	B, C, D, BB	98	1.7	95	2.1
23	4	B, C, D, BC	100	1.5	97	1.9
24	4	B, C, D, BD	98	1.7	93	2.3

Possibly more interesting results of this study are that, even when only four detectors are used, it is considerably better to locate one or two of them at the back.[24] Tables 4 and 5 yield the percentage of "hits" and the error of the calculated activity both for using the 186 keV peak alone and for the simultaneous use of the 92 and 186 keV peaks. Each line gives the results for another arrangement of the four detectors. The detectors in the front are assigned by A, B, C, and D, while the detectors at the back are denoted by BA, BB, BC, and BD. Detector A is located opposite the upper part of the right lung. Detector B is located opposite the upper part of the left lung. Detector C is located opposite the lower part of the left lung, and detector D is located opposite the lower part of the right lung. When measuring very low energy γ photons, as, for example, when measuring contamination with ^{239}Pu, the detectors at the back hardly see anything. However, Tables 4 and 5, in which each line giving a different arrangement of the four detectors shows that, even for natural uranium, with not very high energy photons, positioning of one or two, out of the four, detectors at the back leads to considerably more accurate results. This is clearer in the case of only one photon energy or if the library used were counted for a relatively short time (6 h in Table 5 compared to 12 h in Table 4).

3 NEUTRON SOURCES

The main difference between neutron sources and γ sources is that, while γ sources emit photons of discrete energies, mainly one or two major lines, the radionuclidic neutron source, both fission sources or (α, n) sources, emits a spectrum of neutrons, mainly fast ones with an average of more than 1 MeV. Another difference is that it is quite easy to measure the spectrum of the γ photons emitted from the box, and the resolution of the energy spectrum is quite high, while, on the other hand, measurement of the spectrum of the neutrons is quite a task, and the resolution of the energy spectrum is quite low. Taking these two differences and the much larger sensitivity of usual neutron detectors to thermal neutrons than to more energetic neutrons, it is almost impossible to measure only neutrons that reach the detectors without interaction with the box and the surrounding materials, as was done for γ photons. Not only the interaction with the box but also the interaction with the shield of the experimental system should be considered, as the detectors must be shielded from outside radiation and the researcher must be protected from the neutron irradiation. This shield, if made from hydrogen-containing material, not only shields the detector from outside

Table 5 The average percentage of "hits" and the average error in the activity calculation with four detectors for 56 points (112 measurements, average of six combinations)

Table combination number	Number of detectors	Detectors	92 + 186 keV		186 keV	
			Number of "hits" (%)	Error in activity (%)	Number of "hits" (%)	Error in activity (%)
1	**4**	**A, B, C, D**	**79.8 ± 3.0**	**10**.2	**71.8 ± 1.9**	**16.0**
2	4	BA, BB, BC, BD	77.0 ± 4.8	13.1	66.7 ± 3.4	18.3
3	4	A, B, BA, BB	91.8 ± 1.2	3.6	86.5 ± 3.4	4.6
4	4	A, D BA, BD	92.7 ± 1.4	4.4	83.5 ± 2.3	14.8
5	4	B, C, BB, BC	95.2 ± 1.0	3.0	88.0 ± 2.8	6.4
6	4	A, B, BC, BD	95.0 ± 1.8	2.3	93.7 ± 2.9	2.4
7	4	C, D, BA, BB	91.3 ± 1.5	3.7	90.7 ± 2.9	3.8
8	4	C, D, BC, BD	93.7 ± 1.4	2.4	89.8 ± 3.5	3.4
9	4	A, B, D, BA	93.7 ± 0.5	3.0	86.0 ± 2.8	5.7
10	4	A, B, D, BB	96.3 ± 1.2	2.6	92.7 ± 2.5	3.4
11	4	A, B, D. BC	97.8 ± 1.8	2.1	95.7 ± 1.8	2.6
12	4	A, B, D, BD	94.7 ± 1.4	2.5	89.7 ± 3.7	3.5
13	4	A, B, C, BA	92.7 ± 1.2	3.2	85.3 ± 3.4	5.3
14	4	A, B, C, BB	92.3 ± 1.0	3.4	87.8 ± 3.5	4.8
15	4	A, B, C, BC	92.7 ± 3.3	3.3	92.2 ± 2.4	3.3
16	4	A, B, C, BD	95.5 ± 1.9	2.3	93.0 ± 2.8	2.6
17	4	A, C, D, BA	93.0 ± 0.9	3.8	86.8 ± 2.5	5.5
18	4	A, C, D, BB	95.2 ± 1.6	3.3	89.7 ± 1.2	4.6
19	4	A, C, D, BC	96.3 ± 1.4	2.6	93.3 ± 1.9	3.0
20	4	A, C, D, BD	92.5 ± 1.4	2.9	88.7 ± 2.9	3.7
21	4	B, C, D, BA	92.7 ± 1.4	3.2	86.0 ± 1.9	4.3
22	4	B, C, D, BB	91.5 ± 1.2	3.3	88.5 ± 2.5	4.0
23	4	B, C, D, BC	95.8 ± 1.8	2.5	92.5 ± 1.4	3.2
24	4	B, C, D, BD	91.3 ± 2.1	2.8	86.7 ± 2.0	3.6

radiation but also increases the count-rate at the detector due to reflection of thermal neutrons from the shielding material.

3.1 Two Neutron Detectors

It was found[25] that, when two ^{3}He detectors were positioned on two parallel faces of a box with a small neutron source in the box on the line connecting the two detectors, the ratio of the two count-rates was an exponential function of the distance between the source and one of these faces. The exponential dependence was found to be independent of the medium in the box, from air to highly hydrogenated material, although the coefficient of the exponentiality depended on the medium. This is the same as that described earlier for two γ detectors in the general case and hence the same equation should apply:

$$x = -\frac{a}{2} \cdot \left(1 - \frac{\ln[R(x)]}{\ln[R(0)]}\right) \tag{29}$$

where x is the distance from the box's face, a is the length of the box (the distance between the two faces on which the detectors are positioned), and R is the measured ratio of the two count-rates. $R(x)$ is the ratio for the unknown source and $R(0)$ is the ratio for an external source positioned on the face of the box ($x = 0$). It was found that Equation (29) yields quite accurate results, as can be seen in Table 6. In this table, $\Delta x/a$ is the fractional error in the location. As can be seen in this table, the calculated location is quite accurate except for the cases where the source is positioned very close to one of the two faces of the box. This is probably due to thermal neutrons coming to the close detector only by scattering from the shielding, as the neutrons coming directly from the source are too energetic to react with the ^{3}He in the detector. In order to get more accurate results, 5-cm polypropylene slabs were positioned between the detectors and the box.[26,27]

3.2 More than Two Detectors

When the unknown neutron source is not on the line (actually a plane as the detectors are quite long) connecting the two detectors, it is found that Equation (29) gives the correct result only if $R(0)$ is measured for the external source positioned on the face of the box in a plane parallel to the detectors' plane but passes through the unknown source. Therefore, now the problem is to find this plane. Two solutions are found to this problem. In the first solution, three pairs of ^{3}He

Table 6 The measured (x_{meas}) and calculated (x_{cald}) source-to-detector distance (cm) for an AmBe source

Wood			Paraffin			Air		
x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$
0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000
3.3	2.8	0.012	4.0	2.4	0.040	4.0	4.8	0.020
7.0	5.9	0.026	8.0	6.0	0.050	8.0	8.8	0.020
10.3	9.2	0.027	12.0	10.5	0.037	12.0	12.4	0.009
13.8	13.0	0.019	16.0	14.3	0.044	16.0	16.2	0.004
17.5	17.0	0.013	20.0	19.1	0.023	20.0	19.7	0.007
20.7	20.6	0.002	24.0	24.0	0.000	24.0	23.3	0.018
24.0	21.7	0.058	28.0	28.7	0.018	28.0	26.8	0.030
27.8	28.2	0.011	32.0	33.4	0.034	32.0	30.2	0.044
31.2	32.0	0.019	36.0	37.0	0.025	36.0	33.8	0.054
35.0	35.4	0.010	40.0	39.3	0.018	39.0	38.0	0.024
40.6	39.8	0.020						
Foamplast			**Paper**			**Concrete tiles**		
x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$	x_{meas}	x_{cald}	$\Delta x/a$
0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000
3.0	4.1	0.028	2.4	1.5	0.022	4.8	4.8	0.001
5.0	6.7	0.041	5.2	3.3	0.047	10.0	9.8	0.006
9.0	10.3	0.031	7.8	6.0	0.046	12.1	12.2	0.001
13.0	14.0	0.025	12.0	10.6	0.036	16.8	17.4	0.015
17.0	17.6	0.015	16.0	15.5	0.012	21.7	22.5	0.019
21.0	20.8	0.004	21.5	20.9	0.016	26.6	27.4	0.021
25.0	24.2	0.020	25.0	25.0	0.000	29.0	29.8	0.019
27.8	27.0	0.020	29.0	29.3	0.007	33.6	34.3	0.017
30.6	29.9	0.018	31.5	32.4	0.022	38.5	38.3	0.005
33.4	33.1	0.007	34.0	36.4	0.061			
36.2	34.6	0.041	36.5	38.0	0.039			
39.0	38.3	0.016	39.0	39.9	0.023			

detectors are positioned on two parallel faces of the box.[27,28] It is found that the pair that gives the highest geometric mean of the count-rates is the closest to the unknown source and hence the external source should be positioned in this plane. However, this still gives an error of one-sixth of the width of the box and, consequently, for larger boxes it will require more detectors. A better solution that requires only four detectors involves the moving of the external source on different faces by an iterative method.[29]

The iterative method was applied to localize the unknown neutron source by finding successive approximations to the solution, starting from an initial estimate regarding the unknown source plane (y coordinate). First, the x coordinate of the unknown source was calculated, using Equation (29), from the results of the two detectors on the y axis (detectors 1 and 2, which are on parallel faces of the box, as can be seen in Figure 5). Both $R(x)$ and $R(0)$ were measured by detectors 1 and 2. Measurements were made of the unknown source alone and together with a known (calibrating) source placed outside the box; $R(0)$ was the difference between the two measurements. In this first step, $R(0)$ was measured with the known

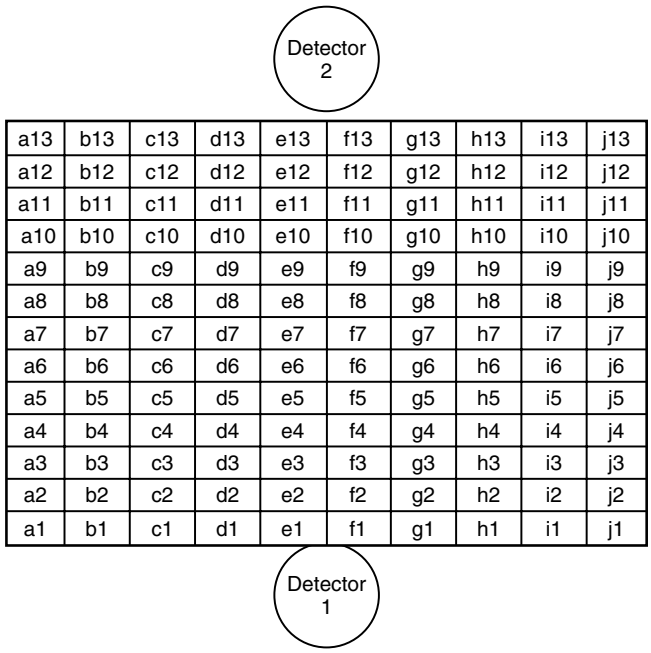

Figure 5 Schematic top view of the source positions measured in the 40×40 cm^2 sample cells. The sample was divided to 10 planes parallel to the detectors' plane.

Table 7 Iteration calculations of the source-to-detector distance (cm) for row b of the experimental results

	y_{real}	x_{real}	Before iterations	Initial estimate	First iteration		Second iteration		Third iteration		After iterations	
			y	x	y	x	y	x	y	x	y	x
b01	4	0	10.95	3.79	6.04	1.54	**6.04**	–	–	–	6.04	1.54
b02	4	3	9.27	5.64	5.23	2.29	3.18	**2.29**	–	–	3.18	2.29
b03	4	6	7.97	7.58	5.31	4.68	3.33	**4.68**	–	–	3.33	4.68
b04	4	9	6.25	10.16	3.14	7.86	**3.14**	–	–	–	3.14	7.86
b05	4	12	4.85	12.41	3.22	10.63	**3.22**	–	–	–	3.22	10.63
b06	4	15	3.93	15.20	3.22	14.07	**3.22**	–	–	–	3.22	14.07
b07	4	20	3.29	19.61	3.29	19.52	**3.29**	–	–	–	3.29	19.52
b08	4	25	3.98	23.92	3.31	24.84	**3.31**	–	–	–	3.31	24.84
b09	4	28	4.77	26.40	4.13	27.90	3.24	**27.90**	–	–	3.24	27.90
b10	4	31	6.27	29.16	4.92	31.29	3.48	**31.29**	–	–	3.48	31.29
b11	4	34	7.75	31.70	5.31	34.44	3.50	**34.44**	–	–	3.50	34.44
b12	4	37	9.11	33.71	5.42	36.92	3.31	**36.92**	–	–	3.31	36.92
b13	4	40	10.42	35.46	7.28	37.61	5.48	39.07	3.52	**39.07**	3.52	39.07
Average absolute error (cm)			2.95	1.91	1.19	0.84	–	–	–	–	0.79	0.73

Table 8 Pre- and postiteration distance calculations between the real source position and the calculated one

Row, y_{real} →		a, $y = 0$		b, $y = 4$		c, $y = 8$		d, $y = 12$		e, $y = 16$	
Line	x_{real}	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations
1	0	10.36	0.00	7.92	2.56	5.60	0.07	3.39	0.27	1.74	1.56
2	3	8.06	3.74	5.89	1.08	4.08	1.06	2.63	1.21	1.60	1.24
3	6	6.07	3.37	4.27	1.48	2.74	1.45	2.03	1.47	1.65	1.41
4	9	3.98	3.37	2.53	1.43	1.41	1.39	1.09	0.95	1.27	1.07
5	12	2.08	2.85	0.94	1.58	0.33	1.29	0.33	0.96	0.83	0.75
6	15	0.88	1.39	0.21	1.21	0.34	1.12	0.19	0.63	0.69	0.63
7	20	0.59	0.78	0.81	0.85	0.90	0.93	0.37	0.39	0.55	0.55
8	25	1.36	0.19	1.08	0.71	0.80	0.89	0.44	0.59	0.30	0.23
9	28	2.50	0.38	1.77	0.76	1.18	0.89	0.74	0.33	0.68	0.54
10	31	4.09	0.68	2.92	0.60	1.90	0.70	1.05	0.35	0.78	0.65
11	34	5.99	0.78	4.40	0.67	2.87	0.79	1.45	0.67	1.13	0.92
12	37	8.11	0.35	6.07	0.69	4.14	0.54	2.41	0.42	1.10	0.73
13	40	10.24	3.30	7.86	1.05	5.68	0.86	3.41	0.55	1.63	1.36
Average		4.95	1.63	3.59	1.13	2.46	0.92	1.50	0.68	1.07	0.90

Row, y_{real} →		f, $y = 20$		g, $y = 24$		h, $y = 28$		i, $y = 32$		j, $y = 36$	
Line	x_{real}	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations	Distance before iterations	Distance after iterations
1	0	1.38	0.73	3.20	0.30	5.70	0.24	8.09	0.65	10.47	2.87
2	3	1.22	0.92	2.16	1.03	3.90	1.31	5.89	1.50	8.19	1.30
3	6	1.67	1.51	2.00	1.28	3.04	1.13	4.34	1.56	6.33	1.61
4	9	1.10	0.92	1.35	0.88	1.99	0.89	3.08	1.08	4.61	1.61
5	12	1.20	1.13	0.83	0.59	1.04	0.70	1.76	0.90	3.00	1.50
6	15	1.07	1.04	0.32	0.25	0.24	0.75	0.68	0.75	1.49	1.00
7	20	1.32	1.32	0.47	0.47	0.14	0.16	0.14	0.16	0.77	0.78
8	25	1.32	1.28	0.79	0.54	0.75	0.27	1.11	0.39	2.17	1.15
9	28	1.10	0.99	1.00	0.34	1.43	0.21	2.00	0.41	3.28	1.33
10	31	1.28	1.12	1.35	0.50	2.05	0.62	3.01	1.02	4.69	1.54
11	34	1.50	1.32	1.95	0.93	3.04	1.01	4.62	1.10	6.35	2.94
12	37	1.40	1.07	2.51	0.86	4.08	0.87	5.99	1.17	8.16	1.20
13	40	1.57	1.15	3.31	0.42	5.46	0.28	7.78	0.42	10.39	0.84
Average		1.32	1.12	1.63	0.64	2.53	0.65	3.73	0.86	5.38	1.51

(calibrating) source positioned at the center of the y axis (on the central plane (initial guess)), connecting the two detectors 1 and 2 (at $x = 0$, 5 cm from detector $2^{(2)}$). This $R(0)$ was assigned $R(0)_x$. Then, the calibrating source was moved to the x axis (right side of the box, $y = 0$) at the calculated x (not to the exact value, but to one of the 13 values of the cells; to the cell closest to the calculated value) and $R(0)$ was measured again using the two detectors on the y axis (detectors 3 and 4, on the other two parallel faces of the box). This $R(0)$ was assigned $R(0)_y$ and was used to calculate the y coordinate, according to Equation 29, using $R(x)$, also from detectors 3 and 4. Then, the calibrating source was moved to the y axis at the calculated y (as close as possible to the calculated y cell). $R(0)_x$ was measured again and a new value of x was calculated. This process was repeated until either x or y was the same as in the previous iteration.

Table 7 shows the various calculations of x and y done by iterations. Table 7 presents, as an example, the iteration calculations of row b (Figure 5) of the experimental results. For most points, two iterations were performed after the initially estimated calculation of x, until x or y recurred. It was found that one iteration (two consecutive calculations of y and x) after the initial estimate already yielded a major improvement to the accuracy.

Table 8 shows the distance between the real source position and the calculated one, $\sqrt{(y_{\text{cald}} - y_{\text{real}})^2 + (x_{\text{cald}} - x_{\text{real}})^2}$ both for x and y calculated with central $R(0)$ for both coordinates and the distances calculated after completing all the iterations. One to four iterations are required, depending on the source's position. Iteration calculations reduce the average calculated distance between the real source position and the calculated one from 2.46 to 1 cm and reduce the maximal distance from above 10 to <4 cm.

REFERENCES

1. O. Presler, O. Pelled, U. German, Y. Leichter, Z.B. Alfassi, 'Determination of a Source in a Box with Two Detectors. I Non-absorbing Media', *Nucl. Instrum. Methods*, **A 491**, 314 (2002).
2. A. Notea, 'The Ge(Li) Spectrometer as a Point Detector', *Nucl. Instrum. Methods*, **91**, 513 (1971).
3. K. Debertin, R.G. Helmer, *γ - and X-Ray Spectrometry with Semiconductor Detectors*, Elsevier, New York, 250–255, 1988.
4. O. Presler, U. German, O. Pelled, Z.B. Alfassi, 'The Validity of the Virtual Point Detector Concept for Absorbing Media', *Appl. Radiat. Isot.*, **60**, 213 (2004).
5. O. Presler, O. Pelled, U. German, Y. Leichter, Z.B. Alfassi, 'Off-center Efficiency of HPGe Detectors', *Nucl. Instrum. Methods Phys. Res.*, **484A**, 462 (2002).
6. Z.B. Alfassi, O. Pelled, U. German, 'The Virtual Point Detector Concept for HPGe Planar and Semi-planar Detectors', *Appl. Radiat. Isot.*, **64**, 574 (2006).
7. S. Mahling, I. Orion, Z.B. Alfassi, 'The Dependence of the Virtual Point Detector on the HPGe Detector Dimensions', *Nucl. Instrum. Methods Phys. Res.*, **A557**, 544 (2006).
8. O. Presler, U. German, V. Pushkarski, Z.B. Alfassi, 'Virtual Point Detector: on the Interpolation and Extrapolation of Scintillation Detectors Counting Efficiencies', *Nucl. Instrum. Methods Phys. Res.*, **A565**, 704 (2006).
9. Z.B. Alfassi, N. Lavi, O. Presler, V. Pushkarski, 'HPGe Virtual Point Detector for Disk Radioactive Sources', *Appl. Radiat. Isot.*, **65**, 253 (2007).
10. O. Presler, U. German, H. Golan, Z.B. Alfassi, 'Determination of a Source in a Box with Two Detectors. The General Case', *Nucl. Instrum. Methods Phys. Res.*, **A527**, 632 (2004).
11. J. Feldman, R. Presler, M. Segal, Y. Sabov, O. Presler, W. Weinstein, U. German, Z.B. Alfassi, 'The Localization of Gamma Emitting Point Source in a Long Rectangular Medium Using Two Detectors on the Parallel Faces', *to be published*.
12. M.A. Lopez, T.N. Bravo, 'A Low Energy Germanium Detector System for Lung Counting at the WBC Facility of Ciemat', *Radiat. Prot. Dosim.*, **89**, 221 (2000).
13. J.L. Webb, M. Gadd, F. Bronsen, O. Tench, 'An Evaluation of Recent Lung Counting Technology', *Radiat. Prot. Dosim.*, **89**, 325 (2000).
14. K. Rhoads, J.A. Mahaffey, C.L. Sanders, 'Distribution of Inhaled PuO_2 in Rat and Hamster Lung', *Health Phys.*, **42**, 645 (1982).
15. G.H. Kramer, L.C. Burns, S. Yiu, 'Lung Counting: Evaluation of Uncertainties in Lung Burden Estimation Arising from a Heterogeneous Lung Deposition Using Monte Carlo Code Simulations', *Radiat. Prot. Dosim.*, **74**, 173 (1997).
16. O. Pelled, U. German, G. Pollak, Z.B. Alfassi, 'Estimation of Errors Due to Inhomogeneous Distribution of Radionuclides in Lungs, *Nucl. Instrum. Methods Phys. Res.*, **A564**, 491 (2006).
17. Z.B. Alfassi, 'The Vector Analysis of Multi Measurements Identification', *J. Radioanal. Nucl. Chem.*, **266**, 245 (2005).
18. S.E. Stein, D.R. Scott, 'Optimization and Testing of Mass Spectra Library Search Algorithms for Compound Identification', *J. Am. Soc. Mass Spectrom.*, **5**, p.859 (1994).
19. O. Pelled, S. Tzroya, U. German, G. Haquin, Z.B. Alfassi, 'Locating a ''Hot Spot'' in the Lungs when Using an Array

of Four HPGe Detectors', *Appl. Radiat. Isot.*, **61**, 107 (2004).

20. O. Pelled, S. Tzroya, U. German, A. Abraham, Z.B. Alfassi, 'Improved Localization of a ''Hot Spot'' in the Lungs for an Array of Four HPGe Detectors – the Simultaneous Use of Two Gamma Energies', *Nucl. Instrum. Methods Phys. Res.*, **A551**, 553 (2005).
21. O. Pelled, A. Abraham, U. German, Z.B. Alfassi, 'Localization of a ''Hot Spot'' of Uranium in the Lungs by an Array of Four HPGe Detectors. The Effect of Variation in the Isotopic Composition', *Nucl. Instrum. Methods Phys. Res. A*, **A584**, 406 (2008).
22. Z.B. Alfassi, 'On the Normalization of a Mass Spectrum for Comparison of Two Spectra', *J. Am. Soc. Mass Spectrom.*, **15**, 335 (2004).
23. S.E. Stein, 'An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data', *J. Am. Soc. Mass Spectrom.*, **10**, 770 (1999).
24. O. Pelled, Ph.D. thesis, presented to Ben Gurion University, Beer Sheva, Israel, 2006 approved 2006.
25. S. Dubinski, A. Talmor, O. Presler, A. Tshuva, I. Yaar, I. Orion, Z.B. Alfassi, 'The Determination of a Neutron Source Position in an Unknown Homogeneous Medium; the Planar Case', *Nucl. Instrum. Methods Phys. Res.*, **A548**, 555 (2005).
26. S. Dubinski, O. Presler, Z.B. Alfassi, 'The Determination of a Neutron Source Position in an Unknown Homogeneous Medium; the 2D Case', *Radiat. Prot. Dosim.*, (2007) DOI:10.1093/rpd/ncm048.
27. S. Dubinski, O. Presler, Z.B. Alfassi, 'The Localization of a Neutron Source in an Unknown Homogeneous Medium; the 3D Case', *Nucl. Instrum. Methods Phys. Res.*, **574A**, 315 (2007).
28. O. Presler, S. Dubinski, Z.B. Alfassi, 'The Determination of a Neutron Source Position in an Unknown Homogenous Medium, the 3D Case', *Radiat. Prot. Dosim.*, (2007) DOI:10.1093/rpd/ncm050.
29. S. Dubinski, O. Presler, Z.B. Alfassi, 'An Iterative Method for the Localization of a Neutron Source in a Large Box (container)', *Nucl. Instrum. Methods Phys. Res.*, **583A**, 439 (2007).

Charged Particle Activation Analysis

Karel Strijckmans
Laboratory Analytical Chemistry, Institute Nuclear Sciences, Gent University, Belgium

1 Introduction **2**
2 Interaction of Charged Particles with Matter **2**
2.1 Stopping Power 2
2.2 Range 2
2.3 Stopping Power and Range Data 3
3 Charged Particle Induced Nuclear Reactions **3**
3.1 *Q*-value and Threshold Energy 3
3.2 Coulomb Barrier 4
3.3 Nuclear Reaction Cross-section 4
3.4 Cross-section Data: Excitation Function and Activation Curve 4
3.5 Radionuclides Formed by Charged Particle Induced Reactions 5
4 Interferences **6**
4.1 Nuclear Interference 6
4.2 Spectral Interference 7
4.3 Matrix Interference 7
5 Standardization **7**
5.1 General 8
5.2 Trace Element Determination in the Bulk of a Sample 9
5.3 Surface Characterization of Thin Layers 10
5.4 Conclusions 10
6 Charged Particle Irradiation **10**
6.1 Accelerator 10
6.2 Heat Release in a Target 11
6.3 Target Holder 12
6.4 Direct Beam Intensity Monitoring 12
6.5 Quantitative Beam Intensity Monitoring 13
7 Activity Measurement **14**
7.1 Radionuclides Formed and Decay Radiation Measured in Charged Particle Activation Analysis 14
7.2 γ-Spectrometry 14
7.3 Positron Counting 15
8 Procedure **16**
8.1 General 16
8.2 Trace Elements in the Bulk of a Sample 17
8.3 Surface Characterization of Thin Layers 18
8.4 Radiochemical Separation 18
9 Data Handling **20**
9.1 γ-Spectrum Analysis 20
9.2 Decay Curve Analysis 20
9.3 Quantitation 21
10 Perspectives **21**
10.1 Charged Particle Activation Analysis 21
10.2 Charged Particle Activation for the Development of Other Analytical Methods 21
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **22**

Charged particle activation analysis (CPAA) is an analytical method for elemental analysis. It is based on charged particle (CP) induced nuclear reactions producing radionuclides, that are identified and quantified by their characteristic decay radiation. CPAA allows trace element determination in the bulk of a solid sample, as well as characterization of a thin surface layer, i.e. determination of mass thickness or composition.

Advantages of CPAA are: (1) good precision, (2) low detection limit, (3) outstanding accuracy, (4) absolute, i.e. reference method, (5) independent, i.e. based on a totally different principle than more common methods for elemental analysis (atomic/optical emission and mass spectrometry), and hence not subject to the same systematic errors, (6) traceability, (7) not subject to surface contamination for trace element determination in the bulk, (8) if instrumental analysis is not possible it is not subject to reagent blank errors, and (9) errors due to nonquantitative yield (and even nonreproducible yield) can be corrected for. Disadvantages of CPAA are its inherent complexity and costs, it is not suitable for liquid samples, and heating occurs during irradiation. Speciation, depth profiling, or scanning is not possible.

This article covers (1) the principles on the slowing down of CPs in matter, nuclear reactions, and interferences; (2) experimental details concerning irradiation and activity measurement; (3) conceptual and experimental procedures and data handling, from the problem to the final result; (4) perspectives. References to relevant databases are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

made. A systematic overview of all possible CPAA applications is beyond the scope of this article.

1 INTRODUCTION

CPAA is an analytical method for elemental analysis. Speciation (i.e. discrimination between different oxidation states of the analyte element or between different compounds) is *not* possible. CPAA is based on CP induced nuclear reactions producing radionuclides and their characteristic decay radiation is measured. The radionuclide is identified by measuring its energy and/or half-life, i.e. qualitative analysis. Quantitative analysis is performed by measuring the number of particles or photons emitted, i.e. radioactivity.

CPAA provides a good precision, low detection limits, and accuracy. It can be considered as an absolute (reference) and an independent method for elemental analysis of solid samples. There is no need for standards that are comparable to the samples, or for standards where the analyte content is determined by another analytical method. Traceability conditions are fulfilled, except for the stopping power correction. CPAA is based on a totally different principle than more common methods for elemental analysis (e.g. atomic or optical emission and mass spectrometry), and hence is not subject to the same systematic errors. It is not subjected at all to errors related to dissolution of the sample and chemical separation (e.g. preconcentration). If instrumental analysis is not possible, then radiochemical CPAA can be applied. Errors due to reagent blanks do not exist. Errors due to surface contamination (e.g. determination of bulk oxygen in aluminum) can be avoided easily. Finally, errors due to nonquantitative yield, and even nonreproducible yield, can be corrected for.

CPAA can be applied for trace element determinations as well as for characterization of thin surface layers. It has proven its outstanding capabilities for the determination of trace light elements (boron, carbon, nitrogen, and oxygen) in the bulk of high purity metals and semiconductors. Recently, CPAA has been applied for the characterization of surface layers, such as the determination of mass thickness or composition (stoichiometry). The disadvantages of CPAA are its inherent complexity and cost, it is not suitable for liquid samples, and heating occurs during irradiation. Neither depth profiling nor scanning is possible for surface characterization.

2 INTERACTION OF CHARGED PARTICLES WITH MATTER

Most CPs do not induce nuclear reactions and are slowed down in matter. The energy loss is described by the stopping power, and the penetration depth in matter by the range.

2.1 Stopping Power

The stopping power of a target for a CP is the energy loss of the CP for an amount of target atoms. In CPAA it is most convenient to use mass stopping power, i.e. the energy loss per unit mass thickness, as given in Equation (1).

$$S = \frac{-\mathrm{d}E}{\mathrm{d}x} \tag{1}$$

where S is the mass stopping power, in $\mathrm{MeV\,g^{-1}\,cm^2}$; E is the energy, in MeV; and $x = \rho l$, the mass thickness, in $\mathrm{g\,cm^{-2}}$ (ρ = mass density in $\mathrm{g\,cm^{-3}}$ and l = thickness in cm).

The stopping power is determined by the nature of the CP, energy of the CP, and nature of the target, as given in Equation (2).

$$S_\mathrm{a} = Z_\mathrm{a}^2 S_\mathrm{p} \left(\frac{E_\mathrm{a}}{A_\mathrm{a}} \right. \tag{2}$$

where Z_a is the atomic number and A_a is the mass number. The subscripts ''a'' and ''p'' refer to a CP and proton, respectively. Figure 1(a) and Equation (2) show that the stopping power for 20-MeV deuterons is equal to the one for 10-MeV protons, and the stopping power for 30-MeV helium-3 particles and 40-MeV alpha particles is four times that of 10-MeV protons. The stopping power decreases with the CP energy ($\sim 1/E$) (Figure 1b) and for increasing atomic number of the target (Figure 1c). This decrease is steep for low-Z targets.

Stopping power data for elemental matter can be calculated based on the Bethe formula. For mixtures, the additivity rule of Bragg and Kleeman (Equation 3) is applied.

$$S = \sum_{i=1}^{n} w_i S_i \tag{3}$$

where w_i is the mass fraction of the ith component, S_i is the mass stopping power of the ith component, and n is the number of components. For compounds, deviations from the Bragg rule may occur due to chemical binding effects. For the energy range of interest in CPAA (>1 MeV) these deviations can be neglected.

2.2 Range

As a consequence of the stopping power phenomenon, CPs lose their energy and are stopped at a depth called the ''range''. From Equation (1) the mass range, R, for a

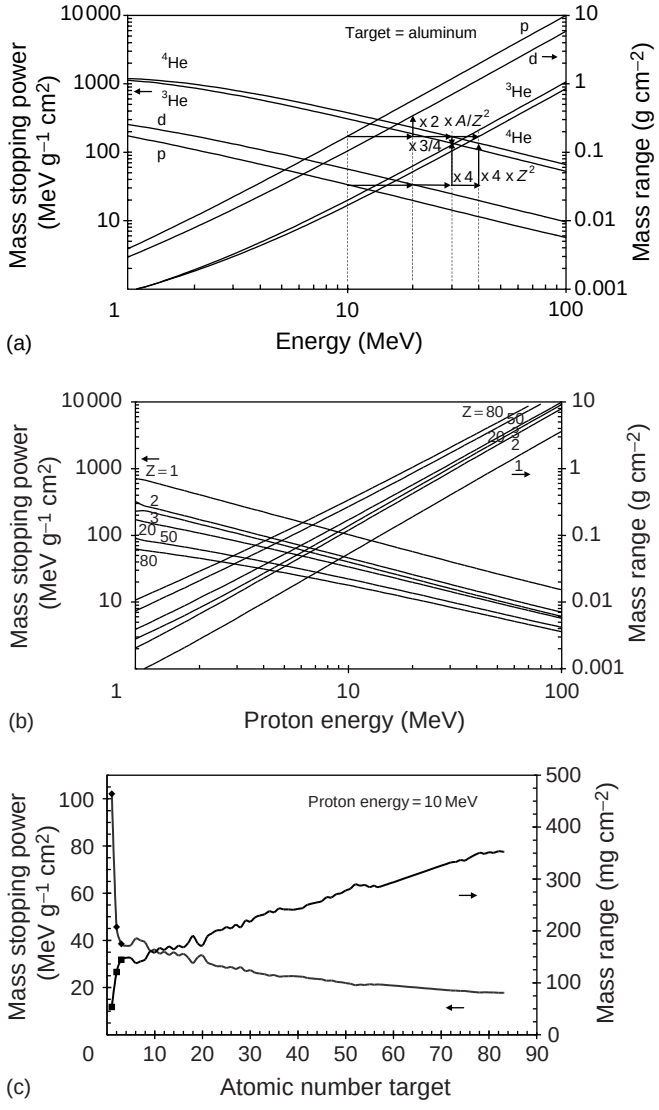

Figure 1 Mass stopping power and mass range for: (a) protons, deuterons, helium-3, and α-particles in an aluminum target as a function of the energy; (b) for elemental targets with atomic numbers $Z = 1, 2, 3, 20, 50$, and 80 as a function of the proton energy; and (c) for 10-MeV protons as a function of the atomic number of an elemental target.

CP with incident energy E_i can be calculated according to Equation (4).

$$R = \int_0^R \mathrm{d}x = \int_{E_i}^0 \frac{\mathrm{d}E}{\mathrm{d}E/\mathrm{d}x} = \int_0^{E_i} \frac{\mathrm{d}E}{S} \tag{4}$$

where R is the mass range, in $\mathrm{g\,cm^{-2}}$; and E_i is the incident CP energy, in MeV. Consequently the range is also determined by the nature of the CP, the energy of the CP, and the nature of the target, as given in Equation (5).

$$R_a = \frac{A_a}{Z_a^2} R_p \left(\frac{E_a}{A_a} \right. \tag{5}$$

Figure 1(a) and Equation (5) show that the range of 20-MeV deuterons is twice that of 10-MeV protons, the range of 30-MeV helium-3 particles is 75% that of 10-MeV protons, and the range of 40-MeV α-particles is equal to that of 10-MeV protons. The range increases with the CP energy ($\sim E^2$) (Figure 1b) and for increasing atomic number of the target (Figure 1c). The increase is steep for low-Z targets.

For CPAA the CP interaction with matter can be described simply: CPs traverse the target undeflected (i.e. nearly straight lines), the beam intensity (the number of CPs per unit time) is almost unchanged, and all CPs are stopped at the same depth in the target (at the range). Consequently, linear versus projected range, and range straggling, should not be considered for CPAA.

2.3 Stopping Power and Range Data

Stopping power and range data can be obtained from the SRIM (Stopping and Range of Ions in Matter) PC software package of Ziegler, IBM Research.[1] The principles are described by Ziegler et al.[2] To calculate ranges expressed in length (as in SRIM) the mass density is required. For mixtures and compounds (not listed in the compound dictionary) the mass density should be introduced, by the user, into the SRIM package. (An updated review of *Nuclear Data for Charged Particle Activation* is available at http://allserv.rug.ac.be/~kstrykmn/nuclear.)

3 CHARGED PARTICLE INDUCED NUCLEAR REACTIONS

CPAA is based on a nuclear reaction $\mathrm{A + a \longrightarrow B + b + } Q$, or A(a,b)B, where A is the stable target nuclide (at rest), "a" is a CP (accelerated), B is a radionuclide, "b" is the particle(s) or photon emitted, and Q is the Q-value. Two important characteristics are the minimum CP energy required to induce such a nuclear reaction, and the probability that this reaction will proceed. As CPs are slowed down when traversing matter, as described in section 2, the latter characteristic should be known as a function of the CP energy.

3.1 *Q*-value and Threshold Energy

The Q-value is the energy released by one single nuclear reaction. It is related to the difference between the resting masses of reactants and products, as given in Equation (6).

$$Q = (m_A + m_a - m_B - m_b)c^2 \tag{6}$$

where m is the nuclidic mass and c is the velocity of light. Nuclidic masses can be replaced by atomic masses,

as electron masses are cancelled, at least if this also is done for ''a'' and ''b'' (e.g. for a (p,α) reaction the atomic masses of ^{1}H and ^{4}He, respectively, are used). A nuclear reaction can be endoergic ($Q < 0$) or exoergic ($Q > 0$). To induce endoergic reactions the CPs should have a minimum energy, i.e. the threshold energy E_t, which is slightly higher than $-Q$. The compound nucleus, formed by collision of the CP ''a'' and the target nucleus A, retains a fraction of the kinetic energy of the CP, i.e. $m_a/(m_A + m_a)$. This fraction is ''lost'' to compensate for the shortfall of mass or energy for an endoergic reaction.

Approximating the nuclidic mass by the mass number A for the recoil energy correction, the threshold energy is given by Equation (7).

$$E_t = -\frac{A_A + A_a}{A_A} Q \qquad \text{for } Q < 0$$

$$E_t = 0 \qquad \text{for } Q \geq 0 \tag{7}$$

where E_t is the threshold energy. For exoergic nuclear reactions the threshold energy is zero by definition.

Q-values and threshold energies can be calculated from the Audi and Wapstra ''adopted masses'' 1993[3] or 1995[4] with the on-line program Q-calc, at the Lund Nuclear Data WWW Service (Lund University, Sweden), at the National Nuclear Data Center (NNDC, Brookhaven National Laboratory, USA), and at the International Atomic Energy Agency (IAEA, Austria),[5] or as a Fortran-77 program[6] for a VAX-VMS system.

3.2 Coulomb Barrier

The Coulomb barrier also determines the minimum CP energy needed to induce a nuclear reaction. The coulombic repulsive force between the target nucleus and the CP dominates at ''large'' distances, and increases when the CP approaches the target nucleus. At some particular distance (i.e. the sum of the radii of CP and target nucleus) the attractive nuclear force balances the Coulomb repulsive force. The decrease in the CP kinetic energy is given by the Coulomb barrier, as in Equation (8).

$$E_C \approx \frac{A_A + A_a}{A_A} \frac{Z_A Z_a}{\sqrt[3]{A_A} + \sqrt[3]{A_a}} \tag{8}$$

where E_C is the Coulomb barrier, in MeV; and $\sqrt[3]{A}$ is proportional to the nucleus radius. The Coulomb barrier for ^{3}He and ^{4}He particles is about twice as high as that for protons or deuterons, and it increases with the atomic number of the target nucleus, roughly $Z_A^{2/3}$, as shown in Figure 2.

The kinetic energy – the CP lost by the Coulomb barrier – is released again when a nuclear reaction occurs. Consequently, the Coulomb barrier influences the energetics of a nuclear reaction only in that a CP must have a kinetic energy higher than the Coulomb barrier before the reaction can occur. Quantum mechanical treatment of the problem explains that the reaction probability (see section 3.3) for CPs with a kinetic energy lower than the Coulomb barrier is not exactly zero, but very low, and rapidly drops as the CP energy decreases.

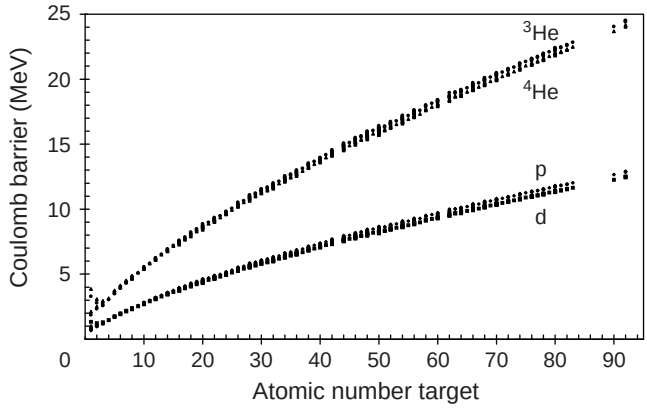

Figure 2 Coulomb barrier of all stable nuclides for proton, deuteron, helium-3, and helium-4 particle induced reactions as a function of the atomic number of the target nucleus.

3.3 Nuclear Reaction Cross-section

The probability of a nuclear reaction is expressed as the (nuclear reaction) cross-section, which has the dimensions of area. This originates from the simple picture that the probability for a reaction between the target nucleus and the incident CP is proportional to the geometric cross-section that the target nucleus presents to a beam of CPs. The average cross-section is somewhere near $10^{-28}\,m^2$. The barn ($1\,b = 10^{-28}\,m^2$) is used as a unit for nuclear reaction cross-section.

The cross-section for a particular reaction (also called the partial reaction cross-section) depends on the energy of the CP. It is zero for CPs below the threshold energy of that reaction. If the Coulomb barrier is higher than the threshold energy (i.e. always the case for exoergic reaction, as $E_t = 0$ for $Q > 0$), then the cross-section is very low in the energy interval between the threshold energy and the Coulomb barrier. For CP energies exceeding the threshold energy and Coulomb barrier, the cross-section increases up to a maximum (typically 1 b), and decreases as more complex reactions become competitive. Two annotated examples are given in Figure 3(a) and (b).

3.4 Cross-section Data: Excitation Function and Activation Curve

As CPs are slowed down in matter, knowledge of the cross-section as a function of the energy is required

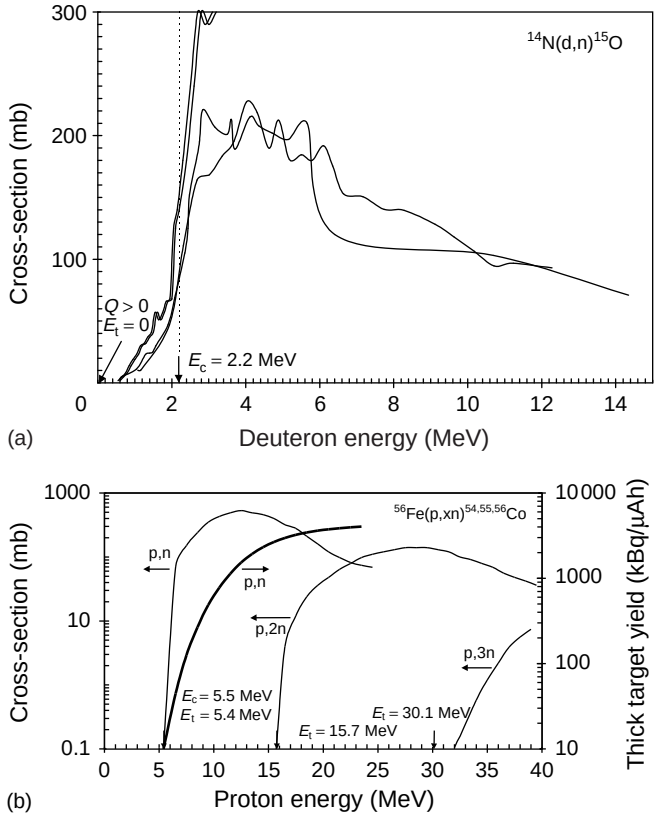

Figure 3 Excitation functions. (a) The $^{14}N(d,n)^{15}O$ reaction is exoergic ($Q > 0$), consequently its threshold energy is zero. Below the Coulomb barrier (2.2 MeV) the cross-section is low, but not zero. It increases steeply around the Coulomb barrier energy. (b) The $^{56}Fe(p,n)^{56}Co$ reaction is endoergic ($Q = -5.3$ MeV), hence its threshold energy equals 5.4 MeV. The Coulomb barrier equals 5.5 MeV. The cross-section is zero below 5.4 MeV, increases to a maximum at 12–13 MeV, and decreases at higher energies, as the $^{56}Fe(p,2n)^{55}Co$ reaction ($E_t = 15.7$ MeV) becomes competitive. For the (p,2n) reaction the reaction cross-section also increases up to a maximum at 27–30 MeV, and decreases at higher energies, as the $^{56}Fe(p,3n)^{54}Co$ reaction ($E_t = 30.1$ MeV) becomes competitive. For thick target yield (TTY), see section 3.5.

for CPAA. However, absolute data are not required (sections 5.2 and 5.3). Therefore the *excitation function* (or curve) indicates the absolute cross-section as a function of the energy, while the *activation curve* indicates a relative data set.

Experimental excitation functions can be obtained from the EXFOR (or CSISRS) database at NNDC mirrored at IAEA and the Nuclear Energy Agency (NEA), France.[7] For the data sets in Figure 3(a), obtained from this database, the agreement is fairly good, but this is *not* representative for the vast majority of CP induced reactions. Actually the $^{14}N(d,n)^{15}O$ reaction has been studied in great detail, because it is extremely important for positron emission tomography (PET), an imaging technique for quantifying the oxygen metabolism, blood flow and blood volume in nuclear medicine.

If no experimental data are available, excitation functions can be estimated using the very simple equations and data from Keller et al.[8]

The experimental procedure needed to obtain an activation curve is quite simple if the target element (A) is available as "thin" foil. "Thin" means much smaller than the range. A stack of "thin" foils (total thickness $\geq$ range) is irradiated, and the foils are measured individually. After decay correction the count rate measured is proportional to the cross-section, as will be shown in Equation (17). Excitation function data require absolute activity measurements, absolute beam intensity monitoring, and absolute energy measurement of the incident CP beam. However, an activation curve (relative cross-section data) is sufficient for CPAA.

In contrast to neutron induced reactions, no experimental data are available for most of the CP induced reactions. Additionally, if different data sets are available, their agreement is often unsatisfactory. Therefore approximate methods for standardization have been developed in CPAA.

3.5 Radionuclides Formed by Charged Particle Induced Reactions

Which radionuclides are formed by irradiation of a particular matrix by a CP with a particular energy, and what is their approximate activity level?

The first question can be answered by observing a chart of nuclides. This represents all nuclides with relevant data for CPAA as chemical symbol, mass number, isotopic abundance (stable nuclides), half-life, decay mode, and decay energies (radionuclides). For proton irradiation, for example, the reaction types given in Figure 4 should be applied to all stable isotopes of all elemental components of the matrix. In the inventory of possible nuclear reactions two categories can be rejected: those leading to stable nuclides, and those with a threshold energy higher than the CP energy. Reactions for which the Coulomb barrier is higher than the CP energy induce only low activity levels, but cannot be excluded. An annotated example is given in Figure 4 for a 10-MeV proton irradiation of nitrogen.

The induced activity is proportional to the number of CPs per time unit, irradiation time (not linearly), number of target nuclides, and reaction cross-section. The number of target nuclides is inversely proportional to the relative atomic mass of the element, and proportional to the mass fraction of the element in the target, the isotopic abundance of the target nuclide, and the activated depth, i.e. the range of the target at the incident energy minus

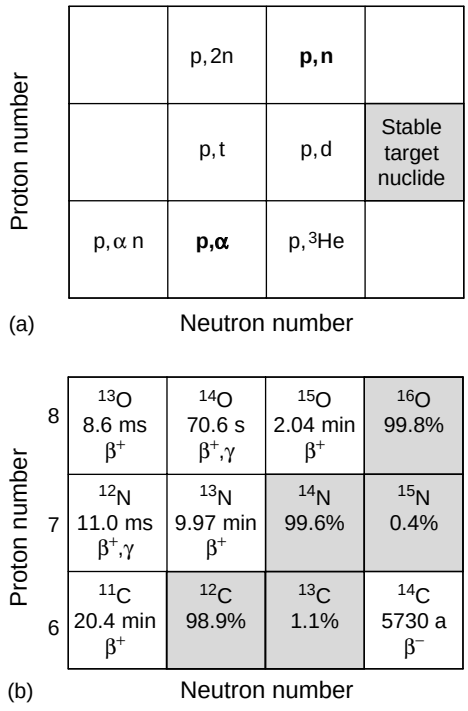

Figure 4 (a) Reaction scheme for proton irradiation to be applied to a chart of nuclides. In general, for a proton energy below 10 MeV only the (p,n) and (p,α) reactions have to be considered; for a proton energy up to 20 MeV the (p,2n), (p,d), (p,t), (p,^{3}He), and (p,αn) reactions also have to be taken into account. (b) Part of a chart of nuclides: irradiation of nitrogen by 10-MeV protons mainly induces ^{14}O and ^{11}C by a (p,n) and (p,α) reaction, respectively, from ^{14}N, i.e. the most abundant stable isotope of nitrogen, and to a negligible extent ^{15}O by a (p,n) reaction from ^{15}N. The (p,α) reaction has not been considered for CPAA as it produces ^{12}C, a stable nuclide.

the range at the threshold energy or the Coulomb barrier, whichever is higher. This will be treated in more detail in section 5.1. A rough estimate of the induced activity, i.e. the thick target yield (TTY), can be calculated from Equation (9).

$$A \approx 380 \frac{1}{Z_a} I \frac{1}{A_r} w\theta (1 - e^{-\lambda t_i}) \overline{\sigma} \left(R_{E_i} - R_{E_{t/C}} \right) e^{-\lambda t_w} \quad (9)$$

where

A = activity, in kBq,
Z_a = atomic number of the CP (1 for p and d; 2 for ^{3}He and ^{4}He (α-particles)),
I = beam intensity, in μA,
A_r = relative atomic mass of the target element,
w = mass fraction of the target element in the matrix, in %,
θ = isotopic abundance, in atom %,
λ = $\ln(2)/t_{1/2}$, decay constant of the radionuclide (where $t_{1/2}$ is the half-life),
t_i = irradiation time,
t_w = waiting time,
$\overline{\sigma}$ = average reaction cross-section, in mb (generally, $100 < \overline{\sigma} < 1000$),
R = mass range, in g cm^{-2} (section 2.2 and Figure 1),
E_i = incident energy,
$E_{t/C}$ = max(E_t, E_C) (sections 3.1 and 3.2).

An example of the TTY as a function of the incident energy is given in Figure 3(b) for the (p,n) reaction.

A chart of nuclides is available as:

- a wall chart or booklet,
  - Kernforshungszentrum Karlsruhe (Germany), (see reference 9)
  - General Electric & KAPL, Knoll Atomic Power Lab. (USA),[10]
- a PC program,
  - NuChart from IAEA (1994), no further updates,[11]
  - Nucleus from AMDC, Atomic Mass Data Center (France),[12]
  - MacNuclide from Center for Nuclear Information Technology (for Mac),[13]
- a PC program in combination with a database on the internet, Isotope Explorer (formerly VuENSDF) from LBNL, Lawrence Berkeley National Laboratory (USA) and Lund University (Sweden),[14]
- on-line databases on the internet,
  - Lund Nuclear Data WWW Service from Lund University (Sweden),[15]
  - jvNuBase from AMDC, Atomic Mass Data Center (France),[16]
  - KAERI Table of the Nuclides from KAERI, Korea Atomic Energy Research Institute (also mirrored).[17]

4 INTERFERENCES

The determination of an analyte element is based on the reaction A(a,b)B, where A is a stable nuclide of that analyte element and B is a radionuclide (section 3). Three different types of interference can be distinguished: nuclear, spectral, and matrix.

4.1 Nuclear Interference

A radionuclide B is also formed from an element other than the analyte element by a nuclear reaction C(c,d)B. C is a stable nuclide of the interfering element.

The CP "c" is mostly identical to "a", but nuclear interferences induced by secondary particles, e.g. fast or thermalized neutrons produced by (a,n) reactions, have to be considered also. Nuclear reactions can only be avoided by a proper choice of the CP energy if the threshold energy E_t of the interfering reaction C(c,d)B exceeds both the threshold energy and the Coulomb barrier of the analyte reaction A(a,b)B. To obtain high sensitivity, the CP energy is chosen at or just below the threshold energy of the interfering reaction.

Interference by the $^{14}N(d,t)^{13}N$ reaction (Figure 4b) could affect the determination of carbon by the $^{12}C(d,n)^{13}N$ reaction. As the threshold energy of the interfering reaction (4.9 MeV) exceeds that of the analyte reaction (0.3 MeV), as well as the Coulomb barrier (2.0 MeV), carbon can be determined sensitively and interference free by irradiation with 4.9-MeV deuterons. It is obvious that the difference in Coulomb barrier is largely insufficient to avoid nuclear interferences, and that the reaction $^{13}C(d,2n)^{13}N$ does not interfere with the determination of carbon.

4.2 Spectral Interference

A radionuclide D is formed by the nuclear reaction C(c,d)D and the measurement does not resolve the activity of B and D. Spectral interferences can be avoided by:

(1) The proper choice of the incident CP energy with respect to the threshold energy and the Coulomb barrier (analogous to nuclear interference).

(2) Selective measurement of the radionuclide D with respect to radionuclide B by spectrometry or decay curve analysis (sections 7 and 9.2). The former is possible if the γ-lines differ by more than the energy resolution of the spectrometer used. The latter is possible if the half-lives of the radionuclides C and D are sufficiently different (about a factor of 2 for comparable activities). If D is short-lived, i.e. its half-life is $\geq$10 times lower as compared to B, then selective measurement (interference $< 0.4\%$) is possible for an irradiation and waiting time equal to one half-life of B (for comparable relative atomic mass, mass fraction, isotopic abundance, cross-section, and range, Equation 9).

(3) Radiochemical separation of B from D, if both are not radioisotopes, i.e. radionuclides of the same element (section 8.4).

Nitrogen can be determined by the $^{14}N(p,n)^{14}O$ as well as by the $^{14}N(p,\alpha)^{11}C$ reaction (Figure 4b). The former reaction leads to ^{14}O, a β^+,γ-emitter. Measuring the 2313-keV γ-ray with a Hewlett-Packard Ge spectrometer, the former reaction only undergoes interference by other γ-emitters within the energy resolution ($\approx$2 keV) of that spectrometer. There could be interference with the latter reaction by, for example, the $^{13}C(p,n)^{13}N$, $^{14}N(p,n)^{14}O$, $^{15}N(p,n)^{15}O$, and $^{16}O(p,\alpha)^{13}N$ reactions, all leading to β^+ (positron) emitters. Measurement of their annihilation radiation (section 7.1) is aspecific and thus could cause spectral interference. As the threshold energies (3–6 MeV) and the Coulomb barriers are not sufficiently different, the interference cannot be avoided by a proper choice of the proton energy. Measurement of the activity as a function of time followed by decay curve analysis (section 9.2) leads to the initial activity of each component if the activity levels are comparable within 1 to 3 orders of magnitude. It is obvious that the other reactions [(p,n), (p,α)] on the analyte element (N) also cause spectral interference (in contrast to nuclear interference). Alternatively, ^{11}C can be radiochemically separated by combustion in oxygen and trapping of the carbon dioxide formed, as the other radionuclides are not radioisotopes of carbon. The latter is called radiochemical analysis (section 8.4) or determination, in contrast to instrumental analysis.

4.3 Matrix Interference

Matrix interference is a special case of spectral interference. As CPAA is primarily a method for trace element determinations or thin-layer characterization, the activity level of the main matrix components could be several orders of magnitude higher than the activity of the analyte element, e.g. 10^6 for a trace element determination at the $\mu g\,g^{-1}$ or ppm level. Then instrumental analysis is almost impossible, even for quite different γ-energies and/or half-lives. Instrumental analysis is only possible for a proper choice of the CP energy, as an alternative for a highly specific radiochemical separation. It is obvious that the determination of light elements (boron, carbon, nitrogen, oxygen) in a high-Z sample is favored by the Coulomb barrier (section 3.2).

5 STANDARDIZATION

Standardization or calibration in CPAA is performed in a relative way, i.e. a sample and a standard are both irradiated and measured. To preserve the advantage of CPAA as an "absolute" or "independent" method, there is no need for a standard similar to the sample and certified for the analyte concentration or mass thickness. The standard is pure elemental matter or a pure compound with exactly known stoichiometry (section 8.1). In a relative method there is no need for absolute data, and errors are cancelled.

For trace element determination in the bulk of a sample, the experimental data (activity) are related to the mass fraction (or concentration) expressed in, for example, mg kg^{-1} (SI), µg g^{-1}, ppm (parts per million) or µg kg^{-1}, ng g^{-1}, ppb (parts per billion).

Surface characterization of thin layers stands for the determination of the partial mass thickness. For an Al_2O_3 layer the partial/total mass thickness is the amount of Al/Al_2O_3 per surface unit, expressed in µg cm^{-2}. From the partial mass thickness the total mass thickness, or thickness or composition (mass fraction or stoichiometry), can be calculated (section 10).

As standardization for trace element determinations in the bulk of a sample and surface characterization have a common and a specific part, the following is divided into three sections. The word ''target'' will be replaced by ''sample'' or ''standard'' in sections 5.2 and 5.3.

5.1 General

The number of radionuclides produced per unit time by irradiation of an infinitesimally thin target is determined by the balance of the increase due to nuclear reactions (proportional to the number of CPs per time unit (i.e. beam intensity, the cross-section and the number of target nuclides per surface unit) and the decrease due to radioactive decay ($-\lambda N$) (Equation 10).

$$\frac{\mathrm{d}N}{\mathrm{d}t} = I\sigma n\,\mathrm{d}l - \lambda N \tag{10}$$

where

N = number of radionuclides,
t = time,
I = number of CPs per unit time or beam intensity,
σ = partial nuclear reaction cross-section,
n = number of target nuclides per unit volume,
l = thickness of the target,
λ = $\ln(2)/t_{1/2}$, decay constant of the radionuclide formed (where $t_{1/2}$ = half-life).

From Equation (10), the number of radionuclides formed and the induced activity after an irradiation time t_i can be calculated. On condition that the beam intensity is constant, this yields Equation (11).

$$A = \lambda N = Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\sigma\,\mathrm{d}x \tag{11}$$

where

A = activity in Bq,
w = mass fraction of the activated element in the target,
M = molar mass of the activated element,
$\mathrm{N_A}$ = Avogadro's constant,
θ = isotopic abundance of the target nuclide,
t_i = irradiation time.

A CP beam is stopped completely in a ''thick'' target, in which the mass thickness D is at least equal to the range R. To calculate the activity for such a ''thick'' target one has to integrate $\sigma\,\mathrm{d}x$ in Equation (11), using the stopping power (Equation 1), as the cross-section has to be expressed as a function of the energy. The calculation is given by Equation (12).

$$\int_0^R \sigma\,\mathrm{d}x = \int_{E_i}^{0} \frac{\sigma\,\mathrm{d}E}{\mathrm{d}E/\mathrm{d}x} = \int_{E_t}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \tag{12}$$

where R is the mass range of the CP in the target, in g cm^{-2}; E_i is the incident CP energy, in MeV; E_t is the threshold energy, in MeV; and S is the mass stopping power, in MeV g^{-1} cm^2. The integration limit $E = 0$ (corresponding to $x = R$) may be replaced by the threshold energy E_t, as the cross-section is zero for $E < E_t$.

If the excitation function is not known, an energy $E_{t/C}$ can be defined as the maximum of threshold energy and Coulomb barrier. The cross-section can be supposed to be zero below $E_{t/C}$, while above $E_{t/C}$ it is constant. Then the following approximation (Equation 13) is obtained from Equations (2) (range definition) and (12).

$$\int_0^R \sigma\,\mathrm{d}x = \int_{E_t}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \approx \overline{\sigma}\left(R_{E_i} - R_{E_{t/C}}\right) \tag{13}$$

if $\sigma(E) = 0$ for $0 \le E \le E_{t/C}$; and $\sigma(E) = \overline{\sigma}$ for $E_{t/C} \le E \le E_i$; where $\overline{\sigma}$ = average cross-section, and $E_{t/C} = \max(E_t, E_C)$.

From Equations (11), (12), and (13) the activity induced in a thick target (i.e. TTY) is given by Equation (14), which approximates to Equation (15).

$$A = Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\int_{E_t}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \tag{14}$$

$$A \approx Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\overline{\sigma}\left(R_{E_i} - R_{E_{t/C}}\right) \tag{15}$$

The target is assumed to be homogeneous, and the decrease in beam intensity (due to nuclear reactions) is negligible. Consequently, neither w nor I is a function of x or E.

For a ''thin'' target, i.e. a target that nearly reduces the CP energy, where the mass thickness D is much lower than the mass range R, the induced activity is given by Equation (16), which approximates to Equation (17).

$$A = Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\int_{E_o}^{E_i} \frac{\sigma\,\mathrm{d}E}{S} \tag{16}$$

$$A \approx Iw\frac{1}{M}\mathrm{N_A}\theta\left(1 - \mathrm{e}^{-\lambda t_i}\right)\sigma_0 D \tag{17}$$

where $E_o \approx E_i$, the outgoing energy, in MeV; σ_0 is the cross-section for $E_o < E < E_i$; and $D \ll R$, the mass thickness of the ''thin'' target, in g cm^{-2}.

5.2 Trace Element Determination in the Bulk of a Sample

A relative method is applied, whereby a ''thick'' sample and a ''thick'' standard are irradiated and measured separately. As the nature of sample and standard is different, a correction, F, has to be made owing to the different stopping power (or range) of sample versus standard. Applying Equations (14) and (15) for, respectively, a sample ''x'' and a standard ''s'', and assuming the sample and standard to have equal natural abundance, the mass fraction of the analyte in the sample yields Equation (18), in which F is given by Equation (19), which approximates to Equation (20). (To comply with SI the ''mass fraction'' w is used, while in the CPAA literature often ''concentration'' c is used.)

$$w_x = w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} F \tag{18}$$

where the subscripts x and s refer to a ''thick'' sample, and a ''thick'' standard, respectively.

$$F = \frac{\int_{E_t}^{E_i} \frac{\sigma \, dE}{S_s}}{\int_{E_t}^{E_i} \frac{\sigma \, dE}{S_x}} \tag{19}$$

$$F \approx \frac{\left(R_{E_i} - R_{E_{t/C}}\right)_s}{\left(R_{E_i} - R_{E_{t/C}}\right)_x} \tag{20}$$

The mass fraction (concentration) of the analyte in the sample w_x can be calculated from:

- the mass fraction of the analyte element in the standard w_s,
- the ratio of the activity in sample and standard at the end of irradiation A_x/A_s,
- the ratio of the beam intensity for standard and sample I_s/I_x,
- the decay constant of the radionuclide formed λ,
- the F-factor, i.e. a correction for different stopping power of sample versus standard.

The F-factor can be calculated exactly by Equation (19), requiring:

- relative cross-section data σ from the threshold energy E_t up to the incident energy E_i,
- relative stopping power data for standard and sample S_s, S_x in the same energy interval.

If no cross-section data are available, the F-factor can be approximated by Equation (20). This approximation yields quite good results if one or more of the following conditions (in decreasing order of importance) are fulfilled:

- the atomic number of sample and standard are similar,
- the threshold energy or the Coulomb barrier for the nuclear reaction is high,
- the incident energy is high.

The ''atomic number'' of a compound is, in principle, that of the element with the same stopping power. Practically, the atomic numbers of the elemental components can be averaged: mass fraction weighed. For simplicity Equation (13) was derived supposing a ''step-wise'' excitation function, which does not agree with reality. A detailed study[18] showed that this approximation can also be obtained by assuming that S_x/S_s is a constant as a function of the energy. If the above mentioned conditions are fulfilled, the latter assumption holds better, as can be observed from the parallel lines in Figure 1(b).

Several approximate methods, not requiring cross-section data, have been proposed and evaluated.[18] Equation (19) is known as the first method of Ricci and Hahn.[19] The second method of Ricci and Hahn[20] is fundamentally false. The method of Chaudri et al.[21] yields nearly the same approximation as the first method of Ricci and Hahn. Chaudri's method makes use of stopping power data instead of range data, as shown in Equation (21).

$$F \approx \frac{\left(S_{E_M}\right)_x}{\left(S_{E_M}\right)_s} \tag{21}$$

where $E_M = (E_i + E_{t/C})/2$. The ''average stopping power method'' is an outstanding approximate method with a negligible error that can be calculated.[22,23] However it needs cross-section data, and therefore there is no reason to prefer it to Equation (19). The method has more intellectual than practical merits.

The internal standardization method, not requiring stopping power data for the sample, has more practical applications.[24] The major composition of a sample is often not known, or the composition changes during irradiation. Samples can lose their organic fraction due to heating, and consequently positive errors are made because the stopping power of hydrogen is 3–5 times higher than for the other elements (Figure 1c). Also, approximate methods not requiring cross-section data have been proposed.

The ''two reactions method'' does not require stopping power data.[25] Its main advantage is that accuracy of the

method is not influenced by the accuracy of the stopping power data. Experimentally, the method is quite complex.

Also, the standard addition method does not require stopping power data.[26,27] As CPAA is mainly intended for solid samples, the applicability of this method is limited to powdered samples.

A critical review is given by Strijckmans.[28]

5.3 Surface Characterization of Thin Layers

As for trace element determinations, a relative method is also applied. A ''thin'' surface layer is irradiated and measured together with a ''thin'' or ''thick'' standard. From Equations (16) and (17) the partial mass thickness, using a ''thin'' standard, yields Equation (22), which approximates to Equation (23).

$$w_x D_x = w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} \int_{E_{os}}^{E_i} \frac{\sigma/\sigma_0 \, dE}{S} \qquad (22)$$

$$w_x D_x \approx w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} D_s \qquad (23)$$

where wD is the partial mass thickness; $D \ll R$, the total mass thickness; and the subscripts x and s refer to a ''thin'' surface layer and a ''thin'' standard, respectively.

For a ''thin'' surface layer (order of magnitude μm to nm) the approximation made in Equation (17) holds. For the ''thin'' standard (order of magnitude 10 to 100 μm) one has to check if the cross-section is still constant in the energy interval $E_{o_s} - E_i$. Although Equation (23) does not contain cross-section data, a knowledge of such data, at least relatively in the above mentioned interval, is required in order to evaluate the equation's applicability.

If thin standards are not available, analogous formulae can be obtained from Equations (14), (15), and (17) for a ''thick'' standard. The partial mass thickness using a ''thick'' standard yields Equation (24), which approximates to Equation (25).

$$w_x D_x = w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} \int_{E_t}^{E_i} \frac{\sigma/\sigma_0 \, dE}{S} \qquad (24)$$

$$w_x D_x \approx w_s \frac{A_x I_s \left(1 - e^{-\lambda t_{is}}\right)}{A_s I_x \left(1 - e^{-\lambda t_{ix}}\right)} \frac{\overline{\sigma}}{\sigma_0} \left(R_{E_i} - R_{E_{t/C}}\right)_s \qquad (25)$$

where the subscripts x and s refer to a ''thin'' surface layer and a ''thick'' standard, respectively. Equation (24) requires at least relative cross-section data. From Equation (25) no approximation can be written that does not require cross-section data. The average cross-section in the energy interval $E_{t/c}$ to E_i is never comparable to the cross-section at the incident energy (actually in the very narrow interval E_o to E_i). For trace element analysis in the bulk of a sample such an approximation was possible (section 5.2, and Equations 18 and 20), and it results in accurate analyses observing the conditions mentioned in section 5.2.

Table 1 Standardization in CPAA

Determination	Cross-section data known	No cross-section data known
Trace element determination in the bulk of a sample (mass fraction)	Equations (18) and (19)	Equations (18) and (20) or (21)[a]
Surface characterization of a thin layer (mass thickness)	–	–
Thick standard	Equation (24)	–
Thin standard	Equation (22)	–
Very thin standard[b]	–	Equation (23)[c]

[a] If $Z_x \approx Z_s$ and/or $E_{t/c} \gg 0$, and/or $E_i \gg 0$.
[b] If $\sigma = \sigma_0$ for $E_o < E < E_i$.
[c] Cross-section data only in the energy interval E_o to E_i.

5.4 Conclusions

The standardization formulae to be used are summarized in Table 1. For trace element determinations in the bulk of a sample (mass fraction), accurate results can be obtained without a knowledge of cross-section data if one or more of the conditions are fulfilled. For surface characterization (mass thickness determination) accurate results can also be obtained without a knowledge of cross-section data if a ''very thin'' standard is used.

6 CHARGED PARTICLE IRRADIATION

6.1 Accelerator

CPs have to be accelerated to an energy higher than the Coulomb barrier and lower than the threshold energy of reactions that are more complex than (a,n) and (a,α). The lower limit is to obtain high reaction cross-sections, and consequently high sensitivity; the upper limit is to avoid nuclear interferences by these complex reactions. For protons and deuterons the ideal energy range should be 5–25 MeV, while for helium-3 and α-particles it should be twice as high. A cyclotron is an ideal accelerator for this energy range, although a tandem Van de Graaff also covers the lower energy range.[29]

A tandem Van de Graaff is a DC linear accelerator, the DC source being an electrostatic Van de Graaff generator. The tandem principle is related to the use of the same voltage to accelerate the particles (e.g. protons) twice. First, negative hydrogen ions (H^-) are accelerated to

5 MeV by applying 5 MV from the electrostatic generator; then the ion is stripped from its electrons, and the same reversed voltage is applied again to accelerate to 10 MeV.

The tandem principle is applied a thousand-fold in AC accelerators. A cyclotron is an AC circular accelerator. A high-frequency electric field accelerates the CPs, following a spiral path, due to the magnetic field applied. Ideally, an isochronous cyclotron should be a multi-particle and variable energy accelerator. This implies that the magnetic field and the high-frequency electric field can be tuned, and that the magnetic field as a function of the radius can be adjusted (to compensate for relativistic mass increase during acceleration). Typically, an isochronous cyclotron has radial or spiral sectors in the magnet poles, a prerequisite for beam focusing. Cyclotrons that accelerate helium-3 particles should be equipped with a recuperation system. Typically, the energy can be tuned by a factor of 5–10.

However, the vast majority of cyclotrons in the world is the two-particle and fixed energy type, the so-called ''baby cyclotron''. Their high-frequency electric and magnetic fields are fixed, and the magnetic field increases with the radius by the shape of the magnet poles. Protons or deuterons can be accelerated, the deuteron energy being half the proton energy. Such cyclotrons are negative-ion machines (H^- ions are accelerated, stripped, and the protons formed leave the magnetic field). Extraction of negative ions with a stripping foil is technically very simple compared to extraction of positive ions. These are extracted with a deflector: a strong DC electric field ''extracts'' the accelerated beam out of the magnetic field. The stripping foil (in contrast to a deflector) can be moved into the cyclotron, extracting a lower energy beam. Energy tuning of up to a factor of 2 is possible. Baby cyclotrons are widely used in nuclear medicine departments for PET, i.e. a functional imaging technique using biologically active molecules labeled with cyclotron-produced, short-lived positron emitters.

The beam requirements are not very severe. The energy should be well-defined (within 0.1 MeV is largely sufficient) to control interferences by their threshold energy and/or Coulomb barrier. The beam energy should be reproducible within several days, because this influences slightly the sensitivity. The beam intensity should be tunable from 0.01 to 5 μA, which is largely under the upper limit of these accelerators, and kept constant during irradiation (Equation 11 was calculated assuming that the beam intensity is constant). The unfavorable beam characteristics of a cyclotron versus a Van de Graaff accelerator, such as pulsed beam and energy resolution, are of no importance in CPAA.

6.2 Heat Release in a Target

CP irradiation can cause substantial heating to the sample. Indeed, standards can be irradiated for shorter times and/or at lower beam intensity. The heat release (power) in a ''thick'' target is given by the energy of one single CP multiplied by the number of CPs per unit time (Equation 26).

$$Q = E\frac{I}{Z_a e} \tag{26}$$

where

Q = heat release (power), in W;
E = CP energy, in J;
I = beam intensity, in A;
Z_a = atomic number of the CP;
e $\approx 1.6 \times 10^{-19}$ C (the elementary charge);

or more practically, according to Equation (27),

$$Q = \frac{EI}{Z_a} \tag{27}$$

where Q is the heat release (power), in W; E is the CP energy, in MeV; and I is the beam intensity, in μA. For a ''thin'' target E should be replaced by the energy decrease for the CP traversing the sample, which can be written as the mass thickness of the target and its stopping power (Equation 28).

$$Q = \frac{(E_i - E_o)I}{Z_a} = \frac{SD}{Z_a} \tag{28}$$

where E_i is the incident energy, in MeV; E_o is the outgoing energy, in MeV; S is the mass stopping power, in MeV g^{-1} cm^2; D is the mass thickness of the ''thin'' target, in g cm^{-2}.

Heat is not released uniformly in the sample. It is obvious that the volume is determined by the irradiation surface and the CP range in the target. The depth distribution is given by the stopping power, as shown in Figure 5 for 10-MeV protons in a ''thick'' aluminum target. One can observe the Bragg peak at the very low energy side, i.e. for a depth approaching the range. Two conclusions can be drawn. For ''thick'' targets, heating is more pronounced at the lower energy side, i.e. below the Coulomb barrier, and consequently not in the activated (and thus analyzed) part of the sample. The most efficient back-side cooling is for a sample thickness not exceeding the range. Remember that for trace element determinations in the bulk of a sample one has to always irradiate ''thick'' samples and standards. For surface characterization of thin layers (nm to μm) the heat release is several orders of magnitude lower. It is obvious that also the heat release in the substrate should be considered.

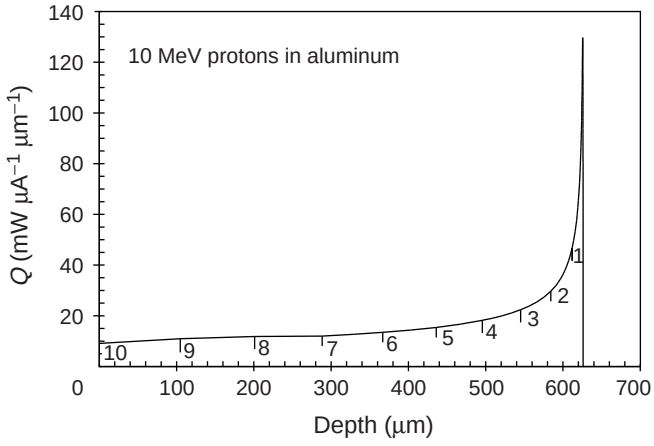

Figure 5 Heat (i.e. power) release as a function of the depth for an aluminum target irradiated with 10-MeV protons. The proton energy decrease is also indicated. The heat release in $mW\,\mu A^{-1}\,\mu m^{-1}$ is equal to the numerical value of the stopping power expressed as $keV\,\mu m^{-1}$, according to Equation (28).

6.3 Target Holder

CPAA has been applied for solid samples (massive or powdered), such as semi-conductors, metals and alloys, ceramics, geological materials, and solid environmental samples (aerosols, soils). It has not been applied for aqueous solutions because optical (atomic) and mass spectrometric methods of analysis are much more suited.

A simple target holder design is a water cooled copper or aluminum base plate, on which a ''thick target'' is mounted, that can be irradiated in the vacuum system of the cyclotron. For surface characterization, the target should be backed (e.g. by aluminum for proton activation) to obtain ''thick'' target conditions, and so avoid activation of the target holder. For powdered samples with poor thermal conductivity, irradiation in a helium atmosphere is recommended.[30] An annotated example is shown in Figure 6.

6.4 Direct Beam Intensity Monitoring

Although the beam intensity I is quantified a posteriori by I-monitors (see below), direct I-monitoring is necessary to keep it constant during irradiation, a prerequisite for obtaining Equation (11). Therefore the target holder should be electrically insulated ($>10^{10}\,\Omega$) from the beam transport system. Moreover, the cooling water should be deionized, as this is the case for the cooling system of the cyclotron. The target holder (i.e. A + B or A + C in Figure 6) is connected to the ground by an ammeter that will measure in the μA range, as shown in Figure 7(a). During irradiation the target tends to charge positively,

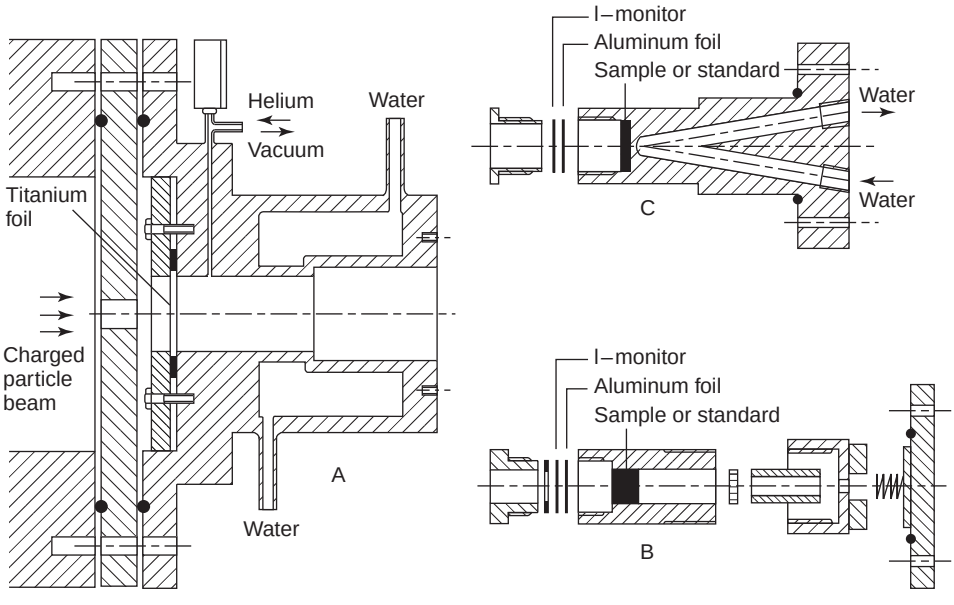

Figure 6 Target system for irradiation in a helium atmosphere.[30] A titanium foil separates the vacuum of the cyclotron from the helium atmosphere. A powdered target is loaded in a dedicated target holder (B), together with a beam intensity monitor foil and an aluminum foil. A disk target is loaded in a dedicated target holder (C), together with a beam intensity foil and an aluminum foil. Powders are loaded from the back-side (B) and, after irradiation, quantitatively unloaded from the front side. Disk targets with good thermal conductivity are also water cooled from the back-side. The target holder (B or C) is brought in a water cooled set-up (A), which is evacuated and filled with helium. (Reproduced by permission of Elsevier Science, from N. De Brucker et al., *Anal. Chim. Acta*, **220**, 93–102 (1989).)

which is balanced by the ''electron'' current measured. By impact of the CPs on the target, secondary electrons are emitted towards grounded parts (e.g. diaphragm) of the beam transport system, which causes positive systematic errors in the *I*-measurement. To avoid emission of secondary electrons (and their positive systematic errors for *I*-measurement) two experimental set-ups are used. A ring-shape electrode at negative potential (−100 V, as the electron energy does not exceed 100 eV) is inserted between the diaphragm and the target, as shown in Figure 7(b). Alternatively, the target holder is electrically connected to a tube, preventing the escape of secondary electrons over a wide spatial angle, as shown in Figure 7(c). A suitable diaphragm prevents the beam from hitting both devices as well as the target holder itself.

6.5 Quantitative Beam Intensity Monitoring

Knowledge of the ratio of the beam intensity for standard and sample (Equations 18, 22–24) is necessary for quantitative analysis in CPAA. There are three reasons for not using direct beam intensity measurement: (1) no absolute beam intensity data are required, (2) to obtain accurate data (no systematic errors due to secondary electrons and/or due to the beam hitting the target holder), and (3) to obtain precise data at low beam intensities (which is mostly the case for standards) compared to the blank. Sample and standard are covered prior to irradiation with a ''thin'' foil (i.e. *I*-monitor) of exactly the same thickness. After irradiation both *I*-monitor foils are measured. From Equation (17) the ratio of beam intensities, I_s/I_x, is given

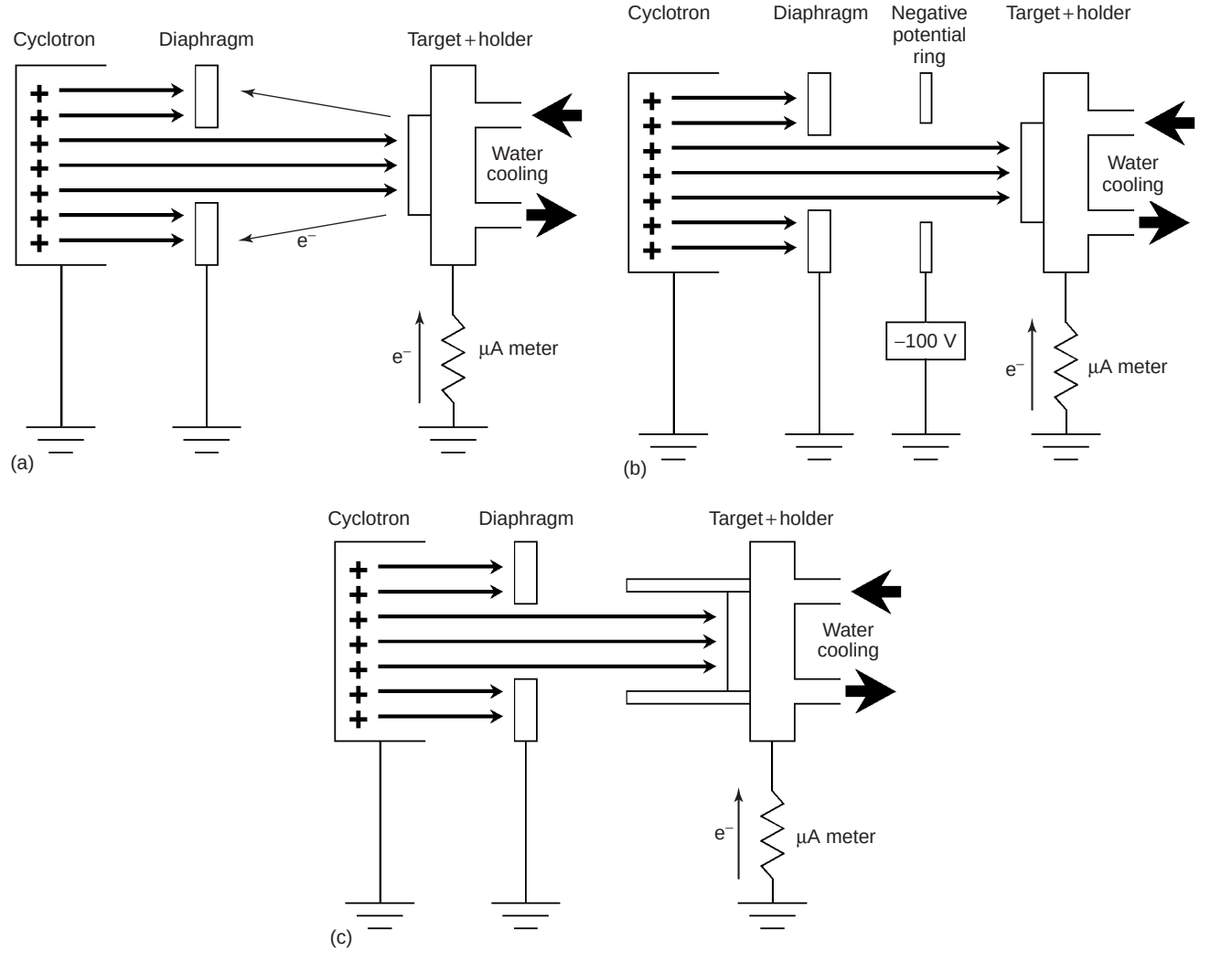

Figure 7 Direct beam intensity monitoring:[28] emission of secondary electrons causes systematic positive errors (a) that can be avoided by a ring shape electrode at negative potential (b) or a tube preventing escape of secondary electrons over a wide spatial angle (c). (Reproduced by permission of Wiley, from K. Strijckmans, 'Charged Particle Activation Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, J. Wiley & Sons, Chichester, 1994.)

by Equation (29).

$$\frac{I_s}{I_x} = \frac{A_s\left(1 - e^{-\lambda t_{is}}\right)}{A_x\left(1 - e^{-\lambda t_{is}}\right)} \quad (29)$$

where I is the beam intensity; A is the activity at the end of irradiation; $\lambda = \ln(2)/t_{1/2}$, the decay constant of the radionuclide (where $t_{1/2}$ = half-life); and t_i is the irradiation time. The subscripts x and s refer to the sample and standard, respectively.

Pure metal foils are the obvious choice as I-monitors because they are: (1) good thermal conductors, (2) monoelemental (the γ-spectra obtained are simple and thus less subject to spectral interference), and (3) available in different thickness.

During irradiation of the I-monitor, recoil nuclides are formed that may leave the I-monitor and enter sample or standard. These recoil nuclides may interfere with the measurement of the sample (or standard) activity, by nuclear interference (if the I-monitor foil is the analyte element, which is the worst possible choice) or by spectral interference. Therefore, a catcher foil is inserted between the I-monitor and the sample or standard that is not activated, e.g. aluminum for proton irradiation.

Sometimes several extra foils are inserted between the I-monitor and sample or standard to reduce the energy to just below the threshold energy or Coulomb barrier of an interfering reaction. This is also the case for a series of standards, when the samples are etched after irradiation (section 8.2).

7 ACTIVITY MEASUREMENT

Section 7.1 explains which radionuclides are formed by CP induced reactions, and gives arguments for measuring γ-spectra, where possible. If this is not possible, measuring the decay curve of annihilation radiation is described. Sections 7.2 and 7.3 give more details about instrumentation.

7.1 Radionuclides Formed and Decay Radiation Measured in Charged Particle Activation Analysis

CP induced reactions produce radionuclides that decay by positron emission and/or electron capture (EC) because they contain one proton in excess compared to stable nuclides. Figure 4(b) shows stable nuclides ^{12}C, ^{14}N, ^{16}O, for which the number of protons (Z) equals the number of neutrons ($A - Z$). In contrast, ^{14}O formed by the $^{14}N(p,n)^{14}O$ reaction is unstable, because of an unbalanced proton/neutron number (8/6). ^{14}O decays by positron emission, i.e. transmutation of a proton into a neutron (that stays in the nucleus), a positron, and a neutrino, both of which are emitted. In this way a stable nuclide ^{14}N is formed again. A positron (β^+) is a positive electron, the anti-matter of an electron. A neutrino has no charge and a negligible mass, and consequently is nearly impossible to detect. The difference in energy level between the unstable radionuclide ^{14}O and the excited state of the nuclide ^{14}N ($E_{\beta_{max}}$) is randomly distributed between the positron and the neutrino. The excited state of ^{14}N further decays to the ground state by photon emission, i.e. γ-rays. Consequently, the energy spectrum of a positron is continuous, from zero to a specific maximum energy ($E_{\beta_{max}}$), while γ-rays are mono-energetic. To identify radionuclides (i.e. qualitative analysis) specifically (i.e. without interferences), γ-spectrometry is preferred.

An alternative to positron emission is EC (or ε). As the decay radiation produced is Auger electrons and/or X-rays, which can easily be absorbed by the sample itself, (in contrast to the γ-rays) they are not used in CPAA.

A limited number of positron emitters decay directly to the ground state without γ-emission. This is the case for the determination of light elements (boron, carbon, nitrogen, oxygen), which is actually the most important application of CPAA. As can be seen from Figure 4(b), the above mentioned reaction is an exception: ^{15}O (determination of N), ^{13}N (determination of C and N), and ^{11}C (determination of B and N) are all pure positron emitters. This is also the case for ^{18}F (determination of O). Selective and sensitive measurement of positron emitters is possible by measuring their annihilation radiation. In positron annihilation a positron loses its kinetic energy, annihilates with an electron, and two annihilation photons (γ-rays) are emitted in opposite direction. According to the mass–energy equivalence law, their energy is 511 keV, corresponding to the mass of an electron at rest. It is obvious that the information about the characteristic maximum energy (and thus the identity of the radionuclide) is lost during the annihilation process. The positron emitter must now be identified by its characteristic half-life. Therefore the activity of a mixture of positron emitter(s) is measured as a function of time, i.e. the decay curve. Decay curve analysis is detailed in section 9.2.

7.2 γ-Spectrometry

γ-Spectrometry can be performed by an NaI scintillation detector and a Ge semiconductor detector. To reduce spectral interferences as much as possible, only a Ge detector is used because of its superior energy resolution characteristics compared to the NaI detector. The Ge detector is a diode, reverse biased. One γ-ray interacts with the detector by Compton scattering or the photoelectric effect. The photo- or Compton electron formed causes ionization in the detector and many electron–positive-hole pairs are formed. The electrons and

positive holes are collected by the electric field applied, and produce a charge pulse. The charge is proportional to the γ-energy (photoelectric effect) or lower (Compton scattering). After linear amplification and analog-to-digital conversion (ADC), the multi-channel analyzer shows a digital spectrum with photopeak(s) corresponding to the γ-energy/energies, and, at the lower energy side, a Compton continuum.

The characteristics of a Ge detector are described by the performance towards the 1333-keV γ-ray of ^{60}Co: (1) energy resolution, i.e. the full width at half-maximum (fwhm), 2 keV or better; (2) peak to Compton ratio, i.e. the ratio of the photopeak height to the Compton continuum at 1 MeV, typically 40–60; and (3) relative detection efficiency, i.e. relative to a $7.6 \times 7.6\,cm^2$ ($3'' \times 3''$) NaI detector at 25 cm source–detector distance, up to 50%. The absolute detection efficiency is of the order of 10^{-3}.

The energy calibration of the γ-spectrometer is carried out by calibration sources covering the whole energy range of interest (0.1–2 MeV), such as ^{241}Am, ^{57}Co, ^{24}Na and ^{60}Co. Activity calibration, i.e. determination of the absolute detection efficiency, is not necessary as absolute activity measurements are not required in CPAA (Equations 18, 22–24). Quantitative analysis requires correction for ''counting losses'' or ''dead-time'', i.e. the spectrometer does not detect (is ''dead'') a γ-ray (a ''count'') because it is still processing the previous one. As radioactive decay is random, the counting-loss probability increases exponentially with the count rate. Therefore, the dynamic range of a γ-spectrometer is limited. Different systems are in common use for the dead-time correction: the ''live-timer'', which is standard in any spectrometer but not applicable for short-lived isotopes, in contrast to the very simple ''pulser method'',[31] the dead time stabilizer (DTS),[32] and the loss-free counting (LFC) module.[33]

Samples and standards should be measured at the same detection efficiency, which is determined by the source-to-detector geometry, and self-absorption as well. For instrumental analysis of solid samples, it is mostly sufficient to measure samples and standards positioned at exactly the same distance from the detector, with the irradiated side towards the detector. Because of the limited range of CPs in matter, the self-absorption of sample and standard (and hence the difference in self-absorption) is very low or negligible, except for very low γ-energies and high-Z targets. This can be roughly estimated, assuming a point source absorbed by half the range, using mass attenuation coefficients compiled by Hubbell.[34,35] For radiochemical analysis (section 8.4) the geometry and self-absorption of the sample (solution or precipitate) can be quite different compared to the standard (a ''thick'' or ''thin'' foil). The relative (i.e. sample vs standard) detection efficiency has to be determined experimentally. A tracer, preferably the mono-elemental standard or, if not, a radionuclide emitting almost the same γ-energy, is irradiated, measured in the standard geometry, brought into the same chemical and physical form as the sample, and measured again. The chemical form should be as similar as possible with respect to its atomic number, Z.

7.3 Positron Counting

Pure positron emitters have to be measured by their 511-keV annihilation photons, emitted in opposite direction (section 7.1). Although this can be measured by a Ge detector, it is more convenient to use an NaI detector because: (1) the detection efficiency is at least equal, even for small NaI detectors, compared to the state-of-the-art Ge detectors; (2) an NaI detector is much cheaper than a Ge detector; (3) a Ge detector should be cooled with liquid nitrogen but an NaI detector does not need cooling; (4) there is no advantage in obtaining the better energy resolution of the Ge detector, as all positron emitters produce annihilation radiation of the same energy. Therefore, there is no need for a multi-channel analyzer. It is sufficient to use a single-channel analyzer that selects all γ-rays measured in the 0.4–0.6 MeV energy interval.

The signal-to-noise ratio can also be improved by the use of a $\gamma-\gamma$ coincidence set-up. Two NaI detectors are used, facing each other, and the β^+ emitter is placed between them. The annihilation photons are emitted in opposite direction, and can be detected simultaneously in both detectors. A coincidence circuit selects all simultaneous events (typically within 40 ns). The system is very insensitive for all γ-radiation other than annihilation photons, even if present in the sample, for positron sources outside the gap between both detectors and for background radiation or random noise. However, the detection efficiency is lower compared to a single detector, because both annihilation photons should produce a photoelectric effect in each detector. The detection efficiency sharply decreases when the source is moved away from the space between the two detectors. Identical geometry for both sample and standard is still more important than for Ge spectrometry. The relative (sample vs standard) detection efficiency can be determined experimentally, as explained in section 7.1, using a pure positron source like ^{68}Ge/^{68}Ga ($t_{1/2} = 270$ days), which is commercially available.

For tuning the set-up (energy range, coincidence time interval) and checking its stability (constant detection efficiency), long-lived positron sources like ^{22}Na ($t_{1/2} =$ 2.6 years, also a γ-emitter) or ^{68}Ge/^{68}Ga ($t_{1/2} = 270$ days) can be used.

As annihilation only occurs after the positron slows down, the positron source should be enveloped by an absorber (of any value of Z). An absorber thickness

equal to the positron range for the characteristic maximum energy of the positron emitter (0.2–0.9 g cm^{-2} for 0.5–2 MeV positrons) is largely sufficient for two reasons: (1) the energy spectrum of the positron emitter is continuous from zero to the characteristic maximum energy; and (2) the projected range (i.e. trajectory projected on the original direction) is much smaller than the linear range (i.e. total path length) for positrons. Indeed, positrons are more deflected when interacting with electrons than with protons. Of course, the annihilation radiation can be absorbed also, therefore a low-Z absorber (plastic, aluminum) is used. The γ absorption of 1-mm aluminum for annihilation photons is 2%.

As for γ-spectrometry, the dynamic range of a γ–γ coincidence set-up is limited by ''counting losses'' or ''dead-time''. A correction is possible, as given in Equation (30), if the dead-time is known, i.e. the time the set-up needs to process one coincident event.

$$R_c = \frac{R_m}{1 - R_m\tau} \tag{30}$$

where R_c is the count rate corrected for dead-time, R_m is the measured count rate, and τ is the dead-time. This equation is only valid for a non-extensible dead-time, i.e. one that is constant as a function of the count rate. Although this is not the case for a γ–γ coincidence set-up, this correction can be applied for counting losses up to 10%, e.g. for $\tau = 10^{-6}$ min per count and $R < 10^5$ counts per min. The dead-time can be determined experimentally by repeated measurements of a short-lived radionuclide that is radioisotopically pure and has a well-known half-life. Irradiation of pure graphite or polyethene with deuterons (<12 MeV) produces the pure positron emitter ^{13}N (Figure 4b) with a half-life of 9.965 ± 0.004 min. Measurements at low count rates, with negligible counting losses, are extrapolated to high count rates (R_c) in Equation (30). The dead-time, τ, can then be calculated from the measured count rate R_m. Alternatively, the best fitted dead-time can be found with a decay curve analysis program (section 9.2) by trial and error.

8 PROCEDURE

The conceptual and experimental procedures to be followed for trace element determinations in the bulk of a sample and for surface characterization are described as a whole, making reference to the other sections if appropriate. The common part is given in the next section. Finally, a radiochemical separation can improve the detection limit and precision.

8.1 General

The first step is to make an appropriate choice for nuclear reaction and CP energy to avoid possible nuclear, spectral, or matrix interference (section 4). For trace element analysis in the bulk of a sample the matrix is the main component of the sample; for surface characterization the main component is the substrate of the surface layer.

The tools are data for threshold energy, Coulomb barrier, cross-section or TTY (if these are not available then Equation 9 may be used), stopping power or range, and a chart of the nuclides (sections 2 and 3). Examples are given in section 4. Activation curves can be obtained experimentally as described in section 3.4.

To preserve the advantage of CPAA as an ''absolute'' or ''independent'' method, no standards may be used that are certified or calibrated by another method. Mono-elemental standards are ideal. Therefore, pure elemental standards (i.e. metals) or a pure compound with exactly known stoichiometry can be used to obtain accurate data for the mass fraction for the analyte element as well as for stopping power. The standard should be stable during irradiation (when heating occurs, as described in section 6.2) in a vacuum. Some organic and hygroscope compounds cannot be used, not only because the mass fraction of the analyte element increases during irradiation, but more importantly because the stopping power decreases as the hydrogen content decreases. Indeed, the stopping power for hydrogen is 3–5 times higher than for the other elements (Figure 1c). Both phenomena lead to positive errors. Not only should the standard itself be stable during irradiation, but so also should the radionuclide formed or any compound of it. Boric acid is a good boron standard for the nuclear reaction ^{10}B(p,α)^{7}Be, but not for ^{10}B(p,n)^{11}C, as the ^{11}C formed volatilizes as carbon dioxide. Boron metal must be used for the latter reaction. To avoid nuclear or spectral interferences the mono-elemental standard should be very pure. Ideally, only activity from the analyte element is found in the γ-spectrum or decay curve.

For irradiation, samples and standards are covered with an I-monitor foil, a catcher foil for recoil nuclides from the I-monitor foil, and, if necessary, extra foils for fine tuning the incident energy on sample or standard (section 6.5). Never use the analyte element as an I-monitor. The thickness of the catcher foil is at least equal to the range of the recoil nuclides. The energy of the recoil nuclides can be calculated exactly, but is always smaller than $E_a + Q$. So the maximum range can be calculated as described in section 2.3. In practice a few μm is sufficient.

The use of an I-monitor, a catcher foil and more extra foils slightly reduces the incident energy on the

sample and the standard. Using the definition of stopping power (Equation 1) this can be calculated by using Equation (31).

$$E_o = E_i - SD \tag{31}$$

where E_o is the outgoing energy, in MeV; E_i is the incident energy, in MeV; S is the mass stopping power, in MeV g^{-1} cm^2; and D is the mass thickness of the "thin" foil, in g cm^{-2}. For each foil with mass thickness D, the outgoing energy E_o of the previous foil is the incident energy E_i of the following foil. For "thin" foils the stopping power is considered constant in the energy interval $E_o - E_i$. The mass thickness of the foils has to be determined experimentally by weighing and measuring its surface, rather than using its nominal or measured thickness multiplied by a tabulated mass density. For an aluminum foil 10×10 cm^2 with a 5-μm nominal thickness, measured with an uncertainty of 0.1 cm and 0.1 mg, the mass thickness yields 1.35 ± 0.03 mg cm^{-2}. The precision on the energy, calculated by Equation (1), yields about 3 keV for α-particles ($S \approx 100$ MeV g^{-1} cm^2), which is largely sufficient. The precision is much better for protons, deuterons, and helium-3 particles, for foils with an atomic number higher than that of aluminum (i.e. foils with a lower stopping power) or for foils with a mass density higher than aluminum or thicker than 5 μm (i.e. foils with a higher mass).

Irradiations are carried out under vacuum or helium, monitoring the beam intensity as described in sections 6.3 and 6.4. For the standards, shorter irradiation times and/or beam intensities are applied, as the amount of analyte element irradiated is orders of magnitude higher than for the sample. To avoid radioactive contamination from the standards, it is good practice to irradiate the samples prior to the standards.

After irradiation, a waiting time, also called the decay or cooling time, is observed for sample handling, chemical etch (section 8.2) or radiochemical separation (section 8.4).

Measurements of samples and standards are done by γ-spectrometry (section 7.2) or positron counting (section 7.3), I-monitors by γ-spectrometry. Time management is important to improve signal-to-noise ratio (and thus the detection limit and repeatability) and to avoid or suppress spectral or matrix interferences. For an irradiation time equal to one half-life one obtains half of the activity for an infinite irradiation time (Equation 9). For a waiting time equal to one half-life one obtains half of the activity for no waiting time (Equation 9). Shorter irradiation or waiting times decrease the analyte activity compared with short-lived interfering activities. Longer irradiation or waiting times decrease the analyte activity compared with long-lived interfering activities. The same holds for the measuring time, simply replacing "activity" by "number of counts" (see Equation 34 in section 9.3).

Data handling is described in section 9.

8.2 Trace Elements in the Bulk of a Sample

To obtain high sensitivity, samples and standards are irradiated at an energy slightly above the maximum cross-section, i.e. when the TTY reaches a plateau. For the determination of iron by the $^{56}Fe(p,n)^{56}Co$ reaction (Figure 3b) the proton energy is about 15–20 MeV. The initial choice is further refined or rejected by considerations of nuclear, spectral and matrix interferences, as explained in section 8.1.

Samples and standards are always "thick" targets (thicker than the range, section 5.2). However, for backside cooling the thickness should not exceed the range too much (sections 6.2 and 6.3). For massive samples (not powders) interference from the surface can be avoided by chemical etch prior to irradiation, at least if further surface contamination can be avoided until the end of irradiation. It is obvious that chemical contamination after irradiation does not interfere any more. For the determination of carbon, nitrogen or oxygen the sample may be contaminated after chemical etch and prior to irradiation. For example, the alumina layer found in air may interfere with trace oxygen determination in the bulk of pure aluminum. The alumina layer is also activated and the oxygen content is found to be too high. Chemical etch *after* irradiation is then the method of choice. The activated surface layer will be removed and replaced by an inactive surface layer that will no longer interfere. The thickness to be removed is not determined by the thickness of the oxide layer, but by the range of the recoil nuclides formed from oxygen in the alumina layer. The energy of the recoil nuclides is less than $E_a + Q$, and its range can be calculated as described in section 2.3. It is good practice to check experimentally the apparent oxygen content compared with the amount removed from the surface layer: a steep decrease followed by a plateau at a few μm. Chemical etch after irradiation reduces the incident energy on the sample, i.e. the sample as it will be measured. Etching the standard is not convenient and may be impossible for powdered standards. A set of two standards is irradiated at two different energies by insertion of additional foil(s) between the recoil foil and the standard. That energy interval should cover the incident energies for all the samples after etch, which is not always very reproducible. Then the standard activity for each individual sample can be obtained by linear interpolation of the set of standards. Energy calculations are done using Equation (31). The mass thickness of the removed surface layer can be obtained by measuring the dimensions of the sample

and weighing prior to irradiation, rather than etch, and after etch. Linear interpolation is justified if the incident energy is chosen in the plateau of the TTY.

8.3 Surface Characterization of Thin Layers

To obtain high sensitivity, samples and standards are irradiated around the energy for maximum cross-section. If Equation (23) is applied, the cross-section should be nearly constant for the energy interval of the ''very thin'' standard (sections 5.1 and 5.3). The initial choice is further refined or rejected by considerations of nuclear, spectral and matrix interferences, as explained in section 8.1.

For surface layer characterization samples are always ''thin''; the standards can be ''thick'', ''thin'', or ''very thin'', each requiring appropriate standardization, as explained in section 5.3. A ''thin'' standard does not mean that the standard should be comparable to the sample thickness (the order of magnitude is μm–nm), or that its thickness or composition should be determined by another method. ''Thin'' standards are typically 10–100 μm thick. Their mass thickness has to be determined experimentally by weighing and measuring the surface, as explained in section 8.2. For a standard that is not available as a foil, a virtual ''thin'' standard can be used. A set of two ''thick'' standards is irradiated at two different energies, one without and the other with an additional foil between the recoil foil and the standard. The difference in standard activities is equal to a ''thin'' standard with a thickness that yields the same energy reduction as the additional foil. The mass thickness of the virtual ''thin'' standard can be calculated from Equation (32).

$$D_s = D_f \frac{S_f}{S_s} \tag{32}$$

where the subscripts s and f refer to the virtual ''thin'' standard and the foil covering the second ''thick'' standard, respectively. As the virtual ''thin'' standard activity results from a small difference in activities, both standards should be measured with good counting statistics. Moreover the virtual ''thin'' standard should be as thick as possible and still comply with the ''nearly constant cross-section'' condition.

Heating of the sample during irradiation is described in section 6.2. To avoid activation of the target holder, samples and ''thin'' standards should be backed by a ''thick'' thermal conductor, e.g. aluminum for proton irradiation.

8.4 Radiochemical Separation

Instrumental analysis is possible if spectral or matrix interference can be avoided by an appropriate choice of the incident energy and/or the measuring conditions (section 4) and/or time management for irradiation and measurement (section 8.1). If not, the radionuclide B formed from the analyte element A has to be separated radiochemically from an interfering radionuclide D formed from an interfering element C. The latter technique is called radiochemical analysis. This section deals with some major differences between a radiochemical separation and a common chemical separation as used in nonnuclear methods of analysis.

For most CP induced reactions, the atomic number of the radionuclide B is different from that of the analyte element A. This is the case for (p,n), (p,α), (d,n), (d,α), (^{3}He,n), (^{3}He,d), (α,n), and (α,d) reactions, but not for (p,d) and (^{3}He,α) reactions (section 3.4 and Figure 4). The chemical separation to be developed for CPAA is thus different from that for all nonnuclear analytical methods and for some other methods based on activation analysis, such as thermal and fast neutron activation analysis using the (n,γ) or (n,2n) reaction, respectively, and photon activation analysis using the (γ,n) reaction. In principle, it is also not necessary to separate the matrix, but rather the radionuclide(s) formed from the matrix element(s). Again, the atomic number of a radionuclide formed from a matrix element is generally different from that of the matrix element itself. Considering only the chemical separation involved, CPAA is an independent analytical method, not subject to the same systematic errors as other analytical methods. Radiochemical separation has three important advantages over conventional chemical separation:

1. An inactive carrier can be added for the elements to be separated (B and D). This avoids the difficulties of a chemical separation at the trace level.
2. Reagent impurities (or blanks) do not influence the detection limit capabilities of the analytical method.
3. Separations need not be quantitative or even reproducible (see section 8.4.2).

For the choice and development of a radiochemical separation the following points have to be considered:

1. *Selectivity of the separation*. As pointed out in section 4, the induced matrix activity (D) can be orders of magnitude higher than the activity to be measured (B). It is not possible to resolve such a low activity in the presence of such a high activity, as the dynamic range of the measuring equipment is limited (sections 7.2 and 7.3). Therefore the decontamination factor (the ratio of D before separation to that after separation) should be up to 10^4.
2. *Quantitative nature of the separation*. Quantitative recovery of B is to be preferred (99% or more).

If the separation is not quantitative and not even reproducible, then determination of the yield for each individual separation is possible. Two approaches can be followed: the addition, before separation, of an accurately known amount of an inactive/active carrier and measurement of the mass/activity of the carrier after separation. It is obvious that the radionuclide added and the analyte radionuclide should be different isotopes of the same element.

3. *Speed of the separation*. To obtain an optimum detection limit, the time needed to perform the whole separation procedure should not exceed a few half-lives of the radionuclide to be measured.
4. *Inactive carrier*. It is not only an advantage, but mostly also a necessity, to add inactive carriers of the elements to be separated. Indeed, if the atomic number of the radionuclide B is different from that of the analyte element A (which is usually the case; see above), this radionuclide B is produced ''carrier free''. As the number of radionuclides is given by

$$N = \frac{At_{1/2}}{\ln(2)} \tag{33}$$

where N is the number of radionuclides, A is the activity (in Bq) and $t_{1/2}$ is the half-life (in s), the amount of substance is 10^{-16} mol for $A = 1$ kBq and $t_{1/2} = 1$ day. It is clear that if there is no accidental addition of inactive element B (e.g. as impurity in the matrix or reagents) it is not possible to separate chemically such a low mass.

5. *Detection efficiency of measurement*. To obtain a high sensitivity, the activity should be measured after separation with the highest possible detection efficiency. This is also determined by the geometry (size and distance from the detector) of the source. Separation procedures ending with volumes of solution are less favorable than those ending with a precipitate.
6. *Repeatability of the measurement*. The repeatability of the measurement is determined by the counting statistics and also by the possibility of reproducing measurements with exactly the same detection efficiency. The former is expressed as the relative standard deviation calculated from counting statistics (i.e. $C^{-1/2}$, where C is the number of counts measured). For a high detection efficiency, one has to use a small source to detector distance, which is less reproducible geometrically. Obtaining a good repeatability and a high detection efficiency are thus contradictory demands.
7. *Self-absorption of the source*. Self-absorption of the source is in principle no problem, as long as the absorption is identical for all samples and standards, which is mostly the case for instrumental analysis (section 7.2). Self-absorption can be important for low-energy γ-rays or for sources containing components of high atomic number.
8. *Separation of the matrix*. Although in principle not necessary, it may be convenient to separate the (inactive) main matrix components and so reduce the mass of material involved in the separation procedure. The following separation process is then faster, leads to smaller volumes of solution or precipitate and allows further evaporation of a solution to a smaller volume if necessary. As a consequence, the detection efficiency is improved and the self-absorption is lowered substantially or becomes negligible.
9. *Health physics*. The activity level required to perform precise measurements (i.e. at the kBq level) provides no health physics risk, as doses are of the order of nSv h^{-1}. The induced matrix activity, however, may be much higher. It is good practice to first separate the main matrix activity, choosing a procedure that limits manual intervention to the strict minimum. Ion exchange chromatography, for instance, is a better choice than manual solvent extraction.

The experimental development of a radiochemical separation is supported by the use of radiotracers. The chemical separation is simulated step by step with the addition of radiotracers (and inactive tracers as well) for the element to be separated (B) and for the interfering element (D), in order to check quantitativeness and selectivity. These radiotracers are preferably produced by thermal neutron activation, if such a facility is available, rather than by CP activation. Indeed, production by, for example, a (p,n) reaction yields a radiotracer that has to be separated from the target material, whereas for production by an (n,γ) reaction the target material can be used as the inactive carrier. The tracer has to be brought to the same chemical form as the radionuclide to be separated. This is often not possible for the dissolution step preceding the actual separation procedure. So, the choice of the dissolution procedure should guarantee quantitative recovery of the radionuclide to be measured.

For each separation step, three measurements can be made: the initial activity, as a reference; and, after separation, the two fractions in which the major and minor activity are collected. From the measurement of the minor fraction the most precise yield can be calculated. However, it is not always possible to perform all measurements with the same detection efficiency. For a precipitation, for example, the reference measurement is made on the solution before precipitation. After precipitation and filtration, a fraction of the filtrate (equal

in volume to the reference) is taken and the yield for the ion to be precipitated can be calculated very precisely. To check coprecipitation (e.g. of matrix activity) the activity of the precipitate is measured. As the detection efficiency is not comparable with that of the reference measurement, the result has to be considered as an estimate. This is largely sufficient for the purpose of evaluating the selectivity of this separation.

There is no need to apply radiochemical separation to mono-elemental standards. However, standards and samples have to be measured at identical detection efficiency. Therefore the standards are brought into exactly the same geometrical form as the samples after separation. If the last step of a separation is ion chromatography, the standards are dissolved and the total volume is brought exactly to the volume of the eluate. Possible differences in self-absorption have to be considered. The eluate may contain an acid other than the one used to dissolve the standard; the eluate may contain only the added inactive carrier (B), whereas the standard contains a larger amount of the analyte element (A).

9 DATA HANDLING

This section follows chronologically from the previous one. Once the irradiation and measurement data (γ-spectrum or decay curve) have been collected, they have to be analyzed and the result (i.e. trace element concentration in the bulk of a sample, mass thickness of a surface layer, composition of a surface layer) has to be calculated.

9.1 γ-Spectrum Analysis

The manufacturers of γ-spectrometers provide suitable programs to resolve the spectrum, i.e. find the significant peaks and calculate their energy, resolution (i.e. fwhm, and comparison to the detector characteristics), net count area, standard deviation (from counting statistics), and detection limit (in number of counts). Once such data have been obtained, it is not sufficient to search only for the most intense γ-ray of the analyte radionuclide. The whole spectrum has be analyzed, starting at the high energy side of the spectrum, using databases with γ-energies and γ-intensities (number of γ-rays emitted per 100 disintegrations). It should be noted that the detection efficiency of a γ-detector is a function of the γ-energy. For high energy γ-rays ($E_\gamma > 1022\,\text{keV}$) single and double escape peaks can be observed at an energy equal to $E_\gamma - 511\,\text{keV}$ and $E_\gamma - 1022\,\text{keV}$. Two intense γ-rays, or γ-rays emitted in cascade, can give rise to sum-peaks. All this can provide useful information concerning the main, minor or trace components in the matrix, which can cause nuclear, spectral and/or matrix interference. Peak broadening is evidence of spectral interference, although no peak broadening does not imply the absence of spectral interference. In case of doubt, decay curve analysis (section 9.2) can be applied to check for spectral interference. All possible reactions on all stable isotopes for an element observed in the spectrum have to be checked, as explained in section 3.4.

Erdtmann and Soyka published a table with γ-energies and γ-intensities, ordered by radionuclide as well as energy.[36] This database is also available at the Lund Nuclear Data WWW Service (Lund University, Sweden).[37] The production mode can be specified as ''CP reaction''. The database is not updated. The NuDat (Nuclear Data) database is retrievable by, and can be sorted by, radionuclide, half-life, γ-energy and/or γ-intensity. The NuDat database at the Lund Nuclear Data WWW Service (Lund University, Sweden) was last updated in January 1996, and will not be updated in the future. In 1999, it is actually the only NuDat base where the production mode can be specified as ''CP reaction''. The NNDC databases are updated three to four times a year, and mirrored at IAEA and NEA.[38] The same database is also obtainable as a PC/DOS program: PCNUDAT.[39] The Table of Isotopes (TOI) database is available as a book and a CD-ROM,[40] and on the Internet at the Lund Nuclear Data WWW Service (Lund University, Sweden).[41] Both are updated yearly.

9.2 Decay Curve Analysis

Positron counting, followed by decay curve analysis, is only applied when the analyte radionuclide is a pure positron emitter. Unfortunately, all CP induced reactions induce positron emitters, and consequently the decay curve may be very complex. In contrast to γ-spectra, which may contain hundreds of γ-rays, the number of components in a decay curve is rather limited (<10). Ideally, a decay curve analysis should answer the questions: (1) how many components are present; (2) what is the half-life of each component; and (3) what is its activity? As most positron emitters are also γ-emitters, their possible presence in the decay curve can be checked by γ-spectrometry and avoided by a proper choice of irradiation and measurement conditions. Moreover, the number of pure positron emitters is rather limited and the number and identity of possible components in the decay curve are mostly known. A decay curve analysis that provides an answer to the second question (qualitative analysis) and the third question (quantitative analysis) is acceptable. By weighed linear regression analysis the third question can be answered, the goodness of fit indicating if the initial assumptions about the number and half-life of

the components are correct. The half-life cannot be fitted because the decay curve is a sum of exponential functions, thus the half-life parameter is not linear. Assuming an initial value for the half-life, the problem can be linearized, and so, by iteration, the best half-life can be fitted. Components that are not present can be rejected because their activity is extremely low or negative. Such an algorithm has been developed by Cumming,[42] and applied successfully. More recently, Prony's improved method has been implemented, tested extensively, and compared to Cumming's method by Schreurs et al.[43] This method also answers the first question, but requires data points equidistant in time.

9.3 Quantitation

Standardization is done in a relative way, the sample and standard being irradiated and measured under, principally, identical conditions. As these identical conditions are not fulfilled in practice, some corrections or normalizations are performed first.

The number of counts measured are corrected for:

1. Counting losses (dead-time), according to sections 7.2 and 7.3, and background (for γ-spectrometry this is mainly the Compton continuum of other γ-rays), using a decay curve analysis program.
2. Differences in measurement (t_w, t_m) conditions (Equation 34), using a decay curve analysis program. The normalized count rate represents the count rate for a decay time zero, i.e. at the end of irradiation. In Equations (22–24) it can replace A for equal detection efficiency of sample and standard.

$$R = \frac{\lambda C\, e^{\lambda t_w}}{1 - e^{-\lambda t_m}} \tag{34}$$

   where
   R = normalized count rate, the number of counts per unit time;
   $\lambda = \ln(2)/t_{1/2}$ decay constant (where $t_{1/2}$ = half-life);
   C = number of counts measured;
   t_w = waiting time;
   t_m = measuring time.
3. Differences in irradiation conditions (I, t_i), corrected for by Equation (18), (22), (23) or (24) (sections 5.2 and 5.3).
4. If applicable, differences in detection efficiency, due to different sample/standard-to-detector geometry and/or self-absorption for sample versus standard (section 7.2).

Finally, difference between the mono-elemental standard and the sample is corrected by Equation (19), (20) or (21) for trace element analysis and Equation (22), (23) or (24) for surface characterization, as explained in sections 5.2 and 5.3. Each individual result can be completed by the standard deviation, calculated from counting statistics, which can be compared to the reproducibility.

For surface characterization, primarily a partial mass thickness is obtained. The total mass thickness can be calculated if the mass fraction (or stoichiometry) is known. The thickness can also be calculated if the mass density is known. Be aware that the mass density can be much lower than the tabulated values, because less dense structures can be deposited, depending on the production mode. The partial mass thickness can also be used to calculate the mass fraction if the total mass thickness is known or if all mass fractions are determined. Also the stoichiometry can be calculated from partial mass thickness data.

10 PERSPECTIVES

10.1 Charged Particle Activation Analysis

During the last two decades CPAA has proved its outstanding capabilities in trace element determinations in solid samples. For the determination of light elements CPAA is still unique. Blondiaux et al. reviewed its applications.[44] Recently CPAA has been proposed for surface characterization[45] and experimentally validated.[46] This method, despite its inherent limitations, is expected to be a reference method, because of its outstanding accuracy, precision and detection limit.

10.2 Charged Particle Activation for the Development of Other Analytical Methods

In addition to CPAA, the production of carrier-free (section 8.4) radionuclides by charged particle activation (CPA) can be a magnificent tool in the development of other analytical methods. A radioactive tracer is a molecule labeled with a radionuclide. A carrier-free radionuclide or tracer is one with a high specific activity, i.e. activity per mass unit. At very low concentrations some analytical methods require preconcentration or chemical separation prior to measurement. These methods can be developed using these tracers, because the extremely low amount of analyte is almost unchanged. The tracer experiments can be performed at the same concentration levels as the final analyses. The activity levels required are very low, and do not present a significant risk. Using rather short-lived radionuclides, there is no waste problem either. An example[47,48] is the determination of platinum by inductively coupled plasma mass spectrometry (ICPMS), using ^{191}Pt.

ACKNOWLEDGMENTS

Grateful acknowledgements are made to the FWO, Fund for Scientific Research, Flanders, Belgium for financial support.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Conversion
CP	Charged Particle
CPA	Charged Particle Activation
CPAA	Charged Particle Activation Analysis
DTS	Dead Time Stabilizer
EC	Electron Capture
fwhm	Full Width at Half-maximum
IAEA	International Atomic Energy Agency
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LFC	Loss-free Counting
NEA	Nuclear Energy Agency
NNDC	National Nuclear Data Center
PET	Positron Emission Tomography
TOI	Table of Isotopes
TTY	Thick Target Yield

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Photon Activation Analysis • Radiochemical Neutron Activation Analysis • Radiotracer Methods

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • γ-Spectrometry, High-resolution, for Radionuclide Determination • Nuclear Detection Methods and Instrumentation

REFERENCES

1. http://www.research.ibm.com/ionbeams/SRIM
2. J.F. Ziegler, J.P. Biersack, U. Littmark, *The Stopping and Range of Ions in Solids*, Pergamon Press, New York, Vol. 1, 1985.
3. G. Audi, A.H. Wapstra, 'The 1993 Atomic Mass Evaluation (I) Atomic Mass Table', *Nucl. Phys.*, **A565**, 1–65 (1993).
4. G. Audi, A.H. Wapstra, 'The 1995 Update to the Atomic Mass Evaluation', *Nucl. Phys.*, **A595**, 409–480 (1995).
5. • http://nucleardata.nuclear.lu.se/database/masses *Select:* reaction calculator
   • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* UTIlities – *select:* QCAlc
   • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* UTIlities – *select:* QCAlc
6. • http://www.nea.fr/abs/html/uscd1200.html *Apply for QCALC, i.e. a VAX-VMS program, only for registered users at* http://www.nea.fr/html/signon.html
7. • http://www.nndc.bnl.gov/nndc/exfor/
   • http://www-nds.iaea.or.at/exfor/
   • http://www.nea.fr/html/dbdata/x4/welcome.html *Only for registered users at* http://www.nea.fr/html/signon.html
   • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* DATa bases – *select:* CSISRS
   • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* DATa bases – *select:* CSISRS
   • telnet://db.nea.fr *User:* NEADB – *password:* GUEST – *select:* Data services – *select:* EXFOR
8. K.A. Keller, J. Lange, H. Münzel, *Q-values and Excitation Functions of Nuclear Reactions, Part c: Estimation of Unknown Excitation Functions and Thick Target Yields for p, d, ^{3}He and α Reactions*, ed. H. Schopper, Springer-Verlag, Berlin, 1974.
9. W. Seelmann-Eggebert, G. Pfennig, H. Münzel, H. Klewe-Nebenius, *Karlsruher Nuklidkarte – Chart of the Nuclides*, 5th edition, Kernforschungszentrum Karlsruhe, 1981.
10. J.R. Parrington, H.D. Knox, S.L. Breneman, E.M. Baum, F. Feiner, *Chart of the Nuclides*, 15th edition, General Electric Co. & Knolls Atomic Power Lab., 1996.
11. • ftp://iaeand.iaea.or.at *User:* NDSOPEN – *password:* GUEST – *change directory:* UD4:[SCR.NUCHRT] – *download:* DISKx.ZIP ($x = 1, 2, 3$)
    • http://www.nea.fr/abs/html/iaea1320.html *Apply for* NUCHART; *only for registered users at* http://www.nea.fr/html/signon.html
12. • http://www-csnsm.in2p3.fr/amdc/nubcons_en.html
    • ftp://csnftp.in2p3.fr/pub/AMDC/nucleus/README.NUC
13. http://www.best.com/~sdv/MacNuclide2
14. http://ie.lbl.gov/isoexpl/isoexpl.htm
15. http://nucleardata.nuclear.lu.se/nucleardata/nd.asp?page=Charts
16. http://csnwww.in2p3.fr/amdc/jvnubase
17. • http://hpngp01.kaeri.re.kr/CoN/index.html
    • http://www.dne.bnl.gov/CoN/index.html
    • http://sutekh.nd.rl.ac.uk/CoN/
    • http://www.nes.ruhr-uni-bochum.de/CoN/index.html

18. C. Vandecasteele, K. Strijckmans, 'Standardization in Charged Particle Activation Analysis', *J. Radioanal. Chem.*, **57**, 121–136 (1980).
19. E. Ricci, R.L. Hahn, 'Theory and Experiments in Rapid, Sensitive Helium-3 Activation Analysis', *Anal. Chem.*, **37**, 742–751 (1965).
20. E. Ricci, R.L. Hahn, 'Sensitivities for Activation Analysis of 15 Light Elements with 18-MeV Helium-3 Particles', *Anal. Chem.*, **39**, 794–797 (1967).
21. M.A. Chaudri, G. Burns, E. Reen, J.L. Rouse, B.M. Spicer, 'A Method for Charged Particle Activation Analysis and its Application to Fluorine Determination by The $^{19}F(p,\alpha\gamma)^{16}O$ Reaction', *J. Radioanal. Chem.*, **37**, 243–253 (1977).
22. K. Ishii, M. Valladon, J.L. Debrun, 'The Average Stopping Power Method for Accurate Charged Particle Activation Analysis', *Nucl. Instrum. Methods*, **150**, 213–219 (1978).
23. K. Ishii, M. Valladon, C.S. Sastri, J.L. Debrun, 'Accurate Charged Particle Activation Analysis: Calculation of the Average Energy in the Average Stopping Power Method', *Nucl. Instrum. Methods*, **153**, 503–505 (1978).
24. K. Casteleyn, K. Strijckmans, 'Internal Standardization in Charged Particle Activation Analysis', *Nucl. Instr. Meth., B*, **68**, 161–165 (1992).
25. K. Ishii, C.S. Sastri, M. Valladon, B. Borderie, J.L. Debrun, 'Two Reactions Method for Accurate Analysis by Irradiation with Charged Particles', *Nucl. Instr. Meth.*, **153**, 507–509 (1978).
26. K. Masumoto, M. Yagi; 'Simultaneous Determination of Phosphorus, Chlorine, Potassium and Calcium in Several Control Serums by Alpha-particle Activation Analysis using the Internal Standard Method', *J. Radioanal. Nucl. Chem. Art.*, **109**, 449–458 (1987).
27. M. Yagi, K. Masumoto, 'Instrumental Charged-particle Activation Analysis of Several Selected Elements in Biological Materials using the Internal Standard Method', *J. Radioanal. Nucl. Chem. Art.*, **111**, 359–369 (1987).
28. K. Strijckmans, 'Charged Particle Activation Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, J. Wiley & Sons, Chichester, 215–252, 1994.
29. K. Strijckmans, 'Charged Particle Accelerators', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, J. Wiley & Sons, Chichester, 64–74, 1994.
30. N. De Brucker, J. Dewaele, K. Strijckmans, C. Vandecasteele, 'Determination of Thallium in Zinc by Proton Activation Analysis', *Anal. Chim. Acta*, **220**, 93–102 (1989).
31. K. Strijckmans, C. Vandecasteele, J. Hoste, 'Non-destructive Proton Activation Analysis for Nitrogen in Refractory Metals', *Anal. Chim. Acta*, **89**, 255–266 (1977).
32. J. Bartošek, G. Windels, J. Hoste, 'A New Circuit for Dead-time Compensating Timer', *Nucl. Instrum. Methods*, **103**, 43–44 (1972).
33. G.P. Westphal, 'Real-time Correction of Counting Losses in Nuclear Pulse Spectroscopy', *J. Radioanal. Chem.*, **70**, 387–410 (1982).
34. J.H. Hubbell, 'Photon Mass Attenuation and Energy-absorption Coefficients from 1 keV to 20 MeV', *Int. J. Appl. Radiat. Isot.*, **33**, 1269–1290 (1982).
35. • http://physics.nist.gov/PhysRefData/Xcom/Text/XCOM.html
    • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* DATa bases – *select:* XRAY
    • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* DATa bases – *select:* XRAY
    • http://www.nea.fr/abs/html/dlc-0174.html *Apply for* ZZ-XCOM, *i.e. a PC program, only for registered users at* http://www.nea.fr/html/signon.html
36. G. Erdtmann, W. Soyka, *The Gamma Rays of the Radionuclides*, Verlag Chemie, Weinheim, 1979.
37. http://nucleardata.nuclear.lu.se/database/es
38. • http://www.nndc.bnl.gov/nndc/nudat
    • telnet://telnet.nndc.bnl.gov *Login:* NNDC – *authorization code:* GUEST – *select:* DATa bases – *select:* NUDAT
    • telnet://iaeand.iaea.or.at *Login:* IAEANDS – *authorization code:* GUEST – *select:* DATa bases – *select:* NUDAT
    • telnet://db.nea.fr *User:* NEADB – *password:* GUEST – *select:* DATa bases – *select:* NUDAT
39. http://www.nndc.bnl.gov/nndcscr/pc_prog/ *Only* NDTxx_16.EXE *and* RADTIONS.ZIP *are required*.
40. R.B. Firestone, V.S. Shirley, C.M. Baglin, S.Y.F. Chu, J. Zipkin, *Table of Isotopes*, 8th edition (Book and CD-ROM + yearly update), J. Wiley & Sons, Chichester, 1996.
41. http://nucleardata.nuclear.lu.se/database/toi/
42. J.B. Cumming, 'CLSQ, the Brookhaven Decay Curve Analysis Program', in *Applications of Computers to Nuclear and Radiochemistry*, ed. G.D. O'Kelley, Symp. Proc. Gatlinburg, Tennessee, Office Technical Services, Dept. of Commerce, USA, 25–33, 1962.
43. S. Schreurs, K. Strijckmans, J.-P. François, *Comparison of Prony's Improved and Cumming's Methods for the Analysis of Multicomponent Radioactive Decay Curves*, submitted for publication, (1999).
44. G. Blondiaux, J.L. Debrun, C.J. Maggiore, 'Charged Particle Activation Analysis', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, J.C. Barbour, C.J. Maggiore, J.W. Mayer, Material Research Society, Pittsburgh, 205–230, 1995.
45. K. Strijckmans, 'Charged-particle Activation Analysis', in *Surface Characterisation: a User's Sourcebook*, eds. D. Brune, R. Hellborg, H.J. Whitlow, O. Hunderi, Wiley-VCH, Weinheim, 169–175, 1997.
46. K. De Neve, K. Strijckmans, R. Dams, 'Feasibility Study on the Characterization of Thin Layers by Charged

Particle Activation Analysis (CPAA)', *Anal. Chem.*, **72** (2000).

47. M. Parent, R. Cornelis, F. Alt, K. Strijckmans, R. Dams, 'Use of Pt-191 Radiotracer for the Development of Enrichment Procedures to Detect Natural Levels of Platinum in Biological and Environmental Materials', *Biol. Trace Elem. Res.*, **43–45**, 109–115 (1994).
48. M. Parent, K. Strijckmans, R. Cornelis, J. Dewaele, R. Dams, 'Production of Pt-191 Radiotracer with High Specific Activity for the Development of Preconcentration Procedures', *Nucl. Instrum. Methods B*, **86**, 355–366 (1994).

Cyclic Activation Analysis

Xiaolin Hou

Risø National Laboratory for Sustainable Energy, Technical University of Denmark, Roskilde, Denmark

1 Introduction **1**
2 Theory **2**
2.1 Equation of Cyclic Activation Analysis 2
2.2 Selection of Time Parameters 3
2.3 Estimation or Confirmation of the Half-life of Nuclide 6
2.4 Dead Time and Pileup Correction 6
2.5 Cumulative Activation Analysis 7
3 Classification and Facilities of Cyclic Activation Analysis **7**
4 Application of Cyclic Activation Analysis **8**
4.1 Determination of Se, F, and O in Biological Materials 8
4.2 Determination of Pb, F, Sc, and Ag in Environmental Samples 10
4.3 Determination of O, Sc, Si, Al, Ag, Au, Rh, Hf, Pb, and U in Geological and Industrial Materials 11
5 Conclusion and Perspective **12**
Abbreviations and Acronyms **12**
Related Articles **13**
References **13**

Cyclic activation analysis (CAA) is a method of activation analysis for elemental analysis, in which a sample is irradiated, decayed, counted, and then irradiated again. This process is repeated for a number of cycles, and the spectra from each counting are summed together to give one final total spectrum. By this process, the counts of a short-lived nuclide of interest are considerably increased, and the analytical sensitivity of elements is significantly improved. The most commonly used CAA is the cyclic neutron activation analysis (CNAA) by irradiation with the thermal, epithermal, and fast neutrons produced from nuclear reactor, accelerator, and isotopic neutron source. The nuclear reactor can supply a much high neutron flux and is most often used for this purpose. At least 20 elements produced short-lived nuclides (half-life <100 s) by thermal neutron bombardment, and more than 10 elements produced nuclides with half-life of 100–600 s. These can be determined by thermal and epithermal neutron CAA. This technique has been widely applied in biological, environmental, geological, and industrial studies, and most often measured elements include Se, F, Pb, Hf, Sc, O, Ag, and Rh.

The advantages of CAA, as compared with conventional activation analysis, include significant improvement in the detection limit, analytical precision, and accuracy for the elements by using short-lived nuclides; short experimental time and increased analytical number of samples per unit time; capability of estimation or confirmation of the half-life of the short-lived nuclide; and determination of the degree of homogeneity of a sample. However, the application of CAA is limited by the number of elements determined, because only some of the elements determined by conventional activation analysis can be determined by this method. In addition, dead time and pileup are serious problems in CAA and must be corrected. The principle, selection of optimal experiment conditions, detection limit, analytical precision of CAA, as well as the dead time and pileup corrections are discussed in this article. Some applications of this method are highlighted.

1 INTRODUCTION

Activation analysis has been well documented as a powerful technique for elemental analysis (*see* **Instrumental Neutron Activation Analysis**, **Charged Particle Activation Analysis**). Its high sensitivity and accuracy, simultaneous multielement analytical capability, and matrix-free and nondestructive analysis make it wide applicable in a variety of research fields such as biological, environmental, geological, industrial, archeological, and forensic studies. Of all kinds of activation analyses, reactor neutron activation has been the mostly used, and as many as 70 elements can be easily determined. However, for many elements, an analytical period from days to weeks is needed because the nuclides detected, which are produced by reaction of their stable isotope with neutron, have longer half-life, such as ^{75}Se (120 days), ^{46}Sc (84 days), ^{110m}Ag (250 days), and ^{181}Hf (45 days). To attain the required sensitivity, a long-time irradiation, delay, and counting are necessary. This reduces its competitive capability to other analytical techniques, such as inductively coupled plasma mass spectrometry (ICP–MS). The use of the short-lived nuclides of element in preference to their longer-lived activation products can significantly reduce the total experimental time, which in turn can reduce the analytical period and increase the number of samples measured per day, and make activation analysis more cost-effective and competitive. On the other hand, several elements can be determined

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

by neutron activation analysis (NAA) only through the measurement of short-lived nuclide, such as ^{20}F (11 s), ^{19}O (27 s), and ^{207m}Pb (0.8 s). Therefore, NAA using short-lived nuclides has been given more attention.

The saturation activation of short-lived radionuclides is reached quickly on irradiation, and there is no further increase in activity with time. Similarly, the counting is also limited by half-life since the activity of a short-lived radionuclide decays very quickly, and further counting cannot accumulate any more counts of radionuclides of interest; on the contrary, it increases the background derived from long-lived interfering nuclides. In addition, owing to the short half-life, the number of radioactive nuclei (N^*) formed in saturation is small ($N^* = N\sigma\Phi/\lambda = N\sigma\Phi t_{1/2}/\ln 2$, where N is the number of target nuclei, σ is the activation cross section, Φ is the neutron flux, λ is the decay constant, and $t_{1/2}$ the half-life of the formed radioactive nuclide). Consequently, the counting has a large statistical error, and both the analytical accuracy and detection limit are unsatisfied. It is possible, however, to overcome this by reirradiating sample and counting it again, once the initial activity has been decayed away. The number of counts of the nuclide of interest can be significantly increased by repeating this process of irradiation-decay-counting-decay and irradiation again for a number of cycles, i.e. cyclic activation, and summing the spectra from each counting to give a final spectrum. By this CAA, the sensitivity, precision, and accuracy will be considerably improved.

The history of CAA can be traced back to the beginning of 1960s. In 1960 and 1961,[1,2] Anders first reported a technique where a sample is cycled between the irradiating source and the radiation detector, in order to improve both the sensitivity and statistical certainty of the elemental analysis. In his experiment, he repeatedly irradiated the sample with neutrons from a Be target of a 2-MeV Van de Graff accelerator and counted on an NaI(Tl) detector for two consecutive periods of equal length. By subtracting the spectra acquired from each other, the resulting γ-ray difference spectrum represented the contribution from only very short-lived nuclides and suppressed the contribution of longer-lived components in the γ-spectra of the activated sample. The method was documented by determining fluorine using faster neutron reaction of ^{19}F(n, α)^{16}N, and Se, Ag, Hf, F, and O by thermal neutron (n, γ) reactions. Preliminary investigation indicated the usefulness of the technique for the determination of 18 elements, such as O, F, Na, Sc, Ge, Se, Br, Y, Rb, Rh, Ag, In, Er, Hf, W, Yb, Ir, and Au. However, the term *cyclic activation analysis* is not mentioned in this article. In 1966, Caldwell et al.[3] suggested the use of a similar technique as part of a combination neutron experiment for remote elemental analysis. The term *cyclic activation analysis* was first used by Givens et al. in 1968,[4] when they measured ^{16}N and ^{24m}Na using a technique similar to that of Anders. However, the irradiating source, rather than the sample, was cycled electronically.[5] Using this facility, the nuclides with half-life down to a few milliseconds, such as ^{24m}Na(20 ms), can be utilized for analytical purposes. In 1969, Tani et al.[6] used a similar experimental facility to observe the γ-spectra of ^{205}Pb ($\sim$4 ms) and ^{207m}Pb (800 ms).

In order to obtain the high sensitivity required while using CNAA as a useful trace element analysis technique, Ozek[7] and Spyrou et al.,[8] in 1973, first used nuclear reactor as a neutron source to supply a neutron flux more than three orders of magnitude higher than other sources used by the earlier workers, such as an accelerator and isotope sources for CNAA. They published their results on the measurement of lead through ^{207m}Pb in environmental samples.[8] Since 1970s, many facilities for CAA have been set up, and methodology and application studies were carried out.[9–16] This article is an updated version of the previous one published in the first edition of this book in 2000.[17]

2 THEORY

2.1 Equation of Cyclic Activation Analysis

The principle of CAA has been described by Given et al.[4] and Spyrou.[18] In this method, the sample is irradiated for a short period of time, and after a delay period from the end of irradiation, the radiation emitted is counted for a short period of time, then the sample is irradiated again and the entire process is repeated for a number of cycles (Figure 1). The detected radiations at each counting period are summed and finally a total cumulative detector response is obtained. The cycle period T is given by Equation (1):

$$T = t_i + t_d + t_c + t_w \tag{1}$$

where t_i is the time of irradiation, t_d is the delay time, i.e. the time between the end of irradiation and the start of counting, which is usually the time required to transfer the sample from the irradiation position to the counting station, or to transfer the irradiation source to the sample station, t_c is the counting time, and t_w is the waiting time, i.e. the time between the end of counting and the start of the next irradiation.

The detector response (or the number of counts) for the first cycle is given by Equation (2):

$$D_1 = \frac{N\sigma\Phi I\varepsilon}{\lambda}(1 - e^{-\lambda t_i})e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \tag{2}$$

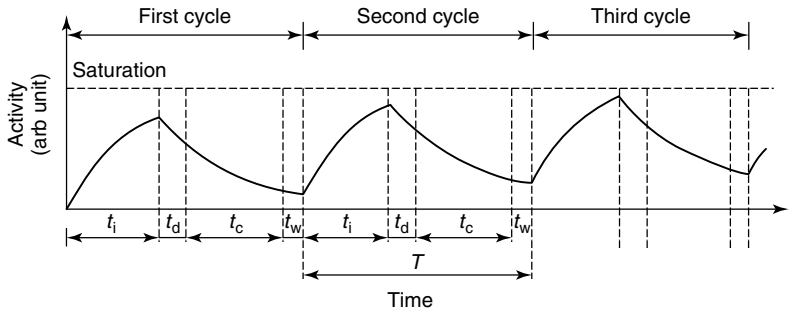

Figure 1 Time parameters of CAA and the variation of the activity of nuclide with time and cyclic number.

where the definitions of σ, Φ, and λ are the same as above, I is the intensity of the radiation of interest, and ε is the counting efficiency of the detector.

In the second counting period, the detector response is the same number of counts due to the second irradiation in addition to what was left from the first irradiation, expressed by Equation (3):

$$D_2 = D_1 + D_1 e^{-\lambda T} = D_1(1 + e^{-\lambda T}) \tag{3}$$

Similarly, for the nth cycle, which can be expressed by Equation (4)

$$D_n = D_1(1 + e^{-\lambda T} + e^{-2\lambda T} + e^{-3\lambda T} + \cdots + e^{-(n-1)\lambda T}) \tag{4}$$

Then the total cumulative detector response in all n cycles is given by Equation (5):

$$\begin{aligned} D_c &= \sum_{i=1}^{n} D_i \\ &= D_1 \left[\frac{n}{1 - e^{-\lambda T}} - \frac{e^{-\lambda T}(1 - e^{-n\lambda T})}{(1 - e^{-\lambda T})^2} \right] \\ &= \frac{N \Phi \sigma I \varepsilon}{\lambda} \left[\frac{n}{1 - e^{-\lambda T}} - \frac{e^{-\lambda T}(1 - e^{-n\lambda T})}{(1 - e^{-\lambda T})^2} \right] \\ &\quad \times (1 - e^{-\lambda t_i}) e^{-\lambda t_d} (1 - e^{-\lambda t_c}) \end{aligned} \tag{5}$$

Equation (5) gives the basic relationship for CAA.

2.2 Selection of Time Parameters

The detector response may be maximized by a proper selection of the parameters in Equation (5). For a given total experiment time, T_t ($T_t = nT = mt_{1/2}$), the maximal value of D_c occurs when $t_d = t_w = 0$ and $t_i = t_c$. However, the fact is that t_d and t_w are not zero and their values ultimately depend on the transfer system used. When the transfer is done manually, the transfer time can be relatively long in the value of 10 s to 1 min; in this case, the method sometimes is referred to as *pseudocyclic activation analysis* (PCAA).[12,19,20] In some fast transfer system, where the transfer is completed electronically and through computer control, the transfer time is of the order of 0.1 s and less.[21,22] The transfer time can be made very small when there is no physical transfer of the sample but a pulsed irradiation is used, which was suggested as the "real" CAA.[3,4] For the following discussion, unless otherwise stated, it is assumed that $t_d = t_w = 0$ and $t_i = t_c$.

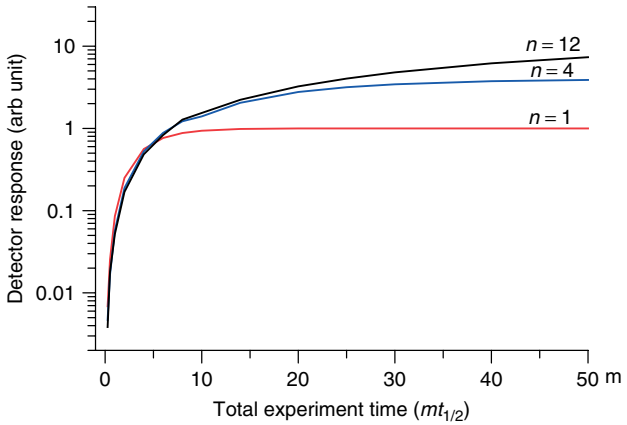

Figure 2 Variation of the conventional and cumulative signal with total experiment time, $t_d = t_w = 0$, $t_i = t_c$ and n is the number of cycles.

Figure 2 shows the variation of detector response for the nuclides of interest with the total experimental time ($T_t = mt_{1/2} = nT$) in conventional and cyclic case with different cyclic numbers. It is clear that below a certain experimental time, conventional activation is to be preferred to cyclic activation, and the cross over point of curves between conventional and cyclic activation shifts to longer experimental time with an increasing number of cycles. After the cross over point, cyclic activation has a higher detector response than that of conventional one-shot irradiation. For a certain total experiment time, such as $T_t = 10t_{1/2}$, the cumulative detector response increases with increasing number of cycles in the beginning, and there is an optimum cyclic number, after which the detector response does not increase and slowly decreases (Figure 3). In addition, for conventional single-shot activation analysis, the detector response increases quickly

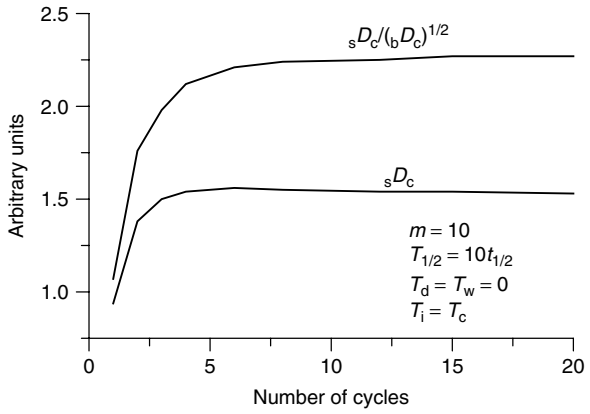

Figure 3 Variation of signal and signal-to-noise ratio $({}_sD_c/\sqrt{{}_bD_c})$ for cyclic and conventional activation with the number of cycles. Here, the underlying background to signal of interest is provided by the activity of nuclide with half-life equal to 10 times half-life of nuclide of interest ($T_{1/2} = 10_{1/2}$).

with increase in the total experiment time in the beginning and 95% of saturation activity is obtained before $5t_{1/2}$, following which a further increase in the total experiment time becomes insignificant (Figure 2). However, for CAA, this time point becomes longer. For $n = 4$, 95% of saturation activity is obtained after $15t_{1/2}$, and for $n = 12$, after a total experiment time of $50t_{1/2}$, the activity still increases with the increase in total experiment time.

However in CAA, the radionuclides produced from elements in the matrix, which has a longer half-life than the nuclide of interest, will decay slowly between irradiations and therefore the underlying background will increase through successive irradiation. Consequently, the detection of an element does not depend solely on the signal from the nuclide of interest, but also depends on the background contribution from the matrix elements.

The purpose of CAA is to improve the analytical precision and detection limit for the elements of interest. The analytical precision is usually expressed by relative standard deviation of a signal S measured above a background B (Equation 6):

$$\delta = \frac{\sqrt{S + 2B}}{S} \tag{6}$$

For the cumulative spectrum of n cycles, it is expressed as Equation (7):

$$\delta n = \frac{\sqrt{({}_sD_c + 2{}_bD_c)}}{{}_sD_c} \tag{7}$$

where ${}_sD_c$ is the cumulative detector response for the nuclide of interest, ${}_bD_c$ is the cumulative detector response for the nuclide making the major background contribution to the signal from the nuclide of interest.

The detection limit L_d is usually given by Equation (8):

$$L_d = 2\sqrt{B} \qquad \text{(counts)}$$

$$\text{or} \quad L_d = \frac{2\sqrt{B}}{S/m} = \frac{2m\sqrt{B}}{S} \qquad (\mu g) \tag{8}$$

where S/m is the counts per unit mass element of interest.

For CAA with n cycles, the detection limit is given by Equation (9):

$$L_d = \frac{2\sqrt{{}_bD_c}}{{}_sD_c/m} = \frac{2m\sqrt{{}_bD_c}}{{}_sD_c} \tag{9}$$

Therefore, $(\sqrt{{}_bD_c}/{}_sD_c)$ or $({}_sD_c/\sqrt{{}_bD_c})$ and $(\sqrt{{}_sD_c + 2{}_bD_c}/{}_sD_c)$ or $({}_sD_c/\sqrt{{}_sD_c + 2{}_bD_c})$, not the detector response for the nuclide of interest, must be optimized, when the most suitable cyclic timing parameters are chosen for a nuclide.

In Figures 4 and 5, the signal-to-noise ratio $({}_sD_c/\sqrt{{}_bD_c})$, and the ratio of signal to the square root of sum of ${}_sD_c$ and $2{}_bD_c$, $({}_sD_c/\sqrt{{}_sD_c + 2{}_bD_c})$, are plotted versus total experiment time $(mt_{1/2})$ for comparison of cyclic with conventional activation analysis. In this case, for illustrative purpose, assuming that only one long-lived nuclide with a half-life $(T_{1/2})$ of 10 times half-life of nuclide of interest $(T_{1/2} = 10t_{1/2})$ is considered to contribute to the underlying background of the signal. It can be seen that the background plays a significant role in emphasizing the difference between conventional and cyclic cases. With increase in the total experiment time, $({}_sD_c/\sqrt{{}_bD_c})$ and $({}_sD_c/\sqrt{{}_sD_c + 2{}_bD_c})$ have a maximum and this maximum shift to a longer experimental time with increasing number of cycles (Figures 4 and 5). Comparing with Figure 2, the cross over point between conventional and cyclic activation in Figure 4 is shifted to a lower value of m,

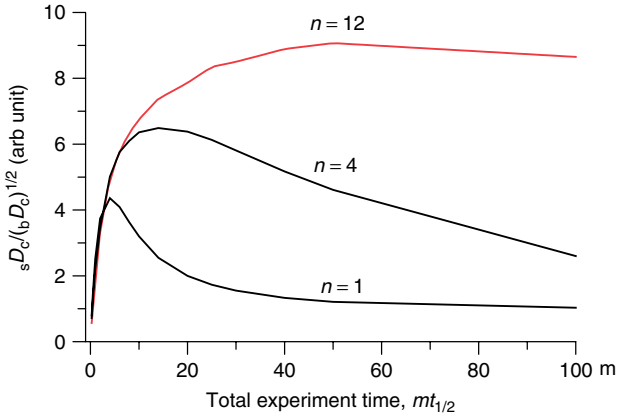

Figure 4 Variation of signal-to-noise ratio $({}_sD_c/\sqrt{{}_bD_c})$ for cyclic and conventional activation with the total experiment time $(T_{1/2} = 10t_{1/2})$.

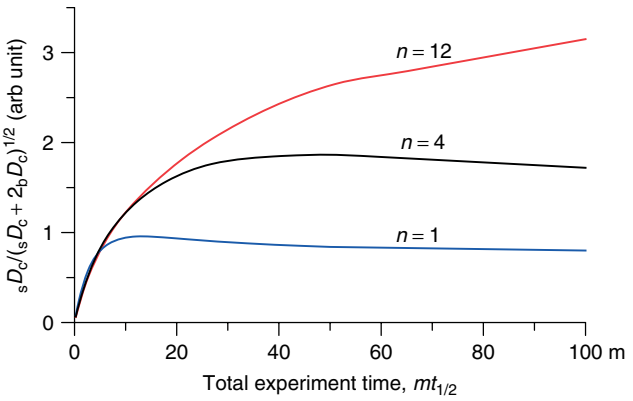

Figure 5 Variation of $({}_{s}D_{c}/\sqrt{{}_{s}D_{c}+2{}_{b}D_{c}})$ for cyclic and conventional activation with the total experiment time $(T_{1/2}=10t_{1/2})$.

i.e. cyclic activation becomes advantageous for shorter total experiment time than the case just considering the detector response.

Figure 6 shows the variation of signal-to-noise with the total experimental time (number of cycle) under a certain cycle period $(T=2t_{1/2})$. It is clear that in this case, signal-to-noise ratio increases always with increasing total experiment time.

To obtain a better cyclic advantage factor over conventional activation, it is obvious that one must increase the total experiment time $(mt_{1/2})$. Choice of the best total experiment time for cyclic activation is therefore a compromise between the time available for sample analysis and the detection limit to be achieved. Whenever the total experiment time is chosen, the number of cycles (n) or cycle period, $T=T_{t}/n$, has to be decided. In Figure 3, the variation of the signal $({}_{s}D_{c})$ and the signal-to-noise ratio $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ were plotted as a function of the number of cycles $(n=T_{t}/T)$. It can be seen that after a certain cycle, $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ and ${}_{s}D_{c}$ increase very slowly, and with further increase in the number of cycles these become

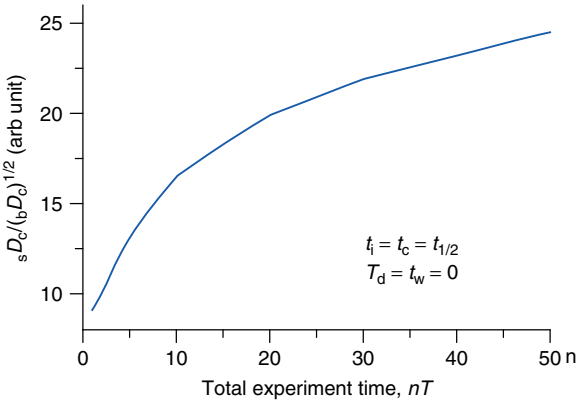

Figure 6 Variation of signal to noise $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ with the total experiment time (nT) for a certain cycle period $(T=2t_{1/2})$.

insignificant. After a maximum occurs, ${}_{s}D_{c}$ decrease slowly. For example, if $T=10t_{1/2}$ and $T_{1/2}=10t_{1/2}$, 95% of saturation activity of nuclides can be obtained after four cycles and 95% of a maximum of $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ after eight cycles. In addition, in order to maximize signal $({}_{s}D_{c})$, the optimal number of cycles is always smaller than that required for maximizing the signal-to-noise ratio $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$, and it also increases with the increase in the half-life of the background for the same total experiment time. However, it becomes less pronounced for the half-life of the background more than $100t_{1/2}$.

Spyrou et al.[18] and Spyrou and Kerr[23] investigated the effect of the transfer time (t_{d} and t_{w}) on $({}_{s}D_{c})$ and $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$, they found that the cross over point between cyclic and conventional activation was shifted to a longer experiment time with increasing transfer time. It means that CAA becomes more preferable to the conventional method with decreasing transfer time, and the transfer time is an important factor effecting the detection limit and precision.

In real situation, many long-lived nuclides from matrix elements, not just one, will contribute to the underlying background of the peak of the nuclide of interest after activation. For calculating these contributions and optimizing the parameters of cyclic activation, many computer programs have been developed.[24–26] In Tout and Chatt's program,[25] by entering total experiment time, transfer time (t_{d}, t_{w}), the mass of sample and relevant nuclear data of matrix elements contributing to the background activity, as well as the half-life of the nuclide of interest, $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ can be calculated for individual matrix elements and the total background. The optimal irradiation and counting time (or number of cycles) and its ranges were finally given for maximum and 95% of maximum of $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$. Their results indicated that a single value of t_{i} and t_{c} can be used to produce results close to the maximal $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ for nuclides with a large range of half-lives, i.e. for a total experiment time of 5 min, $t_{i}=t_{c}=6$–7 s will produce values of more than 90% of maximal $({}_{s}D_{c}/\sqrt{{}_{b}D_{c}})$ for nuclides with half-lives in the ranges 3–22 s. This means that a number of short-lived nuclides can be measured at close to their best detection limits by using only one set of CAA.

In Al-Mugrabi and Spyrou's program,[26] besides the contribution from the photopeaks of nuclides produced by matrix elements to the background, the Compton continuum, single and double escape peaks, and bremsstrahlung were also considered. By entering the information in Tout and Chatt's program, as well as by entering neutron flux, mass of sample and nuclear data of nuclide of interest and detector parameters, not only the optimized activation conditions were produced but also simulated γ-spectrum, detection limits, and precision for the elements of interest were presented.

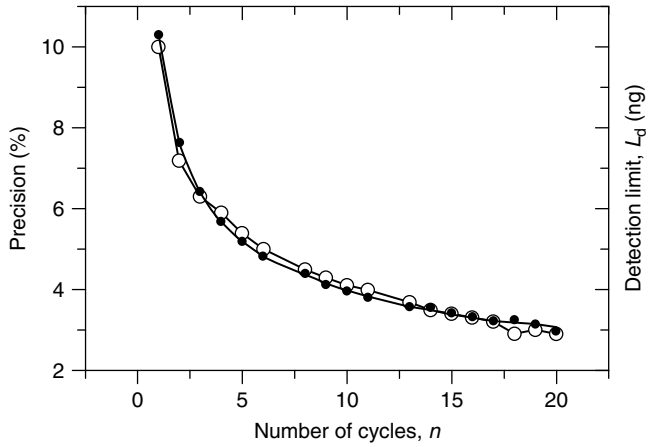

Figure 7 Variation of analytical precision and detection limit of CNAA for Se in human hair. $t_i = t_c = 30$ s, $t_d = t_w = 3$ s.

Figure 7 shows the experimentally measured variation of detection limit and analytical precision of CNAA for Se in human hair, with cyclic period of $T = t_i + t_d + t_c + t_w = 30 + 3 + 30 + 3 = 66$ s. It can be clearly seen that CNAA can significantly improve the detection limit and precision.

2.3 Estimation or Confirmation of the Half-life of Nuclide

Another function of CAA, estimation and confirmation of the half-life of nuclide by using the data obtained during the CAA, was suggested by Spyrou.[18,27,28]

For large n, $(1 - e^{-n\lambda T})$ tends to unity, Equation (5) can be reduced as

$$D_c = D_1 \left[\frac{n}{(1 - e^{-\lambda T})} - \frac{e^{-\lambda T}}{(1 - e^{-\lambda T})^2} \right] = \frac{nD_1}{(1 - e^{-\lambda T})} - \frac{D_1 e^{-\lambda T}}{(1 - e^{-\lambda T})^2} \quad (10)$$

where D_c is a linear function of n with all other parameters constant. Plotting D_c as a function of n, the slope of the line is given by Equation (11):

$$a = \frac{D_1}{(1 - e^{-\lambda T})} \quad (11)$$

and the intercept by Equation (12):

$$b = \frac{D_1 e^{-\lambda T}}{(1 - e^{-\lambda T})^2} \quad (12)$$

The half-life of the nuclide measured can be calculated by Equation (13):

$$t_{1/2} = \frac{T \ln 2}{\ln (1 - a/b)} \quad (13)$$

For this purpose, it is necessary to store the spectra of each individual cycle instead of using the accumulative spectrum.

2.4 Dead Time and Pileup Correction

CAA is usually employed for the determination of short-lived nuclides in a matrix of long-lived activation products. The activity of a sample not only changes considerably during a counting period owing to short-lived nuclide decay but also increases from cycle to cycle because of the matrix activity owing to the accumulation of longer-lived products. Consequently, a rapidly changing dead time is encountered in a counting period. The basic Equation (14) for dead time correction was given by Schonfeld[29]:

$$C = \int_0^{t_c} A_0 e^{-\lambda t} [1 - DT(t)] \, dt \quad (14)$$

where C is the actual acquired net counts in photopeak of interest, A_0 is the true initial photopeak count rate, and $DT(t)$ is the fractional analyzer dead time at time t. In order to implement this correction, the variation of fractional dead time during the counting period must be known. By the least square fitting of the experimental data, Egan et al.[30] found that $DT(t)$ is an exponential function of t, and it can be expressed by Equation (15):

$$DT(t) = B + Ce^{-kt} \quad (15)$$

where B, C, and k are constants. Hence, a correction factor (f) for dead time in a counting period can be obtained using Equation (16).

$$f = \frac{\int_0^{t_c} A_0 e^{-\lambda t} \, dt}{\int_0^{t_c} A_0 e^{-\lambda T} (1 - B - Ce^{-kt}) \, dt} \quad (16)$$

For cyclic activation with n cycles, the correction factor (F_n) can be expressed by Equation (17):

$$F_n = \frac{\sum_{n=1}^{n} \int_0^{t_c} A_0 e^{-\lambda t} dt}{\sum_{n=1}^{n} \int_0^{t_c} A_0 e^{-\lambda t} (1 - B - Ce^{-kt})_n dt} \quad (17)$$

$DT(t)$ not only varies with t in a counting period but also differs from one cycle to another. F_n is also a function of n.

It was observed that the correction factor for the dead time, f, is approximately equal to the ratio of clock time

to live time obtained from the multichannel analyzer clock data in each counting period.[30] Hence, it is usually used to correct for dead time after each cycle. In this case, it is important to store the γ-spectra of each cycle period and correct the counts of interest nuclide for dead time before summation of individual spectrum.

With increasing analyzer dead time, the pulse pileup loss (also called *summing effects* or *coincidence losses*) will become more serious. It can be overcome satisfactorily by introducing a pulser with constant frequency into the system.[30] The ratio of the areas under the pulser peaks, after dead time correction, was measured and the pileup correction factor was defined as F_p (Equation 18):

$$F_p = \frac{\text{Pulser peak area (at } \sim 0\% \text{ dead time)}}{\text{Pulser peak area (at sample dead time)}} \qquad (18)$$

However, pileup correction, unlike dead time, is implemented on the cumulative spectrum.

Wyttenbach[31] suggested another way to correct the pileup loss by using the real time (cock time), t_c, and the live time of the measurement, T_c (Equation 19):

$$F_p = \frac{P}{P_0} = 1 - \frac{2\tau}{\nu}\left(\frac{t_c}{T_c} - 1\right) \qquad (19)$$

where P_0 is the true photopeak count rate without coincidence losses, P is the actual photopeak count rate including pileup, ν is a constant of the detector system, and τ/ν can be measured by plotting P/P_0 versus t_c/T_c and subsequently be used to correct other spectra. However, it should be considered that actually τ depends on the energy of the γ ray measured.[32] In addition, the use of a loss-free counting system with pileup rejection will avoid this problem.[33]

2.5 Cumulative Activation Analysis

Guinn[34] suggested another method similar to CAA, which is based on the irradiation and counting of a number of replicates of the same sample and the individual spectra are summed to give one total spectrum. As the activity derived from activated matrix elements is the same for each cycle, the background count and the dead time in each cycle does not increase with the increasing number of replicates. Hence, this will lead to a better precision and detection limit than that of normal CAA for short-lived nuclides. The main advantage of this method is that a large representative sample can be measured even in the case of material, which activates to form high radioactive products.

By this method, Parry[35,36] measured rhodium, silver, and gold in some geological samples and reference materials and found that the detection limit was improved by a factor of 4.4–4.7 using 20 replicates in accordance with the theoretical factor of $n^{1/2}$. The main disadvantage of this method is the long time required for the preparation of n samples for each material to be analyzed, and the application is limited by the amount of samples available.

For overcoming the relatively high error resulted from inhomogeneity of sample, Farooqi et al.[37] suggested combining this technique (called *mass fraction*) with normal CAA to analyze fluorine in diet sample. In addition to good precision, a lower detection was also obtained.

3 CLASSIFICATION AND FACILITIES OF CYCLIC ACTIVATION ANALYSIS

According to source-target-detector system, Spyrou[18] classified CAA in the following five types:

1. An ''external'' beam of incident radiation, which can be mechanically chopped or electronically pulsed with stationary target, such as pulsed neutron source (or called *neutron generator*) as used by Givens et al.,[4,5] in which the waiting time can be as short as 0.5 ms. This type of CAA was termed *real* CAA. However, the radiation flux supplied in this system is usually low, i.e. the total 4π neutron flux is only $1.16 \times 10^{11}\,\text{n}\,\text{s}^{-1}$ in the facility used by Given et al.[4,5]
2. An ''external'' beam providing a continuous source of incident radiation, with the target moving into and out of the beam either linearly or by rotation. The typical source of this type of CAA is neutron source produced by an accelerator with which, in addition to fast neutron, thermal neutrons can be supplied. Further, the energy of neutron can also be chosen, which is very useful to avoid some interfering nuclear reactions. The neutron flux supplied usually ranges from 10^6 to $5 \times 10^{11}\,\text{n}\,\text{cm}^{-2}\,\text{s}^{-1}$. Most fast neutron CAA has been carried out in this kind of facility.[1,2,14,38,39]
3. An ''enclosed'' beam where the target is made to oscillate either cyclically or linearly into and out of the beam. The main radiation source in this type of CAA is the core of reactor. Not only can a high neutron flux (10^{11}–$10^{15}\,\text{n}\,\text{cm}^{-2}\,\text{s}^{-1}$) be used but also the thermal, epithermal, and fast neutron activation can be carried out. By detection of γ ray by coaxial HpGe (or Ge(Li)) and low-energy photon detector and delay neutron by neutron detector, most elements can be determined with a high accuracy and a good detection limit. Most CAA work has been implemented using this type of system.[7–13,18–20,23,25,27,30,37] Figure 8 shows a schematic diagram of this system. Similar facilities

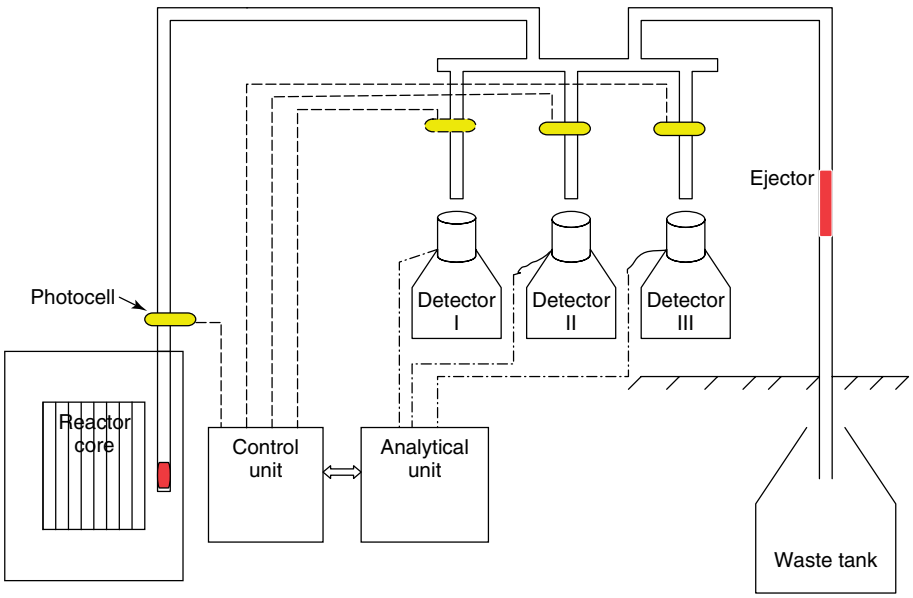

Figure 8 Schematic diagram of CNAA system.

have been set up for CNAA in such institution as University of London Reactor Center, UK,[10] Dalhousie University, Canada,[9] MNSR, China Institute of Atomic Energy,[13] MNSR, Syria,[40] MNSR, Shenzhen University, China (Personal communication), JRRS-SM, Japan Atomic Energy Research Institute, Japan,[41] Atominstitut, Vienna,[12] University of Missouri, USA.[39]

4. The case of an isotopic source, where the target remains stationary but the source is made to oscillate to and from the target position. The most commonly used isotope neutron source is Am–Be and ^{252}Cf neutron source.[42,43] A useful application of this system is in vivo NAA. The neutron flux supplied in this system is only 10^4–10^6 n cm^{-2} s^{-1}, and the detection limit is therefore high.
5. A primary source creating a secondary source, which in turn provides the irradiation beam for the target. An example of this case is the neutron interrogation system described by Mardani,[44] where a neutron beam is made to impinge on nuclear fuel, which on fission produces neutrons that act as the secondary source for the activation of a continuously circulating fluid (freon) containing fluorine. The reaction ^{19}F(n, α)^{16}N was used to determine the amount of fissile material in the fuel element by cyclic measurement of ^{16}N (7.2 s).

In all types of CAA, most work has been carried out using CNAA. Depending on the neutron source used, CNAA can be classified as reactor, neutrons produced by accelerator and neutron generator, and isotopic source CNAA. On the basis of the energy of neutron, it can be classified as thermal, epithermal, and fast neutron CAA; on the basis of the radiation emitted from the activated nuclide, it can be classified as γ ray and delay neutron CAA. Table 1 lists the elements that can be determined using different types of CNAA. The nuclear parameters of elements probably detected by cyclic thermal NAA and cyclic fast neutron activation analysis (FNAA) using their short-lived nuclides ($T_{1/2} < 100$ s) are listed in Tables 2 and 3, respectively. In addition to the elements listed in Tables 2 and 3, some elements with nuclides of medium half-lives ($t_{1/2} > 100$ s), such as Al, Mg, Ca, V, Cu, S, Mn, Na, Cl, K, Ti, Co, Br, I, In, Sr, U, and Th, may occur in the cumulative γ-spectrum or can be further counted a few minutes after the last cycle. These can, therefore, also be determined simultaneously.

4 APPLICATION OF CYCLIC ACTIVATION ANALYSIS

CAA was first suggested for the analysis of geological materials,[3–5] but with its development and maturity, this technique has been used in every field of elemental analysis, and there are numerous articles dealing with the application of CAA. Some main applications are discussed below.

4.1 Determination of Se, F, and O in Biological Materials

Selenium is known as a *biological essential trace element*, its concentration in normal biological tissues is lower than 1.0 μg g^{-1}, and a sensitive analytical technique is

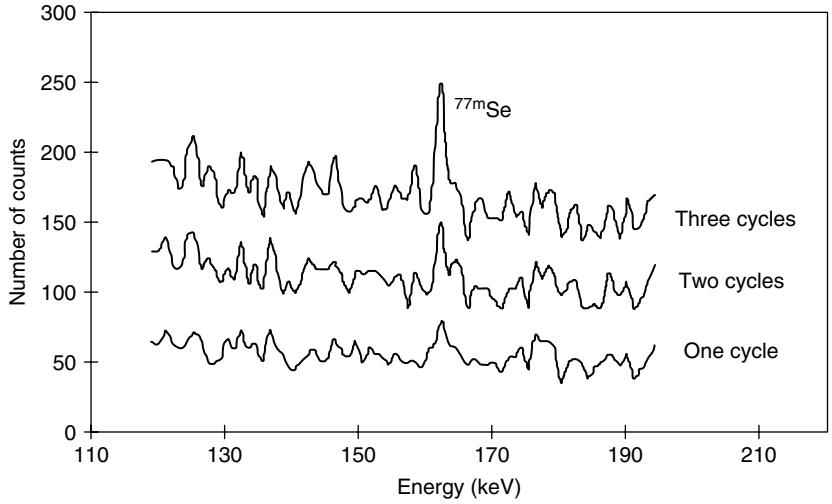

Figure 9 γ spectra of biological sample for ^{77m}Se by single-shot NAA and CNAA.

Table 1 Classification of CNAA and elements detected

Neutron	Source	Radiation measured	Elements detected
Thermal	Reactor, accelerator	γ ray	O, F, Sc, Ge, Se, Rb, In, Rh, Pd, Ag, Sb, Dy, Hf, Ce, Er, Yb, Ir, W, Pb, U
Thermal	Reactor	Delay neutron	U, Th
Epithermal	Reactor	γ ray	O, F, Ge, Rh, Ag, Ce, Er, Hf, Pt, Ir
Fast	Reactor, accelerator, isotopic source, neutron generator	γ ray	B, O, F, S, Cl, Si, Na, Mg, Ti, Ge, As, Se, Br, Rb, Y, In, Ba, Ce, Hf, Nd, Sm, Er, Tb, W, Au, Pb

required for its quantitative determination. NAA is a sensitive method for the determination of Se. In most cases, analysis is performed by measuring ^{75}Se, which has a longer half-life (120 days), and therefore needs a relatively long analytical period (weeks to months). The uses of short-lived isotope ^{77m}Se can considerably reduce the analytical time to a few minutes and thus the analytical expense. However, the detection limit and analytical precision are usually not good enough by conventional single-shot irradiation due to quick decay of ^{77m}Se. CNAA satisfactorily solved this problem, and therefore has been widely used for this purpose.[13,19,20,30] Figure 9 shows γ-spectra of ^{77m}Se in biological sample acquired by conventional single-shot NAA and CNAA. Although the background is increased, the signal of ^{77m}Se and the signal-to-background ratio are significantly improved by CNAA. Shi et al.[48] compared the conventional NAA with CNAA for the determination of Se in different food samples. In general, the detection limit of Se by measuring ^{77m}Se ($t_i : t_d : t_c = 30\,s : 10\,s : 30\,s$) is better than that by measuring ^{75}Se ($t_i : t_d : t_c = 17\,h : 16\,days : 12\,h$) using the same neutron flux by a factor of 1.5, and the detection limit using CNAA with three cycles is two times better than the single-shot NAA.

Figure 7 shows the detection limit and analytical precision for Se in human hair by conventional and cyclic NAA, respectively. It can be clearly seen that both detection limit and analytical precision were significantly improved by CNAA. Considering the interference from ^{38m}Cl ($t_{1/2} = 0.715\,s$), the optimal time parameters of CNAA of biological samples for Se given by Egan et al.[30] are $t_i = 19.5\,s$, $t_d = 3\,s$, $t_c = 18\,s$, and $t_w = 1\,s$. Under these conditions, the samples with Se concentrations as low as $10\,ng\,g^{-1}$ have been analyzed,[49] and usually 10 samples can be treated in 1 h. CNAA has also been used for the determination of Se in urine[50] and nail,[51] and as low as $1.8\,ng\,g^{-1}$ Se in urine samples has been measured.[50] In MNSR, China Institute of Atomic Energy, more than 2000 varieties of biological samples, such as human and animal tissues and body fluids, diet, and vegetation, have been analyzed for Se using this method in the last few years.

Fluorine is another human essential trace element and a bone seeker. Dental caries and osteoporosis were thought to be related to the intake of F and

Table 2 Nuclear data of elements (nuclides) detected by thermal and epithermal neutron CAA using short-lived nuclides ($t_{1/2} < 100\,s$)[45,46]

Element	Nuclei produced	Abundance of target nuclei	Half-life of produced nuclei (s)	σ_0 (b)	I_0 (b)	Main γ-ray energy (keV)
O	^{19}O	0.002	26.9	0.000161	0.00081	197.14
F	^{20}F	1.00	11.03	0.0095	0.039	1633.60
Cl	^{38m}Cl	0.242	0.715	0.05	0.38	671.3
Sc	^{46m}Sc	1.00	18.75	9.6	4.9	142.53
Ge	^{75m}Ge	0.365	47.7	0.143	0.35	139.6
Ge	^{77m}Ge	0.078	52.9	0.092	1.0	215.48
Se	^{77m}Se	0.090	17.45	21	16	161.93
Rb	^{86m}Rb	0.722	61.2	0.05	1.16	556.17
Tc	^{100}Tc	–	15.8	20	30	539.5
Rh	^{104}Rh	1.00	42.3	134	1275	555.8
Pd	^{107m}Pd	0.273	20.9	0.013	0.2	214.9
Ag	^{110}Ag	0.482	24.6	89	112	657.76
In	^{116m}In	0.957	14.1	87	–	1293.6
Sb	^{124m}Sb	0.427	93.0	0.035	0.93	645.86
Ce	^{139m}Ce	0.0025	56.4	0.15	2	757.0
Dy	^{165m}Dy	0.282	75.6	1698	425	515.5
Er	^{167m}Er	0.336	2.27	15	10	207.8
Yb	^{177m}Yb	0.127	6.41	3.8	–	104
Pt	^{199m}Pt	0.072	13.6	0.3	–	319
Hf	^{179m}Hf	0.273	18.68	53	1039	216
Ir	^{192m}Ir	0.373	87	300	1060	58
W	^{183}W	0.265	5.65	20	600	107.9
Pb	^{207m}Pb	0.241	0.80	0.03	0.1	570

the concentration in the tissues. Only one isotope, ^{20}F, can be used for its determination by NAA. Owing to very short half-life of ^{20}F (11.2 s), CNAA was therefore widely used for its determination.[9,11,37,52] Under the conditions of $t_i = t_c = 10\,s$ and $t_d = t_w = 2\,s$, a detection limit of $1–10\,\mu g\,g^{-1}$ for diet sample can be obtained by 14 cycles.[37] For the same condition, the detection limit of F in bone was reported to be $52\,\mu g\,g^{-1}$ [52]; this is mainly attributed to the high background under ^{20}F γ-ray peak (1633 keV), which is due to high ^{28}Al signal (1779 keV) in the spectra for bone samples. Since the reactions $^{23}Na(n, \alpha)^{20}F$ and $^{20}Ne(n, p)^{20}F$ can also produce ^{20}F, and the concentration of Na is much high in biological samples, the sodium interference must be corrected for. In addition, this interference can be significantly reduced using well-thermalized neutrons, such as irradiating in a thermal neutron irradiation facility in research reactors. Fluorine was actually the first element to be determined in CAA.[1,2] In addition to Se and F, De Silva and Chatt[20] also determined other elements, such as Rb, Ag, Sc, Al, Ca, Cl, Cu, K, Mg, Na, V, and I, in biological samples simultaneously by using one analytical condition; Cheng et al.[45] has also determined Ag and Sc in nail samples using CNAA.

Oxygen is a major component element of human and animal tissues and it is closely related with the metabolism of energy and mass. On the basis of low natural abundance of ^{18}O, Hou and Jiang[53] studied the metabolism of oxygen in mouse body by tracer with enriched isotope of ^{18}O and CNAA by measuring ^{19}O (26.9 s) produced by $^{18}O(n, \gamma)^{19}O$ reaction. Under time parameters of $t_i = t_c = 20\,s$, $t_d = t_w = 3\,s$ and six cycles, 0.01% of variation of the concentration of oxygen in animal blood was observed. Fast neutron reaction $^{19}F(n, p)^{19}O$ may interfere with the determination of ^{18}O; however, the low concentration of F in tissues (except bone and tooth) and low flux of fast neutron in the reactor make this interference not a serious problem.

4.2 Determination of Pb, F, Sc, and Ag in Environmental Samples

Lead is well known as a *toxic element* and its distribution as a pollutant is widespread. It is, therefore, of interest to determine its concentration in a variety of environmental materials. Lead cannot be determined by thermal NAA unless a very short-lived radionuclide ^{207m}Pb was used.[54] However, the half-life of ^{207m}Pb of only 0.8 s makes it difficult to be measured using conventional NAA method. By using a fast transfer pneumatic system, Egan and Spyrou[22] determined lead in some environmental samples. Under the conditions, $t_i = t_c = 2\,s$, $t_d = t_w = 1\,s$, and 50 cycles, the sensitivity in an interference-free matrix

Table 3 Nuclear data for elements (nuclides) detected by fast neutron CAA using short-lived nuclides ($t_{1/2} < 100\,s$)[46,47]

Element	Reaction	Abundance of target nuclei	Half-life of produced nuclei (s)	σ (mb)	Main γ-ray energy (keV)
B	$^{10}B(n, p)^{11}Be$	0.200	13.8	3.3	2125
O	$^{16}O(n, p)^{16}N$	0.99756	7.1	40	6129
F	$^{19}F(n, p)^{19}O$	1.00	27.1	20	197.14
Na	$^{23}Na(n, p)^{23}Ne$	1.00	38	44	440
	$^{23}Na(n, \alpha)^{20}F$	1.00	11.0	150	1633.60
Mg	$^{25}Mg(n, p)^{25}Na$	0.10	59.6	45	583
	$^{26}Mg(n, \alpha)^{23}Ne$	0.11	37.6	75	440
S	$^{34}S(n, p)^{34}P$	0.042	12.4	73	2127
Cl	$^{37}Cl(n, \alpha)^{24}P$	0.2423	12.4	101	2127
Ti	$^{46}Ti(n, p)^{46m}Sc$	0.080	18.7	48	142.53
Cr	$^{53}Cr(n, p)^{53}V$	0.0950	96.0	48	1006.2
Fe	$^{58}Fe(n, p)^{58}Mn$	0.0031	65	7	810.8
Ge	$^{76}Ge(n, p)^{76}Ga$	0.077	29	9.5	562.9
	$^{76}Ge(n, 2n)^{75m}Ge$	0.078	48	1000	139.6
As	$^{75}As(n, p)^{75m}Ge$	1.00	48	16	139.6
Se	$^{80}Se(n, p)^{80}As$	0.50	16	16	665.9
	$^{78}Se(n, 2n)^{77m}Se$	0.235	17.5	840	161.93
Br	$^{79}Br(n, n')^{79m}Br$	0.5069	4.9	653	207
Rb	$^{87}Rb(n, 2n)^{86m}Rb$	0.2783	61.2	750	556.17
Ru	$^{100}Ru(n, p)^{100}Tc$	0.126	15.8	15	539.5
Rh	$^{103}Rh(n, \alpha)^{100}Tc$	1.00	15.8	11	539.5
Pd	$^{104}Pd(n, p)^{104}Rh$	0.11	42.3	2.7	555.8
	$^{108}Pd(n, 2n)^{107m}Pd$	0.267	21.3	517	215
Nd	$^{141}Nd(n, 2n)^{141m}Nd$	0.2713	62.4	0.5	756.5
Sm	$^{144}Sm(n, 2n)^{143m}Sm$	0.031	66	540	754
Tb	$^{159}Tb(n, 2n)^{158m}Tb$	1.00	10.5	451	110
Er	$^{168}Er(n, 2n)^{167m}Er$	0.271	2.28	700	207.8
Au	$^{197}Au(n, n')^{197m}Au$	1.00	7.7	291	278
Hf	$^{180}Hf(n, 2n)^{179m}Hf$	0.352	18.7	600	216
Pb	$^{208}Pb(n, 2n)^{207m}Pb$	0.523	0.8	1650	570

was found to be 5 μg, and a detection limit of 30 μg g^{-1} was reported for biological matrix.

Owing to a high-resonance integral cross section of reaction $^{19}F(n, \gamma)^{20}F$, Parry et al.[55] determined fluorine in moss and soil sample with an improved detection limit by using epithermal neutron and cyclic activation.

Silver and scandium are usually determined by using their long-lived nuclides, ^{110m}Ag and ^{46}Sc. However, the use of CNAA not only reduces the analytical time but also improves the detection limit of these two elements by using their short-lived nuclides, ^{110}Ag and ^{46m}Sc. Chatt et al.[9,20] and Spyrou et al.[23] have used CNAA to analyze some environmental samples for Sc and Ag as well as other elements, such as Dy, Hf, Se, Al, V, etc.

In addition, fast neutron CAA was also used for the determination of Pb and F in environmental samples by using reactions $^{208}Pb(n, 2n)^{207m}Pb$ and $^{19}F(n, p)^{19}O$; a comparable detection was reported for the determination of F by epithermal neutron CAA and fast neutron CAA.[55] In addition to Pb and F, fast neutron CAA has also been used for many other elements such as As, Ce, Nd, and Hf.[39]

4.3 Determination of O, Sc, Si, Al, Ag, Au, Rh, Hf, Pb, and U in Geological and Industrial Materials

CAA was proposed early by Caldwell et al.[3] for the elemental analysis of lunar and planetary samples. Using this technique, [4,5] they successfully detected O, Si, Mg, and Al by fast neutron reactions, $^{16}O(n, p)^{16}N$, $^{28}Si(n, p)^{28}Al$, $^{24}Mg(n, p)^{24m}Na$, and $^{27}Al(n, \alpha)^{24m}Na$. Joshi and Agrawal[16] determined Pb and Na in soil samples by using fast neutron reactions $^{208}Pb(n, 2n)^{207m}Pb$ and $^{23}Na(n, \alpha)^{20}F$ under the condition of $t_i = t_c = 2.5\,s$, $t_d = t_w = 0.4\,s$ and 14 cycles. Their results indicated that besides Pb and Na, other eight elements, such as O, Si, Mg, As, Ce, Hf, Nd, and In, could also be determined by this method. Chatt et al.[9] determined Hf in Zirconium wire and Pb in brass by thermal neutron CAA, the detection limits reported are 0.93 μg g^{-1} and 3.6%, respectively. Parry[35,36] analyzed some geological materials for Rh,

Ag by thermal neutron CAA and Au by fast neutron CAA; she also analyzed F in industrial samples by used fast neutron CAA.[55] Owrang et al.[56] reported a fast neutron CAA method for the in situ determination of oxygen in oil by using a neutron generator, a few milligrams of oxygen in oil sample has been determined, using this technique, the concentration of oxygen in metal oxides could also be measured.

Uranium is usually determined by measuring ^{239}Np produced by ^{238}U(n, γ)^{239}U(β^-)^{239}Np reaction. However, by measuring delay neutron after thermal neutron irradiation and cyclic activation, the detection limit can be significantly improved. In addition to geological sample, it can also analyze super-pure material, such as single-crystal silicon for uranium in nanogram per gram level. Song[57] used this technique to analyze a variety of geological and industrial samples for uranium; a detection limit of 0.1 ng g^{-1} was reported.

5 CONCLUSION AND PERSPECTIVE

Since it was first proposed in the beginning of 1960s and rapidly developed in 1970s and 1980s, CAA has become a useful and mature technique for elemental analysis. Although some new CAA facilities were installed, in recent years, only some limited improvement on methodology was made. However, with the increase in CAA facilities, this technique is being applied in more wide fields, and with the increasing interest in using short-lived nuclides for activation analysis, CAA is attracting more attention. However, CAA, as one of the activation analytical techniques, can only be a supplement and not instead of the conventional activation analysis.

Comparing with conventional activation analytical techniques, CAA offers several advantages: (i) significant improvement of detection limit, analytical precision and accuracy for elements determined by using short-lived nuclides; (ii) short experimental time, hence reduced analytical expense and increased analytical capacity; (iii) capability of determining the degree of homogeneity of a sample by analyzing several portions of it for single cycles (cumulative activation); (iv) simultaneous determination of short- and medium-lived nuclides; and (v) estimating or confirmation of half-life of short-lived nuclides. However, CAA also has some disadvantages, e.g. (i) limited number of elements analyzed, i.e. only <50% of elements analyzed by conventional NAA can be determined by CNAA; (ii) some special equipment or changes of normal equipment used in conventional activation analysis are needed; (iii) dead time and pileup correction must be made; (iv) interference from impurity elements in shuttle rabbit material cannot be avoided and a special compressed gas (such as N_2, not air) should be used to transfer sample for reduced interference from ^{41}Ar and other nuclides in activated air.

In comparing with nonnuclear analytical techniques, such as ICP–MS, owing to rapidity, nondestructive analysis, and in situ analysis, CAA as one of the activation analysis techniques is more competitive. However, with the improvement of ICP–MS technique (especially nondestructive analysis by laser ablation inductively coupled plasma mass spectrometry (LA–ICP–MS)) and its popularity, the application of CAA (including all activation analysis techniques) is becoming less.

ABBREVIATIONS AND ACRONYMS

σ	Activation cross section
C	Actual acquired net counts in photopeak of interest
f	Correction factor for dead time in one cycle
F_n	Correction factor for dead time of n cycles
F_p	Correction factor for pileup
t_c	Counting time
S/m	Counts per unit mass element of interest
T	Cycle period ($T = t_i + t_d + t_c + t_w$)
CAA	Cyclic activation analysis
CNAA	Cyclic neutron activation analysis
λ	Decay constant ($\lambda = \ln 2/t_{1/2}$)
t_d	Delay time (the time interval between the end of irradiation and the start of counting)
D_1	Detector response (or the number of counts)
D_n	Detector response in nth cycle
ε	Efficiency of the detector
FNAA	Fast neutron activation analysis
DT(t)	Fractional analyzer dead time at time t
$t_{1/2}$	Half-life of the nuclide of interest
ICP–MS	Inductively coupled plasma mass spectrometry
I	Intensity of the radiation of interest
LA–ICP–MS	Laser ablation inductively coupled plasma mass spectrometry
NAA	Neutron activation analysis
n	Number of cycles
N	Number of target nuclei
N^*	Number of radioactive nuclei
PCAA	Pseudocyclic activation analysis
L_d	Qualitative detection limit

Φ	Radiation (neutron) flux intensity
δ	Relative standard deviation
I_0	Resonance integral cross section
σ_0	Thermal neutron activation cross section
P	The actual photopeak count rates including pileup
P_0	The true photopeak count rates without coincidence losses
T_c	The live time of the measurement
${}_bD_c$	The cumulative detector response for the nuclide making the major background contribution to the signal from the nuclide of interest
t_i	Time of irradiation
D_c ($_sD_c$)	Total cumulative detector response in all n cycles
T_t	Total experiment time
A_0	True initial photopeak count rate
t_w	Waiting time (the time interval between the end of counting and the start of irradiation)

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides

Mass Spectrometry **(Volume 13)**
Inorganic Substances, Mass Spectrometric in the Analysis of

REFERENCES

1. O.U. Anders, 'Determination of Fluorine by Neutron Activation', *Anal. Chem.*, **32**(10), 1368–1369 (1960).
2. O.U. Anders, 'Use of Very Short-lived Isotopes in Activation Analysis', *Anal. Chem.*, **33**(12), 1706–1709 (1961).
3. R.L. Caldwell, W.R. Mills Jr, L.S. Allen, P.R. Bell, R.L. Heath, 'Combination Neutron Experiment for Remote Analysis', *Science*, **152**, 457–465 (1966).
4. W.W. Givens, W.R. Mills, R.L. Caldwell, 'Cyclic Activation Analysis', *Proceedings of International Conference on Modern Trends in Activation Analysis, 1968*, National Bureau of Standard, Gaithersburg, Maryland, 139–147, 1969.
5. W.W. Givens, W.R. Mills, R.L. Caldwell, 'Cyclic Activation Analysis', *Nucl. Instrum. Methods*, **80**, 95–103 (1970).
6. A. Tani, Y. Matsuda, Y. Yuasa, N. Kawai, 'Repeat Pulse Activation Analysis for Measurement of Short Half Life Radionuclides', *Radiochem. Radioanal. Lett.*, **1**, 155–161 (1969).
7. F. Ozek, *'Cyclic Activation and its Applications'*, University of Surrey, M.Sc. Thesis, 1973.
8. N.M. Spyrou, K. Ingle, F. Ozek, 'Cyclic Activation in Trace Element Analysis of Environmental Samples', *Proceedings of 2nd International Conference on Nuclear Methods in Environmental Research*, Columbia, Missouri, 106–120, 1974.
9. A. Chatt, K.N. DeSilva, J. Holzbecher, D.C. Stuart, R.E. Tout, D.E. Ryan, 'Cyclic Neutron Activation Analysis of Biological and Metallurgical Samples', *Can. J. Chem.*, **59**, 1660–1664 (1981).
10. G.D. Burholt, E.A.Y. Caesar, T.C. Jones, 'The Fast Cyclic Activation System for Neutron Activation Analysis in the University of London Reactor', *Nucl. Inst. Method*, **204**, 231–134 (1982).
11. V.L. Spate, J.S. Morris, C.K. Baskett, M.M. Mason, C.L. Reams, D. Hunter, W.C. Willett, 'Determination of Fluorine in Human Nails Via Cyclic Instrumental Neutron Activation Analysis', *J. Radioanal. Nucl. Chem., Articles*, **179**(1), 27–33 (1994).
12. S.S. Ismail, F. Grass, G.P. Westphal, 'A New Efficient Set-up for Cyclic and Pseudocyclic Short Time Activation Analysis With High Count Rates', *J. Radioanal. Nucl. Chem.*, **233**(1–2), 149–153 (1998).
13. X.L. Hou, H.A. Das, 'Accuracy, Precision and Sensitivity in Cyclic INAA with Short-lived Radionuclides', *J. Radioanal. Nucl. Chem.*, **223**(1–2), 67–72 (1997).
14. P. Ila, 'Determination of Total Oxygen in Standard Rocks by Cyclic Activation Analysis Using 14 MeV Neutron Generator', *J. Radioanal. Nucl. Chem., Articles*, **148**(2), 415–422 (1991).
15. S.E.B. Nonie, K. Randle, A.G. Blackband, 'Development of a computer-based system for Fast Neutron Cyclic Activation Analysis', *J. Radioanal. Nucl. Chem., Articles*, **167**(1), 89–95 (1993).
16. G.C. Joshi, H.M. Agrawal, 'Multielemental Analysis of Soil Samples by Fast Neutron Activation', *J. Radioanal. Nucl. Chem., Articles*, **198**(2), 475–485 (1995).
17. X.L. Hou, 'Cyclic Activation Analysis', in *Encyclopedia of Analytical Chemistry*, ed. R.A. Meyers, John Wiley & Sons Ltd, Chichester, 12447–12459, 2002.
18. N.M. Spyrou, 'Cyclic Activation Analysis, A Review', *J. Radioanal. Chem.*, **61**(1–2), 211–242 (1981).

19. A. Chattopadhyay, K.N. DeSilva, 'Pseudo-Cyclic Neutron Activation analysis of Ag, F, Rb, Sc and Se in Biological Samples', *Trans. Am. Nucl. Soc.*, **32**, 185–186 (1979).

20. K.N. DeSilva, A. Chatt, 'A Method to Improve Precision and Detection Limits for Measuring Trace Elements Through Short-lived Nuclides', *J. Trace Microprobe Technol.*, **1**(3), 307–337 (1982–1983).

21. H.U. Fanger, R. Pepelnik, W. Michaelis, 'Fast Neutron Activation Analysis with Short-lived Nuclides', *J. Radioanal. Chem.*, **61**, 147–163 (1981).

22. A. Egan, N.M. Spyrou, 'Determination of Lead Via Lead-207m Using Cyclic Activation and A Modified Sum-coincidence System', *Anal. Chem.*, **48**(13), 1959–1962 (1976).

23. N.M. Spyrou, S.A. Kerr, 'Cyclic Activation: the Measurement of Short-lived Isotope in the Analysis of Biological and Environmental Samples', *J. Radioanal. Chem.*, **48**, 169–183 (1979).

24. M. Colin, C. Friedli, P. Lerch, 'Heavy Ions Activation Analysis: Theoretical Approach for Optimizing Cyclic Analysis', *Nucl. Instrum. Methods, Phys. Res.*, **B10**, 1062–1064 (1985).

25. R.E. Tout, A. Chatt, 'The Effect of Sample Matrix on Selection of Optimum Timing Parameters in Cyclic Neutron Activation Analysis', *Anal. Chim. Acta*, **133**, 409–419 (1981).

26. M.A. Al-Mugrabi, N.M. Spyrou, 'The use of Simulation for the Optimization of the Signal-to-noise Ratio in Cyclic Activation Analysis', *J. Radioanal. Nucl. Chem.*, **110**(1), 67–77 (1987).

27. N.M. Spyrou, C. Adesanmi, M. Kidd, L.G. Stephens-Newsham, A.Z. Ortaoval, F. Ozek, 'Usefulness of Thermal and Epithermal Cyclic Activation Analysis with Reactor System', *J. Radioanal. Chem.*, **72**(1–2), 155–182 (1982).

28. A.Z. Ortaovali, F. Ozek, I. Celenk, N.M. Spyrou, 'Study on Some Problems and Application of Cyclic Activation Analysis', *J. Radioanal. Chem.*, **61**(1–2), 175–182 (1981).

29. E. Schonfeld, 'Alpha? A Computer Program for the Determination of Radioisotopes by Least Squares Resolution of the Gamma-ray Spectra', *Nucl. Instrum. Methods*, **42**, 213–218 (1966).

30. A. Egan, S.A. Kerr, M.J. Minski, 'Determination of Selenium in Biological Materials Using ^{77m}Se and Cyclic Activation Analysis', *Radiochem. Radioanal. Lett.*, **28**(5–6), 369–378 (1977).

31. A. Wyttenbach, 'Coincidence Losses in Activation Analysis', *J. Radioanal. Chem.*, **8**, 335–343 (1971).

32. Z.B. Alfassi, A. Tsechansky, A.P. Kushelevsky, 'On the Correction of the Self Coincidence in Gamma Ray Spectra Dependence of Resolution Time on Energy', *J. Radioanal. Chem.*, **55**, 135–139 (1980).

33. K. Heydron, E. Damsgaard, 'Validation of a Loss-free Counting System for Neutron Activation Analysis with Short-lived Indicators', *J. Radioanal. Nucl. Chem.*, **215**(2), 157–160 (1997).

34. V.P. Guinn, 'Cyclic Nuclear Activation Analysis', *Radiochem. Radioanal., Lett.*, **44**, 133–140 (1980).

35. S.J. Parry, 'Cumulative Neutron Activation Analysis for the Improved Detection of Short-lived Nuclides', *J. Radioanal. Chem.*, **75**(1–2), 253–258 (1982).

36. S.J. Parry, 'Cumulative Neutron Activation Analysis of Short-lived Radionuclides for the Analysis of Large Samples in Mineral Exploration', *J. Radioanal. Nucl. Chem.*, **112**(2), 383–386 (1987).

37. A.S. Farooqi, W. Arshed, O.A. Akanle, C. Jeynes, N.M. Spyrou, 'Fluorine Determination in Diet Samples Using Cyclic INAA and PIGE Analysis', *J. Radioanal. Nucl. Chem.*, **161**(1), 71–78 (1992).

38. S.E. Nonie, K. Randle, 'The Multielement Potential of Fast Neutron Cyclic Activation Analysis', *J. Radioanal. Nucl. Chem.*, **185**(1), 35–44 (1994).

39. G.C. Joshi, H.M. Agrawal, 'Multielemental Analysis of Soil Samples by Fast Neutron Activation', *J. Radioanal. Nucl. Chem.*, **198**(2), 475–484 (1995).

40. I. Khamis, I. Othman, M. Nasri, M. Bakkour, 'Transfer Time Optimization of a Rapid Cyclic Instrumental Neutron Activation Analysis for Trace Element Detection', *Review of Scientific Instruments*, **77**(2), 1492–1494 (2001).

41. C. Yonezawa, T. Kurosawa, H. Matsue, 'Rapid Determination of Trace Amounts of Fluorine in Silicon Nitride by Instrumental Neutron Activation Analysis', *Bunseki Kagaku*, **47**(1), 829–834 (1998).

42. I.P. Matthews, N.M. Spyrou, 'Multielement Analysis of Bulk Matrices by Measurement of Prompt and Delay Gamma-ray as Well as Cyclic Activation Using An Isotopic Neutron Source', *Int. J. Appl. Radiat. Isot.*, **33**(1), 61–68 (1982).

43. S.J. Ryde, W.D. Morgan, A. Sivyer, C.J. Evans, J. Dutton, 'A Clinical Instrument for Multielement in vivo Analysis by Prompt, Delayed and Cyclic Neutron Activation Using ^{252}Cf'', *Phys. Med. Biol.*, **32**(10), 1251–1271 (1987).

44. K. Mardani, 'Neutron Interrogation of Fissile Material Using A Remote Detection System', University of Surrey, M.Sc. Thesis, 1978.

45. IEAE, 'Practical Aspects of Operation, A Neutron Activation Analysis Laboratory', IAEA-TECDO-564, Vennia, 1991.

46. J.R. Parrington, H.D. Knox, S.L. Breneman, E.M. Baum, F. Feiner, *'Chart of the Nuclides'*, 15th edition, General Electric Co. and KAPL Inc., 1996.

47. Z.B. Alfassi, *Activation Analysis*, Vol. II, CRC Press, Boca Raton, FL, 1990.

48. Y. Shi, E.E. Sullivan, J. Holzbecher, A. Chatt, 'Determination of Selenium in Canadian Food Items by Cyclic

Instrumental Neutron Activation Analysis', *Biol. Trace Elem. Res.*, **71–72**, 377–386 (1999).

49. A. Chatt, H.S. Dang, B.B. Fong, C.K. Jayawickreme, L.S. MaDowell, D.L. Pegg, 'Determination of Trace Elements in Food by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **124**(1), 65–77 (1988).
50. M. Navarrete, L. Cabrera, T. Martinez, G. Revel, J.P. Meyer, J. Guarner, 'Relationship Between the Cervical Uterine Cancer Evolution and Selenium Concentration in Urine Determined by NAA', *J. Radioanal. Nucl. Chem.*, **217**(2), 167–169 (1997).
51. T.P. Cheng, J.S. Morris, S.R. Koirtyohann, V.L. Spate, C.K. Baskett, 'The Analysis of Human Nail for 24 Elements via k0 and Cyclic Neutron Activation Analysis', *Nucl. Instrum. Methods*, **A353**, 457–460 (1994).
52. S.A. Kerr, N.M. Spyrou, 'Fluorine Analysis of Bone and Other Biological Materials: A Cyclic Activation Method', *J. Radioanal. Chem.*, **44**, 159–173 (1978).
53. X.L. Hou, W.J. Jiang, *Study of Metabolism of Oxygen in Mice by ^{18}O Tracer and Cyclic Neutron Activation Analysis*, Annual Report of China Institute of Atomic Energy, Chinese Atomic Energy Press, Beijing, 1993.
54. M. Wiemick, S. Amiel, 'Use of Very Short-lived Nuclide in Non-destructive Activation Analysis with A Fast Shuttle Rabbit, II Determination of Lead by Means of ^{207m}Pb', *J. Radioanal. Chem.*, **3**, 393–403 (1969).
55. S.J. Parry, R. Benzing, K.L. Bolstad, E. Steinnes, 'Epithermal/ Fast Neutron Cyclic Activation Analysis for the Determination of Fluorine in Environmental and Industrial Samples', *J. Radioanal. Nucl. Chem.*, **244**(14), 67–72 (2000).
56. F. Owrang, H. Mattsson, A. Nordlund, 'A Method for In-situ Quantification of Oxygen in Oil Using fast Neutron Activation Analysis', *Kerntechnik*, **69**(1–2), 51–57 (2004).
57. Q.X. Song, 'Improvement in the Determination of Trace Uranium by Repeating Activation and Delayed Neutron Counting, Determination of The Uranium Content of Multi-crystalline and Single-Crystal Silicon Samples', *Chinese J. Nucl. Techn.*, **10**(6), 35–46 (1987).

Elastic Recoil Detection Analysis

Patrick Trocellier
Commissariat à l'Énergie Atomique – Service de Recherches de Métallurgie Physique, Centre d'Études Nucléaires de Saclay, 91191 Gif sur Yvette, France

Timo Sajavaara
Department of Physics, University of Jyväskylä, Finland

1 Introduction **2**
2 Scattering Process and Conversion of Energy Spectrum to Depth Profile in Elastic Recoil Detection Analysis **2**
2.1 Kinematics 2
2.2 Cross Section 3
2.3 Energy–Depth Relationships 4
2.4 Recoil Yield 5
2.5 Depth Resolution 5
3 Conventional Elastic Recoil Spectrometry **5**
3.1 Glancing (Reflection) Geometry 5
3.2 Transmission Geometry 5
3.3 Data Processing 6
3.4 Standard Matrices and Ion-beam-induced Damage 6
3.5 Microbeam-induced Elastic Recoil Detection Analysis 7
4 Variant Recoil Methods **7**
4.1 $\Delta E-E$ Telescope 7
4.2 Time-of-flight Elastic Recoil Detection Analysis 8
4.3 Coincidence Spectrometry 9
4.4 Electromagnetic Cross Filters ($E \times B$) 10
5 High-energy and Heavy-ion Elastic Recoil Spectrometry **11**
5.1 High-energy ^{4}He-induced Elastic Recoil 11
5.2 High-energy Heavy-ion-induced Elastic Recoil 11
6 Application Examples of Elastic Recoil Detection Analysis for Profiling Hydrogen and other Elements **15**
6.1 Polymer Sciences 15
6.2 Semiconductor Materials 16
6.3 Thin Films 17
6.4 Interface Reactivity 17
6.5 Other Application Examples 18
7 Conclusion **18**
Acknowledgments **19**
Abbreviations and Acronyms **19**
Related Articles **19**
References **19**
Further Reading **24**

In 1976, a Canadian group described in detail for the first time a new ion beam analytical method based on the elastic recoil of target nuclei collided with high-energy heavy incident ions. In this case, 25–40-MeV ^{35}Cl impinged on a multilayer C or Cu (backing)/LiF or LiOH/Cu (30–150 nm)/LiF or LiOH and H, Li, O, and F recoiled atoms were detected. These exemplified the main characteristics of elastic recoil detection analysis (ERDA): its sensitivity to depth distribution and its ability to detect light elements in heavy substrates. In 1979, the use of megaelectronvolt energy ^{4}He beams permitted the use of ERDA to be extended to depth profiling of hydrogen isotopes in the near-surface region of solids.

ERDA has rapidly been revealed to be an excellent alternative to resonant nuclear reaction spectrometry (see **Nuclear Reaction Analysis***) for hydrogen determination in solids. Despite its less advantageous performance with respect to its lower depth resolution, lower analyzable depth, comparable sensitivity, and more restricting irradiation and detection geometry, some ERDA features have made its development in ion beam analysis (IBA) laboratories worldwide easier; these are simultaneous access to ^{1}H and ^{2}H depth distributions, access to single-ended Van de Graaff accelerators compared with tandem accelerators or cyclotrons, and the ability to be combined with Rutherford backscattering spectrometry (RBS) (see* **Rutherford Backscattering Spectroscopy***).*

The development of detection devices and the analytical capabilities offered by high-energy heavy-ion-induced ERDA in material sciences for profiling light, medium, and high mass number elements give this method a wide area in which to progress. The main advantage of heavy-ion ERDA and quite unique feature among analysis techniques is the fact that all sample elements can be depth profiled in one measurement by single detector telescope. By means of Monte Carlo (MC) simulations, the interpretation and reliability of the results have increased greatly over the last few years.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

1 INTRODUCTION

In 1976, L'Ecuyer et al.,[1] from the INRS-Energie Varennes (Québec), presented a paper that described in detail for the first time a new ion beam analytical method based on the elastic recoil of target nuclei collided with high-energy heavy incident ions, in this case, 25–40-MeV ^{35}Cl impinging on a multilayer C or Cu (backing)/LiF or LiOH/Cu (30–150 nm)/LiF or LiOH.[1] The work exhibited two of the main characteristics that established the technique of ERDA: sensitivity to depth distribution and the ability to detect light elements in heavy substrates.

The need to use high-energy heavy ions was a limitation to the development of this new analytical tool. In 1979, Doyle and Peercy published the first real example of the use of ERDA to determine hydrogen isotope depth profiles in the near-surface region of solids.[2]

At the same time, resonant nuclear reactions induced by energetic heavy ions were strongly promoted for the quantitative measurement of ^{1}H. ^{7}Li, ^{11}B, ^{15}N, and ^{19}F ion beams between 3 and 20 MeV were successfully shown to be very efficient for hydrogen depth profiling.[3–7]

ERDA has been revealed to be an excellent alternative to resonant nuclear reaction spectrometry for hydrogen determination in solids. Despite its less advantageous performance with respect to lower analyzable depth, comparable sensitivity, and more restricting irradiation and detection geometry, some ERDA features have made its use and development in IBA laboratories worldwide fast; these are simultaneous detection of not only ^{1}H and ^{2}H but also other elements, rapid development of detectors, and the great availability of small single-ended and tandem accelerators. As a result of detector development, especially in the case of TOF- and gas-ionization detectors, energies even below 10 MeV can now be used for high-resolution depth profiling.

In October 1996, Tirira, Serruys, and Trocellier, a group of French authors from Saclay published the first ion beam monograph entirely dedicated to ERDA.[8] This book has been used as the basis for the technical part of the present article and also as a bibliographical source.

In the following, we establish the theoretical basis of elastic recoil phenomenon, describe the experimental configurations generally adopted for elastic recoil spectrometry, and illustrate the analytical capabilities of ERDA for quantitative measurements in solids. Section 2 is devoted to general considerations of elastic collisions induced by megaelectronvolt ^{4}He and heavier incident ions in a hydrogenated material. Section 3 presents the practical characteristics of elastic recoil spectrometry in terms of irradiation and detection geometry, analytical performance, and data processing. Section 4 gives details of the main methodological variations from conventional ERDA using a single surface barrier detector and more recent experimental improvements. Section 5 is devoted to the use of high-energy ^{4}He ions and heavy ions for carrying out light-element profiling in solids. Section 6 contains some application examples of ERDA in a wide variety of fields from polymer science to semiconductor or thin-film technology.[9–11]

The concluding section of this article attempts to focus on the relevance of ERDA in comparison with resonant nuclear reaction spectrometry and secondary ion mass spectrometry (SIMS) (*see* **Secondary Ion Mass Spectrometry as Related to Surface Analysis**) for hydrogen determination in the near-surface region of solids. It also indicates the major trends in ERDA development since 1997.

2 SCATTERING PROCESS AND CONVERSION OF ENERGY SPECTRUM TO DEPTH PROFILE IN ELASTIC RECOIL DETECTION ANALYSIS

2.1 Kinematics

When an incident ion (mass m) with kinetic energy E_0 (a momentum P_0) interacts with a target atom (mass M) at rest through a purely elastic collision, the application of the conservation principles for total energy and momentum allows us to write Equations (1) and (2) (Figure 1):

$$E_0 = E + E' \tag{1}$$

$$P_0 = p + p' \tag{2}$$

with E, p, E', and p' the respective kinetic energies and momenta of the scattered incident ion and the recoiled target atom. Introducing θ and ϕ, the respective laboratory angles of scattering for m and M (Figure 1), we obtain Equations (3) and (4) given by

$$E' = K' E_0 \tag{3}$$

where

$$K' = \frac{4 m M \cos^2 \phi}{(m + M)^2} \tag{4}$$

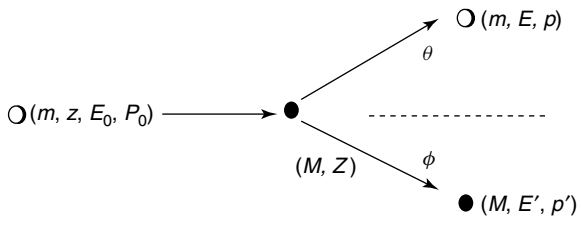

Figure 1 Schematic view of the elastic collision m/M.

and the angular relationship, Equation (5):

$$\tan\theta = \frac{\sin(2\phi)}{(m/M) - \cos(2\theta)} \tag{5}$$

If $m > M$, as for example when ^{1}H is the target atom and $^4He^+$ is the incident ion, the incident ion cannot be scattered from its incident direction through an angle greater than θ_{max}. This value is directly derived from the above equations, Equation (6):

$$\sin\theta_{max} = \frac{M}{m} \tag{6}$$

For a ^{4}He/^{1}H elastic collision, $\sin\theta_{max} \approx 0.26$ and $\theta_{max} = 14.47^\circ$. For a usual recoil angle $\phi = 30^\circ$, $K' = 0.48$, thus 2-MeV $^4He^+$ gives protons recoiling from the target surface an energy of 960 keV.

The kinematics of the elastic collision is described in detail in two IBA handbooks by Chu[12] and Tirira.[8]

2.2 Cross Section

2.2.1 Rutherford Scattering

The classical Rutherford formalism developed for a purely coulombian scattering leads to the well-known basic formula, Equation (7):

$$\left(\frac{d\sigma}{d\Omega}\right)^{cm} = \left[\frac{zZe^2}{4E_0^{cm}\sin^2(\xi/2)}\right]^2 \tag{7}$$

where $(d\sigma/d\Omega)^{cm}$ is the differential scattering cross section expressed in the center of mass system, z and Z are the respective atomic numbers of the incident ion and the target atom, e the elementary charge, E_0^{cm} the kinetic energy of the incident ion in the center of mass, and ξ the scattering angle in the center of mass.

In the laboratory system, Equation (7) becomes Equation (8):

$$\delta = \left(\frac{d\sigma}{d\Omega}\right)^{L} = \left(\frac{zZe^2}{4E_0}\right)^2 \left[\frac{m}{M} + 1\right]^2 \frac{1}{\cos^3\phi} \tag{8}$$

It is obvious that

- δ exhibits a minimum value for $\phi \approx 0$;
- δ tends to infinity when ϕ tends to $\pi/2$; in this case, the projectile/target interaction cannot be described by the Rutherford model;
- δ rapidly decreases as E_0 increases; and
- when $M \gg m$, δ is independent of the mass ratio.

In ERDA, Rutherford behavior is mainly observed for heavy-ion beams colliding with light target atoms (see Section 5) and for a ^{4}He ion beam below 1 MeV.

2.2.2 Non-Rutherford Cross Section

The Rutherford formula cannot be applied for elastic scattering/recoil spectrometry when the interaction potential between the incident ion and the target atom deviates from a pure coulomb potential. This can happen in two different cases. First, if the energy is high enough and secondly, in the low-energy case, if the two nuclei are screened by inner-shell electrons. The high energy at which elastic scattering/recoil cross sections begin to deviate from classic Rutherford behavior is called the *threshold energy*. It corresponds to the limit above which nuclear forces affect the interaction potential. Several authors have reviewed theoretical models combining a pure coulomb potential with a weak-perturbing nuclear contribution, as for example a Yukawa-like potential.[13] Experimental measurements of the threshold energy have been made for several projectile–target nucleus couples[14] and analytical formulations have been extracted.[15]

Using R_0, the distance of closest approach as a free parameter, the threshold energy is expressed by Equations (9) and (10):

$$E_{th} = \left(\frac{zZe^2}{2R_0}\right)\left(1 + \frac{m}{M}\right)\left(1 + \frac{1}{\cos\phi}\right) \tag{9}$$

where, for example,

$$R_0 = C_1 + C_2(m^{1/3} + M^{1/3}) \tag{10}$$

$C_1 = 6.003 \times 10^{-1}$m, $C_2 = 0.864 \times 10^{-15}$m, and $5.2 < (m^{1/3} + M^{1/3}) < 8.6$ (after Räisänen).[15] C_1 and C_2 are constants and m and M are the mass numbers. Applying this formalism, the threshold energy is 57.37 MeV for the couple ^{35}Cl/^{12}C and 469.60 MeV for the couple ^{127}I/^{16}O.[15]

In the case of heavy incident ions, the smaller cross sections due to the screening effect in low-energy scattering needs to be taken into account in the analysis. According to the analytical formula derived by Andersen et al.[16] the deviation from the pure Rutherford cross section is 21.7, 10.4, and 1.5% for Cu recoils detected at 40° and ejected by 10 MeV ^{197}Au, ^{127}I, and ^{35}Cl incident ions, respectively.

2.2.3 ^{4}He/^{1}H Collision

Historically, the scattering of protons by ^{4}He target nuclei was first investigated in the range 1–20 MeV for E_p. A phase shift analysis derived from quantum mechanics was thus applied to evaluate the cross section.[17,18]

The reverse collision can be treated by the same approach in the range 1–4 MeV for ^{4}He. The decomposition of wavefunctions in partial waves can be restricted

Table 1 Parameters obtained for A_l and Γ_l in the case of a $^4He/^1H$ elastic collision up to $E_{He} = 4\,MeV$

Partial wave	Scattering length A_l (fm)	Effective range Γ_l (fm)
$S_{1/2}$	3.35 ± 0.01	6.45 ± 0.01
$P_{1/2}$	-17.8 ± 0.8	-0.43 ± 0.05
$P_{3/2}$	-37.6 ± 1.5	-0.62 ± 0.03

to $S_{1/2}$, $P_{1/2}$, and $P_{3/2}$ contributions to describe the scattering of ^4He by ^1H. This method consists in determining nuclear phase shifts by using physical models such as the optical model or the *R*-matrix theory.[19–23] Tirira and Bodart have used experimental data below 4 MeV to extend previous phase shift studies.[24]

The best set of theoretical parameters (scattering length and effective range), derived by Tirira and Bodart[25] is presented in Table 1.

For ^4He energy values below 3 MeV and ϕ values below 30°, i.e. the most frequently used experimental conditions, several points with respect to the functional dependence of the non-Rutherford cross section $\delta_{nR} = (d\sigma/d\Omega)^L$ should be noted:

- δ_{nR} exhibits a minimum value for an energy E_{min} in the range 2.4–3 MeV for a recoil angle less than 30° ($E_{min} = 2.41\,MeV$ at 0°). As the angle increases from 0°, the cross-section minimum moves toward larger energy values.
- δ_{nR} has its lowest value at $\phi = 0^\circ$ for $E_{He} < E_{min}$.
- δ_{nR} has its highest value at 0° for $E_{He} > E_{min}$.
- δ_{nR} appears to reach the same constant value around $E_{He} = 3\,MeV$ for $0 < \phi < 40^\circ$.
- δ_{nR} varies very slightly near 0° with respect to E_{He}.
- δ_{nR} varies slightly in the $0–30^\circ$ range between 2 and 3 MeV.
- δ_{nR} shows a quasi-Rutherford behavior for $E_{He} <$ 1 MeV.

Comparing previously published data sets and their own experimental results, Tirira and Bodart derived a simple analytical expression allowing the elastic recoil differential cross section in the laboratory to be calculated for any $E_{He} - \phi$ couple, Equation (11)[25]:

$$\ln\left(\frac{d\sigma}{d\Omega}\right)^L = A_1 E_{He} + A_2 + A_3 E_{He}^{-1} + A_4 E_{He}^{-2} \qquad (11)$$

Table 2 contains different data sets for A_i values, for $0 \le \phi \le 40^\circ$ and $1 \le E_{He} \le 4\,MeV$.

Other binary collisions such as $^4He/^2H$, $^4He/^3H$, and $^3He/^1H$ have been treated in the same way.[26–30] A detailed discussion on this subject and the corresponding analytical expressions are given by Tirira et al.[8]

Table 2 Practical data sets for a rapid elastic recoil differential cross-section calculation[25]

ϕ ($^\circ$)	A_1	A_2	A_3	A_4
0	0.7651	1.7201	5.6116	−1.7011
5	0.7581	1.7321	5.6302	−1.7148
10	0.7366	1.7716	5.6797	−1.7527
15	0.6994	1.8492	5.7417	−1.8049
20	0.6449	1.9807	5.7890	−1.8568
25	0.5732	2.1840	5.7888	−1.8906
30	0.4779	2.4758	5.7117	−1.8897
35	0.3651	2.8682	5.5349	−1.8408
40	0.2349	3.3687	5.2445	−1.7350

From ref 25. © Elsevier 2006.

2.3 Energy–Depth Relationships

An incident ion with energy E_0 impinges on a target surface with an angle of incidence α with respect to the normal to the surface. Let us consider an elastic collision occurring at depth x beneath the target surface, the scattering direction makes an angle θ with the incident direction and the recoil angle is ϕ. The path of the incoming ion before collision is simply expressed by Equation (12):

$$L_{in} = \frac{x}{\cos\alpha} \qquad (12)$$

Its outgoing path is given by Equation (13):

$$L_{out} = \frac{x}{\cos\alpha} \qquad (13)$$

where the outgoing angle has been chosen to be equal to the incoming one. The path of the recoiled target atom is then given by Equation (14):

$$L_r = \frac{x}{\cos\beta} \qquad (14)$$

where $\beta = \pi - \phi - \alpha$.

The energy of the incident ion just before the collision is given by Equation (15):

$$E_0(x) = E_0 - \int_0^{x/\cos\alpha} S(E)\,dl \qquad (15)$$

where dl is the differential path. The energy of the scattered incident ion leaving the target is given by Equation (16):

$$E_1(x) = K E_0(x) - \int_0^{x/\cos\alpha} S(E)\,dl \qquad (16)$$

and the energy of the recoiled target atom is given by Equation (17):

$$E_2(x) = K' E_0(x) - \int_0^{x/\cos\beta} S_r(E)\,dl \qquad (17)$$

where $S(E)$ and $S_r(E)$ are the stopping powers for the incident ion and the recoiled target atom, respectively, in the target material considered.

2.4 Recoil Yield

For a thin target, the number of recoiled target atoms emitted in the depth range $x - x + \mathrm{d}x$ through an angle ϕ in the solid angle $\Delta\Omega$ is given by Equation (18):

$$\frac{\mathrm{d}N(x)}{\mathrm{d}x} = \frac{N_0 Q\, \mathrm{d}\sigma(E,\phi)}{\mathrm{d}\Omega\Delta\Omega} \tag{18}$$

where N_0 is the target atomic density and Q the total number of incident ions. The total recoil yield is then obtained by integrating Equation (18) on the thickness t of the target sample, considering the differential cross section as constant along t.

For a thick target, the analytical expression giving the recoil yield is slightly more complex because it has to take into account the variation of the differential cross sections with energy and all the energy spreading effects induced by particles traveling in matter.

2.5 Depth Resolution

Depth resolution δR_x strongly depends on total energy resolution of the system δE_T. This factor includes

- the detector energy resolution δE_d;
- the fluctuations in projectile energy δE_p;
- the energy spreading due to angular deviations δE_θ;
- the geometrical energy spreading δE_g;
- the energy straggling during incoming and outgoing paths δE_s;
- the multiple scattering contribution δE_{ms};
- the contribution from Doppler effect δE_D;
- the energy spreading due to the absorber δE_a; and
- the energy spreading due to surface roughness δE_{rou}.

A detailed description of the methods used for evaluating each contribution has been given by Tirira et al.[8]

3 CONVENTIONAL ELASTIC RECOIL SPECTROMETRY

3.1 Glancing (Reflection) Geometry

Figure 2 summarizes the experimental configuration called *glancing* or *reflection* geometry.

Several points should be noted:

- The beam spot size is enhanced by a factor $(1/\cos\alpha)$ with respect to normal incidence owing to the oblique incidence α.

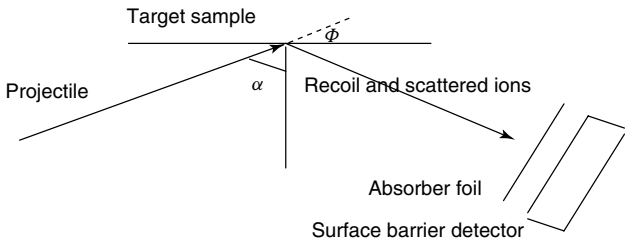

Figure 2 Schematic view of glancing geometry.

- The surface barrier detector is preceded by an absorber foil dedicated to stopping the ions scattered in the direction ϕ.
- The analyzable depth is rather limited owing to the glancing geometry. The range from the surface is obtained by multiplying the true incident ion range in the target by a factor $(\cos\alpha)$.
- This experimental configuration is absolutely prohibited with rough surface targets.

3.2 Transmission Geometry

Figure 3 summarizes the experimental configuration called *transmission* geometry.

Transmission geometry was the first experimental configuration used for ERDA measurements.[1,31–33]

Several points should be noted:

- The thickness of the target and the incident energy of the ^{4}He beam are strongly dependent, typically 20–25 μm for 3–3.5 MeV.[34]
- The presence of an absorber foil in front of the surface barrier detector is not necessary because the target itself slows down the scattered ions while they travel along their path until they are stopped.
- Practically, the surface barrier detector has to be placed as close as possible to the 0° direction, but not directly in view of the incident beam direction to avoid severe irreversible damage in the case of target breaking.
- The analyzable depth is only limited by the range of recoiling protons; it can reach up to 6–7 μm for polymer or silica-based materials.[34,35]

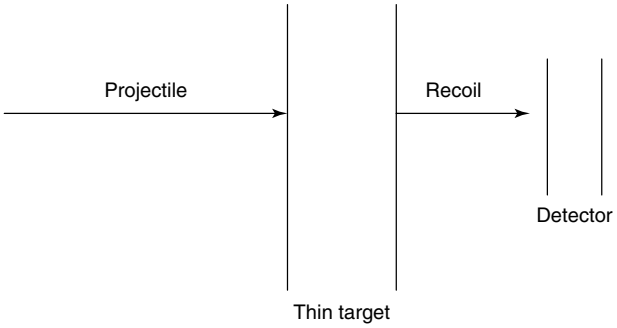

Figure 3 Schematic view of transmission geometry.

Table 3 Performances of ERDA hydrogen measurement

ERDA geometry	Analyzable depth (μm)	Depth resolution (nm)	
		Surface	Deep
Glancing	0.5–1	10	40
Transmission	6–8	30	25

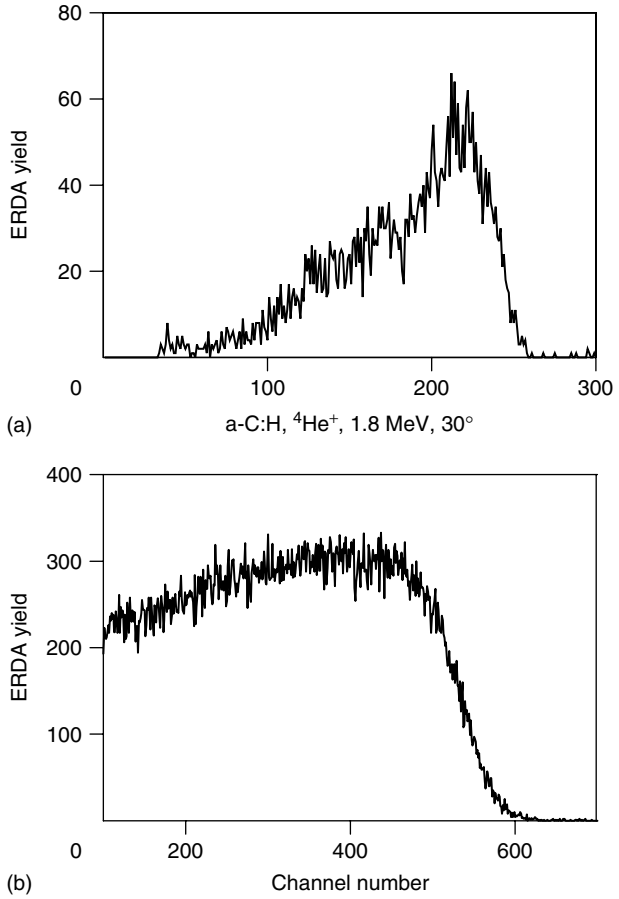

Figure 4 Examples of ERDA spectra obtained with a 1.8 or a 3.5-MeV ^{4}He ion millibeam. (a) Thin hydrogenated NiC/Ti multilayer coated on a silicon substrate in reflection geometry and (b) thick hydrogenated amorphous carbon layer in reflection geometry.

Table 3 permits comparison of the analytical performances offered by the two available geometries for hydrogen measurement.

Figure 4 gives two different examples of ERDA spectra obtained in glancing geometry for a hydrogenated NiC/Ti multilayer and a thick amorphous hydrogenated carbon layer.

3.3 Data Processing

Conventional ERDA data are treated in the same way as RBS, using a reconstitution-comparison-iteration computer program. The first step in the process is to define the experimental configuration (incident energy, geometry, detector characteristics, and target description). Then a simulated spectrum is built using the collision kinematic equation, the energy-loss formalism, the elastic recoil cross-section table, and straggling and multiple scattering subroutines.[36] Finally, the simulated spectrum is compared to the experimental data and adjusted in terms of elemental distributions, using a classical least-squares method until it matches correctly.

Several authors have proposed specific computer codes dedicated to ERDA interpretation.[36–38] Our purpose, in this section, is not to describe in detail any specific approach. The references listed above constitute excellent starting points to get to know classical IBA tools. An IAEA intercomparison[39] of some of the most used computer codes was published recently.

In the case of heavy-ion elastic recoil detection analysis (HI-ERDA), another approach known as *slab analysis* is often used instead of simulation and comparison. In the energy spectrum, a channel corresponds to a certain depth slice. The depth is calculated using scattering kinematics and either experimental or, more often, semiempirical stopping powers. The yield at different depths is then normalized using stopping powers and scattering cross sections for a given incident ion and sample atoms. If all the sample elements can be analyzed, the final depth profiles can be normalized to unity to obtain the atomic ratios. All the effects degrading depth resolution that are mentioned above are still visible in the depth profiles after slab analysis.

3.4 Standard Matrices and Ion-beam-induced Damage

IBA techniques are considered to be quantitative even without the use of reference samples provided that the differential cross section of the interaction observed is accurately known in the energy range investigated and also provided the experimental geometry (target and detector) is fully described.

However, the quantitative determination of any elemental depth distribution in a complex target material generally requires the use of a reference sample. This standard (or model sample) is investigated either to simplify the calculation procedure or to assess the profile extraction method. Moreover, finding a good hydrogen reference sample remains a particular problem owing to its high mobility in solids even under standard temperature conditions. The use of reference samples is highly recommended for accurate results in the case of HI-ERDA also; even in the case of thin films when all the elements present can be detected and concentrations normalized to unity.

Polymers would represent the most practical type of hydrogen reference samples because of their large H

content, their variety of formulation, their thickness range (from 3 μm up to several hundred micrometers) and their well-known core composition homogeneity. Nevertheless, there are two severe limitations:

- the thickness heterogeneity of thin films having thickness less than 20 μm and
- the sensitivity of polymer structure to ion beam bombardment.

Elemental losses are impossible to prevent and in practise polymers have to be replaced by hydrogen-implanted monocrystals (Si or Ge, for example). The versatility of implantation conditions (energy and fluence) is of major interest but conservation of the hydrogen profile for long periods of time is not guaranteed. The hydrogen mobility in silicon at 25 °C is far from negligible. Metal hydrides such as TiH_2 or Ti_xD_y (with $x = 1$ and $y \approx 1.8$) could be an alternative to implanted crystals,[40] but control of the stoichiometry is rather difficult. Hydrogenated amorphous carbon or silicon-carbide layers appear now to be a better compromise than hydride layers.[11]

The choice of a multielemental reference sample requires a complete description of the investigated sample. Routine methods such as electron microprobe analysis (*see* **Scanning Electron Microscopy in Analysis of Surfaces**) must be very efficient to ensure the lateral composition homogeneity of the target material. The most important criterion is to choose a reference material with a similar composition in order to be able to neglect corrections for stopping power. Ceramics, glasses, or thin layers are the most useful materials for IBA standards provided that their sensitivity to ion bombardment has been checked before use.

Several effects can be inferred by an incident ion beam impinging on a radiation-sensitive material: electron excitation, ionization, atomic displacement, bond breaking, formation of charge defect centers, temperature increase, growth of an electrical gradient, and surface sputtering.[38,41] The main consequences of these phenomena is the modification of the target composition and particularly the loss of certain elements of interest such as hydrogen, alkali metals, or halogen elements.[42] An extensive review of beam damage is presented by Tirira et al.[36] A comprehensive bibliography is given and concrete examples discussed.

3.5 Microbeam-induced Elastic Recoil Detection Analysis

Microbeam-induced elastic recoil detection analysis (μERDA) is the only routinely available technique for hydrogen determination using a nuclear microprobe. Experiments in transmission geometry are, in general, easier to carry out than experiments in reflection geometry.

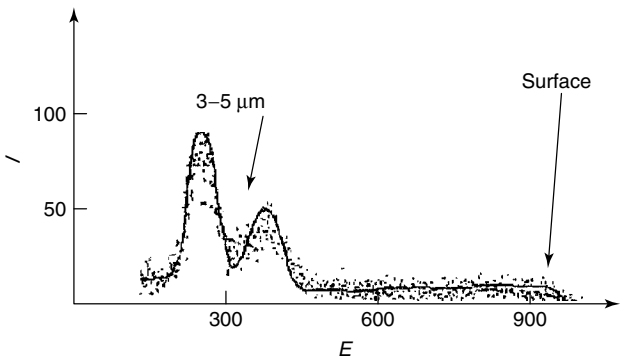

Figure 5 Thin polyimide foil (25 μm) investigated by μERDA in transmission geometry after high-energy heavy-ion irradiation.

Transmission geometry is used in polymer science applications or in geological applications[34,43–45] and reflection geometry is chosen for thin-layer characterization.[11]

Figure 5 gives an example of an ERDA spectrum obtained in transmission geometry on a thin polyimide film (25 μm) preirradiated with 50-MeV ^{32}S ions and measured with a 3.05-MeV $^4He^+$ microbeam (5×5 μm^2, 200 pA). The heavy-ion irradiation induces the formation of hydrogen bubbles in the core of the modified polymer structure. Two of these hydrogen-rich zones are found in the spectrum between 3 and 5 μm from the surface.

4 VARIANT RECOIL METHODS

The energy discrimination between a scattered incident ion and a recoiled target atom is one of the major problems to be solved in ERDA. It can be based on the difference of energy loss in a given thickness of a well-known material. When $m > M$ ($^4He/^1H$ collision), the scattered ^4He loses a higher fraction of its energy than the recoil ^1H reaching the detector. The use of an absorber foil (polymer or Al) in front of the detector allows discrimination by stopping the scattered events. This simple technique is currently the most adopted one for ERDA measurement (see Section 3) in transmission geometry, in which the target itself plays the role of the absorber foil. When $m < M$, this system is completely inefficient.

4.1 ΔE–E Telescope

The second way to solve the scattered/recoil event ambiguity is to replace the absorber foil by a thin detector (10–20 μm). The device constituted by a couple of detectors is called a solid-state $\Delta E - E$ telescope.[46,47]

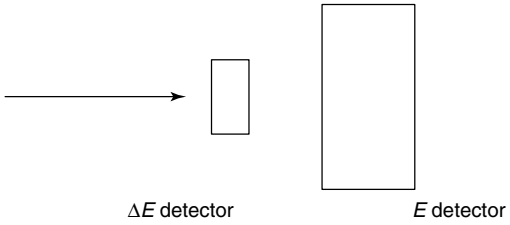

Figure 6 Schematic design of a $\Delta E-E$ detector.

Scattered ions are stopped in the ΔE detector and recoil events are detected in the E detector ($\geq 100\,\mu m$). Figure 6 is a scheme of such a detection device.

The solid ΔE detector can be replaced by a gas-ionization chamber.[47–49] Telescopes have been used for ERDA measurements with ^{4}He beams[46] as well as with heavy-ion beams.[50–53]

The mass and charge separation ability of a telescope are simply described by Equations (19) and (20):

$$\frac{\Delta M}{M} = \left[\left(\frac{\Delta E_R}{E_R}\right)^2 + \left(\frac{\Delta dE}{dE}\right)^2 + \left(\frac{\Delta dx}{dx}\right)^2\right]^{1/2} \quad (19)$$

$$\frac{\Delta Z}{Z} = \frac{1}{2}\left[\left(\frac{\Delta E_R}{E_R}\right)^2 + \left(\frac{\Delta dE}{dE}\right)^2 + \left(\frac{\Delta dx}{dx}\right)^2\right]^{1/2} \quad (20)$$

where dE is the energy loss in the thin detector and dx is the thickness of the thin detector.

The presence of an absorber foil in front of the detector in conventional ERDA induces a high straggling contribution in the total depth resolution. The use of a $\Delta E-E$ telescope eliminates this straggling contribution because the recoil energy can be simply deduced by summing the energy loss in the thin detector ΔE and the residual energy E_R. Only the straggling contribution from the detector dead layers cannot be suppressed. For example, a depth resolution better than 25 nm has been obtained for oxygen determination in native oxide layers using a 43-MeV ^{63}Cu beam.[51] Siegele et al. have obtained a depth resolution better than 20 nm on stainless steel with 130-MeV ^{127}I ions.[54] A value of 10 nm has been obtained by Assman.[55]

Figure 7 illustrates the increased selectivity offered by the use of a solid-state telescope for simultaneous detection of ^{1}H, ^{2}H, and ^{3}H with 4-MeV ^{4}He$^+$ in a titanium hydride layer.[8]

4.2 Time-of-flight Elastic Recoil Detection Analysis

The determination of mass separation between scattered and recoil ions by measuring the difference in their time of flight (TOF) for a given path length was first introduced in the mid-1980s.[56–58] The principle of TOF spectrometry is based on the measurement of particle

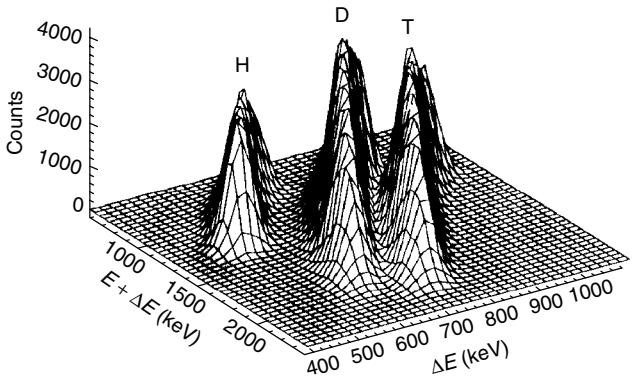

Figure 7 Three-dimensional plot of hydrogen, deuterium, and tritium distributions in a titanium hydride sample bombarded with 4-MeV helium ions. (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

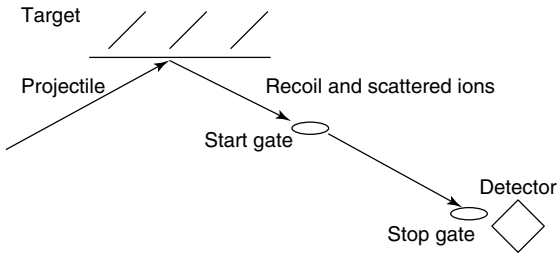

Figure 8 Schematic view of a TOF spectrometer.

flight time between a start gate and a stop gate giving two distinct signals. Simultaneously, the total energy of the particles is measured by means of a solid-state or gas-ionization detector. Figure 8 schematically describes a TOF spectrometer.

The flight time $t_f = t_2 - t_1$ of a particle with mass M and energy E_R traveling along a path of length L is given by Equation (21):

$$t_f = t_2 - t_1 = L\left(\frac{M}{2E_R}\right)^{1/2} \quad (21)$$

If the TOF spectrometer is set up with an entrance window of thickness x, Equation (20) becomes Equation (22):

$$t_f = L\left\{\frac{M}{2[E_R - S(E)x]}\right\}^{1/2} \quad (22)$$

where $S(E)$ is the stopping power of the window material for recoil ion M at energy E_R.

The time detector is generally an electrostatic mirror protected by a thin carbon foil ($0.5–20\,\mu g\,cm^{-2}$) so that the energy recorded by the surface barrier detector, which is also often used as the stop-signal detector, is given by Equation (23):

$$E = E_R - S(E)x \quad (23)$$

where x is the thickness of one or, in the case of two timing gates, two carbon foils. The carbon foil of the first electrostatic mirror serves as the entrance window for the TOF spectrometer. Typical flight time values are in the range 30–100 ns for a flight path of about 50 cm. Normally, the energy spectra used in the analysis is converted from timing signal. The two main reasons for this are (i) the normally better resolution for other elements than for H and He and (ii) the linear calibration for all the elements, which is not the case for a solid-state detector. A well-known limitation of the TOF-E technique is the poor detection efficiency (typically 10–60%) of timing gates for H, which is strongly stopping force and therefore, energy dependent. An elegant way to get around this limitation is to use the fact that for elements heavier than B, the detection efficiency is very close to 100%, and if the timing window is long enough, almost all noncoincident events seen only by the E detector are hydrogen. By summing up coincident events from the H curve and noncoincident events, 100% detection efficiency is obtained for H. This technique is described in detail in two round-robin characterizations where H was also studied.[59,60]

Figure 9 gives an example of the application of a TOF spectrometer for the ERDA characterization of a glass ceramic with an 84-MeV ^{127}I beam.[8] It can be seen that the mass resolution is less than 1 for low masses, and that ^{10}B and ^{11}B isotopes are clearly distinguishable. Nevertheless, mass resolution decreases with increasing mass so that ^{27}Al and ^{28}Si cannot be separated. This is mostly due to poor energy resolution of the solid-state detector. The mass resolution can be drastically improved by introducing a gas-ionization detector equipped with a thin SiN entrance window.[61] However, Al and Si can be readily separated with a $\Delta E-E$ gas detector as shown by Siegele et al.[54]

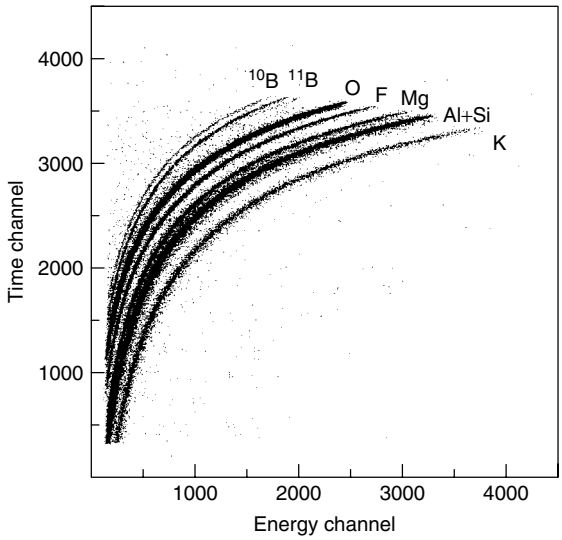

Figure 9 TOF spectrum obtained for a glass ceramic bombarded by a 84-MeV ^{127}I beam. (Reproduced with the kind permission of Plenum Press.)

A characteristic feature for TOF-ERDA is the fact that depth resolution at the surface improves when the energy of incident ion is lowered. This is due to the fact that the elemental energy spectra used in the analysis are converted from TOF spectra. Even with a simple TOF-E telescope, all elements below Ne can be separated from each other with projectile (Cl, Br, and I) energies below 20 MeV.

4.3 Coincidence Spectrometry

In the first ERDA experiments, from Cohen in 1972[31] to L'Ecuyer in 1976,[1] transmission geometry together with coincident detection of scattered and recoil particles were applied to achieve mass selectivity, to reduce the background, and to improve the sensitivity.

By detecting both particles, four parameters can easily be obtained: E_S, E_R, θ, and ϕ. Thus the mass of the recoil nucleus M and the interaction depth x can be deduced according to Figure 10. One variation of this technique, proton-proton scattering utilized by Reichart et al., allows sensitive 3D profiling of hydrogen down to ppm-levels.[62]

Two families of ERDA coincidence methods can be distinguished: those carefully adjusting θ and ϕ to achieve mass selectivity and to reduce the background, and those carefully measuring E_S and E_R to eliminate θ and ϕ.

Depth information is the ultimate goal of the first family of methods; this can be deduced from energy measurement. Large solid detectors are generally used in the second family of methods, and as a result, depth resolution is sacrificed. Figures 11 and 12 give one example of the application of each method.

Partial separation of ^{63}Cu and ^{65}Cu has been obtained by Klein et al.[63] using a pair of position-sensitive detectors at the GSI heavy-ion microprobe (Darmstadt,

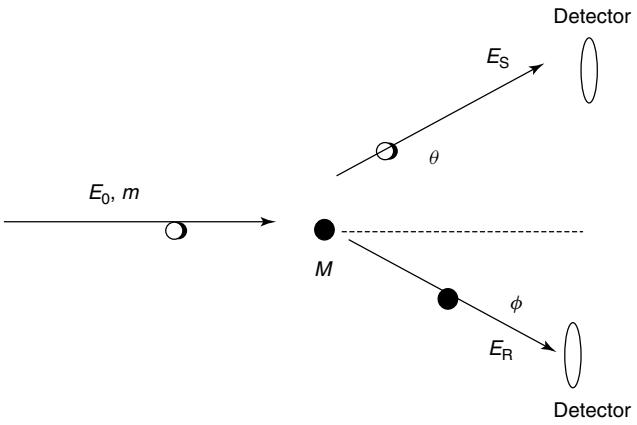

Figure 10 Typical coincidence configuration.

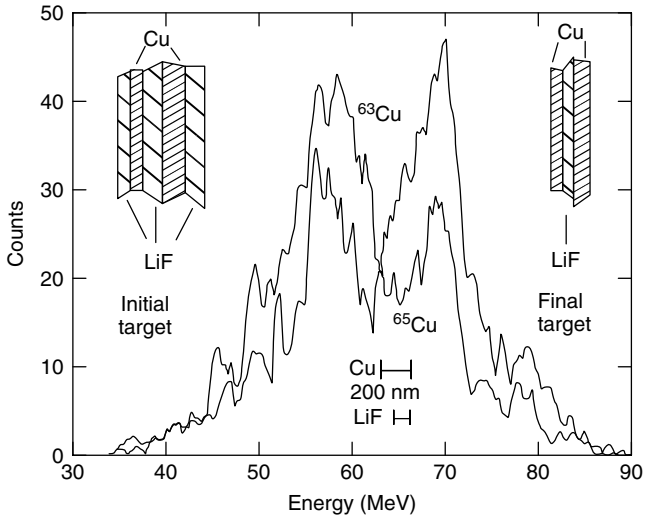

Figure 11 Coincident elastic recoil detection spectra of two thin Cu films separated by LiF measured with ^{58}Ni ions.[63] The final target is different from the initial one because the outer LiF layers have been partially decomposed during the irradiation. (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

Germany) on a sandwich target (LiF/Cu/LiF/Cu/LiF) with 1.4-MeV amu^{-1} ^{58}Ni ions. The dominant peak in Figure 12(a) corresponds to recoiled carbon atoms; recoiled oxygen is seen in the spectrum at higher He energies. Figure 12(b) is similar to Figure 12(a), but rotated by 180°, the surface line corresponds to the scattering process at the rear surface of the target.

4.4 Electromagnetic Cross Filters ($E \times B$)

The Canadian group from Varennes was the first to propose the use of an electromagnetic cross filter as an achromatic mass and charge selector and not only as a Wien filter.[65,66] A typical $E \times B$ (electric and magnetic crossed fields) device for ERDA measurement is shown in Figure 13.

When a particle passes through the $E \times B$ filter, Equations (24) and (25) describe its exact movement:

$$m\frac{dv_y}{dt} = q(E - v_x B) \tag{24}$$

$$m\frac{dv_x}{dt} = q_{v_y} B \tag{25}$$

Assuming $v_x = v_0$ and $v_y = 0$, they reduce to Equation (26):

$$m\frac{dv_y}{dt} = q(E - v_0 B) \tag{26}$$

At the exit of the filter, the particle is deviated by an angle α given by Equation (27):

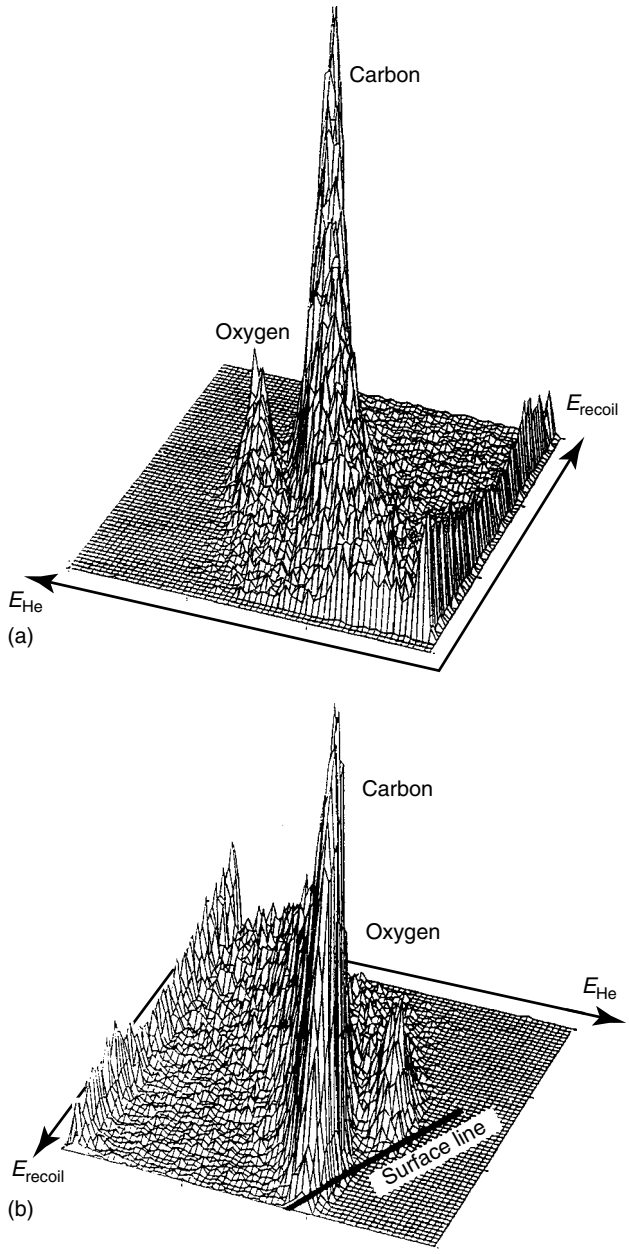

Figure 12 Coincidence recoil spectrum of a free-standing polycarbonate film obtained using a ^{4}He beam.[64] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

$$\sin\alpha \approx \alpha = \frac{qL(E - v_0 B)}{mv_0^2} \tag{27}$$

where L is the total length of the $E \times B$ filter.

The filter should deflect particles of all energies by the same angle; this condition is satisfied when Equation (28) applies:

$$\frac{d\alpha}{dE} = 0 \quad \text{or} \quad E = \frac{Bv_0}{2} \tag{28}$$

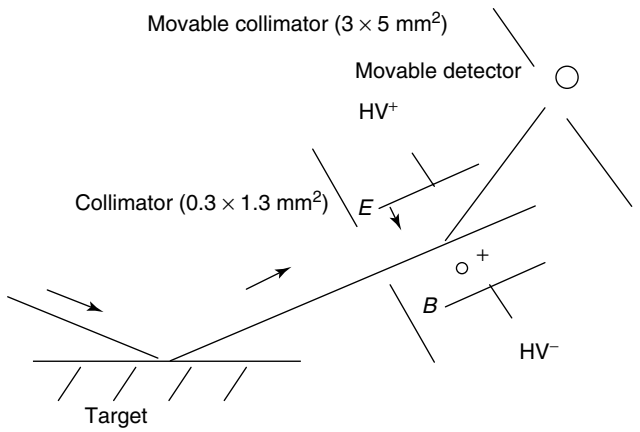

Figure 13 Schematic view of a $E \times B$ device used in ERDA.

Typical values for L, B, and E are, respectively, 70–80 mm, 1 kG, and 1.5–2 kV cm^{-1}.

Electromagnetic cross filters have been evolving since 1995 and more sophisticated devices have appeared such as the $B - E \times B - B$ filter proposed by Serruys, which has been described previously.[8]

The Canadian group from INRS Varennes and A. Chevarier's group in France (Lyon University) have collaborated together on the use of the $E \times B$ filter to characterize amorphous hydrogenated layers. They have shown that hydrogen desorption induced by 2.5-MeV $^4He^+$ with or without previous deuterium implantation was smaller for a-SiC:H than for a-C:H layers.[11] Figure 14 compares the hydrogen distributions in a-SiC:H and a-C:H layers after a deuterium implantation at a level between 2.2 and 3.5×10^{16} D cm^{-2}.

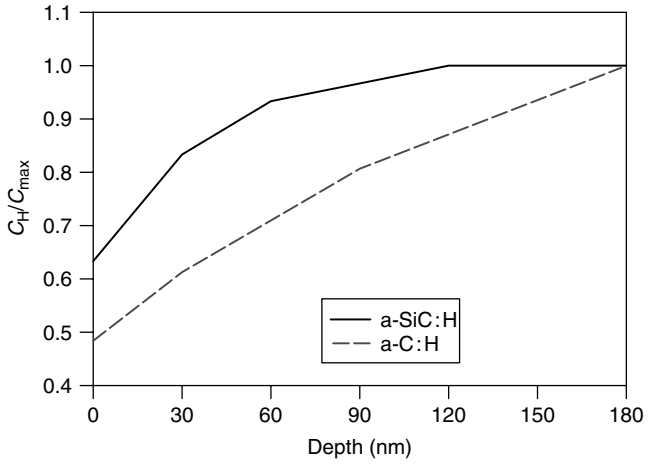

Figure 14 Comparison of the hydrogen distributions in a-SiC:H and a-C:H layers after a deuterium implantation between 2.2 and 3.5×10^{16} D cm^{-2} measured with a 2.5-MeV ^{4}He beam.

5 HIGH-ENERGY AND HEAVY-ION ELASTIC RECOIL SPECTROMETRY

As discussed in Section 2, low-energy ^{4}He (2–4 MeV)-induced ERDA is adapted to profile target nuclei lighter than the projectile: generally ^{1}H and ^{2}H. Using a specific coincidence detection device allows the researcher to extend the profiled mass range to target nuclei heavier than the projectile (see Section 3). There are two other means to improve the range of analyzable elements and the total explored depth. The first consists in increasing the energy of the incident ^{4}He beam to improve the energy separation between the contributions of the different target components and also to take advantage of the increase in the non-Rutherford cross section for light elements.[67,68] The second consists in increasing both the mass and the energy of the incident ion beam; this technique is called *high-energy HI-ERDA*.[48,69–71] The selectivity is thus largely improved; this was the configuration adopted by L'Ecuyer et al. in 1976 to investigate multilayered structures.[1]

5.1 High-energy ^{4}He-induced Elastic Recoil

When ^{4}He incident energy increases above 1 MeV, Rutherford scattering is no longer an applicable model. High-energy ^{4}He scattering in light elements has a non-Rutherford behavior and cross-section resonances exist for several couples ^{4}He/target nucleus, as shown in Figure 15 for carbon and oxygen.

Figure 16 shows the ERDA spectrum for ^{12}C recoil of an Al/C sandwich target, obtained using a 30 MeV α-particle beam in transmission geometry.

5.2 High-energy Heavy-ion-induced Elastic Recoil

Equation (4), describing the kinematics of the collision m/M, with $m \gg M$ becomes Equation (29):

$$E' = 4E_0 \cos^2 \phi \frac{M}{m} \tag{29}$$

The interaction cross section follows Rutherford behavior as long as the incident energy is below the threshold energy defined by Equation (9).

The analytical expression of the differential recoil cross section given by Equation (8) becomes Equation (30):

$$\frac{d\sigma}{d\Omega} = \left(\frac{zZe^2}{2E_0}\right)^2 \left[\frac{(m/M)^2}{\cos^3 \phi}\right] \tag{30}$$

Several points should be noted:

- The differential cross section is nearly independent of the recoil mass because the (Z/M) ratio ranges

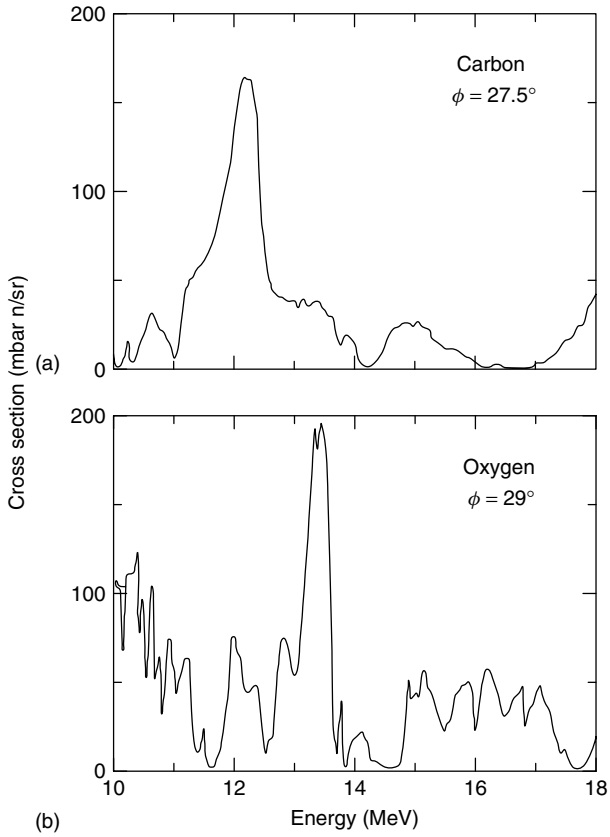

Figure 15 Scattering cross section for (a) $^4He/^{12}C$ and (b) $^4He/^{16}O$ collisions between 10 and 18 MeV.[72–74] (Reproduced by courtesy of H. Rijken and with the kind permission of Plenum Press.)

between 0.4 and 0.5 for all the elements except hydrogen.

- For hydrogen recoils $(Z/M) = 1$ and the differential cross section is increased by a factor of four relative to the other recoiled target atoms.
- The sensitivity of high-energy HI-ERDA increases strongly with increasing z. Anyhow, if a sophisticated detection system like $\Delta E - E$ or TOF-E is used in the analysis, the high-z projectile does not necessarily provide the lowest detection limit, but that can be obtained with lighter projectiles giving more energy and therefore better separation in the 2-D histogram for the recoils.
- The selectivity of high-energy HI-ERDA decreases with increasing Z. For example, C, N, and O can be separated well in stainless steel using a 136-MeV ^{127}I beam,[75] but transition elements such as Cr, Fe, and Ni cannot be properly separated using 136-MeV ^{127}I and the Bragg counter as shown in Figure 17. Nevertheless, Siegele et al. have shown that this separation is possible at 136 MeV, using a telescope with a gas chamber as the ΔE detector.[76]

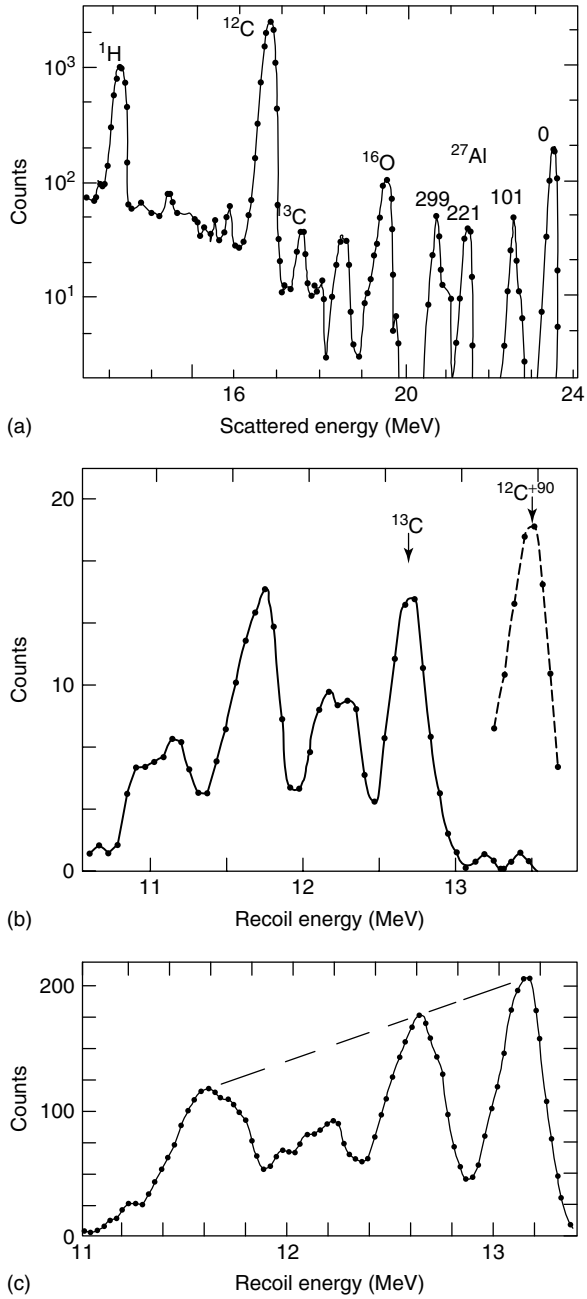

Figure 16 (a) Scattered ^4He spectrum, (b) ERDA spectrum for ^{13}C recoils, and (c) ERDA spectrum for ^{12}C recoils of an Al/C sandwich target, obtained using a 30-MeV α-particle beam in transmission geometry.[67] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

- The differential cross section increases with increasing recoil angle ϕ (factor 5 between 30 and 60°), but simultaneously the recoil energy is reduced. Generally, the detector system is located around 40°.

High-energy HI-ERDA requires specific detection devices such as those described in Section 4 (TOF

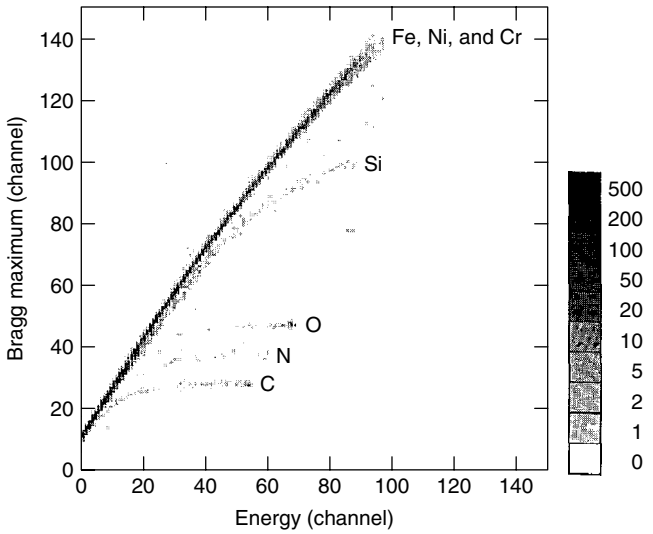

Figure 17 Two-dimensional plot (maximum energy loss/recoil energy) obtained on stainless steel with 136-MeV ^{127}I.[75] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

spectrometer or telescope). Electrostatic or magnetic spectrometers and ionization chambers are also in use.[68] Data processing is generally based on two or more parameter data acquisition and the elemental information is expressed in matrix form.

The analytical performance offered by high-energy HI-ERDA, particularly depth resolution, depends strongly on the total energy resolution of the detection system, on the nature and energy of incident ions and recoil particles, and on the depth where the collision has occurred. Dollinger et al. have reported a surface depth resolution of 8 nm for hydrogen determination in a thin carbon film bombarded by 120-MeV ^{197}Au ions in transmission geometry using a large magnetic spectrograph.[69] For carbon/boron multilayers deposited on a thick silicon substrate, a surface depth resolution of less than 1 nm is achieved in reflection geometry using the same beam parameters.[69] With low-energy ions and standard TOF-E telescope Brijs et al. demonstrated a depth resolution of 2 nm at the surface for H, N, and Si when characterizing very thin $SiO_2/Si_3N_4/SiO_2$ film stacks using glancing angles.[77] Measurement using 6-MeV ^{35}Cl incident ions and very glancing $3 + 35 = 38$ angle in reflection geometry is presented in Figure 18.

Normal ERDA geometry with the detector around 40° also offers good possibilities for the use of forward scattered incident ions, in addition to recoils, in the analysis. One possibility is to use light incident ion and do forward scattering analysis. The benefit of this is the high cross section and good depth resolution because of the small energy spread due to multiple scattering and kinematics. An example of this is the

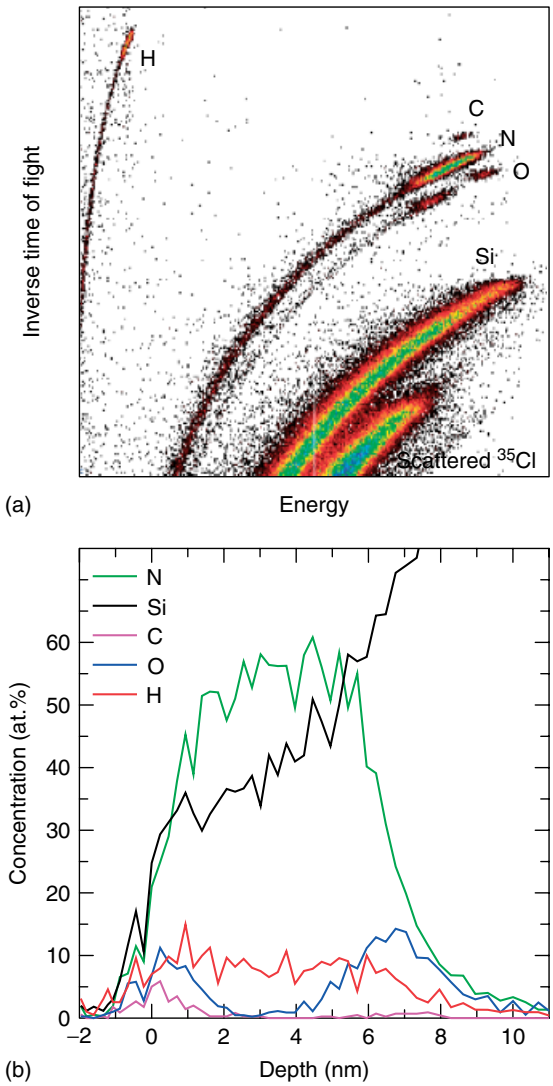

Figure 18 A metalorganic chemical vapor deposited Si_3N_4 film with some H and O impurities measured with 6-MeV ^{35}Cl beam using 3.2° incident and 35° recoiling angles. (a) A TOF-E histogram of the measurement showing all the sample elements and (b) the depth profiles converted from elemental TOF spectra using slab analysis approach. Density of 3.3 g cm^{-3} was used in conversion of the depth scale to nanometers.

study of 6-nm HfSiO and 25-nm NiSi films on silicon using 3–8-MeV ^{16}O ions as projectiles.[78] The use of scattered incident ions with recoiled target atoms in the analysis is demonstrated in Figure 19 where an atomic layer deposited Li–La–O film was measured using a 16-MeV ^{63}Cu beam. The recoiled La atoms suffered greatly from multiple scattering and could not be used in the analysis without proper MC simulations, but as La is the only heavy element present, scattered ^{63}Cu could be used instead. As two different scattering cross sections were used, the scattering angle had to be known very accurately.

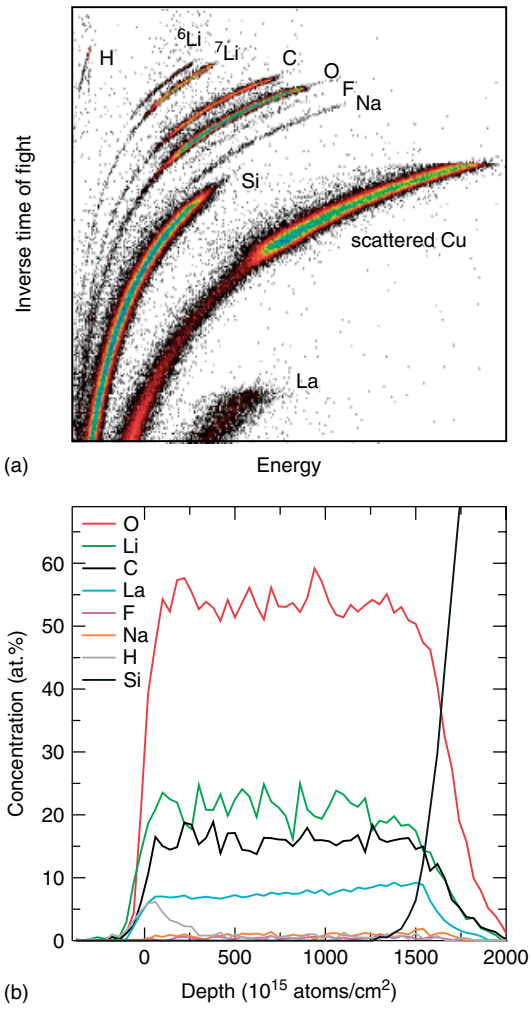

Figure 19 An atomic layer deposited Li–La–O film with nominal thickness of 165 nm measured with a 16-MeV ^{63}Cu beam and 19° incident and recoiling angles. (a) A TOF-E histogram of the measurement showing all the sample elements and (b) corresponding depth profiles. The scattered incident ions were used in the analysis of La profile. Note the low-energy tails due to multiple scattering.

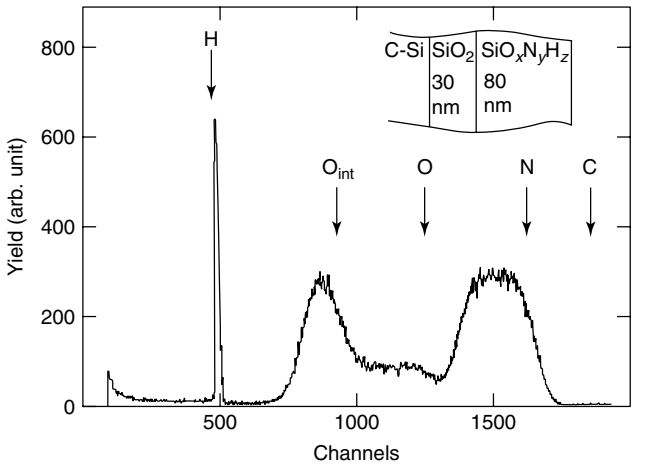

Figure 20 ERDA spectrum of an 80-nm silicon oxynitride film coated on 30-nm SiO_2 deposited on Si obtained with 30-MeV ^{28}Si ions.[80] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

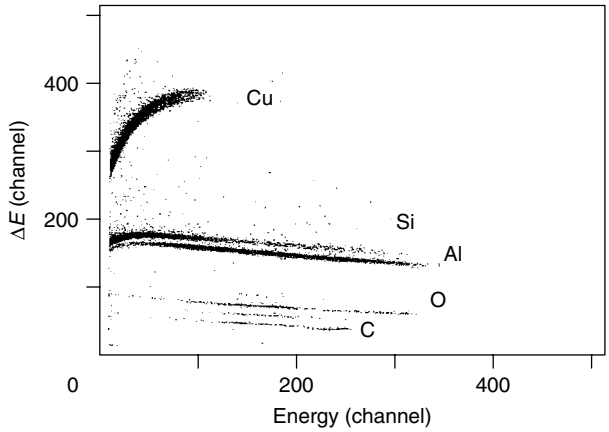

Figure 21 $\Delta E-E$ matrix and recoil energy spectra for a 15-nm Al/Cu multilayer deposited on silicon obtained by using a 200-MeV ^{197}Au ion beam.[81] (Reproduced with the kind permission of Elsevier Science and Plenum Press.)

The main limitation on analytical performance in high-energy HI-ERDA (depth resolution, sensitivity, or analyzed depth) is attributed to the behavior of the target material under ion bombardment. In the case of films with a thickness of the order of nanometers, the sputtering phenomenon, which is directly proportional to the nuclear energy loss, has to be taken into account when analyzing the measurements. In the case of thicker films, especially of polymers, a much more visible effect is the loss of light elements throughout the film thickness. In the case of porous low-*k* materials, the losses were found to be directly proportional to the energy loss of incident ions in the films and therefore to minimize the losses, the heaviest possible ion had to be used in the analysis.[79]

Figures 20 and 21 give two different examples of the use of high-energy HI-ERDA. The first example deals with the analysis of an 80-nm silicon oxynitride film coated on 30-nm SiO_2 deposited on Si with 30-MeV ^{28}Si ions.[80] The second example is dedicated to the characterization of a 15-nm Al/Cu multilayer deposited on silicon by using a 200-MeV ^{197}Au ion beam.[81]

Another general feature of HI-ERDA that has a great effect on the analysis reliability is multiple scattering, which, for more detailed analysis, can be divided to multiple and plural scattering. This multiple scattering causes angular spread, which can be described with analytical functions.[82–84,85] Equally important in the analysis are events where the scattering occurs closer to the sample atom nucleus resulting in a large angle

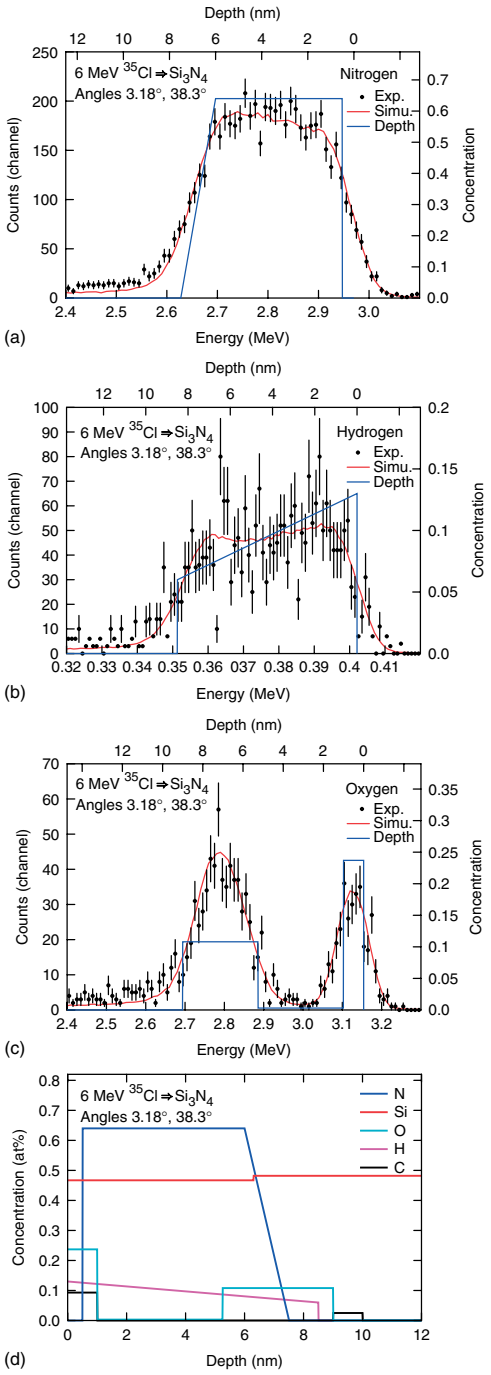

Figure 22 Monte Carlo simulations of the energy spectra for (a) nitrogen, (b) hydrogen, and (c) oxygen for the same measurement and data as in Figure 18. Black dots are experimental yields, red lines MC simulations, and blue lines depth profiles used for MC simulations. In (d) all depth profiles used for reproducing experimental spectra are collected in one depth profile.

scattering. Although these events are scarce, they can play an important role in the analysis. These scatterings are referred as *plural scattering*. Due to the nature of these events, no analytical function has been used to simulate them and they require full MC simulations. Multiple and plural scattering effects are strongly enhanced if the studied sample contains heavy elements; to carry out a very detailed analysis, the MC-simulation program should be used. The most sophisticated one at the moment is MCERD by Kai Arstila,[86] which can, due to several computational enhancements, do full simulations in roughly the same time as it takes to measure the sample. In Figure 22 an example of the use of MCERD in the analysis is shown for the same measurement as in Figure 18. The experimental energy spectra are well reproduced; the low-energy tails, and reliable depth profiles are also obtained. The conclusion of rising Si content deeper in the film drawn on the basis of slab analysis (Figure 18) can be proven to be wrong by MC-based analysis. In addition to MC simulations, also Molecular Dynamics simulations have a great potential in interpreting IBA results.[87]

6 APPLICATION EXAMPLES OF ELASTIC RECOIL DETECTION ANALYSIS FOR PROFILING HYDROGEN AND OTHER ELEMENTS

Four main application fields can be essentially distinguished for ERDA measurement, using either a low-energy ^{4}He beam, higher energy ^{4}He, or heavier ions: polymer science, microelectronics, thin films, and interface reactions. Each of these fields is illustrated in this section.

6.1 Polymer Sciences

Polymers are the most hydrogenated media encountered among solids. Three complementary topics of interest have given rise to the development of ERDA-based studies: the investigation of polymer surface properties, the hydrogen distribution in polymer blends and the irradiation behavior of polymers.

Many application examples have been presented and discussed by Tirira et al.[8] Only some of them are presented in this article.

For example, Chou et al. have studied the modifications induced by radio frequency or microwave treatment under a D_2O atmosphere of polyimide films. They showed that the polyimide surface exhibited a *D* uptake proportional to the time of treatment.[88] Moreover, a correlation has appeared between the surface concentration of hydroxy groups and the development of surface morphology during plasma treatment.

Green et al. studied the segregation of deuterated copolymer (molecular weight *M* in the range 87 000–520 000) at the interface of a homopolymer blend

(polystyrene/poly(vinyl methyl ether) with a molecular weight between 100 000 and 200 000) using 2.8-MeV ^{4}He ions.[89] They have shown that the diffusion coefficient of deuterated polymer varies as M^{-2}.

Composto et al. measured the diffusion coefficient of polymer molecules (deuterated polystyrene and deuterated poly(xylenyl ether) chains) in a complex polymeric structure using D as a tracer.[90] Diffusion coefficients have shown to be independent of the matrix molecular weight and to decrease as M^{-2}.

Several authors have applied ERDA to evaluate the degradation of polymer composition under charged-particle bombardment.[91] Wang et al. have investigated the modification of the electrical conductivity of high-temperature and high-performance polyetheretherketone (PEEK) and polyethersulfone polymers by ion implantation (50 and 180 keV As and Xe ions). A two-component regime has been found in which dominance depends on the implantation fluence ($<10^{15}$ ions cm^{-2} or $>10^{17}$ ions cm^{-2}). Another example of ERDA study of polymer degradation was given in Section 3. Figure 5 corresponds to the application of μERDA for polymer characterization after high-energy heavy-ion bombardment. In this case, 50-MeV ^{32}S ions were previously used to irradiate thin polyimide foils (25 μm). Bombardment-induced bond breaking leads to the formation of hydrogen-rich zones (micrometric bubbles or association of nanometric bubbles) trapped in the degraded polymer structure, as seen in Figure 5.

6.2 Semiconductor Materials

In this area, ERDA measurements are generally dedicated to profiling successive thin layers coated on a crystalline semiconductor substrate (Si, Ge, or GaAs) having different compositions and thicknesses.

The hydrogen profile is measured using a conventional approach with ^{4}He ions as shown by Wang et al.[92] or Barbour et al.[93] Other light elements such as C, N, and O have been profiled using high-energy HI-ERDA.[80]

Wang et al. have profiled hydrogen in semi-insulating polysilicon films. Figures 23 and 24 show the hydrogen recoil spectra obtained on films immersed in a gas flow reactant mixture (N_2O/SiH_4) in various ratios or annealed between 700 and 1000 °C for 30 min[92] for a given gas reactant ratio. The film deposition rate decreases when the N_2O/SiH_4 ratio increases owing to the increase in oxygen concentration. The H content of the film decreases when the annealing temperature increases owing to the densification and restructuring of the film.

Barbour et al. have shown that for silicon nitride film deposition by electron cyclotron resonance, the hydrogen content of the film was minimized when the stoichiometry was closest to Si_3N_4.[93]

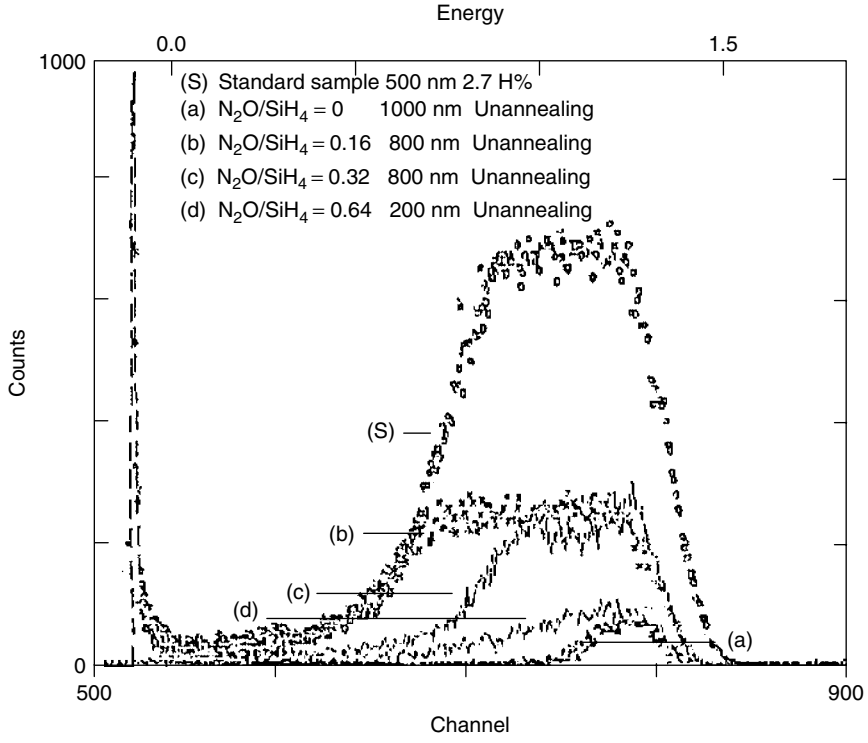

Figure 23 Hydrogen recoil spectra for polysilicon samples immersed in a N_2O/SiH_4 flow with various N_2O/SiH_4 ratios. (Data reproduced from Wang et al.,[92] with the kind permission of Elsevier Science and Plenum Press.)

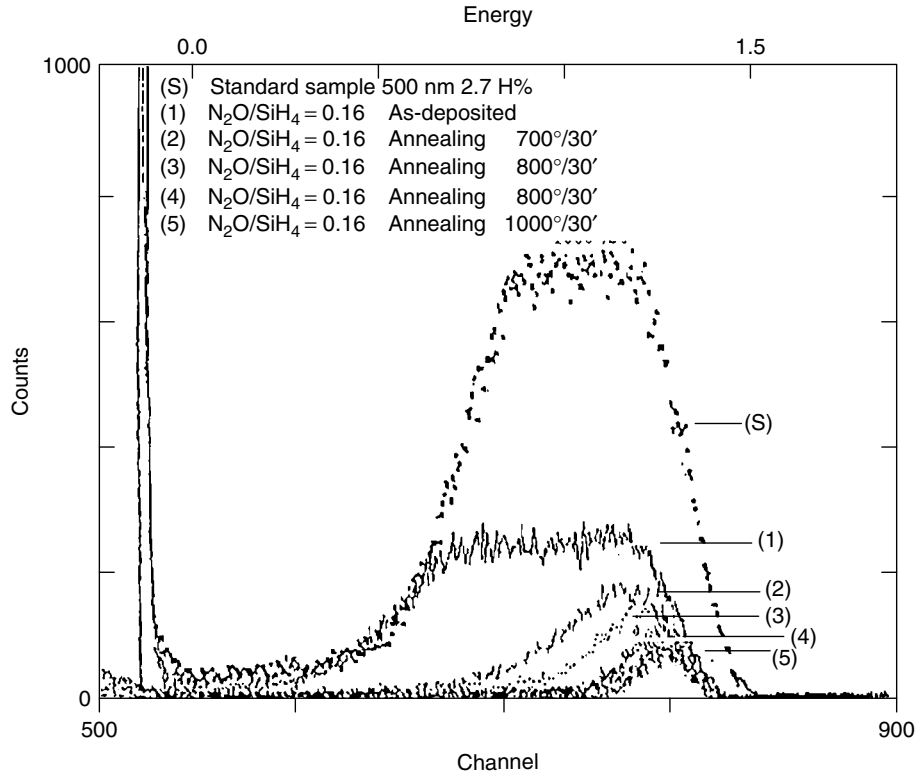

Figure 24 Hydrogen recoil spectra obtained for various annealing temperatures. (Data reproduced from Wang et al.,[92] with the kind permission of Elsevier Science and Plenum Press.)

6.3 Thin Films

In the case of thin-film monolayered or multilayered structures, ion beam methods have been revealed to be powerful characterization tools, particularly considering the HI-ERDA[94] and coupling of RBS with ERDA.[95–97]

The development of materials based on hydrogenated amorphous carbon ''diamond-like'' or silicon-carbide layers, because of their promising tribological properties, strongly influenced the interest in ERDA in the early 1990s.[98–101]

The relationship between the optical and electrical properties of hydrogenated amorphous silicon carbon films and their stoichiometry have been determined combining RBS and ^{4}He-induced ERDA by Compagnini et al.[95] A hydrogen concentration of 30 atom% was shown to increase the optical energy gap by a factor of two and the electrical resistivity by several orders of magnitude.

Figure 25 gives three recoil spectra obtained on a:C–H or a:C–D layers. Hydrogen implantation in the a:C–D layer was shown to induce a large surface D release,[99] extending to at least 80% of the analyzed depth (0–0.3 μm). Microelectronics is an important driving force in the development of this method. IBA has played an important role in the development of films replacing SiO_2 as a gate oxide[102] and the role of IBA in the film development has been very important. Another important thin-film application related to microelectronics has been Cu diffusion barriers like TiN[103] and TaN.[104,105] During the development process, deposited films normally contain substantial amounts of H, C, and O due to the precursors used – HI-ERDA can easily provide this information.

6.4 Interface Reactivity

This field of application corresponds to the experimental study of atomic transport mechanisms near a solid/solid or a solid/liquid interface, for example, oxide corrosion behavior,[106,107] wear properties of treated metal, amorphous layers, or glass surfaces.[108–113] Intensive use of ERDA has been reported by the Canadian group from Varennes[114] and by the French group from the University of Lyon.[11] These two groups have collaborated in studying some candidate materials for fusion reactor device coatings that act as facings for the plasma. Essentially, they studied hydrogen loss under irradiation and annealing conditions.

Figure 26 shows the erosion of the hydrogen depth profile in an a:C–H layer after $^4He^+$ bombardment at

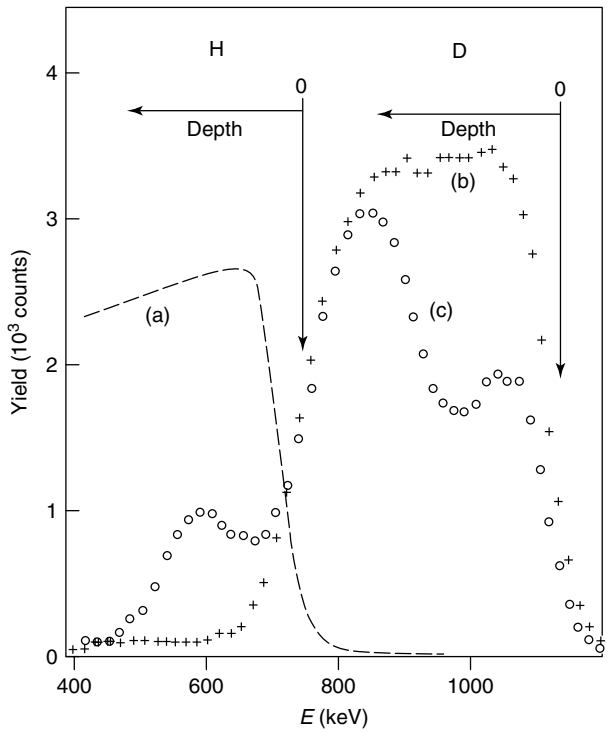

Figure 25 Energy spectrum of recoil protons and deuterons obtained with a 2.6-MeV $^4He^+$ beam (incidence angle = 75°, recoil angle = 30°). (a) Proton reference from unimplanted a:C–H layer, (b) deuteron reference from unimplanted a:C–D layer, (c) implanted area of the same layer in (b) after implantation with 11-keV H^+ at a fluence of 1.8×10^{17} ions cm^{-2}. The depth scale is approximately 0–0.3 μm. (Data from Boutard et al.,[99] with the kind permission of Plenum Press.)

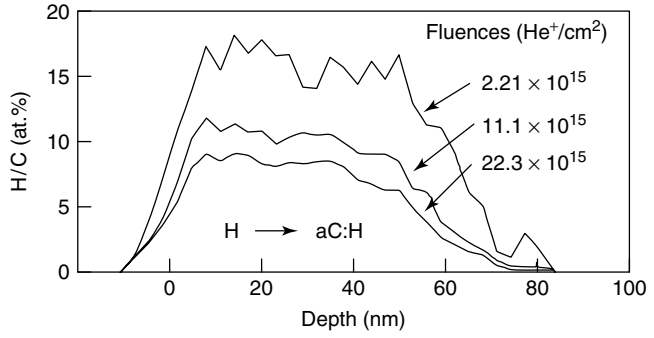

Figure 26 Evolution of the hydrogen depth profile measured by the $E \times B$ technique in an a:C–H layer bombarded with $^4He^+$ between 2.21×10^{15} and 2.23×10^{16} ions cm^{-2}. (Data reproduced from Ross and Richard,[115] with the kind permission of Plenum Press.)

different fluences. The $E \times B$ technique is used with a 350-keV incident helium beam.[115]

Ross and Richard have shown that hydrogen depletion is negligible in Si, SiC, and Sc, higher in Be, Si_3N_4, and TiC, and very high in C.

6.5 Other Application Examples

Other conventional ERDA or high-energy HI-ERDA applications have been reported since the late 1980s. For example, Mosbah et al. successfully proposed the use of ERDA in transmission geometry to measure the water content of melt inclusions trapped in minerals.[44] Their first published data concerned inclusions from Pantelleria, San Pietro, and Guadeloupe: water contents ranged between 3.4 and 11.8 wt% and agreed well with theoretical calculations. Naitoh et al. have studied the reordering of an aluminum-adsorbed silicon surface like $\sqrt{3} \times \sqrt{3}$ Al/Si (111) induced by hydrogen sorption.[116] They found that saturation occurs for a hydrogen coverage of about 1.3×10^{15} H cm^{-2}. One totally new application for the TOF-ERDA telescope was introduced by Zhang et al. in 2001.[117] For doing stopping force measurements, they installed a movable thin film between the second timing gate and energy detector and could then obtain continuous stopping curves for this film. This idea has now been used in many places for stopping force measurements, and has given better stopping values for the IBA community.

7 CONCLUSION

Three of the most striking characteristics of the ERDA method are its wide usable energy range from 350 keV to 1 GeV, its strong versatility in terms of detection configurations and variants, and with HI-ERDA, the possibility of depth profiling all the sample elements from H to Bi in a single measurement.

Besides the fact that it remains a delicate method, ERDA has featured in IBA-related conferences as the most promising and evolving technique for quantitative determination of elemental depth profiles in solids, from hydrogen to medium and even high mass numbers. In this approach, no prior knowledge of the sample is required and good results can be obtained without standard samples. The basic physics of the elastic collision process has made substantial progress since the late 1980s, especially when considering differential cross-section measurements and models or data-processing methods.

Moreover, the spectacular development of high-energy HI-ERDA since the beginning of the 1990s must be emphasized. It is based on the strong theoretical and experimental capabilities derived from nuclear physics research (experimental devices, technical procedures, and theoretical models) that have brought a constant source of improvement to ERDA users for materials science applications. More and more IBA laboratories worldwide have developed specific facilities for applying ERDA to various problems in this field. In addition to

high-energy HI-ERDA, special TOF-E telescopes have been constructed for laboratories with smaller (1.7–3-MV terminal voltage) tandem accelerators, making this unique analysis technique available to more users.

The different application examples presented and discussed in the above sections constitute a good sample of the use of ERDA.

ERDA is the most practical IBA technique available for profiling hydrogen isotopes in the near-surface region of solids from 100 nm to 10 μm with a depth resolution of 20 nm for conventional ERDA and even 1 nm for HI-ERDA at the surface. The use of resonant nuclear reactions induced by high-energy heavy-ion beams to profile hydrogen seems to be restricted to specific applications and particularly to samples able to support high fluences of 0.4–0.5-MeV amu^{-1} ^{15}N ions.[118] A round-robin characterization of ultrathin to thin (1–100 nm) aluminum oxynitride (AINO) films was organized and the participants used different IBA techniques.[60] This paper gives a good overview of the possibilities and limitations of present IBA techniques in the case of thin films. Considering the analytical capabilities of SIMS, it must be noted that besides its excellent depth resolution and its very strong sensitivity, SIMS is a destructive technique and the calibration of the erosion rate of the target under investigation remains difficult. Nevertheless, the technical improvements encountered with the SIMS machines of the last generation tended to enhance the competition between these two methods.

The future of ERDA looks bright. In three decades, it has matured from something new and exciting to a widely used and trusted technique. In the beginning, the technique was utilized in big laboratories mostly concentrating on nuclear physics, but having plenty of useful instruments around. In the next phase, smaller laboratories, which had mostly concentrated on RBS adopted the conventional ERDA technique, but the instrumentation development was still concentrated on bigger institutes that had mostly high-energy heavy beams available. This millennium, the development has again been toward smaller accelerators, but now using HI-ERDA usually with TOF-E detectors. In these laboratories, the use of 5–20-MeV heavy incident beams have replaced, in many applications, the previously intensively used RBS and nuclear reaction analysis (NRA). This development toward lower energies is likely to continue as large tandem accelerators are not constructed anymore and old ones are being dismantled.

It is not difficult to predict that, in the future, accurate ERD analysis will be performed by means of MC-based simulation software capable of handling properly both multiple and plural scattering, and also all the instrumental effects that broaden the energy spectra. With the aid of this type of analysis, the surface depth resolution, and especially the reliability of analysis deeper in the sample, will be pushed to a new level.

ACKNOWLEDGMENTS

Dr. Kai Arstila is acknowledged for the MC simulations.

ABBREVIATIONS AND ACRONYMS

ERDA	Elastic Recoil Detection Analysis
HI-ERDA	Heavy-ion Elastic Recoil Detection Analysis
IBA	Ion Beam Analysis
MC	Monte Carlo
μERDA	Microbeam-induced Elastic Recoil Detection Analysis
NRA	Nuclear Reaction Analysis
PEEK	Polyetheretherketone
RBS	Rutherford Backscattering Spectrometry
SIMS	Secondary Ion Mass Spectrometry
TOF	Time-of-flight

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission) • Rutherford Backscattering Spectroscopy

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

REFERENCES

1. J. L'Ecuyer, C. Brassard, C. Cardinal, J. Chabbal, L. Deschênes, J.P. Labrie, B. Terreault, J.G. Martel, R. Saint Jacques, 'An Accurate and Sensitive Method for the Determination of the Depth Distribution of Light Elements in Heavy Materials', *J. Appl. Phys.*, **47**, 381–382 (1976).
2. B.L. Doyle, P.S. Peercy, 'Technique for Profiling ^{1}H with 2.5 MeV Van de Graaff Accelerators', *Appl. Phys. Lett.*, **34**, 811–813 (1979).
3. G.M. Padawer, E.J. Schneid, 'A Lithium Microprobe Technique for Measuring the Concentration Profile of Hydrogen at Surfaces', *Trans. Am. Nucl. Soc.*, **12**, 493–494 (1969).

4. E. Ligeon, A. Guivarch, 'A New Utilization of ^{11}B Ion Beams: Hydrogen Analysis by ^{1}H(^{11}B, α)$\alpha\alpha$ reaction', *Radiat. Eff.*, **22**, 101–105 (1974).
5. W.A. Lanford, H.P. Trautvetter, J.F. Ziegler, J. Keller, 'New Precision Technique for Measuring the Concentration Versus Depth of Hydrogen in Solids', *Appl. Phys. Lett.*, **28**, 566–568 (1976).
6. D.A. Leich, T.A. Tombrello, 'A Technique for Measuring Hydrogen Concentration Versus Depth in Solid Samples', *Nucl. Instrum. Methods*, **108**, 67–71 (1973).
7. J.P. Thomas, C. Pijolat, M. Fallavier, 'Emploi de la Réaction Résonnante ^{1}H(^{15}N, $\alpha\gamma$)^{12}C pour L'obtention de Profils de Concentration de l'hydrogène dans les Matériaux', *Rev. Phys. Appliquée*, **13**, 433–439 (1978).
8. J. Tirira, Y. Serruys, P. Trocellier, *Forward Recoil Spectrometry: Applications to Hydrogen Determination in Solids*, Plenum, New York, 1996.
9. L. Calcagno, G. Foti, 'Interdiffusion in Polystyrene Crosslinked by keV ion Irradiation', *J. Appl. Phys.*, **71**, 3216–3217 (1992).
10. T. Sajavaara, J. Jokinen, K. Arstila, J. Keinonen, 'TOF-ERDA Spectrometry Applied for the Analysis of Be Migration in (100) GaAs', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **139**, 225–229 (1998).
11. G. Barbier, '*Etude du Comportement de l'hydrogène dans des Matériaux Amorphes Hydrogénés de Type a-C: H et a-SiC: H Devant Faire Face au Plasma des Réacteurs à Fusion*', Thèse, Université Lyon I, 10 April 1997.
12. W.K. Chu, J.W. Mayer, M.A. Nicolet, *Backscattering Spectrometry*, Academic Press, New York, 1978.
13. M. Bozoian, 'Threshold of Non-Rutherford Nuclear Cross-section for Ion Beam Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **56/57**, 740–743 (1991).
14. P. Goppelt, B. Gebauer, D. Fink, M. Wilpert, Th. Wilpert, W. Bohne, 'High Energy ERDA with Very Heavy Ions Using Mass and Energy Dispersive Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **68**, 235–240 (1992).
15. J. Räisänen, E. Rauhala, J. Knox, J.F. Harmon, 'Non-Rutherford Cross-section in Heavy Ion Elastic Recoil Spectrometry: 40–70 MeV ^{32}S Ions on Carbon, Nitrogen and Oxygen', *J. Appl. Phys.*, **75**, 3273–3277 (1994).
16. H.H. Andersen, F. Besenbacher, P. Loftager, W. Möller, 'Large-Angle Scattering of Light Ions in the Weakly Screened Rutherford Region', *Phys. Rev. A*, **21**, 1891–1901 (1980).
17. G. Freier, E. Lampi, W. Sleator, J.H. Williams, 'Angular Distribution of 1 to 3.5 MeV Protons Scattered by ^{4}He', *Phys. Rev.*, **75**, 1345–1347 (1949).
18. G.R. Miller, G.C. Phillips, 'Scattering of Protons from Helium and Level Parameters in ^{5}Li', *Phys. Rev.*, **112**, 2043–2047 (1958).
19. R.A. Arndt, D.D. Long, L.D. Roper, 'Nucleon–alpha Elastic-scattering Analyses', *Nucl. Phys. A*, **209**, 429–446 (1973).
20. G.R. Satchler, W. Owen, A.J. Elwyn, G.L. Morgan, R. Walte, 'An Optical Model for the Scattering of Nucleons from ^{4}He at Energies Below 20 MeV', *Nucl. Phys. A*, **112**, 1–30 (1968).
21. C.L. Critchfield, D.C. Dodder, 'Phase Shifts in Protonalpha Scattering', *Phys. Rev.*, **76**, 602–605 (1949).
22. D.C. Dodder, G.M. Halle, N. Jarmie, K. Witte, 'Tables of Phase Shifts and Experimental Observables for ^{4}He Elastic Scattering', Report LA-6389-MS, Los Alamos Research Laboratories, 1968.
23. T. Stammbach, R. Walter, '*R*-matrix Formulation and Phase Shifts for N-^{4}He Scattering for Energies up to 20 MeV', *Nucl. Phys. A*, **180**, 225–240 (1972).
24. J. Tirira, P. Trocellier, J.P. Frontier, P. Trouslard, 'Theoretical and Experimental Study of Low-energy ^{4}He-induced ^{1}H Elastic Recoil with Application to Hydrogen Behavior in Solids', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 203–207 (1990).
25. J. Tirira, F. Bodart, 'Alpha-proton Elastic Scattering Analyses up to 4 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **74**, 496–502 (1996).
26. G. Terwagne, G.G. Ross, L. Leblanc, 'Cross-section Measurements of the ^{1}H(^{3}He, ^{1}H)^{3}He Reaction at 20° and 30° between 1.9 and 3 MeV', *J. Appl. Phys.*, **79**, 8886–8891 (1996).
27. V. Quillet, F. Abel, M. Schott, 'Absolute Cross-section Measurements for H and D Elastic Recoil Using 1 to 2.5 MeV ^{4}He Ions and for the ^{12}C(d, p)^{13}C and ^{16}O(d, p_1)^{17}O Nuclear Reactions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **83**, 47–61 (1993).
28. A. Galonsky, R.A. Douglas, W. Haeberli, M.T. McEllistrem, H.T. Richards, 'Deuteron-Helium Differential Scattering Cross-section', *Phys. Rev.*, **98**, 586–599 (1955).
29. F. Besenbacher, I. Stensgaard, P. Vase, 'Absolute Cross-section for Recoil Detection of Deuterium', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 459–463 (1986).
30. J.A. Sawicki, 'Measurements of the Differential Cross-sections for Recoil Tritons in ^{4}He-^{3}T Scattering at Energies between 0.5 and 2.5 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 123–127 (1988).
31. B.L. Cohen, C.L. Fink, J.H. Degnan, 'Nondestructive Analysis for Trace Amounts of Hydrogen', *J. Appl. Phys.*, **43**, 19–25 (1972).
32. F.A. Smidt, A.G. Pieper Jr, 'Studies of the Mobility of Helium in Vanadium', *J. Nucl. Mater.*, **51**, 361–365 (1974).
33. J.A. Moore, I.V. Mitchell, M.J. Hollis, J.A. Davies, L.M. Howe, 'Detection of Low-mass Impurities in Thin Films

Using MeV Heavy-ion Elastic Scattering and Coincidence Detection Techniques', *J. Appl. Phys.*, **46**, 52–61 (1975).

34. J. Tirira, P. Trocellier, J.P. Frontier, 'Analytical Capabilities of ERDA in Transmission Geometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 147–150 (1990).
35. J. Tirira, P. Trocellier, J.P. Frontier, P. Massiot, J.M. Costantini, V. Mori, '3D Hydrogen Profiling by Elastic Recoil Detection Analysis in Transmission Geometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **50**, 135–139 (1990).
36. J. Tirira, J.P. Frontier, P. Trocellier, Ph. Trouslard, 'Development of a Simulation Algorithm for Energy Spectra of Elastic Recoil Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **54**, 328–333 (1991).
37. E. Kótai, 'Computer Methods for Analysis and Simulation of RBS and ERDA Spectra', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **85**, 588–596 (1994).
38. M. Mayer, 'SIMNRA Users Guide', Report IPP 9/113, Max-Planck-Institut für Plasmaphysik, Garching, 1997.
39. N.P. Barradas, K. Arstila, G. Battistig, M. Bianconi, N. Dytlewski, C. Jeynes, E. Kótai, G. Lulli, M. Mayer, E. Rauhala, E. Szilágyi, M. Thompson, 'Summary of ''IAEA intercomparison of IBA software''', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **266**, 1338–1342 (2008).
40. P. Trocellier, M. Le Pioufle, 'La Microsonde Nucléaire du Centre d'Etudes de Bruyéres-le-Châtel en Sciences des Matériaux', *Rev. Sci. Technique de la Défense*, **1**, 137–146 (1994).
41. O. Auciello, R. Kelly, *Ion Bombardment Modification of Surfaces*, Elsevier, New York, 1984.
42. A. Dunlop, F. Rullier-Albenque, C. Jaouen, C. Templier, J. Davenas, *Materials under Irradiation*, Transactions Technical Publications, Brookfield, 1993.
43. B.P. Doyle, R.D. Maclear, S.H. Connell, P. Formenti, I.Z. Machi, J.E. Butler, P. Schaaff, J.P.F. Sellschop, E. Sideras-Haddad, K. Baruth-Ram, '3D Micro-ERDA Microscopy of Trace Hydrogen Distributions in Diamond Using a 2-D-PSD with Event Reconstruction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **130**, 204–210 (1997).
44. M. Mosbah, R. Clocchiatti, J. Tirira, J. Gosset, P. Massiot, P. Trocellier, 'Study of Hydrogen in Melt Inclusions Trapped in Quartz with a Nuclear Microprobe', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **54**, 298–303 (1991).
45. R.J. Sweeney, V.M. Prozesky, K.A. Springhorn, 'The Quantitative Determination of H in Geological Materials (Silicates) by Elastic Recoil Detection Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **130**, 623–627 (1997).
46. V.M. Prozesky, C.L. Churms, J.V. Pilcher, K.A. Springhorn, R. Behrisch, 'ERDA Measurement of Hydrogen Isotopes with a $\Delta E - E$ Telescope', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **84**, 373–379 (1994).
47. D.K. Avasthi, D. Kabiraj, A. Bhagwat, G.K. Mehta, V.D. Vankar, S.B. Ogale, 'Simultaneous Detection of Light Elements by ERDA with Gas-ionization/Si $\Delta E - E$ Telescope', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **93**, 480 (1994).
48. J.P. Stoquert, G. Guillaume, M. Hage-Ali, J.J. Grob, C. Ganter, P. Siffert, 'Determination of Concentration Profiles by Elastic Recoil Detection with a $\Delta E - E$ gas Telescope and High Energy Incident Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **44**, 184–194 (1989).
49. H. Timmers, R.G. Elliman, G.R. Palmer, T.R. Ophel, D.J. O'Connor, 'The Development of a Facility for Heavy-ion Elastic Recoil Detection Analysis at the Australian National University', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 611–615 (1998).
50. W.M. Arnoldbik, C.T.A.M. de Laat, F.H.M.P. Habraken, 'On the Use of a $\Delta E - E$ Telescope in Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 832–835 (1992).
51. W.M. Arnoldbik, F.H.P.M. Habraken, 'Elastic Recoil Detection', *Rep. Prog. Phys.*, **56**, 859–902 (1993).
52. J.A. Davies, J.S. Forster, S.R. Walker, 'Elastic Recoil Detection Analysis with Heavy Ion Beams', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 594–602 (1998).
53. D. Plachke, H.D. Carstanjen, R.M. Emrick, H. Nolte, H. Huber, W. Assmann, R. Grötzschel, 'Channeling and Blocking Measurements on Quasicrystals', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 816–821 (1998).
54. R. Siegele, J.A. Davies, J.S. Forster, H.R. Andrews, 'Forward Elastic Recoil Measurements Using Heavy Ions', *J. Appl. Phys.*, **76**, 4524–4532 (1994).
55. W. Assmann, H. Huber, C. Steinhausen, M. Dobler, H. Glucker, A. Weidiger, 'Elastic Recoil Detection Analysis with Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **89**, 131–139 (1994).
56. R. Groleau, S.C. Gujrathi, J.P. Martin, 'Time-of-flight System for Profiling Recoiled Light Elements', *Nucl. Instrum. Methods Phys. Res.*, **218**, 11–15 (1983).
57. J.P. Thomas, M. Fallavier, D. Ramdane, N. Chevarier, A. Chevarier, 'High-resolution Depth Profiling of Light Elements in High Atomic Mass Materials', *Nucl. Instrum. Methods Phys. Res.*, **218**, 125–128 (1983).
58. H.J. Whitlow, G. Possnert, C.S. Petersson, 'Quantitative Mass and Energy-dispersive Elastic Recoil Spectrometry: Resolution and Efficiency Considerations', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **27**, 448–457 (1987).
59. G. Boudreault, R.G. Elliman, R. Grötzschel, S.C. Gujrathi, C. Jeynes, W.N. Lennard, E. Rauhala, T. Sajavaara, H. Timmers, Y.Q. Wang, T.D.M. Weijers, 'Round Robin: Measurement of H Implantation Distributions in Si by Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **222**, 547–566 (2004).

60. N.P. Barradas, N. Added, W.M. Arnoldbik, I. Bogdanović-Radović, W. Bohne, S. Cardoso, C. Danner, N. Dytlewski, P.P. Freitas, M. Jakšić, C. Jeynes, C. Krug, W.N. Lennard, S. Lindner, Ch. Linsmeier, Z. Medunić, P. Pelicon, R.P. Pezzi, C. Radtke, J. Röhrich, T. Sajavaara, T.D.M. Salgado, F.C. Stedile, M.H. Tabacniks, I. Vickridge, 'A Round Robin Characterisation of the Thickness and Composition of Thin to Ultra-thin AINO Films', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **227**, 397–419 (2005).

61. C. Kottler, M. Döbeli, F. Glaus, M. Suter, 'A Spectrometer for Low Energy Heavy Ion ERDA', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **248**, 155–162 (2006).

62. P. Reichart, G. Dollinger, A. Bergmaier, G. Datzmann, A. Hauptner, H.J. Körner, 'Sensitive 3D hydrogen microscopy by proton proton scattering', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **219–220**, 550–554 (2004).

63. S.S. Klein, P.H.A. Mutsaers, B.E. Fischer, 'Mass Selection and Depth Profiling by Coincident Recoil Detection for Nuclei in the Middle-mass Region', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **50**, 150–153 (1990).

64. H.C. Hofsäss, N.R. Parikh, M.L. Swanson, W.K. Chu, 'Depth Profiling of Light Elements Using Elastic Recoil Coincidence Spectrometry (ERCS)', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 151–156 (1990).

65. G.G. Ross, B. Terreault, G. Gobeil, G. Abel, C. Boucher, G. Veilleux, 'Inexpensive, Quantitative Hydrogen Depth Profiling for Surface Probes', *J. Nucl. Mater.*, **128/129**, 730–733 (1984).

66. G.G. Ross, L. Leblanc, 'Depth Profiling of Hydrogen and Helium Isotopes by Means of the ERD E × B Technique', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **62**, 484–492 (1992).

67. S.S. Klein, 'Separate Determination of Concentration Profiles for Atoms with Different Masses by Simultaneous Measurement of Scattered Projectile and Recoil Energies', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 464–468 (1986).

68. H.A. Rijken, S.S. Klein, L.J. Van Ijzendorn, M.J.A. de Voigt, 'Elastic Recoil Detection Analysis with High Energy Alpha Beams', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **79**, 532–535 (1993).

69. G. Dollinger, T. Faestermann, P. Maier-Komor, 'High Resolution Depth Profiling of Light Elements', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 422–427 (1992).

70. B. Terreault, J.G. Martel, R.G. Saint Jacques, J. L'Ecuyer, 'Depth Profiling of Light Elements in Materials with High Energy Ion Beams', *J. Vac. Sci. Technol.*, **14**, 492–500 (1977).

71. B.L. Doyle, N.D. Wing, 'The Sandia Nuclear Microprobe', SANDIA Report SAND82-2393, 1982.

72. E.B. Carter, G.E. Mitchell, R.H. Davis, 'Elastic Scattering of Alpha Particles by ^{12}C in the Bombarding Energy Range 10–19 MeV', *Phys. Rev.*, **133**, B1421–B1433 (1964).

73. M.K. Mehta, W.E. Hunt, R.H. Davis, 'Scattering of Alpha Particles by Oxygen II: Bombarding Energy Range 10–19 MeV', *Phys. Rev.*, **160**, 791–802 (1967).

74. H.A. Rijken, *'Detection Methods for Depth Profiling of Light Elements Using High Energy Alpha Particles'*, Thesis, University of Eindhoven, 1993.

75. R. Siegele, J.A. Davies, J.S. Forster, H.R. Andrews, 'Forward Elastic Recoil Measurements Using Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **90**, 606–610 (1994).

76. R. Siegele, H.K. Haugen, J.A. Davies, J.S. Forster, H.R. Andrews, 'Forward Elastic Recoil Measurements Using Heavy Ions', *J. Appl. Phys.*, **76**(8), 4524–4532 (1994).

77. B. Brijs, T. Sajavaara, S. Giangrandi, T. Janssens, T. Conard, K. Arstila, K. Nakajima, K. Kimura, A. Bergmaier, G. Dollinger, A. Vantomme, W. Vandervorst, 'The Analysis of a thin $SiO_2/Si_3N_4/SiO_2$ Stack: A Comparative Study of Low-energy Heavy-ion Elastic Recoil Detection, High-resolution Rutherford Backscattering and Secondary Ion Mass Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **249**, 847–850 (2006).

78. T. Sajavaara, B. Brijs, S. Giangrandi, K. Arstila, A. Vantomme, W. Vandervorst, 'Analysis of Thin High-k and Silicide Films by Means of Heavy Ion Time-of-flight Forward-scattering Spectrometry', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **249**, 292–296 (2006).

79. S. Giangrandi, B. Brijs, T. Sajavaara, H. Bender, F. Iacopi, A. Vantomme, W. Vandervorst, 'Irradiation-induced Damage in Porous Low-k Materials During Low-energy Heavy-ion Elastic Recoil Detection Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **249**, 189–192 (2006).

80. F.H.P.M. Habraken, 'Light Element Depth Profiling Using Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **68**, 181–189 (1992).

81. W. Assmann, H. Huber, Ch. Steinhausen, M. Dobler, H. Glückler, A. Weidinger, 'Elastic Recoil Detection Analysis with Heavy Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **89**, 131–139 (1994).

82. P. Sigmund, K. Winterborn, 'Small-Angle Multiple Scattering of Ions in the Screened Coulomb Region. I. Angular Distributions', *Nucl. Instrum. Methods*, **119**, 541–557 (1974).

83. E. Szilágyi, F. Paszti, 'Theoretical calculation of the depth resolution of IBA methods', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **85**, 616–620 (1994).

84. E. Szilágyi, L. Wielunski, F. Pászti, 'Theoretical Approximation of Energy Distribution of Elastically Recoiled Hydrogen Atoms', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 701–706 (1998).

85. G. Amsel, G. Battistig, A. L'Hoir, 'Small angle multiple scattering of fast ions, physics, stochastic theory and numerical calculations', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **201**, 325–388 (2003).

86. K. Arstila, T. Sajavaara, J. Keinonen, 'Monte Carlo Simulation of Multiple and Plural Scattering in Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **174**, 163–172 (2001).

87. K. Nordlund, 'Molecular dynamics for ion beam analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **266**, 1886–1891 (2008).

88. N.J. Chou, A.D. Marwick, R.D. Goldblatt, L. Li, G. Coleman, J.E. Heidenreich, J.R. Paraszczak, 'An Isotope Approach to Characterization of Microwave-Water-Plasma-Modified Polyimide Surfaces', *J. Vac. Sci. Technol., A*, **10**, 248–253 (1992).

89. P.F. Green, D.B. Adolf, L.R. Gilliom, 'Dynamics of Polystyrene/Poly(Vinyl Methyl Ether) Blends', *Macromolecules*, **24**, 3377–3382 (1991).

90. R.J. Composto, E.J. Kramer, D.M. White, 'Reptation in Polymer Blends', *Polymer*, **31**, 2320–2328 (1990).

91. Y. Wang, S.S. Mohite, L.B. Bridwell, R.E. Gield, C.J. Sofield, 'Modification of High-temperature and High-performance Polymers by Ion Implantation', *J. Mater. Res.*, **8**, 388–403 (1993).

92. Y. Wang, B. Huang, D. Cao, J. Cao, D. Zhu, K. Shen, 'Analysis of Hydrogen in Oxygen-doped Polysilicon by $^4He^+$ – H Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **84**, 111–113 (1994).

93. J.C. Barbour, H.J. Stein, O.A. Popov, M. Yoder, C.A. Outten, 'Silicon Nitride Formation from a Silane-Nitrogen Electron Cyclotron Resonance Plasma', *J. Vac. Sci. Technol., A*, **9**, 480–484 (1991).

94. M. Putkonen, T. Sajavaara, L. Niinisto, J. Keinonen, 'Analysis of ALD-processed Thin Films by Ion-beam Techniques', *Anal. Bioanal. Chem.*, **382**, 1791–1799 (2005).

95. G. Compagnini, L. Calcagno, G. Foti, 'Properties of Fully Implanted Si_xC_{1-x}: H Alloys', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **80/81**, 978–981 (1993).

96. R.G. Gordon, U. Riaz, D.M. Hoffman, 'Chemical Vapor Deposition of Aluminum Nitride Thin Films', *J. Mater. Res.*, **7**, 1679–1684 (1992).

97. H.Z. Wu, T.C. Chou, A. Mishra, D.R. Anderson, J.K. Lampert, 'Characterization of Titanium Nitride Thin Films', *Thin Solid Films*, **191**, 55–67 (1990).

98. M. Wang, K. Schmidt, K. Reichelt, H. Dimigen, H. Hübsch, 'The Properties of Titanium-containing Amorphous Hydrogenated Carbon Films', *Surf. Coat. Technol.*, **47**, 691–695 (1991).

99. D. Boutard, W. Möller, B.M.U. Scherzer, 'Hydrogen Isotope Exchange in Amorphous Hydrocarbon Layers: Plasma-deposited a-C:H and Ion Saturated Carbon', *Radiat. Eff. Defects Solids*, **114**, 281–289 (1990).

100. R.W. Paynter, G.G. Ross, C. Boucher, J.F. Pageau, B. Stansfield, 'Analysis of a-C/B:H Layers Deposited in TdeV', *J. Nucl. Mater.*, **196–198**, 553–557 (1992).

101. A. Raveh, L. Martinu, S.C. Gujrathi, J.E. Klemberg-Sapieha, M.R. Wertheimer, 'Structure–Property Relationships in Dual-frequency Plasma Deposited on Hard a-C:H Films', *Surf. Coat. Technol.*, **53**, 275–282 (1992).

102. G. Dollinger, A. Bergmaier, L. Goergens, P. Neumaier, W. Vandervorst, S. Jakschik, 'High Resolution Elastic Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **219**, 333–343 (2004).

103. M. Juppo, P. Alen, M. Ritala, T. Sajavaara, J. Keinonen, M. Leskelä, 'Atomic Layer Deposition of Titanium Nitride Thin Films Using Tert-butylamine and Allylamine as Reductive Nitrogen Sources', *Electrochem. Solid State Lett.*, **5**, C4–C6 (2002).

104. P. Alen, M. Juppo, M. Ritala, T. Sajavaara, J. Keinonen, M. Leskelä, 'Tert-butylamine and Allylamine as Reductive Nitrogen Sources in Atomic Layer Deposition of TaN Thin Films', *J. Mater. Chem.*, **17**, 107–114 (2002).

105. P. Alen, M. Juppo, M. Ritala, T. Sajavaara, J. Keinonen, M. Leskelä, 'Atomic Layer Deposition of Ta(Al)N(C) Thin Films Using Trimethylaluminum as a Reducing Agent', *J. Electrochem. Soc.*, **148**, G566–G571 (2001).

106. G.W. Arnold, H.R. Westrich, W.H. Casey, 'Application of Ion Beam Analysis (RBS and ERD) to the Surface Chemistry Study of Leached Minerals', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 542–546 (1992).

107. H.J. Matzke, A. Turos, 'Mechanisms and Kinetics of Leaching of UO_2 in Water', *Solid State Ionics*, **49**, 189–194 (1991).

108. C. Neelmeijer, R. Grötzschel, E. Hentschel, R. Klabes, A. Kolitsch, E. Richter, 'Ion Beam Analysis of Steel Surfaces Modified by Nitrogen Ion Implantation', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 242–249 (1992).

109. H.J. Whitlow, E. Johansson, P. Anders Ingemarsson, S. Hogmark, 'Recoil Spectrometry of Oil-additive-associated Compositional Changes in Sliding Metal Surfaces', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **63**, 445–450 (1992).

110. J. Rickards, R. Trejo-Luna, M.E. Ortiz, E. Andrade, E. Chavez, E.P. Zironi, H. del Castillo, M. Sanchez, 'Hydrogen Pick-up in Ion Nitrided Steels', *Mater. Sci. Technol.*, **9**, 536–537 (1993).

111. R. Lappalainen, E.L. Wang, J.P. Hirvonen, 'Application of Nuclear Methods to the Study of Transfer Wear of PTFE on Stainless Steel and Silicon', *Surf. Interface Anal.*, **16**, 385–389 (1990).

112. H. Ronkainen, J. Likonen, J. Koskinen, 'Tribological Properties of Hard-carbon Films Produced by the Pulsed

Vacuum Arc Discharge Method', *Surf. Coat. Technol.*, **54/55**, 570–575 (1992).

113. F.L. Freire, C.M. Lepienski, C.A. Achete, Jr 'Characterization by Rutherford Backscattering, Elastic Recoil and Nuclear Reaction Analysis of Near-surface Modifications of Glasses Submitted to a DC Potential', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **68**, 227–230 (1992).

114. F. Schiettekatte, G.G. Ross, A. Chevarier, N. Chevarier, A. Plantier, 'Ion Beam Induced Depth Profile Modification of H, D and He Implanted in Be, C and Si', *Nucl. Instrum. Methods Phys. Res., Sect. B.*, **132**, 607–619 (1997).

115. G.G. Ross, I. Richard, 'Influence of the Ion-Beam-induced Desorption on the Quantitative Depth Profiling of Hydrogen in a Variety of Materials', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 603–607 (1992).

116. M. Naitoh, H. Ohnishi, Y. Ozaki, F. Shoji, K. Oura, 'Hydrogen-induced Reordering of the Si-(111)-$\sqrt{3} \times \sqrt{3}$-Al Surface Studied by ERDA/LEED', *Appl. Surf. Sci.*, **60/61**, 190–194 (1992).

117. Y. Zhang, G. Possnert, H.J. Whitlow, 'Measurements of the Mean Energy-Loss of Swift Heavy Ions in Carbon with High Precision', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **183**, 34–47 (2001).

118. M.W. Ruckman, M. Strongin, W.A. Lanford, W.C. Turner, '^{15}N Hydrogen Depth Profiling Measurements of Candidate Superconducting Supercollider Beam Pipe Materials', *J. Vac. Sci. Technol., A*, **13**(4), 1994 (1995).

FURTHER READING

D.K. Avasthi, W. Assmann, H. Huber, H.D. Mieskes, H. Nolte, 'Online Monitoring of Ion Induced Modifications by ERDA Using a Large Area Position-sensitive Detector Telescope', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **142**, 117–121 (1998).

W. Bohne, J. Röhrich, G. Röschert, 'The New Time-of-flight ERDA Set-up at the HMI-Berlin', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **139**, 219–224 (1998).

J.A. Davies, J. Forster, S.R. Walker, 'High-energy Heavy Ion Beams in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **139**, 120–127 (1998).

M.B. Lewis, E.H. Lee, 'Residual Gas and Ion Beam Analysis of Ion-irradiated Polymers', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **61**, 457–465 (1991).

A.M. Stoneham, 'Radiation Effects in Insulators', *Nucl. Instrum. Methods Phys. Res., Sect. A*, **91**, 1–11 (1994).

Elemental Analysis by Isotope Dilution

T.M. Bahrainwala and Z.R. Turel
The Institute of Science, Mumbai, India

1 Introduction 1
2 History 2
3 Theory 2
3.1 Principle of Isotope Dilution Analysis 2
3.2 Instruments Used 4
4 Statistical Evaluation of the Method with Respect to Accuracy, Precision and Sensitivity of Isotope Dilution Analysis 5
4.1 Factors Affecting Precision 5
4.2 Factors Affecting Accuracy 5
4.3 Factors Affecting Sensitivity 5
5 Limitations of the Method 5
6 Application of Isotope Dilution Analysis to the Determination of Some Representative Elements 5
6.1 Boron 5
6.2 Carbon 6
6.3 Nitrogen 6
6.4 Oxygen 6
6.5 Phosphorus 6
6.6 Sulfur 6
6.7 Argon 6
6.8 Potassium 7
6.9 Calcium 7
6.10 Chromium 7
6.11 Iron 7
6.12 Cobalt 7
6.13 Copper 8
6.14 Zinc 8
6.15 Arsenic 8
6.16 Selenium 8
6.17 Silver 8
6.18 Cadmium 9
6.19 Indium 9
6.20 Antimony 9
6.21 Tellurium 9
6.22 Iodine 9
6.23 Lanthanum 9
6.24 Hafnium 10
6.25 Tungsten 10
6.26 Rhenium 10
6.27 Gold 10
6.28 Mercury 10
6.29 Thallium 11
6.30 Lead 11
6.31 Uranium 11
6.32 Plutonium 12
7 Application of Isotope Dilution Analysis to Some Multielemental Determinations 12
Acknowledgments 14
Abbreviations and Acronyms 14
References 14

The measurement of extremely small traces of materials, which cannot be determined by normal analytical techniques, becomes possible and even relatively rapid and simple if the element can be made radioactive by the action of neutrons or charged particles. Nuclear techniques possess exceptional advantages in having extraordinary sensitivity, selectivity in detection and determination, speed and simplicity.

One of the nuclear techniques applied in trace and ultratrace analysis is isotope dilution analysis (IDA) which was first introduced by Hevesy and Hobbie in 1932. The great advantage of the method lies in the fact that it can be used even when the procedure involves loss of material or does not allow a complete separation of the element or compound. The isotope dilution method has found several applications in the determination of geological ages, fission yields and other problems in nuclear physics and chemistry.

Various techniques involved in IDA such as substoichiometry, liquid chromatography (LC) and radiometric titration are discussed along with the application of this technique to the estimation of elements in complex matrices as reported in the literature.

1 INTRODUCTION

The measurement of extremely small traces of materials which cannot be determined by normal analytical techniques becomes possible and even relatively simple and rapid if the element can be made radioactive by the action of neutrons or charged particles.

Nuclear techniques possess exceptional advantages in having extraordinary sensitivity in detection and determination, speed and simplicity. The certainty in identification is excellent as it involves half-life measurement, type of disintegration and energy measurement of the emitted radiation, which are unique properties of an isotope. The sensitivity in many cases is as high as 10^{-10} g and in special cases it attains 10^{-12} g.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Isotopes are elements (nuclides) having the same atomic number and different mass number. They are classified as stable isotopes (SIs) and radioisotopes (RIs). In IDA, a RI is used as the 'marked' analyte. It has the same behavior as the analyte and is called a 'tracer', 'spike' or 'label'.

One of the nuclear techniques applied in trace and ultratrace analysis is IDA. In the isotope dilution method the isotopic composition of an element remains constant throughout a series of chemical reactions. A SI may be used instead of a radioactive isotope, in which case the concentration must be determined by mass spectrometry (MS) or by some other appropriate procedure. The great advantage of the method lies in the fact that it can be used even when the procedure involves loss of material or does not allow a complete separation of the element or compound. It is therefore particularly useful in the measurement of rare earths, fission products and complex mixtures of organic or biomedical compounds.

IDA and reverse isotope dilution analysis (RIDA) have found useful applications with SIs and radioactive isotopes. With the development of MS, and the availability of separated SIs and radioactive isotopes, the isotope dilution method has found several applications in the determination of geological ages, fission yields and in other problems in nuclear physics. Accuracies of a few per cent may be obtained for some elements that are present in concentrations as low as parts per billion (ppb) or parts per trillion (ppt).

2 HISTORY

The term IDA was first introduced in 1940 by Rittenburg and Foster. Hevesy and Hobbie[1,2] first introduced the method for determining traces of lead present in a mineral in about one part in 10^6; lead was measured by adding Ra D and then depositing PbO_2 anodically. Haissinsky[3] used a similar technique for measuring potassium in the presence of other alkali metals. Similarly, ^{35}S was used to analyze a mixture of benzyl derivatives of sulfides, sulfones and sulfoxides.[4] Other measurements reported in the literature during the initial stages of the development of the method include naphthalene in coal tar,[5] γ-hexachlorobenzene,[6] penicillin,[7] vitamin B_{12} and other organic compounds. In the case of penicillin, the molecule was labelled with ^{13}C.

A great advantage of IDA is that it does not require complete separation of the analyte from the sample. However, the quantity of the analyte separated from the matrix must be determined using an appropriate method, e.g. by gravimetry or spectrometry. This is a disadvantageous feature of the classical form of IDA. To overcome this problem, substoichiometric isotope dilution analysis (Subst-IDA) was proposed in 1958 by Ruzicka and Stary[8] where only the measurement of radioactivities is required. The procedure is rapid and easy, and effects of interfering substances are lowered.[8]

In 1974, the 'sub- and superequivalence (SSE) method' was introduced into IDA.[9] It does not need such a strict reliance on the stability constant as the Subst-IDA mentioned above. Various analytical techniques, such as solvent extraction, ion exchange and precipitation, have been employed for the separation.

3 THEORY

3.1 Principle of Isotope Dilution Analysis

The various types of IDA such as classical IDA, Subst-IDA and sub- and superequivalence isotope dilution analysis (SSEIDA) are explained below using mathematical equations. Their characteristics are also explained.

3.1.1 Classical Isotope Dilution Analysis

In this method, an unknown amount (x) of analyte in the sample is spiked with a RI (A = radioactivity; weight is negligible compared with the analyte in the sample). The specific activity of this analyte is $S_x = A/x$. To the sample, a known amount (y) of the analyte is added, so the specific activity of the mixture becomes $S_{x+y} = A/(x + y)$. An appropriate amount of reagent which reacts with the analyte is added next, and the pure product is separated. To remove any impurity from the matrix, purification must be carried out repeatedly. After that, the weight (m_{x+y}) and the radioactivity (a_{x+y}) of the product are measured. As the specific activity before and after the separation does not change, we have Equation (1):

$$\frac{A}{x+y} = \frac{a_{x+y}}{m_{x+y}} \tag{1}$$

The unknown quantity x to be determined is calculated from Equation (2):

$$x = \frac{Am_{x+y}}{a_{x+y} - y} \tag{2}$$

When the sample contains other components having properties similar to the analyte, it is difficult to obtain complete (100%) recovery of the pure analyte. In such a case, a yield correction would be made in the conventional analysis. However, this correction is not needed in IDA because only the specific activity needs to be known, as shown in the above equations. Therefore, complete separation and recovery are not required and, as a result, repeated purification steps may be made

to remove impurities. This is the major advantage of IDA. The weight m_{x+y} of the pure separated analyte must be determined by an appropriate method; however, this operation is troublesome and time-consuming. This situation is improved by Subst-IDA.

3.1.2 Substoichiometric Isotope Dilution Analysis

In this method, two samples containing an unknown amount (x) of the analyte are spiked ($S_x = A/x$) as in classical IDA, and standard solutions containing known amounts (y, $3y$) of the analyte are added to them. The specific activities of the diluted solutions become $S_{x+y} = A/(x+y)$ and $S_{x+3y} = A/(x+3y)$. To each, substoichiometric amounts (M) of the reagent are added, then the products are separated from unreacted species. Assuming that the quantities (m_{x+y}, m_{x+3y}) separated from each sample are the same, namely $m_{x+y} = m_{x+3y} = m$, as the reagent used reacted quantitatively with the reactants, and the radioactivities obtained as a_{x+y}, a_{x+3y}, Equations (3) and (4) hold good as the specific activity does not change before and after the separation:

$$\frac{A}{x+y} = \frac{a_{x+y}}{m} \tag{3}$$

$$\frac{A}{x+3y} = \frac{a_{x+3y}}{m} \tag{4}$$

From these equations, we obtain Equation (5):

$$x = \frac{y(3a_{x+3y} - a_{x+y})}{a_{x+y} - a_{x+3y}} \tag{5}$$

Equation (5) shows that x is obtained from the radioactivity without measurement of the weight m. This is a major advantage of Subst-IDA. Another point is that the amount of reagent used is substoichiometric compared with the quantity of the analyte and this leads to a decrease in consumption of the reagent by the interfering substances which react with reagent in the same way as the analyte, resulting in improvements in accuracy, precision and sensitivity.

The reagent must react with the analyte almost completely (nearly 100%) and be separated completely if the assumption $m_{x+y} = m_{x+2y}$ is to hold good. For this purpose, the stability constant of the product analyte and the reagent must be sufficiently large. Not many reagents satisfy this condition. This is a disadvantage of Subst-IDA, and to overcome this problem, SSEIDA was proposed.

3.1.3 Sub- and Superequivalence Isotope Dilution Analysis

The sample containing an unknown amount (x) of the spiked analyte (A = radioactivity, weight is negligible compared with x) is placed in each test tube in the first series. They are isotopically diluted by the addition of regularly increasing known amounts iy ($i = 0, 1, 2, \ldots$) of the analyte so that their specific activities become $A/(x+iy)$. In the second series, k times the unknown amount of the first series is placed in each test tube (kA = radioactivity, kx = amount of analyte); they are not isotopically diluted. All solutions in both series are brought to the same volume and same acidity by adding the appropriate amount of solvent.

A definite amount of the reagents is added to each of them. The quantity of reagent is not always necessarily substoichiometric and is allowed to be larger (superequivalent) than that of the analyte. After the separation of the products (first series, m_{x+iy}; and second series, m_{kx}), their radioactivities are measured (first series, a_{x+iy}; second series, a_{kx}). As the specific activities do not change before and after the separation, we have Equations (6) and (7)

$$\frac{A}{x+iy} = \frac{a_{x+iy}}{m_{x+iy}} \tag{6}$$

$$\frac{kA}{kx} = \frac{a_{kx}}{m_{kx}} \tag{7}$$

for the first and second series, respectively. From these equations, we obtain Equation (8):

$$\frac{a_{kx}}{a_{x+iy}} = \left(\frac{m_{kx}}{m_{x+iy}} \quad \left(\frac{1}{x} \quad iy + \frac{m_{kx}}{m_{x+iy}}\right.\right. \tag{8}$$

The quantities m_{x+iy} and m_{kx} generally depend on the concentration of the analyte in the solution, namely $(x+iy)/V$ and kx/V. However, if one sample (jth increment) in the first series has the same concentration (isoconcentration point) of the analyte as that in the second series, the degree of reaction with the reagent and the degree of separation must be the same. This means that Equation (9) holds:

$$m_{x+jy} = m_{kx} \tag{9}$$

In this special case (at the isoconcentration point), Equation (10) is derived from Equations (6) and (7):

$$\frac{a_{kx}}{a_{x+jy}} = k \tag{10}$$

Substituting Equations (9) and (10) into Equation (8), Equation (11) is obtained:

$$x = jy(k-1) \tag{11}$$

Thus, to determine x, it is necessary to find j. The value j is virtual and does not necessarily have to be an integer (the value of x is obtained by a graphical method). The number of samples of the second series is not necessarily

large, but several samples are generally used to obtain a reliable value by averaging.

As SSEIDA does not impose strict conditions on the reactivity and the quantity of the reagent, it has a wide range of choice of reagents compared with Subst-IDA. On the other hand, determination by classic IDA and Subst-IDA is possible using one or two samples (except for the multiple method), but more than four samples are necessary in SSEIDA to find the isoconcentration point. Therefore, this method demands more time and effort, and is not suitable for the analysis of small quantities of samples.

As is known from Equations (6) and (7), the effect of interfering substances is not negligible because their quantities in the matrix are different in the first and second series. Apart from the this disadvantage, SSEIDA seems to be an excellent method when sensitive, precise and accurate determination by Subst-IDA is difficult.

3.1.4 Stable Isotope Dilution Analysis

The principle of stable IDA is almost the same as that of radioisotope dilution analysis (radio-IDA). The difference is that radio-IDA makes use of the change of specific activity but stable IDA must depend on the change in the isotope ratio. A mass spectrometer is the only instrument sensitive enough to measure the isotope ratio, so this method is also called isotope dilution mass spectrometry (IDMS).

Two isotopes (1 and 2) are selected from the various isotopes comprising the analyte to monitor the isotope ratio. Their isotopic abundance in a sample containing an unknown amount (x) of the analyte are taken as $\theta_{x(1)}$, $\theta_{x(2)}$. To the sample, a known amount (y) of the analyte having an enriched SI of composition different from that of the sample is added, where the isotopic abundances of the SI are $\theta_{y(1)}$, $\theta_{y(2)}$, respectively. The isotope ratio, R_{x+y}, of 1 and 2 in the mixture after mixing with the diluent is calculated using Equation (12):

$$R_{x+y} = \frac{x\theta_{x(1)} + y\theta_{y(1)}}{x\theta_{x(2)} + y\theta_{y(2)}} \tag{12}$$

From this equation, we obtain Equation (13):

$$x = \frac{[\theta_{y(1)} - R_{x+y}\theta_{y(2)}]y}{R_{x+y}\theta_{x(2)} - \theta_{x(1)}} \tag{13}$$

As $\theta_{x(1)}$, $\theta_{x(2)}$, $\theta_{y(1)}$, $\theta_{y(2)}$ and R_{x+y} $\{= [m_{x+y}, \theta_{x+y(1)}] / [m_{x+y}, \theta_{x+y(2)}]\}$ of product m_{x+y}, which is separated from matrix elements, are determined by MS and y is known, x is calculated from Equation (13).

3.2 Instruments Used

3.2.1 Radioisotope Dilution Analysis

Some radio-IDA needs to use RIs, so instruments for measuring radioactivity are necessary. An appropriate instrument is selected depending on the kind and energy of radiation.

For example, a NaI(Tl) scintillation counter is commonly used to measure γ-rays from RIs such as ^{125}I and a liquid scintillation counter is used for low-energy β-rays from ^{3}H, ^{14}C, etc. A Geiger–Müller counter is used for high-energy β-rays from ^{204}Tl etc. A high-resolution HPGe semiconductor detector is able to measure γ-rays of various energies at the same time.

3.2.2 Stable Isotope Dilution Analysis

Mass spectrometers are used in stable IDA to measure isotopic ratios. The instrument mainly consist of three parts, for ionization, separation of ions and detection. For ionization, glow discharge, inductively coupled plasma (ICP) and corona discharge are suitable for solid, liquid and gaseous samples. MS combined with gas chromatography (GC) and LC are useful for preliminary separation and concentration of the analyte.

An example of Subst-IDA for the determination of iodine in common salts using ^{131}I as tracer is given below.

About 1 g, accurately weighed, of powdered salt sample was dissolved in 25 mL of doubly distilled water. Carrier solution (10–60 μg of iodide as NaI) and a suitable aliquot of ^{131}I were added and the whole system was warmed to about 50 °C. After cooling, 5 mL of tartaric acid and 2 mL of potassium iodate solutions were added. This mixture acts as a mild oxidizing agent and oxidizes only I^- to I_2 and leaves chloride as such. Initially an attempt to use H_2SO_4 instead of tartaric acid was made, but this resulted in large-scale oxidation of chloride, causing interference in the substoichiometric stage. The iodine liberated was immediately extracted into CCl_4. To the organic phase 0.1 mL of $AgNO_3$ solution (in substoichiometric amount) and 10 mL water were added. Typically about 15 μg or its integral multiple amounts of Ag were added for substoichiometric extraction. This gave approximately 80% extraction. After thorough mixing, a colloidal solution of AgI was formed in the aqueous phase, to which, after separation, 1 mL of sodium thiosulfate solution was added. This resulted in the formation of NaI in the aqueous phase, which was then counted on a 1.75×2.00 in a well-type NaI(Tl) detector.

When analyzing actual samples of common salts, however, a standard solution of NaCl with a known concentration of iodine (as NaI) was simultaneously analyzed according to the standard procedure. Finally, the specific activities of the standard and the sample

aliquots were compared and the concentration of iodine was calculated.

In order to ascertain the accuracy and precision of the method, several suitable mixtures of NaCl and NaI were analyzed. To a solution of nearly 1 g of NaCl, an aqueous solution containing 3–60 μg of NaI was added and its iodine content was determined.

Iodine is important as an essential trace element for humans, and its deficiency leads to goiter, a prominent disorder of the thyroid gland.[10,11] The important sources of iodine are foodstuffs and beverages that contain seafoods and common salt. In order to circumvent the deficiency of iodine in many parts of the world, several countries have launched a program of dietary supplementation. Even such programs have not always been successful in countering goiter in some countries.[12] In India, the consumption of seafood is very low and common salt is the only source of iodine. The Government of India has now launched a program for the consumption of iodized salts by the general public. Iodine has been determined by other methods also.[13–15]

4 STATISTICAL EVALUATION OF THE METHOD WITH RESPECT TO ACCURACY, PRECISION AND SENSITIVITY OF ISOTOPE DILUTION ANALYSIS

4.1 Factors Affecting Precision

The precision of the weighing of the separated analyte is important. However, with Subst-IDA weighing is not required.

Errors in counting radioactivity also affect precision. The measurement should last long enough to lower the error E to below 1% according to equation (14):

$$T = \frac{10\,000}{E^2 r} \tag{14}$$

where T = period of measurement and r = counting rate.

4.2 Factors Affecting Accuracy

The tracer and carrier (diluent) must behave in the same manner as the analyte in the sample in order to obtain a constant specific activity throughout the procedure. For this purpose, properties such as valency state of the element and the kind of reagent must be made uniform at an early stage of analysis. Isotopic equilibrium between the tracer and the analyte must be complete.

Loss of analyte (in unknown amount) may be caused during dissolution by adsorption on the container wall, vaporization, etc. When an unknown amount of an analyte is lost before dilution with the tracer, it strongly affects the accuracy of the method. Therefore, it is desirable to mix the tracer and diluent with the sample in the earliest stages of the operation. Loss after exchange does not affect the accuracy.

In each analysis, a simple clean room is desirable to prevent contamination from the reagents, apparatus, atmosphere, etc. which affects the accuracy. It must be ensured that the radioactivity comes from only one type of nuclide, i.e. the sample is radiochemically pure. When there is another element which emits radiation similar to that of the tracer in the sample, it affects the accuracy. In these cases, the tracer should be purified.

4.3 Factors Affecting Sensitivity

The volume of the reaction mixture should be minimized to increase the sensitivity. The presence of interfering ions affects the sensitivity, but the effect is less in Subst-IDA and SSEIDA, so the sensitivity of these methods is higher than that of classical IDA.

5 LIMITATIONS OF THE METHOD

The use of the proposed method for the determination of amounts of metals larger than 1 mg is not advantageous, because physicochemical analytical methods may be used whose precision is, under these conditions, better than the precision of radiometric methods, which is controlled by the statistical character of the radioactive decay. The sensitivity of the method is limited by two factors: (a) the organic reagent employed cannot be diluted infinitely and (b) at present, RIs of sufficiently high specific activity are not at hand for all cases. The latter limitation has been solved to a certain extent by the development of procedures for the preparation of carrier-free RIs.

6 APPLICATION OF ISOTOPE DILUTION ANALYSIS TO THE DETERMINATION OF SOME REPRESENTATIVE ELEMENTS

6.1 Boron

A method has been described for determining traces of boron in silicon.[16] The sample was dissolved in aqueous sodium hydroxide in the presence of ^{10}B tracer, and boron was separated by modification of Morrison and Rupp's method of electrolysis through a cation-exchange membrane. Boron was finally determined by making isotope dilution measurements with a mass spectrometer with a thermal ionization source. The method has the advantage that quantitative separation of boron was not

necessary. Amounts of boron down to 0.001 μg can be determined to within ±30%.

6.2 Carbon

The sensitivity and precision of field-desorption MS are greatly improved by using an isotopically labelled standard compound and averaging the m/z values for a large number (e.g. 100 or 210) of spectral scans.[17] Cationization involves the production of molecular ions with enhanced m/z values, e.g. when a mixture of glucose, [^{13}C] glucose and NaI was dissolved in water, and the solution was applied to a high-temperature field-desorption emitter, the spectrum contained peaks for the species $(M + Na)^+$. The use of such peaks, with correction for [^{13}C] glucose in the nominally unlabelled glucose, permits the determination of glucose with a mean error of ±7%. The method is considered to show promise for biomedical analysis.

Small amounts of carbon in metals have been determined by the isotope dilution method by using an omegatron.[18] From 0.4 ppm to 6% of carbon in steel, iron, copper, tin, nickel, tungsten or molybdenum or in tungsten carbide was determined by heating the sample (in the form of a thin ribbon) at 1400 K in an atmosphere containing a known concentration of ^{13}CO. After 3 h the concentration of ^{13}CO and ^{12}CO in this atmosphere were determined by MS using a dynamic mass spectrometer (omegatron). The coefficients of variation were <10%.

A method utilizing IDMS has been developed for the determination of carbon in sodium.[19] The details of the method are described for the determination of elemental carbon employing Van Slyke oxidation, although, by choice of spike material, other forms of carbon may be determined. Accuracy and precision in the ranges of 50 ± 10 to 150 ± 25 μg of carbon have been demonstrated. The determination was not affected by the loss of generated carbon dioxide or by contamination from any source except carbon.

6.3 Nitrogen

The method described has been applied to the determination of nitrogen in Nb, W, Mo, La, Ce, Re, Fe, Cr and Ni.[20] It is based on MS measurement of changes (effected by dilution with the nitrogen from the sample) in the isotopic composition of ^{15}N used as an indicator. Nb was used as carrier for the indicator (forming nitrides) for improving the isotopic exchange between indicator and sample. The sensitivity was 1 ppm and the precision was 3% for 5 g samples containing 10 ppm of nitrogen.

6.4 Oxygen

Oxygen has been determined in iron and steel by a modified isotope dilution technique.[21] Previous discrepancies in the determination of oxygen in steel by an isotope dilution and vacuum fusion technique were traced to adsorption of C^{18}O on deposits in the furnace. By treating the specimen at dull-red heat with a measured volume of oxygen enriched in ^{18}O and allowing it to cool before lowering it into the graphite crucible, good agreement was obtained between the two methods.

6.5 Phosphorus

In the method described, ^{32}P-labeled sodium phosphate was used to determine 10–50 ppm of P in 3–5 g samples of triuranium octaoxide.[22] The error was <3% and the reproducibility was satisfactory. The method depends on the precipitation of the P as molybdophosphate, which is converted into $MgNH_4PO_4$ for counting.

The use of the isotope dilution technique to eliminate problems associated with demands for a high load capacity in column-coupling capillary isotachophoresis was studied.[23] Determinations of phosphate present in a model mixture and in white wine served for the evaluation of this approach to quantitative analysis in isotachophoresis. A high selectivity of the analysis with acceptable accuracy and precision of the analysis of complex ionic mixtures in a short time are obvious advantages of this approach.

6.6 Sulfur

Adsorbed sulfate in soil was determined by adding a small quantity of carrier-free ^{35}S as sulfate to a solution in equilibrium with the soil.[24] The adsorbed sulfate was calculated using the isotopic dilution law from total added activity and the activity in the equilibrium solution. Total added activity and activity in the equilibrium fractions were determined by suspending soil in water overnight, adding a solution containing ^{35}S, diluted to an activity close to 6.7 kBq, shaking the suspension for 2 h and centrifugation at 9000 rpm for 10 min. One part of the supernatant was used for liquid scintillation analysis and mixed with Quick Safe A. The remaining supernatant was filtered through a 0.2 μm membrane filter and sulfate was determined using a Dionex 2000i ion chromatograph equipped with a AS4 anion-exchange separation column. Phosphate-extractable sulfate was extracted with 10 mM calcium dihydrogenphosphate. Part of the supernatant was filtered through a 0.2 μm filter and analyzed by ion chromatography as before. The two methods gave corresponding results for fresh soil but for air-dried soil the extraction method yielded approximately 25% more adsorbed sulfate than the ^{35}S method.

6.7 Argon

Radiogenic Ar was determined in rock samples by the method of isotope dilution with an $AlCl_3$ standard

containing ^{38}Ar.[25] After exposure to a neutron flux, 1 g of $AlCl_3$ contained $(4.48 \pm 0.14) \times 10^{-5}$ mL of ^{38}Ar. Rock samples mixed with 0.5 g of irradiated $AlCl_3$ were placed in a silica tube attached to a vacuum system, and the isotope content of the released Ar was determined in a mass spectrometer. The ^{40}Ar content of the rock sample was calculated with a correction for atmospheric Ar.

6.8 Potassium

Small amounts of K have been determined by IDA, involving an exhaustive chemical procedure.[26] A method for the determination of K in minerals difficult to decompose by chemical means is described. Procedures for beryl and tourmaline are given. A method has also been described for the mass spectrometric determination of K in minerals containing <1% of K. The results agree reasonably well with those obtained by chemical methods.

6.9 Calcium

The IDMS determination of low concentrations of Ca in minerals has been discussed.[27] The sample was decomposed with $HF-HClO_4$ after adding ^{42}Ca, then total Ca was separated from K and various other metals on a column of strongly acidic ion-exchange resin, complete removal of K being essential. The Ca content was calculated from the ratio of the peak intensity of ^{40}Ca to that of ^{42}Ca, as determined by MS. The dependence of the analytical error on the MS measurements is discussed. Calcium (0.15–0.004%) was determined in two feldspars and one lepidolite. The results (together with their precision) are reported.

An isotope dilution technique utilizing thermal ionization MS has been developed for the accurate determination of calcium in synthetic and serum samples at the 100 mg g^{-1} level.[28] Calcium was separated from a serum matrix by destruction of the organic matter with $HClO_4$ and HNO_3 followed by ion-exchange separation from interferences using AG50W-X8 100–200-mesh resin. An MS isotopic analysis procedure was developed using a $Ca(NO_3)_2$ solution deposited on Re sample filaments in a triple-filament ion source. The relative error between calculated and experimentally determined concentrations in synthetic calcium solutions was $\leq$0.1%. The 95% limit of error for a single analysis was approximately 0.2% for synthetic and serum samples. A comparison of the isotope dilution data with concurrently determined atomic absorption data from several clinical and independent laboratories is presented.

6.10 Chromium

A method has been described for the determination of chromium and its enriched SI in human urine by IDMS.[29] A volatile chelate was formed using trifluoroacetylacetone (TFA) and the fragment ions corresponding to $Cr(TFA)^{2+}$ in the m/z 356–360 region were monitored. The chelate was thermally stable and exhibited no memory effects when the isotope ratios changed. The detection limit for the method was 0.03 ng g^{-1} of Cr and the accuracy was verified using certified reference materials (CRMs) and by an independent method. The method was highly specific for chromium, owing to the combined properties of the chelating agent, chromatographic column and mass-specific detector. In addition to total chromium determinations, the method can also be used to quantitate enriched SIs of chromium used as metabolic tags in tracer experiments in human nutrition studies.

6.11 Iron

IDA by ion exchange has been used for the determination of traces of iron.[30] A theoretical evaluation of this method is given, from which the conditions for the determination of a number of metals can be predicted. The selectivity of the method was further increased by the use of masking and precipitating agents. The proposed method has been verified by the determination of iron using EDTA and Dowex 50 cation exchanger. The average precision of mean values for amounts of 10^{-7} g of Fe per 5 mL was 1.2%, for amounts of 10^{-8} g of Fe per 5 mL it was 4.0% and for amounts of 10^{-9} g of Fe per 5 mL it was 4.1%. Even large excesses of many metals did not interfere in the analysis. The method has been applied to the analysis of NaI(Tl) crystals. The procedure which has been developed is very simple, and is far more sensitive than colorimetry, activation analysis or other methods used up to now for determining traces of iron.

6.12 Cobalt

IDA has been combined with solvent extraction and spectrophotometry to produce an effective general method for the determination of constituents that will not extract quantitatively.[31] Combining isotope dilution with extraction of the cobalt–2-nitroso-1-naphthol complex into chloroform resulted in an accurate and reliable method for the determination of small amounts of cobalt in alloys of high copper content and in ingot irons. ^{60}Co was used as a radioactive tracer.

An isotope dilution method for the determination of cobalt employing ^{60}Co as the tracer and a spectrophotometric procedure based on the extraction of cobalt thiocyanate with methyl isobutyl ketone (MIBK) has

been developed.[32] The extraction was carried out under neutral or slightly basic pH conditions and this made the procedure virtually specific for cobalt. The method was suitable for the determination of cobalt at concentrations from a few parts per million upwards in biological materials, vitamin B_{12}, salts, metals, etc. When a well-type scintillation detector was used for measuring the radiation of ^{60}Co in liquid samples, the procedure was relatively simple and rapid. The results obtained in testing the behavior of a number of metals and the various stages of the extraction of cobalt thiocyanate with MIBK in the presence of ^{60}Co as tracer are also presented.

6.13 Copper

Copper has been determined in serum samples by radio-IDA and incomplete extraction.[33] On decomposition of serum from a venous blood sample, Cu was labeled with ^{64}Cu. Dithizone in CCl_4 was used for extracting Cu at substoichiometric levels. The original Cu concentration was calculated from Ruzicka's equation. The method is compared with a colorimetric method.

6.14 Zinc

A new separation method using ion exchange based on complexation in a resin matrix has been used for the substoichiometric determination of Zn by the isotope dilution method using a chelating agent-loaded resin.[34] A resin loaded with 8-quinolinol-5-sulfonic acid was also shown to be applicable to the selective preconcentration of chalcophile elements in natural water samples. An additional advantage of this resin was that the ion-exchange capacity can be varied according to the desired purpose and therefore substoichiometric separation may be achieved. The possibility of the determination of zinc with substoichiometric amounts of 8-quinolinol-5-sulfonic acid was investigated.

The comparative extractability of Zn with potassium salts of ethyl, propyl, butyl, pentyl and benzyl xanthates in the pH range 3.5–9.0 into chloroform has been studied, employing a sensitive and rapid substoichiometric radiochemical method.[35] The extent of reproducibility was tested in each case. The effect of associated ions on the extraction was studied. The amount of Zn present in the standard solutions was determined employing each xanthate separately. The Zn content present in geological water samples in and around Tirupati was determined by the method developed and compared with the values obtained by atomic absorption spectroscopy (AAS).

6.15 Arsenic

Arsenic has been determined in biological and environmental standard reference materials (SRMs) by substoichiometric solvent extraction using toluene-3,4-dithiol in benzene.[36] A radiochemical solvent extraction procedure has been developed for the determination of As(III) using ^{76}As tracer. It is based on the complexation of As(III) with toluene-3,4-dithiol at pH 2 and subsequent extraction in benzene. The effects of various parameters such as pH, time of equilibration, nature of solvent, quantitative character and interferences were studied. The method has been further developed into Subst-IDA for the determination of As at <1 mg levels and employed for the analysis of several environmental and biological SRMs from the National Institute of Standards and Technology (NIST) (USA), the International Atomic Energy Agency (IAEA) (Vienna) and NIES (Japan).

6.16 Selenium

Selenium has been determined in plant materials by a fluorimetric method employing ^{73}Se.[37] An exhaustive chemical procedure is described. A 0.05% solution of purified naphthalene-2,3-diamine in 0.1 M HCl (0.5 mL) was employed for the fluorimetric method. The fluorescence was compared against a reagent blank. The isotope dilution results were used to correct for losses during the combustion and extraction. Of 20 metal ions tested, serious interference was observed from Cr^{3+}, Sb^{3+} and Sn^{4+}. The standard deviation (SD) for the determination of 0.2 μg of Se was 0.04 μg.

The determination of Se by the SSE method of IDA involves the extraction of a test solution of Se labeled with ^{75}Se.[38] Sodiumdiethyldithiocarbamate was used as reagent and CCl_4 as a solvent. The activity was counted on an NaI(Tl) detector. The Se concentration was calculated from a calibration graph of concentration against the ratio A to A' where A = activity in the organic phase and A' = repeated measurement in a fresh aliquot. For 1.92 ppm of Se, the SD was 0.1371 ppm.

A method is described for the determination of 10^{-2} ppm levels of Se in organic samples.[39] The method consists of a micromodification of the spectrophotometric method employing 3,3′-diaminobenzidine as reagent for extracting selenium(IV), and incorporates an isotope dilution technique to compensate for the unavoidable losses that occur during the course of the determination of Se.

6.17 Silver

Isotope dilution electroanalysis has been used for the determination of silver at microgram levels.[40] The method consists in electrolysis at a controlled potential with a silver-wire cathode and a platinum wire anode and the measurement of the specific activity of the Ag deposited on the cathode. Two microelectrolysis cells were connected in series, one containing an amount

of inactive Ag (x) plus a known amount of ^{110}Ag (y) and the other containing the same amount (y) of ^{110}Ag. In the presence of Cu^{2+}, Ni^{2+} and Cd^{2+}, Ag may be determined in the range 10^{-3}–10^{-6} g mL^{-1} of electrolyte with a coefficient of variation of 2.3% at the 10^{-5} g mL^{-1} level.

6.18 Cadmium

Cadmium has been determined in SRMs.[41] The method involves four polyethylene samples with nominal Cd contents from 40 to 400 ppm, which were supplied as small wire cuts each of about 10 mg. Each sample was mixed for 1 h and bottled in 30 g units. Homogeneity was tested by taking 15 pieces of 60–250 μg from each of the four bottles and analyzing them by solid-sampling Zeeman-corrected AAS, with measurement at 326.1 nm. These results were tested statistically. Samples VDA-001, -002 and -003 gave normally distributed results and homogeneity factors and minimum representative sample weights (13–27 mg) were calculated. The fourth sample, VDA-004, was not normally distributed, with some microheterogeneity, but could still be used if sample weights ≥1 g were used. For certification, the Cd contents were determined by IDMS. Samples were oxidized with H_2SO_4–HNO_3–H_2O_2, the Cd was plated on to Pt and the deposit uncertainties were in the range of 1.2–12 ppm. Certified values were given and were confirmed.

The feasibility of isotope dilution inductively coupled plasma mass spectrometry (IDICPMS) has been established by determination of Cd in soil extracts from the ^{106}Cd to ^{111}Cd isotope ratio measured before and after addition of ^{106}Cd,[42] reasonable agreement with results by graphite furnace atomic absorption spectroscopy (GFAAS) being obtained. A program in Basic for individual and off-line isotope dilution calculations is appended.

6.19 Indium

Subst-IDA has been described employing an ion-exchange method for the determination of traces of indium.[43] A highly selective, substoichiometric method for the determination of indium has been developed. The procedure consists in the addition of a substoichiometric amount of EDTA to the test solution labeled with a known amount of radioindium, followed by separation of the complex formed from the excess of unreacted indium on a Dowex-50 column. Using a preliminary cupferron extraction, the procedure became highly selective. The average precision of the determination of indium for an amount of 1.1 μg in 10 mL was 1.6% in the presence of an excess of many metals. The maximum sensitivity of the method was about 5×10^{-5} μg per 10 mL.

6.20 Antimony

A rapid and selective substoichiometric radiochemical method has been developed for the determination of microgram amounts of antimony employing potassium ethyl xanthate as a reagent and chloroform as an extractant from sulfuric acid medium.[44] The effect of associated ions on the extraction was studied. The method developed was successfully applied to determine the antimony content in standard solutions and synthetic mixture with an average error of ±2.07%.

6.21 Tellurium

A radiometric method based on redox substoichiometry has been developed for the determination of tellurium.[45] The oxidation of tellurium(IV) to tellurium(VI) with potassium dichromate was employed as the substoichiometric reaction, followed by tributyl phosphate extraction of the unreacted tellurium(IV) for substoichiometric isolation. The oxidation of tellurium(VI) with dichromate under the substoichiometric conditions was incomplete. The chemical yield of tellurium(VI) produced by the oxidation of tellurium(IV) with dichromate was 83%. It was found that a tellurium content of 50–160 μg could be determined with an accuracy of ±1.3% by means of the comparison method.

6.22 Iodine

A Subst-IDA method was developed for the determination of iodine in different brands of common salt.[46] An aqueous salt solution containing ^{131}I tracer and NaI as carrier was oxidized by tartaric acid and KIO_3 and the liberated iodine was extracted with CCl_4. To the extract an aqueous solution of $AgNO_3$ was added in substoichiometric amount to obtain a colloidal solution of AgI. On adding sodium thiosulfate solution, the NaI so formed passed into aqueous solution, which was then counted. Several different brands of salt were analyzed. The method was especially suitable for the determination of microgram amounts of iodide in the presence of excess of chloride.

6.23 Lanthanum

The Subst-IDA method has been used for the determination of La in environmental samples using cupferron in chloroform.[47] Dried powder was fused with sodium peroxide and 99.9% scandium oxide as holdback carrier. The fusion mixture was allowed to melt with occasional stirring, the resulting cake was cooled in H_2O_2 and 6 M HCl before dilution with water. The ^{140}La tracer was treated with La carrier solution, water and 0.02 M aqueous cupferron. Ammonium buffer of pH 9.0 was added

and the resulting solution was shaken with chloroform for 2 min. After equilibration, a brown-red complex was formed with clear phase separation. A portion of the organic and aqueous phases were counted at the 487 and 1596 MeV photopeaks of ^{140}La using an NaI(Tl) well-type detector in conjunction with a multichannel analyzer (MCA). The calibration graphs were linear for 22–550 ng and 1.75–44 μg of La. The method was applied substoichiometrically to four cement dust particulate samples and SRMs of environmental importance; the relative SD was <8%. Interference masking and elimination methods are discussed.

6.24 Hafnium

Hafnium has been determined in Ze-4 and Ze-6 zirconium metal standards by stable IDA.[48] Hf values obtained for Ze-4 and Ze-6 standards were 77.4 ± 3.9 (tentative) and 166 ± 17 ppm, respectively.

Hafnium has been separated from zirconium by a chromatographic method and determined by IDA.[49] The method is based on the different stabilities of the sulfate complexes. The best separation was obtained by using 0.65 N H_2SO_4 for washing the cationic column. With 0.7 N H_2SO_4 the separation was significantly quicker, although not quantitative, but this did not interfere with the method of isotope dilution. The relative error was <1% of Hf in mixture containing approximately 10% of Hf, but higher if the Hf concentration was <3%. The method was used for the determination of Hf in eudialyte and Zr and also in mixed oxides of Zr and Hf of various concentrations.

6.25 Tungsten

An isotope dilution procedure using ^{185}W for the determination of W in high-alloy steels has been described.[50] Tungsten was determined spectrophotometrically with quinol in H_2SO_4 after homogeneous precipitation from HNO_3–H_2O_2. No further chemical separation was required for Mo–V–W steels, but for steels containing Nb, Ta and/or Ti, chemical separation of the elements that would interfere with subsequent spectrophotometric determination was necessary. The method was successfully tested on a number of National Bureau of Standards (NBS) steels to which various amounts of Ta and Ti were added.

6.26 Rhenium

Rhenium has been determined by a surface ionization isotope dilution method.[51] Down to 1 μg of Re can be determined by MS by heating (at 2200 °C) a tungsten filament on which the metal has been deposited. Re in 10 mg samples of Re–W alloys (3–25% of Re) has been determined by an isotope dilution method (with use of ^{187}Re) with a SD of ±2%.

A radiotracer method has been used for the determination of Re.[52] Radiometric correction and interpolation methods have been suggested for rhenium determination using a ^{188}Re (16.9 h) radiotracer. The radioactive isotope was obtained in an ^{188}Re generator consisting of a glass column filled with aluminum oxide with a parent isotope ^{188}W (69.4 days) firmly adsorbed. The isotope ^{188}Re formed was eluted from the generator with a 0.9% solution of sodium chloride. Isotopic equilibrium among stable and radioactive rhenium (in the form of perrhenate ions) was obtained just after mixing the two solutions. A scintillation detector with an NaI crystal was used for γ-counting.

Rhenium has been extracted with a tetraphenylphosphonium complex into dichloroethane. Liquid scintillation counting was used. By this method, 1–100 μg of rhenium(VII) in the analyzed sample was determined. The relative SD was 0.05–0.09% over the whole concentration range.

6.27 Gold

Gold has been determined by radioactivation analysis accompanied by the isotope dilution method.[53] The sample (containing M_x of the element to be determined) and the standard (containing M_s) were irradiated under identical conditions and then each was mixed with a large amount of carrier, an aliquot was separated and their activities A_x and A_s were measured. The value of M_x is expressed by $M_x = M_s A_x / A_s$. This relationship has been exemplified by the determination of Au (0.03–3.4 ppm) in metallic Cu with rhodamine B (CI Basic Violet 10) solution as extracting agent.

6.28 Mercury

IDA by solvent extraction has been employed for the highly selective determination of trace amounts of Hg.[54] Mercury has been determined in amounts of 10^{-6}–10^{-7} g mL^{-1} with an average precision of ±0.5% and in amounts of 10^{-8}–10^{-9} g mL^{-1} with an average precision of ±2.5%. The determination of smaller amounts of mercury was subject to large errors because of the lower specific activity of the radiomercury employed. In the analysis, even significant excesses of metals which also form extractable complexes with dithizone do not interfere. The method is very rapid and simple, because it consists of a single extraction of the analyzed solution and measurement of the activity of the extract. The procedure developed is far more precise and sensitive than the colorimetric method.

A new principle for the substoichiometric determination of traces of Hg by IDA has been reported.[55] The

method was based on the dithizone–carbon tetrachloride extraction of Hg from the test solution (e.g. a solution of an ore or a mineral water). The necessity for determining the chemical yield was avoided. The main advantage of the isotope dilution method is the use of low activities.

Malonic anilide has been employed for the substoichiometric extraction of Hg into a mixture of ether and MIBK and for the subsequent determinations in industrial effluents from the Durgapur area, W.B., India.[56]

A radiometric method based on Subst-IDA has been developed for the determination of trace amounts of Hg using α-thiopicolin-*o*-anisylamide as a substoichiometric reagent.[57] A preconcentration procedure has also been developed based on the adsorption of Hg(II) as a dithizone complex on microcrystalline naphthalene which, when used in conjunction with the proposed radiometric method, enhances its sensitivity and selectivity. As little as 0.2 μg of Hg could be determined in an aqueous phase of 80 mL. The method has been successfully applied to the determination of trace amounts of Hg in city waste incineration ash, zinc metal and cadmium oxide samples.

6.29 Thallium

Thallium has been determined by SSEIDA.[58] The general method has been applied to the determination of Tl. A constant substoichiometric amount of EDTA was used, and the unchelated Tl^{3+} was removed by precipitation with aqueous ammonia before an equal volume of the solution was removed for counting. Common anions did not interfere, and interfering cations were eliminated by extracting Tl(III) from 6 M HCl into diethyl ether.

Thallium has been determined using brilliant green by SSEIDA.[59] The 1:1 complex between $TlCl_4$ and the brilliant green (**1**) cation was extracted into benzene; this avoided the need for large excess of the reagent and yielded a stable colour with a high *E* value. In the subequivalence test, the aqueous medium, containing radioactive plus inactive Tl (carrier), was made 2.5 M in HCl, **1** was added in substoichiometric amount and the extraction was carried out for 9 min. The benzene extract was evaporated and the residue was counted with a Geiger–Müller tube. In the superequivalence variant, twice as much radioactive (but no inactive) Tl and more **1** was used. The amount of active Tl was found from the point of intersection of isotope dilution curves. For 0.5–105 μg of active Tl(III), the errors ranged from ±1.2 to 2.3%. Preliminary extraction of Tl into diethyl ether from 6 M HCl medium and washing the extract with 3 M HCl prevented interference from Fe(III), Ga(III) and In(III). Interference from Au(III) and Sb(V) was avoided by reducing these with metallic Cu before extraction. Sulfate, nitrate and acetate interfered in the formation of the Tl complex.

6.30 Lead

Ultramicro amounts of lead have been determined by IDMS.[60] After isotopic equilibration with approximately 0.5 μg of ^{208}Pb, the sample, e.g. pure water or snow (20–100 g), was mixed with silica gel (13 μg) and H_3PO_4 (127 μg) as stabilizing agents, and part of the mixture was loaded on to a rhenium filament for MS. The Pb concentration of the sample was calculated from the ratio of $^{206}Pb^+$ to $^{208}Pb^+$. The coefficient of variation was approximately 3% for several measurements of added ^{208}Pb (0.003–1.8 μg), 3–7% for river water (approximately 0.07 ppb) and 1.3% for mountain snow (0.86 ppb). The instrument produced 1 fA of ion current from each nanogram of added $^{208}Pb^+$, the detection limit being 0.1 aA. The method was also applicable to the analysis of reagents, e.g. ammonia solution, hydrochloric acid and perchloric acid.

The application of displacement and exchange reactions in IDA of Pb has been studied.[61] The efficiency of separation of metal ions combining a displacement or exchange reaction and a liquid–liquid extraction procedure was calculated as a function of parameters such as pH, the ratio of the concentration of reagent and analyte and the ratio of the concentration of reagent and interfering ions. The resulting equations can be used to calculate favorable conditions for multielement IDA.

6.31 Uranium

A procedure and instrumentation for the automatic measurement of isotopic ratios with a double collector magnetic mass spectrometer have been used for U-isotope analysis.[62] Essentially, measurement was achieved by discriminating in favor of the desired peak plateau for ion current integration during the high-voltage variation. The peak resolution was virtually unaffected by off-set drift in the electronic system. Before each ratio measurement, the off-set values were determined and digitally stored for final compensation when isotopic ratios were calculated with an arithmetic unit.

Traces of uranium have been determined in metallic sodium by IDA and a spectrophotometric method.[63] Sodium (5 g) was evaporated (within 5 h) at a pressure of 10^{-4} Torr at 350 °C in a stainless-steel crucible. There was no loss of U (approximately 1 μg) under these conditions. An isotope dilution method was used to determine 0.001–1 ppm of U with the use of an NBS standard isotope. For >1 ppm of U, the U was reduced with Zn amalgam to U(IV), which was determined spectrophotometrically with arsenazo III after extracting

Fe(III) with diethyl ether. No interference was caused by the small amount of Fe dissolved from the crucible.

6.32 Plutonium

Plutonium has been determined in a NIST SRM by IDA.[64] SRM 4351 human lung was acknowledged to be inhomogeneous. The value of a single analysis for $^{239}Pu + ^{240}Pu$ could lie within a wide range, but the accuracy of the analysis can be corroborated by comparing the corresponding $^{238}Pu/^{239}Pu + ^{240}Pu$ or $^{240}Pu/^{239}Pu$ values with the relationships between these ratios and the total plutonium concentration. The $^{238}Pu/^{239}Pu + ^{240}Pu$ ratio was reported by activity and the $^{240}Pu/^{239}Pu$ ratio by atoms in aliquots of the SRM 4351. The data give an experimental insight into the inhomogeneity of the material, augmenting the purely statistical indicators that have existed so far. An intercomparison exercise on $^{239}Pu + ^{240}Pu$ in SRM 4351 was carried out. The objectives were to evaluate the capabilities of the participating laboratories for Pu measurements in human tissues and to improve the characterization of the $^{239}Pu + ^{240}Pu$ concentration value for the SRM. The results were published, but the values of ^{239}Pu and of ^{240}Pu as distinct from $^{239}Pu + ^{240}Pu$ figures were not reported.

7 APPLICATION OF ISOTOPE DILUTION ANALYSIS TO SOME MULTIELEMENTAL DETERMINATIONS

IDMS has been used for the certification of SRMs.[65] Isotope dilution has been applied at the NBS, employing thermal ionization or spark source MS, to determine accurately the concentration of trace elements in natural materials. These techniques have been applied, in conjunction with separation by ion exchange, chelation by resins, electrodeposition and chemical extraction. The certification of a range of biological and environmental SRMs, including the contents of Pb and trace elements in water, 15 elements in coal and fly ash, U in bovine liver and Hg in water, is presented.

IDMS has been used for the trace analysis of microelectronically relevant heavy metals in high-purity titanium.[66] Because titanium is increasingly used in microelectronic devices, an IDMS method has been developed for the reliable determination of traces of U, Th, Cu, Pb, Cd, Cr, Ni and Fe in high-purity titanium primary materials. The measurement of isotope ratios was carried out with a thermal ionization quadrupole mass spectrometer using positive thermal ions formed by a single- or double-filament ion source, except for thorium, where an ICP mass spectrometer was applied. Different separation techniques (ion exchange, chromatography, extraction, electrolytic deposition, coprecipitation) were used for the trace element–matrix separation and for the specific isolation of the trace elements to be determined. The detection limits obtained were U, Th = 0.07, Cu = 1, Cd = 1.7, Ni = 4, Pb = 6 and Fe = 35 ng g^{-1}. Three titanium samples of different purity were analyzed with concentrations in the following ranges: U, Th = $<0.07 \times 10^{-3}$–0.09, Cd = <0.002–0.7, Cu, Ni, Pb, Cr = 0.01–30 and Fe = $7–6 \times 10^{3}$ ng g^{-1}. The IDMS results for one titanium sputter target were compared with those for two different laboratories which showed the urgent necessity for the application of independent and reliable analytical methods.

Multielemental IDA by means of radiometric titration for the determination of Cu and Pb has been discussed.[67] A theoretical concept was derived for multielement IDA in a liquid–liquid extraction system. The practical performance was based on radiometric titration. For this purpose, a prototype of a titration/extraction vessel with flow injection was constructed. For calculations of the initial concentrations of the elements of interest from the experimental data, a computer program was developed.

Niobium and tantalum have been quantitatively determined by IDA.[68] Preliminary separations of Nb and Ta from each other and from Ti and Zr can be made by precipitation with benzene–selenous acid and extraction with 8-hydroxyquinoline before determination by means of the radioactive isotopes ^{95}Nb and ^{182}Ta. Nb was separated from Zr in 1 M HCl containing 0.5% of citric acid and 1.5% of ammonium selenate and Ta from Zr by increasing the citric acid concentration to 1.5%. Nb was separated from Ta, Zr and Ti by extraction at pH 4.0 to 4.5 with 2% 8-hydroxyquinoline solution in chloroform.

Uranium–plutonium metallic spikes for IDMS accountancy measurements, preparation and characterization have been discussed.[69] The advantages were discussed of the use of a solid $^{235}U–^{239}Pu$ alloy rather than the conventional $^{233}U–^{242}Pu$ solution to provide tracers for the IDMS determination of U and Pu in spent nuclear fuel solutions for reprocessing. The compositions [all percentages (w/w)] U–0.5% Pu–50% Ti, U–0.5% Pu–25% each Nb and Zr and U–1.0% Pu–2.3% Nb were chosen on account of their good cold formability, and were prepared by inductive levitation melting, subsequently being cold-rolled to a thickness of approximately 0.1 mm. Metallography of these alloys showed no significant heterogeneity, and γ-ray spectrometry showed that the relative SD of the Pu distribution in portions composed of 0.1 g of U and 1 mg of Pu (the usual added masses) was 0.4% for the U–Pu–Nb–Zr alloy and 0.1–0.2% for the other two.

Sample preparation approaches have been discussed for the IDICPMS certification of reference materials.[70] The application of IDA to inductively coupled plasma

mass spectrometry (ICPMS) is discussed. The requirements for accurate sample preparation and instrumental analyte–matrix–plasma interaction for the attainment of precise isotope ratio measurements are presented. The method is illustrated by the determination of Mg, Cd, Mo and Pb in synthetic samples, Cu and Cd in Zn ore and Cu and Mo in sewage sludge. The integration of procedures into the certification program at NIST is described.

Simultaneous radiochemical determination of chemically similar elements by the concentration-dependent distribution method has been described.[71] The method with two substoichiometric systems can be characterized by the following features. The selectivity of the two systems must be different, and desirably each of them should react preferentially with a different element. It is sufficient to use only one tracer for the two systems, but the use of two tracers could be advantageous in some cases. The analysis is based on the fact that if phase distributions in the two systems occur under strictly standard conditions, a set of two distribution ratios corresponds to the two values of the unknown concentrations. The calculation of the unknown concentrations from the two distribution ratios can be carried out using (i) theoretically derived equations based on the equilibria involved, (ii) empirical or semiempirical relationships and (iii) calibration plots. The calibration plots can be used in any case; methods (i) and (ii) can be utilized under conditions required for sufficiently strict validity of the corresponding equations. The method was applied for the determination of Rb and Cs.

The partial isotope separation of U and Fe by means of solvent extraction has been discussed.[72] The separation of isotopes is of great practical importance and, hence, it is worthwhile to explore the possibility of isotope separation by means of the simple method of solvent extraction.

Subst-IDA has been applied for the determination of trace elements in liquid samples, of the carrier content in RI solutions and of concentrations of organic reagents.[73] Cu in mineral acids and in a ZnSe single crystal was determined by substoichiometric extraction with dithizone. Values of 1.8 and 0.018 ppm in nitric acid and of 1.4 and 0.44 ppm in ZnSe were obtained. Cu and P carrier contents in ^{64}Cu and ^{32}P solutions were determined by substoichiometric extractions with dithizone in carbon tetrachloride and with molybdic acid into MIBK in a series of solutions, adding various amounts of Cu and P carrier. An analogous method was applied for the analysis of dithizone and diethyldithiocarbamate solutions. The method was also applied to the determination of ^{60}Co radioactivity in environmental samples. The analytical results for water samples are described.

Molybdenum and nickel have been determined in SRMs using the internal normalization technique for high-accuracy IDA.[74] General exact equations and iteration techniques have been developed for internal normalization to eliminate the effect of thermal fractionation on isotope ratio measurements, and therefore IDA, by thermal ionization MS. The techniques were applicable to more than 20 elements, and have been extensively applied to the determination of Mo in ore concentrates (55% Mo) and silicate trace standards (50 and 500 ppm Mo). The SD of all internally corrected Mo isotope ratio measurements was 0.1%. The Mo sample size was 40 μg, but the normalization technique should apply to microgram and smaller samples with a more sensitive ion detection system. Procedures are described for the chemical separation of Mo from matrix interferences and for the MS determination of Mo. Application of the techniques to Ni in three pollution SRMs is described.

A new variation of IDA based on double labeling carried out on filter-paper has been described.[75] The determination of microgram and submicrogram amounts of silver (labeled with ^{110m}Ag) by precipitation with iodide (labeled with ^{131}I) and of Ca (labeled with ^{45}Ca) by precipitation with phosphate (labeled with ^{32}P) are given as examples. In the first example, the two radioelements (^{110m}Ag and ^{131}I) were measured by γ-spectrometry and in the second method (^{45}Ca and ^{32}P) by a β-absorption method. The results show the usefulness of the method.

Coprecipitation has been used as a means of separation in Subst-IDA.[76] It is shown with examples that by applying coprecipitation for separation and by using complexometric and oxidimetric or reductometric reactions as the substoichiometric reaction, a wide range of complexometric determinations of metal ions (e.g. with EDTA) and many types of redox determinations are opened up for further investigation. The method is simple, not time-consuming and offers many possibilities owing to the freedom to select a suitable precipitate.

A semiquantitative method of isotope dilution in the microgram and submicrogram ranges using the ring oven method has been described.[77] Radioactive tracer elements were added to the sample solution and separated by the ring oven technique. The separated ring of precipitate was counted and the original concentration determined from a calibration graph. Examples are cited for the determination of calcium, iron and phosphate.

Ca and Mg have been determined by isotope dilution at the microgram level.[78] Natural calcium contains 96.97% of ^{40}Ca and 2.06% of ^{44}Ca and the tracer isotope used contained 2.9% of ^{40}Ca and 97% of ^{44}Ca. Precautions were necessary to avoid contamination, e.g. from atmospheric dust and from the vessels used. A solution of known concentration (1 mL) mixed with a known amount of tracer isotope was evaporated under nitrogen to one drop, which was applied to the double Ta filament of the spectrometer and evaporated to dryness.

The residue was covered with ammonium oxalate (as ionization activator) by similar evaporation of three drops of a saturated solution, and the ratio of the two isotopes was determined. Amounts of approximately 1 μg of Ca could be determined readily with a SD of 2.5%. The sensitivity under the conditions used was approximately 0.01 μg. Determination of 10^{-4}–10^{-5} μg is considered possible. Interference due to ^{40}K can be eliminated by heating the filament to a temperature below that necessary for the determination of Ca. Natural Mg (78.6% of ^{24}Mg and 10.11% of ^{25}Mg) was determined by mixing $Mg(NO_3)_2$ solution (1 mL) with the tracer isotope (98.5% of ^{25}Mg) and BeO (1 mg) as activator, and applying the concentrated solution to a double rhenium filament. From solutions containing 30.34 and 3.034 μg mL^{-1} of natural Mg, recoveries were 30.44 ± 0.39 and 3.016 ± 0.086 μg mL^{-1}, respectively. Under suitable conditions, the measurement of approximately 10^{-4} μg was considered possible. Use of a double instead of a single filament increased the sensitivity approximately 100-fold.

Application to the determination of elements and chemical compounds by SI dilution has been reported.[79] A review is presented on the use of this technique for the determination of H, rare gases, C, N, O, S, alkali, alkaline earth and heavy metals, the actinium series and compounds containing these elements.

ACKNOWLEDGMENTS

Thanks are due to Dr J.P. Mittal, Director, Chemical Group, BARC and Dr D.D. Sood, Director, Radiochemistry and Isotope Group, BARC for some of the papers on IDA required for this work. Sincere thanks are due to Shri Taher for typing the manuscript in spite of his busy schedule.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
CRM	Certified Reference Material
GC	Gas Chromatography
GFAAS	Graphite Furnace Atomic Absorption Spectroscopy
IAEA	International Atomic Energy Agency
ICP	Inductively Coupled Plasma
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IDA	Isotope Dilution Analysis
IDICPMS	Isotope Dilution Inductively Coupled Plasma Mass Spectrometry
IDMS	Isotope Dilution Mass Spectrometry
LC	Liquid Chromatography
MCA	Multichannel Analyzer
MIBK	Methyl Isobutyl Ketone
MS	Mass Spectrometry
NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
radio-IDA	Radioisotope Dilution Analysis
RI	Radioisotope
RIDA	Reverse Isotope Dilution Analysis
SD	Standard Deviation
SI	Stable Isotope
SRM	Standard Reference Material
SSE	Sub- and Superequivalence
SSEIDA	Sub- and Superequivalence Isotope Dilution Analysis
Subst-IDA	Substoichiometric Isotope Dilution Analysis
TFA	Trifluoroacetylacetone

REFERENCES

1. G. Hevesy, R. Hobbie, 'Lead Content of Rocks', *Nature*, **128**, 1083 (1931).
2. G. Hevesy, R. Hobbie, 'Determination of the Lead Content of Rocks. Application of the Methods of Radioactive Indicators to Analytical Chemistry', *Z. Anal. Chem.*, **88**, 1 (1932).
3–7. M. Haissinsky, *Nuclear Chemistry and Its Applications*, Addison-Wesley, Reading, MA, 1964.
8. J. Ruzicka, J. Stary, *Substoichiometry in Radiochemical Analysis*, Pergamon Press, Oxford, 1958.
9. H. Yoshioka, K. Hasegawa, 'Radioanalytical Application of Sub- and Super-equivalence Method of Isotope Dilution Analysis', *J. Radioanal. Nucl. Chem.*, **160**, 417–423 (1992).
10. E.J. Underwood, *Trace Elements in Human and Animal Nutrition*, Academic Press, New York, 281–322, 1977.
11. A.S. Prasad, *Trace Elements and Iron in Human Metabolism*, John Wiley & Sons, Chichester, 66–75, 1978.
12. S.V. Neveda, U.P. Kulkarni, S.M. Rao, *Trace Analysis and Technological Development*, ed. M. Sankar Das, Wiley Eastern, New Delhi, 257–274, 1983.
13. I.M. Kolthoff, P.J. Elving, E.B. Sandell, *Treatise on Analytical Chemistry*, Part II, Section A, Interscience, New York, 371–372, 1961.
14. P. Allain, Y. Mauras, C. Douge, L. Jaunault, T. Delaforte, C. Beaugrand, 'Determination of Iodine and Bromine in Plasma and Urine by ICP-MS', *Analyst*, **813**, 115 (1990).
15. K. Han, W.F. Koch, K.W. Pratt, 'Improved Procedure for the Determination of Iodide by Ion-chromatography

with Electro-chemical Detection', *Anal. Chem.*, **731**, 59 (1987).
16. D.C. Newton, J. Sanders, A.C. Tyrrell, 'The Determination of Boron in Silicon by Isotope Dilution', *Analyst*, **85**, 870–877 (1960).
17. W.D. Lehmann, H.R. Schulten, 'Quantitative Field Desorption Mass Spectrometry. III. Stable Isotope Dilution and Cationisation', *Angew. Chem.*, **16**(3), 184–185 (1977).
18. O.Yu. Begak, V.V. Fedorov, 'Determination of Small Contents of Carbon in Metals by Isotope Dilution Method with Use of an Omegatron', *Zh. Anal. Khim.*, **32**, 624–626 (1977).
19. K.Y. Eng, R.A. Meyer, C.D. Bingham, 'Determination of Carbon in Sodium by Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **36**, 1832–1834 (1964).
20. K.B. Orlova, E.N. Vitol, 'Determination of Nitrogen in Metals by Isotope Dilution Method', *Zh. Anal. Khim.*, **20**, 694–699 (1965).
21. J.P. Burden, C.R. Masson, 'Modified Isotope Dilution Technique for the Determination of Oxygen in Iron and Steel', *J. Iron Steel Inst.*, **202**(1), 28–31 (1964).
22. F.W. Lima, L.T. Atalla, 'Isotope Dilution Method for the Determination of Phosphorus in Uranium Compounds at the ppm Level', *An. Acad. Bras. Cienc.*, **33**(2), 149–160 (1961).
23. D. Kaniansky, P. Rajec, P. Havasi, A. Svec, F. Macasek, 'Isotope Dilution Analysis in Column Coupling Isotachophoresis', *J. Radioanal. Nucl. Chem. Lett.*, **104**(3), 157–182 (1986).
24. E. Karlton, 'Determination of Adsorbed Sulfate with Isotope Dilution of Sulfur (^{35}S) with Calcium Dihydrogen Phosphate Extraction', *Commun. Soil Sci. Plant Anal.*, **25**(3–4), 207–214 (1994).
25. Yu.M. Artemov, K.G. Knorre, V.P. Strizhov, 'Isotope Dilution with Use of a Standard in the Solid Phase (Applied to Argon)', *Geokhimya*, **12**, 1155–1156 (1963).
26. Yu.A. Shukolyukov, I.I. Matreeva, 'Determination of Small Amounts of Potassium by Isotope Dilution Method', *Zh. Anal. Khim.*, **16**, 544–548 (1961).
27. K.G. Heumann, E. Kubassek, W. Schwabenbauer, 'Mass Spectrometric Isotope Dilution Determination of Low Concentration of Calcium in Minerals', *Fresenius Z. Anal. Chem.*, **287**(2–3), 121–127 (1977).
28. L.J. Moore, L.A. Machlan, 'High Accuracy Determination of Ca in Blood Serum by Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **44**, 2291–2296 (1972).
29. C. Veillon, K.Y. Patterson, 'Determination of Natural and Isotopically Enriched Chromium in Urine by Isotope Dilution Gas Chromatography/Mass Spectrometry', *Anal. Chem.*, **66**, 856–860 (1994).
30. J. Stary, J. Ruzicka, 'Isotope Dilution Analysis by Ion Exchange. I. Determination of Traces of Iron', *Talanta*, **8**, 775–781 (1961).
31. W.D. Ralph, T.R. Sweet, I. Mencis, 'Determination of Small Amounts of Cobalt by Isotope Dilution Analysis', *Anal. Chem.*, **34**, 1992–1994 (1962).
32. K.F. Sporek, 'Determination of Small Amounts of Cobalt Using Isotope Dilution with Cobalt-60', *Anal. Chem.*, **33**, 754–758 (1961).
33. C.A. Bizollon, 'Determination of Copper in Serum by Radioisotope Dilution and Incomplete Extraction', *Ann. Pharm. Fr.*, **23**, 461–472 (1965).
34. H. Akaiwa, H. Kawamoto, K. Ogura, 'Substoichiometric Determination of Zinc by the Isotope Dilution Method Using a Chelating Agent-loaded Resin', *Talanta*, **24**, 394–395 (1977).
35. P.C. Reddy, B. Rangamannar, 'Comparative Substoichiometric Extraction of Zinc with Potassium Salts of Ethyl, Propyl, Butyl, Pentyl and Benzyl Xanthates', *J. Radioanal. Nucl. Chem. Lett.*, **200**(3), 181–190 (1995).
36. N.L. Chutke, M.N. Ambulkar, R.G. Weginwar, A.N. Garg, 'Substoichiometric Isotope Dilution Analysis of Arsenic in Biological and Environmental Standard Reference Materials by Solvent Extraction Using Toluene-3,4-dithiol in Benzene', *J. Radioanal. Nucl. Chem. Art.*, **185**(1), 145–156 (1994).
37. P. Cukor, J. Walzcyk, P.K. Lott, 'Application of Isotope Dilution Analysis to the Fluorimetric Determination of Selenium in Plant Material', *Anal. Chim. Acta*, **30**, 473–482 (1964).
38. J. Lesny, J. Tolgyessy, E.H. Klehr, 'Sub and Super Equivalence Method of Isotope Dilution Analysis. IV. Determination of Selenium', *Radiochem. Radioanal. Lett.*, **28**(1), 77–82 (1977).
39. W.J. Kelleher, M.J. Johnson, 'Determination of Traces of Selenium in Organic Matter. Combined Spectrophotometric and Isotope Dilution Method', *Anal. Chem.*, **33**, 1429–1432 (1961).
40. J. Ruzicka, P. Benes, 'Isotope Dilution Electroanalysis. II. Microgram Scale Determination of Silver', *Collect Czech. Chem. Commun.*, **26**, 1784–1791 (1961).
41. J. Pauwel, A. Lamberty, P. De Bievre, K.H. Grobecker, C. Bauspiess, 'Certified Reference Materials for the Determination of Cadmium in Polyethylene', *Fresenius' J. Anal. Chem.*, **349**, 409–411 (1994).
42. S.E. Long, 'Isotope Dilution Inductively Coupled Mass Spectrometry', *UK At. Energy Res. Establ. Memo.*, AERE-M 3540, **16** (1986).
43. J. Ruzicka, J. Stary, 'Isotope Dilution Analysis by Ion Exchange. II. Substoichiometric Determination of Traces of Indium', *Talanta*, **11**, 691–696 (1964).
44. P.C. Reddy, B. Rangamannar, 'Substoichiometric Radiochemical Determination of Antimony with Potassium Ethyl Xanthate', *J. Radioanal. Nucl. Chem. Lett.*, **187**, 107–112 (1994).
45. J. Suzuki, H. Yoshioka, A. Ohya, T. Kambara, 'Radiometric Analysis Based on Redox Substoichiometry. I.

Principles and Application for the Determination of Tellurium', *J. Radioanal. Nucl. Chem., Art.*, **109**(1), 101–107 (1987).
46. V. Singh, A.N. Garg, 'Substoichiometric Isotope Dilution Analysis Method for the Determination of Iodine in Common Salts Using ^{131}I Tracer', *Analyst*, **119**, 1417–1419 (1994).
47. R.G. Weginwar, N.L. Chutke, A.N. Garg, 'Substoichiometric Isotope Dilution Analysis of La in Environmental Samples Using Cupferron in Chloroform', *Sci. Total Environ.*, **130**, 219–230 (1993).
48. F.W. Walker, 'Determination of Hf in Ze-4 and Ze-6 Zirconium Metal Standards by the Stable Isotope Dilution and Neutron Activation Method', *US At. Energy Comm. Rep.*, KAPL-M-FW-4, **26** (1961).
49. G.M. Kolosova, Y.P. Cheng, M.M. Senyarin, 'Chromatographic Separation of Hf from Zr and the Determination of Hf by Isotope Dilution', *Zh. Anal. Khim.*, **15**, 346–366 (1960).
50. G. Leliaert, J. Hoste, Z. Eeckhaut, 'Determination of Tungsten in High Alloy Steels by Isotope Dilution', *Recl. Trav. Chim. Pays-Bas*, **79**, 557–566 (1960).
51. R.L. Walker, 'Rhenium Analysis by Surface Ionisation Isotope Dilution Method', *Int. J. Mass Spectrom. Ion Phys.*, **10**(2), 207–209 (1972).
52. M.N. Belkova, A.M. Chekmarev, V.I. Shamaev, I.D. Troshkina, 'Determination of Rhenium by Radiotracer Methods of Analysis', *J. Radioanal. Nucl. Chem., Art.*, **173**(1), 3–8 (1993).
53. K. Kudo, N. Suzuki, 'Radioactivation Analysis Accompanied by the Isotope Dilution Method. I. Determination of Traces of Gold', *J. Chem. Soc. Jpn. Pure Chem. Sect.*, **86**, 1050–1053 (1965).
54. J. Ruzicka, J. Stary, 'Isotopic Dilution Analysis by Solvent Extraction. III. Highly Selective Determination of Trace Amounts of Hg', *Talanta*, **8**, 535–538 (1961).
55. J. Ruzicka, J. Stary, A. Zeeman, 'New Principle of Activation Analysis Separations. VI. Substoichiometric Determination of Traces of Hg by Activation Analysis and by Isotope Dilution', *Zh. Anal. Khim.*, **19**, 932–936 (1964).
56. J. Dutta, S. Basu, 'Substoichiometric Extraction of Mercury with Malonic Anilide', *J. Radioanal. Nucl. Chem. Lett.*, **200**, 491–495 (1995).
57. D. Sandhya, M.S. Subramanian, 'Determination of Mercury by Substoichiometric Isotope Dilution Analysis', *Radiochim. Acta*, **65**, 105–109 (1994).
58. V.R.S. Rao, P.C. Rao, G. Tataiah, 'Determination of Tl by the Sub and Super Equivalence Method of Isotope Dilution Analysis', *Radiochem. Radioanal. Lett.*, **29**(5), 261–266 (1977).
59. V.S.R. Rao, P.C. Rao, G. Tataiah, 'Determination of Tl Using Brilliant Green [C.I. Basic Green 1] by the Sub and Super Equivalence Method of Isotope Dilution Analysis', *Radiochem. Radioanal. Lett.*, **29**(5), 261–266 (1977).
60. M. Masayo, N. Seiji, F. Kiyoshi, H. Makoto, 'ID-MS Determination of Ultra-micro Amounts of Lead', *Bunseki Kagaku*, **26**(2), 81–85 (1977).
61. P.C.A. Ooms, U.A.Th. Brinkman, H.A. Das, 'Applications of Displacement and Exchange Reactions in Neutron Activation and Isotope Dilution Analysis', *J. Radioanal. Chem.*, **46**, 255–264 (1978).
62. W. Ehrfeld, D. Froehlich, O. Fromheim, U. Knapp, 'Automatic Isotope Ratio Measurements with a Double-collector Magnetic Mass Spectrometer', *Kernforschungszentrum Karlsruhe Ber.*, KFK-2344, **14** (1977).
63. T. Masao, M. Yuji, T. Yukio, O. Misao, K. Takuji, 'Determination of Traces of Uranium in Metallic Sodium', *Bunseki Kagaku*, **20**, 1085–1091 (1971).
64. D.S. Popplewell, W. McCarthy, G.J. Ham, 'The Isotopic Composition of Plutonium in the National Institute of Standards and Technology Standard Reference Material 4351, Human Lung', *Radioanal. Nucl. Chem.*, **172**(2), 299–304 (1993).
65. L.J. Moore, H.M. Kingston, T.J. Murphy, P.J. Paulsen, 'Use of Isotope Dilution Mass Spectrometry for Certification of Standard Reference Materials', *Environ. Int.*, **10**(2), 169–173 (1984).
66. B. Beer, K.G. Heumann, 'Trace Analysis of Microelectronically Relevant Heavy Metals in High Purity Titanium with Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **65**, 3199–3203 (1993).
67. P.C.A. Ooms, G.P. Leendertse, H.A. Das, U.A.Th. Brinkman, 'Multielement Isotope Dilution Analysis by Means of Radiometric Titration', *J. Radioanal. Chem.*, **67**(1), 5–14 (1981).
68. L.P. Alimarin, G.N. Bilimovich, 'Quantitative Determination of Niobium and Tantalum by Isotope Dilution Method', *Collect. Czech. Chem. Commun.*, **26**(1), 255–261 (1961).
69. J.M. Orea Rocha, C.D. Ingelbrecht, A.J. Criado Portal, 'Uranium–Plutonium Metallic Spikes for Isotope Dilution Mass Spectrometry Accountancy Characterization', *Nucl. Instrum. Methods Phys. Res., Sect. A*, **334**, 96–100 (1993).
70. E.S. Beary, P.J. Paulsen, J.D. Fassett, 'Sample Preparation Approaches for Isotope Dilution Inductively Coupled Plasma Mass Spectrometric Certification of Reference Materials', *J. Anal. At. Spectrom.*, **9**, 1363–1369 (1994).
71. M. Kyrs, J. Halova, 'Simultaneous Radiochemical Determination of Chemically Similar Elements by the Method of Concentration Dependent Distribution. I. Principles and Theory of the Method of Two Substoichiometric Systems', *J. Radioanal. Chem.*, **52**(1), 53–67 (1979).
72. R.H. Betts, J. Bron, 'A Discussion of Partial Isotope Separation by Means of Solvent Extraction', *Sep. Sci.*, **12**, 635–639 (1977).

73. K. Kudo, K. Kobayashi, 'Determination of Trace Elements in Radioactive Materials by Substoichiometric Isotope Dilution Analysis', *J. Radioanal. Chem.*, **53**(1–2), 163–172 (1979).
74. L.J. Moore, L.A. Machlan, W.R. Shields, E.L. Garner, 'International Normalization Techniques for High Accuracy Isotope Dilution Analysis – Application to Molybdenum and Nickel in Standard Reference Materials', *Anal. Chem.*, **46**, 1082–1089 (1974).
75. H. Weisz, V. Krivan, 'Isotope Dilution Analysis in the Micro and Submicro Range by Double Labelling and Precipitation on Paper', *Talanta*, **16**, 823–826 (1969).
76. H.L. Polak, 'Coprecipitation as a Means of Separation in Substoichiometric Isotope Dilution Analysis. I. General Considerations and Introduction', *J. Radioanal. Chem.*, **9**, 241–249 (1971).
77. H. Weisz, D. Klockow, 'Semi-quantitative Isotope Dilution Analysis in the Microgram and Sub-microgram Ranges by the Ring Oven Method', *Mikrochim. Ichnoanal. Acta*, **5**, 1082–1093 (1963).
78. R. Delwiche, M. Vosters, 'Determination of Calcium and Magnesium on the Microgram Scale by Isotope Dilution', *Bull. Soc. Chim. Belg.*, **73**(3–4), 255–267 (1964).
79. Z. Dobkowski, 'Stable Isotope Dilution Analysis. II. Application to the Determination of Elements and Chemical Compounds', *Chem. Anal. (Warsaw)*, **10**, 519–530 (1965).

Instrumental Neutron Activation Analysis: Gamma Lines Table

S. Mark and Z.B. Alfassi
Ben Gurion University, Beer Sheva, Israel

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation[a]	Gamma line intensity	Other lines (keV) (intensities, %)
100	Nd-150	Pm-151	28.40 h		2.53	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
100.1	Ta-181	Ta-182	115.0 d	2099.74	14.1	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
100.9	Eu-153	Eu-154m	46.0 m		25.0	31.8 (5.7), 35.8 (14), 40.9 (8.1), 41.5 (35), 47 (4.5), 68.2 (37)
101.8	Pd-110	Pd-111m	5.5 h	0.43	0.27	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
102.1	Te-130	Te-131m	30 h	0.676	7.91	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
102.3	Gd-160	Gd-161	3.70 m	16.832	14.0	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 165.2 (2.60), 283.6 (6), 314.9 (22.9), 360.9 (60.6), 480.1 (2.70)
102.5	W-182	W-183m	5.15 s		2.35	46.5 (5.9), 52.6 (6.7), 58.0 (36.5), 59.3 (63.5), 67.2 (21.5), 69.1 (5.46), 99.1 (8.6), 107.9 (18.2), 160.5 (4.9)
103.1	Se-80	Se-81m	57.3 m	3.968	9.79	275.9 (1.3), 290.1 (1.0), 552.4 (0.15), 566 (0.38), 828.3 (0.47)
103.2	Sm-152	Sm-153	46.7 h	5500.2	28.3	40.9 (17.4), 41.5 (31.3), 47.0 (9.57), 48.3 (2.74), 69.7 (5.25)
103.2	Gd-152	Gd-153	242 d	220	21.8	40.9 (34), 41.5 (61.5), 47.0 (18.7), 48.3 (5.40), 69.7 (2.34), 83.4 (0.21), 97.4 (30.1)
104.3	Sm-154	Sm-155	22.1 m	124.85	75.1	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 141.4 (2.03), 245.7 (3.76)

[a] Probability of formation = cross-section of formation × natural abundance of appropriate isotope.

(*continued overleaf*)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
103.2	Gd-152	Gd-153	242 d	220	21.8	40.9 (34), 41.5 (61.5), 47.0 (18.7), 48.3 (5.40), 69.7 (2.34), 83.4 (0.21), 97.4 (30.1)
104.3	Sm-154	Sm-155	22.1 m	124.85	75.1	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 141.4 (2.03), 245.7 (3.76)
104.5	Yb-176	Yb-177m	6.41 m		76.6	51.4 (21.2), 52.4 (37.5), 59.3 (12.2), 61.0 (3.26), 227.0 (12.2)
104.7	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
104.8	Nd-150	Pm-151	28.40 h		3.52	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
105.3	Sm-154	Eu-155	4.68 yr		21.8	6.3 (8.5), 42.3 (6.97), 43.0 (12.5), 48.7 (3.87), 86.5 (32.7)
105.4	Hf-176	Hf-177m$_1$	1.08 s		16.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
105.4	Lu-176	Hf-177m$_1$	1.08 s		16.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
105.4	Lu-176	Lu-177m	160.9 d	18.2	12.2	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
105.4	Hf-176	Hf-177m$_2$	51.4 m		16.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
105.9	Mo-100	Mo-101	14.6 m	1.916	0.25	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
105.9	Re-187	Re-188m	18.6 m	100.16	10.8	59.7 (18), 61.1 (31), 63.6 (21.3), 69.2 (11), 71.2 (2.7), 92.5 (5.15), 156.0 (0.6), 169.5 (0.1)
107.9	W-182	W-183m	5.15 s		18.2	46.5 (5.9), 52.6 (6.7), 58.0 (36.5), 59.3 (63.5), 67.2 (21.5), 69.1 (5.46), 99.1 (8.6), 102.5 (2.35), 160.5 (4.9)
108.1	Th-232	Th-233	22.3 m	740	0.36	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
108.2	Dy-164	Dy-165m	1.26 m	47 940	3.01	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 153.8 (0.242), 361.7 (0.534), 515.5 (1.53)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
108.5	Ba-130	Ba-131m	14.6 m	0.265	55.2	31.8 (13.7), 32.2 (25.4), 36.4 (7.5), 37.3 (1.79), 79 (1.19)
108.8	Ru-96	Ru-97	2.9 d	1.6	0.11	18.3 (58.5), 20.7 (11.7), 215.7 (85.8), 324.5 (10.2), 460.6 (0.12), 569.3 (0.88)
109.3	Te-124	Te-125m	58.0 d	0.192	0.330	27.2 (33.8), 27.5 (63), 31 (18.2), 31.7 (3.68), 35.5 (6.66)
109.8	Yb-168	Yb-169	32.02 d		18.0	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 118.2 (1.93), 130.5 (11.5), 177.2 (22), 198 (36), 307.7 (11.1)
109.8	Er-168	Er-169	9.30 d	52.26	0.0013	8.4 (0.017), 49.8 (0.0036), 50.7 (0.0064), 57.5 (0.0021), 59.1 (0.0006), 118.2 (0.001)
109.9	O-18	O-19	27.1 s	0.000032	2.54	197.1 (95.9), 1356.8 (50.4), 1444.1 (2.64), 1554 (1.39)
111	Xe-124	Xe-125m	57 s	2.8	61.8	29.5 (17.9), 29.8 (33.2), 33.6 (9.75), 34.4 (2.14), 141 (19.6)
111.5	Th-232	Th-233	22.3 m	740	0.11	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
111.6	Er-170	Er-171	7.52 h	84.93	20.5	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 116.7 (2.30), 124.0 (9.10), 295.9 (28.9), 308.3 (64.4)
111.7	Ir-193	Ir-194m	171 d	376.2	8.85	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
112.5	Ga-71	Ga-72	14.1 h	187.929	0.12	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
113	Hf-176	Hf-177m_2	51.4 m		27.3	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
113	Lu-176	Lu-177m	160.9 d	18.2	21.7	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
113	Hf-176	Hf-177m_1	1.08 s		27.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
113	Lu-176	Lu-177	6.71 d	5460	6.60	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 208.4 (11), 249.7 (0.21), 321.3 (0.22)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
113	Lu-176	Hf-177m$_1$	1.08 s		27.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
113	Yb-176	Lu-177	6.71 d		6.60	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 208.4 (11), 249.7 (0.21), 321.3 (0.22)
113.8	Yb-174	Yb-175	4.19 d	2067	1.93	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 137.7 (0.12), 144.9 (0.33), 282.5 (3.08), 396.3 (6.55)
113.8	Hf-174	Hf-175	70 d	62.4	0.31	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 229.6 (0.77), 318.9 (0.17), 343.4 (86.6), 353.6 (0.23), 432.8 (1.56)
114	Mo-92	Mo-93m	6.85 h	0.089	0.68	17.4 (18.8), 19.7 (3.7), 263.1 (56.7), 684.7 (99.7), 949.8 (0.12), 1363.0 (0.79), 1477.1 (99.1)
114.3	Nd-148	Nd-149	1.73 h	14.4	19.0	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
116.7	Er-170	Er-171	7.52 h	84.93	2.30	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 124.0 (9.10), 295.9 (28.9), 308.3 (64.4)
116.8	Nd-150	Nd-151	12.4 m	6.768	45.9	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
118.2	Yb-168	Yb-169	32.02 d		1.93	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 130.5 (11.5), 177.2 (22), 198 (36), 307.7 (11.1)
118.2	Er-168	Er-169	9.30 d	52.26	0.001	8.4 (0.017), 49.8 (0.0036), 50.7 (0.0064), 57.5 (0.0021), 59.1 (0.0006), 109.8 (0.0013)
121.1	Se-74	Se-75	119.8 d	46.62	17.3	66.1 (1.14), 96.7 (3.48), 136 (59), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
121.3	Nd-146	Pm-147	2.62 yr		0.0028	39.5 (0.0006), 40.1 (0.0011), 45.4 (0.0003), 46.6 (0.0001), 76.2 (0.0001), 197.4 (0.0001)
121.5	Zn-70	Zn-71g	2.4 m	0.0514	3.0	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
121.5	Zn-70	Zn-71m	3.94 h	0.005	2.9	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
121.6	Lu-176	Lu-177m	160.9 d	18.2	5.87	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
121.6	Yb-176	Yb-177	1.9 h		3.4	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 150.4 (20), 1080.1 (5.5), 1241.4 (3.4)
121.8	Eu-151	Eu-152	13.33 yr	2 82 020	28.4	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
121.8	Eu-151	Eu-152m$_1$	9.32 h	1 57 740	7.16	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 344.3 (2.44), 841.6 (14.5), 963.4 (11.9)
122.7	Hf-178	Hf-179m$_2$	25.1 d	867.2	26.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
122.7	Re-185	Re-186	3.78 d	4188.8	0.72	58.0 (1.55), 59.3 (2.69), 61.5 (1.11), 63.0 (1.92), 67.2 (0.91), 69.1 (0.23), 71.3 (0.66), 73.4 (0.17), 122.7 (0.72), 137.2 (9.20), 767.5 (0.029)
123.1	Eu-153	Eu-154	8.8 yr		40.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
123.8	Ba-130	Ba-131	11.8 d	0.9328	29.2	30.6 (27.9), 31 (51.6), 35 (15.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
124	Er-170	Er-171	7.52 h	84.93	9.10	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 116.7 (2.30), 295.9 (28.9), 308.3 (64.4)
124.6	Xe-126	Xe-127m	69.2 s	0.0405	69.1	29.5 (15.4), 29.8 (28.6), 33.6 (8.43), 34.4 (1.85), 172.5 (37.9)
125.4	W-184	W-185	75.1 d	55.206	0.019	59.7 (0.015), 61.1 (0.025), 69.2 (0.008), 71.2 (0.002)
126.1	Cr-54	Cr-55	3.5 m	0.8496	0.0017	1528 (0.037), 2252.5 (0.0031)
127.2	Mo-100	Tc-101	14.2 m		2.86	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627 (0.39), 715.6 (0.69), 720 (0.23), 842.8 (0.23), 928.6 (0.13)
127.5	Cs-133	Cs-134m	2.91 h	250	12.6	11.2 (0.94), 30.6 (8.95), 31.0 (16.5), 35.0 (4.88), 35.8 (1.10), 138.7 (0.0039)
128.5	Lu-176	Hf-177m$_1$	1.08 s		20.2	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
128.5	Hf-176	Hf-177m_2	51.4 m		20.2	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
128.5	Hf-176	Hf-177m_1	1.08 s		20.2	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
128.5	Lu-176	Lu-177m	160.9 d	18.2	15.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
129.4	Os-190	Os-191	15.4 d	102.96	25.7	63.3 (16), 64.9 (27.6), 73.5 (9.51), 75.6 (2.45)
129.4	Pt-190	Ir-191m	4.94 s		25.7	41.8 (0.0056), 63.3 (16), 64.9 (27.6), 73.5 (9.51), 75.6 (2.45), 82.4 (0.02)
129.4	Pt-190	Pt-191	2.90 d	1.95	3.2	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
129.4	Os-190	Ir-191m	4.94 s		25.7	41.8 (0.0056), 63.3 (16), 64.9 (27.6), 73.5 (9.51), 75.6 (2.45), 82.4 (0.02)
129.6	Ru-104	Rh-105m	45 s		20.0	20.2 (34.5), 22.8 (7.14)
129.6	Ru-104	Ru-105	4.44 h	5.984	5.6	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
129.7	Kr-78	Kr-79m	50 s	0.0595	27.5	12.6 (31.8), 14.1 (5.48)
129.7	Pt-194	Pt-195m	4.02 d	2.961	2.83	30.9 (2.28), 65.1 (22.6), 66.8 (38.7), 75.7 (13.3), 77.9 (3.51), 98.9 (11.4)
129.8	Sr-84	Sr-85m	67.66 m	0.336	0.15	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 151.2 (12.8), 231.9 (84.4), 238.8 (0.275)
129.8	Kr-84	Kr-85m	4.48 h	5.13	0.30	12.6 (3.39), 13.4 (1.80), 14.1 (0.58), 15.0 (0.32), 151.2 (75), 304.9 (14)
130.2	Hg-196	Au-197m	7.80 s		3.12	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 201.8 (1.1), 279.0 (70.9), 409.1 (0.11)
130.2	Pt-196	Au-197m	7.80 s		3.12	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 201.8 (1.1), 279.0 (70.9), 409.1 (0.11)
130.2	Pt-196	Pt-197m	1.57 h	1.265	0.10	53.0 (1.08), 65.1 (13.6), 66.8 (23.8), 68.8 (0.40), 75.7 (8), 77.9 (2.26), 279.0 (2.3), 346.5 (11.2)
130.4	Hg-196	Hg-197m	23.8 h	18	0.22	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 133.9 (34.1), 165.0 (0.27), 279.0 (5.0)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
130.5	Yb-168	Yb-169	32.02 d		11.5	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 177.2 (22), 198 (36), 307.7 (11.1)
131.5	W-184	W-185m	1.67 m	0.061	4.34	9 (12), 58.0 (2.2), 59.3 (3.9), 65.9 (5.82), 173.7 (3.30)
133	Hf-180	Hf-181	42.4 d	443.52	43.0	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 136.2 (6.1), 345.9 (14.0), 482.0 (85.5)
133.6	Ba-130	Ba-131	11.8 d	0.9328	2.17	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
133.9	Hg-196	Hg-197m	23.8 h	18	34.1	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 130.4 (0.22), 165.0 (0.27), 279.0 (5.0)
134.2	W-186	W-187	23.9 h	1081.08	10.3	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 479.5 (25.3), 551.6 (5.89), 618.4 (7.27), 685.8 (31.6), 772.9 (4.77)
135.5	Pt-192	Pt-193m	4.33 d	1.738	0.11	12.6 (0.66), 65.1 (4.30), 66.8 (7.38), 75.7 (2.53), 77.9 (0.67)
136	Kr-78	Kr-79g	35.04 h	2.1	1.0	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
136	Se-74	Se-75	119.8 d	46.62	59.0	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
136	Ge-74	Ge-75m	47.7 s	6.205	0.020	9.9 (23.6), 11.0 (3.47), 61.9 (0.12), 139.7 (38.8)
136.2	Hf-180	Hf-181	42.4 d	443.52	6.1	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 133.0 (43.0), 345.9 (14.0), 482.0 (85.5)
137.2	Re-185	Re-186m	2.0×10^5 yr	12.716	9.20	8.6 (16.0), 10.0 (25.0), 11.7 (3.8), 59.0 (18.6), 59.3 (2.96), 63.0 (1.92)
137.2	Re-185	Re-186	3.78 d	4188.8	9.20	58.0 (1.55), 59.3 (2.69), 61.5 (1.11), 63.0 (1.92), 67.2 (0.91), 69.1 (0.23), 71.3 (0.66), 73.4 (0.17), 122.7 (0.72), 137.2 (9.20), 767.5 (0.029)
137.4	Br-81	Br-82g	35.30 h	12.8206	0.14	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
137.7	Yb-174	Yb-175	4.19 d	2067	0.12	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 144.9 (0.33), 282.5 (3.08), 396.3 (6.55)
138.3	In-115	In-116m_1	54.15 m	7753.32	3.29	416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
138.7	Cs-133	Cs-134m	2.91 h	250	0.0039	11.2 (0.94), 30.6 (8.95), 31.0 (16.5), 35.0 (4.88), 35.8 (1.10), 127.5 (12.6)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
138.9	Nd-150	Nd-151	12.4 m	6.768	8.72	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
138.9	Os-192	Os-193	30.5 h	82	4.27	63.3 (3.74), 64.9 (6.46), 73.0 (3.2), 73.5 (2.22), 460.5 (3.95)
139.7	Ge-74	Ge-75m	47.7 s	6.205	38.8	9.9 (23.6), 11.0 (3.47), 61.9 (0.12), 136.0 (0.020)
140.5	Mo-98	Tc-99m	6.01 h		87.7	18.3 (6.72), 20.7 (1.35), 142.6 (0.041)
140.5	Mo-98	Mo-99	2.75 d	3.1369	90.7	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
141	Xe-124	Xe-125m	57 s	2.8	19.6	29.5 (17.9), 29.8 (33.2), 33.6 (9.75), 34.4 (2.14), 111 (61.8)
141.4	Sm-154	Sm-155	22.1 m	124.85	2.03	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 104.3 (75.1), 245.7 (3.76)
142.5	Fe-58	Fe-59	44.503 d	0.3968	1.03	192.2 (3.11), 334.8 (0.26), 1099.3 (56.5), 1291.6 (43.2)
142.5	Sc-45	Sc-46m	18.7 s	980	56.0	4.1 (6.6), 4.5 (0.88), 889.2 (100), 1120.5 (100)
142.6	Zn-70	Zn-71m	3.94 h	0.005	5.6	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
142.6	Mo-98	Tc-99m	6.01 h		0.041	18.3 (6.72), 20.7 (1.35), 140.5 (87.7)
144.9	Yb-174	Yb-175	4.19 d	2067	0.33	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 137.7 (0.12), 282.5 (3.08), 396.3 (6.55)
145.3	Xe-126	Xe-127	36.41	0.27	4.29	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 172.1 (25.5), 202.9 (68.3), 375 (17.2)
145.4	Ce-140	Ce-141	32.50 d	50.433	48.4	35.5 (4.88), 36 (8.91), 40.7 (2.66), 41.8 (0.7)
146.1	Hf-178	Hf-179m$_2$	25.1 d	867.2	26.3	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
146.8	Ta-181	Ta-182m	15.8 m	1.029	35.6	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 171.6 (46.8), 185.0 (23.4), 318.4 (6.55)
147.2	Lu-176	Lu-177m	160.9 d	18.2	3.45	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
149.1	Ru-104	Ru-105	4.44 h	5.984	1.74	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
149.7	Te-130	Te-131m	30 h	0.676	20.5	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
149.7	Te-130	Te-131	25.0 m	9.126	68.9	28.3 (3.7), 28.6 (6.9), 32.3 (2), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)
150.4	Yb-176	Yb-177	1.9 h		20.0	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 121.6 (3.4), 1080.1 (5.5), 1241.4 (3.4)
150.8	Cd-110	Cd-111m	48.6 m	1.75	30.3	23.1 (34.1), 26.2 (7.25), 245.4 (94)
151.2	Kr-84	Kr-85 m	4.48 h	5.13	75.0	12.6 (3.39), 13.4 (1.80), 14.1 (0.58), 15.0 (0.32), 129.8 (0.30), 304.9 (14)
151.2	Sr-84	Sr-85g	64.84 d	0.196	0.001	13.4 (50.2), 15.0 (8.80), 514 (99.3), 868.1 (0.12)
151.2	Kr-84	Kr-85g	10.72 yr	2.394	0.0001	362.8 (0.0001), 514.0 (0.434)
151.2	Sr-84	Sr-85m	67.66 m	0.336	12.8	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 129.8 (0.15), 231.9 (84.4), 238.8 (0.275)
152.4	Ta-181	Ta-182	115.0 d	2099.74	7.17	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
153.3	Hf-176	Hf-177m_1	1.08 s		21.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
153.3	Lu-176	Hf-177m_1	1.08 s		21.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
153.3	Hf-176	Hf-177m_2	51.4 m		21.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
153.3	Lu-176	Lu-177m	160.9 d	18.2	16.6	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
153.8	Dy-164	Dy-165m	1.26 m	47 940	0.242	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 108.2 (3.01), 361.7 (0.534), 515.5 (1.53)
155	Re-187	Re-188	16.98 h	4569.8	14.9	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
155.9	Nd-148	Nd-149	1.73 h	14.4	5.93	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
156	Re-187	Re-188m	18.6 m	100.16	0.6	59.7 (18), 61.1 (31), 63.6 (21.3), 69.2 (11), 71.2 (2.7), 92.5 (5.15), 105.9 (10.8), 169.5 (0.1)
156	Sn-116	Sn-117m	13.61 d	0.0882	2.11	25 (18.7), 25.3 (35.1), 28.5 (9.92), 29.1 (1.91), 158.6 (86.4)
156.4	Ge-76	Ge-77	11.3 h	0.468	0.76	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
156.4	Ta-181	Ta-182	115.0 d	2099.74	2.73	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
158.2	Xe-134	Xe-135	9.08 h	2.704	0.29	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 249.8 (90.4), 358.4 (0.22), 408 (0.36), 608.2 (2.91)
158.4	Pt-198	Au-199	3.14 d		36.9	49.8 (0.33), 68.9 (4.47), 70.8 (7.59), 80.2 (2.62), 82.5 (0.72), 208.2 (8.38)
158.4	Hg-198	Hg-199m	42.6 m	0.1818	52.5	68.9 (19.0), 70.8 (32.3), 80.2 (11.1), 82.5 (3.08), 374.1 (13.8), 413.4 (0.027)
158.6	Cd-116	Cd-117m	3.36 h	0.1867	109.0	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
158.6	Sn-116	Sn-117m	13.61 d	0.0882	86.4	25 (18.7), 25.3 (35.1), 28.5 (9.92), 29.1 (1.91), 156 (2.11)
159	Te-122	Te-123m	119.7 d	2.86	84.0	27.2 (14.1), 27.5 (26.2), 31 (7.55), 31.7 (1.53), 88.5 (0.087), 247.5 (0.0003)
159.1	Ge-76	Ge-77	11.3 h	0.468	0.22	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
159.4	Ca-46	Sc-47	3.34 d	0.00296	68.0	–
159.7	Ge-76	Ge-77m	52.9 s	0.78	11.3	9.9 (3.75), 10.5 (0.13), 11.0 (0.55), 194.8 (0.48), 215.5 (20.9)
160.3	Sn-122	Sn-123	129.2 d	0.0046	0.002	1021.0 (0.002), 1030.2 (0.031), 1088.6 (0.6)
160.3	Sn-122	Sn-123m	40.08 m	0.828	85.6	26.1 (3.06), 26.4 (5.72), 29.7 (1.63), 30.4 (0.32), 381.4 (0.042), 541.8 (0.02)
160.5	W-182	W-183m	5.15 s		4.9	46.5 (5.9), 52.6 (6.7), 58.0 (36.5), 59.3 (63.5), 67.2 (21.5), 69.1 (5.46), 99.1 (8.6), 102.5 (2.35), 107.9 (18.2)
160.6	Xe-132	Xe-133	5.29 d	10.76	0.064	30.6 (13.3), 31 (24.5), 35 (7.25), 35.8 (1.64), 79.6 (0.6), 81 (35.9)
160.7	Hf-178	Hf-179m$_1$	18.7 s	1436.3	2.79	54.6 (16), 55.8 (28.1), 63.2 (9.39), 65.0 (2.4), 214.1 (95.2), 374.8 (0.005)
162	Se-76	Se-77m	17.5 s	198	52.4	11.2 (20.0), 12.5 (3.14)
162	Ge-76	As-77	38.83 h		0.16	88 (0.27), 239.1 (1.59), 250 (0.41), 520.8 (0.42)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
162.4	In-115	In-116m_2	2.18 s	7753.32	36.6	24.1 (28.3), 27.4 (6.13)
162.5	Th-232	Th-233	22.3 m	740	0.32	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
163.5	Ru-104	Ru-105	4.44 h	5.984	0.15	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
163.9	Xe-130	Xe-131m	11.9 d	1.845	1.96	29.5 (15.1), 29.8 (28.1), 33.6 (8.26), 34.4 (1.81)
163.9	Te-130	Xe-131m	11.9 d		1.96	29.5 (15.1), 29.8 (28.1), 33.6 (8.26), 34.4 (1.81)
165	Hg-196	Hg-197m	23.8 h	18	0.27	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 130.4 (0.22), 133.9 (34.1), 279.0 (5.0)
165.2	Gd-160	Gd-161	3.70 m	16.832	2.60	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 283.6 (6), 314.9 (22.9), 360.9 (60.6), 480.1 (2.70)
165.8	Ce-138	Ce-139	137.7 d	0.275	79.9	33 (22.6), 33.4 (41.5), 37.8 (12.3), 38.7 (3.02)
165.8	Ba-138	Ba-139	1.38 h	25.812	23.8	33 (1.38), 33.4 (2.54), 37.8 (0.76), 38.7 (0.19), 1254.7 (0.033), 1420.5 (0.261)
167.7	Nd-150	Pm-151	28.40 h		7.84	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
169.1	Th-232	Th-233	22.3 m	740	0.34	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
169.1	Pd-110	Pd-111m	5.5 h	0.43	0.42	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
169.5	Re-187	Re-188m	18.6 m	100.16	0.1	59.7 (18), 61.1 (31), 63.6 (21.3), 69.2 (11), 71.2 (2.7), 92.5 (5.15), 105.9 (10.8), 156.0 (0.6)
169.8	Hf-178	Hf-179m_2	25.1 d	867.2	18.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
170.7	Th-232	Th-233	22.3 m	740	0.13	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
170.7	Mg-26	Mg-27	9.46 m	0.42058	0.84	843.7 (71.8), 1014.4 (28.2)
170.7	Nd-150	Nd-151	12.4 m	6.768	3.81	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
171.3	Pd-110	Ag-111m	64.8 s		0.11	22.1 (15.6), 25.0 (3.30), 59.8 (0.531), 245.4 (0.46), 620.1 (0.11)
171.6	Ta-181	Ta-182m	15.8 m	1.029	46.8	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 146.8 (35.6), 185.0 (23.4), 318.4 (6.55)
171.9	Lu-176	Lu-177m	160.9 d	18.2	4.75	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
172.1	Xe-126	Xe-127	36.41	0.27	25.5	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 145.3 (4.29), 202.9 (68.3), 375 (17.2)
172.2	Pt-190	Pt-191	2.90 d	1.95	3.52	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
172.2	Pd-110	Pd-111m	5.5 h	0.43	33.5	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
172.5	Xe-126	Xe-127m	69.2 s	0.0405	37.9	29.5 (15.4), 29.8 (28.6), 33.6 (8.43), 34.4 (1.85), 124.6 (69.1)
173.3	Pd-110	Pd-111	23.4 m	2.22	0.11	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
173.7	W-184	W-185m	1.67 m	0.061	3.30	9 (12), 58.0 (2.2), 59.3 (3.9), 65.9 (5.82), 131.5 (4.34)
174.4	Lu-176	Hf-177m$_1$	1.08 s		16.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
174.4	Hf-176	Hf-177m$_2$	51.4 m		16.4	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
174.4	Hf-176	Hf-177m$_1$	1.08 s		16.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
175.1	Nd-150	Nd-151	12.4 m	6.768	7.19	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
176.2	Ga-69	Ga-70	21.15 m	100.968	0.297	1039.2 (0.673)
176.3	Sn-124	Sb-125	2.73 yr		6.79	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
177	Hf-176	Hf-177m$_1$	1.08 s		4.47	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
177	Lu-176	Lu-177m	160.9 d	18.2	3.40	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
177	Hf-176	Hf-177m$_2$	51.4 m		4.47	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
177	Lu-176	Hf-177m$_1$	1.08 s		4.47	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
177.2	Nd-150	Pm-151	28.40 h		3.58	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
177.2	Yb-168	Yb-169	32.02 d		22.0	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 130.5 (11.5), 198 (36), 307.7 (11.1)
177.3	Ge-76	Ge-77	11.3 h	0.468	0.17	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
179.4	Ta-181	Ta-182	115.0 d	2099.74	3.15	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
179.7	Mo-100	Tc-101	14.2 m		0.64	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
180.2	Kr-78	Kr-79g	35.04 h	2.1	0.10	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
180.3	Au-197	Au-198m	2.30 d		50.8	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 204.1 (41.5), 214.9 (76.9), 333.8 (15.0)
181.1	Mo-98	Mo-99	2.75 d	3.1369	6.08	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
182.2	Dy-156	Dy-157	8.10 h	1.98	2.2	43.7 (24.1), 44.5 (43.2), 50.3 (13.8), 51.7 (3.8), 60.8 (0.4), 83.0 (0.6), 265.3 (0.3), 326.2 (93.2)
183.6	Ru-104	Ru-105	4.44 h	5.984	0.10	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
184.1	Mo-100	Tc-101	14.2 m		1.69	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
184.4	Ho-165	Ho-166m	1.2×10^3 yr	350	73.9	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
185	Ta-181	Ta-182m	15.8 m	1.029	23.4	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 146.8 (35.6), 171.6 (46.8), 318.4 (6.55)
185.8	Pt-198	Pt-199	30.8 m	26.64	3.26	68.8 (2.0), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
186.7	Os-189	Os-190m	9.90 m	0.0041	69.9	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 361.1 (95.2), 502.6 (97.8), 616.1 (98.5)
187.4	Mo-100	Mo-101	14.6 m	1.916	0.47	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
188.4	Xe-124	Xe-125	16.9 h	13.7	54.9	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 243.4 (28.8), 453.8 (4.23), 846.5 (1.03)
188.9	Pd-108	Pd-109m	4.69 m	4.76	55.7	21.1 (21.7), 23.9 (4.56)
188.9	Se-82	Se-83m	70.4 s	0.3666	0.17	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
190.3	Kr-80	Kr-81m	13.0 s	10.2375	67.0	12.6 (14.9), 14.1 (2.56)
190.5	Th-232	Th-233	22.3 m	740	0.13	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
191.4	Pt-196	Pt-197	18.3 h	17.71	3.68	67.0 (0.99), 68.8 (1.69), 77.3 (17.2), 77.9 (0.58), 80.2 (0.16), 268.8 (0.23)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
191.4	Hg-196	Hg-197	2.67 d	462	0.49	67.0 (21.2), 68.8 (36.2), 77.3 (18.1), 77.9 (12.5), 80.2 (3.37), 268.7 (0.038)
191.6	In-113	In-114m	49.51 d	34.668	16.0	23.1 (3.91), 24.1 (28.1), 27.4 (6.08), 558 (3.4), 725.2 (3.4)
191.7	Pt-198	Pt-199	30.8 m	26.64	2.38	68.8 (2.0), 185.8 (3.26), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
191.9	Mo-100	Mo-101	14.6 m	1.916	18.8	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
192.2	Fe-58	Fe-59	44.503 d	0.3968	3.11	142.5 (1.03), 334.8 (0.26), 1099.3 (56.5), 1291.6 (43.2)
192.8	Hf-178	Hf-179m$_2$	25.1 d	867.2	20.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
194.8	Ge-76	Ge-77m	52.9 s	0.78	0.48	9.9 (3.75), 10.5 (0.13), 11.0 (0.55), 159.7 (11.3), 215.5 (20.9)
194.8	Ge-76	Ge-77	11.3 h	0.468	1.68	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
194.9	Th-232	Th-233	22.3 m	740	0.16	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
195.9	Mo-100	Mo-101	14.6 m	1.916	2.86	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
196.6	Xe-128	Xe-129m	8.87 d	0.91	4.59	29.5 (36.1), 29.8 (66.9), 33.6 (19.7), 34.4 (4.32), 39.6 (7.52)
197	Tb-159	Tb-160	72.3 d	2550	5.24	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
197.1	O-18	O-19	27.1 s	0.000032	95.9	109.9 (2.54), 1356.8 (50.4), 1444.1 (2.64), 1554 (1.39)
197.4	Nd-146	Pm-147	2.62 yr		0.0001	39.5 (0.0006), 40.1 (0.0011), 45.4 (0.0003), 46.6 (0.0001), 76.2 (0.0001), 121.3 (0.0028)
198	Yb-168	Yb-169	32.02 d		36.0	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 130.5 (11.5), 177.2 (22), 307.7 (11.1)
198.6	Ge-74	Ge-75	83 m	12.41	1.17	66.0 (0.11), 264.6 (11.3), 419.1 (0.18), 468.8 (0.22), 617.7 (0.11)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
198.6	Se-74	Se-75	119.8 d	46.62	1.47	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
200.6	Te-130	Te-131m	30 h	0.676	7.53	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
201.8	Lu-175	Lu-176	3.7×10^{10} yr	681.8	84.7	54.6 (9.45), 55.8 (16.5), 63.2 (5.53), 65.0 (1.42), 88.3 (13.1), 306.9 (93.3), 401.1 (0.84)
201.8	Pt-196	Au-197m	7.80 s		1.1	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 279.0 (70.9), 409.1 (0.11)
201.8	Hg-196	Au-197m	7.80 s		1.1	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 279.0 (70.9), 409.1 (0.11)
202.5	Y-89	Y-90m	3.19 h	0.1	96.5	479.5 (90.6), 682.0 (0.32)
202.9	Xe-126	Xe-127	36.41	0.27	68.3	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 145.3 (4.29), 172.1 (25.5), 375 (17.2)
202.9	Te-126	Te-127m	109 d	2.55	0.057	27.2 (10.4), 27.5 (19.3), 28.3 (0.4), 28.6 (0.75), 31 (5.56), 31.7 (1.13), 32.3 (0.22), 57.6 (0.5), 88.3 (0.084), 360.3 (0.132), 417.9 (0.969)
202.9	Te-126	Te-127	9.35 h	17.055	0.058	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 215.1 (0.039), 360.3 (0.135), 417.9 (0.993)
203.7	Hg-204	Hg-205	5.20 m	2.924	2.2	415.6 (0.13), 1218.7 (0.006)
204.1	Zr-94	Nb-95m	3.61 d		2.36	16.6 (36), 18.6 (6.97), 235.7 (25.1), 582.1 (0.055)
204.1	Zr-94	Nb-95	34.98 d		0.015	561.9 (0.015), 765.8 (99.9)
204.1	Au-197	Au-198m	2.30 d		41.5	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 180.3 (50.8), 214.9 (76.9), 333.8 (15.0)
204.2	Lu-176	Hf-177m$_1$	1.08 s		17.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
204.2	Hf-176	Hf-177m$_1$	1.08 s		17.8	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
204.2	Lu-176	Lu-177m	160.9 d	18.2	13.6	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
204.2	Hf-176	Hf-177m$_2$	51.4 m		17.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
205.8	Ir-191	Ir-192	73.83 d	23 275.2	3.20	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
205.8	Ir-191	$Ir-192m_2$	241 yr	3.73	3.20	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
207.8	Er-166	Er-167m	2.28 s	504	41.7	48.2 (5.40), 49.1 (9.60), 55.6 (3.07), 57.2 (0.85)
208.1	Nd-148	Nd-149	1.73 h	14.4	2.54	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
208.2	Pt-198	Au-199	3.14 d		8.38	49.8 (0.33), 68.9 (4.47), 70.8 (7.59), 80.2 (2.62), 82.5 (0.72), 158.4 (36.9)
208.3	Se-82	Se-83g	22.5 m	0.04888	1.85	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
208.4	Lu-176	Lu-177m	160.9 d	18.2	59.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
208.4	Lu-176	$Hf-177m_1$	1.08 s		73.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
208.4	Hf-176	$Hf-177m_2$	51.4 m		73.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
208.4	Lu-176	Lu-177	6.71 d	5460	11.0	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 249.7 (0.21), 321.3 (0.22)
208.4	Hf-176	$Hf-177m_1$	1.08 s		73.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
208.4	Yb-176	Lu-177	6.71 d		11.0	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 249.7 (0.21), 321.3 (0.22)
208.5	Kr-78	Kr-79g	35.04 h	2.1	0.78	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
209	Ge-76	Ge-77	11.3 h	0.468	0.89	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
209	Te-128	Te-129	69.6 m	6.34	0.17	27.8 (16.64), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
211	Ge-76	Ge-77	11.3 h	0.468	29.2	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
211.3	Nd-148	Nd-149	1.73 h	14.4	25.9	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
212	Mo-100	Mo-101	14.6 m	1.916	0.51	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
212.2	Te-120	Te-121m	154.0 d	0.032	81.4	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
213.4	Hf-177	Hf-178m_2	31 yr	0.00000186	80.9	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
213.4	Hf-177	Hf-178m_1	4.0 s	20.46	80.9	54.6 (20.1), 55.8 (35.2), 63.2 (11.8), 65.0 (3.01), 88.9 (62), 93.2 (17.3), 325.6 (93.9), 426.4 (96.9)
214	Hf-176	Hf-177m_2	51.4 m		40.9	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
214.1	Hf-178	Hf-179m_1	18.7 s	1436.3	95.2	54.6 (16), 55.8 (28.1), 63.2 (9.39), 65.0 (2.4), 160.7 (2.79), 374.8 (0.005)
214.5	Hf-176	Hf-177m_2	51.4 m		8.56	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
214.5	Hf-176	Hf-177m_1	1.08 s		8.56	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
214.5	Lu-176	Lu-177m	160.9 d	18.2	6.52	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
214.5	Lu-176	Hf-177m$_1$	1.08 s		8.56	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
214.9	Pd-106	Pd-107m	21.3 s	0.355	68.7	21.1 (16.0), 23.9 (3.35)
214.9	Au-197	Au-198m	2.30 d		76.9	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 180.3 (50.8), 204.1 (41.5), 333.8 (15.0)
215.1	Te-126	Te-127	9.35 h	17.055	0.039	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 202.9 (0.058), 360.3 (0.135), 417.9 (0.993)
215.2	Hf-179	Hf-180m	5.5 h	4.67	81.4	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 332.3 (94.4), 443.2 (82.8), 500.7 (14.5)
215.5	Ge-76	Ge-77	11.3 h	0.468	27.1	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
215.5	Ge-76	Ge-77m	52.9 s	0.78	20.9	9.9 (3.75), 10.5 (0.13), 11.0 (0.55), 159.7 (11.3), 194.8 (0.48)
215.6	Tb-159	Tb-160	72.3 d	2550	4.02	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
215.7	Ru-96	Ru-97	2.9 d	1.6	85.8	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 324.5 (10.2), 460.6 (0.12), 569.3 (0.88)
215.9	Ho-165	Ho-166m	1.2×10^3 yr	350	2.63	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
216.1	Ba-130	Ba-131	11.8 d	0.9328	19.9	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
216.4	Tl-205	Tl-206m	3.75 m		83.6	247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
216.7	Hf-177	Hf-178m$_2$	31 yr	0.00000186	63.7	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
217	Hf-178	Hf-179m$_2$	25.1 d	867.2	8.78	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
217	Kr-78	Kr-79g	35.04 h	2.1	2.40	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
218.1	Lu-176	Lu-177m	160.9 d	18.2	3.28	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
218.9	Zr-96	Zr-97	17.0 h	0.0632	0.17	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
219.1	Ge-76	Ge-77	11.3 h	0.468	0.1	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
221.5	Br-81	Br-82g	35.30 h	12.8206	2.25	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
221.8	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
222.1	Ta-181	Ta-182	115.0 d	2099.74	7.56	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
225.1	Ru-104	Ru-105	4.44 h	5.984	0.12	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
225.2	Se-82	Se-83g	22.5 m	0.04888	31.9	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
226	Gd-158	Gd-159	18.56 h	62.1	0.21	43.7 (5.54), 44.5 (9.93), 50.3 (3.06), 51.7 (0.89), 58.0 (2.27), 348.2 (0.22), 363.6 (10.8)
227	Yb-176	Yb-177m	6.41 m		12.2	51.4 (21.2), 52.4 (37.5), 59.3 (12.2), 61.0 (3.26), 104.5 (76.6)
228.5	Hf-176	Hf-177m$_2$	51.4 m		48.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
228.5	Lu-176	Hf-177m$_1$	1.08 s		48.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
228.5	Hf-176	Hf-177m$_1$	1.08 s		48.0	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
228.5	Lu-176	Lu-177m	160.9 d	18.2	36.6	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
229.3	Ta-181	Ta-182	115.0 d	2099.74	3.64	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
229.6	Hf-174	Hf-175	70 d	62.4	0.77	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 318.9 (0.17), 343.4 (86.6), 353.6 (0.23), 432.8 (1.56)
231.5	Se-82	Se-83m	70.4 s	0.3666	0.33	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
231.6	Ce-142	Ce-143	33.0 h	10.526	2.04	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 293.3 (42.8), 350.6 (3.27), 490.4 (2.13), 664.5 (5.6), 722 (5.32)
231.9	Sr-84	Sr-85m	67.66 m	0.336	84.4	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 129.8 (0.15), 151.2 (12.8), 238.8 (0.275)
233.2	Xe-132	Xe-133m	2.19 d	1.345	9.95	29.5 (16), 29.8 (29.7), 33.6 (8.73), 34.4 (1.92)
233.7	Mo-100	Tc-101	14.2 m		0.27	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
233.9	Hf-176	Hf-177m$_2$	51.4 m		7.24	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
233.9	Lu-176	Hf-177m$_1$	1.08 s		7.24	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
233.9	Lu-176	Lu-177m	160.9 d	18.2	5.51	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
233.9	Hf-176	Hf-177m$_1$	1.08 s		7.24	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
235.7	Zr-94	Zr-95	64.03 d	0.86227	0.24	16.6 (0.42), 18.6 (0.66), 724.2 (43.7), 756.7 (55.4)
235.7	Zr-94	Nb-95m	3.61 d		25.1	16.6 (36), 18.6 (6.97), 204.1 (2.36), 582.1 (0.055)
236.6	Hf-178	Hf-179m$_2$	25.1 d	867.2	18.3	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
237.4	Hf-177	Hf-178m$_2$	31 yr	0.00000186	8.79	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
238.3	Mo-100	Tc-101	14.2 m		0.31	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
238.8	Sr-84	Sr-85m	67.66 m	0.336	0.275	13.4 (50.2), 14.1 (1.23), 15.0 (1.23), 15.9 (0.22), 129.8 (0.15), 151.2 (12.8), 231.9 (84.4)
239.1	Ge-76	As-77	38.83 h		1.59	88 (0.27), 162 (0.16), 250 (0.41), 520.8 (0.42)
239.6	Ba-130	Ba-131	11.8 d	0.9328	2.42	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
240.1	Nd-150	Pm-151	28.40 h		3.61	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
240.2	Nd-148	Nd-149	1.73 h	14.4	3.94	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
240.9	Te-130	Te-131m	30 h	0.676	7.57	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
243.4	Xe-124	Xe-125	16.9 h	13.7	28.8	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 188.4 (54.9), 453.8 (4.23), 846.5 (1.03)
244.7	Eu-151	Eu-152	13.33 yr	2 82 020	7.51	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
245.4	Pd-110	Pd-111	23.4 m	2.22	0.48	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
245.4	Pd-110	Ag-111m	64.8 s		0.46	22.1 (15.6), 25.0 (3.30), 59.8 (0.531), 171.3 (0.11), 620.1 (0.11)
245.4	Pd-110	Pd-111m	5.5 h	0.43	0.45	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
245.4	Pd-110	Ag-111	7.45 d		1.24	23.1 (0.18), 96.7 (0.2), 342.1 (6.68)
245.4	Cd-110	Cd-111m	48.6 m	1.75	94.0	23.1 (34.1), 26.2 (7.25), 150.8 (30.3)
245.7	Sm-154	Sm-155	22.1 m	124.85	3.76	40.9 (4.92), 41.5 (8.88), 47.0 (2.71), 104.3 (75.1), 141.4 (2.03)
246.5	Pt-198	Pt-199	30.8 m	26.64	2.16	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
247.2	Tl-205	Tl-206m	3.75 m		12.0	216.4 (83.6), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
247.5	Te-122	Te-123m	119.7 d	2.86	0.0003	27.2 (14.1), 27.5 (26.2), 31 (7.55), 31.7 (1.53), 88.5 (0.087), 159 (84)
247.9	Eu-153	Eu-154	8.8 yr		6.60	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
249.4	Ba-130	Ba-131	11.8 d	0.9328	2.84	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 373.2 (14.1), 486.5 (2.09), 496.3 (47.1)
249.7	Lu-176	Lu-177m	160.9 d	18.2	6.08	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
249.7	Lu-176	Hf-177m$_1$	1.08 s		7.97	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
249.7	Hf-176	Hf-177m$_1$	1.08 s		7.97	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
249.7	Hf-176	Hf-177m$_2$	51.4 m		7.97	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
249.7	Yb-176	Lu-177	6.71 d		0.21	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 321.3 (0.22)
249.7	Lu-176	Lu-177	6.71 d	5460	0.21	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 321.3 (0.22)
249.8	Xe-134	Xe-135	9.08 h	2.704	90.4	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 358.4 (0.22), 408 (0.36), 608.2 (2.91)
250	Ge-76	As-77	38.83 h		0.41	88 (0.27), 162 (0.16), 239.1 (1.59), 520.8 (0.42)
250.6	Te-128	Te-129	69.6 m	6.34	0.35	27.8 (16.64), 209 (0.17), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
254.2	Zr-96	Zr-97	17.0 h	0.0632	1.10	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
254.3	Ce-136	Ce-137m	34.4 h	0.1805	10.9	33 (29.5), 33.4 (54.2), 34.3 (15.6), 34.7 (28.7), 37.8 (16.1), 38.7 (3.94), 39.2 (8.53), 40.2 (2.2), 447.2 (3.03)
254.7	Ge-76	Ge-77	11.3 h	0.468	0.20	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
255	Sn -112	Sn-113	115.1 d	0.71	1.85	24.1 (79.8), 27.4 (17.3), 391.7 (64.2), 638 (0.001)
255.7	Nd-150	Nd-151	12.4 m	6.768	15.3	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
257.5	Hf-178	Hf-179m_2	25.1 d	867.2	3.2	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
257.6	Hf-177	Hf-178m_2	31 yr	0.00000186	16.6	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
260.9	Cd-114	Cd-115	2.23 d	86.16	1.94	24.1 (33.1), 27.4 (7.15), 336.2 (49.7), 492.4 (8.03), 527.9 (27.5)
261.3	Kr-78	Kr-79g	35.04 h	2.1	12.7	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
262.8	Ru-104	Ru-105	4.44 h	5.984	6.49	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
263.1	Mo-92	Mo-93m	6.85 h	0.089	56.7	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 684.7 (99.7), 949.8 (0.12), 1363.0 (0.79), 1477.1 (99.1)
263.7	Cd-112	Cd-113m	14.1 yr	0.96	0.023	23.1 (0.035), 26.2 (0.0074)
264.1	Ta-181	Ta-182	115.0 d	2099.74	3.64	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
264.4	Ge-76	Ge-77	11.3 h	0.468	51.0	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
264.6	Se-74	Se-75	119.8 d	46.62	59.1	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 279.5 (25.2), 303.9 (1.34), 400.6 (11.6)
264.6	Ge-74	Ge-75	83 m	12.41	11.3	66.0 (0.11), 198.6 (1.17), 419.1 (0.18), 468.8 (0.22), 617.7 (0.11)
265.3	Dy-156	Dy-157	8.10 h	1.98	0.3	43.7 (24.1), 44.5 (43.2), 50.3 (13.8), 51.7 (3.8), 60.8 (0.4), 83.0 (0.6), 182.2 (2.2), 326.2 (93.2)
265.7	Bi-209	Bi-210m	3.0×10^6 yr	1.4	51.0	70.8 (3.9), 72.9 (6.6), 82.5 (2.3), 304.8 (28.0), 649.8 (2.9)
265.7	Bi-209	Bi-210	5.01 d	1.9	0.0001	11.1 (0.023), 13.4 (0.023), 15.8 (0.0043), 76.9 (0.003), 79.3 (0.006), 89.6 (0.002), 304.8 (0.0001)
265.7	Tl-205	Tl-206m	3.75 m		86.2	216.4 (83.6), 247.2 (12), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
267.7	Nd-148	Nd-149	1.73 h	14.4	6.03	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
268.1	Ge-76	Ge-77	11.3 h	0.468	0.3	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
268.2	Ba-134	Ba-135m	28.7 h	0.3872	15.6	31.8 (15.4), 32.2 (28.4), 36.4 (8.4), 37.3 (2.01)
268.7	Hg-196	Hg-197	2.67 d	462	0.038	67.0 (21.2), 68.8 (36.2), 77.3 (18.1), 77.9 (12.5), 80.2 (3.37), 191.4 (0.49)
268.8	Pt-196	Pt-197	18.3 h	17.71	0.23	67.0 (0.99), 68.8 (1.69), 77.3 (17.2), 77.9 (0.58), 80.2 (0.16), 191.4 (3.68)
268.8	Lu-176	Lu-177m	160.9 d	18.2	3.43	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
268.9	Hf-178	Hf-179m_2	25.1 d	867.2	11.0	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
270.2	Nd-148	Nd-149	1.73 h	14.4	10.7	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
272	Pd-110	Pd-111m	5.5 h	0.43	0.16	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
272.3	Zr-96	Zr-97	17.0 h	0.0632	0.25	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
273.3	Cd-116	Cd-117	2.49 h	0.3735	27.9	24.1 (4.58), 89.7 (3.26), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
273.5	Br-81	Br-82g	35.30 h	12.8206	0.79	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
275.2	Nd-150	Pm-151	28.40 h		6.61	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
275.9	Se-80	Se-81	18.5 m	26.288	0.87	290.1 (0.68), 552.4 (0.10), 566.0 (0.26), 828.3 (0.32)
275.9	Se-80	Se-81m	57.3 m	3.968	1.3	103.1 (9.79), 290.1 (1.0), 552.4 (0.15), 566 (0.38), 828.3 (0.47)
276	Kr-80	Kr-81g	2.1×10^5 yr	15.75	3.6	11.9 (46.2), 13.3 (7.3)
276.1	Ba-132	Ba-133m	38.9 h	0.0505	17.5	12.3 (1.5), 31.8 (15.3), 32.2 (28.2), 36.4 (8.34), 37.3 (1.99), 632.5 (0.01)
276.4	Ba-132	Ba-133	10.54 yr	0.6565	7.29	30.6 (34.4), 31 (63.5), 35 (18.8), 35.8 (4.24), 53.2 (2.2), 79.6 (2.43), 81 (32.8), 302.9 (18.6), 356 (62.3)
277.3	Hf-176	Hf-177m_2	51.4 m		75.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
278.4	Te-128	Te-129	69.6 m	6.34	0.53	27.8 (16.64), 209 (0.17), 250.6 (0.35), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
279	Hg-196	Hg-197m	23.8 h	18	5.0	67.0 (2.0), 68.9 (13.2), 70.8 (16.7), 77.9 (1.1), 80.2 (6.07), 82.5 (1.59), 130.4 (0.22), 133.9 (34.1), 165.0 (0.27)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
279	Hg-196	Au-197m	7.80 s		70.9	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 409.1 (0.11)
279	Pt-196	Au-197m	7.80 s		70.9	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 409.1 (0.11)
279	Pt-196	Pt-197m	1.57 h	1.265	2.3	53.0 (1.08), 65.1 (13.6), 66.8 (23.8), 68.8 (0.40), 75.7 (8), 77.9 (2.26), 130.2 (0.1), 346.5 (11.2)
279.2	Hg-202	Hg-203	46.58 d	145.53	81.5	70.8 (3.75), 72.9 (6.35), 82.5 (2.2), 84.9 (0.63)
279.5	Se-74	Se-75	119.8 d	46.62	25.2	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 264.6 (59.1), 303.9 (1.34), 400.6 (11.6)
279.8	Dy-164	Dy-165	2.33 h	28 200	0.50	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
280.1	Ru-104	Rh-105	35.36 h		0.17	21.1 (0.31), 306.1 (5.13), 318.9 (19.2)
280.5	Ho-165	Ho-166m	1.2×10^3 yr	350	30.1	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
281.3	Te-128	Te-129	69.6 m	6.34	0.15	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
281.8	Lu-176	Lu-177m	160.9 d	18.2	14.0	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
281.8	Lu-176	$Hf\text{-}177m_1$	1.08 s		18.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
281.8	Hf-176	$Hf\text{-}177m_2$	51.4 m		18.4	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
281.8	Hf-176	$Hf\text{-}177m_1$	1.08 s		18.4	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
282.5	Yb-174	Yb-175	4.19 d	2067	3.08	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 137.7 (0.12), 144.9 (0.33), 396.3 (6.55)
283.6	Gd-160	Gd-161	3.70 m	16.832	6.00	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						165.2 (2.60), 314.9 (22.9), 360.9 (60.6), 480.1 (2.70)
284.3	Te-130	I-131	8.02 d		6.06	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 364.5 (81.2), 637 (7.27), 722.9 (1.8)
285.9	Nd-148	Pm-149	2.21 d		3.10	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 590.9 (0.68), 830.5 (0.03), 833.2 (0.03)
289.3	Ga-71	Ga-72	14.1 h	187.929	0.19	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
289.8	Pd-110	Pd-111	23.4 m	2.22	0.10	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
289.8	Pd-110	Pd-111m	5.5 h	0.43	0.99	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
290.1	Se-80	Se-81	18.5 m	26.288	0.68	275.9 (0.87), 552.4 (0.10), 566.0 (0.26), 828.3 (0.32)
290.1	Se-80	Se-81m	57.3 m	3.968	1.0	103.1 (9.79), 275.9 (1.3), 552.4 (0.15), 566 (0.38), 828.3 (0.47)
293.3	Ce-142	Ce-143	33.0 h	10.526	42.8	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 350.6 (3.27), 490.4 (2.13), 664.5 (5.6), 722 (5.32)
293.5	Ir-193	Ir-194	19.15 h	6897	2.54	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
295	Ru-102	Ru-103	39.25 d	38.24	0.25	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 443.8 (0.32), 497.1 (89.5), 557.0 (0.832), 610.3 (5.64)
295.1	Hf-176	Hf-177m_2	51.4 m		69.0	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
295.9	Er-170	Er-171	7.52 h	84.93	28.9	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 116.7 (2.30), 124.0 (9.10), 308.3 (64.4)
295.9	Ir-191	Ir-192m_1	1.44 m	11 190	0.0018	8 (0.35), 9.2 (8.0), 10.8 (11.6), 12.5 (2.1), 58 (0.039), 316.5 (0.0097)
295.9	Ir-191	Ir-192m_2	241 yr	3.73	28.7	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
295.9	Ir-191	Ir-192	73.83 d	23 275.2	28.7	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
296.1	Se-82	Se-83g	22.5 m	0.04888	0.27	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
296.2	Hf-176	Hf-177m_2	51.4 m		6.51	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
296.2	Lu-176	Lu-177m	160.9 d	18.2	4.96	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
296.5	Lu-176	Hf-177m_1	1.08 s		6.51	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
296.5	Hf-176	Hf-177m_1	1.08 s		6.51	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
296.8	Hf-177	Hf-178m_2	31 yr	0.00000186	9.83	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
297.9	Er-162	Ho-163m	1.09 s		77.5	46.7 (3.19), 47.5 (5.69), 53.8 (1.79), 55.3 (0.50)
298	Xe-136	Xe-137	3.83 m	2.314	0.12	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
298.6	Tb-159	Tb-160	72.3 d	2550	27.4	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
299	Lu-176	Hf-177m_1	1.08 s		2.36	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
299	Hf-176	Hf-177m_2	51.4 m		23.6	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
299	Hf-176	Hf-177m_1	1.08 s		2.36	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
299.5	Kr-78	Kr-79g	35.04 h	2.1	1.57	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
300.6	Nd-150	Nd-151	12.4 m	6.768	1.93	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
300.7	Ho-165	Ho-166m	1.2×10^3 yr	350	3.79	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
300.7	Ir-193	Ir-194	19.15 h	6897	0.35	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
302.9	Ba-132	Ba-133	10.54 yr	0.6565	18.6	30.6 (34.4), 31 (63.5), 35 (18.8), 35.8 (4.24), 53.2 (2.2), 79.6 (2.43), 81 (32.8), 276.4 (7.29), 356 (62.3)
303.9	Se-74	Se-75	119.8 d	46.62	1.34	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 400.6 (11.6)
304.8	Bi-209	Bi-210	5.01 d	1.9	0.0001	11.1 (0.023), 13.4 (0.023), 15.8 (0.0043), 76.9 (0.003), 79.3 (0.006), 89.6 (0.002), 265.7 (0.0001)
304.8	Bi-209	Bi-210m	3.0×10^6 yr	1.4	28.0	70.8 (3.9), 72.9 (6.6), 82.5 (2.3), 265.7 (51.0), 649.8 (2.9)
304.9	Kr-84	Kr-85 m	4.48 h	5.13	14.0	12.6 (3.39), 13.4 (1.80), 14.1 (0.58), 15.0 (0.32), 129.8 (0.30), 151.2 (75)
305.5	Hf-176	Hf-177m_1	1.08 s		2.38	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
305.5	Lu-176	Hf-177m_1	1.08 s		2.38	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
305.5	Hf-176	Hf-177m_2	51.4 m		22.38	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
306.1	Ru-104	Rh-105	35.36 h		5.13	21.1 (0.31), 280.1 (0.17), 318.9 (19.2)
306.3	Kr-78	Kr-79g	35.04 h	2.1	2.60	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
306.8	Mo-100	Tc-101	14.2 m		88.0	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
306.9	Lu-175	Lu-176	3.7×10^{10} yr	681.8	93.3	54.6 (9.45), 55.8 (16.5), 63.2 (5.53), 65.0 (1.42), 88.3 (13.1), 201.8 (84.7), 401.1 (0.84)
307.7	Yb-168	Yb-169	32.02 d		11.1	49.8 (54.2), 50.7 (95.5), 57.5 (30.8), 59.1 (8.30), 63.1 (45), 93.6 (2.70), 109.8 (18), 118.2 (1.93), 130.5 (11.5), 177.2 (22), 198 (36)
308.1	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
308.3	Er-170	Er-171	7.52 h	84.93	64.4	49.8 (13.5), 50.7 (23.8), 57.5 (7.66), 59.1 (2.07), 111.6 (20.5), 116.7 (2.30), 124.0 (9.10), 295.9 (28.9)
308.4	Ir-191	Ir-192m_2	241 yr	3.73	29.7	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
308.4	Ir-191	Ir-192	73.83 d	23 275.2	29.7	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
311.3	Mo-100	Tc-101	14.2 m		0.24	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
311.4	Pd-108	Pd-109	13.46 h	219.6	0.032	22.1 (28.5), 25.0 (6.02), 88.0 (3.61), 647.3 (0.024)
311.5	Hf-176	Hf-177m_2	51.4 m		58.8	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
312.7	K-41	K-42	12.36 h	9.8258	0.35	899 (0.053), 1524.6 (18.6)
314.9	Gd-160	Gd-161	3.70 m	16.832	22.9	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 165.2 (2.60), 283.6 (6), 360.9 (60.6), 480.1 (2.70)
316	Hf-178	Hf-179m_2	25.1 d	867.2	19.7	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
316.4	Ru-104	Ru-105	4.44 h	5.984	10.9	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
316.5	Ir-191	Ir-192m_1	1.44 m	11 190	0.0097	8 (0.35), 9.2 (8.0), 10.8 (11.6), 12.5 (2.1), 58 (0.039), 295.9 (0.0018)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
316.5	Ir-191	Ir-192	73.83 d	23 275.2	82.9	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
316.5	Ir-191	Ir-192m$_2$	241 yr	3.73	82.9	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
316.8	Tl-205	Tl-206m	3.75 m		9.5	216.4 (83.6), 247.2 (12), 265.7 (86.2), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
317	Pt-198	Pt-199	30.8 m	26.64	4.87	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 493.8 (5.73), 543.0 (14.8), 714.5 (1.86)
318	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
318.4	Ta-181	Ta-182m	15.8 m	1.029	6.55	56.3 (26.8), 57.5 (46.8), 65.2 (15.8), 67.0 (4.02), 146.8 (35.6), 171.6 (46.8), 185.0 (23.4)
318.6	Zn-68	Zn-69	56 m	18.8	0.0012	872 (0.0002)
318.6	Zn-68	Zn-69m	13.76 h	1.353	0.0013	8.6 (1.90), 9.6 (0.26), 438.6 (94.8), 574.1 (0.033)
318.9	Ru-104	Rh-105	35.36 h		19.2	21.1 (0.31), 280.1 (0.17), 306.1 (5.13)
318.9	Hf-174	Hf-175	70 d	62.4	0.17	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 343.4 (86.6), 353.6 (0.23), 432.8 (1.56)
319	Lu-176	Lu-177m	160.9 d	18.2	10.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
319.4	Nd-146	Nd-147	10.98 d	24.066	1.95	38.2 (13), 38.7 (23.5), 43.8 (7.05), 44.9 (1.96), 91.1 (27.9), 531 (13)
320.1	Cr-50	Cr-51	27.7 d	69.165	9.83	–
320.1	Ti-50	Ti-51	5.8 m	0.9308	93.0	608.5 (1.18), 928.6 (6.88)
321.3	Yb-176	Lu-177	6.71 d		0.22	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 249.7 (0.21)
321.3	Lu-176	Lu-177	6.71 d	5460	0.22	54.6 (1.68), 55.8 (2.94), 63.2 (0.98), 65.0 (0.25), 71.6 (0.16), 113.0 (6.6), 208.4 (11), 249.7 (0.21)
324	Ir-193	Ir-194m	171 d	376.2	2.0	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
324.5	Ru-96	Ru-97	2.9 d	1.6	10.2	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 215.7 (85.8), 460.6 (0.12), 569.3 (0.88)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
325.6	Hf-177	Hf-178m$_1$	4.0 s	20.46	93.9	54.6 (20.1), 55.8 (35.2), 63.2 (11.8), 65.0 (3.01), 88.9 (62), 93.2 (17.3), 213.4 (80.9), 426.4 (96.9)
325.6	Hf-177	Hf-178m$_2$	31 yr	0.00000186	93.9	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
326.1	Ru-104	Ru-105	4.44 h	5.984	1.05	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
326.2	Dy-156	Dy-157	8.10 h	1.98	93.2	43.7 (24.1), 44.5 (43.2), 50.3 (13.8), 51.7 (3.8), 60.8 (0.4), 83.0 (0.6), 182.2 (2.2), 265.3 (0.3)
326.5	Nd-148	Nd-149	1.73 h	14.4	4.56	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
326.7	Hf-176	Hf-177m$_2$	51.4 m		65.4	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
327.7	Hf-176	Hf-177m$_1$	1.08 s		23.7	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
327.7	Lu-176	Hf-177m$_1$	1.08 s		23.7	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
327.7	Lu-176	Lu-177m	160.9 d	18.2	18.1	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
327.7	Hf-176	Hf-177m$_2$	51.4 m		23.7	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
327.7	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
328.4	Ir-193	Ir-194	19.15 h	6897	13.0	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
328.4	Ir-193	Ir-194m	171 d	376.2	92.8	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
328.8	La-139	La-140	40.28 h	892.12	20.7	432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
330.4	Zr-96	Zr-97	17.0 h	0.0632	0.11	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
330.9	Ru-104	Ru-105	4.44 h	5.984	0.66	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
332.1	Sn-124	Sn-125m	9.52 m	0.728	97.5	26.1 (0.53), 26.4 (1), 29.7 (0.28), 589.6 (0.2), 643 (0.16), 1404 (0.7), 1483.9 (0.18), 1615.3 (0.12)
332.1	Sn-124	Sn-125	9.64 d	0.0224	1.31	469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
332.3	Hf-179	Hf-180m	5.5 h	4.67	94.4	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 215.2 (81.4), 443.2 (82.8), 500.7 (14.5)
333.7	Mo-100	Mo-101	14.6 m	1.916	0.78	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
333.8	Au-197	Au-198m	2.30 d		15.0	67.0 (25.1), 68.8 (42.9), 77.9 (14.7), 80.2 (4.0), 97.2 (69.0), 180.3 (50.8), 204.1 (41.5), 214.9 (76.9)
334.3	Te-130	Te-131m	30 h	0.676	9.53	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
334.8	Fe-58	Fe-59	44.503 d	0.3968	0.26	142.5 (1.03), 192.2 (3.11), 1099.3 (56.5), 1291.6 (43.2)
336.2	Cd-114	Cd-115	2.23 d	86.16	49.7	24.1 (33.1), 27.4 (7.15), 260.9 (1.94), 492.4 (8.03), 527.9 (27.5)
336.2	Cd-114	In-115m	4.49 h		45.8	24.1 (27.7), 27.4 (6), 497.4 (0.047)
336.6	Ga-71	Ga-72	14.1 h	187.929	0.11	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
337.6	Ge-76	Ge-77	11.3 h	0.468	0.22	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
338.7	Ge-76	Ge-77	11.3 h	0.468	0.63	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
338.8	Ir-193	Ir-194m	171 d	376.2	55.1	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
340.1	Nd-150	Pm-151	28.40 h		22.4	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
340.2	Se-82	Se-83g	22.5 m	0.04888	0.41	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
341.6	Hf-176	Hf-177m_1	1.08 s		2.19	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
341.6	Lu-176	Hf-177m_1	1.08 s		2.19	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
341.6	Hf-176	Hf-177m_2	51.4 m		2.19	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
342.1	Pd-110	Ag-111	7.45 d		6.68	23.1 (0.18), 96.7 (0.2), 245.4 (1.24)
343.4	Hf-174	Hf-175	70 d	62.4	86.6	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 318.9 (0.17), 353.6 (0.23), 432.8 (1.56)
344.3	Eu-151	Eu-152	13.33 yr	2 82 020	26.6	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
344.3	Eu-151	Eu-152m_1	9.32 h	1 57 740	2.44	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 121.8 (7.16), 841.6 (14.5), 963.4 (11.9)
344.5	Cd-116	Cd-117	2.49 h	0.3735	17.9	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
344.7	Kr-78	Kr-79g	35.04 h	2.1	0.24	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
344.9	Nd-150	Pm-151	28.40 h		2.10	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
344.9	Zn-64	Zn-65	244.3 d	36.936	0.003	770.6 (0.003), 1115.5 (50.7)
345.9	Hf-180	Hf-181	42.4 d	443.52	14.0	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 133.0 (43.0), 136.2 (6.1), 482.0 (85.5)
346.5	Pt-196	Pt-197m	1.57 h	1.265	11.2	53.0 (1.08), 65.1 (13.6), 66.8 (23.8), 68.8 (0.40), 75.7 (8), 77.9 (2.26), 130.2 (0.1), 279.0 (2.3)
346.9	Co-59	Co-60	5.272 yr	1680	0.0076	1173.2 (99.9), 1332.5 (100)
347.6	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
348.2	Dy-158	Dy-159	144.4 d	4.3	0.001	43.7 (26.9), 44.5 (48.2), 50.3 (14.9), 51.7 (4.25), 58.0 (2.22), 79.5 (0.005)
348.2	Gd-158	Gd-159	18.56 h	62.1	0.22	43.7 (5.54), 44.5 (9.93), 50.3 (3.06), 51.7 (0.89), 58.0 (2.27), 226.0 (0.21), 363.6 (10.8)
350	Ru-104	Ru-105	4.44 h	5.984	0.28	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
350.2	Ru-104	Ru-105	4.44 h	5.984	1.0	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
350.6	Ce-142	Ce-143	33.0 h	10.526	3.27	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 490.4 (2.13), 664.5 (5.6), 722 (5.32)
351.2	Pt-190	Pt-191	2.90 d	1.95	3.36	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
353	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
353.6	Hf-174	Hf-175	70 d	62.4	0.23	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 318.9 (0.17), 343.4 (86.6), 432.8 (1.56)
355.4	Zr-96	Zr-97	17.0 h	0.0632	2.27	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
356	Ba-132	Ba-133	10.54 yr	0.6565	62.3	30.6 (34.4), 31 (63.5), 35 (18.8), 35.8 (4.24), 53.2 (2.2), 79.6 (2.43), 81 (32.8), 276.4 (7.29), 302.9 (18.6)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
356.7	Se-82	Se-83m	70.4 s	0.3666	17.3	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
356.7	Se-82	Se-83g	22.5 m	0.04888	68.6	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
357.5	Pd-102	Pd-103	16.96 d	3.468	0.0221	20.1 (63.8), 22.8 (13.2), 39.7 (0.0683), 357.5 (0.0221), 497.1 (0.004)
357.9	Pd-110	Pd-111m	5.5 h	0.43	0.40	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
358.1	Rh-103	Rh-104	42.3 s	13 500	0.016	19.2 (0.24), 21.7 (0.050), 555.8 (1.99), 1237 (0.066)
358.4	Xe-134	Xe-135	9.08 h	2.704	0.22	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 249.8 (90.4), 408 (0.36), 608.2 (2.91)
359.9	Pt-190	Pt-191	2.90 d	1.95	6.00	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 409.4 (8.00), 456.5 (3.36), 538.9 (13.7)
359.9	Th-232	Th-233	22.3 m	740	0.12	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
360.3	Te-126	Te-127m	109 d	2.55	0.132	27.2 (10.4), 27.5 (19.3), 28.3 (0.4), 28.6 (0.75), 31 (5.56), 31.7 (1.13), 32.3 (0.22), 57.6 (0.5), 88.3 (0.084), 202.9 (0.057), 417.9 (0.969)
360.3	Te-126	Te-127	9.35 h	17.055	0.135	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 202.9 (0.058), 215.1 (0.039), 417.9 (0.993)
360.9	Gd-160	Gd-161	3.70 m	16.832	60.6	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 165.2 (2.60), 283.6 (6), 314.9 (22.9), 480.1 (2.70)
361.1	Os-189	Os-190m	9.90 m	0.0041	95.2	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 186.7 (69.9), 502.6 (97.8), 616.1 (98.5)
361.7	Dy-164	Dy-165	2.33 h	28 200	0.84	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
361.7	Dy-164	Dy-165m	1.26 m	47 940	0.534	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 108.2 (3.01), 153.8 (0.242), 515.5 (1.53)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
362.6	Hf-178	Hf-179m_2	25.1 d	867.2	38.5	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
362.8	Kr-84	Kr-85g	10.72 yr	2.394	0.0001	151.2 (0.0001), 514.0 (0.434)
363.6	Gd-158	Gd-159	18.56 h	62.1	10.8	43.7 (5.54), 44.5 (9.93), 50.3 (3.06), 51.7 (0.89), 58.0 (2.27), 226.0 (0.21), 348.2 (0.22)
364.5	Te-130	I-131	8.02 d		81.2	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 284.3 (6.06), 637 (7.27), 722.9 (1.8)
365.4	Ag-109	Ag-110m	249.9 d	226.4	0.11	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
365.7	Ho-165	Ho-166m	1.2×10^3 yr	350	2.55	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
366.3	Ni-64	Ni-65	2.52 h	1.3832	4.61	507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
366.5	Mo-98	Mo-99	2.75 d	3.1369	1.16	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
366.9	Cd-116	Cd-117m	3.36 h	0.1867	3.33	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
367.4	Lu-176	Lu-177m	160.9 d	18.2	3.17	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
367.4	Ge-76	Ge-77	11.3 h	0.468	13.3	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
367.9	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
370.1	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
371.6	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
371.6	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
373.2	Ba-130	Ba-131	11.8 d	0.9328	14.1	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 486.5 (2.09), 496.3 (47.1)
374.1	Hg-198	Hg-199m	42.6 m	0.1818	13.8	68.9 (19.0), 70.8 (32.3), 80.2 (11.1), 82.5 (3.08), 158.4 (52.5), 413.4 (0.027)
374.8	Hf-178	Hf-179m$_1$	18.7 s	1436.3	0.005	54.6 (16), 55.8 (28.1), 63.2 (9.39), 65.0 (2.4), 160.7 (2.79), 214.1 (95.2)
375	Xe-126	Xe-127	36.41	0.27	17.2	28.3 (25), 28.6 (46.5), 32.3 (13.5), 33 (2.81), 57.6 (1.22), 145.3 (4.29), 172.1 (25.5), 202.9 (68.3)
376.7	Pd-110	Pd-111m	5.5 h	0.43	0.97	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
376.7	Pd-110	Pd-111	23.4 m	2.22	0.44	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
377.9	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
378.5	Lu-176	Lu-177m	160.9 d	18.2	29.9	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
378.5	Lu-176	Hf-177m$_1$	1.08 s		39.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
378.5	Hf-176	Hf-177m$_1$	1.08 s		39.3	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
378.5	Hf-176	Hf-177m$_2$	51.4 m		39.3	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
379.3	Mo-100	Mo-101	14.6 m	1.916	0.32	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
381.2	Mo-100	Mo-101	14.6 m	1.916	0.30	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
381.2	Ga-71	Ga-72	14.1 h	187.929	0.27	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
381.4	Sn-122	Sn-123m	40.08 m	0.828	0.042	26.1 (3.06), 26.4 (5.72), 29.7 (1.63), 30.4 (0.32), 160.3 (85.6), 541.8 (0.02)
385	Lu-176	Lu-177m	160.9 d	18.2	3.16	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
385	Hf-176	Hf-177m_2	51.4 m		4.15	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
385	Hf-176	Hf-177m_1	1.08 s		4.15	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
385	Lu-176	Hf-177m_1	1.08 s		4.15	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
386.3	Zn-70	Zn-71m	3.94 h	0.005	93.0	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
388.4	Sr-86	Sr-87m	2.80 h	8.2824	81.8	13.4 (0.15), 14.1 (8.75), 15.9 (1.58)
389	Kr-78	Kr-79g	35.04 h	2.1	1.52	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
389.2	Se-82	Se-83g	22.5 m	0.04888	0.62	225.2 (31.9), 356.7 (68.6), 485.7 (2.26), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
389.9	Zn-70	Zn-71m	3.94 h	0.005	2.6	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
390	Zn-70	Zn-71g	2.4 m	0.0514	3.8	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
390.8	Ir-193	Ir-194m	171 d	376.2	35.1	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
391.3	Pd-110	Pd-111m	5.5 h	0.43	5.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
391.4	Se-82	Se-83m	70.4 s	0.3666	0.10	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
391.7	Sn -112	Sn-113	115.1 d	0.71	64.2	24.1 (79.8), 27.4 (17.3), 255 (1.85), 638 (0.001)
391.7	Sn -112	In-113m	1.66 h		64.2	24.1 (20.1), 27.4 (4.35)
391.9	Pt-198	Pt-199m	13.6 s	0.194	84.7	32 (2.78), 65.1 (2.01), 66.8 (3.45), 75.7 (1.18), 77.9 (0.31)
393.3	Mo-100	Tc-101	14.2 m		0.10	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
393.4	Ru-104	Ru-105	4.44 h	5.984	3.73	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
393.4	Xe-136	Xe-137	3.83 m	2.314	0.14	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
396.3	Yb-174	Yb-175	4.19 d	2067	6.55	53.0 (1.15), 54.1 (2.02), 61.2 (0.67), 63.0 (0.17), 113.8 (1.93), 137.7 (0.12), 144.9 (0.33), 282.5 (3.08)
397.6	Kr-78	Kr-79g	35.04 h	2.1	9.50	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
398.6	Zn-70	Zn-71g	2.4 m	0.0514	0.61	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
398.9	Mo-100	Mo-101	14.6 m	1.916	0.90	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
400.4	Zr-96	Zr-97	17.0 h	0.0632	0.32	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
400.6	Se-74	Se-75	119.8 d	46.62	11.6	66.1 (1.14), 96.7 (3.48), 121.1 (17.3), 136 (59), 198.6 (1.47), 264.6 (59.1), 279.5 (25.2), 303.9 (1.34)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
401.1	Lu-175	Lu-176	3.7×10^{10} yr	681.8	0.84	54.6 (9.45), 55.8 (16.5), 63.2 (5.53), 65.0 (1.42), 88.3 (13.1), 201.8 (84.7), 306.9 (93.3)
402.3	Nd-150	Nd-151	12.4 m	6.768	1.97	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
402.6	Kr-86	Kr-87	76.3 m	0.0519	49.6	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
408	Xe-134	Xe-135	9.08 h	2.704	0.36	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 249.8 (90.4), 358.4 (0.22), 608.2 (2.91)
408.7	Mo-100	Mo-101	14.6 m	1.916	1.60	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
409.1	Hg-196	Au-197m	7.80 s		0.11	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 279.0 (70.9)
409.1	Pt-196	Au-197m	7.80 s		0.11	67.0 (7.10), 68.8 (12.1), 77.3 (0.30), 77.9 (4.18), 80.2 (1.13), 130.2 (3.12), 201.8 (1.1), 279.0 (70.9)
409.4	Pt-190	Pt-191	2.90 d	1.95	8.00	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 456.5 (3.36), 538.9 (13.7)
409.8	Hf-178	Hf-179m_2	25.1 d	867.2	20.9	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
410.9	Ho-165	Ho-166m	1.2×10^3 yr	350	11.7	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
411.1	Eu-151	Eu-152	13.33 yr	2 82 020	2.23	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
411.8	Au-197	Au-198	2.70 d		95.5	68.9 (0.81), 70.8 (1.38), 80.2 (0.47), 82.5 (0.13), 675.9 (1.06), 1087.7 (0.23)
413.4	Hg-198	Hg-199m	42.6 m	0.1818	0.027	68.9 (19.0), 70.8 (32.3), 80.2 (11.1), 82.5 (3.08), 158.4 (52.5), 374.1 (13.8)
413.5	Pd-110	Pd-111m	5.5 h	0.43	1.7	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
413.5	Ru-104	Ru-105	4.44 h	5.984	2.22	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
413.7	Lu-176	Lu-177m	160.9 d	18.2	17.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
415.5	Pd-110	Pd-111m	5.5 h	0.43	1.5	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
415.6	Hg-204	Hg-205	5.20 m	2.924	0.13	203.7 (2.2), 1218.7 (0.006)
416.3	Ge-76	Ge-77	11.3 h	0.468	20.6	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
416.9	In-115	In-116m$_1$	54.15 m	7753.32	29.2	138.3 (3.29), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
417.9	Te-126	Te-127m	109 d	2.55	0.969	27.2 (10.4), 27.5 (19.3), 28.3 (0.4), 28.6 (0.75), 31 (5.56), 31.7 (1.13), 32.3 (0.22), 57.6 (0.5), 88.3 (0.084), 202.9 (0.057), 360.3 (0.132)
417.9	Te-126	Te-127	9.35 h	17.055	0.993	28.3 (0.031), 28.6 (0.057), 32.3 (0.017), 57.6 (0.03), 202.9 (0.058), 215.1 (0.039), 360.3 (0.135)
418.5	Hf-176	Hf-177m$_1$	1.08 s		28.1	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
418.5	Hf-176	Hf-177m$_2$	51.4 m		28.1	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
418.5	Lu-176	Lu-177m	160.9 d	18.2	21.4	63.2 (20.1), 113.0 (21.7), 153.3 (16.6), 208.4 (59.4), 228.5 (36.6), 281.8 (14.0), 319.0 (10.4), 327.7 (18.1), 378.5 (29.9), 413.7 (17.4), 418.5 (21.4)
418.5	Lu-176	Hf-177m$_1$	1.08 s		28.1	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
419.1	Ge-74	Ge-75	83 m	12.41	0.18	66.0 (0.11), 198.6 (1.17), 264.6 (11.3), 468.8 (0.22), 617.7 (0.11)
419.8	Ge-76	Ge-77	11.3 h	0.468	1.16	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
421.7	Mo-100	Mo-101	14.6 m	1.916	0.56	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
423.5	Nd-148	Nd-149	1.73 h	14.4	7.48	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
423.5	Nd-150	Nd-151	12.4 m	6.768	6.43	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
426.4	Hf-177	Hf-178m$_2$	31 yr	0.00000186	96.9	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
426.4	Hf-177	Hf-178m$_1$	4.0 s	20.46	96.9	54.6 (20.1), 55.8 (35.2), 63.2 (11.8), 65.0 (3.01), 88.9 (62), 93.2 (17.3), 213.4 (80.9), 325.6 (93.9)
427.9	Sn-124	Sb-125	2.73 yr		29.4	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
428.4	Ga-71	Ga-72	14.1 h	187.929	0.20	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
432.5	La-139	La-140	40.28 h	892.12	2.99	328.8 (20.7), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
432.6	Mo-100	Mo-101	14.6 m	1.916	0.12	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
432.8	Hf-174	Hf-175	70 d	62.4	1.56	53.0 (26.4), 54.1 (46.3), 61.2 (15.4), 63.0 (3.96), 89.4 (2.34), 113.8 (0.31), 229.6 (0.77), 318.9 (0.17), 343.4 (86.6), 353.6 (0.23)
433.2	Ce-136	Ce-137	9.0 h	1.197	0.065	10.6 (0.6), 33.0 (21.8), 33.4 (40.1), 37.8 (11.9), 38.7 (2.92), 436.6 (0.334), 447.2 (2.24)
433.9	Ag-107	Ag-108	2.41 m	1932.14	0.48	21.1 (1.29), 23.9 (0.27), 618.9 (0.25), 633 (1.75)
433.9	Ag-107	Ag-108m	127 yr	17.1	90.7	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 614.4 (91.2), 633.0 (0.15), 722.9 (91.3)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
434.2	Cd-116	Cd-117	2.49 h	0.3735	9.79	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
436.1	Er-162	Er-163	75.0 m	2.66	0.028	46.7 (22.4), 47.5 (39.8), 53.8 (12.5), 55.3 (3.52), 439.9 (0.027), 1113.5 (0.049)
436.6	Ce-136	Ce-137	9.0 h	1.197	0.334	10.6 (0.6), 33.0 (21.8), 33.4 (40.1), 37.8 (11.9), 38.7 (2.92), 433.2 (0.065), 447.2 (2.24)
438.6	Zn-68	Zn-69m	13.76 h	1.353	94.8	8.6 (1.90), 9.6 (0.26), 318.6 (0.0013), 574.1 (0.033)
439.3	Pd-110	Pd-111m	5.5 h	0.43	0.23	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
439.4	Ge-76	Ge-77	11.3 h	0.468	0.19	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
439.9	Er-162	Er-163	75.0 m	2.66	0.027	46.7 (22.4), 47.5 (39.8), 53.8 (12.5), 55.3 (3.52), 436.1 (0.028), 1113.5 (0.049)
439.9	Ne-22	Ne-23	37.2 s	0.42412	33.0	1636.5 (1), 2076.4 (0.1)
441	Th-232	Th-233	22.3 m	740	0.23	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
442.4	Se-82	Se-83g	22.5 m	0.04888	1.09	225.2 (31.9), 356.7 (68.6), 485.7 (2.26), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
443.2	Hf-179	Hf-180m	5.5 h	4.67	82.8	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 215.2 (81.4), 332.3 (94.4), 500.7 (14.5)
443.8	Ru-102	Ru-103	39.25 d	38.24	0.32	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 497.1 (89.5), 557.0 (0.832), 610.3 (5.64)
444	Eu-151	Eu-152	13.33 yr	2 82 020	3.12	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
445.7	Nd-150	Pm-151	28.40 h		4.01	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
446.8	Ag-109	Ag-110m	249.9 d	226.4	3.66	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
447.2	Ce-136	Ce-137m	34.4 h	0.1805	3.03	33 (29.5), 33.4 (54.2), 34.3 (15.6), 34.7 (28.7), 37.8 (16.1), 38.7 (3.94), 39.2 (8.53), 40.2 (2.2), 254.3 (10.9)
447.2	Ce-136	Ce-137	9.0 h	1.197	2.24	10.6 (0.6), 33.0 (21.8), 33.4 (40.1), 37.8 (11.9), 38.7 (2.92), 433.2 (0.065), 436.6 (0.334)
447.7	Th-232	Th-233	22.3 m	740	0.15	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
448.7	Mo-100	Mo-101	14.6 m	1.916	0.69	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
451.5	Ho-165	Ho-166m	1.2×10^3 yr	350	3.09	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
451.6	Se-82	Se-83g	22.5 m	0.04888	0.82	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
452.3	Te-130	Te-131m	30 h	0.676	5.6	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
452.3	Te-130	Te-131	25.0 m	9.126	18.2	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)
453.1	Zn-70	Zn-71g	2.4 m	0.0514	0.18	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
453.1	Zn-70	Zn-71m	3.94 h	0.005	1.1	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
453.3	Tl-205	Tl-206m	3.75 m		81.0	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
453.7	Hf-178	Hf-179m_2	25.1 d	867.2	66.0	122.7 (26.9), 146.1 (26.3), 169.8 (18.9), 192.8 (20.9), 217.0 (8.78), 236.6 (18.3), 268.9 (11), 316.0 (19.7), 362.6 (38.5), 409.8 (20.9), 453.7 (66)
453.8	Xe-124	Xe-125	16.9 h	13.7	4.23	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 188.4 (54.9), 243.4 (28.8), 846.5 (1.03)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
454	Hf-177	Hf-178m$_2$	31 yr	0.00000186	16.3	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
454.4	Pd-110	Pd-111m	5.5 h	0.43	1.3	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
455.5	Xe-136	Xe-137	3.83 m	2.314	31.2	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
456.5	Pt-190	Pt-191	2.90 d	1.95	3.36	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 538.9 (13.7)
457.2	Tl-205	Tl-206m	3.75 m		21.5	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
457.4	Se-82	Se-83g	22.5 m	0.04888	3.50	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
459.2	Th-232	Th-233	22.3 m	740	0.15	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
459.6	Te-128	Te-129m	33.6 d	0.4755	4.54	27.2 (8.18), 27.5 (15.2), 27.8 (10.4), 31 (4.39), 695.9 (2.9)
459.6	Te-128	Te-129	69.6 m	6.34	7.14	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
460.5	Os-192	Os-193	30.5 h	82	3.95	63.3 (3.74), 64.9 (6.46), 73.0 (3.2), 73.5 (2.22), 138.9 (4.27)
460.6	Ru-96	Ru-97	2.9 d	1.6	0.12	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 215.7 (85.8), 324.5 (10.2), 569.3 (0.88)
461.4	Ge-76	Ge-77	11.3 h	0.468	1.20	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
463.1	In-115	In-116	14.10 s	3828.8	0.25	1252.6 (0.031), 1293.5 (1.30)
463.4	Sn-124	Sb-125	2.73 yr		10.5	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
465.8	Lu-176	Hf-177m$_1$	1.08 s		3.11	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
465.8	Hf-176	Hf-177m$_1$	1.08 s		3.11	105.4 (16), 113.0 (27.3), 128.5 (20.2), 153.3 (21.8), 174.4 (16.4), 204.2 (17.8), 208.4 (73), 228.5 (48), 281.8 (18.4), 327.7 (23.7), 378.5 (39.3), 418.5 (28.1)
465.8	Hf-176	Hf-177m$_2$	51.4 m		3.11	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
468.1	Ir-191	Ir-192m$_2$	241 yr	3.73	48.1	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
468.1	Ir-191	Ir-192	73.83 d	23 275.2	48.1	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
468.8	Ge-74	Ge-75	83 m	12.41	0.22	66.0 (0.11), 198.6 (1.17), 264.6 (11.3), 419.1 (0.18), 617.7 (0.11)
469	Mo-100	Mo-101	14.6 m	1.916	0.12	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
469.4	Ru-104	Ru-105	4.44 h	5.984	17.3	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
469.8	Sn-124	Sn-125	9.64 d	0.0224	1.38	332.1 (1.31), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
470.1	Ru-104	Ru-105	4.44 h	5.984	0.18	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
470.5	Te-120	Te-121m	154.0 d	0.032	1.4	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
470.5	Te-120	Te-121	16.8 d	0.192	1.4	26.1 (21.5), 26.4 (48.6), 29.7 (11.5), 30.4 (2.26), 37.1 (0.12), 65.5 (0.26), 507.6 (17.7), 573.1 (80.3)
472.7	Se-82	Se-83g	22.5 m	0.04888	0.14	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
475.4	Ge-76	Ge-77	11.3 h	0.468	0.94	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
475.4	Cs-133	Cs-134	2.06 yr	2650	1.46	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
478	Re-187	Re-188	16.98 h	4569.8	1.04	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
479.5	W-186	W-187	23.9 h	1081.08	25.3	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 551.6 (5.89), 618.4 (7.27), 685.8 (31.6), 772.9 (4.77)
479.5	Y-89	Y-90m	3.19 h	0.1	90.6	202.5 (96.5), 682 (0.32)
480.1	Gd-160	Gd-161	3.70 m	16.832	2.70	43.7 (13.6), 44.5 (24.5), 50.3 (7.56), 51.7 (2.15), 56.3 (3.80), 102.3 (14), 165.2 (2.60), 283.6 (6), 314.9 (22.9), 360.9 (60.6)
482	Hf-180	Hf-181	42.4 d	443.52	85.5	56.3 (9.50), 57.5 (16.6), 65.2 (5.59), 133.0 (43.0), 136.2 (6.1), 345.9 (14.0)
482.9	Ir-193	Ir-194m	171 d	376.2	96.9	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
484.5	Cd-114	Cd-115m	44.6 d	1.03	0.290	933.8 (2), 1290.6 (0.89)
484.6	Ir-191	Ir-192	73.83 d	23 275.2	2.17	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
484.6	Ir-191	$Ir\text{-}192m_2$	241 yr	3.73	3.16	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 588.6 (4.58), 604.4 (8.33), 612.4 (5.43)
485.7	Se-82	Se-83g	22.5 m	0.04888	2.26	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
485.9	Pd-110	Pd-111m	5.5 h	0.43	0.49	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
486.5	Ba-130	Ba-131	11.8 d	0.9328	2.09	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 496.3 (47.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
487	La-139	La-140	40.28 h	892.12	45.9	328.8 (20.7), 432.5 (2.99), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
487.3	Zn-70	Zn-71g	2.4 m	0.0514	0.12	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
487.3	Zn-70	Zn-71m	3.94 h	0.005	62.3	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
487.4	Te-128	Te-129	69.6 m	6.34	1.32	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 802.1 (0.18), 1083.9 (0.46), 1111.6 (0.18)
489.2	Ca-46	Ca-47	4.54 d	0.00296	6.74	530.4 (0.1), 767 (0.19), 807.9 (6.89), 1297.1 (74.9)
489.5	Ru-104	Ru-105	4.44 h	5.984	0.54	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
490.4	Ce-142	Ce-143	33.0 h	10.526	2.13	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 350.6 (3.27), 664.5 (5.6), 722 (5.32)
490.8	Th-232	Th-233	22.3 m	740	0.17	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
492.4	Sm-144	Sm-145	340 d	2.17	0.0031	38.2 (39.7), 38.7 (72), 43.8 (21.6), 44.9 (6), 61.3 (12.4)
492.4	Cd-114	Cd-115	2.23 d	86.16	8.03	24.1 (33.1), 27.4 (7.15), 260.9 (1.94), 336.2 (49.7), 527.9 (27.5)
492.7	Te-130	Te-131	25.0 m	9.126	4.84	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 602 (4.2), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)
493.8	Pt-198	Pt-199	30.8 m	26.64	5.73	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 543.0 (14.8), 714.5 (1.86)
495	Hf-177	Hf-178m_2	31 yr	0.00000186	68.7	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
496.3	Ba-130	Ba-131	11.8 d	0.9328	47.1	30.6 (27.9), 31 (51.6), 35 (15.2), 123.8 (29.2), 133.6 (2.17), 216.1 (19.9), 239.6 (2.42), 249.4 (2.84), 373.2 (14.1), 486.5 (2.09)
497	Mo-100	Mo-101	14.6 m	1.916	0.56	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
497.1	Ru-102	Ru-103	39.25 d	38.24	89.5	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 443.8 (0.32), 557.0 (0.832), 610.3 (5.64)
497.1	Pd-102	Pd-103	16.96 d	3.468	0.0040	20.1 (63.8), 22.8 (13.2), 39.7 (0.0683), 357.5 (0.0221), 497.1 (0.004)
497.4	Cd-114	In-115m	4.49 h		0.047	24.1 (27.7), 27.4 (6), 336.2 (45.8)
498.4	Sb-123	Sb-124m_2	20.2 m	0.811	22.0	602.7 (22), 645.8 (22), 1101.0 (0.3)
498.4	Sb-123	Sb-124m_1	60.2 d	1.58	20.0	602.7 (20), 645.8 (20), 1101.0 (0.3)
499	Th-232	Th-233	22.3 m	740	0.21	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
499.3	Ru-104	Ru-105	4.44 h	5.984	2.03	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
499.7	Mo-100	Mo-101	14.6 m	1.916	1.47	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
500.1	Ru-104	Ru-105	4.44 h	5.984	0.55	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
500.7	Hf-179	Hf-180m	5.5 h	4.67	14.5	54.6 (9.87), 55.8 (17.2), 57.5 (48.4), 63.2 (5.78), 65.0 (1.48), 93.3 (17), 215.2 (81.4), 332.3 (94.4), 443.2 (82.8)
502.6	Os-189	Os-190m	9.90 m	0.0041	97.8	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 186.7 (69.9), 361.1 (95.2), 616.1 (98.5)
505.1	Mo-100	Mo-101	14.6 m	1.916	1.2	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
505.9	Mo-100	Mo-101	14.6 m	1.916	11.8	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
507.6	Te-120	Te-121m	154.0 d	0.032	17.7	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
507.6	Zr-96	Zr-97	17.0 h	0.0632	5.05	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
507.6	Te-120	Te-121	16.8 d	0.192	17.7	26.1 (21.5), 26.4 (48.6), 29.7 (11.5), 30.4 (2.26), 37.1 (0.12), 65.5 (0.26), 470.5 (1.4), 573.1 (80.3)
507.8	Ni-64	Ni-65	2.52 h	1.3832	0.29	366.6 (4.61), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
508.8	Pr-141	Pr-142	19.13 h	760	0.023	641.2 (0.002), 1575.7 (3.7)
508.9	Pd-110	Pd-111m	5.5 h	0.43	0.16	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
508.9	Pd-110	Pd-111	23.4 m	2.22	0.21	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
510	Se-82	Se-83g	22.5 m	0.04888	44.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
510.1	Mo-100	Mo-101	14.6 m	1.916	0.98	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
511	Cu-63	Cu-64	12.7 h	311.265	35.8	1345.8 (0.48)
511.6	Zn-70	Zn-71m	3.94 h	0.005	28.4	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
511.6	Zn-70	Zn-71g	2.4 m	0.0514	32.0	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
512.8	Mo-100	Mo-101	14.6 m	1.916	1.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
513.4	Zr-96	Zr-97	17.0 h	0.0632	0.55	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
513.7	Ru-104	Ru-105	4.44 h	5.984	0.20	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
514	Kr-84	Kr-85g	10.72 yr	2.394	0.434	151.2 (0.0001), 362.8 (0.0001)
514	Sr-84	Sr-85g	64.84 d	0.196	99.3	13.4 (50.2), 15.0 (8.80), 151.2 (0.001), 868.1 (0.12)
514.1	Mo-100	Mo-101	14.6 m	1.916	0.81	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
515.4	Mo-100	Mo-101	14.6 m	1.916	0.51	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
515.5	Dy-164	Dy-165m	1.26 m	47 940	1.53	45.2 (2.67), 46.0 (4.77), 46.7 (0.10), 47.5 (0.18), 52.1 (1.49), 53.5 (0.42), 108.2 (3.01), 153.8 (0.242), 361.7 (0.534)
516.1	Mo-100	Tc-101	14.2 m		0.11	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
519.3	Pd-110	Pd-111m	5.5 h	0.43	0.14	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
520.3	Ge-76	Ge-77	11.3 h	0.468	0.28	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
520.4	Se-82	Br-83	2.39 h		0.067	529.6 (1.4), 552.6 (0.017)
520.8	Ge-76	As-77	38.83 h		0.42	88 (0.27), 162 (0.16), 239.1 (1.59), 250 (0.41)
523	Kr-78	Kr-79g	35.04 h	2.1	0.25	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
524.2	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
525.3	Kr-78	Kr-79g	35.04 h	2.1	0.43	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
525.6	Pd-110	Pd-111m	5.5 h	0.43	1.2	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
526.6	Xe-134	Xe-135m	15.6 m	0.0312	80.5	29.5 (3.92), 29.8 (7.28), 33.6 (2.14), 34.4 (0.47), 786.9 (0.004)
527.4	Hf-176	Hf-177m_2	51.4 m		7.14	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
527.9	Cd-114	Cd-115	2.23 d	86.16	27.5	24.1 (33.1), 27.4 (7.15), 260.9 (1.94), 336.2 (49.7), 492.4 (8.03)
529.6	Se-82	Br-83	2.39 h		1.40	520.4 (0.067), 552.6 (0.017)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
529.8	Ho-165	Ho-166m	1.2×10^3 yr	350	10.3	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
530.4	Ca-46	Ca-47	4.54 d	0.00296	0.1	489.2 (6.74), 767 (0.19), 807.9 (6.89), 1297.1 (74.9)
531	Nd-146	Nd-147	10.98 d	24.066	13.0	38.2 (13), 38.7 (23.5), 43.8 (7.05), 44.9 (1.96), 91.1 (27.9), 319.4 (1.95)
531.4	Mo-100	Tc-101	14.2 m		1.02	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
533.1	Se-82	Se-83g	22.5 m	0.04888	4.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
533.5	Mo-100	Mo-101	14.6 m	1.916	0.40	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
535	Hf-177	Hf-178m_2	31 yr	0.00000186	8.86	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
538.9	Pt-190	Pt-191	2.90 d	1.95	13.7	10 (36), 63.3 (38.3), 64.9 (66.1), 73.5 (22.7), 75.6 (5.85), 82.4 (4.9), 96.5 (3.28), 129.4 (3.2), 172.2 (3.52), 351.2 (3.36), 359.9 (6.00), 409.4 (8.00), 456.5 (3.36)
539.3	Ru-104	Ru-105	4.44 h	5.984	0.11	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
540.5	Nd-148	Nd-149	1.73 h	14.4	6.58	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
541.8	Sn-122	Sn-123m	40.08 m	0.828	0.020	26.1 (3.06), 26.4 (5.72), 29.7 (1.63), 30.4 (0.32), 160.3 (85.6), 381.4 (0.042)
543	Pt-198	Pt-199	30.8 m	26.64	14.8	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 714.5 (1.86)
545.1	Mo-100	Tc-101	14.2 m		5.98	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
545.8	Dy-164	Dy-165	2.33 h	28 200	0.16	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						361.7 (0.84), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
547	Pd-110	Pd-111m	5.5 h	0.43	0.27	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
547	Pd-110	Pd-111	23.4 m	2.22	0.37	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
551.6	W-186	W-187	23.9 h	1081.08	5.89	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 618.4 (7.27), 685.8 (31.6), 772.9 (4.77)
552.2	Pd-110	Pd-111m	5.5 h	0.43	0.26	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
552.4	Se-80	Se-81m	57.3 m	3.968	0.15	103.1 (9.79), 275.9 (1.3), 290.1 (1.0), 566 (0.38), 828.3 (0.47)
552.4	Se-80	Se-81	18.5 m	26.288	0.10	275.9 (0.87), 290.1 (0.68), 566.0 (0.26), 828.3 (0.32)
552.6	Se-82	Br-83	2.39 h		0.017	520.4 (0.067), 529.6 (1.4)
552.9	Cd-116	Cd-117m	3.36 h	0.1867	125.0	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
554.3	Br-81	Br-82g	35.30 h	12.8206	70.6	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
555.8	Rh-103	Rh-104m	4.34 m	1000	2.38	19.2 (0.28), 20.2 (55.1), 22.8 (11.4), 51.4 (48.2), 77.5 (2.07), 97.1 (3.0), 767.8 (0.10)
555.8	Rh-103	Rh-104	42.3 s	13 500	1.99	19.2 (0.24), 21.7 (0.050), 358.1 (0.016), 1237 (0.066)
556.1	Rb-85	Rb-86m	1.02 m	3.82501	98.2	13.4 (0.88), 15.0 (0.16)
556.4	Pd-110	Pd-111m	5.5 h	0.43	0.19	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
557	Ru-102	Ru-103	39.25 d	38.24	0.832	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 443.8 (0.32), 497.1 (89.5), 610.3 (5.64)
558	In-113	In-114m	49.51 d	34.668	3.40	23.1 (3.91), 24.1 (28.1), 27.4 (6.08), 191.6 (16), 725.2 (3.4)
558	In-113	In-114	71.9 s	16.69	0.07	23.1 (2), 26.2 (0.42), 575.7 (0.004), 1300 (0.14)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
558	Ge-76	Ge-77	11.3 h	0.468	15.2	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
559.1	As-75	As-76	26.32 h	450	44.7	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
559.2	Ru-104	Ru-105	4.44 h	5.984	0.11	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
561.9	Zr-94	Nb-95	34.98 d		0.015	204.1 (0.015), 765.8 (99.9)
562.4	Ir-193	Ir-194m	171 d	376.2	69.9	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
563.2	Cs-133	Cs-134	2.06 yr	2650	8.38	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
563.2	As-75	As-76	26.32 h	450	1.17	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
564	Sb-121	Sb-122	2.70 d	335.2	70.8	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 692.6 (3.68), 1141.1 (0.57), 1256.8 (0.77)
564.2	Tl-205	Tl-206m	3.75 m		10.0	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
564.4	Cd-116	Cd-117m	3.36 h	0.1867	14.7	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
565.7	Dy-164	Dy-165	2.33 h	28 200	0.13	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 620.6 (0.10), 633.4 (0.57), 715.3 (0.53)
566	Se-80	Se-81	18.5 m	26.288	0.26	275.9 (0.87), 290.1 (0.68), 552.4 (0.10), 828.3 (0.32)
566	Se-80	Se-81m	57.3 m	3.968	0.38	103.1 (9.79), 275.9 (1.3), 290.1 (1.0), 552.4 (0.15), 828.3 (0.47)
566.2	Zn-70	Zn-71m	3.94 h	0.005	0.19	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
566.6	Mo-100	Mo-101	14.6 m	1.916	0.73	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
569.3	Ru-96	Ru-97	2.9 d	1.6	0.88	18.3 (58.5), 20.7 (11.7), 108.8 (0.11), 215.7 (85.8), 324.5 (10.2), 460.6 (0.12)
569.3	Cs-133	Cs-134	2.06 yr	2650	15.4	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
571	Ho-165	Ho-166m	1.2×10^3 yr	350	5.81	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
571.3	As-75	As-76	26.32 h	450	0.14	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
571.7	Mo-100	Mo-101	14.6 m	1.916	0.19	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
571.9	Se-82	Se-83g	22.5 m	0.04888	4.46	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
573.1	Te-120	Te-121m	154.0 d	0.032	80.3	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
573.1	Te-120	Te-121	16.8 d	0.192	80.3	26.1 (21.5), 26.4 (48.6), 29.7 (11.5), 30.4 (2.26), 37.1 (0.12), 65.5 (0.26), 470.5 (1.4), 507.6 (17.7)
574.1	Zn-68	Zn-69m	13.76 h	1.353	0.033	8.6 (1.90), 9.6 (0.26), 318.6 (0.0013), 438.6 (94.8)
574.2	Hf-177	Hf-178m_2	31 yr	0.00000186	83.6	88.9 (62), 93.2 (17.3), 213.4 (80.9), 216.7 (63.7), 237.4 (8.79), 257.6 (16.6), 296.8 (9.83), 325.6 (93.9), 426.4 (96.9), 454.0 (16.3), 495.0 (68.7), 535.0 (8.86), 574.2 (83.6)
574.9	Zn-70	Zn-71m	3.94 h	0.005	0.11	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
575	Pd-110	Pd-111m	5.5 h	0.43	3.0	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
575.1	Ru-104	Ru-105	4.44 h	5.984	0.84	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
575.7	In-113	In-114	71.9 s	16.69	0.004	23.1 (2), 26.2 (0.42), 558 (0.07), 1300 (0.14)
580	Pd-110	Pd-111m	5.5 h	0.43	0.63	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
580	Pd-110	Pd-111	23.4 m	2.22	0.837	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
581.6	Se-82	Se-83g	22.5 m	0.04888	0.34	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
582.1	Zr-94	Nb-95m	3.61 d		0.055	16.6 (36), 18.6 (6.97), 204.1 (2.36), 235.7 (25.1)
582.5	Ge-76	Ge-77	11.3 h	0.468	0.74	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
583.6	Pd-110	Pd-111m	5.5 h	0.43	0.25	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
586.3	Te-130	Te-131m	30 h	0.676	1.97	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
587.4	Ga-71	Ga-72	14.1 h	187.929	0.12	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
588.6	Ir-191	Ir-192m_2	241 yr	3.73	4.58	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 604.4 (8.33), 612.4 (5.43)
588.6	Ir-191	Ir-192	73.83 d	23 275.2	4.58	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 604.4 (8.33), 612.4 (5.43)
589.2	Ir-193	Ir-194	19.15 h	6897	0.14	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
589.6	Sn-124	Sn-125m	9.52 m	0.728	0.2	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 643 (0.16), 1404 (0.7), 1483.9 (0.18), 1615.3 (0.12)
590.1	Mo-100	Mo-101	14.6 m	1.916	5.6	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
590.9	Mo-100	Mo-101	14.6 m	1.916	16.4	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
590.9	Nd-148	Pm-149	2.21 d		0.68	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 285.9 (3.10), 830.5 (0.03), 833.2 (0.03)
591.8	Eu-153	Eu-154	8.8 yr		4.83	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
593.4	Se-82	Se-83g	22.5 m	0.04888	0.75	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
595.2	Th-232	Th-233	22.3 m	740	0.16	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
595.4	Pd-110	Pd-111m	5.5 h	0.43	0.12	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
596.1	Zn-70	Zn-71m	3.94 h	0.005	27.9	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
600.5	Ir-193	Ir-194m	171 d	376.2	62.3	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
600.6	Sn-124	Sb-125	2.73 yr		17.8	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
600.9	Ga-71	Ga-72	14.1 h	187.929	5.54	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
602	Te-130	Te-131	25.0 m	9.126	4.20	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 948.5 (2.26), 997.2 (3.34), 1147 (4.96)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
602.4	Zr-96	Zr-97	17.0 h	0.0632	1.4	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
602.7	Sb-123	Sb-124	60.2 d	175.07	98.4	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
602.7	Sb-123	$Sb\text{-}124m_1$	60.2 d	1.58	20.0	498.4 (20), 645.8 (20), 1101.0 (0.3)
602.7	Sb-123	$Sb\text{-}124m_2$	20.2 m	0.811	22.0	498.4 (22), 645.8 (22), 1101.0 (0.3)
603	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
604.4	Ir-191	$Ir\text{-}192m_2$	241 yr	3.73	8.33	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 612.4 (5.43)
604.4	Ir-191	Ir-192	73.83 d	23 275.2	8.33	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 604.4 (8.33), 612.4 (5.43)
604.7	Cs-133	Cs-134	2.06 yr	2650	97.6	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
606.1	Kr-78	Kr-79g	35.04 h	2.1	8.10	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
606.3	Br-81	Br-82g	35.30 h	12.8206	1.17	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
606.5	Hf-176	$Hf\text{-}177m_2$	51.4 m		11.6	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
606.6	Sn-124	Sb-125	2.73 yr		5.02	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
606.8	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
608.2	Xe-134	Xe-135	9.08 h	2.704	2.91	30.6 (1.46), 31 (2.7), 35 (0.8), 35.8 (0.18), 158.2 (0.29), 249.8 (90.4), 358.4 (0.22), 408 (0.36)
608.3	Mo-100	Mo-101	14.6 m	1.916	1.07	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
608.5	Ti-50	Ti-51	5.8 m	0.9308	1.18	320.1 (93), 928.6 (6.88)
609.2	Se-82	Se-83g	22.5 m	0.04888	3.09	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
609.3	Ni-64	Ni-65	2.52 h	1.3832	0.14	366.6 (4.61), 507.8 (0.29), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
610.3	Ru-102	Ru-103	39.25 d	38.24	5.64	20.1 (7.11), 22.8 (1.47), 53.3 (0.38), 295.0 (0.25), 443.8 (0.32), 497.1 (89.5), 557.0 (0.832)
611.6	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
612.4	Ir-191	Ir-192m_2	241 yr	3.73	5.43	63.0 (2.71), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (3.16), 588.6 (4.58), 604.4 (8.33)
612.4	Ir-191	Ir-192	73.83 d	23 275.2	5.43	63.0 (2.17), 65.1 (2.71), 66.8 (4.65), 205.8 (3.20), 295.9 (28.7), 308.4 (29.7), 316.5 (82.9), 468.1 (48.1), 484.6 (2.17), 588.6 (4.58), 604.4 (8.33)
614.4	Ag-107	Ag-108m	127 yr	17.1	91.2	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 433.9 (90.7), 633.0 (0.15), 722.9 (91.3)
614.4	Ge-76	Ge-77	11.3 h	0.468	0.56	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
616.1	Os-189	Os-190m	9.90 m	0.0041	98.5	61.5 (5.71), 63.0 (9.85), 71.3 (3.39), 186.7 (69.9), 361.1 (95.2), 502.6 (97.8)
616.2	Br-79	Br-80g	17.4 m	435.934	6.70	639.2 (0.24), 665.6 (1.15), 704.0 (0.19)
616.2	Br-79	Br-80m	4.42 h	121.656	7.17	37.1 (39.1), 48.9 (0.33), 639.2 (0.26), 665.6 (1.23), 704 (0.2)
617.7	Ge-74	Ge-75	83 m	12.41	0.11	66.0 (0.11), 198.6 (1.17), 264.6 (11.3), 419.1 (0.18), 468.8 (0.22)
618.4	W-186	W-187	23.9 h	1081.08	7.27	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 551.6 (5.89), 685.8 (31.6), 772.9 (4.77)
618.9	Ag-107	Ag-108	2.41 m	1932.14	0.25	21.1 (1.29), 23.9 (0.27), 433.9 (0.48), 633 (1.75)
619.1	Br-81	Br-82g	35.30 h	12.8206	43.1	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
620.1	Pd-110	Pd-111	23.4 m	2.22	0.11	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
620.1	Pd-110	Ag-111m	64.8 s		0.11	22.1 (15.6), 25.0 (3.30), 59.8 (0.531), 171.3 (0.11), 245.4 (0.46)
620.2	Zn-70	Zn-71m	3.94 h	0.005	56.7	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
620.3	Ag-109	Ag-110m	249.9 d	226.4	2.78	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
620.6	Dy-164	Dy-165	2.33 h	28 200	0.10	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 633.4 (0.57), 715.3 (0.53)
621.6	Se-82	Se-83g	22.5 m	0.04888	3.22	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
622	Ir-193	Ir-194	19.15 h	6897	0.33	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
623.2	Pd-110	Pd-111m	5.5 h	0.43	0.59	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
623.2	Pd-110	Pd-111	23.4 m	2.22	0.28	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
624.8	Ge-76	Ge-77	11.3 h	0.468	0.17	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
625.6	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
626.2	Ag-109	Ag-110m	249.9 d	226.4	0.23	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
627	Mo-100	Tc-101	14.2 m		0.39	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
630	Ga-71	Ga-72	14.1 h	187.929	24.8	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
631.2	Se-82	Se-83m	70.4 s	0.3666	0.47	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
631.8	Cd-116	Cd-117m	3.36 h	0.1867	2.80	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
631.8	Ge-76	Ge-77	11.3 h	0.468	6.59	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
632.3	Ru-104	Ru-105	4.44 h	5.984	0.15	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
632.5	Ba-132	Ba-133m	38.9 h	0.0505	0.01	12.3 (1.5), 31.8 (15.3), 32.2 (28.2), 36.4 (8.34), 37.3 (1.99), 276.1 (17.5)
632.8	Pd-110	Pd-111m	5.5 h	0.43	3.4	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
633	Ag-107	Ag-108	2.41 m	1932.14	1.75	21.1 (1.29), 23.9 (0.27), 433.9 (0.48), 618.9 (0.25)
633	Ag-107	Ag-108m	127 yr	17.1	0.15	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 433.9 (90.7), 614.4 (91.2), 722.9 (91.3)
633	Re-187	Re-188	16.98 h	4569.8	1.25	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
633.4	Dy-164	Dy-165	2.33 h	28 200	0.57	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 715.3 (0.53)
634.4	Ge-76	Ge-77	11.3 h	0.468	1.97	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
635.1	Re-187	Re-188	16.98 h	4569.8	0.15	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
635.9	Sn-124	Sb-125	2.73 yr		11.3	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
637	Te-130	I-131	8.02 d		7.27	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 284.3 (6.06), 364.5 (81.2), 722.9 (1.8)
638	Sn -112	Sn-113	115.1 d	0.71	0.0010	24.1 (79.8), 27.4 (17.3), 255 (1.85), 391.7 (64.2)
638.2	Hf-176	Hf-177m_2	51.4 m		20.1	214.0 (40.9), 228.5 (48), 277.3 (75.8), 295.1 (69), 299.0 (23.6), 305.5 (22.38), 311.5 (58.8), 326.7 (65.4), 378.5 (39.3), 418.5 (28.1), 638.2 (20.1)
638.7	Ru-104	Ru-105	4.44 h	5.984	0.22	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
639.2	Br-79	Br-80g	17.4 m	435.934	0.24	616.2 (6.70), 665.6 (1.15), 704.0 (0.19)
639.2	Br-79	Br-80m	4.42 h	121.656	0.26	37.1 (39.1), 48.9 (0.33), 616.2 (7.17), 665.6 (1.23), 704 (0.2)
641.2	Pr-141	Pr-142	19.13 h	760	0.002	508.8 (0.023), 1575.7 (3.7)
642.8	Mo-100	Mo-101	14.6 m	1.916	1.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
643	Sn-124	Sn-125m	9.52 m	0.728	0.16	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 1404 (0.7), 1483.9 (0.18), 1615.3 (0.12)
645.1	Ir-193	Ir-194	19.15 h	6897	1.16	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
645.6	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
645.8	Sb-123	Sb-124	60.2 d	175.07	7.24	602.7 (98.4), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
645.8	Sb-123	Sb-124m_1	60.2 d	1.58	20.0	498.4 (20), 602.7 (20), 1101.0 (0.3)
645.8	Sb-123	Sb-124m_2	20.2 m	0.811	22.0	498.4 (22), 602.7 (22), 1101.0 (0.3)
646.1	Os-184	Os-185	93.6 d	60	80.8	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 717.4 (4.11), 874.8 (6.59), 880.5 (4.98)
647.3	Pd-108	Pd-109	13.46 h	219.6	0.024	22.1 (28.5), 25.0 (6.02), 88.0 (3.61), 311.4 (0.032)
649.8	Bi-209	Bi-210m	3.0×10^6 yr	1.4	2.9	70.8 (3.9), 72.9 (6.6), 82.5 (2.3), 265.7 (51.0), 304.8 (28.0)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
650.4	Pd-110	Pd-111m	5.5 h	0.43	0.41	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
650.4	Pd-110	Pd-111	23.4 m	2.22	0.55	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
652.7	Ru-104	Ru-105	4.44 h	5.984	0.30	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
654.7	Pd-110	Pd-111m	5.5 h	0.43	0.14	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
654.8	Nd-148	Nd-149	1.73 h	14.4	7.95	114.3 (19), 155.9 (5.93), 211.3 (25.9), 240.2 (3.94), 267.7 (6.03), 270.2 (10.7), 326.5 (4.56), 423.5 (7.48), 540.5 (6.58), 654.8 (7.95)
656.2	Ru-104	Ru-105	4.44 h	5.984	2.03	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
657	As-75	As-76	26.32 h	450	0.14	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
657.7	Ag-109	Ag-110m	249.9 d	226.4	94.7	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
657.7	Ag-109	Ag-110	24.6 s	4157.07	4.49	21.1 (0.17), 23.9 (0.037), 815.3 (0.038), 1125.7 (0.015)
657.9	Zr-96	Zr-97	17.0 h	0.0632	106.0	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
657.9	Zr-96	Nb-97	72.1 m		98.5	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
660.7	Mo-100	Mo-101	14.6 m	1.916	0.22	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
661.6	Xe-136	Cs-137	30.14 yr		85.1	31.8 (2.02), 32.2 (3.72), 36.4 (1.1), 37.3 (0.26)
661.6	Xe-136	Ba-137m	2.55 m		89.9	31.8 (2.13), 32.2 (3.93), 36.4 (1.16), 37.3 (0.28)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
661.6	Ba-136	Ba-137m	2.55 m	0.0785	89.9	31.8 (2.13), 32.2 (3.93), 36.4 (1.16), 37.3 (0.28)
664.5	Ce-142	Ce-143	33.0 h	10.526	5.60	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 350.6 (3.27), 490.4 (2.13), 722 (5.32)
664.8	Se-82	Se-83g	22.5 m	0.04888	3.22	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
665	Te-130	Te-131m	30 h	0.676	4.32	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
665.3	As-75	As-76	26.32 h	450	0.39	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
665.6	Br-79	Br-80m	4.42 h	121.656	1.23	37.1 (39.1), 48.9 (0.33), 616.2 (7.17), 639.2 (0.26), 704 (0.2)
665.6	Br-79	Br-80g	17.4 m	435.934	1.15	616.2 (6.70), 639.2 (0.24), 704.0 (0.19)
666.8	Zn-70	Zn-71g	2.4 m	0.0514	0.90	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
668.5	Pd-110	Pd-111m	5.5 h	0.43	0.93	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
669.8	Th-232	Th-233	22.3 m	740	0.68	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
670.5	Ho-165	Ho-166m	1.2×10^3 yr	350	5.83	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
671.4	Sn-124	Sb-125	2.73 yr		1.80	27.2 (21.5), 27.5 (40), 31.7 (2.34), 35.5 (5.9), 176.3 (6.79), 427.9 (29.4), 463.4 (10.5), 600.6 (17.8), 606.6 (5.02), 635.9 (11.3), 671.4 (1.8)
672.5	Re-187	Re-188	16.98 h	4569.8	0.11	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 829.5 (0.41), 931.3 (0.56), 1610.4 (0.10)
673.1	Ge-76	Ge-77	11.3 h	0.468	0.63	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
673.8	Kr-86	Kr-87	76.3 m	0.0519	1.89	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
673.9	Se-82	Se-83m	70.4 s	0.3666	15.1	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
675.9	Au-197	Au-198	2.70 d		1.06	68.9 (0.81), 70.8 (1.38), 80.2 (0.47), 82.5 (0.13), 411.8 (95.5), 1087.7 (0.23)
676.4	Ru-104	Ru-105	4.44 h	5.984	15.5	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
676.6	Ag-109	Ag-110m	249.9 d	226.4	0.14	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
677.6	Ag-109	Ag-110m	249.9 d	226.4	10.7	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
677.8	Nd-150	Nd-151	12.4 m	6.768	2.72	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
679.4	Se-82	Se-83g	22.5 m	0.04888	1.10	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
680.5	Tl-205	Tl-206m	3.75 m		79.3	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 803.1 (0.0055), 1021.5 (62.9), 1139.9 (5.2)
682	Y-89	Y-90m	3.19 h	0.1	0.32	202.5 (96.5), 479.5 (90.6)
684.7	Mo-92	Mo-93m	6.85 h	0.089	99.7	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 949.8 (0.12), 1363.0 (0.79), 1477.1 (99.1)
685.8	W-186	W-187	23.9 h	1081.08	31.6	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 551.6 (5.89), 618.4 (7.27), 772.9 (4.77)
687	Ag-109	Ag-110m	249.9 d	226.4	6.49	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
687.8	Ir-193	Ir-194m	171 d	376.2	59.1	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
690.6	Zr-96	Zr-97	17.0 h	0.0632	0.25	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
692.6	Sb-121	Sb-122	2.70 d	335.2	3.68	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 564 (70.8), 1141.1 (0.57), 1256.8 (0.77)
694.2	Pd-110	Pd-111m	5.5 h	0.43	1.9	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
695.6	Mo-100	Mo-101	14.6 m	1.916	7.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
695.9	Te-128	Te-129m	33.6 d	0.4755	2.90	27.2 (8.18), 27.5 (15.2), 27.8 (10.4), 31 (4.39), 459.6 (4.54)
698.3	Br-81	Br-82m	6.13 m	119.8233	0.024	46.0 (0.24), 776.5 (0.2), 1474.8 (0.016)
698.3	Br-81	Br-82g	35.30 h	12.8206	27.9	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
698.5	Ge-76	Ge-77	11.3 h	0.468	0.22	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
699.2	Zr-96	Zr-97	17.0 h	0.0632	0.12	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
701.8	Mo-100	Mo-101	14.6 m	1.916	0.38	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
702.6	Nb-93	Nb-94	2.03×10^4 yr	100	100.0	871.1 (100)
702.6	Nb-93	Nb-94m	6.26 m	15	0.0030	16.6 (37.1), 18.6 (7.18), 41.0 (0.08), 871.1 (0.48)
703.8	Pd-110	Pd-111m	5.5 h	0.43	0.63	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
703.8	Zr-96	Zr-97	17.0 h	0.0632	0.93	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
704	Br-79	Br-80m	4.42 h	121.656	0.20	37.1 (39.1), 48.9 (0.33), 616.2 (7.17), 639.2 (0.26), 665.6 (1.23)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
704	Br-79	Br-80g	17.4 m	435.934	0.19	616.2 (6.70), 639.2 (0.24), 665.6 (1.15)
705.2	Ge-76	Ge-77	11.3 h	0.468	0.10	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
706.1	Se-82	Se-83g	22.5 m	0.04888	0.2	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
706.7	Ag-109	Ag-110m	249.9 d	226.4	16.7	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
708.1	Ag-109	Ag-110m	249.9 d	226.4	0.28	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
709.8	Pd-110	Pd-111	23.4 m	2.22	0.13	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
711.7	Ho-165	Ho-166m	1.2×10^3 yr	350	59.3	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
712.1	Se-82	Se-83g	22.5 m	0.04888	2.67	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
712.3	Ge-76	Ge-77	11.3 h	0.468	0.78	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
713	Mo-100	Mo-101	14.6 m	1.916	3.38	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
713.8	Sb-123	Sb-124	60.2 d	175.07	2.40	602.7 (98.4), 645.8 (7.24), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
714.3	Ge-76	Ge-77	11.3 h	0.468	6.77	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
714.5	Pt-198	Pt-199	30.8 m	26.64	1.86	68.8 (2.0), 185.8 (3.26), 191.7 (2.38), 246.5 (2.16), 317.0 (4.87), 493.8 (5.73), 543.0 (14.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
715.3	Dy-164	Dy-165	2.33 h	28 200	0.53	46.7 (2.66), 47.5 (4.74), 53.8 (1.49), 55.3 (0.42), 94.7 (3.58), 279.8 (0.50), 361.7 (0.84), 545.8 (0.16), 565.7 (0.13), 620.6 (0.10), 633.4 (0.57)
715.6	Mo-100	Tc-101	14.2 m		0.69	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
717.4	Os-184	Os-185	93.6 d	60	4.11	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 646.1 (80.8), 874.8 (6.59), 880.5 (4.98)
717.6	Nd-150	Pm-151	28.40 h		3.99	100.0 (2.53), 104.8 (3.52), 167.7 (7.84), 177.2 (3.58), 240.1 (3.61), 275.2 (6.61), 340.1 (22.4), 344.9 (2.1), 445.7 (4.01), 717.6 (3.99)
718	Se-82	Se-83g	22.5 m	0.04888	16.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
718.9	Pd-110	Pd-111m	5.5 h	0.43	0.17	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
720	Mo-100	Tc-101	14.2 m		0.23	127.2 (2.86), 184.1 (1.69), 306.8 (88.0), 531.4 (1.02), 545.1 (5.98), 627.0 (0.39), 715.6 (0.69), 720.0 (0.23), 842.8 (0.23), 928.6 (0.13)
721.4	Zn-70	Zn-71g	2.4 m	0.0514	0.54	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
722	Ce-142	Ce-143	33.0 h	10.526	5.32	35.5 (18.5), 36 (33.8), 40.7 (10.1), 41.8 (2.67), 57.4 (12.3), 231.6 (2.04), 293.3 (42.8), 350.6 (3.27), 490.4 (2.13), 664.5 (5.6)
722.8	Sb-123	Sb-124	60.2 d	175.07	11.3	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
722.9	Te-130	I-131	8.02 d		1.80	29.5 (1.36), 29.8 (2.53), 80.2 (2.62), 284.3 (6.06), 364.5 (81.2), 637 (7.27)
722.9	Ag-107	Ag-108m	127 yr	17.1	91.3	21.1 (54.3), 22.1 (1.35), 23.9 (11.5), 25.0 (0.28), 79.1 (6.46), 433.9 (90.7), 614.4 (91.2), 633 (0.15)
723.3	Eu-153	Eu-154	8.8 yr		19.7	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
724.2	Ru-104	Ru-105	4.44 h	5.984	46.7	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
724.2	Zr-94	Zr-95	64.03 d	0.86227	43.7	16.6 (0.42), 18.6 (0.66), 235.7 (0.24), 756.7 (55.4)
724.8	Pd-110	Pd-111m	5.5 h	0.43	0.27	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
725.2	In-113	In-114m	49.51 d	34.668	3.4	23.1 (3.91), 24.1 (28.1), 27.4 (6.08), 191.6 (16), 558 (3.4)
733	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
735.1	Se-82	Se-83g	22.5 m	0.04888	0.82	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
735.7	Ga-71	Ga-72	14.1 h	187.929	0.37	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
736.4	Nd-150	Nd-151	12.4 m	6.768	7.19	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
739.4	Mo-98	Mo-99	2.75 d	3.1369	12.1	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 777.8 (4.36), 822.8 (0.13), 961.0 (0.10)
739.5	Mo-100	Mo-101	14.6 m	1.916	0.30	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
740.1	As-75	As-76	26.32 h	450	0.12	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
742	Ce-142	Pr-143	13.58 d		0.0001	–
743.4	Zr-96	Nb-97m	53 s		98.0	16.6 (1.10), 18.7 (0.21)
743.4	Zr-96	Zr-97	17.0 h	0.0632	92.6	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
743.6	Ge-76	Ge-77	11.3 h	0.468	0.17	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
744.3	Ag-109	Ag-110m	249.9 d	226.4	4.66	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
745.6	Pd-110	Pd-111m	5.5 h	0.43	0.12	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
745.7	Ge-76	Ge-77	11.3 h	0.468	0.91	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
748.1	Cd-116	Cd-117m	3.36 h	0.1867	4.5	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
749.9	Ge-76	Ge-77	11.3 h	0.468	0.84	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
751.8	La-139	La-140	40.28 h	892.12	4.41	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
752.3	Ho-165	Ho-166m	1.2×10^3 yr	350	13.2	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
753	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
753.4	Zn-70	Zn-71m	3.94 h	0.005	3.3	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
754.2	Ce-138	Ce-139m	56.4 s	0.00375	92.4	34.3 (1.58), 34.7 (2.9), 39.2 (0.86), 40.2 (0.22)
756.7	Zr-94	Zr-95	64.03 d	0.86227	55.4	16.6 (0.42), 18.6 (0.66), 235.7 (0.24), 724.2 (43.7)
756.9	Eu-153	Eu-154	8.8 yr		40.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
762.2	Pd-110	Pd-111m	5.5 h	0.43	1.2	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
763.9	Ag-109	Ag-110m	249.9 d	226.4	22.4	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
764.4	Th-232	Th-233	22.3 m	740	0.12	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
765.3	Tb-159	Tb-160	72.3 d	2550	2.04	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
765.8	Zr-94	Nb-95	34.98 d		99.9	204.1 (0.015), 561.9 (0.015)
766.7	Ge-76	Ge-77	11.3 h	0.468	0.74	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
767	Ca-46	Ca-47	4.54 d	0.00296	0.19	489.2 (6.74), 530.4 (0.1), 807.9 (6.89), 1297.1 (74.9)
767.5	Re-185	Re-186	3.78 d	4188.8	0.029	58.0 (1.55), 59.3 (2.69), 61.5 (1.11), 63.0 (1.92), 67.2 (0.91), 69.1 (0.23), 71.3 (0.66), 73.4 (0.17), 122.7 (0.72), 137.2 (9.20)
767.8	Rh-103	Rh-104m	4.34 m	1000	0.10	19.2 (0.28), 20.2 (55.1), 22.8 (11.4), 51.4 (48.2), 77.5 (2.07), 97.1 (3.0), 555.8 (2.38)
770.6	Zn-64	Zn-65	244.3 d	36.936	0.003	344.9 (0.003), 1115.5 (50.7)
771.3	Zn-70	Zn-71m	3.94 h	0.005	2.0	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
771.8	As-75	As-76	26.32 h	450	0.12	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
772.9	W-186	W-187	23.9 h	1081.08	4.77	59.7 (8.09), 61.1 (14), 69.2 (4.78), 72.5 (12.9), 134.2 (10.3), 479.5 (25.3), 551.6 (5.89), 618.4 (7.27), 685.8 (31.6)
773.7	Te-130	Te-131m	30 h	0.676	38.1	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
774.2	Mo-100	Mo-101	14.6 m	1.916	0.33	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
775.8	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
776.5	Br-81	Br-82m	6.13 m	119.8233	0.20	46.0 (0.24), 698.3 (0.024), 1474.8 (0.016)
776.5	Br-81	Br-82g	35.30 h	12.8206	83.4	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
777.8	Mo-98	Mo-99	2.75 d	3.1369	4.36	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 822.8 (0.13), 961.0 (0.10)
778.3	Mo-100	Mo-101	14.6 m	1.916	0.96	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
778.8	Ho-165	Ho-166m	1.2×10^3 yr	350	3.33	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
778.9	Eu-151	Eu-152	13.33 yr	2 82 020	13.0	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
781.3	Ge-76	Ge-77	11.3 h	0.468	0.96	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
782.5	Te-130	Te-131m	30 h	0.676	7.76	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
784.8	Ge-76	Ge-77	11.3 h	0.468	1.24	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
786.4	Ga-71	Ga-72	14.1 h	187.929	3.20	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
786.9	Xe-134	Xe-135m	15.6 m	0.0312	0.004	29.5 (3.92), 29.8 (7.28), 33.6 (2.14), 34.4 (0.47), 526.6 (80.5)
790.1	Mo-100	Mo-101	14.6 m	1.916	0.13	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
793.8	Te-130	Te-131m	30 h	0.676	13.8	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
794.3	Ge-76	Ge-77	11.3 h	0.468	0.26	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
795.8	Cs-133	Cs-134	2.06 yr	2650	85.4	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
796.5	Cd-106	Cd-107	6.50 h	0.25	0.065	22.1 (89), 25 (18.8), 93.1 (4.68), 796.5 (0.065), 828.9 (0.163)
797.5	Nd-150	Nd-151	12.4 m	6.768	5.51	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
797.8	Pd-110	Pd-111m	5.5 h	0.43	0.97	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
799	Se-82	Se-83g	22.5 m	0.04888	16.0	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
799	Se-82	Se-83m	70.4 s	0.3666	1.23	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
800.3	Sn-124	Sn-125	9.64 d	0.0224	1.0	332.1 (1.31), 469.8 (1.38), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
801.9	Cs-133	Cs-134	2.06 yr	2650	8.73	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 1038.6 (1), 1167.9 (1.8), 1365.1 (3.04)
802.1	Te-128	Te-129	69.6 m	6.34	0.18	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 1083.9 (0.46), 1111.6 (0.18)
803.1	Tl-205	Tl-206	4.20 m		0.0055	–
803.1	Bi-209	Po-210	138.38 d		0.0012	10.5 (0.013), 12.6 (0.11), 14.9 (0.0026)
803.1	Bi-209	Tl-206	4.20 m		0.0055	–
804.3	Mo-100	Mo-101	14.6 m	1.916	1.00	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
804.5	Zr-96	Zr-97	17.0 h	0.0632	0.65	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
807.9	Ca-46	Ca-47	4.54 d	0.00296	6.89	489.2 (6.74), 530.4 (0.1), 767 (0.19), 1297.1 (74.9)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
810.2	Ga-71	Ga-72	14.1 h	187.929	2.01	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
810.3	Ho-165	Ho-166m	1.2×10^3 yr	350	63.3	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
810.4	Ge-76	Ge-77	11.3 h	0.468	2.15	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
813.4	Ge-76	Ge-77	11.3 h	0.468	0.12	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
814.3	Kr-86	Kr-87	76.3 m	0.0519	0.16	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
815.3	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
815.3	Ag-109	Ag-110	24.6 s	4157.07	0.038	21.1 (0.17), 23.9 (0.037), 657.7 (4.49), 1125.7 (0.015)
815.8	La-139	La-140	40.28 h	892.12	23.6	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
818	Ag-109	Ag-110m	249.9 d	226.4	7.32	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
818.7	In-115	In-116m$_1$	54.15 m	7753.32	11.5	138.3 (3.29), 416.9 (29.2), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
822	Ru-104	Ru-105	4.44 h	5.984	845.9	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
822.5	Sn-124	Sn-125	9.64 d	0.0224	3.99	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
822.8	Mo-98	Mo-99	2.75 d	3.1369	0.13	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 961.0 (0.10)
822.8	Te-130	Te-131m	30 h	0.676	6.10	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
823.1	Ge-76	Ge-77	11.3 h	0.468	0.57	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
826.2	Co-59	Co-60m	10.47 m	2040	0.0083	6.9 (27.2), 7.7 (3.67), 58.6 (2.02), 1332.5 (0.25)
827.8	Br-81	Br-82g	35.30 h	12.8206	24.2	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
828.3	Se-80	Se-81m	57.3 m	3.968	0.47	103.1 (9.79), 275.9 (1.3), 290.1 (1.0), 552.4 (0.15), 566.0 (0.38)
828.3	Se-80	Se-81	18.5 m	26.288	0.32	275.9 (0.87), 290.1 (0.68), 552.4 (0.10), 566 (0.26)
828.3	Pd-110	Pd-111m	5.5 h	0.43	0.16	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
828.9	Cd-106	Cd-107	6.50 h	0.25	0.163	22.1 (89), 25 (18.8), 93.1 (4.68), 796.5 (0.065)
829.5	Re-187	Re-188	16.98 h	4569.8	0.41	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 931.3 (0.56), 1610.4 (0.10)
829.8	Zr-96	Zr-97	17.0 h	0.0632	0.22	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
830.5	Nd-148	Pm-149	2.21 d		0.030	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 285.9 (3.10), 590.9 (0.68), 833.2 (0.03)
830.6	Ho-165	Ho-166m	1.2×10^3 yr	350	10.7	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
831.8	Cd-116	Cd-117	2.49 h	0.3735	2.26	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
832	Kr-78	Kr-79g	35.04 h	2.1	1.26	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
833.2	Nd-148	Pm-149	2.21 d		0.030	39.5 (0.06), 40.1 (0.11), 45.4 (0.033), 285.9 (3.10), 590.9 (0.68), 830.5 (0.03)
833.4	Cu-65	Cu-66	5.1 m	66.9011	0.18	1039.2 (8), 1333.2 (0.003)
834	Ga-71	Ga-72	14.1 h	187.929	95.6	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
835.7	Pd-110	Pd-111m	5.5 h	0.43	0.20	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
835.7	Pd-110	Pd-111	23.4 m	2.22	0.27	22.1 (17.0), 25.0 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459 (0.56)
836.4	Kr-86	Kr-87	76.3 m	0.0519	0.77	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
836.5	Se-82	Se-83g	22.5 m	0.04888	15.9	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
841.6	Eu-151	Eu-152m$_1$	9.32 h	1 57 740	14.5	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 121.8 (7.16), 344.3 (2.44), 963.4 (11.9)
842.8	Mo-100	Tc-101	14.2 m		0.23	127.2 (2.86), 184.1 (1.69), 306.8 (88), 531.4 (1.02), 545.1 (5.98), 627 (0.39), 715.6 (0.69), 720 (0.23), 842.8 (0.23), 928.6 (0.13)
843.2	Ge-76	Ge-77	11.3 h	0.468	0.20	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
843.7	Mg-26	Mg-27	9.46 m	0.42058	71.8	170.7 (0.84), 1014.4 (28.2)
845.4	Kr-86	Kr-87	76.3 m	0.0519	7.34	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
845.9	Ru-104	Ru-105	4.44 h	5.984	0.62	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
846.5	Xe-124	Xe-125	16.9 h	13.7	1.03	28.3 (28.9), 28.6 (53.7), 32.3 (15.6), 33 (3.25), 55 (5.93), 188.4 (54.9), 243.4 (28.8), 453.8 (4.23)
846.8	Mn-55	Mn-56	2.58 h	1330	98.9	1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
848.9	Xe-136	Xe-137	3.83 m	2.314	0.62	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
852	Ru-104	Ru-105	4.44 h	5.984	0.15	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
852.2	Te-130	Te-131m	30 h	0.676	21.0	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
853.2	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
854.9	Zr-96	Zr-97	17.0 h	0.0632	0.33	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
856.1	Ca-48	Ca-49	8.72 m	0.2057	0.13	1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
859.1	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
860.4	Cd-116	Cd-117m	3.36 h	0.1867	7.89	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
861.1	Ga-71	Ga-72	14.1 h	187.929	0.91	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
862.8	Pd-110	Pd-111m	5.5 h	0.43	0.14	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
866.6	Se-82	Se-83g	22.5 m	0.04888	8.85	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
867.4	Eu-151	Eu-152	13.33 yr	2 82 020	4.21	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
867.6	As-75	As-76	26.32 h	450	0.13	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
867.8	La-139	La-140	40.28 h	892.12	5.59	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
868.1	Sr-84	Sr-85g	64.84 d	0.196	0.12	13.4 (50.2), 15.0 (8.80), 151.2 (0.001), 514 (99.3)
869.7	Mo-100	Mo-101	14.6 m	1.916	0.34	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
871.1	Nb-93	Nb-94m	6.26 m	15	0.48	16.6 (37.1), 18.6 (7.18), 41 (0.08), 702.6 (0.003)
871.1	Mo-100	Mo-101	14.6 m	1.916	1.80	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
871.1	Nb-93	Nb-94	2.03×10^4 yr	100	100.0	702.6 (100)
872	Zn-68	Zn-69	56 m	18.8	0.0002	318.6 (0.0012)
873.2	Eu-153	Eu-154	8.8 yr		11.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
874.8	Os-184	Os-185	93.6 d	60	6.59	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 646.1 (80.8), 717.4 (4.11), 880.5 (4.98)
875.2	Ge-76	Ge-77	11.3 h	0.468	0.74	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
875.9	Ru-104	Ru-105	4.44 h	5.984	2.47	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822 (845.9), 875.9 (2.47), 969.4 (2.08)
877.4	Mo-100	Mo-101	14.6 m	1.916	3.40	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
878.2	Ru-104	Ru-105	4.44 h	5.984	0.47	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
879.4	Tb-159	Tb-160	72.3 d	2550	30.0	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
880.5	Os-184	Os-185	93.6 d	60	4.98	59.7 (20.5), 61.1 (35.5), 69.2 (12), 71.2 (3.05), 646.1 (80.8), 717.4 (4.11), 874.8 (6.59)
880.7	Cd-116	Cd-117	2.49 h	0.3735	3.96	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
882.1	Pd-110	Pd-111m	5.5 h	0.43	0.20	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
883.5	Mo-100	Mo-101	14.6 m	1.916	0.70	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
883.6	Se-82	Se-83g	22.5 m	0.04888	7.75	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
884.7	Ag-109	Ag-110m	249.9 d	226.4	72.9	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
887	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
887.8	Se-82	Se-83g	22.5 m	0.04888	4.73	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
888.7	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
889.2	Sc-45	Sc-46	83.82 d	1740	100	4.1 (6.6), 4.5 (0.88), 142.5 (56), 1120.5 (100)
890.1	Th-232	Th-233	22.3 m	740	0.14	86.5 (2.7), 92.3 (0.6), 94.7 (0.8), 95.9 (0.98), 108.1 (0.36), 162.5 (0.32), 169.1 (0.34), 441 (0.23), 499 (0.21), 669.8 (0.68), 890.1 (0.14)
894.2	Ga-71	Ga-72	14.1 h	187.929	9.88	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
896.3	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
896.5	Ge-76	Ge-77	11.3 h	0.468	0.12	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
898	Rb-87	Rb-88	17.8 m	3.339	14.5	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
899	K-41	K-42	12.36 h	9.8258	0.053	312.7 (0.35), 1524.6 (18.6)
901	Ge-76	Ge-77	11.3 h	0.468	0.11	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
903.6	Mo-100	Mo-101	14.6 m	1.916	0.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
907	Ge-76	Ge-77	11.3 h	0.468	0.90	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
907.3	S-36	S-37	5.0 m	0.003	0.058	3103.3 (94.1), 3740.4 (0.23)
907.6	Ru-104	Ru-105	4.44 h	5.984	0.52	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
909.2	Sr-88	Sr-89	50.55 d	0.47908	0.0095	–
909.2	Sr-88	Y-89m	16.1 s		99.1	14.9 (0.45), 16.8 (0.083)
910	Te-130	Te-131m	30 h	0.676	3.28	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
910.1	Zn-70	Zn-71m	3.94 h	0.005	0.31	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
910.3	Zn-70	Zn-71g	2.4 m	0.0514	7.8	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
913.8	Ge-76	Ge-77	11.3 h	0.468	0.35	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
915.5	Sn-124	Sn-125	9.64 d	0.0224	3.85	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
919.6	La-139	La-140	40.28 h	892.12	2.68	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 925.2 (2.68), 1596.5 (95.4), 2521.7 (3.43)
923.1	Ge-76	Ge-77	11.3 h	0.468	0.65	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
924.2	Ga-71	Ga-72	14.1 h	187.929	0.14	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
925.2	La-139	La-140	40.28 h	892.12	2.68	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 1596.5 (95.4), 2521.7 (3.43)
925.5	Ge-76	Ge-77	11.3 h	0.468	0.74	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
928.6	Ti-50	Ti-51	5.8 m	0.9308	6.88	320.1 (93), 608.5 (1.18)
928.6	Mo-100	Tc-101	14.2 m		0.13	127.2 (2.86), 184.1 (1.69), 306.8 (88), 531.4 (1.02), 545.1 (5.98), 627 (0.39), 715.6 (0.69), 720 (0.23), 842.8 (0.23), 928.6 (0.13)
928.9	Ge-76	Ge-77	11.3 h	0.468	0.99	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
931.3	Re-187	Re-188	16.98 h	4569.8	0.56	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 1610.4 (0.10)
931.4	Cd-116	Cd-117m	3.36 h	0.1867	3.64	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
933.3	Mo-100	Mo-101	14.6 m	1.916	0.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
933.7	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
933.8	Cd-114	Cd-115m	44.6 d	1.03	2.00	484.5 (0.29), 1290.6 (0.89)
934.2	Mo-100	Mo-101	14.6 m	1.916	3.40	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
934.8	Kr-78	Kr-79g	35.04 h	2.1	0.13	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
937.5	Ag-109	Ag-110m	249.9 d	226.4	34.3	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
938.7	Ir-193	Ir-194	19.15 h	6897	0.59	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
939.4	Ga-71	Ga-72	14.1 h	187.929	0.26	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
939.4	Ge-76	Ge-77	11.3 h	0.468	0.27	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
943.3	Se-82	Se-83g	22.5 m	0.04888	0.89	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
943.5	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
944.7	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
946.7	Kr-86	Kr-87	76.3 m	0.0519	0.13	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
948.5	Te-130	Te-131	25.0 m	9.126	2.26	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 997.2 (3.34), 1147 (4.96)
949.8	Mo-92	Mo-93m	6.85 h	0.089	0.12	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 684.7 (99.7), 1363.0 (0.79), 1477.1 (99.1)
950.9	Ho-165	Ho-166m	1.2×10^3 yr	350	3.07	184.4 (73.9), 280.5 (30.1), 410.9 (11.7), 529.8 (10.3), 571.0 (5.81), 670.5 (5.83), 711.7 (59.3), 752.3 (13.2), 810.3 (63.3), 830.6 (10.7), 950.9 (3.07)
952.1	Br-81	Br-82g	35.30 h	12.8206	0.37	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
956.7	Zn-70	Zn-71m	3.94 h	0.005	0.19	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
961	Mo-98	Mo-99	2.75 d	3.1369	0.10	18.3 (9.68), 20.7 (1.94), 40.6 (1.05), 140.5 (90.7), 181.1 (6.08), 366.5 (1.16), 739.4 (12.1), 777.8 (4.36), 822.8 (0.13)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
962.3	Tb-159	Tb-160	72.3 d	2550	10.0	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
963.4	Eu-151	Eu-152m$_1$	9.32 h	1 57 740	11.9	39.5 (7.48), 40.1 (13.5), 45.4 (4.09), 121.8 (7.16), 344.3 (2.44), 841.6 (14.5)
964.1	Eu-151	Eu-152	13.33 yr	2 82 020	14.6	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
964.7	Zn-70	Zn-71m	3.94 h	0.005	4.7	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
964.8	Zn-70	Zn-71g	2.4 m	0.0514	0.77	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
966.2	Tb-159	Tb-160	72.3 d	2550	25.5	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
968.3	Sb-123	Sb-124	60.2 d	175.07	1.84	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
969.4	Ru-104	Ru-105	4.44 h	5.984	2.08	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822 (845.9), 875.9 (2.47), 969.4 (2.08)
970.5	Ga-71	Ga-72	14.1 h	187.929	1.10	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
971.4	Zr-96	Zr-97	17.0 h	0.0632	1.10	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
974.7	Zn-70	Zn-71m	3.94 h	0.005	0.35	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
975.2	Pd-110	Pd-111m	5.5 h	0.43	0.15	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
980.6	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
982.2	Xe-136	Xe-137	3.83 m	2.314	0.21	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
987.9	Se-82	Se-83g	22.5 m	0.04888	0.62	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
987.9	Se-82	Se-83m	70.4 s	0.3666	15.3	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
988.2	Mo-100	Mo-101	14.6 m	1.916	0.19	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
988.6	Zn-70	Zn-71m	3.94 h	0.005	1.2	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
995.9	Se-82	Se-83g	22.5 m	0.04888	1.3	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
996.3	Eu-153	Eu-154	8.8 yr		10.3	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
996.3	Pd-110	Pd-111m	5.5 h	0.43	0.25	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
996.5	Ge-76	Ge-77	11.3 h	0.468	0.10	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
997.2	Te-130	Te-131	25.0 m	9.126	3.34	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 1147 (4.96)
997.2	Ag-109	Ag-110m	249.9 d	226.4	0.13	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
997.6	Se-82	Se-83m	70.4 s	0.3666	1.28	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
999.9	Ga-71	Ga-72	14.1 h	187.929	0.80	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1001.7	Ta-181	Ta-182	115.0 d	2099.74	2.09	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1004.8	Eu-153	Eu-154	8.8 yr		17.9	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
1006.5	Zn-70	Zn-71m	3.94 h	0.005	0.74	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1007.4	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1007.6	Br-81	Br-82g	35.30 h	12.8206	1.31	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1011	Mo-100	Mo-101	14.6 m	1.916	2.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1011.4	Zn-70	Zn-71m	3.94 h	0.005	0.68	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1011.8	Ir-193	Ir-194m	171 d	376.2	3.6	111.7 (8.85), 324.0 (2.0), 328.4 (92.8), 338.8 (55.1), 390.8 (35.1), 482.9 (96.9), 562.4 (69.9), 600.5 (62.3), 687.8 (59.1), 1011.8 (3.6)
1012.5	Mo-100	Mo-101	14.6 m	1.916	12.8	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1014.4	Mg-26	Mg-27	9.46 m	0.42058	28.2	170.7 (0.84), 843.7 (71.8)
1016.4	Nd-150	Nd-151	12.4 m	6.768	2.92	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
1017.5	Ru-104	Ru-105	4.44 h	5.984	0.32	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
1018.6	Mo-100	Mo-101	14.6 m	1.916	0.64	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1020	Mo-100	Mo-101	14.6 m	1.916	0.47	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1020.6	Se-82	Se-83m	70.4 s	0.3666	1.97	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660 (1.78), 2051.4 (11)
1021	Sn-122	Sn-123	129.2 d	0.0046	0.002	160.3 (0.002), 1030.2 (0.031), 1088.6 (0.6)
1021.3	Zr-96	Zr-97	17.0 h	0.0632	1.34	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1021.5	Tl-205	Tl-206m	3.75 m		62.9	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1139.9 (5.2)
1022.5	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1024.5	Zr-96	Zr-97	17.0 h	0.0632	1.16	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1024.5	Zr-96	Nb-97	72.1 m		1.08	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
1025.7	Kr-78	Kr-79g	35.04 h	2.1	0.16	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
1029.1	Cd-116	Cd-117m	3.36 h	0.1867	11.7	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1030.2	Sn-122	Sn-123	129.2 d	0.0046	0.031	160.3 (0.002), 1021.0 (0.002), 1088.6 (0.6)
1030.5	Se-82	Se-83m	70.4 s	0.3666	20.9	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1036.5	Se-82	Se-83g	22.5 m	0.04888	0.2	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1038.6	Cs-133	Cs-134	2.06 yr	2650	1.00	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1167.9 (1.8), 1365.1 (3.04)
1039.2	Ga-69	Ga-70	21.15 m	100.968	0.673	176.2 (0.297)
1039.2	Cu-65	Cu-66	5.1 m	66.9011	8	833.4 (0.18), 1333.2 (0.003)
1042.1	Se-82	Se-83g	22.5 m	0.04888	1.03	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1044	Br-81	Br-82g	35.30 h	12.8206	27.4	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1045.2	Sb-123	Sb-124	60.2 d	175.07	1.85	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1368.2 (2.36), 1691.0 (49.0), 2091.0 (5.61)
1049.8	Mo-100	Mo-101	14.6 m	1.916	0.35	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1050.7	Ga-71	Ga-72	14.1 h	187.929	6.91	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1051.7	Cd-116	Cd-117	2.49 h	0.3735	3.79	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1303.3 (18.4), 1576.6 (11.2), 1723.1 (2.01)
1053.6	Se-82	Se-83m	70.4 s	0.3666	1.49	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1061.4	Lu-175	Lu-176m	3.65 h	1597.36	0.0007	54.6 (2.87), 55.8 (5.02), 63.2 (1.68), 65.0 (0.43), 82.1 (0.007), 88.3 (8.86), 1159.3 (0.0014)
1061.7	Ge-76	Ge-77	11.3 h	0.468	0.14	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1063.4	Se-82	Se-83m	70.4 s	0.3666	3.39	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1063.4	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1064.1	Se-82	Se-83g	22.5 m	0.04888	5.90	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1064.2	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1065.9	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1066	Cd-116	Cd-117m	3.36 h	0.1867	23.1	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1067.1	Sn-124	Sn-125	9.64 d	0.0224	9.05	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1087.7 (1.11), 1089.2 (4.28), 2001.8 (1.79)
1076.6	Rb-85	Rb-86	18.66 d	30.81659	8.78	–
1080.1	Yb-176	Yb-177	1.9 h		5.5	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 121.6 (3.4), 150.4 (20), 1241.4 (3.4)
1080.8	Ge-76	Ge-77	11.3 h	0.468	0.23	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1081.4	Br-81	Br-82g	35.30 h	12.8206	0.63	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1082	Se-82	Se-83g	22.5 m	0.04888	2.68	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1083.9	Te-128	Te-129	69.6 m	6.34	0.46	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1111.6 (0.18)
1085.2	Ge-76	Ge-77	11.3 h	0.468	5.72	211.0 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1085.9	Eu-151	Eu-152	13.33 yr	2 82 020	9.92	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
1087.7	Sn-124	Sn-125	9.64 d	0.0224	1.11	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1089.2 (4.28), 2001.8 (1.79)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1087.7	Au-197	Au-198	2.70 d		0.23	68.9 (0.81), 70.8 (1.38), 80.2 (0.47), 82.5 (0.13), 411.8 (95.5), 675.9 (1.06)
1088	Pd-110	Pd-111m	5.5 h	0.43	0.19	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1088.6	Sn-122	Sn-123	129.2 d	0.0046	0.6	160.3 (0.002), 1021.0 (0.002), 1030.2 (0.031)
1089.2	Sn-124	Sn-125	9.64 d	0.0224	4.28	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 2001.8 (1.79)
1097.3	In-115	In-116m$_1$	54.15 m	7753.32	56.2	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1293.5 (84.4), 1507.4 (9.96), 1753.8 (2.46)
1098.5	Pd-110	Pd-111m	5.5 h	0.43	0.1	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1099.3	Fe-58	Fe-59	44.503 d	0.3968	56.5	142.5 (1.03), 192.2 (3.11), 334.8 (0.26), 1291.6 (43.2)
1101	Sb-123	Sb-124m$_1$	60.2 d	1.58	0.3	498.4 (20), 602.7 (20), 645.8 (20)
1101	Sb-123	Sb-124m$_2$	20.2 m	0.811	0.3	498.4 (22), 602.7 (22), 645.8 (22)
1102.1	Te-120	Te-121m	154.0 d	0.032	2.54	26.1 (26), 26.4 (48.6), 27.5 (18.8), 29.7 (13.9), 65.5 (0.26), 212.2 (81.4), 470.5 (1.4), 507.6 (17.7), 573.1 (80.3), 1102.1 (2.54)
1107.4	Zn-70	Zn-71m	3.94 h	0.005	2.8	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1109.3	Zn-70	Zn-71g	2.4 m	0.0514	0.16	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1110.3	Se-82	Se-83g	22.5 m	0.04888	0.41	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1110.5	Zr-96	Zr-97	17.0 h	0.0632	0.11	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1111.6	Te-128	Te-129	69.6 m	6.34	0.18	27.8 (16.64), 209 (0.17), 250.6 (0.35), 278.4 (0.53), 281.3 (0.15), 459.6 (7.14), 487.4 (1.32), 802.1 (0.18), 1083.9 (0.46)
1112.1	Eu-151	Eu-152	13.33 yr	2 82 020	13.6	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
1113.5	Er-162	Er-163	75.0 m	2.66	0.049	46.7 (22.4), 47.5 (39.8), 53.8 (12.5), 55.3 (3.52), 436.1 (0.028), 439.9 (0.027)
1115.1	Kr-78	Kr-79g	35.04 h	2.1	0.37	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5), 606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
1115.5	Ni-64	Ni-65	2.52 h	1.3832	14.8	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1481.8 (23.5), 1623.4 (0.47), 1724.9 (0.39)
1115.5	Zn-64	Zn-65	244.3 d	36.936	50.7	344.9 (0.003), 770.6 (0.003)
1115.9	Pd-110	Pd-111m	5.5 h	0.43	1.0	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1116	Se-82	Se-83m	70.4 s	0.3666	0.54	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1119.3	Xe-136	Xe-137	3.83 m	2.314	0.11	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1120	Zn-70	Zn-71g	2.4 m	0.0514	2.2	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1120.5	Sc-45	Sc-46	83.82 d	1740	100	4.1 (6.6), 4.5 (0.88), 142.5 (56), 889.2 (100)
1121.3	Ta-181	Ta-182	115.0 d	2099.74	35.0	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1122.1	Nd-150	Nd-151	12.4 m	6.768	4.59	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
1125	Ge-76	Ge-77	11.3 h	0.468	0.11	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1125.5	Te-130	Te-131m	30 h	0.676	11.4	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
1125.7	Ag-109	Ag-110	24.6 s	4157.07	0.015	21.1 (0.17), 23.9 (0.037), 657.7 (4.49), 815.3 (0.038)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1129.9	As-75	As-76	26.32 h	450	0.14	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1139.8	Zn-70	Zn-71m	3.94 h	0.005	0.2	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1139.9	Tl-205	Tl-206m	3.75 m		5.2	216.4 (83.6), 247.2 (12), 265.7 (86.2), 316.8 (9.5), 453.3 (81), 457.2 (21.5), 564.2 (10), 680.5 (79.3), 803.1 (0.0055), 1021.5 (62.9)
1141.1	Sb-121	Sb-122	2.70 d	335.2	0.57	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 564 (70.8), 692.6 (3.68), 1256.8 (0.77)
1144.5	Ca-48	Ca-49	8.72 m	0.2057	0.11	856.1 (0.13), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
1147	Te-130	Te-131	25.0 m	9.126	4.96	28.3 (3.7), 28.6 (6.9), 32.3 (2), 149.7 (68.9), 452.3 (18.2), 492.7 (4.84), 602 (4.2), 948.5 (2.26), 997.2 (3.34)
1147.9	Zr-96	Zr-97	17.0 h	0.0632	2.64	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1148.9	Te-130	Te-131m	30 h	0.676	1.9	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
1150.8	Ir-193	Ir-194	19.15 h	6897	0.59	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
1151.8	Ge-76	Ge-77	11.3 h	0.468	0.18	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1159.3	Lu-175	Lu-176m	3.65 h	1597.36	0.0014	54.6 (2.87), 55.8 (5.02), 63.2 (1.68), 65.0 (0.43), 82.1 (0.007), 88.3 (8.86), 1061.4 (0.0007)
1161	Mo-100	Mo-101	14.6 m	1.916	3.97	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1163.3	Pd-110	Pd-111m	5.5 h	0.43	0.32	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1167.9	Cs-133	Cs-134	2.06 yr	2650	1.80	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1365.1 (3.04)
1169	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1173.2	Co-59	Co-60	5.272 yr	1680	99.9	346.9 (0.0076), 1332.5 (100)
1175.4	Kr-86	Kr-87	76.3 m	0.0519	1.11	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1177.9	Tb-159	Tb-160	72.3 d	2550	15.5	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
1180.6	Nd-150	Nd-151	12.4 m	6.768	15.3	116.8 (45.9), 138.9 (8.72), 175.1 (7.19), 255.7 (15.3), 423.5 (6.43), 736.4 (7.19), 797.5 (5.51), 1016.4 (2.92), 1122.1 (4.59), 1180.6 (15.3)
1183.5	Ir-193	Ir-194	19.15 h	6897	0.30	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
1184.2	Mo-100	Mo-101	14.6 m	1.916	0.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1186.6	Mo-100	Mo-101	14.6 m	1.916	1.03	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1189	Ta-181	Ta-182	115.0 d	2099.74	16.4	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1191.7	Se-82	Se-83g	22.5 m	0.04888	4.18	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1193.3	Ge-76	Ge-77	11.3 h	0.468	2.43	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1199.9	Tb-159	Tb-160	72.3 d	2550	2.36	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1200	Mo-100	Mo-101	14.6 m	1.916	1.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1200.1	Pd-110	Pd-111m	5.5 h	0.43	0.30	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1206.6	Te-130	Te-131m	30 h	0.676	9.73	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
1206.9	Se-82	Se-83g	22.5 m	0.04888	0.89	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1209.9	Mo-100	Mo-101	14.6 m	1.916	0.13	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1212.7	As-75	As-76	26.32 h	450	1.63	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1215.2	Ga-71	Ga-72	14.1 h	187.929	0.79	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1215.4	Ge-76	Ge-77	11.3 h	0.468	0.12	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1216	As-75	As-76	26.32 h	450	3.84	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1218.7	Hg-204	Hg-205	5.20 m	2.924	0.006	203.7 (2.2), 415.6 (0.13)
1221.4	Ta-181	Ta-182	115.0 d	2099.74	27.4	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1225.9	Se-82	Se-83g	22.5 m	0.04888	1.2	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1228.5	As-75	As-76	26.32 h	450	1.39	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1230.9	Ga-71	Ga-72	14.1 h	187.929	1.45	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1231	Ta-181	Ta-182	115.0 d	2099.74	11.6	67.7 (41.3), 100.1 (14.1), 152.4 (7.17), 156.4 (2.73), 179.4 (3.15), 222.1 (7.56), 229.3 (3.64), 264.1 (3.64), 1121.3 (35), 1189.0 (16.4), 1221.4 (27.4), 1231.0 (11.6)
1234.6	Cd-116	Cd-117m	3.36 h	0.1867	11.0	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1237	Rh-103	Rh-104	42.3 s	13 500	0.066	19.2 (0.24), 21.7 (0.05), 358.1 (0.016), 555.8 (1.99)
1241.4	Yb-176	Yb-177	1.9 h		3.4	53.0 (3.8), 54.1 (6.7), 61.2 (2.2), 121.6 (3.4), 150.4 (20), 1080.1 (5.5)
1242.2	Ge-76	Ge-77	11.3 h	0.468	0.38	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1245.2	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510.0 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1249.4	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1251.1	Mo-100	Mo-101	14.6 m	1.916	4.61	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1252.6	In-115	In-116	14.10 s	3828.8	0.031	463.1 (0.25), 1293.5 (1.30)
1254.7	Ba-138	Ba-139	1.38 h	25.812	0.033	33 (1.38), 33.4 (2.54), 37.8 (0.76), 38.7 (0.19), 165.8 (23.8), 1420.5 (0.261)
1256.8	Sb-121	Sb-122	2.70 d	335.2	0.77	25 (0.5), 25.3 (0.93), 27.2 (0.088), 27.5 (0.16), 28.5 (0.26), 29.1 (0.5), 564 (70.8), 692.6 (3.68), 1141.1 (0.57)
1259.3	Se-82	Se-83g	22.5 m	0.04888	1.85	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1260.1	Ga-71	Ga-72	14.1 h	187.929	1.13	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1260.2	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1263.9	Ge-76	Ge-77	11.3 h	0.468	0.80	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1266.1	Si-30	Si-31	2.62 h	0.3317	0.07	–
1268.6	Zr-96	Zr-97	17.0 h	0.0632	0.17	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1268.6	Zr-96	Nb-97	72.1 m		0.16	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
1271.9	Tb-159	Tb-160	72.3 d	2550	7.60	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
1273.2	Xe-136	Xe-137	3.83 m	2.314	0.23	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1274.5	Eu-153	Eu-154	8.8 yr		35.5	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
1276.1	Zr-96	Zr-97	17.0 h	0.0632	0.97	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1276.8	Ga-71	Ga-72	14.1 h	187.929	1.56	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1280	Ge-76	Ge-77	11.3 h	0.468	0.16	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1282.5	Pd-110	Pd-111m	5.5 h	0.43	1.0	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1282.7	Zn-70	Zn-71m	3.94 h	0.005	0.27	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1286.3	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1290.6	Cd-114	Cd-115m	44.6 d	1.03	0.890	484.5 (0.29), 933.8 (2)
1290.7	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1291.6	Fe-58	Fe-59	44.503 d	0.3968	43.2	142.5 (1.03), 192.2 (3.11), 334.8 (0.26), 1099.3 (56.5)
1293.3	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1293.5	In-115	In-116	14.10 s	3828.8	1.30	463.1 (0.25), 1252.6 (0.031)
1293.5	In-115	In-116m$_1$	54.15 m	7753.32	84.4	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1507.4 (9.96), 1753.8 (2.46)
1293.7	Ar-40	Ar-41	1.83 h	65.736	99.1	1677.2 (0.052)
1293.8	Se-82	Se-83g	22.5 m	0.04888	1.6	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1297.1	Ca-46	Ca-47	4.54 d	0.00296	74.9	489.2 (6.74), 530.4 (0.1), 767 (0.19), 807.9 (6.89)
1299.1	Se-82	Se-83g	22.5 m	0.04888	5.83	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1300	In-113	In-114	71.9 s	16.69	0.14	23.1 (2), 26.2 (0.42), 558 (0.07), 575.7 (0.004)
1303	Se-82	Se-83m	70.4 s	0.3666	0.93	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11)
1303.3	Cd-116	Cd-117	2.49 h	0.3735	18.4	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1576.6 (11.2), 1723.1 (2.01)
1304	Mo-100	Mo-101	14.6 m	1.916	2.78	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1305.9	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1306.7	Zn-70	Zn-71m	3.94 h	0.005	0.11	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1309.3	Ge-76	Ge-77	11.3 h	0.468	0.46	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1311.4	Zn-70	Zn-71m	3.94 h	0.005	0.10	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1312.1	Tb-159	Tb-160	72.3 d	2550	2.97	197.0 (5.24), 215.6 (4.02), 298.6 (27.4), 765.3 (2.04), 879.4 (30.0), 962.3 (10.0), 966.2 (25.5), 1177.9 (15.5), 1199.9 (2.36), 1271.9 (7.60), 1312.1 (2.97)
1312.8	Ge-76	Ge-77	11.3 h	0.468	0.34	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1314.3	Mo-100	Mo-101	14.6 m	1.916	0.23	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1316.9	Se-82	Se-83g	22.5 m	0.04888	4.12	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1317.5	Br-81	Br-82g	35.30 h	12.8206	26.9	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1319.7	Ge-76	Ge-77	11.3 h	0.468	0.28	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1321.3	Ru-104	Ru-105	4.44 h	5.984	0.20	262.8 (6.49), 316.4 (10.9), 393.4 (3.73), 413.5 (2.22), 469.4 (17.3), 676.4 (15.5), 724.2 (46.7), 822.0 (845.9), 875.9 (2.47), 969.4 (2.08)
1322.2	Zn-70	Zn-71m	3.94 h	0.005	0.23	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1325.7	Mo-100	Mo-101	14.6 m	1.916	0.26	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1332.1	Kr-78	Kr-79g	35.04 h	2.1	0.44	217.0 (2.4), 261.3 (12.7), 299.5 (1.57), 306.3 (2.6), 389.0 (1.52), 397.6 (9.5),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						606.1 (8.10), 832.0 (1.26), 1115.1 (0.37), 1332.1 (0.44)
1332.5	Co-59	Co-60m	10.47 m	2040	0.25	6.9 (27.2), 7.7 (3.67), 58.6 (2.02), 826.2 (0.0083)
1332.5	Co-59	Co-60	5.272 yr	1680	100	346.9 (0.0076), 1173.2 (99.9)
1333.2	Cu-65	Cu-66	5.1 m	66.9011	0.003	833.4 (0.18), 1039.2 (8)
1333.6	V-51	V-52	3.75 m	488.775	0.59	1434.1 (100), 1530.7 (0.12)
1334.3	Ag-109	Ag-110m	249.9 d	226.4	0.13	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1336.3	Mo-100	Mo-101	14.6 m	1.916	0.18	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1338	Kr-86	Kr-87	76.3 m	0.0519	0.63	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1339.3	Cd-116	Cd-117m	3.36 h	0.1867	2.1	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1339.4	Mo-100	Mo-101	14.6 m	1.916	0.12	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1341.2	Se-82	Se-83g	22.5 m	0.04888	5.70	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1345.8	Cu-63	Cu-64	12.7 h	311.265	0.48	511 (35.8)
1346.1	Mo-100	Mo-101	14.6 m	1.916	1.03	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1352.5	Se-82	Se-83g	22.5 m	0.04888	4.80	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1355.9	Mo-100	Mo-101	14.6 m	1.916	1.67	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1356.8	O-18	O-19	27.1 s	0.000032	50.4	109.9 (2.54), 197.1 (95.9), 1444.1 (2.64), 1554 (1.39)
1362.7	Zr-96	Zr-97	17.0 h	0.0632	1.35	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1363	Mo-92	Mo-93m	6.85 h	0.089	0.79	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 684.7 (99.7), 949.8 (0.12), 1477.1 (99.1)
1365.1	Cs-133	Cs-134	2.06 yr	2650	3.04	31.8 (0.24), 32.2 (0.44), 36.4 (0.13), 475.4 (1.46), 563.2 (8.36), 569.3 (15.4), 604.7 (97.6), 795.8 (85.4), 801.9 (8.73), 1038.6 (1), 1167.9 (1.8)
1366.3	Rb-87	Rb-88	17.8 m	3.339	0.11	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1368.2	Sb-123	Sb-124	60.2 d	175.07	2.36	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1691.0 (49.0), 2091.0 (5.61)
1368.3	Ge-76	Ge-77	11.3 h	0.468	3.18	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1368.5	Na-23	Na-24	14.96 h	53	100	2753.9 (99.9), 3867.3 (0.052)
1377.7	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1379.4	Ho-165	Ho-166	26.80 h	6300	0.93	48.2 (2.86), 49.1 (5.07), 55.6 (1.62), 57.2 (0.45), 80.6 (6.20), 1581.9 (0.18), 1662.4 (0.12)
1380.4	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1380.8	Zn-70	Zn-71m	3.94 h	0.005	0.36	121.5 (2.9), 142.6 (5.6), 386.3 (93), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1382.4	Rb-87	Rb-88	17.8 m	3.339	0.77	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1382.6	Kr-86	Kr-87	76.3 m	0.0519	0.29	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1382.7	Mo-100	Mo-101	14.6 m	1.916	1.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1384.3	Ag-109	Ag-110m	249.9 d	226.4	24.3	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1388.5	Pd-110	Pd-111m	5.5 h	0.43	0.41	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1388.5	Pd-110	Pd-111	23.4 m	2.22	0.54	22.1 (17.0), 25 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
1389.9	Kr-86	Kr-87	76.3 m	0.0519	0.19	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1394.8	Mo-100	Mo-101	14.6 m	1.916	0.61	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1404	Sn-124	Sn-125m	9.52 m	0.728	0.7	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 643 (0.16), 1483.9 (0.18), 1615.3 (0.12)
1408	Eu-151	Eu-152	13.33 yr	2 82 020	20.8	121.8 (28.4), 244.7 (7.51), 344.3 (26.6), 444.0 (3.12), 778.9 (13), 867.4 (4.21), 964.1 (14.6), 1085.9 (9.92), 1112.1 (13.6), 1408.0 (20.8)
1408.9	Ca-48	Ca-49	8.72 m	0.2057	0.63	856.1 (0.13), 1144.5 (0.11), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
1414.2	Mo-100	Mo-101	14.6 m	1.916	0.50	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1418.6	Mo-100	Mo-101	14.6 m	1.916	0.88	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1420.5	Ba-138	Ba-139	1.38 h	25.812	0.261	33 (1.38), 33.4 (2.54), 37.8 (0.76), 38.7 (0.19), 165.8 (23.8), 1254.7 (0.033)
1420.6	Se-82	Se-83g	22.5 m	0.04888	4.10	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1430	Mo-100	Mo-101	14.6 m	1.916	0.14	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1431.6	Mo-100	Mo-101	14.6 m	1.916	0.36	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1432.9	Cd-116	Cd-117m	3.36 h	0.1867	13.4	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
1434.1	V-51	V-52	3.75 m	488.775	100	1333.6 (0.59), 1530.7 (0.12)
1435.8	Se-82	Se-83g	22.5 m	0.04888	0.82	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1439.1	As-75	As-76	26.32 h	450	0.33	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1440.8	Mo-100	Mo-101	14.6 m	1.916	0.16	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1444.1	O-18	O-19	27.1 s	0.000032	2.64	109.9 (2.54), 197.1 (95.9), 1356.8 (50.4), 1554 (1.39)
1447.4	Se-82	Se-83g	22.5 m	0.04888	0.48	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1452.6	Ge-76	Ge-77	11.3 h	0.468	0.11	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1453.6	As-75	As-76	26.32 h	450	0.13	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1459	Pd-110	Pd-111	23.4 m	2.22	0.56	22.1 (17.0), 25 (3.58), 59.8 (0.553), 70.4 (0.78), 245.4 (0.48), 376.7 (0.44), 580.0 (0.837), 650.4 (0.55), 1388.5 (0.54), 1459.0 (0.56)
1459	Pd-110	Pd-111m	5.5 h	0.43	0.42	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1460.8	K-39	K-40	1.28×10^9 yr	195.846	10.7	3 (0.95)
1464	Ga-71	Ga-72	14.1 h	187.929	3.55	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1468.9	Ir-193	Ir-194	19.15 h	6897	0.19	66.8 (0.41), 293.5 (2.54), 300.7 (0.35), 328.4 (13.0), 589.2 (0.14), 622.0 (0.33), 645.1 (1.16), 938.7 (0.59), 1150.8 (0.59), 1183.5 (0.30), 1468.9 (0.19)
1474.8	Br-81	Br-82m	6.13 m	119.8233	0.016	46.0 (0.24), 698.3 (0.024), 776.5 (0.2)
1474.8	Br-81	Br-82g	35.30 h	12.8206	16.6	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1475.2	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1475.8	Ag-109	Ag-110m	249.9 d	226.4	3.99	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1476	Zn-70	Zn-71m	3.94 h	0.005	0.6	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1476.5	Ge-76	Ge-77	11.3 h	0.468	0.23	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1477.1	Mo-92	Mo-93m	6.85 h	0.089	99.1	17.4 (18.8), 19.7 (3.7), 114.0 (0.68), 263.1 (56.7), 684.7 (99.7), 949.8 (0.12), 1363.0 (0.79)
1479	Ge-76	Ge-77	11.3 h	0.468	0.20	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1481.8	Ni-64	Ni-65	2.52 h	1.3832	23.5	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1623.4 (0.47), 1724.9 (0.39)
1483.9	Sn-124	Sn-125m	9.52 m	0.728	0.18	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 643 (0.16), 1404 (0.7), 1615.3 (0.12)
1485.9	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1495.6	Ge-76	Ge-77	11.3 h	0.468	0.47	211 (29.2), 215.5 (27.1), 264.4 (51.0), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1505	Ag-109	Ag-110m	249.9 d	226.4	13.1	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1507.4	In-115	In-116m$_1$	54.15 m	7753.32	9.96	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1753.8 (2.46)
1514.1	Mo-100	Mo-101	14.6 m	1.916	0.19	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1515.6	Zr-96	Zr-97	17.0 h	0.0632	0.13	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1515.6	Zr-96	Nb-97	72.1 m		0.12	657.9 (98.5), 1024.5 (1.08), 1268.6 (0.16), 1515.6 (0.12)
1517.8	Mo-100	Mo-101	14.6 m	1.916	0.22	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1520.4	Mo-100	Mo-101	14.6 m	1.916	0.24	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1523	Mo-100	Mo-101	14.6 m	1.916	0.29	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1524.6	K-41	K-42	12.36 h	9.8258	18.3	312.7 (0.35), 899 (0.053)
1526.6	Mo-100	Mo-101	14.6 m	1.916	0.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1528	Cr-54	Cr-55	3.5 m	0.8496	0.037	126.1 (0.0017), 2252.5 (0.0031)
1530.3	Mo-100	Mo-101	14.6 m	1.916	0.27	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1530.7	V-51	V-52	3.75 m	488.775	0.12	1333.6 (0.59), 1434.1 (100)
1531.2	Kr-86	Kr-87	76.3 m	0.0519	0.36	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1532.5	Mo-100	Mo-101	14.6 m	1.916	5.96	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1538.8	Ge-76	Ge-77	11.3 h	0.468	0.13	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1548.7	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1554	O-18	O-19	27.1 s	0.000032	1.39	109.9 (2.54), 197.1 (95.9), 1356.8 (50.4), 1444.1 (2.64)
1554.8	Se-82	Se-83g	22.5 m	0.04888	2.54	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1558.5	Se-82	Se-83m	70.4 s	0.3666	1.20	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1562.3	Ag-109	Ag-110m	249.9 d	226.4	1.18	657.7 (94.7), 677.6 (10.7), 687.0 (6.49), 706.7 (16.7), 763.9 (22.4), 818 (7.32), 884.7 (72.9), 937.5 (34.3), 1384.3 (24.3), 1505.0 (13.1)
1568.1	Ga-71	Ga-72	14.1 h	187.929	0.20	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1571.6	Ga-71	Ga-72	14.1 h	187.929	0.82	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1573.7	Ge-76	Ge-77	11.3 h	0.468	0.62	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1575.7	Pr-141	Pr-142	19.13 h	760	3.7	508.8 (0.023), 641.2 (0.002)
1576.6	Cd-116	Cd-117	2.49 h	0.3735	11.2	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1723.1 (2.01)
1576.8	Xe-136	Xe-137	3.83 m	2.314	0.10	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1578	Kr-86	Kr-87	76.3 m	0.0519	0.13	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1581.9	Ho-165	Ho-166	26.80 h	6300	0.18	48.2 (2.86), 49.1 (5.07), 55.6 (1.62), 57.2 (0.45), 80.6 (6.20), 1379.4 (0.93), 1662.4 (0.12)
1589.7	Mo-100	Mo-101	14.6 m	1.916	0.28	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1596.5	La-139	La-140	40.28 h	892.12	95.4	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 2521.7 (3.43)
1596.5	Eu-153	Eu-154	8.8 yr		1.83	123.1 (40.5), 247.9 (6.60), 591.8 (4.83), 723.3 (19.7), 756.9 (40.5), 873.2 (11.5), 996.3 (10.3), 1004.8 (17.9), 1274.5 (35.5), 1596.5 (1.83)
1596.7	Ga-71	Ga-72	14.1 h	187.929	4.24	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1599.3	Mo-100	Mo-101	14.6 m	1.916	1.75	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1609.2	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1610.4	Re-187	Re-188	16.98 h	4569.8	0.10	61.5 (1.36), 63.0 (2.35), 71.3 (0.81), 73.4 (0.20), 155.0 (14.9), 478.0 (1.04), 633.0 (1.25), 635.1 (0.15), 672.5 (0.11), 829.5 (0.41), 931.3 (0.56)
1611.2	Kr-86	Kr-87	76.3 m	0.0519	0.10	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1612.5	Xe-136	Xe-137	3.83 m	2.314	0.12	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1615.3	Sn-124	Sn-125m	9.52 m	0.728	0.12	26.1 (0.53), 26.4 (1), 29.7 (0.28), 332.1 (97.5), 589.6 (0.2), 643 (0.16), 1404 (0.7), 1483.9 (0.18)
1622.6	Ca-48	Sc-49	57.4 m		0.01	1761.9 (0.05)
1623.4	Ni-64	Ni-65	2.52 h	1.3832	0.47	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1724.9 (0.39)
1631.6	Zn-70	Zn-71g	2.4 m	0.0514	0.38	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54), 910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1633.7	F-19	F-20	11.0 s	0.96	100.0	3332.5 (0.009)
1636.5	Ne-22	Ne-23	37.2 s	0.42412	1.0	439.9 (33), 2076.4 (0.1)
1642.4	Cl-37	Cl-38	37.18 m	10.49159	31.6	2167.6 (42.4)
1650.3	Br-81	Br-82g	35.30 h	12.8206	0.79	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1651.3	Pd-110	Pd-111m	5.5 h	0.43	0.68	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1660	Se-82	Se-83m	70.4 s	0.3666	1.78	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660 (1.78), 2051.4 (11.0)
1662.4	Ho-165	Ho-166	26.80 h	6300	0.12	48.2 (2.86), 49.1 (5.07), 55.6 (1.62), 57.2 (0.45), 80.6 (6.20), 1379.4 (0.93), 1581.9 (0.18)
1662.4	Mo-100	Mo-101	14.6 m	1.916	0.68	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1664.6	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1673.9	Mo-100	Mo-101	14.6 m	1.916	1.69	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1677.2	Ar-40	Ar-41	1.83 h	65.736	0.052	1293.7 (99.1)
1680.8	Ga-71	Ga-72	14.1 h	187.929	0.90	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1684.2	Se-82	Se-83g	22.5 m	0.04888	0.2	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1691	Sb-123	Sb-124	60.2 d	175.07	49.0	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 2091.0 (5.61)
1691.1	Pd-110	Pd-111m	5.5 h	0.43	1.2	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1694.5	Se-82	Se-83m	70.4 s	0.3666	0.74	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1709.8	Ge-76	Ge-77	11.3 h	0.468	0.29	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
1710.9	Ga-71	Ga-72	14.1 h	187.929	0.39	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1712.8	Mo-100	Mo-101	14.6 m	1.916	0.20	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1715.9	Se-82	Se-83g	22.5 m	0.04888	0.6	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1719.7	Ge-76	Ge-77	11.3 h	0.468	0.38	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1721.9	Pd-110	Pd-111m	5.5 h	0.43	0.32	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1723.1	Cd-116	Cd-117	2.49 h	0.3735	2.01	24.1 (4.58), 89.7 (3.26), 273.3 (27.9), 344.5 (17.9), 434.2 (9.79), 831.8 (2.26), 880.7 (3.96), 1051.7 (3.79), 1303.3 (18.4), 1576.6 (11.2)
1724.9	Ni-64	Ni-65	2.52 h	1.3832	0.39	366.6 (4.61), 507.8 (0.29), 609.3 (0.14), 1115.5 (14.8), 1481.8 (23.5), 1623.4 (0.47)
1727.2	Ge-76	Ge-77	11.3 h	0.468	0.14	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1740.5	Kr-86	Kr-87	76.3 m	0.0519	2.04	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1750.5	Zr-96	Zr-97	17.0 h	0.0632	1.34	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1753.8	In-115	In-116m$_1$	54.15 m	7753.32	2.46	138.3 (3.29), 416.9 (29.2), 818.7 (11.5), 1097.3 (56.2), 1293.5 (84.4), 1507.4 (9.96)
1754.9	Mo-100	Mo-101	14.6 m	1.916	0.46	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1755	N-15	N-16	7.13 s	0.00000888	0.13	2741 (0.76), 6129 (68.8), 7115.2 (5)
1759.6	Zn-70	Zn-71m	3.94 h	0.005	0.93	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
1759.8	Mo-100	Mo-101	14.6 m	1.916	0.98	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1761.9	Ca-48	Sc-49	57.4 m		0.05	1622.6 (0.01)
1768.2	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1775.2	Pd-110	Pd-111m	5.5 h	0.43	0.44	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1779	Se-82	Se-83m	70.4 s	0.3666	0.71	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
1779	Al-27	Al-28	2.246 m	23.1	100.0	–
1779.6	Br-81	Br-82g	35.30 h	12.8206	0.12	221.5 (2.25), 554.3 (70.6), 619.1 (43.1), 698.3 (27.9), 776.5 (83.4), 827.8 (24.2), 1007.6 (1.31), 1044.0 (27.4), 1317.5 (26.9), 1474.8 (16.6)
1779.8	Rb-87	Rb-88	17.8 m	3.339	0.22	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1779.9	Se-82	Se-83g	22.5 m	0.04888	1.9	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1783.4	Xe-136	Xe-137	3.83 m	2.314	0.41	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1787.7	As-75	As-76	26.32 h	450	0.33	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
1810.7	Mn-55	Mn-56	2.58 h	1330	27.2	846.8 (98.9), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
1827.1	Se-82	Se-83g	22.5 m	0.04888	1.4	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1836	Rb-87	Rb-88	17.8 m	3.339	22.1	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
1837.6	Ga-71	Ga-72	14.1 h	187.929	0.21	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1840.3	Mo-100	Mo-101	14.6 m	1.916	1.37	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
1842.6	Kr-86	Kr-87	76.3 m	0.0519	0.14	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
1846.4	Ge-76	Ge-77	11.3 h	0.468	0.16	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
1847.6	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1851.6	Zr-96	Zr-97	17.0 h	0.0632	0.35	355.4 (2.27), 507.6 (5.05), 602.4 (1.4), 657.9 (106), 743.4 (92.6), 1021.3 (1.34), 1024.5 (1.16), 1147.9 (2.64), 1362.7 (1.35), 1750.5 (134)
1854.4	Se-82	Se-83g	22.5 m	0.04888	1.5	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1861.1	Ga-71	Ga-72	14.1 h	187.929	5.25	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1871.2	Se-82	Se-83g	22.5 m	0.04888	1.4	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1877.9	Ga-71	Ga-72	14.1 h	187.929	0.23	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1894.8	Se-82	Se-83g	22.5 m	0.04888	7.75	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1904.4	Zn-70	Zn-71g	2.4 m	0.0514	0.17	121.5 (3.0), 390.0 (3.8), 398.6 (0.61), 511.6 (32), 666.8 (0.90), 721.4 (0.54),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						910.3 (7.8), 964.8 (0.77), 1120.0 (2.2), 1631.6 (0.38)
1916.3	Xe-136	Xe-137	3.83 m	2.314	0.10	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)
1920.2	Ga-71	Ga-72	14.1 h	187.929	0.16	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1970.8	Pd-110	Pd-111m	5.5 h	0.43	0.59	70.4 (8.5), 172.2 (33.5), 391.3 (5.1), 575.0 (3.0), 632.8 (3.4), 694.2 (1.9), 762.2 (1.2), 1115.9 (1.0), 1282.5 (1.0), 1691.1 (1.2)
1973.2	Se-82	Se-83g	22.5 m	0.04888	0.62	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
1991.3	Ga-71	Ga-72	14.1 h	187.929	0.11	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
1997.3	Cd-116	Cd-117m	3.36 h	0.1867	26.2	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
2000.1	Ge-76	Ge-77	11.3 h	0.468	0.53	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2000.9	Te-130	Te-131m	30 h	0.676	2.01	102.1 (7.91), 149.7 (20.5), 200.6 (7.53), 240.9 (7.57), 334.3 (9.53), 773.7 (38.1), 782.5 (7.76), 793.8 (13.8), 822.8 (6.1), 852.2 (21), 1125.5 (11.4), 1206.6 (9.73)
2001.8	Sn-124	Sn-125	9.64 d	0.0224	1.79	332.1 (1.31), 469.8 (1.38), 800.3 (1.0), 822.5 (3.99), 915.5 (3.85), 1067.1 (9.05), 1087.7 (1.11), 1089.2 (4.28)
2011.9	Kr-86	Kr-87	76.3 m	0.0519	2.88	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2028.1	Mo-100	Mo-101	14.6 m	1.916	0.10	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2029.4	Ga-71	Ga-72	14.1 h	187.929	0.12	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24),

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2032.1	Mo-100	Mo-101	14.6 m	1.916	6.94	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2038.4	Mo-100	Mo-101	14.6 m	1.916	0.21	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2041.2	Mo-100	Mo-101	14.6 m	1.916	2.11	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2045.2	Se-82	Se-83g	22.5 m	0.04888	0.69	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2051.4	Se-82	Se-83m	70.4 s	0.3666	11.0	356.7 (17.3), 673.9 (15.1), 987.9 (15.3), 997.6 (1.28), 1020.6 (1.97), 1030.5 (20.9), 1053.6 (1.49), 1063.4 (3.39), 1660.0 (1.78), 2051.4 (11.0)
2072.4	Se-82	Se-83g	22.5 m	0.04888	0.27	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2076.4	Ne-22	Ne-23	37.2 s	0.42412	0.1	439.9 (33), 1636.5 (1)
2077.2	Ge-76	Ge-77	11.3 h	0.468	0.22	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2085.4	Se-82	Se-83g	22.5 m	0.04888	0.55	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2088.8	Mo-100	Mo-101	14.6 m	1.916	0.79	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2089.6	Ge-76	Ge-77	11.3 h	0.468	0.33	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2091	Sb-123	Sb-124	60.2 d	175.07	5.61	602.7 (98.4), 645.8 (7.24), 713.8 (2.40), 722.8 (11.3), 968.3 (1.84), 1045.2 (1.85), 1368.2 (2.36), 1691.0 (49.0)
2096.3	As-75	As-76	26.32 h	450	0.66	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39),

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
						1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
2096.4	Cd-116	Cd-117m	3.36 h	0.1867	7.44	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
2109.5	Ga-71	Ga-72	14.1 h	187.929	1.04	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2110.8	As-75	As-76	26.32 h	450	0.39	559.1 (44.7), 563.2 (1.17), 665.3 (0.39), 1212.7 (1.63), 1216 (3.84), 1228.5 (1.39), 1439.1 (0.33), 1787.7 (0.33), 2096.3 (0.66), 2110.8 (0.39)
2111.2	Rb-87	Rb-88	17.8 m	3.339	0.12	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2112.8	Mo-100	Mo-101	14.6 m	1.916	0.15	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2113	Mn-55	Mn-56	2.58 h	1330	14.3	846.8 (98.9), 1810.7 (27.2), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
2114.3	Mo-100	Mo-101	14.6 m	1.916	0.45	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2118.9	Rb-87	Rb-88	17.8 m	3.339	0.44	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2126.1	Ge-76	Ge-77	11.3 h	0.468	0.19	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2167.3	Se-82	Se-83g	22.5 m	0.04888	0.34	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2167.6	Cl-37	Cl-38	37.18 m	10.49159	42.4	1642.4 (31.6)
2201.7	Ga-71	Ga-72	14.1 h	187.929	25.9	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2214.3	Ga-71	Ga-72	14.1 h	187.929	0.18	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
2223.2	Mo-100	Mo-101	14.6 m	1.916	0.17	191.9 (18.8), 505.9 (11.8), 590.1 (5.6), 590.9 (16.4), 695.6 (7.20), 1012.5 (12.8), 1161 (3.97), 1251.1 (4.61), 1532.5 (5.96), 2032.1 (6.94)
2228.9	Ca-48	Ca-49	8.72 m	0.2057	0.19	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
2252.5	Cr-54	Cr-55	3.5 m	0.8496	0.0031	126.1 (0.0017), 1528 (0.037)
2290.2	Se-82	Se-83g	22.5 m	0.04888	9.26	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2317.7	Zn-70	Zn-71m	3.94 h	0.005	0.65	121.5 (2.9), 142.6 (5.6), 386.3 (93.0), 487.3 (62.3), 511.6 (28.4), 596.1 (27.9), 620.2 (56.7), 753.4 (3.3), 964.7 (4.7), 1107.4 (2.8)
2322.8	Cd-116	Cd-117m	3.36 h	0.1867	7.86	158.6 (109), 552.9 (125), 564.4 (14.7), 860.4 (7.89), 1029.1 (11.7), 1066 (23.1), 1234.6 (11), 1432.9 (13.4), 1997.3 (26.2), 2322.8 (7.86)
2337.4	Se-82	Se-83g	22.5 m	0.04888	3.43	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2341.6	Ge-76	Ge-77	11.3 h	0.468	0.45	211 (29.2), 215.5 (27.1), 264.4 (51), 367.4 (13.3), 416.3 (20.6), 558.0 (15.2), 631.8 (6.59), 714.3 (6.77), 1085.2 (5.72), 1368.3 (3.18)
2371.7	Ca-48	Ca-49	8.72 m	0.2057	0.49	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 3084.4 (92.1), 4071.9 (7), 4738.2 (0.21)
2408.5	Kr-86	Kr-87	76.3 m	0.0519	0.23	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2419.9	Se-82	Se-83g	22.5 m	0.04888	0.41	225.2 (31.9), 356.7 (68.6), 510 (44.3), 718 (16.3), 799 (16), 836.5 (15.9), 866.6 (8.85), 883.6 (7.75), 1894.8 (7.75), 2290.2 (9.26)
2491	Ga-71	Ga-72	14.1 h	187.929	7.68	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2507.8	Ga-71	Ga-72	14.1 h	187.929	12.8	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)

(*continued overleaf*)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
2515	Ga-71	Ga-72	14.1 h	187.929	0.25	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2521.7	La-139	La-140	40.28 h	892.12	3.43	328.8 (20.7), 432.5 (2.99), 487.0 (45.9), 751.8 (4.41), 815.8 (23.6), 867.8 (5.59), 919.6 (2.68), 925.2 (2.68), 1596.5 (95.4)
2522.9	Mn-55	Mn-56	2.58 h	1330	0.99	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2657.5 (0.65), 2959.8 (0.31), 3369.6 (0.17)
2554.8	Kr-86	Kr-87	76.3 m	0.0519	9.23	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2558.1	Kr-86	Kr-87	76.3 m	0.0519	3.92	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2577.7	Rb-87	Rb-88	17.8 m	3.339	0.19	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2621.1	Ga-71	Ga-72	14.1 h	187.929	0.13	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2657.5	Mn-55	Mn-56	2.58 h	1330	0.65	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2959.8 (0.31), 3369.6 (0.17)
2677.9	Rb-87	Rb-88	17.8 m	3.339	2.02	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2734	Rb-87	Rb-88	17.8 m	3.339	0.11	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
2741	N-15	N-16	7.13 s	0.00000888	0.76	1755 (0.13), 6129 (68.8), 7115.2 (5)
2753.9	Na-23	Na-24	14.96 h	53	99.9	1368.5 (100), 3867.3 (0.052)
2811.4	Kr-86	Kr-87	76.3 m	0.0519	0.32	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
2844	Ga-71	Ga-72	14.1 h	187.929	0.43	600.9 (5.54), 630 (24.8), 834 (95.6), 894.2 (9.88), 1050.7 (6.91), 1596.7 (4.24), 1861.1 (5.25), 2201.7 (25.9), 2491.0 (7.68), 2507.8 (12.8)
2849.8	Xe-136	Xe-137	3.83 m	2.314	0.18	298 (0.12), 393.4 (0.14), 455.5 (31.2), 848.9 (0.62), 982.2 (0.21), 1273.2 (0.23), 1576.8 (0.1), 1612.5 (0.12), 1783.4 (0.41), 2849.8 (0.18)

Energy (keV)	Irradiated target	Radio-nuclide	Half-life	Probability of formation	Gamma line intensity	Other lines (keV) (intensities, %)
2959.8	Mn-55	Mn-56	2.58 h	1330	0.31	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 3369.6 (0.17)
3009.4	Rb-87	Rb-88	17.8 m	3.339	0.25	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
3084.4	Ca-48	Ca-49	8.72 m	0.2057	92.1	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 4071.9 (7), 4738.2 (0.21)
3103.3	S-36	S-37	5.0 m	0.003	94.1	907.3 (0.058), 3740.4 (0.23)
3214	B-11	B-12	2.02×10^{-2} s	0.44	0.0006	4437.1 (1.28)
3218.5	Rb-87	Rb-88	17.8 m	3.339	0.22	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
3308.5	Kr-86	Kr-87	76.3 m	0.0519	0.45	402.6 (49.6), 673.8 (1.89), 836.4 (0.77), 845.4 (7.34), 1175.4 (1.11), 1338.0 (0.63), 1740.5 (2.04), 2011.9 (2.88), 2554.8 (9.23), 2558.1 (3.92)
3332.5	F-19	F-20	11.0 s	0.96	0.009	1633.7 (100)
3369.6	Mn-55	Mn-56	2.58 h	1330	0.17	846.8 (98.9), 1810.7 (27.2), 2113 (14.3), 2522.9 (0.99), 2657.5 (0.65), 2959.8 (0.31)
3486.5	Rb-87	Rb-88	17.8 m	3.339	0.13	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
3740.4	S-36	S-37	5.0 m	0.003	0.23	907.3 (0.058), 3103.3 (94.1)
3867.3	Na-23	Na-24	14.96 h	53	0.052	1368.5 (100), 2753.9 (99.9)
4071.9	Ca-48	Ca-49	8.72 m	0.2057	7	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4738.2 (0.21)
4437.1	B-11	B-12	2.02×10^{-2} s	0.44	1.28	3214 (0.0006)
4738.2	Ca-48	Ca-49	8.72 m	0.2057	0.21	856.1 (0.13), 1144.5 (0.11), 1408.9 (0.63), 2228.9 (0.19), 2371.7 (0.49), 3084.4 (92.1), 4071.9 (7)
4742.7	Rb-87	Rb-88	17.8 m	3.339	0.15	898.0 (14.5), 1382.4 (0.77), 1779.8 (0.22), 1836.0 (22.1), 2118.9 (0.44), 2577.7 (0.19), 2677.9 (2.02), 3009.4 (0.25), 3218.5 (0.22), 4742.7 (0.15)
6129	N-15	N-16	7.13 s	0.00000888	68.8	1755 (0.13), 2741 (0.76), 7115.2 (5)
7115.2	N-15	N-16	7.13 s	0.00000888	5.0	1755 (0.13), 2741 (0.76), 6129 (68.8)

Instrumental Neutron Activation Analysis

Zeev B. Alfassi
Ben Gurion University, Beer Sheva, Israel

1 Introduction 1

2 Basic Nuclear Physics 1
2.1 Radioactive Decay 1
2.2 Kinetics of Decay of Radioactive Nuclides 3
2.3 Kinetics of Chain Decays 4
2.4 Kinetics of Formation of Radioactive Nuclides by Irradiation 4

3 The Shape of a γ Spectrum 7

4 Neutron Sources 9
4.1 Sample Introduction 10

5 Instrumental Neutron Activation Analysis 10
5.1 Techniques 10
5.2 Calculation 11
5.3 Nuclear Interferences 13
5.4 A Test Case: Instrumental Neutron Activation Analysis of Trace Elements in Silicon 14
5.5 An Example of Interference Removal in Instrumental Neutron Activation Analysis: Determination of Platinum in Aerosols 15
5.6 Epithermal Instrumental Neutron Activation Analysis 16
5.7 Fast Neutron Instrumental Neutron Activation Analysis 20
5.8 Instrumental Neutron Activation Analysis with Emissions Other than γ Emission 22
5.9 Depth Profiling by Instrumental Neutron Activation Analysis 23
5.10 Practical Cases of Instrumental Neutron Activation Analysis 24

Abbreviations and Acronyms 27

Related Articles 28

References 28

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Analysis of the elemental content of a sample by irradiation with neutrons and measurement of the induced radioactivity are described.

1 INTRODUCTION

Chemical analysis by nuclear activation is an elemental analysis, i.e. it determines the contents of the various elements in the analyte sample, but cannot tell in what chemical form (compounds, valence states, etc.) they are present. The analysis is based on a reaction of the analyte element with nuclear projectiles (neutrons, accelerated small charged particles, e.g. protons, or γ-photons) [Equation 1]:

$$\text{target} + \text{projectile} \longrightarrow \text{light product} + \text{heavy product} \quad (1)$$

In instrumental neutron activation analysis (INAA) the projectile is a thermal neutron and the measurement of the concentration of the element is done via the heavy product, if it is a radionuclide which emits γ-rays. The amount produced of the radionuclide is proportional to the number of the target atoms.

2 BASIC NUCLEAR PHYSICS

This section gives only a brief description of basic knowledge. Readers who are interested in further detail should consult common textbooks.[1–6]

2.1 Radioactive Decay

Most radioactive decays occur via the transformations neutron → proton and proton → neutron and are called β-decays. They are characterized by the mass number, ΔA, which remains unchanged, whereas the atomic number is changed by ±1. There are three transformations of this type. A neutron is transformed into a proton by a β^- process [Equation 2]:

$$n \longrightarrow p^+ + \beta^- + \bar{\nu} \quad (2)$$

The proton remains in the nucleus, whereas the electron (β^-) and the antineutrino ($\bar{\nu}$) are emitted from the nucleus, with high kinetic energies. The electron is written as β^- and not e^-, to show that it is not an atomic electron. A proton can be transformed into a neutron by two processes. One process is β^+ decay, in which a proton is transformed into a neutron and a positron ($\beta^+ = e^+$), an entity similar in mass to an electron but which has an opposite charge [Equation 3]:

$$p^+ \longrightarrow n + \beta^+ + \nu \quad (3)$$

In this process there is also an emission of a neutrino (ν). Another process is the reaction of a proton with one of the surrounding electrons in the atom to yield a neutron. This process is called electron capture (EC). The neutrino and antineutrino are very small particles with no charge, and consequently they are very difficult to detect. They are not important in activation analysis, except for their effect on the β energies, which will be explained later.

Most radionuclides are not pure β emitters, but they emit simultaneously also a γ-photon (more accurately, the γ emission occurs in a very short time of $<10^{-10}$ s after the β^- emission). γ-Rays are electromagnetic waves, like light and radiowaves, but they have much higher energies (much shorter wavelengths). They have usually higher energies than X-rays, although the main difference between X-rays and γ-rays is their sources. X-rays are due to atomic transitions (transitions between different energy levels of electrons), whereas γ-rays are due to nuclear transitions (transitions between different energy levels of the nucleons). In most cases β^- decay does not yield the ground state of the product nuclide, but forms it in an excited state. The excited-state nuclide decays very rapidly (in most cases) to the ground state by the emission of either γ-rays or atomic electrons called conversion electrons. The emitted conversion electrons have no importance in activation analysis and we can neglect their emission and refer only to the emission of γ-rays.

A very important difference between β processes and γ decay (besides the difference in entities of the emitted particles, electrons vs photons, and the difference in ΔA, ± 1 in the β process and 0 in the γ decay) is the fact that in β processes there is emission of two particles, β^- and $\bar{\nu}$ or β^+ and ν, whereas in γ decay there is emission of only one photon. There are cases of the emission of a few photons from the same nuclide, but these are successive emissions and not a simultaneous emission as in the β processes. The importance of this difference lies in the fact that each nuclear decay is a transformation between two discrete energy states, resulting in a definite energy released in the process. If only one particle is emitted from the nucleus, this particle has a defined energy. When two particles are emitted, the released energy is distributed between them, and each particle can have a spectrum of energies ranging from zero up to a maximum energy, equal to the released energy. Since the emitted γ-photons have definite energies, they can be used in order to identify their emitters, by the measurement of the energy of the photons. As β-particles do not have definite energies, they usually cannot be used to identify their emitters.

In INAA, we need to identify the different radionuclides (in order to know the nuclide from which they were formed), in addition to the measurement of their activities. This is the reason why INAA employs almost exclusively the measurement of the spectrum of emitted γ-photons. In a few cases, where the indicator radionuclide (IRN) is a pure β^- emitter, it can be measured only in cases where there are only few (2–3) nonspecific γ-emitter IRNs in the sample with different half-lives. Their separate activities can be measured by measuring the activities at different times (measuring the decay curve) and extracting the various activities from the time dependence of the measured activity.

We said that the γ-photons emitted are due to the de-excitation of the nuclide produced in the β process. However, this de-excitation does not always have to be by only one photon. Moreover, not all β decays lead to the same level of excitation. These two facts together with the de-excitation by emission of conversion electrons (mainly from low-lying levels) lead to the possibilities that one radionuclide can have more than one kind (energy) of photon and that the number of photons should not be equal to the number of nuclides which have decayed (disintegrated). The number of photons of specific energy emitted per 100 disintegrated nuclides is called the intensity of the γ line, expressed as a percentage. Let us consider different cases of γ line intensity in four examples of radionuclides.

1. ^{28}Al decays completely to the 1.778 MeV excited state of ^{28}Si. This level decays to the ground state by only one photon. This means that ^{28}Al has only one γ line of 1.778 MeV with an intensity of 100%. This can be expressed by a diagram (called a decay scheme) of decreasing energies as shown in Scheme 1.

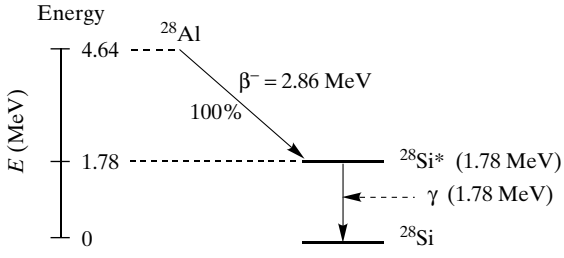

Scheme 1

2. ^{24}Na decays almost completely (>99.8%) to one level of ^{24}Mg (Scheme 2). However, this level decays to the ground state only by two successive photons. The first photon of 2.75 MeV is followed by a second photon of 1.39 MeV. This decay scheme explains why ^{24}Na has two γ lines of 2.75 and 1.39 MeV, each with almost 100% intensity.

3. ^{27}Mg decays to two different excited levels, resulting in two β^- with different energies, 1.59 MeV (31%) and 1.75 MeV (69%) (Scheme 3). The higher excited level (1.01 MeV) decays only 98% directly to the

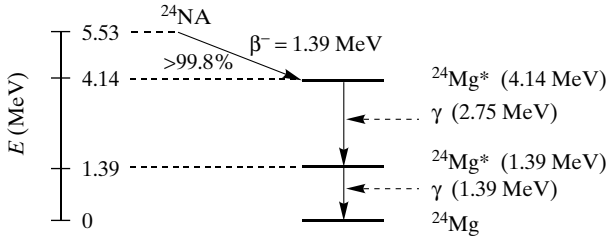

Scheme 2

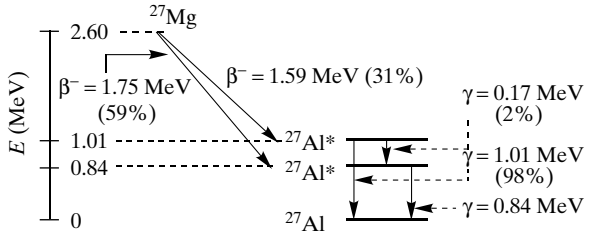

Scheme 3

ground state while 2% ($0.31 \times 0.02 = 0.006 = 0.6\%$ from the total nuclides disintegrated) decays to the lower excited state. The lower excited level decays completely to the ground state by one photon. This decay scheme explains why ^{27}Mg has three γ lines of 1.01 MeV (31%), 0.84 MeV (69%) and 0.18 MeV (0.6%).

4. ^{38}Cl decays directly to three different states of ^{38}Ar, both to two excited states of ^{38}Ar and to the ground state of ^{38}Ar (Scheme 4). This is the reason for ^{38}Cl having three different energy β^- particles (remember, it is only maximum energy, since β^- leads to a spectrum of β^- energies, up to E_{maximum}), 4.81 MeV (53%), 2.77 MeV (9%) and 1.77 MeV (38%). The higher excited state (3.77 MeV) decays almost completely to the lower excited state and only very little (0.06% of the 38%) decays directly to the ground state.

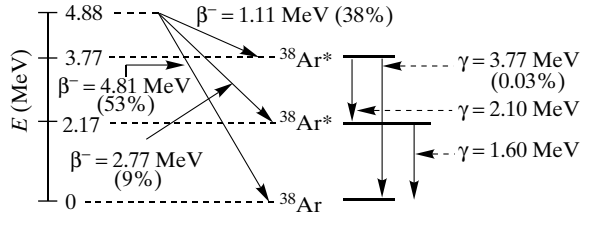

Scheme 4

Consequently ^{38}Cl has three γ lines, 3.77 MeV (0.03%), 2.10 MeV (38%) and 1.60 MeV (38% + 53% = 91%). These data together with more physical data on the nuclides can be found in books collecting all the decay schemes of the nuclides,[7] but for application of radionuclides it is sufficient to use the collection of all the γ lines of the radionuclides without the detailed decay scheme.[8–10] These collections are for all known radionuclides, but for most of cases in activation analysis it is better to use a smaller collection of radionuclides, only of those potentially formed by the specific kind of activation analysis.[11–13]

2.2 Kinetics of Decay of Radioactive Nuclides

The decay of radioactive nuclides is a statistical process. Thus, the number of atoms decaying per unit time (rate of decay) is proportional to the number of atoms of that specific radionuclide present in the sample. If the number of nuclei (atoms) of a specific radionuclide at time t is N^*, their rate of disappearance (decay or disintegration) is given by Equation (4):

$$-\frac{dN^*}{dt} = \lambda N^* \tag{4}$$

The constant λ, which is different for each radionuclide, is called the decay constant of this radionuclide. Integration of Equation (4) leads to the decay equation of radionuclides, i.e. the equation describing the time dependence of the number of atoms of the specific radionuclide [Equation 5]:

$$N^*(t) = N_0^* e^{-\lambda t} \tag{5}$$

where N_0^* is the number of atoms at time chosen as $t = 0$. Equation (5), which is the same as the integrated equation of any first-order chemical process, indicates the existence of a constant lifetime for the half of the atoms, which is called the half-life. It means that independently of the value of N_0^*, it takes the same time (for a specific radionuclide) for disintegration of half of the atoms, leaving only $N_0^*/2$ atoms of that radionuclide. Substituting $N^*(t)$ by $N_0^*/2$ in Equation (5) leads to the correlation between the half-life ($t_{1/2}$) and the decay constant (λ) [Equation 6]:

$$t_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda} \tag{6}$$

Equation (5), the decay equation, can be written also with $t_{1/2}$ instead of λ [Equation 7]:

$$N^*(t) = N_0^*(0.5)^{t/t_{1/2}} \tag{7}$$

The advantage of Equation (7) over Equation (5) is in that it gives a better 'feeling' for the extent of the decay. Thus, for example, if the half-life is $t_{1/2} = 2.5$ days, we know that after 2.5 days only half of the original atoms remain, and after 5 days the number of original atoms is reduced to one quarter. Equation (5) was easier to use when people were using tables of exponents. Nowadays

with computers there is no difference which equation is used. In tables of data, only $t_{1/2}$, the half-life, is given. In order to use Equation (5), λ should be calculated from Equation (6).

2.3 Kinetics of Chain Decays

In some cases a radionuclide decays not to a stable nuclide but to another radionuclide setting a chain of decays. Each chain is terminated in a stable nuclide. For a chain of decays as shown in Equation (8):

$$\mathrm{R}_1 \xrightarrow{\lambda_1} \mathrm{R}_2 \xrightarrow{\lambda_2} \mathrm{R}_3 \xrightarrow{\lambda_3} \cdots \longrightarrow \mathrm{R}_n \xrightarrow{\lambda_n} \mathrm{S} \qquad (8)$$

where R denotes a radionuclide and S stands for a stable isotope, the number of atoms R_k at time t assuming that at time zero only $R_1 \neq 0$ and is equal to R_1^0 is given by the Bateman equation [Equations 9 and 10]:

$$R_n = C_1 \mathrm{e}^{-\lambda_1 t} + C_2 \mathrm{e}^{-\lambda_2 t} + \cdots + C_n \mathrm{e}^{-\lambda_n t} \qquad (9)$$

$$R_n = \sum_{i=1}^{n} C_i \mathrm{e}^{-\lambda_i t} \qquad (10)$$

where [Equations 11 and 12]:

$$C_1 = \frac{\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_{n-1} N_1^0}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1) \cdots (\lambda_n - \lambda_1)} \qquad (11)$$

$$C_2 = \frac{\lambda_1 \lambda_2 \cdots \lambda_{n-1} N_1^0}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2) \cdots (\lambda_n - \lambda_2)} \qquad (12)$$

and generally [Equation 13]:

$$C_k = \frac{\prod_{i=1}^{n-1} \lambda_i}{\prod_{i=1,\neq \mathrm{k}}^{n} (\lambda_i - \lambda_k)} N_1^0 \qquad (13)$$

This is the solution of n simultaneous differential equations [Equations 14 and 15]:

$$\frac{\mathrm{d}R_1}{\mathrm{d}t} = -\lambda_1 R_1 \qquad (14)$$

$$\frac{\mathrm{d}R_k}{\mathrm{d}t} = \lambda_{k-1} R_{k-1} - \lambda_k R_k \quad (k = 2, 3, \ldots, n) \qquad (15)$$

2.4 Kinetics of Formation of Radioactive Nuclides by Irradiation

When a thin target is bombarded with a beam of projectiles, the rate of nuclear transformation (the number of nuclides produced per unit time) is proportional to the beam intensity (I = number of incident particles per unit time), to the target nucleus density (n = number of target nuclei per unit volume) and to the thickness of the target ($\mathrm{d}x$). The proportionality to the thickness of the target is limited to sufficiently small $\mathrm{d}x$, such that both the intensity of the beam and the energy of the projectiles remain practically unchanged. The proportionality constant is denoted σ and is called the reaction cross-section. This term comes from a simple model, assuming that each geometrical collision leads to a nuclear reaction. In this case, σ is the geometrical cross-section of the collision pair, $p(r_1 + r_2)^2$ [Equation 16]:

$$\frac{\mathrm{d}N^*}{\mathrm{d}t} = \sigma I n \, \mathrm{d}x \qquad (16)$$

Since the nucleus radius is a few femtometers (10^{-13} cm), the cross-sections are of the order of 10^{-24} cm^2. This is the reason why it has been customary to express cross-sections in units of barns, where 1 barn (b) = 10^{-24} cm^2.

Equation (16) assumes that the beam cross-section is smaller than the target size facing the beam, and consequently each projectile particle is transversing through the target. In the case that the target size is smaller than the beam cross-section, the beam intensity, I, should be replaced by the product ϕA, where ϕ is the beam flux (number of projectile particles per unit area and per unit time) and A is the area of the target facing the beam [Equation 17]:

$$\frac{\mathrm{d}N^*}{\mathrm{d}t} = \sigma \phi A n \, \mathrm{d}x \qquad (17)$$

$A \, \mathrm{d}x$ is the volume of the target and consequently $nA \, \mathrm{d}x$ is the total number of target atoms, N [Equation 18]:

$$\frac{\mathrm{d}N^*}{\mathrm{d}t} = \sigma \phi N \qquad (18)$$

Since we are speaking of a reaction of a specific nuclide, N is the total number of the atoms of that nuclide. Equation (18) is used mainly for irradiation in a nuclear reactor, where the target is located inside a uniform flux of neutrons. Equation (10) is used in cases of a narrow beam from charged particle accelerators.

In the case of prompt activation analysis, the only important number is the number of nuclei which were transformed by the nuclear reaction. Hence the number of nuclei transformed is given by the integration of Equation (16) or (18) over the irradiation time. However, for delayed activation analysis, the important factor is the number of nuclei decaying during the measurement period, which is delayed and done some time after the end of irradiation. In calculating the number of nuclei of the newly formed radionuclide at the end of the irradiation, it would be wrong to integrate those equations. The radionuclides are not only formed during the irradiation, but also are decaying. Combining Equations (18) and (4) leads to the complete equation for the rate of the change

of the number of radioactive nuclei during the irradiation in a constant flux of projectiles (ϕ) [Equation 19]:

$$\frac{dN^*}{dt} = \sigma\phi N - \lambda N^* \tag{19}$$

Integration of Equation (19) with the initial condition $N_0^* = 0$ leads to the equation for the number of radioactive nuclei at the end of irradiation for time t_i [Equation 20]:

$$N_{EOI}^* = \frac{\sigma\phi N}{\lambda}(1 - e^{-\lambda t_i}) \tag{20}$$

where t_i is the irradiation time and EOI stands for end of irradiation. In the product λt_i, λ should have the same units as t_i, whereas in the term before the parentheses the units should be the same as the time unit of ϕ.

For cases where $\lambda t_i \ll 1 (t_i \ll t_{1/2})$, the term in parentheses is approximately equal to λt_i ($e^{-\lambda t_i} \approx 1 - \lambda t_i$ for $\lambda t_i \ll 1$) and N_{EOI}^* is equal to $\sigma\phi N t_i$, i.e. the rate of formation multiplied by the irradiation time, since the decay is negligible.

For an irradiation time that is long relative to the half-life of decay ($t_i \gg t_{1/2}$, $\lambda t_i \gg 1$), the term in parentheses is equal to 1, which is its maximum value. This is the maximum number of nuclei which can be formed. Longer irradiation does not increase the number of radioactive nuclei, since the rate of formation by the nuclear reaction is equal to the rate of disappearance by radioactive decay. This statement assumes that N does not change substantially during the irradiation, which is common for the usual values of ϕ (10^{15} n s^{-1} cm^{-2}), σ (1 b) and t_i (<10 days). For the extreme cases $\Delta N/N$ is still usually $\leq 10^{-3}$. This maximum number of radionuclei formed is termed the saturation value or the number at saturation, N_{sat}^* [Equation 21]:

$$N_{sat}^* = \frac{\sigma\phi N}{\lambda} \tag{21}$$

The saturation value can be increased by increasing the flux of the bombarding particles or the amount of the irradiated nuclei (the mass of the target), but not by longer irradiation. Usually we speak of the mass of a specific element in the target and not of the number of atoms of this element, so it is desirable to transform N, the number of atoms, into m, the mass of specific element in the activated samples. The number of atoms N in m mg of element with atomic weight M is $(m/M)N_A$, where N_A is Avogadro's number. In contrast to the usual chemical measurements, where all natural isotopes of the same element react in almost the same way, this is not true for nuclear activation. For example, let us consider chlorine. Chlorine has two stable natural isotopes, ^{35}Cl with 75.77% abundance and ^{37}Cl with 24.23% abundance. Irradiation with thermal neutrons leads to an (n,γ) reaction forming the two radionuclides, ^{36}Cl and ^{38}Cl. ^{36}Cl cannot be used as IRN for activation analysis owing to absence of γ radiation (^{36}Cl is a pure β emitter) and its long half-life (the IRN is a radionuclide used for the identification and quantification of an element in activation analysis). Hence, it is not the total number of chlorine atoms that should be used in Equations (18–21) but rather the number of ^{37}Cl atoms alone. This is done by multiplying N by the abundance of the parent of the IRN (0.2423 for ^{37}Cl). Using f for the abundance, Equation (21) takes the form of Equation (22):

$$N_{sat}^* = \frac{m}{M} N_A f \frac{\sigma\phi}{\lambda} \tag{22}$$

The number of the nuclei of the IRN at the end of irradiation is obtained from Equation (16) as Equation (23):

$$N_{EOI}^* = N_{sat}^*(1 - e^{-\lambda t_i}) \tag{23}$$

The measurement of the γ activity (the rate that the γ-rays are emitted by the activated sample) is done not immediately after the end of irradiation, but rather after some time, which is called the decay time, t_d. The decay time (called also 'cooling' time) can be fairly short, being only the time required to transfer the sample from the irradiation port to the measurement station, or can be long. Long decay times are used when we are interested in IRNs with medium to long half-lives (they are also called medium- to long-lived IRNs). The high activity of the short-lived radionuclides at short times after irradiation obscures the activity of the longer lived radionuclides. Consequently, in order to measure medium- and long-lived IRNs, their measurement is delayed. This is the reason why the decay time is called also 'cooling time'. For determining the concentrations of many elements in the sample, the γ activity in the irradiated sample is measured several times (at least three) after different decay times. Since the number of atoms of the IRN at the beginning of the decay time is N_{EOI}, the number at the end of the decay period, t_d, is given by Equation (24) [according to Equation 5]:

$$N_{EOD}^* = N_{SOC}^* = N_{EOI}^* e^{-\lambda t_d} = N_{sat}^*(1 - e^{-\lambda t_i})e^{-\lambda t_d} \tag{24}$$

The subscripts EOD and SOC stand for end of decay and start of counting, respectively. The measurement of the γ activity is called counting, since it is done by the counting of the pulses formed in the γ-ray detection systems, due to the interaction of the γ-photons with the detector.

The counting is done during a time t_c called the counting time. The number of atoms of the radionuclides still undecayed at the end of the counting time, N_{SOC}^*, is given by Equation (25) [according to Equation 5]:

$$N_{EOC}^* = N_{SOC}^* e^{-\lambda t_c} = N_{sat}^*(1 - e^{-\lambda t_i})e^{-\lambda t_d}e^{-\lambda t_c} \tag{25}$$

The number of atoms decaying in the counting period (ΔN_c) is given by the difference in the number of atoms at the beginning of the counting period, N^*_{SOC}, and the number of atoms at the end of this period, N^*_{EOC} [Equation 26]:

$$\Delta N_c = N^*_{SOC} - N^*_{EOC} = N^*_{sat}(1 - e^{-\lambda t_i})e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \quad (26)$$

The actual number of events of decay recorded by the detection system, is smaller than ΔN_c owing to three different factors: (i) the intensity of the specific γ line (number of photons emitted per 100 atoms decayed) is in many cases $< 100\%$; (ii) not every γ-photon emitted reaches the detector, as the photons are emitted isotropically and only those in the direction of the detector will hit the detector; and (iii) some of the photons reaching the detector will not interact with it, and others will lose only part of their energy in the detector, and consequently will not be recognized as originating from the specific IRN. The last two factors are combined in a factor called the geometric efficiency of the detector, ε. This factor depends on the detector, the energy of the γ line and the distance of the sample from the detector. For large samples, ε depends also on the sample size. The factor ε is determined experimentally by counting the activity of calibrated standards (which can be bought commercially), for which the rate of disintegration is known accurately. The rate of decay of the standards is given on its certificate at specific date and time. Its activity at the time of calibrating our system is calculated using Equation (5). Note that Equation (5) applies both to the number of atoms in the sample and to the activity of the sample since the rate is $-dN^*/dt$, which is given by λN^*. Hence the number of counts recorded by the detection system, C, is given by Equation (27):

$$C = \varepsilon I_\gamma \Delta N_c = \varepsilon I_\gamma N^*_{sat}(1 - e^{-\lambda t_i})e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \quad (27)$$

where I_γ is the intensity of the specific γ line. Thus, for activation in a uniform flux, the number of measured counts is given by Equation (28):

$$C = mN_A f\sigma\varepsilon I_\gamma \phi \frac{1 - e^{-\lambda t_i}}{\lambda} e^{-\lambda t_d}(1 - e^{-\lambda t_c}) \quad (28)$$

For a beam smaller than the target, ϕ is replaced by I/A, where A is the area of the target, usually expressed by $m_{sample}/(\rho\, dx)$. It should be noted that here the mass is the total mass of the sample and not the mass of the determined element; $\rho\, dx$ is the common way in nuclear physics to express thickness and its units are grams per square centimeter.

Equation (28) shows that the number of counts recorded is proportional to the mass of the element responsible for the specific γ line, and this is the basis of the activation method since the measurement of the specific γ line yields the mass in the sample of the specific original nuclide.

For some elements, mainly gold and tantalum, the radionuclides produced by the (n,γ) reaction with the reactor neutrons have exceptionally high cross-sections (e.g. for ^{198}Au the cross-section is 2.5×10^4 b). Owing to the high cross-section and although their concentration is very limited, a radionuclide is formed by absorption of two neutrons (in two consecutive steps and not in one interaction) in the stable nuclide. For example, for gold, which has only one stable isotope, ^{197}Au, the process is illustrated by Equation (29):

$$\begin{array}{ccccc} ^{197}\text{Au} & \longrightarrow & ^{198}\text{Au} & \longrightarrow & ^{199}\text{Au} \\ & & ^{198}\text{Hg(stable)} & & ^{199}\text{Hg(stable)} \end{array} \quad (29)$$

For ^{198}Au, the rate of formation is given by Equation (30):

$$\frac{d[^{198}\text{Au}]}{dt} = [^{197}\text{Au}]\sigma_1\phi - (\lambda_1 + \sigma_2\phi)[^{198}\text{Au}] \quad (30)$$

Equation (30) is the same as Equation (19) substituting λ for $\alpha_1 = \lambda_1 + \sigma_2\phi$. Hence the number of ^{198}Au atoms at the end of irradiation is given by Equation (31):

$$\begin{aligned} N(^{198}\text{Au})_{EOI} &= \frac{\sigma_1\phi[^{197}\text{Au}]}{\lambda_1 + \sigma_2\phi}[1 - e^{-(\lambda_1+\sigma_2\phi)t}] \\ &= \frac{\sigma_1\phi[^{197}\text{Au}]}{\alpha_1}(1 - e^{-\alpha_1 t}) \end{aligned} \quad (31)$$

For ^{199}Au, the rate of formation is given by Equation (32):

$$\frac{d[^{199}\text{Au}]}{dt} = \sigma_2\phi[^{198}\text{Au}] - \lambda_2[^{199}\text{Au}] \quad (32)$$

This is similar to the set of the differential Equations (15), differing in that the constants in the positive term in one equation differ from those in the negative term in the next one, as not all of the disappearance of R_i leads to R_{i+1}; however the same method of Bateman can be used also here. The solution of Equation (32), substituting in it Equation (31), yields Equation (33):

$$[^{199}\text{Au}] = \frac{\sigma_1\sigma_2\phi_2}{\alpha_1\lambda_2}[^{197}\text{Au}^0]\left(1 - \frac{\lambda_2 e^{-\alpha_1 t} - \alpha_1 e^{-\lambda_2 t}}{\lambda_2 - \alpha_1}\right. \quad (33)$$

This equation neglects the disappearance of ^{199}Au due to the absorption of neutrons. An equation which includes it was given by Heydorn,[14] but its contribution is negligible for all fluxes used in activation analysis. The general treatment can be found in Friedlander et al.[3] (p. 201).

3 THE SHAPE OF A γ SPECTRUM

In optical (absorption and emission) spectrometry we know that if a species has only one excited state its spectrum will appear as a single line without any background. This is due to the yes/no quantum characteristic of the energy transfer. The energy difference between the two states is either transferred or not transferred, but it cannot be transferred in parts. A photon of light in the infrared, visible or ultraviolet range can be absorbed totally or not at all, but it cannot lose only part of its energy. The situation is different for high-energy γ-rays. Low-energy X-rays interact with matter by the photoelectric process. In this process the γ-photon interacts with one electron of the material, losing all its energy to this electron. The transferred energy is higher than the binding energy of the electron, and thus the electron is ejected from the atom with kinetic energy equal to this difference ($h\nu_{\text{photon}}$ − binding energy of the electron). However, since the range of high kinetic energy electrons is considerably shorter than that of γ-rays, the electrons will lose their kinetic energy inside the detector material and from the point of view of the detector the complete energy of the photon have been absorbed in it, resulting in the same energy absorbed for all photons (assuming that all photons are of the same energy). The situation is different for higher energy γ-rays. The cross-section (probablity of reaction = rate constant) for the photoelectric process decreases with increasing energy of the photon more than that of the Compton scattering. For higher energy photons the main interaction with matter is via Compton scattering. In germanium the photoelectric effect is the dominant process for γ interaction up to about 200 keV, whereas from 200 keV upwards the Compton scattering becomes more important. It should be stressed that also below 200 keV the Compton scattering occurs but its contribution to the interaction of the photon with the germanium is less than that of the photoelectric process. Its contribution increases with the energy although its cross-section decreases with the energy, since it decreases less than that of the photoelectric process. In the Compton scattering the photon interacts with what might be called a free electron (mainly electrons from outer shells whereas the photoelectric process occurs with inner shell electrons, mainly from the innermost one, the K shell). The photon transmits only part of its energy. The energy that the electron receives is more than sufficient to eject it from the atom and it moves with the excess energy as kinetic energy, losing it by collisions in a very short distance. The photon retains part of its energy, by changing its frequency since $E = h\nu$, and scattered in a different direction. In the Compton process the energy of the original photon is shared between two particles (the ejected electron and the scattered photon), and consequently there is a continuous distribution of energies of the scattered photon. In order to conserve both energy and momentum, the γ-photon cannot lose all its energy (in the photoelectric effect the momentum conservation is compensated by the recoiling of the atom from which the electron was ejected, whereas in Compton scattering it is an interaction with a 'free' electron). The maximum energy that the photon can lose is given by Equation (34):

$$\frac{E_\gamma}{1 + m_0c^2/2E_\gamma} \tag{34}$$

where E_γ is the energy of the initial photon and m_0 is the electron rest mass; m_0c^2 is the rest mass of an electron in energy units and it is equal to 0.511 MeV. The main difference between the photoelectric and Compton processes with respect to their responses in the detector material is that the electrons (in both processes) lose all their energy in the detector, owing to their short range, while the scattered photon might escape from the crystal without further interaction. In γ-ray spectrometry we are measuring the full-energy peak, called the photopeak, as this energy will be the only one to appear if the only interaction process is photoelectric absorption. The photopeak results either from the γ-photons losing all their energy by photoelectric absorption or from a Compton scattering process followed by photoelectric absorption of the scattered lower energy γ-photon in the detector (the photoelectric absorption can be after one scattering or several scatterings, all of them within the detector material). However, the scattered γ-photon from the Compton process (actually from the several consecutive Compton processes) might escape from the detector crystal, leaving in the detector less energy than the full energy peak. This escape of scattered photons not only reduces the photopeak, but its main disadvantage is the formation of a background for lower energy peaks, since part of the energy (those given to the electron) is absorbed in the detector and the detector sees them as counts of lower γ energy than the actual energy of the γ-photon emitted by the IRN. The scattered photons do not have discrete energy and can be between a minimum given by Equation (34) and a maximum given by the original energy of the photon (minus the small value of the binding energy of the photon, which is not completely a free electron). For one γ energy source the Compton scattering produces a continuum background ranging from zero up to a maximum, called the Compton edge, given by Equation (35). This results from the fact that the minimum energy that the scattered photon can have is not zero but rather is given by Equation (35):

$$E_{\text{minimum of scattered}\,\gamma} = E_\gamma - \frac{E_\gamma}{1 + m_0c^2/2E_\gamma} = \frac{E_\gamma m_0c^2}{2E_\gamma + m_0c^2} \tag{35}$$

For $E_\gamma \gg m_0c^2$, E_{minimum} approaches a value of $m_0c^2/2$, i.e. 256 keV. Thus, for high-energy γ the Compton edge will be separated from the photopeak by about 256 keV. For a one-energy source the range between the photopeak and the Compton edge is almost free of counts. However, most real samples have many IRNs having different γ energies and the Compton continuum stretches from zero up to the Compton edge of the highest γ energy.

If the energy of the measured photon is >1.022 MeV, a third process can also be operative in the interaction of the photon with the detector – a process called pair production. In this process, the photon energy is transformed, under the influence of the field of a nucleus, into matter in the form of an electron–positron pair (transformation of energy into a matter–antimatter pair). Since the rest mass of each electron or positron is 0.511 MeV, the threshold of this reaction is 1.022 MeV to conserve energy. Although the threshold energy is 1.022 MeV, the cross-section for pair production is very low, for the energy range of <1.5–1.6 MeV, and can be ignored. In this process of pair production, the excess energy is shared between the kinetic energies of the electron and the positron. Since the ranges of positrons and electrons are very short, this excess energy ($E_\gamma - 1.022$ MeV) will be deposited in the detector. When the positron has lost all its kinetic energy, it reacts with an electron in an annihilation reaction to form two γ-photons of 0.511 MeV each (transformation of matter–antimatter into electromagnetic energy). Each of these 511-keV photons can either escape from the detector without any interaction, or lose part of its energy by Compton scattering, or lose all its energy in the detector (either by photoelectric absorption or by succesive Compton scatterings and photoelectric absorption). If one of the 511-keV photons deposits all its energy in the detector, while the other one totally escapes from the detector, the energy absorbed in the detector will be 511 keV less than the full energy peak. With γ-photons of above 1.6 MeV we usually see also this peak of $E_\gamma - 511$ keV, where E_γ is the photopeak energy. This peak is called single-escape (SE) peak, due to one 511-keV photon escaping from the detector. Another peak in the spectrum is the double-escape (DE) peak, rising from the two 511-keV photons escaping from the detector without any interaction with the detector material. The energy of this peak is $E_\gamma - 1.022$ MeV. When analyzing a γ-ray spectrum to find the radionuclides in the analyzed spectrum, every time we find a peak with energy $E_\gamma > 1.6$ MeV, which we know belongs to some radionuclide, we have to remember that in the list of the found peaks there are some which do not belong to other radionuclides but are either SE or DE peaks. We saw that ^{28}Al has a γ energy of 1.778 MeV. Hence we expect to see in the ^{28}Al γ-ray spectrum peaks of 1.778 MeV, 1.269 MeV (SE = 1.778 − 0.511) and 0.758 MeV (DE peak). ^{24}Na has two γ lines of 2.75 and 1.39 MeV. We expect to have four lines in the γ-ray spectrum, 2.75 MeV, 2.24 MeV (SE), 1.73 MeV (DE) and 1.39 MeV. Although 1.39 MeV is above the threshold of 1.022 MeV for pair production, its cross-section for this process is low and pair production contributes very little to the interaction of the 1.39-MeV photon with the detector. Hence in most cases we will not see the SE and DE peaks of this photon.

Figure 1 shows the measured γ-ray spectrum of ^{24}Na. The ratio of escape peaks to the photopeak depends on the detector size (influencing the probability of the γ escape) and on the energy of the γ-photon (affecting the chance that the photon will interact by pair production). In order to know whether a peak is a photopeak or SE or DE, it is preferable to start the analysis of the spectrum from the highest energy peak. For each peak we can remove from the list of peaks its SE and DE peaks.

An important artefact in many γ-ray spectra is the sum peak, a peak which is due to the summation of two γ-photons. If two photons are interacting with the detector, and the time between them reaching the crystal is smaller than a time characteristic of the detector, the detector treats both of them as a single photon and the energy deposited is the sum of the energies of the two photons. The two photons can be due to different radionuclides,

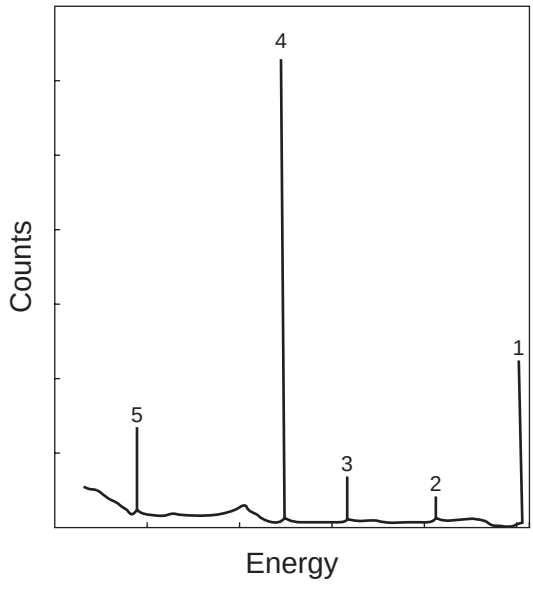

Figure 1 γ-Ray spectrum of ^{24}Na radioisotope taken with an HPGe detector. (1) Photopeak of 2754 keV; (2) SE peak of 2243 keV (2754–511 keV); (3) DE peak at 1732 keV (2754–1022 keV); (4) photopeak at 1369 keV; (5) 511-keV positron annihilation.

or to the same radionuclide which has more than one γ line. In the case of different radionuclides the sum peak is refer to 'pile-up' and can be reduced (relative to the photopeaks) by removing the measured sample further from the detector or by analyzing a smaller sample. A lower rate of counting increases the average time between two photons reaching the detector, thus decreasing the chances of pile-up. In the spectrum of only one γ line there is also a peak with energy given by Equation (35). This peak is due to 180° backscattering from external sources, mainly the shielding of the detector, of the original photon. The backscattered Compton photon reacts with the detector, giving a peak of its full energy. For a source with various γ-ray energies, this backscatter peak can range from 70 to 256 keV, thus adding to the Compton continuum rather than forming a distinct peak.

The Compton background is an intrinsic background due to the physical characteristics of the detector and the interaction of radiation with matter. Another source of background is external, due either to radioactive sources other than our sample or to cosmic radiation. This external background can be minimized by the use of shielding, done mainly with lead owing to its high Z and mainly its high density. Usually the detector is located inside a lead 'castle'. Lead shielding configurations suitable for a germanium detector are available commercially, usually in the form of a hollow cylinder with a sliding lid or revolving door for access.

In many cases where the Compton continuum interferes with the measurements of mainly small peaks, it can be reduced by the use of the Compton suppression method. In this method, the germanium detector is located inside an annular scintillation detector or inside a ring of scintillation detectors. Many of the scattered Compton γ-rays which escape from the germanium detector (in the case of the Compton continuum) deposit some energy in the surrounding detectors, whereas those photons which lose all their energy in the germanium detector do not cause any signal in the scintillation detector(s). The system is operated in the anticoincidence mode, i.e. the multichannel analyzer (MCA) is recording only the events which occur only in the germanium detectors. If within a fixed time the pulse in the germanium detector is followed with a pulse in the scintillation detector (indicating that the germanium pulse is in the Compton continuum and not part of a photopeak), this event is discarded. The scintillation detector(s) are usually NaI(Tl), although in some systems it is a crystal of bismuth germanate (BGO) ($Bi_4Ge_3O_{12}$) owing to its higher efficiency for absorption of γ-rays (since it has a high-Z element). However, BGO crystals are more expensive and are not used in most systems. A schematic diagram of a Compton suppression system is given in Figure 2.

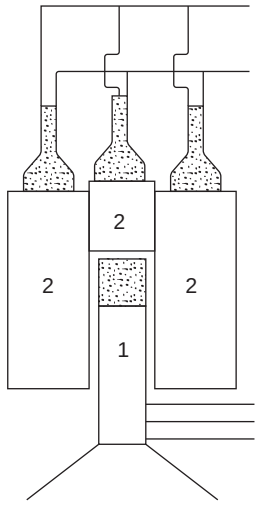

Figure 2 Schematic diagram of anticoincidence Compton suppression γ-spectrometer. (1) HPGe detector with liquid nitrogen cooling; (2) NaI(Tl) detectors with photomultipliers.

4 NEUTRON SOURCES

Three main sources of neutrons are available: (1) research nuclear reactors (nuclear reactors used as neutron sources), (2) ion and electron accelerators including neutron generators and (3) radioactive sources.

Research nuclear reactors have the highest neutron fluxes, but are limited concerning their price and availability. Consequently, nuclear reactors will be used for INAA of very minute amounts, which is the main subject of this article. Research nuclear reactors are usually large devices in which fissionable material, almost exclusively ^{235}U, is fissioned into two nuclides with simultaneous emission of neutrons which induce further fissions in a chain reaction. The fission-produced neutrons are very energetic. The cross-section for neutron-induced fission of fissionable nuclides increases with decreasing energy of the neutrons, and in order to increase the neutron activity, moderators which slow the neutrons are added to the reactor. To reflect back some of the neutrons which leak from the reactor core, reflectors are used. The fission process releases large amounts of energy, mainly due to the stopping of the two recoiling fissioned particles, and the system is cooled by a coolant (either liquid or gas). Nuclear reactors are categorized according to their fuel, moderator, coolant, reflector and configuration.

Almost all research nuclear reactors (neutron sources) are heterogeneous reactors in which the fuel is in the form of rods. The fuel is enriched ^{235}U (natural uranium has only 0.7% of ^{235}U, the fissile material). Most research reactors have 93–99% ^{235}U. Many of the reactors have rods which are U–Al alloys, but some

of the newer designs (mainly those converted to 20% ^{235}U) are of the uranium silicide type. TRIGA reactors operate with uranium–zirconium hydride fuel, which, owing to their large negative temperature coefficient of reactivity, allow the operation of the reactor in pulses. In the light water reactor (LWR), ordinary water (H_2O) is used both as a moderator and as a coolant. The reflector is mainly graphite but there are also Be and H_2O reflected reactors. The construction is either pool type or tank-in-pool type. Owing to the relative high cross-section for capturing thermal neutrons by H atoms and therefore the relatively small amount of moderator, the flux of neutrons in LWRs always contains a large fraction of fast and epithermal neutrons. The available powers are in the range 10–5000 kW with neutron fluxes of 5×10^{10}–1.5×10^{14} n cm^{-2} s^{-1}. Many reactors are unique in their design, but there are some commercial types which are more common, e.g. the American TRIGA and the Canadian Slowpoke. The TRIGA reactor is a popular multipurpose research reactor. About 50 of them are operating with power levels of 18 kW–3 MW fluxes of 7×10^{11}–3×10^{13} n cm^{-2} s^{-1}. The most common types are those of 250 kW and 1 MW. They are of the pool type, graphite reflected with uranium–zirconium hydride fuel with ^{235}U enrichment of 10–70%. The Slowpoke reactor is a low-power (20 kW) reactor designed specifically as a teaching aid with additional purposes of activation analysis and the production of small amounts of radioisotopes. The system is designed to operate remotely. It can be provided with up to five irradiation sites in the core with a flux of 10^{12} n cm^{-2} s^{-1} and five further tubes outside the reflector with half of that flux.

Heavy water research reactors are tank types. They contain usually enriched uranium fuel. Heavy water is used as both moderator and cooler, and neutrons are reflected from the heavy water and graphite. Owing to the low cross-section for thermal neutron absorption by D and O, and owing to the higher mass of moderator, they are characterized by a well-thermalized neutron flux (very little epithermal and fast neutron fluxes except inside the core). Owing to the lower moderation power of D compared with H, the physical size of heavy water reactors is larger and hence they have a large available irradiation volume. Their power is usually between 10 and 26 MW (fluxes of up to 2×10^{14} n cm^{-2} s^{-1}).

4.1 Sample Introduction

The way of introducing a sample into the neutron flux depends on the physical structure of the reactor. It is essential that the introduction of the sample will not affect the operation of the reactor. The irradiation site may be within the reactor core or outside in the moderator/reflector region. If the reactor is an open-pool type with access from above, vertical tubes or ropes can be installed to lower the samples down inside the core or close to the side. The samples are closed in sealed ampules in order not to be in contact with the water surrounding the core. The ampules are usually made of aluminum owing to its corrosion resistance and short-lived activation products. For short irradiations a polyethylene capsule can also be used. The manual loading of samples is neither quick nor reproducible in time when short irradiations are performed. In order to have quick and reproducible sample introduction, mechanical systems are used. Two types of mechanical systems are used: chain-driven racks and pneumatic devices. The latter is the most common, owing to the shorter loading and unloading time and less maintenance and failures of operation. In the pneumatic device, the sample is pushed along a tube with pressurized gas (air or nitrogen). The transfer time depends on the pressure and the distance transferred. In many systems the transit time is 1 s or less. The pneumatic device can be automated very easily. Figure 3 depicts schematically the pneumatic system.

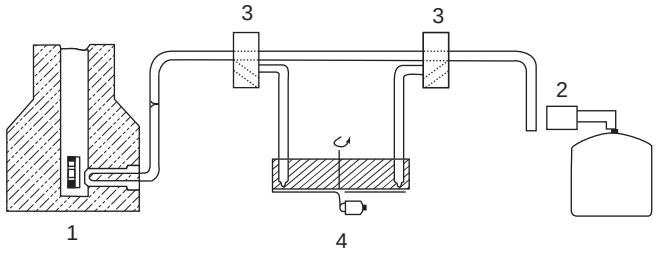

Figure 3 Schematic layout of a fast transfer system for INAA. (1) Reactor; (2) γ-spectrometer; (3) valves; (4) sample changer; the valves can allow also direct transfer of rabbit from reactor to detector or indirect transfer through the sample changer for decay.

With closed-tank reactors, the irradiation is done either by a pneumatic device or with neutron beams.

5 INSTRUMENTAL NEUTRON ACTIVATION ANALYSIS

5.1 Techniques

INAA is one of the simplest techniques for trace element analysis. Samples are inserted into polyethylene or polypropylene cans for short irradiation or into a quartz vial which is inserted in an aluminum can for long irradiation. PVC cans or glass vials cannot be used, since they will be become very radioactive owing to the formation of ^{38}Cl or ^{24}Na, respectively. For a short irradiation time each sample is 'sent' alone

to the reactor by means of a pneumatically operated tube – the rabbit system. For long irradiation times, many samples can be lowered into the reactor and left there for the required time. After the samples have been removed from the reactors, they are counted (this means taking a γ-ray spectrum) immediately or the γ-ray spectrum is taken after some 'cooling' (decay) time, to reduce the high radioactivity level due to short-lived radionuclides. Before counting, the vials are washed to remove surface contaminants. In some cases the samples are counted inside the vial, whereas in others the samples are taken out of the vial before the measurement, owing to interferences from trace elements in the vial.

5.2 Calculation

In general, the first calculation step consists of peak identification, i.e. identification of which peaks are photopeaks (discarding SE, DE and sum peaks) and identification of the radioactive species which are producing them. There are tables of γ-ray energies listed in order of increasing energies for ease of identification.[8–13] Some tables give the energies of all known radionuclides whereas others give them only for radionuclides which can be formed in one form of nuclear activation, e.g. in a nuclear reactor. For better identification the tables give the half-life of this radionuclide, its probability of formation and other γ lines of the same radionuclide. Thus, if we find one γ peak of some radionuclide, but not its other line/lines, it means a wrong identification. The different lines also have to be checked to be in the reported ratios, taking into account the relative intensities of the γ lines and the efficiency of the detector at those energies. Nowadays most MCAs are computer (PC)-based and have in their data a nuclide library. However, most libraries check only for the energy of the photopeak with a preset tolerance of 1–2 keV. Some commercial computer programs check also for the presence (in the right ratios) of the other lines of the same radionuclide. Most programs do not treat the identification through half-lives. It should be remembered that SE and DE energies are not listed in most libraries and this can lead to misidentifications. Figure 4 gives an example of γ-ray spectra measured from an activated sample.

The second step in the calculation is calculation of the net counts under the peak (plotted as number of counts in each channel per channel number). Various methods exist for the evaluation of the net peak 'area' (really it is the net peak counts). The simpler methods define the channels which belong to this peak and sum all the counts in that peak [Equation 36]:

$$\text{total} = \sum_{i=L}^{u} C_i \tag{36}$$

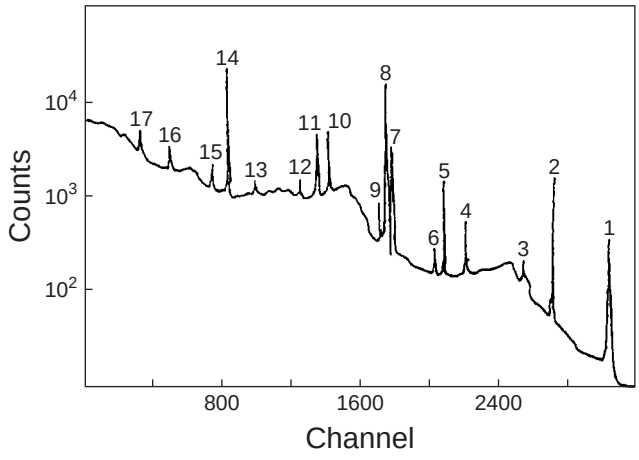

Figure 4 γ-Ray spectrum of a neutron-activated geological sample measured with an HPGe system 15 min after the end of irradiation. (1) 3084 keV ^{49}Ca; (2) 2754 keV ^{24}Na; (3) 2573 keV SE of ^{49}Ca; (4) 2243 keV SE of ^{24}Na; (5) 2113 keV ^{56}Mn; (6) 2062 keV DE of ^{49}Ca; (7) 1811 keV ^{56}Mn; (8) 1779 keV ^{28}Al; (9) 1732 keV DE of ^{24}Na; (10) 1434 keV ^{52}V; (11) 1369 keV ^{24}Na; (12) 1268 keV SE of ^{28}Al; (13) 1014 keV ^{27}Mg; (14) 846 keV ^{56}Mn + ^{27}Mg; (15) 757 keV DE of ^{28}Al; (16) 511 keV β^+ annihilation; (17) 320 keV ^{51}Ti.

where C_i is the counts in channel number i, and L and u are the lower and upper channels belonging to the peak. However, this total count value contains also the background (Compton and external), which should be subtracted. The background is estimated by taking the average number of counts in 3–5 channels on both sides of the peak, multiplying by the number of channels in the peak [Equation 37]:

$$\text{background} = \frac{u - L + 1}{2 \times 3} \left(\sum_{i=L-3}^{L-1} C_i + \sum_{i=u+1}^{u+3} C_i \right) \tag{37}$$

The value of 3 in this equation can be increased or decreased, but it is about the optimum value; lower numbers will increase the statistical error in the evaluation of the background, and larger numbers will increase the possibility that the following channels belong already to other peaks. The radioactive decay is not an absolute process but rather of a statistical nature, and consequently the larger the number of counts the more accurate is the result. The standard deviation of each decay count (as long as the counting time is small compared with the half-life of the measured radionuclide) is equal to the square root of the number of counts. The number of net counts is given by Equation (38):

$$N = \text{net counts} = \text{total} - \text{background} = T - B \tag{38}$$

The standard deviation of the net counts, σ_N, and the relative standard deviation, $\%\sigma_N$, are given by

Equations (39) and (40):

$$\sigma_N = \sqrt{T+B} = \sqrt{N+2B} \tag{39}$$

$$\%\sigma_N = \frac{100\sqrt{T+B}}{N} = \frac{100\sqrt{N+2B}}{N} \tag{40}$$

It should be kept in mind that the rate of counting is limited both by the rate of the process in the detector (which causes 'pile-up' or a sum peak in the case of high rates of counts) and by the rate of the sorting-out of pulses by the MCA and the electronic system. During the 'busy' period of the electronic system, they cannot acquire another signal, so that each signal reaching the system during its 'busy' period will be discarded. This 'busy' period is called the dead time and the actual measurement time (live time) is smaller than the real time (measured by the clock). Most MCA systems contain a device for correction of the dead time. However, these devices are accurate only up to about 10% dead time. It is usually recommended not to measure with dead times above 10%. In spite of this recommendation, there are cases where measurements with high dead times are inevitable, e.g. the measurement of a small amount of a short-lived IRN in the presence of a considerably higher activity of longer lived radionuclide.[15] In this case the correct live time can be measured[16] by the use of a pulse generator with a known rate at which its pulses are introduced to the same electronic system (the height of the pulse is chosen in a range were there are no γ signals), calculating the live time from the ratio of the measured to expected counts from the pulse generator. Instead of using a pulse generator, the longer lived radionuclide can be measured after partial decay or preferably at larger distance.[17]

5.2.1 Calibration and Calculation

The amount of a nuclide in a sample can be calculated from the counts of the radionuclide produced from its neutron activation by the use of Equation (28) if the physical constants $\sigma, f, I_\gamma, \phi$ and λ are known and the geometry factor ε is measured by using calibrated radioactive standards, which can be bought commercially. These standards are calibrated in a 4π detector, i.e. a detector which surrounds the source from all directions such that any β-particle or photon interacts with the detector. The geometry factor ε depends on the distance of the source from the detector and the energy of the γ line. For a fixed distance, ε is a maximum type function of the energy, the maximum being at 100–150 keV. For higher energies than the maximum, $\log\varepsilon$ is approximately a linear function of log(energy), although for more accurate interpolation a quadratic dependence of $\log\varepsilon$ on log(energy) is used. Standard sources often used for calibration are ^{241}Am (60 keV), ^{131}Ba (81, 302, 356 and 383 keV), ^{137}Cs (661 keV), ^{60}Co (1773 and 1332 keV), ^{22}Na (1275 keV) and ^{88}Y (898 and 1836 keV). These standard radioactive sources are used also for the calibration of the number of channels vs. the appropriate γ energy. The standard radioactive sources are supplied with a certificate, certifying their activity at a fixed date. The activity is given either in microcuries [$1\,\mu\text{Ci} = 3.7 \times 10^4$ dps (disintegrations per second)] or in kilobecquerels ($1\,\text{kBq} = 10^3$ Bq, 1 Bq = 1 dps, $1\,\mu\text{Ci} =$ 37 kBq). The activity of the source at the measurement is calculated by the use of Equation (41) which is derived from Equation (2):

$$A(t) = A_0 \exp(-\lambda t) \tag{41}$$

where $A(t)$ is the activity at the measurement time, A_0 is the certified activity and t is the time elapsed from the date of certification to the measurement.

The parameters f, I_γ and λ are tabulated and known very accurately. However, σ is a function of the energy of the neutrons and the real term in Equation (28) should not be $\sigma\phi$ but rather the integral $\int_0^\infty \sigma(E)n(E)\,\mathrm{d}E$, where $n(E)$ is the energy – flux density or the flux distribution function (number of neutrons per second per square centimeter for which their energy is in the range $E + \mathrm{d}E$ to E), $\phi = \int_0^\infty n(E)\,\mathrm{d}E$. These functions $n(E)$ and $\sigma(E)$ are not always known and sometimes they change with time and with location in the reactor. In order to overcome this problem, the use of comparators has been applied to calculate the results of neutron activation analysis (NAA). A comparator is a standard which contains a known amount of the element to be determined in the sample. If the sample (denoted by subscript s) and the comparator (denoted by subscript c) are irradiated simultaneously and counted under the same conditions, then Equation (26) gives Equation (42):

$$m_\mathrm{s} = m_\mathrm{c}\frac{C_\mathrm{s}}{C_\mathrm{c}}\frac{D_\mathrm{c}}{D_\mathrm{s}}\frac{M_\mathrm{c}}{M_\mathrm{s}} \tag{42}$$

where D is the decay factor [$D = \exp(-\lambda t_\mathrm{d})$] and M is the measurement time factor [$M = 1 - \exp(-\lambda t_\mathrm{c})$], where t_c is the counting time, not related to the subscript c representing the comparator. This method of comparators is the most accurate one, and it is the best one to use if the concentration of a single element (or even a few elements) in the sample is required. It should be noted that the sample and the comparators are of similar size and composition to compensate for unequal self-absorption of the measured γ-photons in the sample or comparator. This problem of self-absorption is almost negligible if both sample and comparators are small. The main problem in the use of a comparator for each element is when INAA is used for multielement analysis, which is the common case. In some samples more than 25 elements

are determined simultaneously (more accurately, by three irradiations for different periods of times followed by 5–6 countings at different times after the irradiations for various periods of times).[18–23] For these multielement determinations it is difficult to prepare comparators for all the elements, and further it become impractical owing to the large amount of space required for all the comparators. Thus for multielement analysis either multielemental comparators have to be used, or one or two element comparators are used for calculation of other elements. Some multielement standards (comparators) of different types (geological, biological, etc.) have been prepared either by groups of scientists or by governmental or international organizations. However, such a standard [standard reference material (SRM)] is not recommended as a primary standard owing to some large uncertainties in some elements and problems with homogeneity of small samples (of the SRMs) used in INAA.

The use of one or two elements as comparators for all elements has become more widely used in recent years. The advantage of using two elements or one element with two isotopes which are activated is that the use of two activated nuclides, if chosen well, can give more information on the energy distribution of the neutron flux in the reactor.[24–26] The most popular method is the so-called k_0 method, in which k_0 values were determined for most of the elements according to Equation (43):

$$k_0 = \frac{B_c}{B_s} \tag{43}$$

where [Equation 44]

$$B = \frac{C}{fI_\gamma \sigma_{th}} \tag{44}$$

and the subscripts c and s stand for comparator and sample/element, respectively, and σ_{th} is the cross-section for reaction with thermal neutrons. The most popular comparator is Zr, which has two activatable isotopes, used to monitor the thermal and epithermal flux simultaneously. The calculation of the amount of the different elements involve both the thermal and epithermal neutrons.

5.3 Nuclear Interferences

There are two main kinds of interferences in the calculation of trace element concentration by INAA. The first is the formation of the same radionuclide from two different elements. This would be almost impossible if all neutrons were thermal neutrons since then the only possible reaction is (n,γ). However, all reactors have some higher energy neutrons, which can induce other reactions. Examples of this case are the determination of Mg in the presence of Al, Cr in the presence of Fe or Al in the presence of Si or P. Mg can be determined only through its less abundant isotope $^{26}Mg(n,\gamma)^{27}Mg$. The cross-section for this reaction is rather low, making this measurement not very sensitive. It suffers interference since ^{27}Mg can be formed from nonthermal neutrons reacting with Al in an (n,p) reaction, $^{27}Al(n,p)^{27}Mg$. Similarly, the determination of chromium by the $^{50}Cr(n,\gamma)^{51}Cr$ reaction suffers interference by the nonthermal neutron reaction $^{54}Fe(n,\alpha)^{51}Cr$. Al is determined by the ^{28}Al,$^{27}Al(n,\gamma)^{28}Al$ reaction, which can be produced also by the reactions $^{31}P(n,\alpha)^{28}Al$ and $^{28}Si(n,p)^{28}Al$. In a flesh sample it was found that the contribution of P to ^{28}Al is more than 10 times that of the ^{28}Al from Al, in most reactors.[27–31]

Even for only thermal neutrons in few cases a specific IRN can be formed from two elements, in cases where the IRN can be formed also by consecutive reaction with two neutrons. Pt has several IRNs but all of them except one are short-lived and suffer from interferences from the major element background. Hence, for minute amounts of Pt it is determined by ^{199}Au formed in the process shown by Equation (45):

$$^{198}Pt \longrightarrow {}^{199}Pt \longrightarrow {}^{199}Au(3.139\,d,\ 158.3\,keV\gamma line) \tag{45}$$

However ^{199}Au is formed also by a two-neutron reaction with ^{197}Au, since ^{198}Au has a very high cross-section for neutron capture (2.65×10^4 b)[32,33] [Equation 46]:

$$^{197}Au \longrightarrow {}^{198}Au \longrightarrow {}^{199}Au \tag{46}$$

Another kind of interference is from two radionuclides having very close γ lines. An example of this kind of interference is the 846.8-keV line of ^{56}Mn and the 843.8-keV line of ^{27}Mg. Nowadays very good spectrometers are able to separate these two peaks if they have comparable activities but this is impossible for low-resolution germanium detectors or when one activity is much higher. If they cannot be separated by the spectrometer there are two ways to overcome this interference:

1. Since the half-life of ^{56}Mn (2.56 h) is longer than that of ^{27}Mg (9.45 min), decay of 2 h will leave practically only ^{56}Mn (the activity of ^{27}Mg will decrease by a factor of 6647 whereas the ^{56}Mn activity will be decreased by a factor of less than two). Thus measurements of the sample after the irradiation and after decay of 2–3 h will give the activities of both ^{56}Mn (from the delayed measurement) and ^{27}Mg (by subtracting the corrected activity of ^{56}Mn from the first measurement).
2. The activities of both radionuclides can be calculated using the fact that ^{27}Mg also has another γ line at 1014 keV. The ratio of the activities of the two lines

can be measured in a pure sample of Mg, and thus the 843.8-keV activity of ^{27}Mg can be calculated from the 1014-keV activity. Subtracting this activity from the combined (843.8 + 846.8)-keV peak yields the activity of ^{56}Mn 846.8-keV peak.

5.4 A Test Case: Instrumental Neutron Activation Analysis of Trace Elements in Silicon

Silicon has three naturally stable isotopes, ^{28}Si (92.21%), ^{29}Si (4.70%) and ^{30}Si (3.09%). Thus only ^{30}Si produces a radionuclide by the usual (n,γ) reaction of thermal neutrons. ^{31}Si produced by the ^{30}Si(n,γ)^{31}Si reaction has a 2.62-h half-life and very few γ-rays, with an intensity of only 0.07% (1.26 MeV). Considering the low abundance, the relatively low cross-section (0.28 b) and mainly the low intensity of the γ-rays, ^{31}Si does not present a large problem in instrumental analysis by γ-ray spectrometry. However, owing to the presence of fast and epithermal neutrons in the reactor, the main activity induced in silicon is due to ^{28}Al, ^{29}Al and ^{27}Mg formed by the reactions ^{28}Si(n,p)^{28}Al, ^{29}Si(n,p)^{29}Al and ^{30}Si(n,α)^{27}Mg. As the half-lives of these radionuclides are short (2.3, 6.6 and 9.5 min, respectively), INAA is possible for radionuclides with half-lives longer than 1–2 h. In any case, the activity induced in nuclides with shorter half-lives is usually not sufficient to measure the required purity for electronic-grade semiconductors, owing to the small amount of radioactive atoms which can be formed until reaching saturation. As there is no background from the Si matrix after about a day or two and since the concentrations of the trace elements are low, the ways to increase the sensitivity (decrease the lower limit of detection) are to irradiate with a larger total number of neutrons (higher flux, longer irradiation times or both). In the high flux reactor (HFR) at Petten, The

Table 1 Limits of detection of trace elements in Si by INAA (ppb by weight)[a]

Element	Ref. 35	Ref. 36	Refs. 37 and 38	Ref. 39	Element	Ref. 35	Ref. 36	Refs. 37 and 38	Ref. 39
Na	0.5	0.1	0.3	0.1	Sb	<0.18	0.01	0.02	0.006
Mg		3000			Te	0.004		0.2	
Al		500			I	8			
Cl		5			Cs	0.0005	0.05		
K	0.15	100	0.6		Ba	0.06	10	1.5	
Ca	10	2000			La	0.00015	0.01	0.0007	
Sc	0.00003	0.003	<0.002		Ce	0.0009		0.1	
Ti	3	1000	30		Pr	0.006			
V			10		Nd	0.01			
Cr	0.002	1	0.02	0.1	Sm	0.00003	0.01	0.0004	
Mn	15	0.1			Eu	0.00007	0.05	0.002	
Fe	0.3	50	4	1.6	Gd	0.006			
Co	0.0005	0.1	0.1	0.001	Tb	0.0001		0.006	
Ni	0.15	2000	3		Dy	300			
Cu	<0.56	0.5	0.1	0.1	Ho	0.003			
Zn	0.015	1.0	0.2	0.3	Er	0.04			
Ga	0.0015		0.05		Tm	0.003			
Ge	1				Yb	0.0002		0.001	
As	<0.033	0.02	0.001	0.008	Lu	0.0004			
Se	0.002		0.1		Hf	0.003	20	0.005	
Br	<0.0024	0.02	0.01		Ta	0.005	0.05	0.01	
Rb	0.01				W	<0.025	0.02	0.0001	
Sr	0.15		4		Re	0.0003			
Y	200				Os	0.0007			
Zr	0.15	500	2		Ir	0.000004		0.0002	
Nb	1.5				Pt	0.004		0.005	
Mo	0.006		0.01		Au	<0.0016	0.0005	0.0001	0.0008
Ru	0.0015		0.1		Hg	0.0006		0.7	
Pd	0.06				Tl	3			
Ag	0.003	0.01	0.08		Pb	300			
Cd	0.015		0.3		Th	0.0002	0.1	0.0015	
In	0.004		0.3		U	0.001	0.3	0.02	
Sn	0.2								

[a] Verheijke et al.[35] irradiated wafers of 15 mm diameter and 0.5 mm thickness (20.6 g) with a flux of 4×10^{13} n cm^{-2} s^{-1} for 72–96 h and counted for 6–8 h; Lindstrom[36] irradiated samples of 100 mg for up to 6 h at 5×10^{13} n cm^{-2} s^{-1}; Revel et al.[37,38] used 1 g of Si and irradiated for 72 h at 2.3×10^{14} n cm^{-2} s^{-1}; Fujingawa and Kudo[39] irradiated 5–10-g samples for 11 d at 2×10^{13} n cm^{-2} s^{-1}.

Netherlands, a special irradiation facility was constructed for irradiation of silicon samples from Philips, in which silicon wafers of up to 15 cm in diameter can be irradiated with 4×10^{13} thermal $n\,cm^{-2}\,s^{-1}$ and the irradiation is done for 72–96 h.[34,35] They also used large Ge(Li) detectors (100–150 cm^3) and counted for 8–16 h, and obtained the lowest detection limits. Table 1 summarizes some of the limits of detection given in the literature for an Si matrix.[35–39] For some elements, e.g. Cl and Mn, shorter irradiations were used, between 0.5 and 2.0 h. A special correction has to be done in the case of the determination of Na owing to the reaction $^{28}Si(n,p\alpha)^{24}Na$, as was shown by Niese[40] by irradiating very pure Si in a reactor core with and without a Cd cover to absorb the thermal neutrons. The contribution of the $^{28}Si(n,p\alpha)^{24}Na$ reaction to ^{24}Na in a silicon matrix depends on the spectrum of the neutrons in the reactor core, as very high energy is required for this reaction and the Q value exceeds that of a 14-MeV neutron generator.[40] Niese found[40] that the contribution of the $^{28}Si(n,p\alpha)^{24}Na$ reaction is equal to 0.7 ppb. For the HFR at Petten, The Netherlands, the contribution is 0.55[35] or 0.6 ppb.[34] Revel et al.[37] found this contribution of the $^{28}Si(n,p\alpha)^{24}Na$ reaction to be <0.4 ppb. Haas et al.[41] warned against the common use of wrapping the sample for irradiation in Al foil, as they showed that there is recoil of ^{24}Na from the Al wrapping to the sample, due to the high recoil energy from the $^{27}Al(n,\alpha)^{24}Na$ reaction. In many other uses it is not important, but owing to the very low concentration of Na in pure Si, this contribution might be considerable.

Several important elements cannot be determined in this method, including mainly the light elements H, Li, Be, B, C, N, O and F as they do not form radioisotopes or they are too short lived. Phosphorus cannot be determined in this way owing to a lack of γ emission from the only produced ^{32}P; however, ^{32}P β-rays can be determined destructively by the extraction of the phosphomolybdic complex in the presence of hold-back carriers of Ta and Au.[42] The chemical yield of the extraction was found to be $83.8 \pm 4\%$ and the decontamination factors were 79–6500 for the different elements studied. The limit of detection was found to be 3×10^{-11} g. They carried out their study[42] for phosphorus-doped silicon and Alfassi and Yang[43] showed that for this case, there is no need for chemical separation and 10–20 d of cooling are sufficient to ensure that the only β emitter is ^{32}P (with an accuracy of 1–2%, which is much better than the reproducibility of the extraction yield). The limit of detection should be about the same, as both methods use liquid scintillation counting for the measurement of the β activity, although Alfassi and Yang used a discriminator which reduces the detection limit by a factor of two.

5.5 An Example of Interference Removal in Instrumental Neutron Activation Analysis: Determination of Platinum in Aerosols[33]

Pt is emitted into the air by corrosion of the catalytic converters used in cars to reduce the emission of CO, NO_x and unburned hydrocarbons. In order to determine the concentration of Pt in the air, a high-volume air-pump pumps air through a filter and the amount deposited on the filter (together with the aerosols) is determined. Pt can be determined via several radionuclides. ^{191}Pt suffers from very low abundance (0.01%) of the parent stable nuclide, ^{190}Pt. ^{197}Pt and ^{199}Pt suffer from relatively short half-lives (18.3 h and 30.8 min). If a measurement is made shortly after the irradiation, these IRNs could not be observed owing to the large amounts of Compton-scattered photons from the major elements Na, Cl and Br. Thus Pt can be determined in minute amounts only via ^{199}Au [3.14 d, 158.4 keV (36.9%)], which is formed by the sequence of reactions in Equation (47):

$$^{198}Pt(7.2\%) \longrightarrow {}^{199}Pt \longrightarrow {}^{199}Au \qquad (47)$$

^{199}Au suffers from the two kinds of interferences mentioned earlier: (1) it can be formed also from other elements besides Pt and (2) another radionuclide formed from other elements have a very close γ line. In the following paragraphs we will show how these interferences can be overcome.

1. ^{199}Au can be formed also from the successive absorption of two neutrons by ^{197}Au owing to the very high cross-section for the (n,γ) reaction of ^{198}Au (26 500 b) and the high cross-section for the (n,γ) reaction of ^{197}Au (98.7 b) [Equation 48]:

$$^{197}Au \longrightarrow {}^{198}Au \longrightarrow {}^{199}Au \qquad (48)$$

   Since ^{198}Au is a radionuclide with a close half-life to ^{199}Au (2.69 d), it can be measured simultaneously with ^{199}Au, and can be used to calculate the concentration of Au in the sample and hence to calculate the contribution of Au to the measured peak of ^{199}Au. Subtracting this contribution, we obtain the counts of ^{199}Au formed from ^{198}Pt and hence the concentration of Pt.
2. Another problem is the very close γ lines of ^{199}Au (158.4 keV) and ^{47}Ca (159.3 keV). If similar counts are obtained for the two radionuclides, the γ spectrum shows clearly two distinct peaks, although the lower one (158.4 keV) does not go to zero, and hence there is lower accuracy in the calculation of the area under each peak (i.e. the counts due to each radionuclide). The situation is worse when the number of counts

of one of the radionuclides is considerably higher (ratio > 4), in which case the spectrum does not show two distinct peaks but one peak with a shoulder, more or less visible. In order to separate this composite peak into its two contributions, two methods can be used. An important remark concerning this measurement is not to use a computer convolution program to calculate the area of this peak, as it will be programmed with a fitting function which is characteristic for a pure photopeak. In this case it is better to calculate the area by a simple integration of the counts under the peak, subtracting a trapezoid for the background correction.[15] The two methods for separation of the contributions are based on (i) the different temporal behaviors of the two contributions and (ii) another γ line which allows the calculation of the contribution of one radionuclide to the mixed 158.4–159.3 keV peak, facilitating the calculation of the contribution of the other one by subtraction.

(a) Temporal separation: although ^{199}Au and ^{47}Sc have close half-lives which usually will not allow the temporal separation of their contributions, their parents [as none of them is produced directly from the (n,γ) reaction and are products of β decay of the (n,γ) products] have considerably different half-lives and hence the two contributions can be separated by measuring twice and solving the two equations (obtained by writing equations for the expected counts, writing x and y for the masses of Pt and Ca) with the two unknowns m_{Pt} and m_{ca} (or x and y, $x = m_{Pt}$ and $y = m_{ca}$) [Equations 49 and 50]:

$$^{198}\text{Pt} \longrightarrow {}^{199}\text{Pt} \longrightarrow {}^{199}\text{Au} \qquad (49)$$

$$^{46}\text{Ca} \longrightarrow {}^{47}\text{Ca} \longrightarrow {}^{47}\text{Sc} \qquad (50)$$

The disadvantage of this method is the need to make two measurements, hence the counting time required is more than doubled, since due to the decay between the two measurements the counting time of the second one must be longer than the first one in order to obtain the same statistical accuracy.

(b) ^{47}Sc does not have a γ peak without interference to calculate its contribution to the composite peak; however, its parent ^{47}Ca has a good γ line at 1297 keV without interference and this γ line can be used to measure the concentration of Ca and consequently its contribution via ^{47}Sc to the 158–159-keV peak.

The above contributory interferences are the only ones in a well-moderated nuclear reactor (heavy water moderated), while in most research reactors (light water moderated) the high-energy neutron flux will lead to additional interferences by the ^{47}Ti(n,p)^{47}Sc and ^{199}Hg(n,p)^{199}Au reactions. These interferences can be removed by additional irradiation inside a box made of B or Cd, which are thermal neutron absorbers, as explained in sections 5.6 and 5.7. However, this correction is not mandatory in most cases. The ^{199}Hg(n,p)^{199}Au reaction can be neglected as the cross-section is lower than for (n,γ) and the fast flux is smaller than the thermal flux and the Hg concentration is low. The ^{47}Ti(n,p)^{47}Sc reaction can be a greater problem owing to the higher concentration of the more abundant Ti. However, since Ti has several isotopes, the measurement of ^{48}Sc, which can be formed only by the ^{48}Ti(n,p)^{48}Sc reaction, can be used to calculate the contribution of the ^{47}Ti(n,p)^{47}Sc reaction.

5.6 Epithermal Instrumental Neutron Activation Analysis

In the usual INAA, the whole reactor neutron energy spectrum is used. However, in some cases, the use of part of the neutron spectrum is preferable; these systems are characterized by large differences in the activation cross-sections for the desired and the interfering nuclides in the various parts of the energy spectrum. The required trace elements are activated with part of the neutron spectrum while the interfering major elements are activated more strongly with the other parts of the spectrum, and thus we prefer to avoid this second part.

The neutron energy spectrum in a nuclear reactor is usually divided, for convenience, into three portions, and their relative abundances are dependent on the reactor structure. The most abundant fraction is the one of thermal neutrons, i.e. those neutrons which are in thermal equilibrium with the moderator atoms. Their most expected energy is equal to kT (where k is Boltzmann's constant and T is the neutron temperature), which at room temperature is equal to about 0.025 eV. The neutrons with energy above that of the thermal neutrons are divided into fast neutrons, those which are directly from fission and have not been moderated at all with energy mainly above 1 MeV, and epithermal neutrons, i.e. partly moderated and having energy between tenths of an electronvolt and 1 MeV. When the whole reactor neutron energy spectrum is used for activation, the main contribution is from the thermal neutrons, also owing to their usually higher cross-section [(n,γ) reactions].

A special case where advantage can be taken of the epithermal and fast neutrons is the case where the required trace elements are activated more strongly

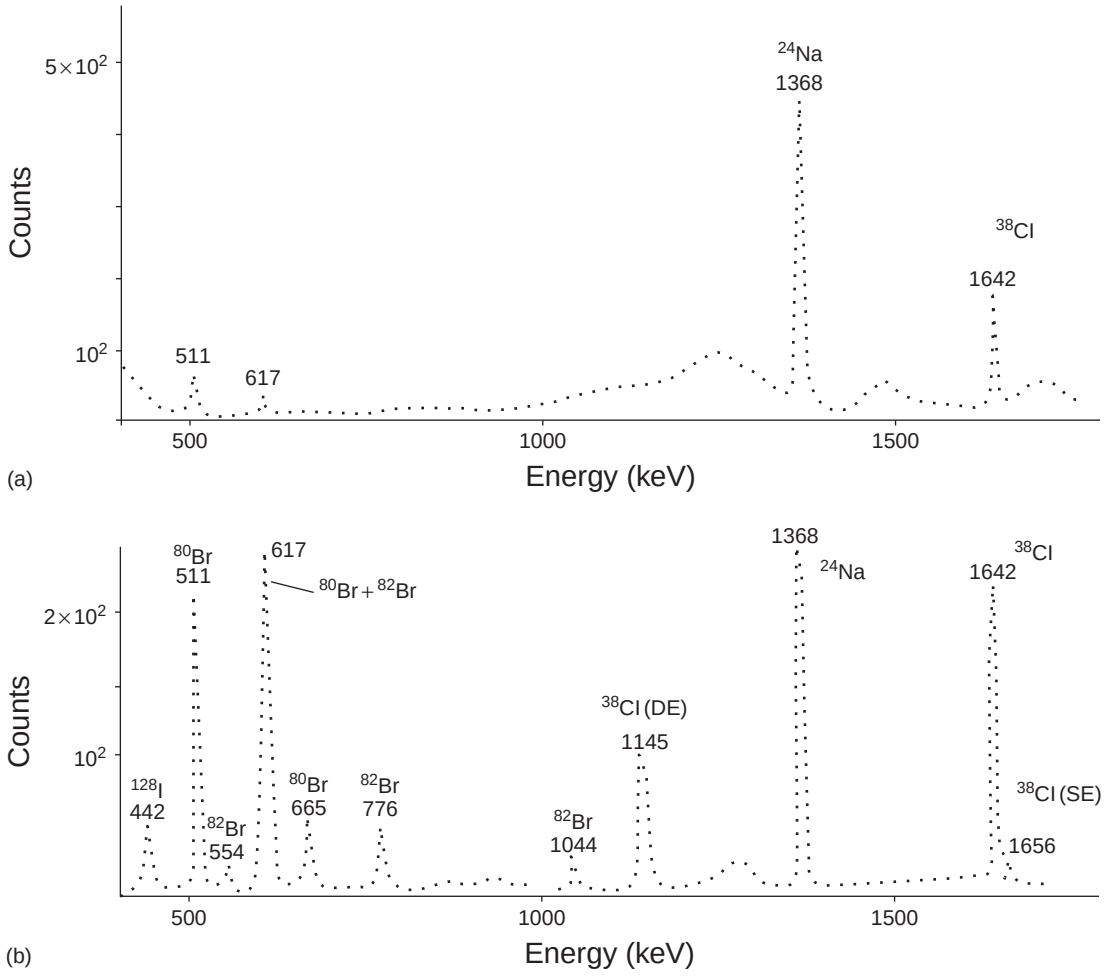

Figure 5 γ-Ray spectrum of a neutron-activated blood sample taken with a Ge(Li) detector. (a) Irradiated bare with reactor neutrons; (b) irradiated in a Cd shield.

relative to the major elements by the epithermal or fast neutrons. Most major elements in geology and biology follow the $1/v$ cross-section rule (their activation cross-section is inversely proportional to the square root of the neutron energy) throughout the whole energy spectrum. On the other hand, many of the less abundant elements have, in addition to their thermal activation, large activation cross-section resonances in the epithermal energy region and consequently can be activated preferentially in this region. Similarly, several of the less common elements can be activated by other neutron reactions besides the common (n,γ) reaction. These (n,p) and (n,α) reactions require higher energy than thermal reactions (and in most cases they do not occur also in the epithermal region and are induced only by fast neutrons).

A simple example of the advantage of using neutron filters (thermal neutron absorbers) in activation analysis is seen in Figure 5(a) and (b), which shows the γ-ray spectra of a sample of blood serum, activated with reactor neutrons once within a cadmium wrapping and once without any absorber (bare irradiation).[44] In the case of activation without a Cd absorber, the Compton-scattered photons of the major elements Na and Cl cover the peaks of bromine and iodine (bromine can be determined only after a delay of several days while the shorter-lived iodine cannot be determined instrumentally and can be determined only after chemical separation). The activation with epithermal neutrons (Cd cover) shows clearly the peaks of Br and I. A more complicated (geological) sample irradiated with and without a Cd cover is shown in Figure 6.

Owing to the presence of both thermal and epithermal neutrons, the product $\sigma\phi$ in Equation (28) should be replaced by the term $\phi_{th}\sigma_{th} + \phi_{epi}I_0$, where ϕ_{th} and σ_{th} are the flux and cross-section of the reaction with thermal neutrons and ϕ_{epi} and I_0 [the resonance integral of the

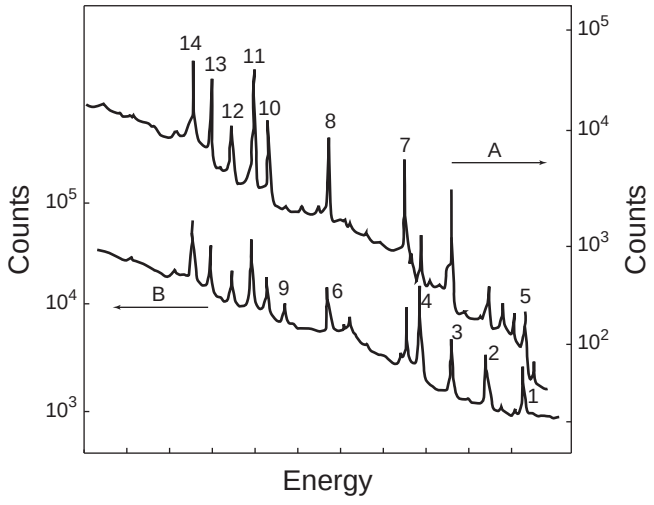

Figure 6 γ-Ray spectrum of neutron-activated air-dust after 10 d of cooling. (A) Inside a Cd sheet; (B) irradiated bare. (1) 1732 keV DE of ^{24}Na; (2) 1595 keV ^{140}La; (3) 1474 keV ^{82}Br; (4) 1369 keV ^{24}Na; (5) 1691 keV ^{124}Sb; (6) 1099 keV ^{59}Fe; (7) 1317 keV ^{82}Br; (8) 1044 keV ^{82}Br; (9) 889 keV ^{46}Sc; (10) 828 keV ^{82}Br; (11) 777 keV ^{82}Br; (12) 690 keV ^{82}Br; (13) 619 keV ^{82}Br; (14) 554 keV ^{82}Br.

(n,γ) reaction] are defined by Equations (51) and (52):

$$\phi_{epi} = \frac{\text{total epithermal flux}}{\int_{0.55\,eV}^{1\,MeV} dE/E} \tag{51}$$

$$I_0 = \int_{0.55\,eV}^{1\,MeV} \frac{\sigma(E)}{E}\,dE \tag{52}$$

The boundaries of 0.55 eV and 1 MeV were chosen since neutrons with energies above 1 MeV are categorized as fast neutrons rather than epithermal neutrons, and 0.55 eV as it is the cut-off in neutron absorption by Cd, the main filter used to capture thermal neutrons.

The epithermal activation properties of a nuclide can be expressed by the ratio I_0/σ_{th} or by means of the absorber ratio. The absorber ratio is the ratio of the activities of this nuclide irradiated once bare (without a cover) and once covered by an absorber of thermal neutrons. Thus the cadmium ratio is given by Equation (53):

$$R_{Cd} = \frac{\phi_{th}\sigma_{th} + \phi_{epi}I_0}{\phi_{th}I_0} = 1 + \frac{\phi_{th}\sigma_{th}}{\phi_{epi}I_0} \tag{53}$$

The most important fact is not only how the insertion of an absorber of the thermal neutrons influences the activity of the specific measured nuclide, but also how it influences the interfering nuclides. Hence it would be advantageous to analyze an element by epithermal neutron activation analysis (ENAA) rather than by using the whole spectrum of reactor neutrons for activation [reactor neutron activation analysis (RNAA)] only if its ratio of resonance integral to thermal neutron cross-section, I_0/σ_{th}, is larger than this ratio for the interfering elements. Several criteria have been suggested to measure the advantage of ENAA over RNAA, as follows.

5.6.1 Brune and Jirlow's Advantage Factor

Brune and Jirlow[45] suggested the use as an advantage factor for ENAA activation of the ratio between the cadmium ratio of the measured nuclide and the interfering nuclide [Equation 54]:

$$F_{BJ} = \frac{R_{Cd}^{i}}{R_{Cd}^{0}} \tag{54}$$

where R_{Cd} is the cadmium ratio as defined previously and the superscripts 0 and i stand for the measured and the interfering nuclides, respectively. This is the most commonly used advantage factor and several tables of this factor for many nuclides have appeared in the literature for cadmium absorbers and for boron absorbers.[11,45–48]

5.6.2 Parry's 'Improvement Factor'

Parry[49] pointed out that while the advantage factor describes well the increase in the signal-to-noise ratio, it does not consider the decrease in the activity of the analyzed element due to the elimination of the activation by the thermal neutrons and hence does not treat the larger error resulting from the lower counting statistics. Parry suggested that the true criterion should be the improvement in the detection sensitivity. The lower detection limit, L_D, for a radioactivity measurement, i.e. the minimum signal which can be detected above the background at the 95% confidence level is given by Equation (55):[50]

$$L_D = 2.33\sqrt{B} \tag{55}$$

where B is the background activity. The minimum detected mass in activation analysis (sensitivity) is given by Equation (56):

$$M_D = \frac{A}{L_D} = \frac{A}{2.33\sqrt{B}} \tag{56}$$

where A is the specific activity of the analyzed element under the experimental conditions for the activation and detection. Since A is proportional to the activity of the analyzed element and since B is due mainly to the interfering nuclide, Parry suggested that the improvement factor, f_p, is given by Equation (57):

$$f_p = \overline{\frac{R_{Cd}^{i}}{R_{Cd}^{0}}} \tag{57}$$

5.6.3 Bem and Ryan's Advantage Factor

Bem and Ryan[51] in 1981 followed the same trend of thinking as Parry; however, they suggested that the advantage factor should describe the improvement of the relative standard deviation of the counts. If the net counts (baseline corrected counts) is N and the background count is B, the advantage factor is given by Equation (58):

$$f_{BR} = \frac{\left(\frac{1}{N} + \frac{2B}{N^2}\right)^{1/2}_{0}}{\left(\frac{1}{N} + \frac{2B}{N^2}\right)^{1/2}_{i}} \tag{58}$$

where the subscripts i and 0 have the same meaning as the superscripts earlier.

When the background is mainly due to the interfering radionuclide, Bem and Ryan's advantage factor is equal to that of Parry. The correlation of Parry's advantage factor with that of Brune and Jirlow can be seen from Equation (59):

$$f_p = \frac{f_{BJ}}{\sqrt{R^i_{Cd}}} \tag{59}$$

Therefore, Parry's advantage factor is always smaller than that of Brune and Jirlow.

5.6.4 Tian and Ehmann's Generalized Advantage Factor

Tian and Ehmann[52] criticized Bem and Ryan's criterion (and consequently Parry's) on the grounds that in practice, in INAA, the limit on the number of counts is not due to the activity of the analyte, but rather to problems associated with the high counting rate. A high count rate leads to inferior resolution and also to problems of pile-up and dead time of the electronic system. In order to overcome these problems, the samples are measured far from the detector or are irradiated for shorter times. When the sample is activated with only epithermal neutrons, the total activity of the sample is reduced considerably, and hence the counting efficiency can be increased by using a smaller sample–detector distance, or the total counts can be increased by using larger samples or longer irradiation times. If the increase in counts (due to either count efficiency or size of the sample or length of irradiation) is given by G^2, then the generalized Tian and Ehmann's advantage factor is given by Equation (60):

$$f_{TE} = G\ \frac{\overline{R^i_{Cd}}}{R^0_{Cd}} \tag{60}$$

If $G = 1$, $f_{TE} = f_p$, and if $G = \sqrt{R^i_{Cd}}$, then $f_{TE} = f_{BJ}$.

The last case is the practical one, since both in the thermal activation and in the epithermal activation, the counting efficiency is usually chosen to obtain the maximum total counting rate allowable by the dead-time correction device. This generalized approach of Tian and Ehmann[52] gives a firmer basis for the widely used definition of Brune and Jirlow.

5.6.5 Thermal Neutron Absorbers

The main absorbers for the thermal neutrons are cadmium and boron owing to their high cross-section for reaction with thermal neutrons. The variation of the cross-section with the energy is different for Cd and for B, hence an intelligent choice of an absorber (also sometimes called a filter) will lead to optimized detection of some nuclides.[53] In some cases, one absorber is used and in others a combination of two absorbers, e.g. Cd + B.[49,53,54] In some experiments, filters which absorb the epithermal neutrons in some regions are used, allowing more selectivity for some elements.[55]

Since the main absorbers are boron and cadmium, it is very important to compare them from the technical point of view. The absorber can be used as a covering sheet, wrapping the sample in it, mainly in the case of cadmium for which metallic sheets are commercially available, or as a capsule built from these materials, or as a mixture with the sample, used with B_2O_3,[56] or as a permanent lined installation inside one of the irradiation ports of the reactor. The use of absorber-lined ports has the disadvantages that (1) scattered thermal neutrons can come from angles which are not covered by the lining of the absorber, leading to lower absorber ratios, whereas when the absorber is used as a capsule or wrapping, it is covered from all angles, and (2) the capsule in which the sample is held during the pneumatic transfer and which usually is made of polyethylene leads to partial thermalization of the epithermal neutrons. On the other hand, the use of an absorber-built capsule or wrapping suffers from the disadvantages of the absorption reactions. Cadmium is activated to short- and long-lived nuclides and the unloading and unpacking of the sample for medium- and long-lived radionuclide measurement faces radiation safety problems owing to the high radiation dose. Short-lived ($t_{1/2} < 20$–30 s) radionuclides cannot be measured at all in a cadmium capsule since the short half-life prohibits the safe unpacking of the vessel and the activity of the absorber is too high to allow measurement together with the filter.

While the absorption of neutrons by ^{10}B does not lead to radioactive products, the reaction $^{10}B(n,\alpha)^7Li$ is very exoergic ($Q = 2.792$ MeV) and the samples are heated considerably. Stroube et al.[57] found that in the 20-MW reactor at the former US National Bureau of

Standards (NBS), thermal heating of the boron nitride vessel limited the length of irradiation for freeze-dried foods to 4 s and prevented completely safe irradiation of wet food. When biological samples are irradiated, this heating accelerates the thermal decomposition of organic compounds, producing high pressure in the sample container, when airtight sealed, which may explode and contaminate or even ruin the irradiation port. In other cases, elements may be volatilized and lost. In order to avoid these effects, the time of irradiation should be limited. Williamson et al.[54] measured the temperature inside a polyethylene rabbit inserted into a Cd-lined irradiation port and found that the temperature reached an equilibrium value of 90 °C in about 7 min. When a BN capsule was irradiated in the same position, the measured temperature was 120 °C in about 3 min and continued to rise. Ehmann et al.[58] irradiated rock samples in a boron carbide filter for 20 h, keeping the sample in heat-sealed quartz ampules. Quartz, being a poor thermal conductor, keeps the sample from being highly heated; this solution is good for geological samples but will probably not suffice for biological samples. Chisela et al.[59] studied the temperature in a sintered BC capsule in an air-cooled irradiation facility and found the capsule to reach steady-state temperatures of 163 and 194 °C for 4.0- and 5.0-MW reactors, respectively. When a permanent installation from powdered B_4C was made with water coolant, the temperature reached not more than 50 °C.

The use of a permanent lining of absorber also has the disadvantage of reducing the total flux of the neutrons in the reactor and of excessive use of the nuclear fuel. A possible advantage of boron over cadmium is the reuse of the same filter in subsequent irradiation. Cadmium filters usually cannot be reused, at least immediately, owing to the long-lived radioactivity produced in cadmium during irradiation. The activity produced in boron filters is small and is due only to contamination in the boron. However, the use of boron filters is limited in many cases to not high total doses of neutrons owing to structural failure of the capsule, probably owing to excessive formation of helium gas from the $^{10}B(n,\alpha)^7Li$ reaction. Cadmium filters are easily made from metallic cadmium sheets of about 1 mm thickness. Boron is a difficult material to machine and Stuart and Ryan[60] prepared boron shields by forming a mixture of boron carbide powder and paraffin wax. The mixture was heated to 70 °C (paraffin melting point = 56 °C) and cast into cylindrical forms. A central hole was made in the form as it solidified by using a heated rod of appropriate diameter. The hole was not made through the whole length of the cylinder, in order to obtain a cylindrical capsule with a central cavity closed at one end. The other end was closed with a top made from the same material. Parry[47] used the same method, but instead of using boron carbide, she used B powder. It should be mentioned that the paraffin causes a small thermalization of the epithermal neutrons. The best machinable refractory boron compound is BN,[61] and consequently many of the studies with boron filters were made using BN capsules. Ehmann et al.[58] suggested that BN should not be used owing to the relatively high cross-section of 1.81 b for the $^{14}N(n,p)^{14}C$ reaction which leads to the formation of an appreciable amount of the long-lived radioactive ^{14}C, but rather boron carbide, another refractory compound of boron, should be used. Unfortunately, boron carbide is extremely hard and is not machinable, and a capsule can be made only by the hot pressing process.

One of the disadvantages of boron filters is the impurities found in boron powder, as discussed in detail by Bem and Ryan.[51] However, if a boron capsule is used together with a permanently installed Cd lining, the interferences due to the activities of ^{28}Al, ^{56}Mn and ^{38}Cl from the boron contaminants are significantly reduced.[47] Both cadmium and boron have high absorption cross-sections for low-energy neutrons; however, the energy dependences of the cross-sections differ considerably. Cadmium approaches a perfect sharp filter for the thermal region and has some resonance in the epithermal range, whereas boron behaves as almost a perfect $1/v$ absorber with no sharp energy cut-off. Although the cross-section for neutron capture by boron is lower than the cross-section for absorption by cadmium in the lower energy range of 0.01–1 eV, it can be compensated for by using thicker boron absorbers. A 0.25 cm thick boron shield is sufficient to stop practically all the thermal neutrons. The effective cut-off energy is almost independent of the thickness of the Cd absorber whereas it increases considerably with increasing thickness of the boron absorber.

5.7 Fast Neutron Instrumental Neutron Activation Analysis

The reactions that occur with fast neutrons with energy usually in the megaelectronvolt range should be looked upon in two ways: (1) the use of these reactions for the determination of some elements and (2) the possible interference of these reactions in the determination of some elements by (n,γ) reactions, due to the formation of the same nuclide as was shown earlier dealing with nuclear interferences. These interferences can be solved only by the use of double irradiation, one with a bare core and one inside a Cd or B filter, and calculating the contribution of each element. The same treatment is usually applied with the use of (n,p) and (n,α) reactions in the determination of some elements. The main advantage of these reactions is that they produce nuclides different from those produced by (n,γ) reactions. Consequently,

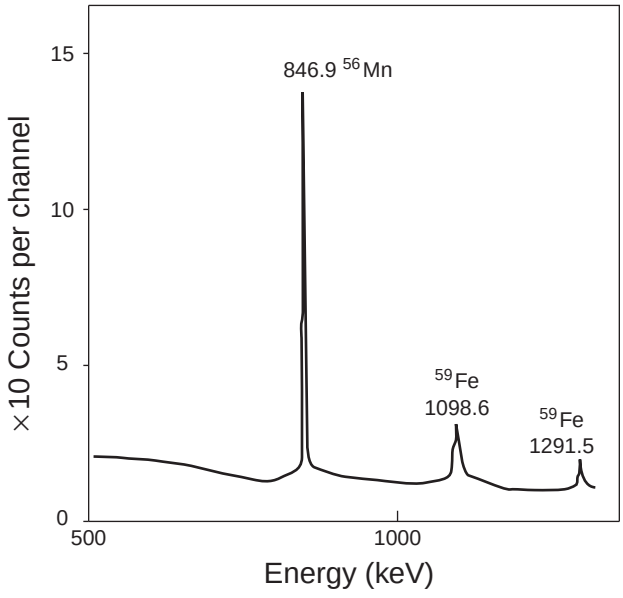

Figure 7 γ-Ray spectrum of iron irradiated with reactor neutrons (irradiation for 10 min, decay for 5 min and count for 6 min).

they may lead to a faster determination in the case of producing a short-lived nuclide rather than the long-lived nuclide produced in (n,γ) reactions. In other cases, they may allow the determination of elements which cannot be measured via (n,γ) reactions since the radionuclide produced is only a β emitter.

5.7.1 Rapid Determination of Iron

Figure 7 shows the γ-ray spectrum of an iron sample irradiated for a short time.[62] As can be seen, the peak of 847-keV γ-rays of ^{56}Mn due to the ^{56}Fe(n,p)^{56}Mn reaction is considerably higher than the ^{58}Fe(n,γ)^{59}Fe 1099-keV peak. Hence the use of (n,p) reactions for the determination of iron has a higher sensitivity in the case of short irradiation and counting. However, ^{56}Mn is formed also from manganese by the ^{55}Mn(n,γ)^{56}Mn reaction. The concentrations of both iron and manganese can be found by double irradiation, one sample with reactor neutrons (without any filter) and one sample with epithermal neutrons (with a cadmium absorber). If the specific activity (measured counts under the experimental set-up per gram of the element) for irradiation with reactor neutrons will be F_R and M_R for iron and manganese, respectively, and similarly for epithermal neutrons F_E and M_E, then the activity of 1 g of sample containing P_F% of iron and P_M% of manganese will be given by Equations (61) and (62):

$$C_R = \frac{F_R P_F + M_R P_M}{100} \tag{61}$$

$$C_E = \frac{F_E P_F + M_E P_M}{100} \tag{62}$$

where C_R and C_E are the activities induced by reactor neutrons and epithermal neutrons, respectively. The solution of these two equation gives Equations (63) and (64):

$$P_F = \frac{M_E C_R - M_R C_E}{\Delta} \tag{63}$$

$$P_M = \frac{F_R C_E - F_E C_R}{\Delta} \tag{64}$$

where $\Delta = 100(F_R M_E - F_E M_R)$.

5.7.2 Determination of Phosphorus and Silicon

Thermal neutron activation cannot be used for the determination of phosphorus and silicon. A radiative capture (n,γ) reaction with the only stable isotope of P leads to the formation of ^{32}P, which is a pure β emitter. In the case of silicon [stable isotopes ^{28}Si (92.2%), ^{29}Si (4.7%) and ^{30}Si (3.1%)], the only radionuclide produced by the (n,γ) reaction is ^{31}Si, which is almost only a β emitter. Its very low intensity of γ-rays (1266 keV, 0.07%) together with the low abundance of ^{30}Si and the low cross-section for radiative capture (0.11 b) permit only the determination of relatively large amounts of silicon. However, activation with epithermal neutrons leads also to the formation of ^{28}Al via both ^{31}P(n,α)^{28}Al and ^{28}Si(n,p)^{28}Al reactions and of ^{29}Al by the ^{29}Si(n,p)^{29}Al reaction. ^{28}Al is also produced by the ^{27}Al(n,γ)^{28}Al reaction. This leads to a procedure for the determination of Si from the activity of ^{29}Al, which is produced only from silicon, using the activities of ^{28}Al from activation with a Cd filter and without a filter to determine the concentration of both aluminum and phosphorus.

However, the activity of ^{29}Al produced for Si is almost two orders of magnitude less than the activity of ^{28}Al produced from it. Hence the use of ^{29}Al will both limit the minimum amount of silicon that can be determined and will reduce the accuracy of the measurement. Another problem associated with the measurement of ^{29}Al is that its main γ line is the 1273-keV line, which suffers from the interference of the SE peak of the more abundant ^{28}Al at 1268 keV.

However, when the concentration of Si is high, silicon can be determined by the ^{29}Si(n,p)^{29}Al reaction, as was done by Hancock[63] for the measurement of silicon in pottery using a Cd shield to decrease the formation of ^{28}Al. The samples were allowed to decay for 17–20 min before counting, to decrease the ^{28}Al activity further. Ördögh et al.[64] measured in that way the concentration of silicon in very small inhomogeneous lymph-node samples. The concentration

of P was determined spectrophotometrically by the molybdenum blue method and hence the concentration of Si and Al can be found from the ^{28}Al activity induced by both epithermal activation (Cd cover) and reactor neutron irradiation. Another way was suggested by Alfassi and Lavi,[65] who used the simultaneous determination of ^{27}Mg and ^{28}Al, each of them for both reactor activation and irradiation with only epithermal neutrons, to measure simultaneously Mg, Al, Si and P. Each of these radionuclides can be formed by three reactions (Scheme 5).

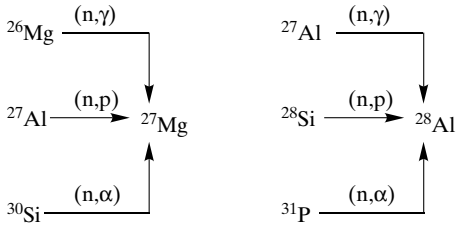

Scheme 5

The concentration of the four elements Mg, Al, Si and P can be obtained from a solution of the four equations for the four measured activities (844 keV due to ^{27}Mg and 1778 keV due to ^{28}Al, each without a filter and with a cadmium absorber). The specific activity (measured counts per gram of the element under the experimental set-up) will be assigned by three letters, the first giving the target element, the second the element formed and the third (R or E) indicating whether the activation has been done by reactor neutrons (without an absorber) or by epithermal neutrons (with a cadmium absorber). Thus, for example, SAR means the specific activity of ^{28}Al produced from silicon by reactor neutrons. The activities measured per gram of sample are also represented by three letters, the first one is always C, and the second and the third have the same meanings as for specific activities. If the concentrations of Mg, Al, Si and P are given by f_{Mg}, f_{Al}, f_{Si} and f_P in weight fractions, then the four appropriate equations are represented by Equations (65–68):

$$CMR = MMRf_{Mg} + AMRf_{Al} + SMRf_{Si} \quad (65)$$

$$CME = MMEf_{Mg} + AMEf_{Al} + SMEf_{Si} \quad (66)$$

$$CAR = AARf_{Al} + SARf_{Si} + PARf_P \quad (67)$$

$$CAE = AAEf_{Al} + SAEf_{Si} + PAEf_P \quad (68)$$

5.8 Instrumental Neutron Activation Analysis with Emissions Other than γ Emission

As written earlier, in most delayed activation analysis only the γ emission is measured, since it is difficult to distinguish between β^- emitters, owing to their extended spectra rather than monoenergetic emission, X-ray emission is less penetrating and neutron emission is very rare. However, all these emissions can be used in special cases. For β^-, it was shown that both for biological samples[66] and most geological samples,[67] owing to the higher concentration of P the main relatively long-lived β^- emitter is ^{32}P, and after 10 days more than 95% of the β^- emission is due to ^{32}P, and hence the total β^- measurement can be used for the determination of P.

5.8.1 X-Rays[68]

X-rays suffer from two disadvantages, which can be turned to advantage in special cases. First, the range of X-rays is fairly short and only those emitted from the first few micrometers of the sample can be detected. This is disadvantageous when the whole bulk concentration is needed. However, it is advantageous when only the first few micrometers are important and when in them the concentrations of the measured elements are higher. Second, fewer nuclides emit X-rays (mainly after EC or internal conversion) than those emitting γ-rays. This is disadvantageous since fewer elements can be determined in this method. However, if the major elements of the sample, which interfere with the measurement of the trace elements, do not emit X-rays, the few elements which emit X-rays can be determined even at low concentration owing to the lower background.

Owing to the low penetration of X-rays, the samples used for X-ray measurements should be very thin. The X-rays can be due to internal conversion (X-rays of the Z of the IRN), EC (X-rays of $Z-1$ element) or internal conversion after a β decay ($Z+1$ element). In γ-ray spectrometry the electrons emitted by the active sample are stopped by a thin Perspex or aluminum foil as the electrons emitted by the active sample are less penetrating than γ-rays. In the case of low-energy X-rays, they are less penetrating than some of the electrons and the electrons should be removed by a magnetic field.[69–71] X-ray measurements have been used mainly for the determination of rare earths in geological samples,[72–74] copper[75] and bromine[76] in biological and geological samples, but also other elements, such as Nb in steel[77] and Co and Hg in Dead Sea water,[78] have been determined.

5.8.2 Neutrons[79]

Delayed neutrons are emitted from only very few radionuclides. In many cases the radionuclide decays by both β^- and neutron emission. For example, ^{87}Br decays 2.8% by neutron emission while ^{88}Br decays 6.6% by neutron emission and ^{88}Se only 0.8%. ^{90}Br decays 25% by neutron emission. At least 57 radionuclides which emit neutrons are known, but none of them is

formed by an (n,γ) reaction since the nuclides should be more than one neutron further from the line of stability in order to emit neutrons. However, those radionuclides can be formed in a neutron-induced fission of fissile nuclides such as ^{235}U, since the heavy elements have a higher percentage of neutrons in the nucleus. This is the reason why activation analysis followed by measurement of neutrons is used for the determination of only U and Th. Since other elements do not produce neutron emitters, the background from other elements is very small. The neutrons are measured (after their moderation in a paraffin layer) by BF_3 counters. These are gaseous ionization chambers which measure the α-particles formed by the ^{10}B(n,α)^{7}Li reaction. Outside α-particles are not able to penetrate the counter walls. At least 57 delayed neutron emitters are formed in the neutron-induced fission with half-lives ranging from 0.08 to 55.6 s. In order to get rid of the ^{17}N (4.16 s) and to reduce the error due to wrong time determination, which is more prominent for short-lived radionuclides, the determination of the emitted neutrons is done only for those which have half-lives longer than about 10 s, by delaying the counting for 30–60 s.[80,81]

5.9 Depth Profiling by Instrumental Neutron Activation Analysis

Depth profiling involves the measurement of the concentration of a trace element as a function of its distance from the surface of the sample. This can be done by a destructive method in which the whole sample is activated and then divided into a number of sections of varying distances from the surface, either by gradual slow dissolution or by gradual milling and grinding (for each layer, its thickness and γ or β activity of the various elements are measured). Many samples are too important to be destroyed in the measurement. Using (n,α) or (n,p) reactions, the depth profiling can be done nondestructively. The emitted charged particle (α or p) has a definite kinetic energy. This particle, while transversing in the material, from its place of formation inside the sample to the detector positioned in front of the sample, loses part of its energy, depending on the distance it traveled in the sample. From the measured energy, the distance from the surface can be calculated. Thus, the energy spectrum measured for this particle is related to the concentration depth profile of the element, which is responsible for this nuclear reaction.

In order to allow the measurement of low concentrations, the nuclear reaction should have a high cross-section of the order of at least 10 b. These high cross-sections rule out most elements and only very few elements can be determined.

The method was developed in 1972 by Ziegler et al.[82] for the determination of boron due to the high cross-section of the ^{10}B(n,α)^{7}Li ($\sigma = 3837$ b). This cross-section is five orders of magnitude larger than that of almost any other stable nuclide for (n,α) reactions (except ^{6}Li). The measured sample is placed in a vacuum chamber (pressure $< 5 \times 10^{-6}$ Torr). They needed only about 10^{11} neutrons to detect a concentration of 100 ppm of boron in a 2-cm Si wafer. The ^{7}Li is produced in both the ground state (6%) and in the 479-keV excited state (94%). The energy is distributed between the α-particle and the ^{7}Li according to their masses (1471 and 839 keV for ^{4}He and ^{7}Li in the reaction leading to the excited states). The bombarding thermal neutron passes through the Si wafer (170 µm thick) with negligible attenuation, so that the reaction efficiency is independent of the depth. The neutron beam was collimated to produce a spot 1 cm radius at the target chamber (neutron flux 2.3×10^8 n cm^{-2} s^{-1}). The targets were mounted at a 45° angle to the neutron beam. The α-particles were detected by a surface barrier detector subtending an angle of 19 msr and was specially fabricated to have a low γ and electron background and to be able to measure the spectrum of the α-particles in the presence of ^{7}Li, although this led to poorer energy resolution for the α-particles [full width at half-maximum (fwhm) $\approx$ 20 keV instead of 12–14 keV]. The depth of the boron is determined by the energy loss of the 1471-keV α-particles using the known specific energy loss, dE/dx. An energy, E_f, parallel to depth is given by Equation (69):

$$\text{depth} = \int_{1471}^{E_f} \left(\frac{dE}{dx}\right)^{-1} dE \qquad (69)$$

Ziegler et al.[82] measured the energy spectra of 10^{15}–10^{16} ^{10}B atoms per square centimeter and calculated a concentration resolution of 3 ppm and a depth resolution of 20 nm. Müller et al.[83] used the high-flux reactor of ILL, Grenoble, France, with a collimated flux of 1.04×10^9 n cm^{-2} s^{-1} and 1.5 cm diameter, a detector with resolution of better than 16 keV for the 1471-keV peak. For a counting time of 1000 min, 10^{11} ^{10}B atoms per square centimeter will lead to about 110 integral counts. This is enough for measuring total boron concentration but not for profile determination and they estimated that for profile determination the minimum required concentration is 10^{12} ^{10}B atoms per square centimeter. They measured the concentration depth profiles of boron in silicon wafers implanted with 10^{12}, 10^{13} and 10^{14} ^{10}B atoms per square centimeter and the accuracy of the depth determination was about 25 nm. It should be remembered that this is for ^{10}B atoms whereas they constitute only 20% of the natural boron atoms. Kvitek et al.[84] studied the boron concentration profile in Si using the spectra of both

α-particles and ^{7}Li particles. As can be expected from the higher dE/dx, the spatial resolution obtained with the 839-keV ^{7}Li particle is better than that obtained by using the 1431-keV ^{4}He line (by about 30%). However, in a later work, they explained[85] that in spite of the better resolution of the ^{7}Li profiling, the fact that the ^{7}Li line is located in the high-background region of the measured spectra often makes the profile analysis by this method difficult. In order to obtain better resolution they suggested measuring the α-spectra with different tilting angles.

Biersack et al.[86] discussed the achievable depth resolution and the main sources of uncertainties. They pointed out four main inherent uncertainties: (1) the uncertainty in the angle θ between the point of reaction and point of detection as the particle travel distance L, $L = x/\cos\theta$, where x is the depth of the target atom; (2) the uncertainty in the energy resolution of the detector; (3) the uncertainty in energy loss due to energy loss straggling, assuming a fixed pathlength, L; and (4) uncertainty due to multiple angular scattering. They concluded that for (n,α) reactions, a depth resolution of 5 nm is feasible, and 10-nm depth resolution is feasible for (n,p) reactions.

Apart from B, the only element which can be determined in this way using the natural abundance is lithium with 7.4% ^{7}Li, having a cross-section of 940 b for the reaction ^{6}Li(n,α) T, with an α energy of 2055 keV and a T energy of 2728 keV. Helium can be determined not for natural He but for ^{3}He-doped substrates (stable isotope, natural abundance 0.0001%). Similarly, Na and Be can be determined with the radioactive isotopes ^{22}Na and ^{7}Be, respectively, owing to their high cross-section.

Downing et al.[87] gave a detailed description of the neutron depth profiling (NDP) system and summarized the available data in various systems throughout the world. Special attention was paid to minimizing the contamination of the thermal neutron beams with fast neutrons and γ-rays. The amount of the γ-rays and fast neutrons was reduced by filtering the beam through 200 mm of sapphire single crystal and an 80-mm long silicon single crystal, which led to a ratio of cadmium to gold of $>10^{4}$ and a γ intensity of 200 mrad h^{-1}.

The NDP technique is quantitative and has few interferences in silicon-based materials. As a result, it is often used for calibration for other techniques. Other instrumental methods, such as secondary ion mass spectrometry (SIMS), Fourier transform infrared (FTIR) spectrometry, and Auger electron spectroscopy (AES) have greater sensitivity than NDP in most applications and better resolution in some matrices, but each of the other techniques is subject to highly interactive chemically and electronically induced artifacts. NDP can be used to calibrate the other methods and, after calibration, the better resolution of the other method can be used. Rutherford backscattering spectroscopy (RBS) and NDP complement each other as they both nondestructively analyze different compositions: RBS measures the concentration profile of heavy atoms in matrices of light atoms whereas NDP is limited to some light elements.

5.10 Practical Cases of Instrumental Neutron Activation Analysis

In section 5.4 a test case of INAA of Si was given. Si is one of the best matrices for INAA, owing to the low radioactivity induced in it by neutron irradiation, and consequently very low detection limits were obtained. This is not the case for all matrices. The following section gives details of several examples of analyses of various matrices by INAA.

Greenberg et al.[88] described the procedures adopted by the NAA group of the National Institute of Standards and Technology (NIST) in the USA. The samples are first packaged in a cleaned container. The container, depending on the type of the sample and the length of irradiation, can be polyethylene bags or vials, sealed by heating, or flame-sealed high-quality quartz (to eliminate as much as possible of the activatable sodium) or aluminum cans. The samples are irradiated in a well-moderated (by D_2O) neutron flux of 10^{14} n cm^{-2} s^{-1} using a pneumatic tube for transferring the sample to the reactor. The length of the irradiation, decay and counting depends on the analyzed elements. For one or a few elements the timings are optimized for these elements. Thus, for the determination of mercury in urine the samples (in quartz ampules) are irradiated for 1 h and counted after decay both of 5 and 20 d in order to measure the Hg content via the formation of both ^{197}Hg ($t_{1/2} = 64.1$ h, $E_\gamma = 77$ and 191 keV) and ^{203}Hg ($t_{1/2} = 46.59$ d, $E_\gamma = 279$ keV). The long decay is used to reduce the activities of Na, Cl and Br.

To measure Cl and Br in various oils, they used short irradiation of a few minutes and measured the activity of ^{38}Cl ($t_{1/2} = 37.2$ min, $E_\gamma = 1642.2$ and 2167.6 keV) after a decay of several minutes. The concentration of Br was determined via ^{82}Br ($t_{1/2} = 35.3$ h, $E_\gamma = 554.3$ and 776.5 keV) after a decay of 24 h. For lower detection limits of bromine, irradiation from several hours up to 24 h was applied.

For a complete multielement analysis (20–40 elements which can be determined by NAA), the elements are divided into three groups according to the half-lives of their IRNs: 2 min –15 h, 0.5–5 d and >5 d.

Some elements appear in more than one group. The usual scheme of irradiations, decays and counting is followed. The same sample can be used for the two irradiations.

Table 2 Elemental concentrations (ppm, unless % is indicated) determined by INAA vs NBS (NIST) certified values

Element	SRM 1648 certified		SRM 1572 certified		Method
	INAA	value	INAA	value	
Ag	6.5 ± 0.5				C
Al(%)	3.5 ± 0.1	3.42 ± 0.11			A
As	119 ± 5	115 ± 10	31 ± 0.3	3.1 ± 0.3	B
Ba	750 ± 60		21.4 ± 1.2		A
Br	515 ± 30		8.2 ± 0.3		A, B
Ca(%)	5.9 ± 0.5		3.19 ± 0.10	3.15 ± 0.10	A
Cd	72 ± 6	75 ± 7			B
Ce	54 ± 3		0.28 ± 0.08		C
Cl(%)	0.46 ± 0.02				A
Co	17.9 ± 0.5		0.02 ± 0.006		C
Cr	410 ± 10	403 ± 12	0.74 ± 0.18	0.8 ± 0.2	C
Cs	3.5 ± 0.2		0.098 ± 0.010		C
Eu	0.81 ± 0.08		0.010 ± 0.003		C
Fe	$39\,200 \pm 800$	$39\,100 \pm 1000$	92 ± 13	90 ± 10	C
Hf	4.3 ± 0.3				C
Hg			0.079 ± 0.016	0.08 ± 0.02	A
I	20 ± 5				A
In	1.00 ± 0.07				A
K(%)	1.00 ± 0.06	1.05 ± 0.01	1.76 ± 0.03	1.82 ± 0.06	A, B
La	43 ± 2		0.19 ± 0.03		B
Mg(%)	0.85 ± 0.08				A
Mn	800 ± 20				A
Na(%)	0.41 ± 0.02	0.425 ± 0.002			B
Rb	53 ± 9		5.0 ± 0.3	4.84 ± 0.06	C
Sb	46 ± 2		0.041 ± 0.010		B, C
Sc	6.3 ± 0.2		0.011 ± 0.001		C
Se	27 ± 2	27 ± 1			C
Sm	4.2 ± 0.4		0.052 ± 0.006		B
Sr			94 ± 3	100 ± 2	C
Th	7.5 ± 0.3		0.015 ± 0.004		C
Ti(%)	0.41 ± 0.02				A
V	133 ± 7	140 ± 3			A
W	4.9 ± 0.6				B
Zn	4800 ± 200	4760 ± 140	28.3 ± 0.8	29 ± 2	C

The accuracy and precision of their determination can be found in Table 2, which gives their results for two NBS SRMs together with the certified values. The uncertainties listed represent the estimated overall analytical uncertainties at the 95% confidence level. SRM 1648 is Urban Dust and SRM 1572 is Citrus Leaves. In the column for the method, A, B and C refer to the previous scheme.

Cunningham and Stroube[89] applied INAA to 240 food composites. They irradiated with a neutron flux of 4.9×10^{13} n cm^{-2} s^{-1}. Two irradiations were done on each sample, one for 15 s and one for 4–5 h. For the short irradiation, a 2-min decay time and a 10-min counting time were used. For long irradiation, the decay times varied from 5 d to 4 weeks and counting times were 6–8 h. In the short irradiation they determined Ca, Cl, Cu, Ga, I, In, K, Mg, Mn, Na, S, Sr, Tl and V and in the long irradiation Ag, As, Au, Ba, Br, Cd, Ce, Co, Cr, Cs, Eu, Fe, Hf, La, Lu, Mo, Rb, Sb, Sc, Se, Sm, Ta, Th, W, Yb and Zn.

Sun and Jervis[90] used a Slowpoke reactor, which is a nuclear reactor specially designed mainly for teaching, with activation analysis and the production of small amounts of radioisotopes in mind. The reactor operates at a low power of 20 kW and has a maximum flux of 10^{12} n cm^{-2} s^{-1} (Sun and Jervis used a flux of 2.5×10^{11} n cm^{-2} s^{-1}). An advantage of this type of reactor is its high stability (1–2% over long periods), which allows the irradiation of standards only every few days and not necessarily with each sample. Sun and Jervis[91] determined 35 elements in rocks and soils using two irradiations and three countings. Al, Ba, Ca, Cl, Dy, I, Mg, Mn, Sr, Ti, U and V were determined by 1-min irradiation followed by 5-min counting after a decay of 10–15 min. Long irradiations of 16 h were followed by 30-min counting after 4–9-d decay (As, Br, K, La, Na,

Sb, Sm, W and Yb) and 100-min counting after a decay of 15–20 d (Ce, Co, Cr, Cs, En, Fe, Hf, Ln, Nd, Rb, Sb, Sc, Se, Sn, Ta, Tb, Th, Yb and Zn). Large relative standard deviations between several samples were found for Cl (32%), Sb (24%), Ti (18%), Ba (16%) and Br (14%), but for most elements the average relative standard deviation was <10%. A similar scheme was used by the same authors for samples of coal and ashes, except that the short irradiation time for coal samples was increased to 5 min.

Iskander[92] measured the concentrations of 28 elements in tobacco and cigarette paper. Samples were irradiated for two irradiation periods, the first for 3 min and the second for 8 h with a neutron flux of 2×10^{12} n cm^{-2} s^{-1} from a TRIGA Mark I reactor. The sample was irradiated for 3 min and was counted for 3 min after a 1-min decay (to measure Al, V, Ti, Ca and Mg) and after an additional decay of 26 min (total decay time 30 min) was counted again for 1000 s (to determine Cl and Mn). Also, the long-irradiated samples were counted twice: 4000 s after a decay of 12 h (to determine K, Na, As, Br and La) and 40 000 s after a decay of 21 d (Ba, Rb, Th, Cr, Ce, Hf, Fe, Sb, Sr, Ni, Sc, Se, Zn, Cs, Co and Eu).

McOrist et al.[93,94] studied trace element concentrations in various Australian opals. They irradiated their samples twice, for 1 min at a thermal flux of 2.5×10^{13} n cm^{-2} s^{-1} and for 18 h at a thermal flux of 5×10^{12} n cm^{-2} s^{-1}. Each irradiated sample was counted twice. The short-irradiated samples were counted for 10 min after a 20-min decay and for 45 min after a decay of 24 h. After the long irradiation, the samples were counted for 2 and 4 h after 7 and 28 d of decay, respectively.

Wallner and Katzlberger[95] studied the concentrations of 14 elements in spruce tree rings and needles from an area where an Al refinery used to be. Al, Mn, Ca and Cl were determined by 3–7 min of irradiation at a thermal neutron flux of 5×10^{10} n cm^{-2} s^{-1}. They calculated that for Al, assuming that the main interference for its 1779-keV line is from the Compton-scattered photons of the

Table 3 INAA timings for various elements in biological samples

Element	IRN	γ-ray of IRN	Half-life (keV) of IRN	Timings system	Interfering reactions
Na	24Na	15.02 h	1369	2, 3	24Mg(n,p), 27Al(n,a)
Mg	27Mg	9.46 m	1014	1	27Al(n,p)
Al	28Al	2.24 m	1779	1	28Si(n,p), 31P(n,a)
Si	31Si	2.62 h	1266	2	31P(n,p), 34S(n,a)
S	37S	5.0 min	3103	1	
Cl	38Cl	37.3 min	1642	1, 2	
K	42K	12.36 h	1525	2	42Ca(n,p)
Ca	49Ca	8.72 m	3084	1	
Sc	46Sc	83.8 d	889	4	46Ti(n,p)
V	52V	3.76 m	1434	1	
Cr	51Cr	27.70 d	320	4	54Fe(n,a)
Mn	56Mn	2.58 h	847	2	56Fe(n,p)
Fe	59Fe	44.6 d	1099	4	
Co	60Co	5.27 y	1173	4	63Cu(n,a)
Ni	58Co	70.8 d	811	4	
Cu	66Cu	5.10 m	1039	1	66Zn(n,p)
Zn	65Zn	244.1 d	1116	4	
As	76As	26.3 h	559	3	79Br(n,a)
Se	75Se	118.5 d	265	4	
Br	82Br	35.34 h	776	3	
Rb	86Rb	18.8 d	1077	4	
Sr	87mSr	2.80 h	388	2	
Mo	99Mo-99mTc	66.02 h	141	3	
Cd	115Cd-115In	53.4 h	336	3	
Sb	122Sb	2.68 d	564	3	
Sb	124Sb	60.20 d	1691	4	
I	128I	24.99 m	443	2	
Cs	134Cs	2.06 y	796	4	
Ba	139Ba	82.9 min	166	2	
Hg	203Hg	46.8 d	279	4	

The timing systems are:	system:	1	2	3	4
	Time of irradiation:	1–10 min	2–20 min	10 h	10 h
	Time of counting:	1–2 min	1 h	5–8 d	30 d
	Time of decay:	2–5 min	10 min	1–2 h	3–5 h

Table 4 Detection limits found for INAA of various biological samples

Biological sample	Irradiation time[a]	Decay time	Counting time	Detection limits (ppb)
1. Human blood[b]	5 d	7 mon	2 h	Co(4), Cs(4), Fe(10,000), Se(20)
	24 h, E	21 d	2 h	Br(1000), Fe(1500), Rb(300), Se(100), Zn(500)
2. Hair[c]	30 s	10 s	20 s	Ag(140), Cl(13,000), F(23,000), Se(160)
	10 m	5 m	5 m	Al(1600), Ba(7600), Ca(52,000), Cu(3,800), I(260) K(16,000), Mn(750), Na(4,100), S(0.28%), Zn(3,500)
	16 h	2 d	50 m	As(46), Au(2), Sb(45)
	16 h	21 d	50 m	Co(97), Hg(200), Ag(140), Ba(7,600), Se(160), Zn(3,500)
3. Meat, fish and poultry[d]	15 s	2 m	10 m	Ca(82,000), Cl(2,400), Cu(8,300), K(0.1%), Mg(0.013%), Mn(1,900), Na(260), V(81)
	4 h	5–28d	7 h	Ag(120), As(650), Br(130), Cd(170), Co(23) Cr(350), Cs(39), Eu(3.7), Fe(6,800), Rb(430), Sb(32), Sc(1.4), Se(300), Zn(460)

[a] d means days, h means hours, s = seconds, m = minutes, mon = months, E means irradiation with epithermal neutrons.
[b] F. Chisela and P. Bratter, *Anal. Chim. Acta.*, **188**, 85 (1986).
[c] A. Chatt, M. Sayjad, K.N. De Silva and C.A. Secord in *IAEA Health-related Monitoring of Trace Element Polluants using Nuclear Techniques*, IAEA TECDOC-330, IAEA, Vienna 1985, p. 33.
[d] W.C. Cunningham and W.B. Stroube, Jr, *Sci. Tot. Environ.* **63**, 29 (1987).

2113-keV line of ^{56}Mn, the optimum irradiation time is 4 min. However, for samples with a low Ca content longer irradiations up to 7 min were used. Na, K, Rb, Ba, Zn, Fe, Sc, Cr, Co and Br were determined by 10 h of irradiation with a thermal neutron flux of 8×10^{13} n cm^{-2} s^{-1}. The samples were counted three times, after a decay of 4–5 d (Na and K), a decay of 7 d (Br and Ca) and a decay of 14 d (Ba, Co, Cr, Fe, Rb, Sc and Zn).

Oddone et al.[96] studied the elemental content of Anatolian obsidians, volcanic glasses, by INAA using both thermal and epithermal neutrons. Mg, K, Ca, Ti, Mn, As and Dy were determined by 2 min of irradiation at a thermal neutron flux of 1×10^{12} n cm^{-2} s^{-1} in a TRIGA Mark II reactor, 250 kW. The sample was counted six times after decays of 0.5, 10, 30, 120, 480 and 960 min. The longer the decay time, the longer was the counting. Al and Si were determined by repeating the short-time irradiation for a sample wrapped in a 10-mm thick Cd foil. Both Al and Si were determined from the 1779-keV line of ^{28}Al, using the epithermal neutron activation reaction ^{28}Si(n,p)^{28}Al. Na, Ca, Sc, Cr, Fe, Co, Zn, Se, Rb, Zr, Nb, Sb, Ba, La, Ce, Nd, Sm, Eu, Gd, Tb, Ho, Tm, Yb, Lu, Hf, Ta and Th were determined by irradiation for 20 h. The samples were counted five times after decays of 3, 6, 24, 48 and 96 d. Epithermal neutron INAA was applied also with long irradiation (using samples inside a 1.5-mm thick Cd box) in order to reduce the interferences from elements with low resonance integrals and high thermal cross-section (small I_0/σ ratio). The samples were irradiated for 40 h and counted five times after decays of 3, 6, 12, 24 and 48 d. The epithermal neutron INAA improved the sensitivity and accuracy of the measurement of Zn, Se, Rb, Zr, Nb, Sb, Cs, Ba, Gd, Tb, Ho, Tm, Yb, Hf, Ta and Th. Ni, Sr and U, which could not be measured by thermal INAA, could be determined by the epithermal INAA.

Sato[97] suggested four different schemes of timings for the determination of various elements in biological samples (three times of irradiation with the last one being counted twice, after different decay times). His suggestion for the various elements is given in Table 3. Table 4 gives three examples from the literature, for the irradiation of biological samples together with their timings and the limits of detection of some elements via INAA.

ABBREVIATIONS AND ACRONYMS

AES	Auger Electron Spectroscopy
BGO	Bismuth Germanate
DE	Double-escape
EC	Electron Capture
ENAA	Epithermal Neutron Activation Analysis
FTIR	Fourier Transform Infrared
fwhm	full width at half-maximum
HFR	High Flux Reactor
INAA	Instrumental Neutron Activation Analysis
IRN	Indicator Radionuclide
LWR	Light Water Reactor
MCA	Multichannel Analyzer
NAA	Neutron Activation Analysis
NBS	National Bureau of Standards
NDP	Neutron Depth Profiling
NIST	National Institute of Standards and Technology
RBS	Rutherford Backscattering Spectroscopy
RNAA	Reactor Neutron Activation Analysis

SE	Single-escape
SIMS	Secondary Ion Mass Spectrometry
SRM	Standard Reference Material

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis

Nuclear Methods **(Volume 14)**
Instrumental Neutron Activation Analysis: Gamma Lines Table • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides

REFERENCES

1. K.S. Krane, *Introductory Nuclear Physics*, John Wiley, New York, 1988.
2. B.G. Harvey, *Introduction to Nuclear Physics and Chemistry*, 2nd edition, Prentice Hall, Englewood Cliffs, NJ, 1969.
3. G. Friedlander, J.W. Kennedy, E.S. Macias, J.M. Miller, *Nuclear and Radiochemistry*, 3rd edition, John Wiley, New York, 1981.
4. G.R. Choppin, J. Rydberg, *Nuclear Chemistry*, Pergamon Press, Oxford, 1980.
5. W.D. Ehmann, D.E. Vance, *Radiochemistry and Nuclear Methods of Analysis*, John Wiley, New York, 1991.
6. A. Vertes, I. Kiss, *Nuclear Chemistry*, Elsevier, Amsterdam, 1987.
7. R.B. Firestone, V.S. Shirley, C.M. Baglin, Y.S. Chu, J. Zipkin, *Table of Isotopes*, 8th edition, John Wiley, New York, 1996.
8. G. Erdtmann, W. Soyka, *The Gamma Rays of the Radionuclides*, Verlag Chemie, Weinheim, 1979.
9. U. Reus, W. Westmeier, *At. Data Nucl. Data Tables*, **29**, 1–192, 193–406 (1983).
10. E. Browne, R.B. Firestone, V.S. Shirley, *Table of Radioactive Isotopes*, John Wiley, New York, 1986.
11. Z.B. Alfassi, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 35, Vol. II, 1990.
12. H.P. Yule, in *Nondestructive Activation Analysis*, ed. S. Amiel, Elsevier, Amsterdam, 113, 1981.
13. M.D. Glascock, *Tables for Neutron Activation Analysis*, 4th edition, University of Missouri, Columbia, MO, 1996.
14. K. Heydorn, *Radiochim. Acta*, **3**, 161 (1964).
15. Z.B. Alfassi, B. Rietz, *Analyst*, **119**, 2407 (1994).
16. O.U. Anders, *Anal. Chem.*, **32**, 1368 (1960); **33**, 1706 (1961).
17. Z.B. Alfassi, N. Lavi, *Nucl. Instrum. Meth. Phys. Res. A*, **365**, 190 (1995).
18. L.I. Zhuk, A.A. Kist, I.N. Mikholskaya, N.S. Osinskaya, T. Tilayev, S.I. Tursunbayev, S.V. Agzamova, *J. Radioanal. Nucl. Chem.*, **120**, 369 (1988).
19. R.G.V. Hancock, M.D. Grynpas, B. Alpert, *J. Radioanal. Nucl. Chem.*, **110**, 283 (1987).
20. W.C. Cunningham, W.B. Stroube, Jr, *Sci. Total Environ.*, **63**, 29 (1987).
21. R. Dams, J.A. Robbins, K.A. Rahn, J.W. Winchester, *Anal. Chem.*, **42**, 861 (1970).
22. A. Wyttenbach, S. Bajo, L. Tobler, *J. Radioanal. Nucl. Chem.*, **114**, 137 (1987).
23. A. Simons, S. Landsberger, *J. Radioanal. Nucl. Chem.*, **110**, 555 (1985).
24. F. De Corte, A. Speecke, J. Hoste, *J. Radioanal. Chem.*, **3**, 205 (1969).
25. A. Simonits, F. De Corte, J. Hoste, *J. Radioanal. Chem.*, **31**, 467 (1976).
26. F. De Corte, S. Jovanovic, A. Simonits, L. Moens, J. Hoste, *Atomkernenergie, Suppl.*, **44**, 641 (1984).
27. Z.B. Alfassi, N. Lavi, *J. Radioanal. Nucl. Chem.*, **150**, 183 (1991).
28. M.D. Glascock, W.Z. Tian, W.D. Ehmann, *J. Radioanal. Nucl. Chem.*, **92**, 379 (1985).
29. Z.B. Alfassi, N. Lavi, *Analyst*, **109**, 959 (1984).
30. S. Landsberger, A. Arendt, B. Keck, M. Glascock, *Trans. Am. Nucl. Soc.*, **56**, 230 (1988).
31. B. Kratochvil, N. Motkosky, M.S.M. Duke, D. Ng, *Can. J. Chem.*, **65**, 1047 (1987).
32. X. Lin, K. Heydorn, B. Rietz, *J. Radioanal. Nucl. Chem.*, **160**, 85 (1992).
33. Z.B. Alfassi, T.U. Probst, B. Rietz, *Anal. Chim. Acta*, **360**, 243 (1998).
34. M.J.J. Theunissen, H.J.J. Jaspers, J.M.G. Hansen, M.L. Verheijke, *J. Radioanal. Nucl. Chem.*, **113**, 391 (1987).
35. M.L. Verheijke, H.J.J. Jaspers, J.M.G. Hansen, M.J.J. Theunissen, *J. Radioanal. Nucl. Chem.*, **113**, 397 (1987).
36. R.M. Lindstrom, in *Microelectronic Processing: Inorganic Materials Characterization*, ed. L.A. Casper, AES Symposium Series, No. 295, American Chemical Society, Washington, DC, 294, 1986.
37. G. Revel, N. Deschamps, C. Dardenne, J.L. Pastol, B. Hania, H. Dinh Ngugen, *J. Radioanal. Nucl. Chem.*, **85**, 137 (1984).
38. G. Revel, in *Euroanalysis VII Conference Proceedings 1987*, ed. E. Roth, Editions Physiques, Paris, 167, 1988.
39. K. Fujingawa, K. Kudo, *J. Radioanal. Chem.*, **52**, 411 (1979).
40. S. Niese, *J. Radioanal. Chem.*, **38**, 37 (1977).
41. E.W. Haas, R. Hofmann, F. Richter, *J. Radioanal. Chem.*, **69**, 219 (1982).
42. H. Jaskolska, L. Rowlinska, *J. Radioanal. Chem.*, **26**, 31 (1975).

43. Z.B. Alfassi, M.H. Yang, *J. Radioanal. Nucl. Chem.*, **132**, 99 (1989).
44. N. Lavi, Z.B. Alfassi, *Analyst*, **109**, 361 (1984); Z. Etzion, Z. Alfassi, N. Lavi, R. Yagil, *Biol. Trace Elem. Res.*, **12**, 411 (1987).
45. D. Brune, K. Jirlow, *Nukleonik*, **6**, 242 (1964).
46. E.S. Gladney, D.R. Perrin, J.P. Balagna, C.L. Warner, *Anal. Chem.*, **52**, 2128 (1980).
47. S.J. Parry, *J. Radioanal. Chem.*, **72**, 195 (1982); *J. Radioanal. Nucl. Chem.*, **81**, 143 (1984).
48. Z.B. Alfassi, *J. Radioanal. Nucl. Chem.*, **90**, 151 (1985).
49. S.J. Parry, *J. Radioanal. Chem.*, **59**, 423 (1980).
50. L.A. Currie, *Anal. Chem.*, **40**, 587 (1986).
51. H. Bem, D.E. Ryan, *Anal. Chim. Acta*, **124**, 373 (1981).
52. W.Z. Tian, W.D. Ehmann, *J. Radioanal. Nucl. Chem.*, **84**, 89 (1984).
53. F. Rossitto, M. Terrani, S. Terrani, *Nucl. Instrum. Methods*, **103**, 77 (1972).
54. T.G. Williamson, P.E. Penneche, B. Hostica, J.S. Brenizer, T.L. Ngugen, *J. Radioanal. Nucl. Chem.*, **114**, 387 (1987).
55. R.K. Tokay, M. Skalberg, G. Skarnemak, *J. Radioanal. Nucl. Chem.*, **96**, 265 (1985).
56. R.J. Rosenberg, *J. Radioanal. Chem.*, **62**, 145 (1981).
57. W.B. Stroube, Jr, W.C. Cunningham, G.J. Lutz, *J. Radioanal. Nucl. Chem.*, **112**, 341 (1987).
58. W.D. Ehmann, J. Brückner, D.M. McKown, *J. Radioanal. Chem.*, **57**, 491 (1980).
59. F. Chisela, D. Gawlik, P. Bratter, *J. Radioanal. Nucl. Chem.*, **98**, 133 (1987).
60. D.C. Stuart, D.E. Ryan, *Can. J. Chem.*, **59**, 1470 (1981).
61. H. Soreq, H.C. Griffin, *J. Radioanal. Chem.*, **79**, 135 (1983).
62. Z.B. Alfassi, N. Lavi, *J. Radioanal. Nucl. Chem.*, **84**, 363 (1984).
63. R.G.V. Hancock, *J. Radioanal. Chem.*, **69**, 313 (1982).
64. M. Ördögh, E. Orban, G. Miskovitz, J. Appel, E. Szabo, *Int. J. Appl. Radiat. Isot.*, **25**, 61 (1974).
65. Z.B. Alfassi, N. Lavi, *Analyst*, **109**, 959 (1984).
66. E. Steinnes, *Talanta*, **14**, 753 (1967).
67. E. Steinnes, *Anal. Chim. Acta*, **57**, 457 (1971).
68. M. Mantel, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 111, Vol. 1, 1990.
69. M. Mantel, S. Amiel, Z.B. Alfassi, *Anal. Chem.*, **50**, 441 (1978).
70. Z.B. Alfassi, T. Biran-Izak, M. Mantel, *Nucl. Instrum. Methods*, **151**, 227 (1978).
71. L.B. Graman, *J. Radioanal. Nucl. Chem.*, **99**, 75 (1986).
72. J. Hertogen, R. Gijbels, *Anal. Chim. Acta*, **56**, 61 (1971).
73. M. Mantel, S. Amiel, *J. Radioanal. Chem.*, **16**, 127 (1973).
74. P.A. Baedecker, J.J. Rowe, E. Steinnes, *J. Radioanal. Chem.*, **40**, 115 (1977).
75. D.R. Allen, W. Brookhart, *Anal. Chem.*, **46**, 1297 (1974).
76. M.S. Rapaport, M. Mantel, R. Nothman, *Anal. Chem.*, **51**, 1356 (1979).
77. M. Mantel, C. Shenberg, M.S. Rapaport, *J. Radioanal. Chem.*, **75**, 145 (1982).
78. M. Stiller, M. Mantel, M.S. Rapaport, *J. Radioanal. Nucl. Chem.*, **83**, 345 (1984).
79. Z.B. Alfassi, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 97, Vol. 1, 1990.
80. S. Amiel, *Anal. Chem.*, **34**, 1683 (1963).
81. C. Shenberg, Y. Nirel, Z. Alfassi, Y. Shiloni, *J. Radioanal. Nucl. Chem.*, **114**, 207 (1988).
82. J.F. Ziegler, G.W. Cole, J.E.E. Baglin, *J. Appl. Phys.*, **43**, 3809 (1972).
83. K. Müller, R. Henkelmann, H. Borofka, *Nucl. Instrum. Methods*, **129**, 557 (1979).
84. J. Kvitek, V. Hnatowicz, P. Kotas, *Radiochem. Radioanal. Lett.*, **24**, 205 (1976).
85. J. Cerevna, V. Hnatowicz, J. Hoffmann, Z. Kosina, J. Kvitek, P. Onheiser, *Nucl. Instrum. Methods*, **188**, 185 (1981).
86. J.P. Biersack, D. Fink, R. Henkelmann, K. Müller, *Nucl. Instrum. Methods*, **149**, 93 (1978).
87. R.G. Downing, R.F. Fleming, J.K. Langland, D.H. Vincent, *Nucl. Instrum. Methods*, **218**, 47 (1983).
88. R.R. Greenberg, R.F. Fleming, R. Zeisler, *Environ. Int.*, **10**, 129 (1984).
89. C.C. Cunningham, W.B. Stroube, Jr, *Sci. Total Environ.*, **63**, 29 (1987).
90. J.X. Sun, R.E. Jervis, *J. Radioanal. Nucl. Chem.*, **89**, 553 (1985).
91. J.X. Sun, R.E. Jervis, *J. Radioanal. Nucl. Chem.*, **114**, 89 (1987).
92. F.Y. Iskander, *J. Radioanal. Nucl. Chem.*, **89**, 511 (1985).
93. G.D. McOrist, A. Smallwood, J.J. Fardy, *J. Radioanal. Nucl. Chem.*, **185**, 293 (1994).
94. G.D. McOrist, A.S. Smallwood, *J. Radioanal. Nucl. Chem.*, **223**, 9 (1997).
95. G. Wallner, Ch. Katzlberger, *J. Radioanal. Nucl. Chem.*, **223**, 139 (1997).
96. M. Oddone, Z. Yegingil, G. Bigazzi, T. Ercan, M. Ozdogan, *J. Radioanal. Nucl. Chem.*, **224**, 27 (1997).
97. T. Sato, in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, FL, 323–358, Vol. II, 1990.

Neutron Radiography for the Analysis of Plant–Soil Interactions

Brett H. Robinson, Ahmad Moradi and Rainer Schulin
Institute of Terrestrial Ecosystems, ETH, Zürich, Switzerland

Eberhard Lehmann
Paul Scherrer Institut, Villigen, Switzerland

Anders Kaestner
Paul Scherrer Institut, Villigen, Switzerland

1 Introduction 1
2 Neutron Radiography Configuration 1
3 Neutron Interactions with the Plant–Soil System 2
4 Configuration of Soil–Plant System For Optimal Imaging with Neutron Radiography 3
5 Practical Considerations 5
6 Application 6
7 Conclusions and Outlook 6
Related Articles 7
References 7

Neutron radiography (NR) can be used to quantify the spatial distribution of water in the soil–plant system with high precision and good spatial resolution. This property of neutron imaging results from the high interaction probability of hydrogen nuclei with slow neutrons. If there is a sufficient difference between the water content of the soil and roots, neutron radiographs can reveal plant roots and show root development. NR is noninvasive, and the radiation dose needed to image plant roots in soil does not affect plant development. Quantification of the soil's water content often requires correction for neutron-scattering artifacts. Root visibility is proportional to root thickness, and is inversely related to the width of the sample container and the water and organic matter contents of the ambient soil. Ideally, the soil should have low organic matter content and low water content but still permit the normal development of plant roots. Currently, the availability of neutron-imaging facilities limits the widespread application of NR to soil and root studies. However, technological development and increased investment will result in NR becoming a standard method for some soil–plant analyses.

1 INTRODUCTION

Terrestrial plants provide humanity with food and oxygen. Above- and belowground plant processes are equally important: while assimilation through photosynthesis occurs in the aboveground portions, the uptake of water and mineral nutrients occurs belowground. We have a limited understanding of some belowground processes because soil hinders the observation of plant roots. It is difficult to measure root development and water flux in the root zone without disturbing root growth or using artificial systems. Existing techniques include minirhizotrons, which are transparent plastic tubes inserted into the ground to view the roots, e.g. by using a video camera.[1] Minirhizotrons interfere with the root environment and only provide an incomplete picture. X-ray radiography has insufficient contrast to reveal root–water interactions.[2]

Willatt et al.[3] showed that NR could reveal roots and root zone processes, without greatly perturbing the system, thus allowing sequential measurements. This is a critical advantage of NR over other techniques. However, the available technology in the 1970s gave images of insufficient quality and exposed the plants to radiation doses that were potentially harmful. Recent technological developments, especially improved beam collimation, detection systems, and image-processing techniques, have allowed the production of images with much higher contrast and spatial resolution while reducing the plants' radiation exposure. These advances open up the possibility of using this technique to study root system development in soils and simultaneously monitor soil moisture distribution in near real time. Here, we describe the state of the art of NR as it relates to the analysis of plant–soil interactions.

2 NEUTRON RADIOGRAPHY CONFIGURATION

Figure 1 shows the configuration of a typical NR facility. NR requires a neutron source, which may be a reactor, the target of an elemental particle accelerator, or a neutron-emitting isotope. Thermal (12–100 meV) and cold (0.12–12 meV) neutrons are preferable for NR investigations. Therefore, epithermal, intermediate, and fast neutrons from the neutron source must be slowed

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

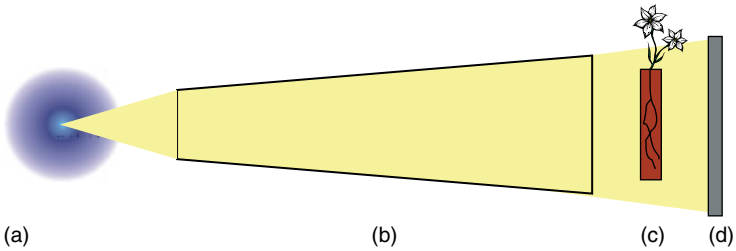

Figure 1 Configuration required for neutron imaging: (a) neutron source, (b) collimator, (c) sample, and (d) detector.

down using a moderator such as heavy water. Neutrons enter a collimator that forms a neutron beam with specific geometric properties. The collimator may also contain filters that modify the energy spectrum of the beam or reduce the beam's content of γ rays. The neutron beam is transmitted through the sample onto a plane position-sensitive detector, which is usually a scintillation screen. A charge-coupled device (CCD) camera coupled to the detector records a two-dimensional image that is a projection of the object on the detector plane. The advent of high-resolution digital cameras has allowed fast imaging and increased resolution. Digital imaging techniques that use a CCD camera combined with image-processing tools nowadays yield quantifiable images with a resolution of about 100 μm. Lehmann et al.[4] obtained a much higher spatial resolution with a special setup for microtomography. The required exposure time for such images is in the order of seconds. Images acquired using imaging plates or films (used until 1995) required several minutes of exposure to the potentially damaging neutron beam. The radiation dose received per image using modern techniques is about 0.003 mSv,[5] which is some two orders of magnitude less than the minimum value of 0.2 mSv h^{-1} found to affect plant growth.[6]

3 NEUTRON INTERACTIONS WITH THE PLANT–SOIL SYSTEM

NR is based on the Beer–Lambert exponential law of attenuation of radiation passing through matter[7]:

$$I = I_0 \exp(-\Sigma_{\text{sample}} d) \quad (1)$$

where I is the attenuated radiation (neutron) flux (cm^{-2} s^{-1}), after an incident neutron flux I_0 passes through a material of thickness d (cm) with an attenuating coefficient Σ (cm^{-1}), which is a characteristic of the material. The attenuation coefficient, also called the *macroscopic cross section*, is related to the tabulated microscopic cross section σ (cm^2) as

$$\Sigma = N\sigma \quad (2)$$

with a nuclear density N (M). When a sample is placed in a neutron beam, heterogeneities in the composition and thickness of the sample result in variations in the intensity of the transmitted beam. Unlike X rays, neutron radiation interacts with atomic nuclei. There is no systematic change in the neutron attenuation coefficient with atomic number or mass. Each isotope has a specific neutron cross section, σ, which is also energy dependent. Hydrogen has a neutron cross section some 10 times greater than deuterium and also greater than many other elements in the soil–plant system.

NR reveals structures in plant–soil systems owing to differences in the Σ values of the system's components. Table 1 shows a list of the chemical elements in the plant–soil system, along with their abundances and relative neutron cross sections. In both plant and soil, hydrogen is responsible for more than 90% of the neutron attenuation.

Some hydrogen is associated with organic molecules in the system; however, most hydrogen is water borne. Plant roots may thus be distinguished from soil due to their higher water content (θ). The gravimetric water content, θ ($g\,g^{-1}$) of plant roots generally ranges between 0.7 and 0.95 $g\,g^{-1}$, while that of soils at field capacity usually ranges between 0.12 and 0.3 $g\,g^{-1}$. The structures that NR reveals in the soil–plant system are sensitive to θ.

The high neutron attenuation coefficient of hydrogen is an important advantage of NR over X rays when applied to soil–plant system because the difference in water content allows the visualization of roots. The X-ray attenuation coefficients of root and soil components are less distinct (data not shown); therefore, the resulting radiograph has less contrast. Figure 2 shows that NR provides a better contrast between roots and soil than X rays.

The attenuation coefficient (Σ) results from two types of neutron interactions with matter: absorption (Σ_a) and scattering (Σ_s) (Figure 3a). Hydrogen attenuates neutrons primarily by noncoherent elastic scattering.[11] Neutron scattering causes deviations from the exponential law of attenuation for thicker samples (more than a few millimeters) because some neutrons are multiple scattered into the detector plane, thus producing

Table 1 Chemical elements, listed in order of abundance, in the plant[8]–soil[9] system, along with their relative neutron attenuation coefficients.[10] The plant and soil are assumed to have water contents of 0.8 and 0.2 g g^{-1}, respectively

	Plant (mol kg^{-1})	Soil (mol kg^{-1})	Element Σ (cm^{-1})	Plant neutron attenuation (cm^{-1})	Soil neutron attenuation (cm^{-1})
Hydrogen	100	24	3.4	344	82
Oxygen	50	35	0.17	8.5	6
Silicon	71E–4	9.3	0.11	<0.1	1.02
Carbon	7.5	1.1	0.56	4.2	0.6
Aluminum	7.4E–4	2.1	0.1	<0.1	0.21
Potassium	5.1E–2	0.93	0.06	<0.1	<0.1
Sodium	8.7E–4	0.82	0.09	<0.1	<0.1
Nitrogen	0.21	3.6E–2	0.43	<0.1	<0.1
Calcium	2.5E–2	0.19	0.08	<0.1	<0.1
Iron	3.6E–4	0.19	1.2	<0.1	0.22
Magnesium	1.6E–2	7.8E–2	0.15	<0.1	<0.1
Phosphorus	1.3E–2	1.2E–2	0.12	<0.1	<0.1
Titanium	4.1E–6	2.4E–2	0.6	<0.1	<0.1
Sulfur	6.2E–3	2.3E–3	0.06	<0.1	<0.1
Manganese	1.8E–4	2.8E–3	1.2	<0.1	<0.1
Chlorine	5.6E–4	1.6E–3	1.3	<0.1	<0.1
Zinc	6.1E–5	5.3E–4	0.35	<0.1	<0.1
Boron	3.7E–4	1.8E–4	102	<0.1	<0.1
Copper	1.9E–5	1.6E–4	1.1	<0.1	<0.1
Molybdenum	2.1E–7	5.2E–5	0.52	<0.1	<0.1
Total	158	74		356	91

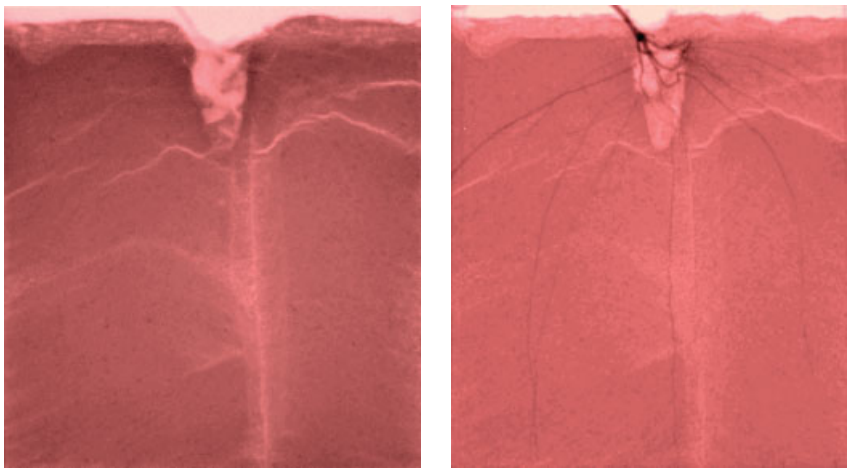

(a) (b)

Figure 2 Comparison of contrast between roots and soil in X-ray (120 keV) radiograph (a) and neutron radiograph (b) of the same sample. While soil heterogeneity and soil cracks are more visible in the X-ray radiograph, neutron radiography provides better contrast between soil and roots.

an artificially high signal. While the significance of neutron scattering is of little importance in the detection of roots in soil, it can be problematic when quantifying θ of the system. Algorithms such as the quantitative neutron imaging (QNI)[12] correct for the nonlinearity arising from neutron scattering. The QNI iteratively reconstructs the image by overlapping point-scattered functions calculated using a Monte Carlo simulation.

4 CONFIGURATION OF SOIL–PLANT SYSTEM FOR OPTIMAL IMAGING WITH NEUTRON RADIOGRAPHY

A neutron radiograph is the result of all the neutron attenuation processes that occur when the neutron beam passes through a sample. The Σ value of the soil depends on θ and the neutron attenuation properties of its solid components. Soils that are high in iron or organic matter

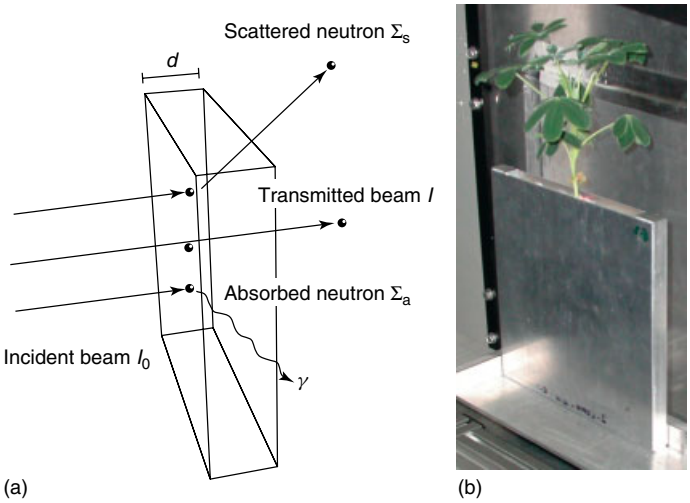

Figure 3 Neutron interactions with the sample (a) and an example of a plant sample (*Lupinus albis*) mounted in front of the detector (b). The dimensions of the sample container are 150 mm × 150 mm × 12 mm.

Table 2 Properties of various plant growth media with regard to NR

	Bulk density ($g\,cm^{-3}$)	Water content at 1 bar ($g\,g^{-1}$)	Σ (dry material) (cm^{-1})	Σ (water content at −1 bar) (cm^{-1})	Notes
Perlite	0.125	1.08	0.1	4.5	–
Porous glass beads	0.49	0.16	1.6	2.5	High Σ due to boron in glass
Ferrous mine tailings	1.4	0.08	1	1.6	–
Loam	1.2	0.17	0.5	1.5	–
Peat	0.58	0.09	0.75	1.4	–
Loamy sand	1.3	0.09	0.3	0.75	Normal root development
Fine quartz sand	1.5	0.01	0.25	0.35	Root development perturbed
Coarse quartz sand	1.45	0.01	0.25	0.3	

are unsuitable for NR because their high Σ values obscure the visualization of plant roots. The value of θ of the soil at the time of measurement should be as low as possible without inducing water stress in the plants. In practice, this represents the soil's θ at a water potential of −1 bar. Silica sands have a low inherent Σ and a low θ at −1 bar. However, Menon et al.[5] showed that the high density and sharp edges of this material perturb normal root growth. Moradi et al. (unpublished data) tested a variety of plant growth media for their suitability in NR (Table 2).

Table 2 shows that loamy sand, which permits normal root development, has a relatively low Σ at a θ at a water potential of −1 bar. Loamy sand has a higher θ at field capacity (ca. $0.35\,g\,g^{-1}$) than quartz sand (ca. $0.15\,g\,g^{-1}$). Plants can thus be left longer in loamy sand without irrigation.

Root visibility is proportional to root thickness, and inversely related to the width of the sample container.[13–17] The minimum detectable root thickness increases exponentially as the thickness of the soil profile increases. Moradi et al. (unpublished data) showed that using the loamy sand in Table 2, with a soil θ of $0.16\,g\,g^{-1}$ (ca. −1 bar) and an average root θ of $0.85\,g\,g^{-1}$, the minimum detectable root thickness R (mm) is related empirically to the slab thickness T (mm) according to the function:

$$R = 0.0034T^{1.68} \quad (3)$$

Roots are most easily visualized in thin containers. However, this condition is not conducive to plant growth since the walls of the container restrict root development. Therefore, one needs to find a balance between the ease of root visualization and the restriction of normal plant development.

For many herbaceous species, a soil profile with a thickness of 12 mm provides enough space for relatively normal root development.[5] This gives a minimal detectable root thickness of 0.22 mm (Equation 2). Such a setup allows the visualization of the skeleton of the root system; however, fine roots (<2 mm in diameter) are undetectable.

5 PRACTICAL CONSIDERATIONS

The sample container should be made of a material with a low Σ, such as aluminum ($\Sigma = 0.1\,\mathrm{cm}^{-1}$). Importantly, the system should not contain high concentrations of cobalt, which can form the persistent radioactive isotope Cobalt-60, with a half-life of about five years, upon exposure to neutron radiation.

Filling the sample container with soil as homogeneously as possible provides better contrast for root visualization. Filling from the side of the container rather than from the top results in less structural heterogeneity associated with variations in the particle size distribution and thus pore size distribution, which are visible in the resulting neutron radiograph because of variations in soil water content. Figure 4 shows the effect of filling patterns caused by pouring soil from the top of the container.

The suitability of NR to investigate root systems differs among plant species. As described earlier, thick roots with a high θ are more easily resolved than finer, drier roots. In a 10-mm-thick soil profile, the minimum detectable root size is ca. 0.16 mm (Equation 2). However, many plant species produce finer roots, which would not be visible in this system. Decreasing the thickness of the soil profile may permit the visualization of these roots, but may cause unnatural root growth patterns due to confinement. Moradi et al. (unpublished data) reported good root visibility in some members of the Fabaceae and Asteraceae families, whereas resolution was insufficient in Brassicacae, Solonaceae, and Poaceae. In principle, larger species such as small trees could be investigated using NR. However, this would require a larger container for nonperturbed growth. Consequently, the spatial resolution would decrease and one could only resolve large structural roots.

The water content of the soil and the plant should be monitored so that the soil water content is low enough at the time of imaging to provide adequate contrast yet not so low that the plant becomes water stressed. If a series of measurements are to be taken, then one can calculate the soil θ in the zone of interest from the results of each radiograph. This requires scattering correction and a water quantification algorithm calibrated for the particular plant–soil system. Water quantification using NR has the advantage over a gravimetric measurement because it can be used to determine the water content of the soil in the zone of interest, rather than providing an average value of a soil profile that may have a heterogeneous moisture distribution.

The neutron beam formed in the collimator is not perfectly homogeneous. This also holds for the detector system. Therefore, there are spatial variations in intensity of the resulting radiography that are caused by the beam, rather than the sample. This requires that each radiograph be corrected by normalizing the image by a "flat field" or open-beam image with no sample. Similarly, noise generated by the camera assembly should be removed. A

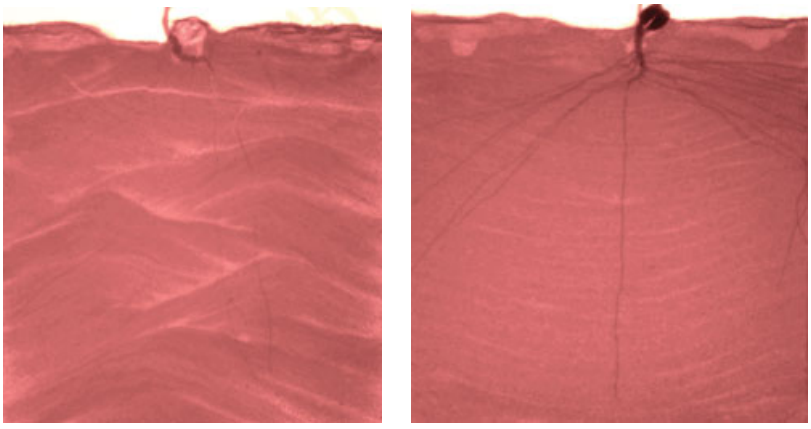

(a) (b)

Figure 4 Soil heterogeneity resulting from filling patterns can affect the image quality. The radiographs show the roots of *Cicer arietinum* in identical soil with high heterogeneity (a) and low heterogeneity (b).

standard first step in image analysis is normalization, i.e. to transform recorded raw information relative values:

$$I' = NORM\frac{I_{raw} - I_{dark}}{I_{openbeam} - I_{dark}} \quad (4)$$

where I_{raw} is the image as registered by the camera, I_{dark} is the dark noise image without the beam, $I_{openbeam}$ contains the spatial field variation of the beam without object, and NORM is a factor to bring the resulting value into the valid range of the image-processing tools.

The background gray-level intensity within a neutron radiograph, caused, for example, by variations in the soil water content, may vary so much that a single global threshold cannot differentiate the roots satisfactorily. Menon et al.[5] overcame this problem by modifying an algorithm originally developed for resolving blood vessels in retinal images.[18]

6 APPLICATION

The application of NR can enhance the study of root development and root–soil interactions by revealing the location of roots over time without disturbance. Plant root development is a function of the plant species and the nature of the soil into which they penetrate. Soil components, such as organic matter, nutrients, and contaminants, occur heterogeneously. NR is an ideal tool to study how roots interact with patches of low or high concentration. When roots encounter a patch or discontinuity in soil, they may proliferate, wither, or continue growing unaffected. Figure 5(a) and (b) shows how NR can reveal plant responses to a patch of nickel, a toxic heavy metal, in soil. Such experiments, for example, may reveal species that avoid contaminant hotspots and reduce the risk of plant contaminant uptake. Similarly, understanding the mechanisms by which crop plants interact with patches of nutrients may aid the development of treatments to improve crop, and subsequently human, nutrition.

Perhaps more importantly, NR can quantify the spatial distribution and flux of water in the plant–soil system in near real time, in combination with root imaging. This permits the study of root water uptake and the effect of roots on the passage of water through soil. Figure 6 shows a series of images detailing water infiltration into a profile and water uptake by plant roots. Such studies have a wide variety of potential applications, such as selecting vegetation to minimize leaching from contaminated sites and the optimization of irrigation and fertilization systems.

7 CONCLUSIONS AND OUTLOOK

The quantification of root mass and soil water content using NR of plant–soil interactions has considerable scope for improvement, particularly, the processing of raw images obtained at the neutron facility. Refinement of root segmentation algorithms described by Menon et al.[5] would greatly enhance the accuracy and precision of the technique. One obvious drawback of NR is that, at present, it requires access to a specialized facility, of which there are only a few available worldwide with the desirable performance. Competition for beam time is fierce, since NR finds applications in many fields of science. However, the usefulness of NR indicates that it may follow the same pattern of development as synchrotron radiation facilities. Initially, financial

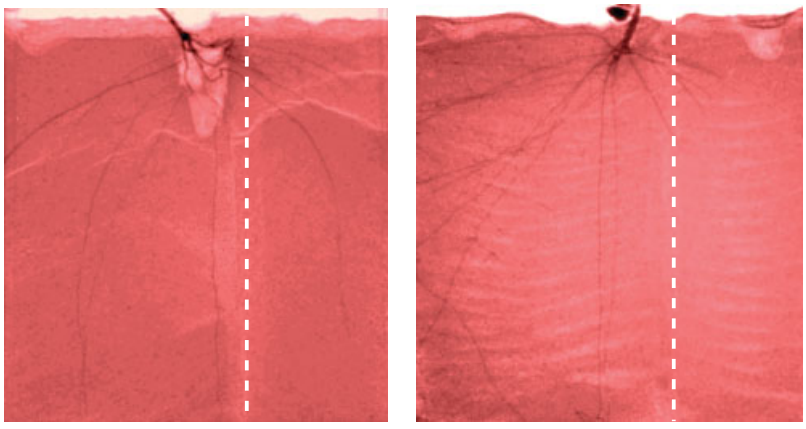

Figure 5 Neutron radiographs of two 150 mm × 150 mm × 120 mm slabs filled with sandy loam. The area to the right of the dotted line was spiked with 125 mg kg^{-1} Ni. The roots of *Cicer arietinum* (b) avoid the high-Ni zone, while the roots of *Berkheya coddii* (a) are unaffected. Neither plant showed any differences in the aboveground portions relative to their respective controls.

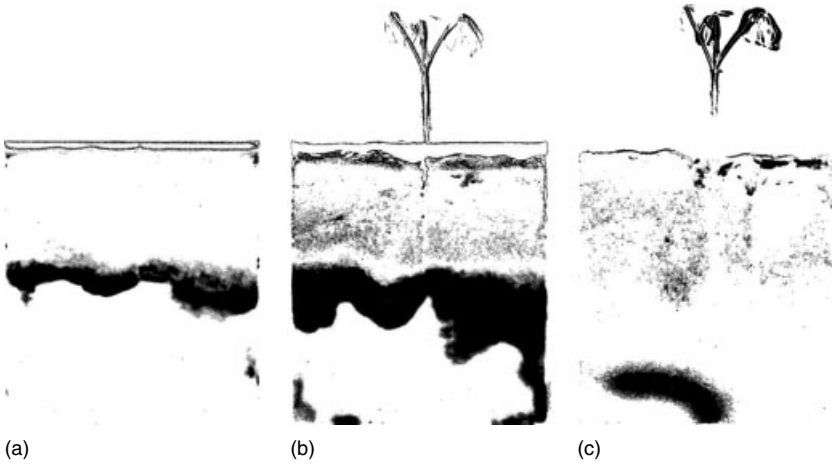

Figure 6 The change in water content (both positive and negative) of a 150 mm × 150 mm × 120 mm container, in which *Lupinus albus* was growing, after the infiltration of 10 mL of water ((a) 1–2 min, (b) 2–5 min, (c) 28–40 min). The wet front shows up as a dark band, while root water uptake is visible in b and c as a discontinuous gray area.

and technical constraints limited their application, but their usefulness ensured subsequent technological development and capital injection, and thus they became commonplace and the standard equipment for some analyses.

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Nuclear Reaction Analysis • Prompt γ-Neutron Activation Analysis • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis • Soil Instrumental Methods

REFERENCES

1. R.O. Kuchenbuch, K.T. Ingram, 'Image Analysis for Non-destructive and Non-invasive Quantification of Root Growth and Soil Water Content in Rhizotrons', *J. Plant Nutr. Soil Sc.*, **165**(5), 573–581 (2002).
2. P.J. Gregory, D.J. Hutchison, D.B. Read, P.M. Jenneson, W.B. Gilboy, E.J. Morton, 'Non-invasive Imaging of Roots with High Resolution X-ray Micro-tomography', *Plant Soil*, **255**(1), 351–359 (2003).
3. S.T. Willatt, R.G. Struss, H.M. Taylor, 'In situ Root Studies Using Neutron Radiography', *Agron. J.*, **70**(4), 581–586 (1978).
4. E.H. Lehmann, G. Frei, G. Kuhne, P. Boillat, 'The Micro-setup for Neutron Imaging: A Major Step Forward to Improve the Spatial Resolution', *Nucl. Instrum. Methods A*, **576**(2–3), 389–396 (2007).
5. M. Menon, B. Robinson, S.E. Oswald, A. Kaestner, K.C. Abbaspour, E. Lehmann, R. Schulin, 'Visualization of Root Growth in Heterogeneously Contaminated Soil Using Neutron Radiography', *Eur. J. Soil Sci.*, **58**(3), 802–810 (2007).
6. A. Real, S. Sundell-Bergman, J.F. Knowles, D.S. Woodhead, I. Zinger, 'Effects of Ionizing Radiation Exposure on Plants, Fish and Mammals: Relevant Data for Environmental Radiation Protection', *J. Radiol. Prot.*, **24**(4A), A123–A137 (2004).
7. S. Kasperl, P. Vontobel, 'Application of an Iterative Artefact Reduction Method to Neutron Tomography', *Nucl. Instrum. Meth. A*, **542**(1–3), 392–398 (2005).
8. F.B. Salisbury, C.W. Ross, *Plant Physiology*, Wadsworth, Belmont, CA, 1992.
9. B. Mason, C. Moore, *Principles of Geochemistry*, John Wiley & Sons, Hong Kong, 1982.
10. PSI (2008) '*Comparison Between Neutron and X-ray Attenuation*', from http://neutra.web.psi.ch/images/diag_x_n_e.gif.
11. D.G. Allis, B.S. Hudson, 'Inelastic Neutron Scattering Spectrum of Cs-2[B12H12]: Reproduction of its Solid-state Vibrational Spectrum by Periodic DFT', *J. Phys. Chem. A*, **110**(10), 3744–3749 (2006).
12. R.K. Hassanein, *Correction Methods for the Quantitative Evaluation of Thermal Neutron Tomography*, ETH Zurich, Switzerland, 2006.
13. T.M. Nakanishi, M. Matsubayashi, 'Nondestructive Water Imaging by Neutron Beam Analysis in Living Plants', *J. Plant Physiol.*, **151**(4), 442–445 (1997).

14. J. Furukawa, T.M. Nakanishi, H. Matsubayashi, 'Neutron Radiography of a Root Growing in Soil with Vanadium', *Nucl. Instrum. Meth. A*, **424**(1), 116–121 (1999).

15. T.M. Nakanishi, Y. Okuni, J. Furukawa, K. Tanoi, H. Yokota, N. Ikeue, M. Matsubayashi, H. Uchida, A. Tsiji, 'Water Movement in a Plant Sample by Neutron Beam Analysis as well as Positron Emission Tracer Imaging System', *J. Radioanal. Nucl. Chem.*, **255**(1), 149–153 (2003).

16. U. Matsushima, Y. Kawabata, M. Hino, P. Geltenbort, B.M. Nicolai, 'Measurement of Changes in Water Thickness in Plant Materials Using Very Low-energy Neutron Radiography', *Nucl. Instrum. Meth. A*, **542**(1–3), 76–80 (2005).

17. T.M. Nakanishi, Y. Okuni, Y. Hayashi, H. Nishiyama, 'Water Gradient Profiles at Bean Plant Roots Determined by Neutron Beam Analysis', *J. Radioanal. Nucl. Chem.*, **264**(2), 313–317 (2005).

18. A. Hoover, V. Kouznetsova, M. Goldbaum, 'Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response', *IEEE Trans. Med. Imag.*, **19**(3), 203–210 (2000).

Nuclear Reaction Analysis

Guy Demortier
Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium

1 Introduction: Role of Nuclear Reaction Analysis in Comparison with Other Ion Beam Analysis Techniques 1
2 Fundamentals of Nuclear Reactions 4
3 Kinematics of Nuclear Reactions 5
4 Identification of Signals for Analytical Purposes 8
5 Examples of Application 11
5.1 Analysis of Hydrogen in Various Materials 11
5.2 Depth Profiles of F, Na in SnO_2 Coated Glass 14
5.3 Stoichiometry of Y–Ba–Cu–O Superconductors 15
5.4 Three-dimensional Microanalysis of a Gold–Silicon Eutectic Alloy 16
5.5 Determination of Traces of Sulfur in Ancient Gold Artefacts 19
5.6 Particle-induced X-ray Emission, Particle-induced γ-ray Emission and Nuclear Reaction Analysis for Elemental Analysis of a Biological Reference Material 23
Abbreviations and Acronyms 23
Related Articles 23
References 23

Nuclear reaction analysis (NRA) is one among the numerous analytical techniques involving the irradiation of a material with energetic particles. In NRA methods we exclude any remnant radioactivity together with interactions leading to the scattering of the incident particle only: this last procedure leads to RBS (Rutherford backscattering spectroscopy) and ERD (elastic recoil detection).

In NRA the nature of the emitted particle used as an analytical signal is obviously different from that of the incident one. In order to avoid permanent radioactivity, the energy of the projectile is maintained at a sufficiently low value (below 6 MeV, easily available using a small accelerator). Consequently, due to the large coulomb barrier around heavy nuclei, only light nuclei may be easily identified (atomic mass less than or equal to 30).

After discussing the kinematics and fundamentals of the interactions, we present examples of applications in material sciences, archaeological and biological materials.

1 INTRODUCTION: ROLE OF NUCLEAR REACTION ANALYSIS IN COMPARISON WITH OTHER ION BEAM ANALYSIS TECHNIQUES

When a material is irradiated with charged particles like protons, deuterons, α-particles or other heavier ions, at energies ranging from a few hundred kiloelectron volts to a few megaelectron volts, the emission of photons, charged particles and neutrons gives rise to various spectroscopic methods of elemental analysis. The great majority of interactions of the incident charged particles with atoms of the irradiated material take place with atomic electrons: the incident particle progressively loses its energy ($-\mathrm{d}E$) in the material. The energy loss $\mathrm{d}E$ is expressed by Bethe's law of stopping power $S(E)$ Equation (1):

$$S(E) = -\frac{\mathrm{d}E}{\rho\,\mathrm{d}x} = K\frac{z^2 m}{E_\mathrm{i}}\frac{Z}{A}\ln\frac{aE}{I} \qquad (1)$$

where z, m and E_i are respectively the charge, mass and the energy of the incident particle, where ρ, I, Z and A, are respectively the density, mean ionization potential, atomic number and the atomic mass of the material, and a is an adjustable parameter. K comprises only physical constants. When the incident particle has sufficient energy to ionize atoms from any shell, $S(E)$ decreases with the increase in the incident energy (Table 1). Exceptions to this general rule occur at low incident particle energy, as seen for He^4 and He^3 ions in Table 1, when the projectile cannot eject inner shell electrons, the main consequence being that $Z_\mathrm{eff} < Z$. Therefore $S(E)$, in those cases, is smaller than the expected value.

The incident particle comes to rest after crossing a distance R, known as the range. This incident particle does not experience an appreciable deviation when interacting with atomic electrons. Interaction with one atomic nucleus only rarely causes deviation of the incident particle from a straight line. The range is almost the total distance that the particle travels following a straight line in the material. The range (R) increases with the incident projectile energy ($R \sim E^{1.8}$), and decreases with an increase in its charge and its mass. Table 2 gives some data on ranges of protons, deuterons, ^{3}He and α-particles (R) for a selection of light, medium and high Z elements.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Stopping power of light ions in materials in keV mg^{-1} cm^{-2}

Energy (MeV)	C	Al	Cu	Ag	Au
			Protons		
0.2	5.9 (E + 2)	3.6 (E + 2)	2.0 (E + 2)	1.5 (E + 2)	1.1 (E + 2)
0.5	3.7 (E + 2)	2.5 (E + 2)	1.6 (E + 2)	1.1 (E + 2)	7.6 (E + 1)
1	2.3 (E + 2)	1.7 (E + 2)	1.2 (E + 2)	8.7 (E + 1)	6.0 (E + 1)
2	1.4 (E + 2)	1.1 (E + 2)	8.1 (E + 1)	6.3 (E + 1)	4.5 (E + 1)
3	1.0 (E + 2)	8.4 (E + 1)	6.3 (E + 1)	5.0 (E + 1)	3.7 (E + 1)
5	7.0 (E + 1)	5.8 (E + 1)	4.4 (E + 1)	3.7 (E + 1)	2.8 (E + 1)
7	5.4 (E + 1)	4.5 (E + 1)	3.5 (E + 1)	2.9 (E + 1)	2.3 (E + 1)
			Deuterons		
0.2	7.5 (E + 2)	4.2 (E + 2)	2.4 (E + 2)	1.7 (E + 2)	1.2 (E + 2)
0.5	5.4 (E + 2)	3.4 (E + 2)	1.9 (E + 2)	1.4 (E + 2)	9.5 (E + 1)
1	3.7 (E + 2)	2.5 (E + 2)	1.6 (E + 2)	1.1 (E + 2)	7.6 (E + 1)
2	2.3 (E + 2)	1.7 (E + 2)	1.2 (E + 2)	8.7 (E + 1)	6.0 (E + 1)
3	1.8 (E + 2)	1.4 (E + 2)	9.4 (E + 1)	7.2 (E + 1)	5.1 (E + 1)
5	1.2 (E + 2)	9.5 (E + 1)	7.0 (E + 1)	5.6 (E + 1)	4.0 (E + 1)
7	9.3 (E + 1)	7.5 (E + 1)	5.7 (E + 1)	4.6 (E + 1)	3.4 (E + 1)
			^{3}He		
0.2	2.1 (E + 3)	1.2 (E + 3)	7.0 (E + 2)	5.2 (E + 2)	3.7 (E + 2)
0.5	2.2 (E + 3)	1.3 (E + 3)	7.8 (E + 2)	5.6 (E + 2)	4.0 (E + 2)
1	1.8 (E + 3)	1.2 (E + 3)	7.0 (E + 2)	5.1 (E + 2)	3.5 (E + 2)
2	1.2 (E + 3)	8.7 (E + 2)	5.6 (E + 2)	4.1 (E + 2)	2.9 (E + 2)
3	9.3 (E + 2)	7.0 (E + 2)	4.7 (E + 2)	3.5 (E + 2)	2.5 (E + 2)
5	6.5 (E + 2)	5.0 (E + 2)	3.6 (E + 2)	2.8 (E + 2)	2.0 (E + 2)
7	5.0 (E + 2)	4.0 (E + 2)	2.9 (E + 2)	2.3 (E + 2)	1.7 (E + 2)
			^{4}He		
0.2	2.0 (E + 3)	1.1 (E + 3)	6.4 (E + 2)	4.7 (E + 2)	3.5 (E + 2)
0.5	2.3 (E + 3)	1.4 (E + 3)	7.7 (E + 2)	5.7 (E + 2)	4.0 (E + 2)
1	2.0 (E + 3)	1.3 (E + 3)	7.5 (E + 2)	5.4 (E + 2)	3.8 (E + 2)
2	1.4 (E + 3)	1.0 (E + 3)	6.2 (E + 2)	4.5 (E + 2)	3.1 (E + 2)
3	1.1 (E + 3)	8.2 (E + 2)	5.3 (E + 2)	3.9 (E + 2)	2.7 (E + 2)
5	8.0 (E + 2)	6.1 (E + 2)	4.2 (E + 2)	3.2 (E + 2)	2.2 (E + 2)
7	6.2 (E + 2)	4.9 (E + 2)	3.5 (E + 2)	2.7 (E + 2)	1.9 (E + 2)

Note that the particle range increases with the atomic weight of the material owing to a dependence on Z/A.

During this slowing down of the projectile, a statistical transfer of energy to the atomic electrons of the material takes place. This statistical behavior of the energy loss of the projectile gives rise to energy straggling: an unequal energy loss from identical particles crossing the same target material under identical conditions. The relation between the energy loss and the achieved depth is therefore subject to some inaccuracy. For a very small energy loss, straggling can be calculated using the Landau and Vavilov theory.[1] Energy straggling is not symmetrical around the average value. For monoenergetic protons of 1 MeV crossing a thickness of 2.4×10^{-5} g cm^{-2} ($\approx$90 nm) of pure Al, the mean energy loss is 4.2 keV and straggling on this energy loss has a width at half maximum of 2.7 keV (Figure 1a).[2] For 10^{-3} g cm^{-2}, the energy loss of 1 MeV protons is 400 keV and straggling gives rise to a Gaussian distribution with a width at half maximum of 27 keV (Figure 1b). The relation between energy and depth achieved in a material is then defined in this case with an accuracy of 7%, which is relatively much lower than the inaccuracy at low incident energy: the greater the energy loss, the smaller the statistical uncertainty in the relation between depth with respect to loss. In most practical cases achievable by NRA, the depth profiling analysis is limited to regions of a few hundred nanometers, where the Landau and Vavilov calculation theory is valid, to a few micrometers where the Gaussian profile is effective. In this last case, the straggling, δ, is simply given by Equation (2):

$$\delta = 0.93z\sqrt{\frac{Z}{A}}\sqrt{x} \tag{2}$$

where x is the depth achieved in μm, z is the atomic number of the incident particle, Z and A are the atomic number and the atomic weight of the target, respectively.[3,4] Straggling is then the same for all particles of equal z (deuteron or proton, α or ^{3}He). The dependence of straggling on the energy of the particle is

Table 2 Range of light ions in materials in mg cm^{-2}

Energy (MeV)	C	Al	Cu	Ag	Au
			Protons		
0.2	3.1 (E − 1)	5.4 (E − 1)	9.4 (E − 1)	1.3 (E 0)	1.8 (E 0)
0.5	9.6 (E − 1)	1.5 (E 0)	2.6 (E 0)	3.7 (E 0)	5.2 (E 0)
1	2.7 (E 0)	3.9 (E 0)	6.4 (E 0)	8.8 (E 0)	1.3 (E + 1)
2	8.4 (E 0)	1.1 (E + 1)	1.7 (E + 1)	2.3 (E + 1)	3.2 (E + 1)
3	1.7 (E + 1)	2.2 (E + 1)	3.1 (E + 1)	4.0 (E + 1)	5.7 (E + 1)
5	4.1 (E + 1)	5.1 (E + 1)	7.0 (E + 1)	8.8 (E + 1)	1.2 (E + 2)
7	7.3 (E + 1)	9.1 (E + 1)	1.2 (E + 2)	1.5 (E + 2)	2.0 (E + 2)
			Deuterons		
0.2	3.2 (E − 1)	5.8 (E − 1)	9.9 (E − 1)	1.3 (E 0)	1.8 (E 0)
0.5	7.9 (E − 1)	1.4 (E 0)	2.4 (E 0)	3.3 (E 0)	4.6 (E 0)
1	1.9 (E 0)	3.1 (E 0)	5.3 (E 0)	7.3 (E 0)	1.0 (E + 1)
2	5.4 (E 0)	7.9 (E 0)	1.3 (E + 1)	1.8 (E + 1)	2.5 (E + 1)
3	1.0 (E + 1)	1.4 (E + 1)	2.2 (E + 1)	3.0 (E + 1)	4.4 (E + 1)
5	2.4 (E + 1)	3.2 (E + 1)	4.7 (E + 1)	6.2 (E + 1)	8.8 (E + 1)
7	4.4 (E + 1)	5.6 (E + 1)	7.9 (E + 1)	1.1 (E + 2)	1.4 (E + 2)
			^{3}He		
0.2	2.2 (E − 1)	3.8 (E − 1)	6.1 (E − 1)	7.9 (E − 1)	1.0 (E 0)
0.5	3.3 (E − 1)	5.7 (E − 1)	9.6 (E − 1)	1.3 (E 0)	1.7 (E 0)
1	5.7 (E − 1)	9.4 (E − 1)	1.6 (E 0)	2.2 (E 0)	3.0 (E 0)
2	1.2 (E 0)	1.9 (E 0)	3.2 (E 0)	4.3 (E 0)	6.1 (E 0)
3	2.2 (E 0)	3.2 (E 0)	5.1 (E 0)	6.9 (E 0)	9.9 (E 0)
5	4.8 (E 0)	6.6 (E 0)	1.1 (E + 1)	1.3 (E + 1)	1.9 (E + 1)
7	8.3 (E 0)	1.1 (E + 1)	1.6 (E + 1)	2.1 (E + 1)	3.0 (E + 1)
			^{4}He		
0.2	2.6 (E − 1)	4.5 (E − 1)	7.3 (E − 1)	9.3 (E − 1)	1.2 (E 0)
0.5	3.7 (E − 1)	6.5 (E − 1)	1.1 (E 0)	1.4 (E 0)	1.9 (E 0)
1	5.9 (E − 1)	9.9 (E − 1)	1.7 (E 0)	2.3 (E 0)	3.1 (E 0)
2	1.2 (E 0)	1.8 (E 0)	3.1 (E 0)	4.2 (E 0)	6.0 (E 0)
3	1.9 (E 0)	2.9 (E 0)	4.8 (E 0)	6.6 (E 0)	9.3 (E 0)
5	4.0 (E 0)	5.8 (E 0)	9.1 (E 0)	1.2 (E + 1)	1.7 (E + 1)
7	6.9 (E 0)	9.5 (E 0)	1.4 (E + 1)	1.9 (E + 1)	2.7 (E + 1)

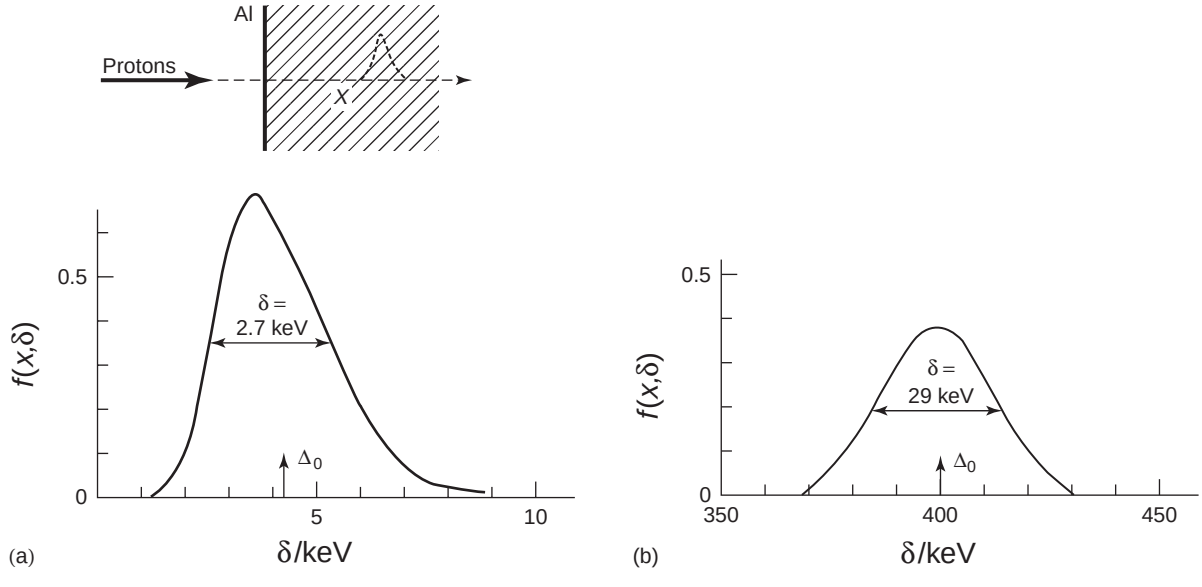

Figure 1 Energy straggling of 1 MeV protons for an energy loss of 4.2 keV (a) and 400 keV (b) after crossing a thickness of 2.4×10^{-5} g cm^{-2} (a) or 2×10^{-3} g cm^{-2} (b) in Al. (Adapted from G. Deconninck,[2] page 72.)

totally included in the value of x. For compound materials, Z and A are simply replaced by the mean values of Z and A.

The interaction of the incident particle with the atomic electrons of the target induces ionization. When ionization takes place in an inner electronic shell, subsequent rearrangement in order to return to atomic equilibrium leads to the emission of characteristic X-rays: the analytical technique based on this atomic process is called PIXE (particle-induced X-ray emission) (see **PIXE (Particle-induced X-ray Emission)**). PIXE is a very sensitive method for bulk and trace analysis of elements with an atomic number Z greater than 20.[5] PIXE may be also used to study elements from Na to K, but the absorption of low energy characteristic X-rays (1–3 keV) from these light elements into thick samples generally restricts the potential of the method. Analysis of low Z elements would then only be satisfactory for thin samples or to probe these elements in a very narrow layer at the surface of a bulk material.

When the incident particle passes close to the atomic nucleus, it may be elastically scattered. At very low impact parameter, a particle of mass m lower than the mass, M, of the collided nucleus, is scattered at an angle close to 180°. The spectroscopy of this backscattered particle (RBS) may easily be used for depth profiling of elements up to thicknesses close to 30% of their own range. The energy of the scattered particle at an angle close to 180° is related to the mass of the collided nucleus, Equation (3)

$$E_{\text{scatt}} = \left(\frac{M-m}{M+m}\right)^2 E_{\text{i}} \qquad (3)$$

and to the depth in the material at which the scattering took place. This depth can be determined using the stopping power of the target material. Furthermore, as the cross-section for such elastic scattering is proportional to the Z^2 of the target atoms, RBS is particularly powerful for the study of high Z elements (see **Rutherford Backscattering Spectroscopy**).

The probability of coulomb excitation of a target nucleus leading to the emission of a characteristic γ-rays[6] is much lower than for the backscattering and inner shell ionization phenomena. Consequently, PIGE (particle-induced γ-ray emission) is of particular interest for the study of light elements (see **Particle-induced γ-Ray Emission**) where PIXE and RBS suffer from a lack of sensitivity and accuracy. For these light target elements, the characteristic X-rays are of such low energy that they are absorbed into the material itself and RBS offers a much lower cross-section (because it is proportional to Z^2). When using heavy projectiles, the forward elastic scattering of light target nuclei give rise to another method of analysis called ERD (see **Elastic Recoil Detection Analysis**). If the energy of the incident particle is maintained below 5–10 MeV, no delayed radioactivity may be induced as would be the case in CPAA (charged particle activation analysis) typically performed at incident energies greater than 10 MeV (see **Charged Particle Activation Analysis**). By using all these analytical techniques simultaneously (or sequentially), powerful characterization of many materials may be achieved. In sections 5.3 and 5.6 of this article, we will discuss several complementary applications of nuclear reactions with other ion beam analysis (IBA) techniques.

2 FUNDAMENTALS OF NUCLEAR REACTIONS

In discussing NRA we will concentrate on prompt reactions leading to a transmutation of the target atom and then excluding RBS (see **Rutherford Backscattering Spectroscopy**), ERD (see **Elastic Recoil Detection Analysis**) and part of PIGE (see **Particle-induced γ-Ray Emission**) where the target nucleus does not change during the interaction. We will also exclude delayed γ-ray emission from a transmuted nucleus (see **Charged Particle Activation Analysis**).

The concise form of a nuclear reaction as expressed by nuclear physicists is

$$\text{A (a, bc) B}$$

where A is the target nucleus, a the incident projectile, b (and possibly c) the emitted particles or photons, B the residual nucleus which is generally a heavy particle carrying a very low energy which remains in the target material and cannot be detected.

In this respect, a reaction named

$$^{27}\text{Al(p, }\gamma)^{28}\text{Si}$$

would represent the capture of a proton in an ^{27}Al nucleus leading to the formation of an excited ^{28}Si which immediately decays to a stable ^{28}Si by the emission of a cascade of γ-rays. The spectroscopy of these γ-rays may be used as analytical signals.

The reaction

$$^{15}\text{N(p, }\alpha\gamma)^{12}\text{C}$$

would be induced by protons on ^{15}N leading to the production of a ^{12}C nucleus in an excited state after the emission of an α-particle. The spectroscopy of these γ-rays emitted by the excited ^{12}C or/and the spectroscopy of the emitted α-particle could be used for analytical purposes.

The interactions of a loosely bound compound projectile (such as a deuteron) give rise to the emission of energetic protons or neutrons after stripping of the incident deuteron. The breakdown of the incident deuteron (stripping reaction) may be written as

$$^{12}C(d, p_0)^{13}C \quad \text{or} \quad ^{12}C(d, p_i\gamma)^{13}C^*$$

The former represents the stripping of an incident deuteron by a ^{12}C nucleus, leading to the capture of a neutron in ^{12}C to produce a ^{13}C nucleus in its fundamental state. The emitted proton has a maximum energy $E(p_0)$. Various other ways could leave the residual nucleus of ^{13}C in one of its ($i = 1, 2, 3, \ldots$) excited states. The corresponding energies $E(p_i)$ of the p_i protons are then lower than $E(p_0)$. The stored energy in the excited $^{13}C^*$ nuclei will shortly afterwards appear as a cascade of characteristic γ-rays of this residual nucleus of ^{13}C. The energy of the emitted protons $p_0, p_1, p_2, p_3, \ldots$ is quantized and the differences in their energies corresponds to the various quantum states of the residual nucleus. Several groups of protons of different energies, as well as γ-rays, may be used for analytical purposes. Stripping of a deuteron with the capture of the proton is also possible. A neutron is then emitted, but in practice the spectroscopy of these emitted neutrons is much less easy than the spectroscopy of charged particles or photons, so its use in NRA is very limited.

3 KINEMATICS OF NUCLEAR REACTIONS

A fundamental parameter of a nuclear reaction is the Q value. It represents the balance between the total rest mass of interacting particles and those of emitted ones, Equation (4):

$$Q = (m_A + m_a - m_B - m_b)c^2 \tag{4}$$

where c is the velocity of light in vaccum. A positive Q value is a characteristic of a nuclear reaction in which some mass is converted into kinetic energy of emitted particles. This Q value added to the incident particle energy E_a is shared by both outgoing particles B and b, Equation (5):

$$Q = E_b + E_B - E_a \tag{5}$$

The partition between E_B and E_b may be calculated by using conservation laws: conservation of total momentum because the interaction takes place without any influence of external forces; conservation of total mass–energy because the isolated system formed by the incident particle and the target nucleus is being considered. In terms of E_a and E_b, one finds Equation (6):

$$Q = E_b\left(1 + \frac{m_b}{m_B}\right) - E_a\left(1 - \frac{m_a}{m_B}\right) - \frac{2}{m_B}\left[E_a E_b m_a m_b\right]^{1/2}\cos\theta \tag{6}$$

When b is the lightest emitted particle, its energy E_b in the direction θ relative to the incident beam of particles a is given by Equation (7):

$$E_b^{1/2} = \frac{\cos\theta(m_a m_b E_a)^{1/2} \pm \{m_a m_b E_a \cos^2\theta + (m_B + m_b)[m_B Q + (m_B - m_a)E_a]\}^{1/2}}{m_B + m_b} \tag{7}$$

As $E_b^{1/2}$ must be a positive quantity, the plus sign must only be kept if the following condition is met:

$$(m_B + m_b)|m_B Q + (m_B - m_a)E_a| > 0$$

This is always the case for positive Q values. For negative Q values, the minimum E_a energy giving rise to the emission of b at $\theta = 0°$ is given by Equation (8):

$$E_{a\,\min} = \|Q\|\frac{m_B + m_b}{m_B + m_b - m_a} \tag{8}$$

For higher E_a energies, two different energies may be attributed to the same particle b emitted at the same angle θ but in a limited energy range of E_a and only for the ejection of b in the forward direction. We will ignore this special behavior which could be inconvenient for elemental analysis.

A list of selected reactions for the study of light isotopes ($A \leq 20$) is given in Table 3. Resonant reactions which may be used for depth profiling (see sections 5.1 and 5.2 below) are particularly interesting and a suggestion for the best E_a incident energies is given.

For hydrogen, helium and lithium isotopes in a material, the detection of coincidental b and B particles may sometimes be used to solve chemical interference problems (superposition of signals from particles emitted with the same energy but arising from different B nuclei). As E_b energies are also dependent on the emission angle θ, a specific geometrical assembly may be required for solving interference problems.

The geometry of a nuclear reaction used for analysis in thick target materials is given in Figure 2. The incident particle "a" hits the sample at point I, crosses a thickness $x/\cos\alpha_i$, losing some of its energy (mainly by interaction with atomic shell electrons) and induces a nuclear reaction at P. The emitted particle with an energy $E_b(\theta)$ to be calculated for $E_a = E_a(0) - \Delta E(x/\cos\alpha_i)$ is produced at P and has to cross the distance $x/\cos\alpha_o$ before emerging from the sample surface at O. The outgoing energy at O is equal to $E_b(\theta) - \Delta E_b\,(x/\cos\alpha_o)$.

Table 3 Useful nuclear reactions for the analysis of light elements (hydrogen to fluorine)[a]

Element	Reaction	Q (MeV)	Recommended energy for depth profiling (MeV)	Refs.
Hydrogen	$D(d,p)T$	4.033		8
	$D(^3He,p)^4He$	18.352	0.64	9
	$D(^3He,\alpha)^1H$	18.352	0.64	
Helium	$^3He(d,p)^4He$	18.352	0.40	9
	$^3He(d,\alpha)^1H$	18.352	0.40	
Lithium	$^6Li(p,\alpha)^3He$	4.02		10
	$^6Li(p,^3He)^4He$	4.02		
	$^6Li(d,\alpha)^4He$	22.374		11
	$^6Li(^3He,p)^8Be$	18.786		12
	$^7Li(p,\alpha)^4He$	17.347		13
Beryllium	$^9Be(p,\alpha)^6Li$	2.125	0.33	14
	$^9Be(d,t)^8Be$	4.496		15
	$^9Be(d,\alpha)^7Li$	7.153		16
	$^9Be(^3He,p)^{11}B$	10.322		17
	$^9Be(^3He,\alpha)^8Be$	18.912		18
	$^{10}Be(^3He,p)^{12}C$	19.693		12
Boron	$^{10}B(d,\alpha)^8Be$	17.818		19
	$^{10}B(\alpha,p)^{13}C$	4.063	1.567	20, 21
	$^{11}B(p,\alpha)^8Be$	8.582	0.163 and 0.650	22, 23
	$^{11}B(^3He,p)^{13}C$	13.185		24
	$^{11}B(^3He,d)^{12}C$	10.464		24
	$^{11}B(\alpha,p)^{14}C$	0.784	Higher than 2.5 MeV owing to low Q value	21, 25
Carbon	$^{12}C(d,p)^{13}C$	2.722		26
	$^{12}C(^3He,p)^{14}N$	4.779		27
	$^{12}C(^3He,\alpha)^{11}C$	1.856		28
	$^{13}C(d,p)^{14}C$	5.951		29
Nitrogen	$^{14}N(d,p)^{15}N$	8.610		30
	$^{14}N(d,\alpha)^{12}C$	13.574		31, 32
	$^{15}N(p,\alpha)^{12}C$	4.966	0.429	33
	$^{15}N(d,\alpha)^{13}C$	7.687		34
Oxygen	$^{16}O(d,p)^{17}O$	1.917		35
	$^{16}O(d,\alpha)^{14}N$	3.11		36
	$^{16}O(^3He,p)^{18}F$	2.032		27, 37
	$^{16}O(^3He,\alpha)^{15}O$	4.914		37
	$^{18}O(p,\alpha)^{15}N$	3.980	0.629	11, 38
	$^{18}O(d,\alpha)^{16}N$	4.247		38
Fluorine	$^{19}F(p,\alpha)^{16}O$	8.114	0.340; 1.347	39
	$^{19}F(d,p)^{20}F$	4.374		40
	$^{19}F(d,\alpha)^{17}O$	10.031		40
	$^{19}F(\alpha,p)^{22}Ne$	1.675	2.45	41

[a] For applications of these nuclear reactions to elemental analysis see Demortier.[60]

The angle θ is simply $\alpha_i + \alpha_o$. As the most intense fraction of the emitted particles are elastically scattered a particles (see **Rutherford Backscattering Spectroscopy**), some calibrated absorber has to be inserted between the surface sample and the detector in order to stop them all and then to accept only b particles induced by a true nuclear reaction, when the target nucleus A is transmuted to another one, B. Reactions with large Q values and detection of b particles of mass lower than the mass of a are particularly useful characteristics of nuclear reactions suitable for analysis. A thin absorber will stop all the scattered a particles but will not greatly affect the energy of the lighter emitted b particles (see Equation 1).

In special cases corresponding to a large Q value and a low E_a energy, one may expect to detect b particles selectively with a higher mass than the incident particle a even after crossing the dedicated absorber. Reactions $^6Li(d,\alpha)^4He$, $^7Li(p,\alpha)^4He$, $^9Be(^3He,\alpha)^8Be$, $^{10}B(d,\alpha)^8Be$,

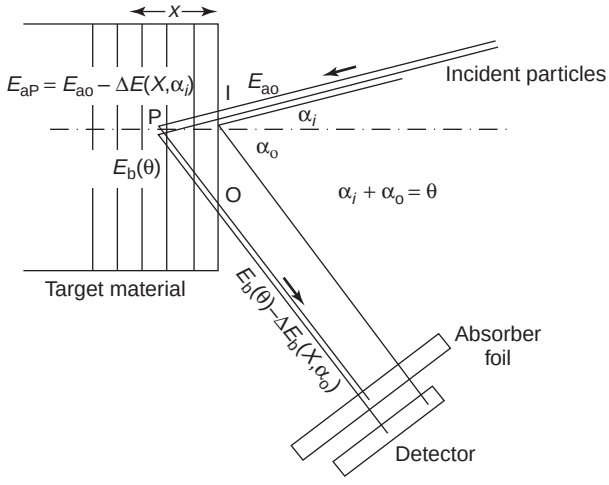

Figure 2 Parameters involved in the study of thick materials by the nuclear reaction A(a, b)B.

$^{14}N(d,\alpha)^{12}C$, and $^{19}F(p,\alpha)^{16}O$ shown in Table 3 would offer this possibility for an incident E_a energy in the region of 1 MeV, as would $^{11}B(p,\alpha)^{8}Be$, $^{18}O(p,\alpha)^{15}N$, $^{19}F(p,\alpha)^{16}O$, when using incident protons in the range of a few hundred kiloelectron volts.

The energetic scale of the compound nucleus C* made up with all the nucleons of A and a (and then also B and b) may be represented schematically (Figure 3).

Quantum states have their own energy width $\Gamma = \Delta E$ (and their associated lifetime ΔT related by $\Gamma \Delta T \geq \hbar$). Low-lying states (bound states) can only decay through γ-ray emissions: their width is of the order of 10^{-3} eV. At much higher excitation energies, known as positive energy states, the width is typically of the order of 0.1–100 keV. These levels may be formed by several typical configurations of nucleons, in particular (A + a), (B + b) or (C_i^*). Each of these configurations is associated with a typical width Γ_a, Γ_b, Γ_γ. The sum of all possible configurations Γ_i is related to the total width Γ: the higher Γ, the shorter the lifetime of the level.

When the incident particle a is sent on the nucleus A with a precise energy E_{ar}, the probability of forming an excited nucleus is greater than for any other E_a energy slightly greater or lower than E_{ar}. For particles of energy E_{ar}, the probability of the emission of a particle b (and/or a cascade of characteristic γ-rays) is also enhanced.

Owing to competing phase shifts in the quantum mechanical behavior of nuclear reactions, the cross-section may sometimes be very low in the vicinity of a resonant energy. The discontinuities in the cross-section values of nuclear reactions induced by a on A give rise to some potential for depth profiling of A nuclei: signals induced in the detector by b or c particles will be particularly intense when arising from those regions in the target material where $E_a = E_{ar}$. If a material is irradiated with particles a of energy greater than E_{ar}, the region where E_a will reach this particular energy E_{ar} will be achieved at a defined depth below the surface. This depth may be calculated by using the $S(E_a)$ values of

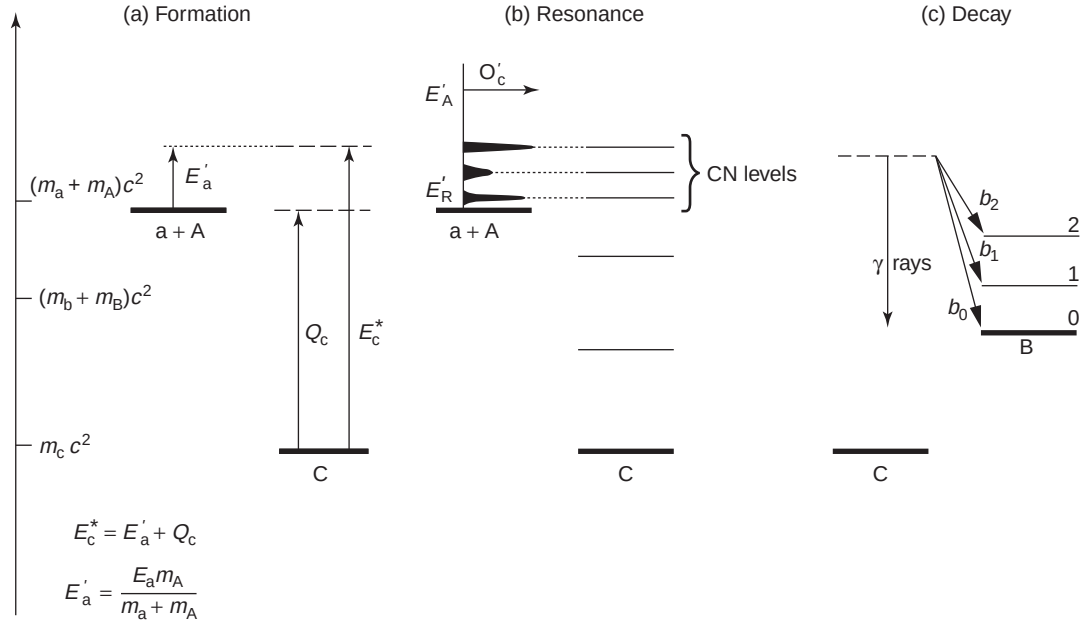

Figure 3 Mass–energy scheme of a resonant nuclear reaction. The formation in various levels of an excited compound nucleus (CN) (b) by the interaction of an incident particle a on a target nucleus A takes place only for discrete incident E'_a energies (E'_a calculated in the center of mass system). Particles b of various energies (b_0, b_1, ...) are consequently emitted to give rise to a final nucleus B which could appear in various states (c). (From G. Deconninck, LARN.[2])

Table 1 and the corresponding detected E_b values may be calculated by using Equation (7) and the necessary correction for the energy loss in the outgoing direction (see Figure 2) which is to be calculated by using the corresponding $S(E_b)$ values of Table 1.

The depth resolution is directly related to the total width of the resonance: the narrower the resonance, the more precise will be the depth resolution. Useful resonant energies E_{ar} for depth profiling are limited for several reasons:

1. The cross-section at the resonant energy must be intense in comparison with its value outside the resonance.
2. The resonant energy must be well isolated in order to avoid superposition of signals originating from various regions below the sample surface where several resonant energies may be reached.
3. The resonance width must be of the order of (or less than) the energy resolution of the whole experimental set-up (taking into account the accelerator energy resolution, detector resolution, detector solid angle accepting particles emitted with various θ and then various energies ...).
4. The energy straggling of the incident particles into a buried layer under the surface must be sufficiently small, as should be the energy straggling of the emitted particles along its outgoing path and into the necessary absorber (inserted between the target and the detector to stop elastically scattered a particles).

The detection of emitted γ-rays from the residual nucleus does not suffer from the last uncertainty (straggling in energy of the emitted particle) but γ-rays may be affected by Doppler shift when the emitting residual nucleus decays before being stopped in the material. Doppler broadening of emitted γ-rays only arises if the lifetime of the recoiling C^* or B^* nuclei is shorter than the time required to stop this recoil nucleus in the sample target.

4 IDENTIFICATION OF SIGNALS FOR ANALYTICAL PURPOSES

Modern particle detectors of protons, deuterons, ^{3}He and α-particles with energies in the megaelectron volt range are now exclusively solid-state detectors: p–n junctions made with crystals of silicon.[7]

The full energy of the detected particle is converted into pairs of electrons and holes: the number of pairs is strictly proportional to the deposited energy. A reverse polarization of the p–n junction is maintained by an adjustable bias across the desired detector depth, which is at least the range of the particles of maximum energy to be detected. This range may vary from 5 μm for 1-MeV α-particles to 0.6 mm for 7-MeV protons. The energy resolution (width at half maximum of a peak induced by monoenergetic particles) is generally in the order of 8–15 keV for the whole range of these useful energies.

For the highest energies, a succession of detectors mounted as a telescope may be necessary. The associated electronics of these telescopes is tailored in order to collect all the signals induced in the successive detectors coincidentally. For energetic particles, the time delay between the signals produced in two distant thin detectors can be taken into account by measuring the time between the two short signals; in this last case, the time-of-flight of the particle is used. For distant detectors, this time interval is used to identify the mass of the particle only. A third detector, or a set of successive detectors which stop the particle completely, is necessary to determine the energy of the charged particle. The resolution of time-of-flight systems is indeed insufficient to measure the energy of the energetic b particles accurately.

The kinetic energy (converted into a proportional signal in the solid-state detector and possibly the time-of-flight) is a characteristic quantity which allows experimentalists to identify the nucleus which was participating in the nuclear reaction. If this nuclear reaction took place at the surface of the sample and if the emitted particle were directly collected in the backward direction, the energy of this detected particle would be simply given by Equation (7), where E_a is the incident energy delivered by the particle accelerator. In contrast if the reaction took place below the surface, the incident particle reaches this region with a lower energy $E'_a < E_a$. E'_a can be calculated by using the target stopping power along the entrance pathlength $x/\cos\alpha_1$ and E'_b can be recalculated using the same Equation (7). The emitted b particle would also lose some of its energy along the outgoing trajectory at a distance $x/\cos\alpha_o$ which can also be calculated using the appropriate stopping power of the material for this particle b. Dividing the whole distribution of signals from $E_{b\,max}$ to $E_{b\,min}$ into energy intervals with a width of the order of the energy resolution of the experimental set-up (including energy straggling for both incident and emitted particles), it is possible to determine the depth profile of thicknesses ranging from 10 nm to a few microns. The experimentalist must have the ability to certify the origin of signals and to exclude eventual interfering elements (see Figure 4).

Figure 4 shows that particles with identical energies may be emitted with the same incident energy E_a and at the same angle θ by two different A nuclei. Changing

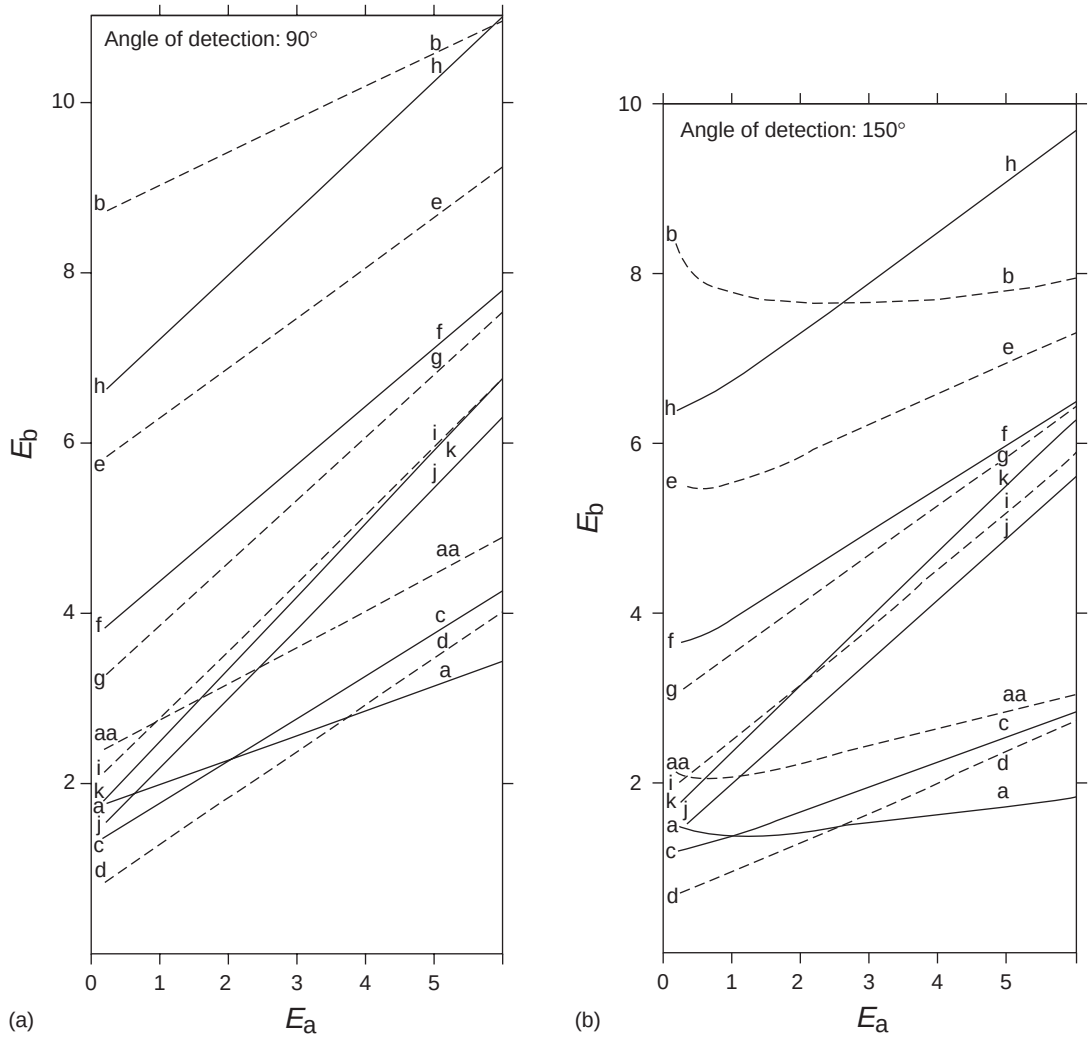

Figure 4 (a) Relation between the energy E_b of emitted α-particles at $\theta = 90°$ and the incident energy E_a for the most useful nuclear reactions induced by protons on ^{6}Li(a), ^{7}Li(b), ^{9}Be(c), ^{10}B(d), ^{11}B(e), ^{15}N(f), ^{18}O(g), ^{19}F(h), ^{23}Na(i), ^{27}Al(j) and ^{31}P(k). Line aa refers to reaction ^{6}Li(p,^{3}He)α which appears simultaneously with reaction ^{6}Li(p,α)^{3}He. (b) Relation between the energy E_b of emitted α-particles at $\theta = 150°$ and the incident energy E_a for the most useful nuclear reactions induced by protons on ^{6}Li(a), ^{7}Li(b), ^{9}Be(c), ^{10}B(d), ^{11}B(e), ^{15}N(f), ^{18}O(g), ^{19}F(h), ^{23}Na(i), ^{27}Al(j) and ^{31}P(k). Line aa refers to reaction ^{6}Li(p,^{3}He)α which appears simultaneously with reaction ^{6}Li(p,α)^{3}He.

the energy E_a or the angle of detection are alternatives that may be used to deal with the interferences. It may be helpful to use two detectors at two different angles simultaneously for complicated analyses: interferences between reactions at two different A nuclei do not take place at both angles (see Figure 4a and b).

In addition to the measurement of the energy of b particles which identify the target nuclei, the measurement of the intensity of signals with identical energies gives the concentration of these nuclei of interest. The intensity of such signals is governed by the cross-section of the particular nuclear reaction: the greater the cross-section, the better the capability of the nuclear reaction to perform accurate (due to statistics) and sensitive elemental determinations. Knowledge of all characteristics of the nuclear reaction (presence of resonances, angular distributions of emitted particles and elimination of interfering reactions by selecting the best angle of detection) is of prime importance to achieving confident analyses.

All these data may be found in various nuclear data tables[42–45] and handbooks.[46,47]

For γ-ray detection, large volume (40–100 cm^3) Ge(Li) photon detectors of high energy resolution (1.2 keV at 1.5 MeV) are used. They are cooled at liquid nitrogen temperature. Various processes of interaction (photoelectric effect, Compton effect, pair creations) give rise to an extended spectrum from which only the full

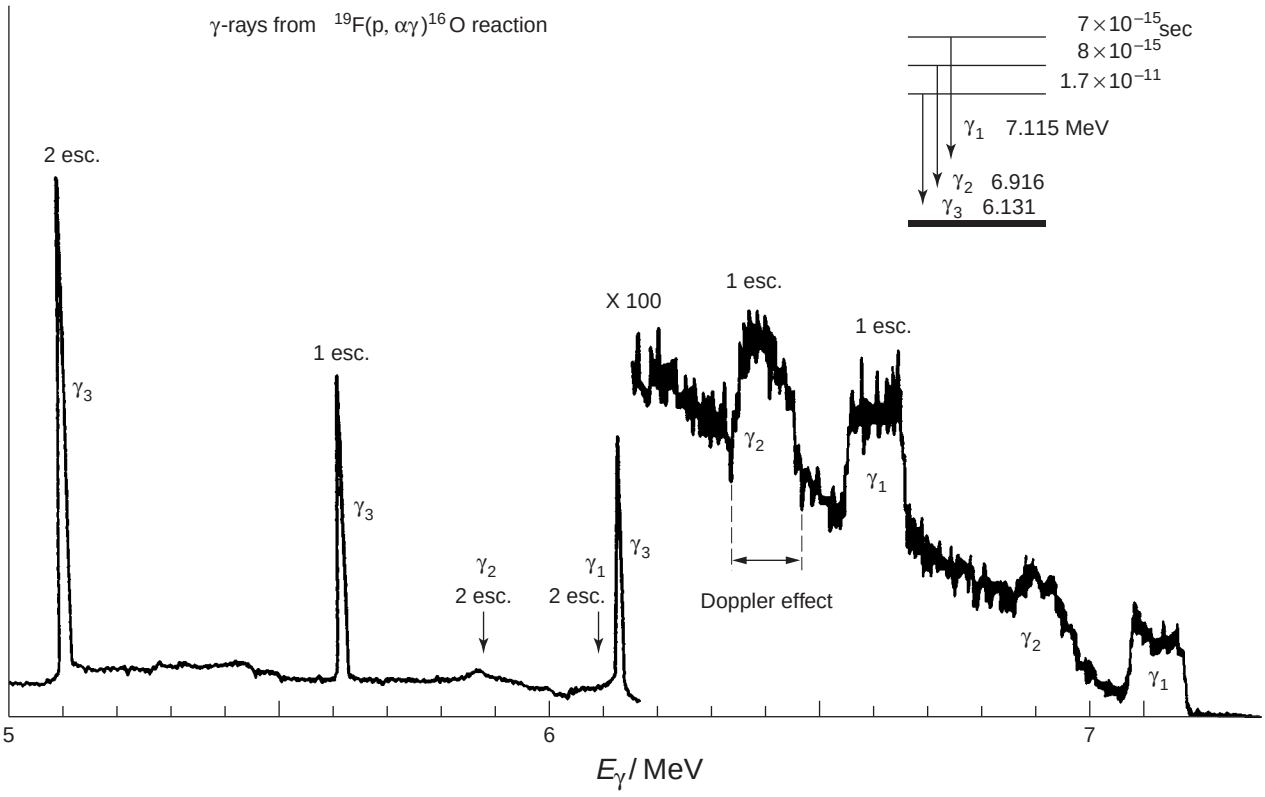

Figure 5 High-energy part of a γ-ray spectrum obtained with a solid-state Ge(Li) detector during the irradiation of a material containing fluorine. The γ_1, γ_2 and γ_3 peaks are produced by (p,αγ) reaction leading to excited ^{16}O. Both γ_1 and γ_2 peaks are broadened by the Doppler effect because of the short lifetime of the corresponding levels of the oxygen nucleus which decays before being at rest in the material. The sharp γ_3 peaks do not suffer this broadening. Each characteristic γ-ray gives rise to signals in three regions: the full energy peak and two escape peaks corresponding to the loss of one or two of the annihilation photons (511 keV) produced after the initial pair creation. (From G. Deconninck, LARN.[2])

energy peak is selected for γ-ray identification. The efficiency of such a detector is quite high for γ-rays of 2 MeV or less but may also be used at higher energies (Figure 5).

An alternative would be a large NaI or BGO (bismuth germanate oxide) detector (about 1000 cm^3) with a very high detection efficiency but with a much lower energy resolution. This kind of detector is particularly useful for depth profiling using (p,γ) or (p,αγ) resonant reactions involving γ-ray energies of more than 2 MeV. In contrast to particle spectroscopy whose energy is a complicated function depending on $E_a, \theta, x/\cos\alpha_1$ and $x/\cos\alpha_2$, γ-ray spectroscopy only concerns measurement of the energy of the excited state of the promptly decaying excited B^* nucleus. To profile one element which could give rise to a resonant nuclear reaction (see Table 3) at a particular E_{ar} energy, the E_a energy is progressively increased to allow this resonant reaction to take place at various depths below the surface. For increasing values of $(E_a - E_{ar})$ the resonant reaction would produce γ-rays at depths calculated using the stopping power (Table 1) of the incident particle a only.

Examples of depth profiling using γ-ray or particle spectroscopy will be given in sections 5.1 and 5.6 of this article.

The analytical investigation of large specimens or fragile materials which cannot be introduced in the vacuum sometimes requires the use of external beam geometry. This technique is now widely familiar to PIXE (see **PIXE (Particle-induced X-ray Emission)**) and PIGE users (see **Particle-induced γ-Ray Emission**) but may be also tailored for NRA when the outgoing particle has a sufficiently high energy (i.e. for (d,p) reactions) by comparison with the energy of the incident particle. The incident beam crosses a thin foil of metal or of mylar before reaching the sample situated outside the vacuum vessel of the accelerator. The distance in the air between the target and the detector may be finely tuned in order to stop the elastically backscattered a particles completely and allow only b particles to reach this detector. The energy straggling of this incident particle in the exit window and in the air, and the energy straggling of the emitted particle in the sample and also in the air does not allow, in this particular case, depth profiling to be performed.

5 EXAMPLES OF APPLICATION

Many laboratories in the world use NRA in the field of materials science. We have selected a few typical results. The first concerns the study of hydrogen distribution in the depth of materials by using a methodology involving the detection of γ-rays produced by irradiation of energetic ^{15}N ions to induce a resonant nuclear reaction at various depths in the sample. General problems related to energy straggling, cross-sections and performances of γ-ray detectors will be discussed.

The second results concern the depth profiling of F and Na in SnO_2 coated glass (an insulating material) using simultaneous RBS and NRA techniques at two particular resonant proton energies for (p,$\alpha\gamma$) reactions. The third application benefits from simultaneous PIXE and NRA for a complete characterization of all the components of a YBaCuO superconductor using a microbeam of low energy deuterons.

We will then present the use of a (d,p) reaction for the study of microregions in the grain boundaries of Au-Si alloys and for the determination of traces of sulfur in ancient gold jewellery artefacts. Nuclear reactions are indeed not sensitive to high Z nuclei owing to the high coulomb barrier, but they offer a unique technique (free of interference) for studying light elements in heavy matrices. The last application concerns the determination of all the major and minor elements and a majority of trace elements (in the region of a few micrograms per gram) in a reference material of biological interest.

5.1 Analysis of Hydrogen in Various Materials

The use of megaelectron volt ion beams to analyze materials for hydrogen has been more interesting and more important than even the early researchers in this field realised. The reasons for this are simple:

1. Hydrogen is probably the most common contaminant element, especially in thin film materials.
2. Hydrogen has important effects on the chemical, physical and electrical properties of many materials.
3. The analysis for hydrogen is difficult or impossible by most traditional analytical methods.

Megaelectron volt ion beams to probe for hydrogen may be used in two different approaches. The first is NRA, in which megaelectron volt ions bombard a sample, inducing nuclear reactions between the incident ion and hydrogen in the target. The second is ERD in which megaelectron volt ions bombard a sample and through elastic collisions cause hydrogen ions to recoil out of the sample with the number of these recoils used to determine the amount of hydrogen in the target. This last problem (of ERD) is treated in **Elastic Recoil Detection Analysis**. The basic principle for hydrogen analysis using NRA is to use nuclear reactions between protons and light nuclei but in the reverse direction: light nuclei are incident particles and hydrogen is the collided nucleus.

While many nuclear reactions are possible for hydrogen analysis, the reactions most commonly used to probe for ^{1}H in material are those induced by ^{15}N and ^{19}F.[43] These reactions are similar in many ways: they both have low-energy isolated resonances above a few megaelectron volts. Another reaction which is capable of profiling much deeper into samples uses a ^{7}Li beam. The important parameters for most useful reactions for H analysis are given in Table 4.

The use of the ^{15}N reaction as a probe for hydrogen is given here as an illustration of this approach. The nuclear reaction is:

$$H(^{15}N, \alpha\gamma)^{12}C$$

The very narrow and intense resonance is shown in Figure 6. The γ-ray intensity at the resonant energy is four orders of magnitude higher than the one off-resonance.

A beam of ^{15}N ions is incident on the sample being analyzed and the yield of characteristic γ-rays of ^{12}C is measured. Because this is a resonant reaction, the cross-section is large at the resonance energy (E_{res}) and small off-resonance. Therefore, if a sample is bombarded with ^{15}N ions at the resonance energy, the yield of γ-rays is proportional to the hydrogen only present at the surface

Table 4 Nuclear reaction cross-section characteristics

Reaction	$^{7}Li + H$ (ref. 49)	$^{15}N + H$ (ref. 50)	$^{15}N + H$ (ref. 50)	$^{19}F + H$ (ref. 50)	$^{19}F + H$ (ref. 50)
Resonance energy (MeV)	3.07	6.385	13.35	6.418	16.44
Cross-section (σ_0) at resonance (mb)	4.8	1650	1050	88	440
Resonance width (Γ) (keV)	81	1.8	25.4	44	86
$\sigma_0\Gamma$(mb keV)	389	2970	26 700	3870	37 800
Relative yield	0.13	1.000	9.0	1.3	12.7
Energy of next resonance (MeV)	7.11	13.35	18.0	9.1	17.6
γ-Ray energy (MeV)	17.7, 14.7	4.43	4.43	6.13, 6.98, 7.12	6.13, 6.98, 7.12

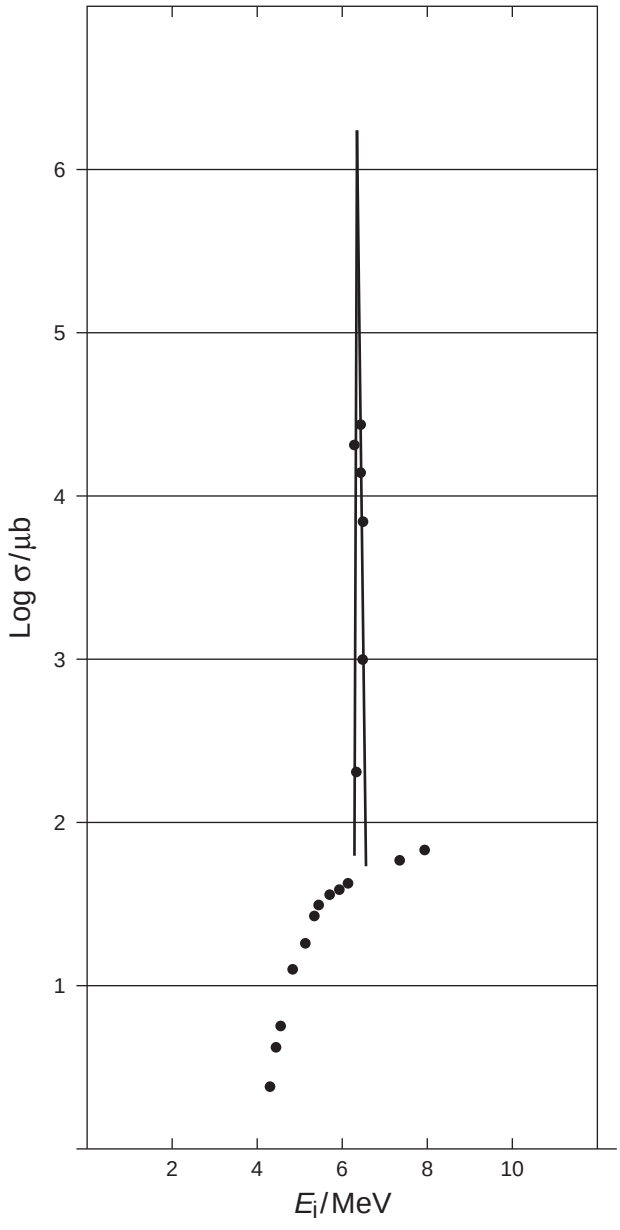

Figure 6 Variation of the cross-section of the reaction $H(^{15}N,\alpha\gamma)^{12}C$ for increasing ^{15}N incident ions. The cross-section increases by at least four orders of magnitude at the resonance energy ($E_i = 6.385$ MeV). (Data are taken from Horn and Lanford.[51])

of the sample. If the beam energy is raised above the resonant energy, there is no longer a reaction with surface hydrogen (because the ^{15}N ions are above the resonant energy) but as the ions lose energy when penetrating the sample, they reach the resonant energy at a particular depth: the yield of γ-rays is then proportional to the hydrogen content at this depth.

Characteristic γ-rays from the reaction are detected by a large (and therefore efficient) scintillation detector (NaI), or a BGO detector, located close behind the irradiated samples.

The γ-ray yield is proportional to the energy integrated area under the Breit–Wigner formulae for a resonant cross-section, i.e. proportional to $\sigma_0\Gamma$, where σ_0 is the cross-section at the resonance energy and Γ is the width of the resonance. The yield of γ-rays is also inversely proportional to the energy loss of the incident ions. This is because the yield of γ-rays is proportional to the amount of H per cubic centimeter within the resonance detection window (i.e. when the energy is on-resonance) and this is proportional to the thickness of the resonance detection window. This thickness is inversely proportional to dE/dx.

As in all resonant nuclear reaction profiling, a measurement of the number of γ-ray signals induced by the nuclear reactions versus incident energy is made. A known standard material containing a known quantity of hydrogen at a well-defined depth is used for quantitative determination of depth profiles in the material to be studied.

The matrix of this reference material may be different from the matrix of the material under investigation. All the matrix effects are contained in dE/dx which is matrix-dependent.

The depth resolution (Δx) of the method depends on the measured full width at half maximum (fwhm) of the hydrogen profile, for an infinitely thin layer of hydrogen at a certain depth in the sample. This is then determined by the width of the reaction resonance (Γ) and the fluctuations in the energy (ΔE) of the bombarding ions at this depth in the sample. ΔE has contributions from the energy spread in the incident proton beam (ΔE_b), which depends on the performance of the accelerator, energy straggle (ΔE_s) (see Equation 2) and, for detection of γ-rays, the Doppler energy broadening (ΔE_D). Since in most cases these are all Gaussian distributions, the total energy width is obtained by adding them in quadrature, Equation (9):

$$\Delta E^2 = \Delta E_b^2 + \Delta E_s^2 + \Delta E_D^2 \qquad (9)$$

For the depth profile at the extreme surface, the energy straggling is not Gaussian and another more sophisticated procedure is involved.[52] Some additional explanation is given for Doppler effect. Hydrogen atoms in materials are not stationary but vibrate. For a narrow nuclear resonance, this motion can shift the reaction on- and off-resonance. For example, the width of the ^{15}N resonance is 1.8 keV, whereas the Doppler broadening is in the range 5–15 keV, depending on the strength of the hydrogen bond. This contribution is nevertheless usually small compared to straggle, except when analyzing very thin hydrogen surface layers where the Doppler effect is the dominant source of broadening.

The ideal nuclear reaction for profiling would be one with an infinitely narrow resonance reaction with a zero cross-section off-resonance. The ^{15}N nuclear reaction is nearly this ideal. The off-resonance cross-section is more than four orders of magnitude smaller than the peak resonance cross-section. For other resonance reactions, the off-resonance cross-sections are generally much larger.

However, even in this favorable case, there are situations where the nonzero off-resonance cross-section needs to be taken into account. Thus, a layer of TiH_2 on clean Si was profiled.[51] The results of this profile are shown in Figure 7. The counts above 7.2 MeV essentially all come from the off-resonance cross-section.

The correction for such off-resonance effects is simple. Since the off-resonance yield is proportional to the on-resonance energy γ-ray yield (integrated between 6.4 and 7.2 MeV in Figure 6) times the off-resonance cross-section, once the proportionality constant is known, these off-resonance counts are subtracted from the total counts. In the case of ^{15}N NRA, this correction is Equation (10):

$$Y_{\text{off res}} = (1.28 \times 10^{-5})I \tag{10}$$

where I is the energy integrated on-resonance yield (in counts keV).[51]

This correction is simply due to the fact that no other resonance at higher energy (the next one being at $E_{\text{inc}} = 18$ MeV) in the $^{15}N + p$ reaction gives a contribution.

The problem of sensitivity or limit of detection is discussed now. When making NRA measurements where the reaction product is a high-energy γ-ray, the detectors are also sensitive to background radiations, such as cosmic rays and, in some cases, accelerator-generated radiation. This background is the main factor in the limitation of the sensitivity of NRA. There are several ways in which sensitivity can be improved, the first being by high beam current. Since the external background is independent of the beam current whereas the real count rate increases linearly with this beam current, sensitivity improves linearly with beam current. However, this approach can be used only up to the limit of the beam current available in a particular accelerator or the limit of beam current

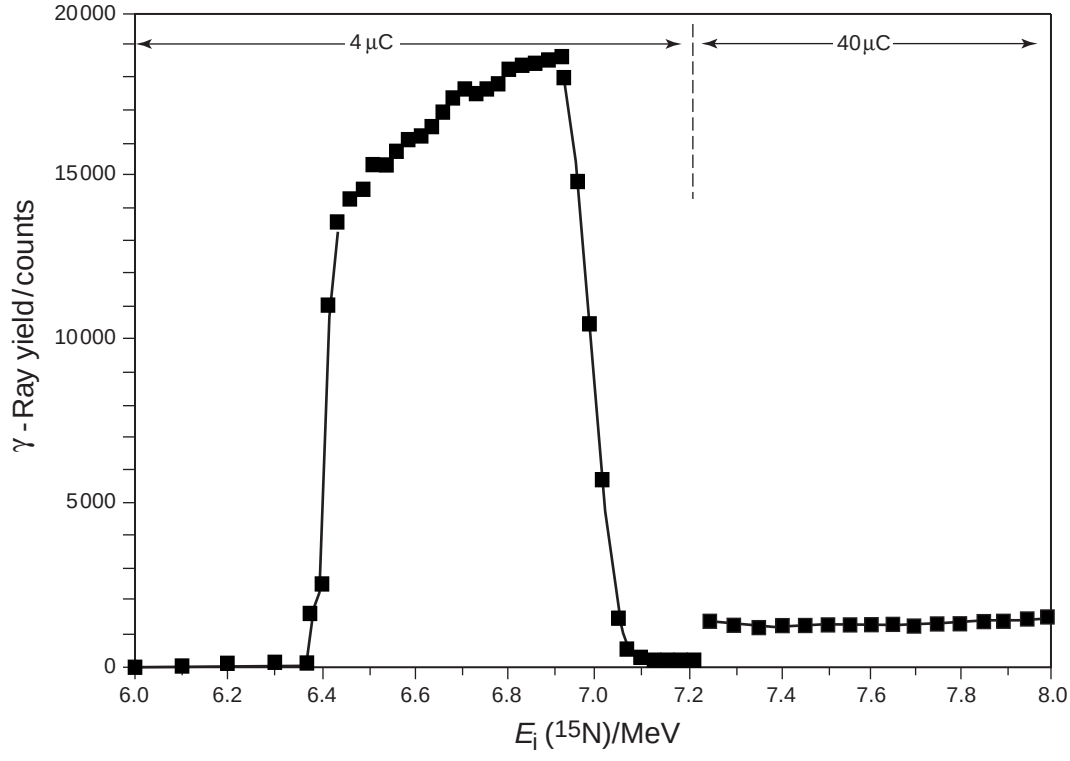

Figure 7 Use of the resonant energy of ^{15}N ions for hydrogen depth profiling. The sample is a coating of TiH_2 on a Si substrate. The sharp increase in the γ-ray yield appears when the incident ^{15}N energy reaches the resonant energy: the production of γ-rays takes place at the sample surface. When the ^{15}N ions incident energy exceeds 7 MeV, the resonant energy is only reached in the Si substrate. At energies higher than 7.2 MeV, the number of incident ^{15}N ions was increased by a factor of ten to appreciate the continuum (out-resonance) counting rate. (Reprinted from W. Lanford, 'Analysis for Hydrogen by Nuclear Reaction and Energy Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 65–82. Copyright (1992), with permission from Elsevier Science.[48])

that a particular sample will withstand. The last limitation may be avoided by rastering the beam along a large area on the surface for samples where the surface homogeneity is guaranteed.

Another approach is to build material and electronic shields to reduce the background in the γ-ray detectors. This approach is now much easier with the advent of bismuth germinate BGO scintillation detectors which are progressively replacing the usual NaI scintillator. BGO detectors are much more efficient (for a given size) than NaI and are much less sensitive to neutrons generated by cosmic rays.

In good cases, sensitivities of the order of micrograms per gram and depth resolutions of a few nanometers are possible. However, the practical limit achievable is dependent on the nature of the sample being analyzed and on the quality of the accelerator (energy resolution, beam current intensity, etc.) and its associated equipment.

Historically, NRA has been used in many studies of hydrogen in materials. However, because of the wide availability of new facilities for He ERD and because of the potential of high-energy heavy ion ERD, this alternative method is becoming more popular (see **Elastic Recoil Detection Analysis**).

Nuclear reactions may also be used for the study of the other isotopes of hydrogen. Some of the reactions that have been used include: ^{2}H(d,p)T, ^{2}H(d,n)^{3}He, ^{3}He(d,p)^{4}He, ^{3}H(d,n)^{4}He and ^{3}H(p,n)^{3}He. For all reactions giving rise to neutrons, the analytical technique involves the detection of the associated charged particle.

The most widely used reaction for deuterium analysis is the D(^{3}He,p)^{4}He reaction[53] which has a positive Q-value (18.35 MeV) making identification of the reaction products easy and the elimination of interfering signals comfortable using a coincidence technique with the detection of both particles simultaneously emitted: p and ^{4}He. Depth information is obtained by the energy loss of the outgoing ^{4}He particles. The cross-section for this reaction is not large (70 mb at 0.75 MeV) compared to the elastic cross-section (see **Elastic Recoil Detection Analysis**), so count rate problems are of concern.

5.2 Depth Profiles of F, Na in SnO_2 Coated Glass

A typical application of simultaneous RBS and resonant NRA is the study of fluorine and sodium migration in a SnO_2 layer deposited on a glass substrate.[54] The depth profiles are measured by proton-induced resonant reactions. The depth calibration which is essential to localize the interface is done by proton backscattering on Sn. As the SnO_2 layer is thin (200 nm), the same stopping power may be used for incoming protons and outgoing elastically scattered protons on Sn, the heaviest element in the sample. The fwhm of the Sn signals is therefore twice the actual energy loss (ΔE_p) of the incoming protons in the SnO_2 layer. ΔE_p then gives the exact position of the interface of SnO_2 with the bulk glass in terms of energy, without using the stopping power parameters. RBS is therefore used to calibrate the resonant reaction excitation curves.

Profiling of sodium is carried out by using the ^{23}Na(p,$\alpha\gamma$)^{20}Ne at the resonance energy $E_R = 1010.5$ keV (resonance width $\Gamma_R < 0.5$ keV). γ-Rays of 1634 keV are detected in a 4 in $\times$ 4 in well NaI detector offering maximum efficiency. The target sample is irradiated into the well of the NaI detector. As it is essential to reduce the irradiation duration to prevent sodium migration induced by the beam, the beam is rastered on the sample in order to avoid local overheating of the glass. The experimental excitation curve is displayed in Figure 8(b). A small surface concentration is observed followed by a steep edge at $E_p = E_R + 14.6$ keV, and then a constant concentration of 16 atom% (the bulk concentration of Na in glass). The RBS spectrum was collected simultaneously. Figure 8(a) shows this RBS spectrum for the SnO_2 layer on glass under the same experimental conditions as used for the profiling, i.e. $E_p = 1.015$ MeV. The width $2\Delta E_p$ of the Sn signal as measured on the RBS spectrum is 29.2 keV (twice 14.6 keV) and can be related directly to the step in the Na depth profile. The exact localization of the concentration step with respect to the interface is therefore known directly, without any calculation using physical parameters such as the stopping power of the Sn and O mixture. Therefore, any other element that could be present in the layer and whose measurement is not accessible by either RBS or NRA methods will not disturb the measurement.

Profiling of fluorine is done using the ^{19}F(p,$\alpha\gamma$)^{16}O at resonance energy $E_r = 340$ keV ($\Gamma_R = 2.4$ keV). The γ-rays in the range of 6–8 MeV are also detected using the large NaI scintillator. The excitation curve is displayed in Figure 8(d) showing a shallow concentration followed by a huge peak localized in the SnO_2 layer, close to the interface with the glass substrate, and just in front of the sodium step. The energy loss $\Delta E'_p$ corresponding to the interface was, once again, measured by the backscattering spectrum which was collected simultaneously in the same way as in Figure 8(c). Once more, the width of the Sn peak is twice the energy loss of the protons measured by the new RBS spectrum. Using these two resonant nuclear reactions, it was then possible to localize the relative position of (i) the fluoride layer (completely in the SnO_2 surface layer), (ii) the interface between SnO_2 and the glass, and (iii) the sodium step which is located entirely in the glass substrate.

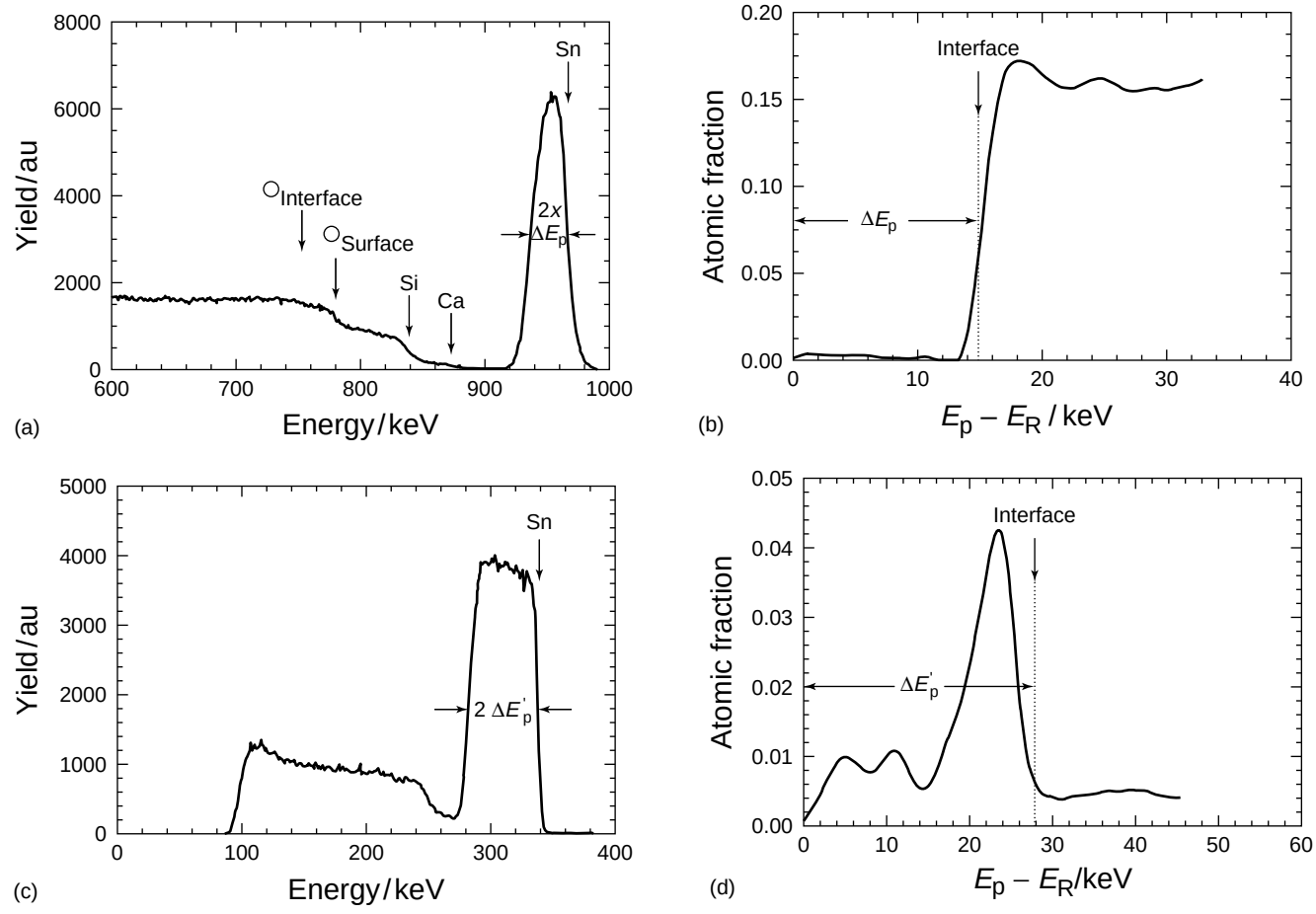

Figure 8 Determination of Na profile by the $^{23}Na(p,\alpha\gamma)^{20}Ne$ reaction around the resonant energy (1010.5 keV). The Sn thickness was determined at 1 MeV by RBS (a) and the step in the γ-ray production indicated that Na is only present in the glass substrate behind the SnO_2 coating (b). Similarly, the F profile is determined by the $^{19}F(p,\alpha\gamma)^{16}O$ reaction around the resonant energy at 340 keV. The corresponding RBS spectrum at 340 keV (c) gives the same SnO_2 thickness as in (a) and the F profile (d) indicates that this element is mainly present in the SnO_2 coating. (Reprinted from G. Terwagne, G. Deconninck, 'Concentration Profile of Light Elements Near an Interface', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 153–155, Copyright (1992), with permission from Elsevier Science.[54])

5.3 Stoichiometry of Y–Ba–Cu–O Superconductors

In addition to crystallographic characterization and measurement of thermal and electrical properties of superconducting materials, rapid and nondestructive analytical techniques involving milliprobe facilities were developed to search for the best conditions for their synthesis and to test the stoichiometric composition of $YBa_2Cu_3O_{6.5+x}$.

Irradiation of pellets made from a powder superconductor with 2 MeV deuterons may give, at the same time, the relative bulk concentration of all the elements.[55] A beam of deuterons is focused on a narrow region (5 μm) of the pellet. A solid-state particle detector, situated at 135° relative to the incident beam, collects protons from (d,p) reactions on O and C. A low carbon content is always possible from the precursors used in the synthesis ($BaCO_3$) or from a deposit on the sample surface by the incident deuteron beam of some residual atoms from oil used in the pumping set-up. The elastically scattered deuterons are stopped in a mylar foil of appropriate thickness inserted between the irradiated surface and the detector. Simultaneously, a Si(Li) detector collects X-rays, L X-rays of Ba and K X-rays of Cu and Y which are used as analytical signals for the rest of the elements. A second solid-state detector is installed at 170° relative to the incident beam and is used to collect the elastically backscattered deuterons in order to check the homogeneity of the Y, Ba and Cu distribution at depth. A flat plateau of Ba, Y and Cu in the RBS spectrum would certify this homogeneity. The deuteron beam is rastered on the surface. The integrated time of the whole measurement is of the order of 1 h to create a map of the four elements in an area of 40 × 40 spots of 5 μm wide.

A typical proton spectrum is shown in Figure 9(a). Carbon signals from surface contamination appear in region D. Carbon signals from the bulk material (region C) are to be subtracted from regions of oxygen signals (regions A and B). The subtraction of the carbon contribution in regions A and B of oxygen in the bulk (A) and at the surface (B) is performed by calculating the relative contribution of carbon in (C) and in the corresponding regions A and B of the proton spectrum collected on a pure carbon homogeneous standard sample (Figure 9b).

Using the pure RBS cross-section on Cu, Y and Ba, the stoichiometric determination of heavy compounds (expected to be $Y_1Ba_2Cu_3$) was $Y_{1.01}Ba_2Cu_{3.01}$ (Ba_2 as reference).

The accuracy of the measurement of O concentration (7 ± 1) by RBS was not sufficient because of superposition of the O signals on to the huge contribution of the heavier elements. The accuracy of the O content using results of the (d,p) reaction, is of the order of 3% (statistical accuracy).

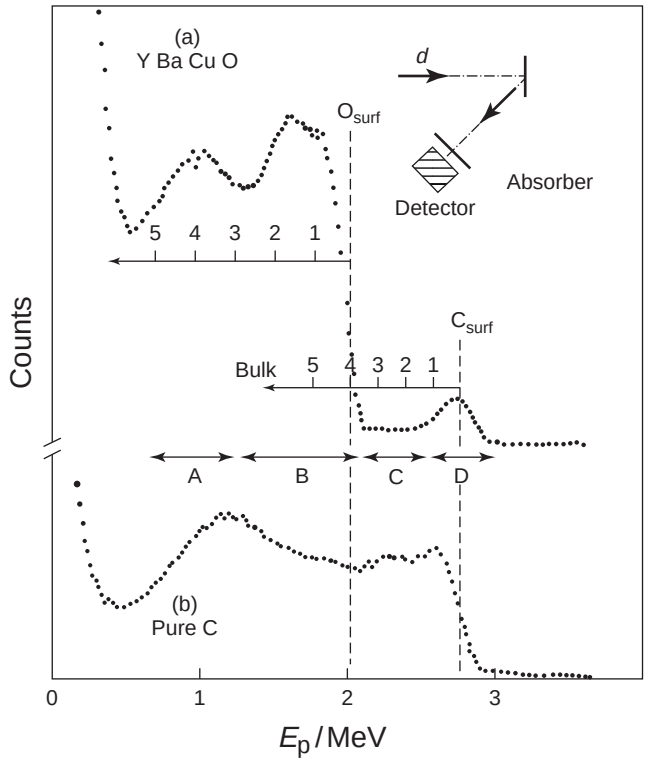

Figure 9 Proton spectra from (d,p) reactions on C and O. The incident deuteron energy is 2 MeV and the mylar absorber thickness is 10 mg cm^{-2}. Oxygen may be profiled down to 5–6 mg cm^{-2} below the surface. (Reprinted from G. Demortier et al., 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors With Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496, Copyright (1988), with permission from Elsevier Science.[56])

Materials showing the strongest Meissner effect (levitation of a small magnet) at liquid nitrogen temperature are also those showing the best homogeneity in Cu, Ba and Y concentrations in $5 \times 5 \times 5$ μm regions analyzed (Figure 10a and b). Possible local nonhomogeneities in several samples show that Ba and Y are related.[56]

Ternary diagrams of three of the four components (Y, Ba, Cu, O) indicate that, as far as the most homogeneous material is concerned, local discrepancies from the ideal $Y_1Ba_2Cu_3$ composition are mainly observed in the Y content, the reason lies only in the poor statistical accuracy on K X-rays of Y induced by 2 MeV deuterons.[56] An improvement in the accuracy of Y determination is achieved when the same sample is studied by PIXE induced by 3 MeV protons. The distribution of the points in a ternary diagram involving O, Cu and Ba as extracted from numerical results on the homogeneous material (Figure 11a) is not of statistical origin, but indicates some nonhomogeneity in the microscopic distribution of oxygen. For the second (nonhomogeneous) sample, the nonhomogeneity is so great that the points are scattered far from the ideal position: the center of the triangle. Nonhomogeneities in the O concentration of Figure 11(b) cannot be correlated with the presence of any specific oxide (CuO, Y_2O_3, BaO, etc.). For the homogeneous sample, the local concentration of oxygen exhibits variations of about 10% around the expected value of 6.7. The reason could be the presence, in the synthesized sample, of copper in two different valence states (2^+ and 3^+).

The experimental accuracy on the relative concentrations of Cu, Ba and O, using simultaneously PIXE and (d,p) reactions, was less than 3%.

5.4 Three-dimensional Microanalysis of a Gold–Silicon Eutectic Alloy

A film of silicon was deposited on pure polycrystalline gold foils, previously rolled down to thicknesses ranging from 10 to 20 μm. The deposition was performed with an electron gun at low pressure. The amount of evaporated silicon was adjusted in order to obtain films up to 1 μm thick on the gold substrate. This substrate was maintained at 400 °C during the silicon deposition, a temperature which is slightly higher than the eutectic temperature (363 °C). At this temperature a liquid Au–Si mixture diffuses, during the deposition procedure, in the gold foils (Figure 12) but only along well-defined paths, the gold grain boundaries. K X-rays of Si induced by PIXE **PIXE (Particle-induced X-ray Emission)** clearly indicate that a gold–silicon alloy has diffused on the surface of the foil by forming characteristic leaf-shaped decorations starting from the boundaries of the gold grains which were flattened by the rolling procedure. Several of these

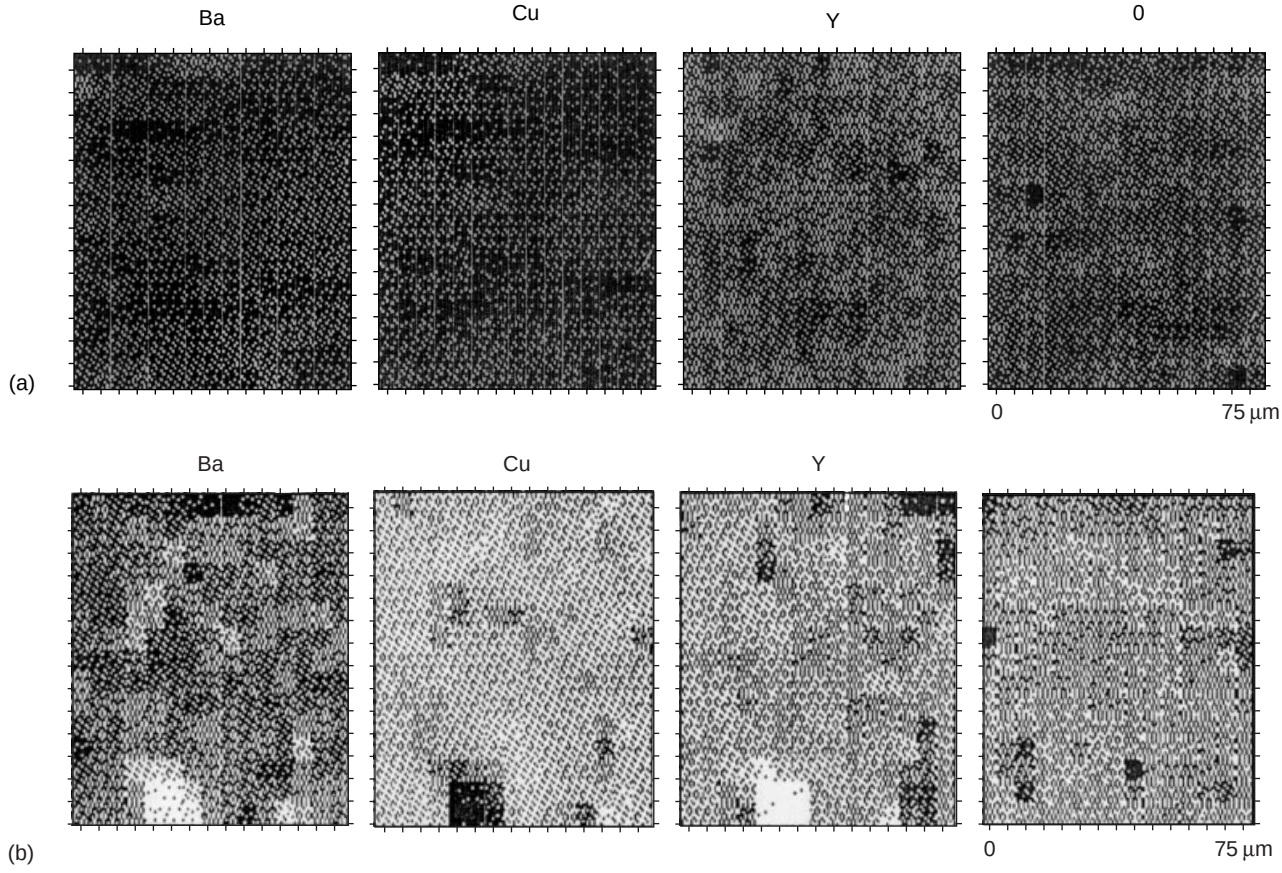

Figure 10 Concentration maps of Ba, Cu, Y and O in bulk superconductors. Concentrations are presented by 64 gradations of gray. The maximum values are, respectively, in each pixel: 8192 for Ba(L), 4096 for Cu(K_α), 256 for Y(K_α) and 1024 for O (protons). The beam diameter is less than 5 μm and the distance between steps is 10 μm. (a) Homogeneous YBaCuO sample; (b) Nonhomogeneous YBaCuO sample. (Reprinted from G. Demortier et al., 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors With Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496. Copyright (1988), with permission from Elsevier Science.[56])

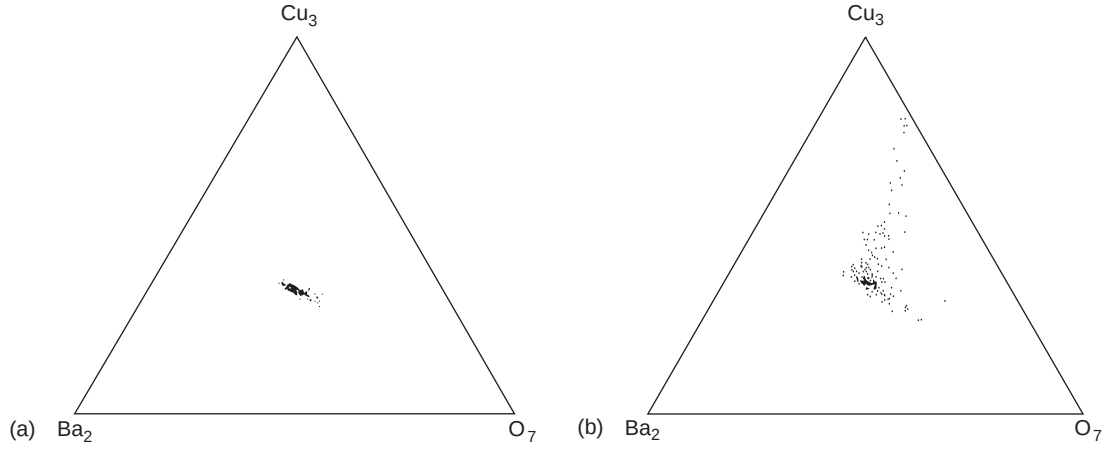

Figure 11 Ternary diagram of Cu, Ba and O for a homogeneous sample (a) and a nonhomogeneous one (b). (Reprinted from G. Demortier et al., 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors With Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496. Copyright (1988), with permission from Elsevier Science.[56])

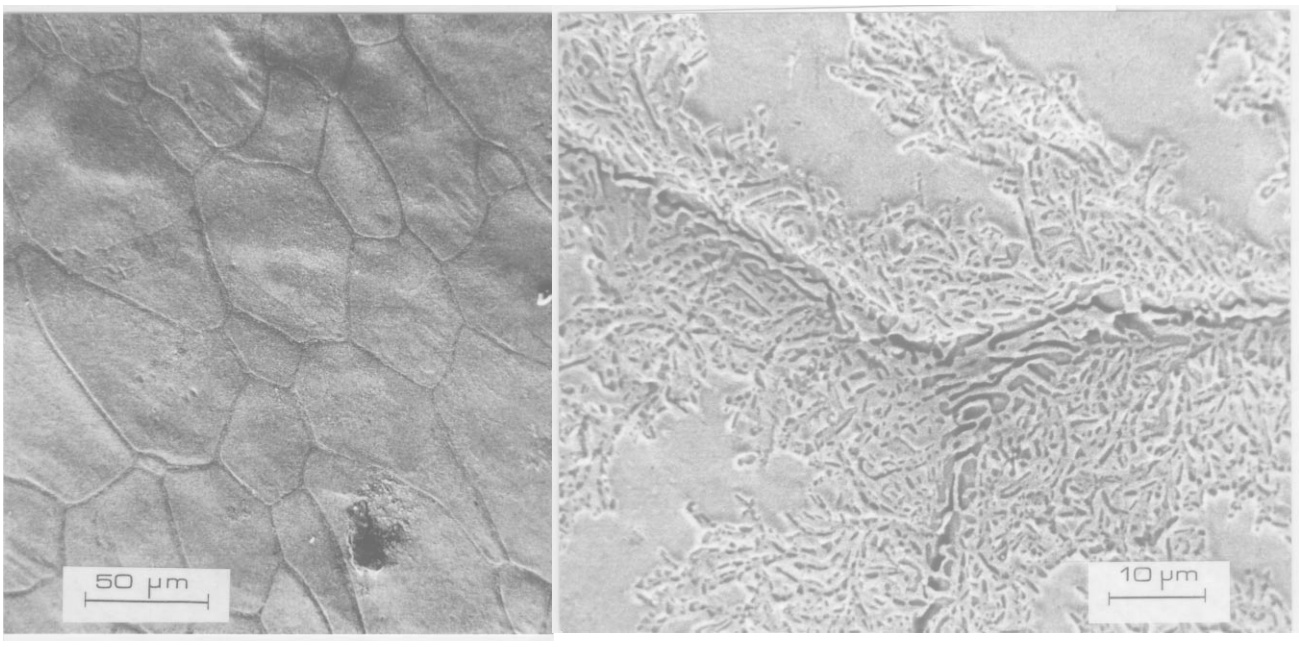

Figure 12 Electron micrographs of a Au–Si eutectic alloy formed by Si deposition in the grain boundaries of a 12 μm thick sheet of gold. (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

flattened grains may occupy the whole thickness of the foil. Therefore, the diffusion of the liquid phase took place through the whole gold foil thickness. The diffusion is so fast (a few minutes at such a low temperature cannot be explained by a process of solid state diffusion) that nothing other than a liquid phase of gold–silicon eutectic alloy could diffuse. The atomic concentration of Si in this eutectic alloy is 19% (about 3.2% Si by weight).

The depth profile of Si in gold is taken using a (d,p) reaction.[57] The nuclear reaction $^{28}Si(d,p_0)^{29}Si$ ($Q = 6.25\,keV$) (leaving the residual ^{29}Si nuclei in their fundamental energy level) is mainly governed by neutron capture involving no exchange of angular momentum ($\ell = 0$). As a consequence, the maximum intensity of the proton emission is achieved in the forward direction. The cross-section at $\theta = 0^\circ$ of this nuclear reaction at $E_d = 2.8\,MeV$ is quite large at $6\,mb\,sr^{-1}$. We then chose to detect the emitted protons in the forward direction. In this special arrangement, the proton detector is situated into the incident deuteron beam (Figure 13). As incident deuterons of 2.8 MeV cannot cross the coulomb barrier of a Au nucleus, the (d,p) nuclear reaction can only take place with light nuclei. Therefore, the detected proton spectra contain only information on Si. A 20 μm thick absorber of pure gold is inserted between the sample to be studied and the detector in order to stop the incident deuteron beam completely even when the beam would cross a hole in the gold–silicon sample. The choice of a material containing only a heavy element (Au) as the absorber prevents any production of additional protons in this gold absorber during the deuteron irradiation.

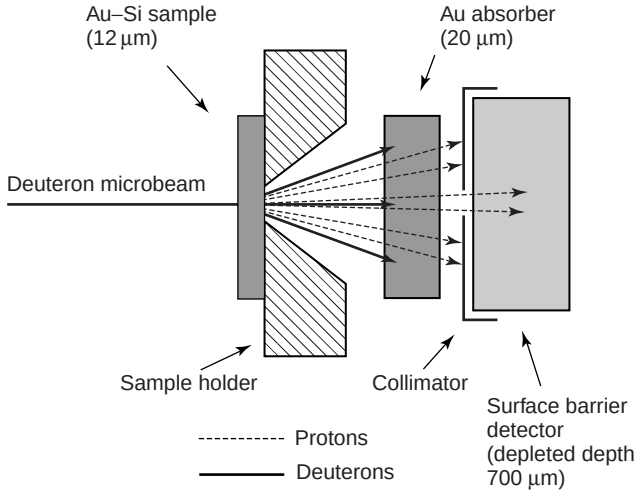

Figure 13 Special experimental set-up used to perform depth profiling of Si in a thin gold foil treated for Si deposition. (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

The quantitative determination of silicon at different depths below the surface is based on the comparison of proton spectra obtained when ultrapure silicon and gold–silicon samples in the same geometrical

arrangement are bombarded. The silicon reference material is a thin foil (75 μm thick) of Si used in semiconductor technology. The corresponding depths in pure silicon and in gold–silicon alloys are calculated using the relative stopping power of pure silicon and that of the eutectic gold–silicon alloy (Figure 14). To scan the sample, a proton microbeam (5 μm × 5 μm) is rastered on the surface along 50 μm × 50 μm regions, a surface involving more than the area of one single flattened grain.

If the deuteron beam hits a gold grain boundary in which the diffusion of the eutectic alloy took place, a proton spectrum like that of Figure 14(a) may be observed

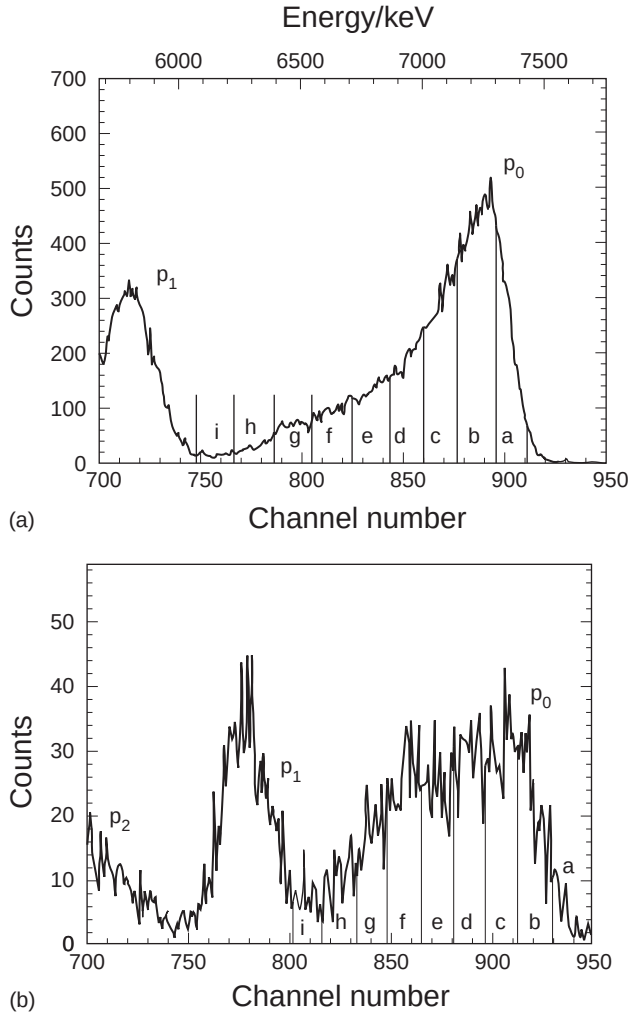

Figure 14 Experimental procedure to obtain depth profiling of silicon from the decomposition in nine slices from the p_0 contribution. (a) Pure 75 μm Si sample; (b) 12 μm thick Au foil processed for silicon diffusion (irradiation of the rear face with 2.8 MeV deuterons). (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

showing a more important contribution in the medium and the low energy parts of the p_0 broad ''peak'' than when this deuteron beam hits a uniform silicon sample (Figure 14b). The depth distribution of Si in this particular case indicates that the relative concentration of Si is increasing from the surface to a region buried at several microns below the surface.

The map of Figure 15 corresponds to the silicon distribution at various depths below the surface irradiated at the rear part of the prepared gold foil. Figure 15(a) reflects the Si concentration at the bottom of the prepared foil which is reached by the most energetic protons. The regions of the proton spectrum shown in Figure 14(a) show schematically the corresponding regions in the depth of the 12-μm thick irradiated specimens. The incident beam was directed onto the face of the gold foil opposite to the face on which the Si deposition was made. In any region the silicon concentration does not exceed a few percent, a concentration well below that of the eutectic composition. This eutectic concentration is observed in Figure 15(i) which corresponds to grain boundary regions at the point where deuterons exit and at the surface where the silicon was deposited.

Nevertheless, in regions where silicon is detected only a liquid phase with a composition close to that of the eutectic could be admitted. The diffusion takes place so rapidly (a few minutes) that no appreciable solid state diffusion could be expected. Each concentration of silicon lower than 19% (as observed in nearly all parts of each map) can be understood by considering that the size of a grain boundary is narrower than the area of the incident deuteron beam. Scans like those of Figure 15 may then be converted into measurements of the width of the grain boundary filled with the eutectic alloy, the ratio of the measured silicon concentration on the eutectic composition being the ratio of the width of the grain boundary to the beam diameter.

5.5 Determination of Traces of Sulfur in Ancient Gold Artefacts

Investigations of the composition of archaeological artefacts made from precious metals require the use of nondestructive analytical techniques. PIXE and PIGE have been successfully applied in our laboratory for topographical elemental analysis of Fe, Cu, Zn, Ag, Cd, Sn and Au in uncorroded surfaces of jewellery items. The composition of the material in narrow regions of tiny solders on antique gold artefacts may give valuable information about the workmanship of ancient goldsmiths. In various papers, the hypothesis that in ancient times cadmium sulfide (a yellow mineral with a color close to that of gold) may have been ''alloyed'' with gold to obtain a material suitable for brazing has

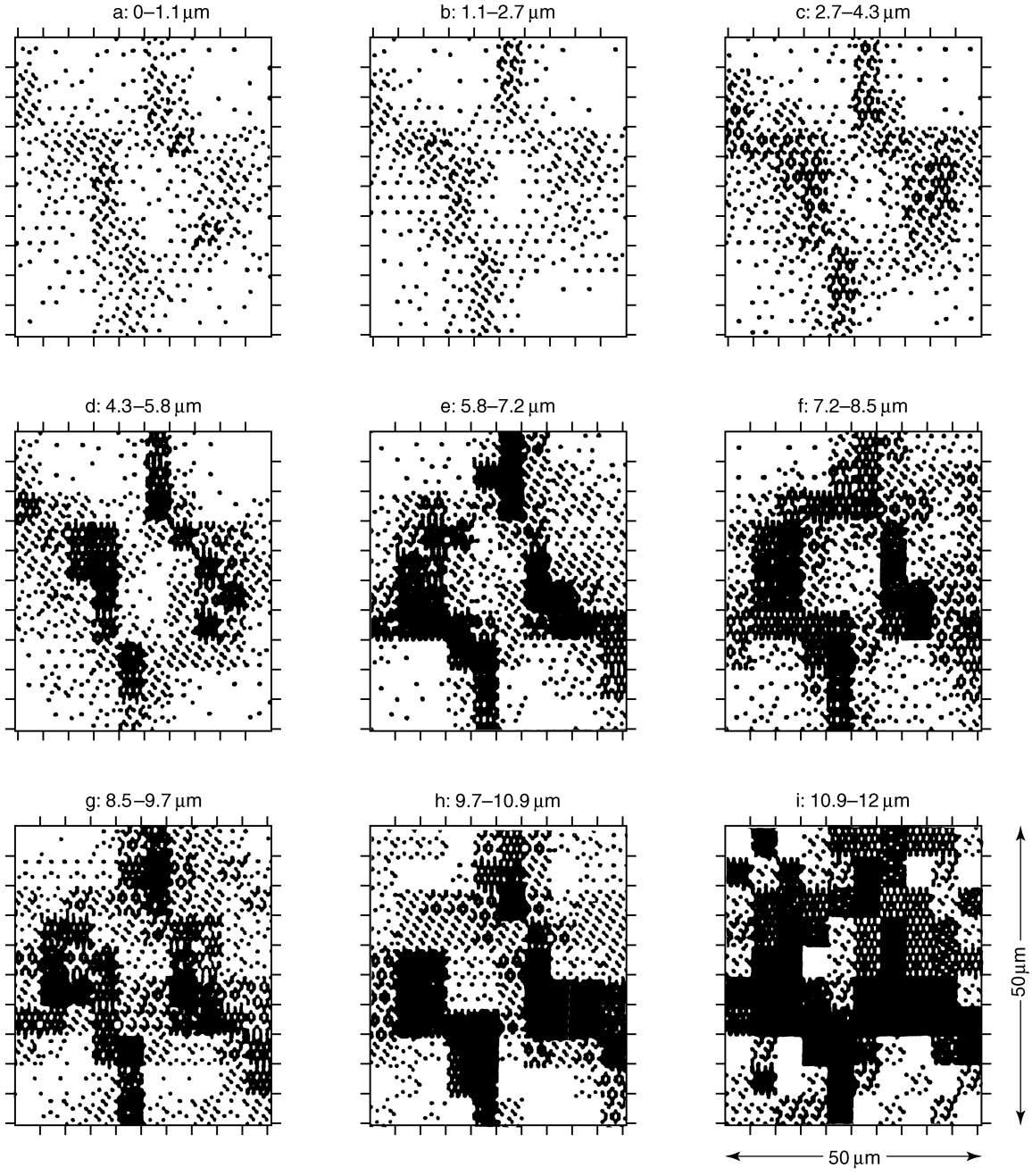

Figure 15 Set of nine maps corresponding to the nine slices of Figure 14. Symbols (a) to (i) refer to the nine layers in the 12 μm thick sample. (Reprinted from S. Mathot, G. Demortier, 'Three-dimensional Nuclear Microanalysis in Materials Science', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **77**, 312–319. Copyright (1993) with permission from Elsevier Science.[55])

been discussed.[57] To test this hypothesis an experimental procedure for analyzing traces of sulfur in gold artefacts was developed. It is indeed expected that during the antique alloying procedure, traces of sulfur from cadmium sulfide were introduced in the brazing alloy.

Detection of energetic protons induced by deuteron bombardment of the specimen under investigation was achieved. The detector was situated at an angle of 135° with respect to the incident deuteron beam in order to facilitate the study of thick specimens often with an irregular shape. Sheets of mylar of uniform thickness (2.9 mg cm^{-2} each) were inserted between the target and the detector in order to stop the scattered deuterons completely. Owing to the fact that most of the (d,p) reactions have a large positive Q value, the proton energies are not significantly reduced by these absorbers.

The best experimental condition for sulfur determination is obtained at $E_d = 1.9$ MeV, in which case the cross-section is sufficient and the p_0 and p_1 peaks are well resolved.[57] Since the cross-section for 1.7 MeV deuterons is greater than at 1.9 MeV, working conditions at 1.9 MeV (on a thick sample) allow us to check for the presence of sulfur in the bulk and not only at the surface. Possible interfering elements (with their appropriate Q values) are within the scope of (d,p) reactions on S: Si, Mg, Cl, Ca (for p_0) and Al, P, Cl, K, Mg (for p_1).

The choice of detection angle and incident deuteron energy was then a determining factor in eliminating the

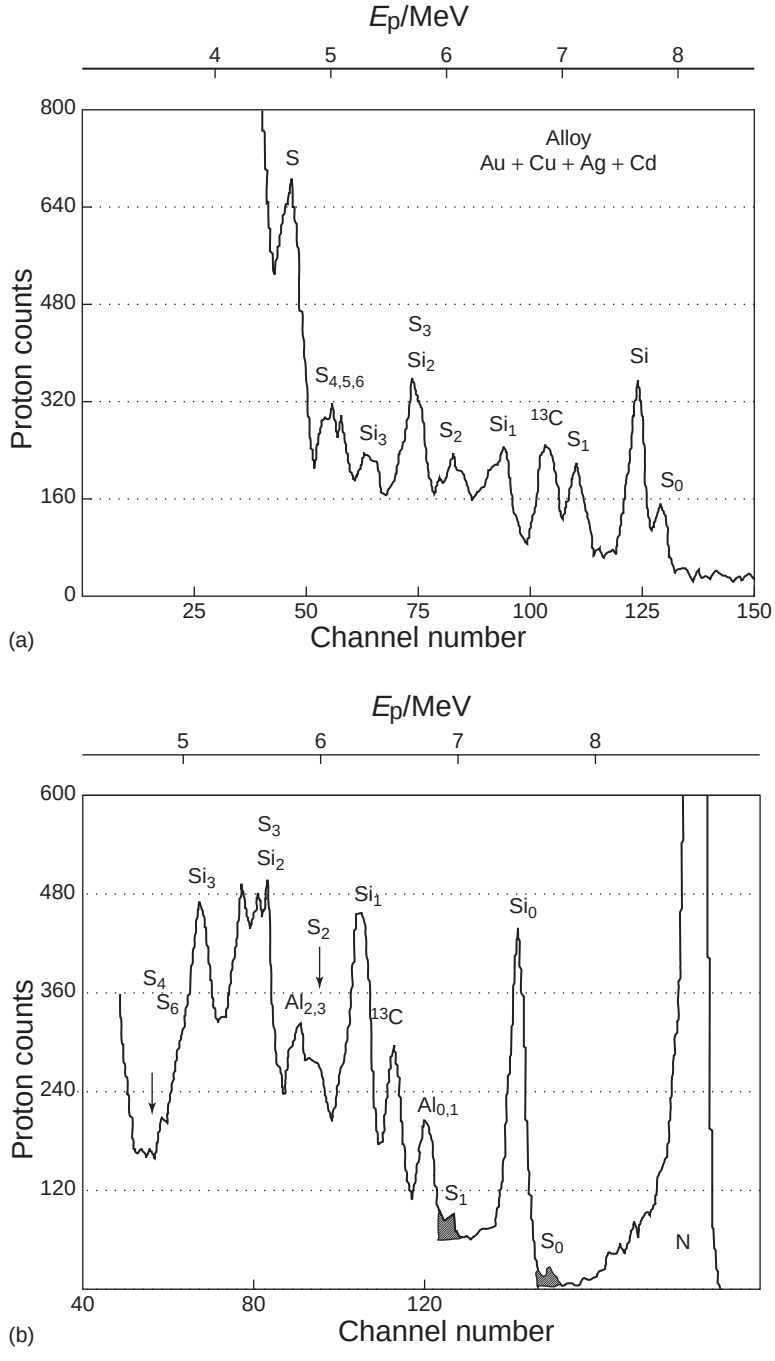

Figure 16 Experimental results from a (d,p) reaction induced on gold alloys prepared by an antique procedure (a) and in a soldering region of an antique jewellery artefact of the 9th century A.D. (b) Several peaks indicate the presence of Si and S. (Reprinted from G. Demortier, A. Gilson, 'Determination of Traces of Light Elements in Gold Artifacts Using Nuclear Reactions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **18**, 286–290. Copyright (1987) with permission from Elsevier Science.[57])

possible interferences. Detection of p_0 and p_1 protons provided a check for an interference-free investigation. This is because of the presence of both p_0 and p_1 signals for S outside the regions for Al($p_0 + p_1$) and (Si p_0) and the relative intensities of these peaks in comparison with a reference sample.

Sulfur and silicon (which were always introduced in metals when heating ores in antique metallurgy) are of crucial importance in understanding ancient technologies. Sulfur (in blende and greenockite) and silicon (in many minerals) may be expected to be present in narrow parts of the solder formed during the antique brazing of gold. Figure 16(a) shows the analysis of a soldering alloy that we prepared using this antique procedure: gold–cadmium alloys (with low melting point) are obtained by direct dissolution of CdS in a previously melted drop of gold. The concentrations of S and Si detected are, respectively, 120 ± 35 ppm and 50 ± 15 ppm. Note also the presence of C (from ^{13}C) as an extra signal arising from the experimental set-up, i.e. a carbon (from oil in the pumping system) deposit on the target during the deuteron irradiation. The mylar absorber must have a thickness of 25 mg cm^{-2} in order to eliminate also protons from

Figure 17 Small spherical pearl (diameter 11 mm) (9th century A.D.) decorated with granules, and which could be a part of a necklace, found in Syria with other Byzantine jewellery items.

Table 5 Analysis of biological reference materials by NRA, PIGE and PIXE

Element	IBA techniques				Bowen's values[58]	
	E_i in keV	Reaction	Particle or photon detected	Concentration	Mean	Ranges
B	450	(p,α)		46.2 ± 5 ppm	48.3	39–56
	700	(p,α)		45.1 ± 4 ppm		
C	1020	(d,p)	p_0	$43.7 \pm 2\%$	41.8%	
N	1500	(d,α)	α_0	$4.76 \pm 0.3\%$	4.31%	
O	900	(d,p)	p_0	$43 \pm 3\%$	5.15%[a]	
F	2000	(p,p′γ)	γ110 keV	7.14 ± 2 ppm	4.92	4.2–6.2
	2000	(p,p′γ)	γ197 keV	6.23 ± 2 ppm		
	2200	(p,αγ)	γ5100 keV	6.8 ± 3 ppm		
	2200	(p,αγ)	γ6100 keV	6.1 ± 3 ppm		
Na	1500	(p,p′γ)	γ439 keV	2390 ± 120 ppm	2506	1220–3250
	1900	(p,p′γ)	γ439 keV	2630 ± 100 ppm		
	2000	(p,p′γ)	γ439 keV	2510 ± 80 ppm		
	2100	(p,p′γ)	γ439 keV	2480 ± 70 ppm		
	2200	(p,p′γ)	γ439 keV	2540 ± 70 ppm		
	2000	(p,p′γ)	γ585 keV	1920 ± 250 ppm	1572	1370–1700
	2200	(p,p′γ)	γ585 keV	1580 ± 150 ppm		
P	700	PIXE	K_α	$0.47 \pm 0.04\%$	0.4489	0.402–0.481
Cl	700	PIXE	K_α	$0.37 \pm 0.02\%$	0.3415	0.218–0.445
K	700	PIXE	K_α	$2.70 \pm 0.2\%$	2.46	2.06–2.93
Ca	700	PIXE	K_α	$4.05 \pm 0.1\%$	4.085	
Mn	1700	PIXE	K_α	13.0 ± 2 ppm	14.7	12.6–18.0
Fe	1700	PIXE	K_α	113 ± 7 ppm	118.3	88–157
Cu	1700	PIXE	K_α	4.9 ± 0.5 ppm	4.99	3.6–6.5
Zn	1700	PIXE	K_α	32.5 ± 2 ppm	33.2	30–38
Br	1700	PIXE	K_α	26.0 ± 2 ppm	26.1	
Rb	1700	PIXE	K_α	49.0 ± 4 ppm	52.2	49–57
Sr	1700	PIXE	K_α	100.2 ± 5 ppm	98.9	65–150

[a] This low oxygen concentration, determined by neutron activation analysis, is not reliable with the expected value for a biological sample.

(d,p) reactions on ^{12}C. The concentration of Cd could be determined simultaneously by detection of characteristic X-rays induced by the same deuteron beam. The presence of residual sulfur (only at trace level) indicates that at least 95% of S was lost during the alloying procedure.

Figure 16(b) was obtained during the irradiation of a region of a soldered gold granule on a hollow gold pearl found in Syria (Hauran) with other more prestigious objects of Byzantine jewellery (9th century A.D.) (Figure 17).

The information from S (at concentration 30 ± 7 ppm) cannot be easily isolated from the large contribution of Si at the surface. The width of the peaks of Si present in this ancient artefact is much narrower than those observed in pure Si and in the soldering alloys formed by the methods available in antiquity (Figure 16a). The presence of traces of Si and often also traces of Al, can be attributed to dust inclusions in narrow regions of ''porous'' solders.

5.6 Particle-induced X-ray Emission, Particle-induced γ-ray Emission and Nuclear Reaction Analysis for Elemental Analysis of a Biological Reference Material

Bowen's kale,[58] a powder obtained by dry freezing this vegetable, was widely distributed in various laboratories to be certified for a reference material. Using various IBA techniques, it was possible to determine most of the main elements and traces quantitatively. Table 5 summarizes these results.[59]

NRA induced by protons and deuterons was used for the determination of B, C, N, F and Na. The choice of a specific incident energy and the selected identification signal were determined for interference-free determination.

Compositions of various elements (from P or Ca) were obtained at low energy PIXE in order to minimize the X-ray absorption into the target itself. Traces of higher Z elements were found using PIXE at 1.7 MeV in order to improve the sensitivity. Determination of hydrogen was not possible due to the rapid destruction of the sample powder during heavy ion irradiation (for ERD or NRA using ^{15}N). The concentration of oxygen was established at 43%, a more reliable value than the previous determination using neutron activation analysis. The total content of this reference material using PIXE, PIGE and NRA exclusively (hydrogen not included) is $98.9 \pm 3.5\%$.

ABBREVIATIONS AND ACRONYMS

BGO	Bismuth Germanate Oxide
CPAA	Charged Particle Activation Analysis
ERD	Elastic Recoil Detection
fwhm	Full Width at Half Maximum
IBA	Ion Beam Analysis
NRA	Nuclear Reaction Analysis
PIGE	Particle-induced γ-ray Emission
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy

RELATED ARTICLES

Coatings **(Volume 2)**
Coatings Analysis: Introduction

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in • Noble Metals, Analytical Chemistry of

Surfaces **(Volume 10)**
Ion Scattering Spectroscopy in Analysis of Surfaces

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission) • Rutherford Backscattering Spectroscopy

REFERENCES

1. P. Vavilov, 'Ionization Losses of High Energy Heavy Particles', *Sov. Phys. JETP*, **5**, 749–751 (1997).
2. G. Deconninck, *Introduction to Radioanalytical Physics*, Elsevier Scientific, Amsterdam, 1978.
3. G. Deconninck, Y. Fouilhe, 'Energy Spreading Calculations and Consequences', in *Ion Beam Surface Layer Analysis*, eds. O. Meyer, K. Käppeler, G. Linker, Plenum Press, 97, Vol. 1, 1975.
4. J.M. Harris, M.A. Nicolet, 'Energy Straggling of ^4He Ions Below 2 MeV in Al, Ni and Au', *Phys. Rev. B*, **11**, 1013 (1975).
5. S.A.E. Johannsson, J.L. Campbell, K.G. Malmqvist (eds.), *Particle-induced X-ray Emission Spectrometry (PIXE)*, John Wiley & Sons, New York, 1995.
6. G. Deconninck, G. Demortier, F. Bodart, 'Prompt Gamma-ray Spectroscopy and its use in Elemental Analysis', *Atomic Energy Rev.*, Suppl. 2, 1981.
7. W.S.C. Williams, *Nuclear and Particle Physics*, Clarendon Press, Oxford, 1991.
8. N. Jarmie, J. Seogrove, 'Nuclear Data Tables', Los Alamos Scientific Laboratory Report, LA-2014, 1957.
9. J.L. Yarnell, R.H. Lovberg, W.R. Stratton, 'Angular Distribution of the Reaction $^3He(d,p)^4He$ Between 240 keV and 3.65 MeV', *Phys. Rev.*, **90**, 292 (1953).

10. A.J. Elwyn, R.E. Holland, C.N. Davids, L. Meyer-Shützmeister, F.P. Moorning, W. Ray, Jr, 'Cross Sections for ^{6}Li(p,^{4}He)^{3}He Reaction at Energies Between 0.1 and 3.0 MeV', *Phys. Rev. C*, **20**, 1984 (1979).
11. B. Maurel, 'Théorie stochastique de la perte d'énergie dans la matière des particules chargées. Application aux courbes d'excitation autour de résonances nucléaires étroites et à la mesure des profils de concentration près de la surface des solides', PhD Thesis, University of Paris VII, 1981.
12. J.P. Schiffer, T.W. Bonner, R.H. Davis, F.W. Prosser, Jr, 'Study of the Reaction Mechanism for (^{3}He,p) Reactions with ^{6}Li; ^{11}B and ^{13}C', *Phys. Rev.*, **104**, 1064 (1980).
13. Y. Cassagnou, J.M.F. Jeronymo, A. Sadeghi, P.D. Forsyth, 'The ^{7}Li(p,α)α Reaction', *Nucl. Phys.*, **33**, 449 (1962).
14. A.J. Sierk, T.A. Tombrello, 'The ^{9}Be(p,α) and (p,d) Cross Sections at Low Energies', *Nucl. Phys. A*, **210**, 341 (1973).
15. M. Shaanan, R. Kalish, V. Richter, 'Measurement of Profiles by ^{9}Be(p,α)^{6}Li Reaction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **16**, 56 (1986).
16. J.A. Biggerstaff, R.F. Hood, H. Scott, M.T. McEllistrem, 'Differential Cross Sections for (d,α) and (d,t) Reactions in ^{9}Be', *Nucl. Phys.*, **36**, 631 (1962).
17. E.A. Wolicki, H.D. Holmgren, R.L. Johnston, E. Geer-Illsley, 'Differential Cross Sections for the ^{9}Be(^{3}He,p)^{11}B Reaction', *Phys. Rev.*, **116**, 1585 (1959).
18. B. Bilwes, R. Bilwes, J.L. Ferrero, A. Garcia, 'Mécanisme de la réaction ^{9}Be(^{3}He,α)^{8}Be(0) et résonances géantes isoscalaires du ^{12}C', *J. Physique*, **39**, 805 (1978).
19. K.H. Purser, B.H. Wildenthal, 'The B^{10}(d,α)Be^{8} Reaction and the 2.94 MeV State in ^{8}Be', *Nucl. Phys.*, **44**, 22 (1963).
20. L.C. McIntyre, Jr, J.A. Leavitt, M.D. Ashbaugh, Z. Lin, J.O. Stoner, Jr, 'Determination of Boron Using the B(α,p)C Nuclear Reaction at Incident Energies near 3 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 221 (1992).
21. A. Armigliato, G.G. Bentini, G. Ruffini, G. Battaglin, G. Della Mea, A.V. Drigo, 'Analysis of Boron Predeposited Silicon Wafers by Combined Ion Beam Techniques and X-ray Microanalysis', *Nucl. Instrum. Methods*, **149**, 653 (1978).
22. E. Ligeon, A. Bontemps, 'Nuclear Reaction Analysis of Boron and Oxygen in Silicon', *J. Radioanal. Chem.*, **12**, 335 (1972).
23. J.P. Scanlow, G. Farell, M.C. Ridgway, R. Valizadeh, 'Nuclear Reaction Analysis of Shallow Boron Implanted in Silicon', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **16**, 479 (1986).
24. H.D. Holmgren, E.A. Wolicki, R.L. Johnston, '^{11}Be(^{3}He,p)^{13}C and ^{11}B(^{3}He,d)^{12}C Reactions', *Phys. Rev.*, **114**, 1281 (1959).
25. C. Olivier, J.W. McMillan, T.B. Pierce, 'An Investigation of the Use of the Nuclear Microprobe for Examining Distributions of Boron Isotopes', *J. Radioanal. Chem.*, **31**, 515 (1976).
26. E. Kashy, R.R. Perry, J.R. Risser, 'Excited States in ^{14}N from ^{12}C(d,d)^{12}C, ^{12}C(d,p$_0$)^{13}C and ^{12}C(d,p$_1$)^{13}C*', *Phys. Rev.*, **117**, 1289 (1960).
27. G. Terwagne, 'Cross Section Measurements of the Reactions Induced by ^{3}He Particles on Carbon', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **122**, 1 (1997).
28. H.M. Kuan, T.W. Bonner, J.R. Risser, 'An Investigation of the ^{12}C + ^{3}He Reactions at Bombarding Energies Between 1.8 and 5.4 MeV', *Nucl. Phys.*, **51**, 481 (1964).
29. J.B. Marion, G. Weber, 'Compound Nucleus Effects in Deuteron Reactions ^{12}C(d,p)^{13}C', *Phys. Rev.*, **103**, 167 (1956).
30. G. Amsel, D. David, 'La microanalyse de l'azote par l'observation directe de réactions nucléaires: applications', *Rev. Phys. Appl.*, **4**, 383 (1969).
31. W. Koenig, A. Riccato, R. Stock, P. Cuzzocrea, E. Perillo, M. Sandolini, G. Spadaccini, 'Cross Section Measurements and Angular Distribution Study of ^{14}N(d,α)^{12}C Nuclear Reaction', *II Nuovo Cimento A*, **39**, 9 (1977).
32. T. Ishimatsu, T. Nakashima, N. Kato, S. Morita, 'The ^{14}N(d,α_0) and ^{14}N(d,α_1) Reactions from 1 to 3.5 MeV', *J. Phys. Soc. Jpn.*, **17**, 1189 (1962).
33. F.B. Hagendorn, J.B. Marion, 'Further Investigation of ^{15}N + p Reactions', *Phys. Rev.*, **108**, 1015 (1957).
34. J.A. Sawicki, J.A. Davies, T.E. Jackman, 'Absolute Calibration of the ^{14}N(d,α_0)^{13}C and ^{15}N(p,α_0)^{12}C Reaction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 530 (1986).
35. F. Seiler, C.H. Jones, W.J. Anzick, D.F. Herring, K.W. Jones, 'The Elastic Scattering of Deuterons by ^{16}O from 0.65 to 2.0 MeV', *Nucl. Phys.*, **45**, 647 (1963).
36. A. Turos, L. Wielunski, A. Barcz, 'Use of the Nuclear Reaction ^{16}O(d,α)^{14}N in the Microanalysis of Oxide Surface Layers', *Nucl. Instrum. Methods*, **111**, 605 (1973).
37. S.Y. Tong, W.N. Lennard, P.F.A. Alkemada, I.V. Mitchell, 'Absolute Surface Carbon Coverage Determination via the ^{12}C(^{3}He,p)N^{14} Reaction', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **45**, 91 (1990).
38. G. Amsel, 'Spectroscopie par détecteurs à semidconducteur des réactions ^{16}O + d, ^{18}O + p', *Ann.Phys.*, **9**, 297 (1964).
39. D. Dieumegard, B. Maurel, G. Amsel, 'Microanalysis of Fluorine by Nuclear Reactions. I. ^{19}F(p,α_0)^{16}O and ^{19}F(p,$\alpha\gamma$)^{16}O Reactions', *Nucl. Instrum. Methods*, **168**, 93 (1980).
40. B. Maurel, G. Amsel, D. Dieumegard, 'Microanalysis of Fluorine by Nuclear Reactions. II. ^{19}F + d Reactions', *Nucl. Instrum. Methods*, **191**, 349 (1981).
41. M. Vos, D.O. Boerma, P.J.M. Smulders, S. Oosterhoff, 'Defect and Dopent Depth Profiles in Boron-implanted Silicon Studied with Channeling and Nuclear Reaction Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **17**, 234 (1986).
42. S. Fiarman, S.S. Hanna, 'Energy Levels of Light Nuclei, A = 3', *Nucl. Phys. A*, **251**, 1–64 (1975).
43. S. Fiarman, W.E. Meyerhof, 'Energy Levels of Light Nuclei, A = 4', *Nucl. Phys. A*, **206**, 1–64 (1973).

44. P.M. Endt, C. Van der Leun, 'Energy Levels of Light Nuclei, A = 21–44', *Nucl. Phys. A*, **310**, 1–752 (1978).
45. F. Ajzenberg-Selove, 'Energy Levels of Light Nuclei, A = 5–10', *Nucl. Phys. A*, **227**, 1–244 (1974); 'Energy Levels of Light Nuclei, A = 16–17', **281**, 1–148 (1977); 'Energy Levels of Light Nuclei, A = 18–20', **475**, 1–198 (1987); 'Energy Levels of Light Nuclei, A = 11–12', **506**, 1–158 (1990); and 'Energy Levels of Light Nuclei, A = 13–15', **523**, 1–196 (1991).
46. J.W. Mayer, E. Rimini (eds.), *Ion Beam Handbook for Material Analysis*, Academic Press, New York, 1997.
47. J. Tesmer, M. Nastasi (eds.), *Handbook of Modern Ion Beam Materials Analysis*, MRS, Pittsburgh, 1995.
48. W. Lanford, 'Analysis for Hydrogen by Nuclear Reaction and Energy Recoil Detection', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 65–82 (1992).
49. P. Trocellier, Ch. Engelmann, 'Hydrogen Depth Profile Measurements Using Resonant Nuclear Reactions: a Review', *J. Radioanal. Nucl. Chem.*, **100**, 117 (1986).
50. F. Xiong, F. Rauch, C. Shi, Z. Zhou, R.P. Livi, T.A. Tombrello, 'Hydrogen Depth Profiling in Solids: a Comparison of Several Resonant Nuclear Reaction Techniques', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **27**, 432 (1987).
51. K.M. Horn, W. Lanford, 'Observation of the Bond-dependent Doppler Broadening of the $p(^{15}N,\alpha\gamma)^{12}C$ Nuclear Reaction', *Nucl. Instrum. Methods*, **29**, 609 (1983).
52. G. Amsel, B. Maurel, 'High Resolution Techniques for Nuclear Reaction Narrow Resonance Width Measurements and for Shallow Depth Profiling', *Nucl. Instrum. Methods*, **218**, 183 (1988).
53. P.A. Pronko, J.G. Pronko, 'Depth Profiling of ^{3}He and ^{2}H in Solids Using the $^3He(d,p)^4He$ Resonance', *Phys. Rev. B*, **9**, 2870 (1974).
54. G. Terwagne, G. Deconninck, 'Concentration Profile of Light Elements Near an Interface', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 153–155 (1992).
55. S. Mathot, G. Demortier, 'Silicon Diffusion in Gold Grain Boundaries', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **50**, 52–56 (1990); and 'Three-dimensional Nuclear Microanalysis in Materials Science', **77**, 312–319 (1993).
56. G. Demortier, F. Bodart, G. Deconninck, G. Terwagne, Z. Gabelica, E.G. Derouane, 'Stoichiometric Characterization of Y–Ba–Cu–O Superconductors with Nuclear Probes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **30**, 491–496 (1988).
57. G. Demortier, A. Gilson, 'Determination of Traces of Light Elements in Gold Artifacts Using Nuclear Reactions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **18**, 286–290 (1987).
58. H.J.M. Bowen, 'Problems in the Elementary Analysis of Standard Biological Materials', *J. Radioanal. Chem.*, **19**, 215 (1972).
59. J. Baijot-Stroobants, G. Debras, G. Demortier, 'Analyse quantitative globale et distribution superficielle du carbone, de l'azote et de l'oxygène dans des matrices biologiques par réactions nucléaires de basse énergie', *Radioanal. Radiochem. Lett.*, **27**, 271–280 (1976).
60. G.Demortier (ed.), 'Microanalysis of Light Elements (Hydrogen to Neon) Using Charged Particle Accelerators', Proceedings of the Third International Conference on Chemical Analysis, Namur, July 1991, *Nucl. Instrum. Methods Phys. Res., Sect. B*, **66**, 1–309, 1992.

Particle-induced γ-Ray Emission

P.H.A. Mutsaers and M.J.A. de Voigt
Cyclotron Laboratory, Department of Applied Physics, Eindhoven University of Technology, and Center of Plasma and Radiation Physics, Den Dolech 2, P.O.Box 513, 5600 MB, The Netherlands

1 Introduction 1
2 Basic Nuclear Physics 2
2.1 Principles of Nuclear Reactions 2
2.2 γ-Ray Emission 5
2.3 Principles of Depth Profiling 7
3 Particle-induced γ-Ray Emission 7
3.1 Basic Principles of Particle-induced γ-Ray Emission 7
3.2 Possible Reactions and Their γ-Ray Energies 8
3.3 Sensitivity 8
3.4 Background 12
3.5 Experimental Set-up and Procedures 12
3.6 Depth Profiling 14
4 Application of the Particle-induced γ-Ray Emission Technique 15
5 Conclusions 15
Abbreviations and Acronyms 16
Related Articles 16
References 16

Particle-induced γ-ray emission (PIGE) is commonly applied for quantitative analysis of materials. γ-Rays, characteristic of the elements, are emitted after nuclear reactions. These reactions are induced by particle beams (e.g. protons) from accelerators. The reaction conditions and probabilities are explained, as well as the γ-ray properties and probabilities of emission. Knowledge of those parameters enables concentrations of elements to be deduced from the γ-ray spectra, which under certain conditions are also depth dependent. Therefore the samples must be small and flat. The experimental set-up and procedures are given together with a table with the most intense proton reactions and the energies of the emitted γ-rays. Various examples are given to illustrate the method in different applications.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

PIGE or prompt radiation analysis (PRA) is an analytical technique making use of γ-rays that are emitted during or immediately after bombardment of a sample with an ion beam. Here, discussion is restricted to prompt γ-ray emission. These are γ-rays that are detected during irradiation, in contrast to activation techniques where radiation is detected after irradiation.[1] The technique has been used since the 1960s for analytical purposes. Sippel and Glover[2] showed that PIGE with protons could be used for the determination of Mg and F in geological samples. Pierce et al.[3] showed that deuterons could be used to determine C and Pierce et al.[4] used 4 MeV protons to determine the amount of Si in several kinds of steel. At that time, the detection of γ-rays was mostly performed with NaI(Tl) detectors which have a relatively poor energy resolution, thereby limiting the analytical technique to the determination of one or two elements simultaneously. The development of Ge(Li) detectors with a much better energy resolution around 1970 made it possible to determine a number of elements in the sample simultaneously. Since then, studies have been carried out[5–9] with light as well as heavy projectiles that showed that PIGE with a Ge(Li) detector is a powerful and sensitive method of multielemental analysis. In general, protons and α-particles are used as projectiles, but measurements have also been done with d, t, ^{3}He, Li, B, N, F and Cl projectiles.

Besides the determination of concentrations in (thick) samples, PIGE can also be used to determine depth profiles by using narrow resonances in a cross-section of nuclear reactions. In general, PIGE is a technique that is less often used than other ion-beam analysis techniques like particle-induced X-ray emission (PIXE) or Rutherford backscattering spectroscopy (RBS). In PIXE or RBS spectra, elements appear in a relatively simple and systematic way, making interpretation and analysis of the spectra straightforward. In PIGE, the situation is more complex since even different isotopes of the same element result in completely different γ-spectra. Knowledge about the energy of resonances, the energy of the γ-rays, and interfering reactions in the sample must be available for each nuclide of interest. Interpretation and analysis of PIGE spectra can thus be much more complex than for PIXE or RBS spectra. This means that PIXE and RBS can be seen as general techniques to determine the composition of an unknown sample, while PIGE is used when the composition of the sample is more or less known and a γ-spectrum can be used to quantify the concentration or determine a concentration profile. PIXE (see **PIXE (Particle-induced X-ray Emission)**) is a technique that is generally used to determine elements

with $Z > 12$. PIGE is often used as a complementary technique to PIXE to determine light elements.

Although PIGE is regarded as a less general and versatile technique compared to PIXE, PIGE can have strong advantages. This is particularly the case for light elements ($Z < 11$) that cannot be analyzed or are hard to analyze with PIXE. Moreover, PIGE produces totally different spectra for elements of neighboring Z, removing peak-overlap problems between neighboring elements in the spectra. PIGE is only sensitive to specific isotopes, limiting the general applicability. It is, however, a powerful tool when isotopic analysis is required. Also, compared to similar PIXE spectra, minimal pile-up problems are present in PIGE spectra when there is no significant background radiation. Although PIGE is typically less sensitive than PIXE, the peaks in a PIGE spectrum are well separated and there is no need for absorption corrections for the γ-rays leaving the sample. This simplifies the experimental arrangement and the analysis of the spectra and can make the technique more attractive.

2 BASIC NUCLEAR PHYSICS

2.1 Principles of Nuclear Reactions

PIGE is based on nuclear reactions induced by accelerated beams that hit the sample of interest. After the reaction, nuclei in the sample remain in the excited states and decay by emitting γ-rays. The energies of these γ-rays are characteristic of the elements (isotopes). The two measured quantities thus yield the elemental identification and concentration, respectively. The technique is thus quantitative, or semi-quantitative when standards are used.

For a nuclear reaction to occur between a beam particle and a sample nucleus it is necessary that the beam energy exceeds the Coulomb barrier. The barrier energy (in MeV) can be estimated from Equations (1) and (2):

$$E_c = 1.44 \frac{Z_1 Z_2}{d} \tag{1}$$

$$\text{with} \qquad d = 1.25 \;\; A_1^{1/3} + A_2^{1/3} \tag{2}$$

where Z_1 and Z_2 are the proton numbers of the beam particles and the sample nuclei, respectively. The quantity d in fm (1 fm $= 10^{-15}$ m) is approximately the sum of the two respective radii, with A_1 and A_2 the respective mass numbers. The approximation is based on the notion that nuclear forces have very short ranges ($\sim$1 fm). Thus the two colliding nuclei must almost touch each other. If the beam energy is smaller than E_c then the beam particle will be deflected by the nucleus. γ-Rays can still be produced in this case by Coulomb excitation of the nucleus. This, however, is not very likely for light ion beams and will not be discussed further because light-ion induced reactions will be treated as applying most frequently. These reactions encompass inelastic scattering, capture (or compound nuclear) reactions and particle transfer. For protons they are denoted as: (p,p$'\gamma$), (p,γ) and, for instance, (p,$\alpha\gamma$), respectively.

At low energy, $< E_c$, only atomic excitations occur, giving rise to the emission of X-rays. This process is exploited in the PIXE method, as discussed in **PIXE (Particle-induced X-ray Emission)**.

Equation (1) results in $E_c = 1.2$ MeV for protons on ^{7}Li, 4 MeV on ^{28}Si, 5.2 MeV on ^{40}Ca and 13.6 MeV on ^{208}Pb. For α-particles on ^{28}Si, a beam energy of at least 7 MeV is required. It is concluded that a proton beam energy of at least 1 MeV is required and single-ended Van de Graaff accelerators, tandem Van de Graaff accelerators and cyclotrons are used for PIGE experiments.

The most commonly applied nuclear process in PIGE is the compound nuclear reaction, schematically written as Equation (3)

$$\mathrm{a} + \mathrm{A} \longrightarrow \mathrm{C}^* \longrightarrow \mathrm{B} + \mathrm{b}, \tag{3}$$

with a representing the beam particle, A the target nucleus, C* the compound nucleus (CN) in excited state that may decay by γ-rays or form a final product nucleus B and particle b. The product nucleus B, when formed in an excited state, B*, decays by γ-rays. The latter γ-rays are called reaction γ-rays, while those of C* are γ-rays of the CN. Those γ-rays are characteristic of the sample nucleus A (see Figure 1), in particular from the well-known discrete low-lying states.

The CN model is based on the assumption that the probability of decay into any specific set of final products is independent of the means of formation of the CN. The decay of the CN level only depends on its energy. If the excitation energy is above the particle binding energy then it surely will decay via the particle channel, because particle decay, owing to its strong interaction, is much more likely than γ-ray decay with its electromagnetic interaction.

The excitation energy, E_x, of the CN C* can be calculated from Equations (4) and (5):

$$E_x = Q + \frac{E_p m_A}{m_A + m_a} \tag{4}$$

$$\text{with} \qquad Q = [m_C - (m_A + m_a)]c^2 \tag{5}$$

with Q the reaction Q value in MeV, when the masses are expressed in atomic mass units, u, and the speed of light squared as $c^2 = 931.5\ \mathrm{MeV\,u^{-1}}$. In exothermic reactions mass is converted into energy and Q is positive, while

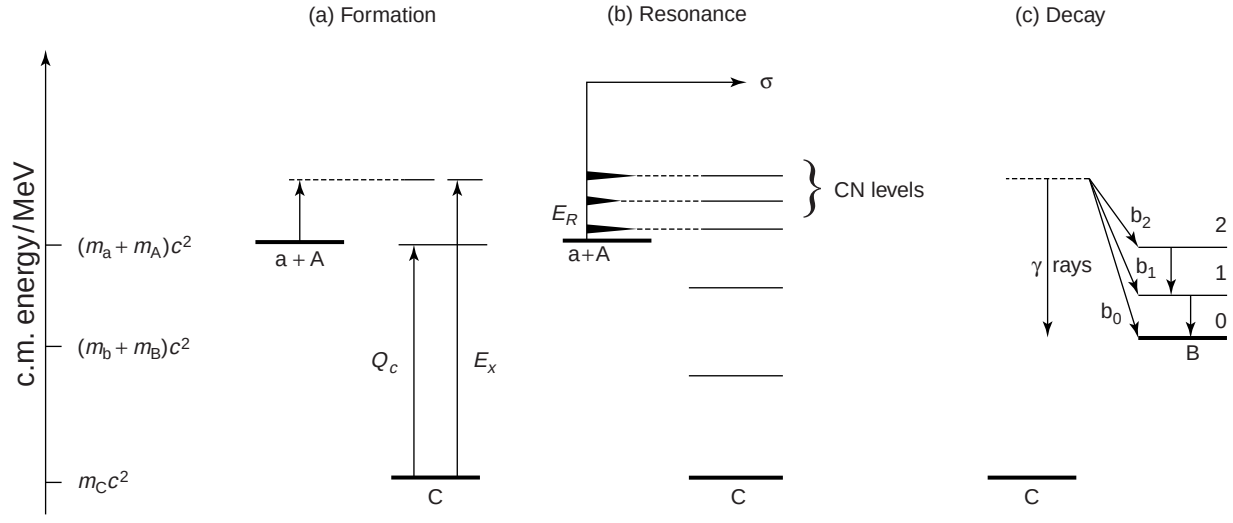

Figure 1 Formation (a) of CN C* in an excited state (CN level) (b) and its decay either by γ-rays or by particle emission b (c) to form the reaction product B* in excited states or in the ground state B. In (b) resonance levels are indicated with their schematic cross-sections σ. On the vertical axis the center-of-mass (c.m.) energy is indicated. (The figure is adapted from Deconninck.[1])

in endothermic reactions Q is negative. In Equation (4) the laboratory beam energy E_p is corrected by the mass factor of the two initial particles to obtain the energy in the center-of-mass system as an internal energy. The external energy or kinetic energy, E_{kin}, of the center-of-mass itself is written as Equation (6)

$$E_{kin} = \frac{E_p m_a}{m_A + m_a} \tag{6}$$

This energy is completely transferred to the recoiling CN as kinetic energy. The recoil energy may be of importance if the nuclear states involved in the decay of the CN have lifetimes shorter than the slowing down time of the recoil in the sample. In that case the γ-rays are emitted in flight and may show a shift and broadening in energy (see section 2.2).

Another important factor in PIGE analysis is the yield of γ-rays, which is directly connected to the reaction and emission probabilities. Here the reaction probability will be treated, while γ-ray emission is discussed in section 2.2.

The reaction probability is commonly expressed by the reaction cross-section σ, which represents a fictive surface at the nucleus through which the initial particle has to penetrate to induce a particular reaction. In many cases the cross-section exhibits a resonance-like behavior as a function of beam energy (Breit–Wigner shape), such as in proton capture or (p,γ) reactions at relatively low energy. The width Γ (in eV) of the resonance is a measure of the probability of forming that particular CN level and of decaying via γ-ray emission. The width Γ is related to the lifetime, τ, of the CN level via Equation (7), the uncertainty relation of Heisenberg:

$$\Gamma\tau = \frac{h}{2\pi} \tag{7}$$

To distinguish the formation of the CN level from the decay probability, Γ_p is introduced as the formation width (or probability) and Γ_γ as the decay width. The reaction cross-section can then be expressed as a function of beam energy in Equation (8)

$$\sigma_{p,\gamma}(E) = \pi\lambda\!\!\!^{-2} g \frac{\Gamma_p \Gamma_\gamma}{(E - E_R) + \Gamma^2/4} \tag{8}$$

Here E_R is the resonance energy at the maximum of the cross-section, $\lambda\!\!\!^{-}$ is the reduced wavelength of the initial particle and g is a statistical factor, determined by the spins, I, j and J of the nuclear particles A, a, and C*, respectively, Equation (9):

$$g = \frac{2J + 1}{(2j + 1)(2I + 1)} \tag{9}$$

For details see Deconninck.[1] If the resolution of the beam is larger than the total width Γ, the total area I_m under the resonance is obtained by integrating the cross-section over a broad energy range around the resonance, Equation (10):

$$I_m = 2\pi\lambda\!\!\!^{-2} g \frac{\Gamma_p \Gamma_\gamma}{\Gamma} \tag{10}$$

A Breit–Wigner resonance and its integral are shown in Figure 2.

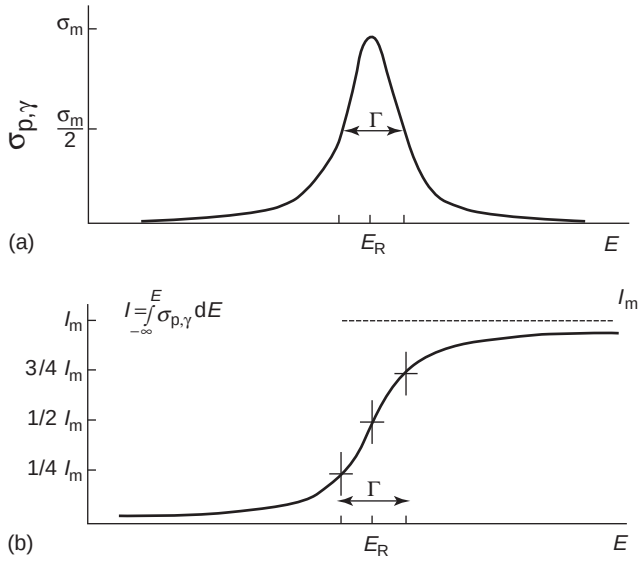

Figure 2 Breit–Wigner resonance shape curve (a) and its integral curve (b) are illustrated along with some of the relevant parameters discussed in the text. The maximum cross-section at the resonance energy E_R is indicated by σ_m. (The figure is taken from Deconninck.[1])

In the case where Γ is small (keV) the factor $(2J+1)\Gamma_P\Gamma_\gamma/\Gamma$ is proportional to the integrated cross-section. Thus the factor is proportional to the sensitivity of the resonance reaction. The total width Γ is contained in several nuclear data tables for many resonance levels in nuclei.[9–11] From this it can be concluded that the value of the cross-section is not only important for sensitivity, but also for knowledge of resonance structures (Equation 8) and, for practical applications, for the beam energy spread.

There are many other possible nuclear reactions yielding γ-rays that are resonant or nonresonant. Examples often used for elemental analysis are the inelastic scattering (p,p′γ), and further (p,$\alpha\gamma$) and (p,nγ) with protons. With other beam particles, commonly used reactions are (d,pγ), (d,nγ) and (α,nγ). At relatively low beam energies, in the few megaelectron volts region, the (p,p′γ) and the (p,$\alpha\gamma$) reactions also show resonant structures. Therefore Equation (8) can be used to find the cross-section by replacing Γ_γ for the exit channel by Γ_p and Γ_α, respectively. In the cases where the cross-section does not show resonant behavior, but varies only slowly with beam energy, the cross-section can be taken to be constant or averaged over the relevant energy region.

For nonresonant reactions, Equation (11) can be used to calculate the γ-ray yield $Y(E_0)$ at an incident energy E_0 of particle x on a target containing element A with atomic mass A_A (in g mol^{-1}), concentration c_A and with a cross-section $\sigma_{x,\gamma}(E)$, according to Equation (8), to produce particular γ rays:

$$Y(E_0) = N_p c_A A_A^{-1} N_{av} e(E_\gamma) \frac{\Omega}{4\pi} \int_{E_0}^{E_f} \frac{\sigma_{x,\gamma}(E)}{\varepsilon(E)}\, dE \qquad (11)$$

where N_p is the number of particles x hitting the target, which can be deduced from the measured beam current and time. N_{av} is the Avogadro number, $e(E_\gamma)$ is the efficiency of the detector for the particular γ-ray and Ω is the solid angle of the detector (in steradians). The cross-section $\sigma_{x,\gamma}(E)$ (cm^2, note 1 barn $= 10^{-24}$ cm^2) and the stopping cross-section $\varepsilon(E)$ (MeV g^{-1} cm^2) are integrated over the energy from E_0 to E_f, the energy of the particle x when leaving the target at the back side. Equation (11) also represents the thick-target yield, by putting $E_f = 0$, corresponding to the incident particle coming to a complete stop in the sample. The thick-target yield in literature is often expressed in units of $(\mu C\,sr)^{-1}$, that is in counts per quantity of electric charge on the target and per unit of solid angle. This can be derived easily from Equation (11) by dividing out N_p (expressed as charge) and Ω. The cross-sections and the stopping cross-sections of particle x slowed down in the sample have been tabulated in various publications for many particles, energies and stopping materials (see e.g. Nuclear Data Tables[12] for reaction cross-sections and Ziegler et al.[13] for stopping powers). If they are not tabulated then the known values must be interpolated or, in the worse case, measured. In Equation (11) the attenuation of γ-rays in the target, in possible absorbers and in the detector window is neglected, but it must be taken into account when γ-ray energies are low.

For detailed formulas in cases of sharp and broad resonances the reader is referred to Deconninck.[1] To circumvent the problems involved in the use of absolute cross-sections and stopping powers, internal or external standards are often used to compare with the measurement on the real sample. In the case where internal standard elements are added, their concentrations and cross-sections must be known precisely. In the case of an external standard, a sample is prepared with exactly the same elements as the sample of interest, at known concentrations. In the first case only one measurement is needed, but integrations for the different elements (Equation 11) are still necessary. However, beam current and detector solid angle are canceled. In the second case the integrations are canceled, but beam current (and solid angle, if different) need to be measured.

For thin samples the cross-section can be considered constant. It is convenient to add a standard element with mass number A_s and with known concentration c_s to the sample. When comparing the γ-ray yields of element A and standard S, not only are the solid angles and number of beam particles canceled, but so too are the stopping powers because in both cases they are the same. The

unknown concentration c_A can easily be calculated from the measured γ-ray yields from Equation (12):

$$c_A = c_s \frac{Y_A \sigma_s e_s A_A}{Y_s \sigma_A e_A A_s} \tag{12}$$

2.2 γ-Ray Emission

γ-Ray emission is governed by the electromagnetic interaction and transitions are described by connecting the initial and final nuclear quantum states via the electromagnetic multipole operator. The electromagnetic field can be expanded in terms of multipoles, i.e. monopole, dipole, quadrupole, octupole, and so on. They can be of electric or magnetic character, with certain constraints. The multipole moments have the form of Equation (13):

$$M_l \propto r^l Y_l^m(\theta, \phi) \tag{13}$$

here l represents the order of the multipole moment and $Y_l^m(\theta, \phi)$ is the spherical harmonic function. The electromagnetic operator, which acts on the initial and final states involved and gives rise to the γ-ray transition, also has the basic form of Equation (13). For a detailed treatment of the electromagnetic features important for γ-ray emission the reader is referred to text books, for example by Krane[14] and Ejiri and de Voigt.[15] Here we will restrict ourselves to the more practical rules that determine the nature and intensity of γ-rays.

The electromagnetic transition probability $T(E(M), l)$, according to Equation (7) is given as Equation (14):

$$T(E, (M), l, if) = \frac{\Gamma(E, (M), l, if)}{\hbar} = \frac{\Gamma_\gamma}{\hbar} = \frac{1}{\tau} \tag{14}$$

The radiation width Γ_γ of Equation (7) is thus equivalent to the transition probability T, which depends on the electric (E) or magnetic (M) character, on the multipolarity l of the radiation and on the structure of the initial and final nuclear states (if). The multipolarity l of the transition corresponds to the order l of the multipole moment and of the multipole operator (Equation 13). The transition probability depends on external factors, i.e. the energy E_γ of the transition, the multipolarity l and the nuclear radius R, as well as on the internal factors (i.e. the nuclear structure of the initial and final states). The latter information is not always available and varies strongly from nucleus to nucleus and even within a nucleus from state to state. Therefore we will discuss the general consequences of the external factors and refer the reader to the various nuclear data compilations for the internal factors; i.e. for nuclei with mass 21–44 to the compilation of Endt[10] (and references quoted therein) and for nuclei below mass 21 to the compilation of Ajzenberg–Selove[11] (and references quoted therein).

The importance of the external and internal factors is that they also determine the value of Γ_γ and thereby the cross-section $\sigma_{x,\gamma}$, according to Equation (8) and thus the yield $Y(E_0)$, according to Equation (11). A simplified indication for the external factor T_{ex} can be obtained from Equations (15) and (16):

$$T_{ex}(E) \propto \left(\frac{E_\gamma}{\hbar c}\right)^{2l+1} R^{2l} \tag{15}$$

and

$$T_{ex}(M) \propto \left(\frac{E_\gamma}{\hbar c}\right)^{2l+1} R^{2l-2} \tag{16}$$

Noting that $\hbar c = 197\,\text{MeV fm}$, that E_γ is of the order of 1 MeV and that the nuclear radius is about 3 fm for light nuclei, one concludes that the transition probability decreases rather strongly with increasing multipolarity of the radiation. For each unit of l the difference is about a factor of 10^4. The electric transitions are more probable than the magnetic ones of the same multipolarity by a factor of R^2, which is about a factor of 10 for light nuclei. Note that these are only rough estimates and that the internal factor may change the picture quite drastically. For instance in collectively rotating nuclei, in the strongly deformed regions, the electric quadrupole (E_2) transition is strongly enhanced. Nevertheless, it is probably safe to state that in general the sequence of probability for transitions to occur is E_1, M_1, E_2, M_2, E_3, M_3, E_4, M_4, and so on. The transition probability increases strongly with the transition energy, particularly for the higher multipolarities. This is illustrated in Figure 3, where absolute transition probabilities are given for one-particle transitions, i.e. transitions caused by one (valence) proton in the nucleus.

The energy of the γ-ray is simply the energy difference between the initial and final states, Equation (17):

$$E_\gamma = E_i - E_f, \tag{17}$$

here E_i and E_f represent the excitation energies of the initial and the final states, respectively. In this formula the recoil energy of the de-exciting nucleus is neglected. It can be calculated from the γ-ray energy and mass, M, of the nucleus as Equation (18):

$$E_{rec} = \frac{E_\gamma^2}{2Mc^2} \tag{18}$$

Noting that the numerator is of the order of megaelectron volts and the denominator of gigaelectron volts, E_{rec} is only about $10^{-3} E_\gamma$, thus mostly of the order of kiloelectron volts, which can be neglected.

The multipolarity of the γ-radiation is restricted by the spins (and parities) of the initial and final states I_i (π_i)

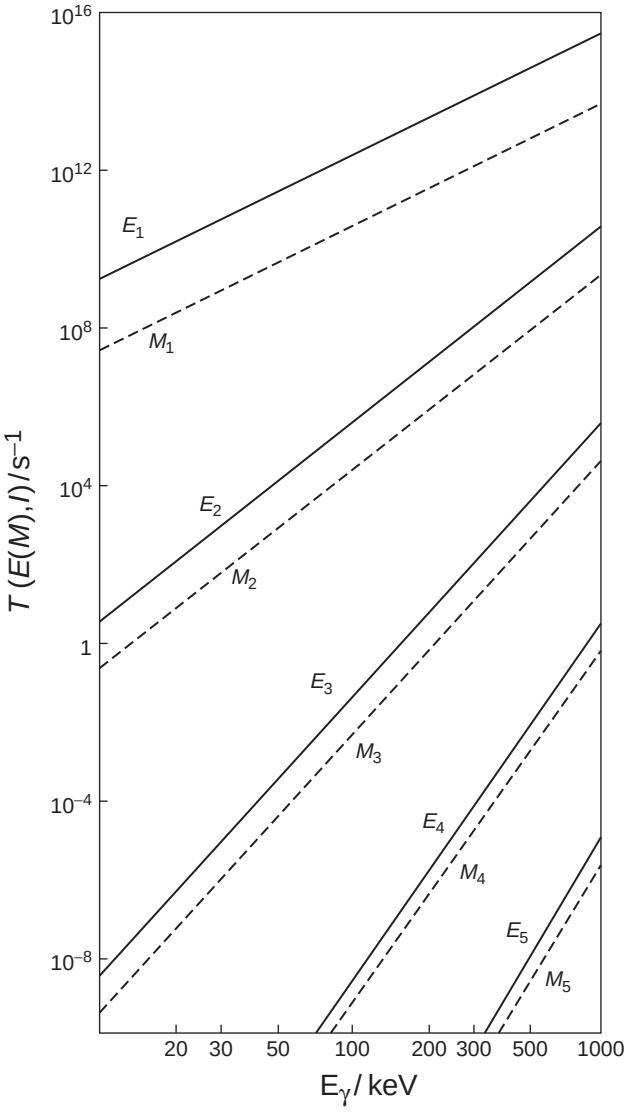

Figure 3 One-particle transition probabilities as functions of the γ-ray energies for various multipole radiations.

and I_f (π_f), respectively. The multipole radiation of order l (Equation 13) carries l units ($\hbar$) of angular momentum per photon. Conservation of angular momentum and the notion that its couplings are vectorwise leads to the rule, Equation (19):

$$|I_i - I_f| \leq l \leq I_i + I_f \tag{19}$$

It thus appears that for certain spin combinations more than one multipolarity is allowed. In that case the lowest allowed multipolarity is the most probable one with mixing of the next higher one, for which l is one unit larger. The parities of the initial and final states, π_i and π_f, are also important for the electric or magnetic character. Conservation of parity, with π_l the parity of the emitted γ-ray with multipolarity (angular momentum) l, leads to the rule, Equation (20):

$$\pi_l = \pi_i \pi_f \tag{20}$$

The parity π_l of electric transitions is determined by Equation (21):

$$\pi_l(E) = (-1)^l \tag{21}$$

The parity π_l of magnetic transitions is determined by Equation (22):

$$\pi_l(M) = (-1)^{l+1} \tag{22}$$

The consequences of the latter two rules are that electric transitions with even multipolarity l and magnetic transitions with odd l have even parity; thus M_1, E_2, M_3, E_4, M_5, E_6, and so on have even parity. Consequently E_1, M_2, E_3, M_4, E_5, M_6, and so on have odd parity. With the rules shown in Equations (19)–(22) and from the known spins and parities of the initial and final states the character of the radiation can be determined and thereby the probability of the radiation and its intensity as far as these factors are concerned. The following examples will illustrate the simplicity of the rules:

Transition
$1^+ \rightarrow 0^+$ M_1 radiation only possible
$1^- \rightarrow 0^+$ E_1
$2^+ \rightarrow 0^+$ E_2
$3^+ \rightarrow 2^+$ $M_1, E_2, \ldots, M_5$. In practice only M_1, E_2 mixed.
$1^- \rightarrow 2^+$ E_1, M_2, E_3. In practice only E_1, M_2 mixed.

Note that the transition $0^+ \rightarrow 0^+$ cannot proceed via γ-radiation because the intrinsic spin of a photon is 1^-. In this particular case the transition is fully converted to conversion electrons. All other transitions are also for a minor fraction converted, but the conversion coefficients are rather small, particularly for low multipolarity radiation and high transition energies. For instance, at 300 keV for E_1 radiation the conversion is already below 1%, but for M_4 radiation, electrons and γ-rays have about the same probability. For further detail see, for example, Krane[14] and Ejiri and de Voigt.[15]

It should be realized that γ-rays generally are not emitted isotropically in space. Asymmetry can be as much as 50%. The shape of the angular distribution depends on the angular momentum transfer in the reaction and on the population of the initial state and the final nucleus, as well as on the character (multipolarity) of the radiation. Quadrupole radiation, for instance, has its maximum along the direction of the beam, while dipole radiation peaks in directions perpendicular to the beam direction. For absolute values, measured angular distributions must be used or can be calculated on the basis of known

or assumed angular momentum alignments (see e.g. Krane[14]).

Finally, the energy shift and broadening of the γ-ray lines in the spectrum, which are due to the Doppler effect, should be mentioned. When a γ-ray is emitted during the flight of the excited nucleus through the sample or in vacuum behind the sample, the energy measured will be Doppler shifted. This is the case when the lifetime of the nuclear excited state is smaller than the average slowing down time of the recoiling nucleus in the sample. This time depends on the initial velocity of the recoil and on the slowing down material, but generally has values in the range of 10^{-8}–10^{-14} s. The measured γ-ray energy is then given by Equation (23):

$$E_\gamma = E_{\gamma 0} \quad 1 + \frac{v}{c}\cos\theta \tag{23}$$

where $E_{\gamma 0}$ is the unshifted energy, θ the angle of observation with respect of the beam direction and v is the velocity of the recoiling nucleus. In the case of compound nuclear reactions the initial velocity can be calculated easily from Equation (6) as Equation (24):

$$\frac{v}{c} = 0.046\,35\frac{\sqrt{m_a E_p}}{m_A + m_a} \tag{24}$$

The masses are given in mass numbers and the energy in MeV, resulting in the ratio v/c. Almost all nuclear lifetimes of compound states, mostly highly excited states, are very short, of the order of 10^{-14}–10^{-15} s. The corresponding de-exciting γ-rays are thus fully Doppler shifted and broadened, depending on the angle of emission and on the detector solid angle. Low-lying excited states may have much longer lifetimes and the corresponding γ-rays will not be affected by the Doppler effect. The resulting sharp lines are generally more useful for elemental analysis.

2.3 Principles of Depth Profiling

Charged-particle induced resonance reactions offer the possibility of obtaining the depth in the sample from which the γ-radiation is emitted and thus determining the depth profile of the elements. The leading principle is based on the fact that the beam particles are slowed down in the sample and thus the beam energy has to be increased to match the value of the resonance. The amount of increase ΔE (keV) depends on the depth Δx (cm) and the two quantities are related via the stopping power S (keV cm^{-1}) by Equation (25):

$$\Delta x = \frac{\Delta E}{S} \tag{25}$$

In compilations (see e.g. Ziegler et al.[13]) the stopping cross-section ε is often given in units of keV atom^{-1} cm^2. Here ε is defined by Equation (26) as:

$$\varepsilon = \frac{\Delta E}{N\Delta x} \tag{26}$$

where N is the atomic density in the sample (at cm^{-3}). It should be noted that the depth resolution depends on the straggling in the sample, on the resonance width and on the spread in beam energy. It is therefore important to select resonances that are as narrow as the beam spread (keV).

Thus, the depth Δx can be calculated from the difference between the resonance and the beam energy, the stopping cross-section and the atomic density. The atomic density can be expressed in terms of the mass density ρ (g cm^{-3}) as Equation (27):

$$N = \frac{\rho N_{av}}{M} \tag{27}$$

where N_{av} is the Avogadro number (6.022045×10^{23} mol^{-1}) and M is the molecular or atomic weight.

The stopping cross-sections are tabulated (for example in Ziegler et al.[13]) for elements. In the case of a compound a simple rule called Braggs rule may be applied, Equation (28):

$$\varepsilon = \sum_i c_i \varepsilon_i \tag{28}$$

where c_i are the mass concentrations or the atomic fractions in a molecule.

3 PARTICLE-INDUCED γ-RAY EMISSION

3.1 Basic Principles of Particle-induced γ-Ray Emission

In PIGE, samples are bombarded with a beam consisting of light particles (protons, deuterons, tritons, ^{3}He or ^{4}He (α)) or heavy ions. The energy of the incident particles is usually in the range of 1–4 MeV amu^{-1}. The beam enters a vacuum chamber, in which the sample is placed, through a set of diaphragms and the emitted γ-rays are detected by, for example, a hyperpure Ge or Ge(Li) detector. The pulses from the detector are analyzed by a multichannel analyzer (MCA). A more detailed description of the experimental set-up can be found in section 3.5. A typical PIGE energy spectrum is shown in Figure 4.

This type of energy spectrum consists of a number of peaks corresponding to the γ-rays related to the elements present in the sample. These characteristic peaks are superimposed on a background that originates from Compton scattering of γ-rays in the detector. The number of counts in a peak of a certain element corresponds

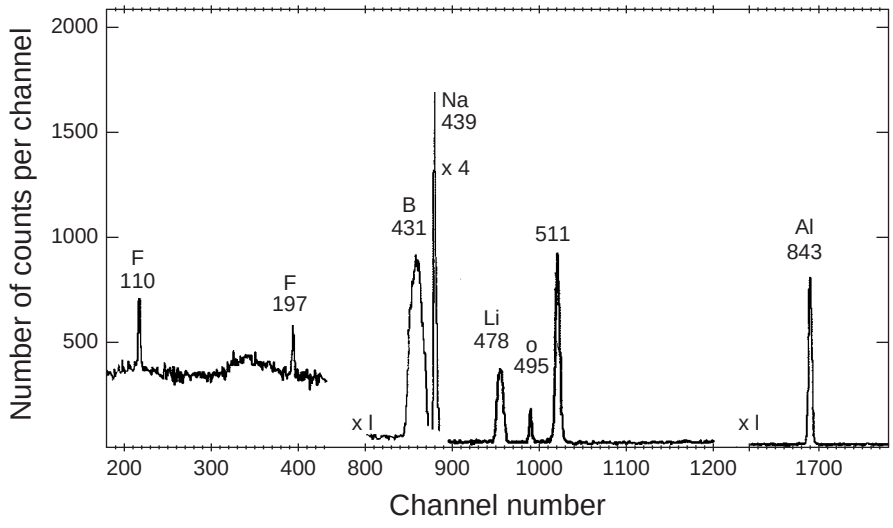

Figure 4 Example of a PIGE energy spectrum of a potassium feldspar (ST-2B) sample. The sample was measured with 1.8 MeV protons. Energies marked on peaks are in keV.[8]

to the amount of that element in the sample under investigation. All the parameters that determine the number of counts in a certain peak are known or can be measured. Using computer programs and databases, an absolute determination of the amounts of the elements in the sample is possible.

A complication in the spectra can be that the same γ-rays could result from two different reactions. For example, consider the reactions $^{26}Mg(p,\gamma)^{27}Al$ and the $^{27}Al(p,p'\gamma)^{27}Al$ which both give γ-rays with an energy of 843 keV and 1023 keV. Magnesium should then be identified by the $^{25}Mg(p,p'\gamma)^{25}Mg$ reaction which produces γ-rays of 586 keV. Aluminum can be identified by the $^{27}Al(p,\gamma)^{28}Si$ reaction with γ-rays of 1778 and 2836 keV.

Prompt γ-rays can originate from different types of nuclear reactions such as $(p,p'\gamma)$, $(p,\alpha\gamma)$, or (p,γ) for protons, and $(d,n\gamma)$, $(\alpha,\alpha'\gamma)$ and so on for other projectiles. When an element consists of more than one isotope, γ-rays can be produced in any of the isotopes and this will show up in the accompanying spectra as more lines belonging to one element but different isotopes. This leads to isotopic determination.

3.2 Possible Reactions and Their γ-Ray Energies

In a PIGE spectrum, the peaks are characteristic of different elements or even different isotopes of elements present in the sample under investigation. In Table 1, the most intense proton reactions together with the energy of the emitted γ-rays are presented for elements with $4 < Z < 20$. In Table 2, the most intense proton reactions for a number of heavy element isotopes are presented. The elemental contents are obtained from isotopic contents, taking the natural abundances into account.

3.3 Sensitivity

A general discussion about the sensitivity of PIGE for different kinds of sample is difficult to give since it depends on the content of the major elements (the matrix), the minor elements and the trace elements in the sample. There are several ways to define the sensitivity of an analytical method.

The first and commonly used way to define sensitivity is by the minimal detectable concentration or limit of detection (LOD). It is defined by Equation (29)

$$\mathrm{LOD} = k\sigma_b, \tag{29}$$

where σ_b is the standard deviation of the background under a (certain interval of a) peak of interest in a spectrum, and k is a factor reflecting the desired confidence level.[19] It can be interpreted more easily with the help of Figure 5.

In this figure, the energy interval ΔE is taken as the detector resolution and is equal to the full width at half-maximum (fwhm) of the peak. I_Z is the number of counts in the peak and I_b is the number of counts in the background under the peak. Both are the number of counts in the energy interval ΔE. The standard deviation σ_b is taken to be the standard deviation of the background counts I_b. The confidence level $k = 3$, gives a confidence level of 99.98% for normally distributed measurements. This means that a peak is considered to be detectable if, Equation (30):

$$I_Z \geq 3\sqrt{I_b} \tag{30}$$

Table 1 Most intense possible proton reactions for light element isotopes with $4 < Z < 20$, together with the energy of the emitted γ-rays[16,17]

Isotope	Reaction	E_γ	Isotope	Reaction	E_γ	Isotope	Reaction	E_γ
^{7}Li	^{7}Li(p,nγ)^{7}Be	429	^{25}Mg	^{25}Mg(p,p′γ)^{25}Mg	390	^{33}S	^{33}S(p,p′γ)^{33}S	841
	^{7}Li(p,p′γ)^{7}Li	478		,,	585		,,	1472
^{9}Be	^{9}Be(p,γ)^{10}B	415		,,	975	^{34}S	^{34}S(p,γ)^{35}Cl	1219
	,,	718		,,	990		^{34}S(p,p′γ)^{34}S	2127
	,,	1022		,,	1380	^{35}Cl	^{35}Cl(p,p′γ)^{35}Cl	1219
	,,	7477		,,	1612		,,	1763
	^{9}Be(p,αγ)^{6}Li	3562		,,	1965		^{35}Cl(p,γ)^{36}Ar	1970
^{10}B	^{10}B(p,αγ)^{7}Be	429		^{25}Mg(p,γ)^{26}Al	417		,,	2208
	^{10}B(p,p′γ)^{10}B	718	^{26}Mg	^{26}Mg(p,γ)^{27}Al	844	^{37}Cl	^{37}Cl(p,nγ)^{37}Ar	1410
^{11}B	^{11}B(p,p′γ)^{11}B	2125		,,	1014		,,	1611
^{12}C	^{12}C(p,γ)^{13}N	2366		^{26}Mg(p,p′γ)^{26}Mg	1809		^{37}Cl(p,p′γ)^{37}Cl	1727
	,,	3511	^{27}Al	^{27}Al(p,p′γ)^{27}Al	171		^{37}Cl(p,αγ)^{38}Ar	2127
^{13}C	^{13}C(p,γ)^{14}N	1635		,,	844		^{37}Cl(p,γ)^{38}Ar	670
	,,	2313		,,	1014		,,	1642
	,,	8062		,,	1720		,,	2168
	^{13}C(p,p′γ)^{13}C	3089		,,	2211	^{39}K	^{39}K(p,γ)^{40}Ca	755
^{14}N	^{14}N(p,p′γ)^{14}N	2313		,,	2734		^{39}K(p,nγ)^{41}Ca	2522
	^{14}N(p,γ)^{15}O	6793		,,	2981	^{41}K	^{41}K(p,p′γ)^{41}K	980
^{15}N	^{15}N(p,αγ)^{12}C	4439		,,	3004		,,	1294
^{16}O	^{16}O(p,γ)^{17}F	495		^{27}Al(p,γ)^{28}Si	1779		^{41}K(p,γ)^{42}Ca	313
^{17}O	^{17}O(p,p′γ)^{17}O	871		,,	2839		,,	899
^{18}O	^{18}O(p,p′γ)^{18}O	1982		^{27}Al(p,αγ)^{24}Mg	1369		,,	1525
^{19}F	^{19}F(p,p′γ)^{19}F	110	^{28}Si	^{28}Si(p,γ)^{29}P	1384		^{41}K(p,αγ)^{38}Ar	2168
	,,	197		,,	4343		^{41}K(p,nγ)^{41}Ca	1943
	,,	1236		^{28}Si(p,pγ)^{28}Si	1779		,,	2010
	,,	1349	^{29}Si	^{29}Si(p,γ)^{30}P	677		,,	2576
	,,	1357		,,	709		,,	2604
	,,	1459		^{29}Si(p,p′γ)^{29}Si	755	^{42}Ca	^{42}Ca(p,p′γ)^{42}Ca	1525
	^{19}F(p,αγ)^{16}O	6129		,,	1273	^{43}Ca	^{43}Ca(p,p′γ)^{43}Ca	373
	,,	6917		,,	2028	^{44}Ca	^{44}Ca(p,p′γ)^{44}Ca	1157
	,,	7114	^{30}Si	^{30}Si(p,γ)^{31}P	1266		^{44}Ca(p,γ)^{45}Sc	364
^{23}Na	^{23}Na(p,p′γ)^{23}Na	440		,,	2233		,,	531
	,,	1636		^{30}Si(p,p′γ)^{30}Si	2235		,,	543
	,,	1951	^{31}P	^{31}P(p,p′γ)^{31}P	1266		,,	720
	,,	2391		,,	2233	^{48}Ca	^{48}Ca(p,nγ)^{48}Sc	371
	^{23}Na(p,αγ)^{20}Ne	1634		^{31}P(p,γ)^{32}S	2230		,,	520
	^{23}Na(p,γ)^{24}Mg	1369		^{31}P(p,αγ)^{28}Si	1779		,,	779
	,,	4238	^{32}S	^{32}S(p,γ)^{33}Cl	811			
^{24}Mg	^{24}Mg(p,γ)^{25}Al	452		,,	2035			
	^{24}Mg(p,p′γ)^{24}Mg	1369		^{32}S(p,p′γ)^{32}S	2230			

Systematic studies concerning sensitivities have been performed by Deconninck and Demortier[5,6] with protons, Clark et al.[20] with protons and α-particles, Borderie et al.[9,21] with tritons and α-particles, Giles and Peisach[7] with α-particles and Borderie et al.[22] with heavy ions. Of about 70 elements studied, 40 can be determined with high sensitivity (<100 ppm). Trace element sensitivity (ppm range) can be obtained for the light elements Li, Be, B, N, O, F, Na and P.[23]

In Figure 6, it can be seen that protons and α-particles are the best choice for the determination of light elements. Heavy ions can be used for the determination of medium and heavy elements at a level around 50 ppm. Accurate determination of concentrations in samples depends on the LOD and on the presence of major and minor elements in the sample. As a rule of thumb, accurate determinations can be performed at concentration levels around 10 times the LOD.

The second way to define sensitivity is to express it as the minimum absolute amount that can be detected. The absolute amount can be calculated from the concentration of an element if the areal mass density (g/cm^2) of the sample is known. The minimal amount of light elements is in the range 10^{-10}–10^{-11} g. The lower limit in the absolute amount is obtained with the ion microprobe technique (see section 3.5). Here, a beam diameter below 1 μm can

Table 2 Most intense possible proton reactions for elements (isotopes) with $Z > 30$ together with the energy of the emitted γ-rays[18]

Isotope	Reaction	E_γ	Isotope	Reaction	E_γ	Isotope	Reaction	E_γ
^{69}Ga	^{69}Ga(p,γ)^{70}Ge	175	^{86}Sr	^{86}Sr(p,γ)^{87}Y	793	^{111}Cd	^{111}Cd(p,p$'\gamma$)^{111}Cd	342
	,,	668	^{87}Sr	^{87}Sr(p,γ)^{88}Y	232	^{112}Cd	^{112}Cd(p,p$'\gamma$)^{112}Cd	617
	,,	1040	^{88}Sr	^{88}Sr(p,γ)^{89}Y	909	^{113}Cd	^{113}Cd(p,p$'\gamma$)^{113}Cd	298
	,,	1708		,,	1313	^{114}Cd	^{114}Cd(p,p$'\gamma$)^{114}Cd	558
	^{69}Ga(p,p$'\gamma$)^{69}Ga	318		,,	1507	^{115}In	^{115}In(p,nγ)^{115}In	115
^{71}Ga	^{71}Ga(p,nγ)^{71}Ge	175		,,	1745		,,	497
	,,	327	^{89}Y	^{89}Sr(p,γ)^{90}Zr	420	^{121}Sb	^{121}Sb(p,p$'\gamma$)^{121}Sb	508
	,,	391		,,	562		,,	573
	,,	500		,,	891	^{123}Sb	^{123}Sb(p,p$'\gamma$)^{121}Sb	160
	^{71}Ga(p,p$'\gamma$)^{71}Ga	391		,,	2186		,,	382
^{70}Ge	^{70}Ge(p,γ)^{71}As	147		,,	2319		,,	542
^{72}Ge	^{72}Ge(p,γ)^{73}As	254	^{90}Zr	^{90}Zr(p,γ)^{91}Nb	1082	^{126}Te	^{126}Te(p,p$'\gamma$)^{126}Te	666
	,,	361			1208	^{127}I	^{127}I(p,p$'\gamma$)^{127}I	145
^{73}Ge	^{73}Ge(p,nγ)^{73}As	254	^{92}Zr	^{92}Zr(p,γ)^{93}Nb	656		,,	172
	,,	361		,,	744		,,	203
^{74}Ge	^{74}Ge(p,γ)^{75}As	199	^{94}Zr	^{94}Zr(p,p$'\gamma$)^{94}Zr	919		,,	375
	,,	265	^{93}Nb	^{93}Nb(p,γ)^{94}Mo	449		,,	418
	,,	280		,,	703		,,	629
	,,	596		,,	742	^{181}Ta	^{181}Ta(p,p$'\gamma$)^{181}Ta	136
	^{74}Ge(p,p$'\gamma$)^{74}Ge	596		,,	850		,,	165
^{76}Ge	^{76}Ge(p,p$'\gamma$)^{76}Ge	563		,,	871		,,	302
^{75}As	^{75}As(p,p$'\gamma$)^{75}As	199		^{93}Nb(p,p$'\gamma$)^{93}Nb	742	^{182}W	^{182}W(p,p$'\gamma$)^{182}W	100
	,,	280	^{95}Mo	^{95}Mo(p,p$'\gamma$)^{95}Mo	204	^{183}W	^{183}W(p,p$'\gamma$)^{183}W	292
	,,	573	^{96}Mo	^{96}Mo(p,p$'\gamma$)^{96}Mo	778	^{184}W	^{184}W(p,p$'\gamma$)^{184}W	111
	^{75}As(p,γ)^{76}Se	559	^{97}Mo	^{97}Mo(p,p$'\gamma$)^{97}Mo	481	^{186}W	^{186}W(p,p$'\gamma$)^{186}W	123
	^{75}As(p,nγ)^{75}Se	112	^{98}Mo	^{98}Mo(p,p$'\gamma$)^{98}Mo	787	^{194}Pt	^{194}Pt(p,p$'\gamma$)^{194}Pt	329
	,,	287	^{100}Mo	^{100}Mo(p,p$'\gamma$)^{100}Mo	535	^{195}Pt	^{195}Pt(p,p$'\gamma$)^{195}Pt	211
^{76}Se	^{76}Se(p,p$'\gamma$)^{76}Se	559	^{99}Ru	^{99}Ru(p,p$'\gamma$)^{99}Ru	90		,,	239
^{77}Se	^{77}Se(p,p$'\gamma$)^{77}Se	239	^{100}Ru	^{100}Ru(p,p$'\gamma$)^{100}Ru	540	^{196}Pt	^{196}Pt(p,p$'\gamma$)^{196}Pt	356
^{78}Se	^{78}Se(p,p$'\gamma$)^{78}Se	614	^{101}Ru	^{101}Ru(p,p$'\gamma$)^{101}Ru	127	^{198}Pt	^{198}Pt(p,p$'\gamma$)^{198}Pt	407
^{80}Se	^{80}Se(p,p$'\gamma$)^{80}Se	666	^{102}Ru	^{102}Ru(p,p$'\gamma$)^{102}Ru	475	^{197}Au	^{197}Au(p,p$'\gamma$)^{197}Au	191
	,,	276	^{104}Ru	^{104}Ru(p,p$'\gamma$)^{104}Ru	358		,,	269
^{79}Br	^{79}Br(p,p$'\gamma$)^{79}Br	217	^{103}Rh	^{103}Rh(p,p$'\gamma$)^{103}Rh	295		,,	279
	,,	306		,,	357		,,	548
	,,	523	^{102}Pd	^{102}Pd(p,p$'\gamma$)^{102}Pd	556	^{198}Hg	^{198}Hg(p,p$'\gamma$)^{198}Hg	412
	^{79}Br(p,γ)^{80}Kr	617	^{104}Pd	^{104}Pd(p,p$'\gamma$)^{104}Pd	556	^{199}Hg	^{199}Hg(p,p$'\gamma$)^{199}Hg	158
^{81}Br	^{81}Br(p,p$'\gamma$)^{81}Br	276	^{105}Pd	^{105}Pd(p,p$'\gamma$)^{105}Pd	280		,,	208
	^{81}Br(p,nγ)^{81}Kr	190		,,	442	^{200}Hg	^{200}Hg(p,p$'\gamma$)^{200}Hg	368
^{85}Rb	^{85}Rb(p,p$'\gamma$)^{85}Rb	130	^{108}Pd	^{108}Pd(p,p$'\gamma$)^{108}Pd	434	^{203}Tl	^{203}Tl(p,p$'\gamma$)^{203}Tl	279
	,,	151	^{110}Pd	^{110}Pd(p,p$'\gamma$)^{110}Pd	374	^{205}Tl	^{205}Tl(p,p$'\gamma$)^{205}Tl	204
	^{85}Rb(p,γ)^{86}Sr	1077	^{107}Ag	^{107}Ag(p,p$'\gamma$)^{107}Ag	325		,,	416
	,,	1854		,,	423			
	^{85}Rb(p,nγ)^{85}Sr	232	^{109}Ag	^{109}Ag(p,p$'\gamma$)^{109}Ag	311			
^{87}Rb	^{87}Rb(p,nγ)^{87}Sr	485		,,	415			
	,,	1229	^{110}Cd	^{110}Cd(p,p$'\gamma$)^{110}Cd	658			

be used,[24] making detection of an absolute amount of 10^{-15}–10^{-16} g possible, although measurement times can be very high if concentration distributions have to be determined.

The third way is to use the thick-target yield (Equation 11) as a way to compare different elements, projectiles, energies and so on. The advantage of this method is that PIGE is most often used for the analysis of thick samples. A thick sample means a sample in which the incident particle is completely stopped. Generally, in thin samples the sensitivity is poor compared to thick samples. There is a strong energy dependence of the excitation curve, and the sample weight and composition should be known accurately.[25]

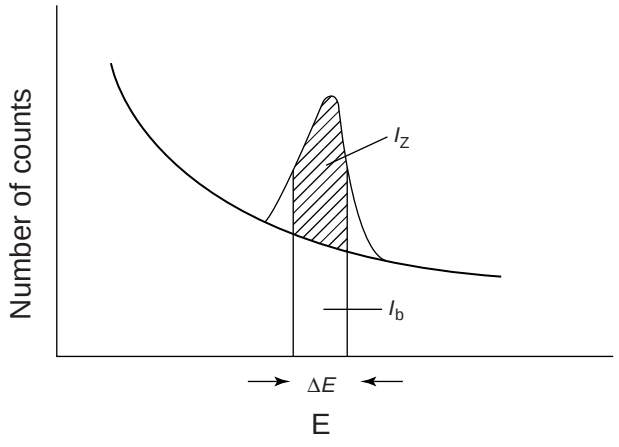

Figure 5 Limit of detection.

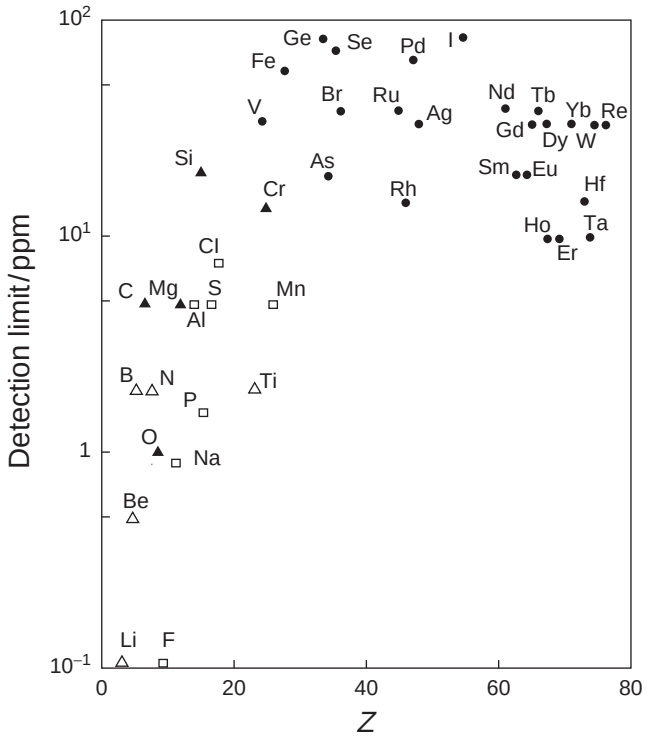

Figure 6 LOD of a number of elements using PIGE for different projectiles (adapted from Borderie[23] and references therein). ●, ^{35}Cl at 55 MeV; , p; , t; △, α; light projectiles at $E \leq 5$ MeV.

Several authors have published tables with thick-target yields.[7,16–18,25,26]

Kenny et al.[25] measured the absolute thick-target yield for several elements at incident proton energies of 2.0 and 2.5 MeV. They measured the yields for several γ-ray lines for elements ranging from F to Au. In Figure 7 the thick-target yield is plotted as a function of the atomic number Z for the strongest γ-ray line of an element. It can be seen that the yield, and thus the sensitivity, is highest for the light elements.

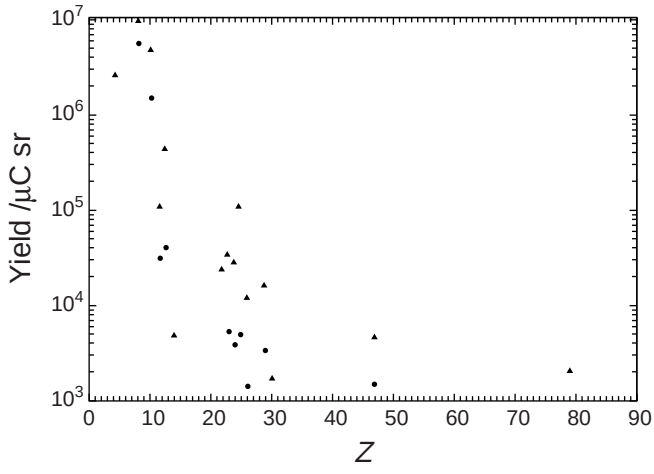

Figure 7 Thick-target yields of the strongest γ-rays for a proton energy E_p of 2 MeV(●) and 2.5 MeV( ). (The figure is drawn with data taken from Kenny et al.[25])

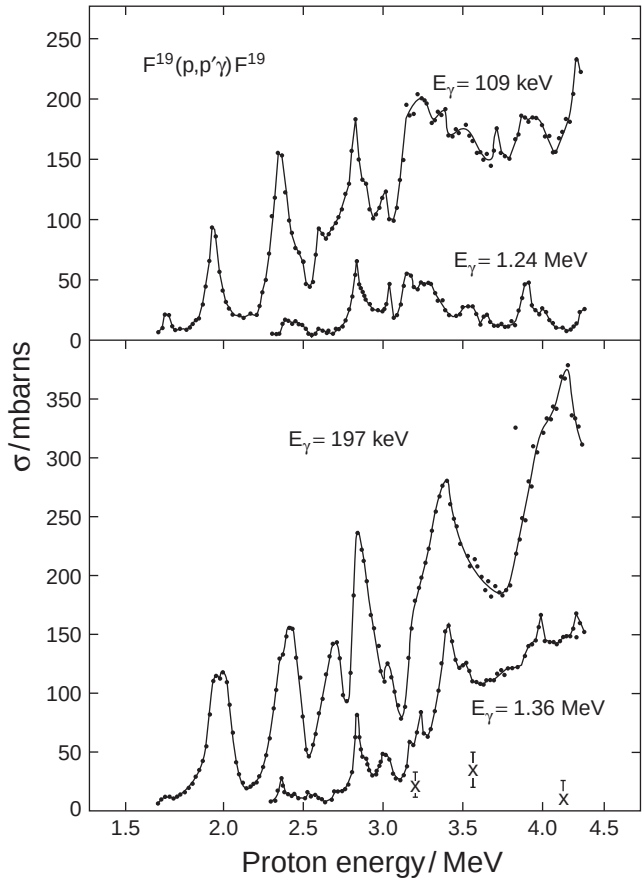

Figure 8 Cross-section for the production of 0.109, 0.197, 1.24 and 1.36 MeV γ-rays from the inelastic proton scattering reaction on fluorine (from Ranken et al.[27]).

Other datasets are presented by Anttila et al.[16] and Kiss et al.[17] They present thick-target yields for several proton energies between 1.0 and 4.2 MeV for isotopes with $4 < Z < 21$. Thick-target γ-ray yields for heavy elements ($Z > 30$) were determined by Räisänen and Hänninen[18] for 1.7 and 2.4 MeV protons. Thick-target γ-ray yields for light elements with 2.4 MeV He^+ are given by Lappalainen[26] and prompt γ-rays generated by 5 MeV α-particles were investigated by Giles and Peisach[7] who determined the sensitivity of elements ranging from lithium to hafnium.

In the following, a few examples of excitation curves (cross-section versus particle energy) for light elements will be given. In Figure 8 the cross-section for the $^{19}F(p,p'\gamma)^{19}F$ inelastic scattering reaction is given[27] for four different γ-rays (109 keV, 197 keV, 1.24 MeV and 1.36 MeV).

For magnesium, excitation functions of ^{24}Mg, ^{25}Mg and ^{26}Mg were determined by Boni et al.[28] Taking into account the isotopic abundances, the 1369 keV γ-ray from $^{24}Mg(p,p'\gamma)^{24}Mg$ has the highest cross-section. As an example, this excitation function is given in Figure 9 for beam energies from 2.8 to 3.8 MeV.

A proper excitation curve is measured such that angular distribution effects are averaged out. For this purpose a large solid-angle detector is often positioned at an angle of 55° with respect to the beam direction.

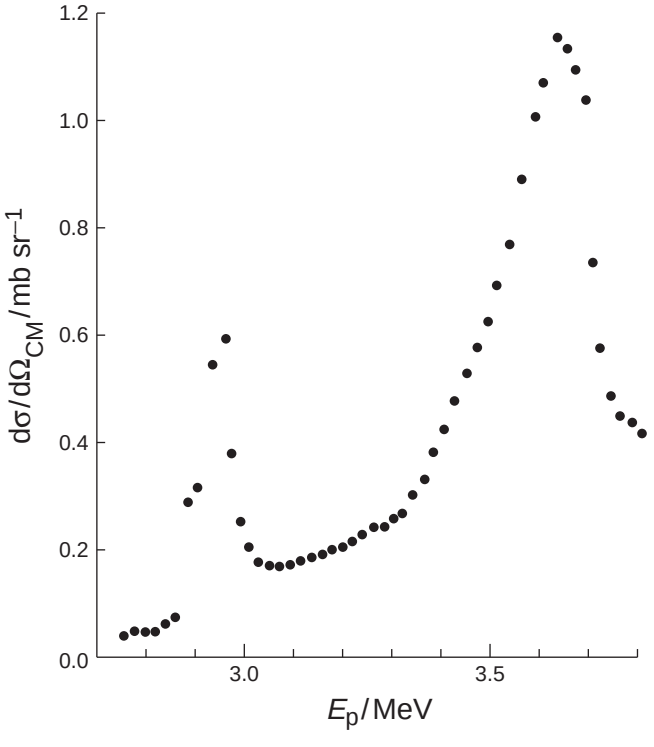

Figure 9 Excitation function for the 1369 keV γ-ray from $^{24}Mg(p,p'\gamma)^{24}Mg$, $\theta_{lab} = 90°$ (from Boni et al.[28]).

3.4 Background

The LOD of an element can be calculated with the help of Equation (30) and is determined by the background radiation under the peak. The background can be caused by minor or major elements in the sample under investigation but also by the experimental set-up and/or the detector.

The background caused by the minor and major elements is dependent on the type and energy of the projectiles. The background is mainly caused by Compton scattering in the detector, depending on the response function and on the γ-rays reaching the detector. These γ-rays can be generated by competing nuclear reactions in the sample and in the surrounding materials.

γ-Radiation can also be produced by reactions taking place in the diaphragms in front of the vacuum chamber, the (aluminum) beam pipe, the vacuum chamber itself, and the Faraday cup used to measure the beam current. In the design of the vacuum chamber care has to be taken to use materials that do not produce large numbers of γ-rays. Light ions, particularly those induced by deuterons, cause more severe background effects than heavy ions. This is mainly due to the lower Coulomb barrier and to the many reaction channels that are open. Precautions have to be taken so that materials do not contain contaminations that can produce γ-rays. Another problem could be natural radioactivity in the surroundings of the experimental set-up. Although the intensity generally is too low to be of any influence, in time-consuming experiments or in cases where a very low detection limit is required, this background can be of importance. To prevent any influence of the surroundings, lead shielding of the detector should be used. Aged lead exhibits less activity than new material.

3.5 Experimental Set-up and Procedures

In this section a typical experimental arrangement to perform PIGE experiments will be presented together with some experimental procedures.

3.5.1 Accelerator and Beam Guidance System

Most PIGE experiments on light elements are performed with energies between 1 and 4 MeV amu^{-1}. This energy is also ideally suited to performing simultaneous measurements with PIXE and/or RBS. In this case a single-ended Van de Graaff accelerator or pelletron accelerator is well suited to perform the PIGE experiments. If reactions are to be used where higher energies and/or heavy ions such as ^{15}N are necessary, tandem accelerators or cyclotrons should be used. In order for the accelerator to perform depth profiling, an easily changeable beam energy and

good energy resolution are required. This excludes a cyclotron for depth profiling, unless improvement of the beam quality is obtained, for example by means of a dispersive system.[24]

Usually, the accelerator is connected to a so-called analyzing magnet which selects a beam with a desired energy for a specific ion. The calibration of the analyzing magnet is usually done by means of a suitable resonant nuclear reaction. Nonlinearity and hysteresis effects of the magnet can make the calibration procedure and the setting of the magnetic field to the correct value critical if great accuracy in the beam energy is required.

Sometimes a proton beam is used[29] with an increased energy spread to make the excitation function more or less flat without any resonance structure. In this case it is not necessary to know the exact value of the beam energy from the accelerator.

The beam is transported to the irradiation chamber through evacuated pipes by means of bending magnets which bend the beam and electrostatic quadrupole lenses which focus the beam. The size of the beam on the target is usually in the range 0.5–5 mm in diameter but can be focused down to below 1 μm in a microprobe set-up.

3.5.2 *Irradiation Chamber*

Many possible layouts of the irradiation chamber are used to perform PIGE analyses. In Figure 10, a possible layout is presented where most of the commonly used features are included. The targets can be mounted on a one- or two-dimensional translation stage. This allows for multiple samples to be mounted and measured without breaking the vacuum by the application of micrometers or stepper motors. Moreover, the surface of the sample can be scanned if inhomogeneities in the sample are expected. One target position is mostly occupied with fluorescent quartz. It can be used to observe and align the beam spot on the target. For the majority of applications it is useful to be able to detect several types of radiation and particles from the sample. Besides a Ge(Li) detector to detect high energy γ-rays (PIGE), a Si(Li) detector to detect X-rays (PIXE) and a surface barrier detector to detect scattered particles (RBS) are usually mounted in the irradiation chamber.

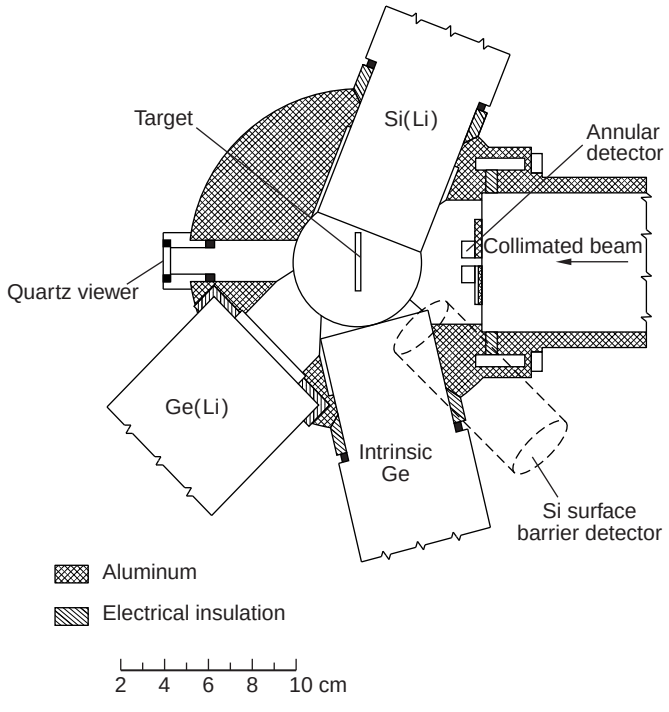

Figure 10 Example layout of a vacuum chamber to perform PIGE experiments (from Borderie[23] and references therein).

For quantitative measurements, the beam charge can be determined in several ways. With thin samples in which the incident particles are not stopped, a Faraday cup behind the sample can be used to determine the beam charge accurately. With thick samples a rotating vane in front of the sample can be used or the beam charge can be determined from a beam current measurement from the sample itself. In all cases, care has to be taken to prevent secondary electrons from the target reaching the Faraday cup by applying secondary electron suppressor rings near the target and/or in front of the Faraday cup.

3.5.3 *Detectors*

The γ-rays in a PIGE experiment can be measured either by a scintillator (NaI(Tl)) or by bismuth–germanium–oxide (BGO), a Ge or a Ge(Li) detector. A detailed description of these detectors can be found in a textbook by Knoll.[30] The NaI(Tl) detector has a relatively poor energy resolution (about 6% at 662 keV) combined with a high efficiency. It can be used in experiments where the energy resolution is not critical but efficiency is important, for example in depth profiling experiments where high efficiency is needed. BGO detectors have the advantage of a higher efficiency than NaI(Tl) for a given size crystal and a better peak-to-Compton ratio. The compact set-up makes BGO well suited to exploiting weak resonances. The superior energy resolution of a Ge detector (<2 keV at 1.33 MeV), a relatively high efficiency and a large peak-to-Compton ratio means that the Ge detector is mainly used in experiments to determine the bulk concentration of complex samples. The energy resolution is high enough to distinguish γ-rays from competing reactions and to analyze interfering peaks in the spectrum. For low-energy γ-rays, use of a planar hyperpure germanium detector can be very useful.[23] It has an energy resolution of about 500 eV at 122 keV.

3.5.4 Electronics

The detectors are connected to charge-sensitive preamplifiers and the resulting pulses are amplified in a spectroscopy amplifier. Shaping and filtering in these amplifiers are used to improve the signal-to-noise ratio. Next, the pulses are fed into a MCA. The analog-to-digital converter (ADC) part of the MCA usually has a resolution of 12 bit (4096 channels), although sometimes a higher resolution is used. The conversion time of the ADC has to be fast enough to prevent dead time problems, especially during high-count-rate experiments. Usually, a conversion time below 10 μs is applied.

3.5.5 Calibration and Spectrum Analysis

First, the γ-ray detection system has to be calibrated. This can be carried out quite easily by using a set radioactive source such as ^{60}Co, ^{208}Tl or ^{207}Bi, depending on the energy region of interest. As an example, in Figure 11, a calibration spectrum is presented. Several peak-fitting routines exist to find the significant peaks and to determine accurately their (channel) positions in the spectrum. To analyze a spectrum, first the peaks have to be identified and possible overlap problems solved. The high energy resolution of a Ge(Li) detector and the limited number of γ-rays means that the peaks in a PIGE spectrum are usually well separated without overlap problems. The actual calibration is usually performed by fitting a linear (or higher) polynomial function between the energy of the peaks in the spectrum and their corresponding positions. The second step is to determine the intensities of the peaks. This is generally performed by a computer code. This computer code is usually also equipped with an automatic peak search algorithm making analysis of PIGE spectra easy to perform. Finally, the intensities of the peaks combined with a sensitivity calibration give the concentration of the isotopes present in the sample.

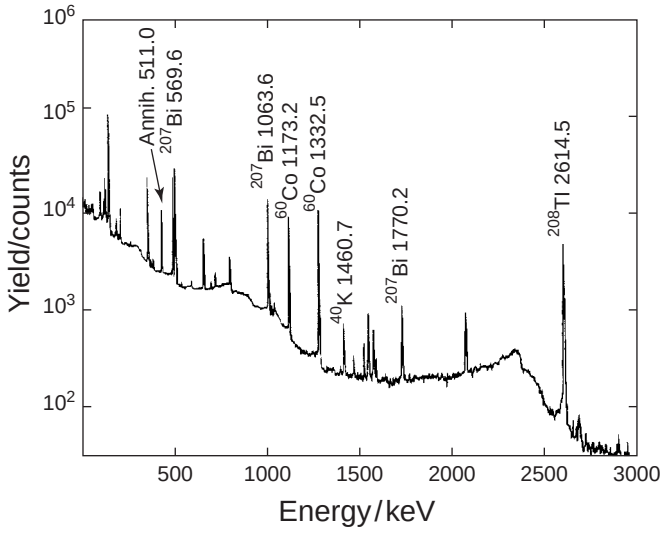

Figure 11 γ-Ray energy calibration spectrum of a Ge(Li) detector using several radioactive sources as indicated at the relevant peaks (from Hirvonen[31]).

The precise beam energy has to be determined, particularly for depth profiling. This can be done by a number of reactions with narrow and strong resonances, depending on the energy of interest; e.g. the reaction ^{27}Al(p,γ)^{28}Si with a resonance energy of 991.90 ± 0.04 keV and ^{15}N(p,$\alpha\gamma$)^{12}C with 429.57 ± 0.09 keV.

3.5.6 Target Preparation

Generally, the target preparation of PIGE samples is easy: just use them as they are. Some geological samples can be polished into sections, 20 μm thick, which allow investigation with an optical microscope.[23] Powdered samples can be pressed into pellets and analyzed directly. Biological materials can be freeze-dried and measured. When liquids have to be analyzed, more complicated procedures have to be followed or an external beam has to be used (see below). In microbeam analysis, care has to be taken not to lose the integrity of the sample during preparation. In the analysis of light elements in biological samples this can be especially difficult to accomplish.

3.5.7 External Beam

Samples such as liquids, wet biological samples or objects with large dimensions cannot be placed inside an evacuated irradiation chamber. These samples can be analyzed with an external beam set-up where the beam exits the beam pipe through a thin foil which is often cooled. The sample is often placed in a helium atmosphere to decrease target heating and to increase the particle range of the protons (and thus diminish straggling) since stopping for hydrogen in helium is low.[23]

3.5.8 Microbeam

A focused microbeam can be used to determine lateral distributions of elemental concentrations on a scale of about 1 μm. A microbeam set-up requires an object and aperture diaphragm combined with a number (2–4) of magnetic quadrupoles to focus the beam down to micrometer sizes.[24] The focusing principles, problems, and applications can be found in a textbook by Watt and Grime.[32]

3.6 Depth Profiling

Several isotopes have the potential for depth profiling, for instance ^{1}H, ^{13}C, ^{15}N, ^{18}O, ^{19}F, ^{22}Ne, ^{23}Na, ^{24}Mg,

^{26}Mg, ^{27}Al, ^{29}Si, ^{30}Si. Most of these light nuclei have sharp and strong resonances in the excitation curve. The instrumentation for performing depth profiling is discussed in section 3.5. Additional requirements for the accelerator are a good energy resolution and the capability to change the energy easily. High-energy resolution of the beam is required in applications where narrow resonances are used. A detailed description of depth profiling can be found in Hirvonen.[31]

4 APPLICATION OF THE PARTICLE-INDUCED γ-RAY EMISSION TECHNIQUE

Possible applications range from geological and archeological samples, steel samples, dust and aerosol samples to biomedical samples. In this section only a few examples are presented of PIGE analyses to serve as illustrations.

Van IJzendoorn et al.[33] used the PIGE technique to quantify thin layers of SiF_x that were a result of reactive ion etching of Si wafers with a CF_4 plasma. The quantification is important to understand the etching process. The ^{19}F(p,p′γ)^{19}F reaction was used to determine F on the Si wafer. A proton energy of 2.78 MeV was chosen to suppress a Si reaction and thus limit the Compton background. An example of a γ-ray spectrum is given in Figure 12.

In Figure 13 an example of the analysis of a hafnium plate is presented.[18] The hafnium plate is bombarded with 10 mC of 2.4 MeV protons. In the spectrum the following elements can be seen: O (150 ppm), Na (0.3 ppm), Al (30 ppm) and P (5 ppm). Also heavier elements including Zr (2.8%), Fe (100 ppm) and Cu (<50 ppm) can be observed.

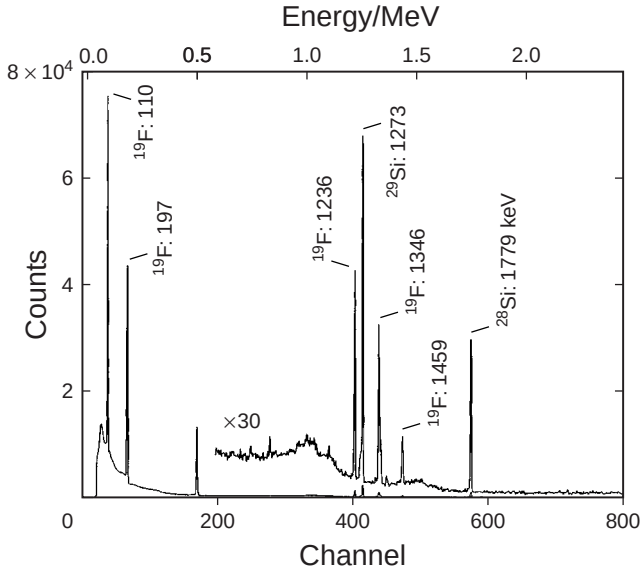

Figure 12 γ-Ray spectrum of a 120-nm thick CF_2 layer on Si measured with 2.8-MeV protons (from van IJzendoorn et al.[33]).

An example of depth profiling with the ^{15}N(p,αγ)^{12}C reaction is given in Figure 14. Here, the diffusion of 40 keV ^{15}N in evaporation-deposited Ni was measured.[34] The nitrogen segregates to the surface and to the Ni–Ta interface at increased temperatures. The nitrogen distribution is very narrow and the broadening is due to experimental resolution at this depth. This is clearly seen at the measured depth profiles.

5 CONCLUSIONS

We have shown that PIGE offers a simple, sensitive and quantitative multielemental analysis technique. With

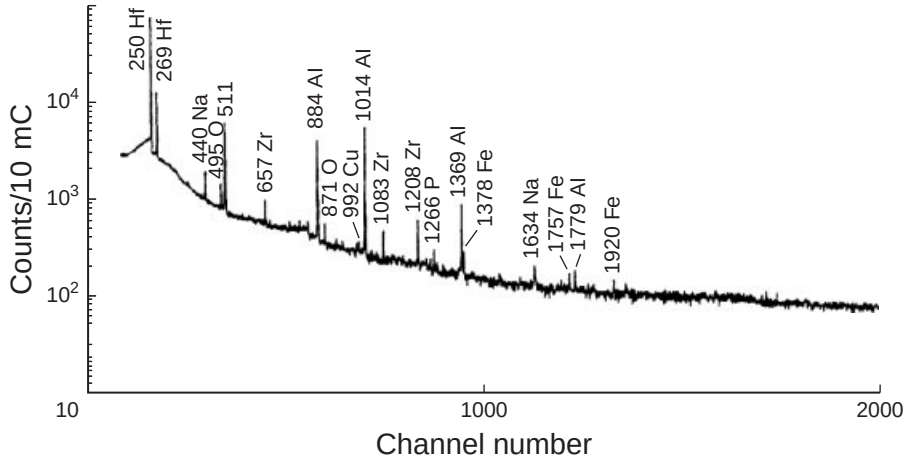

Figure 13 γ-Ray spectrum of a hafnium plate after bombardment with 2.4-MeV protons. A lead absorber in front of the detector was used to decrease the low-energy γ-rays (from Räisänen and Hänninen[18]). The annihilation peak is 511 keV.

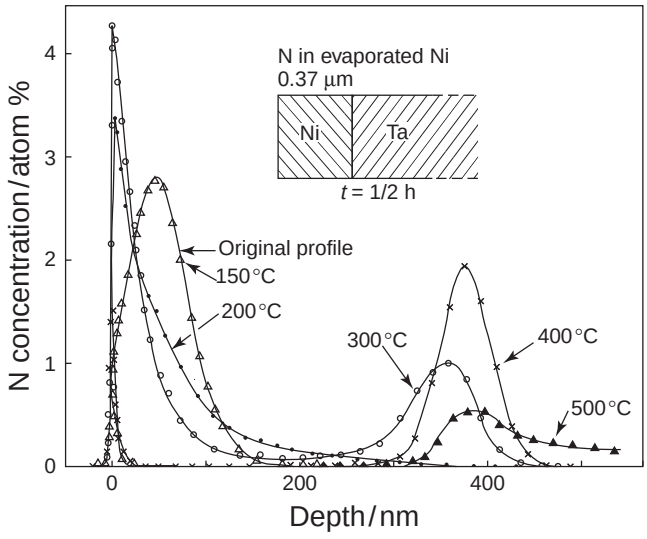

Figure 14 ^{15}N profiles in evaporated Ni on Ta after annealing. The original profile is that of the 40 keV ^{15}N implants (from Hirvonen[31] and Lappalainen and Anttila[34]). The top layer of Ni is 370-nm thick. The measured ^{15}N profiles show increased yields at the surface and at the Ni–Ta interface at increased temperatures (from Lappalainen and Anttila[34]).

tunable accelerators and a good beam energy resolution, depth profiles can be obtained as deep as several micrometers, depending on the material and beam energy. The technique can be combined easily with PIXE and RBS for simultaneous additional information. The strength of PIGE is found particularly for very light elements where other techniques fail or experience difficulties. Sensitivities can reach the parts per million range, depending on the element of interest and on the matrix composition.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
BGO	Bismuth–Germanium–Oxide
CN	Compound Nucleus
fwhm	Full Width at Half-maximum
LOD	Limit of Detection
MCA	Multichannel Analyzer
PIGE	Particle-induced γ-Ray Emission
PIXE	Particle-induced X-Ray Emission
PRA	Prompt Radiation Analysis
RBS	Rutherford Backscattering Spectroscopy

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • PIXE (Particle-induced X-ray Emission)

REFERENCES

1. G. Deconninck, *Introduction to Radioanalytical Physics*, Akademia Kiado, Budapest, 1978.
2. R.F. Sippel, E.D. Glover, 'Sedimentary Rock Analysis by Charged Particle Bombardment', *Nucl. Instrum. Methods*, **9**, 37–48 (1960).
3. T.B. Pierce, P.F. Peck, W.M. Henry, 'The Rapid Determination of Carbon in Steels by Measurement of the Prompt Radiation Emitted During Deuteron Bombardment', *Analyst*, **90**, 339–345 (1965).
4. T.B. Pierce, P.F. Peck, D.R.A. Cuff, 'Application of Inelastic Proton Scattering to the Rapid Determination of Silicon in Steels', *Anal. Chim. Acta*, **39**, 433–436 (1967).
5. G. Deconninck, 'Analysis by Prompt X- and γ-ray Detection', *J. Radioanal. Chem.*, **12**, 157–169 (1972).
6. G. Deconninck, G. Demortier, Laboratoire Interfacultaire d'Analyses par Reactions Nucleaires Reports 725 and 752, LARN, Namur, Belgium, 1972 and 1975.
7. I.S. Giles, M. Peisach, 'A Survey of the Analytical Significance of Prompt γ-rays Induced by 5 MeV Alpha-particles', *J. Radioanal. Chem.*, **50**, 307–360 (1979).
8. R.B. Boulton, G.T. Ewan, 'Simultaneous Analysis of Light Elements Using Prompt Nuclear Reaction γ-rays', *Anal. Chem.*, **49**, 1297–1304 (1977).
9. B. Borderie, J.N. Barrandon, 'New Analytical Developments in Prompt γ-ray Spectrometry with Low-energy Tritons and Alpha Particles', *Nucl. Instrum. Methods*, **156**, 483–492 (1978).
10. P.M. Endt, 'Energy Levels of $A = 21–44$ Nuclei (VII)', *Nucl. Phys. A*, **521**, 1–830 (1990).
11. F. Ajzenberg-Selove, 'Energy Levels of Light Nuclei $A = 18–20$', *Nucl. Phys. A*, **392**, 1–216 (1983).
12. 'Nuclear Data and Nuclear Data Sheets', Academic Press, New York and London.
13. J.F. Ziegler, B.F. Biersack, U. Littmark, *The Stopping and Ranges of Ions in Solids*, Pergamon Press, New York, 1985.
14. K.S. Krane, *Introductory Nuclear Physics*, J. Wiley & Sons, New York, 1998.
15. H. Ejiri, M.J.A. de Voigt, *γ-Ray and Electron Spectroscopy in Nuclear Physics*, Clarendon Press, Oxford, 1989.
16. A. Anttila, R. Hänninen, J. Räisänen, 'Proton-induced Thick-target γ-Ray Yields for the Elemental Analysis of the $Z = 3–9$, 11–21 Elements', *J. Radioanal. Chem.*, **62**, 293–306 (1981).
17. A.Z. Kiss, E. Koltay, B. Nyako, E. Somorjai, A. Anttila, J. Räisänen, 'Measurements of Relative Thick Target

Yields for PIGE Analysis on Light Elements in the Proton Energy Interval 2.4–4.2 MeV', *J. Radioanal. Chem.*, **89**, 123–141 (1985).
18. J. Räisänen, R. Hänninen, 'Heavy Element ($Z > 30$) Thick-target γ-Ray Yields Induced by 1.7 and 2.4 MeV Protons', *Nucl. Instrum. Methods*, **205**, 259–268 (1983).
19. International Union of Pure and Applied Chemistry, 'Nomenclature, Symbols, Units and their Usage in Spectrochemical Analysis', *Spectrochim. Acta*, **33**, 241–245 (1978).
20. P.J. Clark, G.F. Neal, R.O. Allen, 'Quantitative Multielement Analysis Using High Energy Particle Bombardment', *Anal. Chem.*, **47**, 650–658 (1975).
21. B. Borderie, J.L. Pinault, J.N. Barrandon, *Analusis*, **5**, 282 (1977).
22. B. Borderie, J.N. Barrandon, B. Delaunay, M. Basutçu, 'Use of Coulomb Excitation by Heavy Ions (^{35}Cl, 55 MeV) for Analytical Purposes: Possibilities and Quantitative Analysis', *Nucl. Instrum. Methods*, **163**, 441–451 (1979).
23. B. Borderie, 'Present Possibilities for Bulk Analysis in Prompt γ-Ray Spectrometry with Charged Particles', *Nucl. Instrum. Methods*, **175**, 465–482 (1980).
24. P.H.A. Mutsaers, 'Nuclear Microprobe Design', *Nucl. Instrum. Methods, B*, **113**, 323–329 (1996).
25. M.J. Kenny, J.R. Bird, E. Clayton, 'Proton Induced γ-Ray Yields', *Nucl. Instrum. Methods*, **168**, 115–120 (1980).
26. R. Lappalainen, A. Anttila, J. Räisänen, 'Absolute α-induced Thick-target γ-Ray Yields for the Elemental Analysis of Light Elements', *Nucl. Instrum. Methods*, **212**, 441–444 (1983).
27. W.A. Ranken, T.W. Bonner, J.H. McCrary, 'Energy Dependence of $F^{19} + p$ Reactions', *Phys. Rev.*, **109**, 1646–1651 (1958).
28. C. Boni, E. Cereda, G.M. Braga Marcazzan, V. de Tomasi, 'Prompt γ Emission Excitation Functions for PIGE Analysis of Li, B, F, Mg, Al, Si and P in Thin Samples', *Nucl. Instrum. Methods B*, **35**, 80–86 (1988).
29. C. Boni, A. Caridi, E. Cereda, G.M. Braga Marcazzan, 'PIXE–PIGE Analysis of Thin Fly-ash Samples', *Nucl. Instrum. Methods, B*, **45**, 352–355 (1990).
30. G.F. Knoll, *Radiation Detection and Measurement*, J. Wiley & Sons, New York, 1979.
31. J.-P. Hirvonen, 'Nuclear Reaction Analysis: Particle-γ Reaction', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, PA, 1995.
32. F. Watt, G.W. Grime, *Principles and Applications of High-energy Ion Microbeams*, Hilger, Bristol, 1987.
33. L.J. van IJzendoorn, M. Haverlag, A. Verbeek, J. Politiek, M.J.A. de Voigt, 'Quantification of Fluorine in SiF_x Etch Residues on Silicon with Proton Induced γ-Emission', *Nucl. Instrum. Methods, B*, **113**, 411–414 (1996).
34. R. Lappalainen, A. Anttila, 'Annealing Behaviour of Implanted Nitrogen in Bulk and Evaporated Ni', *Appl. Phys. A*, **42**, 263–267 (1987).

Photon Activation Analysis

Kazuyoshi Masumoto
High Energy Accelerator Research Organization, Tokyo, Japan

Christian Segebade
Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany

1 Introduction **2**
1.1 Analytical Use of Electron Accelerators 2
1.2 Principle of Photon Activation Analysis 2
1.3 Short History of Photon Activation Analysis 3
1.4 Textbook and Data Collections for Photon Activation Analysis 4

2 Photonuclear Reaction **4**
2.1 Characteristics of Photonuclear Reactions 4
2.2 Bremsstrahlung for Photon Activation 4
2.3 Photonuclear Reaction Yield 5

3 Analytical Application of Photon Activation **7**
3.1 Conditions of Photon Activation Analysis 7
3.2 Characteristics of Photon Activation Analysis 8

4 Analytical Procedures **8**
4.1 Procedure for Light-element Analysis 8
4.2 Procedure for Multielement Analysis 9
4.3 Sample Preparation, Standard and Flux Monitoring 9
4.4 Irradiation 10
4.5 Measurement and Evaluation 13

5 Interference Management **14**
5.1 Interference Reactions 14
5.2 Peak Overlapping 14
5.3 Other Errors 15

6 Quantitative Methods for Quality Assurance **16**
6.1 Internal Standard Method 16
6.2 Internal Standard Method Coupled with the Standard Addition Method 17
6.3 The Stable Isotope Dilution Method 18

7 Application **19**
7.1 Analysis of Light Elements 19
7.2 Multielement Analysis of Biological Materials 19
7.3 Analysis of Geological and Environmental Materials 19
7.4 Multielement Analysis of Raw Materials and Industrial Products 20

8 Comparison with Other Activation Analysis Methods **20**
8.1 Irradiation 20
8.2 Comparison of Nuclear Reactions Induced 21
8.3 Sensitivity and Selectivity of Elements to be Determined and Applicability to Various Kinds of Materials 21

9 Conclusion **22**

Abbreviations and Acronyms **22**

Related Articles **22**

References **22**

From the large number of analytical methods, activation analysis techniques are the only ones which are based upon nuclear reaction. The material sample studied is exposed to high-energy radiation which can be partly absorbed by a nucleus in the sample. Thus the nucleus is excited to a high energy level which can decay through quasi-prompt emission of a nuclear particle or photon. The product nuclide produced is mostly radioactive, and so emits delayed radiation. Both this and the aforementioned prompt radiation can be measured using appropriate radiation detectors. By evaluating the energy and the count rate of the particles detected, qualitative and quantitative analyses of the target material under study can be performed. Thus it is clear that elements only, not chemical species, can be determined directly. A large variety of particles can be used for activation, namely uncharged ones (neutrons, photons) or charged particles like protons, deuterons, tritons and even heavier ones. Mostly thermal neutrons from nuclear research reactors are used since this technique offers the highest average analytical sensitivity. During photon activation, the target nucleus is activated by photonuclear reaction. This is induced to ''normal'' material at high energies, usually not below about 10 MeV. The photonuclear reaction data of the elements suggest an activation energy around 30 MeV with respect to analytical sensitivity and interfering reactions, respectively. This energy is best achievable with bremsstrahlung sources like high-power linear accelerators or microtrons.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Favorable irradiation parameters are: 30 MeV electron energy at 100–150 μA mean electron beam current. With the help of suitable radiation spectrometers, e.g. high-resolution germanium detectors connected to appropriate pulse processing electronics, photon (γ or characteristic X-ray) spectra can be taken by which simultaneous multicomponent analyses can be carried out without chemical separations, sometimes even nondestructively. Moreover, partly extreme sensitivities can be achieved, and some elements can be analyzed whose determinations are difficult or impossible using other techniques, e.g. light elements like carbon, nitrogen, oxygen and fluorine. A further advantage is the relative freedom from blanks in many cases; after bremsstrahlung exposure, undesirable surface contaminants can be removed from the sample, and the recontamination that eventually occurs is inactive, and thus can be disregarded. Since the activation and measuring process is independent of the chemical status of the component studied, a large variety of matrices can be analysed. Photon activation has been applied in several areas including:

- *geo- and cosmochemistry;*
- *environmental, biological and medical science;*
- *industrial product and high-purity material analysis;*
- *archaeological and forensic science;*
- *certification of reference materials.*

The disadvantages of the method are common to all activation analysis techniques, e.g. the instrumental equipment costs. The cost of a high-performance germanium spectrometer is about US $30 000, and this does not include the permanent costs of maintenance and liquid nitrogen supply. Also, additional personnel qualifications are required for radioactive laboratory work. Finally, the handling of radioactive waste unavoidably produced during activation analysis might be problematic in some cases.

1 INTRODUCTION

1.1 Analytical Use of Electron Accelerators

Electron accelerators have been used not only for nuclear and particle physics but also in many other areas, such as in material, biological, geological and archeological sciences. In the analytical use of electron accelerators, photons such as synchrotron radiation and bremsstrahlung induced by high-energy electrons have been used more frequently than an electron beam itself.

Synchrotron radiation is generated by high-energy electrons circulating in the synchrotron. Synchrotron radiation energy is distributed continuously covering the infrared (IR) to X-ray region and interacts with molecules or atoms. In particular, the X-ray region of synchrotron radiation has been applied extensively to the field of analytical science using X-ray photoelectron spectroscopy (XPS), extended X-ray absorption fine structure (EXAFS), and X-ray fluorescence (XRF); these methods are useful to determine elements and their chemical states. Recently, many facilities using electron storage rings have been constructed for the exclusive use of synchrotron radiation.

However, bremsstrahlung is radiated in a forward direction of the electron beam when the electron is decelerated by Coulomb interactions in a material. Bremsstrahlung has a continuous energy distribution with a maximum energy equal to the electron acceleration energy. As the photon energy is significantly higher than that of synchrotron radiation the bremsstrahlung photon interacts with the nucleus. The cross-sections of photonuclear reactions are about a hundred times larger than that of nuclear reaction induced by electrons of the same energy. Therefore, bremsstrahlung from high-energy electrons can well be used for radioisotope production and particularly for activation analysis. Usually, photon activation analysis (PAA) is carried out using an electron linear accelerator or a microtron which can supply mean electron beam currents of some tens to hundreds of microamperes at energies between 6 to say 60 MeV.

1.2 Principle of Photon Activation Analysis

In PAA, unknown samples are irradiated with high-energy photons. Then, radionuclides produced by photonuclear reactions in the sample (the (γ,n)-type reaction occurs most probably; see below, section 2) are identified by their radiation emission energy, and the intensity of radioactivity is measured. Each radionuclide represents an element originally present in the sample studied, and the intensity of radioactivity is proportional to the abundance of this element.

In general, unknown samples and comparative reference standards are irradiated simultaneously or sequentially, equipped with flux monitors. After correction of the activating photon flux received by the sample and the standard, respectively, the contents of the components of the sample can be calculated by the specific activity ratios of the sample and the comparative standard. Usually, a γ-ray spectrum of sample is measured with a high-resolution spectrometer as described below (see section 1.2.2). In the spectrum, the elements present in the sample are represented by respective peaks whose areas are proportional to the radioactivity of the element to be determined. The peak area ratio of the sample and the standard is commonly used for the calculation of the elemental content. In the following, the different radiation measurement devices commonly used in PAA are described briefly.

1.2.1 Scintillation Spectrometer

A scintillation detector consists of a large, cylindrical thallium (Tl)-doped NaI crystal which is shielded by aluminum housing (3 in. both in height and diameter being the standard size). Incident photon energy emitted from a radioactive sample is converted into visible light by the crystal. A photocathode plus photomultiplier (PM) tube is connected to the crystal. These convert the light flash into an electric pulse signal whose height is proportional to the absorbed photon-ray energy. These signals are processed further (amplified, reshaped and more) by a sequence of amplifiers (preamplifier, spectroscopy amplifier). The signals thus produced have values of several volts, depending upon the incident energy, are Gauss-shaped and short enough (4 μs in the normal approach) to enable undisturbed processing of high pulse frequencies, generated by higher activities of the measured material. The signals are processed further in a multichannel analyzer which produces a spectrum whose lines represent the respective energies emitted by the sample measured.

1.2.2 Coaxial Germanium Spectrometer

This device is used for high-resolution measurement of γ-rays. The basic set-up is similar to that of the NaI spectrometer described above. A high-purity germanium single crystal is applied instead the scintillation crystal plus PM tube. The other pulse-processing electronics are the same as described above. The resolution power exceeds that of an NaI spectrometer by about two orders of magnitude. Therefore, it is indispensable for the measurement of complex multicomponent spectra as mostly occur during instrumental multielement activation analysis.

1.2.3 Low-energy Photon Spectrometer

A low-energy photon detector (LEPD) is commonly used for detecting characteristic X-rays and γ-rays below 100 kV. The LEPD is made of a thin crystal of germanium (typically 10 mm thickness, 30 mm diameter). All the other devices in the spectrometer are the same as used in the coaxial spectrometer described above. The detection efficiency for X-rays is high, because X-rays are fully absorbed even by a thin crystal. Most of the photoactivation product isotopes of medium to heavy elements decay by electron capture (EC), because the (γ,n) reaction products are mostly neutron-deficient nuclides. EC decay is always accompanied by characteristic X-ray emissions whose energies are those of the respective decay products. The γ-ray spectra of medium to heavy elements are often more complex than the X-ray spectra, as shown by Segebade and Weise.[1] Therefore, low-energy photon spectrometry offers a complementary method to γ-ray spectrometry. The LEPD detection sensitivities for 36 elements were compared with that of a usual Ge detector.[2] For quantitative analysis, self-attenuation of X-rays in a sample should be always estimated by the factor

$$F(E) = \frac{1 - \exp[-\mu(E)d]}{\mu(E)d}$$

where d is the sample thickness and $\mu(E)$ is the total attenuation coefficient of the sample material at a given energy. Sato et al.[3] carefully checked the self-absorption effect for the analysis of biological materials and the validity of the method was verified by analyzing several biological materials.

1.3 Short History of Photon Activation Analysis

The pioneering work on photonuclear reactions was done by Chadwick and Goldhaber[4] in 1934. In this work, the photodisintegration of deuterons was induced by 2.62 MeV γ-rays from ^{208}Tl as the excitation source. The radioisotope excitation method can only be applied to limited examples such as several isomers produced by (γ,γ′)-reactions and the photodisintegration of deuterium and beryllium.

Since 1950, electron accelerators have been constructed in many research institutions. These primarily served nuclear physics purposes, but an intense systematic study of PAA also began. In 1954, Basile et al.[5] used a betatron for the determination of oxygen in organic acids. In the beginning, PAA was recognized as an analytical tool for light elements such as C, N and O, which cannot be determined by neutron activation analysis (NAA) using nuclear reactors.

First, Geiger-Müller counters were used for product nuclide radiation counting. In the 1940s, the energy resolution power of scintillation crystal detectors was recognized. Mostly Tl-doped NaI crystals were used for γ-counting, and they are still used for special applications (e.g. light element analysis).

Since the beginning of the 1960s, high-resolution γ-ray spectrometry developed dramatically with the introduction of the Ge-based semiconductor detector, the multichannel analyzer and, somewhat later, the computer. In activation analysis, multielement analysis frequently can be performed easily without chemical separation. PAA has been extensively applied to a large variety of materials as a complementary tool for NAA, and it has been recognized that PAA is suitable for multielement analysis of geological, biological and environmental materials. The analysis of light elements in highly purified materials has been also very important in the field of material science and technology and will be of urgent interest also in the future.

Electron accelerators of comparatively low energy (a few tens of megaelectron-volts) and high power (a few kilowatts) have been used for PAA. However, a limited number of electron accelerators suitable for PAA are available for analysts since modern accelerators are designed for the production of much higher energies to primarily meet the demands of nuclear and particle physicists.

1.4 Textbook and Data Collections for Photon Activation Analysis

In 1972, Engelmann[6] introduced PAA in his book *Advances in Activation Analysis*, in which the principle of PAA and the application to light element analysis were explained in detail. *Photon Activation Analysis* by Segebade et al.[7] is a comprehensive textbook of PAA, containing 1206 references and many useful tables, such as a list of usable and competing photonuclear reactions and a list of γ-rays and X-rays. Kushelevsky[8] presented a short guide of PAA in *Activation Analysis*, where the principles and applications of the method are concisely summarized.

Toms[9] compiled the photonuclear reaction products and their γ-ray energies whilst Williams et al.[10,11] collected γ-ray spectra of 40–44 MeV photon activation products. Lutz and Segebade[12] published a two-dimensional (half-life versus γ-ray energy) map of photonuclear reaction products and Kato[13] obtained sensitivities of 71 elements in PAA using 30 MeV bremsstrahlung. These basic data have been very helpful for users of PAA.

2 PHOTONUCLEAR REACTION

2.1 Characteristics of Photonuclear Reactions

A photonuclear reaction is the interaction between a nucleus and photon. The cross-section of a photonuclear reaction is mainly dependent on the probability of photon absorption. A schematic representation of the cross-section of the photon absorption is shown in Figure 1.[7] The inelastic reaction which is expressed by (γ,γ') can be induced by the irradiation of photon energies even below a few mega electron volts and target nuclides are excited from the ground state to some isomer levels. The cross-section integrals of (γ,γ') reactions are very small, and only a few radioisotopes which are useful for PAA are produced. A photonuclear reaction is usually characterized by the ''giant dipole resonance'' or ''giant resonance'', because this type of reaction has the largest cross-section among several types of photonuclear reactions. When the wavelength of the photon becomes similar to the diameter of nucleus, the photon can be absorbed by the target nucleus through the electric dipole resonance mechanism with a high degree of probability. This phenomenon is interpreted by the fact that a collective oscillation of all protons against all neutrons is induced by the electromagnetic wave. The maximum cross-section region is located at about 14 MeV for heavy elements to about 25 MeV for light elements. After photon absorption, one or two neutrons and/or protons or higher-order particles can be released because the binding energy of one nucleon is about 10 MeV. At about 60 MeV where the wavelength approximately equals the diameter of the deuteron, a neutron–proton pair in a nucleus is likely to absorb the photon. As the result of this reaction, called a ''quasi-deuteron reaction'', a proton and neutron pair is emitted. In this energy region, two or more nucleons can also be released from the target nucleus by a direct process and many kinds of reaction products are formed. The cross-sections of the direct process are much smaller than that of the (γ,n) reaction of the giant resonance region. Above 150 MeV, the total photoabsorption cross-section increases again because of the photo-meson production. As the spallation and fragmentation reactions are also induced, various reaction products are produced. However, the latter two types of photoreaction are not suitable for analytical purposes, but rather have to be considered as sources of interference.

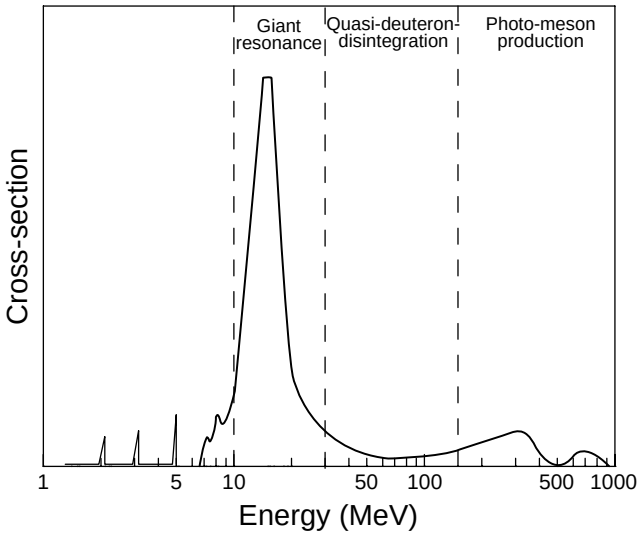

Figure 1 Schematic representation of the total cross-section for the absorption of photons by nuclei. (Reproduced by permission from Walter de Gruyter GmbH & Co.)

2.2 Bremsstrahlung for Photon Activation

Several radioisotopes (e.g. ^{124}Sb) were used as γ-ray excitation sources and applied to the analysis of deuterium, beryllium and several isomeric states. Photon fluxes from

such radioisotopes are usually not sufficient for practical elemental analysis purposes.

Bremsstrahlung is used preferably for activation in PAA. Bethe and Heitler[14] theoretically discussed the radiation loss of electrons. Interaction between electrons and matter can be explained as follows. For low atomic number (Z) target material and low-energy electrons, the major partition of energy loss is due to ionization. For high-Z target and high-energy electron, radiation energy loss is increased. As the radiation loss is proportional to Z^2, elements with higher atomic number more efficiently produce bremsstrahlung. Radiation length is defined as the thickness in which the energy of the electron is reduced to $1/e$ of its incident energy. The radiation lengths of air, aluminum and lead are 304 m (36.7 g cm^{-2}), 8.9 cm (24.0 g cm^{-2}) and 5.6 mm (6.37 g cm^{-2}) respectively. The cross-sections of ionization and radiation loss become equal at a critical energy which is about $800/Z$ MeV, where Z is the atomic number of the target. Therefore radiation loss becomes predominant for electrons above 10 MeV, when a heavy element, e.g. platinum, is used as target material.

In the case of a thin target, bremsstrahlung is radiated strongly to the forward direction, i.e. the direction of the incident electron beam. A thick target is used in practice for activation analysis. In this case, the angular distribution of bremsstrahlung is broadened, and the percentage of low-energy photons increases because of the multiple scattering of the electron in a thick target. Figure 2 shows the calculated photon flux for electron energies of 25, 30 and 35 MeV and a mean electron beam current of 100 μA.[15] The spectrum of bremsstrahlung is calculated by Schiff's equation.[16] The tungsten target is 6 mm (two radiation lengths) thick, which is the range of 35 MeV electrons (thus, all incident electrons are absorbed). The spectrum is integrated over 5° from the direction of the electron beam.

In PAA, the most suitable irradiation energy is 20–30 MeV because the radiated bremsstrahlung spectrum covers the giant resonance region of most of the nuclides. Under such irradiation conditions, reaction products induced by the (γ,n) and (γ,p) reactions can be conveniently used for quantitative element determination.

2.3 Photonuclear Reaction Yield

In PAA, the reaction yield is used for the estimation of sensitivity of the pertinent element and the magnitude of nuclear interference. When a photonuclear reaction is induced by the irradiation of bremsstrahlung of a certain maximum energy the photonuclear reaction yield Y can be derived by the integral in Equation (1) by using the bremsstrahlung spectrum and the excitation function for the relevant reaction,

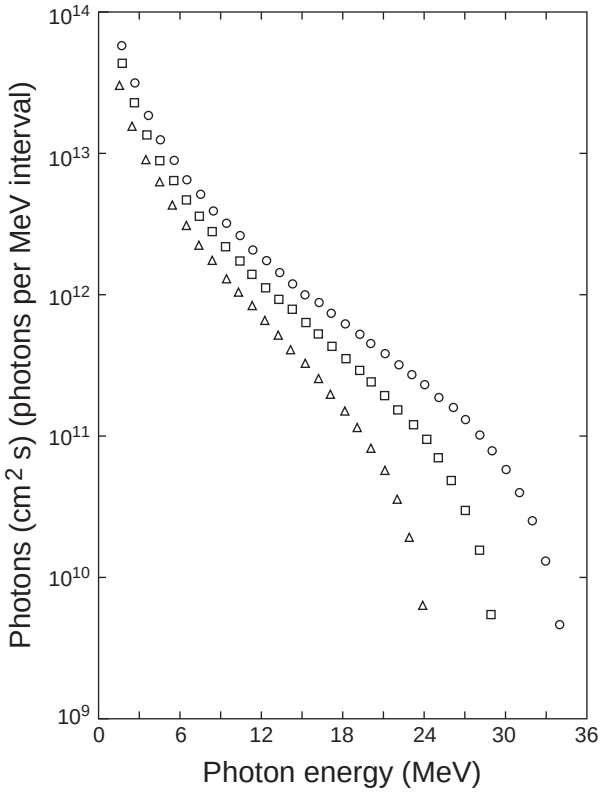

Figure 2 Calculated flux for (○) 35-, ( ) 30-, and (△) 25-MeV electrons for 100 μA average beam current and a subtended angle of 5°. (Reprinted with permission from G.J. Lutz, 'Calculation Sensitivities in Photon Activation Analysis', *Anal. Chem.*, **41**, 426. Copyright (1969) American Chemical Society.)

$$Y = N_0 \int_{E_{th}}^{E_{max}} \sigma(k)\Phi(k, E_{max})k^{-1}\,dk \qquad (1)$$

where N_0 is Avogadro's number, E_{th} is the threshold energy for the reaction, E_{max} is the bremsstrahlung maximum energy, $\sigma(k)$ is the cross-section in square centimeters per nucleus, and $\Phi(k, E_{max})/k$ is the number of photons at a given energy k. If the bremsstrahlung spectrum is normalized to 1 roentgen and a target amount of 1 mole is used, a yield can be obtained per mole per roentgen.

When equivalent quanta, Q, are introduced, the yield per equivalent quanta for the given reaction, σ_Q, is given by Equation (2),

$$\sigma_Q = Q^{-1} \int_{E_{th}}^{E_{max}} \sigma(k)\Phi(k, E_{max})k^{-1}\,dk \qquad (2)$$

where Q is given by

$$Q = E_{max}^{-1} \int_{0}^{E_{max}} \Phi(k, E_{max})\,dk \qquad (3)$$

According to the definition, Q is a hypothetical number of photons per second per square centimeter assuming maximum energy of all photons.

The general feature of a photonuclear reaction yield at the bremsstrahlung maximum energy of 30 MeV is shown in Figure 3 where the yields of various types of photonuclear reactions are plotted against the atomic number.[13] The yields (mol^{-1} R^{-1}) of (γ,n), (γ,2n) and (γ,3n) reactions increase with mass number. The (γ,2n) reaction yield is about one order of magnitude smaller than that of (γ,n) reaction for heavy elements. The (γ,p) reaction shows maximum yield at atomic number 20 and decreases with atomic number. In the case of higher-order charged particle emission reaction, yield curves show a similar shape as that of (γ,p) reaction.

The reaction yields of the (γ,n),[17] (γ,p)[18] and (γ,α)[19] reactions were measured with 20 MeV bremsstrahlung irradiation. The difference of the yields between 20 and 30 MeV are small. The reaction yields were also measured systematically at 60 MeV irradiation.[20] The yields of the single particle emission reactions, such as (γ,n), (γ,p) and (γ,α), are almost equal to the results of 30 MeV irradiation. The yields of (γ,2n), (γ,3n), (γ,4n), (γ,pn) and (γ,αn) reactions are remarkably higher. The (γ,2p) reactions are observed for several light elements. The emission of more particles, up to nine, was observed at 60 MeV.

To observe the dependency of the photonuclear reaction yield upon energy and mass number, the functions of the yield (σ_Q; mb) for the (γ,n), (γ,p) and (γ,pn) reactions are shown in the form of two-dimensional graphs in Figures 4, 5 and 6. In these figures, x- and y-axes show the target mass number and bremsstrahlung

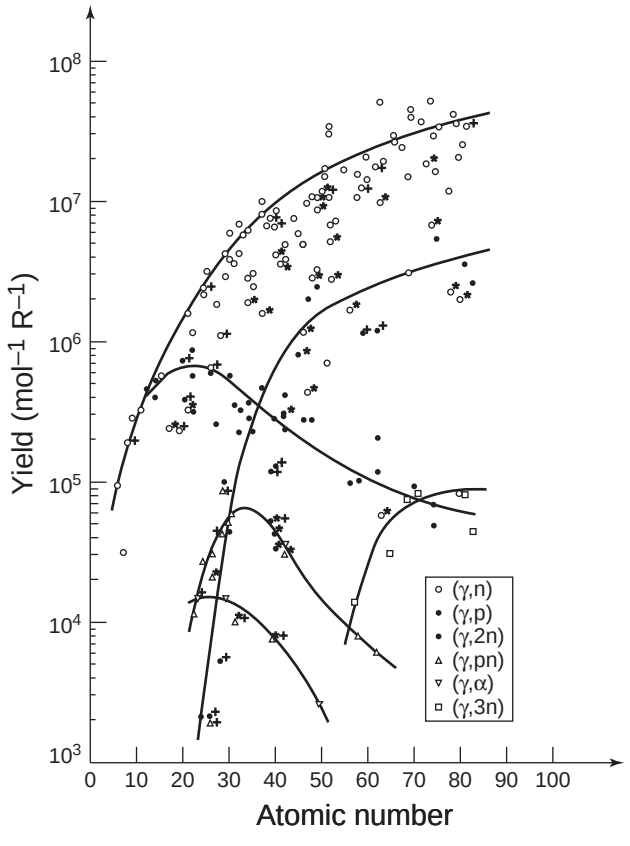

Figure 3 Yields of photonuclear reactions as a function of atomic number with 30 MeV bremsstrahlung. (Reproduced by permission of Akademica Kiado, Budapest.)

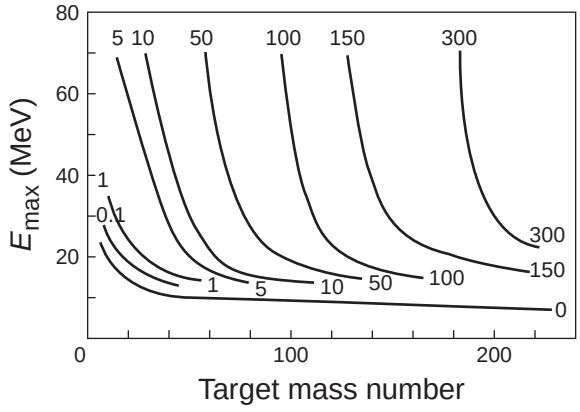

Figure 4 Yields of the (γ,n) reactions as a function of bremsstrahlung maximum energy and target mass number. The numerical values in the figure are yields per equivalent quanta (in mb). (Reproduced from K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 573 (1978) by permission of North-Holland Publishing Co.)

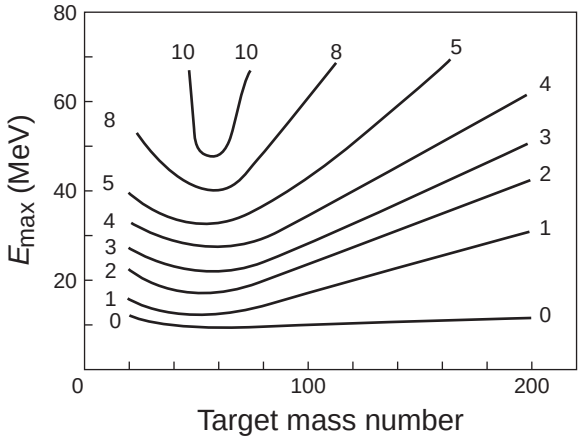

Figure 5 Yields of the (γ,p) reactions as a function of bremsstrahlung maximum energy and target mass number. The numerical values in the figure are yields per equivalent quanta (in mb). (Reproduced from K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 573 (1978) by permission of North-Holland Publishing Co.)

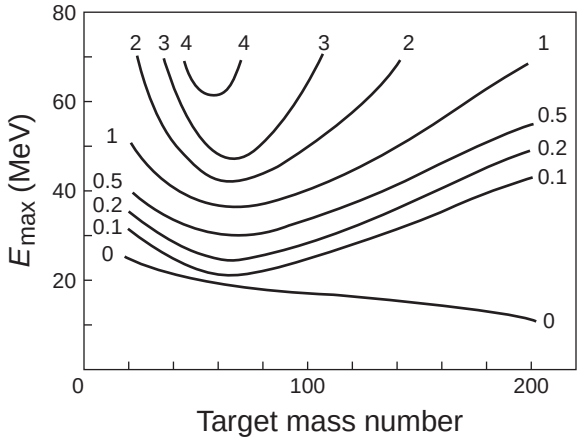

Figure 6 Yields of the (γ,pn) reactions as a function of bremsstrahlung maximum energy and target mass number. The numerical values in the figure are yields per equivalent quanta (in mb). (Reproduced from K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 573 (1978) by permission of North-Holland Publishing Co.)

maximum energy, respectively.[21] In each figure, function graphs of the respective reaction yields are drawn. The yield "zero" marks the calculated mass threshold.

The (γ,n) reactions start to take place in the 10 MeV region, and the yield increases rapidly with energy up to 30 MeV. The curves are relatively parallel to the energy axis above 30 MeV. It is found that the yield is dependent on the mass number rather than on the excitation energy in this energy region.

The charged particle emission reactions such as (γ,p) and (γ,pn) show quite different behavior. The functions run through a maximum for mass numbers between 50 and 60. The reaction yields decrease with increasing mass number at the given bremsstrahlung maximum energy. This is due to the Coulomb potential barrier restriction of high-Z nuclei. Below 50 MeV, the yield of certain mass numbers is dependent on the bremsstrahlung maximum energy. The yield of the (γ,pn) reactions shows similar patterns to that of the (γ,p) reactions though the former have higher threshold and lower yields than the latter.

Similar charts of photonuclear reaction cross-section and the trends of integrated cross-section were also given by Barbier.[22] He also applied the simple calculation method of induced activity. The excitation curves for the photofission reaction, the photoproduction of π-mesons and other various reactions up to the gigaelectron-volt region are shown in this paper.

The (γ,n) reaction is useful for the determination of light to heavy elements and the (γ,p) reaction is favorable for the analysis of medium-Z (from Na up to Sn) elements.

3 ANALYTICAL APPLICATION OF PHOTON ACTIVATION

3.1 Conditions of Photon Activation Analysis

At constant photon flux during irradiation, radioactivity induced in an element to be determined can be calculated as shown in Equation (4):

$$A = \frac{wN_0\theta}{AW}\sigma_Q Q[1 - \exp(-\lambda t)] \qquad (4)$$

where w is the amount of element, θ is the isotopic abundance of the target isotope, AW is the relative atomic mass of element, σ_Q is the yield (cm^2 per equivalent quanta), Q is the flux density (equivalent quanta $cm^{-2}\,s^{-1}$), λ is the decay constant of the induced radioisotope, and t is the irradiation time.

In order to produce bremsstrahlung efficiently, metals of high Z and high melting point, such as platinum and tungsten, are recommended as a converter material. Electron energy which is higher than about 15 MeV is indispensable to induce giant resonance reactions. Beam currents of several tens to hundreds of microamperes are necessary to induce radioactivity giving sufficient detection sensitivity. Betatrons and synchrotrons are not suitable for PAA due to their low beam currents. In Van de Graaf accelerators, beam current is in the microampere region, but the acceleration energy is low. However, many types of electron linear accelerators and microtrons are suitable for PAA.

As an example, consider the induced radioactivity estimated using Equation (1). The equivalent quanta obtained under the irradiation conditions mentioned above are about $10^{14}\,cm^{-2}\,s^{-1}$. The yield of (γ,n) reaction is around 100 mb ($10^{-25}\,cm^2$) for elements in the relative atomic mass region of around 100 (say molybdenum through silver). When $1\,\mu g\,g^{-1}$ of an element in a sample of 100 mg is analyzed, the number of target nuclides is about 10^{14}. After irradiation for a duration equal to 1/10 of the half-life of the radioisotope, induced activity is about 10^3 disintegrations s^{-1}. When the detection efficiency is 1%, 10 counts s^{-1} can be obtained. To obtain γ-ray peak counts within a counting error of 1%, a counting time of 1000 s is necessary at least. Therefore, it seems to be efficient to simply increase the irradiation time, flux, sample mass, detector efficiency and counting time. However, in the instrumental approach (analysis without radiochemical separation), the detection limit cannot always be improved so simply since the background activity might also be increased. The choice of irradiation energy and the cooling time is more effective to attain better detection limits.

Sensitivity data have been published which were obtained for several conditions of irradiation and

measurement, e.g. irradiation energies of 20 MeV,[17] 30 MeV,[13] 40–44 MeV[23] and 110 MeV.[24] Lutz calculated sensitivities for electron energies of 25, 30 and 35 MeV.[15]

3.2 Characteristics of Photon Activation Analysis

Since PAA is a complementary method to the more familiar NAA, the characteristics of both are compared in the following.

3.2.1 Transmittance of High-energy Photons

As high-energy photons can penetrate and activate mostly without notable matrix effects, PAA is suitable for bulk analysis of various kinds of materials and for simultaneous irradiation of several samples. The photon flux decreases with distance from the conversion target. Moreover, the bremsstrahlung beam exhibits a sharp flux density gradient perpendicular to the beam axis. Thus the activating radiation doses of samples and comparison standards should be monitored.

3.2.2 Systematic Trend of Photonuclear Reaction Yields

The (γ,n) reactions are mostly utilized in PAA. As shown in Figures 3 and 4, the threshold energies of the (γ,n) reactions are about 10 MeV for heavy elements and about 20 MeV for light elements. The (γ,n) reaction yields of heavy elements are up to two orders of magnitude higher than those of light elements at the bremsstrahlung maximum energy of 30 MeV. However, light elements are usually major components and heavy elements exist at trace levels in natural products. As the radioactivity induced in an element is roughly proportional to the product of the yield and the abundance of the pertinent element, the differences in radioactivities induced in the light and heavy element fractions are frequently of the same order of magnitude. Such a systematic trend of photonuclear reaction yields gives the well-balanced sensitivities for the instrumental analysis of natural samples.

3.2.3 Interference by Self-shielding Effect, Matrix Activity and Fission Products

In NAA, the self-shielding effect and intense radioactivity sometimes interfere with trace element analysis, as there are several nuclides which have anomalously large neutron capture cross-sections. Typical examples of self-shielding are Cd and B. In NAA, after neutron exposure, ^{24}Na, ^{32}P and ^{82}Br are major activities in biological materials and ^{56}Mn and ^{24}Na are major activities in rock, dust, soil and related samples because of their large neutron activation cross-sections. Rare-earth metals such as Eu and transition metals such as Co also lead to intense radioactivities and self-shielding. In PAA, such a self-shielding effect is mostly negligible, and the problem of matrix activity is not so frequent because of the penetrability of bremsstrahlung and, respectively, the systematic character of photonuclear reaction yields as mentioned above.

When U and/or Th are present in a sample in large amounts, neutron-induced fission products often interfere with the analysis of medium to heavy elements because some of the nuclides formed by fission are the same as those used for NAA of another element. A respective interference might also occur in PAA, but not to the extent as in NAA.

In NAA, the determination of Mg, Ti, Ni, Y, Zr, Nb, I, Tl, Pb and the light elements is very difficult. However, these elements are easily determined by PAA because photonuclear reaction products of these elements have suitable nuclear characteristics for photon spectrometry. More than 30 elements in geological and environmental samples can be determined instrumentally. To reduce radioactivities from light elements and to improve the signal-to-background ratio for heavy elements, 20 MeV irradiation is sometimes preferable to 30 MeV irradiation. Most of the medium to heavy elements can be determined at concentration levels of micrograms per gram and below.

3.2.4 Characteristics of the (γ,n) Reaction Products

Radioisotopes induced by the (γ,n) reaction decay with positron emission and/or EC in the most cases. Several light elements, such as carbon, nitrogen, oxygen and fluorine, lead to purely positron-emitting nuclides. These can be determined by the coincidence counting method (see below, section 4.1). Many of the (γ,n) products from medium to heavy elements decay by EC. In EC decay, the detection of characteristic X-rays of the decay product is convenient for analytical purposes. Low-energy photon spectrometers (see above) are useful to detect characteristic X-rays and low-energy γ-rays.

4 ANALYTICAL PROCEDURES

4.1 Procedure for Light-element Analysis

As mentioned above, the photonuclear activation products of some light elements like C, N, O and F are pure positron emitters without any characteristic photon emission. The annihilation photon quanta ($E = 511$ keV) produced through reaction of positrons with electrons in close vicinity can be measured most favorably with a pair of γ-ray detectors facing one another and electronically switched in coincidence.[6,7] Since the annihilation

quanta, each representing the rest mass of an electron, are emitted from the source at an angle of 180°, the pulse processing electronic system is set up to exclusively count quanta coincidentally absorbed in both detectors. Thus, radiation background is suppressed to a great extent. Because the photon energy is unspecific the activation product (^{11}C, ^{13}N, ^{15}O, ^{18}F respectively) usually has to be separated from the sample matrix after bremsstrahlung exposure. The different appropriate radiochemical separation procedures are described in detail in the literature.[3,4] Heat extraction of radiocarbon, radionitrogen and radiooxygen followed by collection of the respective radiofractions with appropriate agents, as well as vapor distillation of radiofluorine, are utilized mostly. The radiochemical purity of the separated fraction is checked by recording a decay curve during photon counting and comparing the half-life thus obtained with the literature value. Analyses without chemical separations are possible in exceptionally favorable cases only.

The practical analytical procedure typically might be as follows: The analytical sample is prepurified and wrapped in aluminum foil. The comparison standard (usually the pure element, e.g. graphite in the case of carbon analysis, or a stoichiometrically well-determined compound) is also wrapped, and both samples are stacked in an irradiation container, e.g. an aluminum rabbit. This is then transported pneumatically to the irradiation position. The exposure period is adjusted according to the respective half-life and to the expected concentration to be determined. After irradiation, the rabbit is transported back into the laboratory where the samples are unpacked, taken out of the aluminum foil and eventually etched with a cleaning agent to remove surface contaminants. Thereafter, the sample is subject to radiochemical separation. Then the collecting substance containing the radionuclide to be measured is brought into the counting position between the detector pair, followed by multiple countings at a fixed cycle. Counting cycle and data acquisition are performed by a personal computer. The comparison standard is then treated likewise, eventually omitting the cleaning step. The element content is then calculated by comparison of the integral count numbers of the sample and the standard, respectively.

4.2 Procedure for Multielement Analysis

Multielement analysis of natural samples by the comparative method is described as a typical case of PAA. The comparative method (see section 4.1) is mostly used in PAA. When a sample is irradiated together with a comparative standard that contains w_{std} of the element to be determined the unknown amount (w_x) of element in the sample is obtained by Equation (5):

$$w_x = w_{std} \left(\frac{A_x}{A_{std}} \right) \left(\frac{\phi_{std}}{\phi_x} \right) \quad (5)$$

where A_x and A_{std} are the activities induced from the element to be determined and ϕ_x and ϕ_{std} are fluxes of bremsstrahlung irradiated on the sample and the comparative standard respectively.

Usually, the γ-ray peak area ratio of the sample and the standard is used instead of the activity ratio (A_x/A_{std}). In this case, the counting geometry (such as the shapes of sample and standard, the distance from sample and detector) must be equal. As the radioactivity decays according to its half-life after irradiation and also during measurement, the radioactivity of sample and standard should be normalized to the same time, usually the end of irradiation. When the obtained peak count number (C) is normalized to the end of irradiation, the initial activity (A) is calculated by Equation (6):

$$A = \frac{C\lambda}{\exp(-\lambda t_c)[1 - \exp(-\lambda t_m)]} \quad (6)$$

where λ is the decay constant $[\ln(2)/t_{1/2}]$ for the radioisotope, t_c is the cooling time after irradiation, and t_m is the measuring time.

4.3 Sample Preparation, Standard and Flux Monitoring

For the comparative method, a standard containing a well-known amount of each component to be analyzed must be available. This might be a multielement or a single element comparator. It is desirable that such a standard and samples have similar elemental compositions. These materials should be prepared under strictly controlled conditions following preparation procedures which enable maximum accuracy of the respective element concentrations. Certified reference materials (CRMs) should not be used for highly precise concentration calculations, but rather as materials analysed for quality control of the obtained results, the accuracy in particular.

As various kinds of CRMs are issued by many organizations in the world, a standard suitable for nearly each analytical purpose can be chosen. These CRMs are analyzed by several analytical methods at several laboratories. However, most of these certified values are averaged values with standard deviations. Usually only limited numbers of elements are certified. There are several examples where the concentration of the element to be determined is not certified or its concentration is largely different from that in the sample. In order to avoid this, two or more CRMs can be used for cross-checking. When natural products are used as CRMs, homogeneity and stability should be checked carefully. Also limited

amounts of CRMs have been supplied. In most cases the use of CRMs as comparators is not recommended, since high-precision results as required in certification analyses cannot be expected.

Laboratory standards have been useful to solve the above problems. A filter paper doped with several standard solutions of elements to be determined is often used in NAA. This method cannot be recommended for PAA because a small amount of residual acid might decompose the cellulose paper during irradiation. Moreover, the homogeneity of filter paper is questionable. Phenol–formaldehyde resin[25] and gelatine[26] were used as the base materials for activation analysis of biological materials. An artificial standard of polyacrylamide gel matrix was also developed by utilizing the copolymerization reaction of acrylamide in a mixed solution containing the appropriate amounts of elements.[27] Polyacrylamide has high tolerance for radiation and a similar matrix composition to biological materials. For the analysis of rock samples, multielement solution was gelatinized by hydrolysis of tetraethylsilicate.[28,29] As the gelatinization takes place in acidic conditions, most of the standard solution of metals can be used directly. Silica gel has high stability for irradiation. The elemental abundance in such synthetic standards is easily and accurately obtained as the weight ratio of the added amount of element to the final weight of gel, and elemental composition can be adjusted freely according to samples to be analyzed and elements to be determined. The homogeneity of these standards is usually satisfactory because the mixed solution containing several elements is solidified as is.

As photon flux and energy spectrum show pronounced gradients, bremsstrahlung irradiated on each sample should be monitored individually. To estimate flux, sample and standard are stacked alternately or irradiated with flux monitors. The accuracy of the former method is good because the specific activity of each element in the standard is plotted against the irradiation position and the specific activity induced in the sample can be estimated. However, the measurement of samples and standards irradiated is very time-consuming. In the latter method, the monitor reaction data are largely different from the reaction for analysis of each element. Therefore, it is recommended to choose the monitor reaction that has a similar excitation function to the reaction for analysis, and a product radioisotope that has a half-life similar to those of the radioisotopes to be determined.

4.4 Irradiation

An example of an irradiation set-up[30] is shown in Figure 7. The beam profile can be observed with a thin beryllium oxide disk of 0.2 mm in thickness and the beam position is precisely adjusted with a platinum position monitor made up of a thin wire and a small flat tip connected to an electronic microammeter. The platinum tip is positively charged by secondary electron emission when hit by the electron beam. Therefore, the beam position can be pointed exactly to the center of sample by scanning for the maximum current of the microammeter. The beam intensity is indirectly monitored with a ferrite-core-loaded toroidal coil coupled with a current integrator after rectification, because electric current is induced on the toroidal coil when a pulsed beam is passing through the core. This core monitor has been calibrated with a pulse generator. The induced pulses are also monitored

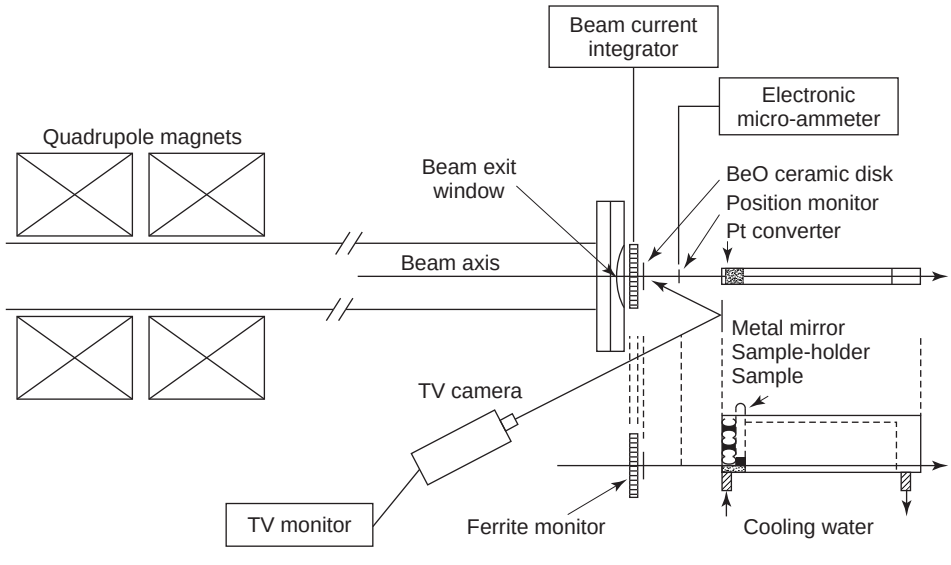

Figure 7 Schematic drawing of the irradiation system. (Reproduced by permission of Akademica Kiado, Budapest.)

with an oscilloscope. Four pieces of platinum plate of 0.5 mm thickness are placed inside the sample holder to serve as the conversion target. The converter and sample are cooled with water to avoid melting of materials caused by high incident beam power. The beam power, which is the product of electron energy and average beam current, turns to heat finally. In this facility, irradiation, for activation analysis has usually been performed at 30 MeV electron energy and 150 μA average current. Hence, 4.5 kW of beam power is focused within a diameter of a few millimeters.

Bremsstrahlung does not cause such significant radiation damage on the sample. Electrons penetrating the converter cause radiation damage and local heating which creates severe problems especially for biological samples. To reduce effects caused by penetrating electrons, a graphite block has been inserted between the converter and sample. Light elements absorb electrons efficiently without reducing the flux of bremsstrahlung from the converter considerably. Also, a bending magnet placed behind the converter to deflect electrons passing through the converter has been applied. In this case, samples should be kept at a good distance from the converter to avoid unwanted electron bombardment. Another possibility to avoid electron-induced damage is to use a thick target which absorbs all electrons (see above).

Secondary neutrons are mainly emitted from the converter and also from surrounding materials. These neutrons are predominantly induced by photonuclear reactions in the giant resonance region. It is known that energy spectra of photoneutrons are predicted by the statistical model of the nucleus, and the angular distribution is mainly isotropic.[31] Photoneutrons are retarded by the collision with surrounding materials and thermalized by water and other light material in the vicinity. Such neutrons cause the (n,γ), (n,p) and (n,α) reactions in a sample.

The spatial distributions of bremsstrahlung and photoneutrons are shown in Figure 8.[32] Figure 8 shows the results of Pt (2 mm) plus graphite (10 mm) and Pt (2 mm), W (3 mm) and graphite (10 mm). The solid lines show the distribution of activity of ^{196}Au induced by the ^{197}Au(γ,n)^{196}Au reaction which corresponds to the flux of bremsstrahlung. In Figure 8, the spatial distribution of neutrons is drawn by dotted lines, according to the ^{198}Au activity induced by the ^{197}Au(n,γ)^{198}Au reaction. According to the result of Engelmann,[6] the best thickness of

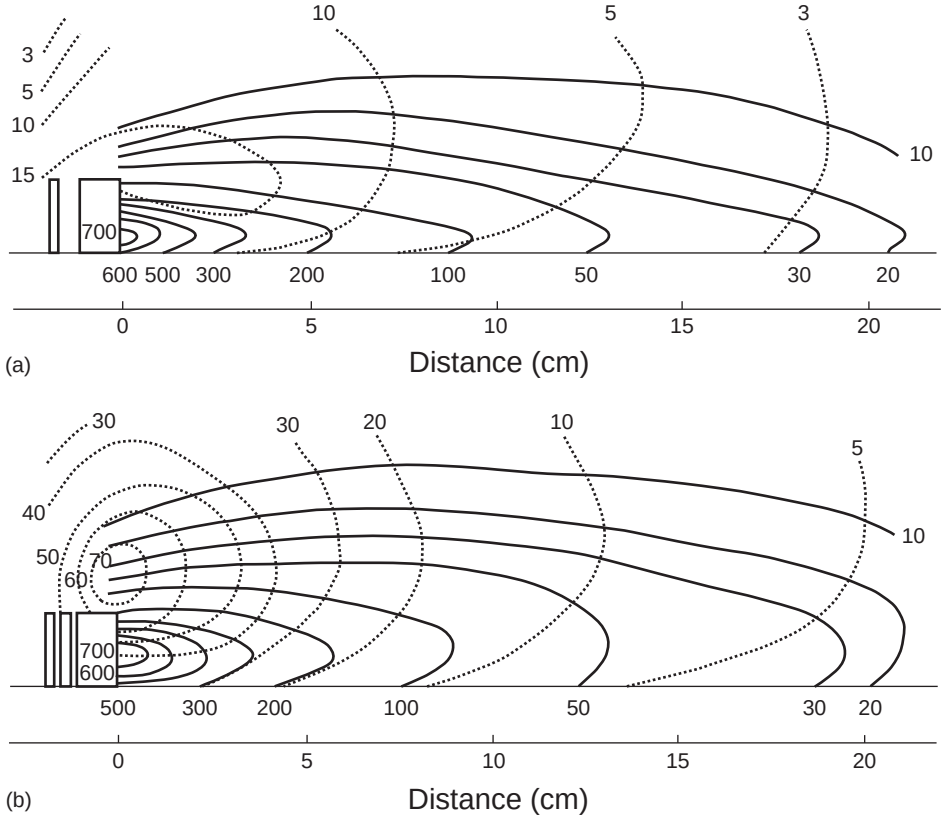

Figure 8 Distribution of the intensity of bremsstrahlung and photoneutrons in sample holder by the irradiation of 30 MeV electrons: (a) 2 mm Pt converter + 10 mm graphite; (b) 2 mm Pt + 3 mm W converter + 10 mm graphite. Straight line: ^{196}Au activity (354 keV γ-ray peak counts $s^{-1}mg^{-1}$). Dotted line: ^{198}Au activity (412 keV γ-ray peak counts $s^{-1}mg^{-1}$).

platinum converter to induce (γ,n) reactions is 3 g cm^{-2} at 30 MeV and 4.5 g cm^{-2} at 45 MeV. As the thickness of 2 mm is 4.3 g cm^{-2}, the converter thickness presented in Figure 8(a) is suitable for activation. In this case, electron energy is reduced by half of the initial energy. The target combination of Figure 8(b) is used for the irradiation of biological materials. The additional converter of 3 mm thick W plate can reduce the initial electron energy by a factor of six, and the graphite block of 10 mm in thickness can fully absorb electrons passing through the converter. As this graphite block absorbs only 3 to 4% of photons at the energy of 10 to 30 MeV, photon flux is not suppressed significantly. As shown in Figure 8(b), the bremsstrahlung beam produced by the converter of 5 mm in thickness is broadened, and the photoneutron flux is increased. The full widths at half-maximum of bremsstrahlung induced by 2- and 5-mm-thick converters are 12 and 16 mm, respectively, directly behind

Table 1 Pertinent nuclear data for PAA of natural products

Element	Target nuclide	Isotopic abundance (%)	Reaction	Product nuclide	Half-life[a]	γ-Ray energy (keV)
C	^{12}C	98.89	(γ,n)	^{11}C	20.39 m	511
N	^{14}N	99.63	(γ,n)	^{13}N	9.97 m	511
O	^{16}O	99.76	(γ,n)	^{15}O	1.18 m	511
F	^{19}F	100	(γ,n)	^{18}F	1.830 h	511
Na	^{23}Na	100	(γ,n)	^{22}Na	2.602 y	1274.5
Mg	^{25}Mg	10.0	(γ,p)	^{24}Na	14.66 h	1368.6
Si	^{30}Si	3.1	(γ,p)	^{29}Al	6.6 m	1273.4
P	^{31}P	100	(γ,n)	^{30}P	2.50 m	511
Cl	^{35}Cl	75.77	(γ,n)	^{34m}Cl	32.2 m	146.4, 2127.7
K	^{39}K	93.26	(γ,n)	^{38}K	7.64 m	2167.7
Ca	^{44}Ca	2.09	(γ,p)	^{43}K	22.3 h	372.8, 617.5
Ca	^{48}Ca	0.187	(γ,n)	^{47}Ca	4.536 d	1297.1
Sc	^{45}Sc	100	(γ,n)	^{44}Sc	3.93 h	1157
Ti	^{47}Ti	7.4	(γ,p)	^{46}Sc	83.83 d	889.3, 1120.5
Ti	^{48}Ti	73.7	(γ,p)	^{47}Sc	3.341 d	159.4
Ti	^{49}Ti	5.4	(γ,p)	^{48}Sc	1.821 d	983.5, 1037.5, 1312.0
Cr	^{52}Cr	73.79	(γ,n)	^{51}Cr	27.704 d	320.1
Mn	^{55}Mn	100	(γ,n)	^{54}Mn	312.2 d	834.8
Fe	^{57}Fe	2.15	(γ,p)	^{56}Mn	2.578 h	846.8
Co	^{59}Co	100	(γ,n)	^{58}Co	70.92 d	810.8
Ni	^{58}Ni	68.3	(γ,n)	^{57}Ni	1.503 d	1377.6
Cu	^{65}Cu	30.8	(γ,n)	^{64}Cu	12.701 h	511.0, 1345.8
Zn	^{66}Zn	27.9	(γ,n)	^{65}Zn	244.1 d	1115.5
Zn	^{68}Zn	18.8	(γ,p)	^{67}Cu	2.580 d	184.6
Ga	^{69}Ga	60.1	(γ,2n)	^{67}Ga	3.261 d	184.6, 300.2
As	^{75}As	100	(γ,n)	^{74}As	17.78 d	595.9
Se	^{76}Se	9.0	(γ,n)	^{75}Se	119.77 d	136.0, 264.7
Br	^{79}Br	50.69	(γ,2n)	^{77}Br	2.3765 d	239.0, 520.6
Rb	^{85}Rb	72.17	(γ,n)	^{84}Rb	32.9 d	881.7
Sr	^{88}Sr	82.6	(γ,n)	^{87m}sr	2.80 h	388.4
Y	^{89}Y	100	(γ,n)	^{88}Y	106.61 d	898.1, 1836.1
Zr	^{90}Zr	51.5	(γ,n)	^{89}Zr	3.268 d	909.2
Nb	^{93}Nb	100	(γ,n)	^{92m}Nb	10.15 d	934.5
Mo	^{100}Mo	9.6	(γ,n)	^{99}Mo	2.7477 d	140.5
Sb	^{123}Sb	42.7	(γ,n)	^{122}Sb	2.70 d	564.4
I	^{127}I	100	(γ,n)	^{126}I	13.0 d	388.6, 666.4
Cs	^{133}Cs	100	(γ,n)	^{132}Cs	6.48 d	667.7
Ba	^{136}Ba	7.85	(γ,n)	^{135m}Ba	1.20 d	268.3
Ce	^{140}Ce	88.5	(γ,n)	^{139}Ce	137.66 d	165.9
Tm	^{169}Tm	100	(γ,n)	^{168}Tm	93.1 d	198.2, 815.9
Tl	^{203}Tl	29.5	(γ,n)	^{202}Tl	12.23 d	439.6
Pb	^{204}Pb	1.42	(γ,n)	^{203}Pb	2.169 d	279.2
U	^{238}U	99.275	(γ,n)	^{237}U	6.75 d	208

[a] y, years; m, months; d, days, h, hours.

the graphite block. Through quasi-elastical scattering the photoneutrons are also distributed behind the graphite block. The spatial distribution and neutron thermalization are influenced by the irradiation conditions. To estimate the interference caused by photoneutrons it is recommended that neutron monitors be applied adjacent to the samples during each radiation cycle.

High irradiation energy is preferable to obtain better sensitivity. However, as several types of photonuclear reactions are increasingly induced at the incident electron energy above 30 MeV, nuclear interference problems arise at higher irradiation energy.

4.5 Measurement and Evaluation

The elements and nuclear data for PAA of typical natural samples are listed in Table 1. Since each radioisotope has a specific half-life, detection sensitivity and signal-to-background ratio, which depend upon both matrix activity and radioactivity to be analyzed, are strongly decay-time dependent. Figure 9 shows the time dependency of detection limits of nine elements in coal fly ash (NIST Standard Reference Material (SRM) 1633a) irradiated at 25 MeV for 2 h. The detection limit is defined as the concentration of element which yields the count number equal to 3σ of background counts under pertinent γ-ray photopeak. As shown in Figure 9,[32] the detection limit strongly varies with cooling time. For multielement analysis of natural samples, γ-ray spectra of samples should be measured several times at appropriate intervals after irradiation. In Figure 10,[32] the detection limits of 26 nuclides are plotted against a suitable cooling time in the case of coal fly ash quoted above. The degree of interference from γ-ray overlapping also changes according to the half-lives of pertinent nuclides. Therefore, the cooling time should be carefully adjusted in order to analyze each element under the best possible conditions.

As a result, typical γ-ray measurements for the analysis of the type of natural products studies are performed preferably in the following periods: (1) a few minutes after irradiation for K, Cl; (2) a few hours after irradiation for Ti(^{47}Sc), Sc, Fe, Sr; (3) a few days after irradiation for Mg, Ca(^{43}K), Ti(^{48}Sc), Ni, Ba, Pb; (4) 1 week after irradiation for Ca(^{47}Ca), Zn(^{67}Cu), Cs, Zr, Mo, Sb, U; and (5) several weeks after irradiation for Na, Ti(^{46}Sc), Cr, Mn, Co, Zn(^{65}Zn), As, Rb, Nb, Y, I, Ce, Tl. Two or more radioisotopes are available for the determination of several elements, e.g. Ca, Ti and Zn.

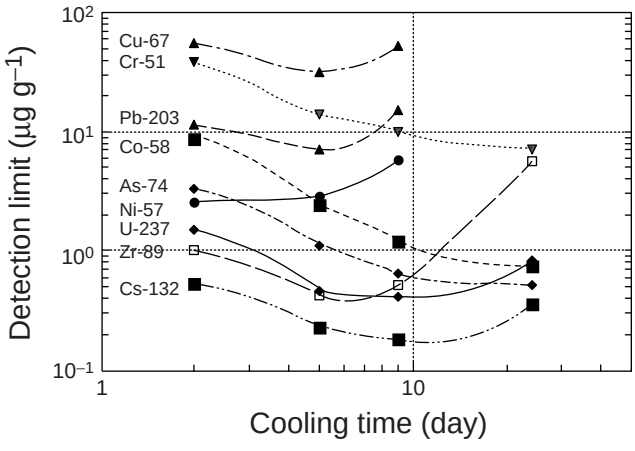

Figure 9 Cooling time dependence of the detection limit of nine elements in coal fly ash.

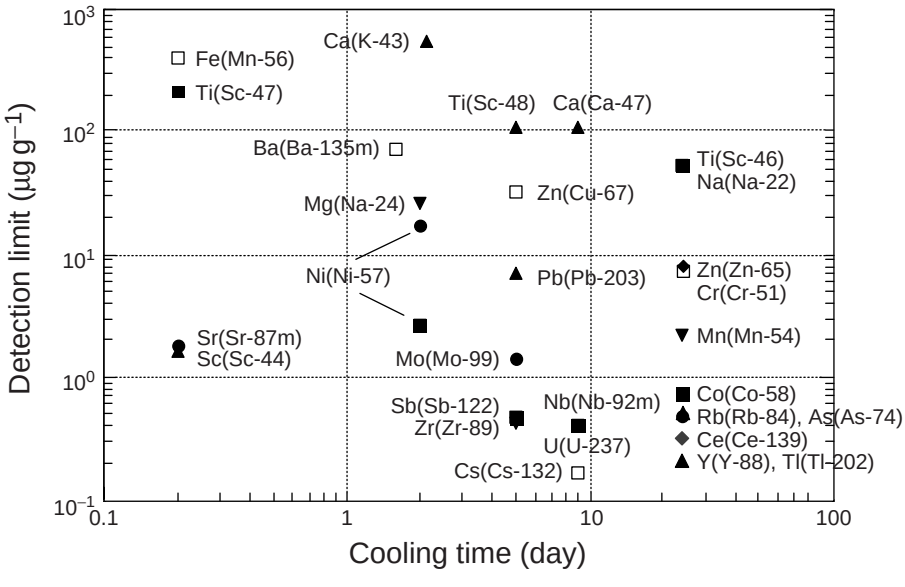

Figure 10 Detection limits of 26 nuclides obtained at a suitable cooling time in coal fly ash.

5 INTERFERENCE MANAGEMENT

5.1 Interference Reactions

In PAA, a radioisotope used for determination can be produced by another element. The interference reactions have been discussed by Kato et al.,[20,21] Williams and Hislop,[23] Segebade et al.[33] and Miyamoto et al.[34]

Typical examples of nuclear interference are listed in Table 2. For example, $^{55}Mn(\gamma,n)^{54}Mn$ and $^{52}Cr(\gamma,n)^{51}Cr$ reactions are used for the determination of Mn and Cr. However, ^{54}Mn and ^{51}Cr can be produced by $^{56}Fe(\gamma,pn)^{54}Mn$ and $^{56}Fe(\gamma,\alpha n)^{51}Cr$ reactions from Fe. The degree of interference is strongly dependent upon the irradiation conditions and, in particular, the Mn/Fe and Cr/Fe abundance ratios. Hence, when natural materials are analyzed the interference of Cr and Mn might be crucial since usually Fe is in significantly higher concentrations than the former elements in these materials. For the determination of Na(^{22}Na) through $^{23}Na(\gamma,n)^{22}Na$ the analytical reaction suffers interference from $^{24}Mg(\gamma,pn)^{22}Na$ and $^{27}Al(\gamma,\alpha n)^{22}Na$. The interferences from Mg and Al should be checked as well. The interference reactions are always accompanied by charged particle emission. As these reactions are suppressed by the nuclear Coulomb barrier, interference reactions are not so severe for heavier elements.

Photoneutrons also cause interference problems for the analysis of several elements. For example, ^{24}Na from $^{25}Mg(\gamma,p)^{24}Na$ for the determination of Mg and ^{56}Mn from $^{57}Fe(\gamma,p)^{56}Mn$ for the determination of Fe suffer interference from $^{23}Na(n,\gamma)^{24}Na$ and $^{55}Mn(n,\gamma)^{56}Mn$, respectively. The distribution of photoneutrons is critically dependent on the irradiation condition and set-up as shown in Figure 8. Table 3 shows the weight ratio of the amount to yield the same activity from interference and analytical reaction. The weight ratios (yield of analytical reaction/yield of interfering reaction; calculated for equal element masses) were obtained at several irradiation positions using a 5 mm thick converter. The weight ratios in interference cases caused by photonuclear reactions decrease with the distance from the graphite block. Hence, the interference yields increase with the distance from the converter. In this interference case, the change of the weight ratio reflects the change of bremsstrahlung spectrum. This is due to the larger scattering cross-section of lower photon energies. In the case of photoneutron interference, the fast neutron flux rapidly decreases with the distance from the converter; consequently, the interference yield by slow neutrons becomes more significant than that by fast neutrons.

The nuclear reaction yields of the analytical reaction and the interference reaction were studied systematically.[21] The results are listed in Table 4 for the SRM orchard leaves (NIST SRM 1571) and in Table 5 for the geochemical reference sample JB-1 (basalt) issued by the Geological Survey of Japan. The magnitude of interference, expressed as the percentage of the activity induced by the interference reaction to the total activity, increases with the irradiation energy.

Table 2 Interference reactions in PAA

Element	Analytical reaction	*Q*-value (MeV)	Interfering reaction	*Q*-value (MeV)
C	$^{12}C(\gamma,n)^{11}C$	18.7	$^{16}O(\gamma,\alpha n)^{11}C$	25.9
			$^{14}N(\gamma,p2n)^{11}C$	31.2
F	$^{19}F(\gamma,n)^{18}F$	10.4	$^{23}Na(\gamma,\alpha n)18F$	20.9
Na	$^{23}Na(\gamma,n)^{22}Na$	12.4	$^{24}Mg(\gamma,pn)^{22}Na$	19.2
			$^{27}Al(\gamma,\alpha n)^{22}Na$	22.5
Mg	$^{25}Mg(\gamma,p)^{24}Na$	12.1	$^{27}Al(\gamma,2pn)^{24}Na$	31.4
			$^{27}Al(n,\alpha)^{24}Na$	–[a]
			$^{23}Na(n,\gamma)^{24}Na$	–[a]
Cl	$^{35}Cl(\gamma,n)^{34m}Cl$	12.6	$^{39}K(\gamma,\alpha n)^{34m}Cl$	19.9
K	$^{39}K(\gamma,n)^{38}K$	13.1	$^{40}Ca(\gamma,pn)^{38}K$	21.4
Sc	$^{45}Sc(\gamma,n)^{44}Sc$	11.3	$^{46}Ti(\gamma,pn)^{44}Sc$	21.7
Ti	$^{48}Ti(\gamma,p)^{47}Sc$	11.4	$^{51}V(\gamma,\alpha)^{47}Sc$	10.3
			$^{48}Ca(\gamma,n)^{47}Ca\rightarrow{}^{47}Sc$	10.0
Cr	$^{52}Cr(\gamma,n)^{51}Cr$	12.0	$^{56}Fe(\gamma,\alpha n)^{51}Cr$	19.7
			$^{54}Fe(n,\alpha)^{51}Cr$	–[a]
Mn	$^{55}Mn(\gamma,n)^{54}Mn$	10.2	$^{56}Fe(\gamma,pn)^{54}Mn$	20.9
			$^{54}Fe(n,p)^{54}Mn$	–[a]
Fe	$^{57}Fe(\gamma,p)^{56}Mn$	10.6	$^{55}Mn(n,\gamma)^{56}Mn$	–[a]
			$^{59}Co(\gamma,2pn)^{56}Mn$	28.0
Co	$^{59}Co(\gamma,n)^{58}Co$	10.5	$^{60}Ni(\gamma,pn)^{58}Co$	20.0
			$^{63}Cu(\gamma,\alpha n)^{58}Co$	16.2
Zn	$^{68}Zn(\gamma,p)^{67}Cu$	10.0	$^{71}Ga(\gamma,\alpha)^{67}Cu$	5.3
As	$^{75}As(\gamma,n)^{74}As$	10.2	$^{76}Se(\gamma,pn)^{74}As$	19.8
Rb	$^{85}Rb(\gamma,n)^{84}Rb$	10.5	$^{86}Sr(\gamma,pn)^{84}Rb$	20.1
Sr	$^{88}Sr(\gamma,n)^{87m}Sr$	11.1	$^{89}Y(\gamma,pn)^{87m}Sr$	18.2
Y	$^{89}Y(\gamma,n)^{88}Y$	11.5	$^{90}Zr(\gamma,pn)^{88}Y$	19.8

[a] Photoneutron reaction.

5.2 Peak Overlapping

Another type of interference is the spectral overlap of peaks whose energies are so close that the spectrometer cannot resolve them. These might be due to signals of different nuclides or one product nuclide whereby the former case is more critical than the latter. However,

Table 3 Weight ratio of several interference reactions (converter thickness = 5 mm)

Interfering reaction	Analytical reaction	Distance from graphite block (cm)		
		0	6	14
$^{56}Fe(\gamma,pn)^{54}Mn$	$^{55}Mn(\gamma,n)^{54}Mn$	92	71	60
$^{55}Mn(n,\gamma)^{56}Mn$	$^{57}Fe(\gamma,p)^{56}Mn$	0.29	0.20	0.18
$^{23}Na(n,\gamma)^{24}Na$	$^{25}Mg(\gamma,p)^{24}Na$	29	19	17
$^{27}Al(n,\alpha)^{24}Na$	$^{25}Mg(\gamma,p)^{24}Na$	204	185	165

Table 4 Interference in PAA of NIST SRM 1571 orchard leaves

Element	Abundance ratio	Magnitude of interference (%)[a] E_{max}		
		30 MeV	40 MeV	60 MeV
Na(^{22}Na)	Mg/Na = 75.6	51.9	88.7	92.5
	Al/Na = 4.99	0.38	6.2	13.9
Mg(^{24}Na)	Al/Mg = 0.066	0.031	0.09	0.56
Mn(^{54}Mn)	Fe/Mn = 3.3	2.0	13.0	20.2
Cr(^{51}Cr)	Fe/Cr = 130	0.76	37.3	54.2
Zr(^{89}Zr)	Mo/Zr = 15.4	0.073	0.95	7.1
Fe(^{56}Mn)	Mn/Fe = 0.303	9.77[b]		
Mg(^{24}Na)	Na/Mg = 0.0132	0.000059[b]		

[a] (Activity from interfering element)/(Total activity) × 100.
[b] Interference caused by photoneutron.

Table 5 Interference in PAA of JB-1 (basalt) issued from the Geological Survey of Japan

Element	Abundance ratio	Magnitude of interference (%)[a] E_{max}		
		30 MeV	40 MeV	60 MeV
Na(^{22}Na)	Mg/Na = 2.25	2.83	17.7	26.7
	Al/Na = 3.70	0.28	4.69	10.7
Mg(^{24}Na)	Al/Mg = 1.65	0.78	2.2	12.1
Ti(^{47}Sc)	Ca/Ti = 8.32	12.2	10.8	8.1
	V/Ti = 0.0266	0.12	0.13	0.13
Mn(^{54}Mn)	Fe/Mn = 54.1	31.1	71.1	80.6
Cr(^{51}Cr)	Fe/Cr = 155	0.96	41.3	58.5
Y(^{88}Y)	Zr/Y = 6.00	0.83	–	4.1
	Nb/Y = 0.59	–	–	0.067
Zr(^{89}Zr)	Mo/Zr = 0.15	0.00071	0.0094	0.075
Nb(^{92m}Nb)	Mo/Nb = 1.5	0.096	0.53	0.95
Fe(^{56}Mn)	Mn/Fe = 0.303	0.66[b]		
Mg(^{24}Na)	Na/Mg = 0.0132	0.20[b]		

[a] (Activity from interfering element)/(Total activity) × 100.
[b] Interference caused by photoneutron.

the use of doublet or multiplet peaks from one product nuclide should also be avoided. Typical examples for peak overlap interference by different nuclides are listed in Table 6. This type of interference mainly occurs during analysis of medium to heavy elements. To avoid the interference the difference of half-lives of product nuclides can be effectively used in the following cases: Co/Ca, As/Ca, Tl/Zn and Pb/Hg. However, this frequently entails excessive time consumption in the analytical task. On the other hand, if one nuclide emits two or more γ-ray energies, the peak ratio method can be applied to subtract the interfering peak area. Typical examples are Zn/Ga, Nb/Fe, Cs/I, I/Y and U/Ga. The latter method reduces the time consumption, but a greater analytical error is induced.

Nowadays the γ-ray overlapping interference is not so problematic due to the dramatic improvement of the energy resolution power of modern photon spectrometers; a modern Ge detector offers peak half-widths of 1.8 keV or less at 1333 keV (^{60}Co). Moreover, several highly efficient peak deconvolution programs have been developed.

5.3 Other Errors

As mentioned in section 3.1, the induced radioactivity is proportional not only to sample mass but also to the cross-section and isotopic abundance of the sample components, flux of bremsstrahlung and saturation factor. In the case of γ-ray measurement, the peak count number obtained is dependent on the counting time, γ-ray branching ratio of radioisotope, detection efficiency of γ-ray and others. In the above factors, sample mass, irradiation time and counting period can be adjusted to reduce statistical errors and to improve precision and detection limit. However, the improvement of these factors is often restricted by the experimental conditions, such as number of samples and limited access to irradiation and measurement facilities. It is important to store the peak counts until the counting error becomes small enough.

Activation analysis includes the following steps: (1) sample preparation; (2) irradiation; (3) measurement; and (4) analysis. To improve accuracy, systematic errors

Table 6 Typical interferences caused by γ-ray overlapping

Element	Analytical reaction	Half-life	γ-ray (keV)	Interfering reaction	Half-life	γ-ray (keV)	Other γ-ray (keV)
Co	^{59}Co(γ,n)^{58}Co	70.92 d	810.8	^{48}Ca(γ,n)^{47}Ca	4.536 d	807.9	1297.1
Zn	^{68}Zn(γ,p)^{67}Cu	2.580 d	184.6	^{69}Ga(γ,2n)^{67}Ga	3.261 d	184.6	300.2
As	^{75}As(γ,n)^{74}As	17.78 d	595.9	^{44}Ca(γ,p)^{43}K	22.3 h	593.4	372.8, 617.5
Nb	^{93}Nb(γ,n)^{92m}Nb	10.15 d	934.5	^{54}Fe(γ,pn)^{52}Mn	5.591 d	935.5	744.2, 1434.1
I	^{127}I(γ,n)^{126}I	13.0 d	388.6	^{89}Y(γ,2n)^{87}Y	3.346 d	388.4	484.9
Cs	^{133}Cs(γ,n)^{132}Cs	6.48 d	667.7	^{127}I(γ,n)^{126}I	13.0 d	666.4	388.6
Tl	^{203}Tl(γ,n)^{202}Tl	12.23 d	439.6	^{70}Zn(γ,n)^{69m}Zn	13.76 h	438.6	
Pb	^{204}Pb(γ,n)^{203}Pb	2.169 d	279.2	^{204}Hg(γ,n)^{203}Hg	46.60 d	279.2	
U	^{238}U(γ,n)^{237}U	6.75 d	208.0	^{69}Ga(γ,2n)^{67}Ga	3.261 d	209.0	300.2

in each step should be checked carefully. Nuclear interference and γ-ray overlapping are avoidable, but some systematic errors, such as flux monitoring and counting geometry, cannot be cancelled out by quantitative methods. In the following section, the internal standard method, the standard addition method and the stable isotope dilution activation analysis used for highly accurate analyses are described.

6 QUANTITATIVE METHODS FOR QUALITY ASSURANCE

6.1 Internal Standard Method

The most reliable method for monitoring the activating flux received by a sample is the method using an internal standard because the flux monitoring component is inherent in both the sample and the comparative standard. This method was called the internal reference method by Tsai et al.[35] and Yagi and Masumoto[36] or the internal monitor method by Oka et al.,[37] in order to differentiate the slight change of the definition of each technique applied.

First, the comparative standard in which the abundance ratio (K) of the element (a) to be determined and the internal standard element (b) are accurately known is prepared. It is not necessary to know the absolute abundance of each element. The accurately weighed amount (w_b) of the internal standard is added to the sample and mixed homogeneously. Then, appropriate amounts of the sample and the comparative standard are irradiated. The activity ratio of a pertinent element to the internal standard is measured for the sample (R) and the comparative standard (R^*), respectively, and the relevant equations are shown as Equations (7) and (8).

$$R = \frac{A_a}{A_b} = \frac{(w_a N_0 \theta_a / AW_a)\sigma_a Q[1 - \exp(-\lambda_a t_a)]}{(w_b N_0 \theta_b / AW_b)\sigma_b Q[1 - \exp(-\lambda_b t_b)]} \quad (7)$$

$$R^* = \frac{A_a^*}{A_b^*} = \frac{(w_a^* N_0 \theta_a / AW_a)\sigma_a Q^*[1 - \exp(-\lambda_a t_a)]}{(w_b^* N_0 \theta_b / AW_b)\sigma_b Q^*[1 - \exp(-\lambda_b t_b)]} \quad (8)$$

The sample weight and flux (Q) are eliminated by introducing the activity ratio. Again, by including the ratio of the two activity ratios (R^*/R) mentioned above, nuclear reaction yields, saturation factor, isotope abundance and relative atomic mass can be cancelled out. Using Equation (9), abundance ratios of the sample (w_b/w_a) and the comparative standard (K) are left.

$$\frac{R^*}{R} = K\frac{w_b}{w_a} \quad (9)$$

Therefore, the abundance of element a (w_a) can be calculated by Equation (10):

$$w_a = K w_b \left(\frac{R^*}{R} \right. \quad (10)$$

When the ratio is obtained as the γ-ray peak area ratio, Equation (11) can be used.

$$\frac{R^*}{R} = \frac{(C_a^*/\phi(\gamma_a)\varepsilon(\gamma_a)\Omega^*)/(C_b^*/\phi(\gamma_b)\varepsilon(\gamma_b)\Omega^*)}{(C_a/\phi(\gamma_a)\varepsilon(\gamma_a)\Omega)/(C_b/\phi(\gamma_b)\varepsilon(\gamma_b)\Omega)} \quad (11)$$

where $\phi(\gamma)$ is the branching ratio of the γ-ray, $\varepsilon(\gamma)$ is the detector efficiency at the γ-ray energy, Ω is the solid angle of the measurement, and C is the peak area. The asterisk designates the standard. Finally, all these factors can be cancelled out by applying the double ratio. The double ratio can be expressed as two γ-ray peak area ratios.

In this method, it is only necessary to know the abundance ratio (K) of the comparative standard and amount of internal standard element (w_b) added to the sample, respectively. It is not necessary to weigh the sample and the comparative standard irradiated, to monitor the photon flux on sample and standard by the flux monitors, or to adjust strictly the geometrical condition of measurement, because this formula only needs the activity ratio. However, the latter statement is valid in limits only since the functions of counting efficiencies of the analytical signal and the internal monitor signal over the distance of the sample from the detector might differ. This difference can be significant particularly in the close vicinity of the detector.

It is essential that the irradiation energy for both the sample and the standard is the same. The difference of matrix effect between the sample and the comparative standard cannot be compensated by this method.

As to the comparison of the internal standard method with the single comparator method[38] as frequently utilized in NAA: in the single comparator method, the same amount of an element (a) to be determined and the comparator element (b) are irradiated and the activity ratio of element a to b is measured in advance. The activity ratio (C), which is regarded as constant, can be expressed as shown in Equation (12).

$$C = \frac{(\theta_a / AW_a)\sigma_a[1 - \exp(-\lambda_a t_a)]}{(\theta_b / AW_b)\sigma_b[1 - \exp(-\lambda_b t_b)]} \quad (12)$$

To determine an element in the sample, a sample and single comparator pair is irradiated simultaneously and the abundance of the element (w_a) can be calculated by the activity ratio (A_a/A_b), the amount of comparator (w_b) and constant as in Equation (13):

$$w_a = C w_b \left(\frac{A_a}{A_b} \right. \quad (13)$$

Although the quantitative equation of the internal standard method is analogous to that of the single comparator method, the principles of both methods are quite different from each other. The accuracy of the single comparator method is mostly unsatisfactory in PAA, because it is very difficult to keep the activity ratio constant. The energy spectrum for every irradiation and the beam current during one irradiation cannot be considered constant for every accelerator.

For highly pure materials, one of the main matrix components can be used conveniently as internal standard without adding an exotic element. This method was applied for the analysis of impurity or minor constituent in metals or alloys by Kato et al.[35,37] If a matrix element (atomic number Z) leads to isotopes of $(Z-1)$ or $(Z-2)$ by the $(\gamma,\mathrm{p}x\mathrm{n})$ or $(\gamma,\alpha x\mathrm{n})$ reactions, these radioisotopes can be used for internal standardization. Then, a $(\gamma,x\mathrm{n})$ product from neighboring elements (atomic number $Z-1$ or $Z-2$) can be chemically separated together with isotopes of $(Z-1)$ or $(Z-2)$ induced from the matrix. As these internal standards chemically behave like the $(\gamma,x\mathrm{n})$ products from impurity, their activity ratio is always constant. Therefore, the unknown amount can be determined without correction of chemical yield.

When the internal standard is added to the sample, the following conditions are required: none or a negligibly small amount of internal standard exists in the sample; only one or two γ-rays emit from the internal standard; and γ-rays from the internal standard do not interfere with γ-rays to be analyzed. Sc, Y and Sm have been used routinely as internal standards for a variety of samples by Segebade et al.,[39] and Ce was used for the trace element analysis of aluminum alloys by Yagi and Masumoto.[36]

To maintain maximum accuracy it is desirable that the half-life and γ-ray energy of the internal standard are similar to those of radioisotopes to be analyzed, and that the cross-section of the photonuclear reaction of the internal standard is similar to that of an element to be determined because the full energy spectrum of bremsstrahlung cannot be accurately monitored by using only one photonuclear reaction and also the photon flux tends to fluctuate in each irradiation.

6.2 Internal Standard Method Coupled with the Standard Addition Method

To achieve higher analytical accuracy without using any comparative standard and to compensate matrix effects arising from irradiation and measurement, the internal standard method coupled with the standard addition method has been proposed. This method was initiated by Lariaert et al.[40] in NAA as the internal standard method. After Lariaert et al. applied it to NAA of minor elements in high alloy steels,[41] Yagi and Masumoto[42] revived it by the full use of high-resolution γ-ray spectrometry and a new sample preparation technique. In order to distinguish it from the method dealt with in the previous section, this method is called the internal standard method coupled with the standard addition method. In this section, the definition and characteristics of this method are described in some detail.

In this method, duplicate samples are taken. The accurate amount (w_0) of element to be determined is added to one portion of sample. The sample and the standard addition sample are irradiated simultaneously and the activity ratios of the pertinent element and internal standard are measured for both sample (R) and standard addition sample (R^*).

The activity ratio of the standard addition sample can be expressed by Equation (14):

$$R^* = \frac{A_a^*}{A_b^*} = \frac{[(w_a + w_0)N_0\theta_a/AW_a]\sigma_a Q^*[1-\exp(-\lambda_a t_a)]}{(w_b N_0\theta_b/AW_b)\sigma_b Q^*[1-\exp(-\lambda_b t_b)]} \tag{14}$$

Then, by calculating the double ratio (R^*/R), the amount of internal standard element (w_b) is also cancelled out. As shown in Equation (15), a very simple relation is obtained.

$$\frac{R^*}{R} = \frac{w_a + w_0}{w_a} \tag{15}$$

The unknown amount of w_a is calculated using Equation (16):

$$w_a = \frac{w_0}{[(R^*/R) - 1]} \tag{16}$$

In this method, it is not necessary to know the amount of internal standard present in the sample (w_b). Only the amount of element added to the sample (w_0) is used for calculation. If an appropriate internal standard does not exist in the sample, the same amount of external element is added to the sample and the standard addition sample.

This method was applied to multielement analyses and one or more elements were used simultaneously as internal standards. For PAA of biological, geological and environmental samples, Na, Mg, Ca, Ti and Fe, usually inherent in natural samples, are useful internal standards because the concentrations of these elements are relatively high and half-lives and γ-rays of reaction products are convenient for the measurement. This method has been applied to PAA and charged particle activation analysis (CPAA) of biological materials,[42–44] sediment,[45,46] soil,[47,48] fly ash[49] and alloys.[50–52]

The use of two or more internal standards is effective to cross-check the analytical results.[53] The homogeneity of elements and internal standards in the sample and the standard addition sample is always checked by comparing with the standard deviations of activity ratios (R^* and R).

If $\sigma(R^*)$ is always larger than $\sigma(R)$ of a certain element, only this element was not mixed homogeneously at the standard addition step. If $\sigma(R)$ is always larger than $\sigma(R^*)$ of a certain element, this element is not homogeneous in the sample itself.

This method combines the characteristics of the internal standard method and that of the standard addition method. In the standard addition technique, comparative standards are not necessary. This method can eliminate several systematic errors, which are involved in the comparative method, such as the matrix composition of sample, the geometrical conditions of irradiation and measurement, and the analysis of the γ-ray spectrum.

In order to check whether loss or contamination of an element is caused during sample preparation, another portion of sample is prepared in the same procedure as the standard addition sample without adding an element to be determined and the double ratio (R^*/R) is also measured. When this ratio is <1, the pertinent element might be lost during the standard addition procedure. When this ratio is >1, contamination has to be considered. The analysis of CRMs by this method is useful to check the degree of interference reaction. If the content of one element is systematically higher than the certified value, the positive deviation indicates interference by other component(s) in the sample. As this method is capable of self-checking it is very useful for the analysis of reference materials.

6.3 The Stable Isotope Dilution Method

Stable isotope dilution activation analysis utilizes isotope analysis. For example, Ce has two stable isotopes, ^{140}Ce (88.48%) and ^{142}Ce (11.08%), which can produce ^{139}Ce and ^{141}Ce, respectively, by (γ,n) reactions. When an enriched isotope ^{142}Ce is added to the sample, the activity ratio of ^{139}Ce and ^{141}Ce changes. The degree of the activity ratio change is relative to the abundance of Ce in the sample and the added amount of ^{142}Ce. The conditions of this method are that the elements to be determined consist of at least two stable isotopes, and these two or more isotopes produce radioisotopes that can be detected separately by γ-ray spectrometry.

Stable isotope dilution activation analysis is performed using the following steps: the enriched isotope j is used as a spike, in which the isotopic abundances (θ_s) of i and j are known; the enriched isotope (w_s) is added to the sample; and the sample and the spiked sample are irradiated simultaneously. The activity ratio induced by two stable isotopes (i and j) is measured for the sample (R_x) and the spiked sample (R_m) using Equations (17) and (18).

$$R_x = \frac{A_{ix}}{A_{jx}} = \frac{(w_x\theta_{ix}/AW_x)\sigma_i[1-\exp(-\lambda_i t_i)]}{(w_x\theta_{jx}/AW_x)\sigma_j[1-\exp(-\lambda_j t_j)]} \quad (17)$$

$$R_m = \frac{A_{im}}{A_{jm}} = \frac{(w_x\theta_{ix}/AW_x + w_s\theta_{is}/AW_s)\sigma_i[1-\exp(-\lambda_i t_i)]}{(w_x\theta_{jx}/AW_x + w_s\theta_{js}/AW_s)\sigma_j[1-\exp(-\lambda_j t_j)]} \quad (18)$$

where the subscripts x, s and m indicate the sample, an enriched isotope and the spiked sample, respectively, w_x is the weight of element in the sample, w_s is the spiked amount of an enriched isotope, AW is the relative atomic mass, R is the activity ratio (isotope i/isotope j), and θ_j and θ_i are the isotopic abundances of isotopes i and j respectively.

Again the ratio of these activity ratios is used as shown in Equation (19):

$$\frac{R_m}{R_x} = \frac{(w_x\theta_{ix}/AW_x + w_s\theta_{is}/AW_s)(w_x\theta_{ix}/AW_x)}{(w_x\theta_{jx}/AW_x + w_s\theta_{js}/AW_s)(w_x\theta_{jx}/AW_x)} \quad (19)$$

Then, the equation is slightly simplified to Equation (20):

$$\frac{R_m}{R_x} = \frac{1 + (w_s/w_x)(\theta_{is}/\theta_{ix})(AW_x/AW_s)}{1 + (w_s/w_x)(\theta_{js}/\theta_{jx})(AW_x/AW_s)} \quad (20)$$

From Equation (20), the unknown amount (w_x) of element can be obtained using Equation (21):

$$w_x = w_s\left(\frac{AW_x}{AW_s}\right)\left(\frac{\theta_{js}}{\theta_{jx}}\right)\frac{(R_m/R_x) - (\theta_i/\theta_j)_s/(\theta_i/\theta_j)_x}{1-(R_m/R_x)} \quad (21)$$

This equation was introduced by Masumoto and Yagi[54] in 1983.

When highly enriched isotopes can be used as spikes, the abundance ratio $(\theta_i/\theta_j)_s/(\theta_i/\theta_j)_x$ becomes much smaller than the activity ratio (R_m/R_x). Then, a simple equation can be obtained, as shown by Equation (22):

$$w_x = w_s\left(\frac{AW_x}{AW_s}\right)\left(\frac{\theta_{js}}{\theta_{jx}}\right)\frac{1}{(R_x/R_m)-1} \quad (22)$$

As the unknown amount of an element can be determined by the activity ratio of two radioisotopes, the stable isotope dilution method is regarded as another type of internal standard method. In order to reduce error, the ratio (R_m/R_x) should preferably range from 0.3 to 0.7.[55]

In PAA, this technique was proposed for the determination of Ca, Zn and Ce.[54] Furthermore, it was applied to PAA[56] and CPAA[57] of Sr in biological materials. In the case of isotope dilution, it is not necessary to determine the chemical yield when the element to be determined is separated before or after irradiation. Hence, Sr was separated chemically before irradiation. Simultaneous determination of Ca, Rb, Sr and Ce was tried in PAA[58] in order to show the applicability of this technique in multielement analysis.

7 APPLICATION

7.1 Analysis of Light Elements

One of the important features of PAA is the analysis of light elements, such as C, N, O and F. In the book by Engelmann et al.,[59] PAA techniques for light element analysis have been described in detail and compared with chemical methods and CPAA. The radioisotopes produced by light elements are positron emitters. In order to analyze such elements, chemical separation after irradiation is mostly necessary. Fedoroff et al.[60] systematically studied the separation of radiocarbon in molten salts and expanded their studies to the simultaneous determination of C and N.[61]

In this section, some analytical examples of carbon and fluorine are described.

7.1.1 Determination of Carbon

Yoshioka et al.[62,63] determined the carbon impurity in copper and gallium arsenide. In PAA, trace amounts of carbon can be determined by detecting ^{11}C produced by $^{12}C(\gamma,n)^{11}C$. As ^{11}C is a positron emitter with a half-life of 20.39 min, a rapid chemical separation of radioactive carbon is necessary after irradiation (yet faster separation schemes have to be applied for oxygen analyses via ^{15}O; half-life = 2 min). Copper samples were oxidized with Sn as a flux material, and GaAs with Pb_3O_4. $^{11}CO_2$ produced in the combustion process was collected as $Ba^{11}CO_3$. The separation procedure was completed within one half-life of ^{11}C after irradiation. In the case of Cu and GaAs, the sensitivity is 0.02 $\mu g\,g^{-1}$ or 10^{15} atoms cm^{-3}, respectively.

Carbon impurity in a highly purified semiconductor has been usually analyzed with IR spectrophotometry. This method uses the relationship between the IR absorbance and the carbon concentration. The absorption of certain wavelengths of IR light is effected by the resonance with a vibration mode of a chemical bond between carbon and the substrate. A nuclear reaction, however, is not influenced by the chemical state of carbon. This fact suggests that the calibration factor between carbon concentration versus absorbance of IR, which had been used formerly, should be reconsidered.

7.1.2 Determination of Fluorine

The ion-selective electrode (ISE) is commonly used for determining fluorine in an aqueous solution. However, trace amounts of fluorine are difficult to detect by an ISE. In PAA, fluorine can be determined by the $^{19}F(\gamma,n)^{18}F$ reaction. In the case of high-level sodium contents in the sample, e.g. geological material, the irradiation energy should be reduced to 20 MeV because of the interfering nuclear reaction $^{23}Na(\gamma,\alpha n)^{18}F$. For the sensitive analysis of fluorine, pyrolysis and lanthanide precipitation of fluorine-18 after 20 MeV irradiation have been proposed.[64] The detectable concentration level was 0.06 $\mu g\,g^{-1}$ and rocks containing less than 10 $\mu g\,g^{-1}$ of fluorine were analyzed.

7.2 Multielement Analysis of Biological Materials

PAA of biological materials was reviewed by Kuttemperoor[65] and Sato.[66] By the development of high-resolution Ge detectors, PAA has been widely recognized as an efficient multielement determination technique of several biological materials, as reported by Kato et al.[67,68]

Biological samples, especially several organs of animals and blood, usually contain Na and P at high concentration levels. As the neutron activation cross-section of $^{23}Na(n,\gamma)^{24}Na$ is very large and the Compton scattering by γ-rays from ^{24}Na causes very high background in the γ-ray spectrum, ^{24}Na often masks other γ-rays emitted by minor and trace elements in NAA. Furthermore, since ^{32}P due to phosphorus produces high-energy β-ray emission followed by bremsstrahlung, detection of trace elements is frequently perturbed, especially in blood and bone samples.

In the case of PAA, high intensity radiation produced in biological samples is due to ^{11}C, ^{13}N and ^{15}O induced by the (γ,n) reaction from carbon, nitrogen and oxygen, respectively. Fortunately, γ-ray measurement can be carried out after a comparatively short cooling period, say about 3 h after irradiation, as the longest half-life of the aforementioned nuclides is 20 min (^{11}C). Even after 3 h, several short- to medium-lived radioisotopes such as ^{29}Al from Si, ^{38}K from K, ^{34m}Cl from Cl, ^{56}Mn from Fe and ^{87m}Sr from Sr can often be measured. Moreover, the carbon reaction can be suppressed by activating at an energy below its (γ,n) threshold (about 20 MeV). Many elements which can hardly be determined by NAA, such as Mg, Ca, Ni, Zn, Sr, Zr, Mo, I and Pb, are easily analyzed by PAA. In 1984, PAA of Zr in biological samples was performed without chemical separation.[69] The radioisotopes of medium half-life, such as ^{24}Na from Mg and ^{43}K and ^{47}Ca from Ca, sometimes show up as a source of undesired Compton background. To remove background activities, Yamashita and Suzuki[70] proposed a group separation scheme using complex formation of metals with tropolone-5-sulfonic acid and collecting on anion-exchange resin.

7.3 Analysis of Geological and Environmental Materials

PAA has been applied for the analysis of geological and environmental materials. About 20 to 30 elements can be determined instrumentally and simultaneously. The list of determinable elements covers major to trace

elements, essential to toxic elements, transition metals, heavy metals, alkali and alkaline-earth metals, halogens and others. These elements are Na, Mg, Si, Cl, K, Ca, Sc, Ti, Cr, Mn, Fe, Co, Ni, Zn, As, Rb, Sr, Y, Zr, Nb, Mo, Sb, I, Cs, Ba, Ce, Tl, Pb and U. The pioneering work of multielement analysis of rock and soil was done by Kato et al.[71,72] and Chattopadhyay and Jervis[73] respectively. PAA of atmospheric matter was reported by Aras et al.[74] and Kato et al.[75]

Geological and environmental samples contain various elements with high neutron capture cross-section, such as Na, Sc, Mn, Co, Br and Eu. In NAA, these elements induce high intense radioactivity even at trace levels. In PAA, radioactivities induced by major elements such as Na, Mg, Ca, Ti, Mn and Fe are relatively low in most cases. As the (γ,p) reactions are the most efficient reactions of Mg, Ca, Ti and Fe, the radioactivities induced by these elements can be reduced by lowering the irradiation energy well below 30 MeV.

Several selected elements advantageously detectable by PAA were also determined. Determination of Ni by a coincidence technique,[76] determination of Nb and Y for the characterization of standard rocks,[77] and determination of I, Tl and U in environmental materials[78] are typical examples.

Deep-sea sediment[79,80] and volcanic rocks[81] were analysed by PAA to discuss the sedimentation process and the distribution of minor elements during the formation of rock, respectively.

7.4 Multielement Analysis of Raw Materials and Industrial Products

Gijbels and Hertogen[82] reviwed PAA of ore and minerals. Albert[83] discussed trace analysis using PAA in a review article of nuclear methods. As mentioned before, NAA of certain matrices is sometimes very difficult, because of the self-shielding problem and high radioactivity caused. Typical examples are rare-earth elements, cadmium and boron, which also have large neutron capture cross-sections, and uranium ore and phosphate, which lead to fission products. To verify the capability of PAA in these fields, lanthanides in boron carbide were determined instrumentally by PAA.[84] Detection limits of these elements were found in the submicrogram range. Twenty-eight impurities in boron were also determined by PAA.[85]

8 COMPARISON WITH OTHER ACTIVATION ANALYSIS METHODS

In this section, the different features, advantages and disadvantages of the three activation analysis techniques PAA, CPAA and NAA are compared regarding the irradiation technique, nuclear reactions induced as well as sensitivity and selectivity achievable. Only the basic methods are considered; special techniques like nuclide-induced photodisintegration or 14 MeV NAA are excluded from discussion.

8.1 Irradiation

8.1.1 Photon Activation Analysis

The energy distribution of bremsstrahlung is continuous from zero to the incident electron energy. As high-energy photons have a high penetration power for most materials, a stack of several samples can be irradiated simultaneously. The matrix effect normally is very small in PAA. The energy distribution is not changed drastically during penetration. As the photon flux is decreased exponentially by scattering and absorption in the matter, the photon flux density on each sample should be monitored. Although the incident electron beam is focused, bremsstrahlung cannot be focused. As already touched on, bremsstrahlung has a nonisotropic spatial distribution which depends on the converter thickness and the geometry of the electron beam.

8.1.2 Charged Particle Activation Analysis

As the charged particles are strongly retarded in a sample by scattering, particles are stopped at a certain range. Both the induced radioactivity and the predominant reaction type induced change along with the depth since both flux and energy of the respective particle change. Therefore, range or stopping power of charged particles in each material should be taken into account for the quantitative analysis. It is difficult to irradiate a stack of samples. However, CPAA can efficiently be applied to analyze surface layers of materials. Hence, the surface contamination problem is more imminent than in NAA or PAA. Local heating of the materials by charged particle is a severe problem for the analysis of organic and other sensitive materials. Charged particles can be focused even on extremely small areas. Thus CPAA can be used favorably for the analysis of small spots, e.g. in the study of areal elemental distributions.

8.1.3 Neutron Activation Analysis

Since the neutron has no electrical charge it can penetrate the bulk of samples more easily. The absorption cross-sections of neutrons are strongly dependent on the nuclear properties of each element. The energy distribution of reactor neutrons shows Maxwell–Boltzmann distribution according to the temperature in the reactor. In most cases, many samples can be irradiated simultaneously

and homogeneously in a capsule. The self-shielding effect of certain elements, such as Cd, should be considered because the neutron flux is locally depressed in the presence of these elements. Neutron beams cannot be focused except, to a certain limited extent, with a neutron beam tube installed at the reactor.

8.2 Comparison of Nuclear Reactions Induced

8.2.1 Photon Activation Analysis

As mentioned above, photonuclear reactions are mainly controlled by the photoabsorption process (giant resonance, see above) giving rise to the (γ,n)-type reaction. The resonance energy ranges from about 10 MeV for heavy elements to about 20 MeV for light elements. The yield of the photoneutron reaction increases with atomic number. For analysis, (γ,n) and (γ,p) reactions are utilized normally. Higher-order reactions such as (γ,αn), (γ,np) or (γ,T) mostly appear as sources of interference. In practice, 30 MeV bremsstrahlung has proven to be optimal in most applications.

8.2.2 Charged Particle Activation Analysis

In CPAA, protons, deuterons, α and ^{3}He particles are commonly used as probes. In the case of charged particle activation, the Coulomb barrier effect should be considered. For heavier elements, higher excitation energies are needed to penetrate the coulomb barrier and to induce nuclear reactions. As several reaction types can be induced at this excitation energy many different radioisotopes are observed and the probability of nuclear interference is considerable. Since the higher atomic number elements are more likely to be subject to nuclear interference than light ones, CPAA is not normally suitable for the analysis of heavy elements.

In the case of deuteron and ^{3}He activation, many nuclear reactions are exothermic and can be induced at low excitation energy. Such nuclear reactions have been applied to the analysis of light elements. The irradiation particle and irradiation energy can be selected quasi-freely for the selective analysis of certain elements.

8.2.3 Neutron Activation Analysis

In NAA, thermal and epithermal neutrons are used. Neutrons are absorbed by elements according to the $1/v$ law and by resonance reactions. The thermal neutron capture cross-section is inversely proportional to the velocity of the neutron, and this is known as "the $1/v$ law". The resonance reactions induced in the epithermal region are often used for the analysis of specific elements using cadmium shielding of the sample because cadmium has very high absorption cross-section for thermal neutrons. The analysis of light element is very difficult. Unlike in PAA, there is no regularity in the neutron absorption reactions. Nonetheless, NAA can be used for multielement analysis of a large variety of materials.

8.3 Sensitivity and Selectivity of Elements to be Determined and Applicability to Various Kinds of Materials

8.3.1 Photon Activation Analysis

PAA is used for the analysis of light elements such as C, N, O and F, medium elements such as Ti, Ni, Sr, Zr, Nb, and Y, and heavy elements such as Pb, which are very difficult to analyze by NAA. Furthermore, PAA is suitable for the analysis of matrices which contain some elements of high neutron capture cross-section, e.g. boron and noble metals. As photonuclear reaction cross-sections systematically increase with atomic number, PAA can be applied to the analysis of heavier elements in light matrices, such as natural materials (rocks, soils, dusts, organics, etc.). The sensitivity for trace analyses is mostly sufficient (microgram per gram region and below), although not as good as in NAA for the average case. Using appropriate equipment and with the exception of a few interference cases (see above), PAA is highly selective due to discrete energy signals of the respective activation products.

8.3.2 Charged Particle Activation Analysis

CPAA is advantageous particularly for light element analysis. CPAA has often been applied to the impurity analysis of industrial materials. The shape of the sample is critical, so the application is somewhat restricted. The average sensitivity is comparable to that of PAA whilst the selectivity is frequently limited due to numerous nuclear reactions eventually induced in the individual element.

8.3.3 Neutron Activation Analysis

NAA can detect many elements instrumentally, except for light elements. Extremely low detection limits are achievable for a good number of elements, e.g. several rare earth elements, noble metals and others; the respective detection limits are lower partly by orders of magnitude, as compared with PAA and CPAA. Thus, NAA is suitable for sensitive analysis of trace elements in various types of materials. However, interference through eventual excessive matrix activities have to be taken into account. Moreover, in the normal case, only small sample masses should be irradiated because of the intense radioactivity induced. Thus, problems of representativeness might arise. As to the selectivity, see section 8.3.1.

9 CONCLUSION

PAA offers sufficient sensitivity for the analysis of a great number of elements covering the entire periodic table. A large variety of sample types can be analyzed. The quality of PAA results can be improved by application of quantitative methods, such as the internal standard method or the stable isotope dilution method. In the case of the classical comparative method, the use of synthetic multielement standards of acrylamide or silica matrix yields significant improvement of accuracy.

In PAA (also in NAA), samples can be measured repeatedly. A laboratory automation system using a small robot for sample changing[86,87] has released analysts from the tedious work of sample handling and data acquisition during γ-ray measurement. Moreover, many ''human'' sources of error can be avoided. Such systems also assist researchers of many other fields using activation analysis and other radioactivity measurements.

Although other, nonactive analytical methods underwent dramatic improvement in recent years concerning sensitivity and other features, PAA and other activation techniques can offer specific advantages, e.g. the analysis of elements difficult to analyse by other methods. Moreover, purely instrumental and even nondestructive analyses are state-of-the-art in PAA.

Thus, as trace characterization is still important in material science, environmental science and other fields, PAA can contribute to the analysis of light to heavy elements.

ABBREVIATIONS AND ACRONYMS

CPAA	Charged Particle Activation Analysis
CRM	Certified Reference Material
EC	Electron Capture
EXAFS	Extended X-ray Absorption Fine Structure
IR	Infrared
ISE	Ion-selective Electrode
LEPD	Low-energy Photon Detector
NAA	Neutron Activation Analysis
PAA	Photon Activation Analysis
PM	Photomultiplier
SRM	Standard Reference Material
XPS	X-ray Photoelectron Spectroscopy
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis

Nuclear Methods **(Volume 14)**
Charged Particle Activation Analysis • Elemental Analysis by Isotope Dilution • Instrumental Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
γ-Spectrometry, High-resolution, for Radionuclide Determination • Nuclear Detection Methods and Instrumentation

REFERENCES

1. Ch. Segebade, H.P. Weise, 'Comparison of Sensitivity Estimates for Low Energy Photon and Classical Gamma-ray Spectroscopy Applied to Photon Activation Analysis', *J. Radioanal. Chem.*, **45**, 209–220 (1978).
2. T. Kato, T. Sato, K. Mori, 'Analytical Sensitivities for Thirty-six Elements in Cellulose Matrix by Means of High-energy Photon Activation and Low-energy Photon Detection', *Anal. Sci.*, **6**, 103–106 (1990).
3. T. Sato, K. Mori, T. Kato, 'Self-absorption Correction for Low-energy Photons in Instrumental Photon Activation Analysis of Biological Materials', *Anal. Sci.*, **82**, 217–221 (1991).
4. J. Chadwick, M. Goldhaber, 'A Nuclear Photo-effect: Disintegration of the Diplon by γ-rays', *Nature*, **134**, 237 (1934).
5. R. Basile, J. Hure, P. Leveque, C. Schuhl, *C. R. Acad. Sci.*, **239**, 422 (1954).
6. C. Engelmann, 'Photon Activation Analysis', in *Advances in Activation Analysis*, eds. J.M.A. Lenihan, V.P. Guinn, Academic Press, New York, 1–55, 1972.
7. C. Segebade, H.-P. Weise, G.J. Lutz, *Photon Activation Analysis*, Walter de Gruyter, Berlin, 1988.
8. A.P. Kushelevsky, 'Photon Activation Analysis', in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, 219–237, 1990.
9. M.E. Toms, 'A Compilation of Photonuclear Reaction Products and Associated Gamma-energies', *J. Radioanal. Chem.*, **20**, 177–381 (1974).
10. P. Anderson, J.S. Hislop, D.R. Williams, 'High Resolution γ-spectra of 40–44 MeV γ Photon Activation Products. Part 1. The Elements Sodium to Molybdenum', AERE – R 7823, HMSO, London, 1974.
11. D.R. Williams, 'High Resolution γ-spectra of 40–44 MeV γ Photon Activation Products. Part 2. The Elements Ruthenium to Uranium', AERE – R 9021, HMSO, London, 1978.
12. G.J. Lutz, Ch. Segebade, 'Two-dimensional (Half-life vs. Energy) Map of Gamma-rays Emitted by Radionuclides Produced Through Photonuclear Reactions', *J. Radioanal. Chem.*, **33**, 303–337 (1976).
13. T. Kato, 'Quantitative Evaluation of the Photon-activation Method in Analysis', *J. Radioanal. Chem.*, **16**, 307–327 (1973).

14. H. Bethe, W. Heitler, 'On the Stopping of Fast Particle and on the Creation of Positive Electrons', *Proc. R. Soc. London*, **A134**, 83–112 (1934).
15. G.J. Lutz, 'Calculation Sensitivities in Photon Activation Analysis', *Anal. Chem.*, **41**, 424–427 (1969).
16. L.I. Schiff, 'Energy-angle Distribution of Thin Target Bremsstrahlung', *Phys. Rev.*, **83**, 252–253 (1951).
17. Y. Oka, T. Kato, K. Nomura, T. Saito, 'The Yields of Radioactivities Induced by (γ,n) Reactions with 20 MeV Bremsstrahlung', *Bull. Chem. Soc. Jpn.*, **40**, 575–579 (1967).
18. Y. Oka, T. Kato, K. Nomura, T. Saito, H.-T. Tsai, 'A Study on the Yields of (γ,p) Reactions with 20 MeV Bremsstrahlung', *Bull. Chem. Soc. Jpn.*, **41**, 380–384 (1968).
19. Y. Oka, T. Kato, K. Nomura, T. Saito, H.-T. Tsai, 'A Study on the Yields of (γ,α) Reactions with 20 MeV Bremsstrahlung', *Bull. Chem. Soc. Jpn.*, **41**, 2660–2662 (1968).
20. T. Kato, K. Masumoto, N. Sato, N. Suzuki, 'The Yields of Photonuclear Reactions for Multielement Photon Activation Analysis', *J. Radioanal. Chem.*, **32**, 51–70 (1976).
21. K. Masumoto, T. Kato, N. Suzuki, 'Activation Yield Curves of Photonuclear Reactions for Multielement Photon Activation Analysis', *Nucl. Instrum. Methods*, **157**, 567–577 (1978).
22. M. Barbier, 'Radioactivity Induced by High Energy Electrons and Photons', in *Induced Radioactivity*, North-Holland, Amsterdam, 214–249, 1969.
23. D.R. Williams, J.S. Hislop, 'High Resolution γ-spectra of 40–44 MeV γ Photon Activation Products. Part 3. Data Section', AERE – R 9022, HMSO, London, 1978.
24. E. Ricci, 'Sensitivities for Photon Activation Analysis with Thick-target, 110-MeV Electron Bremsstrahlung', *Anal. Chem.*, **46**, 615–618 (1974).
25. M.A. Kolmi'tsev, T.S. Ambardanashvili, V.Yu. Dundua, 'Comparison Standards for Neutron Activation Analysis on a Base of Phenol–Formaldehyde Resin', *J. Radioanal. Chem.*, **20**, 549–557 (1974).
26. D.H. Anderson, J.J. Murphy, W.W. White, 'Gelatin Multicomponent Trace Element Reference Material', *Anal. Chem.*, **48**, 116–117 (1976).
27. K. Masumoto, N. Suzuki, 'A New Trace Element Standard Reference Material with Polyacrylamide Gel Matrix for Radioactivation Analysis', *Radiochem. Radioanal. Lett.*, **42**, 99–106 (1980).
28. J.W. Mitchell, L.D. Blitzer, T.Y. Kometani, T. Gills, L. Clark, Jr, 'Homogeneously Doped Silica Matrices for Trace Element Standards in Neutron Activation Analysis', *J. Radioanal. Chem.*, **39**, 335–342 (1977).
29. A.R. Date, 'Preparation of Trace Element Reference Materials by a Co-precipitated Gel Technique', *Analyst*, **103**, 84–92 (1978).
30. M. Yagi, K. Kondo, 'A New Irradiation System for Photon Activation Analysis by Electron Accelerator', *J. Radioanal. Chem.*, **36**, 247–253 (1977).
31. G.A. Price, 'Energy Spectra and Angular Distributions of Photoneutrons from Heavy Nuclei', *Phys. Rev.*, **93**, 1279–1285 (1954).
32. K. Masumoto, K. Sakamoto, Y. Miyamoto, J.H. Zaid, A. Kajikawa, H. Haba, 'The Application of KURRI-Linac to Photon Activation Analysis of Environmental Materials', Abstract from the 40th Symposiumon Radiochemistry, Riken, Japan, 192–193, 1996.
33. Ch. Segebade, G.J. Lutz, H.P. Weise, 'Quantitative Evaluation of Interference Reactions in Photon Activation Analysis. The Production of Identical Radionuclides', *J. Radioanal. Chem.*, **39**, 179–184 (1977).
34. Y. Miyamoto, H. Haba, A. Kajikawa, K. Masumoto, T. Nakanishi, K. Sakamoto, 'Interferences in Neutron and Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **239**, 165–175 (1999).
35. H.-T. Tsai, K. Kato, Y. Oka, 'Photon Activation Analysis for Iodine by Using Cesium as the Internal-reference Element', *Bull. Chem. Soc. Jpn.*, **43**, 2482–2486 (1970).
36. M. Yagi, K. Masumoto, 'A New Internal Reference Method for Activation Analysis and its Application. Determination of Ti, Cr, Fe, Cu, Ga and Zr in Aluminum Alloys by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **84**, 369–380 (1984).
37. Y. Oka, T. Kato, H.-T. Tsai, K. Nomura, 'An Internal-monitor Method for Activation Analysis of Silver in Palladium with Thermal Neutrons', *Bull. Chem. Soc. Jpn.*, **41**, 329–333 (1968).
38. F. Girrardi, G. Guzzi, J. Pauly, *Anal. Chem.*, **37**, 1085–1092 (1965).
39. Ch. Segebade, M. Kuhl, B.F. Schmitt, R. Neider, 'Some Remarks on the State of Photon Activation Analysis and the Use of Internal Standards', *J. Radioanal. Chem.*, **72**, 665–696 (1982).
40. G. Laliaert, J. Hoste, Z. Eeckhaut, 'Neutron Activation Analysis Using an Internal Standard', *Nature*, **182**, 600 (1958).
41. G. Laliaert, J. Hoste, Z. Eeckhaut, 'Activation Analysis of Vanadium in High Alloy Steels Using Manganese as Internal Standard', *Anal. Chim. Acta*, **19**, 100–107 (1958).
42. M. Yagi, K. Masumoto, 'A New Internal Standard Method for Activation Analysis and its Application. Determination of Co, Ni, Rb and Sr in Pepperbush by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **84**, 316–331 (1984).
43. K. Masumoto, M. Yagi, 'Simultaneous Determination of P, Cl, K and Ca in Several Control Serums by Alpha-particle Activation Analysis Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **109**, 449–458 (1987).
44. M. Yagi, K. Masumoto, 'Instrumental Charged-particle Activation Analysis of Several Selected Elements in Biological Materials Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **111**, 359–369 (1987).

45. K. Masumoto, M. Yagi, 'Highly Accurate and Precise Multielement Determination of Environmental Samples by Means of Photon Activation Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **100**, 287–301 (1986).
46. M. Yagi, K. Masumoto, 'Instrumental Photon Activation Analysis of Environmental Materials Using the Internal Standard Method', *J. Radioanal. Nucl. Chem.*, **109**, 337–346 (1987).
47. K. Masumoto, M. Yagi, 'Instrumental Photon Activation Analysis of Soil Samples Using the Internal Standard Method Coupled with the Standard Addition Method', *J. Radioanal. Nucl. Chem.*, **116**, 375–388 (1987).
48. K. Masumoto, M. Yagi, 'Revaluation of the Internal Standard Method Coupled with the Standard Addition Method Applied to Soil Samples by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **121**, 131–139 (1988).
49. K. Masumoto, M. Yagi, 'Instrumental Photon Activation Analysis of Coal Fly Ashes Using the Internal Standard Method Coupled with the Standard Addition Method', *J. Radioanal. Nucl. Chem.*, **152**, 55–66 (1991).
50. M. Yagi, K. Masumoto, 'Simultaneous Determination of Ti, Cr, Fe, Cu, Ga and Zr in Aluminum Alloys by Charged Particle Activation Analysis Using Internal Standard', *J. Radioanal. Nucl. Chem.*, **91**, 379–387 (1985).
51. K. Masumoto, M. Yagi, 'Determination of Phosphorus in Low-alloy Steels by Alpha-activation Analysis', *J. Radioanal. Nucl. Chem.*, **130**, 243–250 (1989).
52. K. Masumoto, M. Hara, D. Hasegawa, E. Iino, M. Yagi, 'Photon and Proton Activation Analysis of Iron and Steel Standards Using the Internal Standard Method Coupled with the Standard Addition Method', *J. Radioanal. Nucl. Chem.*, **217**, 247–253 (1997).
53. K. Masumoto, 'Application of the Internal Standard Method Coupled with the Standard Addition Method in Photon Activation Analysis to the Trace Characterization of Environmental Materials', *J. Radioanal. Nucl. Chem.*, **203**, 37–46 (1996).
54. K. Masumoto, M. Yagi, 'Stable-isotope Dilution Activation Analysis and Determination of Ca, Zn and Ce by Means of Photon Activation', *J. Radioanal. Chem.*, **79**, 57–68 (1983).
55. M. Yagi, 'Stable Isotope Dilution Activation Analysis', in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, 131–143, Vol. 1, 1990.
56. M. Yagi, K. Masumoto, 'Isotope Dilution Activation Analysis for Special Samples in which the Self-shielding Effect is Negligible. Determination of Sr by Means of Photon Activation', *J. Radioanal. Nucl. Chem.*, **90**, 91–103 (1985).
57. K. Masumoto, M. Yagi, 'Determination of Sr in Biological Materials by Charged Particle Activation Analysis Using Stable-isotope Dilution Method', *J. Radioanal. Nucl. Chem.*, **91**, 369–378 (1985).
58. K. Masumoto, M. Yagi, 'Simultaneous Determination of Ca, Rb, Sr and Ce in Environmental Materials by Photon Activation Analysis Using the Stable Isotope Dilution Method', *J. Radioanal. Nucl. Chem.*, **99**, 287–297 (1986).
59. C. Engelmann, G. Kraft, J. Pauwels, C. Vandecasteele, *Modern Methods for the Determination of Non-metals in Non-ferrous Metals*, Walter de Gruyter, Berlin, 1985.
60. M. Fedoroff, C. Loos-Neskovic, G. Revel, 'Radiochemical Separations in Molten Salts. Determination of Carbon in Metals by Photon Activation', *J. Radioanal. Chem.*, **55**, 219–232 (1980).
61. M. Fedoroff, C. Loos-Neskovic, J.C. Rouchaud, V.N. Samosyuk, B.A. Chapyzhnikov, 'Simultaneous Determination of Nitrogen and Carbon in Silicon by Photon Activation', *J. Radioanal. Chem.*, **88**, 45–49 (1985).
62. A. Yoshioka, K. Nomura, M. Takeya, K. Shimura, K. Masumoto, M. Yagi, 'Determination of Carbon in Copper by Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **122**, 175–182 (1988).
63. A. Yoshioka, K. Nomura, O. Kawakami, K. Shimura, K. Masumoto, M. Yagi, 'Determination of Carbon Impurity in Gallium Arsenide Crystals by Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **148**, 201–209 (1991).
64. H. Sakurai, M. Fukaya, Y. Hayashibe, Y. Sayama, K. Masumoto, T. Ohtsuki, 'Determination of Fluorine in Standard Rocks by Photon Activation Analysis', *J. Radioanal. Nucl. Chem.*, **217**, 267–271 (1997).
65. V.Z. Kuttemperoor, 'Photon Activation Analysis of Biological Materials', in *CRC Handbook Series in Clinical Laboratory Science*, ed. D. Seligson, CRC Press, Boca Raton, 209–284, 1977.
66. T. Sato, 'Activation Analysis of Biological Materials', in *Activation Analysis*, ed. Z.B. Alfassi, CRC Press, Boca Raton, 344–345, 1990.
67. T. Kato, N. Sato, N. Suzuki, 'Multielement Photon Activation Analysis of Biological Materials', *Anal. Chim. Acta*, **81**, 337–347 (1976).
68. N. Sato, T. Kato, N. Suzuki, 'Multielement Determination in Tobacco Leaves by Photon Activation Analysis', *J. Radioanal. Chem.*, **36**, 221–238 (1977).
69. T. Sato, 'The Determination of Zirconium in Biological Materials by Photon Activation Analysis', *J. Radioanal. Nucl. Chem. Lett.*, **86**, 141–150 (1984).
70. M. Yamashita, N. Suzuki, 'Photon Activation Analysis of Trace Metals in Biological Materials Via Collection of Metal Tropolone-5-sulfonate Complexes onto Anion-exchange Resin', *J. Radioanal. Chem.*, **60**, 73–85 (1980).
71. T. Kato, I. Morita, N. Sato, 'Non-destructive Photon-activation Analysis of Standard Rocks with 30 MeV Bremsstrahlung', *J. Radioanal. Chem.*, **18**, 97–110 (1973).

72. N. Sato, T. Kato, N. Suzuki, 'Multielement Photon-activation Analysis of Rocks Materials with 30 MeV Bremsstrahlung', *Radiochim. Acta*, **21**, 63–68 (1974).
73. A. Chattpadhyay, R.E. Jervis, 'Multielement Determination in Market-garden Soils by Instrumental Photon Activation Analysis', *Anal. Chem.*, **46**, 1630–1639 (1974).
74. N.K. Aras, W.H. Zoller, G.E. Gordon, G.J. Lutz, 'Instrumental Photon Activation Analysis of Atmospheric Particulate Material', *Anal. Chem.*, **45**, 1481 (1974).
75. T. Kato, N. Sato, N. Suzuki, 'Non-destructive Multielement Photon-activation Analysis of Environmental Materials', *Talanta*, **23**, 517–524 (1976).
76. K. Masumoto, N. Suzuki, 'Selective Coincidence Spectrometry in the Nondestructive Determination of Nickel in Geological Materials with High-energy Photon Activation', *J. Radioanal. Chem.*, **46**, 121–135 (1978).
77. T. Kato, K. Masumoto, 'Determination of Niobium and Yttrium in GSJ Basalt by Photon Activation Analysis', *Geostand Newslett.*, **5**, 167–170 (1981).
78. K. Masumoto, T. Ohtsuki, Y. Miyamoto, J.H. Zaidi, A. Kajikawa, H. Haba, K. Sakamoto, 'Photon Activation Analysis of Iodine, Thallium and Uranium in Environmental Materials', *J. Radioanal. Nucl. Chem.*, **239**, 495–500 (1999).
79. T. Kato, N. Sato, N. Suzuki, 'Multielement Analysis of Deep-sea Sediments by Photon Activation', *Bull. Chem. Soc. Jpn.*, **50**, 1930–1933 (1977).
80. T. Kato, M. Kato, 'Multielement Determination in Floor Sediments of the Japan Sea by Nondestructive Photon Activation Analysis', *J. Radioanal. Chem.*, **57**, 105–112 (1980).
81. T. Yoshida, K. Masumoto, K.-I. Aoki, 'Photon-activation Analysis of Standard Rocks Using an Automatic γ-ray Counting System with a Micro-robot', *J. Jpn. Assoc. Min. Pet. Econ. Geol.*, **81**, 406–412 (1986).
82. R.H. Gijbels, J. Hertogen, 'Activation Analysis of Ores and Minerals', *Pure Appl. Chem.*, **49**, 1555–1573 (1977).
83. Ph. Albert, 'Nuclear Methods in Trace and Ultra-trace Analysis: Past Experiences and Future Possibilities', *Pure Appl. Chem.*, **54**, 689–713 (1982).
84. Ch. Berthelot, G. Carraro, V. Verdingh, 'The Use of Photon Activation Analysis for the Determination of Sm, Eu, Gd and Dy in Boron Carbide', *J. Radioanal. Chem.*, **50**, 185–194 (1979).
85. T. Kato, K. Nagase, H. Yokobayashi, 'Determination of Trace Elements in Boron by Instrumental Photon Activation Analysis', *Anal. Sci.*, **1**, 307–311 (1985).
86. M. Yagi, K. Masumoto, M. Muto, 'An Automatic Gamma-ray Spectrometer Equipped with a Micro-robot for Sample Changing', *J. Radioanal. Nucl. Chem.*, **98**, 31–38 (1986).
87. K. Masumoto, S.-I. Ushino, M. Mutoh, M. Yagi, T. Yoshida, 'An Automatic Measurement System for Radiochemical Study Using Small Robots and PC-LAN', *Lab. Robotics Automat.*, **6**, 21–27 (1994).

PIXE (Particle-induced X-ray Emission)

P.A. Mandò
Dipartimento di Fisica dell'Università and INFN, Firenze, Italy

1 Introduction **1**
1.1 General Features 1
1.2 Brief Historical Overview and Review Literature 3
1.3 Comparison with other X-ray-based Analytical Techniques 3

2 In-depth Review of the Technique **5**
2.1 Basic Aspects of X-ray Emission from Atomic Shells 5
2.2 Quantitative PIXE Analysis 6
2.3 Minimum Detection Limits in PIXE Analysis 11
2.4 Ion Accelerators and Sample Irradiation 13
2.5 External Beam Set-ups 15
2.6 Detection of X-rays, Data Acquisition and Reduction 17
2.7 Micro-PIXE 21

3 Survey of the Main Analytical Applications **22**
3.1 Environmental Pollution Studies 22
3.2 Biomedical Applications 24
3.3 Applications to Earth Science 26
3.4 Applications to Art and Archaeology 27
3.5 Applications to Historical Problems 29

Abbreviations and Acronyms **30**

Related Articles **30**

References **30**

Particle-induced X-ray emission (PIXE) is the most popular among the ion beam analysis (IBA) techniques, which are based on the use of the specimen to be analyzed as a target for a beam of accelerated particles. The detection of the radiation induced by the beam bombardment is then used to discriminate and quantify the presence of the different elements in the specimen. In PIXE, what is exploited is in particular the X-rays emitted from the target, whose energies are characteristic of the emitting atomic species. After a general, simple description of the main features of PIXE, with a short historical overview and a brief comparison to other X-ray fluorescence (XRF) techniques, the article covers in greater depth all the specific aspects of this ''nuclear'' technique. The basic aspects of X-ray emission from the atoms are first recalled; then, the extraction of quantitative compositional data from thin and thick specimens is explained, and a discussion is given of the excellent performance of PIXE in terms of minimum detection limits (MDLs), and of the factors affecting them. A technical description then follows of how proper beams for PIXE are produced and of the experimental set-ups commonly used, with particular emphasis on the external-beam arrangements. The X-ray detector characteristics, the electronics for constructing the energy spectra, and the software processes for their deconvolution, leading to the extraction of quantitative data, are then briefly described. The last section surveys the main analytical applications of PIXE in various fields (environmental monitoring, biomedicine, earth sciences, cultural heritage), with no intent of exhaustiveness but rather with the purpose of focusing on when and why PIXE may be particularly suitable in each of them.

1 INTRODUCTION

1.1 General Features

PIXE is a technique for the elemental analysis of a sample. This is used as a target for the bombardment with a beam of accelerated particles; the interactions of the beam particles with the target atoms lead to the emission of X-rays of characteristic energies, through the detection of which the target composition can be deduced. The PIXE technique is often also referred to as *proton*-induced X-ray emission, owing to the fact that in this technique protons are almost universally chosen to induce X-ray emission. For this reason, the following discussion always considers the case of protons; no difference of either theoretical or practical relevance occurs anyway with beams of other particles. The PIXE technique belongs to the category of IBA techniques, and is perhaps the most widely used among them.

The method makes it possible to detect simultaneously and with great sensitivity almost the full range of elements present in the analyzed sample. In practice, some limitations partially restrict the number of detectable elements, but the multi-element capability is actually retained to a very large extent. As a rule of thumb (to be better specified below), one might state that in a single measurement – typically lasting few minutes – all elements with $Z > 10$ are detected with MDLs down to trace levels.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The working principle of PIXE is quite simple. Using a small (usually electrostatic) accelerator, a beam of particles is produced with an energy of typically 2–3 MeV. The beam particles, when impinging on a target, interact with atoms and nuclei of the target itself: in particular, the inner shells of the target atoms may be ionized. The inner shell vacancy is promptly filled (within a time interval of the order of 10^{-15} s) by an outer electron, and the difference $\Delta E = E_o - E_i$ of the electron binding energies E_o and E_i in the outer and inner shell is released. One of the processes through which the energy ΔE can be released is just the emission of an X-ray of that energy. A competing mechanism (the Auger effect) is the energy transfer to another electron of an outer shell, which is consequently ejected with a kinetic energy $E_k = \Delta E - E_j$, where E_j is the binding energy of the electron to which ΔE is transferred. Figure 1 shows schematically the processes of X-ray and Auger electron emission.

The energy of the emitted X-ray (or Auger electron) is clearly characteristic of the emitting atomic species, being the difference of electron binding energies that are indeed characteristic of the atomic species. Thus, the energy spectroscopy of the X-rays from the sample during particle beam bombardment provides an analysis of its composition. Such an analysis is *multi-elemental*, because all the atomic species present in the target undergo interactions with the beam particles, and therefore emit characteristic X-rays. It is also *quantitative*, because during an irradiation the number of interactions of the beam particles with the atoms of a given species is obviously proportional to the abundance of atoms of that species in the target. Under most circumstances (better specified in the following) the analysis is also totally *nondestructive*, because the beam particle–target atom interactions do not produce significant alterations in the target and involve only a tiny fraction of the target atoms. In any case, the possibility of further analysis (by the same or different techniques) is guaranteed for the same specimen.

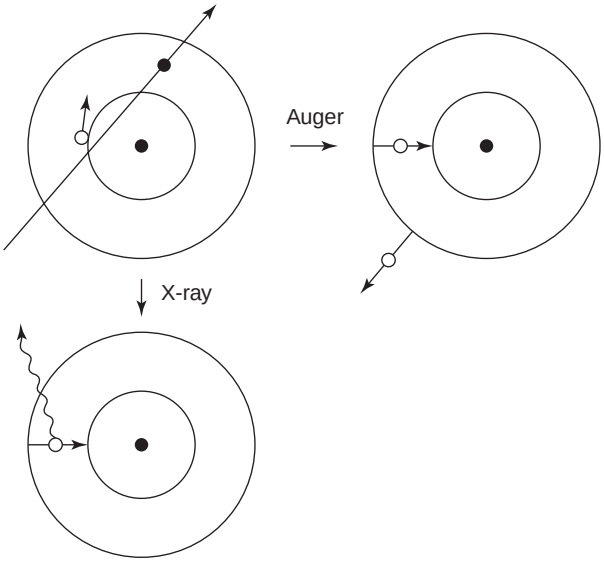

Figure 1 Diagram of an inner shell ionization induced by proton bombardment, followed by Auger electron or X-ray emission (alternative to each other). (Reproduced by permission from Johansson et al.[1])

The principle of operation of PIXE is the same as that exploited to perform elemental analysis through electron probe microanalysis (EPMA) or XRF, but with a different ionizing agent (protons rather than electrons or primary electromagnetic radiation, respectively). This difference has consequences in terms of important analytical parameters such as MDLs, range of detected elements, or sample depth probed (as briefly discussed later).

The possibility of performing PIXE while keeping the sample in the atmosphere (*nonvacuum* or *external* PIXE), and that of miniaturizing the beam size down to 1 μm or even less (micro-PIXE), make it a very versatile technique. All kinds of samples may be quantitatively analyzed: solid, gaseous, or liquid (the latter two with an external beam). Elemental mapping may also be performed by scanning the beam on the sample during analysis (or moving the sample under the beam).

In PIXE, the detection and discrimination of X-rays emitted under beam bombardment is usually performed through the use of solid state Si(Li) detectors (or, much less frequently, hyperpure Ge detectors), which discriminate the different X-rays by their different energies. The detector response signals are proportional to the energy releases within the active volume of the detector; this way of discriminating different X-rays is therefore called energy-dispersive spectroscopy (EDS). The energy resolution is generally adequate for unambiguous discrimination. Conventionally, resolution is given as the full width at half-maximum (fwhm) of the peak corresponding to the 5.9-keV X-ray from Mn. A typical value is around 160 eV, but even better resolution may be reached, especially for small-surface detectors and when a long time constant is used for the linear shaping of the detector signals. To give an idea of the kind of information collected, Figure 2 shows two typical X-ray spectra obtained with Si(Li) detectors in a PIXE measurement of a sample of aerosol. We return later to a discussion of the characteristics of the spectra (in particular the background shape) and to the way of treating partially overlapping lines. Here the intention is to point out that the multi-element analytical capability is a real feature of PIXE with EDS, and not just a theoretical claim.

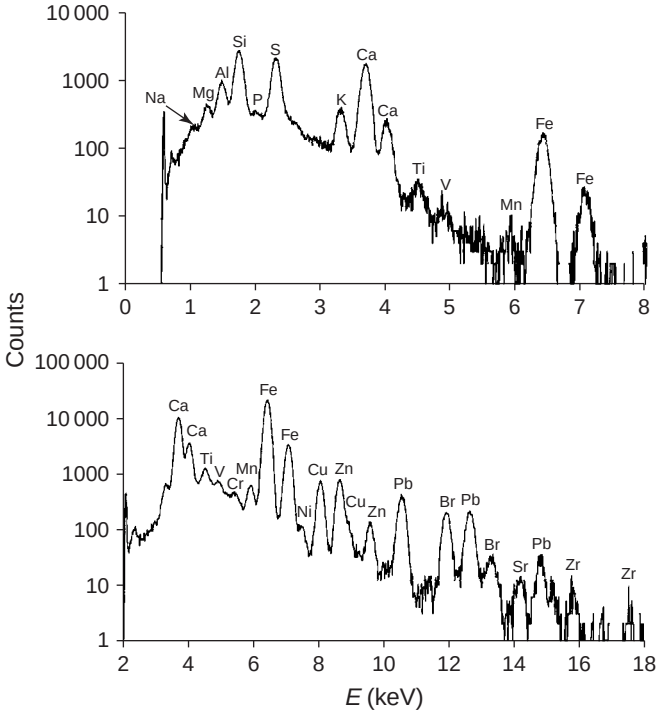

Figure 2 PIXE spectra from a sample of aerosol collected on a Millipore® filter. The spectra refer to the same sample, but have been recorded simultaneously by two Si(Li) detectors in different geometries (see Figure 11, and the discussion of two-detector set-ups in section 2.6.3). The spectrum acquisition time was about 7 min, at a proton beam current of 5 nA.

By using Bragg spectrometers that exploit X-ray diffraction from a crystal, wavelength-dispersive spectroscopy (WDS) may also be performed, which allows a much better energy resolution (by a factor of 50 or so). The WDS technique is clearly at a premium to solve close multiplets of X-rays of similar energies from different elements, but a serious drawback is a drastic drop of detection efficiency. In typical PIXE set-ups for analytical purposes, WDS is therefore never accomplished, because one of the great analytical pros of the technique, i.e. its high sensitivity, would be lost. After all, even with EDS most ambiguities may be solved by using other X-rays from the same element as markers of its presence. Thus, the use of WDS is left in general to studies of basic atomic physics, such as changes of electron binding energies as a function of chemical state, rather than to analytical applications. For this reason, we will not discuss WDS any further in this context.

1.2 Brief Historical Overview and Review Literature

The PIXE technique was established as a universal analytical technique by Johansson et al. in the early 1970s. Its popularity and the range of applications have constantly increased over the years. Today (1999), tens of laboratories throughout the world currently use PIXE for elemental analysis in fields as diverse as environmental pollution, biology and medicine, earth sciences, material sciences, art and archaeology, historical document examination, and forensic science. Comprehensive descriptions of the theoretical basis of the technique and its practical use in the main fields of application are available.[1,2] Dedicated conferences on ''Particle-induced X-ray Emission and its Analytical Applications'' have been held every third year since 1977; PIXE sections have been and are present in many other, more general, series conferences (e.g. International Conference on Ion Beam Analysis, European Conference on Accelerators Applied to Research and Technology, and others). Specific conferences on nuclear microprobes (where micro-PIXE applications are widely represented), and more recently on the applications in the field of biomedicine (bio-PIXE Symposia), are also periodically held.

The literature on PIXE work is distributed over several journals dealing with the fields of the specific applications. However, the above-mentioned Proceedings of the International PIXE Conferences – published by North Holland as separate issues of the review *Nuclear Instruments & Methods in Physics Research, Section B* – are a good source of concentrated information concerning both applications and basic physical and technological advancements. The journal *International Journal of PIXE* (published by World Scientific Publishing Co Pte Ltd, Singapore) contains topical and review articles on the various fields of application.

1.3 Comparison with other X-ray-based Analytical Techniques

1.3.1 Electron Probe Microanalysis

1.3.1.1 Minimum Detection Limits The absence of a primary bremsstrahlung (BS) continuum in the PIXE spectra (section 2.3.1), with the consequent low background in the energy spectra, is the main reason for much better (i.e. lower) detection limits in PIXE with respect to EPMA. In the latter, primary BS gives rise to a considerable, continuous background that extends to 20–30 keV (20–30 kV being the typical accelerating voltage used in the electron gun). This background is therefore present over all the energy region of interest for elemental analysis by X-ray line spectra. As X-ray production cross-sections for EPMA and PIXE are comparable under optimum conditions, the much lower backgrounds in PIXE spectra translate into better MDLs, typically lower by more than two orders of magnitude in PIXE with respect to EPMA.

1.3.1.2 Probed Depth The range of electrons typically used in EPMA ($E_e = 20–30$ keV) is of the order

of a few micrometers at most, in light-Z matrices. This is consequently also the investigated depth from surface with EPMA, for all elements. As discussed below, in typical PIXE ($E_p = 2 - 3$ MeV), the range traversed by protons is about 10 times larger; for higher-Z elements (detected through higher-energy X-rays) the probed depth is consequently much larger in PIXE than for EPMA. For the detection of lighter elements the effects of X-ray self-absorption (section 2.2) become dominant after only a few micrometers and the difference in the probed depth between the two techniques becomes smaller.

1.3.1.3 Space Lateral Resolution With an electron beam it is relatively standard to obtain beam spots on the target of less than 1 μm. Elemental mapping by scanning electron microscopy (SEM) is therefore easily performed at an EPMA facility. Standard PIXE is performed with larger beam cross-sections, but the possibility of micro-PIXE down to beam sizes below one micrometer exists, as briefly described in section 2.7. This makes it possible to perform trace-element two-dimensional mapping of a sample. However, it should be pointed out that EPMA yields excellent space resolution as a standard feature, whereas micro-PIXE requires sophisticated tuning and cannot therefore be considered a routine analytical technique.

1.3.1.4 Possibility of Nonvacuum Analysis With EPMA, it is mandatory to analyze the sample in vacuum. When large objects are investigated, this implies taking a sample of the material, which may be problematic when analyzing precious items. With PIXE, external beams are easily used (most practical aspects of the experiment get easier) which makes the analysis of large objects feasible with no need to take samples and no damage at all.

1.3.2 X-ray Fluorescence

1.3.2.1 Minimum Detection Limits Both PIXE and XRF can be used for trace-element analysis down to parts per million (ppm) levels. The difference between the two is the behavior of the sensitivity curve with Z. Generally speaking, due to the different ionization mechanisms (and the consequent trends of X-ray production cross-sections) XRF has lower MDLs for higher-Z elements than PIXE, whereas the analytical capability of PIXE extends to lower-Z elements. However, for both techniques the sensitivity curve can be varied by changing the energy of the ionizing agent (primary electromagnetic radiation and protons, respectively). The XRF method may become totally blind to higher-Z elements when choosing an exciting radiation suitable for obtaining lower MDLs for lower-Z elements, and vice-versa, whereas PIXE preserves the total range of detected elements even when optimized (by varying the proton energy) with a better sensitivity to lighter or heavier elements.

1.3.2.2 Probed Depth XRF probes considerably greater depths than standard energy PIXE as far as the heavier elements are concerned. For the lighter-Z elements (under conditions when XRF can detect them) it is again the self-absorption of low-energy X-rays – rather than the excitation mechanism – that determines the effective probed depth. Thus the difference is smaller.

1.3.2.3 Lateral Resolution The possibility of obtaining real microbeams for PIXE has been mentioned in section 1.3.1. As far as XRF is concerned, until recently the typical spot over which XRF analysis was carried was of the order of square centimeters. The analysis of much smaller spots required strong collimation, and this drastically decreased the exciting radiation intensity and therefore made measurement times unduly long. However, the recent development of focusing techniques for X-radiation (based on capillary light-guides suitable in the range of X-ray wavelengths) has opened up new possibilities of miniaturizing the exciting X-ray beams for XRF, while still preserving acceptable intensities. The useful spot size obtained with these techniques is now of the order of some tenths of a square millimeter.

1.3.2.4 Nonvacuum Analysis As with PIXE, XRF can be easily accomplished in nonvacuum conditions (although commercial XRF equipment is usually designed for in-vacuum analysis). In this respect there is no technical difference between the two techniques.

1.3.2.5 Nonlaboratory Analysis Portable XRF equipment has been developed, mostly based on the use of an X-ray tube as the source of electromagnetic exciting radiation. Although some limitations exist when performing fully quantitative analyses – especially for light elements – the opportunity to perform in situ analyses on non-transportable materials is considerable. The need for a particle accelerator to produce the beam that excites X-ray emission makes conventional PIXE an intrinsically nonportable technique. However, in the late 1990s portable PIXE equipment, based on the use of a purely α-emitting source (^{210}Po), has been developed and tested with success.[3,4] The present limitation of this apparatus is the low intensity of the α-beam due to isotropic emission of the particles from the source. This makes the measurements much more time-consuming than in the laboratory. Further developments in the relative source–sample–detector geometry are underway to improve the performance of this interesting variant to standard PIXE.

2 IN-DEPTH REVIEW OF THE TECHNIQUE

2.1 Basic Aspects of X-ray Emission from Atomic Shells

Let us first recall the spectroscopic nomenclature for atomic shells and for X-rays corresponding to electron transitions between them.

The potential energy is set to zero for free electrons, whereas electrons in the atomic shells have increasingly negative energies the more bound they are. The innermost electronic shell (principal quantum number $n = 1$) is the K shell, and corresponds to the most tightly bound electrons; therefore, for any given element, K electrons have the minimum energy value. Examples of K-level energies are −1072 eV for Na, −4038 eV for Ca, −7113 eV for Fe, −11 103 eV for Ge, −25 514 eV for Ag, −37 441 eV for Ba, and −88 005 eV for Pb.

The second shell (principal quantum number $n = 2$) is the L shell, with subshells L_1, L_2 and L_3 corresponding to $(l = 0,\ j = 1/2)$, $(l = 1,\ j = 1/2)$, $(l = 1,\ j = 3/2)$, respectively. Examples of level L_1, L_2, and L_3 energies are: −63.3, −31.1, and −31.0 eV for Na; −438, −350, and −346 eV for Ca; −846, −721, and −708 eV for Fe; −1414, −1248, and −1217 for Ge; −3806, −3524, and −3351 eV for Ag; −5988, −5624, and −5247 eV for Ba; and −15 860, −15 200, and −13 035 eV for Pb.

The sequence of other shells is then M, N, and O, with principal quantum numbers $n = 3$, 4, and 5, respectively, with less tightly bound electrons (this applies to higher-Z elements having a large number of electrons).

Once a vacancy is created in an inner shell (by any ionizing agent), not all electronic transitions from outer shells or subshells are possible. Transition selection rules arise from quantum mechanical laws, so that only certain transitions are permitted, each having a well defined probability.

In the case of ionization from the L or superior shells, the Coster–Kronig effect must be taken into account. For the L shell, this consists of a nonradiative transfer of a vacancy, primarily created in the L_1 or L_2 subshell, to a higher L subshell (L_2 or L_3 for primary L_1 ionization, L_3 for primary L_2 ionization). Coster–Kronig transitions occur before the filling from an outer electron takes place; the electron transition energy is then varied. The probabilities of L_1 to L_2 (f_{12}), L_1 to L_3 (f_{13}) and L_2 to L_3 (f_{23}) Coster–Kronig transitions are high (up to over 50% for the L_1 to L_3 transition), and vary with atomic number. Similar and even more complicated effects take place when the primary ionization occurs in a shell higher than L. Although the Coster–Kronig effect is important from the point of view of basic atomic physics, it has no practical consequence for quantitative analyses based on induced X-ray classification, such as PIXE, EPMA, and XRF. In fact, it can be treated as an effective change of ionization probability of the various subshells and automatically taken into account when performing quantitative PIXE measurements, for which the knowledge of ionization cross-sections is essential (section 2.2).

Once an ionization has been produced in an inner shell, deexcitation of the atom can be accomplished by X-ray or Auger electron emission. The ratio of the number of X-ray emissions to the total number of ionizations produced in the given shell is called the fluorescence yield relative to that shell and is usually indicated with w (w_K, w_L, etc.). For the L and higher shells, subshell ionizations and consequently subshell fluorescence yields can be considered separately.

Fluorescence yields may be calculated theoretically and measured experimentally.[5–8] They are a function of the atomic number Z and Figure 3 shows the behavior of w_K and w_L versus Z.

The nomenclature for X-rays emitted as a consequence of an outer-electron transition following inner-shell ionizations recalls both the primary vacancy location and the filling electron provenance. These are referred to as K-series, L-series, M-series, etc. X-rays, depending on the initially ionized shell. Within each series, the further specification is then related to the provenance shell of the electron. K_α and K_β X-rays correspond to the filling electron coming from the L and superior shells, respectively, with still further distinctions ($K_{\alpha1}$, $K_{\alpha2}$, $K_{\beta1}$, $K_{\beta2}$, $K_{\beta3}$) depending on the provenance subshells. Figure 4 gives a summary of the X-ray nomenclature for the main lines of the K and L series.

Because electron binding energies increase with Z, X-ray energies also increase with the atomic number. Figure 5 shows the trend for X-ray energies of the K and L main lines as a function of Z.

Within each X-ray series, the intensity ratio of the different transitions varies with Z.

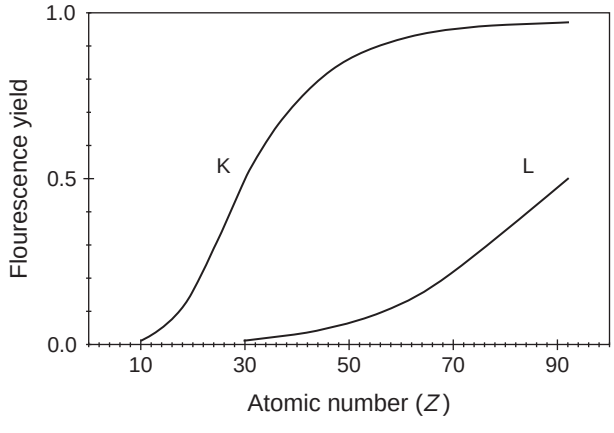

Figure 3 Trend of the K- and L-fluorescence yields versus Z. (Reproduced by permission from Johansson and Campbell.[2])

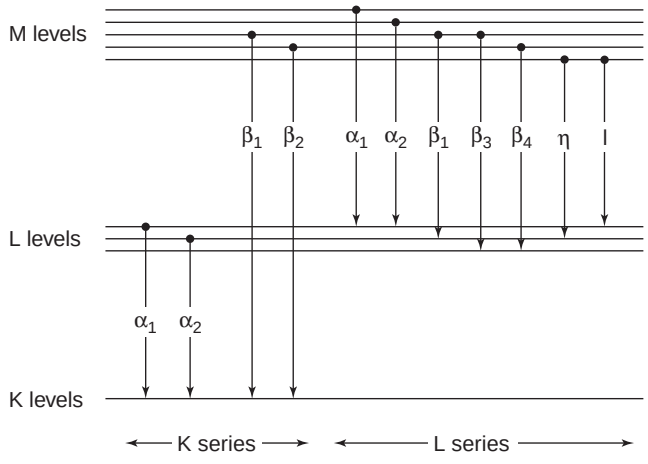

Figure 4 Summary of the main X-ray transitions to the K and L levels and of the corresponding nomenclature. (Reproduced by permission from Johansson et al.[1])

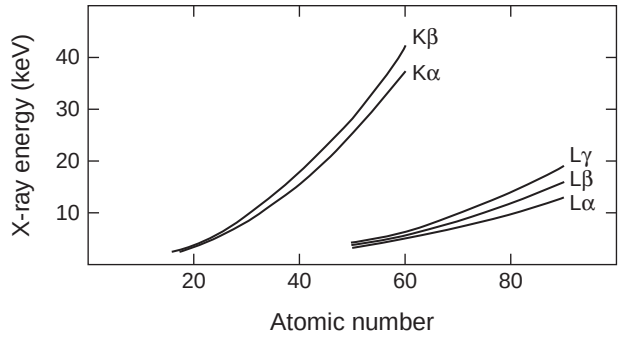

Figure 5 Energies of the main K- and L-series X-rays as a function of the element Z. (Reproduced by permission from Johansson et al.[1])

The ratio $I(K_\beta)/I(K_\alpha)$ of the two main groups (β and α) for K-series X-rays increases rapidly from 1.3% for Al (the lowest-Z element where K_β rays are possible) to 12% for K, then much more slowly (14% for Zn, 19% for Zr, 22% for Sn, etc.). The ratio $I(L_\beta)/I(L_\alpha)$ of the two main groups of the L-series X-rays decreases from about 70% in Sn to about 55% in the highest-Z elements; contrary to the K-series case, the $I(L_\beta)/I(L_\alpha)$ ratio also displays a dependence on the ionizing particle energy, but this effect is not very pronounced and is always known with sufficient accuracy for the purpose of spectra deconvolution.

2.2 Quantitative PIXE Analysis

Although a unified mathematical treatment of the procedure for a quantitative PIXE analysis might be given for all kinds of samples, it is easier to understand the principle by first examining the case of thin samples. An operational definition of "thin" and "thick" for a PIXE measurement is given below; for the moment, when referring to a thin specimen let us think of a few atomic layers.

2.2.1 Thin Specimens

The proton interactions with the inner shell electrons of the target atoms produce ionizations. The number of ionizations $N_{\mathrm{ion},j}(Z)$ produced in a given shell j ($j = \mathrm{K,L,M,\ldots}$) for a given atomic species Z in the target is obviously proportional to the target concentration of that atomic species, to the number of particles N_p having passed through the target and to the thickness t of the sample (Equation 1):

$$N_{\mathrm{ion},j}(Z) = \sigma_{\mathrm{ion},j} N_Z t N_\mathrm{p} \tag{1}$$

where N_Z is the number of atoms of atomic number Z per unit volume. The term $\sigma_{\mathrm{ion},j}$ is the *ionization cross-section* for the given shell j. As seen from Equation (1), it has the physical dimensions of an area, and is clearly a measure of the probability of the interaction giving rise to the specific ionization. Except for the case of the K shell, ionization cross-sections are made up of more terms, corresponding to ionizations of the various subshells, for instance σ_{L_1}, σ_{L_2}, and σ_{L_3}.

Measured values of K- and L- ionization cross-sections have been reported in a large number of papers; a much smaller amount of data is available for the M-shell case.

For the K shell, a collection of measured values is available[9] and a set of reference values based on statistical analysis of a compendium of published data has been compiled.[10] With the exception of very low atomic numbers, the results of the so-called ECPSSR (see below for a justification of the acronym) theory of the particle–electron interaction[11,12] agree within about 3% [13] with the best experimental data. This theory (developed from the pioneering work of Brandt and Lapicki[14,15]), treats the deflection and velocity change of the particles due to the Coulomb (C) field, the perturbation of the atomic stationary state (PSS) by the projectile, relativistic effects (R), and energy loss (E) during the collisions between beam particles and atomic electrons. For each element, the cross-section data (reference or theoretical) may be fitted as a function of proton energy E_p with a fifth-order logarithmic polynomial.[16,17] A universal treatment over a wide range of elements and proton energies with just six fitted parameters altogether, yielded – in the early stages of development of PIXE – a reasonable although not completely satisfactory reproduction of the tabulated data. However, using a distinct set of polynomial coefficients for each atomic number, tabulations can be reproduced within 1%.

The situation is more complex for the L shell, in which the ionization cross-sections differ and have a different dependence on the proton energy for the three subshells. As a consequence, the relative intensities between the various L lines also have a dependence on proton energy, which is not the case for K lines. Coster–Kronig transitions further complicate the overall picture; thus, comparison between experimental values (extracted for instance from the compilation of L X-ray production cross-sections in Cohen[18]) and theoretical predictions (such as the ECPSSR approach[19]) is more difficult and the agreement less satisfactory.

The trend of $\sigma_{\mathrm{ion,K}}$ and $\sigma_{\mathrm{ion,L}}$ with proton bombarding energy, taken from ECPSSR calculations for some elements, is shown in Figure 6. A strong decrease with increasing Z is apparent, especially for the K-shell ionization cross-sections. A strongly increasing trend with increasing E_{p} occurs within the range of proton bombarding energies usually adopted. The cross-section reaches a maximum when the projectile velocity matches that of the ejected electron. Further increase of projectile energy results in a slow decrease in cross-section. It is apparent from Figure 6 that the increase with E_{p} is larger for higher values of Z so that, in general, working at higher energies makes the analysis of heavier elements comparatively more sensitive.

For the M shell, much less systematic work has been done. However, although little use is normally made in PIXE of the detected M-series X-rays for the quantification of the high-Z elements, knowledge of M-subshell ionization cross-sections, and consequently of the intensity ratios of M lines, would be important.

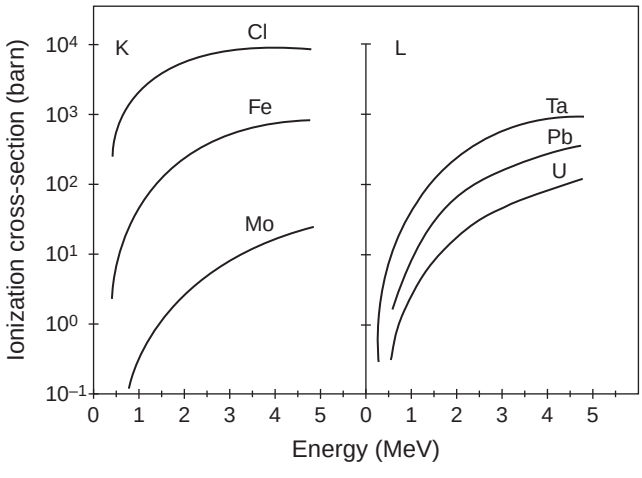

Figure 6 K- and total L-ionization cross-sections (in barn; 1 barn = 10^{-24} cm^2) for some elements versus proton bombarding energy. The displayed values are an ECPSSR calculation, which however is quite representative of the reference values from a compilation of the available experimental data. (Reproduced by permission from Johansson and Campbell.[2])

Indeed, in X-ray spectra from complex samples, the M lines from high-Z elements overlap the K or L lines of lighter elements so that a safe deconvolution of the multiplets must rely also on (experimental or theoretical) knowledge of the intensity ratios of the various M lines.

The product of ionization cross-section and fluorescence yield for a given shell gives the cross-section σ_{X_j} for the production of an X-ray of the corresponding series; from what has been said it follows that σ_{X_j} is also a function of Z and E_{p}. As stated above, within each series the branching ratio among the different lines is a slowly varying function of Z and, for the L shell, this also depends a little on the particle bombarding energy.

In terms of X-rays produced, the equivalent of Equation (1) is given by Equation (2):

$$N_{\mathrm{X}_j}(Z) = \sigma_{\mathrm{X}_j} N_Z t N_{\mathrm{p}} \tag{2}$$

The angular distribution in the emission of the produced X-rays is isotropic with respect to the beam incidence.

Equation (2) can be rewritten in terms of areal density $\rho_Z t$ of the element Z (mass per unit area normal to the beam, with ρ_Z standard volumic mass density) as:

$$N_{\mathrm{X}_j}(Z) = \sigma_{\mathrm{X}_j} \frac{N_{\mathrm{Av}} \rho_Z t}{A} \frac{Q}{e} \tag{3}$$

where N_{Av} is Avogadro's number and A the atomic mass of the species. In Equation (3), the number of protons N_{p} passing through the target is more conveniently expressed by integrated beam charge Q divided by unit charge e of the particles. Indeed, Q is usually a measurable quantity (see section 2.4.3).

The capability of performing a fully quantitative analysis with PIXE in essence derives from Equation (3).

In a thin target, for any given element the X-ray production cross-sections are constant throughout the target (because the beam energy loss is negligible, i.e. all interactions happen at the same proton energy, say E_0), and the number of detected X-rays is a well-defined fraction of those produced (isotropic emission of all X-rays). From Equation (3), one obtains the number Y_0 of X-rays from the element Z, detected during a measurement, simply by taking the overall detection efficiency into account. The latter is the product of a geometrical factor (related to the solid angle Ω covered by the detector) and of an intrinsic efficiency $\varepsilon_{\mathrm{det}}$, a function of the X-ray energy (and therefore of Z) depending on the detector characteristics (section 2.6.1):

$$Y_{0\mathrm{X}_j}(Z) = \sigma_{\mathrm{X}_j}(Z, E_0) \frac{\Omega}{4\pi} \varepsilon_{\mathrm{det}} \frac{N_{\mathrm{Av}} \rho_Z t}{A} \frac{Q}{e} \tag{4}$$

Equation (4) can be reorganized as follows (Equation 5):

$$Y_{0X_j}(Z) = \left(\sigma_{X_j}(Z, E_0)\frac{\Omega}{4\pi}\varepsilon_{\text{det}}\frac{N_{\text{Av}}}{Ae}\right) Q\rho_Z t \qquad (5)$$

or (Equation 6):

$$Y_{0X_j}(Z) = \eta Q \rho_Z t \qquad (6)$$

by defining an efficiency factor η for the given X-ray, beam energy, detector and geometry (Equation 7):

$$\eta = \eta(Z) = \sigma_{X_j}(Z, E_0)\frac{\Omega}{4\pi}\varepsilon_{\text{det}}\frac{N_{\text{Av}}}{Ae} \qquad (7)$$

From Equation (6) it is easily seen that – *provided* η *is known* – the areal density $\rho_Z t$ for a given element can be obtained by measuring the corresponding X-ray yield on the detector and the integrated beam charge on the target. Knowledge of η is in principle possible even in absolute terms, because the physical factors in Equation (7) are either known (σ, see above) or measurable (Ω, ε_{det}, see section 2.6). In practice, rather than relying on the absolute knowledge of the terms in Equation (7) for determining the efficiency factors $\eta(Z)$, the latter are obtained experimentally through Equation (5) itself. This is performed by bombarding a series of thin standards containing known quantities $\rho_Z t$ of the various elements, and measuring the corresponding X-ray yields and the integrated beam charge. The measurements on the standards must be performed using exactly the same geometry, with the same absorption layers (if any) and at the same beam energy as on the samples. The analytical procedure for the determination of elemental concentrations in unknown thin samples is therefore a relative procedure.

Figures 7 and 8 show the curves for the efficiency factor as a function of X-ray energy (i.e. Z), for the K and L series, respectively, in a typical detector geometry. As a result of the very large variation of X-ray production cross-section over Z, the efficiency factor in a given geometry also spans a wide range of values, drastically decreasing for the detection of X-rays of higher-Z elements within the same series. In addition, the intrinsic efficiency of Si(Li) detectors progressively drops at X-ray energies above 20–25 keV, which corresponds to K-series X-rays of elements around $Z = 50$. This would create a loss of sensitivity to the detection of heavier elements; the problem is to a large extent overcome by the fact that L-series X-rays (produced with much larger cross-sections, see Figure 6) can be used as markers of their presence. At these values of Z the L-series X-rays are well within the useful range for an Si(Li) detector

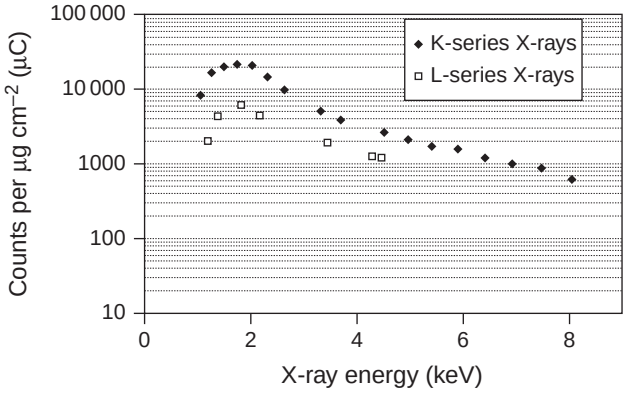

Figure 7 Trend of efficiency factors versus X-ray energies, for K-series and L-series X-rays, with an Si(Li) detector covering a solid angle of 2.5 msr (~12.5 mm^2 active area, ~7 cm distance) and no explicit absorbers. Experimental data obtained with thin standards certified to ±5% accuracy.

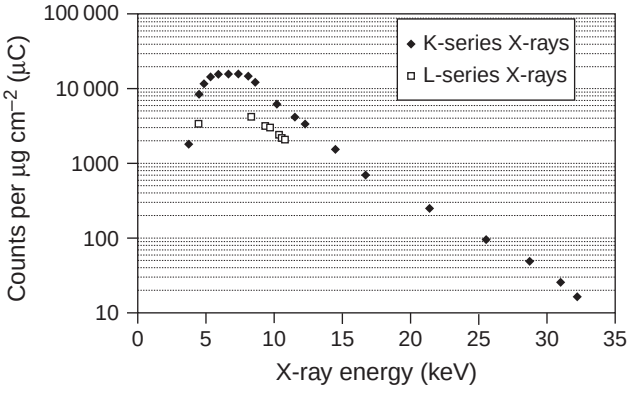

Figure 8 Same as Figure 7, but for a larger solid angle Si(Li) detector (~70 msr) having a 400-µm Mylar® absorber in front.

(see section 2.6.1), and this holds true for L-series X-rays up to the highest Z values. It will also be seen that in the higher-energy region of the spectrum the background is much lower, thus improving detection of weak lines. Further expedients to make detection limits more homogeneous through the widest possible range of Z are discussed in section 2.6.3.

2.2.2 Nonthin Specimens

Before seeing how quantitative analysis may be performed in the general case, it is useful to consider more carefully what "thin" and "nonthin" mean for PIXE analysis. The two physical effects that concur to determine whether a specimen is nonthin as far as PIXE analysis is concerned are:

- the beam energy loss while penetrating the sample;
- the self-absorption of X-rays produced in the interior of the sample.

A sample is thin if one may neglect these effects (the so-called matrix effects) and assume therefore:

- that all interactions happen at the same beam energy;
- that X-rays produced by the beam inside the target, emitted within the acceptance solid angle of the detector, have negligible probability of being lost to detection due to interactions inside the sample material itself.

By the expression ''thin sample'', we mean that the quantitative results obtained from its PIXE analysis by neglecting matrix effects are correct. Of course, this is never absolutely true and the validity of the thin-target approximation depends on the degree of accuracy one is looking for. Another important fact to bear in mind is that the thickness below which a sample may be considered thin depends both on its composition and on the X-ray energy that is observed (i.e. on the element, the abundance of which one is quantifying). Indeed, both beam stopping power and X-ray attenuation due to self-absorption depend on the average Z of the material; X-ray attenuation also strongly depends on the X-ray energy, increasing dramatically with decreasing X-ray energy. As a result of the latter point, a given physical thickness may be still considered ''thin'' as far as detection of a higher-energy X-ray is concerned, whereas it is already ''nonthin'' for the detection of a lower-energy X-ray from a lighter element. Approximately below 10–20 $\mu g\,cm^{-2}$ a sample of a medium-Z matrix can be assumed to be thin (to about 1% or better) for the quantification of all elements usually detected in PIXE (X-ray energies above 1 keV, corresponding to Na K_α). For the detection of higher-Z elements, detected through X-rays in the energy range above 6–7 keV, even a thickness of 2–300 $\mu g\,cm^{-2}$ might be considered thin.

When the sample thickness is such that the consequences of beam energy loss and X-ray absorption cannot be neglected, the extraction of quantitative information from the X-ray intensities in PIXE spectra is more complicated, but can still be dealt with. Indeed, both the physics of beam particle–target atom collisions, producing the energy loss, and that of X-ray interactions, producing the self-absorption effect, are well known and can be properly taken into account. The X-ray yield is obtained by integration of the contributions from the different depths inside the target. For each of them, one has to consider the proper X-ray production cross-section (which is a function of E_p and therefore of the depth from surface) and the reduction in the detected X-ray intensity, which also depends on the depth at which they are produced, because those X-rays produced deeper in the sample are more affected by self-absorption.

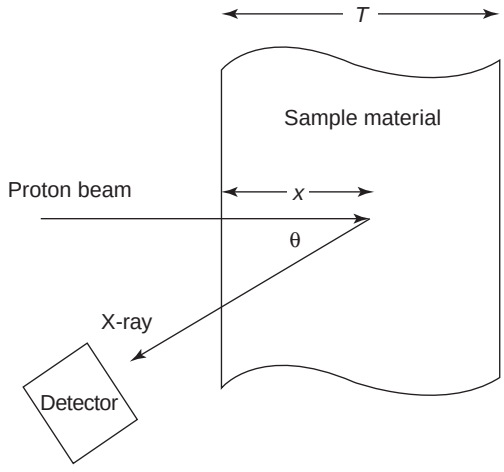

Figure 9 Diagram of a thick target analysis (see Equation 8).

The detected X-ray yield for a given element is obtained by the following generalization of Equation (4):

$$Y_{X_j}(Z) = \frac{\Omega}{4\pi}\varepsilon_{det}\frac{Q}{e}\frac{N_{Av}}{A}\rho_Z \int_0^T \sigma_{X_j}(Z, E_p)e^{-\frac{\mu x(E_p)}{\cos\vartheta}}\,dx \tag{8}$$

In Equation (8) T is the finite thickness of the sample and the exponential term accounts for the self-absorption of X-rays produced at a general depth x from the surface. The X-ray intensity attenuation coefficient μ is a function of the matrix composition and of the X-ray energy. The usual detection geometry (Figure 9) has been assumed in Equation (8), where ϑ is the angle at which the detector is placed with respect to the target normal.

In Equation (8), one should consider that x and E_p can each be expressed as a function of the other through the beam mass stopping power $S = dE_p/\rho dx$ (a function of the matrix composition and of E_p). Equation (8) can therefore be expressed as Equation (9),

$$Y_{X_j}(Z) = \frac{\Omega}{4\pi}\varepsilon_{det}\frac{Q}{e}\frac{N_{Av}}{A}\frac{\rho_Z}{\rho}\int_{E_0}^{E_0-\Delta E}\sigma_{X_j}(Z, E_p) \times e^{-\frac{\mu x(E_p)}{\cos\vartheta}}\frac{dE_p}{S(E_p)} \tag{9}$$

where E_0 is the initial beam energy (at target entrance) and ΔE is the total beam energy loss through the target thickness T (Equation 10):

$$\Delta E = \int_0^T S\,dx \tag{10}$$

In general, for any X-ray, a matrix correction factor F_{X_j} is introduced, defined as the ratio between the ideal yield Y_0 that would be obtained from the same target, in the absence of matrix effects, and the actual yield Y

(Equation 11):

$$F_{X_j}(Z) = \frac{Y_{0X_j}(Z)}{Y_{X_j}(Z)} \tag{11}$$

Therefore, from Equations (4) and (8), one obtains

$$F_{X_j}(Z) = \frac{T\sigma_{X_j}(Z, E_0)}{\int_0^T \sigma_{X_j}(Z, E_p(x))e^{-\frac{\mu x}{\cos\vartheta}}dx} \tag{12}$$

By recalling Equation (6) for Y_0, Equation (11) implies (Equation 13):

$$(\rho t)_Z = \frac{Y_{X_j}(Z)}{Q\eta_{Z,E_0}}F_{X_j}(Z) \tag{13}$$

where η is the efficiency factor, resulting from the measurements on thin standards at the surface beam energy E_0. In principle, therefore, the quantitative analysis of nonthin samples can be obtained from the detected yields $Y_{X_j}(Z)$, by considering the samples as thin (use of thin target efficiency factors) and then applying the proper correction factors F to account for matrix effects.

The problem remains as to how to obtain proper correction factors. This is a priori a problem, but can be solved quite well in many cases, such as when analyzing a homogeneous target of known thickness. The equivalent to the exact knowledge of target thickness T is a much more frequent case than might be thought at first sight. In fact, whenever the thickness, although unknown, is larger than the proton range R_p (of the order of 50–100 $\mu g\,cm^{-2}$), for the calculation of F it can be replaced by R_p itself, or by any larger value. In this case one talks about "infinitely thick" targets, where of course this term only implies that $T > R_p$. Even before this limit is reached, at least for the lower-Z elements, the exponential term in Equations (8) and (12) may be so small after a certain depth that the expression under the integral goes to zero and contributions from deeper layers can be neglected. In simple terms, for the lower-Z elements the dominant matrix effect is self-absorption rather than the decrease in production cross-section due to the beam stopping down; their X-rays continue to be produced at depths from which they will not reach the detector due to the self-absorption effect. Actually, as far as the detection of lower-Z elements is concerned, the target becomes infinite much before the condition $T > R_p$ is fulfilled.

It is also necessary to make an assumption about the matrix composition to evaluate the integral in Equation (12), because both μ and S are matrix dependent. If this composition is not known, as is usually the case in analytical work, the adopted technique then becomes an iterative procedure. The zeroth-order approximation of the matrix composition is obtained by neglecting all the matrix effects, i.e. by using Equation (13), with $F(Z) = 1$ for all values of Z. The matrix composition obtained by the areal densities thus deduced for the detected elements is then used to compute the stopping power and absorption coefficients to be used in Equation (12). From that, a first-order approximation of the matrix composition is obtained by reworking Equation (13), and this composition is then used again to compute the stopping power and absorption coefficients to be used in Equation (12), and so on iteratively. The procedure rapidly converges and a good estimate of the actual composition can be given. Two aspects must be pointed out in this respect.

First, the assumption is made of a homogeneous composition as a function of depth through the explored depth range (recall that we are talking about tens of micrometers at most). However, this assumption may be completely wrong, and the only way to check its validity is to perform further independent analyses with other techniques (or even simultaneously with other IBA techniques, or performing subsequent PIXE runs at different proton bombarding energies).

Second, the various approximations of the matrix composition, if based on just the PIXE-detected elements, can be largely defective because very light elements are hidden from the analysis (typically, this happens for $Z < 11$, but the actual limit depends on the specific set-up). The undetected fraction can be important and affects the actual matrix effects, which depend on all the elements present in the material. In many practically important circumstances the undetected fraction may be known independently, on the basis of separate measurements such as simultaneous measurements with other IBA techniques. In other cases, it can be indirectly deduced by the concentrations of the detected fraction itself; this is the case for instance in the analysis of rocks of volcanic origin. Here all elements can be considered to be present in the form of oxides, so that each element bears an amount of oxygen in a well-defined proportion to the detected element's abundance. Under these circumstances, the undetected fraction is well estimated and a good approximation of the whole matrix composition can be used to compute the factors F from Equation (12). Finally, it should be remembered that both the stopping power and the absorption coefficients depend on the matrix composition, but not so dramatically as to alter significantly the final values obtained by the iterative procedure of Equation (12) if a minor change in the matrix composition assumption is made. Obviously, this affects the different elements in a different manner: as a rule of thumb, lighter elements are more affected by incorrect assumptions about matrix composition. As for any other quantitative analytical technique, it must be

kept in mind here that the more unverified assumptions are made, the less accurate may be the final quantitative result.

If the sample thickness is larger than about 50 μm, the uncertainty about its value is irrelevant, because from the point of view of PIXE this is actually an infinitely thick target for the quantification of all elements. However, for a target of an intermediate unknown thickness, the quantitative analysis becomes problematic because, in principle, there may be different combinations of composition and thickness leading to the same detected X-ray yields. Some independent guess often may be made on either of the two, thus leading to reasonable results also in this case; however, it must be stressed that this remains the hardest situation to deal with for an accurate quantitative analysis.

2.3 Minimum Detection Limits in PIXE Analysis

A high value of the efficiency factor η for a given element, leading to large statistics for the corresponding X-ray peak in the PIXE spectrum (Equation 6), is obviously important for a sensitive determination of that element. However, this is not the only requisite, as the MDL depends on the signal-to-noise ratio.

In general, the performance in terms of the MDLs improves for any analytical method when the signal is detected with higher efficiency and when the noise is low. Because the cross-sections for the production of X-rays by accelerated particles are high, the number of X-rays produced by PIXE per unit quantity of a given atomic species is large. This means that the signal is good. The noise from which the signal must emerge is, in the case of PIXE (as well as in any other spectroscopic technique), the continuous background in the spectra. As can be seen in Figure 2, a continuous background is actually present in the X-ray energy spectra from a PIXE measurement, and characteristic X-ray peaks must be discriminated from the statistical fluctuations of this background. Let us therefore examine in some detail the reasons for the presence of this continuous background and the possibilities for the operator to reduce it.

2.3.1 Background in PIXE Spectra

There are several physical reasons for the presence of the continuous background in the PIXE spectra.

In the lower-energy part the continuous spectrum is higher and arises mainly from atomic electrons ejected by the interactions of the beam particles with the target. During their stopping down within the target, these electrons emit BS radiation. This is a consequence of a basic law of electromagnetism, that when a particle bearing electric charge undergoes an acceleration, it emits energy in the form of electromagnetic radiation. The intensity of the emitted BS is proportional to the square of the acceleration; its energy has a continuous spectrum extending up to the initial energy of the emitting particle. The electron BS in the lower part of the PIXE spectrum is called secondary BS, as opposed to the primary BS produced directly by the protons in their stopping down. It is easy to understand that:

- In PIXE, secondary BS has a much higher intensity than primary BS, which is essentially negligible. The forces responsible for both proton and secondary electron acceleration have equal intensity (being the Coulomb forces of interaction with atomic electrons) but the mass of protons is much larger than the mass of electrons. The proton acceleration and the corresponding BS intensity are consequently much lower.
- The maximum energy of the secondary BS is in the ratio $4m_e : M$ to that of the incident particles (with m_e and M being the electron and proton mass, respectively), this being the maximum energy transfer from incident particles to free electrons. Indeed, the maximum velocity acquired by the electrons, occurring for central collisions, is equal to twice that of the incident particles, so that Equation (14) holds:

$$E_{\max,e} = \frac{1}{2} m_e 4 v_p^2 = 4 \frac{m_e}{M} \frac{1}{2} M v_p^2 = \quad 4 \frac{m_e}{M} \quad E_p \qquad (14)$$

In the case of protons, the maximum energy value of the secondary BS is therefore around 4.3 and 6.5 keV for $E_p = 2$ MeV and 3 MeV, respectively. Consequently, under typical measuring conditions the background due to BS is practically zero in the higher part of the energy spectrum, i.e. the region of the X-ray lines used for the quantification of medium- and high-Z elements. This is just the region where X-rays are produced with lower cross-sections by PIXE. Therefore, where the signal is lower, the noise is fortunately lower also, and good MDLs are preserved. Many studies (both theoretical and experimental) have been carried out on the physics leading to the BS emission in PIXE.[20,21] An important practical feature is that BS emission is not isotropic, but has a maximum at around 90° to the beam direction. Therefore, to improve the performance in terms of MDLs, one should avoid placing the X-ray detector at around this angle. The most frequently used geometry is with the detector as backwards as possible to the beam direction, typically at 135° or more.

Some background continuum exists even in the higher energy part of the PIXE spectrum. It originates mainly from the Compton interactions of the higher-energy gamma rays produced by the incident particles either

in the target or in other materials before hitting the target (collimators, exit window in the external beam set-up, etc.). Compton interactions produce a partial energy release of the electromagnetic radiation in the material where they occur; the fraction of the full gamma energy that is released depends on the scattering angle in the Compton interaction. The resulting spectrum is therefore continuous and practically flat throughout the energy region of interest for PIXE, i.e. 1–30 keV.

The subject of gamma ray production in a PIXE experiment has been covered in detail.[22] The intensity of Compton background is generally low, and depends on:

- The detector material and active volume. It increases with increasing Z and this is one reason why silicon detectors are in general preferred to germanium ones. As far as the detector's volume is concerned, in the range of detector sizes that are used, larger ones produce a higher Compton background because of an increased probability of gamma ray interactions within their volume.
- The composition of the target itself and of the other materials with which the beam particles can interact. For low-Z materials, nuclear interactions with the beam protons are much more probable due to the closer approaches of the protons to the target nuclei, which in turn derives from the lower Coulomb repulsion. At the proton bombarding energies used in PIXE, most nuclei of low-Z elements can be excited inelastically and emit as a consequence gamma rays, which produce the Compton background in the PIXE spectra. The intensity of such emissions depends on the element, but also on the thickness of the emitting material and beam energy, due to the existence of resonances in the nuclear interaction cross-sections. Under typical conditions, fluorine and sodium are the elements giving rise to the higher gamma ray yield.

The Compton background arising from the production of gamma rays of elements within the target is unavoidable and the only expedient to lower it is to avoid the presence of unnecessary material around the detector, to decrease the probability of Compton scattering of the gamma rays in the surrounding materials. More can be done instead to decrease the contribution due to gamma rays originating from primary beam interactions on materials other than the target itself. First, low-Z materials such as aluminum should be avoided as constituents of chambers, collimators, scatterers, exit windows (in external beam set-ups), etc. Second, the points along the beam path that may constitute possible gamma ray sources arising from beam interactions should be effectively shielded towards the X-ray detector.

2.3.2 X-ray Line Interference

In addition to the presence of the continuous background, there is another factor that may worsen – in specific circumstances only – the MDL of some elements. This problem is common to all analytical techniques based on X-ray spectrometry (i.e. also XRF and EPMA) and it is recalled here just for the sake of completeness. This limiting factor consists of energy overlaps between X-rays originating from different elements. When the energy difference between two X-rays is much smaller than the detector resolution, they can hardly be discriminated by the spectra deconvolution codes. If one of the two elements giving rise to the overlapping X-rays is present in much larger abundance than the other, it can mask the presence of the weaker one even in the case of somewhat larger energy differences, within the tails of the strong peak. Thus, the possibility of detection of the weaker line is considerably worsened. Examples of this problem are:

- K_β lines of element Z with K_α of element $(Z+1)$, for $20 < Z < 30$;
- K line of Na (1.041 keV) with L lines of Zn (1.020 keV);
- K line of Mg (1.254 keV) with L lines of As (1.282 keV);
- K_α line of S (2.308 keV) with M lines of Pb (centroid at 2.346 keV) and to a lesser extent with those of Hg (centroid at 2.224 keV);
- K_α line of Ti (4.509 keV) with L_α lines of Ba (centroid at 4.465 keV);
- K_α line of As (10.532 keV) with L_α lines of Pb (centroid at 10.551 keV).

When the interference between two lines cannot be safely deconvoluted, the quantification of the two elements must be done on the basis of other lines of their spectra. This often implies that minor-intensity X-ray lines come into play and the MDL for the corresponding element increases.

2.3.3 Minimum Detection Limits

In the light of all the above considerations concerning high X-ray production cross-sections and detection efficiencies, and low continuous background in the energy spectra, it turns out that PIXE is basically a very sensitive technique. Figure 10 gives an idea of the MDLs of PIXE under typical analytical conditions. It should be emphasized that this is just a general indication, and that in any given analytical situation the actual MDL for a given element may be different. The MDL can depend on specific features of the sample under analysis, such as whether it is thinner or thicker, or the specific matrix material. Experimental conditions also have an

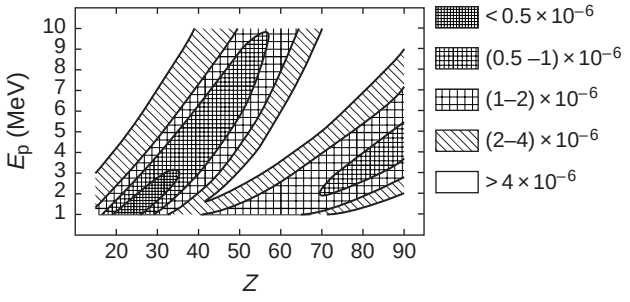

Figure 10 Attainable MDL as a function of proton bombarding energy and atomic number, in typical conditions. (Reproduced by permission from Johansson and Johansson.[17])

effect: beam energy, beam current intensity, and total time of target exposure to the beam all affect the analytical performance in terms of the MDL. Under certain circumstances, the attainable MDL can be much better than suggested by Figure 10. An example is when searching for ultratrace elements of medium Z in a very low-Z matrix. The opposite occurs in the presence of a high abundance of elements that give rise to X-rays interfering with those of the elements searched for.

When dealing with MDLs, both relative (in terms of minimum detectable mass over the total mass of the sample) and absolute values should be considered. Indeed another great quality of PIXE is that absolute quantities as small as picograms of the elements in the detectable range can be quantified in measurements lasting only a few minutes. This makes PIXE a very useful technique when nondestructivity is a mandatory requirement (as in the case of precious items such as works of art), or when the quantity of material to be analyzed is very small (as in some applications related to environmental pollutants).

2.4 Ion Accelerators and Sample Irradiation

2.4.1 Beam Production and Transportation

Most of the particle beams for PIXE analysis are produced by small Van de Graaff accelerators, with maximum terminal voltages of 2–4 MV. Accelerators of this kind often exist in nuclear physics divisions of universities or other research institutions, being the equipment used for nuclear spectroscopy studies in the 1960s and 1970s. In these situations the initial cost of setting up centers for IBA and specifically PIXE has been limited to the dedicated irradiation chambers (or the specimen holders for external set-ups, see below) and to the detection systems, which are not especially expensive. This certainly favored, in the 1980s, the start-up of PIXE activities in many laboratories. However, these accelerators are becoming less supported by their manufacturers. A new generation of small electrostatic accelerators (based on voltage generators of both Van de Graaff and Cockroft-Walton type) have been installed in some laboratories either to replace older ones or to start up new PIXE (and other IBA) activities. Most of these new accelerators are of the tandem type. In tandem accelerators, two acceleration tubes are placed on opposite sides of the positive high-voltage terminal, located in the centre of the machine, inside a tank which is gas filled at high pressure (this is needed to avoid voltage discharges to ground). The ions are produced by an external negative-ion source and then injected into the first acceleration tube from the low-voltage side. After acceleration to the positive voltage terminal, the ions undergo a charge inversion through the stripping of the outer electrons (produced by passing the beam through proper thin carbon foils or gas flow), so that they are further accelerated in the second tube towards ground. In this way, a lower terminal voltage is sufficient to produce a beam of sufficient energy. For instance, with a 2-MV tandem, proton beams up to 4 MeV, or α-beams up to 6 MeV can be produced. Further advantages of tandems derive from the presence of the external ion source (contrary to the case of single-ended electrostatic accelerators, where the ion source must be on the high-voltage terminal inside the high-pressure tank). With different sources, it is possible to accelerate any kind of ions, including the heavier ions that may be used for other IBA techniques. In addition, one can perform accelerator mass spectroscopy (AMS) for the search of ultratrace elements or rare isotopes in a sample. For this purpose, the sample is used as the material to be ionized in the source; the best-known case is the quantification of ^{14}C for archaeological dating. Together, these factors have initiated start-up of several new laboratories (or the modernization of previously established ones) based on tandem machines, where both PIXE and other ion-beam-based applications are performed.

In standard PIXE, the beam from the accelerator is led to hit the sample by conventional beam transportation systems, mostly based on magnetic deflection and focusing. The typical energy stability of beams produced by this kind of accelerator is quite good, with a residual voltage ripple normally less than 1 kV rms at maximum voltage. For PIXE applications this is not a real need, however, because the trend of X-ray production cross-sections with beam energy is rather smooth. Also, the beam currents that the accelerators might produce (up to tens of microamps) are not exploited, due to the large value of the production cross-sections. Typically, PIXE measurements are performed at beam current intensities of 20–30 nA at most (but much less than 1 nA for certain applications in which precious samples might otherwise be damaged by a more intense beam; see section 3).

Thus, in both respects (energy stability and current intensity) the accelerator performance is never a limitation for standard PIXE. This is less true for micro-PIXE set-ups, in which a much higher beam brightness is needed and high energy stability is at a premium. However, the standard performance of typical accelerators is already adequate for micro-PIXE; thus, a new accelerator of special design is not required just to implement a micro-PIXE facility.

The beam transportation line is obviously under vacuum. Typical values are around 10^{-6} mbar, although worse vacuum levels may be tolerated for PIXE applications (but not for micro-PIXE). Visual inspection of the beam cross-section is often provided along the beam line by retractable quartz viewers, whose fluorescence is observed through viewports by means of TV cameras. This makes the tuning of focusing and deflection systems much easier. For the same purpose, retractable Faraday cups are often available at critical points along the beam line.

The beam cross-section at the target may need to be either as small as possible (for a microbeam, section 2.5), or larger so as to average over a wider area of interest, which may be even some tens of square millimeters. In the latter case, it is necessary to have a uniform distribution of the beam current intensity over the probed area, in order to correctly average out inhomogeneities in the target composition.

A uniform beam current density can be achieved in several ways. One possibility is to pass the beam through a thin metal foil (some hundreds of micrograms per square centimeter) placed one or two meters upstream. Then the multiple small-angle scattering within the foil widens the angular aperture of the beam, and an adjustable collimator picks up its central, homogeneous part over the desired size. A side effect with this technique is an energy loss and straggling of the beam, but this is not a problem in PIXE and can be perfectly taken into account in the quantitative analysis.

Alternatively, the beam may be defocused by untuning the magnetic quadrupoles present on the beam line for focusing, and then collimating close to the target. Although this technique is simpler, it is less effective because it is difficult to monitor the actual homogeneity of the beam intensity over the selected area. A third method uses a well focused beam, and rasters with an x-y scan over the selected sample area. The disadvantage of this technique is that, in the presence of target inhomogeneities, the X-ray count rate from the different parts may vary considerably during the raster. Consequently, the pile-up corrections in fitting the PIXE spectra (section 2.6.2) may become problematic, as the corrections applied by the data reduction software are based on the assumption of a time-uniform count rate.

All collimators along the beam transportation lines are generally made in materials producing no detectable X- or gamma rays. Tantalum is the usual choice far from the target point because of its good dissipation of the heat generated by beam stopping. The production of K-series X-rays (around 60 keV) and of gamma rays due to Coulomb excitation of the nucleus is not intense, whereas copious L X-rays produced are easily absorbed by placing lead shielding in the line of sight to the detectors. Close to the target, the preferred material is graphite, which does not emit X-rays if sufficiently pure.

2.4.2 Irradiation Chambers

In conventional PIXE, irradiation is performed with the sample in vacuum. This is a limitation in several important kinds of application, such as when large samples are to be analyzed and taking samples of material is not allowed, or when keeping a specimen in vacuum could alter its composition. To overcome these problems, many laboratories have implemented external beam irradiation facilities (section 2.5).

The irradiation chamber is often made in aluminum, for economical reasons, but this material is not best suited because of its relatively high cross-section for the production of gamma rays. The regions of the chamber (as well as of the interior of the beam line) that are exposed to the beam (due to scattering) should be lined with nonemitting materials, such as ultrapure graphite.

Most chambers have a multisample capability in order to make the analysis faster. More or less sophisticated means of automatic sample changing are available. Usually a glass window allows the operator to view the samples inside the chamber with a video camera, both during sample changing and during irradiation, to check sample integrity under the beam.

The usual detection geometry is with the X-ray detector as backwards as possible, because the secondary BS is lower (section 2.3.1). The X-ray detector may be placed outside the vacuum chamber, viewing the sample through a window made as thin as possible so as to limit absorption of low-energy X-rays. Alternatively the detector cryostat may even be incorporated within the chamber, sometimes with the possibility of a windowless geometry based on a sliding vacuum seal (section 2.6.1). A means of inserting a variety of X-ray absorbers of different thickness and material must be provided, to enable changes to the detection efficiency curve versus X-ray energy according to the specific analytical demands of the experiment (section 2.6.2).

Additional detectors are often present in the chamber to complement PIXE analysis with the simultaneous use of other types of ion-beam-based techniques. These usually include one or more surface barrier silicon detectors

with which to implement backscattering spectroscopy. When using protons as projectiles, it is not always safe to assume ''Rutherford'' elastic scattering as valid, i.e. cross-sections may not always be computed from purely Coulomb projectile–nucleus interactions. In the proton elastic scattering on light-Z elements, the effects of the nuclear force on cross-sections may be nonnegligible or even dominant, due to the close approaches that can take place between protons and target nuclei. In spite of this, however, useful quantitative information can be extracted (corroborating and sometimes adding to that obtained by PIXE), provided that experimental data (including the nuclear interaction effects) are available for the cross-sections. Inelastic scattering can also provide useful information. A germanium detector may be faced to the irradiation chamber for the simultaneous detection of gamma rays: PIGE. Because gamma ray energies are much higher than X-ray energies, and consequently the absorption of electromagnetic radiation is not a problem, the germanium detector need not be placed inside the chamber.

2.4.3 Measurement of Beam Current and Charge Integration

During an irradiation, it is important to monitor the beam current because the sample may be damaged if safe levels of intensity are not maintained. The safety level varies with sample types, being much lower in general for electrically and thermally nonconductive materials. Equally important is accurate integration of the beam current during each run, as this is often a fundamental parameter for correct quantification of the obtained data (see Equations (3), (6), and (13)).

With thin specimens that allow the beam to pass through them with negligible energy loss, the traditional technique may be used of both measuring and integrating the beam current in a Faraday cup. The cup must be obviously insulated from the beam line. The cup material or lining must be such as to avoid emission of detectable X- and gamma rays, for the same reasons given above for the materials of beam line, collimator, and irradiation chamber. Typically graphite is used. Care should be taken that the cup receives all the particles that have traversed the beam; with samples of intermediate thickness, a relevant overall scattering may occur and the Faraday cup geometry must ensure an adequate acceptance angle. Also, the phenomenon of secondary electron escape from the cup may be relevant, which could give rise to an overestimate of the actual beam current. To prevent this effect, a negatively charged ring may be used as a guard electrode at the entrance to the cup, to reject outgoing electrons. Alternatively they can be deflected and thus recollected through the use of small magnet assemblies (electrical bias and small magnets will not affect the much heavier and energetic incoming particles). An accuracy as good as 1% can be achieved through the use of good-quality current meters and charge integrators.

With thick samples that stop the beam, the current must be measured and integrated on the sample itself, which requires the samples to be made electrically conductive if not already so (in order to avoid sample charging and consequent abrupt discharges). This may be achieved by a thin surface carbon coating, as in the case of EPMA. Secondary electrons escaping from the sample may be a much more relevant effect than with Faraday cups, due to the open geometry; they must be recollected by using a negatively biased suppressor electrode at a few hundred volts potential. Alternatively, the target itself may be positively biased with respect to the chamber.

With thick samples – or even with samples of intermediate thickness, causing a large scattering of the beam – alternatives to the Faraday cup may be used to measure the beam current. A thin self-supporting foil can be inserted along the beam path before the sample, and the count rate of backscattered particles, measured by a proper particle detector, can be taken as an indirect measurement of beam current intensity. Alternatively, the yield of some other interaction product of the beam particles with the foil material – such as X-rays themselves – may be used as an indicator of the beam current. In some external beam set-ups, the beam line exit window itself is used. When using such procedures, the problems of correct quantification of the current connected to secondary electron emission are completely overcome. After a calibration they may even provide a good absolute measurement of the current, provided that the metal foil does not change its thickness during irradiation due to beam-induced damage. A variant of this method is to measure the yield of backscattered particles (or other beam-induced radiation) on a rotating vane intercepting the beam at a sufficiently high frequency that, at the end of the measurement, the current sampling procedure is significant. Typically, 1 Hz is sufficient since run times are of the order of minutes.

2.5 External Beam Set-ups

An increasingly popular variant of conventional PIXE is external or ''in-air'' PIXE. This, as well as other IBA techniques, can be performed simply by passing the particle beam through a thin window, thus hitting the sample in its natural environment. The beam size can be determined by an aperture in the last section of the beam path within the vacuum, whereas the X-ray detectors are external. The sample is placed a few millimeters from the window and, by using a positioning stage (sometimes with micrometric control), can be moved in the plane

perpendicular to the beam for selective analysis. The external environment is often saturated with helium, to reduce beam energy loss and straggling with respect to air. The helium may be simply flooded in front of the exit window, with no containment, or be held in a simple closed environment. The use of helium has the further significant advantage that the transmission of low-energy X-rays is much higher than in air, which increases the possibility of detecting lighter elements. As an example, more than 96% of the 1.04-keV K X-rays of sodium are still transmitted when traversing 3 cm of helium, whereas for a similar distance in air the transmission would be only 2.4×10^{-6}; for the X-rays of sulphur (2.31 keV), the transmissions in 3 cm of helium and air are 99.7% and 28%, respectively. A disadvantage is that helium can penetrate the thin beryllium window of the Si(Li) detectors (section 2.6.1). This produces a progressive worsening of the vacuum levels inside the cryostat, which may necessitate a pump and bake procedure, typically after some months of use. To overcome this problem, the use of hydrogen rather than helium has also been reported (detector windows are much less permeated by the hydrogen molecules), but has not gained popularity for the fear of accidents due to its inflammability.

Figure 11 is a sketch of a simple external beam set-up, with helium flooding in an unenclosed environment and two Si(Li) detectors.

Topics relevant to external beam analysis have been reviewed in several articles.[23–25]

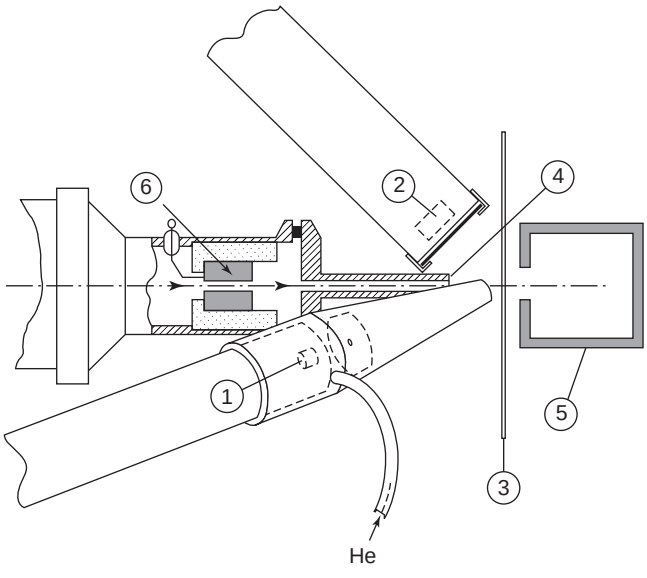

Figure 11 Diagram of a two-detector, external beam set-up (viewed from above): (1, 2) X-ray detectors, (3) sample for analysis, (4) beam exit window, (5) Faraday cup, (6) graphite collimator.

The main advantages of an external beam set-up are:

- ease of handling and changing of the sample – even large objects may be analyzed, and the specimen positioning is not problematic;
- drastically reduced risk of beam-induced damage to the sample – the problems of local heating are much smaller, especially when using a helium flow, which is a very efficient heat remover;
- electrical charging of insulated samples does not occur – this avoids the need for surface coating by graphite, as is usually done with samples in vacuum;
- ease of installation and change to the set-up (type, number and position of detectors, tilting angle of the sample, provisions for monitoring camera, etc.).

The first three of these points have encouraged the use of external beam set-ups for the analysis of fragile and/or precious objects, such as items of artistic, archaeological, and historical interest (sections 3.4 and 3.5). However, the external beam set-up is advantageous for most PIXE applications. Only for applications requiring a microbeam below about 10–20 μm is the external beam set-up not appropriate. Indeed, the scattering in the window and in the external path produces a beam halo. Thus, even with the best focusing, the actual beam spot size on the target cannot be smaller than a minimum value (depending on several factors, mainly the distance travelled by the beam in atmosphere). Some inconvenience with an external beam may also arise for other IBA techniques, such as Rutherford backscattering spectroscopy (RBS) or nuclear reaction analysis (NRA), when operated simultaneously with PIXE. As in those techniques the accuracy of the analytical information is directly related to the beam energy, the energy straggling due to the exit window and to the residual path in the atmosphere worsens the quality of the data obtained. However, in the late 1990s thinner beam exit windows have become available (see below). This, and the use of helium outside the window, has enabled the advantages of the external beam set-up to be retained without excessive drawbacks.

In external beam set-ups, the exit windows for the beam must be mechanically strong enough to sustain the pressure drop from the vacuum environment of the accelerator beam lines to atmosphere, and also exhibit a high resistance to radiation damage. Certain materials, such as Mylar®, which are mechanically strong enough, cannot tolerate even weak beams without early rupture. A typical material for beam exit windows is Kapton®, a polymer of composition $(C_{22}H_{10}N_2O_4)_n$. At a thickness of 8 μm, Kapton® typically withstands a beam of 10 nA with a cross-section of 1 mm^2 for 2–3 days, after which a change of the window is a wise precaution. The energy losses in

8 μm of Kapton® are 90 keV and 125 keV, respectively, for 3 MeV and 2 MeV protons; the fwhm energy straggling is about 20 keV (uncorrected Bohr estimate). Further energy loss and straggling before hitting the target are typically of the same order in air, or much smaller in helium. The overall values for both the energy loss and straggling (the former sum up linearly, the latter combine quadratically) are therefore of little consequence to PIXE, due to the smooth trend of X-ray production cross-sections with beam energy. Therefore, the possibility of a quantitative analysis is fully retained with external PIXE.

Another material sometimes used for exit windows is aluminum (about 10 μm thick), with a durability of months before changing the window. However, a higher continuous background is produced in the spectra due to beam-induced gamma rays from the window. Its use is therefore not recommended when bombarding very thin samples, where the X-ray yield is relatively weaker. With thick samples, an aluminum window may be a good choice.

The successful use of ultrathin windows of Si_3N_4, available down to a thickness of 0.1 μm, has been reported.[26] This material exhibits a very high resistance to the beam and the thin foils still withstand the pressure drop for a surface area of a few square millimeters. It has been employed therefore to extract focused proton microbeams down to 20 μm, with negligible halo due to the extreme thinness and to a very compact geometry with the specimen quite close to the window, in a helium atmosphere.

2.6 Detection of X-rays, Data Acquisition and Reduction

2.6.1 Working Principle and Characteristics of Si(Li) Detectors

Typical detectors used in PIXE are lithium drifted silicon crystals acting as diodes. In these solid-state nuclear detectors a free charge-depleted volume is created on one side of the junction by applying a reverse bias with a convenient high voltage power supply (some hundreds of volts, up to 1 kV). The ionizing radiation that interacts in this volume produces free hole–electron pairs that are promptly swept by the electric field – due to the voltage bias – towards the collecting electrodes. As the number of electron–hole pairs created inside the active volume is proportional to the energy release of the ionizing radiation (an average energy value of 3.6 eV is needed to create a hole–electron pair), the total charge collected on the electrodes is also proportional to the released energy. The current pulses through the detector produce, at the preamplifier output, voltage steps of amplitudes proportional to the energy releases. These voltage signals, when appropriately amplified and shaped, are then amplitude-analyzed to construct the detected X-ray energy spectra (section 2.6.2).

The Si(Li) detectors must work at a low temperature (typically −170 °C). Otherwise, when voltage biased, the inverse current of the diode (increasing exponentially with increasing temperature) would damage the detector itself and/or the first stage of signal preamplification. In particular, the compensation due to the presence of the drifted Li ions might be lost in a short time due to their mobilization (induced by the electric field). Si(Li) detectors can instead be safely stored at room temperature when they are not under bias. Protection circuits should always be used (and are indeed available in modern detectors) to prevent the high voltage supply from actually applying the bias until a sufficiently low diode temperature is reached.

Because of the need for a low operation temperature, the crystals are kept inside a cryostat under vacuum and cooling is typically achieved by conduction through a cold finger kept in contact with liquid nitrogen in a Dewar. The option of a Peltier cooling system is also available, which makes the assembly much more compact at the expense of less-effective cooling (which implies a slightly worse energy resolution). A vacuum is first created in the cryostat at time of manufacture. A proper amount of molecular sieves, in thermal contact with the cold finger within the cryostat, passively maintain a sufficient level of vacuum when the detector is cooled. In general, this situation may be maintained for several years, after which servicing (pump and bake procedure) is needed to restore the vacuum.

The cryostat wall in front of the detector (the entrance window) must be very thin if low-energy X-rays are to be transmitted with negligible probability of being absorbed by the window itself before reaching the detector's active volume. The window should also be light-tight because silicon detectors are photosensitive, producing a large response to ambient light. Traditional cryostat windows are made of beryllium, of a thickness down to 8 μm. They can stand the pressure drop over a surface area of about 1 cm^2, but they must be handled with great care because they are very fragile. Any puncture of the window may be catastrophic, with sudden ventilation of the detector's cryostat. This in turn may have very serious consequences in that the detector system can be destroyed if the diode was under bias – the sudden vapor condensation on its surface gives rise to current paths, which may spoil the passivated surface that maintains the bias voltage. In addition, the large currents may burn the entrance FET (field effect transistor) of the preamplifier.

Typical active areas of the Si(Li) detectors are from 10 to 100 mm^2 and the thickness of the depleted volume is from 3 to 6 mm. As far as the intrinsic efficiency

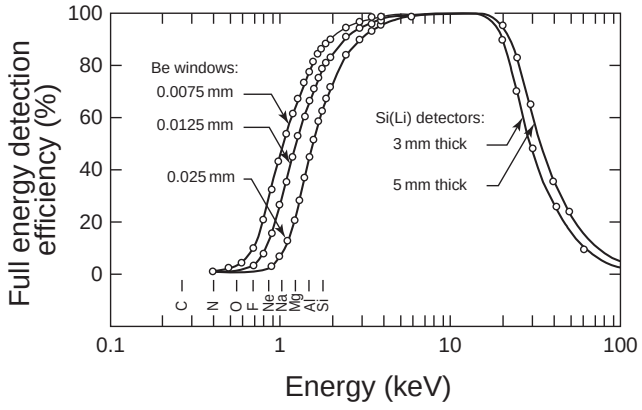

Figure 12 Intrinsic efficiency of Si(Li) detectors versus X-ray energy. The three curves on the left side refer to detector entrance windows of different thickness; the two curves on the right refer to different overall thickness of the detector's active volume. (Reproduced by permission of EG&G Ortec.)

ε_{det} is concerned (see Equation (4) and the following discussion), Si(Li) detectors have a 100% intrinsic efficiency over an X-ray range from 2–3 keV to about 20 keV. This means that in this energy range an X-ray reaching the entrance window will lose all its energy in interactions inside the active volume, thus producing a full-energy peak in the spectrum. Figure 12 shows the intrinsic efficiency of typical Si(Li) detectors as a function of X-ray energy, for various entrance windows and detector active thicknesses. The detection efficiency decreases at both lower and higher energies. On the high-energy side, this is in connection with the limited active thickness of the detector – with smaller detector thicknesses and higher X-ray energies, the probability of losing all the energy within the active volume decreases. On the low-energy side, the intrinsic efficiency decrease is due to the increasing probability of X-ray absorption in the detector entrance window.

Materials other than beryllium have also been used to build extra-low absorption windows that make the detection of very light elements possible, down to carbon. An alternative when the operator wishes to extend analysis to these elements is the possibility of window removal. This option can only be used when working with an in-vacuum set-up, and through some mechanical design to allow for vacuum-tight insertion of the cryostat inside the irradiation chamber. Removal of the window can only be done once a sufficiently good level of vacuum is reached in the chamber, in order to avoid vapor condensation on the cold detector surface.

It should be remembered that the severe problems of self-absorption within the sample itself (section 2.2.2) for the very-low-energy X-rays involved in the detection of these light elements (276 eV for carbon) makes their analysis only qualitative anyway. This consideration is more critical the lower the atomic number. As a general rule, relatively good quantification can be made only when starting from Na ($E_X = 1041$ eV); however, this also depends on the kind of sample (whether it is extremely thin or not, the nature of matrix, etc.) and on the degree of accuracy required.

Besides the effect of the cryostat window, at lower X-ray energies the detection efficiency is very sensitive to any absorption layers introduced intentionally or due to dead surface layers on the detector. The latter may occur when the vacuum level inside the cryostat worsens because of small leaks – a layer of condensed vapor is formed on the front surface of the cold detector inside the cryostat. Low-energy efficiency can also be drastically reduced by purposely interposing absorption filters along the path from target to detector. For example, with a Mylar® absorber of only 10-μm thickness the transmission of the 1.04 keV X-rays of Na is 3.3%, that of the 1.74 keV X-rays of Si is 47%, that of the 2.3 keV X-rays of S is 72%, and 92% of the 3.69 keV X-rays of Ca are transmitted. With 100 μm of Mylar®, transmissions are $<10^{-14}$, 0.05%, 3.5%, and 43%, respectively, for the X-rays of the four elements. It is obvious that by using different absorber material and thickness one can drastically modify the curve of detection efficiency versus X-ray energy. As explained in section 2.6.3, this property is often exploited to make the MDLs more homogeneous over a wider range of Z.

2.6.2 Spectra Acquisition and Data Reduction

The voltage signals produced by the charge preamplifiers are first shaped and amplified to make them suitable for amplitude classification, which is the basis for the acquisition of the energy spectra. A quasi-Gaussian pulse shape is normally used, which can be synthesized by different linear networks including successive stages of differentiation and integration of the original quasistep signal from the preamplifier.

The time constant used must be long enough to allow for complete collection of the charge created by the ionizing radiation inside the active volume of the detector. Incomplete charge collection would result in the so-called ballistic deficit effect, which produces low-energy tails to the line shapes in the energy spectra, increasing linearly with peak energy. Also, because of the prevalence of series noise, long time constants favor a better signal-to-noise ratio, i.e. better energy resolution.

At the same time, the time constant should be short enough to avoid a long time occupancy of the signal baseline by the pulse, in order to make handling of higher count rates possible. For the usual quasi-Gaussian pulse shapes a baseline recovery to better than 1% is achieved in a time equal to 3–5 times the time constant.

When the average time interval between successive pulses gets smaller than about 10–20 times the baseline occupation of a single pulse, there begins to be a non-negligible, increasing probability that a pulse arises before the previous one has completed its recovery to the baseline. This generates a wrong classification of the pulse amplitude (the pile-up effect) with consequent distortion of the accumulated energy spectra. Minor distortions may also be generated by imperfect tunings of the electronic chain, such as improper adjustment of the pole-zero cancellation or of the baseline active recovery circuits.

The choice of time constant is dictated by the main requirement in terms of performance (as outlined above, high count rate capability and good energy resolution are somehow conflicting). In general, a compromise between the two needs is given by a time constant value of 6–10 μs, but longer values also may be used when extremely good resolution is the main goal and the low count rate makes this possible. Of course, everything depends on the specific experimental situation – target composition and thickness, beam current, detector size and geometry, etc.

The linearly shaped pulses are amplitude analyzed by an analog-to-digital converter (ADC), which is then coupled to a computer where a distribution histogram of the digitized amplitudes is constructed. Because all the signal handling processes are linear, the histogram represents the spectrum of the radiation energy losses inside the detector. Thus, the full-energy X-ray interactions give rise to peaks in the spectrum. The energy scale of the spectrum can be calibrated by using a source of X- or gamma rays of known energies for a spectrum acquisition.

To extract quantitative information about elemental abundance in the sample, the peak areas in the X-ray energy spectra must be evaluated as accurately as possible. The various aspects of the spectrum deconvolution procedures are essential for quantitative analysis. In particular, modeling of the continuous background shape and deconvolution of peak multiplets have been widely studied. Several software packages are available to the analyst, including provisions for taking thick target matrix effects into account. Extended reviews of these aspects are available.[2,27]

The first step of the process is in general the modeling of the background. As the background is higher in the lower-energy X-ray region (section 2.3.1), where in addition the X-ray density (and hence the probability of overlaps) is higher, this is the most crucial energy region to be considered. Basically, two different strategies may be adopted: the assumption of an analytical (in general polynomial) function for the shape of the continuum; and an empirical subtraction of its contribution by a smoothing–filtering procedure that eliminates the variations of the counts per channel that are slower than a preset value. To be considered as not arising from true lines, these variations must meet certain criteria, which take into account the amplitude of the statistical fluctuations and the assumed peak widths. Once the contribution of the continuous background has been eliminated by this filtering procedure, the peak areas are determined from the filtered spectra (rather than from the direct spectra).

For the purpose of the detection of peaks, even in the presence of multiplets, and of the evaluation of peak areas, two aspects are essential: (a) knowledge of the detector's response to a single line, i.e. the peak shape; and (b) access to a complete and accurate database of X-ray energies and relative intensities for all the elements in the periodic table, so that after energy calibration the peaks can be assigned to the corresponding elements.

Considering the first point above, the line shapes of the full-energy X-ray peaks in the spectrum are basically Gaussian. The effect of electronic noise that adds randomly to the signal pulse height, and the stochastic nature of the process of electron–hole pair creation in the detector's active volume both contribute to this shape. X-rays losing identical energies inside the active volume do not always produce exactly the same number of pairs, and this fluctuation in turn results in voltage pulses of variable amplitudes and therefore in peak widening. Due to the statistical nature of pair production, this contribution to the fluctuations is proportional to the square root of the number of electron–hole pairs created, i.e. of the energy loss. The total width of a peak in the X-ray spectrum is then given by an expression such as Equation (15),

$$\Delta E = \sqrt{A + BE_X} \tag{15}$$

where A is the square of the electronic noise contribution, B is the square of the so-called Fano-factor relating fluctuations to the square root of energy loss, and E_X is the X-ray energy corresponding to the peak. This behavior of ΔE versus X-ray energy is taken into account when deconvoluting complex spectra.

However, the actual spectral line shape may be more complicated than Gaussian if ballistic deficit and/or pulse pile-up effects are present. Even in the absence of these effects, for high precision deconvolution an accurate representation of the detector's response is needed. Even minor effects give rise to lineshape distortions, which have to be taken into account, in particular to properly detect and quantify small peaks on the tails of larger ones. Figure 13 illustrates the main contributions to the lineshape. Several deviations from the ideal Gaussian shape can occur to a larger or smaller extent depending on the specific detector.[28]

A major effect that must be mentioned is the presence of an additional satellite peak – an escape peak – at an energy lower by 1.74 keV than the full energy peak.

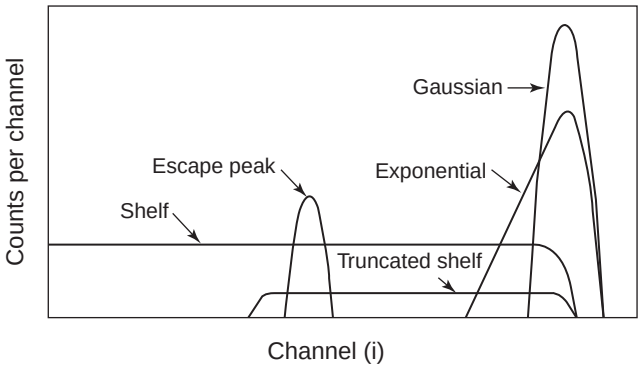

Figure 13 Diagram of the different lineshape components for a Si(Li) detector in response to an X-ray. (Reproduced by permission from Campbell.[28])

Escape peaks result from primary X-ray interactions that ionized – by fluorescence – the K shell of the silicon atoms in the detector crystal. The resulting K-series X-rays of silicon (1.74 keV) may then escape the detector before releasing their energy to the crystal. In these cases, the total energy released to the crystal is 1.74 keV below the primary X-ray energy.

The relative intensity of the escape peak with respect to the corresponding full-energy peak increases with decreasing X-ray energy. This is because the average depth of interaction of the primary X-ray from the detector surface is lower at lower X-ray energies, and it is then less probable that the silicon X-rays produced by fluorescence are recaptured within the crystal. The relative intensity of the Si escape peak to the full-energy peak falls from about 1% at 5 keV to about 0.1% at 10 keV. It is important to keep in mind the presence and intensity of the escape peaks because they may overlap the full-energy peaks of other elements, giving rise to some confusion in the spectra deconvolution if not properly taken into account. Examples of possible mistakes between an escape peak and a full-energy peak which may be important to an evaluation are:

- Ca K_α escape ($3.69 - 1.74 = 1.95$ keV) with the P K_α line (2.00 keV)
- Cr K_α escape ($5.41 - 1.74 = 3.67$ keV) with the Ca K_α line (3.69 keV)
- Fe K_β escape ($7.05 - 1.74 = 5.31$ keV) with the Cr K_α line (5.41 keV).

2.6.3 Optimization of Detection Limits over a Wide Range of Z

As discussed in section 2.2.1 (Figures 7 and 8), the physics of X-ray production in PIXE unbalances the detection efficiency in favor of the low- and medium-Z elements. This holds true even after using the favorable circumstance that L-series rather than K-series X-rays can be used for the quantification of higher-Z elements. This unbalance is a problem when detection of low or trace quantities of high-Z elements is required in the presence of a lower-Z matrix. Figure 7 shows that, if the count rate due to the low-energy X-rays emitted by low-Z elements is kept within tolerable limits, the total number of counts of X-rays from medium- and high-Z elements is too small within reasonable measurement times. In this situation two strategies may be adopted.

The first uses two X-ray detectors simultaneously, which complement each other in terms of the efficiency curve and the solid angle covered. One detector is chosen to cover the widest possible solid angle (which increases efficiency regardless of the X-ray energy), but is prevented from obtaining a large count rate from low-energy X-rays by shielding its surface with an absorber of appropriate thickness (this reduces the efficiency for just the low-energy X-rays). This detector is thus optimized for the medium- and high-Z elements but, due to the absorber, it is practically blind to the lightest elements (see the efficiency curve in Figure 8). The second detector, with no absorber but covering a much smaller solid angle, is then used to simultaneously provide analytical information concerning the low-Z elements. This case is represented by the efficiency curve of Figure 7. The combined effect of the two-detector system, when both L- and K-series lines are used, is that a much broader range of Z is covered with more homogeneity in the detection efficiency.

By changing the absorber on the higher-Z detector and the solid angle of the lower-Z detector (by using appropriate collimators), it is possible to optimize the global detection efficiency over the two detectors to the widest range of Z, tailored to the particular analytical problem (which trace elements are of interest, the matrix composition, etc.).

A cheaper alternative to a two-detector system is the use of a so-called funny filter on just one detector. A funny filter is a nonhomogeneous absorber, which can be constructed from a stack of foils having central apertures of different sizes. In this way different absorptions can be achieved for different solid angles. Thus the detection efficiency for higher-energy X-rays (which are little or not at all affected by the presence of the absorption layers) can be actually made larger than for the lower energies. In other terms, referring to Equation (7), the parameter Ω becomes an increasing function of X-ray energy, getting higher for higher-Z elements. A judicious choice of the funny filter geometry and composition can compensate effectively for the drop in X-ray production cross-section at higher Z. The drawback is that the actual shape of the function η versus E_X becomes more irregular than with a homogeneous absorber (or no

explicit absorber). This results in a more difficult modeling of the efficiency curve $\eta(E_X)$, which in turn is essential to obtain accurate quantitative results with PIXE. In practice, when using a funny filter a large number of standard reference samples (one for each element over a much larger Z range) is needed to experimentally determine the detailed behavior of $\eta(E_X)$. Even so, the uncertainty remains larger because the impact of possible inaccuracies in the knowledge of the areal densities of the individual standards is no more attenuated by the effect of ''averaging'' with neighboring-Z standards that is instead allowed when in the presence of a smooth behavior with Z.

2.7 Micro-PIXE

Depending on the specific analytical problem, one may need a relatively large size of probing proton beam (up to tens of square millimeters, when it is necessary to get an average composition on a wide surface), or a very small one (when the surface distribution of the elemental concentrations has to be mapped with the highest possible detail). When the beam diameter is reduced to about 100 μm or less, one generally talks about microbeams. In the simplest microbeam set-ups, the cross-section of the beam on the target is defined by collimation, by placing slits or apertures upstream, not far from the target. By this procedure, beam sizes as small as 100–200 μm can be obtained; this represents a sufficiently small size for many applications requiring detailed information on spatial distribution of elemental concentrations.

Sometimes an even smaller beam cross-section is necessary, and any further reduction of the beam size by brute collimation is no longer effective. At very small apertures, scattering of the particles by the slit or collimator edges produces significant beam halos over an area larger than that geometrically defined by the aperture. To obtain beam sizes down to 1 μm or below, strong focusing is necessary. The production of well-defined and sufficiently intense beams (1 nA of proton current is not exceptional) of the order of 1 μm is now a standard feature of set-ups based on strong focusing and the technical aspects of these set-ups have been widely reviewed.[29–31] Essentially, a microbeam is obtained through a beam-optics system which includes an object (a small, micrometrically adjustable slit aperture illuminated by the beam from the accelerator) and a set (doublet, triplet or quadruplet) of magnetic quadrupoles acting as a whole on the particle beam as a positive lens. The beam is focused on the target (image of the system), the distance of which from the quadrupoles is much smaller than the object–lens distance. In this way, high demagnification is obtained. In order to reduce the geometrical aberrations, the beam divergence at the entrance of the quadrupole system is strongly limited by collimator slits (Figure 14).

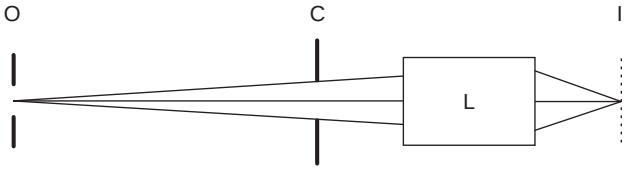

Figure 14 Diagram of a microbeam-forming system, with object slits O, collimator slits C to limit beam divergence, positive lens L (obtained with a doublet, triplet or quadruplet of quadrupoles) and image plane I where the beam is focused and the target must be placed. (Reproduced by permission from Grime and Watt.[30])

Almost invariably, the microbeam facility includes a system to fast scan the beam two-dimensionally over the sample, with a raster area having sides of the order of up to several hundred micrometers. This is accomplished by magnetic deflection with coils prior to the quadrupole focusing. Figure 15 is a photograph of the final section of a proton microprobe, including collimator slits, scanning coils, triplet of magnetic quadrupoles and irradiation camera, with micrometric adjustment of the target position and viewing microscope from outside.

In the presence of scanning coils, data acquisition is performed in the so-called list-mode. This means that when an X-ray is detected, a triplet of data is stored: the X-ray energy and the two coordinates of the beam spot position at the detection instant (deduced by the current values of the scanning coils at the instant of X-ray detection). The analysis of the triplets, which can be performed on line, includes the possibility of obtaining elemental maps (plotted as the number of detected X-rays of any given energy as a function of position over the whole scanned area). Alternatively, it is possible to

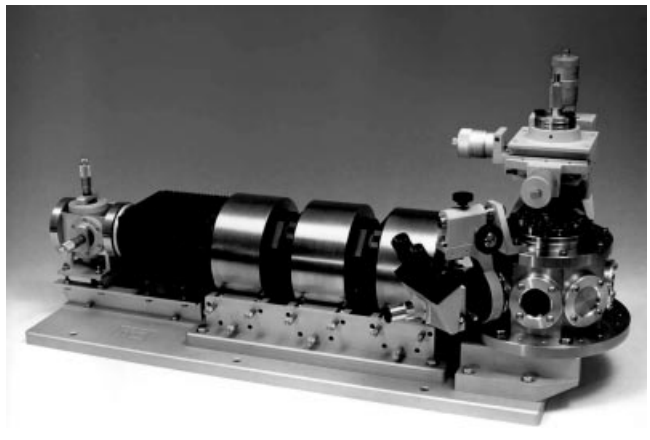

Figure 15 Photograph of a proton microprobe system, showing the final section. From left to right: the collimator slits (micrometrically adjustable), the scanning coils, the lens (made in this case of a triplet of quadrupoles), the target chamber. (Reproduced by permission of Oxford Microbeams Ltd.)

reconstruct the X-ray spectrum detected over any portion of the scanned area (down to pixels that correspond to the limits of lateral resolution, i.e. to 1 μm or even less). This is similar to the result of EPMA; however, the better MDLs offered by PIXE (section 1.3.1) make it possible to extend elemental mapping even to trace elements.

When pushing the limits of lateral resolution to their extremes, a number of factors come into play which limit the possibilities of obtaining ever smaller beam spot sizes.

First, the complex aberration effects of the probe-forming lens system have to be diagnosed and minimized. These include:

- astigmatism, due to the lack of cylindrical symmetry of the quadrupoles;
- higher-order multipole contamination of the quadrupole field, deriving from inaccuracies of construction of the quadrupoles (deviations from fourfold symmetry);
- spherical aberrations, coming into play only if higher currents are needed, which implies the acceptance of a larger beam divergence;
- rotational, translational and tilt misalignments of the quadrupole multiplet assembly.

Second, external sources can also produce broadening of the beam spot. Among these, an important role is played by the accelerator characteristics:

- beam brightness at the object slits, which in turn depends on many factors connected with the ion source, accelerator, and up-stream beam optics design;
- accelerator energy instabilities, which produce the equivalent of chromatic aberrations.

Other relevant factors affecting the quality of the microbeam are:

- scattering phenomena, both at the slits and due to residual gas in the beam line due to poor pumping;
- stray electromagnetic fields that may be caused by several sources in the vicinity of the beam line (transformers, motors, power supplies, etc.);
- mechanical effects such as vibrations (e.g. induced by pumps) in the target area, or even slow drifts of the target during irradiations, such as those caused by thermal effects.

One final point concerning PIXE with proton microbeam set-ups is that although a lateral resolution as small as 1 μm or less can be obtained, one should not forget that the depth investigated by PIXE at standard bombarding energies (2–3 MeV) is up to some tens of micrometers (sections 1.3 and 2.2). Consequently, in thick targets the compositional information is an averaged value over a larger size. Thus full advantage of ultrahigh lateral resolution is only obtained with very thin targets (such as with certain biological specimens), or in samples where the elemental map of the surface layer is also representative of the structure in depth (as in the analysis of electronic devices).

3 SURVEY OF THE MAIN ANALYTICAL APPLICATIONS

The analytical applications of PIXE are innumerable. Regular international conferences are held to give scientists of the IBA community the opportunity to exchange ideas on the technical developments of PIXE (improved databases of fundamental parameters, detection systems, data reduction codes, etc.), and on new analytical applications. New ideas continue to be proposed, especially when PIXE is combined with other IBA and nonIBA techniques, which is performed more and more. The following discussions outline some general considerations concerning groups of applications, with no claim to being exhaustive. The intention is to point out the peculiar analytical aspects of PIXE that are specifically important in each of the main applications.

3.1 Environmental Pollution Studies

These applications were among the first to be successfully performed during the early years of PIXE development. Both water pollution and air pollution are currently investigated, with more emphasis on the latter. The importance of aerosol analysis is obvious, both in view of air quality assessment in urban and industrial areas (impact on the health of humans, animals and plants), and as a tool in the investigation of general atmospheric phenomena such as short- and long-distance air transportation, visibility, and climate.

Since the early stages of aerosol investigations by PIXE, several reasons for a widespread use of this technique became apparent.

- Variations in the atmospheric aerosol composition can take place over very short timescales (of the order of an hour) or on longer, seasonal, timescales. The same applies to spatial variations: air transportation and deposition phenomena can produce extremely different results depending on the site of measurement. This large variability calls for vast numbers of analyses, and a method such as PIXE, capable of yielding fast quantitative results, lends itself to these

investigations much better than does conventional chemical analysis.

- All elements are potentially important in aerosol analysis, because many can be toxic, and because any compound, bound in particles of the appropriate size, can have an effect on visibility (by the scattering of light). Also, the simultaneous detection of groups of elements, occurring in particular ratios, can be a signature of the source of pollution. Their detection at different sites may constitute an important indicator of long-range air transportation phenomena. In performing the necessary multi-element quantitative analyses, the PIXE method offers advantages in terms of time and cost, with respect to chemical methods or optical spectroscopies. Particularly important is the possibility of complementing PIXE with the simultaneous use of other IBA techniques to extend the range of detectable elements, especially the lighter ones.
- The total quantity of aerosol mass available for compositional analysis is quite small. Typical aerosol mass concentrations in air range from a few to over one hundred micrograms per cubic meter. Traditional ways of collecting aerosol are by filtration over long periods of time, using large pumps and large-size filters in order to collect at least one or two milligrams of aerosol for the analysis of only some elements or compounds. The sensitivity of PIXE, for which overall quantities of about a microgram are adequate for a multi-element analysis, considerably reduces collection times and makes the use of much smaller pumping set-ups possible.
- It is important to obtain separate information on the composition of aerosols of different sizes. The particle size distribution is not homogeneous, and the relevance of particles of different sizes is quite different. As far as the impact on health is concerned, this is clearly connected to the penetration of particles into the breathing apparatus, with smaller ones reaching deeper levels and therefore being potentially more dangerous. As to visibility problems created by aerosols, the size-dependence of their effect, being connected to the scattering of light, is even more obvious. Finally, the size distribution also provides information on the sources of pollution. For all these reasons, sampling methodologies providing size-fractionation of the aerosol are highly desirable. These have been developed using inertial impactors that exploit the differing inertias of different sized particles when following air flow paths in special collectors. These samplers may also provide time-sequence deposition of the size-fractionated aerosol on streaks (typically, about 1 cm high) whose total length may correspond up to 1 week of aerosol collection or even more; each section of about 1 mm width corresponds to a short sampling period of typically 1 h. PIXE can be used to analyze these sections selectively, thereby reconstructing the time behavior of the aerosol composition. The overall quantity of aerosol deposited on each section is minute, typically of the order of 10 μg. The high absolute sensitivity of PIXE is therefore exploited to the highest degree. No other technique enables the analyst to obtain such a high time-resolution of the variation in aerosol composition, with simultaneous sensitive detection of a very large number of elements in runs lasting a few minutes. In addition, after PIXE measurements the samples remain available for alternative analyses, to corroborate or complement the data obtained by PIXE.
- The analysis of aerosol collected on ''traditional'' filters is also important. Daily sampling with these filters is routinely performed by public authorities in many western cities, but in most cases the only parameter measured is the total mass of aerosol deposited over the day. The collected aerosol is analyzed compositionally only for specific campaigns aimed at obtaining information about a few specific components; in these cases, with traditional analysis, the samples are destroyed. The nondestructive PIXE analysis of time sequences of these filters allows instead the trend of air pollution to be followed over long timescales without losing them for further tests. The reconstruction of aerosol composition has been reported in urban environments for periods of months or years.[32,33] It is obvious that over these timescales the use of 1-h resolution streakers would result in unmanageable quantities of data. Also with traditional filtering procedures, therefore, PIXE provides a precious tool in both widening the range of elements detected from each filter and preserving the samples for further investigation.

Let us now summarize the typical procedure for a PIXE analysis of aerosols.

After sampling on streak (for high time-resolution analysis), or on one-day filters (for studies of trends over long timescales), proton irradiation is performed with runs typically lasting a few minutes per sample.

For the analysis of streaks, the ''sample'' is each section of the streak and use is made of relatively small-size beams (typically 1 mm; microbeams are not needed). After completion of a run, the next section is immediately exposed to the beam by simply shifting the streaker by the proper distance, often in an automated process. For the daily filters, larger beams (made homogeneous over their cross-section by one of the methods outlined at the

end of section 2.4.1) are used and/or scans are performed over the whole area.

The currents used are typically 5–40 nA. Both in-vacuum and external beams have been used for this kind of measurement. External beams are preferable both to facilitate positioning, changing and/or scanning of the samples, and to minimize the risk of selective loss of some aerosol components when under the beam.

The spectra are normalized to the integrated beam charge on a Faraday cup behind the sample (the aerosol deposit plus substrate are thin enough to let the beam pass through), or to alternative normalization parameters (section 2.4.3).

The deduction of areal densities of the different elements on the deposit is then performed by simple comparison with thin standards. In the thin-target approximation, only the very lowest-Z detected elements may be slightly underestimated, due to the onset of self-absorption effects (section 2.2.2) within the particle grains. The latter effects can be of some relevance (up to 50%) only for the larger-size fraction of the particulate and for elements from Na to Si. However corrections can be applied, and/or these elements may be simultaneously detected by alternative IBA techniques, such as particle-induced gamma ray emission (PIGE), which also may produce data on lower-Z elements undetected by PIXE, such as fluorine.[34,35]

The elemental concentration in the air is finally obtained through knowledge of the sampling parameters (area of pumping aperture, flow rate, duration of sampling).

In a typical collection campaign, the amount of data accumulated is huge (30–40 elemental concentrations from each sample, separation according to particle size, hundreds of streaker sections (or daily filters) analyzed). Therefore, a statistical analysis must be performed. Principal component analysis (PCA) or factor analysis (FA)[36,37] are routinely performed on the data sets. These processes group the detected elements into factors according to the similarity of the behavior over time of their concentrations, thereby indicating pollution sources, reconstructing their relative weights in the total particulate mass, and also disentangling possible contributions from different sources to the same element.

Owing to the large amount of data that can be collected in a relatively easy way, large-scale monitoring of environmental quality makes wide use of PIXE data. For instance, the US monitoring network ''IMPROVE'' is largely based on PIXE-determined elemental concentrations in air.

Hundreds of papers have been published on PIXE measurements of aerosols. Specific results and review articles are published on the occasions of the triennial International PIXE Conferences.[38] Several reviews are available.[2,39,40]

3.2 Biomedical Applications

Biomedical applications of PIXE have been reviewed in detail together with exhaustive lists of references.[2,41,42] Also for biomedical work, a large number of specific papers are published in the Proceedings of the International PIXE Conferences,[38] Nuclear Microprobe Conferences, and the more recent topical series of bio-PIXE Conferences.[43,44]

The main reason for the great success of PIXE in biomedical applications is linked to the very nature of biological materials, where light elements – undetected by PIXE – are the dominant component. This favors attaining the best results in terms of MDLs for higher-Z elements. Provided that the samples can tolerate the relatively intense beam currents, MDLs well below the ppm level are easily reached. This trace element capability makes it possible to obtain precious information on physiological and pathological processes in biomedicine, where the oligo-elements often play a key role. The problem is in the correct preparation of the samples, because the biological matrix is easily damaged by irradiation and elements such as hydrogen and oxygen may be lost to a large extent due to heating. However, when these problems are adequately tackled, biomedical applications of PIXE can yield important results.

The second feature of PIXE that is important in biomedical applications is the high spatial resolution, i.e. the possibility of mapping trace element distributions in the tissues with a high definition. If a microbeam is used, the intracellular distribution of trace elements may be revealed.

To obtain good results in biomedical PIXE applications, close interdisciplinary collaboration between physicists and biologists is important at all stages of the experiment. This includes experiment design, sample preparation, data collection and processing, and interpretation of the results.

Traditionally, applications of PIXE in biomedicine have focused on the detection of trace quantities of transition elements and heavier elements in tissues. More recently, interest has shifted to the detection of lighter elements, such as aluminum, potassium and calcium. Often, the low concentration levels of these elements, which are well within the analytical capabilities of PIXE, are not easy to detect with alternative techniques, especially when dealing with small tissue samples or at the cellular level.

The sampling strategy is crucial to obtaining significant biodata from any analytical technique and must be adapted to the specific features of the chosen methodology. An essential point to bear in mind is that a large number of samples is necessary because of the large variability of trace element concentrations in biological

tissues. In this respect, PIXE is a good choice because it can analyze large numbers of samples in a relatively short time, in addition to preserving them for further complementary analysis. The latter point is an advantage in biomedical investigations because the bio-availability of trace elements may be quite different, depending on the chemical form in which the elements are bound. Thus further analysis of the same sample can be performed to yield this essential chemical bonding information.

Bio-PIXE experiments are usually performed on samples collected from living matter (biopsies). To avoid or minimize ethical problems connected with the acquiring of such samples, many investigations are performed on dermatological samples, such as nails or hair, as biological indicators of a physiological or pathological state of the individuals from which they were taken. The use of body fluids, such as sweat, tears, urine or blood, also implies minimum invasion in the sampling procedure. In any case, even when dealing with biopsies, the high absolute sensitivity of PIXE, requiring the collection of only minute amounts of tissue, is a good reason for selecting this technique. However, due to the very small amounts of sampled material involved, special care must be taken to avoid contamination from the surgical instruments used for the biopsies.

The need for small amounts of sampled material is beneficial to studies aimed at clarifying the processes of forest decline. Drill cores are collected from trees, small enough as not to harm them, and the radial trend of contaminant elements from environmental agents studied along the tree rings. One exploits here the fact that, in plants, only the external layer is living matter, interacting with the environment, and that each annual tree ring records the elemental content resulting from equilibrium with the environment for the corresponding year. Past environmental conditions can be thus reconstructed by simple PIXE scans on the small samples collected.

Sample preparation for bio-PIXE analysis depends on whether macro-analysis or micro-analysis is intended.

In macro-analysis, the desired spatial resolution in the detected elemental concentration is no smaller than a fraction of a millimeter; sometimes an average over a larger dimension is requested. When this is the case, many biological materials can be analyzed with a minimum of preparation, especially when an external beam set-up is used. Examples of materials analyzed in this way are botanical samples such as leaves, wood, and tree rings, or animal (or human) specimens such as bones, nails, hair, and teeth. Direct beam bombardment is possible in these cases, but no internal standard can be included and the absolute quantitative analysis may be difficult with PIXE alone because the matrix material is to a large extent undetected. However, in these studies it is often the relative behavior of minor or trace element concentrations that is searched for, so the matrix composition can be unimportant. Another point to be kept in mind when extracting analytical information from these untreated thick samples is that the total quantity of interrogated material volume is small, in the milligram range, due to the essentially surface character of PIXE analysis (section 2.2.2). Caution must be used in evaluating the representativeness of the results, which may be questioned if the material is not homogeneous to the macroscopic scale. Finally, surface roughness can also cause inaccuracies in the quantitative data, especially as far as lower-Z elements are concerned.

Soft tissues are often dried before analysis. The technique of lyophilization by freeze-drying is most common and performed by commercially available equipment. It has been shown that, for macro-analysis, the original composition is not affected by this procedure (whereas the spatial distribution of elements in micro-analysis could be affected if crystals are formed or thawing occurs). After drying, the tissues may be pulverized, homogenized and compressed to pellets for analysis. In this process, the material can be mixed with ultra-pure carbon powder, to ensure that the sample is electrically conductive. This avoids problems connected with insulated samples during bombardment. Throughout all these steps, great care must be taken to avoid contamination, such as from the pressing surfaces of the pelletizer.

To further increase sensitivity to the trace elements, the organic matrix mass can be reduced by well-established techniques such as wet ashing, dry ashing, low-temperature ashing, and acid digestion. The possible selective loss of some elements (e.g. chlorine and bromine in dry ashing) must be considered with care.

The preparation of biological samples for micro-PIXE analysis is different. It is clear that, for the result to be significant in terms of biological information about trace element distribution at the cellular or sub-cellular level, the spatial structure must be preserved during sample preparation and irradiation. Experience gained in the preparation of samples for electron microscopy is only of partial help here, because in those preparations chemicals are often added to increase visual contrast, and these may contaminate or selectively remove certain elements. Cryo-preparation techniques have been developed instead that lead to almost instantaneous freezing of the cells with preservation of a life-like state in terms of the elemental distribution. The frozen material must then be maintained at low temperature (below 130 °C) and the analysis performed on thin sections obtained with an ultracryomicrotome. As these sections show very little image contrast in using optical microscopy, their positioning under the microbeam with traditional optical methods is difficult. Contrast enhancement by

histochemical treatments is ruled out for fear of contamination, so alternative techniques have been developed. For instance, because the samples are thin, scanning transmission ion microscopy (STIM) can be used to properly direct the beam scan over the sample for microanalysis.

3.3 Applications to Earth Science

Review treatments of PIXE applications in the earth sciences have been given in several articles. (See references 2 [Chapter 3], 45, 46)

The impact of PIXE for the earth sciences has increased during the 1990s, with developments in microbeam technology. Micro-PIXE has early been acknowledged by geologists as an ideal complement to traditional compositional investigations using EPMA. The very low MDLs attainable by PIXE, even on the micrometric scale, make it very attractive to extend to trace elements the studies of elemental partitioning between different phases in a rock, or the zoning effects within a single crystal.

Whereas the MDLs of EPMA can be pushed at most to hundreds of ppm, micro-PIXE with beams of the order of 10 μm can reach, with relative ease, a factor of more than 100-fold better, i.e. ppm sensitivity. The attraction for Earth scientists is that this extra sensitivity is achieved with the same sample preparation (thin sections glued on a glass support) and using the same analytical principle used for EPMA (i.e. X-ray spectroscopy), with which they are so familiar. In the preparation of thin sections for PIXE, it must be remembered that the penetration of 2–3 MeV protons is larger than for 20–30 keV electrons (section 1.3.1). Therefore, thicker sections must be cut (about 70–80 μm; the standard thickness for EPMA is 30 μm), to prevent the protons from reaching the backing. With this precaution, both techniques can be used on the very same sample. In addition to providing additional information on trace elements, PIXE also provides quantification of the major and minor elements so that the results can be compared to EPMA data for the latter elements. What is crucially important is that the additional trace-element analysis provided by PIXE is achieved in a nondestructive manner, unlike alternative highly sensitive techniques with high spatial resolution, such as mass spectroscopies SIMS (secondary ion mass spectroscopy) or LAMS (laser ablation mass spectroscopy).

In comparing EMPA and PIXE results for major and minor elements, the differences in the probed depth and in spatial resolution (section 1.3.1) must be considered. This comparison is useful in that it adds reliability to the overall analytical data. The quantification of elemental concentrations from the PIXE spectra can be as accurate as 1–2%, for two reasons. First, the sample thickness is infinite and the problems connected with knowledge of the integration interval, discussed in section 2.2.2, are avoided. Second, the undetected fraction (very light elements) is in most cases (such as for silicate rocks) easy to estimate just on the basis of the detected fraction; or even all elements are actually detected, as in sulfide rocks. In silicate rocks the elements can be considered to be present in the form of oxides, thus the only undetected element is oxygen and its abundance can be estimated from the stoichiometric ratios to the detected elements. In practice, an initial estimate is done of the total oxygen abundance (which is often around 40%), in order to make a first approximation of the matrix effects in the computations described in section 2.2.2. Iterative procedures then lead to a self-consistent dataset of all elemental abundances, oxygen included.

Typical problems that have been addressed by micro-PIXE in the field of earth sciences are trace element partitioning between phases, which provides important clues to the physical and chemical parameters of the environment at the time of rock formation, and mineral zoning effects as regards trace elements. The latter frequently do not follow the pattern of major elements, and the different behavior is of relevance to geologists. These investigations require a beam of small dimensions; however, this does not necessarily imply the use of a microbeam. Beam sizes of a few tens of microns may often be adequate. One should keep in mind that the depth within the mineral, interrogated by the proton beam, is of the order of 10 μm, or more for the medium-to-heavy elements typically searched for in trace quantities. A higher lateral resolution would therefore bring no benefit (section 2.7). The relatively broad size of microbeam required means that no specially refined tuning is needed on the set-up.

More traditional, broader beam sizes have been used to investigate rocks to obtain bulk composition data. Here, the obvious competitor to PIXE is XRF which, when operating on a larger spatial scale, is capable of yielding trace element abundance. Indeed, the same kind of problems are addressed by XRF and PIXE, and even the same samples can be used. As the problem in bulk composition analysis is the inhomogeneous structure of the rocks, the latter have to be thoroughly ground to a very fine powder and homogenized. The final sample preparation step is usually pressing the material into pellets, which are then exposed to the X-ray or proton beam. Due to the trace element capability of both techniques, great care must be taken to avoid contamination by the instruments used for grinding and pressing. Although in principle both XRF and PIXE are capable of yielding global information on composition, it must be said that the former technique has better MDLs, if properly applied, for the higher-Z elements, whereas PIXE is better for the lighter ones. Another difference lies in the different depth investigated, but

this is not relevant for homogenized samples because the pellets are effectively infinitely thick for both techniques. Finally, a common feature of XRF and PIXE analyses, which is disappointing to geologists, is that the important problem of rare earth patterns cannot be investigated. On the one hand, the detection efficiency for the K-lines of rare earth elements (which are sufficiently energy separated from each other) is too low, so that they cannot be detected at their low levels of concentrations. On the other hand, the L-series X-rays, which might instead be detected with more efficiency, cannot be safely deconvoluted because the overlaps of their multiplets are too close. Some work has been done to overcome these problems. Higher-energy PIXE has been attempted, but the advantage of the increase in K-series X-ray production cross-sections is accompanied by a much higher continuous background, induced in the spectra by the onset of many nuclear reactions. The latter lead to emission of gamma rays, producing Compton interactions in the detector (section 3.1). Also the attempt for a proper deconvolution of the close-lying multiplets of L-series X-rays was basically unsuccessful, as even small differences in the assumed X-ray lineshape in the spectrum fitting procedure alter the deduced relative intensities of the various elements.

When considering broad-beam PIXE applications in the earth sciences, not only XRF but all the other traditional techniques can usually be applied as well. This is because there is usually plenty of material to analyze and the nondestructive character of XRF and PIXE becomes less important. The advantage of XRF and PIXE is in their multi-element capability, which makes the compositional characterization of samples from bulk materials much faster. However, as a summary consideration one should admit that the real benefit of PIXE to the earth sciences is only when using microbeams for trace-element distribution studies.

3.4 Applications to Art and Archaeology

Recent surveys of PIXE applications in the field of art and archaeology can be found.[47,48] These applications probably started later than those mentioned above, but have now been developed to a large extent as a result of more widespread use of external beam set-ups. The importance of PIXE measurements (and IBA in general) in this domain is demonstrated by the installation of a dedicated accelerator in the Laboratoire de Recherche des Musées de France, in the basements of the buildings of the Louvre Museum in Paris.[49–51] With the prominent use of PIXE analyses among others, this accelerator has produced significant contributions to a better understanding of the materials and techniques used in the past for the production of works of art. Intense and significant activity in this field has also been performed in many other laboratories throughout the world.

The problems addressed by PIXE analysis in this field fall into two categories: first, an advancement of the knowledge of materials and production technologies used in works of art in ancient times; second, the investigation of materials prior to restorations, so as to appropriately choose the restoring procedure, and avoid incompatibilities and irreversible effects between materials.

The motivating forces within the first category can be many. Reconstructing the material composition of artifacts may lead to:

- deeper insight into the technical developments of the past, allowing the historian to better understand the advance in the technological skills of artisans at various sites and times;
- discovery of the sources of supply of raw materials, thus contributing to the solution of historical problems, such as the existence of certain trade routes at the time of manufacture of the ancient objects under analysis;
- indirect dating through compositional analogies with dated materials and, under favorable circumstances, lending support to attributions to specific artists, providing a criterion for authentication (or refutation).

To the second category of problems, namely the restoration of works, should be added the significant contribution to a better understanding of the deterioration of the works of art.

Whatever the reason for analyzing the material of an object of artistic or archaeological relevance, in almost all cases a mandatory requirement is that no damage must occur, and quite often the physical removal of some material for analysis is forbidden. This raises a difficulty in that the inherent inhomogeneity of the materials used to produce works of art (whether one is dealing with paints, or metal alloys, or terracotta, or any other) calls for an analysis extended to many samples. The representativeness of the results might be otherwise questioned. Hence there is a need to perform such analyses with techniques capable of yielding a full compositional characterization of artifacts in a fast and nondestructive way. PIXE meets these requirements probably better than any other analytical technique. With the exception of metal objects, external set-ups are needed so that the nondamaging character of the technique is thus fully preserved provided that sufficiently low beam currents are used. Low beam intensities, on the other hand, are feasible in most investigations because of the nonthin targets (i.e. the quantity of material is relatively large) and because the problem is often not the

search for trace elements. Typical beam currents used in this kind of application are well below 1 nA, sometimes down to a few picoamps.

In terms of risks of damage, the most sensitive precious materials are glass, porcelains, and any vitrified surface. With these materials, even with external beams, currents of a few tens of picoamps can sometimes produce a visible stain after a measurement. This sort of damage can often be tolerated, but the curator of the object must be warned about the risk. This damage is attributed to atomic dislocations in the structure of the material, giving rise to color centres. Stains induced in this manner have been often reported to fade away after some days or weeks, due to a self-annealing process, even at room temperature. To minimize the effect, whenever the sample size makes it possible, a diffused beam over a larger area may be employed, or a smaller-size beam may be scanned across the surface (or the surface moved under the beam).

Although all kinds of materials may be analyzed by PIXE, most published work has concentrated on gold artifacts (coins, jewels), temperas used in miniatures in illuminated manuscripts or incunabula, and terracotta. Relatively little PIXE activity has concerned the materials of traditional paint on wood or canvas. This is because PIXE is a surface analysis technique, and little or nothing can be discovered about the deeper layers. Thus, if PIXE is performed on the paintings as they are, the protective varnish, which is invariably laid over the color in paintings on wood or canvas, prevents access to the paint materials. Material sampling is seldom allowed; alternatively, the varnish could be removed, but this is only done on the occasion of restorations, and is not normally permitted for general investigations, such as when trying to characterize the color palette of a given painter or school.

For paints, the field where PIXE is best exploited is that of illuminated ancient books. Here, no varnish was used because the paints were believed to be naturally protected by remaining unexposed to direct light and atmospheric agents for most of the time. Much PIXE work has concentrated on this subject,[25,52–54] and unexpected results of art-historical importance have been found; for instance, the widespread use of ultramarine blue has been detected in Italian scriptoria dating from early medieval times. Until this finding it was believed that this pigment (obtained by a sophisticated refining technique from lapislazuli, a semiprecious stone only found in Afghanistan) was reserved for very important works, at least until the early Renaissance, because of the high cost of import from a distant site.

All PIXE work on miniature paints has been performed with external beam set-ups. This is essential with manuscripts, because of handling considerations and because only by the use of an external beam (in a helium flow) can the measurements be really nondamaging, in all senses.

The surface character of PIXE analysis also explains why most of the published work on metal products concerns gold. The surface patina layer covering ancient objects made of other metals is a problem to PIXE, unless the material under study is the patina itself (which can in fact be the case sometimes). As far as gold items are important work includes that by Demortier,[55–57] which led among other things to the discovery of skilful soldering techniques as early as 3000 years ago. This work demonstrated the use of various gold alloys with decreasing melting temperatures, which avoided the unsoldering of nearby previously soldered details. These measurements have been performed both with external beams and with in-vacuum microbeams. For this kind of object, indeed, no disadvantage (damage, or difficult handling due to large size) comes from the use of a vacuum set-up, whereas the use of an in-vacuum microbeam provides sometimes a better tool of analysis.

Terracotta has been often investigated by PIXE analysis. The information searched for in this case is usually a provenance clue, often based on the trace-element content of the material. Thus, the PIXE quality most exploited is its very low MDLs. When dealing with terracotta, the compositional analysis is usually performed on fragments, whose esthetical value is irrelevant; therefore, higher beam currents may be used with no fear of possible visible damage, and the sensitivity to trace elements achieved in runs lasting a few minutes. However, due to the surface character of PIXE analysis, caution must be exercised when deriving conclusions from the detection of traces – surface contamination might be the reason for finding trace concentrations of unexpected elements. These risks must be carefully evaluated and procedures for reducing them adopted (such as analysis of cleaved surfaces of the fragments, or preliminary cleaning with appropriate noncontaminant processes). The statistical analysis of detected trace-element abundances over large series of samples may also help in discriminating significant results from accidental findings. An alternative to direct irradiation is to sample some material, which in small amounts is often allowed when dealing with fragments. In these cases, the material can be powdered and pressed into pellets for analysis. The same considerations as for the analysis of rocks with macro-PIXE (section 3.3) apply in this respect, in particular about the alternatives to PIXE for these analyses.

An example of extensive PIXE investigations of this kind is the analysis of pottery and obsidian artifacts performed in Australia.[58] A large number of pottery samples have been analyzed with a PIXE/PIGE set-up; their composition was clearly correlated to local

sources of clay, which made it possible to establish relationships between prehistoric settlements and their trade connections. The obsidian artifacts, which were largely used by prehistoric populations as tools for everyday life or for hunting, were prepared using material collected from local deposits of volcanic origin. The compositional characterization of hundreds of these tools by PIXE and PIGE, followed by statistical analysis for pattern recognition, made clear separations into different groups possible. Because the populations took their tools with them when migrating from one land to another (even over distances of thousands of kilometers), the archaeologists were able to reconstruct the processes of migration by matching the composition of obsidian tools to that of volcanic flows in quite different places.

3.5 Applications to Historical Problems

Although related to the previous section, the studies described below have been separated out because, for them, the use of PIXE has been particularly useful and in certain respects unique. The study of ancient written documents, namely of the composition of paper, parchment, and especially ink, is a tool to characterize and differentiate them. Such characterizations address problems of chronology, authenticity, or technical procedures for the production of books in ancient times. For the study of inks in particular, PIXE is an exceptionally suited tool owing to its high absolute sensitivity, its nondestructivity, the possibility of using beams well below 1 mm in diameter in external set-ups and to perform simultaneous quantitative analysis of many elements. All these features make it really useful for these problems.

In a study of one exemplar of the Gutenberg Bible,[59] the paper and ink were analyzed by an external 1-mm^2 beam. The ink was notable for its high Cu and Pb content, and their weight ratio was used to characterize the ink. It was thus possible to reconstruct the sequence of printing of the various pages of the book.

Another study of the same group[60] concerned the so-called Vinland map, which is believed to be a pre-Columbus document giving the first cartographic evidence of North America. However, its authenticity had been debated and after a chemical analysis (which was performed on a forcedly limited number of tiny samples of ink, as they had to be picked up from the map), it was concluded that it must be a forgery because of a high content of TiO_2, which has become available only in recent times. An extensive PIXE analysis of ink from this document, performed instead on a very large number of spots, with no need for picking up material and no damage, showed that the amount of titanium was in fact quite small, thus re-opening the possibility that the map is authentic. In this case, the nondestructive character of a PIXE analysis in situ was the winning card.

Ink analysis by PIXE has also been used to indirectly reconstruct the chronology of documents, as in the following example. During his life Galileo produced, besides the published works, hundreds of notes, hand-written on loose sheets of paper, where he wrote propositions and sketched diagrams about his ideas on the problems of motion and on astronomy. These sheets are obviously undated and conflicting opinions have been raised within the community of science historians about their chronological order. Establishing their dates on the basis of an objective criterion is important for the reconstruction of the development of Galilean thought. Analysis with PIXE of the ink composition in some of Galileo's dated hand-written documents – such as letters – first showed[61] that a correlation can be established between different ink compositions and periods of Galileo's life. The inks are of the so-called metallo-gallic type and the quantitative ratios between the various metallic components (Fe, Ni, Cu, Zn, Pb) vary from one manuscript to another, so that they can be used to characterize different periods. The idea is then to characterize the ink of the undated propositions in the loose sheets, and attribute them to the different periods on the basis of composition matches to the dated documents. Even when an absolute date cannot be established, an important objective is a relative dating, i.e. differentiation between the inks of the various undated documents. Indeed, most of the controversies concerned whether certain notes had been written at the same or at different times. Using these criteria, both absolute and relative dating have been possible for a number of these loose sheets and controversies between historians of science have been solved.[62,63] It is important to keep in mind that this contribution is to be considered in the frame of a global approach where the historical and textual evidence is of course of primary importance. This calls for the need for integrated teams of researchers from the various fields to work together throughout all the phases of the investigation.

From a technical point of view, PIXE analysis of the inks in documents is easily performed with external beams. The typical beam size used is between 200 μm and 1 mm, with no need for microbeam set-ups. Because the elements of interest are not trace elements in the inks, but are major or minor components, the beam currents may be kept very low (typically 100–200 pA) and no damage occurs. In extracting quantitative data, two circumstances must be considered. First, the beam will also excite X-rays from the paper (or parchment) substrate, so that the result refers to an effective target that is a combination of ink and substrate. This is taken into account by measurements on spots of uncovered substrate, and appropriate subtraction of the backing contribution, after beam charge normalization of the

ink plus substrate and substrate-only spectra. Second, the ink actually penetrates the substrate, to an extent depending on many factors such as substrate porosity, ink fluidity at the very instant of writing, and even ageing phenomena. These effects can lead to ambiguities in that the target thickness and the matrix composition of the effective target are to some extent uncertain. However, as far as the ratio of the elements of interest (medium or heavy metals) is concerned, it can be shown that the value deduced with no matrix correction is only weakly dependent on these parameters. Therefore, the accuracy of the thin target approximation (sections 2.2.1 and 2.2.2) remains sufficient for a meaningful ink discrimination (about 10% for the measured ratios). The differences between different inks are much larger than that, even with changes of orders of magnitude having been observed.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
AMS	Accelerator Mass Spectroscopy
BS	Bremsstrahlung
EDS	Energy-dispersive Spectroscopy
EPMA	Electron Probe Microanalysis
FA	Factor Analysis
FET	Field Effect Transistor
fwhm	Full Width at Half-maximum
IBA	Ion Beam Analysis
LAMS	Laser Ablation Mass Spectroscopy
MDL	Minimum Detection Limit
NRA	Nuclear Reaction Analysis
PCA	Principal Component Analysis
PIGE	Particle-induced Gamma Ray Emission
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectroscopy
STIM	Scanning Transmission Ion Microscopy
WDS	Wavelength-dispersive Spectroscopy
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Proton-induced X-ray Emission in Environmental Analysis

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • Ion Scattering Spectroscopy in Analysis of Surfaces

Mass Spectrometry **(Volume 13)**
Tandem Mass Spectrometry: Fundamentals and Instrumentation

Nuclear Methods **(Volume 14)**
Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Rutherford Backscattering Spectroscopy

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist (eds.), *Particle-induced X-Ray Emission Spectrometry (PIXE)*, John Wiley & Sons Inc., Chichester, 1995.
2. S.A.E. Johansson, J.L. Campbell, *PIXE: a Novel Technique for Elemental Analysis*, John Wiley & Sons Inc., Chichester, 1988.
3. G. Pappalardo, J. De Sanoit, A. Musumarra, C. Calvi, C. Marchetta, 'Feasibility Study of a Portable PIXE System Using a ^{210}Po Alpha Source', *Nucl. Instrum. Methods.*, **B109/110**, 214 (1996).
4. L. Pappalardo, 'A Portable PIXE System for the In-situ Characterisation of Black and Red Pigments in Neolithic, Copper Age and Bronze Age Pottery', *Nucl. Instrum. Methods.*, **B150**, 576 (1999).
5. W. Bambynek, B. Crausemann, R.W. Fink, H.U. Freund, H. Mark, C.D. Swift, R.E. Price, P. Venugopalo Rao, 'X-ray Fluorescence Yields, Auger, and Coster-Kronig Transition Probabilities', *Rev. Mod. Phys.*, **44**, 716 (1972).
6. M.H. Chen, B. Crasemann, H. Mark, 'Relativistic K-shell Auger Rates, Level Widths, and Fluorescence Yields', *Phys. Rev.*, **A21**, 436 (1980).
7. M.O. Krause, 'Atomic Radiative and Radiationless Yields for K and L Shells', *J. Phys. Chem. Ref. Data*, **8**, 307 (1979).

8. J.H. Hubbel, National Institute of Standards and Technology Report NISTIR 89-4144, U.S. Dept. of Commerce, Gaithersburg, MD, 1989.
9. G. Lapicki, 'Cross-sections for K-shell X-ray Production by Hydrogen and Helium Ions in Elements from Beryllium to Uranium', *J. Phys. Chem. Ref. Data*, **18**, 111 (1989).
10. H. Paul, J. Sacher, 'Fitted Empirical Reference Cross Sections for K-shell Ionization by Protons', *At. Data Nucl. Data Tables*, **42**, 105 (1989).
11. M.H. Chen, B. Crasemann, 'Relativistic Cross Sections for Atomic K- and L-shell Ionisation by Protons, Calculated from a Dirac–Hartree–Slater Model', *At. Data Nucl. Data Tables*, **32**, 217 (1985).
12. M.H. Chen, B. Crasemann, 'Atomic K-, L- and M-shell Cross Sections for Ionization by Protons: a Relativistic Hartree–Slater Calculation', *At. Data Nucl. Data Tables*, **41**, 257 (1989).
13. J.L. Campbell, W.J. Teesdale, J.-X. Wang, 'Accuracy of Thick-target Micro-PIXE Analysis', *Nucl. Instrum. Methods.*, **B50**, 189 (1990).
14. W. Brandt, G. Lapicki, 'L-shell Coulomb Ionization by Heavy Charged Particles', *Phys.Rev.*, **A20**, 465 (1979).
15. W. Brandt, G. Lapicki, 'Energy-loss Effect in Inner-shell Coulomb Ionization by Heavy Charged Particles', *Phys.Rev.*, **A23**, 1717 (1981).
16. K.R. Akselsson, S.A.E. Johansson, 'X-ray Production by 1.5–11 MeV Protons', *Z. Phys.*, **266**, 245 (1974).
17. S.A.E. Johansson, T.B. Johansson, 'Analytical Applications of Particle-induced X-ray Emission', *Nucl. Instrum. Methods.*, **137**, 473 (1976).
18. D.D. Cohen, 'K and L Shell X-ray Cross Sections for Use in PIXE Analysis Systems', *Nucl. Instrum. Methods.*, **B49**, 1 (1990).
19. D.D. Cohen, M. Harrigan, 'K- and L-shell Ionisation Cross Sections for Protons and Helium Ions Calculated in the ECPSSR Theory', *At. Data Nucl. Data Tables*, **33**, 255 (1985).
20. F. Folkmann, C. Gaarde, T. Huus, K. Kemp, *Nucl. Instrum. Methods.*, **116**, 487 (1974).
21. K. Ishii, S. Morita, 'Continuous Backgrounds in PIXE', *Int. J. PIXE*, **1**(1), 1 (1990).
22. S.A.E. Johansson, 'Optimization of the Sensitivity in PIXE Analysis', *Int. J. PIXE*, **2**(1), 33 (1992).
23. E.T. Williams, 'PIXE Analysis with External Beams: Systems and Applications', *Nucl. Instrum. Methods.*, **B3**, 211 (1984).
24. B.L. Doyle, D.S. Walsh, S.R. Lee, 'External Micro-ion-beam Analysis (X-MIBA)', *Nucl. Instrum. Methods.*, **B54**, 244 (1991).
25. P.A.Mandò, 'Advantages and Limitations of External Beams in Applications to Arts and Archaeology, Geology and Environmental Problems', *Nucl. Instrum. Methods.*, **B85**, 815 (1994).
26. T. Calligaro, J.-C. Dran, B. Moignard, J. Salomon, 'Development of an External Beam Nuclear Microprobe on the AGLAE Facility of the Louvre Museum', *Nucl. Instr. Meth.*, **B161–163**, 328 (2000).
27. J.L. Campbell, 'Instrumentation, Fundamentals, and Quantification – Extraction of X-ray Intensities from PIXE Spectra', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 2.6, 1995.
28. J.L. Campbell, 'Instrumentation, Fundamentals, and Quantification – Detector Lineshape and Escape Peaks', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 2.4.3, 1995.
29. F. Watt, G.W. Grime (eds.), *Principles and Applications of High Energy Ion Microbeams*, Hilger, Bristol, UK, 1987.
30. G.W. Grime, F. Watt, *Beam Optics of Quadrupole Probe Forming Systems*, Hilger, Bristol, UK, 1984.
31. F. Watt, G.W. Grime, 'The High-energy Ion Microprobe', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 3, 1995.
32. J. Miranda, 'Studies of Atmospheric Aerosols in Large Urban Areas Using PIXE: an Overview', *Nucl. Instrum. Methods.*, **B109/110**, 439 (1996).
33. I. Orlic, X. Wen, T.H. Ng, S.M. Tang, 'Two Years of Aerosol Pollution Monitoring in Singapore: a Review', *Nucl. Instrum. Methods.*, **B150**, 457 (1999).
34. C. Boni, A. Caridi, E. Cereda, G.M.B. Marcazzan, 'A PIXE/PIGE Setup for the Analysis of Thin Samples', *Nucl. Instrum. Methods.*, **B47**, 133 (1990).
35. F. Calastrini, P. Del Carmine, F. Lucarelli, P.A. Mandò, P. Prati, A. Zucchiatti, 'External Beam PIGE for Fluorine Determination in Atmospheric Aerosol', *Nucl. Instrum. Methods.*, **B136/138**, 975 (1998).
36. G.D. Thurston, J.D. Spengler, 'A Quantitative Assessment of Source Contributions to Inhalable Particulate Matter Pollution in Metropolitan Boston', *Atmos. Environ.*, **19**(1), 9 (1985).
37. E. Swietlicki, S. Puri, H.C. Hansson, 'Urban Air Pollution Source Apportionment Using a Combination of Aerosol and Gas Monitoring Techniques', *Atmos. Environ.*, **30**(15), 2795 (1996).
38. K.G. Malmqvist (ed.), *Proceedings of the Eighth International Conference on PIXE and its Analytical Applications*, Lund (Sweden), June 1998, *Nucl. Instrum. Methods.*, **B150**, 1999.
39. E. Koltay, 'Elemental Analysis of Atmospheric Aerosols: Results and Perspectives of the PIXE Technique', *Int. J. PIXE*, **1**(2), 93 (1990).
40. T.A. Cahill, 'Compositional Analysis of Atmospheric Aerosols', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 5, 1995.

41. K.G. Malmqvist, 'Quantitative PIXE Analysis of Biomedical Material – Sample Preparation, Irradiation and Quality Control', *Nucl. Instrum. Methods.*, **B49**, 184 (1990).
42. K.G. Malmqvist, 'Biological and Medical Applications', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 4, 1995.
43. K. Ishii, H. Orihara, S. Morita (eds.), *Proceedings of the International Symposium on Bio-PIXE*, Sendai (Japan), July 1992, in *Int. J. PIXE*, **2**(3,4), 1992.
44. Z. Zengh, Z. Hu, L. Wei (eds.), *Proceedings of the International Symposium on Bio-PIXE*, Beijing (China), August 1996, in *Int. J. PIXE*, **6**(1,2), 1996.
45. J.D. MacArthur, Xin-Pei Ma, 'A Review of Particle-induced X-ray Emission in Geology', *Int. J. PIXE*, **1**(4), 311 (1991).
46. J.L. Campbell, 'Applications in Earth Sciences', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 6, 1995.
47. K.G. Malmqvist, 'Applications in Art and Archaeology', in *Particle-induced X-Ray Emission Spectrometry (PIXE)*, eds. S.A.E. Johansson, J.L. Campbell, K.G. Malmqvist, John Wiley & Sons Inc., Chichester, Chapter 7, 1995.
48. M.A. Respaldiza, J. Gómez Camacho (eds.), *Applications of Ion Beam Analysis to Arts and Archaeometry*, Secretariado de Publicaciones de la Universidad de Sevilla, 1997.
49. Ch. Lahanier, G. Amsel, Ch. Heitz, M. Menu, H.H. Andresen (eds.), *Proceedings of the International Workshop on Ion Beam Analysis in the Arts and Archaeology*, Pont-à-Mousson (France), February 1985, *Nucl. Instrum. Methods.*, **B14**, 1 (1986).
50. G. Amsel, M. Menu, J. Moulin, J. Salomon, 'The 2 MV Tandem Pelletron Accelerator of The Louvre Museum', *Nucl. Instrum. Methods.*, **B45**, 296 (1990).
51. M. Menu, T. Calligaro, J. Salomon, G. Amsel, J. Moulin, 'The Dedicated Accelerator-based IBA Facility AGLAE at the Louvre', *Nucl. Instrum. Methods.*, **B45**, 610 (1990).
52. P. Del Carmine, M. Grange, F. Lucarelli, P.A. Mandò, 'Particle-induced X-ray Emission with an External Beam: a Non-destructive Technique for Material Analysis in the Study of Ancient Manuscripts', in *Studi e Testi*, eds. M. Maniaci, P.F. Munafò, Biblioteca Apostolica Vaticana, Città del Vaticano, 7–27, Vol. 358, 1993.
53. M. Bernasconi, R. Cambria, L. Dal Poz, P. Del Carmine, M. Grange, F. Lucarelli, J.D. MacArthur, P.A. Mandò, 'Analyse des couleurs dans un groupe de manuscrits enluminés du XII au XV siècle avec l'emploi de la technique PIXE', in *Studi e Testi*, eds. M. Maniaci, P.F. Munafò, Biblioteca Apostolica Vaticana, Città del Vaticano, 57–101, Vol. 358, 1993.
54. C. Neelmeijer, W. Wagner, H.P. Schramm, U. Thiel, 'De re metallica (G.Agricola) – IBA on air', *Nucl. Instrum. Methods.*, **B99**, 390 (1995).
55. G. Demortier, 'Microbeam Applications in Archaeology', in *Principles and Applications of High Energy Ion Microbeams*, eds. F. Watt, G.W. Grime, Hilger, Bristol, UK, 1987.
56. G. Demortier, 'Ion Beam Analysis of Gold Jewelry', *Nucl. Instrum. Methods.*, **B64**, 481 (1992).
57. G. Demortier, 'Ion Beam Studies of Archaeological Gold Jewellery Items', *Nucl. Instrum. Methods.*, **B113**, 347 (1996).
58. P. Duerden, E. Clayton, J.R. Bird, D.D. Cohen, 'Recent Ion Beam Studies on Archaeology and Art', *Nucl. Instrum. Methods.*, **B14**, 50 (1986).
59. B.H. Kusko, T.A. Cahill, R.A. Eldred, R.N. Schwab, 'Proton Milliprobe Analyses of the Gutenberg Bible', *Nucl. Instrum. Methods.*, **B3**, 689 (1984).
60. T.A. Cahill, R.N. Schwab, B.H. Kusko, R.A. Eldred, G. Möller, D. Dutschke, D.L. Wick, A.S. Pooley, 'The Vinland map, Revisited: New Compositional Evidence on Its Inks and Parchment', *Anal. Chem.*, **59**, 829 (1987).
61. L. Giuntini, F. Lucarelli, P.A. Mandò, W. Hooper, P.H. Barker, 'Galileo's Writings: Chronology by PIXE', *Nucl. Instrum. Methods.*, **B95**, 389 (1995).
62. F. Lucarelli, P.A. Mandò, 'Recent Applications to the Study of Ancient Inks with the Florence External-PIXE Facility', *Nucl. Instrum. Methods.*, **B109–110**, 644 (1996).
63. P. Del Carmine, L. Giuntini, W. Hooper, F. Lucarelli, P.A. Mandò, 'Further Results from PIXE Analysis of Inks in Galileo's Notes on Motion', *Nucl. Instrum. Methods.*, **B113**, 354 (1996).

Prompt γ-Neutron Activation Analysis

Carlos Oliveira, José Salgado, and Frederico G. Carvalho
ITN-Nuclear and Technological Institute, Sacavém, Portugal

1 Introduction **1**
2 Theory and Applicability of the Method **2**
2.1 Physical Principles 2
2.2 Detectability and Sensitivity 5
2.3 Prompt Versus Delayed Neutron Activation Analysis 6
3 Equipment **7**
3.1 General 7
3.2 Neutron Sources 7
3.3 Shielding 9
3.4 Detectors 11
3.5 Signal Processing and Data Handling 11
4 Instrument Design and Calibration **13**
4.1 Experimental Calibration 13
4.2 k_0 Method 13
4.3 Monte Carlo Library Least Squares Technique 14
4.4 Monte Carlo Simulation of Calibration Curves 14
4.5 Design Optimization 15
5 Applications of the Method **15**
5.1 Introduction 15
5.2 Borehole Logging 15
5.3 Industrial On-line Applications 16
5.4 Medical In Vivo Applications 17
5.5 Determination of Environmental Contaminants 19
5.6 Explosives and Drugs Detection Systems 19
5.7 Other Applications 19
6 Present Trends of Development **20**
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **21**

Prompt γ-neutron activation analysis (PGNAA) is a multielemental analytical technique based on the detection of γ-rays emitted by excited nuclei following neutron radioactive capture or neutron inelastic scattering. The accumulated count rate relative to the γ-ray energy characteristic of a certain isotope depends on its concentration. Measurements take place during sample irradiation. The technique can provide accurate information on the elemental composition of materials in a broad range of cases: raw materials, mineral exploration, contaminants, detection of forbidden products, such as drugs and explosives, inspection of containers and applications in health care. Among the main advantages of the PGNAA technique are: faster results; nondestructiveness; flexible sample requirements; bulk sample analysis; in situ application by direct on-line measurement; detectable unique signatures for most elements; improved accuracy for short-lived elements; low residual activity. On the other hand, the most significant limitations are: higher detection limits (DLs) as compared with delayed γ-neutron activation analysis (DGNAA); longer irradiation times; single (one-cycle) sample irradiation; impracticability of enhancing a specific isotope contribution by radiochemical or physical (half-life) separation techniques; requirement for relatively larger samples.

1 INTRODUCTION

PGNAA is a powerful analytical technique that can provide accurate information on the elemental composition of materials in a broad range of cases: raw materials, mineral exploration, contaminants, detection of forbidden products, such as drugs and explosives, inspection of containers (detection of plutonium) and applications in health care.

When a material is bombarded with neutrons, nuclear interactions in the sample may lead to the formation of excited compound nuclei that lose their excitation energy by the emission of one or more γ-rays (prompt γ-rays) and eventually one or more particles. PGNAA is based on the detection of the emitted γ-rays. The γ-ray energy identifies the element while the intensity of the line is a measure of the element concentration.

The analysis is the result of a combination of a number of different processes or steps:

- neutron emission by a neutron source (radioisotope, neutron generator, nuclear reactor, spallation source);
- neutron interaction with the sample's constituent nuclei;
- emission of γ-rays following neutron capture or neutron inelastic scattering;
- γ-ray attenuation in the material;

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

- detection of γ-rays (solid state or scintillation detector);
- conversion of detector counts in percent weight for the elements being analyzed based on a previous calibration of the system.

These steps and processes will be addressed in the following sections.

The PGNAA technique is best suited for measuring elements: (1) with very high neutron capture cross-section; (2) elements that decay too rapidly to be measured by conventional DGNAA; and (3) elements that through neutron reactions generate only stable isotopes or radioactive isotopes with a low probability of γ-decay.

The main advantages of PGNAA are: faster results; nondestructiveness (sample preparation is not needed); flexible sample requirements; bulk sample analysis (representative average for nonhomogeneous materials); in situ application by direct on-line measurement; detectable unique signatures for most elements; improved accuracy for short-lived elements; low residual sample activity.

On the other hand, the most significant handicaps are: higher DLs as compared with DGNAA, which restricts the number of elements that can be measured at trace concentrations; longer irradiation times; single (one-cycle) sample irradiation; impracticability of radiochemical or physical (half-life) separation to enhance specific isotope contributions; requirement for relatively larger samples.

The gauge response of a PGNAA analyzer is determined in a complex manner by the system parameters – geometry, source activity and neutron energy spectrum, and detector efficiency – and by certain properties of the analyzed sample such as bulk density and water content. It follows that the count rate caused by a given γ-ray line is a nonlinear function of the concentration of the element to be measured. There are, in principle, several methods for calibrating a PGNAA gauge. Experimental calibration methods use standards of known composition. An alternative method is the use of Monte Carlo codes to simulate the sample spectra and/or to establish calibration curves.

Trends in the evolution of PGNAA relate to both the instruments of the technique and the range of its applications.

The use of advanced neutron sources, improved spectrometers (Compton suppression and pair spectrometer), nuclear imaging techniques, sophisticated calibration methods and the extension of PGNAA to new materials and sample types are some of the trends which indicate that this technique has a promising future.

2 THEORY AND APPLICABILITY OF THE METHOD

Neutrons are electrically neutral; they are not affected by atomic electrons or by the positive charge of the nucleus. Neutrons can pass through the atomic cloud and interact directly with the nucleus; they are highly penetrating in most materials. On the other hand, the attenuation coefficient of γ-rays for most elements decreases monotonically with energy in the range from 0.1 to 10 MeV so that γ-rays with energy > 2 MeV are also relatively penetrating. Thus neutron–gamma based techniques do not suffer from the limitations of penetration of X-rays and β-rays and are favored for investigating bulky materials. Sample preparation is not required. All the major elements can be assayed provided a high-resolution detector is used.

Neutron activation analysis (NAA) is an important elemental analytical technique with many applications in different fields: industry, medicine, resource exploration, forensic work, and so on.

2.1 Physical Principles

When a material is bombarded with neutrons, a neutron can be captured by a nucleus, forming a compound nucleus. The compound nucleus remains for a certain time in a highly excited state due to the addition of the binding energy of the neutron to the nucleus. It will lose this energy by the emission of one or more γ-rays (prompt γ-rays) and eventually one or more particles,

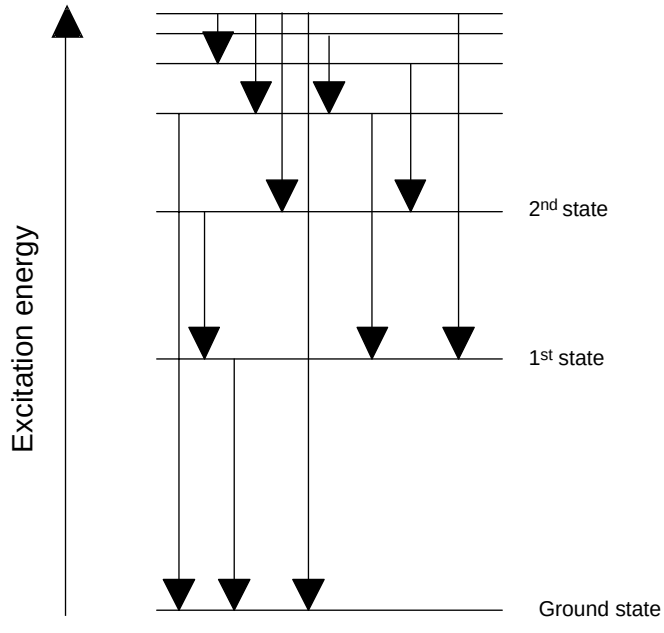

Figure 1 Decay scheme of an excited nucleus.

namely one neutron (inelastic scattering), which can be accompanied by several γ-rays. After the emission of this primary radiation, the daughter nucleus can remain unstable (radioactive), decaying to the ground state by a β-decay process, followed by γ-emission (delayed γ-rays).

As shown in Figure 1, the de-excitation of the compound nucleus may take the form of a stepwise process, which can be very complex, originating a spectrum with a large number of lines (89 for Si, 449 for Cl, 113 for Ca and 187 for Fe). Most of these lines have a very small relative intensity. The large number of emitted photons is due to (1) the high binding energy of the compound nucleus and (2) the narrowing of the energy gap between levels with increasing excitation energy. Two other causes of the increase in the number of lines in a spectrum must also be taken into account: (1) every isotope has its own decay scheme and (2) the high-energy γ-photons may interact through pair production in the detector. The subsequent positron annihilation originates the emission of two 0.511 MeV γ-rays. One or both of these γ-rays may escape the detector. This means that every γ-line may be split in three lines (the full f, with energy E_γ; the single escape s, with energy $E_\gamma - 0.511$ MeV; and the double escape d, with energy $E_\gamma - 1.022$ MeV. The γ-ray spectrum is a signature of the sample elemental composition. The intensity of the lines is proportional to the weight content of the respective element.

Prompt γ-rays are emitted within 10^{-15}–10^{-14} s of the interaction; delayed γ-rays originate in the decay of induced radioactivity. Two types of activation analysis can, then, be used: PGNAA when information is derived from prompt γ-rays and conventional instrumental neutron activation analysis (INAA) or DGNAA, where information is derived from delayed γ-rays (Figure 2). In the second case, the irradiated samples are normally removed from the irradiation position to detect the delayed γ-rays.

PGNAA is based on the analysis of a few γ-lines for each element. The selection of these lines takes into account their intensity and interference from other lines.

Two physical mechanisms predominate in PGNAA: thermal neutron capture reactions (n,γ) and fast neutron inelastic scattering (n,n′γ).

Thermal neutron capture	${}^{A}_{Z}\mathrm{X}(\mathrm{n},\gamma){}^{A+1}_{Z}\mathrm{X}^{*}$	${}^{27}_{13}\mathrm{Al}(\mathrm{n},\gamma){}^{28}_{13}\mathrm{Al}$ $E_\gamma = 7.724$ MeV

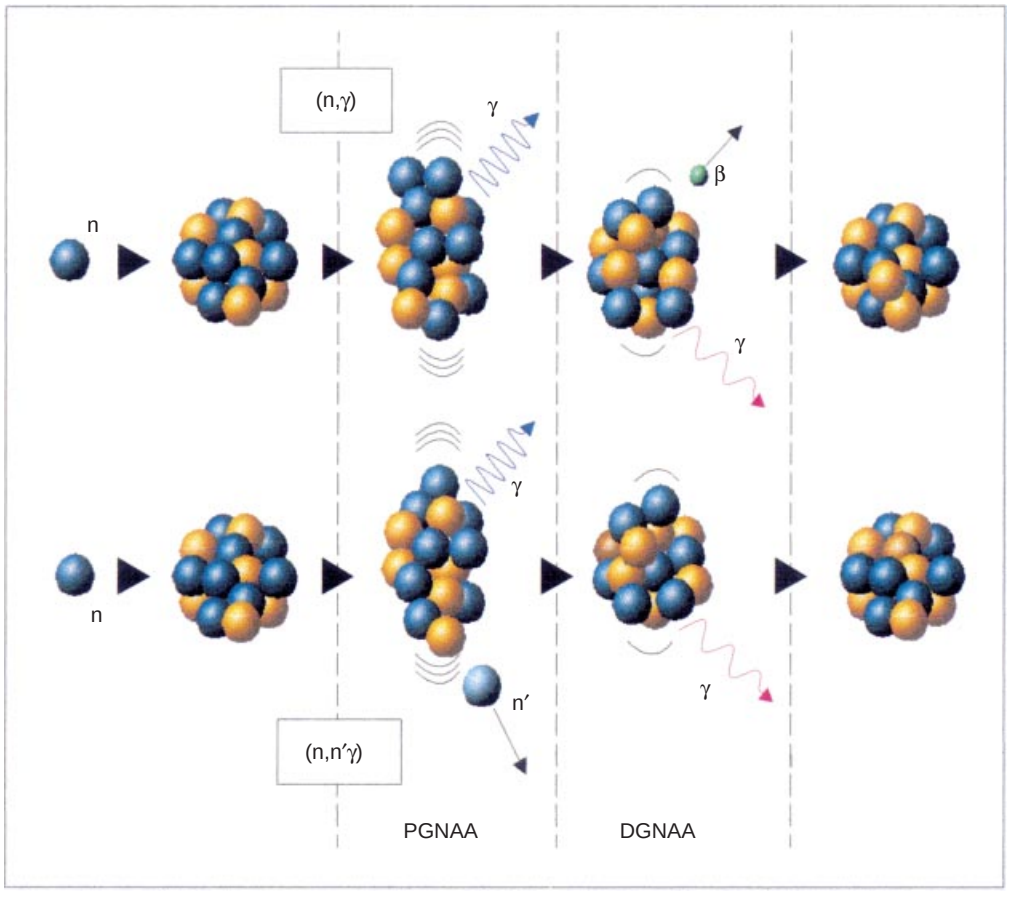

Figure 2 Basic processes in PGNAA and DGNAA.

Table 1 Microscopic cross-section of selected nuclei for thermal neutron capture and 14 MeV inelastic neutron scattering reactions and prompt γ-ray energies[a]

Reaction	$\sigma_{c,th}$ (barn)	$\sigma_{i,14\,MeV}$ (barn)	Dominant line energies E_γ (MeV)
$^{1}_{1}H(n,\gamma)^{2}_{1}H$	0.32	–	2.223
$^{12}_{6}C(n,\gamma)^{13}_{6}C$	0.0034	–	4.945, 3.684, 1.261
$^{12}_{6}C(n,n'\gamma)^{12}_{6}C$	–	0.425	4.439
$^{16}_{8}O(n,n'\gamma)^{16}_{8}O$	–	0.474	6.128, 6.049, 7.117
$^{14}_{7}N(n,\gamma)^{15}_{7}N$	0.075	–	10.828, 6.321, 5.532
$^{32}_{16}S(n,\gamma)^{33}_{16}S$	0.491		0.841, 5.420, 2.380
$^{32}_{16}S(n,n'\gamma)^{32}_{16}S$		0.500	2.230, 4.282
$^{27}_{13}Al(n,\gamma)^{28}_{13}Al$	0.232	–	7.724
$^{56}_{26}Fe(n,\gamma)^{57}_{26}Fe$	2.26	–	7.646, 7.632
$^{56}_{26}Fe(n,n'\gamma)^{56}_{26}Fe$		0.749	0.847, 1.238
$^{40}_{20}Ca(n,\gamma)^{41}_{20}Ca$	0.43		1.943, 6.420
$^{40}_{20}Ca(n,n'\gamma)^{40}_{20}Ca$	–	0.243	3.736, 3.904

$1\text{barn} = 10^{-28}\,\text{m}^2$.
[a] Reproduced from Mikesell et al.[1] with permission from IAEA.

Inelastic scattering $\quad {}^{A}_{Z}X(n, n'\gamma)^{A}_{Z}X^{*} \quad {}^{208}_{82}Pb(n,n'\gamma)^{208}_{82}Pb$

$$E_\gamma = 2.615\,\text{MeV}$$

For every nucleus there is a definite probability (cross-section) of thermal neutron capture ($\sigma_{c,th}$) and/or inelastic scattering ($\sigma_{i,14\,MeV}$), as is shown in Table 1.

The larger the cross-section the larger the number of nuclear reactions for a given isotopic abundance in the sample and a given neutron flux and, generally, the higher the sensitivity of the method.

For a homogeneous sample the count rate, R_{ij}, in γ-peak i from element j is given by the expression in Equation (1):

$$R_{ij} = \frac{N_A w_j}{4\pi M_j}\rho I_{ij}\varepsilon_i \int_{E_n}\int_{V}\Phi(E_n,\vec{r})\sigma_j(E_n) \times \int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dE_n\,dV \quad (1)$$

where N_A is the Avogadro number; w_j, the weight content of element j; M_j, the atomic mass; ρ, the bulk density ($m_j = w_j \times \rho$ is the mass of element j); I_{ij}, the number of photons pertaining to element j, with energy E_i, emitted per neutron captured; ε_i the detector efficiency for photons of energy E_i; $\Phi(E_n, \vec{r})$, the energy-dependent neutron flux, at position $\vec{r}$; $\sigma_j(E_n)$, the average element microscopic cross-section for the nuclear reaction that originates the γ-line of interest; $\vec{r}_d$, the path in the medium along the line drawn through volume element dV and the detection position; $\mu(E_i)$, the mass attenuation coefficient; and S, the detector surface.

When the energy dependence of the neutron flux inside the medium does not vary with position, the flux dependence on E_n and r can be separated. This happens, for instance, throughout the volume of a moderating material such as coal[2,3] when using ^{241}Am-Be, ^{252}Cf or 14 MeV neutron sources. Also the thermal and fast neutron fluxes can be formally separated and the neutron flux can be written as Equation (2):

$$\Phi(E_n,\vec{r}) = \Phi^0_{th}\Phi_{th}(E_n)\Phi_{th}(\vec{r}) + \Phi^0_F\Phi_F(E_n)\Phi_F(\vec{r}) \quad (2)$$

where th and F refer, respectively, to the thermal and the fast neutron fluxes. The energy dependence of the thermal component can be approximated by a Maxwell distribution and this will be indicated in the following formula by substituting M for th in $\Phi_{th}(E_n)$.

Equation (1) can thus be rewritten as Equation (3):

$$R^{I,C}_{ij} = \frac{N_A m_j}{4\pi M_j} I_{ij}\varepsilon_i\Phi^0_{th,F}\int_{E_n}\Phi_{M,F}(E_n)\sigma_j(E_n)\,dE_n \times \int_{V}\Phi_{th,F}(\vec{r})\int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dV \quad (3)$$

where the superscripts I and C refer to (n,n′γ) and (n,γ) reactions, respectively. Equation (3) gives the count rate, R^I_{ij}, arising from the fast component of the neutron flux or gives the count rate, R^C_{ij}, coming from the thermal component of the neutron flux. For most elements the thermal neutron capture predominates. In fact, in most applications the information derived from the detector count rate originates in thermal neutron reactions (thermal PGNAA); in certain special cases, however, fast neutrons are used to obtain information on the sample composition (fast PGNAA). Taking into account that the Maxwellian distribution has the expression in Equation (4):

$$\Phi_M(E_n) = \frac{1}{(kT)^2}E\exp\left\{-\frac{E}{kT}\right\} \quad (4)$$

where k is the Boltzmann constant and T the neutron temperature, Equation (3) takes the form, Equation (5):

$$R^C_{ij} = \frac{N_A m_j}{4\pi M_j} I_{ij}\varepsilon_i\Phi^0_{th}\int_{E_n}\frac{1}{(kT)^2}E_n\exp\left\{-\frac{E_n}{kT}\right\}\sigma_j(E_n)\,dE_n \times \int_{V}\Phi_1(\vec{r})\int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dV = \frac{N_A m_j}{4\pi}\frac{\sigma_{j,0}I_{ij}}{M_j}\varepsilon_i\Phi^0_{th}\int_{V}\Phi_1(\vec{r})\int_{s}\frac{\exp\{-\mu(E_i)r_d\}}{r_d^2}\,d\vec{r}_d\,dV \quad (5)$$

where $\sigma_{j,0}$ is the element averaged $2200\,\text{m s}^{-1}$ neutron capture cross-section.

This equation shows that the count rate from a given line is proportional to the sample content of element j. However when a perturbation of the fast neutron flux is

caused by the sample, as happens currently in applications using isotopic neutron sources and large samples, the volume integral depends on the sample bulk density and composition, particularly on the water content. Fast neutrons emitted by the neutron source are generally moderated inside the sample. An increase of the bulk density and/or of the water content (or hydrogen content), will increase the thermal neutron flux and, consequently, the reaction rate. On the other hand, increasing the bulk density will increase the attenuation of photons inside the sample. The two effects have, thus, opposite signs and can cancel each other. An optimized design of the measuring arrangement (geometry, proper choice of sample size, amount of external thermalization) can make the volume integral essentially independent of sample density and composition.

In certain cases it is important to take into account the contribution of the epithermal neutrons to the detector count rate. To this end the thermal neutron cross-section can be substituted by the quantity shown in Equation (6):

$$g\sigma_{j,0} + I'_0 \frac{\Phi_{\text{epi}}}{\Phi_{\text{th}}} \tag{6}$$

where Φ_{epi} is the epithermal flux, g is the Wescott constant and I'_0 the resonance integral, Equation (7):

$$I'_0 = \int_{E_{\text{th}}}^{\infty} \sigma(E) \frac{\mathrm{d}E}{E} \tag{7}$$

2.2 Detectability and Sensitivity

In Equation (5) the quantity $S_{\text{f}} = I\sigma/M$ is constant for a given γ-ray pertaining to a given element. S_{f} is called *sensitivity factor* or *index* and is a useful measure of the relative sensitivity of the method for different elements under the same experimental conditions. Table 2 shows selected sensitivity factors of different elements when using gamma energies above 2.223 MeV.

Examination of the data shows that certain elements can be very easily detected by PGNAA. These include H, B, Cl, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, Gd, Dy, Au and Hg.

For thermal PGNAA to be appropriate for the detection of an element it is not sufficient that the element has a high sensitivity factor. It is important that the useful gamma lines of the element show a high spectral contrast,

Table 2 Sensitivity factor of various elements for thermal PGNAA

Element	Cross-section σ (barn)	Atomic mass, M	Energy (MeV)	Intensity, I (γ/n)	Sensitivity
H	0.332	1.008	2.223	1.00	0.329
B[a]	0.103	10.811	4.443	0.7596	0.00724
C	0.0034	12.011	4.945	0.6764	0.00019
N	0.0747	14.007	10.829	0.1412	0.00075
Na	0.400	22.990	6.395	0.2218	0.00386
Al	0.230	26.981	7.724	0.2743	0.00234
Si	0.160	28.086	4.934	0.6269	0.00357
Cl	33.2	35.453	6.111	0.2000	0.1873
K	2.10	39.102	7.770	0.0670	0.00360
Ca	0.430	40.08	6.420	0.3889	0.00417
Sc	27.2	44.96	8.175	0.1183	0.0716
Ti	6.10	47.90	6.760	0.2417	0.0308
V	5.04	50.94	7.163	0.1314	0.0130
Cr	3.10	52.00	8.884	0.2697	0.0161
Mn	13.3	54.94	7.244	0.1213	0.0294
Fe	2.55	55.85	7.645	0.2412	0.0110
			7.631	0.2851	0.0130
Co	37.2	58.93	6.876	0.0821	0.0518
Ni	4.43	58.71	8.999	0.3774	0.0284
Cu	3.79	63.54	7.915	0.3082	0.0184
Zn	1.10	65.37	7.863	0.1058	0.00178
Ag	63.6	107.87	5.698	0.0116	0.00683
Cd	2450	112.40	5.824	0.0213	0.464
Gd	4.9×10^4	157.25	6.750	0.0225	70.1
Dy	930	162.50	5.607	0.0293	0.168
Au	98.8	196.97	6.251	0.0552	0.0278
Hg	376	200.59	5.966	0.1386	0.2598
Pb	0.17	207.19	7.368	0.9405	0.00077

[a] In the case of B, the sensitivity can be enhanced (approx. 73) if the 0.48 MeV γ-ray from the reaction ^{10}B(n,α) is detected. This is usually only done when the low energy gamma background does not disturb the measurement.

i.e. a high ratio of the line intensity to that of neighboring lines pertaining to other elements.

The sensitivity factor gives qualitative information only: elements with high sensitivity factors will be more easily detected in a complex matrix. However, it is necessary to quantify the feasibility of measuring an element. Two quantities are normally considered: the absolute sensitivity (in counts $s^{-1}\,mg^{-1}$) and the DL (in $\mu g\,g^{-1}$).

The DL is related to the absolute sensitivity, S, through Equation (8):[4]

$$DL = 3.29\frac{\sqrt{R_b/t}}{S} \tag{8}$$

where R_b is the background count rate and t is the counting time.

Actual values of DL (or S) for a given arrangement and element are determined experimentally.

The two quantities, S and DL, depend on intrinsic sample characteristics related to nuclear properties of the elements (cross-section; number of γ-rays emitted and atomic mass; matrix interferences) and on parameters of the measuring arrangement (neutron spectrum and flux, sample dimensions, detector efficiency and resolution, background). For example, a reactor-based facility using a cold neutron source, guide tubes and a high-resolution germanium detector with Compton suppression, will normally show higher absolute sensitivity (lower DLs) than an arrangement using an isotopic neutron source for on-line measurement of large samples conveyed on a belt.

The actual DLs vary from one arrangement to another, depending on the neutron flux in the sample, the distance between sample and detector and the background due to shielding. The DL of an element can be affected by interferences from other elements present in the sample or in shielding or structural materials.

The elemental DLs for reactor-based PGNAA facilities are usually much lower than those of facilities with isotopic sources, used in on-line measurements.

As an example, Table 3 shows the DLs of the PGNAA facility installed at the Cold Neutron Research Facility of the National Institute of Standards and Technology (NIST).[5]

Table 3 Estimated spectral interference-free DLs for a 1 g sample counted for 24 h

Range ($\mu g\,g^{-1}$)	Element
0.01–0.1	B, Cd, Sm, Gd
0.1–1	Hg
1–10	H, Cl, In, Nd
10–100	Na, S, K, Ti, V, Cr, Mn, Co, Ni, Cu, As, Se, Br, Mo, Au
100–1000	Mg, Al, Si, P, Ca, Fe, Zn
1000–10 000	C, N, F, Pb

2.3 Prompt Versus Delayed Neutron Activation Analysis

As a laboratory analytical technique for routine sample analysis, PGNAA is generally complementary to the conventional delayed NAA systems installed at most research reactors. In contrast to DGNAA, the method is in principle applicable to all elements. As irradiation and measurement are carried out simultaneously, true multielement analysis is possible in a single measurement, at the expense of extremely complex γ-ray spectra.

DGNAA however is used intensively and is preferred for the determination of a large number of elements where no restrictions apply to handling and preparation of the sample. This is so because of the possibility of using higher neutron fluxes and of postirradiation counting that allow for background reduction and lower DLs.

In PGNAA arrangements installed at reactor facilities a wide variety of samples, e.g. liquids, solids, gases, museum objects, hazardous materials, can be irradiated using extracted beams that have fewer limitations on sample size and composition. In all cases neutron fluxes used in PGNAA are currently one or more orders of magnitude lower than those used in DGNAA.

The PGNAA technique is best suited for measuring elements: (1) with extremely high neutron capture cross-section; (2) elements which decay too rapidly to be measured by DGNAA; and (3) elements which produce only stable isotopes or radioactive isotopes with low gamma decay probability.

Sample preparation is not needed for PGNAA, and the results are obtained immediately. The irradiated samples may be large, which is advantageous when a representative average for nonhomogeneous material is needed. (The accuracy of the results may be affected by the presence of inhomogeneities that influence significantly the neutron flux distribution in the sample bulk and/or the position-dependent gamma attenuation.) However, large-sized samples may lead to problems due to neutron self-shielding, and specially due to neutron scattering, particularly in hydrogen.

The main advantages of the PGNAA technique are summarized in Table 4. Foremost among the advantages are nondestructiveness and the fact that nearly all elements yield a detectable signal. On the other hand, the most significant handicap is higher DLs, which restricts the number of elements that can be measured at trace concentrations and the long irradiation times often required.

Table 4 Advantages and disadvantages of PGNAA compared with DGNAA

Advantages	Faster results
	Nondestructiveness
	Low residual activity
	Flexible sample requirements
	Bulk sample analysis
	In situ application (on-line measurements)
	Detectable unique signatures for most elements
	Improved accuracy for short-lived elements
Disadvantages	Single (one-cycle) sample irradiation
	Longer irradiation time required
	Cannot use half-lives to enhance some particular isotope contribution
	Most elements not detectable at trace levels
	Relatively large sample required

While comparing different NAA techniques mention should be made of pulsed NAA, a technique that borrows some features from both PGNAA and for DGNAA. Pulsed neutron sources, e.g. neutron generators, are used so that between pulses the delayed gamma emission from the sample can be detected separately from the prompt gammas. This approach can lead to enhanced signal-to-background ratios in specific situations.

3 EQUIPMENT

3.1 General

According to the physical principle of PGNAA an operational arrangement must include a neutron source, a detection system with appropriate electronics and adequate shielding for radiological safety purposes. Generally, a dedicated computer forms the interface between the equipment and the user.

The specific objective for which the PGNAA device is designed (nature of the samples, elements to be measured, type of nuclear reactions, expected γ-ray spectrum) determines the type of source, the detector and the electronics as well as the manner in which they are assembled. An arrangement for on-line control of an industrial process in a factory or an analyzer installed at the end of a neutron guide tube have quite different designs.

3.2 Neutron Sources

The choice of the neutron source is determined not only by requirements concerning nuclear parameters such as neutron energy and flux, and half-life (when an isotopic source is considered), but also by the specific conditions regarding its utilization. Different kinds of sources are available: isotopic sources, neutron generators, nuclear reactors and spallation sources.

3.2.1 Isotopic Neutron Sources

There are two types of isotopic neutron source used in PGNAA: (α,n) sources and fission sources.

Sources of the first type are based in (α,n) reactions of isotopes that have large cross-sections for neutron production. Beryllium is an outstanding example. The sources consist of a mixture of beryllium and an α-emitter such as polonium or americium, the whole contained in a sealed capsule:

$$^{4}_{2}\mathrm{He} + ^{9}_{4}\mathrm{Be} \longrightarrow ^{1}_{0}\mathrm{n} + ^{12}_{6}\mathrm{C}$$

However, only a small number of α-particles (1 in 10^4) is effective in the production of neutrons so that the yield of the sources is relatively small. Also a background of γ-rays is present that is inconvenient for the application of the technique. Table 5 shows the main characteristics of several beryllium (α,n) sources.

Taking into account the neutron yield, the half-life of the α-emitter and the γ-background, the ^{241}Am isotope is currently preferred. There are ^{241}Am-Be sources commercially available up to 925 GBq.[6]

Table 5 Be (α,n) source characteristics

Nuclide	$T_{1/2}$	$E\alpha$ (MeV)	Yield ($n\,s^{-1}/10^6\,\alpha$)	Yield ($n\,s^{-1}\,Ci^{-1}$)	Yield ($n\,s^{-1}\,g^{-1}$)	Av. energy (MeV)	γ-dose at 1 m for $10^6\,n\,s^{-1}$ ($mGy\,h^{-1}$)
^{238}Pu	89 y	5.50	114	2.8×10^6	4.5×10^7	4.0	<0.01
^{239}Pu	24.4 y	5.14	65	1.6×10^6	1.2×10^5	4.59	<0.01
^{210}Po	138 d	5.30	73	2.5×10^6	1.1×10^{10}	4.54	<0.001
^{241}Am	458 y	5.48	82	2.2×10^6	6.5×10^8	4.46	0.01
^{244}Cm	18.1 y	5.79	100	3×10^6	2.5×10^8	4.31	<0.01
^{242}Cm	163 d	6.10	118	4×10^6	$\approx 10^{10}$	4.16	<0.01
^{226}Ra	1620 y	7.7 to 4.8	502	1.1×10^7	1.5×10^7	3.94	0.5
^{227}Ac	22 y	7.4 to 5.7	702	1.5×10^7	1.7×10^9	3.87	0.07

Spontaneous fission sources can be found in the region of the transuranium elements where the α-decay instability competes with spontaneous fission. However the half-lives of the transuranium elements are generally very short. The only nuclide with an acceptable half-life for use in applications is ^{252}Cf ($T_{1/2} = 2.645$ y). This nuclide decays by α emission in 97% of the cases and by spontaneous fission in the remaining 3%. The average number of neutrons emitted per fission is 3.76. A source containing 3.7×10^{10} Bq of ^{252}Cf has a neutron yield of 4.3×10^{9} n s^{-1}, which is three orders of magnitude higher than that of the more interesting (α,n) sources. The γ-dose at 1 m from a source emitting 10^{6} n s^{-1} is 0.0007 mGy h^{-1} (a factor of 10 smaller than for a ^{241}Am-Be source). The average neutron energy of a ^{252}Cf source is 2.3 MeV.

In Figure 3 neutron spectra and γ-spectra for Cf and Am-Be sources are plotted for comparison.[7]

The high specific activity of the Cf source is an advantage over the (α,n) sources. Other advantages are small size and low heat and gas generation, which decreases the risk of source damage for long-term operation.[8] These radioisotope sources are used in PGNAA devices designed for field work.

3.2.2 Neutron Generators

Sealed tube neutron generators[9,10] are based on the following reactions:

$$^{2}_{1}\mathrm{H} + {}^{3}_{1}\mathrm{H} \longrightarrow {}^{4}_{2}\mathrm{He}(3.6\,\mathrm{MeV}) + {}^{1}_{0}\mathrm{n}(14.1\,\mathrm{MeV})$$

$$^{2}_{1}\mathrm{H} + {}^{2}_{1}\mathrm{H} \longrightarrow {}^{3}_{2}\mathrm{He}(0.8\,\mathrm{MeV}) + {}^{1}_{0}\mathrm{n}(2.45\,\mathrm{MeV})$$

In generators emitting 14.1-MeV neutrons, a deuterium ion beam (or a mixture of deuterium and tritium ions) is accelerated up to between 70 and 300 kV and made to impinge on a tritiated target (or a deuterated and tritiated target). In the latter case 2.45-MeV neutrons are also produced. However, the emission of 14.1-MeV neutrons dominates. It is possible to obtain an emission rate of 10^{11} n s^{-1}.

Tubes emitting 2.45-MeV neutrons are based on deuterium–deuterium reactions, i.e. a beam of deuterium ions is accelerated up to a few hundred kilovolts and impinges on a deuterium target. Neutron emission of such tubes may reach 10^{10} n s^{-1}.

A great advantage of neutron generator tubes is the possibility of on/off switching, i.e. the production of neutrons is stopped whenever required. Another interesting characteristic is the possibility of pulsing the neutron beam.

The lifetime of a sealed tube is limited by target heating and target sputtering from ion beam impurities.

A new neutron generator source has been presented: the inertial electrostatic confinement (IEC) device. The

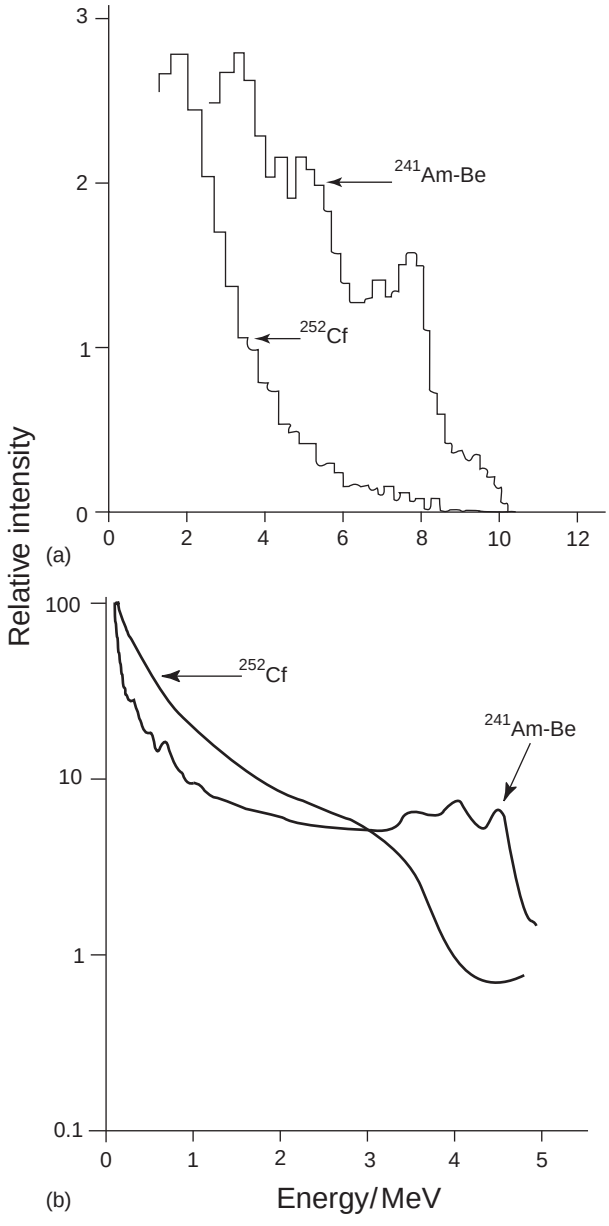

Figure 3 Neutron spectrum (a) and gamma spectrum (b) for ^{241}Am-Be and ^{252}Cf neutron sources. (Reproduced by permission from Amersham Buchler.)

IEC device is a neutron source where deuteron ions are accelerated producing fusion reactions as they react with a deuterium plasma target. Commercially available units deliver 10^{7} of 2.45-MeV n s^{-1} during steady-state operation. The neutron emission is isotropic. It is a practical low-to-medium-yield source for laboratory and industrial applications.[11]

3.2.3 Nuclear Reactors

Nuclear reactors are still the most important neutron sources. The thermal neutron flux of a reactor source

is of the order of 10^{13} n cm^{-2} s^{-1} up to 10^{15} n cm^{-2} s^{-1}. The neutron energy spectrum can be divided into three components: fast, epithermal and thermal. Their relative importance is dependent on the reactor structure. Usually, the thermal neutrons constitute the most intense fraction. Thermal neutrons are in thermal equilibrium with the atoms of the reactor moderator; their most probable energy equals kT (where k is Boltzmann's constant and T is the neutron temperature). At room temperature the most probable energy is 0.025 eV. Epithermal neutrons have energies from $5kT$ to 0.8 MeV. Fast neutrons have energies >0.8 MeV.

PGNAA facilities are usually installed at the extraction end of a neutron beam tube (sometimes using a neutron guide).[5,12–26]

The performance can be enhanced by using subthermal or "cold" neutrons.[12–14,16,19] These neutrons have energies lower than 10 meV. Cold neutrons are extracted from a cooled moderator placed in the neighborhood of the reactor core. The moderator may be liquid hydrogen or deuterium at about 30 K. Occasionally, liquid methane is also used at temperatures of about 100 K. The cold neutron absorption cross-section for most elements is about two to three times greater than that for thermal neutrons ($1/v$ behavior).

Owing to their long wavelength, about 0.3 nm, cold neutrons can be bent out of the flight path followed by higher energy radiations by using curved neutron guides. As a consequence, such cold neutron beams display a much lower γ and fast neutron background. Figure 4 shows the Cold Neutron Research Facility at NIST.

3.2.4 Neutron Spallation Sources

The operation of a neutron spallation source is, in principle, similar to that of an X-ray tube.[27,28]

Protons are accelerated to a high energy, in the range of hundreds of megaelectronvolts (typically 800 MeV up to 1 GeV) and directed on to a heavy metal target, such as uranium, tungsten, lead or a liquid metal (mercury). The result is excitation of the target nuclei. Their deexcitation is accomplished by the emission of several neutrons per incident proton (the number shows an energy dependence and may reach 30). Most of the neutrons produced have energies similar to those of a fission spectrum.

Neutron spallation sources can be continuous or pulsed. A typical value of the neutron flux from a continuous spallation source is 10^{13}–10^{14} n s^{-1} cm^{-2}.

Figure 5 shows the Paul Scherrer Institute spallation neutron source SINQ.[29]

3.3 Shielding

The shielding of a PGNAA facility must be designed according to the neutron source used. Shielding problems for on-line systems are completely different from those of an analyzer installed at the end of a neutron guide tube in a reactor hall. The shielding has two major functions: reducing the detector background so as to enhance the signal to background ratio and guaranteeing radiological safety.

Radiation coming directly from the neutron source as well as that originating in interactions within the sample and components of the system (structure, shielding) can be classified in four groups: (1) fast and epithermal neutrons, (2) thermal and cold neutrons, (3) primary γ-radiation and (4) secondary γ-radiation.

3.3.1 Fast and Epithermal Neutrons

The absorption cross-section of fast and epithermal neutrons is small for all materials. For this reason, it is necessary to slow down the neutrons before capture. The most effective shield for these neutrons is a mixture of a hydrogenous material and a thermal neutron absorber.

3.3.2 Thermal and Cold Neutrons

A few elements, such as boron, lithium, cadmium and gadolinium, have a high cross-section for thermal and cold neutrons. Compounds of these elements can be used to obtain an effective shielding against these neutrons. Neutron absorption in boron and cadmium is followed by the emission of γ-rays. This emission must be taken into account in the design of the shielding of the facility.

3.3.3 Primary γ-rays

Primary γ-rays are emitted directly by the neutron source. The problem is particularly acute in on-line analyzers using isotopic neutron sources. The source is generally surrounded by a layer of a strong γ-ray absorber such as lead or bismuth.

3.3.4 Secondary γ-rays

Secondary γ-rays originate in (n,γ) or (n,n′γ) reactions that take place within the sample or in shielding and structural materials. Only γ-rays from the sample directed towards the detector are wanted for the analysis. Adequate shielding must be provided to absorb all other γ-rays.

3.3.5 Composite Shielding

A good shielding material for neutrons is a mixture of a hydrogenous material and a good thermal neutron

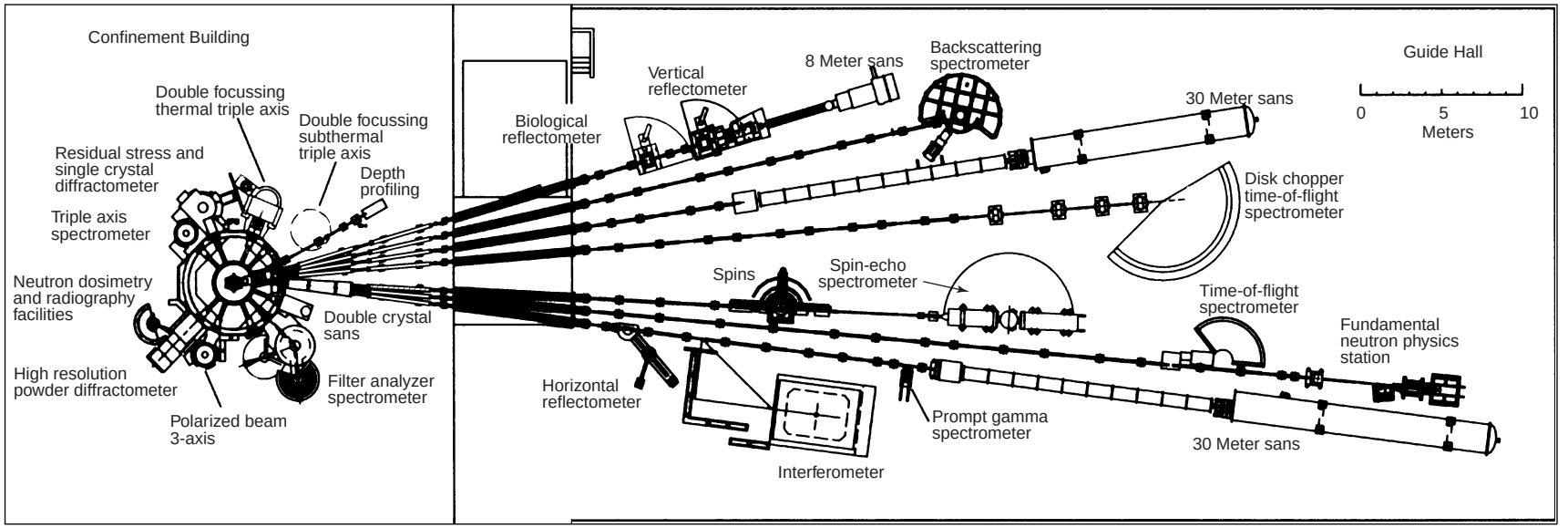

Figure 4 Experimental facilities around the NIST Reactor. A PGNAA facility is represented. (Reproduced from Paul et al.[12] with permission.)

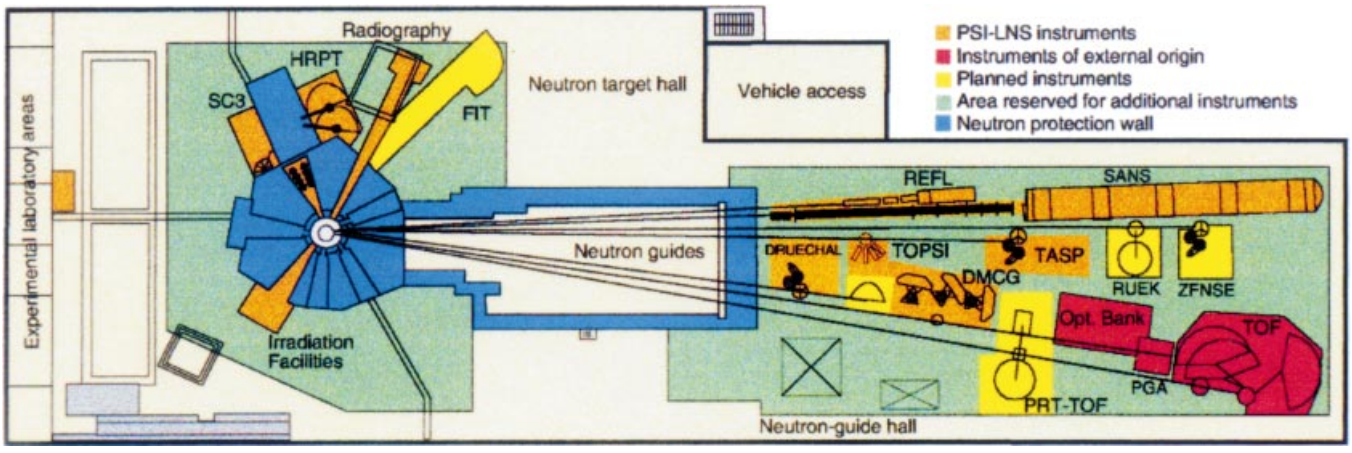

Figure 5 Experimental facilities around the Paul Sherrer Institute spallation neutronen quelle (SINQ). A PGNAA facility is represented on the right-hand side. (Reproduced from Paul Scherrer Institute Annual Report[29] with permission.)

absorber. Polyethylene with addition of boron or lithium has proven to make a good shielding because of the physical and chemical properties of the composite: high density of hydrogen atoms, chemical inertness and ease of machining. Lithium–polyethylene materials are very useful as detector shielding because lithium generates no capture γ-rays. Other hydrogenous materials, such as water or paraffin, are not advisable for industrial applications because of fire hazards and the possibility of leakage that can jeopardize safety.

A combination of lead and boron–polyethylene makes a material with good shielding properties for mixed neutron–γ fields.

3.4 Detectors

The criteria for the choice of a γ-ray detector must take into consideration efficiency, resolution, sensitivity to neutron damage and maintenance.

Essentially two types of detector are used in PGNAA. Scintillation detectors based on a thallium-doped sodium iodide detector, (NaI(Tl)), or a bismuth germanate detector ($Bi_4Ge_3O_{12}$), (BGO), and solid-state germanium detectors, normally the n-type hyper-pure germanium detector (HPGe). Resolution, efficiency and maintenance care are different for the two detector types.

The HPGe detector shows the best resolution, approximately 0.1% fwhm (full width at half-maximum). The NaI(Tl) scintillator has a resolution of about 5% at 2 MeV, or 3% at 10 MeV; the BGO scintillator has the poorest resolution, about a factor of two lower than that of NaI(Tl). The resolution of the HPGe detector enables the separation of closely spaced γ-lines in complex spectra.

Regarding efficiency, HPGe detector units currently used normally have lower efficiencies than BGO and NaI(Tl) detector units. However, HPGe detectors can have relative efficiencies (compared to a standard NaI(Tl) detector) of the order of 1, if a large enough germanium crystal is used, with the corresponding high cost. The HPGe detector is the most sensitive to neutron interactions within the crystal that promote the formation of traps causing resolution degradation. Owing to the mechanism of transport of charged particles inside the crystal, the n-type detector is less sensitive to neutron damage than the p-type.

Finally, HPGe detectors must be operated at low temperature (100 K) requiring expensive maintenance services. This is a drawback for on-line applications in industry.

Frequently an association of different detectors is used to optimize the arrangement (HPGe/NaI(Tl) or HPGe/BGO).

Detectors used for on-line industrial PGNAA are in general scintillation detectors. Reactor-based analyzers predominantly use HPGe detectors.

3.5 Signal Processing and Data Handling

The main objectives of signal processing and data handling at a PGNAA facility are to maximize the pulse throughput rates and minimize the γ-ray background without detriment to the energy resolution.

The characteristics of the signal processing system must be adapted to the type of detector and to the expected γ-ray count rates. Signal processing and data handling are carried out by a γ-spectroscopy unit.

The basic components of the γ-spectroscopy unit are the preamplifier, amplifier, analog to digital converter (ADC) and multichannel analyzer (MCA).

The function of a preamplifier is to extract the signal from the detector without significantly degrading

the intrinsic signal-to-noise ratio. The preamplifier is located as close as possible to the detector to minimize noise.

A germanium detector represents an extremely high capacitive impedance source (10 GΩ, 10–50 pF), generating a small charge ($Q = 2.7 \times 10^{-13}$ C) in the detection of a 5-MeV particle. The small output signal requires that the preamplifier does not add appreciable noise.

The absorption of a γ-ray by the detector produces a current pulse at the preamplifier input. The preamplifiers are typically of charge sensitive configuration and are coupled to the detector through the gate of a field effect transistor (FET). Two types of these preamplifiers can be used with Ge detectors: one employs the dynamic charge restoration method (resistor capacitor feedback) and the other the pulsed charge restoration method (pulsed optical or transistor reset) to discharge the integrator.

In many high count rate and high-resolution applications the transistor reset preamplifiers (TRP) offer the best performance.

Nowadays advanced hybrid-electronics manufacturing capability allows the incorporation of the detector element, preamplifier and high-voltage filter into a compact streamlined system.

Where the resolution is not a vital system parameter it is not necessary to use Ge detectors. In their place scintillation detectors can be used. A scintillation detector coupled to a photomultiplier tube generates moderately large signals at the output thus relaxing restrictions on the noise contribution from the preamplifier. They can be also associated with charge sensitive preamplifiers.

A main amplifier with sufficient linearity and large enough bandwidth extracts the energy information from the preamplifier output and converts it to a shape that can be easily analyzed by an ADC. The first stage of the amplifier (differentiator) reduces the decay time and blocks the low frequency or direct current components in the input signal. The following stage (integrator) recovers the energy information from the signal. This yields Gaussian-shaped pulses. The amplitude of each pulse is directly proportional to the energy of the γ-radiation.

To process an incoming pulse completely the amplifier must be adjusted for a time constant (TC) about 10 times the input pulse rise time. A shorter TC stops the processing before it has reached its maximum amplitude, which means loss of resolution in the data. But a longer TC means longer amplifier processing time, which reduces the throughput capabilities of the amplifier: if a new pulse reaches the amplifier before processing the previous one, a pile-up occurs and the piled up pulses are invalid events. Although there are ways to prevent the processing of these pulses, the net effect is decreased system output. As a shorter shaping TC minimizes pile-up, a TC of 2 μs is recommended for good resolution with Ge detectors. Its maximum throughput capability is just under 20 kcps (kilocounts per second), which is achieved at an input count rate of about 50 kcps. If the input rate exceeds that value, throughput actually decreases owing to pulse pile-up.

For high count rates the use of a gated integrator (GI) will allow a gain of the order of 10 kcps in throughput.

In order to ensure that the piled-up events are not processed by the ADC, additional circuits must usually be considered. These circuits are called pile-up rejector/live time corrector circuit (PUR/LTC). When a new pulse arrives before the previous one is processed, the new pulse is rejected.

The ADC unit converts the amplitude of each event into a digital value to be processed by the MCA or computer. Two types of ADC are available: the Wilkinson ADC and the fixed conversion time (FCT) ADC. The FCT ADCs can offer conversion times significantly shorter than those from the Wilkinson type at least for higher energy events. The choice depends on the specific application taking into account channel resolution and signal amplitude.

The MCA or a personal computer-based MCA acquisition board collects and stores the data, provides its visualization on a monitor, performs mathematical operations to generate the final results and transfers the raw and/or the treated data to an external storage device.

Current cheap memory circuitry provides adequate numbers of data channels and counting capacities as well as short enough data storage times (no more than a couple of microseconds) to deal with the highest count rates encountered in practice.

Hanna et al.[30] and Wormald[31] proposed the use of compound detectors for Compton suppression and pair spectrometry. The use of coincidence/anticoincidence techniques with compound detectors can provide simpler spectra by selecting only the useful events. Pair production and Compton scattering are the most important interactions for high-energy γ-rays. If the interaction occurs within the sensitive volume of the detector and the resulting secondary γ-rays escape the detector, a smaller amplitude pulse is recorded.

Schematically, pair spectrometers use three detectors: a central one in which the primary radiation is detected by pair production, and two others for detecting the 0.511-MeV annihilation photons. An event is recorded by using coincidence techniques if a photon of $E_\gamma - 1.022$ MeV is registered in the central detector and, simultaneously, two 0.511 MeV photons are detected in the others. This method improves the high-energy tail of the spectrum. Wormald proposed a close packing array of cylinders or

hexagonal prisms to form a large collecting area adapted to the shape of the sample.

The Compton suppression spectrometer uses anticoincidence techniques to eliminate events resulting from Compton scattering: only signals which do not have a coincidence signal in another detector are recorded. This technique greatly simplifies the low-energy part of the spectra. A γ-ray spectrometer operating simultaneously in Compton suppression and pair production modes acquires information in three modes: single mode, Compton suppression mode and pair mode. Analysis of the complex γ-ray spectrum can be performed by comparing the three spectra. The spectrometer performance is enhanced.

A wide variety of application programs is available for data handling of the spectra (nuclide identification, background and interference corrections, peak intensity and percent weight content calculation).

4 INSTRUMENT DESIGN AND CALIBRATION

The gauge response of a PGNAA analyzer is determined in a complex manner by the system parameters – geometry, source activity and neutron energy spectrum, and detector efficiency – and by certain properties of the analyzed sample such as bulk density and water content. It follows that the count rate caused by a given γ-ray line is a nonlinear function of the concentration of the element to be measured.

The density and the dimensions of the sample have a direct effect on reaction rates in the sample volume and on the attenuation of γ-rays and thus on the γ-count rate at the detector. For a fixed geometry and composition of the sample the thermal neutron flux depends on the bulk density. When the density increases, the thermal neutron distribution is ''compressed'' spatially with a higher maximum shifted towards the region of the sample adjacent to the neutron source. On the other hand, the relevant γ-rays on average travel a longer way to the detector inside the sample and are subject to a stronger attenuation. These two effects act in opposite directions upon the γ-count rate and can cancel each other for particular geometries.

Furthermore, the hydrogen content of the sample may also have a significant effect on the neutron flux distribution.

In principle, the composition of an unknown sample can be derived by solving Equations (3) or (5). In this direct or absolute method the concentration of each element is calculated using fundamental parameters, such as cross-sections, number of photons emitted by the element per neutron captured, detector efficiency, γ-ray attenuation coefficients in the sample and solid angle viewed by the detector.[32,33] However, this method is only applicable in a small number of cases owing to uncertainties in the available microscopic nuclear parameters, the unavailability of others and the influence of composition, density and sample volume on the count rate from a given γ-ray line. In activation analysis of large samples a number of correction factors must be applied (neutron absorption and moderation, γ-ray attenuation within the sample and detector efficiencies).

The simplest case is that of thin samples irradiated in a thermal or cold neutron beam. Neutron scattering and γ-absorption in the sample are, then, negligible thus reducing the influence of sample composition and homogeneity on the count rate.

The problem becomes more complicated for on-line gauges using isotopic neutron sources and large samples. Generally, there is neutron thermalization inside the sample so that the thermal neutron flux distribution is dependent on the sample composition, particularly on hydrogen content (moisture) and bulk density.

There are, in principle, several methods for calibrating a PGNAA gauge.

4.1 Experimental Calibration

Experimental calibration establishes the correlation between the γ-ray peak areas and the known chemical composition, density and water content of the samples. The experimental determination of the calibration curves demands long laboratory work, using a large number of standards with well-known compositions, matrix and geometry that closely match those of the sample.[34] The variation in a given parameter can affect the γ-count rate due to a given element, although the concentration of the element remains unchanged. It is a very time-consuming approach, because new standards must be prepared for each sample type.

4.2 k_0 Method

The k_0 method is used extensively in delayed NAA.[35] The k_0-factor is the ratio of the absolute sensitivity divided by the detection efficiency of the γ-rays of the element to be measured compared with that of a comparator element, s, Equation (9):

$$k_{0j} = \frac{R_j/m_j\varepsilon_j}{R_s/m_s\varepsilon_s} = \frac{\theta_j I_j \sigma_j/M_j}{\theta_s I_s \sigma_s/M_s} \tag{9}$$

where θ is the isotopic abundance of the nuclide.

In the absence of significant γ-ray self-absorption and for elements with neutron cross-section varying as $1/v$,

the k_0-factor is independent of neutron energy, sample matrix and geometry.

Measurement of k_0-factors allows in principle accurate determinations of relative concentrations of every element contained in the sample without the use of the comparator. In effect, the relative concentration of element j to element i in the sample is given by the formula, Equation (10):

$$\rho_{j,i} = \frac{m_j}{m_i} = \frac{R_{\gamma j}/\varepsilon_j k_{0j}}{R_{\gamma i}/\varepsilon_i k_{0i}} \tag{10}$$

When the concentration of one of the elements is known, the absolute concentration of any other element can be assessed. Otherwise it is necessary to measure only the sample and a single standard of similar properties (geometry, neutron scattering and γ-absorption), containing the comparator element to determine absolute concentrations.

There are some problems that can affect the accuracy of the k_0 measurements. Among them are deviations from the $1/v$ behavior of the neutron cross-sections, gamma attenuation in thick samples or with large concentrations of high-Z elements and heterogeneities in the sample.

The k_0 method has been applied by several authors to PGNAA measurements.[36–39]

4.3 Monte Carlo Library Least Squares Technique

Gardner et al. have developed the Monte Carlo library least squares method (MCLLS) for PGNAA.[40–42]

According to this model the PGNAA gauge is a ''black box''. The gauge response, R is represented by the function, Equation (11):

$$R = f(x, u, v, w, \ldots) \tag{11}$$

where x is the parameter to be measured and u, v and w represent variables that characterize the measurement interferences. Then, Equation (12):

$$x = F(R, u, v, w, \ldots) \tag{12}$$

In the linear least squares (LLS) model the spectrum of the sample is a linear combination of individual components. The content of each channel can be expressed as, Equation (13):

$$b_i = \sum_{j=1}^{m} x_j a_{ij} + E_i \quad i = 1, n \tag{13}$$

where b_i is the unknown sample spectrum or counting rate in channel i of the sample; x_j is the amount of component j in the sample; a_{ij} are the library spectra or counting rate per unit amount of component j in channel i; and E_i is the random error in channel i due to statistical counting rate fluctuations.

If the system is linear, i.e. if the gauge response to an individual component does not depend on other parameters, the concentration, x_i can be obtained by minimizing the reduced chi-square with respect to each amount.

However, as the PGNAA analyzers are usually nonlinear, the MCLLS approach has been developed. This is an iterative method that requires that a very accurate Monte Carlo code enables the calculation of the spectra of a sample given its composition, density and geometry. To this end the entire arrangement and the sample are simulated.

Simulations are performed in order to generate the elemental library spectra of all elements present in the sample. The model consists of the following steps:

1. The Monte Carlo code generates the pulse height spectrum of a sample whose composition and density are similar to that of the unknown sample.
2. The prompt γ-ray spectrum for each element is recorded separately to serve as a library spectrum.
3. The linear library least squares analysis is then performed using the spectra of the unknown sample and the calculated library spectra. The elemental composition obtained is compared with the initially assumed values.
4. If the calculated elemental composition of the unknown sample is close enough to the values assumed for the Monte Carlo simulation, these are taken as the final measured values. If not, the LLS values are taken for the next iteration and steps 1 through 3 are repeated until the linearity assumption is valid.

4.4 Monte Carlo Simulation of Calibration Curves

Another method relies on the Monte Carlo simulation of the measuring arrangement with samples having different compositions, densities, water contents and thicknesses, in order to establish the correlation between the count rate for unit concentration of an arbitrary element and the sample parameters.[43] A comparison is then carried out between the simulated calibration curves and the experimental results obtained with a few standards. A regression analysis is carried out to adjust the simulated calibration curves to the experimental points. Although Monte Carlo simulation studies do not entirely suppress the need to carry out experimental calibration work, they considerably reduce the number of calibration standards required. This method has been used by

Oliveira et al.[44,45] using the MCNP (Monte Carlo N-particle) code[46] for PGNAA analysis of coals and cement raw materials.

4.5 Design Optimization

The crucial part of the design of a PGNAA system is that of ensuring a well-defined correspondence between the output of the system and the elemental composition of the sample. Design optimization is particularly important in the case of equipment for field use with isotopic neutron sources. With fast neutron sources, the thermal flux within the sample volume strongly depends on the moderating properties of the sample, which in turn depend on its density and the number of hydrogen atoms per unit volume, i.e. the water content. On the other hand, the average path length to the detector of γ-rays generated in the sample strongly depends on the sample dimensions. Thus the γ-ray count rate at the detector depends on the sample bulk density and dimensions and on the spatial distribution of the thermal flux.

The purpose of design optimization is to reduce the complexity of the functional dependence of the system response relative to the chosen calibration parameters. This can be achieved in a certain range of sample parameters by acting on the sample geometry, namely the sample thickness, and on the effective neutron source spectrum modified by the introduction of a moderating material around the source or between source and detector, which will substantially influence the thermal neutron flux and its spatial distribution inside the sample.

Monte Carlo simulation is a useful method for speeding up design work and system optimization. The influence of the different parameters on the specific count rate from an element can be established by means of the simulation. Optimum design parameters mean that the gauge response is less sensitive to measurement interferences, such as the ones referred to above (fluctuations of density and water content). To reach this objective, the variance of the measured variables must be minimized with respect to both the gauge response variance and the variances of each measurement interference (see Equation 12).

5 APPLICATIONS OF THE METHOD

5.1 Introduction

PGNAA is a powerful analytical technique that can provide accurate information about the elemental composition of materials in a broad range of cases: raw materials, mineral exploration, contaminants, detection of forbidden products, such as drugs and explosives, inspection of containers (detection of plutonium) and applications in health care. In the following sections the main fields of application of the method are briefly reviewed.

5.2 Borehole Logging

Since the early 1980s, the most interesting advances in nuclear logging techniques concerned the development of spectroscopy techniques, especially for the analysis of γ-ray spectra.[47]

Measurements of the energy spectrum of γ-rays were first applied to the detection of natural radioactivity from potassium and of unstable isotopes of the thorium and uranium decay series. Later on, neutron sources coupled to energy-sensitive detectors were used to observe γ-rays from neutron-induced reactions.

Nowadays, spectroscopy of γ-rays from thermal neutron capture and the inelastic scattering of fast neutrons is commonly used in the study of wellbore geophysics. In this type of well logging, a gauge consisting of a neutron generator or a radioisotopic fast neutron source, and γ-ray detectors is introduced through the formations in boreholes. There are applications for exploration and exploitation of mineral resources (oil, gas and coal) as well as for the study of geochemical processes.

The main interest of the oil and gas industry in this technique derives from its ability to locate porous zones that may contain hydrocarbons and are free of clay minerals whose presence would inhibit fluid flow. At the same time it must be ensured that oil and gas production do not contaminate potable water supplies. The coal industry is interested in in situ control of coal quality.

The set of detectable elements provides valuable information on the nature of rock matrix or geological formation. Previously, this information could only be obtained through discrete core analyses.

However, the proper determination of elemental concentrations requires reliable information on geophysical parameters as well as specific spectrometric information related to the individual elements. Several approaches have been developed.[48–50]

Elemental concentrations obtained from logging measurements have been routinely used to provide sedimentary mineralogy, cation-exchange capacity, permeability and other rock and fluid properties. A comparison between mineralogy derived from logging measurements and that derived from laboratory analyses of core samples is shown in Figure 6.

Information from borehole logging during coal exploration can give an early indication of coal quality by allowing the direct determination of the concentration of sensitive elements and the derivation of important

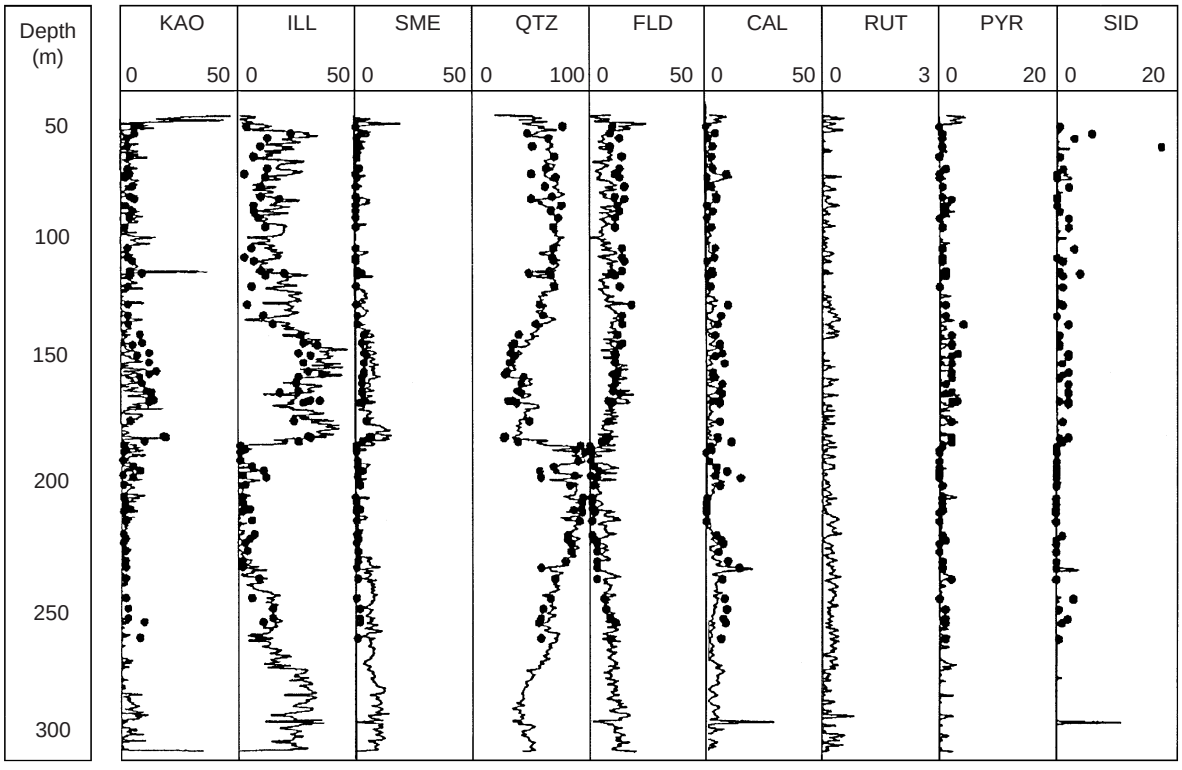

Figure 6 Example of the comparison between mineralogy derived from logging measurements and that derived from laboratory analyses of core samples. The continuous curve is the result from elemental logging measurements. The circles are the results from laboratory analyses of core material. KAO, kaolinite; ILL, illite; SME, smectite; FLD, feldspar; QTZ, quartz; CAL, calcite; RUT, rutile; PYR, pyrite; SID, siderite. (Reproduced from Clayton and Schweiter, *Nucl. Geophys.*, **7**, 158, Copyright (1993)[47] with permission from Elsevier Science.)

economic parameters such as ash content and calorific value.

5.3 Industrial On-line Applications

On-line industrial applications of PGNAA are particularly important in the quality control of coal and in the fabrication of cement. Equipment for bulk analysis of coal and cement is presently commercially available. Other on-line applications in raw material control have been implemented, e.g. the determination of Al in bauxite ores[51] or conveyor belt sorting of aluminum alloys in scrap.[52]

The PGNAA technique surpasses the X-ray fluorescence (XRF) technique owing to the higher penetrating power of the neutron–gamma radiation, which as a consequence has the potential to analyze the whole bulk of the material stream. In these applications radioisotopic sources (generally ^{252}Cf) or neutron generators are used.

The raw material can flow through a vertical chute where the analysis process occurs; generally the chute is lined with an extremely low-friction material to ensure plug free operation. Alternatively, the raw material can be conveyed on a belt. In this case physical integration of the equipment in a previously existing process line is simpler and less costly. Problems related to ''dead zones'' which always appear in the vertical chute are also avoided in this configuration.

The use of coal as an energy source has severe environmental consequences (ash and SO_2 emissions). Good quality control in all phases of the coal fuel cycle will limit the adverse effects. PGNAA can be used in all phases of the coal cycle: mining, coal preparation and power station operation, for the determination of ash, moisture content and specific energy. Among the major coal elements, hydrogen and iron give the best response; sulfur and many of the ash elements (Si, Al, Ti, Ca, Mg, Na, K) are also satisfactorily detected. Carbon responds poorly; however, as its concentration is high it can be detected. Oxygen can hardly be detected in the γ-spectrum due to the element's very small thermal neutron capture cross-section. However, oxygen and carbon can be easily detected using (n,n$'\gamma$) reactions, whose threshold energies are, respectively, 6.13 MeV and 4.44 MeV. This implies the use of a 14-MeV neutron generator or of ^{241}Am-Be as the neutron source.[53]

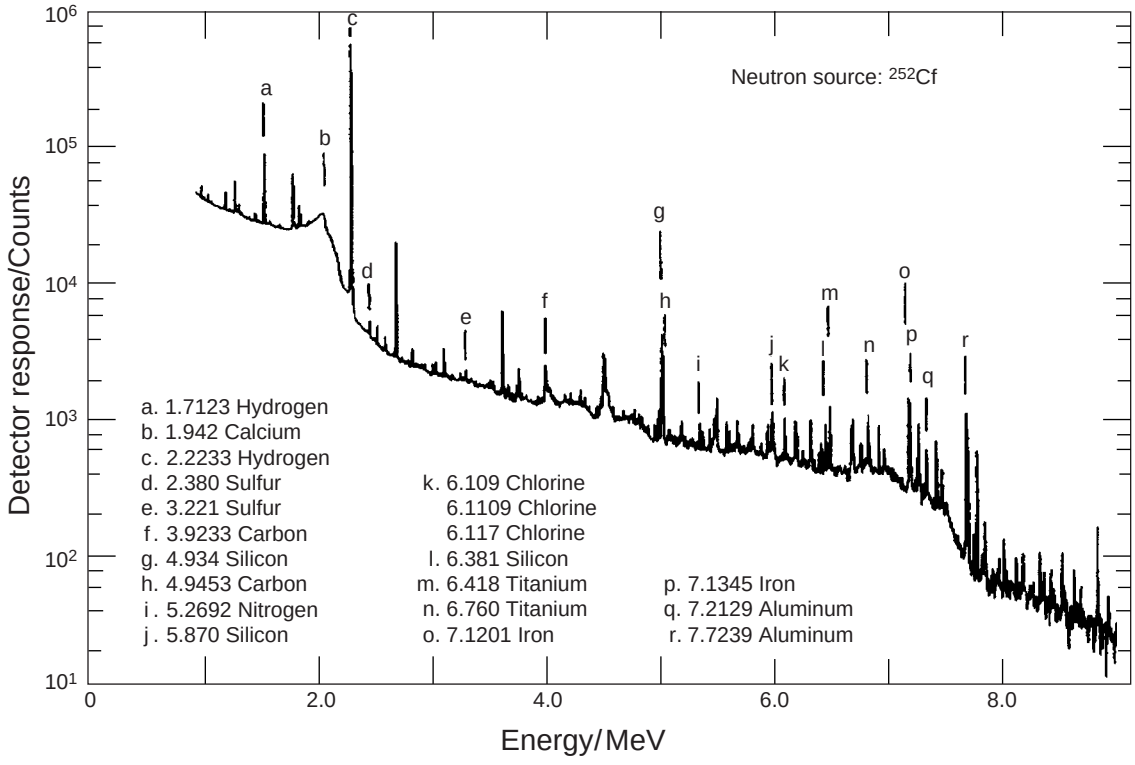

Figure 7 An experimental prompt γ-ray spectrum of a sub-bituminous coal. (Reproduced with permission from Y.L. Yuan, R.P. Gardner, K. Verghese, *Nucl. Technol.*, **77**(1), 105 (1987), Copyright 1987 by the American Nuclear Society, La Grange Park, Illinois.)

Figure 7 shows an experimental prompt γ-ray spectrum of a coal.

The cement industry is fundamentally interested in the measurement of silicon, aluminum, iron and calcium oxides. All the elements involved (except oxygen) can be determined by PGNAA.[54]

The most common application in the fabrication of cement is raw mix proportioning, with the analyzer positioned just ahead of the raw mill. Real-time results can be used to make frequent adjustments of the component feeders to keep the raw mix close to reference composition; another application is the analysis of raw material as it enters the preblending pile. As an alternative to mechanical samplers, analyzers can reduce maintenance and labor costs. A third application is the sorting of raw material coming from the quarry. When synchronized with a downstream diverter, the analyzer can be used to sort one-minute increments of material into save or reject piles.

5.4 Medical In Vivo Applications

In vivo neutron activation analysis (IVNAA) is a unique direct in vivo method for the multielemental analysis of the living human body. Various other radiation-based techniques (radiography, isotopic imaging, radiotracer dilution) reflect only relative changes in tissue density or volume. Table 6 shows the elements measured by activation analysis.[55]

The prompt γ-variety of IVNAA can provide data on the total or partial body content of the following elements: oxygen, carbon, hydrogen, nitrogen, calcium, sodium, chlorine, silicon, iron, cadmium and mercury.

Although PGNAA instruments were initially developed in nuclear research laboratories, methodologies have advanced to a stage where facilities were designed and built exclusively for clinical application in hospitals and health centres.

The choice of the neutron source is based on elements to be measured, the required degree of uniformity of the activation, the acceptable level of accuracy and the admissible radiation dose. Cf and Pu-Be are the most used sources; some facilities are equipped with neutron generators or cyclotrons. Large NaI(Tl) scintillators are generally used as detectors, although in certain facilities germanium or BGO detectors are preferred.

Considerable effort has been expended in reducing the absorbed dose without decreasing the precision of results. The substitution of Pu-Be sources by Cf contributes to this end. The range of exposure doses in existing facilities depends in part on the element to

Table 6 Elements measured by in vivo activation analysis[a]

Body element	Reference man		Technique	E_γ (MeV)
	g	%		
Total body				
Oxygen	43 000	61	pulsed	6.10
Carbon	16 000	23	pulsed	4.44
Hydrogen	7000	10	prompt	2.23
Nitrogen	1800	2.6	prompt	10.83
Calcium	1000	1.4	delayed	3.10
			prompt	6.42
Phosphorus	780	1.1	delayed	1.78
Sodium	100	0.14	delayed	2.75
			prompt	6.40
Chlorine	95	0.12	delayed	6.11
			prompt	2.23
Partial body				
Silicon	18	trace	prompt	1.78
Iron	4.2	trace	delayed	0.85
			prompt	0.847
Cadmium	<0.15	trace	prompt	0.56
Aluminum	<0.08	trace	delayed	1.78
Mercury	<0.002	trace	prompt	0.37
Lithium	<0.01	trace	prompt	($^3_1$H counting)

[a] Reproduced from Ellis[55] with permission.

be analyzed, the precision required and the detection sensibility of the counting system. The Swansea system is shown in Figure 8.[56] The dose is often less than, or comparable to, many routine diagnostic radiographic procedures.

The initial application that has influenced the subsequent development of IVNAA is the determination of total body calcium content in patients suffering from various metabolic diseases. Since calcium is predominantly located in the skeleton, a measure of this element would provide a direct examination of the total skeletal mass. Clinical research studies have investigated a number of metabolic disorders in which there is a primary or secondary alteration of mineral metabolism: osteoporosis, osteomalacia, renal osteodystrophy, Paget's disease, Cushing's syndrome, acromegaly, thyroid and parathyroid disorders, myotonic dystrophy, thalassemia and alcoholic cirrhosis.

The body nitrogen measurement provides a direct determination of body protein, which is the essential component of the body's lean tissue mass. Changes in protein mass can reflect changes in nutritional status. Hence clinical studies involving nitrogen measurements in patients with progressive diseases may be important in the following cases: postoperative surgery, protein malnutrition, cancer renal failure, cardiovascular disease, total parental nutrition, obesity and growth deficiencies.

Body carbon measurements can provide an index of the body fat.

Cadmium measurements have been carried out in industrial workers. Kidney concentrates cadmium and has been identified as the initial target organ in toxicological studies. The association between cadmium burdens and various diseases related to kidney malfunctions have also

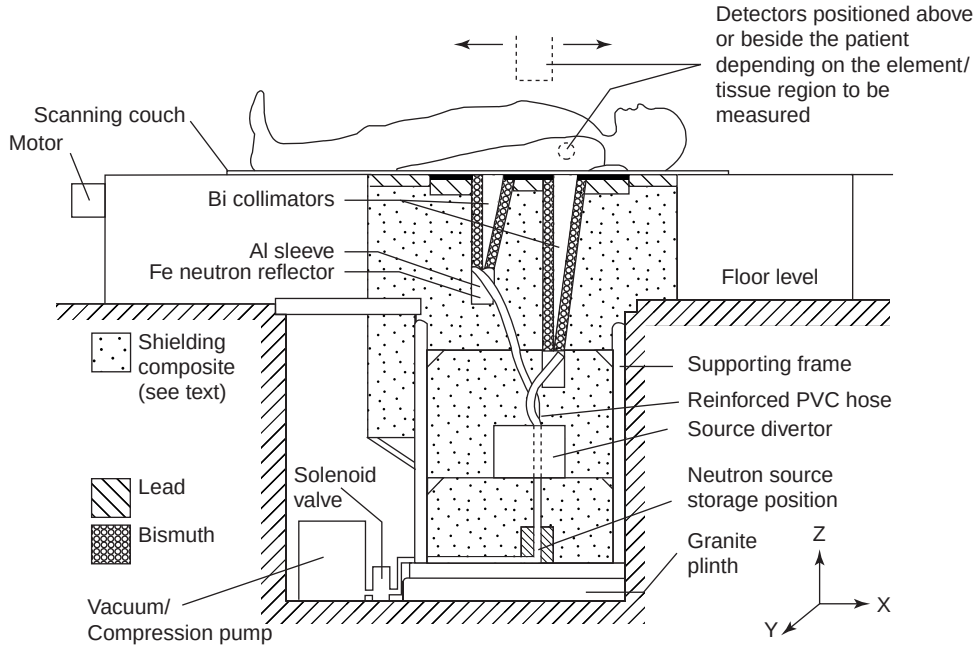

Figure 8 The Swansea IVNAA system. (Reproduced from Natto et al., *Int. J. Appl. Radiat. Isot.*, **49**, 546, Copyright (1998)[56] with permission from Elsevier Science.)

been investigated. Several studies involving nondialyzed patients in renal failure have shown that cadmium was a cofactor in the cause of the renal disease.

5.5 Determination of Environmental Contaminants

Elemental analysis of toxic elements in contaminated materials is essential for solving environmental problems. PGNAA can also be applied to this purpose in the determination of major and minor elemental components and even of trace elements.

Dulloo et al.[57] used a 14-MeV pulsed generator for the nondestructive assay of mercury, cadmium and lead. The DLs for a 600 s run are 15 ppm, 170 ppm and 8600 ppm for Cd, Hg and Pb respectively.

Yonezawa et al.[23] carried out multielement analysis of environmental samples by PGNAA with cold and thermal neutrons in a reactor facility. Cd was detected in mussels, river sediments, pond sediments, vehicle exhaust particulates and unpolished rice flour; Hg was detected in human hair.

Hiep and Tan[22] used the technique for analyzing the concentration of Al, Fe, Ti, Si, Ca and K in airborne dust samples.

5.6 Explosives and Drugs Detection Systems

Considerable progress has been made in research and development of nuclear-based techniques for the detection of explosives and drugs. The technique must be noninvasive, sensitive to the smallest effective amounts of explosives and drugs and give rise to a minimum of false alarms. An effective contraband inspection system must be capable of determining the relevant chemical elements and their spatial distribution in the inspected object.[58]

Chemical explosives have high nitrogen and oxygen contents. Pure narcotics are characterized by high content of H and C, moderate content of O and small content of N. However, narcotics hydrochlorides also contain small amounts of Cl. While high nitrogen density is a good indicator of explosives, the detection of a low oxygen to carbon ratio combined with the presence of N and Cl is a strong indication of the presence of drugs.

The PGNAA technique enables the three-dimensional mapping of explosives and narcotics in luggage. This is accomplished by irradiating parts of the interrogated object with short pulses of fast monoenergetic neutrons (Figure 9). The three-dimensional distribution can be obtained by scanning the sample.[59]

5.7 Other Applications

Many other applications have been carried out in different fields using PGNAA.

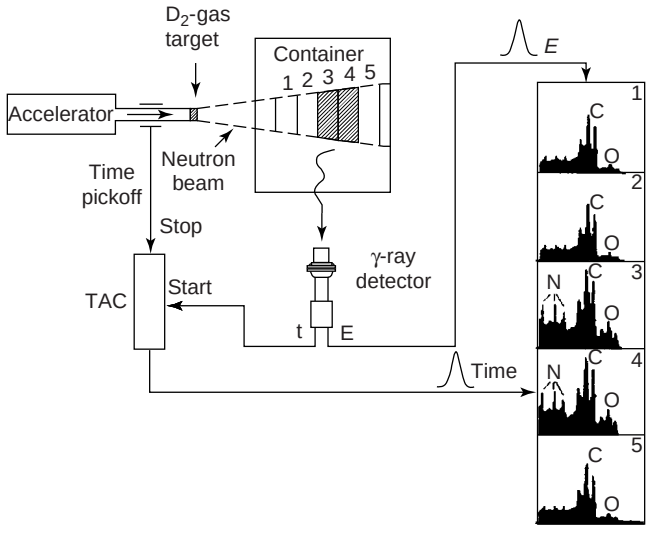

Figure 9 Principle of three-dimensional multielemental mapping. The γ-ray spectra shown on the right are measured in the five indicated layers. (Reproduced from Sawa[58] with permission.)

Anderson et al.[60] described the determination of H, B, Cl, K, Na, S, Ca and Cd in foods using a thermal neutron beam. The advantage of this type of analysis lies in its sensitivity to elements of nutritional or toxicological interest while no sample preparation is needed, thus eliminating the risk of volatilization losses.

Kuno et al.[61] applied PGNAA to the determination of the concentration of B, H, S and Si in on-shore sediments in order to investigate the distribution and migration of chemical elements. The measurements were complemented by DGNAA results.

Caffrey et al.[62] applied a portable PGNAA device with an isotopic source to the identification of the contents of munitions, chemical storage containers, chemical warfare materials in stockpiles, etc.

Prettyman et al.[63] investigated the use of the technique as a diagnostic tool related to matrix characterization for nondestructive assay and plutonium surveillance. The technique can help to confirm item descriptions (for example, plutonium chloride versus plutonium oxide).

Spyrou et al.[64] proposed a method to determine the concentration of a large number of elements by measuring the prompt or delayed γ-rays and applying the principles of reconstructive tomography to obtain the spatial distribution of elemental concentration in a selected plane or slice through the sample. This method has been employed in the examination of radioactive waste as well as in elemental analysis of biomedical and bioenvironmental samples.

The contents of major and some trace elements in large archaeological materials have been determinated by Sueki et al.[24]

6 PRESENT TRENDS OF DEVELOPMENT

According to Gardner[52] ''PGNAA is a rapidly emerging important new technology and measurement approach''. However, further development of its applications in different branches of human activity require improvements in neutron sources, signal processing and data handling as well as in nuclear data.

The introduction of a new point source generator based on the IEC principle may have a promising future in on-line field applications of PGNAA. Apart from sharing the on-off switching capability with conventional neutron generator devices, it offers longer life and lower costs.

Industrial on-line process control requires short response times. For a PGNAA analyzer this means, normally, high count rates. High-rate data acquisition systems must be improved.

Enhancement of signal to background noise ratio leads to lower statistical uncertainty of the results thus contributing to reducing response times. This objective can be achieved by proper system design optimization. Monte Carlo methods can play an important role in this respect.

More accurate measurements and thus better data regarding microscopic nuclear parameters, such as cross-sections, number of photons emitted per neutron captured by the relevant isotope and γ-ray attenuation coefficients in different materials are important for simulation studies carried out with a view to optimizing design and calibration of PGNAA analyzers.

Although certain physical constraints (neutron flux perturbations and γ-ray attenuation) are difficult to overcome, the development and application of the k_0 method to large samples would be welcome.

The MCLLS technique, which has been applied to coal, cement raw materials, oil well logging, vitrified nuclear waste and conveyor belt sorting of aluminum alloys in scrap, can also be successfully applied to other fields.

The nuclear imaging technique using a pulsed fast neutron analyzer for the detection of explosives concealed in airline luggage has potential application in the future to a variety of other situations.

ACKNOWLEDGMENTS

The authors wish to thank our colleagues Carlos Cruz for helpful discussions concerning signal processing and data handling, António Falcão for his contribution in graphics design of Figures 1 and 2 and Teresa Pires for her improvements in graphics in Figures 3, 6, 7, 8 and 9.

ABBREVIATIONS AND ACRONYMS

ADC	Analog to Digital Converter
BGO	Bismuth Germanate Detector
DGNAA	Delayed γ-Neutron Activation Analysis
DL	Detection Limit
FCT	Fixed Conversion Time
FET	Field Effect Transistor
fwhm	Full Width at Half-maximum
GI	Gated Integrator
HPGe	Hyper-pure Germanium Detector
IEC	Inertial Electrostatic Confinement
INAA	Instrumental Neutron Activation Analysis
IVNAA	In Vivo Neutron Activation Analysis
LLS	Linear Least Squares
MCA	Multichannel Analyzer
MCLLS	Monte Carlo Library Least Squares Method
MCNP	Monte Carlo N-particle
NAA	Neutron Activation Analysis
NaI(Tl)	Thallium-doped Sodium Iodide Detector
NIST	National Institute of Standards and Technology
PGNAA	Prompt γ-Neutron Activation Analysis
PUR/LTC	Pile-up Rejector/Live Time Corrector Circuit
SINQ	Spallation Neutronen Quelle
TC	Time Constant
TRP	Transistor Reset Preamplifiers
XRF	X-ray Fluorescence

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Radiochemical Methods **(Volume 14)**
Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation

REFERENCES

1. J.L. Mikesell, F.E. Senftle, A.B. Tanner, 'Borehole Elemental Analysis of Coal Using Fast Neutrons, Gamma, X-ray and Neutron Techniques for the Coal Industry', Proceedings of an Advisory Group Meeting, Vienna, 1984, IAEA, Vienna, 25–42, 1986.
2. M.R. Wormald, C.G. Clayton, 'In-situ Analysis of Coal by Measurement of Neutron-induced Prompt γ-rays', *Int. J. Appl. Radiat. Isot.*, **34**(1), 71–82 (1983).
3. C. Oliveira, J. Salgado, 'A Monte Carlo Simulation Study of the Space Dependence of Prompt Gamma-ray Volume Emission Rate in Bulk Coal Irradiated with an Am-Be Source', Proceedings of the Conference on Radioisotope Application and Radiation Processing in Industry, Leipzig, ed. D. Luther, Institute of Isotope and Radiation Research (2), 737–743, 1989.
4. D.L. Anderson, M.P. Failey, W.H. Zoller, W.B. Walters, G.E. Gordon, R.M. Lindstrom, 'Facility for Non-destructive Analysis for Major and Trace Elements Using Neutron-capture Gamma-ray Spectrometry', *J. Radioanal. Chem.*, **63**(1), 97–119 (1981).
5. R.L. Paul, R.M. Lindstrom, A.E. Heald, 'Cold Neutron Prompt Gamma-ray Activation Analysis at NIST – Recent Developments', *J. Radioanal. Nucl. Chem.*, **215**(1), 63–68 (1997).
6. 'Industrial Gauging and Analytical Instrumentation Sources', Amersham Catalogue S96/89, Amersham, UK.
7. E.A. Lorch, 'Neutron Spectra of ^{241}Am/B, ^{241}Am/Be, ^{241}Am/F, ^{242}Cm/Be, ^{238}Pu/^{13}C and ^{252}Cf Isotopic Neutron Sources', *Int. J. Appl. Radiat. Isot.*, **24**, 585–591 (1973).
8. Y.A. Karelin, Y.N. Gordeev, V.I. Karasev, V.M. Radchenko, Y.V. Schimbarev, R.A. Kuznetsov, 'Californium-252 Neutron Sources', *Appl. Radiat. Isot.*, **48**(10–12), 1563–1566 (1997).
9. P. Bach, 'Sealed Tubes Neutron Generators', *International Conference Proceedings on Nuclear Techniques for Analytical and Industrial Applications*, Mykonos, June 1992, eds. G. Vourvopoulos, T. Paradellis, Western Kentucky University, Bowling Green, KY, 397–412, 1992.
10. J.J. Kehayias, H. Zhuang, 'Use of a Pulsed Neutron Generator for the In Vivo Measurement of Carbon, Oxygen and Hydrogen Content in Humans', *International Conference Proceedings on Nuclear Techniques for Analytical and Industrial Applications*, Mykonos, June 1992, eds. G. Vourvopoulos, T. Paradellis, Western Kentucky University, Bowling Green, KY, 265–284, 1992.
11. G.H. Miley, J. Sved, 'The IEC Plasma-target-based Neutron Source', *Appl. Radiat. Isot.*, **48**(10–12), 1557–1561 (1997).
12. R.L. Paul, R.M. Lindstrom, D.H. Vincent, 'Cold Neutron Prompt Gamma Activation Analysis at NIST: A Progress Report', *J. Radioanal. Nucl. Chem.*, **180**(2), 263–269 (1994).
13. C. Yonezawa, M. Margara, H. Sawahata, M. Hoshi, Y. Iyo, E. Tachikawa, 'Prompt Gamma-ray Analysis using Cold and Thermal Guided Neutron Beams', *J. Radioanal. Nucl. Chem.*, **193**(1), 171–178 (1995).
14. K. Unlu, C.R. Martinez, B.W. Wehring, 'Prompt Gamma Activation Analysis with the Texas Cold Neutron Source', *J. Radioanal. Nucl. Chem.*, **193**(1), 145–154 (1995).
15. G. Molnar, T. Belgya, L. Dabolczi, B. Fazekas, Z.S. Revay, A. Veres, I. Bikit, Z. Kiss, J. Östör, 'The New Prompt Gamma-activation Analysis Facility at Budapest', *J. Radioanal. Nucl. Chem.*, **215**(1), 111–115 (1997).
16. B.W. Wehring, K. Unlu, C.R. Martinez, 'Application of Cold Neutron Prompt Gamma Activation Analysis at the University of Texas Reactor', *Appl. Radiat. Isot.*, **48**(10–12), 1343–1348 (1997).
17. S. Alam, M.A. Zaman, S.M. Islam, M.H. Ahsan, 'Extraction of Pure Thermal Neutron Beam for the Proposed PGNAA Facility at the Triga Research Reactor of AERE, Savar, Bangladesh', *Nucl. Method Instrum.*, **B83**, 235–244 (1993).
18. O.K. Harling, J.M. Chabeuf, F. Lambert, G. Yasuda, 'A Prompt Gamma Neutron Activation Analysis Facility Using a Diffracted Beam', *Nucl. Method Instrum.*, **B83**, 557–562 (1993).
19. C. Yonezawa, A.K. Wood, 'Prompt γ-ray Analysis of Boron with Cold and Thermal Neutron Guided Beams', *Anal. Chem.*, **67**(24), 4466–4470 (1995).
20. K. Sueki, K. Kobayashi, W. Sato, H. Nakahara, T. Tomizawa, 'Nondestructive Determination of Major Elements in a Large Sample by Prompt γ-ray Neutron Activation Analysis', *Anal. Chem.*, **68**(13), 2203–2209 (1996).
21. V.H. Tan, 'Utilisation of Horizontal Experimental Channels of the Dalat Reactor for Research and Applications', Proceedings 10th Pacific Basin Nuclear Conference, Kobe, Japan, Atomic Energy Society of Japan and Japan Atomic Industrial Forum, 10–16, 1996.
22. N.T. Hiep, V.H. Tan, 'Thermal Neutron Capture Prompt Gamma Analysis for the Element Determination in Environmental Objects', Proceedings 10th Pacific Basin Nuclear Conference, Kobe, Japan, Atomic Energy Society of Japan and Japan Atomic Industrial Forum, 1353–1357, 1996.
23. C. Yonezawa, H. Matsue, M. Hoshi, 'Multi-element Analysis of Environmental Samples by Cold and Thermal Guided Neutron Induced Prompt Gamma-ray Measurement', *J. Radioanal. Nucl. Chem.*, **215**(1), 81–85 (1997).
24. K. Sueki, Y. Oura, W. Sato, H. Nakahara, T. Tomizawa, 'Analysis of Archaeological Samples by the Internal Monostandard Method of PGAA', *J. Radioanal. Nucl. Chem.*, **234**(1–2), 27–31 (1998).
25. K. Riley, O. Harling, 'An Improved Prompt Gamma Neutron Activation Analysis Facility using Focused Diffracted Neutron Beam', *Nucl. Method Instrum.*, **B143**, 414–421 (1998).
26. Y. Oura, H. Nakahara, K. Sueki, W. Sato, T. Tomizawa, 'Completely Non-destructive Elemental Analysis of Bulky Samples by PGA (Prompt Gamma Analysis)', 13th

Radiochemical Conference, Marianske Lazne-Jachymov, Czech Republic, Czech Chemical Society, I.M. Marci Spectroscopic Society and Czech Radiological Society, Abstract, 136, 1998.
27. M.S. Zucker, N. Tsoupas, P.E. Vanier, U. von Wimmersperg, S.F. Mughabghab, E. Schmidt, 'Spallation Neutron Production Measurements', *Nucl. Sci. Eng.*, **129**, 180–186 (1998).
28. J.R. Haines, J. Alonso, B.R. Appleton, T.A. Gabriel, 'Overview of the National Spallation Neutron Source', *Am. Nucl. Soc.*, **76**, 377–378 (1997).
29. 'Theme 1996: Neutrons for research; SINQ' Paul Scherrer Institute Annual Report 1996, 19–33 (1996).
30. A.G. Hanna, R.M. Brugger, M.D. Glascock, 'The Prompt Gamma Neutron Activation Analysis Facility at MURR', *Nucl. Method Instrum.*, **188**, 619–627 (1981).
31. M.R. Wormald, 'Pair Spectrometer NaI(Tl) Array for Neutron-induced Prompt Gamma-ray Analysis', *Nucl. Geophys.*, **3**(4), 373–380 (1989).
32. T. Gozani, G. Reynolds, H. Bozorgmanesh, E. Elias, T. Maung, V. Orphan, 'Nuclear Assay of Coal', Electric Power Research Institute, Palo Alto, CA, Vol. 1, 1980.
33. T. Gozani, 'Physics of Recent Applications of PGNAA for On-line Analysis of Bulk Minerals', Proceedings of the 5th International Symposium on Capture Gamma-ray Spectroscopy Related Topics, Knoxville, TN, American Institute of Physics, Woodbury, NY, 1984.
34. W. Herzog, 'Prompt Gamma Neutron Activation Analysis of Hard Coal, Raw Lignite and a Raw Glass Mixture', *Nucl. Geophys.*, **3**(4), 467–473 (1989).
35. F. de Corte, 'The k_0-standardization Method. A Move to the Optimization of Neutron Activation Analysis', Aggrégé Thesis, University of Gent, 1987.
36. R.L. Paul, 'The Use of Element Ratios to Eliminate Analytical Bias in Cold Neutron Prompt Gamma-ray Activation Analysis', *J. Radioanal. Nucl. Chem.*, **191**(2), 245–256 (1995).
37. R.L. Paul, R.M. Lindstrom, 'Measurement of k_0-Factors of Prompt Gamma-ray Activation Analysis', 2nd International k_0 Users Workshop, Ljubliana, Jozef Stefan Institute, Ljubliana, Slovenia, 55, 1996.
38. R.M. Lindstrom, R. Fleming, R. Paul, E. Mackey, 'The k_0 Approach in Cold Neutron Activation Analysis', Proceedings of the International k_0 Users Workshop, Gent, ed. F. de Corte, Institute of Nuclear Science, University of Gent, 121, 1992.
39. M. Rossbach, F. Corte, 'Introducing the k_0-Factor into Prompt Gamma Cold Neutron Activation Analysis', Proceedings of the International k_0 Users Workshop, Gent, ed. F. de Corte, Institute of Nuclear Science, University of Gent, 125, 1992.
40. C.M. Shyu, R.P. Gardner, K. Verghese, 'Development of the Monte Carlo-library Least Squares Method of Analysis for Neutron Capture Prompt Gamma-ray Analyzers', *Nucl. Geophys.*, **7**(2), 241–267 (1993).
41. R.P. Gardner, P. Guo, Q. Ao, C.L. Dobbs, 'Black Box Radiation Gauges and Analyzers: Dream Or Reality?', *Appl. Radiat. Isot.*, **48**(10–12), 1273–1288 (1997).
42. R.P. Gardner, A. Sood, Y.Y. Wang, L. Liu, P. Guo, R.J. Gehrke, 'Single Peak Versus Library Least-squares Analysis Methods for the PGNAA Analysis of Vitrified Waste', *Appl. Radiat. Isot.*, **48**(10–12), 1331–1335 (1997).
43. C.G. Clayton, C.F. Coleman, 'Analysis of Neutron Flux Distributions Generated by Sources of Fast Neutrons in a Range of Bituminous Coals', *Int. J. Appl. Radiat. Isot.*, **36**(10), 757–788 (1985).
44. C. Oliveira, J. Salgado, F.G. Carvalho, 'Optimization of PGNAA Instrument Design for Bulk Coal Analysis', *Nucl. Geophys.*, **7**(2), 285–303 (1993).
45. C. Oliveira, J. Salgado, 'Calibration Curves of a PGNAA System for Cement Raw Material Analysis using the MCNP Code', *Appl. Radiat. Isot.*, **40**(12), 1685–1689 (1998).
46. J.E. Briesmeister (ed.), 'MCNP – A General Monte Carlo Code for N-Particle Transport Code, LA-12625-M, Version 4B Manual. RSICC, Oak Ridge, TN, 1997.
47. C.G. Clayton, J.S. Schweitzer, 'A Review of Aspects of Nuclear Geophysics'. *Nucl. Geophys.*, **7**(2), 143–171 (1993).
48. R.C. Odom, S.M. Bailey, R.D. Wilson, 'Benchmarking Computer Simulations of Neutron-induced Gamma-ray Spectroscopy for Well Logging', *Appl. Radiat. Isot.*, **48**(10–12), 1321–1328 (1997).
49. R. Hertzog, L. Colson, B. Seeman, M. O'Brien, D. McKeon, P. Wraight, J. Grau, E. Ellis, J. Schweitzer, M. Herron, 'Geochemical Logging with Spectrometry Tools', Technical Symposium, Society of Petroleum Engineers, Dallas, Paper SPE16792, Richardson, TX, 1987.
50. J.A. Grau, J.S. Schweitzer, D.V. Ellis, R.C. Hertzog, 'A Geological Model for Gamma-ray Spectroscopy Logging Measurements', *Nucl. Geophys.*, **3**(4), 351–359 (1989).
51. Y. Idris, I.I. Funtua, I.M. Umar, S.B. Elegba, 'Neutron Activation Analysis of Aluminium with an Am-Be Neutron Source-application for Bauxite Investigation in Nigeria', *Appl. Radiat. Isot.*, **49**(1–2), 41–42 (1998).
52. R.P. Gardner, P. Guo, A. Sood, C.W. Mayo, R.J. Gehrke, C.L. Dobbs, 'Monte Carlo Aided Treatments of the Non-linear Inverse PGNAA Measurement Problem for Various Continuous On-line Applications', *J. Radioanal. Nucl. Chem.*, **233**(1–2), 105–107 (1998).
53. 'Nuclear Techniques in the Coal Industry', Proceedings of a Final Research Coordination Meeting Held in Krakow, IAEA-TECDOC-845, 1994.
54. R.C. Woodward, 'On-line Elemental Analysis', *World Cement*, **20**(8), 272–274 (1989).
55. K.J. Ellis, 'In Vivo Activation Analysis: Present and Future Prospects', *J. Radioanal. Nucl. Chem.*, **169**(2), 291–300 (1993).
56. S.A. Natto, D.G. Lewis, S.J.S. Ryde, 'Benchmarking the MCNP Code for Monte Carlo Modelling of an In Vivo

Neutron Activation Analysis System', *Appl. Radiat. Isot.*, **49**(5–6), 545–547 (1998).

57. A.R. Dulloo, F.H. Ruddy, T.V. Congedo, J.G. Seidel, R.J. Gehrke, 'Detection Limits of a Laboratory Pulsed Gamma Neutron Activation Analysis System for the Non-destructive Assay of Mercury, Cadmium, and Lead', *Nucl. Technol.*, **123**, 103–112 (1998).
58. Z.P. Sawa, 'PFN GASGA Techniques for Detection of Explosives and Narcotics', International Conference Proceedings on Nuclear Techniques for Analytical and Industrial Applications, eds. G. Vourvopoulos, T. Paradellis, Mykonos, June, 1992, 335–358, 1992.
59. T. Gozani, 'Neutron Based Non-intrusive Inspection Techniques', International Conference on Neutron in Research and Industry, ed. G. Vourvopoulos, Mykonos, June 1996, 174–181, 1997.
60. D.L. Anderson, W.C. Cunningham, G.H. Alvarez, 'Multielement Analysis of Food by Neutron Capture Prompt Gamma-ray Activation Analysis', *J. Radioanal. Nucl. Chem.*, **167**(1), 139–144 (1993).
61. A. Kuno, M. Matsuo, B. Takano, C. Yonezawa, H. Matsue, H. Sawahata, 'Neutron Induced Prompt Gamma-ray and Instrumental Neutron Activation Analyzes of Urban Estuarine Sediments', *J. Radioanal. Nucl. Chem.*, **218**(2), 169–176 (1997).
62. A.J. Caffrey, R.J. Gehrke, R.C. Greenwood, J.K. Hartwell, K.M. Krebs, G.D. McLaughlin, 'US Army's Experience with the PINS Chemical Assay System', *Trans. Am. Nucl. Soc.*, **72**, 117 (1995).
63. T.H. Prettyman, L.A. Foster, S. Parish, 'Applications of PGNAA to Plutonium Surveillance', *Trans. Am. Nucl. Soc.*, **77**, 146–147 (1997).
64. N.M. Spyrou, J.M. Sharaf, S. Rajeswaran, E. Mesbahi, 'Determination of the Elemental Distribution in a Sample Using Neutron Induced Gamma-ray Emission Tomography', *J. Radioanal. Nucl. Chem.*, **217**(2), 273–241 (1997).

Radiochemical Neutron Activation Analysis

Kaj Heydorn*
Risø National Laboratory, Roskilde, Denmark

1 Introduction **1**
1.1 Principle and Definition of Radiochemical Neutron Activation Analysis 1
1.2 Relationship to Instrumental Neutron Activation Analysis 2
1.3 Advantages and Limitations 2
2 Theory and Operating Principles **2**
2.1 Sample and Comparator 3
2.2 The Activation Process 4
2.3 Radiochemical Separation 4
2.4 Methodology of Counting 5
2.5 Data Processing 6
2.6 Recovery Correction 7
2.7 Final Calculation of Analytical Result 8
3 Instrumentation **10**
3.1 Irradiation Systems 10
3.2 Counting Equipment 11
4 Quality Assurance **11**
4.1 Statistical Control 12
4.2 Verification Methods 13
4.3 Traceability Chain 14
4.4 Uncertainty Budget 15
4.5 Statement of Uncertainty 16
5 Perspective and Future Developments **16**
5.1 Analytical Metrology 17
5.2 Certification Analysis 17
5.3 Sampling Constants 17
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Radiochemical neutron activation analysis (RNAA) is a method for trace-element determination based on the measurement of an indicator radionuclide, chemically separated from a neutron-activated sample. Unlike instrumental neutron activation analysis (INAA) the RNAA is therefore not a nondestructive method, and at the same time the amount of work put into the analysis of each sample makes RNAA much more expensive than INAA. This is the cost that has to be paid for determining selected elements at much lower concentrations and superior accuracy, while maintaining matrix independence and most of the other excellent characteristics of INAA. One of these characteristics is the small self-shielding during irradiation with neutrons, which opens the possibility of analyzing also large samples without any pretreatment.

Particularly for the lowest levels of trace elements RNAA provides several important advantages over alternative analytical methods:

a) the absence of a reagent blank, and
b) the corresponding insensitivity to contamination after activation
c) the controlled addition of carrier to avoid losses by adsorption etc., and
d) the corresponding possibility of determining chemical yield or recovery.

When carefully and thoughtfully carried out, RNAA qualifies as a so-called definitive method, one of the reasons being its direct traceability to a comparator of the pure element to be determined.

1 INTRODUCTION

The use of neutrons for analytical purposes goes back to the classical paper by Hevesy and Levi,[1] who discovered in 1936 in Copenhagen that activation of a sample of presumably pure Y_2O_3 by means of a Ra–Be neutron source gave rise to a radionuclide that could only be attributed to Dy; by comparing with samples of pure Dy they were even able to determine the actual quantity of Dy in the particular sample.

1.1 Principle and Definition of Radiochemical Neutron Activation Analysis

Although himself a chemist Hevesy never applied radiochemistry to any activated samples for analytical purposes; this was first done by Seaborg and Livingood,[2] who determined Ga after the addition of carrier and radiochemical separation of iron activated at the Berkeley cyclotron.

RNAA is a method for the determination of elements, and its full power was not realized until after the second World War, when the availability of nuclear reactors was brought to the attention of the analytical community. A clear presentation of the principles of RNAA, including carrier addition and yield determination, was

* Current affiliation: Technical University of Denmark, Lyngby, Denmark

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

first made by Boyd,[3] who still provides the fundamental description of what was later referred to as *comprehensive* RNAA:[4]

1. activation with neutrons
2. carrier addition
3. radiochemical separation
4. measurement of activity
5. determination of yield.

After half a century RNAA has reached maturity, the principles are well understood and no major technological developments are expected to take place in the foreseeable future. At the same time, it is also improbable that any other method will be found that could provide the same special characteristics so valuable for ascertaining the quality of trace element analysis.

1.2 Relationship to Instrumental Neutron Activation Analysis

The outstanding characteristic of neutron activation analysis in comparison with other methods of analysis is the timescale associated with the lifespan of the excited species or characteristic radionuclides, often referred to as *indicators*, which permit the introduction of a chemical separation between the end of activation and the beginning of measurement.

In purely instrumental neutron activation analysis this decay period is used only for simple physical operations, such as separating the sample from its container, taking an aliquot by volume or weight, cleaning the surface of the sample or adjusting its shape or volume before measurement is begun. In INAA the preservation of the integrity of the sample permits the determination of many elements in the same sample.

The radiochemical separation carried out during the decay period is characteristic of RNAA and serves to reduce or eliminate unwanted radionuclides interfering with the measurement of the indicator, so as to minimize the uncertainty associated with the determination of a particular element. RNAA is therefore basically a method for the determination of only a small number of elements in the same sample.

Over the years successful attempts have been made to reduce the need for radiochemical separations by advances in instrumentation for counting, supplemented with the use of γ-radiation or charged particles rather than neutrons for activation.[5] For highest accuracy at the lowest levels RNAA is still needed as a reference method.

1.3 Advantages and Limitations

RNAA is clearly not a nondestructive method, and at the same time the amount of work put into the analysis of each sample makes RNAA much more expensive than INAA. This is the cost that has to be paid for determining selected elements at much lower concentrations and superior accuracy in comparison with INAA, while maintaining most of the other excellent characteristics of this method, see **Instrumental Neutron Activation Analysis**, including the small self-shielding during irradiation with neutrons, which opens up the possibility of analyzing large samples as well without any pretreatment.[6]

Particularly for the lowest levels of trace elements RNAA provides several important advantages over alternative analytical methods:

1. absence of a reagent blank
2. corresponding insensitivity to contamination after activation
3. controlled addition of carrier to avoid losses by adsorption etc.
4. corresponding possibility of determining chemical yield or recovery.

The presence of a controlled amount of carrier during the radiochemical separation process, regardless of the original content in the sample, makes the a priori precision in the uncertainty budget independent of the concentration of the element to be determined, the *determinand*, see section 4.4. Any method that has been found to be useful for chemical separation of the indicator element from the material to be analyzed by any other analytical method, could also be used in RNAA. In this article we are therefore concerned only with the evaluation of the *performance characteristics* of the potential separation methods (see section 2.3) not with the methods themselves.

When carefully and thoughtfully carried out, RNAA qualifies as a so-called definitive method,[7] one of the reasons being its direct traceability to a comparator of the pure element to be determined.

2 THEORY AND OPERATING PRINCIPLES

The special features of RNAA can be expressed as a combination of separate, almost independent unit operations, and Figure 1[8] illustrates the entire process from the sample to be analyzed to the result of the analysis.

The RNAA sensitivity S given in Equation (1) for a particular determinand element X is expressed as the number of counts per unit weight of X, and it depends upon the half-life of its indicator $T_{1/2}$, the duration of the activation t_i, the decay time t_c, and the counting time t_m

$$S = \frac{C_x(1 - e^{-\lambda t_i})\, e^{-\lambda t_c}\, (1 - e^{-\lambda t_m})}{\lambda} \qquad (1)$$

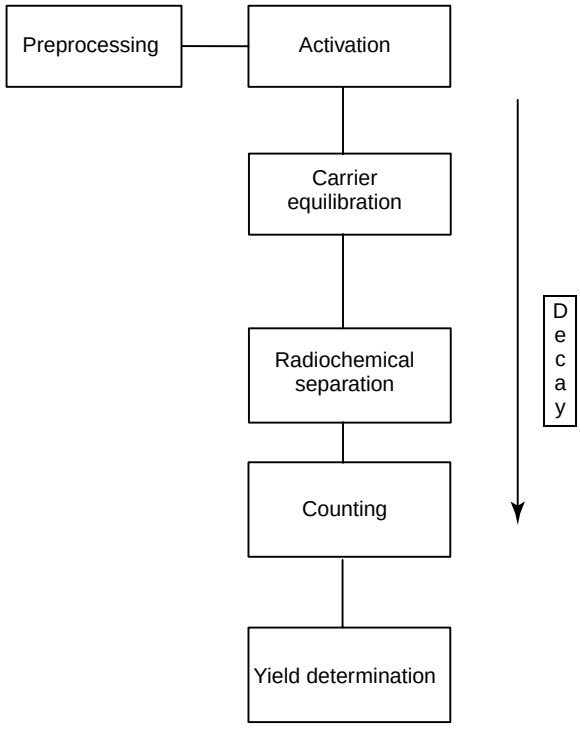

Figure 1 Schematic representation of RNAA methodology.

where $\lambda T_{1/2} = \ln 2$, and the rate constant (Equation 2)

$$C_x = P_i R_c P_m \qquad (2)$$

has to be determined experimentally from a known quantity of *comparator*.

The rate of formation P_i is proportional to the neutron flux density during activation, see section 2.2, but its magnitude varies by orders of magnitude from element to element. The recovery R_c and the detection efficiency P_m will be discussed later.

Table 1 lists those elements with corresponding indicators which have rate constants larger than that of Zn.[9a]

Table 1 Elements with high sensitivity in neutron activation analysis

Element	Indicator	Half-life	γ-Energies (keV)
Aluminum	^{28}Al	2.2 min	1779
Antimony	^{122}Sb	2.8 d	564
	^{124}Sb	60.2 d	603, 1691
Arsenic	^{76}As	26.5 h	559, 657
Bromine	^{82}Br	35.5 h	554, 777, 619
Cesium	^{134}Cs	2.06 y	569, 605, 796
Cobalt	^{60}Co	5.27 y	1173, 1333
Copper	^{64}Cu	12.7 h	511
Dysprosium	^{165}Dy	2.33 h	95, 280, 362
Europium	^{154}Eu	8.59 y	723, 873, 1275
Gallium	^{72}Ga	14.1 h	834, 2202
Gold	^{198}Au	2.69 d	412
Hafnium	^{175}Hf	70 d	344
	^{181}Hf	42.4 d	134, 482
Holmium	^{166}Ho	26.9 h	81, 1379
Indium	^{116}In^{m}	54 min	1097, 1294
Iodine	^{128}I	25.1 min	443
Lanthanum	^{140}La	40.2 h	329, 487, 1596
Manganese	^{56}Mn	2.58 h	847, 1811, 2113
Mercury	^{203}Hg	46.6 d	279
Praseodymium	^{142}Pr	19.1 h	1576
Rhenium	^{186}Re	90.6 h	137
Ruthenium	^{103}Ru	39.3 d	497
Samarium	^{153}Sm	46.8 h	103
Scandium	^{46}Sc	83.8 d	889, 1121
Selenium	^{75}Se	120 d	136, 265
Silver	^{110}Ag^{m}	250 d	658, 885, 1384
Sodium	^{24}Na	15.0 h	1368, 2754
Tantalum	^{182}Ta	115 d	1121, 1221
Terbium	^{160}Tb	73 d	299, 879, 965
Thorium	^{233}Pa	27 d	312
Tungsten	^{187}W	24 h	480, 686
Vanadium	^{52}V	3.8 min	1434
Ytterbium	^{175}Yb	4.2 d	283, 396
Zinc	^{65}Zn	244 d	1115

2.1 Sample and Comparator

In RNAA the analytical sample is just as prone to contamination before activation, as it is in all other analytical methods. The characteristic insensitivity to contamination applies only after completion of the activation process, and it is therefore important that any conditioning or preprocessing of a sample be kept at an absolute minimum.

Some reactors require biological samples to be dried or lyophilized before activation in order to reduce formation of gases by radiolysis of the water content; this may, however, require special tracer studies to ascertain that no loss of determinand is incurred.[10] At low levels of determinand any such preprocessing necessitates the introduction of a blank to be analyzed together with a sample in order to monitor possible contamination occurring before activation.

Although it is possible to carry out neutron activation analysis without activating an isotopic comparator together with the sample,[9a] this is rarely done for RNAA. The limited number of elements that can be determined in the same sample makes the use of individual comparators much more attractive, because of the greater freedom it gives with respect to both activation and counting conditions.

A comparator is prepared from a weighed quantity of the element to be determined or a compound with an accurately known stoichiometry,[11] chosen so that no other significant activity will be present at the time of measurement.

The validity of the comparator requires that it be activated under exactly the same conditions as the sample with respect to both time and space. It must be physically as similar as possible to the sample at the time of counting, i.e. after completion of the radiochemical separation: same counting vial, same volume, similar matrix, etc. However, the comparator should *not* be processed, in order to eliminate any possible risk of loss, which means that $R_c \equiv 1$, and the rate constant for the comparator C_x^* is reduced to Equation (3)

$$C_x^* = P_i P_m \tag{3}$$

The comparator need not be identical or even similar to the sample, as long as it is exposed to the same fluence of neutrons. The sample, therefore, can be of any size or shape from up to perhaps several kilograms[12] to practically weightless quantities, still using the same comparator. In some cases it may be appropriate to weigh the analytical sample, as well as to take an aliquot of the comparator after completion of the activation.

2.2 The Activation Process

Activation of a sample by neutrons could be carried out with a number of devices, such as accelerators, cyclotrons, radioactive sources,[13] or even spallation sources; nuclear reactors remain, however, the most suitable source of neutrons for RNAA.

Research reactors are found in many countries,[14] and most of them offer irradiation services for neutron activation analysis with neutron flux densities up to 10^{14} neutrons $cm^{-2}\,s^{-1}$, which is more than sufficient for the determination of the elements listed in Table 1. More important is the spatial flux homogeneity that is needed to ascertain that samples and comparator irradiated together are exposed to exactly the same number of neutrons.

Irradiation of an element with a relative atomic mass A_r at a constant neutron flux density φ for a period of time t_i, produces a specific activity of:

$$\frac{N_A \sigma_a \varphi (1 - e^{-\lambda t_i})}{A_r} \text{Bq g}^{-1}$$

where σ_a is the appropriate cross-section of the element for the formation of a specific radionuclide, and N_A is the Avogadro number.

The *saturation activity* achieved after an infinite irradiation time is numerically equal to the rate of formation of the indicator nuclide (Equation 4)

$$P_i = \frac{N_A \sigma_a \varphi}{A_r} \tag{4}$$

Some small reactors with modest neutron flux densities like the Canadian Slowpoke type[15] may be so stable in time that sample and comparator need not even be irradiated simultaneously. In most nuclear reactors, however, changes in neutron flux exposure during irradiation may take place as a result of the introduction or removal of other samples, movement of control rods, change of power level, moderator temperature, or virtually intermittent operation, as well as many other things beyond the control of the analyst. These changes will affect indicator isotopes differently because of their different half-lives; however, the simultaneous irradiation of a comparator of the element to be determined, the determinand, will automatically account for such changes.

Subject to the needed sensitivity the duration of irradiation should be kept as short as possible[9b] in order to:

1. keep the total activity of the sample low enough for proper handling without excessive shielding and time-consuming use of remote handling equipment;
2. minimize radiation decomposition of the sample, which may result in the loss of halogens, mercury,[16] or other elements, when the irradiation capsule is opened;
3. reduce the influence of long-lived activated matrix components on the choice of counting conditions;
4. limit the nuclear interference from double neutron capture in some elements.[17]

2.3 Radiochemical Separation

During the period between the end of activation and the beginning of measurement the analytical sample is subjected to a radiochemical separation, which should be completed preferably within one mean life of the indicator.[9b] The performance characteristics of the radiochemical separation therefore must include the time needed to complete the task, which in practice limits the use of RNAA to elements with indicator half-lives in excess of a few minutes.

Sample decomposition in RNAA presents the same problems as in all other methods of chemical analysis;[18] to ascertain carrier equilibration the procedure should include a complete oxidation–reduction cycle. The process can hardly be completed in less than several minutes, which is just feasible for V in human serum with an indicator half-life of 3.8 min,[19] but barely enough for Al with 2.2 min half-life.[20]

The absorption of a neutron gives rise to the emission of one or more capture γ-rays with energies of several MeV, which causes a nuclear recoil energy of $536 \times E_\gamma^2 / A_r$ eV which is enough to break a chemical bond. Because of this Szilard–Chalmers effect the activated atom no longer retains its original position in a chemical entity, and RNAA is therefore incapable of determining any speciation of an element. By

the same token, however, postirradiation carrier equilibrium does not necessarily require complete sample decomposition, but only complete dissolution. This is an important difference from other reference methods, such as isotope dilution mass spectrometry (IDMS), that rely on complete isotopic exchange between an added isotopic tracer and the total content of an element in the sample.

The addition of carrier is required not only for all elements that need to be determined, or rather their indicators, but also for some interfering elements, from which a good separation is important. Such a *hold-back carrier* could prevent adsorption or coprecipitation that might result in poor and irreproducible separation from interfering radionuclides.

The aim of a radiochemical separation is to achieve a high and reproducible recovery of the indicator radionuclide at the same time as a very low yield of all other radionuclides. The efficiency of the separation is expressed by the *separation factor*,[21] K, which, as shown in Equation (5), is the ratio between the yield of the interfering species, r_i, and the recovery of the determinand

$$K = \frac{r_i}{R_c} \tag{5}$$

K is usually determined by adding a suitable radioactive tracer together with carriers to an unirradiated sample which is then processed and measured exactly as a normal sample.[22] Experimentally determined separation factors are presented in Table 2[9c] for a simple radiochemical separation of Cu in biological materials. A good separation will have separation factors better than 10^{-4}, but in extreme cases a separation factor better than 10^{-7} has been demonstrated.[38] It is worth pointing out that the use of highly effective radiochemical separations with good separation factors makes possible the determination of several other elements with less sensitivity than those listed in Table 1, such as Cr, Fe, Mo, and Cd.[22,23]

Complete separation of the indicator from all other radionuclides in the activated sample is required when the indicator has to be measured by its emission of β-particles in order to achieve sufficient sensitivity.[24] In most cases of γ-emitting indicators a high separation factor is needed only for the dominant interfering elements. Such *interference removal* is often needed for activated matrix components such as ^{24}Na, ^{42}K, and ^{82}Br that dominate the γ-spectra of activated biological samples for days or weeks after the end of irradiation. Successful removal of ^{24}Na is possible by means of hydrated antimony pentoxide,[25] and other matrix indicators may be removed by precipitation with thioacetamide, extraction of Fe with ether, or other conventional methods of chemical separation.[26] It should be recalled that nuclear interference from matrix elements cannot be eliminated by a radiochemical separation, but requires a preirradiation separation.

Each combination of determinand element and matrix composition may have its own optimum radiochemical separation method that minimizes the effort and maximizes the quality of the final result. A general discussion of the factors involved in such an optimization must be based on the separation factors needed for the particular analytical task. This paves the way for the use of less specific *group separations*, the use of *scavengers* to collect the elements to be determined, adsorption on solid oxides, activated carbon etc. Some of these methods seem to work best without carrier addition, but without determination of the recovery of the indicator these methods would suffer from an intrinsic lack of credibility.

2.4 Methodology of Counting

A radiochemically separated sample that contains only the indicator radionuclide could be measured by any suitable instrument with a sufficiently high detection efficiency, and in this respect it would be hard to beat a simple Geiger-Müller (GM) counter or liquid scintillation counter which detects β-particles with close

Table 2 Experimentally determined interferences in a method for the determination of Cu in biological materials by RNAA[9c]

Interfering element	Radionuclide	Separation factor K	Effective value f	mg of element equivalent to 1 µg of Cu
Sodium	^{24}Na	1.4×10^{-5}	0.033	2000
Zinc	^{65}Zn	1.4×10^{-3}	−0.016	40
Gallium	^{72}Ga	3.7×10^{-1}	−0.13	0.02
Arsenic	^{76}As	1.3×10^{-3}	0.93	0.8
Bromine	^{80}Br^{m}	5.0×10^{-5}	0.32	60
Antimony	^{122}Sb	2.6×10^{-2}	0.27	0.15
Tungsten	^{187}W	1.5×10^{-3}	0.96	0.7

to 100% efficiency. It is, however, possible only in very special cases[24,27] to prove that the sample does not contain any other radionuclides that might influence the measurement. In other cases the required high radionuclidic purity may be achieved only at the cost of a lower recovery of the indicator.

Most radiochemically separated samples therefore contain other radionuclides than the indicator and therefore must be measured with instruments that can distinguish between the radionuclides present, usually by measurement of a complete γ-spectrum, where individual radionuclides can be identified by their different γ-energies.

Among such detectors the Na(Tl) scintillation detector is available in the largest sizes, which may even be used to surround the sample; it therefore offers the highest detection efficiency, P_m. Better γ-energy resolution is obtained with the germanium semiconductor detector, which will reduce interference from accompanying radionuclides, but in return carries a higher cost and a lower efficiency.

From the point of view of minimum interference the *optimum decay time*[9b] from end of activation to measurement is given by Equation (6)

$$t_c = \frac{1}{\lambda} - t_i \tag{6}$$

when the indicator is closer to saturation than any of the other radionuclides.

For the sensitivity of the analysis a lower efficiency P_m can be compensated by an increase in *counting time* t_m. However, the increasing influence of radionuclides with longer half-lives on the measurement increases the uncertainty of the measurement of the indicator, so that for a small signal on a high base line nothing is gained[9d] by counting for more than approximately 1.81 half-lives.

Good detection efficiency requires both samples and comparators to be counted close to the detector; in some cases it may even be expedient to count in a well counter. This results in considerable γ-ray and X-ray coincidence losses and corresponding, but identical, changes in the spectra of sample and comparator. Unless such coincidences give rise to spectral interference, the close counting geometry creates no bias in the determination of the indicator. However, P_m becomes highly dependent upon small differences in counting geometry, as well as on the uniformity of the materials. The reproducibility of the counting geometry should therefore be carefully checked,[28] and the identity of the sample and comparator geometry must be ascertained. For optimum results it is recommended that all samples and comparators be counted in liquid form, so as to ensure homogeneity; by measuring the same volumes in identical containers the counting geometry can be kept constant within <1%.

2.5 Data Processing

The processing of γ-spectra in RNAA is very different from the methods employed in INAA (see **Instrumental Neutron Activation Analysis**).

First, we are only interested in the indicator nuclide and therefore need not identify all other radionuclides in the spectrum, unless they contribute to the spectrum in the vicinity of the indicator γ-energies, such as those listed in Table 1.

We are, however, interested in the best possible quantitation of the indicator with the minimum uncertainty and without interference, even close to the limit of detection. In the evaluation of small-to-invisible photo peaks on a high Compton continuum, the determination of the number of counts that can with a high degree of certainty be attributed to the indicator is best carried out by simple peak integration according to Covell, and not by fitting the small number of counts to an assumed Gaussian or other peak shape.

The optimum choice of peak boundaries for peak area evaluation depends upon the full width at half-maximum (FWHM) and the relative height of the photo peak to the base continuum. The optimum half-width has a limiting value of 0.90758 FWHM for photo peak heights approaching zero, and Figure 2 shows optimum half-widths for peak area integration as a function of relative peak heights and FWHM.[29]

The FWHM is a function of γ-energy E

$$(\mathrm{FWHM})^2 = wE + w_0 \tag{7}$$

where the coefficients w and w_0 in Equation (7) are determined from the actual spectrum by weighted linear regression of FWHM^2 for the largest photo peaks on their corresponding γ-energies or channel numbers. These peaks would be the same as used for energy calibration by simple linear regression of γ-energies on channel numbers and do not necessarily include any photo peak of the indicator nuclide.

The FWHM corresponding to the chosen photo peak of the indicator is now used in combination with Figure 2 to select the peak boundaries for optimum precision of the partial peak area evaluation according to Covell. The partial peak area is then converted to the full peak area by assuming a Gaussian shape with the assumed FWHM, which is a purely mathematical operation that does not change the relative precision of the measurement. The final result of the data processing is the full peak area in counts together with its associated standard deviation, calculated on the basis of a Poisson distribution for each individual channel content. An implementation of this approach has been verified by comparison with IAEA (International Atomic Energy Agency) test spectra with

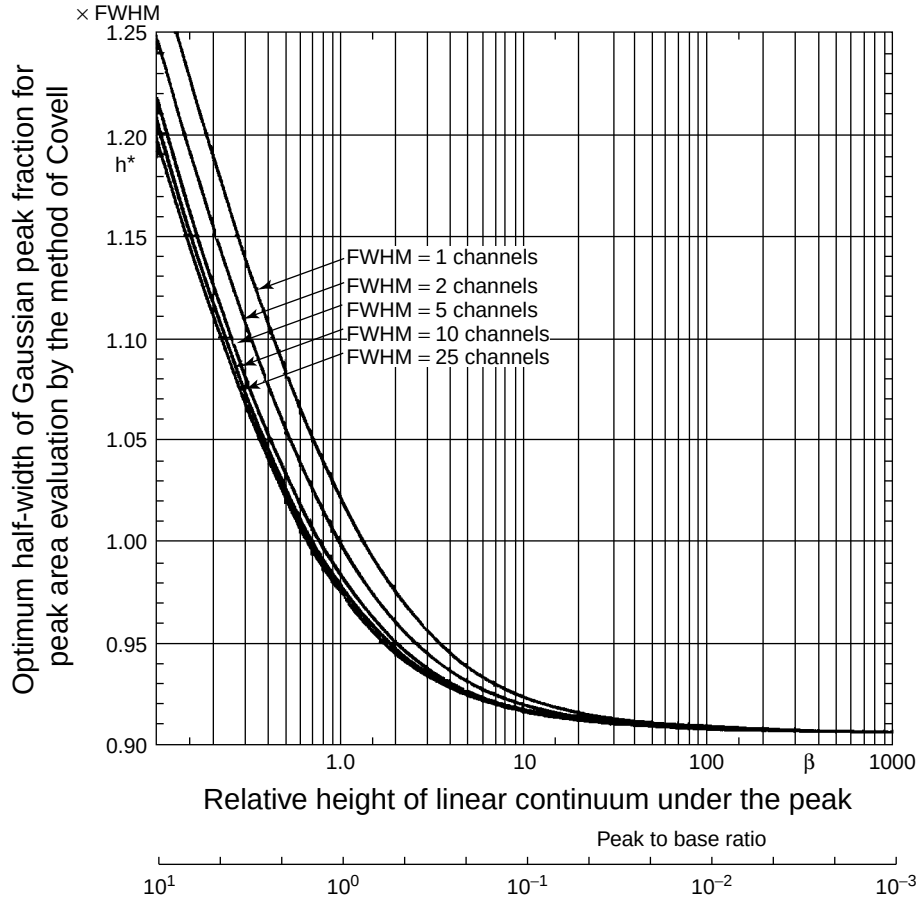

Figure 2 Optimum half-width of Gaussian peak fraction for peak area evaluation by the method of Covell as a function of height of linear continuum under the peak for selected values of FWHM. (Reprinted with permission from K. Heydorn, W. Lada, *Anal. Chem.*, **44**, 2314.[29] Copyright 1972, American Chemical Society.)

known peak areas, as well as with actual spectra with known ratios.[28]

The possible interference from other radionuclides is not estimated by attempting to resolve doublets or multiplets, but by the measurement of the effective values f of potentially interfering elements.[30] By irradiating a known quantity of an interfering element exactly like a sample together with a comparator of the determinand, and using the peak boundaries for the determinand, the *effective value* is determined as the ratio between counts per unit mass of interfering element relative to determinand. For elements with effective values in excess of 0.01 separation factors should be determined in order to specify the assumptions of the RNAA.

Table 2 lists effective values for seven elements suspected of interfering with the determination of Cu by RNAA.[9c] This allows the user to judge the performance characteristics of the method for the analysis of actual samples.

Interference from the formation of the indicator nuclide during activation from another element by (n,p), (n,α) reactions, or by double neutron capture, is usually referred to as *nuclear interference* and can be corrected for only by the simultaneous activation of an additional comparator of the interfering element.[17] A particularly treacherous interference occurs when the indicator nuclide is also produced by fission of uranium, which may necessitate a separate determination of this element[31] in order to correct for this positive bias.

Sometimes the absence of interference may be demonstrated by the determination of the same element using two different indicator nuclides.[32]

2.6 Recovery Correction

The determination of recovery or chemical yield of the analytical separation is an outstanding property of RNAA and enables the method to estimate the true content of the determinand in the sample.

In routine applications of highly standardized radiochemical separations the chemical yields are often determined by adding radioactive tracers of the determinands

to an inactive sample before performing the separation, exactly as described for the determination of interferences. The use of such a priori chemical yields instead of individual recoveries is justified only when the yields are close to 100% and the correction therefore quite small.[33]

In research applications, where every analysis might differ from the rest, recovery is usually not quantitative and is susceptible to considerable variation from sample to sample, which means that recoveries must be determined for each individual analysis. In that case much larger corrections for chemical yield can be accepted, while still maintaining statistical control.

A comparison of the amount of carrier in the separated sample with the originally added quantity by almost any analytical method could be used for the determination of recovery, and in many applications gravimetry has been used. However, such chemical methods might jeopardize the principle of *not* processing the added carrier reference comparator sample, and therefore a direct instrumental method without any additional treatment and a minimum of handling of the separated sample is clearly preferable, and simple radioanalytical methods based on either a short re-irradiation or the addition of a radioactive tracer, are often the methods of choice:

1. With carrier added at the milligram level the irradiation time can be very short, and the measurement of the induced radioactivity can be very simple.[34] By simultaneous irradiation and direct comparison between added and recovered carrier there is no need to know the amount of carrier accurately, provided the content in the sample is negligible in comparison to the added quantity of carrier.
2. Addition of a radioactive tracer together with the carrier opens the possibility of determining recovery at the same time as measurement of the indicator. Clearly, the tracer should not be produced by neutron activation and the amount added should be chosen carefully in order not to disturb the measurement.[9e] A list of radionuclides used as tracers for the determination of individual recoveries of added carrier is presented in Table 3. If they are truly carrier-free they may even be added to the sample before activation, so that any losses that might occur as a result of preirradiation processing would be included in the recovery.[10]
3. Instead of adding the radioactive tracer to the sample before separation it may be added after completion of the separation and the measurement of the indicator, in order not to cause spectral interference. It is now possible to correct for recovery during separation by comparing the specific activities of the carriers added to the separated sample and the carrier reference comparator sample by using *substoichiometric* methods.[35]

It has been proposed that an internal standard of another element be used to correct for recovery;[36] however, this is limited to the cases where this element yields a radionuclide isotopic with the indicator nuclide.

If carefully thought out, the individual determination of recovery for each sample requires very little additional effort in comparison to the radiochemical separation.[37]

2.7 Final Calculation of Analytical Result

The calculation of results obtained by RNAA is basically similar to that of INAA; however, the proper utilization of the various measurements for correcting initial results and estimating their uncertainty may become quite complicated. It is, therefore, strongly recommended that

Table 3 Indicators and tracers for individual yield corrections in RNAA

Determinand	Indicator	Half-life	γ-Energy (keV)	Tracer	Half-life	γ-Energy (keV)
Antimony	^{122}Sb	2.8 d	564	^{125}Sb	2.8 y	408
Arsenic	^{76}As	26 h	559	^{74}As	17.8 d	596
Cadmium	^{115}Cd	53.5 h	528	^{109}Cd	463 d	88
Cobalt	^{60}Co	5.3 y	1332	^{57}Co	271 d	122
Copper	^{64}Cu	12.7 h	511	^{67}Cu	62 h	184
Gallium	^{72}Ga	14 h	834	^{67}Ga	78 h	93
Gold	^{198}Au	2.7 d	412	^{195}Au	183 d	99
Iodine	^{128}I	25 min	443	^{126}I	13.0 d	388
Manganese	^{56}Mn	2.6 h	847	^{54}Mn	312 d	835
Mercury	^{197}Hg	65 h	78	^{203}Hg	47 d	279
Selenium	^{75}Se	120 d	265	^{81}Se^{m}	57 min	103
Tin	^{123}Sn^{m}	40 min	160	^{113}Sn	115 d	392
Tungsten	^{187}W	24 h	480	^{181}W	121 d	65
Uranium	^{239}Np	56 h	β	^{237}Np	2×10^{6} y	α
Vanadium	^{52}V	3.8 min	1434	^{48}V	16.0 d	984

calculation be done in the form of a spread-sheet, not only in order to avoid making mistakes, but also because it facilitates the implementation of quality control, as well as the smooth propagation of errors from one stage to the next.

As an example let us pretend that we have carried out RNAA for two determinands in the same sample, and for simplicity let us assume that both can be determined from the same γ-spectrum, but with some interference from one indicator to the other. We also assume that comparators of both determinands were irradiated together with the sample. The data available for calculating the concentration of determinands in a sample may be organized as shown in Table 4.

Calculations are now carried out in the following sequence:

1. Photopeak counts are corrected to count rates at the time of pile-out by multiplication with a decay correction for

Table 4 Data used to calculate concentration of determinands

Cell content	Column
Sample identification	A
Irradiation time t_i	B
Date and time of pile-out	C
Weight of sample	D
Recovery 1	E
estimated standard deviation	F
Recovery 2	G
estimated standard deviation	H
Counting time t_m	K
Date and time of counting	L
Photopeak counts 1. determinand	M
estimated standard deviation	N
Photopeak counts 2. determinand	Q
estimated standard deviation	R
Photopeak counts 3. determinand	U
estimated standard deviation	V
Comparator 1. determinand	AA
Content of determinand	AD
Counting time t_m^*	AK
Date and time of counting	AL
Photopeak-1 counts	AM
estimated standard deviation	AN
Photopeak-2 counts	AQ
estimated standard deviation	AR
Comparator 2. determinand	BA
Content of determinand	BD
Counting time $t_m^\#$	BK
Date and time of counting	BL
Photopeak-1 counts	BM
estimated standard deviation	BN
Photopeak-2 counts	BQ
estimated standard deviation	BR

t_c = date and time of counting − date and time of pile-out

$$\text{decay correction} = \frac{\lambda e^{\lambda t_c}}{1 - e^{-\lambda t_m}} \quad (8)$$

In Equation (8) λ is the decay constant for the radionuclide associated with the photopeak in question.

Decay corrections are conveniently placed in the columns immediately after the counting data:

	Columns		
• Sample	O	S	X
• Comparator	AO	AS	
• Comparator	BO	BS	etc.

2. The uncorrected content of 1 determinand in the separated sample and its corresponding standard deviation are calculated as:

$$\text{content} = \text{AD} \times \frac{\text{M}}{\text{AM}} \times \frac{\text{O}}{\text{AO}}$$

and placed in column CA

$$\text{CV}^2 = \left(\frac{\text{N}}{\text{M}}\right)^2 + \left(\frac{\text{AN}}{\text{AM}}\right)^2$$

$$\text{standard deviation} = \text{SQRT}(\text{CV}^2) \times \text{content}$$

and placed in column CB

3. The uncorrected content of 2 determinand in the separated sample and its corresponding standard deviation are calculated as:

$$\text{content} = \text{BD} \times \frac{\text{Q}}{\text{BQ}} \times \frac{\text{S}}{\text{BS}}$$

and placed in column CC

$$\text{CV}^2 = \left(\frac{\text{R}}{\text{Q}}\right)^2 + \left(\frac{\text{BR}}{\text{BQ}}\right)^2$$

$$\text{standard deviation} = \text{SQRT}(\text{CV}^2) \times \text{content}$$

and placed in column CD

4. Correction of the influence of 2 determinand on the result for the 1 determinand requires the calculation of the effective value:

$$\text{effective value} = \frac{\text{AD}}{\text{BD}} \times \frac{\text{BM}}{\text{AM}} \times \frac{\text{BO}}{\text{AO}}$$

and placed in column CE

$$\text{CV}^2 = \left(\frac{\text{AN}}{\text{AM}}\right)^2 + \left(\frac{\text{BN}}{\text{BM}}\right)^2$$

$$\text{standard deviation} = \text{SQRT}(\text{CV}^2) \times \text{effective value}$$

and placed in column CF

5. The corrected content of the 1 determinand in the separated sample is now calculated as:

   content = CA − CC × CE

   and placed in column CG

   $$CV^2 = \left(\frac{R}{Q}\right)^2 + \left(\frac{BR}{BQ}\right)^2 + \left(\frac{BN}{BM}\right)^2$$

   $$\text{standard deviation} = \text{SQRT}\left(CA \times \frac{N}{M}\right)^2 + CV^2(CC \times CE)^2 + \left(CG \times \frac{AN}{AM}\right)^2$$

   and placed in column CH

   At this stage it should be checked whether CG is negligible in comparison with the quantity of added carrier; if not the recovery E should be reduced by subtracting their ratio.

6. The real concentration of determinand in the sample material and its uncertainty are finally calculated from the original sample weight and the recovery of the appropriate carrier in the separated sample:

   $$\text{concentration} = \frac{CG}{D \times E}$$ is placed in column CK

   $$CV^2 = \left(\frac{F}{E}\right)^2 + \left(\frac{CH}{CG}\right)^2$$

   standard deviation = SQRT(CV^2) × ABS(CK)

   and placed in column CL

An example of actual results obtained from the determination of Pt in a reference material from the Bureau Communautaire de Réference (BCR) certified reference material (CRM) 186 pig kidney[38] is shown in Table 5, which also indicates the spreadsheet columns, from which these results are taken. The indicator for Pt is ^{199}Au, formed by the decay of ^{199}Pt, but also from neutron capture in ^{198}Au, the indicator for gold. The correction for nuclear interference is in this case comparable to the total amount of ^{199}Au and therefore leads to very large relative uncertainties in the final results for Pt.

Corrections for nuclear interference from uranium in the determination of La, Ce, Sm, Nd, Zr, Mo, Ru, Te, etc. may be carried out in a similar way,[9f] except that the concentration of U in the sample has to be determined separately.

This paradigm can easily be modified to accommodate other situations in RNAA with additional determinands and γ-spectra, using the same general approach to calculation of results and their uncertainties.

Corrections for possible blank values, such as those associated with preirradiation treatment of the sample,[33] must be carried out by calculating the blank value exactly as if it were a sample and then subtracting from the final result for the sample itself.

3 INSTRUMENTATION

Radiochemical separation requires a radioanalytical laboratory of Type C,[39] which is essentially just a good quality chemical laboratory with a radiochemical hood and a little shielding to reduce radiation exposure.

Requirements for irradiation facilities and counting equipment do not go beyond those needed for INAA, but the emphasis on some of the features and specifications is shifted because of the nature of the separated sample and the low levels of activity of the indicator nuclides.

3.1 Irradiation Systems

In order to benefit most from the properties of RNAA it should be possible to activate rather large samples without any preceding treatment that might lead to losses or contamination. As with INAA it is important that the entire sample and associated comparators be exposed to identical neutron flux densities for exactly the same time.

This is best achieved in a system that permits rapid transport in and out of the irradiation zone, while providing rotation of the irradiation container during activation.[40] In order to minimize sample decomposition

Table 5 Determination of the platinum content in BCR CRM 186 pig kidney by RNAA[38]

Sample size(g)	Recovery (%)	Apparent platinum (ng) ± standard deviation	Observed gold (ng) ± standard deviation	Effective value ratio ± standard deviation	Actual platinum (ng) ± standard deviation
2.23	76.4	13.6 ± 0.8	3.206 ± 0.004	4.50 ± 0.04	−0.9 ± 0.8
2.38	92.5	13.6 ± 0.5	3.713 ± 0.004	3.50 ± 0.03	0.6 ± 0.5
2.19	72.2	10.7 ± 0.9	3.161 ± 0.003	3.84 ± 0.03	−1.4 ± 0.9
2.25	95.7	18.5 ± 0.9	4.436 ± 0.005	4.36 ± 0.05	−0.8 ± 0.9
D	E	CA ± CB	CC ± CD	CE ± CF	CG ± CH

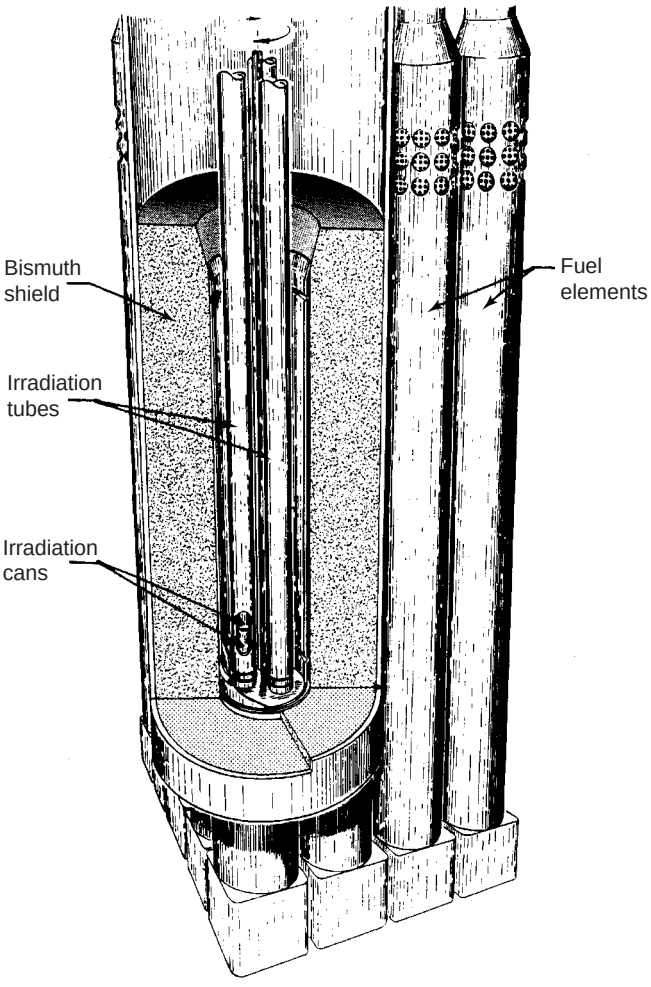

Figure 3 Cut-away of the rotating rig facility in the Danish DR 3 reactor. The three irradiation tubes in the rotating assembly are surrounded by 50 mm of bismuth in order to reduce γ-ray exposure. Carbon dioxide gas is used for transportation of the irradiation cans in and out of the reactor during operation.[9] (Reprinted with permission from K. Heydorn, *Neutron Activation Analysis*, Chapter 3, Uniscience 1984. Copyright CRC Press, Boca Raton, FL.)

the thermal neutrons should be accompanied by a minimum of epithermal and fast neutrons, as well as γ-radiation.

In Figure 3 is shown a drawing of such a facility in the Danish Reactor DR 3, which permits simultaneous irradiation of three containers of about 30 cm^3, while they are rotated in a thermal neutron flux density of 3×10^{13} neutrons cm^{-2} s^{-1} shielded by 50 mm of bismuth to reduce the γ-ray intensity.

3.2 Counting Equipment

A complete radiochemical separation of the indicator nuclide would allow use of the very sensitive but *nonspecific* detection of β-particles by very simple counting equipment. However, no separation is perfect, and in practice the less sensitive, but high-resolution detection of γ-rays is preferable, wherever possible.

With the availability of germanium semiconductor detectors having efficiencies comparable to, or even surpassing, that of the $3'' \times 3''$ sodium-iodide scintillation detector, the instrumentation for RNAA becomes similar to that needed for INAA with the same electronic equipment and multichannel analyzer with at least 16 K (K = 1024) channels.

The *detection efficiency* P_m determines the sensitivity of determination S, while an *improved resolution* of the detection system improves both the precision and accuracy of the determination by:

1. facilitating the detection of unexpected interferences
2. minimizing the correction for known spectral interferences
3. reducing the standard deviation of the photopeak area
4. improving the detection limit.

Similar improvements may sometimes be achieved by the use of two or more detectors in combination with coincidence/anticoincidence circuitry to obtain spectra providing improved detection limits with only a slight loss of sensitivity.[41]

As previously discussed the close counting geometry of the separated sample must be exactly equal to that of the comparator, calling for an accurately reproducible counting position to ascertain exactly equal counting efficiency and coincidence losses for both. A counting assembly with such reproducible counting positions for liquid samples is shown in Figure 4.

The usually low activity of the separated samples makes corrections for dead time and pile-up less critical, and a simple 50–60 Hz pulser with a narrow peak in the highest channels of the spectrum is a reliable way of making such corrections. The use of digital signal processing (DSP) improves resolution at high count rates, but seems to give no particular advantages at low count rates.

4 QUALITY ASSURANCE

While RNAA is potentially capable of providing analytical results without bias and with known uncertainty, the method clearly requires no less skill and experience than other methods. Quality assurance is therefore a sine qua non, and fortunately the method is particularly well suited for modern methods of quality control/quality assurance (QC/QA) based on the so-called BIPM (Bureau International des Poids et Mesures) philosophy.[42]

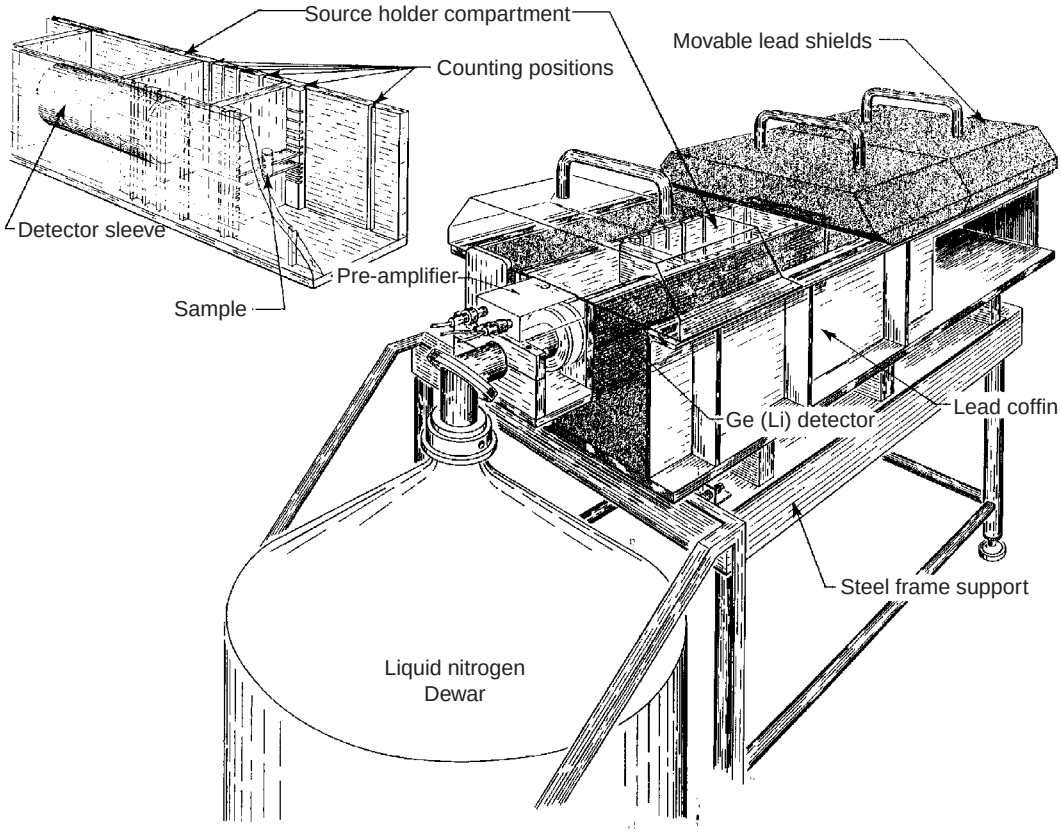

Figure 4 Horizontal lead coffin for counting with a germanium semiconductor detector. The perspex box shown as an insert illustrates that the modifications required by using different detectors do not affect the source compartment. The coffin offers 100 mm of lead shielding, but is open at the detector end that usually faces a wall. (Reprinted with permission from K. Heydorn, *Neutron Activation Analysis*, Chapter 3, Uniscience 1984.[9] Copyright CRC Press, Boca Raton, FL.)

The implementation of this line of thought is described in a basic ISO (International Organization for Standardization) document,[43] and for use in analytical chemistry in a Eurachem Guide;[44] none of these documents, however, provide any guidance with respect to RNAA or to activation analysis in general.

4.1 Statistical Control

The analytical method is in a state of statistical control when its known sources of variation fully account for the observed variability among replicate determinations.

The *a posteriori variability* of n replicate analyses is expressed as the variance s^2 of results y_i in Equation (9)

$$s^2 = \frac{\sum^{n}(y_i - y_0)^2}{n-1} \tag{9}$$

where y_0 is their mean value.

When the analytical method is completely defined, it is possible to identify and evaluate all sources of variation by statistical or other methods, referred to as uncertainty components of type A or B.[43] Their influence on the analytical results is calculated by the law of error propagation on the basis of variances $\widehat{\sigma}^2$ of p individual components x_i (Equation 10)

$$\widehat{\sigma}_0^2 = \sum_{1}^{p} \left(\frac{\partial y}{\partial x_i} \right)^2 \widehat{\sigma}_i^2 \tag{10}$$

This resulting variability may be calculated without actually carrying out any analyses and is therefore referred to as the *a priori precision*.[45]

In RNAA and other methods where the process of counting stochastic events is part of the analytical method the statistics of counting has to be compounded with the a priori precision, even though its contribution is not known until an actual measurement has been carried out. The calculation of *counting statistics* is based on the Poisson distribution for the number of counts detected from a radioactive source, and this assumption has been amply demonstrated to be fulfilled in neutron activation analysis. Each photopeak area is associated with a standard

deviation, σ_C, calculated from the same spectrum and by the same computer program. The total expected a priori variability of radioanalytical results then becomes Equation (11)

$$\widehat{\sigma}^2 = \widehat{\sigma}_0^2 + \sigma_C^2 \tag{11}$$

where σ_C^2 expresses the contribution from counting statistics.

Agreement between a priori and a posteriori variability means that the analytical method is in a state of statistical control; this is tested by calculating the statistic T by Equation (12)

$$T = \sum_1^n \frac{(y_i - \widehat{\mu})^2}{\widehat{\sigma}_i^2} \tag{12}$$

where $\widehat{\mu}$ is calculated by Equation (13)

$$\widehat{\mu} = \frac{\sum_1^n y_i \widehat{\sigma}_i^{-2}}{\sum_1^n \widehat{\sigma}_i^{-2}} \tag{13}$$

as the weighted mean of n observations. If there is no significant difference between the expected–a priori–and the actual–a posteriori–standard deviations the statistic T is closely approximated by a χ^2-distribution with $n-1$ degrees of freedom.[46] Significant deviations from this distribution show that there is a disagreement between the sources of variability expected to influence the result according to Equation (11) and the observed variability. The process of identifying the sources and magnitudes of variability that influence the actual analytical results is called the Analysis of Precision.[47]

4.2 Verification Methods

Verification means to verify the absence of errors, and the analysis of precision is an important tool for ascertaining the absence of unknown sources of variability. It is also possible to use this statistic to demonstrate the absence of systematic errors.[48]

The most important condition for relying on the analysis of precision is that the standard deviations σ_i of the analytical results are correctly estimated for each individual measurement. Hence, it must be ascertained that the program used to calculate the results and its associated uncertainty correctly takes into account the contributions from counting statistics and other known uncertainty components. This was tested by a series of experiments involving replicate activation and counting of identical samples under a variety of conditions, covering the entire field of potential application.[28]

Table 6 Verification of calculated counting statistics for photopeak areas covering almost two orders of magnitude and having FWHM ranging from 2.8 to 15 channels

γ-Energy (keV)	Total peak area counts ± SEM[a]	Replicates n	Statistic T
245	43 440 ± 72	29	23.23
344	103 578 ± 95	29	45.95
368	3 166 ± 38	29	26.29
411	7 281 ± 34	29	22.02
444	9 244 ± 39	29	27.40
689	1 673 ± 30	29	23.07
779	21 074 ± 52	27	30.55
867	6 150 ± 44	23	25.25
964	19 102 ± 49	19	29.64
1112	15 362 ± 48	17	25.66
1213	1 505 ± 25	14	10.58
1299	1 629 ± 19	13	8.53
1408	19 319 ± 45	13	16.01
All energies		300	314.18

[a] SEM = Standard error of the mean = Standard deviation/$\sqrt{n}$.

As an example it was considered necessary to ascertain that counting statistics were correctly calculated for spectra recorded with different gains, which means that identical photopeaks had different widths expressed as FWHM. By counting the same ^{152}Eu source up to 30 times at different gain settings and using the most intense γ-line at 122 keV as a comparator the results shown in Table 6 were obtained. For each set of photopeak areas and associated counting statistics, σ_C, the statistic T was calculated in accordance with Equations (12) and (13). All photopeak areas covering an intensity range of two orders of magnitude and values of FWHM from 2.8 to 15 channels were found to be in good statistical control with the possible exception of the 344 keV photopeak. This photopeak offers the highest precision next to the comparator peak and is therefore providing the highest sensitivity for the detection of unexpected sources of variability; however, no correlation with FWHM could be observed and it was therefore concluded that, although there might be a small additional random source of variation, there was no systematic effect associated with the changes in photopeak width.

Once it has been verified that all known sources of variation are correctly accounted for in the calculated standard uncertainty of the analytical result, it becomes possible to detect and identify unknown sources of variability by judicious use of the analysis of precision.[47] Verification of the absence of unknown sources of variability in the analytical results, however, does not exclude the presence of a constant bias or proportional error. Only comparison with results obtained by independent methods or the analysis of materials with known composition can be used for such verification.

Table 7 Results for selenium in mg/kg by neutron activation analysis of a series of BCR certified biological reference materials

Certified reference material			Analytical result	
BCR code	Type of material	Confidence interval	Mean ± SEM	Ratio ± Standard deviation
CRM 184	Bovine muscle	183 ± 12	179 ± 3	0.98 ± 0.03
CRM 185	Bovine liver	446 ± 13	453 ± 27	1.02 ± 0.06
CRM 186	Pig kidney	10 300 ± 500	10 860 ± 170	1.05 ± 0.03
CRM 189	Wholemeal flour	132 ± 10	127 ± 3	0.96 ± 0.04
CRM 278	Mussel tissue	1660 ± 40	1647 ± 25	0.99 ± 0.02
CRM 279	Sea lettuce	593 ± 32	600 ± 22	1.01 ± 0.05
CRM 281	Rye grass	28 ± 4	22.2 ± 2.5	0.79 ± 0.11
			Weighted mean	1.00 ± 0.01
			T	8.17
			Degrees of freedom	6
			$P(\chi^2 \geq T)$	0.23

The preferred way of verifying the absence of such errors is to analyze a series of matrix-matched reference materials with certified concentrations of the determinand for the entire range of concentrations of interest. Results of such a verification analysis for selenium by INAA and RNAA are shown in Table 7 giving results for seven biological reference materials certified by BCR and covering a concentration range of more than two decades.[28] On the assumption that the confidence intervals for the certified values represent the standard uncertainty with a coverage factor $k = 2$, we can combine the uncertainties with the standard uncertainties for the actual results. The ratios between the observed and the expected results are calculated together with their associated standard deviations and presented in the last column of Table 7.

The weighted mean of these ratios is calculated according to Equation (13) and the distribution of the results is tested by the T-test from Equation (12). In this case the test shows no disagreement with the expected ratio of one, and therefore no detectable systematic error. Within the range of concentrations and matrices investigated we can therefore disregard blank or contamination problems, as well as interference or calibration errors.

4.3 Traceability Chain

Traceability is characterized by an *unbroken chain* of comparisons that relates the analytical result to a national or international realization of the appropriate SI units.[49] Each step of the traceability chain must be associated with an expression of uncertainty in accordance with the ISO document,[43] so that the combined uncertainty of the analytical result can take this into account.

Traceability is not so difficult to establish in the physical measurement of a property of a particular sample, including elemental analysis by INAA. Traceability in chemical analysis is much more complicated for two reasons:

- The traceability chain is broken as soon as a chemical separation takes its beginning, i.e. already at the sample dissolution stage.
- There are a huge number of chemical species that may require determination, some being unavailable in pure condition.

In this respect RNAA is superior to other analytical methods in keeping the traceability chain unbroken because:

- carrier addition and equilibration take place right from the dissolution stage, thereby restoring the traceability chain by the determination of the recovery of added carrier;
- the number of elements that can be determined by RNAA is not more than approximately 70; the majority of these can be used directly as comparators in the form of pure elements or compounds with known stoichiometry.[11]

It must, however, be remembered that RNAA determines a particular isotope of the element to be determined and therefore relates directly only to the same isotope in the comparator. Thus, the indicator radionuclide represents the element as such only if the isotopic composition of the comparator exactly equals that in the sample. For the most sensitive elements listed in Table 1 this condition is essentially always fulfilled; but in special cases the isotopic traceability becomes of major concern.

Neutron activation analysis was for example used for the determination of the specific activity of ^{125}I[50] and ^{129}I[51] using ^{126}I, ^{128}I, and ^{130}I as indicators. In this case total iodine content had a completely different isotopic composition from naturally occurring iodine and could

only be determined by using individual comparators for all nuclides, each having its own traceability chain.

The uncertainty with which a result can be reported is thus limited by the uncertainty of the traceability chain, which should therefore be kept at a level where its contribution is small in comparison with other uncertainty components. This is usually achieved by the use of certified reference materials from internationally recognized sources like BCR or NIST (National Institute of Standards and Technology).

In a method like RNAA with inherently small uncertainties and a short traceability chain the contribution to the uncertainty from the traceability chain can be kept at a level of less than 1% when several different pure isotopic comparators are used.[52] The use of the so-called k_0 method[53] based on the use of comparators of elements different from the determinand also represents a break in the traceability chain and is therefore discouraged in the context of using RNAA for certification purposes.

4.4 Uncertainty Budget

When standard deviations, σ_i, have been estimated for all uncertainty components, p, discussed in section 2, the uncertainty budget is composed simply by listing the individual items of the summation in Equation (10). Common to all RNAA budgets are sample weight, quantity of comparator, half-life of indicator, neutron flux density variation, measurements of decay and counting times, amount of carrier, and recovery. Another important common feature is that the radiochemical separation as such does not contribute to the uncertainty in any other way than expressed by the recovery, which means that the entire a priori standard uncertainty is independent of the concentration of the determinand.

All these major uncertainty components are evaluated by statistical methods and therefore belong to Type A.[43] Other uncertainty components may have to be evaluated by inference from other methods or by previous experience; such contributions are referred to as Type B. In RNAA, uncertainty components of Type B might include several potential losses or gains of determinand or comparator that are not otherwise accounted for, such as

- sample contamination before or during irradiation
- loss of determinand before complete carrier equilibration
- leaching from or adsorption to the irradiation container walls
- influence of moisture etc. on sample weight
- effects of radiolysis
- incomplete sample dissolution.

Many of these uncertainty components apply to both sample and comparator, and estimated corrections should in principle be applied to the results of the analysis. The comparator should have a much higher concentration of determinand than the sample and therefore be much less susceptible to contamination than the sample, but losses might occur from

- incomplete transfer from the irradiation container
- diffusion, adsorption, or precipitation during irradiation.

The experienced analyst should be able to reduce the magnitude of all these effects to the extent that even with an estimated Type B combined standard uncertainty of 100% of the correction to be applied, there would be no significant influence on the uncertainty budget.

In a budget for the determination of chlorine by RNAA[48] all Type A contributions are listed with their individual standard uncertainties, as shown in Table 8. Based on the analysis of precision and the application of the T-statistic from Equation (12) it had been concluded that no detectable contribution from Type B uncertainty components could be detected.

Since RNAA is characterized by the absence of reagent blanks the uncertainty budget contains only relative sources of variation, which facilitates the comparison of the items: it is here immediately recognized that the variation in fluence between sample and comparator is the main contributor to the combined uncertainty, whereas weighings have no detectable effect. The influence of half-life depends entirely on the time lapse between the measurement of the separated sample and the comparator. In this particular case the recovery is determined by re-irradiation, which means that the contributors from its measurement are exactly the same as for the sample itself.

Even in a complicated method like RNAA the combined uncertainty is almost always dominated by one or two uncertainty components, one of which is often the contribution from counting statistics. As pointed out above, this contribution is not known until an actual analysis has been carried out; therefore it does not strictly

Table 8 Sources of variability and their estimated standard uncertainties[48]

Parameter	Mean	σ_a	σ_r (%)
Sample weight	2 g	0.2 mg	
Comparator	2 g	0.2 mg	
Concentration	0.2 mg g^{-1}	–	<0.5
Half-life	$T_{1/2} = 37.24$ min	0.05 min	
Neutron flux density	2.5×10^{13} n cm^{-2} s^{-1}	–	2.3
Decay correction	$t_1 - t_2 = 40$ min	0.1 min	
Carrier	0.2 cm^3	0.6 mm^3	
Photopeak area	>22 000 counts	≅150 counts	<0.7

belong to the a priori precision, but must of course be included in the combined uncertainty.

At very low levels of the determinand the a priori precision becomes insignificant, and the uncertainty is completely attributed to the counting statistics; this contribution can be influenced by the choice of detector, counting geometry and counting time, thereby making an a priori estimation less meaningful.

4.5 Statement of Uncertainty

Results presented without uncertainty are useless because no conclusions can be based on them; results with incorrect uncertainty are dangerous because false and misleading conclusions may be drawn.

The tradition of quoting only uncertainty from counting statistics when reporting results obtained by neutron activation analysis is therefore dangerous and should be supplemented with the a priori precision obtained from the uncertainty budget.

Both the analytical result and its uncertainty must be given in SI units with the appropriate subdivisions, and it is preferable to state the *standard uncertainty*, so that statistical tests and inferences can be made directly from the numbers presented. This applies to all results, including those that are not significantly different from zero, or even negative numbers.

As an example we shall take the numbers for platinum in radiochemically separated samples of BCR 186 that were presented in Table 4. These numbers are typical for analytical results, obtained as a difference between two almost equally large numbers: without any real difference between these numbers the probability of getting positive or negative results is the same, and getting three negative and one positive result from four experiments as in this case, has a probability of 25%. The quoted standard uncertainties confirm that the individual numbers are not significantly different from zero.

Thus, there is no reason to regard these numbers as in any way suspicious and therefore they should also be treated in exactly the same way as all other results. Concentrations are calculated exactly as usual, and results are presented in Table 9 together with their uncertainties. Application of the T-test from Equation (12) shows no significant difference between the results, which can therefore be pooled in accordance with Equation (13). The final result is closer to zero than any of the individual results and its standard uncertainty is smaller.

The standard uncertainty provides a confidence interval corresponding to approximately 67% for a normal distribution and this changes very little with deviations from this distribution. However, in many practical situations this confidence interval is not generally appreciated, and it has been decided to multiply the standard uncertainty with a coverage factor, k,[43] which is frequently taken to be a factor of 2. This gives a coverage of 95% for a normal distribution, but this number may change considerably with deviations from this assumption. Under all circumstances the coverage factor must be reported for all analytical results, so that standard uncertainties can be recalculated before use of the data.

Table 9 Concentrations of Pt in BCR 186 pig kidney found by RNAA[38]

Sample size (g)	Recovery (%)	Actual platinum (ng) ± standard deviation	Calculated concentration ($ng\,g^{-1}$) ± standard deviation
2.23	76.4	-0.9 ± 0.8	-0.5 ± 0.5
2.38	92.5	0.6 ± 0.5	0.3 ± 0.2
2.19	72.2	-1.4 ± 0.9	-0.9 ± 0.6
2.25	95.7	-0.8 ± 0.9	-0.4 ± 0.4
D	E	CG ± CH	CK ± CL
		Weighted mean	-0.06 ± 0.18
		$T = 5.70$ at 3 degrees of freedom	

Instead of reporting negative results or results with large uncertainties some laboratories prefer to present their results simply as *n.d. (not detected)*. In other cases, compliance with legal requirements or limits necessitates reporting safe *upper limits*, which are conventionally set as three times the standard uncertainty. Such deflation of the information content is, however, strongly discouraged and should always be accompanied by the reporting of the original data, positive or negative, with their associated standard uncertainty.

Proper reporting of analytical results makes possible the direct application of the analysis of precision and other statistical tools to help in deciding whether a difference between sets of measurement has been observed or not and thereby helps to avoid drawing false conclusions.

5 PERSPECTIVE AND FUTURE DEVELOPMENTS

A top priority in the analytical perspective is the introduction of the BIPM philosophy[42] in analytical chemistry courses at the university level and the implementation of the corresponding ISO Norm[43] at the practical level, so that analytical results can assume their rightful place in the hierarchy of measurement science. This process has been hindered by the traditionally inexact and ambiguous chemical terminology,[54] and the pontification of incommensurate terms and concepts in some ISO standards.[55] Publication of the EURACHEM guide[44] has been a

very significant step forward, but its recommendations are not yet mandatory even for laboratories seeking accreditation. The methodology of RNAA has contributed significantly to revealing such inadequacies[56] and continues to pioneer the introduction of generally recognized measurement concepts and definitions in analytical chemistry.

5.1 Analytical Metrology

The purpose of analytical metrology is the introduction and implementation of metrology in analysis, based on the successful use of the BIPM philosophy in physical measurements, which at any rate are assuming a dominant position in contemporary analysis. In metrology we require correction for all known errors before reporting a measurement, while the uncertainty of the correction must be included in the uncertainty budget. This is often neglected in analytical measurement, where even corrections for blank value or recovery are not always made.[57]

This has been brought out very clearly by comparisons between traditional chemical analyses and purely instrumental methods, where the latter frequently yields higher results for the content of trace elements or other determinands, because of incomplete dissolution of the sample or loss of elements. At the lowest levels this has sometimes been compensated by the inadvertent contamination of the sample during chemical processing or inadequate blank correction.

Correction for incomplete recovery that can always be made in RNAA by measurement of the chemical yield of the carrier is much less straightforward in alternative methods, but nevertheless has to be done during the development stage of the analytical method by using radioactive or stable tracers.[58]

Correction for incomplete sample dissolution can be made by separate analysis of the solid residue by an instrumental method such as INAA.[59]

If such corrections are small their contribution to the uncertainty budget is also small and may be based on sound judgement as a Type B component to the combined uncertainty.

5.2 Certification Analysis

The certification of reference materials for elemental content is required for purposes of international traceability and the mutual recognition of laboratory accreditation needed in environmental and commercial activities. No effort is spared by the leading suppliers of CRMs to make sure that the certified concentrations are reliable, and each of them have developed procedures for performing certification analyses that minimize the risk of erroneous or misleading certification.[60]

The aim is to certify the *total amount* of an element present in the material, regardless of its chemical or physical form, and the certified value should be as close as possible to the true value, which means that corrections must be made for all known or assumed biases. In order to minimize the risk of overlooking unknown or unexpected biases, the use of several independent analytical methods for determining the same element is the basis for all serious certifications.

The absence of a reagent blank, the possibility of determining the recovery of added carrier, and the maintenance of statistical control make neutron activation analysis very attractive for the certification of low concentrations of the elements listed in Table 1. Together with its direct traceability to the pure element this makes neutron activation analysis a potentially definitive method.[7]

The combination of RNAA with INAA using different indicators and different comparators even opens up the possibility of self-validation[32] by fulfilling most of the requirements for completely independent methods of analysis. This introduces the possibility of certifying reference materials for elements for which no alternative reliable methods are available.

These properties of RNAA further emphasize the applicability of this method for certification purposes, and Table 10 presents examples of elements determined in a variety of reference materials, certified by BCR and covering the entire range of trace element concentrations from $0.01\,mg\,kg^{-1}$ to 0.01%. Particularly since the introduction of natural materials for certification, NIST and all other major suppliers have counted on neutron activation analysis not only for elements with

Table 10 Use of RNAA by BCR for the certification of reference materials in the European Union

Element	Materials	Examples of certification	
		BCR code	$mg\,kg^{-1}$
Antimony	Animal feed	CRM 281	0.047
	Metallic copper	CRM 074	0.58
Arsenic	Animal feed	CRM 274	0.132
	Animal tissue	CRM 278	5.9
	Metallic copper	CRM 074	0.78
	Coal	CRM 181	27.7
	Plants	CRM 279	3.09
	Sediment	CRM 320	76.7
Cadmium	Animal tissue	CRM 422	0.017
	Coal	CRM 040	0.11
	Milk powder	CRM 151	0.101
	Hair	CRM 397	0.521
	Sewage sludge	CRM 146	77.7
	Incineration ash	CRM 176	470

(*continued overleaf*)

Table 10 (*continued*)

Element	Materials	Examples of certification	
		BCR code	mg kg^{-1}
Calcium	Milk powder	CRM 063	12 600
Chlorine	Fresh water	CRM 399	50.5
Chromium	Animal feed	CRM 281	2.14
	Coal	CRM 040	31.3
	Leaves	CRM 100	8.0
	Sewage sludge	CRM 597	203
	Fly ash	CRM 038	192
Cobalt	Animal feed	CRM 402	0.178
	Coal	CRM 040	7.8
Copper	Animal feed	CRM 414	29.5
	Animal tissue	CRM 186	31.9
	Cereals	CRM 191	2.6
	Milk powder	CRM 150	2.23
	Plants	CRM 279	13.14
	Sewage sludge	CRM 144	713
Iodine	Animal feed	CRM 129	0.167
	Animal tissue	CRM 422	4.95
	Milk powder	CRM R63	0.81
Iron	Fresh water	CRM 398	0.029
	Milk powder	CRM 150	11.8
Manganese	Animal tissue	CRM 422	0.543
Mercury	Animal feed	CRM 482	0.48
	Animal tissue	CRM 185	0.044
	Coal	CRM 180	0.123
	Milk powder	CRM 151	0.101
	Sediment	CRM 277	1.77
	Soil	CRM 141R	0.25
	Sewage sludge	CRM 146R	8.62
	Fly ash	CRM 038	2.10
Molybdenum	Animal feed	CRM 281	0.84
Phosphorus	Metallic copper	CRM 017	6.9
	Milk powder	CRM R63	11 100
	Leaves	CRM 101	1690
Selenium	Animal feed	CRM 274	1.03
	Animal tissue	CRM 184	0.183
	Metallic copper	CRM 074	0.37
	Coal	CRM 182	0.68
	Cereals	CRM 189	0.132
	Plants	CRM 279	0.593
	Sediment	CRM 320	0.214
	Incineration ash	CRM 176	41.2
Thallium	Incineration ash	CRM 176	2.85
Zinc	Animal tissue	CRM 278	76
	Coal	CRM 040	30.2
	Plants	CRM 279	51.3
	Soil	CRM 143R	1063
	Sewage sludge	CRM 145R	2137

high sensitivity, but also for elements where alternative methods are not entirely satisfactory;[61] this situation is not likely to change in the foreseeable future.

5.3 Sampling Constants

The characterization of a natural CRM must include information about possible heterogeneity, either in the form of a minimum sample size or – preferably – as a statement of the maximum sampling constant.[62] All reference materials are carefully homogenized during their preparation, which means that they may be considered uniform and therefore can be characterized by the Ingamells' sampling constant[63] (Equation 14)

$$K_S = R^2 W \tag{14}$$

where R is the coefficient of variation observed by the analysis of replicate samples with weight W. With a known sampling constant the contribution of heterogeneity to the variability of results is accounted for by adding a term to the a priori precision from Equation (11) (Equation 15)

$$\hat{\sigma}_w^2 = \hat{\sigma}^2 + \hat{\mu}^2 \frac{K_S}{10^4 W} \tag{15}$$

which also shows that the sampling constant can be determined by analyzing a number of samples by an analytical method in statistical control.[47]

Present reference materials have sampling constants of the order of 1 g,[40] but modern methods of analysis use much smaller samples and therefore need materials with smaller sampling constants. Sampling constants need to be determined separately for each determinand,[64] and the use of methods that are in statistical control regardless of sample size is therefore in great demand. For many elements both RNAA and INAA can be used on samples in the milligram to microgram range without problems and therefore can be used to advantage in future verification of very small sampling constants just by activating the samples in very high neutron flux densities.

ABBREVIATIONS AND ACRONYMS

BCR	Bureau Communautaire de Réference
BIPM	Bureau International des Poids et Mesures
CRM	Certified Reference Material
DSP	Digital Signal Processing
FWHM	Full Width at Half-Maximum
GM	Geiger-Müller
IDMS	Isotope Dilution Mass Spectrometry
INAA	Instrumental Neutron Activation Analysis
ISO	International Organization for Standardization
NIST	National Institute of Standards and Technology
QC/QA	Quality Control/Quality Assurance
RNAA	Radiochemical Neutron Activation Analysis
SEM	Standard Error of the Mean

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Instrumental Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Radiotracer Methods

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Nuclear Detection Methods and Instrumentation

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. G. Hevesy, H. Levi, 'The Action of Neutrons on the Rare Earth Elements', *K. Dan. Vidensk. Selsk., Mat.-Fys. Medd.*, **XIV**(5), 3–34 (1936).
2. G.T. Seaborg, J.J. Livingood, 'Artificial Radioactivity as a Test for Minute Traces of Elements', *J. Am. Chem. Soc.*, **60**, 1784–1786 (1938).
3. G.E. Boyd, 'Method of Activation Analysis', *Anal. Chem.*, **21**, 335–347 (1949).
4. 'Definitions and Terms', *J. Radioanal. Nucl. Chem.*, **113**, 539–541 (1987).
5. W.D. Ehmann, D.E. Vance, *Radiochemistry and Nuclear Methods of Analysis*, J. Wiley & Sons, New York, 253–308, 1991.
6. K. Heydorn, 'INAA – Application and Limitation', *Fresenius' J. Anal. Chem.*, **337**, 498–502 (1990).
7. R. Dybczynski, B. Danko, 'A Definitive Method for the Determination of Trace Amounts of Cobalt in Biological Materials by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **182**, 43–59 (1994).
8. K. Heydorn, 'Neutron Activation Analysis' in *Encyclopedia of Medical Devices and Instrumentation*, ed. J.G. Webster, J. Wiley & Sons, New York, Vol. 3, 2041–2046, 1988.
9. K. Heydorn, *Neutron Activation Analysis for Clinical Trace Element Research*, CRC-Press, Boca Raton, FL, (a) 75–83, Vol. 1, (b) 101–103, Vol. 1, (c) 120–122, Vol. 1, (d) 129–131, Vol. 1, (e) 145–149, Vol. 1, (f) 42–45, Vol. 2, 1984.
10. K. Heydorn, E. Damsgaard, B. Rietz, 'Systematic Differences in the Determination of Vanadium in Standard Reference Material 1571 Orchard Leaves', *Anal. Chem.*, **52**, 1045–1049 (1980).
11. J.R. Moody, R.R. Greenberg, K.W. Pratt, T.C. Rains, 'Recommended Inorganic Chemicals for Calibration', *Anal. Chem.*, **60**, 1203A–1218A (1988).
12. P. Bode, R.M.W. Overwater, J.J.M. De Goeij, 'Large-sample Neutron Activation Analysis: Present Status and Prospects', *J. Radioanal. Nucl. Chem.*, **216**, 5–11 (1997).
13. K. Heydorn, *Handbook on Metals in Clinical and Analytical Chemistry*, Marcel Dekker, Basle, 181–194, 1994.
14. *Nuclear Research Reactors in the World*, International Atomic Energy Agency, Vienna, 1998.
15. D.E. Ryan, A. Chatt, J. Holzbecher, 'Analysis for Trace Elements with a SLOWPOKE Reactor', *Anal. Chim. Acta*, **200**, 89–100 (1987).
16. K. Heydorn, E. Damsgaard, 'Gains or Losses of Ultratrace Elements in Polyethylene Containers', *Talanta*, **29**, 1019–1024 (1982).
17. Lin Xilei, K. Heydorn, B. Rietz, 'Limit of Detection for the Determination of Pt in Biological Material by RNAA using Electrolytic Separation of Gold', *J. Radioanal. Nucl. Chem.*, **160**, 85–99 (1992).
18. R.R. Greenberg, H.M. Kingston, R.L. Watters, K.W. Pratt, 'Dissolution Problems with Botanical Reference Materials', *Fresenius' J. Anal. Chem.*, **338**, 394–398 (1990).
19. A.R. Byrne, J. Versieck, 'Vanadium Determination at the Ultratrace Level in Biological Reference Materials and Serum by Radiochemical Neutron Activation Analysis', *Biol. Trace Elem. Res.*, **27**, 529–540 (1990).
20. A.J. Blotcky, J.P. Claasen, F.R. Roman, E.P. Rack, S. Badakhsh, 'Determination of Aluminum by Chemical and Instrumental NAA in Biological Standard Reference Materials and Human Brain Tissue', *Anal. Chem.*, **64**, 2910–2913 (1992).
21. A. Mizuike, 'Separations and Preconcentrations', in *Trace Analysis, Physical Methods*, ed. G.H. Morrison, Interscience, New York, 103–159, 1965.
22. D. van Renterghem, R. Cornelis, R. Vanholder, 'Radiochemical Determination of Twelve Trace Elements in Human Blood Serum', *Anal. Chim. Acta*, **257**, 1–5 (1992).
23. R.R. Greenberg, R. Zeisler, 'A Radiochemical Procedure for Ultratrace Determination of Chromium in Biological Materials', *J. Radioanal. Nucl. Chem.*, **124**, 5–20 (1988).
24. J. Kucera, M. Vobecky, L. Soukal, D. Zákoucky, D. Vénos, 'Low Level Determination of Thallium in Biological and Environmental Reference Materials by RNAA Using Several Counting Methods', *J. Radioanal. Nucl. Chem.*, **217**, 131–137 (1997).
25. F. Girardi, E. Sabbioni, 'Selective Removal of Radiosodium from Neutron-activated Materials by Retention on Hydrated Antimony Pentoxide', *J. Radioanal. Chem.*, **1**, 169–178 (1968).
26. Z.B. Alfassi, C.M. Wai (eds.), *Preconcentration Techniques for Trace Elements*, CRC Press, Boca Raton, FL, 1992.
27. E. Damsgaard, K. Heydorn, 'Synthesis of Precision for the Certification of Phosphorus in Biological Materials by

Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **113**, 267–275 (1987).
28. K. Heydorn, 'Validation of Neutron Activation Analysis Techniques', in *Quality Assurance for Environmental Analysis*, ed. Ph. Quevauviller et al., Elsevier, Amsterdam, 89–110, 1995.
29. K. Heydorn, W. Lada, 'Peak Boundary Selection in Photopeak Integration by the Method of Covell', *Anal. Chem.*, **44**, 2313–2317 (1972).
30. K. Heydorn, E. Damsgaard, 'Simultaneous Determination of Arsenic, Manganese, and Selenium in Biological Materials by Neutron Activation Analysis', *Talanta*, **20**, 1–11 (1973).
31. B. Danko, R. Dybczynski, 'Determination of Molybdenum and Uranium in Biological Materials by Radiochemical Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **216**, 51–57 (1997).
32. A.R. Byrne, 'Some Improvements in the Quality of NAA Procedures and the Reliability of the Results', *Biol. Trace Elem. Res.*, **44**, 529–537 (1994).
33. R.R. Greenberg, R. Zeisler, H.M. Kingston, T.M. Sullivan, 'Neutron Activation Analysis of the NIST Bovine Serum Standard Reference Material using Chemical Separations', *Fresenius' J. Anal. Chem.*, **332**, 652–656 (1988).
34. K. Heydorn, 'Use of the National Physical Laboratory Ionization Chamber Type 1383A in Neutron Activation Analysis', *J. Radioanal. Chem.*, **10**, 245–256 (1972).
35. K. Masumoto, 'Isotope Dilution Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, John Wiley & Sons, New York, 511–523, 1994.
36. A.R. Byrne, M. Dermelj, 'An Endogeneous Standard, Radioisotopic Ratio Method in NAA', *J. Radioanal. Nucl. Chem.*, **223**, 55–60 (1997).
37. A. Lagerwaard, J.R.W. Woittiez, J.J.M. de Goeij, 'An Independent Accurate Reference Method for the Determination of Chromium in Biological Materials', *Fresenius' J. Anal. Chem.*, **351**, 786–789 (1995).
38. B. Rietz, K. Heydorn, 'Determination of Gold and Platinum in Biological Materials by Radiochemical Neutron Activation Analysis using Electrolytic Separation of Gold', *J. Radioanal. Nucl. Chem.*, **174**, 49–56 (1993).
39. *Safe Handling of Radionuclides*, Safety Series No. 1, International Atomic Energy Agency, Vienna, 1973.
40. K. Heydorn, E. Damsgaard, 'The Determination of Sampling Constants by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **110**, 539–553 (1987).
41. S. Landsberger, S. Peshev, 'Compton Suppression Neutron Activation Analysis: Past, Present and Future', *J. Radioanal. Nucl. Chem.*, **202**, 201–224 (1996).
42. P. Giacomo, 'News from Bureau International de Poids et Mesures', *Metrologia*, **17**, 73 (1981).
43. *Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1993.
44. Eurachem Working Group, *Quantifying Uncertainty in Analytical Measurement*, BSI, London, 1995.
45. K. Heydorn, 'Detection of Systematic Errors by the Analysis of Precision', *Accuracy in Trace Analysis*, National Bureau of Standards Publication 422, Washington, DC, 127–139, 1976.
46. K. Heydorn, K. Nørgaard, 'Analysis of Precision of Activation Analysis Methods', *Talanta*, **20**, 835–842 (1973).
47. K. Heydorn, 'Quality Assurance and Statistical Control', *Microchim. Acta*, **III**, 1–10, (1991).
48. K. Heydorn, 'Detecting Errors in Micro and Trace Analysis by Using Statistics', *Anal. Chim. Acta*, **283**, 494–499 (1993).
49. *Certification of Reference Materials – General and Statistical Principles*, ISO Guide 35, ISO, Geneva, 1989.
50. K. Heydorn, 'Determination of the Specific Activity of Carrier-free ^{125}I Preparations by Neutron Activation Analysis', *J. Radioanal. Chem.*, **3**, 225–232 (1969).
51. Susan Parry et al., 'The Determination of ^{129}I in Milk and Vegetation Using Neutron Activation Analysis', *Sci. Total Environ.*, **174**, 351–360 (1995).
52. R.R. Greenberg, 'Accuracy in Standards Preparation for Neutron Activation Analysis', *J. Radioanal. Nucl. Chem.*, **179**, 131–139 (1994).
53. F. de Corte, 'The Traceability of the Neutron Activation Analysis k_0-Standardization Method', *J. Trace Microprobe Tech.*, **5**, 115–134 (1987).
54. K. Heydorn, 'Letter to the Editor', *Trends Anal. Chem.*, **11**, II (1992).
55. 'Accuracy (trueness and precision) of Measurement Methods and Result – Part 1: General Principles and Definitions', ISO Standard 5725-1, International Organization for Standardization, Geneva, 1994.
56. K. Heydorn, 'Comments on Definitions of Analytical Terms', *Accred. Qual. Assur.*, **2**, 204–205 (1997).
57. P. Willetts, R. Wood, 'Measurement Uncertainty – A Reliable Concept in Food Analysis and for the Use of Recovery Data?', *Accred. Qual. Assur.*, **3**, 231–236 (1998).
58. IUPAC, 'Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement', 1995 IUPAC Technical Report, 1998.
59. K. Heydorn, E. Damsgaard, 'Indirect Determination of Soluble Trace Elements in Certified Reference Materials by Neutron Activation Analysis', *Mikrochim. Acta*, **119**, 297–304 (1995).
60. K. Heydorn, 'Procedures for Certification of BCR Environmental Reference Materials', in *Harmonization of Health Related Environmental Measurements using Nuclear and Isotopic Techniques*, IAEA, Vienna, 157–173, 1997.
61. R.R. Greenberg, 'The Role of Neutron Activation Analysis in the Development and Certification of NIST

Environmental Standard Reference Materials', in *Harmonization of Health Related Environmental Measurements using Nuclear and Isotopic Techniques*, IAEA, Vienna, 137–144, 1997.

62. 'Contents of Certificates of Reference Materials', *ISO Guide 31*, International Organization for Standardization, Geneva, 1981.
63. C.O. Ingamells, P. Switzer, 'A Proposed Sampling Constant for Use in Geochemical Analysis', *Talanta*, **20**, 547–568 (1973).
64. A. Chatt, R.R. Rao, C.K. Jayawickreme, L.S. McDowell, 'Determination of Sampling Constants for Selenium in Biological Reference Materials by Neutron Activation', *Fresenius' J. Anal. Chem.*, **338**, 399–407 (1990).

Radiochemical Separation Schemes for Multielement Determination

Zeev B. Alfassi
Ben Gurion University of the Negev, Beer Sheva, Israel

1 Introduction 1
2 Separation in Material Sciences 1
3 Geological and Environmental Samples 5
4 Biological Samples 12
Abbreviations and Acronyms 14
Related Articles 14
References 14

Radiochemical activation analysis is a more laborious analytical method than instrumental analysis, but can be used to determine elements at lower concentrations. In most cases, the sample to be analyzed is separated into groups of elements rather than individual elements, so that the elements do not interfere with one another during the measurement of radionuclides content. Schemes for separation into multielement groups are given for metals, biological, geological and environmental samples.

1 INTRODUCTION

While instrumental neutron activation analysis (INAA) is more popular than RNAA (radiochemical neutron activation analysis), owing to its simplicity, RNAA can determine much lower concentrations than INAA, because the interferences from other radionuclides are reduced considerably. Whereas INAA involves only activation and activity measurement, RNAA also includes dissolution of the sample and chemical separation of the sample solution. In some cases, the interferences can be removed by decay (''cooling'') and INAA can be used. However, in many cases the IRNs (indicator radionuclides) are shorter lived than those of the interferences, and in other cases the required decay time is too long to be practical. Owing to the chemical dissolution, separation and recovery yield determination, RNAA involves more person hours and should be avoided when it is possible. Heydorn, in his article (**Radiochemical Neutron Activation Analysis**), shows why RNAA is so important in verification of the concentration of trace elements in reference materials. However, RNAA is not just used for validation of reference materials, although this is clearly one of its main uses. RNAA involves irradiation and counting, dealt with in the article **Instrumental Neutron Activation Analysis**, as well as dissolution of the samples and separation of the dissolved samples into various groups. The dissolution step is much simpler than for other methods of analysis, because the final measurement is the determination of the radionuclides formed by the irradiation. As the dissolution step is done after the irradiation is carried out, no contamination during the dissolution step can influence the final results. In addition, as the dissolution of the sample is done together with that of carriers, the danger of losing trace elements by adsorption is considerably lower. On the other hand, owing to the radioactivity of the irradiated sample, the dissolution, if performed many times, should be done behind lead shields. In principle every chemical separation that is used in any other analytical method can be used in RNAA. But since the separation very rarely involves a single element, an appropriate group of elements which do not interfere with each other is different for various determination methods. This is the reason that it was decided that it was worth collecting together in this article the main methods of chemical separation into groups that are used in RNAA. In some cases, even the separation into groups is not needed, and it is sufficient to remove the major interferences. Egger and Krivan[1] in their determination of trace elements in aluminum, removed only the interference of ^{24}Na (by sorption on a hydrated antimony pentoxide, HAP, column) and were then able to measure 38 trace elements, all in concentrations of less than 1 ppb. For most other analytical methods the bulk of Al might need to be removed.

An additional aspect which is different for RNAA with respect to all other determination methods, is that in RNAA the final determination can be performed while the trace elements are adsorbed onto a separation column, whereas all other methods require the elution of the trace elements from the column.

2 SEPARATION IN MATERIAL SCIENCES

Park et al.[2] studied the concentration of 13 trace elements in high-purity molybdenum by separating them into three groups, separated from the major interferences which are Mo and W. A cation exchange column (Dowex™ 50) in dilute HCl medium is used to remove Mo and W, which are only weakly adsorbed on this column, while many of the impurity elements are strongly absorbed. Na and K are eluted by using additional dilute

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

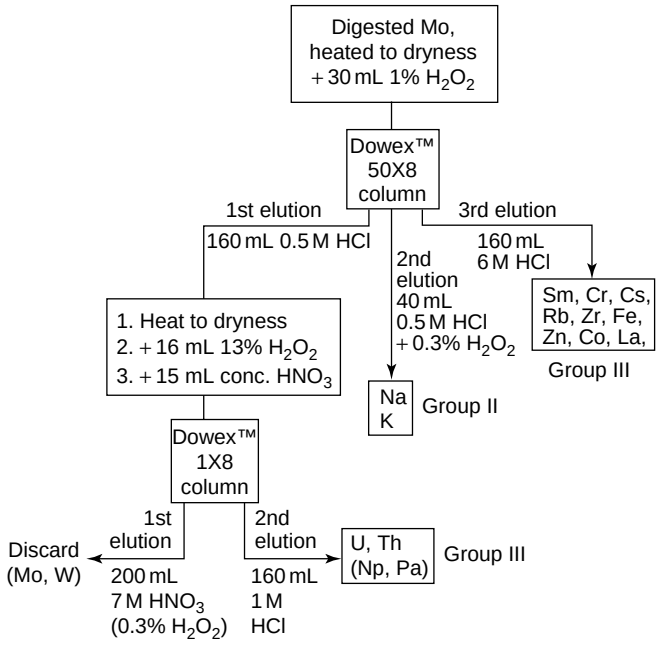

Figure 1 Park et al.[2] method for separation of 13 elements into three groups for determination of trace elements in molybdenum. (Reproduced from Park et al.[2] with permission.)

HCl. Nine trace elements were eluted with 6-M HCl. Together with Mo and W, ^{239}Np and ^{233}Pa, which are the IRNs for U and Th, due to β^- decay following the (n,γ) reaction, were also eluted, respectively, Equation (1)

$$^{238}\mathrm{U(n,\gamma)}^{239}\mathrm{U} \xrightarrow{\beta^-} {}^{239}\mathrm{Np};$$
$$^{232}\mathrm{Th(n,\gamma)}^{233}\mathrm{Th} \xrightarrow{\beta^-} {}^{233}\mathrm{Pa} \qquad (1)$$

Np and Pa were separated from Mo and W using anion exchange (Dowex™-1) in strong nitric acid media. Pa and Np are strongly absorbed on the anion exchange resin, while molybdenum and tungsten are not. Figure 1 gives a schematic representation of the process of separation of Park et al.[2] Before counting, each eluate was concentrated by partial evaporation.

Theimer and Krivan[3] developed a group separation for the determination of 20 elements in high-purity molybdenum (Figure 2). They developed two modifications. The first one involves removal of Mo and its daughter Tc and the main contamination of W, by absorption on anion exchange resin (Dowex™ 1X8). Five other elements are also absorbed on the column, but 20 elements are eluted and can be determined simultaneously in the eluate. The second modification allows determination of only 11 elements, although separation is effected into two groups. The advantage of this special modification is the separation of ^{233}Pa (the IRN of Th) with only Sc, enabling determination of small amounts of Th. Both modifications

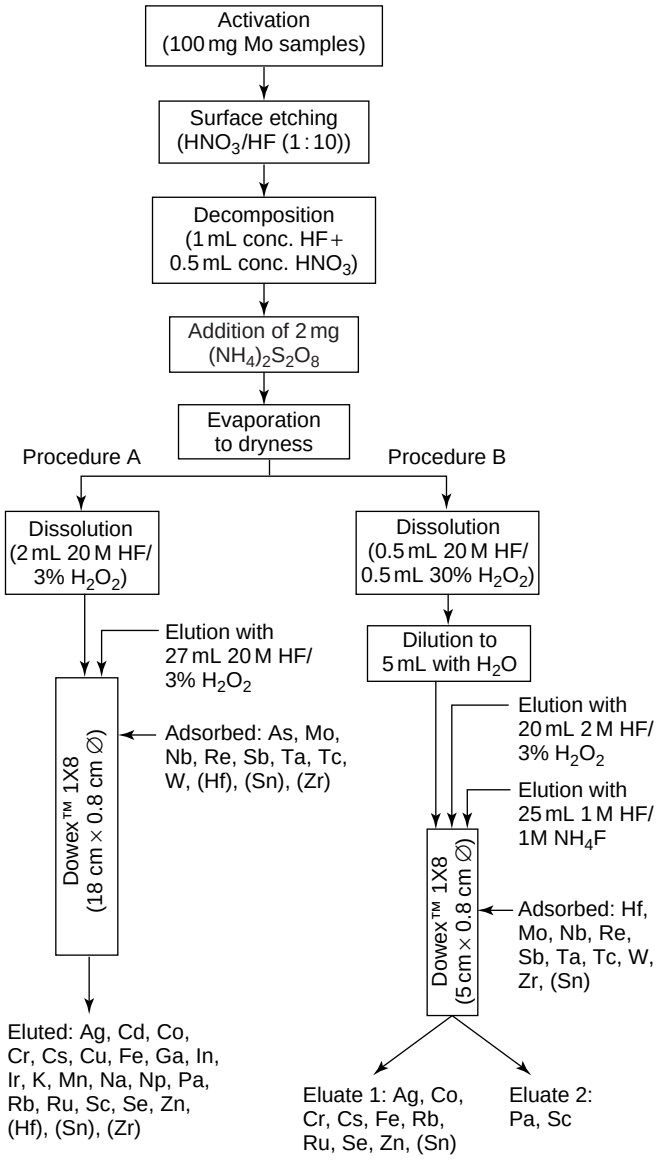

Figure 2 Flow chart of the two procedures for radiochemical separation of trace elements in molybdenum. (Reproduced from Theimer and Krivan,[3] with permission.)

are based on elution with various concentrations of HF, similar to all the procedures developed by Krivan et al. for determination of trace elements in metals. The flow chart for the two radiochemical separation procedures is given in Figure 2. The two methods are based on Mo, Tc, and W being absorbed on an anion exchange column in HF media, while the trace elements are eluted with HF. For selective separation of Pa (together with Sc only), lower concentrations of HF are used. The elements given in parentheses are eluted but not completely, so that they are found both on the column and in the eluate. The elements determined via medium-lived IRNs and which had

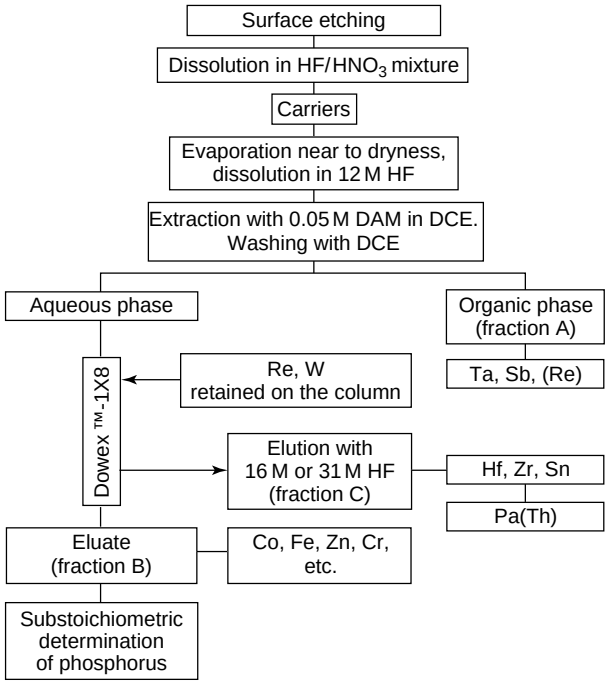

Figure 3 Flow chart for the radiochemical separation of trace elements in tungsten. (Reproduced from Caletka et al.[4] with permission.)

to be in eluate 1 according to their distribution ratios, as for example Cu, Ga, K, Mn and Na, do not appear in procedure B, because about 1% of the technetium is eluted with this fraction, masking the activity of these IRNs. However, if $(NH_4)_2S_2O_8$ is added to the sample solution prior to its evaporation, in order to oxidize all technetium species to the heptavalent state, less than 0.01% of the technetium is found in fraction 1, and the other elements can also be determined in this fraction.

A similar technique was developed by Krivan's group to determine the content of 19 elements in high-purity tungsten.[4] However, a modification was done in order to measure the concentrations of Ta and Sb, which in the previous processes were adsorbed on the anion exchange resins together with Mo, Tc and W, and could not be determined. Ta and Sb are extracted with organic solvent (dichloroethane, DCE) after forming complexes with diantipyrilmethane (DAM), prior to the introduction of the solution of the dissolved irradiated tungsten onto the anion exchange column. In order to measure the P content via ^{32}P, a β^--only emitter, ^{32}P is substoichiometrically extracted as molybdophosphate with tetraphenylarsonium chloride in dichloromethane.[5] The flow chart for this separation is given in Figure 3.

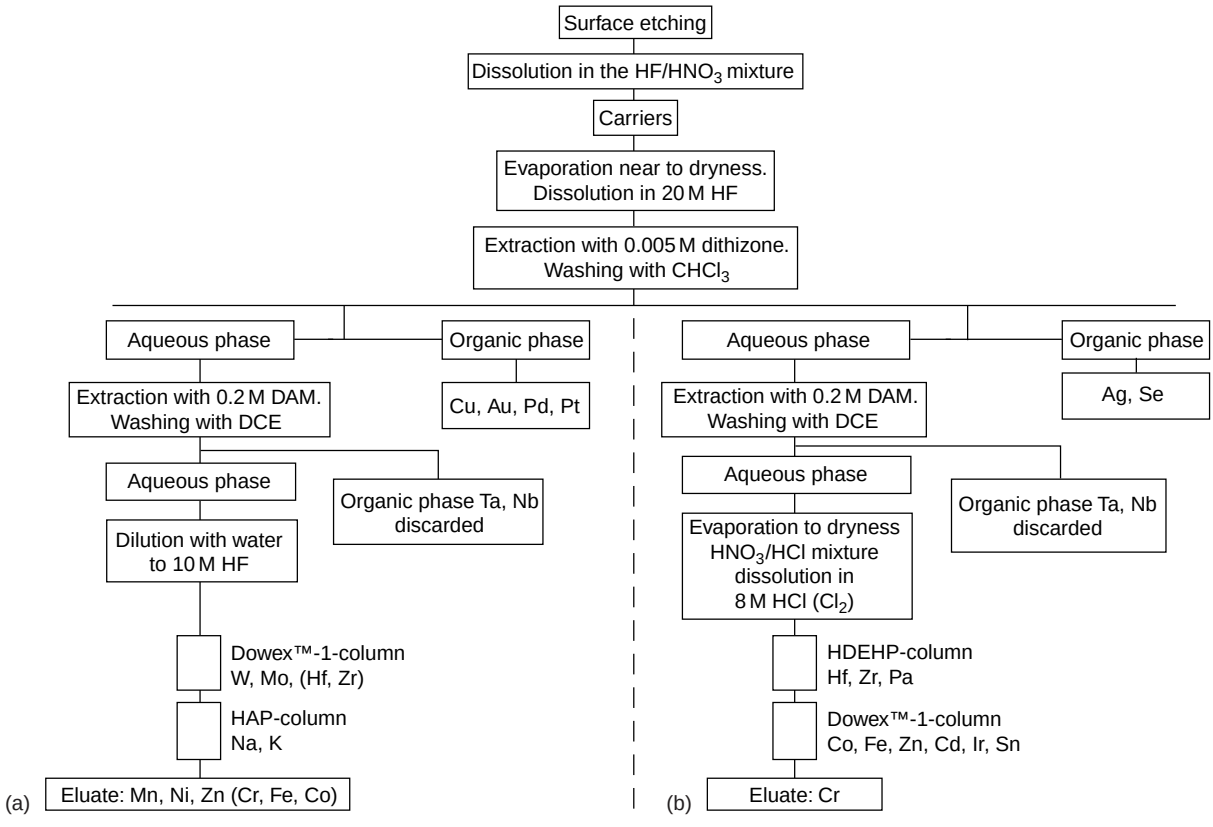

Figure 4 Two procedures for RNAA of trace elements in niobium (a) for medium-lived IRNs; (b) for long-lived IRNs. HDEHP, di(2-ethylhexyl)-orthophosphoric acid. (Based on figures from Caletka et al.[6] with permission.)

Caletka et al.[6] developed a RNAA procedure for determination of 26 elements in niobium by irradiation and processing of two samples, one for measurement of medium-lived IRNs (the radionuclide used for the determination of an element in NAA), by irradiation for 12 h. The second sample was irradiated for 5 days in order to determine the long-lived IRNs. In the determination of medium-lived IRNs, they wanted to measure ^{56}Mn ($t_{1/2} = 2.58$ h) and ^{65}Ni ($t_{1/2} = 2.52$ h) as well. This is possible because the main activity produced from the matrix (^{94m}Nb) is short lived ($t_{1/2} = 6.24$ min) and one hour of cooling leaves a sample that is not too radioactive to handle. For matrixes where the main activity is longer lived, IRNs with half-lives of a couple of hours can be determined only by pre-irradiation separation of the matrix, or treatment in special hot laboratories, where very high radioactivity can be handled. Different schemes of separation were devised for the medium- and long-lived IRNs, however, they have many common features. Both use solvent extractions followed by separation on columns. The same two solvent extractions were used in the two

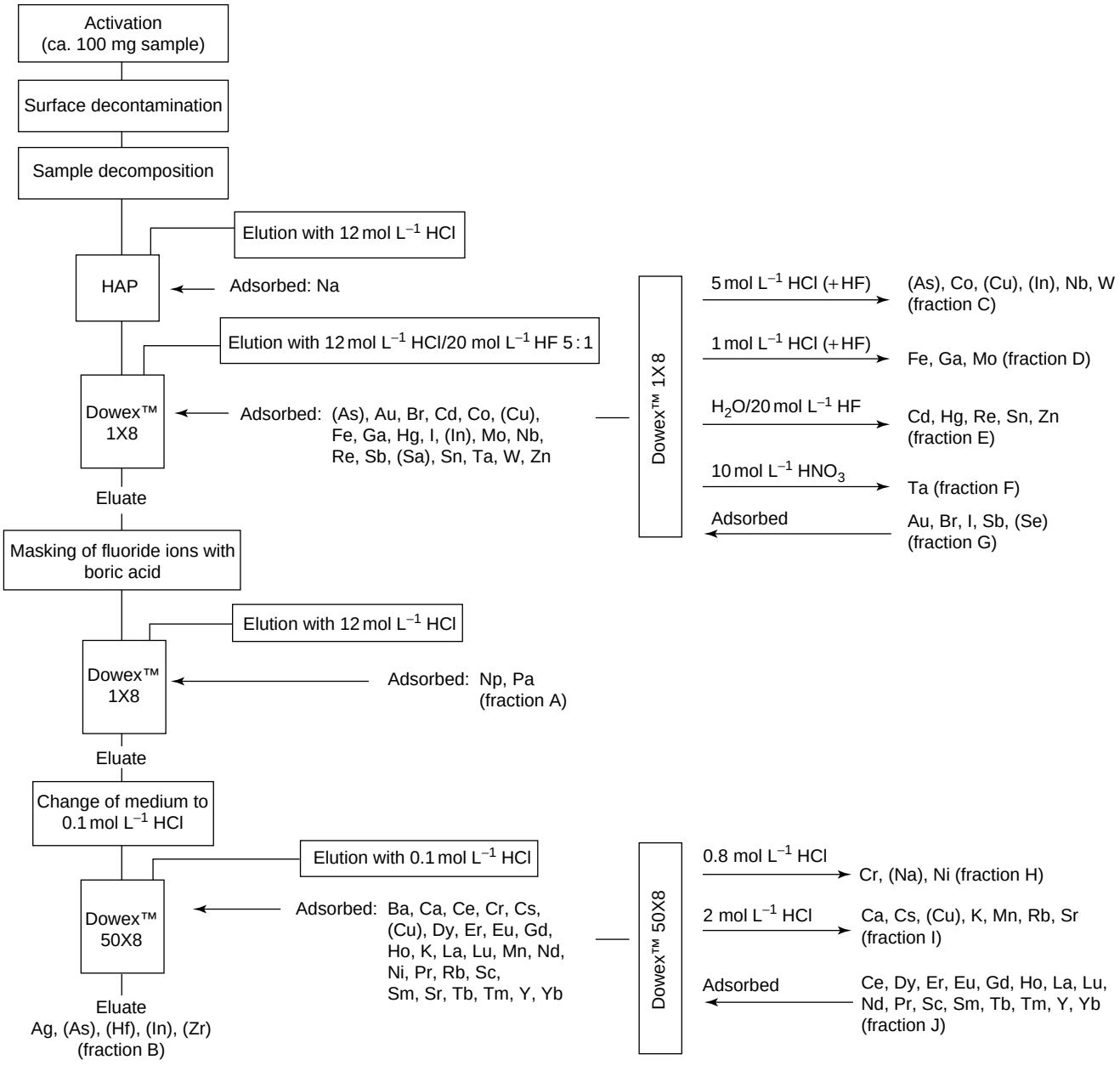

Figure 5 Procedure for separation of 52 elements into 10 fractions in irradiated ultrapure aluminum. (Reproduced from Egger and Krivan[7] with permission.)

procedures, and the only difference is in the columns and acids used for separation of the last aqueous phase into three groups. The first extraction step is for cations which form chelates with dithizone and are extracted with chloroform. This fraction includes Cu, Au, Pd and Pt in the medium-lived IRNs, and Ag, Se in the long-lived IRN samples. The second extraction of metals is for those which form complexes with DAM. The organic solvent is DCE. This extraction is used to remove tantalum and niobium, and the organic phase containing them is discarded. The elements remaining in the aqueous phase are separated into three groups using two columns in series. One group is adsorbed on each column, and the third group contains the elements which were eluted from the two columns. For the long-lived samples, the first column is the reversed phase di-(2-ethylhexyl) orthophosphoric acid on a solid support and then a column of anion exchange resin-Dowex™-1. For the medium-lived IRNs, the anion exchange-Dowex™-1 is the first column, followed by a HAP column. In high-concentration HCl solution, HAP retains only the Na ions; however, under the conditions used here, K also remained on the HAP column. The flow chart of the separation is given in Figure 4.

A thorough determination of the trace elements in ultrapure metal was performed for aluminum.[7] Previously, we mentioned that 38 elements were determined in levels of parts per billion, by selective removal of ^{24}Na with a HAP column. Later, the same group extended this study, and in order to improve the limits of detection, an additional separation into 10 groups was done.[7] Forty-three elements have limits of detection below 10 ppb, and for U and Th the limits of detection are 50 ppt (5×10^{-11}). The separation scheme is given in Figure 5. First, Na^+ is retained on HAP. With 12-M HCl elution, this is the only cation to be retained. The eluate from the HAP column is loaded onto an anion exchange and washed with 12-M HCl–20-M HF, 5 : 1. Twenty elements from the eluate are adsorbed on the anion exchanger resin (Dowex™ 1X8), and later divided into five groups, by elution of four different groups with different eluants. The fifth group is the one that is retained on the column after the fourth elution. The original eluate from the Dowex™ 1X8 column is masked by addition of boric acid and the solution is introduced into another Dowex™ 1X8 anion-exchanger column. The masking of the F^- anions by boric acid leads to retention of Pa and Np (IRNs of Th and U) on the column. Further elements are separated on a cation exchanger. The solution is evaporated to dryness, dissolved in 0.1-M HCl and introduced into a cation exchanger Dowex™ 50X8 and washed with 0.1-M HCl. The eluate is one group of five elements, while the 27 elements remaining on the column are separated into three groups by two elutions. The flow chart for the separation (Figure 5) shows the separation of 52 elements as studied by radiotracers. Not all of them were found in the ultrapure Al. Only those not in brackets were actually found in the irradiated Al.

3 GEOLOGICAL AND ENVIRONMENTAL SAMPLES

Many studies have been done on the determination of precious metals in geological samples. Nadkarni and Morrison[8,9] digested the irradiated sample

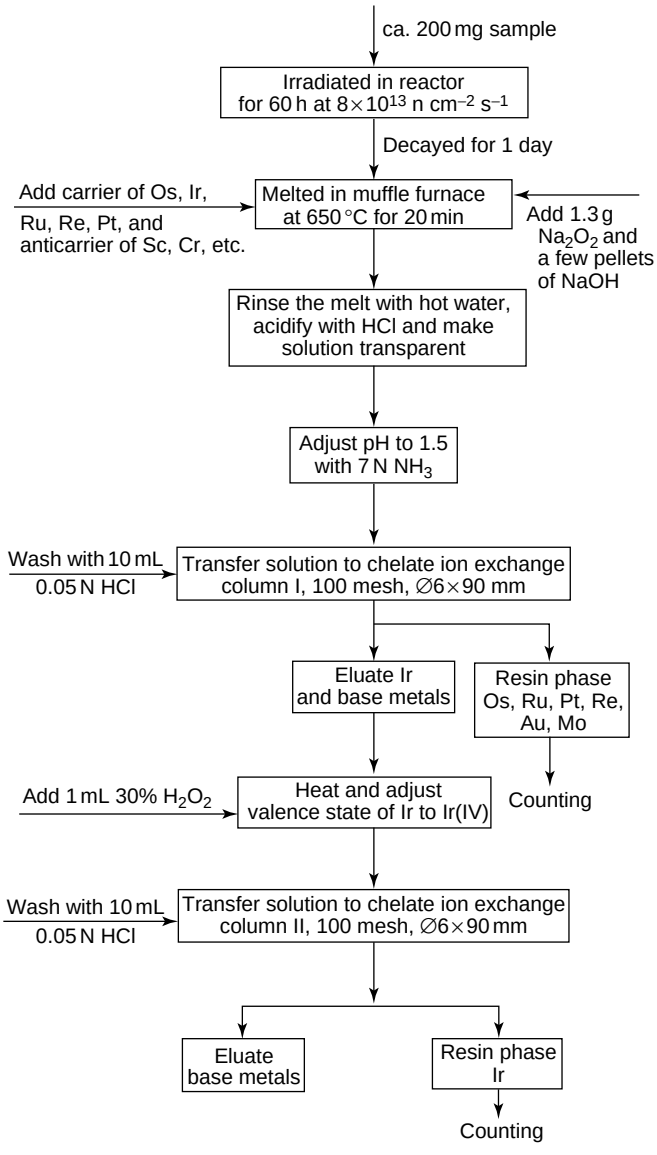

Figure 6 Radiochemical separation procedure of noble metals based on chelate ion-exchange resin. (Reproduced from Chai et al.[14] with permission.)

(200–500 mg) by peroxide fusion, and separated the noble metals on a Srafion NMRR selective chelating ion-exchange resin. However, yields are not well reproducible and reirradiation of each sample is necessary to determine them.[10] Only Pd, Pt, Ir and Au can be determined by this process. The use of the ion exchanger Srafion NMRR was questioned by Stockman,[11] who could not repeat the results of Nadkarni and Morrison and attributed this to changes in the manufacture of the resin. There are also some previous data which give contrasting results for the specificity of this resin for d^8 ions.[12,13]

Chai et al.[14] developed two different methods for radiochemical separation of the noble metals. They found that Nadkarni and Morrison were right and the Srafion NMRR resin can be used to retain Os and Ir. Ten milliliters of 0.05-M HCl eluted 100% of Sc, Cs, and Fe, whereas only 10% of Os was eluted. Nadkarni and Morrison reported that thiourea can be used to elute the noble metals, whereas Stockman argued that thiourea elutes only gold, while Ir remains on the resin. Chai et al.[14] found that Ir and Re can be eluted with NH_3, whereas only one-quarter of Os was eluted and no Au or Pt. Chai et al.[14] found that all platinum group elements (PGE), including Au, are retained by various productions of this resin. Chai et al.[14] suggested a scheme for radiochemical separation of noble metals, which is given in Figure 6. They found that their digestion process forms mainly Ir(III), which is not absorbed onto the chelate resin. However, oxidation of the chelated Ir with hot H_2O_2 transforms all Ir to the +4 oxidation state, which is absorbed on another column with the same resin.

An alternative process for the separation of the noble metals, suggested by Chai et al.,[14] is based on extraction of noble metals with long-chain primary amines (19–23 carbon atoms) in DCE. All the noble metals are extracted in high yields into the organic phase. The base metals were only negligibly extracted.

Another group besides the noble metals, which is determined many times by RNAA, is the rare earth elements (REE) group. Zilliacus et al.[15] described two procedures for the separation of the REE group from irradiated geological samples. A simple and faster method was developed for samples with concentrations above 0.5 ppm, while a method with more separation steps was developed for samples with lower concentrations. A more time-consuming process is required to obtain a clean REE fraction. The two processes are described in Figure 7. Both methods are based on fusion with

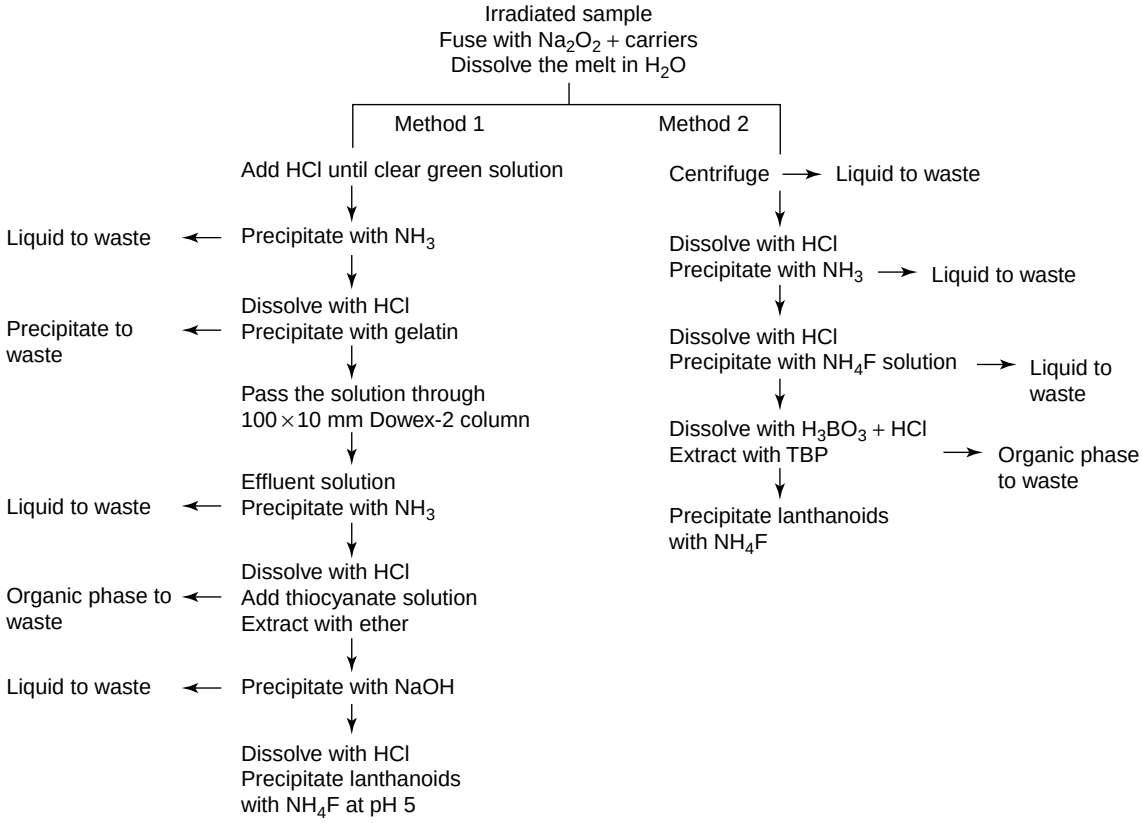

Figure 7 Separation scheme for the analysis of REE in geological samples (Reproduced from Zilliacus et al.[15] with permission.)

Na_2O_2 and cycles of precipitation as hydroxides and as fluorides. In the simple method there is one more step for removal of Sc and other metals by extraction with tributylphosphate (TBP). For the lower concentration samples, SiO_2 is removed by precipitation with gelatin in acidic solution, the main interferences are adsorbed on an anion exchange column in a HCl medium, while the REE are eluted. Sc is removed by extraction with ether from SCN^- solution. Purification of the REE fraction is done by a cycle of precipitation of hydroxides and fluorides. A similar process was also developed by Laul et al.[16] They did not use the Sc removal step by extraction; however, in the final step, they performed three cycles of hydroxides–fluoride precipitation.

A similar method was developed by Wandless and Morgan.[17] The method is based on the precipitation of the REE hydroxides both at high pH and at pH 9 in ammoniacal solution. SiO_2 is removed by precipitation with gelatin in acidic solution. Fe is removed by an anion exchanger in the Cl^- form from 8-M HCl solution. Zr, Sc and Hf are removed by an anion exchanger in the SCN^- form from 0.8-M SCN^- + 0.5-M Cl^- solution. The only remaining interference is Cr^{3+}, which is removed by precipitation of the REE with excess 8-M NaOH. The complete scheme is given in Figure 8.

Parthasarathy et al.[18] separated the REE into two groups rather than into one group, as in all other processes. They separated a group of light REE and a group of heavy REE. The process is different from all those previously described, and the flow chart is given in Figure 9.

Morrison et al.[19] described a procedure for chemical group separations, which, with the use of a Compton suppression Ge(Li) detector, can determine 45 elements in geological samples. The flow chart for this process is given in Figure 10. The first group is the one including volatile elements and volatile fluorides. Elements adsorbed to HAP in 8-M HCl formed the second group. Those adsorbed to anion exchange resin in 8-M HCl formed the third and fourth groups. The separation between these two groups is done by elution with 0.5-M HCl. The elements not adsorbed on the anion exchange in the 8-M HCl are separated into two groups by extraction with TBP. Laul et al.[20] developed a RNAA method for trace elements in terrestrial rocks and stony meteorites. They stated that their main emphasis was to minimize chemical procedure and maximize the γ-ray spectrometry aspect. However, their separation is longer than that of Morrison et al.,[19] and they measure only 16 elements. The flow chart of their separation is given in Figure 11. The

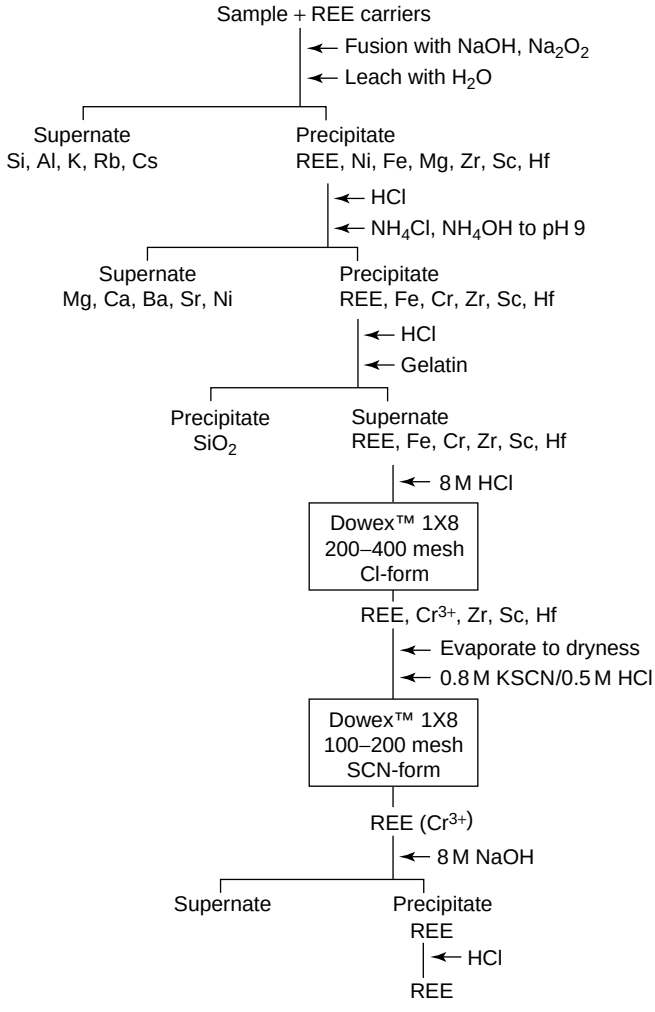

Figure 8 Separation procedure flow chart for RNAA of REE in geological samples. (Reproduced from Wandless and Morgan[17] with permission.)

scheme involves four different separation techniques: distillation, solvent extraction, ion-exchange columns, and precipitation.

Smets et al.[21] developed a group separation for RNAA determination of 24 elements in a wide variety of silicate rocks and minerals. The samples were decomposed in a $HF-HNO_3$ mixture in a PTFE (polytetrafluoroethylene) lined bomb, and were separated into soluble and insoluble fluorides. The soluble fluorides were separated into three groups by sequential elution (0.1-M HF, 3-M HCl + acetone, 12-M HCl) from a cation exchange column. The insoluble fluorides (Ca, Sr, Ba, REE and part of Rb and Cs) were dissolved and purified from iron and scandium activities by extraction with TBP. The separation scheme is given in Figure 12. The elements written in large type were determined, whereas those elements in small

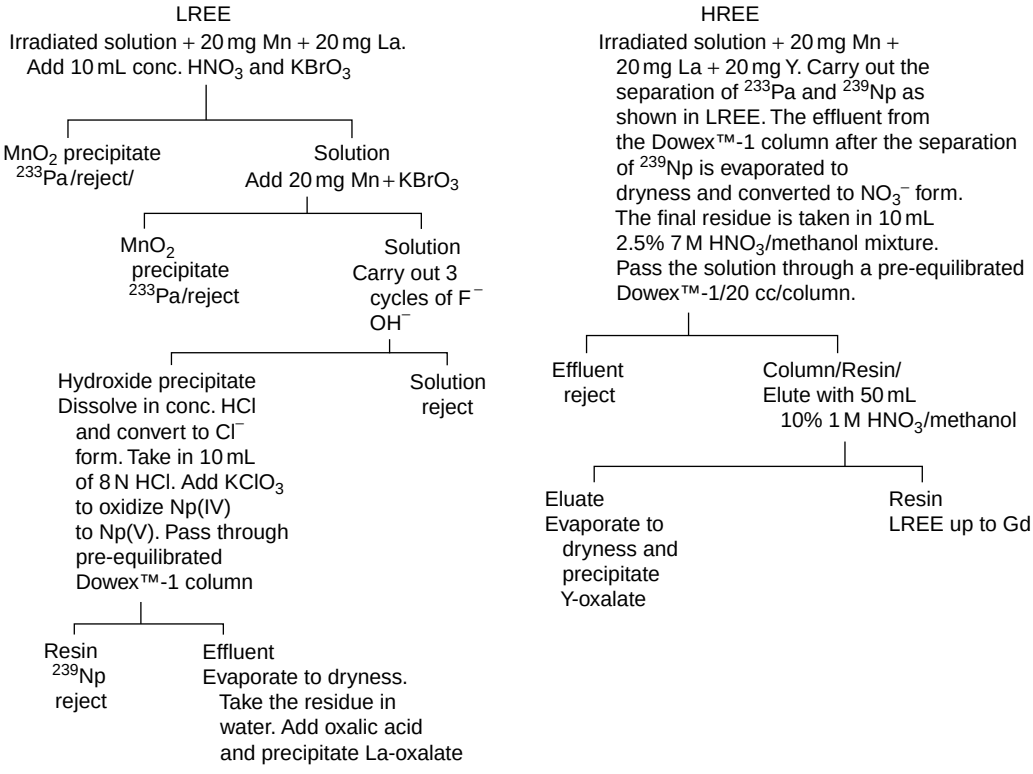

Figure 9 Flow chart for RNAA of REE by separation of two fractions, light and heavy REE (LREE and HREE). (Reproduced from Parthasarathy et al.[18] with permission.)

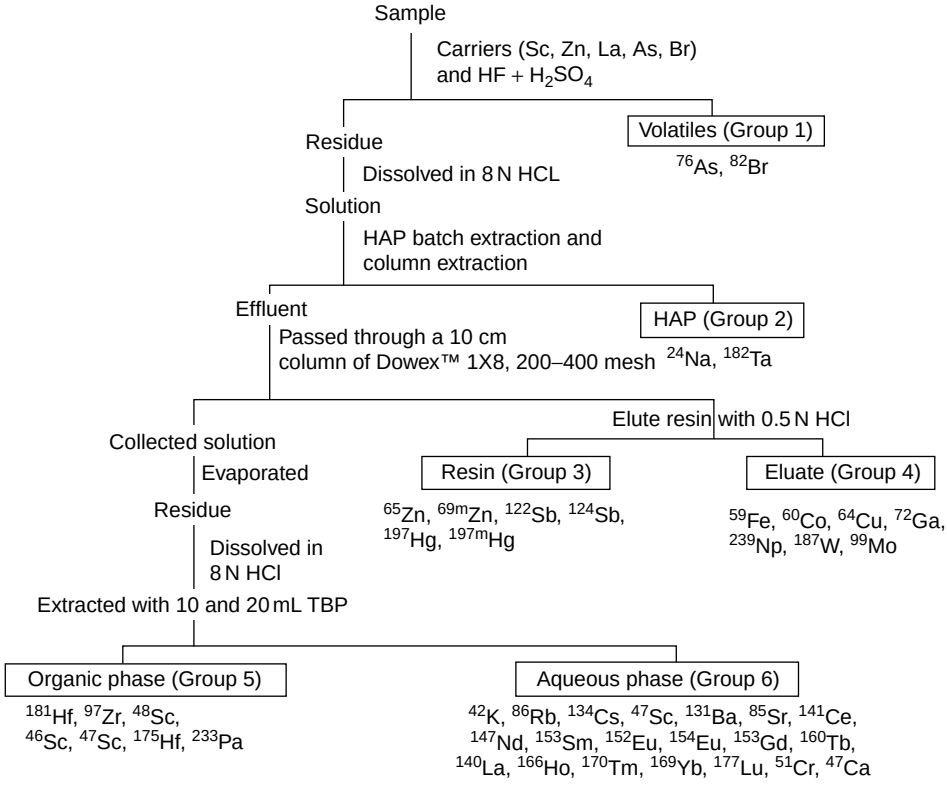

Figure 10 Flow chart for chemical separation of 45 trace elements in geological samples into six groups appropriate for simultaneous counting. (Reproduced from Morrison et al.[19] with permission.)

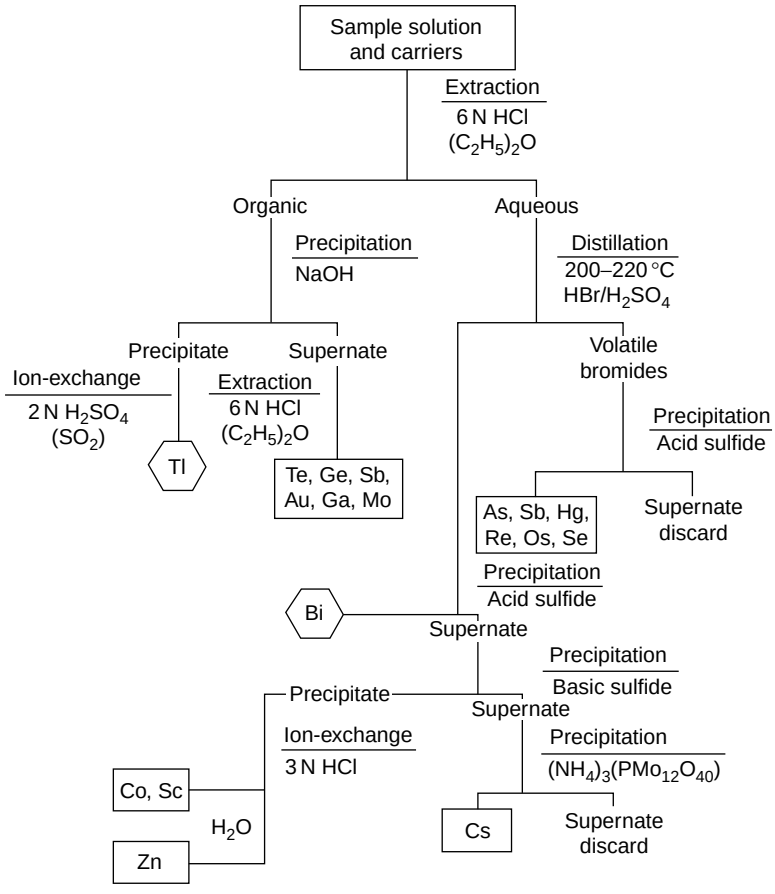

Figure 11 Chemical separation of trace elements in geological samples, leading to determination of 16 elements. (Reproduced from Laul et al.[20] with permission.)

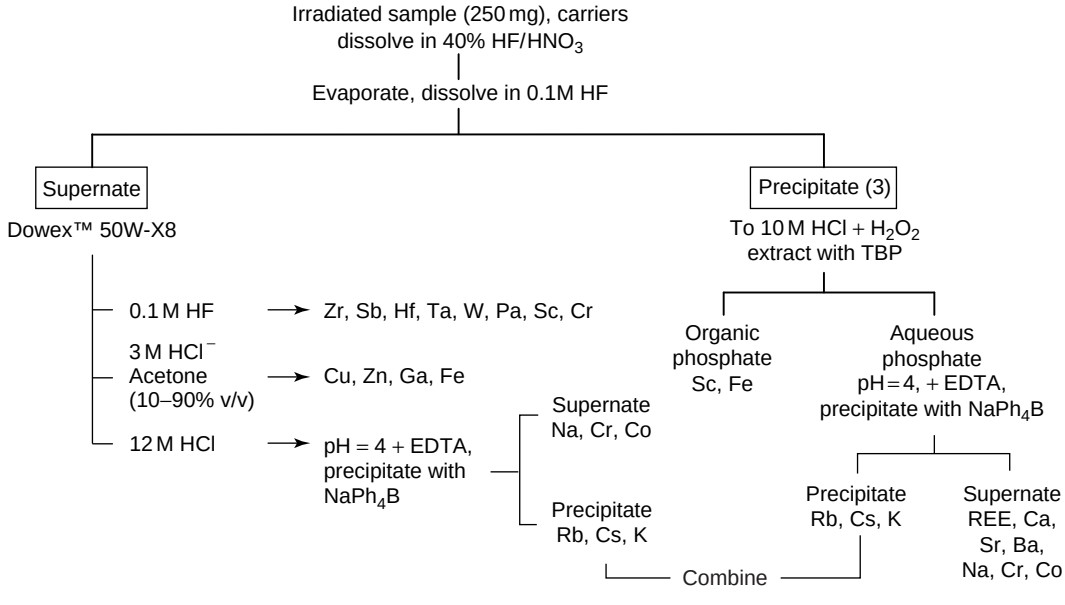

Figure 12 Chemical separation procedure for groups of trace elements in geological samples. (Reproduced from Pietra et al.[22] with permission.)

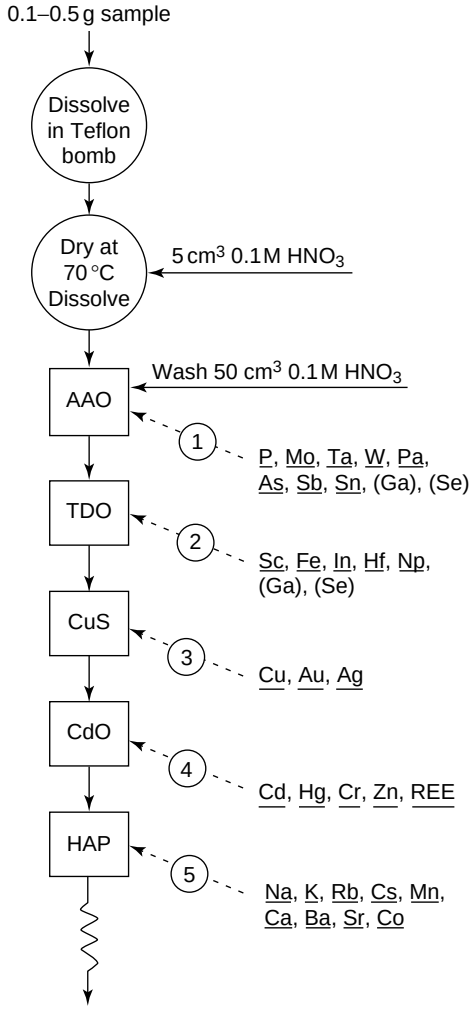

Figure 13 Radiochemical separation scheme for the determination of 39 elements from geological samples. (Reproduced from Pietra et al.[22] with permission.)

type are found in more than one fraction and cannot be determined. In some samples their concentration is too high and they might interfere with the determination of the other elements in the group. For these cases, the authors developed a more thorough scheme, in which for each fraction there is a method of removing the interference. However, these further steps should be done only if γ-spectrometry will not be sufficient.

Pietra et al.[22] developed 22 different RNAA procedures for environmental and biological samples. Many of these procedures are for several elements or for single elements; however, some are for group separations. The most extensive procedure (50 elements) was used only for biological samples. The next most extensive (39 elements) procedure was used for geological samples, and the separation scheme is depicted in Figure 13. The samples were dissolved in acid mixtures in a Teflon bomb, and, after drying, dissolved in 0.1-M HNO_3 which was also used as the only eluant. The trace elements were separated into five groups by the use of five different columns: acidic aluminum oxide (AAO), tin dioxide (TDO), copper sulfide (CuS), cadmium oxide (CdO) and HAP. The separation time for this procedure is about 2 h. All the IRNs were counted in the columns without elution. In another scheme where the geological material was dissolved in 6-M HF, the HAP column was used first, followed by the cation exchange column, and later an anion exchange column can be used to separate the mixture into four groups in which 35 elements can be determined. As in the previous scheme, in order to simplify the procedure and to make it easier for automation, only one eluant (6-M HF) is used.

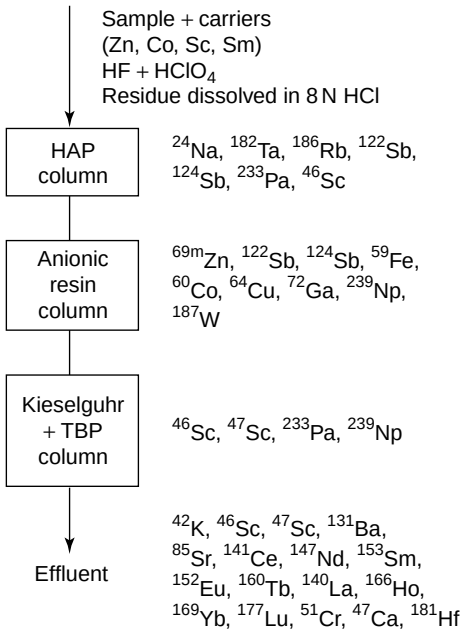

Figure 14 Simple group separation scheme for geological samples. (Reproduced from Vasconcellos and Lima[23] with permission.)

Vasconcellos and Lima[23] developed a procedure that is between several elements and a group separation. It is based on the use of three columns: HAP, anionic resin column and a reverse phase chromatography column with TBP on a Kieselguhr support. Several elements appeared in more than one fraction, but all the REE appear in one group and can be measured. The same is true for Na, Ta, Zn, Fe, Co, Cu, Ga, K, Ba, and Cr. The flow chart for the separation is given in Figure 14.

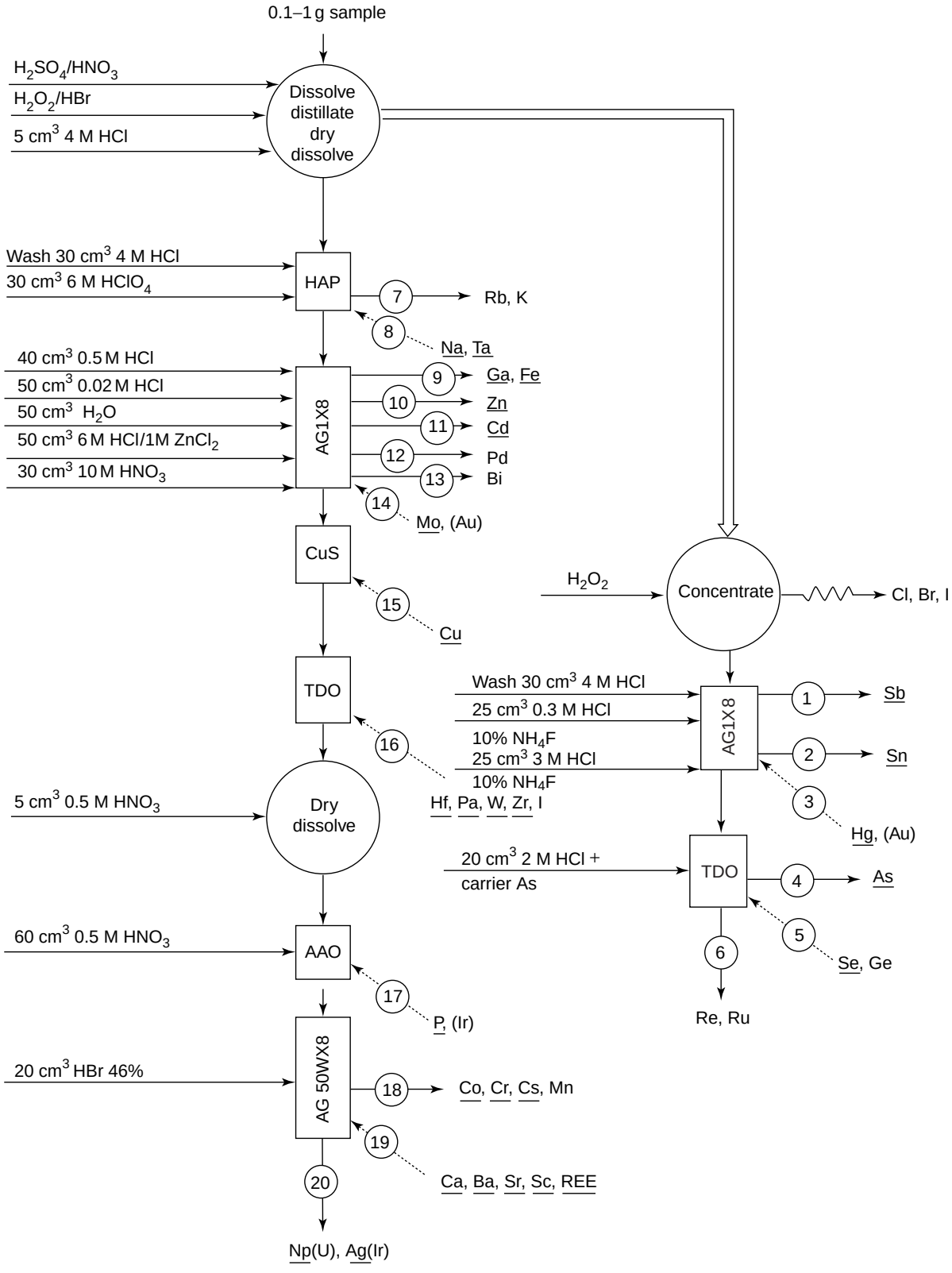

Figure 15 Pietra et al.[22] scheme for radiochemical separation for the determination of 50 elements in biological samples. (Reproduced from Pietra et al.[22] with permission.)

4 BIOLOGICAL SAMPLES

Many group separations were developed for biological materials. Some of them divide the sample into a few groups measured on a Ge(Li) or HPGe detector, while others perform very elaborate separations into many groups to enable activity measurements on very small amounts using a well-type NaI(Tl) detector. Figure 15 shows the flow chart of the Pietra et al.[22] procedure for the determination of 50 elements. The procedure starts with distillation of the elements (Cl, Br, I, Os) and the chlorides–bromides (Sb, Sn, Hg, Au, As, Se, Ge, Re, Ru). The chlorides–bromides are separated into six groups (not more than two elements in the same group) by the use of two columns (first an anion exchanger and then TDO) with successive elutions. An arrow from the column means elution. Elements with arrows to the column are retained on the column. Elements in brackets are present in more than one group. The nonvolatile elements/chlorides/bromides are separated using successive elutions with six columns: anion exchanger, CuS, TDO, AAO, and cation exchanger. The procedure of Pietra et al.[22] for the determination of 39 elements using five

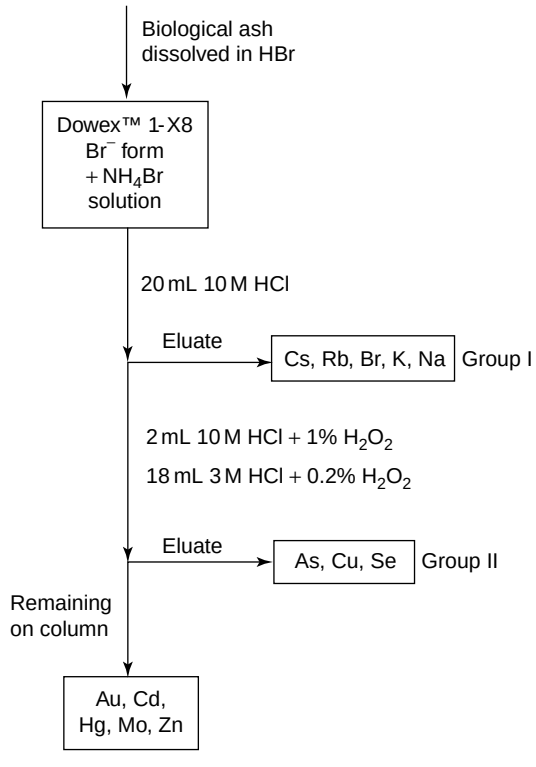

Figure 16 Determination of 10 trace elements in biological specimens by separation into three groups[24] (Na, K and Br were not determined).

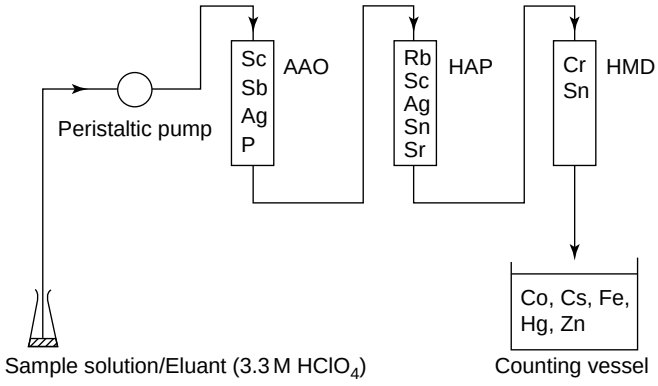

Figure 17 Determination of 16 trace elements in biological samples by separation into four groups using three columns. AAO, acidic aluminum oxide; HAP, hydrated antimony pentoxide; HMD, hydrated manganese dioxide. (Reproduced from Xilei et al.[25] with permission.)

columns with a single eluant, which is described for geological material, was also used for biological material (Figure 13).

Although 50 elements can be separated and determined, most studies do not need so much information, and most studies measured 10–20 elements by more simple methods. Pietra et al.[22] have several schemes for this purpose. van Renterghem and Cornelis[24] determined 10 elements in human serum. Their procedure divides 13 elements (Na, K, and Br were not determined in the serum) into three groups by using an anion exchange column in the Br^- form with two elutions. The scheme is shown in Figure 16.

Sixteen trace elements were determined by the same group using three columns and one eluant,[25] as can be seen in Figure 17.

Schuhmacher et al.[26] developed an almost complete scheme, in order to enable the activities to be measured with a well-type NaI(Tl) detector, to measure the contents of 25 trace elements in small biological samples (1–15 mg). The scheme is based on 14 consecutive columns divided into three Groups (3,7,4). For all columns in one group, the same eluant was used. Between the second and third group the eluate is automatically titrated to pH 2. The scheme is shown in Figure 18.

Figure 19 gives the scheme of Yeh et al.[27] for group separation, measuring 21 elements, not including Na and Br, which can also be measured. This scheme includes distillation of Br as an element, a HAP column, As and Se distillation as bromides, and an AAO column. A more extensive review on RNAA can be found in Alfassi.[28]

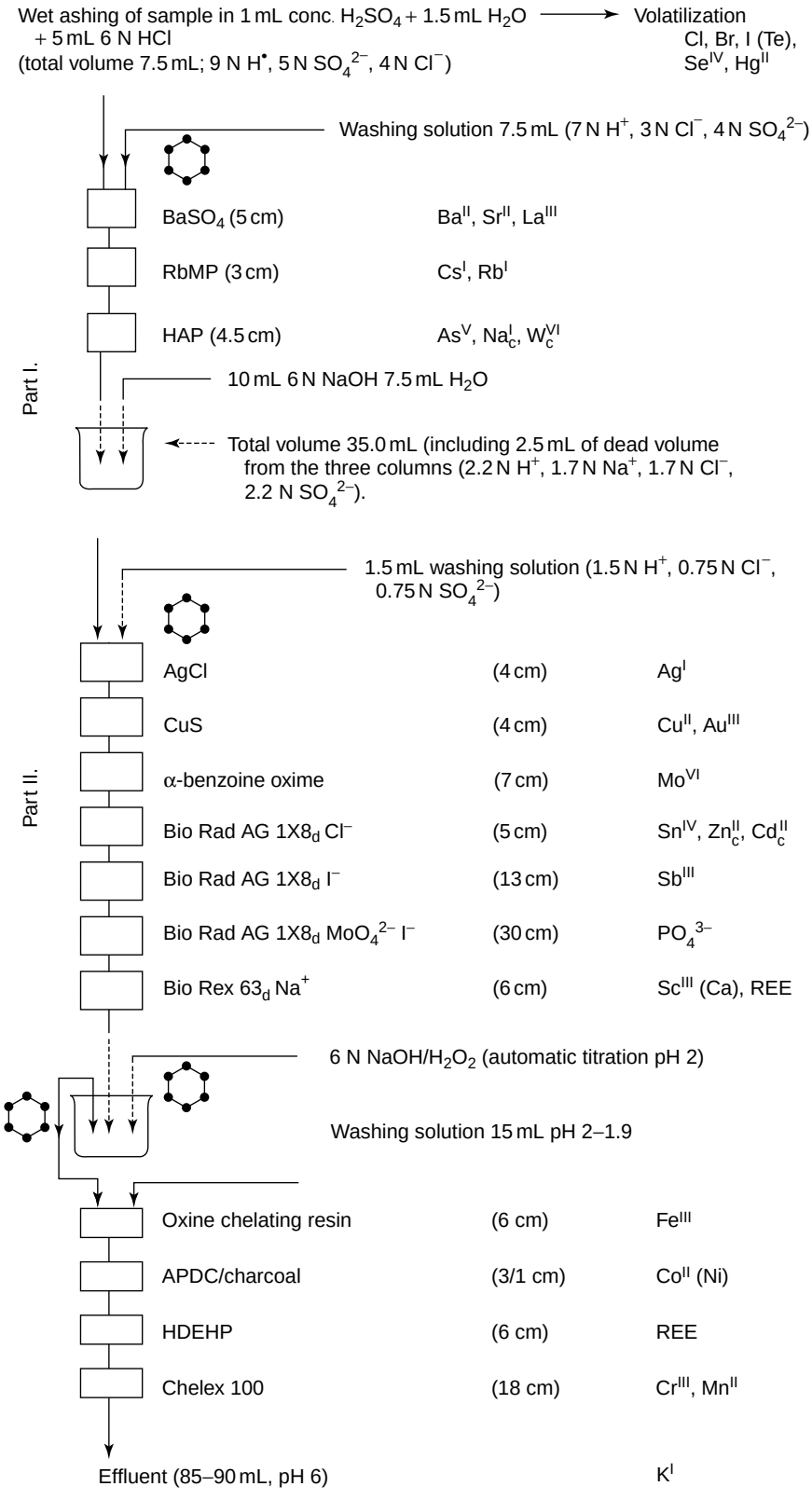

Figure 18 Half automated procedure for almost complete separation of 25 elements in biological samples. RbMP, Rubidium meta phosphate; APDC, ammonium pyrrolidine dithiocarbamate. (Reproduced from Schuhmacher et al.[26] with permission.)

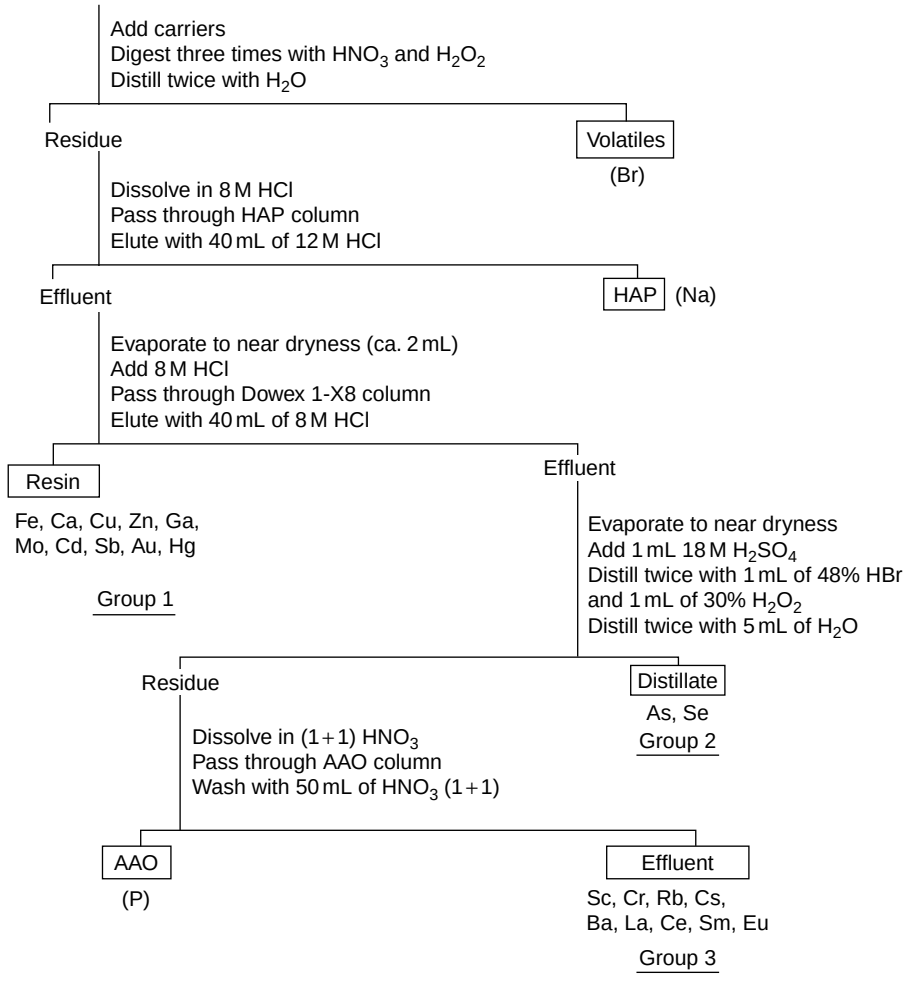

Figure 19 Yeh et al.[27] scheme for RNAA determination of 21 elements. (Reproduced from Yeh et al.[27] with permission.)

ABBREVIATIONS AND ACRONYMS

AAO	Acidic Aluminum Oxide
APDC	Ammonium Pyrrolidine Dithiocarbamate
DAM	Diantipyrilmethane
DCE	Dichloroethane
HAP	Hydrated Antimony Pentoxide
HDEHP	Di(2-ethylhexyl)-orthophosphoric Acid
HMD	Hydrated Manganese Dioxide
INAA	Instrumental Neutron Activation Analysis
IRN	Indicator Radionuclide
PGE	Platinum Group Elements
RbMP	Rubidium Meta Phosphate
REE	Rare Earth Elements
RNAA	Radiochemical Neutron Activation Analysis
TBP	Tributylphosphate
TDO	Tin Dioxide

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

REFERENCES

1. K.P. Egger, V. Krivan, *Fresenius' Z. Anal. Chem.*, **323**, 827 (1986).

2. K.S. Park, N.B. Kim, Y.S. Kim, K.Y. Lee, H.W. Choi, Y.Y. Yoon, *J. Radioanal. Nucl. Chem.*, **123**, 585 (1988).
3. K.H. Theimer, V. Krivan, *Anal. Chem.*, **62**, 2722 (1990).
4. R. Caletka, R. Hausbeck, V. Krivan, *J. Radioanal. Nucl. Chem.*, **120**, 305 (1988).
5. R. Caletka, C. Vorwalter, V. Krivan, *Anal. Chim. Acta*, **141**, 393 (1982).
6. R. Caletka, W.G. Faix, V. Krivan, *J. Radioanal. Chem.*, **72**, 109 (1982).
7. K.P. Egger, V. Krivan, *Fresenius' Z. Anal. Chem.*, **331**, 394 (1988).
8. R.A. Nadkarni, G.H. Morrison, *Anal. Chem.*, **46**, 232 (1974).
9. R.A. Nadkarni, G.H. Morrison, *J. Radioanal. Chem.*, **38**, 435 (1977).
10. S.J. Parry, *Analyst*, **105**, 1157 (1980).
11. H.W. Stockman, *J. Radioanal. Chem.*, **78**, 307 (1983).
12. S.L. Law, *Science*, **174**, 285 (1971).
13. R.A.A. Muzzareli, R. Rocchetti, *Anal. Chim. Acta*, **70**, 465 (1974).
14. C.F. Chai, S.L. Ma, X.Y. Mao, K.N. Liao, W.C. Liu, *J. Radioanal. Nucl. Chem.*, **114**, 281 (1987).
15. R. Zilliacus, M. Kaistilla, R.J. Rosenberg, *J. Radioanal. Chem.*, **71**, 323 (1982).
16. J.C. Laul, E.A. Lepel, W.C. Weimer, N.A. Wognan, *J. Radioanal. Chem.*, **69**, 181 (1982).
17. G.A. Wandless, J.W. Morgan, *J. Radioanal. Nucl. Chem.*, **92**, 273 (1985).
18. P. Parthasarathy, H.B. Desai, S.R. Kaysasth, *J. Radioanal. Nucl. Chem.*, **105**, 277 (1986).
19. G.H. Morrison, J.T. Gerard, A. Travesi, R.L. Currie, S.F. Peterson, N.M. Potter, *Anal. Chem.*, **41**, 1633 (1969).
20. J.C. Laul, D.R. Case, M. Wechter, F. Schmidt-Bleek, M.E. Lipschutz, *J. Radioanal. Chem.*, **4**, 241 (1970).
21. T. Smets, J. Hertogen, R. Gijbels, J. Hoste, *Anal. Chim. Acta*, **101**, 45 (1978).
22. R. Pietra, E. Sabbioni, M. Gallorini, E. Orvini, *J. Radioanal. Nucl. Chem.*, **102**, 69 (1986).
23. M.B. Vasconcellos, W.F. Lima, *J. Radioanal. Chem.*, **44**, 55 (1978).
24. D. van Renterghem, R. Cornelis, in *Trace Element Analytical Chemistry in Biology and Medicine*, eds. P. Bräter, P. Schramel, Walter de Gruyter, Berlin, 55, 1988.
25. L. Xilei, D. van Renterghem, R. Cornellis, L. Mees, *Anal. Chim. Acta*, **211**, 231 (1988).
26. J. Schuhmacher, W. Maier-Borst, H. Hauser, *J. Radioanal. Chem.*, **37**, 503 (1977).
27. S.J. Yeh, P.Y. Chen, C.N. Ke, S.T. Hsu, S. Tanaka, *Anal. Chim. Acta*, **87**, 119 (1976).
28. Z.B. Alfassi, in *Determination of Trace Elements*, ed. Z.B. Alfassi, VCH, Weinheim, 309, 1994.

Radiotracer Methods

Hendrik A. Das
Netherlands Energy Research Foundation (ECN), and University of Amsterdam, Amsterdam, The Netherlands

1 Introduction 1
2 Principle and Formulation 1
2.1 General 1
2.2 Radiotracer Experiments in Closed Systems 2
2.3 Radiotracer Experiments in Open Systems 3
2.4 Limitations of Radiotracer Methods 3
2.5 Practical Applications of Radiotracers 4
3 Measurement of Spatial Distributions 5
3.1 Survey 5
3.2 Profiling in Solids 5
3.3 Radiography and Tomography 6
3.4 Diffusion in Wetted Granular Solids 6
3.5 Diffusion in Liquids 8
4 Observations on Phase Equilibria and Related Kinetics 8
4.1 Survey 8
4.2 Speciation-controlled Diffusion 8
4.3 Kinetics of Colloid Association 9
4.4 Evaporation Kinetics of Traces of Organic Pollutants 9
Related Articles 10
References 10

Small amounts of radionuclides of suitable half-life can be used to observe the kinetics of mass transport and isotopic exchange in the equilibrium distribution of a (micro-)component between coexistent phases.

Quantitative assay of radioactivity is sensitive, precise and accurate and usually performed instrumentally. Though handling of radionuclides is restricted to authorized laboratories only, the combination of its potential and simplicity makes the radiotracer method an indispensable part of modern analytical chemistry.

1 INTRODUCTION

Small changes in isotopic composition by addition of minute quantities of either stable isotopes or radionuclides offers the possibility of labeling ions and molecules without significant shifts in their thermodynamic properties. This ''tracer'' principle enables the observation of kinetics, including isotopic exchange and equilibrium distributions.

The measurement of stable tracers is performed by mass spectrometry, usually on small aliquots of solution. As a rule, assay of radiotracers may be achieved by direct β or γ counting of relatively large subsamples. In case of α-active radiotracers, sample pretreatment is mandatory to obtain a thin-layered counting aliquot.

Radiotracers are applied in authorized laboratories only. Owing to the sensitivity of detection, the amount of activity, expressed in disintegrations per second per Becquerel (Bq) can be at quite a modest rate, usually below 10^5 Bq. Such quantities can be with handled with simple precautions in the many low-level tracer laboratories in hospitals, industry and university laboratories.

The most frequent application of the method is in medical diagnosis, followed by metabolic studies, environmental transport and availability experiments, including pilot plant simulations. The optimization of benchtop analytical procedures constitutes another established range of uses. This text concentrates on the principles of the radiotracer method and its nonmedical applications. A few illustrative cases have been taken from environmental studies, material science and analytical procedure development.

2 PRINCIPLE AND FORMULATION

2.1 General

The basic idea of the radiotracer method is the supposed identical behavior of stable and radioactive nuclides. Thus the presence of a tiny amount of an isotopic radionuclide, negligible in weight but well able to be detected by its radiation, enables observation of the ''labeled'' element. The chemical species of the radiotracer should be equal to that of the bulk mass in question. Thus both ions and (labeled organic) molecules are used. The small influence of the mass difference can be neglected in the large majority of applications. It becomes significant for light elements only, from about oxygen to elements of lower atomic mass, and in some artificial arrangements for isotope separation.

The signal obtained from a radiotracer depends on the concentration and the specific activity of the compound or phase which is being measured. In the case of a closed system, a radiotracer experiment gives information on the rate of the net mass-transport which affects the concentration, and on the rate of isotopic exchange which

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

influences the specific activity of the compound of interest. In open systems this distinction is not generally made but the variation of the signal is considered to be proportional to the changes in concentration.

2.2 Radiotracer Experiments in Closed Systems

The aim of radiotracer experiments in closed systems is to observe the interaction between two or more phases and compounds and, eventually, their equilibrium situation. Usually an amount of radioactivity is added as a component of one phase or compound and its quantity measured as a function of time. If the reaction between two compounds in one phase is examined, separation has to be performed prior to the measurement. Often an aliquot is taken for counting.

The result is a number of counts, A, related to the specific activity, a (disintegrations $s^{-1} g^{-1}$ or disintegrations $s^{-1} mol^{-1}$) and the concentration, c ($g\,L^{-1}$ or M) by Equation (1)

$$A = \gamma\, a\, c \tag{1}$$

The constant of proportionality, γ (in counts disintegration^{-1} s L or counts disintegrations^{-1} s mol), accounts for the volume of the aliquot and the counting conditions. At the beginning of the experiment Equation (2) follows

$$A_0 = \gamma\, a_0\, c_0 \tag{2}$$

and thus, Equation (3)

$$\frac{A}{A_0} = \frac{a}{a_0}\frac{c}{c_0} \tag{3}$$

If we now write Equation (4)

$$f_A = \frac{A}{A_0}, \quad f_a = \frac{a}{a_0} \quad \text{and} \quad f_c = \frac{c}{c_0} \tag{4}$$

it follows that Equation (5)

$$f_A = f_a f_c \tag{5}$$

and thus, with $df/dt = f'$, Equation (6)

$$f'_A = f_a f'_c + f_c f'_a \tag{6}$$

Equation (6), the activity balance, together with the mass balance, forms the basis of all radiotracer experiments in a closed system. In kinetic studies the third obvious equation is the rate balance: at equilibrium the net rate of two opposing reactions must be zero.

Without net mass transport $f_c = 1$ and $f'_c = 0$ and thus $f'_A = f'_a$ and $f_A = f_a$. In this case, a gradual isotopic exchange can be observed. If on the other hand there is no isotopic exchange, the reverse holds and $f_A = f_c$, enabling the measurement of net mass transport.

When both a and c change during the experiment, extra information is needed to solve Equation (6). This may take the form of a separate determination of f_c or a logical prediction of f_a. The various possibilities are summarized in Figure 1.

A clear distinction must be made between the amount of radiotracer which is present in the system and that which is available: in leaching experiments on radioactivated solids the available amount increases nonlinearly with time.

In general, the moment of introduction of the radioactivity does not coincide with the beginning of the

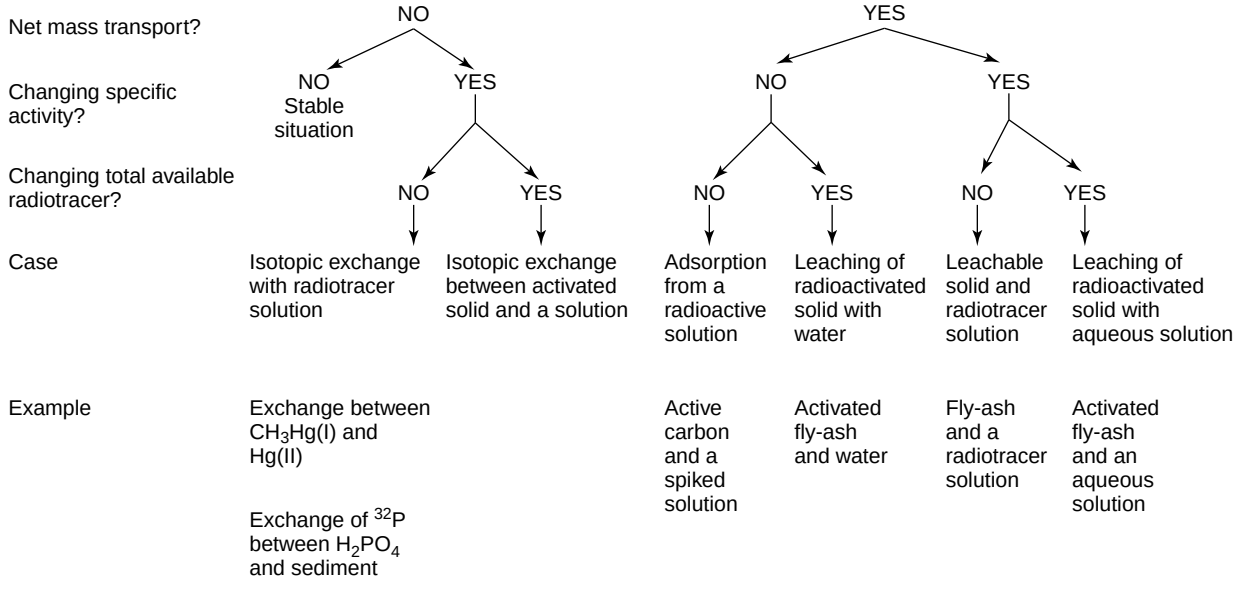

Figure 1 Survey of radiotracer experiments in a closed system.

interaction. By varying that moment, a (decreasing) ratio between net mass transfer and isotopic exchange may be observed.

Applications of radiotracer experiments to closed systems refer to interactions between compounds in one phase or different, adjacent, phases. As isotopic exchange or net mass transfer may be requested, this makes four possible combinations.

2.2.1 Isotopic Exchange

Generally, the reaction is of the type $AB + A^*C \longleftrightarrow A^*B + AC$. The asterisk indicates the radiotracer. By definition, the time of its introduction is taken as zero. By plotting the logarithm of the exchanged fraction against time, the rate constant can be obtained.[1–3] A similar formalism is applicable to isotopic exchange between two adjacent phases.

2.2.2 Net Mass Transport and Final Spatial Distribution

The case of net mass transport at a constant specific activity is by far the most frequent among radiotracer applications. A complication arises when the total available amount of radioactivity changes with time, as in the leaching of radioactivated or labeled solids. The activity of the leachant is measured as a function of time. It is obvious that the specific activity will remain constant only when the fresh eluent does not contain this element of interest. The mass balance may be written as Equation (7)

$$V\frac{\mathrm{d}c}{\mathrm{d}t} = GF(t) \tag{7}$$

where V is the volume of eluent (mL), c is the concentration in the eluent ($\mathrm{g\,mL^{-1}}$), t is the time (h), G is the amount of solid (g) and $F(t)$ is the specific mass transfer function ($\mathrm{g\,g^{-1}\,h^{-1}}$). Measurements of c as a function of t yields $F(t)$.

Table 1 gives the empirical expression for $F(t)$ and those for f_A, the number of counts as a function of time for a leachant without or with some initial concentration of the analyte.

2.3 Radiotracer Experiments in Open Systems

Radiotracers are used in open environmental systems to measure flow rate and/or time of residence as well as their distributions. This is invariably done by a one-time (δ-function) injection. Either the radioactivity is added to the system or induced by activation in small aliquots taken from the system. Unlike experiments in a closed system, f_c only is considered. Thus $f_a = 1$ and $f_A = f_c$, implying that the radiotracer remains in its original phase.

In principle, there are no special rules involved in the use of a radiotracer as the tagging agent. Thus the formalism developed by Danckwerts[4] and expanded in many textbooks, like that of Levenspiel,[5] may be used. The distribution of the radiotracer over the effluent, $f(t)$, is determined. From this the first and the second moment, average residence time and its variance are calculated. These measurements are mandatory in column leaching studies as the flow pattern influences the activity elution curve.

2.4 Limitations of Radiotracer Methods

Applications of radiotracers are restricted by fundamental as well as practical reasons. The limited choice of radiotracers and their half-lives and the isotope effects in reaction rates and equilibrium concentrations can be regarded as fundamental restrictions. The requirements for the operation of a radiochemical laboratory and the mandatory safety precautions in field experiments are of a practical nature.

The finite number of useful radionuclides and the limitations in their availability are reflected by any commercial catalogue. The most important consequence is the predominance of β-counting, usually by liquid scintillation, in physiological tracer experiments.

The influence of the mass difference between the natural isotopic mixture and the radiotracer can be expressed in terms of empirically determined correction factors for equilibria and reaction rates.[6,7] In elements with atomic masses greater than oxygen, the effect is usually lower than 5%.

Table 1 Functions f_c and f_A for the interaction between an activated leachable solid and an aqueous solution

c_0	$= 0$	$\neq 0$
$F(t)$	$k_1 e^{-\alpha t} - k_2(c - c_{eq})$	
$f_c = c/c_{eq}$	$\frac{k_1}{c_{eq}} \frac{e^{-\alpha t} - e^{\frac{-k_2 G}{V} t}}{\left(\frac{k_2 G}{V} - \alpha\right)} + \left(1 - e^{\frac{-k_2 G}{V} t}\right.$	$\frac{k_1}{c_{eq}} \frac{G}{V} \frac{e^{-\alpha t} - e^{\frac{-k_2 G}{V} t}}{\left(\frac{k_2 G}{V} - \alpha\right)} + \left(1 - e^{\frac{-k_2 G}{V} t}\right. + \frac{c_0}{c_{eq}} e^{\frac{-k_2 G}{V} t}$
$f_A = A/A_{eq}$	identical to f_c	$\left\{\frac{[c_0 V/G + k_1/\alpha](1 - e^{-\alpha t})}{c_0 V/G + k_1/\alpha(1 - e^{-\alpha t})}\right\} f_c$

2.5 Practical Applications of Radiotracers

The numerous applications of radiotracers can be systematized by division into three broad categories:

- optimization and performance of laboratory determinations by chemical separation
- measurement of spatial distributions
- observation of phase equilibria and related kinetics.

Some major trends are given below.

2.5.1 Application of Radiotracers in Laboratory Determinations by Chemical Separation

Optimization and, eventually, actual performance of trace determinations in the analytical laboratory may be done efficiently by means of radiotracers. Inorganic applications usually rely on γ-emitting radionuclides, while organic (i.e. biological and medical) experiments are mostly based on β-emitters. A second distinction is that between off- and on-line measurements. Obviously, the latter are preferred whenever they are feasible.

An example of an inorganic application is met in the determination of Hg^{II} in water by reduction with $SnCl_2$ and aeration; the escaping mercury vapor is trapped on active carbon.[8] Figure 2(a–c) illustrate to this procedure.

Radionuclides used in the separation and determination of organic trace compounds by high-performance liquid chromatography and off- or on-line counting are listed in Table 2; Figure 3 shows the flow-sheet of the apparatus.[9,10]

The choice between off- and on-line counting depends on the scope of the experiment and the specific activity of the radiotracer. The advantages of off-line counting are the variable counting time, the stable and low count rate of the background and the absence of any influence of chemiluminescence. The accompanying disadvantages comprise the inevitable ''breaking-up'' of the originally continuous chromatogram, the risk of irreproducible fractions and the time-consuming laborious procedure. Off-line scintillation counting is convenient when a small number of well-separated peaks are expected and if the amount of radioactivity, governed by the specific activity, is low.

Flow-through detection with a water-miscible liquid scintillator implies its continuous addition, usually at flow rates between 0.1 and 1 mL min^{-1} and subsequent mixing. The detector cell is ~50 µL in volume, thus appreciably larger than the usual ~10 µL cells of other detection methods. The advantages of flow-through counting are the automatic procedure and the high counting efficiency. The obvious disadvantage is band broadening in the cell.

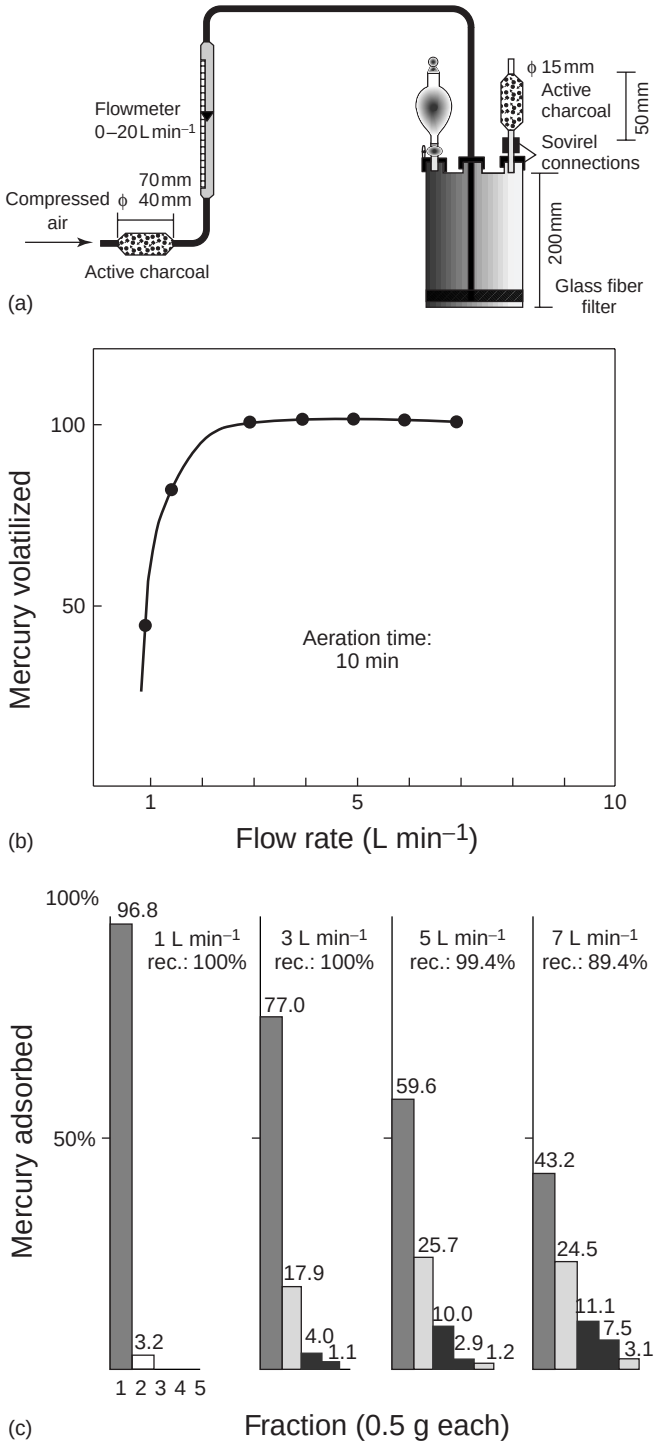

Figure 2 Determination of mercury in water samples. (a) The aeration vessel. (b) Influence of flow rate on volatilization of mercury. (c) Influence of flow rate on adsorption of mercury on activated charcoal. rec. = recovery.

Solvent segmentation based on postcolumn extraction/segmentation of the column eluate with a suitable liquid scintillation is the best option.[11]

Table 2 Radionuclides used in liquid scintillation counting

	E_γ(MeV)	E_β(keV)	$T_{1/2}$	Specific activity[a] (TBq mmol^{-1})	Range[b] (μm)
^{3}H		19	12.4 years	1	4
^{11}C	0.511		20.4 min		
^{14}C		155	5760 years	2 GBq mmol^{-1}	264
^{13}N	0.511		9.96 min		
^{15}O	0.511		2.03 min		
^{18}F	0.511		109.7 min		
^{32}P		1710	14.3 days	200	7870
^{35}S		167	87.4 days	30	302
^{36}Cl		700	3×10^5 years	3 KBq mmol^{-1}	2707
^{45}Ca		258	164 days	45 GBq mmol^{-1}	636
^{59}Fe	1.1; 1.2	500; 1600	44.6 days	40 GBq mmol^{-1}	
^{63}Ni		67	100 years	20 GBq mmol^{-1}	54
^{67}Ga	0.18; 0.30		78 h	25	
^{99m}Tc	0.14		6 h		
^{125}I	0.035	(X-ray)	60 days	75	
^{131}I	0.36	610	8 days	550	
^{140}La	0.49; 1.6	1380	40 h	5 GBq mmol^{-1}	

[a] Specific activities given are about the maximum values specified for commercially available radiolabeled compounds or elements. TBq mmol^{-1} unless stated otherwise.
[b] The maximum range given is calculated for pure β-emitters in water from the formula: Range in g cm^{-2} $= 0.11[(1 + 22.4\,E_\beta^2)^{1/2} - 1]$ with E_β in MeV.

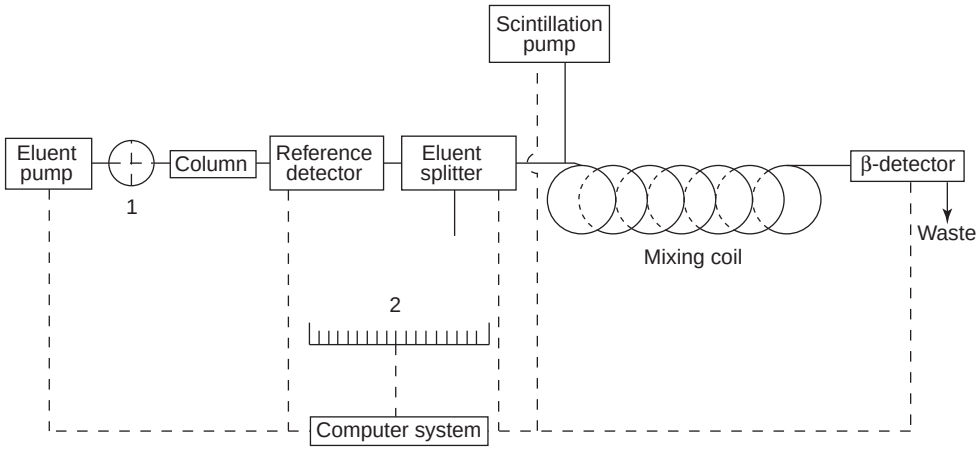

Figure 3 Apparatus for flow-through liquid scintillation counting. 1 = injector, 2 = fraction collector.

3 MEASUREMENT OF SPATIAL DISTRIBUTIONS

3.1 Survey

In assessing spatial distribution, radiotracers are most often used to measure net mass transfer of a previously spiked component within a known matrix. Thus radioanalysis is just one of the options. Conversely, the measurement of isotopic exchange is open to isotopic methods only. Here radioanalysis has to compete with mass spectrometry in measuring coefficients of self-diffusion.

The oldest application in spatial distributions is the study of self-diffusion in metals.[1–3] Relevant uses are met in the profiling of doping agents in semiconductors, tomography and the determination of diffusion coefficients in wetted granular solids and in liquids. In industrial pilot plants and factories, radiotracers are used to measure linear velocity and dispersion as mentioned in section 2.3.

3.2 Profiling in Solids

Flat solid surfaces of spiked samples are probed for their trace concentration profiles by abrasion or controlled

dissolution, followed by measurements on the material removed or the remaining aliquot. In general the first approach gives better precision but the actual advantage over measurement on the remaining aliquot depends strongly on the concentration profile.

The thickness of the layer removed in one step usually varies between 0.5 and 10 μm with a lower limit of ~0.1 μm. The radiotracer method thus fills the gap between macroscopic removal of surface layers and the physical methods for the submicron range.

A severe limitation is imposed by the required properties of the radiotracer and the mandatory combination of a radiochemical and solid-state laboratory which is found only in some university institutes and a few industrial laboratories. The use of special polishing apparatus with copper discs is also implied to obtain a precision of less than 1 μm.[11] The amount removed is determined by microbalance weighing at a precision of ≤10 μg.

Counting of the remaining aliquot is restricted to β-emitting radionuclides and is performed with a solid anthracene detector. Collection of the abrased material on the copper discs and the surface of the remaining aliquot is feasible within 1%. In the case of silicon HF dissolution is necessary which precludes the use of a liquid scintillator for β-emitters and restricts detection to Cerenkov counting and individual quenching correction.

3.3 Radiography and Tomography

There is no possibility or need to discuss the whole field of present day radiotracer-based radiography. Its use in biology and medicine as thin layer radiochromatography is obvious. As a relatively new feature, the use of storage phosphor screens can be mentioned, as this has increased the sensitivity and widened the applicability of double labeling procedures.[12]

From the beginning of the 1980s, computer-aided tomography of three-dimensional objects became a new branch of radionuclide application.[13] As with radiography, this text does not endeavor to treat this new specialization in detail; only principles are signaled.

Tomography is either based on the variation of the object's own γ-count rate with the position towards the detector or on the change in attenuation of transmitted photons or neutrons. The first approach may be based on previous neutron activation[14] or an added tracer; the last option belongs to the medical imaging techniques which will not be considered here. Industrial and nuclear waste tomography use the same principle.[15] The full width at half maximum of the scanning of a ^{198}Au-source was ~7 mm in the early experiments. Much better spatial resolution is obtainable in examining used fuel elements through a <1 mm slit of ~1 m length, as performed at ECN, Petten.

Using a ^{241}Am (59.5 keV) and a ^{137}Cs (661 keV) source in conjunction with collimators of 1–5 mm and distances of 50–150 mm resulted in full width at half maximum of 5–40 mm.[16] Comparable values (9–18 mm) were obtained in transmission and ^{131}I-based emission experiments with phantoms.[17] Off-line measurements on a neutron-activated bone sample by way of ^{24}Na, $T^{1/2} = 15$ h, gave a spatial resolution of 1–2 mm.[18]

On-line measurements in the 1.8×10^{12} cm^{-2} s^{-1} neutron beam at the Grenoble HFR gave ~7 mm full width at half maximum for $CdCl_2$ test pellets of that size.[19]

Apart from the routine measurements on used fuel elements, as performed at ECN, no systematic application of radionuclide-based tomography is reported: simulated fuel[19] and a test loop experiments with ^{24}Na[20] are scarce examples. In sharp contrast, the use of radioactive sources in industrial trouble-shooting as "go devils" is an established practice.

3.4 Diffusion in Wetted Granular Solids

Probably the most practical consequence of differences in element speciation is the variation in environmental mobility. Obviously the water content and the apparent diffusion coefficient of an inert tracer, i.e. one which does not interact with the solid, have to be determined. These data will then serve as reference values.

The use of a miniaturized system for diffusion measurements was first demonstrated in the case of traces of ^{239}Pu.[21,22] Then it was applied to self-diffusion under equilibrium conditions.[23]

Two identical cylinders are pressed together in a syringe. One part contains the radionuclide; the diffusion into the other section is measured. If a "weightless" spike is used, operation is under equilibrium conditions. This procedure yields the coefficient of self-diffusion, which for trace constituents is equal to the effective diffusion coefficient. When a component is involved which is present in solution only the apparent diffusion coefficient(s) obtained reflect(s) the tortuosity and constructivity of the void fraction(s). Experimentally, these two effects combine into one "restriction factor".

If a component is present in both the solid and the liquid phase and if the added radiotracer exchanges rapidly over the phase border, its diffusion will reflect the distribution ratio. When the element involved is present in more than one chemical form the result of the diffusion experiment will depend on the rate of isotopic exchange between the added spike and the (other) species. Only in the limiting cases of either very rapid exchange or no exchange at all, radiotracer diffusion will yield valuable information. Rapid exchange leads to the determination of the mass-averaged apparent diffusion coefficient. No exchange at

all implies the determination of the apparent diffusion coefficient of the spiked chemical form. The simple diffusion Equation (8) applies to a radiotracer which is exclusively present in solution and in one chemical form only and for the case of semi-infinite media.

$$\left(\frac{\partial(ac)}{\partial t}\right)_x = D\left(\frac{\partial^2(ac)}{\partial x^2}\right)_t \tag{8}$$

where a is the specific activity ($\mathrm{Bq\,g^{-1}}$), c is the concentration ($\mathrm{g\,L^{-1}}$), D is the coefficient of self-diffusion ($\mathrm{m^2\,s^{-1}}$), x is the position (m) and t is time (s).

At chemical equilibrium, for a ''weightless'' spike, this expression reduces to Equation (9)

$$\left(\frac{\partial a}{\partial t}\right)_x = D\left(\frac{\partial^2 a}{\partial x^2}\right)_t \tag{9}$$

From a plot of $\ln a$ against x^2 and curve fitting to the solution of Equation (9) for a finite system[23] the apparent diffusion coefficient together with information about the eventual coexistence of more than one flow-channel and their relative importance are obtained.

Under equilibrium conditions and with rapid isotopic exchange Equation (10) gives the experimentally apparent D-value

$$D_e = \frac{D/r}{1 + (m_S/m_L)} \tag{10}$$

where D refers to the ''true'' or unretarded diffusion coefficient in an aqueous solution, r to the mentioned restriction factor and m_S/m_L to the ratio of the exchangeable masses of the element involved in the solid and liquid phases. Whether the assumption of fast isotopic exchange is valid or not can be verified by varying the time lag between the application of the radiotracer to one compartment and the actual joining of the two sections. If this has no influence on the result, the condition is fulfilled.

As an inert radiotracer, tritiated water is used. Environmental investigations refer most often to anions as these feature the highest mobilities. Suitable radiotracers are $^{74}AsO_3^{III}$, $^{75}SeO_3^{II}$, and $^{75}SeO_4^{II}$, $^{99}MoO_4^{II}$ and $^{124}SbO_3^{III}$. It can be concluded that a specific activity of $\leq$10 mCi g^{-1} is sufficient for adding a ''weightless'' spike of $\leq$20 µg g^{-1}.

Figure 4 summarizes the procedure. About 3 g of the sample material is mixed with 0.9 mL of tracer solution and carefully homogenized. The diffusion tube consists of polythene with internal diameter 8 mm and length 70 mm. The interface of the two segments is gently marked with a tiny drop of ^{170}Tm solution.

After a storage time of 5–15 days, the combined segments are taken out, frozen in liquid nitrogen and sliced into 0.5–1 mm coupes. These are dried, weighted and counted. Data are processed to find the most likely D-value. Figure 5 gives the result for the diffusion of tritiated water through a fly-ash. The estimated value of the apparent diffusion coefficient is expressed as pDe. The usual error is 0.1 unit.

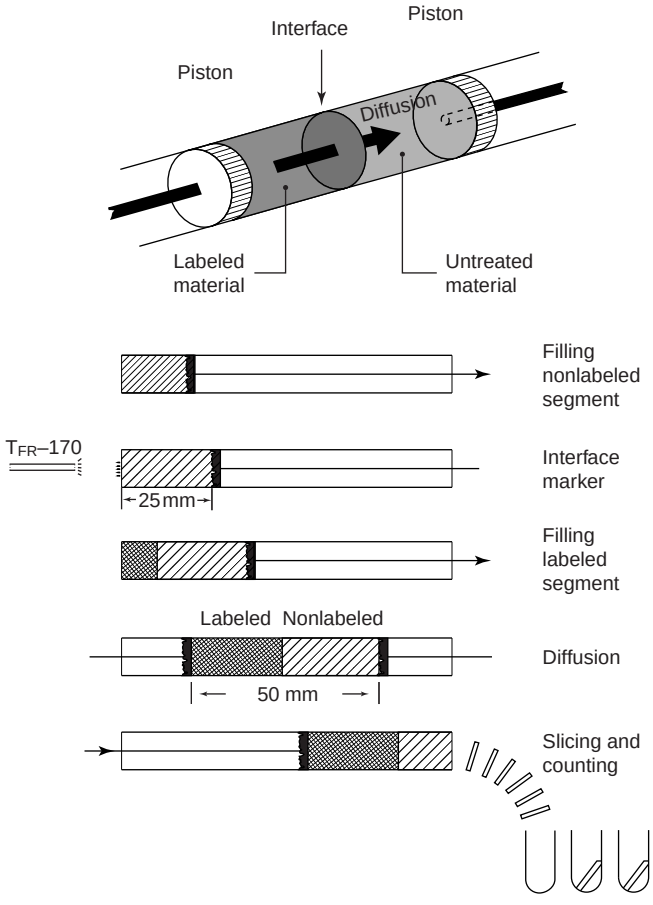

Figure 4 Measurement of the apparent diffusion coefficient After Das.[23]

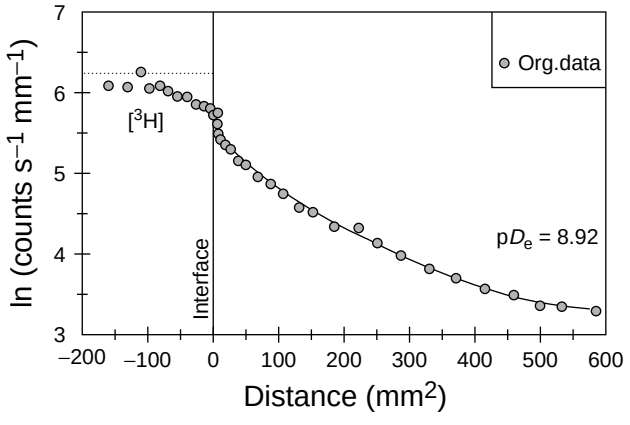

Figure 5 The diffusion of tritiated water in fly-ash. After Das.[23]

3.5 Diffusion in Liquids

The precise and accurate determination of diffusion coefficients of trace constituents (either of ionic or of molecular nature) in dilute, mostly aqueous, solutions is one of the obvious uses of radiotracers in the field of spatial speciation.

The main advantage of radiotracer experiments here, as in wet granular solids, is the possibility of measuring at zero diffusion gradient.[24] Two techniques have been put forward, based respectively on a glass frit diaphragm[25] and an open capillary within a large tank.[24]

Initially, a diaphragm cell of $\sim$100 mL was used.[28] Later, the capillary procedure came into use, offering the advantage of a small spiked volume of $\geq$10 μL per capillary. It implies stirring the outside solution and thus requires rigorous precautions against turbulent flow and mechanical dragging of the spiked liquid from the capillaries. In the final version just one capillary is used, mounted within the well of a bored-through plastic scintillator which monitors continuously the remaining radioactivity. The whole apparatus is mounted in a thermostated room. With a (small) correction for the variation of the counting geometry with proceeding diffusion, a precision of $\sim$0.2% can be reached.[26,27]

Reports on the use of the capillary procedure are found in Kepak et al.[29] and Podhajecky et al.,[30] while a reexamination of the diaphragm procedure is given by Podhajecky et al.[31]

4 OBSERVATIONS ON PHASE EQUILIBRIA AND RELATED KINETICS

4.1 Survey

Within a system in (apparent) equilibrium, elements often occur in different physical phases. The potential use of radiotracers in measuring these distributions, and the kinetics with which they are reached, is obvious. The early, often ''classical'', applications are mentioned in most textbooks on radiochemistry. The number of literature reports on the subsequent use in all branches of science, technology and medicine is enormous. Virtually all applications refer to the measurement of net mass transport (cf. section 2.2).

Current literature abounds with (more or less) defined procedures for physical speciation,[32–41] particulate matter and colloids in natural waters[42–54] and availability measurements on trace metals in granular solid wastes.[55–57] Eventually these investigations pertain to the occurrence and distribution of radionuclides from nuclear bomb tests and waste reprocessing.[58–67] Quoted references give some examples only.

A methodical application is comparison of analytical techniques on water containing particulate matter.[68] Equally, preconcentration and storage have been examined with the help of radiotracers. The contact time between water sample and container, whether polythene or glass, should not exceed more than a few minutes. In situ immobilization, eventually combined with preconcentration, is imperative. Scavenging with active carbon after adsorption of a suitable chelating agent is a possibility.[42] Collection on Chelex-100 or on a combination of an anion and cation exchanger is well known; a survey is given by Florence.[69]

Freezing, in combination with freeze-drying, is the best way of preserving the sample in its original state.[70] Acidification to pH $\sim$ 1–2 is justified only after thorough elimination of particulate matter, to avoid desorption. Using ^{203}Hg, it was found that Hg^{II} desorbs completely at pH = 0.1 with a half-life of $\sim$30 min.[42]

The quest for well-defined and generally applicable speciation schemes has been successful in a few cases only. The separation of colloids from fresh water samples by the hollow fiber technique[47,52] is an example; the standard leaching test on coal fly-ash is another,[55] although a recent comparison revealed some systematic discrepancies between the various modes.[71] Sequential leaching of soils by a series of increasingly aggressive solutions has been applied in a multitrade of schemes, all of limited significance. Several articles[72–82] giving examples have been published.

Applications of radiotracers to adsorption and desorption measurements and in isotopic exchange experiments were mentioned in section 2. Three practical applications are considered here:

- measurement of speciation controlled diffusion through soils[82,84]
- kinetic association experiments with colloids in water[85]
- evaporation kinetics of traces or organic pollutants from surface water.[86]

4.2 Speciation-controlled Diffusion

Figure 6 outlines the apparatus used. The ''steady state'' experiment imposes a linear gradient over the test sample but it takes a considerable time to reach it.[83] Typical experiments with sand samples take $\sim 10^3$ h. The radionuclide concentration in the unspiked reservoir is always much lower than that in the spiked compartment. This eliminates back diffusion.

The concentration in the unspiked cell is monitored until its rate of increase becomes constant. Relevant mathematics are given by Crank.[83] The crucial feature is that $\partial^2 c/\partial x^2$ is zero and thus $\partial c/\partial x$ is constant.

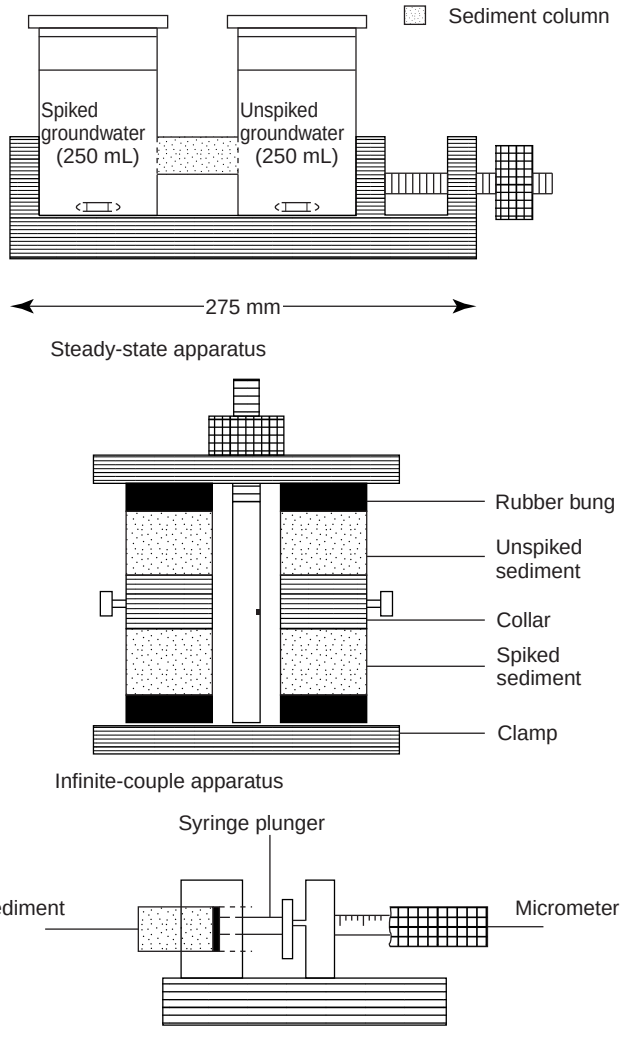

Figure 6 Equipment for steady state diffusion through wet sand.[82] (J.J.W. Higgo, Iodine Speciation and Diffusion in a Sand–Groundwater System, DG Science, Research. Development, 1991. Reproduced by permission of the British Geological Survey.)

"Infinite couple" or "back-to-back" geometry has two cylinders of material, one of them spiked, pressed together. It thus resembles the array discussed in section 4.1, except for its much larger dimensions. These enable the use of large aliquots of groundwater, taken from a separate equilibration tank.[86]

The effective diffusion coefficient and distribution constant was determined for both I^- and IO_3^-, spiked with $^{125}I(T^{1/2} = 60\,\text{day})$, as a function of the solid/solute ratio. Data were found to follow Freundlich isotherms.

4.3 Kinetics of Colloid Association

The use of spikes in radiotracers to determine recoveries in physical and chemical separations is obvious. Their most apparent application is met in the experiments on

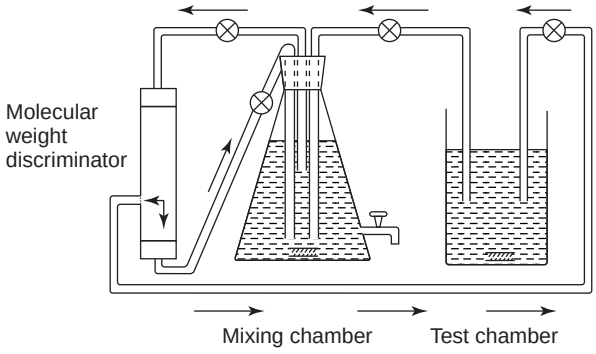

Figure 7 Equipment for the measurement of trace element association with colloids. After Salbu.[47]

hollow fiber preconcentration, using an "Amicon" apparatus, of naturally occurring radionuclides.[47,48,66,85–88] Cartridges with nominal molecular weight cut-off of 10^5, 10^4 and 10^3 are applied. Isolated fractions are analyzed by γ-spectrometry.

Kinetic association experiments are performed with the apparatus shown schematically in Figure 7. The sample is introduced into the test chamber and transported by a peristaltic pump to a mixing chamber and separator. The ultrafiltrate returns to the test chamber while the retained colloids go to the mixing chamber. The decrease in tracer radioactivity in the test solution reflects dilution, association with naturally occurring colloids and, eventually, sorption to the equipment.

It is found that equilibrium is usually reached in 1–2 h. The majority of naturally occurring ^{60}Co and added ^{65}Zn is in the $>10^4$ D fraction.

An alternative to dialysis may be found in electrophoresis.[33,64] Free liquid electrophoresis was applied to speciation studies of radium in natural waters using ^{224}Ra as the radiotracer.[64] Results indicate that in 10^{-2} M chloride solutions the Ra^{II} cation is the main chemical species between pH = 2 and pH = 7.

4.4 Evaporation Kinetics of Traces of Organic Pollutants

Radiotracers are applied to environmentally important kinetics. In evaporation measurements, the overall mass transfer coefficient, K, in $\text{mol}\,\text{min}^{-1}\,\text{cm}^{-2}$ has to be obtained.[89] To this end, evaporation experiments are conducted with labeled trace constituents from open vessels with a known gas–liquid surface area. The concentration of the (volatile) trace compound decreases exponentially with a characteristic constant KS/V_L for a given surface area S (cm^2) and liquid volume V_L (mL).

In a closed flask, the equilibrium concentration in the solution reflects the thermodynamic potential of the

trace constituent in the liquid phase. By measuring this equilibrium as a function of the initial concentration, eventual deviation from ideality can be detected.

''Carrier-free'' ^{3}H- or ^{14}C-labeled compounds are used to spike stock solutions of the chemicals of interest. With commercially available specific activities of $\sim$1 TBq ^{3}H (m mol^{-1}), a liquid volume of $\sim$50 mL, an aliquot size of 2 mL and a counting time of 100 min, the limit of determination at $\sigma_{rel} \sim 1\%$ is 10^{-2} ppb. For ^{14}C-labeled compounds of $\sim$2 GBq ^{4}C (m mol^{-1}) the lower limit is $\sim$5 ppb.

RELATED ARTICLES

Forensic Science **(Volume 5)**
Atomic Spectroscopy for Forensic Applications

Nuclear Methods **(Volume 14)**
Elemental Analysis by Isotope Dilution • Instrumental Neutron Activation Analysis • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Rutherford Backscattering Spectroscopy • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Actinides and other Alpha-emitters, Determination of • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Mass Spectrometry of Long-lived Radionuclides

REFERENCES

1. M. Haïssinsky, *Nuclear Chemistry and its Applications* (English translation), Addison-Wesley, Reading MA, 536–565, 1964.
2. H.A.C. McKay, *Principles of Radiochemistry*, Butterworths, London, 393–445, 1971.
3. G. Friedlander, J.W. Kennedy, J.M. Miller, *Nuclear and Radiochemistry*, 2nd edition, J. Wiley & Sons, New York, 191–220, 1964.
4. J.M. Smith, *Chemical Engineering Kinetics*, 3rd edition, McGraw-Hill Int. Book Company, London, 268–297, 1981.
5. O. Levenspiel, *Chemical Reaction Engineering*, J. Wiley & Sons, New York, 242–308, 1962.
6. L. Melander, *Isotope Effects on Reaction Rates*, Ronald Press, New York, 46–64, 1960.
7. M. Haïssinsky, *Nuclear Chemistry and its Applications* (English translation), Addison-Wesley, Reading, MA, 640–669, 1964.
8. H.A. van der Sloot, H.A. Das, 'The Determination of Mercury in Water Samples from the Environment by Neutron Activation Analysis', *Anal. Chim. Acta*, **73**, 235–244 (1974).
9. A.C. Veltkamp, H.A. Das, R.W.Frei, U.A.Th. Brinkman, E.C.N. Petten, 'On-line Low-level Radiometric Detection of [14C]Remoxipride in Liquid Chromatographic Effluents. Application to Urine Samples', *J. Chromatogr.*, **384**, 357–369 (1987).
10. A.C. Veltkamp, H.A. Das, U.A.T. Brinkman, 'Direct and Reverse Flow-through Heterogeneous Scintillation Counting of Radiolabeled Amino Acids Using Post-column Solvent Segmentation', *J. Chromatogr.*, **508**(2), 301–308 (1990).
11. L.M.L.J. Leblans, M.L. Vetcheyke, *Philips Tech. J.*, **25**, 103–106 (1963).
12. R.R. Johnston, S.C. Pickett, D.L. Barker, 'Quantitative Double-label Autoradiography Using Storage Phosphor Imaging', *Electrophoresis*, **11**, 355–360 (1990).
13. G. Davies, N.M. Spyrou, I.G. Hutchinson, J. Huddleston, 'Applications of Emission Tomography in the Nuclear Industry', *Nucl. Instrum. Methods Phys. Res., Sect. A*, **A242**(3), 615–619 (1986).
14. T.B. Pierce, J. Huddleston, I.G. Hutchinson, 'A Study of the Possible use of Tomographic Techniques for Examining the Distribution of Activity in Irradiated Samples', *Radiochem. Radioanal. Lett.*, **48**, 81–92 (1981).
15. D.J. Gentle, N.M. Spyrou, 'Region of Interest of Tomography in Industrial Applications', *Nucl. Instr. Methods*, **A299**, 534–537 (1992).
16. T.B. Pierce, J. Huddleston, I.G. Hutchinson, 'The Preliminary Investigation into the Potential Use of Tomographic Techniques for the Examination of Mechanical Components', *J. Radioanal. Chem.*, **79**, 287–301 (1983).
17. J.M. Sanders, N.M. Spyrou, 'Improvement of Image Quality in Emission and Transmission Tomography by Radiation of Scattered Photons', *Nucl. Instr. Methods*, **A221**, 93–97 (1984).
18. N.M. Spyrou, 'Prompt and Delayed Radiation Measurements in the Elemental Analysis of Biological Materials: The Case for Neutron Induced Gamma-ray Emission Tomography', *J. Radioanal. Nucl. Chem.*, **110**(2), 641–653, (1987).
19. N.M. Spyrou, Kusminazto, E.G. Nicolau, '2D Reconstruction of Elemental Distribution Within a Sample Using Neutron Capture Prompt Gamma-rays', *J. Radioanal. Nucl. Chem.*, **112**, 57–64 (1987).
20. D.J. Finnigan, F. McFall, J. Huddleston, I.G. Hutchinson, 'Application of Emission Tomography to the Study of Solute Concentration Processes in PWR Steam Generators', *Water Chem. Nucl. React. Syst.*, **5**(2), 182–183 (1990).

21. F. Schreiner, S. Fried, A.M. Friedman, 'Measurement of Radionuclide Diffusion in Ocean Floor Sediment and Clay', *Nucl. Technol.*, **59**, 429–438 (1982).
22. B. Torstenfeld, 'Migration of the Fission Products Strontium, Technetium, Iodine and Cesium in Clay', *Radiochim. Acta*, **39**(2), 97–104 (1986).
23. H.A. Das, H.A. van der Sloot, G.J. de Groot, R.C.H. Steneker, 'Determination of the Apparent Diffusion Coefficients in Water Containing Granular Solids', *Int. J. Appl. Radiot. Isot.*, **40**, 615–619 (1989).
24. H.J.M. Bowen, *Chemical Applications of Radioisotopes*, London, Methuen, 1969.
25. R. Mills, 'A Remeasurement of the Self-diffusion Coefficients of Sodium Ion in Aqueous Sodium Chloride Solutions', *J. Am. Chem. Soc.*, **77**, 6116–6119 (1955).
26. R. Mills, E.W. Godbole, 'The Precision Measurement of Single Ion Diffusion Coefficients', *Austr. J. Chem.*, **11**, 1–8 (1958).
27. R. Mills, E.W. Godbole, 'The Precision Measurement of Single Ion Diffusion coefficients', *Austr. J. Chem.*, **12**, 102–103 (1959).
28. R. Mills, A.W. Adamson, 'The Measurement of Self-diffusion in Electrolyte Solutions', *J. Am. Chem. Soc.*, **77**, 3454–3458 (1955).
29. F. Kepak, J. Teply, J. Kriva, 'The Study of the Effect of Gramma Radiation on the Ionic and Colloidal Solutions of ^{144}Ce(III) at Trace Concentration by the Method of Self-diffusion and Centrifugation', *Radiochem. Radioanal. Lett.*, **21**, 365–370 (1975).
30. P. Podhajecky, P. Benes, A. Gosman, 'Radiotracer Analysis of the Physicochemical State of Trace Elements in Aqueous Solutions. I. Capillary Diffusion Method', *J. Radioanal. Chem.*, **51**(2), 315–324 (1979).
31. P. Podhajecky, A. Gosman, P. Benes, 'Radiotracer Analysis of the Physicochemical State of Trace Elements in Aqueous Solutions. IV. Diaphragm Diffusion Method', *J. Radioanal. Chem.*, **57**(2), 253–262 (1980).
32. M. Kersten, U. Foerstner, 'Effect of Sample Pretreatment on the Reliability of Solid Speciation Data of Heavy Metals – Implications for the Study of Early Diagenetic Processes', *Mar. Chem.*, **22**(2–4), 299–312 (1987).
33. P. Benes, M. Kuncova, J. Slovak, P. Ramos, 'Analysis of the Interaction of Radionuclides with Solid Phase in Surface Waters Using Laboratory Model Experiments: Methodical Problems', *J. Radioanal. Nucl. Chem.*, **125**(2), 295–315 (1988).
34. V. Spevackova, J. Kucera, 'Trace Element Speciation in Contaminated Soils Studied by Atomic Absorption Spectrometry and Neutron Activation Analysis', *Int. J. Environ. Anal. Chem.*, **35**(4), 241–251 (1989).
35. R. Chester, F.J. Lin, K.J.T. Murphy, 'A Three Stage Sequential Leaching Scheme for the Characterization of the Sources and Environmental Mobility of Trace Metals in the Marine Aerosol', *Environ. Technol. Lett.*, **10**(10), 887–900 (1989).
36. A. Bermond, G. Sommer, 'Simulation of Heavy Metals Extraction in Soil Samples Compared with Experimental Results', *Environ. Technol. Lett.*, **10**(11), 989–994 (1989).
37. P.M.V. Nirel, F.M.M. Morel, 'Pitfalls of Sequential Extractions', *Water Res.*, **24**(8), 1055–1056 (1990).
38. M. Sager, R. Belocky, R. Pucsko, 'On the Determination of the Speciation of Major and Trace Elements in Sediments by Sequential Solution Techniques', *Acta Hydrochim. Hydrobiol.*, **18**(2), 157–173 (1990).
39. S. Dudka, A. Chlopecka, 'Effect of Solid-phase Speciation on Metal Mobility and Phytoavailability in Sludge-amended Soil', *Water, Air, Soil Pollut.*, **51**(1–2), 153–160 (1990).
40. A.M. Ure, 'Trace Elements in Soil: Their Determination and Speciation', *Fresenius' J. Anal. Chem.*, **337**(5), 577–581 (1990).
41. A.M. Ure, 'Trace Element Speciation in Soils, Soil Extracts, and Solutions', *Mikrochim. Acta*, **2**(1–6), 49–57 (1991).
42. H.A. van der Sloot, 'Neutron Activation Analysis of Trace Elements in Water Samples After Preconcentration on Activated Carbon', Report ECN-1, 1976, Petten, The Netherlands.
43. J.C. Duinker, R.F. Nolting, H.A. van der Sloot, 'The Determination of Suspended Metals in Coastal Waters by Different Sampling and Processing Techniques (Filtration, Centrifugation)', *Neth. J. Sea Res.*, **13**(2), 282–297 (1979).
44. B. Salbu, H.E. Bjoernstad, N.S. Lindstroem, E.M. Brevik, J.P. Rambaek, J.O. Englund, K.F. Meyer, H. Hovind, P.E. Paus, 'Particle Discrimination Effects in the Determination of Trace Metals in Natural Fresh Waters', *Anal. Chim. Acta*, **167**, 161–170 (1985).
45. A. Mizuike, 'Recent Developments in Trace Metal Speciation in Fresh Water', *Pure Appl. Chem.*, **59**(4), 555–564 (1987).
46. J.H.D. Vangenechten, N.A. Chughtai, J. Bierkens, O.L.J. Vanderborght, 'Similarity of Americium-241 and Iron-59 Speciation in Selected Freshwaters and of their Adsorption on Crayfish Exoskeleton', *J. Environ. Radioact.*, **5**(4), 275–286 (1987).
47. B. Salbu, H.E. Bjoernstad, E. Lydersen, A.C. Pappas, 'Determination of Radionuclides Associated with Colloids in Natural Waters', *J. Radioanal. Nucl. Chem.*, **115**(1), 113–123 (1987).
48. B. Salbu, 'Radioactive Tracer Techniques in Speciation Studies', *Environ. Technol. Lett.*, **8**(8), 381–392 (1987).
49. N. Breward, D. Peachy, 'The Development of Portable Equipment to Study Physical and Chemical Phases in Natural Waters', DOE Report RW/888, 102.
50. J. John, B. Salbu, E.T. Gjessing, H.E. Bjoernstad, 'Effect of pH, Humus Concentration, and Molecular Weight on Conditional Stability Constants of Cadmium', *Water Res.*, **22**(11), 1381–1388 (1988).

51. M. Obdrzalek, P. Benes, 'Analysis of the Speciation of Lead in River Waters Using a Radiotracer Method', *Radioisotopy*, **29**(1–2), 202–209 (1988).
52. S.C. Apte, M.J. Gardner, D.T.E. Hunt, 'An Evaluation of Dialysis as a Size-based Separation Method for the Study of Trace Metal Speciation in Natural Waters', *Environ. Technol. Lett.*, **10**(2), 201–212 (1989).
53. A. Mizuike, 'Preconcentration of Trace Elements in Water for Speciation Studies', *J. Chin. Chem. Soc.*, (Taipei), **37**(2), 117–124 (1990).
54. K. Hirose, 'Chemical Speciation of Trace Metals in Seawater: Implication of Particulate Trace Metals', *Mar. Chem.*, **28**(4), 267–274 (1990).
55. ANS Standards Committee, Working Group ANS 16.1. Report ANSI-ANS 16-1-1986.
56. H.A. van der Sloot, 'A Standard Leaching Test for Combustion Residues', Report BEOP-31, Arnhem, The Netherlands, 1984. Available from ECN.
57. IAEA-Tecdoc 626, 'Guidelines for Leaching Studies on Coal Fly Ash and Other Solid Wastes', Vienna, July 1991.
58. H.A. Vos, 'The Speciation of Radionuclides in Sediment and Soils', Report ARL TRO 58, 1981, Yallambie, Victoria 3085.
59. I. Neretnieks, *Diffusivities of Some Dissolved Constituents in Compacted Wet Bentonite Clay*, Royal Institute of Technology, Stockholm, Sweden, 1982.
60. A. van Dalen, J. Wijkstra, 'Distribution and Diffusion Measurements in Deep Ocean Sediment Samples', *Radiochim. Acta*, **38**(3), 141–144 (1985).
61. A. van Dalen, J. Wijkstra, 'Neptunium and Americium Distribution and Diffusion Measurements in Deep-ocean Sediment Samples', *J. Less-Common Met.*, **122**, 563–567 (1986).
62. B. Törstenfelt, 'Migration of the Fission Products Strontium, Technetium, Iodine and Cesium in Clay', *Radiochim. Acta*, **39**(2), 97–104 (1986).
63. P. Vilks, D.J. Drew, 'The Effect of Colloids on Actinide Migration', in *Proceedings of 2nd International Conference of Radioactive Waste Management*, Canadian Nuclear Society, Toronto, 667–673, 1986.
64. P. Benes, 'Speciation Procedures', in *The Environmental Behavior of Radium*, IAEA, Vienna, 279–283, 1990.
65. G. Riise, H.E. Bjoernstad, H.N. Lien, D.H. Oughton, B. Salbu, 'A Study on Radionuclide Association with Soil Components Using a Sequential Extraction Procedure', *J. Radioanal. Nucl. Chem.*, **142**(2), 531–538 (1990).
66. B. Salbu, 'Preconcentration and Fractionation Techniques in the Determination of Trace Elements in Natural Waters, their Concentration and Physico-chemical Form', University of Oslo, Norway, 1984.
67. 'Materials Research Society Symposium Proceedings: Scientific Basis for Nuclear Waste Management', MRS, Pittsburgh, Pennsylvania, 1982.
68. J. Smith, J. Nelissen, R. van Griekeren, 'Comparison of Preconcentration Procedures for Trace Metals in Natural Waters', *Anal. Chim. Acta*, **111**, 215–226 (1979).
69. T.M. Florence, 'The Speciation of Trace Elements in Waters', *Talanta*, **29**, 345–364 (1982).
70. G. Tölg, 'Extreme Trace Analysis of the Elements', *Talanta*, **19**, 1409–1521 (1972).
71. H.A. van der Sloot et al., 'Comparison of Different Regulatory Leaching Test Procedures for Waste Materials', Report ECN-C-91-082. Petten (NH), The Netherlands, 1991.
72. A. Tessier, P.G.C. Campbell, M. Bisson, 'Sequential Extraction Procedure for the Speciation of Particulate Trace Metals', *Anal. Chem.*, **51**(7), 844–851 (1979).
73. V. Spevackova, J. Kucera, 'Trace Element Speciation in Contaminated Soils Studied by Atomic Absorption Spectrometry and Neutron Activation Analysis', *Int. J. Environ. Anal. Chem.*, **35**(4), 241–251 (1989).
74. R. Chester, F.J. Lin, K.J.T. Murphy, 'A Three Stage Sequential Leaching Scheme for the Characterization of the Sources and Environmental Mobility of Trace Metals in the Marine Aerosol', *Environ. Technol. Lett.*, **10**(10), 887–900 (1989).
75. A. Bermond, G. Sommer, 'Simulation of Heavy Metals Extraction in Soil Samples Compared with Experimental Results', *Environ. Technol. Lett.*, **10**(11), 989–994 (1989).
76. P.M.V. Nirel, F.M.M. Morel, 'Pitfalls of Sequential Extractions', *Water. Res.*, **24**(8), 1055–1056 (1990).
77. M. Sager, R. Belocky, R. Pucsko, 'On the Determination of the Speciation of Major and Trace Elements in Sediments by Sequential Solution Techniques', *Acta Hydrochim. Hydrobiol.*, **18**(2), 157–173 (1990).
78. J.M. Martin, P. Nirel, A.J. Thomas, 'Sequential Extraction Techniques, Promises and Problems', *Mar. Chem.*, **22**, 313–341 (1987).
79. M. Kersten, U. Foerstner, 'Effect of Sample Pretreatment on the Reliability of Solid Speciation Data of Heavy Waters', *Mar. Chem.*, **22**, 299–312 (1987).
80. H. Umeki, A. Suzuki, R. Kiyose, 'A Leach Model For Safety Assessment', *Adv. Ceramics*, **20**, 523–529 (1986).
81. D.C. Adriano (ed.), *Biogeochemistry of Trace Metals*, Lewis Publishers, Boca Raton, 1992.
82. J.J.W. Higgo, 'Iodine Speciation and Diffusion in a Sand–Groundwater System', Report EUR 13277 EN, 1991.
83. J. Crank, *The Mathematics of Diffusion*, Clarendon Press, Oxford, 1975.
84. D. Haigh, 'The Groundwater Chemistry of the BGS In Situ Migration Experiment', British Geological Survey, Report WE/08/16.
85. B. Salbu, 'Size Fractionation Techniques Combined with INAA for Speciation Purposes', *J. Radioanal. Nucl. Chem.*, **112**(1), 169–174 (1987).

86. H.A. Das, W.J.M. Hegeman, C.H. van der Weyden, 'Radiotracer Measurement of the Volatalization of Organic Trace Constituents from Water', *Isotopes Environ. Health Stud.*, **32**, 341–346 (1996).
87. C.J.M. Kramey, 'Approaches in Chemical Speciation Studies in Marine Waters', *Elsevier Oceangr. Series*, **43**, 3–36 (1986).
88. J.A. Cox, K. Slonawska, D.K. Gatchell, A.G. Hiebert, 'Metal Speciation by Donnan Dialysis', *Anal. Chem.*, **56**, 650–657 (1984).
89. R.B. Bird, W.E. Stewart, E.N. Lightfoot, *Transport Phenomena*, J. Wiley & Sons, New York, 1960.

Rutherford Backscattering Spectroscopy

Kenji Kimura
Kyoto University, Kyoto, Japan

1 Introduction **1**
2 Fundamentals **2**
2.1 Kinematics of Ion–Atom Collisions 2
2.2 Rutherford Cross-section 3
2.3 Non-Rutherford Cross-sections 4
2.4 Stopping Power 5
2.5 Energy-loss Straggling 6
3 Standard Rutherford Backscattering Spectroscopy **6**
3.1 Apparatus 6
3.2 Thin Film Analysis 7
3.3 Depth Profiling 7
3.4 Computer Simulation 8
3.5 Channeling Effect 8
3.6 Comparison to Related Techniques 9
4 Advanced Techniques **10**
4.1 Heavy Ion Rutherford Backscattering Spectroscopy 10
4.2 High-resolution Rutherford Backscattering Spectroscopy 11
4.3 Microprobe Rutherford Backscattering Spectroscopy 12
4.4 Light Element Analysis by Non-Rutherford Scattering 12
Abbreviations and Acronyms **13**
Related Articles **13**
References **13**

Rutherford backscattering spectroscopy (RBS) is one of the most powerful techniques for measuring elemental depth profiles. It allows quantitative and nondestructive analysis with a reasonable depth resolution. In conventional RBS (combination of 1–4 MeV He ions with a silicon surface-barrier detector), the typical depth resolution is about 10 nm and the typical mass resolution is about 1 u for light elements (mass number $M < 20$) and about 20 u for heavy elements (M ca. 200). Sensitivity of 100 ppm can be easily obtained for heavy elements in light hosts although analysis of light elements in heavy hosts is difficult. In the present article, physical concepts which RBS relies on are briefly introduced. Some examples of analyses are demonstrated and several advanced techniques which have been developed to improve capabilities of the conventional RBS are discussed.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

The analysis of the surface and near-surface region of solids has become an important field of science and technology. RBS is one of the most powerful methods among a number of analyzing techniques. In RBS, high-energy ion beams, usually H or He ions with energies in the range 1–4 MeV, are used as probes. A sample is irradiated by the ion beam. Almost all ions penetrate deep inside the sample up to ca. 10 μm until they entirely lose their kinetic energy. During the penetration some ions collide with the target atoms and are subject to elastic Coulomb scattering (Rutherford scattering) between the projectile and the target nuclei and may be backscattered from the sample. The energy of the ion backscattered from the target atom depends on the target atom mass. This allows compositional analysis of a surface region of several micrometers by measuring the energy spectrum of the backscattered ions. Figure 1 illustrates a schematic set-up of RBS and an example of an energy spectrum for a two-element thin film on a low-mass substrate. The ions scattered from each element form a separated peak. The number of target atoms in the film can be derived from the peak yields and the peak width gives the film thickness. Table 1 summarizes the features of standard RBS and can be used as a rough guide.

The method was first employed by nuclear physicists to analyze their targets in 1951.[1] Extensive application to material science began in the late 1960s. The technique has been refined and now constitutes one of the most common methods for compositional analysis of the surface region. RBS has many outstanding features:

1. RBS is a nondestructive method of analysis.
2. It allows quantitative analysis of all elements simultaneously except for hydrogen.
3. RBS has a depth resolution typically of about 10 nm, which can be improved up to monolayer resolution in the surface region with special equipment.
4. The measurement time is typically as short as several minutes.
5. Sensitivity of RBS is about 100 ppm for heavy elements although it is not so good for light elements.

RBS relies on the following physical concepts:

1. The kinematic factor of the elastic scattering which describes the reduction of incident energy in a collision between the probe ion and the target atom.

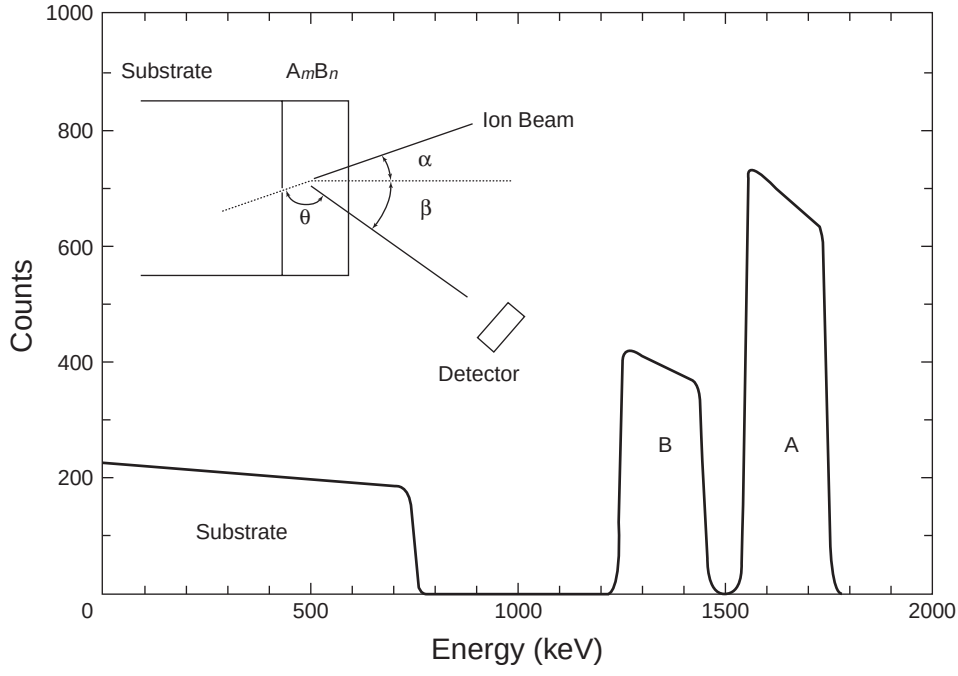

Figure 1 RBS spectrum for a two-element compound film (A_mB_n) on a low-mass substrate. The inset shows the schematic set-up of RBS.

Table 1 Important parameters and capabilities of standard RBS

Probe ion	H, He
Ion energy	1–4 MeV
Beam current	1–100 nA
Detector	SSB detector
Depth resolution	ca. 10 nm (easily improved to 2 nm using a grazing-angle technique)
Range	ca. 10 μm
Element range	>1 for H ion >4 for He ion
Mass resolution	1 u for light elements ($M_2 < 40$ u) with 2 MeV He ion 20 u for heavy elements (M_2 ca. 200 u) with 2 MeV He ion
Sensitivity	ca. 100 ppm for heavy elements in light hosts

SSB, silicon surface barrier.

The resulting energy of the scattered ion increases with target atom mass. This allows identification of the target atom by measuring the scattered ion energy.

2. The differential scattering cross-section which gives the probability of scattering. This allows basic quantitative analysis without a standard sample.
3. The stopping power which is defined by the energy loss of the ion per unit path length inside the target. The energy of the backscattered ion depends on the depth from which the ion was scattered because the path length is proportional to the depth. This allows the depth profiling of elements in the target.
4. The energy-loss straggling which is the fluctuation of the energy loss arising from the statistical feature of the energy-loss process. This determines the intrinsic depth resolution.

There are numerous reviews on RBS[2–4] and for the state of the art reader can refer to relevant conference proceedings such as ion beam analysis (IBA), Conference on the Application of Accelerators in Research and Industry (CAARI), European Conference on Accelerator Applications in Research and Technology (ECAART) and Nuclear Microprobe Technology and Applications (NMTA). Most have been published in *Nuclear Instruments and Methods* in the Physics Research section B.[5]

2 FUNDAMENTALS

2.1 Kinematics of Ion–Atom Collisions

When an ion is scattered from a target atom at an angle θ, the ratio of the scattered-ion energy E to the incident energy E_0 can be calculated using the laws of conservation of energy and momentum, Equation (1)

$$K = \frac{E}{E_0} = \frac{(M_2^2 - M_1^2 \sin^2\theta)^{1/2} + M_1 \cos\theta}{M_1 + M_2}^2 \qquad (1)$$

where M_1 and M_2 are masses of the incident ion and the target atom, respectively. The binding energy of the target atom and its thermal vibration are neglected in the derivation of Equation (1). These effects are negligibly small in collisions between ions with megaelectron volt energy and target atoms. It is clear that the energy ratio K, called the kinematic factor, is determined by the mass ratio M_2/M_1 and the scattering angle θ. The basic idea of RBS relies on this equation. Figure 2 shows the kinematic factor as a function of the mass ratio M_2/M_1. The kinematic factor, and thus the energy of the scattered ion, increases monotonically with the mass ratio indicating that the target atom mass can be determined from the observed energy of the scattered ion.

The mass resolution is defined by Equation (2)

$$\delta M_2 = \frac{\delta E}{E_0} \left(\frac{\mathrm{d}K}{\mathrm{d}M_2} \right)^{-1} \tag{2}$$

where δE is the overall energy resolution of the RBS system. The dominant contribution to δE comes from the energy resolution of a SSB detector which is commonly used in RBS. Figure 3 displays the calculated mass resolution for 2-MeV H, He, and C ions, where the typical energy resolutions of the SSB detector (i.e. $\delta E = 10$, 15, and 50 keV for H, He, and C ions, respectively) are used for δE. The results at different scattering angles are also shown for 2-MeV He ions. A mass difference of 1 u can be resolved for light elements ($M_2 < 40$ u) with 2-MeV He ions. The resolution, however, becomes worse very rapidly with increasing M_2 and reaches δM_2 ca. 20 u at M_2 ca. 200 u.

For better mass resolution, heavy ions are usually employed. Although the energy resolution of the SSB detector is worse for heavy ions, the resulting mass resolution for C ions is better than for He ions and $\delta M_2 =$

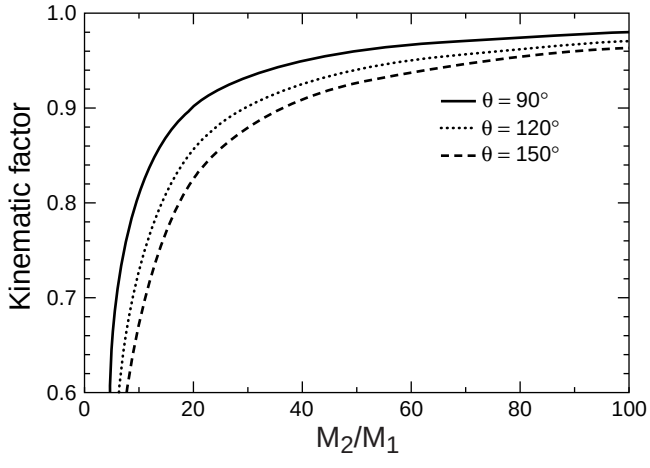

Figure 2 Kinematic factor K as a function of the ratio of the target mass to the ion mass M_2/M_1.

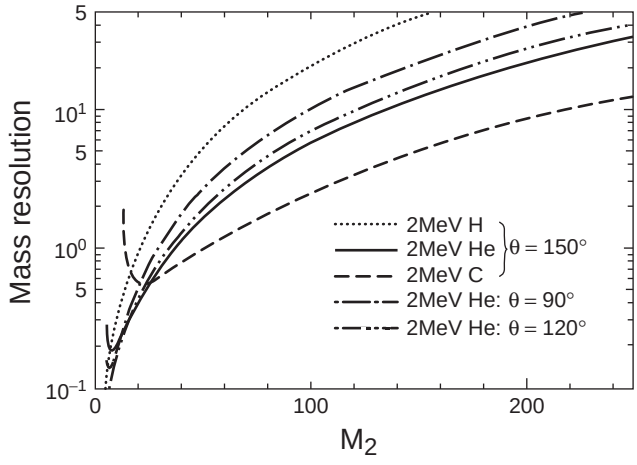

Figure 3 Mass resolution for 2-MeV H, He and C ion beams as a function of target mass M_2. Typical energy resolution of the SSB detector is used in the calculation.

1 u is achieved up to M_2 ca. 50 u with 2-MeV C ions. The resolution can be improved considerably using the time-of-flight (TOF) technique for energy analysis instead of the SSB detector in heavy ion RBS (HIRBS). The details of the heavy ion RBS will be discussed in section 4.1.

2.2 Rutherford Cross-section

The scattering cross-section σ is an effective area associated with the colliding particles. When projectile ions impinge on a thin target film of thickness t, the scattering yield is given by Equation (3)

$$Y_{\mathrm{tot}} = \frac{Nt\sigma Q}{\cos\alpha} \tag{3}$$

where N is the volume density of the target atom, Q the number of incident ions, and α is the incident angle (the definition of α is shown in Figure 1). In RBS, the ions scattered at a particular scattering angle θ are measured by a detector with a finite acceptance solid angle $\Delta\Omega$. The number of these ions can be calculated (Equation 4) with an angular differential cross-section $\mathrm{d}\sigma/\mathrm{d}\Omega$

$$Y(\theta) = Nt \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \frac{\Delta\Omega Q}{\cos\alpha} \tag{4}$$

The differential cross-section for Coulomb scattering (Rutherford cross-section) is given by an analytical formula (Rutherford formula), Equation (5)

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \right)_{\mathrm{R}} = \left(\frac{Z_1 Z_2 e^2}{4E} \right)^2 \times \frac{4 \left\{ \sqrt{(M_2^2 - M_1^2 \sin^2\theta)} + M_2 \cos\theta \right\}^2}{M_2 \sin^4\theta \sqrt{M_2^2 - M_1^2 \sin^2\theta}} \tag{5}$$

where Z_1, Z_2 are atomic numbers of incident ion and target atom, respectively and E is the incident energy. This simple formula guarantees to extract quantitative data about abundances of elements from RBS spectra. The cross-section is proportional to Z_2^2, indicating high sensitivity for heavy elements in RBS. A useful number in evaluating the Rutherford cross-section is $e^2 = 14.4\,\text{eV Å}$ ($1\,\text{Å} = 10^{-10}\,\text{m}$). For example, 1 MeV He^+ ions of total dose 1 μC are incident on a Si film of $t = 1000\,\text{Å}$ at $\alpha = 0°$. The differential cross-section is calculated to be $(d\sigma/d\Omega)_R = 1.126 \times 10^{-8}\,\text{Å}^2/sr$ at $\theta = 150°$. Using a typical acceptance angle of SSB detectors, $\Delta\Omega = 5\,\text{msr}$ the yield is given by Equation (6)

$$Y = 4.994 \times 10^{-2}\,\text{Å}^{-3} \times 1000\,\text{Å} \times 1.126 \times 10^{-8}\,\text{Å}^2/sr^{-1} \times 5 \times 10^{-3} sr \times \frac{1 \times 10^{-6}}{1.602 \times 10^{-19}} \approx 17\,500 \quad (6)$$

The associated statistical error is $1/\sqrt{17\,500} = 0.8\%$. This means the thickness of the film can be determined with an accuracy less than 1% in this case.

2.3 Non-Rutherford Cross-sections

The Rutherford cross-section can be used in situations where the incident ion energy is in the order of $1\,\text{MeV u}^{-1}$. The cross-section deviates from the Rutherford formula in both high-energy and low-energy regions. On the low-energy side, the screening effect due to the bound electrons in the target atom reduces the cross-section from the Rutherford formula. The cross-section cannot be given by a simple analytical formula like Equation (5). The screened Coulomb potential is given by Equation (7)

$$V(r) = \frac{Z_1 Z_2 e^2}{r}\phi \quad \frac{r}{a} \quad (7)$$

where ϕ is the screening function and a is the screening length. There are several analytical formulas for the screening function given by a sum of exponential terms, Equation (8)

$$\Phi(z) = \sum_i^n \alpha_i \, e^{-\beta_i z} \quad (8)$$

where $\{\alpha_i\} = \{0.1, 0.55, 0.35\}$, $\{\beta_i\} = \{6.0, 1.2, 0.3\}$ for Molière's screening function[6] ($a = 0.8853 a_B (Z_1^{1/2} + Z_2^{1/2})^{-2/3}$ is commonly used) and $\{\alpha_i\} = \{0.181\,75, 0.509\,86, 0.280\,22, 0.028\,171\}$, $\{\beta_i\} = \{3.1998, 0.942\,29, 0.402\,90, 0.201\,62\}$ for Ziegler–Biersack–Littmark's screening function ($a = 0.8853 a_B (Z_1^{0.23} + Z_2^{0.23})^{-1}$ is used).[7]

Using the screening function, the differential cross-section can be calculated numerically. Algorithms for rapid computations are proposed by Mendenhall and Weller.[8] A useful analytical formula for the screening correction valid for large scattering angles is given by L'Ecuyer et al.,[9] Equation (9)

$$\frac{d\sigma/d\Omega}{(d\sigma/d\Omega)_R} = 1 - \frac{0.049 Z_1 Z_2^{4/3}}{E_{CM}} \quad (9)$$

where E_{CM} is the center-of-mass kinetic energy in kilo-electron volts. Examples of the results calculated with Equation (9) together with the result of exact numerical calculation are shown in Figure 4. The deviation from the Rutherford cross-section is large for heavy elements and the correction formula for Equation (9) is reasonably good for both heavy and light elements. Although the deviation for 1-MeV He–Au scattering is about 4% and the Rutherford formula may not be used in a precise measurement, Equation (9) can be safely used even in this situation. It must, however, be noted that Equation (9) does not contain the scattering angle θ and is valid only for large θ. At smaller scattering angles, the screening effect becomes large.

On the high-energy side, the incident ion may approach to the target nuclei within the range of the nuclear force. This causes deviation from the Rutherford cross-section. The deviation is large for light elements because the Coulomb barrier of the nucleus is low for light elements. An empirical formula to estimate the threshold energy, above which the deviation from the Rutherford cross-section for $160° < \theta < 180°$ is larger than 4%, is given by Leavitt and McIntyre,[10] Equations (10–12)

$$E_{th} = (0.12 \pm 0.01)Z_2 - (0.5 \pm 0.1)\,\text{MeV for H} \quad (10)$$

$$E_{th} = (0.25 \pm 0.01)Z_2 + (0.4 \pm 0.2)\,\text{MeV for He} \quad (11)$$

$$E_{th} = (0.330 \pm 0.005)Z_2 + (1.4 \pm 0.1)\,\text{MeV for Li} \quad (12)$$

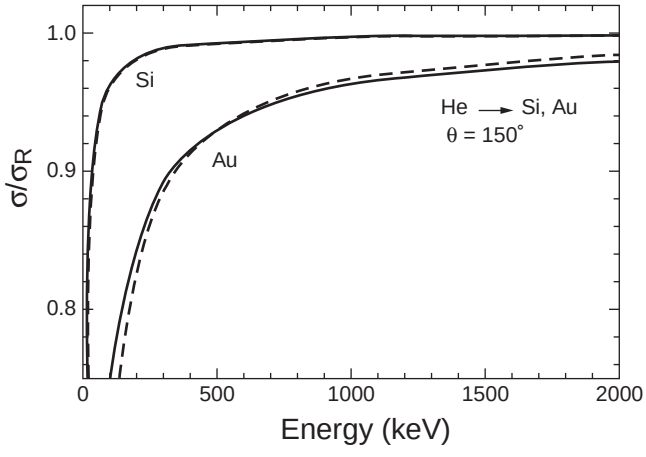

Figure 4 Screening correction for the scattering cross-section as a function of the incident ion energy. The results of the numerical calculation (solid lines) as well as those of the analytical formula (dashed lines) are shown for He–Si and He–Au scattering.

In contrast to the correction for screening effect, no practical methods exist for rapid accurate estimation of the nuclear force correction. The cross-section depends on both the ion energy and the scattering angle in a complicated manner (e.g. Figure 14). Nevertheless, high-energy ions are useful in some cases, especially in analysis of light elements. The standard RBS is less sensitive for light elements due to the small Rutherford cross-sections of light elements (the cross-section is proportional to the square of the target atomic number). The non-Rutherford cross-sections for light elements are frequently enhanced by one or two orders of magnitude while those for heavy elements may remain as a Rutherford cross-section. This improves the sensitivity for light elements. The details will be discussed in section 4.4.

2.4 Stopping Power

An energetic ion passing through a solid loses its energy primarily through excitation of electrons. The rate of the energy loss per unit path length, $\mathrm{d}E/\mathrm{d}x$, is called stopping power and understanding the stopping power has been a subject of extensive studies. Using the first-order perturbation theory, the stopping power is given by[11] Equation (13)

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{4\pi Z_1^2 Z_2 N e^4}{mV^2} \ln\left(\frac{mV^2}{I}\right. \tag{13}$$

where m is the electron mass, V the velocity of the ion, and I ($\approx 10 \times Z_2$ eV) is the mean excitation energy of the target electrons. This equation, called the Bethe–Bloch formula, is used as a guide in semi-empirical fitting of stopping power data to provide semi-empirical formulas. A number of tables are available to calculate stopping powers reasonably accurately.[7,12–14] Figure 5 shows examples of the stopping powers of Si and Au for H and He ions calculated with the tabulated values given by Andersen and Ziegler.[12,13] The stopping power has a broad maximum around 1 MeV for He ions and around 100 keV for H ions. The maximum energy and the maximum stopping power increases with both projectile and target atomic number. The stopping power for megaelectron-volt He ions is several times larger than that for megaelectron-volt H ions. This indicates that better depth resolution is available for He ions.

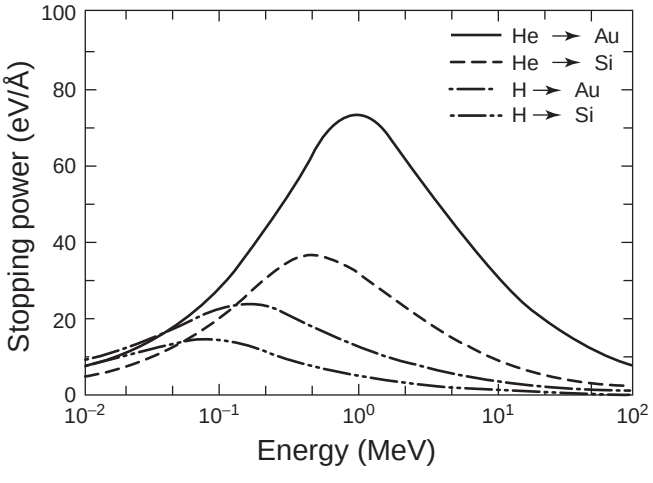

Figure 5 Stopping powers of Si and Au for H and He ions as a function of the ion energy.

In order to estimate the stopping power of a compound, it is convenient to introduce stopping cross-section defined by Equation (14)

$$\varepsilon = \frac{1}{N}\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right. \tag{14}$$

The stopping power of a compound $\mathrm{A}_m\mathrm{B}_n$ can be calculated using Bragg's rule, which is based on the simple assumption that the constituent elements act independently in the energy loss process, Equation (15)

$$\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right._{\mathrm{A}_m\mathrm{B}_n} = N_{\mathrm{mol}}[m\varepsilon(\mathrm{A}) + n\varepsilon(\mathrm{B})] \tag{15}$$

where $\varepsilon(\mathrm{A})$ and $\varepsilon(\mathrm{B})$ are the stopping cross-sections of the constituents and N_{mol} is the molecular density. This neglects the effect of chemical bonding but gives reasonably accurate values for megaelectron-volt light ions.

The relationship between the energy of the scattered ion and the depth at which the ion is scattered can be calculated using the stopping power, the kinematic factor, and the geometry of the scattering. Neglecting the variation of the stopping power along the ion path, the energy of the ion scattered from a target atom at a depth t is given by Equation (16)

$$E(t) = K \quad E_0 - \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right._{\mathrm{in}} \frac{t}{\cos\alpha} \quad - \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right._{\mathrm{out}} \frac{t}{\cos\beta} \tag{16}$$

where the subscripts in and out refer to the inward path and outward path, respectively, and α and β are the incident and exit angles, respectively. For a thin surface layer, the variation of the stopping power is small and the use of an average value is a good approximation. The variation in energy δE corresponds to a variation in depth δt through Equation (17)

$$\delta E = -[S]\delta t \tag{17}$$

where $[S]$ is commonly called the energy-loss factor and is given by Equation (18)

$$[S] = \quad K\frac{(\mathrm{d}E/\mathrm{d}x)_{\mathrm{in}}}{\cos\alpha} + \frac{(\mathrm{d}E/\mathrm{d}x)_{\mathrm{out}}}{\cos\beta} \tag{18}$$

The depth resolution depends on the overall energy resolution, the stopping power, and the geometry. The energy resolution is normally composed of three contributions: the energy spread of the incident beam δE_{in}, the detector resolution δE_d, and the energy loss straggling δE_{s} (the last of these, δE_{s} will be discussed in detail in the next section). Assuming these contributions are independent, the overall resolution is given by Equation (19)

$$(\delta E)^2 = (\delta E_{\text{in}})^2 + (\delta E_{\text{d}})^2 + (\delta E_{\text{s}})^2 \tag{19}$$

For example, in the case of 2-MeV He on Au at $\alpha = 0°$ and $\beta = 30°$ ($\theta = 150°$), the energy-loss factor is calculated to be 135 eVÅ^{-1}. Using a typical energy resolution in the standard RBS, $\delta E = 15$ keV and a depth resolution of 11 nm is obtained. The depth resolution can be improved up to ca. 2 nm simply by using a grazing-angle technique, usually with a grazing exit angle, for example β ca. 85°. In the grazing-angle geometry, the acceptance angle of the SSB detector should be much smaller than the grazing angle. Otherwise the distribution of β and with it the distribution of $[S]$ makes the depth resolution worse.

2.5 Energy-loss Straggling

The energy-loss process described in the previous section has a stochastic feature. The slowing down of the ion is accompanied by energy spreading which is called energy-loss straggling. For long path length, the energy distribution approaches a Gaussian distribution. The standard deviation Ω of the distribution for the high-energy limit was given by Bohr,[15] Equation (20)

$$\Omega_{\text{B}}^2 = 4\pi Z_1^2 Z_2 e^4 N t \tag{20}$$

and has been extended to lower energies by many authors. A useful analytical formula was given by Lindhard and Scharff,[16] Equation (21)

$$\frac{\Omega^2}{\Omega_{\text{B}}^2} = L(\chi) = \begin{cases} 0.68\chi^{1/2} - 0.08\chi^{3/2} & \chi \leq 3 \\ 1 & \chi \geq 3 \end{cases} \tag{21}$$

where $\chi = (V/v_{\text{B}})^2/Z_2$, V is the ion velocity, and v_{B} ($= 2.2 \times 10^8$ cm s^{-1}) the Bohr velocity. More refined models were developed by Bonderup and Hvelplund,[17] Chu,[18] and Besenbacher et al.[19]

For compounds, an additive rule similar to Bragg's rule was proposed,[18] Equation (22)

$$\frac{\Omega^2(\text{A}_m\text{B}_n)}{N_{\text{mol}} t} = \frac{m\Omega^2(\text{A})}{N_{\text{A}} t} + \frac{n\Omega^2(\text{B})}{N_{\text{B}} t} \tag{22}$$

The contribution of the energy-loss straggling to the energy resolution is given by Equation (23)

$$(\delta E_{\text{s}})^2 = 8 \ln 2 \left(K^2 \Omega_{\text{in}}^2 + \Omega_{\text{out}}^2\right) \tag{23}$$

where $8 \ln 2$ is a conversion factor from the standard deviation to the full width at half maximum, and Ω_{in} and Ω_{out} are the straggling in inward and outward paths. For megaelectron-volt H and He ions incident on layers $\leq$100 nm, the straggling is small (e.g. Ω_{B} ca. 2.7 keV for He–Si at $t = 100$ nm) compared to the resolution of the SSB detector (δE_d ca. 15 keV for megaelectron-volt He) and can be neglected in the obtainable depth resolution in the standard RBS. The contribution of the energy-loss straggling, however, increases with $t^{1/2}$ and the straggling sets a fundamental limit to the depth resolution at deeper layers.

3 STANDARD RUTHERFORD BACKSCATTERING SPECTROSCOPY

3.1 Apparatus

In the standard RBS, 1–4-MeV He ion beams and a SSB detector are commonly used. Although both the mass and depth resolutions are poorer for H ions than for He ions, H ions are sometimes used in RBS in order to analyze deeper regions. Ions generated by an ion source are accelerated by an electric field. The most widely used devices for the production of a megaelectron-volt ion beam are the Van de Graaff-type and Cockcroft- and Walton-type accelerators. The accelerated ions are analyzed by a magnet to eliminate contaminant ion species and to select the ion energy. The calibration of the analyzing magnet is usually done using resonant nuclear reactions. After passing the analyzing magnet, the ion beam is collimated by apertures to a size of ca. 1 mm and introduced to a scattering chamber where targets are mounted on a manipulator. The beam current is usually monitored by target current or a beam monitor system installed between the final aperture and the target. The typical beam current is of the order of 10 nA. For a channeling measurement, the beam divergence angle and the precision of the manipulator should be less than 1 mrad.

A SSB detector is installed inside the chamber to measure the energy spectrum of scattered ions. The SSB detector is basically a Schottky barrier diode. A reverse bias is applied to make a depletion region of the order of 100 μm which is greater than the range of the mega-electron-volt light ions. The electron-hole pairs created by the ion in the depletion region are swept by the high electric field and collected. The quantity of charge collected bears a linear relationship to the total energy spent to create electron–hole pairs (about 3.6–3.7 eV per electron–hole pair). The linearity between the collected charge and the ion energy is quite good for megaelectron-volt light ions. The signal from the detector is amplified

by a preamplifier and a linear amplifier and sent to a pulse-height analyzer (PHA). The PHA stores pulses of a given height in a given channel. The relation between the channel number and the ion energy can be calibrated by measuring ions of known energy, for example α particles from ^{241}Am. It should be noted that the energy lost in the entrance window as well as the energy lost to nuclear scattering do not contribute to the pulse height. The equivalent thickness of the window is about 100 nm, which corresponds to about 30 keV loss for 1-MeV He. These effects are serious for heavy ions and must be taken into account in the conversion procedure from pulse height to ion energy, especially in the low-energy region.

3.2 Thin Film Analysis

An example of RBS spectrum for a high-T_c superconductor thin film deposited on a Si substrate is depicted in Figure 6.[20] The film was prepared by ion beam sputtering of $Bi_2Sr_2CaCu_2O_7$. The peaks for Bi, Sr, Cu, and Ca are well separated and the oxygen peak can be clearly seen on top of the Si signal. The integrated counts Y_i for each element can be accurately determined. The areal density of element i is obtained from Equation (4) as

$$n_i = N_i t = \frac{Y_i \cos\alpha}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_i \Delta\Omega Q} \tag{24}$$

Although precise measurements of the ion energy E, the scattering angle θ, the detector acceptance angle $\Delta\Omega$, and the incident ion number Q are necessary to estimate the absolute areal density n_i, an average stoichiometric ratio can be obtained without these values, Equation (25)

$$\frac{N_i}{N_j} = \frac{Y_i}{Y_j}\frac{(\mathrm{d}\sigma/\mathrm{d}\Omega)_j}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_i} \approx \frac{Y_i}{Y_j}\left(\frac{Z_j}{Z_i}\right)^2 \tag{25}$$

The composition of the film shown in Figure 6 can be determined to be $Bi_1Sr_{1.6}Ca_{0.9}Cu_{1.4}O_7$ using this equation.

In order to estimate the absolute areal density n_i using Equation (24), a standard sample implanted with a heavy element of a known areal density N_{ref} is sometimes used, Equation (26)

$$n_i = N_i t = \frac{Y_i(\mathrm{d}\sigma/\mathrm{d}\Omega)_{\mathrm{ref}}}{Y_{\mathrm{ref}}(\mathrm{d}\sigma/\mathrm{d}\Omega)_i} N_{\mathrm{ref}} \approx \frac{Y_i}{Y_{\mathrm{ref}}}\left(\frac{Z_{\mathrm{ref}}}{Z_i}\right)^2 N_{\mathrm{ref}} \tag{26}$$

where Y_{ref} is the yield of the standard sample measured under the same conditions. An alternative conventional method to estimate the absolute areal density is to make use of the signal of substrate material, Equation (27)

$$n_i = N_i t = \frac{Y_i}{y_s}\frac{\delta E_{\mathrm{ch}}}{[S]_S} N_s \frac{(\mathrm{d}\sigma/\mathrm{d}\Omega)_s}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_i} \approx \frac{Y_i}{y_s}\frac{\delta E_{\mathrm{ch}}}{[S]_S} N_s \left(\frac{Z_s}{Z_i}\right)^2 \tag{27}$$

where y_s is the height of the energy spectrum at the leading edge of the substrate, δE_{ch} the corresponding energy for one channel, N_s the atomic density of the substrate, $(\mathrm{d}\sigma/\mathrm{d}\Omega)_s$ the differential cross-section of the substrate atom and $[S]_S$ the energy-loss factor for the substrate.

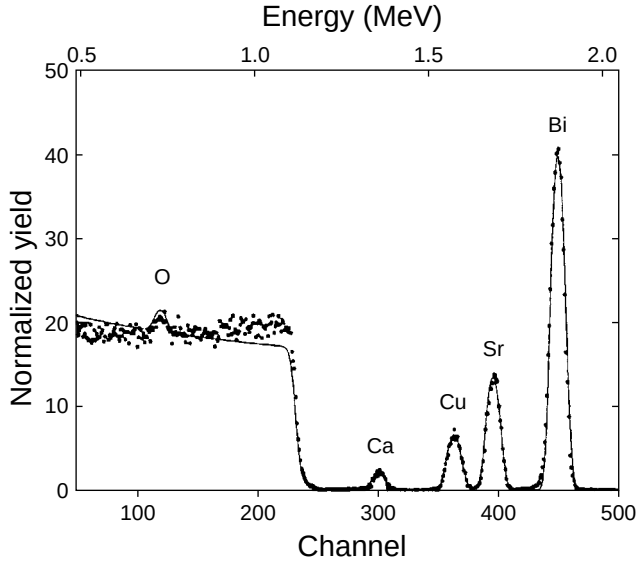

Figure 6 RBS spectrum of a high-T_c superconductor thin film $Bi_1Sr_{1.6}Ca_{0.9}Cu_{1.4}O_7$ deposited on a Si substrate. Signals of Bi, Sr, Cu, and Ca are seen as separated peaks. (Reproduced by permission of Elsevier Science B.V. from Climent-Font et al.[20])

3.3 Depth Profiling

When the stoichiometry changes with depth, the variation of the composition can be determined from the RBS spectrum. The stoichiometry at a depth t can be derived from the corresponding yields of all compositional elements at t, Equation (28). In Equation (28) F_i is the atomic fraction of element i at a depth t.

$$F_i = \frac{y_i[S]_i \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_i}{\sum_j y_j[S]_j \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_j} \approx \frac{y_i[S]_i / Z_i^2}{\sum_j y_j[S]_j / Z_j^2} \tag{28}$$

This procedure is not simple because the energy-loss factor $[S]_i$ is a function of the composition. For dilute concentrations of an impurity, the stopping power is simply determined by the matrix and the depth profiling

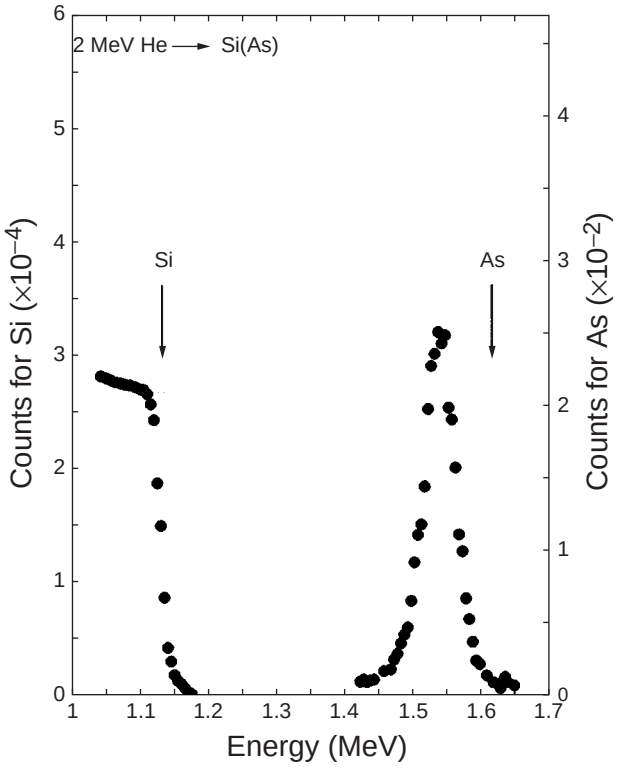

Figure 7 RBS spectrum of 2-MeV He ions backscattered from Si implanted with 1.2×10^{15} As ions cm^{-2} at 200 keV. The arrows indicate the energy of ions scattered from surface Si and As atoms. (Reproduced by permission of Plenum Press from Mayer et al.[2])

is straightforward. Figure 7 shows an example of such an RBS spectrum for Si implanted with 1.2×10^{15} As ions cm^{-2} at 200 keV measured with 2-MeV He ions.[2] There is a separated peak of As at ca. 1.55 MeV which corresponds to t ca. 150 nm. The width of the As peak at ca. 60 keV is much larger than the energy resolution of the SSB detector showing that the observed As peak profile reflects the actual As distribution in Si. The energy-loss factors for Si and As are calculated to be 44.3 and 46.0 eV Å^{-1} respectively. Using the peak yield of As, $y_{As} = 250$ and the Si yield at the corresponding depth, $y_{Si} = 28\,000$, the peak concentration can be calculated from Equation (28) as Equation (29)

$$F_{As} \approx \frac{250 \times 46/33^2}{(250 \times 46/33^2 + 28\,000 \times 44.3/14^2)} = 0.0017 \tag{29}$$

Note that the sensitivity of RBS can easily reach 100 ppm level for heavy elements.

The RBS spectrum shown in Figure 7 also provides a good example of how to estimate areal density of the implanted atoms using Equation (27). Using the total yield of the As peak $Y_{As} = 3350$ and $\delta E_{ch} = 5$ keV, the areal density of As is given by Equation (30)

$$n_{As} \approx \frac{3350}{28\,000} \frac{5\,\text{keV}}{44.3\,\text{eV Å}^{-1}} 4.994 \times 10^{-2}\,\text{Å}^{-3} \left(\frac{14}{33}\right)^2 = 0.12\,\text{Å}^{-2} \tag{30}$$

The value obtained agrees well with the number of implanted As ions measured in the implantation procedure. This good agreement clearly shows the capability for quantitative analysis in RBS.

3.4 Computer Simulation

When the signals from different elements overlap in a spectrum, derivation of the target composition as a function of depth is not straightforward although a basic derivation can be obtained using Equation (28). A powerful approach to analyzing RBS spectra is by computer simulation. There are a number of computer simulation programs to analyze the RBS spectrum such as RUMP,[21] GISA,[22] and SCATT/HYPRA.[23] In these programs, the target composition is changed until the calculated spectrum matches the measured spectrum. As with other inverse problems, this procedure may obtain a local minimum. Recently, a simulated annealing algorithm was successfully applied to the analysis of RBS spectra.[24] It tends asymptotically to the absolute minimum rather than to a local minimum with any initial compositions. This allows fully automatic analysis without time-consuming human intervention. Figure 8 shows examples of the different steps of the fit to a RBS spectrum of a multilayered iron–cobalt silicide generated by the simulated annealing algorithm. The final result agrees reasonably well with the original spectrum with a simple initial composition (Figure 8a) generated automatically by the program.

3.5 Channeling Effect

The energy spectrum from a single crystal aligned with a low-index axis or plane parallel to the incident beam direction is very different from those discussed before. The ion experiences a series of correlated small angle collisions with crystal atoms and cannot approach to the atomic row or plane over a certain distance. This phenomenon, called channeling, drastically reduces the yield of Rutherford scattering. The analyzing procedures described in the previous sections cannot be applied to the RBS spectra measured under channeling conditions. In a simple RBS measurement, a single crystal target is usually rotated during measurement to avoid channeling effects.

Although the channeling is undesirable for RBS, it provides a unique method for studying lattice defects.

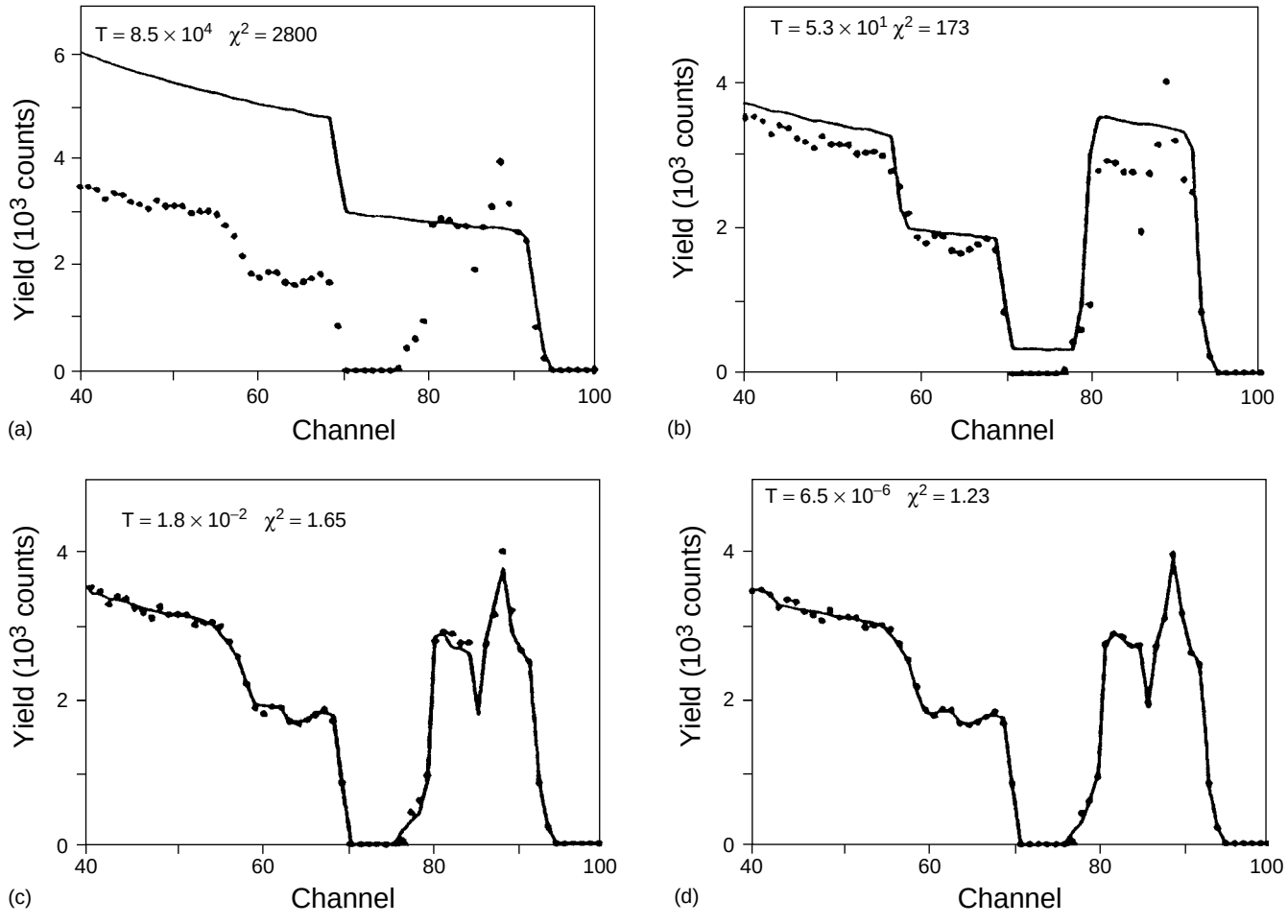

Figure 8 Differential steps of the fit from (a) the initial guess to (d) the final results obtained in simulated annealing analysis of a RBS spectrum of a multilayered iron–cobalt silicide. (Reproduced by permission of American Institute of Physics from Barradas et al.[24])

Channeling ions might stop channeling upon scattering by lattice defects (dechanneling) or even be directly backscattered from the defects (direct scattering). These effects depend on the structure of the defect. Impurity atoms, for example, may cause direct scattering depending on both the location of the impurity atom and the channeling axis as shown in Figure 9.[25] The substituted atoms cannot be seen by channeling ions but the interstitial atoms may be seen by the ions channeling through particular channels. Thus the site of the impurity atom can be roughly determined by channeling measurements at several major axes. For more detailed information, the lattice site of the interstitial atom can be determined within an accuracy of 0.01 nm by measuring the incident-angle dependence of the backscattering yield from the impurity atoms around the channeling direction.

Other defects also cause dechanneling and the energy dependence of the dechanneling probability is different in each case, that is $\propto \sqrt{E}$ for dislocation, $\propto E^0$ for stacking fault, and $\propto E^{-1}$ for point defect.[26] Thus the type of major defect can be determined by measuring the energy dependence of the dechanneling probability. Other useful applications of channeling are described in the literature.[27–29]

3.6 Comparison to Related Techniques

The low-energy version of RBS is called ion scattering spectroscopy (ISS).[3] Rare gas ions with energy of several kiloelectron volts and electrostatic analyzers are commonly used. ISS is sensitive to surface atoms because of the large neutralization probability of ions scattered from subsurface atoms and that is used extensively for compositional and structure analysis of surfaces. In order to study subsurface regions, the TOF technique is employed to measure energy spectra of both ions and neutral atoms. An alternative method is the use of alkali

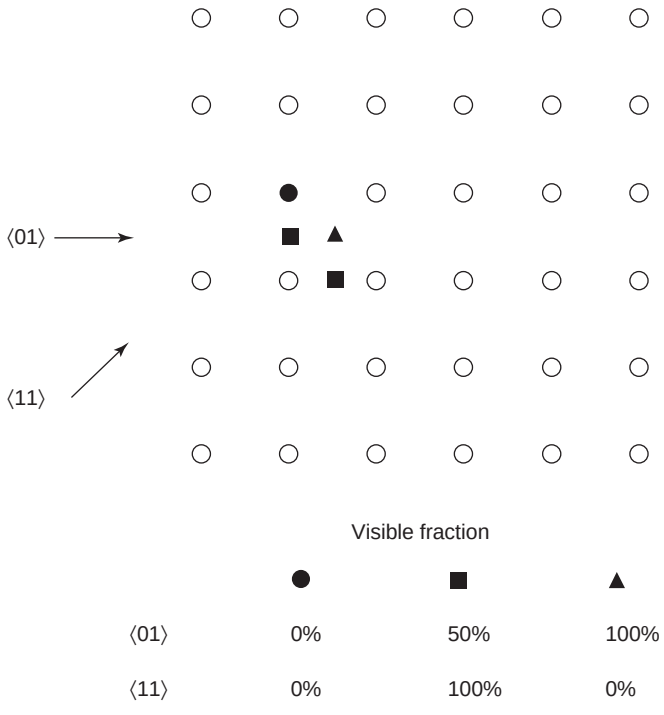

Figure 9 Two-dimensional model illustrating how the channeling effect can be used to locate foreign atoms in a crystal.

metal ions for which the neutralization probability is small even for the ions scattered from the subsurface region.[30]

The sensitivity of RBS is not sufficient for determination of trace elements. Particle-induced X-ray emission (PIXE) provides quantitative analysis of trace elements present at parts per million levels.[31] Also see the article **PIXE (Particle-induced X-ray Emission)** in this publication. H ions of 1–3 MeV are used to produce inner-shell vacancies in target atoms. Characteristic X-rays emitted via refilling the inner-shell vacancies by outer-shell electrons are detected by an X-ray detector. Compared with the similar technique of electron probe microanalysis (EPMA), the background due to the brehmsstrahlung is many orders of magnitude smaller and a sensitivity of better than 0.1 ppm is easily obtained for many elements in low-mass materials. Although the sensitivity of PIXE is excellent, PIXE does not have depth resolution.

Elastic Recoil Detection (ERD) is a similar technique to RBS.[32] Also see the article **Elastic Recoil Detection Analysis** in this publication. In ERD, recoiled ions are detected instead of the scattered ions. While RBS is suitable for analysis of heavy elements, ERD provides quantitative depth profiles of light elements, especially hydrogen. Before the first introduction of ERD,[33] nuclear reaction analysis (NRA) was predominantly used to measure depth profiles of light elements.[34] Also see the article **Nuclear Reaction Analysis** in this publication. Nowadays, ERD is commonly used because of its simplicity and ease of use although the standard ERD has poorer sensitivity and depth resolution than NRA.

4 ADVANCED TECHNIQUES

4.1 Heavy Ion Rutherford Backscattering Spectroscopy

As was discussed in section 2.1, use of heavy ions improves mass resolution. Heavy ions also have an advantage of better depth resolution than light ions because the stopping power is proportional to the square of the ion charge. Figure 10 shows an example of heavy ion RBS (HIRBS) spectrum of a thin $Y_1Ba_2Cu_3O_{7-x}$ film using 22-MeV ^{12}C ions and a SSB detector.[35] Although the energy resolution of the SSB detector is degraded for C ions, the separation of each element is perfect and, moreover, two copper isotopes, ^{63}Cu and ^{65}Cu, are clearly resolved, which is usually impossible in the standard RBS with megaelectron-volt He ions.

If the TOF technique is employed to measure the energy spectrum instead of the SSB detector, the energy resolution and thus both the mass and depth resolutions are improved considerably (typical energy resolution of TOF detectors is better than 1%). Figure 11 depicts an example of a HIRBS spectrum (25 MeV ^{35}Cl beam) of a 48 nm layer of GaAs capped by a 26 nm Pd layer on Si substrate measured with a TOF detector.[36] The two Ga isotopes as well as several Pd isotopes are well-resolved.

Although the excellent mass resolution is clear in HIRBS, there are several disadvantages of HIRBS: (1) The cross-section deviates from the Rutherford formula particularly for heavy target elements. (2) Light

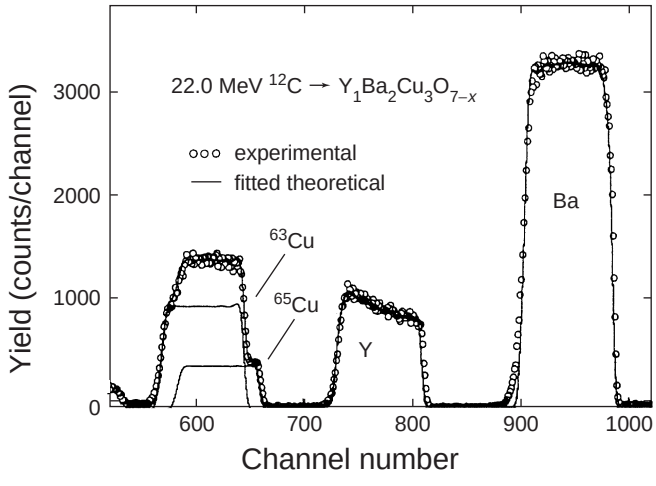

Figure 10 Heavy ion RBS spectrum of the $Y_1Ba_2Cu_3O_{7-x}$ high-T_c superconductor film measured with 22 MeV C ions. (Reproduced by permission of Elsevier Science B.V. from Rauhara et al.[35])

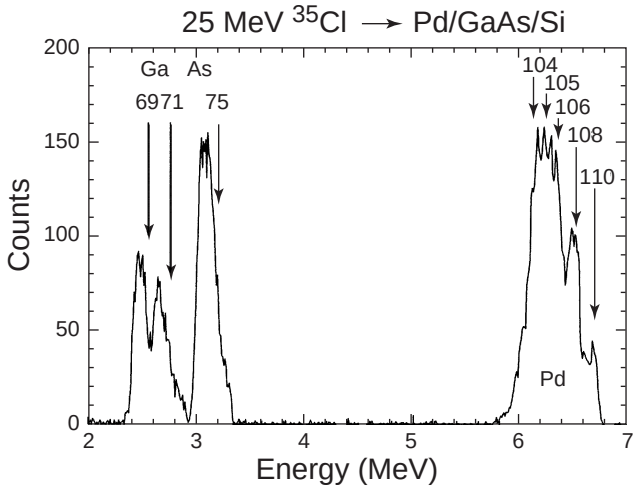

Figure 11 Heavy ion RBS spectrum (25-MeV ^{35}Cl beam) of a 48-nm layer of GaAs covered by 26 nm of Pd measured with a TOF detector. (Reproduced by permission of Elsevier Science B.V. from Döbeli et al.[36])

elements are less sensitive than standard RBS; in particular, elements lighter than the probe ion cannot be analyzed. (3) Irradiation of heavy ions causes significant radiation damage not only in targets but also in the SSB detectors.

4.2 High-resolution Rutherford Backscattering Spectroscopy

Improvement in the depth resolution up to atomic level has been tried by several groups.[37–46] An overview of existing high-resolution spectrometers for IBA is given in the literature.[45] It was shown that both a high-resolution spectrometer (energy resolution of ca. 0.1% is easily achieved for electrostatic or magnetic spectrometers) and a grazing-angle technique are essential for monolayer resolution in RBS.[44] Figure 12 shows examples of HRBS spectra of PbTe (100) and PbTe (111) measured with a magnetic spectrometer. The arrows labeled Pb and Te indicate the energies of the ions elastically scattered from Pb and Te atoms. There are several peaks at ca. 285.5, ca. 283.3 and ca. 281.1 keV in Figure 12(a). These peaks corresponds to the Pb atoms in the first, second, and third atomic layers, respectively. Note that even the ions scattered from the topmost atomic layer show inelastic energy loss because the electronic surface is outside the atomic surface. There are also small peaks at ca. 278 and ca. 275.8 keV corresponding to the Te atoms in the first and second layers. The yield ratio of the Pb peak to the Te peak is close to the cross-section ratio indicating that the PbTe (100) is stoichiometric. The spectrum of PbTe (111) shown in Figure 12(b) is basically the same as the PbTe (100) except for the yield of the first Pb

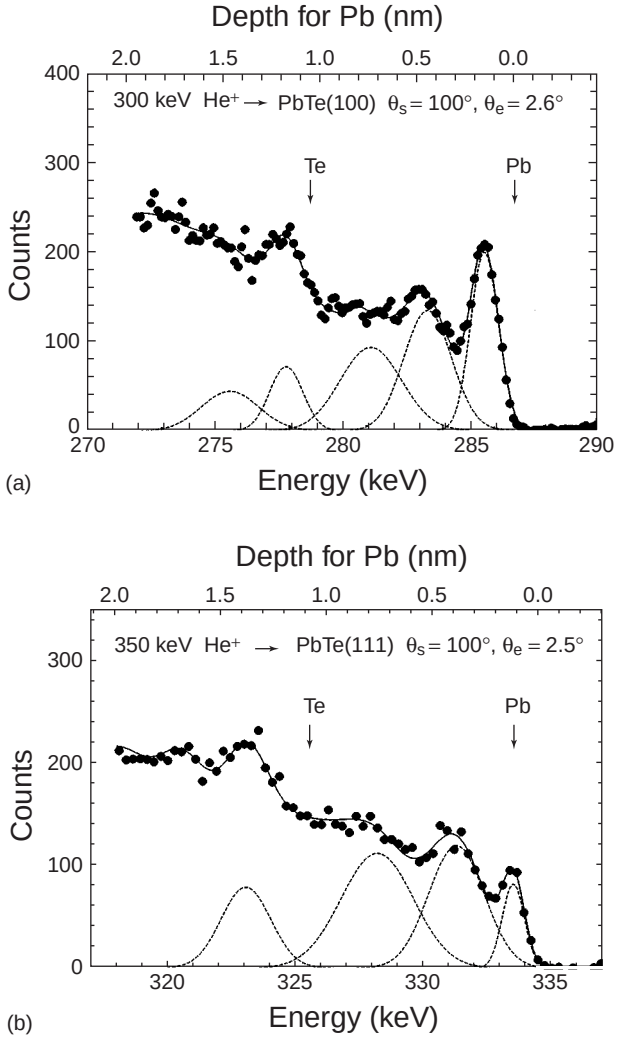

Figure 12 High-resolution rutherford backscattering spectroscopy (HRBS) spectra of (a) PbTe (001) and (b) PbTe (111) surfaces. Surface atomic layers can be resolved as separated peaks.

peak. The yield of the first Pb peak is about 25% of the second and third Pb peaks and the inelastic energy loss for the first Pb peak (ca. 1 keV) is smaller than that for the first Te peak (ca. 2.1 keV). From these results, it can be concluded that the PbTe (111) surface is terminated by the Pb layer and the atomic density of the topmost Pb layer is reduced to about 25% of the bulk value.

Combination of heavy ions with a high-resolution spectrometer may provide better depth resolution. In addition to the large radiation damage, however, the charge-state distribution of the scattered heavy ions complicates the analysis because the scattered ions of all charge states cannot simultaneously be measured by the electrostatic or magnetic spectrometers. An optimal probe ion for HRBS has been discussed and

the sub-megaelectron-volt He ion was concluded to be the best one.[46]

Disadvantages of HRBS are the relatively small acceptance angle $\Delta\Omega$ (typically less than 1 msr) and a narrow energy window of the spectrometer (typically several percent of the ion energy). These lead to a long acquisition time and possible radiation damage. Spectrometers with a wide acceptance angle and a wide energy window have been developed by several groups[47,48] and not only a high-resolution spectrometer but also a high-resolution RBS system with a compact sub-megaelectron-volt ion accelerator are now commercially available.

4.3 Microprobe Rutherford Backscattering Spectroscopy

Finely focused high-energy ion beams turn ion beam spectroscopy into ion beam microscopy.[49] Development of a high-energy ion microprobe was started as early as the late 1960s. There are more than 40 facilities which have high-energy ion microprobes. Magnetic quadrupole lenses are commonly used to form microprobes of the order of ca. 1 μm at a beam current of ca. 100 pA. In scanning the microprobe across a specimen and detecting a variety of ion-scattering processes, such as RBS, PIXE elastic recoil, nuclear reaction, and so on, microscope images are obtained. These images display features of the specimen that cannot readily be imaged by other techniques. The combination of the microprobe with RBS provides three-dimensional mapping of elements in the specimen.

Figure 13 shows an example of the RBS mapping and tomographic image of 1 μm size WSi_x wiring lines.[50] A beam of 300-keV Be^{2+} ions (beam size ca. 0.1 μm) was rastered over a $30 \times 30\,\mu m^2$ area and RBS spectra were measured at 128×128 points. The mapping image is made with the W signal of the observed RBS spectra.

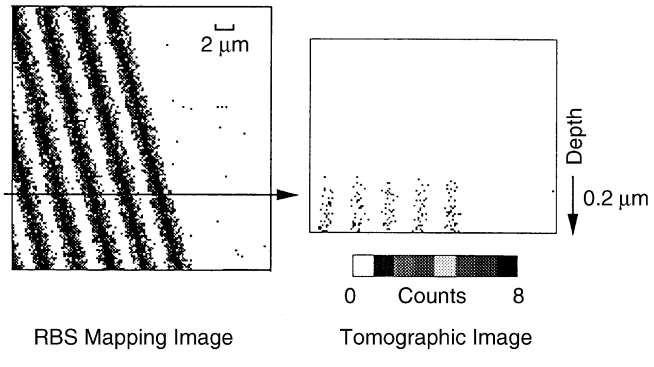

Figure 13 RBS mapping and tomographic image of 1-μm-sized WSi_x wiring lines measured with a microprobe. (Reproduced by permission of Elsevier Science B.V. from Park et al.[50])

The tomographic image shows the cross-sectional image along the line shown in the mapping image, which is obtained by converting the each RBS spectrum to a depth distribution of W. The WSi_x lines are clearly resolved in both the mapping and the tomographic images.

4.4 Light Element Analysis by Non-Rutherford Scattering

When the ion energy approaches the Coulomb barrier height of the target nucleus, the ion penetrates the Coulomb barrier and a compound nucleus in a highly excited state may be formed. If the incident energy of the ion matches one of the excited levels, the reaction probability is considerably changed and often enhanced. The excited compound nucleus thus produced may decay via re-emission of the incident particle. This phenomenon, called elastic resonance scattering, has a large cross-section sometimes several orders of magnitude larger than the Rutherford cross-section. Figure 14 displays an example of the non-Rutherford cross-section for $^4He-^{12}O$ scattering as a function of the incident He energy.[51] There are several resonances with enhancement factors of 100–200. Because the cross-sections for heavy elements remain Rutherford values in this energy region, the relative sensitivity for oxygen is improved by a factor of ca. 100. However, the cross-section varies in a very complicated manner with energy and also with the scattering angle. The analysis of energy spectra has to be performed with care. There is an energy region, 8.3–8.9 MeV, where the cross-section varies slowly. Using this energy region the analysis of oxygen is as simple as standard RBS with an enhancement factor of about 35.

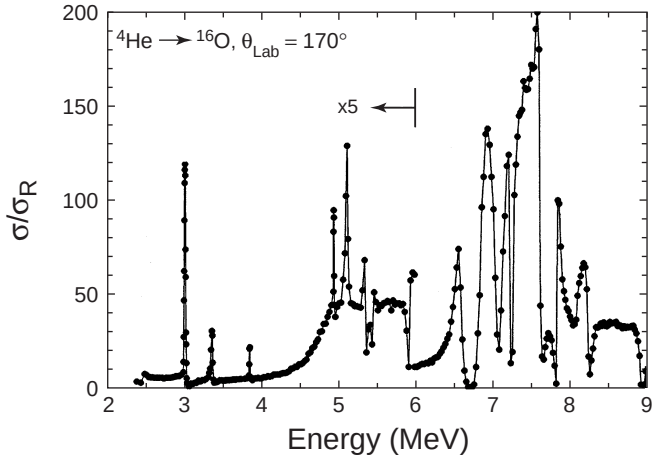

Figure 14 Measured 170° laboratory cross-sections for $^4He-^{16}O$ expressed in terms of the Rutherford cross-sections. (Reproduced by permission of Elsevier Science B.V. from Cheng et al.[51])

A sharp and isolated resonance is seen at $E_R = 3.045$ MeV. When the energy of the incident He ions is larger than E_R, the resonance occurs at a certain depth where the ion energy matches E_R after energy loss in the incoming path, that is, only the oxygen at a certain depth can be seen by the resonance. Thus a sharp resonance can be used for depth profiling of oxygen by sweeping the incident ion energy. Other examples of sharp resonances used in non-RBS are the 4.265 MeV resonance in the $^4He-^{12}C$ cross-section and 2.525 MeV resonance in the $^1H-^9Be$ cross-section. Useful data on non-Rutherford cross-sections for H and He incident ions are compiled in the literature.[52]

ABBREVIATIONS AND ACRONYMS

CAARI	Conference on the Application of Accelerators in Research and Industry
ECAART	European Conference on Accelerator Applications in Research and Technology
EPMA	Electron Probe Microanalysis
ERD	Elastic Recoil Detection
HIRBS	Heavy Ion Rutherford Backscattering Spectroscopy
HRBS	High-resolution Rutherford Backscattering Spectroscopy
IBA	Ion Beam Analysis
ISS	Ion Scattering Spectroscopy
NMTA	Nuclear Microprobe Technology and Applications
NRA	Nuclear Reaction Analysis
PHA	Pulse-height Analyzer
PIXE	Particle-induced X-ray Emission
RBS	Rutherford Backscattering Spectroscopy
SSB	Silicon Surface Barrier
TOF	Time-of-Flight

RELATED ARTICLES

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Elastic Recoil Detection Analysis • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • PIXE (Particle-induced X-ray Emission)

REFERENCES

1. A.B. Brown, C.W. Snyder, W.A. Fowler, C.C. Lauritsen, 'Excited States of the Mirror Nuclei, Li^7 and Be^7', *Phys. Rev.*, **82**, 159–181 (1951).
2. J.W. Mayer, M.-A. Nicolet, W.-K. Chu, in *Material Characterization Using Ion Beams*, eds. J.P. Thomas, A. Cuchard, Plenum Press, New York, 333–366, 1978.
3. L.C. Feldman, J.W. Mayer, *Fundamentals of Surface and Thin Film Analysis*, North-Holland, Amsterdam, 1986.
4. J.R. Tesmer, M. Nastasi (eds.), *Handbook of Modern Ion Beam Materials Analysis*, Material Research Society, Pittsburgh, 1995.
5. M.F. De Silva, J.C. Soares (eds.), 'Proceedings of the 13th International Conference on Ion Beam Analysis', Lisboa/Portugal 1997, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 136–138.
6. G. Molière, 'Theorie der Streuung schneller geladener Teilchen I Einzelstreuung am abgeshirmten Coulomb-Feld', *Z. Naturforsch, A: Phys. Sci.*, **2**, 133–145 (1947).
7. J.F. Ziegler, J.P. Biersack, U. Littmark, 'The Stopping and Ranges of Ions in Solids', in *The Stopping and Ranges of Ions in Matter*, ed. J.F. Ziegler, Pergamon Press, New York, Vol. 1, 1985.
8. M.H. Mendenhall, R.A. Weller, 'Algorithms for the Rapid Computation of Classical Cross-Sections for Screened Coulomb Collisions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **58**, 11–17 (1991).
9. J. L'Ecuyer, J.A. Davies, N. Matsunami, 'How Accurate Are Absolute Rutherford Backscattering Yields', *Nucl. Instrum. Methods*, **160**, 337–346 (1979).
10. J.A. Leavitt, L.C. McIntyre, Jr., 'Backscattering Spectrometry', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 37–81, 1995.
11. F. Bloch, 'Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie', *Ann. Phys. (Leipzig)*, **16**, 285–320 (1933).
12. H.H. Andersen, J.F. Ziegler, 'Hydrogen Stopping Powers and Ranges in All Elements', in *The Stopping and Ranges of Ions in Matter*, ed. J.F. Ziegler, Pergamon Press, New York, Vol. 3, 1977.
13. J.F. Ziegler, 'Helium Stopping Powers and Ranges in All Elemental Matter', in *The Stopping and Ranges of Ions in Matter*, ed. J.F. Ziegler, Pergamon Press, New York, Vol. 4, 1977.
14. ICRU Report 49, 'Stopping Powers and Ranges for Protons and Alpha Particles', International Commision on Radiation Units and Measurements, Bethesda, 1993.
15. N. Bohr, 'The Penetration of Atomic Particles through Matter', *Kgl. Danske Videnskab. Selskab. Matt-Fys. Medd.*, **18**(8), (1948).
16. J. Lindhard, M. Scharff, 'Energy Loss in Matter by Fast Particles of Low Charge', *Kgl. Danske Videnskab. Selskab. Matt-Fys. Medd.* **27**(15), (1953).
17. E. Bonderup, P. Hvelpund, 'Stopping Power and Energy Straggling for Swift Protons', *Phys. Rev. A*, **4**, 562–569 (1971).

18. W.K. Chu, 'Calculation of Energy Straggling for Protons and Helium Ions', *Pys. Rev. A*, **13**, 2057–2060 (1976).
19. F. Besenbacher, J.U. Andersen, E. Bonderup, 'Straggling in Energy Loss of Energetic Hydrogen and Helium Ions', *Nucl. Instrum. Methods*, **168**, 1–16 (1980).
20. A. Climent-Font, C. Palacio, M.T. Fernández-Jiménez, D. Díaz, 'Characterization of BiSrCaCuO Thin Films by the Complementary Use of IBA and AES Analytical Techniques', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **118**, 108–112 (1996).
21. L.R. Doolittle, 'A semi-automatic Algorithm for Rutherford Backscattering Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 227–231 (1986).
22. J. Saarilahti, E. Rauhala, 'Interactive Personal Computer Data Analysis of Ion Backscattering Spectra', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 734–738 (1992).
23. M.D. Strathman, S. Bauman, 'Angle-Resolved Imaging of Single Crystal Materials with MeV Helium Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 840–845 (1992).
24. N.P. Barradas, C. Jeynes, R.P. Webb, 'Simulated Annealing Analysis of Rutherford Backscattering Data', *Appl. Phys. Lett.*, **71**, 291–293 (1997).
25. J.W. Mayer, L. Eriksson, J.A. Davies, *Ion Implantation in Semiconductors*, Academic Press, New York, 1970.
26. Y. Quéré, 'Dechanneling of Fast Particles by Lattice Defects', *J. Nucl. Mater.*, **53**, 262–267 (1974).
27. L.C. Feldman, J.W. Mayer, S.T. Picrau, *Materials Analysis by Ion Channeling*, Academic Press, New York, 1982.
28. D.S. Gemmel, 'Channeling and Related Effects in the Motion of Charged Particles through Crystals', *Rev. Mod. Phys.*, **46**, 129–227 (1974).
29. M.L. Swanson, 'Channeling', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 231–300, 1995.
30. E. Taglauer, W. Englert, W. Heiland, D.P. Jackson, 'Scattering of Low-Energy Ions from Clean Surfaces: Comparison of Alkali- and Rare-Gas-Ion Scattering', *Phys. Rev. Lett.*, **45**, 740–743 (1980).
31. S.A.E. Johansson, J.L. Campbell, *PIXE: A Novel Technique for Elemental Analysis*, Wiley, New York, 1988.
32. W.A. Arnold Bik, F.H.P.M. Habraken, 'Elastic Recoil Detection', *Rep. Prog. Phys.*, **56**, 859–902 (1993).
33. J. L'Ecuyer, C. Brassard, C. Cardinal, J. Chabbal, L. Deschenes, J.P. Labrie, B. Terrault, J.G. Martel, R. St. Jacques, 'An Accurate and Sensitive Method for the Determination of the Depth Distribution of Light Elements in Heavy Materials', *J. Appl. Phys.*, **47**, 381–382 (1976).
34. W.A. Lanford, 'Nuclear Reactions for Hydrogen Analysis' in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 193–204, 1995.
35. E. Rauhara, J. Saarilahti, N. Nath, 'Protons, Helium and Heavy ions in Backscattering Analysis of High-T_c Superconductors', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **61**, 83–90 (1991).
36. M. Döbeli, P.C. Haubert, R.P. Livi, S.J. Spicklemire, D.L. Weathers, T.A. Tombrello, 'A Time-of-Flight Detector for Heavy Ion RBS', *Nucl. Instrum. Methods Phys. Res. Sect. B*, **56/57**, 764–767 (1991).
37. E. Bøgh, 'Rutherford Scattering of Protons in the Surface Layers of a Tungsten Single Crystal', *Phys. Rev. Lett.*, **19** 61–63 (1967).
38 J.K. Hirvonen, G.K. Hulbler, 'Application of a High-Resolution Magnetic Spectrometer to Near-Surface Material Analysis', in *Ion Beam Surface Layer Analysis*, eds. O. Meyer, G. Linker, F. Käppeler, Plenum, New York, 457–469, 1976.
39. A. Feuerstein, H. Grahmann, S. Kalbitzer, H. Oetzmann, 'Rutherford Backscattering Analysis with Very High Depth Resolution using an Electrostatic Analyzing System', in *Ion Beam Surface Layer Analysis*, eds. O. Meyer, G. Linker, F. Käppeler, Plenum, New York, 471–481, 1976.
40. T. Oku, J. Kanasaki, N. Matsunami, N. Itoh, K. Matsuda, M. Aoki, 'Proton Energy Loss Spectroscopy for Surface Layer Analysis in the Monolayer Regime', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **15**, 142–145 (1986).
41. J. Vrijmoeth, P.M. Zagwijn, J.F.M. Frenken, J.F. van der Veen, 'Monolayer Resolution in Medium-Energy Ion-Scattering Experiments on the NiSi2 (111) Surface', *Phys. Rev. Lett.*, **67**, 1134–1137 (1991).
42. Th. Enders, M. Rilli, H.D. Carstanjen, 'A High-Resolution Electrostatic Spectrometer for the Investigation of Near-Surface Layers in Solids by High-Resolution Rutherford Backscattering with MeV Ions', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **64**, 817–824 (1992).
43. W.M. Arnoldbik, W. Wolfswinkel, D.K. Inia, V.G.G. Verleun, S. Lobner, J.A. Reinders, F. Labohm, D.O. Boerma, 'A High Resolution Magnetic Spectrograph for Ion Beam Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **118**, 567–572 (1996).
44. K. Kimura, K. Ohshima, M. Mannami, 'Monolayer Analysis in Rutherford Backscattering Spectroscopy', *Appl. Phys. Lett.*, **64**, 2232–2234 (1994).
45. D.O. Boerma, 'Report on the Round Table Session on: Near Surface Analysis and Charge Exchange Studies with High Resolution Spectrographs', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 42–46 (1998).
46. K. Kimura, K. Nakajima, M. Mannami, 'Some Applications of High-resolution RBS and ERD using a Magnetic Spectrometer', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 1196–1202 (1998).
47. W.A. Lanford, B. Anderberg, H. Enge, B. Hjorvarsson, 'Compact Broad Range Magnetic Spectrometer for Use in Ion Beam Analysis', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **136–138**, 1177–1182 (1998).

48. K. Kimura, M. Kimura, Y. Mori, M. Maehara, H. Fukuyama, 'Development of a Compact High-resolution RBS System for Monolayer Analysis', in *Application of Accelerators in Research and Industry*, eds. J.L. Duggan, I.L. Morgan, American Institute of Physics, Woodbury, 500–503, 1999.
49. M.B.H. Breese, D.N. Jamieson, P.J.C. King, *Material Analysis Using a Nuclear Microprobe*, Wiley, New York, 1996.
50. Y.K. Park, T. Kishimoto, M. Takai, 'Microanalysis of a submicron patterned WSi_x Structure using a Nuclear Microprobe', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **130**, 534–538 (1997).
51. H.-S. Cheng, H. Shen, J. Tang, F. Yang, 'Cross Sections for 170° Backscattering of ^{4}He from Oxygen in the Energy Range of 2.0–9.0 MeV', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **83**, 449–453 (1993).
52. R.P. Cox, J.A. Leavitt, L.C. McIntyre, Jr, 'Non-Rutherford Elastic Backscattering Cross Sections', in *Handbook of Modern Ion Beam Materials Analysis*, eds. J.R. Tesmer, M. Nastasi, Materials Research Society, Pittsburgh, 481–508, 1995.

Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

M. Borsaru
CSIRO-Exploration and Mining, Kenmore, Queensland, Australia

1 Introduction 1
2 Scattering Cross-section 2
3 Scattering and Absorption of γ-Radiation 2
3.1 The Three Principal Processes 2
3.2 Other Interactions of γ-Rays With Matter 3
3.3 Density and Composition Measurements With γ-Rays 4
4 Natural γ-Radiation 5
5 Scattering and Absorption of Neutrons 6
5.1 Neutron Sources 7
5.2 Interaction of Neutrons with Matter 7
5.3 Macroscopic Treatment of Slowing Down 8
6 Radiation Detectors 8
7 The Determination of Water Content in Soil and Boreholes Using Neutrons 9
7.1 Definition of Porosity 9
8 The Monte Carlo Method 9
9 Nuclear Techniques for In Situ Analysis 10
9.1 Passive Logging – Natural γ-Ray 10
9.2 Active Logging 11
9.3 Oil Well Logging 11
9.4 Uranium 13
9.5 Coal 14
9.6 Iron Ore 14
Abbreviations and Acronyms 14
Related Articles 15
Appendix 1 15
The Compton Effect 15
Elastic Scattering of Fast Neutrons 15
References 15

Nuclear techniques are widely used in geological and mineral applications. Owing to the high penetration of γ-radiation and neutrons, nuclear techniques have found many applications, including well logging, bulk analysis and on-stream analysis. The first part of this article presents the physics of the interaction of γ-rays and neutrons with matter. Applications based on these interactions to well logging for oil, uranium, coal and iron ore are presented in the second part of the article. Applications based on the natural radioactivity of rocks are also covered.

1 INTRODUCTION

The interactions of γ-radiation and neutrons with matter form the basis of a large number of nuclear techniques developed for chemical analysis or used to infer properties of interest in logging. Nuclear techniques have the ability to measure properties of materials without contact and, because of the high penetration of both γ-radiation and neutrons, enable measurements to be made through walls. In borehole-logging applications, these two types of penetrating radiation can traverse the pressure housing of the logging tools and the formation and return a signal carrying information about the surrounding formation and the rock matrix.

This article presents the principles of the interaction of γ-rays and neutrons with matter and outlines geological/mineral applications. The emphasis is on nuclear borehole logging. Because of the deep penetration of neutrons and γ-rays, nuclear techniques are well suited to borehole-logging applications. The volume of matter investigated by nuclear radiation is much larger than the core analyzed from a cored borehole, so a more representative sample is analyzed.

Well logging grew from the need of the petroleum industry to evaluate hydrocarbon formations. Today, borehole logging is a mature technology, and nuclear logging is well established in the oil, gas, uranium and coal industries. It is also making inroads in the mineral industry. It is now very difficult to imagine these industries without logging.

Nuclear logging, employing in situ γ-ray spectroscopy, can measure the chemical composition of the earth formation, thereby providing an indication of the dominant mineralogy. This is vital information in oil well logging. In situ γ-ray spectroscopy replaces the time-consuming chemical analysis of the core in a laboratory. Bulk density, which is another important property of an earth formation, is measured by γ-ray scattering. Some minerals contain the natural radioactive elements potassium, thorium and uranium. The detection of their γ-radiation indicates the presence of these elements, and quantitative measurements can be made of their concentration.

Nuclear applications are an integral part of today's measurement technologies, contributing to increased

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

efficiency and large financial savings in industry. Although a mature technology, new applications are still being found.

2 SCATTERING CROSS-SECTION

For the study of neutron and γ-ray scattering it is convenient to introduce the scattering cross-section. For a very thin target bombarded with neutrons or γ-rays, the number of scattered particles is very small compared with the number of particles in the beam, as shown in Equation (1):

$$-\frac{\mathrm{d}N}{N} = A\,\mathrm{d}x = n\sigma\,\mathrm{d}x \tag{1}$$

where $-\mathrm{d}N$ is the number of scattered particles, N is the number of incoming particles, $\mathrm{d}x$ is the thickness of the target and A is a proportionality constant dependent on the number of nuclei per unit volume n and the nature of the nuclei, represented by σ, the scattering cross-section. For a thicker target, the number of particles passing through undeflected, N, is obtained by integrating Equation (1), illustrated in Equation (2):

$$N = N_0\,\mathrm{e}^{-n\sigma x} \tag{2}$$

where N_0 is the total number of particles in the beam which hit the target. The total number of particles scattered from the beam is shown in Equation (3):

$$N_{\mathrm{sc}} = N_0(1 - \mathrm{e}^{-n\sigma x}) \tag{3}$$

When the target contains several isotopes of different cross-sections, $n\sigma$ is replaced by $n_1\sigma_1 + n_2\sigma_2 + n_3\sigma_3 + \cdots$

3 SCATTERING AND ABSORPTION OF γ-RADIATION

γ-Rays are electromagnetic radiation, as are X-rays, light, radio waves, etc. For γ-rays or neutrons, the processes by which they interact with matter are such that a single event will remove the γ-ray or neutron from the beam. According to Equation (2), the intensity of the beam hitting a target will decrease exponentially, as illustrated in Equation (4):

$$N = N_0\,\mathrm{e}^{-n\sigma x} = N_0\,\mathrm{e}^{-\mu x} \tag{4}$$

where the constant μ is called the absorption coefficient and has dimension $(\text{length})^{-1}$. It is also common to measure absorber thickness in $\mathrm{g\,cm^{-2}}$ and write Equation (4) as Equation (5):

$$N = N_0\,\mathrm{e}^{-\mu_{\mathrm{m}}\xi} \tag{5}$$

where $\xi = \rho x$ is the mass per unit area and where the mass absorption coefficient μ_{m} is, as illustrated in Equation (6):

$$\mu_{\mathrm{m}} = \frac{\mu}{\rho} = \frac{n\sigma}{\rho} = \frac{N_{\mathrm{A}}\sigma}{M_{\mathrm{A}}} \tag{6}$$

N_{A} is Avogadro's number, M_{A} is the atomic weight of the element in the absorber and ρ is the density. If more elements are present in the absorber, then Equation (7) applies:

$$\mu_{\mathrm{m}} = N_{\mathrm{A}} \sum_i \frac{\alpha_i\sigma_i}{M_{\mathrm{A}i}} \tag{7}$$

where α_i is the abundance ratio of the ith element.

The interaction of γ-radiation with matter in the energy range from 50 keV to 50 MeV is dominated by three processes: (1) photoelectric effect, (2) Compton effect and (3) pair production.

3.1 The Three Principal Processes

3.1.1 The Photoelectric Effect

In the photoelectric effect, predominant at low energies, the γ-ray (photon) gives all its energy to a bound electron. The electron, to overcome its binding energy in the atom, uses part of the energy and the rest is taken as kinetic energy. The energy of the emitted electron is therefore $E_{\mathrm{e}} = E_\gamma - E_{\mathrm{b}}$, where E_γ is the γ-ray energy and E_{b} is the binding energy of the electron in the atom, usually the k-shell energy. The probability of photon interacting with an electron in a given orbit is higher when the energy of the photon is equal to the binding energy and decreases with increasing energy. The electron gives up its energy to other atoms and eventually falls back into an orbit. The X-rays emitted following the capture of the electrons into orbits interact with the matter again and again until all the energy is absorbed either at very low-energy orbital levels or other mechanisms which are outside the scope of this article.

The cross-section of the photoelectric absorption per atom is a function of photon energy, E, illustrated in Equation (8), and is approximated[1] as being proportional to:

$$\sigma(E) \approx \frac{Z^{4.5}}{E^n},\ 2.5 \le n \le 3.5 \tag{8}$$

The mass absorption coefficient for the photoelectric effect is shown in Equation (9):

$$\mu_{\mathrm{mph}} = \rho N_{\mathrm{A}} \sum_i \frac{\alpha_i\sigma_i(E)}{M_{\mathrm{A}i}} \tag{9}$$

The contribution of the photoelectric effect at low energies is significant for heavy elements because of

their high atomic number. For lighter elements, the contribution of the photoelectric effect is reduced as a result of the linear relationship between its cross-section and $Z^{4.5}$.

3.1.2 The Compton Effect

The Compton effect predominates around 1 MeV. It implies scattering of photons by the atomic electrons. The photon is deflected from its trajectory with or without loss of energy, as shown schematically in Figure 1 (ν is the frequency of the photon of wavelength λ, where $\nu = 1/\lambda$). This happens at photon energies much larger than the electron binding energies so that, theoretically, the electrons can be considered free. Expressions for the energy $h\nu'$ of the scattered photon, the energy T of the scattered electron and the relationship between the angles of scattering shown in Figure 1 are given in Appendix 1. The cross-section for the Compton scattering was calculated by Klein and Nishina.[2] It is well described by Davisson[3] and many other books of nuclear physics or quantum mechanics.

3.1.3 The Pair Production Effect

In the pair production effect, a photon disappears with the creation of an electron/positron pair. The total kinetic energy of the resultant particles is equal to the photon energy minus the mass energy of the electron/positron pair created by the interaction (1.02 MeV) and therefore the pair production effect cannot take place below 1.02 MeV. Figure 2 shows schematically the pair production effect. The pair production takes place in the field of a nucleus or electron with no change of the state of the nucleus or its electrons. The nucleus must be present to absorb some of the momentum of the photon. Pair production cannot take place in the absence of the nucleus or electron because the conservation of energy and conservation of momentum cannot be simultaneously observed.

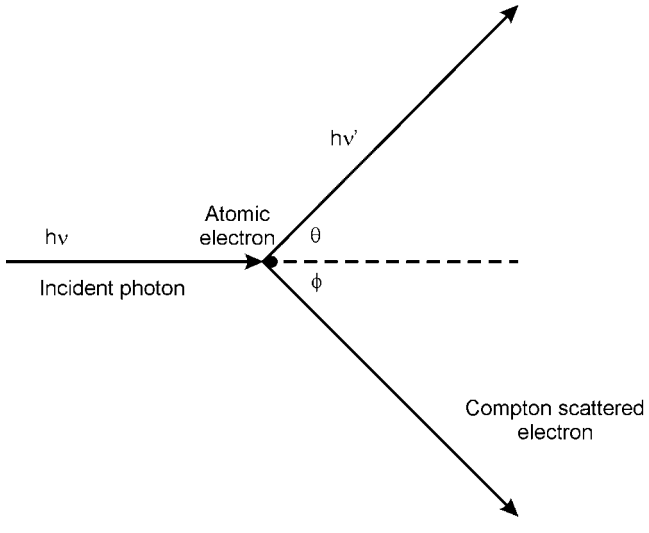

Figure 1 Compton effect.

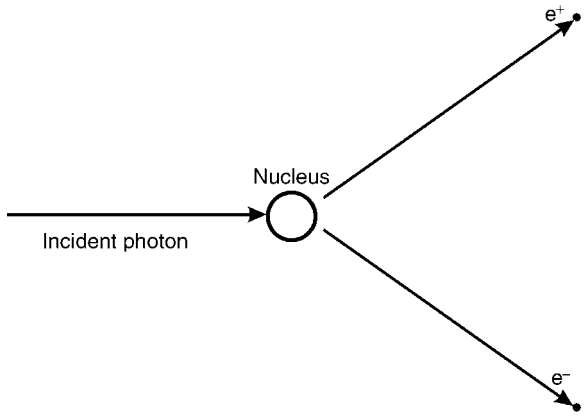

Figure 2 Pair production effect.

3.2 Other Interactions of γ-Rays With Matter

The photoelectric, Compton and pair production effects are the major interactions of γ-rays with matter and contribute most of the total γ-ray attenuation coefficient in the energy region being considered. However, they are not the only interactions. Other interactions of interest are resonance scattering of γ-rays and photonuclear reactions, also called γ-activation. There are other interactions (Thomson, Delbrück, Rayleigh) which occur at low energies. They are treated in more specialized books and are outside the scope of this general article.

3.2.1 Resonance Scattering

γ-Ray resonance scattering is an elastic process that takes place via an excited state of a stable nucleus. If the energy of the γ-ray corresponds exactly to the energy level of the stable nucleus, the nucleus can absorb this γ-ray and become an unstable excited nuclear state. The nucleus will regain its stable state by emitting a γ-ray that is theoretically of the same energy as the one absorbed. In practice, the γ-ray emitted by the excited nucleus at rest has less energy than the photon that excited the nucleus, E_r, because of the recoil energy taken by the nucleus. If M is the mass of the emitting atom, conservation of energy and momentum leads to the expression for the energy of the emitted photon $h\nu$: $h\nu = E_r - (E^2/2Mc^2)$. The same amount of kinetic energy is transferred to the nucleus which is excited by the γ-ray, so that the total energy displacement is $\Delta E = E^2/Mc^2$. In general, ΔE is large by comparison with the width of the energy level Γ and the system is out of resonance, in other words

the cross-section for resonance absorption is small. From Heisenberg's principle, $\Gamma(\text{eV}) = 6.58 \times 10^{-16}/\tau$, where τ is the lifetime of the excited state in seconds. This is why it took a long time to prove the existence of γ-ray resonance scattering.

Mössbauer discovered that an atom embedded in a crystal lattice can emit or absorb γ-radiation without loss from the recoil because the momentum is taken up by the crystal as a whole with negligible energy loss.

There are a few methods of producing resonance scattering. The change of quantum energies can be achieved by moving the absorber and the radiation source with respect to each other (based on the Doppler effect). Another practical technique to obtain γ-rays of the precisely defined energy for resonance scattering is to choose a radioisotope source that decays via the excited state of the chosen element. For example, if the element chosen for γ-ray scattering is ^{60}Ni, a ^{60}Co source may be used. ^{60}Co decays by β- and γ-ray emission via excited states of ^{60}Ni. Matching the radioactive source to the chosen element is essential for γ-ray resonance. Resonance scattering does not take place when the γ-ray source is in a solid state because the recoil energy losses during the emission of γ-rays make it deficient in energy by several tens of keV. This deficiency may be overcome by using a gaseous source. γ-Rays from gaseous sources are Doppler broadened so that about 1% of the γ-rays are in resonance.

γ-Ray resonance scattering has found practical applications. Mössbauer spectroscopy is a powerful technique for elemental analysis based on this effect. Copper and nickel are two favorable elements for analysis by γ-ray scattering using vapor sources of $^{65}ZnI_2$ and $^{60}CoBr_2$ respectively.

3.2.2 Photonuclear Reactions

γ-Rays of high energies can produce nuclear reactions with atomic nuclei. If the incident γ-ray energy exceeds a threshold energy, it is possible to remove particles from stable nuclei. Each element is characterized by a particular threshold energy. The most frequent nuclear reactions induced by high-energy γ-rays in the energy range of interest are (γ,n) processes. The maximum cross-section for light elements is at 20–25 MeV and $\cong$15 MeV with other elements. The photoneutron reaction has found practical applications mostly for the determination of low-atomic-number elements. The technique is selective if the (γ,n) threshold energies are selected appropriately.

The only reactions to have threshold energies less than 2.5 MeV are the (γ,n) reactions for ^{9}Be (threshold = 1.67 MeV) and ^{2}H (2.22 MeV). The (γ,n) reaction with beryllium has been used in different techniques of analysis for this element. ^{124}Sb is a suitable γ-ray source for beryllium analysis. It has a half-life of 60 days and emits γ-rays of 1.69 and 2.09 MeV. The average energy of neutrons from the ^{9}Be(γ,n) reaction using a ^{124}Sb source is approximately 24 keV. The neutrons are moderated in a hydrogenous moderator and detected with thermal neutron detectors such as $^{10}BF_3$ or ^{3}He counters.

3.3 Density and Composition Measurements With γ-Rays

The three major interactions of γ-radiation with matter have found wide industrial applications. Here we review the basis of density and composition determination.

The Compton mass attenuation coefficient μ_c for photon energies much larger than the electron binding energies can be expressed as (Equation 10):

$$\mu_c = \left(\frac{\rho N_A Z}{M_A}\right) e_\mu \tag{10}$$

where $\rho N_A/M_A$ represents the number of atoms per cubic centimeter, Z is the number of electrons in the atom and e_μ is the Compton scattering cross-section per electron. For a γ-ray of given energy (Equation 11):

$$\mu_c = \frac{Z}{M_A}\rho \tag{11}$$

Except for hydrogen, the Z/M_A ratio for low-Z elements (which form the Earth's crust) is very nearly a constant equal to $\frac{1}{2}$. For hydrogen, $Z/M_A = 1$, and this introduces an error when hydrogenous material is present in matter investigated. If the energy of γ-rays is greater than 300 keV, one can neglect the photoelectric effect and assume that the interaction is by the Compton effect. Equation (4) can be used to derive Equation (12):

$$\log N = \log N_0 - \mu_c x = \log N_0 - \frac{Z}{M_A}\rho x \tag{12}$$

For a radioactive source of long half-life, N_0 is constant. By plotting a graph of $\log N$ vs density (ρ), one obtains a straight line if the thickness χ of the materials of different densities is constant. The density is extracted from this linear relationship.

Concept of Equivalent Atomic Number, Z_{eq} In order to facilitate measurement of the chemical composition of a medium by γ-scattering, Czubek[1] introduced a quantity called the equivalent atomic number (Z_{eq}) of the scattering medium in terms of photoelectric and Compton coefficients. We have seen that the Compton and photoelectric absorption coefficients (μ_c, μ_{ph}) for given γ-ray energies are proportional to $(Z/M_A)\rho$ and $(Z^{4.5}/M_A)\rho$ respectively. Therefore, $\mu_{ph}/\mu_c = \text{const.}Z^{3.5}$. One can

show that, for a mixture of elements (Equation 13):

$$Z_{\mathrm{eq}} = Z = \sqrt[3.5]{\frac{\sum_i W_i \frac{Z_i^{4.5}}{M_{\mathrm{A}i}}}{\sum_i W_i \frac{Z_i}{M_{\mathrm{A}i}}}} \tag{13}$$

where Z_{eq} is a substitute for Z for a multielement medium. With the assumption that $Z_i/M_{\mathrm{A}i}$ is a constant for low-Z elements,

$$Z_{\mathrm{eq}} = \sqrt[3.5]{\sum_i W_i Z_i^{3.5}} \tag{14}$$

Equation (14) shows the strong dependence of the Z_{eq} of a medium on the constituent element with the highest atomic number. This dependence can be used to determine the concentration of a high-Z element in a matrix of low-Z elements, e.g. the percentage of Fe in iron ore, ash in coal, etc. (%Fe or %ash are proportional to Z_{eq}).

Practical Methods of Density and Z_{eq} Measurements When there is access to both sides of the medium under investigation, the density and Z_{eq} are determined by measuring the direct attenuation of γ-radiation passing through the medium. A collimated γ-ray beam of energy greater than 300 keV is employed for density measurements. The γ-ray sources used are usually ^{60}Co and ^{137}Cs. By using γ-rays of both low energy and high energy in a medium with a high photoelectric absorption coefficient, one can determine both the photoelectric and Compton absorption coefficients of the medium and thus Z_{eq} from Equation (13).

When only one side of the medium is available, e.g. boreholes, roads or thick walls, the detector and the source are placed on the same side of the medium. Shielding is used to stop the γ-rays from the source reaching the detector. Figure 3 shows schematically the geometry suitable for this application together with the γ-ray energy spectrum recorded by the detector. The γ-rays from the source undergo successive Compton scattering, losing energy. The γ-rays can reach the detector after one or multiple scatterings. Below 300 keV the photoelectric absorption increases with decreasing γ-ray energy. This spectrum is called the backscattered spectrum. It ranges from almost the source energy down to an energy at which photoelectric absorption reduces the spectral intensity to almost zero. The high-energy region of the spectrum, above 300 keV, is a function of the electronic density of the medium, and below 300 keV both photoelectric and Compton effects are present. Theory shows that the ratios of the count rates recorded in two energy windows, at high and low energy, is a function only of the Z_{eq} of the medium. Borehole logging techniques for the determination of Fe grade in iron ore deposits or ash in coal have been developed on this basis.[4,5]

4 NATURAL γ-RADIATION

All rocks and soils contain a great number of radioactive elements that emit γ-radiation. The three main sources of natural γ-rays are potassium, decay products in the uranium series and decay products in the thorium series.

Of the three naturally occurring isotopes of potassium, only ^{40}K is radioactive; it has a half-life of 1.28×10^9 years and isotopic abundance of 0.0118%, which gives a specific activity of 31.4 Bq per g of natural potassium. ^{40}K decays to ^{40}Ca and emits a γ-ray of energy 1.46 MeV. Because many rock-forming minerals contain potassium, it is the commonest natural γ-ray encountered in nature.

^{238}U and ^{235}U are more abundant among the isotopes of uranium, with ^{238}U making up 99.3% of the total. Each isotope of uranium decays by α-emission to an isotope of thorium. Thorium and its daughters are also radioactive and a decay chain of the parent isotope is produced. ^{238}U has a half-life much longer than its radioactive decay products, and in an undisturbed medium a secular equilibrium is established. A radioactive decay series is said to be in a state of secular equilibrium when the number of atoms of each daughter being produced in the series is equal to the number of atoms of that daughter being lost by radioactive decay. ^{214}Bi is one of the isotopes produced in the ^{238}U chain and decays emitting a 1.765 MeV γ-ray. This γ-ray can be used for the determination of uranium concentration in a uranium deposit if the deposit is in equilibrium. Uranium is widespread in the environment. Most crustal rocks contain some uranium, averaging about 2.7 ppm or 33 Bq kg^{-1}. Phosphate-bearing rocks used as fertilizer contain high concentrations of uranium. It is also found in high concentration in oceans, ranging from about 2 to 3.7 mg m^{-3}.

^{232}Th has a decay chain similar to the decay chains of uranium. One prominent γ-ray of energy 2.6 MeV, emitted by the daughter product ^{208}Tl, is usually used for the determination of thorium concentration. Like uranium, thorium is widely found in rocks. The average thorium content of the continental upper crust is 9–10 ppm, about four times greater than that of uranium. However, having lower activity than uranium, the radioactivity concentration of thorium and uranium is about the same. High concentrations of thorium ($\cong$10%) are found in monazite, a rare-earth mineral. Owing to its high level of radioactivity, people living close to monazite sands can be exposed to high doses of radiation.

Measurement of natural radiation has found many applications. Because of the high level of natural

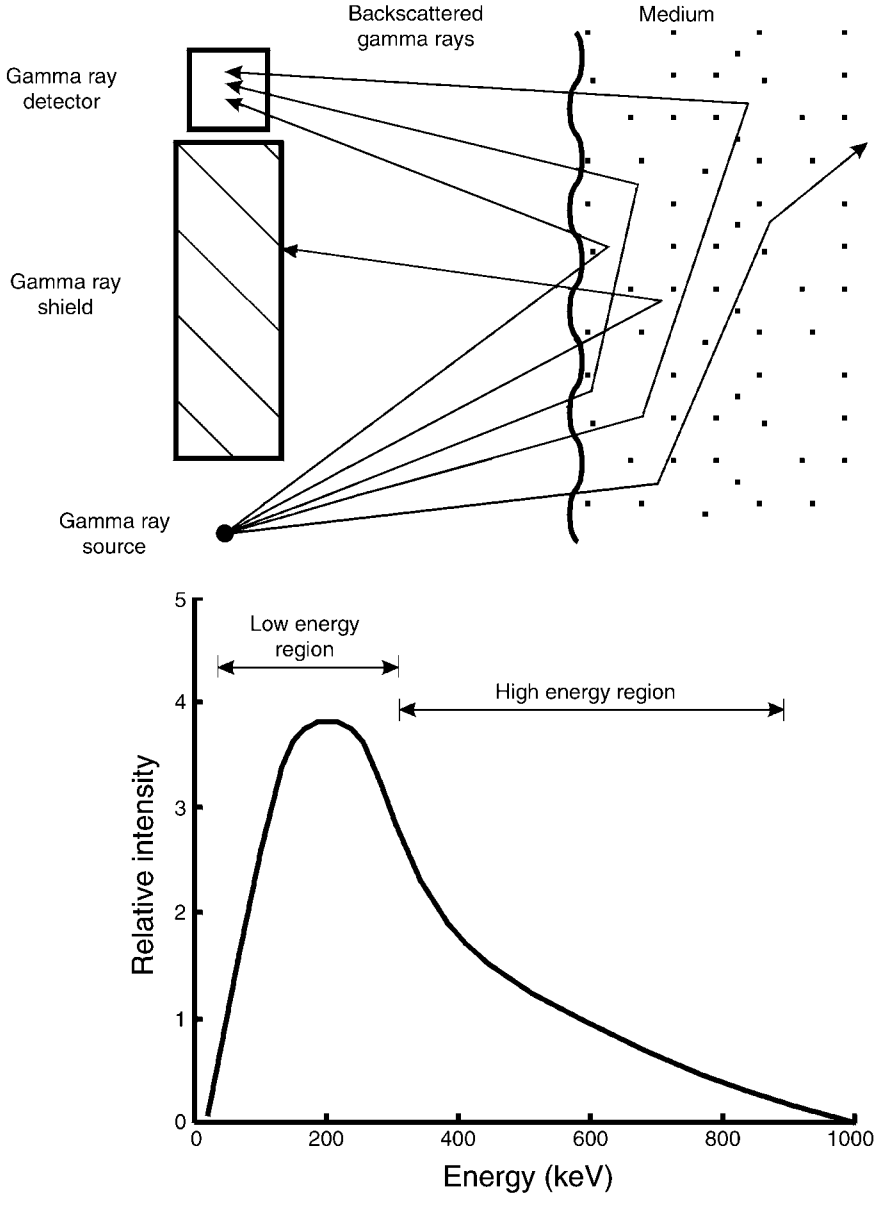

Figure 3 Schematic representation of the backscattered γ-rays and a backscattered γ-ray spectrum.

radioactivity in shales it is used to distinguish between shales and other sediments. It is also used in the exploration and mining of uranium. A technique to determine the ash content of coal on conveyor belts has also been developed. It is based on the fact that uranium, thorium and potassium are associated with the mineral mater in coal and the γ-ray activity of coal is correlated with the ash content.[6] An interface gauge for the simultaneous monitoring of the green liquor and red mud interface and the profile of the solids concentration has been developed for the alumina industry. It has been shown that, subsequent to the digestion of bauxite in caustic soda, traces of radioactivity originally present in bauxite remain with the red mud and the green liquor is free of radioactivity.[7] The gauge consists of a number of γ-ray counters arranged inside a Teflon-coated steel tube, which is suspended vertically inside settling tanks.

5 SCATTERING AND ABSORPTION OF NEUTRONS

The neutron is a neutral particle having a mass approximately equal to the mass of a proton (about 1800 times the rest mass of an electron). The neutron has a magnetic moment and therefore there is an electromagnetic

interaction between the neutron and the atomic electrons. However, this interaction is small and can be neglected when studying the scattering and stopping of neutrons in matter. In the present discussion we will assume that the neutron interacts exclusively with the atomic nuclei. Consequently, neutron interactions in matter are rarer than γ-ray interactions and neutron ranges are longer.

The neutron energies are arbitrarily classified as:

- fast neutrons – above 500 keV
- intermediate – 1 keV to 500 keV
- slow – below 1 keV.

Slow neutrons are further subdivided as

- epithermal neutrons – from 0.1 eV to 1 keV
- thermal neutrons – below 0.1 eV.

5.1 Neutron Sources

The most common neutron sources used in applications are: (1) isotopic sources, (2) neutron generators and (3) nuclear reactors.

5.1.1 Isotopic Sources

Isotopic sources use (i) (α,n) reactions, (ii) photo-neutron reactions or (iii) the fission process.

Neutron Sources Based on the (α,n) Reaction The principal reaction leading to neutron emission is the reaction: $^{9}\mathrm{Be} + \alpha \longrightarrow {}^{12}\mathrm{C} + \mathrm{n}$. This source consists of an α-emitting nuclide, e.g. Am, Pu, Po or Ra, mixed with beryllium. The α-emitter and the beryllium are usually in powder form and are doubly sealed in stainless-steel capsules.

Neutrons Emitted by Photonuclear Reactions The binding energy of the last neutron in ^{9}Be is only 1.67 MeV, and neutrons can therefore be released from ^{9}Be by photodisintegration, i.e. by bombardment with γ-rays of energy 1.67 MeV or more.

Fission Sources The spontaneous fission of ^{252}Cf, which is an artificial isotope produced in nuclear reactors, is a source of neutrons. The Cf nucleus splits spontaneously into two fission products and emits one or more neutrons.

5.1.2 Neutron generators

Intense neutron radiation can be obtained by bombarding various targets with charged particles from a Van de Graaff accelerator or from a cyclotron. Commercially available neutron generators use one of the following reactions: $^{2}\mathrm{H} + {}^{2}\mathrm{H} \longrightarrow {}^{3}\mathrm{H} + \mathrm{n}$ (D–D reaction) or $^{2}\mathrm{H} + {}^{3}\mathrm{H} \longrightarrow {}^{4}\mathrm{He} + \mathrm{n}$ (D–T reaction). Neutron generators based on these reactions produce neutrons of energies 2.6 MeV and 14 MeV respectively.

5.1.3 The Nuclear Fission Reactor

The nuclear fission reactor provides neutron fluxes of high intensity close to the reactor core.

5.2 Interaction of Neutrons with Matter

A fast neutron suffers many interactions in the process of losing its energy to reach thermal energy. This process is called slowing down. The neutron may disappear during this process as a result of nuclear reactions when the neutron impinges on nuclei forming the scattering medium. The probability of scattering from or reaction with a single nucleus is called cross-section (σ). The unit for σ is the barn and is equal to 10^{-24} cm^2.

There are four types of interactions that can occur between the neutron and the surrounding nuclei: (1) particle reactions (n,*x*), e.g. (n,α), (n,p), (n,*x*n) (for neutrons of high energy only), (2) inelastic scattering (n,n′γ) (the neutron energy must be above a threshold characteristic for each element, normally above 1 MeV), (3) elastic scattering and (4) radiative capture.

5.2.1 (n,x) Reactions

A nuclear reaction involves the release of a charged particle or more than one neutron. The final reaction product cannot be the same as the target nucleus. Characteristic γ-rays are emitted and the neutron reactions can be used to identify the presence of particular elements in materials or for the determination of their concentration.

5.2.2 Inelastic Scattering

A neutron can interact with a nucleus of atomic number Z and mass number A to form a compound nucleus in an excited state of mass $A + 1$. The compound nucleus decays very rapidly to the ground state by emitting a γ-ray, and it then emits a neutron. The γ-rays that are produced are characteristic of the energy levels of the compound nucleus and can be used to identify the target nucleus.

5.2.3 Elastic Scattering

Elastic scattering is the most important interaction for neutrons produced by isotopic sources (average energy below 4.5 MeV). This interaction is also known as ''billiard ball'' scattering. Although the neutron reactions and inelastic scattering reduce the neutron population, their importance is modest by comparison with neutron scattering. Inelastically scattered neutrons can suffer large

energy loss; however, they constitute a small fraction of the neutrons slowing down to low energies. Most of the neutrons reach low energies through repeated elastic collisions. In elastic scattering, a percentage of the incident neutron's kinetic energy is transferred to the recoiling nucleus and the neutron will have less kinetic energy after collision.

A useful quantity in the theory of the neutron slowing-down process is the percentage decrease in the neutron energy as the result of a collision. This is usually expressed as the *average logarithmic energy decrement* ξ, also known as neutron *lethargy*. For nuclei with mass $A > 10$, which are common in the Earth's crust, $\xi = 2/(A + 2/3)$. For hydrogen $\xi = 1$. Typical values of ξ for a few common nuclei are: H(1), C(0.158), O(0.12), Al(0.072), Si(0.070), Ca(0.049), Fe(0.035). The lethargy is used to calculate the mean number of scatterings required to slow down the neutron. For example, the average number of scatterings calculated for a few elements to slow down a neutron of 14 MeV, produced by a neutron generator, to thermal energy (0.01 eV) are: H(19), C(112), O(154), Al(290), Si(297), Fe(539). These examples show the unique position of hydrogen in slowing down neutrons. Its position is further enhanced when taking into account scattering cross-sections. Most of the nuclei abundant in the Earth's crust have scattering cross-sections of several barns. The proton's scattering cross-section between $\approx 10^5$ eV and 0.5 eV is about 20 barn.

The cross-section for elastic scattering, the relationship between the neutron energies before and after scattering and the theoretical expressions for lethargy are given in Appendix 1.

5.2.4 Radiative Capture

Once slowed down to thermal energies, neutrons diffuse through the medium without further loss of energy until their life is terminated by other processes such as radiative capture. In this process, the thermal neutron enters the nucleus, producing a compound nucleus in an excited state, which then decays to the ground state by the emission of one or more γ-rays that are characteristic of the capture nucleus. The radiation capture process only takes 10^{-11} s, which is virtually instantaneous compared with the initial slowing down and diffusion process, which may take several hundred microseconds. The product nucleus may be stable or it may decay (with a half-life of between a few microseconds and a number of years) to another product nucleus, often with β-particle emission. This is called thermal neutron activation.

5.3 Macroscopic Treatment of Slowing Down

For a moderating medium consisting of a homogeneous collection of different elements, the elastic macroscopic scattering cross-section Σ_s can be written as $\Sigma_s = \sum_i n_i \sigma_{si}$, where n_i is the number of nuclei of the ith element per cubic centimeter and i represents the existing elements. The same applies to the absorption cross-section $\Sigma_a \cdot \Sigma$ is analogous to μ in γ-ray attenuation. For a beam of neutrons passing through a material of thickness x, $N = N_0 e^{-\Sigma x}$ where N represents the number of neutrons that have had no interaction. The mean free path λ is defined as $\lambda = 1/\Sigma$. For a neutron density of N neutrons per cubic centimeter, the number of interactions per cubic centimeter is $N\nu\Sigma$, where ν is the velocity of the neutron. The quantity $N\nu$ is called the flux ϕ and has units of neutrons per square centimeter per second. $\langle\xi\rangle$ for the moderating medium is the weighted-average cross-section taken over all i nuclei: $\langle\xi\rangle = \sum_i (n_i \sigma_{si} \xi_i)/\Sigma_s$.

The theory of neutron slowing down and diffusion is presented in books dealing with nuclear reactor theory. It was developed by Glasstone and Edlund[8] for nuclear reactors. The theory was later developed for nuclear logging by other authors. Neutron slowing down ends when the neutrons reach thermal equilibrium with the atoms and molecules of the medium. Once neutrons have reached thermal energies, they diffuse through a medium in the same way that one gas diffuses through another, except that their lifetime is limited by absorption. If there was no absorption, the distribution of neutron energies would follow the Maxwell–Boltzmann gas distribution law; the most probable energy, defined as the energy corresponding to the peak of the distribution curve, would be kT, where T is the temperature in Kelvin (absolute temperature) and k is the Boltzmann constant (8.61×10^{-5} eV K^{-1}). For a temperature of 25 °C, kT is about 0.025 eV. The neutron population is given by the diffusion equation which describes conservation of neutrons: Production + scattering in − absorption − leakage = 0.

The Boltzmann transport equation derived for gases also applies to neutrons. Solving the Boltzmann equation is not an easy task. For more accurate results, the neutrons are treated as belonging to several energy groups (multigroup diffusion theory).

6 RADIATION DETECTORS

The γ-ray detectors most commonly used in nucleonic gauges can be divided into three categories: gas-filled counters, scintillation detectors and solid-state detectors. The gas-filled counter is one of the oldest detectors for nuclear radiation. The Geiger–Mueller counter is still used in nuclear applications for its reliability and low cost. The scintillation detector is one of the most widely used detectors. Popular scintillators are NaI(Tl), CsI(Tl)

or CsI(Na) and $Bi_4Ge_3O_{12}$ (commonly referred to as BGO). The solid state detector is used in applications when high-energy resolution is required.

Common neutron detectors are the boron counter, based on the (n,α) reaction $^{10}B + n \longrightarrow {}^{7}Li + \alpha + \gamma$, the ^{3}He detector, based on the (n,p) reaction $n + {}^{3}He \longrightarrow {}^{3}H + p$ and special scintillators developed for neutron detection.

7 THE DETERMINATION OF WATER CONTENT IN SOIL AND BOREHOLES USING NEUTRONS

The principle of the determination of water content is the slowing down of fast neutrons. Fast neutrons emitted from a neutron source successively undergo the processes of slowing down, thermalization and diffusion. Hydrogen primarily determines the slowing down power of the medium, but it does not dominate its neutron transport property because of the presence of elements with very large neutron absorption cross-sections (e.g. Cl, Cd, In, Sm, Gd, B). These elements serve to keep the scattered neutrons in the vicinity of the source.

7.1 Definition of Porosity

It is common to express the water content in terms of porosity. The bulk density of an ideal formation consisting of uniformly distributed, fluid-filled pores in a rock matrix is $\rho_B = \Phi\rho_f + (1 - \Phi)\rho_{ma}$. Here, ρ_f and ρ_{ma} are the fluid and rock matrix densities, respectively, and Φ is the porosity. The porosity is given by Equation (15):

$$\Phi = \frac{\rho_{ma} - \rho_B}{\rho_{ma} - \rho_f} \tag{15}$$

When neutrons travel through matter from the point source, their fluxes decrease with distance. The rate of decrease is governed by a ''scaling unit'', which is a property of the medium. The scaling units are the slowing-down length L_s, which is a measure of the ability of the bulk medium to reduce the neutron energy to a final energy in a given distance, the diffusion length of thermal neutrons L_d (the distance that thermal neutrons travel until they are absorbed), and the migration length L_m of the neutrons from their original state as primary neutrons to the point of their absorption in the medium. These distance variables are related by the equation: $L_m^2 = L_s^2 + L_d^2$. Experimental data show that the values of L_s and L_d depend mainly on the porosity of the rocks. This is because there is a certain degree of saturation of the pore space by water (or hydrocarbons when applied to well logging).

A neutron moisture gauge consists of a neutron source and one or two neutron detectors. The instruments are normally calibrated in terms of porosity or water content. The shape of the calibration curve relating detector count rate to porosity or water content is influenced by source–detector distance/geometry. There are two basic approaches to measuring the porosity: the short-spaced (source–detector) sonde used for soil moisture determinations and the long-spaced sonde used for borehole logging.

8 THE MONTE CARLO METHOD

The Monte Carlo method is a numerical procedure based on statistical theory. It is used to calculate the distribution of γ-rays or neutrons produced by one or more γ-ray or neutron sources in a medium of known chemical composition. Particle fluxes and response functions (reaction rates) in selected volume elements can be calculated with fair accuracies. The technique is useful in the design of nucleonic gauges to predict the optimum source–detector–shielding configuration. The gauge response for a particular configuration can be tested in a short time at low cost on PCs using Monte Carlo calculations. Thus, tedious, time-consuming and expensive laboratory measurements can be bypassed.

The concept of the Monte Carlo technique is tracking the transport through matter of neutrons or γ-rays by simulating the statistical nature of the interaction processes. This is based on the fact that the macroscopic cross-sections may be interpreted as a probability of interaction per unit distance traveled by the neutron or γ-ray. In Monte Carlo calculations, a set of neutrons/γ-rays histories is generated by following individual neutrons/γ-rays through successive collisions. The locations and results of collisions are determined from the range of possibilities by sets of random numbers. A large library of experimental nuclear data is used in the program to select the random numbers that influence the particle history. These include the direction of the first flight on leaving the source, the choice of interaction from those possible at each collision, the mean free path between collisions and the angle of emergence after collision. Tracking the particle finishes if the particle is either absorbed, passes a lower energy limit set by the user or crosses an outer boundary. Then the history of a new particle starts. In order to decrease the statistical errors of predictions obtained from Monte Carlo calculations, the tracking of many particle histories is needed. This is becoming more and more affordable, in both time and cost, with the advent of the new generation of powerful PCs. Codes for Monte Carlo calculations on PCs are readily available and

the technique has become a powerful tool in nucleonic gauge design and nuclear applications.

9 NUCLEAR TECHNIQUES FOR IN SITU ANALYSIS

Nuclear logging is well established and used routinely in the oil, gas, uranium and coal industries. It is essential for the oil and gas industries, in which very deep holes are drilled (thousands of meters). Owing to the deep penetration of γ-radiation and neutrons, nuclear logging can locate the presence of oil or gas behind the well casing.

The laboratory analysis of core samples retrieved from boreholes and nuclear logging are complementary. Although the core can provide all the information that can be extracted from a borehole, nuclear logging is able to provide information almost instantaneously. The volume of rock sampled by nuclear borehole logging is also much larger than the core samples and thus provides better sampling statistics, especially in heterogeneous deposits. There are a number of books published in this field: Hallenburg;[9] Hearst and Nelson;[10] Tittman;[11] Ellis.[12] There are other books dealing with more specific aspects of nuclear logging. More recently, review papers have been published by Ellis et al.,[13] Mills et al.,[14] and Borsaru.[15]

Nuclear borehole logging techniques are either passive (natural γ) or active. In passive logging, the natural radiation in the borehole is measured by an appropriate detector, whereas in active logging an artificial radioactive source provides the radiation measured by the detector. Figure 4 shows a schematic diagram of a basic borehole logging system for nuclear logging. The borehole logging tool shown in the figure is fitted with one radiation detector. The source can be either a γ-ray source or a neutron source. In the absence of the source the tool is used for passive logging. Nuclear logging can be classified according to the radioactive source employed in the logging tool.

9.1 Passive Logging – Natural γ-Ray

The *γ-ray spectral log* technique provides a spectrometric measurement of the γ-radiation of rocks whereas the γ-log measures the intensity of the total natural γ-radiation

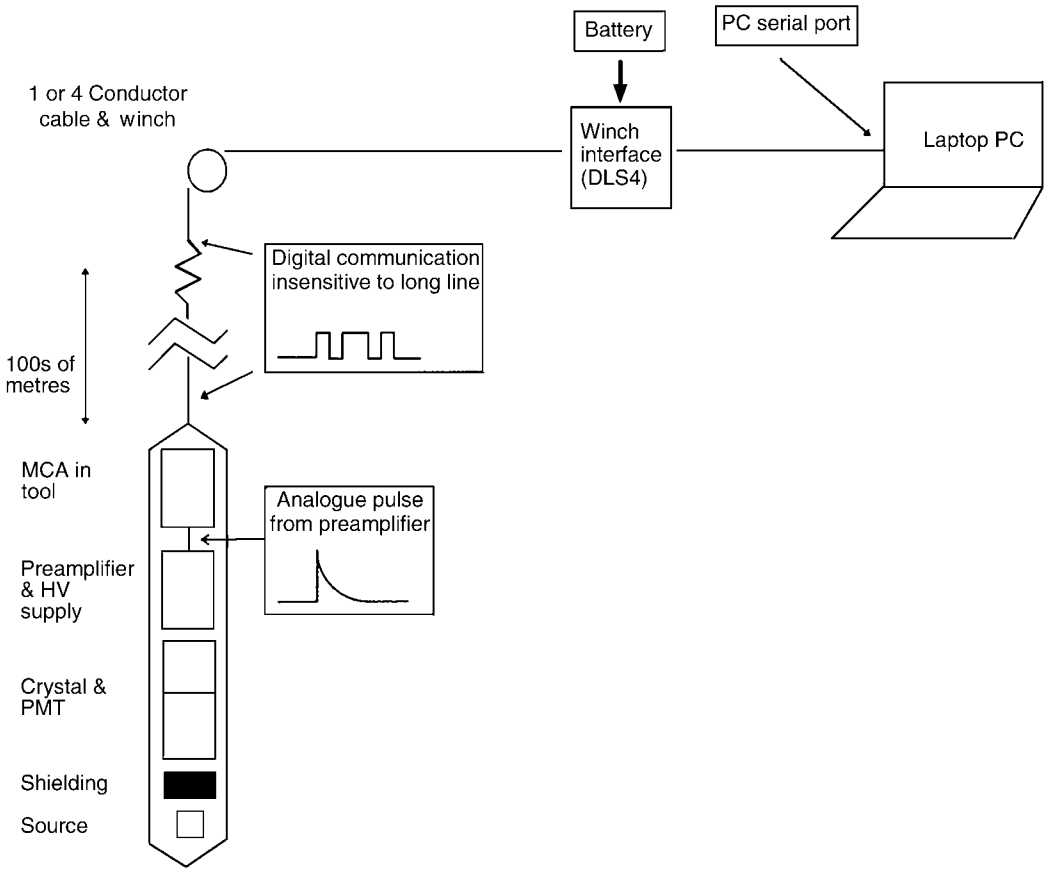

Figure 4 Schematic representation of a nuclear logging system.

of the rock. These techniques have widespread applications in petroleum and uranium exploration, potash, phosphate, coal, heavy minerals, iron ore and other mineralizations. In the γ-ray spectral log, the potassium, uranium and thorium components of the natural radiation are determined and the K, U, Th, U/K, Th/K and U/Th measurements are used for borehole lithology determinations and correlations. A review of γ-ray spectral logging has been written by Fertl.[16]

9.2 Active Logging

The backscattered γ-ray technique was developed for measuring the rock density, providing information on the type of rock encountered by the borehole, and in special cases for grade control. Figure 5 shows schematically a backscattered γ-ray spectrum obtained with a ^{60}Co γ-ray source. The backscattered spectrum is dependent on the intensity and energy of the primary γ-ray source employed in the logging tool and the type of rock intersected by the borehole. Czubek[1] reviewed the theory of γ–γ logging. The theory shows that the spectral intensity at high energy (>180 keV) is determined by the electron density, and at low energies by both the electron density and Z_{eq} (see Equations 13 and 14). The count rate in the density region is related to the bulk density of the rock interrogated by the γ-radiation. The physical basis for the density measurement is that the Compton cross-section per electron is essentially independent of the atom in which the electron is bound, and in this energy region the only interaction of consequence is Compton scattering. The electron density is directly proportional to bulk density. The ratio of counts recorded in energy windows at high and low energies is proportional to Z_{eq}. Since Z_{eq} is strongly dependent on the element of highest atomic number, e.g. Fe in iron ore, even a small variation of that element in a low-atomic-number matrix will cause significant variation in Z_{eq}. This is the principle of determining Fe grade in iron ore, or ash in coal, by the γ–γ technique. Mathew and Anderson[17] have shown that the ratio of count rates in two energy windows at high energies (>500 keV) is proportional to the borehole diameter. In order to extract the most information from the backscattered γ-ray spectrum, the γ–γ logging must be spectrometric and automatic gain stabilization must be incorporated in the system.

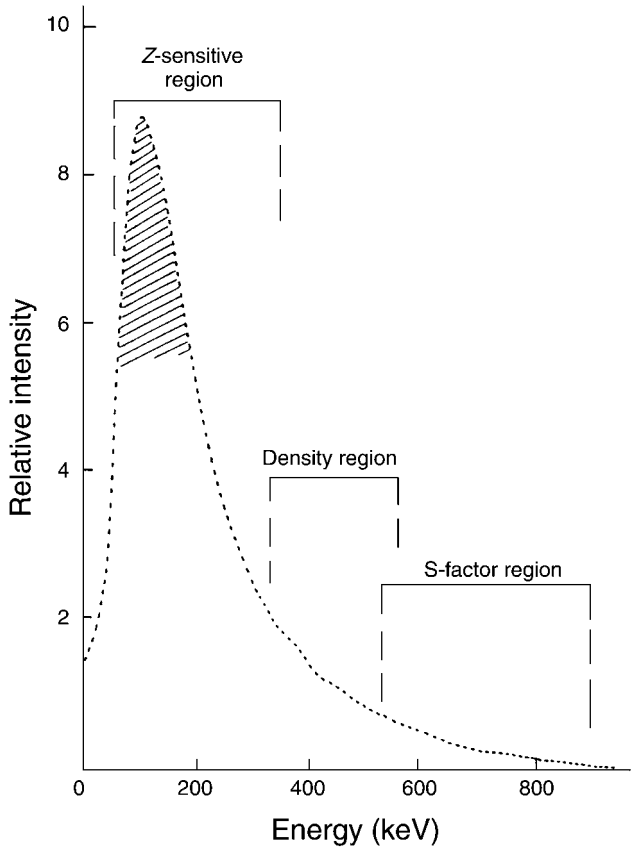

Figure 5 Typical γ–γ spectrum obtained with a ^{60}Co γ-ray source. (Reproduced from 'Use of Backscattered Gamma Radiation for Determining Grade of Iron Ores in Blast Holes and Development Drill Holes', *Proceedings of the Australasian Institute of Mining and Metallurgy*, (1977). Reprinted with permission of the Australasian Institute of Mining and Metallurgy.)

Prompt γ neutron activation, neutron inelastic scattering and *neutron activation* are useful techniques in borehole logging. They require the use of a neutron source. The type of technique chosen depends on the particular application, the nuclei involved and the energy of the neutrons. When the neutron source is a neutron generator, both neutron inelastic scattering and prompt γ neutron activation can be performed simultaneously with the neutron generator operating in a pulsed mode. This technique is largely used in oil well logging.

9.3 Oil Well Logging

9.3.1 γ-Ray Source

The bulk density of rock formations is an important parameter in oil well logging. The new generation density-logging tool, the litho-density tool, measures both the formation's bulk density and its photoelectric absorption, using a ^{137}Cs γ-ray source. This is carried out by recording the count rates in the Z-sensitive and density regions of the backscattered γ-ray spectrum shown in Figure 5. The photoelectric absorption P_e gives a measure of the average atomic number of the formation, which provides an indication of the lithology (sandstone, limestone, dolomite, etc.). The tool has two detectors at different spacings from the γ-ray source and has gain stabilization. Porosity can also be determined from the bulk density and P_e.

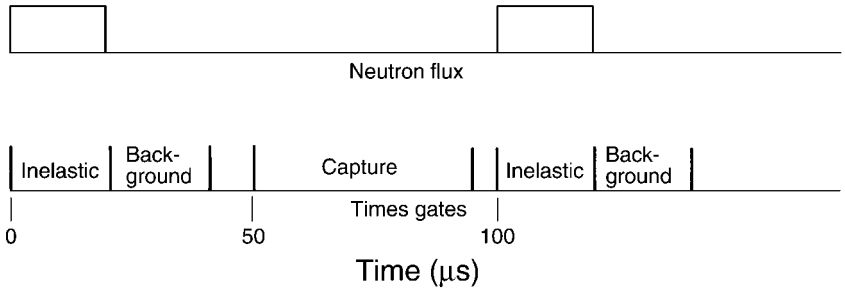

Figure 6 Timing of gates for a logging tool using a neutron generator.

9.3.2 Neutron Generator

The neutron generator is well established in oil well logging. Owing to the high energy of the neutrons produced by the generator (14 MeV), the depth of investigation is greater than that for radioactive neutron sources. It can also excite many nuclear reactions that cannot be produced by steady-state sources. However, probably the most important feature of the neutron generator is that it can be pulsed, and this allows the development of special techniques that cannot be achieved with radioactive sources.

Neutron Die-away or Neutron Lifetime This is an old technique developed for the determination of the chlorine content of formation fluid. A pulsed neutron source is needed and the neutron generator is an ideal one. The basis of this technique is to measure the change in neutron flux with time after a neutron burst. The fast neutrons emitted during the short neutron burst are slowed down to thermal energy and diffuse in the formation. They die when absorbed in a capture process by nuclei from the formation, and capture γ-rays are released. The number of capture γ-rays is proportional to the number of thermal neutrons present in the formation. The determination of neutron die-away is based on measuring the ratio of γ-ray count rates at two different times after the fast neutrons produced by the burst have decayed to thermal neutrons.[18] The die-away log provides information to distinguish oil (low-capture cross-section with resulting slow γ-ray die-away rate) from saline water (high-capture cross-section and fast die-away rate) behind the steel casing in wells.

γ-ray Spectroscopy Using a Neutron Generator Using a neutron generator in a pulsed mode makes it possible to separate the γ-rays released by neutron inelastic scattering from the γ-rays generated by the neutron capture process. Inelastic scattering γ-rays are emitted within a few nanoseconds of the fast neutron interaction and occurs during the neutron pulse. Capture γ-rays are produced after tens of microseconds by thermal neutrons traveling through the formation. Figure 6 shows schematically the gate timing of this tool. The tool is used to measure the carbon/oxygen and calcium/silicon ratios. By measuring both ratios, one can distinguish carbon in calcium carbonates from that in hydrocarbons and thus the presence of oil can be established. The energy of γ-rays released from neutron inelastic scattering with C, O, Ca and Si are 4.44, 6.1, 3.7 and 1.78 MeV respectively. A background spectrum is recorded in a time window of the same length, but with the generator turned off (see Figure 6). This is a capture spectrum and is subtracted from the inelastic spectrum.[19]

Schlumberger developed the Geochemical Logging Tool (GLT™). This tool makes the following spectral measurements: natural γ-rays produced by K, Th and U; Al by neutron activation with a ^{252}Cf neutron source; and elements in the rock formation measured by capture γ-ray spectrometry using a pulsed neutron generator.[20] The tool incorporates three scintillation detectors, a ^{252}Cf neutron source and a neutron generator. Grau and Schweitzer[21] developed a technique capable of measuring the absolute concentration of a number of elements present in the formation. The technique assumes that the composite spectrum measured in boreholes can be represented by a linear combination of known standard spectra. The weighted least-squares method of fitting the entire spectrum with the set of standard spectra extracts the chemical composition. The measured spectrum is decomposed into elemental standards. Carbon and oxygen are not measured by the GLT™ but are accounted for by the chemical combination of each element measured in the capture spectrum as a single oxide or carbonate. The tool is used to derive mineralogy and petrophysical parameters from the elemental concentrations.

9.3.3 Steady-state Neutron Sources

Am–Be is used for porosity measurements. Single-detector tools employing mostly epithermal neutron detection or two-detector neutron sondes employing thermal neutron detection are commonly used in oil well logging for porosity measurements.

9.3.4 Natural γ-Ray Spectral Logging

Natural γ-ray spectral logging provides information about the presence of clays in the formation matrix. The tendency is to replace the natural γ-ray tools recording the count rates in the three K/U/Th windows with fully spectrometric tools, which record the whole spectrum in 256 channels.

9.3.5 Measurements While Drilling

Measurements while drilling are used to gather information on the formation being drilled in real time. The detectors are mounted in the drilling string near the bit. The detectors are either scintillators or Geiger–Mueller tubes that have been ruggedized to withstand the drilling process. γ-Ray, neutron porosity and density measurement-while-drilling (MWD) tools have been developed and are widely used.

9.4 Uranium

9.4.1 Natural γ-Ray Logging

γ-Rays emitted from the uranium minerals are routinely used for both quantitative and qualitative determination of uranium in exploration or ore-body evaluation. The γ-rays are ideal to use as intrinsic indicators for uranium as they occur naturally, sample relatively large volumes of material and are simple to detect and identify. The decay series of ^{238}U (excluding some short-lived isotopes) is:

$$^{238}U \xrightarrow{4.5\times10^{9}\,a} {}^{234}U \xrightarrow{2.5\times10^{5}\,a} {}^{230}Th \xrightarrow{8\times10^{4}\,a} {}^{226}Ra \xrightarrow{1620\,a} {}^{222}Rn \xrightarrow{3.8\,d} \ldots {}^{214}Pb \xrightarrow{27\,min} {}^{210}Pb \xrightarrow{20\,a} {}^{206}Pb(\text{stable})$$

When secular equilibrium is reached after some 10^6 years, the activity of each isotope, except ^{206}Pb, in the decay series is equal. Disequilibrium can occur throughout the chain if one or more of the daughter products is lost by any process other than radioactive decay. Since each daughter product is an element with its own characteristic physical and chemical properties, it may behave differently within a given environment. For example, ^{222}Rn is a gas. Also, the solubilities of radium, uranium and thorium differ, and preferential leaching of elements may occur.

Total γ-ray Logging This is probably the easiest logging technique for quantitative analysis. The total number of γ-rays can be recorded with either a Geiger counter or a scintillator. The logging probes are calibrated in special uranium test pits. Since this method records γ-rays from many daughters of uranium, it can only be applied for quantitative analysis if both the test pit and uranium ore body are in secular equilibrium.

Accurate Spectral γ-ray Logging The principle of this technique for the in situ quantitative determination of uranium[22] is measurement of the γ-radiation (1001 keV) produced during the decay of protactinium-234 (^{234}Pa) to uranium-234. In the uranium decay series, ^{238}U first decays to ^{234}Th, which decays to ^{234}Pa, which in turn decays to ^{234}U. Because the half-lives of all three reactions are short, it is unlikely that in nature the ^{238}U and ^{234}Pa will become separated. Thus, the determination of the ^{234}Pa grade from the count rate of the 1001 keV γ-ray gives a measure of the ^{238}U grade in the ore body. This method is independent of the amount of disequilibrium in the ore body. A high resolution (intrinsic Ge detector) is necessary to measure the count rate of the 1001 keV, which is weak in comparison with other γ-ray lines in the natural γ-spectrum.

9.4.2 Delayed Fission Neutrons

Fission occurs when uranium nuclei interact with neutrons and the nuclei split into fission products, emitting neutrons. Two to three percent of the fission products are unstable and β-decay, forming other nuclei. By doing this, each fission product emits one neutron. These are called delayed neutrons and are emitted with different half-lives, between 0.2 and 50 s. This is the basis of the delayed fission neutron technique for quantitative analysis of uranium and is independent of the state of equilibrium of the uranium ore body. The technique consists of irradiating the uranium ore with neutrons, removing the neutron source and counting the delayed neutrons after the source neutrons disappear.

The neutron generator is well suited for this application.[23] The generator is pulsed and then turned off. When the neutrons produced by the generator die away, the delayed neutrons are counted. The effect of diffusion of the neutrons in the medium must be accounted for. Background subtraction is achieved using the reaction $^{17}O(n,p)^{17}N \longrightarrow {}^{16}O + n$.

^{252}Cf is also used as a source of neutrons[24] by shuttling the source inside the moving probe between two detectors; while one detector is counting the delayed neutrons, the source is irradiating the ore body. Owing to the much lower energy of the neutrons produced by ^{252}Cf, no background subtraction is needed (the energy of the neutrons is below the threshold energy of the $^{17}O(n,p)$ reaction). Correction for diffusion of the neutrons in the uranium ore is needed.

9.4.3 Prompt Fission Neutrons

The neutrons emitted immediately after fission can also be used to measure the uranium concentration.[25,26] The

technique uses a neutron generator as the primary source of neutrons. An epithermal neutron detector is gated off during a neutron pulse and starts counting epithermal neutrons immediately after the pulse, while thermal neutrons are present and causing fission. The uranium concentration is determined from the ratio of epithermal neutrons to (i) thermal neutrons or (ii) thermal neutron die-away time.

9.5 Coal

Nuclear borehole logging is used routinely in the exploration for coal and its acceptance in the mining stage for coal quality control is increasing. The parameters determining the economic value of coal are the thickness and depth of the coal seams and the quality of coal. Coal quality is determined by the ''raw ash'' content, calorific value, rank, moisture content, volatiles and the elemental composition of the ash. These parameters are determined from the coal core in the laboratory. However, this complete information is not always needed. In many cases it is sufficient to measure only thickness, depth and the raw ash content of the coal seams, and this can be achieved by nuclear logging in ''easier to drill open holes''. Natural-γ, $\gamma-\gamma$ and prompt neutron$-\gamma$ techniques are all used for coal logging. Natural-γ is used to delineate the coal seams based on the fact that coal is low in natural radiation compared with the interseam sediments.

The $\gamma-\gamma$ Technique This technique was developed for the coal mining industry in the 1970s.[27,28] It is used mostly for the delineation of coal seams based on the difference in density between coal and sediments. The ash content of the coal seams can be derived indirectly from the correlation between ash and density. Good density–ash correlations are found for many coal deposits. However, this correlation is not universal. A direct measurement of ash in coal is achieved using the spectrometric $\gamma-\gamma$ technique.[5] This technique relies on the fact that variations in the ash content of coal correspond to changes in Z_{eq}. When the ash content and Z_{eq} are uniquely correlated, ash content can be measured by measuring changes in Z_{eq}. This is carried out by calculating the P_z ratio between count rates recorded in two broad spectral windows at high and low energies in the backscattered γ-ray spectrum. This measurement requires gain stabilization.

The Prompt Neutron$-\gamma$ Method for Determination of Ash in Coal Owing to its high hydrogen content, coal is an excellent medium for the neutron capture technique. The neutrons emitted by the neutron source are thermalized by colliding with the hydrogen nuclei present in coal and subsequently interact with the nuclei in the coal matrix. The basis for ash determination is the correlation that exists between ash and the main constituents of ash, i.e. Al, Si and Fe or a combination of two of these elements. Charbucinski et al.[29] developed a method based on the neutron$-\gamma$ technique for the determination of ash content, depth and thickness of coal strata. The technique was later extended for the measurement of Fe in coal.[30]

Coal Face Ash Analyzer Coal ash determination at the coal face falls into the category of in situ determination. A portable instrument capable of monitoring the quality of exposed coal seams and differentiating between coal seams and ''look-alike'' coal sediments is useful for the coalmining industry. Wesolinski and de Jesus[31] developed a technique for the determination of ash on the coal face based on the simultaneous measurement of the backscattered γ-radiation at two different energies. They developed a portable instrument that can locate the coal–seam interface and measure the ash content on the coal face. The activity of the γ-ray sources employed in the instrument was 185 MBq. Borsaru et al.[32] developed a coal face analyzer based on the measurement of natural γ-rays. The technique relies on the correlation between the natural γ-radiation of the coal and its ash content. Another coal face analyzer based on the backscattered $\gamma-\gamma$ technique was developed more recently by Borsaru et al.[33] It uses two microsources (^{133}Ba and ^{137}Cs) of total activity 2.2 MBq. By employing such low-activity γ-ray sources, the analyzer does not need special shielding and the user is not exposed to unacceptable levels of radiation.

9.6 Iron Ore

The backscattered $\gamma-\gamma$ technique has been employed for the simultaneous measurement of iron grade, density and borehole diameter in wide (310 and 380 mm) dry blast holes in iron ore deposits.[4] The primary γ-ray source was 165 MBq ^{60}Co. The backscattered γ-ray spectrum was stabilized using a ^{60}Co microsource. The borehole diameter was measured using the ratio (S-factor) between count rates in energy windows at intermediate energies (400–700 keV) and at high energies (700–950 keV).[17] The source-to-detector separation was 70 cm. The iron grade determination was based on the P_z technique described earlier. The accuracy for ore grade obtained in 1.5 m splits was 1.9% Fe in ore in the range from 35% to 69% Fe and 3% Fe for ores in the range of iron content 9–69% Fe.

ABBREVIATIONS AND ACRONYMS

MWD Measurement-while-drilling

RELATED ARTICLES

Steel and Related Materials **(Volume 10)**
Iron Ore, Sample Preparation and Analysis of

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Photon Activation Analysis • Prompt γ-Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

APPENDIX 1

The Compton Effect

The following expressions (Equations 16–18) can be derived for the energy $h\nu'$ of the scattered photon, the energy T of the scattered electron and the relationship between the angles of scattering shown in Figure 1:

$$h\nu' = \frac{h\nu}{1+\alpha(1-\cos\theta)} \tag{16}$$

$$T = h\nu\frac{\alpha(1-\cos\theta)}{1+\alpha(1-\cos\theta)} \tag{17}$$

$$ctg\phi = (1+\alpha)tg\tfrac{1}{2}\theta \tag{18}$$

where $\alpha = h\nu/m_0c^2$ and m_0 is the rest energy of the electron. According to Equation (16), the energy after scattering is not independent of the energy before scattering. The mean relative energy loss decreases with decreasing γ-ray energy.[1] Equation (16) shows that the energy of the scattered γ-ray is a function of the angle of scattering, and the maximum energy that an electron can acquire is less than the total energy of the γ-ray: it is, at most, 225 keV.

Elastic Scattering of Fast Neutrons

The cross-section for elastic scattering for fast neutrons[34], as shown in Equation (19), is approximately:

$$\sigma_{el} \approx \pi\left(R + \frac{\lambda}{2\pi}\right)^2 \tag{19}$$

where $R_0 = 1.4 \times 10^{-13}$ cm and $R = R_0A^{1/3}$ is approximately equal to the radius of the nucleus with mass number A. $\lambda/2\pi$ is the wavelength equal to the radius of the neutron and is equal to $4.44 \times 10^{-13}/E_n$, where E_n is the neutron energy in MeV. When a neutron interacts elastically with a nucleus, the relation between the neutron energy before and after scattering, E_1 and E_2 respectively, is illustrated in Equation (20) as:

$$\frac{E_2}{E_1} = \frac{A^2 + 2A\cos\theta + 1}{(A+1)^2} \tag{20}$$

where A is the mass of the target nucleus and θ is the scattering angle in the center-of-mass system. Equation (20) reveals four features of elastic scattering: the fractional energy loss depends only on the mass ratio and scattering angle and is independent of energy. The minimum possible value $E_{2\,min}$ of E_2 (corresponding to $\theta = 180^\circ$) is $E_{2\,min} = [(A-1)/(A+1)]^2$ and is the result of a head-on collision. For $\theta = 0$, a glancing collision, $E_2 = E_1$, and no energy loss occurs. When $A = 1$, corresponding to collision with a hydrogen nucleus, for $\theta = 180^\circ$ (head-on collision), $E_2 = 0$ and all the neutron's kinetic energy is transferred to the recoiling proton.

The percentage decrease in the neutron energy as the result of a collision, expressed as the *average logarithmic energy decrement* ξ, is known as neutron *lethargy* and is defined in Equation (21) as:

$$\xi = \ln\left(\frac{\overline{E_1}}{\overline{E_2}}\right) = \frac{\int_{E_1}^{E_{2\,min}} [\ln(E_1/E_2)]\,P(E_2)\,dE_2}{\int_{E_1}^{E_{2\,min}} P(E_2)\,dE_2} \tag{21}$$

where $P(E_2)$ is the probability that the scattered neutron will have energy E_2. When the integrations are performed, it is found that the mean lethargy increase per collision, ξ, illustrated in Equation (22) is:

$$\xi = 1 + \frac{(A-1)^2}{2A}\ln\frac{A-1}{A+1} \tag{22}$$

REFERENCES

1. J.A. Czubek, 'Advances in Gamma-gamma Logging', in *Nuclear Geophysics, Selected Papers on Applications of Nuclear Techniques in Mineral Exploration, Mining and Process Control*, ed. C.G. Clayton, Pergamon Press, Exeter, 153–172, 1983.
2. O. Klein, Y.Z. Nishina, 'Über die streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac', *Z. Physik*, **52**, 853 (1929).
3. C.M. Davisson, 'Interaction of γ-radiation with Matter', in *Alpha-, Beta- and Gamma-ray Spectroscopy*, ed. K. Siegbahn, North-Holland, Amsterdam, 33–78, Vol. 1, 1965.
4. P.L. Eisler, J. Charbucinski, M. Borsaru, A.P. Indrans, 'Spectrometric Gamma-gamma Borehole Methods for

Iron Ore Mining', Proceedings of the Second International Symposium on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications, 6–8 October, Golden, Colorado, USA, 189–201, 1987.
5. M. Borsaru, J. Charbucinski, P.L. Eisler, S.F. Youl, 'Determination of Ash Content in Coal by Borehole Logging in Dry Boreholes using Gamma-gamma Methods', *Geoexploration*, **23**, 503–518 (1985).
6. P.J. Mathew, 'On-stream Coal Ash Analysis Based on Natural Gamma Ray Activity', Nuclear Techniques in the Exploration and Exploitation of Energy and Mineral Resources, IAEA-SM-308/89, 93–103, 1991.
7. P.J. Mathew, 'A Natural γ-Ray Gauge for Green Liquor and Red Mud', Proceedings of the International Conference on Applications of Radioisotopes and Radiation in Industrial Development, 1–3 March, Bombay, 1984, eds. S.M. Rao, A.B. Majali, R.G. Deshpande, T.S. Murthy, Wiley Eastern, Bombay, 154–163, 1984.
8. S. Glasstone, M.C. Edlund, *The Elements of Nuclear Reactor Theory*, Van Nostrand, Princeton, NJ, 1952.
9. J.K. Hallenburg, *Geophysical Logging for Mineral and Engineering Applications*, Pennwell Books, Penwell Publishing, Tulsa, OK, 1984.
10. J.R. Hearst, P.H. Nelson, *Well Logging for Physical Properties*, McGraw-Hill, New York, 1985.
11. J. Tittman, *Geophysical Well Logging*, Academic Press, Orlando, FL, 1986.
12. D. Ellis, *Well Logging for Earth Scientists*, Elsevier, New York, 1987.
13. D.V. Ellis, J.S. Schweitzer, J.J. Ullo, 'Nuclear Techniques for Subsurface Geology', *Ann. Rev. Nucl. Part. Sci.*, **37**, 213–241 (1987).
14. W.R. Mills, D.C. Stromswold, L.S. Allen, 'Advances in Nuclear Oil Logging', *Nucl. Geophys.*, **5**, 209–227 (1991).
15. M. Borsaru, 'Nuclear Techniques for In Situ Evaluation of Coal and Mineral Deposits', *Nucl. Geophys.*, **7**, 555–574 (1993).
16. W.H. Fertl, 'Gamma-ray Spectral Logging: A New Evaluation Frontier', *World Oil*, **196**(4), 79–91, **196**(5), 87–98, **196**(6), 147–155, **196**(7), 189–204, **197**(2), 57–63, **197**(4), 99–112 (1983).
17. P.J. Mathew, M.R. Anderson, 'A Gamma-gamma Method for Measuring the Diameters of Air-filled Boreholes', Borehole Geophysics for Mining and Geotechnical Applications, ed. P.G. Killeen, paper 85-27, 269–275, Geological Survey of Canada, Canada, 1986.
18. A.A. Youmans, E.C. Hopkinson, R.A. Bergen, H.L. Oshry, 'Neutron Lifetime: A New Nuclear Log', *J. Petr. Tech.*, **16**, 319 (1964).
19. W.E. Schultz, H.D. Smith, 'Laboratory and Field Evaluation of a Carbon/Oxygen (C/O) Well Logging System', *J. Petr. Tech.*, **26**, 1103–1110 (1974).
20. R. Hertzog, L. Colson, B. Seeman, M. O'Brien, H. Scott, D. McKeon, P. Wraight, J. Grau, D. Ellis, J. Schweitzer, M. Herron, 'Geochemical Logging with Spectrometry Tools', *SPE Form. Eval.*, **4**, 153–162 (1989).
21. J.A. Grau, J.S. Schweitzer, 'Elemental Concentrations from Thermal Neutro Capture Gamma-ray Spectra in Geological Formations', *Nucl. Geophys.*, **3**, 1–9 (1989).
22. B. Tanner, R.M. Moxham, F.E. Senftle, 'Assay for Uranium and Determination of Disequilibrium by Means of In Situ High-resolution γ-Ray Spectrometry', US Geol. Survey Open-File Rept., USA, 77–571, 1977.
23. W.W. Givens, W.R. Mills, C.L. Dennis, R.L. Caldwell, 'Uranium Assay Logging using a Pulsed 14-MeV Neutron Source and Detection of Delayed Fission Neutrons', *Geophysics*, **41**, 468 (1976).
24. D.K. Steinman, J. Stokes, J.A. Adams, R. Voigt, C.S. Pepper, K.V. Laubscher, W.J. Smith, J. John, '252Cf-based Delayed-neutron Logging System for In Situ Assay of Uranium Ore', Trans. Soc. Prof. Well Log Analysts 21st Logging Symp., paper I, 1980.
25. R.W. Barnard, 'Interpretation of Prompt Fission Neutron Uranium Borehole Logs', *Trans. Am. Nuc. Soc.*, **28**, 107 (1978).
26. D.R. Humphreys, R.W. Barnard, H.M. Bivens, D.H. Jensen, W.A. Stephenson, J.H. Weinlein, 'Uranium Logging with Prompt Fission Neutrons', *Int. J. Appl. Rad. Isotopes*, **34**, 231 (1983).
27. J.R. Samworth, 'The Radiation Density Log Applied to the Resolution of Thin Beds in Coal Measures', SPWLA 3rd European Symposium, London, paper R, 1974.
28. D.R. Reeves, 'Application of Wireline Logging Techniques to Coal Exploration', Proceedings of the First International Coal Exploration Symposium, London, 18–21 May, ed. W.L.G. Muir, 112–128, 1976.
29. J. Charbucinski, S.F. Youl, P.L. Eisler, M. Borsaru, 'Prompt Neutron-gamma Logging for Coal Ash in Water-filled Boreholes', *Geophysics*, **51**, 1110–1118 (1986).
30. M. Borsaru, M.S. Biggs, W.J.F. Nichols, 'Neutron-gamma Logging for Iron in Coal and Implications for Estimating the Ash Fusion Characteristics at Callide Mine', *Nucl. Geophys.*, **7**, 539–545 (1993).
31. E.S. Wesolinski, A.S.M. de Jesus, *A Portable Coal Face Ash Monitor Based on Dual Energy Gamma Radiation*, Nuclear Techniques in the Exploration and Exploitation of Energy and Mineral Resources, IAEA, Vienna, 33–46, 1991.
32. M. Borsaru, C. Ceravolo, P. Waddington, Gu Wenhao, 'A Coal Face Ash Analyzer Based on Natural Gamma-ray Activity', *Nucl. Geophys.*, **6**, 383–390 (1992).
33. M. Borsaru, C. Ceravolo, G. Carson, T. Tchen, 'Low Radioactivity Portable Coal Face Ash Analyzer', *Appl. Radiat. Isot.*, **48**, 715–720 (1997).
34. R.D. Evans, *The Atomic Nucleus*, McGraw-Hill, New York, 1955.

Radiochemical Methods: Introduction

Rolf J. Rosenberg
VTT Chemical Technology, Espoo, Finland

1 Introduction 1
2 Properties of Radionuclides 1
2.1 Nuclides and Radioactive Decay 1
2.2 Decay Modes 2
3 Medical Use of Radionuclides 3
4 Industrial Use of Radionuclides 3
5 Use of Radionuclides in Research 3
6 Origin of Radionuclides in the Environment 3
7 Measurement of Radionuclides 4
Abbreviations and Acronyms 5
References 5

1 INTRODUCTION

Roentgen announced the discovery of X-rays in 1895. This led to extensive use and study of this phenomena, which had already been studied since 1859 without the realization of its true nature. By 1897, a number of injuries caused by X-rays had already been reported from laboratories all over the world.[1] In the same year Bequerel discovered radioactivity and soon after, in 1900, the first skin burn caused by radioactivity concentrated from uranium ore, was produced. The Curies learned to produce radium, which found use in luminous paints in instruments and clocks during and after World War I. The health effects of radium were unknown to the public and workers involved in painting licked the brushes to point them. This digestion of radium, and the use of radium for medical purposes, led to many deaths from bone cancer and aplastic anemia. As late as 1944–1951 a radium compound was used as a cure for tuberculosis in Germany. It has been estimated that more than 100 persons died from the effects of radium. Another cause of a number of deaths from cancer caused by radioactivity is radon and its daughters. In Eastern Europe many mines for heavy metals also contain uranium. Radon and its daughters that are incorporated into uranium caused a fatal lung disease, which was later identified as cancer. Information about this did not reach miners in the USA. Therefore the intense uranium mining that took place in the USA, starting after World War II, also caused numerous fatalities from lung cancer.

These cases have been studied thoroughly leading to an understanding of the effects of radiation and how to avoid it. Thus, early experiences with the use of nuclear reactors and in building atomic bombs during World War II could be managed with a good safety record. The health effects of the atomic bombs in Hiroshima and Nagasaki have also been thoroughly investigated, yielding a great deal of information. As a consequence the health effects of ionizing radiation are well known. Based on this knowledge, international safety regulations have been set on allowed doses, allowed radionuclide concentrations in water, food and air and so on. It can safely be said that the regulations are such that human beings are much better protected against radioactivity than against chemical toxicity.

After World War II the cold war resulted in the nuclear weapons race. Both the USA and the Soviet Union conducted nuclear weapons tests above ground. The fallout of fission and activation products from these tests spread over the northern hemisphere. This started research in all countries, in which the environmental pollution and transport of these radionuclides were studied. Much work was done in the development of analytical methods for the determination of radionuclides. The nuclear industry, nuclear power plants and reprocessing plants during their normal operation caused some additional release of radionuclides. Major releases were caused by accidents in the Soviet Union, USA and UK. New threats have appeared with the disintegration of the Soviet Union. Lack of funds has increased the risks of accidental releases from nuclear submarines, reprocessing plants and nuclear power plants.

All these are good reasons for continuous monitoring of radionuclides in the environment. New reasons for measurement of low activities and long-lived radionuclides have developed in connection with research on the disposal of spent fuel and reprocessing waste. As a consequence, new methods of making these measurements are continuously being developed. In addition to the classical radiometric methods, new mass spectrometric methods are published.

2 PROPERTIES OF RADIONUCLIDES

2.1 Nuclides and Radioactive Decay

An element is identified by the number of positively charged protons in its nucleus and the consequent number of negatively charged electrons.[2] In addition to protons,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the nucleus contains neutrons, which do not carry any charge. Neutrons and protons have an atomic weight of unity. The mass of electrons is negligible compared with the mass of protons or neutrons. All elements consist of isotopes. Each isotope of an element is identified by the number of neutrons. Thus all isotopes of a specific element contain the same number of protons, which is determined by the atomic number, Z. The mass (M) of the isotope is the sum of the number of protons and neutrons in the nucleus. A specific combination of protons and neutrons is called a nuclide. A nuclide is identified by its Z and M values. The chemical symbol gives the Z and the mass is given as a superscript on the left side of the chemical symbol. ^{238}U is uranium with mass 238; it is called uranium-238. ^{3}H is called hydrogen-3, or tritium. Most elements, but not all of them, have several stable isotopes, in a ratio that is constant. Thus nuclides with the same Z are called isotopes, nuclides with the same mass are called isobars and nuclides with the same number of neutrons are called isotones.

In addition to the stable isotopes, all elements have radioactive isotopes with variable half-lives. A radionuclide is unstable, which means that the nucleus decays releasing energy and particles. This is seen as emission of radiation of various kinds. Part of the energy is released in the form of electromagnetic radiation, like γ-rays, and partly as kinetic energy of emitted particles. This energetic ionizing radiation is what makes radionuclides hazardous for living organisms. The rate of decay is characteristic for each radionuclide. If there are N atoms the rate is $-\mathrm{d}N/\mathrm{d}t$, which is proportional to N according to Equation (1)

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\lambda N \tag{1}$$

Instead of the disintegration constant λ, the half-life T is commonly used. The relation is, Equation (2)

$$\lambda T = \ln 2 = 0.693 \tag{2}$$

The half-lives of the different radionuclides vary between fractions of seconds to billions of years. The energy released is expressed in units of electron volts, eV. This is the energy required to transport an electron through a potential difference of 1 V. Because nuclear binding energies are of the order of 8 MeV per nucleon, nuclear reactions always release significant amounts of energy. Therefore the kinetic energies of the emitted particles usually vary from a few kiloelectron volts to several megaelectron volts.

The amount of an element is usually expressed as a weight unit like kilogram or gram, or a concentration unit like grams per kilogram, grams per liter or such. Lately the more scientific unit, the mole, is used, in order to relate the quantity to the number of atoms or molecules. The amount of stable nuclide is usually expressed as isotopic percentage, that is the ratio of the total number of atoms of the element. The amounts of very long-lived radionuclides are also expressed in this way, for instance ^{238}U and ^{235}U. Of course the amount of both stable nuclides and radionuclides may be expressed in absolute mass units or as the number of nuclides. Particularly when the amount of a long-lived radionuclide has been determined by mass spectrometry, it is not uncommon to give the results in mass concentration units.

The mass of a radionuclide with a short half-life is very small compared with its activity. The activity also has the benefit of being proportional to its toxicity, which is the usual reason for determining the concentration. The methods most commonly used for the determination of the radionuclide are radiometric, which give the activities directly. For these reasons the amounts of radionuclides are usually expressed in units of radioactivity. The activity of a radionuclide is expressed as disintegrations per unit of time. The basic unit is the Bequerel, Bq. This is the number of disintegrations per second. Greater and smaller units are kBq, MBq, GBq and mBq.

An older unit of activity, seen in older texts and still in use in the USA, is the curie (Ci). One Ci is 3.7×10^{10} Bq. Because the unit is quite large it is common to use smaller units like mCi, μCi, nCi and pCi especially in expressing environmental activities.

2.2 Decay Modes

Radionuclides decay in different ways. Isomeric transitions (IT), that is transitions between different energy levels of the same nuclide, result in emission of energy only, normally in the form of a γ-ray, $^{60m}\mathrm{Co} \longrightarrow {}^{60}\mathrm{Co} + \gamma$. This usually occurs in connection with emission of particles, but separate IT with measurable half-lives are also known. When the transition energy is greater than 1.022 MeV, pair production may occur. This means that an electron positron (positively charged electron) pair is emitted instead of the γ-ray. There is also a third mode of IT: internal conversion in which an electron from the electron cloud carries the de-excitation energy and a conversion electron is emitted.

β-Decay is the most common form of decay of a radionuclide. Most radionuclides formed by fission or by neutron activation reactions are neutron rich, and they decay mainly by emitting a negatively charged β-particle, that is a β^--particle or radiation. In most cases, but not always, the daughter nuclide is left in an isomeric state which decays either immediately or with a measurable half-life by emitting one or several γ-rays.

Proton-rich nuclides, usually produced by γ-ray or charged particle reactions, disintegrate by positron- (β^+)

emission or by electron capture. In electron capture the charge of the nucleus is adjusted by capturing a negative electron instead of emitting a positive β-particle. In this case the only measurable radiation is the X-ray which is emitted when the captured electron is replaced by another from a higher energy level.

Heavy nuclides, $Z > 82$ with a few exceptions, may decay by emission of an α-particle, which is a ^{4}He nucleus. The α-particles usually carry much energy.

The most heavy elements, $Z > 90$, may undergo spontaneous fission in which the nucleus is divided into two fractions of comparable, but not equal, size. Two to three neutrons are also emitted, and often also γ-radiation.

3 MEDICAL USE OF RADIONUCLIDES

As has been indicated above radionuclides were used for their believed health effects early in the period following the discovery of radionuclides. One of the most important applications is the use of radiation for cancer treatment. In the beginning large ^{60}Co sources were used to treat cancer. These pose a potential environmental hazard because the ^{60}Co sources have to be disposed of in some way. ^{60}Co sources have largely been replaced by linear accelerators. The radiation doses have to be measured and controlled with high accuracy.

Different short-lived radionuclides, inserted into humans in a number of ways, are used for both treatment and diagnostics. The radionuclides are usually used as labeled compounds. The substrate on to or into which they are adhered are chosen so that they transport the radionuclides into specific organs of the human body. They can then be determined by measuring the gamma radiation which they emit. Gamma cameras enable the measurement of the radionuclide distribution in two or three dimensions. These measurements reveal tumors and other kinds of deficiencies in the human body. The radioisotopes used in medical applications are usually so short-lived that they do not pose any environmental hazards.

Radionuclides are also used as labels in radioimmunoassay, which is used for the determination of biologically active compounds.

4 INDUSTRIAL USE OF RADIONUCLIDES

The production of nuclear weapons and the use of nuclear energy have been the largest industrial-scale activities involving radionuclides. The first nuclear weapons were based on enriched uranium, but soon plutonium was used instead. The plutonium is produced by irradiating uranium in research reactors. The plutonium has to be chemically separated from the uranium and large amounts of radionuclides are produced by fission of uranium during the irradiation. The use of nuclear energy involves radioactivity during uranium mining and reprocessing of the fuel. These activities cause large amounts of radioactive waste, which have to be stored in a safe way. This has not always been successful, as will be indicated below. Uranium and plutonium and their isotopes have to be measured during the different stages of the processes.

Other industrial uses involve radioactive sources in a number of gauges. Radiography using ^{60}Co and ^{137}Cs sources is in use. In industrial plants such parameters as thickness and density are commonly measured in the process using gauges based on radioactive sources. These usually do require calibration measurements and after dismantling they need to be disposed of in a safe way.

5 USE OF RADIONUCLIDES IN RESEARCH

Radionuclides are produced in research reactors and accelerators for use in medicine, as tracers and in analytical chemistry. The medical uses have already been mentioned. Radioactive tracers are used as labels in biochemical and chemical research in order to study chemical reactions. Radioactive tracers are also used to study industrial processes, transport of water and air and wear of engines, to mention some applications.

Nuclear reactions and activation analysis are used for materials research and trace element analysis of environmental, geological, archaeological, biological and industrial materials.

6 ORIGIN OF RADIONUCLIDES IN THE ENVIRONMENT

Radionuclides occur naturally in the environment.[1] They can be found in the atmosphere, in all natural waters, in rocks, and so on. Usually the amounts are small, but sometimes quite large activities can be found. All living systems contain radionuclides. Most of them are of natural origin, but after World War II an increasing amount of radioactivity has been distributed in the environment by human activity.

Uranium and thorium are elements which exist in amounts in the microgram per gram range in rocks, particularly granitic rocks. Both are radioactive consisting mainly of the isotopes ^{235}U with a half-life of 7.13×10^8 years, ^{238}U with a half-life of 4.468×10^9 years and ^{232}Th with a half-life of 1.405×10^{10} years, sufficiently

long half-lives to enable the isotopes to have survived the billions of years in which the earth has existed. These long-lived radionuclides support a long line of daughter nuclides with variable half-lives, the uranium–actinium, uranium–radium and thorium series. These series have the gas radon, Rn, as an intermediate isotope. As an inert gas, radon emanates and is thus spread into houses and the atmosphere in areas of high uranium content in the bedrock. For instance in Finland, 75% of the radiation dose to the population from natural sources is caused by ^{222}Rn and its daughters. Because of the volatility of radon, its daughters ^{210}Pb and ^{210}Po may be found in the atmosphere. A part of the ^{210}Po is directly volatilized when coal is burned in power plants. Some coals contain significant amounts of uranium and its daughters. These three natural series all end of in stable nuclides of lead. Thus ^{206}Pb, ^{207}Pb and ^{208}Pb are at least partly of radiogenic origin, and therefore the lead isotope ratios are not constant, which is opposite to the case for most other isotope ratios. As uranium and thorium are natural constituents of rocks, they can also be found in all materials containing stone, such as the concrete walls in houses.

Another original radionuclide in the bedrock is ^{40}K, with a half-life of 1.26×10^9 years. ^{40}K is mixed with all the potassium of the earth in a constant ratio. Therefore all rocks, building materials, waters and living systems also have their share of ^{40}K.

New radionuclides are constantly produced in small amounts by cosmic ray-induced nuclear reactions in the atmosphere. More than ten different nuclides are produced. The most significant ones in respect of quantity are ^{3}H, ^{10}Be, ^{14}C, ^{26}Al, ^{26}Cl and ^{39}Ar.

Radionuclides have significant uses in medicine, industrial processes, industrial measurement systems and science. Thus radionuclides are produced artificially by different kinds of accelerators and nuclear reactors. However, most of these radionuclides have short half-lives and the contamination of the environment from these sources is insignificant. Artificial nuclides that have been incorporated into the environment mainly originate from nuclear weapons tests and the nuclear energy fuel cycle. In nuclear weapons tests and nuclear power production the main process is fission of ^{235}U (reactors) or ^{239}Pu (bombs). In both instances radionuclides are produced via two kinds of nuclear reaction. The main process is the fission itself, but significant amounts of radionuclides are also produced by activation via neutrons which are released in the fission process. During fission, all radionuclides in the mass range 80–110 and 125–155 are produced. The most significant ones in regard to safety are ^{131}I with a half-life of 8 days, ^{90}Sr with a half-life of 28.5 years and ^{137}Cs with a half-life of 30.2 years. The neutron activation reaction produces some rather short-lived nuclides like ^{54}Mn (312 days) and ^{60}Co (5.3 years), but the most significant with respect to safety are the long-lived actinides, such as ^{239}Pu (24 110 years). The nuclear weapons tests performed in the atmosphere during the 1960s severely contaminated the globe with these radionuclides. In 1986 the Chernobyl accident contaminated large areas of Europe with significant amounts of radioactivity. Starting from World War II, local contamination has occurred in several locations dealing with nuclear weapons production and reprocessing of nuclear fuel especially in the USA, UK and Russia. Although local, the radionuclide contamination from Windscale (Sellafield), UK, can be seen in the ocean up to the North Pole. Compared to these releases, releases from nuclear power plants are insignificant. Research connected with the final disposal of spent nuclear fuel has increased interest in the behavior of other long-lived nuclides produced in nuclear fuel. In addition to ^{239}Pu, nuclides like ^{99}Tc, ^{129}I and ^{237}Np have been increasingly studied.

In Europe the disintegration of the Soviet Union has introduced new needs for radionuclide measurements through threats from the disintegration of nuclear submarines in the Barents Sea and the poor maintenance of nuclear power plants.

7 MEASUREMENT OF RADIONUCLIDES

Originally radionuclides were always measured by radiometric methods, that is by measuring the radiation emitted at the decay.[3] This is convenient for several reasons. In most cases it is the simplest way of doing the measurement. The sensitivity is excellent, the instrumentation usually of low cost, and instrumental detection is often possible. Short-lived radionuclides have a high activity, whereas the mass of the nuclide is usually very low. The signal per unit weight material obtained is very high for short-lived nuclides. This can be illustrated by the following examples. One Bq corresponds to 0.3 pg of ^{137}Cs (half-life 30 years), 27 pg of ^{226}Ra (half-life 1599 years) and 152 ng of ^{129}I (half-life 1.57×10^7 years). Therefore nonradiometric methods have not been considered for the determination of important radionuclides like ^{90}Sr, ^{137}Cs, ^{241}Am and the different Cm isotopes. Radiometric methods also give the measure of the property sought directly, that is the radiological toxicity. Therefore radiometric methods are dealt with first and more thoroughly. A number of compilations and conference proceedings describe methods to determine radionuclides.[4–16]

In this encyclopedia, radiometric methods and their applications are described in the following articles:

Nuclear Detection Methods and Instrumentation
γ-Spectrometry, High-resolution, for Radionuclide Determination

β-Particle Emitters, Determination of
Actinides and other Alpha-emitters, Determination of

However, nuclear weapons testing and the nuclear industry, especially nuclear waste research, have introduced the need for measurement of long-lived radionuclides. In assessing the long-term environmental effects of nuclear power production, the disposal of radioactive waste in general and spent fuel in particular, a number of long-lived radionuclides have to be considered. These are ^{99}Tc, ^{129}I, ^{237}Np and ^{239}Pu.

For very long-lived nuclides the specific activity may be so low that detection limits are too poor for certain types of sample. When the mass (or number of atoms per unit weight) becomes significant compared to the intensity of the emitted radiation per unit weight of the radionuclide, other mass-sensitive detection methods may be more favorable. An example of this is given by Toole et al.[17] While the α-spectrometric detection limit is 0.1 mBq for all actinides, the detection limit for inductively coupled plasma mass spectrometry changes from 0.004 μBq to 84 mBq when the half-life of the nuclide changes from 1.4×10^{10} to 6.5×10^3 years (^{232}Th and ^{240}Pu respectively).

McMahon[18,19] and Rosenberg[20] have discussed all kinds of mass analytical techniques feasible for the determination of long-lived radionuclides. However, in normal cases the radionuclide has to be determined in a mixture with its stable isotopes with a very unfavorable mass ratio, or in a mixture with its radioactive isotopes. This means that methods capable of isotope analysis are needed. The only exception is ^{99}Tc, because it is the only isotope of Tc found in nuclear fuel. There are several references to the use of a number of analytical techniques for the determination of ^{99}Tc compiled by Long and Sparks.[21]

In practice the methods used are neutron activation analysis and mass spectrometry. These methods are described in this encyclopedia in the following articles:

Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides
Mass Spectrometry of Long-lived Radionuclides
Actinides and other Alpha-emitters, Determination of

In order to understand the behavior of radionuclides in the environment, it is not always sufficient to know their concentrations in different samples. There is also a need to know their physicochemical forms, their species. Therefore we have included one article describing methods for speciation of radionuclides:

Speciation of Radionuclides in the Environment

ABBREVIATIONS AND ACRONYMS

IT Isomeric Transitions

REFERENCES

1. M. Eisenbud, *Environmental Radioactivity*, Academic Press, USA, 1973.
2. H.A.C. McKay, *Principles of Radiochemistry*, Butterworths, England, 1971.
3. Glenn F. Knoll, *Radiation Detection and Measurement*, John Wiley & Sons, USA, 1979.
4. United Nations Scientific Committee on the Effects of Atomic Radiation, Sources and Effects of Ionizing Radiation, UNSCEAR 1993 Report to the General Assembly, with scientific annexes, United Nations, New York, 1993.
5. Generic Models and Parameters for Assessing the Environmental Transfer of Radionuclides from Routine Releases, Safety Series No. 57, Procedures and Data, IAEA, Vienna, 1982.
6. Monitoring of Airborne and Liquid Radioactive Releases from Nuclear Facilities to the Environment, Safety Series No. 46, Recommendations, IAEA, Vienna, 1978.
7. DOE Methods for Evaluating Environmental and Waste Management Samples, eds. S.C. Goheen, M. McCulloch, B.L. Thomas, R.G. Riley, D.S. Sklarew, G.M. Mong, S.K. Fadeff, Pacific Northwest Laboratory, Richland, Washington, 1994.
8. Environmental Measurements Laboratory, Procedures Manual, eds. N.A. Chieco, D.C. Bogen, E.O. Knutson, Environmental Measurements Laboratory, US Department of Energy, November 1990.
9. Rapid Methods for Measuring Radioactivity in the Environment, IAEA, Vienna, 1971.
10. Standard Methods for the Examination of Water and Wastewater, eds. A.E. Greenberg, L.S. Clesceri, A.D. Eaton, American Public Health Association, USA, 1992.
11. Organisation of Surveys for Radionuclides in Food and Agriculture, FAO, Rome, 1962.
12. S.K. Fadeff, S.C. Gohen, R.G. Riley, M.H. Carter, 'Radiochemistry Methods in DOE Methods for Evaluating Environmental and Waste Management Samples: Addressing New Challenges', *Radioactivity Radiochem.*, **5**(4), 16–25 (1994).
13. H.L. Krieger, E.L. Whittaker, Prescribed Procedures for Measurement of Radioactivity in Drinking Water, US Environmental Protection Agency, EPA-600/4-80-032, 1980.
14. Proceedings of the Third International Conference on Low-level Measurement of Actinides and Long-lived Radionuclides in Biological and Environmental Samples, *J. Radioanal. Nucl. Chem. Articles*, **156**, 1992.

15. Proceedings of the Fourth International Conference on Low-level Measurement of Actinides and Long-lived Radionuclides in Biological and Environmental Samples, *J. Radioanal. Nucl. Chem. Articles*, **177**(1) and **182**(1), 1994.
16. Proceedings of the Fifth International Conference on Low-level Measurement of Actinides and Long-lived Radionuclides in Biological and Environmental Samples, *J. Radioanal. Nucl. Chem. Articles*, **197**(1,2), 1995.
17. J. Toole, A.S. Hursthouse, P. Mcdonald, K. Sampson, M.S. Baxter, R.D. Scott, D.K. Mckay, 'The Determination of Actinides in Environmental Samples by ICP/MS, Plasma Source Mass Spectrometry', *Royal Society of Chemistry*, Special Publication no. 85, 155–162, 1990.
18. A.W. McMahon, 'A Critical Intercomparison of the Techniques for the Determination of Low Levels of Long-lived Radionuclides', United Kingdom Atomic Energy Authority HARWELL report, AERE R 13 617, 116p., 1989.
19. A.W. McMahon, 'An Intercomparison of Non-radiometric Methods for the Measurement of Low Levels of Radionuclides', *Appl. Radiat. Isot.*, **43**, 289–303 (1992).
20. R.J. Rosenberg, 'Non-conventional Measurement Techniques for the Determination of Some Long-lived Radionuclides Produced in Nuclear Fuel, A Literature Survey', *J. Radioanal. Nucl. Chem. Articles*, **171**(2), 465–482 (1993).
21. S.E. Long, S.T. Sparks, A Literature Survey of Techniques for the Analytical Determination of Technetium-99, AERE R 12 742, January, 36p., 1988.

Actinides and other Alpha-emitters, Determination of

Luis León Vintró and Peter I. Mitchell
University College Dublin, Dublin, Ireland

1 Introduction 1
2 Yield Monitors: Necessity and Choice 2
2.1 Monitors for the Analysis of Thorium 2
2.2 Monitors for the Analysis of Uranium 3
2.3 Monitors for the Analysis of Neptunium 3
2.4 Monitors for the Analysis of Plutonium 3
2.5 Monitors for the Analysis of Americium and Curium 3
3 Matrix Pretreatment 4
3.1 Sampling 4
3.2 Sample Preparation 5
3.3 Filtration/Ultrafiltration 6
3.4 Preconcentration 6
4 Radiochemical Separation and Purification 10
4.1 Ion Exchange 10
4.2 Solvent Extraction 13
4.3 Example of a Scheme for the Separation and Purification of Actinide Nuclides 15
4.4 Extraction Chromatography 17
5 Chemical Speciation Analysis 18
6 Source Preparation 20
6.1 Sources Suitable for Counting by α-Spectrometry 20
6.2 Sources Suitable for Liquid Scintillation α-Spectrometry 21
6.3 Sources Suitable for Mass Spectrometric Counting 21
7 Radiometry 21
7.1 α-Spectrometry 21
7.2 Liquid Scintillation α-Spectrometry 25
8 Mass Spectrometry 26
8.1 Inductively Coupled Plasma Mass Spectrometry 26
8.2 Thermal Ionization Mass Spectrometry and Resonance Ionization Mass Spectrometry 26
8.3 Accelerator Mass Spectrometry 27
9 Quality Assurance/Quality Control 28
9.1 Traceability and Standardized Solutions 28
9.2 Field and Reagent Blank Analyses 28
9.3 Intercomparison and Reference Materials 28
10 Perspectives and Future Developments 29
Acknowledgments 29
Abbreviations and Acronyms 29
Related Articles 29
References 30

To assay an actinide or other α-emitter at the low concentrations encountered in many environmental samples necessitates chemical separation of the actinide of interest from the bulk matrix, purification by the elimination of other interferants, mounting onto a suitable substrate for counting, and quantification by radiometric, mass spectrometric or other means. In this contribution, the chain of analysis from sampling to counting is critically reviewed and many of the most important pitfalls and sources of error are highlighted. Specifically, the pretreatment, preconcentration, separation and purification steps required for the assay of thorium, uranium, neptunium, plutonium, americium and curium in a wide range of environmental matrices are examined in detail. Emphasis is given to the need for appropriate radiochemical yield monitors and to the practical application of ion exchange, solvent extraction and extraction chromatography in the purification of these elements. This is complemented by a description of methods for the preparation of quality sources suitable for counting by α-spectrometry, liquid scintillation α-spectrometry and mass spectrometry. In addition, an overview is given of the operational principles and systems used in α-radiometry and mass spectrometry, including a discussion of the practical limits of detection and quantitation achievable with these systems. Finally, the importance of traceability and good quality assurance in precision actinide assay is emphasized.

1 INTRODUCTION

The actinide series of the periodic table comprises the naturally occurring thorium, protactinium and uranium nuclides, and the transuranium elements neptunium to lawrencium, inclusive. Small traces of naturally produced transuranium nuclides remain or are being generated in nature,[1–3] but the overwhelming ''weight'' of present environmental inventories of these elements derive from human activities, related mainly to the nuclear fuel cycle

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and to military applications. Although direct quantitative measurement of some of these nuclides is possible by non-destructive techniques such as γ- and X-ray spectrometry, most of them decay by the emission of α- or β-particles, or by processes such as electron capture and isomeric transitions without accompanying γ-emissions. For these nuclides, direct measurement in the presence of the host matrix is not possible, as the particles emitted are rapidly absorbed by the matrix before they can be detected. At environmental level, even those nuclides with γ- and X-ray emissions may be difficult to detect, given the minute concentrations involved, the comparatively low sensitivity of γ- and X-ray detection systems, and interference from other photon-emitters in the sample. In such cases, radiochemical separation of the nuclides of interest from the bulk matrix, purification by the elimination of other interferants and preparation in a form suitable for counting are essential. Accurate determination of actinide concentrations at these low levels depends heavily on the analyst's ability to prepare extremely pure analytes, free of any gravimetric or radiometric interferences which could lead to complications in the interpretation of the measured spectra.

Since most of the actinides spontaneously decay by α-emission, high-resolution α-spectrometry has become the most widely used technique for the determination of these nuclides. However, a number of alternative techniques, such as inductively coupled plasma mass spectrometry (ICPMS) and accelerator mass spectrometry (AMS), are becoming increasingly popular.

Over the years, numerous radiochemical separation schemes, based primarily on coprecipitation, ion exchange or solvent extraction, have been proposed and adapted for different matrices and types of analysis. In this contribution, an up-to-date review of many of these separation schemes is presented, together with a detailed description of those schemes most widely used today. Although the techniques described are valid in general and can be applied to the analysis of a wide range of concentration levels, our discussion here will focus on the analysis of α-emitting radionuclides present at trace or indicator level in environmental matrices. Specifically, attention will be paid to the determination of environmental concentrations of the naturally occurring thorium and uranium α-emitting nuclides, and the transuranium nuclides neptunium, plutonium, americium and curium.

2 YIELD MONITORS: NECESSITY AND CHOICE

Since the separation of the nuclides of interest from a host matrix involves physical and chemical operations which are unlikely to be quantitative or reproducible from one sample to another, precise determination of a particular actinide's concentration in an environmental sample requires the use of a chemical yield monitor with which to assess the overall recovery of the operations involved. Because the actinides have no stable isotopes, other radioactive isotopes of these elements are employed by most laboratories as chemical yield monitors.

The basic requirements and some of the difficulties encountered when using chemical yield monitors in actinide analyses at environmental concentrations have been discussed in detail by several authors.[4–6] Essentially, a suitable yield monitor should have the same chemical behavior as that of the determinand at all stages of the analysis, should be in complete physical and chemical equilibrium with the determinand from the earliest possible stage of the process, and should not interfere with its measurement. Further limitations are imposed by half-life, production costs, purity of the yield monitor and, most importantly, the presence of the intended tracer in the environmental materials to be analysed. The choice of yield monitor and the suitability and availability problems encountered are best illustrated by considering each element separately.

2.1 Monitors for the Analysis of Thorium

Of the thorium isotopes, α-emitting ^{228}Th ($t_{1/2}$ = 1.91 years), ^{230}Th ($t_{1/2} = 7.7 \times 10^4$ years) and ^{232}Th ($t_{1/2} = 1.41 \times 10^{10}$ years) are widely used as geochronometers and as powerful tracers of oceanographic processes. In sediment dating, ^{228}Th is frequently used as a yield monitor for ^{230}Th and ^{232}Th. Its use, however, is predicated on the absence of ^{228}Th in the sample in question. When, as is often the case, significant amounts of ^{228}Th are present in the sample, or when this nuclide is the object of the analysis, an alternative yield monitor must be employed. This is usually ^{234}Th ($t_{1/2} = 24.1$ days), which has traditionally been employed as an isotopic yield monitor for the radiochemical determination of ^{228}Th, ^{230}Th and ^{232}Th. Although ^{234}Th can be prepared easily from a ^{238}U solution with good isotopic purity, the main difficulties associated with the use of this nuclide are its relatively short half-life, its occurrence in the sample, and the fact that it decays by both β- and γ-emission, thus requiring a separate radiometric determination.

Another alternative is ^{229}Th, which is being employed increasingly.[7] The main advantages of this nuclide as a monitor are its long half-life ($t_{1/2} = 7340$ years) and its decay by α-emission, which simplifies counting, requiring a single radiometric determination. The complexity (i.e. multicomponent nature) and energy spread of its α-''peak'', and its close proximity to the ^{230}Th ''peak'' (actually an unresolved doublet) require the preparation

of quality sources and the avoidance of ''overspiking'' (i.e. the addition of an excessive activity of yield monitor relative to that of the determinand/analyte) so that interference from spectral degradation is kept to a minimum.[4]

2.2 Monitors for the Analysis of Uranium

Uranium analysis is usually based on the determination of the naturally occurring ^{234}U, ^{235}U and ^{238}U nuclides. An ideal yield monitor is ^{232}U ($t_{1/2} = 72$ years), whose α-''peak'' is well resolved from those of the three nuclides of interest. Samples should be measured as soon as possible after radiochemical separation, as ^{228}Th, ingrown from the decay of ^{232}U, interferes spectrometrically with its ^{232}U precursor.

2.3 Monitors for the Analysis of Neptunium

The only significant α-emitting nuclide of neptunium in the environment is ^{237}Np. As there is no other α-emitting neptunium isotope available, most radiochemical determinations of ^{237}Np employ the β-emitter ^{239}Np ($t_{1/2} = 2.36$ days) as yield monitor. Again, a separate radiometric determination is required, usually by β-counting, to determine the radiochemical yield (the daughter of ^{239}Np is a γ-emitter). The practicalities associated with the use of ^{239}Np as a yield monitor have been discussed in some detail by Harvey and Lovett.[4] Suffice that ^{239}Np can readily be obtained by ''milking'' from a solution of its precursor, ^{243}Am ($t_{1/2} = 7380$ years), with which ^{239}Np is in secular equilibrium. Solvent extraction and/or ion exchange can be used to separate ^{239}Np from ^{243}Am (see section 4.2.3), at which point the ^{239}Np is added to the sample as yield monitor.[8,9] It is imperative that the separation of the two be quantitative and this must be checked each time. Alternatively, ^{243}Am, with ^{239}Np in equilibrium, can be added directly to the sample without prior separation.[10] The first approach has the disadvantage of requiring weekly separation and calibration. On the other hand, the second approach relies on the complete elimination of americium in a single analytical step at a precisely determined time.

Np-235 ($t_{1/2} = 396$ days) has been used on occasion as an alternative yield monitor for neptunium analysis.[11] Although its half-life is somewhat longer than that of ^{239}Np, availability difficulties and the fact that it must be assayed separately by X-ray spectrometry has limited its application.

2.4 Monitors for the Analysis of Plutonium

The three most commonly used yield monitors in the analysis of plutonium are ^{236}Pu, ^{242}Pu and ^{244}Pu. Pu-236 is produced with accelerators by charged particle irradiation of suitable targets, for example, via the reactions ^{234}U(α, 2n)^{236}Pu and ^{235}U(d, n)^{236}Np(β^-)^{236}Pu.[6] The main advantage of this method is that the yield monitor is produced in a pure state, free of other α-emitting plutonium nuclides. There are, however, several disadvantages. The α-''peak'' of ^{236}Pu lies to higher energy than those of the plutonium nuclides of environmental interest and care must be taken to avoid tailing of the ^{236}Pu ''peak'' into that of ^{238}Pu as a result of overspiking with ^{236}Pu or poor resolution due to degraded source quality.[4,12] Further, the short half-life of ^{236}Pu ($t_{1/2} = 2.85$ years) not only limits its shelf life but also requires the periodic removal of ingrown daughters (i.e. $^{232}\text{U} \rightarrow {}^{228}\text{Th} \rightarrow {}^{224}\text{Ra} \rightarrow {}^{220}\text{Rn} \rightarrow {}^{216}\text{Po} \rightarrow {}^{212}\text{Pb} \rightarrow {}^{212}\text{Bi}$), some of which can interfere with both ^{236}Pu and ^{238}Pu. Also, its short half-life prohibits its use in mass spectrometric analysis.

Pu-242 is produced by successive neutron capture in ^{239}Pu, with chemical purification of the plutonium followed by mass separation of the ^{242}Pu from other isotopes.[6,13] In contrast to ^{236}Pu, the ^{242}Pu tracer is not isotopically pure and corrections, albeit small, must be made for certified impurities of α-emitting ^{238}Pu, ^{239}Pu, ^{240}Pu and, if measured, β-emitting ^{241}Pu. The α-''peak'' of ^{242}Pu lies lower in energy than that of any other α-emitting plutonium nuclide of environmental interest and, consequently, moderate overspiking is not a problem. However, underspiking must be avoided as the degraded tails of the unresolved ^{239}Pu and ^{240}Pu peaks can contribute to the ^{242}Pu ''peak''.[4] The long half-life of ^{242}Pu ($t_{1/2} = 3.76 \times 10^5$ years) makes it suitable for both radiometric and mass spectrometric measurement, and in practice means that a solution of ^{242}Pu remains free of daughter impurities for a substantial period of time.

Pu-244 ($t_{1/2} = 8.26 \times 10^7$ years) is also produced by neutron irradiation, but is even more difficult to obtain in a pure state than ^{242}Pu. Its low specific activity precludes its use in α-spectrometry, but makes it very attractive for mass spectrometric analysis.

2.5 Monitors for the Analysis of Americium and Curium

Since americium and curium have very similar chemical properties and do not interfere spectrometrically with each other, they are often analysed together using a single yield monitor, most commonly ^{243}Am. This approach is valid provided that a radiochemical procedure is employed that does not discriminate between americium and curium at any stage. As in the case of ^{242}Pu, the α-emissions of ^{243}Am lie lower in energy than those of the americium or curium nuclides of interest, and care must be taken to avoid underspiking which could result in the degraded tail of ^{241}Am interfering with the ^{243}Am ''peak''. The daughter of ^{243}Am ($t_{1/2} = 7380$ years) is ^{239}Np ($t_{1/2} = 2.36$ days), a short-lived β-emitter which, in

turn, decays to ^{239}Pu. If, as is often the case, the sample is analyzed for both plutonium and americium, the ^{239}Pu impurity introduced by the addition of ^{243}Am as yield monitor must be taken into account, as must the ^{241}Am impurity derived from the decay of the ^{241}Pu impurity present in the ^{242}Pu yield monitor.

3 MATRIX PRETREATMENT

3.1 Sampling

The procurement of truly representative samples is a subject in itself and represents, undoubtedly, one of the most difficult and limiting aspects of environmental analysis. The selection of the matrices to be analysed, the sampling methodology to be adopted, the number of samples required, and the sample sizes needed to ensure representativeness are just some of the issues to be considered when planning a sampling campaign.[14–17] Environmental materials routinely analysed for actinide nuclides include aqueous samples (e.g. seawater, freshwater, groundwater), soil, sediment, foodstuffs, milk, human and animal tissues, bones, urine, air filters, seaweed, vegetation, etc. Key features of the procedures employed to sample some of these materials are discussed below.

3.1.1 Water Samples

The analysis of actinides in freshwater or seawater requires that a representative sample at a prescribed sampling depth be taken and that the sample remain unaltered until the water is retrieved and treatment commences. For these reasons, all water samplers should be fabricated from clean, inert materials in order to avoid contamination and minimize adsorption on those surfaces in contact with the sample. When sampling deep waters, the sampler should fill or exchange its contents with the surrounding water at the sampling depth and then be sealed completely by a triggering mechanism.[18]

Near-surface samples are often taken from at least 2 m below the surface using peristaltic or deck pumps made from clean, inert materials. Deeper waters are usually sampled using specially designed bottles (e.g. Niskin bottles, Gerard barrels, etc.). The sample volume required for a satisfactory analysis is obviously a function of radionuclide concentration; it may also depend on the nature of the analysis to be performed.

When the separate analysis of suspended particulate and filtered water is required, the samples should be filtered immediately after collection (see section 3.3). Filtered (or, indeed, unfiltered) samples should be stored in poly(vinyl chloride) (PVC) tanks or polyethylene carboys and pretreatment should commence as soon as possible after filtration (sampling) to avoid losses by adsorption onto container walls. Depending on the nature of the analysis, the addition of HCl or other mineral acids and/or more complex mixtures may be required to stabilize the samples.

In order to avoid the disastrous consequences of cross-contamination, the storage tanks should be thoroughly cleaned after use or, where possible, replaced by unused ones. An alternative is to fit disposable polyethylene liners inside the containers. In the case of equipment that, of necessity, has to be reused, care should be taken to clean all surfaces which may have been in contact with the samples as thoroughly as possible before reuse.

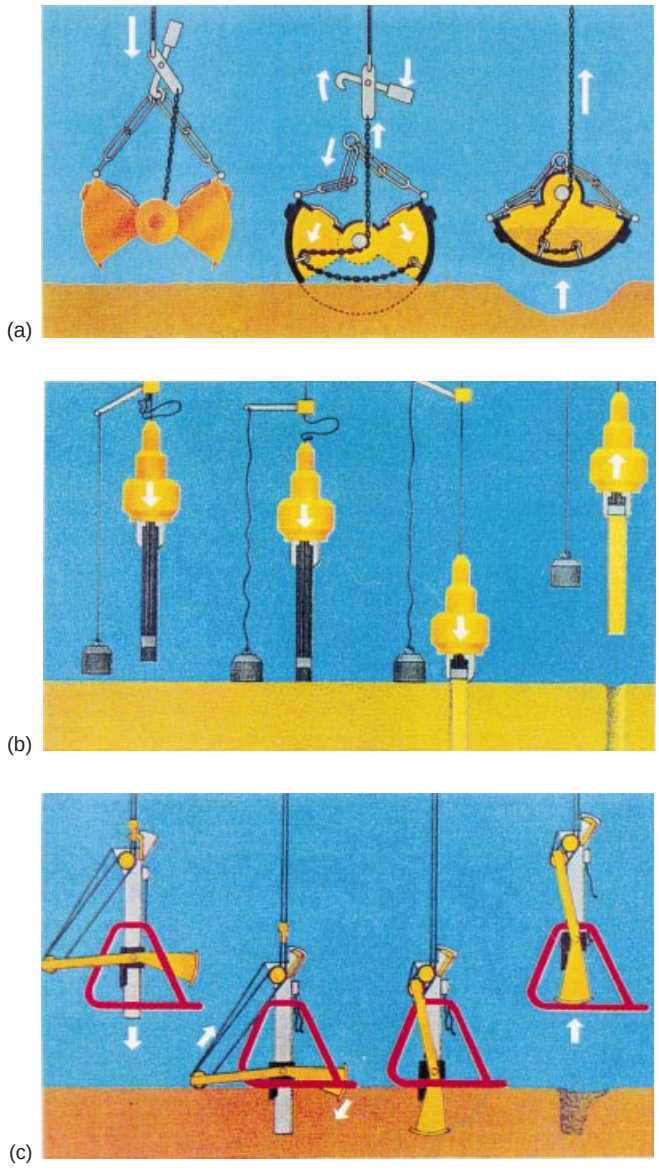

Figure 1 Operational principle of (a) grab samplers, (b) piston corers and (c) box corers.

3.1.2 Sediment Samples

Sampling of surface sediments is most frequently done with dredgers or grab samplers. Dredgers, although useful for retrieving coarse sediments and gravels, suffer from severe limitations if representative samples are required. Sediments are collected from a stretch of seabed rather than at a particular point and much of the fine-grained fraction is lost as the dredger is brought to the surface. These problems are partially solved with the use of grab samplers (Figure 1a), which include jaws that drive into the sediment and close, retaining the full sample. However, loss of the fine-grained fraction is not uncommon when complete closure is not achieved due to a shell, pebble or rock fragment becoming trapped in the jaws, and there is little control of the sampling depth. If, as is often the case, preservation of the sediment is a major consideration, coring devices, rather than dredgers and grabs, must be used. The types of corers most frequently employed to sample seabed and lacustrine sediments are the gravity corer, the piston corer and the box corer.

Gravity corers are simple devices, with no moving parts to jam or fail. The penetration of a hollow tube or cylinder into the sediment bed depends only on the weights attached above the tube, and a small hydrographic winch is usually enough for lowering and retrieving such a corer. The base of the coring tube contains an annulus of slightly curved, spring metal fingers which prevents the loss of sample upon pull-out. This is aided by a check valve at the top of the corer, which allows water to flow through the barrel as it penetrates the bed and which closes upon retrieval. The operational principle of the piston corer is quite different from that of the gravity corer, in that the sampling barrel in a piston corer is activated and driven into the bed by an automatic ejection mechanism after the surface of the bed has been reached (Figure 1b).

Box corers consist of a weighted cubic box, mounted on a housing frame, which is allowed to sink into the sediment, again with the assistance of weights. At this point, a forked arm is triggered which pivots underneath the bottom of the sampling box, enabling the retrieval of a large quantity of practically undisturbed sediment (Figure 1c). Back on deck, subsamples may be taken by driving hollow PVC tubes vertically through the retrieved sediment ''monolith''. For more comprehensive and detailed descriptions of sediment sampling techniques, the reader is referred to McQuillin and Ardus,[19] Moore and Heath,[20] and Stowe.[21]

Sediment cores must be extruded and sectioned as soon as possible after recovery. It is recommended that an outer annulus or rind of sediment be removed from each section or layer in order to eliminate contamination due to smearing, which is difficult to avoid during sampling and extrusion.[22]

3.1.3 Soil Samples

If, as is often the case, soil samples are used to estimate accumulated deposition from, for example, global fallout, sampling should be carried out in open, flat, undisturbed areas.[17] Meteorological conditions, vegetation and human/animal activity should be carefully considered when selecting a particular site. In the case of contaminated sites, systematic sampling should be carried out to establish the distribution pattern of the contamination. The type of sampling device or approach used clearly depends on the characteristics of the soil in question (e.g. granulometry, porosity, moisture content, mineralogical composition, etc.). Coring is the preferred method for cumulative deposition studies. The main limitation is, of course, the presence of large quantities of stones, roots or other material which hamper the penetration of the corer to the required depth. In these cases, alternative methods such as scooping within a defined template are recommended. A detailed description of these methods can be found in Chieco.[17]

3.1.4 Biological Samples

In the case of biological samples (including foodstuffs, marine and terrestrial vegetation, human and animal bones, tissues, urine, etc.), the amount of material available is generally larger than that required for analysis and, thus, care must be taken to ensure that a representative subsample is acquired.[23] This can be achieved by homogenization of a large or composite sample prior to subsampling. Replicate analyses should be carried out to demonstrate the adequacy of the homogenization and subsampling methods employed. Wet and dry ashing techniques, described below, are normally used to eliminate organic matter as samples in an inorganic form are more readily solubilized.

3.2 Sample Preparation

Sample preparation covers all the preparatory steps that must be carried out prior to chemical separation and purification. Importantly, these steps include the complete homogenization of the sample as well as the equilibration (isotopic exchange) of yield monitor(s) with determinand(s) and, in certain cases, the stabilization of their respective oxidation states.

In the case of water samples, the initial step in the analysis is designed to achieve full equilibration between determinand(s) and yield monitor(s) prior to preconcentration. This is particularly important for multivalent actinides such as plutonium, which normally present in aqueous media as a mixture of oxidation states exhibiting quite distinct sorption/coprecipitation properties. The most common method of achieving

equilibration is by immediate acidification of the sample, followed by the addition of a reducing agent such as Na_2SO_3[24] or $Na_2S_2O_5$.[25] Appropriate quantities of yield monitor(s) and carrier(s), for example Fe, Mg, Ca, Mn, are added and the sample is mixed for several hours to ensure thorough homogenization. If the suspended particulate phase and the solution phase are to be analyzed separately, filtration should be carried out immediately after sampling and prior to acidification.

In the case of urine samples, equilibration between determinand(s) and yield monitor(s) is achieved by, for example, acidification with aqua regia and the repeated addition of H_2O_2 in the course of wet ashing (see section 3.4.2.2).

Solid samples such as soil, sediment and biota are normally dried in an oven to constant weight, ground, blended or milled to a fine powder, and thoroughly mixed to ensure homogeneity. In some cases, a defined heating cycle is recommended.

3.3 Filtration/Ultrafiltration

Microfiltration is the process of removing particulate matter from aqueous or gaseous samples by passage through a microporous medium such as a membrane filter. Although micrometer-sized particles can also be removed by the use of nonmembrane or depth filters, that is fibrous media, only a membrane filter having a precisely defined pore size can ensure quantitative retention. Screen filters are so named because they retain particles on the surface of the membrane through which the aqueous (or gaseous) sample is forced to pass. High quality filters are readily available from a number of leading manufacturers.

In the case of water, it has become conventional to define the suspended particulate phase as the material retained by a 0.45-μm pore-size filter. This is, however, an operational cut-off, since a variety of smaller particles and colloids are known to be present in great abundance in all natural waters.[26–28] When the analysis of smaller particles in the colloidal-size range is required, these particles can be concentrated by, for example, tangential-flow ultrafiltration using polysulfone membrane cassettes with nominal molecular weight limits (NMWLs) as low as 0.8–1 kDa (1 kDa ~ 1.2 nm).

The ultrafiltration assembly is normally operated in concentration mode. In this mode, macromolecules and colloids are retained by the membrane, flowing tangentially across its surface and returning to the original sample container, the contents of which are referred to as the retentate. Molecules and particles smaller than the pore size (NMWL) of the membrane are able to pass through it and these (the permeate) are directed to a separate container. Both fractions can then be treated separately in the same manner as filtered water.

3.4 Preconcentration

Following the sample preparation stage, samples are usually preconcentrated to reduce bulk or volume, after which the separation chemistry can commence. Preconcentration may involve the ashing of solid samples to remove organic material, the evaporation of liquid samples to reduce their volume, the acid leaching of solid samples followed by evaporation of the leachates, bulk coprecipitation, etc.[29] Since the actual steps involved depend on the type of matrix being processed, and these are numerous, we have confined ourselves here to some examples chosen to highlight the preconcentration steps required for the assay of actinides in some of the most frequently analyzed materials, namely water, sediment, soil and biological matter. The reader should, nevertheless, note that alternative or additional steps may be required for certain other types of matrices (e.g. urine, bone, etc.), the details of which have recently been reviewed elsewhere.[17,23]

3.4.1 Preconcentration of Aqueous Samples

One of the difficulties associated with the analysis of actinides in sea- and freshwater is the requirement, more often than not, to process large volumes in order to achieve the required analytical precision. The logistic and practical problems associated with the handling and in situ processing of large volumes should not be underestimated. Numerous techniques have been proposed over the years for the preconcentration of actinides from large volumes of water. These are generally based on the coprecipitation of actinides with suitable carriers or their in situ sorption onto mineral surfaces.

3.4.1.1 Actinide Preconcentration by Coprecipitation
Although a variety of methods have been developed based on coprecipitation with, for example, manganese dioxide,[30] magnesium hydroxide,[31] bismuth phosphate[32,33] and ferrous hydroxide,[25,34] coprecipitation of actinides with ferric hydroxide[35] is most widely used, particularly for transuranium analyses.[36–41]

In essence, the previously filtered (or, sometimes, unfiltered) sample is acidified with 12 M HCl to pH ~ 1.6. Upon the addition of appropriate amounts of yield monitor(s), reducing agent(s) and Fe^{3+} carrier (>1 mg per litre of sample), the sample is vigorously stirred for at least 3 h in order to ensure complete equilibration. The coprecipitation of the actinides with ferric hydroxide is achieved by adjusting the pH of the sample to between 7.5 and 8.0 using sodium hydroxide or ammonia. Care must be taken during the pH adjustment to prevent the formation of insoluble calcium and magnesium carbonates. The sample is left to settle overnight (~12 h) and the $Fe(OH)_3$ precipitate, containing the actinide nuclides, is collected

by filtration through, for example, a cellulose nitrate screen filter. The filter is then dissolved in a concentrated solution of HNO_3 and HCl (e.g. 1:1 by volume), the solution evaporated to dryness and the residue, after conversion to nitrate or chloride form, redissolved in the appropriate medium for further purification.

The ferric hydroxide coprecipitation method cannot be used if the analysis of neptunium is also a requirement. In aqueous solution, neptunium is present as the pentavalent neptunyl ion, which is soluble even at alkaline pH and does not coprecipitate with ferric hydroxide.[42] In this case, an alternative coprecipitation step, such as coprecipitation with barium sulfate following reduction with ferrous ammonium sulfate,[43] is recommended.

3.4.1.2 Actinide Preconcentration by In Situ Sorption onto Metal Oxides There are obvious practical difficulties in separating radioelements from large volumes of water, as we have already noted. However, a number of techniques for the in situ separation of specific radioelements from bulk samples have been developed over the years, which obviate the need to collect and store large volumes. All of these methods have one thing in common in that the sample is pumped quickly through a cell or sampler containing a chemical agent which separates the radioelement of interest from the bulk liquid.

Sorption on, for example, manganese dioxide[30,44–49] or aluminum oxide[50,51] has been used to preconcentrate actinide elements from large volumes of water. Sorption samplers capable of processing up to 4000 L of water with a single cartridge (or charge) have been developed[52] which removes the need to collect and transport large volumes of water while, at the same time, providing a substantial improvement in analytical precision due to the larger volumes that can be processed in relatively short periods of time. Problems such as sample deterioration during handling and storage are virtually eliminated, as the sampler operates on water pumped directly from the water body in question.

Caution, however, should be exercised when using sorption samplers, as laboratory and field tests have shown that actinide retention by these systems is not necessarily quantitative. In experiments using Al_2O_3 as the sorbing agent, Nevissi and Schell[53] demonstrated that while ^{207}Bi, ^{210}Po, ^{155}Eu and ^{241}Am were quantitatively retained on a single sorption bed (200 mesh Al_2O_3 powder, ~6 mm thick), ^{60}Co and $^{239,240}Pu$ were not. To quantify ^{60}Co and $^{239,240}Pu$, both of which exist in natural waters in more than one valence state simultaneously, it proved necessary to pass the sample through a set of three (or more) beds arranged in series and to measure the retention of both nuclides on each of the beds. Clearly, the viability of the sorption cell approach is based on the assumption that none of the radioelement of interest is present in a nonsorbable form. There is evidence to suggest that this is not always the case.

3.4.2 Dry and Wet Ashing/Fusion

Fundamental to any radiochemical analysis is the need to prepare the analyte in a form suitable for separation chemistry, ideally, fully dissoluted or solubilized. A combination of dry and wet ashing greatly facilitates the leaching of soluble components from a solid matrix, in that any organic matter present is readily eliminated and adsorbed radionuclides can be taken into solution. Accordingly, dry and wet ashing are usually the first operations to be carried out in most radiochemical separations. Experience, however, shows that dry and wet ashing, on their own, are seldom sufficient to completely solubilize a sample, and additional steps may have to be employed, particularly in the case of samples containing actinides in a refractory form.

3.4.2.1 Dry Ashing Dry ashing is normally the first preconcentration step in treating biological samples such as seaweeds, soft animal tissues and bones. Its main purpose is the oxidation of organic matter by slow combustion, until the sample is converted into a nonvolatile inorganic residue comprising metallic oxides as well as nonvolatile phosphates, silicates and the like.

In the course of dry ashing atmospheric oxygen acts as an oxidizing agent, converting organic carbon into CO_2. Some elements are prone to volatilization, especially when certain other species are present in the matrix. For example, the chloride ion can react with lead producing volatile components such as $PbCl_2$. Moreover, if polonium is present, it can be volatilized very easily given its low melting point of 254 °C.[54] Consequently, volatilization must be guarded against when carrying out certain analytical determinations.

Preparatory to dry ashing, an appropriate quantity of sample is dehydrated at a suitable temperature (traditionally 105 °C, though temperatures as low as 80 °C have become common) to constant mass in a ventilated oven (this step is necessary to determine the dry mass of the sample, which is the reference mass to which the mass activity of the analyte is normally referred). Following this, an accurately measured quantity of the appropriate yield monitor is added by pipetting little droplets of the solution as evenly as possible over the surface of the sample and gently drying under an infrared lamp.

Dry ashing is effected by placing the sample in a preweighed porcelain crucible and heating it (partially covered) in a muffle furnace while slowly raising the temperature from approximately 150 to 500 °C in, for example, 50 °C increments over a period of 48 h. Ashing is continued until a powdery, white ash forms, indicating

the complete removal of carbon. The additional time necessary for this process is typically about 24 h.[17,55] Upon completion of the ashing process, the sample is weighed again, this time to determine the ash weight. Nonreproducibility in the dry weight to ash weight ratio for a set of similar samples is a useful indicator of possible losses due to deflegration in the course of combustion.

Dry ashing is advantageous in that it provides a simple means of substantially reducing sample bulk (a factor of between 4 and 10 is typical for soft tissues). However, it is not to be recommended if the nuclides of interest are volatile at ashing temperatures. A good example is the measurement of ^{210}Po in shellfish, where the edible portions should only be wet ashed. Dry ashing should not lead to any significant losses in the case of the actinides, all of which have melting points above 500 °C.

A complication arises from the fact that reactions can take place within the ash at high temperatures which result in the formation of refractory metals and other refractory compounds. Violent ignition of the sample surface may produce such high temperatures. Since the actinides are known to form refractory oxides at these temperatures, it is recommended that the temperature of the furnace be increased very slowly in order to prevent sudden ignition bursts. Equally, the furnace temperature should not be allowed to exceed 550 °C.

3.4.2.2 Wet Ashing Wet ashing is a very useful method to be applied in cases where the organic content of the sample is not too high (e.g. mineral-rich sediments or previously dry-ashed samples). The technique can also be used to digest precipitates retained by cellulose microfilters simply by wet ashing the filters together with their contents. Although it is a more tedious procedure than dry ashing, it is often preferred as there is little or no risk of volatilization and refractory compound formation. Moreover, yield monitors and carriers can be added directly to the solution, becoming rapidly equilibrated with indigenous ions in the course of the (wet) ashing process.[17,54–58]

Decomposition and dissolution of samples by wet ashing occurs because most electropositive metals, as well as many metal oxides, carbonates, sulfides, etc., are soluble in dilute acids. More electronegative metals and some of the commonest minerals (e.g. aluminates, silicates) require treatment with hot concentrated mineral acids, which provide the necessary high proton concentration.[54]

The most commonly used reagent in wet ashing is concentrated (16 M) HNO_3. Given its high oxidizing power and acid strength, most digestions can be carried out using this reagent alone. Alternatives include concentrated (12 M) HCl, a mixture of three parts of 12 M HCl to one part of 16 M HNO_3 (aqua regia),[41,59] or 40% HF.[60] The oxidizing power of these acids can be enhanced by the addition of small volumes of H_2O_2 during the digestion process.[61,62]

Hydrofluoric acid is a weak, nonoxidizing acid. However, the fluoride ion is a most powerful complexing anion, capable of combining with refractory elements to the point of making them soluble. The main application of HF is the solubilization of siliceous materials by volatilization of silicon as SiF_4 while the rest of the matrix is brought into solution. For this reason, HF digestion is an appropriate procedure with which to treat soil samples from nuclear bomb test sites, where plutonium and americium may have become incorporated in silicates and polysilicates during the explosion/fireball. In addition, when using HF as a digestant, uranium in a hexavalent state can be eliminated following the formation and volatilization of UF_6. There are, however, certain difficulties associated with the use of HF, including the requirement to use poly(tetrafluoroethylene) (PTFE) or platinum beakers instead of borosilicate glass vessels, the need to remove completely the fluoride ion after digestion in order to prevent it from interfering at a later stage in the radiochemical analysis, and the high toxicity of this acid.

When the complete oxidation of organic matter that proves resistant to digestion by the means outlined above is desired (e.g. lipids), treatment with small additions of hot concentrated (98%) H_2SO_4 is recommended (H_2SO_4 is a very strong oxidizing agent, which possesses powerful dehydrating properties). In addition, the digestant can be heated to relatively high temperatures by virtue of the higher boiling point of H_2SO_4 compared to that of HNO_3 or HCl. However, its use is not recommended in the presence of significant quantities of Ca, Sr, Ba and Pb, as they can form sparingly soluble ($CaSO_4$) or totally insoluble ($SrSO_4$, $BaSO_4$ and $PbSO_4$) sulfates. In these cases, an alternative procedure is to use perchloric acid ($HClO_4$), which dissolves most metals and alloys. However, potassium, rubidium and cesium perchlorates are insoluble in aqueous solution, and the use of concentrated $HClO_4$ is restricted by the ever-present risk of explosion.

Various wet ashing procedures, applicable to different sample matrices, have been described in detail in the literature.[17,56,60–63] Here, we shall confine ourselves to a typical methodology which has proved its worth in our laboratory.

Wet ashing of biological samples follows immediately upon dry ashing. Spiked samples are removed from the muffle furnace and allowed to cool to room temperature in their original crucibles. Each sample is transferred to a beaker containing 16 M HNO_3 (typically 100 mL). The crucible is rinsed twice with 16 M HNO_3, and the rinses added to the beaker, which is transferred to a hot plate.

The mixture is then boiled under reflux (i.e. covered with a watch glass to minimize evaporation) and stirred using a Teflon-coated magnetic stirring bar. Small additions of H_2O_2 are made regularly during the digestion, which is allowed to proceed for a minimum of 6 h. The watch glass and the magnetic stirrer are removed and the sample is evaporated to near dryness. The presence of unoxidized organic matter is evidenced by the evolution of brown fumes and the formation of a black residue. The further addition of small quantities of 16 M HNO_3 and H_2O_2, and repeated evaporation to dryness until the residue appears to be almost pure white, is sufficient to eliminate residual organic matter. The sample can now be redissolved in 100 mL of 8 M HNO_3. It is our practice to microfilter the sample at this stage prior to purification by solvent extraction or ion exchange.

The digestion procedure for sediments and soils is somewhat simpler. A small quantity of sediment (10–20 g) is dried at 105 °C to constant mass and digested with 100 mL of hot aqua regia spiked with appropriate yield monitors. After a minimum of 6 h of continuous leaching, the sample is allowed to cool to room temperature and the solid residue is removed by microfiltration. The filtrate is evaporated to dryness and, after conversion to nitrate form, re-dissolved in 8 M HNO_3, at which stage it is ready for further purification. Note that this procedure does not involve the complete dissolution of the sample; it is designed purely to leach actinides that have been (weakly) adsorbed on the surface of sediment particles, as opposed to actinides incorporated within the mineral lattice.

Heated pressure vessels and microwave digestion units are being used increasingly to accelerate the digestion and to achieve the complete dissolution of small (gram-sized) samples.[64–68]

The digestion of soil samples containing actinides in a refractory form requires a more aggressive attack, which can be achieved by digesting the samples in a mixture of 40% HF and 16 M HNO_3, 1 : 3 by volume. HF provides the complexing anion necessary to break down silicate compounds, while HNO_3 provides the oxidizing power. As stressed above, digestions must be carried out in PTFE or platinum-coated dishes. Upon completion of the process, samples must be evaporated fully to dryness and converted to the nitrate form. Re-dissolution in 8 M HNO_3, followed by filtration, should be carried out in order to obtain a clear/pure solution suitable for further processing.

Certain precipitates, for example $Fe(OH)_3$, can usually be dissolved in 16 M HNO_3 after a 6-h digestion, as described above for biological samples. However, dissolution of other precipitates, such as NdF_3, can be quite difficult, particularly when large quantities of precipitate are involved (e.g. plutonium chemical speciation analysis; see section 5), as NdF_3 is practically insoluble in concentrated HNO_3. Nevertheless, Lovett and Nelson[69] have reported that the dissolution of small quantities of NdF_3 can be effected with ~70 mL of 16 M HNO_3 to which a small quantity of boric acid (H_3BO_3) and an equal volume of deionized/distilled water have been added.

3.4.2.3 Fusion Total dissolution of mineral or biological samples is indicated in cases where actinides are known to be present in a refractory form as a result of their exposure to high temperatures, or where they are mixed with Cr, Si and Fe oxides which are intractable to normal dissolution. In such cases, low and erratic recoveries and nonreproducible results are likely when using wet ashing as the sole solubilization technique. The lack of reproducibility arises from the failure to achieve isotopic equilibrium between monitor and determinand.[70] For a comparison of wet ashing and fusion techniques in the determination of actinide radionuclides, the reader is referred to a specific study of this topic by Smith et al.[71]

One of the most commonly employed fusion methods involves the use of a potassium fluoride/pyrosulfate solution[72–74] in order to volatilize silica as SiF_4. Another technique is based on the dissolution of mineral samples with lithium metaborate.[75] The potassium fluoride/pyrosulfate fusion is carried out by sprinkling 3 g of anhydrous potassium fluoride over a 0.5-g sample of analyte, previously digested in a mixture of 0.5 mL of 16 M HNO_3 and 3 mL of 48% HF. The resulting ''cake'' is melted over a blast burner until a clear solution is obtained. Following cooling, 3.5 mL of concentrated H_2SO_4 are added and the solution heated until all the fluoride gases have been expelled. Upon the appearance of fumes of SO_3, the sample is treated with 2 g of sodium hydrogen sulfate in order to obtain a clear pyrosulfate solution. The latter is evaporated to near dryness, the residue re-dissolved in diluted HCl and then treated as described in section 3.4.3 below.

In the lithium metaborate fusion method, 2 g of spiked sample are combined with 6 g of $LiBO_2$ in a graphite crucible and heated at 950 °C for 30 min. Upon cooling, the $LiBO_2$ – silicate mixture thus formed is immediately dissolved in 1 M HNO_3, at which point it also can be coprecipitated as described below.

3.4.3 Coprecipitation

Following leaching or dissolution, the sample is filtered and the filtrate, containing the actinides of interest, is converted to the appropriate medium for a coprecipitation step in which these actinides are not only further preconcentrated but also purified from other leached elements such as the alkalis and alkaline earths. Coprecipitation either with $Fe(OH)_3$ in an alkaline medium,[36,38,76] or

with calcium oxalate in an acidic medium at a pH between 1.5 and 4.5[39,62,77,78] is often used. The latter is recommended for samples with a high calcium or phosphate content (e.g. bone), in which case the actinides do not coprecipitate efficiently with $Fe(OH)_3$.[79] Actinides can also be coprecipitated with rare-earth fluorides such as LaF_3, NdF_3 and CeF_3,[80,81] and with bismuth phosphate[82] and barium sulfate.[83,84]

To illustrate, consider the case of coprecipitation with ferric hydroxide. Here, the sample in solution is transferred to a beaker and an appropriate quantity of Fe^{3+} carrier (100 mg) added if necessary. This step may be omitted in the case of soils and sediments, as they already contain enough natural iron (in the case of water, which may have been treated at an earlier stage with iron or neodymium, further addition of carrier is unnecessary). Coprecipitation is effected by slowly adding an appropriate quantity of 35% ammonium hydroxide solution to the sample with constant stirring until a flocculant mass of brown $Fe(OH)_3$ precipitate is formed.[36] After adjusting the pH to at least 8, the sample is allowed to stand for 2 h at room temperature prior to centrifugation. The supernatant is discarded and the precipitate dissolved in a small quantity of either 16 M HNO_3 or 12 M HCl, depending on the purification procedure to follow. It is then evaporated to dryness and re-dissolved in 100 mL of the appropriate medium (e.g. 8 M HNO_3, 9 M HCl, 1 M HNO_3–90% methanol) prior to further purification.

4 RADIOCHEMICAL SEPARATION AND PURIFICATION

Although the sample pretreatment steps described above separate the actinides of interest from the bulk matrix, a large number of potential interferants (i.e. stable elements and natural or artificial radioelements) are still present in solution along with the actinides and must be removed if high quality, undegraded sources are to be obtained. Over the years, various separation schemes, primarily based on the use of ion exchange and solvent extraction techniques, have been proposed and adapted for different matrices and types of analyses.

4.1 Ion Exchange

The separation by ion exchange of actinides from other elements and from each other is based on differences in their sorption characteristics on conditioned ion exchange resins. Exchange resins are insoluble porous lattices with attached ionic functional groups linked to exchangeable counterions of opposite charge.[85] The charge of the functional group dictates whether the resin will act as a cation or anion exchanger. During the ion exchange process, when resin and sample are put in contact, counterions are replaced by sample ions having the same charge. Neutral molecules and those having the same charge as the functional group are not retained and pass through the resin without adsorption. The exchanged ions can selectively be desorbed by the introduction of an ion of higher affinity for the exchangeable sites or a high concentration of an ion of equivalent or lower affinity.

Typical cation exchange resins are formed from a cross-linked styrene–divinylbenzene copolymer with attached anionic functional groups, for example SO_3^{2-}, which forms the well known AG 50 W cation exchanger **(1)**. Attached to the functional group are exchangeable counterions opposite in charge to that of the functional group. The counterion is normally H^+, thus forming the sulfonic acid group SO_3H^- with the functional group SO_3^{2-}. Counterions are substituted by similarly charged sample cations during the cation exchange process because the counterions have a lower selectivity for the resin's functional groups than the sample ions to be sorbed. The cation exchange process can be summarized by Equations (1) and (2):

$$zR - SO_3H + Me^{z+} \Longleftrightarrow Me(R - SO_3)_z + zH^+ \quad (1)$$

$$\frac{[Me(R - SO_3)_z][H^+]^z}{[R - SO_3H]^z[Me^{z+}]} = K_c \gg 1 \quad (2)$$

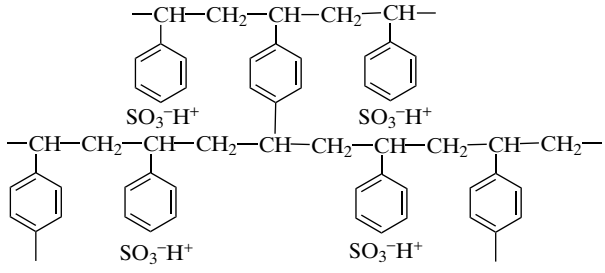

(1)

where z is the electric charge of the sample ion Me and R symbolizes the styrene–divinylbenzene organic group.

Many of the features described above are common to cation and anion exchange resins. The fundamental difference between a cation and an anion exchanger is that, in the latter case, the styrene–divinylbenzene copolymer is derivatized with a cationic functional group, for example, $CH_2N^+(CH_3)_3$, forming the AG 1 anion exchanger **(2)**. Anion exchange resins are normally available in two ionic forms, namely the nitrate form with NO_3^- as the counterion and the chloride form with Cl^- as the counterion. Conversion from one form into the other is usually achieved by gently washing the resin with an HCl or HNO_3 solution of the appropriate concentration.

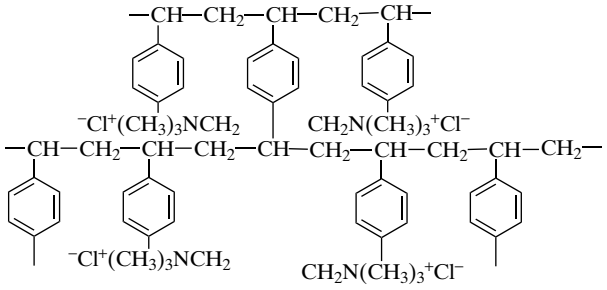

(2)

4.1.1 Separation of Thorium by Ion Exchange

Most of the procedures for the separation of thorium from other actinides are based on the inability of thorium to form chloride complexes. A clean separation of thorium from the rest of the actinides can be achieved by anion exchange from a nitric medium.[60,76] In this medium (typically 8 M HNO_3), thorium is strongly sorbed together with plutonium and neptunium, while iron, trivalent actinides, uranium, the bulk of rare earths and divalent metals, unable to form nitrate anionic complexes, pass through without being adsorbed.[35,86] Thorium can then be separated from plutonium and neptunium by elution of thorium using 7–8 M HCl. In this medium, thorium is not retained by the column, while plutonium and neptunium, which form strong chloride complexes, remain sorbed.

For samples free of plutonium and neptunium, cation exchange on an AG 50W resin from a 9 M HCl medium can be employed as an alternative. Thorium(IV) and cerium(III) are sorbed on the column, whereas americium, curium and other lanthanides pass straight through. Calcium, strontium and other divalent cations are not retained by the column either, because Equation (3)

$$2R - SO_3H + Me^{2+} \Longleftrightarrow Me(R - SO_3)_2 + 2H^+ \quad (3)$$

is not favored at high proton concentrations, as shown by Equation (4)

$$[Me(R - SO_3)_2] = \frac{K_c[R - SO_3H]^2[Me^{2+}]}{[H^+]^2} \cong \frac{\text{constant}}{[H^+]^2} \quad (4)$$

where the concentration of Me^{2+} can be considered constant by virtue of its continuous supply while the sample is being passed. Thorium can be eluted from the column with 0.75 M oxalic acid ($C_2H_2O_4$).[60]

4.1.2 Separation of Uranium by Ion Exchange

Separation of uranium from other actinides by anion exchange is based on its different sorption behavior in HNO_3 and HCl media.[60,87] In HCl medium (9 M), uranium (together with plutonium, neptunium, polonium and iron) is strongly adsorbed as the complex anion $UO_2Cl_4^{2-}$, while thorium and trivalent actinides pass through the column without adsorption. Elution of uranium is achieved using a 7–8 M HNO_3 solution. In this medium, uranium is not retained by the column, whereas plutonium and neptunium are.

A difficulty associated with this procedure is the co-elution of iron and polonium with the uranium fraction. To avoid problems associated with excessive iron at the source preparation stage and the presence of spectrometrically interfering polonium isotopes, the three elements are eluted sequentially using appropriate column volumes of 7–8 M HNO_3.[88] Iron is eluted first, followed by uranium and, finally, polonium. Alternatively, iron can be removed from the column (together with plutonium) by reducing these elements with iodide.[89] When co-eluted with a single HNO_3 wash, uranium and iron are often separated from each other by solvent extraction with isopropyl ether from an 8 M HCl medium.[60]

Iron can also be separated from uranium by sorption of uranium onto an anion exchange resin from an ammonium acetate medium at pH 4.5–5.0.[90] In this medium, iron passes through the resin without adsorption, while uranium is retained. Uranium can then be eluted with dilute HCl.

4.1.3 Separation of Neptunium by Ion Exchange

The levels of neptunium (in reality ^{237}Np) in the environment are low compared with those of most other actinides. This probably explains why comparatively few papers have been published over the years on methods for the assay of neptunium in environmental matrices.[9–11,91–93] When α-spectrometry is used to determine ^{237}Np at trace level, the key step is the purification of ^{237}Np from other interfering nuclides and, in particular, from ^{234}U, ^{231}Pa and ^{230}Th.

For example, Yamamoto et al.[9] improved on the method of Sakanoue[94] (which is based on solvent extraction of neptunium with tri-*n*-octylamine (TOA) and its coprecipitation with LaF_3) by introducing further purification steps involving anion exchange. In this method, neptunium(IV) is adsorbed from an 8 M HNO_3 medium saturated with $Al(NO_3)_3$. The column is washed successively with 8 M HNO_3, 10 M HCl and 0.1 M HI–10 M HCl to remove uranium, thorium and plutonium, respectively. Finally, neptunium is eluted from the column with 4 M HCl. Further purification (in order to eliminate any remaining traces of ^{234}U) can be effected by converting the eluate from the previous column to a 4 M CH_4COOH medium and passing it through a column of AG 1 anion exchange resin (acetate form). In this medium, neptunium is not adsorbed on the column, whereas uranium is strongly sorbed.[9]

In an alternative approach, favored by Schell and Tobin,[60] neptunium is sorbed, together with plutonium and thorium, from an 8 M HNO_3 medium and, after the elimination of thorium and plutonium by rinsing with 9 M HCl and 0.1 M NH_4I–10 M HCl, respectively, eluted with 3 M HCl and a few drops of HF.

4.1.4 Separation of Plutonium by Ion Exchange

Although plutonium is readily sorbed on both cation and anion exchangers from acidic media, anion exchange from a nitric or hydrochloric medium provides the most efficient way to separate this element from other actinides and has been adopted in the majority of separation schemes published in the literature. Indeed, the strong tendency of plutonium to form anionic complexes, the variation in the stability of these complexes with oxidation state and the reluctance of most metals to form anionic complexes, make it possible to separate plutonium from many impurities by adsorbing it as an anionic complex and then selectively desorbing it by controlling the acid concentration and elutant composition.

Plutonium in all valences and concentrations is readily adsorbed onto cation exchange resins from acidic solutions. However, the degree of adsorption varies depending on the oxidation state of plutonium (decreasing in the order Pu(IV) > Pu(III) > Pu(VI)) and the acid concentration (decreasing with increasing acid molarity). The common practice is, therefore, to adsorb plutonium from relatively dilute acid and elute it with a concentrated one. Mathew et al.[33] used this approach to separate plutonium from bismuth introduced via a bismuth phosphate coprecipitation step. Plutonium was sorbed from a 0.75 M HCl medium onto a Dowex 50 WX-8 resin (H^+ form), while Bi^{3+} and the alkalis and alkaline earths passed through without adsorption. Plutonium was then eluted with 12 M HCl and further purified by anion exchange.

As we have noted above, the most widely used separation schemes are based on anion exchange from HCl or HNO_3 medium. From the former, plutonium is adsorbed as the $PuCl_6^{2+}$ species. An efficient separation of plutonium can be achieved by adsorption from 9 M HCl.[95,96] In this medium, Pu(IV), Np(IV) and U(VI) are strongly sorbed, while Th(IV) and trivalent actinides pass through the column without adsorption. Iron and polonium, which also form anionic complexes in this medium, are co-adsorbed along with plutonium. After the column is rinsed with 9 M HCl in order to remove nonadsorbed elements from the dead volume, iron, uranium and polonium are sequentially stripped using 7.2–8.0 M HNO_3.[24,88] Elution of plutonium is then achieved using an appropriate elutant. Although 1.2 M HCl–0.7% H_2O_2 is widely used,[36,38,41,88,96,97] alternative elutants must be employed if pure solutions of plutonium are required, as in this medium neptunium is co-eluted along with plutonium. Selective desorption of plutonium has been achieved using 9 M HCl–0.05 M NH_4I[24] or 6 M HCl–0.25 M HF.[87,95] The iodide ion in strong hydrochloric media reduces plutonium from the tetravalent to the trivalent state. In the trivalent state, plutonium does not form anionic complexes and is released from the column, while Np(IV) and U(VI) remain adsorbed.

A similar scheme is appropriate for the purification of plutonium from a nitric medium (normally 8 M HNO_3). In this medium, tetravalent plutonium is left as the hexanitrato complex, $Pu(NO_3)_6^{2-}$, readily exchangeable on the anion exchange column, while iron, trivalent actinides, uranium, the bulk of rare earths and divalent metals (unable to form anionic complexes) pass through without being adsorbed.[35,86] Neptunium and thorium, which do form anionic complexes, are co-adsorbed along with plutonium. Separation of thorium from plutonium is achieved by rinsing the column with a moderately concentrated (7–8 M) solution of HCl (thorium is not adsorbed in this medium and, thus, is eluted from the column). As in the previous scheme, several solutions can be used to elute plutonium, such as 12 M HCl–0.1 M NH_4I,[33,41,60,69,98] 10 M HCl–0.1 M HI,[39] 0.36 M HCl–0.01 M HF[99] and 0.35 M HNO_3.[100] Finally, if required, neptunium can be eluted from the column with, for example, 1 M HCl.[101]

4.1.5 Separation of Americium and Curium by Ion Exchange

Transplutonium elements such as americium and curium can be preconcentrated in strongly basic cation exchangers from dilute solutions of HCl or HNO_3,[102] or from 2–6 M HCl in the presence of at least 40% alcohol.[103] Unfortunately, the large number of other elements co-adsorbed with these radioelements effectively limits the use of cation exchangers to the preconcentration stage rather than the purification stage. Nevertheless, good separation of the transplutonium elements from the lanthanides (but not from other elements) has been achieved by sorption in a 1 M HCl medium and elution with 6 M HCl.[104] Under these conditions, americium and curium are rapidly eluted from the column, while the lanthanides are only very slowly separated.

Separation of certain actinides from lanthanides can be achieved using cation exchangers at high temperatures and complexing agents such as lactic acid or α-hydroxyisobutyric acid.[42,78,105] However, this technique requires sophisticated equipment and is not practical for the routine analysis of environmental samples. Anion exchange represents a much simpler and more versatile alternative.

Transplutonium elements are only slightly sorbed on anion exchangers in dilute mineral acid media. However, they are strongly sorbed from solutions containing 80–90% alcohol at acid concentrations of 0.1–1 M.[103,106] This property has been used to separate americium and curium from other elements which are poorly sorbed under these conditions (i.e. elements with $Z < 56$, as well as Hf, Ta, Re, U and Np). The most effective results have been obtained with solutions of methanol containing HNO_3. For example, Guseva et al.[106] used 0.5 M HNO_3–95% methanol, while other authors have employed 1 M HNO_3–93% methanol[38,61,97] or 1 M HNO_3–90% methanol.[36,78] Pu(IV), Th(IV), Pb(II) and Bi(II), which also have high distribution coefficients in these media, are co-adsorbed with the transplutonium nuclides and, accordingly, are normally separated prior to this step. Lighter lanthanides are also co-adsorbed, but heavier lanthanides (above gadolinium) pass through without sorption.[102] Unfortunately, some of the lighter lanthanides (i.e. cerium, lanthanum and neodymium) dominate the total lanthanide content of most environmental samples and an additional purification step must be included in order to eliminate them. One such step is to rinse the column with an ammonium thiocyanate solution in methanol[61,97,107] or, alternatively, to pass the sample through a second anion exchanger in nonaqueous HCl–methanol medium.[78]

Americium and curium are finally co-eluted using 1.5 M HCl–86% methanol,[36] 0.5 M HNO_3–80% methanol[103] or 8 M HNO_3.[78] The use of HCl–methanol has the added advantage of not eluting any traces of Pu, Po, Th, U, Bi and Pb that might still be on the resin.[61,97]

4.2 Solvent Extraction

Solvent extraction techniques[108] are based on the capacity of different elements to form stable complexes with organic ligands. The resulting complexes are insoluble in water, but readily soluble in benzene, xylene, cyclohexane, carbon tetrachloride, *n*-heptane and other organic solvents. When nonmiscible aqueous and organic phases are equilibrated with each other, these large nonpolar or weakly polar uncharged complexes are preferentially transferred from the aqueous to the organic phase,[109] while uncomplexed elements remain in the aqueous phase.

The extractability of the actinides diminishes as a function of valency in the sequence IV > VI > III > V and, for this reason, actinides are usually extracted in the tetravalent state. The efficiency of the extraction also depends on the acidity of the aqueous phase. By varying the acidity and composition of this phase, selective extraction of actinides and other elements which form complexes of differing stabilities can be effected.

Solvent extraction of actinides is usually carried out by means of an organic complexing agent in combination with an organic solvent. The choice of complexing agent depends on the nature of the element to be extracted and includes: amines such as trilauryl amine (TLA), TOA, tri-*iso*-octylamine (TIOA) and methyl-tri-capryl ammonium chloride (Aliquat-336); organophosphorus compounds such as di-2-ethyl hexyl phosphoric acid (HDEHP), tri-*n*-octyl phosphine oxide (TOPO), tri-*n*-butylphosphate (TBP) and dibutyl-*N*,*N*-diethyl carbamyl phosphate (DDCP); and other compounds such as β-diketone thenoyl trifluoroacetone (TTA) and ethylenediaminetetraacetic acid (EDTA).

4.2.1 Separation of Thorium by Solvent Extraction

As in the case of ion exchange, the inability of thorium to form stable chloride complexes has been used to separate thorium from iron and other actinides by solvent extraction.[76] In a separation scheme using TIOA in xylene (1 : 9 by volume), thorium in an 8 M HCl medium remains in the aqueous phase, together with trivalent actinides, while plutonium and neptunium are extracted into the organic phase.[60] Following contact of the aqueous phase with TOPO in heptane, thorium is extracted into the organic phase, while americium and curium remain in the aqueous phase. The now purified thorium is finally back-extracted using 0.6 M H_2SO_4.

TBP in a range of organic solvents is also used frequently to extract thorium, together with uranium, neptunium and plutonium, from nitrate solutions.[76] Back-extraction of thorium is achieved by contacting the organic phase with an HCl solution. Alternatively, thorium can be complexed with EDTA before extraction with TBP, leaving thorium in the aqueous phase. HDEHP in toluene is also effective in extracting thorium from nitric, hydrochloric and perchloric acid solutions, as is TTA in benzene or toluene. Extraction at pH 1–2 from the latter gives an efficient separation of thorium from aluminum and uranium, as well as from the alkaline and rare earths.[76] Separation of the extracted actinides from one another can be achieved by selective stripping from the organic phase with the appropriate mineral acid and molarity, and/or adjustment of oxidation state.[110]

4.2.2 Separation of Uranium by Solvent Extraction

Solvent extraction from acid solutions into organic acids, ketones, ethers, esters, alcohols and organophosphorus compounds has been used extensively to separate uranium from other actinides and from a wide range of other elements. Extraction of uranium from a nitric medium (3–6 M) into TBP is particularly useful, as it allows separation of this element from iron, which is problematic by

anion exchange (see section 4.1.2). In nitric medium, iron (together with thorium and protactinium) is co-extracted with uranium. However, iron can be removed by stripping the organic phase with dilute nitric acid.[76] Uranium can be recovered from the organic phase by back-extraction into solutions such as ammonium carbonate, sodium carbonate and oxalic acid.

TIOA in xylene can be used to extract uranium from an 8 M HCl medium.[60,111] After the addition of a few drops of 30% H_2O_2, uranium, together with plutonium and neptunium, are extracted into the organic phase, leaving thorium and trivalent actinides in the aqueous phase. Plutonium, neptunium and uranium are then sequentially stripped from the organic phase with 0.05 M NH_4I–8 M HCl, 0.02 M HF–4 M HCl and 0.1 HCl, respectively.

TTA is a suitable complexing agent for uranium, forming a complex which can be extracted from aqueous solutions at pH 3 or higher using benzene as solvent.[76] TOPO in cyclohexane or carbon tetrachloride has also been used for the extraction of uranium from a nitric medium and for the separation of this element from iron.

4.2.3 Separation of Neptunium by Solvent Extraction

Separation of neptunium from other actinides by solvent extraction can be achieved using a TOA–xylene (1 : 10 by volume) solution as extractant.[94] In this method, HI is used to reduce neptunium to the tetravalent state in 10 M HCl medium under heat (70–80 °C). After cooling, neptunium is extracted into TOA–xylene medium, while thorium, americium and plutonium (converted to the trivalent state by HI) remain in the aqueous phase. The organic phase, containing neptunium and uranium, is washed with 10 M HCl and neptunium is back-extracted with 0.1 M HF–1 M HCl.

A similar procedure, based on the extraction of neptunium, plutonium and uranium into TIOA–xylene, followed by selective stripping of these elements, has already been discussed (section 4.2.2). An alternative approach, employing the TIOA–xylene combination as a first purification step prior to anion exchange, has recently been reported by Chen et al.[92] Here, following successive coprecipitation steps (see section 3.4.3) with Fe^{3+} and $K_2S_2O_5$, neptunium is extracted from a 9 M HCl medium, leaving plutonium in the aqueous phase. Neptunium is then back-extracted with 2 M HCl and a few drops of NaCl, co-precipitated as $Fe(OH)_2$–$Np(OH)_4$, and dissolved in 12 M HCl prior to purification by anion exchange.

Sumiya et al.[112] used solvent extraction with TTA–xylene to purify neptunium following anion exchange from an acetic medium, while Rameback and Skalberg[113] reported an effective method to separate transuranics from large amounts of uranium using HDEHP. In Rameback and Skalberg's procedure, uranium, neptunium, plutonium, americium and curium are extracted into HDEHP from a nitric medium. After the elimination of americium and curium with a 5 M HNO_3 wash, neptunium is back-extracted with 1 M HNO_3 containing hydroxylamine hydronitrate, leaving plutonium in the organic phase. Plutonium, if required, can be recovered by back-extraction with 3 M HCl containing Ti(III). This method has the advantage of completely eliminating uranium from the neptunium assay, and has been applied successfully to the determination of ^{237}Np in samples of spent nuclear fuel by ICPMS.[113]

4.2.4 Separation of Plutonium by Solvent Extraction

Solvent extraction of plutonium is most often carried out from a nitric medium, since the partition coefficients are generally higher than in a hydrochloric medium. Plutonium separation can be accomplished with a number of complexing agents, mainly organophosphorus compounds, amines and several chelating agents.

Among the organophosphorus compounds, HDEHP has gained widespread use and popularity in the separation of actinides. In a solvent extraction procedure introduced by Murray and Statham[104] and designed to separate plutonium, americium and curium from environmental samples of seawater, sediment and marine biota, plutonium is efficiently extracted from an 8 M HNO_3 medium into a 0.45 M solution of HDEHP in *n*-heptane, leaving trivalent americium and curium in the aqueous phase. The organic phase, containing the plutonium, is washed with several portions of 9 M HCl to remove thorium and much of the iron, prior to the back-extraction of plutonium using a freshly prepared solution of 8 M HCl–0.1 M NH_4I. In another scheme (for the separation of transuranium nuclides from large amounts of uranium), neptunium, plutonium, americium, curium and uranium are all extracted from HNO_3 medium into HDEHP, after which the transuranium nuclides are back-extracted sequentially.[113] Americium and curium are back-extracted with 5 M HNO_3, neptunium with 1 M HNO_3 containing hydroxylamine hydronitrate and, finally, plutonium with 3 M HCl containing Ti(III). In yet another separation scheme, actinides and lanthanides are extracted from an $HClO_4$ medium into HDEHP.[84,110] Trivalent actinides and lanthanides are stripped from the organic phase with 8 M HNO_3, before plutonium is back-extracted with a 4 M HNO_3 solution after reduction to the trivalent state with 2,5-di-*tert*-butylhydroquinone.

The most useful amines are tertiary amines such as TLA, TOA, TIOA and Aliquat-336. Singh et al.[79,114] used TLA as an extractant for the separation of uranium, thorium and plutonium from environmental

and biological samples. In this method, uranium and plutonium are extracted from 10 M HCl into a 20% TLA solution in xylene, leaving thorium in the aqueous phase. Plutonium is back-extracted by reduction to Pu(III) using a 0.05 M NH_4I solution in 8 M HCl, leaving uranium in the organic phase. Alternatively, plutonium in the tetravalent state can be extracted with 25% TLA in xylene from a 3 M HNO_3 medium.[70] The organic phase must be washed twice with 10 M HCl to remove traces of thorium and then with 8 M HNO_3 to remove iron and uranium, before plutonium is stripped from the organic phase using 2 M H_2SO_4.

Hashimoto et al.[115] used TOA in xylene to separate trace amounts of plutonium from macro amounts of uranium in 8 M HNO_3 solution. Butler[111] developed a technique for the separation of plutonium, neptunium and uranium using TIOA. Chloride complexes of the three elements are extracted from an 8 M HCl medium into a 10% TIOA solution in xylene after the addition of a few drops of 30% H_2O_2 to adjust plutonium to the tetravalent state. Because thorium is unable to form anionic chloride complexes, it remains in the aqueous phase. Plutonium is stripped from the organic phase by reduction to Pu(III) using a warm, freshly made solution of 8 M HCl–0.05 M NH_4I, leaving neptunium and uranium in the organic phase.

Fisenne and Perry[116] used Aliquat-336 for the separation of plutonium from biological samples. Here, tetravalent plutonium is extracted from 8.5 M HNO_3 into a 30% Aliquat-336 solution in toluene. Uranium and thorium are then eliminated by washing with HNO_3 and HCl, and plutonium back-extracted with 1 M HCl–0.01 M HF.

Of the many chelating agents which form extractable complexes with plutonium, TTA in benzene, xylene or carbon tetrachloride is the most often used. Plutonium(IV) is most efficiently extracted with this agent from nitric, hydrochloric or perchloric medium. The maximum extraction of plutonium with 0.1 M TTA in benzene takes place in the acidity range 0.5–1.0 M. Many ions accompany the extraction, including Zr(IV), Hf(IV), Ce(IV), Sn(IV), Th(IV), U(VI), Np(IV), Nb(V) and Fe(III).[117,118] However, by adjusting the acidity of the solution it is possible to separate Pu(IV) from a number of other elements. For example, Bunzl and Kracke[77] used TTA as extractant in the determination of plutonium in biological samples. In this method, plutonium is extracted from 1 M HNO_3 into 5% TTA in benzene and back-extracted with 8 M HCl.

4.2.5 Separation of Americium and Curium by Solvent Extraction

Separation of americium and curium by solvent extraction is usually performed from an acidic medium using an organic phosphoric acid derivative as complexing agent. As noted above, Murray and Statham[104] used HDEHP to separate plutonium, americium and curium from environmental samples. In this method, the lack of extractability of trivalent transplutonium elements from a strongly acidic solution (8 M HNO_3) in HDEHP is first used to separate americium and curium from other readily extractable elements (e.g. plutonium and uranium). The americium and curium are further purified from other interfering nuclides by extraction into HDEHP from an HNO_3 medium adjusted to pH 2, followed by repeated washes with 0.075 M HCl and back-extraction with 9 M HCl. Other schemes using organophosphorus reagents include the extraction of transuranium elements into HDEHP from a perchloric medium, with separation of americium and curium from plutonium and thorium by back-extraction into HNO_3,[84,110] and separation of americium and curium from thorium in an 8 M HCl medium using TOPO.[60] Bifunctional organophosphates such as DDCP possess a much greater extractability towards trivalent americium and curium, and have also been employed in numerous separation schemes in order to remove major interfering elements such as calcium, lead and radium.[38,41,59,101]

A number of extraction methods for the purification of americium make use of the more unusual oxidation states of this element, that is Am(V) or Am(VI). However, the conversion and stability of americium to these higher oxidation states is difficult and their application to large numbers of environmental samples is unrealistic. A detailed review of methods for the preparation of americium in an oxidized form and its extraction in numerous media can be found in Schulz.[119]

4.3 Example of a Scheme for the Separation and Purification of Actinide Nuclides

An example of a successful scheme combining anion exchange and solvent extraction for the separation and purification of actinides in environmental samples is described in detail in this subsection. The complete scheme is summarized in Figures 2–4, inclusive. In brief, anion exchange from an 8 M HNO_3 medium[35] is first used to separate and purify thorium, plutonium and neptunium. The raffinate from this step is converted to 9 M HCl medium and passed through a second anion exchange column in order to separate and purify uranium, and eliminate iron.[60] The raffinate from this second column is, in turn, converted to HNO_3–methanol medium and passed through a third anion exchange column in order to purify americium and curium.[36] If further purification of the separated actinides is required, this can be achieved by, for example, solvent extraction into TIOA–xylene[111] or HDEHP–heptane.[104]

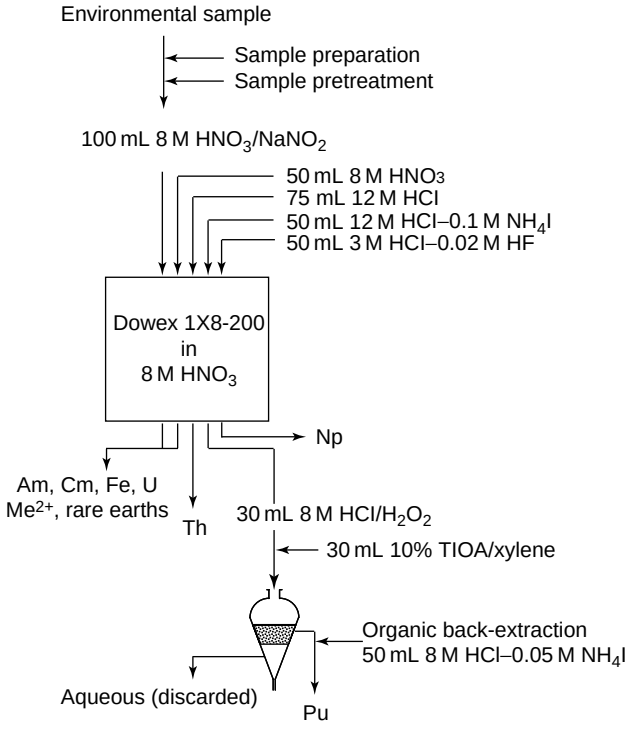

Figure 2 Example of a separation and purification scheme for plutonium, thorium and neptunium nuclides by ion exchange and solvent extraction.

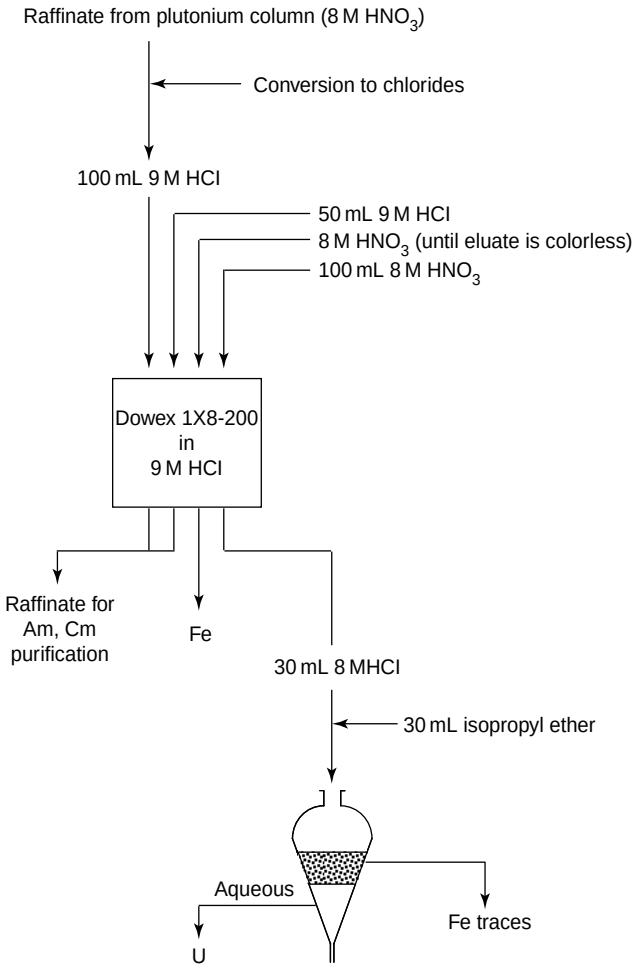

Figure 3 Example of a separation and purification scheme for uranium nuclides by ion exchange and solvent extraction.

4.3.1 Detailed Method

Following appropriate preconcentration steps (see section 3), the sample is conditioned in 8 M HNO_3 medium; this is the starting point of the separation scheme shown in Figure 2. Upon conversion of plutonium to the tetravalent state by the addition of $NaNO_2$, the sample is passed through an anion exchange resin (e.g. Dowex 1X8, 80 mm by 7 mm diameter), previously conditioned in 8 M HNO_3, at a flow rate of 1–2 mL min^{-1}. Following the sample, an additional 50 mL of 8 M HNO_3 is passed to purge the dead volume of the column. Raffinate and wash are retained for subsequent uranium, americium and curium analysis. Thorium, plutonium and neptunium are then sequentially stripped from the column using 75 mL of 12 M HCl, 50 mL of warm 12 M HCl–0.1 M NH_4I, and 50 mL of 3 M HCl–0.02 M HF, respectively.

If further purification of plutonium is indicated, some additional steps are required. Firstly, the plutonium eluate is evaporated to dryness, the residue dissolved in HNO_3 (to eliminate iodine) and HCl, evaporated to dryness again, and finally redissolved in 30 mL of 8 M HCl after conversion to the chloride form by successive additions of 12 M HCl. A few drops of H_2O_2 are added to convert plutonium to the tetravalent state, after which the solution is shaken vigorously for 5 min with TIOA–xylene extractant, previously equilibrated with fresh 8 M HCl. Plutonium and any traces of neptunium and uranium are extracted into the organic phase. The aqueous phase, containing small traces of iron and thorium, is discarded. Plutonium is then back-extracted from the organic phase using a freshly made (and warm) solution of 8 M HCl–0.05 M NH_4I.

A similar approach can be used to further purify neptunium. In this case, neptunium is extracted into TIOA–xylene from an 8 M HCl medium together with any traces of plutonium and thorium. Plutonium is eliminated first by stripping with an 8 M HCl–0.05 M NH_4I solution as above, and neptunium subsequently recovered by back-extraction into a 4 M HCl–0.05 M HF solution.

Purification of uranium from americium and curium is carried out by anion exchange in a 9 M HCl medium (Figure 3) as follows. The combined 8 M HNO_3 raffinate and wash from the first exchange column is evaporated

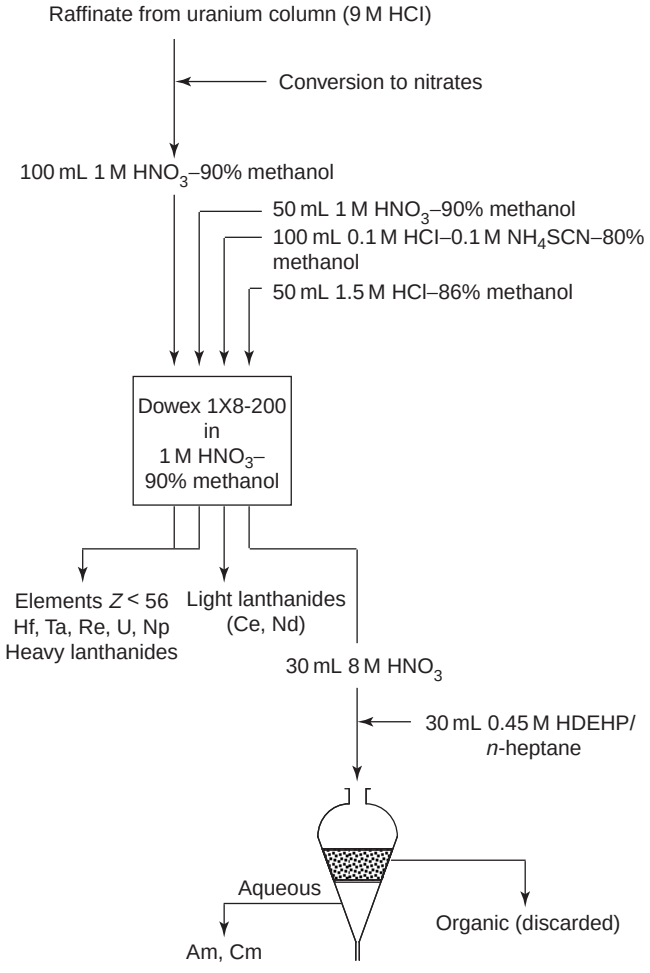

Figure 4 Example of a separation and purification scheme for americium and curium nuclides by ion exchange and solvent extraction.

to dryness, redissolved in 100 mL of 9 M HCl and passed through an anion exchange column (e.g. Dowex 1X8, 80 mm by 7 mm diameter) previously equilibrated in the same medium. Uranium, together with iron and polonium, are adsorbed in the column, while americium, curium and the lanthanides are not. Following a 50-mL 9 M HCl wash to clear the dead volume of the column, iron, uranium and polonium are sequentially eluted using appropriate quantities of 8 M HNO_3. In samples with high iron content, 8 M HNO_3 is passed until the eluate from the column is almost colourless, indicating the elimination of most of the iron. This fraction of the eluate can be discarded. An additional 100 mL of 8 M HNO_3 is then passed to elute uranium. As there may still be some small traces of iron present in the eluate, further purification can be accomplished by solvent extraction of residual iron into isopropyl ether from an 8 M HCl medium.

The purification of americium and curium is also done by anion exchange, this time from an HNO_3–methanol medium (Figure 4). The combined 9 M HCl raffinate and wash from the second exchange column is evaporated to dryness, redissolved in 100 mL of 1 M HNO_3–90% methanol and passed through an anion exchange column (e.g. Dowex 1X8, 80 mm by 7 mm diameter) previously equilibrated in this medium. A 100-mL wash of 0.1 M HCl–0.1 M NH_4SCN–86% methanol is then passed through the column to elute lanthanides as well as any traces of uranium. Americium and curium are finally eluted from the column using 50 mL of 1.5 M HCl–86% methanol.

Further purification of americium and curium can be effected by an additional solvent extraction step. By converting the eluate to an 8 M HNO_3 medium and shaking with 0.45 M HDEHP in *n*-heptane, residual traces of uranium, plutonium and thorium are extracted into the organic phase, while americium and curium are left in the aqueous phase.

In summary, the scheme described above permits the simultaneous determination of thorium, uranium, neptunium, plutonium, americium and curium, and has proved very reliable for the assay of actinides in a wide range of environmental matrices, for example, seawater, soil, sediment and biological materials. Chemical recoveries in the range 70–90% are routinely achieved with this method.

4.4 Extraction Chromatography

With the development of new, highly selective resins in recent years, the technique of extraction chromatography, which combines the selectivity of solvent extraction with the speed and convenience of chromatography, has been used increasingly in the low-level assay of actinides.[120,121] In extraction chromatography, an organic extractant (e.g. TOPO, HDEHP, TBP, Aliquat-336) is supported on an inert porous resin (e.g. microporous polyethylene) by adsorption or by the capillary effect. The coated resin is transferred to a column and preconditioned in the usual manner. As the sample is passed, the actinide(s) of interest is selectively extracted onto the resin. After elimination of impurities (as in ion-exchange chromatography), actinides are eluted (back-extracted) using appropriate elutants (extractants).

Resins using the quaternary amine Aliquat-336 as extractant (e.g. Eichrom's TEVA resin) have proved to be very effective for the separation of tetravalent actinides as anionic complexes, and for the separation of most actinides from interfering lanthanides.[120,122,123] Resins using dianylamylphosphonate as extractant (e.g. Eichrom's UTEVA resin) have shown great selectivity for uranium and tetravalent actinides, extracted as neutral complexes,[122,124–128] while resins based on the synergistic effect of a combined octylphenyl-*N*,*N*-di-isobutyl carbamoylphosphine oxide/TBP extractant

(Eichrom's TRU resin) have proved highly specific for the extraction of transuranium nuclides from a nitric medium.[17,122,129–131]

Despite the higher cost of commercial extraction chromatography resins compared to conventional ion exchange resins and the implications in quality assurance/quality control terms which the implementation of new protocols in accredited laboratories represent, the reduction in analysis time, reagent costs and resulting chemical wastes can offer attractive advantages over more traditional techniques.

5 CHEMICAL SPECIATION ANALYSIS

It has long been recognized that the mobility and bioavailability of any element in the environment, particularly the aquatic, is strongly influenced by its physical and chemical form. Of no element is this more true than for plutonium, which can exist in solution in at least four different oxidation states simultaneously. These states include plutonium(III), plutonium(IV), plutonium(V) and plutonium(VI), of which plutonium(IV) and plutonium(V) predominate in oxygenated waters as the reduced (hydrolyzed) $Pu(OH)_4$ and the oxidized PuO_2^+ forms, respectively.[132–134] While plutonium in reduced form is highly particle-reactive, being quickly adsorbed by suspended particles, sediments and natural colloids, plutonium in oxidized form is much more soluble and can be transported over great distances within the solution phase. Clearly, accurate data on the speciation of plutonium or, for that matter, any other element exhibiting multivalent behavior, are essential when there is a need to model its transfer within the aquatic environment and predict its likely impact on man.

Field studies have confirmed that, in the absence of complexing agents, oxidized forms of plutonium predominate in the dissolved ($<0.45\,\mu m$) phase in natural waters.[135–138] The presence of significant levels of complexing agents (such as dissolved organic carbon) results in a redistribution in the partitioning of plutonium between the dissolved and particulate phases, and in the proportions of the various chemical species present in the dissolved phase. According to theoretical speciation calculations, Pu(III) should predominate in the dissolved phase under reducing conditions.[139,140] Indeed, colloidal organic carbon at concentrations commonly encountered in natural waters ($1–20\,mg\,L^{-1}$) has been shown to strongly interact with Pu(III,IV) species and inhibit their sorption onto suspended particles.[141,142] Speciation studies of plutonium adsorbed onto suspended particulate and sediment have shown plutonium to be present almost entirely in reduced chemical form.[135,138]

In the case of americium and curium, it is well established that the solution chemistry of both elements is dominated by the trivalent Am(III) and Cm(III) forms. Indeed, the dominance of highly insoluble forms of americium and curium such as $Am(OH)_2^+$ and $Cm(OH)_2^+$ in alkaline solutions is predictable.[143] Measurements on the chemical speciation of americium in filtered water from the north-eastern Irish Sea appear to confirm that americium is present almost exclusively as Am(III).[144] Accordingly, a much greater proportion of americium (and curium) than plutonium can be expected to be in a highly insoluble, hydrolyzed form, leading to a stronger affinity for suspended particulate matter, colloids and sediments.

In solution, thorium exists only in the tetravalent state, characterized by a strong sedimentary affinity. In contrast, under normal oxidizing conditions, hexavalent $(UO_2)^{2+}$ is the stable oxidation state of uranium and presents in the form of soluble carbonate complexes.[145] Under moderately strong reducing conditions, however, uranium is reduced to the tetravalent state. The same is true for neptunium, which in normal oxidizing conditions exists in the pentavalent form as $(NpO_2)^+$, but can be reduced to Np^{4+} under reducing conditions.[4]

The determination of the oxidation state distribution of actinides in aqueous media is based on the use of multiple yield monitors (ideally, one for each of the oxidation states present) throughout the entire oxidation state separation procedure. To illustrate, we shall consider the case of plutonium in a natural water system in some detail. For many purposes, ''reduced'' plutonium can be regarded as the sum of the plutonium in oxidation states (III) and (IV), and ''oxidized'' plutonium as the sum in oxidation states (V) and (VI). This assumption is based on the similarity in behavior between plutonium(III) and (IV) species and reduced elements such as La(III) and Th(IV), and between plutonium(V) and (VI) species and oxidized elements such as Np(V) and U(VI).[133] These similarities constitute the basis of a technique for the sequential separation of the lower (III,IV) and higher (V,VI) oxidation states of plutonium in natural waters by coprecipitation with rare earth fluorides,[135] which is described in some detail below.

To separate the reduced and oxidized species of plutonium, it is necessary to use two plutonium yield monitors in pure, but different, chemical forms. These chemical forms should be identical to those of the species under examination. This allows corrections to be made for the (usually) small carry-overs that invariably take place between the reduced and oxidized species in the course of analysis. The recommended chemical forms are Pu^{4+} for the reduced species and either PuO_2^+ or PuO_2^{2+} for the oxidized species. A full description of the methods use to prepare plutonium tracers in these oxidation states

can be found elsewhere;[41,69,146] here, a brief outline only will be given.

Essentially, a purified solution of, for example, ^{242}Pu(IV), is prepared by passing ^{242}Pu in 8 M HNO_3 through an anion exchange column (e.g. AG 1X8, NO_3^- form). Plutonium is then eluted with an appropriate volume of 0.1 M HNO_3 and the eluate combined immediately with an equal volume of 16 M HNO_3 to give ^{242}Pu(IV) in approximately 8 M HNO_3.

Preparation of the oxidized tracer requires the use of an electrolytic cell (Figure 5). A solution of 2–3 mL of 0.8 M HNO_3, containing a different plutonium isotope to that of the reduced tracer (e.g. ^{236}Pu or ^{244}Pu), is placed in the anodic compartment of the cell, whilst the cathodic compartment is filled to the same level with pure 0.8 M HNO_3. Direct current is passed through the cell at 3.5 V for about 16 h, leaving some two-thirds of the activity added at the outset in the anodic acompartment as oxidized Pu(VI).

Solutions of pentavalent plutonium can be prepared by extracting Pu(VI) from a dilute acetate solution with TTA, irradiating the organic phase with fluorescent light in order to induce the photo-reduction of plutonium(VI) to plutonium(V), and back-extracting Pu(V) with 0.1 M sodium acetate.[147] It is also possible to prepare Pu(V) directly by electrolytic oxidation in 1 M HCl medium.

Returning to the determination of the oxidation state distribution of plutonium in, for example, natural waters, the first action is to microfilter the sample, preferably in situ. Following this, and without delay, two isotopic monitors in different oxidation states, e.g. ^{242}Pu(IV) and ^{236}Pu(VI), are added to the sample which is made 0.8 M in HNO_3, 0.25 M in H_2SO_4, 0.0005 M in $K_2Cr_2O_7$ and 0.0007 M in Nd. The addition of a holding oxidant ($K_2Cr_2O_7$) prevents the reduction of the higher oxidation states by reducing agents used in the next step, in which neodymium, added as $Nd(NO_3)_3$, is precipitated as NdF_3 having made the sample 0.25 M in HF. The precipitate, containing the reduced plutonium species only, is collected by filtration and retained for analysis. The oxidized species, which have remained in solution, are subsequently reduced by the addition of ferrous ammonium sulfate and coprecipitated with NdF_3 by again making the sample 0.0007 M in Nd. This precipitate is also collected by filtration and retained for analysis of the oxidized fraction. Excellent separation between the oxidation state groups is normally achieved, with carry-overs, typically, being a few percent or less. Although small, these carry-overs should be corrected for; this is easily done, as they are readily quantifiable.

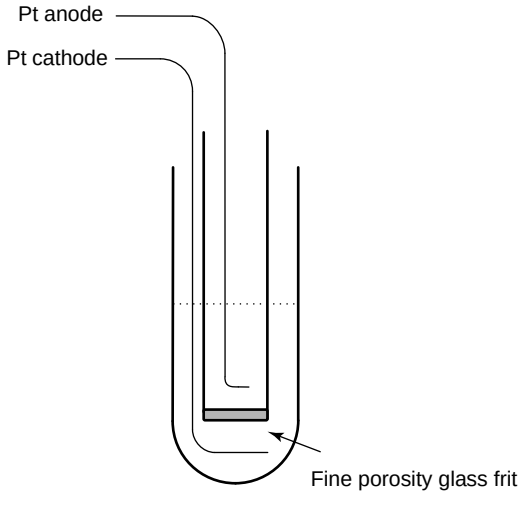

Figure 5 Electrolytic cell used in the preparation of a Pu(VI) yield monitor.

It should be stressed that the sequential coprecipitation method described above cannot distinguish plutonium(III) from plutonium(IV), nor plutonium(V) from plutonium(VI), due to the oxidizing effect of $K_2Cr_2O_7$. Although calculations based on thermodynamic considerations suggest that PuO_2^+ is the predominant oxidized species (at least in seawater), confirmation of this prediction is not possible using the above technique alone. However, in principle, it is possible to determine the percentage of plutonium(VI) in the oxidized fraction if a separate experiment is performed, using an identical sample of water, in which plutonium in oxidation states (III), (IV) and (V) is removed by the addition of excess calcium carbonate (~20 mg of $CaCO_3$ per liter of sample is usually sufficient). Since the scavenging of plutonium(V) by calcium carbonate is not 100% efficient, the experiment should also be carried out with dual monitors, one to label the Pu(V) fraction and the other Pu(VI). It is assumed that the removal of Pu(III) and Pu(IV) by $CaCO_3$ is quantitative. Alternatively, Pu(VI) can be scavenged with Pu(III) and Pu(IV) using hydrated silicon dioxide.[133]

Upon return to the laboratory, the filters containing the precipitate of each plutonium fraction are separately dissolved in concentrated HNO_3 in the presence of small quantities of boric acid (this facilitates the elimination of excess fluoride). Each solution is then diluted to an appropriate volume with distilled water and neutralized with ammonia to a pH of ~9. The resulting precipitates are separated by centrifugation and redissolved in 16 M HNO_3. After evaporation, the residues are finally redissolved in 8 M HNO_3, and 1 mL of saturated $Al(NO_3)_3$ solution in 8 M HNO_3 added to enhance the separation of plutonium from neodymium in the subsequent purification step (anion exchange as described in section 4.3).

6 SOURCE PREPARATION

6.1 Sources Suitable for Counting by α-Spectrometry

Following radiochemical purification, the actinide of interest must be mounted on a suitable surface or substrate prior to α-spectrometry. This is one of the most critical steps in the whole process, as the quality of the source is often the limiting factor in determining the quality of the observed spectrum.[148] Ideally, a source of good quality should approximate closely to a near infinitely thin, massless, uniformly active deposit in order to minimize energy degradation caused by (1) energy-loss straggling, (2) variations in the path lengths of the emitted particles due to nonuniformity in the thickness of the source, and (3) local variations in the stopping power due to the inhomogeneous composition of multi-element sources.[149]

Several techniques have been developed to prepare sources of spectrometric quality including vacuum sublimation, electrospraying, evaporation, direct or electrostatic precipitation and electrodeposition. Although the best sources are produced by vacuum sublimation,[150] the low efficiency of this method (1–25% depending on the active source area and source uniformity required) has limited its use in the analysis of low activity environmental samples which demand near quantitative recovery.[151] Instead, electrodeposition has been the method most commonly employed to obtain undegraded spectra with good deposition yields, providing rugged sources suitable for storage for long periods of time.

Electrodeposition of actinides from an aqueous solution involves the cathodic deposition of their hydrolysis products following the passage of a direct current through an appropriate electrolyte.[151] Plating from a variety of different electrolyte media has been reported in the literature, though acidic solutions containing ammonium sulfate,[152,153] ammonium chloride[154] or ammonium chloride/ammonium oxalate mixtures[69,155] have become the most widely used. While a sulfate medium is reported to yield marginally higher recoveries, it has been suggested that better quality sources are produced with an ammonium chloride/ammonium oxalate medium.[24] Further, plating from the latter requires somewhat shorter plating times and is not as critically dependent on pH. Accordingly, we have chosen to discuss electrodeposition from a chloride/oxalate medium here.

A typical electrolytic cell is depicted in Figure 6. It consists of a high quality, polypropylene sleeve (disposable), threaded into a stainless-steel support, which acts as both an electrical contact and a heat sink. A highly polished stainless-steel disk (20-mm diameter), mounted on this support, acts as the cathode. The anode is made of platinum and consists of a thin flat disc,

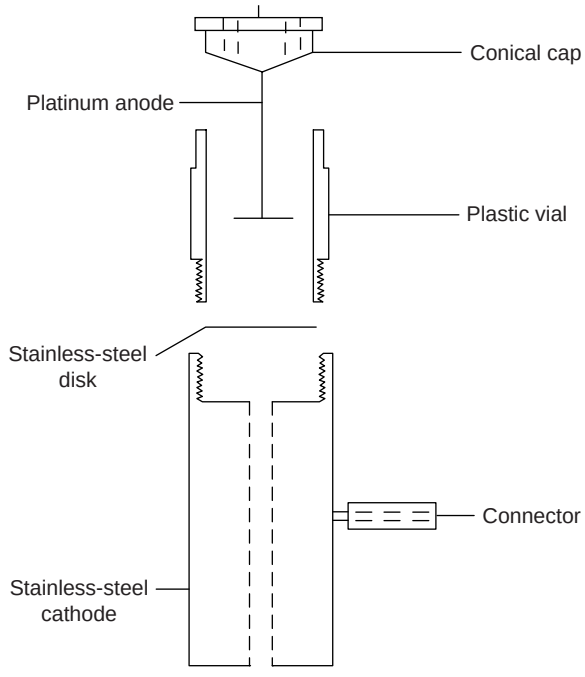

Figure 6 Schematic of a typical electrodeposition cell.

11 mm in diameter, held about 8 mm from the cathode by a platinum wire threaded through a PTFE lid shaped to facilitate the return of condensed vapor to the electrolyte.

Although the use of platinum as cathodic material has been reported to yield better quality sources, the high cost of this metal prevents its use in routine analysis, unless the discs are cleaned and reused. This is not advisable when undertaking trace level analyses. The better quality of platinum-supported sources as compared to stainless-steel is said to be due to the dissolution and subsequent replating of elemental constituents of the stainless-steel discs during electrodeposition.[102] However, it has been reported that the use of a chloride–oxalate medium is not accompanied by any noticeable cathodic corrosion.[155] Some workers have recommended the use of a rotating cathode for best results. However, the practice has not been widely adopted.

To prepare the source for α-spectrometry, the purified solution containing the actinide of interest in, for example, nitrate medium is evaporated to dryness and the nitrate salts eliminated by the repeated addition and evaporation of small amounts of HCl. The residue is redissolved in a solution comprising 1 mL of 3 M HCl and 8 mL of 1 M ammonium chloride–0.01 M ammonium oxalate, together with a few drops of thymol blue pH indicator. The pH of the solution is adjusted to ~2.0 by the dropwise addition of a 50% (v/v) solution of NH_4OH (thymol blue indicator becomes salmon-pink in the pH range 2.0–2.3). This pH value is within the range found to yield maximum deposition.[41] The solution is

transferred to the electrolytic cell and plating carried out by passing a direct current of 1.0 A through the cell for 45–60 min. One minute before switching off the current, the solution is neutralized by the addition of 1 mL of concentrated ammonia in order to prevent redissolution of the deposited actinide upon removal of the electric field. Once the cell is disconnected and the system disassembled, the electroplated disc is rinsed gently with distilled water and acetone, dried at room temperature and heated at 500 °C for 30 s in order to better fix plated actinide hydroxides by conversion to their more stable oxides and to volatilize any remaining traces of polonium.

6.2 Sources Suitable for Liquid Scintillation α-Spectrometry

Liquid scintillation α-spectrometry[156] offers an attractive alternative to conventional α-spectrometry in cases where the total α-activity or the activities of energetically well-resolved actinides are to be assayed, in that it provides virtually 100% counting efficiency and eliminates the problem of self-absorption. Sources are prepared by incorporating the purified analyte nuclide(s) in a liquid scintillation cocktail, that is, an organic solvent (e.g. toluene, benzene, xylene) containing a fluor (e.g. POPOP, PPO) and, in many though not all cases, an energy transfer agent such as naphthalene.[157] It is a fundamental requirement of this counting method that the analyte be distributed homogeneously within the cocktail. As the analyte is usually in aqueous form, sources can be prepared simply by mixing the analyte solution with an aqueous-accepting scintillation cocktail and shaking vigorously for a few seconds. Unfortunately, the introduction of an aqueous sample to an aqueous-accepting scintillator is not to be recommended if the lowest background, best (α-) energy resolution, and best separation of β-/γ- from α-pulses are desired, that is, if color quenching and chemical quenching are to be avoided. Instead, the use of an extractive scintillator (an aqueous-immiscible organic solution containing both the scintillator components and a liquid–liquid extraction reagent for transferring the analyte of interest to the cocktail) is now the method of choice in many laboratories, particularly when there is a requirement to discriminate between α- and β-emitting nuclides.[156]

6.3 Sources Suitable for Mass Spectrometric Counting

Procedures for the preparation of sources suitable for counting by mass spectrometry are clearly dependent on the nature of the analyte phase (solid, liquid or gas) and on the type of spectrometry proposed, for example, ICPMS, thermal ionization mass spectrometry (TIMS) and AMS.

In the case of ICPMS, by far the most common method of sample introduction reported is that of pneumatic nebulization, in which a fine aerosol of analyte droplets is injected into the plasma.[158] This method is, however, very inefficient and alternative techniques such as electrothermal volatilization (ETV) and laser ablation are increasingly being used.

ETV from a graphite rod or a refractory wire is a very convenient method of introducing small samples of a few microliters to the plasma. It produces a transient signal lasting a few seconds while the jet of vapor from the heated surface traverses the plasma.[158] Laser ablation is employed for the direct analysis of solid samples. A single laser shot with an energy of approximately 0.5 J focused on the surface of the sample can ablate 50–100 μg of material in the form of vapor and condensed microparticulates.[158]

In TIMS, the analyte is evaporated on to a rhenium or tungsten filament (with an appropriate salt to enhance the ionization process), prior to heating and measurement.[76,159]

In AMS, the analyte is dispersed in an iron oxide matrix, produced by dissolving the analyte in HNO_3, adding iron nitrate, evaporating to dryness, and baking at 800 °C to convert the nitrate to oxide. The iron oxide matrix is then mixed with aluminum powder (which serves as both an electrical and a thermal conductor) and pressed firmly into a suitably designed sample holder.[160,161]

7 RADIOMETRY

7.1 α-Spectrometry

α-Spectra from thin sources are nowadays measured with passivated (ion-) implanted planar silicon (PIPS) detectors. The PIPS detector is essentially a thin wafer of partially or totally depleted silicon formed by reverse biasing a p–n junction. In the depleted region of the junction, which constitutes the sensitive volume of the detector, α-particles lose most of their energy in electronic excitation and ionization, with the production of electron–hole pairs by electron transfer from the valence band to the conduction band of the crystal. These charge carriers are swept out to the edges of the depleted region by the applied electric field, generating an electronic pulse whose amplitude is directly proportional to the number of charge carriers collected.

Ideally, the amplitude of the pulse should be strictly proportional to the energy of the incident particle. In practice, however, a number of physical processes from which no charge is collected and the statistical nature of the processes involved in the generation of the signal give rise to a detector response with nonzero

variance, limiting the energy resolution of the system.[162] Electronic energy loss straggling in the dead layer (i.e. layer at the surface of the detector through which α-particles must pass before reaching the depleted volume), nonelectronic losses in the depleted region, electron–hole pair statistics and electronic noise contribute to the width and asymmetric shape observed in the pulse height spectrum of monoenergetic α-particles.[149,163] Since no improvement in the intrinsic response of the detector to the incoming radiation can be achieved, a major goal in the design and operation of these detectors is to make the contributions to the peak width arising from energy losses in dead layers (in both detector and source) and electronic noise as small as possible.

Modern PIPS detectors approach the theoretical resolution limits of silicon detectors very closely, making them the preferred choice for most applications in high resolution α-spectrometry. The planar process used in the fabrication of PIPS detectors combines passivation, photolithographic patterning and doping techniques to produce high quality radiation detectors with thin entrance windows and low electronic noise.

The window (face contact) of a PIPS detector consists of a very thin (<50 nm) heavily doped p^+ layer, usually produced by low-energy implantation of boron ions into the silicon wafer. This implanted window is substantially thinner than the delicate metallic contacts of earlier silicon surface barrier detectors, which consisted of a thin evaporated layer of gold (~80 nm) or aluminum (~200 nm). It is also a more rugged and stable surface, which can be readily cleaned with mild solvents. The thinner window of the PIPS detector results in improved resolution, even at close source-to-detector spacings. This is important in environmental studies, where the activities involved are often very low and maximum efficiency with minimal loss of resolution is highly desirable.

The electronic noise of a semiconductor junction-type detector is basically determined by its reverse current and capacitance.[164] The reverse current not only sets a limit for the detection of low-energy charged particles but also contributes to the degradation of the energy resolution. The reverse current is the sum of the diffusion current, surface leakage current and so-called bulk generation current. The passivation step, in which a silicon dioxide layer is created in the surface of the silicon wafer by thermal oxidation, suppresses the surface leakage current,[165] reducing the reverse current of a modern (450 mm^2) PIPS detector to ~10 nA.

PIPS detectors are available with active areas between 25 and 3000 mm^2 and depletion depths between 100 and 700 μm. Since the electronic noise of the preamplifier increases with increasing capacitance,[162] improved energy resolution can be achieved with detectors of smaller active area and greater depletion depth. This is, however, at the expense of detection efficiency, which diminishes with diminishing active area, and of reverse current, which increases with the increasingly large voltages required to extend the depth of the depleted region.[166]

7.1.1 α-Spectrometric Set-up

The excellent operational characteristics of PIPS detectors do not, of themselves, guarantee the best possible resolution. The long counting times needed to record the spectra of low activity sources with good statistics also require reliable, drift-free, high stability electronics. An example of a multiple detector system, specially designed for low-level analysis, is depicted in Figure 7. The system provides each PIPS detector with vacuum, bias supply and preamplifier/amplifier, routing the output signal from the latter into a dedicated segment of a multichannel analyzer (MCA) via an analog-to-digital converter (ADC). The vacuum in each chamber is maintained at ~0.7 Pa by means of a rotary vacuum pump. The bias voltage applied to each detector is set to the value recommended by the manufacturer (typically 40–60 V) and interlocked to a vacuum gauge to ensure that bias is not supplied under poor vacuum conditions.

Sources are normally mounted horizontally in source holders fabricated from low-background materials. A reverse bias of 10 V DC is applied to each source/holder

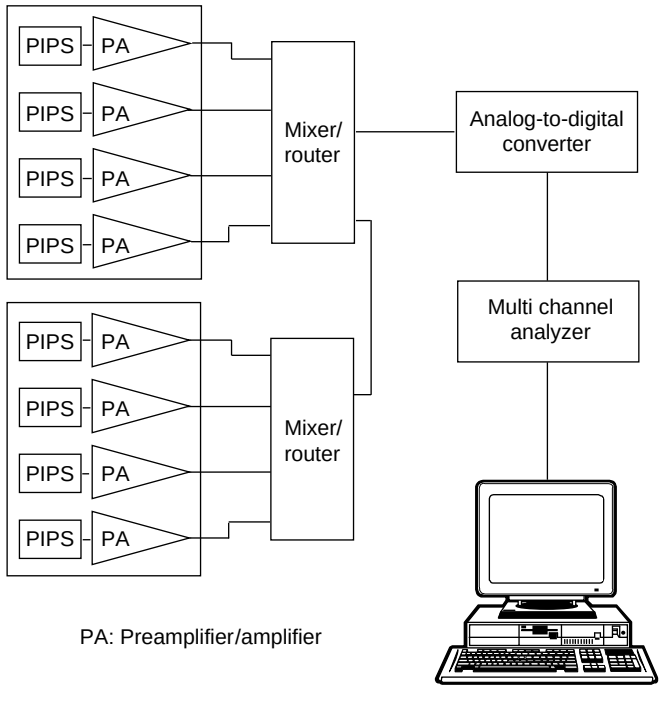

Figure 7 A typical α spectrometric set-up.

which, together with the small amount of air intentionally left in the counting chamber (equivalent to ~16 mg cm^{-2}), helps to prevent detector contamination by so-called α-recoil (i.e. contamination due to recoiling fragments which become embedded in the detector following α-decay). For routine low-level actinide analysis with 450-mm^2 detectors, a typical source-to-detector spacing is ~5 mm. At this separation, an energy resolution (full width at half-maximum) of 18–20 keV is readily achievable, provided that the source is of good quality. For analyses requiring the highest resolution, it is customary to use a detector of smaller active area (say 100 mm^2) and to increase the source-to-detector spacing to ~10–15 mm in order to improve collimation. In this configuration, energy resolutions of 10–12 keV are attainable with high quality electroplated sources.

In the arrangement shown in Figure 7, the pulses from each detector are processed by a preamplifier/amplifier module incorporating a low-noise, charge-sensitive preamplifier with a diode-protected field effect transistor (FET) front end, and a linear shaping amplifier which provides the necessary level of signal conditioning for high-resolution spectrometry. The output from each amplifier is fed to a mixer/router, digitized by an ADC and routed to one of the memory segments (typically 512 channels each) of a personal computer-based MCA.

7.1.2 α-Spectra

If the radiochemical purification and source preparation protocols described above are successfully implemented, the resulting spectra are usually quite simple, consisting of a small number of well defined peaks on a small, discrete background. Typical α-spectra of thorium, uranium, plutonium and americium are shown in Figures 8–11, inclusive. In each spectrum, one of the ''peaks'' (actually an unresolved or partially resolved doublet or multiplet) corresponds to the tracer nuclide added to monitor the radiochemical yield. In the absence of spectral interferences, the analysis reduces to the integration of the number of counts under each ''peak'' and the subtraction of the corresponding background. The activity of the analyte nuclide is then obtained simply from the ratio of its peak area to that of the tracer peak multiplied by the known activity of the tracer (in becquerels).

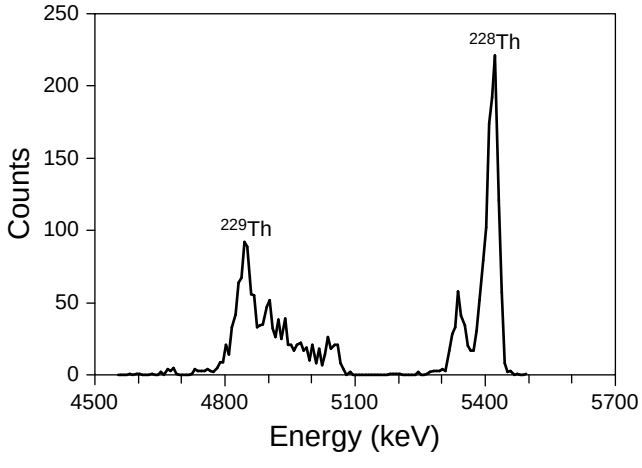

Figure 8 Pulse height spectrum of a typical thorium sample with ^{229}Th as yield monitor.

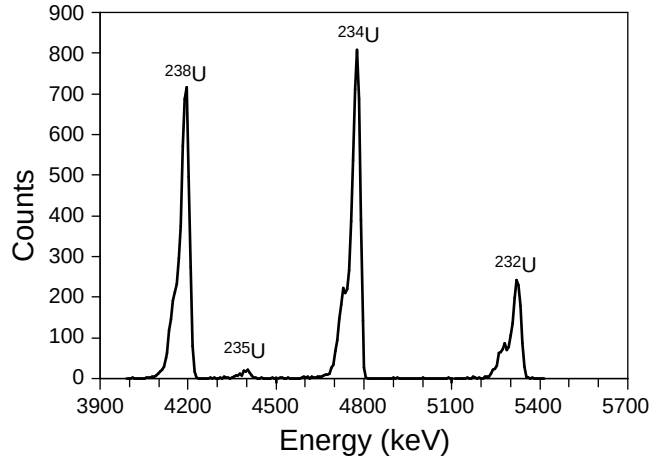

Figure 9 Pulse height spectrum of a typical uranium sample with ^{232}U as yield monitor.

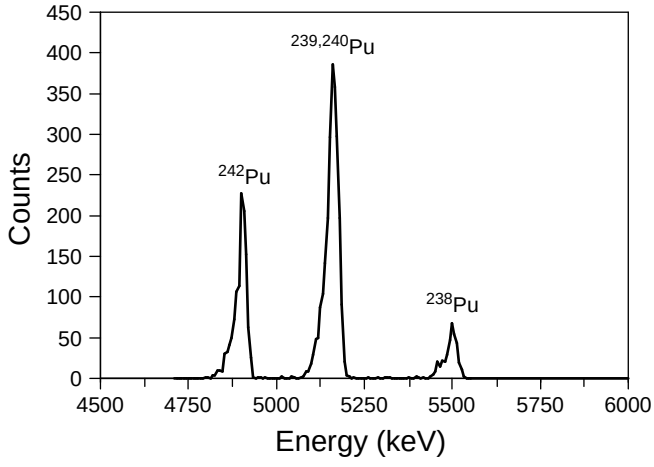

Figure 10 Pulse height spectrum of a typical plutonium sample with ^{242}Pu as yield monitor.

7.1.3 Background Measurements

The background of a PIPS detector is generally very low, of the order of 10^{-5}–10^{-6} s^{-1} for a region of interest of say 100 keV. With use, however, the background tends to increase by nuclide evaporation or recoil contamination. Since background subtraction may be of significance,

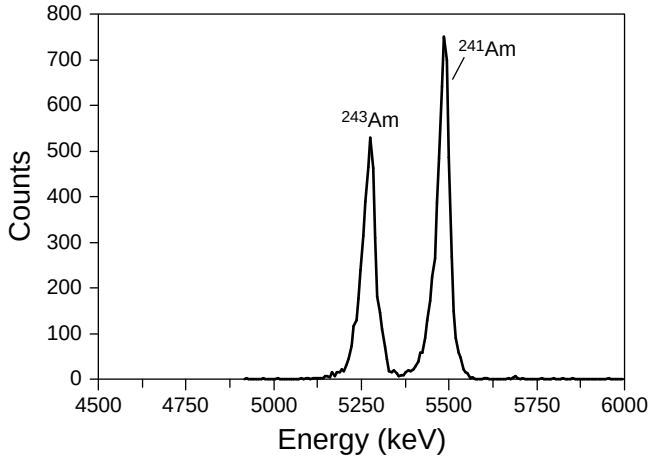

Figure 11 Pulse height spectrum of a typical americium/curium sample with ^{243}Am as yield monitor.

especially at the low activities often encountered in environmental samples, periodic background measurements must be carried out to check for any contamination that may have arisen (e.g. in the course of analysis or laboratory operations) and ensure accurate background corrections.

7.1.4 Spectral Interferences

Spectral interferences are considered to occur when the α-emissions from different nuclei overlap in energy. This may be because the various peaks have similar energies and cannot readily be resolved from one another, or because of the tailing of well-resolved peaks to lower energies when counting ''thick'' (i.e. poor quality) sources. Spectral conflicts increase the uncertainty of the analysis to the point of making it impossible for spectra (and, by extension, sources) affected by severe interferences. In extreme cases, the best cure is the careful repurification of the source, followed by replating and recounting.

In some cases, the small energy separation between the α-emissions of two or more isotopes of a particular element results in the creation of unresolved (or only partially resolved) multiplets. This type of overlap is unavoidable and is usually dealt with by quoting the summed activities of the composite peak (e.g. $^{239+240}$Pu, $^{243+244}$Cm).

More difficult to unravel are the overlaps that arise from the incomplete removal of certain impurity nuclides, e.g. members of the natural radioactive decay series. For example, in the case of thorium spectra, the rapid ingrowth of a ^{224}Ra peak (intensity 5.1%) under the ^{228}Th peak should be taken into account by measuring the main ^{224}Ra peak (intensity 94.5%) and estimating the contribution from ^{224}Ra in the ^{228}Th window on the basis of the relative intensities of the two ^{224}Ra peaks. Moreover, when using a mixed ^{232}U/^{228}Th tracer, care must also be taken to account for the presence of indigenous ^{228}Th, which originates from the ^{232}Th $\rightarrow$ ^{228}Ra $\rightarrow$ ^{228}Th natural decay chain. One way to do this is to assume that the ^{228}Th and ^{232}Th are in secular equilibrium. This is not always the case, however, and the analysis of an unspiked replicate sample to establish the actual ^{228}Th/^{232}Th activity ratio becomes imperative, unless one is certain of the equilibrium assumption. The problem does not arise when ^{229}Th is used as yield monitor.

The analysis of uranium spectra using ^{232}U as monitor is straightforward, in that the ^{238}U, ^{235}U, ^{234}U and ^{232}U peaks are usually well resolved from one another. Care must be taken, however, to count the sources soon after sample purification, as short-lived daughter nuclides ingrow quickly, yielding peaks, some of which cannot be resolved from the peaks of interest.[76]

In the case of transuranium spectra, the main interferences arise from the presence of (1) a ^{234}U peak in the ^{242}Pu window, (2) ^{210}Po in the ^{243}Am window, (3) ^{228}Th in the ^{238}Pu or ^{241}Am window (the energies being similar), (4) ^{224}Ra in the ^{243,244}Cm window, and (5) ^{212}Bi in the ^{242}Cm window.

When measuring plutonium in samples of marine origin, the contribution of ^{234}U to the ^{242}Pu peak can be corrected for by using the well-known ratio of ^{234}U to ^{238}U in this environment, namely 1.14 ± 0.01, to determine the number of counts in the ^{242}Pu window due to ^{234}U. For samples of other origin, sources must be stripped and repurified until essentially no ^{238}U appears in the spectrum.

Th-228 interference in the ^{238}Pu or ^{241}Am window is corrected for by assuming secular equilibrium between ^{228}Th and its daughter nuclides ^{224}Ra, ^{220}Rn, ^{216}Po and ^{212}Bi after the appropriate ingrowth period (i.e. at least 2 weeks post-electrodeposition). The same correction can be applied to the ^{224}Ra interference in the ^{243,244}Cm window and to the ^{212}Bi interference in the ^{242}Cm window.

The contribution to the peaks of interest from possible trace impurities in the yield monitors and chemical reagents used should also be taken into account by the periodic analysis of appropriate reagent blanks. Finally, when operating outside the normal laboratory environment, for example, when in the field or aboard a research vessel, the analysis of field blanks should be an essential component of the overall analytical program.

7.1.5 Calculation of Activity

Once the area under each peak is corrected for background, field/reagent blank and spectral interference, the mass activity (or activity concentration) of the analyte

nuclide can be determined using Equation (5):

$$C = \frac{N_S A_T}{N_T m} \tag{5}$$

where C is the mass activity (or activity concentration) in the sample (Bq unit^{-1}), N_S is the net number of counts in the analyte peak, A_T is the activity of the yield monitor added to the sample (Bq), N_T is the net number of counts in the yield monitor peak, and m is the size of the sample in appropriate units (i.e. kg, m^{-3}). The chemical recovery of the sample can also be determined using Equation (6):

$$R(\%) = \frac{N_T 100}{A_T t \Omega} \tag{6}$$

where t is the counting interval (s), and Ω is the fractional solid angle. The uncertainty associated with the calculated activity is easily determined using quadratic error propagation theory.

7.1.6 Limits of Detection and Quantitation

The limit of detection (LOD) is the lowest concentration of an analyte that the analytical process can reliably detect. The LOD in most instrumental methods is based on the relationship between the gross analyte signal, S_t, the field blank, S_b, and the variability in the field blank, σ.[167] The LOD has been variously defined by the extent to which the gross signal exceeds S_b:

$$S_t - S_b \geq K_d \sigma$$

where K_d is a constant factor. It is recommended that detection should be based on a minimal value for K_d of 3, so that the LOD is located at 3σ above the gross blank signal, S_b. This is illustrated in Figure 12, which depicts the region of detection of analyte in the gross signal as $S_t \geq S_b + 3\sigma$, and similarly for the net signal as $S_x \geq S_d$.

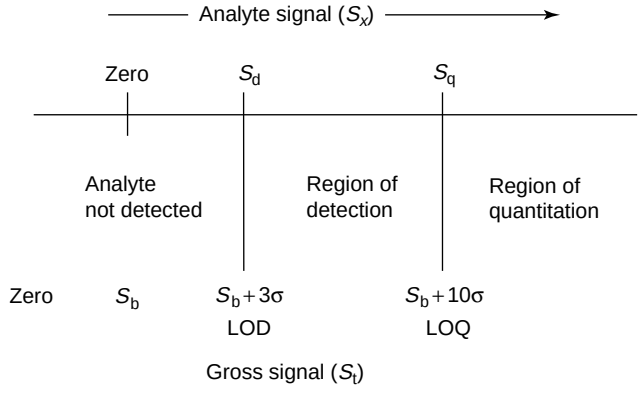

Figure 12 Schematic of the minimal criteria recommended by the American Chemical Society's Committee on Environmental Improvement. (Source: ACS, 1980.[167])

Thus, in α-spectrometry, the lower limit to detection, expressed in concentration units, would then be obtained from the net signal using Equation (7)

$$\text{LOD(Bq unit}^{-1}) = \frac{S_d}{\Omega m t R} = \frac{3\sigma}{\Omega m t R} \tag{7}$$

where, again, Ω is the fractional solid angle, m is the sample size (unit), t is the counting interval (in seconds) and R is the fractional radiochemical yield or recovery.

While a value of $K_d = 3$ is considered minimal, higher values may be required since this value implies definite risks of 7% for false positive (concluding that the analyte is present when it is absent) and false negative (the reverse) decisions. This is recognition that the numerical or statistical significance of the apparent analyte concentration increases as the analytical signal increases above the LOD. In other words, there is need for a second, more stringent definition, namely a limit of quantitation (LOQ), located above the measured average field blank, S_b, by the following definition:

$$S_t - S_b \geq K_q \sigma$$

In this case it is recommended that the minimum value for K_q be 10.[167] The region of quantitation is also depicted in Figure 12. Combined, the two definitions define the region of detection $S_q > S_x \geq S_d$ and the region of quantitation $S_x \geq S_q$.

In the context of low-level assay by α-spectrometry, a representative LOQ (calculated on the basis of an average field blank count rate of 5×10^{-6} s^{-1}, a fractional solid angle of 0.28, a recording time of 2×10^6 s (about 3 weeks), and a radiochemical yield of 80%) would be ~50 μBq. For a water sample of say 100 L, this is equivalent to an activity concentrations of ~0.5 mBq m^{-3}, while for a solid sample of 10 g it represents ~5 mBq kg^{-1}.

7.2 Liquid Scintillation α-Spectrometry

In liquid scintillation spectrometry, as explained above, the sample is placed in direct contact with a liquid scintillation cocktail, normally composed of one or more fluorescent solutes in an organic solvent. The energy of the α- or β-particles emitted by the source is lost in excitation and ionization of solvent molecules. Activated solvent molecules subsequently transfer their excess energy to dissolved organic scintillator molecules, which are brought to an excited state. When these scintillator molecules return to their natural or ground state, energy is lost in the form of light and heat. The intensity of the light produced in the scintillation cocktail is directly proportional to the energy of the particle. The flash of light is simultaneously detected by two photomultiplier tubes operating in coincidence mode, giving rise to an electronic pulse whose amplitude is

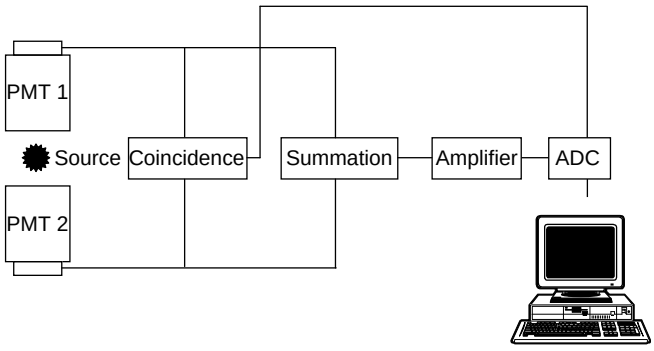

Figure 13 Simplified block diagram of a liquid scintillation counter. PMT, photomultiplier tube.

linearly related to the energy of the particle. A simple block diagram of a modern liquid scintillation counter is shown in Figure 13. The purpose of the coincidence unit is to reduce background caused by photomultiplier tube noise. The summation circuit adds the signals from both photomultipliers, optimizing the signal-to-noise ratio for optimum detection.

Since the sample is in direct contact with the scintillation cocktail, it can influence the intensity of the light produced in the scintillation process. This phenomenon of reduced light intensity is termed quenching and has to be taken into account, especially when counting low-energy β-emitters. Traditionally, quenching is classified into two broad categories, namely chemical quenching and color quenching.[156,168] Chemical quenching occurs when excited solvent molecules fail to transfer energy to scintillator molecules (the fluor) or instead transfer their energy to nonlight-producing molecules, whereas color quenching occurs when the emitted light is absorbed or attenuated before it can reach the photomultiplier tubes.

α-Particles, because of their charge and high decay energies, dissipate most of their energy in inelastic collisions with solvent molecules, with much of the energy being wasted in nonradiative losses. As a result, the fluorescence conversion efficiency for these particles is approximately an order of magnitude lower than for β-particles,[168] complicating the discrimination between α- and β-events in spite of the fact that α-particles are much more energetic than β-particles.

In practice, discrimination between α- and β-events is based on the characteristics of the light pulses produced by these particles in a scintillation cocktail. The high specific ionization of α-particles gives rise to a delayed fluorescence component which results in light pulses of longer duration than those created by β-particles.[169] On this basis, α- and β-events can be discriminated by analysing the shape of the respective light pulses.

8 MASS SPECTROMETRY

8.1 Inductively Coupled Plasma Mass Spectrometry

In recent years, an increasing number of researchers have explored the possibilities of ICPMS as an alternative technique for the determination of actinide nuclides at trace level.[123,170–175] ICPMS is an analytical technique that performs elemental analysis with excellent sensitivity and high sample throughput. ICPMS systems rely on the introduction of sample molecules into an argon plasma at temperatures of 6000–10 000 K, where they are dissociated and the resulting atoms ionized. Normally, liquid samples (e.g. a purified actinide in an appropriate acid medium) are transferred by a peristaltic pump to a nebulizer, which feeds a fine aerosol of analyte droplets into the plasma. Analyte ions are extracted from the plasma into a mass spectrometer (held at high vacuum) through a pair of orifices, known as sampling and skimmer cones. These ions are focused by a series of ion lenses into a magnetic quadrupole analyser, which separates the ions on the basis of their mass/charge ratio. The ion count is recorded for each mass number using, for example, a continuous dynode channel electron multiplier. Other detectors, such as discrete dynode multipliers and the Daly detector, have also been used to a limited extent.

A realistic LOQ for a conventional ICPMS system employing a pneumatic nebulizer and a quadrupole spectrometer is $\sim 2 \times 10^9$ atoms, at the present time. To illustrate, in the case of ^{239}Pu, this is equivalent (in activity terms) to an LOQ of $\sim$2 mBq, which is a factor of $\sim$40 higher (i.e. poorer) than conventional high-resolution α-spectrometry. However, ETV of the sample can extend this limit down to $\sim$0.02 mBq or slightly better, and new high-resolution ICPMS systems are just becoming available with a claimed LOQ approaching 10^6 atoms. Using ETV ICPMS, LOQ values of 30 fg and 9 fg for ^{238}U and ^{232}Th, respectively, have recently been reported, both being blank limited.[176]

In summary, for the assay of actinides, the latest ICPMS systems appear to be at least as sensitive as high-resolution α-spectrometry, except for the shorter-lived α-emitters (i.e. $t_{1/2} < 100$ years), where α-spectrometry remains pre-eminent. ICPMS offers the great advantage of reducing counting times to a few minutes compared to the weeks-long counting times associated with α-spectrometry.

8.2 Thermal Ionization Mass Spectrometry and Resonance Ionization Mass Spectrometry

In TIMS, samples are deposited on specially treated filaments, usually fabricated from rhenium or tantalum. The filament is heated slowly under closely controlled conditions to ionize the sample, which is then introduced

in the mass spectrometer. Because of isobaric and chemical interferences, it is necessary to first carry out a high quality chemical separation of the analyte element of interest.

LOQs with TIMS of 2×10^6 atoms for thorium and 8×10^6 atoms for plutonium have been reported.[177,178] This is at least two orders of magnitude more sensitive than that attainable for the long-lived isotopes of these elements with conventional α-spectrometry.

Resonance ionization mass spectrometry (RIMS) has also been shown to be a highly sensitive technique for the detection of trace elements. The method is based on the step-wise excitation of atoms from a defined state to highly excited states by resonant absorption of photons, followed by ionization with subsequent mass analysis of the ions. For neptunium, a detection limit of 4×10^8 atoms has been reported using this technique.[179] RIMS has also been employed for the determination of uranium and plutonium.[180,181] A detection limit for plutonium of $\sim 10^7$ atoms has been claimed.

8.3 Accelerator Mass Spectrometry

AMS is a highly sensitive technique for isotopic analysis in which atoms extracted from a sample are ionized, accelerated to high energies (mega electron-volts), separated according to their momentum, charge and energy, and individually counted (having been identified as having the correct atomic number or mass). The principal difference between AMS and conventional mass spectrometry is the energy to which the ions are accelerated (mega electron-volts as compared to kiloelectron-volts in conventional mass spectrometry). The practical consequence of operating at higher energies is that ambiguities in identification of atomic and molecular ions with the same mass are removed. This has made it feasible to use AMS to measure isotopic ratios as low as 1 in 10^{15}. Moreover, this sensitivity can be achieved with sample sizes of a milligram or less (containing as few as 10^6 atoms of the isotope of interest) within a measurement time of about an hour.

Although AMS is a relatively new and rapidly expanding field, various authors have already presented excellent reviews of the technology involved and the applications possible.[182–190] Accordingly, only a brief description of the main features of actinide measurement by AMS will be given here.

In essence, the ion source (see section 6.3) produces a beam of (usually negative) ions (charge q), which are accelerated to energy E (typically 100–400 keV) by a pre-accelerator and then separated according to their atomic mass (M) using a uniform magnetic field (B) directed normal to the plane of motion (momentum analysis). The radius of curvature (r) of the path of each ion is determined by Equation (8)

$$2\left(\frac{ME}{q^2}\right. = k_1(Br)^2 \tag{8}$$

At this stage, ions with the correct ME/q^2 value are injected into a tandem accelerator, where they are accelerated to energies of some tens of mega electron-volts. The ions also pass through a gas or foil stripper, which removes one or more electrons from the ions, forming multiply charged positive ions (charge $= +nq$). In addition, molecular ions with $n \geq 3$ are dissociated during the stripping process. The now positive ions are accelerated to ground potential before leaving the accelerator. Two or more magnetic and/or electrostatic selectors, placed after the accelerator, allow further rejection of unwanted ions. Magnetic analyzers select the correct ions on the basis of their ME/q^2 value, while electrostatic analyzers select the correct ions on the basis of their energy-to-charge ratio (E/q). A radial electrostatic field of constant magnitude (ε) deflects the ions along a circular path of radius ρ defined by Equation (9)

$$2\left(\frac{E}{q}\right. = k_2\varepsilon\rho \tag{9}$$

In some systems, a velocity selector (Wien filter) is used instead of, or in addition to, an electrostatic analyzer. Finally, ions having the required mass and atomic number are identified and recorded using a suitable detector (e.g. longitudinal-field ionization chamber, time-of-flight system, Faraday cup, silicon detector). Such positive identification is necessary because ions other than those of the analyte nuclide may also reach the detector.

Several researchers have explored the application of AMS to the detection of heavy elements at trace level. Using AMS, Kilius et al.[191] measured $^{230}Th/^{232}Th$ ratios in uranium ore, thorium metal and uranium metal with high precision. Zhao et al.[192] used AMS to measure ^{236}U in uranium ore at a level of $\sim 10^{-10}$ relative to ^{238}U. More recently, Purser et al.[193] discussed in detail the design and operation of a novel AMS system, incorporating a 3-MV tandem accelerator, for the assay of actinide elements. It seems that the measurement of ^{236}U in high-grade ore presents the greatest challenge, because the beam analysis system must completely eliminate ions of the other uranium isotopes (which are orders of magnitude more abundant in the sample material) since the detection system does not provide effective discrimination between ^{236}U and ^{235}U.

The absolute sensitivity of the AMS technique for actinide assay is perhaps best illustrated by the results of Fifield et al.,[160,161] who demonstrated the capacity

of the technique to measure plutonium at the 10^6 atom level. For ^{239}Pu, this limit is also at least two orders of magnitude lower than that attainable with conventional α-spectrometry.

9 QUALITY ASSURANCE/QUALITY CONTROL

Although the application of well-tested analytical methods and good laboratory practices is an essential prerequisite for the production of high quality data, it is not, of course, a guarantee of the reliability of the data, and checks on the precision and absolute accuracy of the results should be part of every analytical program. In fact, the strengthening of national and international regulations, as well as the growth in the level of analytical work contracted out to commercial laboratories, require that radioactivity measurements be ''traceable'' to national and/or international standards, and that such traceability be supported by records of measuring equipment calibration, radiochemical procedures followed and quality assurance programs implemented.

Internal quality control can be achieved by means of the careful monitoring of field and reagent blanks, regular background determinations, periodic recalibration of tracers and detecting systems, and analysis of reference materials (RMs). The accuracy of the results obtained by a laboratory may be tested by participation in international intercomparison exercises such as those organized by the International Atomic Energy Agency as part of its Analytical Quality Control Services program. Intercomparison exercises are also organized by other organizations (e.g. National Institute of Standards and Technology (NIST)).

9.1 Traceability and Standardized Solutions

Traceability is defined as the property of a result of a measurement whereby it can be related to appropriate standards, generally national or international standards, through an unbroken chain of comparisons. The purpose of requiring traceability is to ensure that measurements are accurate representations of the specific quantity subject to measurement within the uncertainty of the measurement.

In low-level radioactivity analyses, reliable measurements must be traceable to recognized national and/or international measurement standards. A number of national laboratories, such as the National Physical Laboratory in the UK, NIST in the USA, Laboratoire Primaire des Rayonnements Ionisants in France and Physikalisch-Technische Bundesanstalt in Germany, provide low-level radionuclide standards directly traceable to absolute primary radioactivity standards. These and other national standards laboratories regularly participate in validation and intercomparison exercises organized under the auspices of the International Bureau for Weights and Measures to ensure traceability between different national standards.

α-Emitting radionuclide solutions for a number of actinides (e.g. ^{228}Th, ^{229}Th, ^{232}U, ^{233}U, ^{234}U, ^{235}U, ^{236}U, ^{236}Pu, ^{237}Pu, ^{238}Pu, ^{239}Pu, ^{240}Pu, ^{242}Pu, ^{244}Pu, ^{237}Np, ^{241}Am, ^{243}Am and ^{244}Cm) are available from these organizations and widely used as chemical yield monitors, as internal standards for radiochemical assay validation and for calibration of measuring instruments (e.g. α-spectrometers and mass spectrometers). These certified and traceable solutions are normally supplied in flame-sealed borosilicate ampoules, together with certificates detailing reference date of measurement, mass of solution, radioactivity concentration, assessment of uncertainty, method(s) of measurement, chemical composition of the solution, and recommended radionuclide half-life at the time of issue.

9.2 Field and Reagent Blank Analyses

The analysis of a field (or laboratory/reagent) blank to determine the true blank signal to be subtracted from the gross analyte signal is not a straightforward exercise. Ideally, a field blank should have the same composition as the sample, except that it should be free of analyte. Unfortunately, this is seldom the case in practice, and analysts all too often have to process blanks whose composition only crudely approximates that of the sample. Nevertheless, field and reagent blank analyses are essential, as they provide the only reliable means of determining whether contamination has been introduced into the samples as a consequence of in-field/in-laboratory contamination or via the chemical reagents used. For this reason, blank analyses should always be carried out with each batch of environmental samples processed.

9.3 Intercomparison and Reference Materials

Analysis of appropriate RMs represents the best and most direct way to investigate bias during analysis or to test the reliability of an analytical method.[194] An RM is a material or substance for which one or more properties are sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials.[195] A certified reference material (CRM) is an RM where one or more of its properties are certified by a technically valid procedure, accompanied by or traceable to a certificate or other documentation which is issued by a certifying body.[195]

A number of CRMs for radionuclide determination are available from NIST, covering a variety of radionuclides and environmental matrices in different concentration ranges. Uncertified natural matrix RMs are also available from the International Atomic Energy Agency in Vienna and from that Agency's International Laboratory of Marine Radioactivity in Monaco. Materials on offer include soils, marine sediments, algae, biota and human and animal bone.

10 PERSPECTIVES AND FUTURE DEVELOPMENTS

There is little doubt that the use of mass spectrometric techniques for the assay of actinides at trace or environmental level will continue to increase at the expense of α-spectrometry. However, the latter will remain the method of choice for actinides and other α-emitters with half-lives shorter than $\sim$100 years, owing to its much higher sensitivity for such nuclides. Similarly, the technique of extraction chromatography is becoming increasingly popular for the sequential separation of actinides from different matrices, with the development of highly selective coated resins. Benefits claimed for this technique include higher chemical yields, reduction in analysis time with consequent savings in labor costs, and significant reductions in the quantities of waste chemicals produced.

One of the main drawbacks of analytical radiochemistry is the labor-intensive nature of the operations involved. Although attempts, some quite successful, have been made to part-automate the analytical steps required to separate and purify various actinides from natural matrices, it appears that industry is still some way from developing dedicated and fully automated systems with the full range of capabilities and the flexibility essential in what is, after all, one of the most demanding areas in analytical chemistry.

Standard RMs and, in particular, fully traceable reference solutions are vital tools in the complementary fields of analytical radiochemistry, α-spectrometry and mass spectrometry. Unfortunately, there is reason to question the future availability of some of the most important tracers presently in general use, for example ^{242}Pu and ^{244}Pu. It is, therefore, incumbent on the scientific community, in association with international and national standards laboratories, to take the necessary initiative to protect the continued availability of as wide a selection of these tracers as possible. This should include surveying the user community as to their future needs. It is likely that one of these needs will be for tracers for chemical speciation analysis, a field of study that has attracted growing interest in recent years.

ACKNOWLEDGMENTS

We gratefully acknowledge the many colleagues and associates who generously advised and assisted us over the years in implementing a number of the analytical schemes and protocols described above in our own laboratory.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
AMS	Accelerator Mass Spectrometry
CRM	Certified Reference Material
DDCP	Dibutyl-*N*,*N*-diethyl Carbamyl Phosphate
EDTA	Ethylenediaminetetraacetic Acid
ETV	Electrothermal Volatilization
FET	Field Effect Transistor
HDEHP	Di-2-ethyl Hexyl Phosphoric Acid
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LOD	Limit of Detection
LOQ	Limit of Quantitation
MCA	Multichannel Analyzer
NIST	National Institute of Standards and Technology
NMWL	Nominal Molecular Weight Limit
PIPS	Passivated (Ion-) Implanted Planar Silicon
PTFE	Poly(tetrafluoroethylene)
PVC	Poly(vinyl Chloride)
RIMS	Resonance Ionization Mass Spectrometry
RM	Reference Material
TBP	Tri-*n*-butylphosphate
TIMS	Thermal Ionization Mass Spectrometry
TIOA	Tri-*iso*-octylamine
TLA	Trilauryl Amine
TOA	Tri-*n*-octylamine
TOPO	Tri-*n*-octyl Phosphine Oxide
TTA	β-Diketone Thenoyl Trifluoroacetone

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Detection and Quantification of Environmental Pollutants • Heavy Metals Analysis in Seawater and Brines • Inductively Coupled Plasma Mass Spectrometry in Environmental Analysis • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Liquid Chromatography/Mass Spectrometry in Environmental Analysis •

Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites

Food **(Volume 5)**
Sample Preparation for Food Analysis, General

Liquid Chromatography **(Volume 13)**
Ion Chromatography

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History • High-resolution Mass Spectrometry and its Applications • Isotope Ratio Mass Spectrometry

Nuclear Methods **(Volume 14)**
Radiotracer Methods

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Mass Spectrometry of Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation • Speciation of Radionuclides in the Environment

General Articles **(Volume 15)**
Literature Searching Methodology • Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. G.T. Seaborg, M.L. Perlman, 'Search for Elements 93 and 94 in Nature. Presence of 94^{239} in Pitchblende', *J. Am. Chem. Soc.*, **70**, 1571–1573 (1948).
2. D.F. Peppard, G.W. Mason, P.R. Gray, J.F. Mech, 'Occurrence of the $(4n + 1)$ Series in Nature', *J. Am. Chem. Soc.*, **74**, 6081–6084 (1952).
3. W.A. Myers, M. Lindner, 'Precise Determination of the Natural Abundance of ^{237}Np and ^{239}Pu in Katanga Pitchblende', *J. Inorg. Nucl. Chem.*, **33**, 3233–3238 (1971).
4. B.R. Harvey, M.B. Lovett, 'The Use of Yield Tracers for the Determination of Alpha-emitting Actinides in the Marine Environment', *Nucl. Instrum. Methods Phys. Res.*, **223**, 224–234 (1984).
5. B.R. Harvey, M.B. Lovett, 'Yield Tracer Applications in the Determination of Man-made Radionuclides at Low Environmental Concentrations', in *Low-level Measurements of Man-made Radionuclides in the Environment*, Proc. of the Second International Summer School, eds. M. García-León, G. Madurga, La Rábida, Huelva, Spain, 25 June–6 July 1990, World Scientific, Singapore, 217–238, 1991.
6. S.M. Jerome, D. Smith, M.J. Woods, S.A. Woods, 'Metrology of Plutonium for Environmental Measurements', *Appl. Radiat. Isot.*, **46**(11), 1145–1150 (1995).
7. M.E. Wrenn, N.P. Singh, S.A. Ibrahim, N. Cohen, 'Thorium-229 as an Isotopic Tracer for the Radiochemical Determination of Thorium Isotopes in Biological Samples', *Anal. Chem.*, **59**(12), 1712–1713 (1978).
8. R.J. Pentreath, B.R. Harvey, 'The Presence of Np-237 in the Irish Sea', *Mar. Ecol. Prog. Ser.*, **6**, 243–247 (1981).
9. M. Yamamoto, K. Chatani, K. Komura, K. Ueno, 'Development of Alpha-ray Spectrometric Techniques for the Measurement of Low-level ^{237}Np in Environmental Soil and Sediment', *Radiochim. Acta*, **47**, 63–68 (1989).
10. E. Holm, M. Nilsson, in *Radioelement Analysis: Progress and Problems*, ed. W.S. Lyon, Ann Arbor Sci. Publ., Ann Harbor, MI, 231, 1980.
11. G. Rosner, R. Winkler, M. Yamamoto, 'Simultaneous Radiochemical Determination of Np-237 and Np-239 with Np-235 as a Tracer, and Application to Environmental Samples', *J. Radioanal. Nucl. Chem.*, **173**(2), 273–281 (1993).
12. I.K. Kressin, W.D. Moss, H.F. Sculte, 'Plutonium-242 Versus Plutonium-236 as an Analytical Tracer', *Health Phys.*, **28**, 41–47 (1975).
13. P.G. Johnson, R.D. Carver, R.J. Dupzyk, 'Preparation of Plutonium-242 for Use as an Alpha Spectrometry and Mass Spectrometry Tracer', *Health Phys.*, **33**, 301–302 (1977).
14. L.H. Keith, *Principles of Environmental Sampling*, American Chemical Society, Washington, DC, 1988.
15. J.R.P. Clarke, C.L.P. Thomas, 'The Benefits of Good Sampling', *Valid Anal. Measure. Prog. Bull.*, **15**, 13–14 (1996).
16. C.L.P. Thomas, H. Schofield, *Sampling Source Book*, Butterworth Heinemann, Oxford, 1996.
17. N.A. Chieco, *The Procedures Manual of the Environmental Measurements Laboratory*, 28th edition, ed. N.A. Chieco, HASL-300, US Department of Energy, New York, 1997.
18. K. Grasshoff, 'Sampling and sampling techniques', in *Methods of Seawater Analysis*, eds. K. Grasshoff, M. Ehrhardt, K. Kremling, Verlag Chemie, Weinheim, 1–19, 1983.
19. R. McQuillin, D.A. Ardus, *Exploring the Geology of Shelf Seas*, Graham & Trotman Limited, London, 158–176, 1977.
20. T.C. Moore, G.R. Heath, 'Sea-floor Sampling Techniques', in *Chemical Oceanography*, 2nd edition,

eds. J.P. Ryley, R. Chester, Academic Press, London, 75–126, Vol. 7, 1978.
21. K.S. Stowe, *Ocean Science*, John Wiley & Sons, New York, 152–156, 1979.
22. B.R. Harvey, M.B. Lovett, 'Identification and Control of Contamination, Interference and Other Errors in Low-level Radionuclide Analysis', in *Low-level Measurements of Man-made Radionuclides in the Environment*, Proc. of the Second International Summer School, eds. M. García-León, G. Madurga, La Rábida, Huelva, Spain, 25 June–6 July 1990, World Scientific, Singapore, 239–262, 1991.
23. B. Carpenter, J. Cobb, J. Toole, G. Longworth, 'The Radiochemical Analysis Laboratory', in *The Radiochemical Manual*, ed. G. Longworth, AEA Technology plc, Harwell, 55–116, 1998.
24. M.B. Lovett, S.J. Boggis, P. Blowers, 'The Determination of Alpha-emitting Nuclides of Plutonium, Americium and Curium in Environmental Materials: Part 1. Seawater', Aquatic Environment Protection Report 7, Ministry of Agriculture, Fisheries and Food, Directorate of Fisheries Research, Lowestoft, 1990.
25. Q. Chen, A. Aarkrog, S.P. Nielsen, H. Dahlgaard, H. Nies, Y. Yu, K. Mandrup, 'Determination of Plutonium in Environmental Samples by Controlled Valence in Anion Exchange', Risø Report M-2856, Risø National Laboratory, Denmark, 1991.
26. M.L. Wells, E.D. Goldberg, 'Occurrence of Small Colloids in Seawater', *Nature*, **353**, 342–344 (1991).
27. M.L. Wells, E.D. Goldberg, 'Marine Submicron Particles', *Mar. Chem.*, **40**, 5–18 (1992).
28. M.L. Wells, E.D. Goldberg, 'Colloid Aggregation in Seawater', *Mar. Chem.*, **41**, 353–358 (1993).
29. J.D. Eakins, 'The Application of Radiochemical Separation Procedures to Environmental and Biological Materials', *Nucl. Instrum. Methods Phys. Res.*, **223**, 194–199 (1984).
30. K.M. Wong, G.S. Brown, V.E. Noshkin, 'A Rapid Procedure for Plutonium Separation in Large Volumes of Fresh and Saline Water by Manganese Dioxide Coprecipitation', *J. Radioanal. Chem.*, **42**, 7–15 (1978).
31. V.F. Hodge, 'Simple Recovery of Plutonium, Americium, Uranium and Polonium from Large Volumes of Ocean Water', *Anal. Chem.*, **46**, 1334–1336 (1974).
32. K.C. Pillai, R.C. Smith, T.R. Folsom, 'Plutonium in the Marine Environment', *Nature*, **203**, 568–571 (1964).
33. E. Mathew, V.M. Matkar, K.C. Pillai, 'Determination of Plutonium, Americium and Curium in Environmental Samples', *J. Radioanal. Chem.*, **62**, 264–278 (1981).
34. H.D. Livingston, D.R. Mann, V.T. Bowen, 'Double-tracer Studies to Optimize Conditions for the Radiochemical Separation of Plutonium from Large Seawater Samples', in *Reference Methods for Marine Radioactivity Studies II*, Technical Report Series No. 169, STI/DOC/10/169, International Atomic Energy Agency, Vienna, 1975.
35. K.M. Wong, 'Radiochemical Determination of Plutonium in Seawater, Sediments and Marine Organisms', *Anal. Chim. Acta*, **56**, 355–364 (1971).
36. E. Holm, R. Fukai, 'Determination of Americium and Curium by Using Ion-exchange in Nitric Acid–Methanol Medium for Environmental Analysis', *Talanta*, **23**, 853–855 (1976).
37. T. Nakanishi, M. Yajima, M. Senaga, M. Takei, A. Ishikawa, K. Sakamoto, M. Sakanoue, 'Determination of ^{239,240}Pu in Seawater', *Nucl. Instrum. Methods Phys. Res.*, **223**, 239–242 (1984).
38. S. Ballestra, J. Gastaud, J.J. Lopez, 'Radiochemical Procedures Used at IAEA-ILRM Monaco for Measuring Artificial Radionuclides Resulting from the Chernobyl Accident', in *Low-level Measurements of Man-made Radionuclides in the Environment*, Proc. of the Second International Summer School, eds. M. García-León, G. Madurga, La Rábida, Huelva, Spain, 25 June–6 July 1990, World Scientific, Singapore, 395–415, 1991.
39. A. Yamato, 'Radiochemical Separation Techniques for the Determination of Low Level Artificial Radionuclides in Marine Samples', in *Low-level Measurements of Man-made Radionuclides in the Environment*, Proc. of the Second International Summer School, eds. M. García-León, G. Madurga, La Rábida, Huelva, Spain, 25 June–6 July 1990, World Scientific, Singapore, 195–216, 1991.
40. Y. Yu, H.E. Bjørnstad, B. Salbu, 'Determination of Plutonium-239 + Plutonium-240 and Plutonium-241 in Environmental Samples Using Low-level Liquid Scintillation Spectrometry', *Analyst*, **117**, 439–442 (1992).
41. J. Vives i Batlle, 'Speciation and Bioavailability of Plutonium and Americium in the Irish Sea and other Marine Ecosystems', PhD Thesis, National University of Ireland, Dublin, 1993.
42. P. De Regge, R. Boden, 'Review of Chemical Separation Techniques Applicable to Alpha Spectrometric Measurements', *Nucl. Instrum. Methods Phys. Res.*, **223**, 181–187 (1984).
43. C.W. Sill, 'Determination of Gross Alpha, Plutonium, Neptunium and/or Uranium by Gross Alpha Counting on Barium Sulfate', *Anal. Chem.*, **52**, 1452 (1980).
44. H.D. Livingston, J.K. Cochran, 'Determination of Transuranic and Thorium Isotopes in Ocean Water: in Solution and in Filterable Particles', *J. Radioanal. Nucl. Chem.*, **118**(2), 299–308 (1987).
45. D.R. Mann, L.D. Surprenant, S.A. Casso, 'In Situ Chemisorption of Transuranics from Seawater', *Nucl. Instrum. Methods Phys. Res.*, **223**, 235–238 (1984).
46. M.T. Crespo, M.L. Aceña, E. García-Toraño, 'Adsorption of Some Actinide Elements on MnO_2', *Sci. Total Environ.*, **70**, 253–263 (1988).
47. M.T. Crespo, J.M. Los Arcos, C.E. Granados, M.L. Aceña, 'On the Concentration and Determination of Plutonium in Natural Waters by Adsorption on MnO_2 Filters', *J. Radioanal. Nucl. Chem.*, **130**(1), 99–110 (1989).

48. M.T. Crespo, J.L. Gascon, M.L. Aceña, 'Techniques and Analytical Methods in the Determination of Uranium, Thorium, Plutonium, Americium and Radium by Absorption on Manganese Dioxide', *Sci. Total Environ.*, **130**, 383–391 (1993).
49. J.L. Gascón, M.T. Crespo, M.L. Aceña, 'Anomalies in Thorium Adsorption on Manganese Dioxide', *J. Radioanal. Nucl. Chem.*, **139**, 249–260 (1990).
50. A. Nevissi, W.R. Schell, 'Efficiency of a Large Volume Water Sampler for Some Radionuclides in Salt and Freshwater', in *Proc. Fourth National Symposium on Radioecology*, ed. C.E. Cushing, Jr, Corvallis Oregon, 12–14 May 1975, The Ecological Society of America, Spec. Publ. No. 1, Dowden, Hutchinson and Ross Inc., Stroudsburg, PA, 277–282, 1976.
51. W.R. Schell, A. Nevissi, D. Huntamer, 'Sampling and Analysis for Pu and Am in Natural Waters', *Mar. Chem.*, **6**, 143–153 (1978).
52. W.B. Silker, R.W. Perkins, H.C. Rieck, 'A Sampler for Concentrating Radionuclides from Natural Waters', *Ocean Eng.*, **2**, 49–55 (1971).
53. A. Nevissi, W.R. Schell, 'Distribution of Plutonium and Americium in Bikini Atoll Lagoon', *Health Phys.*, **28**, 539–547 (1975).
54. R.G. Anderson, *Sample Pretreatment and Separation*, Analytical Chemistry by Open Learning Series, ed. N.B. Chapman, John Wiley & Sons, London, 1987.
55. H.A. Boyd, B.C. Eutsler, J.F. McInroy, in *Actinides in Man and Animals*, ed. M.E. Wrenn, RD Press, Salt Lake City, 43, 1981.
56. J.H. Harley, 'Manual of Standard Procedures', Health and Safety Laboratory Report NYO-4700-3, New York, 1970.
57. J.W.T. Meadows, J.S. Schweiger, B. Mendoza, R. Stone, 'Procedure for Plutonium Analysis of Large (100 g) Soil and Sediment Samples', in *Reference Methods for Marine Radioactivity Studies II*, International Atomic Energy Agency, Vienna, 89–90, 1975.
58. N.P. Singh, M.E. Wrenn, 'Tracers and Methods for Determining Thorium and Uranium in Biological Samples', in *Actinides in Man and Animals*, ed. M.E. Wrenn, RD Press, Salt Lake City, 53, 1981.
59. S. Ballestra, R. Fukai, 'An Improved Radiochemical Procedure for Low Level Measurements of Americium in Environmental Matrices', *Talanta*, **30**(1), 45–48 (1983).
60. W.R. Schell, M.J. Tobin, 'Quality Assurance Manual Submitted to the Pennsylvania D.E.R. Appendix A. Radiochemical Procedures', Report to the Pennsylvania Department of Energy Resources, University of Pittsburgh, A4–A48, 1989.
61. A. Yamato, 'An Anion Exchange Method for the Determination of Americium-241 and Plutonium in Environmental Samples', *J. Radioanal. Chem.*, **75**(1–2), 263–273 (1982).
62. K. Sekine, T. Imai, A. Kasai, 'Liquid–Liquid Extraction Separation and Sequential Determination of Plutonium and Americium in Environmental Samples by Alpha Spectrometry', *Talanta*, **34**(6), 567–570 (1987).
63. N.P. Singh, S.A. Ibrahim, N. Cohen, M.E. Wrenn, 'Solvent Extraction Method for Determination of Plutonium in Soft Tissue', *Anal. Chem.*, **50**, 357–360 (1979).
64. W.R. Alexander, T.M. Shimmield, 'Microwave Oven Dissolution of Geological Samples: Novel Application in the Determination of Natural Series Radionuclides', *Radioanal. Nucl. Chem. Lett.*, **145**(4), 301–310 (1990).
65. M. Roig, M.M. Ribera, G. Rauret, 'Application of the Microwave-oven to the Pretreatment of Macrosamples in Environmental Radioactivity Monitoring', *J. Radioanal. Nucl. Chem.*, **190**(1), 59–69 (1995).
66. J.G. Sengupta, N.B. Bertrand, 'Direct ICP-MS Determination of Trace and Ultratrace Elements in Geological Materials after Decomposition in a Microwave Oven. I. Quantification of Y, Th, U and the Lanthanides', *Talanta*, **42**(11), 1595–1607 (1995).
67. J.S. Alvarado, T.J. Neal, L.L. Smith, M.D. Erickson, 'Microwave Digestion of Plant Tissue and the Subsequent Determination of Trace Lanthanide and Actinide Elements by Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **322**(1–2), 11–20 (1996).
68. J.A. Sánchez Cabeza, P. Masqué, I. Ani-Ragolta, '^{210}Pb and ^{210}Po Analysis in Sediments and Soils by Microwave Acid Digestion', *J. Radioanal. Nucl. Chem.*, **61**, 209–217 (1998).
69. M.B. Lovett, D.M. Nelson, 'Determination of Some Oxidation States of Plutonium in Seawater and Associated Particulate Matter', in *Techniques for Identifying Transuranic Speciation in Aquatic Environments*, Ispra, Italy, 24–28 March 1980, International Atomic Energy Agency, Vienna (STI/PUB/613), 27–35, 1981.
70. N.P. Singh, 'Radiochemical Procedures for the Determination of Actinides in Biological and Environmental Samples', in *Low Level Measurements and their Applications to Environmental Radioactivity*, Proc. I International Summer School, eds. M. García León, G. Madurga, La Rábida, Huelva, Spain, 28 September–9 October 1987, World Scientific, Singapore, 471–490, 1988.
71. L.L. Smith, F. Markin, T. Ten Kate, 'Comparison of Acid Leachate and Fusion Methods to Determine Plutonium and Americium in Environmental Samples', Argonne National Laboratory, Argonne Report ANL/ACL 92/2, 1992.
72. C.W. Sill, 'Determination of Thorium and Uranium Isotopes in Ores and Mill Tailings by Alpha Spectrometry', *Anal. Chem.*, **49**(4), 618–621 (1977).
73. L. Wetter, F. Markun, T. Tenkate, 'Determination of Refractory Actinides in Environmental Samples. Is Total Dissolution Necessary?', in *Proc. 34th Annual*

Conference on Bioassay, Analytical and Environmental Radiochemistry, Las Vegas, 1988.
74. C.W. Sill, D.S. Sill, 'Sample Dissolution', *Radioact. Radiochem.*, **6**(2), 8–14 (1995).
75. M. Gascoyne, J.P.A. Larocque, 'A Rapid Method of Extraction of Uranium and Thorium from Granite for Alpha Spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **223**, 250–252 (1984).
76. A.E. Lally, 'Chemical Procedures', in *Uranium-series Disequilibrium: Applications to Earth, Marine, and Environmental Sciences*, eds. M. Ivanovich, R.S. Harmon, Clarendon Press, Oxford, 95–126, 1992.
77. K. Bunzl, W. Kracke, 'Simultaneous Determination of Plutonium and Americium in Biological and Environmental Samples', *J. Radioanal. Nucl. Chem.*, **115**(1), 13–21 (1987).
78. M.E.D. Bains, P.E. Warwick, 'The Separation of Actinides from Lanthanides by Anion Exchange in Methanol/Hydrogen Chloride Medium, and its Applications to Routine Separation', *Sci. Total Environ.*, **130/131**, 437–445 (1993).
79. N.P. Singh, C.J. Zimmerman, L.L. Lewis, M.E. Wrenn, 'Quantitative Determination of Environmental Levels of Uranium, Thorium and Plutonium in Bone by Solvent Extraction and Alpha Spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **223**, 558–562 (1984).
80. S.G. Thompson, G.T. Seaborg, 'The First Use of Bismuth Phosphate for Separating Plutonium from Uranium and Fission Products', in *Progress in Nuclear Science. Series III. Process Chemistry*, eds. F.R. Bruce, J.M. Fletcher, H.H. Hyman, J.J. Katz, Pergamon Press, London, Vol. 1, 1956.
81. R.P. Bernabee, 'A Rapid Method for the Determination of Americium–Curium, Plutonium and Thorium in Biological and Environmental Samples', *Health Phys.*, **44**(6), 688–691 (1983).
82. E.E. Campbell, W.D. Moss, 'Determination of Plutonium in Urine by Ion Exchange', *Health Phys.*, **11**, 734–742 (1965).
83. C.W. Sill, K.W. Puphal, F.D. Hindman, 'Simultaneous Determination of Alpha-emitting Nuclides of Radium Through Californium in Soil', *Anal. Chem.*, **46**(12), 1725–1737 (1974).
84. R.L. Williams, G.E. Grothaus, 'Determination of Actinides at the Radiological and Environmental Sciences Laboratory', *Nucl. Instrum. Methods Phys. Res.*, **223**, 200–203 (1984).
85. W. Rieman, H.F. Walton, *Ion-exchange in Analytical Chemistry*, Pergamon Press, Oxford, 1970.
86. K.M. Wong, V.E. Noshkin, B.T. Bowen, 'Radiochemical Procedures Now Used at WHOI for the Analysis of Strontium, Antimony, Rare Earths, Caesium and Plutonium in Seawater Samples', in *Reference Methods for Marine Radioactivity Studies*, eds. Y. Nishiwaki, R. Fukai, International Atomic Energy Agency Report 118, Vienna, 119–127, 1970.
87. R.F. Anderson, A.P. Fleer, 'Determination of Natural Actinides and Plutonium in Marine Particulate', *Anal. Chem.*, **54**, 1142–1147 (1982).
88. P. Germain, P. Guegueniat, S. May, G. Pinte, 'Measurement of Transuranic Elements, Chiefly ^{237}Np (by Neutron Activation Analysis), in the Physical and Biological Compartments of the French Shore of the English Channel', *J. Environ. Radioact.*, **5**, 319–331 (1987).
89. V.R. Casella, C.T. Bishop, A.A. Glosby, EPA-600/7–80-019, 1980.
90. S.G. Bhat, D. Krishmaswami, D. Lal, W.S. Moore, 'Thorium-234/Uranium-238 Ratios in the Ocean', *Earth Planet. Sci. Lett.*, **5**, 483–491 (1969).
91. J. Kuwabara, M. Yamamoto, S. Oikawa, K. Komura, D.J. Assinder, 'Measurements of Tc-99, Cs-137, Np-237, Pu Isotopes and Am-241 in Sediment Cores from Intertidal Coastal and Estuarine Regions in the Irish Sea', *J. Radioanal. Nucl. Chem.*, **240**(2), 593–601 (1999).
92. Q. Chen, H. Dahlgaard, S.P. Nielsen, A. Aarkrog, 'Determination of ^{237}Np in Sediments and Seawater', in *Proc. Int. Symp. on Marine Pollution*, IAEA-TECDOC-1094, International Atomic Energy Agency, Vienna, 346–347, 1999.
93. D.J. Assinder, 'A Review of the Occurrence and Behaviour of Neptunium in the Irish Sea', *J. Environ. Radioact.*, **44**(2–3), 335–347 (1999).
94. M. Sakanoue, 'Transuranium Nuclides in the Environment', *Radiochim. Acta*, **42**, 103–112 (1987).
95. M.C. de Bortoli, 'Radiochemical Determination of Plutonium in Soil and other Environmental Samples', *Anal. Chem.*, **39**(3), 375–377 (1967).
96. N.A. Talvitie, 'Radiochemical Determination of Plutonium in Environmental and Biological Samples by Ion Exchange', *Anal. Chem.*, **43**(13), 1827–1830 (1971).
97. E. Holm, S. Ballestra, R. Fukai, 'A Method for Ion-exchange Separation of Low Levels of Americium in Environmental Samples', *Talanta*, **26**, 791–794 (1979).
98. H. Mussalo, T. Jaakkola, J.K. Miettinen, K. Laiho, 'Distribution of Fallout Plutonium in Southern Finns', *Health Phys.*, **39**, 245–255 (1980).
99. I.K. Kressin, 'Separation of Plutonium in Urine Without Sample Ashing for Determination by α-Spectrometry', *Anal. Chem.*, **53**, 1270–1274 (1981).
100. H.W. Kirby, W.E. Sheenan, 'Determination of ^{238}Pu and ^{241}Am in ^{239}Pu by Alpha-spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **223**, 356–359 (1984).
101. IAEA, *Measurement of Radionuclides in Food and the Environment*, International Atomic Energy Agency, Vienna, 1989.
102. P.E. Warwick, I.W. Croudace, R. Carpenter, 'Review of Analytical Techniques for the Determination of ^{241}Am in Soils and Sediments', *Appl. Radiat. Isot.*, **47**, 627–642 (1996).

103. B.F. Myasoedov, I.A. Lebedev, 'Separation and Determination of Transplutonium Elements', *CRC Crit. Rev. Anal. Chem.*, **15**(4), 347–394 (1985).
104. C.N. Murray, G. Statham, 'Applications of a Solvent Extraction Procedure Using Di-2-ethyl Hexyl Phosphoric Acid (HDEHP) for the Separation of Some Transuranic Elements in Environmental Samples', *Dtsch. Hydrogr. Z.*, **29**, 69–75 (1976).
105. G.R. Choppin, R.J. Silva, 'Cation Exchange Separation of Rare Earth Elements Using Alpha-hydroxyisobutyrate', *J. Inorg. Nucl. Chem.*, **3**, 153–154 (1956).
106. L.I. Guseva, I.A. Lebedev, B.F. Myasoedov, G.S. Tikhomirova, 'Isolation of Transplutonium Elements on Ionites in Aqueous/Alcohol Media', *J. Inorg. Nucl. Chem., Suppl.*, **20**, 55–59 (1976).
107. M. Sakanoue, 'The Analysis of Transuranic Nuclides in the Environment', in *Low Level Measurements and their Applications to Environmental Radioactivity*, Proc. I International Summer School, eds. M. García León, G. Madurga, La Rábida, Huelva, Spain, 28 September–9 October 1987, World Scientific, Singapore, 382–391, 1988.
108. G.H. Morrison, H. Freiser, *Solvent Extraction in Analytical Chemistry*, Wiley & Sons, New York, 1957.
109. M. Taube, *Plutonium, a General Survey*, Kernchemie in Einzeldarstellungen 4, Verlag Chemie, Weinheim, 1974.
110. R.P. Bernabee, D.R. Percival, F.D. Hindman, 'Liquid–Liquid Extraction Separation of Plutonium and Americium', *Anal. Chem.*, **52**(14), 2351–2353 (1980).
111. F.E. Butler, 'Rapid Bioassay Methods for Plutonium, Neptunium, and Uranium', *Health Phys.*, **15**, 19–24 (1968).
112. S. Sumiya, S. Morita, K. Tobita, M. Kurabayashi, 'Determination of Tc-99 and Np-237 in Environmental Samples by Inductively-coupled Plasma-mass Spectrometry', *J. Radioanal. Nucl. Chem.*, **177**(1), 149–159 (1994).
113. H. Rameback, M. Skalberg, 'Separation of Neptunium, Plutonium, Americium and Curium from Uranium with Di-(2-ethylhexyl)-phosphoric Acid (HDEHP) for Radiometric and ICP-MS Analysis', *J. Radioanal. Nucl. Chem.*, **235**(1–2), 229–233 (1998).
114. N.P. Singh, C.J. Zimmerman, L.L. Lewis, M.E. Wrenn, 'Simultaneous Determinations of Alpha-emitting Isotopes of Uranium and Plutonium in Bone', *J. Radioanal. Nucl. Chem.*, **81/2**, 247–253 (1984).
115. T. Hashimoto, K. Taniguchi, H. Sugiyama, T. Sotobayashi, 'Simultaneous Determination of Uranium and Plutonium Isotopes in Soils by Means of Single Alpha Spectrometry', *J. Radioanal. Chem.*, **52**(1), 133–142 (1979).
116. I.M. Fisenne, P.M. Perry, 'The Determination of Plutonium in Tissue by Aliquat-336 Extraction', *Radiochem. Radioanal. Lett.*, **33**(4), 259–264 (1978).
117. A.M. Poskanzer, B.M. Foreman, 'A Summary of TTA Extraction Coefficients', *J. Inorg. Nucl. Chem.*, **16**, 323–336 (1961).
118. M.S. Milyukova, N.I. Gusev, I.G. Sentyurin, I.S. Sklyarenko, *Analytical Chemistry of Plutonium*, Ann Arbor–Humphrey Science Publishers, London, 1969.
119. W. Schulz, *The Chemistry of Americium*, Technical Information Center, Energy Research and Development Administration, US Department of Commerce, Springfield, 1976.
120. G.J. Ham, 'Determination of Actinides in Environmental Materials Using Extraction Chromatography', *Sci. Total Environ.*, **173/174**, 19–22 (1995).
121. C. Testa, D. Desideri, M.A. Meli, C. Roselli, 'Extraction Chromatography in Radioecology', *Radioact. Radiochem.*, **2**(4), 46–55 (1996).
122. R. Pilviö, 'Separation of U, Th, Pu, Am and Cm with TRU and UTEVA Resins', Eichrom User's Group Meeting, Karlsruhe, Germany, 1997.
123. Y. Muramatsu, S. Uchida, K. Tagami, S. Yoshida, T. Fujikawa, 'Determination of Plutonium Concentration and its Isotopic Ratio in Environmental Materials by ICP-MS after Separation Using Ion-exchange and Extraction Chromatography', *J. Anal. Atom. Spectrom.*, **14**(5), 859–865 (1999).
124. S.L. Maxwell, 'Rapid Actinide Separation Methods', *Radioact. Radiochem.*, **8**(4), 36–44 (1998).
125. I.W. Croudace, P. Warwick, R. Taylor, S. Dee, 'Rapid Procedure for Plutonium and Uranium Determination in Soils Using a Borate Fusion Followed by Ion-exchange and Extraction Chromatography', *Anal. Chim. Acta*, **371**(2–3), 217–225 (1998).
126. P. Goodall, C. Lythgoe, 'Rapid Separation of Uranium and Plutonium by Extraction Chromatography for Determination by Thermal Ionisation Mass Spectrometry', *Analyst*, **124**(3), 263–269 (1999).
127. H.E. Carter, P. Warwick, J. Cobb, G. Longworth, 'Determination of Uranium and Thorium in Geological Materials Using Extraction Chromatography', *Analyst*, **124**(3), 271–274 (1999).
128. P.E. Warwick, I.W. Croudace, A.A. Dale, 'An Optimised and Robust Method for the Determination of Uranium and Plutonium in Aqueous Samples', *J. Appl. Radiat. Isot.*, **50**(3), 579–583 (1999).
129. Berne, 'Use of Eichrom's TRU Resin in the Determination of Am, Pu and U in Air Filter and Water Samples', Environmental Measurements Laboratory Report, EML-575, 1995.
130. R.A. Boll, G.K. Sweitzer, R.W. Gerber, 'An Improved Actinide Separation Method for Environmental Samples', *J. Radioanal. Nucl. Chem.*, **220**(2), 201–206 (1997).
131. C.A. Moody, S.E. Glover, D.B. Stuit, R.H. Filby, 'Preconcentration and Separation of Thorium, Uranium, Plutonium and Americium in Human Soft Tissue by Extraction Chromatography', *J. Radioanal. Nucl. Chem.*, **234**(1–2), 183–187 (1998).

132. J.W. Morse, G.R. Choppin, 'Laboratory Studies of Plutonium in Marine Systems', *Mar. Chem.*, **20**, 73–89 (1986).
133. K.A. Orlandini, W.R. Penrose, D.M. Nelson, 'Pu(V) as the Stable Form of Oxidised Plutonium in Natural Waters', *Mar. Chem.*, **18**, 49–57 (1986).
134. G.R. Choppin, A. Kobashi, 'Distribution of Pu(V) and Pu(VI) in Seawater', *Mar. Chem.*, **30**, 241–247 (1990).
135. D.M. Nelson, M.B. Lovett, 'Oxidation States of Plutonium in the Irish Sea', *Nature*, **276**, 599–601 (1978).
136. M.A. Wahlgren, K.A. Orlandini, 'Comparison of the Geochemical Behaviour of Plutonium, Thorium and Uranium in Selected North American Lakes', *Environmental Migration of Long-lived Radionuclides*, IAEA-SM-257/89, International Atomic Energy Agency, Vienna, 757–774, 1982.
137. P.I. Mitchell, J. Vives Batlle, T.P. Ryan, W.R. Schell, J.A. Sánchez-Cabeza, A. Vidal-Quadras, 'Studies on the Speciation of Plutonium and Americium in the Western Irish Sea', in *Radionuclides in the study of Marine Processes*, eds. P.J. Kershaw, D.S. Woodhead, Elsevier Applied Science, London, 37–51, 1991.
138. P.I. Mitchell, J. Vives i Batlle, A.B. Downes, O.M. Condren, L. León Vintró, J.A. Sánchez-Cabeza, 'Recent Observations on the Physico-chemical Speciation of Plutonium in the Irish Sea and the Western Mediterranean', *Appl. Radiat. Isot.*, **46**(11), 1175–1190 (1995).
139. D. Rai, R.J. Serne, J.L. Swanson, 'Solution Species of Plutonium in the Environment', *J. Environ. Qual.*, **9**, 417–420 (1980).
140. K.L. Nash, J.M. Cleveland, T.F. Rees, 'Speciation Patterns of Actinides in Natural Waters: a Laboratory Investigation', *J. Environ. Radioact.*, **7**, 131–157 (1988).
141. D.M. Nelson, R.P. Larsen, W.R. Penrose, 'Chemical Speciation of Plutonium in Natural Waters', in *Environmental Research on Actinide Elements*, eds. J.E. Pinder, III, J.J. Alberts, K.W. McLeod, R.G. Scheckhise, US Department of Energy, New York, 27–48, 1987.
142. K.A. Orlandini, W.R. Penrose, B.R. Harvey, M.B. Lovett, M.W. Findlay, 'Colloidal Behaviour of Actinides in an Oligotrophic Lake', *Environ. Sci. Technol.*, **24**(5), 706–712 (1990).
143. G.R. Choppin, 'Aspects of Nuclear Waste Disposal of Use in Teaching Basic Chemistry', *J. Chem. Educ.*, **71**(10), 826–829 (1994).
144. R.J. Pentreath, D.S. Woodhead, P.J. Kershaw, D.F. Jeffries, M.B. Lovett, 'The Behaviour of Plutonium and Americium in the Irish Sea', *Rapp. P.-V. Reun. Cons. Int. Explor. Mer.*, **186**, 60–69 (1986).
145. R.F. Anderson, 'A Method for Determining the Oxidation State of Uranium in Natural Waters', *Nucl. Instrum. Methods Phys. Res.*, **223**, 213–217 (1984).
146. D. McCubbin, K.S. Leonard, 'A Preliminary Study to Assess the Effect of Some Seawater Components on the Speciation of Plutonium', *J. Radioanal. Nucl. Chem.*, **172**, 363–370 (1993).
147. A. Saito, R.A. Roberts, G.R. Choppin, 'Preparation of Solutions of Tracer Level Plutonium(V)', *Anal. Chem.*, **57**, 390–391 (1985).
148. K. Kandiah, 'Instrumental Factors which affect the Observed Spectrum', *Nucl. Instrum. Methods Phys. Res.*, **223**, 277–284 (1984).
149. E. Steinbauer, G. Bortels, P. Bauer, J.P. Biersack, I. Ahmad, 'A Survey of the Physical Processes which Determine the Response Function of Silicon Detectors to Alpha Particles', *Nucl. Instrum. Methods Phys. Res.*, **A339**, 102–108 (1994).
150. G. Bortels, C. Hurtgen, D. Santry, 'Nuclide Analysis on Low-statistics Alpha-particle Spectra: an Experimental Verification for Pu Isotopes', *Appl. Radiat. Isot.*, **46**(11), 1135–1144 (1995).
151. A.E. Lally, K.M. Glover, 'Source Preparation in Alpha Spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **223**, 259–265 (1984).
152. N.A. Talvitie, 'Electrodeposition of Actinides for Alpha Spectrometric Determination', *Anal. Chem.*, **44**(2), 280–283 (1972).
153. L. Hallstadius, 'A Method for the Electrodeposition of Actinides', *Nucl. Instrum. Methods Phys. Res.*, **223**, 266–267 (1984).
154. R.F. Mitchell, 'Electrodeposition of Actinide Elements at Tracer Concentrations', *Anal. Chem.*, **32**, 326–328 (1960).
155. K.M. Puphal, D.R. Olsen, 'Electrodeposition of Alpha-emitting Nuclides from a Mixed Oxalate–Chloride Electrolyte', *Anal. Chem.*, **44**(2), 284–289 (1972).
156. W.J. McDowell, B.L. McDowell, *Liquid Scintillation Alpha Spectrometry*, CRC Press, Boca Raton, FL, 1994.
157. J.W. McKlveen, W.J. McDowell, 'Liquid Scintillation Alpha Spectrometry Techniques', *Nucl. Instrum. Methods Phys. Res.*, **223**, 372–376 (1984).
158. A.L. Gray, 'The Origins, Realization and Performance of ICP-MS Systems', in *Applications of Inductively Coupled Plasma Mass Spectrometry*, eds. A.R. Date, A.L. Gray, Chapman and Hall, Glasgow, 1–42, 1989.
159. V.E. Noshkin, C. Gatrousis, 'Fallout ^{240}Pu and ^{239}Pu in Atlantic Marine Samples', *Earth Planet. Sci. Lett.*, **22**, 111–117 (1974).
160. L.K. Fifield, R.G. Cresswell, M.L. di Tada, T.R. Ophel, J.P. Day, A.P. Clacher, S.J. King, N.D. Priest, 'Accelerator Mass Spectrometry of Plutonium Isotopes', *Nucl. Instrum. Methods Phys. Res.*, **B117**, 295–303 (1996).
161. L.K. Fifield, A.P. Clacher, K. Morris, S.J. King, R.G. Cresswell, J.P. Day, F.R. Livens, 'Accelerator Mass Spectrometry of the Planetary Elements', *Nucl. Instrum. Methods Phys. Res.*, **B123**, 400–404 (1997).
162. P. Burger, K. De Backker, W. Schoenmaekers, 'Performance of Large Area Particle Detectors', in *Solid State Imagers and Their Applications*, SPIE, The International Society for Optical Engineering, Bellingham, USA, 38–44, Vol. 591, 1985.

163. P. Bauer, G. Bortels, 'Response of Si Detectors to Electrons, Deuterons and Alpha Particles', *Nucl. Instrum. Methods Phys. Res.*, **A299**, 205–209 (1990).
164. P. Burger, W. Schoenmaekers, 'PIPS-technology for α, β and γ Measurements', in *Aerosol Measurements and Nuclear Accidents: a Reconsideration*, Commission of the European Communities, EUR 11755, Brussels, 136–152, 1988.
165. J. Kemmer, 'Improvement of Detector Fabrication by the Planar Process', *Nucl. Instrum. Methods Phys. Res.*, **226**, 89–93 (1984).
166. J.R. Batten, 'Evaluation of Passivated Ion Implanted Planar Silicon Detectors for the Spectroscopy and Assay of Low Energy Electrons', *Nucl. Instrum. Methods Phys. Res.*, **A277**, 478–487 (1989).
167. American Chemical Society's Committee on Environmental Improvement, 'Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry', *Anal. Chem.*, **52**, 2242–2249 (1980).
168. C.J. Passo, G.T. Cook, *Handbook of Environmental Liquid Scintillation Spectrometry: Compilation of Theory and Methods*, Packard, Meriden, 1994.
169. F.D. Brooks, 'Development of Organic Scintillators', *Nucl. Instrum. Methods Phys. Res.*, **162**, 477–505 (1979).
170. C.K. Kim, Y. Tagaku, M. Yamamoto, K. Kawamura, K. Shiraishi, Y. Igarashi, S. Igarashi, H. Takayama, N. Ikeda, 'Determination of ^{237}Np in Environmental Samples Using Inductively Coupled Plasma–Mass Spectrometry', *J. Radioanal. Nucl. Chem.*, **132**, 131 (1989).
171. T.J. Shaw, R. Francois, 'A Fast and Sensitive ICP-MS Assay for the Determination of ^{230}Th in Marine Sediments', *Geochim. Cosmochim. Acta*, **55**, 2075–2078 (1991).
172. J.S. Crains, L.L. Smith, 'Enhanced Measurement of Long-lived Actinides by Inductively-coupled Plasma Mass Spectrometry', *J. Radioanal. Nucl. Chem.*, **194**, 133–139 (1995).
173. A.E. Eroglu, C.W. McLeod, K.S. Leonard, D. McCubbin, 'Determination of Plutonium in Seawater Using Co-precipitation and Inductively Coupled Plasma Mass Spectrometry with Ultrasonic Nebulisation', *Spectrochim. Acta, Part B*, **53**(6–8), 1221–1233 (1998).
174. S. Sturup, H. Dahlgaard, S.P. Nielsen, 'High Resolution Inductively Coupled Plasma Mass Spectrometry for the Trace Determination of Plutonium Isotopes and Isotope Ratios in Environmental Samples', *J. Anal. Atom. Spectrom.*, **13**(12), 1321–1326 (1998).
175. I. Rodushkin, P. Lindahl, E. Holm, P. Roos, 'Determination of Plutonium Concentrations and Isotope Ratios in Environmental Samples with a Double-focusing Sector Field ICP-MS', *Nucl. Instrum. Methods Phys. Res.*, **A423**(2–3), 472–479 (1999).
176. J.B. Truscott, L. Bromley, P. Jones, E.H. Evans, J. Turner, B. Fairman, 'Determination of Natural Uranium and Thorium in Environmental Samples by ETV-ICP-MS after Matrix Removal by On-line Solid Phase Extraction', *J. Anal. Atom. Spectrom.*, **14**(4), 627–631 (1999).
177. K.O. Buesseler, J.E. Halverson, 'The Mass Spectrometric Determination of Fallout ^{239}Pu and ^{240}Pu in Marine Samples', *J. Environ. Radioact.*, **5**, 425–444 (1987).
178. J.E. Halverson, 'A Review of the Applications of Mass Spectrometry to Low Level Radionuclide Metrology', *Nucl. Instrum. Methods Phys. Res.*, **223**, 349–355 (1984).
179. J. Riegel, R. Deissenberger, G. Herrmann, S. Kohler, P. Sattelberger, N. Trautmann, H. Wendeler, F. Ames, H.J. Scheerer, F.J. Urban, 'Resonance Ionization Mass-spectrometry for Trace Analysis of Neptunium', *Appl. Phys. B*, **56**(5), 275–280 (1993).
180. D.L. Donohue, J.P. Young, 'Detection of Plutonium by Resonance Ionisation Mass Spectrometry', *Anal. Chem.*, **55**, 378–379 (1983).
181. D.L. Donohue, D.H. Smith, J.P. Young, H.S. McKown, C.A. Pritchard, 'Isotopic Analysis of Uranium and Plutonium Mixtures by Resonance Ionisation Mass Spectrometry', *Anal. Chem.*, **56**, 379–381 (1984).
182. A.E. Litherland, 'Accelerator Mass Spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **B5**, 100–108 (1984).
183. D. Elmore, F.M. Phillips, 'Accelerator Mass Spectrometry for Measurement of Long-lived Radioisotopes', *Science*, **236**, 543–550 (1987).
184. W. Wölfli, 'Advances in Accelerator Mass Spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **B29**, 1–13 (1987).
185. L.K. Fifield, T.R. Ophel, G.L. Allan, J.R. Bird, R.F. Davie, 'Accelerator Mass Spectrometry at the Australian National University's 14UD Accelerator: Experience and Developments', *Nucl. Instrum. Methods Phys. Res.*, **B52**, 233–237 (1990).
186. K.W. Allen, 'Accelerator Mass Spectrometry (AMS) of Heavy Elements ($M \geq 40$)', *J. Phys. G: Nucl. Part. Phys.*, **17**, S349–S354 (1991).
187. M. Suter, 'Accelerator Mass Spectrometry – Fascinating Applications and their Technical Challenges', *Nucl. Instrum. Methods Phys. Res.*, **B64**, 321–329 (1992).
188. E. Boraetto, D. Berkovits, R. Delmas, R.R. Johnson, A. Kaufman, M. Magaritz, M. Paul, M. Pourchet, 'Measurements of Anthropogenic Radionuclides in Environmental Samples', *Nucl. Instrum. Methods Phys. Res.*, **B92**, 350–356 (1994).
189. J. Rucklidge, 'Accelerator Mass Spectrometry in Environmental Geoscience. A Review', *Analyst*, **120**, 1283–1290 (1995).
190. L.K. Fifield, 'Accelerator Mass Spectrometry and its Applications', *Rep. Prog. Phys.*, **62**, 1223–1274 (1999).
191. L.R. Kilius, N. Baba, M.A. Garwan, M. Litherland, M.J. Nadeau, J.C. Rucklidge, G.C. Wilson, X.L. Zhao, 'AMS of Heavy Ions with Small Accelerators', *Nucl. Instrum. Methods Phys. Res.*, **B52**, 357–365 (1990).
192. X.L. Zhao, M.J. Nadeau, L.R. Kilius, A.E. Litherland, 'The First Detection of Naturally-occurring U-236

with Accelerator Mass-spectrometry', *Nucl. Instrum. Methods Phys. Res.*, **B92**, 249–253 (1994).
193. K.H. Purser, L.R. Kilius, A.E. Litherland, X.L. Zhao, 'Detection of U-236: A Possible 100-Million Year Neutron Flux Integrator', *Nucl. Instrum. Methods Phys. Res.*, **113**, 445–452 (1996).
194. S. Ballestra, D. Vas, 'Analytical Quality Control – Quality Assurance IAEA-ILMR Intercomparison Programmes', in *Low-level Measurements of Man-made Radionuclides in the Environment*, Proc. of the Second International Summer School, eds. M. García-León, G. Madurga, La Rábida, Huelva, Spain, 25 June–6 July 1990, World Scientific, Singapore, 293–313, 1991.
195. ISO, 'Terms and Definitions Used in Connection with Reference Materials', ISO Guide 30, ISO, Geneva, 1981.

β-Particle Emitters, Determination of

David R. Caprette
Rice University, Houston, USA

1 Introduction **1**
2 β-Decay and Interactions **2**
2.1 Atomic Theory and Radioactive Decay 2
2.2 β-Decay 3
2.3 Interactions with Matter 5
3 β-Particle Detection and Quantitation **7**
3.1 General Properties of Instrumentation 7
3.2 Choice of Detection Method 9
3.3 Error Analysis, Statistics, and Quality Control 9
4 Scintillation Detectors **12**
4.1 Principles 12
4.2 Liquid Scintillation Detectors 13
4.3 Quenching and Quench Correction 13
4.4 Organic Crystal Detectors 18
5 Gas-filled Detectors **18**
5.1 Ionization Chambers 18
5.2 Proportional Counters 18
5.3 Geiger–Müller Counters 19
6 Solid-state Detectors **19**
7 Visual Imaging **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **21**

Radioisotopes have unstable nuclei that decay in a statistically predictable manner with the release of radiation in the form of particles or electromagnetic waves. In β-decay "fast electrons" called β-particles are released, which have a spectrum of kinetic energies ranging from near zero to a characteristic maximum (E_{max}). E_{max} can vary by orders of magnitude among different types of emitters. The low mass of β-particles causes them to lose energy quickly through molecular collisions, producing excitation of valence electrons, ionization of atoms, and in some cases specialized radiation (bremsstrahlung and Čerenkov radiation, and Compton scattering). Detection of β-particle emissions is complicated by the wide range of possible energies and the weakness of some emissions, so that a single type of detector is inadequate for the analysis of radiation from all β-particle emitters.

Detectors are characterized by a sensitive volume in which particles interact with matter to produce a signal, an example of which is the emission of photons. The signal can produce a continuous current or a pulse. In pulse mode it is desirable to choose a detector such that the pulse is proportional to the energy of the particle, giving the instrument the property of energy discrimination. This property permits spectral analysis and methods of quantitation that are based in part on energy spectra. Liquid scintillation counting (LSC) is the most accurate and widely used method for detecting low-energy β-particles. Usually the information obtained by LSC must be corrected for signal loss due to absorption of incident particles or emitted photons. This signal loss is called quenching. For quantitation of low-energy β-radiation, quench corrections are usually made using the sample channels ratio (SCR) or an external standard method. All such calculations are subject to error and derived values are subject to the compounding of precision error during mathematical manipulation of the data.

Scintillation counting with organic crystal detectors is the method of choice for high-energy β-particle emitters. Gas-filled detectors are of limited value for β-particle detection, although windowless proportional counters are appropriate for detection of low-energy particles, especially in gaseous form. When neither energy discrimination nor precise quantitation is required, Geiger–Müller counters are acceptable for high-energy β-particle detection. They can detect all but the least energetic of low-energy emitters, although Geiger–Müller counters are not practical for the analysis of activity from such emitters because the signal is very weak, and the energies from incident particles are frequently absorbed before reaching the sensitive volume. Silicon-barrier semiconductors support energy discrimination and are of value for some applications. Visual imaging with photographic film is used for localization of β-particle emitter activity, as in labeling of biological tissues or macromolecules.

1 INTRODUCTION

This article is designed to assist a new investigator in selecting appropriate instrumentation for detection, qualitative analysis, and/or quantitation of sources of β-radiation for research or for quality control applications such as contamination monitoring for safety purposes. The target group includes investigators who are unfamiliar with radiation detection or who have

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

experience in measurement of α- and/or γ-radiation but are not familiar with the requirements for accurate detection and quantitation of β-particle-emitting radionuclides.

Prior to selection of instrumentation, the investigator should be acquainted with the nature of β-decay and how β-particles interact with molecules of surrounding materials. Properties of detectors in general are discussed, and terminology of radiation detection is introduced. Criteria for the choice of detection method are presented along with a table outlining the advantages and disadvantages of the available detection methods for β-particles. Principles of operation and general operating procedures for useful detection systems are presented.

2 β-DECAY AND INTERACTIONS

2.1 Atomic Theory and Radioactive Decay

An atom exists as a positively charged nucleus surrounded by a number of negatively charged orbiting electrons. The nucleus consists of one or more positively charged protons, which give the atom its unique atomic number (Z), and a number of uncharged particles of similar mass, called neutrons. The identity of an element and its characteristic properties are conferred by its atomic number. The number of protons and neutrons combined give an element its characteristic atomic weight (A).[1]

Most elements can exist in the form of different isotopes, because atoms can have the same atomic number but different numbers of neutrons. The isotopes of a given element all have the same atomic number but vary in atomic mass. Different isotopes of an element are designated by the nearest integer to the atomic mass. For example, the most common form of phosphorus, with atomic number 15, has an atomic mass of 30.98 g mol^{-1} and is designated phosphorus-31. Phosphorus-32 is a less common isotope of phosphorus. It has the same number of protons, hence the same atomic number, but has one more neutron than phosphorus-31. Stable isotopes are those for which spontaneous changes in atomic structure either are not known to occur, or occur at a negligible rate (see below for a more precise definition). Eighty-one elements with one or more commonly occurring stable isotopes are known, with atomic numbers ranging from 1 (hydrogen) to 83 (bismuth).[2]

Unstable isotopes are any of those for which radioactive decay, in the form of changes in atomic mass, loss of energy, ejection of particles or photons, and/or atomic rearrangements, is known to take place. Atomic decay can result in the emission of one or more of several different types of ionizing radiation, of which β-radiation is one category. While the term isotope is used to refer to elements with constant atomic number but differing atomic masses, nuclide is a more general term used to distinguish any specific nuclear species such as tritium (hydrogen-3), phosphorus-32, or oxygen-16. The term radionuclide refers to any nuclide that is known to undergo radioactive decay.

The Heisenberg uncertainty principle requires that it is impossible to predict when a particular atom will decay. Thus, efforts to quantify ionizing radiation rely heavily on probability theory. For a known quantity of a radionuclide, it is predictable that within a given period of time a certain percentage of the material will undergo decay. The fundamental law of radioactive decay[3] (Equation 1) predicts the rate of disintegration of a number of radioactive nuclei (N) from the decay constant (λ) for the nuclide.

$$\frac{dN}{dt} = -\lambda N \tag{1}$$

Perhaps a more familiar measure of decay rate is the half-life, which is the time required for one-half of the atoms of a radionuclide to decay. For a starting amount A_0 of radionuclide with decay constant λ, the amount A remaining after time t is given by the decay equation (Equation 2).

$$A = A_0 e^{-\lambda t} \tag{2}$$

From Equation (2), when $A = A_0/2$, the elapsed time t gives the half-life of the radionuclide. Typical half-lives range from nanoseconds to several billion years. A stable nuclide can now be defined as an isotope for which no decay is observed to take place, or for which the half-life exceeds 5×10^8 years.

A reported half-life value for a radionuclide incorporates some assumptions. As a rule, the half-life of a radionuclide is independent of chemical bonding, the state of the matter, or environmental conditions such as temperature and pressure. However, chemical bonding has been observed to influence the half-life of radionuclides that undergo a particular type of decay, namely electron capture (section 2.2.3).[2] A radionuclide may undergo more than one type of decay before a stable nuclide is attained. Furthermore, different paths to the stable nuclide or lowest energy state (ground state) are possible. Even with alternative decay schemes, a rate of decay can be measured and reported.

The traditional unit of radioactivity is the curie (Ci), in which 1 Ci consists of 3.7×10^{10} atomic disintegrations per second (dps). This unit was developed from the best estimate for the decay rate of 1 g of pure radium-226.[3] Research laboratories usually work with levels of activity on the order of millicuries (mCi) or microcuries

(μCi). Since 1975, the becquerel (Bq) has replaced the curie as the preferred unit of radioactivity. One becquerel is defined as one disintegration per second, or 2.703×10^{-11} Ci. Thus a working amount of radioactivity in a laboratory would be measured in kilobecquerels (kBq) or megabecquerels (MBq). These units refer to the rate at which nuclei decay, which is not the same as the rate of emission of a particular type of energy. This qualification is necessary since atoms of a particular radionuclide may decay by different mechanisms, producing different types of radiation.

Radionuclides are seldom available as pure sources. Generally they are diluted in a larger concentration of the stable form of the same element. The specific activity of a source is defined as the amount of radioactivity per unit mass or volume of the material. Typical units of specific activity include curies per gram (Ci g^{-1}) or becquerels per gram (Bq g^{-1}).

The electron-volt (eV) is the traditional unit for measurement of the energy of radiation.[3] One electron-volt is defined as the kinetic energy given to an electron as it is accelerated through a potential difference of one volt. Characteristic energies of ionizing radiations are in the range of thousands or millions of electron-volts (keV or MeV, respectively). The energy of radiation from a specific source may be distributed in a continuous spectrum, in which case the maximum energy (E_{max}) is reported and/or the spectrum is represented graphically. When all emissions of a type have the same energy the emission is termed monoenergetic.

2.2 β-Decay

When a decay process results in a change in the atomic number Z of a nuclide, but no change in atomic mass A, the event is categorized as a β-decay. The change in Z takes place as a neutron is converted to a proton, or vice versa. Thus the atom is actually converted into a different element. β-Decay typically results in the ionization and/or excitation of electrons. An orbital electron exists in a stable bound state at a discrete energy level. Under quantum theory, a limited number of energy states is allowed for each electron. The process of excitation takes place when an orbital electron absorbs a photon whose energy is exactly equal to the difference between that electron's energy state and a higher one, and the electron is excited to the higher energy state. De-excitation takes place when an excited electron drops to a lower discrete energy state, emitting a photon. The energy needed to remove an electron from its orbital is called its binding energy. When an electron absorbs energy in excess of its binding energy it is ejected from the atom and becomes a free electron, forming an ion pair in a process called ionization.

Two other major types of radioactive decay are α and γ, respectively. Upon decay, an α-particle emitter ejects a helium nucleus, which is a particle consisting of two protons and two neutrons. The atom therefore loses two atomic number units and four atomic mass units. As with a β-particle emitter, an α-particle emitter is converted to a different element upon disintegration. However, unlike a β-particle emitter there is a significant change in atomic mass.

γ-Decay usually is secondary to one of the previously discussed types of decay. The emission of particles during α- or β-decay usually leaves the new nucleus in an excited energy state. γ-Decay involves a transition of the excited nucleus to a lower state, possibly the ground state, resulting in emission of a γ-ray photon. No change is seen in either atomic number or mass following γ-photon emission.

Three common types of β-radiation are recognized. In two of these modes the radiation is in the form of either a fast electron or a positron. The term ''fast electron'' refers to a free electron (i.e. one that is not in orbit about an atomic nucleus) with kinetic energy. In both types of decay the charged particle is called a β-particle, and its kinetic energy ranges from zero to a maximum that is characteristic of the radionuclide. A third common type of decay, electron capture, does not involve release of a particle but does result in emission of X-rays and Auger electrons (section 2.2.3). Other rare types of decay have been included in the category of β-particle emission but will not be treated in this article.

2.2.1 Fast Energetic Electron (β^-) Decay

A plot of neutron-to-proton ratio for stable isotopes reveals that for nuclear stability to be maintained, the neutron-to-proton ratio increases with increasing atomic number from 1 up to 1.5 for the element bismuth (atomic number 83). Elements with atomic number greater than bismuth are all unstable. β^--Decay is characteristic of nuclides with an unstable neutron-to-proton ratio in the nucleus, exceeding that of a stable nucleus.[1,2] That is, nuclei with an ''excess'' of neutrons typically undergo β^--decay. There is not necessarily a one-to-one correspondence between number of protons and number of neutrons, thus the word ''excess'', used here, means ''in excess of the number required to maintain a stable nucleus''. For example, the most common form of hydrogen consists of a proton with one orbital electron. Deuterium, a stable isotope, has one neutron in addition to its proton. The nucleus of tritium, which is unstable, consists of a proton and two neutrons.

During decay of a β^--particle emitter, a neutron is converted to a proton, an electron, and an electron antineutrino ($\bar{\nu}$). The mass of the atom remains unchanged

but the atomic number in this type of decay increases by one. For example, the common stable isotope of carbon has mass number 12 and atomic number 6. Carbon-14, an unstable isotope, has an atomic mass of 14, and an atomic number of 6. Following β^--decay the mass number remains 14 but the atomic number increases to 7, and thus the carbon atom is converted to a nitrogen atom (nitrogen-14). Similarly, tritium has one proton and two neutrons, hence mass number 3. Upon decay, the nucleus is converted to a stable isotope of helium (helium-3), with two protons and one neutron. In β^--decay the energy of disintegration can be calculated from the atomic mass energy difference between the radionuclide and the new nuclide that results from the decay, and the binding energy of the ejected electron.

The electron that is ejected from the nucleus during β^--decay is a β^--particle, also called a ''fast electron'' in reference to its kinetic energy. The kinetic energy of each β^--particle is a fraction of the energy released in the disintegration process, with the remaining energy imparted to the antineutrino. The energy of disintegration is randomly shared among three entities, namely the nucleus, the β^--particle, and the antineutrino. Thus, for any given type of β^--particle emitter, the kinetic energies of the fast electrons vary in a continuous spectrum from zero, when all of the energy not imparted to the nucleus is given to the antineutrino, to a characteristic maximum in which all the energy is acquired by the β^--particle and none by the antineutrino.[4] Each type of β^--particle emitter has a unique energy spectrum with a characteristic energy maximum E_{max} (Table 1), while emitters that decay by electron capture produce monoenergetic emissions.

For a comprehensive list of β^--particle emitters and properties the reader is referred to published tables of isotope data.[6] For purposes of discussion, emitters with energy maxima of less than 500 keV are considered to be of low energy, while emitters with energy maxima greater than 1000 keV are considered to be high-energy sources. When measuring sources with energies between these values, instrumentation that is suitable for either range should be considered. Tritium should be considered separately as its energy maximum is nearly an order of magnitude lower than the next weakest β^--source.

Table 1 Selected radioisotopes of interest in biological sciences[5]

Nuclide	E_{max} (keV)	Half-life
Tritium (hydrogen-3)	18.6	12.26 years
Carbon-14	156	5730 years
Magnesium-28	459	21.1 h
Phosphorus-32	1710	14.28 days
Sulfur-35	167	87.4 days
Chlorine-36	709 (98.1%)	3.01×10^6 years
Potassium-40	1314	1.27×10^9 years
Calcium-45	257	164 days
Rubidium-87	690 (8.8%)	18.7 days
	1770 (91.2%)	
Strontium-89	554 (~0.01%)	50.5 days
	1463 (~100%)	
Iodine-131	606 (89.7%)[a]	8.06 days

[a] Several other β^--decay modes are omitted for brevity.

2.2.2 *Fast Positron (β^+) Decay*

Positron decay is characteristic of nuclides with an unstable excess of protons.[2] In this decay mode, a proton is converted into a neutron, a positron (β^+), and a neutrino (ν). The mass number again remains the same, but the atomic number is reduced by one. The positron is also called a β-particle, just as a fast electron is called a β-particle in β^--decay, but the symbol is written as β^+. An example of a β^+-emitter is sodium-22, half-life 2.6 years.[1] With sodium-22 about 90.5% of decay processes result in emission of a β^+-particle, while 9.5% occur by electron capture (described in section 2.2.3).

A positron is very short-lived, typically surviving less than a microsecond.[1] Annihilation with release of energy takes place the instant the positron encounters an electron, which usually occurs within the source itself (Figure 1). In the annihilation process the energy in the masses of the two particles is converted to electromagnetic radiation, and the two photons are emitted in opposite directions (i.e. anticoincident photons), conserving linear momentum.[4] The kinetic energy associated with either incident particle is negligible compared with its resting mass energy of 0.51 MeV. Thus, detection of γ-photons with energy of 0.51 MeV is indicative of positron decay. Readers interested in detection of annihilation photons are encouraged to consult articles on γ-radiation determination.

2.2.3 *Electron Capture*

An alternative decay mode resulting from an excess of protons involves the capture of an electron from the

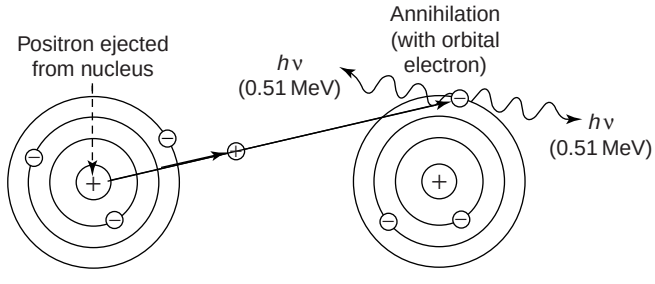

Figure 1 Positron annihilation with release of anticoincident photons.

electron shell of the nuclide. As with the other forms of β-decay the atomic number changes without a change in the atomic mass; however, no β-particle is emitted.[2] The atomic number goes up by one and although the correct number of orbital electrons is now present, a vacancy is left in one of the inner shells. When the vacancy is filled by an electron from one of the outer shells, monoenergetic X-rays are emitted with energy that is characteristic of the emitter. Alternatively, rearrangement of electrons following the capture results in transfer of excitation energy to an outer electron, which is then ejected from the atom. Such electrons are called Auger electrons, and have energies equal to the difference between the starting excitation energy of the atom and the binding energy of the orbit from which the electron was ejected.[3] Characteristic energies of Auger electrons are low relative to those of β-particles.

2.3 Interactions with Matter

The primary focus of this section will be interactions involving β^--particles. In addition, a phenomenon known as Compton scattering, which is characteristic of the interaction of photon radiation with matter, will be discussed because of the significance of Compton scattering in the use of external standards for scintillation counting (section 4.3.3.3).

In these discussions the emitted particles will be called the incident particles while the particles with which they interact will be called target particles. This article will at times use the term collision to describe an interaction between two subatomic components. It should be noted here that the term collision, while convenient, does not accurately describe interactions of radiation with matter. Particles do not physically collide as billiard balls collide in classic physics experiments; the interactions are more complex, although superficially they resemble events that we commonly call collisions. For example, the most common target particles of ionizing radiation are orbital electrons. An interaction between a β-particle and an orbital electron is in the form of an electrostatic force. In addition to electrons, target ''particles'' can also include nuclei, nucleons, or whole atoms. In all cases, the interactions are best described in terms of quantum physics, but for the purposes of this article the term collision will suffice.[4]

β-Particles can cause ionization and excitation of atoms and molecules as well as types of interaction that are not characteristic of heavy particle interaction. These include emission of bremsstrahlung (section 2.3.2) and Čerenkov (section 2.3.3) radiations.[1]

β-Particles tend to follow a tortuous path when encountering matter because, since they are themselves electrons, the mass of the particle is equal to that of an orbital electron. β-Particles tend to lose energy at a slow rate compared with heavy charged particles such as α-particles. Multiple interactions are possible, with the number of likely interactions and changes in direction depending on the initial kinetic energy of the particle (Figure 2).[3]

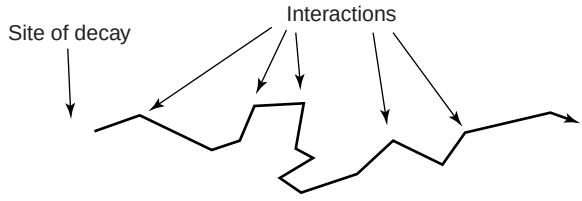

Figure 2 Hypothetical path of a β-particle in air.

2.3.1 Ionization and Excitation

Passage of a β-particle through a medium results in ionization and/or excitation of atoms. Ionization takes place when the energy transferred to an orbital electron exceeds the binding energy of the electron, causing the target particle to be ejected from the atom and creating an ion pair (Figure 3). Since a β-particle and an electron are of equal resting mass the result is a considerable loss of energy on the part of the incident β-particle and scattering of both the incident particle and electron in new directions. The secondary electrons that are so produced have sufficiently high kinetic energies to cause additional ionization and/or excitation events.

Typically about half of the energy in a β-particle is dissipated through ionization while the other half is dissipated through excitation.[4] When the energy transferred to an orbital electron is less than the binding energy the result can be excitation of the electron to a higher energy state, as shown in Figure 4. De-excitation results in the emission of a photon.

2.3.2 Bremsstrahlung Radiation

Bremsstrahlung radiation (braking radiation) becomes important as the atomic number of the radionuclide

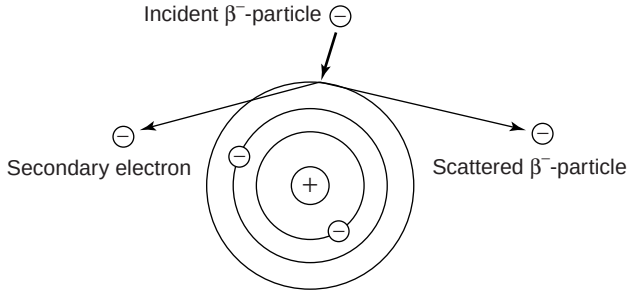

Figure 3 Ionization of target particle by incident β^--particle.

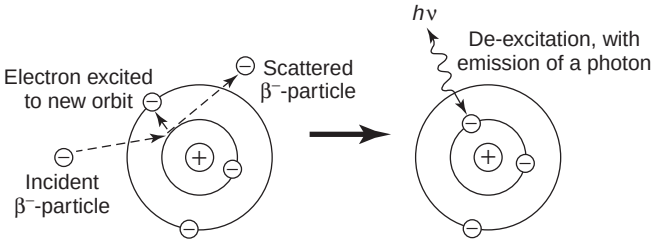

Figure 4 Excitation of an orbital electron by incident β-radiation (left) followed by de-excitation (right).

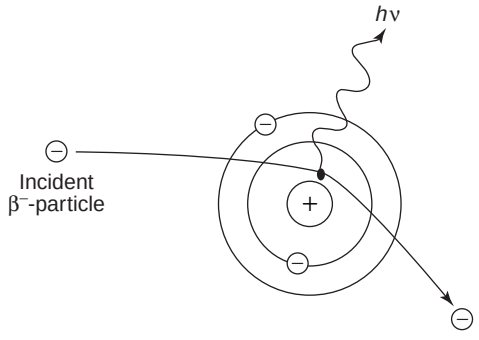

Figure 5 Bremsstrahlung radiation.

increases and as the energy of the β-particle increases. As a fast electron of sufficient energy approaches a nucleus, the electron is accelerated along a different path by electrostatic interactions. The encounter causes the electron to lose energy, which is emitted as electromagnetic radiation (Figure 5).[5] Such radiation can escape from detectors, a situation that becomes significant when energies exceed a few megaelectron-volts. Typical energies of β-particle emitters are no more than a few megaelectron-volts, and in this energy range bremsstrahlung contributes relatively little to the dissipation of energy of decay.[3]

2.3.3 Čerenkov Radiation

Very high energy emitters can produce β-particles that travel faster than the speed of light in a particular medium, such as water or clear acrylic.[1] This produces a kind of "shock wave" similar to that produced when an object exceeds the speed of sound in air. As the particle is slowed by the medium, visible photons in the blue-white region of the spectrum are emitted. Čerenkov radiation is responsible for the blue glow in water used as a cooling medium for nuclear reactors.

The detection of Čerenkov radiation is of interest in high-energy physics, and is beyond the scope of this article. Specialized devices used for detection of Čerenkov radiation are based on detection of this "Čerenkov light".

2.3.4 Compton Scattering

Compton scattering results from the interaction of a photon with an electron in the absorbing material. The effect deserves mention here because of the use of γ-ray sources as external standards in scintillation counting.[3,7] Any photon with sufficient kinetic energy, such as those emitted by γ-ray or X-ray sources, can cause the Compton effect. Energy is transferred from the incident ray to a target electron, leaving the photon with less energy than prior to the "collision" and ejecting an electron with kinetic energy nearly equal to that lost by the incident ray (Figure 6). There is much variation in the amount of energy transferred to the target electron during such events, as well as variation in direction of the electrons. Electrons that are scattered in this manner are known as recoil or Compton electrons.

Compton electrons interact with matter in a manner identical to β⁻-particles. The energy distribution due to Compton electrons is characterized by an abrupt increase in the distribution of energies near the high-energy end of

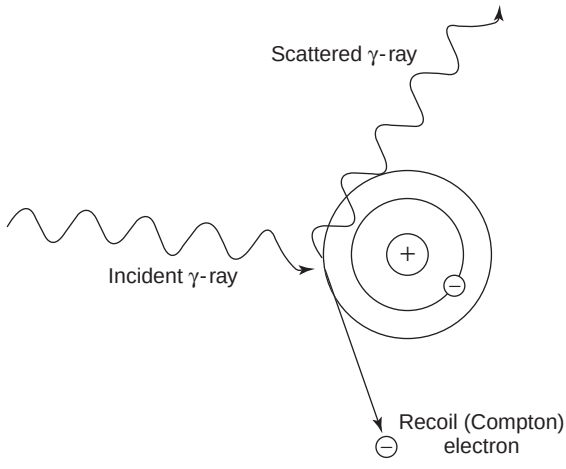

Figure 6 Scattering of orbital electrons by incident γ-radiation (Compton effect).

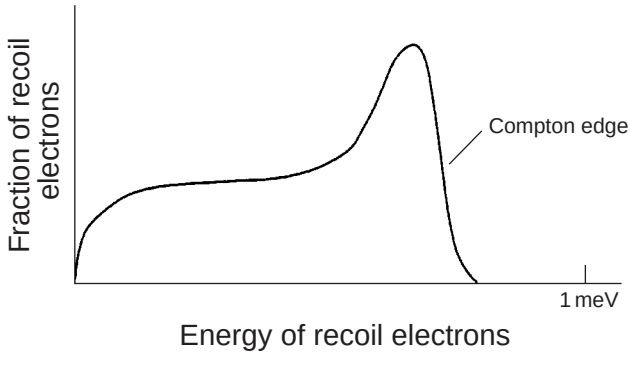

Figure 7 Hypothetical energy distribution of Compton electrons.

the spectrum followed by an even more abrupt decrease known as a Compton edge (Figure 7). These properties are exploited in the use of external gamma sources for quench correction (section 4.3.3.3).

3 β-PARTICLE DETECTION AND QUANTITATION

The principles on which detection of β-radiation is based include production of a measurable signal by ionization and/or de-excitation, where the size of the signal is proportional to the number of ion pairs or photons produced. Many factors participate in the determination of the number of such events per particle; however, for quantitation purposes the size of the signal is treated as proportional to the initial kinetic energy of the β-particle.

3.1 General Properties of Instrumentation

3.1.1 Detectors

An instrument designed for the detection and measurement of β-particle emissions consists of a detector (in which ionizing radiation produces a signal) and a signal processing mechanism.[4] A detector has three principal components. Radioactive decay takes place in a sensitive volume in which interactions between ionizing radiation and matter take place, producing ion pairs and subsequent events that can be detected. The sensitive volume requires a housing, the properties of which can be critical to the purpose of detection. The final component is a mechanism for delivery of a signal to a processing system. The sensitive volume can be a gas (ionization chambers, proportional counters, Geiger–Müller counters), a liquid (liquid scintillation counters), or a solid (semiconductor detectors, organic crystal detectors).

When ionizing radiation passes through a gas it can leave a trail of positive and negative ions, which migrate in an electric field. This generates a current that is directly proportional to the number of ion pairs produced. The generation of electric current by the formation of ion pairs in an electric field is the basis for detection by gas-filled detectors and semiconductor detectors. Excitation of electrons by either β-particles or secondary electrons leads to fluorescent de-excitation, which can be detected by a light-sensitive device such as a photomultiplier tube (PMT).[4] Scintillation counters and Čerenkov detectors both use light emission as a signal source.

3.1.2 Modes of Operation

A detector can be operated in current mode, in which the detector measures the average direct current resulting from β-particle emissions and subsequent interactions within a sensitive volume. This form of signal processing does not discriminate between individual events, but rather the signal is an average current that depends on the number of events per unit time and the average number of charge carriers produced per event.[3] For most laboratory applications a current mode of operation is inappropriate, since no information about the energy distribution of decay events is available. A current mode of operation is suitable when the intent is simply to monitor high-energy radiation from a known source, for example to detect fluctuations in radiation levels in a nuclear reactor.[4]

Most detectors operate in pulse mode, in which each quantum of radiation that produces a measurable signal is recorded as a separate event.[3] A pulse mode of operation can return the number of β-particles produced per unit time as well as the distribution of energies of the particles. The information obtained is essential to most forms of analysis. In pulse mode a detector returns a count rate, or counts per minute (cpm), the value of which is proportional to, but seldom equal to, the rate of actual radioactive decay (disintegrations per minute, or dpm).

3.1.3 Detector Efficiency

Efficiency (ε) of detection is a critical property of any radiation detector. Efficiency can be defined as the ratio of events recorded per unit time to the actual number of atomic disintegrations taking place in the sample during the same time period. The latter value is usually reported as dpm, or less frequently as dps. In pulse mode, efficiency would therefore be reported as cpm/dpm. Efficiency is affected by the properties of the detector as well as the type and energy of the emitter, the physical or chemical form of the emitter, and the surrounding gas, solvent molecules, or impurities that may interfere with detection or contribute spurious counts. Factors that may contribute to the efficiency with which β-decay is recorded are described next.

The geometry of a detector puts some limit on the number of radiations that are emitted in the direction of the sensitive volume. For example, in LSC photons must strike the sensitive surface of a PMT to record a count. Signals will be missed due to geometric constraints unless the sensitive surface completely surrounds the sample and all parts of the vial containing the sample are transparent to photons. In proportional counters samples can be introduced as a film on a support surface with only one side exposed to the sensitive volume. This would be described as a 2π counter, describing the angle subtended by the detector. If the sample is exposed on all sides of a support, the counter is described as having 4π geometry.

Some fraction of radiations may not reach the sensitive volume for reasons other than the geometry of the detector. The same types of interactions that take place between β-radiation and matter in the sensitive volume can take place within the sample itself, involving the radioactive nuclide, nonradioactive isotopes of the nuclide, contaminating materials, solvent molecules, etc. The result can be self-absorption, namely scattering of energy from radioactive decay before the signal leaves the sample. Depending on detector type, radioactive materials may be separated from the sensitive volume by a barrier called a window. A properly designed window should be transparent to measured radiation, but some attenuation may take place due to interactions of ionizing radiation with materials of the window itself. Loss of signal can also result from interactions with air molecules between the sample and the window. These types of losses, called material attenuation, are a major problem with detection of β-radiation, particularly from low-energy sources. In fact, the fraction of radiation from low-energy β-particle emitters that reaches the sensitive volume is usually quite small unless the sample is incorporated directly into the sensitive volume itself.

Efficiency is affected by the probability of the radiation actually interacting with the sensitive medium and producing a signal. This is of less concern when measuring β-radiation, since typically nearly 100% of charged particles do react with the sensitive volume.[4]

The efficiency of actual data recording is affected by several intrinsic factors not related to the sample itself, including detector dead time losses, coincidence losses, and incomplete capture. Dead time losses stem from the fact that the detector and/or signal processing mechanism may be inactive for a period of time following detection of an event.[5] Coincidence losses result from the occurrence of two or more nearly simultaneous decay events which are recorded as a single event of higher energy. Coincidence losses are a major factor when specific activities are very high with correspondingly high count rates.

3.1.4 Background

Radiation from sources other than the source to be analyzed contribute to the sum total of counts recorded. Typical sources of this background radiation include radioactive impurities in the constituent materials of the detector itself, as well as in the environment surrounding the detector. All materials including the atmosphere itself contain some background activity. Sources can be nearby or far away depending on the type of emitter. For example, a common source of background radiation in LSC is potassium-40, present as a small fraction of the natural potassium that is a component of glass vials. Cosmic radiation also produces a constant background radiation. To the extent that background radiation is not distinguishable from experimental activity, counts due to background are unavoidably included with counts recorded from the intended source.[3]

If emissions from the sample to be analyzed are low enough that the background contribution is significant, correction can be made by counting samples of identical composition but without the radionuclide of interest. For example, in LSC background counts are determined by counting a vial with the same volume and type of cocktail, but containing ''cold'' samples (samples treated identically to the labeled samples, but without radioactivity). Modern instruments can be programmed to automatically subtract background counts. However, background radiation is normally so low relative to source emissions that its contribution can be ignored.[4]

Under some circumstances background is sufficiently high to be a problem. In laboratories near accelerators, reactors, or in environments with extensive use of high-energy emitters, appropriate shielding may be necessary to reduce background radiation. Unfortunately, the shielding itself is often a source of background radiation due to impurities in the construction materials.[4] Shielding does not protect experiments in which very low levels of radiation must be detected. When the source is extremely weak, the investigator may have to resort to use of statistics to detect radiation from the source to be analyzed, over and above the counts due to background.

Spurious counts can be produced by such phenomena as chemiluminescence, which is sometimes encountered in LSC. PMTs are sensitive to fluctuations in supplied voltage and may deliver spontaneous pulses.

3.1.5 Signal Processing

The simplest pulse-mode processors only provide an output of cpm, and do not resolve the energies of individual emissions. A portable Geiger–Müller counter usually provides an analog output in the form of a galvanometer needle or audible clicks. Sophisticated processors can sort measured energies for individual emissions into specific channels or windows. Here the term window should be distinguished from its other use, to describe any barrier between a radioactive source and the sensitive volume of the detector. Rather than lumping all counts, regardless of energy of the source emission, into a single count total, each detected event can be logged as a count into one of two or more channels, each representing a specific range of apparent energies. By obtaining count totals for specific energy ranges, activity due to different emitters within a sample can be determined (see section 4.3.5), and signal losses can often be corrected (see section 4.3).

Processors can perform complex calculations to correct for signal losses resulting from detector geometry, inefficient capture, and other causes. Processors frequently include coincidence circuits to ensure that the pulse originated in the sensitive volume rather than in the detection device itself. Background radiation can be counted and automatically subtracted from experimental counts, and statistical calculations can be made to ensure consistent performance and efficient use of the instrumentation.

3.2 Choice of Detection Method

No single detection method is favorable for both low-energy and high-energy β-particle detection and measurement. Low-energy β-particles are readily stopped by physical barriers such as transparent windows between the source and the sensitive volume. Emitters with very low energy, especially tritium, have an extremely limited range in air alone. Even within a sensitive volume, a portion of the energy of a low-energy β-particle may be lost through interactions that produce no signal, such as absorption by impurities in the source. On the other hand, high-energy emitters require a sensitive volume of greater density and stopping power, so that all or most of the energy will be dissipated within the sensitive volume. For example, in a small volume a β-particle with high energy may penetrate the entire sensitive volume and collide with the wall of the detector itself. The escape of both high-energy β-particles and secondary radiation complicates attempts at energy discrimination.

The following questions should be considered when selecting a detection method.

- What specific emitters will be analyzed?
- Will detection be limited to low-energy β-particle emitters or low- and high-energy emitters?
- Will the same instrument be needed for detection of other types of radiation?
- Will it be necessary to count one type of emitter in the presence of another? (Detection methods that are suitable for β-particle determination may be entirely inappropriate for measurement of other types of radiation.)
- What is the nature of the application? For example, is simple detection all that is required, such as for contamination monitoring or detection of the presence/absence of label, is quantitation of activity necessary for analytical methods such as radioimmunoassays, or is there a need to obtain an energy spectrum for an unknown source?

The physical state of the radionuclide (gaseous, liquid, solid) in part determines the suitability of a particular method as well.

Ionization chambers are generally used for measurement of gaseous samples, but they are not suitable for detection of β-particle emitters. Many common β-radiation sources in laboratory use emit in the low-energy range, with E_{max} of less than 0.5 MeV. Provided the source is liquid or can be dissolved in liquid, LSC is by far the most suitable method of detection for most such applications. Proportional counters are also used for detection of low-energy emitters. Geiger–Müller counters can efficiently detect only the higher energy emissions and are not suitable for quantitation of low-energy β-particle activity, or for energy discrimination in any case. High-energy (>1 MeV) emitters can be detected using Geiger–Müller counters and spectra analyzed using scintillation counting with organic crystals. Table 2 lists recommended instrumentation for high- and low-energy β-particle determination. As suggested in the table, LSC offers the most versatile method of β-particle detection, and is the recommended detection method for the vast majority of individual laboratory applications.

3.3 Error Analysis, Statistics, and Quality Control

3.3.1 Counting Statistics

When any variable that is subject to statistical fluctuations is measured repeatedly, variation is observed among

Table 2 Recommended instrumentation for determination of β-particle emitters

Form	Purpose	Recommended instrumentation
Low-energy β-particles		
Soluble gas, liquid, solid	Detection, quantitation, spectroscopy	Liquid scintillation counter most preferred; proportional counter suitable
Gas, vapor-coated solid	Detection, quantitation, spectroscopy	Windowless proportional counter; Si-barrier semiconductor
High-energy β-particles		
Any form	Spectroscopy	Scintillation (organic crystal) preferred; Si-barrier semiconductor
Any form	Detection, quantitation only	Liquid scintillation counter preferred; proportional counter

measurements. Provided the experiment is well controlled and there is no bias in the sampling method, the values of individual measurements should be distributed symmetrically about a single value, called the sample mean or experimental mean $\bar{x}_e$. The determination of counting rates is subject to statistical principles as well. If a sample with relatively long half-life of known activity is counted for a fixed period of time and the count is repeated several times, each count typically yields a different value due to statistical variation.

In the field of statistics the true mean $\bar{x}$ is the average value of a measurement that would be obtained if every possible experiment could be conducted. In effect, the total number of samples (n) would be infinity. The experimental mean is an estimate of the true mean obtained by repeated sampling (Equation 3).

$$\bar{x}_e = \frac{\sum x_i}{n} \tag{3}$$

The distribution of individual data points (x_i) about the experimental mean can be represented by a single number, the variance σ^2 (Equation 4).

$$\sigma^2 = \frac{\sum (x_i - \bar{x}_e)^2}{n} \tag{4}$$

When a large number of samples is taken (usually >30) then a Gaussian or normal distribution can be used as a model for the distribution of individual data points about a mean. Applying a normal distribution to an experimental sample, the square root of the variance (called the standard deviation, σ) can be used to estimate the probability that an experimental mean is within a given percentage difference of the true mean. In this model the probability that the true mean is within one standard deviation of an experimental mean is 68.3%. The probability that it is within two standard deviations is 95.5%, and for three standard deviations it is 99.9%. To obtain a probability of less than 0.05 that an experimental mean is within a predetermined fraction of the true mean, sufficient data must be collected so that $2\sigma/\bar{x}_e$ is one-half of that fraction.

The notation σ^2 applies to variation among an infinite number of samples, that is, σ^2 is the variance that would be obtained if all possible experiments were to be performed. Experimental data represent a subset of this universe, thus a calculated variance is only an estimate of σ^2. Furthermore, σ^2 is underestimated when calculated from a data set, thus a truer estimate of σ^2 is given by the experimental variance s^2 (Equation 5). The experimental standard deviation s is then used as an estimate of σ.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \tag{5}$$

Suppose the investigator wishes to be confident that an experimental mean is within 1% of the value of the true mean, with a confidence level of <0.05 (<5% probability that the experimental mean is outside this limit). Then the true mean must be within the range $\bar{x}_e \pm 0.005\bar{x}_e$, that is, ±0.5%. To reach the desired level of confidence, sufficient data must be collected so that twice the standard deviation is equal to $0.005\bar{x}_e$, or, $\sigma = 0.0025\bar{x}_e$. For example, suppose an experimental mean is determined to be 40 000. Data points should be collected until $\sigma = (0.0025)(40\,000) = 100$.

The application of this method of assurance to a set of repeated measurements of the same variable is fairly straightforward. In counting radioactivity, however, there is a complication. An estimate of the actual rate of atomic disintegrations (dpm) is made for a sample by counting the activity for a set period of time, at the end of which the total counts divided by minutes of counting is used to determine cpm; cpm are then corrected to obtain dpm (these terms were defined in section 3.1.3). It would be trivial and time-consuming to make repeated counts, average them, and obtain mean cpm, variance, and standard deviation in the manner of repeated sampling. In fact, to the extent that statistical theory is based upon the probability of occurrence of real events rather than rates or other derived quantities, this approach would be incorrect.[3] An estimate of cpm in a sample is best determined from a single count x, collected over a set period of time. Thus the number of samples is one. The experimental mean is in fact equal to x, and s^2 cannot be calculated. For this situation the best estimate of the true variance is given by Equation (6).

$$\sigma^2 \cong s^2 = \bar{x} = x \tag{6}$$

The standard deviation is therefore best estimated by Equation (7).

$$\sigma \cong \sqrt{s^2} = \sqrt{x} \tag{7}$$

Using this model to obtain a probability of less than 0.05 that a cpm estimate is within a predetermined percentage of the cpm that correspond to the true dpm for the sample, twice the square root of the total counts must be less than or equal to half that percentage. For example, for a desired accuracy of 1%, twice the square root of total counts must be less than or equal to 0.5% of the total. That level is reached when total counts reach 160 000 (Equation 8).

$$\frac{2\sqrt{160\,000}}{160\,000} = 0.005 \tag{8}$$

When activities of samples vary considerably, the maximum counting time should be set in order to obtain a reasonably accurate count for the samples of lowest

expected activity. The instrument can be programmed to interrupt counting of a more active sample and move to the next sample when the desired level of accuracy is obtained. For example, programmable Beckman liquid scintillation counters allow the operator to select a "2σ (%)" value, which is simply 200 divided by the square root of the total count. Selection of a 2σ (%) value of 1.00 causes the instrument to cease counting when the total count reaches 160 000. Note that this method of error estimation is independent of the time of counting, and relies only on the total count obtained. In the example above counting would be interrupted whether it took 5 s or 5 min to reach 160 000. This allows for more efficient operation by avoiding prolonged counts and returns results more rapidly than if all counts were performed for a fixed interval.

In obtaining derived values from counts, such as ratios, rates, or other numbers that require various functional manipulations, the investigator must be aware that any uncertainty in the original measurement is carried throughout the process. The magnification of error in the course of obtaining calculated values is called error propagation.[1] The use of the same specific count more than once in arriving at a calculation should be avoided since this complicates error determination.[3] Let A and B equal two separate counts, with a and b the respective errors, so that count A can be represented as $(A \pm a)$ and count B can be represented as $(B \pm b)$. When multiplying or dividing either quantity by a constant, the error is multiplied or divided by the same constant. The error resulting from either summing A and B or subtracting one from the other is given by $\pm(a^2 + b^2)^{1/2}$. The product of A and B yields the error $\pm(a^2B^2 + b^2A^2)^{1/2}$. For A divided by B the error is given by Equation (9).

$$\text{error} = \pm\left(\frac{a^2}{B^2} + \frac{A^2}{B^4}b^2\right)^{1/2} \tag{9}$$

3.3.2 *Quality Control for Instrumentation Performance*

Instrumentation can be periodically checked by taking a series of counts while keeping the conditions as consistent as possible. Typically this might be 20–50 one-minute counts. An experimental mean is calculated and a chi-squared (χ^2) test conducted to determine if the experimental data fit a Gaussian distribution. The value of the χ^2 statistic is given by Equation (10).

$$\chi^2 = \frac{1}{\bar{x}_e}\sum(x - \bar{x}_e)^2 \tag{10}$$

If the individual counts do represent a normal distribution then the sample variance s^2 should equal the mean, which is assumed to be equal to the true variance σ^2. If the number of individual counts is n, then χ^2 is related to s^2 by the relationship in Equation (11), in which $n - 1$ represents the number of statistical degrees of freedom.

$$\chi^2 = \frac{(n-1)s^2}{\bar{x}_e} \tag{11}$$

The deviation of the ratio of s^2 to $\bar{x}_e$ from unity indicates the deviation of the sample data from the theoretical distribution. To estimate the probability that the difference in distribution of experimental data from the statistical model is not due to chance the χ^2 statistic should be calculated, and the probability determined from a χ^2 distribution table.

3.3.3 *Quality Control for Radiation Safety*

High-energy β-particle emitters such as phosphorus-32 release β-particles that can travel in air up to 6 m from the source, the length of a typical laboratory bench. Thus high-energy emitters represent an exposure hazard that requires shielding for protection of personnel. An acrylic shield with a thickness of 1 cm or greater effectively stops β-radiation from phosphorus-32. A survey meter (portable Geiger–Müller counter) can be used to detect contamination of benches, sinks, or instrumentation by a high-energy emitter. Radiation from low-energy emitters may travel as little as a few inches in air, and may not be sufficiently energetic to penetrate the window of a Geiger–Müller counter. Such emitters do not represent a major exposure hazard, but certainly represent an ingestion or direct contact hazard.

The form of the radionuclide affects the hazard level of the substance. For example, carbon-14 bound to carbon dioxide represents a very low hazard since carbon dioxide is quickly diluted in air and is not sequestered in the human body. Radioactive iodine vapors are in fact absorbed and accumulate in the thyroid gland to levels that can pose considerable risk. Tritium is the weakest β-particle emitter in common use, but its incorporation into biomolecules that are involved in biosynthetic pathways creates a moderate ingestion hazard. Tritium is undetectable by any device in which a window separates the source material from the sensitive volume.

In most laboratories using β-type radionuclides, wipe tests for contamination are regularly conducted. Most safety protocols require that wipe tests back up results obtained from a Geiger–Müller counter when using higher energy emitters. Wipe tests are essential for detection of contamination from very-low-energy emitters such as tritium and carbon-14. To conduct a wipe test following normal clean-up, a piece of filter paper or a cotton swab is dipped in an appropriate solvent and then wiped over the surface to be tested, usually covering 100 cm^2. The activity is then counted either by LSC (any β-particle

emitter) or organic crystal scintillation counting (high-energy emitter). Safety protocols always specify an action level: a significant number of cpm over background at which an area is determined to be contaminated. When the action level is exceeded, the contaminated area should be cleaned and re-tested.

4 SCINTILLATION DETECTORS

4.1 Principles

4.1.1 The Scintillation Process

When an incident emission contacts molecules of any material, it excites a number of electrons to a higher energy state, resulting in the emission of ultraviolet (UV) or visible light upon decay of the electrons to the normal state. The light flash that results is called a scintillation event, and materials that emit the light are called scintillants.[4] In scintillation counting, materials are selected so that the energy from the β-particle emission is efficiently transferred to molecules of the scintillant and the event can be detected by measuring the light emitted following de-excitation of electrons in the scintillant. Theoretically any material is a potential scintillant, however materials used for radiation detection are selected based on several criteria.

- To be a useful material for β-particle detection, scintillation must take place immediately upon interaction with a β-particle, that is, it must have the property of fluorescence. Delayed emission of photons, called phosphorescence, would make quantitation difficult or impossible.
- To relay information efficiently, upon de-excitation the scintillant must emit a large fraction of the incident energy as light rather than heat (a radiative process).
- A material used as a scintillant must not absorb its own emissions. The material must be transparent to the wavelengths emitted during the scintillation process.

4.1.2 Scintillator Materials

Organic scintillators include crystals, liquids, and plastic. β-Particle-emitting radionuclides can be measured using organic crystals such as anthracene and stilbene.[3] β-Radiation can also be assayed by direct incorporation of the source in plastic during polymerization. However, organic liquids consisting of solvent and dissolved scintillator molecules are most often employed for determination of β-activity. Solvents of choice include toluene, xylene, and 1,4-dioxan. The system may include a single organic scintillator, in which case it is known as a binary liquid scintillator, or it may include a primary and a secondary solute to improve the efficiency of detection of events.[4] Useful primary solutes include 2,5-diphenyloxazole (PPO) and 2-phenyl-5-(4-biphenylyl)-1,3,4-oxadiazole (PBD). The most commonly used secondary solute is 1,4-di-(2-[5-phenyloxazolyl])-benzene (POPOP).

Inorganic scintillators include crystals, glasses, and gases.[4] Alkali halide crystals such as sodium iodide or cesium iodide are the most widely used in scintillation detectors. They are more commonly used for measurement of γ-radiation, since energy transfer from γ-radiation is much more in a crystalline medium than in a liquid medium. They can be used for high-energy β-particle measurement as well, but an organic scintillator is preferred for high-energy β-particle detection.

4.1.3 Detection and Quantitation

Light emitted by the scintillation process is often in the UV range, requiring specialized UV-transparent materials or conversion to visible light. Visible light emitted by detection systems is usually too weak to be seen with the unaided eye, so sensitive PMTs are used to record scintillation events. Because weak light emissions are measured, the samples must be counted in complete darkness. Scintillation counters are constructed with a conveyor system that indexes samples, one by one, into a dark chamber containing the detection apparatus.

Photons emitted by scintillation materials enter a PMT through a semitransparent window, then generate free electrons through interaction with photosensitive material on the inside of the window. The electrons are accelerated toward an electrode called a dynode which is coated with material that is capable of emitting secondary electrons. Secondary electrons are produced through a series of dynodes of increasing positive charge, producing an amplification cascade resulting in the collection of a large number of electrons at the anode. A PMT has a fixed gain so that it functions as a high-gain amplifier. Because the pulse emitted by the PMT is proportional to the initial number of incident photons, the use of PMTs permits energy discrimination by the collector apparatus.

In scintillation counting, signals of different pulse heights are sorted into channels or windows which are determined by discriminator settings. A channel consists of a specific range of pulse heights and is determined by upper and lower discriminator settings, which are simply upper (ULs) and lower limits (LLs) for pulse heights that will be recorded in the channel. Discriminator settings are divided arbitrarily throughout the entire range of possible pulse heights–for example, from 0 to 1000. Usually two or three channels can be established for a count.

4.2 Liquid Scintillation Detectors

LSC is widely employed in the life sciences, including biomedical and biochemical research, when assay of low-energy β-particle-emitting radionuclides is essential. In LSC the source is dissolved in an organic solvent along with the scintillation materials. The incorporation of the radiation source into a liquid brings the source as close as possible to the scintillant molecules, thus making it possible to accurately record radiation from weak sources. The solvent system must be capable of dissolving the primary and/or secondary solute as well as the radiation source. In biology and medicine the source of radiation is often in an aqueous solvent, so the carrying capacity for water in the organic solvent can be a limiting factor. Modern systems contain emulsifiers and other additives that improve the capacity of the solvent for aqueous materials. A complete liquid system including solvent, additives, and primary and secondary solutes is called a cocktail.

In a binary liquid system, the β-particle interacts with the solvent molecules, causing ionization and excitation. Energy transfer among solvent molecules continues through molecular collisions throughout the medium.[7] Collisions with primary solute molecules result in excitation and de-excitation and the emission of photons. The wavelengths of photons emitted by primary solutes are usually within the UV range. The walls of glass vials would absorb such emissions, so special materials such as silica windows are needed to permit detection of UV light by a PMT. Secondary solutes serve as wavelength shifters, absorbing emissions from the primary solutes and emitting light in the visible range. Such emissions are not absorbed by clear glass. To serve as a secondary solute the material must have a lower excitation potential than the primary solute, which in turn has a lower excitation potential than the solvent molecules.

In addition to the problem of dissolving the source material, a disadvantage of LSC is the reduction or loss of signal due to (1) absorption of excitation energy by chemicals in the cocktail, (2) absorption of light by colored impurities in the system, or (3) radiation quenching due to differences in the effectiveness of energy transfer by radiations of different types and intensities.

4.3 Quenching and Quench Correction

4.3.1 Chemical Quenching

In the process of chemical quenching, impurities in the system absorb excitation energy, preventing transfer of energy from the solvent to the primary solute. The energy is converted to heat rather than light, and the signal is reduced. When weak radiation emissions are involved, the signal often goes undetected. Common sources of chemical quenching include the original solvent, often water, acidic or basic solutes that may have been used to solubilize the source, various salts and other impurities, and the source molecules themselves.

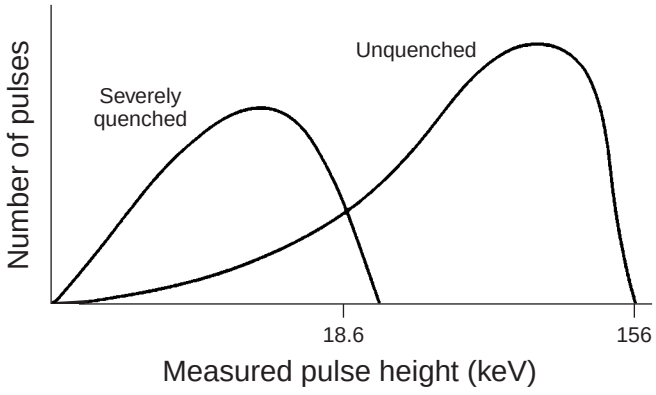

Figure 8 Estimated pulse heights for the same amount of carbon-14 activity in an unquenched and quenched sample, respectively. Note that the pulse height scale is nonlinear.

Chemical quenching may be unavoidable, particularly if a primary cause is the source molecule itself. Fortunately, chemical quenching is correctable using the SCR or Compton edge method. Figure 8 represents the distribution of pulse heights that would be observed from an unquenched sample versus a severely quenched sample.

4.3.2 Color Quenching

Color quenching results from the absorption of light emitted by either the primary or secondary solute, or both, by colored compounds in the source materials or solvent. A common source of color quenching in biomedical work is hemoglobin, for example. The degree of quenching depends on the properties of the absorbing substance, thus it is problematic to establish a quench curve as for other forms of quenching. Color quenching must be prevented by removal or bleaching of the interfering substance before the sample has been prepared for counting.

4.3.3 Quench Correction

4.3.3.1 Internal Standardization In the internal standardization method for quench correction, the efficiency of counting is measured directly by adding a known quantity of radioisotope to each sample, and relating cpm to known dpm.[7] Each sample is first counted in the normal manner, then a standard quantity of the same radionuclide is added to each vial and the vials are re-counted. The efficiency of counting is given by Equation (12). The standard must be the same type of emitter as is present in

the unknown sample, to ensure that the quenching effect is the same on both. To avoid statistical error in determining efficiency, the specific activity of the standard must be considerably greater than that of the sample.

$$\frac{(\text{cpm for standard + sample}) - (\text{cpm for sample only})}{\text{known dpm of the standard}} \quad (12)$$

Although this method has the advantages of simplicity and rapidity, there are disadvantages. Once the standard has been added it cannot be removed, so the sample is irreversibly altered. The solvent containing the standard may actually introduce an additional quenching agent. Additional hazards are created by the need to open sample vials and add additional substances. The use of expendable standards with high specific activity rather than sealed quenched standards presents a disposal problem.

4.3.3.2 Sample Channels Ratio Quench correction by the SCR takes advantage of two situations. First, β-particles are released with energies ranging from zero to some characteristic maximum. Second, quenching reduces the recorded pulse height in response to higher energy decay, in addition to abolishing the signal from the lowest energy events. Quenching causes a shift in the proportion of pulses in the high-energy portion of the spectrum to the total recorded pulses. For a single radionuclide, the ratio of cpm in the high-energy range of the spectrum to total cpm is proportional to the efficiency of counting. This method of quench correction is very effective for single-labeled samples, but is not suitable for correction of counts from dual-labeled samples.

In practice, two counting channels are selected. Let channel 1 encompass the entire range of possible pulse heights for the radionuclide. Channel 2 constitutes the range of pulse heights from some intermediate value (LL) to the maximum pulse height. Modern liquid scintillation counters that are equipped with programming automatically select appropriate channels for specific radioisotopes. A set of sealed quenched standards of known dpm is counted to yield channel 1 cpm and the SCR for each standard. Instead of sealed standards, quenched standards can be prepared by the operator. The quenching agent need not be the same agent responsible for quenching the experimental samples. To prepare quenched standards, equal activities of nuclide are added to each of a number of vials containing equal amounts of scintillation cocktail. Progressively increasing amounts of a quenching agent such as carbon tetrachloride are then added to the vials. A quench curve is prepared by calculating fractional efficiency for each standard (Equation 13) and plotting fractional efficiency against SCR.

$$\text{fractional efficiency} = \frac{\text{recorded channel 1 cpm}}{\text{known dpm in the sample}} \quad (13)$$

To convert cpm to dpm for a sample, the cpm for channel 1 and the SCR are recorded. Fractional efficiency of counting is determined from the quench curve. If the background radiation is significant, then the average background cpm (determined from unlabeled samples) are subtracted from the total channel 1 cpm and the remaining cpm are divided by the fractional efficiency to obtain dpm.

The energy level that establishes an UL or LL for a channel is called a discriminator setting. The choice of discriminator that establishes an LL for channel 2 depends on the expected extent of quenching.[7] If the LL is set too high, then all of the SCR measurements will be near zero for moderate to severely quenched samples. On the other hand, too low a limit attenuates the relationship for slightly quenched samples, affecting the accuracy of the dpm determination. In Figure 9, distributions of pulse heights and a set of discriminator settings are depicted for moderately quenched and severely quenched samples. For moderately quenched samples this choice of discriminator setting is appropriate. Decreasing the LL for channel 2 would lower the sensitivity of changes in SCR to changes in counting efficiency. However, with this LL, SCR values for severely quenched samples would all be zero despite differences in the degree of quenching of individual samples. A lower discriminator setting, giving channel 2 a broader range, would be required for such samples.

In addition to the problem of adjusting discriminator settings to compensate for differences in quench range, quench correction by SCR can require long counting times for samples with low specific activity. The statistical accuracy of the SCR depends on the accuracy of channel 2

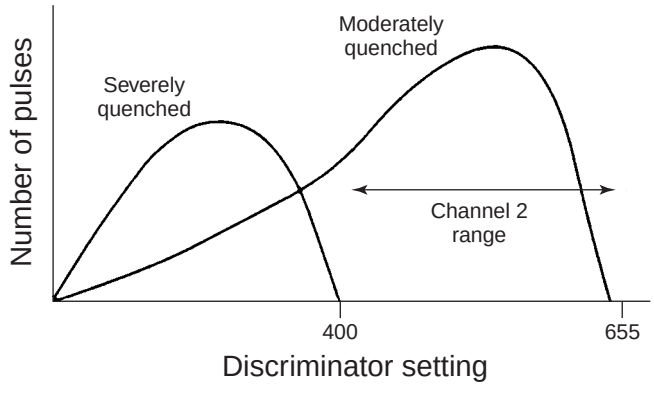

Figure 9 Choice of discriminator setting must be determined by the expected range of quenching in experimental samples.

counts, a number that is smaller and therefore less reliable than total counts.

4.3.3.3 Quench Correction Using an External Standard An external standard consists of a high-energy γ-source that can be brought in proximity to the sample vial and separated from the vial by shielding when the sample is actually counted. A fraction of the γ-photons that strike the vial and contents undergo Compton scattering, and the Compton electrons subsequently produced can excite solvent molecules in an identical manner to β-particles emitted by the internal source.

The earliest form of quench correction using an external γ-photon source involved measuring the counts in a specific pulse height range (channel).[7] It was called the external standard counts per minute (ESCPM) method. To use this method, quenched standards are prepared as for the SCR method and a series of standards is counted to obtain counting efficiency and the external standard count. Correction for any counts from the samples themselves that overlap the external standard window must be made. As for SCR, a standard curve of efficiency versus ESCPM is prepared and efficiencies of counting individual samples are determined from the curve. Also as with SCR the choice of discriminator settings depends on the level of quenching expected. Too high a discriminator setting for severely quenched samples results in values of zero for the ESCPM of most samples, while too low a setting for moderately quenched samples results in too great a loss of resolution.

The ESCPM method is seldom used at present because of serious limitations on preparation of the samples. The number of scattered electrons, and therefore the results for a single data point, depends on the position of the γ-source relative to the sample as well as the volume and electron density of the sample. Error accumulates as a result of preparation of samples containing different materials and amounts of material.

The external standard channels ratio (ESCR) method is similar to the SCR method in that it yields a ratio rather than a direct count. The effects of varying source position, sample volume, and sample density cancel out, since they all equally affect counts in both channels. The choice of discriminator settings again depends on the necessary quench range. As with the ESCPM, sample counts must be subtracted from external standard counts when the sample itself contributes to counts in either of the channels selected. As with the previous methods, dpm for samples is determined by plotting counting efficiency versus the ESCR and dividing measured cpm by the fractional efficiency corresponding to the ESCR for each sample.

A very effective method of quench monitoring that appears to be superior to either of the previous methods using an external γ-ray source is measurement of the pulse height corresponding to the Compton edge of the scattered electrons.[7] Quenching affects the pulse height distribution in response to Compton scattering as it affects the pulse height distribution for a β-particle emitter; that is, pulse heights produced when photons are detected by the PMTs are reduced by quenching. For practical purposes counting efficiency is plotted versus one-half the value of the pulse height of the Compton edge.

Beckman Instruments, Inc., starting with its LS 8000 series liquid scintillation counters, included the option to determine a number called an "H#" for quenched samples. The instruments introduce the external standard and rapidly take counts in a series of narrow channels to determine the pulse height corresponding to the inflection point of the Compton distribution for the sample. The pulse height positions of a typical set of quenched and unquenched samples are illustrated in Figure 10. The H# is the difference in measured pulse heights between the inflection points of the quenched sample and an unquenched sample. In Figure 10 the H# would be $780 - 513 = 267$. To set up a standard curve, H# values are determined for quenched standards and fractional efficiency is plotted versus H#.

4.3.4 Example of Single Label Counting

4.3.4.1 Example of Quench Correction for Carbon-14 Using the Sample Channels Ratio Method All data were obtained using a Beckman LS 3801 liquid scintillation counter (Beckman Instruments, Inc., Fullerton, CA, USA). Progressively quenched standards, each containing 51 000 dpm carbon-14 activity, were counted in two channels. Discriminator settings for channel 1 were: LL, 0; UL, 655. This range includes pulse heights corresponding to β-particle energies from zero to E_{max} for carbon-14 (UL of 655 includes signals from β-particles with energy

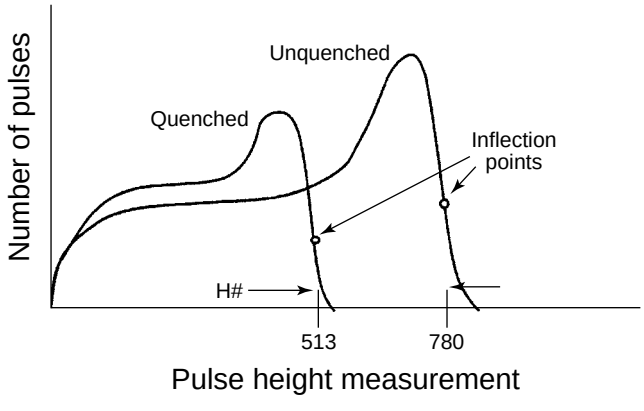

Figure 10 Shift in pulse height of the inflection point of the Compton edge for quenched and unquenched samples.

up to 156 keV). An LL of 400 was selected for channel 2, with a UL of 655, and both channel 1 counts and SCR were determined. Similar data were collected with the channel 2 LLs set at 300 and 500 respectively (Figure 11). The scale is arbitrary. To accomplish quench correction it is not necessary to know the energy level that corresponds to a particular LL. Fractional efficiencies were calculated from channel 1 cpm using Equation (14).

$$\text{dpm} = \frac{\text{channel 1 cpm}}{\text{fractional efficiency}} \tag{14}$$

LL values of 300 and 500 were clearly unacceptable. The former reduced the accuracy of the curve by attenuating the possible range of SCR values. With the latter setting, the efficiency of counting moderate to severe samples could not be distinguished. With an LL of 400 there is still some loss of accuracy at the low end of the curve, so LL = 375 was finally selected (Figure 12).

An experimental sample yielded an SCR of 0.34 and channel 1 cpm of 17 250. From Figure 12, the fractional efficiency corresponding to an SCR of 0.34 is 0.82. Activity in the sample was determined by dividing 17 250 by 0.82 to give 21 036 dpm.

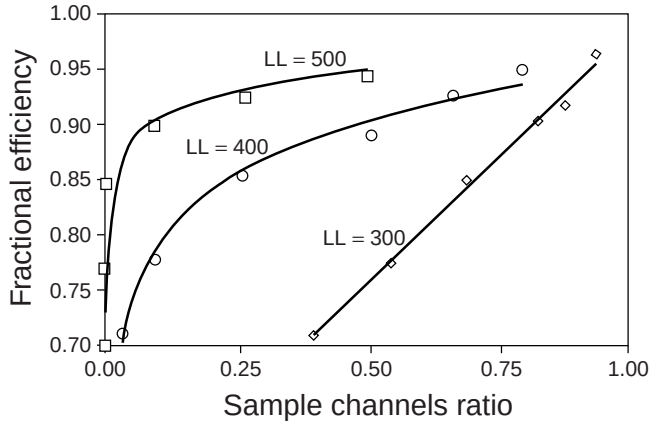

Figure 11 Three possible quench curves with different LLs for channel 2.

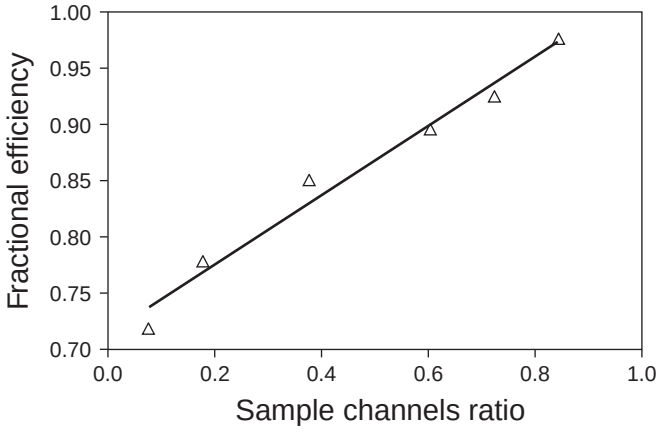

Figure 12 Quench curve for carbon-14 samples with LL = 375.

4.3.5 Dual Label Counting

If there is considerable disparity in E_{max} between two nuclides A and B, then reliable determinations of activity can be obtained for both emitters, even when they are included in the same sample. For measurement of activity in such dual-labeled samples, the ULs and LLs for channels 1 and 2 should be set as illustrated in Figure 13. Note that the discriminator settings are chosen so that the channels do not overlap, which was not the case for single-labeled samples.

Channel 2 returns only those counts due to the presence of the higher energy emitter, nuclide B. Channel 1 includes counts from nuclides A and B. The cpm due to nuclide B can be determined by carrying out the following two steps. First, a sample containing a known amount of B is counted and the ratio (R) of channel 2 cpm to known dpm for B is determined from the count. Second, the experimental samples are counted, and for each sample the channel 2 cpm are divided by R to get cpm for nuclide B.

The cpm recorded in channel 1 include all cpm due to source A as well as ''spillover'' cpm due to source B. The spillover cpm must be calculated and subtracted from the channel 1 cpm to determine what proportion of the count rate is due to source A. Spillover is determined by subtracting channel 2 cpm from the total cpm value for source B that was calculated previously. Subtracting spillover cpm from channel 1 cpm gives the cpm due to source A. If quenching is not significant, this method suffices as an estimate of dpm for both emitters. With significant quenching, an external standard can be used to estimate dpm (see section 4.3.5.1).

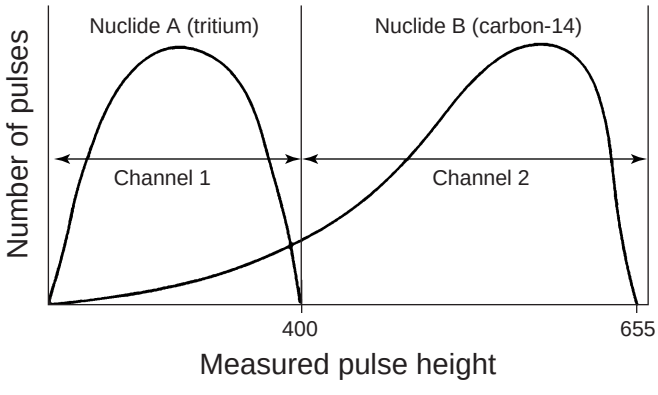

Figure 13 Selection of channels for dual-label counting of tritium and carbon-14. Note that the horizontal scale is not linear.

All of the manipulations described here are subject to error propagation, and this should be taken into account (section 3.3). The proportion of the lower energy nuclide should be high so that spillover counts are a fairly small fraction of channel 1 counts (see section 4.3.5.1).

4.3.5.1 Example of Counting Dual-labeled Samples Containing Tritium and Carbon-14 Figures 14 and 15 are quench curves that were developed using the H# method for single-labeled quenched standards containing carbon-14 and tritium, respectively. Channels were set up as illustrated in Figure 13. Fractional efficiency for each nuclide was calculated by dividing cpm by known dpm in the sample. This applies to carbon-14 as well as tritium, even though "spillover" counts from carbon-14 into channel 1 were not counted.

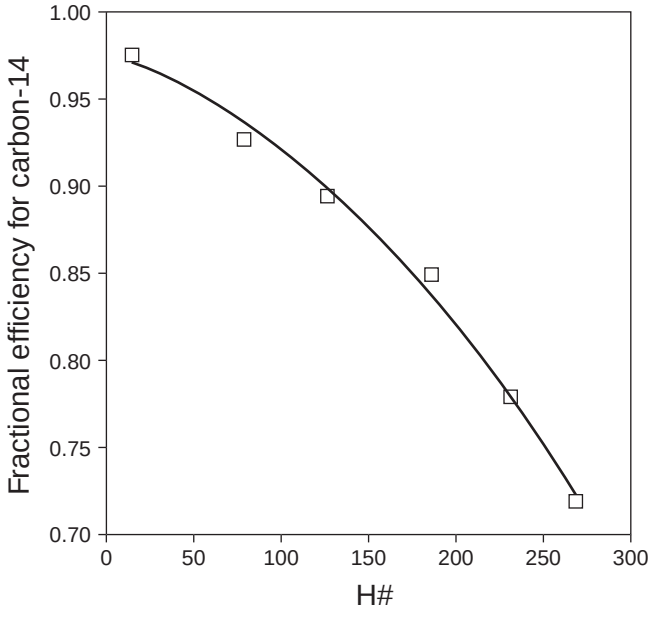

Figure 14 Quench curve for carbon-14 standards using H#.

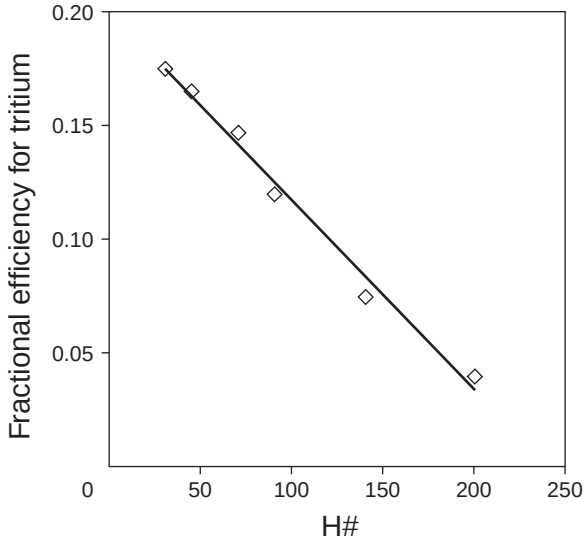

Figure 15 Quench curve for tritium standards using H#. Note that the efficiency of counting for tritium is considerably less than that for carbon-14.

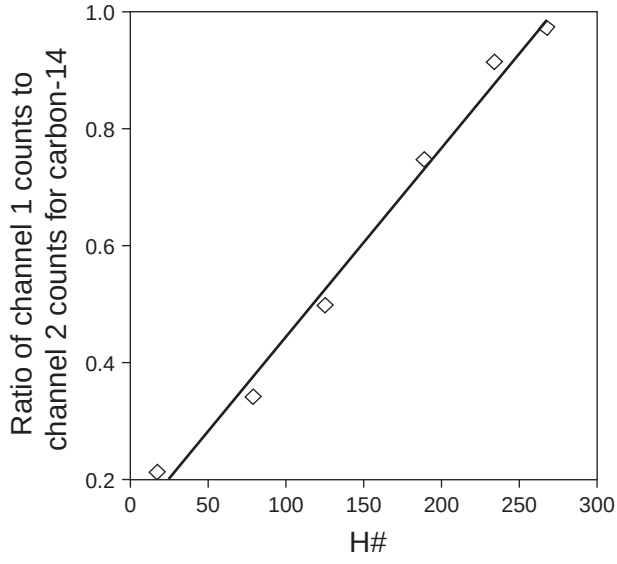

Figure 16 Plot of channel 1 to channel 2 cpm ratio versus H# for carbon-14.

A third plot (Figure 16) is also necessary, namely a plot of the ratio of channel 1 to channel 2 counts versus H# for the quenched carbon-14 standards. This plot is used to estimate "spillover" counts of carbon-14 into the tritium channel.

To obtain dpm for carbon-14 in a mixed sample, H# is used with the quench curve of Figure 14 to obtain the fractional efficiency for the sample; dpm are then estimated using Equation (15).

$$\text{dpm}(^{14}\text{C}) = \frac{\text{channel 2 cpm}}{\text{fractional efficiency}} \qquad (15)$$

Estimation of the contribution of tritium requires the use of the curves in Figures 15 and 16. To determine spillover cpm (not dpm) of carbon-14 into the tritium channel, the ratio (R) of channel 1 to channel 2 counts is determined from the H# of the unknown using the curve in Figure 15. Spillover cpm are then given by Equation (16).

$$\text{spillover cpm} = (\text{channel 2 cpm})(R) \qquad (16)$$

Spillover cpm are subtracted from channel 1 cpm to get cpm due solely to the tritium label, and then the quench curve in Figure 15 is used to estimate dpm for tritium.

A dual-labeled sample gave channel 1 counts of 44 583 and channel 2 counts of 23 444, with H# = 150. From Figures 14 and 15 the fractional efficiencies are 0.88

for carbon-14 and 0.07 for tritium. Figure 16 gives a value for R of 0.60. From Equation (15) the total dpm contributed by carbon-14 are $23\,444/0.88 = 26\,640$. From Equation (16) the spillover cpm $= (23\,444)(0.60) = 14\,066$. Subtracting spillover counts from channel 1 cpm yields $44\,583 - 14\,066 = 30\,517$ cpm for tritium. Correcting for quenching, tritium dpm are given by $30\,517/0.07 = 435\,957$, which should be reported as 436 000 dpm with rounding to significant digits.

Note that in this example it was not necessary to subtract background radiation, which averaged 20 cpm per sample. For high count rate samples, background radiation can be ignored. Background radiation must be subtracted for samples with low activity and, to the extent that background radiation adds experimental error, confidence in the accuracy of the result is reduced.

4.4 Organic Crystal Detectors

For detection of high-energy β-particle emitters, organic crystals provide effective detectors. The most commonly used materials are anthracene ($C_{14}H_{10}$), *trans*-stilbene ($C_{14}H_{12}$), and quaterphenyl ($C_{24}H_{18}$).[4] The source is brought into a cavity in the crystal, and β-radiation from the source causes excitation and de-excitation of the π electrons in the benzene ring structures. Excited electrons lose some excess energy through thermal interactions until they reach a minimal energy in the excited state, from which they de-excite with emission of a photon. The absorption spectra and emission spectra do not match, therefore the crystal is transparent to its scintillations. As with LSC, data are collected with a PMT, and energy discrimination is possible.

5 GAS-FILLED DETECTORS

A gas-filled detector contains a gas mixture that is situated between two electrodes, namely a positively charged anode and a negatively charged cathode. The filling material is usually a noble gas such as argon or helium, although other gases have been used as well.[4] When ionizing radiation passes through the gas it produces pairs of ions, consisting of positive ions that are attracted to the cathode and electrons that are attracted to the anode. Ions are collected at the electrodes, producing an electric current. Radiation can be detected and recorded as discrete events, charge can be accumulated to measure a total amount of radiation, or a constant current that is proportional to the rate of atomic disintegration can be measured.

Three general types of gas-filled detectors, namely ionization chambers, proportional counters, and Geiger–Müller counters, are available. All three will be described, although only proportional counters are well-suited to β-particle detection.

5.1 Ionization Chambers

When β-particles pass through a gas, leaving trails of ion pairs, the numbers of positive ions and ejected electrons produced are proportional to the initial kinetic energy of the particle. When an electric field is applied to the gas, the ions are attracted to the electrodes and an electric current is produced. A simple ionization chamber might consist, for example, of a cylindrical volume of gas with the outer wall serving as the cathode and an inner wire running the length of the cylinder axis to serve as the anode. Each ionizing event produces a measurable discrete current pulse of specific height.[2]

If the voltage drop between electrodes is small, the ion pairs recombine before electrons can reach the anode, and most of the signal is lost. In a voltage range of approximately 40–200 V cm^{-1} nearly all electrons reach the anode, and the instrument reaches saturation. This voltage range allows effective detection of α-radiation. α-Particles retain much of their initial kinetic energy following an interaction with a gas molecule, thus many ion pairs are produced per event, and the current that is produced is relatively large. It is nearly impossible to detect β-particles in this range, however. The mass of an electron is so small that β-particles quickly lose kinetic energy in a gas and the current pulses produced by β-particle emissions are nearly undetectable.

5.2 Proportional Counters

A proportional counter operates on a principle similar to that of an ionization chamber, but the electric field is stronger. As electrons are accelerated toward the anode they collide with other molecules with sufficient energy to produce secondary fast electrons, amplifying the signal by a factor of 10^3–10^5 depending on voltage and field strength. This proportional range is approximately 200–800 V cm^{-1}. Proportional counters yield information that can be analyzed for pulse height distribution as well as total counts. Dead times for proportional counters are of the order of a few microseconds. Thus when count rates are less than 10 kBq or so (1 kBq $= 10^3$ dps), the probability of an event going undetected is negligible. Dead time is a concern when count rates approach 100 kBq.

Solid or gaseous samples can be introduced into the chamber or gas can be passed through the chamber, in which case the instrument is called a flow counter. Because self-absorption is a problem with solid samples, solids should be introduced as thin films. This can be accomplished by electroplating or vapor coating, or if the material is soluble by applying a film of solution

to a surface and allowing the solvent to evaporate. Windowless proportional counters are quite acceptable for counting low-energy β-particle emissions, especially carbon-14 and tritium. The dimensions of proportional counters limit the degree to which energy from a β-particle is dissipated within the sensitive volume.[3] The number of ion pairs produced from a high-energy particle is proportional to only a small fraction of the energy, the rest being lost through collision with the wall opposite the sample. Thus proportional counters can be used for quantitation of activity of higher energy emitters, but energy discrimination with such emitters is not possible.

5.3 Geiger–Müller Counters

When the voltage drop between electrodes in a gas-filled counter is of the order of kilovolts, a single ionization event leads to an avalanche of photon production and secondary ionizations that results in an easily measurable pulse of several volts.[2] The pulse height is independent of the initial kinetic energy of the radiation. Gas-filled devices that operate in this range are called Geiger–Müller counters, and are usually filled with argon or helium. Since the signal is the same regardless of the intensity of the radiation, Geiger–Müller counters yield only counts of activity and cannot be used for energy resolution. Electric field strengths for Geiger–Müller counters can range from 900 to 1400 $V\,cm^{-1}$, depending on the geometry of the chamber, type of fill gas, etc.

In the high electrical field strength of a Geiger–Müller counter, a large number of positive ions are formed from the filling gas as well as electrons.[3] When the positive ions reach the cathode after a typical dead time of 100–500 μs, they combine with electrons at the cathode surface. Frequently this process results in the liberation of an electron from the surface which in turn can trigger a full discharge. Such recycling can result in a continuous output of pulses from one decay event. The problem is solved by adding a second gas, a quench gas, to the filling gas. Quench gases include ethyl alcohol and ethyl formate, which are ultimately consumable and limit the useful lifetime of the chamber. Alternative quench gases include halogen gases such as chlorine or bromine. A suitable quench gas has a lower ionization energy than the filling gas, so that the filling gas tends to transfer positive charge to the quench gas as the charge drifts toward the cathode. Thus, the positive ions that eventually reach the cathode are derived from the quench gas. When positive ions from the quench gas recombine with electrons at the cathode, the energy of recombination is dissipated by dissociation or excitation of quench gas molecules, not filling gas molecules.[1] The proportion of quench gas to filling gas molecules is small, so the event ''dies out'' rather than producing a full discharge. The free energies of dissociation and recombination of preferred quench gases are relatively small, and such gases strongly absorb photons that are liberated in the process, further preventing secondary discharges.

Geiger–Müller counters are fairly simple devices and are suitable for high-energy β-particle detection when only a count rate is needed, and for low-energy emissions (except for tritium) if a suitable window is selected. Because such counters are usually equipped with a window that absorbs much of the energy from fast electrons, they are of limited value for low-energy β-particle determination.

6 SOLID-STATE DETECTORS

The density of a solid is approximately 1000 times greater than that of a gas, therefore many more information carriers can be packed into a relatively small volume. This gives the advantage of relative compactness. A single ionization can produce many more carriers in semiconductor material than in other media, thus detectors utilizing semiconductors are capable of greater energy resolution than any other type of detector.[3] A number of materials are used, but for β-particle determination silicon has been the material of choice.

The basic operating principles of solid-state detectors are similar to those of gas-filled detectors, but the gas is replaced by a solid consisting of a semiconductor material. Typical semiconductor material has a resistivity intermediate between highly conductive metals and insulating material.[5] Most of the electrons in silicon, for example, are valence electrons–that is, they are associated with specific atoms. When the material is sandwiched between two electrodes and a voltage is applied across the material, very little current can be produced. However, when an electron in semiconductor material is given a discrete energy by ionizing radiation, it is moved to an energy level that leaves it independent of any individual atom and leaves a vacancy called a hole. Thus ionizing radiation produces electron–hole pairs which both move toward an anode, thus conducting electricity and producing a pulse. Ionizing radiation of varying energies produces a continuum of pulses, permitting spectral analysis.

The addition of impurities to semiconductor material can result in a lowering of the energy required to produce an electron–hole pair. For example, phosphorus has five valence electrons and indium has three. Incorporation of a small amount of phosphorus into a silicon chip gives the material an excess of negative charge and considerably lowers the energy needed to raise a valence electron to the conduction band. Such material is called n-type, for negative. Indium contributes an excess of holes, which can be treated as an excess of positive charge, producing

p-type material. When p-type material is coated with n-type material, electrons move from the n-type material to the p-type material and combine with holes to produce a region in which there are no charge carriers. This surface barrier is of very high resistance, permitting a high voltage to be placed across the device. The reverse bias resulting from placement of a positive electrode on the n-type side and a negative electrode on the p-type side increases the zone of depletion of charge carriers. When the n-type surface is exposed to ionizing radiation, current pulses are produced as electrons are excited to the conduction band.[7]

Silicon barrier detectors are especially useful for charged particle (α or β) spectroscopy.

7 VISUAL IMAGING

Visual imaging methods are used in tracer experiments for detection of labeled structures or molecules. While such methods are not used for direct quantitation of β-activity, they deserve mention here because of their importance in research, especially in the life sciences. In addition, films are used in dosimetry, most commonly in the form of a film badge. A film badge is sensitive to radiation that can penetrate air, such as γ- and high-energy β-particle emissions. Incident radiation forms silver ions in the grains of a film or emulsion, recording the position of impact just as such photographic materials record the impact of photons from visible light.

Common applications in research include autoradiography, in which a film or emulsion is exposed to material containing labeled substances. The locations of labeled structures or molecules can be determined by superimposing the photographic image over the image of the material itself. Examples are thin sections of labeled tissues, or electrophoretic gels containing labeled macromolecules. β-Particle emitters are also used to label antibodies and other molecules that bind specific sites in such materials, allowing identification of structures and macromolecules by visual imaging. A large body of literature under the subject of autoradiography is available in this area.

ABBREVIATIONS AND ACRONYMS

cpm	Counts Per Minute
dpm	Disintegrations Per Minute
dps	Disintegrations Per Second
ESCPM	External Standard Counts Per Minute
ESCR	External Standard Channels Ratio
LL	Lower Limit
LSC	Liquid Scintillation Counting
PMT	Photomultiplier Tube
SCR	Sample Channels Ratio
UL	Upper Limit
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Immunoassay Techniques in Environmental Analyses

Environment: Water and Waste cont'd **(Volume 4)**
Proton-induced X-ray Emission in Environmental Analysis • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
Immunoassays in Forensic Toxicology • X-ray Fluorescence in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
Radiation Hybrid Mapping

Polymers and Rubbers **(Volume 9)**
Positron Annihilation Spectroscopy of Polymers and Rubbers

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces

Mass Spectrometry **(Volume 13)**
Discrete Energy Electron Capture Negative Ion Mass Spectrometry • Electron Ionization Mass Spectrometry

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Instrumental Neutron Activation Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis • Radiochemical Separation Schemes for Multielement Determination • Radiotracer Methods • Rutherford Backscattering Spectroscopy • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Actinides and other Alpha-emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination

• Mass Spectrometry of Long-lived Radionuclides • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation • Speciation of Radionuclides in the Environment

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

REFERENCES

1. W.H. Ehmann, D.E. Vance, *Radiochemistry and Nuclear Methods of Analysis*, Wiley, New York, 1991.
2. K.H. Lieser, *Nuclear and Radiochemistry: Fundamentals and Applications*, Wiley, New York, 1997.
3. G.F. Knoll, *Radiation Detection and Measurement*, Wiley, New York, 1979.
4. W.H. Tait, *Radiation Detection*, Butterworths, Boston, 1980.
5. G.R. Choppin, J.-O. Liljenzin, J. Rydberg, *Radiochemistry and Nuclear Chemistry*, 2nd edition, Butterworth-Heinemann Ltd, London, 1995.
6. C.M. Lederer, J.M. Hollander, I. Perlman, *Table of Isotopes*, 6th edition, Wiley, New York, 1967.
7. D.L. Horrocks, *Applications of Liquid Scintillation Counting*, Academic Press, New York, 1974.

γ-Spectrometry, High-resolution, for Radionuclide Determination

Harri I.K. Toivonen[1] and Mika T. Nikkinen
Radiation and Nuclear Safety Authority, Helsinki, Finland

1 Introduction **1**
2 Semiconductor Detectors **2**
2.1 Photon Interaction with Matter 3
2.2 Production of Electrons and Holes 4
2.3 Properties of Semiconductors 5
2.4 p–n Junction 6
2.5 Charge Collection 8
2.6 High-purity Germanium Detector 9
3 Energy Resolution of the Spectrometer **11**
3.1 Full Width at Half-maximum 11
3.2 Uncertainty of Charge Production 12
3.3 Uncertainty of Charge Collection 12
3.4 Electronic Noise 12
4 Germanium Detector and its Operation **13**
4.1 Counting Geometry and Detector Selection 13
4.2 Sample Preparation 15
4.3 Cooling Systems 16
4.4 Detector Shielding 17
5 Measurement Electronics **18**
5.1 Preamplifier 18
5.2 Amplifier 19
5.3 Analog-to-digital Converter 20
5.4 Spectrum Stabilization 20
5.5 Digital Signal Processing 20
6 γ-Spectrum Analysis **21**
6.1 Characteristics of the Spectrum 21
6.2 Calibrations 21
6.3 Peak Location 25
6.4 Peak Area Determination 25
6.5 Minimum Detectable Activity 27
6.6 Radionuclide Identification and Activity Calculations 28
7 In Situ γ-Spectrometry **31**
7.1 Principles of In Situ Measurement 31
7.2 Ground-level γ-Spectrometry 32
7.3 Airborne Spectrometry 33
8 Quality Assurance System **34**
8.1 Standards and Procedures 34
8.2 Technical Quality Requirements 35
8.3 Software Validation 35
8.4 Intercomparison and Performance Evaluation 37
9 Discussions **37**
Acknowledgments **38**
Abbreviations and Acronyms **39**
Related Articles **39**
Appendix 1 **39**
Radionuclide Standards 39
Natural Radionuclides 39
Man-made Radionuclides 40
Noble Gases 42
References **42**

High-resolution γ-spectrometry is a nondestructive assay method that gives low detection limits for many radioactive nuclides. γ-Spectrometers use an electric field to collect the free charge carriers, electrons and holes, in a semiconductor detector. Modern detectors are based on high-purity germanium (HPGe). The material properties of germanium containing extremely low concentrations of impurity atoms are excellent at low temperature (77 K) for creating and collecting charges in the sensitive region of the detector. The major components of the spectrometer are the detector, cooling system, shielding system, signal processing unit, and analysis software. The charge created at the detector is directly proportional to the energy deposited in the detector by the radiation. The weak charge pulse is transformed, shaped and amplified to a voltage pulse, which is then converted to digital form. A multichannel analyzer (MCA) collects and stores the spectrum, which comprises counts partitioned among different channels according to energy dissipated in the detector. The spectrum contains peaks that can be unequivocally associated with certain radionuclides. Peak areas are proportional to the activity of the nuclide. Meticulous calibrations are required to operate the instrument properly. Software plays a major role in the analysis of the spectra, and interactive human work and artificial intelligence, i.e. an expert system, are needed to handle complex cases. Stable elements can be analyzed through neutron activation and subsequent γ-spectrometric determination. γ-Spectrometry is not a sensitive method for radionuclides that have a long half-life (≫30 years).

1 INTRODUCTION

Nondestructive analyses of material structure and composition are fascinating techniques that have

[1] H. Toivonen is at the Preparatory Commission for the Comprehensive Nuclear Test-Ban-Treaty Organization, Vienna International Centre, PO Box 1250, A-1400 Vienna, Austria.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

applications in many human activities. X-ray fluorescence, for example, is a widely used method to characterize impurity concentrations within a sample matrix. The sample atoms are excited by photons from an X-ray generator or from a radioactive source. Prompt X-rays are then emitted in the rearrangement of the K and L shells. These X-rays, characteristic to each element and usually of low energy (<30 keV), are detected with a spectrometer. Similarly, neutron activation and subsequent γ-spectrometry may reveal the presence of other impurity atoms in a sample matrix. Furthermore, the environment and the samples collected from the environment may contain trace amounts of man-made radionuclides or natural radionuclides that can be detected, identified and quantified directly with a high-resolution γ-spectrometer. The energy range of interest extends up to 3 MeV (10 MeV in special applications). The analytical capability of the γ-spectrometry is often superior to other assay methods.

Ionizing radiation is detected through its various interactions with matter. The atoms and the molecules of the absorbing medium are excited or ionized. The ionization events result in the creation of free charge carriers that drift in an electric field. The de-excitation of the excited and ionized atoms may also result in the emission of luminescent photons. Luminescence is used in scintillation detectors, whereas collection of free charge carriers is the key operational principle of gas-filled detectors, such as the ionization chamber and proportional counter, and of solid-state detectors based on semiconductors.

In a semiconductor detector, a photon produces an electrical signal (pulse) of size proportional to the energy deposited in the detector, over a wide energy range (0.1 keV to 10 MeV). In a γ-spectrum, these pulses are detected as narrow peaks at the characteristic energy of the emitting nuclide.

γ-Spectrometry is a discipline or domain in which many dimensions have to be taken into account. The true interpretation of a multipeak spectrum originating from multiline γ-emitters requires understanding of the phenomena involved in the detection process. The complexity of the measurement and the analysis is very often underestimated in practice and, consequently, invalid procedures are implemented that distort the identification results and the activity or concentration calculations.

The analyst must understand the basic physics behind the γ-spectrum and be able to select and set up the instruments for best performance in a cost-efficient manner. Of particular importance is the energy resolution of the spectrometer. High resolution in optimized counting geometry paves the way for low detection limits, is an indication of good quality of the measurement, and is an essential prerequisite for nuclide identification and quantification in complex spectra.

2 SEMICONDUCTOR DETECTORS

The sharply defined atomic electron energy levels are broadened into bands of energy states in a crystal. The outer electrons of the atoms are bound in an energy band, known as the valence band. The higher energy state, the conduction band, is separated from the valence band by an energy gap. An electron can be transferred from the valence band to the conduction band under the influence of external or internal (thermal) energy greater than the energy gap. Within the conduction band electrons can move freely. Silicon and germanium have most favorable material properties for radiation detection (Table 1).

Semiconductor research became intensive in the 1950s. It was soon realized that detectors made of germanium always need cooling, whereas silicon detectors could also

Table 1 Properties of silicon and germanium[1,2]

	Si at 300 K	Si at 77 K	Ge at 77 K
Atomic number	14	14	32
Atomic weight	28.09	28.09	72.60
Density ($kg\,m^{-3}$)	2330		5320[a]
Dielectric constant	12	12	16
Average energy per electron–hole pair (eV)	3.62	3.76	2.96, 2.98
Band gap (eV)	1.12	1.16	0.74
Intrinsic carrier density (m^{-3})	1.5×10^{16}		2.4×10^{19}[a]
Intrinsic resistivity ($\Omega\,m$)	2300		0.47[a]
Electron mobility ($m^2\,V^{-1}\,s^{-1}$)	0.135	2.1	3.6
Hole mobility ($m^2\,V^{-1}\,s^{-1}$)	0.048	1.1	4.2
Fano factor		0.084–0.16	0.057–0.13

[a] $T = 300$ K.

be operated in room temperature (owing to different band gaps). The early detectors were intended for charged particle measurements because the achievable sensitive thickness was only a few hundred micrometers, which was sufficient to absorb energy from particles but not from the more penetrating photons.

In the early 1960s silicon detectors were manufactured using lithium diffusion and drift techniques for the compensation of material impurities. Silicon detectors several millimeters thick could be produced, and were successfully applied to X-ray spectrometry. However, the small photoelectric absorption cross-section because of the low atomic number makes silicon ineffective for γ-spectrometry at high energies (Figure 1).

The impurity compensation techniques using lithium atoms was utilized in 1962 by Freck and Wakefield[4] to produce a detector, known as the lithium-drifted germanium (Ge(Li)) detector. The first detectors were designed for planar geometry but soon the coaxial shape

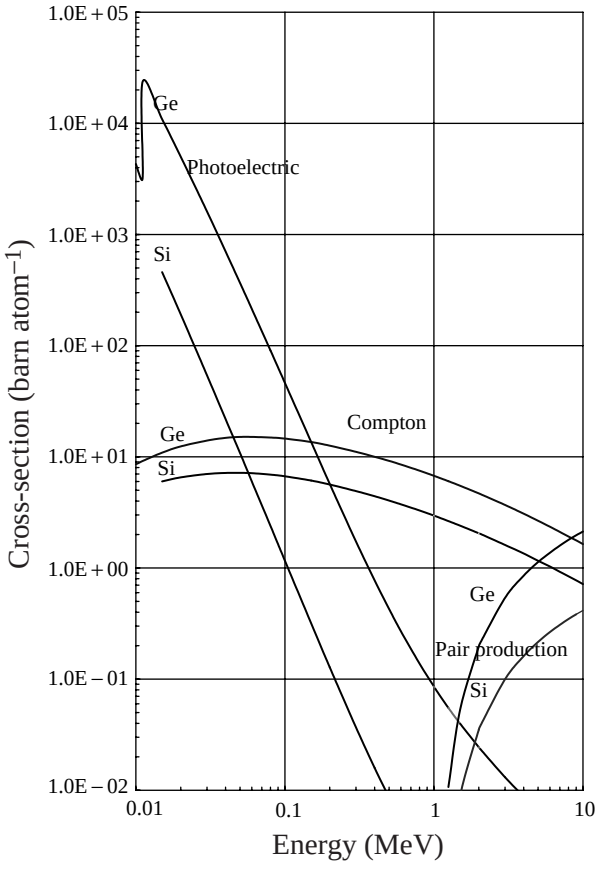

Figure 1 Cross-sections of photoelectric absorption, Compton scattering and pair production for Ge and Si[3] (1 barn = 10^{-28} m^{-2}). Note the large difference in the photoelectric absorption cross-sections between Ge and Si. Germanium has a K absorption edge at 11 keV; escape of characteristic X-rays may occur after photoelectric absorption, particularly at low energies in small detectors.

was adopted to construct large-volume Ge(Li) detectors for γ-spectrometry. These detectors provided very high-energy resolution, and revolutionized the analysis of γ-emitting radionuclides. Sample preparation became simple because radiochemical purification was no longer needed before sample counting. One drawback was that the Ge(Li) detectors had to be kept at low temperatures all the time in order to prevent the high-mobility Li atoms drifting away and damaging the detector. Ge(Li) detectors introduced the age of high-resolution γ-spectrometry, but their production ceased after the development of ultra-HPGe.

In the early 1970s germanium was purified to a level that made it possible to construct detectors without lithium compensation. Nowadays germanium crystals with extremely low impurity concentrations (10^{15}–10^{16} atoms per m^3) can be produced. In 1998 detector-grade germanium was manufactured by only three companies (EG&G Instruments USA, Tennelec, USA, and Hoboken, Belgium). Unlike the Ge(Li) detector, the HPGe detector can be stored and manipulated at room temperature, although for measurement it too must be cooled using either liquid nitrogen (77 K) or an electrical system.

2.1 Photon Interaction with Matter

The photoelectric effect, Compton scattering, and pair production are the major mechanisms through which energetic photons interact with matter. By these processes the energy of the incident photons is transferred to the kinetic energy of electrons, which in turn produce charge carriers, free electrons and positively charged holes, in the detector medium. Each mode of interaction produces different effects within the resulting γ-spectrum.

In photoelectric absorption, the incident photon of energy E interacts with a bound electron. The entire energy of the photon is absorbed, and an electron (K or L) is ejected with energy $E - E_B$, where E_B is the binding energy of the electron. However, E_B is quickly transferred to a characteristic X-ray as a result of a subsequent transition between the energy levels of the electrons. The deposition of the entire energy of the photon in the detector produces a prominent peak in the spectrum, known as the full-energy peak, photo-peak or γ-peak. When the interaction occurs near the detector surface, the characteristic X-ray may escape from the detector. In a γ-spectrum acquired with a germanium detector, a small peak may sometimes be found 10 keV below the full-energy peak (Figure 1).

The photoelectric absorption cross-section is proportional to $Z^5/E^{3.5}$ (or to $Z^{4.5}/E^3$), where Z is the atomic number.[1,5] Therefore, heavy atoms absorb γ-radiation much more effectively than light atoms. An ideal detector

material has a high Z value. This is why silicon is not well suited for γ-spectrometry at high energies.

The incident photon may also be scattered by an electron. The process is known as Compton scattering. The initial photon energy is then shared by the electron and the scattered photon. This secondary photon may be absorbed in a subsequent photoelectric event, may be further scattered, or may leave the detector without further interaction. In the last case only the energy transferred to the electron is absorbed. Because of variability of energy transfer in scattering events, Compton scattering produces a smooth baseline continuum in the γ-spectrum. This continuum increases detection limits of radionuclides (see section 6.5). Compton scattering is the dominant interaction process in germanium between energies 150 keV and 8 MeV (Figure 1).

In pair production, the energy of the incident photon must exceed twice the rest mass energy of an electron (2×511 keV). The incident photon is completely absorbed and an electron–positron pair is created in the electric field of the nucleus of the absorbing material. Pair production may occur also in the field of an electron, but the probability is low. The electron and the positron transfer their kinetic energy to the medium. Annihilation of the positron in germanium produces two γ-rays that must also be absorbed in the detector to produce a full-energy peak. However, one or both of these photons may escape the detector.

When both photons created in the annihilation leave the detector without interaction, a *double escape peak* may be found in a γ-spectrum at energy of 1.022 MeV below the photo-peak. If only one of the two photons leaves the detector, a *single escape peak* may be produced at energy of 511 keV below the photo-peak. However, the single escape peak and the 511 keV annihilation peak are broader than the photo-peaks nearby because of the Doppler effect. At the point of annihilation, neither the electron nor the positron is at complete rest. The conservation of energy and momentum implies that the photons have slightly different energies. The degree of Doppler broadening depends on the annihilation environment, i.e. on the detector material. The resulting annihilation and single escape peaks are typically 1–2 keV broader than the photo-peaks at the same energies.

2.2 Production of Electrons and Holes

In a perfect semiconductor there are no electron energy levels within the band gap (Figure 2). Therefore, to raise a bound electron from the valence band to the conduction band a minimum energy, E_g, the energy difference between the top of the valence band and the bottom of the conduction band, is needed. Thermal excitation

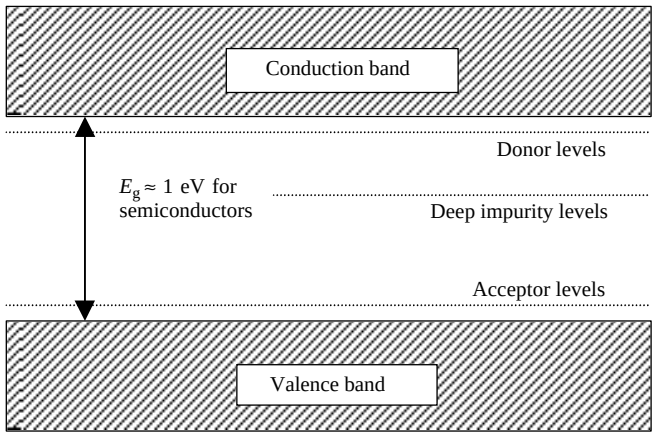

Figure 2 Energy levels in a semiconductor. Donor or acceptor impurities, such as phosphorus and boron atoms, occupy sites in a semiconductor crystal replacing germanium atoms and they have energy levels near the conduction band and the valence band respectively. Some impurities have energy levels deep in the band gap and they may act as trapping centers of charge carriers in a detector.

or radiation interacting with semiconductor material can break the bonds of the electrons. For each free electron there is a missing electron or hole in the valence band. The electrons in the conduction band and the holes in the valence band are mobile in the electric field.

The interaction of photons with semiconductor medium is a complicated cascade. Not all of the energy of the primary photon is consumed for the ionization of the medium. Some of the energy of the incident photon is transferred to the lattice as phonons, and the average energy W needed to produce an electron-hole pair is therefore larger than the band-gap energy E_g (Table 1). The non-ionizing processes dominate only at very low energies, and, consequently, W does not depend on the energy or type of the radiation.

Absorption of energy E produces n electron–hole pairs, on average, in the sensitive region of the detector as illustrated in Equation (1):

$$n = \frac{E}{W} \tag{1}$$

As a result of energy transfer from ionizing radiation to the semiconductor medium, tightly bound electrons of the atoms in the lattice may be raised to energy levels well above the bottom of the conduction band. The energies of such excited electrons and the embedded holes are transferred to other electrons and holes through a series of interactions. Finally, the holes are at the top of the valence band and the electrons at the bottom of the conduction band. Eventually the excited electrons in the conduction band return to the valence band and then the number of electron–hole pairs is again dominated by thermal

excitation. In an electric field, however, the electrons and holes migrate according to the field direction.

The number of electron–hole pairs formed in the detector medium plays a major role in detector resolution. If their amount would follow Poissonian statistics, the standard deviation would be (Equation 2)

$$\sigma_{n,\text{Poisson}} = \sqrt{n} = \sqrt{\frac{E}{W}} \tag{2}$$

The energy uncertainty, i.e. the standard deviation of the full energy absorption would then be (Equation 3)

$$\sigma_E = W\sigma_{n,\text{Poisson}} = \sqrt{WE} \tag{3}$$

In reality, however, the resolution of the detector is much better than the prediction of Equation (3). Simple Poissonian statistics do not describe the formation of electron–hole pairs. A correction factor, known as the Fano factor, F, must be introduced, as illustrated in Equations (4) and (5):

$$\sigma_n^2 = F\sigma_{n,\text{Poisson}}^2 = F\frac{E}{W} \tag{4}$$

Therefore,

$$\sigma_E = W\sigma_n = \sqrt{FWE} \tag{5}$$

If all of the energy transferred to the semiconductor medium went into creating electron–hole pairs, there would be no statistical fluctuation, and the Fano factor would be zero. On the other hand, if only a small fraction of the absorbed energy were used to produce electron–hole pairs, the fluctuations would be truly random, and the Poissonian statistics would be valid. For the semiconductors, the real situation is between these two extremes. The exact value of the Fano factor is not known, and the experimental values vary considerably (Table 1). The Fano factor seems to be somewhat smaller for germanium than for silicon; in addition, the W values are smaller for germanium (Table 1). Thus, Equation (5) predicts that the germanium detectors have slightly better energy resolution than the silicon detectors.

In a scintillation counter, such as the sodium iodide detector, the energy of the primary photon is transferred to light photons, and these photons in turn produce photoelectrons at the cathode of a photomultiplier. However, the efficiency of the photoelectron production is low (about 10%) in this random multistep process and, consequently, the Poissonian statistics are valid, i.e. $F = 1$. Furthermore, production of a photoelectron at the photocathode requires energy of about 170 eV. Thus, the peak widths of the spectra acquired with a NaI detector are very large compared with the high resolution of the germanium detector. The NaI peak widths are 20–30 times greater than HPGe peak widths, as calculated from Equation (5) using Fano factors of 0.06–0.12 for germanium. In practice, depending on the type of detectors and quality of the electronics, the resolution difference may be as large as a factor of 50.

The overall resolution of the detector is determined by charge production, uncertainty of charge collection, and noise of electronic devices (section 3).[6,7]

2.3 Properties of Semiconductors

Thermally generated free electrons and holes give rise to electrical conduction in the crystal. This *intrinsic* semiconduction is characterized by equal concentration of electrons and holes. In Equation (6) the concentration n_i of electrons and holes depends on temperature as[8]

$$n_i = CT^{3/2}\mathrm{e}^{-E_\mathrm{g}/2kT} \tag{6}$$

where C is a factor independent of temperature T and k is the Bolzmann constant ($1.3805 \times 10^{-23}\,\mathrm{J\,K^{-1}}$). Values of n_i can be derived from conductivity measurements. At room temperature, n_i is $1.5 \times 10^{16}\,\mathrm{m^{-3}}$ for silicon and $2.4 \times 10^{19}\,\mathrm{m^{-3}}$ for germanium.[8]

The intrinsic carrier concentration determines the lower limit of the leakage current in a detector. To minimize leakage, i.e. to reduce the number of electrons in the conduction band, the detector has to be cooled to a low temperature. Cooling is necessary for all germanium detectors and for high-resolution silicon X-ray detectors.

2.3.1 Impurity Doping

Adding impurity atoms of the fifth or third group of the periodic table increases electrical conductivity of an intrinsic semiconductor. Phosphorus or arsenic atoms, for example, need only four of their five valence electrons to complete covalent bonds to Ge or Si atoms. The extra electron is very loosely bound, has an energy level near the conduction band, and is easily moved to the conduction band, thus increasing the conductivity of the material. Impurity atoms that donate electrons to the conduction band are known as donors. Germanium with such impurity is said to be of n-type (negative donor impurities). Similarly, some atoms, such as boron, accept electrons from the valence band, leaving behind a hole that can move through the crystal. These impurity atoms are known as acceptors. Germanium with such impurity is said to be of p-type (positive acceptor impurities).

A real crystal contains both donor and acceptor impurities. These impurity atoms tend to neutralize each other, i.e. the electrons of the donors fall into the acceptor levels and no free charge carriers are formed. The properties of the semiconductor depend on the type of impurity that is present in excess. Let N_D

and N_A be the donor and acceptor concentrations. The net concentration, $N_D - N_A$ (n-type) or $N_A - N_D$ (p-type), and the dominating impurity element determine the quality of the crystal. The material is said to be *compensated*, if $N_D = N_A$. In practice a small imbalance exists between donor and acceptor concentrations.

In an n-type crystal the electron concentration, n_e, is increased in the conduction band. Because of this, recombination disturbs the equilibrium amounts of electrons and holes. As a result, the hole concentration, n_h, is lower than the charge carrier concentration, n_i, of the intrinsic material, see Equation (7), in such a way that

$$n_e n_h = n_i^2 \tag{7}$$

If a region is heavily doped with donors, the number of holes is reduced to a negligible level. Thus, in an n-type material the electrical conductivity is determined by electrons. Similarly, the motion of holes dominates conductivity in a p-type material.

The heavily doped layers are called as n+ or p+ layers. The concentration of the doping atoms can be as high as $10^{24}\,m^{-3}$. These layers provide blocking contacts that are necessary for the functioning of a detector (see section 2.4). Evaporation and diffusion or ion bombardment are utilized for the manufacture of the blocking contacts in commercial detectors.

2.3.2 Charge Traps

Charge trapping and recombination are the result of impurities or defects in the crystal. Impurities, such as zinc and cadmium, within the lattice have energy levels deep in the band gap. These impurities may trap charge carriers temporarily, thus preventing them from contributing to the electric signal of the preamplifier. The trapping efficiency is proportional to the Boltzmann factor, $\exp(-E/kT)$, where T is the operating temperature and E the binding energy of the trap. At low temperatures the trapping is more enhanced than at room temperature. Some impurities can trap both electrons and holes, which will then annihilate. This process is often more probable than direct recombination through the band gap. Structural defects of the lattice can also trap charge carriers.

The charge-carrier life time, τ, is a measure of the quality of the crystal. In a good crystal τ is of the order of 0.1–1 ms. Detector manufacturing processes may shorten carrier life time but it is still sufficient for the proper operation of the detector because the charge collection time is below 1 μs (see section 2.5).

2.3.3 Motion of Charge Carriers

In an electric field, electrons and holes move in opposite directions in the crystal, with drift velocities directly proportional to the electric field strength. The constant of proportionality is called "mobility". The mobility of the electrons, μ_e, is of the same order of magnitude as the mobility of the holes, μ_h. Both types of charge carrier have therefore a significant influence on the signal of the detector. The lower the temperature the higher the mobility ($T^{-3/2}$).[1,5,8] Therefore, cooling a germanium detector not only reduces leakage current but also improves the efficiency of the charge collection.

Resistivity, ρ_p, of a p-type semiconductor depends on mobility of the holes and on concentration, N_A, of the acceptor material[1,5] as illustrated in Equation (8):

$$\rho_p = \frac{1}{\mu_h e N_A} \tag{8}$$

Similarly, the mobility of electrons and the donor concentration, N_D, dominate the resistivity in an n-type crystal.

Since the quality of a crystal depends on mobility and carrier life time, a quantity $\mu\tau$ is sometimes used as a quality index. Assuming a life time of 20 μs for the charge carriers in a germanium detector, the product $\mu\tau$ would be about $7.2 \times 10^{-5}\,m^2\,V^{-1}$ for the electrons and $8.4 \times 10^{-5}\,m^2\,V^{-1}$ for the holes (Table 1). In other semiconductor materials (section 9), except silicon, the product $\mu\tau$ is several orders of magnitude lower than for germanium.

2.4 p–n Junction

To detect the small transient current produced by the energy absorbed in the detector, the background or the leakage current must be negligible (hence conductors are not suitable for radiation detection). Leakage is small in insulators but they have other problems, such as large W, small μ, and charge losses through trapping. These problems are avoided in the semiconductors. The leakage current is extremely low at low temperatures (10^{-9}–10^{-12} A at 77 K).[1,9] However, a bias voltage applied to a pure semiconductor material would produce a leakage current several orders of magnitude larger than the peak current (a few microamperes) caused by the absorption of photons and subsequent motion of electrons and holes. Consider, for example, a high-purity silicon wafer with resistivity of 50 000 Ωcm (see intrinsic material in Table 1), thickness of 0.1 cm and surface area of 1 cm^2. A small potential difference of 50 V across the wafer would produce a current of 10 mA.[7] Therefore, silicon and germanium alone cannot be used as a detector without special arrangements.

An essential requirement for a detector is that the positive and the negative contacts must not inject holes and electrons into the sensitive area of the detector. Blocking contacts are needed to prevent free exchange

of electrons and holes between the electrodes and the sensitive volume of the detector. A p–n junction beneath the surface of the detector provides this blocking. A thin layer of heavily doped n-type germanium formed at a surface of p-type crystal prevents holes moving into the detector bulk material, even when a voltage is applied across the junction. Positive voltage must be connected to the n-side of the junction, otherwise the system would be a good conductor (major charge carriers attracted by voltage of right polarity). This arrangement of connecting positive voltage to the negative type semiconductor, known as *reverse-biased junction*, reduces leakage current drastically but allows collecting the free charges produced in the detector material by the energy absorbed from the incident photon. In detector manufacture, blocking contacts are formed on both electrodes; n+ material is used for the positive contact and p+ material for the negative contact.

A diode is a simple and familiar application of the p–n junction. The junction conducts current when the voltage is connected in the forward direction, but it conducts very little when biased in the reverse direction. This reverse bias voltage drives the electrons away from the junction to the n-side and the holes to the p-side (Figure 3).

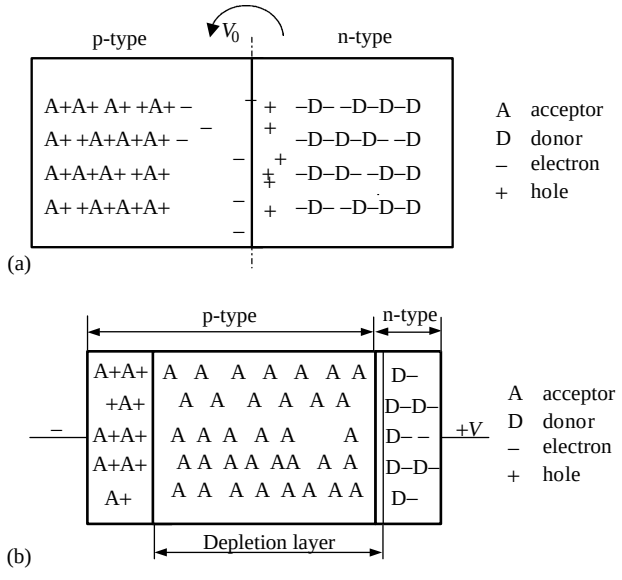

Figure 3 (a) p–n junction without external bias voltage in germanium lattice containing acceptor and donor impurities. A few electrons and holes diffuse across the junction forming a charged dipole layer and consequently, an electric potential difference (0.4 V for Ge and 0.7 V for Si). This potential difference prevents further diffusion of charge carriers across the junction. (b) Reverse bias voltage applied to p–n+ junction where the n-type material is heavily doped. A region is formed that is completely depleted of the charge carriers. Because $N_D \gg N_A$, the depletion layer extends mainly to the p-side, i.e. $d_A \gg d_D$ (see text).

Thus a large depletion region may be formed. When the donor concentration on the n-side of the junction is much higher than the acceptor concentration on the p-side, the diffusing electrons travel a large distance into the p-type material before recombination with holes; the diffusion of the holes into the n-side is far less efficient. Because the net charge must be zero, as illustrated in Equation (9):

$$N_A d_A = N_D d_D \tag{9}$$

where d_A is the depletion thickness inside the acceptor layer and d_D the depletion thickness inside the donor layer. The depleted volume is the sensitive region of the detector. In that region the concentration of free charge carriers is suppressed by several orders of magnitude. The only remaining charges, the ionized donor sites and the filled acceptor sites, are immobile and do not contribute to the conductivity. The depletion region has a very high resistivity compared with the materials on either side of the p–n junction. Thermal excitation produces some electrons and holes but the electric field sweeps them away immediately.

In a detector, the ionizing radiation produces electrons and holes in the depletion region. At the same time, the leakage current of the reverse biased junction is small, allowing the collection and measurement of the small transient current in an external circuit. The total free charge, i.e. the integral of the current, is proportional to the amount of electron–hole pairs formed, and, consequently, to the energy absorbed in the detector.

A detector is only partially depleted when the depletion region does not extend to the edges of the crystal. Most detectors, however, operate with very high bias voltage and, therefore, the depletion region extends through the full thickness of the semiconductor material (excluding the thin n+ or p+ layers). Such a detector is fully depleted.

The electric potential difference ϕ caused by the charge distribution q in the region of the reverse biased junction can be calculated from the Poisson equation (Equation 10)

$$\nabla^2 \phi = -\frac{q}{\varepsilon} \tag{10}$$

where ε is the dielectric constant of the medium, i.e. $K\varepsilon_0$ ($K = 16$ for Ge and the permittivity of vacuum $\varepsilon_0 = 8.85 \times 10^{-12}\,\mathrm{C\,V^{-1}\,m^{-1}}$).

Equation (10) is easy to solve in one-dimensional slab geometry.[1] Assuming a constant charge distribution of eN_D and $-eN_A$ on the n+ and p-side of the junction, respectively, the total thickness d of the depletion layer at voltage V is illustrated in Equation (11) as:

$$d \cong \sqrt{\frac{2\varepsilon V}{eN_A}} \tag{11}$$

If the p-side doping would be dominant, N_D should be substituted for N_A. Utilizing Equation (8), allows expressing d as a function of resistivity, see Equation (12):

$$d \cong \sqrt{2\varepsilon\mu_h\rho_p} \tag{12}$$

For a large depletion region, it is necessary that the semiconductor material has high resistivity. This can be achieved, as shown by Equation (11), only if the impurity concentration of the material is very low. In this respect germanium is the best material available; in a modern detector there are, on average, fewer than one impurity atom for 10^{12} Ge atoms.

The depletion region resembles a capacitor (charges separated by an insulator). The larger the depletion thickness, the smaller the capacitance C. Equation (13) shows capacitance C for a unit area

$$C = \frac{\varepsilon}{d} = \sqrt{\frac{\varepsilon e N_A}{2V}} \tag{13}$$

The capacitance of the detector should be as small as possible because this improves energy resolution (section 3.4). Equation (13) shows that the capacitance is minimized by increasing bias voltage until the detector is fully depleted. At larger voltages the capacitance remains constant; the operation of the detector is stable and, furthermore, the charge collection efficiency is improved at the edges of the detector. In a large crystal, V is typically 2–5 kV. The capacitance is smaller in coaxial geometry than in planar geometry for detectors of equal diameter and thickness (see sections 2.6 and 4.1).

2.5 Charge Collection

The rise time of the pulse signal from the semiconductor detector varies considerably depending on the location where the charge carriers are generated within the crystal. In a large detector the electrons or the holes have to travel a few centimeters. Therefore, the charge collection time at saturation velocity of $10^5\,m\,s^{-1}$ is about 100–500 ns, and, consequently, the pulse rise time is of the same order of magnitude. It is necessary that the shaping time of these pulses in the amplifier is much larger than the longest rise time expected for the detector. Otherwise the energy resolution could be substantially deteriorated as a result of reduced signal amplitudes (''ballistic deficit'').

Figure 4 shows the shape of the signal in a depleted planar detector where the electric field is large enough for the charge carriers to reach their saturation velocity. As a result of formation of electron–hole pairs, a constant electric current flows in the external circuit until all free charges have been collected. In practice the detector response is not linear because the free charges are not generated at a point but rather along the track of the ionizing photons and secondary electrons. Furthermore, the electric field may be low at the corners of a cylindrical detector and thus the drifting velocity is low. Such slow pulses may deteriorate energy resolution, causing low-energy tails on the peaks in a γ-spectrum. In addition, charge trapping can reduce pulse amplitude and possible detrapping may affect the leading edge of the pulse. Therefore, the shape and rise time of the pulse vary greatly. In most applications, however, exact timing is usually not required.

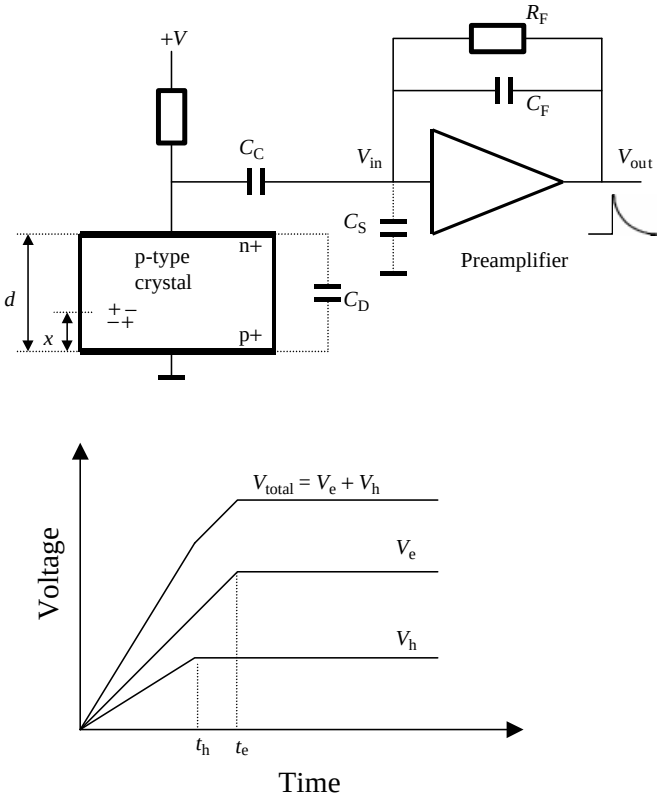

Figure 4 Pulse shape in a planar p-type detector using an AC-coupled preamplifier. The ionizing radiation produces n electron–hole pairs that start to migrate towards the electrodes. Time intervals t_h and t_e are needed to collect all holes and electrons, respectively. The charge collection times vary depending on the location where the charges are created within the detector. The transient current is integrated by the total capacitance $C = C_D + C_S$ (detector and preamplifier) producing a linear ramp voltage that approaches a maximum of $V_{total} = ne/C$. The feedback capacitor C_F establishes the charge gain for the detector ($V_{out} = ne/C_F$). In the AC-coupling one electrode can be grounded. However, the load resistance and the input stray capacitance to the ground worsen the signal-to-noise ratio. Furthermore, the coupling capacitor C_C causes extra differentiation of the signal. A DC-coupled configuration provides better energy resolution and is normally used in γ-spectrometry with germanium detectors (see Figure 11).

Table 2 Properties of a large coaxial germanium detector manufactured by EG&G ORTEC in 1997

Dimensions		Absorbing layers	
Crystal diameter	89.8 mm	Aluminum	1.0 mm
Crystal length	54.4 mm	Inactive germanium (n+)	700 μm
Core diameter	7.2 mm		
End cap to crystal	4 mm		

Performance specifications at 1.33 MeV using digital spectrum processing

	Quality assurance data sheet by the manufacturer	Verification measurements by the user
Resolution (fwhm)	1.97 keV	1.81 keV
Relative efficiency	100%	101%
Peak shape (fwtm/fwhm)[a]	1.93	1.92
Peak-to-Compton ratio	77	88

[a] Perfect Gaussian curve, 1.82.

2.6 High-purity Germanium Detector

Equation (11) allows estimation of the maximum impurity concentration in a semiconductor detector of given size. In a small detector with a depletion thickness of 1 cm and reverse bias voltage of 1000 V, the required net impurity atom concentration should be below $2 \times 10^{16}\,m^{-3}$. Nowadays an order of magnitude improvement is possible, and thus large and efficient detectors can be manufactured. Detectors with a diameter of 10 cm and a height of 5 cm, or even larger volumes, are now commercially available (see Table 2). Extreme purification has been successful only for germanium.

The manufacture of HPGe detectors, including purification of germanium and production of blocking contacts, is a complicated process and technical details are not always available in the open scientific literature. In principle, bulk germanium, produced in large quantities for electrical industry, is further purified by heating and subsequent removal of the molten zone where the impurities tend to concentrate. The process has to be repeated several times. Lithium evaporation and diffusion are used to produce an n+ layer at the outer surface of a ''standard'' coaxial p-type crystal, or, alternatively, donor atoms can be implanted using an accelerator. The p+ contact at the inner side of the coaxial p-type crystal is produced by ion implantation of boron or other acceptor atoms. Injection of electrons can also be prevented by a metal–semiconductor surface barrier.[1] The n+ contact may have a thickness or dead layer of 500–700 μm, which significantly attenuates photons with energies below 100 keV (Table 3). However, the thickness of the p+ ion-implanted contact might be only 0.3 μm. If this surface is on the outer part of the crystal, the attenuation of the photons is negligible down to a few keV.

Table 3 Transmission (%) of photons through the n+ dead layer in germanium (p-type detector) and through Al windows (data calculated with XGAM).[3] If low-energy photons need to be measured, the outer dead layer of the detector has to be small (0.3 μm). Aluminum enclosure with thickness of 1 mm suits photons above 30 keV, whereas a thin beryllium window or a polymer film has to be used for photons below 20 keV

Energy (keV)	Ge (0.7 mm)	Al (1 mm)	Al (1.5 mm)
10	0	0.1	0
20	0	42	27
30	7	76	66
40	11	87	82
50	31	92	88
60	49	96	94
80	72	97	95
100	83	97	96

Recent detector development has made extremely thin n+ coating possible as well.

2.6.1 Planar Detectors

A germanium detector in planar geometry is typically intended for the detection of low energy photons. In a p-type detector the depletion region begins at the n+ contact. At the voltage, predicted by Equation (11), the detector comes fully depleted. At higher voltages the electric field inside the detector increases in a uniform way everywhere in the detector At the electric field of $10^5\,V\,m^{-1}$ (77 K) the electrons reach their saturation velocity of $10^5\,m\,s^{-1}$.[10] For holes, a larger field would be required, but this is usually not possible because of breakdown and surface leakage.

In a modern planar detector the rear contact does not cover the full area of the wafer. This design reduces detector capacitance compared with a conventional planar system. Reduced capacitance offers lower noise in the

preamplifier and, consequently, improved energy resolution. Such a detector with a thin beryllium window is very efficient in the energy range of 3–100 keV. Large-area planar detectors with sufficient thickness (30 mm) have very high counting efficiency for environmental samples in a wide energy range (see section 4.1). They are more efficient for nuclides that emit photons with energy below 500 keV than coaxial detectors of much larger volumes.

2.6.2 Coaxial Detectors

Commercial large-volume detectors are traditionally constructed in cylindrical or coaxial format. The outer surface of the cylinder is one electrode, and the second contact is the inner cylindrical surface manufactured by removing the core of the crystal (Figure 5). Usually, only a part of the core is removed so that the outer electrode extends over the other flat end of the crystal. In these close-ended configurations the electric field lines are not radial, as they would be in true coaxial configuration. Thus the charge collection efficiency is somewhat reduced compared with the ideal geometry. There are regions where the electric field is weak and thus the charge collection is not efficient. However, the close-ended configuration provides advantages that compensate for these problems. The planar surface is not vulnerable to leakage currents and provides a good entrance window for weakly penetrating radiation. Distortion of the electric field can be minimized by extending the central hole to near the front surface. In addition, the corners, i.e. those areas where the electric field is weak, can be removed during detector manufacture (''bulletization'').

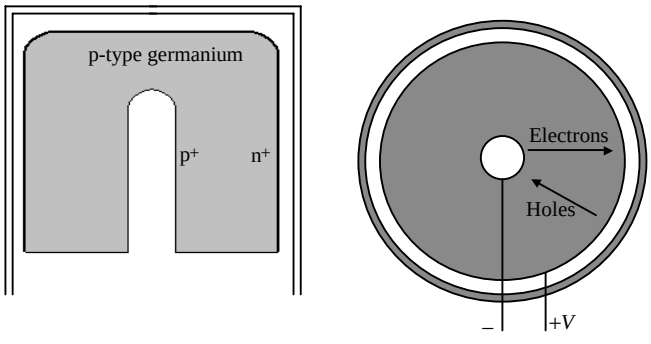

Figure 5 Coaxial p-type germanium detector. The charge carriers, formed in the pure depleted p-type material, are collected by the electrodes. However, the dead layer, i.e. the thick n+ layer, effectively absorbs low energy photons. In an n-type detector, the thin p+ layer is facing outwards. With a thin beryllium window the response of an n-type detector extends below 5 keV. In this detector the holes are collected by the outside electrode and they have, on average, a shorter distance to travel. This reduces efficiency of hole trapping and consequently, an n-type detector is not as vulnerable to radiation damage as a p-type detector of equal size.

A well detector provides near 4π counting geometry for small sources. The energy resolution is, however, usually less than that of the coaxial detectors (Table 4). On request at the purchase phase, the resolution can be improved using special material balancing in the manufacture.

The detector crystal must be kept clean. Traces of impurity at the crystal surface may create leakage currents. The detector is mounted in a housing that maintains a vacuum and provides thermal insulation and protection against surface impurities. Often the preamplifier is also located inside this housing.

To calculate the electric field and potential of the detector, the Poisson Equation (10) has to be solved in cylindrical geometry.[1] The voltage needed to fully deplete the detector is directly proportional to the acceptor concentration in a p-type crystal. The capacitance of the detector with height h (m) is illustrated in

Table 4 Properties of some HPGe detectors commercially available. Data by EG&G ORTEC[11] and Canberra[9]

	Coaxial p-type	Coaxial n-type	Short coaxial n-type	Well p-type	Planar p-type
Window	500–600 μm[a]	0.3 μm[b]	0.3 μm[b]	0.3 μm[b]	0.3 μm[b]
Energy range	40 keV[c] to 10 MeV	3 keV to 10 MeV	3 keV to 1 MeV	10 keV to 10 MeV	3 keV to 300 keV[d]
Sizes	10–150% (relative efficiency)	10–100% (relative efficiency)	36–70 mm (diameter)	<400 cm^3 (active volume)	8–80 mm (diameter)
Resolution at 5.9 keV		660–1300 eV	300–500 eV		150–480
Resolution at 122 keV	850–1500 eV		580–720 eV	1200–1400 eV	480–750
Resolution at 1.33 MeV	1.7–2.4 keV	1.8–2.7 keV		2.0–2.3 keV	
Peak/Compton	34 : 1–90 : 1	34 : 1–64 : 1			
fwhm/fwtm	1.9–2.0	1.9–2.3			
fwhm/fwfm	2.6–3.1	2.6–3.3			

[a] n+ impurity region at the entrance, Al enclosure not included.
[b] Entrance contact produced by ion implantation.
[c] 3 keV for an extended range detector with very thin outer n+ layer and Be window.
[d] A thick planar detector (30 mm) has excellent energy response below 500 keV and its useful energy range extends up to 2 MeV.

Equations (14) and (15)

$$C = \frac{2\pi\varepsilon h}{\ln(r_2/r_1)} \tag{14}$$

or

$$C(\text{pF}) = \frac{890h}{\ln(r_2/r_1)} \tag{15}$$

For good energy resolution, the detector capacitance should be as small as possible. Therefore, the ratio of core-to-crystal radii, r_2/r_1, should be minimized. For example, the capacitance of the detector with dimensions in Table 2 is expected to be 20 pF.

The shape and the dimensions of the detector have to be chosen according to the needs of the application (section 4.1). A large-volume detector is not the best solution for all applications.

3 ENERGY RESOLUTION OF THE SPECTROMETER

The energy resolution qualities of germanium detectors provide an efficient means of detecting and resolving a peak of a nuclide from baseline continuum or from a neighboring peak in the γ-spectrum (see Figure 17). The photo-peak pulses are concentrated into a few channels, and thus a peak may be found above the background noise that would otherwise mask the primary signal, as often happens in low-resolution spectrometry.

Uncertainty of charge production, incomplete charge collection and noise of the electronic equipment determine the total energy resolution of a measuring system based on a semiconductor detector. These uncertainties are independent of each other and are therefore additive. Thus, the width w of a peak can be divided into components as[1,5] illustrated in Equation (16)

$$w^2 = w_\text{p}^2 + w_\text{c}^2 + w_\text{n}^2 \tag{16}$$

where w_p, w_c and w_n refer to the uncertainty in production of charge carriers, in charge collection, and in electronic noise respectively.

3.1 Full Width at Half-maximum

In high-resolution γ-spectrometry, the resolution of the measuring systems is described by the full width at half-maximum (fwhm) of the photo-peak (Figure 6). In low-resolution spectrometry, however, it is common practice to express the resolution in percentages (fwhm normalized by the photo-peak energy). For example, a good NaI detector may have an energy resolution of 6% at 662 keV that corresponds to fwhm of 40 keV.

An ideal photo-peak can be described by the Gaussian distribution, illustrated in Equation (17), as

$$C = \frac{A}{\sqrt{2\pi\sigma}}\, \text{e}^{\frac{-(E-E_0)^2}{2\sigma^2}} \tag{17}$$

where C is the number of counts per energy interval, E the energy, E_0 the energy of the peak centroid, σ the standard deviation of the distribution, and A the peak area. The baseline under the peak has to be subtracted in a proper manner. Often a constant or linear approximation across the baseline of the peak is sufficient. Spectrum analysis programs fit a Gaussian curve or a modified Gaussian

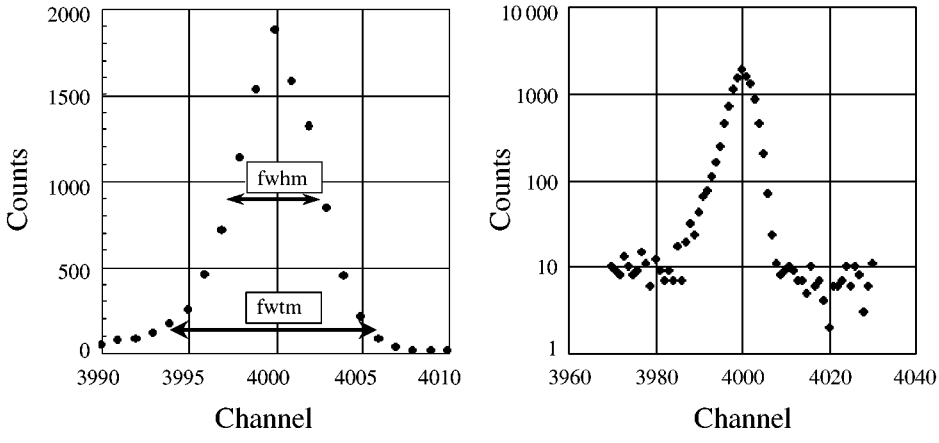

Figure 6 Verification of resolution and peak shape. The measurement was performed with a ^{60}Co source of 5.6 kBq (the detector is described in Table 2). Both curves represent same data. One channel corresponds to 0.33 keV. fwhm can be estimated in different ways. The most straightforward solution is based on peak area (11740) and peak height (1863). Equation (20) gives an estimate of 5.34 channels or 1.79 keV for the fwhm. The logarithmic display of the pulse height distribution reveals that the shape of the peak is not completely symmetric. At the rising part of the peak there is a clear tail that is caused by incomplete charge collection or electronic noise. Fine tuning of the measuring system should be adopted to further improve the quality of the spectrometer.

curve to the measured distribution describing the peak shape (section 6.4).

Equation (17) gives the relationship between the fwhm and σ and the result is Equation (18):

$$\text{fwhm} = 2\sqrt{2\ln 2}\,\sigma = 2.355\sigma \tag{18}$$

Therefore, Equation (16) can be written as shown in Equation (19):

$$\text{fwhm}^2 = \text{fwhm}_\text{p}^2 + \text{fwhm}_\text{c}^2 + \text{fwhm}_\text{n}^2 \tag{19}$$

The fwhm can be estimated from the calibration data by simple interpolation. For high accuracy, the peak areas should be large ($>10^4$ counts). This general method does not depend on any assumptions of the peak shape. If it is appropriate to assume a Gaussian shape, a convenient way to calculate the fwhm is based on the peak area A (channel-by-channel summing and appropriate baseline subtraction) and peak height $C(E_0)$, shown in Equation (20), as in Equations (17) and (18):

$$\text{fwhm} = \frac{2.355}{\sqrt{2\pi}}\frac{A}{C(E_0)} = 0.939\frac{A}{C(E_0)} \tag{20}$$

Full width at one-tenth maximum (fwtm) and full width at one-fiftieth maximum (fwfm) describe the peak shape near the baseline more appropriately than the fwhm does. fwtm and fwfm should not be calculated by the area method, however. Their ratio with fwhm is compared with the perfect Gaussian ratios, for which fwtm/fwhm = 1.82 and fwfm/fwhm = 2.38. In an optimum counting system these ratios should be below 1.9 and 2.5 respectively.[11]

3.2 Uncertainty of Charge Production

The uncertainty of charge production in a semiconductor detector was discussed in section 2.2. Table 5 shows the fwhm_p values for the Fano factors between 0.06–0.10 (see Table 4). These uncertainties are idealistic lower limits of the detector resolution.

The average energy W to create an electron–hole pair depends on temperature ($\Delta W/W = 2.53 \times 10^{-4}$ per K).[12] Temperature variations during the measurement shift the position of the peak and, thus, the final peak at the end of the acquisition period is broadened. Temperature increase of one degree moves the position of a peak near 1 MeV by 0.25 keV. Therefore, high-resolution measurement is only possible after temperature equilibrium is reached. A sufficiently long time (1 day) should be allowed for cooling the detector, and the temperature of the surroundings should be kept unaltered during the counting of the sample. A digital spectrum stabilizer is one option for compensation of temperature drifting (section 5.4).

Table 5 Contribution of charge production statistics to fwhm of a germanium detector for Fano factors of 0.06 and 0.10

Fano 1factor	fwhm_p Equation (5)	fwhm_p at 122 keV	fwhm_p at 477 keV	fwhm_p at 1.33 MeV
0.06	$0.99\sqrt{E}$	350 eV	0.68 keV	1.14 keV
0.10	$1.28\sqrt{E}$	450 eV	0.88 keV	1.48 keV

The temperature of the crystal can be recorded using the connection in the detector that gives the HV shutdown signal. The internal temperature resistor changes from 100 Ω at 77 K to 540 Ω at room temperature. With 1 mA current the voltage changes from 100 mV to 540 mV, i.e. $2\,\text{mV K}^{-1}$. So an external 10-bit analog-to-digital converter (ADC) with the proper input range (say 1 V) would give a resolution of 1 mV or 0.5 K. The temperature measurement is implemented in commercial detectors, but only for the protection of the detector against warming and subsequent damage of the preamplifier (detectors built before 1998). However, the continuous demand toward higher quality creates pressure to include various state-of-health indicators in the measurement system, including recording of the temperature of the crystal.

3.3 Uncertainty of Charge Collection

Incomplete charge collection may be a problem in large detectors where the charge carriers have to travel long distances. The peaks in the spectrum may have a low-energy tail that makes spectrum analysis more difficult and masks small peaks near a larger photo-peak. Impurities or crystal defects trap charge carriers. In a strong electric field some trapped charges may be released within the rise time of the preamplifier signal, whereas other charges do not contribute to the pulse at all.

The manufacturers design the detectors to avoid regions with low electric field, and the detectors are operated with a bias voltage as high as possible. In modern detectors the efficiency of charge collection is very near 100%. However, spectrum analysis software should be able to cope with peaks that have tails.

3.4 Electronic Noise

Electronic noise in the preamplifier and in the other signal processing equipment causes uncertainties in pulse height and shape that has significant effects on the overall resolution of the spectrometer. The preamplifier produces most of the noise. Any uncertainty in the front end of the signal processing is magnified by the main amplifier. The contribution of the noise to the resolution can be measured by injecting a suitable charge (Ee/W) to the input of the preamplifier. This charge may be generated by a radiation source and a detector,

or by a pulse generator through a capacitor (known as charge terminator). For estimating the system gain, V_0/E, the output pulse height V_0 at the main amplifier has to be measured. The preamplifier noise is equivalent to the root-mean-square voltage V_{rms} at the output of the main amplifier in the absence of any pulses,[5] see Equation (21)

$$\text{fwhm}_n = 2.355\frac{E}{V_0}V_{rms} \quad (21)$$

The capacitance of the detector and the capacitance of the cabling between the detector and the preamplifier should be as small as possible. Typically, the external capacitance is between 10–40 pF (section 2.5). Other sources of noise are the input field effect transistor (FET), the resistance connected to the input, and the leakage currents from the detector and the FET.

The electronic noise is a constant, independent of energy. Therefore, it dominates total fwhm at low energies. Above 1 MeV the uncertainty of charge production and collection is more important than the noise from electronic equipment.

The electronic noise has two major components. Parallel noise (step noise) results from detector leakage current and from thermal noise in the feedback resistor of the preamplifier. A low count rate, large feedback resistor (1000 MΩ), cooling, and a short time constant in the amplifier (section 5.2) minimize the parallel noise. Some preamplifiers do not need a feedback resistor at all (section 5.1). Series noise (delta noise) depends on the total capacitance at the preamplifier input. To reduce series noise, the main amplifier should be operated with long time constant. This requirement is in conflict with the parallel noise optimization. In practice a time constant of 3–6 μs minimizes the total noise. The FET and the feedback resistor are the critical parts of the preamplifier and for reducing noise, they are normally mounted together with the detector for the operation at low temperature.

The noise of the electronics is dominating the total resolution at energies below 300–500 keV. A state-of-the-art general-purpose detector has fwhm of 750 eV at 122 keV (data sheet by the manufacturer); in practice the resolution could be as good as 600 eV. Such performance opens completely new possibilities for γ-spectrometric nuclide determination. For example, the peaks of ^{235}U (185.7) and ^{226}Ra (186.1 keV) may be resolved using high-quality deconvolution algorithms. The task is inaccurate or impossible for detectors with a noise level above 1000 eV (see Figure 17).

4 GERMANIUM DETECTOR AND ITS OPERATION

4.1 Counting Geometry and Detector Selection

Sufficiently high resolution and counting efficiency are the key requirements for germanium detectors, and their specifications should be chosen according to the particular application. A thick large-volume detector is a good solution for high-energy photons (>1 MeV), but it is not necessarily the optimum choice for normal environmental analysis where the emphasis is more on low and medium energies (<1 MeV). Increasing the volume of the detector improves counting efficiency but, unfortunately, also increases the background count rate, with a consequent reduction in the detection capability. The cost is reduced by choosing a detector with optimum dimensions.

Detector manufacturers specify products according to the concept of relative efficiency, where the efficiency of HPGe detector is quoted relative to that of an 3″ × 3″ NaI detector for the 1.33 MeV photons of ^{60}Co, with a 25-cm source detector distance. The relative efficiency is the ratio, expressed as a percentage, of the corresponding detection probabilities of the two detectors with full-energy peaks. A ''100% detector'' has a counting efficiency of 1.2×10^{-3} for 1.33 MeV photons of ^{60}Co at a source detector distance of 25 cm.[13] Although relative efficiency is the basic quantity for specifying the performance of the detector, this parameter is far too simple to be used as the key specification in the purchase of a detector. The overall counting geometry, including the shape of the sample and the dimensions of the detector, must be optimized.

The environmental samples may contain high-energy background nuclides, such as ^{40}K (1460.8 keV) or ^{208}Tl (2614.5 keV), which make the detection of low-energy peaks more difficult. The Compton continuum in a γ-spectrum results from Compton interactions in the detector and its surroundings (section 2.1). The peak-to-Compton ratio is defined as the ratio of the height of 1.33 MeV ^{60}Co peak to the average Compton plateau between energies 1.040 MeV and 1.096 MeV. An ideal detector to determine low amounts of unknown radionuclides has a high counting efficiency, good energy resolution (fwhm < 2 keV at 1.33 MeV and <0.8 keV at 122 keV), large peak-to-Compton ratio (>70), and optimized dimensions for the sample and energy of interest.

The coaxial p-type detector has been the standard choice for environmental γ-spectrometry (section 2.6), whereas n-type detectors provide detection over a larger energy range but are more expensive. The greater efficiency of the n-type detector at low energies also increases the summing of X-rays. If nuclides with photon energies below 60 keV, such as ^{241}Am (59.5 keV) or ^{210}Pb

(46.5 keV), need to be detected and quantified, a thorough cost-efficiency analysis should be performed to choose between a coaxial n-type detector and a planar p-type detector. An extended range coaxial p-type detector and a thick planar detector should also be considered in these comparisons (Table 4). The window material must be chosen according to the energy range of interest (Table 3).

The relative efficiency R_{eff} (%) is directly proportional to the volume of the detector V (cm^3). Approximate expressions are illustrated by Equations (22) and (23)[9,11]

$$R_{\text{eff}} = \frac{V}{4.3} \tag{22}$$

and $$R_{\text{eff}} = KD^{\alpha}L^{\beta} \tag{23}$$

where $K = 243.21$, $\alpha = 2.8155$, $\beta = 0.7785$ and D is diameter (decimeters) and L length (decimeters). A 100% detector has a volume of about 400 cm^3. The price of the detector is proportional to the volume and, consequently, to the relative efficiency. A cost–benefit analysis is certainly warranted to optimize the performance of the counting system.

The first step in detector selection is to define the energy range of interest and decide at which energy, i.e. for which nuclide the detection capability should be optimized. The sample geometry plays a major role in the selection of the dimensions of the crystal.

Let us consider the selection of an HPGe crystal for an air surveillance program that is intended to detect trace amounts of fission products. The energy range of interest is 60–1800 keV. Furthermore, the system specifications require that the minimum acceptable relative efficiency is 40%, and ^{140}Ba is the key nuclide for which the detection limit is defined (section 6.5). In high-volume air sampling the area of the filter is about 0.1 m^2. For efficient counting, this area has to be reduced to the size that is of the same order of magnitude as the detector dimensions. In the laboratory the sample can be pressed into tablet form, with a diameter of 45 mm and thickness of about 6 mm (this is referred later as a "small" sample). Another sampling device might produce samples that have a diameter of 80 mm and thickness of 10 mm ("large" sample). The task is to find the optimum detector for both types of samples. The counting efficiency should be high, but the detector volume should be as small as possible (to reduce costs and baseline in the spectrum).

The primary photons of ^{140}Ba have an energy of 537 keV; more than 90% of these photons (axial direction) interact in a germanium crystal with a thickness of 55 mm (Figure 7). Increasing the thickness (volume) of the detector would not greatly affect the photo-peak count rate (although the first interactions are mainly Compton scattering). However, a larger thickness would increase the level of baseline in the spectrum because the air sample and the surrounding construction materials contain natural radionuclides (^{208}Tl and ^{40}K) which emit high-energy photons. The required detection limit can therefore be attained without incurring the cost of a large-volume detector.

Relative efficiency as a function of detector dimensions is shown in Figure 8. This diagram, together with Figure 7, can be used for the selection of the detector. The optimization of the detector is based on the dimensions

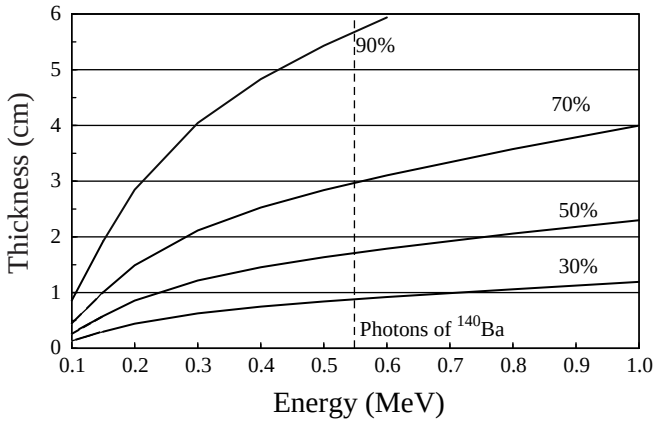

Figure 7 Transmission of collimated photons through germanium. In environmental measurements (see Figure 10), the photons enter the detector at various angles which in most counting configurations increases the probability of interaction within the crystal volume. The initial thickness of the crystal can be chosen from this diagram. For example, there is not much advantage to have a crystal thicker than 55 mm for the detection of 537 keV photons emitted by ^{140}Ba.

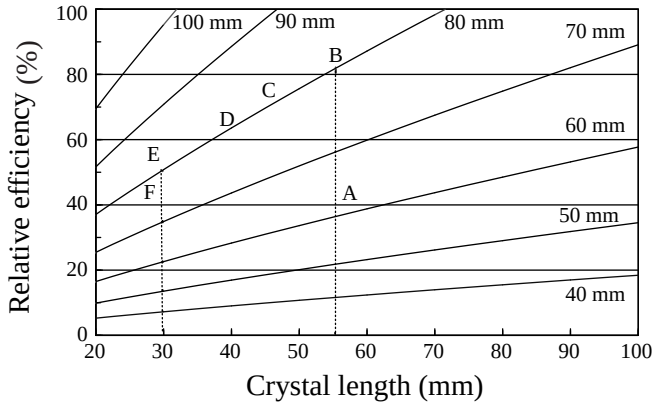

Figure 8 Relative efficiency of a germanium detector as a function of crystal dimensions. The curves are calculated by Equation (23). The parameter above the curves is the diameter of the crystal. This diagram together with Figure 7 can be utilized for the selection of a detector according to the dimensions of the sample and the energy interval of interest. For example, a 40% detector could have very different dimensions. The diameter could vary between 50–80 mm and the length between 20–100 mm. (For further discussion, see text.)

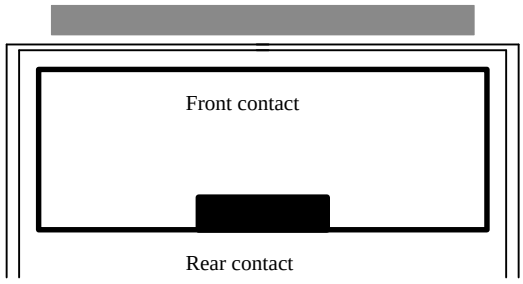

Figure 9 Planar counting geometry. For many environmental applications, the planar detector provides good resolution and optimum counting efficiency with minimum baseline. The diameter and the length of the detector have to be chosen according to the sample diameter and nuclides of interest (energy) respectively. The counting efficiency is maximized in planar geometry because there is no need for bulletization and central core. The front contact can be manufactured with little dead germanium beyond the active region, which provides high counting efficiency also at low energies.

of the sample and the energy interval of interest. A 40% detector (A) with thickness of 55 mm and diameter of 62 mm would fulfil the minimum requirements for the small samples. However, the counting geometry would be inefficient for the large samples, and therefore the detection limit would be high. Detector B with diameter of 80 mm would be a good choice for both cases. It may cost twice as much as the detector A, however. Detectors C and D are thinner (45 and 40 mm), but they may still give lower detection limits for ^{140}Ba than the thicker detectors; in a critical application this should be verified experimentally or theoretically, perhaps by Monte Carlo simulations. Detector D may cost 10 000–15 000 US$ less than the detector B. Detectors E and F could be built in planar geometry (Figure 9). Equipment suppliers should be requested to offer these different types of detectors. A tendering process based on relative efficiency alone is not sufficient to guarantee good detection performance.

Detectors already present in a laboratory should also be optimized for their particular application. Sample counting geometry might have to be modified to suit the dimensions of the detector.

Until recently detector manufactures seem to have had little interest in producing standard detectors with similar characteristics, and customers have been happy to receive better than expected performance (in terms of relative efficiency). Every detector is thus an individual, and has to be meticulously calibrated by the user.

Modern detectors can be cloned with high accuracy. These detectors first became available commercially in 1998. Such detectors can be calibrated by the manufacturer or by special laboratories for a few counting geometries. Because the properties of the standard detectors are well known, reliable efficiency curves can be calculated for any type of sample (section 6.2.2.2). Furthermore, this computational calibration approach provides the means to take into account the variability of the sample density.[14,15] A broken standard detector can be replaced quickly, whereas it may take several weeks to recover from the breakdown of a nonstandard detector because of the calibration requirements (for verifying detector performance see quality assurance procedures in section 8).

Optimization of the detector dimensions gives the best performance with the lowest costs. The adoption of commercial, optimized, standard detectors should give several advantages to the users. The capital costs and the running costs could be greatly reduced and, furthermore, the quality of the measurements could be improved.

4.2 Sample Preparation

Sampling strategies and sample preparation techniques are crucial for the success of environmental monitoring programs. To achieve low detection limits, the sample volume has to be reduced to match the detector dimensions. An ideal sample would be a point source of zero mass. In practice, the whole sample, or a small aliquot, has to be placed in a container of optimized dimensions.

A narrow beaker (Figure 10) with a diameter far less than the diameter of the detector suits samples of small volume ($<30\,cm^3$). To avoid self-absorption at low energies in samples with high density (soil, sediment, etc.), a wide beaker should be considered for volumes larger than $15–20\,cm^3$.[14,15]

If the diameter of the container and the detector are of roughly equal size, the container can be used for a large range of sample volumes, up to $300–500\,cm^3$, depending on the diameter of the detector. A Marinelli container (Figure 10) provides superior counting efficiency for very large volumes.

Environmental samples have different origins, for example gaseous, aquatic, soil, air filter, food and biota sample, with various shapes and volumes. The

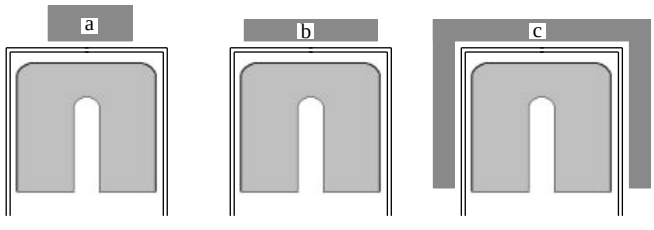

Figure 10 Different counting geometries for cylindrical detectors. For high counting efficiency, the sample should be compacted to a small volume (a). A wide beaker (b) is a better choice for larger samples or for high-density materials. Marinelli beaker (c) has to be used for very large volumes.

laboratories must develop procedures to cope with these different samples. Standard counting geometries, for which detector efficiency calibrations have been performed, are necessary.

Surface water may be sampled with open-mouthed bottles, whereas immersion-type devices can be opened at the selected depth below the surface. Pumping devices allow collecting large volumes of water. Water samples are vaporized in a warm closet (105 °C). A good practice is to use porcelain evaporating basin covered with double cellophane foil.

Bulk water samples should be filtered to trap the sediment. Filters containing suspended particulates must be dried before measurement. Both water and filtered sediment should be analyzed because sorption in sediment and accumulation in biota may enrich radionuclide concentrations.

Some fungi, lichens, and moss are sensitive indicators of environmental contamination because they concentrate radionulides by a factor of 10–100 (compared with soil concentration). Biota samples should be weighed (wet samples) and ashed in a furnace to accomplish complete decomposition. An aliquot of blended ash in a small beaker is a good sample counting geometry.

There are two practices for ashing. Dry ashing is performed by charring the sample at 400–450 °C and/or ashing it at 600 °C. Wet ashing uses acids to chemically destroy organic components in the sample. Dry ashing is suitable only if the element of interest is nonvolatile. Iodine, for example, tends to escape from the sample during the heating. Wet ashing utilizes concentrated nitric, hydrochloric and sulfuric acids, with heating to temperatures above 100 °C. Because of high acid concentration and temperature, sample handling must be performed with great care.

A coring tool is needed for soil and peat sampling. The tool could be a smooth steel bore tube, with soils removed by internal pressure, or a modified golf-course hole cutter, fitted with extended blades that can be removed to collect intact and whole samples. For radionuclide depth distribution analyses, the cores can be sectioned at suitable intervals. Each sample must be placed into an air-tight container to retain moisture. At the laboratory the samples are screened to remove objects larger than 2 mm in diameter, weighed and dried, usually for a few days at temperatures of 105 °C, and then crushed and homogenized (particle size $\ll 100\,\mu m$). Samples are again weighed and placed in standard containers (the volumes are typically between 50 and 150 cm^3). Particular attention has to be paid to the quality assurance of the detector efficiency calibration (see section 6.2).

Air filter samples have a large surface area. The goal of the sample preparation is to produce a tablet with diameter equal to or smaller than the detector diameter (section 4.1). Before pressing, the sample could be wrapped in a thin cellulose paper. Such a procedure results in a sample which has a clean outer surface. Therefore, several samples can easily be piled for one composite sample. Another elegant sample treatment method is to utilize a robot that cuts the filter into small, cloned pieces. Because no pressing device is used, the sample slices are ideal also for autoradiography, which may reveal the particulate nature of the debris and allows isolating single particles for more detailed analysis, including γ-spectrometry.

If extremely low detection limits are required for certain radionuclides, methods of radiochemistry could be applied to get rid of the sample matrix and disturbing inherent natural radioactivity. For example, the following procedure could be used for cesium separation. A small amount of stable cesium carrier (20 mg) is added to the sample, which is then wet ashed with nitric and hydrochloric acids. Cesium is coprecipitated by ammonium molybdophosphate. The resulting homogeneous, thin precipitation has a small mass and contains only cesium, without any interfering nuclides present.

By choosing the appropriate radiochemical procedures, it is possible to achieve significant improvements in detection sensitivity. However, one of the major attractive features of γ-spectrometry is that fundamentally it is a nondestructive technique. Therefore, chemical manipulation of samples should only rarely be resorted to. If extensive chemical treatment is required, then other techniques, such as inductively coupled plasma mass spectrometry, should be considered.

4.3 Cooling Systems

The most common cooling system is based on an isolated dewar filled with liquid nitrogen in thermal contact with the detector. The crystal and the front-end components of the preamplifier are housed in a clean vacuum chamber, which is attached to or inserted into a dewar.

The detector chamber and the preamplifier are packaged together in a compact aluminum cylinder, known as ''end cap''. Cooling is provided through a pedestal that connects the copper ''cold finger'' to the liquid nitrogen reservoir. Part of the detector housing is evacuated to provide thermal insulation. ''Sweating'' of the detector or the dewar is an indication of vacuum insulation problems.

Electrical cooling is operationally simpler than a liquid nitrogen system. Although the investment costs are higher, the running costs are lower during the detector lifetime. The electrically refrigerated cryostat should therefore be considered as an alternative in all applications, even where a supply of liquid nitrogen is available.

The first electrically cooled detectors utilized closed-cycle freon and helium refrigeration. The quality of the equipment was not sufficient to provide γ-spectra with resolution that is achievable with liquid nitrogen cooling. In addition, they had moving parts, which made maintenance necessary at short intervals. Advances in the development of electrical cooling systems have solved the operational and resolution problems encountered in the past, and nowadays, the quality of the γ-spectra is excellent. Commercial equipment uses an air-cooled compressor, which supplies refrigerant to the cold head capsule. The compressed refrigerant expands and cools the detector. Because the cold head has no moving parts, the resolution does not deteriorate because of microphonics, the maintenance costs are low, and the system has a long expected service time (>5 years). The cooling system weighs 20–30 kg and the power consumption is about 500 W. No external pumping is needed. The refrigerant (<100 g) is flammable but is chlorofluorocarbon (CFC)-free.[9,11]

Electrically refrigerated cryostats are designed to operate at ambient temperatures of 10–35 °C. Unlike liquid nitrogen-cooled detectors, which have a stable temperature as a result of the boiling nitrogen, electrically cooled detectors are prone to temperature drifting, with resulting resolution deterioration, unless the ambient air temperature is carefully controlled. The monitoring of the temperature of the crystal at short intervals, perhaps every 10 minutes, provides an external state-of-health signal that can be measured at low cost (section 3.2).

4.4 Detector Shielding

Shielding of the detector is of vital importance in low-level counting because it reduces the amount of background radiation reaching the detector. The environmental background comes mainly from ^{40}K in natural potassium, from nuclides in the decay chains of uranium and thorium, and from cosmic radiation. The background reduction is an art of its own and may contain various passive and active measures. Often the sample itself is the dominating source of background. A practical goal in environmental applications is to attenuate external background radiation by a factor of 1000.

A laboratory dedicated for low-level counting should be designed using materials of low natural activity. Concrete is often the main construction material, and it should be made of sand with low activity of potassium (<100 Bq kg^{-1}), radium (<5 Bq kg^{-1}) and thorium (<5 Bq kg^{-1}). Neutrons, induced by cosmogenic muons, interact with the germanium crystal and the surrounding materials generating photons that may be detected in low-level counting. Neutron fluence rate is reduced in a laboratory that is located underground or, at some extent, in ground floor of a large building. The background related to radon daughters can be reduced, if needed, by flushing nitrogen or clean air into the surroundings of the detector.

The shielding of the detector is usually constructed of lead. Steel is another possibility, but it has some serious drawbacks. Firstly, to attenuate 1 MeV of photons by a factor of 1000, requires 147 mm of steel but only 87 mm of lead.[3] Secondly, modern steel is often contaminated with ^{60}Co; pre-World War II steel might be expensive and difficult to get. Thirdly, Compton scattering is more dominant in steel than in lead, increasing backscatter from the shielding.

The conventional design of detector shielding is based on 100–150 mm of lead. A thicker lead shield does not provide further background reduction because of the enhanced interaction with cosmic rays. Backscatter is minimized by making the detector shield distance as large as possible (>100 mm).

The photoelectric absorption of γ-radiation by lead inevitably produces fluorescent X-rays. The shielding should therefore have an inner layer that is designed to absorb low-energy photons. A typical "graded-Z" shield consists of cadmium and copper or tin. Cadmium absorbs X-rays of lead (72.8 keV, 74.9 keV, 84.8 keV, and 87.3 keV). The copper layer absorbs the X-rays from cadmium (22 keV) and provides further attenuation of X-rays from the lead.

The thickness of the liner materials varies widely; a commercial design, intended to attenuate Pb X-rays by a factor of 100,[11] has a liner with 0.5 mm of cadmium and 16 mm of copper, whereas another design is based on tin (1 mm) and copper (1.5 mm).[9] Cadmium is a toxic material and it has a high neutron cross-section resulting in prompt emission of 558 keV photons; furthermore, the neutron activation product ^{109}Cd emits 88 keV photons. Because of difficulties in material handling and possible inconveniences in spectrum analysis, it should be carefully assessed whether cadmium really is needed for the shielding. In practice, a simple shield liner based on tin (1–2 mm) and copper (1–2 mm) is often sufficient for γ-spectrometry of environmental samples, particularly if the low-energy photons are not of major interest.

Shielding material may contain natural ^{210}Pb, and the shield liner absorbs its 46.5 keV radiation. The ^{210}Bi, decay product of ^{210}Pb, is a beta emitter (E_{max} = 1161 keV), however, and its bremsstrahlung generates a continuum up to 500 keV.[9] A careful design of a lead shield would contain two types of lead. The outer layer could be ordinary lead, whereas the inner layer (10–30 mm) could contain special old lead with low ^{210}Pb content (<20 Bq kg^{-1});[9] the aged lead should be much older than the half-life of ^{210}Pb (22 years). However, if the

sample itself, such as an air filter, contains large amounts of natural beta and γ-emitters, the special lead liner does not improve detection limits of fission products or other nuclides of interest.

Active background shielding is based on anticoincidence counting designed either for the reduction of Compton continuum or for the removal of the contribution of high-energy cosmic radiation. The latter system, known as veto detector,[5] is relevant for the counting of environmental samples only if the overall background is dominated by the counting environment. The veto detector is installed above the lead shielding, and it may improve nuclide detection limits by a factor of two. However, the guard detector itself must be shielded against environmental radiation (false coincidences with the HPGe detector). The system is complex and expensive, and, consequently, it is used only in special applications.

Compton suppression systems utilize a large ''guard'' detector (NaI detector for example) that surrounds the germanium detector. If a primary photon leaves only part of its energy in the germanium detector and the escaping photon interacts with the guard detector, the processing of the HPGe signal is stopped by the guard system that sends a triggering pulse for closing the input gate of the recording electronics.

Active shielding may reduce the Compton continuum by a factor of ten and, consequently, the peak-to-Compton ratio may increase to 1000 : 1. A background suppression of 10 translates to the reduction of the detection limit by a factor of three (section 6.5). This improvement would be crucial for low-level determinations of some nuclides, ^{137}Cs in particular. However, the anti-Compton systems are technically complex and can handle only small samples. The most serious drawback is the suppression of full-energy peaks of nuclides that emit photons in cascade (^{60}Co, etc.). Peaks are suppressed also by random coincidences in the HPGe detector and in the guard detector. Because of these drawbacks, the counting system should not only register the suppressed spectrum but also the original nongated spectrum.

5 MEASUREMENT ELECTRONICS

The signal processing of an HPGe detector is based either on traditional analog electronics or on modern compact digital devices. These two technologies handle the signals coming from the detector in different ways.

In the analog device, the charges created and collected in the detector are transformed, shaped and amplified to a large voltage pulse. This pulse is thereafter converted to a digital form by a fast and stable ADC.

In the digital device, the pulse generated by the preamplifier is digitized and there is no need for an amplifier or ADC. The most advanced digital equipment can handle overlapping signals normally rejected by the analog amplifier due to false signal shape. Digital signal processing also provides excellent temperature stability for peak position and shape.

5.1 Preamplifier

The main functions of the preamplifier are to convert the charge generated at the detector to a voltage pulse, to amplify the signal and to transfer the pulse to the main amplifier.

The charge created in the detector is so weak (5.4×10^{-14} C MeV^{-1}) that it cannot be used directly for the radiation detection. Preamplifiers designed for γ-spectrometry are charge-sensitive and produce output pulses with amplitude proportional to the integrated charge, i.e. to the energy received from the incident photon. An advantage of the charge-sensitive preamplifier is that a change in the input capacitance has no affect on the output voltage (Figures 4 and 11). The total capacitance of the detector has, of course, a crucial importance for resolution (section 3.4)

The preamplifier is located close to the detector to avoid noise caused by the capacitance in the wires between the preamplifier and the detector. The pulse from the preamplifier has a short rise time but a long decay time (around 50–100 μs). Therefore, pulses with different charge collection times can be handled. A preamplifier does not usually have pulse-shaping capability.

The rising edge of the preamplifier pulse carries the information on radiation energy. The main amplifier extracts this information for further processing. However, at high count rates, the piling of pulses in a typical preamplifier with RC-feedback may be a problem. The average DC voltage level may become too large to preserve linearity between charge and pulse height. At very high count rates, the DC voltage may rise to near the supply voltage of the preamplifier (24 V). Then, saturation or paralysis has to be taken into account, particularly in applications intended for emergency preparedness. Count rate performance is described in terms of energy rate; a typical preamplifier has a safe working level below 10^5 MeV s^{-1}.

The noise of the preamplifier is reduced in a design with no feedback resistor, and where the charges accumulate in the capacitor C_F (Figure 11). This results in a stepwise rise in the output voltage; each step is the result of a separate charge collection process. From time to time the output has to be reset to the zero level to prevent the system exceeding the saturation level. This kind of resetting is performed using optical feedback or transistor reset.

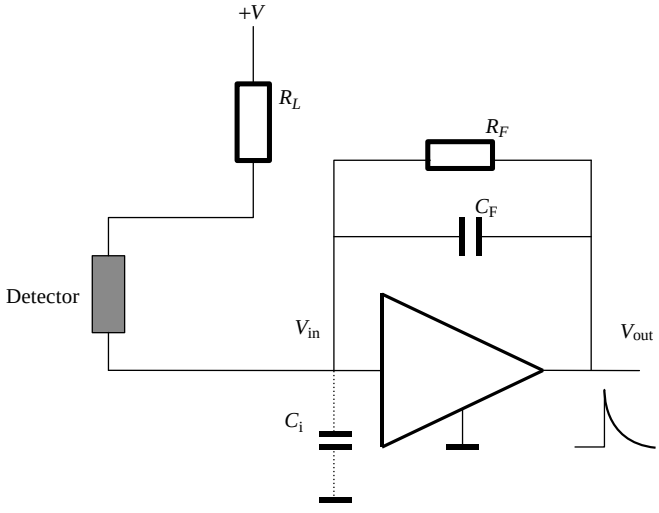

Figure 11 Simplified diagram of a DC-coupled preamplifier. Because the amplification of the operational amplifier is very large ($\gg (C_i + C_F)/C_F$), the output voltage is nearly independent of the input capacitance C_i. Therefore, the output voltage is directly proportional to the initial charge Q created at the detector ($V_{out} = Q/C_F$). To provide operational convenience, the high voltage is normally supplied via the preamplifier.

Optical feedback provides the lowest noise, but transistor feedback is useful in high-count rate systems.

5.2 Amplifier

The purpose of the amplifier is to convert the output signal from the preamplifier into a form that is suitable for the application, for the ADC in particular. Signal amplification is only one of the roles of the amplifier. Pulse shaping is the primary function of the instrument. Another important feature is the rejection of piled pulses at high count rate. Furthermore, pole-zero cancellation (Figure 12) and baseline restoration (see later) are needed to compensate for undesirable consequences of the signal processing. The modern amplifier is a very sophisticated instrument which would more appropriately be named as signal processor or pulse processor.

Since the preamplifier signal decays slowly toward the baseline, each new pulse is added on the tail of the previous pulse. The amplifier restores the baseline reference level between pulses by differentiation (Figure 12). To reduce noise and to measure the amplitude accurately, the signal has to be integrated. The differentiator is a high-pass filter and the integrator is a low pass filter. The initial long preamplifier signal is converted to a short pulse that can be handled by the ADC. The shaping time constant (Figure 12) has to be optimized for each counting system. For small HPGe detectors the optimum time constant is 2–4 μs, whereas the large detectors need a longer period

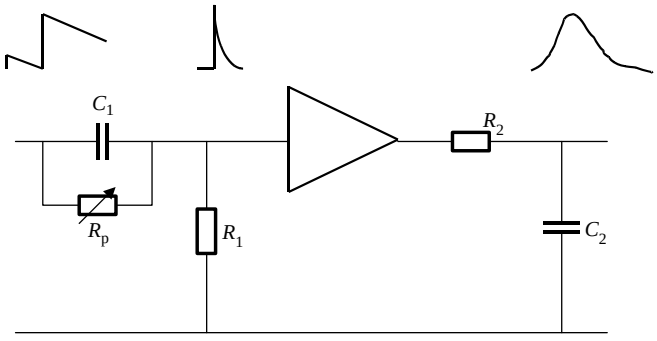

Figure 12 Passive CRRC-shaping of the signal. The shaping time constant $\tau = R_1C_1 = R_2C_2$ of an amplifier is typically 1–10 μs. The pulse has a maximum at 1.2τ.[9] Without compensation for the long decay time of the preamplifier pulse, there is a small amplitude undershoot at 6τ or 7τ. The resistor R_p corrects this problem. The process is known as pole-zero cancellation. This correction matches the differentiator to the slow fall time of the preamplifier pulse. In practice the pole-zero cancellation is performed using a point source close to the detector and by checking that the amplifier output baseline is restored to zero level optimally.

of 4–10 μs. Reducing shaping time increases detector throughput at the expense of resolution.

The shape of the output pulse is often semi-Gaussian, and the height of the pulse is proportional to the original energy absorbed by the detector. The semi-Gaussian shape is achieved by a single differentiator followed by two active integrators. The active circuits are superior to passive RC integrators, particularly for pulse symmetry and for signal-to-noise ratio.

Pile-up, also known as random coincidence summing, is a process where two or more photons arrive to the detector simultaneously or almost simultaneously. At high count rate the preamplifier pulses may be piled. If such a signal goes through the amplifier, the ADC may then misinterpret the ill-conditioned pulse. Therefore the piled pulses have to be rejected using a special pile-up-rejector (PUR) which ignores pulses that are too close to each other.[5] PUR improves the quality (peak-to-baseline ratio) of the spectrum by reducing the summing effect of two independent photons.

The unipolar output pulse of the amplifier has to be normalized to a DC reference that is usually 0 V. Baseline stabilization ensures that the successive pulses are precisely quantified. The simplest baseline restorer is just a CR differentiator. The correction circuit guarantees that the average signal area above ground equals the average signal area below ground. To prevent the introduction of noise, the time constant of the differentiator must be at least 50 times larger than the shaping time constant of the amplifier. At low count rate the baseline between the pulses remains near the ground potential. However, at high count rate the baseline must shift down. This can be

corrected by a gated baseline restorer that connects the CR circuit to ground only between the pulses.

5.3 Analog-to-digital Converter

ADC is intended for the conversion of the amplifier pulses to a digital form. The ADC converts the height of each analog pulse to a number, i.e. to a sequence of bits representing the energy of the incident photon. The output of the ADC specifies the channel in the MCA at which one count will be stored for each output pulse of the same size. The final result of the process is a pulse height distribution, i.e. an energy spectrum.

In the ''Wilkinson ADC'', the pulse charges a capacitor. Thereafter the capacitor is discharged at a uniform rate. The time of the discharge is proportional to the initial maximum charge on the capacitor. The discharge time is measured with high precision using a clock (100–500 MHz) that receives a gate signal at the beginning and at the end of the discharge. The Wilkinson ADC has excellent linearity but the digitization time varies according to the pulse height.

In the ''successive approximation ADC'', a fixed time is used for the analog-to-digital (AD) conversion. The initial analogue pulse is stretched and then compared several times with a changing reference voltage. During the process the channel that represents the pulse is generated bit by bit. For example, a conversion process with 12 stages has a resolution of 12 bits, i.e. 4096 channels. The conversion time in the successive approximation ADC is typically 10 μs, but the fastest instruments need less than 1 μs.[9,11]

Conversion gain of the ADC refers to the number of channels in the spectrum. In high-resolution γ-spectrometry this setting is between 4096 and 16384 channels (14 bits). The dynamic range at which the ADC is working can be reduced using lower and higher level discriminators which select the voltage range for the ADC. These settings are practical if there are lots of pulses beyond the energy range of interest. All the pulses converted in vain increase dead time of the measurement system.

Dead time is the time required for pulse shaping and AD conversion. During this period the system is not able to handle new incoming pulses. The dead time depends on the pulse rate and equipment conversion times. The dead time and its complement, the live time, are calculated during the spectrum acquisition by the computer that controls the measurement. The live time is the actual time during which the system collects counts. For proper timing, a live time clock is a common solution that works well at low count rates because all MCA dead time losses are manifested at its input gate. A further development is the Gedcke–Hale method that compensates also for the losses due to the leading edge pile-up in the amplifier.[5,11]

To verify the dead time reported by the measuring equipment, the actual throughput of the system can be checked in a simultaneous measurement using a sample and a standard source. The procedure is as follows:

1. Verify that the sample does not produce a peak at the same location as the standard source.
2. Measure the pulse rate of the standard source for the peak of interest. The pulse rate is calculated as the ratio of the counts detected at the peak divided by the live measuring time reported by the hardware.
3. Repeat the measurement for the sample and the standard source simultaneously.
4. Compare pulse rates above. The system hardware reports correct live time (and dead time) if the pulse rates are equal.

Another possibility to verify or to calculate the dead time is the following procedure:

1. As above.
2. Measure the count rate from the calibration source. Verify that the dead time is zero (a weak source). Calculate the peak count rate as the ratio of the peak area to the real measuring time.
3. Repeat the measurement for the sample and the standard source simultaneously. Calculate the peak count rate as the ratio of the peak area to the real measuring time.
4. Calculate the live time and the dead time (in percentages) by dividing the pulse rates in points 3 and 2.

The standard source above can be replaced by a high-quality pulse generator.

5.4 Spectrum Stabilization

A digital stabilizer may be needed for an MCA in unstable conditions, such as when temperatures may vary. The most common way to stabilize the spectrum is to use a radioactive source that has one peak at low energy and another peak at high energy. The amplifier zero and gain and the ADC are then adjusted in real time to compensate for any drifting. Another possibility is to use a pulse generator. For low-level counting this method is preferred because it produces no Compton continuum (section 2.1). On the other hand, from the operational point of view a pulse generator is much more complicated and unreliable than a simple point source.

5.5 Digital Signal Processing

Traditional γ-spectrum acquisition hardware includes several devices. Firstly, the preamplifier converts the

charges created in the detector to a voltage pulse. Secondly, an amplifier shapes the pulse to a form suitable for the ADC and, finally, the ADC determines the height of the pulse in digital units. Each one of these steps produces noise and consumes some time, resulting in broadening of peaks and a reduction in pulse throughput of the system.

Recently, instruments based on a digital signal processor (DSP) have been developed. The DSP combines functions of amplifier and ADC. To enhance the signal-to-noise ratio, the pulse obtained from the preamplifier is converted to triangular, trapezoid or quasi-trapezoid shape and the result is digitized directly to channel addresses. The shape of the pulse can be adjusted to fit the detector individually. Particularly, the duration of the pulse ''flat top'' can be set to compensate for the problems of detector charge collection and subsequent ballistic deficit. Also, the pulse rise and fall time can be set. These features together with the elimination of separate amplifier and ADC reduce noise in the measurement system, with a consequent improvement in resolution. In addition, the throughput of the measurement system rises, because the pulse shaping requires less time than with the traditional pulse shaping electronics. The DSP can easily reach a throughput of 100 000 counts per second, which can only be matched by the best and most expensive amplifiers and ADCs. In the future, commercial DSP may also have the power to resolve the overlapping preamplifier pulses in real time.

6 γ-SPECTRUM ANALYSIS

6.1 Characteristics of the Spectrum

The analysis of a γ-spectrum is based on photo-peaks; the analysis software locates the peaks, determines the peak areas (counts) and thereafter calculates peak emission rates (counts per second corrected for detector efficiency), associates peaks with radionuclides, and calculates their activities. It is important to know if the activity of a radionuclide is above a statistical detection threshold. Therefore, the detection limits are usually determined during the analysis phase as well. Many software packages are available for performing spectrum analyses.[16]

A typical γ-spectrum consists of many components. Most of the baseline in the spectrum is formed by photon scattering. Around the region of 150–300 keV, the backscattered radiation is manifested as a broad peak or an increased baseline. The peaks in the spectrum have many origins (Figure 13).

6.2 Calibrations

Four different calibrations have to be performed for γ-spectrometric measurements:

- energy calibration, i.e. channel vs. energy;
- efficiency calibration, i.e. γ-peak area and source activity relationship versus energy;
- shape calibration; i.e. peak shape versus energy;

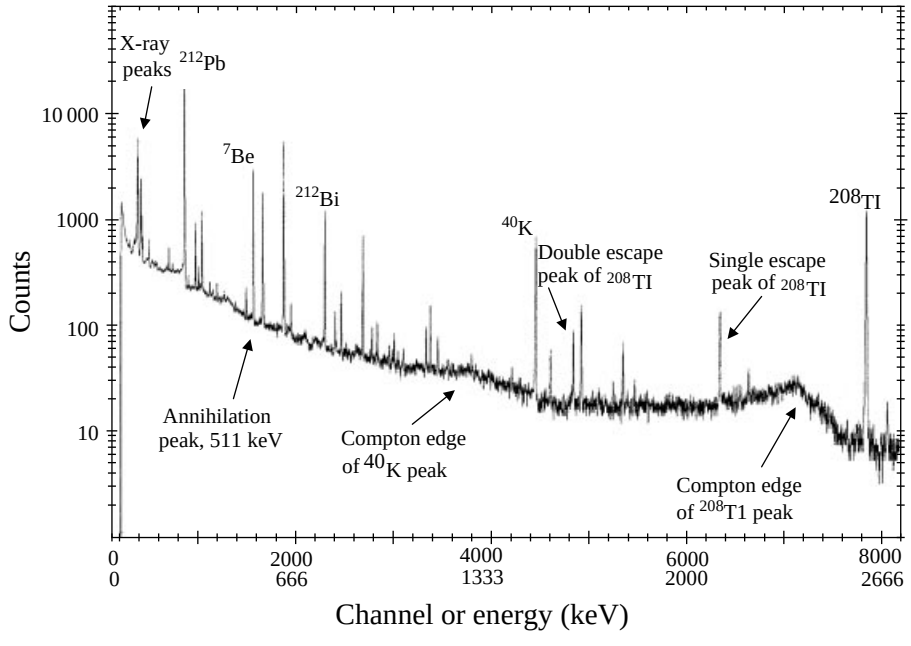

Figure 13 Typical air filter spectrum containing normal natural radionuclides. Analysis software, such as Shaman,[17] is designed to explain the origin of all photo-peaks, escape peaks, sum peaks, X-ray peaks and X-ray escape peaks.

- total efficiency calibration, i.e. total number of photons detected per source activity versus energy.

The high quality of the calibrations is of utmost importance, because all the analysis results rely on them.

6.2.1 Energy Calibration

Energy calibration must be performed before any other calibration or analysis procedure. The calibration represents the relationship between the MCA channels and the photon energies. The calibration procedure is quite straightforward because the resulting channel number versus energy plot is very close to a straight line. Energy calibration can be generated using knowledge of γ-energies contributing to spectral peaks (Figure 14). The energy calibration has to be recalculated periodically to maintain the correct energy-channel relationship. The amplifier gain tends to drift due to environmental reasons (temperature, etc.), although this can be compensated during spectrum acquisition using digital stabilization (section 5.4).

6.2.2 Efficiency Calibration

6.2.2.1 The Empirical Calibration Method The empirical calibration method is a straightforward but laborious procedure. In the empirical method the calibration source resembles the actual sample, with the chemical composition and size being similar. Many γ-spectrum analysis programs support a semiautomatic process for the efficiency calibration. A fully automated procedure is also possible.[19]

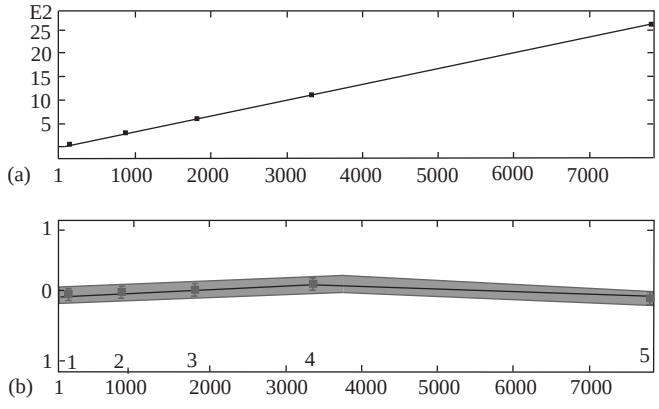

Figure 14 Energy calibration (a) and its nonlinearity (b) as seen on actual SAMPO 90 display.[16,18,19] MCA channel number is on the x-axis and energy (keV) on y-axis (E2 = $\times$100). The nonlinearity plot shows the distance between the data points and the calibration curve, i.e. the distance between the energy-channel pair and the line fitted to the data. When the nonlinearity is much below 1 keV, the calibration is close to perfect. In the present calibration, the energy-channel pairs are calculated automatically using peaks of the naturally occurring radioisotopes (46.5–2614.5 keV).

The calibration source is added evenly to the sample matrix. Self-absorption is taken into account only if the calibration source has the same density as the real sample. Otherwise self-absorption correction has to be introduced at the data analysis phase.

The calibration source is usually a liquid containing many isotopes. To prevent the effect of coincidence summing (section 6.6), the calibration sources have to be selected carefully. The best choice is to use single-line sources, such as ^{241}Am and ^{137}Cs (Appendix 1). The calibration liquid has to be diluted to a suitable concentration (dead time below 10%), and it must be handled cautiously. Even a minor contamination of the detector or the shielding may disturb or prevent low-level sample analysis.

The calculation of the efficiency calibration curve is straightforward from the results of the measurements performed. Only the peak count rate, calibration source activity and photon branching ratio (obtained from the nuclear data) have to be taken into account. For example, if the ^{137}Cs calibration source with activity of 10 kBq has been used, and the calibration peak count rate is 100 cps, the efficiency at the energy of 661.7 keV (^{137}Cs) is $100/(10000 \times 0.85) = 0.0118$. The measured calibration points are usually fitted to logarithmic polynomials, but this mathematical relationship between efficiency and energy has no firm theoretical background and does not help to understand the calibration. However, the log–log plot of the calibration data must form a straight line above 200 keV; any deviation from this line is an indication of erroneous calibration measurement.

6.2.2.2 The Semiempirical Calibration Method The semiempirical calibration method requires that measurements are performed to generate one or more efficiency curves for a simple fixed geometry. Point sources with well-known activities (uncertainty <5%) are placed in a predefined position at a distance larger than 10 cm from the detector end cap (this arrangement avoids problems with true coincidence correction, see section 6.6). The calibration is then performed with a small set of nuclides (Figure 15), such as ^{241}Am (59.5 keV), ^{57}Co (122.1 keV), ^{137}Cs (661.7 keV) and ^{60}Co (1173.2 keV, 1332.5 keV). To have more confidence in the result, more nuclides should be used; for example the calibration data points from ^{88}Y (898.1 keV, 1836.1 keV) and ^{141}Ce (145.4 keV) would be valuable additions to the measured efficiency curve. The activities of the nuclides should be between 5 kBq and 500 kBq and more than 10^4 counts should be acquired in each peak (to allow a counting error of <1%). The point source measurements provide complete calibration curves for a wide energy range. Thereafter the final efficiency calibration curve is generated by software which

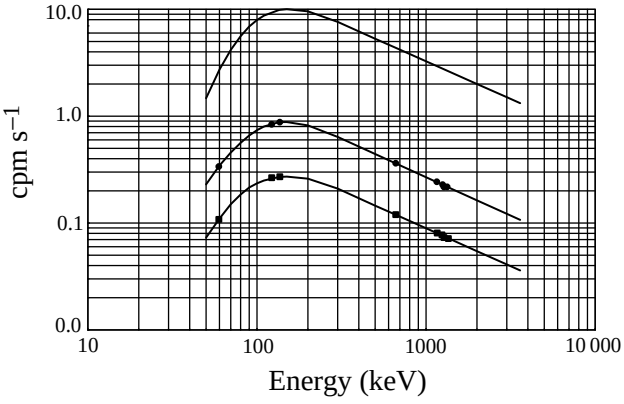

Figure 15 Peak efficiency calibration (counts per minute for emission rate of one photon per second). The upper curve is the final calibration curve for the sample of interest (air filter sample with diameter of 77 mm, height of 9 mm and density of 100 kg m^{-3}). This response was calculated from the other two experimental curves.[14,15] The measurements were performed using reference point sources (^{241}Am, ^{57}Co, ^{137}Cs and ^{60}Co) at distances of 131 mm and 251 mm from the end cap of the detector. The activities of the point sources were between 5.5 kBq and 385 kBq. Four X-ray images of the detector were taken to verify the detector dimensions (Table 3) given in the manufacturer's data sheet and to find out the exact location of the detector inside the housing. These data together with the beaker dimensions are needed for the peak efficiency calculation of the sample geometry. (Reproduced by permission of the Radiation and Nuclear Safety Authority.[20])

converts the measured efficiency curves to the geometry of interest (Figure 15). The computation involves a straightforward integration over the detector and sample materials.[14,15] This method provides a means of taking into account variability of sample dimensions and density. However, it requires accurate information on the detector itself and on the surrounding materials, including the sample.

The calculated peak efficiency calibration must be verified with a sample comparison measurement using an HPGe detector with well-known characteristics. The overall inaccuracy of the efficiency curve is related to activity inaccuracies of the point sources; the method has a typical uncertainty of 5% over the energy range of 50–3000 keV. Reliable and validated software, such as Decca,[15] is necessary for the success of the method.

The semiempirical calibration method is simple, reliable, and provides a way of obtaining calibration curves for different samples. This method should be chosen, whenever possible, avoiding awkward geometry-specific measurements that are prone to errors and contain a substantial risk for detector contamination. In the future, detector manufactures may provide point source calibration curves for their standard detectors and software for the user-specific calibration needs.

6.2.2.3 The Computational Calibration Method The computational calibration method includes Monte Carlo calculations and requires no measurement. All the calculations are performed using information about the geometry, the sample and the surrounding materials. The detector and the sample need to be described in detail; the window materials, the dead layer, the vacuum around the detector and the properties of the sample and its container are crucial information for the success of the Monte Carlo calculations. Computational methods are likely to become increasingly important in the future as hardware and software capability improves, and as standard detectors with well-known characteristics become available. However, modeling of the pulse acquisition system is also warranted. Otherwise the overall uncertainty of the simulations is above 5–10%, which is unacceptable for operational use.

6.2.3 Peak Shape Calibration

Peak shape calibration is required for peak location and fixed-width peak area determination. For optimum performance, the peak location algorithm requires input of the width of the peak.

The simplest method for shape calibration is to calculate the fwhm directly from the data for a number of γ-peaks (section 3.1). Another method is to perform free-width peak fitting for selected calibration peaks. This method requires that the γ-peaks are not part of a multiplet.

The peak width calibration is not sufficient for the determination of the shape of γ-peaks that have tails. If they are not properly taken into account, the peak area determination is inaccurate. The tailing parameters have a slight energy dependence.[18] The tails in the γ-peak occur typically in the lower energy side (section 3.1). Sometimes, there can also be a tail in the higher end of the γ-peak as a result of high count rate problems, with pulse pile-up effects in the electronics.

In addition to normal photo-peaks, the spectrum contains other peaks. Sum peaks, single escape peaks, X-ray peaks, and the annihilation peak are slightly wider than the photo-peaks at the same energy. Such peaks should not be used in the generation of the calibration curve.

The spectrum analysis program has to construct a mathematical relationship between fwhm and energy. The following functions have been used for the shape calibration, Equations (24–27):

$$y = a + bE \tag{24}$$

$$y = a + bE + cE^2 \tag{25}$$

$$y = a + b\sqrt{E} \tag{26}$$

$$y = a + b\sqrt{E} + cE + \cdots \tag{27}$$

where $y = \text{fwhm}$, and a, b and c are fitting parameters. None of these equations have any theoretical justification. The peak width is a smoothly increasing function of energy. Therefore, all of these equations seem to describe the data reasonably well, at least when justified by the naked eye. Equation (26) is used widely; it has, however, the inherent feature of falling off rapidly at low energies, overestimating at medium energies, and underestimating at high energies. Equation (27) corrects these shortcomings, and describes the data well, but it has the same drawback as the other equations; there is no interpretation for the fitting parameters that would refer directly to the quality of the equipment.

Equation (16) shows that the variances of charge production, charge collection and electronic noise are additive, but not the standard deviations. The variance of the charge production is proportional to the energy (Equation 5). The noise of the equipment is constant, independent of energy, and dominates the total resolution at low energies. It is not clear what function would describe the uncertainty of charge collection fortuitously; this effect is very small for modern detectors, particularly for detectors with small dimensions. To improve the quality of the fit, however, a correction factor proportional to E^2 could be used, if needed, for large detectors. Then, the fwhm would be described, as illustrated in Equation (28), by

$$\text{fwhm} = \sqrt{a + bE + cE^2} \tag{28}$$

and the curve fitting would be performed by the function (as shown in Equation 29)

$$y = \text{fwhm}^2 = a + bE + cE^2 \tag{29}$$

The components of the fwhm for charge production and noise are shown in Equations (30) and (31) as

$$\text{fwhm}_\text{p} = \sqrt{bE} \tag{30}$$

$$\text{fwhm}_\text{n} = \sqrt{a} \tag{31}$$

In addition, in Equation (32), the constant b is constrained according to Equations (5), (18) and (30) as

$$b = 2.355^2 FW = 5.546FW \tag{32}$$

The acceptable values of b are limited by $F(0.058–0.12)$ and $W(2.96\,\text{eV})$. The noise component fwhm_n could be estimated directly (Equation 21). Measurement is not necessarily a practical approach.

Similarly, the contribution of charge collection could be described, Equation (33), by

$$\text{fwhm}_\text{c} = \sqrt{c}E \tag{33}$$

The interpretation of c is not necessarily meaningful. A fit without the second-order term ($c = 0$) may give an estimate for b that is somewhat too large. This is an indication of imperfect charge collection in the detector. The correction term ($c > 0$) improves the fit and reduces the value of b to an acceptable range.

Briefly, in curve fitting, there are two options which have theoretical justification:

1. apply Equation (29) with $c = 0$;[6] if the fit is bad, include second order term ($c > 0$);
2. apply Equation (29) with constraints for a, b and c; the value of a is a constant or it may be constrained to a range given by the error estimate of the noise measurement; the value of b should be constrained to a range of 0.94–2.0 eV; $c \geq 0$.

Typical behavior of the peak width can be seen in Figure 16.

6.2.4 *Total Efficiency Calibration*

Total efficiency calibration is required for coincidence correction (section 6.6). This calibration curve represents the relationship between all the photons detected in the detector and the source activity. The calibration depends on source geometry and source detector distance. As most of the photons are detected at the Compton continuum, the surrounding material also has an effect on this calibration. Only single-line emitters should be selected. Emitters with two lines can be used but the mean energy of the γ-peaks has to be selected for the calibration energy.

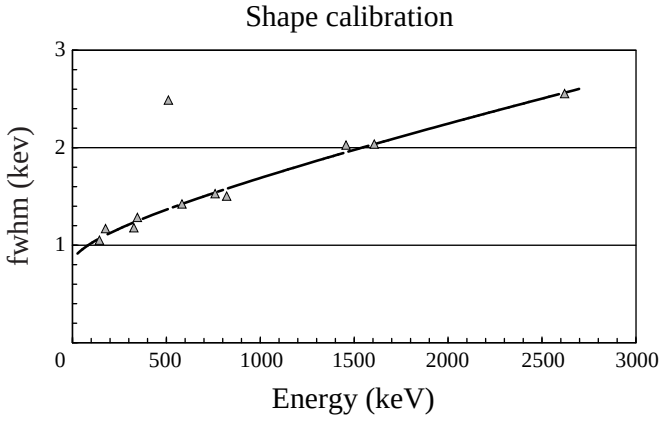

Figure 16 Peak widths of a typical γ-spectrum. A square root polynomial of degree 3 is fitted to the data (Equation 27). Equation (28) gave exactly similar fit (difference less than 0.02 keV). $\text{fwhm}_\text{n} = 0.92\,\text{keV}$ (Equation 31) and $F = 0.11$ (Equation 32). The separate dot represents the width of the annihilation line (511 keV). The annihilation line is about twice as broad as the γ-peaks at the same energy.

6.3 Peak Location

The most common method of finding γ-peaks in the spectrum is to calculate the smoothed second difference and to determine if there is a peak above the search threshold.[21] Peak detection is based on signal-to-noise ratio of peak to baseline, and the result is presented as standard deviations. In the peak location process, one also has to be aware of the peak shape which is specific to the detector used. The peak width is the essential parameter. The shape of the filter varies in different applications. Other peak location methods are fast Fourier transformation for locating the nominal values in the spectrum and a correlation method based on integer summing.[5,6]

Because the data are discrete pulse counts with stochastic fluctuations, statistical criteria have to be applied to the peak search. The generalized second difference, illustrated by Equation (34)

$$dd_i = \sum_{j=-k}^{k} c_j n_{i+j} \tag{34}$$

and its standard deviation are calculated for each channel i as shown in Equation (35):

$$sd_i = \sqrt{\sum_{j=-k}^{k} c_j^2 n_{i+j}} \tag{35}$$

where n_{i+j} is the pulse count at channel $i+j$, and c_j is a constant multiplier which can be obtained using different filters, such as a Gaussian function with a distribution similar to the γ-peak.[5,6,18,19] Standard deviation of pulses at each channel n_i is assumed to be the square root of n_i. If the γ-peaks have pure Gaussian shape, with standard deviation s, on a constant baseline b, then (Equation 36)

$$n_{i+j} = a \exp\left(\frac{-(i+j)^2}{2s^2}\right) + b \tag{36}$$

where a is a constant. The peak is located at the channel where the smoothed second difference has its negative minimum. This is the same channel where the first difference changes the polarity from positive to negative.

The peak search threshold is a statistical limit the peak area has to exceed for its acceptance. The peak significance tells approximately how many standard deviations the peak area is above the background area defined by the baseline. The search algorithm finds a peak if its significance is above the selected peak search threshold. Usually this parameter has a value of 3.0 "sigmas". A different set-up between 2.5–10 may be more appropriate in special applications. A good peak search algorithm can disregard too narrow peaks and Compton edges because of their incorrect shape.

A library oriented peak search is useful for radionuclides of particular importance. In this method, peaks are inserted to the spectrum at the energies of interest. Thereafter, statistical tests are used to decide whether the peaks are really present.

6.4 Peak Area Determination

The simplest way to determine the γ-peak area is to sum the channel contents over the peak and to subtract the baseline by taking into account the baseline level on both sides of the γ-peak. The method is feasible for nonoverlapping γ-peaks.

An accurate method for peak area determination is to fit a shape function into the spectral data. The peak shape is often Gaussian but many programs use tailing parameters to improve the fit.[18] The baseline below the peak is either linear, parabolic, or a step function. There are several methods of describing the step function below the γ-peak. For normal γ-peaks the step is higher on the low-energy side. For the escape peak the step can be higher on the high-energy side of the peak.

Incomplete charge collection, charge trapping and recombination, electronic noise and random summing at high pulse rate tend to broaden the γ-peak. The tails, particularly at the low energy side of the γ-peak, have to be taken into account during peak area determination.

In general, the fitting procedures use the least-squares techniques to minimize the weighted square sum of the fitting residuals (Equation 37):

$$\chi^2 = \sum_{i=k-a}^{k+m} \frac{(n_i/f_i)^2}{n_i} \tag{37}$$

where i is the channel number; k is the center channel of the peak; a and m are the limits for the fitting interval; n_i refers to the counts, and f_i to the fitted function at the channel i.

The baseline of the spectrum should be included in this fitting procedure. The minimized function is calculated as follows in Equation (38):

$$\chi^2 = \sum_{i=k-a}^{k+m} \frac{\left(n_i - b_i - \sum_{j=1}^{q} f_{ij}\right)^2}{z_i} \tag{38}$$

where b_i is the baseline function, q number of peaks in the fitting interval, and z_i is the weighting factor (usually n_i or smoothed value of n_i).

Fitted functions differ in different software. The following possibilities are quite common:

- peak shape can be either Gaussian or Gaussian with exponential tails;

- the baseline below the peak can be either linear, parabolic or step function;
- the step function may have various definitions which have been developed using the data itself, not the physical phenomenon producing the data.

The fitting procedures vary as well:

- linear fixed-width (peaks are not allowed to move during the fitting procedure);
- nonlinear fixed-width (peak position and height are allowed to move during the fitting procedure);
- free-width (there is no limitation on peak width, the peak width is compensating fluctuations caused by uncertainties in the peak width calibration and overlapping peaks);
- mixed (the user can select different approaches for each peak during the interactive peak fitting procedure).

The fixed-width peak fitting procedure requires that the peak shape calibration is properly set up for the spectrum prior to the analysis. The free-width fitting procedure does not need this information, but on the other hand, is not using all the information available to treat the overlapping peaks and, therefore, the error estimates are increased due to the poorly-shaped peaks. Fixed-width fitting makes it possible to resolve highly complex multiplets (Figure 17).

Residual peak search is possible only for the fixed-width fitting procedure. The missing peak can be inserted if during the fitting procedure the analysis software, or the user in the interactive fitting mode, notices a missing component by checking the residual between the fitted curves and the actual data.

Good software allows to view the residuals simultaneously with the fitted peak. High residuals (above 4 sigma) are an indication of a missing peak. The residuals can also reveal inaccuracies in the peak shape calibration, particularly if the peak tails are omitted for detectors which clearly have non-Gaussian peak shapes. Other visible structures, such as ''residual waves'' are an indication of incorrect peak width calibration.

6.4.1 *Background Subtraction*

Some radionuclides, such as ^{40}K, ^{232}Th, ^{235}U and ^{238}U, commonly contribute to the background of the spectrum. If these nuclides are of interest in the analysis, the constant background has to be subtracted from the calculated activity. Subtracting the two spectra is bad practice in the statistical sense because of the merging two different distributions. The subtraction therefore has to be done

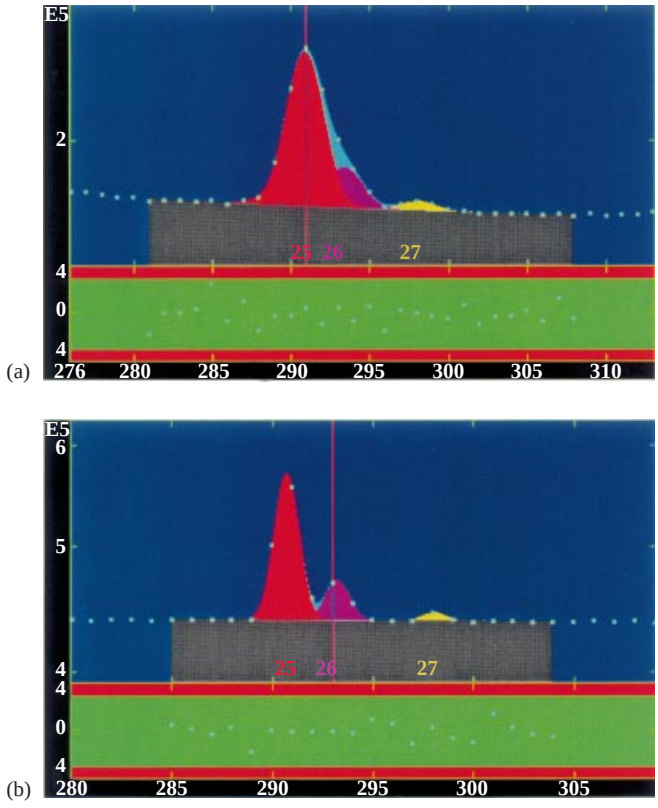

Figure 17 Deconvolution of a multiplet at 163–166 keV. (a) Part of a spectrum measured using a coaxial detector with fwhm of 1.3 keV at 165 keV. The peaks, labeled as 25, 26 and 27, refer to ^{140}Ba (162.8 keV near channel 291), ^{136}Cs (164.0 keV near channel 293) and ^{139}Ce (165.9 keV near channel 298) respectively. The deconvolution was performed in the interactive mode with SAMPO90.[19] The analysis of the multiplet is difficult because two peaks are close to each other. (b) The same multiplet in a synthetic spectrum generated for the state-of-the-art modern planar detector with fwhm of 0.7 keV at 165 keV. This performance provides means to resolve the peaks at the initial automated analysis phase. The overall reliability is greatly improved, nuclide identification is straightforward and activity calculation is more accurate. The goodness of the fit, i.e. the residual values between fitted functions and actual data in unit of standard deviations (σ), is shown in the lower window of both figures. All data points and the fitted curve are within $\pm 3\,\sigma$ at the region of interest.

after the peak area determination. The net area A_n of the peak is shown in Equation (39) as

$$A_n = A_s - \frac{T_s}{T_b} A_b \tag{39}$$

where A_s is the peak area of the γ-spectrum of interest; A_b is the peak area of the background and T_s and T_b are the corresponding live measuring times. This approach assumes that the same detector and the same electronics set-up are used both for the background and the actual sample acquisition. The corresponding error for this

reduced peak area is illustrated in Equation (40)

$$\Delta N = \sqrt{\Delta S^2 + \left(\frac{T_s}{T_b}\Delta B\right)^2} \tag{40}$$

where ΔS and ΔB are the errors of the areas in the actual measurement and in the background measurement, respectively.

6.5 Minimum Detectable Activity

The existence of a γ-peak has a probabilistic character; so the probability level at which the selected radionuclide or γ-peak exists has to be defined. In γ-spectrometry this is often calculated by the method of Currie,[22] as explained below. This method is based on statistical considerations only, and calculations use strict statistical rules applied to the properties of the measured spectrum. The analyst selects a probability level which all the peaks have to fulfil for reliable detection. The resulting minimum detectable activity (MDA) is used to discriminate against calculated activities less than this MDA value.

Consider a hypothesis that a peak is observed at a certain place in the spectrum. Let α be the probability that this observation is spurious. Then $1 - \alpha$ is the probability that this observation is correct. The critical limit or the decision limit is the area of the γ-peak for its detection with a given probability level. It is defined as illustrated in Equation (41)

$$L_C = k_\alpha \sigma_0 \tag{41}$$

where k_α is the abscissa of the standardized normal distribution at the selected probability level $(1 - \alpha)$, and σ_0 is the standard deviation of the net signal when its expected value equals zero. Consider now that this observation procedure is already performed. Let β be the probability that a peak is not seen at the selected energy, but there still is a peak. When a nuclide detection is warranted at the selected probability level, the nuclide activity has to exceed the detection limit, shown in Equation (42), of

$$L_D = L_C + k_\beta \sigma_D \tag{42}$$

where k_β is the abscissa of the standardized normal distribution at the selected probability level $(1 - \beta)$, and σ_D is the standard deviation associated to this limit. The variance of the net signal illustrated in Equation (43), is

$$\sigma^2 = \sigma_{S+B}^2 + \sigma_B^2 \tag{43}$$

where σ_{S+B} is the standard deviation of the gross signal and σ_B is the adjacent standard deviation of the baseline. When there is only background present, the expected value of the signal is zero, giving in Equation (44):

$$\sigma_0^2 = 2\sigma_B^2, \sigma_B = \sqrt{B} \tag{44}$$

and Equation (45)

$$L_D = L_C + k_\beta \sqrt{L_D + \sigma_0^2} \tag{45}$$

Finally, L_D can be solved using Equations (41), (42) and (45) as shown in Equation (46):

$$L_D = L_C + \frac{k_\beta^2}{2}\left(1 + \sqrt{1 + \frac{4}{k_\beta^2}\left(L_C + \left(\frac{L_C}{k_\alpha}\right)^2\right)}\right) \tag{46}$$

With the selected probability level of 95% for both β and α the detection limit can be calculated. In this case k_α and k_β are 1.65. If there are no interfering peaks at the same energy range, the decision limit L_C, illustrated in Equation (47), is:

$$L_C = k_\alpha \sqrt{2}\sigma_B = k_\alpha \sqrt{2B} = 1.65\sqrt{2B} \approx 2.3\sqrt{B} \tag{47}$$

At 95% confidence level ($k_\alpha = k_\beta = k = 1.65$) the detection limit L_D, illustrated in Equation (48), is

$$L_D = k^2 + 2L_C = 2.71 + 4.65\sqrt{B} \tag{48}$$

At 99% confidence level ($k = 2.33$) the detection limit, illustrated in Equation (49), is

$$L_D = 5.42 + 6.59\sqrt{B} \tag{49}$$

If there is an interfering γ-peak, the standard deviation is as shown in Equation (50)

$$\sigma_0 = \sqrt{2B + \left(\frac{T_s}{T_b}\right)\left(A_b + \left(\frac{T_s}{T_b}\right)^2 A_b\right)} \tag{50}$$

where A_b is the area of the background peak (obtained using long-term background data).

The sampling and the sample preparation process also have an influence on the detection limit. Some of the nuclides may be lost or removed before counting. In air sampling, for example, the particle collection efficiency may be significantly below 100%. However, if the sampling is optimized, i.e. the system has short tubes without any bends, and the filter material is of good quality, the losses of small particles (<10 μm) are negligible.

6.5.1 Baseline Width Selection for Minimum Detectable Activity Calculations

The baseline definition in the original Currie formulas were intended for a single channel analyzer.[22] Therefore, the method expects the baseline to cover exactly the

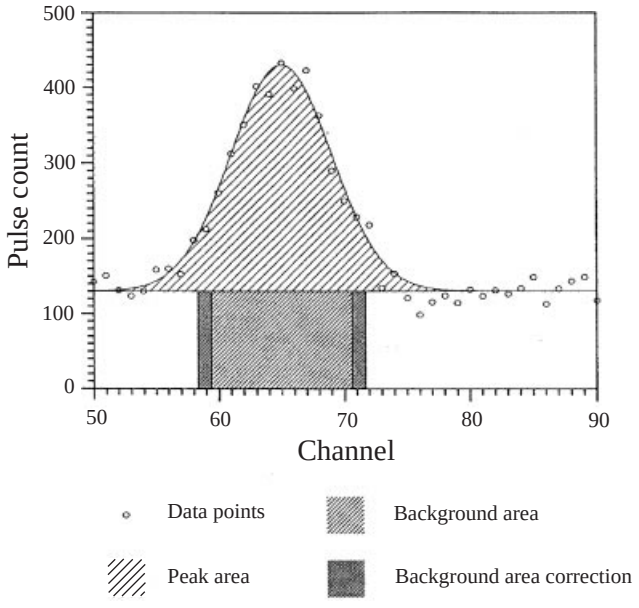

Figure 18 Gamma-peak and its baseline for the calculation of the MDA.[17] For this peak the fwhm is about nine channels. The background area (±0.59 fwhm) is multiplied by 1.193 (background area correction) to compensate for the fact that only 83% of the peak area is within the ±0.59 fwhm area region. (Reproduced by permission of T.T. Hakulinen.[17])

counts that are "below" the peak. Broadening the baseline width too much increases MDA to unrealistic values.

The optimum baseline width below the γ-peak is $\pm 1.4\sigma$, as derived analytically by optimizing the signal-to-noise ratio.[17] Only 83% of the peak area is included within $\pm 1.4\sigma$ (assuming Gaussian peak). Therefore, the baseline should be extended by a factor of 1.193 to cover the whole peak baseline area (Figure 18). As a result, the optimum and justified baseline width is ±0.71 fwhm (Equation 18).

For a typical environmental sample, MDA in a γ-spectrum is about 10 mBq for nuclides that emit photons of high abundance, such as ^{137}Cs. An example of the results of the MDA calculations is given in Table 6 for an air sample. The selection of the counting time is of crucial importance for the optimum detection capability. In principle, several spectra should be saved during the data acquisition for optimizing the MDA for different nuclides with different half-lives.

The surety of peak existence can be categorized according to the peak significance, which is defined as the peak area divided by L_D according to the definition in Equation (48) with probability level of 95%. An uncertain or "gray" zone can be defined as $(0.5–1.4) \times L_D$. A found peak can be regarded as a false peak (type I error), if its area is below the lower limit, which in fact is L_C, as defined in Equation (47) at the probability level of 95%. The peak can be regarded as a true peak if its area is above the upper limit, which in fact is L_D at 99% confidence level, as defined in Equation (49). The analysis of the gray peaks, particularly if they are associated to radionuclides, requires special attention from the analyst. The gray zone may have to be enlarged or diminished according to the application and software; the areas of small peaks and L_D vary according to the algorithm used.

6.6 Radionuclide Identification and Activity Calculations

After the peak areas have been calculated, and their emission rates are determined using the efficiency calibration, it is time to identify the radionuclides contributing to the γ-spectrum. This involves qualitatively deciding which nuclides are present and thereafter quantitatively determining their activities. All identification procedures start by matching the observed peak energies with energies of photon transitions by known radionuclides. This gives a preliminary set of candidate nuclides, among which one hopefully has a subset forming the correct solution. There are three basic approaches for resolving the preliminary candidate set:

1. Calculate the activity of each candidate nuclide based on the γ-peak of largest abundance that is not interfering with other candidate nuclides. In this case the analyst has to select the appropriate γ-lines specific to the radionuclides of interest.
2. Form a matrix of all candidate nuclides and spectrum peaks with branching ratios as coupling constants and minimize it in the least-squares sense. This yields activity estimates for all included candidate nuclides. Normally precalculation discrimination of candidates is required, which can be based on simple rules taking into account the energy tolerances, other γ-lines of the radionuclides and the decay periods.[18,19,24]
3. Use an expert system approach using a comprehensive set of both heuristic rules and objective quantitative information on the measurement and radionuclides involved to reduce the set of candidate nuclides as much as possible.[17,25–27] Afterwards, calculate the activities using the methods above.

Parent–daughter relationships have to be taken into account for radionuclides from the same decay chain. Otherwise, short-lived nuclides produced by longer living parents would be erroneously discarded or the activity calculation would be incorrect.

Least-squares minimization of the candidate matrix is used to solve the problem of interfering radionuclides (Table 7). In the candidate matrix the associated nuclides with unknown activities are in one dimension and the

Table 6 Typical detection limits in air sampling for a high-volume sampler.[23] MDA is calculated from the Currie equation (Equation 48) using baseline width of 1.42 times the fwhm.[a] MDA (Bq) is converted to MDC (Bqm^{-3}) using the total amount of air ($22500\,m^3$) that passed through the filter during the 24 hour sampling period. The sample was counted in Williams geometry (Figure 10a) using a 40% p-type HPGe

Nuclide	Energy (keV)	Half-life	MDC (μBqm^{-3})			
			No decay 4 h counting	4 h decay 20 h counting	24 h decay 24 h counting	72 h decay 72 h counting
Co-60	1173.2	5.27a	14	1.3	1.0	0.51
Co-60	1332.5	5.27a	15	1.3	0.8	0.46
Zr-95	724.2	64.4d	26	2.3	1.8	0.9
Zr-95	756.7	64.4d	21	1.8	1.4	0.72
Nb-95	765.8	35.2d	12	1.1	0.8	0.47
Zr-97	743.4	16.8h	23	2.9	5.0	45
Mo-99	739.5	2.76d	98	9	9	11
Mo-99/Tc-99m	140.5	2.76d	7.0	0.9	0.74	0.8
Tc-99m	140.5	6.02h	31	11	83	57 000
Ru-103	497.1	39.4d	11	1.1	0.8	0.41
Ru-106	621.8	368d	105	10	7.0	3.5
I-131	364.5	8.04d	13	1.4	1.0	0.64
Te-132	228.2	3.25d	10	1.4	1.1	1.1
I-133	529.5	20.3h	19	2.7	3.8	20
Cs-134	604.7	2.06a	11	1.0	0.71	0.38
Cs-136	818.5	13.0d	13	1.1	0.8	0.50
Cs-137	661.6	30.1a	12	1.2	0.8	0.43
Ba-140	537.4	12.8d	53	5.3	3.6	2.2
Ce-141	145.5	32.4d	12	1.4	0.9	0.50
Ce-143	293.3	1.40d	26	3.8	4.4	11
Ce-144	133.5	284d	49	5.8	3.7	1.9
U-237	101.1	6.75d	22	2.9	2.0	1.4
U-237	207.9	6.75d	35	4.3	3.2	2.2
Np-239	103.7	2.36d	26	3.8	3.1	3.8

[a] The original calculations were performed using baseline width of 2.5 fwhm.

Table 7 Activity calculation for interfering radionuclides by least-squares minimization. Some peaks in the spectrum may be associated to several radionuclides but most of these nuclides have multiple γ-lines. Therefore, the spectrum can have ''subsets'' of peaks which have correlation with other peaks. The fraction each nuclide is explaining of each peak has correlation with other nuclides. Labels X in the present hypothetical example refer to the radionuclides contributing to the peaks involved. Numerical least-squares minimization is the most practical way to solve this problem, i.e. the computer finds the optimum shares for each radionuclide. Modern software packages are able to perform this calculation[17,24,25]

	100 keV	220 keV	1020 keV	2540 keV	2780 keV
Nuclide 1	X		X		
Nuclide 2		X			X
Nuclide 3		X	X		
Nuclide 4		X		X	X

emission rates of the found photo-peaks are in the other dimension; the branching ratios of the respective nuclide transitions are the coupling constants. The coupling matrix and the emission rate vector are weighted by the errors of the peak emission rates.

The construction of the candidate matrix is straightforward, but calculating the matrix inversion is not always a simple task.[17] The matrix can be underdetermined, overdetermined, or even completely singular. Therefore strong numerical methods, such as singular value decomposition (SVD), have to be used.[28]

Correction for the radionuclide decay has to be included in the activity calculations for all the phases of the sample preparation and acquisition. The correction term for the decay during counting is illustrated in Equation (51):

$$C_C = \frac{\lambda t_c}{1 - e^{-\lambda t_c}} \tag{51}$$

where t_c is counting time and λ is the decay constant.

The correction term for the waiting before the data acquisition, i.e. the decay correction during the waiting time t_w, is illustrated in Equation (52):

$$C_w = e^{+\lambda t_w} \tag{52}$$

During the sampling time t_s some of the radionuclides in the sample have decayed. The correction term is

illustrated in Equation (53):

$$C_S = \frac{\lambda t_s}{1 - e^{-\lambda t_s}} \quad (53)$$

For the air filter sample, the average activity concentration a (Bq m^{-3}) of a nuclide is shown in Equation (54)

$$a = \frac{\lambda t_s A_w}{(1 - e^{-\lambda t_s}) V_{tot}} \quad (54)$$

where V_{tot} is the total volume of air and A_w is the activity at the end of the sampling.

Self-absorption correction is essential if the dimensions and the density of the samples vary. The photons are absorbed in the sample matrix in different ways. This uncertainty is a source of fluctuations in the γ-peak areas. If the efficiency calibration is based on the computational methods, an individual efficiency curve can be calculated for each sample according to the dimensions and the density of the sample.[14,15] The self-absorption correction is of particular importance in the applications where accurate nuclide activity ratios are required.

Background matrix correction is similar to the self-absorption correction but carried out for true background nuclides present in the structures of the measurement room. An approximate correction function can be estimated a priori,[17] or more accurately from the background measurements.

Coincidence summing is an important phenomenon at high-count rates or close geometries. True coincidence summing occurs when two or more γ-transitions in a decay scheme are in cascade and the half-lives of the intermediate states are shorter than the time resolution of the detector. Therefore, the cascade photons seem to create too small peaks, and furthermore, a coincidence peak is formed at the energy equal to the sum of the component energies. In some cases, when the intermediate half-lives are favorable, both photons can be rejected by the PUR or active Compton suppression hardware.

The calculation of the coincidence correction is straightforward.[6,26,29] Let us consider a simple decay scheme with two photons in cascade (Figure 19). If no photon summing would occur, the ideal peak count rate S_{21}^* for the transition from level two to level one would be affected only by the nuclide activity, the peak efficiency of the detector and the branching ratios (x_{21} and N_2). However, because of photon summing, the observed count rate S_{21} is as illustrated in Equation (55)

$$S_{21} = S_{21}^*(1 - e_{t,10}) \quad (55)$$

where $e_{t,10}$ is the total efficiency of the detector in the counting geometry involved for the transition from level

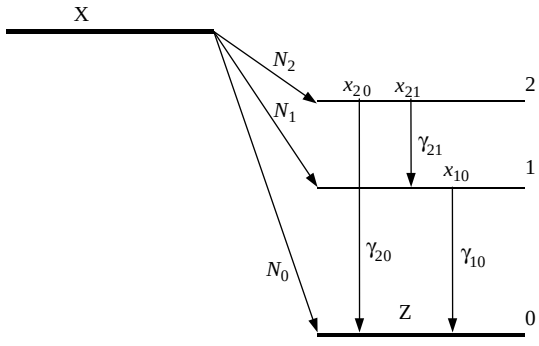

Figure 19 Principle of coincidence correction. The nuclide X decays to ground state and to two short-lived isomeric states of nuclide Z. N_i is the fraction (percentage) of direct decays to energy level i and x_{ij} is the branching ratio for the transition from level i to level j. An example of this type of decay scheme is ^{60}Co and its daughter ^{60}Ni. Level 2 has two different transitions: directly to ground state or to level 1 which decays immediately to the ground state. Sometimes, particularly in close-counting geometries, the two photons in cascade, γ_{21} and γ_{10}, deposit their energies, or part of their energies, simultaneously in the detector. Therefore, the true peak count rate of γ_{20} is increased compared with the prediction of the peak efficiency of the detector, whereas the count rate of the other two photons γ_{21} and γ_{10} is decreased. This true summing effect must be corrected in the activity calculation.

one to level zero, i.e. $e_{t,10}$ is the overall probability that this photon is counted. The coincidence summing correction is defined in Equation (56) as

$$C_{21} = \frac{S_{21}^*}{S_{21}} = \frac{1}{1 - e_{t,10}} \quad (56)$$

Similar but more complex correction factors can be derived for other transitions and for different decay schemes.[6,26,29] The correction factors vary between 0.4–2.3 and they depend on the source–detector distance.[26] At distances larger than 15 cm, no correction is needed.

The theory of coincidence correction is well known. Therefore, it is astonishing that only very few commercial software packages are able to treat coincidence summing effects.

Random summing occurs when the count rate is sufficiently high for two photons to interact simultaneously with the detector. Random summing is not a problem in low-level counting.

6.6.1 Unidentified Peaks

If a spectrum contains unidentified peaks after a routine analysis, the following procedures could be implemented:

1. Check if the peak is a single or double escape peak (511 keV or 1022 keV below a large γ-peak above 1022 keV).

2. Check if the peak is a result from random summing by two large peaks at low energies.
3. Check if the peak comes from a nuclide which is in parent–daughter relationship with nuclides already identified.
4. Check if the peak could be associated to cosmic radiation (139.7 keV of ^{75m}Ge for example), or to other external origin.
5. Check if the peak could be associated to X-ray escape (10 keV below a large photo-peak).
6. Check other nuclide candidates at the same energy. Start the procedure from the highest unidentified peak. The energy look-up tolerance could be widened to account for possible errors in the energy calibration.

6.6.2 Rule-based Nuclide Identification

Sophisticated expert systems, such as Shaman,[17,25–27] have several advantages compared with classical, algorithmic identification procedures:

1. Comprehensive radionuclide library. No tailoring of the library is necessary because the expert system can use a variety of heuristic rules to reduce the number of candidate nuclides during identification. This makes it possible to use a library of all known radionuclides and transitions (80 000 γ- and X-ray lines), which in principle gives the system the ability to correctly identify any sample composition. Classical identification algorithms get into trouble with spurious identifications if the library is too large (more than 1000 γ-lines).
2. Rule-driven identification process rather than predetermined algorithm. Management of knowledge and the handling of various complex phenomena is straightforward, including
   (a) decay chain handling and various activity corrections;
   (b) nuclide-specific rules taking into account production modes of nuclides;
   (c) handling of escape and coincidence sum peaks;
   (d) rules specific to the MDA calculations;
   (e) peak-specific rules based on MDA, peak width, and peak significance, for example, and
   (f) sample and measurement-specific rules for including the analyst's specific knowledge of the measurement (nuclides that can or cannot be present, etc.).
3. Versatile and yet conservative rule base. By correctly setting the expert system parameters, the system can be tuned to the identification problem precisely, which has not generally been possible for traditional identification methods.

The identification result by an expert system is excellent because of advantages of the combined heuristic rules and the complete nuclide library. In a typical γ-spectrum, 99% of the γ-peaks can be identified routinely without any human interference.[27] Combination of rule-based identification, robust activity calculation and high-quality quantitative analysis with strong multiplet resolving capability results in the competent γ-spectrum analysis package.

7 IN SITU γ-SPECTROMETRY

7.1 Principles of In Situ Measurement

γ-Spectrometry performed in the environment is a rapid method for detecting and quantifying radioactivity in the air and in the ground. Large areas can be surveyed rapidly after a nuclear accident using portable equipment in vehicles, helicopters or airplanes. Combining of concentration or deposition results with satellite navigation data provides a powerful tool for the assessment of radiation hazard.

The term in situ γ-spectrometry has been adopted for measurements that are performed close to the source in the environment. This technique has been used since the 1960s for uranium exploration and environmental monitoring. The reactor incident at Three Mile Island in 1979, and the Chernobyl accident in 1986, convinced radiation safety authorities and other organizations to adopt quick and reliable methods for the determination of radionuclides, and in situ γ-spectrometry is now used widely for estimating levels of environmental contamination.

Besides providing prompt results on radiation hazard at the early stage of a nuclear accident, the in situ method reduces data variability, because soil sampling with later laboratory measurement is prone to nonrelevant spatial source variations. The field of view is particularly large in airborne measurements. In fall-out mapping, the depth distribution of the radionuclides may be unknown. Various assumptions have to be made which, in turn, are the main source of uncertainty in the result. Therefore, some ground or air sampling should be performed to verify the in situ results.

The theory of in situ γ-spectrometry was developed by Beck et al. in 1972.[30] The basic problem is to determine the ratio between the full energy peak count rate N_f and the source term S_x, as illustrated in Equation (57):

$$\frac{N_f}{S_x} = \frac{N_0}{\varphi}\frac{N_f}{N_0}\frac{\varphi}{S_x} \tag{57}$$

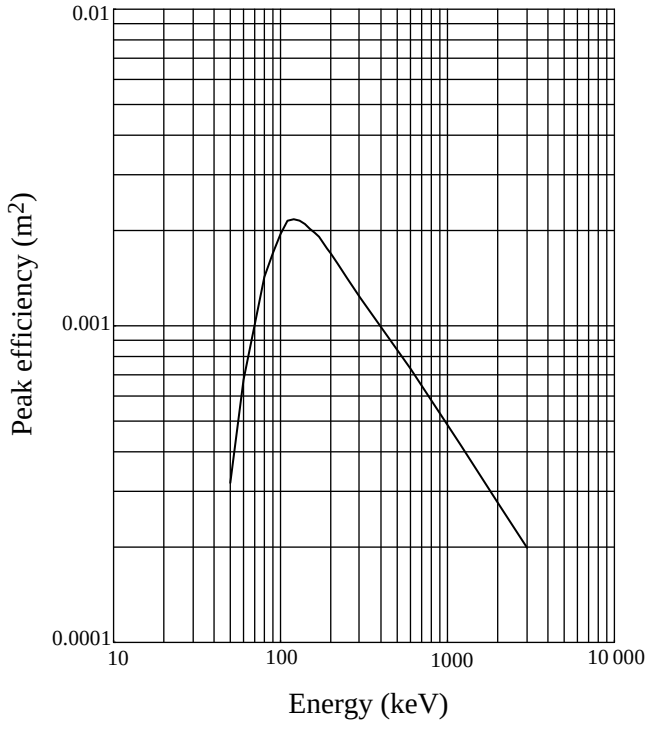

Figure 20 Full-energy peak count rate per unit fluence rate, N_0/φ, for photons arriving from the axial direction of the detector ($L = 59.3$ mm, $D = 59.7$ mm, 37% HPGe).[20]

where N_0/φ is the photo-peak efficiency (Figure 20), i.e. the full-energy peak count rate (s^{-1}) per unit fluence rate ($s^{-1}m^{-2}$), for photons arriving from the axial direction of the detector ($\theta = 0$). N_f/N_0 is the angular correction factor that takes into account that the fluence is not normal to the detector face. φ/S_x is the ratio between the fluence rate and activity concentration, expressed as activity per unit area S_a ($Bq\,m^{-2}$), activity per unit mass S_m ($Bq\,kg^{-1}$) or activity per unit volume S_v ($Bq\,m^{-3}$). φ/S_x does not depend on the properties of the detector and can be calculated for various geometries.[20,31,32]

7.2 Ground-level γ-Spectrometry

7.2.1 Radionuclides in Terrain

The term N_f/N_0 in Equation (57) depends on the detector characteristics and the source geometry. It is computed as a weighted average of the normalized full-energy peak count rate per unit fluence rate as a function of angle, $N_f(\theta)$, over the fluence rate distribution.[30] The sensitivity of the detector is maximized if it faces downward ($\theta = 0$). Sometimes this might be impractical, particularly in mobile measurements.[33] The arrangement of detector facing upward with the dewar underneath reduces peak count rates because the dewar attenuates photon fluence from the ground underneath. Because the angular distribution of the fluence is peaked toward the horizontal direction, however, the total efficiency reduction is only a few percent for surface distribution of activity, and 10–20% for uniform source distribution.[34]

If the detector chosen for the in situ measurements has equal length and diameter ($L/D = 1$), it may be possible to neglect the angular correction factor. In this case the detector should be calibrated using photon fluence from the side-wall direction ($\theta = 90^\circ$). Angular correction is necessary for a detector that is installed in a complex surrounding, for example inside a vehicle.

The source distribution has to be known or a reasonable approximation has to be used for the determination of the radionuclides in the ground. The radionuclides are usually assumed to have homogeneous concentration in the horizontal plane, but the depth distribution can vary. The natural radionuclides, ^{40}K and the progenies of ^{238}U and ^{232}Th, often have a uniform concentration, whereas man-made radionuclides may have an exponential concentration profile. Fresh fallout is a thin layer on top of the ground. The uncertainty analysis of the in situ γ-spectrometry has shown that most of the error of the method is related to the depth distribution of the radionuclides in the soil rather than to detector characteristics, counting statistics or other measuring parameters.[35]

To compute φ/S_a, an exponential model for the source depth profile is widely used.[31] This method has only one parameter, the relaxation mass per unit area (β) for describing the real nuclide distribution. In recent fallout β varies between 0.1 and 3 $g\,cm^{-2}$, but it can be up to 20 $g\,cm^{-2}$ for aged events (>10 years). Fluence rates in air at a height of 1 m above ground are available for natural radionuclides distributed homogeneously and for man-made nuclides as a function of β.[31]This information together with the calibration data allows calculating the peak response of the detector (s^{-1} per $kBq\,m^{-2}$) for various nuclides (Figure 21).

The MDA per unit area is almost similar for radionuclides with equal decay yield. The weaker attenuation of high-energy photons is compensated by a lower peak response of the detector (see Figure 20). ^{131}I or ^{137}Cs contamination below 0.1 $kBq\,m^{-2}$ can be detected with a counting time of only 30 min (22% HPGe).[31]

7.2.2 Radionuclides in Air

In situ γ-spectrometry of airborne radionuclides provides many advantages that are most useful at the early stages of a nuclear accident. The results are immediately available and no sampling is needed. Detection of Xe isotopes is part of the measuring procedure; it is not a separate complex task (noble gases are the major source of external exposure in some accident scenarios).

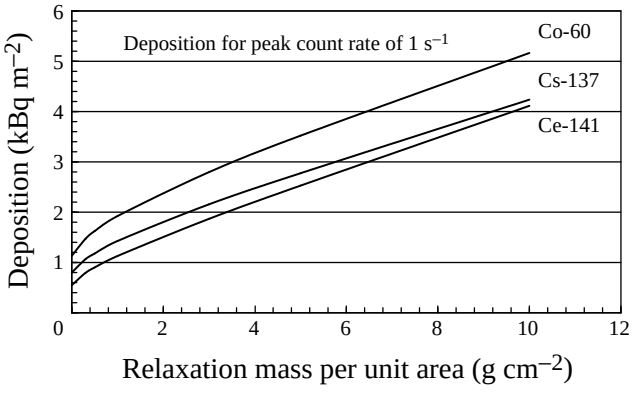

Figure 21 Deposition which gives a peak count rate of $1\,s^{-1}$ as function of source depth distribution for detector at height 1 m above ground. No angular correction; for HPGe detector characteristics, see Figure 20.

In a nuclear emergency, the count rate may be so high that a counting period of a few tens of seconds can give adequate results. Data acquisition is easy to perform in short consecutive sequences; spectrum analysis and report generation can also be automated and performed in a few seconds.

Radionuclide concentration in air is a complex dynamic quantity that may change in short intervals. Far away from the source, the plume is usually wide enough to assume a simple geometry. The computation of φ/S_v is a straightforward task for an infinite slab plume.[20,31]

The airborne contamination can be monitored with a detector that has no shielding. This approach is practical when the goal is to detect the presence of airborne man-made activity rather than to quantify the concentrations. Often the airborne activity produces in a few hours a deposition that generates a larger fluence rate than the cloud itself. For quantitative concentration estimates, the side walls of the detector should be shielded. The accuracy is further improved if the shielding has a lid for background measurements.

The isotopes of iodine, ^{131}I, ^{132}I, ^{133}I, are the major inhalation hazard in a nuclear reactor accident. The inhalation exposure of the population may require countermeasures at a concentration level above $1000\,Bq\,m^{-3}$. The in situ γ-spectrometry can detect contamination as low as $1–10\,Bq\,m^{-3}$ in 10–30 min.[20,31]

7.3 Airborne Spectrometry

7.3.1 Ground Survey

Gamma-rays penetrate up to several hundred meters in air. Airborne survey of the environment is possible at an the altitude of 100 m (or even 300 m), although a lower altitude of 30–60 m is more suitable for detailed radionuclide mapping (Table 8). The equipment must have the capability to measure at short intervals. In some applications, such as geological survey and search of lost or stolen radiation sources, the spectra acquisition time is about 1 s, whereas to reduce data volume for fallout mapping a longer counting period of 10–60 s would be more appropriate.

Table 8 Photon fluence rate ($m^{-2}\,s^{-1}$) in air per unit fallout ($m^{-2}\,s^{-1}$) for fresh debris ($\beta = 0$) as a function of altitude[31]

Energy (keV)	Height (m)		
	1	50	100
100	1.72	0.130	0.0324
500	2.01	0.268	0.102
1000	2.15	0.365	0.165

Low-resolution γ-spectrometry with large NaI crystals (volume >10 L) has a high sensitivity. The count rate from a 25 L NaI array is typically 50 times larger than from a 100% HPGe detector (volume ratio). However, the superior energy resolution of the HPGe detector provides reliable peak quantification and unambiguous nuclide identification for complex spectra. Therefore, mobile high-resolution γ-spectrometry is the only tool that can produce reliable nuclide-specific fallout maps immediately after a nuclear accident. Low-resolution spectrometry is useful at the later phases of the accident when most of the short-lived nuclides have decayed or in accident scenarios where only one or a few nuclides are involved.

The mobile measurements are much more difficult to perform than the laboratory measurements. The monitoring equipment has to be operated in harsh conditions where factors such as temperature or vibrations may interfere. Furthermore, the coordinates have to be recorded and attached to the measured spectra. The amount of data produced can be as large 100 Mbytes per hour. Special software is needed to analyze the data and to produce nuclide-specific maps on the area searched.[36]

Since the Chernobyl accident, several countries, particularly in Europe, have created mobile units that can perform the data acquisition and analysis promptly.[37] The detection limits of airborne surveys are far below safety limits applied for the protection of the public (Table 9). Recent developments in detector manufacture have provided HPGe detectors with a large diameter (>80 mm). These detectors, which are planar rather than coaxial (section 4.1), reduce detection limits drastically, particularly for low-energy nuclides. Such a mobile system is very powerful in safeguard measurements, either in national or in international applications.

Table 9 Detection limits for an airborne measuring system installed in a helicopter (70% p-type HPGe detector, integration time 15 s, flight altitude 60 m). The MDA is calculated using the Currie equation (Equation 48)[22,36]

Nuclide	Energy (keV)	MDA (kBq m^{-2})
^{241}Am	59.5	25
^{57}Co	122.4	3.3
^{137}Cs	661.6	1.8
^{60}Co	1332.5	0.65

7.3.2 Tracking of a Radioactive Plume

Determination of nuclide-specific concentrations within a plume is difficult using in situ γ-spectrometry. The concentration pattern is a complex three-dimensional distribution which changes continuously with meteorological conditions. However, in situ spectrometry is useful for locating the radioactive plume and for identifying the release nuclides. Detection of the radioactive cloud before the nuclides deposit to the ground gives more time for possible countermeasures.

Two different scenarios should be addressed. The plume may have clear boundaries because the source is near, or the plume has traveled a long distance and has dispersed in the atmosphere over a large frontier. A helicopter is useful for the close source studies, whereas a fixed-wing aircraft is preferable for a large area survey. The detection system could have an HPGe detector and two NaI crystals separated by lead shielding. The NaI-detector system provides information on whether the helicopter or the aircraft is over the plume, below the plume, or within the plume. Besides plume detection, it is possible to determine the axis, the boundaries and the height of the radioactive cloud.

A rough quantitative estimate of the release may be possible.[38] However, more realistic goals for the flying team are plume detection, key nuclide identification (iodine) and sampling for later laboratory analyses.

8 QUALITY ASSURANCE SYSTEM

A quality assurance system is intended to guarantee the high quality of the products and services. Repeated measurements and analyses must produce the same, correct and reliable results. Anomalous results must be recognized in all circumstances. Furthermore, the means to control the whole chain of custody of the samples provide confidence in the overall analytical capability.

Any laboratory seeking external contracts will be expected to have appropriate accreditation. Comprehensive guidance to the accreditation process is provided by the national accreditation bodies, such as the National Institute of Standards and Technology (NIST) in the USA or the National Measurement Accreditation Service (NAMAS) in the UK. The process itself may take a long time, including development of quality requirements, their implementation, preparation of quality documentation, training, and internal and external audits. Although a laboratory would not seek any formal certificate, it is worth setting up a tailored quality assurance system applying international standards, such as ISO 9001 or ISO 17025.[39,40]

8.1 Standards and Procedures

The quality assurance system sets requirements for the management and the technology. The international and national standards are an excellent support for developing the requirements and the procedures. The management requirements[40] address the following points:

- organization
- structure of the quality system
- document control
- request, tender, and contract review
- subcontracting services
- purchasing services and supplies
- services to the clients
- complaints by the clients
- control of nonconforming work
- corrective action
- preventive action
- records
- internal audits and
- management reviews.

The technical requirements and procedures are intended to verify that the instruments are working properly, nuclear data are valid, standards are fit for the purpose and are traceable to international standards, and the spectrum analysis is correct. High quality is achieved if the measurement and analysis are performed according to validated and well-documented standard procedures. Any operational change must be properly validated before adoption to routine use.

The same set of most recent and validated nuclear data should be used in all operations of the laboratory. Changes must be documented. The International Atomic Energy Agency (IAEA) provides a small well-evaluated data set that is useful for energy and efficiency calibrations.[41] The National Nuclear Data Center at Brookhaven National Laboratory, USA, maintains an Evaluated Nuclear Structure Data File (ENSDF). Appendix 1 is a subset of this file for radionuclides that are of importance in environmental γ-spectrometry.

High-quality data is also available through the Internet (see, for example, http://www.fysik.lu.se/nucleardata/).

The accuracy of the measurements depends on the quality of the calibration standards. The calibration sources should be purchased from accredited laboratories that can demonstrate traceability to a national standard. In the past, calibrations were performed, and are often still performed, in the geometry that mimics the sample dimensions and density. This kind of work requires calibrated solutions, is difficult to carry out (dilutions), is prone to contamination, and provides calibration only to one fixed type of sample. A far better way is a semiempirical approach based on standard point sources (section 6.2).

Each spectrometer must have a log book. All essential events, such as equipment fine-tuning or adjustment, must be documented. A maintenance program is intended for checking calibrations on a weekly or even on a daily basis. Energy and shape calibrations should be verified regularly, although they can be calculated from the data itself if required (Table 10). Checking of the efficiency calibration is of vital importance because there is no way to verify this from the data. A procedure should be established to count and analyze a mixture of sources in suitable intervals, such as ^{57}Co (122 keV) and ^{60}Co (1332 keV), for a short period, for example 10 min.

The laboratory must record the receipt of samples, the analysis process and the results. The analyst verifies that documented procedures have been followed. The essential information about the analysis must be documented in the laboratory notebook. The complete analysis process should be possible to check at any time. This goal is made easier to reach by software that saves the complete measurement and analysis configuration.

8.2 Technical Quality Requirements

Quality criteria and metrics play a key role for the quantitative evaluation of the measurement and analysis. Metrics should be applied to daily operations, performance evaluation on regular intervals and to software validation.

Evaluation metrics and their implementation are an important element of the quality assurance system. The definition of the metrics is followed by the development of implementation procedures and software for producing the reports required. The proposed criteria and procedures should not be too expensive, nor too difficult to achieve technically.

The following issues are of importance for the metrics development:

- requirements
- applicability of the metrics
- measurable and nonmeasurable aspects
- sources of data
- success criteria
- distribution of tasks
- frequency of evaluation and
- procedures for evaluation.

Control charts are perhaps the simplest technical quality elements. A control chart refers to a plot of the quantity involved as a function of time, including upper and lower warning and action limits (confidence limits are set often to 95% and 99.7%). This procedure is intended to monitor changes over time. Table 11 provides an example of subjects that could be evaluated regularly either through a control chart or using more advanced metrics.

8.3 Software Validation

The quality of software can be assessed through validation tests and developing evaluation metrics for the

Table 10 An example of a quality assurance report generated during the analysis of a γ-spectrum. Energy and shape calibration were recalculated from the measured spectrum itself. The following peaks, which are present in the sample spectrum, were used to assure the energy calibration: 238.63 keV (^{212}Pb), 477.61 keV (^{7}Be), 583.19 keV (^{208}Tl), 860.56 keV (^{212}Pb), 1460.80 keV (^{40}K). The report shows that the original calibration data provided by the laboratory are out of range at large energies (cf. requirements in Table 11). The ratio fwtm/fwhm tends to increase as a function of energy indicating charge collection problems at large energies; this deficiency is manifested as additional tails in the Gaussian peak (see Figure 6). In this case the original calibration data should not be used for the analysis and the software must have capability to cope with peak tails

Energy (calib.) keV	Energy (recalc.) keV	Energy difference keV	fwhm (calib.) keV	fwhm (recalc.) keV	fwhm difference keV	fwtm keV	fwtm/fwhm
238.57	238.63	0.06	0.94	0.95	0.01	1.74	1.82
477.77	477.61	−0.16	1.23	1.18	−0.05	2.21	1.87
510.94	510.75	−0.19	1.27	1.52	0.25A	3.48	2.29
583.47	583.19	−0.28	1.33	1.29	−0.04	2.37	1.83
1462.06	1460.80	−1.26!	1.93	1.91	0.01	3.93	2.05
2617.81	2614.64	−3.18!	2.47	2.31	−0.15!	5.20	2.25

A, annihilation peak,
!, warning, calibration is out of range.

Table 11 Evaluation metrics for γ-spectrometry. Whenever possible, the subject should be described in a control chart containing upper and lower warning limits and action limits. Further quantitative requirements must be developed according to the specific needs of the application

Subject	Measurement	Requirement
Sampling	Specific to the type of sample; sampling time, minimum volume etc. Proportion of requirements met	Set appropriate figures
Decay time	Proportion of requirement met	Set limits for minimum and maximum duration
Measurement time	Proportion of requirement met	Set limit for minimum counting time
Time before reporting	Proportion of requirement met	Set limit for the reporting delay
Energy calibration verification	Distribution of the maximum difference in energy between sample data and calibration data; mean and SD	<0.5 keV for all channels
Shape calibration verification at the major peak (specific to the sample)	Proportion of the requirements met (comparison with the calibration value) fwhm distribution of the major peak;[a] mean and SD	$\mathrm{fwhm_{meas}} - \mathrm{fwhm_{cal}} < 0.1\,\mathrm{keV}$
Peak tail verification for all peaks with area > 2000 counts	Distribution of fwtm/fwhm for the major peak; mean and SD Distribution of fwtm/fwhm for all peaks; mean and SD	Mean < 1.9
Efficiency calibration verification	Areas of the large peaks (>2000 counts) in check source counting	
Baseline sensitivity (MDA) for the nuclide of interest;	Proportion of requirements met	Develop nuclide-specific requirement according to the sample and operational procedures (decay time counting time)
Detection capability at different background	Plot MDA as a function of the activity of the principal natural nuclide, such as ^{212}Pb or ^{40}K For air sampling: Plot MDC as a function of ^{212}Pb concentration and perform curve fitting to $\sqrt{a + b^{212}\mathrm{Pb}}$; Calculate MDC from this curve for ^{212}Pb concentrations of (a) $10\,\mathrm{mBq\,m^{-3}}$ (small background); rank the stations (b) $100\,\mathrm{mBq\,m^{-3}}$ (large background)	
Data availability (samples arrived and analyzed/expected figure)	Availability percentage	>95%
Breakdown of equipment	Distribution of repair time; mean and SD	
Peak search	Distribution of number of peaks found per spectrum for similar sample types; mean and SD	
Peak fitting (if free-width algorithm is used)	Distribution of $\mathrm{fwhm/fwhm_{cal}}$; mean and SD ($\mathrm{fwhm_{cal}}$ refers to the calibration curve)	
Peak identification in automated mode	Proportion of peaks identified;[b] mean and SD	
Nuclide identification in automated mode	Proportion of nuclides identified;[c] mean and SD	
False peaks in automated mode	Distribution of number of peaks removed manually (type I errors) per spectrum; mean and SD Distribution of true peaks found per spectrum; mean and SD (type I error correction)	
Missed peaks in automated mode	Distribution of number of peaks added manually (type II errors) per spectrum; mean and SD	
False nuclides in automated mode	Distribution of number of nuclides removed manually (type I errors) per spectrum; mean and SD	
Missed nuclides in automated mode	Distribution of number of nuclides added manually (type II errors) per spectrum; mean and SD	

[a] Known to be present in the spectrum, such as ^{7}Be in air filter samples.
[b] Peak-ID percentage = 100% × [1 − unexplained peaks/(peaks found − type I peaks)].
[c] Nuclide-ID percentage = 100% × (1 − missing nuclides/nuclides present).

operations. Validation tests are of particular importance in assessing the quality of the algorithms and the functional capability of the overall measurement and analysis system.

Software validation is a complex issue. Validation tests can be designed to focus on the performance of the individual components of the software, such as the peak search capability, or they could emphasize the overall analytical capability.

Some commercial programs have ISO 9000 certification. Many quality issues, such as documentation, are at high levels in these products. The certification, however, does not guarantee that the algorithms of the software are correct or are capable of performing the specific task of the laboratory at an adequate level. In fact, experience has shown that such software may indeed contain serious errors which are very difficult to correct because of certification or accreditation requirements.

Test spectra are necessary for software validation. These spectra should be accompanied by a certificate that specifies the nuclides, peak positions and peak areas (counts); the baseline should be realistic for the samples of interest. In general, there are no test spectra available that have universal acceptance by the users, software developers and regulatory bodies. It is not a simple task to prepare small peaks above certain continuum without violating statistics of the superimposed spectrum. The test spectra should be generated for the samples of interest, i.e. they should resemble routine spectra of the laboratory.

Test spectra can be generated by measurements or directly by software. An elegant possibility is to spike routine spectra with relevant radionuclides and then to test the performance of the software against this well-characterized nuclide composition. The statistical properties of the channel contents should be correct, and no model-dependent unrealistic features should be included.

The IAEA distributed in 1976 a set of nine artificial spectra. Another set of eight experimental spectra were generated for the IAEA software intercomparison in 1995.[16] These spectra are now available from the IAEA. The International Electrotechnical Commission (IEC) has plans to produce test spectra, and some software vendors have their own test spectra but these are not generally available. Further international coordination and cooperation are necessary to provide test spectra for the quality assessment of software products.

Software for data processing is intended to detect large and small peaks, to identify the nuclides and to quantify the activities in a reliable manner. The metrics outlined in Table 11 suggests criteria for the technical performance of the automated operations and the interactive analysis.

The peak search capability should be sensitive and reliable. Only a few spurious peaks (type I error) should be identified, and real peaks should not be missed (type II error). The nuclide identification performance is another important performance indicator of the quality of the analysis. Here, in particular, the type I and type II error rates should be small. Type II error is more severe than type I error. In interactive analysis it is far simpler to remove peaks and nuclides than to locate and identify new peaks and nuclides.

8.4 Intercomparison and Performance Evaluation

The use of test samples and participation in intercomparison exercises give confidence in the overall analytical performance. The intention is to test the reliability and accuracy of the system rather than the performance of its individual components. Differences between laboratories may reveal systematic errors in the analytical capability, such as the omission of the true coincidence correction in the low-level counting in close geometries.

Performance evaluation of radionuclide measurement and analysis, either on a short-term or on a long-term basis, is an essential quality element. A systematic approach is necessary, but, above all, the quality culture of the laboratory is determined by the commitment of the staff to the goals, principles, and procedures defined by the management.

9 DISCUSSIONS

Determination techniques of radionuclides using γ-spectrometry have been developed to a high level since the early steps in the 1960s. The modern trends in γ-spectrometric measurements and analysis approach the level of complete automation. It will not take many years before the customers can buy a complete solution to the nondestructive determination of various nuclides, including calibrations, analysis and reliable quality assurance system.

Germanium can be purified to extremely low concentrations of impurity atoms. The continuous improvement of raw material makes it possible to produce larger and larger detectors. The trend toward higher counting efficiency will inevitably continue. The commercial availability of standard germanium detectors makes calibrations easy and reliable. The counting system can be applied to a new type of sample simply by calculating the new efficiency calibration curve without any measurement. The quality system is greatly simplified because calibrations and their documentation are eliminated, or at least partially eliminated. A large fraction of the quality issues is transferred from the users to the companies selling detectors, associated electronics and software.

Although germanium detectors dominate nuclide determination techniques, there is an incessant need to

develop a high-resolution detector (or at least an essentially better detector than NaI) that could be operated at room temperature. To reduce the leakage current, such a detector must have a band gap larger than 1 eV. The band gap should not, however, be too large, or the energy resolution would be deteriorated beyond the useful range. The charge carrier mobility should also be good and the charge collection efficient. High-Z materials, such as CdTe and HgI_2, are suitable for production of small detectors. The volumes range from 20 mm^3 to a few cm^3.[2,42] Intensive research efforts aim to develop room temperature detectors with larger dimensions. Adding Zn seems to improve the quality of the detector. A comprehensive review on $Cd_{1-x}Zn_xTe$ detectors is given by James et al.[43]

The CdTe detectors have a better energy resolution than the NaI detectors, but their performance is by no means near the quality of a cooled germanium detector. Typical fwhm of the ^{137}Cs peak (661.6 keV) is less than 2 keV for HPGe detector, 40–60 keV for the NaI detector and 15–25 keV for the CdTe detector. The resolution of the CdTe detector can be improved (<5 keV) through Peltier cooling (<0 °C). The peak-to-Compton ratio for CdTe and CdZnTe detectors is poor because of the small volume of the detector as the photons tend to escape the crystal.

Charge collection and long-term stability are severe problems of CdTe detectors.[42] Owing to these difficulties, the pulse rise time varies more than in germanium detectors. Advanced digital signal processing mitigates this drawback, but despite all efforts the CdTe detectors are still individuals. The variation of the material is difficult to handle.

The room temperature detectors have many applications, such as uranium enrichment measurement (below 186 keV), spent fuel verification measurement (^{137}Cs, 661 keV)[42] and emergency preparedness monitoring (^{131}I, 364 keV) in a nuclear power plant.

The demand of society for sensitive and reliable environmental monitoring requires low detection limits in assay systems. Indeed, the possibility to determine low amounts of radionuclides provides information on the releases of industrial operations, especially from the nuclear fuel cycle. Many applications, such as safeguards verification, benefit from the improved detection and identification capability. The nondestructive analysis has the inherent disadvantage that the measurement is performed together with the disturbing materials in the sample. The background nuclides within the sample are the dominating source of inaccuracy in many analyses. New methods have to be found to enrich nuclides of interest in the samples collected from the environment.

Air sampling and subsequent γ-spectrometric analysis of the filter are often the most sensitive methods for searching for nuclear signatures in the environment. The air sampler concentrates the airborne radionuclides to a small sample. The face velocity of the air through the filter is typically about 1 ms^{-1}, whereas particles are deposited on surfaces with a velocity about three orders of magnitude smaller. The advantages of the enormous concentration capability cannot be fully utilized because of the similar enrichment of background nuclides, radon and thoron progenies in particular. This deterioration of the detection capability would be reduced if more advanced sampling methods were available. The man-made nuclides are often attached to larger particles than the background nuclides. It is possible to design equipment that would separate out the small background nuclides ($\ll$1 μm) from the main air stream. An impactor could sort the particles according to their size. However, the flow rate through such a device is small and thus the overall detection capability is not improved. A virtual impactor, i.e. a device that sorts particles according to their size without using impaction plates, would allow a much larger flow to pass through the system. Design of such an instrument for high-volume air sampling would improve the detection capability of man-made nuclides.

Particle fractionation provides ideal samples not only for γ-spectrometry but for other analyses as well, such as α-spectrometry, without resorting to radiochemistry. In the future, the nuclide determination techniques are not only improved through counting and analysis equipment but also by adoption of advanced sampling systems.

γ-spectrum analysis has been, and it still is, a difficult task requiring human intervention. Recent development, based on a data-driven approach rather than on straightforward adoption of rigid algorithms, may ultimately provide complete explanation of a γ-spectrum, with every peak explained. An advanced expert system (section 6.6) can already perform automatically a better analysis than many analysts can do with traditional interactive software.

ACKNOWLEDGMENTS

We thank Dr Lars-Erik DeGeer and Dr Murray Matthews, Comprehensive Nuclear-Test-Ban Treaty Organisation, Vienna, and Mr Jarmo Ala-Heikkilä and Mr Timo Hakulinen, Helsinki University of Technology, for valuable comments on the manuscript. Mr Arto Leppänen and Mr Tapani Honkamaa, Radiation and Nuclear Safety Authority, Finland, have provided much of the data on the detector calibrations.

ABBREVIATIONS AND ACRONYMS

AD	Analog-to-digital
ADC	Analog-to-digital Converter
CFC	Chlorofluorocarbon
DSP	Digital Signal Processor
ENSDF	Evaluated Nuclear Structure Data File
FET	Field Effect Transistor
fwfm	full width at one-fiftieth maximum
fwhm	full width at half-maximum
fwtm	full width at one-tenth maximum
Ge(Li)	Lithium-drifted Germanium
HPGe	High-purity Germanium
IAEA	International Atomic Energy Agency
IEC	International Electrotechnical Commission
MCA	Multichannel Analyzer
MDA	Minimum Detectable Activity
NAMAS	National Measurement Accreditation Service
NIST	National Institute of Standards and Technology
PUR	Pile-up-rejector
SVD	Singular Value Decomposition

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis • Sampling Considerations for Biomonitoring

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Radiochemical Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Mass Spectrometry of Long-lived Radionuclides • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation • Speciation of Radionuclides in the Environment

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

APPENDIX 1

Radionuclides, their half-lives, main emission energies (keV) and emission yield (percentage). A subset of 100 radionuclides and 245 lines was taken from the library of ENSDF maintained by the Brookhaven National Laboratories. The data was downloaded from Shaman.[17,25,26]

Radionuclide Standards

Nuclide	Half-life	Energy (keV)	Yield (%)
Na-22	2.6019a	511.0	179.789
		1274.5	99.944
Mn-54	312.12d	834.8	99.975
Co-57	271.8d	14.4	9.680
		122.1	85.900
		136.5	10.330
Co-60	5.2714a	1173.2	99.900
		1332.5	99.982
Zn-65	243.9d	1115.5	50.700
Y-88	106.65d	898.0	93.699
		1836.1	99.199
Ag-108m	127a	433.9	90.500
		614.3	89.800
		722.9	90.800
Cd-109	1.26656a	88.0	3.610
Sn-113	115.09d	391.7	64.000
Cs-137	30.1a	661.7	85.209
Ba-133	10.52a	81.0	34.099
		276.4	7.164
		302.9	18.330
		356.0	62.049
		383.9	8.940
Eu-152	13.542a	121.8	28.400
		244.7	7.490
		344.4	26.600
		778.9	12.960
		964.1	14.340
		1085.9	9.920
		1112.1	13.550
		1408.0	20.870
Eu-154	8.592a	123.1	40.400
		722.3	20.000
		873.2	12.090
		996.3	10.340
		1004.8	17.900
		1274.5	34.400
Hg-203	46.612d	279.2	81.459
Am 241	432.7a	59.5	35.900

Natural Radionuclides

Nuclide	Half-life	Energy (keV)	Yield (%)
Be-7	53.29d	477.6	10.520
K-40	1.277E + 09a	1460.8	10.670
Ge-75m	47.7s	139.7	38.799

(*continued overleaf*)

Natural Radionuclides (*continued*)

Nuclide	Half-life	Energy (keV)	Yield (%)
Tl-208	3.053m	510.8	22.600
		583.2	84.500
		860.6	12.420
		2614.5	99.160
Tl-210	1.3m	296.0	79.000
		799.6	98.959
		1070.0	12.000
		1210.0	17.000
		1316.0	21.000
Pb-210	22.3a	46.5	4.250
Pb-211	36.1m	404.9	3.780
		832.0	3.520
Pb-212	10.64h	238.6	43.299
		300.1	3.280
Pb-214	26.8m	242.0	7.500
		295.2	18.500
		351.9	35.799
Bi-211	2.14m	351.1	12.910
Bi-212	1.00917h	727.3	6.580
		785.4	1.102
		1620.5	1.490
Bi-214	19.9m	609.3	44.799
		768.4	4.800
		1120.3	14.800
		1238.1	5.860
		1764.5	15.360
Rn-220	55.6s	549.7	0.110
Rn-222	3.8235d	510.0	0.076
Ra-223	11.435d	269.6	13.700
Ra-224	3.66d	241.0	3.970
Ra-226	1600a	186.1	3.500
Ac-228	6.15h	338.3	11.300
		911.2	26.600
		969.0	16.200
Th-227	18.72d	236.0	12.300
Th-228	1.9131a	84.5	1.266
		131.6	0.135
		216.0	0.263
Th-232	1.405E + 10a	63.8	0.267
		140.9	0.027
Th-234	24.1d	63.3	4.500
		92.4	2.600
		92.8	2.600
Pa-234	6.7h	131.2	20.000
		569.5	10.000
		883.2	15.000
		926.0	11.000
		927.1	11.000
		946.0	12.000
Pa-234m	1.17m	766.4	0.210
		1001.0	0.590
U-235	7.038E + 08a	143.8	10.900
		185.7	57.500
U-238	4.468E + 09a	49.6	0.064

Man-made Radionuclides

Nuclide	Half-life	Energy (keV)	Yield (%)
Na-24	14.959h	1368.6	100.000
		2754.1	99.944
K-42	12.36h	1524.7	18.080
Sc-46	83.81d	889.3	99.984
		1120.5	99.987
Ti-44	49a	67.9	94.400
		78.3	96.199
Cr-51	27.702d	320.1	10.080
Mn-56	2.5785h	846.8	98.900
		1810.7	27.200
		2113.1	14.300
Fe-59	44.496d	1099.3	56.500
		1291.6	43.200
Co-58	70.82d	810.8	99.447
Ni-65	2.52h	1115.5	14.800
		1481.8	23.500
Br-82	1.47083d	554.3	70.800
		619.1	43.400
		698.4	28.500
		776.5	83.500
		827.8	24.000
		1044.0	27.200
		1317.5	26.500
		1474.9	16.320
Zr-95	64.02d	724.2	44.150
		756.7	54.500
Zr-97	16.9h	743.3	94.800
Nb-94	20 300a	702.6	97.900
		871.1	99.900
Nb-95	34.97d	765.8	99.790
Nb-95m	3.61d	235.7	24.100
Mo-99	2.7475d	140.5	4.520
		181.1	6.080

Man-made Radionuclides (*continued*)

Nuclide	Half-life	Energy (keV)	Yield (%)
		739.6	12.130
		778.0	4.340
Mo-101	14.61m	191.9	18.799
		505.9	11.800
		590.9	16.400
		1012.5	12.800
Tc-99m	6.01h	140.5	89.059
Ru-97	2.9d	215.7	86.199
		324.6	10.200
Ru-103	39.26d	497.1	91.000
Ru-105	4.44h	316.4	11.100
		469.4	17.500
		676.4	15.700
		724.3	47.299
Rh-101	3.3a	127.2	73.000
		198.0	71.000
		325.2	13.400
Rh-102	207d	475.1	39.000
		511.0	29.440
Rh-102m	2.9a	475.1	95.000
		631.3	56.000
		697.5	44.000
		766.8	34.000
		1046.6	34.000
		1112.8	19.000
Rh-105	1.47333d	306.1	5.100
		318.9	19.100
Pd-109	13.46h	88.0	3.600
Ag-110m	249.76d	657.8	94.599
		677.6	10.350
		706.7	16.440
		763.9	22.290
		884.7	72.699
		937.5	34.360
		1384.3	24.280
		1505.0	13.040
Ag-111	7.45d	245.4	1.240
		342.1	6.700
Cd-107	6.5h	93.1	4.800
Sn-125	9.64d	332.1	1.310
		469.9	1.380
		822.5	3.990
		915.6	3.850
		1067.1	9.040
		1089.2	4.280
		2001.8	1.790
Sb-122	2.7d	564.2	69.300
		692.7	3.780
Sb-124	60.2d	602.7	97.800
		722.8	10.760
		1691.0	47.299
Sb-125	2.73a	427.9	29.400
		463.4	10.450
		600.6	17.780
		606.6	5.020
		635.9	11.320
Te-132	3.204d	228.2	88.000
I-123	13.2h	159.0	83.300
I-131	8.04d	364.5	81.199
I-132	2.295h	522.7	16.000
		630.2	13.300
		667.7	98.699
		772.6	75.599
		954.6	17.600
I-133	20.8h	529.9	87.000
I-134	52.6m	595.4	11.200
		621.8	10.600
		847.0	95.400
		884.1	64.900
		1072.6	15.000
I-135	6.57h	1131.5	22.740
		1260.4	28.900
		1678.0	9.620
		1791.2	7.770
Cs-134	2.062a	563.2	8.380
		569.3	15.430
		604.7	97.599
		795.8	85.400
		801.9	8.730
Ba-140	12.752d	537.3	24.389
La-140	1.6781d	328.8	20.600
		487.0	44.299
		815.8	22.900
		1596.2	95.400
Ce-139	137.64d	165.9	79.886
Ce-141	32.501d	145.4	48.200
Ce-143	1.37663d	57.4	11.700
		293.3	42.799
Ce-144	284.893d	133.5	11.090
Pr-144	17.28m	696.5	1.342
		2185.7	0.694
Nd-147	10.98d	91.1	27.900
		531.0	13.100

Noble Gases

Nuclide	Half-life	Energy (keV)	Yield (%)
Ar-41	1.822h	1293.6	99.099
Kr-79	1.46d	261.3	12.700
		511.0	14.210
		606.1	8.100
Kr-85	10.756a	514.0	0.434
Kr-85m	4.48h	151.2	75.000
		304.9	14.000
Kr-87	1.27167h	402.6	50.000
		845.4	7.300
		2554.8	9.200
Kr-88	2.84h	196.3	26.000
		834.8	13.000
		1529.8	10.900
		2195.8	13.200
		2392.1	34.599
Xe-133	5.243d	81.0	38.000
Xe-133m	2.19d	233.2	10.000
Xe-135	9.14h	249.8	90.199
Xe-135m	15.29m	526.6	80.500

REFERENCES

1. G.F. Knoll, *Radiation Detection and Measurement*, 2nd edition, John Wiley & Sons, New York, 1989.
2. E. Sakai, 'Present Status of Room Temperature Semiconductor Detectors', *Nucl. Instrum. Methods*, **196**, 121–130 (1982).
3. NIST Standard Reference Database 8, 'X-ray and Gamma-ray Attenuation Coefficients and Cross-sections Database, version 2.01', 1988.
4. D.V. Freck, J. Wakefield, 'Gamma-ray Spectrum Obtained with a Lithium-drifted p-i-n Junction in Germanium', *Nature*, **193**(4816), 669 (1962).
5. G. Gilmore, J.D. Hemingway, *Practical Gamma-ray Spectrometry*, John Wiley & Sons, New York, 1995.
6. K. Debertin, R.G. Helmer, *Gamma- and X-ray Spectrometry with Semiconductor Detectors*, North Holland, Amsterdam, 1988.
7. S.S. Kapoor, V.S. Ramamurthy, *Nuclear Radiation Detectors*, Wiley Eastern Limited, New Delhi, 1986.
8. Bertolini, A. Coche (eds.) *Semiconductor Detectors*, North Holland, Amsterdam, 1968.
9. Product Catalogue of Canberra Industries, Inc., Meriden, 1997.
10. G. Ottaviani, C. Canali, A. Alberigi Quaranta, 'Charge Carrier Transport Properties of Semiconductor Materials Suitable for Nuclear Radiation Detectors', *IEEE Trans. Nucl. Sci.*, **NS-22**, 192–204 (1975).
11. Product Catalogue of EG&G ORTEC, Inc, Oak Ridge, 1997.
12. R.H. Pehl, F.S. Goulding, D.A. Landis, M. Lenzlinger, 'Accurate Determination of the Ionization Energy in Semiconductor Detectors', *Nucl. Instrum. Methods*, **59**, 45–55 (1968).
13. IEEE Standard Test Procedures for Germanium Gamma-ray Detectors, ANSI/IEEE Standard 325, 1986.
14. S. Klemola, 'Optimization of Sample Geometries in Low-level Gamma-Spectroscopy', *Nucl. Instrum. Methods, A*, **369**, 578–581 (1996).
15. H. Aaltonen, S. Klemola, F. Ugletveit, 'Validation of a Method for Computer Calculation of Germanium Detector Efficiencies', *Nucl. Instrum. Methods Phys. Res. A*, **339**, 87–91 (1994).
16. M. Blaauw, V. Osorio Fernandez, P. van Espen, G. Bernasconi, R. Capote Noy, H. Manh Dung, N.I. Molla, 'The 1995 IAEA Intercomparison of γ-ray Spectrum Analysis Software', *Nucl. Instrum. Methods, A*, **387**, 416–432 (1997).
17. T.T. Hakulinen, 'Knowledge Based System for Identification of Gamma Sources', Helsinki University of Technology, TKK-F-B147, Helsinki, 1993.
18. J.T. Routti, S.G. Prussin, 'Photopeak Method for the Computer Analysis of Gamma Ray Spectra from Semiconductor Detector', *Nucl. Instrum. Methods*, **72**, 125 (1969).
19. P.A. Aarnio, M.T. Nikkinen, J.T. Routti, 'SAMPO 90 High Resolution Interactive γ-spectrum Analysis Including Automation with Macros', *J. Radioanal. Nucl. Chem.*, **160**(1), 289–296 (1992).
20. T. Honkamaa, H. Toivonen, M. Nikkinen, 'Monitoring of Airborne Contamination Using Mobile Equipment', Radiation and Nuclear Safety Authority, STUK-A130, Helsinki, 1996.
21. M.A. Mariscotti, 'A Method for Automatic Identification of Peaks in the Presence of Background and its Application to Spectrum Analysis', *Nucl. Instrum. Methods*, **50**, 417–532 (1967).
22. L.A. Currie, 'Limits for Quantitative Detection and Quantitative Determination', *Anal. Chem.*, **40**(3), 586–593 (1968).
23. Leppänen, H. Toivonen, 'Detection of Man-made Gamma-emitting Radionuclides in the Presence of Radon Progeny Aerosols', Proceedings of the Fourteenth International Conference on Nucleation and Atmospheric Aerosols, Helsinki, 674–677, 1996.
24. P.A. Aarnio, M.T. Nikkinen, J.T. Routti, 'γ-Spectrum Analysis Including NAA with SAMPO for Windows', *J. Radioanal. Nucl. Chem.*, **193**(1), 179–185 (1995).
25. P.A. Aarnio, J.J. Ala-Heikkilä, T.T. Hakulinen, J.T. Routti, 'Expert System for Nuclide Identification in γ-spectrum Analysis', *J. Radioanal. Nucl. Chem.*, **193**, 219–227 (1995).
26. J.J. Ala-Heikkilä, 'Expert System for Nuclide Identification in Environmental γ-Spectra', Helsinki University of Technology, TKK-F-B159, Helsinki, 1995.

27. J. Ala-Heikkilä, T. Hakulinen, P. Aarnio, M. Nikkinen, H. Toivonen, 'Evaluation of Expert System Shaman in Processing Gamma-ray Spectra at the Comprehensive Test Ban Treaty Prototype International Data Center, Helsinki University of Technology, TKK-F-B171, Helsinki, 1997.
28. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, 'Numerical Recipes in C – The Art of Scientific Computing', Cambridge University Press, New York, 1988.
29. K. Debertin, U. Schotzig, 'Coincidence Summing Corrections in Ge(Li)-Spectrometry at low Source-to-Detector Distances', *Nucl. Instrum. Methods*, **158**, 471–477 (1979).
30. H.L. Beck, J. DeCampo, C. Gogolak, 'In Situ Ge(Li) and NaI(Tl) Gamma-ray Spectrometry', Report HASL-258. US Department of Energy, Environmental Measurement Laboratory, New York, 1972.
31. ICRU. International Commission of Units and Measurements, 'Gamma-ray Spectrometry in the Environment', ICRU Report 53, ICRU Bethesda, 1994.
32. R.R. Finck, 'High Resolution Field γ-spectrometry and its Application to Problems in Environmental Radiology', Doctoral Dissertation, Lund University, 1992.
33. T. Honkamaa, H. Toivonen, M. Nikkinen, 'Fallout Mapping Using Advanced Measuring Vehicle', in *Rapid Radioactivity Measurements in Emergency and Routine Situations*, Proceedings of the National Physical Laboratory, Teddington, UK, 257–263, 1999. *Radioact. Radiochem.*, (in press).
34. I.K. Helfer, K.M. Miller, 'Calibration Factors for Ge Detectors Used for Field Spectrometry', *Health Phys.*, **55**, 15–29 (1988).
35. W. Sowa, E. Martini, K. Gehrcke, P. Marschner, M. Naziry, 'Uncertainties of In Situ γ-spectrometry for Environmental Monitoring', *Radiat. Prot. Dosim.*, **27**, 93–101 (1989).
36. M. Nikkinen, M. Kettunen, H. Toivonen, 'Mobile Radiation Detection Using γ-spectrometers', IAEA-SM-351 (International Atomic Energy Agency), Symposium on International Safeguards, Vienna, 1998.
37. Nordic Nuclear Safety Research Program, 'Resume95: Rapid Environmental Surveying Using Mobile Equipment', NKS/EKO-3, Copenhagen, 1997.
38. K. Saito, R. Sakamoto, M. Tsutsumi, T. Nagaoka, S. Moriuchi, 'Prompt Estimation of Release Rates of Gaseous Radioactivity from a Nuclear Plant Using Aerial Survey, *Rad. Prot. Dosim.*, **22**, 77–85 (1988).
39. ISO 9001. International Organization for Standardization. Quality Systems. Model for Quality Assurance in Design/Development, Production, Installation and Servicing, 1987.
40. ISO 17025. International Organization for Standardization. General Requirements for the Competence of Testing and Calibration Laboratories, 1998.
41. International Atomic Energy Agency, X-ray and Gamma-ray Standards for Detector Calibration, IAEA-TECDOC-619, IAEA, Vienna, 1991 (available also on diskette from IAEA Nuclear Data Section).
42. R. Arlt, D.E. Rundquist, 'Room Temperature Semiconductor Detector for Safeguards Measurements', *Nucl. Instrum. Methods*, **380**, 455–461 (1996).
43. R.B. James, T.E Schlesinger, J. Lund, M. Schieber, '$Cd_{1-x}Zn_xTe$ Spectrometers for Gamma and X-ray Applications', *Semiconduct. Semimetals*, **43**, 335–381 (1995).

Mass Spectrometry of Long-lived Radionuclides

H.J. Sabine Becker and Hans-Joachim Dietze
Central Department of Analytical Chemistry, Research Centre, Jülich, Germany

1 Introduction 1
2 Application of Different Mass Spectrometric Techniques for Determination of Radionuclides 2
2.1 Methods of Solid-state Mass Spectrometry 2
2.2 Inductively Coupled Plasma Mass Spectrometry 5
3 Conclusions 11
Abbreviations and Acronyms 12
Related Articles 12
References 12

Inorganic mass spectrometric methods are powerful multielement analytical techniques for sensitive determination at the trace and ultratrace level, and for isotope ratio measurements in different materials (e.g. conducting, semiconducting and nonconducting solid samples; technical, environmental, biological, geological and water samples). An inorganic mass spectrometer consists of the following parts: the ion source for the evaporation, atomization and ionization of samples, the analyzer for the mass (and energy) separation of ion beams which are extracted from the ion source, and the ion detection system. The detection limits of solid-state mass spectrometric techniques by thermal ionization mass spectrometry (TIMS), glow discharge mass spectrometry (GDMS), laser ablation inductively coupled plasma mass spectrometry (LA/ICPMS) for the direct analysis of solid samples were determined down to the sub-nanogram per gram level. Resonance ionization mass spectrometry (RIMS) and accelerator mass spectrometry (AMS) are highly sensitive, selective ultrasensitive mass spectrometric techniques with detection limits up to 10^6 atoms.

Inductively coupled plasma mass spectrometry (ICPMS) has been applied as the most important and sensitive mass spectrometric technique with its multielement capability for the determination of long-lived radionuclides in solid and aqueous samples, with detection limits the sub-picogram per liter. Isotope ratios can be determined by ICPMS using single ion collectors and multicollectors with a precision of 0.02% and 0.006% relative standard deviation (RSD) respectively. This review discusses the most important inorganic mass spectrometric techniques and their application for the quantitative determination of long-lived radionuclides in metals, alloys, and nonconductors, as well as in environmental, biological and geological samples, and radioactive waste materials. The application of these methods to isotope ratio measurements is also discussed.

1 INTRODUCTION

Inorganic mass spectrometric methods, with their ability to provide a very sensitive multielement determination of trace and ultratrace elements and precise isotopic analysis, have been increasingly established for the determination of long-lived radionuclides[1–4] especially in environmental materials such as waters, and geological and biological samples,[5–15] nuclear materials and radioactive waste,[16–20] high-purity materials,[21,22] and ceramics and glass.[9,23,24] Such mass spectrometric measurements of long-lived radionuclides are of additional interest in areas such as radiobioassay, environmental monitoring, decontamination and environmental remediation, health, safety, nuclear waste characterization (radioactive waste control) and management of radioactive waste for storage and disposal. The applications fields of inorganic mass spectrometry as a universal and extremely sensitive analytical method for the simultaneous determination of long-lived radionuclides in the trace and ultratrace range, and their isotope ratio measurements are already very wide and will increase as the sensitivity and precision improve and ever-decreasing detection limits are possible. The important applications of long-lived radionuclides in analytical chemistry are summarized in Table 1. The determination of the concentration and precise isotopic analysis of radioactive elements (e.g. the terrestrial sources of radioactivity, ^{238}U, ^{235}U, ^{232}Th and their decay nuclides) by inorganic mass spectrometry is used for environmental research, in solid-state research and for controlling the purity of, for example, high-purity metals, alloys, semiconductors and insulators for microelectronics. Furthermore, the measurement of contamination and enrichment of selected radionuclides (e.g. ^{129}I as an environmental indicator of nuclear accidents, ^{99}Tc or ^{79}Se as fissions products, and ^{237}Np, ^{239}Pu, ^{240}Pu, ^{241}Am) at ultralow levels during the environmental monitoring required as a result of

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

Table 1 Applications of mass spectrometry for the determination of long-lived radionuclides

Environmental monitoring	Solid-state research and material controlling	Radioactive waste	Geochronology	Isotope dilution
Natural radioactivity contamination in environment (fallout from nuclear weapons tests, nuclear power plants)	*Natural radioactivity (in high-purity metals semiconductors and ceramics)*	*From nuclear reactor*	*Dating of geological samples*	*Determination of element concentrations*
⇓	⇓	⇓	⇓	⇓
^{99}Tc, ^{129}I, Pu, ^{237}Np, Th, U, ^{241}Am	Th, U	Recycling, final storage ^{79}Se, ^{99}Tc, Th, U, Pu, Am	$^{87}Rb \xrightarrow{\beta^-} {}^{87}Sr$ $^{147}Sm \xrightarrow{\alpha} {}^{143}Nd$ $^{187}Re \xrightarrow{\beta^-} {}^{187}Os$ $^{238}U \xrightarrow{\text{fission}} {}^{206}Pb$	Ultratrace analysis ^{129}I, ^{235}U, ^{230}Th

fallout from nuclear weapons testing, the operation of nuclear power plants or after nuclear accidents are of increasing interest. Long-lived radionuclides have also been be used for tracer experiments in biological and geological research, and the application of isotopic dilution methods using long-lived radioactive isotopes.[25] On the other hand, the determination of long-lived radionuclides by the isotope dilution procedure can be carried out with the application of stable, enriched isotopes.[26] Among the existing analytical physical methods (such as optical emission spectrometry, atomic absorption spectrometry, radioanalytical techniques, neutron activation analysis, and X-ray spectrometry) inorganic mass spectrometry is preferred for the determination of very low concentrations of radionuclides and their isotope ratios at the ultratrace level. Inorganic mass spectrometry has been developed into a congruent method of the well-established classical radioanalytical techniques because of easier sample preparation steps and excellent detection capability, especially for the characterization of radionuclides with half-lives $>10^4$ years in environmental samples and waste materials with very low radioactivity.

The development of analytical methods for the determination of long-lived radionuclides at ultratrace concentration levels in environmental samples and radioactive waste materials from nuclear reactors has focused on improving microanalytical techniques in order to reduce the sample volume, to improve the precision (RSD) of measurements and the detection limits. For example, quality control of radioactive waste packages requires powerful and fast analytical methods which allow many samples to be measured in a short time, with a high degree of accuracy and precision. Using conventional radiochemical methods for the determination of long-lived radionuclides at low concentrations, a careful chemical separation of the analyte, e.g. by extraction or chromatography, is often necessary. Such complete separation steps of the interferents and enrichment of long-lived radionuclides in order to achieve low detection limits are generally time-consuming.

This review discusses the most important mass spectrometric techniques and their application for quantitative determination and isotope ratio measurements of long-lived radionuclides.

2 APPLICATION OF DIFFERENT MASS SPECTROMETRIC TECHNIQUES FOR DETERMINATION OF RADIONUCLIDES

2.1 Methods of Solid-state Mass Spectrometry

In trace element determination and isotope analysis of long-lived radionuclides in solids, very sensitive multielement solid state mass spectrometric methods allow a direct analysis of the sample material without any chemical sample preparation. The mass spectrometric analytical techniques with multielemental capability used for the determination of radionuclides are LA/ICPMS,[27–29] GDMS[30–32] and secondary ion mass spectrometry (SIMS).[5,6,33] TIMS,[34,35] AMS[36] and RIMS[37,38] have been used for sensitive ultratrace analysis and precise determination of isotopic ratios of selected long-lived radionuclides in solid samples. Because the mass spectrometric technique is determined by the evaporation of solid sample (e.g. thermal evaporation, laser ablation, evaporation and atomization of sample in plasmas, or using electron beam or ion bombardment) and ionization method applied (e.g. electron impact ionization, multiphoton ionization using a laser beam and ionization

during the sputtering process) the ion separation system to be used is essentially defined by the physical properties of the ions formed (e.g. initial energy) and the ionization itself (e.g. formation of disturbing molecular and cluster ions, often with great intensity). Both static magnetic sector fields (single-focusing) or combinations of electric and magnetic sector fields (double-focusing) are used for ion separation as well as dynamic ion separation systems (e.g. quadrupole mass spectrometer, time-of-flight (TOF) mass spectrometer, Fourier transform mass spectrometer and ion trap mass spectrometer). Ion detection can be performed electrically (with photomultiplier, channel plates, Channeltron or Faraday cup).

The main advantage of solid state mass spectrometry is that a possible danger of contamination during sample preparation and the occurrence of numerous interferences in the mass spectra of aqueous solutions after digestion of sample with acids (in comparison to ICPMS) can thus be reduced to a minimum. However, quantification in solid state mass spectrometry generally proves difficult if no suitable standard reference materials are available.

2.1.1 Thermal Ionization Mass Spectrometry

TIMS is a well recognized analytical technique for the precise and accurate measurement of isotopic ratios for different elements with a precision of better than 0.01%.[35] In TIMS a small volume (1–10 μL) of aqueous sample solution, which contains some ng to μg of the analyte to be analyzed, is deposited on a cleaned filament surface (e.g. high-purity Re) and evaporated to dryness. The most frequently applied technique in TIMS works with two filaments which are arranged opposite each another. One of the filaments is used for the evaporation of sample by thermal heating and the other for ionization of evaporated atoms and molecules on the hot filament surface. Ions formed by thermal ionization possess low initial energies (0.1–0.2 eV), therefore, generally, single magnetic sector field mass spectrometers were used for ion separation. At present, two magnetic sector field instruments (e.g. MAT 262, Finnigan MAT, Germany or Sector 54 from Micromass, UK) are commercially available on the analytical market. Instruments with a multiple collector are important for precise isotope analysis and precise determination of concentration.

Mass discrimination in the instrument (e.g. ion optical system or ion detector) and mass fractionation effects (caused during the evaporation of sample, where the measured isotope ratio changes with time) are limiting factors for the accuracy of measured isotopic ratios in TIMS. A fundamental paper on fractionation correction in TIMS is that by Habfast.[39] Different effects could be considered by different internal calibration or by using isotopic standards (e.g. from the National Institute of Standards and Technology (NIST), Gaithersburg, USA; or Institute of Reference Materials and Measurements (IRMM), Geel, Belgium) with well-known isotopic ratios of an element of interest.

The capability of TIMS in precise isotopic ratio measurements, as the major application field of this analytical technique, is used in accurate trace element determination of radiogenic elements for determining element concentration by the isotope dilution method using high-enriched isotopic spikes. Interesting investigations of TIMS and IDA were performed in Heumann's group by the determination of U and Th (and other elements) in, e.g. high-purity aluminum[40] and refractory metals,[41] and silicides of Mo, W and Ta and silicon oxide[42] after trace–matrix separation. Impurities of the radioactive elements U and Th and their decay products at the picogram per gram level in high-purity refractory metals and their silicides, which have been increasingly used for gate electrodes, interconnections and diffusion barriers in integrated circuits, affect the performance of integrated circuits. An excellent review of the isotope dilution technique using TIMS, including applications for geological dating and for the determination of long-lived radionuclides is given by Heumann.[43]

TIMS has been extensively used by Aggarwal and Crain[44] for the accurate determination of half-life of a number of transactinium isotopes (e.g. ^{241}Pu, ^{242}Pu, ^{243}Am, ^{242}Cm) and of the fission yield of stable and long-lived fisson products in thermal neutron induced fission of ^{233}U, ^{239}Pu and ^{241}Pu.

The application of mass spectrometry in the determination of U and Th in high-purity metals and alloys is summarized in Table 2.

TIMS requires an often time-consuming sample preparation (digestion of solid samples and trace–matrix separation) and therefore has been replaced by the more sensitive ICPMS in the past few years.

2.1.2 Glow Discharge Mass Spectrometry

GDMS was established as a powerful and efficient analytical method for the direct trace element and depth profile analysis of solids.[49] This mass spectrometric technique uses an argon gas glow discharge at a pressure of 0.1–10 torr as an ion source. The cathode surface consisting of the sample material is sputtered by Ar^+ ions which are formed in a low-pressure plasma and accelerated towards the cathode. Sputtered neutral particles are ionized in the glow discharge plasma (''negative glow'') by Penning and/or electron impact ionization and charge exchange processes. For the direct analysis of solid samples the commercial direct current glow discharge

Table 2 Application of mass spectrometry in the determination of U and Th in metals and alloys

Samples	Method and equipment	Elements	Measured concentration	Limit of detection	Reference
Aluminum	TIMS "THQ" isotope dilution	U, Th after trace–matrix separation	0.67 ng g^{-1} (U)– 1.5 ng g^{-1} (Th)	0.018 ng g^{-1} (U) 0.06 ng g^{-1} (Th)	Beer and Heumann[40]
Molybdenum Tungsten	TIMS "MAT 261" isotope dilution	U, Th after trace–matrix separation	0.02 ng g^{-1} (Th)	0.006 ng g^{-1} (U) 0.008 ng g^{-1} (Th)	Herzner and Heumann [41]
Titanium	TIMS "THQ" isotope dilution	U, Th after trace–matrix separation	<0.07–90 ng g^{-1} (U, Th)	0.07 ng g^{-1} (U) 0.07 ng g^{-1} (Th)	Beer and Heumann[45]
Tantalum	ICPQMS ELAN 5000	U, Th after trace–matrix separation	0.11 μg g^{-1} (Th) 0.17 μg g^{-1} (U)	0.01 μg g^{-1} (Th) 0.02 μg g^{-1} (U)	Panday et al.[46]
Cobalt	TIMS "THQ" isotope dilution	U, Th after trace–matrix separation	120 ng g^{-1} (U)	0.007 ng g^{-1} (U) 0.017 ng g^{-1} (Th)	Beer and Heumann[47]
Zircalloy	ICPQMS ELAN 5000	U, Th	1.8 μg g^{-1} (Th) 0.47 μg g^{-1} (U)	0.01 μg g^{-1} (Th) 0.02 μg g^{-1} (U)	Panday et al.[48]

mass spectrometer VG-9000 (VG-Elemental, Thermo Instruments, UK), a double-focusing sector field mass spectrometer with Nier–Johnson geometry, is available for the determination of trace elements in electrically conducting materials with detection limits in the nanogram per gram concentration range and lower with reproducibility of about ±10%. The analysis of nonconducting materials by dc GDMS is difficult due to charge-up effects on the sample surface. Different techniques such as mixing nonconducting powdered samples with a high-purity metal powder (or high-purity graphite) or the use of a secondary cathode have therefore been applied for the analysis of electrically insulating samples by dc GDMS. Betti[31] used the VG 9000 for different applications to characterize radioactive waste materials in electrically conducting and nonconducting materials, e.g. for different types of nuclear fuels, alloys containing Pu and U, cladding materials, nuclear waste glasses.

2.1.3 Laser Ablation Inductively Coupled Plasma Mass Spectrometry

To an increasing extent LA/ICPMS is the method of choice for the direct analysis of solid samples. This analytical technique used the evaporation of sample material by a focused or defocused laser beam, mostly using neodymium yttrium aluminum garnet (Nd : YAG) lasers with $\lambda = 1064$ nm and $\lambda/2 = 532$ nm or $\lambda/4 = 266$ nm, in an inert gas atmosphere (e.g. Ar) under normal pressure and post-ionization of evaporated and ablated material in an inductively coupled plasma of the ion source of an ICPMS. Most commercial laser ablation systems (e.g. LSX-200, CETAC, USA and LUV 266, Merchantek, USA) are coupled to low-resolution quadrupole analyzers. In our laboratory a noncommercial laser ablation system was coupled to a quadrupole-based ICPMS ("ELAN 6000", Perkin Elmer SCIEX, Canada) and double-focusing sector field mass spectrometer ("ELEMENT", Finnigan MAT, Germany) for the characterization of solid radioactive waste materials.[28] The experimental arrangement of LA/ICPMS is shown in Figure 1. For the determination of long-lived radionuclides in nonconducting materials, a synthetic laboratory standard with a concrete matrix was doped with a low concentration of long-lived radionuclides (e.g. ^{99}Tc, ^{129}I, ^{232}Th, ^{233}U, ^{237}Np, ^{238}U). Figure 2 shows a part of mass spectrum of long-lived radionuclides in concrete measured by LA/ICPMS. The detection limits determined for Tc, U and Np in a blank concrete sample were in the low pg g^{-1} range. The detection limits are lower by more than one order of magnitude using double-focusing sector field ICPMS in comparison to quadrupole LA/ICPMS.

Table 3 summarizes some selected applications of mass spectrometry in the trace and ultratrace analysis of long-lived radionuclides in nonconductors and solid radioactive waste.

2.1.4 Resonance Ionization Mass Spectrometry

In RIMS one or more lasers are tuned precisely to the wavelength required for the excited states and ionization of evaporated atoms in order to obtain a highly selective ionization of analyte. For the formation of a neutral gas by evaporation of solid or liquid samples all well-known sources of evaporation (e.g. thermal vaporization on a hot Re filament, laser and spark ablation, glow discharge, ion bombardment, evaporation using electron beam) were applied.

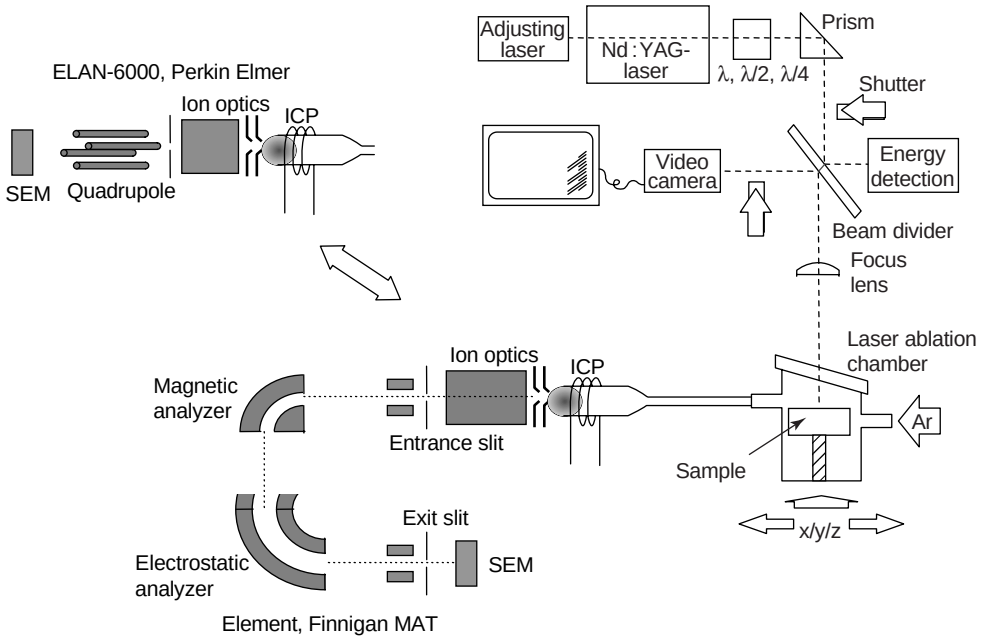

Figure 1 Experimental arrangement of LA/ICPMS using a double-focusing sector field ICPMS or quadrupole-based ICPMS. (Reproduced with permission of Gastel et al.[28])

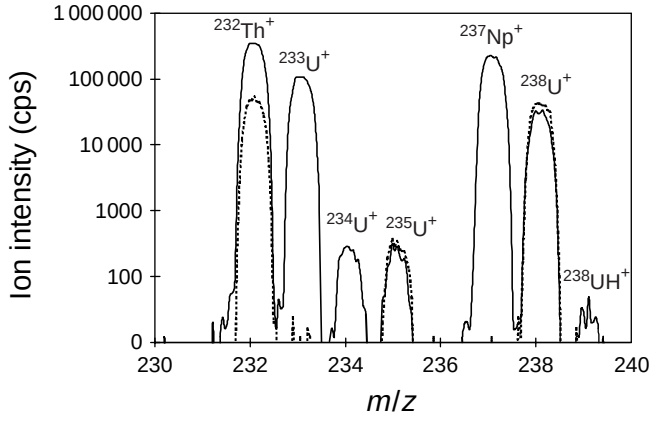

Figure 2 Part of a mass spectrum of long-lived radionuclide in concrete measured by LA/ICPMS. (Reproduced with permission of Gastel et al.[28])

RIMS is a highly selective and sensitive method for trace analysis of radiotoxic elements in environmental, biological and technical samples. The determination of Pu after electrolytic separation from soil, air filters or urine was described by Erdmann et al.[51] with detection limits of 10^6–10^7 atoms. Nunnemann et al.[13] determined Pu isotope ratios in environmental samples in order to distinguish between Pu from nuclear power plants and from nuclear fallout after nuclear weapons tests. The possibility of determining ^{89}Sr and ^{90}Sr at the ultratrace level in environmental samples by using RIMS has been reported by Wendt et al.[38]

2.1.5 Accelerator Mass Spectrometry

AMS is a highly selective (and extremely expensive) ultrasensitive mass spectrometric technique. The sensitive determination of exotic radionuclides such as ^{14}C, ^{3}He, ^{10}Be, ^{26}Al, ^{32}Si, ^{129}I or ^{205}Pb is carried out by ion sputtering (e.g. using a Cs^+ primary ion source). The sputtered secondary analyte ions were extracted into a two-stage mass filter. The mass-separated ions were accelerated in a tandem accelerator to energies of some MeV and were stripped in a following gas target or foil stripper in order to separate isobaric atomic and molecular ions. After the deceleration of ions and a final mass separation the ions were sensitively detected.

Recently, AMS was able to determine ^{236}U/^{238}U isotope ratios in the range 10^{-8}–10^{-14} for evidence of ^{236}U in environmental samples.[52] With AMS absolute detection limits of 10^6 atoms are reached for isotopes such as ^{239}Pu.[53]

With its capability of providing isotopic abundance ratios as low as 10^{-15} for very small samples, AMS was developed for the detection of extremely low concentrations of long-lived radionuclides for research in geochronology and archaeology.

2.2 Inductively Coupled Plasma Mass Spectrometry

As the most popular analytical technique for fast multielement determination of elements in the trace and ultratrace concentration range in aqueous solutions [54,55] ICPMS has been increasingly used for the determination

Table 3 Application of mass spectrometry in trace and ultratrace analysis of long-lived radionuclides in nonconductors, in solid and aqueous radioactive waste, and in standard solutions

Sample	Method and equipment	Radionuclides	Measured concentration	Limit of detection	Reference
Silicide of Mo, W and Ta	TIMS	U after trace–matrix separation	0.1 ng g^{-1}	1 pg g^{-1} (U)	Herzner and Heumann[42]
SiO_2	"THQ" Isotope dilution		(U in SiO_2)	9 pg g^{-1} (U)	
UO_2 fuels	ICPQMS "Elan 5000" ion chromatography isotope dilution	^{238}U, ^{237}Np, ^{239}Pu	1 µg L^{-1} (^{237}Np, ^{239}Pu)	0.06 µg L^{-1} (Np) 0.026 mg L^{-1} (Pu)	Barrero Moreno et al.[50]
Concrete, radioactive waste samples	Double-focusing LA/ICPMS "ELEMENT"	^{99}Tc, ^{232}Th, ^{233}U, ^{235}U, ^{238}U, ^{237}Np	0.03 µg g^{-1} – 6.1 µg g^{-1}	In synthetic concrete laboratory standard 0.02 ng g^{-1} (^{236}U) 1.4 ng g^{-1} (^{99}Tc)	Becker et al.[27]
High-level waste sludge, as a molten glass (for disposal)	ICPQMS VG PlasmaQuad	^{79}Se, ^{93}Zr, ^{107}Pd, ^{239}Pu, ^{243}Am, ^{244}Cm	0.44 µg g^{-1} (^{243}Am) 0.19 µg g^{-1} (^{244}Cm) 0.85 µg g^{-1} (^{242}Pu)	No data	Bibler et al. [9]
Glass	ICPQMS VG PlasmaQuad	U, Th after trace–matrix separation	0.1 – 1 µg g^{-1} (U, Th)	3 ng g^{-1} (U) 1 ng g^{-1} (Th)	Rohr et al. [24]
Standard solution, radwaste samples	DIHEN/ICPQMS "Elan 6000"	^{226}Ra, ^{230}Th, ^{233}U, ^{237}Np, ^{239}Pu, ^{241}Am	Isotope ratios $^{235}U/^{238}U$ at 0.5 ng L^{-1} $^{239}Pu/^{240}Pu$ at 12 ng L^{-1}	0.012 ng L^{-1} (^{238}U) 0.063 ng L^{-1} (^{239}Pu)	Becker et al.[20]
Standard solution, radwaste samples	DF/ICPMS "ELEMENT"	^{226}Ra, ^{230}Th, ^{233}U, ^{237}Np, ^{239}Pu, ^{241}Am	Isotope ratios $^{239}Pu/^{240}Pu$ at 10 pg L^{-1}	0.05 pg L^{-1} (^{241}Am) 0.04 pg L^{-1} (^{239}Pu)	Becker and Dietze[4]

DIHEN, direct injection high-efficiency nebulizer; DF/ICPMS, double-focusing inductively coupled plasma mass spectrometry.

of long-lived radionuclides with very low activity. In contrast to solid-state mass spectrometry, ICPMS is being used for the characterization of long-lived radionuclides in aqueous and solid samples after sample dissolution owing to the simple quantification procedure of the aqueous solution. The principle of ICPMS can be described as follows. The chemical compounds contained in the sample solution are decomposed into their atomic constituents in an inductively coupled argon plasma and ionized at a high degree of ionization (>90% for most chemical elements) with a low fraction of multiply charged ions (≈1%). The positively charged ions are extracted from the inductively coupled plasma (at atmospheric pressure) into the high vacuum of the mass spectrometer via an interface.

At present different commercial double-focusing sector field ICPMS, e.g. ''ELEMENT'' (Finnigan MAT, Bremen, Germany), ''PlasmaTrace 2'' (Micromass Ltd, UK), ''AXIOM'' (VG Elemental, UK) and ''JMS-Plasma X2'' (Joel, Japan), and quadrupole-based ICP mass spectrometers (e.g. Perkin Elmer Sciex, Hewlett Packard, VG Elemental, Varian GmbH analytical instruments, Spectro Analytical Instruments, Micromass) are available on the international market. In the low-resolution mode, the element sensitivity of commercial double-focusing sector field ICPMS is significantly higher than conventional quadrupole ICPMS. The extreme element sensitivity of double-focusing sector field ICPMS permits ultratrace analysis down to the sub-fg mL^{-1} concentration range.[4] In Table 4 the detection limits for double-focusing ICPMS ''ELEMENT'' with plasma-shielded torch[4] and ''PlasmaTrace''[3] with ultrasonic nebulizer are compared. A calibration graph for ^{241}Am to determine the detection limit using the ICPMS ''ELEMENT'' with a plasma-shielded torch is shown in Figure 3.[4] For the determination of long-lived radionuclides the multielement capability is used as a main advantage of powerful ICPMS, which means the actinides are determined quasi-simultaneously by ICPMS. In contrast, if isobaric interferences of long-lived radionuclides with stable isotopes of impurities are expected, a matrix separation and/or an enrichment of long-lived radionuclides is required.

Yamamoto et al.[56] determined ^{99}Tc in environmental and radioactive waste samples, where there was the possibility of isobaric interference by $^{99}Ru^+$ (and $^{98}MoH^+$) ions. These authors suggested avoiding such interference by separating ^{99}Tc using different solvent extraction and purification techniques with anion exchange after leaching of geological samples. They determined ^{99}Tc with an absolute detection limit of 0.25 pg, corresponding to 0.16 mBq, using a double-focusing sector field ICPMS (PlasmaTrace, VG Elemental Ltd.) and analyzed sediment samples from the Irish Sea.

Eroglu and associates[10] investigated the enrichment and separation of ^{99}Tc from seawater by anion exchange and determined a detection limit (for a 14-mL sample volume) of 0.03 ng L^{-1} using a quadrupole-based ICPMS (HP 4500, Hewlett-Packard).

The application of mass spectrometry in trace, ultratrace and isotope analysis of long-lived radionuclides in biological, geological and environmental samples[57–60] is summarized in Table 5.

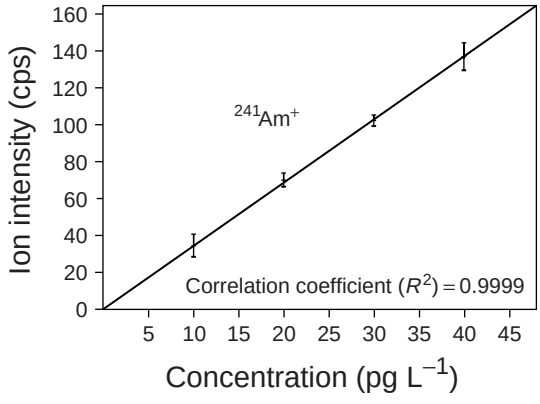

Figure 3 Calibration graph for $^{241}Am^+$ to determine the detection limit. (Reproduced with permission of Becker and Dietze.[4])

Table 4 Detection limits for several actinides in high-purity water by double-focusing sector field ICPMS

Nuclide	Half-life (years)	DF/ICPMS ''ELEMENT''[4]		DF/ICPMS ''PlasmaTrace'',[3]
		Concentration (pg L^{-1})	Activity (kBq L^{-1})	Concentration (pg L^{-1})
^{230}Th	7.5×10^4	0.08	6.2×10^{-8}	20
^{232}Th	1.4×10^{10}	0.1	2.2×10^{-13}	–
^{233}U	1.6×10^5	0.07	2.3×10^{-8}	–
^{238}U	4.5×10^9	0.2	2.4×10^{-12}	20
^{237}Np	2.1×10^{16}	0.05	1.3×10^{-9}	2
^{239}Pu	2.4×10^4	0.04	9.4×10^{-8}	5
^{241}Am	7.3×10^3	0.05	3.6×10^{-7}	–

Table 5 Application of mass spectrometry in trace, ultratrace and isotope analysis of long-lived radionuclides in biological, geological and environmental samples

Sample	Method and equipment	Radionuclides	Measured range	Limit of detection	Reference
Sediment samples, Fish samples	ICPQMS "VG PlasmaQuad PQ2" and ultrasonic nebulizer	^{99}Tc, ^{230}Th, ^{232}Th, ^{230}U, ^{238}U, ^{237}Np, ^{238}Pu, ^{239}Pu, ^{240}Pu, ^{241}Am	pg L^{-1} to μg L^{-1} range	4.0 pg L^{-1} (^{235}U) 1.2 pg L^{-1} (^{230}Th, ^{237}Pu, ^{239}Pu, ^{240}Pu, ^{241}Am)	Chiappini et al.[57]
Biological standard reference materials, Sea and river water	ICPQMS "VG PlasmaQuad PQ2", "ELAN 5000A and ETV" On-line solid phase extraction	^{238}U, ^{232}Th	Biological samples, ng g^{-1}; waters, ng mL^{-1} range	ETV/ICPQMS 0.9 pg mL^{-1} (^{238}U) 0.3 pg mL^{-1} (^{232}Th) ICPQMS (water): 5.4 pg mL^{-1} (^{238}U) 6.2 pg mL^{-1} (^{232}Th)	Truscott et al.[58]
Urine	SIMS "CAMECA IMS4F"	Isotope analysis ^{235}U/^{238}U; ^{239}Pu/^{242}Pu	^{235}U/^{238}U: 0.00023–0.0015 ^{239}Pu/^{242}Pu: 0.0064	0.1 ng L^{-1} (^{238}U) 0.01 ng L^{-1} (^{239}Pu)	Amaral et al.[6]
Soils, sediments	SIMS "CAMECA IMS3", TIMS Isotope dilution	Th, U	2.5–25 μg g^{-1} (U)	sub-μg g^{-1} range	Adrians et al.[5]
Meteorites (Chondrites)	ICPQMS	Th, U	10 ng g^{-1} to 1.7 μg g^{-1}	0.28 ng g^{-1} (Th)	Shinotsuka and Ebihara[59]
Seawater (Irish sea)	ICPQMS HP 4500 Flow injection (micro-column) preconcentration	^{99}Tc^{239}Pu/^{240}Pu	0.87 ng L^{-1} (^{99}Tc) 0.27–0.94 pg L^{-1} (^{239}Pu)	0.03 ng L^{-1} (^{99}Tc) 5 fg L^{-1} (^{239}Pu)	Eroglu et al.[10] Eroglu et al.[11]
River and lake water	ETV/ICPQMS "VG PlasmaQuad PQ2"	^{99}Tc, ^{238}U, ^{236}U, ^{232}Th, ^{230}Th, ^{226}Ra	1.5–6 ng L^{-1} (U) 1.4–26 ng L^{-1} (Tc)	0.6 fg (^{226}Ra) 0.9 fg (^{236}U)	Alvarado and Erickson[60]
Soil, sediment, grass	GDMS VG 9000	^{137}Cs, ^{90}Sr, ^{239}Pu, ^{240}Pu, ^{241}Pu, ^{232}Th	2 pg g^{-1} (^{137}Cs) 1 pg g^{-1} (^{90}Sr)	0.2 pg g^{-1} (^{137}Cs) 0.01 pg g^{-1} (^{90}Sr)	Betti et al.[30]
Soil, grass, milk, urine	RIMS after chemical separation	^{89}Sr, ^{90}Sr	8×10^8 atoms (^{90}Sr)	3×10^6 atoms (^{90}Sr)	Wendt et al.[38]

ETV, electrothermal vaporization.

2.2.1 Solution Introduction in Inductively Coupled Plasma Mass Spectrometry Using Micronebulizers

The development of microanalytical methods for the precise determination of nuclide abundances and concentration of long-lived radionuclides at ultratrace concentration levels in radioactive wastes and also for controlling contamination from radioactive waste in the environment is a challenging task. In order to analyze small sample volumes, micronebulizers (MCN-100, Cetac Technologies, USA and MicroMist, Glass Expansion, Australia)[61] and the DIHEN (J.E. Meinhard Associates, USA)[20] have been increasingly used instead of high solution volume consuming ultrasonic nebulizers for determining long-lived radionuclides by the solution introduction into the inductively coupled plasma of an ICPMS. Owing to direct nebulization of the sample solution in the inductively coupled plasma with the DIHEN (with an analyte transport efficiency into the plasma of 100%), and the reduction of the solution uptake rate and the sample size to the 1 μL min^{-1} and fg range, respectively, very sensitive measurements of long-lived radionuclides in aqueous solutions are possible.[20] Whereas with a quadrupole-based ICPMS using a micronebulizer the detection limits of long-lived radionuclides (^{230}Th, ^{237}Np, ^{238}U, ^{239}Pu and ^{241}Am) in aqueous solutions varied from 0.012 to 0.11 ng L^{-1},[20] in double-focusing sector field ICPMS with a shielded torch the detection limits are determined to be 0.0003–0.0008 ng L^{-1}.[4]

The determination of ^{99}Tc in a highly radioactive evaporator concentrate from a nuclear power plant[62] was performed after complex chemical sample preparation by DF/ICPMS with a shielded torch using the microconcentric MicroMist nebulizer with minicyclonic spray chamber. The detection limit with double-focusing ICPMS was determined as 5 pg L^{-1}, corresponding to an activity of 3 μBq mL^{-1}, in comparative measurements by quadrupole-based ICPMS (ELAN 6000, Perkin Elmer Sciex) the detection limit of ^{99}Tc was 0.1 ng L^{-1}.[4]

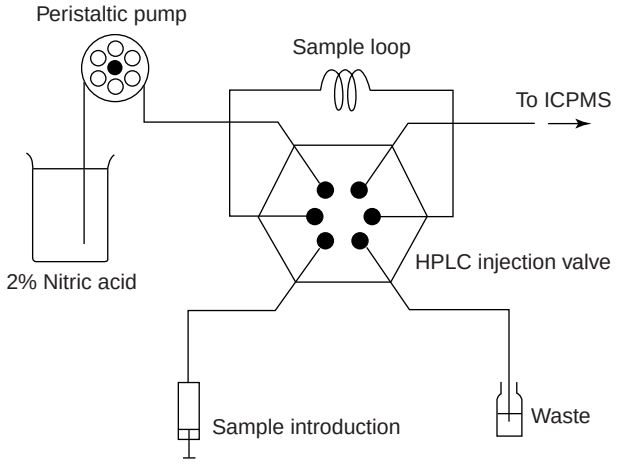

Figure 4 Sample introduction by flow injection. (Reproduced with permission of Becker and Dietze.[4])

2.2.2 *Ultratrace Analysis of Radionuclides in Small Sample Volumes*

The application of flow injection is useful for sample introduction to measure small sample volumes of radioactive waste solutions in order to minimize radioactive contamination in the ICPMS.[4,63] Figure 4 shows the principle of flow injection for solution introduction of small sample volumes. To handle microliter amounts of an aqueous solution of long-lived radionuclides a commercial high-performance liquid chromatography (HPLC) injection valve was coupled to a micronebulizer for small droplet formation and a minicyclonic spray chamber. The small sample volume (sample loop: $>1\,\mu L$) of standard or radioactive waste solution is introduced by this HPLC injection valve into a continuous flow of 2% nitric acid. For example, a ^{237}Np standard solution (sample loop: $20\,\mu L$; Np concentration: $10\,ng\,L^{-1}$ and $100\,ng\,L^{-1}$) was measured with a precision of 2.0% and 1.6% (RSD, $n = 5$), respectively.[4] The transient signals of $^{241}Am^+$ and $^{237}Np^+$ for $10\,ng\,L^{-1}$ and $100\,ng\,L^{-1}$ measured by DF/ICPMS with plasma-shielded torch are shown in Figure 5. Furthermore, the flow injection method was applied in our laboratory for isotope dilution analysis of Th in ICPMS. In this experiment a $2\,\mu g\,L^{-1}$ ^{232}Th solution (continuous flow) was spiked with $20\,\mu L$ of $5\,\mu g\,L^{-1}$ highly enriched ^{230}Th (99.85%) for quantitative Th determination in radioactive solutions.[61]

2.2.3 *Application of Flow Injection Technique, High-performance Liquid Chromatography and Capillary Electrophoresis for Coupling to Inductively Coupled Plasma Mass Spectrometry*

The most important problem in determining long-lived radionuclides in radioactive waste or environmental

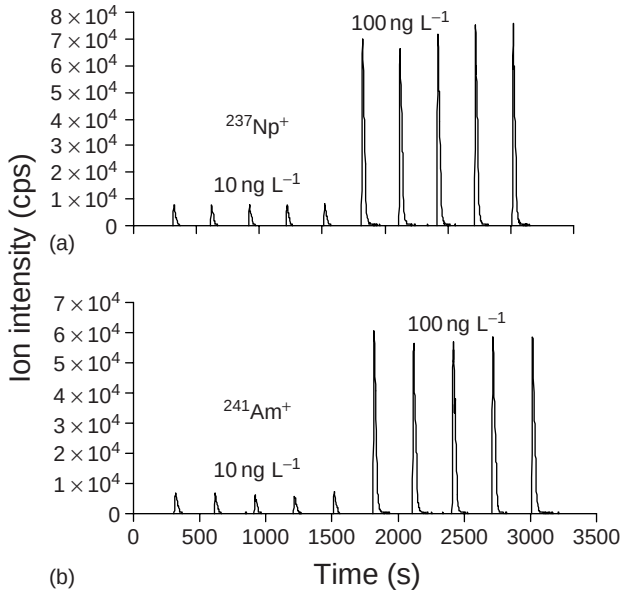

Figure 5 Transient signals of $^{241}Am^+$ and $^{237}Np^+$ (10 and $100\,ng\,L^{-1}$) measured by flow injection (sample loop, $20\,\mu L$). (Reproduced with permission of Becker and Dietze.[4])

samples is the appearance of isobars of radioactive and stable nuclides of the same mass but different atomic number, thus causing interference in the mass spectra. In order to separate isobars, such as long-lived ^{173}Lu from stable ^{173}Yb, on-line HPLC (for chromatographic separation of a lanthanide mixture into the individual elements) coupled to an ICPMS (for the mass spectrometric determination of the nuclides from each of the previously separated elements as a function of their mass) is the method of choice. In past years on-line ion chromatography ICPMS has been increasingly used for the characterization of radioactive materials and environmental samples using quadrupole-based ICPMS.[50,64,65] The use of ion chromatography for the determination of fission products and actinides in nuclear applications is reviewed by Betti.[17] On-line trace enrichment by flow injection using a microcolumn of activated alumina and mass spectrometric determination of uranium in mineral, river and seawater has been described by Dadfarnia and McLeod.[66]

Flow injection with on-line preconcentration using solid-phase adsorption on a mini-column of ^{99}Tc, ^{230}Th and ^{234}U at ultratrace levels in soils is described by Hollenbach et al.[67] Detection limits in the soil for ^{99}Tc, ^{230}Th and ^{234}U were $11\,mBq\,g^{-1}$ ($0.02\,ng\,g^{-1}$), $3.7\,mBq\,g^{-1}$ ($0.005\,ng\,g^{-1}$) and $0.74\,mBq\,g^{-1}$ ($0.003\,ng\,g^{-1}$) respectively.

An analytical procedure has been developed for the determination of spallation nuclides in an irradiated (with 800 MeV protons) Ta target of a spallation neutron source

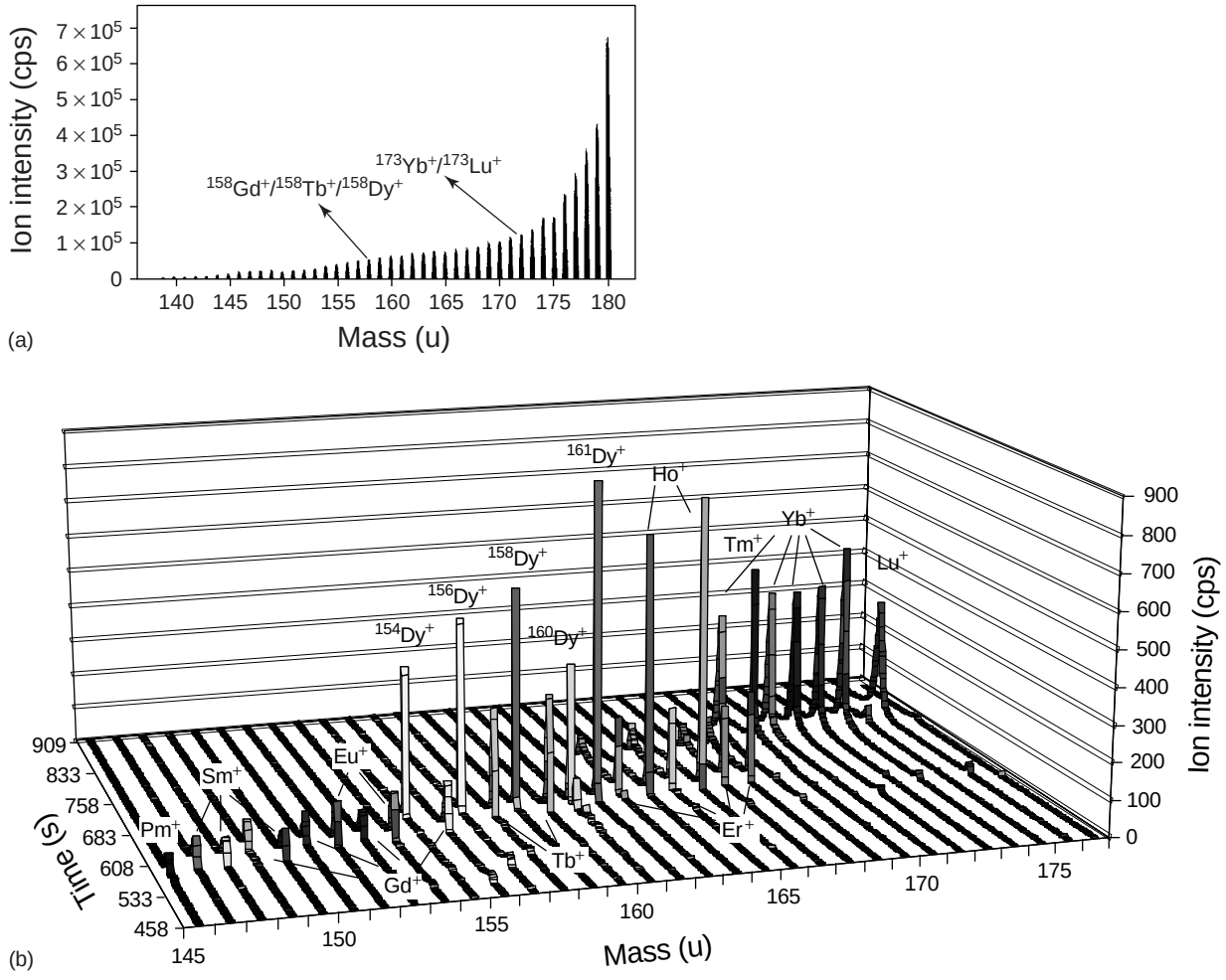

Figure 6 (a) Inductively coupled plasma mass spectrum of spallation nuclides in an irradiated tantalum target without lanthanide separation. (b) Mass spectrum of separated lanthanide elements (from Pm to Lu) in an irradiated tantalum target using on-line CE/ICPMS.

using HPLC coupled on-line to ICPMS.[58,68] Owing to expected isobaric interferences from long-lived radionuclides and stable isotopes in the irradiated Ta target, HPLC was useful for the chemical separation of analytes before the mass spectrometric measurements. The determination of spallation nuclides in irradiated Ta was performed by HPLC coupling on-line with double-focusing sector field ICPMS after dissolution of Ta in an HNO_3/HF mixture and matrix separation by liquid–liquid extraction of the Ta matrix (in order to reduce the high ^{182}Ta activity). The theoretical results of spallation yields of Ta are verified by measurements of the concentration of spallation nuclides of irradiated Ta.[68,69] A further reduction of the sample volume from 100 µL using HPLC/ICPMS in the nL range is interesting for the characterization of high-radioactive solution. Therefore, the on-line coupling of capillary electrophoresis with ICPMS was proposed in our laboratory in order to separate spallation lanthanides in radioactive solutions.[70] Figure 6 shows the ICP mass spectrum of spallation nuclides in the mass range of lanthanides in the irradiated Ta target (a) before lanthanide separation and by on-line CE/ICPMS on about 30 nL sample volume; (b) with lanthanide separation in the mass range from Pm to Lu using the double-focusing sector field ICPMS.

2.2.4 Precise Measurements of Isotope Ratios by Inductively Coupled Plasma Mass Spectrometry

Isotope ratio measurements of radiogenic elements are of great importance in the nuclear industry, where TIMS has occupied a favored position in the last few decades for routine measurements (e.g. quality assurance of fuel material; isotopic composition of U and Pu; reprocessing plant, nuclear material accounting and radioactive waste control).

Table 6 Application of mass spectrometry in isotopic measurements

Samples	Method and equipment	Isotopic ratios	Precision of isotopic analysis (at concentration)	Reference
Standard solutions Uranium metal Waste samples	ICPMS "ELEMENT"	$^{235}U/^{238}U$	0.07% (10 μg L^{-1}) 0.23% (100 ng L^{-1})	Kerl et al.[71]
Standard solutions Waste samples	ICPMS "ELEMENT"	$^{240}Pu/^{239}Pu$ $^{230}Th/^{232}Th = 10^{-3}$	2.0% (Pu: 0.1 ng L^{-1}, 50 fg) 0.34% (5 ng L^{-1})	Becker and Dietze[4]
Waste samples (silt, filter, glass, ash)	LA/ICPMS ELAN 6000	$^{235}U/^{238}U$ $^{230}Th/^{232}Th$	1.6% 1.3%	Becker and Dietze[4]
Standard solutions	ICPMS "Plasma 54" prototype	$^{235}U/^{238}U$	0.03% (1 mg L^{-1})	Taylor et al.[72]
Standard solutions	TIMS "VG 354"	$^{235}U/^{238}U$ $^{242}Pu/^{239}Pu$	0.0068% (100–200 ng U) 0.026% (10–20 ng Pu)	Callis and Abernathey[34]
Nuclear samples	ICPQMS "ELAN 5000" On-line ion chromatography	$^{135}Cs/^{138}Cs$ $^{134}Cs/^{138}Cs$ $^{136}Cs/^{138}Cs$	0.92% (50 μg L^{-1}) 1.38% (50 μg L^{-1}) 0.94% (50 μg L^{-1})	Barrero Moreno et al.[76]
Standard solutions	ICPQMS "ELAN 6000"	$^{235}U/^{238}U$	0.028% (10 μg L^{-1})	Platzner et al.[77]
Standard solutions	ICPMS "Plasma 54"	$^{235}U/^{238}U$	0.014% (U: 1 mg L^{-1})	Walder and Freedmann[78]
Geological samples	ICPMS "Plasma 54"	$^{230}Th/^{232}Th$ $^{234}U/^{238}U$ (0.000054)	0.04% 0.12%	Luo et al.[79]
Environmental samples (sediments, soil, pine needles)	ICPMS "HP 4500" "PMS-2000"	$^{240}Pu/^{239}Pu$	4.1% (at pg mL^{-1} level)	Momoshima et. al.[80]
Environmental samples (sediments)	ICPMS "PlasmaTrace 2"	$^{240}Pu/^{239}Pu$	2%	Stürup et al.[81]
Soil sample	DIHEN/ICPMS "ELAN 6000"	$^{235}U/^{238}U$ (0.022)	0.13% (U: 0.7 μg g^{-1})	Becker et al.[20]
Commercial uranium samples	TIMS "MAT 262-RPQ-Plus"	$^{235}U/^{238}U$ $^{234}U/^{238}U$, $^{236}U/^{238}U$	0.2–0.3% 0.4–3%	Richter et al.[82]
Environmental samples (tree bark, Sellafield)	ICPMS "HP 4500"	$^{235}U/^{238}U$ (0.0053–0.0064)	0.4–2%	Bellis et al.[83]
Spent fuels	ICPMS "VG PlasmaQuad PQ2" TIMS "Sector 54", HPLC	$^{242}Am/^{241}Am$ (0.003) $^{246}Cm/^{244}Cm$ (0.018)	1.3% (ICPMS) 0.7% (ICPMS)	Chartier et al.[16]

For example, Callis and Abernathey[34] developed a rapid high-precision analytical technique for the determination of isotope ratios of U and Pu by total sample volatilization using multiple-filament TIMS with a commercial multicollector instrument (VG-354, VG Isotopes). Run-to-run reproducibilities of <0.02% RSD have been obtained for isotope ratios of U and Pu. Inkret et al.[73] reported on applications of TIMS to the sensitive detection of ^{239}Pu and ^{240}Pu intakes. The determination of Pu concentration in urine samples yielded an average measurement uncertainty of 3.8 μBq per 24 h, a 40-fold improvement over the measurement uncertainties associated with radiochemistry/α-spectroscopy analytical methods.

The application of laser ablation multicollector ICPMS for in situ U and Th isotopic analysis to certified glass standards and naturally occurring opal is described by Stirling et al.[74]

At present TIMS is being increasingly replaced for precise isotope ratio measurements by ICPMS due to its excellent sensitivity and good RSD. For example, Agarande and colleagues investigated the long-term stability of low concentration U reference solutions by using ICPMS.[75] The excellent precision of isotope ratio measurements of long-lived radionuclides by ICPQMS and double-focusing sector field ICPMS has been demonstrated in different studies from our laboratory.[4,8,20,61,71,77] Selected applications of mass spectrometry for the determination of isotope ratios are summarized in Table 6.

3 CONCLUSIONS

In addition to solid-state mass spectrometric techniques, which allow the direct determination of long-lived

radionuclides in the picogram per gram range with a minimum of sample preparation, ICPMS, especially with on-line coupling techniques (ETV, HPLC, ion chromatography, flow injection) with detection limits in the femtogram per milliliter range is an excellent tool for the analysis of aqueous solutions. Numerous, quite different applications demonstrate the excellent capability of inorganic mass spectrometry for determining very low activity radionuclides, due to their low detection limits for the evidence of contamination from radioactive waste in the environment (in biological and medical samples, waters or geological materials).

The significance of ICPMS in precise isotope ratio measurements at ultratrace levels is increasing, especially when multicollector, and/or double-focusing sector field instruments are used.

ABBREVIATIONS AND ACRONYMS

AMS	Accelerator Mass Spectrometry
DF/ICPMS	Double-focusing Inductively Coupled Plasma Mass Spectrometry
DIHEN	Direct Injection High-efficiency Nebulizer
ETV	Electrothermal Vaporization
GDMS	Glow Discharge Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IRMM	Institute of Reference Materials and Measurements
LA/ICPMS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
Nd : YAG	Neodymium Yttrium Aluminum Garnet
NIST	National Institute of Standards and Technology
RIMS	Resonance Ionization Mass Spectrometry
RSD	Relative Standard Deviation
SIMS	Secondary Ion Mass Spectrometry
TIMS	Thermal Ionization Mass Spectrometry
TOF	Time-of-flight

RELATED ARTICLES

Mass Spectrometry **(Volume 13)**
High-resolution Mass Spectrometry and its Applications • Inorganic Substances, Mass Spectrometric in the Analysis of • Isotope Ratio Mass Spectrometry

Radiochemical Methods **(Volume 14)**
β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation • Speciation of Radionuclides in the Environment

REFERENCES

1. J.S. Crain, 'Applications of Inductively Coupled Plasma Mass Spectrometry in Environmental Radiochemistry', *Atom. Spectrosc. Perspectives*, **11**(2), 30–39 (1996).
2. J.S. Becker, H.-J. Dietze, 'Determination of Long-lived Radionuclides by Double-focusing Sector Field ICP Mass Spectrometry', *Adv. Mass Spectrom.*, **14**, 681–689 (1998).
3. C.K. Kim, R. Seki, S. Morita, S. Yamasaki, A. Tsumura, Y. Igarashi, M. Yamamoto, 'Application of High-resolution Inductively Coupled Plasma Mass Spectrometry to the Measurements of Long-lived Radionuclides', *J. Anal. Atom. Spectrom.*, **6**, 205–209 (1991).
4. J.S. Becker, H.-J. Dietze, 'Application of Double-focusing Sector Field ICP Mass Spectrometry with Shielded Torch Using Different Nebulizers for Ultratrace and Precise Isotope Analysis of Long-lived Radionuclides', *J. Anal. Atom. Spectrom.*, **14**, 1493–1500 (1999).
5. A.G. Adriaens, J.D. Fassett, W.R. Kelly, D.S. Simons, F.C. Adams, 'Determination of Uranium and Thorium Concentrations in Soils: Comparison of Isotope Dilution–Secondary Ion Mass Spectrometry and Isotope Dilution/Thermal Ionization Mass Spectrometry', *Anal. Chem.*, **64**, 2945–2950 (1992).
6. A. Amaral, C. Cossonnet, P. Galle, 'The Use of Secondary Ion Mass Spectrometry in Radiotoxicology', *Radiat. Prot. Dosim.*, **79**(1–4), 137–140 (1998).
7. E.H. Bailey, A.J. Kemp, V. Ragnarsdottir, 'Determination of Uranium and Thorium in Basalts and Uranium in Aqueous Solution by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **8**, 551–556 (1993).
8. J.S. Becker, H.-J. Dietze, 'Precise Isotope Ratio Measurements for Uranium, Thorium and Plutonium by Quadrupole-based Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **364**(6), 482–488 (1999).
9. N.E. Bibler, W.F. Kinard, W.T. Boyce, C.J. Coleman, 'Determination of Long-lived Fission and Actinides in Savannah River Site HLW Sludge and Glass for Waste Acceptance', *J. Radioanal. Nucl. Chem.*, **234**(1–2), 159–163 (1998).
10. A.E. Eroglu, C.W. McLeod, K.S. Leonard, D. McCubbin, 'Determination of Technetium in Seawater Using Ion Exchange and Inductively Coupled Plasma Mass Spectrometry with Ultrasonic Nebulization', *J. Anal. Atom. Spectrom.*, **13**, 875–878 (1998).

11. A.E. Eroglu, C.W. McLeod, K.S. Leonard, D. McCubbin, 'Determination of Plutonium in Seawater Using Co-precipitation and Inductively Coupled Plasma Mass Spectrometry with Ultrasonic Nebulization', *Spectrochim. Acta, Part B*, **53**, 1221–1233 (1998).
12. H.N. Edmonds, S.B. Moran, J.A. Hoff, J.N. Smith, R.L. Edwards, 'Protactinium-231 and Thorium-230 Abundances and High Scavenging Rates in the Western Arctic Ocean', *Science*, **280**, 405–407 (1998).
13. M. Nunnemann, N. Erdmann, H.-U. Hasse, G. Huber, J.V. Kratz, P. Kunz, A. Mansel, G. Passler, O. Stetzer, N. Trautmann, A. Waldek, 'Trace Analysis of Plutonium in Environmental Samples by Resonance Ionization Mass Spectroscopy (RIMS)', *J. Alloys Comp.*, **271–273**, 45–48 (1998).
14. P. Roth, W. Werner, I. Wendler, P. Schramel, 'Variation of Natural ^{232}Th Excretion in Non-exposed Persons', *J. Radioanal. Nucl. Chem.*, **226**(1–2), 285–287 (1997).
15. A. Stroh, F. Bea, P.G. Montero, 'Ultratrace-level Determination of Rare Earth Elements, Thorium, and Uranium in Ultramafic Rocks by ICP/MS', *Atomic Spectrosc.*, **16**(1), 7–11 (1995).
16. F. Chartier, M. Aubert, M. Pilier, 'Determination of Am and Cm in Spent Nuclear Fuels by Isotope Dilution Inductively Coupled Plasma Mass Spectrometry and Isotope Dilution Thermal Ionization Mass Spectrometry After Separation by High-performance Liquid Chromatography', *Fresenius' J. Anal. Chem.*, **364**, 320–327 (1999).
17. M. Betti, 'Use of Ion Chromatography (ICP/MS) for the Determination of Fission Products and Actinides in Nuclear Applications', *J. Chromatogr. A*, **789**, 369–379 (1997).
18. W. Kerl, J.S. Becker, W. Dannecker, H.-J. Dietze, 'Application of On-line HPLC/ICP/MS for the Determination of the Nuclide Abundances of Lanthanides Produced via Spallation Reactions in an Irradiated Tantalum Target of a Spallation Neutron Source', *Fresenius' J. Anal. Chem.*, **362**, 433–439 (1998).
19. J.I. Garcia Alonso, 'Determination of Fission Products and Actinides by Inductively Coupled Plasma Mass Spectrometry Using Isotope Dilution Analysis: A Study of Random and Systematic Errors', *Anal. Chim. Acta*, **312**, 57–78 (1995).
20. J.S. Becker, H.-J. Dietze, J.A. McLean, A. Montaser, 'Ultratrace and Isotope Analysis of Long-lived Radionuclides by Inductively Coupled Quadrupole Mass Spectrometry Using a Direct Injection High Efficiency Nebulizer', *Anal. Chem.*, **71**, 3077–3084 (1999).
21. B. Beer, K.G. Heumann, 'Trace Analyses of U, Th and Other Heavy Metals in High-purity Aluminum with Isotope Dilution Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **343**, 741–745 (1992).
22. P. Herzner, K.G. Heumann, 'Ultratrace Analysis of U, Th, Ca and Selected Heavy Metals in High-purity Refractory Metals with Isotope Dilution Mass Spectrometry', *Microchim. Acta*, **106**, 127–135 (1992).
23. M. Fukuda, Y. Sayama, 'Determination of Traces of Uranium and Thorium in (Ba, Sr)TiO_3 Ferroelectrics by Inductively Coupled Plasma Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **357**, 647–651 (1997).
24. U. Rohr, L. Meckel, H.M. Ortner, 'Ultratrace Analysis of Uranium and Thorium in Glass; Part 1: ICPMS, Classical Photometry and Chelate-GC', *Fresenius' J. Anal. Chem.*, **348**, 356–363 (1994).
25. J.S. Crain, L.L. Smith, J.S. Yaeger, J.A. Alvarado, 'Determination of Long-lived Actinides in Soil Leachates by Inductively Coupled Plasma Mass Spectrometry', *J. Radioanal. Nucl. Chemistry*, **194**, 133 (1995).
26. J.I. Garcia Alonso, D. Thoby-Schultzendorff, B. Giovanonne, J.-P. Glatz, G. Pagliosa, L. Koch, 'Characterization of Spent Nuclear Fuel Dissolver Solutions and Dissolution Residues by Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **9**, 1209–1215 (1994).
27. J.S. Becker, M. Gastel, D. Tenzler, H.-J. Dietze, 'LA/ICP/MS for Trace Analysis of Long-lived Radionuclides in Solid Non-conducting Radioactive Waste Samples', *Adv. Mass Spectrom.*, 14, MoPo141 (1998).
28. M. Gastel, J.S. Becker, H.-J. Dietze, 'Determination of Long-lived Radionuclides in Solid Non-conducting Radioactive Waste Materials by LA/ICP/MS', *Spectrochimica Acta B*, **52B**, 2051–2059 (1997).
29. T. Hirata, 'Soft Ablation Technique for Laser Ablation–Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **12**, 1337–1342 (1997).
30. M. Betti, S. Giannarelli, T. Hiernaut, G. Rasmussen, L. Koch, 'Detection of Trace Radioisotopes in Soils, Sediments and Vegetations by Glow Discharge Mass Spectrometry', *Fresenius' Z. Anal. Chem.*, **355**, 642–646 (1996).
31. M. Betti, 'Use of a Direct Current Glow Discharge Mass Spectrometer for the Chemical Characterization of Samples of Nuclear Concern', *Anal. Atom. Spectrom.*, **11**, 855–860 (1996).
32. M. Betti, G. Rasmussen, L. Koch, 'Isotopic Abundance Measurements on Solid Nuclear-type Samples by Glow Discharge Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **355**, 808–812 (1996).
33. A. Amaral, P. Galle, C. Cossonnet, D. Franck, P. Pihet, M. Carrier, O. Stephan, 'Perspectives of Uranium and Plutonium Analysis in Urine Samples by Secondary Ion Mass Spectrometry', *J. Radioanal. Nucl. Chem.*, **226**(1–2), 41–45 (1997).
34. E.L. Callis, R.M. Abernathey, 'High-precision Isotopic Analyses of Uranium and Plutonium by Total Sample Volatilization and Signal Integration', *Intern. J. Mass Spectrom. Ion Proc.*, **103**, 93–105 (1991).
35. I.T. Platzner, 'Modern Isotope Ratio Mass Spectrometry', in *Chem. Anal.*, John Wiley, Chichester, 1–530, Vol. 145, 1997.
36. J. Rucklidge, 'Accelerator Mass Spectrometry in Environmental Geoscience', *Analyst*, **120**, 1283–1290 (1995).
37. M. Nunnemann, N. Erdmann, H.-U. Hasse, G. Huber, J.V. Kratz, P. Kunz, A. Mansel, G. Passler, O. Stetzer,

N. Trautmann, A. Waldek, 'Trace Analysis of Plutonium in Environmental Samples by Resonance Ionization Mass Spectroscopy (RIMS)', *J. Alloys Comp.*, **271–273**, 45–48 (1998).

38. K. Wendt, J.V. Kratz, J. Lantzsch, P. Müller, W. Northershäuser, A. Seibert, N. Trautmann, W. Waldek, K. Zimmer, 'Rapid Ultratrace Determination of ^{89,90}Sr in Environmental Samples by Collinear Laser Resonance Ionization Spectrometry', *Kerntechnik*, **62**, 2–3 (1997).
39. K. Habfast, 'Fractionation Correction and Multiple Collectors in Thermal Ionization Isotope Ratio Mass Spectrometry', *Intern. J. Mass Spectrom. Ion Proc.*, **176**, 133–148 (1998).
40. B. Beer, K.G. Heumann, 'Trace Analyses of U, Th and Other Heavy Metals in High-purity Aluminum with Isotope Dilution Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **343**, 741–745 (1992).
41. P. Herzner, K.G. Heumann, 'Ultra-trace Analysis of U, Th, Ca and Selected Heavy Metals in High-purity Refractory Metals with Isotope Dilution Mass Spectrometry', *Mikrochim. Acta*, **106**, 127–135 (1992).
42. P. Herzner, K.G. Heumann, 'Trace Determination of Uranium, Thorium, Calcium, and Other Heavy Metals in High-purity Refractory Metal Silicides, Niobium and Silicon Dioxide with Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **64**, 2942–2944 (1992).
43. K.G. Heumann, 'Isotope Dilution Mass Spectrometry (IDMS) of the Elements', *Mass Spectrom. Reviews*, **11**, 41–67 (1992).
44. S.K. Aggarwal, H.C. Crain, 'Thermal Ionization Mass Spectrometry', in *Introduction to Mass Spectrometry*, eds. S.K. Aggarwal, H.C. Crain, Indian Society for Mass Spectrometry (ISMAS), Mumbai, 233–250, 1997.
45. B. Beer, K.G. Heumann, 'Trace Analysis of Microelectronically Relevant Heavy Metals in High-purity Titanium with Isotope Dilution Mass Spectrometry', *Anal. Chem.*, **65**, 3199–3203 (1993).
46. V.K. Panday, J.S. Becker, H.-J. Dietze, 'Determination of Trace Impurities in Tantalum by Inductively Coupled Plasma Mass Spectrometry After Removal of the Matrix by Liquid–Liquid Extraction', *Anal. Chim. Acta*, **329**, 153–159 (1996).
47. B. Beer, K.G. Heumann, 'Isotope Dilution Mass Spectrometry of Microelectronically Relevant Heavy Metal Traces in High-purity Cobalt', *Fresenius' J. Anal. Chem.*, **347**, 351–355 (1993).
48. V.K. Panday, J.S. Becker, H.-J. Dietze, 'Trace Impurities in Zircaloys by Inductively Coupled Plasma Mass Spectrometry After Removal of the Matrix by Liquid–Liquid Extraction', *Atom. Spectrom.*, May/June, 97–100 (1995).
49. H.W. Harrison, 'Glow Discharge Mass Spectrometry', in *Inorganic Mass Spectrometry*, eds. F. Adams, R. Gijbels, R. Van Grieken, *Chem. Anal.*, John Wiley, New York, 85–123, Vol. 95, 1988.
50. J.M. Barrero Moreno, M. Betti, J.I. Garcio Alonso, 'Determination of Neptunium and Plutonium in the Presence of High Concentrations of Uranium by Ion Chromatography/Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **12**, 355–361 (1997).
51. N. Erdmann, G. Herrmann, G. Huber, S. Köhler, J.V. Kratz, A. Mansel, M. Nunnemann, G. Passler, N. Trautmann, A. Turchin, A. Waldek, 'Resonance Ionization Mass Spectroscopy for Trace Determination of Plutonium in Environmental Samples', *Fresenius' J. Anal. Chem.*, **359**, 378–381 (1997).
52. X.L. Zhao, L.R. Kilius, A.E. Litherland, T. Beasley, 'AMS Measurement of Environmental U-236 Preliminary Results and Perspectives', *Nucl. Instrum. Meth. B*, **126**, 297–300 (1997).
53. L.K. Fifield, R.G. Cresswell, M.L. di Tada, T.R. Ophel, J.P. Day, A.P. Clacher, S.J. King, N.D. Priest, 'Accelerator Mass Spectrometry of Plutonium Isotopes', *Nucl. Instrum. Meth. B*, **117**, 295–303 (1996).
54. A. Montaser (ed.), *Inductively Coupled Plasma Source Mass Spectrometry*, VCH Publishers, New York, 1998.
55. J.S. Becker, H.-J. Dietze, 'Inorganic Trace Analysis by Mass Spectrometry', *Spectrochimica Acta*, **53B**, 1475–1506 (1998).
56. M. Yamamoto, K. Syarbaini, K. Kofuji, A. Tsumura, K. Komura, K. Ueno, D.J. Assinder, 'Determination of Low-level ^{99}Tc Environmental Samples by High Resolution ICPMS', *J. Radioanal. Nucl. Chem.*, **197**, 185 (1995).
57. R. Chiappini, J.-M. Taillade, S. Brébion, 'Development of a High-sensitivity Inductively Coupled Plasma Mass Spectrometer for Actinide Measurement in the Femtogram Range', *J. Anal. Atom. Spectrom.*, **11**, 497–503 (1996).
58. J.B. Truscott, L. Bromley, P. Jones, E.H. Evans, J. Turner, B. Fairman, 'Determination of Natural Uranium and Thorium in Environmental Samples by ETV/ICP/MS After Matrix Removal by On-line Solid Phase Extraction', *J. Anal. Atom. Spectrom.*, **14**, 627–631 (1999).
59. K. Shinotsuka, M. Ebihara, 'Precise Determination of Rare Earth Elements, Thorium and Uranium in Chondritic Meteorites by Inductively Coupled Plasma Mass Spectrometry: A Comparative Study with Radiochemical Neutron Activation Analysis', *Anal. Chim. Acta*, **338**, 237–246 (1997).
60. J.S. Alvarado, M.D. Erickson, 'Determination of Long-lived Radioisotopes Using Electrothermal Vaporization–Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **11**, 923–928 (1996).
61. J.S. Becker, R.S. Soman, K.L. Sutton, J. Caruso, H.-J. Dietze, 'Determination of Long-lived Radionuclides by Inductively Coupled Plasma Mass Spectrometry Using Different Nebulizers', *J. Anal. Atom. Spectrom.*, **14**, 933–937 (1999).
62. E. Hertendi, Z. Szücs, J. Csongor, J. Gulyás, E. Svingor, P. Ormai, A. Fritz, J. Solymosi, I. Gresitis, N. Vajda, Z. Molnar, 'Application of Scaling Technique for Estimation of Radionuclide Inventory in Radioactive Waste',

406–463, Nuclear Society of Slovenia, Portorož. 'Proceedings of 3rd Regional Meeting: Nuclear Energy in Central Europe', 16–19 September 1996, Portorož, Slovenia.
63. W. Kerl, 'Ultraspuren- und Isotopenanalyse langlebiger Raduionuklide mittels doppel- fokussierender Sektorfeld-ICP-Massenspektrometrie', *Berichte des Forschungszentrum Jülich*, **3605**, 1–22 (1998).
64. J.M. Barrero Moreno, J.I.G. Alonso, P. Arbore, G. Nicolaou, L. Koch, 'Characterization of Spent Nuclear Fuels by Ion Chromatography–Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **11**, 929–935 (1996).
65. Y. Marumatsu, S. Uchida, K. Tagami, S. Yoshida, T. Fujikawa, 'Determination of Plutonium Concentration and Its Isotopic Ratio in Environmental Materials by ICP/MS After Separation Using Ion-exchange and Extraction Chromatography', *J. Anal. Atom. Spectrom.*, **14**, 859–865 (1999).
66. S. Dadfarnia, C. McLeod, 'On-line Trace Enrichment and Determination of Uranium in Waters by Flow Injection Inductively Coupled Plasma Mass Spectrometry', *Appl. Spectrosc.*, **48**, 1331–1335 (1994).
67. M. Hollenbach, J. Grohs, S. Mamich, M. Kroft, E.R. Denoyer, 'Determination of Technetium-99, Thorium-230 and Uranium-234 in Soils by Inductively Coupled Plasma Mass Spectrometry Using Flow Injection Preconcentration', *J. Anal. Atom. Spectrom.*, **9**, 927–933 (1994).
68. W. Kerl, J.S. Becker, H.-J. Dietze, W. Dannecker, 'Application of On-line HPLC/ICPMS for the Determination of the Nuclide Abundances of Lanthanides Produced via Spallation Reactions in an Irradiated Tantalum Target of a Spallation Neutron Source', *Fresenius' J. Anal. Chem.*, **362**, 433–439 (1998).
69. J.S. Becker, W. Kerl, H.-J. Dietze, 'Nuclide Analysis of an Irradiated Tantalum Target of a Spallation Neutron Source Using High Performance Ion Chromatography and Inductively Coupled Plasma Mass Spectrometry', *Anal. Chim. Acta*, **387**, 145–154 (1999).
70. J. Day, S. Becker, H.-J. Dietze, unpublished results.
71. W. Kerl, J.S. Becker, H.-J. Dietze, W. Dannecker, 'Isotopic and Ultratrace Analysis of Uranium by Double-focusing Sector Field ICP Mass Spectrometry', *Fresenius' J. Anal. Chem.*, **359**, 407–409 (1998).
72. P.D.P. Taylor, P. De Bievre, A.J. Walder, A. Entwistle, 'Validation of the Analytical Linearity and Mass Discrimination Correction Model Exhibited by a Multiple Collector Inductively Coupled Plasma Mass Spectrometry by Means of a Set of Synthetic Uranium Isotope Mixtures', *J. Anal. Atom. Spectrom.*, **10**, 395–398 (1995).
73. W.C. Inkret, D.W. Efurd, G. Miller, D.J. Rokop, T.M. Benjamin, 'Application of Thermal Ionization Mass Spectrometry to the Detection of ^{239}Pu and ^{240}Pu Intakes', *Intern. J. Mass Spectrom.*, **178**, 113–120 (1998).
74. C.H. Stirling, J.N. Christensen, A.N. Halliday, 'In-situ ^{238}U–^{234}U–^{230}Th Isotopic Analysis by Laser Ablation Multiple Collector ICPMS', *Mineralogical Magazine*, **62A**, 1462–1463 (1998).
75. M. Agarande, K. Mayer, P. De Bièvre, 'Verification of the Long Term Stability of Uranium Reference Solutions of Low Concentration', *J. Radioanal. Nucl. Chem.*, **207**(2), 395–400 (1996).
76. J.M. Barrero Moreno, M. Betti, G. Nicolaou, 'Determination of Caesium and Its Isotopic Composition in Nuclear Samples Using Isotope Dilution–Ion Chromatography–Inductively Coupled Plasma Mass Spectrometry', *J. Anal. Atom. Spectrom.*, **14**, 875–879 (1999).
77. I.T. Platzner, J.S. Becker, H.-J. Dietze, 'Stability Study of Isotope Ratio Measurements for Uranium and Thorium by ICP/QMS', *Atom. Spectrosc.*, **20**(1), 6–12 (1999).
78. A.J. Walder, P.A. Freedman, 'Isotopic Ratio Measurement Using a Double-focusing Magnetic Sector Mass Analyzer with an Inductively Coupled Plasma as an Ion Source', *J. Anal. Atom. Spectrom.*, **7**, 571–575 (1992).
79. X. Luo, M. Rehkämper, D-Ch. Lee, A.N. Halliday, 'High Precision ^{230}Th/^{232}Th and ^{234}U/^{238}U Measurements Using Energy-filtered ICP Magnetic Sector Multiple Collector Mass Spectrometry', *Intern. J. Mass Spectrom. Ion Proc.*, **171**, 105–117 (1997).
80. N. Momoshima, H. Kakiuchi, Y. Maeda, E. Hirai, T. Ono, 'Identification of the Contamination Source of Plutonium in Environmental Samples with Isotopic Ratios Determined by Inductively Coupled Plasma Mass Spectrometry and Alpha-spectrometry', *J. Radioanal. Nucl. Chem.*, **221**(1–2), 213–217 (1997).
81. S. Stürup, H. Dahlgaard, S.C. Nielsen, *J. Anal. Atom. Spectrom.*, 'High Resolution Inductively Coupled Plasma Mass Spectrometry for the Trace Determination of Plutonium Isotopes and Isotope Ratios in Environmental Samples', *J. Anal. Atom. Spectrom.*, **13**, 1321–1326 (1998).
82. S. Richter, A. Alonso, R. Wellum, P.D.P. Taylor, 'The Isotopic Composition of Commercially Available Uranium Chemical Reagents', *J. Anal. Atom. Spectrom.*, **14**, 889–891 (1999).
83. D. Bellis, R. Ma, C. McLeod, 'Isotopic Analysis of Uranium in Tree Bark', in O. Douard (ed.) *1999 Winter Conference on Plasma Spectrochemistry*, University of Pau, 10th–15th January 1999, Book of Abstracts, p. 164.

Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides

Rolf J. Rosenberg
Technical Research Centre of Finland, Finland

1 Introduction 1
2 Principles of Neutron Activation Analysis 1
3 Neutron Activation Analysis of Radionuclides 2
3.1 General 2
3.2 Technetium-99 2
3.3 Iodine-129 3
3.4 Cesium-135 3
3.5 Neptunium-237 3
4 Standards 4
5 Quality Assurance 4
6 Conclusions 5
Abbreviations and Acronyms 5
Related Articles 5
References 5

This article describes the use of neutron activation analysis (NAA) for the quantitative determination of very long-lived radionuclides in the environment and nuclear waste. Because of its sensitivity and isotope specificity NAA is used for the determination of a few radionuclides with half-lives of the order of 10^4 years or longer. The detection limits for some nuclides are lower than those obtainable with most kinds of mass spectrometry. The method has a substantial freedom from systematic errors.

The benefit of NAA, compared with direct radiometric measurement of the parent nuclide, is based on the fact that neutron activation may produce radionuclides with shorter half-lives and higher specific activities than those of the parent nuclides. An especially significant improvement is obtained when a γ-emitting nuclide is produced from a nuclide which only emits β-particles.

There are a few long-lived radionuclides which need to be determined because they are produced in nuclear fuel, and may pose a hazard in the final disposal of nuclear waste. Of these possible applications, the determination of ^{99}Tc, ^{129}I, ^{135}Cs and ^{237}Np is described here. NAA has proven to be the most sensitive method to determine ^{129}I and a useful method to determine ^{237}Np in environmental samples. Therefore, NAA can be recommended as an analytical method for these nuclides, when a research reactor is available.

1 INTRODUCTION

This article describes the use of NAA for the quantitative determination of very long-lived radionuclides in the environment and nuclear waste. NAA is mainly known as a reliable trace element analysis method for different kinds of samples, but because of its sensitivity and isotope specificity it has also found applications in the determination of a few radionuclides with half-lives of the order of 10^4 years or longer, namely ^{99}Tc, ^{129}I, ^{135}Cs and ^{237}Np.

NAA can be dated to the time of Hevesy and Levi,[1] who published their first paper of the method in 1936. Following the development of nuclear reactors in the early 1950s, the possibilities of applying NAA to trace element analysis were recognized. The invention of high-resolution Ge(Li) detectors in the 1960s and more recent advances in computers and automation during the 1970s and 1980s have made possible application of NAA to research studies involving large numbers of samples.

However, the need for a research reactor for irradiation and the advances in new types of mass spectrometers have decreased the use of NAA in the 1990s. When a research reactor is available, NAA still has some use for the determination of a few long-lived radionuclides, for the following reasons. The detection limits for some nuclides are lower than those obtainable with most kinds of mass spectrometry. The method has a substantial freedom from systematic errors. The physical processes are well understood. High-resolution γ-ray spectroscopy allows the qualitative identification of the nuclides present, and also their quantitative determination. The density of γ-lines is small, allowing their separation, and the energy is high, rendering matrix errors small. The method is complementary to other methods commonly in use, which are therefore likely to give a different bias in the result. The method is not sensitive to contamination, because short-lived nuclides are measured.

2 PRINCIPLES OF NEUTRON ACTIVATION ANALYSIS

The sample to be analyzed is irradiated with neutrons in a nuclear research reactor. The nuclides in the sample undergo nuclear reactions in which the nucleus absorbs a neutron and practically simultaneously some excitation

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

energy in the form of γ-radiation is emitted. This is called an (n,γ)-reaction. Other reactions also occur but they are of lesser significance in NAA. The product nuclide may be either radioactive or stable. The activity of a produced radionuclide is given by Equation (1):

$$A = N[\varphi_{th}\sigma_{th} + \varphi_{epi}I]\left(1 - e^{-\frac{\ln 2 t_i}{T}}\right)\left(e^{\frac{\ln 2 t_d}{T}}\right) \quad (1)$$

where A = activity in disintegrations per second (Bq), N = number of target atoms, φ_{th} = thermal neutron flux, φ_{epi} = epithermal neutron flux, σ_{th} = average thermal neutron cross-section, I = effective resonance integral or epithermal cross-section, T = half-life of product radionuclide, t_i = irradiation time and t_d = decay time between the end of irradiation and middle of measurement. The half-life and the irradiation and decay times have to be expressed in the same units. The neutron flux in the equation is expressed as the number of neutrons per square centimeter per second. The cross-section is expressed in square centimeters; in tabulations it is expresses in barns (1 barn = 10^{-24} cm^2).

New radionuclides with half-lives ranging from seconds to years are produced in these reactions. The activity of the radionuclide produced is measured using high-resolution γ-spectrometry, as described in the article **γ-Spectrometry, High-resolution, for Radionuclide Determination**. The amount of the original radionuclide (number of nuclides, N) can be calculated from Equation (1) when the physical parameters and the detector efficiency as a function of γ-ray energy are known. It can also be calculated from Equation (2) when a standard with a known concentration of the same radionuclide is irradiated together with the sample and measured under the same conditions as the sample:

$$N_s = \frac{N_{st}A_s e^{-\frac{\ln 2 t}{T}}}{A_{st}} \quad (2)$$

where t is the time elapsed between the measurements of the standard and the sample.

3 NEUTRON ACTIVATION ANALYSIS OF RADIONUCLIDES

3.1 General

The benefit of NAA compared with direct radiometric measurement of the parent nuclide is based on the following. For some radionuclides the analytical characteristics can be improved by neutron activation, which produces radionuclides with shorter half-lives than those of the parent nuclides. These have higher specific activities than the parent nuclide, and if the reaction cross-section and other nuclear characteristics are favorable a considerable improvement can be obtained. An especially significant improvement is obtained when a γ-emitting nuclide is produced from a nuclide which only emits β-particles. The reason for this is that the measurement of γ-radiation is more specific than the measurement of β-radiation, which does not have a discrete energy.

Table 1 Improvement of specific activity by neutron activation

Reaction	Half-life of target nuclide (years)	Improvement of specific activity (product/original)
^{99}Tc(n,γ)^{100}Tc	2.1×10^5	2600
^{126}Sn(n,γ)^{127}Sn	$\sim 10^5$	14
^{129}I(n,γ)^{130}I	1.57×10^7	10^6
^{135}Cs(n,γ)^{136}Cs	2×10^6	350
^{226}Ra(n,γ)^{227}Ra	1600	12
^{237}Np(n,γ)^{238}Np	2.14×10^6	3.5×10^5
^{242}Pu(n,γ)^{243}Pu	3.75×10^5	8600

There are a few long-lived radionuclides which need to be determined because they are produced in nuclear fuel, and may pose a hazard in the final disposal of nuclear waste. Neutron activation produces nuclides with shorter half-lives from some of these radionuclides. Table 1 shows the reactions and the improvement in the specific activities when the sample is irradiated for 15 h in a neutron flux of 10^{13} n cm^{-2} s^{-1}.[2]

Of these possible applications, only the determination of ^{99}Tc, ^{129}I, ^{135}Cs and ^{237}Np has been reported and only the determination of ^{129}I and ^{237}Np has been of substantial significance in research.

3.2 Technetium-99

Theoretically, NAA is more sensitive than β-counting for the determination of ^{99}Tc, although in practical work poorer detection limits are observed. The detection limit is not low enough for the analysis of environmental samples, but because NAA has the advantage over β-counting of being specific, it is used for the analysis of nuclear waste.

The NAA of ^{99}Tc can be based on two reactions, ^{99}Tc(n,γ)^{100}Tc and ^{99}Tc(n,n$'$)^{99m}Tc.[3–6] Another method of producing ^{99m}Tc is by the (γ,γ') reaction. The half-life of ^{100}Tc is 15.5 s and it emits γ-rays of 540 and 591 keV with low intensity (7 and 6%, respectively). ^{99m}Tc, on the other hand, has a half-life of 6 h and emits a γ-ray of 140 keV with an intensity 87%. When counting ^{100}Tc, which has a very short half-life, ^{99}Tc has to be separated from the sample prior to irradiation.

For NAA a preirradiation separation procedure may be used. Here the procedure of Bate[3] is described briefly.

^{99m}Tc or ^{95m}Tc is added as tracer for yield determination. Technetium is extracted into cyclohexanone and then back-extracted into water by diluting the cyclohexanone with water and carbon tetrachloride. The water solution is made basic and the Tc is adsorbed on Bio-Rad 1-X1 ion-exchange resin in a small column suitable for neutron irradiation. The resin is irradiated for 11 s and counted for 30 s.

When ^{99m}Tc is counted, the sample can be directly irradiated and followed by a radiochemical separation.[6] No papers on the analysis of real samples using this method have been published.

3.3 Iodine-129

NAA is a significantly more sensitive method than particle counting for the determination of ^{129}I.[7] It has been stated that NAA is 10^6 times more sensitive than β-counting.[8] The best published detection limit is 0.15×10^{-4} mBq. As a consequence, NAA seems to be the most commonly used method for the analysis of ^{129}I as indicated by the large number of publications on the subject.

The analysis is based on the reaction ^{129}I(n,γ)^{130}I. The half-life of ^{130}I is 12.4 h. The main γ-ray is of 536 keV to the extent of 99%. Other prominent lines are 418, 668 and 734 keV. Because of the extremely low concentrations of ^{129}I in environmental samples, the production of ^{129}I through reactions other than (n,γ) is a potential source of error. The production of ^{129}I via the threefold reaction ^{127}I(n,γ)^{128}I(n,γ)^{129}I(n,γ)^{130}I is a potential source of error. The effect of this error has to be controlled and therefore the ^{129}I/^{127}I ratio always has to be determined. This can be done by determining ^{127}I through the ^{127}I(n,γ)^{128}I reaction. The half-life of ^{128}I is 25 min. A more convenient method is to use the reaction ^{127}I(n,2n)^{126}I because the half-life of ^{126}I is 13 days and it can be measured simultaneously with ^{130}I. This is possible when the fast flux density is sufficiently high. Other reactions producing ^{130}I activity also occur. To avoid these errors, ^{130}I (or I) has to be separated from Cs, U and Te prior to the irradiation. Br is especially troublesome owing to the similarity of its chemical properties.

Because of the sources of error described above, a radiochemical separation is almost always required for the NAA determination of ^{129}I, regardless of the sample type. The lower the concentrations, the more elaborate are the separations needed. The separation can be divided into three parts: (1) extraction of ^{129}I from the sample, (2) preirradiation purification and (3) postirradiation purification before counting. When a moderate detection limit is required, step 3 may be omitted.

Soil and vegetation samples may be directly heated in a stream of oxygen in a quartz tube. The evaporated iodine is either adsorbed on activated charcoal[7,9–11] or in a solution of KOH and K_2SO_3.[12] Iodine can also be distilled from a solution as elemental iodine.[13–16] Ion exchange can also be used as a step in the separation of iodine.[13,17] The yield is determined either by adding stable I carrier or ^{125}I tracer. The distillation or combustion step is usually followed by the classical extraction cycle based on oxidation/reduction using CCl_4 as an organic solvent.[10,12,14,17] One novel method is to separate I from Br in the gas phase using a hydrated manganese dioxide (HMD) column.[7,11] The final sample for irradiation can be prepared by evaporation of the sample on to a foil,[12] adsorption on an ion-exchange resin[17] or activated charcoal,[7,11,16] sample irradiation in the liquid form[14] or by evaporation of elemental iodine into a quartz ampule.[10]

Stable iodine is usually added as a carrier for post-irradiation separation. Again the most commonly used method is the oxidation/reduction extraction cycle with carbon tetrachloride. Other methods are combustion followed by HMD separation and adsorption on activated charcoal,[7,11] distillation[18] and ion exchange.[17] Usually a combination of two of these is employed. Finally, the sample is prepared for counting by AgI or PdI_2 precipitation. Sometimes the sample is absorbed on activated charcoal for counting[7,11] or the sample is counted in liquid form.

3.4 Cesium-135

^{135}Cs is a β-emitter with a specific activity of 43 mBq ng^{-1}. Therefore, its determination by radiometric means is very difficult. Very few references on the determination of ^{135}Cs can be found in the literature. Stamm[19] determined ^{135}Cs by NAA in sodium coolant. The determination was based on the ^{135}Cs(n,γ)^{136}Cs reaction with a thermal cross-section of 8.7 barn and a resonance integral of 66 barn. ^{136}Cs has a half-life of 13.7 days and emits several γ-rays, the most important of which are 341, 818 and 1048 keV with respective absolute intensities of 49, 100 and 80%. No details of detection limits were given.

3.5 Neptunium-237

^{237}Np has a specific activity of 26 mBq ng^{-1}. The extremely low concentrations found in nature may be determined by NAA after extensive chemical separations prior to measurement. With a sample of 100 g, the detection limit for NAA is 0.2 mBq kg^{-1}, which means a detection limit of 0.02 mBq per sample.

NAA of ^{237}Np is based on the reaction ^{237}Np(n,γ)^{238}Np. The product nuclide has a half-life of 2.12 days and γ-energies 924 keV (3%), 984 keV (24%), 1026 keV (8%) and 1029 keV (17%). Uranium may interfere

with the analysis by producing ^{237}Np via the reactions ^{238}U(n,2n)^{237}U^{237}Np and ^{235}U(n,γ)^{236}U(n,γ)^{237}U^{237}Np. Although this interference is said to be small,[20] uranium has to be separated from the sample prior to the neutron irradiation. Other reasons for the preirradiation separation are the possibility of using large samples and avoidance of working with high activities. A postirradiation radiochemical separation is also needed in order to remove interfering activities produced from remaining impurities during irradiation. Bromine is especially troublesome. ^{239}Np, which can be produced by neutron irradiation of uranium, is usually used as spike to correct for the yield.

Different separation methods for neptunium have been used.[21] Iron is often extracted into diisopropyl ether[22] or methyl isobutyl ketone (MIBK).[20] Neptunium is purified in a Dowex 1-X 4 anion-exchange column using HCl solution[20,23] or nitric acid.[22,23]

The postirradiation separation procedure may be the same as the preirradiation step, sometimes omitting some steps.[22,23] Extraction with toluene has also been used.[20] The sample may be counted as a solution.

4 STANDARDS

Calibrated standard solutions are available from several commercial suppliers. Both single radionuclides and mixtures of radionuclides may be used, depending on need. For efficiency and energy calibration of γ-spectrometers, it is useful to purchase a mixture in which the energy distribution and intensities are balanced. For energy and point-source efficiency calibration, solid radionuclide sources are convenient to use.

For α- and β-emitting radionuclides, solutions of the individual nuclides are usually used. Those can be purchased from a number of producers of radionuclides both in Europe and the USA.

Reference samples are needed for the quality assurance program. The International Atomic Energy Agency (IAEA) and the Institute for Reference Materials and Measurements (IRMM) in Europe and the National Institute for Standards and Technology (NIST) and the New Brunswick Laboratory in the USA provide standard reference samples of different kinds.

5 QUALITY ASSURANCE

The purpose of any analysis is to obtain results, which can be used for decision making. Therefore, it is the responsibility of the analyst to ensure that the results are reasonably correct and reliable and that the precision and accuracy are known. Modern quality standards also require traceability to such a degree that control of the correctness of all data and calculations by outside persons is possible. The IAEA definition is that it is necessary to provide documentation to show that the analytical results are reliable.[24] All this means that an overall plan for quality assurance has to be made and the procedures and protocols followed in all determinations of radionuclides. The international and national institutions provide all required support for this kind of activity.[24]

The quality control measures are the following: use of adequate procedures and instrumentation; appropriate calibration methods; insertion of control samples in all sample measurements; control of new methods by use of standard reference samples and by analysis of the same sample with different methods; and control of overall quality by participation in national and international intercomparisons (round robins).

The precision can be determined by the analysis of duplicates of the same sample during a period of several days and preferably by different persons. It is more difficult to detect systematic errors. Methods used for this are (1) analysis by several different methods, (2) analysis of reference materials of adequate quality and (3) participation in interlaboratory comparisons.

When a new method is developed, it has to be tested with a suitable standard reference sample. The sample matrix and the radionuclides in the reference sample must be as close as possible to the matrix and radionuclides to be measured in the real samples. If such is not available, the method should be tested by analyzing a sample with several different methods, if possible. Another method is to send a duplicate to another laboratory known for its quality.

In controlling the accuracy of routine work, the following has to be taken into account. The condition of the instruments must be checked at regular intervals. For γ-spectrometers this means that the energy calibration, resolution and efficiency must be checked for each run. This can be done by studying the spectra of the standards and control standards, which must be inserted in each run. The measured values of each parameter are compared with the original set of values. Any significant changes in the values may indicate that something is wrong. The best way to notice trends is to fill the values into a control chart. In a graphic display, trends are easily noticeable. The control level may be three standard deviations.

Control standards are always inserted into each run for general control of accuracy. A good rule is that 10% of all samples are control samples. Certified reference materials are sometimes used for this purpose. However, certified reference materials are expensive and of limited availability. Therefore, many laboratories prepare their own control samples for routine use. Reference materials

are used only occasionally, and when new methods are developed. The results obtained when analyzing the control samples are also plotted on a control chart.

Intercomparisons are organized on an international scale by the IAEA and it is very useful to participate in these. The additional value from these, compared with analyzing reference materials, is that the ''true'' values are not known. Therefore, a more reliable indication of the true accuracy may be obtained. International intercomparisons cannot be organized very frequently, and therefore intercomparisons on a national or other less official levels can also be organized.

6 CONCLUSIONS

^{129}I and ^{237}Np are critical radionuclides in considering the environmental effects of nuclear waste. They are also found in the environment as a consequence of nuclear weapons tests. In addition, ^{237}Np may be used for the production of nuclear weapons and therefore it is a nuclide for which safeguards surveillance is necessary. For all these reasons there is a need to determine these nuclides, sometimes at very low concentrations. NAA has proven to be the most sensitive method to determine ^{129}I and a useful method to determine ^{237}Np in environmental samples. Therefore, NAA can be recommended as an analytical method for these nuclides, when a research reactor is available.

ABBREVIATIONS AND ACRONYMS

HMD	Hydrated Manganese Dioxide
IAEA	International Atomic Energy Agency
IRMM	Institute for Reference Materials and Measurements
MIBK	Methyl Isobutyl Ketone
NAA	Neutron Activation Analysis
NIST	National Institute for Standards and Technology

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Neutron Activation in Environmental Analysis

Field-portable Instrumentation **(Volume 5)**
Radon, Indoor and Remote Measurement of

Nuclear Methods **(Volume 14)**
Charged Particle Activation Analysis • Cyclic Activation Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Photon Activation Analysis • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis

Radiochemical Methods **(Volume 14)**
Radiochemical Methods: Introduction • Actinides and other Alpha-emitters, Determination of • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Mass Spectrometry of Long-lived Radionuclides • Nuclear Detection Methods and Instrumentation • Speciation of Radionuclides in the Environment

REFERENCES

1. G. Hevesy, H. Levi, 'The Action of Neutrons on the Rare Earth Elements', *Det Kgl. Danske Vidensk. Seskab. Mat.-Fysi. Medde.*, **14**(5), 3–34 (1936).
2. R.J. Rosenberg, 'Non-conventional Measurement Techniques for the Determination of Some Long-lived Radionuclides Produced in Nuclear Fuel, a Literature Survey', *J. Radioanal. Nucl. Chem. Articles*, **171**(2), 465–482 (1993).
3. L.C. Bate, 'Determination of Technetium-99 in Mixed Fission Products by Neutron Activation Analysis', in *Radioelement Analysis*, ed. W.S. Lyon, Ann Arbor Science Publishers, Ann Arbor, MI, 175–189, 1980.
4. B.J. Mincher, J.D. Baker, 'The Separation of ^{99}Tc from Mixed Fission Products and Its Determination by Neutron Activation', *J. Radioanal. Nucl. Chem. Articles*, **139**, 273–276 (1990).
5. N. Ikeda, R. Seki, M. Kamemoto, M. Otsuji, 'A Possibility of (n,n′) Activation Analysis for Technetium-99', *Radioisotopes*, **37**, 414–415 (1988).
6. N. Ikeda, R. Seki, M. Kamemoto, M. Otsuji, 'Activation Analysis for Technetium-99 by the Use of a Neutron Excitation Reaction', *J. Radioanal. Nucl. Chem. Articles*, **131**, 65–71 (1989).
7. R.M. Lindstrom, G.J. Lutz, B.R. Norman, 'High-sensitivity Determination of Iodine Isotopic Ratios by Thermal and Fast Neutron Activation', *J. Trace Microprobe Tech.*, **9**, 21–32 (1991).
8. F.P. Brauer, 'Measurement Methods for Low-level ^{129}I Determinations', Pacific Northwest Laboratory Report PNL-SA-16788, Pacific Northwest Laboratory, Richland, WA, 1989.
9. F.P. Brauer, R.S. Strebin, Jr, J.E. Fager, J.H. Kaye, 'Interference Control in Low-level ^{129}I Analysis', *J. Radioanal. Chem.*, **72**, 501–514 (1982).
10. H. Katagiri, O. Narita, A. Yamato, M. Kinoshita, 'Low Level Measurements of ^{129}I in Environmental Samples', *J. Radioanal. Nucl. Chem. Articles*, **138**, 187–192 (1990).

11. G.J. Lutz, H.L. Rook, R.M. Lindstrom, 'Determination of I-129 at Natural Levels by Thermal Neutron Activation Analysis', *J. Trace Microprobe Tech.*, **2**, 33–51 (1984).
12. Y. Muramatsu, S. Uchida, M. Sumiya, Y. Ohmomo, 'Iodine Separation Procedure for the Determination of ^{129}I and ^{127}I in Soil by Neutron Activation Analysis', *J. Radioanal. Nucl. Chem. Lett.*, **94**, 329–338 (1985).
13. N. Ikeda, R. Seki, M. Nagasawa, M. Enoshita, Y. Nakamura, 'Detection of ^{129}I in Some Environmental Samples', *Radioisotopes*, **30**, 162–164 (1981).
14. J.E. Martin, F. Marcinowski, S.K. Cook, 'Optimization of Neutron Activation Analyses of ^{129}I in Low-level Radioactive Waste Samples', *Appl. Radiat. Isot.*, **41**, 727–731 (1990).
15. G.R. Doshi, S.N. Joshi, K.C. Pillai, '^{129}I in Soil and Grass Samples Around a Nuclear Reprocessing Plant', *J. Radioanal. Nucl. Chem. Lett.*, **155**, 115–127 (1991).
16. R. Seki, E. Kimura, T. Takahashi, N. Ikeda, 'Distribution and Behavior of Long-lived Radioiodine in Soil', *J. Radioanal. Nucl. Chem. Articles*, **138**, 17–31 (1990).
17. L.C. Bate, J.R. Stokely, 'Iodine-129 Separation and Determination by Neutron Activation Analysis', *J. Radioanal. Chem.*, **72**, 557–570 (1982).
18. B.T. Wilkins, S.P. Steward, R.O. Major, 'Determination of Long-lived Radionuclides in Environmental Media Around a Nuclear Fuel Reprocessing Plant', *J. Radioanal. Nucl. Chem. Articles*, **115**, 249–262 (1987).
19. H.H. Stamm, 'Determination of ^{135}Cs in Sodium from an In-pile Loop by Activation Analysis', *J. Radioanal. Chem.*, **14**, 367–373 (1973).
20. A.R. Byrne, 'Determination of ^{237}Np in Cumbrian (UK) Sediments by Neutron Activation Analysis', *J. Environ. Radioact.*, **4**, 133–144 (1986).
21. E. Holm, A. Aarkrog, S. Ballestra, 'Determination of ^{237}Np in Large Volume Samples of Sea Water by a Radiochemical Procedure', *J. Radioanal. Chem.*, **115**, 5–11 (1987).
22. C.-K. Kim, R. Seki, N. Ikeda, 'Determination of ^{237}Np in Soil Samples by Neutron Activation Analysis', *Radioisotopes*, **37**, 229–230 (1988).
23. S. May, CH. Engelmann, G. Pinte, 'Neutron Activation Determination of ^{237}Np in Irradiated Experimental Fuels and in Waste Solutions and Distribution Studies in Seawater and Submarine Fauna and Flora of Disposal Areas', *J. Radioanal. Nucl. Chem.*, **113**, 343–350 (1987).
24. 'Measurement of Radionuclides in Food and the Environment, A Guidebook', Technical Reports Series No. 295, International Atomic Energy Agency, Vienna, 1989.

Nuclear Detection Methods and Instrumentation

A. Tsechanski
Ben-Gurion University of the Negev, Israel and Soroka University Medical Center, Israel

1 Introduction **1**
2 Nuclear Detectors **1**
2.1 Gas-filled Detectors 1
2.2 Scintillation Detectors 5
2.3 Semiconductor Detectors 10
3 γ-Ray Spectroscopy **22**
3.1 Instrumentation for Nuclear Spectroscopy 22
3.2 Energy Measurements 23
Abbreviations and Acronyms **24**
Related Articles **25**
References **25**

Nuclear detectors and associated electronics are among the most important components of analytical chemistry involving nuclear methods. Historically, the first and simplest particle detectors were gas-filled counters using an electric field to collect the free charge carriers, electrons and ions, produced in the counter by ionizing radiation. These detectors include ion chambers, proportional counters and Geiger counters. The charge created at the proportional counter is directly proportional to the energy deposited in the detector by a nuclear particle and, therefore, this kind of detector can be used for energy measurements of nuclear radiation (nuclear spectrometry). The main disadvantage of gas-filled detectors is their low efficiency for γ-rays. This is because the gas detection medium is low in density. From this point of view, solid detectors are considerably more advantageous. Of these detectors, the scintillation counter is one of the earliest and most widely used detectors of nuclear radiation. It consists of a detection medium (inorganic or organic scintillator) and a photomultiplier tube (PMT). The incident particle enters the scintillator and interacts with its atoms, raising them to excited states. De-excitation follows very rapidly by emission of light photons. The emitted photons are multiplied in the PMT, resulting in an output pulse of up to several volts. The high efficiency scintillation detectors are widely used in instrumental analytical chemistry including neutron activation analysis (NAA), both prompt neutron activation analysis (PNAA) and delayed neutron activation analysis (DNAA), and in X-ray flourescence analysis (XRFA). Contemporary high-resolution γ-spectrometry is based on solid-state ionization chambers. Only after the successful development of the technology for production of highly purified single crystals of germanium (Ge) and silicon (Si) did it become possible to develop Ge- and Si-based solid-state detectors by forming reverse-biased junctions and cooling the semiconductor to liquid nitrogen temperature (77 K). The most prominent feature of the semiconductor detectors is their excellent energy resolution and availability of large crystals with relative efficiencies up to 120% compared with a standard NaI(Tl) detector. High-resolution γ-ray spectrometry is widely used in evaluation of γ-emitting radioactive nuclides in various samples, in NAA, in a high-sensitivity particle-induced X-ray emission (PIXE) method and others. Nuclear detectors provide a variety of information on detected radiation (energy, type of the radiation, timing data, etc.) in the form of electronic signals. To obtain this information, the signal must be processed by the appropriate nuclear electronic system, which includes power supply, preamplifier, main amplifier and multichannel analyzer (MCA) for digital processing of the analog signals.

1 INTRODUCTION

One of the most important components of analytical chemistry involving nuclear methods is the particle detector used and associated electronics. In many cases, both of these elements prescribe the methodology of the analysis and its applicability, accuracy and limitations. Therefore, a short discussion of nuclear detection methods is essential.

This article presents a concise description of nuclear detectors such as gas-filled detectors including ionization chambers, proportional counters and Geiger counters, scintillation detectors with organic and inorganic scintillators, solid-state detectors including silicon diodes, Si(Li), Ge(Li) and high-purity germanium (HPGe) detectors, along with a short excursion into nuclear electronics and γ-ray spectroscopy. Those interested in comprehensive coverage of the subject are referred to the widely accepted standard textbook in the field by Knoll.[1]

2 NUCLEAR DETECTORS

2.1 Gas-filled Detectors

In general, detection of nuclear radiation is based on registration of the different kinds of radiative interactions with the atoms of the detector material. In most cases,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

such interactions release a large number of secondary electrons, which are collected and shaped into a voltage or current pulse for subsequent analysis by electronic means.

One of the first and best-known types of nuclear detectors, based on a gas as the sensitive medium, is called a gas-filled detector. The simplest detector of this type is the *ionization chamber*, which can simply be regarded as a gas-filled parallel-plate capacitor. When an energetic charged particle passes through a gas inside the ionization chamber, it ionizes a large number of gas molecules or atoms along its track. As a result, electrons and positive ions are produced. The electric field between the plates is strong enough to keep the ions from recombining with the electrons and directs them in opposite directions. The negative free electrons drift toward the positive plate and the positive ions are attracted to the negative plate. When this charge is collected on the plates, an electric pulse is produced. The magnitude of the pulse is easily estimated. The average energy required to produce an electron–ion pair in dry air is about 35 eV ($1\,\text{eV} = 1.602 \times 10^{-19}\,\text{J}$). Thus, a 1-MeV charged particle produces about 3×10^4 electron-ion pairs. For a square $10 \times 10\,\text{cm}$ chamber with a plate separation of 1 cm, the capacitance is $8.85 \times 10^{-12}\,\text{F}$, and therefore the resulting voltage pulse is $1.6 \times 10^{-19}\,\text{C}$ per ion $\times\, 3 \times 10^4$ ion/$(8.85 \times 10^{-12}\,\text{F}) = 0.5\,\text{mV}$. This is a low-amplitude signal, which must be amplified before it can be analyzed by standard electronics.

If both the electrons and ions are completely collected after production of ionization by a charged particle, the final voltage signal is proportional to the total ionization and, therefore, to the charged particle energy. Usually, however, this requires too long a collection time (milliseconds). This is an exceedingly long time for handling reasonably high counting rates, hence the ion chamber is of no use in counting individual particles.

The gas in an ionization chamber should be free from electronegative impurities. Therefore, the use of air in an ion chamber is disadvantageous because air contains oxygen, an extremely electronegative gas. Ion chambers employing air have poor saturation characteristics and require high operating voltages, particularly when chambers are filled to high pressures to obtain high sensitivities. To solve this problem, some filling gas other than air is used. The most common choice is nitrogen or an argon–nitrogen mixture for γ-sensitivity or hydrogen and hydrogen–argon mixture for neutron sensitivity. The gas pressure should be chosen to give an appropriate range of the charged particle, typically between 0.1 and 40 atm (1×10^4–$4 \times 10^6\,\text{N m}^{-2}$).

Ion chambers find wide use as radiation monitors, especially in intense radiation fields. Portable ion chambers are widely used as survey instruments for monitoring β- and X-rays (with a β-shield). Radioactive gases are detected by incorporating them as a component of the filling gas of an ion chamber vented to atmosphere to allow a continuous flow of the gas to be sampled. Tritium monitoring, based on ion chambers, is also possible.[2]

The main disadvantages of ion chambers are the very small output signal and the poor energy resolution compared with proportional counters and semiconductor detectors. To use a gas-filled detector, such as an ionization chamber, for detecting individual nuclear particles, the electric field in the ionization chamber must be increased significantly. In such a case, the electrons drifting to the anode can acquire kinetic energy between two successive collisions sufficient to cause the further (secondary) ionization of gas molecules. The new electrons produced by the secondary ionization are in turn accelerated towards the anode and can produce further ionization. This rapid amplification by production of secondary ionizations is called a Townsend avalanche. Provided the electric field is not too high, the total charge produced is proportional to the initial ionization of the primary particle and therefore to its energy. As a result of development of the gas-discharge avalanche, the charge collected by the electrodes may be up to 10^4–10^5 times that of the initial ionization. Because the pulse height obtained is proportional to the ionization capability of the primary particle, the device is known as a *proportional counter*.

A proportional counter is commonly a cylinder with a very thin wire anode coaxial to a cylindrical cathode. The electric field strength at a radius r in this geometry is given by Equation (1):

$$E(r) = \frac{V}{r \ln(b/a)} \tag{1}$$

where a and b are the radii of the cathode and anode, respectively. The high-field region, therefore, is concentrated very close to the anode wire. The gas multiplication takes place in this intense-field narrow region around the anode, resulting in almost total independence of the gas multiplication factor from the initial position of ionization. In general, the voltage induced on the collecting electrode by a charged particle is proportional to the fraction of voltage through which the charged particle falls before collection. Because about half of the total number of ion pairs is formed within one mean free path of the surface of the anode, the positive ions fall through a much greater potential before collection than the electrons. The signal induced on the anode therefore results mainly from the positive ions rather than the electrons. Because most of the positive ions are formed in the region of high electrostatic field near the central wire, the rise time of the signal is very rapid (of the order of microseconds) and, therefore, the

proportional counter operates at high counting rates of the order of $10^6\ s^{-1}$.

Because of the gas amplification, the signal in the proportional counter is much larger than in an ion chamber. Therefore, the proportional counter is particularly useful for detecting low energy radiation and particularly photons (down to about 0.1 keV). Very thin entry windows (down to 1 μm thick) of plastic, mica or beryllium are employed in these counters. The gas filling of a proportional counter is selected from the noble gases neon, argon, krypton and xenon. In addition, up to 10% of a quench gas (methane is most commonly used) is added to stabilize the operation at high values of gas multiplication. The most common, general purpose, proportional counter is filled with a mixture of 90% argon and 10% methane, known as P-10 gas. Proportional counters for detecting γ-photons by absorption within the gas are filled with the inert gases of higher atomic number (Z) (krypton or xenon).

Proportional counters are designed in sealed tube configuration or as windowless flow counters. In the latter case, the radioactive sample to be counted is introduced directly into the sensitive volume of the detector. The advantage of such counting is obvious – there is no need for an entrance window which otherwise would considerably attenuate or even prevent soft radiations, such as α-particles or low-energy photons, from entering.

With the development of highly stable high-gain amplifiers, proportional counters are the best choice for routine low-level activity sample measurements. Because of their high collection efficiency, gas proportional detectors are unsurpassed for extremely low-level radiation measurements. Because sensitivity is one of the most significant parameters for low-level monitoring, the proportional counter surface area should be considerable. The α/β-counting systems designed specifically for extremely low activity samples use a large diameter (up to 5.7 cm) gas flow proportional detector with an ultrathin window for maximum efficiency or a sealed proportional detector for less stringent requirements.[3]

Proportional counters are used in XRFA, especially if only one element is to be determined and provided that the matrix under consideration does not contain interfering elements. In such a case, the resolution considerations are of minor importance and the peak-to-background ratio becomes the most important factor. Argon is the most commonly used filling gas for the detection of low-energy X-ray photons, and krypton or xenon for higher energies.

If the electric field in a proportional counter is increased to even larger values, a new factor plays an increasingly important role. This factor is a considerable excitation of the gas atoms and molecules. As a result of very fast de-excitation of these molecules, ultraviolet (UV) photons are emitted that, in turn, produce photoelectrons mostly from the counter wall. These photons migrate far from the place of the original avalanche, causing secondary avalanches by photoelectrons, and so the gas in the entire tube is involved in the discharge. As a result, the amplification is very large, up to 10^{12}. Because the entire active volume (gas) of the tube participates in every registration effect, the amount of charge collected on the electrodes is independent of the initial ionization and, virtually, a single primary ion pair can fire the discharge. As a result, there is no information at all on the type and energy of the original radiation. This region of operation is called the *Geiger–Müller region*, and gas counters employing this principle are usually known as Geiger–Müller counters (or simply Geiger counters).

Just after beginning the gas discharge in the Geiger counter, the electrons, being considerably more mobile than positive ions, are effectively collected by the anode, so that the output signal consists mainly of the electrons from the many avalanches. The electron collection time is of the order of a few microseconds. During this short interval, the positive ions, having a lower mobility, barely move away from the avalanche region around the anode. They are left in the counter as a positive charge sheath, which continuously builds up and reduces the electric field intensity in the counter to a level which eventually terminates the multiplication process. The discharge would then be ended after the positive ion sheath arrives at the cathode. However, when the positive ions arrive at the cathode, they cause photon emission, which in turn causes the gas discharge in the counter to start again. To prevent this from occurring, some method of quenching of the discharge must be used.

In general, quenching is implemented by external electronic means or by adding a second type of gas (quenching gas) to the tube. In electronic quenching, some reduction in the counter operating voltage is brought about immediately after the counter pulse appears. The reduction must be enough to bring the counter voltage below the starting point of the second discharge. After allowing time for all the positive ions to reach the cathode and become neutralized (up to 1 ms), the voltage is restored to its original working value.

In the so-called self-quenching counters, the quenching action is accomplished by the addition, to the main (counting) gas, a small amount of a second gas with complex molecules, with a lower ionization potential than the simple counting gas molecules. In addition, when ionized, the quenching gas should have a higher probability of losing its energy by dissociation rather than by emission of UV photons. A typical filling for a Geiger counter is 90% argon and 10% ethanol. As the positive charge sheath, consisting mostly of argon ions, begins to move towards the cathode, collisions occur

with the molecules of the quenching gas, transferring an electron from the ethanol molecule to an argon ion. Thus, only ions of the quenching gas reach the cathode and are neutralized there. However, the energy of the ions that formerly went into emission of a UV photon is now absorbed in the dissociation of the ethanol molecule (a process not possible for a simple argon atom). The quenching gas is thus gradually used up, and the Geiger tube life expectation will be about 10^8–10^9 counts. Another widespread category of Geiger tubes is the so-called halogen counters. Their counting gas is composed of noble gases (Ne, Ar or Kr) with a small quantity (0.1–1%) of halogen vapor (Cl_2, Br_2 or I_2) as a quenching agent. As in the previous case, the diatomic molecules of the halogen gas are dissociated in the quenching process, but the subsequent spontaneous regeneration of the dissociated halogen molecules eliminates, in principle, the need to replace the Geiger tube. However, contamination of the quench gas by reaction products and changes on the anode surface ultimately limit the lifetime of the halogen Geiger counters to about 10^{10} counts, compared with 10^8 typical life expectation counts of counters quenched by organic vapor.

In the period immediately following the Geiger discharge, the positive space charge around the anode effectively reduces the electric field below the ignition point. If another charged particle occurs inside the counter during this period of time, it is not observed, because gas multiplication is prevented. This *dead time* when the tube is unable to respond to radiation is of the order of 50–200 μs. After the dead time there is a progressive recovery when the Geiger counter can respond again but the resulting pulse does not attain its normal magnitude. Registration of such a partial pulse depends on the sensitivity adjustment of the counting electronics. The usual procedure is to arrange the electronic circuit with a defined dead time so that the loss of counts from finite dead time is accurately estimated.

The counting efficiency of the Geiger counter for any charged particle that enters its active volume is close to 100%, because even one charged particle inside the tube is enough to trigger full-scale Geiger discharge. The efficiency for counting γ-rays depends on the probability of interaction of the incident γ-ray with the solid wall of the counter and on the probability that the secondary electron released in this interaction reaches the active volume before the end of its track. Because the probability of γ-interaction within the sensitive layer is small even for high-Z cathode materials, γ-ray counting efficiency is generally low and at 1 MeV is about 1%. Most Geiger counters have a cylindrical stainless-steel cathode and a mica window thickness down to 1.5 mg cm^{-2}; gas pressure of the order of 100 Torr (1.33×10^4 N m^{-2}) is usual.

Geiger counters are used where simple, portable, inexpensive and rugged radiation detectors are needed. A large variety of Geiger counters are available for different applications, including the detection of α-, β- and γ-radiation in separate or mixed radiation fields. Many types of radiation survey meters used in X-ray and γ-monitoring consist of an energy-compensated Geiger tube, a portable high-voltage power supply and counting electronics. They are ideally suited for detecting different radiation levels from radioactive sources and radiation contamination and in X-ray machines in both industrial and hospital environments.

The various operating regions of gas-filled detectors are presented in Figure 1. The amplitude of the observed pulse is given as a function of the applied voltage within the detector for two particles depositing two different amounts of energy within the gas. At very low voltages, the field is insufficient to prevent the recombination of the primary electrons and ions, and the collected charge is less than the original one. As the voltage is increased, recombination is suppressed and the region of ion saturation is achieved. This is the *ionization chamber region*, where the pulse output is proportional to the primary ionization and thus to the energy of the radiation, but is independent of the applied voltage. As the voltage is increased still further, gas multiplication begins in the detector and the observed pulse amplitude increases substantially with the applied voltage. This is the *proportional region*, in which the

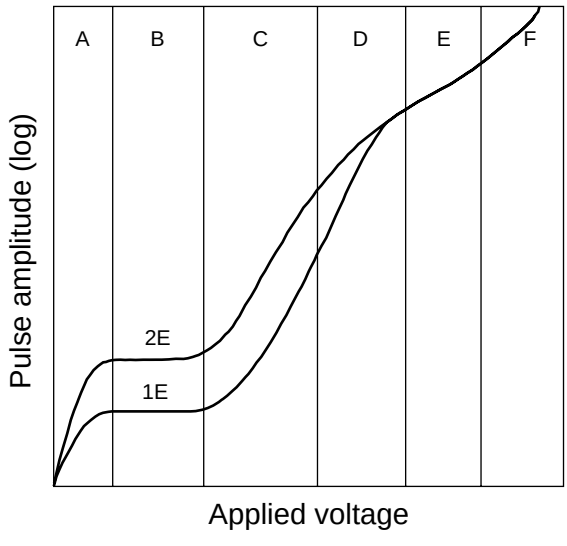

Figure 1 Pulse amplitude as a function of applied voltage for the different operating regions of gas-filled detectors. Plots are given for charged particles of two different energies (E and $2E$): A, pulse amplitude reduced by recombination; B, ionization chamber region; C, proportional counter region; D, limited proportionality region; E, Geiger counter region; F, continuous discharge region.

pulse amplitude rises proportionally to the voltage, but the output pulse remains proportional to the energy of the radiation, thus simplifying identification of the primary radiation and its energy. Further increase of the applied voltage results in increasing concentration of positive ions (space charge) around the anode, which significantly diminishes the electric field within the detector. As a result, some nonlinearities in gas multiplication begin to be observed. This is the region of *limited proportionality*, in which the pulse amplitude still increases with energy of the primary radiation, but not linearly. Finally, as the applied voltage becomes sufficiently high, gas multiplication is determined by the space charge created by the positive ions. Under these conditions, the gas discharge is self-limiting and all output pulses are of the same amplitude, regardless of the kind of primary radiation and its energy. This is the *Geiger–Müller region* of operation characteristic of Geiger counters. For more detailed information on the subject of gas-filled detectors, see general reference books.[4–6]

2.2 Scintillation Detectors

2.2.1 Basic Principles

The main disadvantage of gas-filled detectors is their low efficiency for several types of nuclear radiation and particularly for γ-rays. This arises from the fact that the gas detection medium is low in density. From this point of view, solid detectors are more advantageous, because their higher densities result in acceptable absorption probabilities for detectors of reasonable size. Of the detectors with a solid detection medium, the scintillation counter is one of the earliest and most widely used detectors of nuclear radiation. A schematic representation of the basic elements and processes in a scintillation detector is given in Figure 2. The detector consists of a detection medium (scintillator) and a PMT. The incident particle enters the scintillator and interacts repeatedly with the atoms of the scintillator, raising them to excited states. De-excitation follows very rapidly by emission of photons of visible or near-visible light. The emitted photons strike a photocathode, releasing a maximum of one photoelectron per photon. Finally, these photoelectrons are multiplied in the PMT, resulting in an output signal pulse of up to several volts.

In general, for use as a scintillator for nuclear radiation, the material should possess the following properties:

1. should convert the energy of charged particles into emitted light linearly and with high scintillation efficiency;
2. scintillator should be transparent to its own emitted light;

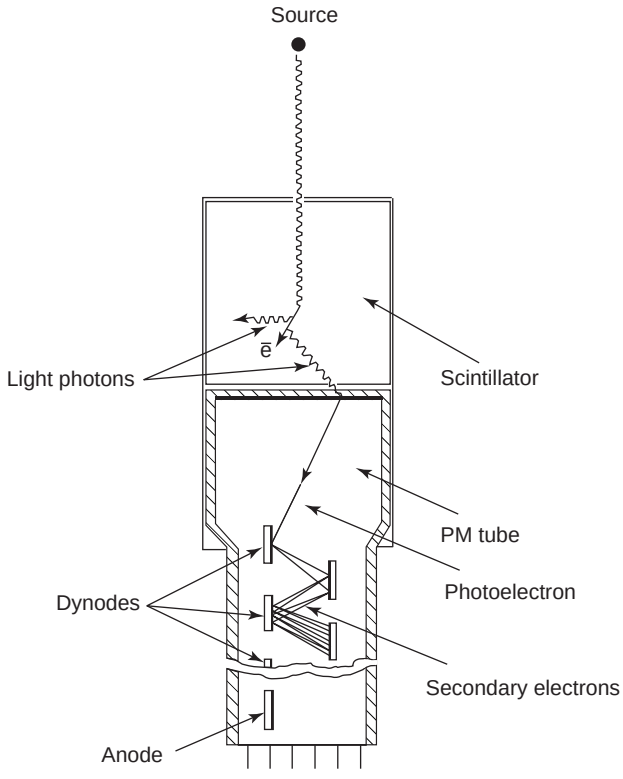

Figure 2 Basic elements and processes in a scintillation detector.

3. decay time of the luminescence should be short;
4. spectral distribution of the emitted light should match as closely as possible the response of the photocathode of the PMT used.

Many materials are known to luminesce under the influence of nuclear radiation. They are in gaseous, liquid or solid form. Scintillating materials are usually divided into organic and inorganic materials. The most widely used inorganic scintillators have generally higher light output and better linearity but slower time response than organic scintillators. The high photoelectric cross-section and high density of inorganic scintillators make them especially valuable for γ-detection whereas organic scintillators are more preferable for charged particle and neutron spectroscopy. Organic scintillators in liquid or plastic form are easily fabricated in large and irregular geometries. The size of inorganic scintillators is strictly limited by the maximum size to which a single crystal may be grown.

A general review of the basic theory, practice and application of scintillation detectors can be found in the exhaustive monograph by Birks.[7] Because the mechanisms of luminescence in organic and inorganic scintillators are different, they are considered separately.

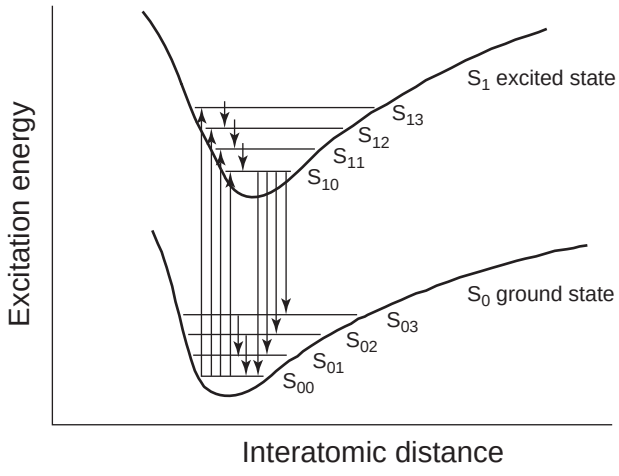

Figure 3 Electronic states structure in an organic scintillator.

2.2.2 Organic Scintillators

The molecules in organic scintillators (whether in the gaseous, liquid or plastic state) are relatively loosely bound and, as a result, retain their individual electronic structure and luminescent features. The properties of organic scintillators can be discussed in term of the discrete excited states of the molecules presented in simplified form in Figure 3. The electronic energy states are represented as a potential minimum. Inside the electron potential minimum is a series of levels with much finer spacing corresponding to various vibrational states of the molecule. Typical spacing of vibrational levels is about 0.15 eV, and the electronic excitation energies are of the order of 3–4 eV. A charged particle passing inside the scintillator interacts with many molecules, losing a few electronvolts at each interaction. Generally, a scintillator molecule absorbs energy either by excitation of the electrons to higher excited states, or by atoms in the molecule vibrating against one another. In Figure 3, the absorption of energy by the molecule is represented by upwards arrows. As a result of the interaction, many vibrational states also become excited. However, because any excited vibrational state is not in thermal equilibrium with its neighbors, it quickly decays (during $\sim$1 ps) to the lowest vibrational state of the electronic excited state S_{10}. Therefore, as a first stage of the scintillation process, a population of excited molecules in the S_{10} state is produced. De-excitation of these molecules (during a few nanoseconds) to one of the vibrational states of the ground-level electronic state produces the principal scintillation light (the downward arrows in Figure 3). The vibrational states of the electronic ground state again decay quickly to the vibrational ground state S_{00}.

The length of the upward arrows in Figure 3, corresponding to photon absorption energies, is greater than that of the downward arrows, corresponding to photon emission energies in a scintillator. This explains why organic scintillators are transparent to their own emission spectrum.

Of the pure organic materials, only anthracene and stilbene (aromatic hydrocarbons with a linked benzene ring structure) have achieved widespread popularity. Anthracene features the greatest light output per unit energy (scintillation efficiency). Stilbene has a lower scintillation efficiency but it exercises pulse-shape discrimination (PSD) properties to distinguish between scintillation induced by heavy charged particles (e.g. protons) and by electrons. Both materials are relatively fragile and may be fabricated as single crystals with a side of no more than several centimeters.

If a small concentration of impurity is added to a bulk scintillator (solvent), in some cases, there is high probability that the excitation energy absorbed mainly by the solvent molecules is transferred ultimately to the impurity molecules. If these molecules de-excite by radiationless transition (mainly by collisions), the process reduces the scintillation efficiency, and the effect is known as ''quenching''. A prominent example is dissolved oxygen which substantially degrades the light output of many scintillators. On the other hand, if the added impurity is an efficient scintillator it may effectively increase the overall fluorescent efficiency. These binary organic scintillators, widely used in liquid and plastic form, represent a separate type of organic scintillator called liquid and plastic solutions. The primary excitation takes place mainly in the host material or solvent which may be toluene, xylene or benzene for liquid scintillators and poly(vinyltoluene) or polystyrene for plastics. The excitation is effectively transferred to solute molecules such as *p*-terphenyl, PPO (2,5-diphenyloxazole) and PBD [2-phenyl-5-(4-biphenylyl)-1,3,4-oxadiazole]. Sometimes a third additive is present in a concentration of a few percent or less and serves as a waveshifter. Its function is to improve matching between the fluorescence spectra and the PMT spectral response and to increase the transparency of the scintillator to its own light. The most commonly used waveshifter is POPOP {1,4-di[2-(5-phenyloxazolyl)]benzene}.

For most organic scintillators, the variation of the light output with electron energy is slightly nonlinear for low energies up to about 0.1 MeV and is linear at higher energies. Heavy charged particles, such as protons or α-particles, are considerably less efficient in producing light in organic scintillators than electrons. This subject is extensively treated in a general reference.[7]

As to the time response of the organic scintillators, this category of scintillators is the fastest. In a simple approach, the time profile of the light pulse is described by a single exponent $I = I_0 e^{-t/\tau}$, where τ is the time

constant describing the decay of the optical levels. In the fastest organic scintillators, $\tau = 1.5–2.5$ ns. This exponent represents the prompt fluorescence responsible for most of the observed scintillation light. In some scintillators, however, the slower components are also observed with a characteristic decay time of several hundred nanoseconds. It should be pointed out that the relative amounts of light in the fast and slow components often depend on the nature of the exciting particle. This dependence is used to discriminate between different kind of particles. This is so-called PSD, used in several applications, but the most common use is for discrimination between γ-rays (i.e. via Compton electrons) and fast neutrons (i.e. via recoil protons).

2.2.3 Inorganic Scintillators

In general, inorganic scintillators are single-crystal insulated materials. A monocrystalline structure is necessary because multiple scattering and absorption of the emitted light at the many faces of a polycrystal would make such a scintillator useless. The energy band structure of a crystalline scintillator includes the lower band, called the valence band, which is generally full and represents the energy states of the electrons strictly bound at crystal sites. The upper band, called the conduction band, contains empty energy states. The forbidden energy band, in which electrons can never be found, is between the valence and conduction bands. In inorganic scintillators, the width of the forbidden band is about 4 eV. Incoming radiation can excite an electron and elevate it from the full valence band across the forbidden band into the conduction band. In the pure crystal, the excited electron loses its energy by emitting a photon and returns into the valence band. Although pure inorganic crystals are used as scintillators, they are inefficient because of the high self-absorption of the emitted light by the crystal itself. To reduce self-absorption of the light and to increase the probability of visible photon emission, a small amount of a suitable impurity, called an activator, is added to the crystal. Such activators create additional energy states within the forbidden band and photon emission takes place between the activator states. Because the photon energy is less than the forbidden bandwidth of the host material, self-absorption cannot occur and the scintillation process results in emission of photons of lower energy lying in the visible range. Another advantage of the change in photon energy from the UV (pure crystal) to the visible (activated crystal) results in better overlap between the emission spectra of inorganic scintillators and the photosensitivity of most photocathodes.

The most notable property of inorganic scintillators is their high efficiency for detecting γ-rays. This follows from the higher density and higher atomic number of inorganic scintillators, resulting in much more efficient detection by the photoelectric effect than in any organic material. Of the inorganic scintillators, the most common material is sodium iodide to which 0.1% mole fraction of thallium is added as an activator. Such NaI(Tl) crystals have a high light yield. They are fabricated in irregular shapes and in large sizes, up to 30 cm. Sodium iodide is deliquescent and therefore must be encapsulated in an airtight container for normal use. The dominant decay time of the scintillation of NaI(Tl) is long ($\sim$230 ns) compared with that of organic scintillators and it is therefore not chosen for applications requiring high counting rates and fast timing. The response of the NaI(Tl) crystal to γ-rays (via the photoeffect) is virtually linear over the practical energy range. This feature, in addition to its excellent light yield, makes for NaI(Tl) the generally accepted standard scintillator for γ-ray spectrometry.

Of other inorganic scintillators, CsI(Tl) and CsI(Na) have become popular for use as γ-detectors. Both crystals have a higher effective atomic number than NaI(Tl) and as a result their γ-ray absorption coefficient is higher. Cesium iodide crystals are less brittle than NaI(Tl) and are not as hygroscopic. CsI crystals can be cut into thin sheets of thickness only sufficient enough to stop heavy charged particles. Because of its low sensitivity to γ-rays and electrons, this configuration is ideally suited for detecting heavy charged particles.

Silver-activated zinc sulfide [ZnS(Ag)] is another inorganic scintillator used for the detection of α-particles and heavy ions. It has a very high luminescent efficiency but is only available as a multicrystalline powder with a limited opacity to its own light. For detecting heavy charged particles, the ZnS(Ag) powder is dispersed as a thin layer (scintillation screen). It is effectively used for fast neutron detection by the proton recoil method. In such a case, detector is a dispersion of zinc sulfide in a hydrogenous medium.

Of the recently developed inorganic scintillation materials, bismuth germanate (Bi_4GeO_{12}) (BGO) cannot be omitted. It was first introduced commercially in 1979. Its prominent feature is a high stopping power (Z of Bi = 83, density = 7.13 g cm^{-3}), resulting in the highest probability of any scintillator for the photoelectric absorption of γ-rays. The light yield is relatively low, only $\sim$21% of that of NaI(Tl). This factor limits low energy applications of BGO and impairs its energy resolution. Therefore, BGO is chosen when the need for high γ-ray counting efficiency outweighs consideration of the energy resolution. The principal decay time of BGO is 300 ns, close to that of NaI(Tl); however, the slower decaying components responsible for the afterglow are more than 100-fold weaker than those in sodium iodide. This feature, together

Table 1 Properties of some common scintillators

Name	Type	Density ($g\,cm^{-3}$)	Refractive index	Light output (% anthracene for γ-rays)	Decay constant main component (ns)	Wavelength of max. emission (nm)	Principal application
Anthracene	Organic crystal	1.25	1.62	100	30	447	α-,β-,γ-, fast neutrons
Stilbene	Organic crystal	1.16	1.626	50	4.5	410	γ-, fast neutrons, PSD
PILOT U	Organic plastic	1.032	1.58	67	1.36	391	Ultrafast timing
NE 213	Organic liquid	0.874	1.508	78	3.7	425	γ-, fast neutrons, PSD
NaI(Tl)	Inorganic crystal	3.67	1.85	230	230	415	γ-
CsI(Tl)	Inorganic crystal	4.51	1.80	95	1000	540	γ-, α-
BGO	Inorganic crystal	7.13	2.15	48	300	505	γ-

with the high γ cross-section of BGO and its nonhygroscopicity, have led to widespread application of BGO as a scintillation scanner in computerized tomography and subsequently in positron emission tomography. In both cases, scintillators must accurately follow rapid changes in photons intensity.

The properties of some most commonly used scintillators are presented in Table 1.[1,8]

2.2.4 Photomultiplier Tubes

A schematic diagram of a PMT is presented in Figure 2. It consists of a photocathode, a series of multiplying and focusing electrodes called *dynodes* and a collector electrode, or anode. A small number of photoelectrons released from the photocathode by the scintillation light are accelerated (and focused) through 100–200 V on to the first dynode. The dynodes are made of alloys having a high probability of secondary electron emission. As a result, an increased number of secondary electrons (a typical value is 4–6 for every primary incident electron) is produced by the first dynode and these, in turn, are accelerated on to the second dynode, where further multiplication takes place. The process of multiplication is repeated on each dynode. If a PMT has N multiplication stages (dynodes) and the multiplication factor for a single stage is δ, then the overall gain K for the PMT is given by Equation (2):

$$K = \alpha\delta^N \tag{2}$$

where α is the fraction of all photoelectrons collected by the multiplier structure (typically $\alpha \approx 1$). A common number of stages is 9–14. With typical value of $\delta = 5$, this results in an overall gain of 10^6–10^9 at the anode. This high gain is sufficient to produce a pulse of several volts on the anode without any external amplification. The multiplication factor δ varies (and generally increases) with the interdynode voltage. Therefore, the overall gain K is a very sensitive function of the high voltage applied to a PMT.

One of the important elements of a PMT is the photocathode, which is responsible for converting incident light photons into photoelectrons. The photocathode is deposited as an opaque or semitransparent layer on the inside of the glass envelope. Historically, the most popular photocathode material was a compound of antimony and cesium which gave a peak sensitivity to light at ~350 nm and a peak quantum efficiency (i.e. number of photoelectrons emitted per incident light photon) of up to 23%. In more recent years, a multialkali photocathode (Na_2KSb activated with Cs) was developed. This photocathode is sensitive from the UV to the red and near-infrared regions of the spectrum and has a high quantum efficiency of up to 30% in the blue region. The multialkali photocathodes have a limited upper temperature range of about 50 °C. Of bialkali photocathodes, K_2CsSb has an even higher peak quantum efficiency. Another prominent feature of the bialkali photocathodes is their significantly lower thermoionic emission than the multialkali or Cs–Sb photocathodes. When a PMT is operated even in complete darkness, electrons are still emitted from the photocathode primarily because of the thermoionic effect. The resulting dark current is amplified by the PMT and sets a limit to the smallest signal observable by the PMT. For this reason, the bialkali photocathodes are preferable to the others.

Materials such as BeO, MgO and Cs_3Sb have been used ($\delta \approx 5$ at typical accelerating voltages between dynodes of 200–300 V) for the electron-emitting surface of the dynodes. More recently, gallium phosphide (GaP) was introduced as a material for the dynode emitting surface, particularly for the first dynode. The multiplication factor δ of GaP material is considerably higher, close to 10 at 200 V, and goes up to 50 at a 1000 V interstage voltage. This feature is valuable in PMT for ultrafast timing applications.

The coupling of a scintillator to a PMT is achieved in many ways. Some detectors integrally connect the PMT to the scintillation crystal enclosed in a light-tight assembly. Others incorporate the scintillator in one container with the PMT mounted so that it can be removed by the user

and reinstalled. In any case, transparent optical grease should be used between the scintillator and the PMT to provide a smooth change in refractive index and minimize internal reflection. Scintillators are mounted directly on the PMT face plate or optically coupled by a light pipe typically made of quartz or glass. Light pipes enhance uniform light collection or when the PMT must be located far from the scintillator. Both the scintillator and the light pipe must be painted with reflective material to enhance light collection. Magnesium oxide is the most commonly used reflective material.

2.2.5 *Application of Scintillation Detectors in Analytical Chemistry*

Liquid scintillators are often used as large-volume detectors with dimensions of up to several meters. In such applications the liquid scintillator may be the only choice from a price standpoint. However, probably the most common application of liquid scintillators is in assaying radioactive material which can be dissolved or dispersed in the scintillator solution. In this case the scintillator acts as a very efficient detector for the emitted charged particles. This so-called liquid scintillation counting method is especially valuable for counting of low energy β-rays of tritium ($E_\beta^{\max} = 18\,\text{keV}$) and ^{14}C ($E_\beta^{\max} = 160\,\text{keV}$). Liquid scintillation counting can also be used for α-active samples with an efficiency close to 100%. Because the α-particle energies from radionuclides are high enough (between about 2 and 11 MeV), the light output is much greater than in the case of low-energy β-particle counting. Therefore, pulse-height discrimination is easily applied to reject β- or low-energy γ-induced effects. The high-energy γ-background is eliminated by PSD. Because of very high counting efficiency, background discrimination possibilities and uniform counting geometry, liquid scintillators are extensively applied for counting low-level activity samples.

Because plastic scintillators are relatively inexpensive, they can be used in large-volume solid scintillator detectors and in a large variety of forms (flat sheets, rods, cylinders, etc.). The α- or β-particle detectors are manufactured by optically cementing a thin plastic scintillator sheet to a thermoplastic light-guide disk. Because of their low sensitivity to γ-radiation (low Z), these detectors are ideally suited to low-level monitoring of α- and β-particles, heavy ions, fission fragments, etc. Another α-particle detector, which features low sensitivity to γ-radiation and large light output, is based on ZnS(Ag) phosphor. The detector consists of a thin layer of ZnS(Ag) bonded to a Perspex thermoplastic disk 6.4 mm thick. The phosphor layer is covered by double aluminated plastic film, permitting use of the detector in full illumination.[9]

In general, organic scintillators are used for the detection and spectrometry of charged particles (electrons, protons, α-particles, etc.). As such they are widely used for the detection and spectrometry of γ-rays through the Compton electrons and of fast neutrons through recoiled protons. To expand the detection possibilities of both liquid and plastic organic scintillators, it is possible to load them with some additional materials. Thus, for instance, boron, lithium or gadolinium is added to give a thermal neutron detector. In the case of boron and lithium, the (n,α) reaction results in direct detection of α-particles in the organic scintillator. In the case of gadolinium, effective capture for thermal neutrons results in β- and γ-activity which is detected in the organic scintillator.

For inorganic scintillators, the scintillation detector most frequently used is the sodium iodide crystal activated with thallium, referred to as NaI(Tl). In many applications involving low-level counting systems, where α-, β-, and γ-rays must be acquired simultaneously from the same sample, an NaI(Tl) detector should be incorporated in the shield assembly to take advantage of coincidence methods to reject background events as much as possible.

Recently, the monitoring of radon and radon progeny levels in households has become of increasing interest. Radon-222 is an inert gas coming from radium-226, has a half-life of 3.8 days and is an α-emitter. Radon is the heaviest known gas and it is ubiquitous in soil and air. The main hazard of radon is from inhalation of the element and its solid daughters, which are collected on dust in the air. Radon accounts for roughly 50% of the total background exposure. The US Environmental Protection Agency (EPA) recommends remediation action if the concentration of radon is consistently above $150\,\text{Bq}\,\text{m}^{-3}$.[10]

One of the techniques widely used for radon monitoring is based on the routine screening of radon levels in air by counting γ-rays coming from the radon progeny nuclides captured in activated charcoal canisters.[11] A $7.6 \times 7.6\,\text{cm}$ ($3 \times 3\,\text{in}$) NaI(Tl) detector in a benchtop, low-background lead shield is used to accommodate charcoal canisters for counting the radon daughters, such as ^{214}Pb ($E_\gamma = 242, 295$, and $352\,\text{keV}$) and ^{214}Bi ($E_\gamma = 609\,\text{keV}$). Depending on its moisture content, charcoal effectively absorbs radon at a predictable rate. By weighing the canister before and after exposure to radon, the amount of adsorbed water is determined. Knowledge of this and the exposure time allows a calibration to be made so that the level of radon daughters captured on the cartridge is converted into the average radon concentration in the air during exposure. The charcoal canister method is within the $\pm 25\%$ error band considered acceptable by the EPA.[12]

Medium-size NaI(Tl) detectors [$5 \times 5\,\text{cm}$ or $7.6 \times 7.6\,\text{cm}$ (2×2 or $3 \times 3\,\text{in}$)] and multiple detector

configurations (up to 64 sodium iodide detectors) are widely used for foodstuff monitoring, especially after the Chernobyl nuclear accident. Foodstuff monitoring is based on routine screening of solid and liquid foodstuffs. It is intended primarily for the quantitative assay of radioactive contamination of foodstuffs by γ-emitting radioisotopes, such as ^{137}Cs ($E_\gamma = 662$ keV) and ^{134}Cs ($E_\gamma =$ 605, 796 keV). To count a foodstuff or any other sample in the optimum and reproducible geometry with a minimum background, the sample should be placed in a Marinelli (re-entrant) beaker surrounding the sodium iodide detector, which is, in turn, placed inside a low-background lead shield. The sodium iodide-based foodstuffs monitor may also be used in other low-level counting applications, such as radon charcoal filter counting, test swipes counting, water monitoring and material monitoring.

Large-size [up to $10 \times 10 \times 40$ cm ($4 \times 4 \times 16$ in)] NaI crystals with integral PMTs are used in whole-body counters for rapid ($\sim$1 min) subject screening. They provide excellent sensitivity for quantitative analysis of fission and activation products at nuclear facilities, such as ^{60}Co, ^{137}Cs, ^{134}Cs, ^{131}I, ^{133}I, ^{54}Mn and ^{95}Zr. Another application field for the large NaI(Tl) or CsI(Na) scintillation crystals is the aerial survey. Radiation detectors are carried on planes or helicopters which cover large areas in relatively short time intervals. The detectors are used to detect γ-rays of ^{40}K ($E_\gamma = 1461$ keV) or daughters from the uranium or thorium radioactive series.

NaI(Tl) detectors are often used in the form of a well detector, that is, the scintillation crystal is shaped to form a cavity. Surrounded by the detector element, such a well is highly efficient in measuring small samples, because the radioactive samples are placed directly in the crystal well.

The high efficiency NaI(Tl) and BGO inorganic scintillators are widely used in NAA, both prompt (PNAA),where measurements of emitted photons occur during the neutron irradiation, and delayed DNAA, where measurements occur following radioactive decay. In this application, both scintillators have the advantage of high Z and high density, resulting in very high interaction cross-sections for emitted photons, especially of higher energies. For instance, the detection efficiency of an NaI(Tl) detector outperforms high-resolution germanium detectors by a factor of at least 10 for $E_\gamma > 8$ MeV. Therefore, a rapid prompt γ-activation analysis scan is better conducted by use of a scintillation detector. Although NaI(Tl) detectors are available in large size (as large as 24 in diameter in γ-ray scanning), the detector itself may become radioactive in neutron irradiation (15 h ^{24}Na and 25 min ^{128}I) through (n,γ) reactions. The high density BGO scintillator, however, acquires no such internal γ contamination, and its detection efficiency further exceeds that of the NaI(Tl) scintillator by another factor of 10 at $E_\gamma > 8$ MeV.[13]

An NaI detector, which has much higher efficiency than the gas-filled proportional counter, can be effectively used as an X-ray detector in XRFA, especially in the cases when one element has to be determined and provided that the matrix itself does not contain interfering elements.

2.3 Semiconductor Detectors

2.3.1 Introduction

Historically, semiconductor detectors were developed as solid-state ionization chambers. Highly insulating diamond crystals were first used as a sensitive material (instead of gas) to obtain a high-field, low-current, solid-state device for detecting nuclear radiation. However, such crystals were quickly abandoned because of inadequate charge collection characteristics of the diamond crystals resulting from the charge trapping at defect centers. Only after the successful development of the technology for the production of highly purified single crystals of silicon (Si) and germanium (Ge) for transistors did it become possible to develop Si- and Ge-based solid-state detectors by forming reverse-biased junctions on these materials. The most prominent feature of the semiconductor detectors is their excellent energy resolution. In addition, they have fast timing capabilities, and their effective thickness and thus their efficiency are varied in accordance with the requirements of the experiment. Detailed information on the subject of semiconductor detectors is available in the current literature.[1,14–17]

2.3.2 Basic Principles

At first sight, it seems that many solids could be used as the sensitive material in solid-state detectors, but this is not so, for a number of reasons. First we have to take into account the noise considerations. Even in the absence of radiation, the conductivity of the active volume is not negligible, so that inevitably a nonzero leakage current across the surface of the solid-state detector exists and obscures the small useful signal. From this point of view, insulators (with resistivity of up to $10^{16}\,\Omega$ cm) are the best choice. However, insulators suffer from such grave shortcomings as charge carrier trapping and formation of a space charge. Under these circumstances, the electrons and holes formed in an insulator would quickly be trapped or recombine and so would induce only a negligible voltage pulse on the detector electrodes. Therefore, semiconductors are the more appropriate choice for solid-state detectors, despite the fact that their resistivity under normal conditions is considerably less than that required by signal-to-noise ratio considerations ($\sim 10^{10}\,\Omega$ cm). Of all semiconductor materials, germanium and silicon

are used almost exclusively in the modern solid-state detectors, because adequate purification technology is available only for these materials. Using this advanced purification technology, silicon and germanium with resistivity of no more than about 50 000 and 50 Ω cm, respectively, are obtained. Both values are too small to allow the straightforward application of a homogeneous block of Si or Ge as a semiconductor detector. Therefore, some special measures are taken to increase sharply the resistivity of the semiconductors. These measures include cooling the semiconductor to liquid nitrogen temperature (77 K), use of a reverse biased p–n junction, compensation by the lithium drift method and other means that will be elaborated later.

2.3.3 Physical Properties of Pure Germanium and Silicon

Both Ge and Si have a crystal structure consisting of a regular repetition of a tetrahedron unit cell with an atom at each vertex. Each atom in such a structure contributes four electrons, so that the atom is tetravalent. The connection between neighboring atoms stems from the fact that each of the four valence electrons of a Ge or Si atom is shared by one of its four adjacent neighbors. This is the so-called electron-pair, or covalent, bond. Because all four valence electrons take part in binding of one atom to another, no free carriers of electricity are available and the crystal behaves as an insulator, at least at very low temperatures (close to 0 K). In such a pure crystal, an electron is allowed to exist only in the valence band, where it is immobile, or in the conduction band, where it is free to move in an applied electric field. At 0 K the width of the forbidden band E_g is 1.21 and 0.78 eV for Si and Ge, respectively. Energies of this magnitude normally cannot be acquired from an applied electric field. Therefore, at absolute zero, the valence band remains full, the conduction band stays empty and the crystals are insulators. As the temperature is raised, an increasing number of the valence electrons can acquire energy greater than the forbidden bandwidth E_g and thus transfer into the conduction band. Once in the conduction band, these electrons are free to move under the influence of an applied electric field. The elevation of an electron to the conduction band leaves a vacancy in the valence band. This vacancy is known as a positive ''hole'' which under the influence of an electric field will move in a direction opposite to that of the electron. The field-induced motion of both electrons and holes contributes to the overall conductivity of the semiconductor. The production of equal amounts of conduction electrons and holes by thermal agitation is known as the intrinsic mechanism of conductivity. In an intrinsic semiconductor, the number of holes p is equal to the number of free electrons n. Whereas thermal excitation produces new hole–electron pairs, other hole–electron pairs disappear as a result of recombination. It is found that the intrinsic concentration n_i ($n = p = n_i$) varies with temperature T according to Equation (3):

$$n_i = AT^{3/2}e^{-E_{g0}/2kT} \tag{3}$$

where E_{g0} is the forbidden bandwidth at 0 K in eV, k is the Boltzmann constant in eV K^{-1} and A is a constant ($A = 1.8 \times 10^{16}$ cm^{-3} for Si and 8.5×10^{15} cm^{-3} for Ge). From Equation (3) it follows that the intrinsic hole or electron concentrations at 300 K are 1.5×10^{10} cm^{-3} in Si, and 2.5×10^{13} cm^{-3} in Ge. Another important conclusion is that the equilibrium intrinsic concentration, being an exponential function of temperature, decreases sharply as the material is cooled.

If an electric field is applied to the semiconductor material, the charge carriers (free electrons and holes) start to drift in the direction dictated by the field. Because of the inevitable inelastic collisions with ions of the crystalline lattice, a steady-state condition is reached in which an ultimate value of drift velocity is attained. This drift velocity v is proportional to both the applied field E and the mobility of the charge carriers (Equations 4a and 4b):

$$v_h = \mu_h E \tag{4a}$$

$$v_e = \mu_e E \tag{4b}$$

where μ_h and μ_e are mobilities of holes and free electrons, respectively. These particles move in opposite directions under the influence of the applied field E, but because they are opposite in sign, both currents are in the same direction. Hence the total current density J is given by Equation (5):

$$J = (n\mu_e + p\mu_h)qE = \sigma E \tag{5}$$

where n and p are the concentrations of free electrons and holes, respectively, and σ is the conductivity. Therefore, we have Equation (6):

$$\sigma = \frac{1}{\rho} = (n\mu_e + p\mu_h)q \tag{6}$$

where ρ is the resistivity of the semiconductor. The mobility μ is a function of temperature and electric field intensity.[18] It varies as T^{-m} over a temperature range of 100–400 K. For Si, $m = 2.5$ (2.7) for electrons (holes), and for Ge, $m = 1.66$ (2.33) for electrons (holes). With regard to the dependence on the electric field intensity, the mobility remains constant up to approximately 10^3 V cm^{-1}. At higher electric field values, the mobility decreases with increase in E, and the carrier speed approaches a constant value of the order of 10^7 cm s^{-1} (saturation velocity).

The forbidden bandwidth E_g in semiconductor also depends on temperature.[18] Experimentally, for Si, we have Equation (7):

$$E_g(T) = 1.165 - 1.66 \times 10^{-4} T \quad (7)$$

and at room temperature (300 K), $E_g = 1.11$ eV. Similarly, for Ge, we have Equation (8):

$$E_g(T) = 0.746 - 2.70 \times 10^{-4} T \quad (8)$$

and at room temperature, $E_g = 0.66$ eV.

The main constants describing the most important physical characteristics of Ge and Si are given in Table 2.[1,19]

In gas-filled detectors, the mobility of free electrons is typically 1000 times greater than that of ions, but in semiconductor materials the mobility of the holes is comparable to that of the electrons (see Table 2). As a result, the holes contribute a significant part of the overall conductivity and so improve considerably the timing characteristics of semiconductor detectors. Moreover, as shown in Table 2, the mobilities of the charge carriers increase by an order of magnitude at the temperature of liquid nitrogen (77 K). Thus the timing characteristics of semiconductors are improved even more by cooling to this temperature. As a result, semiconductor detectors are expected to be among the fastest of all radiation detectors.

From Equations (3) and (6) it follows that the resistivity of semiconductor materials increases very rapidly with cooling. Calculations based on Equations (3) and (6) and data presented in Table 2 show that the intrinsic resistivity of absolutely pure (i.e. without any impurities) Ge and Si increases at 77 K to such an extent that both materials are excellent insulators at this temperature. It may appear at first sight that either material could be successfully used at low temperatures to form the basis of a homogeneous solid-state detector. However, in practice any semiconductor material inevitably contains some small amounts of impurities which cannot be removed by the most advanced contemporary purification technologies. Moreover, the electrical properties of real semiconductor materials are largely dictated by these very small amounts of residual impurities. The intrinsic semiconductor is a very useful theoretical concept but in practice it is virtually unattainable.

2.3.4 *Physical Properties of Real (Doped) Germanium and Silicon*

The effect of the small concentration of impurity present in semiconductor material should be considered, either as a residual amount after the best purification possible or as a small amount intentionally added to the material to change its properties. If a small percentage of trivalent or pentavalent impurity atoms is added to tetravalent Si or Ge, a doped, impure or *extrinsic* semiconductor is formed. The impurity atoms displace some of the host atoms in the crystal lattice. If the dopant atom has five valence electrons, four of the five electrons occupy covalent bonds; the fifth is bound only very loosely to the original impurity atom and, as a result, will be available as a carrier of current. The energy required to detach the fifth electron from the atom is very small, of the order of 0.01 eV for Ge and 0.05 eV for Si. Suitable pentavalent impurities are P, As and Sb. Because such impurities donate excess negative charge carriers (electrons), they are referred to as donor, or n-type, impurities, and the semiconductor material containing them is called an n-type semiconductor. Donor impurities introduce a set of

Table 2 Some basic properties of intrinsic Ge and Si

Property	Ge	Si
Atomic number	32	14
Atomic weight	72.60	28.09
Density (300 K) ($g\,cm^{-3}$)	5.32	2.33
Atoms cm^{-3}	4.41×10^{22}	4.96×10^{22}
Dielectric constant (relative)	16	12
Forbidden energy gap at 0 K (eV)	0.746	1.165
Forbidden energy gap at 300 K (eV)	0.665	1.115
Intrinsic carrier density, n_i at 300 K (cm^{-3})	2.4×10^{13}	1.5×10^{10}
Intrinsic resistivity at 300 K (Ω cm)	45	2.3×10^{5}
Electron mobility at 300 K ($cm^2\,V^{-1}\,s^{-1}$)	3900	1350
Hole mobility at 300 K ($cm^2\,V^{-1}\,s^{-1}$)	1900	480
Electron mobility at 77 K ($cm^2\,V^{-1}\,s^{-1}$)	3.6×10^{4}	2.1×10^{4}
Hole mobility at 77 K ($cm^2\,V^{-1}\,s^{-1}$)	4.2×10^{4}	1.1×10^{4}
Carrier saturation velocity at 300 K ($cm\,s^{-1}$)	5.9×10^{6}	8.2×10^{6}
Carrier saturation velocity at 77 K ($cm\,s^{-1}$)	9.6×10^{6}	10^{7}
Energy per electron–hole pair at 300 K (eV)	(not applicable)	3.62
Energy per electron–hole pair at 77 K (eV)	2.96	3.76

discrete donor states just slightly below the conduction band. Because the energy spacing between the donor levels and the bottom of the conduction band is very small (0.01 eV for Ge and 0.05 eV for Si), almost all of the donor impurities are ionized at room temperature (Equation 3), and therefore, almost all the excess electrons of the donor atoms are raised into the conduction band. Because in most practical cases the impurity atoms are present in concentrations of the order of a few parts per million, the concentration of impurity atoms N_D is large compared with the concentration of intrinsic electrons n_i. Therefore, the free electron concentration in an n-type material becomes completely dominated by the contribution from the donor atoms ($n \approx N_D$). In an n-type semiconductor material, the number of electrons increases and the number of holes decreases, because the larger number of electrons present increases the rate of recombination of electrons with holes.

On the other hand, if a trivalent impurity, such as In, B or Ga, is added to an intrinsic semiconductor, only three of four covalent bonds are formed, leaving a vacancy (hole) in the fourth one. These impurities are known as acceptor, or p-type, impurities. They introduce discrete acceptor states just above the valence band, and the material itself is referred to as a p-type semiconductor because the principal charge carriers are the positively charged holes. Just as in the case of the donor impurity, only a very small amount of an acceptor must be added to change appreciably the conductivity of the semiconductor material. The ionized acceptor sites represent fixed negative charges, which balance the excess hole charges and do not take part in current conduction in the semiconductor. Usually the concentration N_A of acceptor impurities is made large enough to be compared with the intrinsic concentration of holes p_i. In such a case the number of holes is totally determined by the concentration of acceptors ($p \approx N_A$). As mentioned above, n-type impurities decrease the number of holes. Similarly, doping with p-type impurities considerably decrease the concentration of free electrons. The mass-action law controls the number of the free charge carriers in the semiconductor. It states[18] that, under thermal equilibrium, the product of the number of free negative (n) and positive (p) concentrations is a constant, independent of the amount of donor and acceptor impurity doping (Equation 9):

$$np = n_i^2 \tag{9}$$

where n_i is the intrinsic concentration given by Equation (3). The important result following from Equation (9) is that the doping of an intrinsic semiconductor increases the conductivity, and also produces a material in which the electric carriers are either predominantly holes or predominantly electrons. In an n-type semiconductor, the electrons are called the majority carriers and the holes are called minority carriers. In a p-type material, the holes are the majority carriers and the electrons are the minority carriers.

In practice, a semiconductor material always contains both types of impurities and their effects will at least partly cancel one another out, because the electrons produced by the donor impurities recombine with the holes generated by the acceptor impurities. The ultimate character of the semiconductor material (n-type or p-type) depends on whether the number of electrons generated by the donors exceeds the number of holes generated by the acceptors or vice versa.

The action of nuclear radiation on a semiconductor material ultimately results in the creation of electron–hole pairs. The average energy ε necessary to produce an electron–hole pair is found experimentally to be independent of the type and the energy of the incident radiation. The values of ε are 3.62 eV in Si at 300 K and 3.76 eV in Si and 2.96 eV in Ge at 77 K (Table 2). The smallness of the average energy ε compared with the average energy necessary to create an electron–ion pair in a gas (typically about 30 eV) represents an outstanding feature of semiconductor detectors and confers the superior energy resolution of these detectors. This follows from the increased (by about tenfold) number of charge carriers produced in the semiconductor as compared with the gas detector for the same energy deposited in the detector. The relative statistical fluctuation in the number of charge carriers per event becomes smaller as the total number is increased.

Once electrons and holes are formed in a semiconductor as a result of the action of nuclear radiation, these may recombine with one another, be trapped by impurities or imperfection centers or be collected at the electrodes of the detector, producing a useful signal. The charge carrier lifetime is measured as 10^{-2}–10^{-4} s in the purest semiconductor materials available. On the other hand, the mean charge carrier lifetime from direct electron–hole recombination is estimated to be as high as about 1 s. Therefore, direct recombination is rather improbable, and the main factor determining the lifetimes of the charge carriers is the trapping effect. The trapped charge carrier is released back to the relevant band; however, the time delay (detrapping time) is often too long and, as a result, the charge carrier is lost to the charge collection process or is collected with significantly reduced efficiency. On the other hand, if the mean detrapping time is substantially shorter than the charge collection time, then the trap has no effect on the charge collection process. In modern semiconductor detectors, the maximum permissible concentration of trapping centers is of the order of 10^{10} cm^{-3}, corresponding to approximately one

for every 10^{12} atoms of the host material.[19] This results in high charge collection efficiency of the order of 0.999. The corresponding collection time is of the order of 10^{-7} s and carrier lifetime is of the order of 10^{-4}–10^{-5} s.

In summary, the physical properties required for a semiconductor detector material are as follows:

1. high specific electric resistivity to give low leakage current;
2. freedom from recombination and trapping of charge carriers to give efficient (preferably 100%) charge collection;
3. small width of the forbidden band to give a large number of electron–hole pairs;
4. high mobilities of both charge carriers to give a fast time response;
5. high atomic number Z to give a high photoelectric cross-section for γ- and X-ray spectrometry.

2.3.5 Basic Properties of the p–n Junctions

Virtually all semiconductor detectors exploit the remarkable properties of the junction created between two semiconductor materials with different types of conductivity. Such a p–n junction is formed if donor impurities are introduced into one side and acceptors into the opposite side of a single semiconductor crystal. When n- and p-type semiconductors are brought into contact, the electrons from the n-type material diffuse across the junction into the p-type material and the holes from the p-type material into the n-type material because there is a density gradient of the free charge carriers across the junction. Free electrons diffusing from the n-type into the p-type material combine there with the holes and disappear as a result of annihilation. Each free electron that disappears from the n-side of the junction leaves behind an immobile positively charged donor site. Similarly, the holes diffusing in the opposite direction annihilate with the free electrons in the n-type material, leaving behind immobile negatively charged acceptor sites. The accumulated space charge from the immobile donor and acceptor sites creates an electric field which eventually halts further diffusion. Since the region of the junction is depleted of mobile charge carriers, it is called the depletion region, or the space-charge region. If the concentrations of the donors on the n-side of the junction and acceptors on the p-side are approximately the same, the depletion region extends equally into both sides across the junction. However, this is not the case in semiconductor detector practice. Here an intentional strong asymmetry in the doping levels on each side of the junction is produced. In such a case, the depletion layer extends almost entirely into the low-level doping side of the junction.

The configuration just described is a junction diode. The width of the depletion region of such a diode is of the order of a 0.5 μm and the potential difference across the junction (the contact potential) amounts to a few tenths of a volt. An electric field across the junction sweeps the free charge carriers out of the depletion region. As a result, the carrier density concentration remaining in the depletion region is approximately eight orders of magnitude lower than typical carrier densities in the bulk material.[1] Such a low concentration of free charge carriers in the depletion region means that the specific resistivity of this region is very high, as follows from Equation (6). As mentioned, high specific resistivity is one of the main properties required of the material for semiconductor detectors. The p–n junction diode can be used for the detection of nuclear radiation even with no external voltage applied because a nonzero electric field exists across the junction. If a charged particle enters the depletion region and creates electron–hole pairs, these will be swept out of the depletion region by the electric field towards the respective electrodes, and their flow forms a useful electronic pulse. However, the thickness of the depletion region in the unbiased junction diode is very small ($<10^{-4}$ cm). As a result, the active volume of the detector is too small to be of any practical interest. For the same reason, the capacitance of an unbiased junction is high, resulting in very poor signal-to-noise properties of the detector. Therefore, unbiased junction detectors are not used as practical detectors of nuclear radiation.

The thickness of the depletion region of the junction diode is increased considerably by applying a reverse bias to the junction. The reverse bias causes both the holes in the p-side of the junction and the electrons in the n-side to move away from the junction. Consequently, the region of the uncompensated negative charge is spread into the p-side of the junction, and that of uncompensated positive charge is spread into the opposite n-side. Because practical detectors are operated with a very large reverse bias voltage (up to 3000 V) compared with the contact potential, the thickness of the depletion region is increased considerably, reaching more than 0.1 cm for Si diodes. As mentioned, the depletion region width is determined by the biasing voltage and also by the free carrier densities. Because the net charge must be zero, then we have Equation (10):

$$N_A W_p = N_D W_n \tag{10}$$

where N_A and N_D are the acceptor and donor concentrations in the corresponding sides of the p–n junction and W_p and W_n are the depths to which the space charge extends into the p- and n-sides of the junction. In semiconductor detector practice, the n-side of the junction is usually doped to a much greater extent than the p-side. Consequently, the depletion layer extends much further into the p-side than the n-side. In such a case, the width d

of the depletion layer is given by Equation (11):

$$d \approx W_p = (2\varepsilon_o V \mu \rho_d)^{1/2} \quad (11)$$

where ε_o is the dielectric constant of the semiconductor, V is the reverse bias and ρ_d is the resistivity of the doped semiconductor.[1] From Equation (11), it follows that the largest depletion layer is obtained when a semiconductor material of the highest resistivity is chosen and the largest possible reverse bias is applied to the junction. For the material to be of highest resistivity, it must have the highest purity possible. The largest reverse bias is limited by the maximum electric field occurring at the junction. It must be no more than $2 \times 10^4\,\mathrm{V\,cm^{-1}}$ under typical conditions. In addition, the reverse bias should not be increased above the point at which the detector becomes totally depleted.

The pulse magnitude observed in a semiconductor detector after the charge collection process is inversely proportional to the sum of the capacitance of the p–n junction and the input circuit. The signal-to-noise ratio of the system is improved if the capacitance of the junction decreases. Therefore, the largest possible reverse bias is preferable.

2.3.6 Charged-particle Detectors

In most cases, detectors for charged-particle spectrometry are reverse-biased diodes made with highly purified, single-crystal silicon. Silicon is chosen because its energy gap is large enough to ensure that the resistivity is sufficiently high even at room temperature. In addition, the low atomic number of Si is of no importance for charged particle detectors.

For detecting charged particles, a depletion layer must be located as close as possible to the surface of a detector, because the regions outside the depletion layer act as a dead layer for incoming charged particles. There are a number of technologies for manufacturing silicon detectors with surface junction construction. For example, the diffusion method of fabrication starts by diffusing a donor impurity, such as phosphorus, at a high temperature (~800 °C) through the front face of a homogeneous crystal of p-type silicon. A p–n junction is formed at a distance of 0.1–1 μm from the surface of the crystal. Because the superficial n-type layer is heavily doped compared with the original p-type layer, the depletion layer extends primarily into the p-side of the junction. Therefore, the n-side of the junction represents a dead layer. The thickness of the dead layer should be determined experimentally to make possible accurate measurement of the energy of the incoming charged particles.

To avoid a large dead layer and the high temperature involved in the diffusion process, another type of surface junction detectors was developed, the surface barrier detector, which successfully replaced the diffusion detectors in many applications. Typically, a surface barrier detector is a diode made of an etched n-type silicon wafer with a very thin ($40\,\mu g\,cm^{-2}$) layer of gold deposited by evaporation on the Si wafer. The gold layer forms a p-type rectifying contact on the diode. The other surface of the wafer is coated with a layer of $40\,\mu g\,cm^{-2}$ of evaporated aluminum to produce good ohmic contact. When a reverse bias is applied to the contacts (i.e. a negative voltage on the gold electrode), electrons from the vicinity of the gold electrode are swept away, creating a depletion region that forms a sensitive medium for detecting charged particles.

The entrance window (gold layer) of surface barrier detectors is so thin (~0.02 μm) that it is transparent to the photons of visible light. To avoid a high noise level generated by normal light, surface barrier detectors are used in a lightproof environment, usually provided by the vacuum chamber required for most charged particle applications.

Another technique for producing the surface junction involves creating superficial layers of dopants by ion implantation. In this process, the ions of dopants from a small accelerator impinge on the surface of a crystal and penetrate a short distance into the crystal.

Surface junction detectors are available in a wide range of active areas from 7 to 2000 mm^2 and sensitive depths from 10 to 5000 μm and are therefore satisfactory for detecting many charged particles in a wide energy range.

From Equation (11) it follows that the width of the depletion layer is proportional to the applied voltage. According to the magnitude of the applied reverse-bias voltage, the detector is partially depleted, fully (totally) depleted or over-depleted. In the first case, the depth of the depletion region W is less than the physical thickness L of the detector. Partially depleted detectors are therefore sensitive to charged particles incident only on the junction surface of the detector.

Thin, totally depleted silicon detectors are widely used as transmission detectors for identifying charged particles based on their rate of energy loss dE/dx. These are so-called dE/dx detectors. For this application, a very thin detector is chosen so that the particle emerges with minimum energy loss and can be further detected by another detector (telescope arrangement). The dE/dx detector should have dead layers as thin as possible on both the entrance and the exit sides of the detector. Planar, totally depleted, silicon surface barrier detectors are currently available in thicknesses down to 10 μm.

As follows from Equation (11), the width of the depletion layer in the surface junction detector is determined by the resistivity (purity) of the starting silicon wafer and by the bias voltage applicable without

breakdown. The currently available purity of silicon is such that depletion thicknesses are limited to no more than a few millimeters. Greater thickness in silicon is obtained only with intrinsic material or through the use of material compensated by the lithium-drifted process.

Generally, all real semiconductor materials contain impurities of both n- and p-type. Because the electrons originating from donor impurities recombine with holes from acceptor atoms, it is obvious that a partial mutual compensation of impurities takes place. In such a case, the net concentration, $N_D - N_A$, where N_D and N_A are the concentrations of donor and acceptor atoms, respectively, determines the type of semiconductor. If $N_D > N_A$, then it is an n-type material, and vice versa. If, on the other hand, $N_D = N_A$, then the semiconductor behaves like an intrinsic material. Such compensated materials are designated with the letter "i".

The method of impurity compensation based on the process of lithium ion drifting was developed by Pell in 1960.[20] The method is used effectively for compensating p-type silicon and germanium crystals after the crystals have been grown. The lithium drift method is based on the fact that lithium ions, being of small radius, are fast interstitial diffusant in both Ge and Si and act as donor impurities with very low ionization potential (0.033 eV in Si and 0.0043 eV in Ge). The mobility and therefore the coefficient of diffusion of Li ions are some 10^7 times greater than those of ordinary diffusants such as boron and phosphorus.

In both Si and Ge, the material with the highest available purity is of p-type. The process of lithium compensation starts by diffusing lithium, which acts as a donor, into the front surface of a p-type semiconductor. To increase lithium mobility, the temperature of the crystal is elevated to 500 °C. As a result of diffusion, an n-type region is created near the exposed surface of the crystal. The resulting p–n junction is then back-biased, and the positive lithium ions are slowly pulled further by the electric field into the p-region where their concentration increases and approaches the concentration of the p-type impurities in the original material. Simultaneously, the lithium donor concentration decreases in the n-region. The lithium ions, having high mobility, redistribute themselves under the influence of the local fields so that the total space charge in the drifted region tends towards zero at every point. Otherwise the local field would essentially reverse direction and would sweep the lithium ions back in the opposite direction. An equilibrium point is thus established in which the lithium ions go on to drift into the p-region, spreading themselves over the region so that the number of lithium donors compensates exactly for the acceptors of the original material. It has been shown by Pell[20] that any deviation from exact compensation by the lithium ions during the drifting is unstable, and any instantaneous imbalance in lithium concentration is quickly eliminated by itself, thus restoring exact compensation. Using the lithium drifting process, thick compensated regions of up to 5–10 mm in Si and 10–15 mm in Ge are achieved. In this case, the drift process can take hundreds of hours. More detailed information about the lithium drift process is available.[15]

Lithium ions have high mobility at room temperature, particularly in germanium. Therefore, to preserve the lithium concentration required for compensation, the Ge crystal must be cooled to liquid nitrogen temperature (77 K) immediately after the desired compensation is attained. In silicon, however, the lithium ion mobility is lower, so that lithium-drifted silicon detector can be temporarily stored at room temperature without cooling.

As mentioned, the problem of insufficiently thick depletion zone in semiconductor detectors was solved with the developing the lithium drifting process to obtain compensated material. Such silicon based detectors are called *lithium-drifted silicon detectors*, or Si(Li) detectors (pronounced "silly"). Junctions of the Si(Li) detectors are known as p–i–n junctions and their properties differ from the ordinary n–p junctions discussed earlier. Because, theoretically, no net charge exists in the compensated i-region, the resulting electric field is constant across the i-region, dropping sharply to zero at its boundaries. Therefore, the active volume of the detector is determined by the dimensions of the i-region.

The increased thickness of the compensated (active) zone of the Si(Li) detectors (5–10 mm) makes them suitable for the detection and spectroscopy of β-particles and low-energy X- and γ-rays. For the same reason, the bulk generated leakage current of Si(Li) detectors is a significant contributor to the noise. As a result, the detectors in these applications must be cooled to liquid nitrogen temperature. For this purpose they are mounted in a vacuum-tight cryostat incorporating a built-in sorption pump. A thin beryllium or Mylar window allows entry of the low-energy radiation into the detector. The cryostat is mounted on a liquid nitrogen Dewar vessel with a capacity of about 30 L. To diminish the contribution of the electronic noise in the energy resolution of Si(Li) detectors, they are supplied with a low-noise preamplifier incorporating a cooled first stage. Being low-energy γ-ray spectrometers, Si(Li) detectors are extensively used for the observation of the Mössbauer effect, where energies below 40 keV are of principal interest. As X-ray spectrometers they are widely used in source and tube-excited X-ray analysis for atmospheric, air and water pollution studies.

2.3.7 Germanium Detectors

Germanium detectors are mainly used in γ-ray spectroscopy because the atomic number of Ge ($Z = 32$)

is much higher than that of Si ($Z = 14$). The photoelectric probability, depending upon Z^4 to Z^5, is thus about 60 times greater in Ge than in Si. In general, much greater sensitive thicknesses are required to detect γ-rays than charged particles. Before the mid-1970s the required depletion depths were achieved only by counter doping p-type germanium crystals with lithium in the lithium-ion drifting process described above in connection with Si(Li) detectors. These lithium-compensated Ge-detectors are known as Ge(Li) (pronounced "jelly") detectors.

To increase overall detector volume and thus detection efficiency, the Ge(Li) detectors are constructed in cylindrical or coaxial geometry. In true coaxial geometry, lithium diffuses on the outside of the p-type germanium crystal and drifts towards a central core under a strong electrical field, thus producing a cylindrical shell of high resistivity (intrinsic) material. A central core of insensitive p-type germanium extends along the entire length of the cylindrical crystal. In a closed-end coaxial configuration, lithium is also allowed to drift from the front face of the Ge crystal. As a result, the sensitive volume of the detector is increased even more. Coaxial Ge(Li) detectors are also constructed in a well configuration, in which the central insensitive core of the germanium cylinder is removed. Radioactive samples are then placed within the well, thus substantially increasing the counting efficiency. For lower γ-energies (up to several hundred kiloelectronvolts), Ge(Li) detectors are fabricated in a planar configuration, in which the electrical contacts are formed on the flat surfaces of a germanium disk a few centimeters thick.

To reduce the thermal charge carrier generation (electronic noise) to an acceptable level, Ge(Li) detectors must be cooled to liquid nitrogen temperature. The cooling of Ge(Li) detectors prevents lithium precipitation and thus protects from destruction. This is necessary because of the high mobility of lithium ions in germanium at room temperature. For this reason, Ge(Li) detectors must be kept at liquid nitrogen temperature at all times (both in operation and on storage). The germanium crystal and input stage of the preamplifier are installed in a vacuum chamber which is inserted in a liquid nitrogen container (Dewar), just as for Si(Li) detectors (Figure 4). At these temperatures, reverse leakage currents are very low, in the range 10^{-9}–10^{-12} A.[21]

In the mid-1970s, advances in germanium purification technology produced high-purity germanium with an impurity level as low as 10^{10} atoms cm^{-3}. Detectors produced from this ultrapure Ge material are called high-purity germanium detectors, usually abbreviated as HPGe detectors. The outstanding feature of these detectors is that they do not have to be kept at liquid nitrogen temperatures at all times. HPGe detectors must

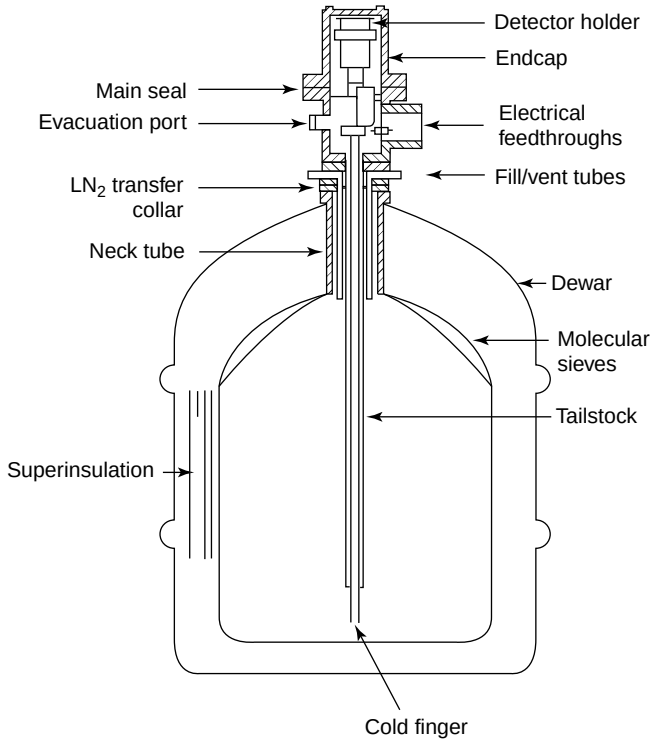

Figure 4 Ge detector attached to a dipstick cryostat inserted in a liquid nitrogen container.[21]

be cooled only when a high voltage is applied. Between uses, these detectors can be held at room temperature.

Because HPGe and Ge(Li) detectors are virtually identical in construction and operation, their main performance characteristics are practically the same. On the other hand, because HPGe detectors are obviously much more convenient to operate than Ge(Li) detectors, they have almost completely replaced Ge(Li) detectors in contemporary γ-ray spectroscopy.

One of the main measures of the performance of nuclear detectors is their resolution. In general, resolution is the ability of a detector to separate two peaks close together in energy. Therefore, the narrower the peak, the better the resolution of the detector. Resolution is measured as the full width of the peak at half the maximum amplitude [full width at half-maximum (fwhm)]. Germanium detectors feature very high energy resolution capability compared with other types of radiation detectors (proportional counters or scintillation counters). Basically, the better energy resolution of Ge detectors is attributed to the small amount of energy required to generate a pair of charge carriers (electron–hole) in germanium. As a result, a greatly increased number of charge carriers is produced, forming the large output signal relative to that of other detectors. At about 3 eV per electron–hole pair (see Table 2), the

Table 3 Representative fwhm values (keV) of Ge detectors compared with various detector types

Counter	Energy (keV)		
	5.9	122	1332
Proportional counter	1.2	–	–
3 × 3 in NaI(Tl)	–	12.0	60
Si(Li)	0.16	–	–
Planar Ge	0.18	0.5	–
Coaxial Ge	–	0.8	1.8

number of charge carriers produced in Ge is about one and two orders of magnitude higher than in gas and scintillation detectors, respectively. The energy resolution of Ge detectors expressed as fwhm in kiloelectronvolts is often specified at 5.9 keV (^{55}Fe), 122 keV (^{57}Co) and 1332 keV (^{60}Co).The representative fwhm values and the resultant improvement in energy resolution of Ge detectors over other types of radiation detectors are presented in Table 3.[21]

The striking difference in energy resolution between a Ge detector and a 3 × 3 in NaI(Tl) scintillation detector is illustrated in Figure 5. The germanium detector used is a coaxial Ge(Li) detector with a 7.5% efficiency [compared with a standard 3 × 3 in NaI(Tl) crystal].

Efficiency is another important parameter of nuclear detectors. In general, the absolute, or total, efficiency of a γ-detector is defined as the ratio of the number of events registered by the detector to the number of γ-rays emitted by the source in all directions. The total efficiency depends on the probability of the interaction of a γ-ray with the detector sensitive medium and on the solid angle subtended by the detector. For historical reasons, a relative detection efficiency of germanium detectors is in common use in γ-ray spectrometry. It is defined at 1.33 MeV (^{60}Co) relative to the efficiency of a standard 3 in diameter, 3 in long NaI(Tl) scintillator; each detector at 25 cm from a point ^{60}Co source. Because the absolute efficiency of a standard 3 × 3 in NaI(Tl) crystal is well known (1.2×10^{-3}), the relative detection efficiency is

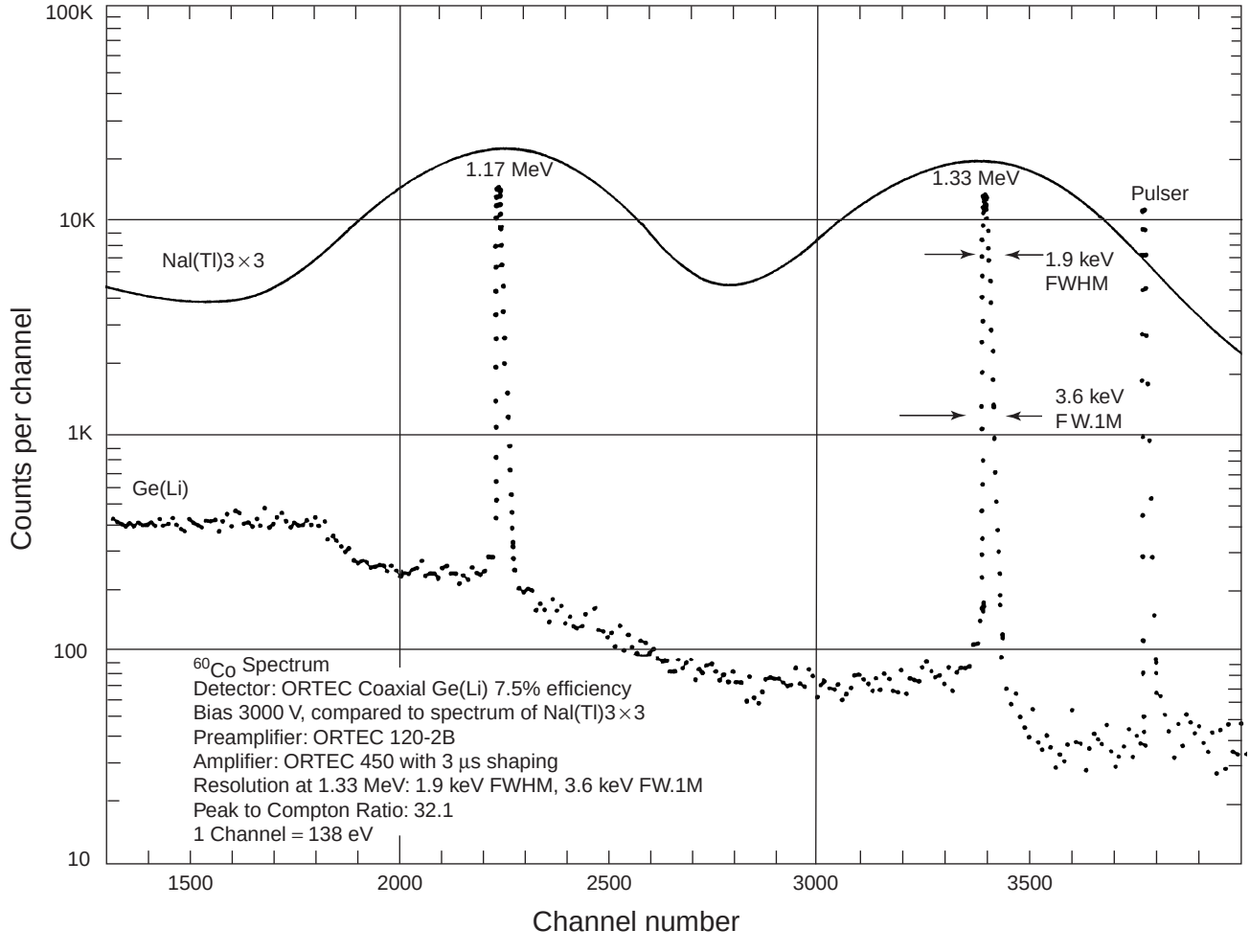

Figure 5 ^{60}Co spectrum for NaI(Tl) and Ge(Li) detectors.[34] (Reprinted by permission from ORTEC, Inc.)

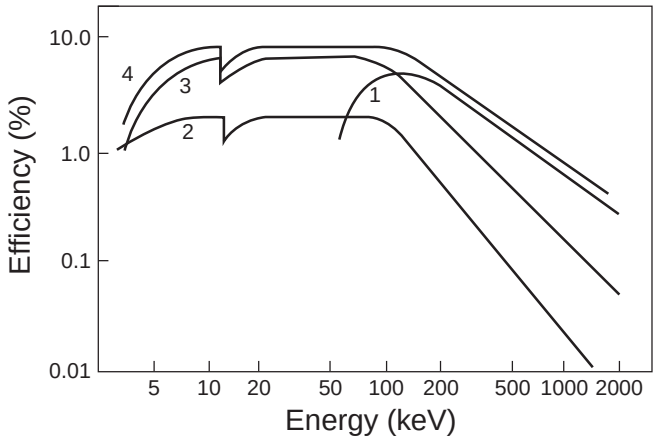

Figure 6 Typical absolute efficiency curves for various Ge detectors with 2.5 cm source to endcap spacing: curve 1, 10% relative efficiency coaxial HPGe detector; curve 2, 200 $mm^2 \times 10$ mm thick low-energy Ge planar detector; curve 3, 10 $cm^2 \times 15$ mm thick low-energy Ge detector; curve 4, 15% relative efficiency reverse-electrode Ge detector.[21]

easily obtained. Coaxial Ge detectors of greater than 100% relative efficiency have been fabricated from germanium crystals ranging up to ~75 mm in diameter.[21] About 2 kg of germanium is required for such a detector.

Typical absolute efficiency curves for various types of germanium detectors are presented in Figure 6. Curve 1 represents a 10% relative efficiency, coaxial, HPGe detector. It is used in general γ-spectroscopy for the energy range from 50 keV to more than 10 MeV. Curve 2 represents a 200 $mm^2 \times 10$ mm thick planar detector. The detector element is made of HPGe. Its cryostat window is made from beryllium for improved sensitivity to low-energy photons. The detector is used mainly in low-energy photon spectrometry in the energy range 3–200 keV. Curve 3 shows the absolute efficiency curve for a 10 $cm^2 \times$ 15 mm thick, low-energy, germanium detector. This type of detector is fabricated with an ultrathin boron implanted *p*-contact on the front face and on the cylindrical wall. The *n*-contact, a lithium-diffused spot on the rear face, is of less than full area, and thus the capacitance of the detector is less than that of a planar device. Because preamplifier noise is a function of detector capacitance, the low-energy germanium detector affords lower noise and consequently better resolution at low and moderate energies than any other detector.[21] For applications involving moderate γ-ray energies (5–500 keV), this type of detector outperforms an expensive large-volume coaxial detector. Curve 4 gives the absolute efficiency data for a 15% relative efficiency, reverse-electrode detector. The main design feature of the detector is the reversal of the electrodes from the configuration in a conventional coaxial detector. The p-type electrode (ion-implanted boron) is on the outside, and the n-type contact (diffused lithium) is on the inside of the Ge crystal. The ion-implanted outside contact is extremely thin compared with a lithium-diffused contact. In conjunction with a thin cryostat window, this results in an energy response down to 5 keV, giving this detector a dynamic range of 2000 : 1 (5 keV to 10 MeV). Another feature of this detector is that it is about 10 times more resistant to radiation damage than conventional coaxial Ge detectors.[21]

2.3.8 Semiconductor Detectors in Analytical Chemistry

High-resolution γ-ray spectroscopy is widely used in the evaluation of γ-emitting radioactive nuclides in various samples, such as soil, air, water and foodstuffs. In most cases, such an evaluation is based on the use of Ge detectors which have a number of benefits in analytical γ-spectroscopy. First, high-resolution, large and superlarge germanium γ-ray detectors are available with relative efficiencies up to 120%. Another benefit is the availability of low and ultralow γ-background cryostats and transplantable detector capsules which are easily switched from one cryostat to another without using any pumps, heater tapes or other accessories.[22]

One purpose of radioactivity measurements at a low level (e.g. environmental samples) is to determine their contamination level and to limit the internal exposure of humans to radioactive species. The permitted levels of radioactivity are relatively low [e.g. following the Chernobyl nuclear accident, ad hoc limits were imposed on the content of radioactive Cs, typically 600 Bq kg^{-1} in food and 370 Bq kg^{-1} in milk[23]]. Other examples of the low-level measurements of radionuclides include the monitoring of the environment and effluent discharge of nuclear power plants and evaluating the natural background. Because of diverse radionuclide content of the low-level samples, it has become standard practice to use high- resolution Ge detectors to resolve, identify and quantify the γ-emitting isotopes present.

The main performance parameters of Ge detectors, relating to low-level and environmental measurements, are efficiency (both at 25 cm and Marinelli efficiency), energy resolution, peak shape and peak-to-Compton ratio.

The efficiency (at 25 cm) of Ge coaxial detectors is described in concise form in the previous section. The Marinelli efficiency is especially important in environmental applications involving measurements of weak activity. In such measurements, the environmental sample to be evaluated comprises large-volume materials, such as different fluids, soil or vegetation. In the early days of environmental studies, it was the practice to reduce the sample volume (e.g. by evaporation, ashing or leaching) in order to obtain a reproducible and optimized sample–detector geometry. To avoid the problems

associated with the sample processing (cost and possible sample destruction), samples are now measured directly in Marinelli beakers without any preprocessing.

As mentioned above in connection with scintillation counting, the Marinelli (re-entrant) beaker allows the counting of an environmental sample in an optimum and reproducible geometry with minimum background by placing it as close to the sensitive detector volume as possible, a prerequisite for all low-level counting systems. This makes it necessary to specify the Marinelli beaker efficiency which differs, obviously, from the detector efficiency at 25 cm. Measurements show[24] that although a large-diameter Ge detector may have a high relative efficiency for a given sensitive volume, when used with a Marinelli beaker geometry, it has an absolute full-energy peak efficiency lower than that of the lower relative efficiency detector. For these reasons, an American Nuclear Society (ANS)/Institute of Electrical and Electronic Engineers (IEEE) Standard has been introduced for the measuring detectors using the Marinelli beaker geometry. According to this standard, a detector is measured using a Marinelli beaker filled with 450 mL of a solid or liquid carrier containing a specified, uniformly dispersed, radioactive material. The solid carrier is preferable because it constitutes a safe and chemically inert sealed source. The filled beaker is designated a Marinelli Beaker Standard Source (MBSS).[24] Employing the Marinelli beaker geometry, detectors are compared to provide meaningful measurements for different users and sites.

Energy resolution and good peak shape (symmetry) are very important parameters in the resolution of complex radionuclide mixtures in low-level counting. As mentioned, energy resolution is characterized by the fwhm at 1.33 MeV. Peak shape, the symmetry of the peak is measured by the full width at one-tenth maximum (fw.1m)/fwhm ratio at 1.33 MeV. Ge detectors typically have fw.1m/fwhm ratios of 1.9 or even less, which is very close to the ultimate value of 1.83 for the theoretical Gaussian peak.

The peak-to-Compton ratio is one of the most important measures of Ge detector performance. It characterizes a detector's ability to distinguish low-energy peaks in the presence of high-energy sources. The peak-to-Compton ratio is obtained by dividing the height of the 1.33 MeV peak by the average Compton plateau between 1.040 and 1.096 MeV. Usually, higher peak-to-Compton values are achieved with better values of energy resolution. Ge detectors with good peak-to-Compton ratios are especially valuable in applications such as NAA and radiation waste analysis. In these applications, the isotopes of interest may have low activity compared with the activities of other isotopes in the sample, and therefore, may be hidden by the Compton plateau.

In the low-level counting applications, obviously, there is a need for Ge detectors with a low radioactive background, both external and internal. The main contributors to the external background are the cosmic rays, ^{40}K from building structures and radon. This kind of background can be markedly reduced by appropriate shielding and by flushing the volume around the detector with a radon-free gas (boiled-off liquid nitrogen is often used). The main source of internal activity is the detector's cryostat. It contains the primordial emitters ^{238}U, ^{232}Th, ^{235}U and their progeny, which emit a mixed spectrum of α-, β-, and γ-rays. In addition to the naturally occurring activities, there are also fission-related activities including ^{137}Cs fallout and ^{60}Co. This background includes both photopeaks and associated Compton continua, and it covers energy range from 46 to 2614 keV.

To achieve the ultimate in low-background counting, the main parts of the cryostat are made of special low background materials, such as magnesium (endcap), oxygen-free high conductivity (OFHC) copper (detector cup, cooling rod), activated charcoal (for cryosorption pumping) and low-cobalt stainless steel (cryostat flange).[19] For the same reason, shielding of Ge detectors is essential to reduce external background radiation. A special low-background virgin lead with steel jacket is often used for this purpose. A 10-cm thickness of lead is sufficient for low-level counting applications covering the energy range from 0 to 2.6 MeV. Still greater thicknesses provide additional attenuation of γ-rays, but also increase the background from undesired cosmic-ray interaction with the lead shield. To reduce the unavoidable K X-ray of the lead (80 keV), a graded-Z shield is used, consisting of materials with decreasing atomic numbers (Z) approaching the detector to absorb the lead X-ray and emit a secondary X-ray of lower energy. Typical graded-Z shields are $_{82}$Pb($\sim$10 cm, $E_x = 80$ keV)$-_{48}$Cd($\sim$0.5 mm, $E_x = 22$ keV)$-_{29}$Cu($\sim$1.5 mm, $E_x = 3$ keV).

For even lower-level measurements, active shielding methods, such as a Compton-suppression spectrometer,[1,25] provide additional enhancement of the signal-to-background ratio and improve still further the minimum detectable activity (MDA). The Compton-supression system is based on a well known and established fact that in the energy range between $\sim$200 keV and several MeV the most dominant interaction process between γ-photons and a Ge detector is Compton scattering, which gives rise to a powerful continuous (Compton) background in the energy spectrum, thus inevitably obscuring the low-activity photo-peaks to be sought in the same energy interval. Therefore, to improve the signal-to-background ratio and the MDA, the Compton continuum must be supressed. This is done by means of anti-Compton shields surrounding a Ge detector and accompanying electronics. Such a system consisting of a

combination of a Ge detector and an active (electronic) anti-Compton shield is called a Compton supression spectrometer. In this spectrometer, the photons Compton scattered by the active region of the Ge detector are detected by means of a large scintillation crystal [NaI(Tl) or BGO] surrounding it. If the signal generated by the scattered photon in the surrounding anti-Compton shield is time-coincident with the signal generated by the Compton electron in the Ge detector itself, then the signal from the shield can be effectively used to reject the registration of Compton-based interactions in the Ge detector. In this way the Compton continuum background can be supressed by an order of magnitude. Systems capable of peak-to-Compton values as high as 940 : 1, depending on the Ge detector chosen, have been reported. The NaI detector used for this purpose was a 9×9 in axial annulus, with a 3×3 in plug detector, the detector system being housed in a $35 \times 35 \times 75$ cm lead shield, with 4-in thick walls, lined with 0.020-in Cd and 0.060-in Cu.[25]

Indubitable benefits of using large, high-efficiency Ge detectors for the quantitative determination of environmental radionuclides have been discussed.[26] It is pointed out that lower detection limits are obtained with higher-efficiency Ge detectors than with smaller ones. A single large detector results in a substantially higher sample throughput than multiple smaller detectors of the same total efficiency. This translates directly into cost-per-sample saving in a low-level counting laboratory.

In a wide range of applications, such as environmental restoration of previously used nuclear sites, assessment of radionuclides deposited during emergencies, decommissioning and routine monitoring of nuclear facilities, there is a need for prompt and economical in situ γ-spectroscopy. Such an in situ γ-spectrometer must be lightweight and portable, yet rugged enough to withstand a broad range of harsh environmental conditions. The best for this application is a portable, high-resolution HPGe detector mounted in a transplantable detector capsule attached to an all-attitude cryostat. A small-capacity Dewar vessel (e.g. 1.2 L) provides at least 12 h of continuous operating time. The measurements are made with the detector positioned 1 m above the ground. The methodology of these measurements was developed by members of the United States Department of Energy (USDOE) Environment Measurement Laboratory.[27,28] This methodology is widely accepted for measurement of radionuclide-specific activity in the ground. It is based on empirically derived correction factors for detector orientation and for angular depth distribution of radioactive sources. By applying these correction factors to measured radionuclide intensity, an estimate of radioactivity per unit area or unit volume can be made.

Having unsurpassed energy resolution, HPGe detectors are widely used as photon spectrometers in NAA. A well-resolved spectrum not only helps separate neighboring photopeaks but also aids in the detection of weak photons on the spectral continuum. This is particularly important for multielemental analysis when many elements in the sample matrix are irradiated by neutrons and emit up to several hundred prompt γ-rays simultaneously. As the HPGe detector is placed as close as possible to the neutron-irradiated sample (prompt activation analysis) to acquire the maximum counts, the detector is inevitably subject to neutron bombardment and excitation. In such a case, a typical prompt γ-ray spectrum of the HPGe detector itself under irradiation by neutrons should be used as a reference for the spectral background during prompt γ-counting.[13] It should be pointed out that because the HPGe detector is subject to neutron irradiation during prompt γ-analysis, various degrees of damage to the detector can result from prolonged exposure to neutrons. For example, an exposure of an HPGe detector to a fast neutron fluence of 10^9 n cm^{-2} is sufficient to risk the high-resolution performance and the detector may become unusable with exposure beyond 10^{10} n cm^{-2} owing to destruction of the crystal lattice.[29]

Another possible application of HPGe detectors is in a high-sensitivity PIXE method in which X-ray emission is used for elemental analysis of all elements from boron to uranium bombarded by high-energy hydrogen or helium atoms with kinetic energies of a few MeV. HPGe detectors can be used up to and above 100 keV, which permits the detection of characteristic K X-rays for all elements. At lower X-ray energies, Si(Li) detectors must be used. Such detectors are sensitive to X-rays with energies between 1 and 30 keV. Above 30 keV the X-ray cross-section falls steeply and it is then not efficient to measure K X-rays from the heavier elements. Instead, L X-rays are used for the analysis of heavy elements, as they are more intense and are more efficiently detected in the Si(Li) detector.[30]

The Si-based semiconductor detectors are used for charged-particle detection and spectroscopy. Ion-implanted, passivated, partially depleted Si detectors are widely used in α-spectroscopy of environmental samples for transuranics and to assess intake of these α-emitting elements. The light-tight option of the ion-implanted silicon detector with a protective polymer front-contact coating is employed for α- and β-continuous air monitoring in an adverse environment.

Because α-particles are heavily attenuated even by air, α-spectroscopy with Si detectors must be performed with the samples in a vacuum chamber. For the same reason, extensive sample pretreatment is often required depending on the radionuclide involved and the medium in which it is contained.[31]

Typical applications of environmentally rugged Si detectors include use in portable survey instrumentation, air monitors, α area monitors, personal α alarms/monitors,

alarm pocket dosimeters and others. Silicon detectors are used for α detection from radon and radon progeny in radon monitors and radon working level meters.

3 γ-RAY SPECTROSCOPY

3.1 Instrumentation for Nuclear Spectroscopy

Nuclear detectors provide a variety of information on detected radiation (energy, type of the radiation, timing data, etc.) in the form of electronic signals. To obtain this information, the signal must be processed by the appropriate nuclear electronics system. However, the subject of nuclear electronics is outside of the scope of this article and interested readers are referred to appropriate reference books.[1,32,33] In what follows we present a short, qualitative description of the process of obtaining the most important characteristic of nuclear radiation, its energy distribution (energy spectrum).

A schematic diagram of the basic electronic equipment for measuring the energies of γ-rays is presented in Figure 7. A photon detector is represented as a capacitor in which charge carriers are generated by interaction of a nuclear particle within the detector. The high voltage provides the detector bias to collect the electric charge. As a result, a small (pulse) current flows, producing the pulse voltage drops across the bias resistor R. The voltage pulse produced enters into the preamplifier isolated from the high voltage by the capacitor C. In general, the preamplifier serves for the initial amplification of the detector output signal. The preamplifier is usually located as close to the detector as possible to reduce excessive cable capacitance, which lowers the pulse height and degrades the rise time of the pulse. Also, such positioning amplifies the signal before it enters a long cable leading to the main amplifier. Because the capacitance of most solid-state detectors depends upon the high-voltage bias, a charge-sensitive preamplifier should be used. The voltage pulse produced at the output of the preamplifier is then be proportional to the collected charge and independent of detector capacitance.

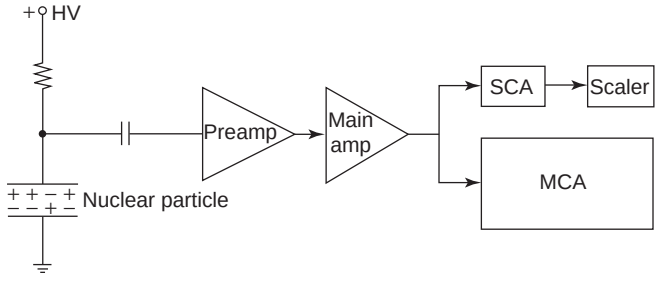

Figure 7 Schematic diagram of electronics for energy measurements of a nuclear radiation.

The main amplifier shapes the pulse for further processing and amplifies the pulse from the preamplifier to make it compatible with the input range of the spectrum analyzer. In energy spectroscopy, the pulse-height information is of primary importance. Therefore, very strict proportionality between the input and output amplitudes of the main amplifier must be maintained. Generally, the pulse at the output of a preamplifier has a fast rise time, ranging from a few nanoseconds up to a few microseconds, and a slow decay time of about 50 μs. The amplitude of this pulse is proportional to the energy of the radiation to be measured. Therefore, the long decay time of the preamplifier output pulse may cause the distortion of the energy information of the second pulse, if it rides on the tail of the first pulse. This is the so-called pile-up effect.

Another shortcoming of tailed pulses is their bad signal-to-noise ratio;[34] in fact, it would be more advantageous to have a Gaussian or triangular pulse shape. For this purpose, RC clipping is used, in which the tailed pulse is differentiated to remove the slowly varying tail. Subsequent integration reduces the noise. The resulting pulse is much shorter and has a near-Gaussian shape with improved signal-to-noise characteristics and count rate capabilities.

Specially shaped pulses from the output of the main amplifier are directed into the pulse-height analysis system. This may consist of either a single-channel analyzer (SCA) and a counter or a MCA (see Figure 7).

The SCA is an electronic device which analyzes and sorts incoming pulses according to their amplitudes. It has both lower-level and upper-level discriminators, which define a certain amplitude range in the main amplifier output. The span between the lower-level and upper-level discriminator thresholds is called the SCA window. Only pulses with amplitudes falling within the SCA window result in a standard logic output pulse from the SCA. Pulses with amplitudes below the lower level setting or greater than the upper-level setting are rejected and do not produce output pulses. With the detector output proportional to energy, the SCA is used to obtain energy spectra of the incident radiation. For this purpose, the lower level and upper level of the SCA are set to give a fixed, narrow window, and full sweeping of the window is performed across the pulse-height range under consideration. The number of pulses counted by scaler (counter) at each step and normalized per unit time is then presented as a histogram of the energy spectrum. However, full data collection and processing by a SCA are tedious and time-consuming. Automatic and fast acquisition of energy spectra data is achieved by the MCA. An MCA can be considered as a succession of SCAs with the same window-width settings. All windows are arranged sequentially in order of increasing energy. Thus, by representing the number of counts of each SCA versus the mean energy of the corresponding window, a

histogram of a number of counts versus energy, i.e. an energy spectrum, is obtained.

In an actual MCA, the succession of SCAs is replaced by a special device called an analog-to-digital converter (ADC). This is the core of an MCA. The ADC measures and digitizes the height of each pulse to be analyzed and determines to which pulse-height window this pulse corresponds. The MCA then takes this window number and adds to a memory channel whose address corresponds to the digitized value. In this way all incoming pulses are sorted according to pulse heights represented by channels. The content of each channel is stored in the memory. The total number of channels into which the voltage range of input pulses (0–10 V) is digitized is called the conversion gain. This gain determines the resolution of the MCA and ranges from 64 up to 16 K in contemporary MCAs.

An input pulse is delayed while checking is performed to see if it is within the selected SCA range. If it is, the input pulse is passed to the ADC. The ADC digitizes its amplitude, generating a number proportional to the pulse voltage. This number represents an address in the memory to which one count is added. The display reads the memory many times per second, producing a point plot of memory contents versus memory locations, which is equivalent to the number of pulses versus voltage, i.e. energy spectrum.

MCAs provide a number of possibilities for manipulation of a spectrum display and the data stored in memory. This includes expanding or contracting, linear or logarithmic display modes and calculational capabilities, such as energy calibration, peak searching, area extraction, background subtraction and many others.

3.2 Energy Measurements

There are three principal kinds of interactions of photons with matter, the photoelectric effect, Compton scattering and pair production. All three processes are strongly dependent on the energy of the photon and the atomic number Z of the material. In the photoelectric process, a photon is absorbed by an atom and, as a result of the absorption, a photoelectron is emitted. Its energy is equal to photon energy minus the binding energy. The photoelectric cross-section increases very strongly with increase in Z of the material and decreases with increase in photon energy. The photoelectron produces the electron–hole pairs in the sensitive medium of the detector, resulting in the output pulse. Electrons have a short range in the crystal and therefore all of their energy is deposited in the detector.

At higher photon energies, the dominant interaction is Compton scattering. In the Compton process, the photon interacts with what might be termed a free electron, losing part of its energy to the electron and scattering to a different direction. Because the energy of the original photon is shared between two particles in Compton scattering (the recoil electron and the scattered photon), there is a continuous distribution of pulse amplitudes of Compton electrons up to some maximum pulse height. This maximum pulse height produces the Compton edge, which is the maximum energy that can be imparted to an electron in head-on collision. It is given by Equation (12):

$$E_e^{max} = \frac{E_0}{1 + m_0c^2/2E_0} \tag{12}$$

where E_e^{max} is the maximum electron energy, E_0 is the incident photon energy, m_0 is the electron rest mass and c is the velocity of light.[1]

Pair production is the third important process of photon interaction with matter. The photon enters the detector and creates an electron–positron pair in the field of a nucleus. From the law of conservation of mass and energy, it follows that the initial photon must have an energy of at least 1.022 MeV ($=2m_0c^2$) because this is the exact energy necessary to create both an electron and a positron.

To show the relative probability of each of the three types of interactions, in Figure 8 graphs of the absorption cross-section for germanium are presented. As shown, at low energies up to about 200 keV, the photoelectric effect is dominant. Compton scattering becomes more important near 200 keV. Above 1.02 MeV, pair production increases in probability.

All three processes determine the response of a solid-state detector to monoenergetic γ-rays. The relative importance of these processes for detector response lies in the proportion of the incident γ-ray energy deposited in the detector-sensitive medium. A typically idealized response of a solid-state detector to monoenergetic γ-rays is presented in Figure 9.

The full-energy peak, or photopeak, results from the γ-photon losing all of its energy by photoelectric absorption or Compton scattering followed by photoelectric absorption in the detector. Compton scattering causes only a fraction of the energy of the initial photon to be deposited in the crystal because the scattered photon can escape from the crystal without further interaction. This process contributes to the energy response of the detector in the so-called Compton continuum, ranging from zero up to a maximum determined by the Compton edge (Equation 12). If the energy of an incident photon is above 1.022 MeV, pair production occurs. The positron produced slows in the crystal to an energy close to that of the atomic electron and annihilates, producing two 0.511-MeV photons. Each of these photons escapes from the detector totally or deposits only part of its energy by Compton scattering. If one of these photons escapes totally, the total energy absorbed is 0.511-MeV less than the original incident photon energy. If both

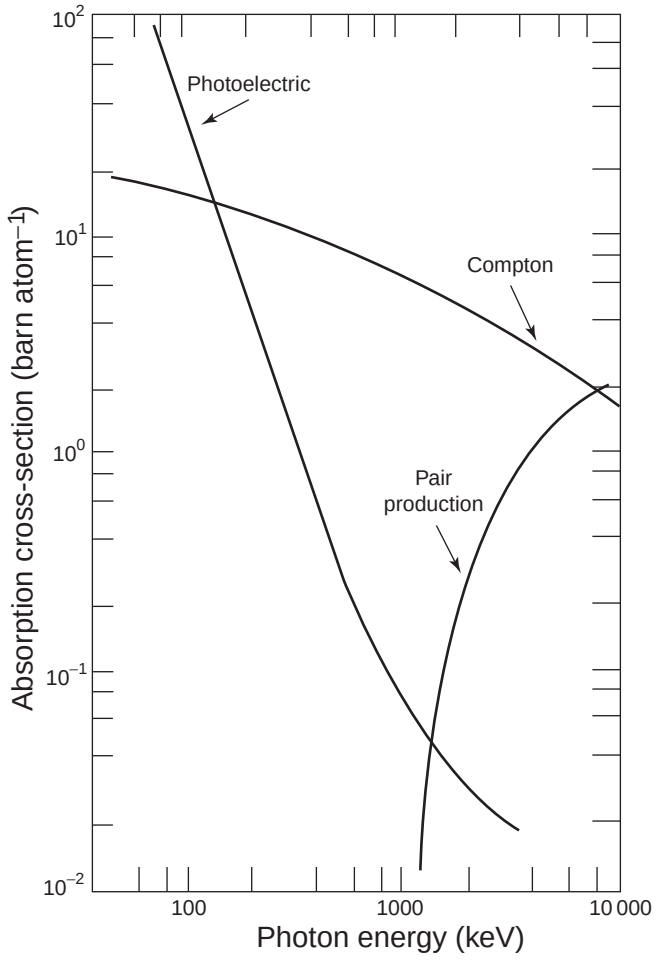

Figure 8 γ-Absorption cross-section in Ge.

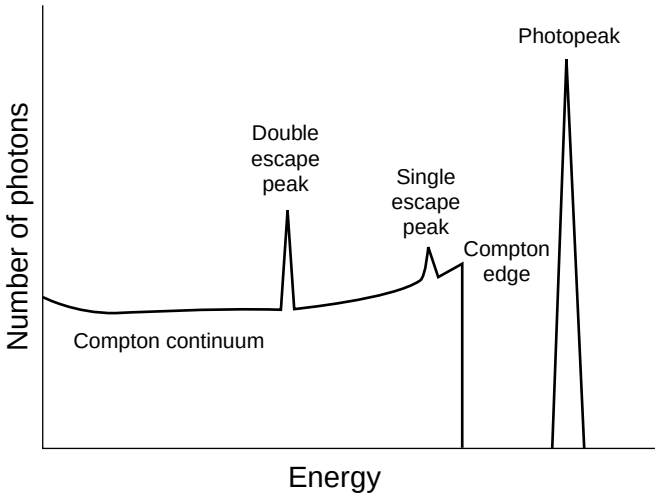

Figure 9 Idealized detector response to monoenergetic γ-photons.

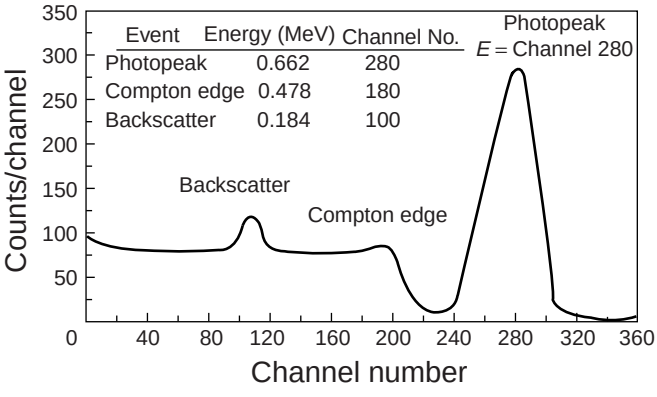

Figure 10 Real spectrum of the monoenergetic γ-photons from ^{137}Cs in a 5 × 5 cm (2 × 2 in) NaI(Tl) scintillator.

annihilation photons escape totally, the total deposited energy is 1.022 MeV less than the original photon energy. Therefore, in the measured spectrum, in addition to the full-energy peak, there are two additional peaks, a single-escape peak at energy $(E_\gamma - 0.511)$ MeV and a double-escape peak at energy $(E_\gamma - 2 \times 0.511)$ MeV (Figure 9). The single- and double-escape peaks appear only if the incident photon energy is above 1.022 MeV.

The detector response presented in Figure 9 is idealized, because it does not take into account broadening of all peaks by the final resolution of the detector and the effect of multiple Compton scattering.

Figure 10 represents a real spectrum of monoenergetic γ-photons from a ^{137}Cs source taken by a 5 × 5 cm (2 × 2 in) NaI(Tl) scintillator. The photopeak at 0.662 MeV is clearly separated from the Compton edge (0.478 MeV). Multiple Compton scattering fills the gap between the Compton edge and the photopeak, forming continuity in the spectrum from zero to the photopeak energy. The final resolution of the detector broadens the photopeak and rounds off the Compton edge and backscatter peak at 0.184 MeV. Backscatter occurs when photons from the source produce Compton interactions in the materials surrounding the detector, e.g. Pb shielding. Backscattered photons are then detected through photoelectric absorption when they enter the detector. The energy of the backscatter peak is $E_0/(1 + 2E_0/m_0c^2)$.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
ANS	American Nuclear Society
BGO	Bismuth Germanate (Bi_4GeO_{12})
DNAA	Delayed Neutron Activation Analysis
EPA	Environmental Protection Agency
fw.1m	full width at one-tenth maximum
fwhm	full width at half-maximum

HPGe	High-purity Germanium
IEEE	Institute of Electrical and Electronic Engineers
MBSS	Marinelli Beaker Standard Source
MCA	Multichannel Analyzer
MDA	Minimum Detectable Activity
NAA	Neutron Activation Analysis
OFHC	Oxygen-free High Conductivity
PIXE	Particle-induced X-ray Emission
PMT	Photomultiplier Tube
PNAA	Prompt Neutron Activation Analysis
PSD	Pulse-shape Discrimination
SCA	Single-channel Analyzer
USDOE	United States Department of Energy
UV	Ultraviolet
XRFA	X-ray Flourescence Analysis

RELATED ARTICLES

Coatings **(Volume 2)**
Atomic Spectroscopy in Coatings Analysis

Environment: Water and Waste **(Volume 3)**
Explosives Analysis in the Environment

Environment: Water and Waste cont'd **(Volume 4)**
Neutron Activation in Environmental Analysis • Proton-induced X-ray Emission in Environmental Analysis • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Field-portable Instrumentation **(Volume 5)**
Radon, Indoor and Remote Measurement of

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Industrial Hygiene **(Volume 6)**
Spectroscopic Techniques in Industrial Hygiene

Peptides and Proteins **(Volume 7)**
X-ray Crystallography of Biological Macromolecules

Polymers and Rubbers **(Volume 9)**
Neutron Scattering in Analysis of Polymers and Rubbers • X-ray Scattering in Analysis of Polymers

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
X-ray Methods for the Study of Electrode Interaction

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Lifetime Measurements, Applications of

Nuclear Methods **(Volume 14)**
Chemical Analysis by Nuclear Methods: Introduction • Charged Particle Activation Analysis • Cyclic Activation Analysis • Elastic Recoil Detection Analysis • Instrumental Neutron Activation Analysis • Instrumental Neutron Activation Analysis: Gamma Lines Table • Nuclear Reaction Analysis • Particle-induced γ-Ray Emission • Photon Activation Analysis • PIXE (Particle-induced X-ray Emission) • Prompt γ-Neutron Activation Analysis • Radiochemical Neutron Activation Analysis • Rutherford Backscattering Spectroscopy • Scattering and Absorption of γ-Rays and Thermalization and Disappearance of Neutrons

Radiochemical Methods **(Volume 14)**
Actinides and other Alpha-emitters, Determination of • β-Particle Emitters, Determination of • γ-Spectrometry, High-resolution, for Radionuclide Determination • Neutron Activation Analysis in the Determination of Very Long-lived Radionuclides

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction • X-ray Photoelectron and Auger Electron Spectroscopy

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview

REFERENCES

1. G.F. Knoll, *Radiation Detection and Measurements*, 2nd edition, John Wiley & Sons, New York, 1989.
2. W.H. Hallenbeck, *Radiation Protection*, Lewis Publishers, Tokyo, 223, 1994.

3. *Canberra Catalog*, 8th edition, Canberra Industries, Meriden, CT, 354, 1989.
4. B.B. Rossi, H.H. Staub, *Ionization Chambers and Counters*, McGraw-Hill, New York, 1949.
5. D.H. Wilkinson, *Ionization Chambers and Counters*, Cambridge University Press, Cambridge, 1950.
6. J.W. Boag, 'Ionization Chambers', in *Radiation Dosimetry*, eds. F.H. Attix, W.C. Roesch, Academic Press, New York, 1–72, Vol II, 1966.
7. J.B. Birks, *The Theory and Practice of Scintillation Counting*, Pergamon Press, Oxford, 1964.
8. *Table of Physical Constants of Scintillators, Scintillator Catalogue*, Nuclear Enterprises, Edinburgh, 50, 1980.
9. *Scintillators for the Physical Sciences*, Nuclear Enterprises, Edinburgh, 1980.
10. W.H. Hallenbeck, *Radiation Protection*, Lewis Publishers, Tokyo, 167, 1994.
11. B.L. Cohen, E.S. Cohen, *Health Phys.*, **45**, 50 (1983).
12. *Instruments and Systems for Nuclear Spectroscopy, Catalog 93/94*, EG&G ORTEC, Oak Ridge, TN, 3–12, 1993–1994.
13. C. Chung, 'Prompt Activation Analysis', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, John Wiley & Sons, Chichester, Chapter 8, 1994.
14. G. Dearnaley, D.C. Northrop, *Semiconductor Counters for Nuclear Radiation*, 2nd edition, John Wiley & Sons, New York, 1966.
15. G. Bertolini, A. Coche (eds.), *Semiconductor Detectors*, Elsevier-North Holland, Amsterdam, 1968.
16. K. Debertin, R.G. Helmer, *Spectrometry with Semiconductor Detectors*, Elsevier Science Publishers, Amsterdam, 1988.
17. *A Handbook of Radioactivity Measurements Procedures*, NCRP Report No. 58, 2nd edition, National Council on Radiation Protection and Measurements, Bethesda, MD, 1985.
18. J. Millman, C.C. Halkias, *Integrated Electronics*, McGraw-Hill Kogakusha, Tokyo, 1972.
19. *Detector and Instruments for Nuclear Spectroscopy, Catalog 91/92*, EG&G ORTEC, Oak Ridge, TN, 2–11, 1991–1992.
20. E.M. Pell, *J. Appl. Phys.*, **31**, 291–302 (1960).
21. *Canberra Catalogue*, 8th edition, Canberra Industries, Meriden, CT, 1989.
22. *Instruments and System for Nuclear Spectroscopy, Catalog 93/94*, EG&G ORTEC, Oak Ridge, TN, 3–2, 1993–1994.
23. *Radioactivity and its Measurement in Foodstuffs*, Food Monitoring Note, EG&G ORTEC, Oak Ridge, TN, 1987.
24. *Instruments for Research and Applied Sciences*, EG&G ORTEC, Oak Ridge, TN, 24, 1980.
25. *Applied Nuclear Spectroscopy*, EG&G ORTEC, Oak Ridge, TN, 62, 1997.
26. R. Keyser, T. Twomey, S. Wagner, *Radioact. Radiochem.*, **1**(2), 47–56 (1990).
27. H.L. Beck, J. De Campo, C.V. Gogolak, *In Situ Ge(Li) and NaI(Tl) γ-ray Spectroscopy*, US Department of Energy, Environmental Measurement Laboratory, New York, HASL-258, 1972.
28. I.K. Helfer, K.M. Miller, *Health Phys.*, **55**, 15–29 (1988).
29. J.F. Cameron (ed.), Technical Report Series, International Atomic Energy Agency, Vienna, (112), 1970.
30. U.A.S. Tapper, W.J. Przybylowcz, H.J. Annegarn, 'Particle-induced X-ray Emission', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, John Wiley & Sons, Chichester, Chapter 13, 1994.
31. *Measurement of Radionuclides in Food and the Environment, a Guidebook*, Technical Reports, Series No. 295, IAEA, Vienna, 1989.
32. P.W. Nicholson, *Nuclear Electronics*, John Wiley & Sons, Chichester, 1974.
33. W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments*, 2nd edition, Springer-Verlag, Berlin, 1994.
34. *Instruments for Research, Catalogue 1002*, ORTEC, Oak Ridge, TN, 1970–1971.

Speciation of Radionuclides in the Environment

Brit Salbu
Agricultural University of Norway, Norway

1 Introduction **1**
2 Theory and Operating Principles **2**
2.1 Source Term Characteristics Including Physicochemical Forms 2
2.2 Dilute Solution Chemistry 3
2.3 Particles, Colloids and Complexes 4
2.4 Association with Solid Phases 5
2.5 Bioavailability 6
3 Sampling and Preanalysis Strategy **6**
4 Fractionation of Radionuclides in Water **7**
4.1 Size-fractionation Techniques 8
4.2 Charge Fractionation Techniques 9
5 Fractionation of Radionuclides in Soils and Sediments **10**
5.1 Surface Analytical Techniques 11
5.2 Size Distribution Pattern 13
5.3 Extraction Techniques – Reversible/Irreversible Sorption Processes 13
6 Dynamic Model Experiments Using Tracers **16**
6.1 Distribution of Species 16
6.2 Kinetics of Transformation Processes 17
7 Speciation Modeling **18**
8 Methods of Analysis **18**
8.1 Low-level Radionuclide Measurements 18
8.2 Mass Spectrometric Methods (Inductively Coupled Plasma Mass Spectrometry, Accelerator Mass Spectrometry) 19
9 Perspective and Future Developments **19**
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **21**
References **21**

To obtain information on physicochemical forms of radionuclides in the environment, speciation techniques should be applied in situ, at-site or in laboratories shortly after sample collection. Among speciation techniques for radionuclides in waters, combined in situ ultrafiltration and exchange chromatography are most promising as radionuclides species are fractionated with respect to nominal molecular mass and change properties simultaneously. However, the fractionation and characterization of chemically well-defined low molecular mass (LMM) species in waters is still a challenge. Surface analytical techniques (e.g. electron microscopy) are useful for characterizing colloids and particles with respect to structure and radionuclide distribution. Recent development within synchrotron-based X-ray microbeam techniques allows detailed information of crystallographic structure (micro-X-ray diffraction (μ-XRD)), oxidation states (micro-X-ray absorption near-edge spectroscopy (μ-XANES)) and volume distribution (μ-XANES tomography) of particle-associated radionuclides on the micrometer scale. Information on radionuclides reversibly or irreversibly associated with solid surfaces (binding mechanisms) is attained by different extraction techniques, while dynamic tracer experiments can provide information on the kinetics involved.

1 INTRODUCTION

To assess short- and long-term consequences of naturally occurring and artificially produced radionuclides released to the environment, information on the distribution of radionuclide species influencing mobility and biological uptake is essential (Figure 1). Radionuclides may be present in different physicochemical forms (i.e. radionuclide species) varying in size (nominal molecular mass), structure and morphology, density, oxidation states and charge properties. LMM species are believed to be mobile and potentially bioavailable, while high molecular mass (HMM) species, such as colloids, pseudocolloids and particles are inert. Thus, development within speciation of radionuclides, i.e. analytical techniques applicable for fractionation, identification and characterization of radionuclide species in waters, soils and sediments, is essential for improving the prediction power of assessment models.

The distribution of species will depend on the source in question, release scenarios, course of events, dispersion processes and physicochemical conditions (e.g. pH, redox, microbial activities) at the deposition site. Following accident releases from nuclear installations, the physicochemical forms of radionuclides will be completely different to those released during normal operations. Hence the distribution of radionuclide species and key properties of specific species, e.g. radioactive particles, may reflect the source of release. After deposition, however, the original distribution of radionuclide species will change with time owing to interactions with natural components (e.g. humic substances, clay minerals) and weathering

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

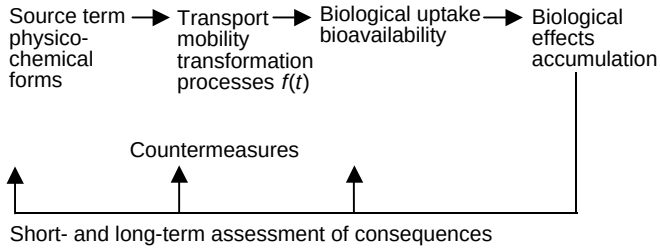

Figure 1 Assessment models depend on source terms (physicochemical forms), mobility and bioavailability of released radionuclides.

processes. LMM species may be irreversibly sorbed on solids, whereas radionuclides associated with radioactive particles can mobilize with time owing to weathering. Therefore, experimental information on transformation processes and kinetics changing the original distribution of mobile and bioavailable species radionuclide species over time is also required if uncertainties associated with transport models, ecosystem transfer models and dose-assessment models are to be reduced.

Similarly to trace element speciation, the analytical strategy includes in situ, at-site or shortly after sampling-fractionation techniques to obtain information on the actual distribution of radionuclide species. For radionuclides present in low concentrations in water, in situ or at-site size-fractionation techniques applicable for large volumes provide information on the distribution of radionuclide species according to size classes, i.e. LMM species, colloids, pseudocolloids and particles. When interfaced with charge fractionation techniques, information on the size distribution of neutral and charged LMM radionuclide species can be obtained simultaneously. For radionuclides in soils and sediments, the mobile fraction includes LMM species in soil waters and also species reversibly sorbed on solid-phases. To distinguish between species reversibly and irreversibly associated in soils or sediments, extraction techniques are useful, whereas for solid radioactive particles sensitive surface analytical techniques can be applied. Information on radionuclide species, interactions and kinetics can also be obtained from dynamic model experiments by adding chemically well-defined radioactive tracer species to soil–water and sediment–water systems, preferably in the field. Although radionuclide speciation is still an analytical challenge within radioecology, major progress has been achieved in recent years, as experience from trace element speciation has been adopted.

2 THEORY AND OPERATING PRINCIPLES

The distribution of radionuclide species in natural waters and in soils and sediments will depend on source term characteristics and trace chemistry phenomena taking place in dilute aquatic solutions, in addition to interactions occurring in heterogeneous soil–water and sediment–water systems at deposition sites.

2.1 Source Term Characteristics Including Physicochemical Forms

The major sources of artificial radionuclides in the environment[1,2] are fallout from 522 atmospheric nuclear weapon tests and authorized or accident releases from the nuclear fuel cycle, especially reprocessing plants (e.g. Sellafield in the UK, La Hague in France and Mayak PA in Russia) and nuclear reactors (e.g. Chernobyl in Ukraine). Releases from satellite, aircraft and submarine accidents and also dumping of radioactive waste at sea have a local impact only.[2] Among artificially produced radionuclides released to the environment, ^{137}Cs, ^{90}Sr and Pu isotopes are of major radiological concern, even though several other long-lived radionuclides (^{99}Tc, ^{129}I, ^{237}Np, ^{3}H, ^{14}C, etc.) are also released from nuclear installations. For dose-assessments, however, the contribution from naturally occurring nuclides can be substantially higher than from artificially produced nuclides,[1–3] in particular from U and Th decay products such as Po and Pb isotopes in alum sheet areas. In groundwaters, natural radionuclide species (LMM species, colloids, particles) released from rocks due to dissolution and desorption processes, erosion and atomic recoil from radioactive decay dominate. In surface waters, the concentrations of cosmogenic radionuclides (e.g. ^{3}H, ^{7}Be, ^{10}Be, ^{14}C), aerosol-associated radon decay products (e.g. ^{210}Po, ^{210}Pb) and artificially produced radionuclides (LMM species, colloids, particles) in fallout or from effluents show spatial and temporal variations, depending on sources and distance from sources, wind direction and water pathways, climate and weather conditions. Sources and processes influencing the speciation of natural occurring radionuclides have been thoroughly discussed previously.[3,4]

Source term characteristics (i.e. activity concentration of individual radionuclides, activity ratios and physicochemical forms of radionuclides) serve as input to dispersion, transport, ecosystem transfer and dose-assessment models. Usually source term information is restricted to inventory estimates (total radioactivity, becquerels), or to activity concentrations of different radionuclides in air (becquerels per cubic meter) or on the ground (becquerels per square meter), while information on the physicochemical forms (e.g. ions, complexes, colloids, particles) is scarce. The distribution of physicochemical forms varying in size (nominal molecular mass distribution), structure and morphology, density, oxidation states and charge properties will depend on the particular source and release scenario, course of events, distance

from the source, dispersion and deposition conditions. During normal operations, effluents from reprocessing plants[5–7] (e.g. Sellafield, UK; La Hague, France; Mayak PA, Russia) and nuclear reactors[8] (e.g. Sweden) contain LMM species, colloids and particles, while volatiles are released to the atmosphere. Condensed particles may form if volatiles are condensed on available aerosol surfaces, during the release and during transport.

At nuclear test sites (e.g. Marshall Islands, Maralinga and Emu sites in Australia; Mururoa, French Polynesia),[9–11] radioactive particles containing fissile material (U and Pu) are frequently observed. Furthermore, particles containing fission and activation products in addition to tranuranic elements are often related to high-temperature accident scenarios, for instance the Windscale pile fire in 1957,[12] the B-52 aircraft crash at Thule in 1968,[13] the satellite Cosmos 954 crash in Canada in 1978[1] and the Chernobyl accident in 1986.[14–17] During an accident, uranium dioxide fuel particles can be oxidized to U_3O_8, being less resistant to weathering than UO_2. Depending on release temperatures, fuel particles may also be depleted[14,15] to variable degrees in volatile radionuclides (e.g. ^{137}Cs and ^{90}Sr). The presence of refractory radionuclides (e.g. ^{95}Zr, ^{144}Ce) indicates release of fuel particles, and the activity ratios may reflect fuel burn-up. However, fuel particles have also been released to the atmosphere at low-temperature (e.g. Windscale piles in the early 1950s).[18,19]

Radionuclides accidentally or intentionally released into the environment are traditionally assumed to be homogeneously distributed as ions, molecules or simple compounds. Thus, assessment of dose rates and corresponding radiological risks from external exposure, inhalation and ingestion are most often derived from average bulk or surface concentrations of radionuclides. As radioactive contaminants may be present as localized heterogeneities (i.e. aggregates or particles, pseudocolloids, colloids), the average bulk concentrations may fail to be representative of the real distribution of the radionuclides.[20] The failure to recognize heterogeneities of radionuclides may lead to analytical inconsistencies, irreproducible results and erratic conclusions, especially due to incomplete dissolution of samples.[9–11,21] Radioanalytical work, carried out in connection with nuclear weapon tests, nuclear accidents and releases from nuclear installation during normal operations, has highlighted the radioanalytical challenges associated in particular with radioactive particles or aggregates and the need to characterize them in terms of their radionuclide composition, structural and morphological properties and physicochemical forms, especially oxidation states of matrix elements. Since a significant part of the activity can be carried by radioactive particles, fractionation techniques should be performed preferentially in the field to differentiate between radionuclide species released from a source.

2.2 Dilute Solution Chemistry

In dilute aquatic systems, trace chemistry phenomena which are negligible at macro-concentrations become significant;[22] the reaction rates may decrease as the probability of effective collisions decreases and reaction mechanisms may change as the chemistry at phase boundaries or the chemistry of colloidal systems becomes predominant. The concentration of most radionuclides in natural waters is very low and the borderlines between categories of radionuclide species ranging from simple ions, molecules or complexes, through hydrolysis products and polymers, colloids and pseudocolloids, to species sorbed or incorporated in suspended inorganic or organic particles and organisms are difficult to establish as there is continual transition between species (Figure 2). As natural waters are dynamic systems, the distribution of species exhibits spatial and temporal variations due to transformation processes. Owing to sorption on available surfaces, complexation with naturally occurring ligands, polymerization and aggregation of colloids, LMM forms are transformed to HMM forms, while desorption, displacement, dissolution and dispersion processes may mobilize LMM forms to the aquatic phase. Thus,

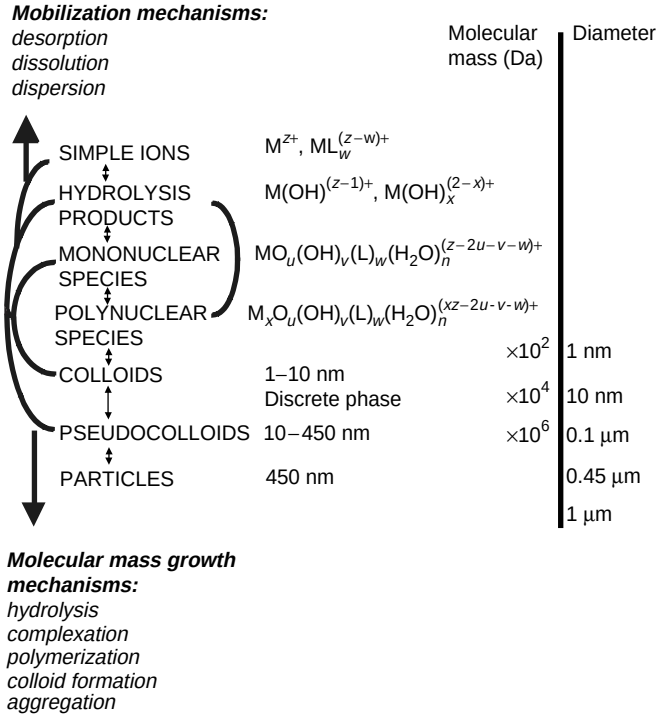

Figure 2 Size classes for the different physicochemical forms of radionuclides in aquatic systems. Transformation processes changing the distribution of species are indicated.

fractionation should take place in situ, at-site or shortly after sampling as storage of water samples prior to fractionation may influence the actual distribution of radionuclide species.

Information on trace chemistry phenomena, including physicochemical forms of radionuclides, interactions and kinetics of importance for mobility and bioavailability, should be based on experimental data from natural systems. The lack of sensitive, species-specific detectors necessitates the use of fractionation techniques similar to those used for trace element speciation purposes. As for trace elements, the distribution of radionuclide species is expected to vary substantially, depending on macrochemical conditions (pH, Eh, ionic strength, temperature) and also on the quality and quantity of interacting components (e.g. humic substances, Fe, Al, Mn).

2.3 Particles, Colloids and Complexes

Owing to weathering, U, Th and decay products such as ^{210}Po and ^{210}Pb may be associated with particulate or colloidal material in aquatic systems.[3] Following the diffusion of radon into the atmosphere, decay products such as ^{210}Po and ^{210}Pb may also associate with aerosol particles and return to the terrestrial environment via wet and dry deposition.[3,4]

Following nuclear weapon tests or nuclear accident releases, a major fraction of radionuclides is associated with particles or aggregates ($\geq$0.45 μm). As observed close to the Chernobyl reactor, more than 90% of radionuclides deposited[14–17] were associated with particles in the micrometers to millimeters range, and deposition densities up to 10^5 particles m^{-2} have been reported.[23] Smaller radioactive particles (micrometer range) were also observed up to 2000 km from the site.[24–26] After deposition, weathering of radioactive particles occurs and associated radionuclides are mobilized with time. Information on the particle composition (e.g. UO_2 fuel), structural changes (e.g. oxidation of UO_2 to U_3O_8 during release or after deposition) and chemical conditions (pH, redox, microbial activity) at the deposition site is essential to assess weathering rates and the subsequent mobilization of associated radionuclides. Based on Chernobyl samples collected over a 10-year period,[27] the particle weathering constant was found to vary from 0.04 to 0.42 yr^{-1}. A low weathering constant was associated with fuel particles released during the initial explosion and deposited to the west of the reactor, whereas a high weathering constant was associated with oxidized U_3O_8 particles released during the fire and deposited to the north of the site. Within the 30-km zone, the weathering rate (becquerels released per unit time) of fuel particles is reported to be slower than the interaction (strong fixation) rate for ^{137}Cs in soils.[28] For ^{90}Sr in soils, however, the rate of interaction (weak association) is significantly lower than the weathering rate. As the potential transfer of released radionuclides within the ecosystem over time will depend on ratios between particle weathering rates and rates of interaction (strong fixation or weak association) of mobilized radionuclides, the mobility and biological uptake of ^{90}Sr are expected to increase with time as weathering proceeds.

Colloidal forms of radionuclides (Figure 2) may be released from a source (e.g. effluents[6]) or radionuclides may form colloids (1–10 nm) in aquatic systems.[22] Furthermore, pseudocolloids are formed by the interaction of reactive radionuclides with surfaces of natural colloidal or pseudocolloidal material (diameter 1–450 nm), being important transporting agents for trace elements and radionuclides in natural water systems, for instance in aerial transport,[29] rain water,[26] soil–water,[30] groundwater,[31,32] runoff and river water,[33] lake water[34] and seawater.[35] Chernobyl-contaminated runoff to a mountain lake was characterized by ^{137}Cs associated with colloidal material, while ^{90}Sr was present in LMM forms.[33] Budget calculations also showed that colloidal ^{137}Cs was retained in the lake, whereas ^{90}Sr as mobile LMM species was transported through the lake and downstream.

The presence of colloidal material is favored under low-temperature conditions, and the surfaces of colloids in natural waters are usually negatively charged, partly due to organic coatings. Changes in chemical conditions, for instance, in mixing zones (fresh waters, estuaries) where waters of different qualities mix may significantly influence the stability of colloids (relationship between attractive van der Waals forces and repulsive Coulomb forces) and aggregation and particle growth occur until sedimentation takes place. Furthermore, polymerization of scavenging elements such as aluminum may occur in mixing zones.[36] Hence trace amounts of charged radionuclide species present in unstable mixing zones such as estuaries may be affected by processes controlled by the macrochemistry of stable elements.

The use of organic ligands (e.g. ethylenediaminetetraacetate, diethylenetriaminepentaacetate, nitrilotriacetate, citrate) for separating radionuclide complexes is well known within radiochemistry. Thus, tables with laboratory-based stability constants for a large variety of radionuclide organic complexes have been developed over decades, often based on radioactive tracers and synthetic organic standards. The interaction with nonpolymeric and noncolloidal organic ligands may take place via carboxylic group interactions (organic salt), covalent bonds with carbon atoms or via electron-donating atoms (O, N, S, etc. coordination complexes) or by overlap of π-electron orbitals (π-complexes). Several mechanisms may be involved simultaneously. Radionuclides

can be released to the environment as stable complexes or chelates[5] or especially for multivalent radionuclides (e.g. actinides such as Th, U, Np, Pu, Am) may form complexes owing to interactions with naturally occurring organic ligands.[37,38] The mobility is usually increased as radionuclides reversibly associated with surfaces of solids may desorb upon the complex formation, and the sorption of organic complexes to negatively charged surfaces is usually low (low distribution coefficient, K_d). Forming lipophilic organic complexes, the transfer through biological membranes is also usually enhanced (i.e. increased bioavailability). Especially in groundwaters, inorganic complexes, e.g. carbonato or mixed hydroxo–carbonato complexes, play an important role in transport of radionuclides such as ^{210}Pb and actinides.[3,39]

Inorganic ligands forming stable complexes or chelates are also frequently applied for the separation of specific radionuclides. A well-known ligand is hexacyanoferrate, used for fractionation of positively charged Cs isotopes from large volumes of waters or used as a countermeasure when administered to animals to reduce the activity concentration of Cs isotopes in animal products.[40] Although information on complex or chelate formation of radionuclides with naturally occurring ligands, especially organic substances such as humic or fulvic substances, has been improved during recent years, the development of fractionation techniques differentiating between radionuclides associated with organic complexing ligands, organic polymers and organic colloids still represents an analytical challenge.

2.4 Association with Solid Phases

Processes occurring in soil–water or sediment–water systems influencing the mobility of trace metals and radionuclides and the transfer to vegetation via root uptake have been intensively studied, especially in cultivated areas, over decades. Several concepts and models have also been developed during recent years.[41–43] Except for sieving techniques and grain-size analysis, however, speciation studies in heterogeneous soil–water and sediment–water systems are still suffering from a lack of well-defined methods.

The term ''mobile species'' in soils and sediments refers to LMM species, e.g. ions or complexes soluble in soil water ($\leq$0.45 μm) and species reversibly associated with solid surfaces (Figure 3). Thus, a fraction of mobile species may be available for active biological uptake. The term ''inert species'' includes radioactive colloids, pseudocolloids and particles deposited in soils or sediments, as well as radionuclides irreversibly bound to or incorporated into mineral lattices. The systems are dynamic; equilibrium may establish relatively rapidly between LMM forms in soil solutions and species

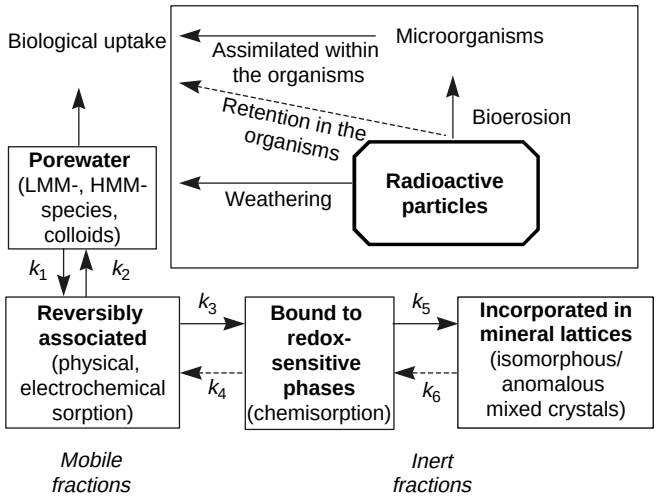

Figure 3 Distribution of radionuclide species in soil–water (sediment–water) systems. Weathering of deposited radioactive particles is included.

reversibly associated with surfaces of solids, but slowly for species penetrating diffuse double layers of colloids and particles into mineral lattices.[44] Thus, equilibrium within systems may not have been established when samples are collected.

The transfer of radionuclides in soil–water or sediment–water systems is frequently modeled using distribution coefficients, K_d (Bq kg^{-1} soil or sediment per Bq L^{-1} or Bq m^{-3} water). Hence, the models refer to equilibrated systems. Information on K_ds is usually based on measurements of total concentrations of radionuclides in contaminated soils or sediments and in waters (e.g. pore water, lake or seawater). Large variations are seen, however, for reported K_d values in the literature,[45] partly owing to the fact that equilibrium has not been established. As the system is dynamic, the distribution of radionuclides between solid and solutions should be described as a time-dependent function. Time-dependent functions are also relevant if radioactive particles are present in soils or sediments (Figure 3). In this case, the apparent K_d will be very high, indicating that soils or sediment will act as an efficient sink. As weathering occurs and radionuclides such as ^{90}Sr are mobilized over time, however, the distribution of nuclides between solid and solution is reduced over time and the observed K_d cannot be considered to be constant.

To improve the resolution in K_d, water can be fractionated with respect to LMM species (section 4), and extraction techniques can be utilized to distinguish between reversibly and irreversibly associated radionuclides in soils and sediments (section 5). Thus, observed K_d can be differentiated between $K_{d,\ reversible}$ and $K_{d,\ irreversible}$ for modeling purposes. Alternatively, laboratory experiments where soil or sediment samples are incubated for

a certain time period with synthetic solutions to which a chemically well-defined tracer has been added have been applied.[41] During recent years, incubation techniques have been further developed to improve prediction models based on K_d values (section 6).

A major analytical problem is the lack of appropriate in situ fractionation techniques to study interactions of radionuclides in soil–water or sediment–water systems. However, techniques can be applied at-site or in the laboratory shortly after sampling, provided that the systems are not manipulated prior to fractionation (e.g. dried or homogenized samples).

2.5 Bioavailability

Within radioecology, the transfer from soils, sediments or water to vegetation, animals or fish is usually described by a transfer coefficient (TC) ($Bq\,kg^{-1}$ vegetation per $Bq\,m^{-2}$ soil), transfer factor (TF) ($Bq\,kg^{-1}$ vegetation per $Bq\,kg^{-1}$ soil) or concentration factor (CF) ($Bq\,kg^{-1}$ animal per $Bq\,kg^{-1}$ fodder or $Bq\,kg^{-1}$ fish per $Bq\,m^{-3}$ water). The coefficients or factors are based on total activity concentrations[44,45] owing to the lack of information on bioavailable radionuclide species present. However, TC, TF and CF show large seasonal variations and the analytical data reported in the literature are scattered. Thus, uncertainties associated with these variables contribute significantly to the overall uncertainties in predicting consequence assessments. The relationship between trace metals or radionuclide species and biological uptake in aquatic organisms has been thoroughly discussed by Simkiss and Taylor[46] and Fisher and Reinfelder.[47] LMM forms, often referred to as ''free ions'' or ''labile'' species in the literature, such as cations, hydrated ions, charged or neutral complexes and organometallic compounds including lipophilic complexes, are potentially able to penetrate biological membranes. Most data on the bioavailability of radionuclides are, however, obtained from model biotest experiments (section 6).

Radionuclides associated with colloids and particles are considered inert and the active biological uptake in plants and animals should be low. Therefore, in areas heavily contaminated with radioactive particles, as seen close to the accidental site in Chernobyl, TC, TF and CF reflecting the transfer to biological systems (e.g. vegetation, animals, fish) should be low. Field experiments performed within the Chernobyl 30-km zone demonstrated also that the ''relative biological availability'' of ^{137}Cs from soil to grass was low compared with that of an added ionic Cs tracer.[48] Furthermore, several studies have confirmed that particle weathering and the subsequent increases in ^{90}Sr mobility are accompanied by increases in soil-to-plant TFs in the 30-km zone,[27] and that the transfer increases with distance from the source. As an alternative to TC or TF, a mobility factor (MF)[49] can be calculated according to Equation (1):

$$\mathrm{MF} = \frac{\text{mobile species } (\mathrm{Bq\,m^{-2}}) \times 100(\%)}{\text{total deposition } (\mathrm{Bq\,m^{-2}})} \qquad (1)$$

where mobile species include the species reversibly associated with solids in soil (or sediment) and species taken up by vegetation from the same site (becquerels per square meter). Thus, the pool of mobile species in soils is assumed to act as a reservoir for root uptake. By including mobile species taken up by vegetation, MF is less sensitive to seasonal variations due to vegetation growth. For particle-contaminated soils (Chernobyl), the mobile fraction, and hence, MF, are small. For samples collected in Ukraine, Belarus and Norway, the calculated MF also demonstrated low mobility of ^{90}Sr close to the reactor, whilst MF increased with increasing distance from the reactor.[49,50]

Following the Chernobyl accident, numerous animal feeding experiments with contaminated fodder have been performed[40] and the quality of data with respect to uptake, accumulation, distribution of radionuclides in biological compartments, excretion and CF values has been significantly improved. In most models, the animal digestion of fodder contaminated with radioactive particles is not taken into account, i.e. CF is expected to be low. However, fuel particles retained as a point source in the gastrointestinal tract may be trapped much longer than previously anticipated and uptake may occur from digestion.[51] In areas contaminated with radioactive particles, the influence on radionuclide speciation from microorganisms due to bioerosion,[52] especially from certain fungi in soils and from benthic organisms in sediments, may be significant. Furthermore, retention of colloids or particles in filtering organisms may represent an important pathway. Therefore, time-dependent variations in observed TC, TF and CF should be expected if radioactive particles are present.

3 SAMPLING AND PREANALYSIS STRATEGY

The analytical results and interpretation of data are closely related to the analytical strategy, sampling, fractionation and measurement techniques chosen, and will depend on

- representative sampling and fractionation of samples;
- in situ, at-site or shortly after sampling;
- preconcentration techniques and chemical yields;
- method of analysis, including determination limits, precision and accuracy.

General aspects associated with sampling, speciation and analytical methods have been discussed in several reviews and textbooks.[53–56] This article will therefore focus on recent developments within the field. To obtain information of radionuclide species in real systems, samples should be fractionated in situ or at-site and storage of samples or manipulations (e.g. preconcentrated water or dried soils) prior to fractionation should be avoided. In addition to characterization of species, information on transformation processes changing the distribution of species over time and kinetics is needed for assessment purposes. Hence, time series should be considered essential. There is, however, still a great need for standardization of sampling and fractionation techniques as quality assurance and accreditation require proper documentation. Furthermore, standard reference material and protocols of relevance for speciation purposes should be developed.

4 FRACTIONATION OF RADIONUCLIDES IN WATER

There are several requirements[56] which should be met by fractionation techniques applicable for speciation purposes:

- fractionation should take place in situ, at-site or shortly after sampling to avoid storage effects;
- rapid fractionation techniques are needed to avoid establishment of equilibria between species retained and species to be separated during the fractionation;
- sorption effects should be minimized;
- clogging effects should be avoided;
- colloids and aggregates in solution should be stable;
- chemical reagents influencing the distribution of species should be avoided.

Owing to the low concentrations of radionuclides in most waters, techniques yielding large fractionated volumes being concentrated prior to analysis are favorable as determination limits are lowered. Size-fractionation techniques allow discrete phases to be retained, leaving mobile and potentially bioavailable LMM species relatively undisturbed. In addition to the analysis of LMM species, information on the size distribution of colloids or particles, structure and morphology, surface distribution of nuclides and oxidation states should also be obtained by further analysis of the retained fraction using advanced solid surface speciation techniques (e.g. electron microscopy with X-ray microanalysis (XRMA), synchrotron-based X-ray microscopic techniques).

A variety of size and charge fractionation techniques have been utilized for speciation studies in the laboratory,[53–56] while techniques applicable in situ or at-site are more limited (Table 1). Procedures for sampling and characterization of colloidal material have, however, been reviewed.[57]

Table 1 In situ/at-site and in-laboratory (small volumes) techniques for speciation of radionuclides

Size-fractionation	Charge fractionation	Solid state speciation
Filtration[a]	Exchange chromatography (cation, anion, adsorption)[a]	Electronmicroscopy techniques (SEM, TEM)
Tangential flow/hollow-fiber ultrafiltration[a]	Liquid–liquid extraction	X-ray-induced spectroscopy (μ-XRD, μ-XAS, μ-XANES, EXAFS, μ-tomography)
Continuous flow centrifugation[a]	Sequential extractions	
In situ dialysis (small volumes)[a]	Electrochemical methods	Laser-induced spectroscopy (LIPAS, LITLS, LAMMA)
Ultracentrifugation	Crown ether chromatography	Mass spectrometry (SIMS, ICPMS, AMS)
Density centrifugation		Electron energy loss spectroscopy
Dialysis		Raman spectroscopy
Gel chromatography		Nuclear magnetic resonance spectroscopy
Combined techniques:		
Filtration/Ion-exchange chromatography[a]		
Tangential flow ultrafiltration/Exchange chromatography[a]		
In situ dialysis with exchange resins (small volumes)[a]		
Electrodialysis (small volumes)		
Chromatographic methods (HPLC, LC) and mass spectrometry (LC/MS/MS, HPLC/ICPMS)		

[a] Applied in situ and/or at site.
AMS, accelerator mass spectrometry; EXAFS, extended X-ray absorption fine structure analysis; HPLC/ICPMS, high-performance liquid chromatography/inductively coupled plasma mass spectrometry; HPLC, high-performance liquid chromatography; ICPMS, inductively coupled plasma mass spectrometry; LIPAS, laser-induced photoacoustic spectrometry; LC/MS/MS, liquid chromatography/tandem mass spectrometry; LC, liquid chromatography; SEM, scanning electron microscopy; SIMS, secondary ionization mass spectrometry; TEM, transmission electron microscopy; μ-XAS, micro-X-ray absorption spectroscopy.

4.1 Size-fractionation Techniques

Among size-fractionation techniques, filtration, tangential flow ultrafiltration and partially continuous-flow centrifugation meet the majority of requirements for in situ or at-site radionuclide speciation studies in natural waters (Table 1). In situ dialysis for collecting LMM species in waters can only be applied in heavily contaminated areas owing to the low volumes applied. As LMM species are essential for assessing mobility and bioavailability of radionuclides, further fractionation according to charge properties should be applied (section 4.2).

4.1.1 Filtration

Filtration of large water volumes through membranes of various types and pore diameters is most commonly used for the separation of particles larger than 450 nm in natural waters. By interfacing the filtration equipment to a peristaltic pump sampling system, in situ filtration of large volumes is attained.[32,58] Radionuclides associated with retained particulate material on filters can be measured directly (γ-emitters) or after radiochemical separations (β-, α-emitters), while nuclides in filtrates can be analyzed after preconcentration. As conventional suction filtration using a vacuum pump is often employed, the pressure should be kept relatively low (20 mmHg) to avoid rupture of algal or bacterial cells.

Experience over decades within trace element speciation has demonstrated methodological effects[56,57] within conventional filtration, such as clogging of membranes, concentration polarization, sorption of species and salt retention. When clogging occurs due to blocking of pores by large particles or bridging of smaller particles retained within the pores, the effective diameter of the filter pores and subsequently the filtration rate will decrease. Hence, species in the filtrate are no longer defined according to the pore diameter. Introducing stirred cells, clogging and concentration polarization are minimized, but mechanical disruption or dispersion of aggregates (e.g. humic substances) may occur. Cellulose ester filters (e.g. Millipore, 0.45 μm) act as depth filters, and clogging is of less importance than for polycarbonate screen filters (Nuclepore, 0.4 μm). Glass-fiber membranes, being poorly defined with respect to pore diameters, should be avoided in radionuclide speciation studies. In low-turbidity water, the most serious source of error is adsorption of species to equipment surfaces. As adsorption loss depends on physicochemical forms of the radionuclide and surface charge of equipment walls (dependent on material, cleaning procedures), the loss may be difficult to account for. Conditioning the equipment with a sample aliquot will, however, reduce sorption losses.

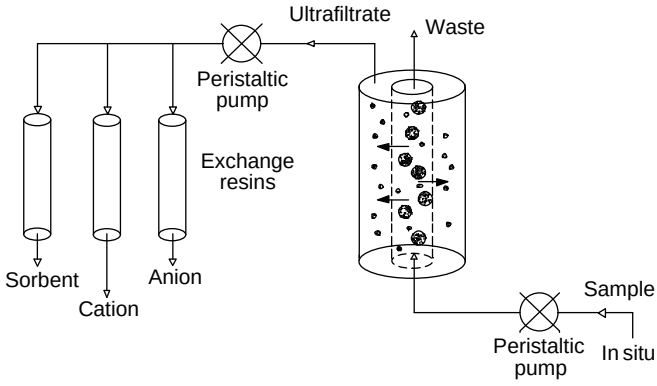

Figure 4 In situ hollow-fiber ultrafiltration combined with on-line chromatography for fractionation of LMM neutral and positively and negatively charged radionuclide species.

4.1.2 Tangential Flow Ultrafiltration

Using a peristaltic pump connected with hollow-fiber[33] or tangential cross-flow[59] systems, large water volumes are transferred directly from the sampling depth into the molecular mass discriminator and ultrafiltrates are collected (Figure 4). Consisting of up to 1000 tubes, the effective surface areas of hollow-fiber cartridges are relatively large. Thus, volumes from about 5 L up to hundreds of liters can be processed with fibers, while even larger volumes can be ultrafiltered using tangential cross-flow systems. Cartridges and tangential cross-flow membranes with nominal molecular weight cut-off values in the range 10^3–10^5 Da are also available. The pore-size distribution of hollow-fiber membranes is relatively narrow, clogging is negligible owing to the high internal flow and sorption is insignificant when a sample aliquot is used for conditioning purposes.[60] The cross-flow cassette membranes are, however, less well defined and memory effects reflecting internal sorption may influence the results. A mass balance approach can be utilized to account for methodological effects. In situ systems can be utilized for the determination of radionuclides associated with colloidal material in open surface waters such as river and lake waters[33] and seawaters[58] [e.g. Irish Sea,[35] Kara Sea[61]]. For waters with limited volumes such as soil waters,[30] for effluents from reactors[8] and reprocessing plants[6] (Sellafield, UK; La Hague, France) or for groundwaters, samples are usually collected a short time prior to fractionation.

Other ultrafiltration techniques (e.g. stirred cell, filtration during ultracentrifugation) are frequently applied in laboratories processing highly contaminated solutions [e.g. effluents[37]] or model experiments using radioactive tracers (section 6). Introducing a magnetic stirrer in conventional ultrafiltration, the concentration polarization is favorably suppressed and the membrane clogging reduced, while mechanical disruption of aggregates may influence the results.

4.1.3 Continuous-flow Centrifugation

Colloidal-sized particles sediment during centrifugation by artificially increasing the force of gravity to counteract the viscosity, mutual repulsion, Brownian movement and the force of shear.[53–56] As the particle separation is based on differences in sedimentation coefficients, K_S, small, dense particles are separated together with larger, less dense particles. For particles with similar K_S, special consideration must be given to the particle shape (e.g. plate-shaped clay minerals versus rod-shaped viruses), particle charge (e.g. organic coatings), the degree of external or internal water (macroporosity) and the viscosity, which is reduced if the temperature during centrifugation increases. For particles differing widely in K_S, the most serious source of error is the tendency of small, e.g. clay, particles to cohere and form coarser particles. Furthermore, sorption during centrifugation due to transport of species to tube walls is difficult to quantify. Despite these effects, continuous-flow systems[62] have previously proved to be a useful system for collecting colloids or particles from large volumes of low-turbidity waters. Hence associated radionuclides can be determined from the retained fraction and further solid surface analysis can be performed.

High-speed centrifugation (about 10 000–20 000 rpm) and ultracentrifugation (30 000–60 000 rpm) are frequently applied in laboratory when small volumes are processed. Furthermore, sucrose density centrifugation (high-speed or ultracentrifugation) for the separation of organic and inorganic colloids, in particular microorganisms from clay minerals in natural waters, should be considered potentially interesting for the speciation of radionuclides in waters within contaminated areas.

4.1.4 In Situ Dialysis

Close to release sites where the contamination of radionuclides is sufficiently high (small volumes to be analyzed), in situ dialysis techniques should be useful for the collection of LMM radionuclide species. The technique has previously proved very valuable for the determination of LMM trace-element species in natural waters.[56] By inserting dialysis bags (pore-size 1–10 kDa) containing up to 100 mL of deionized water in rivers or lakes under stable conditions, equilibrium between LMM species on both sides of the membrane is assumed to be established within 5–7 days. Then, dialysis bags are collected and LMM species in the dialyzate analyzed. In slow-running waters (e.g. groundwater), however, clogging of membranes due to, e.g. algal growth, may seriously influence the results. So far no published data on the in situ dialysis of radionuclides in natural waters seem to be available. However, other dialysis techniques[22] have occasionally been applied for fractionation of radionuclide tracers in laboratory studies (section 6).

4.2 Charge Fractionation Techniques

Charge fractionation techniques are well known within radiochemistry[22] (e.g. exchange chromatography, liquid–liquid extraction, electrochemical techniques) and are useful for the determination of charged radionuclide species in solution. The techniques are frequently applied in laboratory experiments for concentrating radioactive tracers from solutions. However, for natural water systems large volumes are usually needed owing to low concentrations of radionuclides. In situ or rather at-site fractionation of charged radionuclides from large volumes of waters can be obtained by applying chromatographic techniques[63] coupled on-line to a pump system (Table 1). These systems are often used for wastewater cleaning purposes.

4.2.1 Exchange Chromatography

Using cat(an)ion-exchange or chelating resins, simple cat(an)ions initially present in solution, or dissociated from weak complexes or desorbed from solid surfaces, react with the chromatographic reagent and are retained on the resin (electrostatic sorption or chemisorption), while oppositely charged species, neutral undissociated molecules and colloidal material are excluded and eluted.[63] Information on species retained can be obtained from direct measurement of resins (e.g. retention of γ-emitters) or indirectly by the difference[64] between the total concentration in the sample and the concentration in the eluate (Equation 2):

$$C_{\text{retained}} = C_{\text{total}} - C_{\text{eluted}} \quad (2)$$

The fraction of radionuclides retained by resins depends on stability constants and the equilibration time. As the interactions of charged species with resins are time-dependent processes, analytical results depend upon the flow rate of solution through the column. Being essentially a slow fractionation procedure, a shift of chemical equilibria during separation may complicate the interpretation of results. As species differing in size are also fractionated according to charge, the result for unfiltered waters will depend on the experimental conditions (e.g. resins, mesh, flow rates) and the results may be difficult to interpret. Resins may act as a filter and if colloids or particles are retained within the resin, the fraction of charged species estimated from eluates may be overestimated. By removal of particles and colloids prior to exchange chromatography, this effect is avoided.

Cation-exchange resins (e.g. Dowex™, Chelex™, Amberlite™) are frequently applied for retention of positively

charged metal or radionuclide species in fresh waters[36,64] and seawaters.[61,62] Anion-exchange resins (e.g. Dowex-1™) are often used for the retention of negatively charged radionuclide complexes of, for instance, Pu and other actinides in groundwater.[32,38] Chelating resins being species specific, e.g. hexacyanoferrate for chelating Cs^+ ions, are most efficient as interferences from other Cs species or other radionuclide cations can be avoided. For neutral radionuclide species, exchange columns or filter cartridges with sorbents (e.g. active carbon) are frequently used, while XAD™ resins often applied for retention of organometallic species and humic substances should also be considered applicable for radionuclides. For colloids, in particular Pu colloids in seawater, sorption to Al_2O_3 is claimed by Mitchell et al.[35] to be specific. However, retention of other Pu species cannot be excluded.

To account for methodological effects, a mass balance approach can be utilized. It is also essential to underline the fact that chemical yields obtained for well-defined tracers for optimizing the technique may not be relevant as species in real systems may differ significantly from tracer species added.

4.2.2 Redox Techniques

Already in 1978 differentiation with respect to oxidation state of plutonium species in the Irish Sea was performed using different redox agents prior to the coprecipitation of the reduced forms (Pu^{3+}, Pu^{4+}) on rare earth fluorides.[65] Recent data from the Irish Sea and the English Channel illustrate also that a significant fraction of Pu isotopes is present as colloidal material.[66] Similar principles have been used to distinguish between Co^{II} and Co^{III} forms in nuclear discharges to seawater.[5] When strong redox agents are added to samples prior to fractionation, however, transformation of species occurs, e.g. species are desorbed from particle surfaces and colloids may dissolve. Ultraviolet (UV) irradiation of acid oxidizing media (e.g. H_2O_2) has also been recommended for destroying organic material and release associated metals. However, UV irradiation may also influence the size distribution pattern of clay minerals. Therefore, redox agents can be utilized for the determination of total activity concentrations, but should be avoided for speciation purposes.

4.2.3 Combined Fractionation Techniques

Speciation schemes are often based on size (e.g. 0.1–1-μm filters) and charge (e.g. chelating, cation-exchange and anion-exchange resins or sorbents) fractionation techniques performed sequentially. Combined techniques, where species are separated according to size in the colloidal range and charge simultaneously, represent a significant improvement within speciation studies (Table 1). When ultrafiltration is interfaced (on-line) with exchange chromatography (e.g. cation, anion, active C sorbent), radionuclides associated with charged and neutral colloidal species (Figure 4) can be differentiated simultaneously. As the sum of fractions should equal the total concentrations, methodological effects can be accounted for. Although filters implanted with exchange resins have been used for trace metals, clogging of filters may easily affect the results.

The use of dialysis bags (e.g. 100 mL) containing chelating agents is a promising in situ technique for radionuclides in contaminated aquatic systems, as species with molecular weights lower than the membrane pores are separated according to charge.[67] Thus, the fraction retained in the resin represents the LMM cationic (anionic) forms present in waters, and interferences from charged HMM forms are avoided. Careful standardization is needed and membrane effects must be controlled. In running water, however, clogging of dialysis membranes should be minimized. So far, no experiments with dialysis bags containing resins have been reported for radionuclides.

In laboratory, several electrochemical techniques (electrophoresis, electrodialysis) have previously been utilized for determining charged species of radionuclides.[22] Today, the combination of advanced separation techniques applicable for identification of radionuclides associated with for instance volatile and non-volatile organic components [e.g. HPLC/ICPMS and LC/MS/MS] should also be encouraged.

5 FRACTIONATION OF RADIONUCLIDES IN SOILS AND SEDIMENTS

Colloids and particles act as transport agents for radionuclides in water systems and soil and sediment act as sinks for many nuclides released from a source. An inhomogeneous distribution of radionuclides (localized heterogeneities) in soils and sediments also indicates deposition of radioactive aggregates and particles. This can most often be confirmed using autoradiography, followed by advanced surface sensitive techniques. Owing to changes in physical and chemical conditions (i.e. pH, Eh) radionuclides can be remobilized with time and contaminated soils and sediments may act as a diffuse source in the future. Thus, information on the distribution of radionuclide species in pore water, distribution of nuclides between solid and solution (K_d) and binding mechanisms is essential for long-term consequence assessments. Determination of the distribution coefficients, K_d, from total activity concentrations gives no information on processes or mechanisms. Based on extraction techniques or model experiments using well-defined tracers,

however, time-dependent distribution functions can be utilized in process-oriented models.

5.1 Surface Analytical Techniques

To obtain information on colloid or particle characteristics, several potential surface analytical techniques are available in laboratories (Table 1). However, the need for high sensitivity in the microscale range limits the application of several methods. Autoradiography, scanning and TEM and laser spectroscopy are most useful techniques. In addition, the recent developments within synchrotron-based X-ray microscopic techniques prove very promising.

5.1.1 Autoradiography

Heterogeneities containing localized radioactivity (i.e. hot spots) can be observed within contaminated areas using portable Geiger–Müller tubes or NaI detectors.[19,50] Then, qualitative autoradiography can be performed by spreading dried samples (soil, sediment, vegetation) on paper or plastic foils which are placed in close contact with an X-ray- or α-radiation-sensitive film for a certain time of exposure.[9,19,27] For activity levels in the range ≤ 1–10 Bq, the exposure time may reach several months. Using well-defined point sources, quantitative autoradiography can be obtained.[9,16]

5.1.2 Scanning Electron Microscopy with X-ray Microanalysis

Structure information of particles identified, for instance, from autoradiography, can be obtained using SEM interfaced with XRMA, as illustrated for a Chernobyl particle in Figure 5(a–f). Prior to analysis, particles mounted on stubs are usually carbon coated to avoid artefacts due to charging. Using the BEI mode, bright areas reflect particle surfaces containing high-atomic-number elements, while the elemental distribution is attained using X-ray mapping. XRMA also provides information

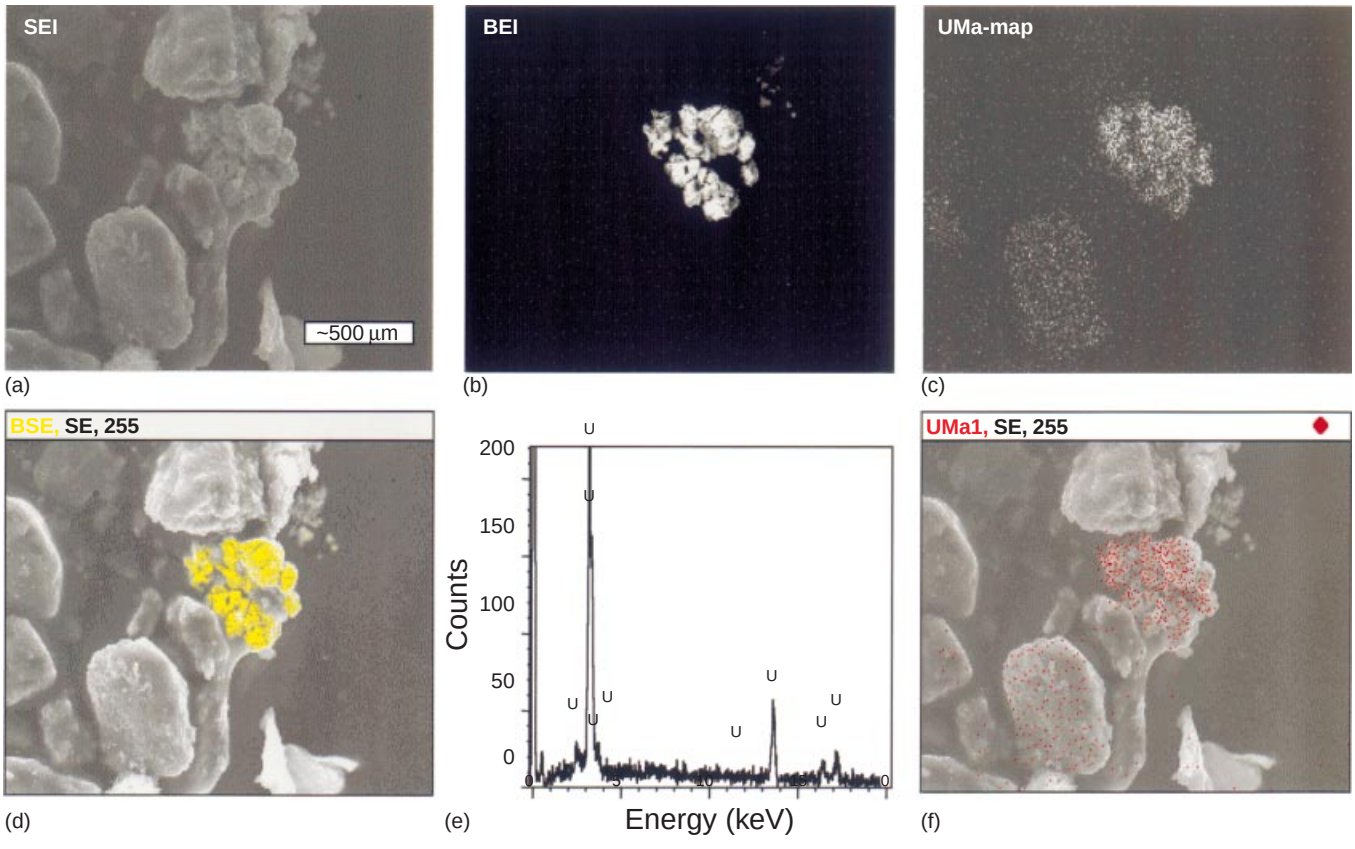

Figure 5 Surface analysis of fuel particles collected within the 30 km zone of Chernobyl using different options in SEM with an X-ray microanalyzer: (a) secondary electron imaging (SEI) mode (bar = 500 µm); (b) backscattered electron imaging (BEI) mode, bright areas reflect the distribution of high atomic number elements; (c) X-ray mapping of the distribution of uranium; (d) superimposed images by combining methods, SEI and BEI modes; (e) elemental analysis by XRMA confirms the presence of U; (f) superimposed images by combining methods, SEI mode and X-ray U mapping.[68]

on the elemental composition at specific particle sites. These techniques have been utilized to characterize radioactive particles from different sources, e.g. flake-like uranium fuel particles released under low-temperature conditions from Windscale, UK, and large uranium fuel aggregates and also crystalline and amorphous single fuel particles released at high-temperature from the Chernobyl reactor.[19] Although all these are U-fuel particles, their behavior in the environment is expected to be different; the transport of spheres is different from that of flakes and the weathering rates of amorphous structures are significantly higher than those of crystalline phases. Radioactive particles, different from those associated with accidental scenarios, have also been identified in contaminated sediments,[61,68] e.g. in the Irish Sea, due to effluents from the Sellafield nuclear installation and in the Stepovogo Fjord at Novaya Zemlya due to dumped radioactive waste.

5.1.3 Transmission Electron Microscopy with X-ray Microanalysis

For colloidal radioactive material, TEM interfaced with XRMA can be utilized for structure and elemental analysis. When droplets of water are transferred to coated grids and carefully dried under a UV lamp prior to analysis, electron-dense structures can be recognized, when compared with a blank (distilled water). Thus, electron-dense structures reflect the presence of colloidal material in effluents[6] from Sellafield and La Hague reprocessing plants (Figure 6a and b). XRMA provides information on the element composition while the distribution of element is obtained from X-ray mapping.

5.1.4 Laser-induced Spectroscopic Methods

Surface analysis of radioactive species can also be performed using laser-induced spectroscopic methods.[4] Owing to the high sensitivity, the most promising technique so far is LIPAS.[69] Using well-tuned laser beams, absorbed photons excite orbital electrons. By decay, the light-absorbing species are heated, and temperature changes can be detected by piezoelectric crystals. Although these techniques often have been used for the determination of low-level concentrations of radionuclides, in particular transuranic elements such as Pu, the technique should be further improved for speciation purposes.

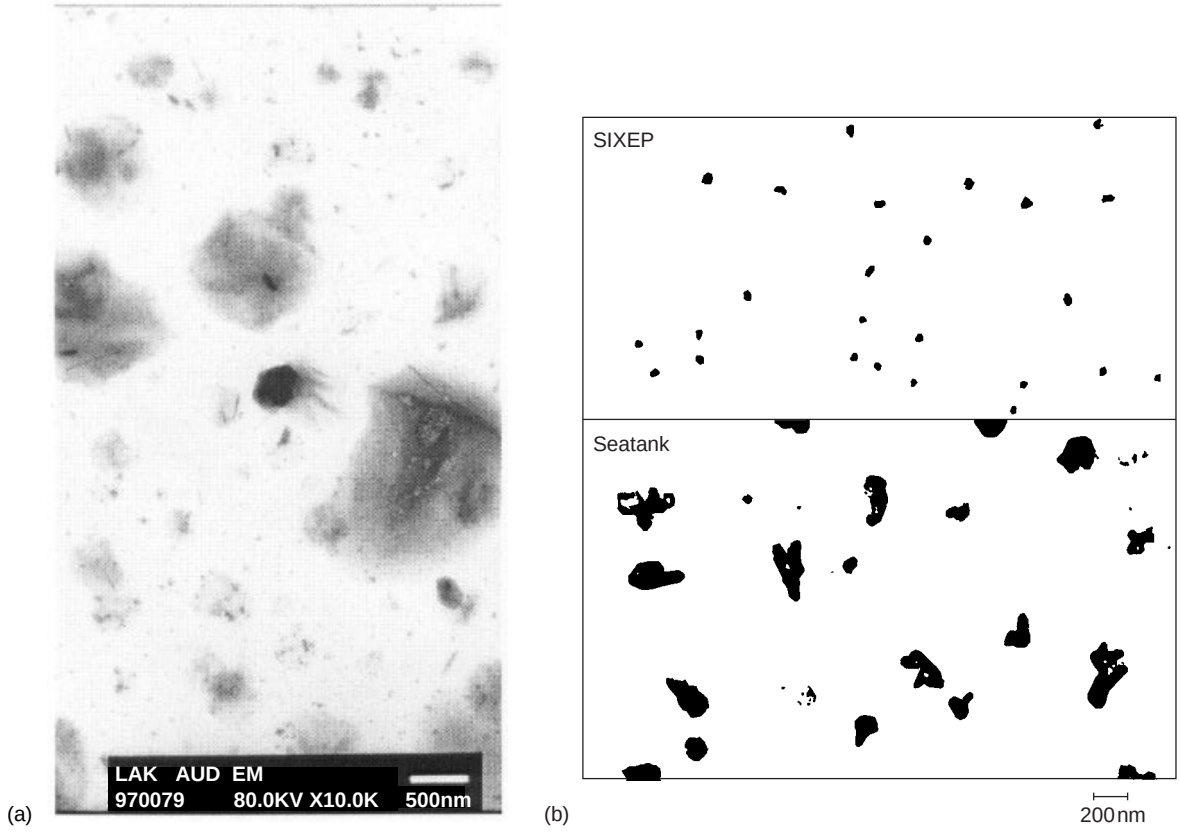

Figure 6 TEM of (a) effluent from the La Hague reprocessing plant, France, and (b) two effluents (SIXEP, Seatank) from the Sellafield reprocessing plant, UK.

5.1.5 Synchrotron-based X-ray Microscopic Techniques

Recent developments within sensitive X-ray microscopic techniques for solid surface analysis have recently been reviewed by Adams et al.[70] Using monochromatic X-ray microbeams, micro-X-ray fluorescence spectroscopy (μ-XRF) and μ-XAS provide information on the elemental distribution on particle surfaces, while μ-XRD gives information on the crystallographic structures on microscopic areas of solid particles. By tuning the energy of the microbeam source and scanning the energy over an absorption edge of an element of interest, X-ray absorption spectrometry (XAS), EXAFS and X-ray absorption near-edge structure spectrometry (XANES) can be utilized for speciation purposes. From μ-XANES data, the oxidation number and coordination state of the studied atom can be derived, while EXAFS provides information on the number of atoms, the atomic number and the distance to neighboring atoms in the surface of the solid.[70]

In the laboratory, efficient micro-X-ray sources based on rotating-anode tubes equipped with capillary X-ray optics can be applied.[70] However, the third-generation synchrotron radiation sources, yielding high-intensity and highly monochromatic X-ray microbeams, have significantly increased the sensitivity and applicability of X-ray microscopic techniques. Using synchrotron-based XANES, the oxidation states of uranium associated or with components in soils and sediments have been well documented.[71,72] Recently, different synchrotron-based X-ray microscopic techniques have been combined by utilizing the micro-X-ray beam at the European Synchrotron Radiation Facility (ESRF), France, to study uranium species in micrometer-sized radioactive particles released from the Chernobyl reactor (unpublished work). From μ-XRD different crystalline forms could be identified, while μ-XANES clearly demonstrated differences in oxidation states of U in released UO_2 fuel; reduced U fuel particles were released during the initial explosion while oxidized U_3O_8 particles were released during the fire. Using microtomography and computerized reconstruction of the particle[73] (Figure 6a and b), information on internal cavities and channels due to the formation of volatiles during operation or the accident could also be observed.

5.2 Size Distribution Pattern

Soil or sediment pore waters (Figure 3) are traditionally collected by squeezing or centrifuging cores, preferentially under in situ temperature and redox conditions. In recent years, porous ceramic cup samplers (vacuum pump suction filtration) with a pore diameter of about 1 μm have been more frequently applied for speciation purposes.[30] Alternatively, dialysis bags, containing water or ion-exchange or chelating resins, have been utilized for the collection of LMM positively charged species in soil–water systems. In both cases, time-dependent sorption of species and clogging of pores may significantly affect the results. For radionuclides associated with the sand, silt and clay fractions in soil and sediments, grain size wet sieving, and radioactivity measurements of the obtained fractions can be useful. Furthermore, individual fractions can be subjected to surface-sensitive analysis and leaching experiments can be performed.

5.3 Extraction Techniques – Reversible/Irreversible Sorption Processes

Numerous articles concerning extraction of radionuclides from soils and sediments have been published in recent years.[21,27,28,41–44,74–77] The application of single extractions and sequential extractions has also recently been reviewed by Kennedy et al.[78] Information on reversible (e.g. physical sorption, electrostatic sorption), and irreversible (e.g. chemisorption) interactions of reactive charged radionuclide species, and the degree of binding to natural components in soils or to deposited radioactive particles (Figure 3) can be attained from extraction experiments. A variety of extraction agents have been applied for extracting radionuclides from soils, in particular from agricultural areas, and from sediments. However, the sequential extraction techniques most often used are modified versions of the method described by Tessier et al.[79] A major objective has been to obtain the bioavailable fraction mimicking the plant uptake. However, no extracting agent so far is able to predict plant uptake without large uncertainties. Reversible and irreversible interactions can be distinguished using sequential extraction of radionuclides in soils or sediments, provided that significant artefacts are not produced. To account for differences in the soil chemistry between individual samples, stable isotopes or analogs in the fractions obtained should also be determined.[49] Comparison of the extraction yields of the radioactive and stable isotopes may also provide information on the degree of isotopic exchange and if the radioactive isotope is associated with particles.

Extraction or sequential extraction techniques performed in laboratories have been subject to criticisms for decades, partly owing to the interpretation of analytical results. Instead of referring to the displacement or dissolution power of the reagents (e.g. pH-sensitive fraction, weakly reduced fraction, weakly oxidized fractions), several authors refer to a presumed mineral phase (e.g. bound to carbonates, bound to Fe and Mn oxides, bound to organics) without proving that the referred phases are present and are fully or partially responding to the specific chemical treatment. Even

though the early sequential extraction schemes were developed to characterize the phases with which contaminants were associated, misinterpretation of data can easily occur. Unless the actual phase is proved to be present, the phrasing should therefore rather refer to the extraction agent applied. Furthermore, the techniques should be standardized if data in the literature are to be compared, and standard reference materials and protocols applicable for speciation purposes should be developed.

5.3.1 Single Extractions of Soil and Sediments

A variety of extraction agents (e.g. electrolytes, complexing agents, acids, redox agents) have been used, and a variety of analytical results have been published. However, the use of an inert, indifferent or weak electrolyte (e.g. 1 M NH_4OAc at soil pH) corresponding to the first step in most sequential extraction procedures has been found useful for studying pollutant and nutrient competition and reversible associations.[78]

5.3.2 Sequential Extraction of Soil and Sediments

Sequential extractions are useful techniques for studying reversible or irreversible interactions of radionuclides with solid phases, when reagents are chosen to differentiate between binding mechanisms (Table 2):

- reversible physical sorption: indifferent inert nonreacting electrolytes;
- reversible electrostatic sorption: pH effect, competing/complexing ligands;
- irreversible chemisorption: red/ox systems and increased temperature.

It is essential that fresh solids are used and that the solid–solution system is relatively unaffected in the initial steps. Thus, electrolytes (e.g. Ca ions) with significant higher ionic strength than the soil solution should be avoided as an initial step as particles may aggregate.

According to Beneš and Majer,[22] desorption mechanisms can be distinguished in displacement reactions; the adsorbed species are displaced by some foreign species and dissolution processes; the physicochemical forms of the desorbed species are changed. By increasing the concentration of an indifferent (inert) electrolyte (e.g. H_2O or NH_4OAc with pH of soils), the fraction of radionuclides which is desorbed by displacement can easily be distinguished from that being released after dissolution induced, for instance, by redox agents. The chosen experimental procedure is crucial for the quality of the analytical results. Using low-speed centrifugation, colloidal material is included in the aqueous phase, whereas with high-speed ultrafiltration most of the colloidal material is included in the solid phase. For organic soils, filtration must often be applied for practical purposes and clogging of filters may occur. The solid : solution ratio is also essential, as interparticle effects may take place at high ratios.

As reversibility is a matter of kinetics, the time needed for reversible processes (e.g. physical or electrostatic sorption) to reach equilibrium may be considerable.[43,80] Thus, the chosen contact time should at least allow rapid surface processes to occur. However, slow diffusion processes are essential when long-term assessments are to be made. Factors influencing the analytical results (solid : solution ratio, stirring, temperature, separation of phases, time of contact) should be standardized to avoid methodological effects (e.g. coprecipitation) and for comparison purposes. When standardized, the precision of the sequential extraction method has proved to be within 10%.[44,49,50,74]

5.3.3 Reversible Interactions

By successively repeating the inert nonreacting electrolyte extraction (e.g. H_2O, NH_4OAc), the reversibility of physical or electrostatic sorption of macrocomponents and also radionuclides can be demonstrated. For mobile species being reversibly associated with solid phases, the chemical yield of successive extractions with aliquots of the same electrolyte should be close to 100%. By extraction 10 times with water, organic carbon (e.g. humic substances) and probably associated radionuclides are continuously released upon each extraction (Figure 7a). Similarly, 90–100% of ^{90}Sr in

Table 2 Sequential extractions; models and reagents

Processes	Model	Agents	Reagents
Physical sorption	Consecutive layers, reversible reaction	Indifferent (inert) electrolyte	H_2O, soil pH NH_4OAc, soil pH
Electrostatic sorption	Monolayer, reversible reaction	Ion-exchangable species, increased acidity (pH)	pH < soil pH NH_4OAc, pH 5
Chemisorption	Monolayer, irreversible reaction	Red/ox agents, increase in temperature	Weak reducing: $NH_2OH{\cdot}HCl$ in HOAc Weak oxidizing: H_2O_2, pH 2 Strong oxidizing: 7 M HNO_3, 80 °C

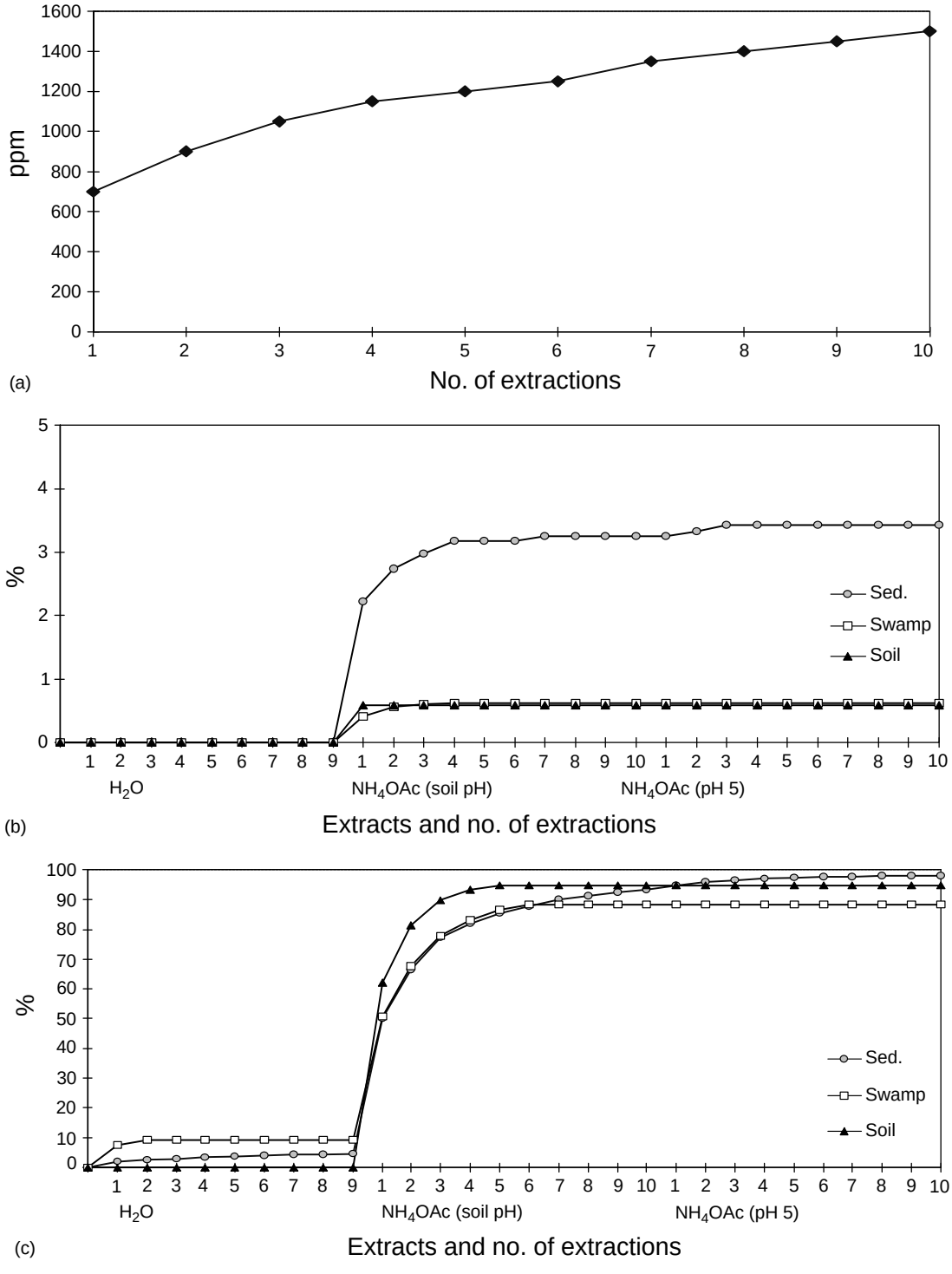

Figure 7 Repeated extraction (×10) with H_2O, NH_4OAc at soil pH, NH_4OAc at pH 5 of (a) organic carbon (extracted with water at soil pH), (b) ^{137}Cs and (c) ^{90}Sr in soils and sediment from Mayak PA, Russia.[7]

contaminated soils, swamp and sediments collected at Mayak PA, Russia, is steadily released upon successive extractions with electrolytes (Figure 7c). For ^{137}Cs (Figure 7b), however, a maximum of 1–3% could be considered reversibly sorbed and exchangeable, i.e. present as mobile species.[7] For Chernobyl soils collected in 1989, the fraction of extractable ^{90}Sr was, however, surprisingly low (Figure 8). Detailed electron microscopy studies and sequential leaching of individual Chernobyl particles and soils containing particles confirmed a very

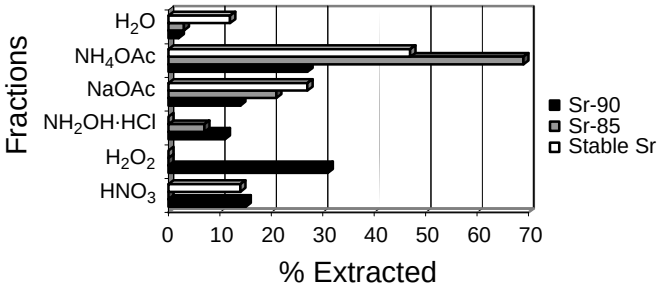

Figure 8 Sequential extraction of ^{90}Sr, stable Sr and the added tracer ^{85}Sr in soils from Chernobyl.[21]

low extractability.[21] By following the increase in the NH_4OAc-extractable ^{90}Sr with time after deposition, the environmental weathering rates of uranium fuel particles could also be determined.[27,81] The weathering rates were primarily dependent on particle composition (e.g. structure, oxidation state) and on soil pH.

5.3.4 Irreversible Interactions

Within the sequential extraction scheme, pH is lowered and the redox power is gradually increased. For irreversibly sorbed species (chemisorption) or species incorporated in the mineral lattice (e.g. crystal defects), rupture of chemical bonds is needed if species are to be released.[22] However, dissolution processes are usually affected by slow kinetics.[80] By successively changing oxidizing reagents, overlapping fractions (low chemical yield in several fractions) may be obtained. Owing to the strong interaction (fixation) to clays, the most effective extracting agent for ^{137}Cs and stable Cs in soil or sediment is usually strongly acidic and oxidizing reagents, e.g. 7 M HNO_3. Unless Sr minerals are present, stable Sr in soils and sediments is usually reversibly associated. Uranium fuel particles are, however, oxidized by H_2O_2 in HNO_3 and associated radionuclides are mobilized. Hence the distribution of extractable radioactive and stable isotopes of, for instance, Cs and Sr will be significantly different. As demonstrated for soils contaminated with fuel particles collected close to the Chernobyl reactor (Figure 8), the enriched H_2O_2–HNO_3 fraction for ^{90}Sr indicates a fuel particle association, and the distribution in extracts is different from that of stable Sr, being NH_4OAc extractable, and of the added $^{85}Sr^{2+}$ ionic tracer. Hence, the weathering rates of fuel particles are a key factor for assessing mobility in soils and uptake of ^{90}Sr in vegetation.

Most extraction schemes involve leaching with strong acid as the final extraction, intended to represent radionuclides associated with the residual mineral fraction. However, mineral phases are attacked at different rates, and temperature and contact time are essential parameters if complete dissolution is to be attained.[21] Therefore, care should be taken when the final analysis is based on radiochemical separations (α- and β-emitters).

6 DYNAMIC MODEL EXPERIMENTS USING TRACERS

Chemically well-defined spectroscopically pure radioactive tracers are efficient tools in radionuclide speciation studies.[22,82] γ-Emitting tracer species can be applied in situ using field lysimeters or in the laboratory to study the behavior of radionuclide species with respect to mobility and biological uptake.[43,83] Furthermore, γ-emitting isotopes of different elements or present as different species of the same element can be studied simultaneously. Using stable isotopes in field studies, samples have to be collected and analyzed in the laboratory. In the field, major natural conditions (composition, temperature, etc.) are maintained, even though some artefacts (e.g. distortion of water flow pattern) can hardly be avoided. However, most tracer experiments are performed in laboratories, under climate-controlled conditions or at room temperature. A variety of size and charge fractionation techniques are applied in tracer experiments as the volumes involved can be small (Table 1). However, methodological effects can easily affect analytical results when small volumes are involved.

6.1 Distribution of Species

Assuming equilibrium in the studied system, autoradiography, scanning techniques or analysis of withdrawn samples are frequently used to identify the distribution of added β- and γ-emitting tracers. After the Chernobyl accident, laboratory experiments were utilized to improve the predicting power of K_d values.[41–43] Samples of soils and sediments were incubated with synthetic solutions, e.g. synthetic fresh water or Ca and Mg or Na and NH_4 solution containing ionic tracers (e.g. Cs isotopes) for a limited time of contact (e.g. 24 h or 7 days). Then K_d values for the tracer (and other major components) were determined after separation of solid and solution or after extractions with weak electrolytes.[41] Assuming equilibrium between species in solution and species reversibly associated with solids, the observed K_d for reversible species can be utilized in short-term assessment. Alternatively, a chemical reagent (e.g. AgTU) is added to soil–water mixtures prior to the addition of the Cs tracer, and according to Wauters et al. AgTU masks the edges of illitic clay minerals to prevent the association of Cs ions.[41,84] Thus, the concept of the radiocesium interception potential (RIP) of the micaceous frayed-edge sites (FES) of soil

or sediment systems has been widely adopted.[41,43,84] For soils and sediments RIP may vary by 2–3 orders of magnitude. Based on 133 soil– or sediment–solution systems, however, the experimental K_ds for ^{137}Cs deviate a factor of <3 from the predicted K_ds for 90% of the observations.[42] Based on this approach, the K_d predictions are based solely on the solid-phase characteristics determined after 24 h of incubation of ionic tracers under laboratory conditions.[42] Hence slow interactions are not taken into consideration. Furthermore, the K_d values obtained from measurements of samples collected in a field contaminated with particles will be different from those obtained from ionic solutions.

Biotest experiments (plant uptake, animal uptake, uptake in aquatic organisms) are most often performed under laboratory conditions. If uptake and excretion rates are well known, TF, MF or CF can be determined after a given time of exposure.[40] Otherwise, dynamic biotest experiments must be performed to ascertain that equilibrium status is reached.

6.2 Kinetics of Transformation Processes

Using size and charge fractionation techniques, changes in the distribution of specific radionuclide species added to a system can be followed over time to obtain information on transformation processes affecting the added species and kinetics involved. Especially in heterogeneous systems (e.g. soil–water or sediment–water systems), no other techniques can provide kinetic information of trace chemistry processes taking place in solutions, at the solid–solution interface and within solids over time.

In dynamic experiments using size and charge fractionation techniques, the formation of radioactive colloids or pseudocolloids from tracer (e.g. transuranic elements) solutions with varying composition can be followed over time.[85] Furthermore, the time-dependent interaction of different species in solution with a variety of solid surfaces can be followed using extraction techniques. Time-dependent interactions of Cs isotopes with solid phases influencing the distribution coefficient K_d in soils have been well documented by Absalom et al.[43] As demonstrated for marine sediments–seawater systems (Figure 9), the interaction of added $^{134}Cs^+$ ions with solids followed at least a two-step function: rapid initial surface interaction and probably slow diffusion into double layers of solids.[86] The interaction and kinetics will depend on the sediment composition and surface structure of the solid phase. Furthermore, interactions and kinetics (Figure 10) will depend on radionuclide species in question, as demonstrated for Pu^{3+}/Pu^{4+} and Pu^{5+}/Pu^{6+} species added to marine sediments.[87] The interaction of Pu^{3+}/Pu^{4+} with sediments is significantly stronger (higher K_d) than for Pu^{5+}/Pu^{6+}. Dynamic K_d experiments can be combined with sequential extractions to distinguish between reversible and irreversible (slow reversible) association of radionuclides. At different time intervals, aliquots are withdrawn and extracted, and information on reversible (mobile) and irreversible (inert) interactions and kinetics involved can be utilized in kinetic models. As illustrated in Figure 11, the time-dependent association of $^{134}Cs^+$ ions with solids can be described as a three-box model,[88] where the kinetics are essential for predicting the soil to water transfer of ionic Cs species.

Similarly, fractionation techniques should be included in biotest experiments, to obtain information on time-dependent uptake of specific radionuclide species. Thus, CF values will vary according to radionuclide species present in the exposure and organisms exposed: high CF is obtained for radiocolloids in filtering organisms, whereas low CF is obtained for fish.[88] Similarly, high CF is obtained for ionic species in fish, but low CF

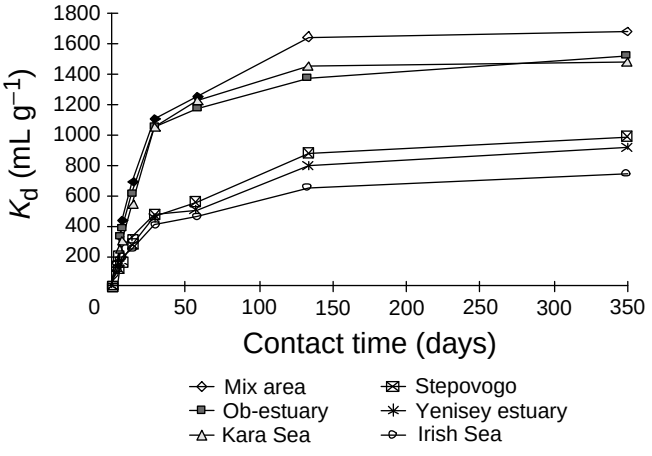

Figure 9 Time-dependent distribution of $^{137}Cs^+$ ions between solid and solution (K_d) in seawater–sediment systems.[86]

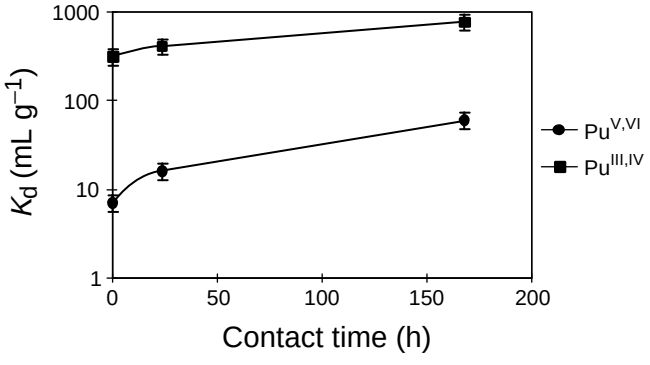

Figure 10 Time-dependent distribution of Pu^{3+}/Pu^{4+} and Pu^{5+}/Pu^{6+} species between solid and solution (K_d) in seawater–sediment systems.[87]

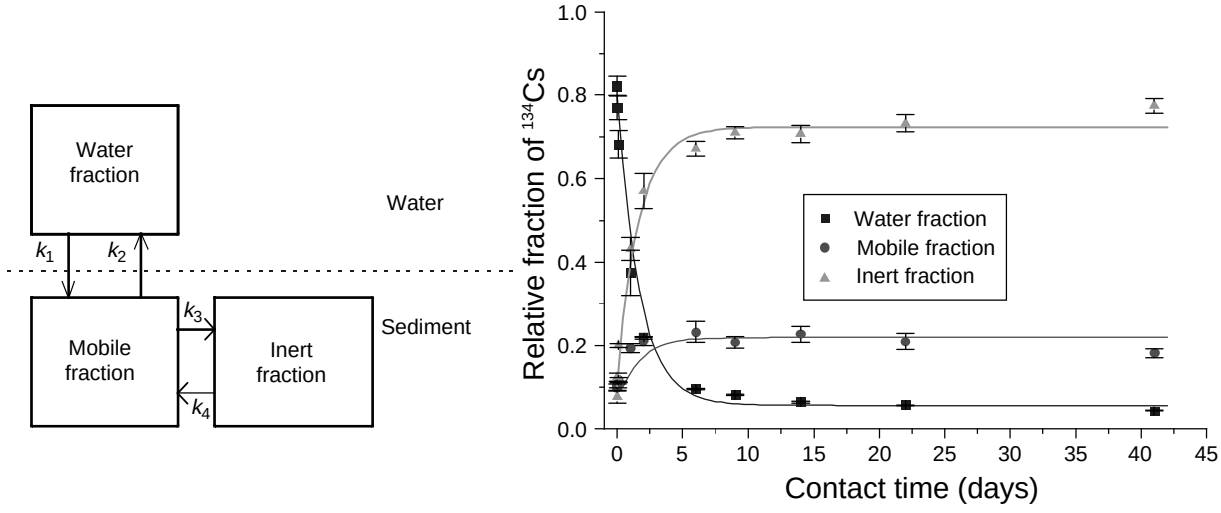

Figure 11 Time-dependent interactions of $^{134}Cs^{+}$ ions in seawater–sediment systems. Three-box dynamic model applied interactions based on dynamic tracer experiments and sequential extractions.[88]

for filtering organisms. As also demonstrated after the Chernobyl accident, the uptake of ionic ^{134}Cs ions in domestic animals was significantly higher than ^{137}Cs from Chernobyl-contaminated fodder.[40]

7 SPECIATION MODELING

In recent years, models based on thermodynamic constants (e.g. stability or complex constants, distribution coefficients, TCs, concentration ratios) have been developed for radionuclide speciation purposes. Usually, input data include the total activity concentration of radionuclides of interest and of selected interacting components in the studied system as well as literature-derived constants. However, several factors may contribute to large uncertainties in the estimates:

- thermodynamic equilibria may not be attained in the studied system owing to slow processes and, especially for unstable mixing zones (e.g. estuaries), such model predictions can hardly be relevant, hence time-dependent functions rather than constants should be derived;
- thermodynamic constants derived from laboratory experiments may not be relevant for natural conditions (e.g. artificial solutions at room temperature);
- literature data may vary considerably (orders of magnitude);
- lack of relevant data on key trace chemistry processes, for instance the behavior of polymers and colloids, implies that most models describe the system only partially.

Hence thermodynamic models predicting mobility and bioavailability of radionuclides from a source may suffer from large uncertainties. To improve the prediction power, experimental information on radionuclide species, transformation and kinetics should be provided and utilized as input data in dynamic models.

8 METHODS OF ANALYSIS

Following fractionation of samples, fractions obtained (e.g. ultrafiltrate) or fractions retained (e.g. particles on membranes) are subjected to analysis of radionuclides and other elements of interest. As the concentration of most radionuclides in obtained fractions usually is low, low-level techniques are essential in speciation studies.

8.1 Low-level Radionuclide Measurements

After preconcentration of γ-emitters from large volumes of fractionated natural waters (e.g. seawater), radionuclides in the concentrate can be determined using high-volume NaI or HPGe detectors. Usually β- and α-emitting radionuclides are analyzed by α- and β-spectrometry or liquid scintillation techniques, after preconcentration with yield monitors and time-consuming radiochemical separations.[89] Using low-level liquid scintillation spectrometers, β- and α-emitting isotopes can be determined simultaneously (e.g. ^{239,240}Pu and ^{241}Pu). However, recent developments within sensitive mass spectrometric techniques (ICPMS, AMS) have proved to be an excellent alternative.

8.2 Mass Spectrometric Methods (Inductively Coupled Plasma Mass Spectrometry, Accelerator Mass Spectrometry)

Mass spectrometric methods for determination of very low concentrations of elemental isotopes, and thereby isotopic ratios, have been developed using a variety of ion sources. Among these are thermal ionization mass spectrometry (TIMS) and SIMS, ICPMS and AMS. Although primarily applied to the determination of stable element isotopes and radiogenic isotopes in geological studies, the techniques can equally well be applied to the measurement of activity concentrations of long-lived low-level radionuclides in various samples using ''isotope dilution'' methods. Excellent reviews of these techniques have been published elsewhere.[90,91] In recent years, emphasis has been put on the development of ICPMS and AMS techniques applicable in particular for actinides.

8.2.1 Inductively Coupled Plasma Mass Spectrometry

In ICPMS, samples are introduced by means of a nebulizer, graphite furnace or laser (laser ablation). It is assumed that the sample is fully ionized in the plasma (6000–8000 K) before being transferred into the mass spectrometer. Using nebulizer systems for sample introduction, however, particle discrimination may occur. By scanning the mass range within milliseconds, ICPMS acts as a highly sensitive multielemental technique. For transuranic elements ICPMS has proved to be most useful, by distinguishing between low levels of nuclear weapon-derived ^{239}Pu and nuclear fuel cycle-derived ^{240}Pu. Hence the isotopic ratios can be utilized for source identification (''fingerprint''). By interfacing, for instance, chromatographic separation systems (e.g. LC or HPLC) with ICPMS, the extreme sensitivity of the combined system should be well suited for radionuclide speciation studies in the laboratory.

8.2.2 Accelerator Mass Spectrometry

Although not as widespread as ICPMS, AMS offers almost unparalleled resolution, precision and ultralow-level detection limits (from 10^4 to 10^6 atoms). The technique requires the preparation of homogeneous solid metallic or metal oxide samples containing the radionuclide of interest together with a suitable isotopic tracer. The ion source is produced by spluttering the solid sample with Cs^+ ions and injecting the negative ions into the accelerator. The beam is then accelerated towards a high-voltage positive potential, stripped of outer electrons and the positively charged ions formed (up to 13+) are accelerated. Then, the isotopes are separated according to their mass/charge ratios in a high magnetic field before counting. The technique has been developed to permit the measurement of heavy radionuclides such as Pu and Np[92] and has been applied successfully to measurement of ^{240}Pu/^{239}Pu isotope ratios in low-level environmental samples.[93]

9 PERSPECTIVE AND FUTURE DEVELOPMENTS

When short- or long-term consequences of radionuclides released to different ecosystems are assessed, the key factors contributing to major uncertainties are poorly defined source term characteristics especially with respect to the physicochemical form of radionuclides, poorly defined time-dependent interactions of different radionuclide species between water and solids [constant distribution coefficients (K_d) for a given nuclide in a defined soil– or sediment–water system] and poorly defined time-dependent transfer of different radionuclide species to biological systems (constant CFs for a given nuclide to a specific organism). As illustrated in Table 3, a release of mobile species will usually imply a low K_d for soils and sediments and a high CF, for instance for fish. If transformation into reactive species occurs with time (e.g. Pu^{5+}/Pu^{6+} reduced to Pu^{3+}/Pu^{4+}), K_d will increase, CF for fish will decrease and the retention in benthic organisms will be the relevant pathway in dose-assessments. Alternatively, if inert radioactive particles are released, the observed K_d values will be significantly higher and CFs (in fish) lower than expected from literature data. When weathering of particles occurs and mobile species are released, K_d will decrease, whilst CF for fish will increase with time. Hence time-dependent changes in the distribution of radionuclide species and subsequently in observed K_ds and CFs will significantly influence the dose assessments.

Fractionation, identification and characterization of radionuclide species in the environment are still a challenge within analytical radiochemistry. Although the analytical strategy for radionuclide speciation includes fractionation in situ, at-site or shortly after sampling, most emphasis has so far been put on filtration of waters, extraction of soils and sediments and model experiments in the laboratory. In addition, radioactive particles have usually been identified from the emission of high-energy γ- or β-radiation only. Hence, the potential for improvements is substantial. Among size-fractionation techniques, in situ high-capacity ultrafiltration (e.g. tangential flow systems) is favorable as particles, colloids and LMM species in natural waters can be distinguished. By interfacing charge fractionation techniques (e.g. chromatography), information on neutral and positively and negatively charged LMM species can be attained. As LMM species

Table 3 Impact of mobile species and hot particles

	Source term →	Transport → processes	Biological → uptake	Dose-assessment
Impact of	Speciation	K_d	CF	mSv
Mobile species	High load of mobile species	Low Increase $f(t)$ when transformed into reactive species interacting with surfaces	High in fish Low in benthic Decrease $f(t)$ in fish for reactive species Increase $f(t)$ in benthic for reactive species	Underestimated short-term traditional dose-assessment Overestimated long-term traditional dose-assessment for reactive species
Hot particles	High load of inert particles	Very high Decrease $f(t)$ when transferred into mobile species (e.g. weathering and mobilization of ^{90}Sr)	Low in fish High in benthic Increase $f(t)$ in fish for mobile species Decrease $f(t)$ in benthic for mobile species	Overestimated short-term traditional dose-assessment Underestimated long-term traditional dose-assessment for mobile species

are believed to be mobile and potentially bioavailable, a major challenge within analytical chemistry is the development of techniques able to differentiate between LMM species in heterogeneous systems.

Soils and sediments may act as sinks for released radionuclides. Owing to remobilization, however, contaminated sediments may also act as a diffuse source of radionuclides in the future. Experimental information on dynamic solid–water interactions in soil and sediment systems is therefore greatly needed. As process information can hardly be obtained from total activity concentration data, techniques distinguishing between radionuclide species reversibly and irreversibly associated with particles should be further exploited. By combining chemical methods (e.g. ultrafiltration, chromatography, extractions) with surface-sensitive techniques, e.g. electron microscopy or X-ray- or laser-based spectroscopy, valuable information on solid-state species and changes in structure and oxidation states could be followed over time. The combination of advanced size and charge fractionation techniques (LC/MS/MS, HPLC/ICPMS) should also be utilized in speciation studies. Hence new insights into trace chemistry processes influencing radionuclide mobility and bioavailability can be implemented in dispersion, transport and dose-assessment models to improve their predictive power.

ACKNOWLEDGMENTS

The author is indebted to colleagues at the Agricultural University of Norway (Dr D.H. Oughton, Dr T. Krekling and Dr P. Borretzen) and the Norwegian Radiation Protection Authority (Dr P. Strand) and also to colleagues in Ukraine, Belarus and Russia for valuable collaboration.

ABBREVIATIONS AND ACRONYMS

AMS	Accelerator Mass Spectrometry
BEI	Backscattered Electron Imaging
CF	Concentration Factor
EXAFS	Extended X-ray Absorption Fine Structure Analysis
FES	Frayed-edge Sites
HMM	High Molecular Mass
HPLC	High-performance Liquid Chromatography
HPLC/ICPMS	High-performance Liquid Chromatography/Inductively Coupled Plasma Mass Spectrometry
ICPMS	Inductively Coupled Plasma Mass Spectrometry
LC	Liquid Chromatography
LC/MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
LIPAS	Laser-induced Photoacoustic Spectrometry
LMM	Low Molecular Mass
MF	Mobility Factor
RIP	Radiocesium Interception Potential
SEI	Secondary Electron Imaging
SEM	Scanning Electron Microscopy
SIMS	Secondary Ionization Mass Spectrometry
TC	Transfer Coefficient
TEM	Transmission Electron Microscopy
TF	Transfer Factor
TIMS	Thermal Ionization Mass Spectrometry
UV	Ultraviolet

XANES	X-ray Absorption Near-edge Structure Spectrometry
XAS	X-ray Absorption Spectrometry
XRMA	X-ray Microanalysis
μ-XANES	Micro-X-ray Absorption Near-edge Spectroscopy
μ-XAS	Micro-X-ray Absorption Spectroscopy
μ-XRD	Micro-X-ray Diffraction
μ-XRF	Micro-X-ray Fluorescence Spectroscopy

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Laser Ablation Inductively Coupled Plasma Spectrometry in Environmental Analysis • Liquid Chromatography/Mass Spectrometry in Environmental Analysis

Field-portable Instrumentation **(Volume 5)**
Solid-phase Microextraction in Analysis of Pollutants in the Field

Particle Size Analysis **(Volume 6)**
Field-flow Fractionation in Particle Size Analysis • Filtration in Particle Size Analysis

Surfaces **(Volume 10)**
Scanning Electron Microscopy in Analysis of Surfaces

Radiochemical Methods **(Volume 14)**
Mass Spectrometry of Long-lived Radionuclides

REFERENCES

1. UNSCEAR, *Sources and Effects of Ionizing Radiation*, United Nations Scientific Committee on the Effects of Atomic Radiation, United Nations, New York, 1993.
2. P. Strand, M. Balonov, A. Aarkrog, M.J. Bewers, B. Howard, Y. Tsaturov, *Arctic Pollution Issues: Radioactive Contamination*, Norwegian Radiation Protection Authority, Østerås, 1997.
3. K.H. Lieser, 'Radionuclides in the Geosphere: Sources, Mobility, Reactions in Natural Waters and Interactions with Solids', *Radiochim. Acta*, **70**, 355–375 (1995).
4. H.R. von Gunten, P. Beneš, 'Speciation of Radionuclides in the Environment', *Radiochim. Acta*, **69**, 1–29 (1995).
5. K.S. Leonard, D. McCubbin, B.R. Harvey, 'Chemical Speciation and Environmental Behaviour of ^{60}Co Discharged from a Nuclear Establishment', *J. Environ. Radioact.*, **20**, 1–21 (1993).
6. B. Salbu, H.E. Bjørnstad, I. Sværen, S.L. Prosser, R.A. Bulman, B.R. Harvey, M.B. Lovett, 'Size Distribution of Radionuclides in Nuclear Fuel Reprocessing Liquids After Mixing with Seawater', *Sci. Total Environ.*, **130/131**, 51–63 (1993).
7. Joint Norwegian–Russian Expert Group for Investigation of Radioactive Contamination in the Northern Areas, 'Sources Contributing to Radioactive Contamination of the Techa River and Areas Surrounding the ''Mayak'' Production Association, Urals, Russia', Norwegian Radiation Protection Authority, Østerås, 1997.
8. B. Salbu, H.E. Bjørnstad, E. Lydersen, A.C. Pappas, 'Determination of Radionuclides Associated with Colloids in Natural Waters', *J. Radioanal. Nucl. Chem.*, **115**, 113–135 (1987).
9. S. Simon, T. Jenner, J. Graham, A. Borcher, 'A Comparison of Macro- and Microscopic Measurements of Plutonium in Contaminated Soil from the Republic of the Marshall Islands', *J. Radioanal. Nucl. Chem.*, **194**, 197–205 (1995).
10. M. Cooper, P. Burns, B. Tracy, M. Wilks, G. Williams, 'Characterization of Plutonium Contamination at the Former Nuclear Weapons Testing Range at Maralinga in South Australia', *J. Radioanal. Nucl. Chem.*, **177**, 161–184 (1994).
11. P.R. Danesi, P. De Regge, J. La Rosa, M. Makarewicz, J. Moreno, Z. Radecki, E. Zeiller, 'Residual Plutonium Isotopes and Americium in the Terrestrial Environment at the Former Nuclear Test Sites of Mururoa and Fangataufa', presented at the 7th International Conference on Low-level Measurements of Actinides and Long-lived Radionuclides in Biological and Environmental Samples, Salt Lake City, September 1998.
12. L. Arnold, *The Windscale Fire 1957, Anatomy of a Nuclear Accident*, Macmillan Press, London, 1992.
13. J.N. Smith, K.M. Ellis, A. Aarkrog, H. Dalsgaard, E. Holm, 'Sediment Mixing and Burial of the ^{239,240}Pu from the Thule, Greenland Nuclear Accident', *J. Environ. Radioact.*, **25**, 134–159 (1994).
14. N.A. Loschilov, V.A. Kashparov, E.B. Yudin, V.P. Protsak, M.A. Zhurba, A.E. Parshakov, 'Physicochemical Forms of the Radioactive Fallout from the Chernobyl Reactor Accident', in *The Radiobiological Impact of Hot Beta-particles from the Chernobyl Fallout: Risk Assessment, Kiev, Ukraine, August, Part I*, IAEA, Vienna, 34–39, 1992.
15. N.A. Loschilov, V.A. Kashparov, V.D. Polyakov, V.P. Protsak, E.B. Yudin, M.A. Zurba, A.E. Parshakov, 'Nuclear-physical Properties of Hot Particles Formed as a Result of Accident at ChNPP', *Radiochemistry*, **4**, 113–119 (1992) (in Russian).

16. A. Bogatov, A. Borovoy, ' ''Hot'' Fuel Particles from Unit 4 and Vicinity: Research Methods and Phenomenological Description', in *The Radiobiological Impact of Hot Beta-particles from Chernobyl Fallout: Risk Assessment, Kiev, Ukraine, August, Part II*, IAEA Vienna, 1–16, 1992.
17. V.D. Kuriny, Y.A. Ivanov, V.A. Kashparov, N.A. Loschilov, V.P. Protsak, E.B. Yudin, M.A. Zurba, A.E. Parshakov, 'Particle Associated Chernobyl Fall-out in the Local and Intermediate Zones', *Ann. Nucl. Energy*, **20**, 415–420 (1993).
18. D. Jakeman, *Notes of the Level of Radioactive Contamination in the Sellafield Area Arising from Discharges in the Early 1950s*, UKAEA, AEEW Report 2104, Atomic Energy Establishment, Winfrith, 1986.
19. B. Salbu, T. Krekling, D.H. Oughton, G. Østby, V.A. Kashparov, T.L. Brand, J.P. Day, 'Hot Particles in Accidental Releases from Chernobyl and Windscale Nuclear Installations', *Analyst*, **119**, 125–130 (1994).
20. K. Bunzl, 'Probability of Detecting Hot Particles in Environmental Samples by Sample Splitting', *Analyst*, **122**, 653–656 (1997).
21. D.H. Oughton, B. Salbu, T.L. Brand, J.P. Day, A. Aarkrog, 'Under-determination of Strontium-90 in Soils Containing Particles of Irradiated Uranium Oxide Fuel', *Analyst*, **118**, 1101–1105 (1993).
22. P. Beneš, V. Majer, *Trace Chemistry of Aqueous Solutions*, Elsevier, Amsterdam, 1980.
23. N.V. Viktorova, E.K. Geiger, *Investigation of the Deposition and Spread of Radioactive Aerosol Particles in the Chernobyl Zone Based on Biological Monitoring*, Report EUR 13 574, Commission of the European Communities, Luxembourg, 223–236 (1990).
24. L. Devell, U. Tovedal, U. Bergstrøm, A. Applegren, J. Chussler, L. Andersson, 'Initial Observations of Fallout from the Reactor Accident at Chernobyl', *Nature*, **321**, 817–819 (1986).
25. T. Raunemaa, S. Lehtinen, M. Kulmala, '2–10 μm Sized Hot Particles in Chernobyl Fallout to Finland', *J. Aerosol Sci.*, **18**, 693–699 (1987).
26. B. Salbu, 'Radionuclides Associated with Colloids and Particles in Rainwaters, Oslo, Norway', in *Hot Particles from the Chernobyl Fallout*, eds. H. von Philipsborn, F. Steinhäusler, Bergbau- und Industriemuseums, Theuern, 83–84, Vol. 16, 1988.
27. V.A. Kashparov, D.H. Oughton, S.I. Zvarich, V.P. Protsak, S.E Levchuk, 'Kinetics of Fuel Particle Weathering and ^{90}Sr Mobility in the Chernobyl 30-km Exclusion Zone', *Health Phys.*, **76**, 251–259 (1999).
28. T.I. Bobovnikova, Y.P. Virchenko, A.V. Konoplev, A.A. Siverina, I. Shkuratova, 'The Role of Chemical Speciation of Radionuclides in Soils in Their Transfer to Surface Runoff', *Sov. Soil. Sci.*, **23**, 5–52 (1991) (in Russian).
29. W.A. McKay, N.J. Pattenden, 'The Transfer of Radionuclides from Sea to Land via the Air – a Review', *J. Environ. Radioact.*, **12**, 49–77 (1990).
30. A.F. Nisbet, B. Salbu, S. Shaw, 'Association of Radionuclides with Different Molecular Size-fractions in Soil Solution; Implications for Plant Uptake', *J. Environ. Radioact.*, **18**, 71–84 (1993).
31. E. Degueldre, G. Longworth, V. Moulin, P. Vilks, C. Ross, G. Bidoglio, A. Cremers, J. Kim, J. Pieri, J. Ramsay, B. Salbu, U. Vuorinen, 'An International Intercomparison Exercise on the Sampling and Characterization of Groundwater Colloid', *Grimsel Colloid Exercise*, Paul Scherrer Institute, PSI-Bericht, No. 39, 68, 1989.
32. E.L. Cooper, M.K. Haas, J.F. Mattie, 'Studies of the Speciation of Plutonium and Other Actinides in Natural Groundwater Using Anion-exchange Resin', *Appl. Radiat. Isot.*, **46**, 1159–1173 (1995).
33. B. Salbu, H.E. Bjørnstad, J.E. Brittain, 'Fractionation of Cs Isotopes and Sr-90 in Snow-melt Runoff and Lake Waters from Contaminated Areas in Norway', *J. Radioanal. Nucl. Chem.*, **156**, 7–20 (1989).
34. K.A. Orlandini, W.R. Penrose, B.R. Harvey, M.B. Lovett, M.W. Findlay, 'Colloidal Behaviour of Actinides in an Oligotropic Lake', *Environ. Sci. Technol.*, **24**, 706–712 (1990).
35. P.I. Mitchell, J.V.I. Batlle, A.B. Downes, O.M. Condren, L.L. Vintro, J.A. Sanches-Cabeza, 'Recent Observation on the Physicochemical Speciation of Plutonium in the Irish Sea and the Western Mediterranean', *Appl. Radiat. Isot.*, **46**, 1175–1190 (1995).
36. B.O. Rosseland, I.A. Blakar, A. Bulger, F. Kroglund, A. Kvellestad, E. Lydersen, D.H. Oughton, B. Salbu, M. Staurnes, R. Vogt, 'The Mixing Zone Between Limed and Acidic River Waters: Complex Aluminium Chemistry and Extreme Toxicity for Salmonoids', *Environ. Pollut.*, **78**, 3–8 (1992).
37. K.S. Leonard, B.R. Harvey, R.J. Woodhead, T. Brooks, D. McCubbin, 'Assessment of an Ultrafiltration Technique for the Fractionation of Radionuclides Associated with Humic Material', *J. Radional. Nucl. Chem.*, **181**, 309–320 (1994).
38. D.E. Robertson, A.J. Schilk, K.H. Abel, E.A. Lepel, C.W. Thomas, S.L. Pratt, E.L. Cooper, P. Hartwig, R.W.D. Killey, 'Chemical Speciation of Radionuclides Migrating in Groundwaters', *J. Radioanal. Nucl. Chem.*, **194**, 237–252 (1995).
39. H. Nitsche, A. Muller, E.M. Standifer, S.R. Deinhammer, K. Becraft, T. Prussin, R.C. Gatti, 'Dependence of Actinide Solubility and Speciation on Carbonate Concentrations and Ionic-strength in Groundwater', *Radiochim. Acta*, **58/59**, 27–32 (1992).
40. K. Hove, P. Strand, B. Salbu, D.H. Oughton, N. Astasheva, A. Sobolev, A. Vasiliev, A. Ratnikov, R.M. Aleksakhin, T. Jigareva, V. Averin, S.K. Firsakova, M. Crick, J.I. Richards, 'Use of Caesium Binders to Reduce Radiocaesium Contamination of Milk and Meat in Belarus, Russia and Ukraine', in *International Symposium on Environmental Impact of Radioactive Releases*, IAEA-SM-339, IAEA, Vienna, 539–547, 1995.

41. J. Wauters, A. Elsen, A. Cremers, A.V. Konoplev, A.A. Bulgakov, R.N.J. Comans, 'Prediction of Solid/Liquid Distribution Coefficients of Radiocaesium in Soils and Sediments. Part 1: a Simplified Procedure for the Solid-phase Characterization', *Appl. Geochem.*, **11**, 589–594 (1996).
42. J. Wauters, A. Elsen, A. Cremers, 'Prediction of Solid/Liquid Distribution Coefficients of Radiocaesium in Soils and Sediments. Part 3: a Quantitative Test of a K_D Predictive Equation', *Applied Geochem.*, **11**, 601–603 (1996).
43. J.P. Absalom, N.M.J. Crout, S.D. Young, 'Modeling Radiocaesium Fixation in Upland Organic Soils of Northwest England', *Environ. Sci. Technol.*, **30**, 2735–2741 (1996).
44. R.N.J. Comans, M. Haller, P. de Preter, 'Sorption of Cesium on Illite: Non-equilibrium Behaviour and Reversibility', *Geochim. Cosmochim. Acta*, **55**, 433–440 (1991).
45. G.M. Desmet, P. Nassimbeni, M. Belli, *Transfer of Radionuclides in Natural and Semi-natural Environments*, Elsevier Applied Science, Barking, 1990.
46. K. Simkiss, M.G. Taylor, 'Transport of Metals Across Membranes in Metal Speciation and Bioavailability in Aquatic Systems' in *Metals in Marine Systems*, eds. A. Tessier, D.R. Turner, John Wiley & Sons, Chichester, 1–44, 1995.
47. J.C. Fisher, J.R. Reinfelder, 'The Tropic Transfer of Metals in Marine Systems', in *Metals in Marine Systems*, eds. A. Tessier, D.R. Turner, John Wiley & Sons, Chichester, 363–406, 1995.
48. Y.I. Bondar, Yu.A. Ivanov, A.G. Ozornov, 'Estimation of Relative Biologic Availability of ^{137}Cs in and of its Total Biologic Availability in Soils at Territory Subjected to Radioactive Contamination', in *The Radiobiological Impact of Hot Beta-particles from the Chernobyl Fallout: Risk Assessment, Kiev, Ukraine, August, Part 1*, IAEA, Vienna, 58–67, 1991.
49. D.H. Oughton, B. Salbu, G. Riise, H. Lien, G. Østby, A. Nøren, 'Radionuclide Mobility and Bioavailability in Norwegian and Soviet Soils', *Analyst*, **117**, 481–486 (1992).
50. B. Salbu, D.H. Oughton, A.V. Ratnikov, T.L. Zhigareva, S.V. Kruglov, K.V. Petrov, N.V. Grebenshakikova, S.K. Firsakova, N.P. Astasheva, N.A. Loshchilov, K. Hove, P. Strand, 'The Mobility of ^{137}Cs and ^{90}Sr in Agricultural Soils in the Ukraine, Belarus and Russia, 1991', *Health Phys.*, **67**, 518–528 (1994).
51. B. Salbu, T. Krekling, K. Hove, D.H. Oughton, V.A. Kashparov, N.A. Astasheva, 'Biological Relevance of Hot Particles Ingested by Domestic Sheep', in *International Symposium on Environmental Impact of Radioactive Releases*, IAEA-SM-339, IAEA, Vienna, 324–325, 1995.
52. G.M. Gadd, 'Influence of Microorganisms on the Environmental Fate of Radionuclides', *Endeavour*, **20**, 150–156 (1996).
53. J.R. Kramer, H.E. Allen (eds.), *Metal Speciation, Theory, Analysis and Application*, Lewis Publishers, Chelsea, MI, 1989.
54. J.W. Patterson, R. Passion (eds.), *Metals: Speciation, Separation and Recovery*, Lewis Publishers, Chelsea, MI, 1990.
55. G.E. Batley (ed.), *Trace Element Speciation: Analytical Methods and Problems*, CRC Press, Boca Raton, FL, 1989.
56. B. Salbu, E. Steinnes (eds.), *Trace Elements in Natural Waters*, CRC Press, Boca Raton, FL, 1995.
57. J.F. McCarthy, C. Degueldre, 'Sampling and Characterization of Colloids and Particles in Groundwater for Studying Their Role in the Subsurface Transport of Contaminants in Environmental Particles', *Environmental Particles*, eds. J. Buffle, H. Vanleeuwen, Lewis Publishers, Chelsea, MI, 1992.
58. M.R. Barman, J.H. Chen, G.J. Wasserburg, 'Th-230–Th-232 Systematics in the Central Pacific Ocean: The Sources and the Fates of Thorium', *Earth Planet. Sci. Lett.*, **139**, 351–363 (1996).
59. W. Buddemeier, J.R. Hunt, 'Transport of Colloidal Contaminants in Groundwater: Radionuclide Migration at the Nevada Test Site', *Appl. Geochem.*, **3**, 535–548 (1988).
60. E. Lydersen, H.E. Bjørnstad, B. Salbu, A.C. Pappas, 'Trace Element Speciation in Natural Waters Using Hollow-fibre Ultrafiltration', in *Speciation of Metals in Water, Sediments and Soil Systems*, ed. L. Landner, Springer, Berlin, 85–97, 1987.
61. B. Salbu, A.I. Nikitin, P. Strand, G.C. Christensen, V.B. Chumichev, B. Lind, H. Fjelldal, T.D.S. Bergan, A.L. Rudjord, M. Sickel, N.K. Valetova, L. Foyn, 'Radioactive Contamination from Dumped Nuclear Waste in the Kara Sea – Results from the Joint Russian–Norwegian Expeditions in 1992', *Sci. Total Environ.*, **202**, 185–198 (1994).
62. C. Jeandel, J.M. Martin, A.J. Thomas, 'Plutonium and other Artificial Radionuclides in the Seine Estuary and Adjacent Areas', in *Techniques for Identifying Transuranic Speciation in Aquatic Environments*, IAEA, Vienna, 89–102, 1981.
63. Y.K. Chau, 'Chromatographic Techniques in Metal Speciation', *Analyst*, **117**, 571–575 (1992).
64. C.T. Driscoll, 'A Procedure for the Fractionation of Aqueous Aluminium in Dilute Acidic Waters', *Int. J. Environ. Anal. Chem.*, **16**, 267–283 (1984).
65. D.M. Nelson, M.B. Lovett, 'Oxidation State of Plutonium in the Irish Sea', *Nature*, **276**, 599–601 (1978).
66. D. Boust, P.I. Mitchell, K. Garcia, O. Condren, L.L. Vintro, G. Leclerc, 'A Comparative Study of the Speciation and Behaviour of Plutonium in the Marine Environment of Two Reprocessing Plants', *Radiochim. Acta*, **74**, 203–210 (1996).
67. G.M. Morrison, 'Approaches to Metal Speciation Analysis in Natural Waters', in *Speciation of Metals in Water,*

Sediment and Soil Systems, ed. L. Landner, Springer Verlag, Berlin, 58–73, 1987.
68. B. Salbu, T. Krekling, D.H. Oughton, 'Characterization of Radioactive Particles in the Environment', *Analyst*, **123**, 843–849 (1998).
69. J.I. Kim, R. Stumpe, R. Klenze, 'Laser Induced Photoacoustic Spectroscopy for the Speciation of Transuranic Elements in Natural Aquatic Systems', *Top. Curr. Chem.*, **157**, 130–179 (1990).
70. F. Adams, K. Janssens, A. Snigirev, 'Microscopic X-ray Fluorescence Analysis and Related Methods with Laboratory and Synchrotron Radiation Sources', *J. Anal. At. Spectrom.*, **13**, 319–331 (1998).
71. H. Nitsche, 'Synchrotron X-ray Absorption Spectroscopy – a New Tool for Actinide and Lanthanide Speciation in Solids and Solution', *J. Alloys Compd.*, **223**, 274–279 (1995).
72. P.M. Bertsch, D.B. Hunter, S.R. Sutton, S. Bajt, M.L. Rivers, 'In Situ Chemical Speciation of Uranium in Soils and Sediments by Micro-X-ray Absorption Spectroscopy', *Environ. Sci. Technol.*, **28**, 980–985 (1994).
73. C. Raven, A. Snigirev, I. Snigereva, P. Spanne, A. Souvorov, V. Kohn, 'Phase-contrast Microtomography with Coherent High-energy Synchrotron X-rays', *Appl. Phys. Lett.*, **69**, 1826–1828 (1996).
74. G. Riise, H.E. Bjørnstad, H.N. Lien, D.H. Oughton, B. Salbu, 'A Study on Radiocaesium Association with Soil Components Using a Sequential Extraction Procedure', *J. Radioanal. Nucl. Chem.*, **142**, 531–538 (1990).
75. M. Vidal, G. Rauret, ' Prediction Capacity of a Sequential Extraction Scheme', *J. Radioanal. Nucl. Chem.*, **181**, 85–95 (1994).
76. K. Bunzl, W. Kracke, W. Schimmack, L. Zelles, 'Forms of Fallout ^{137}Cs and $^{239+240}$Pu in Successive Horizons of a Forest Soil', *J. Environ. Radioact.*, **39**, 55–65 (1998).
77. M. Trautmannsheimer, P. Schramel, R. Winkler, K. Bunzl, 'Chemical Fractionation of Some Natural Radionuclides in a Soil Contaminated by Slags', *Environ. Sci. Technol.*, **32**, 238–243 (1998).
78. V.H. Kennedy, A.L. Sanches, D.H. Oughton, A.P. Rowland, 'Use of Single and Sequential Chemical Extractants to Assess Radionuclide and Heavy Metal Availability from Soils for Root Uptake', *Analyst*, **122**, R98–R100 (1997).
79. A. Tessier, P.G.C. Campbell, M. Bisson, 'Sequential Extraction Procedure for the Speciation of Particulate Trace Metals', *Anal. Chem.*, **51**, 844–851 (1979).
80. J.P. Absalom, S.D. Young, N.M.J. Crout, 'Radio-caesium Fixation Dynamics: Measurement in Six Cumbrian Soils', *Eur. J. Soil Sci.*, **46**, 461–469 (1996).
81. S.V. Krouglov, A.D. Kurinov, R.M. Aleksakhin, 'Chemical Fractionation of Sr-90, Ru-106, Cs-137 and Ce-144 in Chernobyl-contaminated Soils: an Evolution in the Course of Time', *J. Environ. Radioact.*, **38**, 59–76 (1998).
82. B. Salbu, 'Radioactive Tracer Techniques in Speciation Studies', *Environ. Technol. Lett.*, 381–392 (1987).
83. G. Szabo, J. Guczi, A. Nisbet, 'Investigation of The Solid-phase Speciation of Sr-90, Cs-137, Pu-239 and Am-241 in Soils Determined by Extraction and Ultrafiltration Methods', *J. Radioanal. Nucl. Chem.*, **226**, 255–259 (1997).
84. A. Cremers, A. Elsan, P. de Preter, A. Maes, 'Quantitative Analysis of Radiocesium Retention in Soils', *Nature*, **335**, 247–249 (1988).
85. K.H. Lieser, R. Hill, 'Hydrolysis and Colloid Formation of Thorium in Water and Consequences for its Migration Behaviour – Comparison with Uranium', *Radiochim. Acta*, **56**, 37–45 (1992).
86. D.H. Oughton, P. Børretzen, B. Salbu, E. Tronstad, 'Mobilization of Radionuclides from Sediments: Potential Sources to Arctic Waters', *Sci. Total Environ.*, **202**, 155–165 (1997).
87. B. Salbu, P. Børretzen, D.H. Oughton, M. Sickel, L. Skipperud, 'The Influence of Radionuclide Speciation on Distribution Coefficients (K_d) and Concentration Factors (CF)', in *Radionuclides in other Oceans – Distribution, Models and Impacts, Booklet of Abstracts, Norwich/Lowestoft, England*, 328, 1997.
88. P. Børretzen, personal communication.
89. E. Holm, 'Assessment of Beta and Alpha Particle Emitting Nuclides', in *Radioecology, Lectures in Environmental Radioactivity*, ed. E. Holm, World Scientific, Singapore, 295–313, 1994.
90. G.P. Russ, 'Isotope Ratio Measurements Using ICPMS', in *Applications of Inductively Coupled Plasma Mass Spectrometry*, eds. A.R. Date, A.L. Gray, Blackie, Glasgow, 90–114, 1989.
91. K.G. Heumann, 'Isotope-dilution Mass Spectrometry', in *Inorganic Mass Spectrometry*, eds. F. Adams, R. Gijbels, R. van Grieken, Chemical Analysis Series, John Wiley & Sons, Chichester, 301–376, Vol. 95, 1988.
92. L.K. Fifield, R.G. Cresswell, M.L. diTada, T.R. Ophel, J.P. Day, A.P. Clacher, S.J. King, N.D. Priest, 'Accelerator Mass Spectrometry of Plutonium Isotopes', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **117**, 295–303 (1996).
93. D.H. Oughton, personal communication.

Raman Spectroscopy: Introduction

Javier Laserna
University of Málaga, Málaga, Spain

Raman spectroscopy comprises the family of spectral measurements made on molecular media based on inelastic scattering of monochromatic radiation. During this process energy is exchanged between the photon and the molecule such that the scattered photon is of higher or lower energy than the incident photon. The difference in energy is made up by a change in the rotational and vibrational energy of the molecule and gives information on its energy levels. From the beginning much of the theoretical and experimental work in Raman spectroscopy was centered on the fundamentals of inelastic scattering and its application for understanding molecular structure. However, as time elapsed Raman spectroscopy became increasingly important for the advancement of chemical measurements. Certainly, Raman spectroscopy has a special significance to the field of analytical chemistry as a whole, not only because of the impact of the technique itself, but also because its development anticipated a revolution in the way analytical measurements were to be made. The revolution was the insertion of powerful physical methods into a discipline that had been primarily pure chemistry.

The genesis of Raman spectroscopy was in the first quarter of the 20th century when the scattering of monochromatic radiation with change of frequency was predicted theoretically by the Austrian quantum physicist Smekal.[1] The scattering of light by various media had long been studied by Rayleigh in 1871, Einstein in 1910 and others, but no change of wavelength had been observed, with the sole exception of certain types of scattering in the X-ray spectral region observed by Compton.[2] With this background, many scientists were surrounding the idea of inelastic scattering, which was first reported in Calcutta by Raman and coworker Krishnan[3] and almost simultaneously by Landsberg and Mandelstam[4] in Moscow in 1928. Two years later, Raman received the Nobel Prize in Physics for the discovery which bears his name since then.

Some historical details of the developments in Raman spectroscopy both from the theoretical and instrumental points of view are summarized in Figure 1. As with many other branches of analytical chemistry, the development of Raman spectroscopy through the years depended largely on the availability of suitable tools, and significant advances have invariably followed the invention of new instruments. This circumstance can be readily recognized by observation of Figure 1. Developments in Raman spectroscopy occurred slowly during the period from 1930 to 1950, with much of the work immediately following its discovery being devoted to fundamental studies. The rich legacy from the efforts in infrared spectroscopy and the Raman work during this period resulted in formalizing a sound model of molecular vibration dynamics, setting the foundation for Raman scattering as a predictive and interpretative class of spectroscopy. Unfortunately, the basic discoveries made during this period were not followed up in chemical measurements until suitable electronic measuring devices were developed much later. When they were available, advances occurred rapidly in all aspects of Raman spectroscopy from data analysis to instrument miniaturization and, in parallel, the capabilities of Raman-based chemical measurements.

The experimental problems of Raman spectroscopy are the low intensity of the inelastic scattering and the much larger intensity of the Rayleigh scattering. This fact has posed several restrictions to the progress of Raman spectroscopy and has configured to a large extent the instrumentation since the beginning of the experimentation to the present. In their earliest experiments Raman and Krishnan used filtered sunlight as a radiation source and detected the Raman lines of some sixty liquids and gases. They observed the scattering light visually, using a set of compensating colored filters to enhance the optical sensitivity. A more definitive spectrum of carbon tetrachloride exhibiting both the Stokes and anti-Stokes lines recorded photographically using 435.83-nm mercury excitation was published in 1929.[5] It is interesting to note that at these early times, the Raman spectra could be obtained with relatively simpler apparatus than those required for infrared measurements. As a result at the end of the 1930s the situation was that those working on infrared group frequency analysis had often to resort to Raman data collections for reference material since the Raman spectra were more extensive and better catalogued than the corresponding infrared data.[6] Nevertheless, the exposure times for photographically recording the Raman spectra were of several hours.

Early experimental work was directed toward improving the radiation sources. The mercury lamp, filtered to give essentially monochromatic radiation from one of the prominent mercury lines, became the standard source during the 1930s. Later, the mercury Toronto arc lamp became the ultimate source since its introduction in 1952.[7] However, the decisive transformation in the quality of Raman spectra came with invention of the laser in 1960,[8] which was shortly applied as a monochromatic source. In fact, in 1962 Porto and Wood[9] reported the use of a pulsed ruby laser for exciting Raman spectra. The advantages of the laser included its capabilities

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

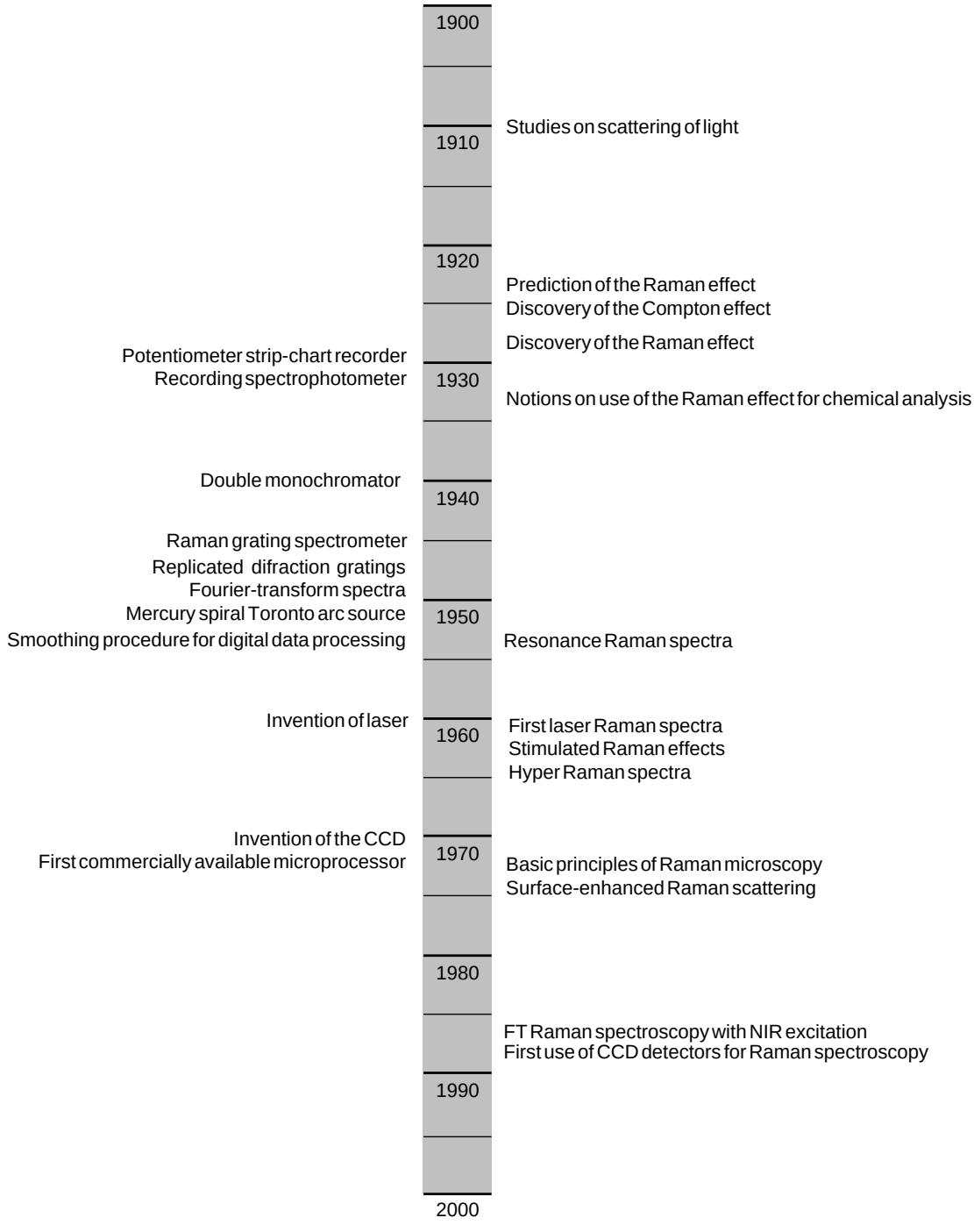

Figure 1 Developments in instrumentation and Raman-related concepts through the 20th century. NIR = near-infrared. (Some data associated with instrumentation have been taken from H.A. Laitinen, G.W. Ewing (eds.), *A History of Analytical Chemistry*, American Chemical Society, Washington, 1977.)

for focusing onto a very small sample, thus enabling excellent spectra to be obtained routinely from materials in short supply. Microgram quantities suffice since then as compared with 10^4–10^6 times these amounts in the pre-laser era. Also, the laser mitigated some problems with stray light, thus improving the detection power and allowing the study of low frequency vibrations. The capabilities of the laser to provide variable wavelength output permitted, in certain circumstances, avoidance of the problems of fluorescence and absorption encountered

with many samples. Although for many years the laser was the weaker component in terms of durability and reliability of the Raman instrument, during the 1970s most Raman instrumentation used Ar^+ lasers of appropriate stability for excitation with the lines of 488 nm and 514.5 nm. Kr^+, He–Ne, cadmium and ruby lasers were also widely employed. Following the demonstration of Fourier transform (FT) Raman spectroscopy in 1986,[10] the use of Nd : YAG lasers operating at 1064 nm has been generalized to decrease the fluorescence level. Optoelectronic devices have progressed dramatically in the past decade as a consequence of major achievements in solid-state technology. As a result compact, efficient, and reliable diode lasers are now available from the visible to the infrared that have been demonstrated to work properly in Raman instruments in combination with suitable filter sets.

Raman spectroscopy always had problems with the high level of elastic scattering, in particular for investigation of lines at short Raman shift. The situation has been handled by fitting the monochromator with two or three dispersion stages. The first commercially available double monochromator incorporated into a spectrophotometer was marketed by 1940 and still today double and triple monochromators are used routinely. These systems may reduce the level of Rayleigh scatter by 10 or more orders of magnitude at Raman shifts of only a few cm^{-1}. The price paid for this earnings has been an increase in the size and price of the instrument and a decrease in the throughput of the optical system. The search for alternatives has resulted in the development of high efficiency holographic notch filters for rejection of Rayleigh light, which make successive dispersion stages unnecessary, thus increasing significantly the luminosity of the Raman experiment.

Although photoelectric devices were available in recording spectrophotometers by World War II, Raman detection for more than three decades was dominated by photography. Advantage was taken of the capability for light integration of photographic emulsions to mitigate the Raman detection problem. During the 1940s and 1950s much progress was made in electronics and reliable photomultipliers were developed. By 1970, high quality photomultiplier tubes were available and were universally used except for high resolution spectroscopy of gases in which photographic recording was still occasionally employed. Photocathode surfaces with efficiencies of 10–20% in the blue and green parts of the spectrum were used. In the red and near-infrared quantum efficiencies were not much better than 1%. To reduce dark current, photomultiplier tubes with small area photocathodes and cooling capabilities were used. Signals were processed in pulse counting mode and digital-to-analog converted for display on chart recorders. This instrument configuration changed with the advent of array detectors. First was the photodiode array and much later the charge-coupled device (CCD). In fact, although the CCD was invented in 1970 by Boyle and Smith,[11] its use as a detector for Raman spectroscopy was first reported in 1987.[12] The bidimensional capabilities of array detectors make simultaneous time-resolved signal acquisition and imaging possible. The cooling technology is now mature enough to ensure extremely long integration times with a very low dark signal. Since a single reading is needed to process the signal accumulated, insignificant levels of the noise from readings and from electronics are produced.

The use of computers in Raman spectroscopy occurred relatively late as compared with other techniques. This served however to gain a vast amount of technology and experience from other fields. The situation is well illustrated with FT techniques, whose evolution has been closely coupled with the development of digital computers. A primitive way of transforming the interferograms in spectra was the direct recording of the interferometer detector output in a voltmeter or ammeter, the measurement of the peak heights by hand and the completion of the necessary computation. It is clear that the mathematical process as originally devised for taking the FT was extremely long and consumed enormous amounts of time. Progress in interferometric spectroscopy was very limited as a result. Although some work was done on analog computers, the situation improved with the use of digital computers. However its generalized use only came with the development of the fast FT, a method which greatly reduces the number of required operations using the symmetry properties due to the equal steps in both the interferogram and the spectrum to achieve the transform. This method was relatively easily made compatible with the processing capabilities of the personal computers. The result was the commercial development of computer-controlled spectrometers capable of rapid acquiring, computing, and displaying high quality and resolution spectral data. When FT Raman was first demonstrated, all these technological barriers were already overcome.

Of particular significance in recent years has been the application of fiber optics in Raman spectroscopy. Fiber optics, integrated optics and microoptics have been brought to a high degree of refinement which, in combination with diode lasers, expand the analytical capabilities of Raman spectroscopy. Fiber optic probes with an ample variety of designs to satisfy the most difficult sampling demands have been described. As a result sophisticated spectroscopic measurements in traditionally inaccessible environments such as high temperature process streams and hazardous sampling locations are now possible.[13]

CURRENT CAPABILITIES OF RAMAN SPECTROSCOPY

At the time of the Raman discovery, analytical chemistry was dealing with problems associated with the analysis of inorganic materials and analytes using wet chemistry methods. Outstanding precision and selectivity levels were achieved in many cases using a depurate technique of sample manipulation. Although the methods of analysis were discussed in terms of accuracy and representivity (anticipating the current discussion of analytical properties and quality control), chemical analysis was considered primarily an instrument of trade.[14] Methods of analysis were carefully chosen on the basis of speed, dependability, and cost. With this panorama, needless to say that little room existed for the use of Raman spectroscopy as an analytical tool. The low efficiency of the inelastic scattering was an added difficulty for the choice of Raman measurements as a viable analytical solution.

The first to recognize that the Raman spectrum could be a powerful resource in chemical analysis was Kohlrausch, who measured the Raman spectra of a wide range of organic compounds as early as in 1931. Later, some qualitative analysis applications were described for monitoring the progress of distillations and for detecting certain chemical groups such as nitriles, conjugated olefins, and aromatics. Soon after the first Raman grating with photoelectric detection was described in 1946 a catalog of Raman spectra of 172 pure hydrocarbons was published. From that time up to the current days of powerful analytical techniques, the growth of the analytical applications of Raman spectroscopy has been spectacular, with capabilities for performing conventional analytical determinations and complex measurement tasks. A wide range of measurement techniques for these purposes are available. Figure 2 shows a schematic diagram of the several faces of modern Raman spectroscopy, including its analytical uses and measurement techniques. It can be used for routine qualitative and quantitative measurements of both inorganic and organic materials, and it is successfully employed to solve complex analytical problems such as determining chemical structures. Gases, vapors, aerosols, liquids and solids can be analyzed. As well as room temperature observations, cryogenic and high temperature measurements can be made, including in situ identification and quantitation of combustion products in flames and plasmas. Raman spectroscopy is one of the few spectrochemical techniques amenable to both laboratory and distance measurements. Open-path optical configurations for remote sensing of atmospheric contaminants are available. Also, Raman excitation and collection using optical fibers in combination with chemical sensors have been adapted for process measurements. Raman spectroscopy can take advantage of the convenience of a range of instrumental configurations, from dispersive to interferometric systems, from monochannel to multichannel detection schemes, and a wide choice of laser systems for convenience of excitation and filter units for spectral purification.

A variety of experimental techniques have been developed. In 1953 the first resonance Raman spectra were reported by Shorigin, and since then the use of resonance enhancement to improve the sensitivity of Raman spectroscopy has become a very popular technique, in particular for those researchers interested in biological problems. Stimulated scattering processes based on high order dielectric susceptibility were reported for the first time in 1962.[15] Several coherent techniques have been developed, of which the most popular is coherent anti-Stokes Raman scattering, which allow the acquisition of high resolution Raman spectra, not limited by the Raman spectrometer but for the line width of the laser line.

Micro-Raman spectroscopy has evolved rapidly since 1966 when it was pointed out that the intensity of Raman light should be independent of sample volume and should remain essentially constant with decreasing sample size down to the dimension determined by the diffraction limit, and hence the wavelength, of the laser excitation.[16] With routine limits of detection in the nanogram range and high molecular selectivity, micro-Raman spectroscopy has now become a major analytical technique of application in both industry and research. Confocal optical designs and imaging capabilities have further extended the power of this approach.

The enormous enhancement of Raman intensity for molecules adsorbed on rough surfaces was first observed in 1974[17] and then explained as a new phenomenon, increasing the Raman cross-section of the adsorbed molecule by 5–6 orders of magnitude.[18,19] The name for this new technique was surface-enhanced Raman spectrometry (SERS). The interest of SERS for analytical purposes resides in its capability to provide information on molecular identity at trace concentration levels.[20]

It is clear from the above that the success of Raman spectroscopy is largely due to successive improvements in experimental techniques and measurement systems. This section is aimed at presenting the current state of the art in instrumentation, assessing the value of established and new instrumental approaches. The section is organized as follows: **Dispersive Raman Spectroscopy, Current Instrumental Designs** is discussed by N. Quy Dao. Progress in the different components of the spectrometer is emphasized. Next, **Fourier Transform Raman Instrumentation** is described by H. Edwards. The importance of the technique for the spectroscopic investigation of interdisciplinary problems and the accessories needed for these purposes are illustrated. The article by J. Andrew is devoted to **Raman Microscopy and Imaging**. Several

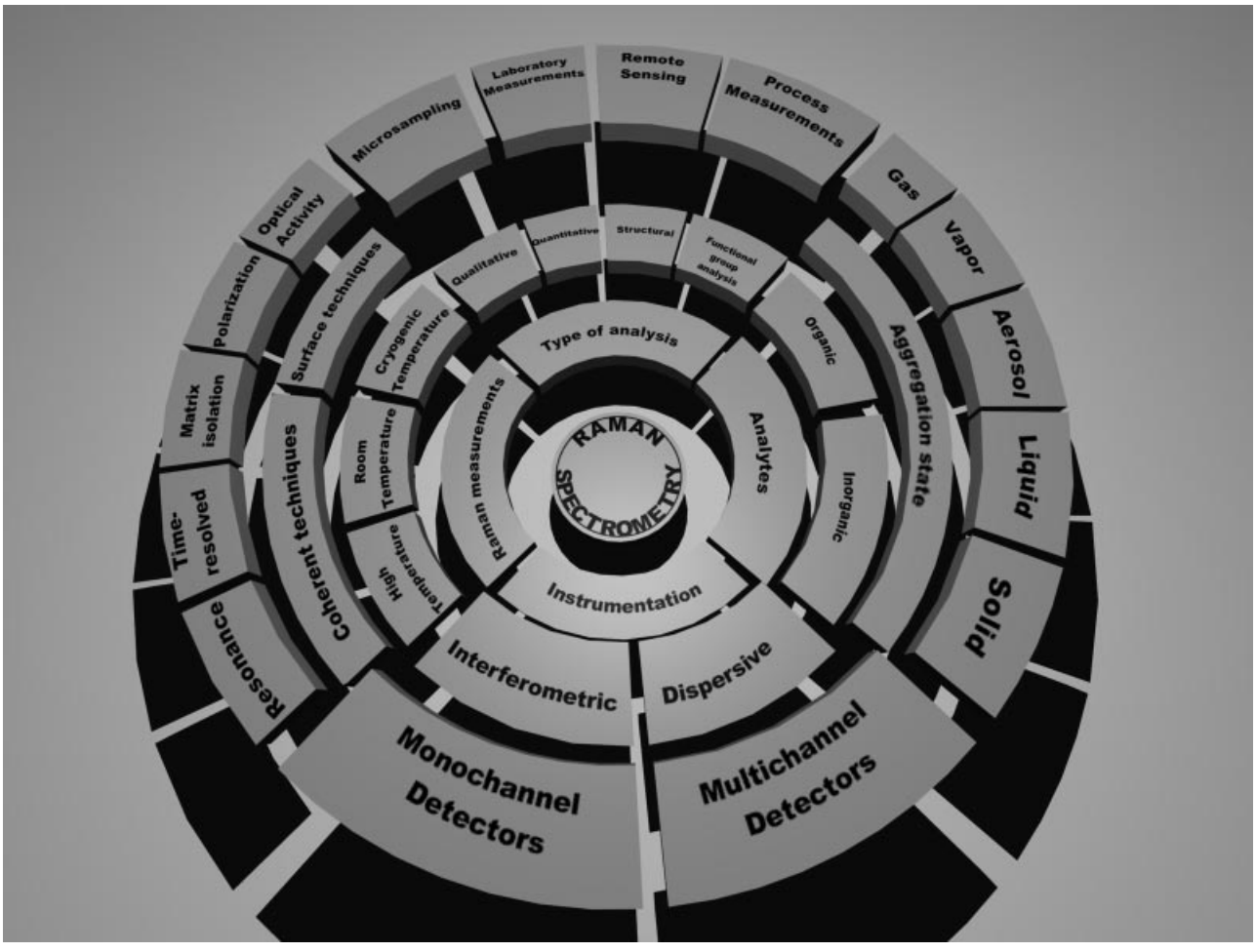

Figure 2 Schematic diagram of the current capabilities of Raman spectroscopy.

approaches for the generation of Raman images are discussed in detail and a comparison is made with different spectral imaging methods. Finally, **Raman Scattering, Fundamentals**, by J. Popp and W. Kiefer, describes linear and nonlinear Raman scattering in both gases and condensed phases.

No attempt has been made to be comprehensive of the broad world of analytical approaches that have derived from the Raman effect. Rather, the section has been designed to give an appraisal of the bases and major means which served to convert Raman spectroscopy in one of the most powerful, versatile and fascinating tools for the investigation of matter.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
FT	Fourier Transform
SERS	Surface-enhanced Raman Spectrometry

REFERENCES

1. A. Smekal, 'The Quantum Theory of Dispersion', *Naturwissenschaften*, **11**, 873 (1923).
2. A. Compton, 'A Quantum Theory of the Scattering of X-rays by Light Elements', *Phys. Rev.*, **21**, 483 (1923).
3. C.V. Raman, K.S. Krishnan, 'The Optical Analog of the Compton Effect', *Nature (London)*, **121**, 711 (1928).
4. G. Landsberg, L. Mandelstam, 'A Novel Effect of Light Scattering in Crystals', *Naturwissenschaften*, **16**, 557 (1928).
5. C.V. Raman, K.S. Khrishnan, 'The Production of New Radiations by Light Scattering', *Proc. R. Soc. (London)*, **122**, 23 (1929).
6. N.M. Jones, 'Analytical Applications of Vibrational Spectroscopy – a Historical Review', *Eur. Spectrosc. News*, **72**, 10 (1987).
7. H.L. Welsh, M.F. Crawford, T.R. Thomas, G.R. Love, 'Raman Spectroscopy of Low-pressure Gases and Vapors', *Can. J. Phys.*, **30**, 77 (1952).

8. T.H. Maiman, 'Stimulated Optical Radiation in Ruby', *Nature (London)*, **187**, 493 (1960).
9. S.P.S. Porto, D.L. Wood, 'Ruby Optical Maser as a Raman Source', *J. Opt. Soc. Am.*, **52**, 251 (1962).
10. T. Hirschfeld, D.B. Chase, 'FT–Raman Spectroscopy: Development and Justification', *Appl. Spectrosc.*, **40**, 133 (1986).
11. W.S. Boyle, G.E. Smith, 'Charge Coupled Semiconductor Devices', *Bell System Tech. J.*, **49**, 587 (1970).
12. S.B. Dierker, C.A. Murray, J.D. Legrange, N.E. Schlotter, 'Characterization of Order in Langmuir–Blodgett Monolayers by Unenhanced Raman Spectroscopy', *Chem. Phys. Lett.*, **137**, 453 (1987).
13. S.M. Angel, T.F. Cooney, H.T. Skinner, 'Applications of Fiber Optics in NIR Raman Spectroscopy', in *Modern Techniques in Raman Spectroscopy*, ed. J.J. Laserna, Wiley, Chichester, 1996.
14. G.E.F. Lundell, 'The Chemical Analysis of Things as They Are', *Ind. Eng. Chem. (Anal. Ed.)*, **5**, 1 (1933).
15. E.J. Woodbury, W.K. Ng, 'Ruby Laser Operation in the Near IR', *Proc. IRE*, **50**, 2367 (1962).
16. M. Delhaye, M. Migeon, 'Effets Raman', *C.R. Acad. Sci. Paris*, **262**, 1513 (1966).
17. M. Fleischmann, P.J. Hendra, A.J. Mcquillan, 'Raman Spectra of Pyridine Adsorbed at a Silver Electrode', *Chem. Phys. Lett.*, **26**, 163 (1974).
18. D.L. Jeanmaire, R.P. van Duyne, 'Surface Raman Spectroelectrochemistry. Part I. Heterocyclic, Aromatic, and Aliphatic Amines Adsorbed on the Anodized Silver Electrode', *J. Electroanal. Chem.*, **84**, 1 (1977).
19. M.G. Albrecht, J.A. Creighton, 'Anomalously Intense Raman Spectra of Pyridine at a Silver Electrode', *J. Am. Chem. Soc.*, **99**, 5215 (1977).
20. J.J. Laserna, 'Combining Fingerprinting Capability with Trace Analytical Detection: Surface-enhanced Raman Spectrometry', *Anal. Chim. Acta*, **283**, 607 (1993).

Dispersive Raman Spectroscopy, Current Instrumental Designs

Nguyen Quy Dao
Ecole Centrale de Paris, Châtenay-Malabry Cedex, France

1 Introduction **2**
2 Theory and Operating Principles **2**
2.1 Physical Principles of Raman Diffusion 2
2.2 Some Early Raman Instruments 3
3 Raman Sources **3**
3.1 Physical Principles of Lasers 3
3.2 Different Types of Laser 3
3.3 Choice of Lasers for Raman Spectroscopy 7
3.4 Precautions in the Manipulation of Laser Radiation 7
4 Detectors **9**
4.1 Photomultiplier Tubes 9
4.2 Multichannel Detectors 10
4.3 Different Types of Noise in Detectors 11
4.4 Choice of Detector Type 13
5 Monochromators **14**
5.1 Filters 14
5.2 Gratings 14
5.3 Various Monochromator Designs 17
6 Raman Devices **19**
6.1 Spectrometers and Spectrographs 19
6.2 Dispersive Raman Microscopes and Raman Imaging 21
6.3 Specific Problems Encountered in Monochromators 21
6.4 Influence of Excitation Wavelength 24
7 Sampling Methods in Raman Spectroscopy **26**
7.1 Sampling Configurations 26
7.2 Polarization Studies 26
7.3 Coupling with Optical Fibers 28
8 Computer Tasks **28**
9 Perspective and Future Developments **29**
Acknowledgments **33**
Abbreviations and Acronyms **33**
Related Articles **33**
References **33**

Raman spectroscopy (RS) is based on an incoherent inelastic scattering process of light by matter, in either its solid, liquid, or gas state. This interaction between ultraviolet (UV), visible (VIS) or near-infrared (NIR) photons with matter gives rise to a Raman spectrum characterizing its different vibrational transitions. These vibrational transitions may be active for the infrared (IR) absorption or the Raman scattering processes, or for both of them, depending on their selection rules and so IR frequencies may be present or absent in the Raman spectrum of a compound and vice versa. IR and RS are therefore closely related.

It is noteworthy that the first Raman spectrometer used by Sir C.V. Raman in 1928 was a dispersive Raman spectrometer (DRS), composed of a mercury lamp when the sun itself was not the exciting source, with a prism monochromator (MC) as the light dispersor and a photographic film as detector. However, it must be emphasized that the Raman effect is extremely weak, and a good Raman spectrum was very difficult to obtain in a reasonable recording time. Progress in the fields of excitation sources, MCs, and of detectors has been constant.

Lasers (mainly gas and solid state) are now currently used as sources. Thanks to the considerable improvements in the other components of a Raman spectrometer, the output power of the laser can now be quite low: a few milliwatts for the excitation light is all that is needed to obtain good Raman signals.

As for MCs, the already highly efficient ruled diffraction gratings (RDGs) were followed by the excellent holographic diffraction gratings (HDGs) used in DRSs. Fourier transform (FT) principles have been also developed, giving birth to FT-Raman spectrometers. Hadamard transform principles have also been used to develop new instruments and Hadamard spectrometers are now being designed.

The ever-present photomultiplier tube (PMT) for the VIS range and semiconductors for the NIR region are largely used as detectors, especially for monochannel and FT-Raman detection. Multichannel detectors with very high quantum efficiency (QE) have appeared on the market opening up new avenues for use of dispersive Raman spectroscopy: High speed and/or compact Raman spectrometers, Raman imaging spectrometers, Raman microscopes, and time-resolved Raman spectrometers are now available using dispersive Raman spectroscopy principles.

The Raman technique is one of the most elegant analytical tools:

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

- *A Raman spectrum gives rich structural information, useful for both fundamental and applied research.*
- *All the states of matter can be studied, whether solid, liquid, gas, glass, or solutions in water and in organic solvents.*
- *There are no sampling problems since a raw material can be studied without any preparation of the sample.*
- *The quantity needed for an examination may be very small: a few microliters in size, a few milligrams to micrograms in weight, or some tens of parts per million in a mixture;*
- *Several derived techniques such as resonance Raman spectroscopy (RRS), time-resolved Raman spectroscopy (TRRS), surface-enhanced Raman spectroscopy (SERS), coherent anti-Stokes Raman Spectroscopy (CARS), hyper-Raman effect, constitute a large panel of methods, where the same instrumentation can sometimes be used giving complementary results.*
- *Raman microscopes and Raman imaging spectrometers are also now available, extending the field of exploration to new domains.*
- *Optical fibers used to transport the excitation and Raman light make possible both remote and in situ analyses. These instruments are now used in industry.*

The present article deals with Dispersive RS and current instrumental designs, but does not tackle other techniques such as the FT-Raman and Raman imaging which are treated in other articles in this Encyclopedia.

1 INTRODUCTION

RS results from the Compton effect, which implies an incoherent diffusion light process. Although the Raman effect was discovered in 1928[1–3] and largely studied in research laboratories during the following decade, the technique was then forgotten for about two decades because Raman measurements were delicate and time-consuming owing to the weakness of the Raman signals. In order to record a good Raman signal, the different components of the Raman spectrometer must be very powerful.

DRSs were the first to be built and remain the most common. They are usually composed of a source, a sample compartment, a MC, a detector, and a dedicated microcomputer.

RS found new life in the 1960s with the discovery of lasers. Lasers make ideal Raman sources with their emission of intense, monochromatic, and polarized light. Laser radiation with wavelengths ranging from UV to NIR is now also available, making it possible to obtain the Raman effect of numerous new compounds which could not be studied otherwise. Moreover, laser diodes (LD) can also be used now as Raman sources.

In the same way, the efficiency of MCs has increased greatly and ghosts and stray light were reduced with the arrival of holographic gratings, while the addition of premonochromators (PMC) (interference filters (IFs) and notch filters (NFs) or gratings) greatly improved resolution. Various MC geometries are now available for recording Raman spectra using either monochannel or multichannel detectors.

With the progress of the electronics industry, the detection of very low light levels has become more accurate. New types of very sensitive detector and especially charge-coupled device (CCD) arrays are now available for different ranges of wavelengths including the very interesting NIR region.

Finally, microcomputers have been added to Raman spectrometers to help with both the collection and the treatment of data thus simplifying the task of obtaining a good Raman spectrum.

A Raman spectrum can now be obtained within a reasonable time (a few seconds). With the latest micro-Raman systems, it is easy to identify microsamples (about 1 μm^3 in volume). TRRS is also available. RRS and SERS are two derived Raman techniques where the presence of traces can be determined at a concentration as low as a few parts per million. New sampling techniques are available and the remote Raman in situ technique utilizing optical fibers to transport the excitation and diffusion light is also being intensively developed.

Progress in instrumentation is constant, and RS is now widely used thanks to its intrinsic advantages and the decreasing cost of instruments, making it a very competitive choice for chemical analyses.

This article is devoted to DRS. Although the different configurations of such an instrument are now well known, various components of DRS have profited from progress in different fields of technology such as optics, electronics, optoelectronics, mechanics, and engineering, giving rise to new Raman sources, detectors, MCs, and sampling techniques. Thanks also to the arrival of powerful microcomputers, several routine tasks can be computer controlled. These are the main subjects treated in the following sections.

2 THEORY AND OPERATING PRINCIPLES

2.1 Physical Principles of Raman Diffusion

The physical principles of RS are treated thoroughly elsewhere in this Encyclopedia, **Raman Scattering, Fundamentals**. Nevertheless, the fact that the Raman effect

is very weak compared to other spectroscopic phenomena such as absorption, fluorescence, or coherent light diffusion needs to be noted. This is the reason why RS remained underdeveloped for so long. With the appearance of lasers, monochromatic, well polarized, and eventually tunable and powerful parallel laser beams constitute a perfect Raman source. Microcomputers also make it possible to control Raman instruments automatically. Most recently, high-performance electronic multichannel detectors[4] have become available. Progress in the making of MCs has also made a giant step. It is now possible to record practically almost all the Raman spectra of compounds that were reputed to be very difficult to obtain. New types of Raman spectrometer such as FT-Raman spectrometers (see **Fourier Transform Raman Instrumentation**) and Hadamard transform Raman spectrometers[5] have also been conceived, bringing new power to RS.

2.2 Some Early Raman Instruments

RS does not need a very sophisticated experimental setup. It was theoretically predicted by Smekal,[6] while the first experimental proof was given by Sir Chandrasekhar Raman, who employed a very simple DRS with a mercury vapor lamp as its excitation source, a prism as its MC and a photographic plate as its detector. However, to obtain a Raman spectrum, several days were necessary. The first Raman spectrum was obtained with a multichannel instrument!

3 RAMAN SOURCES

3.1 Physical Principles of Lasers

The word laser stands for light amplification by stimulated emission (SE) of radiation. Lasers are based on three physical principles: SE, population inversion, and optical resonance amplification.[7]

In contrast to spontaneous emission, where a collection of atoms or molecules in their excited states can randomly emit photons from the excited state to a final state, SE is produced when these atoms or molecules emit their photons prematurely after being hit by photons which have precisely the same energy. SE occurs preferentially in the direction of the applied beam and is perfectly coherent with the primary beam. In practice, a spontaneous emission photon is generated first. This photon releases an avalanche of photons of identical energy by the SE process. The substance in which SE can take place is called an active medium. Usually, the population of excited atoms or molecules at equilibrium is low under normal conditions by virtue of Boltzmann's law and therefore no SE occurs. It is necessary to raise this population to an appropriate level by optical pumping. This operation can be realized using an electrical discharge in the case of gas lasers or a flashlamp or a diode laser for solid-state lasers. In order to amplify the intensity of the beam, a Fabry–Pérot resonance cavity (FPC) is used. This component has the role of amplifying the stimulated beam in the direction of the resonator axis using two reflective mirrors in precise alignment. While all the spontaneous beams which are off-axis are lost, only the beam on-axis is amplified, thus providing a parallel monochromatic laser beam in this direction. The FPC is so adjusted that only the transverse electromagnetic mode (TEM_{00}) can oscillate. Figure 1 illustrates the main components of a laser.

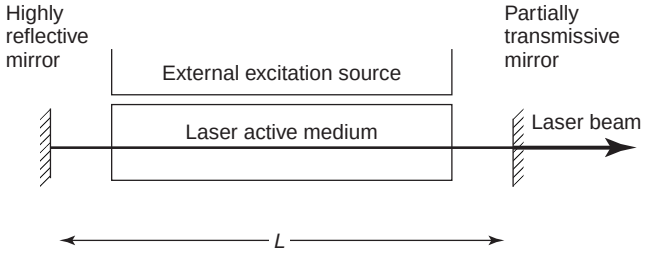

Figure 1 Main components of a laser. The optical distance between the two mirrors of a FPC must be adjusted to the resonance condition $L = K\lambda/2$.

A laser beam is characterized by wavelength or frequency, the power, width, and dispersion, and the monochromaticity. It is broadened by various physical processes. As the intensity distribution of the beam is practically gaussian (Figure 2), the dimension of the beam is defined by its radius taken at $1/e^2$ of its maximum. The beam so limited contains 95% of the total energy of the initial beam. The breadth of the distribution at $I_0/2$ is defined as the full width at half-maximum (fwhm) also called linewidth. The monochromaticity is defined by the Q-factor, Equation (1):

$$Q = \frac{\nu}{\Delta\nu} \tag{1}$$

where ν is the laser frequency and $\Delta\nu$ is its linewidth.

3.2 Different Types of Laser

3.2.1 Gas Lasers

Rare gas lasers are still the most commonly used in RS, in particular He–Ne, Ar^+ and Kr^+ lasers. In some cases, mixed Ar^+/Kr^+ lasers have also been built in order to obtain more laser lines in the same instrument. In the case of He–Ne lasers, the active medium is Ne while excited He atoms transfer their energy to Ne atoms through collision; the existence of various excited energy levels of

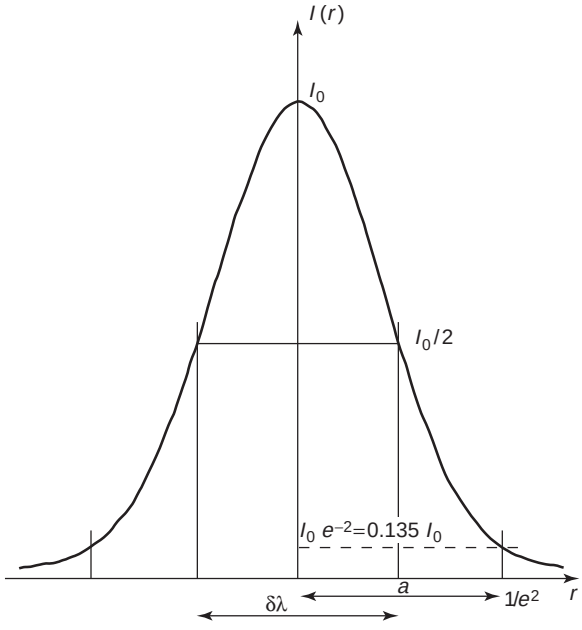

Figure 2 Gaussian intensity distribution $I(r) = I_0 e^{-2}(r/a)^2$, where a = radius of the laser beam, containing 95% of the total energy of the beam and $\delta\lambda$ = fwhm.

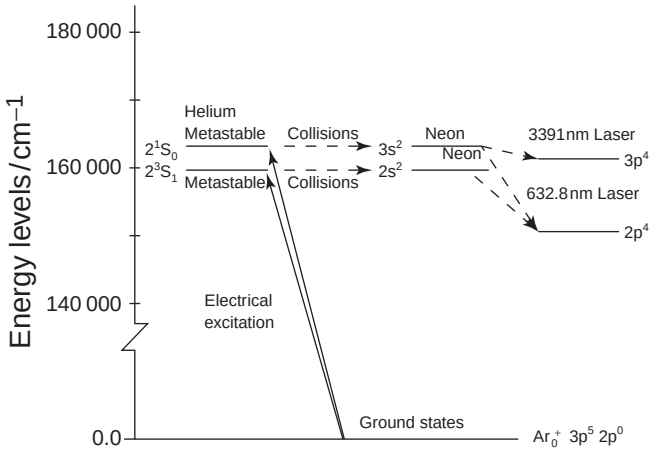

Figure 3 Helium–neon energy level diagram and laser transitions.

Ne gives rise to laser lines by transition to lower levels as shown in Figure 3. These are continuous wave (CW) lasers and the red line at 632.8 nm is routinely used in RS.

In ionized Ar^+ lasers, neutral argon atoms (Figure 4) are first ionized and thus pass to the electronic state $3p^5\ ^2P_0$ of Ar^+. By continuing the electron excitation, the collisions between Ar^+ ions and the energetic electrons populate the 4s and 4p excited levels. Since the lifetime of the 4s level is shorter than the lifetime of 4p, the population of the 4p level increases, and population inversion occurs. The transitions of the two levels 4p

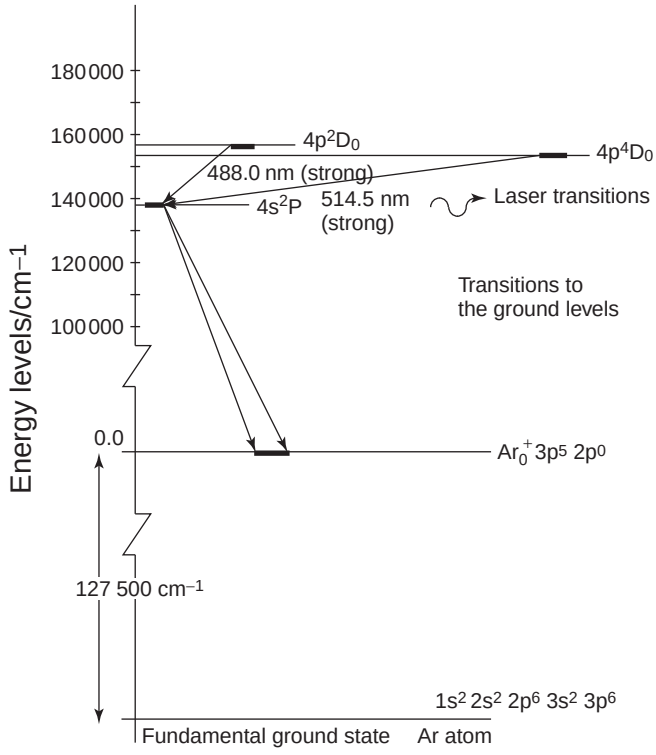

Figure 4 Schematic energy transition of the two strongest laser lines of Ar^+.

$^2D_0$ and 4p $^4D_0$ down to $4s^2\ ^2P$ give rise to the two very strong laser lines of an Ar^+ laser: a blue line at 488.0 nm and and a green line at 514.5 nm. The single line operation is achieved with an intracavity prism assembly. A Brewster angle prism, set between the cavity and the high reflector, is used as MC, allowing only one laser line (Figure 5b). The linewidth is about 5–10 GHz (i.e. 0.16–0.33 cm^{-1}) which is good enough for most routine Raman studies. Two intracavity Brewster angle plates are also added in order to obtain a polarized beam. By tilting the prism, other lines can be obtained. In single-frequency operation, an etalon (which is essentially a bandpass filter) is added and narrower linewidth can be obtained (a few megahertz instead of a few gigahertz) (Figure 5c and d). The Q-factor can be as high as 10^8 in this configuration. These extremely narrow lines are used as sources for high-resolution RS for studying, for example, the rovibrational structures of gas. For broadband operation, the prism is removed (Figure 5a). This can be used as excitation source for dye lasers. The mean beam divergence of a gas laser is about 0.5 mrad.

3.2.2 Liquid Lasers

Liquid lasers are very often tunable lasers, also called dye lasers. Their development is based on the fact that the organic dye solution can absorb strongly in the

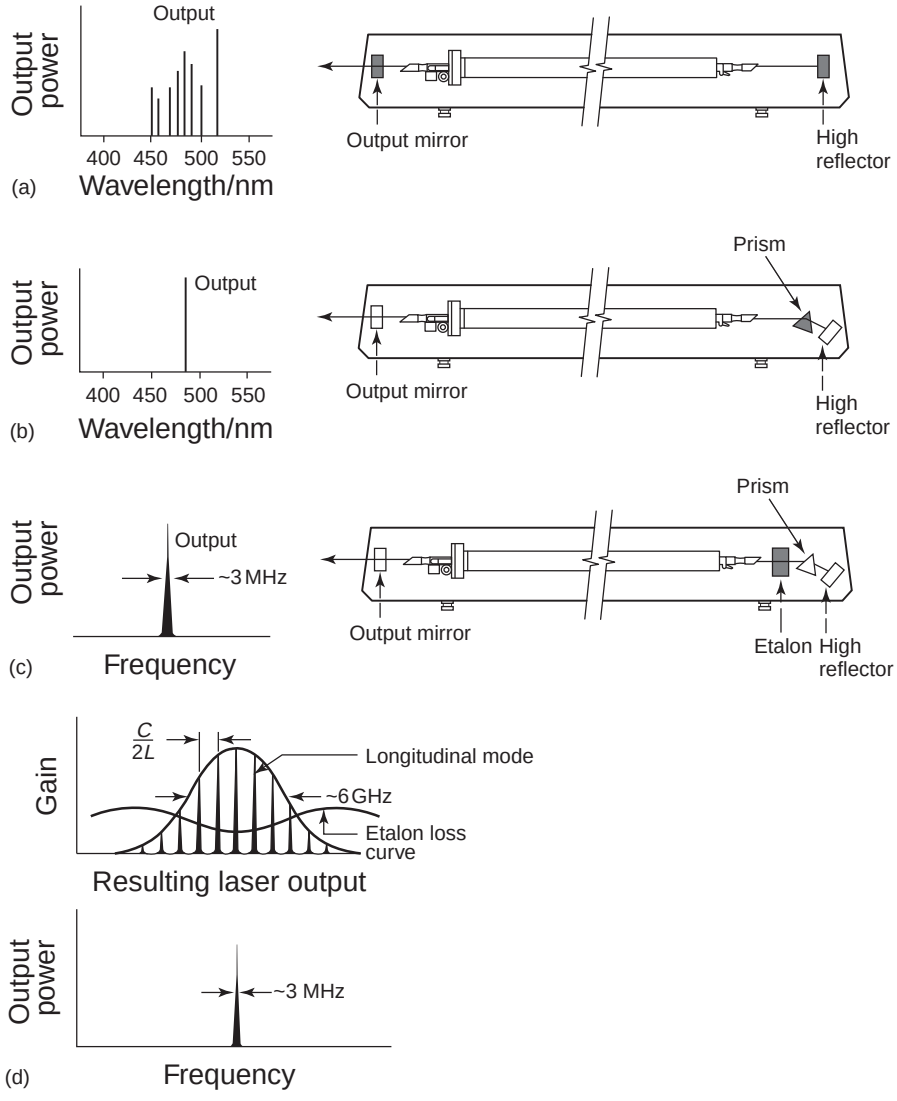

Figure 5 Different line operations. (a) Broadband operation, (b) single-line operation, (c) single-frequency operation, (d) etalon loss minimum tuned to laser gain maximum. (Reproduced by permission of Spectra-Physics Lasers, Inc.)

VIS spectrum and fluoresce in a broad lower frequency range. The dye solution acts as a gain medium. Upon excitation with the intense pump beam (generally a laser line), a population inversion between the ground and excited states occurs. Relaxation to the quasicontinuum of ground state vibrational levels gives rise to an emission of 100 nm or more broad wavelength range. By inserting a wavelength selective element, the dye laser can be tuned over this wavelength range with linewidths down to a few gigahertz. Interest of these lasers is in the possibility of obtaining a precise desired frequency, chosen in the broad fluorescence band of the dye. Depending on the excitation, dye lasers can be pulsed or CW. Such lasers are easy to realise and Figure 6 shows a scheme of a liquid laser cavity. However, with the arrival of tunable solid-state lasers, dye lasers

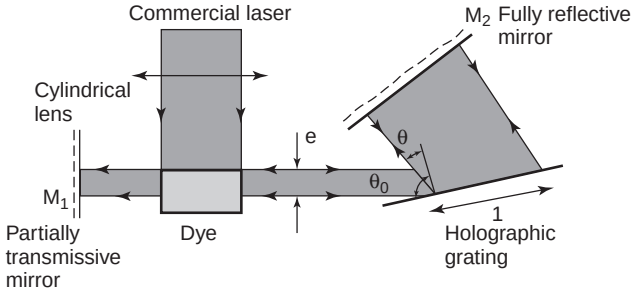

Figure 6 Scheme showing a holographic grating dye cavity for various dyes. θ_0 = flat angle incidence. Laser wavelength λ is determined when θ obeys the diffraction equation: $a(\sin\theta + \sin\theta_0) = k\lambda$.

are less attractive as they are cumbersome and fragile. Dye liquids have a short lifetime and are difficult to

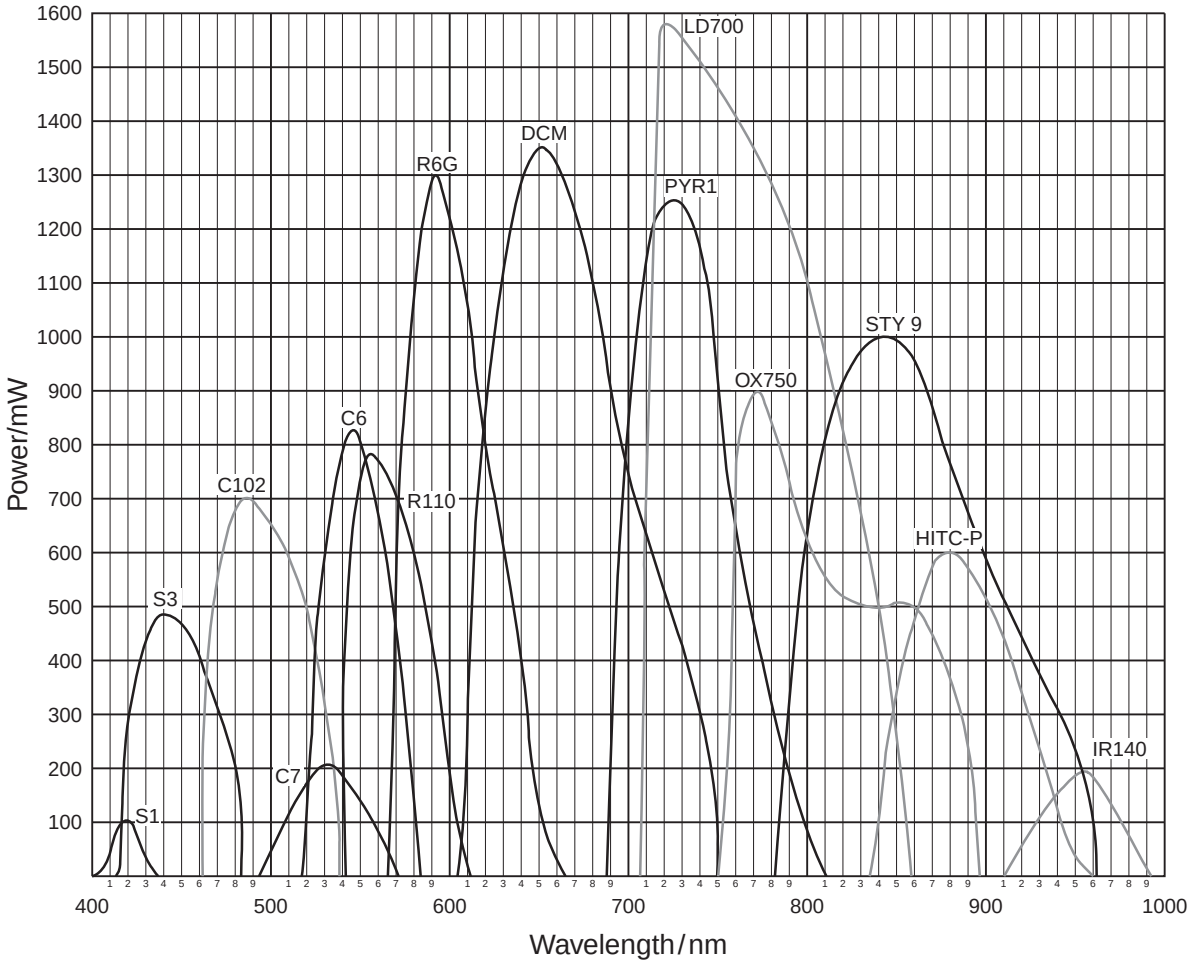

Dye	Pump power (W)	Pump wavelength (nm)	
Stilbene 1 (S1)	1.25	351–364	Argon all lines UV
Stilbene 3 (S3)	2.0	351–364	Argon all lines UV
Coumarin 102 (C102)	2.25	407–423	Krypton all lines violet
Coumarin 7 (C7)	2	476	Argon
Coumarin 6 (C6)	4	488	Argon
Rhodamine 110 (R110)	4	458–514	Argon all lines
Rhodamine 6G (R6G)	4	458–514	Argon all lines
DCM	4	458–514	Argon all lines
Pyridine 1 (PYR1)	7.5	458–514	Argon all lines
LD 700	4	647 and 676	Krypton all lines red
Oxazine 750 (OX750)	4	647 and 676	Krypton all lines red
Styryl 9 (STY 9)	7.5	458–514	Argon all lines
HITC-P	4	647 and 676	Krypton all lines red
IR 140	1.6	752 and 799	Krypton all lines IR

Figure 7 Principal tuning curves of various dyes. (Reproduced by permission of Spectra-Physics Lasers, Inc.)

handle. The change of liquid dyes is a time-consuming operation.

Figure 7 gives the principal tuning curves of various dyes excited with either an Ar^+ (solid curves) or a Kr^+ laser (pale curves).

3.2.3 Solid-state Lasers

Solid-state lasers can be either CW or pulsed. In neodymium lasers, the host lattice for the neodynium ions is either a crystal, usually an yttrium aluminum

garnet crystal (YAG) of formula $Y_3Al_{15}O_{12}$ (Nd : YAG laser), or an amorphous glass (Nd : glass laser). They can be pumped by a flashlamp or more recently, by high-power diode lasers. The principal line of a Nd laser is situated at 1.064 μm in the NIR region. Using frequency doublers or triplers (usually a nonlinear single crystal like potassium dihydrogen phosphate (KDP) or lithium niobate crystals), discrete wavelengths at 0.532, 0.355, and 0.266 μm, respectively are obtained and can be used for excitation in the VIS or UV ranges. Moreover, when associated with dye lasers, various frequency ranges are obtained (0.548–0.580 μm and 0.274–0.290 μm).

Other solid-state lasers also have some specific applications in RS.

3.2.4 Semiconductor Lasers and Diode Lasers

Semiconductor lasers are based on the fact that electrons from a conduction band can be transmitted down to the stable valence band by emitting laser radiation. As the transition energy is rather small, laser light is situated in the IR region. The resonant cavity can be constituted by the end faces of the semiconductor crystal. Titanium : sapphire is a robust tunable laser capable of delivering powerful bands in the red and NIR regions. Figure 8 shows the tuning curves of a Ti : sapphire CW tunable laser.

Diode lasers, based on the same principles, are small in size and robust. These lasers can be stabilized giving good wavelength and intensity reproducibility. However, the linewidth is large (about 10 MHz) and the beam divergence is high (up to 10°). The output frequency and amplitude are a complicated function of driving current and temperature at the p–n junction, and mode-hopping can occur when the laser jumps suddenly from the nth to the $(n+1)$th mode of the cavity while tilting the grating or when the temperature varies.

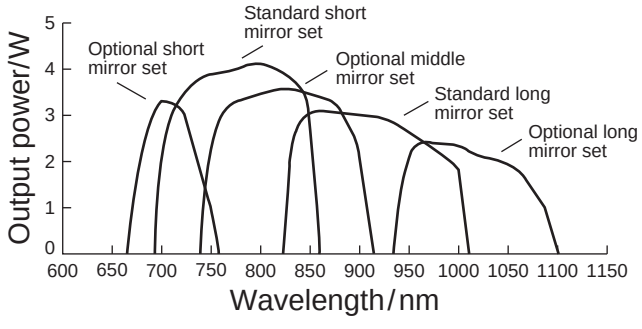

Figure 8 Tuning curves of Ti : sapphire CW tunable laser. (Reproduced by permission of Spectra-Physics Lasers, Inc.)

3.3 Choice of Lasers for Raman Spectroscopy

Lasers are now the universal sources used in RS, thanks to their particularly interesting properties: power, monochromaticity, polarization, and the possibility of obtaining excitation wavelengths ranging from UV to NIR through the VIS region. Practically, however, only the CW laser is used in conventional spectroscopy or in microspectroscopy; pulsed lasers are used for the TRRS technique. Until the 1990s, gas lasers were largely used by Raman spectroscopists. Solid-state lasers, thanks to their robustness, reliability, and power, have tended to supplant gas lasers. However, stabilized diode lasers when associated with corrective optics like a PMC are the ideal source for the instruments of tomorrow.

It is important to remember that Raman intensity is approximately proportional to the frequency to the power of four and the detector response is better when the photon energy is high. However, other physical factors, including electronic transitions, fluorescence, and thermal decomposition also depend on the energy of the excitation beam. Therefore, when no special problem is encountered (fluorescence, decomposition of the samples), the use of blue and green laser lines is advised.

As progress has been made in the optics of MCs and in the efficiency of detectors, it is not necessary to have high power to obtain very good Raman spectra. Lasers with only low or medium power output (from a few milliwatts to some tens of milliwatts) are largely sufficient for Raman applications and costs have been greatly reduced consequently. Single-mode diode lasers have also been tested successfully. However, because of the instability of these lasers, mathematical treatment has been applied to correct the recorded spectra.[8]

Finally, when a specific wavelength is needed as for RRS, tunable lasers are recommended.

UV excitation is often reserved for special cases to avoid fluorescence and also in the cases like the observation of the hyper Raman scattering (HRS) effect or studies of excited states in TRRS. Electronic transitions are indeed frequent for this frequency range, and resonance and nonlinear effects are often observed.

Table 1 gives some characteristics of different lasers with discrete laser lines and Table 2 gives the most usual wavelengths for Raman studies. Other usual lines are the 1064 nm and its harmonics at 532 nm and 355 nm of CW and pulsed YAG-Nd lasers.

3.4 Precautions in the Manipulation of Laser Radiation

Laser light is dangerous for the eyes. As it can be very powerful, all coherent parallel reflection such as reflection

Table 1 Characteristics of different lasers

Laser	Type	Wavelength (nm)	Average power for Raman spect. (mW)	Cooling system	Stability	Pol	Beam diameter at $1/e^2$ (nm)	Beam divergence (mrad)	Applications	Price
He–Ne	Gas (CW)	543.5 632.8 1152.3	→40	Water, air	<5%	Vertical >1000:1	1.1	0.75	Routine	Low
Ar^+	Gas (CW)	488.0 514.5	→4000 →4000	Water, air	<1%	Vertical >100:1	1.5	0.5	Routine	Medium
Kr^+	Gas (CW)	647.1 752.5	→4000 →2000	Water, air	<1%	Vertical >100:1	1.8	0.7	Fluorescent samples	Medium
Nd:YAG	Solid (CW +) pulsed	1064 532[a] 355[b]	→2000	Air	<3%	Vertical >100:1	0.15	1.5	Fluorescent samples, TRRS	Medium
Ti:sapphire	Solid, CW + (tunable) pulsed	720–980	→2000	Water	<3%	Horizontal >100:1	0.95	1.0	RRS	High
Diode	Solid	700–900	→500	–	Variable					Very low
Liquid dye	Liquid (tunable, CW + pulsed)		→100	–						Low

[a] With frequency doubler crystal.
[b] With frequency tripler crystal.

Table 2 Output discrete wavelengths of some lasers used in RS and the relative intensity $I(Ar^+)$ of some usual lines of an Ar^+ laser

λ/nm	Laser	$I(Ar^+)$
333.6	Argon	
337.4	Krypton	
350.7	Krypton	
351.1	Argon	
356.4	Krypton	
363.8	Argon	
406.7	Krypton	
413.1	Krypton	
415.4	Krypton	
437.1	Argon	
454.5	Argon	0.05
457.7	Krypton	
457.9	Argon	0.10
461.9	Krypton	
463.4	Krypton	
465.8	Argon	0.06
468.0	Krypton	
472.7	Argon	0.08
476.2	Krypton	
476.5	Argon	0.22
476.6	Krypton	
482.5	Krypton	
484.7	Krypton	
488.0	Argon	0.75
496.5	Argon	0.30
501.7	Argon	0.15
510.5	Copper	
514.5	Argon	1.00
520.8	Krypton	
528.7	Argon	
530.9	Krypton	
543.5	Helium–neon	
568.2	Krypton	
578.2	Copper	
632.8	Helium–neon	
647.1	Krypton	
657.0	Krypton	
676.5	Krypton	
687.1	Krypton	
752.5	Krypton	
799.3	Krypton	
1060	Nd : glass	
1064	Nd : YAG	
1092.3	Argon	
1152.3	Helium–neon	

on glass envelopes or on crystals under investigation can also damage the eyes. Long-term exposure to diffuse reflections on the sample and/or on optical elements should also be avoided. It is necessary to wear appropriate protective glasses especially when working with NIR radiation. Generally, protective glasses with an optical density (OD) of 2 or 3 greatly diminish the risk of damage while it is still possible to observe the impact of the beam on the sample.

4 DETECTORS

The main problem of detection in RS is the very weak photon signal levels of Raman lines. A good detector needs to be highly sensitive, having a fast response time, good detectivity, a wide linear dynamic range, and high QE over a large frequency range. Although Raman spectra occur over a small range of frequencies in the vicinity of the excitation line, it is interesting when a detector possesses a large response curve as the excitation can vary from the UV to the NIR region. There are two different classes of detector which differ in their physical conception and largely in their use in DRS: the monochannel detectors, that is the PMTs, possess only one detecting element and the multichannel detectors, that is the solid-state array detectors, which have several detecting elements.

The performance parameters of a detector can be defined from the transfer characteristic curve (Figure 9), giving the signal output as a function of the optical input.[9,10] The lower part of the curve (dotted line) represents the detection limit (DL) due to the intrinsic noise of the detector. The upper part (curved line) shows the range where the detector is saturated. The middle part is the linear working range of the detector.

4.1 Photomultiplier Tubes

PMTs are photoemissive detectors. They are essentially composed of two parts: a photoemissive cathode and an

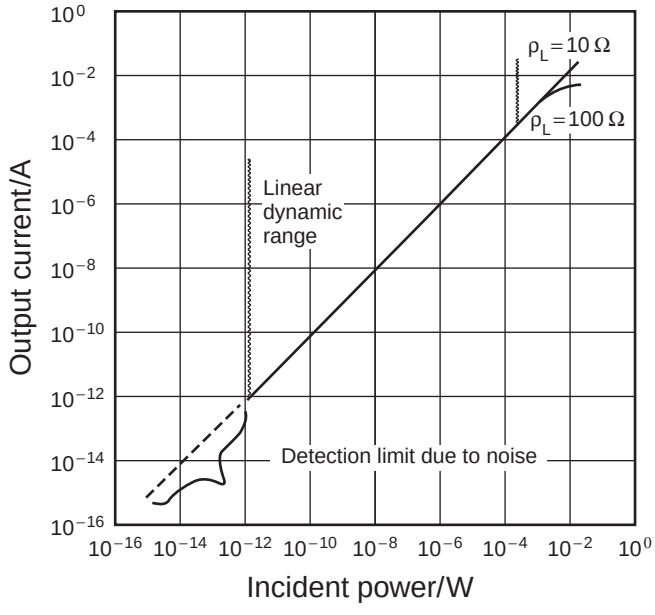

Figure 9 Transfer characteristic curve of a detector. The linear dynamic range is limited at the low level by noise, and at the high level by saturation which depends on the impedance ρ_L of the external circuit. (Reproduced by permission of Oriel Instruments.)

electron multiplier. When photons with sufficient energy hit the photocathode, electrons are ejected. The primary electrons are focused and accelerated by an electrostatic field towards a series of curved electrodes coated with substances able to emit secondary electrons, the dynodes. The primary electrons are then multiplied to become an electron avalanche at the anode. The photoanodic current must be kept below 1 μA to prevent local heating effects and tube fatigue. The electron-multiplying process introduces the least amount of noise and the stability of the detector depends on the fact that the electron beam must follow fairly exact trajectories. Magnetic shields are included in order to increase the stability of the response. Moreover, cooling by Peltier effect diminishes the dark current of the tube. The response time is about 0.5–1 ns. PMTs can be used in a pulsed mode or in direct current (DC). The responsivity as a function of the wavelength is from 180 to 1100 nm in width, and new PMTs are appearing now up to 1500 nm depending on the model. There are several PMT geometries; Figure 10 shows a plan of an end-on PMT.

4.1.1 Direct Current and Photon Counting Devices

The photoanodic current of a PMT is made up of short pulses. Each pulse results from a train of electrons created at the end of the dynode chain by one electron issuing from the photocathode. With analog methods, the anodic current of the PMT is amplified and an average value of intensity is measured during the integration time. Various types of noise are included in the signal, thus diminishing the signal-to-noise ratio (S/N). In the photon converting device (PC) method, only signals with amplitude and duration limited to a predetermined range are considered; in particular the shot noise (the thermal emission of dynodes) is filtered. Therefore only electrons issuing from the photocathode are counted by a numerical counter. The PC method enables very low photon flux, between 10^2 to 10^6 photons s^{-1} to be measured.

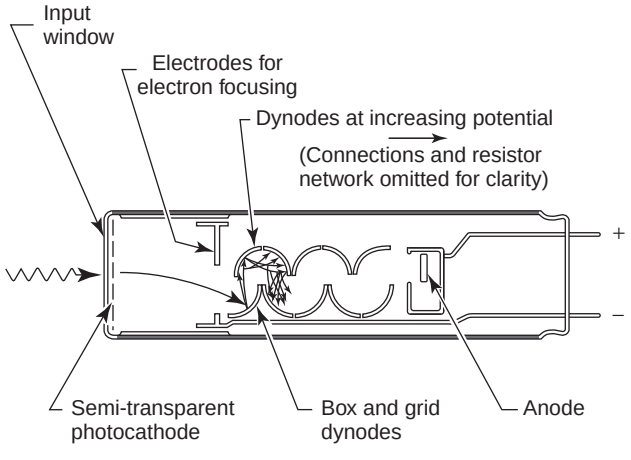

Figure 10 Scheme showing a PMT. (Reproduced by permission of Oriel Instruments.)

4.2 Multichannel Detectors

Multichannel detectors are solid-state detectors based on silicon chip technology. They are divided into two classes: photodiode arrays (PDAs) and CCDs. The major differences lie in the sizes of the sensitive elements (pixels), linear for PDA and two-dimensional arrays for CCD, and in the signal-handling and read-out methods.[11]

PDAs are of linear format. Each pixel has a dimension of $25 \times 2500\,\mu m^2$. PDAs are available in packages with 128, 256, 512 and 1024 pixels. The individual elements respond to the photon flux by producing photocurrents. Those in turn charge individual storage capacitors. The information is given out in a timed sequence. It is important to note that dark current increases with the integration time and it is necessary to know dark current values for each array so that the integration time can be calculated in consequence.

The two-dimensional CCD arrays are built from square elements of $25 \times 25\,\mu m^2$ pixels with some 50 000–1 000 000 pixels ranged in a rectangular array.[12]

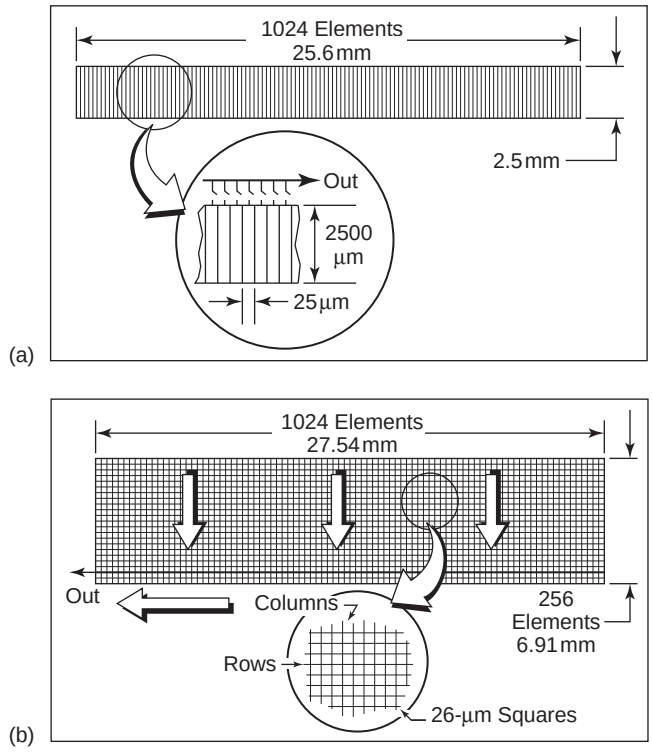

Figure 11 Layout of multichannel detectors. (a) PDA detector and (b) spectroscopic CCD. (Reproduced by permission of Oriel Instruments.)

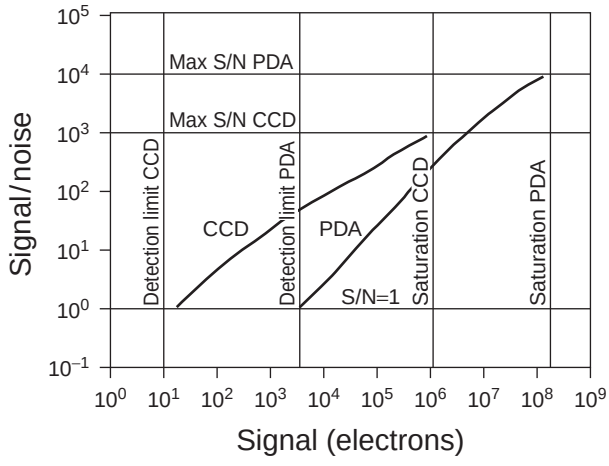

Figure 12 Comparative S/Ns of PDA and CCD detectors. (Reproduced by permission of Princeton Instruments Inc.)

When the pixels of the columns are binned, that is when they are grouped together to form a linear array of tall thin elements, CCD arrays function as linear detectors, with enhanced performances.[13] Figure 11 shows the layout of a PDA and a CCD detector, Figure 12 shows comparative S/Ns as a function of the input signal for the PDA and CCD detectors and Table 3 gives the comparison of PDA and CCD specifications (at 600 nm) as an example. Various CCDs have been built giving good responses for the different ranges of wavelength. It is now possible to obtain excellent CCDs for a working range from the UV to the NIR regions but prices are still high.[14] Figure 13 gives the QE of some PDAs and CCDs as a function of the wavelength range. Typically, CCDs have a sensitivity similar to PMTs but, independently of their multichannel detection possibility, they are not damaged by overexposure to bright light. On a CCD, each pixel is overlaid with a small voltage-carrying element, an electrode. During the illumination of the chip, charges are stored in the pixel well. To collect the data, a sequence of voltages is applied across the electrodes to move the charge row by row down into a shift register at the bottom of the array. This charge is then moved by the same means horizontally to the output node where it is converted to a digital form for processing. This readout method gives extremely low read-out noise, and good separation between signal channels (Figure 14).

CCDs are now largely used for multiplex, fast RS and Raman imaging. The very high QE of the silicon is not obtained practically on the commercial CCD, this is due to the different layers and ''drains'' and the technology and the manufacture of the chips. But some manufacturers propose use of a ''back thinned'' CCD: the CCD is thinned chemically and irradiated directly on the silicon on the back side of the detector. This detector is near the theoretical QE.

4.3 Different Types of Noise in Detectors

When choosing a detector, it is particularly important to know the physical origins of noise in order to avoid it whenever possible. Noise can be extrinsic or intrinsic to the measurement system.[15]

Extrinsic noise involves the transfer of energy from the surroundings to the measurement system and occurs usually in specific frequency ranges. Electric and magnetic fields, radiant energy, mechanical vibration, and cosmic rays can create perturbation. Shielding and grounding the instruments minimizes this kind of noise.

Table 3 Comparison of PDA and CCD specifications (at 600 nm)

	PDA	CCD
Pixel size (binned)	$25 \times 2500\,\mu m^2$	$27 \times 27\,\mu m$ ($27 \times 6912\,\mu m$)
Saturation level (binned)	125×10^6 e	250×10^3 e (810×10^3 e)
Saturation exposure	$>107\,nJ\,cm^{-2}$	$>250\,pJ\,cm^{-2}$
Dark current (25 °C)	<0.1 pA/pixel $<160\,pA\,cm^{-2}$	<0.1 fA/pixel $<16\,pA\,cm^{-2}$
DL	$<3.3\,pJ\,cm^{-2}$ <3800 e	$<3.8\,fJ\,cm^{-2}$ <13 e
Max S/N	10 000 : 1	900 : 1
Dynamic range	32 768 : 1 or 15 bits	65 536 : 1 or 16 bits
Electrons/count	1900	13
Spectral response	180–1100 nm	180–1100 nm (with UV coating)
Linear spectroscopy and imaging	Yes	Yes
Two-dimensional spectroscopy	No	Yes
Shutter required	No	For two-dimensional spectroscopy and imaging

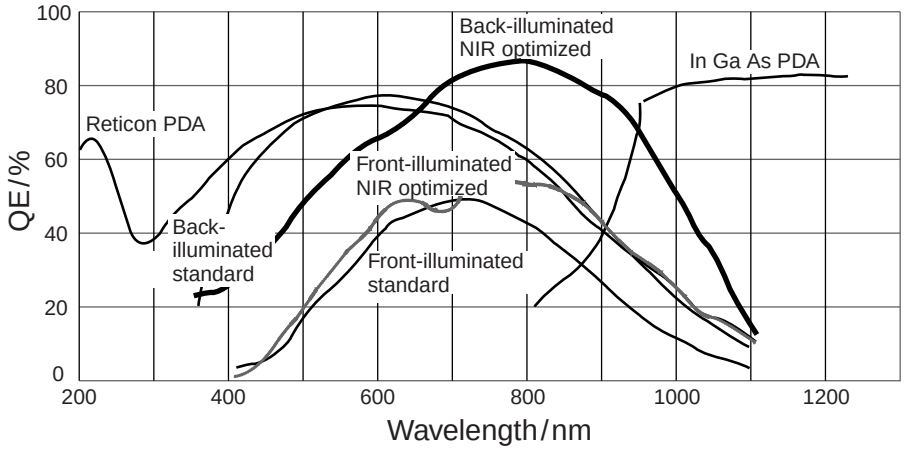

Figure 13 QE of some PDAs and CCDs. (Reproduced by permission of Princeton Instruments Inc.)

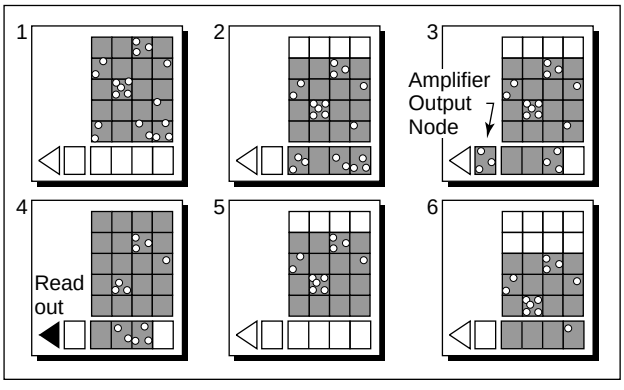

Figure 14 Readout pattern of two-dimensional CCDs. The signal (usually referred to as the charge) is shifted down sequentially and across to the output node. 1–6 show the different steps of the pixels in the same row. (Reproduced by permission of Oriel Instruments.)

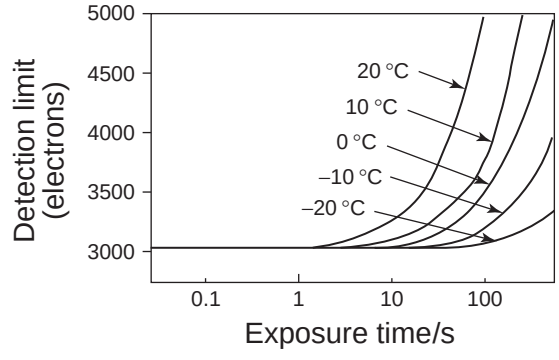

Figure 15 Influence of the temperature on the DL. (Reproduced by permission of Oriel Instruments.)

Intrinsic noise has diverse origins:

- Thermal noise, also known as Nyquist or Johnson noise, is caused by fluctuations caused by the thermal motion of electrons in resistive components. It exists even in the absence of current flow and is represented by Equation (2)

$$V = (4kT\rho\Delta f)^{1/2} \tag{2}$$

where V is the average voltage due to the thermal noise, k the Boltzman constant, T the absolute temperature, ρ the resistance of the electronic device, and Δf the bandwidth of the measured frequencies.

Cooling and narrowing the frequency bandwidth greatly reduce thermal noise. Figure 15 shows as an example the influence of the temperature on the DL of a PDA detector versus exposure time. At $-20\,^{\circ}C$, the DL is reached after 10 s of exposure time while at $20\,^{\circ}C$, noise is much greater and the DL is reached in less than 1 s.

- Shot noise is the statistical fluctuation in the process of photoelectron emission in photoemissive detectors and frequently occurs in PMTs. Generation recombination (GR) noise is a kind of shot noise caused by the statistical fluctuation in the number of charge carriers in semiconductor materials, in which both process generation and recombination create fluctuations. Usually, the magnitude of shot noise is much smaller than that of thermal noise. Shot noise is proportional to the square root of the measurement bandwith Δf and can be reduced by reducing Δf.
- Flicker noise, commonly referred to as $1/f$ noise has two different causes, electrical when associated with potential barriers at electrical contacts and optical where it is due to the fluctuations of the source at low frequency. Flicker noise predominates in measurements from 0 Hz (DC) up to about 300 Hz. It can be eliminated by avoiding the use of low-frequency detection.

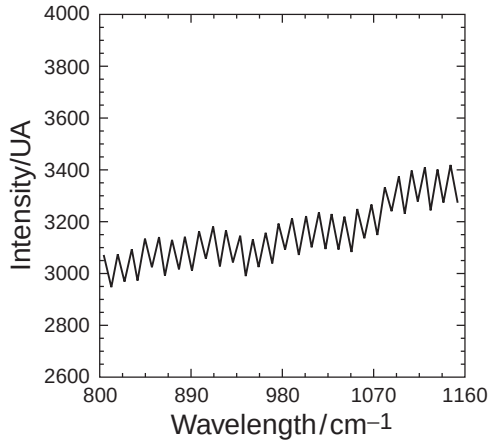

Figure 16 Example of the response curve using a CCD array detector. The saw profile curve is caused by the readout process which is different for the pair pixels and the odd pixels. Corrections can be made using the spectrum of a white source measured under the same conditions.

For solid-state arrays, two other types of noise can also be defined: pixel noise and fixed pattern noise.

- Pixel noise is the noise of an individual pixel and is characterized by the different types of noise already mentioned. It should be noted that this noise increases with the pixel readout rate (two to four times from 50 kHz to 1 MHz) and the temperature.
- Fixed pattern noise is concerned with linear and two-dimensional arrays, where there is always a small difference in the readout values between neighboring pixels. This difference is characteristic of a given array and therefore can be corrected by calibration. Figure 16 shows an example of the response curve using a CCD array where a saw profile curve is observed.

Table 4 sums up the definitions of the different characteristic parameters of a detector.

4.4 Choice of Detector Type

PMT monochannel detection is still largely used in Dispersive RS especially for the investigation of low-frequency ranges necessitating good spectral resolution near the Rayleigh excitation line. This is the case for

Table 4 Definitions of different characteristic parameters of a detector

	Definition	Formula	Unit
Transfer characteristic curve	Diagram representing the output current (Y) versus the incident power or the irradiance of an optical input (X)	Y versus X	A versus W A versus W m^{-2}
Responsivity R	Ability of the detector to measure variations of input power	$R = \dfrac{dY}{dX}$	A W^{-1} or A m^2 W^{-1}
Spectral responsivity, $R(\lambda)$	Responsivity at a given wavelength λ	$R(\lambda) = \dfrac{dY(\lambda)}{dX(\lambda)}$	A m^2 W^{-1}
Total responsivity, R_T	Total output current by total optical flux for the whole wavelength range	$R_T = \dfrac{\int_0^\infty F(\lambda)R(\lambda)\,d(\lambda)}{\int_0^\infty F(\lambda)\,d(\lambda)}$	A W^{-1} or A m^2 W^{-1}
Noise equivalent power (NEP)	Value of X for a given frequency bandwidth i.e. S/N = 1 when Y = RMS noise	$\mathrm{NEP} = \alpha\Delta f/\lambda\eta(\lambda)$	W Hz$^{1/2}$
Detectivity, $D(\lambda)$	Lower performance limit of a detector	$D(\lambda) = 1/\mathrm{NEP}$	Hz$^{1/2}$ W^{-1}
Specific detectivity, D^*	Detectivity power independent of area and bandwidth of the detector	$D^*(\lambda) = D(\lambda)(A\Delta f)^{1/2}$	cm Hz W^{-1}
Dark current Δ_0 and leakage current	Residual signal observed by the detector in absence of light (decreases with temperature)		A
QE	Probability of a single photon producing a detectable photoelectron		%
Dynamic range	Ratio of the largest signal in the linear range to the RMS noise (at a given pixel rate)		
Saturation charge (spectral well capacity)	Largest signal that can be measured for a pixel, owing to its limited well depth		Electrons or coulomb[a]

α = proportionality coefficient, Δf = frequency response bandwidth, $\eta(\lambda)$ = QE at wavelength λ, $F(\lambda)$ = incident monochromatic optical flux at wavelength λ, A = detector sensitive area, RMS, root mean square.

[a] 1 electron charge = 1.602×10^{-19} coulomb.

solid-state physics, for studies of rovibrational structure of gases, and other fine structure studies using PMTs that are particularly efficient in the VIS range. However, for the NIR range, in particular with a 1064-nm excitation of a Nd : YAG laser, solid-state detectors such as Ge or InGaAs are now preferred. This is often the case for FT-Raman spectrometers.

For routine control or studies not requiring very high resolution but where speed and robustness are priorities, solid-state arrays are greatly preferable. PDAs were used as multichannel detectors in the 1980s but have now been replaced advantageously by CCD detectors despite their somewhat greater cost. Commercial multichannel Raman spectrometers are now almost exclusively equipped with CCDs. As the response of these detectors is extended well into the NIR region, NIR dispersive multichannel Raman spectrometers are being developed concurrently with NIR/FT Raman spectrometers. Raman imaging is also possible with the NIR dispersive multichannel Raman method but the corresponding resolution is not as good as with VIS light. It is limited by the laws of diffraction. The resolution of NIR Raman imaging will always be superior to about $10\,\mu m^2$.

5 MONOCHROMATORS

A MC is the principal part of a DRS. It is composed of several mechanical and optical components. The following sections describe the optical components, that is the filters and the gratings.

5.1 Filters

Filters are often used in optical devices to cut off low- or high-frequency range (long-wave or short-wave pass filter) or to transmit or to cut off a very narrow spectral range (bandpass filter or NF). Generally, a filter correctly used on parallel light gives a high throughput, up to 80% of the incident light.

Absorption filters, long- or short-wave filters can be made of colored glass and are used to eliminate different orders of a grating. They have now been advantageously replaced by IFs.

There are two kinds of IF differing in their manufacturing modes: dielectric multilayer filters (DMFs) and holographic filters (HFs). They can be either edge filters, bandpass filters, or NFs. DMFs are based on the interference phenomenon and are formed by several tens of dielectric film layers. The reflection of the light is assured by the refraction index variation from one layer (high index) to the next (low index), each $\lambda/4$ thick. The relatively high reflective index has a square wave profile in a relatively thin layer (ca. $20\,\mu m$). HFs are made using holographic techniques, which greatly improves the IF quality. They have a lower average reflective index with a modulation which varies sinusoidally throughout a thick film ($>50\,\mu m$). When a polychromatic beam hits the filter, interference happens and only the given wavelength beam passes through the filter. For the holographic notch filters (HNFs) in which only a desired narrow wavelength band is absorbed, that is, the notch region, HNFs have uniform transmission, over 70% of the incident light. The cut-on/cut-off edges are very sharp. The notch wavelength range can be tuned by varying slightly the incidence angle of the incident beam. Figure 17 shows the transmittance of a HNF and the influence of the incident angle on the shift of the filtered wavelength range. By increasing the incidence angle, the notch range is red shifted. This characteristic is exploited to lower the recording range near the Rayleigh line. HNF introduces no additional polarization at normal incidence and has a very low effect below 20°.[16]

It must be noted that each filter works for only one predetermined wavelength and that it is necessary to change the HNF when the Raman excitation is different.

HNFs have very interesting applications in the Raman remote sensor head (see later) and as a very efficient filter element in PMC. Figure 18 indicates a possible set-up for a HNF in the PMC of a Raman spectrograph. HNF plays a double role: reflecting the laser line to the sample and rejecting the Rayleigh line and stray light, while allowing only the Raman spectrum to pass towards the MC stage. As it is a very efficient and narrow bandpass filter, Stokes and anti-Stokes Raman lines can be recorded accordingly. To tune the notch region, spacers can be introduced to change the angle of the NF.

The advantages of using HNFs are that the Raman spectrographs are more compact, have better throughput than multistage grating systems, and can be less expensive. Problems of polarization are also minimized. The drawback to using filters is the fluorescence due to the gelatin grains when the laser light hits the filter. To minimize this effect, the filter is placed in between two afocal optics and works by transmission.[17] It is worth noting that another kind of rejection filter has been successfully applied to filter the Rayleigh line.[18,19] This filter is based on a crystalline colloidal array of very small polystyrene spheres with which the Rayleigh light is diffracted according to Bragg's Laws. This filter is not yet commercially available.

5.2 Gratings

Diffraction gratings constitute the principal part of a MC. A grating consists of a regular array of parallel equidistant "reflective" lines (Figure 19).

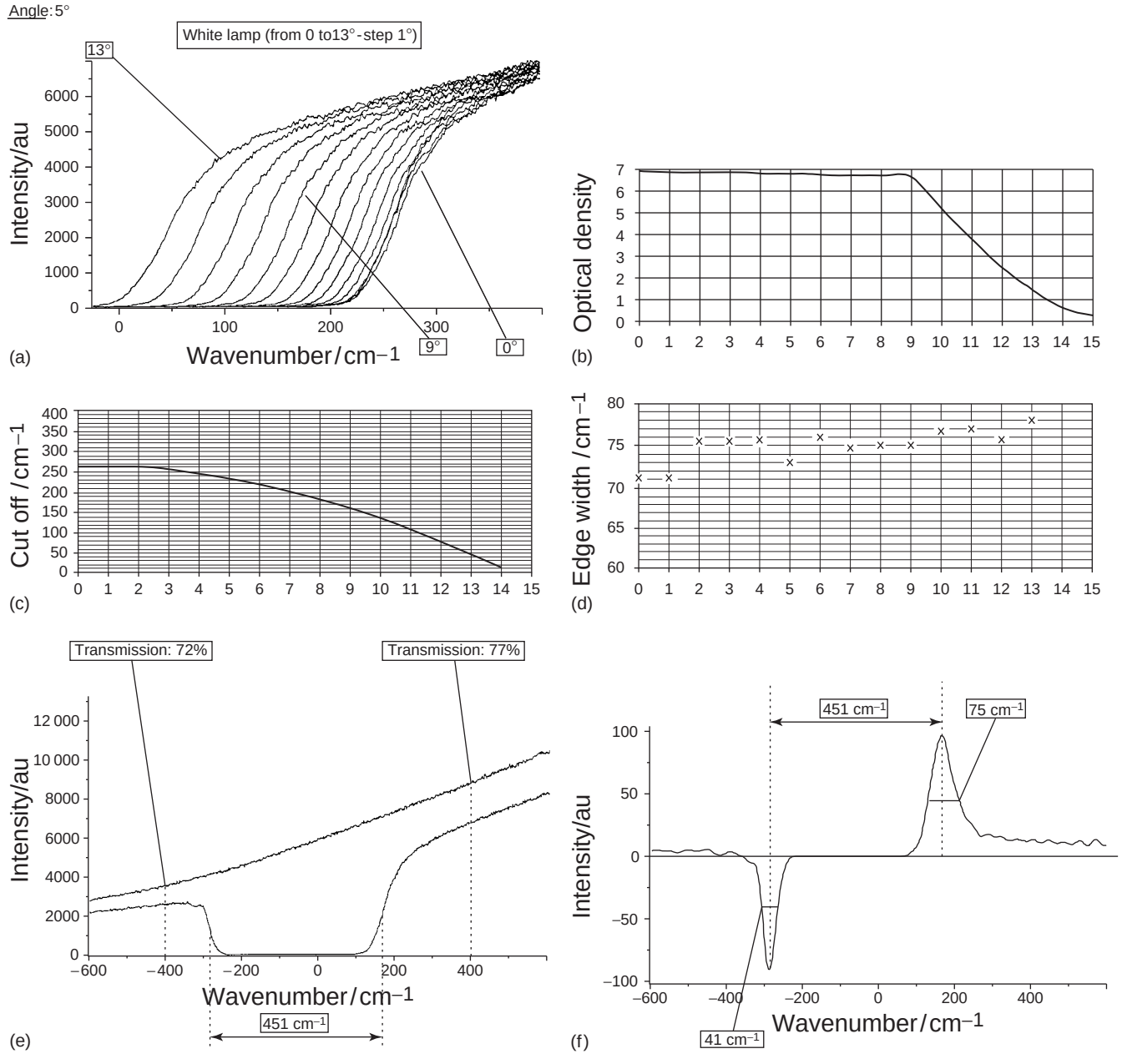

Figure 17 The working conditions of a HNF. (a) By varying the incidence beam angle, the cut-off frequency can be tuned for more than 200 cm^{-1}. (b) From 0° to 9°, the OD is higher than 6, allowing a good elimination of the Rayleigh line. (c) Cut-off position as a function of the incidence beam angle. For 9° incidence, the cut-off position is at 170 cm^{-1}. (d) Spectral edge width on the Stokes side. (e) Characteristic curve of a HNF for the optimum angle (9°), compared to the spectrum of the light source without the HNF (baseline). The transmission is higher than 70%. (f) Definition of the spectral edge width: on the Stokes side: 75 cm^{-1}, on the anti-Stokes side: 41 cm^{-1}. (Reproduced by permission of DILOR.)

These lines can diffract light according to the law of diffraction, Equation (3)

$$\sin i \pm \sin r = \frac{k\lambda}{a} \tag{3}$$

where a is the groove spacing in nm, i is the incidence angle in degrees, r is the diffraction angle in degrees, k is the diffraction order, λ is the incident wavelength in nm and $\pm$ stand for reflection and transmission, respectively.

If n is the groove density or number of grooves per length unit, Equation (4)

$$n = \frac{1}{a} \tag{4}$$

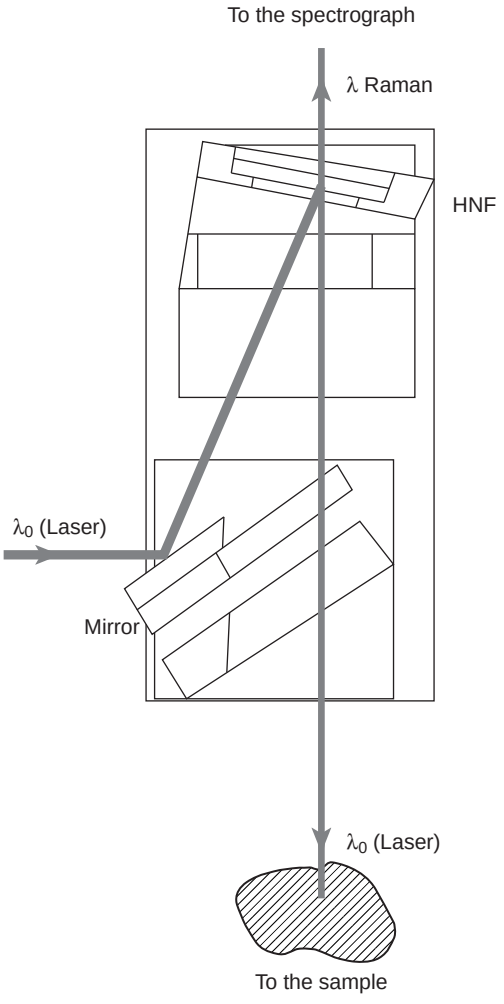

Figure 18 A HNF set-up as PMC in a Raman spectrograph. (DILOR Labram spectrograph, with permission from DILOR.)

which gives Equation (5):

$$\sin i \pm \sin r = kn\lambda \tag{5}$$

i being fixed, and for a given k Equations (3) or (5) give the value of r where a maximum intensity for wavelength λ is found. For $k = 0$, there is no dispersion (zero order). Usually, gratings are used in their first order. In most MCs (see section 5.3), the location of the entrance and exit slits is fixed and the grating rotates around an axis passing through the middle of the grating. The angle D_v, defined by Equation (6),

$$D_v = r - i \tag{6}$$

is an instrumental constant fixed by the constructor. The grating equation can then be written as Equation (7),

$$2\sin\ \frac{i+r}{2}\ \cos\left(\frac{D_v}{2}\right. = kn\lambda \tag{7}$$

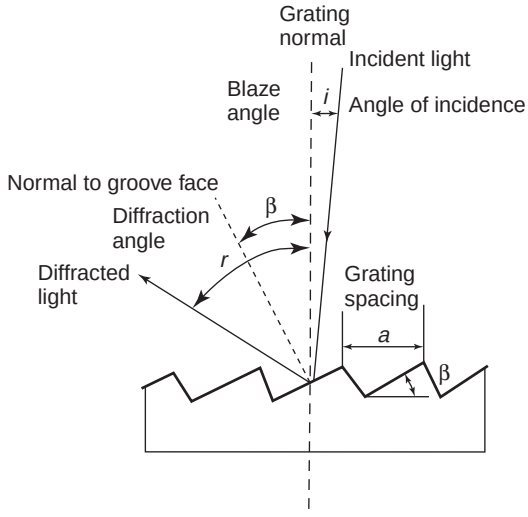

Figure 19 Characteristic angles of a grating. Blaze angle β is the incident angle in which the total energy of the incident light is concentrated in one order.

This formula is used to understand the sinus bar mechanism (see later).

5.2.1 Free Spectral Range

The FSR (free spectral range) is defined as the range of wavelength that is free of overlap from adjacent orders. If v is the wavenumber (Equation 8):

$$v = \frac{1}{\lambda} \tag{8}$$

then Equation (3) could be written (Equations 9–11)

$$v = \frac{kn}{\sin i \pm \sin r} \tag{9}$$

$$v = kv_1 \tag{10}$$

$$\text{with}\ \ v_1 = \frac{n}{\sin i \pm \sin r} \tag{11}$$

For the same angle i and at the same angle of observation r several wavelengths are superimposed in an arithmetic progression $v_1, 2v_1, 3v_1, \ldots$. It can be seen that if the incident light is a continuum of wavelengths, after dispersion by a grating there can be superimposition of other wavelengths $\lambda/2$ in the second order, $\lambda/3$ in the third order, for a given position corresponding to the wavelength $\lambda \ldots$. To prevent the superimposition of the different orders, the diffracted radiation must be contained in a maximum interval called the FSR. To eliminate high-order peaks, edge filters are used.

A grating is characterized by its resolution. The resolution is the ability to separate adjacent spectrum

lines. It is generally defined by Equation (12):

$$R = \frac{\lambda}{\Delta\lambda} \tag{12}$$

$\Delta\lambda$ being the difference in wavelength between two equal intensity spectrum lines that are just separated.

Gratings usually have defects: ghosts and stray light. Ghosts are spurious spectral lines caused by periodic imperfections in the gratings. Stray light comes from two different origins: random nonperiodic vibrations and nonperfect planarity of the reflecting surfaces.

There are two main types of grating distinguished by their fabrication modes. RDGs are made using a ruling engine. A master grating is first built under very rigorous conditions and replicas are reproduced from the master grating. Each grating is individually controlled, and a specification registered certificate giving various characteristics and defects of the grating such as resolution, stray light, and ghosts is delivered. HDGs are made directly by using two monochromatic laser beams producing interference fringes on a photosensitive layer deposited on an optically flat glass. Diffraction gratings can either work by reflection or by transmission.

Efficiency and throughput are the two most important factors in the choice of a grating. Efficiency for various standard gratings varies generally between 50 and 90%. If λ is the center of the spectral region of interest, the maximum efficiency of the grating depends on the parameter λ/a.[20,21]

Gratings are not equally efficient at all wavelengths for various reasons. The efficiency can be tuned by changing the groove facet angles, or shape and depth. The optimization of efficiency by appropriate groove shaping is known as blazing. The blaze wavelength is the wavelength for which the grating is most efficient.

Figure 20 shows a comparison of the efficiency curves of a RDG and a HDG around the blaze wavelength λ_B. In the case where $\lambda/a < 0.2$, which is generally the case in the far-UV domain, the absolute efficiency of the RDG and the HDG drops sharply because of the reflecting factor of the coating material. The efficiencies are comparable for the two types of grating.

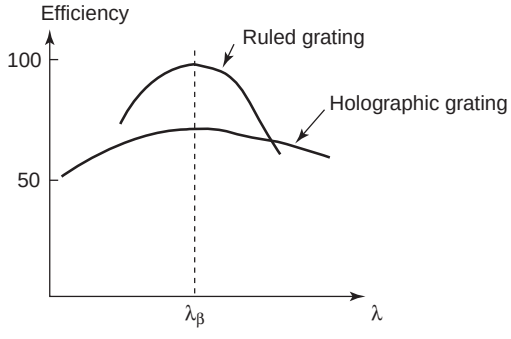

Figure 20 Efficiency curves of holographic and ruled gratings. λ_β = Blaze wavelength. (Reproduced by permission of Instruments S.A.)

It is well known that, when the detector is not photographic and when the spectrum studied is continuous, the throughput of an instrument is proportional to the square of the slit width (all other parameters being fixed). Therefore, a larger density of grooves would increase this quantity. Multiplying the groove density creates a much higher level of ghosts and stray light for the RDG while this is not the case for the HDG (Figure 21). At the same time, because of the complete absence of ghost and lower stray light, HDGs generally have a much higher S/N (Figure 22).

Table 5 sums up the general characteristics of RDGs and HDGs.

5.3 Various Monochromator Designs

5.3.1 Principles of a Monochromator

Figure 23 shows a simplified scheme of a MC. The scattered light issuing from the sample is collected by a lens L1 (or a concave mirror) and is focused on the entrance slit. The parallel beam is collected by L2 (or a concave mirror) and falls onto a dispersing device generally constituted by a grating. The dispersed light is collected by L3. The focused diffracted beam passes through the exit slit situated in the focal plane of L3 and is recorded by a monochannel detector during the

Figure 21 Comparison of stray light levels of (a) ruled and (b) holographic gratings. (Reproduced by permission of Instruments S.A.)

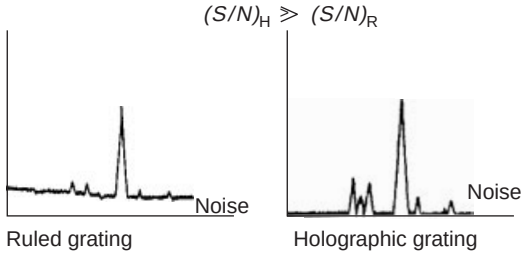

Figure 22 Comparison of the S/N for ruled and holographic gratings. (Reproduced by permission of Instruments S.A.)

Table 5 Characteristics of RDGs and HDGs

		RDG	HDG
Number of grooves per mm	n	20–3600	600–6000
Working region	$\lambda_1-\lambda_2$	UV/VIS/IR	UV/VIS/NIR
FSR	$\Delta\sigma_0$	Increases with n	UV/VIS/NIR
Blaze wavelength	λ_B	In the case of the Littrow configuration ($i = r$) the blaze angle φ_B is defined by $2 \sin \varphi_B = k\lambda_B/a$	UV/VIS/NIR
Efficiency		Given by the constructor for λ_B and Littrow use	UV/VIS/NIR
Ghost and stray light		10^{-6} to 10^{-4} of the parent line intensity	None
S/N		Very good	Excellent
Resolution		Depends upon the width of the grating, the working angles and the wavelength: • the finer the groove spacing, the higher the resolution • depends upon the stray light level	Excellent
Aberrations		No aberration when used with parallel beams	Excellent

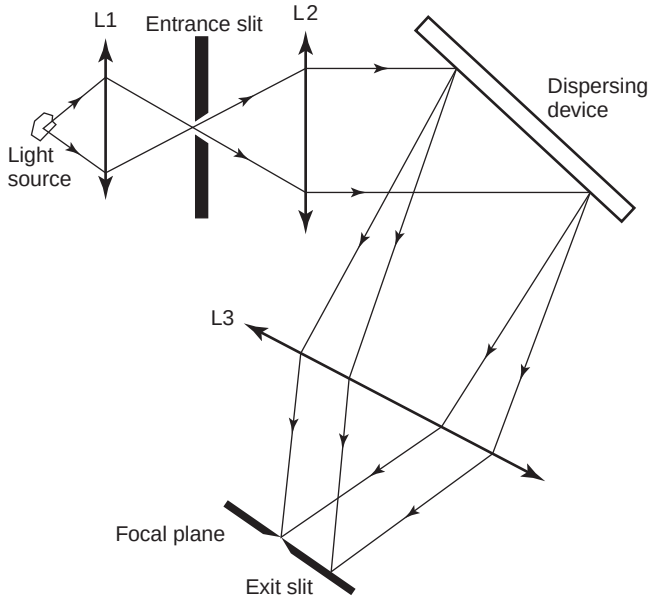

Figure 23 Simplified scheme of a MC. L1, L2, L3 are convergent lenses. With a spectrometer, the dispersing device which is generally a grating turns around on its axis and the diffracted beam pass through the exit slit to be recorded on a monochannel detector. With a spectrograph, the multichannel detector is placed in the focal plane and the whole spectrum range is recorded.

rotation of the grating in the case of a spectrometer. With a spectrograph, the multichannel detector is placed in the focal plane and a spectral range is recorded.

5.3.2 Premonochromators for Lasers

Solid-state lasers or diode lasers, except for the most expensive, do not have the exceptional qualities of a gas laser. The tendency is, however, to use them as Raman sources, since they are robust, cheap, and small in size. The relatively large bandwidths, presence of interference lines, and high background level make the use of IFs or PMC preferable.

IFs are simple and efficient but sometimes they are not available for the exact frequency value and the rejection level at few reciprocal centimeters is not sufficient to work at low frequencies.

Small MCs are often associated with the solid-state lasers to obtain an ideal Raman source. Figure 24 shows the effect of such a PMC on the filtering of undesirable laser lines.

5.3.3 Premonochromators for Raman

A PMC for Raman (Figure 25) is used to eliminate the laser line from the Raman spectrum. It must be followed by a spectrograph to disperse the Raman spectrum.

The advantages of PMC are

- good cut-off (5 cm^{-1} from the exciting line);
- no fluorescence;
- easy to adjust the PMC to the exact value of the diode with the same cut-off quality;
- changing the excitation wavelength is easy by rotating the grating.

5.3.4 Monochromators

There are several MC geometries. The most usual ones are shown in Figure 26.

In the Ebert–Fastié system, a single concave mirror is used as collimator and focus lens. This design gives small and compact high-aperture devices ($f/3$ to $f/5$). In the Czerny–Turner mounting, two small concave mirrors replace the simple mirror in the Ebert–Fastié system. The Czerny–Turner system has a low aperture but has better stray light correction and higher resolution. In the case of a Raman spectrograph, lenses can be used in the place of mirrors.

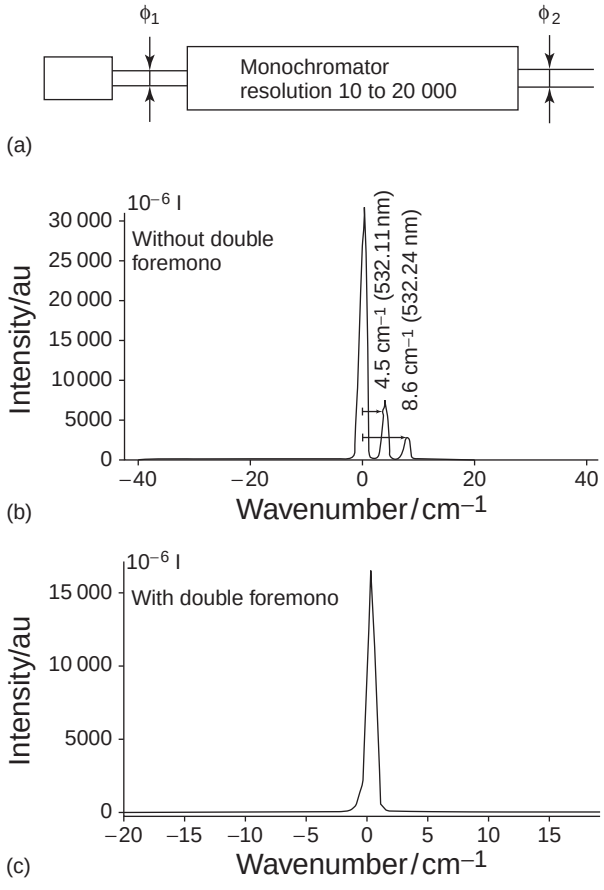

Figure 24 PMC for a laser. (a) The PMC is also used to adapt the small diameter of the laser beam Φ_1 of a low-priced laser, and to reduce the divergence of the output beam (Φ_2 is fixed by the aperture of the entrance optics). (b) Comparison of the spectra of the laser light obtained without a PMC and (c) with a PMC. (Reproduced with permission of DILOR.)

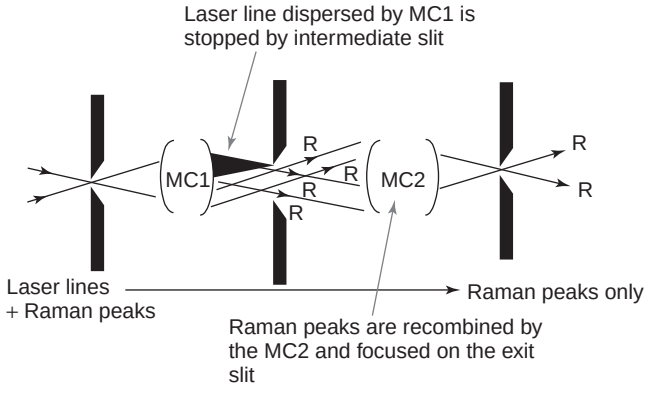

Figure 25 A PMC for Raman is a double subtractive spectrometer (no dispersion on the exit plane), (see Figure 27b). The laser line is eliminated by the intermediate slit.

The different characteristic parameters of a MC are summarized in Table 6.

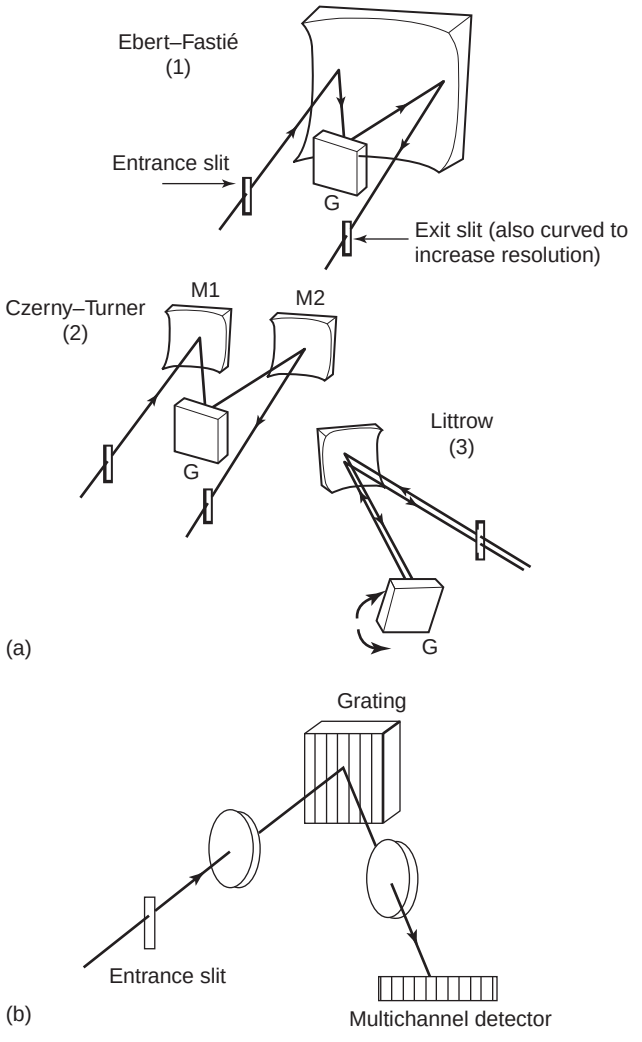

Figure 26 Some classical geometries of MCs. (a) Mirror spectrometers. UV to IR very good efficiency but optical aberrations (coma and astigmatism). Ideal for spectrometer (or UV/VIS spectrometer). (1) Ebert–Fastié configuration, (2) Czerny–Turner configuration, (3) Littrow configuration. (b) Lens spectrograph: very good optical device (flat field without vignettes). Ideal for multichannel detector.

6 RAMAN DEVICES

Raman devices are composed of all the different components described above. The association of these elements forms a spectrometer which must obey specific criteria. Various characteristics of spectrometers are found in Tables 7 and 8 and their defects are defined in Table 9.

6.1 Spectrometers and Spectrographs

In the above sections, the principal components and their characteristics are described for users to choose the

Table 6 Definition of some characteristic parameters of MC

Characteristic	Definition	Symbol and dimension
Resolution or resolving power	Possibility of separating 2 adjacent peaks	$R = \nu/\mathrm{d}\nu$ (dimensionless)
	Smallest frequency range measurable by an instrument	$R = \mathrm{d}\lambda$ or $\mathrm{d}\nu$ (nm or cm^{-1})
Angular dispersion	Angular range $\mathrm{d}\theta$ over which a band $\mathrm{d}\lambda$ is spread	$D_a = (\mathrm{d}\theta/\mathrm{d}\lambda)$ (rad. nm^{-1})
Linear reciprocal dispersion	Wavelength range spread over a unit distance in the focal plane	$D^{-1} = \mathrm{d}\lambda/\mathrm{d}x$ ($\mathrm{nm\,mm}^{-1}$)
Spectral band width	Smallest wavelength interval that the MC can isolate	$s = \mathrm{fwhm} = WD^{-1}$
Light-gathering power	Ability of the collimator mirror to collect light emerging from the entrance slit	$f/\mathrm{number} = f/d$

ν is the frequency expressed in cm^{-1}; S is the spectral bandwidth; W is the exit slit width; f is the focal length of the objective; d is diameter of the entrance aperture.

Table 7 Characteristics of spectrometers

Characteristic	Definition	Unit
Throughput	Photon flux at the exit of the instrument	
Flux ϕ	Energy/time emitted from a light source or slit of given area for a solid angle Q at a given wavelength (or bandpass)	photons s^{-1} or W
Intensity I	Distribution of flux at a given wavelength (or bandpass)/solid angle	$\mathrm{W/steradian}^{-1}$
GE	Maximum beam size an optical instrument can intercept	$\mathrm{d}^2\mathrm{GE} = \mathrm{d}S \cdot \mathrm{d}Q$ $\mathrm{GE} = 4\pi S \sin^2 \alpha/2$
Radiance (luminance) B	Intensity when spread over a given surface	$\mathrm{W/steradian}^{-1}\,\mathrm{cm}^{-2}$
Solid angle Q	Surface intercepted on a sphere by a cone of half-angle α	$\mathrm{d}Q = \mathrm{d}S/r^2 = 2\pi \sin^2 \alpha$ sterad
NA	NA of an objective in a medium of index μ ($\mu = 1$ in air)	$\mathrm{NA} = \mu \sin \alpha$
f/n F number	Objective is opened at nn, f/n. n is related to NA by $\mathrm{NA} = \sin \alpha$ and $n = \mathrm{tg}\,\alpha$	$n = \dfrac{\mathrm{NA}}{\sqrt{1 - \mathrm{NA}^2}}$

NA, numerical aperture; GE, geometric etendue; tg α is the trigonometric function defined as sinus α/cosinus α.

Table 8 Characteristics of some Raman devices

	New double MC (HPG) Raman U 1000	Triple additive MC DILOR Coderg T800	Triple MC 1 double MC in subtractive configuration + 1 spectrometer	Triple MC 1 double MC in subtractive configuration + 1 spectrograph Jobin-Yvon S3000	Labram (confocal)
Grating	HPG 1800 grooves mm^{-1}	HPG 1800 grooves mm^{-1}	600, 1200, 1800 grooves mm^{-1}	600, 1200, 1800 grooves mm^{-1}	HNF + HPG
Focal length (mm)	1000×2	800×3	$320(\times 2) + 640$	$320(\times 2) + 640$	+300 mm
Spectral range (nm)	300–1000	350–900	350–780	350–780	350–980
Stray light	$<10^{-14}$ at 20 cm^{-1}	$<10^{-14}$ measured at 50 cm^{-1} with 1 cm^{-1} slit width	$<10^{-10}$	$<10^{-10}$	$<10^{-9}$
Resolution	120 000	120 000	40 000	15 000	8000–10 000
Slit length	15–20 mm	18 mm	1–20 mm	1–20 mm	1 mm[a]
Detector	PMT + CCD	PMT + photon counting	PMT + photon counting	Intensified PDA or CCD	CCD
Source	190–1000 nm	Any lasers (240–800 nm)	Any lasers (240–800 nm)	Any lasers (400–1000 nm)	Incorporated He–Ne laser
Recordable from excitation line	10 cm^{-1}	1 cm^{-1}	5 cm^{-1}	5 cm^{-1}	50 cm^{-1}

[a] With microscope entrance.
HPG, holographic plane grating.

Table 9 Various defects in MC

Parasites[a]	Comments	Correction
Stray light	Comes from imperfection in the instrument: grating, aperture stop (or slits), detector	HDG, confocal system, good detectors
Ghost	Parasite peaks due to periodical defects in RDG; do not exist in HDG	Use HDG
Slit curvature	Apparent bend in the entrance slit image at the exit plane due to optical aberrations	Double additive or subtractive MC
Coma	Leidel aberration resulting from the use of off-axis mirrors within the MC; wavelength dependent	Mirror corrections

[a] Any radiation, other than the wavelengths of interest, which passes through the MC to the detection system.

components of the Raman device most suited to their problems.

For RS, in order to obtain the lowest stray light rate, double and triple MCs are mounted. Additive double MCs are generally used for monochannel detection. In this case, the aperture of the slits obeys the following conditions: entrance slit = intermediate slit = exit slit.

The result is a very low back-scattering light (Figure 27a).

For very low frequency Raman a triple additive spectrometer permits coverage of the range from a few cm^{-1} to $3500\,cm^{-1}$ including all the laser excitation lines from near-UV to the VIS with a very good resolution using a monochannel detector (Figure 27b). With a multichannel detector it is necessary to use a double subtractive spectrometer (for all the Raman lines) associated with the spectrograph (Figure 27c) or a NF (for a Raman line) and a spectrograph (Figure 27d). Generally to be stigmatic and to cover the large field for the use of multichannel detectors, the spectrograph is refractive (except for UV Raman spectrographs).

The position of the grating for the determination of the wavelength value is activated by a mechanical system. Figure 28 shows the ''sinus bar'' geometry, in which wavelengths are linearly related to the screw displacement. Another system is also used in which the cosecant bar gives a linear relation between the wavenumber and screw displacement.

There are several commercial Raman spectrometers available on the market with two main applications. For research applications, there are powerful and even versatile instruments (which are not mutually exclusive) available except for special studies such as CARS or picosecond TRRS. The latter are quite often assembled by physicists in their own laboratories. For industry, the priority is a robust, simple-to-use miniaturized instrument.

Figure 29 shows a diagram of a typical commercial high-resolution Raman triple MC. Figure 30 represents a mixed version spectrometer/spectrograph, using a PMC mounted with a double subtractive grating.

6.2 Dispersive Raman Microscopes and Raman Imaging

With the dispersive Raman techniques, one of the most interesting possibilities is the study of microscopic samples even when they are within a transparent solid matrix. There are several specific applications for dispersive Raman microscopes (DRMs), in solid state physics, mineralogy, biology, and industry. With the latest multichannel detectors, Raman imaging can also be realised.

One achievement concerns micro-RS using an optical scanning confocal microscope (SCM). Unlike the conventional optical microscope, where the entire field of view is illuminated and imaged simultaneously, the confocal arrangement uses a point light source sharply focused on a diffraction-limited spot by a microscope lens onto the specimen. The aperture matching of the spectrometer is perfectly adjusted by the special use of a spatial filter, which isolates a region exactly coincident with the illuminated spot through a pinhole aperture.

Figure 31 shows a schematic set-up of a confocal microscope and compares the fwhm with a conventional microscope. This configuration gives improved depth of field and axial resolution, better contrast, and good lateral resolution.

New applications can be found for SCMs, for example the examination of impurities inside a host matrix to determine mineral inclusions in host crystals (Figure 32).

6.3 Specific Problems Encountered in Monochromators

6.3.1 Aperture Matching

The GE[22] is a constant of a MC system (Figure 33) and the apertures are optically matched when they have the same GE for the whole system Equation (13):

$$\begin{aligned} GE &= \pi s \sin^2 \frac{\alpha}{2} \\ &= \pi s' \sin^2 \left(\frac{\alpha'}{2}\right. \quad = \pi s'' \sin^2 \left(\frac{\alpha''}{2}\right. \\ &= \pi s^* \sin^2 \left(\frac{\alpha^*}{2}\right. \end{aligned} \tag{13}$$

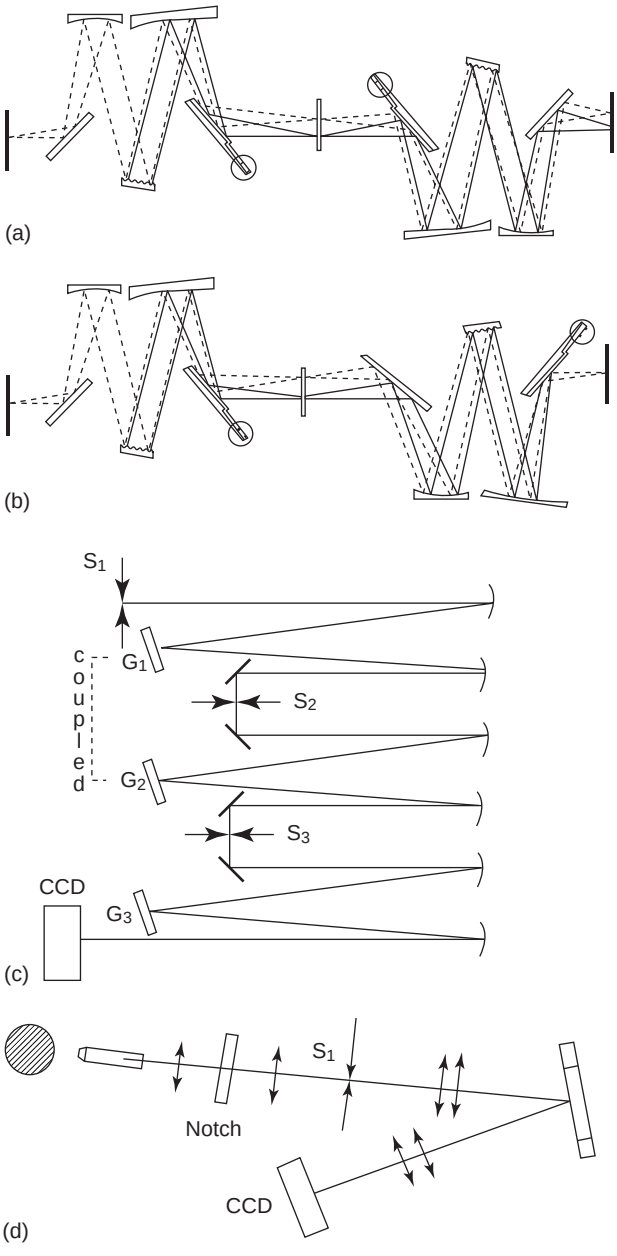

Figure 27 Various configurations of spectrometers and spectrographs. (a) Double additive configuration. In this case, the output spectral band from the first MC is dispersed again by a second MC. An additive double MC spectrometer provides approximately twice the dispersion of a single MC. It is also possible to increase the dispersion again by using a third MC. (Triple MC spectrometer of Figure 29). (b) Double subtractive configuration. The narrow portion of the spectrum passing through the exit slit of the first MC is focused again at the exit of a second MC. The dispersion is slightly less than for a single MC owing to instrumental imperfections (alignment, synchronization, aberrations). In order to obtain the spectrum, it is necessary to disperse this portion again by a third MC (see Figure 30). (c) In the case of a spectrograph, a multichannel detector is used. (d) Notch + spectrograph. This configuration gives rise to a very high luminosity instrument. (Reproduced with permission of Oriel Instruments.)

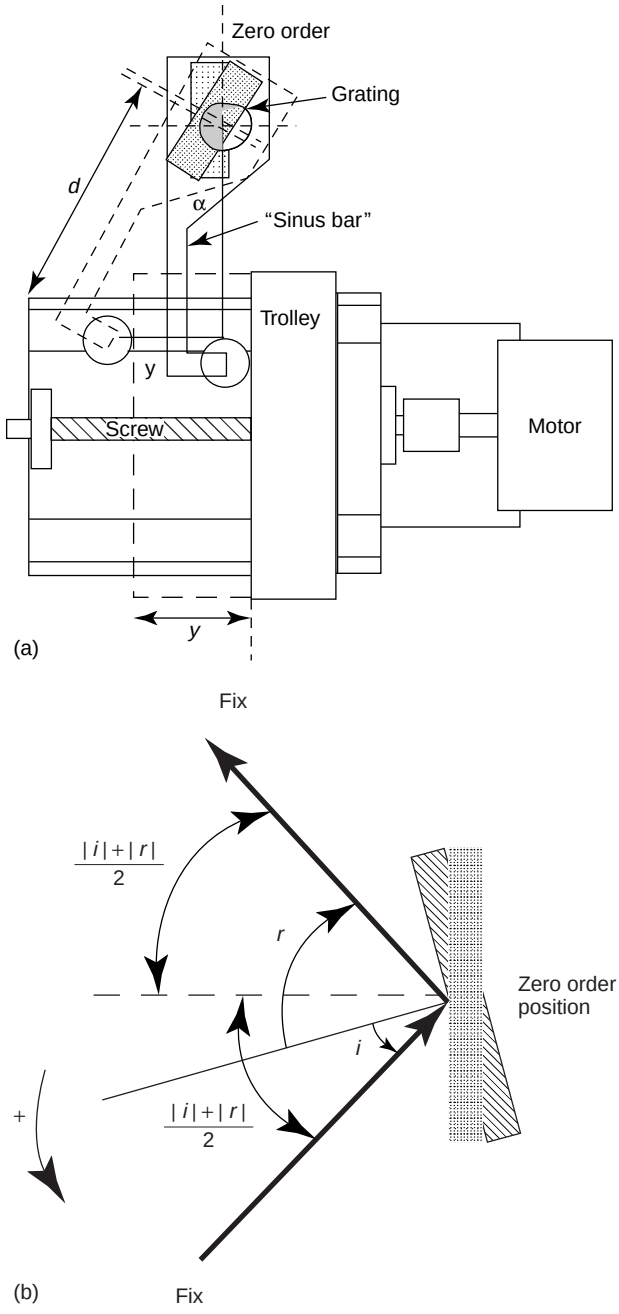

Figure 28 "Sinus bar" geometry. In the case of a double MC, the two gratings are mounted on the same shaft. The axial movement of the trolley corresponds to a rotation of the grating. Following the grating equation $2\sin[(i+r)/2]\cos D_v/2 = kn\lambda$. At the zero order, $(i+r)=0$. When the grating rotates α° from its zero position, $(i+r)=2\alpha$. As $y/d=\sin\alpha$, the grating equation gives $2y/d = kn\lambda/\cos D_v$ and $y = k'\lambda$. The horizontal displacement is a linear function of the wavelength. (Reproduced with permission of DILOR.)

For the user, only the entrance optics are interchangeable, while the different instrument apertures are conceived by the constructor. Therefore, when changing a camera lens

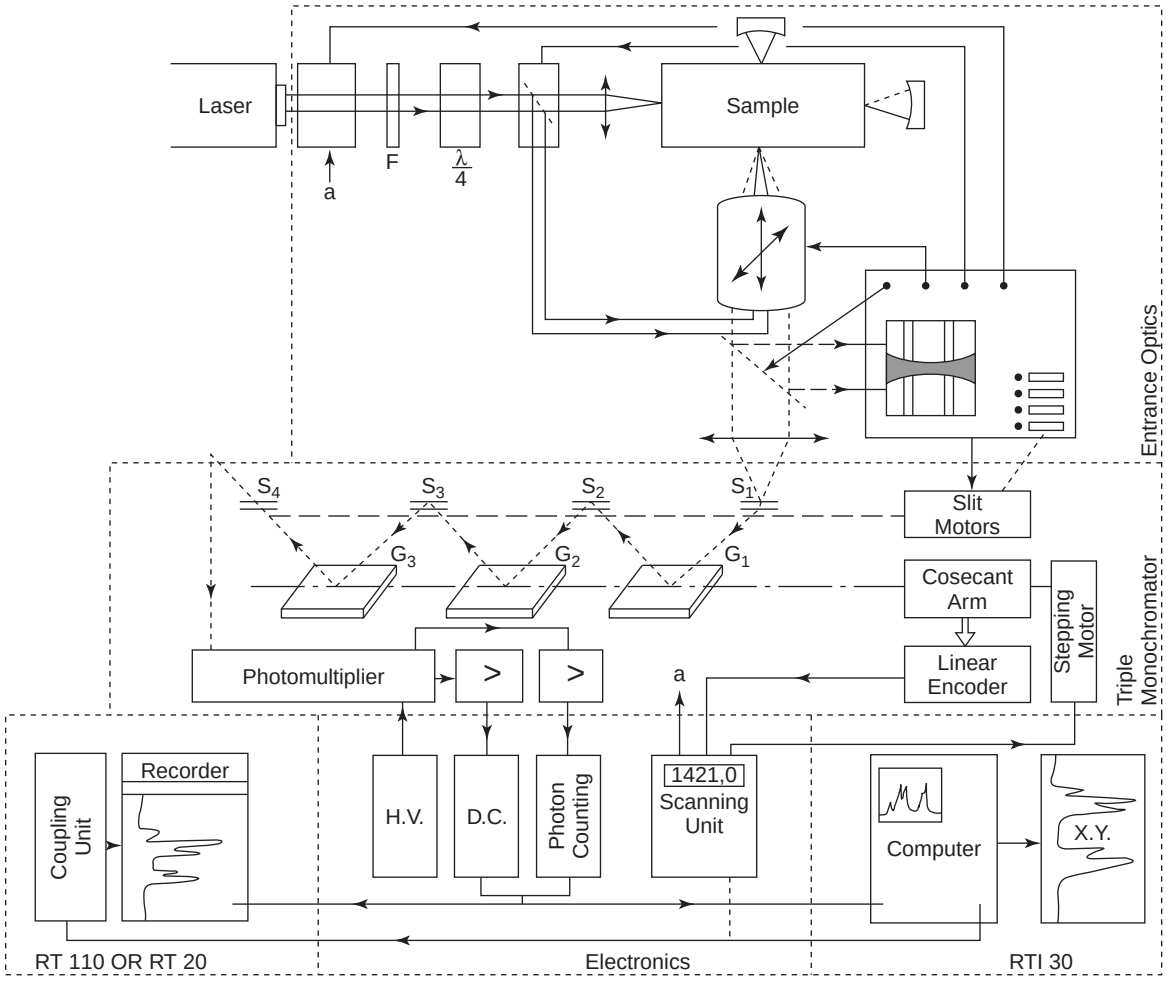

Figure 29 Scheme of a triple MC Raman spectrometer, comprising the laser source, the sample compartment, the triple MC compartment with photomultiplier (PM) detection compartment and the electronics for amplification and recording. S1 and S4 are respectively the entrance and exit slits, S2 and S3 the intermediary slits. G1, G2, G3 are the three gratings which successively disperse the Raman signals issuing from the sample. (Reproduced with permission of DILOR.)

for a microscope objective, care must be taken with the aperture matching.

Figure 34 shows a comparison of the collection efficiency of camera lenses and microscope objectives. As a practical consideration, if a sample is opaque or very thin (penetration width of a few microns), it is better to use a microscope objective with high NA. For a slice of silicon, the intensity recorded with an objective ($\times 100$, NA $= 0.9$) is 23 000 while with a camera lens (50 mm, $f/1.4$) the intensity recorded is only 130. Conversely, for a transparent sample (calcite crystal) with high penetration depth, it is better to use a small magnification microscope objective ($\times 10$), the signal from a larger volume can then be collected. For the sets of objectives for the calcite crystal, the recorded intensities are respectively 16 000 and 28 000.[23]

In this case, the camera objective is better.

6.3.2 Resolution

In single-channel mode, the image of the entrance slit is scanned over the plane of the exit slit. The optimum resolution is obtained when the entrance slit aperture is equal to the exit slit. The slit contribution to the instrument line profile is the convolution of the two slit functions. The line profile is triangular when the exit and the image of the entrance slit widths are equal.[24] Figure 35 shows an example of the influence of the slit aperture on the resolution and intensity of DRS Raman spectra. The recording time increases with resolution.

In multichannel mode, the image of the entrance slit is projected onto the two-dimensional detector, with a magnification factor $G < 1$ depending on the optical configuration of the spectrograph. On the detector, the slit function is trapezoidal.

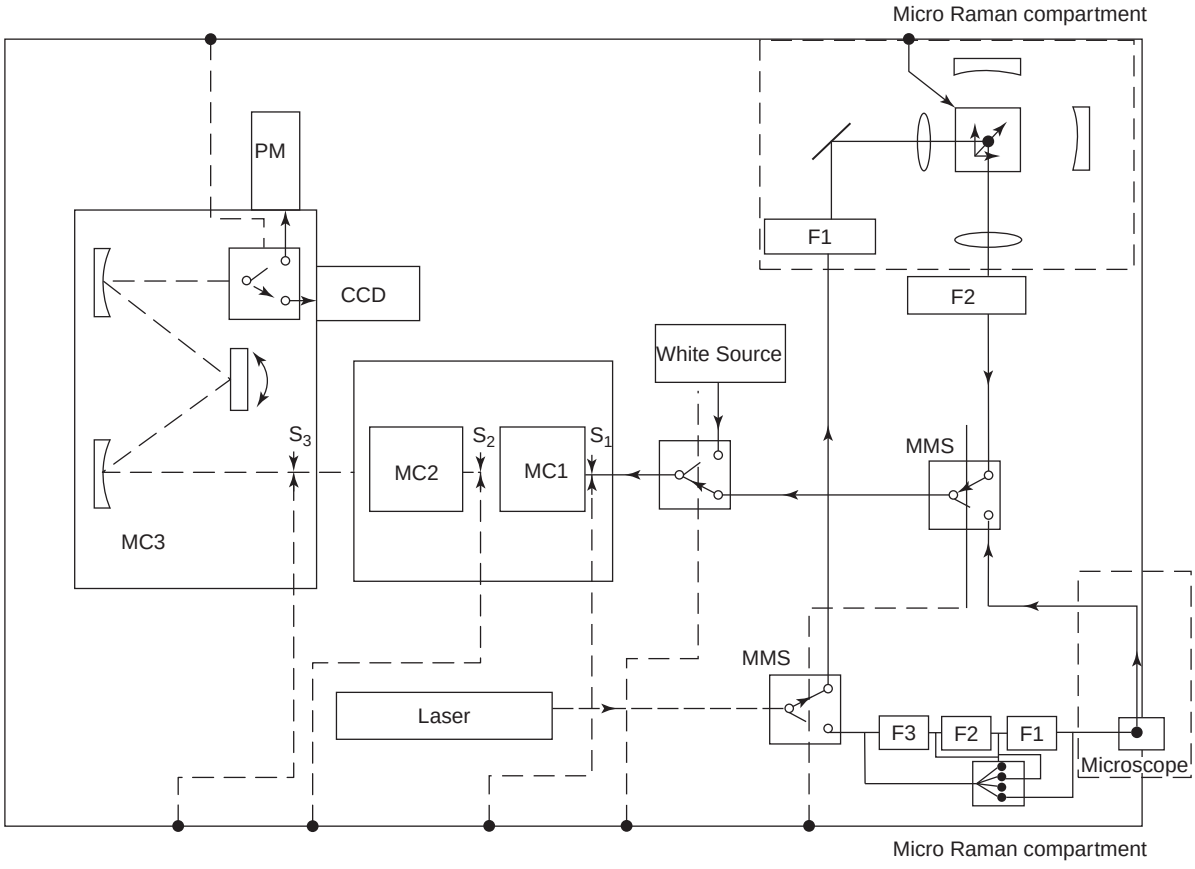

Figure 30 Scheme of a mixed spectrometer/spectrograph. MC1, MC2, MC3 are three MCs, MMS is the micro/macro switch, F1, F2 are filters and S1, S2, S3 are MC slits. Depending on the MMS position, the spectrometer can either function as a triple MC spectrometer together with the PM or a spectrograph with a double PMC in the subtractive geometry and a CCD detector. (Reproduced with permission of Instruments S.A.)

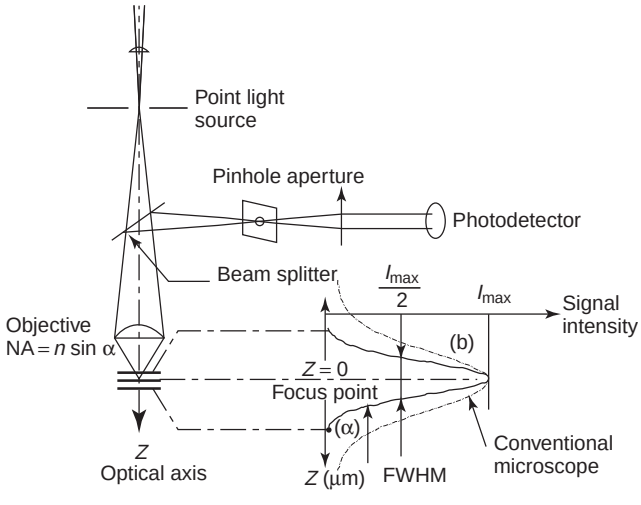

Figure 31 Schematic set-up of a confocal Raman microscope. The pinhole aperture is placed at the conjugate point of the focus point of the microscope objective. Only the diffused light issued from this point passes through the pinhole. The resolution is greatly increased. (Reproduced with permission of DILOR.)

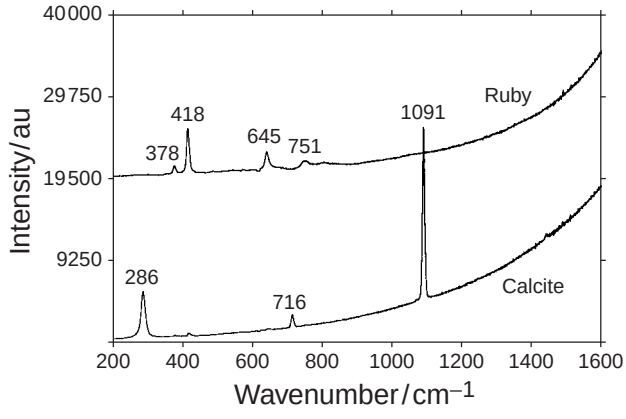

Figure 32 Raman spectrum of an inclusion of calcite in a ruby host as recorded with a confocal Raman microscope.

6.4 Influence of Excitation Wavelength

The intensity of the Raman bands of a compound depends greatly on the excitation wavelength. Numerous compounds are fluorescent because UV/VIS radiations

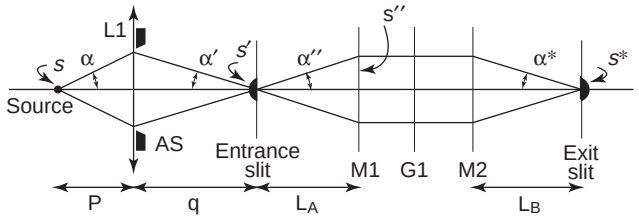

Figure 33 MC system and matching aperture. s is the area of source, s' is the area of entrance slit, s'' is the area of mirror M1, s^* is the area of exit slit, α is the half angle of light collected by L1, α' is the half angle of light submitted by L1, α'' is the half angle of light collected by M1, α^* is the half angle of light submitted by M2, L1 is the lens used to collect light from source, M1 is the spherical collimating Czerny–Turner mirror, M2 is the spherical focusing Czerny–Turner mirror, AS is the aperture stop, G1 is the diffraction grating.

can excite bonding electrons to higher excited states and the de-excitation gives rise to large fluorescence bands. Biological samples also present strong fluorescence bands for VIS radiations owing to the diverse impurities present in these complex samples. In these very common cases, a change in the VIS radiation to NIR excitation can greatly reduce the fluorescence background. This is the main benefit of NIR excitation (Figure 36). In some other cases, fluorescence is caused by doped ions. The Cr^{3+} ions of ruby (corundum doped with Cr^{3+} ions) give rise to a very strong fluorescence, consisting of a very high background and a strong doublet at 692.8 and 694.2 nm in the red region. By exciting with a red wavelength, the Raman spectrum of corundum is recovered with intense fluorescence and is not recordable, while with a green wavelength, good Raman spectra can be obtained as it is far from the fluorescence bands. Figure 37 shows the Raman spectra of a ruby stone with two different excitation wavelengths, in the red region where maximum fluorescence is observed and in the green region where the fluorescence is weaker.

For analytical purposes, it is necessary to be very careful when changing the excitation wavelength. Figure 38

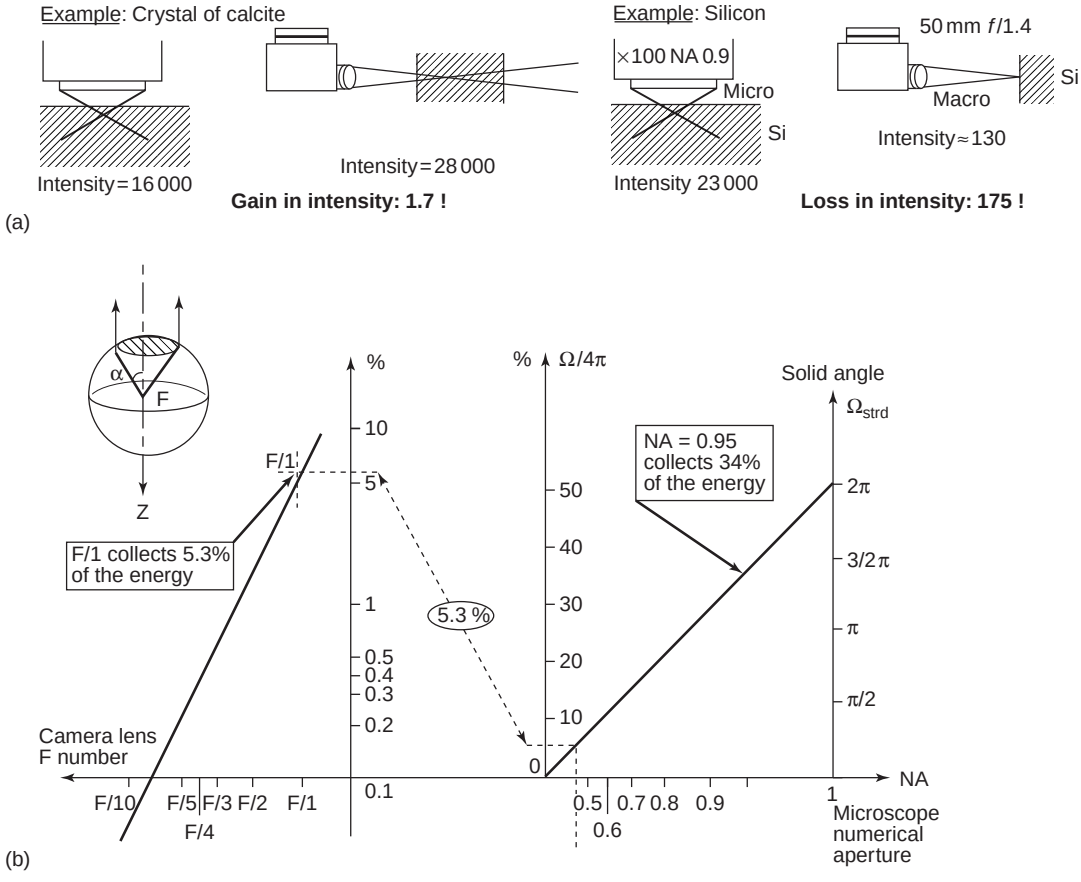

Figure 34 Comparison of the efficiency of collection of camera lenses and microscope objectives. (a) Using a microscope objective ($\times 100$; NA = 0.9) and a macrophoto objective, the gain or loss in intensity depends on the observation modes. In the case of a transparent substance (Calcite), the collection is slightly better for the macrophoto objective, while for an opaque sample (Silicon), the collection is much better for the microscope system. (b) See also graphical representation. (Reproduced with permission of DILOR.)

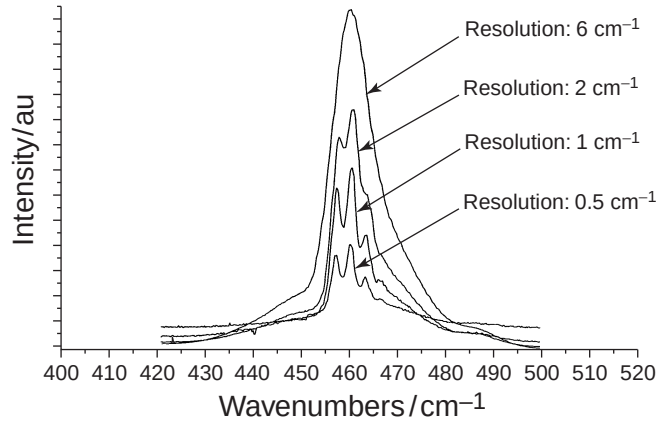

Figure 35 Influence of the slit aperture on the resolution in DRS. When reducing the slit width, the resolution is better while the intensity of the Raman signals decreases. Sample, CCl_4; excitation, 514.5 nm; spectrum from 420–500 cm^{-1}; integration time, 5 s; mode, single-channel sequential; laser power, 100 mW; step, 0.2 cm^{-1}; accumulations, 1.

shows another example, where the Raman spectra of a hard disk recovered with a carbon layer is recorded at the same point with the same optics and with two laser excitations. A doublet is observed for the carbon layer; one at 1368 cm^{-1} and the other at 1575 cm^{-1} caused respectively by the diamond-like and the graphite-like vibrations. The fluorescence backgrounds and the relative intensity of the doublet are different and the deconvolution of the band after background subtraction (Figure 39) gives a completely different intensity ratio of the doublet. The inversion in intensity of the two bands is caused by different preresonance intensity enhancements when excitation wavelengths are different.

Therefore, in analytical applications it is highly advisable to use the same excitation wavelength for the measurement of Raman intensity and for the calibration phase.

7 SAMPLING METHODS IN RAMAN SPECTROSCOPY

Sampling in RS can be very simple because the substance can be examined without any preparation. Improvements have been made, however, in order to make it possible to simplify the sampling procedures even more or to increase the potential of the method. Several set-ups are now quite standard.

7.1 Sampling Configurations

The Raman effect is a diffusion phenomenon spread over 4π steradians of space. To collect the Raman effect, various set-ups can be used. The most common ones are 90° collection and back-scattering geometry (Figure 40) and they can be conveniently realised in DRS.[25]

7.2 Polarization Studies

In studies of liquid and solid substances, it is sometimes important to observe polarization effects. Figure 41 shows the principles of a polarization set-up for studying the polarization of a single crystal. Porto's notation[26] is often adopted in this kind of problem and is shown in Figure 42.

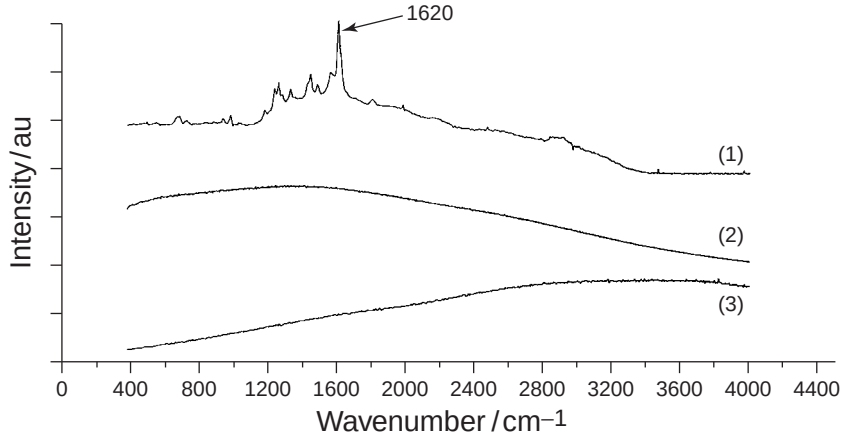

Figure 36 Raman spectra of a black pigment human gallstone using different excitation wavelengths. (1) Excitation line, 1064 nm; Scans, 300; time, 3 s. (2) Excitation line, 632.8 nm; time, 0.1 s; accumulation, 10; grating, 300 lines mm^{-1}; slit, 120; hole, 1000. (3) Excitation line, 514.5 nm; time, 0.1 s; accumulation, 10; grating, 300 lines mm^{-1}; slit, 120; hole, 1000. With NIR excitation (1064 nm), the spectrum of the chemical compound (calcium oxalate) in the calculus is well defined, while with the other two excitation wavelengths in the VIS region, the fluorescence phenomenon completely hides the Raman spectrum of the stone.

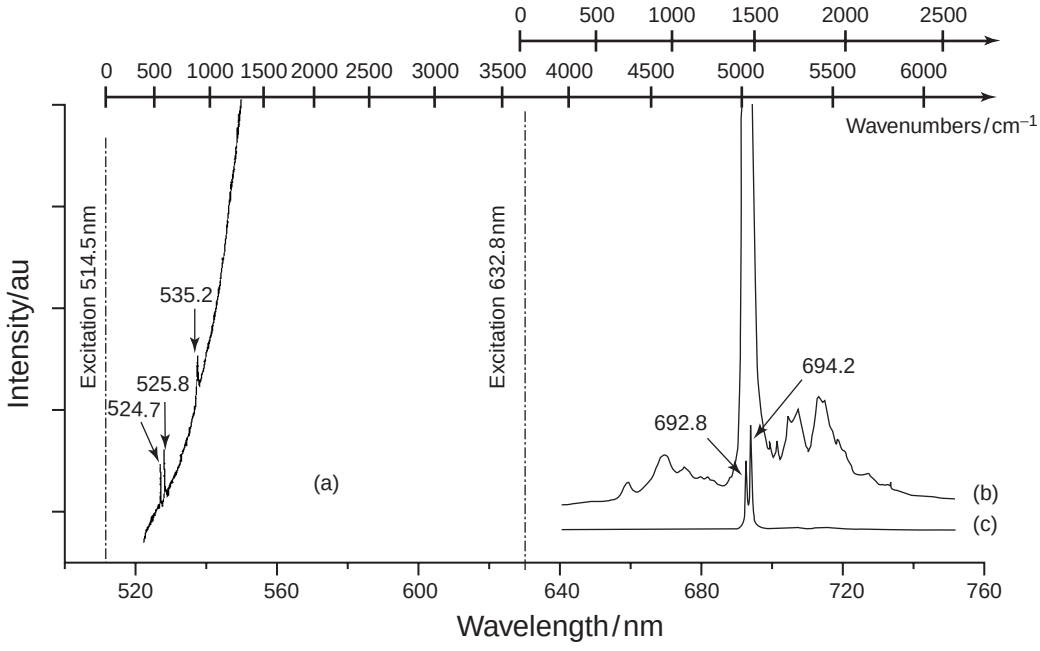

Figure 37 Raman and fluorescence spectra of a ruby stone. Grating, 1800 lines mm^{-1}; slit, 120; hole, 1000; no filter; objective, 50. (a) Time, 5 s; excitation line 514.5 nm; accumulation 35; from 522.5–563.7 nm; condition, notch angle 7.8°. (b) and (c) Time, 0.05 s; excitation line 632.8 nm; accumulation, 10; from 641–751.9 nm; condition, notch angle 11.5°. Ruby is an Al_2O_3 mineral containing Cr^{3+} impurities. Ruby gives a very strong fluorescence in the red region due to chromium ion impurities. When excited by red radiation (632.8 nm in wavelength or 15 803 cm^{-1} in energy), the Raman spectrum of Al_2O_3 is situated in the region of Cr^{3+} emission and is completely hidden by the fluorescence spectrum (b, c). Using green radiation (514.5 nm or 19 436 cm^{-1}), the Raman spectrum of Al_2O_3, which is far from this region (a) is observed: peaks at 524.7 nm (corresponding to a Raman shift of 378 cm^{-1}), 525.8 nm (Raman shift = 418 cm^{-1}) and 535.2 nm (Raman shift = 751 cm^{-1}). Note that the recording time is a 100 times longer for the Raman spectrum (a) than for the fluorescence spectrum, that is the intensity is 100 times weaker for (a) than for (b) using the same ordinate scale.

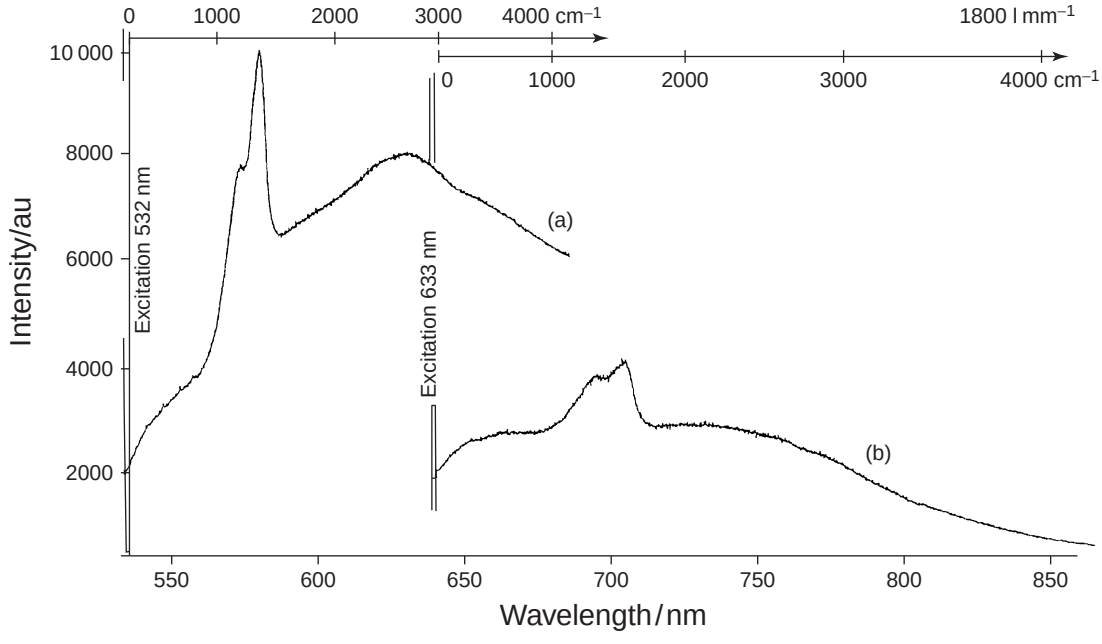

Figure 38 Raman spectra of a hard disk covered with a carbon layer using two different wavelengths: the background and the intensities of the peaks change greatly. (Reproduced with permission of DILOR.)

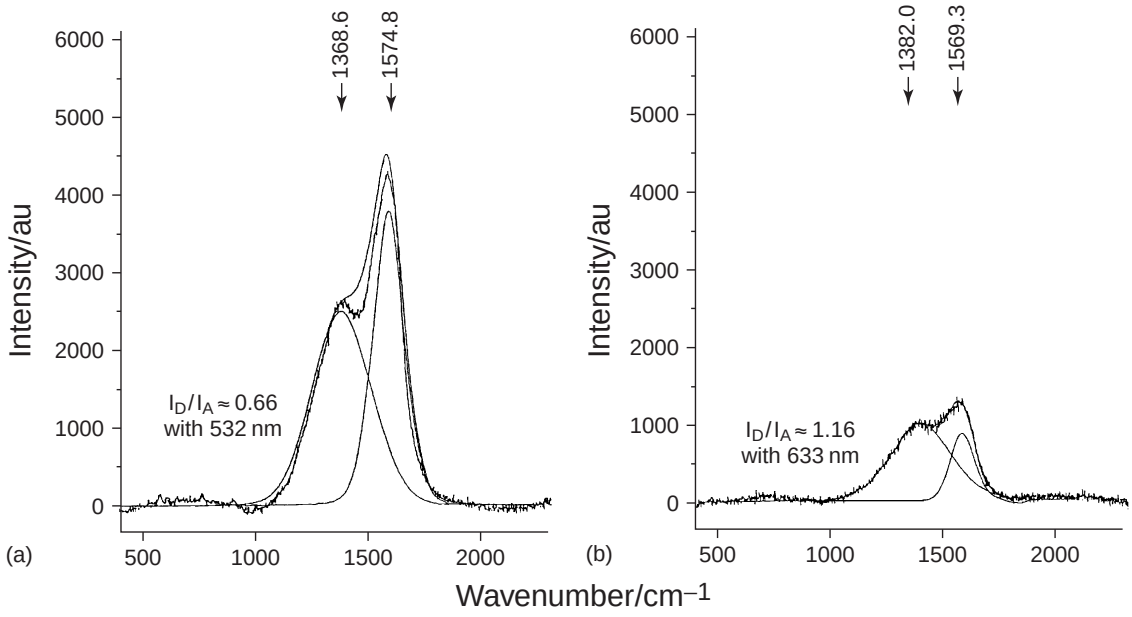

Figure 39 Inversion of intensity measurements of Raman lines due to preresonance effect. (Reproduced with permission of DILOR.)

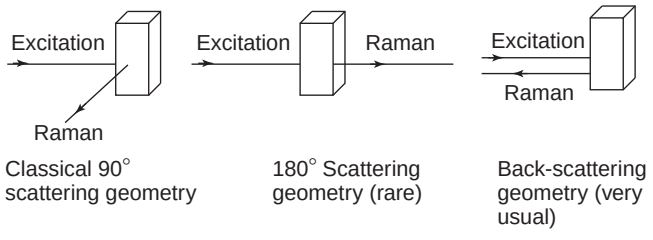

Figure 40 Various geometries for Raman sample observation.

It is also worth noting that even with high-NA microscope objectives, the polarization effect is still important and useful for good physical interpretation.[27]

7.3 Coupling with Optical Fibers

Raman optrodes using optical fibers have been intensively developed, creating even more applications for RS. Remote and in situ measurements with the possibility of site multiplex are the most important uses of this new technique. As the Raman spectrum is situated in the range where silica has good transmission power, optical sensors have been conceived for miniaturized optrodes and for remote and in situ control using optical fibers.[28–35] Optrodes are increasingly used for routine studies. The physical process of transmission of laser light and collection of the Raman signal with optical fibers is shown in Figure 43. Such an optrode needs two optical fibers, one to transport the excitation radiation and the other to transport the Raman spectrum back to the spectrometer. Appropriate filters are necessary to eliminate the Raman fiber spectrum in the excitation fiber and the collection fiber, respectively. Figure 44 shows some optrode probes which have been tested successfully for Raman applications; some of them are now commercially available, others can easily be home-made. Figure 45 shows a comparison of the performance of some optrodes. Figure 46 shows a diagram of a RLFO installation.

8 COMPUTER TASKS

Computers have played an important role in the development of RS. They help to automate several delicate and time-consuming operations, thus improving the manipulation of the instrument.

The microcomputer has different tasks:

- automatic control of the instrument parameters, in particular performing the slit aperture program and moving the cosecant bar for grating wavelength control, recording the detector signal, and presenting the display of the spectrum as a function of the wavenumbers;
- spectrum treatment consisting of operations such as deconvolution, baseline and detector response corrections, addition and subtraction of spectra, decomposition of overlapping peaks, and even holding a library of spectra for identification and analytical purposes.

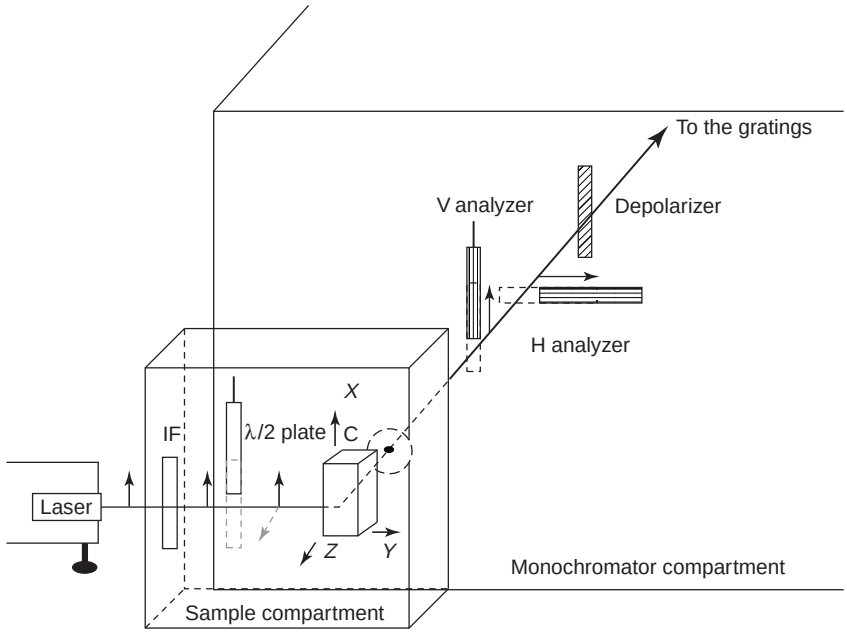

Figure 41 Polarization set-up for studying a single crystal in 90° scattering geometry. IF is the interference filter used to eliminate the plasma lines of the laser, C is the oriented crystal, V, H polarizers are vertical and horizontal polarizers. Depolarizer is a $\lambda/4$ or a scrambler, used to neutralize the polarization effect of the gratings. For a given position of the crystal C (presently *X* vertical), the following polarizations can be obtained, depending on the position of the $\lambda/2$ plate and the V and H polarizers, in the case of a 90° geometry recording.

$\lambda/2$ plate	V analysis	H analysis	Polarization	$\lambda/2$ plate	V analysis	H analysis	Polarization
out	in	out	$Z(XX)Y$	in	in	out	$Z(YX)Y$
out	out	in	$Z(XZ)Y$	in	out	in	$Z(YZ)Y$

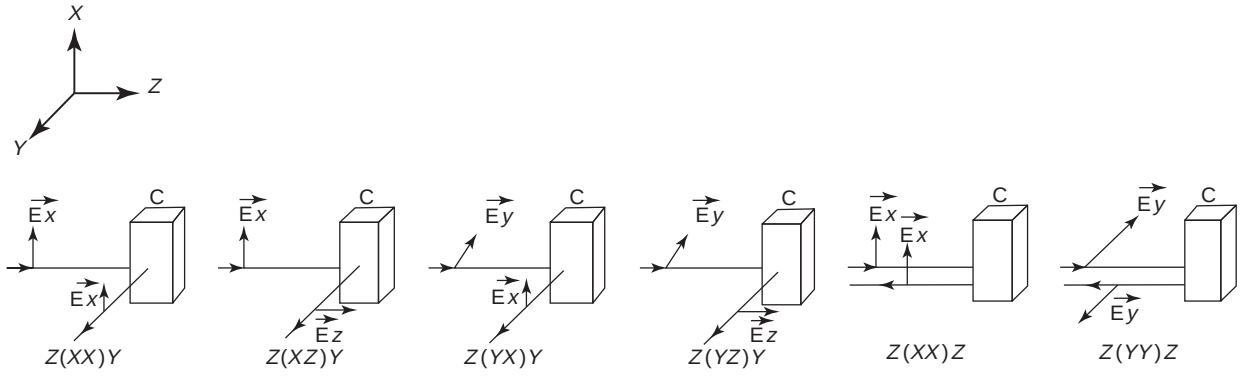

Figure 42 Porto's notation. With one single position of the crystal C, four different polarizations can be obtained. In order to obtain the other two polarizations, it is necessary to rotate the crystal. In the case of a back-scattering set-up, only two polarizations are accessible and never the cross-term polarizations.

9 PERSPECTIVE AND FUTURE DEVELOPMENTS

Raman instrumentation is progressing rapidly and intensively. New types of Raman device are being studied and developed. FT-Raman spectrometers are now well established for specific applications, and Hadamard Raman spectrometers are being tested.[36] However, DRS still retains its predominant role. In the UV/VIS region, only dispersive Raman instruments are at present available. In the NIR region, only monochannel NIR detectors were available until recently and this

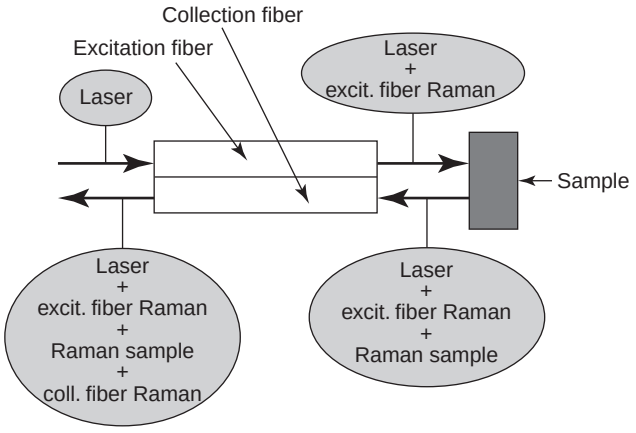

Figure 43 Principle of remote and in situ Raman measurements by optical fibers.

is the reason why FT-Raman spectrometers were first developed, thanks to the well-known Jacquinot (aperture size),[37] Fellgett (frequency multiplex),[38] and Connes advantages (superimposition and co-adding of spectra)[39] of interferometers. However, with progress in CCD technology, NIR multichannel detectors are now available and their performances are remarkable, allowing the extension of dispersive Raman instruments in this region to great advantage. DRMs working in the NIR have only just become available and the specific advantages linked to this method, in particular the ease of changing the excitation lines, the possibility of using diode lasers and solid state lasers open up new applications in this field.

Depending on the specific requirements of the users, the choice of instruments is essential. For physicists who need high resolution, studies of Raman lines near the excitation Rayleigh line, a long focal spectrometer with triple MCs in order to obtain a very high resolution and a low stray light level is advised. Resolution is usually better with monochannel detection, so a photomultiplier with a photon counter is suitable for this kind of study. For TRRS the recording of the spectra needs to be rapid, therefore CCD arrays as detector are often used although more expensive. For industrial applications, resolution is not usually a very important factor. Robustness, rapidity, simplicity, and also cheapness are the main requirements. Very often, a simple HNF as PMC and a HPG as MC is enough for analytical purposes. For routine studies, Raman manufacturers build versatile instruments, grouping a PMC with a double MC in a subtractive configuration, a MC composed of a grating of 300 grooves mm^{-1} for fast examination of the sample and a grating of 1800 grooves mm^{-1} for the final Raman spectrum.

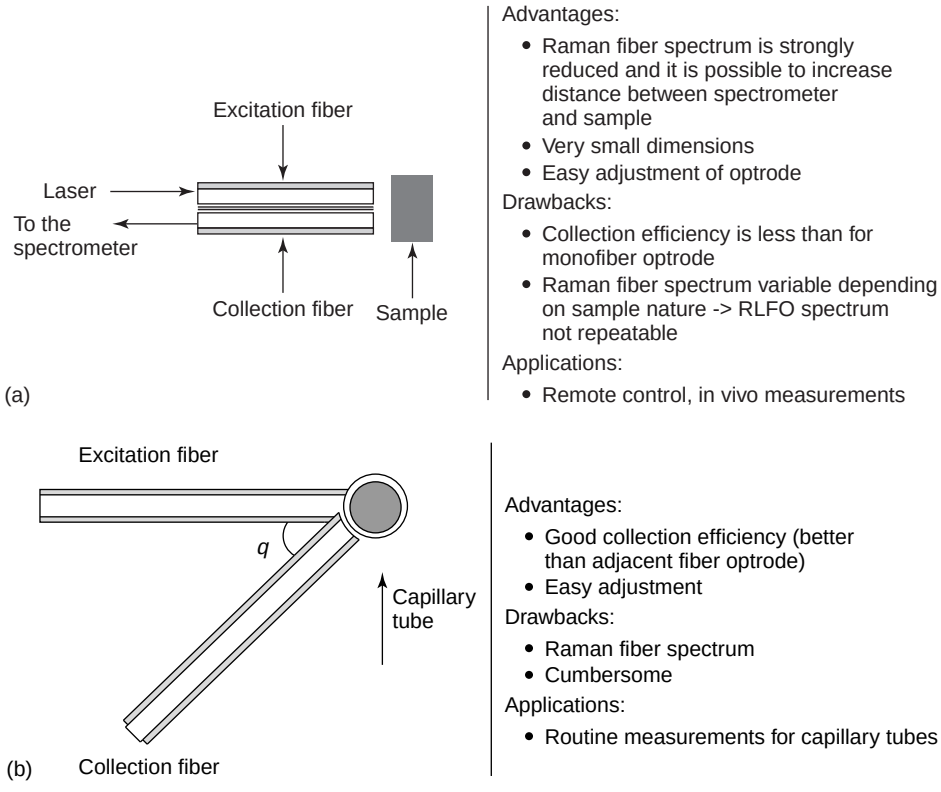

Figure 44 Some optrodes geometries. (a) Two adjacent fiber optrodes characterized by two single fibers whose ends are joined, (b) optrode for capillary tube, (c) handy optrode, (d) DILOR super head, (e) Grin lens optrode, (f) metallized tube optrode.

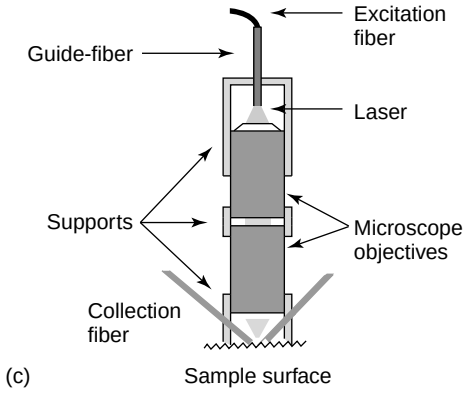

(c)

Advantages:
- Pre-adjustment
- Very handy

Drawbacks:
- Raman fiber spectrum not completely filtered

Applications:
- Surface analyses and remote Raman imaging

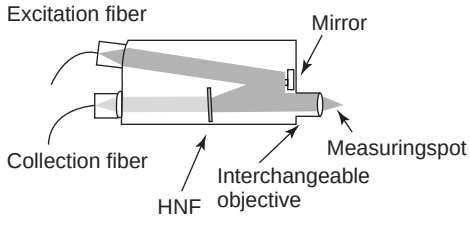

(d)

Advantages:
- Robust
- Pre-adjustment
- Excellent elimination of Raman fiber spectrum
- Interchangeable objective

Drawbacks:
- Cumbersome

Application:
- Remote measurement up to some hundreds of meters

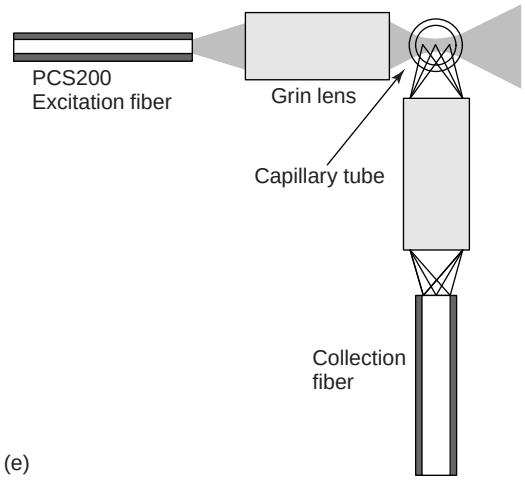

(e)

Advantages:
- Very high efficiency
- Excellent elimination of Raman fiber spectrum when filters are added

Drawbacks:
- Cumbersome
- Expensive

Application:
- Remote measurement for expensive liquids and gases up to some hundred of meters

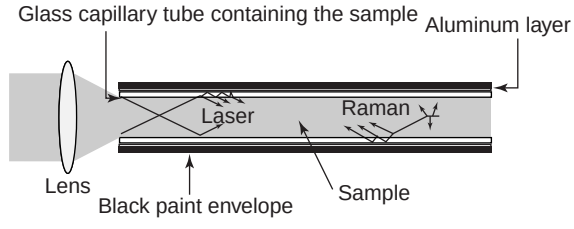

(f)

Advantages:
- Very efficient liquid optrode
- Easy set-up
- Can be associated with other optical fiber optrodes

Drawbacks:
- Can only used for liquids

Application:
- Low concentration solutions

Figure 44 (*Continued*)

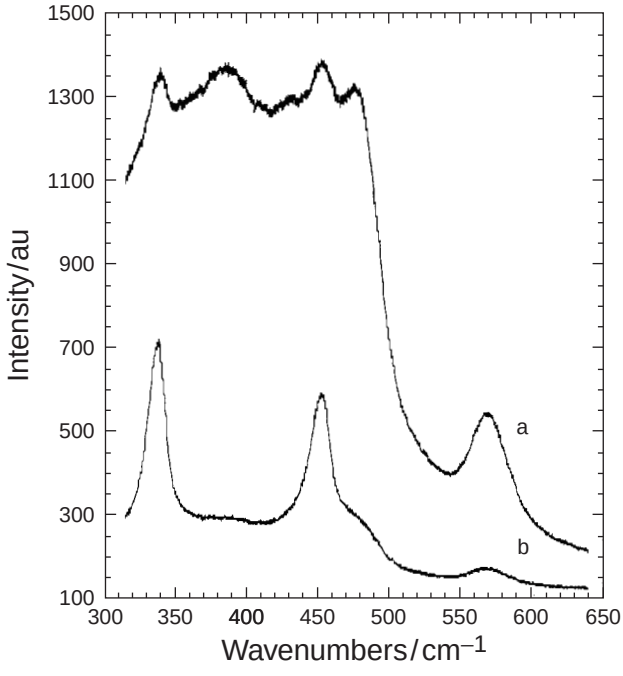

Figure 45 Comparison of the performance of some optrodes by two Raman laser fiber optics (RLFO) spectra of CCl_4 (laser power: 300 mW, collection and excitation fiber lengths: 100 m). (a) Obtained with geometry F2 of Figure 44 and (b) obtained with a Grin lens optrode F5 of Figure 44.

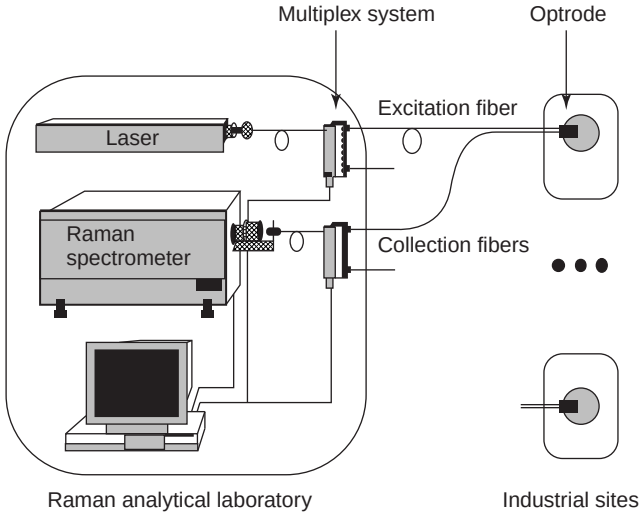

Figure 46 Typical RLFO installation.

RS has become a powerful analytical tool offering numerous possibilities thanks to its intrinsic advantages: that is it is nondestructive, sampling is easy, and remote and in situ control are possible for both qualitative and quantitative analyses. For qualitative analyses, Raman spectra in which Raman lines are narrow and well resolved represent the fingerprints of a chemical compound, so the identification of an unknown substance is fast and accurate. Gases, liquids, and solids can be studied by RS and are easily identified. Raman atlases[40–44] are also now available for numerous chemical compounds and can be included in the computer for an automatic search. For quantitative analytical purposes, Raman techniques can now rival other physical methods. The precision is about 1–3% and the DL can be some 10 ppm when special precautions are taken and special optrodes are used. SERS and RRS are the two methods which enable concentrations as little as parts per billion to be measured. For structural studies, high-resolution RS with the possibility of polarization where necessary, makes possible the investigation of vibrational states of gas, liquids, and solids either alone or complementary to IR and neutron diffusion methods. TRRS is one of the most elegant techniques for appraising the structure of excited states of molecules with a timescale down to the picosecond range. To examine small samples, the Raman microscope and especially Raman imaging are very efficient and fast, and distinctive Raman imaging is now a specific analytical technique. Applications are numerous, in physics, chemistry, biology, mineralogy, gemology, chemical industry, food sciences, petrochemistry, nuclear science, in the arts, environment, and so on. For industrial purposes, RS has made an important breakthrough since there is no sampling and the technique is simple and robust, fast, precise, very user-friendly especially when associated with optical fiber sensors, and has a very good quality/price ratio. Added to these qualities is the huge potential of its remote and in situ measurement possibilities. With the introduction of chemometrics for data treatment, complex problems can be solved, the accuracy and the DL have been greatly improved[45,46] and the problem of transfering analytical data sets between instruments is now being treated.[47]

DRSs were the first Raman instruments to be built and with progress in instrumentation, lasers, multichannel detectors, and gratings, they will continue to be of great use in the future especially with the possibility of the interchange of different components. An exhaustive list of commercial Raman instruments is available on the Internet.[48]

Other papers treating the subject ''Raman spectrometry'' are numerous. A nonexhaustive set of references which can be consulted for more information is given.[49–53]

Other techniques can only complement but not supplant DRS. The future lies in different directions:

- more versatility of DRS instrumentation;
- miniaturization with diode lasers and more robust and compact instruments with high performances are

planned; some portable Raman devices are already available;

- more sensitive detectors to increase the DL in order to achieve TRRS in the picosecond timescale for research in fundamental physics but also to eliminate fluorescence phenomena;
- multiplexing with the help of optical fibers;
- hyphenated analytical methods; Raman spectrometers can be used as detectors for high-performance liquid chromatography;[54]
- Raman microscopy and imaging in the NIR region.

ACKNOWLEDGMENTS

The author gratefully thanks Dr E. Da Silva, Dr M. Daudon, Dr M. Jouan, Dr J.-P. Perchard and Mrs F. Nguyen for having reviewed this article and suggested several improvements. I gratefully acknowledge Mrs P. Salvini's help in realising the tables and figures.

ABBREVIATIONS AND ACRONYMS

CARS	Coherent Anti-Stokes Raman Spectroscopy
CCD	Charge-coupled Device
CW	Continuous Wave
DC	Direct Current
DL	Detection Limit
DMF	Dielectric Multilayer Filter
DRM	Dispersive Raman Microscope
DRS	Dispersive Raman Spectrometer
FPC	Fabry–Pérot Resonance Cavity
FSR	Free Spectral Range
FT	Fourier Transform
fwhm	Full Width at Half-maximum
GE	Geometric Etendue
GR	Generation Recombination
HDG	Holographic Diffraction Grating
HF	Holographic Filter
HNF	Holographic Notch Filter
HPG	Holographic Plane Grating
HRS	Hyper Raman Scattering
IF	Interference Filter
IR	Infrared
KDP	Potassium Dihydrogen Phosphate
LD	Laser Diodes
MC	Monochromator
NA	Numerical Aperture
NF	Notch Filter
NIR	Near-infrared
OD	Optical Density
PC	Photon Converting Device
PDA	Photodiode Array
PMC	Premonochromators
PMT	Photomultiplier Tube
QE	Quantum Efficiency
RDG	Ruled Diffraction Grating
RLFO	Raman Laser Fiber Optics
RMS	Root Mean Square
RRS	Resonance Raman Spectroscopy
RS	Raman Spectroscopy
SCM	Scanning Confocal Microscope
SE	Stimulated Emission
SERS	Surface-enhanced Raman Spectroscopy
S/N	Signal-to-Noise Ratio
TEM	Transverse Electromagnetic Mode
TRRS	Time-resolved Raman Spectroscopy
UV	Ultraviolet
VIS	Visible
YAG	Yttrium Aluminum Garnet Crystal

RELATED ARTICLES

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Fourier Transform Raman Instrumentation • Raman Microscopy and Imaging • Raman Scattering, Fundamentals

REFERENCES

1. C.V. Raman, K.S. Krishnan, 'A New Type of Secondary Radiation', *Nature*, **121**, 501–502 (1928).
2. C.V. Raman, 'A New Radiation', *Indian J. Phys.*, **2**, 387–398 (1928).
3. C.V. Raman, K.S. Krishnan, 'Molecular Spectra in the Extreme Infra-red', *Nature*, **122**, 278 (1928).
4. R.L. McCreery, 'Instrumentation for Dispersive Raman Spectroscopy', in *Modern Techniques in Raman Spectroscopy*, ed. J.J. Laserna, J. Wiley & Sons, Chichester, 41–72, 1996.
5. W.G. Fateley, D.C. Tilotta, 'Raman Spectrometer having Hadamard Electrooptical Mask and Diode Detector', International Patent Classification G01N21/65, PCT/US88/02688 (8 August 1988).
6. A. Smekal, 'The Quantum Theory of Dispersion', *Naturwiss.*, 873–875 (1923).
7. D.L. Andrews, *Lasers in Chemistry*, 3rd edition, Springer, Berlin, 1–95, 1997.
8. K.S. Booksh, C.M. Stellman, W.C. Bell, M.L. Myrick, 'Mathematical Alignment of Wavelength-shifted Optical Spectra for Qualitative and Quantitative Analysis', *Appl. Spectrosc.*, **50**(2), 139–147 (1996).

9. K.W. Busch, M.A. Busch, 'Multielement Detection Systems for Spectrochemical Analysis', in *Chemical Analysis Series*, J. Wiley & Sons, New York, 369–377, Vol 107, 1990.
10. Oriel Corporation, 'Light sources, Monochromators and Spectrographs, Detectors and Detection Systems', *Fiber Optics*, **3–4**, (1994).
11. K.W. Busch, M.A. Busch, 'Multielement Detection Systems for Spectrochemical Analysis Series', in *Chemical Analysis Series*, J. Wiley & Sons, New York, 462–489, Vol 107, 1990.
12. R.L. McCreery, 'CCD Array Detectors for Multichannel Raman Spectroscopy', in *Charge Transfer Devices in Spectroscopy*, eds. J. Sweedler, K. Ratzlaff, M. Denton, VCH, New York, 227–279, 1994.
13. Oriel Instruments, 'New Multichannel Detection Systems', Stratford, CT, 1–5, 1998.
14. Princeton Instruments, 'The Leader in Spectroscopic Detection', Trenton, NJ, 29–93, 1997.
15. H.H. Willard, L.L. Merrit, Jr, J.A. Dean, F.A. Settle, Jr, *Instrumental Methods of Analysis*, 6th edition, Wadsworth Publishing, Belmont, CA, 846–857, 1981.
16. B. Yang, M.D. Morris, H. Owen, 'Holographic Notch Filter for Low-wavenumber Stokes and Anti-Stokes Raman Spectroscopy', *Appl. Spectrosc.*, **45**(9), 1533–1536 (1991).
17. 'DILOR Labram Infinity', Technical Report, DILOR, Lille, 1997.
18. P.L. Flaugh, S.E. O'Donnell, S.A. Asher, 'Development of a New Optical Wavelength Rejection Filter: Demonstration of its Utility in Raman Spectroscopy', *Appl. Spectrosc.*, **38**, 847–850 (1984).
19. S.A. Asher, P.L. Flaugh, S. Hagopian, P. Rundquist, G.R. Washinger, 'Development of a New Rayleigh Line Rejection Filter: The Basis for a New Generation of Raman Instrumentation', Xth International Conference on Raman Spectroscopy, eds. W.L. Peticolas, B. Hudsen, University Printing Department, University of Oregon, Eugene, OR, 20–23, 1986.
20. G.S. Hayat, G. Pieuchard, *Handbook of Diffraction Gratings, Ruled and Holographic*, Instrument SA, Jobin-Yvon, Longjumeau, France, 7–30, 1995.
21. J.M. Lerner, A. Thevenon, *The Optics of Spectroscopy, A Tutorial*, Instrument SA, Jobin-Yvon, Longjumeau, France, 1–56, Vol 2.0, 1988.
22. V.A. Fassel, 'Nomenclature, Symbols, Units and their Usage in Spectrochemical Analysis', *Pure Appl. Chem.*, **30**(1–2), 653–679 (1972).
23. 'DILOR Confocal Laser Raman, Part I: Physical Basis', DILOR, Lille, France, 1–49, 1991.
24. H.H. Willard, L.L. Merrit, Jr, J.A. Dean, F.A. Settle, Jr, *Instrumental Methods of Analysis*, 6th edition, Wadsworth Publishing, Belmont, CA, 35–37, 1981.
25. B. Schrader, 'Raman Spectroscopy of Slight Amounts of Solid and Liquid Substances', *Z. Anal. Chem.*, **197**, 295–309 (1963).
26. S.P.S. Porto, 'Angular Dependence and Polarization Ratio of the Raman Effect', *J. Opt. Soc. Am.*, **56**, 1585–1589 (1966).
27. C. Brémard, P. Dhamelincourt, J. Laureyns, G. Turrell, 'The Effect of High-numerical-aperture Objectives on Polarization Measurements in Micro-Raman Spectrometry', *Appl. Spectrosc.*, **39**(6), 1036–1039 (1985).
28. R.L. McCreery, M. Fleischmann, P. Hendra, 'Fiber Optic Probe for Remote Raman Spectrometry', *Anal. Chem.*, **55**, 146–148 (1983).
29. Nguyen Quy Dao, M. Jouan, P. Plaza, J. Robieux, 'Dispositif Notamment de Multi-mesures Pour Analyzer, à Distance et in situ, un Échantillon par Spectrométrie Raman', Brevet Français (French patent), **84**, 15 047, 1984.
30. S.D. Schwab, R.L. McCreery, 'Versatile, Efficient Raman Sampling with Fiber Optics', *Anal. Chem.*, **56**, 2199–2204 (1984).
31. R.D. McLachlan, G.L. Jewett, J.C. Evans, 'Fiber-optic Probe for Sensitive Raman Analysis', United States Patent, 4 573 761, 1986.
32. A. Cyr, P. Plaza, M. Jouan, E. Laviron, Nguyen Quy Dao, 'Analyse à Distance et in situ à la Surface d'une Électrode, par la Méthode RLFO Utilisant l'effet SERS', *C. R. Acad. Sci. Paris*, **314**(II), 43–49 (1992).
33. Nguyen Quy Dao, M. Jouan, 'The Raman Laser Fiber Optics Method and its Applications (Review)', *Sensors and Actuators B*, **11**, 147–160 (1993).
34. H. Nguyen Quang, M. Jouan, Nguyen Quy Dao, 'Use of a Mono-fiber Optrode in Remote and In Situ Measurements by the Raman Laser Fiber Optics Method', *Appl. Spectrosc.*, **47**(12), 2013–2016 (1993).
35. I.R. Lewis, P.R. Griffiths, 'Raman Spectrometry with Fiber-optic Sampling (Review)', *Appl. Spectrosc.*, **50**(10), 12A–30A (1996).
36. W.G. Fateley, D.C. Tilotta, 'Raman Spectrometer having Hadamard Electro-optical Mask and Diode Detector', International Patent Number W089/01622, 1988.
37. P. Jacquinot, 'The Luminosity of Spectrometers with Prisms, Gratings, or Fabry–Pérot Etalons', *J. Opt. Soc. Am.*, **44**(10), 761–765 (1954).
38. P. Fellgett, 'Theory of Multiplex Interferometric Spectrometry', *J. Phys. Radium*, **19**, 187–191 (1958).
39. J. Connes, P. Connes, 'Near-infrared Planetary Spectra by Fourier Spectroscopy. I. Instruments and Results', *J. Opt. Soc. Am.*, **56**, 896–910 (1966).
40. 'Sadtler Standard Spectra', Sadtler Research Laboratories, Philadelphia, PA, Vols. 1–11, 1976.
41. B. Schrader, *Raman/Infrared Atlas of Organic Compounds*, 2nd edition, ed. F. Ebel, VCH, New York, 1989.
42. Nguyen Quy Dao, M. Daudon, *Infrared and Raman Spectra of Calculi*, Elsevier, Paris, 1–348, 1997.
43. R.A. Nyquist, C.L. Putzig, M.A. Leugers, R.O. Kagel, 'The Handbook of Infrared and Raman Spectra of Inorganic Compounds and Organic Salts', *Infrared and Raman Spectral Atlas of Inorganic Compounds and*

Organic Salts, Academic Press, New York, Vol. 1, (paper support, 1997 ; CD-Rom support, 1998).

44. S.B. Engelsen, 'Raman Spectra of Carbohydrates', www/newton.foodsci.kvl.dk/specarb.html.
45. H. Nguyen Quang, B. Vu Thi, M. Jouan, Nguyen Quy Dao, 'Application of Partial Least-squares Regression to Remote Quantitative Analysis by the Raman Laser Fiber Optics (RLFO) Method', *Analusis*, **20**, 141–147 (1992).
46. P. Weinmann, M. Jouan, Nguyen Quy Dao, B. Lacroix, C. Groiselle, J.P. Bonte, G. Luc, 'Quantitative Analysis of Cholesterol and Cholesterol Ester in Human Atherosclerotic Plaques using NIR Raman Spectroscopy', *Atherosclerosis*, **140**, 81–88 (1998).
47. H. Nguyen Quang, M. Jouan, Nguyen Quy Dao, 'A Simplified Calibration Model of Spectral Data for Quantitative Analyses with Different Raman Spectrometers', *Anal. Chem. Acta*, **379**, 159–167 (1999).
48. J.R. Ferraro, K. Martin, R.J. Jarnutowski, 'Five-year Summary of Commercial Instrumentation for Absorption Spectroscopy (1992–1996)', www.spectroscopymag.com/ferarro.htm.
49. G. Placzek, *Rayleigh and Raman Scattering*, UCRL – TRANS – 526 (L), US Atomic Energy Comm., 206, 1962.
50. B. Schrader, 'New Developments in Raman Spectroscopy', in *Proceedings Colloquium Spectroscopy International*, eds. I. Rubeska, J. Mostecky, Z. Ksandr, Prague, 373–383, 1977.
51. B. Schrader, 'Raman Spectrometry', *Z. Anal. Chem.*, **314**(4), 363–382 (1983).
52. B. Schrader, *Infrared and Raman Spectroscopy: Methods and Applications*, Verlag Chemie, New York, 1995.
53. G. Turrell, J. Corset, *Raman Microscopy: Developments and Applications*, Academic Press, New York, 1996.
54. T.D. Nguyen Hong, M. Jouan, Nguyen Quy Dao, M. Bouraly, F. Mantisi, 'Coupling of High-performance Liquid Chromatography with Raman Spectrometry', *J. Chromatogr. A*, **743**, 323–327 (1996).

Fourier Transform Raman Instrumentation

Howell G.M. Edwards
University of Bradford, Bradford, UK

1 Historical Introduction **1**
2 Experimental Instrumentation for Near-infrared Fourier Transform Raman Spectroscopy **3**
2.1 Lasers 3
2.2 Sample Illumination and Collection Optics 5
2.3 Michelson Interferometer 7
3 Operation of a Fourier Transform Raman Spectrometer **8**
4 Comparison of Conventional Raman and Fourier Transform Raman Spectroscopy **9**
5 Sample Accessories **11**
5.1 Raman Microscopy 11
5.2 Remote Sensing Probes 14
5.3 Hot- and Cold-temperature Accessories 15
5.4 Other Accessories 16
6 Other Instrumental Factors **17**
Acknowledgments **20**
Abbreviations and Acronyms **20**
Related Articles **20**
References **20**

The physical basis of the instrumentation required for the recording of Raman spectra in the near-infrared region of the electromagnetic spectrum using an interferometer and Fourier transform (FT) techniques is described. The advantages and disadvantages of the technique compared with the recording of Raman spectra in the visible or ultraviolet (UV) region are described. The impetus provided by the nondestructive nature of this analytical method, combined with the minimization of fluorescence and the ease of sample presentation to the spectrometer is illustrated with examples from archaeological science, environmental science, and biology; accessories such as a Raman microscopic and a remote sensing fibre-optic probe are described with relevant applications in art conservation, materials analysis, processing and biomedical diagnostics. The potential for future applications of the technique in chemical, physical and biological sciences is outlined.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 HISTORICAL INTRODUCTION

Since the original discovery by Raman in 1928 of the effect which bore his name,[1] spectroscopists had endeavoured to observe Raman scatttering using different wavelengths of excitation. Early restrictions imposed by the photographic plate detectors and irradiance of the sources of radiation effectively led to near-UV, blue and green wavelengths being used and the adoption of mercury arcs, in particular 435.8 and 546.1 nm excitation, as the standard Raman spectroscopic sources of illumination. The search for longer wavelength sources was intrinsically dependent on novel methods of sensitization of photographic emulsions and faster cameras; despite this, some classic work was undertaken using helium lamps and rubidium arcs, which extended the range of excitation sources into the red and made accessible for the first time the recording of Raman spectra of orange and red compounds. The spectroscopic advantages of the interferometer compared with dispersive spectrometers (prisms and gratings) were soon appreciated in the infrared region[2] and in 1964 a seminal paper by Chantry et al.[3] reported the observation of Raman spectra using near-infrared excitation and a Michelson interferometer. However, the inherent weakness of the Raman effect and instrumental deficiencies meant that it was 1986 before FT Raman interferometric spectroscopy using near-infrared excitation was born;[4] in the intervening years, dramatic improvements had been made in radiation sources (lasers), photon detectors, interferometer design and computer-control of Raman spectrometers. It is perhaps also highly significant that the first laser-excited (633 nm) Raman spectrum appeared in this period.[5]

The stage was set for the revitalization of Raman spectroscopy and the last decade has seen the application of the technique to problems in chemical spectroscopy which hitherto had not been dreamed of. Further applications were forthcoming with the advent of Raman microscopy, remote sensing fiber optic couplings and rapid scanning techniques.

Although the major advantages of Fourier transform Raman spectroscopy (FTRS) in the near-infrared was quickly identified with fluorescence minimization or rejection and the ability to record spectra from highly colored materials, several disadvantages were noted. Because of the ν^4 (where ν is the wavenumber of the laser excitation) scattering dependence with observed spectral intensity, the Raman effect is some 25 times weaker at 1064 nm in the near-infrared than it is at 488 nm, the blue radiation of an argon ion gas laser. However, the multiplex advantages of interferometric spectroscopy could be used to offset this (see later) – especially through the use of multiple spectral scanning and co-addition by computer-controlled and laser-referenced instrumentation. A more

subtle disadvantage in near-infrared excitation of Raman spectra involved broad-based absorption by atmospheric water vapor or by aqueous media around 1000 nm and Stokes-shifted wavelengths (Figure 1).[6] However, careful experimental arrangements are now possible in which the Raman spectra of aqueous solutions and of hydrated materials can be obtained successfully using 1064 nm excitation; some examples of this will be provided to demonstrate the quality of these spectra which have been obtained despite the presence of water absorption bands. An obvious disadvantage in near-infrared excitation of Raman spectra becomes manifest when heated samples are being studied, or when sample material starts to absorb at the excitation wavelength; the onset of thermal emission as a broad background can be seen (Figure 2), starting at higher wavenumber shifts, but gradually affecting the lower wavenumber region as the sample temperature increases. It is now generally recognized that the upper temperature limit for the recording of FT Raman spectra of heated materials is about 220 °C; this temperature value refers to the observation of low wavenumber shifts and, as expected, the ν(CH) modes of organic compounds in the spectrum are swamped by thermal emission at much lower temperatures.

In this article, the instrumentation required to undertake FTRS will be described; the importance of the technique in furthering applications of Raman spectroscopy to interdisciplinary problems in

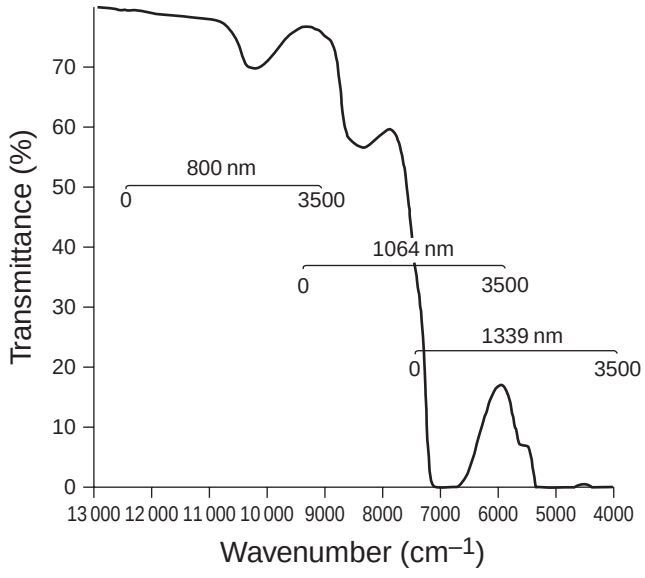

Figure 1 Absorption spectrum of water in the near-infrared region; the line-bars show the 0–3500 cm^{-1} shifts for Stokes Raman spectra excited with 800 nm, 1064 nm and 1339 nm lasers. (Reproduced by permission from Edwards.[6] Copyright J. Wiley & Sons.)

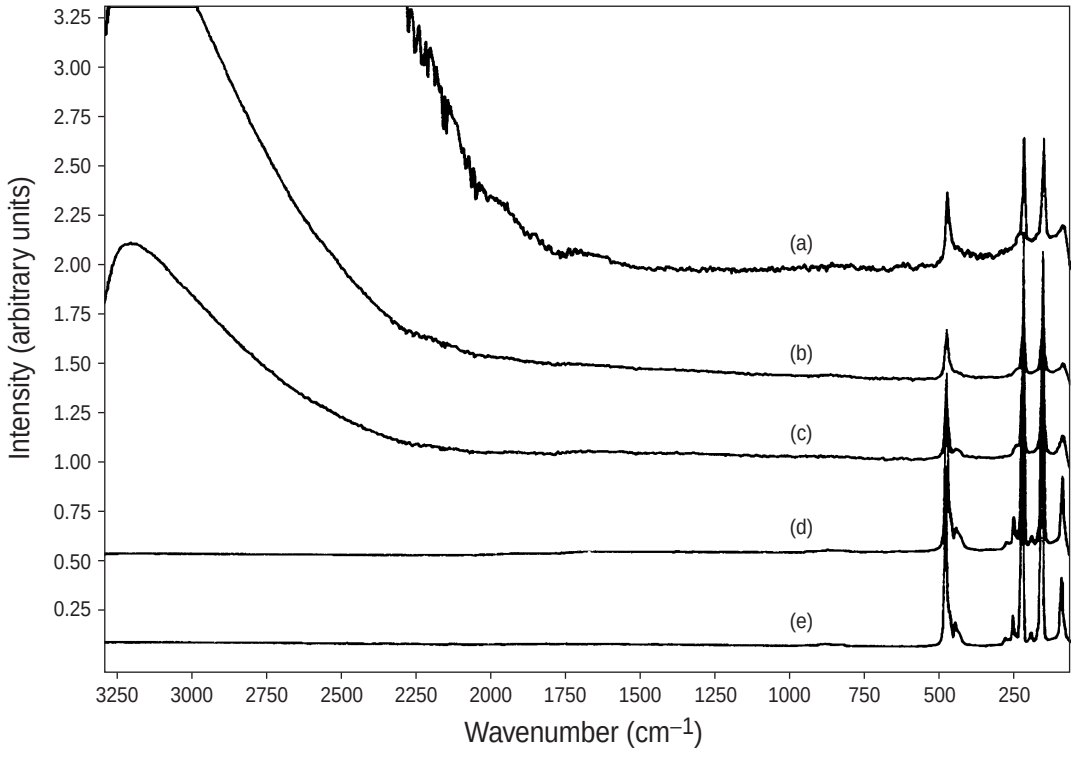

Figure 2 Onset of thermal emission in an FT Raman spectrum from a heated sample; sulfur powder; 1064 nm excitation; 200 scans; 4 cm^{-1} spectral resolution: (a) 200 °C; (b) 175 °C; (c) 150 °C; (d) 100 °C; (e) 25 °C.

archaeological science, environmental science, planetary and extraterrestrial studies, biomedical diagnostics and pharmaceutical science will be illustrated. Several important accessories will be described in relation to some novel spectroscopic problems.

2 EXPERIMENTAL INSTRUMENTATION FOR NEAR-INFRARED FOURIER TRANSFORM RAMAN SPECTROSCOPY

A photograph and schematic diagram of a typical FT Raman spectrometer, in this case a Brüker Equinox 55/S infrared instrument with FRA106 Raman module attachment, is shown in Figure 3. It will be appropriate to describe here the major components of such an instrument and the special accessories which have been constructed to diversify an FT instrument for Raman spectroscopic applications. The other type of FT Raman instrument available commercially is the dedicated, free-standing unit exemplified by the Bruker RFS 100/S system shown in Figure 4.

2.1 Lasers

The selection of a suitable laser wavelength in Raman spectroscopy is normally undertaken with two criteria in mind; the intensity of Raman scattering which can be generated and the avoidance of fluorescence. The former depends on the so-called fourth-power law, in that the intensity of Raman scattered radiation is proportional to ν^4 where ν is the wavenumber of the laser excitation. With commercial lasers now available from the deep UV

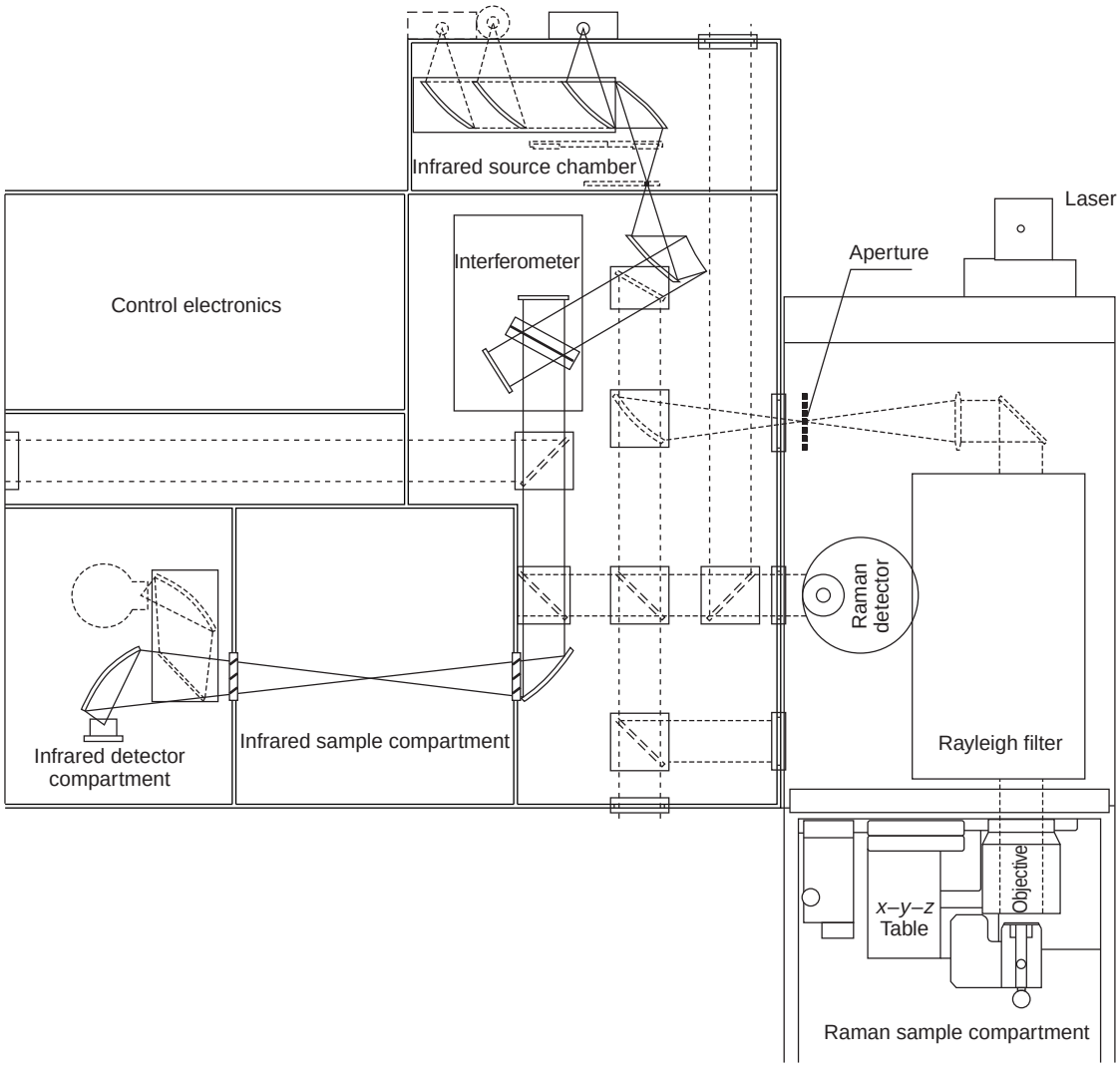

Figure 3 Diagram of a FT infrared spectrometer which has been adapted for FTRS: the Bruker Equinox 55 infrared spectrometer with FRA 106 Raman Module attachment. (Courtesy of Dr. P. Turner, Bruker Spectrospin Ltd., UK.)

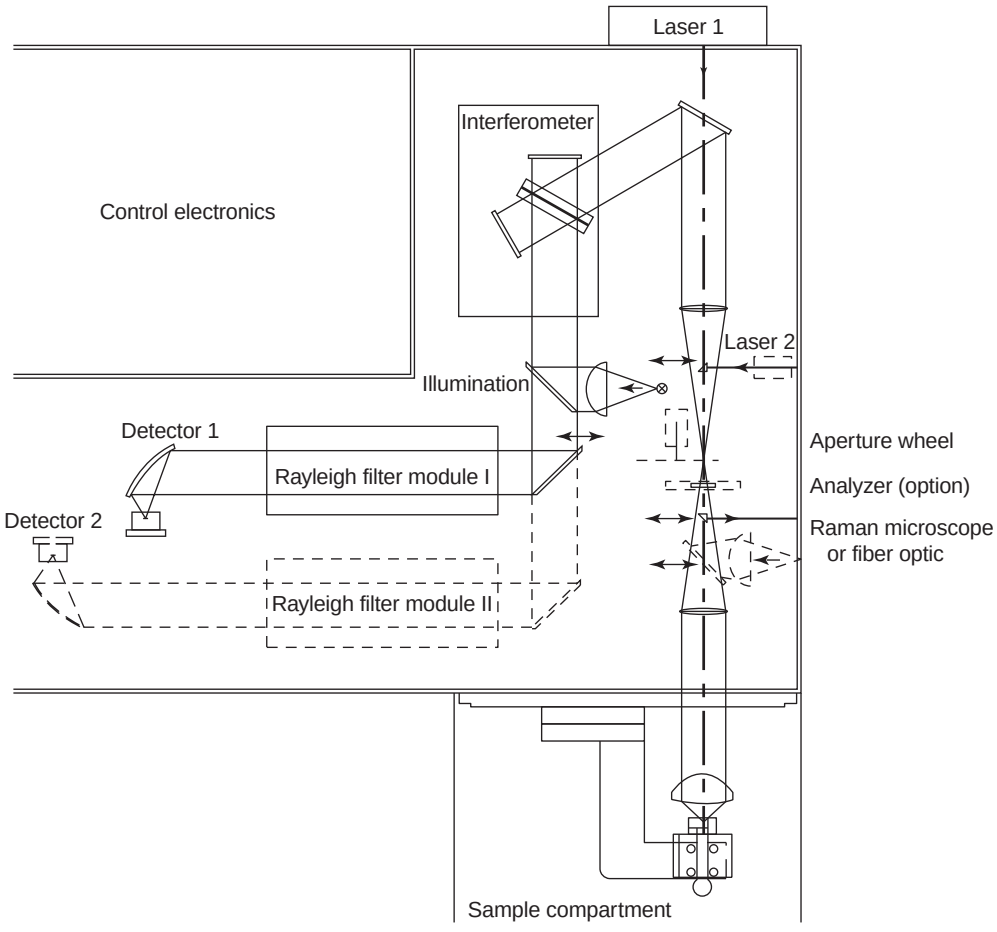

Figure 4 Diagram of dedicated, free-standing FT Raman spectrometer; the Bruker RFS 100 system. (Courtesy of Dr. P. Turner, Bruker Spectrospin Ltd., UK.)

to near-infrared wavelengths, some of which are tunable across the visible region of the electromagnetic spectrum, this means that for similar laser powers, excitation at 250 nm in the UV produces 730 times an intensity of Raman scattering as does excitation at 1300 nm in the infrared, representing the currently used wavelength limits in Raman spectroscopy. Since the Raman effect is so intrinsically weak, therefore, the question should be posed as to why Raman spectroscopists employ instrumentation with near-infrared laser excitation as typified by the topic of this article, FTRS? The answer to this question involves the addressing of the second of the criteria stated above: the minimization of fluorescence – the excitation of which invariably swamps the weaker Raman scattering process. By selection of a laser at higher wavelengths, the energy of excitation is decreased, so that the onset of laser adsorption and the concurrent fluorescence is avoided. This is especially important where complex organic molecules with several low-energy excited states are being studied. Therefore, it is not merely a question of what laser is available and at what cost: the selection of an excitation wavelength is critically important for Raman spectroscopic applications and several examples will be provided to illustrate this aspect. An example is shown for a cyanine dye in Figure 5,[7] excited by lasers operating at 488, 647 and 1064 nm. The move towards longer wavelength visible excitation in Raman spectroscopy is not always successful; even at a helium–neon excitation wavelength of 633 nm, fluorescence can still be troublesome. Using diode lasers at 780 nm or 840 nm, the rejection of fluorescence is better, but now the photodiode detector response to the scattered radiation at these longer wavelengths becomes a problem. This detector efficiency is reduced to almost zero by 1000 nm (1 μm), where the near-infrared properly begins.

The advent of FTRS arose with the ready availability of a Nd^{3+} : YAG (neodymium : yttrium aluminum garnet) laser excitation wavelength in the near-infrared region at 1064 nm; the characteristics of this laser were ideal for Raman spectroscopy in that it was not expensive, and with a frequency of 9395 cm^{-1} the excitation of fluorescence

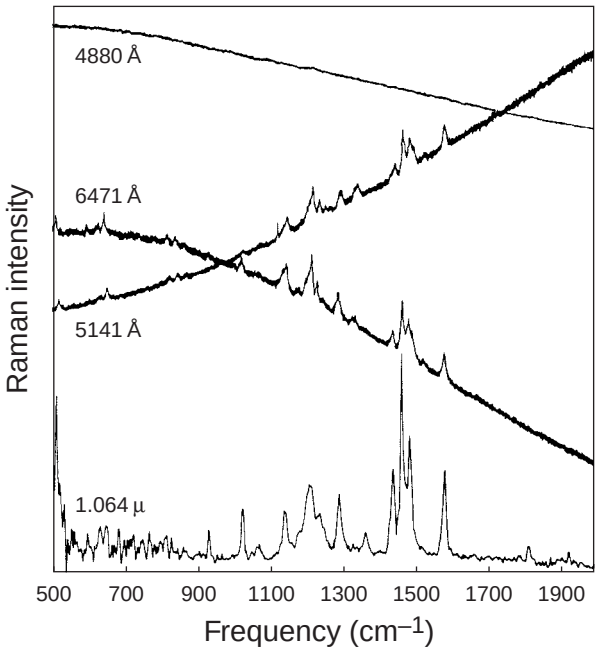

Figure 5 Raman spectra of cyanine dye excited by laser wavelengths at 488, 514, 647 and 1064 nm; the effect of fluorescence is seen to swamp the Raman spectrum completely in the blue. In the yellow and red regions, the Raman spectra can be seen as weaker bands on top of the broad band fluorescence emission. Only in the near-infrared is the Raman spectrum observed without fluorescence effects. (Reproduced by permission from Chase and Rabolt.[7] Copyright Academic Press.)

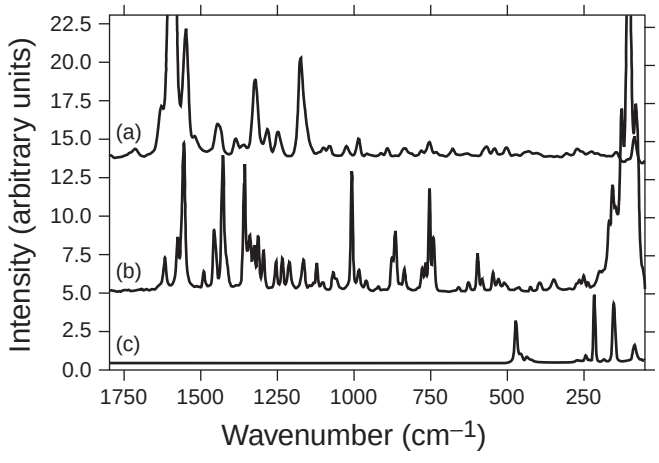

Figure 6 FT Raman spectra of (a) scytonemin; (b) tryptophan and (c) *p*-hydroxybenzaldehyde; 1064 nm excitation, 50–1800 cm^{-1}; 200 scans; 4 cm^{-1} spectral resolution. Note the flat Raman spectral background for the highly fluorescent scytonemin. Using 1064 nm excitation, at the frequency-doubled 532 nm, the fluorescence emission is so intense that it is used in epifluorescence microscopy to locate cyanobacteria.

from organic molecules was significantly reduced. An example is shown in Figure 6 for scytonemin **(1)**, an important natural product which is highly fluorescent, yet which provides an FT Raman spectrum of excellent quality.

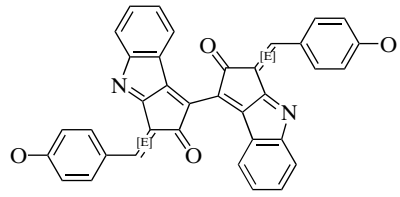

(1)

Several other lasers operating at long wavelengths rival the choice of the Nd^{3+} : YAG option, but using charge-coupled devices (CCDs) with increased quantum efficiencies of detection; these include the Ti–sapphire laser, tunable from about 700 nm to 1000 nm.

The most recent attempt to record FT Raman spectra at relatively longer wavelengths has been reported by Asselin and Chase[8] who used 1300 nm excitation with a Nd^{3+} : YAG laser to record FT Raman spectra of a highly fluorescent polyimide for which Raman spectra at 1064 nm were spoiled by fluorescence emission. The fluorescence was significantly reduced at 1300 nm, but the spectral quality was measurably poorer (Figure 7); this was ascribed to the decrease in Raman scattering intensity between 1064 and 1300 nm (×2), the lower laser power used (×20) and greater background noise in the InSb detector.

Hence, at the current time, the laser of choice for FTRS is the Nd^{3+} : YAG system operating at 1064 nm. The laser is optically pumped by either a lamp or a diode system; in both cases, nonlasing lines are present which need to be filtered lest they interfere with the observation of the Raman spectra with the production of so-called laser-line artefacts. A simple optical notch-filter is normally used to filter out these ''spurious'' features.

Most FTRS is carried out using continuous wave (CW) lasers, in which the stability of the laser minimizes source fluctuation noise[7] and enables the multiplexing of spectral data acquisition to take place. Experiments in FTRS have been carried out with a rapid-pulsing, mode-locked Nd^{3+} : YAG laser system operating at megahertz frequencies with relatively slower (kilohertz) detector responses and have been successful. A major advantage in the use of pulsed laser excitation is the rejection of continuous thermal emission from heated samples in the near-infrared (Figure 8); Figure 8 shows the effect of recording the FT Raman spectrum of $CaSO_4$ at 140 °C using CW and pulsed Nd^{3+} : YAG laser excitation at 1064 nm.

2.2 Sample Illumination and Collection Optics

The purpose of the sample illuminator is the imaging of the laser beam onto the sample, the collection of

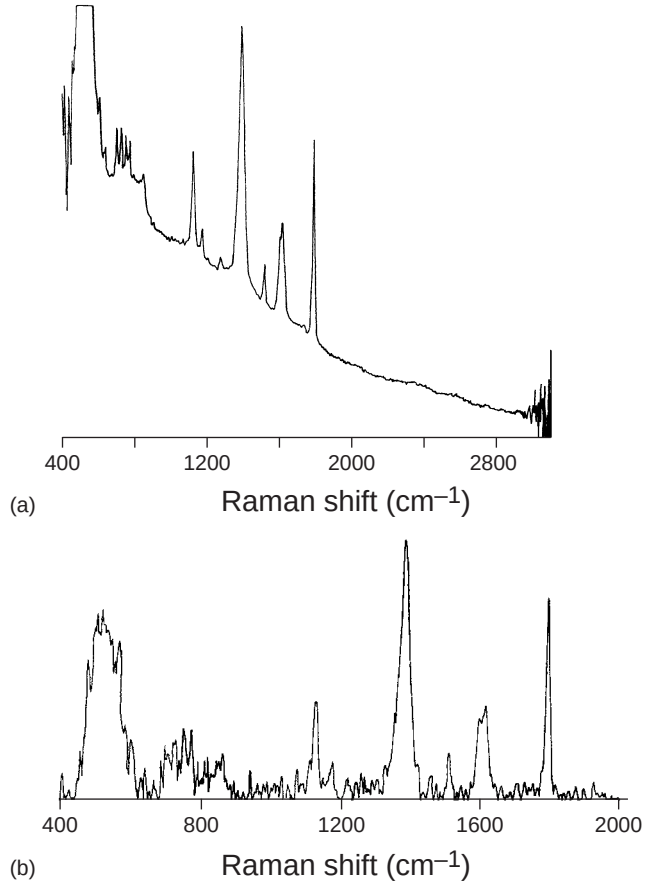

Figure 7 FT Raman spectra of a fluorescent polyimide (a) with 1064 nm excitation and (b) with 1.3 μm excitation. The fluorescence background present at 1064 nm and rejected at 1.3 μm should be noted. (Reproduced by permission from Chase and Rabolt.[7] Copyright Academic Press.)

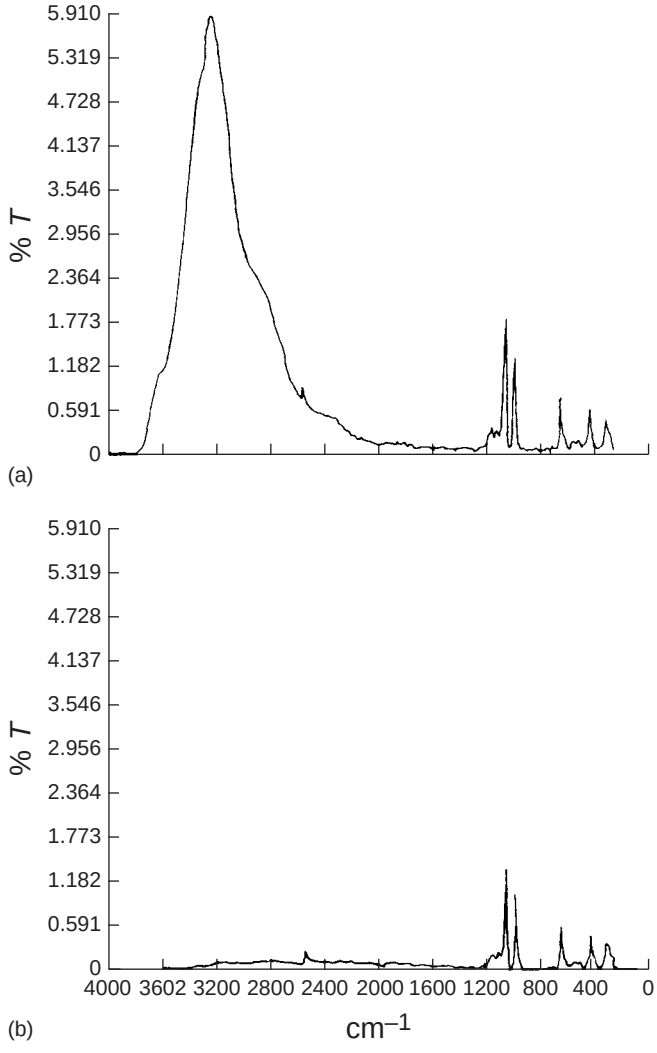

Figure 8 FT Raman spectra of calcium sulfate excitation: (a) CW operation; (b) Q-switched pulsed operation. Note the presence of broad-band infrared emission in the upper spectrum which has been effectively discriminated against in the lower spectrum. (Reproduced by permission from Chase and Rabolt.[7] Copyright Academic Press.)

the scattered radiation and the imaging of this into the interferometer. In FTRS very high imaging and collection efficiencies can be achieved using 180° back scattering and large apertures, commonly $f/0.7$, often using confocal reflectors. If the quantity of sample solid or liquid is small, little of the scattered solid sample is blocked; however, in certain cases, an off-axis collection system may be used to avoid the problem of sample obfuscation of the scattered radiation.

An important instrumental difference between conventional Raman spectrometers and FT Raman interferometers of relevance to a discussion of sample irradiation and collection arrangements relates to the dimensions of the scattered image that is presented to the interferometer. The slit in a conventional spectrometer is a circular aperture in an interferometer; for a typical interferometer operating at 1 cm^{-1} resolution in the near-infrared, the dimension of the entrance aperture is about 4 mm for a spot size at the sample of 1 mm contrasted with a nominal slit width of about 100 μm at 500 nm used in a conventional spectrometer with visible excitation, with a focal cylinder "waist" at the sample of 25 μm. Hence, the extreme focusing of the laser beam in FTRS is neither necessary nor desirable, and potentially large power densities at the focus can be reduced with consequent minimization of sample degradation through thermal effects.

2.2.1 Filters

The scattered radiation in an FT Raman spectrum also contains the Rayleigh scattered radiation at the laser frequency; the latter is up to eight orders of magnitude larger in intensity than the Raman scattering, hence it must be removed lest it causes damage to or saturation of

the detector. Several systems are commonly used for this, including dielectric filtering, in which a multilayer-coated plate is highly absorbing at the laser frequency but transmits at longer wavelengths, typically about 200 cm^{-1} from the laser line. Because of inherent problems in the generation of spectral features and background ''noise'' with dielectric filters, other systems have been discussed, such as holographic notch filters with improved spectral performance. An example of the use of such a filter is seen in the FT Raman spectrum of sulfur powder (Figure 2) where the cut-off is below 50 cm^{-1} from the exciting line. Several sophisticated improvements in optical filter technology involving systems of four interference filters (chevron-type) or semiconductor-edge filters have reduced the cut-off to less than 40 cm^{-1} from the laser line wavelength.

2.3 Michelson Interferometer

The major advantage of FT methods in Raman spectroscopy is undoubtedly the multiplexing of the system compared with the sequential scanning of spectral elements in conventional spectrometer.[9] This multiplex advantage arises from the simultaneous measurement of multiple spectral signals from a single detector signal. The replacement of a conventional diffraction grating by a Michelson interferometer achieved this advantage in a very realizable way. The Michelson interferometer (Figure 9) consists of a beam-splitter set at 45° to the beam of scattered Raman radiation, which is thus split into two mutually orthogonal beams. One of these meets a fixed mirror and is returned along its optical path meeting the other beam which has been reflected at a moving mirror – the position of which affects the phase of the resulting signal, the interferogram. The latter is composed of a series of periodic cosine waves whose amplitudes are dependent on the signal intensity. Analysis of the interferograms is accomplished using Fourier transformations.

The spectral resolution of the interferometer is $1/x$, where x is the pathlength of the interferogram; a 0.5 cm pathlength hence gives a spectral resolution of 2 cm^{-1} at 1064 nm, the maximum achievable with the Bruker IFS66/FRA 106 instrument described earlier. This can be correlated with a conventional spectrometer resolution of 2 cm^{-1} at 500 nm, achieved using a slit width of 100 μm and a reciprocal linear dispersion of 20 cm^{-1} mm^{-1} with holographic or ruled diffraction gratings as exemplified by a Spex Model 1401 double monochromator.

Since radiation from all spectral elements over the desired wavelength range is measured in one interferogram scan, the signal-to-noise ratio (S/N) can be increased interferometrically (Fellgett advantage) by the rapid coaddition of scans; for example, modern interferometers can achieve 60 interferogram scans per second, and calculation reveals that a 30-min spectral accumulation for a 2000-cm^{-1} range therefore gives a S/N improvement of about 50 times over the single-scan dispersive system at 1 cm^{-1} s^{-1}.[9]

Additionally, because of the longer wavelengths in the infrared region the precision required for the mirror movements is readily achievable. However, because of the sensitivity of photomultiplier and CCD detectors in the visible region, FT methods are not so advantageous and UV/visible spectrometers still tend to be dispersive-grating systems. The near-infrared region shows a large frequency range compared with other electromagnetic spectra considered here, (Figure 10).

For Raman spectroscopy, interferometers have several advantages over monochromators of similar aperture where radiation must pass through narrow slits (typically

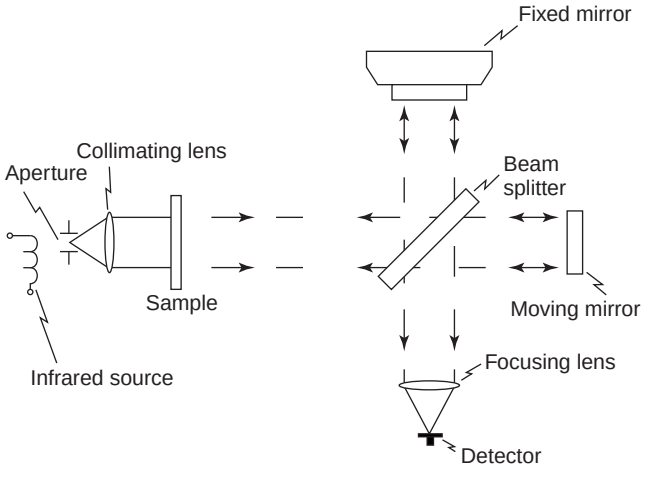

Figure 9 Layout of a Michelson interferometer; the basis of FTRS. (Reproduced by permission from Edwards.[6] Copyright J. Wiley & Sons.)

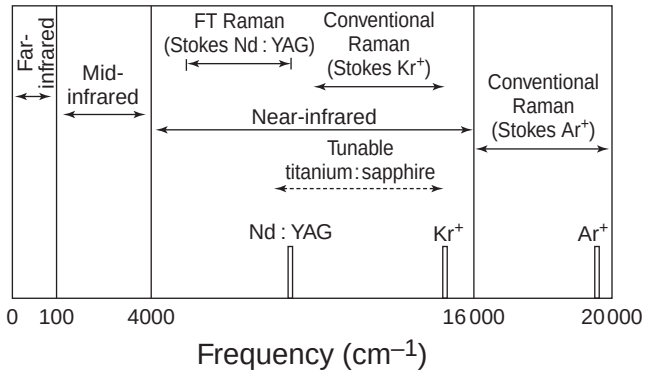

Figure 10 Frequency range (cm^{-1}) of the electromagnetic spectrum with FT Raman and conventional Raman spectroscopic ranges superimposed. (Reproduced by permission from Chase and Rabolt.[7] Copyright Academic Press.)

50 to 100 μm cm^{-1} spectral bandpass) compared with the larger optical conductance of an interferometer (the Jacquinot advantage). For a spectral resolution of 4 cm^{-1}, for example, the étendue (radiation throughput) of an interferometer may be 100 times better than that of the monochromator.[7] Other interferometric properties such as the Connes advantage which arises from the coaddition reproducibility and superimposability of data accumulation, means that in the infrared region the interferometer is supreme for Raman-scattering experiments.

Although it is generally recognized that the interferometer is the key optical component in an FT Raman system, the design leaves little room for adaptability or maneuver. The critical parameters in an interferometer are the transmission and the modulation efficiency.

2.3.1 Transmission

The transmission of the instrument is readily evaluated using laser scattering from standard samples using power measurements at the interferometer entrance aperture and at the detector. Typically, transmissions of about 5–20% are experienced for the near-infrared region; major losses in transmission occur at reflective surfaces and diaphragm apertures. Optical alignment is hence critical, especially of beam-splitters, and the cleanliness of optical surfaces is vital; purging of an interferometer using dry, filtered, recycled nitrogen will assist in the exclusion of dust and water vapor, whose infrared absorptions near 1.3 μm can be troublesome.

2.3.2 Modulation Efficiency

This parameter indicates how effectively the scattered radiation can be analysed. Unmodulated radiation does not only result in a decrease in intensity of the collected signal, but also has an adverse effect on noise levels. The beam-splitter efficiency is critical and for a perfect optical component the modulation efficiency is 100%. A significant problem in this regard is the wavelength dependency of the modulation efficiency; however, instrumental efficiencies of about 80% are readily achievable unless broader ranges of spectral coverage are required. Extension of optical range results in a lower modulation efficiency.

2.3.3 Detectors

The requirements of a detector for FTRS are a high quantum efficiency in the infrared and low background noise, and also a useful spectral range which covers as much as possible of the (Stokes) Raman spectrum. For 1064 nm excitation (9395 cm^{-1}), the 0–3600 cm^{-1} Stokes Raman shift extends to 5800 cm^{-1}. Ideally, a detector should have a high response over the whole range; clearly, in practice, this is not so as Figure 11, which gives the instrument response function for an InGaAs detector at several temperatures, demonstrates. Generally, an extended spectral range will involve an increase in detector noise, which increases the further one goes into the infrared. Cooling the detector to liquid-nitrogen temperatures (77 K) results in significantly lower noise levels, but this is often achieved at the expense of spectral range coverage. For example, Figure 11 shows how the detector sensitivity "cutoff" is decreased from 3400 cm^{-1} on cooling from 250 K to 77 K. The trade-off is in a five-fold improvement in detector sensitivity for the narrower spectral range observed with a detector maintained at low-temperatures.

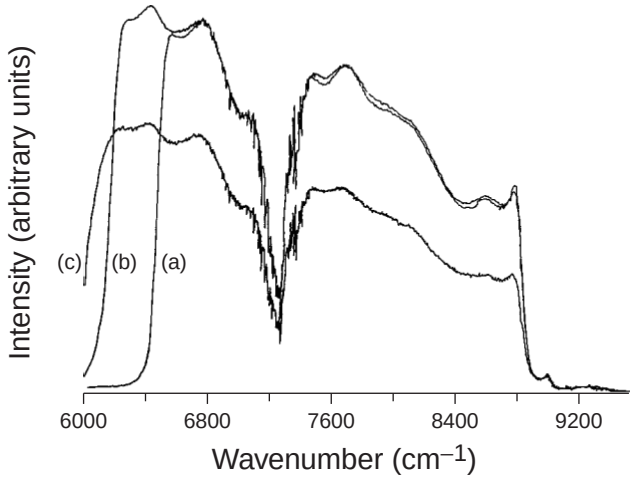

Figure 11 Instrument response function for an InGaAs detector over the 6000–9400 cm^{-1} region in the near-infrared; 1064 nm laser radiation has the wavenumber 9395 cm^{-1}. The variable response of the detector as a function of frequency in the Stokes shift from 9395 cm^{-1} should be noted, as should the sharp cutoff at (a), (b), (c) between 6000 and 6400 cm^{-1}. Spectrum c is the upper temperature response function (250 K) and a is the same response at liquid-nitrogen temperature (77 K), whereas b is intermediate. (Reproduced by permission from Chase and Rabolt.[7] Copyright Academic Press.)

3 OPERATION OF A FOURIER TRANSFORM RAMAN SPECTROMETER

Because of the interdependence of laser source, sample alignment, collection optics setting, interferometer and detector, it is vital to undertake the routine set-up and commissioning of an FT Raman spectrometer before experiments are undertaken to ensure that the instrument is operating at its best. In our laboratories, we use a standard sample of rhombic sulfur, S_8, packed into an aluminum cup.

An initial interferogram of sulfur with a recorded amplitude of 32 K counts per second (cps) for the scattered radiation at a laser power setting of 100 mW is typical. Reduction in signal indicates bad interferometer alignment or a misplaced sample illumination and collection optical arrangement. Signal-to-noise measurements on standard spectra will indicate whether or not the detector is operating with maximum efficiency and sensitivity.

The ratioing of the observed signal to a standard, recorded white-light spectrum is strongly advised for quantitative measurements involving Raman band intensities; for this to be effective, an identical or closely similar optical path for the white-light radiation and that of the Raman experiment should be achieved. White-light radiation is usually generated via a swing-in-mirror situated in the interferometer and use of a tungsten lamp source imaged onto a powdered sample of potassium bromide. Another method involves the collection of scattered white light illumination within the interferometer.

Researchers who approach FTRS from a conventional spectroscopic background using visible excitation will find several points worthy of note:

- Because of the intrinsic lower energies of the infrared laser, higher power levels are feasible in FTRS compared with those used in the visible; this is especially important for highly colored, absorbing, materials and biological materials which are subject to partial degradation or even destruction using visible laser wavelengths.
- It is not as necessary to focus the laser beam in the near-infrared interferometric experiments, whereas this is almost mandatory in their visible counterparts – often, this is accompanied by sample heating effects in the visible.
- The Raman spectra from samples at elevated temperatures are not subjected to the effects of background thermal emission as are their counterparts in the near-infrared. For black-body emission to affect sample Raman spectra in the visible region the temperatures must be high; this is not the case for FTRS, where the realistic experimental upper limit for heated samples is approximately 220 °C for the low wavenumber shift region, and even lower temperatures for vibrational modes in the 3000 cm^{-1} wavenumber shift region.
- Aqueous solutions demonstrate the effect of near-infrared radiation absorption and consequent sample heating which arises from the first overtones of the ν(OH) stretching modes in this region; hence, even if the solute material does not absorb laser radiation at 1064 nm, the aqueous solvent can produce a heating effect.
- In conventional Raman spectroscopy with visible excitation there is frequently a clearly observable spectral improvement in signal-to-noise levels away from the Rayleigh line. Because of the multiflex spectral arrangement in interferometry, however, all signal wavelengths and noise are detected simultaneously. Hence, an interferometer which has not been properly set-up will produce poor signal-to-noise levels throughout the Raman spectra, irrespective of the wavenumber shift region being examined.

4 COMPARISON OF CONVENTIONAL RAMAN AND FOURIER TRANSFORM RAMAN SPECTROSCOPY

The point has already been made in this article that for Raman spectroscopic studies the operation of a system with a near-infrared excitation wavelength (e.g. Nd^{3+} : YAG laser at 1064 nm) obviates the problems associated with fluorescence excitation experienced in the blue and green regions of the visible.[6,10] This is an advantage for the examination of the Raman spectra of fluorescent organic species such as anthracene and polyurethanes and inorganic materials such as uranium salts. In addition, Raman spectra of highly colored absorbing materials such as potassium permanganate, iodine, and metal phosphines have been obtained for the first time using FTRS with infrared excitation. On the negative side, however the ν^4 intensity law for Raman scattering means that the intensity of Raman scattering decreases with increase in wavelength from the visible to the infrared; hence, the Raman scattering from 435.8 nm mercury-arc excitation is about 40 times as intense as that from a Nd^{3+} : YAG laser operating at 1064 nm, power for power. With generally lower powers for infrared lasers compared with the argon-ion system operating near 500 nm, this is compounded even further.

Despite this intrinsic disadvantage, however, the speed and ease of sampling occasioned by the FT spectrometers now make these attractive instruments for both Raman and infrared spectroscopic studies of molecular systems. The advent of FTRS and CCD Raman spectroscopy has truly provided a renaissance for the Raman technique through a wider range of applications now available.

The detailed comparison of conventional disperse Raman with the FT Raman spectroscopic technique is difficult because of the correlation of the many instrumental factors involved in each technique. However, the major differences are as follows:

- Moving to near-infrared excitation at 1064 nm from the visible reduces the Raman scattering intensity

(dependent on the fourth power of the excitation wavelength) by up to 40 times on 435.8 nm excitation.

- An advantage of near-infrared excitation is the reduction in energy of the lasing transition from about 25 000 to 10 000 cm^{-1}, which therefore inhibits the onset of fluorescence; this is of supreme importance in the application of FT Raman techniques to biological materials and to biomedical diagnostics.
- In a scanning Raman monochromator system only a small portion of the dispersed spectrum is analysed at any one time. With the interferometer, however, the Fellgett advantage applies; this is based on the acquisition of all spectral-resolution elements simultaneously. This greater "speed" of acquisition by an interferometer over a dispersing spectrometer is quantified as $\sqrt{r}$, where r represents the number of resolution elements in the spectrum.
- The Jacquinot throughput is much greater for the interferometer than for scanning monochromators in any spectral range since the limitation of the spectral slit is removed. The J stop in a typical interferometer is about 8 mm diameter compared with a typical 0.1 mm (100 μm) slit width in a spectrometer system.
- Internal calibration of the interferometer against a helium–neon laser (Connes advantage) provides exceptional wavenumber reproducibility, which facilitates the superposition of spectral data (accumulation) and data subtraction (background, solvent, etc.). Hence, although all the spectral data in an FT Raman experiment may be provided in one scan of about 2 s, multiple spectral-data accumulation to improve the S/N is desirable. In cases of very weak Raman spectra, especially from biological materials for which the incident laser power has been reduced to less than 20 mW from 1 W to minimize sample degradation, accumulations of many hours duration are acceptably realistic; for example, the first FT Raman spectroscopic analysis of archaeological human tissue of the "Ice-man", a 5200-year-old corpse, has been provided with up to 12 000 accumulations of spectral scans at 4 cm^{-1} resolution from a 16 mg sample without deterioration of spectral resolution or evidence of sample degradation being observed (Figure 12).[11]
- A major advantage of Raman spectroscopy generally for sample investigation is the weak Raman scattering from water molecules; hence, for reactions in aqueous solution, biomolecules, and biopolymeric studies, little sample preparation in the form of desiccation or freeze-drying is necessary. This is particularly crucial for the monitoring of living biological systems by vibrational spectroscopy, where infrared-spectroscopic absorption due to water is strong.

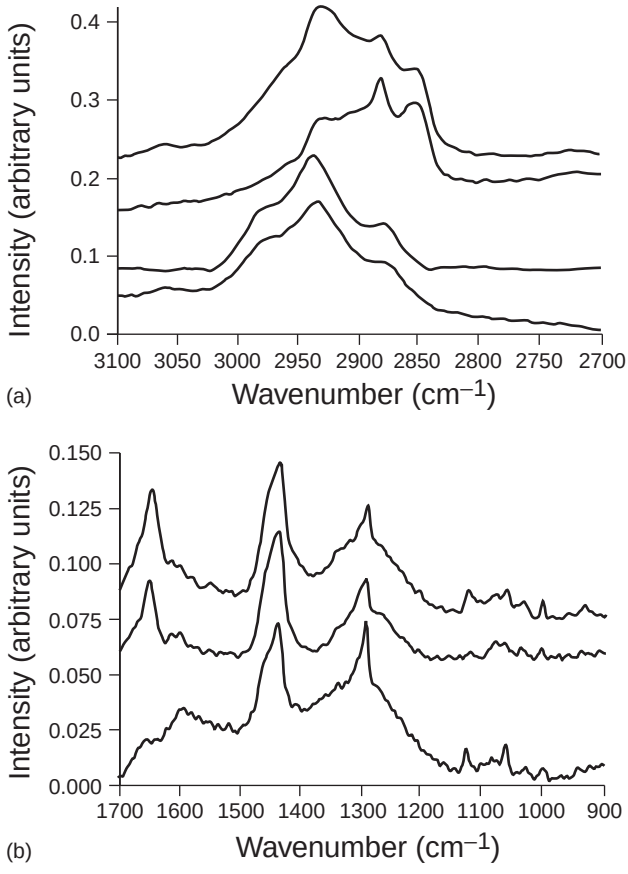

Figure 12 FT Raman spectrum of the skin of the 5200-year-old "Iceman"; 1064 nm excitation, 12 000 scans at 4 cm^{-1} spectral resolution. (a) 2700–3100 cm^{-1}; (b) 900–1700 cm^{-1}. (Reproduced from Williams et al., 'The Iceman: Molecular Structure of 5200 Year Old Skin Characterised by Raman Spectroscopy and Electron Microscopy', *Biochim. Biophys. Acta*, **1246**, 102. Copyright (1995) with permission from Elsevier Science.)

However, as shown in Figure 1, the move of excitation wavelength further into the infrared severely restricts the Raman data that can be acquired in the presence of water, because of increased absorption. This will present a severe limitation to longer-wavelength capability for FT Raman excitation. Even at 1064 nm, the effect of higher-overtone water-based absorption on the higher wavenumber shifts will be manifest. For this reason, the substitution of heavy water (D_2O) is often favored to alleviate partially the absorption problem with 1064-nm excitation in the near-infrared.

- The increased speed of data acquisition and ease of sampling have made the near-infrared excitation of FT Raman spectra attractive for industrial remote-sensing applications; however the extension of FTRS into the realm of fast, real-time data acquisition in nanoseconds is limited by the relatively slow frequency response of the detectors, but time

resolutions of 50 μs through step scanning have been shown to be possible. A critical parameter here, too, is the rather poor S/N generally experienced from single-scan or single-shot Raman spectra.[7] Generally, therefore, the faster detector responses of visible-CCD Raman systems are undoubtedly advantageous at this time for their adoption for industrial analytical measurements in real-time.

A major advantage of Raman spectroscopy, whether conventional or FT technique, is the linear dependence on concentration of molecular species and the observed band intensity.[6] Unlike absorption spectroscopy, where the Beer–Lambert law dictates a logarithmic dependency of concentration on absorption intensity, which demands knowledge of the molar absorptivity ε for the quantification of spectral-intensity data, the linearity of concentration/intensity dependence in the Raman spectrum provides an excellent means of following the growth or depletion of molecular species during a reaction.

In this way, Raman spectroscopy has been developed recently as a technique for monitoring reaction kinetics of polymerizations,[9] such as the living anionic polymerization of styrene and butadiene in hydrocarbon solvents by a butyl-lithium initiator. The information about kinetic and thermodynamic quantities from Raman-spectroscopic measurements in solutions over various temperatures has thus provided detailed information about molecular and species equilibria, the presence of dimers or oligomers in a reaction system, hydrogen bonding in alcohol and water systems, nitrating species in acidic media, and tautomeric stability. All of these applications are of importance in commercial systems and synthetic procedures. An excellent account of the many applications of FTRS to academic, commercial and interdisciplinary problems is provided in the classic work of Hendra et al.[9] Biologically, changes in distribution and composition of amino acid residues in large polypeptides and sulfur bridges in keratotic samples studied by Raman spectroscopy have provided an insight into drug-targeting systems and tissue degeneration, both of which are vital for biomedical diagnostic applications. Likewise, the absorption of drugs across the human skin barrier, transdermal drug delivery, of importance in pharmaceutical therapeutic formulations has been studied successfully by the FT Raman spectroscopic technique and the influence of chemical enhancers on drug-delivery systems quantified.

However, the linearity between the observed intensity and the concentration of scattering species in the Raman spectrum is not always applicable, especially when the wavelength of laser excitation approaches that of an electronic absorption band in the scattering species under investigation. This is the situation that occurs in resonance Raman spectroscopy; the signal strength of Raman scattering from a molecular species is increased by as much as 10^8, and the spectrum is simplified because only the vibrations associated with the chromophore under excitation are able to generate the signals in resonance. For example, in the case of haemoglobin, laser excitation within the haem absorption (400–600 nm) produces a resonance spectrum that contains only the enhanced signals of the haem ring vibrational modes, and no bands arising from the protein component are observed.

5 SAMPLE ACCESSORIES

5.1 Raman Microscopy

Raman microscopy, representing the marriage of one of the oldest pieces of scientific optical equipment (the microscope) and one of the newest (the laser), was born in the mid 1970s when Delhaye and Dhamelincourt[12] constructed the Raman MOLE (molecular optical laser examiner), which operated in the green region of the visible spectrum. The ability to characterize materials in a very small quantity, nondestructively, and typically in the nanogram to picogram range (10^{-9} to 10^{-12} g) and with a viewing ''footprint'' of only 1 or 2 μm, was appreciated instantly and gave rise to many and diverse applications.[6,7,9] With this small size of sample illumination, the potential of Raman microscopy in the area of biological materials was endless; the first Raman spectrum of living cellular material was recorded during this period and Raman imaging was born. However, despite the obvious advantage in being able to record the optical microscopic image and the Raman spectra from spatially resolved specimen areas of about 1–2 μm diameter to give a ''chemical signature'', some problems with the technique still surfaced. Many samples, especially those of biological interest, still fluoresced under blue or green laser excitation and the large laser power densities which could be achieved at samples using even moderately powerful microscope objective lenses resulted in sample degradation.

Messerschmidt and Chase[13] first attempted the recording of FT Raman spectra from microsamples using an infrared reflecting microscope and 1064 nm excitation. A typical commercial FT Raman microscope is shown in Figure 13; the principle is common to several instruments in that the laser radiation from the source is brought via an optical cable to the epi-illuminator of an infrared microscope, focused onto the specimen by the microscope objective; this also acts as the collecting lens for the scattered Raman radiation, which is imaged onto a beam splitter and transferred to the interferometer by another optical cable.

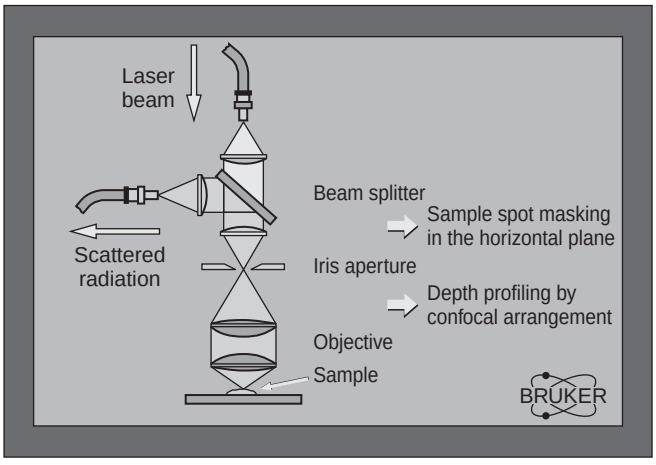

Figure 13 Diagramatic representation of FT Raman microscopic attachment for use with 1064 nm excitation. (Courtesy of Dr. P. Turner, Bruker Spectrospin, UK.)

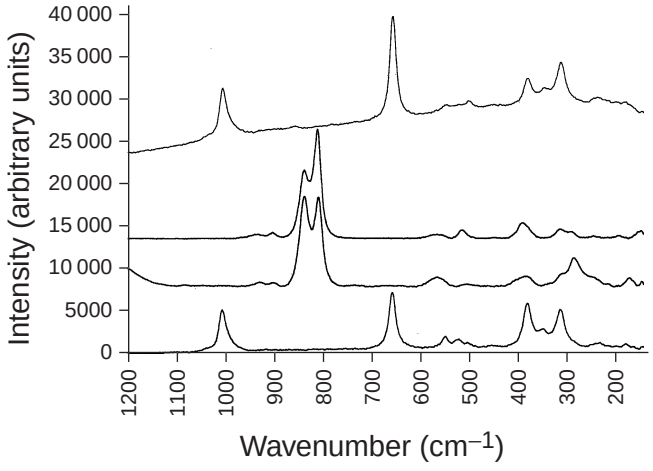

Figure 14 FT Raman microscopic spectra of regions of the Nakhla Martian meteorite, indicating the geological heterogeneity of the specimen. Major features are assignable to olivine, plagioclase and clinopyroxene. (Reprinted from Edwards et al., 'Comparative Raman Microscopy of a Martian Meteorite and Antarctic Lithic Analogs', *Planetary and Space Science*, **47**, 356.[14] Copyright (1999) with permission from Elsevier Science.)

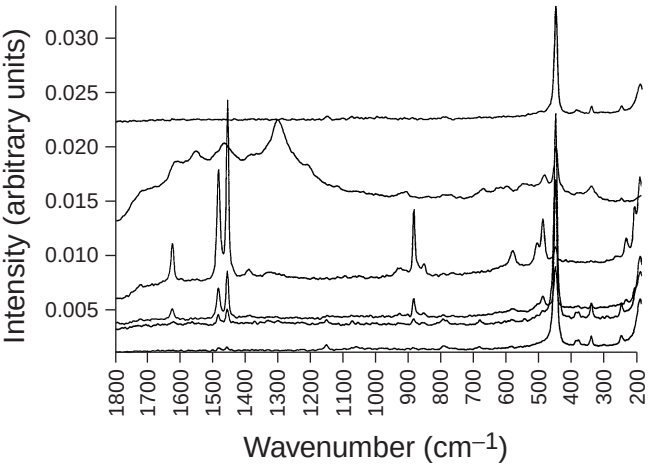

Figure 15 FT Raman microscope study of an Antarctic cryptoendolithic specimen, showing the distribution of geological and biological material in a vertical section. (Courtesy of Dr. D.D. Wynn-Williams, British Antarctic Survey, UK, *J. Raman Spectrosc.*, **28**, 688 (1997). Copyright John Wiley & Sons Limited. Reproduced with permission.)

The "footprint" for illumination and scattering in the microspecimen in the near-infrared is obviously dependent on the power of the objective lens used; typically, using the highest lens power of 100×, a sample diameter of about 8–10 μm is achievable. Although this is of an order of magnitude larger than its visible counterparts, the advantage of fluorescence rejection still pertains in near-infrared FTRS.

Examples of the use of FT Raman microspectroscopy from our own laboratories are provided by the first successful report of the Raman spectrum of ancient human skin tissue (Figure 12) from the Alpine Iceman (dating from 5200 years ago), the whole sample of which was only 16.2 mg. In Figure 12, characteristic bands due to the proteinaceous and lipoidal modes in the human skin are clearly seen and have enabled interpretations to be made of the survivability of organic materials in burial environments.

In Figure 14 is shown a stack plot of micro-Raman spectra from the Nakhla meteorite, one of only 13 meteorites of the Shergotty–Nakhla–Chassigny type identified as having originated in ejecta from asteroidal impact on Mars.[14] The geological heterogeneity of the specimen is clearly seen and major features can be assigned to clinopyroxenes, quartz and olivine. Incidentally, this meteorite is recognized as unique in that it is the only recorded extraterrestrial object which has caused a documented fatality on Earth – namely, a dog, when the meteorite impacted with Earth in Egypt in 1911!

Another very important system of environmental importance which has been studied by Raman microspectroscopy in the near-infrared is a cryptoendolith from the McMurdo Dry Valleys in Antarctica (Figure 15). Here, the living biological organism has colored the interior of Beacon sandstone and has adapted its survival strategy to the production of chemicals, including UV-absorbing pigments, water-storage materials, and cryogenic stabilizers, to combat some of the most extreme environmental conditions on this planet; temperatures as low as −50 °C, 0% humidity, and an enhanced UV-radiation flux reaching the terrestrial surface due to upper atmospheric ozone deflection. Raman spectroscopic studies of a vertical section of these bio-geological systems have revealed

new information about the colonization processes which have current relevance in the debate about the presence of extinct or extant life on Mars.[15]

The localized spatial information provided by FT Raman microspectroscopic studies of pigments used by our ancestors in their decoration of churches or caves has forged a new link between science and art/archaeology.[16] The analysis of the prehistoric cave paintings of the Pecos culture in the Rio Grande/Devils River confluence in North America (Figure 16) has provided novel information about not only the chemicals and minerals used by the ancient artists some 4000 years ago, but also to a hitherto unsuspected ritual in their cave-painting. Figure 17 shows the optical image of a section of a spallated pigmented fragment from Seminole Canyon in which small white particles can be seen embedded in a black pigmented area. The white particles are calcium oxalate monohydrate (whewellite) which is not native to the region's geology, and is attributed to the ancient artists' removal of lichenological deposits from rock surfaces prior to the application of the paint. The presence of organic matter inside the black pigmented layer, and not the red ocher found elsewhere in the caves, assisted in the identification of animal bone-marrow, probably bison or deer, which had been mixed with the paint presumably to give some ritualistic significance to the art work.

Figure 16 North American Indian pictograph from Seminole Canyon, Edwards Plateau, Pecos River, Rio Grande, USA. Age: ca. 4000 years BP. (Courtesy of Professor Jon Russ, University of Arkansas, USA.)

Figure 17 Spallated fragment of black pigment and substrate (41 VV 576) from Seminole Canyon pictograph; the black pigment was identified as carbon from FTRS, the white crystalline material as calcite and calcium oxalate monohydrate (whewellite). (Courtesy of Professor John Russ, University of Arkansas, USA.)

A recent advance in FT Raman microspectroscopy has been the advent of a confocal microscope sample illumination arrangement; the inclusion of optical apertures and iris diaphragms has resulted in a more precise depth-profiling measurement capability, often of the order of only several micrometers, which opens up novel industrial applications and biomedical diagnostics, particularly in the areas of polymer multicoatings, films and sections of prosthetic implants in healthy tissue (e.g. hip replacement joints with biocompatible coatings).

The coupling of an automated $x-y$ drive on the FT Raman microscope stage with computerized data storage acquisition has resulted in the provision of two-dimensional ''scans'' of a fairly smooth sample surface being obtained which reflect surface composition changes. Such an example has been put to good use in the pharmaceutical industry for examination of material heterogeneity; in our own laboratories an example of a crystallite ($\sim$250 μm) of β-oestradiol, which has separated

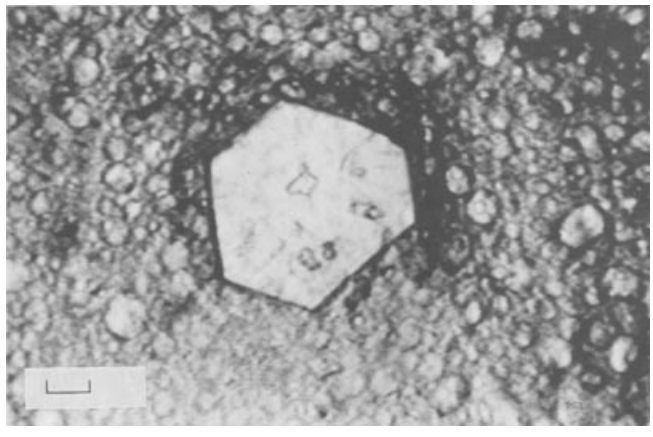

Figure 18 Optical image of oestradiol crystal embedded in gel matrix in an HRT patch. The size of the crystal is about 250 μm on its long axis. (Reprinted from Armstrong et al., *Vibrat. Spectrosc.*, **11**, 107. Copyright (1996) with permission from Elsevier Science.)

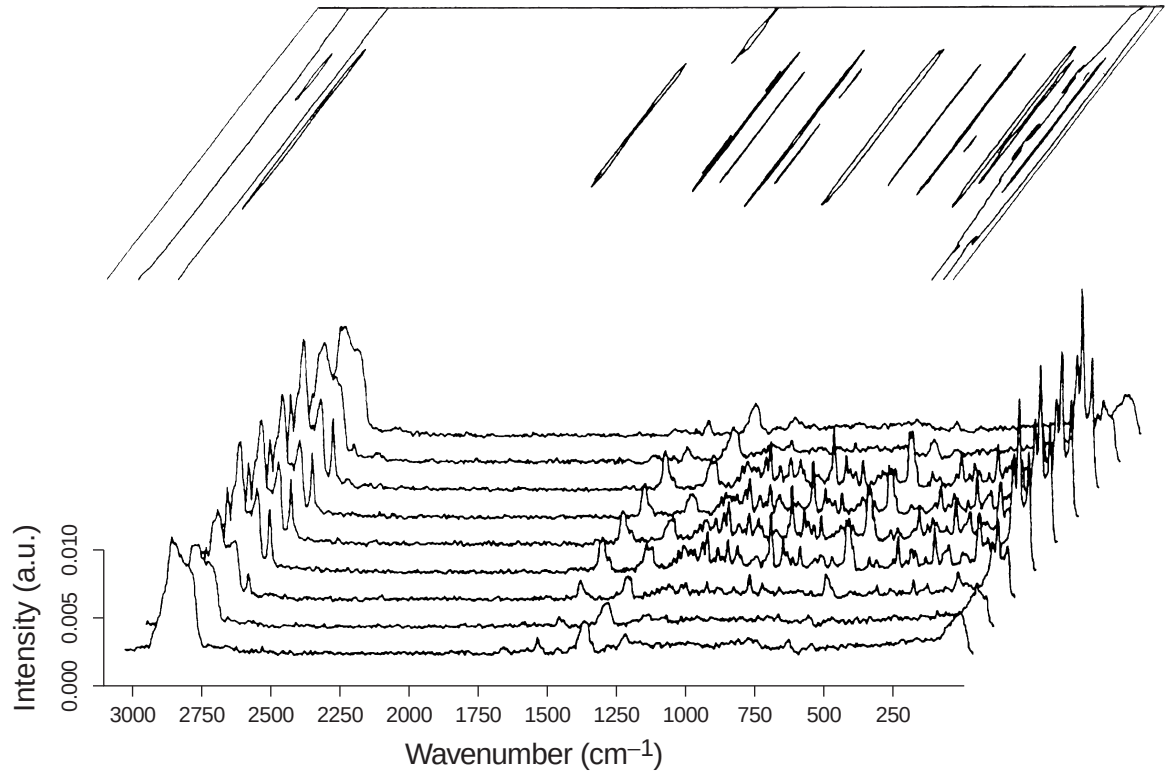

Figure 19 FT Raman microscopic spectra collected from a motorized *xy*-mapping experiment from scans across the crystal and its surroundings in Figure 18. The spectra of the surrounding gel and those of the crystal plus gel can clearly be differentiated. (Reprinted from Armstrong et al., *Vibrat. Spectrosc.*, **11**, 107. Copyright (1996) with permission from Elsevier Science.)

from its carrier gel in a hormone replacement therapy (HRT) patch is shown in Figure 18. The result of an $x-y$ FT Raman microscopic study of this region of the patch is shown in Figure 19; here, the presence of the crystal ''impurity'' separated from the surrounding gel matrix is clearly seen in the presence of oestradiol bands.

5.2 Remote Sensing Probes

Of equal-ranking importance with microscopes in FT Raman spectroscopic instrumental accessories is the remote-sensing probe; here, the existing laser radiation is brought out of the spectrometer sample illuminator housing via a flexible light-pipe, usually constructed of polymer or quartz fibers; and imaged onto the specimen using a focusing lens. These probes are of two types: contact and noncontact – referring to the specimen/lens distance. The construction of the probe geometry is very important and several possibilities are now in commercial use (Figure 20);[17] all have the common feature that the laser beam is brought in to the focusing lens by a central fiber, leaving the scattered radiation to be collected using a concentric series of fibers situated around the incoming laser beam.

The remote sensing probes currently in use with FT Raman interferometers demonstrate effectively the nondestructive sampling advantage of the technique for industrial or research environments. In addition to the advances which have been reported for industrial sampling of pharmaceutical tablets, gels, pharmaceutical preparations and polymer analysis (including determination of the reaction rates in situ in polymerization reactor systems from species concentration measurements) a most exciting development has arisen in the area of biomedical diagnostics.

Here, remote-sensing probes of up to 1-m long have been used to examine in vivo human tissue (Figure 21) for trans-dermal drug delivery studies and to evaluate the Raman spectroscopic technique as a nonsurgical device for the identification of skin disease and skin cancers (Figure 22). In Figures 23 and 24 novel applications to the study of in vitro samples of human tissue of archaeological significance are demonstrated by the analysis of a human skull from a Romano-British burial site (ca. 1600 years old) and the skin of a 30-year-old woman from the Qilakitsoq ice-mummy burials. Both have provided ''first'' examples of this type of analysis using a nondestructive spectroscopic technique.[18]

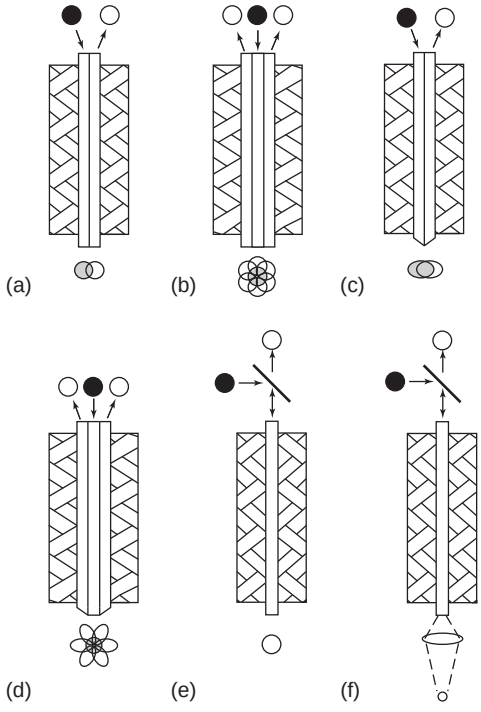

Figure 20 Schematic diagram of some typical fiber-optic Raman probes for remote Raman spectroscopy: (a) 1/1 flat-tipped; (b) 6/1 flat-tipped; (c) 1/1 bevelled; (d) 6/1 bevelled; (e) single fiber probe; (f) single-lens probe. *x* is the number of fibers used in the scattered light collection, and *y* is the number of fibers used in incident laser irradiation. (Reproduced from Angel et al., in Laserna,[17] 390. Copyright J. Wiley & Sons Limited. Reproduced with permission.)

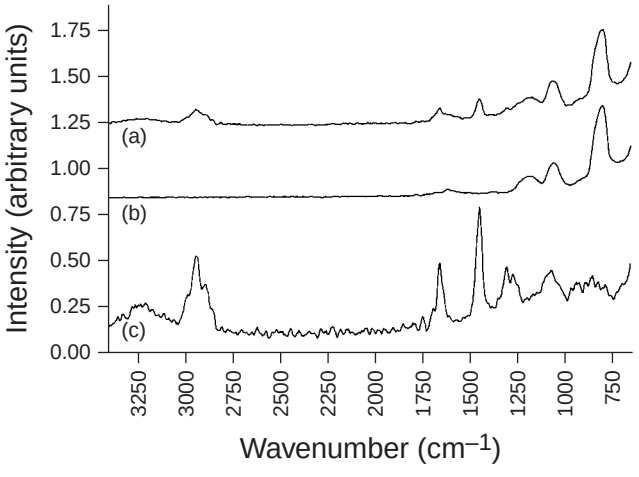

Figure 21 Raman spectrum of human stratum corneum in vivo; recorded using a 1 m silica probe (6/1 arrangement) and a Bruker IFS66/FRA106 FT Raman spectrometer: (a) probe and sample; (b) probe only; (c) subtraction of (a) and (b) spectra with enhancement.

Testament to the interest in remote-sensing accessories for both FT and conventional Raman spectroscopies in

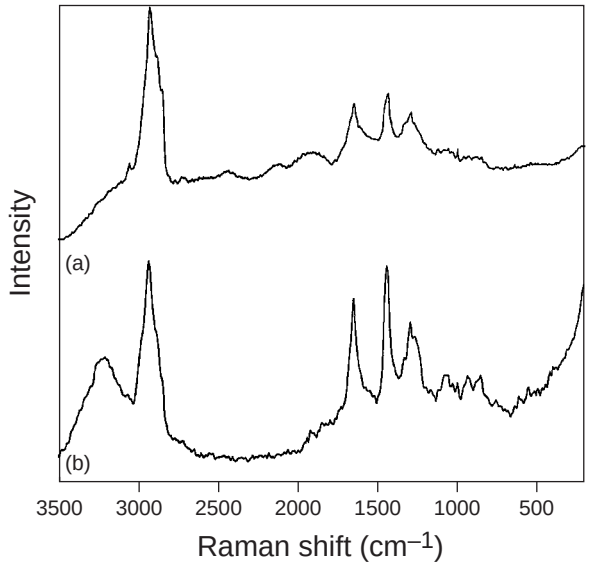

Figure 22 Raman spectra of (a) in vitro and (b) in vivo human stratum corneum – prototype for Raman biomedical diagnostics. The presence of the broad band centred at 3250 cm^{-1} and assigned to water should be noted; this band, which is absent from the in vitro sample, is ascribed to sweat collecting at the sample/probe junction in the in vivo experiment. (Reproduced by permission from *Pharm. Res.*, **10**, 1645 (1993). Copyright Kluwer Academic/Plenum Press.)

Figure 23 Arrangement of Bruker FT Raman spectrometer with remote-sensing probe attachment for the examination of the skull of a Romano–British skeleton (ca. 1600 years BP). The Raman microscopic accessory is shown in the background.

biomedical science is provided by several conferences which are devoted to this theme, including BiOS '99 (San Jose, USA) and Winnipeg 2000 (Canada).

5.3 Hot- and Cold-temperature Accessories

As the number and type of samples accessible to FT Raman spectroscopic study increased, so did the

Figure 24 Ice-mummy of a 30-year-old woman from the Qilakitsoq burial (ca. 500 years BP); note the extremely fine preservation of the body tissue and clothing. (Reproduced by permission from Gniadecka et al.[18] Copyright J. Wiley & Sons Limited. Reproduced with permission.)

requirement for temperature variability.[9] As with conventional Raman techniques, the ability to maintain samples at low temperature assisted in the recording of good quality spectra, especially in cases where specimen decompositions occurred in the laser beam, even at low nominal incident powers.

Chemical information available from quantitative and qualitative studies at variable temperatures was extremely useful for the characterization of the behavior of ions in solution, phase-transitions in solids and the investigation of crystal-growth phenomena. An early industrial use of this facility was provided by the characterization of the degradation and stability of pharmaceuticals with time, under shelf-storage or different environmental and climatic conditions.

In our laboratories, we have developed an infrared spectroscopic environmental chamber for FTRS which has resulted in several novel applications. In situ examination of chemical or physical changes in materials as a function of temperature in the range 20–220 °C (when onset of infrared broad band thermal emission occurs), humidity from 0–100% and with the possibility of utilizing different controlled synthetic atmospheric compositions is achievable.[19] The design of the environmental chamber accessory is shown in Figure 25; a sample of the FT Raman spectra of sulfur recorded in this environmental chamber at temperatures between 25 and 125 °C over the wavenumber range 50–600 cm^{-1} is shown in Figure 26. A detail of the ν(S–S) stretching region of sulfur S_8 during an experiment designed to monitor the rhombic–monoclinic phase change with time at 100 °C is shown in Figure 27, in which the quality of the FT Raman spectra obtained from this accessory is demonstrated.

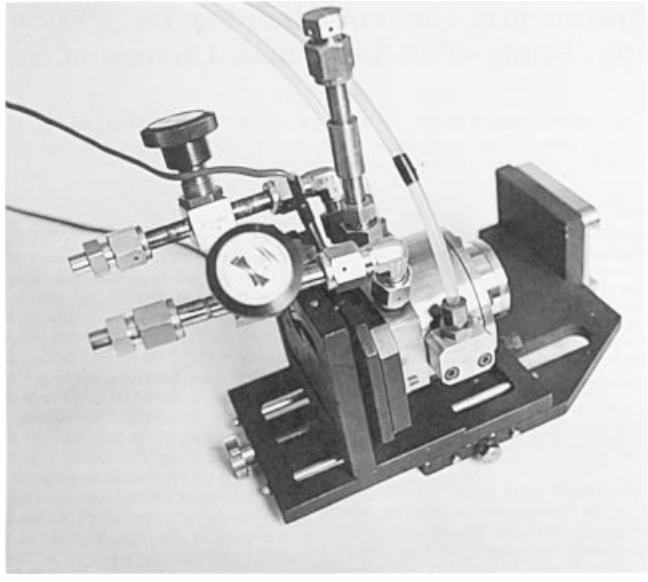

(a)

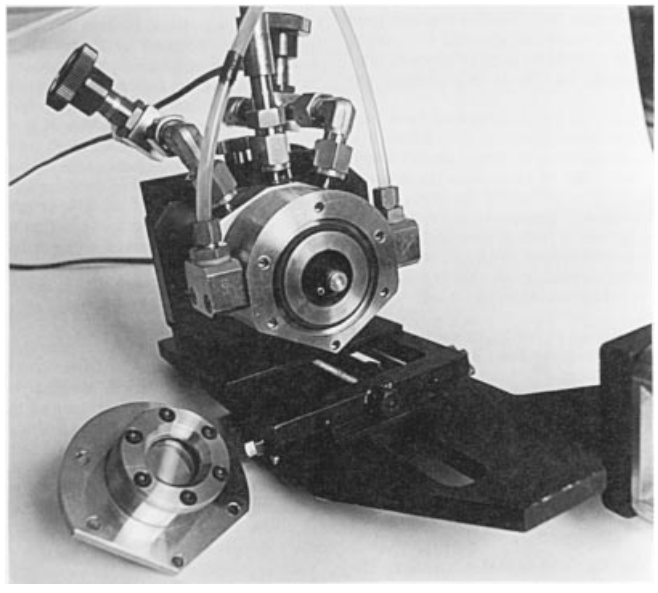

(b)

Figure 25 Environmental chamber for FTRS. (a) Chamber mounted on bracket for accurate location in spectrometer sample bearing of Bruker IFS66/FRA 106 instrument. (b) Front view of chamber showing quartz cover plate removed to expose the sample holder. (Reproduced by permission from Edwards et al.[19] Copyright (1997) Society of Applied Spectroscopy, USA.)

5.4 Other Accessories

There are special cells available for the recording of FT Raman spectra of materials under diverse conditions, including an electrochemical cell, a cryogenic unit and a multisample automatic analyser; the latter holds up to 16

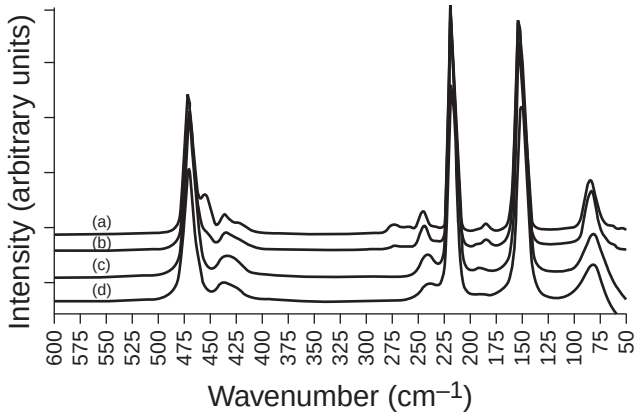

Figure 26 FT Raman spectrum of sulfur in the environmental chamber at different temperatures: (a) 25 °C; (b) 100 °C; (c) 115 °C; (d) 125 °C; 1064 nm excitation, 4 cm^{-1} spectral resolution, 200 scans, 50–600 cm^{-1} range. (Reproduced by permission from Edwards et al.[19] Copyright (1997) Society of Applied Spectroscopy, USA.)

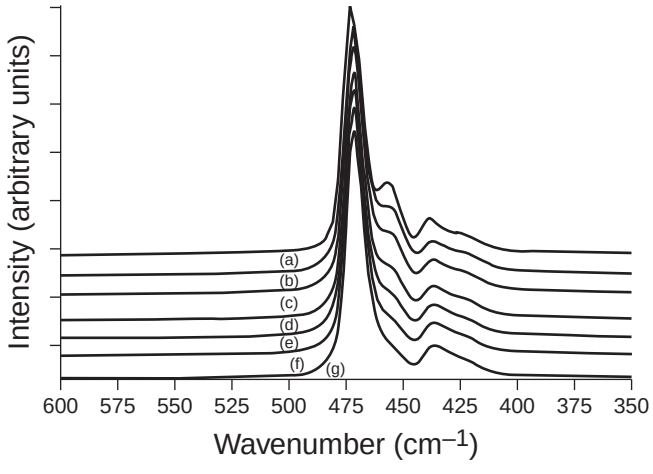

Figure 27 FT Raman spectra of sulfur in the environmental chamber at 100 °C for $t = 0, 30, 60, 120, 210$ and 270 min (spectra b–g respectively) over the 350–600 cm^{-1} range. Spectrum a is the room temperature sample (25 °C) for comparison purposes. Spectroscopic conditions as for Figure 26. (Reproduced by permission from Edwards et al.[19] Copyright (1997) Society of Applied Spectroscopy, USA.)

samples which are held in a remote-controlled carousel for the recording of spectra under sequentially identical conditions – particularly useful for pharmaceutical and polymer specimens. A particularly relevant description of ingeniously designed instrumental accessories for FTRS which take into account the very difficult sampling arrangements of commercially available spectrometers is provided by Hendra et al.[9]

The recording of Raman spectra from the gas phase has provided a challenge for FTRS because of the small sample volumes initiated, the intrinsically low concentration of scatterting species and the low intensity of Raman scattering in the near-infrared. As a result, vibration–rotation and rotational gas phase studies are still almost uniquely confined to visible excitation where high irradiances are achieved through complex multiple-pass cell geometry and the shrewd use of reflectors, often in an intracavity laser environment. A serious drawback to high-resolution gas-phase studies in FTRS is the laser linewidth at 1064 nm; in contrast, mode-selected argon-ion lasers operating at 488 or 515 nm have achievable spectral linewidths (full width at half-maximum) of 0.005 cm^{-1}. However, there is now a report of the FT Raman study of the NO_2/N_2O_4 gas phase equilibrium, which provides an as yet unique example of the application of FTRS to gases where information is not available from other Raman measurements.

Finally, there are also facilities for changing the geometry of sample illumination such that the laser beam is brought to the specimen in a vertical rather than on a horizontal laboratory axis; this, we have found invaluable for the study of extremely fragile, vulnerable samples which cannot be held in the normal configuration. The incident laser can also be brought in at a grazing incidence to the sample, which is essential for the study of thin films and coatings.

6 OTHER INSTRUMENTAL FACTORS

In concluding an account of FT Raman instrumentation, it is appropriate that the facilities provided by the data-acquisition services are also considered as, all too often, these are neglected. Hence, the facilities in a modern, multi-tasking, FT Raman spectrometer for spectral smoothing, baseline subtraction of one or more components from a mixture and comparison of a recorded spectrum with a database for chemometric studies are the most important.

To illustrate the observed effect of some of these operations on the observed Raman spectra, several examples have been chosen from our own laboratory studies.

Figure 28 shows the effect of spectral data accumulation under identical conditions and spectral averaging: the improvement of S/N, especially in the early stages of signal accumulation, is easily confirmed in a living biological system.[20] Incidentally, this stack-plot also demonstrates in a very graphical way that no degradation has occurred during sample irradiation at 1064 nm over a 12-h period. The instrumental internal wavenumber stability over the same period is also obvious from the appearance of the Raman bands, in which no broadening or loss of resolution can be observed.

Figures 29 and 30 show the FT Raman spectra from a study of medieval wall-painting fragments.[21] Medieval

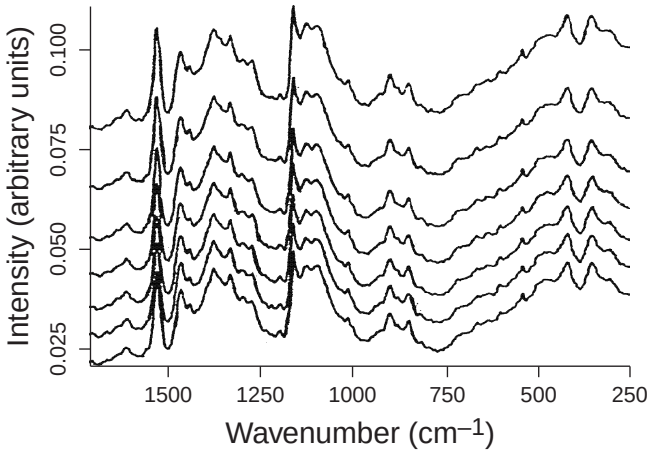

Figure 28 Demonstration of the effect of cumulative scanning and superposition of spectral data in FTRS on the S/N.

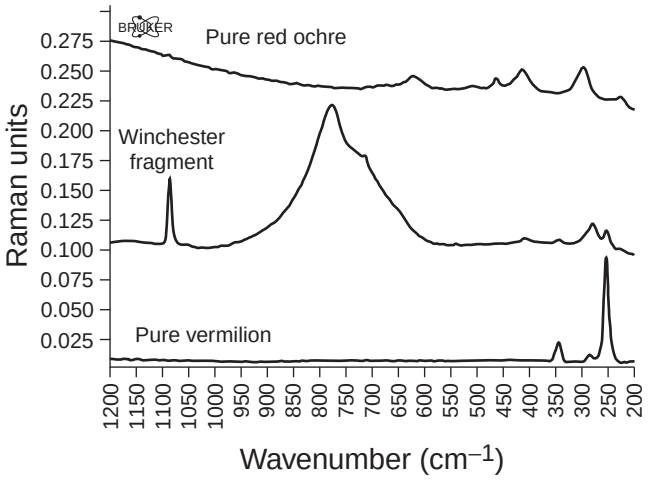

Figure 29 FT Raman stack-plotted spectra of a red-pigmented fragment from the famous ''Entombment of Christ'' wall painting, ca. 1175, in Winchester Cathedral, compared with red ocher and vermilion (cinnabar), respectively. The pigment from the medieval painting clearly is a mixture of cinnabar and red ocher although adulteration of expensive pigments, such as cinnabar and lapis lazuli, was common. The hierarchical significance of the Christ figure in the painting could also suggest that perhaps a later restoration of the damaged artwork might have been undertaken with adulterated materials in times of economic recession in the cathedral. (Reproduced from Edwards et al.[21] Copyright J. Wiley & Sons Limited. Reproduced with permission.)

wall paintings in English churches were practically universal between the twelth and the early sixteenth centuries. Few survive today in anything like a complete state, and most are in fragmentary form, often discovered during repair works when later coverings of limewash, and paint are removed. Scenes commonly depict biblical stories and representations of biblical characters. Certain

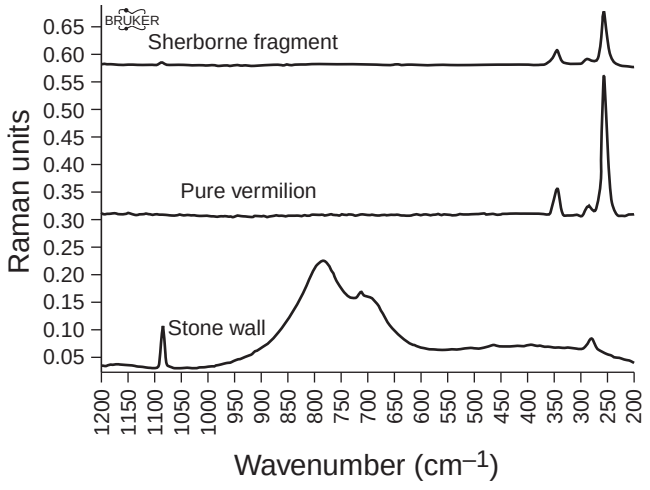

Figure 30 FT Raman stack-plotted spectra of a red-pigmented fragment from the much-deteriorated medieval (ca. 1180) wall paintings in Sherborne Abbey. Unlike the Winchester fragment (Figure 31), the Sherborne fragment consists of pure cinnabar (vermilion). (Reproduced from Edwards et al.[21] Copyright J. Wiley & Sons Limited. Reproduced with permission.)

saints such as the Virgin Mary, St Christopher and St Anne were popular, whilst the east wall of the nave was often reserved for a Last Judgment, or ''Doom'' painting, depicting the final day of reckoning with souls being judged and ushered either into Heaven or Hell; graphic depictions of Hell scenes are common.

Conservation of medieval wall paintings is subject to the same basic principles as other branches of conservation: preservation is the prime objective. However one of the main difficulties with mural paintings is that they have been executed as part of a standing building, and have, in effect, become part of the archaeology of the structure.[22] It is therefore of primary importance to determine both the chemical and physical nature of the materials employed, and that of the substrate, in the formulation of a satisfactory conservation strategy. The interaction of pigments with substratal materials is also important.[16]

From the medieval wall-paintings at Winchester Cathedral (Figure 31) and that at Sherborne Abbey, both dating from ca. 1180, our FT Raman studies have shown the different color palettes used; the red pigment in the Sherborne Abbey artwork consists of pure cinnabar, HgS, whereas that at Winchester Cathedral in contrast is a heavily adulterated mixture of red ocher, Fe_2O_3, and cinnabar. The Raman spectroscopic analysis of wall-paintings involved not only pigments, but the substrates and preparation materials as well. In Figure 32, the Raman spectra of a heterogeneous fragment of Renaissance fresco from the Palazzo Farnese in Caprarola, Italy, painted in 1560 by Zuccari is compared with that

Figure 31 Detail of the ''Entombment of Christ'' wall painting in Winchester Cathedral. (Reproduced from Edwards et al.[20] Copyright J. Wiley & Sons Limited. Reproduced with permission.)

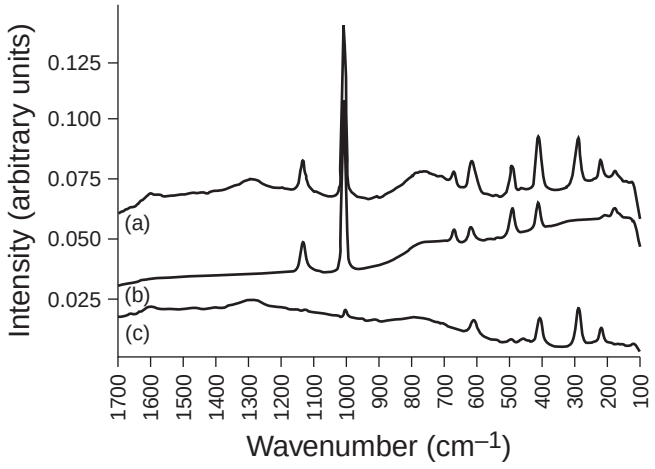

Figure 32 FT Raman spectra stack-plot of (a) pigment and substrate and (b) substrate only, from the 16th century Renaissance frescoes by Zuccari in the Palazzo Farnese, Italy.[23] The spectral subtraction (c) has achieved the purpose of isolating the pigment only, in this case a mixture of red ocher and carbon, with the haematite bands occurring in the low wavenumber region (100–600 cm^{-1}) and the characteristically broader bands due to amorphous carbon occurring at 1300 and 1600 cm^{-1}. There is some conjecture as to the origin of the carbon deposits in the pigment – carbon could have been deliberately added to give a darker red tone to the pigment but we have also noted the presence of carbon particles in pigments at other locations which have arisen from the soot of burning torches or oil lamps used by artists or visitors to the sites.[22,23]

of the substrate.[23] Subtraction of substrate from the ''pigment'' spectrum reveals the Raman spectrum of the true pigment, namely, red ocher. The Renaissance technique involved the intimate soaking of pigment into a wet plaster substrate – the FT Raman spectra of the ''pigment'' therefore contain bands due to substratal minerals as well, and the value of spectral subtraction for the characterization of the pigment is realized. A feature of interest in the subtracted spectra of the pigment is the doublet characteristic of amorphous carbon with components near 1300 and 1600 cm^{-1}; this correlates well with black particles observed visually in specimens under high magnification and may be ascribed to lamp-black or soot, which has been deposited from candles or torches either at the time of application or by later visitors to the site.[22] Features other than pigment and carbon may occasionally be seen in such spectra; an appropriate case is provided by work on the medieval wall-paintings from the Convento de la Peregrina in Sahagun, Spain, where evidence for lichen colonization and biodeterioration has been provided for the first time through FTRS.[22] Bands due to the ν(CO) stretching mode of calcium oxalate monohydrate (whewellite) have been identified in certain areas of the wall-paintings and related to lichen invasion through unglazed apertures in the chapel. Another feature of lichen invasion, the ability to mobilize iron in sandstone substrata and hence the production of areas of differential ''staining'' of substrata, has been identified in the same church through Raman microscopy.

In conclusion, the development of FTRS and microscopy with near-infrared excitation has opened up novel applications in areas of biomedical diagnostics, archaeology, art science, forensic science, environmental science and the interaction of biological materials with geological formations. One of the latest applications which is now being considered is the reduction in size and weight of the Raman spectrometer, excitation laser and detector for an on-board instrument which will accompany small robotic vehicles for the exploration of planetary surfaces. The series of Mars missions being promulagated by NASA and ESA from 2001, which start with small planetary rovers and culminate with a

proposed sample-return mission in 2007/8, and with the Mars Express Beagle II Lander mission in 2003 involve Raman spectroscopists who will have the opportunity to participate in perhaps the most-remote sensing Raman experiments yet undertaken, at a minimum distance of about 30 M miles! Along with experiments[24] planned for the International Space Station in 2001/2 involving FTRS and the evaluation of the survivability of living materials to an extremely harsh space environment, it may be stated that Raman spectroscopy is no longer just applicable within the terrestrial realm and will join infrared spectroscopy as a weapon in the armoury of astronomers, exobiologists, and space scientists.

''*Per Spectroscopica ad Astra*''

ACKNOWLEDGMENTS

The author wishes to record his appreciation to his colleagues in collaboration in some of the international projects cited in this article. In particular, special thanks are due to Professors Fernando Rull, Ole Nielsen, Brian Barry, Jon Russ, Drs Adrian Williams and David Wynn-Williams for their continued support. Also to the funding authorities, ESPRC, NERC, British Antarctic Survey, University of Bradford and the British Council, who made these studies possible.

ABBREVIATIONS AND ACRONYMS

CCD	Charge-coupled Device
CW	Continuous Wave
FT	Fourier Transform
FTRS	Fourier Transform Raman Spectroscopy
HRT	Hormone Replacement Therapy
MOLE	Molecular Optical Laser Examiner
Nd^{3+} : YAG	Neodymium : Yttrium Aluminum Garnet
S/N	Signal-to-noise Ratio
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy, Ex Vivo Tissue Analysis by • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications • Raman Spectroscopy in Analysis of Biomolecules

Process Instrumental Methods **(Volume 9)**
Raman Spectroscopy in Process Analysis

Infrared Spectroscopy **(Volume 12)**
Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Raman Microscopy and Imaging • Raman Scattering, Fundamentals

REFERENCES

1. C.V. Raman, K.S. Krishnan, 'A New Type of Secondary Radiation', *Nature*, **121**, 501–502 (1928).
2. P.R. Griffiths, *Chemical Infrared Fourier Transform Spectroscopy*, Wiley, New York, 1974.
3. G.W. Chantry, H.A. Gebbie, C. Hilsum, 'Interferometric Raman Spectroscopy using Infrared Excitation', *Nature*, **203**, 1052–1053 (1964).
4. D.B. Chase, T. Hirschfeld, 'FT-Raman Spectroscopy: Development and Justification', *Appl. Spectrosc.*, **40**, 133–137 (1986).
5. A. Weber, S.P.S. Porto, 'Rotational Raman Spectroscopy using Laser Excitation', *J. Opt. Soc. Am.*, **55**, 1033–1048 (1965).
6. H.G.M. Edwards, 'Raman Spectroscopy Instrumentation', in *Encyclopedia of Applied Physics*, American Institute of Physics/VCH Publishers, Weinheim and New York, 1, Vol. 16, 1996.
7. D.B. Chase, J.F. Rabolt, *Fourier-transform Raman Spectroscopy: From Concept to Experiment*, Academic Press, San Diego, 1994.
8. K. Asselin, D.B. Chase, 'FT-Raman Spectroscopy at 1.339 Micrometers', *Appl. Spectrosc.*, **48**, 669–701 (1994).
9. P.J. Hendra, C. Jones, G. Warnes, *Fourier Transform Raman Spectroscopy: Instrumentation and Chemical Applications*, Ellis Horwood, Chichester, 1991.
10. J.G. Grasselli, B.J. Bulkin, *Analytical Raman Spectroscopy*, Wiley Interscience, New York, 1991.
11. A.C. Williams, H.G.M. Edwards, B.W. Barry, 'The Iceman: Molecular Structure of 5200 Year Old Skin Characterised by Raman Spectroscopy and Electron Microscopy', *Biochim. Biophys. Acta*, **1246**, 98–105 (1995).
12. M. Delhaye, P. Dhamelincourt, 'Raman Microprobe and Microscope with Laser Excitation', *J. Raman Spectrosc.*, **3**, 33–43 (1975).
13. R. Messerschmidt, D.B. Chase, 'FT-Raman Microscopy at 1064 nm Excitation', *Appl. Spectrosc.*, **43**, 11–14 (1989).
14. H.G.M. Edwards, D.W. Farwell, M.M. Grady, D.D. Wyn-Williams, I.P. Wright, 'Comparative Raman Microscopy of a Martian Meteorite and Antartic Lithic Analogs', *Planet. Space Sci.*, **47**, 353–362 (1999).

15. H.G.M. Edwards, E.M. Newton, 'Application of Raman Spectroscopy to Exobiological Prospecting', in *The Search for Life on Mars*, ed. J.A. Hiscox, British Interplanetary Society, London, 83–88, 1999.
16. H.G.M. Edwards, 'Applications of Raman Spectroscopy to Art and Archaeology', Plenary Lecture 19, in *Proceedings of the XVIth International Conference on Raman Spectroscopy (ICORSVXI)*, Cape Town, South Africa, ed. A.M. Heyns, John Wiley & Sons, Chichester, 81–84, 1998.
17. J.J. Laserna (ed.), *Modern Techniques in Raman Spectroscopy*, John Wiley and Sons, Chichester, 1996.
18. M. Gniadecka, H.C. Wulf, O.F. Nielsen, D.H. Christensen, J.P. Hart Hansen, 'Fourier Transform Raman Spectroscopy of 15th Century Mummies from Qilaqitsok, Greenland', *J. Raman Spectrosc.*, **28**, 179–184 (1997).
19. H.G.M. Edwards, D.W. Farwell, J.M.C. Turner, A.C. Williams, 'Novel Environmental Control Chamber for FT-Raman Spectroscopy: Study of the In Situ Phase Change of Sulfur', *Appl. Spectrosc.*, **51**, 101–107 (1997).
20. F.J. Chu, M.R.D. Seaward, H.G.M. Edwards, 'Novel Applications of FT-Raman Spectroscopy to an Ecological Study of Hong Kong Supralittoral Lichens', *Spectrochim. Acta, Part A*, **54**, 967–982 (1998).
21. H.G.M. Edwards, C. Brooke, J.K.F. Tait, 'An FT-Raman Spectroscopic Study of Pigments of Mediaeval English Wall Paintings', *J. Raman Spectrosc.*, **28**, 95–98 (1997).
22. H.G.M. Edwards, F. Rull Perez, 'Lichen Biodeterioration of the Convento de la Peregrina, Sahagun, Spain', *Biospectrosc.*, **5**, 47–52 (1999).
23. H.G.M. Edwards, E.R. Gwyer, J.K.F. Tait, 'FT-Raman Analysis of Paint Fragments from Biodeteriorated Renaissance Frescoes', *J. Raman Spectrosc.*, **28**, 677 (1997).
24. G. Horneck, D.D. Wynn-Williams, R.L. Mancinelli, J. Cadet, N. Munakata, G. Ronto, H.G.M. Edwards, B. Hock, H. Wanke, G. Reitz, T. Dacher, D.P. Hader, C. Brioullet, 'Biological Experiments on the EXPOSE Facility of the International Space Station', in *Proceedings of the 2nd European Symposium on the Utilisation of the International Space Station, ESTEC*, Noordwijk, The Netherlands, ESA SP-433, 459–468, 1999.

Raman Microscopy and Imaging

Jeremy J. Andrew
Unilever Research US, Edgewater, USA

1 Introduction **2**
2 The Theory of Spectral Imaging **3**
2.1 Generic Spectral Image Data 3
2.2 Raman Microspectroscopy and Imaging 3
2.3 Methodologies within Raman Microscopy and Imaging 5
3 Instrumentation **6**
3.1 Raman Instrument Components 6
3.2 Other Instrument Considerations 8
3.3 Instrumentation for Raman Microspectroscopy 9
3.4 Instrumentation for Raman Imaging 10
3.5 Three-dimensional Imaging Using Raman Spectroscopy 13
4 Experimental Considerations **13**
4.1 Detection Limits 13
4.2 Sampling Conditions 13
4.3 Quantitative Applications 13
4.4 Instrument Response Correction 13
5 Application Examples **14**
5.1 Polystyrene Spheres – a Raman Image Test Sample 14
5.2 Emulsion Systems 14
5.3 Other Systems 15
6 Data Handling and Data Analysis **16**
6.1 Multidimensional Spectral Data Handling and Univariate Analysis 16
6.2 Multivariate Analysis 16
7 Comparison with Other Microscopy and Spectral Imaging Techniques **18**
7.1 Other Alternative Approaches to Spectral Microscopy 19
7.2 Comparison of Raman and Infrared Microscopy 19
7.3 Fourier Transform Raman Microprobes 20
8 Perspective and Future Developments **21**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **22**
References **23**

Raman microscopy and imaging couples the spatial resolution of optical microscopy with the spectral resolution and information content of Raman spectroscopy. Together, a powerful tool for the investigation of chemical, or molecular, heterogeneity is created. Raman microspectroscopy enables the nondestructive analysis of small quantities of sample and the identification of contaminants and inclusions in situ within a matrix – both of great importance and utility in forensic investigations. Perhaps more importantly, it has been recognized that the microheterogeneity of products has an influence on both product performance and appearance. In the analysis of complex, multiphase systems, optical microscopy contrast methods (the majority of which depend upon refractive index changes) are commonly employed to provide information on the structural organization of materials. However, information on the chemical organization of materials is then often based on conjecture. Raman microscopy is arguably the most easily applied of a family of techniques that use the innate spectral signatures of the components as a contrast element and hence provides a method for analyzing chemical heterogeneity in ''real'' systems.

Typically, the information that would be appropriate from an investigation of chemical heterogeneity would be:

- *the number of chemical components present;*
- *the pure spectrum of each component;*
- *spatially resolved concentration information for each individual component.*

Raman microscopy and imaging provides a tool to determine these important factors in complex, multiphase systems. In its simplest form, the Raman microspectroscopy experiment requires a laser to be focused down to a diffraction-limited spot on a sample and the inelastically scattered light is collected from that point and analyzed. This generates a Raman spectrum, which is indicative of the identity and quantity of the molecular species present in that sample volume. The limiting spatial resolution that can be obtained is of the order of 1 μm in the lateral plane and 2 μm in the axial plane, the microscopes used lending themselves to a confocal optical arrangement that provides this level of depth discrimination.

The information content is constrained to that available in the Raman vibrational spectra and so commonly molecules with distinct functional groups are clearly distinguished, whilst it is more difficult to distinguish specifically between molecules with very similar molecular structure and chemical functionality. It is also a commonly held belief that Raman investigations are hampered by

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

fluorescence interference, which obscures the Raman spectrum. However, with modern Raman microscopes there are few samples where this limitation becomes completely defeating.

Several approaches to the generation of Raman images (images where the contrast is based on the chemical or molecular heterogeneity) have been established and are discussed in detail. A comparison is made between these different Raman imaging approaches and also with different spectral imaging methods.

1 INTRODUCTION

Why is Raman microscopy such an important development in analytical science?

There are several reasons. In an industrial context, it is because it has been recognized that the microheterogeneity of products has an influence on both product performance and appearance.[1,2] Hence methods which enable the interrogation of the chemical microheterogeneity in complex, multiphase systems are of great value in understanding how to monitor and influence product properties. Microspectroscopy also enables nondestructive analysis of small quantities of sample and the identification of contaminants and inclusions in situ within a matrix – both of great importance and utility in forensic investigations.

In the analysis of complex, multiphase systems, optical microscopy contrast methods (the majority of which depend upon refractive index changes) are commonly employed to provide information on the structural organization of materials. However, information on the chemical organization of materials is then often based on conjecture.

Whilst there are several approaches to analyzing the chemical heterogeneity of samples, many require the labeling of specific target molecules or the use of phase-specific stains or dyes. However, there is a family of techniques that use the innate spectral signatures of the components and hence provide a method for analyzing ''real'' systems. Members of this family include vibrational (Raman and infrared (IR)) microspectroscopy and X-ray microscopy as well as optical and fluorescence approaches using innate absorption or fluorescence signals. These approaches preclude the argument about labeling, tagging or staining influencing the system under study. Another crucial argument in their favor over labeling methods is the ability to investigate product microstructure in factory-scale processing. Clearly labeling molecules on such large-scale production is not feasible.

Microspectroscopy can therefore provide chemical information on a point basis and be compared to an optical microscopy image. Typically, the information that would be appropriate from such investigation would be:

- the number of chemical components present;
- the pure spectrum of each component;
- spatially resolved concentration information for each individual component.

The information contained within point microspectroscopy is in principle very powerful and provides not just this desired information but the spectroscopy also contains further molecular information such as alkyl chain conformation. In fact, all of the spectral properties associated with bulk spectroscopic investigations are accessible but now with spatial resolution at a level familiar to optical microscopy, that is on the micrometer scale. Perhaps the most widely applicable and useful member of this family is Raman microscopy and imaging.

It is helpful to define more specifically the methodologies possible within the technique of Raman microscopy and imaging, as there has been some confusion with regard to the terminology. In the USA, and in most of the literature, Raman microspectroscopy has been used to describe the concept of obtaining a Raman spectrum using a microscope as the sampling device. In contrast, European workers have tended to use the term Raman microscopy to describe this experiment. The same confusion is observed in the IR literature. The use of Raman microscopy in this context has created much confusion in neighboring disciplines as the term ''microscopy'' is generally expected to be applied to the capture and presentation of an image. If the term Raman microspectroscopy is used to describe the simple collection of a spatially resolved point spectrum, then the term Raman microscopy is reserved for the presentation of a true spectral image. Even with this distinction, there is confusion with regard to the way such spectral images are generated, either by mapping or by ''wide-field'' approaches.

In order to bring greater clarity to the nomenclature, the use of three distinct terms is suggested to describe the various experimental approaches:

1. Microspectroscopy: the use of a microscope as a sampling device to allow spectral information to be obtained from a discrete volume within a sample, preferably registered to some form of conventional optical image. This first approach should strictly not be considered microscopy or imaging as it does not produce an image based upon the spectral contrast but rather uses the microscope as a sampling device for simple point spectroscopy.
2. Mapping and image reconstruction: the use of the microspectroscopy approach coupled to some form

of sample mapping to collect spectral data at many pixels within a sample and subsequently build images based upon the spectral (chemical) heterogeneity.

3. Spectral microscopy: the use of the microscope more conventionally. A wide field of view is imaged using spectral contrast by employing a method of filtering and imaging the entire illuminated area at a particular wavelength or set of wavelengths.

The relative merits of the three approaches are discussed in more detail in section 2.3.

A fourth term, spectral imaging, can then be used to include both experimental approaches of mapping and image reconstruction and spectral microscopy, and this term describes the ability to generate images of the sample where the contrast in the image is derived from the chemical (molecular) heterogeneity. Both experimental approaches result in a spectral image data set that describes the spatial variation of the spectral characteristics of the sample. This provides an excellent representation of the chemical components in a sample and their distribution in a complex, multicomponent, multiphase system. From this data set, an understanding of the molecular interactions or the effects of the molecular distribution on product characteristics can be drawn.

2 THE THEORY OF SPECTRAL IMAGING

It is the aim of spectral imaging to produce images which describe the spatial distribution (and sometimes properties) of molecular species in a sample. The information content and data structure within a spectral image data set are complex and the richness of information even in a single spectral image is quite remarkable. This section describes the generic nature of spectral image data and then turns to specific considerations with respect to Raman microscopy.

2.1 Generic Spectral Image Data

The nature of spectral imaging is complex and the richness of the data obtained is quite staggering.

In an *xy* image plane of a sample, spectral imaging data is actually four-dimensional, each data point having (x,y) coordinates describing its spatial position and wavenumber (or wavelength) and intensity values describing its spectral nature. It is, however, more common to describe spectral imaging only in terms of the spatial dimensionality of the data and consider the spectral dimensions in a different way, so an *xy* spectral image is commonly referred to as a two-dimensional data set. Clearly visualizing, analyzing and interpreting a data set which is four-dimensional in nature requires the use of sophisticated software tools and these are referred to in more detail in section 6.

Spectral images that include full volumetric *xyz* images have been obtained and in a similar way these are conventionally referred to as three-dimensional spectral images. These data sets, which are five-dimensional in nature, create even more challenges for the analysis and visualization of the information contained within them.

Figure 1 shows schematically the nature of spectral image data, as they are collected by either mapping and image reconstruction or by spectral microscopy. It should be stressed that communication of spectral image information is much easier if a more familiar optical microscopy image is obtained to which the spectral information is registered, as shown in Figure 1. This also aids interpretation of the spectral image data.

If we consider a two-dimensional (*xy*) spectral image then the spectral dimensions of the data can be considered in two ways:

1. At each pixel within the *xy* plane a complete spectrum (intensity vs wavenumber or wavelength) is available – spatially resolved spectral information. This representation is more familiar to spectroscopists and also represents the way the data are collected in a point mapping and image reconstruction approach to the generation of a spectral image data set.
2. Alternatively, a spectral image data set can be thought of as a stack of cards, where each card represents an image based on the contrast (intensity) of the spectral signal at each pixel at a particular wavenumber or wavelength – spectrally resolved spatial information. This representation is more familiar to microscopists and also represents the way the data are collected in a spectral microscopy approach to the generation of a spectral image data set.

With either method of data collection or either method of conceiving the spectral dimensionality of the data, the information is equivalent if the field of view, spectral resolution, spectral range, spatial resolution and image fidelity (number of pixels or pixel density) are identical.

However, it is rare that you would choose to analyze a particular sample by both approaches. The relative merits of the two approaches for collecting spectral image data are discussed in section 3.4.

2.2 Raman Microspectroscopy and Imaging

From the first developments of Raman microprobes in the 1970s it has taken more than 20 years for Raman microspectroscopy and Raman imaging to emerge as a

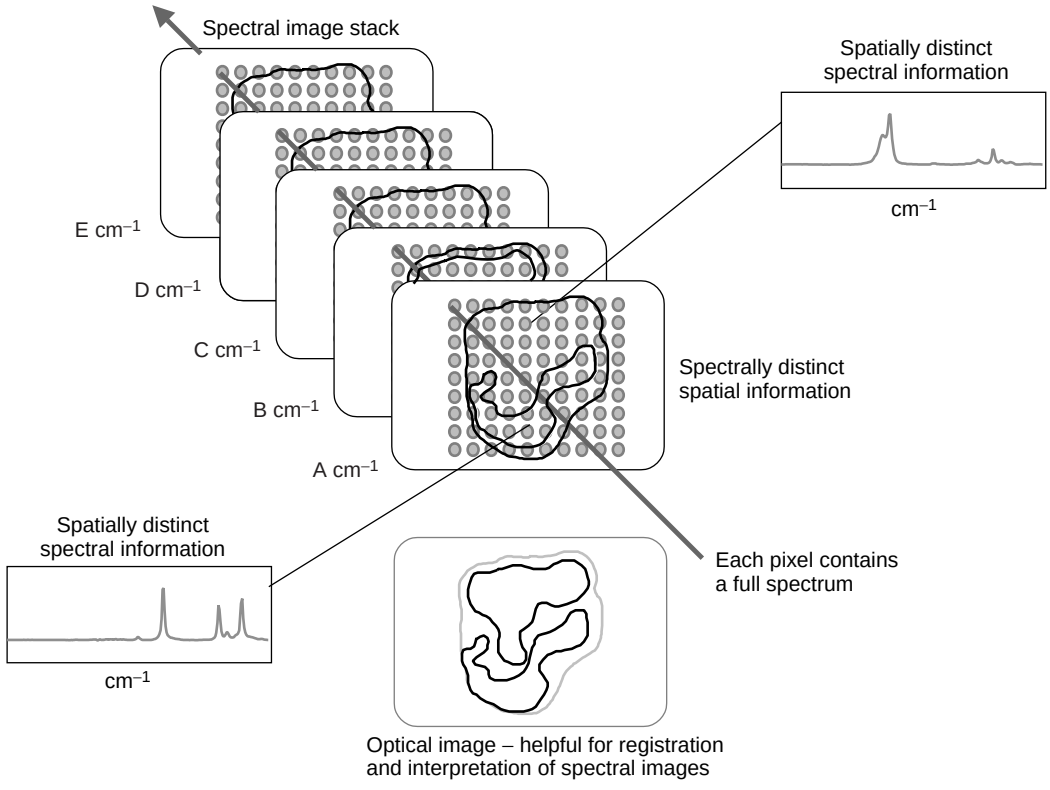

Figure 1 Spectral image data sets.

key analytical technique. The pace has not been set by the theoretical limitations of the technique but rather by the pace of technology development.

The theoretical support for the development of Raman microspectroscopy and Raman imaging has always been very strong. The technique benefits from several key advantages when it comes to understanding heterogeneity in complex, multicomponent samples. These include:

- Innate signals – the Raman effect is inherent to the molecules and provides molecular contrast.
- Ease of sample preparation – sampling is simple, requiring little or no sample preparation, when the conventional microscope backscattering geometry is used. Sampling is further enhanced by the fact that water is a weak Raman scatterer and does not interfere appreciably with the spectra of solutes. Hence biological samples and water-rich product systems lend themselves readily to analysis by this method.
- Spatial resolution – Raman microscopy when operating at the common wavelengths of 780 nm or below, and in a confocal configuration, provides limiting practical spatial resolution of the order of 1 μm in the *xy*-plane and 2 μm in the *z*-direction. This is comparable with the optical resolution familiar to optical microscopists and so the spectral information is readily referenced to an optical image.
- Spectroscopy on a micrometer scale – all of the theory and literature in support of conventional spectroscopy can now be performed at a spatial resolution of 1 μm. The information content in the spectroscopy is not solely limited to intensity but other spectral signatures can be determined with spatial resolution. Examples include the nature of hydrocarbon chain packing and polymer crystallinity and the identification of different polymorphs of the same compound.
- Ability to deal with complex mixtures – the Raman spectrum is a high contrast signal. Therefore, complex mixtures with multiple chemical components can be successfully interrogated with a single experiment. The ability to both discriminate between small differences in molecules (and hence the spectroscopy) and to pick up the presence of molecules at low bulk concentrations is enhanced in a spectral image data set. This is due to the likely spatial variation in the spectral signatures, which provides a second contrast element in the data set. This enables subtle variations in spectroscopy

to be determined. Similarly the likelihood is that materials at low bulk concentrations may be locally at high concentration and hence observable. Both of these advantages are enhanced by the use of sophisticated data analysis tools, which essentially co-add and average the entire data set rather than rely upon the information from a single pixel spectrum.

- Quantitation of the signal – the intensity of Raman scattering is directly proportional to the concentration of scattering species; this is clearly of benefit to quantitative studies. Relative quantitation between species is certainly possible but absolute quantitation requires calibration. The commonly presented distribution maps or images of chemical species based upon intensity should be considered to be equivalent to an arbitrary concentration scale unless much effort is made to quantify the response.

2.3 Methodologies within Raman Microscopy and Imaging

The three distinct approaches within the field of Raman microscopy and imaging given in the Introduction are now discussed in greater detail.

2.3.1 Point Microspectroscopy

In this case, the microscope is used as an optical device to enable microprobe sampling. The spectral information is obtained from a microenvironment, without contribution from the surrounding media (Figure 2).

This is the simplest application of the coupling of Raman with microscope optics. The output is a spectrum, preferably with an optical image which shows the point from where the spectral information was obtained.

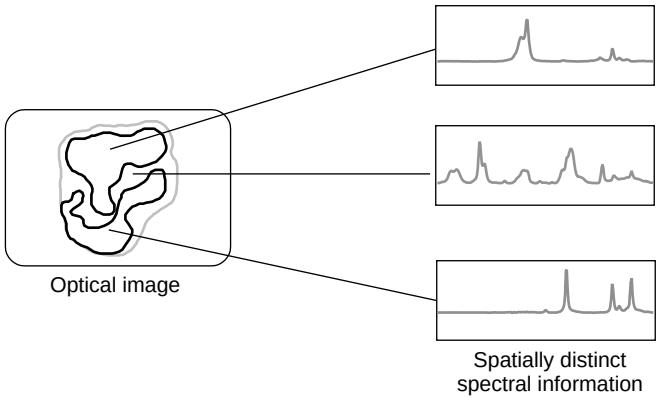

Figure 2 Point microspectroscopy is used to obtain spatially resolved spectral information.

This first approach is often referred to as Raman microscopy. However, the author prefers the term Raman microspectroscopy as strictly this methodology does not produce an image based upon the spectral contrast but rather uses the microscope as a sampling device for the point spectroscopy.

2.3.2 Point Mapping

By scanning an area of the sample under the microscope using a motorized stage, spatially resolved point microspectroscopic information is collected and organized into a spectral image data set (Figure 3). This information can be used to reconstruct chemical images based, for example, on Raman intensity at a particular wavenumber (and hence the distribution of molecular species). This approach is currently preferred for samples containing weak Raman scatterers with inherently low contrast in the spectral image data.

2.3.3 Line Mapping

In the line scanning experiment, the laser is arranged to illuminate a line on the sample and this line is then imaged to the entrance slit of the monochromator and the spectra are obtained spatially resolved along the line at different heights on an imaging detector (Figure 4). There are several instrumental approaches to achieving this and these are described in section 3.4.2.

2.3.4 Point and Line Mapping and Image Reconstruction

The point and line mapping approaches to Raman image reconstruction benefit from the maximum flexibility, that is, the spectral data are complete for each point and chemical images can be constructed from these data in any way appropriate to the spectroscopy.

For example, intensity at chosen wavenumber, intensity ratios, band shifts, band half-widths, etc. can be used to produce chemically specific images. The analysis can then follow the same rigorous spectroscopic principles

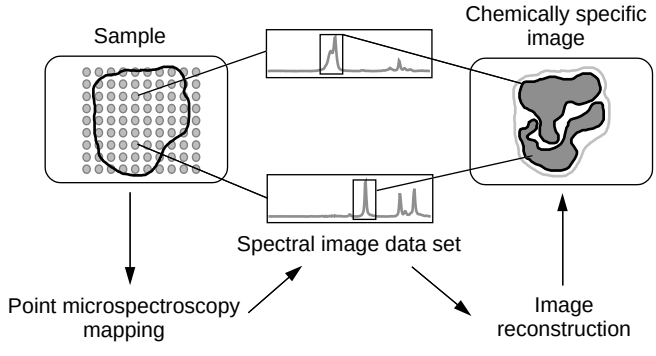

Figure 3 Point mapping of a sample area.

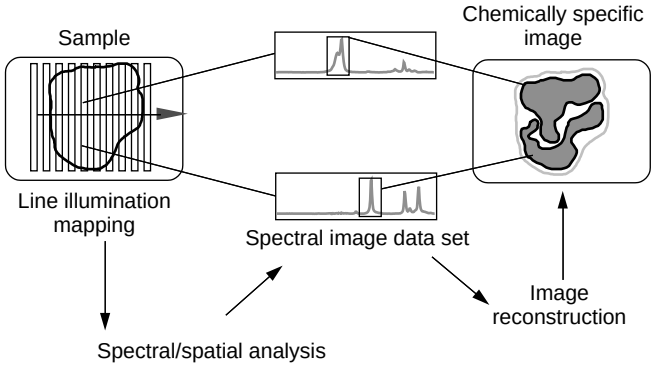

Figure 4 Line mapping of a sample area.

one would apply to single spectra, such as background correction for fluorescence and scattering from the sample. Individual point spectra or the complete spectral image data set can be analyzed to produce semiquantitative information and the spectral resolution is limited only by the monochromator.

The disadvantage is the slow rate of data acquisition – for example, using the simple point scanning/two-dimensional mapping approach, 1 min acquisition per point leads to a total of over 40 h being required to produce a $50 \times 50\,\mu m$ spectral image. However the most recent generation of Raman microscopy instruments with higher light efficiency coupled with more sophisticated data analysis has reduced acquisition times for many practical samples down to 0.5 s per point. An image of $50 \times 50\,\mu m$ at 1-μm spatial resolution can therefore now be collected in approximately 2 h (allowing for additional dead time for stage movement and detector read-out which is not yet optimized). An ultimate goal of 30 min or less for such an image is certainly not unreasonable. The spectral image collection time is not then really prohibitive for the analysis of static systems but even at these data collection rates there is clearly limited application for imaging kinetic phenomena using the point mapping approach.

2.3.5 Spectral Microscopy

By defocusing the laser, or some other means of wide-field or ''global'' illumination, a field of view is exposed to laser irradiation; the image is then captured in the focal plane of the microscope on an imaging detector after some form of spectral discrimination (Figure 5).

The spectral microscopy approach is sometimes referred to as ''global illumination'' or ''true spectral imaging'', in contrast to the mapping and image reconstruction techniques presented above. In principle, either of the approaches – spectral microscopy or spectral mapping and image reconstruction – produce the same spectral image data set if all of the experimental parameters are arranged to be identical. The relative

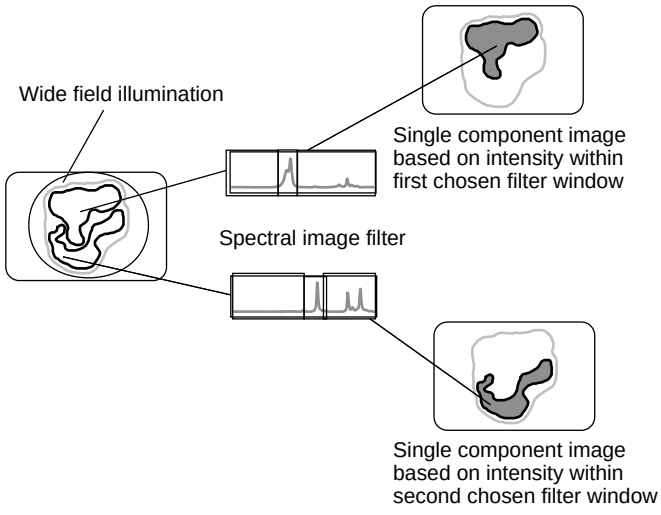

Figure 5 Spectral microscopy.

merits of the different approaches are discussed further in section 3.4.

In practice, if the data set is presented as a series of images based upon the spectral contrast, then a classical microscopist would not distinguish between the mapping or ''true microscopy'' approaches to data collection. (Scanning electron microscopy, for instance, uses a point mapping approach but is conventionally referred to as microscopy and not as a mapping technique.)

3 INSTRUMENTATION

The early 1990s saw significant improvements in the efficiency of every component of Raman instrumentation, taking Raman microscopy out of the academic laboratory and into routine use within mainstream analytical laboratories. It was not long before the potential in Raman imaging was realized.

The hardware for a generic Raman instrument can be classified into five main areas: excitation laser, sampling device, Rayleigh filter, spectral selection and detector. The Raman experiment is then completed by an appropriate analysis method. These components are represented in Figure 6 and the instrumentation described in more detail below.

3.1 Raman Instrument Components

The individual components of a Raman instrument are now described.

3.1.1 Lasers

Owing to the higher efficiency of modern Raman spectrometers the most common excitation sources are

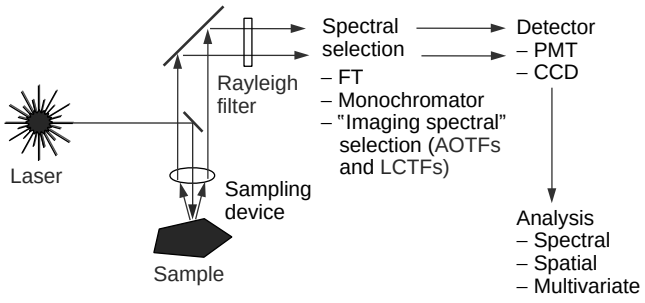

Figure 6 A schematic of the Raman instrument.

now low power, air-cooled helium–neon (HeNe 633 nm) lasers or wavelength-stabilized (external cavity) diode lasers, typically operating in the 780 nm region. Argon ion or krypton ion lasers are still in use and now tend to be low power and air-cooled in contrast to the situation in the 1970s and 1980s where high power, water-cooled lasers were more typical.

In Raman spectroscopy a common problem is sample fluorescence caused by the excitation of electronic transitions in the sample at the same spectral wavelengths which have been used to obtain Raman spectral information. Sample fluorescence is most noticeable when using visible excitation laser sources (such as the argon ion or HeNe lasers). At these wavelengths, the probability of electronic excitation and subsequent fluorescence is high, particularly in anything but pure samples and, as a consequence, their industrial applications have been limited.

One solution to the fluorescence problem is to move the excitation wavelength to 1064 nm, in the near-infrared (NIR), away from the majority of electronic transitions and hence obtain virtually fluorescence-free Raman spectra.[3] The use of NIR excitation would appear to be the ideal source until the wavelength dependence of the Raman scattering efficiency is considered. The intensity of the scattering is inversely proportional to the fourth power of the exciting wavelength.[4,5] Therefore any increase in wavelength dramatically decreases the scattering intensity. The effect that this has on the sensitivity of the Raman experiment has been clearly demonstrated.[6] The 1064-nm excitation has been implemented in the context of Fourier transform (FT) Raman and, although ideal for bulk spectroscopy, it does not prove to be a very practical wavelength for Raman microscopy; this is discussed in section 7.3.

The consensus trade-off between fluorescence suppression and Raman efficiency is now the use of diode laser excitation in the 780-nm region. This is due to the availability of convenient, stable and adequate power diode laser sources and the fact that the Raman signal then remains within the detection window of silicon-based charge-coupled device (CCD) detectors.

3.1.2 Sampling Devices

Sampling devices used in a Raman instrument can be classified into two types. The first category is termed macrosampling devices. The types of vessels used for this purpose include liquid/gas cells (high or low temperature), electrochemical vessels, spinning solid sample holders and more recently fiber optic probes.[7–9] The second category is termed microsampling devices. These include devices where the sampling area is a spot around 1 μm in diameter. The most common microsampling device is a conventional optical microscope and this is described in detail in section 3.3.2.

3.1.3 Rayleigh Filters and Spectral Selection Devices

In traditional instruments in common use up until the late 1980s, double- or triple-pass monochromators were required in order to achieve a reasonable suppression of the Rayleigh scatter at the excitation frequency. The first monochromator was used to filter out the Rayleigh scattered light at the laser wavelength, whilst the second (and third) monochromator was used to provide the spectral dispersion of the scattered light.

Multiple pass monochromators were designed to be used in spectral scanning mode, or single wavelength mode, with point detectors, typically cooled photomultiplier tubes (PMTs). With the advent of multichannel detectors such as intensified diode arrays and subsequently CCD cameras, many multipass monochromators were re-engineered for use with multichannel detectors. However, the inherent optical design for point detection does not match efficiently to a multichannel detection scheme and so the promise of higher efficiency was never fully realized.

A multipass monochromator configuration can still be advantageous for flexibility in multilaser, high resolution laboratories and for investigations very close to the Rayleigh line. However, the consequence of such an arrangement is that the light throughput is extremely low, of the order of 1–2%.

The availability of high efficiency holographic notch and super-notch filters for Rayleigh scattering rejection has now made CCD–single monochromator systems the norm for Raman spectroscopy. Using these filters is simple and highly efficient.[10] The front monochromator of a double monochromator system can be replaced by a holographic filter and a single-pass imaging monochromator optimized for CCD detection is then used to provide the spectral separation. This dramatically improves the light throughput up to a typical 20–25% collection efficiency in single-pass monochromator–microscope sampling systems, whilst allowing a frequency cut-off near the laser line (50–100 cm^{-1}),[11–14] without excessive attenuation at other wavelengths.

3.1.4 Monochromator Design

As discussed above, the initial use of CCD detectors in Raman was hindered by the fact that monochromator design was not aimed towards two-dimensional detectors but at point detectors. As the light gathering efficiency was not matched to a flat field at the detector the true advantage of collecting a spectral range on the CCD, rather than collecting spectra point by point, was not demonstrated.

The new generation of monochromators have been designed so that delivery is optimized to the image plane of the CCD camera and true image quality of the slit is maintained on to the detector. These monochromators are short focal length and can be Czerny–Turner with stigmatism correction, or on-axis stigmatic,[15] both using conventional diffraction gratings. Even more efficient are the new generation of monochromators based on transmission hologram gratings.[16,17] These provide throughput efficiencies of more than 75% and dual hologram technology can provide high spectral resolution and full spectral coverage by projecting two tracks of different regions of the spectrum on to a CCD detector. These represent the best light throughput efficiency available but are less flexible, requiring a new grating for any change in excitation wavelength.

Imaging monochromators also enable multiple source experiments, such as spectral detection of the light from several fiber optics lined up on the slit or the line scanning imaging facility possible on microprobes[18] which is described in detail in section 3.4. The spatially resolved spectra are obtained using the CCD read-out characteristics to independently read each spectrum at different heights on the CCD chip.

3.1.5 Spectral Microscopy–Imaging Filters

There are several approaches to image filter technologies which have been used in spectral microscopy. As this hardware is specific to the microscopy approach it is discussed in section 3.4.4.

3.1.6 Detectors

For many years PMTs were the only detectors used for measurements with scanning Raman spectrometers. The development of CCD cameras provided an ideal detector for Raman dispersive instrumentation because of their much higher quantum efficiency and the advantage that can be obtained from the two-dimensional character of the detector area. The dispersed spectrum can be projected on to the CCD detection chip and spectral information at a wide range of wavelengths can be obtained simultaneously. The resolution is related to the pixel size on the detector and the degree of dispersion caused by the grating. CCD design formats have been optimized for spectral use (such as 1024×256 pixel chips) with the long CCD axis oriented in the spectral direction. More flexible binning and software drivers have been developed specifically for spectroscopic rather than imaging use.

Early CCDs suffered from poor response in the NIR, limiting their use with 780-nm excitation, however back-thinned, deep depletion CCDs with higher quantum efficiencies in the red region of the spectrum are now available.

3.1.7 Advantages of Charge-coupled Device Detectors

As well as the obvious advantage of higher quantum efficiency, there are other advantages of the use of CCDs.

The CCD camera is intrinsically a multichannel device and hence provides a multiplex advantage over using a point detector and spectral scanning. The multiplex advantage is the ability of the system to detect multiple wavelengths instantaneously. A crucial issue is whether there is any multiplex advantage to be gained in Raman spectroscopy.[19] The commonly accepted multiplex advantages are only observed when the experiment is detector noise limited. Raman experiments using CCD detectors often fall into the shot noise limited category, simply due to the low noise characteristics of cooled CCDs, where statistically there is no apparent advantage from using a multiplex technique.

However, there are other factors to consider. The CCD detector removes the requirement for scanning monochromators, naturally increasing the speed of data acquisition through decreased dead time. CCDs have far higher quantum efficiencies than PMT detectors and the noise levels of PMT detectors are higher, which can make equivalent experiments, which are shot noise limited when using a CCD, detector noise limited with PMT detection.

Other advantages are that source power fluctuation (leading to problems with quantitative analysis) are simultaneously observed across the whole spectral range. In CCD systems the fluctuations can average out over long exposure times, which is not possible with single point detectors. By removing the scanning monochromator system, multichannel detection removes drive position uncertainties, which can degrade signal averaging performance and resolves the usual errors of ordinate axis uncertainty in a set of experiments.

3.2 Other Instrument Considerations

Some other considerations should be mentioned with regard to the practical use of Raman instrumentation.

3.2.1 Wavenumber Accuracy

In a dispersive system, the limits on wavenumber accuracy are again related to the optical design, reproducibility of sampling and grating position. Sampling is more easily controlled if a microscope is used as the collection/delivery optics. The grating position reproducibility with CCD–multiplex systems where the grating is held stationary, rather than scanning systems where the grating is moved, is also improved. With appropriate choice of conditions it may not be necessary for the optics to move during a full series of experiments.

Newer monochromator designs such as those based on dual holographic transmission gratings allow for wide spectral coverage at high resolution and negate the need for any moving parts in the monochromator and detection system. Such systems are inherently more stable. Nevertheless it is sensible to calibrate the ordinate axis at appropriate intervals and correct if necessary.

3.2.2 Instrument Response

The development of calibrated white-light sources in Fourier transform Raman spectroscopy (FTRS) to correct for the single beam nature of Raman spectroscopy translates equally importantly to all Raman experiments.[20] Indeed it is perhaps even more critical in dispersive systems if one is changing, for example, gratings and wishes to make comparisons between spectra obtained on the same instrument but with markedly different throughput characteristics. Such calibration systems for microscope devices using National Institute of Standards and Technology traceable standards are now available commercially and it is strongly recommended that these procedures are followed.

3.2.3 Instrument Software

In the past software to manipulate Raman spectra has been basic. The increase in use of off-line analysis packages, such as GRAMS™, and the improving opportunities to convert between file formats make this much less of a consideration. The ease of data acquisition/instrument control is now the major consideration in this area rather than post-collection analysis. However, the analysis of spectral image data sets poses unique problems and is returned to in section 6.

3.3 Instrumentation for Raman Microspectroscopy

In order to obtain an acceptable signal-to-noise ratio (S/N) in a Raman microscope experiment with scanning multipass spectrometers and PMTs, the Raman signal must be integrated for a long time. Therefore, the sample has to withstand a high level of laser irradiation and during these long exposure times the material can suffer from thermal damage or photodegradation. As a consequence, Raman microscopy was of limited applicability until the overall efficiency of Raman spectrometers improved.

The first improvements began to occur in the late 1980s and continued through the early 1990s. Almost every aspect of the Raman instrument's efficiency improved to make not just Raman microspectroscopy but also the Raman mapping and Raman microscopy experiments a practical reality.

There is now a general consensus on Raman microscopy instrumentation for the microspectroscopy and mapping and image reconstruction approaches. The elements of the instrumentation are now described.

3.3.1 Lasers

The major advance has been the development of small, wavelength stabilized, diode lasers in the region of 780 nm. This excitation wavelength proves to be a good compromise between fluorescence suppression and Raman scattering efficiency and the full spectral shift range remains within the CCD detection window. Generally, the combined improvements in the rest of the Raman instrument imply that small, air-cooled lasers such as HeNe lasers at 633 nm can also be used if visible excitation is desired.

3.3.2 Sampling Device – the Microscope

One of the advantages of Raman microscopy is that a conventional optical microscope can be used as the sampling device; an attachment that allows for entry of the laser beam along the optical axis of the microscope and collection of the backscattered light to be returned to the spectrometer has to be added but, to all intents and purposes, the microscope can be arranged to perform as if it were a conventional optical microscope. This allows for images of the sample to be obtained using the array of contrast mechanisms familiar to optical microscopists and the Raman image information is then registered to this. This can be a great advantage when attempting to fully characterize a heterogeneous sample.

In all cases, the laser excitation and Raman scatter is collected through a confocal arrangement to reject out-of-focus light. This has the advantages of minimizing stray light or fluorescence from other parts of the sample not in focus and providing the optimum spatial resolution in both *xy*- and *z*-directions. Most systems use confocal pinholes although one commercial system achieves confocality by having the entrance slit of the spectrometer at the confocal image plane (which provides discrimination in one spatial direction) and then narrowing the binning/read-out region on the

CCD detector to achieve discrimination in the other spatial direction. In certain circumstances it may be advantageous to be able to adjust the pinhole size to achieve higher S/N at the expense of confocal spatial resolution and this is only possible in some of the commercial systems.

Coupling the light between the laser and the microscope for excitation and the microscope and the spectrometer for scattered light can be achieved either by direct optical coupling or by the use of fiber optics. The former is clearly more efficient, whilst the latter provides greater flexibility, in being able to remove the Raman head from the microscope to make adjustments without requiring the system to be realigned.

As with the conventional research optical microscopes, there has been a shift towards infinity corrected optics. There is little advantage in the Raman experiment to using anything but the simplest objectives, as essentially monochromatic light is being brought to a focus on-axis and there is no need to be concerned with wide-field or chromatic image aberrations. In fact, some of the higher quality objectives are less efficient because their light transmission characteristics are lower. With NIR excitation at 780 nm it is advantageous to use NIR objectives optimized for high light throughput in the NIR. However, such objectives may be prohibitively expensive and hence only recommended for extremely low S/N applications.

3.3.3 Rayleigh Filters

The most influential component in the recent revolution in Raman instrumentation has to be the holographic notch filter for Rayleigh rejection.[10–12] It is this small but extremely efficient narrow band rejection filter that is now at the heart of almost all commercial instruments. In the context of Raman microscopy, the use of this filter as a beamsplitter has led to very efficient designs for laser injection and scattered light collection from microscopes.

3.3.4 Monochromators

All commercial systems designed specifically for Raman microscopy are configured around single-pass imaging monochromator systems. There are options based around astigmatism-corrected classical Czerny–Turner designs, stigmatic on-axis spectrographs and the more recent transmission hologram grating designs. All designs are optimized for CCD detection and maintenance of the slit image on to the detector.

3.3.5 Detectors

Again, with no real exceptions, microscope systems are equipped with CCD detectors. There are several options. Back-thinned CCD chips are now manufactured reliably enough to not suffer from gross variations in pixel–pixel quantum efficiency and small variations are easily handled by correct instrument calibration (see section 3.2.2). These chips represent the most effective detectors, particularly the deep-depletion chips which have an enhanced IR region response, giving better efficiency with 780-nm excitation. Conventional front illuminated CCDs are cheaper and work perfectly well with visible excitation, most effectively at 633 nm (HeNe).

CCD camera cooling options are between multistage Peltier (both air and water) and liquid nitrogen. For the vast majority of practical applications, air-cooled CCDs are adequate (most Raman experiments now being shot noise limited). For the most demanding of applications with long integration times of an hour or more, liquid nitrogen cooling may be appropriate.

3.4 Instrumentation for Raman Imaging

The previous section described the instrumentation appropriate for point microspectroscopic investigations. This section describes the additional complexity required to produce Raman images.

3.4.1 Point Mapping

The adaptation of a point microspectroscopic system to point mapping is made simply by the addition of a software-controlled motorized stage.[1,21–25] The software clearly must be interfaced with the spectrometer acquisition software but in most cases this has been dealt with in the current commercial designs.

3.4.2 Line Mapping Instrumentation

The line mapping approach requires two criteria to be met. First, the laser illumination has to be projected across a line on the sample and, second, the scattered light from that line has to be imaged on to the entrance slit of the spectrometer, preserving the spatial resolution along the line.

In the line scanning experiment it is clear that it is best to arrange for the illumination intensity across the line to be uniform in order to confidently interpret the spectral intensity variations.

There are three main methods of illuminating the line on the sample. These are the use of galvanometer-driven spinning mirrors,[26,27] cylindrical lenses[28,29] or the more recent Powell lens optic.[30,31] Even if illumination on to the sample is arranged to be uniform, there will be a transfer function for the line through the microscope and spectrometer optic train that may result in a nonuniform instrument response function. With care this could be calibrated out of the system but is an added

complexity. The direct coupling of the line illumination to the slit of the monochromator implies that confocality (particularly the depth resolution in the z-direction) is lost in the line scanning experiment in the cases using lens optics.

Obtaining a line scanning profile using the spinning mirrors method has the advantage of preserving the confocality in the z-direction[32] but still produces a variation in the illumination intensity along the line. The main advantage of using a cylindrical lens is that the system is easier to align but, along with the loss of confocality, there remain intensity variations in the line at the extremities.[29] However, the use of a Powell lens for line imaging has reduced these illumination intensity variations along the line to less than 10%.[30,31]

3.4.3 Comparison of Line Mapping with Point Mapping

The claim for line mapping is that the method can lead to decreased acquisition times for the same S/N. In a qualitative imaging sense, this is certainly correct. However, point mapping benefits from a simpler on-axis optical design and as a consequence the instrument response at each pixel in the spectral image data set is reliably quantitative. There is no guarantee that this is the case along the line in a line scanning approach. For a quantitative analysis, instrument response corrections have to be made for uneven illumination and uneven light throughput efficiency along the line. The second clear disadvantage with lens-based systems is the loss of confocality, which decreases the depth discrimination in the image. Confocality often proves to be important in real samples.

The major advantage will be achieved when there is excess laser power available that cannot be focused down on to a single point on the sample without power damage. Then the ability to put an appropriate power density simultaneously along a line and collect the spectral information in parallel is clearly desirable.

This method also allows simultaneous collection of data along a one-dimensional spatially resolved axis and brings timescales down to the second/minute range for collection of quality spectral information, acquired across a boundary for example. This may provide interesting opportunities for kinetic studies of interfacial transport phenomena.

The line focusing method can be used to smear out sample heterogeneity and provide a simple bulk spectrum from a region of heterogeneous material. The spectral information collected across the line can be averaged by simply binning the information on the complete vertical (spatial) axis of the CCD chip. This approach results in an improvement in the S/N over point microscopy along the line and subsequent averaging. In all cases there is a decrease in the motorized stage movement (dead) time during image as the stage only has one direction in which to move.

The disadvantages of the line mapping approaches have been indicated above. Essentially, the loss of confocality in the lens-based approaches can be a serious limitation in complex systems. The spinning mirror approach can conserve the confocality but results in added complexity in both the instrumentation and the line-up process. All of the approaches suffer from some degree of line response function (from both uneven illumination and photometric variation in the line transfer function through the optics to the spectrometer).

In summary, there have been only limited attempts to commercialize a line imaging system, based on either spinning mirrors or cylindrical optics.[26,28,29] These instruments have met with very limited success. No commercial devices have yet appeared using Powell lenses, which provide the best option for uniform line illumination.[30,31] It appears that the advantages for line mapping over the simpler point mapping approach have not been valuable enough to make such attempts commercially successful.

3.4.4 Raman Microscopy Instrumentation

In the case of Raman microscopy, there are again two critical criteria. First, a wide field of view has to be illuminated by the laser and, second, the scattered light from that field of view has to be imaged through a spectral section filter on to a camera at the image plane.

In most cases, the illumination of a wide field of view has been achieved by defocusing the laser. This will inevitably result in some uneven illumination. An alternative approach to illumination methods is the use of two nonresonant spinning mirrors[33] and this is claimed to achieve far more even illumination (but at the cost of added instrument complexity).

There are several approaches to the spectral image filtering. The simplest approach is to use an imaging monochromator, with the slit opened to allow the image to pass through the system on to the CCD.[34,35] The images obtained are at very low spectral resolution and so the practical application of this approach is limited. Better resolution, of the order of $25\,cm^{-1}$, is obtained by using dielectric filters as the spectral selection method. A set of filters can be employed which take advantage of the change in transmission maximum of these filters on tilting to cover a wide spectral range with a limited filter set.[24,25,36] A commercial device based upon this technology is available. An obvious alternative is to use a fixed filter and tune the excitation wavelength[37] but this increases the complexity of the laser used.

Electronically tunable image filters include acousto-optic tuneable filters (AOTFs).[38–40] The AOTF essentially creates an anisotropic spectral image filter which causes image blurring at high magnification.[41] This blurring can be eliminated to some degree by image processing. However the AOTF is limited by its minimum 30 cm^{-1} resolution and there have been no attempts to produce a commercial device based upon AOTF technology.

More recent developments include Fabry–Perot approaches[42] and liquid-crystal tuneable filters (LCTFs).[43–45] LCTFs have proved more practical, being rapidly electronically tuneable, with a large acceptance angle, and being shown to be free of image distortion effects. They also provide 8 cm^{-1} resolution at typical excitation wavelengths and so rival the spectral resolution achievable with conventional spectrometers. Two commercial devices are currently available using LCTF technology.

3.4.5 *Comparison of Raman Microscopy with other Imaging Methods*

The Raman microscopy approach, like the line illumination methods, has the advantage in cases where the full laser power cannot be focused down to a small spot on the sample due to laser damage. In practice, because of the decreased laser power per pixel with this approach, it is more common to use spectral microscopy to collect a single wavenumber image or at most a few key selected wavenumber images. In these circumstances, the selected spectral images may be collected relatively quickly. Therefore for inspection of a large number of similar samples for heterogeneity, where there is very clear spectral contrast which can be determined with only a few selected wavelengths, the spectral microscopy approach is ideal.

However, the Raman microscopy approach has its drawbacks which are detailed below:

- *Spectral resolution*. Inherent in the global microscopy approach is the use of filters. Such filtering methods are often limited in their spectral resolution as indicated above and this could be restrictive in many applications. However, the best filters are LCTFs and these are now available with 8 cm^{-1} resolution[43,45] and so rival the spectral resolution that is commonly used with point and line mapping instruments. However, it should be noted that the theoretical light throughput efficiency of LCTFs is 50% maximum (being a polarization-based device) and practically this falls to the region of 20% or worse at most wavelengths.
- *Uneven illumination*. The most common method of global illumination is to defocus the laser on to the sample. The typical Gaussian beam profile and the delivery optics will result in uneven illumination across the image. The distribution profile could be corrected by ratioing the instrument response using the signal from a standard (flat) sample. A patent has recently been applied for which promises to deliver far more uniform illumination across the field of view.[33]
- *Depth of field in the image*. Clearly the spectral microscopy approach does not provide confocal depth discrimination in the spectral image. Deconvolution methods can be used to improve the spatial resolution in the z-direction[46] but have yet to be routinely applied even in the commercial instruments. Therefore this approach is currently best suited to surface inspection or situations where the depth information is not important.
- *Lack of "post-interpretation"*. Images are collected at prechosen wavelengths, rather than full spectral information per point being collected. An intimate knowledge of the spectroscopy of the system (coupled with an assumption that nothing unexpected occurs) is therefore a necessary prerequisite to investigations using spectral microscopy. If subsequent spectral information is required, the sample has to be re-presented to the instrument and point microspectroscopy performed.
- *Quantitative accuracy*. In addition to the problems associated with uneven illumination and transfer function of the image to the camera, all of these spectral filtering methods simply collect all the light present at the chosen wavenumber shift including any background fluorescence. One approach to eliminate this is to take an image at a wavenumber shift "close" to the Raman band of interest but where no Raman scattering is expected. Subtraction of this "background" image or division of the Raman image by the background image (so-called flat-fielding) will then result in a more accurate image relating to the Raman intensity. Owing to the complex nature of backgrounds the quantitative accuracy of such approaches has to be questioned.

The main advantage of the Raman microscopy approach is the speed at which a single Raman image of the intensity contrast at a particular wavenumber can be obtained. Images are often presented at far higher image fidelity (pixel density) than are obtained with the mapping approaches (although the pixel density does not, of course, necessarily equate with spatial resolution, which is more directly related to the excitation wavelength, the numerical aperture of the objective lens and the optical properties of the sample itself).

Therefore the Raman microscopy approach is ideal for two-dimensional imaging with samples of high spectral contrast (where a few key wavenumber images can completely define the sample) and where depth resolution is not an issue. It also lends itself to the study of kinetic processes where spatial heterogeneity is important and to possible three-dimensional investigations which are described in the next section.

3.5 Three-dimensional Imaging Using Raman Spectroscopy

The confocal Raman microscope has the ability, in principle, to perform a depth-profiling function in much the same way as an optical confocal system (with the same limitations associated with scattering samples and laser absorption). Therefore, in principle, the ability to generate three-dimensional spectrally resolved images is available, using the z-axis control of the microscope stage.

Mapping and image reconstruction techniques are going to take several hours to complete, even with the most efficient hardware, but in static systems this may have some practical value. In this case image distortion is observed in the z-axis due to the point spread function of the observed Raman intensity and hence image deconvolution techniques may have to be used on the spectral image data slices to create a more faithful image in three dimensions.

Image deconvolution has also been applied to vertical sets of two-dimensional, wide-field, Raman microscopy images to produce a three-dimensional representation.[46] In any real practical application of three-dimensional Raman imaging, the Raman microscopy approach at a limited set of wavelengths is likely to prove most successful[47] although a direct comparison between the two approaches of Raman microscopy–deconvolution and three-dimensional confocal Raman point mapping has yet to be made.

4 EXPERIMENTAL CONSIDERATIONS

There are several other experimental issues that should be considered with regard to Raman microscopy and these are reviewed here.

4.1 Detection Limits

The detection limits of a range of surface-specific techniques are listed by Aller.[48] For Raman microscopy, the relative limit is listed as 100 ppm but the absolute limit is 10^{-12} g due to the spatial resolution at 1 μm which can be achieved using confocal optics. These detection limits assume that you can concentrate on a particular spot in a system where the analyte in question is known to reside. Spectral quality is then obtained by lengthy data acquisition as appropriate to achieve the 100-ppm discrimination.

A far more practical rule of thumb is that, for most organic species, levels of 1% should be observable by Raman. Specific molecules can vary however by a couple of orders of magnitude with respect to this estimate and specific detection also depends upon background and the presence or absence of other spectrally similar species.

4.2 Sampling Conditions

One of the major advantages of Raman microspectroscopy is that it is essentially a noninvasive, in situ tool, which performs under ambient conditions. The backscatter arrangement of the Raman microscope means that the technique is independent of the surface or substrate on which study is taking place.

The most important consequence of the ability to study systems under ambient conditions with little sample preparation is that Raman investigations will be easily geared up to ''look see'' investigations of chemical heterogeneity.

The poor scattering efficiency of water means that aqueous systems are studied without interference. This contrasts with the mid-IR where large regions of the spectrum are obscured by the IR absorbance spectrum of water.

4.3 Quantitative Applications

Raman spectra are capable of quantitative interpretation. If an internal standard is available, then such use can be extremely confident. However, without an internal standard, photometric accuracy, particularly in microscope systems, has to be considered carefully, as does the effect of polarization when dealing with anisotropic systems.

Nevertheless, *relative quantitation* between species in Raman microspectroscopy is more easily obtained, especially if a system calibration method is employed to ensure accurate comparisons across regions of the spectrum and between different sampling and detection arrangements.[20] The images that are commonly presented of molecular distribution, based on intensity, can certainly be interpreted on a relative concentration scale.

4.4 Instrument Response Correction

One of the underrated successes of FTRS has been to (re)introduce Raman spectroscopists to the fact that Raman is essentially a single beam technique. This became obvious when comparisons of spectra taken on different FT instruments were made. Relative peak heights were affected by the throughput characteristics of the bench (especially the filter systems employed in

early systems). The solution is to use a calibrated light source as a standard illumination source in the sample position collected under instrument conditions and ratio all subsequent spectra by this standard, analogous to the single beam background used in Fourier transform infrared (FTIR) spectroscopy.[20]

This will then correct for instrument effects, such as long-term drifts in response, as long as conditions are matched between the acquisition of the standard and the sample. Such methodology then allows confident comparison between instruments and also between different optical arrangements on the same instrument. This has obvious advantages and is equally applicable and recommended for CCD–monochromator instrumentation such as is used for Raman microscopy.

5 APPLICATION EXAMPLES

Although it is beyond the scope of this article to review the applications matrix in depth, it is helpful to indicate to some degree the richness of the potential application of Raman microscopy and imaging in academic and ''real world'' study. This section highlights a few examples of the application of Raman imaging, concentrating on studies of emulsion systems as an example of the level of information content in spectral imaging data sets and how such information can be extracted.

5.1 Polystyrene Spheres – a Raman Image Test Sample

The classic test sample for any Raman microscopy technique is polystyrene spheres. They can be purchased from standard chemical suppliers and have clear optical and spectral signatures which can be detected.[1,44,46] Given that they can be observed by both optical and Raman methods, they make an ideal spatial resolution test for a Raman microscope system.

In Figure 7 images of polystyrene spheres of a nominal 3.2-μm diameter are shown. Figure 7(a) is a classical bright field transmission optical microscopy image taken by optical microscopy and Figure 7(b) is a Raman intensity image extracted from the complete spectral image data set obtained by point mapping. The spectral image data set is 77×67 pixels at 0.35-μm step size and the Raman image has been created by integrating the intensity of all of the polystyrene bands observed.[1]

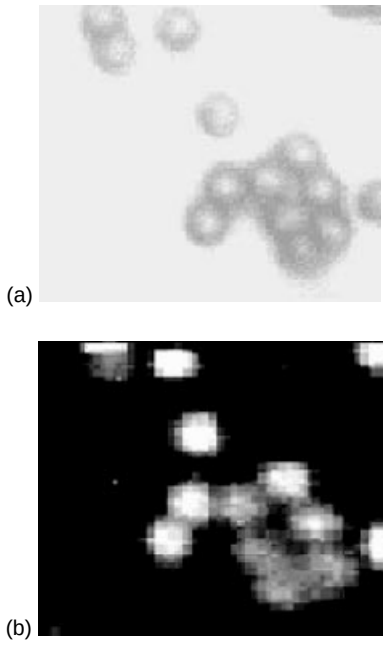

Figure 7 Images of polystyrene spheres (3.2 μm). (a) Bright field image; (b) Raman image. (Reproduced from J.J. Andrew et al. 'Raman Imaging of Emulsion Systems', *Appl. Spectrosc.*, **52**(6), 790–796 (1998) by permission of The Society for Applied Spectroscopy.)

5.2 Emulsion Systems

Emulsion systems, given their water-rich nature, are ideally suited to interrogation by Raman microscopy.[1,2] Figure 8 shows a point microspectroscopic investigation of an emulsion system shown registered to a differential interference contrast optical image. The emulsion has been described in detail in Andrew et al.[1] Such spectra show the presence of spectral (and hence molecular/phase) contrast in the emulsion.

A spectral image of the sample was then obtained by point mapping.[1] Figure 9 shows single wavelength

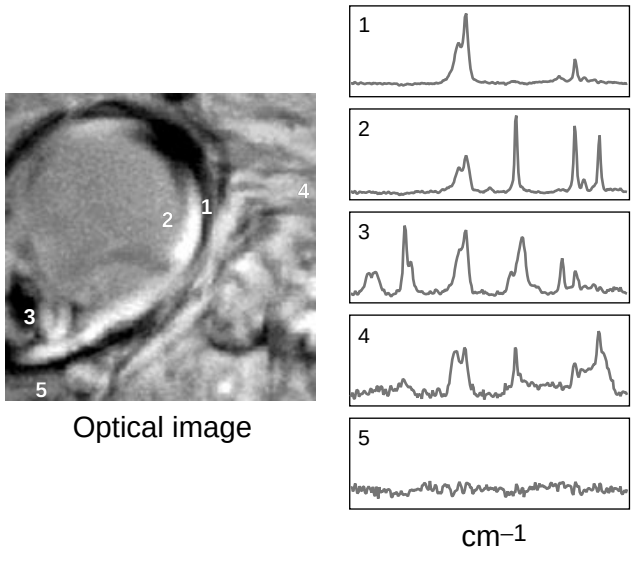

Figure 8 Point microspectroscopy of an emulsion system.

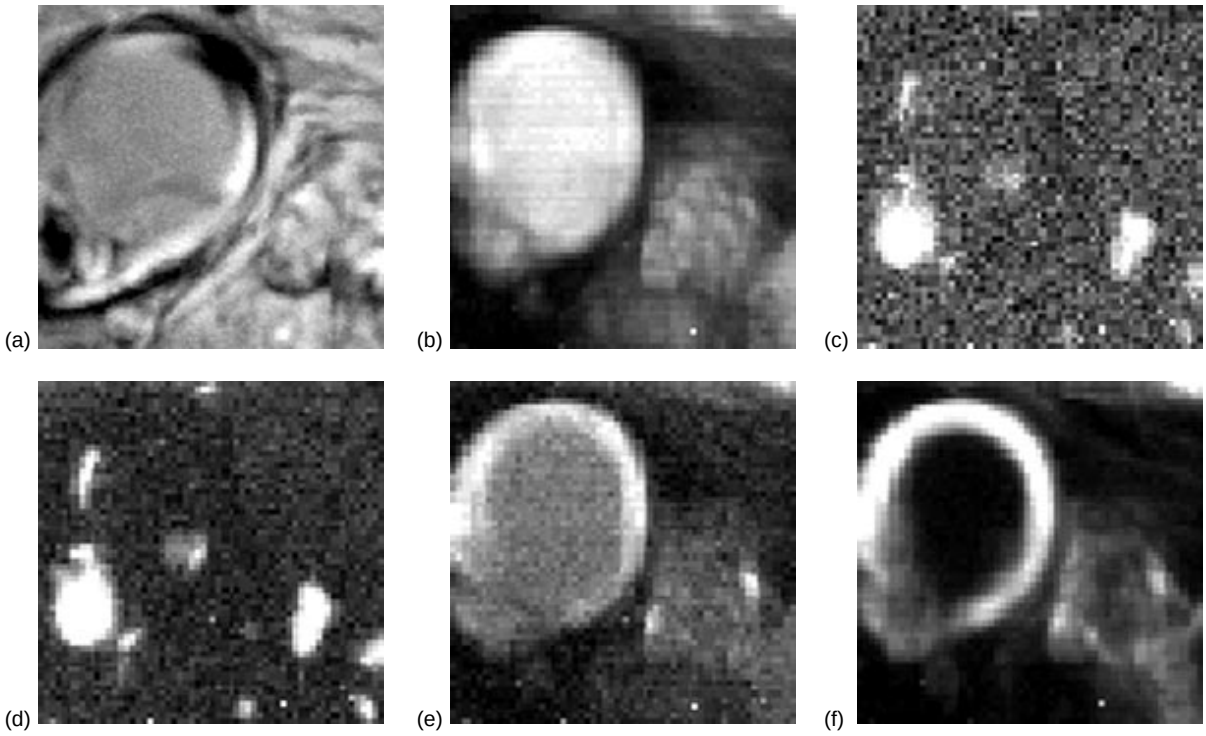

Figure 9 Single wavenumber intensity images of an emulsion data set. (a) Optical image; (b) 1472 cm^{-1} image; (c) 1720 cm^{-1} image; (d) 1621 cm^{-1} image; (e) 1116 cm^{-1} image; (f) 1308 cm^{-1} image. (Reproduced from J.J. Andrew, T.M. Hancewicz, 'Rapid Analysis of Raman Image Data Using Two-way Multivariate Curve Resolution', *Appl. Spectrosc.*, **52**(6), 797–807 (1998) by permission of The Society for Applied Spectroscopy.)

spectral intensity images reconstructed from the spectral image data set obtained. These images represent the intensity at a single wavenumber value and, whilst it is clear that there is spectral image contrast observed, the images show a significant amount of noise.

Such images based upon single wavenumber intensity are a poor way of using the spectral image data. In the polystyrene spheres example, integration of several bands was possible to improve the S/N in the image, as there was only one spectral signature in the data set. In the case of multiple component systems where there is much spectral overlap this is not practical. Other more efficient ways of image reconstruction using the information content in the entire spectral image data set are discussed in section 6.

5.3 Other Systems

Many other examples are seen in the literature of all of the approaches to producing Raman images. Indeed, the richness of the applications matrix is immense and, given that reliable, commercially available, hardware has only recently become available, and is not yet installed in many laboratories, it is expected that the number and range of applications can only grow.

Common examples that have been published in the open literature include analysis of carbon,[49] silicon,[28,50–52] silicon carbide,[53] diamond and diamond-like carbon[54,55] and semiconductors.[56,57] These examples are widespread not only due to their potential practical application in support of the computer industry but also to the fact that such samples have very favorable Raman scattering characteristics.

Microspectroscopy and Raman imaging have been applied to a wide range of polymer science examples beyond simple polystyrene spheres. Examples can be found in the literature;[22–26,39,45,58–61] these cover model systems (polystyrene spheres embedded in other polymer matrices), defects and inclusions in polymers, real-world mixed polymer systems, and polymer laminates (where confocal detection becomes important).

Raman imaging has found applications in other areas of material science,[31,62] and in the study of inorganic systems[63,64] and superconductors.[65–67]

Biological systems have also been imaged using Raman techniques. These include the in situ study of plant cell wall chemistry[68–70] and general cell and tissue chemistry in biochemical research, bacterial classification, and cell biology.[27,71–76] Some attention has also been given to the study of dental adhesive diffusion and dental porcelain in a simulated oral environment.[77,78]

Although few examples have yet appeared in the peer-reviewed literature, it is clear that the pharmaceutical industry can exploit the capability of Raman microscopy and imaging in understanding the properties of fully formulated products. Potential examples include investigating the association of active materials with excipients, particle size after processing in tablets, and the consequent effects on dissolution and active release. This is in addition to more obvious applications such as identification of contaminants or undesirable actives or polymorphs in both bulk drugs and fully formulated products.

Another field where the potential of Raman microspectroscopy has yet to be fully exploited is the area of forensic science.

6 DATA HANDLING AND DATA ANALYSIS

As has been mentioned above, spectral image data are at a minimum four-dimensional in nature and require some degree of interaction in order to be interpreted.

Specific software tools have been developed by several companies to deal with such data. They all contain essentially the same core features but differ in how interactive the investigator can be with the data set. More recently, it has become clear that the best way to deal with the complexity and to get the best from the enormous amount of information in spectral image data sets is to use sophisticated multivariate analysis techniques to reduce the data to their key elements. These issues are now discussed in more detail.

6.1 Multidimensional Spectral Data Handling and Univariate Analysis

At a bare minimum, software must allow for the generation of specific intensity images at chosen wavenumber (wavelength) and the generation of the full point spectra from each pixel within such an image. The manipulation of images, such as producing an image based upon the relative intensity at two wavenumbers (either by subtraction or division), is useful in dealing with backgrounds which are typically present in Raman images. Similarly, the ability to produce an image based on the integrated intensity under a peak can be used to improve S/N in an image of a particular molecular species. Such functions are usually found as integral parts of the spectrometer acquisition software although some off-line packages do exist for post-analysis of data.[1]

Commonly the user has to define the wavenumber(s) of interest prior to observing the spectral contrast image. This presupposes that the interrogator knows where the interesting spectral contrast information is in the data. A useful variation, of use in complex multicomponent, multiphase systems, is to have the ability to rapidly screen through the spectral image data set for spectral contrast (the equivalent of flicking rapidly through the deck of cards in Figure 1). Often it is observed that spectral contrast occurs in unexpected places in the spectrum and this information is useful in establishing a full understanding of the chemical heterogeneity in the sample.[1]

Raman image reconstruction based upon single wavenumber intensity or ratio of intensities at two wavenumbers is essentially a univariate approach. It is only using a small part of the spectral information that is present in the entire data set to image a particular molecular species. In order to obtain reasonable images which represent the molecular distribution, the univariate approach requires high S/N in each individual pixel spectrum and clear spectral separation of bands which are unique to the molecules or phases of interest. This applies constraints to the complexity of the systems that can be studied and also requires long data collection times for the spectral image data sets.[1]

In most systems, spectral signatures of the individual species will be overlapped and not easily separated and visualized in this way. However, it is clear that changes in signal intensity across the wavenumber range will be observed as you go through the spatial elements of the spectral image data sets and that for each species correlations between intensities at many wavenumber locations will be observed. Therefore a more efficient and elegant approach to use in the analysis of these data sets is multivariate analysis and this is described in the next section.

6.2 Multivariate Analysis

Given the problem of having to analyze very large data sets such as those generated by Raman imaging spectrometers, it is natural that data compression techniques such as principal factor analysis (PFA) (also referred to as principal component analysis (PCA)) would be a reasonable place to begin data analysis.[79–82] This has been the traditional way of dealing with so much data at one time.

However, the ultimate goal in spectral image analysis should be to extract as much meaningful chemical information as possible in a reasonable period of time. This means generating real or ''pure'' spectra of individual components or phases in a sample so that a proper chemical identification can be made. And it also means generating real or ''pure'' concentration or intensity information (maps in the case of image data) for each component so that the quantitative relationship between components is preserved at any spatial location on the sample image.

To move PFA analysis, which produces an abstract mathematical solution to the data reduction, to a

more ''real'' representation of the data requires a transformation step following PFA that converts the initial estimates of factors and scores into their ''real'' counterparts. This can be accomplished by applying alternating least-squares (ALS) optimization to the raw factors and scores.[83] The entire optimization process is often referred to as multivariate curve resolution (MCR) based on the curve resolution theory first developed by Lawton and Sylvestre.[84] The MCR process is closely related to other self-modeling algorithms that also use a least-squares transformation to purify variables.[85–90] However, given the typically large data arrays inherent in spectral image data, a quick and efficient method of analysis is paramount. In such instances, standard curve resolution methods prove to be too time-consuming and therefore impractical for routine application.

The theory of MCR has been thoroughly described in the literature, and it has been demonstrated to be an elegant solution to the problem of the analysis of multidimensional spectral data.[83,85,91–96]

Not only does the MCR data analysis approach lead to the automated generation of score plots, which are equivalent to concentration maps for each phase or chemical component but physically and spectroscopically meaningful factors associated with each score map that can be interpreted with conventional spectroscopic principles. Highly overlapping information

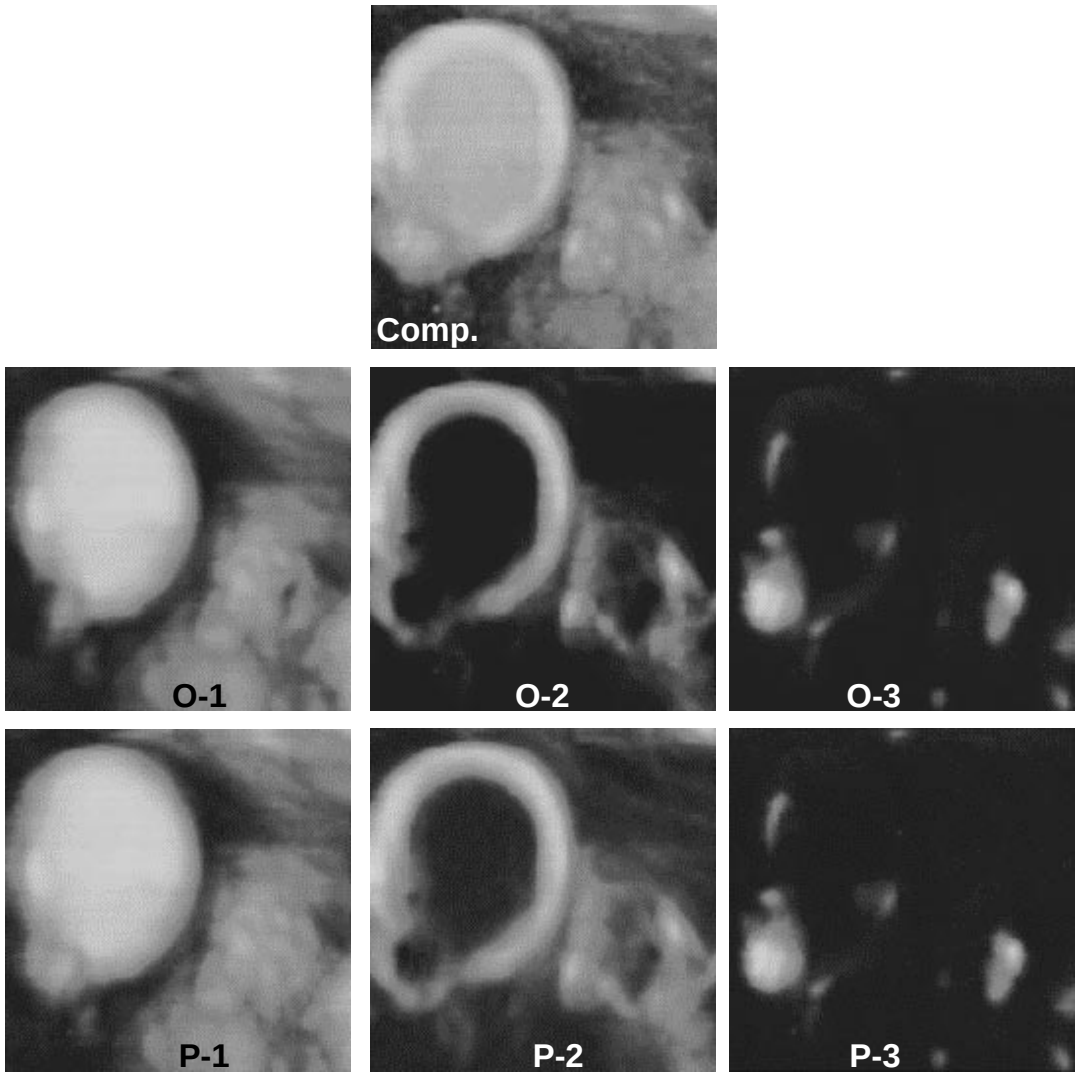

Figure 10 Multivariate score plot images of the emulsion data set shown in Figures 8 and 9. Compo, composite; O, orthogonal projection multivariate curve resolution (OPMCR); P, principal factor multivariate curve resolution (PFMCR). (Reproduced from J.J. Andrew, T.T. Hancewicz, 'Rapid Analysis of Raman Image Data Using Two-way Multivariate Curve Resolution', *Appl. Spectrosc.*, **52**(6), 797–807 (1998) by permission of The Society for Applied Spectroscopy.)

in the spectral dimension can be deresolved taking advantage of the fact that there is contrast in the spatial domain[83] and the spectroscopic interpretation of each factor is greatly simplified over interpreting the images derived from univariate image reconstruction.

Another advantage is that the spectroscopic signal inherent in the data set can effectively be separated from the random noise. This results in higher quality images than could be obtained by the univariate approach, or, alternatively, vastly decreased data acquisition times to obtain the necessary information content in the data set.[83] These two advantages are illustrated in Figures 10–12.

Figure 10 shows score plot images derived by two different multivariate approaches to determining spectral contrast in a spectral image data set.[83] The two approaches, orthogonal projection multivariate curve resolution (OPMCR) and principal factor multivariate curve resolution (PFMCR), are shown to produce equivalent results and both compare favorably with the univariate images shown in Figure 9. The corresponding MCR-generated factors, or spectra, are shown in Figure 11, compared with single spectra extracted from the original data set. Clearly it is valid to interpret the MCR-generated factors by conventional spectroscopic principles.

Figure 12 takes the information generated by the MCR approach for the emulsion system and represents the image data using color coding for the individual phases identified. This is a convenient method of representing the outcome of the MCR analysis.

It is the combined use of these two advantages – speed/automation of data interpretation and tolerance of extremely low S/N – that will make the concept of three-dimensional spatial Raman imaging using hardware confocal/point scanning a practical reality.

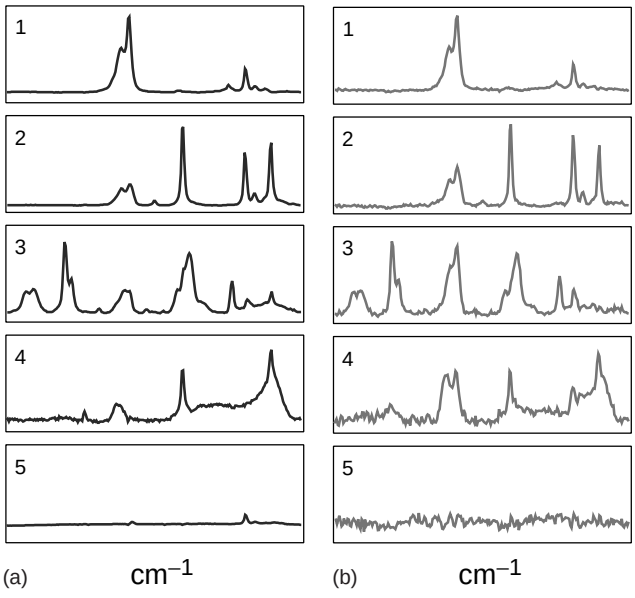

Figure 11 (a) MCR-generated spectra of the emulsion data set shown in Figures 8 and 9 compared with (b) "pure spectra" extracted from the data set.

7 COMPARISON WITH OTHER MICROSCOPY AND SPECTRAL IMAGING TECHNIQUES

Spectral microscopy in all its forms can be considered to be a subset of microscopy techniques in general, although it is more conventionally thought of as a subset of spectroscopy. The overlap with optical microscopy is becoming increasingly blurred and so a few comments on approaches available within this field follow.

Optical microscopy does provide some molecular information in certain circumstances and indeed absorption

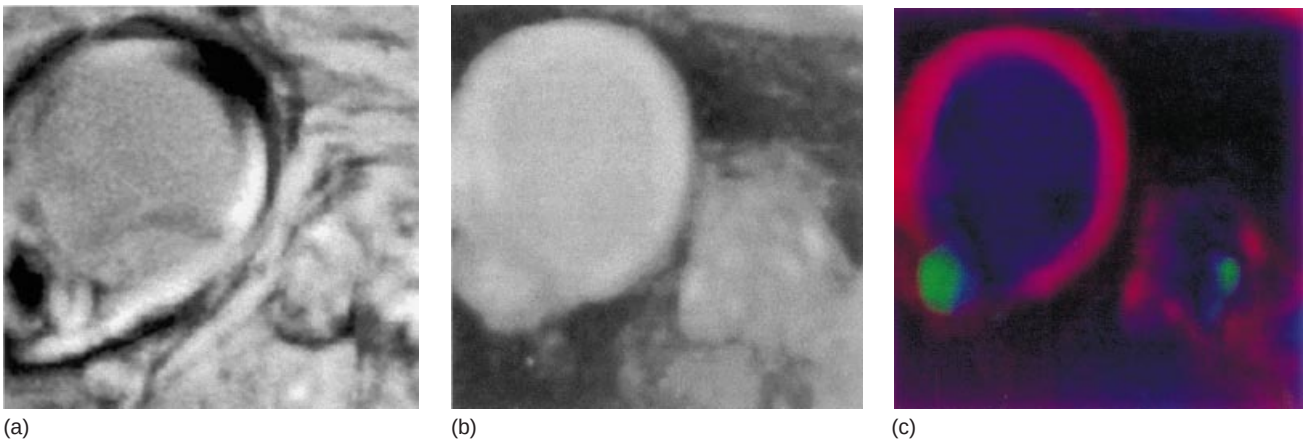

Figure 12 MCR-generated images showing the separation into distinct phases in the emulsion, with each phase color coded. (a) Optical image; (b) composite image; (c) false color composite score map image.

spectroscopy, essentially color contrast, can be used in microscopy as a tool to elucidate the spatial distribution of colored chemical species. Obviously such an approach is limited and the use of phase-specific stains provides a more flexible set of tools to generate contrast in optical microscopy. However, the information obtained is not quantitative and not as specific to the molecular heterogeneity as that obtained by spectral microscopy.

Fluorescence microscopy is usually considered as a subset of optical microscopy due to its wide ranging utility and application in that field. Again specific fluorescent stains for different phases or even macromolecular structures have been developed to allow specific visualization of heterogeneity. Fluorescent tags or labels can be applied to a range of molecules (assuming the synthetic capability is available) and this enables very high sensitivity and specificity for tackling the distribution of single molecular species in complex matrices. Unfortunately, particularly for small molecules, the act of labeling with a fluorescent moiety may change the molecular behavior compared with the untagged system. Also, the expense of labeled molecules (even off-the-shelf, let-alone custom synthesized) prohibits study of large-scale production systems using this approach. Detection of more than one molecule or structure requires multiple label tagging and detection systems, adding to the complexity of sample preparation (although automated detection systems for multiply labeled fluorescent microscopy are now routinely available). Some molecules may be autofluorescent (not requiring labeling to be detected) but this provides a solution in only a few favorable cases and the data obtained may prove to be less easy to interpret.

Spectral microscopy, using the inherent signals of molecules as the contrast element, provides a set of tools to analyze complex mixtures without the need for the complexity of staining, tagging or labeling. Intrinsically it can provide information on all of the molecules in the system in one experiment. The following debate is limited to techniques that fall into this class of methods.

7.1 Other Alternative Approaches to Spectral Microscopy

Raman microscopy has been compared with currently available techniques which can be considered for use in surface and near-surface chemical analysis by Aller.[48] These include electron microprobe (EMP), proton microprobe, IR microscopy, electron spectroscopy for chemical analysis, energy-dispersive X-ray analysis in electron microscopy, Auger electron spectroscopy, electron energy-loss spectroscopy (EELS), surface ionization mass spectroscopy (SIMS), laser microprobe mass spectrometry (LMMS) and field ion microscopy. The scanning probe microscopies, such as scanning tunneling microscopy and atomic force microscopy and other more recent developments such as scanning near-field optical microscopy (SNOM) and spectroscopic variants of SNOM, including Raman,[97–99] should also be mentioned in this regard but are currently of limited applicability.

Chemical, structural and distribution information are the primary concerns when choosing a technique to analyze chemical heterogeneity. Chemical analysis can be characterized either by the elemental (atomic) or the molecular information.

Elemental (atomic) analysis is probably the most familiar method employed as the molecular methods, such as the Raman approach presented here, have yet to be as universally established in analytical laboratories, although they have become increasingly common over the last 5 years. Elemental analysis is limited in understanding heterogeneity, particularly in biological or many common chemical systems where there is little elemental contrast. Another limitation is that, of the techniques listed above for elemental analysis, the majority require vacuum sampling.

Molecular identification is far more widely applicable but such information can only be obtained by EELS, SIMS or LMMS – all under vacuum – and by IR and Raman microscopy, which can be performed under ambient conditions. IR and Raman microscopy therefore occupy a unique niche in the analysis of heterogeneity and the comparison between these two close spectroscopic cousins is now made.

7.2 Comparison of Raman and Infrared Microscopy

In microspectroscopy, the IR microspectrometer systems (conventionally referred to as IR microscopes) attached to conventional FTIR benches have undoubtedly proved their worth in the analysis of small quantities of material such as contaminants and inclusions. Essentially they act as microsampling devices for IR spectroscopy and should not be strictly described as IR microscopes using the definitions given in section 2.3. More recently, IR equivalents of both approaches to acquiring a spectral image data set, mapping and image reconstruction and spectral microscopy as described in section 2.3 have been demonstrated.

IR mapping and image reconstruction again uses the obvious sample scanning approach, shown schematically in Figure 3, using a motorized mapping stage to move the sample pixel by pixel through the sampling area of the IR microscope.

The most elegant implementation of IR spectral microscopy does not use the obvious filter approach but instead employs an imaging step-scan FTIR interferometer to collect images of IR intensity at all retardations of the interferometer. Essentially this creates a two-dimensional spectral image data set of spatially resolved

interferograms. For visualization and interpretation of the spectral content, these data are then reorganized into a conventional two-dimensional spectral image data set (see Figure 1) by Fourier transforming the interferograms at each pixel.

In comparing the two vibrational spectroscopic techniques when used to obtain spectral image data, the Raman approach has a number of advantages over the IR. These are as follows:

1. An *xy* spatial resolution of 1 μm. This is an order of magnitude improvement over IR microspectroscopy in the fingerprint region of the spectrum. The limiting factor here is the diffraction limit of the wavelength of light being focused by the microscope. It should be stressed that this applies to ''best case'' situations. Opaque samples have the potential to ''fuzz'' the theoretical diffraction limit causing less than optimum spatial resolution.
2. A *z* resolution of the order of 2 μm. The IR approach mostly requires the absorption signal to be collected in transmission mode and hence the IR imaging data has no depth resolution.
3. Sample preparation. The ease of alignment and interpretation of data in the 180° backscatter arrangement of a Raman microscope requires little sample preparation. Indeed, the sample preparation is often identical to that performed for conventional epiluminescence microscopy. The use of reflection IR is limited to IR-reflective substrates and even then artifacts are observed in the spectral data. Indeed most IR microscopy experiments have been performed in transmission, which often requires lengthy sample preparation (thin-section microtomy) to produce samples that are able to transmit IR radiation.
4. The ability to study aqueous systems. Water is a weak Raman scatterer but a strong IR absorber. Systems with high water content, such as many biological samples and many industrial products, are not easily analyzed by IR but lend themselves very readily to Raman interrogation.
5. The use of conventional optical microscopes. Raman microscopy at most common excitation wavelengths can be performed on a conventional optical microscopy platform. This enables the sample to be viewed by all of the conventional optical contrast methods and the spectral information registered to these more familiar images. IR microscopy requires a specifically designed microscope employing reflective optics and this restricts you to poor quality bright-field optical microscopy in most cases.

Further comment should be made on the use of reflection IR microscopy.[100,101] Classical IR microscopy in reflection mode suffers from a depth of interrogation that is highly sample dependent and can only be applied to a limited number of (reflective) samples from which enough IR radiation is reflected to be detected. A potential solution to both of these problems in the point microspectroscopy experiment is to use an attenuated total reflectance (ATR) objective[100,101] that has been specifically designed to obtain a near-surface spectrum from a small area on a surface. This has been used successfully in obtaining a defined spectrum from surfaces[102,103] but it is difficult to clearly identify the actual sampled area. The ATR objective requires intimate physical contact between the objective and the surface under investigation and so it is practically most suitable for obtaining a single spectrum. However, one FTIR manufacturer has attempted to automate mapping using the ATR objective and an *xyz* controlled stage (in order to avoid the necessity for preparation of thin samples to perform IR mapping in transmission) but the ATR mapping approach will always be limited to only a few sample types.

However, it still remains that IR spectroscopy and interpretation is more familiar to most scientists than Raman spectroscopy. Certainly the spectral databases are far more exhaustive and readily available. The nature of the relationship between IR absorption and Raman scattering efficiency will mean that for some molecular systems IR will have a significantly better detection limit and contrast sensitivity than Raman and so the ideal situation is, of course, to have both tools available. However, if there is no clear specific reason for biasing towards IR in the systems under study, when faced with a decision of acquiring either Raman or IR microscopy, currently the argument implies that the choice should be for Raman microscopy first.

7.3 Fourier Transform Raman Microprobes

The body of this chapter describes the Raman microscopy technique that has become the de facto standard, that is the use of conventional microscopes, excitation in the region of 780 nm and dispersive spectrographs with CCD detection.

An alternative instrumental approach, at least to point microspectroscopy investigations, is the use of commercially available FT Raman microscopes attached to conventional FT Raman benches.[104]

These have been developed by the major FTIR manufacturers as add-on devices to the successful implementation of FTRS as a bulk spectroscopic technique, using the conventional 1064-nm excitation, at high powers to compensate for the decreased Raman scattering

efficiency at the longer wavelength. Not surprisingly, FT Raman microscopes borrow much of their design and technology from FTIR microscopy.

There are general points to be made about FT Raman microspectroscopy:

1. The power density required at the sample is a real problem in FT Raman microprobe systems where the energy is tightly focused. This often leads to thermal damage of the sample and thermal background observed in the spectrum; cooling the sample can sometimes help with these effects.
2. The best spatial resolution obtained by commercial instruments that has been reported is 4 μm.[104] Practically, in most cases spatial resolution will be worse.
3. Absorption of Raman scattered radiation by water in the region of the NIR used for excitation limits the quantitative potential in aqueous media when performing microspectroscopic measurements.
4. Such systems are limited to point mapping if imaging is to be performed.

It can be seen that many of the advantages which Raman microspectroscopy has over IR are brought into question in an FT Raman microprobe and the technique brings added complications primarily as a consequence of the high power density needed to generate adequate S/N. Nevertheless such systems may be a useful investment if you already have an FT Raman system available to you.

A more recent development has been the use of FT Raman with visible wavelength excitation.[105,106] This approach has been demonstrated as a method to allow low spatial resolution three-dimensional confocal imaging of immiscible liquids in a porous sandstone.[106] However, due to the problems associated with fluorescence when using visible wavelength excitation, it is not expected that the FT visible confocal Raman approach will be universally adopted.

8 PERSPECTIVE AND FUTURE DEVELOPMENTS

Raman microscopy can be considered to be a technique that has now reached a solid level of maturity. The basic design principles of the microscope systems have reached a de facto standard and the components of the system have been optimized such that the Raman experiment is now, in most cases, shot noise limited. Some subtle differences still remain between the major instrument suppliers but, essentially, investing in Raman microscopy hardware when the study of chemical heterogeneity is important should now be considered relatively risk free.

Any further practical improvement in the methodology is now limited to a few areas. It is likely that we will see a more common adoption of spatial multiplexing. That is, line scanning and spectral microscopy are both particularly applicable where depth resolution is not a concern, such as in surface imaging. The more common adoption of these techniques could well be dependent upon the development of much higher powered lasers in the 780-nm region. At higher power levels, the power density at the sample in the point scanning approach would prove to be a problem and hence the efficiency of distributing that power over a wider area of the sample by spatial multiplexing will become attractive.

It is also expected that further examples of full volumetric three-dimensional Raman imaging will become more common. There are some improvements to be made in more efficient instrument driver software to limit dead time in the data collection process during point mapping. Combined with the use of efficient analysis of low S/N data, such as MCR approaches, it will enable the use of point scanning confocal in three spatial dimensions. Further development of axial deconvolution procedures optimized for the wide-field spectral microscopy approach to three-dimensional imaging will take place.

Perhaps most importantly, software algorithms and rules for the most appropriate methods of data analysis in the study of highly complex, severely spectrally overlapped and high-noise/low-signal systems are being developed.[79–83] The routine application of these methods to all forms of spectral image data will widely expand the range of systems amenable to study and improve the efficiency of data interpretation.

One limitation of the current approaches to Raman imaging needs further mention. The spatial dimensions of the Raman imaging experiment are constrained at both the low and high ends. The limiting minimum spatial resolution is clearly controlled by the diffraction limit when using a conventional-microscope-based approach to the order of micrometer dimensions. SNOM-based approaches[97–99] promise to provide Raman images at much improved spatial resolution. However, the light efficiency of such devices limits them to only a few systems of high Raman scattering cross-section.

Perhaps less obviously, there is also a trade-off and limitation at the low spatial resolution/large image size. Essentially, because of the need to maintain high light gathering efficiency, the Raman experiment

is constrained to the use of high numerical aperture objectives. This places a limitation on the best trade-off between spatial resolution and image area covered in the point mapping approach. The sampled area cannot be increased to more than a few micrometers without a serious loss in Raman efficiency. This places a limitation on the maximum step or pixel size when mapping if you are to avoid sub-sampling the area of investigation.

An interesting approach to enable the efficient Raman imaging of large area samples at low spatial resolution is to use imaging fiber bundles.[81,82,107,108] These hold some promise in the analysis of samples where large area heterogeneity is important.

ACKNOWLEDGMENTS

The author wishes to acknowledge his co-workers in Raman at Unilever Research, in particular Allen Millichope, Ian E. Clark, Tom Hancewicz and Shuliang Zhang. Allen and Ian for those pioneering days in the dark, Tom for showing me the error(s) of my univariate ways and Shuliang for his help with the references used in this work and in taking the Raman technique to new heights. I should also acknowledge Mark Browne at Kinetic Imaging, Liverpool, UK, for assistance with software.

ABBREVIATIONS AND ACRONYMS

ALS	Alternating Least-squares
AOTF	Acousto-optic Tuneable Filter
ATR	Attenuated Total Reflectance
CCD	Charge-coupled Device
EELS	Electron Energy-loss Spectroscopy
EMP	Electron Microprobe
FT	Fourier Transform
FTIR	Fourier Transform Infrared
FTRS	Fourier Transform Raman Spectroscopy
IR	Infrared
LCTF	Liquid-crystal Tuneable Filter
LMMS	Laser Microprobe Mass Spectrometry
MCR	Multivariate Curve Resolution
NIR	Near-infrared
OPMCR	Orthogonal Projection Multivariate Curve Resolution
PCA	Principal Component Analysis
PFA	Principal Factor Analysis
PFMCR	Principal Factor Multivariate Curve Resolution
PMT	Photomultiplier Tube
SIMS	Surface Ionization Mass Spectroscopy
S/N	Signal-to-noise Ratio
SNOM	Scanning Near-field Optical Microscopy

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
Raman Spectroscopy in Analysis of Biomolecules

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Forensic Science **(Volume 5)**
Forensic Science: Introduction • Microspectrophotometry in Forensic Science

Pharmaceuticals and Drugs **(Volume 8)**
Vibrational Spectroscopy in Drug Discovery, Development and Production

Process Instrumental Methods **(Volume 9)**
Raman Spectroscopy in Process Analysis

Remote Sensing **(Volume 10)**
Hyperspectral Remote Sensing: Data Collection and Exploitation • Imaging Spectrometry for Geological Applications

Surfaces **(Volume 10)**
Infrared and Raman Spectroscopy in Analysis of Surfaces

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher • Signal Processing in Analytical Chemistry • Soft Modeling of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction • Microspectroscopy • Spectral Data, Modern Classification Methods for

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation • Raman Scattering, Fundamentals

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Multivariate Image Analysis • Quantitative Spectroscopic Calibration

REFERENCES

1. J.J. Andrew, M.A. Browne, I.E. Clark, T.M. Hancewicz, A.J. Millichope, 'Raman Imaging of Emulsion Systems', *Appl. Spectrosc.*, **52**(6), 790–796 (1998).
2. J.D. Louden, H.K. Patel, R.C. Rowe, 'A Preliminary Examination of the Structure of Gels and Emulsions Containing Cetostearyl Alcohol and Cetrimide Using Laser Raman Spectroscopy', *Int. J. Pharm.*, **25**(2), 179–190 (1985).
3. J.G. Grasselli, F. Walder, C. Petty, G. Kemeny, 'Industrial Applications of Raman Spectroscopy', *J. Mol. Struct.*, **294**, 207–210 (1993).
4. G.J. Rosasco, 'Raman Microprobe Spectroscopy', in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Clark, R.J. Hester, Heyden, London, 223–282, Vol. 7, 1980.
5. D.A. Long, *Raman Spectroscopy*, McGraw-Hill, London, 1977.
6. S.M. Mason, N. Conroy, N.M. Dixon, K.P.J. Williams, 'The Development of a Sensitive Near-IR Raman System', *Spectrochim. Acta*, **49A**(5–6), 633–643 (1993).
7. P.J. Hendra, G. Ellis, D.J. Cutler, 'Use of Optical Fibers in Raman Spectroscopy', *J. Raman Spectrosc.*, **19**(6), 413–418 (1988).
8. S.K. Sharma, C.L. Schoen, T.F. Cooney, 'Fiberoptic Remote Raman Probe Design for Use in Monitoring Processes in a High Temperature Oven', *Appl. Spectrosc.*, **47**(3), 377–379 (1993).
9. I.R. Lewis, P.R. Griffiths, 'Raman Spectrometry with Fiber-optic Sampling', *Appl. Spectrosc.*, **50**(10), A12–A30 (1996).
10. N. Everall, 'Use of a Holographic Rejection Filter to Improve the Sensitivity of a Commercial Raman Microprobe', *Appl. Spectrosc.*, **46**(5), 746–748 (1992).
11. C.L. Schoen, S.K. Sharma, C.E. Helsely, H. Owen, 'Performance of a Holographic Supernotch Filter', *Appl. Spectrosc.*, **47**(3), 305–308 (1993).
12. B. Yang, M.D. Morris, H. Owen, 'Holographic Notch Filter for Low Wavenumber Stokes and Anti-stokes Raman Spectroscopy', *Appl. Spectrosc.*, **45**(9), 1533–1536 (1991).
13. J.M. Tesdesco, H. Owen, D.M. Pallister, M.D. Morris, 'Principles and Spectroscopic Applications of Volume Holographic Optics', *Anal. Chem.*, **65**, 441A–449A (1993).
14. M. Kim, H. Owen, P.R. Carey, 'High Performance Raman Spectroscopic System Based on a Single Spectrograph, CCD Notch Filters and a Kr Laser Ranging from the Near-IR to Near-UV Regions', *Appl. Spectrosc.*, **47**(11), 1780–1783 (1993).
15. K.P.J. Williams, G.D. Pitt, B.J.E. Smith, A. Whitley, D.N. Batchelder, I.P. Hayward, 'Use of a Rapid-scanning Stigmatic Raman Imaging Spectrograph in the Industrial Environment', *J. Raman Spectrosc.*, **25**(1), 131–138 (1994).
16. D.E. Battey, J.B. Slater, R. Wludyka, H. Owen, D.M. Pallister, M.D. Morris, 'Axial Transmissive F/1.8 Imaging Raman Spectrograph with Volume-phase Holographic Filter and Grating', *Appl. Spectrosc.*, **47**(11), 1913–1919 (1993).
17. P.A. Walker III, W.K. Kowalchyk, M.D. Morris, 'On-line Raman Spectroscopy of Ribonucleotides Preconcentrated by Capillary Isotachophoresis', *Anal. Chem.*, **67**, 4255–4260 (1993).
18. K.F. Schrum, S.H. Ko, D. Ben-Amotz, 'Description and Theory of a Fiber-optic Confocal and Super-focal Raman Microspectrometer', *Appl. Spectrosc.*, **50**(9), 1150–1155 (1996).
19. B. Schrader, A. Hoffmann, S. Keller, 'Near-infrared Fourier Transform Raman Spectroscopy – Facing Absorption and Background', *Spectrochim. Acta.*, **47A**(9–10), 1135–1148 (1991).
20. F.T. Walder, M.J. Smith, 'Quantitative Aspects of Near-infrared Fourier Transform Raman Spectroscopy', *Spectrochim. Acta.*, **47A**(9–10), 1201–1216 (1991).
21. I.P. Hayward, K.J. Baldwin, D.M. Hunter, D.N. Batchelder, G.D. Pitt, 'Direct Imaging and Confocal Mapping of Diamond Films Using Luminescence and Raman Scattering', *Diamond Relat. Mater.*, **4**(5–6), 617–621 (1995).
22. L. Markwort, B. Kip, E. Da Silva, B. Roussel, 'Raman Imaging of Heterogeneous Polymers – a Comparison of Global Versus Point', *Appl. Spectrosc.*, **49**(10), 1411–1430 (1995).
23. C.R. Kagan, T.D. Harris, A.L. Harris, M.L. Schilling, 'Submicron Confocal Raman Imaging of Holograms in Multicomponent Photopolymers', *J. Chem. Phys.*, **108**(16), 6892–6896 (1998).
24. T. Kasteleiner, R. Evans, J. Yarwood, D. Hodge, R. Swart, 'Raman Microscopic Studies of the Distribution of the Fungicide Chlorthalonil in a PVC Film', *J. Raman Spectrosc.*, **27**(9), 695–698 (1996).
25. M. Claybourn, A. Luget, K.P.J. Williams, 'Raman Microscopy and Imaging of Polymers', *Multidimen. Spectrosc. Polym.*, **598**, 41–60 (1995).
26. L. Markwort, B. Kip, 'Micro-raman Imaging of Heterogeneous Polymer Systems: General Applications and Limitations', *J. Appl. Polym. Sci.*, **61**(2), 231–254 (1996).
27. C.J. De Grauw, C. Otto, J. Greve, 'Line-scan Raman Microspectrometry for Biological Applications', *Appl. Spectrosc.*, **51**(11), 1607–1612 (1997).
28. M. Bowden, D.J. Gardiner, 'Stress and Structural Images of Microindented Silicon by Raman Microscopy', *Appl. Spectrosc.*, **51**(9), 1405–1409 (1997).
29. M. Bowden, D.J. Gardiner, G. Rice, D.L. Gerrard, 'Line-scanned Micro-Raman Spectroscopy Using a

Cooled CCD Imaging Detector', *J. Raman Spectrosc.*, **21**(1), 37–41 (1990).
30. K.A. Christensen, M.D. Morris, 'Hyperspectral Raman Microscopic Imaging Using Powell Lens Line Illumination', *Appl. Spectrosc.*, **52**(9), 1145–1147 (1998).
31. N.L. Jestel, J.M. Shaver, M.D. Morris, 'Hyperspectral Raman Line Imaging of an Aluminosilicate Glass', *Appl. Spectrosc.*, **52**(1), 64–69 (1998).
32. A. Feofanov, S. Shaonov, P. Valisa, E. Da Silva, I. Nabiev, M. Manfait, 'A New Confocal Stigmatic Spectrometer for Micro-Raman and Micro-fluorescence Spectral Imaging Analysis – Design and Applications', *Rev. Sci. Instrum.*, **66**(5), 3146–3158 (1995).
33. E. Da Silva, M. Delhaye, B. Kip, B. Lenain, *Proceedings of 14th International Conference on Raman Spectroscopy*, John Wiley and Sons, Chichester, 1106, 1994.
34. P. Dhamelincourt, F. Wallart, M. LeClercq, A.T. Nguyen, D.O. Landon, 'Laser Raman Molecular Microprobe (MOLE)', *Anal. Chem.*, **51**(3), A414–A421 (1979).
35. P. Dhamelincourt, P. Bisson, 'Principle and Realization of an Optical Microscope Using Raman Effect', *Microsc. Acta*, **79**(3), 267–276 (1977).
36. D.N. Batchelder, C. Cheng, W. Muller, B.J.E. Smith, 'A Compact Raman Microprobe-microscope – Analysis of Polydiacetylene Langmuir and Langmuir-Blodgett Films', *Makromol. Chem. Symp.*, **46**, 171–179 (1991).
37. G.J. Puppels, M. Grond, J. Greve, 'Direct Imaging Raman Microscope Based Upon Tunable Wavelength Excitation and Narrow-band Emission Detection', *Appl. Spectrosc.*, **47**(8), 1256–1267 (1993).
38. P.J. Treado, I.W. Levin, E.N. Lewis, 'High-fidelity Raman Imaging Spectrometry – a Rapid Method Using an Acoustooptic Tunable Filter', *Appl. Spectrosc.*, **46**(8), 1211–1216 (1992).
39. M.D. Schaeberle, C.G. Karakatsanis, C.J. Lau, P.J. Treado, 'Raman Chemical Imaging – Non-invasive Visualization of Polymer Blend Architecture', *Anal., Chem.*, **67**(23), 4316–4321 (1995).
40. S.R. Goldstein, L.H. Kidder, T.M. Herne, I.W. Levin, E.N. Lewis, 'The Design and Implementation of a High-fidelity Raman Imaging Microscope', *J. Microsc., Oxford*, **184**(1), 35–45 (1996).
41. H.R. Morris, C.C. Hoyt, P.J. Treado, 'Imaging Spectrometers for Fluorescence and Raman Microscopy – Acoustooptic and Liquid-crystal Tunable Filters', *Appl. Spectrosc.*, **48**(7), 857–866 (1994).
42. K.A. Christensen, N.L. Bradley, M.D. Morris, R.V. Morrisson, 'Raman Imaging Using a Tunable Dual-stage Liquid-crystal Fabry-Perot Interferometer', *Appl. Spectrosc.*, **49**(8), 1120–1125 (1995).
43. H.R. Morris, C.C. Hoyt, P. Miller, P.J. Treado, 'Liquid Crystal Tunable Filter Raman Chemical Imaging', *Appl. Spectrosc.*, **50**(6), 805–811 (1996).
44. H.R. Morris, C.C. Hoyt, P.J. Treado, 'Imaging Spectrometers for Fluorescence and Raman Microscopy – Acoustooptic and Liquid-crystal Tunable Filters', *Appl. Spectrosc.*, **48**(7), 857–866 (1994).
45. H.R. Morris, B. Munroe, R.A. Ryntz, P.J. Treado, 'Fluorescence and Raman Chemical Imaging of Thermoplastic Olefin (TPO) Adhesion Promotion', *Langmuir*, **14**(9), 2426–2434 (1998).
46. A. Govil, D.M. Pallister, M.D. Morris, '3-dimensional Digital Confocal Raman Microscopy', *Appl. Spectrosc.*, **47**(1), 75–79 (1993).
47. P.J. Treado, personal communication, 1999.
48. A.J. Aller, 'The Challenge of Spectroscopy in the Microanalysis of Biological Surfaces', *Appl. Spectrosc. Rev.*, **27**(3), 223–244 (1992).
49. K. Ray, R.L. McCreery, 'Spatially Resolved Raman Spectroscopy of Carbon Electrode Surfaces: Observations of Structural and Chemical Heterogeneity', *Anal. Chem.*, **69**(22), 4680–4687 (1997).
50. K. Mizoguchi, S. Nakashima, H. Harima, T. Hara, 'Raman Image Study of Defects in Ion-implanted and Post-annealed Silicon', in *Proceedings of the 18th International Conference on Defects in Semiconductors*, Materials Science Forum, Trans Tech Publications, Sweden, 1547–1552, Vol. 196–210, 1995.
51. K. Mizoguchi, S. Nakashima, Y. Sugiura, H. Harima, 'Micro-Raman Characterization of Crystallinity of Laser- recrystallized Silicon Films on SiO_2 Insulators', *J. Appl. Phys.*, **85**(9), 6758–6762 (1999).
52. C.D. Poweleit, A. Gunther, S. Goodnick, J. Menendez, 'Raman Imaging of Patterned Silicon Using a Solid Immersion Lens', *Appl. Phys. Lett.*, **73**(16), 2275–2277 (1998).
53. F. Neri, S. Trusso, C. Vasi, F. Barreca, P. Valisa, 'Raman Microscopy Study of Pulsed Laser Ablation Deposited Silicon Carbide Films', *Thin Solid Films*, **332**(1–2), 290–294 (1998).
54. T. Minabe, K. Hashimoto, A. Fujishima, 'Two-dimensional Raman Imaging Analysis of Microcrystal Diamond Film', *Chem. J. Chin. Univ.-Chin.*, **16**(11), 46–49 (1995).
55. K.W. Nugent, S. Prawer, 'Confocal Raman Strain Mapping of Isolated Single CVD Diamond Crystals', *Diamond Relat. Mater.*, **7**(2–5), 215–221 (1998).
56. F.A. Ponce, J.W. Steeds, C.D. Dyer, G.D. Pitt, 'Direct Imaging of Impurity-induced Raman Scattering in GaN', *Appl. Phys. Lett.*, **69**(18), 2650–2652 (1996).
57. P.D. Wang, C. Cheng, C.M.S. Torres, D.N. Batchelder, 'GaAs Micrometer-sized Dot Imaging by Raman Microscopy', *J. Appl. Phys.*, **74**(9), 5907–5909 (1993).
58. N.J. Everall, 'Measurement of Orientation and Crystallinity in Uniaxially Drawn Poly(ethylene terephthalate) Using Polarized Confocal Raman Microscopy', *Appl. Spectrosc.*, **52**(12), 1498–1504 (1998).
59. M.L. Scheepers, R.J. Meier, L. Markwort, J.M. Gelan, D.J. Vanderzande, B.J. Kip, 'Determination of Free Melamine Content in Melamine–Formaldehyde Resins

by Raman Spectroscopy', *Vibrat. Spectrosc.*, **9**(2), 139–146 (1995).
60. K.P.J. Williams, I.C. Wilcock, 'Raman Spectroscopy of Polymers', *Polym. Polym. Compos.*, **5**(6), 443–449 (1997).
61. S.L. Zhang, J.A. Pezzuti, M.D. Morris, A. Appadwedula, C.M. Hsiung, M.A. Leugers, D. Bank, 'Hyperspectral Raman Line Imaging of Syndiotactic Polystyrene Crystallinity', *Appl. Spectrosc.*, **52**(10), 1264–1268 (1998).
62. K.P.J. Williams, I.C. Wilcock, I.P. Hayward, A. Whitley, 'Applications of State-of-the-art Raman Microscopy and Direct Raman Imaging – Progress in Corrosion, Mineralogy, and Materials Research', *Spectrosc.*, **11**(3), 45–50 (1996).
63. H.W. Messenger, 'Spectral Imaging – Raman Imaging Reveals Details of Mars Meteorite Composition', *Laser Focus World*, **33**(11), 15–15 (1997).
64. J.R. Schoonover, F. Weesner, G.J. Havrilla, M. Sparrow, P.J. Treado, 'Integration of Elemental and Molecular Imaging to Characterize Heterogeneous Inorganic Materials', *Appl. Spectrosc.*, **52**(12), 1505–1514 (1998).
65. K.T. Wu, A.K. Fischer, V.A. Maroni, M.W. Rupich, 'Raman Microscopy Examination of Phase Evolution in Bi(Pb)–Sr– Ca–Cu–O Superconducting Ceramics', *J. Mater. Res.*, **12**(5), 1195–1204 (1997).
66. V.A. Maroni, A.K. Fischer, K.T. Wu, 'Examination of Phase Evolution in High-Tc Superconducting Ceramics by Raman Microspectroscopy and Imaging Raman Microscopy', *Spectrosc.*, **12**(7), 38–45 (1997).
67. V.A. Maroni, A.K. Fischer, K.T. Wu, 'Using Imaging Raman Microscopy to Explore Phase Evolution in Composite Ceramic Superconductors', *Abstr. Papers Am. Chem. Soc.*, **215**, U662–U662 (1998).
68. D.S. Himmelsbach, S. Khahili, D.E. Akin, 'Near-infrared–Fourier Transform–Raman Microspectroscopic Imaging of Flax Stems', *Vibrat. Spectrosc.*, **19**(2), 361–367 (1999).
69. J.S. Bond, U.P. Agarwal, R.H. Atalla, 'Raman Imaging of Lignocellulosic Materials', *Abstr. Papers Am. Chem. Soc.*, **216**, U345–U345 (1998).
70. J.S. Bond, U.P. Agarwal, R.H. Atalla, 'Raman Imaging of Woody Plant Cell Walls', *Abstr. Papers Am. Chem. Soc.*, **217**, U254–U254 (1999).
71. G.J. Puppels, T.C.B. Schut, N.M. Sijtsema, M. Grond, F. Maraboeuf, C.G. deGrauw, C.G. Figdor, J. Greve, 'Development and Application of Raman Microspectrocopic and Raman Imaging Techniques for Cell Biological Studies', *J. Mol. Struct.*, **347**, 477–483 (1995).
72. C. Otto, C.J. deGrauw, J.J. Duindam, N.M. Sijtsema, J. Greve, 'Applications of Micro-Raman Imaging in Biomedical Research', *J. Raman Spectrosc.*, **28**(2–3), 143–150 (1997).
73. N.M. Sijtsema, S.D. Wouters, C.J. deGrauw, C. Otto, J. Greve, 'Confocal Direct Imaging Raman Microscope: Design and Applications in Biology', *Appl. Spectrosc.*, **52**(3), 348–355 (1998).
74. N.M. Sijtsema, J.J. Duindam, G.J. Puppels, C. Otto, J. Greve, 'Imaging with Extrinsic Raman Labels', *Appl. Spectrosc.*, **50**(5), 545–551 (1996).
75. N.J. Kline, P.J. Treado, 'Raman Chemical Imaging of Breast Tissue', *J. Raman Spectrosc.*, **28**(2–3), 119–124 (1997).
76. R. Goodacre, E.M. Timmins, R. Burton, N. Kaderbhai, A.M. Woodward, D.B. Kell, P.J. Rooney, 'Rapid Identification of Urinary Tract Infection Bacteria Using Hyperspectral Whole-organism Fingerprinting and Artificial Neural Networks', *Microbiology*, **144**, 1157–1170 (1998).
77. N.L. Jestel, M.D. Morris, W.J. O'Brien, 'Depth-resolved Raman Microprobe Examination of a Commercial Dental Porcelain Exposed to a Simulated Oral Environment', *Dental Mater.*, **14**(5), 375–382 (1998).
78. D.M. Wieliczka, M.B. Kruger, P. Spencer, 'Raman Imaging of Dental Adhesive Diffusion', *Appl. Spectrosc.*, **51**(11), 1593–1596 (1997).
79. C.A. Drumm, M.D. Morris, 'Microscopic Line-imaging with Principal Component Analysis', *Appl. Spectrosc.*, **49**(9), 1331–1337 (1995).
80. C.A. Hayden, M.D. Morris, 'Effects of Sampling Parameters on Principal Component Analysis of Raman Line Images', *Appl. Spectrosc.*, **50**(6), 708–714 (1996).
81. C.M. Stellman, K.S. Booksh, M.L. Myrick, 'Multivariate Raman Imaging of Simulated and 'Real World' Glass-reinforced Composites', *Appl. Spectrosc.*, **50**(5), 552–557 (1996).
82. C.M. Stellman, K.S. Booksh, A.R. Muroski, M.P. Nelson, M.L. Myrick, 'Principal Component Mapping Applied to Raman Microspectroscopy of Fiber-reinforced Polymer Composites', *Sci. Eng. Comp. Mater.*, **7**(1–2), 51–80 (1998).
83. J.J. Andrew, T.M. Hancewicz, 'Rapid Analysis of Raman Image Data Using Two-way Multivariate Curve Resolution', *Appl. Spectrosc.*, **52**(6), 797–807 (1998).
84. W.H. Lawton, E.A. Sylvestre, 'Mathematical Determination of the Pure Spectra of Two Components in a Two Component Mixture', *Technometrics*, **13**, 617–633 (1971).
85. E.R. Malinowski, *Factor Analysis in Chemistry*, Wiley-Interscience, New York, 1992.
86. R. Tauler, D. Barcelo, 'Multivariate Curve Resolution Applied to Liquid-chromatography Diode Array Detection', *Trac-Trend Anal. Chem.*, **12**(8), 319–327 (1993).
87. M. Maeder, 'Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks', *Anal. Chem.*, **59**(3), 527–530 (1987).
88. H.R. Keller, D.L. Massart, J.O. de Beer, 'Window Evolving Factor Analysis for Assessment of Peak Homogeneity in Liquid Chromatography', *Anal. Chem.*, **65**(4), 471–475 (1993).

89. B. Vandeginste, R. Essers, T. Bosman, J. Reijnen, G. Kateman, '3-Component Curve Resolution in Liquid Chromatography with Multiwavelength Diode-array Detection', *Anal. Chem.*, **57**(6), 971–985 (1985).
90. E.R. Malinowski, 'Window Factor Analysis – Theoretical Derivation and Application to Flow-injection Analysis Data', *J. Chemometr.*, **6**(1), 29–40 (1992).
91. M.A. Sharaf, D.L. Illman, B.R. Kowalski, *Chemometrics*, Wiley-Interscience, New York, 1986.
92. E.R. Malinowski, 'Theory of Error in Factor Analysis', *Anal. Chem.*, **49**(4), 606–612 (1977).
93. H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuhler, 'Calculation of Equilibrium Constants from Multiwavelength Spectroscopic Data. 3. Model-free Analysis of Spectrophotometric and Electron Spin Resonance Titrations', *Talanta*, **32**(12), 1133–1139 (1985).
94. B.G.M. Vandeginste, W. Derks, G. Kateman, 'Multicomponent Self-modeling Curve Resolution in High Performance Liquid Chromatography by Iterative Target Transformation Analysis', *Anal. Chim. Acta*, **173**, 253–264 (1985).
95. F.C. Sanchez, J. Toft, O.M. Kvalheim, D.L. Massart, 'Eigenstructure Tracking Analysis for Assessment of Peak Purity in High Performance Liquid Chromatography with Diode Array Detection', *Anal. Chim. Acta*, **314**(3), 131–139 (1995).
96. S.J. Vanslyke, P.D. Wentzell, 'Real-time Principal Component Analysis Using Parallel Kalman Filter Networks for Peak Purity Analysis', *Anal. Chem.*, **63**(21), 2512–2519 (1991).
97. C.L. Jahncke, M.A. Paesler, H.D. Hallen, 'Raman Imaging with Near-field Scanning Optical Microscopy', *Appl. Phys. Lett.*, **67**(17), 2483–2485 (1995).
98. V. Deckert, D. Zeisel, R. Zenobi, T. Vo-Dinh, 'Near-field Surface Enhanced Raman Imaging of Dye-labeled DNA with 100-nm Resolution', *Anal. Chem.*, **70**(13), 2646–2650 (1998).
99. S. Webster, D.A. Smith, D.N. Batchelder, 'Raman Microscopy Using a Scanning Near-field Optical Probe', *Vibrat. Spectrosc.*, **18**(1), 51–59 (1998).
100. L. Rintoul, H. Panayiotou, S. Kokot, G. George, G. Cash, R. Frost, T. Bui, P. Fredericks, 'Fourier Transform Infrared Spectrometry: a Versatile Technique for Real World Samples', *Analyst*, **123**(4), 571–577 (1998).
101. J.M. Chalmers, N.J. Everall, K. Hewitson, M.A. Chesters, M. Pearson, A. Grady, B. Ruzicka, 'Fourier Transform Infrared Microscopy: Some Advances in Techniques for Characterisation and Structure–Property Elucidations of Industrial Material', *Analyst*, **123**(4), 579–586 (1998).
102. R.K. Dukor, M.N. Liebman, B.L. Johnson, 'A New, Non-destructive Method for Analysis of Clinical Samples with FT/IR Microspectroscopy. Breast Cancer Tissue as an Example', *Cell. Mol. Biol.*, **44**(1), 211–217 (1998).
103. R.J. Markovich, C.A. Evans, C.B. Coscolluela, S.A. Zibas, J. Rosen, 'Spectroscopic Identification of an Amorphous-to-crystalline Drug Transition in a Solid Dispersion SCH 48461 Capsule Formulation', *J. Pharm. Biomed. Anal.*, **16**(4), 661–673 (1997).
104. A.J. Sommer, J.E. Katon, 'Performance Enhancement of a Fourier-transform Raman Microprobe', *Spectrochim. Acta A*, **49**(5–6), 611–620 (1993).
105. C.J.H. Brenan, I.W. Hunter, 'Design and Characterisation of a Visible Light Fourier Transform Raman Spectrometer', *Appl. Spectrosc.*, **49**(8), 1086–1093 (1995).
106. C.J.H. Brenan, I.W. Hunter, J.M. Brenan, 'Noninvasive Confocal Raman Imaging of Immiscible Liquids in a Porous Medium', *Anal. Chem.*, **69**(1), 45–50 (1997).
107. J.Y. Ma, D. Ben-Amotz, 'Rapid Micro-Raman Imaging Using Fiber-bundle Image Compression', *Appl. Spectrosc.*, **51**(12), 1845–1848 (1997).
108. B.J. Marquardt, D.N. Stratis, D.A. Cremers, S.M. Angel, 'Novel Probe for Laser-induced Breakdown Spectroscopy and Raman Measurements Using an Imaging Optical Fiber', *Appl. Spectrosc.*, **52**(9), 1148–1153 (1998).

Raman Scattering, Fundamentals

Jürgen Popp and Wolfgang Kiefer
Universität Würzburg, Würzburg, Germany

1 Introduction **1**
2 Fundamentals of Linear Raman Spectroscopy **2**
2.1 Classical Treatment of Raman Scattering 3
2.2 Short Quantum Mechanical Description of Raman Scattering 8
2.3 Enhancing Raman Signals 9
3 Fundamentals of Nonlinear Raman Spectroscopy **12**
3.1 Spontaneous Nonlinear Raman Scattering: Hyper-Raman Effect 12
3.2 Stimulated Raman Scattering 13
3.3 Coherent Anti-Stokes–Raman Scattering 15
3.4 Femtosecond Time-resolved Nonlinear Raman Scattering 17
4 Instrumentation for Linear Raman Scattering **19**
4.1 Dispersive Raman Spectroscopy 19
4.2 Fourier Transform Raman Spectroscopy 21
4.3 Special Raman Techniques 23
5 Instrumentation for Nonlinear Raman Scattering **23**
5.1 Hyper-Raman Spectroscopy 24
5.2 Coherent Anti-Stokes–Raman Scattering 24
5.3 Stimulated Raman Gain Spectroscopy and Inverse Raman Spectroscopy 30
5.4 Photoacoustic Raman Spectroscopy 31
5.5 Ionization-detected Stimulated Raman Spectroscopy 31
Acknowledgments **33**
Abbreviations and Acronyms **33**
Related Articles **33**
References **34**

An introduction of the fundamentals of linear and nonlinear Raman spectroscopy is given. The Raman effect is the result of inelastic light scattering. A small amount of the photon energy of the incident light wave is modulated by the molecular scattering system. An energy transfer occurs as a result of the coupling between the incident radiation and the quantized states of the scattering system. Depending on the coupling, the incident photons either gain or lose energy. The light, which has less energy than the incident laser light, is named Stokes–Raman scattering, and the radiation, which has more energy, is referred to as anti-Stokes–Raman scattering.

In the case of the coupling between strong laser fields and molecular vibrations the observation of nonlinear Raman effects such as hyper-Raman scattering, stimulated Raman scattering (SRS), coherent anti-Stokes–Raman spectroscopy, the Raman gain spectroscopy, etc., is possible.

Apart from theoretical aspects of Raman spectroscopy an introduction into the instrumentation of linear and nonlinear Raman techniques is provided. For linear Raman spectroscopy two alternate approaches are described: dispersive Raman and Fourier transform Raman (FT-Raman) spectroscopy. Special Raman techniques such as micro-Raman spectroscopy and difference Raman spectroscopy are discussed.

In addition, a review of the instrumentation of several nonlinear Raman methods which are based on the second-order ($\chi^{(2)}$) and the third-order nonlinear susceptibility ($\chi^{(3)}$) is given. These methods include coherent anti-Stokes–Raman scattering (CARS), stimulated Raman gain spectroscopy (SRGS), inverse Raman scattering (IRS), photoacoustic Raman spectroscopy (PARS), and ionization-detected stimulated Raman spectroscopy (IDSRS).

1 INTRODUCTION

The Raman effect was first theoretically predicted by A. Smekal[1] in 1923, followed by quantum mechanical descriptions by Kramers and Heisenberg[2] in 1925, and Dirac[3] in 1927. The first experimental evidence for the inelastic scattering of light by molecules such as liquids was observed by Raman and Krishnan[4] in 1928 and almost simultaneously by Landsberg and Mandelstam from solids.[5] After the experimental verification of the Raman effect, the new scattering technique was mainly used for the investigation of organic and inorganic compounds in liquid or solid state or in solution.[6] With the development of the laser in the 1960s and the appropriate detection devices such as multichannel arrays and charge-coupled devices (CCDs cameras), Raman spectroscopy experienced a dramatic growth in analytical applications. Apart from the spectroscopy of phonons,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the Raman effect has been used for studying other quasi particles in solids such as plasmons and polaritons.[7] The development of high intensity laser systems also led to the discovery of several new nonlinear light-scattering processes.[8–22]

The Raman effect is the result of inelastic light scattering. A small amount of the photon energy of the incident light wave (nowadays only lasers are used as light sources) is modulated by the molecular scattering system. An energy transfer occurs as a result of the coupling between the incident radiation and the quantized states of the scattering system. Depending on the coupling, the incident photons either gain or lose energy. Since only a small percentage of the incident energy is changed, most of the radiation will have the same frequency after the interaction. The latter is called the elastically scattered light or Rayleigh scattering. The light which has less energy compared to the incident laser light is named Stokes–Raman scattering and the radiation which has larger energy is referred to as anti-Stokes–Raman scattering.

Figure 1 shows a simple model of the scattering mechanism. A photon with the energy $\hbar w_L$ is incident on the scattering system with the energy level $\hbar w_R = E_f - E_i$, where i and f label two quantum states. The Stokes–Raman effect results from the transition from the lower energy level (E_i) to a higher one (E_f). The anti-Stokes effect transfers energy from the system to the incident light wave, which corresponds to the transition from a higher energy level (E_f) to a lower one (E_i). Since the anti-Stokes scattering occurs from a thermally excited state (E_f), which according to Boltzmann statistics is less populated than the ground state (E_i), the anti-Stokes intensity is less than the Stokes intensity. Hence, in most cases only the Stokes–Raman spectrum is detected.

The transition observed in Raman spectroscopy from molecules can be between rotational, vibrational or electronic eigenlevels. However, the observation of Raman scattering is in some cases limited by the excitation of fluorescence, which is typically several orders of magnitude stronger than Raman scattering. In solids, Raman scattering can be observed from phonons as well as from other quasi particles.[7] In the following we will primarily discuss Raman scattering from molecules.

2 FUNDAMENTALS OF LINEAR RAMAN SPECTROSCOPY

When light is incident on matter it can interact with the atoms or molecules in different ways. Depending on the energy of the radiation, the photons can be absorbed directly or can be scattered. Absorption of light is most likely if the wavelength of the radiation is in the infrared (IR), visible (VIS) or in the ultraviolet (UV). The IR absorption yields the excitation of rotational and vibrational modes of the molecule, while the UV/VIS absorption results in the excitation of an electronic transition which mostly is followed by fluorescence. Apart from absorption, the incident light can be scattered by atoms and molecules. Investigation of the angular frequency components of the scattered light reveal that most of it is scattered elastically at the same angular frequency. Additionally, there is a small amount of radiation with shifted angular frequency. This

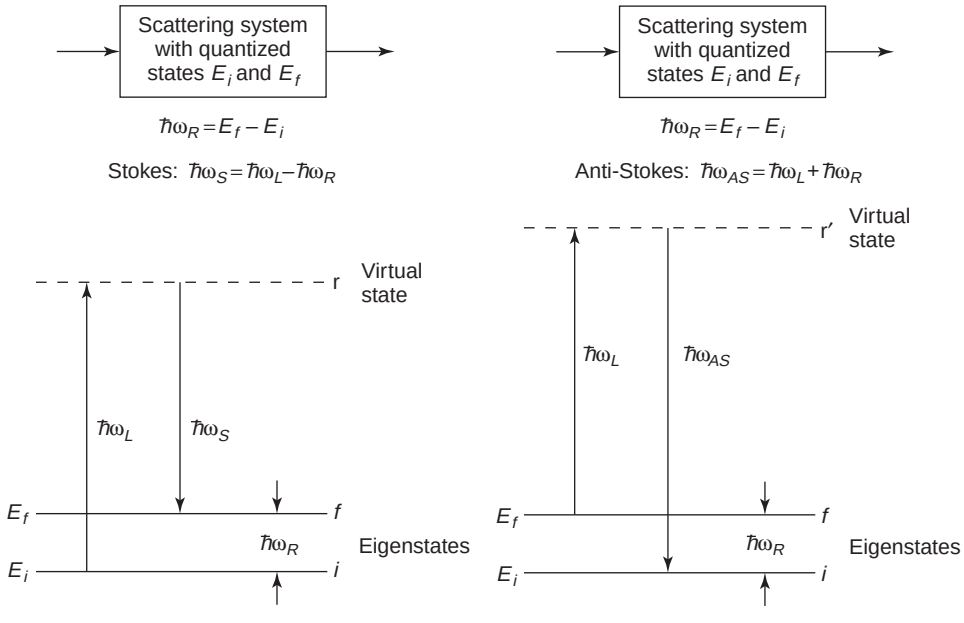

Figure 1 Simple model illustrating Stokes- and anti-Stokes–Raman scattering.

inelastically scattered radiation can occur in the spectrum at lower ($w_S = w_L - w_R$) or higher angular frequencies ($w_{AS} = w_L + w_R$) compared to the excitation angular frequency w_L. The scattered light at w_S and w_{AS} is, as already mentioned above, the so-called Stokes- and anti-Stokes–Raman scattering, respectively. The angular frequency w_R corresponds to a rotational or vibrational molecular transition.

Raman spectroscopy is an important method for investigating molecular vibrations. Thus, Raman spectroscopy gives precise information about the chemical structure of molecules and it also can be used to monitor chemical reactions. Since the scattering efficiency of the nonresonant Raman effect is several orders lower than that of fluorescence, the obtained Raman signal is very weak. The introduction of lasers has stimulated this field of molecular spectroscopy in various manners. Combined with the development of new spectrometers and sensitive detectors, Raman spectroscopy experienced a renaissance. New linear (e.g. resonance Raman spectroscopy (RRS), surface-enhanced Raman spectroscopy (SERS), micro-Raman spectroscopy) and nonlinear Raman spectroscopic techniques (e.g. SRS, CARS) have been developed and successfully used in academic and industrial research. Topics covered by Raman spectroscopy include inorganic and organic chemistry, physics, biology, medicine, geology, etc.

In the following section Raman scattering is described by a simple classic treatment, which gives precise information about the location of Raman frequencies as well as first insights into the selection rules of Raman scattering. However, this classical approach can not yield the proper intensity ratio of Stokes to anti-Stokes–Raman bands. Based on the Boltzmann statistics the experimentally determined ratio of the Stokes to the anti-Stokes–Raman intensity can be used for the determination of the spectroscopic temperature of the sample under investigation. The depolarization ratio ρ will be introduced as a quantity which yields information about the symmetry of molecules, randomly oriented such as in liquids.

2.1 Classical Treatment of Raman Scattering

Most of the experimental results obtained by Raman spectroscopy can be interpreted by a simple classical theory, because many of the mathematical expressions used are valid under certain conditions even in a quantum mechanical treatment.[23] In this classical approach the molecules are described as an ensemble of atoms performing simple harmonic vibrations, without taking into account the quantization of the rotational and vibrational energy levels.

According to the classical theory, the incident electromagnetic field induces an electric dipole moment of the scattering system (molecule or solid). Such an induced dipole moment is oscillating with the frequency of the incident radiation and is acting as a secondary source for electromagnetic radiation. The light is scattered into a solid angle of 4π. The intensity of the scattered radiation can be deduced from the classical theory.

The incident electric (time-dependent) field vector is given by Equation (1):

$$\vec{E} = \vec{E}_0 \cos w_L t \tag{1}$$

which induces the electric dipole moment $\vec{\mu}$, as in Equation (2):

$$\vec{\mu} = \alpha \vec{E} \tag{2}$$

The proportionality quantity α is the polarizability of the scattering system, $\vec{E}_0$ is the amplitude and w_L is the angular frequency of the incident field. α represents a second rank tensor and Equation (2) is the short notation of the following linear equation system given in Equation (3):

$$\begin{aligned} \mu_x &= \alpha_{xx}E_x + \alpha_{xy}E_y + \alpha_{xz}E_z \\ \mu_y &= \alpha_{yx}E_x + \alpha_{yy}E_y + \alpha_{yz}E_z \\ \mu_z &= \alpha_{zx}E_x + \alpha_{zy}E_y + \alpha_{zz}E_z \end{aligned} \tag{3}$$

According to these equations, the three components of the electromagnetic field in Cartesian coordinates and the nine components of the polarizability tensor generate three components of the induced dipole moment. α is

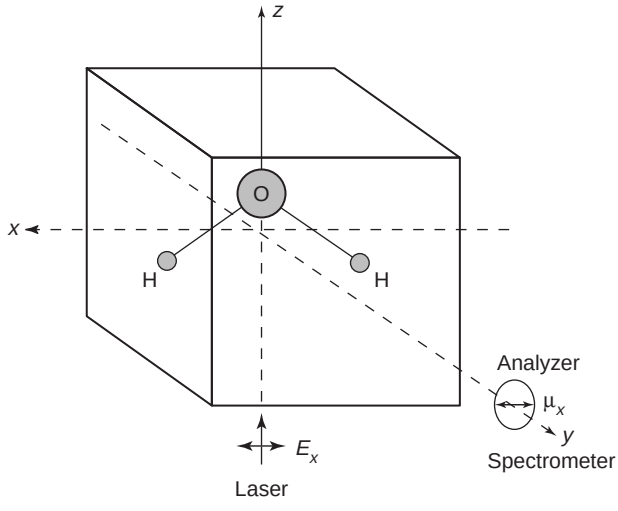

Figure 2 90° Raman scattering of an oriented (with fixed coordinates) scattering system (e.g. oriented water molecule or crystal) illuminated with laser radiation, which propagates in the z-direction and is polarized in the x-direction. The scattered light is observed at 90° in the y-direction, using a polarizer which allows only the observation of x-polarized light.

in many cases a symmetric matrix, and thus there are only six independent components for most molecules (i.e. $\alpha_{\rho\sigma} = \alpha_{\sigma\rho}$).

First we consider the easiest case that only one component of $\overrightarrow{E}$ (e.g. E_x) does contribute to one component of $\vec{\mu}$ (e.g. μ_x). This corresponds to a Raman scattering geometry which is depicted in Figure 2. A water molecule or a crystal is fixed to the xz-plane of the Cartesian coordinate system. The incident radiation propagates in the z-direction and is polarized in the x-direction. According to the interaction of the polarized light wave with the oriented molecule, different dipole components may be excited inside the molecule. The scattered light is observed at 90° in the y-direction. Using a polarizer which allows only the observation of x-polarized light, the observed dipole moment is given in scalar form by Equation (4):

$$\mu_x = \alpha_{xx} E_x \tag{4}$$

In a simple short notation, without using any indices, Equation (4) can be written as Equation (5):

$$\mu = \alpha E \tag{5}$$

or in combination with Equation (1), it becomes Equation (6):

$$\mu = \alpha E_0 \cos w_L t \tag{6}$$

In this case, E_0 represents the amplitude of electric field component which oscillates in the x-direction and induces the corresponding dipole moment.

According to the laws of classical electrodynamics an oscillating electric dipole emits radiation of the same frequency at which the dipole oscillates. If the polarizability α does not change with time ($\alpha = \text{const.}$), then the induced dipole will re-emit the incident radiation of the angular frequency w_L (Rayleigh scattering). However, if the polarizability consists of terms which depend upon vibrational frequencies, these terms will modulate the incident field. According to such time dependent changes in the polarizability ($\alpha = \alpha(t)$) of vibrating molecules, new frequency components will be observed in the scattered radiation. In order to describe a certain lattice vibration or normal mode of a molecule, a harmonic oscillator with the angular frequency w_R is considered. Its time dependent amplitude is given by Equation (7):

$$q = q_0 \cos w_R t \tag{7}$$

In the case that the molecular vibration affects the polarizability and if the vibrational amplitude is small, the polarizability can be expanded in a Taylor series around its equilibrium position $q = 0$, to give Equation (8):

$$\alpha = \alpha(q) = \alpha_0 + \left(\frac{\partial \alpha}{\partial q}\right)_0 q + \cdots \tag{8}$$

where only terms up to first-order are considered. Introducing Equation (7) in (8), and Equation (8) in (2) results in Equation (9):

$$\mu = \alpha_0 + \left(\frac{\partial \alpha}{\partial q}\right)_0 q_0 \cos w_R t \; E_0 \cos w_L t \tag{9}$$

or, alternatively, Equation (10):

$$\begin{aligned} \mu = {} & \underbrace{\alpha E_0 \cos w_L t}_{\text{Rayleigh scattering}} \\ & + \underbrace{\frac{1}{2}\left(\frac{\partial \alpha}{\partial q}\right)_0 q_0 E_0 \cos(w_L - w_R)t}_{\text{Stokes–Raman scattering}} \\ & + \underbrace{\frac{1}{2}\left(\frac{\partial \alpha}{\partial q}\right)_0 q_0 E_0 \cos(w_L + w_R)t}_{\text{anti-Stokes–Raman scattering}} \end{aligned} \tag{10}$$

According to Equation (10) the scattered radiation of the induced dipole moment consists of three components with the angular frequencies w_L, $w_S = w_L - w_R$, and $w_{AS} = w_L + w_R$, which correspond to Rayleigh-, Stokes–Raman, and anti-Stokes–Raman scattering, respectively, as in Equation (11):

$$\mu = \mu^R + \mu^S + \mu^{AS} \tag{11}$$

From this simple derivation one can also deduce that μ^S and μ^{AS} will contribute to the induced dipole moment only if the polarizability does change during the vibration ($(\partial\alpha/\partial q)_0 \neq 0$). From this we can further draw some conclusion about the Raman activity of a certain vibration. As an example for this the molecule CO_2 is considered. A linear molecule such as CO_2 will have $3N - 5$ normal modes whereas nonlinear molecules will have $3N - 6$ normal modes, where N indicates the number of atoms in the molecule. Thus, CO_2 has four normal modes, from which two are degenerate. In Figure 3 both the normal modes and the corresponding changes of the polarizability of CO_2 during a vibration are displayed. Only for the symmetric stretching vibration ν_s of CO_2 $(\partial\alpha/\partial q)_0$ is nonzero. Therefore, this mode is allowed in Raman scattering. Both other modes, the antisymmetric stretching vibration ν_{as} and the degenerate deformation vibration ν_{def} are not observed in a Raman spectrum because the derivative of the polarizability at $q = 0$ vanishes. Those modes are called Raman inactive.

In order to derive an expression for the Raman intensities we consider an oscillating electric dipole with

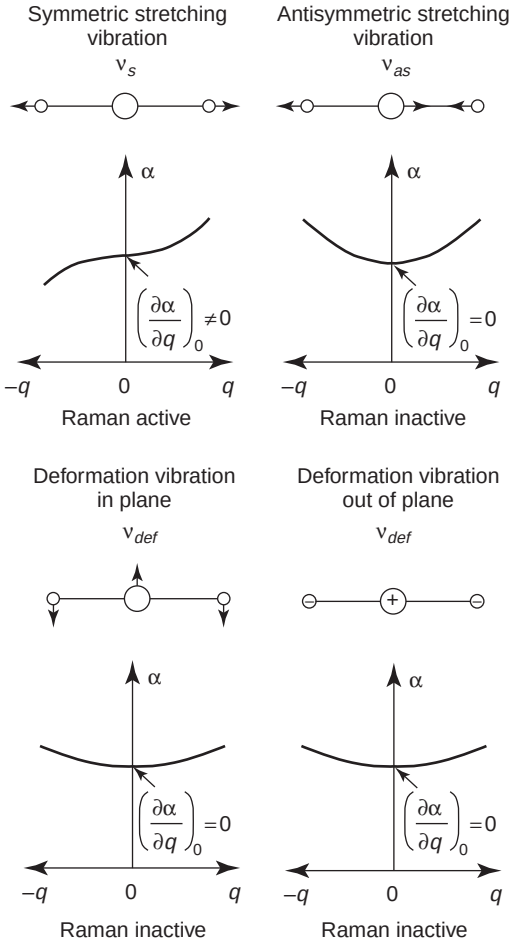

Figure 3 Classical derivation of selection rules for Raman scattering using the molecule CO_2.

angular frequency w, in Equation (12),

$$\mu = \mu_0 \cos wt \tag{12}$$

where μ_0 is the amplitude. The total emitted power derived by classical theory is given by Gerthsen et al.,[24] according to Equation (13):

$$P = \frac{1}{6\pi\varepsilon_0 c^3}|\ddot{\mu}|^2 \tag{13}$$

where $\ddot{\mu}$ is the second derivative of the dipole moment in respect to time, P is the emitted power, ε_0 is the permittivity of vacuum, and c is the speed of light. Since an optical detector can not resolve optical frequencies in the order of $\approx 10^{15}$ Hz, an intensity averaged over time will be detected. Using Equation (12) and averaging over time ($\overline{\cos^2 wt} = 1/2$) gives the average (over the solid angle of 4π) rate of the total emitted power, as in Equation (14):

$$\overline{P} = \frac{w^4}{12\pi\varepsilon_0 c^3}\mu_0^2 \tag{14}$$

where $\overline{P}$ is the total emitted power. Thus, the scattered intensity per solid angle in 90° direction (such as depicted in Figure 2) yields Equation (15)[25]

$$I = \frac{\overline{P}}{8\pi/3} = \frac{w^4}{32\pi^2\varepsilon_0 c^3}\mu_0^2 \tag{15}$$

where I is the scattered intensity per solid angle in the 90° direction. Now, consider a dipole μ which is located at the origin of a Cartesian coordinate system and does vibrate along a fixed direction. From such an oscillating dipole, a detector which is located in the y-direction without any polarizer in front, will detect the μ_x and μ_z but not the μ_y component of the scattered intensity, caused by the transverseness of radiation (see also Figure 4), as in Equation (16):

$$I = \frac{w^4}{32\pi^2\varepsilon_0 c^3}(\mu_x^2 + \mu_z^2) \tag{16}$$

2.1.1 Polarizability

In the same way as an oscillating permanent dipole, an oscillating induced dipole can be treated as a classical electromagnetic emitter. Because of Equation (2) and (3), various components of the induced dipole moment contribute to Raman scattering. First, the different contributions to Raman scattering of molecules with fixed orientation in a Cartesian coordinate system (such as crystals) are considered, followed by randomly oriented molecules (such as gases, liquids, and amorphous solids).

The most common geometry in Raman spectroscopy is a 90° arrangement (see Figure 5). The incident electromagnetic radiation with electric field vector either

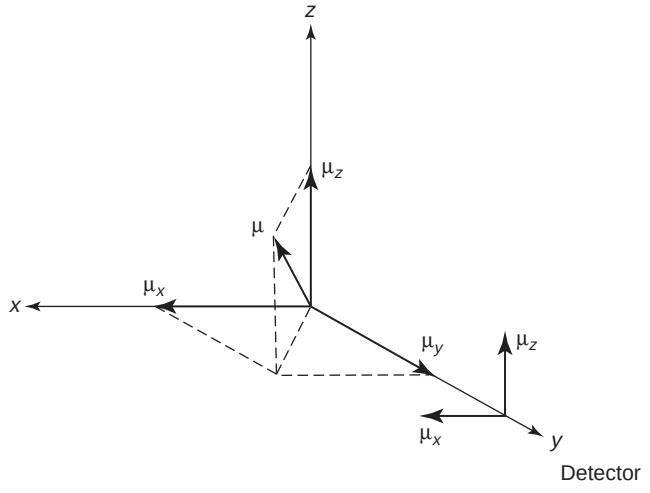

Figure 4 Illustration of the effect of an analyzer in front of a detector (see text).

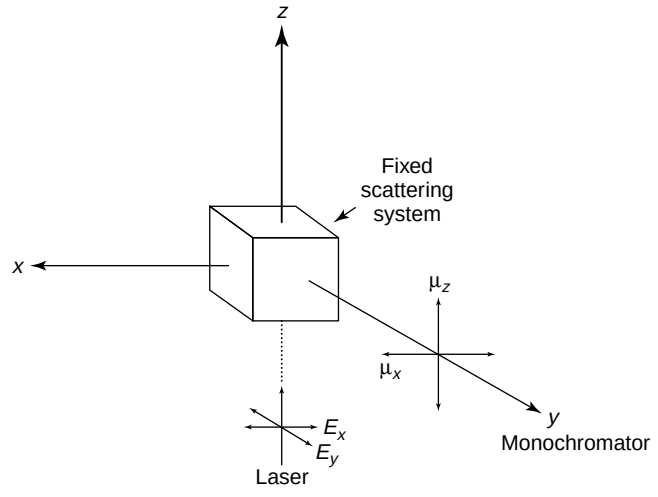

Figure 5 90° Raman scattering geometry of an oriented (with fixed coordinates) scattering system.

in the x- or y-direction propagates along the z-axis. For both polarization cases the induced dipole moments will consist of three components. Detecting the scattered light in the y-direction allows only the observation of the components μ_x and μ_z, according to Equations (17):

$$\begin{aligned} \mu_x &= \alpha_{xx} E_x \\ \mu_z &= \alpha_{zx} E_x \\ \mu_x &= \alpha_{xy} E_y \\ \mu_z &= \alpha_{zy} E_y \end{aligned} \tag{17}$$

Therefore, the entire scattering intensity (including the scattered Rayleigh intensity at $w = w_L$, the Stokes–Raman intensity at $w = w_s$, and the anti-Stokes–Raman intensity at $w = w_{as}$ in the y-direction for x-polarized light is given by Equation (18):

$$\begin{aligned} I_{x(x+z)} &= \frac{w^4}{32\pi^2\varepsilon_0 c^3}(\mu_x^2 + \mu_z^2) \\ &= \frac{w^4}{32\pi^2\varepsilon_0 c^3}(\alpha_{xx}^2 + \alpha_{zx}^2)E_x^2 \end{aligned} \tag{18}$$

and for y-polarized light, Equation (19):

$$I_{y(x+z)} = \frac{w^4}{32\pi^2\varepsilon_0 c^3}(\alpha_{xy}^2 + \alpha_{zy}^2)E_y^2 \tag{19}$$

Using the incident field intensity I_0 instead of the field amplitudes,[25] gives Equation (20):

$$I_0 = \frac{c\varepsilon_0}{2}E_x^2 = \frac{c\varepsilon_0}{2}E_y^2 \tag{20}$$

Combining Equation (20) with Equations (18) and (19) yields Equations (21) and (22):

$$I_{x(x+z)} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0(\alpha_{xx}^2 + \alpha_{zx}^2) \tag{21}$$

$$I_{y(x+z)} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0(\alpha_{xy}^2 + \alpha_{zy}^2) \tag{22}$$

where the first index indicates the polarization of the incident laser beam and the second index characterizes the polarization of the observed scattering. Since in the example above no polarizer was used in the detection line both components (x and z) of the induced dipole are observed. The use of a polarizer in front of the detector results in the suppression of one component, and thus, only one component has to be considered in Equations (21) and (22). In the case of a parallel polarized analyzer in front of the detector Equations (21) and (22) changes, as given by Equations (23) and (24):

$$I_{xx} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 \alpha_{xx}^2 \tag{23}$$

$$I_{yx} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 \alpha_{xy}^2 \tag{24}$$

If a perpendicular polarized analyzer is used the following components of Equations (21) and (22) are observed (Equations 25 and 26):

$$I_{xz} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 \alpha_{zx}^2 \tag{25}$$

$$I_{yz} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 \alpha_{zy}^2 \tag{26}$$

Analyzing scattered light from a crystal whose main axes have been oriented parallel with respect to the axes of the chosen scattering geometry allows directly the investigation of the different components of the polarizability tensor. Since the main axes of a crystal can be oriented in different ways the Porto notation[26] has been introduced: $a(bc)d$, where a and b indicate the propagation direction and the polarization of the incident light wave, respectively, and d and c characterize the direction of observation and the polarization of the scattered light, respectively. In Figure 6 four examples are given for the measurement of the different tensor components (α_{zz}, α_{xz}, α_{xx}, and α_{yx}).

In the following part of this section the Raman scattering of randomly oriented molecules such as gases, liquids, and amorphous solids is discussed. The scattered Raman intensity can be calculated by averaging over all components of every single molecule over the solid angle of 4π. In those cases one can not distinguish

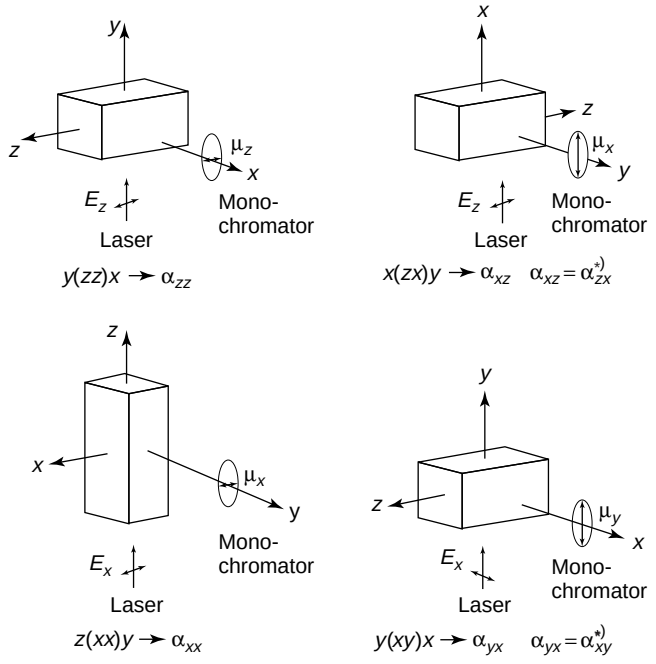

Figure 6 Raman scattering geometries combined with Porto notation[26] for the investigation of the different components of the polarizability tensor of a crystal (*) for symmetric tensors).

experimentally between the different components $\alpha_{\rho\sigma}$ of the polarizability tensor. However, there are two invariants of the tensor components $\alpha_{\rho\sigma}$ which are constant regardless of the orientation of the molecules.[27] The first invariant is the mean value which can be expressed as Equation (27)

$$\overline{\alpha} = \tfrac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \tag{27}$$

and second, the anisotropy, as in Equation (28):

$$\gamma^2 = \tfrac{1}{2}\quad (\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{yy} - \alpha_{zz})^2 + (\alpha_{zz} - \alpha_{xx})^2 + 6(\alpha_{xy}^2 + \alpha_{yz}^2 + \alpha_{zx}^2) \tag{28}$$

According to Wilson et al.[28] the averaged squared scattering tensor elements can be written as Equations (29) and (30):

$$\overline{\alpha_{xx}^2} = \overline{\alpha_{yy}^2} = \overline{\alpha_{zz}^2} = \frac{45\bar{\alpha}^2 + 4\gamma^2}{45} \tag{29}$$

$$\overline{\alpha_{xy}^2} = \overline{\alpha_{xz}^2} = \overline{\alpha_{yz}^2} = \frac{\gamma^2}{15} \tag{30}$$

For a scattering system like in Figure 5 the scattering intensity of randomly oriented molecules results then in Equations (31) and (32):

$$I_{x(x+z)} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 N \frac{45\bar{\alpha}^2 + 7\gamma^2}{45} \tag{31}$$

$$I_{y(x+z)} = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 N \frac{2\gamma^2}{15} \tag{32}$$

where N is the number of molecules in the scattering volume.

2.1.2 Depolarization Ratio

A very useful quantity for the investigation of randomly oriented molecules (gases, liquids, and amorphous solids) is the depolarization ratio ρ. The determination of ρ yields information about the symmetry of vibrations, which makes easier the assignment of experimentally observed Raman vibrations to normal modes of the molecule. The depolarization ratio ρ is calculated from the ratio of two intensities, which can be obtained from Raman measurements with different polarization of incident laser light or the scattered light. Analogous to the scattering geometry depicted in Figure 5, where the incident laser beam propagates in the z-direction and is polarized either in the x- or y-direction, the scattered light is observed at 90° in the y-direction. Depending on the position of a polarizer in front of the detector, the observed dipole moment is either x- or z-polarized. This notation can be replaced by using the terms parallel or perpendicular. For example $I_\parallel$ ($I_\perp$) indicates that the Raman intensity is measured using a polarizer in front of the detector which is oriented parallel (perpendicular) to the polarization of the incident laser light. The scattering intensity can be obtained by Equations (33) and (34):

$$I_{xx} := I_\parallel = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 N \frac{45\bar{\alpha}^2 + 4\gamma^2}{45} \tag{33}$$

$$I_{yx} := I_\perp = \frac{w^4}{16\pi^2\varepsilon_0^2 c^4} I_0 N \frac{\gamma^2}{15} \tag{34}$$

The depolarization ratio is then defined as Equation (35):

$$\rho = \frac{I_\perp}{I_\parallel} = \frac{3\gamma^2}{45\bar{\alpha}^2 + 4\gamma^2} \tag{35}$$

Since both components $\overline{\alpha}$ and γ can be equal to zero, the value of ρ can vary between 0 and 3/4. A Raman band with $\rho = 3/4$ is called depolarized and those with $0 \leq \rho < 3/4$ polarized. Only for totally symmetric vibrations, those which maintain all of the symmetry elements of the molecule, can $\overline{\alpha}$ be nonzero resulting in a depolarization ratio < 3/4. All other modes have $\rho = 3/4$. Thus, totally symmetric vibrations can be easily distinguished from those which are not.

2.1.3 Anti-Stokes versus Stokes–Raman Scattering

According to Figure 1 and Equation (10), a Raman transition can be either from vibrational/rotational ground

state i to the vibrational/rotational excited state f or vice versa. The first process is known as Stokes transition, whereas the second is the so-called anti-Stokes transition. Thus, a Raman spectrum reveals peaks on both sides of the Rayleigh line having identical frequency shifts. Even though the Raman shifts are identical, the observed intensities are quite different. Since the Stokes process corresponds to an energy transfer from an incident laser photon to a molecule and the anti-Stokes process reveals an energy transfer from a molecule to an incident laser photon, the ratio between the Stokes and anti-Stokes intensity provides information about the population probability of a certain energy state. The population probability is given by the Boltzmann distribution and the intensity ratio can be calculated as given in Equation (36):

$$\frac{I_{AS}}{I_S} = \left(\frac{w_L + w_S}{w_L - w_S}\right)^4 g_v \exp\left(\frac{-\hbar(w_L - w_S)}{kT}\right) \tag{36}$$

where I_S and I_{AS} are the intensities of the Stokes and the anti-Stokes field respectively and k is the Boltzmann constant. g_v is the degeneracy factor of the vibration (rotation). (It should be noted that, in a strict quantum mechanical calculation, where the field is also quantized, the exponent of the angular frequency ratio reduces from 4 to 3.) From Equation (36) we can deduce that the intensity ratio strongly depends on the spectroscopic temperature T, and therefore, this ratio can be used for the investigation of the temperature of samples, such as flames.

2.2 Short Quantum Mechanical Description of Raman Scattering

A quantum mechanical treatment of Raman spectroscopy is achieved by time-dependent perturbation theory.[27] More easy and for most cases valid is a semi-classical approach. The results of the semi-classical treatment will be presented in the following. Further, the selection rules will be revisited according to the quantum mechanical description as well as to group theory considerations.

In quantum mechanics the scattering system is not considered as a system consisting of atoms performing harmonic vibrations, but rather as vibrations which are quantized. The allowed energy level of each vibration is given by Equation (37)

$$E_v = \hbar w \left(v + \tfrac{1}{2}\right) \tag{37}$$

where w is the angular frequency of the vibration and v is the vibrational quantum number. A full approach of treating Raman scattering by quantum mechanics is obtained by time-dependent perturbation theory.[27] However, such a treatment is not very simple and most often not necessary. In a semi-classical approach the classical treatment of the dipole oscillation is combined with the quantum mechanical treatment of the scattering system and its changes during a vibration. Under certain conditions,[23] which are fulfilled for most of the molecules and crystals, the polarizability tensor can be replaced by its quantum mechanical analog. This can be concluded if the dipole moment of the transition $f \leftarrow i$ is considered compared with Equation (2), to give Equation (38):

$$\vec{\mu}_{fi} = \int_{-\infty}^{+\infty} \psi_f^* \vec{\mu} \psi_i \, d\tau = \langle \psi_f | \vec{\mu} | \psi_i \rangle \tag{38}$$

ψ_i and ψ_f denote the wave function of the initial and the final quantum state, respectively. In spontaneous Raman scattering the initial quantum state is for example a vibrational ground state, and the final quantum state is an excited vibrational state. The components of the polarizability tensor in Equations (18), (19), (21), and (22) can be replaced by the matrix elements given in Equation (39):

$$(\alpha_{\rho\sigma})_{fi} = \langle \psi_f | \alpha_{\rho\sigma} | \psi_i \rangle \tag{39}$$

Placzek[23] showed further that spatial averaging over squares of the matrix elements $\overline{((\alpha^2_{\rho\sigma})_{fi})}$ results in the same values as if averaging over the squares of the components of the polarizability tensor $\overline{(\alpha^2_{\rho\sigma})}$ is performed. Thus, the Equations (18), (19), (21), (22) and Equation (35) are still valid for this semi-classical approach.

2.2.1 *Selection Rules for Raman Scattering According to the Quantum Mechanical Description as well as to Group Theory Considerations*

In order to learn something about the selection rules in Raman spectroscopy we have to consider Equation (39) once more. A transition $f \leftarrow i$ is allowed if just one of the six matrix elements $(\alpha_{\rho\sigma})_{fi}$, with $\rho\sigma = xx$, yy, zz, xy, xz, yz, is nonzero. Thus, there are six integrals to be calculated in order to determine whether a matrix element is nonzero or not. However, group theory will provide the same information, but with less effort.[28] To get started on this issue, we first consider how symmetry operations of a certain point group affect the components of the polarizability tensor. In Figure 7 the used scattering geometry is depicted, which is essentially the same as in Figure 5. A water molecule has a fixed orientation in the center of the coordinate system. In this figure also all symmetry operations of the point group $mm2 = C_{2v}$ of water are displayed. Depending on the chosen polarization of both the incident light wave $\vec{E}$ and dipole moment $\vec{\mu}$ four components of the polarizability tensor α_{zx}, α_{zy}, α_{xx}, and α_{xy} (see Equation 17) contribute to the Raman scattering in this scattering geometry.

To evaluate, for example, the transformation behavior of the tensor component α_{zx} one needs to consider

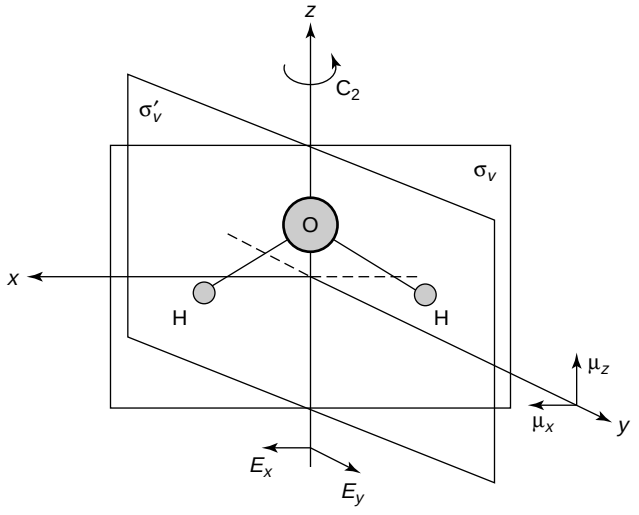

Figure 7 Illustration of the effects of symmetry operations of the point group *mm*2 (Hermann–Mauguin) or C_{2v} on the components of the polarizability tensor.

Table 1 Transformation behavior of the tensor components of the polarizability as a function of symmetry operations for a given point group (e.g. the tensor component α_{zx} transforms as the function zx)

$mm2 = C_{2v}$	E	$C_2(z)$	$\sigma_v(xz)$	$\sigma'_v(yz)$	Function
E_x	E_x	$-E_x$	E_x	$-E_x$	
μ_z	μ_z	μ_z	μ_z	μ_z	
Equation (17 → α_{zx})	α_{zx}	$-\alpha_{zx}$	α_{zx}	$-\alpha_{zx}$	
Cotton[29] B_1	1	−1	1	−1	xz

$mm2 = C_{2v}$: point group; E, $C_2(z)$, $\sigma_v(xz)$, $\sigma'_v(yz)$: symmetry elements; E_x: electric field component in x-direction; μ_z: dipole moment component in z-direction; B_1: irreducible representation; xz: function.

only the transformation of μ_z and E_x (see second equation of (17)). Applying the symmetry operations of the point group $mm2 = C_{2v}$ onto these terms, they will either remain unchanged or their sign will change. Table 1 represents this connection for the tensor component α_{zx}, which transforms as the function zx. Thus, it belongs to the irreducible representation of B_1. Similarly, this is also true for the other tensor components. In general, the components of the polarizability $\alpha_{\rho\sigma}$ transform as the product of their indices, which represent the coordinates $\rho\sigma$, i.e. as the functions xx, xy. This result holds for all point groups.

The selection rules are obtained from the mathematical result, that a certain integral in range between $-\infty$ and $+\infty$ is zero, if the integrand is an odd function. Thus, the integrals of Equation (39) are zero, if the term $\psi_f^* \alpha_{\rho\sigma} \psi_i$ represents an odd function. From the point of view of group theory a transition is observed in Raman scattering, if the product of these three terms belongs to the totally symmetric representation or is contained in it. In the following we restrict our consideration on the fundamental transition $f = 1 \leftarrow i = 0$ of a harmonic oscillator. Since the function ψ_0 is a totally symmetric one, the product of $\psi_f^* \alpha_{\rho\sigma}$ needs also to be totally symmetric. The latter will be fulfilled if $\alpha_{\rho\sigma}$ and the wave function of the first excited vibrational state ψ_f belong to the same symmetry class, that is, the application of all symmetry operations of the point group transforms in the same way.

Whether there are Raman active vibrations or not in a certain point group can be easily seen by considering the "character tables".[28] Consider, for example, the naphthalene molecule which belongs to the *mmm* point group (notation according to Hermann–Mauguin) or to the D_{2h} point group (notation according to Schönflies). Raman active vibrations are only expected from the irreducible representations A_g, B_{1g}, B_{2g}, and B_{3g}, because these transform as the functions x^2, y^2, $z^2(A_g)$, $xy(B_{1g})$, $xz(B_{2g})$, and $yz(B_{3g})$, respectively. Raman transitions in the irreducible representations of A_u, B_{1u}, B_{2u}, and B_{3u} are forbidden.[29]

2.3 Enhancing Raman Signals

The classical spontaneous Raman scattering is an appropriate analytical tool in order to investigate the qualitative as well as the quantitative composition of gases, liquids, or solids. However, there are serious drawbacks which have to be dealt with when using techniques based on spontaneous Raman scattering. First, the conversion efficiency of the Raman effect is poor. Only a portion of 10^{-6} to 10^{-8} laser photons are converted into Raman photons, limiting the detection of molecules with very low concentration. Second, certain molecular systems, such as fluorophore exhibit high fluorescence quantum yields and even weak fluorescence signals may be strong enough to mask the Raman signals.

Therefore, much scientific effort was invested in developing Raman based techniques, which allowed the above mentioned problems to be overcome. The next two chapters deal with two approaches, the resonance Raman effect (RRE) and SERS, with which the Raman intensity can be enhanced by orders of magnitude and/or fluorescence can be quenched.

2.3.1 Resonance Raman Effect

Until now, the energy of the excitation source was considered to be far away from an electronic transition, corresponding to major simplification of Raman theory.

The relevant expression for the Raman intensity in the resonance case has been derived by Kramers and Heisenberg[2] and Dirac[3] from second-order, time-dependent perturbation theory and further developed by Albrecht.[30] Since Raman scattering is a two-photon process, the first step is the virtual absorption of an

incident photon from the initial state i to the entire manifold of eigenstates of the unperturbed molecule, followed by the second step, the virtual emission to the final state f. The components of the transition polarizability tensor are calculated by a summation over all unperturbed eigenstates of the resonant electronic state,[2] according to Equation (40):

$$(\alpha_{\rho\sigma})_{fi} = \sum_r \left\{ \frac{\langle f|\mu_\rho|r\rangle\langle r|\mu_\sigma|i\rangle}{\hbar w_{ri} - \hbar w_L - i\Gamma_r} + \frac{\langle f|\mu_\sigma|r\rangle\langle r|\mu_\rho|i\rangle}{\hbar w_{rf} + \hbar w_L - i\Gamma_r} \right\} \tag{40}$$

where $\langle f|\mu_\rho|r\rangle$ corresponds to the ρth component of the transition dipole moment associated with the transition $f \leftarrow r$, μ_ρ is the dipole moment operator in the ρ direction. The suffixes ρ and σ refer to the molecule-fixed Cartesian vector x, y, and z. w_L and w_{ri} are the angular frequencies of the incident radiation and the angular frequency associated with the transition from the initial state i of the Raman transition to any state r of the unperturbed molecule, respectively. Γ_r is the damping factor related to the lifetime of the state r. Note, that the summation is over all states r of the unperturbed system, with the exclusion of i and f, the states between which the Raman transition takes place.

The scattered RR intensity can be calculated with a semi-classical approach similar to section 2.2. The total intensity scattered on the average into a solid angle of 4π by one freely orientable molecule is given by Equation (41)[31]

$$I_{fi} = CI_0(w_L \pm w_{fi})^4 \sum_{\rho,\sigma} |(\alpha_{\rho\sigma})_{fi}|^2 \tag{41}$$

Here, w_L and w_{fi} are the angular frequency of the incident laser radiation and the angular frequency for Raman transition from state i to state f, respectively. I_0 denotes the incident laser intensity and C is constant which depends on the unit system used.

Two types of RREs can be distinguished: the pre-RRE and the rigorous RRE. In Figure 8, the conventional Raman effect (a), the pre-RRE (b), and the rigorous RRE (c) are depicted. Far away from an electronic absorption, all terms in the denominator and the numerator of the Raman intensity expression such as $w_{el} - w_L$ which depend on the electronic state with the angular frequency w_{el} are negligible (see Equation 40). However, when the photon energy of the exciting laser beam approaches the energy of an electronic dipole transition those terms dominate the observed Raman intensity and thus, the obtained Raman spectrum. In general, the closer the excitation energy approaches the electronic resonance condition, the greater is the intensity of the observed Raman spectrum. Typically, the pre-RRE is observed when the exciting laser angular frequency comes within the high or low angular frequency wing but not under the observable vibrational structure of the electronic absorption band in the Raman scattering process. When the exciting laser angular frequency actually falls within the observable vibrational structure, then the rigorous RRE is observed. In the case that the angular frequency is getting close to an electronic transition, the first term of Equation (40) begins to dominate, and the polarizability as well as the Raman intensity will be dominated by the resonant excited state.

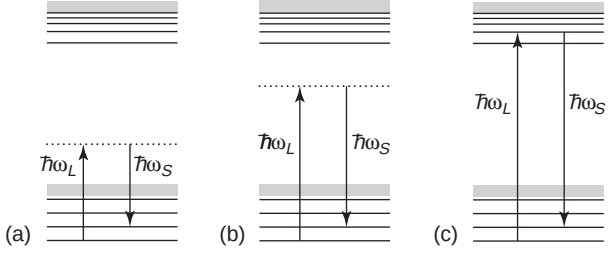

Figure 8 Energy diagram for the conventional Raman effect (a); the pre-RRE (b); and the rigorous RRE (c).

RRS is of great interest in the case of large polyatomic molecules, such as biological molecules, where the absorption is often localized in a particular group of a large molecule, which is the so-called chromophore. The RR enhancement permits the isolation of the Raman bands of a chromophore and structural units located close to it, and therefore gives information about sites of a molecule, which makes an assignment of complicated molecules possible.

Apart from the Raman intensities, the selection rules are also affected under resonance condition. Certain transitions normally forbidden in conventional Raman scattering are allowed in resonance Raman scattering. Those changes provide further information about the molecular system under investigation.

Since the angular frequencies of Raman vibrations depend on the electronic ground state, whereas the scattered Raman intensity is dependent on the absolute location of the incident laser energy to the electronic absorption, information about properties of a molecule in an electronic excited state can be obtained from RRS. The measurement of the scattered Raman intensity as a function of the incident laser energy produces the so-called excitation profile (EP), from which the strength of the interaction between the electronic excited state and the vibrational mode can be estimated.

The change in the selection rules as well as an enhancement of the observed Raman intensity by up to six orders of magnitude give rise to various applications of both pre-RRS and rigorous RRS in chemistry, biochemistry, and biology as well as in physics. For utilizing the RRE no special equipment other than that used for conventional Raman spectroscopy is necessary.

However, it may be suitable to use a tunable laser as excitation source, which allows an appropriate adjustment of the excitation energy lying in the electronic absorption. For example, with the excitation in the UV the minimal detectable concentration of certain species can be lowered by orders of magnitude of five to six using the RRE. For more information on RRS we refer to several reviews.[32]

2.3.2 *Surface-enhanced Raman Scattering*

The second effect which gives rise to an enhancement of Raman signals by up to six orders of magnitude or even more was discovered in the mid-1970s, and is referred to as surface-enhanced Raman scattering. Fleischmann et al.[33] discovered that molecules adsorbed on an electrochemically roughened silver surface give dramatically enhanced Raman signals. Recently, other metals, such as copper or gold, also showed enhanced Raman spectra under certain conditions. Efforts are made for even single molecule detection by means of SERS.[34–36]

Until now, there is no complete theoretical understanding of this surface enhancement. Different groups contributed to the surface enhancement problem.[37–39] In the following, only a brief discussion about the possible enhancement mechanisms will be given. From normal Raman scattering it is well known (see Equation 2), that the transition dipole moment $\vec{\mu}$, which a molecule experiences, depends on the transition polarizability, α, and the strength of the electric field, $\vec{E}$. Therefore, molecules located close to a silver surface can experience either an enhanced electric field or a modified polarizability. According to this distinction the surface theory is divided into two classes: the ''electromagnetic field'' enhancement mechanism, in which the field experienced by the adsorbate molecule is larger than it would be far from the surface, and the ''molecular'' enhancement mechanism, in which α is perturbed by interactions of the adsorbate with the surface.

Classical electromagnetic theory based on the solution of Maxwell's equations indicates that molecules adsorbed to a metal surface will experience a large increase in the electromagnetic field strength compared to the strength of the incident radiation. The extent of the electromagnetic enhancement depends on a number of factors, including the electric properties of the metal, the distance of the molecule from the surface, the orientation of the molecule with respect to the normal of the surface, the energy of the incident radiation, the morphology of the surface, and the size and the geometry of surface roughness. Of particular importance is the surface roughness or curvature, which can be obtained electrochemically or by the use of sols or island films. Calculations showed that the interaction of a light field with a flat (smooth) metal surface can enhance the Raman intensity scattered by an adsorbed molecule by a factor of six.[39] An intensity enhancement of 10^4 to 10^7 can be obtained using roughened metal surfaces, such as small metal particles or rough surfaces of conductive materials. On irradiating rough surfaces, which are usually approximated in theoretical models[40] by isolated spheres or spheroids, with light, the incident electromagnetic field can excite an oscillation of conduction electrons inside the metal. Due to this dipole oscillation the particles become polarized, and start emitting a secondary electric field. Thus, the field at the particle surface is the superposition of the incident field and the generated secondary fields. Since the conduction electrons are spatially confined, they have a characteristic frequency known as the plasma frequency. The quantized state of these plasma oscillations is called a plasmon. If the frequency of the incident field is resonant to such a plasma resonance of the metal, the generated secondary fields become significantly large, and thus a molecule adsorbed on the surface will experience a large enhanced electromagnetic field. The strength of the field enhancement falls off as $1/r^3$ away from the surface and strongly depends on the extent of the curvature of the particle. The frequency of a plasma resonance depends on the metal, on the particle size as well as on the particle shape. Only roughened or particular metals such as silver, copper, or gold can be used as substrate in SERS technology, because only their plasma frequencies are lying in the VIS range or the near IR range of the electromagnetic spectrum.

The transition polarizability α can also be altered as a result of electronic interaction between the adsorbate and the metal. Such an enhancement can be obtained by a direct charge transfer between molecule and metal, excitation of electron–hole pairs requiring atomic scale roughness, or the formation of a chemical bond between the adsorbed molecule and the metal, and are limited to molecules in the first monolayer, being in direct contact with the metal surface. Thus, mechanisms which are modifying α, are the so-called short-range enhancements.

The selection rules for SERS are essentially the same as those for conventional Raman spectroscopy. However, because the local field at the surface is highest in the direction normal to the surface, vibrational modes involving changes in the adsorbate polarizability perpendicular to the surface are preferentially enhanced. A detailed summary of the selection rules for SERS has been given by Creighton et al.[38] The normal direction of the electric field plus the fact that the electromagnetic amplitude falls off rapidly with the distance from the surface allow one to determine the adsorbate orientation with respect to the average surface normal as well as the proximity of the adsorbate functional groups to the surface.

In general a SERS setup is essentially the same one as used for conventional Raman spectroscopy. However, to optimize the electromagnetic surface enhancement effect the used laser frequency has to match the frequency of a plasma resonance. Since the enhancement effect strongly depends on the physical properties of the substrate, a large variety of different types of substrates are presented in the literature. The most common substrates are electrodes, sols, metal films, and silver metal island films. Because SERS provides both rich spectroscopic information and high sensitivity as a result of the large enhancement effect, it is an ideal tool for trace analysis as well as for in situ investigations of various interfacial processes. The method has been applied for the study of a variety of molecules from diatomic ones to molecules of biological importance. Various reviews on SERS have been written to which we refer the interested reader.[32]

3 FUNDAMENTALS OF NONLINEAR RAMAN SPECTROSCOPY

The introduction of lasers, especially of pulsed laser systems (e.g. Nd : YAG (neodymium : yttrium aluminum garnet) laser), has given rise to the development of several new nonlinear optical effects. In 1961 Franken et al.[41] demonstrated the possibility of second harmonic generation, and Kaiser and Garrett[42] observed the two photon absorption. The coupling between strong laser fields and molecular vibrations permit the observation of nonlinear Raman effects such as hyper-Raman scattering,[27] SRS,[8,9] CARS,[27] the Raman gain spectroscopy,[10] the inverse Raman effect,[11] the Raman-induced Kerr effect (RIKE),[12] and the PARS.[13]

The interaction of strong incident laser fields with molecular vibrations can result in an elongation of the electron shell which does not obey Hooke's law any more. Therefore, the induced polarization consists of a linear and a nonlinear contribution. Generally the induced dipole moment $\vec{p}$ in a molecular system is written in the form given in Equation (42):

$$\vec{p} = \alpha \overrightarrow{E} + \frac{1}{2}\beta \overrightarrow{E}\,\overrightarrow{E} + \frac{1}{6}\gamma \overrightarrow{E}\,\overrightarrow{E}\,\overrightarrow{E} + \cdots \qquad (42)$$

where α is the linear polarizability, β the hyperpolarizability, and γ the second order polarizability. The α, β, γ etc. are tensors of rank 2, 3, 4, etc, respectively. $\overrightarrow{E}$ is the electric field, $\overrightarrow{E}\,\overrightarrow{E}$ is the dyadic product of two electric fields, etc. The nonlinear terms of Equation (42) are usually small compared to the linear term, which gives rise to normal, linear Raman scattering. However, when the electric field is sufficiently large, as is the case when for instance a Q-switched laser is focused on the sample, contributions from the second term in Equation (42) are sufficiently intense to be detected.

3.1 Spontaneous Nonlinear Raman Scattering: Hyper-Raman Effect

The importance of the hyper-Raman effect, which stems from the second order term of Equation (42), as a spectroscopic tool results from the fact that it allows the observation of modes, which are accessible neither by IR nor by Raman spectroscopy. Since its discovery by Terhune et al.[14] in 1965, hyper-Raman spectra have been observed in all three states of aggregate. However, reasonable signal-to-noise ratios could only be obtained after the development of high repetitive, high-power lasers and highly sensitive detectors (multichannel diode arrays or CCDs).

The hyper-Raman effect is a three-photon process involving two virtual states of the scattering system. The level scheme for hyper-Rayleigh and hyper-Raman scattering is presented in Figure 9. Two laser photons of the angular frequency w_L are simultaneously scattered to give a photon at angular frequency $2w_L - w_R$ (Stokes hyper-Raman) or $2w_L + w_R$ (anti-Stokes hyper-Raman) when a molecular vibration (or rotation) is excited or destroyed, respectively. Thus, hyper-Raman lines are found in the spectral neighborhood of $2w_L$ (the second harmonic) of the incident laser radiation. It should be mentioned that Rayleigh scattering always occurs, but hyper-Rayleigh scattering occurs only if the scattering material does not have a centre of symmetry. Frequency-doubled radiation consists of hyper-Rayleigh scattering from a pure crystal and consequently a crystal used for frequency doubling must not have a centrosymmetric unit cell.

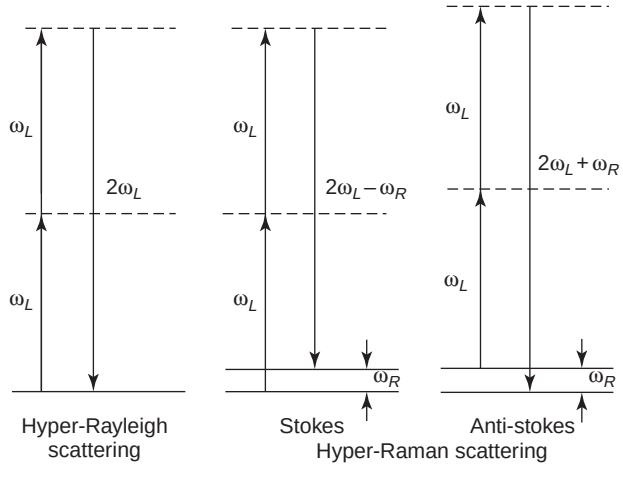

Figure 9 Schematic level diagram for hyper-Rayleigh and hyper-Raman scattering.

The importance of the hyper-Raman effect as a spectroscopic tool results from its selection rules. These are determined by products of three dipole moment matrix elements relating the four levels indicated in Figure 9. It turns out that all IR active modes of the scattering system are also hyper-Raman active. In addition, the hyper-Raman effect allows the observation of ''silent'' modes, which are accessible neither by IR nor by linear Raman spectroscopy. Hyper-Raman spectra have been observed for the gaseous, liquid and solid state. A full description of the theory and practice of hyper-Raman spectroscopy is given by Long.[15,27]

3.2 Stimulated Raman Scattering

At high laser intensities the Raman scattering can become so intense that the Stokes radiation can also influence the molecules. Thus, the molecules interact with two fields: the laser field at angular frequency w_L and the Stokes field at the angular frequency at $w_S = w_L - w_R$. Both fields couple with each other at the angular frequency w_R of the vibrating molecule. According to this parametric coupling, laser energy can efficiently be transferred into the Stokes–Raman field and/or the anti-Stokes–Raman field resulting in the development of an intense SRS signal. Above the threshold about 20 to 30% of the incident laser radiation can be converted into SRS. The major disadvantage of SRS is that only the Raman modes with the highest gain can reach the threshold, and therefore, an SRS spectrum yields less molecular information than spontaneous Raman scattering does. However, this disadvantage can be turned into an advantage by using SRS as a monochromatic, coherent light source for special wavelength ranges.

Both linear Raman scattering and hyper-Raman scattering are spontaneous inelastic light scattering processes. The incident radiation excites with a certain probability one possible molecular vibration in one molecule. Since the vibrational excitation is totally stochastic, common Raman or hyper-Raman spectra are the result of the superposition of stochastic totally independent vibrations from individual molecules.

To obtain coherent molecular vibrations within a certain macroscopic area (e.g. inside the focus area of the laser beam), meaning those molecules are vibrating with a constant phase relation, an intense laser radiation has to be used. Such a process is called SRS. In general, the Raman effect is a very weak process. Its intensity can be sufficiently enhanced by both the RRE and the SERS, as shown above. However, there is another effective way to enhance the strength of Raman signals, which is schematically depicted in Figure 10 and explained as follows. Spontaneous Raman emission depends in a linear way on the input laser intensity. Thus, an intense input laser field will lead to an intense Stokes field (nw_S, where n is the number of Stokes photons), which can stimulate the emission of further Stokes photons, if it is incident on a molecule simultaneously with a laser field at angular frequency w_L. While the incident light beam loses a quantum ($\hbar w_L$) and the molecule is excited by a quantum ($\hbar w_R = \hbar(w_L - w_S)$, a quantum ($\hbar w_S$) is added to the wave at frequency w_S, which consequently becomes amplified. Thus, SRS is a self-amplifying process which is observed in a forward and backward direction of the incident laser pulse. In contrast to spontaneous Raman emission, the conversion efficiency can be about 20 to 30% of the incident radiation. In the case that the generated SRS wave, which is also called first-order SRS, at $w_{1S} = w_L - w_R$ is intense enough and can serve as pump source for another first-order SRS process, multimode SRS at $w_{1S} = w_L - \ell w_R$ is observed. This process is illustrated in Figure 11.

A theoretical description of the stimulated Raman effect in terms of nonlinear Raman susceptibility was presented by Bloembergen[16] soon after the experimental discovery of this effect. He showed that a polarization at Stokes angular frequency w_S is generated via the third-order nonlinear susceptibility $\chi^{(3)}$. Including a degeneracy

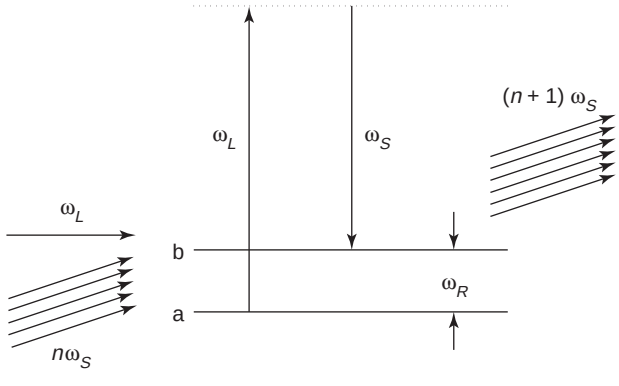

Figure 10 Schematic level diagram for SRS as a quantum process.

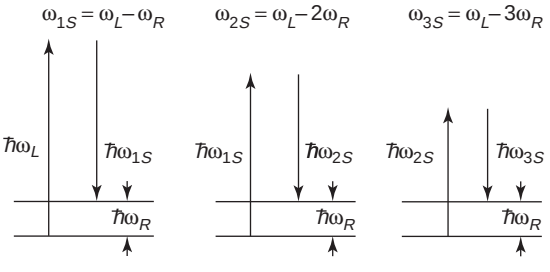

Figure 11 Multimode SRS: first-order SRS (w_{1S}) is pumped by the laser (w_L); second-order SRS (w_{2S}) is pumped by the laser like first-order SRS radiation; third-order SRS (w_{3S}) is pumped by the laser like second-order SRS radiation.

factor, the polarization oscillating at angular frequency w_S is given by Berger et al.[17] (Equation 43):

$$\vec{P}^{(3)}(w_S) = \frac{6}{4}\varepsilon_0 \chi^{(3)}(-w_S; w_L, -w_L, w_S)|\vec{E}(w_L)|^2 \vec{E}(w_S) \tag{43}$$

For the case that the pump angular frequency w_L and the Stokes angular frequency w_S differ by a molecular angular eigenfrequency w_R, i.e. $w_L - w_S = w_R$, the third order nonlinear susceptibility has a negative imaginary part. This leads to a negative absorption coefficient or an exponential gain at the Stokes frequency w_S proportional to laser intensity at pump frequency[19] w_L.

The exponential growth has been shown to depend on the differential Raman cross-section $d\sigma/d\Omega$ (Ω = solid angle) as well as on the linewidth Γ of the molecular transition (w_R) in Equation (44):[18]

$$g(w_S) \propto \left(\frac{d\sigma}{d\Omega} \quad \frac{\Gamma}{(w_L - w_S - w_R)^2 + \Gamma^2}\right. \tag{44}$$

Optimum gain is found at the center of the Raman line where $w_R = w_L - w_S$. There the gain constant for SRS at Stokes frequency is given by Equation (45)

$$g_s \propto \left(\frac{d\sigma}{d\Omega} \quad \frac{1}{\Gamma}\right. \tag{45}$$

From Equation (45) we can immediately recognize that in SRS processes where only one input laser field with frequency w_L is employed, a coherent Stokes wave is generated for those Raman modes which have highest ratio between differential Raman cross-section and linewidth Γ. The latter corresponds to the dephasing time T_2 of the physical system, $\Gamma = 1/T_2$, and reflects the damping of the system. Because of this limitation to modes having the highest ratio between differential Raman cross-section and linewidth and an exponential dependency of the Stokes intensity on the incident laser intensity, SRS as an analytical tool for detailed qualified and quantified information from a sample in an optical cell is not as suitable as spontaneous Raman scattering. However, this disadvantage can be turned into an advantage by using SRS as a monochromatic, coherent light source for special wavelength ranges.

In the classical description of SRS a coherent light beam at Stokes angular frequency w_S is produced by the pump field at $w_L = w_S + w_R$. The distinctive feature of SRS is that an assemblage of coherently driven molecular vibrations provide the means of coupling the two light waves at angular frequency w_L and w_S by modulating the nonlinear susceptibility.

In ordinary Raman scattering the Stokes scattering is generally accompanied by anti-Stokes scattering. Since only a few molecules are found in the vibrational excited state at room temperature, the anti-Stokes–Raman intensity is weak. The strong pumping action of the laser beam changes this situation drastically, so that an appreciable fraction of all molecules in the laser beam are soon available for the anti-Stokes emission. Classically, the anti-Stokes radiation is generated by the interaction of the laser beam with molecular vibration, but the phase of the latter is established by the still more intense Stokes radiation. This is a four-wave mixing process, which is called stimulated anti-Stokes–Raman scattering. Since a pair of laser photons is converted into a Stokes photon and an anti-Stokes photon, no momentum is transferred to the sample. The initial and the final state of this process are identical, therefore, the next two laser photons can be converted in the same molecule without any relaxation time. Thus, the conversion efficiency is high.

As a consequence, an index-matching requirement exists for this four-wave process. The direction of the emission of the anti-Stokes intensity is given by the momentum matching condition,[18] Equation (46)

$$\vec{k}_{AS} = 2\vec{k}_L - \vec{k}_S \tag{46}$$

Since, in condensed phases, there is dispersion of the linear index of radiation, $n = n(w)$ and because Equation (47) holds,

$$|\vec{k}| = \frac{n(w)w}{c} \tag{47}$$

one immediately recognizes that a collinear arrangement between pump, Stokes and anti-Stokes beams is not possible. However, a direction can be found which satisfies the vector Equation (46), but it confines the S and AS photon pair to conical beams coaxial with the laser beam. Figure 12 shows the required construction, the cones being generated by rotating the plane of the figure around $\vec{k}_L$.

The interaction of the electric fields of two input laser beams with angular frequencies w_L and w_S with the electric field associated with coherent molecular vibrations can yield also a gain or a loss in the power of the laser beams. The method where the gain at Stokes frequency is measured is generally referred to as SRGS, whereas the inverse Raman scattering (IRS)

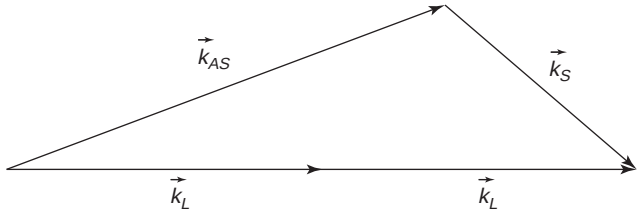

Figure 12 Momentum conservation for stimulated anti-Stokes Raman scattering (representation of Equation 46).

is the terminology used to designate the induced loss at the pump laser frequency. IRS is also often called stimulated Raman loss spectroscopy (SRLS). In Figure 13 a schematic diagram for a few techniques in nonlinear (coherent) Raman spectroscopy is displayed.

Esherick and Owyoung[43] have developed a double resonance method called IDSRS. This technique combines the high sensitivity of resonant laser ionization methods with the advantages of stimulated Raman spectroscopy, e.g. the high spectral resolution. The excitation process, illustrated in Figure 14, can be briefly described as a two-step photoexcitation process followed by ion/electron detection. In the first step two intense narrow-band lasers (w_L, w_P) are used to pump a molecule from its ground state, (a), to a vibrationally excited level,

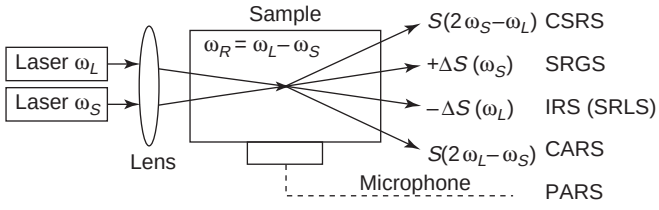

Figure 13 Schematic diagram for a few techniques in nonlinear (coherent) Raman spectroscopy; CSRS, coherent Stokes–Raman scattering; SRGS; IRS; (= SRLS); CARS; PARS.

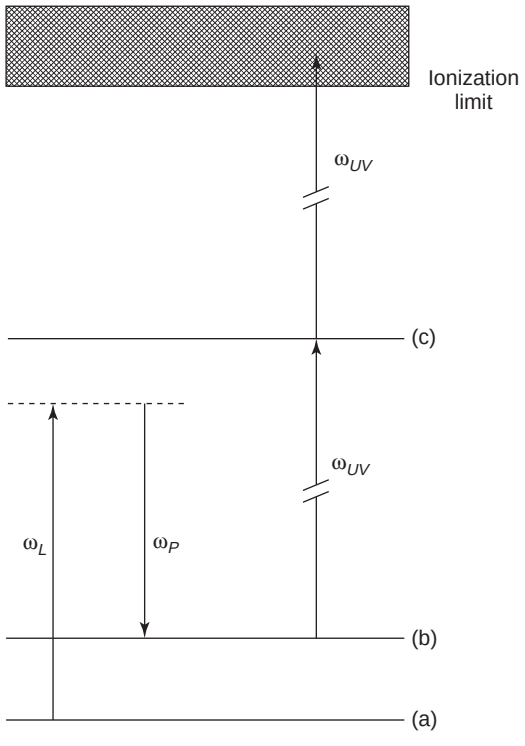

Figure 14 Energy-level diagram illustrating the two excitation steps of IDSRS.

(b), via the stimulated Raman process. The vibrationally excited molecules are then selectively ionized in a second step via a two-photon process by a laser tuned to be resonant with the electronic state, (c), in the molecule. The technique allows an increase in sensitivity of over three orders of magnitude because ions can be detected with higher sensitivity than photons. This technique has become very important when femtosecond lasers are used (see section 3.4). Of all nonlinear Raman spectroscopic techniques CARS is the most often applied. Although a lot of experimental effort is necessary to utilize CARS, the advantages, especially the high CARS intensities, the high spectral and spatial resolution, and the capability of rejecting fluorescence makes CARS a very valuable spectroscopic tool.

The applied laser power of pump and Stokes laser in a typical CARS experiment is in the range 10^4–10^5 W; the produced CARS signal output power can be up to 1 W, whereas the output power in conventional Raman[20] spectroscopy is in the range of 10^{-4} W. In addition the CARS signal is directional with a small solid angle while spontaneous Raman photons are emitted at a solid angle of 4π. The collection angle range is up to five orders of magnitude larger in conventional Raman spectroscopy. Taken together, these two factors imply that CARS is nine orders of magnitude less sensitive to sample fluorescence than spontaneous scattering. The advantage is actually even greater since the CARS signal is at higher frequency than any of the input laser frequencies.

3.3 Coherent Anti-Stokes–Raman Scattering

As shown in the section before a high power laser pulse interacting with matter can excite coherently driven molecular vibrations in conjunction with the generation of an intense coherent Stokes beam. However, only signals from a Raman active mode having the highest gain factor will give rise to stimulated Raman emission. Thus, for molecular spectroscopy with which all Raman active modes should be investigated in order to obtain information about the chemical and structural composition of the sample molecules, excitation with one single strong laser field would certainly not serve the purpose, because only from one particular mode an intense signal will be observed.

This drawback can be avoided when an intense Stokes field is applied externally to the sample by a second laser. In a CARS experiment one laser operates at angular frequency w_L and the second laser at the angular frequency w_S. If the frequency difference between both laser fields coincides with the frequency of a molecular vibration $w_R = w_L - w_S$, a strong coupling of the incident laser field, the Stokes field, and the field generated by the coherently driven molecular vibration via the third order

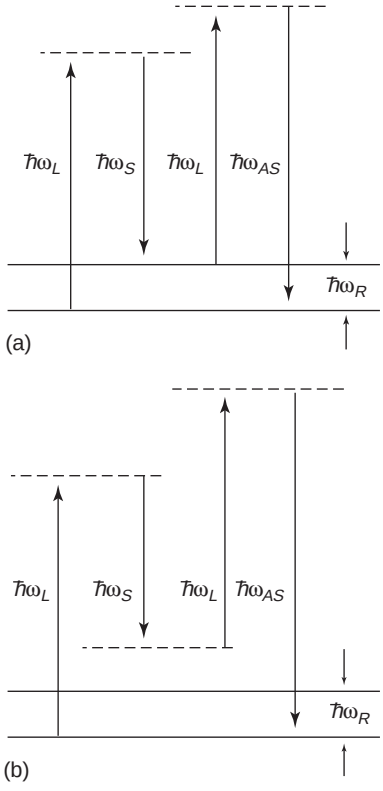

Figure 15 Schematic level diagram for CARS: (a) resonant; (b) non-resonant interaction of the four waves.

nonlinear susceptibility is observed. However, even if the frequency difference does not match a molecular vibration this CARS process can occur, albeit with reduced intensity. The energies for both four-wave mixing processes are illustrated in Figure 15, (a) displays the resonant CARS process, whereas (b) corresponds to the non-resonant CARS process, which always occurs if $w_L - w_S \neq w_R$. Although the resonant CARS process dominates the CARS spectrum it is usually seen against a background emission due to the parametric coupling of the nonresonant process.

According to Figure 15(a) in the first step a vibrational quantum ($\hbar w_R$) is generated by the two quanta ($\hbar w_L$) and ($\hbar w_S$), as in Equation (48)

$$+\hbar w_R = \hbar w_L - \hbar w_S \tag{48}$$

In the second step the vibrational quantum ($\hbar w_R$) is destroyed again by the two quanta ($\hbar w_L$) and ($\hbar w_{AS}$), given by Equation (49)

$$-\hbar w_R = \hbar w_L - \hbar w_{AS} \tag{49}$$

The energy conservation of this process is given by Equation (50):

$$\hbar w_{AS} = 2\hbar w_L - \hbar w_S \tag{50}$$

As shown in section 3.2 using Equations (46) and (47) an anti-Stokes signal w_{AS} is generated only, when the momentum matching condition is fulfilled. In the case of gases, where the dispersion of the index of refraction $n = n(w)$ is negligible, collinear arrangement between pump and Stokes laser beam is sufficient. In condensed phases, however, a slight angle between pump and Stokes beam is required. In contrast to SRS, where the coherent signal is generated with only one input laser beam w_L, the anti-Stokes intensity is not scattered into a cone as described above. Instead, it is scattered in one direction, which lies in the plane given by the two laser beam directions $\vec{k}_L$ and $\vec{k}_S$ and which is determined by the momentum vector diagram shown in Figure 12. A common CARS experiment can be performed by keeping one of the two lasers, e.g. the laser with the angular frequency w_L fixed in frequency and the other one, e.g. the laser at Stokes angular frequency w_S tunable. The coherent anti-Stokes signals are generated each time the frequency difference of the input laser fields matches the molecular frequency of a Raman active transition. Since the CARS signal occurs blue shifted compared to the Rayleigh line, excited fluorescence can not mask the CARS spectrum because of its red shift.

The intensity of generated anti-Stokes radiation can be calculated according to Equation (51)[9,21]

$$I_{AS} = \frac{w_{AS}^2}{n_L^2 n_S n_{AS} \varepsilon_0^2 c^4} \left| \chi_{CARS}^{(3)} \right|^2 I_L^2 I_S L^2 \mathrm{sinc}^2 \left(\frac{\Delta k L}{2} \right. \tag{51}$$

where n_i is the refractive index at the angular frequency w_i (i = L, S, AS), I_L, and I_S are the intensities of the incident pump and Stokes laser beams, respectively, L is the length over which the beams are mixed through the sample (interaction length), and Δk is the wave vector mismatch, i.e. $\Delta k = |2\vec{k}_L - \vec{k}_S - \vec{k}_{AS}|$. Δk is a direct result of the fact that the propagating waves move in and out of phase because of dispersion in the sample. The dependence of I_{AS} on the square of the CARS susceptibility is characteristic for parametric processes. $\chi_{CARS}^{(3)}$ can be directly related to the differential Raman cross-section $\mathrm{d}\sigma/\mathrm{d}\Omega$ as well as to the number density[22] N. We finally note that the phase matching factor in the CARS intensity, i.e. the $\mathrm{sinc}(x) = \sin(x)/x$ function is essentially unity under most experimental conditions, which are usually chosen such that $\Delta k = 0$ for maximum signal.

In order to obtain a large angular separation of the generated CARS beam from the input laser beams, Eckbreth[44] introduced a new beam arrangement, the so-called BOXCARS geometry which still obeys the phase-matching condition. In Figure 16 the phase matching is displayed. In this approach the w_L pump beam is split

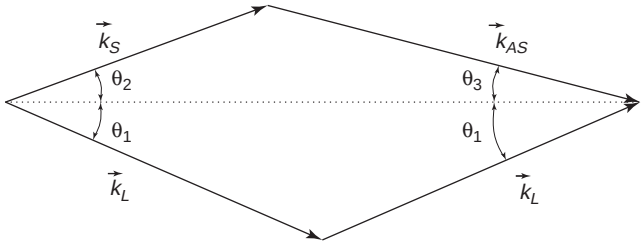

Figure 16 Phase matching for BOXCARS.

into two nonparallel components, which together with the Stokes beam are focused under a larger angle θ_i into the sample. According to Figure 16, the geometry formed by the wave vectors $\vec{k}_L$, $\vec{k}_S$, and $\vec{k}_{AS}$ correspond to a ''box''. From this the name BOXCARS is derived. Note, for the BOXCARS arrangement the laser beams need not reside in the same plane. The phase-matching can also be fullfiled when the two pump beams w_L lie in one plane, and the w_S beam lies in another one (see Figure 17). This type of beam arrangement is known as Folded-BOXCARS[45] and reveals the great advantage that the anti-Stokes beam is completely spatially separated from the incident wave mixing components. However, the disadvantage of this beam arrangement is the small spatial overlap of the laser beams.

In the following we discuss the major advantages but also the limits of CARS. First, in contrast to linear Raman spectroscopy, in a usual CARS experiment the intensity of the CARS signal is four to five orders of magnitude more intense. Thus, in CARS spectroscopy there is in general no detection problem caused by too little intensity if the number density N of the molecules is high enough. Second, the CARS signal occurs as an intense signal beam, which can be discriminated from light resulting from other processes (e.g. fluorescence) even with collinear beam arrangement using filters and a large distance between sample and detector, because the

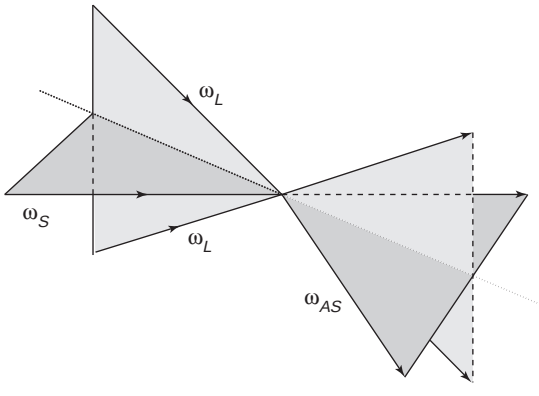

Figure 17 Geometry of laser beams for folded-BOXCARS.

intensity of fluorescence decreases with $1/r^2$. If condensed material is under investigation, the CARS beam can even be separated spatially (as shown above). Third, the CARS process is performed by lasers, which can spatially be focused onto a small region of some micrometer in radius. Thus, the signal is produced only in a small area of the sample which has two consequences: (1) CARS spectroscopy exerts high spatial resolution, and (2) small sample quantities in the range of some micrograms can be investigated easily. Fourth, CARS spectroscopy reveals a high spectral resolution, because, in principle, no spectrometer is necessary in front of the detector. Thus, the spectral resolution is limited by the linewidth of the used laser systems. Fifth, CARS is used for low concentration determination. In the case of non-resonant CARS, meaning that none of the energies of the two lasers hit an electronic resonance, the detection limit is in the range of 0.01 to 0.05 M. However, when the excited virtual states coincide with electronic eigenstates of the molecule, a resonance enhancement of the CARS signal is found. Such a process is referred to as resonance CARS and concentrations in the range 10^{-5}–10^{-7} M can be detected. Sixth, fluorescence of the sample can not mask the CARS signal.

However, CARS can only be performed with high experimental effort. In order to carry out a CARS experiment, it is necessary to avoid any signal fluctuation caused by intensity and frequency instabilities of the lasers, or due to changes of the spatial overlap of excitation laser beams.

Until now, the four-wave mixing process was performed using the three frequencies w_L, w_L, and w_S, which leads to the generation of a fourth wave with the angular frequency w_{AS}. Analogous to the above mentioned coupling the three frequencies w_S, w_S, and w_L also can mix, generating a new wave at the angular frequency, according to Equation (52)

$$2w_S - w_L = w_S - (w_L - w_S) = w_S - w_R \qquad (52)$$

The signal is produced with a shift of w_R on the Stokes side of the w_S line. This type of scattering process is referred to as CSRS. This technique tends to overlap with fluorescence, and furthermore, the CSRS signal is, in principle, weaker than CARS signals. Thus, CARS is used more frequently.

3.4 Femtosecond Time-resolved Nonlinear Raman Scattering

Ultrafast molecular vibrations and rotations are the fundamental motions that characterize chemical bonding and determination of the reaction dynamics at the molecular level. The time scale for these motions are typically 100 fs for vibrations and 100 ps for rotations.

Recent advances in time–domain spectroscopy have enabled spectroscopists to capture chemical processes directly on the time scale of the fastest elementary steps.[46,47]

The investigation of the time-resolved propagation of chemical reactions experienced a tremendous progress within the last two decades. Molecular vibrations and rotations are the fundamental movements in describing chemical bonding and molecular reaction dynamics. The dynamic of a chemical bonding of a molecular system is on a time scale from some femtoseconds to several picoseconds. The development of femtosecond laser systems allows the investigation of physical and chemical processes in real time, revealing deep insight into dynamics of chemical reactions. The fundamental concept of femtosecond laser spectroscopy is based on the generation of molecular wave packets and its observation in real time. Since the pulse time is so short one can coherently excite many vibrational modes at a time. The superposition of these stationary eigenstates is known as a wave packet. The time-dependent development of such a wave packet can be investigated by probing the system with another time-delayed ultra-short laser pulse.[46,47]

Ultrashort laser pulses with a duration of less than about 100 fs allow for a coherent broadband excitation with preparation and detection of vibrational superposition states. The time evolution of these generated wave packets in the ground and the electronically excited state give information on the molecular dynamics. There are different possibilities for studying the molecular dynamics in real time. Typically, the linear response of the optical field in a two-pulse pump-probe scheme is used in order to initiate and probe the wave packets in a certain quantum state.[46–48] In such experiments, the signal detection is performed by laser-induced fluorescence (LIF),[49–51] absorption,[52,53] resonant impulsive stimulated Raman scattering (RISRS),[54] multiphoton ionization (MPI),[55] and by several other methods such as, for example, femtosecond up conversion.[56]

The high intensity of femtosecond laser pulses enables the application of nonlinear spectroscopic techniques such as CARS or CSRS for the investigation of molecular systems. The first results of fs time-resolved CARS of liquids were presented by Leonhard et al.[57] in 1987. The principle of this technique is depicted in Figure 18 using iodine vapor as an example of gaseous matter. Three laser fields, two at angular frequency w_L (pump laser) and one at angular frequency w_S (Stokes laser), are interacting with the molecules. The pump laser has to overlap with vibrational and rotational eigenstates of an electronic excited state of the molecule and the Stokes wave has to be chosen such that the difference of the w_L and w_S matches

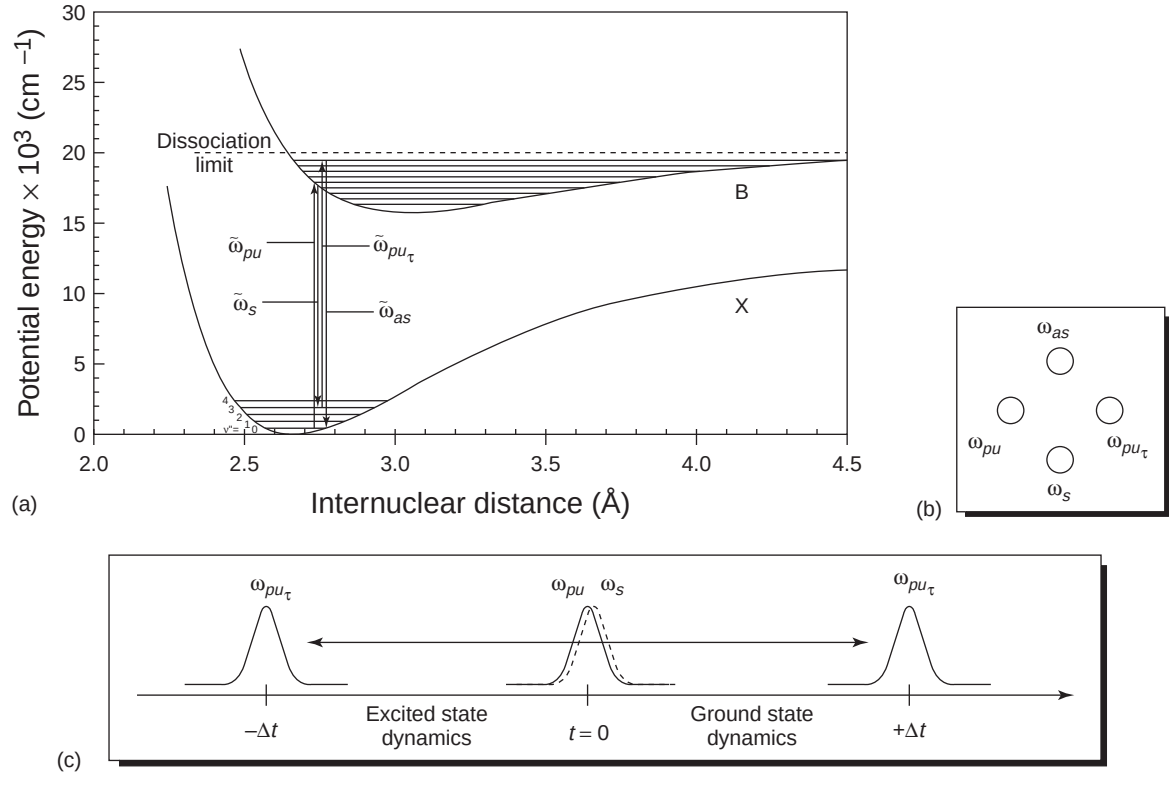

Figure 18 Illustration of the principles of femtosecond-time-resolved CARS spectroscopy. (a) Potential energy-level diagram of iodine with the four wave interaction; (b) spatial arrangement of the four waves; (c) temporal pulse pattern.

an angular frequency w_R of a Raman active vibrational mode. Using delay lines between the three pulses different type dynamics of the molecules under investigation can be probed.[58] In the case that first one pump pulse interacts with the matter, a wave packet is generated in the electronic excited state. If the two other laser pulses at w_L and at w_S enter the sample with a certain time delay (in the fs to ps time range, $\Delta t < 0$) the CARS transient is revealing the dynamic of the electronic excited state (see Figure 18c). However, when one pump and the Stokes pulse simultaneously interact with the sample a wave packet in the electronic ground state is prepared. Thus the delayed pump ($\Delta t > 0$) generates a CARS transient from which the vibrational and rotational dynamic of the ground state can be deduced (see Figure 18c).

In the same way other techniques such as degenerate four-wave mixing (DFWM) or CSRS have been successfully applied.[59]

4 INSTRUMENTATION FOR LINEAR RAMAN SCATTERING

In modern Raman instrumentation there are two alternate approaches: first, the dispersive Raman spectroscopy, and second, the FT–Raman spectroscopy. Both techniques are described in the next two parts of this section. Finally, in the last part, special Raman techniques such as micro-Raman spectroscopy and difference Raman spectroscopy will be discussed.

4.1 Dispersive Raman Spectroscopy

Considering the necessary parts of an experimental Raman setup (see Figure 19) the following components are required to perform a Raman experiment: (1) a light source for the excitation of Raman scattering, (2) a collection optics to collect the Raman scattered light, (3) a monochromator, i.e. a dispersing optical element to separate Raman scattered light into its wavelength components, (4) a detection system to measure the relative intensity of the Raman scattered light. In this section the different parts of a Raman setup will be briefly reviewed. Readers interested in more details are referred to several review articles and book chapters.[60,61]

4.1.1 Light Sources

In the mid-1960s there was a renaissance in conventional Raman spectroscopy by simply replacing the existing light sources (e.g. mercury lamps) by lasers. In general, radiation can be emitted by atoms, ions, or molecules which are in an excited state either spontaneously or the transition from the excited state to the ground state

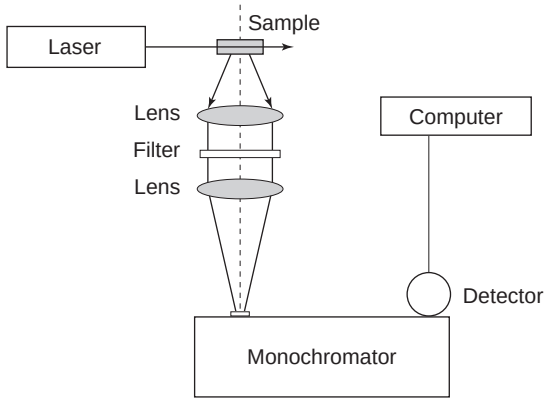

Figure 19 Conventional Raman setup.

can be stimulated by another photon. The latter occurs only if a photon of the same frequency interacts with the matter. Laser action is the result of the stimulated emission process which is giving rise to certain properties of this kind of light source, as follows.

- The outgoing laser photons produce a narrow, and essentially parallel beam, usually with a Gaussian intensity profile.
- The emitted waves are coherent, i.e. being in phase, unlike spontaneous light sources.
- Except for broad-band lasers, laser light is monochromatic.
- The parallel laser light can be easily focused using a lens. The focusing allows the recording of Raman spectra with high spatial resolution (see the micro-Raman spectroscopy section 4.3). However, there is a natural limit in focusing laser light which is called the diffraction limit.
- Apart from continuous wave (CW) lasers pulsed laser systems are also available. In general such a laser first accumulates energy as a population inversion is built up. Using a Q-switch enables the release of the accumulated energy through the emission of a single laser pulse which is usually very short.
- The laser light is usually polarized which results from the construction of the cavity.

In the following paragraph, a short overview of laser light sources used for Raman spectroscopy is given. First, one has to distinguish between CW and pulsed laser systems, and second, there are lasers with a fixed wavelength and those with a tunable wavelength range. For analytical linear Raman spectroscopy mostly CW lasers with fixed wavelength are applied. In the last two decades the argon and the krypton ion laser were used in many Raman laboratories. Enabled by the development of very sensitive micro-Raman setups

the He–Ne laser has also been used as light source for Raman spectroscopy. However, even at the same irradiance Raman spectra obtained from an argon ion laser at 514 nm will be 2.3 to 4.5 times more intense as Raman spectra obtained from a He : Ne laser operating at 633 nm. Such an w^4 dependence of the Raman intensity on the laser wavelength can be understood by considering Equations (33) and (34), where $w_S = w_L - w_R$. Recently, red diode lasers as well as CW Nd : YAG lasers have been used in micro-Raman spectroscopy. In particular, the investigation of biological material such as plant cells, living tissue, needs a light source with a wavelength in the near IR to avoid the excitation of fluorescence. For resonance Raman investigation laser lines in the UV are necessary. Such UV lines are provided for example by an argon ion laser (e.g. 351 and 363 nm or the frequency doubled lines at 244 and 257 nm) or pulsed laser systems such as the third (at 355 nm) and fourth harmonic output (at 266 nm) of a Nd : YAG laser. Since pulsed laser systems can easily provide high pulse intensities those laser systems are very common in nonlinear spectroscopy, such as SRS, CARS. The Nd : YAG with its fundamental at 1064 nm is a powerful, stable light source, which in addition is very reliable. To use the Nd : YAG laser in the VIS region it is necessary to double its frequency. Usually potassium trihydrogen phosphate, potassium dihydrogen phosphate, $LiNbO_3$, and beta-barium borate are used as doubling crystals. Often, the second harmonic (at 532 nm) and the third harmonic (at 355 nm) are used to pump dye lasers, which are tunable laser systems. More than 200 different dyes can be used for this purpose, which provide complete coverage for the range 200 nm to 1 μm. Often a Nd : YAG laser serves as pump for the generation of a population inversion between the electronic ground and excited state of the organic dye molecules. Placing the dye molecules into a cavity built up of two mirrors allows the stimulation of the fluorescence emission. Since the fluorescence occurs over a wide spectral range, monochromatic laser emission can be obtained by using dispersive optics such as a grating, prism or an etalon in the cavity. The tuning range for example for the dye Rhodamine B in ethanol pumped with the second harmonic of the Nd : YAG is between 580 and 630 nm.

In the field of pulsed laser systems a systematic development to shorter pulse length is observed. In the last decade femtosecond laser systems have become more and more popular in modern spectroscopy. Femtosecond laser pulses can be generated by a Ti : sapphire laser. The pulse length of such a solid state device is less than about 100 fs. The output of a Ti : sapphire is tunable between 690 and 1100 nm. Those pulses are usually very weak. Therefore, these pulses need to be amplified for application in Raman spectroscopy.

4.1.2 Optical Components

Closely related to the right choice of laser is the optimal configuration of the external optical components of a Raman setup. For a common Raman measurement there are three optical external devices necessary. First, if a CW laser is used, a filter (e.g. an interference filter) is needed to remove the plasma laser lines from the outgoing beam. Those plasma lines are generally much weaker than the laser line, but their Rayleigh scattered intensity is still of the order of the Raman intensities. Thus, a Raman spectrum will be complicated by those lines. Apart from using such a filter, which usually decreases the laser intensity by about 50%, premonochromators in the laser beam are often applied. A premonochromator is a device consisting of prisms or gratings. Dispersing the incident light allows the spatial separation of the laser line from the plasma lines. In order to use a premonochromator in the UV a quartz prism has to be used. Second, a spherical, microscope objective, or a cylindrical lens is used to focus the parallel laser beam onto the sample. Those lenses may be coated in order to reduce the reflectance of the laser light. Since the Raman intensity is proportional to the irradiance, which is the ratio of the laser power to the area of cross-section, focusing of the laser light increases the obtained Raman signals. However, high irradiance may cause damage to the sample. The minimum beam waist by focusing is fundamentally given by the diffraction limit. Dependent on the used wavelength of the laser, λ, the outgoing laser beam diameter, D, and the focal length of the lens, f, the beam waist, d, can be approximated by Equation (53):[27]

$$d = \frac{4\lambda f}{\pi D} \tag{53}$$

Third, to record Raman spectra, the optimal choice of collection optics is important. Usually the collection system consists of two lenses (see Figure 20). According to Equation (54) the recorded Raman intensity depends on the solid angle over which the scattered light is collected.

$$I_s \propto \mathrm{d}\Omega \tag{54}$$

In order to collect the largest possible solid angle a lens with a high aperture (up to 1 : 0.75) has to be used. The second lens focuses the collected light onto the entrance slit of the spectrometer. To ensure that the spectrometer is not overfilled or underfilled the lens has to match the f number of the spectrometer. The collected Raman intensity can be doubled by using a spherical mirror as depicted in Figure 20.

In general, in Raman spectroscopy two scattering geometries are most common. The scattered light can be collected at 90° or at 180° (back-scattering) to the

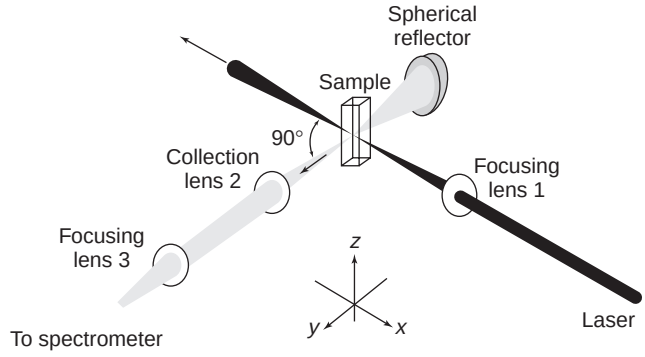

Figure 20 Optical arrangement commonly used in Raman spectroscopy.

incident laser beam. Instead of single lenses it is also very common to use camera objectives for the light collection.

4.1.3 Monochromator

For dispersing the collected light a monochromator is required. Since the generated Raman intensity is about six orders of magnitude weaker than the Rayleigh intensity, a monochromator has to discriminate efficiently the Rayleigh radiation. In Raman spectroscopy, single, double as well as triple monochromators using one, two, or three gratings, respectively, are well established.

Until the development of multichannel detectors (OMA, CCD) the scanning Czerny–Turner double monochromator or additive spectrometers with a photomultiplier as detector were most often applied. The stray light in these instruments is discriminated by two gratings with an intermediate slit. Nowadays the standard configuration is either a triple monochromator or a single monochromator in combination with sharp-cut filters, such as a multilayer dielectric interference filter or diffraction filter. In a triple monochromator the first two stages (subtractive mode) are used for Rayleigh stray light rejection, whereas the third monochromator, the so-called spectrograph, disperses the collected Raman radiation onto a multichannel detector. The advantage of such a device is that the Raman spectrum can be recorded very close (within a few wavenumbers) to the Rayleigh line. The drawback is the poor light transmittance. However, if high light throughput and collection efficiency is required, rather than measuring close to the excitation line, a single grating monochromator with a notch filter will be the appropriate choice.

4.1.4 Detector

The collected radiation is recorded using a detector, which transforms the input radiation into an output signal. The latter can be an electric charge, a current, or a potential. The ratio between output and input of a detector is defined as the responsivity of the detector. One distinguishes between single-channel and multichannel detectors. The most important single-channel detector is the photomultiplier. The physical base of this device is the photoelectric effect. The amplification of the incident photon signal is typical of 10^6 to 10^7 in such a photomultiplier.

In modern Raman setups the photomultiplier tube is replaced by multichannel detector devices. The most common multichannel detectors are intensified diode arrays and CCDs. Both detectors reveal the multiplex advantage, meaning that a large spectral range can be detected simultaneously. Typical intensified diode arrays comprise up to 1024 individual diodes, whereas CCDs are two-dimensional arrays of pixels (e.g. 512×1024).

4.1.5 Sample Preparation

For Raman excitation of liquids and solids, there is usually no special sample preparation necessary. If the measurements are performed in the VIS excitation range ordinary glass capillaries can be used. Also, substances dissolved in water can be investigated in glass capillaries by Raman spectroscopy without any further preparation. This is because water as well as glass reveals only very weak Raman signals. For the investigation of gases it is common to use multi-pass cells, in order to increase the Raman intensity. However, if resonance Raman experiments are carried out in the UV, the glass has to be replaced by quartz. Light-absorbing samples have to be treated with more care in order to avoid thermal decomposition. Kiefer and coworkers have developed several experimental arrangements for the investigation of solids as well as liquids[62–64] using spinning cell techniques.

4.2 Fourier Transform Raman Spectroscopy

In general, routine Raman spectroscopy of highly absorbing samples is associated with two problems. First, the excited fluorescence masks the Raman scattering because it is several orders of magnitude more intense. Second, the absorption of light very often causes a thermal decomposition of the sample. Often the fluorescence is caused by impurities and not by the sample itself. Most biological samples fluoresce when illuminated with VIS laser light.

These problems can be circumvented by using a near-IR wavelength for Raman excitation. In this spectral range the electronic transitions from the electronic ground state occur infrequently, which avoids excitation of fluorescence. Furthermore, the photochemical decomposition of the samples is reduced because of the low photon energy in the near-IR. However, this spectral range is

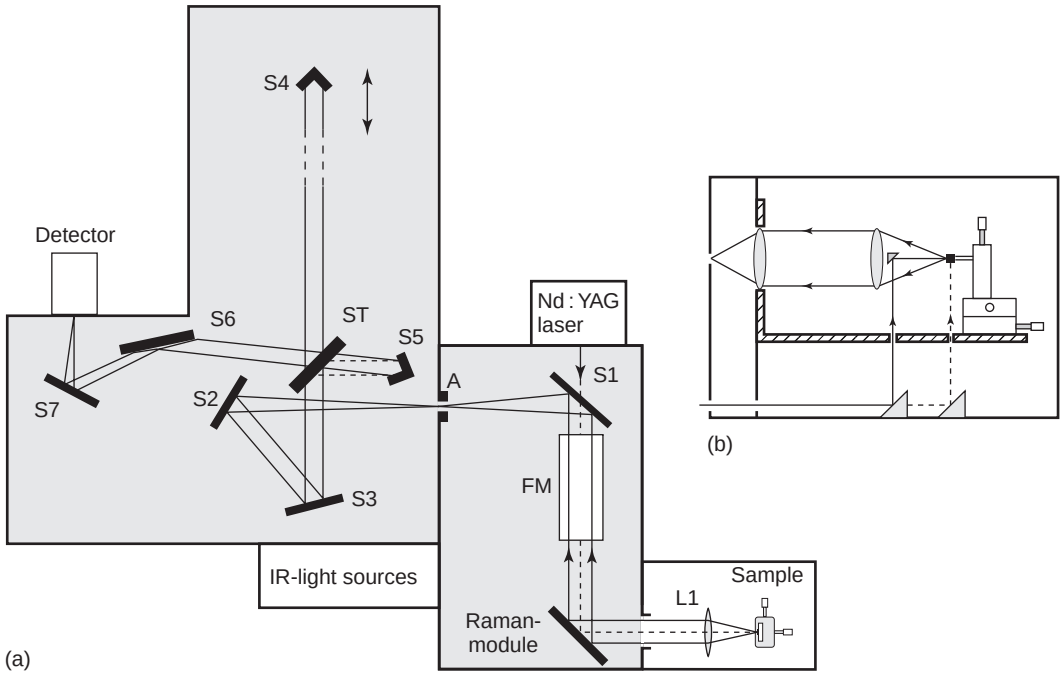

Figure 21 (a) FT-Raman setup. L1, lens; FM, filterbox to remove the laser line; S1–S7, mirrors; A, Jacquinot aperture; ST, beam splitter. (b) Illustration of 90° or 180° backscattering geometry.

due to the w^4 dependence of the Raman cross-section unfavorable for Raman spectroscopy. As an example, the sensitivity of a Raman experiment is reduced by a factor of 18 when changing the excitation wavelength from the 514 nm line of an argon ion laser to the 1064 nm line of a Nd : YAG laser.

Excitation in the near-IR together with FT-Raman spectroscopy overcame the problems of fluorescence, thermal and photochemical decomposition. Furthermore, the use of interferometric optics give rise for achieving multiplex measurements, partially compensating the w^4 dependence. Pioneering work in FT-Raman spectroscopy has been performed by Hirschfeld and Chase[65] in 1986 and Chase[66] in 1987 and several other groups have been active in this field since then.[60,67]

A conventional FT-Raman setup is composed of a light source, the collection optics, an interferometer, and a detector. A scheme of a FT-Raman apparatus and the sample holder is displayed in Figure 21. Usually a CW Nd : YAG laser operating at 1064 nm is used as light source. The single-mode output power lies between 10 mW and 10 W. The laser beam is focused onto the sample with a spot size of approx. 100 µm. The scattered light is collected by an aspherical lens either under 90° or 180° backscattering geometry (see panel (b) of Figure 21). A ''notch'' filter is used to remove the laser line from the collected light. The light is directed into the interferometer which is optimized for the near IR range. The detection is performed with a liquid-nitrogen-cooled germanium or room-temperature InGaAs detector. Usually, a fast Fourier transform (FFT) algorithm transforms the interferogram into a power density spectrum. The resolution of an FT-Raman spectrum is dependent on the distance x the moving mirror travels and the apodization factor used in transforming the spectrum. In an FT-Raman setup the entrance slit is replaced by a spherical pin-hole causing a much higher light throughput (Jacquinot advantage[60]) than is possible in a conventional dispersive Raman spectrometer.

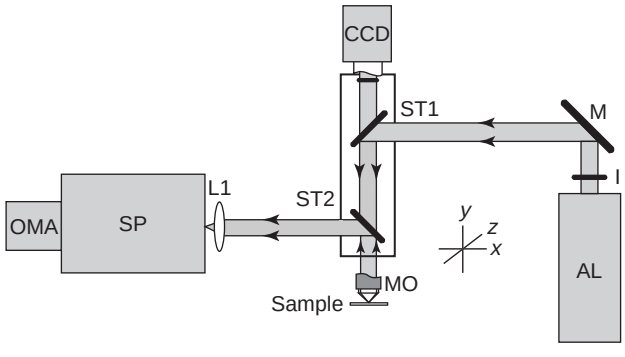

Figure 22 Schematic of a micro-Raman setup. AL, argon ion laser; I, interference filter; M, mirror; ST1, ST2, beam splitter; MO, microscope objective; CCD, CCD camera for sample observation; L1, lens; SP, spectrometer; OMA, optical multichannel array.

4.3 Special Raman Techniques

4.3.1 Micro-Raman Spectroscopy

Nowadays the technique of micro-Raman spectroscopy is a very well established method for both the investigation of micro-samples in the order of picograms or even less and the spectroscopic imaging of the sample surfaces.[68] One of the first micro-Raman setups has been described by Hirschfeld.[69] The author combined an optical microscope with a Raman setup. A scheme of a typical micro-Raman setup is displayed in Figure 22. A microscope objective serves both to focus the laser beam on the sample and to collect the scattered light. Microscope objectives with high numerical aperture and high magnification are used to focus the light to the diffraction limit. The 514 nm line of an argon ion laser can be focused to a spot of about 1 μm. In many cases the focused laser spot produces high-power density typically in the range between 10^4 and 10^6 W cm^{-2}. For Raman spectroscopy of microscopic samples various techniques have been developed (e.g. the rotating sample technique[70] or surface scanning Raman microscopy[71]) in order to avoid heating and decomposition of sensitive, especially absorbing samples. For the investigation of fluorescent micro-samples such as dyes and biological cells, the FT-Raman spectroscopy with long wavelength excitation is also available in combination with a microscope.

4.3.2 Difference Raman Spectroscopy

Raman difference spectroscopy involves simultaneous (or nearly simultaneous) recording of Raman signals from two different samples. This is accomplished by rotating the split cell[72] at the rate of about 1000–2000 rpm. Thus the Raman scattered signals originating from the two compartments of the split cell are alternately detected. After appropriate signal matching, one signal is subtracted from the other resulting in the difference spectrum. This technique can thus be used to examine solute bands without solvent interference, for example, by subtracting the solvent bands from those of the solution. Although it appears that this kind of technique only emulates the double-beam technique already known in IR spectroscopy, the RDS technique has an added advantage and offers a unique application in that it can be used to measure accurately wavenumber shifts or linewidth changes, especially when they are quite small.

The simultaneous recording of two (or more) signals increases remarkably the accuracy of the Raman data. This is primarily because several experimental errors such as reproducibility (backlash) error of the spectrometer from one scan to another, power variation of the exciting laser line with time, adjustment instabilities due to temperature effects, and fluctuations in the response function of the electronic detection system are almost eliminated by applying this technique.[72–76]

Beside the advantages mentioned above, the simultaneous recording of two Raman signals offers several other possibilities. The Raman signals, for instance, from two different polarization measurements can be compared with each other and this permits a direct recording of depolarization ratio[77] or the purely isotropic part[78] of the Raman scattered light, taking a linear combination $(I_{\parallel} - 4/3I_{\perp})$, as a function of wavenumber. In another application, by using circularly polarized light,[79] the circular intensity differential or the reversal coefficient can be recorded in a single scan; a detailed description of all aspects has been given by Kiefer[64] and Asthana and Kiefer.[80]

Considering the above mentioned types of RDS techniques it would be desirable to simultaneously record both the $I_{\parallel}$ and the $I_{\perp}$ from two different samples and it will obviously permit to combine the advantages of difference Raman studies and polarization measurements into one. In order to achieve it, a four-channel gated photon counting system was developed by Eichele et al.[81] which allowed simultaneous recording of four different signals as well as the linear combinations with desired coefficients. Based on this detection system, Laane and Kiefer developed the technique of four-channel Raman difference spectroscopy (FC-RDS).[82] With the advent of multichannel detectors two new concepts for RDS were introduced. Kamogawa and Kitagawa[83] used branched bundles of optical fibers to collect the Raman signals of two samples simultaneously on different rows of the CCD detector. A combination of the established rotating split cell scheme and multichannel detection was introduced by Deckert et al.[84] Here a setup with rotating blinds was used to measure distinct samples. To further increase the accuracy the latter scheme also utilizes the scanning multichannel technique.[85,86]

A very sophisticated technique in RDS was introduced by Martin.[87] He kept the sample fixed and used a rotating mirror device. This offers some unique possibilities, such as comparing crystals, gels, and fibers to water solution, all of the same material. Another advantage is the analysis of as little as 10 μL of scarce sample material. The backscattering RDS technique has been introduced by Eng et al.[88] A multisample system for RDS has been constructed by Savoie et al.[89]

5 INSTRUMENTATION FOR NONLINEAR RAMAN SCATTERING

A review of the instrumentation of several nonlinear Raman methods which are based on the second-order

($\chi^{(2)}$) and the third-order susceptibility ($\chi^{(3)}$) is given in the following. These methods include coherent CARS, SRGS, IRS, PARS, and IDSRS.

5.1 Hyper-Raman Spectroscopy

Hyper-Raman scattering signals are typically five to seven orders of magnitude weaker than those of linear Raman scattering. Larger hyper-Raman signals are produced under resonance conditions[90,91] and in surface-enhanced hyper-Raman scattering.[92–94] Since hyper-Raman scattering is an incoherent process which does not require phase matching, it is preferentially observed at an angle of 90° from the incident laser direction. As for any spontaneous phenomenon, a double monochromator is used in most cases. Initial hyper-Raman spectra were generated with a Q-switched ruby laser operating at ≈1 Hz[14] and detected with a synchronously gated photomultiplier tube and subsequently with optical multichannel detection.[95–97] The use of CW pumped acousto-optically Q-switched Nd : YAG lasers (repetition rates of 5 kHz), synchronously gated photomultiplier tubes,[98,99] and synchronously gated two-dimensional single photon counting detectors[100,101] has improved the signal-to-noise ratio of hyper-Raman spectra. Considerable further improvements have been obtained with mode-locked pulses (at 82 MHz) from a Nd : YAG laser to observe the surface-enhanced hyper-Raman signal from pyridine adsorbed on silver.[93] In these studies hyper-Raman signals were observed with intensities close to spontaneous Raman scattering. They showed that SEHRS has become a useful spectroscopic technique. In view of the recent advances in laser and detector technology, significant improvement in SEHRS sensitivity will come rapidly from the use of intensified CCD cameras for hyper-Raman signal detection and the use of a continuously tunable mode-locked Ti : sapphire laser as the excitation source.[102]

5.2 Coherent Anti-Stokes–Raman Scattering

CARS is the most widely practiced nonlinear Raman spectroscopy technique.[103] The major experimental advantage of CARS (and of most nonlinear coherent Raman techniques) is the large signal produced. In a typical CARS experiment in a liquid or a solid, the applied laser power of the pump and Stokes laser ($10^4 - 10^5$ W) generate an output power up to 1 W,[20] while conventional Raman scattering would give a collected signal power of $\approx 10^{-4}$ W with the same lasers. Since the CARS output is directional, the collection angle can be five orders of magnitude smaller than that needed in spontaneous scattering. Taken together, these two factors imply that CARS is nine orders of magnitude less sensitive to sample fluorescence than spontaneous scattering. The advantage is actually even greater since the CARS signal is at a higher frequency than any of the input laser frequencies. However, there are some disadvantages of CARS: (1) an unavoidable electronic background nonlinearity that alters the line shape and can limit the detection sensitivity; (2) a signal that scales as the square of the spontaneous scattering signal (and as the cube of the laser power), see Equation (51), making the signals from weakly scattering samples difficult to detect; and (3) the need to fulfill the phase matching requirement of section 3.3. While other techniques avoid these difficulties, CARS still remains the most popular coherent nonlinear technique. There are several experimental possibilities for carrying out CARS. Some of the most practical ones will be discussed next.

5.2.1 Scanning Pulsed Coherent Anti-Stokes–Raman Scattering

A typical CARS device for condensed phase spectroscopy which has been developed at the University of Würzburg[104] is shown in Figure 23. Because it contains two tunable (w_L, w_S) laser sources it is also suitable for performing resonance CARS spectroscopy in liquids or in solids.[105–108] For the simultaneous pumping of the two dye lasers a pulsed (10 Hz) Nd : YAG laser with a harmonic generator and a prism harmonic beam separator (Spectra Physics Quanta Ray models GCR-4, HG-4 and PHS-1, respectively) are used. The third harmonic of the Nd : YAG laser with wavelength at 355 nm is split into two beams by a 2 : 1 beam splitter (BS1 in Figure 23). These beams are then used to pump the two dye lasers. They supply the pump (w_L) and Stokes (w_S) beam respectively. The outputs of the dye lasers are corrected for divergence by lenses (L1 and L2) with long focal length. Dye luminescence of the pump laser is removed by a linear dispersing prism arrangement (PR) followed by an aperture (A1). The output of the dye lasers can be chosen to be horizontally or vertically polarized. To achieve exact polarization conditions Glan–Thompson polarizers (P1, P2) are added. The polarization of the Stokes beam can be turned separately by means of a double Fresnel rhombus (FR). The two laser beams are crossed and focused on the sample by an achromatic lens (AL1). The crossing angle can be tuned by changing the distance of the two parallel laser beams. The latter is achieved by moving the phase conditioning mirror (M4) in the way indicated in Figure 23. The CARS signal w_{AS} which is generated in the sample under a certain angle is made parallel to pump and Stokes beam after passing an achromatic lens (AL2). Moving mirror M5 in the direction indicated in Figure 23 while tuning the Stokes laser wavelength leads to a compensation of the change of the observation angle.

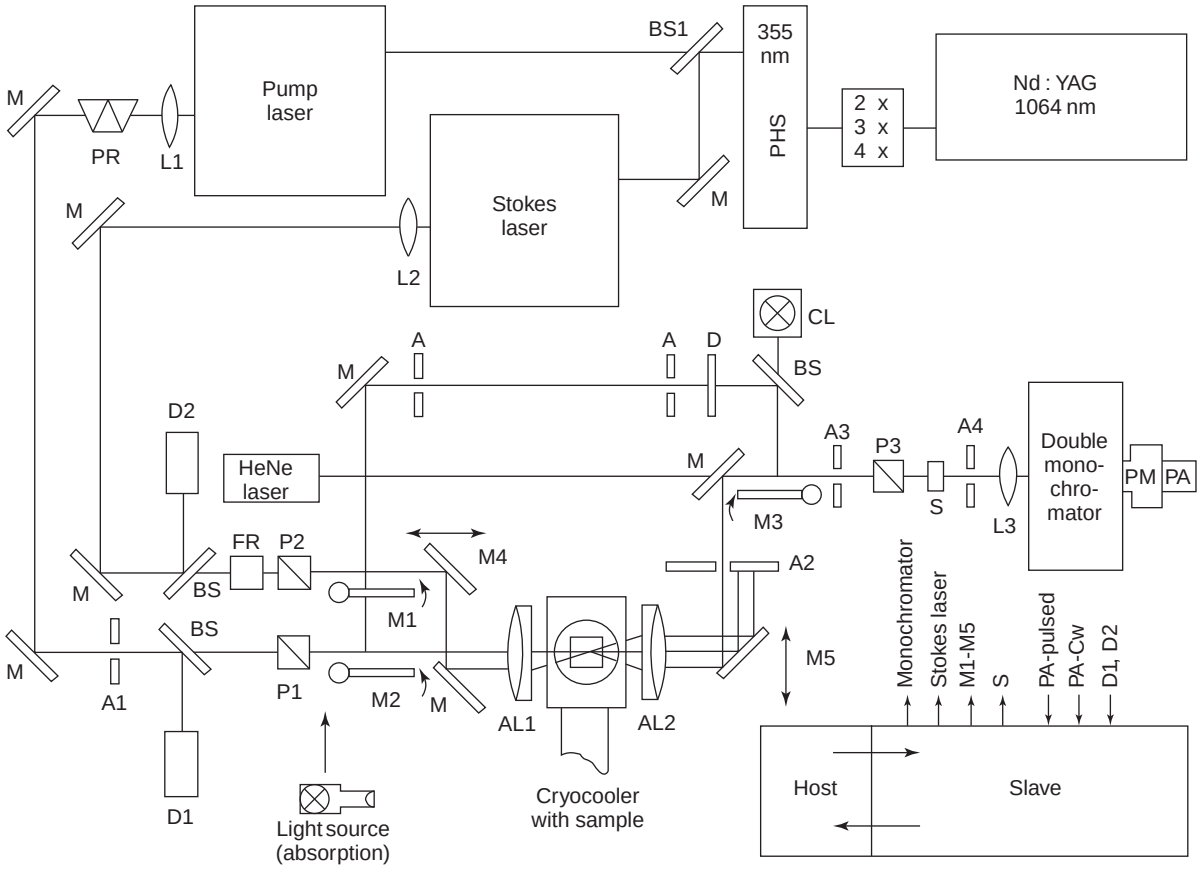

Figure 23 Experimental setup for CARS spectroscopy in condensed phase. A, aperture; AL, achromatic lens; BS, beam splitter; CL, calibration lamp; D, diffuser; D1, D2, diodes; FR, double Fresnel rhombus; L, lens, M, mirror; P, Glan–Thompson polarizer; PA, preamplifier; PHS, prism harmonic separator; PM, photomultiplier; PR, linear dispersing prism arrangement; S, shutter. (Reproduced from Materny et al.[104])

It is important to have the CARS beam in a final fixed position for further rigorous separation from the two dye laser beams. Part of the latter is performed by apertures (A2 and A3). After passing an analyzer (P3), a shutter (S) and a further aperture (A4), the CARS beam is finally focused (by L3) on the entrance slit of a double monochromator (Spex 1403) which suppresses remaining laser or stray light.

For the purpose of alignment a He : Ne laser is installed which fixes the optical axis of the double monochromator. The signal path is aligned along this axis. When the Stokes laser frequency is tuned, mirrors (M4 and M5) are simultaneously moved under computer control to compensate for dispersion (phase-matching). The monochromator can be calibrated by means of several calibration lamps (discharge lamps, CL) when mirror M3 is switched into the right position. Putting mirrors M1 or M2 at 45° with respect to the laser beams, the pump or Stokes beam, respectively, can be directed to the monochromator for exact determination of the laser wavelengths.

The CARS signal is detected by a Model RCA C31024A fast photomultiplier (PM). The intensities of pump and Stokes laser are measured by photodiodes (D1 and D2, respectively). The output signals of PM, D1 and D2 serve as input signals for the computer system. By means of the D1 output the pump laser triggers the complete electronic system for each laser shot. The computer controls the stepping motor of the double monochromator, the wavelength of the Stokes laser, position of mirrors M1–M5 and the shutter. Instead of a solid sample mounted in a cryocooler, as shown in Figure 23, a cell for liquids can also be used at the same place, which then allows CARS spectra of liquids to be obtained. The apparatus can be easily modified for CARS difference spectroscopy, which allows the simultaneous measurement of CARS spectra of two different samples under identical conditions.[109–111] CARS difference spectroscopy can be used similarly as Raman difference spectroscopy[64,72,80] for accurate determination of small frequency shifts and bandwidth changes in liquid mixtures, particularly when the sample is fluorescent.

Experimental problems may arise for CARS spectroscopy on samples with strong and broad absorption in the spectral region of interest. In such cases, the usual measurement procedure in which the CARS spectra are taken in transmission may be limited or even impossible. To overcome these difficulties, a backscattering CARS technique has been developed by Weippert et al.[112] The backward CARS signal is normally negligibly small compared with the forward signal because of a large phase mismatch.[113] However, in the case of strongly absorbing samples the short interaction length L (see Equation 51), which is then in the submicrometer range, causes the phase-matching condition to become less significant. Then, the backscattering CARS signal of strongly absorbing samples can have intensities even higher than those obtained in the conventional way, i.e. in a forward directed beam arrangement.

In order to observe backward scattered CARS signals from the surface of an opaque sample a special CARS set-up has been made, which is essentially the standard configuration as displayed in Figure 23 with specific modifications. The latter is shown in detail in Figure 24. Here, the sample area from Figure 23 is reproduced together with the modification. In this case, the Stokes beam runs parallel above the pump beam. Both are focused by the achromatic lens (AL1) and then deflected downwards by the mirror M6. The common focus is on the surface of the sample. The reflected upcoming beams including the generated CARS beam are reflected by mirror M7 so that after lens AL2 they again travel horizontally. This arrangement permits the CARS study

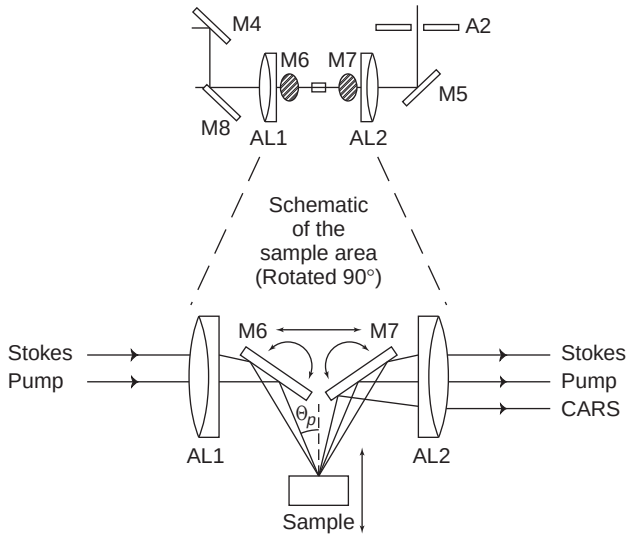

Figure 24 Modification of pulsed CARS setup shown in Figure 23 for backscattering measurements. Only the modified sample area is shown. (Reproduced from Weippert et al.[109])

of the surface of a highly absorbing sample which could be either a solid or a liquid.

Most of the CARS studies performed until now have been on gaseous samples. The main reason for this is the exceptionally high resolution combined with reasonable signal strengths. Although the nonlinear susceptibility $\chi^{(3)}_{CARS}$ of a gas is about two orders of magnitude smaller than that of a typical liquid, the interaction length can be much larger. There is nearly no dispersion of the medium if excitation is away from absorption and, hence, the phase-matching condition (Equation 46) requires a collinear beam arrangement. Since the optical breakdown threshold for gases is orders of magnitude higher than for liquids or solids, the high output power produced by the second harmonic of a Nd : YAG laser ($w_L = 532\,\text{nm}$) and a dye laser (angular frequency w_S) pumped simultaneously by w_L can be profitably used. Thus the CARS signal produced in gases with such a relatively simple CARS apparatus can be nearly as large as in liquids.[103] The high intensity beams with angular frequencies w_L and w_S can be best separated from the incident beams using a prism, dichroic mirrors or interference filters rather than a monochromator as shown in Figure 23.

5.2.2 *Multiplex Coherent Anti-Stokes–Raman Scattering*

Scanning CARS measurements require quite long recording times which may be too long with respect to the evolution of physical phenomena to be studied (e.g. chemical reactions). For such cases multiplex CARS is an attractive alternative to the single-frequency pulsed techniques. In multiplex CARS, the Stokes beam at w_S is produced by a broad band laser (typically $150–200\,\text{cm}^{-1}$) and the full spectrum ($150–200\,\text{cm}^{-1}$) at anti-Stokes angular frequency w_{AS} is recorded in a single laser shot.[114] A spectrograph with an optical multichannel analyzer (preferentially a cooled CCD camera) is needed for this purpose. In this case, however, the resolution is limited by the spectrograph and/or the optical multichannel analyzer employed. It should be noted that the multiplex CARS signal obtained from a single laser shot is comparatively small. However, averaging over many shots gives CARS spectra of similar quality as scanning CARS spectra if the lasers have comparable output power.

5.2.3 *Femtosecond Time-resolved Coherent Anti-Stokes–Raman Scattering*

Time-resolved CARS has produced much important information on the dynamics of molecular vibrations and rotations.[115–118] In particular, vibrational dephasing time constants in liquids and solids could be determined with this technique.[119–122] In these experiments

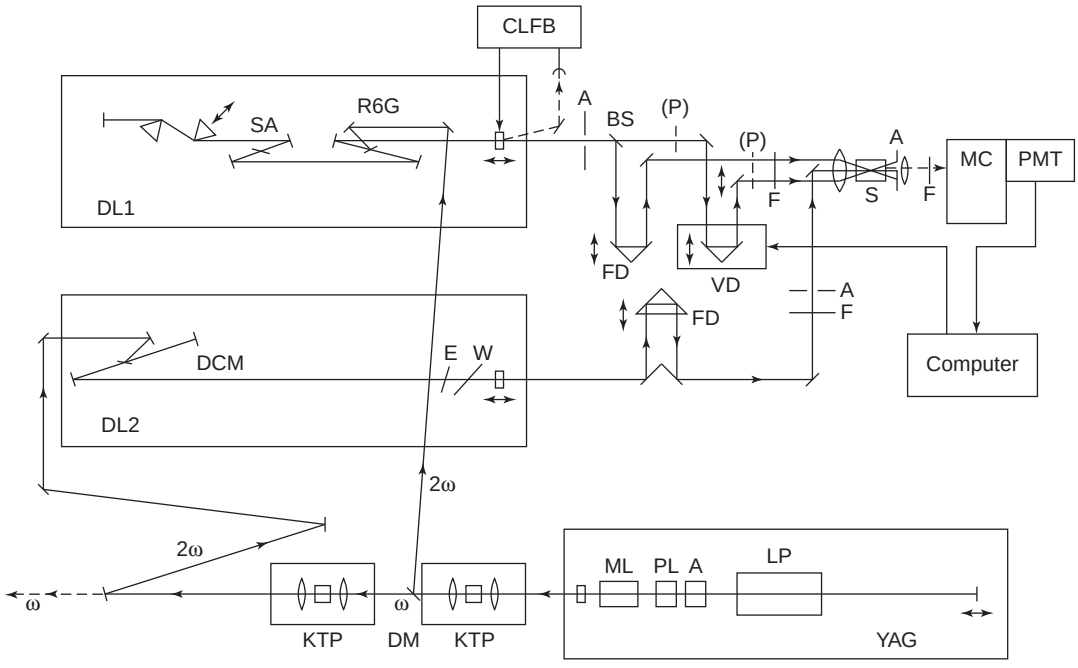

Figure 25 Schematic diagram of a femtosecond time-resolved CARS apparatus. YAG, CW mode-locked Nd : YAG; ML, mode locker; PL, polarizer; As, apertures; LP, laser pot; DM, dichroic mirror; DL1, femtosecond dye laser; SA, saturable absorber; CLFB, cavity-length feed-back system; DL2, picosecond dye laser; W, tuning wedge; E, etalon; FD, fixed delay; VD, variable delay; BS, beam splitter; Ps, half-wave plates (when necessary); Fs, filters; S, sample; MC, monochromator; PMT, cooled photomultiplier. (Reproduced from Okamoto and Yoshihara.[125])

pump and tunable Stokes beams serve to produce a coherent polarization oscillation in a medium at a resonant vibrational frequency. A separate time-delayed probe beam then generates an anti-Stokes signal whose intensity decreases with time due to dephasing and population-changing collisions. For solids and liquids, dephasing times are only a few picoseconds but in low density gases the signals persist for much longer periods.[115]

Recent technical developments in ultrashort-pulse lasers enable the CARS spectroscopist to obtain coherent light pulses shorter than 100 fs. This time scale corresponds to the period of molecular vibrational motions (100 fs $\approx$ 100–150 cm^{-1}) depending on the shape of the pulse, for example). One is therefore able to coherently excite many vibrational modes at a time and monitor relaxation processes in real time.

Femtosecond time-resolved CARS experiments were first reported by Leonhardt et al.[57] Following studies[123,124] showed beating patterns and fast decays of the coherent signal for several molecular liquids. Okamoto and Yoshihara[125] reported the observation of femtosecond time-resolved CARS under various polarization conditions. They also observed time-resolved CARS under preresonance and rigorous resonance conditions, which gave useful information on the vibrational relaxation in electronically excited states. With their femtosecond time-resolved CARS apparatus, the schematic diagram of which is displayed in Figure 25, they were able to study the femtosecond vibrational dephasing of a variety of molecular vibrations,[126–130] among them the C≡N and C≡C stretching vibration in liquid nitriles and liquid alkynes, respectively.

The basis of the experimental femtosecond CARS apparatus developed by Okamoto and Yoshihara[125] which is reproduced in Figure 25 is essentially the same as that of Leonhardt et al.[57] and Zinth and Kaiser[123] with the addition of the possibility to change the polarization of the laser radiation. The main parts of the system are two dye lasers with short pulses and high repetition rates, pumped by a CW mode-locked Nd : YAG laser (1064 nm, repetition rate 81 MHz). The beam of the first dye-laser which produces light pulses with 75–100 fs duration is divided into two parts of equal intensities and used as the pump and probe beam. After fixed (for the pump beam) and variable (for the probe beam) optical delay lines, the radiation is focused onto the sample together with the Stokes radiation produced by the second laser (DL2), which is a standard synchronously pumped dye laser. The anti-Stokes signal generated in the sample is separated from the three input laser beams by an aperture, an interference filter, and a monochromator, and detected by a photomultiplier. For further details we refer to Okamoto and Yoshihara.[125]

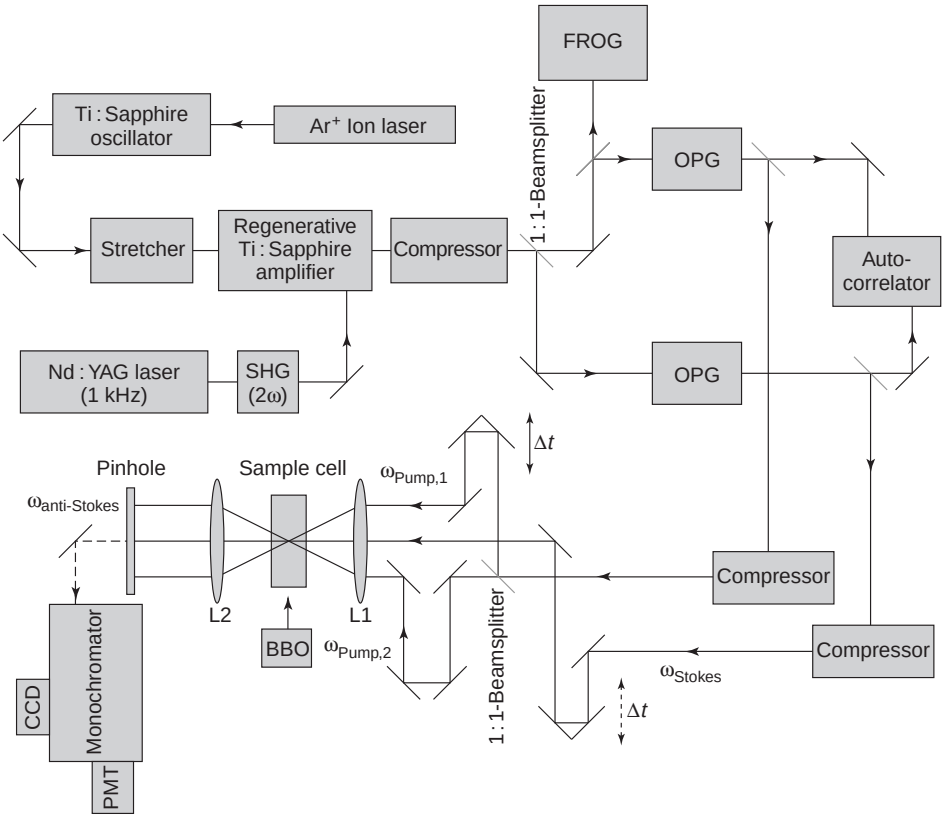

Figure 26 Schematic diagram of a femtosecond time-resolved CARS apparatus for the simultaneous study of ultrafast ground and excited state dynamics of gases. (Reproduced from Schmitt et al.[58])

Although time-resolved CARS is a powerful spectroscopic tool to study the vibrational relaxation of molecular systems, the information on dephasing dynamics via scattering from the isotropic part of the Raman tensor is often obscured by additional contribution of relaxation effects and the nonresonant part of the third order nonlinear susceptibility.[131] An elegant femtosecond three color polarization allows to suppress the unwanted contributions by several orders of magnitude.[132–135]

Recently, femtosecond time-resolved CARS was applied in order to investigate simultaneously the dynamics of the electronic ground and excited state. As model system iodine vapor has been chosen.[58] The experimental setup is shown in Figure 26. The output of a 76 MHz mode-locked Ti : sapphire laser (Coherent MIRA) was passed into a Nd : YAG-pumped regenerative Ti : sapphire amplifier (Clark-MXR) operating at 1 kHz. The pulses were stretched to a duration of $\leq$200 ps before amplification. The output pulses were recompressed yielding less than 100 fs with an energy of $\approx$1.5 mJ at 800 nm. In order to have two different colors available, the 800 nm pulse train was split into two parts by means of a 1 : 1 beam splitter. Using two four-pass optical parametric generators (OPGs) from Light Conversion, two independent wavelengths could be chosen. The laser pulses were produced using sum frequency generation between signal and idler output of the OPG as well as second harmonic generation of either the signal or the idler light. The pulses were finally compressed in double-pass two-prism arrangements resulting in temporal pulse widths of about 70 fs.

One of the OPG outputs was branched by a 1 : 1 beam splitter to produce the two pulses pu and pu_τ (see for notation Figure 18). The two pump beams were then aligned parallel to one another and spatially overlapped at the common focus in the cell by the achromatic lens L1. The Stokes beam S (output of the second OPG) was aligned parallel to and spatially separated from the two pump pulses and passed through the top of lens L1, focusing in the same region as the two pump beams. This folded BOXCARS configuration (see also panel (b) of Figure 18) was employed in order to separate the signal from the incoming pump and probe pulses.[136] In this geometry the phase-matching condition is fulfilled.

The pump pulses and the probe pulse could be delayed to each other by means of Michelson interferometer arrangements. The pump pulse pu and the Stokes pulse S were kept temporally overlapped and fixed. The CARS

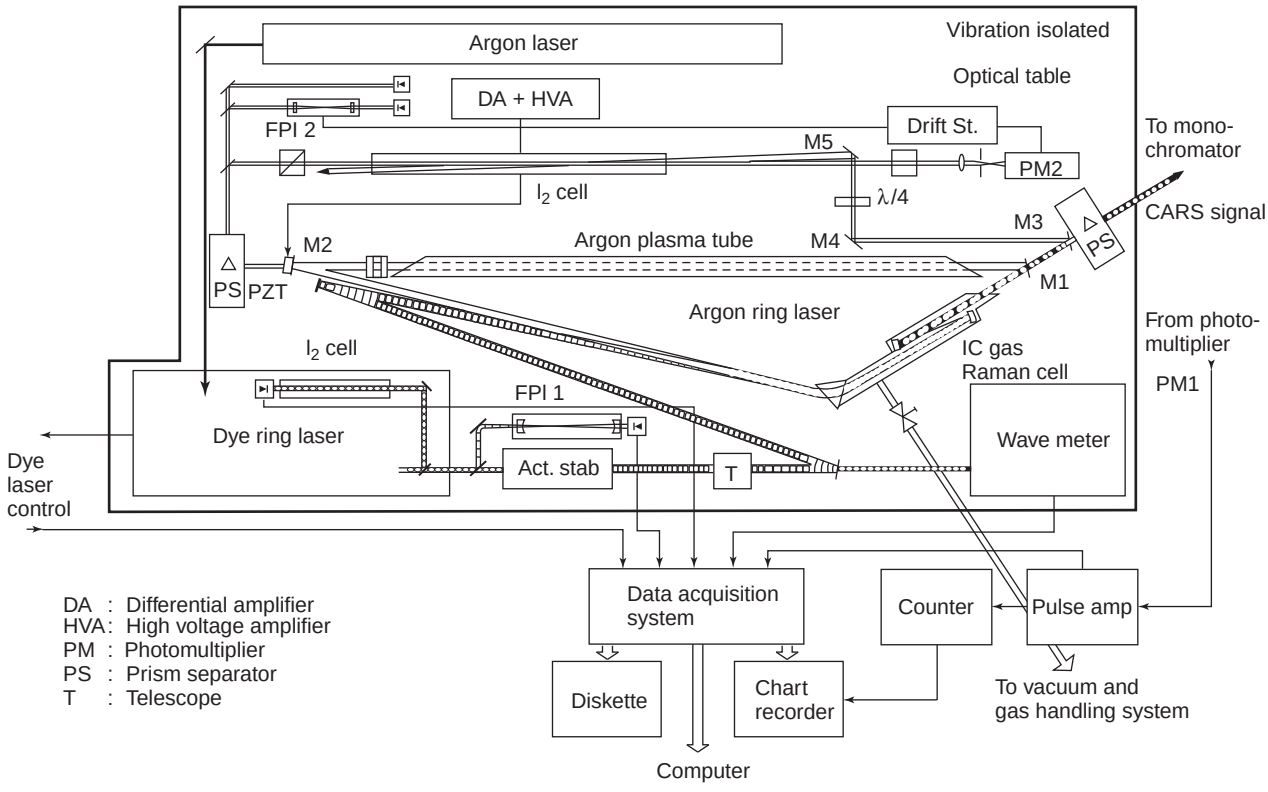

Figure 27 Schematic diagram of a CW CARS spectrometer with actively stabilized argon ion and dye ring lasers.[141]

transients were recorded as a function of delay time $\pm \Delta t$ between the pump pulse pu_τ and the two fixed and time coincident pulses pu and S (see panel (c) of Figure 18). The relative timing between the different pulses was varied with a computer-controlled actuator that allowed for optical delay up to 3 ns with a minimal stepsize of 6 fs. The determination of the position of temporal overlap (time zero) between the different beam pairs was made using a cross-correlation apparatus with second harmonic generation as well as sum frequency mixing in a thin phase-matched BBO crystal.

The femtosecond CARS signal pulse as generated in the sample cell travels in a direction determined by the phase-matching condition. This direction is different from that of the incident laser beams (see panel (b) of Figure 18), and the signal could thus be easily separated by a spatial filter. The CARS signal beam was collimated by a second achromatic lens (L2). After filtering the straylight by means of a monochromator (Acton SpectraPro-500), the anti-Stokes signal was detected by a fast photomultiplier tube (RCA C3 1024 A). The signal-to-noise ratio was enhanced by use of a boxcar integrator (EG&G model 4121B) in gated-integrator mode, as well as by numerical averaging of several pulses.

With this setup the vibrational/rotational dynamics of the iodine molecule in its excited electronic B-state as well as in its ground electronic X-state could be simultaneously studied in one experiment.[58,137]

5.2.4 *High-resolution Continuous Wave Coherent Anti-Stokes–Raman Scattering*

One of the main advantages of CARS and also of other nonlinear Raman spectroscopies is the high spectral resolution that can be achieved in spectra of gases at low pressures. The reason for this is that the instrumental resolving power in these techniques depends only on the convoluted linewidths of the lasers used for excitation, whereas in linear Raman spectroscopy the resolution is mainly determined by the monochromators used to disperse the observed scattered Raman light.

High resolution ($\approx 10^{-3}\,cm^{-1}$) CARS spectra can be achieved with single-mode CW lasers. The obvious disadvantage, however, is the drastically reduced CARS signal level due to the low power levels of CW lasers. The first CW CARS experiment was reported by Barrett and Begley[138] followed by experiments performed by Henesian et al.[139] and Fabelinsky et al.[140] More recently Schrötter et al.,[141] Schrötter and Boquillon,[142] and Frunder et al.[143] have developed a very efficient CW CARS spectrometer with a resolution of $0.001\,cm^{-1}$ (30 MHz). The high efficiency was achieved by intracavity

excitation in an argon ion ring laser modified from a commercial laser head.[143] In this way an intracavity power of 100 to 200 W in a single longitudinal mode at 514.5 nm, selected by an intracavity etalon, was obtained. Figure 27 shows the schematic arrangement of the argon ion ring laser in the CARS spectrometer including the provisions for an active stabilization to a hyperfine component in the absorption spectrum of iodine. The second, tunable laser frequency is provided by a commercial dye ring laser pumped by a second argon ion laser. The beam of the dye ring laser is coupled into the cavity of the argon ion laser at the prism and is focused collinearly with the argon ion laser beam inside a specially constructed gas cell for intracavity CARS excitation. The generated CARS beam is transmitted by the dichroic end mirror of the argon ion laser and spectrally purified by a prism and a grating monochromator. With this sophisticated set-up high resolution CARS spectra can be generated in the spectral region from about 2000 to 3500 cm^{-1}. Several applications to molecules in the gas phase have been reported.[141–144]

5.3 Stimulated Raman Gain Spectroscopy and Inverse Raman Spectroscopy

As outlined in section 3.3, both SRGS and IRS or SRLS fall into the class of coherent nonlinear Raman techniques. SRGS and IRS differ only in that spectra are obtained in the former case by using a Stokes shifted probe laser to scan the gain profile produced by a pump laser source, whereas in the latter case of IRS, an anti-Stokes shifted probe laser is used to scan the absorption profile (loss) on the opposite side of the spectrum relative to the pump frequency.[10,11,145,146] Thus, scanning of the frequency of one of the laser sources relative to the other results in a direct display of the Raman spectrum. The only difference in respect to the spontaneous Raman spectrum will be that, since this is a stimulated process, the signal will scale as the difference in population between the upper and lower level rather than being directly proportional to the lower-level population. The advantages of SRGS and IRS are that the signal (in contrast to CARS) is linearly proportional to the spontaneous Raman scattering cross-section (and to the product of the two laser intensities), and that the phase-matching condition is automatically fulfilled and there is no nonresonant background. The resolution of the nonlinear Raman techniques is limited only by the laser linewidths. These qualities give the stimulated Raman techniques particular appeal under conditions where interference from background luminescence is problematic or in situations where very high resolution is required. The main disadvantage of these techniques, however, is that they are quite sensitive to laser noise. The latter requires high stability in laser power. Pioneering work on stimulated Raman spectroscopies was performed using pulsed high-power laser sources in order to induce appreciable Raman gain or loss that could be easily observed.[11] Later techniques employed stable CW lasers and detect the small $\approx 10^{-5}$ changes in intensity due to Raman gain or loss.[10,145,146] Due to complexity, only a few stimulated Raman gain and loss spectrometers with main application in high resolution molecular spectroscopy have been built since the fundamental developments by Owyoung et al.[145] in 1978. An important improvement has been the realization of the quasi-CW stimulated spectrometer,[10] wherein a single-mode stable CW laser provides a low-noise probe source and a high-power respectively pulsed laser system provides the pump source. Such a scheme is diagrammatically shown in Figure 28 for inverse Raman (Raman loss) spectroscopy.[10] Here, the anti-Stokes shifted probe laser (angular frequency in this case is w_L) is temporally gated on for a time (100 μs) that is long compared to the high power pump pulse (angular frequency w_S, 6 ns), yet short enough to result in a low duty cycle, and thus preclude premature saturation of the detector. As the two beams cross in the sample at their common focus, the Stokes gain or anti-Stokes absorption produced by the pump beam is reflected in an amplitude modulation of the probe beam. Elimination of the background probe level is achieved by high-pass filtering. Such an approach thus incorporates both the large gains in sensitivity using a high-power pump source and the stable noise free advantage of a CW probe laser. Also, since the pump source is repetitively pulsed, a boxcar averager may be used to provide signal averaging which increases the signal-to-noise ratio even further. Similar to pure CW systems, the spectrum is obtained by scanning either the pump or the probe laser, which allows the gain or absorption profile to be displayed directly.

The construction of a spectrometer for high resolution stimulated Raman spectroscopy based on the quasi-CW approach described above, has been reported recently by

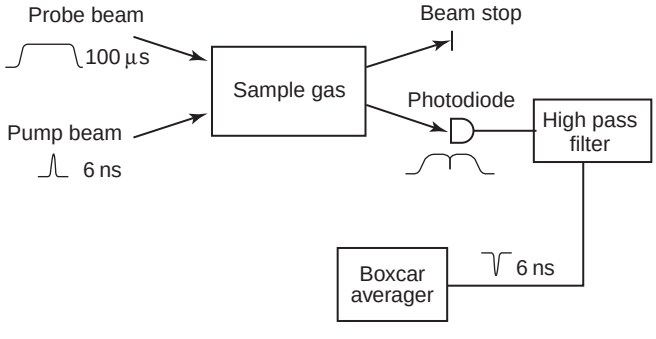

Figure 28 Schematic diagram for quasi-CW IRS. (Reproduced from Owyoung et al.[145])

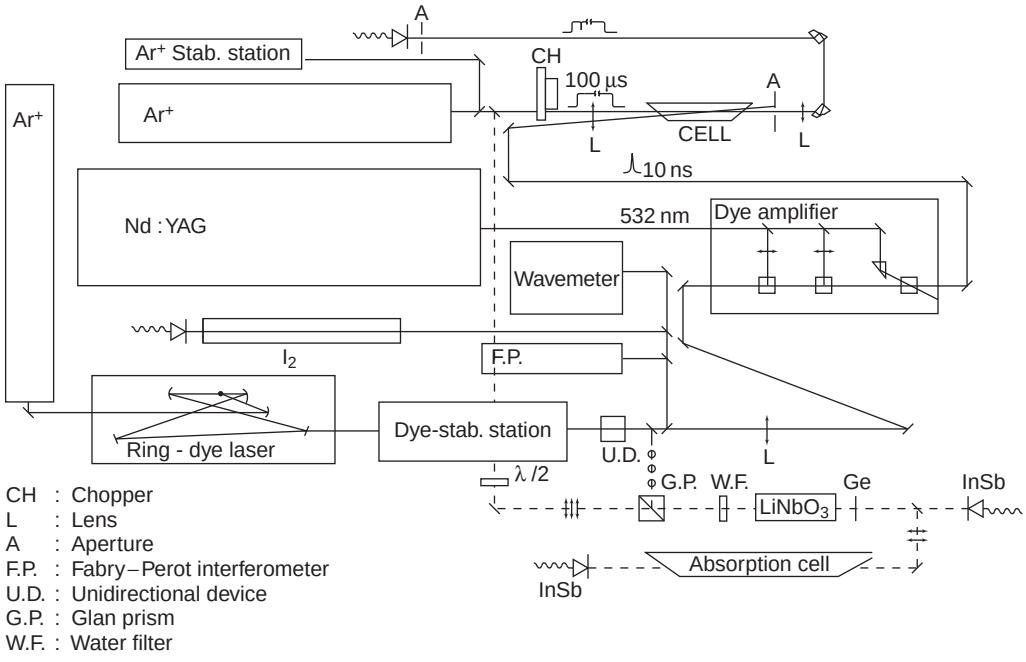

Figure 29 Schematic arrangement of a high-resolution quasi-CW inverse Raman and frequency difference IR spectrometers. (Reproduced from Bermejo et al.[147])

Bermejo et al.[147] Their experimental set-up is shown in Figure 29. It is also suitable for performing high resolution IR absorption spectroscopy. A frequency stabilized single mode argon ion laser delivers the probe beam. The tunable pump beam is produced by means of an argon ion pumped stabilized dye ring laser, whose output is amplified by a three stage dye amplifier which is pumped by a pulsed, frequency doubled Nd : YAG laser. A Fabry–Perot interferometer and a wavemeter serve for the accurate wavelength determination of the dye laser output. Part of the power of the pump and probe laser beams is used for generating a tunable, narrow-band difference frequency by means of non-linear mixing in a $LiNbO_3$ crystal. Hence, simultaneous recording at high resolution IRS and IR difference frequency[148] spectra is possible with this instrumentation. Both resolution and absolute accuracy of the observed Raman transitions are of the order of 10^{-3} cm^{-1}.

5.4 Photoacoustic Raman Spectroscopy

Acoustic detection of the Raman process using CW laser sources was first performed by Barrett and Berry.[149] In PARS the energy is deposited in the sample by the stimulated Raman process and the PARS signal is detected by sensitive acoustic methods. The use of high-power pulsed laser sources again gave great improvements in the sensitivity of this technique. This method is limited by the requirement of a static (''quiet'') sample and by the pressure necessary to transfer the acoustic signal to the microphone. The advantage of the method, which originates from the non-optical nature of the detection, is the absence of a Rayleigh line, because no energy can be deposited into the sample at zero Raman frequency (w_R), which enables to record pure rotational Raman spectra without any interference caused by Rayleigh scattering.[13,150,151]

Since the pioneering work by Barrett and coworkers only a few PARS experiments have been applied to some molecular gases.[152,153] The high potentiality of this technique when combined with the advantage of high-resolution spectroscopy using narrow linewidth lasers has been demonstrated recently by Rotger et al.[154] By a simple modification of their inverse Raman spectrometer,[155] they were able to perform high resolution (0.005 cm^{-1}) PARS and CARS spectroscopy of gases at about 10 torr $\approx$ 1.3 kPa. Their system, which is essentially a CARS set-up where a condenser microphone is inserted in the sample cell just in front of the beam focal point, is shown in Figure 30. Pump (w_L) and Stokes (w_S) lasers are a frequency doubled single mode Nd : YAG laser and a pulsed amplified tunable dye laser, respectively.

5.5 Ionization-detected Stimulated Raman Spectroscopy

IDSRS combines high resolution stimulated Raman spectroscopy with highly sensitive resonance enhanced multi-photon ionization (REMPI) to achieve an increase in sensitivity of over three orders of magnitude. The experimental apparatus to perform IDSRS is shown, in

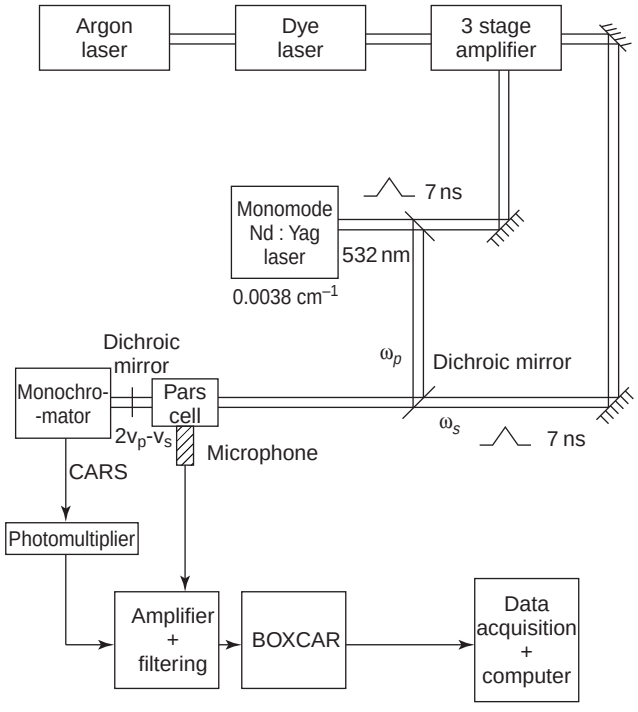

Figure 30 Schematic representation of a PARS/CARS experimental setup. (Reproduced from Rotger et al.[154])

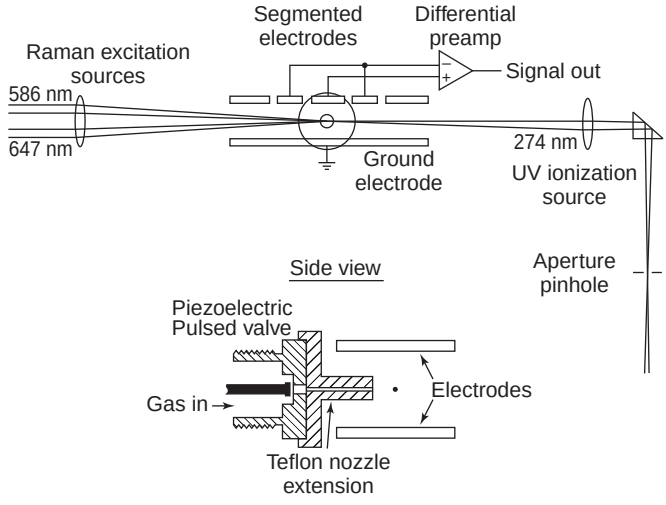

Figure 31 Diagram of the part of the experimental apparatus for IDSRS. (Reproduced from Esherick et al.[156])

part, in Figure 31.[156] The sensitivity of the technique is primarily limited by the level of background ionization present even when the Raman pump sources are off resonance. By using a segmented electrode design in conjunction with a differential amplifier (see Figure 31), it is possible to differentially detect the IDSRS signal relative to the background signal arising from regions where the three input laser beams do not cross. The background signal can thus be suppressed, significantly increasing the overall signal-to-noise ratio. High spectral resolution in the first step of the process is achieved by using two near-Fourier-transform limited bandwidth ($0.002\,cm^{-1}$) pulsed sources to drive the stimulated Raman transition. The latter two light sources originate from a single-mode CW krypton ion laser (647 nm) and a tunable single-mode CW dye laser (in Figure 31, 586 nm), respectively. Both of these sources are pulse amplified up to ≈1 MW in

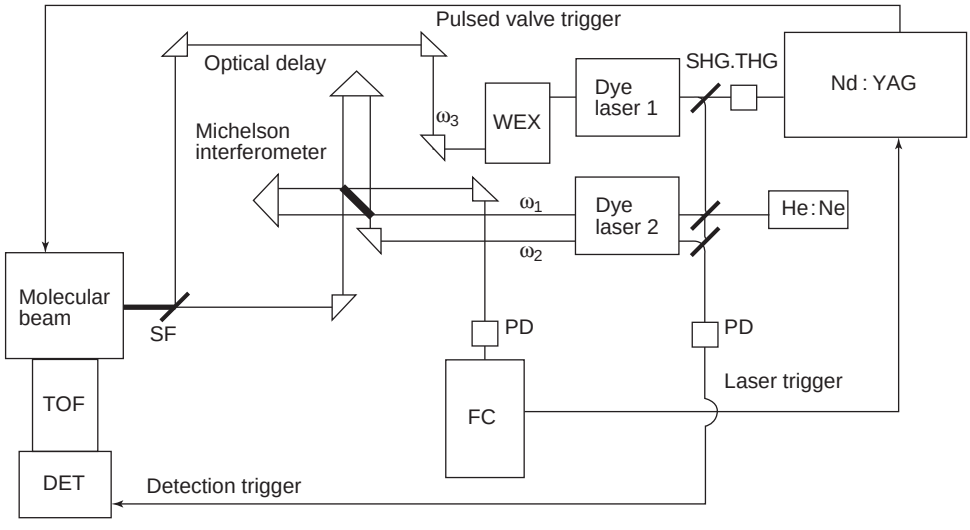

Figure 32 Apparatus configuration for mass-selective interferometric IDSRS experiments. He : Ne, single-mode, frequency stabilized He : Ne laser; SHG, THG, second- and third-harmonic-generating crystals; WEX, wavelength extender for frequency doubling with automatic crystal tracking; PDs, photodiodes; FC, fringe counting electronics for generating trigger pulses from the He : Ne laser output; SF, Suprasil flat or dichroic beam splitter for combining w_1, w_2, and w_3; TOF, DET, time-of-flight tube and ion-detection electronics. (Reproduced from Hartland et al.[159])

Nd : YAG-pumped, dye-amplifier chains. Since the resolution of the Raman spectrum is dependent only on the VIS Raman excitation sources, the linewidth requirements on the UV ionization laser are much less stringent. The frequency doubled output of a commercial Nd : YAG pumped dye laser is therefore sufficient for this source.

High-resolution Raman spectra are obtained by scanning the frequency of one of the Raman excitation sources and monitoring the ionization signal as a function of the frequency difference between this laser and the second, fixed frequency, pump laser. The IDSRS technique has also been applied for N_2.[157]

Since IDSRS involves the detection of ions, it can be readily implemented with mass-selective detection of those ions.[158–160] In fact, ion-mass analysis (by time-of-flight techniques, for example) greatly enhances the capabilities of the IDSRS techniques.

Besides the frequency domain versions of mass-selective IDSRS, in which the frequency is scanned while photo-ions are detected, the same group has also developed Fourier transform versions of this scheme (FT-IDSRS).[159] In this method the two stimulating Raman excitation beams are directed through a Michelson interferometer before being combined with the ionization UV laser field (see Figure 32). Photoions are then detected as a function of interferometer delay. The result is an interferogram that upon Fourier transformation yields a Raman spectrum whose resolution does not depend on the bandwidths of the Raman excitation sources, but, instead, on the delay range of the interferometer scanned in the experiment. Mass-selective IDSRS and FT-IDSRS have been used in a number of studies, including one of the benzene dimer.[160]

ACKNOWLEDGMENTS

The authors wish to thank V. Roman and S. Schlücker for their help in preparing this paper. One of us (JP) gratefully acknowledges the support from the Freistaat Bayern (Bayerisches Habilitations-Stipendium), and WK thanks the German Science Foundation for the financial support through the Sonderforschungsbereich SFB 347 and SFB 410 as well as by the Fonds der Chemischen Industrie.

ABBREVIATIONS AND ACRONYMS

CARS	Coherent Anti-Stokes–Raman Scattering
CCD	Charge-coupled Device
CSRS	Coherent Stokes–Raman Scattering
CW	Continuous Wave
DFWM	Degenerate Four-wave Mixing
EP	Excitation Profile
FFT	Fast Fourier Transform
FT-Raman	Fourier Transform Raman
IDSRS	Ionization-detected Stimulated Raman Spectroscopy
IR	Infrared
IRS	Inverse Raman Scattering
LIF	Laser-induced Fluorescence
MPI	Multiphoton Ionization
Nd : YAG	Neodymium : Yttrium Aluminum Garnet
OPG	Optical Parametric Generator
PARS	Photoacoustic Raman Spectroscopy
REMPI	Resonance Enhanced Multi-photon Ionization
RIKE	Raman-induced Kerr Effect
RISRS	Resonant Impulsive Stimulated Raman Scattering
RRE	Resonance Raman Effect
RRS	Resonance Raman Spectroscopy
SERS	Surface-enhanced Raman Spectroscopy
SRGS	Stimulated Raman Gain Spectroscopy
SRLS	Stimulated Raman Loss Spectroscopy
SRS	Stimulated Raman Scattering
UV	Ultraviolet
VIS	Visible

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Microbiology

Biomolecules Analysis **(Volume 1)**
Raman Spectroscopy in Analysis of Biomolecules

Coatings **(Volume 2)**
Infrared and Raman Spectroscopy and Imaging in Coatings Analysis

Process Instrumental Methods **(Volume 9)**
Raman Spectroscopy in Process Analysis

Infrared Spectroscopy **(Volume 12)**
Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials

Raman Spectroscopy **(Volume 15)**
Raman Spectroscopy: Introduction • Dispersive Raman Spectroscopy, Current Instrumental Designs • Fourier Transform Raman Instrumentation • Raman Microscopy and Imaging

General Articles **(Volume 15)**
Ultrafast Laser Technology and Spectroscopy

REFERENCES

1. A. Smekal, 'Zur Quantentheorie der Dispersion', *Naturwiss.*, **11**(43), 873–875 (1923).
2. H.H. Kramers, W. Heisenberg, 'Über die Streuung von Strahlung durch Atome', *Z. Phys.*, **31**, 681–708 (1925).
3. P.A.M. Dirac, 'The Quantum Theory of Dispersion', *Proc. R. Soc.*, **114A**, 710–728 (1927).
4. C.V. Raman, K.S. Krishnan, 'A New Type of Secondary Radiation', *Nature*, **121**, 501–502 (1928).
5. G. Landsberg, L. Mandelstam, 'Eine neue Erscheinung bei der Lichtzerstreuung in Kristallen', *Naturwiss.*, **16**(28), 557–558 (1928).
6. K.F. Kohlrausch, *Der Smekal-Raman-Effekt*, Springer, Berlin, 1931; Ergänzungsband, 1931–1937, Springer, 1938.
7. H.W. Schrötter, 'Raman Spectroscopy with Laser Excitation', in *Raman Spectroscopy: Theory and Practice*, ed. H.A. Szymanski, Plenum Press, New York, 69–120, 1970.
8. E.J. Woodbury, W.K. Ng, 'Ruby Laser Operation in the Near IR', *Proc. IEEE*, **50**(11), 2367 (1962).
9. P.D. Maker, R.W. Terhune, 'Study of Optical Effects due to an Induced Polarization Third Order in the Electric Field Strength', *Phys. Rev.*, **137A**(3), 801–818 (1965).
10. A. Owyoung, 'High Resolution Coherent Raman Spectroscopy of Gases', in *Laser Spectroscopy IV, Springer Series Opt. Sci., 21*, eds. H. Walther, K.W. Rothe, Springer, Berlin, 175–187, 1979.
11. W.J. Jones, B.P. Stoicheff, 'Inverse Raman Spectra: Induced Absorption at Optical Frequencies', *Phys. Rev. Lett.*, **13**(22), 657–659 (1964).
12. D. Heiman, R.W. Hellwarth, M.D. Levenson, G. Martin, 'Raman-induced Kerr Effect', *Phys. Rev. Lett.*, **36**(4), 189–192 (1976).
13. J.J. Barrett, 'Photoacoustic Raman Spectroscopy', in *Chemical Applications on Nonlinear Raman Spectroscopy*, ed. A.B. Harvey, Academic Press, New York, 89–171, 1981.
14. R.W. Terhune, P.D. Maker, C.M. Savage, 'Measurements of Nonlinear Light Scattering', *Phys. Rev. Lett.*, **14**(17), 681–684 (1965).
15. D.A. Long, 'The Polarizability and Hyperpolarizability Tensors', in *Nonlinear Raman Spectroscopy and its Chemical Applications, Nato Advanced Study Institute Series C, 93*, eds. W. Kiefer, D.A. Long, Reidel, Dordrecht, 99–130, 1982.
16. N. Bloembergen, 'Wave Propagation in Nonlinear Electromagnetic Media', *Proc. IEEE*, **51**, 124–131 (1963).
17. H. Berger, B. Lavorel, G. Millot, 'Nonlinear Raman Spectroscopy', in *Applied Laser Spectroscopy*, ed. D.L. Andrews, VCH, Weinheim, 267–318, 1992.
18. N. Bloembergen, 'The Stimulated Raman Effect', *Am. J. Phys.*, **35**(11), 989–1023 (1967).
19. N. Bloembergen, 'Stimulated Raman Effect and Coherent Raman Scattering', in *Proceedings of the 6th Int. Conf. on Raman Spectr.*, eds. E.D. Schmid, R.S. Krishnan, W. Kiefer, H.W. Schrötter, Heyden, London, 335–344, Vol. 1, 1978.
20. M.D. Levenson, *Introduction to Nonlinear Laser Spectroscopy*, Academic Press, New York, 1982.
21. J.W. Nibler, G.A. Pubanz, 'Coherent Raman Spectroscopy of Gases', in *Advances in Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Wiley, Chichester, 1–50, Vol. 15, 1988.
22. M.A. Yuratich, D.C. Hanna, 'Coherent Anti-Stokes Raman Spectroscopy (CARS). Selection Rules, Depolarization Ratios and Rotational Structure', *Mol. Phys.*, **33**(3), 671–682 (1977).
23. G. Placzek, 'Rayleigh-Streuung und Raman-Effekt', in *Handbuch der Radiologie*, ed. E. Marx, Akad. Verl.-Ges., Leipzig, 209–374, Band VI, Teil II, 1934.
24. C. Gerthsen, H.O. Kneser, H. Vogel, *Physik*, Springer, Berlin, 431–433, 1997.
25. M. Alonso, E.J. Finn, *Fundamental University Physics*, Addison-Wesley, London, 1977.
26. T.C. Damen, S.P.S. Porto, B. Tell, 'Raman Effect in Zinc Oxide', *Phys. Rev.*, **142**(2), 570–574 (1966).
27. D.A. Long, *Raman Spectroscopy*, McGraw–Hill, London, 1977.
28. E.B. Wilson, J.C. Decius, P.C. Cross, *Molecular Vibrations*, Dover Publ., Appendix IV, 1955.
29. F.A. Cotton, *Chemical Applications of Group Theory*, Appendix II A, Wiley, New York, 1990.
30. A.C. Albrecht, 'On the Theory of Raman Intensities', *J. Chem. Phys.*, **34**(5), 1476–1484 (1961).
31. J. Behringer, J. Brandmüller, 'Der Resonanz-Raman-Effekt', *Z. Elektrochem.*, **60**(7), 643–679 (1956).
32. For reviews see citations in the book article by W. Kiefer, 'Applications of Non-classical Raman Spectroscopy: Resonance Raman, Surface Enhanced Raman, and Nonlinear Raman Spectroscopy', in *Infrared and Raman Spectroscopy – Methods and Applications*, ed. B. Schrader, VCH, Weinheim, 465–517, 1995.
33. M. Fleischmann, P.J. Hendra, A.J. McQuillan, 'Raman Spectra of Pyridine Adsorbed at a Silver Electrode', *Chem. Phys. Lett.*, **26**(2), 163–166 (1974).
34. K. Kneipp, Y. Wang, H. Kneipp, I. Itzkan, R.R. Dasari, M.S. Feld, 'Population Pumping of Excited Vibrational-states by Spontaneous Surface-enhanced Raman-scattering', *Phys. Rev. Lett.*, **76**(14), 2444–2447 (1996).
35. K. Kneipp, Y. Wang, H. Kneipp, I. Itzkan, R.R. Dasari, M.S. Feld, 'Single Molecule Detection Using Surface-enhanced Raman Scattering (SERS)', *Phys. Rev. Lett.*, **78**(9), 1667–1670 (1997).
36. S. Nie, S.R. Emory, 'Probing Single Molecules and Single Nanoparticles by Surface-enhanced Raman Scattering', *Science*, **275**, 1102–1106 (1997).
37. R.K. Chang, T.S. Furtak, *Surface Enhanced Raman Scattering*, Plenum, New York, 1982.

38. J.A. Creighton, R.J.H. Clark, R.E. Hester, *Spectroscopy of Surfaces*, Wiley, New York, 1988.
39. M. Moskovits, 'Surface-enhanced Spectroscopy', *Rev. Mod. Phys.*, **57**(3), 783–826 (1985).
40. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation*, Academic Press, New York, 1969.
41. P.A. Franken, A.E. Hill, C.W. Peters, G. Weinreich, 'Generation of Optical Harmonics', *Phys. Rev. Lett.*, **7**(4), 118–119 (1961).
42. W. Kaiser, C.G.B. Garrett, 'Two-photon Excitation in $CaF_2:Eu^{2+}$', *Phys. Rev. Lett.*, **7**(6), 229–231 (1961).
43. P. Esherick, A. Owyoung, 'Ionization-detected Stimulated Raman Spectroscopy', *Chem. Phys. Lett.*, **103**(3), 235–240 (1983).
44. A.C. Eckbreth, 'BOX CARS: Crossed-beam Phase-matched CARS Generation in Gases', *Appl. Phys. Lett.*, **32**(7), 421–423 (1978).
45. J.A. Shirley, R.J. Hall, A.C. Eckbreth, 'Folded Box CARS for Rotational Raman Studies', *Opt. Lett.*, **5**(9), 380–382 (1980).
46. A.H. Zewail, *Ultrafast Dynamics of the Chemical Bond*, World Scientific, Singapore, Vol. 1–2, 1994.
47. J. Manz, L. Wöste, *Femtosecond Chemistry*, VCH, Weinheim, Vols 1–2, 1995.
48. M. Chergui (ed.), *Femtosecond Chemistry*, World Scientific, Singapore, 1996.
49. M.J. Rosker, M. Dantus, A.H. Zewail, 'Femtosecond Real-time Probing of Reactions. I. The Technique', *J. Chem. Phys.*, **89**(10), 6113–6127 (1988).
50. M. Dantus, M.J. Rosker, A.H. Zewail, 'Femtosecond Real-time Probing of Reactions. II. The Dissociation of ICN', *J. Chem. Phys.*, **89**(10), 6128–6140 (1988).
51. L.-E. Berg, M. Beutter, T. Hansson, 'Femtosecond Laser Spectroscopy on the Vibrational Wave-packet Dynamics of the $A^1\Sigma^+$ State of NaK', *Chem. Phys. Lett.*, **253**(3–4), 327–332 (1996).
52. F.E. Doany, R.M. Hochstrasser, B.I. Greene, R.R. Millard, 'Femtosecond-resolved Ground-state Recovery of Cis-stilbene in Solution', *Chem. Phys. Lett.*, **118**(1), 1–5 (1985).
53. T. Kühne, P. Vöhringer, 'Vibrational Relaxation and Geminate Recombination in the Femtosecond-photodissociation of Triiodide in Solution', *J. Chem. Phys.*, **105**(24), 10 788–10 802 (1996).
54. U. Banin, U. Bartana, U. Ruhman, R. Kosloff, 'Impulsive Excitation of Coherent Vibrational Motion Ground Surface Dynamics Induced by Intense Short Pulses', *J. Chem. Phys.*, **101**(10), 8461–8481 (1994).
55. T. Baumert, S. Pederson, A.H. Zewail, 'Femtosecond Real-time Probing of Reactions. 12. Vectorial Dynamics of Transition-states', *J. Chem. Phys.*, **97**, 12 447–12 459 (1993).
56. A. Mokhtari, J. Chessnoy, A. Laubereau, 'Femtosecond Time- and Frequency-resolved Fluorescence Spectroscopy of a Dye Molecule', *Chem. Phys. Lett.*, **155**(6), 593–598 (1989).
57. R. Leonhardt, W. Holzapfel, W. Zinth, W. Kaiser, 'Terahertz Quantum Beats in Molecular Liquids', *Chem. Phys. Lett.*, **133**(5), 373–377 (1987).
58. M. Schmitt, G. Knopp, A. Materny, W. Kiefer, 'Femtosecond Time-resolved Coherent Anti-Stokes Raman Scattering for the Simultaneous Study of Ultrafast Ground and Excited State Dynamics: Iodine Vapor', *Chem. Phys. Lett.*, **270**, 9–15 (1997).
59. M. Schmitt, G. Knopp, A. Materny, W. Kiefer, 'Femtosecond Time-resolved Four-wave Mixing Spectroscopy in Iodine Vapor', *Chem. Phys. Lett.*, **280**, 339–347 (1997).
60. B. Schrader (ed.), *Infrared and Raman Spectroscopy – Methods and Applications*, VCH, Weinheim, 1995.
61. D.L. Gerrard, H.J. Bowley, 'Instrumentation for Raman Spectroscopy', in *Practical Raman Spectroscopy*, eds. D.J. Gardiner, P.R. Graves, Springer, Berlin, 55–76, 1989.
62. N. Zimmerer, W. Kiefer, 'Rotating Surface Scanning Technique for Raman Spectroscopy', *Appl. Spectrosc.*, **28**(3), 279–281 (1974).
63. W. Kiefer, W.J. Schmid, J.A. Topp, 'A Universal Rotating System for Raman Spectroscopy', *Appl. Spectrosc.*, **29**(5), 434–436 (1975).
64. W. Kiefer, 'Recent Techniques in Raman Spectroscopy', in *Advances in Infrared and Raman Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Heyden, London, 1–42, Vol. 3, 1977.
65. T. Hirschfeld, B. Chase, 'FT-Raman Spectroscopy: Development and Justification', *Appl. Spectrosc.*, **40**(2), 133–137 (1986).
66. B. Chase, 'Fourier-transform Raman-spectroscopy', *Anal. Chem.*, **59**(14), A881 (1987).
67. C.G. Zimba, V.M. Hallmark, J.D. Swalen, J.F. Rabolt, 'Fourier Transform Raman Spectrometry of Long-chain Molecules Containing Strongly Absorbing Chromophores', *Appl. Spectrosc.*, **41**(5), 721–726 (1987).
68. M. Bridoux, M. Delhaye, 'Time-resolved and Space-resolved Raman Spectroscopy', in *Advances in Nonlinear Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Wiley, Chichester, 140–153, Vol. 2, 1976.
69. T. Hirschfeld, 'Raman Microprobe: Vibrational Spectroscopy in the Femtogram Range', *J. Opt. Soc. Am.*, **63**(4), 476 (1973).
70. W. Kiefer, H.J. Bernstein, 'Rotating Raman Sample Technique for Colored Crystal Powders. Resonance Raman Effect in Solid Potassium Permanganate', *Appl. Spectrosc.*, **25**(6), 609–613 (1971).
71. M. Lankers, D. Göttges, A. Materny, K. Schaschek, W. Kiefer, 'A Device for Surface-scanning Micro-Raman Spectroscopy', *Appl. Spectrosc.*, **46**(9), 1331–1334 (1992).
72. W. Kiefer, 'Raman Difference Spectroscopy with the Rotating Cell', *Appl. Spectrosc.*, **27**(4), 253–257 (1973).
73. J. Laane, W. Kiefer, 'Determination of Frequency Shifts by Raman Difference Spectroscopy', *J. Chem. Phys.*, **72**(10), 5305–5311 (1980).

74. B.P. Asthana, W. Kiefer, 'Precise Determination of Frequency Shifts by Raman Difference Spectroscopy using Voigt Profile for the Raman Band', *Appl. Spectrosc.*, **37**(4), 334–340 (1983).
75. B.P. Asthana, H. Takahashi, W. Kiefer, 'Study of Vibrational Dephasing and Frequency Shifts in Hydrogen-bonded Systems by Four-channel Raman Difference Spectroscopy', *Chem. Phys. Lett.*, **94**(1), 41–47 (1983).
76. B.P. Asthana, W. Kiefer, E.W. Knapp, 'The Interdependence of Small Linewidth and Frequency Changes of a Vibration in Liquid Mixtures of $CDCl_3$ and CCl_4', *J. Chem. Phys.*, **81**(9), 3774–3778 (1984).
77. W. Kiefer, J.A. Topp, 'A Method for Automatic Scanning of the Depolarization Ratio in Raman Spectroscopy', *Appl. Spectrosc.*, **28**(1), 26–34 (1974).
78. W. Klöckner, W. Kiefer, H.W. Schrötter, 'A Method for Direct Scanning of the Purely Isotropic Part of Raman Scattered Light', *Appl. Spectrosc.*, **31**(1), 52–53 (1977).
79. W. Kiefer, W. Nitsch, 'Techniques for Direct Scanning of the Reversal Coefficient in Raman Spectroscopy', *Appl. Spectrosc.*, **31**(1), 50–52 (1977).
80. B.P. Asthana, W. Kiefer, 'Vibrational Line Profile and Frequency Shift Studies by Raman Spectroscopy', in *Vibrational Spectra and Structure*, ed. J.R. Durig, Elsevier Science, Amsterdam, 67–155, Vol. 20, 1992.
81. H. Eichele, G. Kolb, J. Laane, A. Beckmann, W. Kiefer, 'A Four-channel Photon-counter for Raman Spectroscopy', *J. Phys. E: Sci. Instrum.*, **14**, 1397–1403 (1981).
82. J. Laane, W. Kiefer, 'Application of Four-channel Raman Difference Spectroscopy', *Appl. Spectrosc.*, **35**(4), 428–432 (1981).
83. K. Kamogawa, T. Kitagawa, 'A New Device for Raman Difference Spectroscopy and its Application to Observe Frequency Shifts Due to Isotope Mixing', *J. Phys. Chem.*, **94**(10), 3916–3921 (1990).
84. V. Deckert, W. Liebler, R. Eck, W. Kiefer, 'New Device for Raman Difference Spectroscopy with Multichannel and Scanning Multichannel Detection', *Appl. Spectrosc.*, **51**(7), 939–943 (1997).
85. P. Knoll, R. Singer, W. Kiefer, 'Improving Spectroscopic Techniques by a Scanning Multichannel Method', *Appl. Spectrosc.*, **44**(5), 776–782 (1990).
86. V. Deckert, W. Kiefer, 'Scanning Multichannel Technique for Improved Spectrochemical Measurements with the CCD Camera and its Application to Raman Spectroscopy', *Appl. Spectrosc.*, **46**(2), 322–328 (1992).
87. J.C. Martin, 'Rotating Mirror Device for Nonresonant Laser Difference Spectroscopy', *Rev. Sci. Instrum.*, **5612**(), 2217–2221 (1985).
88. J.F. Eng, R.S. Czernuszewics, T.G. Spiro, 'Raman Difference Spectroscopy via Backscattering from a Spinning Tube and from a Low-temperature Tuning Fork', *J. Raman Spectrosc.*, **16**(6), 432–437 (1985).
89. R. Savoie, S. Alex, M. Langlais, 'Multisample System for Raman Difference Spectroscopy', *Appl. Spectrosc.*, **40**(4), 553–555 (1986).
90. M.A. Washington, A.Z. Genack, H.Z. Cummins, R.H. Bruce, 'Spectroscopy of Excited Yellow Exciton States in Cu_2O by Forbidden Resonant Raman Scattering', *Phys. Rev.*, **15**(B4), 2145–2153 (1977).
91. L.D. Ziegler, J.L. Roebber, 'Resonance Hyper-Raman Scattering Spectra of Ammonia', *Chem. Phys. Lett.*, **136**(5), 377–382 (1987).
92. D.V. Murphy, K.U. von Raben, R.K. Chang, P.B. Dorain, 'Surface-enhanced Hyper-Raman Scattering from SO_3^{2-} Adsorbed on Ag Powder', *Chem. Phys. Lett.*, **85**(1), 43–47 (1982).
93. J.T. Golab, J.R. Sprague, K.T. Karron, G.C. Schatz, R.P. Van Duyne, 'A Surface Enhanced Hyper-Raman Scattering Study of Pyridine Adsorbed Onto Silver – Experiment and Theory', *J. Chem. Phys.*, **88**(12), 7942–7951 (1988).
94. S. Nie, L.A. Lipscomb, S. Feng, N.T. Yu, 'Resonant and Nonresonant Surface-enhanced Hyper-Raman Spectroscopy with a Picosecond Laser-effect of the Excitation Pulse Width', *Chem. Phys. Lett.*, **167**(1–2), 35–40 (1990).
95. J.F. Verdieck, S.H. Peterson, C.M. Savage, P.D. Maker, 'Hyper-Raman Spectra of Methane, Ethane, and Ethylene in Gas Phase', *Chem. Phys. Lett.*, **7**(2), 219–222 (1970).
96. C.M. Savage, P.D. Maker, 'Multichannel Photon Counting Spectrographic Detector System', *Appl. Opt.*, **10**(4), 965–968 (1971).
97. M.J. French, D.A. Long, 'Versatile Computer-controlled Spectrometer for Hyper Rayleigh and Hyper-Raman Spectroscopy', *J. Raman Spectrosc.*, **3**(4), 391–406 (1975).
98. H. Vogt, 'The Hyper-Raman Spectrum of CsI', in *Proc. of 5th Int. Conf. on Raman Spectr.*, eds. E.D. Schmid, J. Brandmüller, W. Kiefer, B. Schrader, H.W. Schrötter, Schulz, Freiburg, 742–743, 1976.
99. W.J. Schmid, H.W. Schrötter, 'Direct Observation of the Twisting Vibration in Tetrachloroethylene Through the Hyper-Raman Effect', *Chem. Phys. Lett.*, **45**(3), 502–503 (1977).
100. H. Vogt, 'Absolute Cross-sections for Hyper-Raman Scattering by Soft Modes in Oxygen-octahedra Ferroelectrics', *Phys. Rev. Lett.*, **36**(B9), 5001–5009 (1987).
101. W.P. Acker, D.H. Leach, R.K. Chang, 'Stokes and Anti-Stokes Hyper-Raman Scattering from Benzene, Deuterated Benzene, and Carbon-tetrachloride', *Chem. Phys. Lett.*, **155**(4/5), 491–495 (1989).
102. N.T. Yu, S. Nie, 'Surface-enhanced Hyper-Raman and FT-Raman Spectroscopy', in *Proc. of XIIIth Int. Conf. on Raman Spectr.*, eds. W. Kiefer, M. Cardona, G. Schaack, F.W. Schneider, H.W. Schrötter, Wiley, Chichester, 37–40, 1992.

103. J.W. Nibler, G.V. Knighten, 'Coherent Anti-Stokes Raman Spectroscopy', in *Topics in Current Physics*, ed. A. Weber, Springer, Berlin, 253–299, Vol. 11, 1978.
104. A. Materny, M. Leuchs, T. Michelis, K. Schaschek, W. Kiefer, 'Resonance CARS Spectroscopy on Diacetylene Single Crystals', *J. Raman Spectr.*, **23**, 99–106 (1992).
105. A. Materny, W. Kiefer, M. Schwoerer, 'Absorption, Luminescence, Resonance Raman, and Resonance Coherent Anti-Stokes Raman-spectroscopy on Unsymmetrically Substituted Diacetylene Single-crystals with Color Zones', *J. Chem. Phys.*, **97**(4), 2237–2246 (1992).
106. A. Materny, W. Kiefer, 'Theoretical-analysis of the Resonance Raman-spectra of Diacetylene Single-crystals', *J. Chem. Phys.*, **97**(2), 841–851 (1992).
107. A. Materny, W. Kiefer, 'Theoretical-analysis of the Resonance CARS Spectra of Diacetylene Single-crystals', *J. Raman Spectrosc.*, **23**(11), 595–603 (1992).
108. M. Leuchs, W. Kiefer, 'Excitation Profile of Coherent Anti-Stokes–Raman Scattering from the MnO_4^- Ion Doped in a $KClO_4$ Crystal at Low-temperature', *J. Chem. Phys.*, **99**(9), 6302–6307 (1993).
109. A. Weippert, W. Kiefer, 'CARS Difference Spectroscopy. 1. Theory for Binary-mixtures', *J. Raman Spectrosc.*, **23**(12), 707–712 (1993).
110. A. Weippert, W. Kiefer, 'CARS Difference Spectroscopy. 2. Determination of Small Frequency-shifts and Band-width Changes in Binary-mixtures', *J. Raman Spectrosc.*, **23**(12), 713–720 (1993).
111. A. Weippert, W. Kiefer, 'CARS Difference Spectroscopy', *Il Nuovo Cimento*, **14D**(10), 1033–1042 (1992).
112. A. Weippert, J.M. Funk, A. Materny, W. Kiefer, 'Experimental-verification of Backscattering CARS Spectroscopy on Opaque Media', *J. Raman Spectrosc.*, **24**(10), 705–708 (1993).
113. M. Pfeiffer, A. Lau, W. Werncke, 'Comparison of Different Phase Conjugate and Backscattering Arrangements in Non-linear Raman Spectroscopy', *J. Raman Spectrosc.*, **21**(12), 835–839 (1990).
114. J.W. Nibler, W.M. Schaub, J.R. MacDonald, A.B. Harvey, 'Coherent Anti-Stokes Raman Spectroscopy', in *Vibrational Spectra and Structure*, ed. J.R. Durig, Elsevier, Amsterdam, Vol. 6, Chapter 3, 1977.
115. A. Laubereau, W. Kaiser, 'Vibrational Dynamics of Liquids and Solids Investigated by Picosecond Light Pulses', *Rev. Mod. Phys.*, **50**(3), 607–665 (1978).
116. K. Duppen, D.P. Weitekamp, D.A. Wiersma, 'Vibrational Dephasing in Molecular Mixed Crystals: a Picosecond Time Domain CARS Study of Pentacene in Naphthalene and Benzoic Acid', *J. Chem. Phys.*, **79**(12), 5835–5844 (1983).
117. S.A. Akhmanov, N.I. Koroteev, S.A. Magintskii, V.B. Morozov, A.P. Tarasevich, V.G. Tunkin, 'Time-domain Coherent Active Raman-spectroscopy of a Free-nitrogen Jet', *J. Opt. Soc. Am.*, **2B**(4), 640–648 (1985).
118. L. Angeloni, R. Righini, E. Castellucci, P. Foggi, S. Califano, 'Temperature-dependent Decay of Vibrational Excitons in K_2SO_4 Crystal Measured by Picosecond Time-resolved CARS', *J. Phys. Chem.*, **92**(4), 983–988 (1988).
119. A. Laubereau, G. Wochner, W. Kaiser, 'Vibrational Dephasing Time of a Molecular Subgroup Investigated by Ultrashort Coherent Probing Techniques', *Chem. Phys.*, **28**(3), 363–369 (1978).
120. N. Kohles, A. Laubereau, 'Investigation of the Rotational Dynamics in Liquids by Time-resolved CARS', *Chem. Phys. Lett.*, **138**(4), 365–370 (1987).
121. P.M. Weber, S.A. Rice, 'Ensemble Dephasing in Vibrationally Excited Jet-cooled Tetrazine and its Complexes with Ar, Kr, and Xe', *J. Chem. Phys.*, **88**(10), 6107–6119 (1988).
122. W.E. Bron, T. Juhasz, S. Mehta, 'New Nonequilibrium Phonon State', *Phys. Rev. Lett.*, **62**(14), 1655–1658 (1989).
123. W. Zinth, W. Kaiser, 'Ultrafast Coherent Spectroscopy', in *Ultrashort Laser Pulses and Applications*, ed. W. Kaiser, Springer, Berlin, 235–278, Topics in Applied Physics, 60, 1988.
124. W. Zinth, R. Leonhardt, W. Holzapfel, W. Kaiser, 'Fast Dephasing Processes Studied with a Femtosecond Coherent Raman System', *IEEE J. Quantum Electron.*, **24**(2), 455–459 (1988).
125. H. Okamoto, K. Yoshihara, 'Femtosecond Time-resolved Coherent Raman-scattering Under Various Polarization and Resonance Conditions', *J. Opt. Soc. Am.*, **7B**(8), 1702–1708 (1990).
126. R. Inaba, H. Okamoto, K. Yoshihara, M. Tasumi, 'Femtosecond Time-resolved Coherent Anti-Stokes–Raman Scattering from Acetonitrile – Solvent Effects on the Vibrational Dephasing of the CN Stretching Band', *Chem. Phys. Lett.*, **185**(1/2), 56–60 (1991).
127. R. Inaba, H. Okamoto, K. Yoshihara, M. Tasumi, 'Observation of the Dephasing of the CN Stretching Vibration in Liquid Nitriles by Femtosecond Time-resolved Coherent Anti-Stokes–Raman Scattering', *J. Phys. Chem.*, **96**(21), 8385–8390 (1992).
128. R. Inaba, H. Okamoto, K. Yoshihara, M. Tasumi, 'Femtosecond Time Resolved Coherent Anti-Stokes Raman Scattering of the C≡C Stretching in Liquid Alkynes', *J. Phys. Chem.*, **97**(30), 7815–7819 (1993).
129. H. Okamoto, R. Inaba, M. Tasumi, K. Yoshihara, 'Femtosecond Vibrational Dephasing of the CN Stretching in Alkanenitriles with Long Alkyl Chains – Dependence on the Chain-length and Hydrogen-bonding', *Chem. Phys. Lett.*, **206**(1–4), 388–392 (1993).
130. H. Okamoto, R. Inaba, K. Yoshihara, M. Tasumi, 'Femtosecond Time-resolved Polarized Coherent Anti-Stokes Raman Studies on Reorientational Relaxation in Benzonitrile', *Chem. Phys. Lett.*, **202**(1–2), 161–166 (1993).
131. P. Aechtner, M. Fickenscher, A. Laubereau, 'Vibrational Dephasing of Simple Liquids Investigated by Picosecond and Femtosecond CARS', *Ber. Bunsen Ges.*, **94**(3), 399–404 (1990).

132. A. Laubereau, H.G. Purucker, 'Time Domain Coherent Raman Techniques', *Il Nuovo Cimento*, **14D**(10), 979–988 (1992).
133. W. Li, H.G. Purucker, A. Laubereau, 'Polarization Interference in Femtosecond CARS', *Opt. Commun.*, **94**(4), 300–308 (1992).
134. M. Fickenscher, H.G. Purucker, A. Laubereau, 'Resonant Vibrational Dephasing Investigated by High-precision Femtosecond CARS', *Chem. Phys. Lett.*, **191**(1/2), 182–188 (1992).
135. H.G. Purucker, V. Tunkin, A. Laubereau, 'Femtosecond 3-Color CARS of Liquids with Magic Polarization Conditions', *J. Raman Spectrosc.*, **24**(7), 453–458 (1993).
136. S. Maeda, T. Kamisuki, Y. Adachi, 'Condensed Phase CARS', in *Advances in Non-linear Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Wiley, Chichester, 253–297, Vol. 15, 1988.
137. M. Schmitt, G. Knopp, A. Materny, W. Kiefer, 'The Application of Femtosecond Time-resolved Coherent Anti-Stokes Raman Scattering for the Investigation of Ground and Excited State Molecular Dynamics of Molecules in the Gas Phase', *J. Phys. Chem.*, **A102**, 4059–4065 (1997).
138. J.J. Barrett, R.F. Begley, 'Low-power CW Generation of Coherent Anti-Stokes Raman Radiation in CH_4 Gas', *Appl. Phys. Lett.*, **27**(3), 129–131 (1975).
139. M.A. Henesian, L. Kulevskii, R.L. Byer, 'CW High Resolution CARS Spectrocsopy of the Q(v_1) Raman Line of Methane', *J. Chem. Phys.*, **65**(12), 5530–5531 (1976).
140. V.I. Fabelinsky, B.B. Krynetsky, L.A. Kulevsky, V.A. Mishin, A.M. Prokhorov, A.D. Savelév, V.V. Smirnov, 'High Resolution CW CARS Spectroscopy of the Q-Branch of the v_2-band in acetylene', *Opt. Commun.*, **20**(3), 389–391 (1977).
141. H.W. Schrötter, H. Frunder, H. Berger, J.P. Boquillon, B. Lavorel, G. Millot, 'High Resolution CARS and Inverse Raman Spectroscopy', in *Advances in Non-linear Spectroscopy*, eds. R.J.H. Clark, R.E. Hester, Wiley, Chichester, 97–148, Vol. 15, 1988.
142. H.W. Schrötter, J.P. Boquillon, 'High Resolution Coherent Anti-Stokes Raman Spectroscopy of Small Molecules', in *Recent Trends in Raman Spectroscopy*, eds. S.B. Banerjee, S.S. Jha, World Scientific, Singapore, 113–126, 1989.
143. H. Frunder, L. Matziol, H. Finsterhölzl, H.W. Schrötter, 'CARS Spectrometer with CW Intra-cavity Excitation for High-resolution Raman Spectroscopy', *J. Raman Spectrosc.*, **17**(1), 143–150 (1986).
144. Th.S. Bican, J. Jonuscheit, U. Lehner, H.W. Schrötter, 'High-resolution Coherent Anti-Stokes Raman Spectroscopy of Ethane', *J. Raman Spectrosc.*, **26**, 699–705 (1995).
145. A. Owyoung, 'High Resolution Coherent Raman Spectroscopy of Gases', in *Laser Spectroscopy IV*, eds. H. Walther, K.W. Rothe, Springer Series in Optical Sciences, Springer, Berlin, 175–187, Vol. 21, 1979.
146. A. Owyoung, 'CW Stimulated Raman Spectroscopy', in *Chemical Applications of Nonlinear Raman Spectroscopy*, ed. A.B. Harvey, Academic Press, New York, 281–320, 1981.
147. D. Bermejo, J. Santos, P. Cancio, J.L. Doménech, C. Domingo, J. Orza, J. Ortigoso, R. Escribano, 'High-resolution Quasi-continuous Wave Inverse Raman Spectrometer – Spectrum of CH_3D in the C–D Stretching Region', *J. Raman Spectrosc.*, **21**(3), 197–201 (1990).
148. A.S. Pine, 'High-resolution Methane v_3-Band Spectra using Stabilized Tunable Difference-frequency Laser System', *J. Opt. Soc. Am.*, **66**, 97–108 (1976).
149. J.J. Barrett, M.J. Berry, 'Photoacoustic Raman Spectroscopy (PARS) using CW Laser Sources', *Appl. Phys. Lett.*, **34**(2), 144–166 (1979).
150. J.J. Barrett, D.F. Heller, 'Theory of Photoacoustic Raman Spectroscopy', in *Nonlinear Raman Spectroscopy and its Chemical Applications*, eds. W. Kiefer, D.A. Long, Reidel, Dordrecht, 563–583, 1982.
151. J.J. Barrett, D.R. Siebert, G.A. West, 'Applications of Photoacoustic Raman Spectroscopy', in *Nonlinear Raman Spectroscopy and its Chemical Applications*, eds. W. Kiefer, D.A. Long, Reidel, Dordrecht, 585–602, 1982.
152. A.M. Brodnikowskii, V.P. Zharov, N.I. Koroteev, 'Principles for Selective Optico-acoustic Compound Detection in Complex Mixtures', *J. Anal. Chem. USSR*, **39**(5), 618–627 (1984).
153. A.B. Duval, D.A. King, R. Haines, N.R. Isenor, B.J. Orr, 'Coherent Raman-Spectroscopy of Glyoxal Vapor', *J. Raman. Spectrosc.*, **17**(2), 177–182 (1986).
154. M. Rotger, B. Lavorel, R. Chaux, 'High-resolution Photoacoustic Raman-spectroscopy of Gases', *J. Raman Spectrosc.*, **23**(5), 303–309 (1992).
155. G. Millot, B. Lavorel, R. Chaux, R. Saint-Loup, G. Pierre, H. Berger, J.I. Steinfeld, B. Foy, 'High-resolution Stimulated Raman-spectroscopy of Methane $^{13}CD_4$ in the Pentad Region', *J. Mol. Spectrosc.*, **127**(1), 156–177 (1988).
156. P. Esherick, A. Owyoung, J. Pliva, 'Ionization-detected Raman Studies of the 1600 cm^{-1} Fermi Dyad of Benzene', *J. Chem. Phys.*, **83**(7), 3311–3317 (1985).
157. M. Leuchs, M. Crew, J. Harrison, M.F. Hineman, J.W. Nibler, 'High-resolution Ionization-detected Raman Gain Spectroscopy of N_2 and C_6H_6', *J. Chem. Phys.*, **105**, 4885–4888 (1996).
158. B.F. Henson, G.V. Hartland, P.M. Felker, V.A. Venturo, 'Time-resolved Raman Spectroscopy of Benzene Dimer – Nanosecond-plus Lifetime at 992 cm^{-1} Vibrational Energy', *J. Chem. Phys.*, **91**(4), 2751–2753 (1989).
159. G.V. Hartland, B.F. Henson, V.A. Venturo, R.A. Hertz, P. Felker, 'Applications of Ionization-detected Stimulated

Raman Spectroscopy in Molecular-beam Studies', *J. Opt. Soc. Am.*, **7B**(9), 1950–1959 (1990).

160. B.F. Henson, G.V. Hartland, V.A. Venturo, R.A. Hertz, P.M. Felker, 'Stimulated Raman-spectroscopy in the ν_1 Region of Isotopically Substituted Benzene Dimers – Evidence for Symmetrically Inequivalent Benzene Moieties', *Chem. Phys. Lett.*, **176**(1), 91–98 (1991).

Thermal Analysis: Introduction

David Dollimore
The University of Toledo, Toledo, USA

1 **Thermal Analysis** 1
2 **Terminology** 1
3 **The Advent of the Computer** 2
4 **Procedural Variables in Thermal Analysis** 2
5 **The Sample Holder** 2
6 **The Rate of Heating** 2
7 **The Atmosphere Surrounding the Sample** 2
8 **The Mass of the Sample** 3
9 **Conclusions** 3
Abbreviations and Acronyms 3

This is a very short article to introduce the subject of thermal analysis and to indicate how the following reviews fall into the general scheme of things. These more detailed reviews on the various topics should be read carefully, as they do cover the subject in depth. In each review a list of references is provided, which should allow the more serious reader to follow the development of each subject by reference to original research articles.

1 THERMAL ANALYSIS

This term covers the use of a variety of techniques to follow the change in any property of a system or sample as the system is subjected to a programmed range of temperatures. The use of such techniques is to be found in many industries. The program of temperature change is generally one of changing the temperature linearly with time. This is the program selected by most of the industrial users as it saves time and hence money to get information in this way. It must be emphasized that the measurement of the property is recorded at the particular temperature indicated and not the property of the system that has been bought back to ambient temperature.

In the reviews that follow, two main thermal analysis techniques are described, namely **Thermogravimetry** and **Differential Scanning Calorimetry and Differential Thermal Analysis**. These need to be defined in order to appreciate the instrument and the experimentation. Thermogravimetry (TG) is possibly the most widely used technique in thermal analysis. It involves the measurement of the mass of a substance as a function of temperature while the substance is subjected to a controlled temperature program. The other technique noted above is calorimetry but goes by two different names, which at first glance is rather confusing. Differential thermal analysis (DTA) is a technique in which the temperature difference between a substance and a reference material is measured as a function of temperature while the system is subjected to a controlled temperature program. Differential scanning calorimetry (DSC) is a technique in which the difference in heat flow to a sample and to a reference is monitored against temperature while the entire system is subjected to a controlled temperature program. Put boldly, if the system operates qualitatively, merely indicating an endothermic or exothermic event, then it will be termed DTA by the instrument manufacturers. If on the other hand the instrument performs as a calorimeter and indicates the enthalpy change for an event, then it termed a differential scanning calorimeter. It is often advantageous to combine two or more of these thermal analysis techniques so that two probing units, e.g. measurement of mass (TG) and changes in heat capacity of the system (DTA or DSC), operate on the same sample. Such a combination goes under the name of **Simultaneous Techniques in Thermal Analysis**. Gas analysis performed when the sample under investigation is subjected to a rising temperature program is seldom used as a single technique but is more often used in conjunction with another thermal analysis technique in a simultaneous fashion. The techniques of thermal analysis are to be found in the literature applied in all fields of science. As an example, the application of such techniques to inorganic chemistry is outlined here as one of the subjects reviewed (see **Inorganic Systems, Thermal Analysis Applications to**).

2 TERMINOLOGY

It has been found necessary to use a common terminology and methods of reporting data obtained on thermal analysis instruments. The International Confederation for Thermal Analysis and Calorimetry (ICTAC) has produced publications and recommendations, on the nomenclature to be used, on calibration methods, and the symbols used to represent the techniques utilized. The symbols and definitions used above are the recommended ICTAC definitions currently in practice at the time of writing. However, the definitions and the abbreviations are constantly under review. New techniques or modifications of older techniques require that this process of a common technology be re-examined.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

3 THE ADVENT OF THE COMPUTER

The computer forms an essential part of any thermal analysis unit. It can be used for instructional purposes. It is also used to control the equipment, i.e. to program the temperature regime, and to record the properties that are being measured. It is used to collect this data for storage, and retrieval as the occasion demands. The computer can be ''loaded'' with programs that allow for an interpretation of the events recorded. Basic data collected from thermal analysis experiments can now be processed in a very short time. These computations would previously have taken up much valuable time as they would have involved manual calculations. Finally, the computer can be used to simulate the behavior of the instrument during a virtual experiment. Thus by inserting the values of the pre-exponential term, the activation energy, and the type of kinetic equation, into the computer, a theoretical simulation of thermal analysis experiments can be achieved. It is also possible to run an electronic library of past thermal analysis experiments for comparison purposes.

4 PROCEDURAL VARIABLES IN THERMAL ANALYSIS

It must be recognized that there are procedural variables that might affect the results of any thermal analysis experiment. Thus a large chunk of wood is logically expected to behave differently to sawdust although the two are probably chemically related. A rock of limestone would give a different result from a ground sample of the same material.

The sample under investigation must be adequately described. The sample is generally a solid and it must be remembered that, alone among the phases, the solid phase to be adequately described requires a detailed record of its past history. Also important are the purity of the sample and its chemical composition, and if possible there should be an attempt at giving its formula. In some instances the sample may be present in a diluted form.

5 THE SAMPLE HOLDER

The geometry and the material used in the construction of the sample holder are important. The sample holder should not interact with the sample during the heating regime. The different thermal conductivity values of the materials used in the sample holder can cause alterations in the heat transfer, which can be apparent in the final thermal analysis plot. The geometric shape of the holder can cause a reaction to change from, for example, a predominantly diffusion mechanism in a narrow holder, to a reaction interface process in a shallow holder.

6 THE RATE OF HEATING

The rate of heating can vary from an isothermal experiment to a heating rate in which several hundred degrees are scanned in a very short time. The response of the sample to such changes can be drastic, and experience and theory have made investigators aware of this fact. This comment also applies to those thermal analysis experiments where the sample is cooled. Very fast cooling, called quenching, can cause a higher temperature state to be preserved at the lower temperature. Very slow heating, called annealing, is most likely going to allow the higher temperature form to transform to a state stable at the lower temperature. There is also the possibility that the design of the equipment might result in thermal lag between different components of the unit and this could be affected by the rate of heating.

7 THE ATMOSPHERE SURROUNDING THE SAMPLE

In thermal analysis experiments the gaseous environment around the sample is going to influence the results. This follows from a casual glance at any textbook on thermodynamics, and the experimenter in thermal analysis should be aware of such a fact. Carrying out the thermal analysis experiment is possible at pressures ranging from a very high vacuum to several hundred times atmospheric pressure. Not all experiments are performed over such an enormous range, but any transformation studied in thermal analysis experiments would be affected by changes in the pressure of the gaseous environment. In fact most experiments on commercial units are studied at atmospheric pressure. These comments apply to both physical transformations and chemical changes. At atmospheric pressure (which is the region where most experiments are run), the nature of the environmental gas and the rate at which the gas is passed over the sample are important. Such features as the thermal conductivity of the gas may affect the temperature range over which the transformation may be noted. If the partial pressure of the product gas over the sample is altered during the thermal analysis experiment, then the temperature range of the transformation may be altered in conformity with Le Chatelier's Principle. The decomposition of calcium carbonate is a case that may be cited. Increasing the partial

pressure of carbon dioxide around the sample will push the decomposition range into higher temperature regions.

8 THE MASS OF THE SAMPLE

The mass of the sample in thermal analysis experiments, and the manner of its distribution, may affect the result. The sample may be present as a fine powder or as a compressed mass (as in a pharmaceutical tablet). This may be summed up, by simply stating that the manner in which the sample is packed is an important factor. It is also a wise precaution to sieve the sample and to investigate the decomposition of the sieved ranges. Again the decomposition of calcium carbonate serves as an example. The small particle size fractions possess a larger surface area than the larger particle size fractions. Hence, the kinetics of the sample can be affected by the initial surface area of the material, and the smaller size fractions will possess a greater chemical reactivity and will decompose in a lower temperature range.

9 CONCLUSIONS

It must be emphasized that the points made above may be substantiated by logical and theoretical reasoning. However, this is no substitute for carrying out actual experiments, as some systems may be more complex than is at first apparent. In such instances some careful experimentation is probably worth more than a simple exercise of theory. Theory has to be based on experiments and not the other way around. The literature on thermal analysis is constantly expanding. *Chemical Abstracts* run a *CA Selects* on thermal analysis. There are several textbooks on the subject, but many of those quoted in review texts of this kind are out of date and not so easily found. It is best to consult the nearest library in order to make an up-to-date choice. *Analytical Chemistry* runs a two-yearly review of the subject, which may prove to be helpful.

ABBREVIATIONS AND ACRONYMS

DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
TG	Thermogravimetry

Differential Scanning Calorimetry and Differential Thermal Analysis

Alan Riga and Ricardo Collins
Cleveland State University, c/o Techon Inc., 6325 Aldenham Dr., Cleveland, USA

1 Introduction **2**
1.1 Differential Scanning Calorimetry: Heat Flux and Power Compensated Methods 2
1.2 Review of Modulated Temperature Differential Scanning Calorimetry 4

2 Experimental Procedures **4**
2.1 Optimization of Differential Scanning Calorimetry Thermal Properties 4
2.2 Calibration, Standards and Standard Methods 4
2.3 Sample Size 5
2.4 Sample Preparation and Pan Selection 6
2.5 Differential Scanning Calorimetry Curve and Sample Analysis 6
2.6 Effect of Sample Specific Heat 7
2.7 Differential Scanning Calorimetry Sensitivity and Resolution 8
2.8 Pressure Differential Scanning Calorimetry 8
2.9 Differential Scanning Calorimetry Method and Thermoplastics 8
2.10 Differential Scanning Calorimetry Method and Thermosets 8
2.11 Differential Scanning Calorimetry Method and an Unknown 9
2.12 Internal Melting Point Standard 9
2.13 Modulated Temperature and Convential Differential Scanning Calorimetry 10
2.14 Examples of Modulated Temperature Differential Scanning Calorimetry Characteristics 10
2.15 Modulated Temperature Differential Scanning Calorimetry Experiments and the Glass Transition 11
2.16 Power Compensated Differential Scanning Calorimetry Experimental Procedures 14
2.17 Quantitative Differential Scanning Calorimetry and Differential Thermal Analysis 16
2.18 Low-temperature Quantitative Analysis 16
2.19 Specific Heat Capacity by Differential Scanning Calorimetry and Modulated Temperature Differential Scanning Calorimetry 18

3 Applications **18**
3.1 Alkane Structure–Property Relationships by Differential Scanning Calorimetry, Thermal Microscopy and X-ray Diffraction Analysis 18
3.2 Oxidative Behavior of Materials by Standard and Pressure Differential Scanning Calorimetry: Polymers and Engine Oils 18
3.3 Inorganic Reactor Deposits by Differential Thermal Analysis, Thermogravimetric Analysis and X-ray Diffraction Analysis 20
3.4 Polymer Characterization by Thermogravimetric Analysis, Differential Thermal Analysis/Differential Scanning Calorimetry, Thermomechanical Analysis, Fourier Transform Infrared and X-ray Diffraction Analysis 22
3.5 Assignment of the Glass Transition Temperature 23
3.6 Copolymer T_g and Molecular Weight 23
3.7 Polymer T_g and Moisture 23
3.8 Effect of High-pressure Gases on the Polymer T_g 23
3.9 Polymer T_g, Crystallization and Fusion 24
3.10 Liquid Crystal Polymer T_g and Internal Reference 24
3.11 Polymer Film T_g and Thermal History 24
3.12 Effect of Inhibitors on Vinyl Monomer Polymerization 24
3.13 Differential Scanning Calorimetry "Fingerprint" Curve Identifies Unknown 24
3.14 Curing and Polymerization of Thermoset Polymers 25
3.15 Polymer Fabrics Identified by Differential Scanning Calorimetry 26

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

3.16 Pharmaceutical Applications: Purity, Polymorphism, Excipients and Lubrication 26
3.17 Food Application: Cocoa Butter 27
3.18 Biochemical Application: Ribonuclease 28
3.19 Hazardous Material: Aromatic Nitro Compound 28
3.20 Characterization of Inorganic Compounds 28
4 Summary **29**
5 Resources **29**
6 American Society for Testing and Materials Methods Used in Differential Scanning Calorimetry and Differential Thermal Analysis **29**
Abbreviations and Acronyms **31**
Related Article **31**
References **31**

The importance of thermal analysis methods in thermal and material science has proved awesome. Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) have had a great impact on material science by enabling the measurement of a great number of physical and chemical properties. These techniques have allowed elucidation of endothermic and exothermic processes at temperatures ranging from very low (−120°C) to high (DSC 600°C and DTA 1800°C).[1–6] Some of the physical properties measured by these thermal analytical methods include melting transitions, crystallization temperatures, enthalpy of fusion and crystallization, specific heat capacity, liquid crystal transitions, vaporization, sublimation, solid–solid transitions, thermal conductivity and the glass transition temperature.[7–14] Chemical properties monitored by DSC and DTA are dehydration, decomposition, oxidative reactions, solid state reactions, chemisorption, combustion, polymerization, curing and catalyzed reactions.[15–19]

In DTA, the temperature difference between the substance and a reference material is measured as a function of temperature, while the substance and reference material are subjected to a controlled temperature program.[20]

In DSC, the difference in energy input into a substance and a reference material is measured as a function of temperature, while the substance and reference material are subjected to a controlled temperature program.[21] Two modes, power compensation DSC and heat flux DSC, can be distinguished, depending on which method of measurement is used.[22,23]

The following is a review of DSC and DTA and their role in thermal science. The root of thermal analysis is the determination of heat and temperature. Heat is a macroscopic quantity and has its origin in molecular motion.[24] DTA measures the temperature difference between a sensitive reactive material and an inert reference as a function of time and temperature. DTA measurements provide information about transition temperatures as well as kinetic time-based phenomena and thermodynamic properties. DSC is a step above DTA in that it measures quantitative heats of reaction and transitions. Both thermal methods can differentiate first order thermodynamic changes like fusion, and second order thermodynamic changes like a glass transition temperature, T_g. This chapter will focus mainly on fundamentals, instrumentation and applications of DSC.

1 INTRODUCTION

1.1 Differential Scanning Calorimetry: Heat Flux and Power Compensated Methods

The function of DTA and DSC instruments and the values they measure will be reviewed. An important innovation of DSC, modulated temperature differential scanning calorimetry (MTDSC), will also be examined in detail.

Some of the first commercial analytical instruments that measure heat flow into or out of a sample as the sample is undergoing a transition have been referred to by a variety of names. They include quantitative differential thermal analysis (QDTA),[25] dynamic differential calorimetry (DDC), dynamic enthalpic analysis (DEA), and DSC. The latter, DSC, is widely accepted as the term used to describe modern instruments that are scanning calorimeters. DSC is defined here based on what it measures: heat flow. The International Confederation of Thermal Analysis[26] has adopted this position in describing DSC. The generic definition of DSC is the accepted definition and is based on equivalency of results obtained, general usage in the literature, comparisons of temperature profiles, and comparison to classical calorimetry.

DSC uses a temperature transducer as its primary sensor. In nearly all commercial differential scanning calorimeters, the heat flow output is based on the difference in temperature between sample and reference as recorded by the temperature sensor. The sensors in differential temperature devices are thermocouples or thermopiles. Thermocouples are also used in DTA and therein lies the confusion of calling the DSC, a DTA. However, there are some fundamental differences between these two techniques.

In conventional DTA, the temperature sensor is placed in the sample, while in the DSC, the temperature sensors

are located external to the sample. Boersma[27] states that this experimental condition, external sensors, must be met in order to make calorimetric measurements. Semiquantitative calorimetric measurements have been accomplished when the DTA conditions operate in both modes, first when the sensor is in contact with the sample, and then when the sensor is removed from the sample. The TA Instruments (TAI) high temperature DTA is capable of either mode.[25] Without sample couple liners present, the instrument performs as a classical nonquantitative DTA. With liners, however, accuracy and precision of calorimetric measurements of 5% can be observed. DSC typically provides calorimetric accuracy and precision to better than 1%. It is desirable to have constant calorimetric sensitivity, but it is not a necessary condition for a DSC instrument. Differential scanning calorimeters with constant calorimetric sensitivity where the calibration constant is not a function of temperature have an output directly in heat flow units of milliwatts.

Differential scanning calorimeters operate by one of two methods, differential temperature and temperature servo measurements. An example of the differential temperature instrument is the TAI DSC 910 system. Figure 1 is a cross-sectional view of this DSC cell. The cell uses a constantan disk as its primary means of heat transfer to the sample and reference positions, as one element of the temperature measuring thermoelectric junctions. The experimental sample and reference are placed in containers or pans that rest on raised platforms on the constantan disk. Heat is transferred through the disk and into the sample and reference via the containers. The differential heat flow to the samples and reference is monitored by chromel–constantan area thermocouples formed by the junction of the constantan disk and a chromel disk that covers the underside of each platform. The differential temperature, T, from the sample and reference thermocouples, is fed to a variable high-gain amplifier where the signal is amplified, electronically scaled to read directly in heat flow units and finally displayed on the y-axis. Chromel and alumel wires are connected to the underside of the chromel disk and the resultant chromel–alumel thermocouple is used to monitor the sample temperature directly.

When a sample is programmed at a constant heating rate through its melt, the heat capacity of the sample increases in theory to infinity and the sample temperature becomes invariant. Therefore, a temperature is developed between the sample and the reference temperature which is constantly rising. This temperature signal is proportional to the heat flowing into the sample endothermically as part of the transition. When the amplified and scaled signal presented to the computer in milliwatts is integrated over time, a value for the heat in millijoules associated with the transition is obtained.[24]

The second type of instrument, the temperature servo system, is shown in Figures 2 and 3.[25] In this design, the sample and reference pans are placed in close contact with platinum resistance thermocouples (PRTs) which are used to measure their temperature at any given moment. In addition, the unit has two individual heaters that control the heat flow to these samples.

When the sample is temperature programmed through its melt, the sample temperature becomes essentially invariant, and the difference temperature signal, ΔT, developed between it and the reference is amplified, scaled and displayed as the output from the device, see Figure 3.[25,28] The process continues when the differential

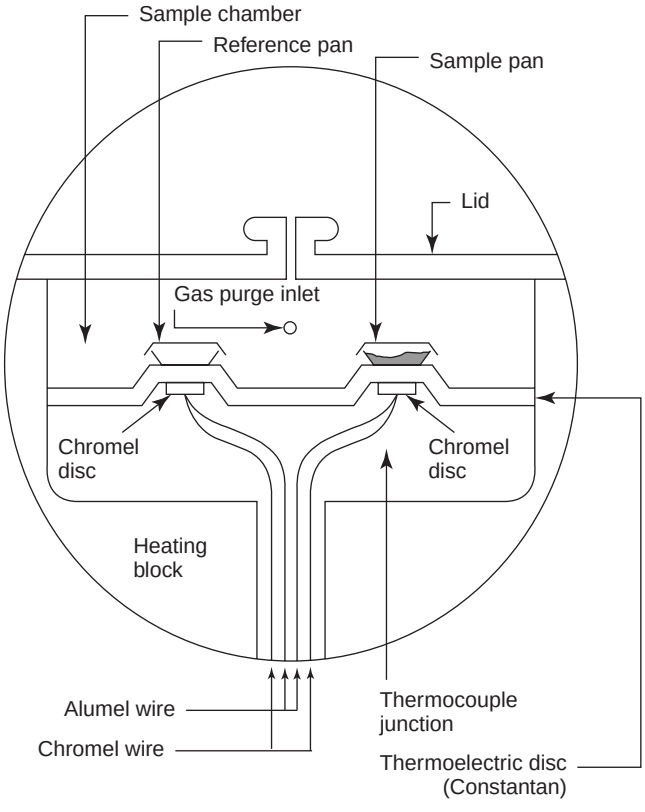

Figure 1 Cross-sectional view of a DSC cell.

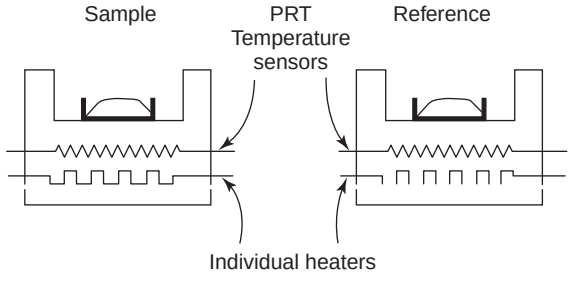

Figure 2 Schematic diagram of power compensated DSC cell.

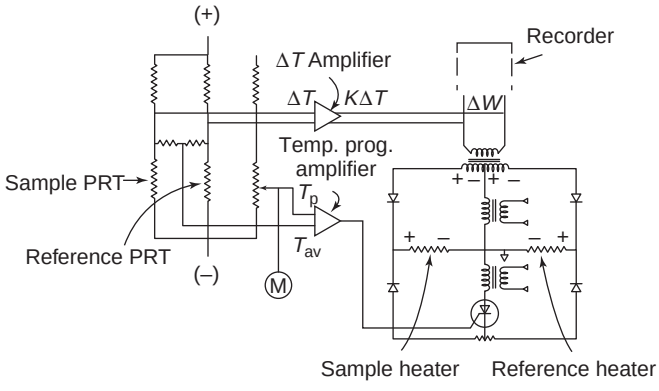

Figure 3 Schematic diagram of temperature servo system.

power is supplied back to the sample heater attempting to reduce the temperature difference between the sample and reference. The operating principle of a servo system requires the restoring signal, in this case a voltage signal to the differential heater, to be proportional to the offset, the temperature difference. If the temperature difference is zero, no differential power could be supplied by the servo system. An actual zero temperature difference can exist only if the servo amplifier gain is infinite; and it is not.

Calorimetry is the scientific field dealing with the measurement of the heat or energy content of various chemical and physical reactions. In a typical classical adiabatic calorimetric experiment, a chemical reaction may be carried out in a sealed container insulated from heat losses to the outside world. From the known heat capacity (C_p) of both the insulated container and its contents and the small measured change in temperature (ΔT) of that system, the heat content of the chemical reaction, H, can be calculated, Equation (1):

$$H = \Delta T \times C_p \tag{1}$$

The experiment is usually designed so that the change in system temperature remains small, usually several Kelvin. Under these conditions C_p is constant. From these experiments we derive the definition of the calorie which is the unit of measurement for heat content or energy, or the amount of heat necessary to raise the temperature of 1 g of water at 15 °C, by 1 °C.

In DSC, the temperature profile of a sample seldom exceeds a few tenths of a degree. Therefore DSC instruments operate well under classical calorimetry conditions, from which thermodynamic data are almost exclusively obtained.[24]

Differential scanning calorimeters based on heat flux or power compensated measures the same quantity, the heat flow into (endothermic) or out of (exothermic) a sample.

1.2 Review of Modulated Temperature Differential Scanning Calorimetry

An innovation in DSC that has become a widely accepted technique is MTDSC.[29–38] It provides the same qualitative and quantitative information about physical and chemical changes as convential DSC. MTDSC, however, provides unique thermochemical data that are not accessible with DSC. MTDSC overcomes the limitations of convential DSC. The effects of baseline slope and curvature are reduced, thereby increasing the sensitivity of the system. Overlapping events such as molecular relaxation and glass transitions can be separated. Heat capacity can be more easily measured using this technique which requires a minimum number of experiments.

MTDSC uses the same convential heat flux DSC cell system. Uniquely, it uses a different heating profile for the sample and reference as supplied by the furnace. Specifically, a sinusoidal modulation or oscillation is superimposed on the convential linear heating ramp. Experimental procedures can vary from method to method for DSC, DTA, MTDSC and pressure differential scanning calorimetry (PDSC).

2 EXPERIMENTAL PROCEDURES

2.1 Optimization of Differential Scanning Calorimetry Thermal Properties

The optimization of DSC is critical in acquiring accurate and precise data. Inspect the DSC cell and assure its good condition. For example, clean the thermal sensors by heating to 500 °C in air to burn off any decomposed materials. Use a brush to wipe off any debris that will interfere with the transfer of heat flow from the sample to the thermal sensor. Always use a purge gas. Typical inert gases are nitrogen and helium. Helium is used to enhance heat transfer. A nominal flow rate is set at 50 mL min^{-1} for most instruments. Oxidative gases are oxygen or air at 21% oxygen and 79% nitrogen.

2.2 Calibration, Standards and Standard Methods

The instrument must be calibrated for temperature and heat flow on a regular basis. American Society for Testing and Materials, ASTM E697[39] can be used to calibrate the DSC temperature (x-axis). ASTM E968[40] is used to calibrate the heat flow (y-axis). The ASTM E1860[41] protocol is used to establish elapsed time, which is especially important when considering International Standardization Organization (ISO) 9000 certification.

An excellent resource of known temperatures and heat of fusion for many standards is given in Table 1,[42] where

Table 1 Enthalpy of melting for standards[a]

Material	Melt temperature (°C)	Heat of fusion		Refs.
		($J\,g^{-1}$)	(+/−)	
Cyclopentane	−151.16 (crystal)	68.69		52
Cyclopentane	−135.06 (crystal)	4.88		52
n-Pentane	−132.66	36.51	0.02	48
n-Heptane	−90.56	138.62		54
Cyclohexane	−87.06	78.70		53
n-Octane	−56.76	180.00		54
Mercury	−38.8344[e]	11.443	0.004	44[i]
n-Decane	−26.66	199.87		54
n-Dodecane	−9.65	214.35		54
H_2O	0.01[e]	335	0.6	43
Benzene	5.53	125.9		47
Cyclohexane	6.54	30.91		53
Diphenyl ether	26.87[d]	101.15	0.10	46
n-Octadecane	28.24	238.76		55
Hexatriacontane	72.14 (crystal)	18.54		55
Hexatriacontane	73.84 (crystal)	59.59		55
Hexatriacontane	75.94	173.38		55
Benzoic acid	122.37[d]	147.4	0.1	45
Indium	156.5985[e]	28.57	0.17	43
Tin	231.928[e]	60.6	0.2	43
Bismuth	271.442[d]	53.07	0.58	43
Cadmium	321.108[d]	55.09	1.4	44
Lead	327.502[d]	23.1	0.3	43
Zinc	419.527[e]	108	0.6	43[g]
Tellurium	449.6	137.0	4	50, 51
Antimony	630.74[d]	163.2	5.2	44[j]
Magnesium	650	362	17	49[h]
Aluminum	660.325[e]	400.1	4.7	44[h]
KBr	734.0	216.0	3.0	51
NaCl	801.0	480.0	10.0	51
Silver	961.78[e]	104.4	3.9	44
Gold	1064.18[e]	63.72	2.1	44
Copper	1084.62[e]	205.4	6.6	44[f,c]
Nickel	1455[d]	297.6		44
Cobalt	1494[d]	274.8	4.3	44[b]
Iron	1538	253	7	49
Palladium	1554[d]	165		

[a] The table summarizes the best known temperatures and heats of melting for many standard materials. These values may be used for temperature and/or cell constant calibration of the TA Instrument's DSC or high temperature DTA.
[b] Sample sublimes.
[c] Sample oxidizes easily.
[d] Temperature taken from the IPTS-1968. *Pure Appl. Chem.*, **22**, 557 (1971).
[e] Temperature taken from the ITS-90. *Metrologia*, **27**, 3–10 (1990).
[f] Reacts with alumina.
[g] Amalgams with aluminum as low as 450 °C, do not heat above 430 °C.
[h] Sample amalgams with platinum.
[i] Shelf-life 1 week at room temperature.
[j] Amalgams with copper.

there are more than 30 standard materials, inorganic, metals and organic chemicals listed. The values cited in Table 1 can be used to establish the DSC cell constants. The latter converts measured values, for example, temperatures and heats of fusion and crystallization to accurate physical properties.

2.3 Sample Size

Thermal properties may be dependent on the specimen size and shape. A common sample size is from 1 to 10 mg. Replicates should be the same size to within 10%. Select the sample shape to ensure good contact

with the pan or container. A uniform size and shape is desirable.

2.4 Sample Preparation and Pan Selection

Heating through transitions erases thermal history at lower temperatures. Mechanical treatment of the sample may produce localized heating or stress in the sample. Cryogenic grinding in a freezer mill creates easily handled powders. The low temperature sampling minimizes treatment effects. Cutting it from the bulk sample with a razor blade or clipper can attain minimum stress on the sample. Minimum temperature and time exposure of the sample is accomplished by the latter techniques. Films can be heat pressed and then paper punched out. However, the effects of heating and oxidation must be considered.

DSC pan or container selection is dependent upon specimen size and shape. One must also consider the maximum temperature of the experiment, if the pan is chemically inert and whether it retains vapors or not. Other factors to consider when selecting an appropriate pan for use is whether it has a high thermal conductivity, low heat capacity, and low cost.

DSC pan types include aluminum, copper, gold, platinum, glass, carbon, stainless steel and mild steel. The form of the pan can also be varied. It can be crimped, hermetically sealed, and contain a molecular leak with a laser drilled pin hole, an autosampler pan and a center raised solid fat index pan to fit over a thermal sensor.

An as-received pan must be cleared in order to remove oils or lubricants used in manufacturing the pan. Organic oils and lubricants can be removed with xylene, acetone, dried with nitrogen gas and stored according to ASTM E-1858-97.[56]

One supplier of aluminum pans used calcium carbonate as lubricant. Excess calcium carbonate confounded oxidative studies where organic oils decomposed to yield acids. The relative oxidative stability varied considerably.[57]

Another vendor did not realize that when their supplier of aluminum pans stamped them out they were embedding small (less than 0.2 mm) particles of iron and copper. These metals are used to accelerate the oxidation of oils.[58,59] These contaminants caused accelerated oxidation and a process that became out of control. The extraneous metals were only detected by scanning electron microscopy (SEM) with energy dispersive X-ray spectroscopy, see Figure 4. Table 2 gives an elemental analysis. When the peak area ratio of iron and copper to chromium was greater than two, the relative oxidation rate (inflection point-first derivative of the exothermic heat flow versus temperature curve) was out-of-control. This particular problem was eliminated by returning to a source of new containers with a high standard of quality control.

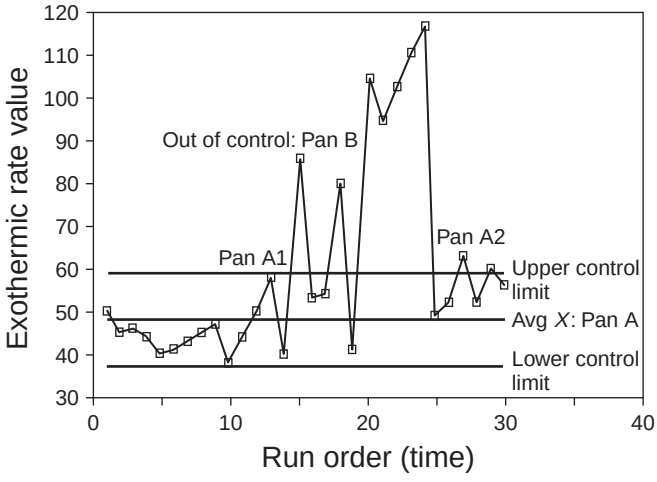

Figure 4 Out-of-control process: exothermic value versus time. , oxidation rate (MW/min); ——, average oxidation rate.

Table 2 Elemental analysis of pan surface

Al pan	Fe	Cr	Cu	Fe/Cr
Poor	414	135	173	3.1
Fair	328	103	136	3.2
Good	315	235	157	1.3

2.5 Differential Scanning Calorimetry Curve and Sample Analysis

A DSC or DTA baseline plot should be flat. The latter can be achieved through calibration with empty pans and application of the electronic correction. All samples have a baseline with an endothermic slope caused by the increasing specific heat capacity of the sample with increasing temperature, see Figure 5.

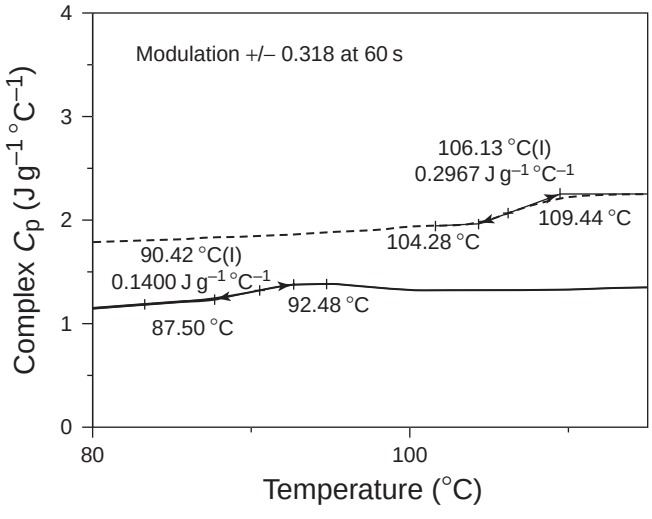

Figure 5 Specific heat capacity increases with temperature. ——, general grade polystyrene (PS) and - - - -, 900 000 MW PS standard.

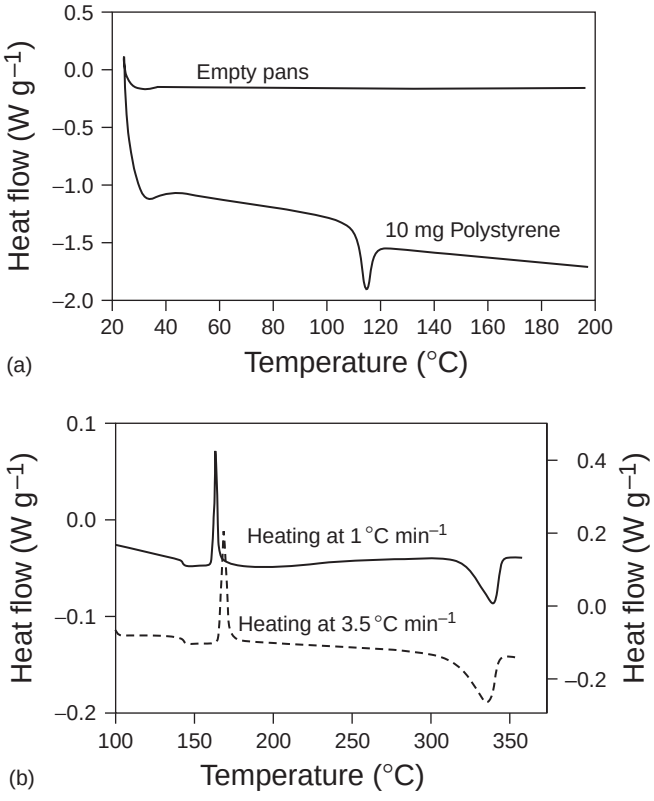

Figure 6 (a) DSC baseline slope. (b) DSC baseline curvature.

DSC baseline curvature cannot normally be corrected by calibration. However, if necessary, it can be eliminated by baseline subtraction, see Figure 6. Baseline slope and curvature can cause an error if integrating peaks over a broad temperature range and can make detection of a weak glass transition temperature, T_g, difficult, Figure 6(a) and (b).

Samples should be kept thin to minimize heat transfer problems. It is better to cut out a representative sample rather than crush it. Stresses could be imparted to the sample prior to thermal analysis. Typical sample mass in DSC is 10–15 mg for polymers. A goal is to achieve a change of 0.1–10-mW heat flow in going through a transition, see Figure 7. Use the lightest, flattest pan possible.

If the sample contains volatiles, put one or more pinholes in the lid of the pan before crimping in order to permit a continuous evaporation process.

2.6 Effect of Sample Specific Heat

Minimize differences in specific heat between the sample and the reference by adding aluminum to the reference pan. This improves the overall baseline; see Figures 8–10. Initiate a DSC analysis 2–8 min before the temperature of interest, depending on the heating rate, for example,

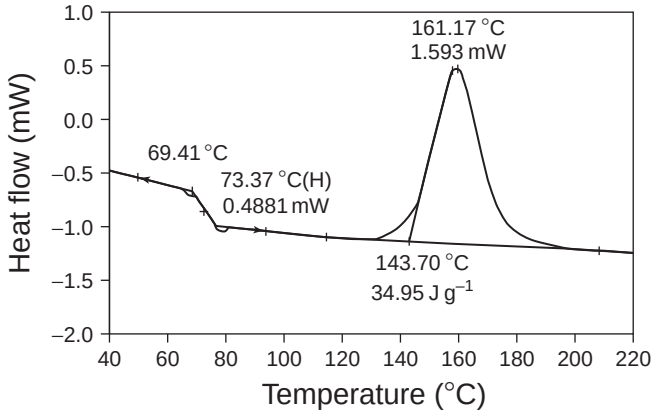

Figure 7 Heat flow change during a transition.

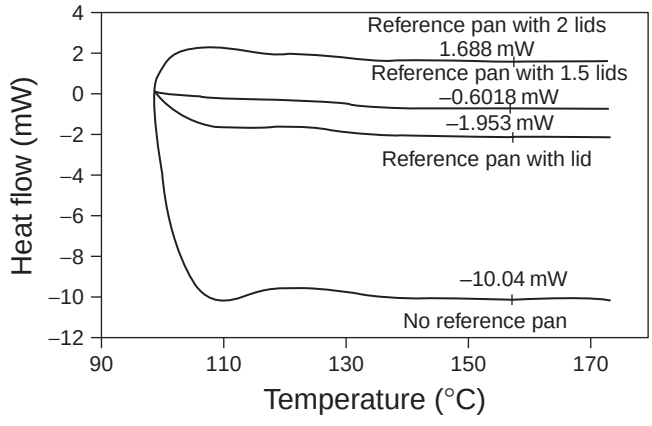

Figure 8 Effect of reference pan weight. Epoxy sample of approx. 10 mg, heating rate 20 °C min^{-1}.

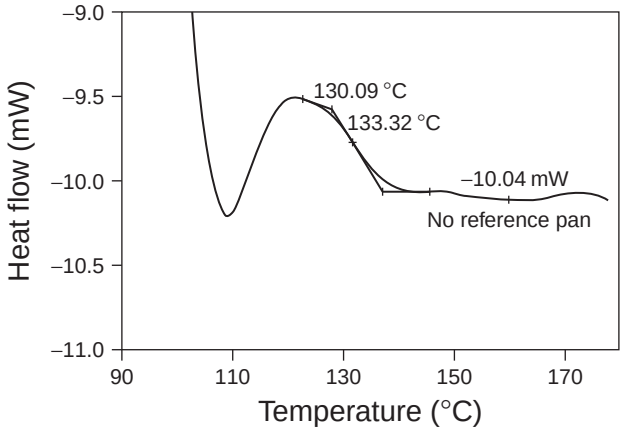

Figure 9 Start up hook and T_g with no reference pan.

5 °C min^{-1} or 20 °C min^{-1}. That is about 40 °C before the temperature of interest (2×20 °C min^{-1} = 40 °C and 8×5 °C min^{-1} = 40 °C).

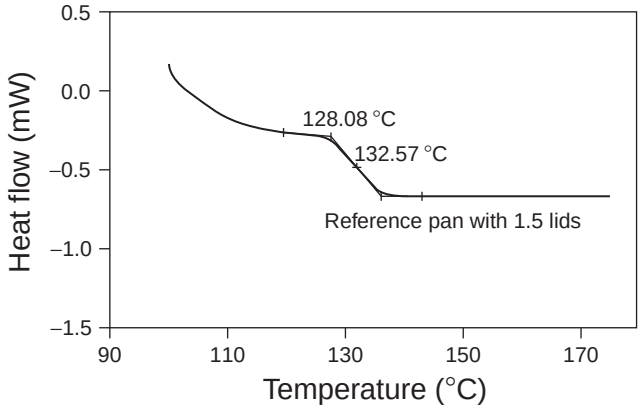

Figure 10 Start up hook and T_g with heat capacity of reference matched to sample.

2.7 Differential Scanning Calorimetry Sensitivity and Resolution

Adjust sample weight, heating rate and purge gas if necessary to improve sensitivity or resolution, see Table 3 and Equation (2)

$$\frac{dQ}{dT} = C_p \times \frac{dT}{dt} + f(T, t) \quad (2)$$

where heat flow = heat capacity × heating rate + kinetic component, and the kinetic component is a function of temperature and time.

2.8 Pressure Differential Scanning Calorimetry

A special pressure cell is needed for both high-pressure as well as low-pressure studies in nitrogen, air or oxygen. The sample of 1–3 mg, is weighed into an open pan as per the ASTM standard test method E-1858 for oxidation induction time (OIT).[56] In the OIT method, the sample, a drug, oil or polymer is heated to an isothermal temperature, 175 °C at 35-MPa oxygen and 195 °C at 14 kPa of air or oxygen. The extrapolated onset temperature is defined as the OIT. A second method, ASTM E-2009, oxidation onset temperature (OOT) is based on a DSC or PDSC.[57] The OOT is defined as the extrapolated onset temperature in a plot of heat flow versus temperature. A thermogravimetric analyzer (TGA) coupled with a differential thermal analyzer, TGA/DTA, has also been used to determine the OOT of hydrocarbons in air. In this study, the sample size was 10 mg, the flow rate of air was 250 mL min^{-1}, heating rate was 10 °C min^{-1}, and the sample pans were aluminum or platinum.[60,61]

Table 3 Basic heat flow Equation (2); parameters, sensitivity and resolution

Parameters	Increase sensitivity	Increase resolution
Sample weight	increase	decrease
Heating rate	increase	decrease
Purge gas	nitrogen	helium

2.9 Differential Scanning Calorimetry Method and Thermoplastics

When analyzing thermoplastics with a standard DSC, the experimental program should include a heat–cool–heat cycle at 10 °C min^{-1}.[62,63] The first heat results are a function of material and an unknown thermal history. The cool cycle gives the sample a known thermal history. These results can be used to compare polymer crystallization properties. Finally, the second heat results are a function of material and a known thermal history. They will be useful for comparison of materials, for example polymers, see Figures 11–13.

2.10 Differential Scanning Calorimetry Method and Thermosets

When analyzing thermoset polymers by convential DSC, the experimental program should include an annealing

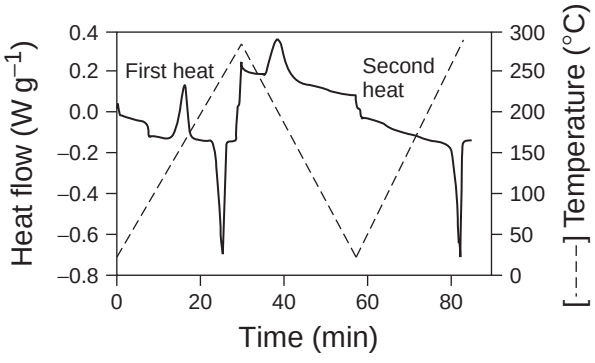

Figure 11 Thermoplastic: heat–cool–heat.

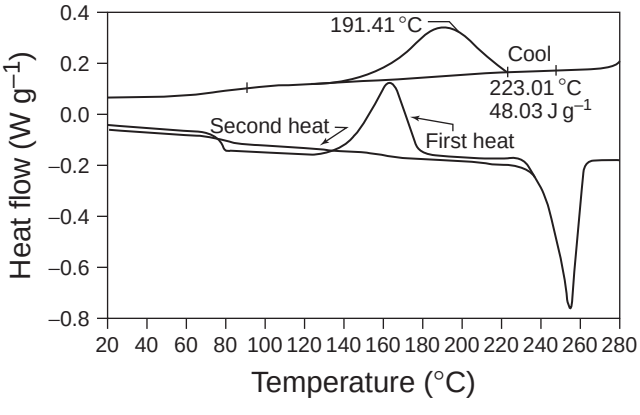

Figure 12 Thermoplastic: heat flow versus temperature.

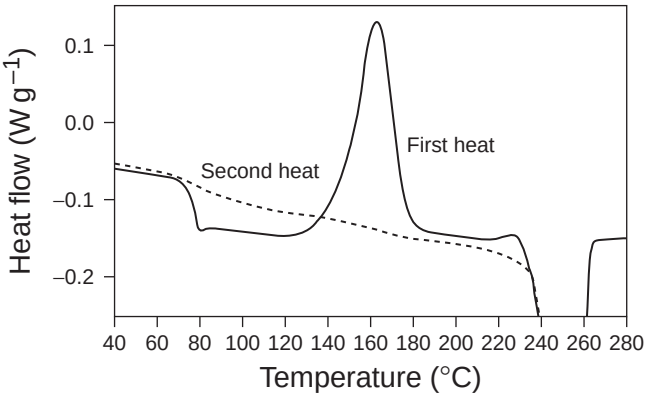

Figure 13 Comparison of first and second heating runs.

period, then heat, quench cool and finally heat at a rate of 10–20 °C min^{-1}. Anneal at approximately 25 °C above the T_g onset to eliminate effects of enthalpic relaxation. The heat portion of the program is to measure the T_g and any residual cure (exothermic polymerization). Do not heat through the decomposition onset temperature; halt the experiment before decomposition occurs. Quench the sample, usually with liquid nitrogen or a dry ice–acetone slurry in a cooling can. This process will give the sample a known thermal history. During the final heating the T_g of the fully cured sample can be measured, see Figures 14–16.

Select an end temperature that does not cause decomposition of the sample in the calorimeter. Degradation products can condense in the cell and cause either corrosion of the cell or baseline problems. Use sealed glass ampoules or stainless steel pans that can take high pressure, >70 MPa, in order to study decomposition by DSC. The use of aluminum hermetically sealed pans can minimize corrosion and destruction of the cell liner. However, knowledge of the onset of degradation assists in lengthening the longevity of a DSC cell.

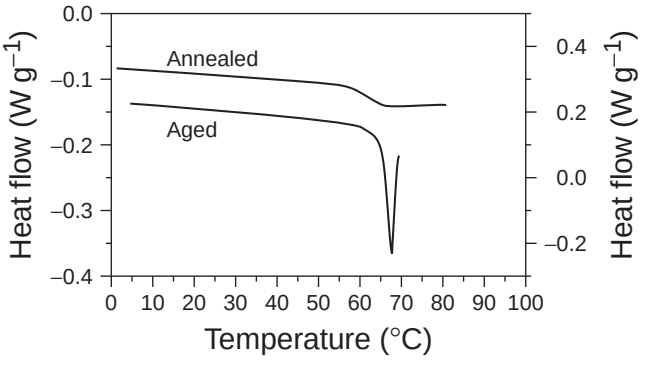

Figure 14 Effect of annealing on shape of glass transition.

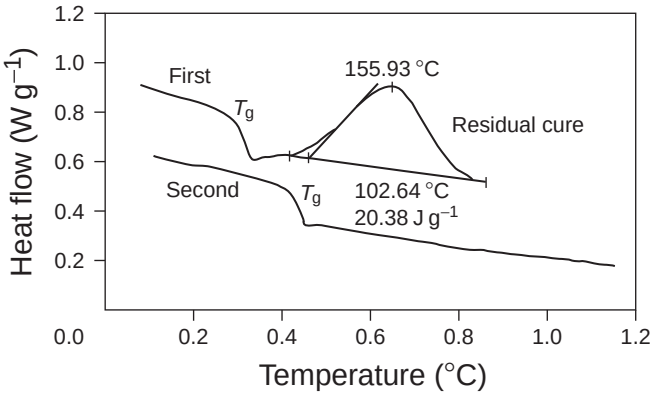

Figure 15 Thermoset: comparison of first and second heating runs.

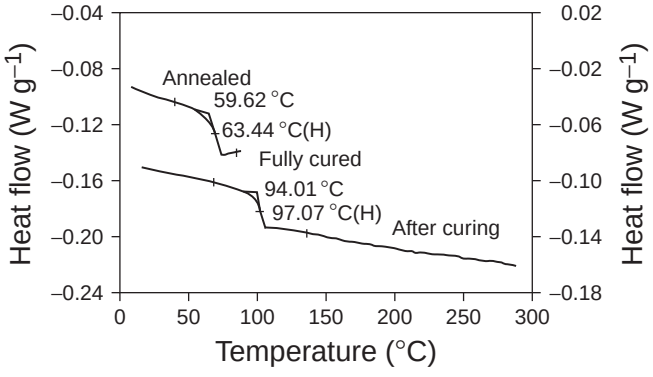

Figure 16 Shift in T_g as a result of curing.

2.11 Differential Scanning Calorimetry Method and an Unknown

When examining an unknown sample, first test the sample by thermogravimetric analysis (TGA). Determine if the sample contains any volatiles and its decomposition temperature. If you do not have a TGA, weigh the sample before and after each DSC run. A 1% weight loss will often show up in the thermal curve as an endothermic peak unless the run is taking place in either air or oxygen where oxidation shows up as an exothermic peak.

2.12 Internal Melting Point Standard

Use an internal melting point standard; see Table 1, to verify transition temperature accuracy. Place the sample in a DSC pan. Invert the lid and crimp it in place or force it into the pan so that good contact is achieved on the top and bottom of the sample. Weigh the internal standard, for example, indium metal, into the inverted lid. Record the DSC thermal curve. The melting profile of indium will be superimposed on the curve, Figure 17.

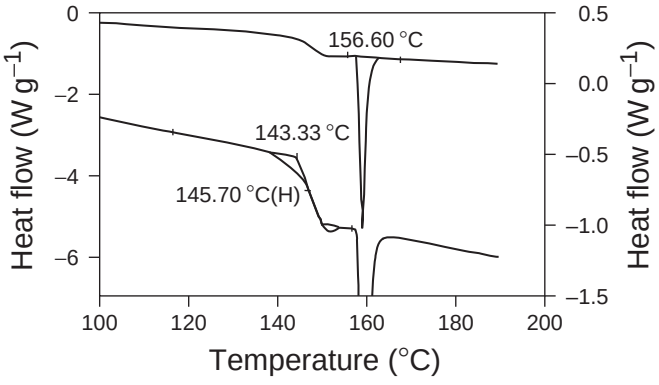

Figure 17 Use of internal melting point standard. Polycarbonate (PC) sample of approx. 10 mg with 1 mg indium, heating rate 5 °C min^{-1}.

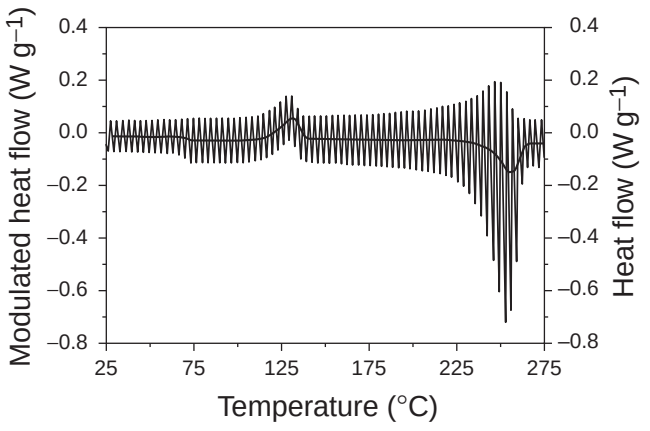

Figure 18 Total heat flow from MTDSC raw signals calculated as the average value of the modulated heat flow signal.

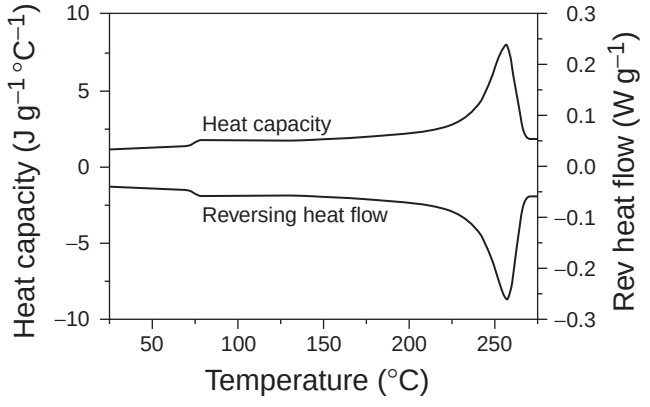

Figure 19 Reversing heat flow from MTDSC raw signals.

2.13 Modulated Temperature and Convential Differential Scanning Calorimetry

How does MTDSC differ from convential DSC? DSC measures the difference in heat flow between a sample and an inert reference. MTDSC measures the same total heat flow. A DSC thermal curve records the data as a function of a linear change in temperature. A MTDSC thermal curve represents the change in temperature as a function of both a linear and a sinusoidal change. The linear change in temperature provides the same information, total heat flow, in both MTDSC and DSC. The sinusoidal change in temperature permits the measurement of heat capacity effects simultaneously with the kinetic effects.

The DSC heating rate can vary from isothermal to 20 °C min^{-1}. The MTDSC heating rate or ramp ranges from isothermal to 5 °C min^{-1}. The MTDSC temperature of modulation amplitude can range from 0.01 to 10 °C. The modulation period can vary from 10 to 100 s or 10–100 mHz in frequency.

In order to compare the two methods, MTDSC and DSC, poly(ethylene terephthalate) (PET) will be examined by the two methods. An MTDSC procedure is:

1. Equilibrate at 0 °C.
2. Modulate 1 °C every 40 s.
3. Maintain isothermal at 0 °C for 5 min.
4. Ramp at 5 °C min^{-1} to 280 °C.

A corresponding DSC protocol is:

1. Equilibrate at 0 °C.
2. Maintain isothermal at 0 °C for 5 min.
3. Ramp at 20 °C min^{-1} to 280 °C.

DSC records only the total heat flow in the sample. The modulated heat flow contains all thermal events occurring in the sample. Fourier transformation analysis of the modulated heat flow signal is used to calculate its average signal continuously. Total heat flow is the same in both techniques at the same average heating rate. The reversing heat flow is the heat capacity component of the MTDSC total heat flow. The nonreversing heat flow is the kinetic component of the heat flow, Figures 18–20.

2.14 Examples of Modulated Temperature Differential Scanning Calorimetry Characteristics

The MTDSC of an as-received molded blend of a PET/acrylonitrilebutadiene terpolymer (ABS) is represented in Figure 21. The total heat flow thermal curve indicates a T_g at about 60 °C and an exothermic event at about 97 °C. The total heat flow T_g is the PET glass transition temperature. The exothermic event is the cold crystallization of PET. The cold crystallization event masked the ABS T_g. However, the reversing heat flow thermal curve clearly differentiates the ABS and PET T_g values. The ABS T_g(e), the

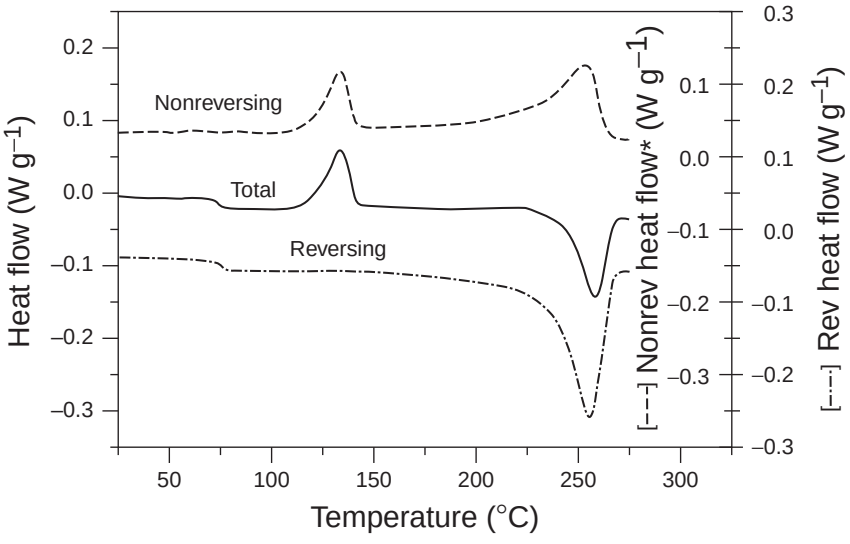

Figure 20 Quench cooled PET/MTDSC.

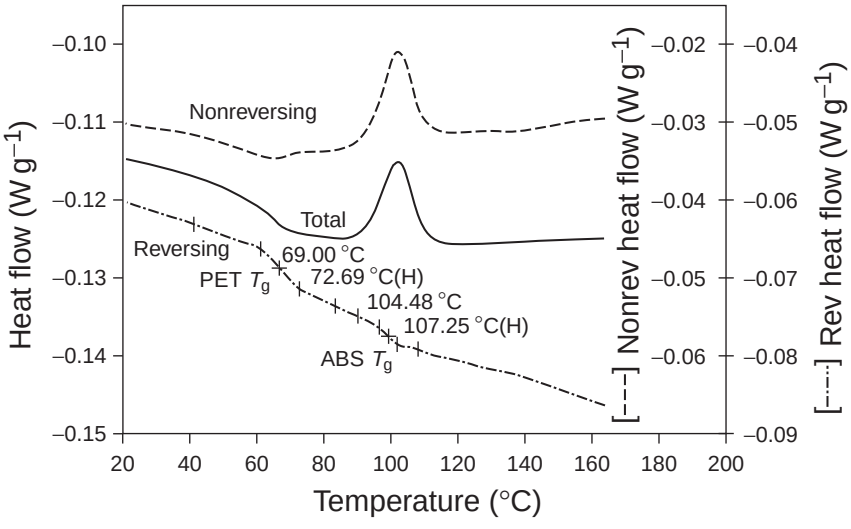

Figure 21 PET/ABS blend ''as received'' MTDSC.

extrapolated onset T_g, is 104.5 °C and the PET T_g(e) is 67.0 °C.

The T_g values and melt temperature of another polymer blend are clearly delineated in their first heat by MTDSC, see Figure 22. This blend contains PET, PC, and high-density polyethylene (HDPE). Overlapping thermal events confound the total heat flow curve. The exothermic PET cold crystallization overlaps the melting peak of HDPE. The reversing heat flow (W g^{-1}) thermal curve for reversible processes has a PET T_g at approximately 72 °C, a peak melt temperature for HDPE at 123 °C and a PC T_g at approximately 138 °C. The nonreversing heat flow, which is representative of nonreversible processes, shows an enthalpic relaxation endotherm at approximately 70 °C and the PET crystallization peak at ca. 115 °C.

2.15 Modulated Temperature Differential Scanning Calorimetry Experiments and the Glass Transition

Next, is a review of the MTDSC experimental factors that affect T_g properties of a thermoplastic and two elastomers. Also determined was the difference between glass transition temperature of elastomers by MTDSC and conventional DSC.

Aubuchon reported that ''the frequency effect seen in MTDSC data is the effect that (occurs when) the heat

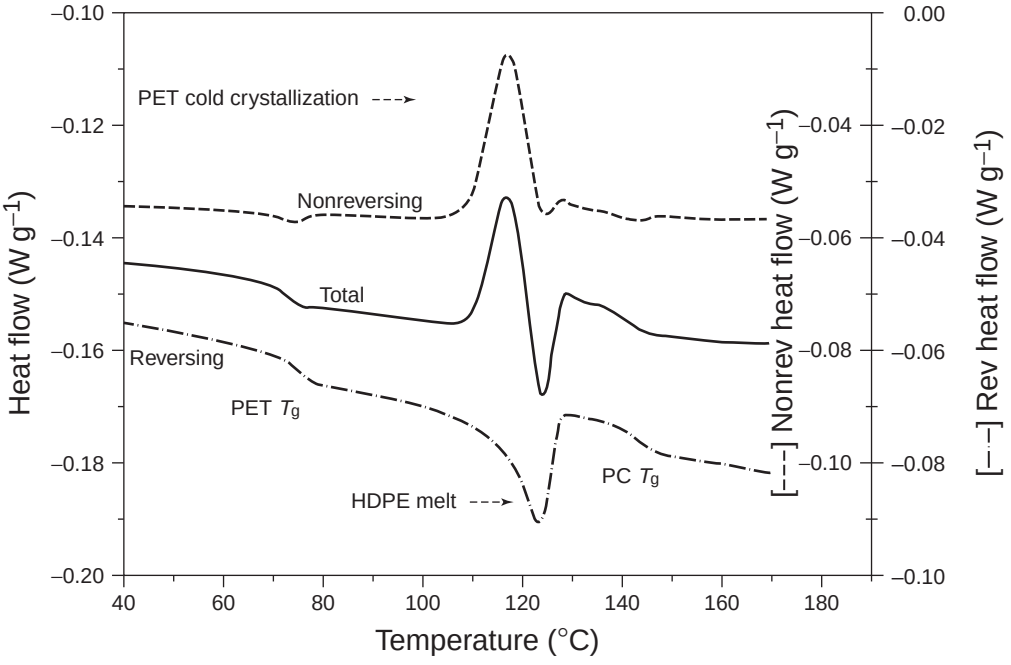

Figure 22 Polymer blend characterization by MTDSC.

capacity and reversing signals shift to higher temperatures with shorter periods (higher frequencies) of temperature modulation.[64] The total heat flow signal does not show any frequency dependence''. Both in dynamic mechanical analysis (DMA) and dielectric thermal analysis (DETA) the temperature shifts are accompanied by orders of magnitude change in frequency. The nonreversing heat flow signal is calculated from the difference of the total heat flow minus the reversing heat flow. Therefore, the nonreversing signal is also dependent on the frequency or period.

The MTDSC has two simultaneous ramps or heat rates. One is the average underlying heating rate (ramp) and the other, the instantaneous heating rate (amplitude). These ramp rates vary with the temperature modulation period and the amplitude. A shift in the extrapolated onset glass transition temperature, T_g(e), of PS with frequency was observed. In the heating cycle, the T_g(e) *decreased* by 1.3 °C with an *increase* of frequency from 25 to 100 s. In the cooling cycle, the T_g(e) *decreased* by 3.4 °C with the same *increase* in frequency. The effect of the amplitude or instantaneous heating rate (0.1, 0.2, and 0.4 °C) on the T_g at a constant heating rate of 1.5 °C min^{-1} showed a variation of 103.7, 103.9 °C (cooling cycle) and 104.4, 104.8 °C (heating cycle). Within experimental error, the T_g did not vary with amplitude for PS. Aubuchon noted that a four-fold increase in frequency (10–40 mHz) or decrease in period (100–25 s) increased the T_g by 3.0 °C. A four-fold increase in amplitude (instantaneous heating rate) increased the T_g(e) by 0.3 °C.

There is a great interest in how the amplitude, heating rate or ramp and period affect the T_g of polymers, for example, polybutadiene (PBD) and polystyrene-cobutadiene (SBR). A two to the third (2^3) factorial designed experiment[65] examined PBD, SBR and a National Institute of Standards and Technology (NIST) high molecular weight PS. The following are the three MTDSC experimental variables and the low and high levels:

- Amplitude, 0.312 and 0.796 °C
- Period, 25 s (40 mHz) and 60 s (17 mHz)
- Ramp, 2.0 and 5.0 °C min^{-1}.

Two sets of experimental variables are considered optimum for MTDSC[66]

1. amplitude, 0.312 °C; period, 60 s (0.017 Hz); and ramp, 2.0 °C min^{-1}
2. amplitude, 0.796 °C; period, 60 s (0.017 Hz); and ramp, 5.0 °C min^{-1}.

The run order for this study was randomized. The following physical properties associated with the glass transition temperature were measured in the reversing mode of the MTDSC for the eight runs and the three polymers (see Figure 23):

- T_g(e), the extrapolated onset glass transition temperature (°C)
- T_g(i), the inflection temperature of the sigmoid-shaped T_g curve (°C)

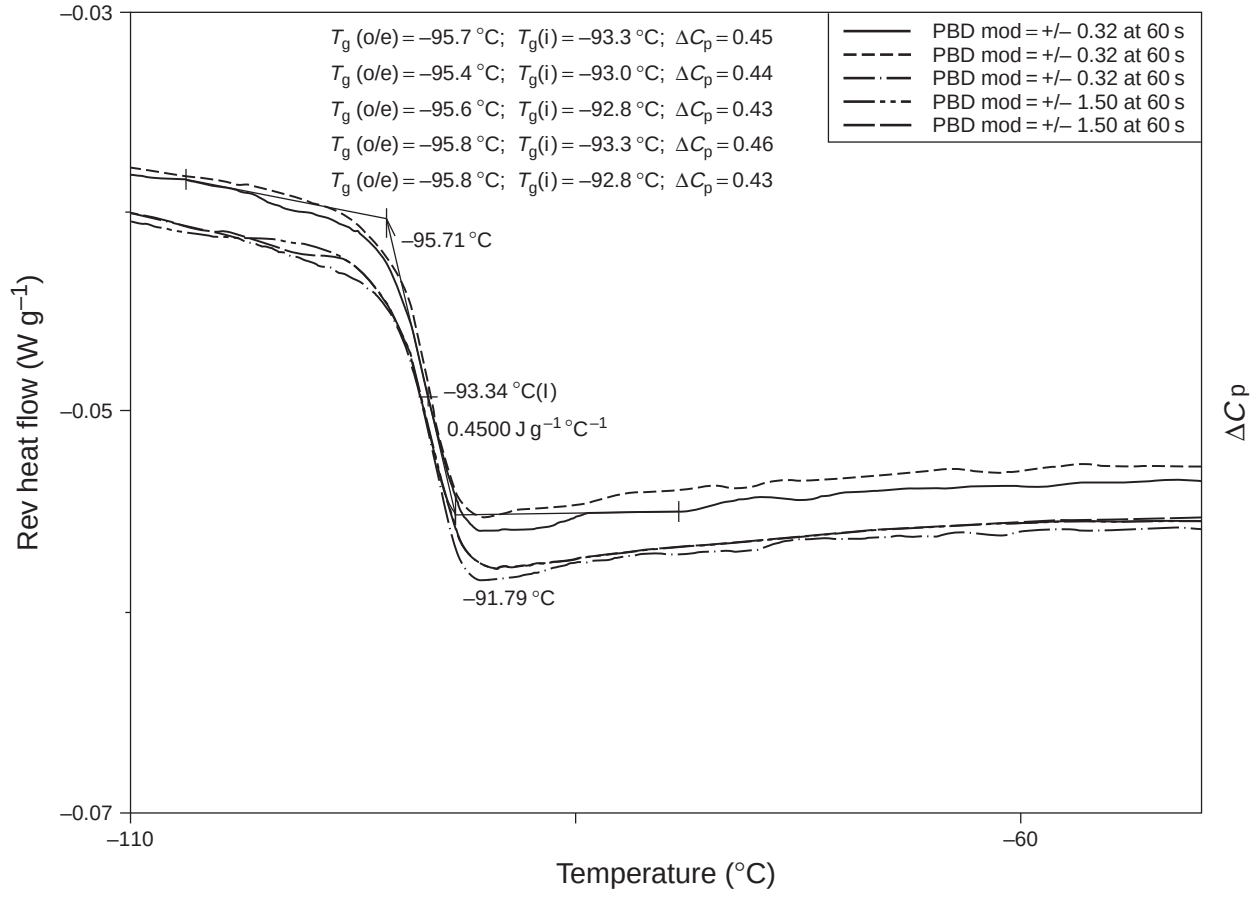

Figure 23 PBD: MTDSC amplitude variation.

- ΔC_p at T_g, the change in heat capacity at the T_g ($J\,g^{-1}\,°C^{-1}$) (y-axis change at T_g)
- ΔT at T_g, the change in temperature at the T_g, measured as the T_g (end) minus the T_g(e) (°C) (x-axis change at T_g), Figure 24.

Statistical and graphics analysis of the data was accomplished with JMP 3.2.[66] A least squares fit model was used and the significance of the data was ranked by the t-test.

The most important MTDSC polymer property studied was the glass transition temperature, T_g(e) or T_g(i). Employing the t-test to evaluate the measured variations in the factorial design the absolute variation of the T_g was $<2\,°C$ and this variation was not significant, Table 4. The interactive variables heating rate and period, heating rate and amplitude and period and amplitude also exhibited small variations in T_g ($<2\,°C$). The standard deviation of the T_g(e) at the 95% confidence level was $\pm 3\,°C$ for the polymers studied. The standard deviation of the T_g(i) at the 95% confidence level was $\pm 2\,°C$. Wide variations in the MTDSC variables, heating rate, amplitude and period did not significantly affect the value of T_g.

However, variation of the period from 25 to 60 s did significantly affect the PS properties, the change in heat capacity at T_g (ΔC_p at T_g) and the change in temperature at T_g (ΔT at T_g). The ΔC_p at T_g varied by 30% and the ΔT at T_g by 25%. Variation of the period also caused similar changes in the ΔC_p at T_g for SBR and PBD.

The t-test[68] for all four glass transition properties of PS varied significantly with the changes in the period (inverse frequency). The properties of the elastomers studied were not affected by the variables, except for the period.

A comparison between the T_g properties of PS as determined by MTDSC and convential DSC is given in Tables 4 and 5 and Figure 24. The MTDSC values in the total heat flow mode were selected from the experimentally designed study and compared to an ASTM DSC test.[69] The ASTM method calls for a heating rate of $10\,°C\,min^{-1}$. The T_g of PS at MTDSC heating rates of 2 and $5\,°C\,min^{-1}$ were compared to standard DSC heating rates of 5, 10 and $20\,°C\,min^{-1}$.

There was a small difference (0.7 and 1.0 °C, respectively) between the T_g(e) and T_g(i) for PS when examined by MTDSC at 2.0 and $5.0\,°C\,min^{-1}$. The T_g(e) found by MTDSC and DSC both at $5.0\,°C\,min^{-1}$ were 101.6 °C and

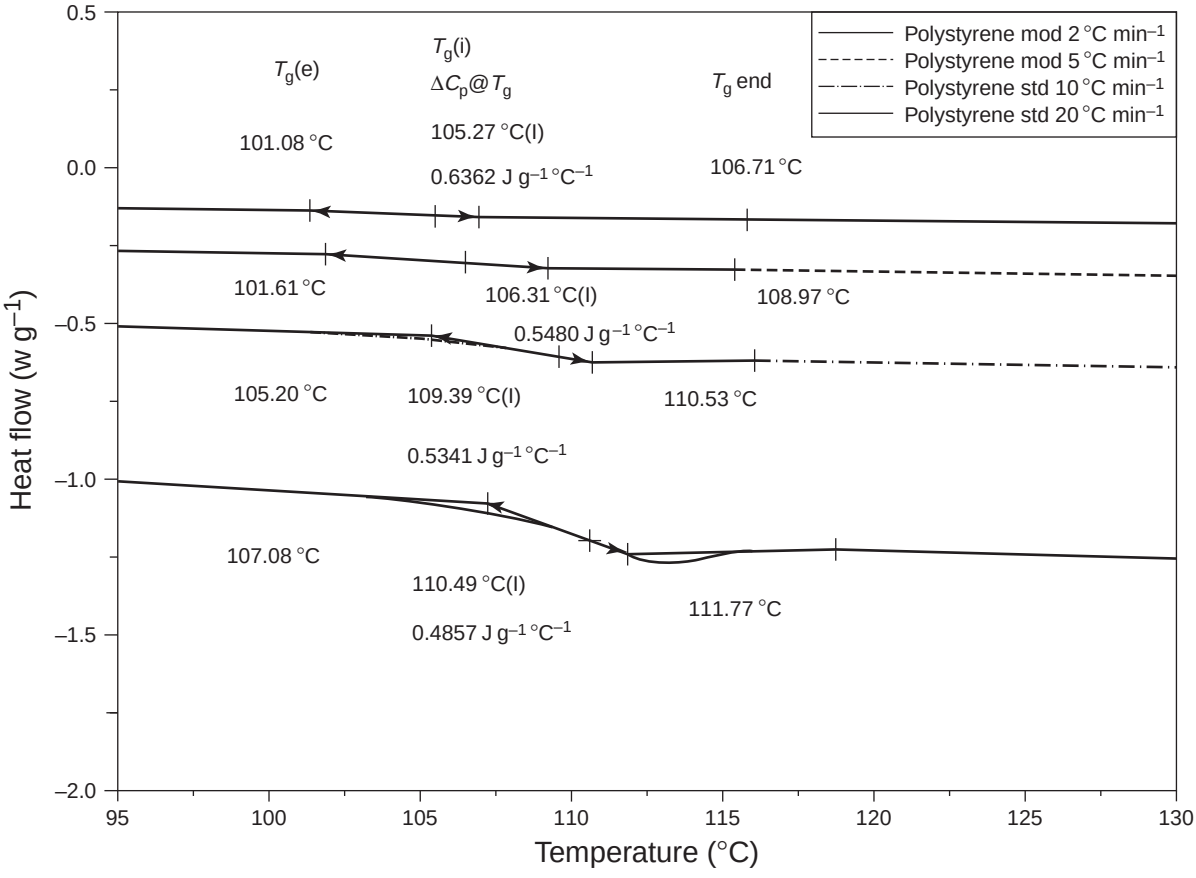

Figure 24 PS standard 900 000 MW: MTDSC plotted against standard DSC.

104.0 °C, a difference of 2.4 °C. The standard DSC T_g(e) increased with increasing heating rate, 5–20 °C min^{-1}, 104.0–107.1 °C. The optimum MTDSC conditions at 2 °C min^{-1} yields a T_g(e) at 101.1 °C and the ASTM standard DSC test at 10 °C min^{-1} had a T_g(e) of 105.2 °C. Therefore, when trying to compare T_g(e) or T_g(i) by MTDSC to the T_g by standard DSC, a 3–4 °C difference should be expected.

The ΔC_p at T_g and ΔT at T_g values were higher for MTDSC than DSC. Based on total heat flow, PS had a ΔC_p at T_g value of 0.54 J g^{-1} °C^{-1} and ΔT at T_g of 5.5 °C. In the MTDSC test the ΔT at T_g for PBD and PS were <6 °C, while the copolymer SBR had a value of 12–15 °C. It appears that MTDSC can distinguish between copolymers and homopolymers. Examining a number of standard thermoplastics substantiated the observation that copolymers have a larger ΔT at T_g than homopolymers.

The MTDSC variables do not significantly affect the measurement of the glass transition temperatures of the polymers studied. The period (frequency) did affect the T_g properties of PS. Only the heat capacity change at T_g was affected by the period variation for PBD and SBR. The MTDSC T_g (reverse heat flow mode) value is more accurate than the standard DSC T_g (total heat flow) measurement. There will be a 3–4 °C higher T_g value when measured by standard DSC compared with the MTDSC value.

2.16 Power Compensated Differential Scanning Calorimetry Experimental Procedures

Next, is a brief review of the power compensated DSC experimental procedures. The optimum experimental conditions for dynamic differential scanning calorimetry (DDSC), DSC and DTA using a Perkin Elmer DSC is 1–75 mg, depending on the sensitivity required.[70,71] The heating rate is 0.1–100 °C min^{-1} depending on the application. A standard heating rate is 20 °C min^{-1}. The experimental conditions for a DDSC are 5–30 mg depending on the sensitivity required. 10 mg is a good default. The repeat pattern is 30 s heat up by 2 °C, 30 s cool down by 1 °C. A wide range of other values works too, but this is a good starting point and one that leads to specific heats within a few percent without any special calibration. A power compensated DSC is

Table 4 Fundamental characteristics of the PS glass transition temperature by MTDSC

Factorial design run order in MTDSC	Experimental variables			Amplitude (+/− °C)	Measured properties (reversed heat)				
	Standard order	Heating rate	Frequency × 10^3 s^{-1} (period, s)		T_g(e) (°C)	T_g(i) (°C)	ΔC_p @ T_g ($J\,g^{-1}\,°C^{-1}$)	ΔT @ T_g (°C)	
3 (optimum)	1	2	17(60)	0.318	106.1	108.1	0.34	4.5	
	1				103.5	107.7	0.29	5.8	
	1				105.6	108.0	0.32	4.2	
					105.1	*107.9*	*0.32*	*4.8*	ave
5	2	5	17(60)	0.318	104.0	108.0	0.36	6.0	
	2				103.5	107.8	0.38	6.4	
					103.8	*107.9*	*0.37*	*6.2*	ave
2	3	2	40(25)	0.318	106.5	107.6	0.24	5.5	
	3				106.8	107.6	0.23	4.2	
					106.7	*107.6*	*0.24*	*4.9*	ave
1	4	5	40(25)	0.318	108.4	110.0	0.15	2.5	
	4				108.9	110.1	0.13	2.0	
	4				108.2	110.1	0.19	3.3	
					108.5	*110.1*	*0.16*	*2.6*	ave
8	5	2	17(60)	0.796	104.9	108.5	0.38	5.7	
	5				105.2	108.0	0.38	5.7	
					105.1	*108.3*	*0.38*	*5.7*	ave
6 (optimum)	6	5	17(60)	0.796	105.7	108.4	0.35	5.8	
					105.1	109.0	0.37	6.0	
					105.4	*108.7*	*0.36*	*5.9*	ave
4	7	2	40(25)	0.796	106.6	110.0	0.25	5.6	
	7				106.1	110.0	0.28	6.4	
					106.4	*110.0*	*0.27*	*6.0*	ave
7	8	5	40(25)	0.796	107.2	110.0	0.24	5.1	
	8				107.3	110.1	0.25	5.3	
					107.3	*110.1*	*0.25*	*5.2*	ave
			average (C)		*106.0*	*108.8*	*0.29*	*5.0*	
			standard deviation +/−	sigma	1.5	1.0	0.08	1.3	
			2 × standard deviation +/−	2 × sigma	3.0	2.0	0.19	2.6	
			% standard deviation +/−		1.5	0.9	27.5	25.0	

Table 5 Fundamental characteristics of the PS glass transition temperature by MTDSC and convential DSC

Sample number	Ramp (°C min^{-1})	Method	MTDSC T_g properties Measured properties (heat flow)				
			T_g(e) (°C)	T_g(i) (°C)	ΔC_p @ T_g ($J\,g^{-1}\,°C^{-1}$)	ΔT @ T_g (°C)	
A3	2	MTDSC (heat flow)	101.1	105.3	0.64	5.6	
	2	MTDSC (reversed heat flow)	*101.1*	*105.3*	*0.64*	*5.6*	
A6	5	MTDSC (heat flow)	101.6	106.3	0.55	7.4	
	5	MTDSC (reversed heat flow)	*101.4*	*105.8*	*0.60*	*6.5*	
B	5	Standard DSC (heat flow)	104.0	107.3	0.50	4.7	
C	10	Standard DSC (heat flow)	105.2	109.4	0.53	5.3	
D	20	Standard DSC (heat flow)	107.1	110.5	0.49	4.7	
		Std DSC average (C)	*105.4*	*109.1*	*0.51*	*4.9*	ave
		standard deviation +/−	1.6	1.6	0.02	0.3	
		% standard deviation +/−	1.5	1.5	4.1	7.1	
		DSC and MTDSC average	*103.1*	*107.1*	*0.56*	*5.7*	ave
		standard deviation +/−	2.4	2.1	0.06	1.0	
		% standard deviation +/−	2.3	1.9	11.1	17.2	

described in Figure 25. A DDSC thermal profile is given in Figure 26.

2.17 Quantitative Differential Scanning Calorimetry and Differential Thermal Analysis

DSC and DTA are used to determine a wide range of physical properties of materials, including the glass transition temperature, melting temperature and solid–solid transition. To quantify these, transition temperatures instrument calibration is necessary with known standard materials. Therefore, calibration of DSC and DTA from −100 to 600 °C consists of heating the calibration material at a controlled rate in a controlled atmosphere through a region of known thermal transition.[72,73] The temperature and the heat flow difference in a heat flux DSC are continuously monitored, that is, between the calibration material and a reference material. A transition is marked by the absorption of energy by the specimen resulting in an endothermic peak in the heating mode of the thermal curve.

For a two-point temperature calibration, select a pure material in the temperature range of interest, see Table 1. The DSC or DTA can be calibrated with some degree of confidence with indium and zinc and the temperature correction factor can be extrapolated 100 °C or from about 50 to 500 °C. However, in order to report transition temperatures in a journal article or in an internal communication, the temperature calibration is best when it is very close to the temperature that needs to be measured. As a minimum, two temperature standards that bracket the temperature of interest should be used.

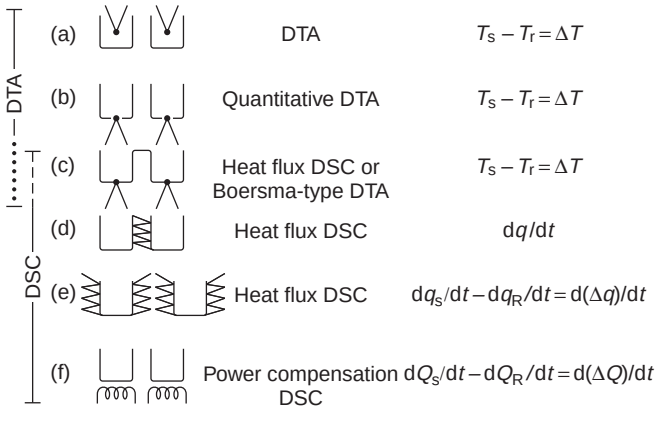

Figure 25 Comparison of DTA, power compensation DSC and heat flux DSC. S is the sample, R is the reference, T is the temperature and q is the heat.

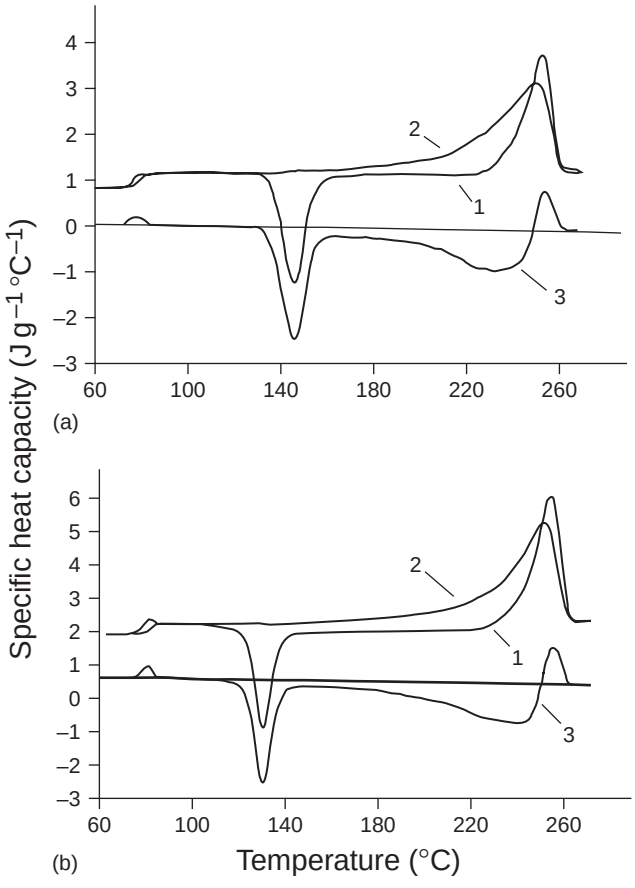

Figure 26 Measured curves of PET ($\beta_0 = 5\,\mathrm{K\,min^{-1}}$, $T_a = 1\,\mathrm{K}$, $t_p = 30\,\mathrm{s}$). (a) Measured with a heat flux DSC. (b) Measured with a modified Perkin–Elmer DSC-7. Calculated from: 1, the underlying heat flow; 2, reversing component; 3, nonreversing component.

2.18 Low-temperature Quantitative Analysis

Low-temperature calibration with mercury is valid. However, mercury is toxic and most researchers prefer to use the melting temperature of spectroscopic grade organic chemicals, for example, octane, −58 °C, decane, −30 °C or dodecane at −10 °C.

DSC specimen containers are typically solvent cleaned aluminum or alumina pans. The pans can also be crimped with a lid or hermetically sealed to preserve the standard material. DTA specimen containers are typically thin walled glass capillaries with a flat bottom. Specimen containers must not react with a standard or sample material or melt in the temperature range of interest. Purge gases, like nitrogen or helium, are used at a flow rate of ca. 50 mL min^{-1}. The mass of the standard is measured to the nearest 0.1 mg. The melting temperatures of calibration materials are given in Table 1. The most common melting temperature standards are indium at 156.6 °C, tin at 231.9 °C and zinc at 419.5 °C.

Table 6 Sapphire specific heat capacity literature values[a,b]

Temperature (°C)	(K)	Specific heat ($J\,g^{-1}\,°C^{-1}$)	Temperature (°C)	(K)	Specific heat ($J\,g^{-1}\,°C^{-1}$)
−183.15	90	0.0949	296.85	570	1.0876
−173.15	100	0.1261	306.85	580	1.0932
−163.15	110	0.1603	316.85	590	1.0987
−153.15	120	0.1968	326.85	600	1.1038
−143.15	130	0.2349	336.85	610	1.1089
−133.15	140	0.2739	346.85	620	1.1137
−123.15	150	0.3134	356.85	630	1.1183
−113.15	170	0.3526	366.85	640	1.1228
−103.15	170	0.3913	376.85	650	1.1271
−93.15	180	0.4291	386.85	660	1.1313
−83.15	190	0.4659	396.85	670	1.1353
−73.15	200	0.5014	406.85	680	1.1393
−63.15	210	0.5356	416.85	690	1.1431
−53.15	220	0.5684	426.85	700	1.1467
−43.15	230	0.5996	446.85	720	1.1538
−33.15	240	0.6294	466.85	740	1.1604
−23.15	250	0.6579	486.85	760	1.1667
−13.15	260	0.6848	506.85	780	1.1726
−3.15	270	0.7103	526.85	800	1.1783
0.00	273.15	0.7180	546.85	820	1.1837
6.85	280	0.7343	566.85	840	1.1888
16.85	290	0.7572	586.85	860	1.1937
26.85	300	0.7788	606.85	880	1.1985
36.85	310	0.7994	626.85	900	1.2030
46.85	320	0.8188	646.85	920	1.2074
56.85	330	0.8373	666.85	940	1.2117
66.85	340	0.8548	686.85	960	1.2159
76.85	350	0.8713	706.85	980	1.2198
86.85	360	0.8871	726.85	1000	1.2237
96.85	370	0.9020	746.85	1020	1.2275
106.85	380	0.9161	766.85	1040	1.2312
116.85	390	0.9296	786.85	1060	1.2348
126.85	400	0.9423	806.85	1080	1.2383
136.85	410	0.9545	826.85	1100	1.2417
146.85	420	0.9660	846.85	1120	1.2451
156.85	430	0.9770	866.85	1140	1.2484
166.85	440	0.9875	886.85	1160	1.2516
176.85	450	0.9975	906.85	1180	1.2548
186.85	460	1.0070	926.85	1200	1.2578
196.85	470	1.0161	976.85	1250	1.2653
206.85	480	1.0247	1026.85	1300	1.2724
216.85	490	1.0330	1076.85	1350	1.2792
226.85	500	1.0409	1126.85	1400	1.2856
236.85	510	1.0484	1176.85	1450	1.2917
246.85	520	1.0557	1226.85	1500	1.2975
256.85	530	1.0627	1276.85	1550	1.3028
266.85	540	1.0692	1326.85	1600	1.3079
276.85	550	1.0756	1376.85	1650	1.3128
286.85	560	1.0817			

Source: D.A. Ditmars et al., *Res. Natl. Bur. Stand.*, **87**(2), 159–163 (1982).

[a] Specific heat capacity measurements by DSC require the use of a well-characterized reference material (usually sapphire) to obtain results. This table summarizes specific heat capacity data (literature values) for sapphire as a function of temperature. These values can be used in the required calculations.

[b] The values in the table were determined by Ginnings and Furukawa of the National Bureau of Standards on aluminum oxide in the form of synthetic sapphire (corundum). The sapphire pieces passed a #10 sieve but were retained by a #40 sieve, and had 99.98–99.99% purity by weight. Heat capacity values below the experimental range were obtained by extrapolation of the Debye equation fitted to the experimental value at the lowest temperature. The units ($J\,g^{-1}mol^{-1}$) are absolute joules per degree per gram mole (molecular weight, 101.9613) at a constant pressure of 1 atm (100 Pa).

For a comprehensive list of available standards, see Table 1.

2.19 Specific Heat Capacity by Differential Scanning Calorimetry and Modulated Temperature Differential Scanning Calorimetry

There is an ASTM method, E-1269-95, that covers the determination of specific heat capacity, C_p, by convential DSC.[74] It is applicable to thermally stable solids and liquids. The normal operating range of the test is from −100 to 600 °C. The total DSC heat flow of an empty sample and reference pan is measured. The C_p is determined from the empty pan heat flow profile and the sample heat profile from the thermal curve. Subtraction of these two thermal curves, sample minus the empty pan, will give C_p as a function of temperature. A sapphire standard is used as the calibrant. It is best to evaluate the multiple heat flow responses to sapphire, in order to determine the precision and accuracy of this method, see Table 6.

Measurements of the C_p by the MTDSC method are direct; that is, the specific heat varies linearly over the desired temperature range.[75,76] When determining C_p by MTDSC, calibrate the cell with sapphire every 10 °C and correct the reversing heat flow output (the calorimetric portion of the total heat flow). A high-molecular weight NIST PS standard is used to verify the sapphire calibration. The repeatability of the C_p measurement is 2–5% of the value at a specific temperature.

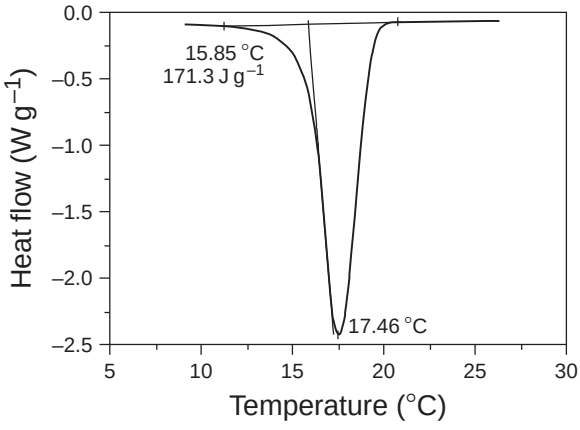

Figure 27 DSC of hexadecane at low temperatures, 2 °C min^{-1}.

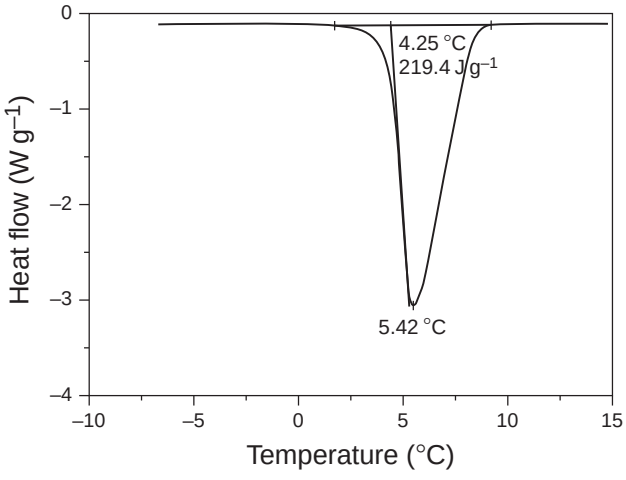

Figure 28 DSC of tetradecane at low temperatures, 2 °C min^{-1}.

3 APPLICATIONS

Some of the DSC and DTA applications reviewed are based on physical and chemical properties of materials. Melting temperature and heat of fusion, crystallization temperature and heat of crystallization, specific heat capacity of polymers, and curing will be included.

3.1 Alkane Structure–Property Relationships by Differential Scanning Calorimetry, Thermal Microscopy and X-ray Diffraction Analysis

DSC and thermal microscopy helped define a commercial product composed of normal alkanes that was involved in a major law suit.[77] The solid-state structures of a number of alkanes have unique crystal structures. These alkanes melt and crystallize below room temperature at more than 60 °C below zero, see Figures 27 and 28. Mixtures of specific alkanes have attributes of pure chemicals, the X-ray diffraction structure (XDS) was very similar, but the melting and crystallization temperatures were statistically different from those expected, see Table 7. The XDS, Miller Index d (002) of a number of even carbon number normal alkanes was linearly related to the corresponding melting and crystallization temperatures, see Figure 29.

3.2 Oxidative Behavior of Materials by Standard and Pressure Differential Scanning Calorimetry: Polymers and Engine Oils

Fundamental knowledge of the oxidative properties of commercial oils is necessary to predict the stability of these fluids. A primary tool to determine the oxidation of fully formulated motor oils, greases, diesel oils, transmission fluids and vegetable oils is DSC and PDSC.[60] A diluted motor oil was used to establish experimentally designed relationships between the DSC variables and the OIT,

Table 7 Melt and crystallization properties of *n*-alkanes by DSC

Normal alkanes		Temperature (°C)			Heat of transition (J g^{-1})		
		T_m	T_n	T_m (lit)	ΔH_m	ΔH_c	ΔH_m[a]
1. Octane	C_8H_{18}	−59	−59	−56	176	181	
2. Decane	$C_{10}H_{22}$	−32	−33	−32	177	177	
3. Dodecane	$C_{12}H_{26}$	−12	−13	−12, −10	182	188	
4. Tetradecane	$C_{14}H_{30}$	4.2	4.3	5.5	220	221	
5. Mixture	$C_{14}-C_{16}$(10 : 90)w	11	11	17[a]	180	178	178[a]
6. Hexadecane	$C_{16}H_{34}$	16	16	18, 16	174	172	
7. Octadecane	$C_{18}H_{38}$	27	27	28	177	175	
8. Eiocosane	$C_{20}H_{42}$	36	36	38	199	201	
9. Docosane	$C_{22}H_{46}$	42	43	43	196	197	

[a] Calculated value based on property additivity of $C_{14}H_{30}$ and $C_{16}H_{34}$.
T_m = melt temperature based ASTM E794-97; T_n = nucleation/crystallization temperature based on ASTM E794-97; ΔH_m = heat of fusion; ΔH_c = heat of crystallization; lit = literature values.

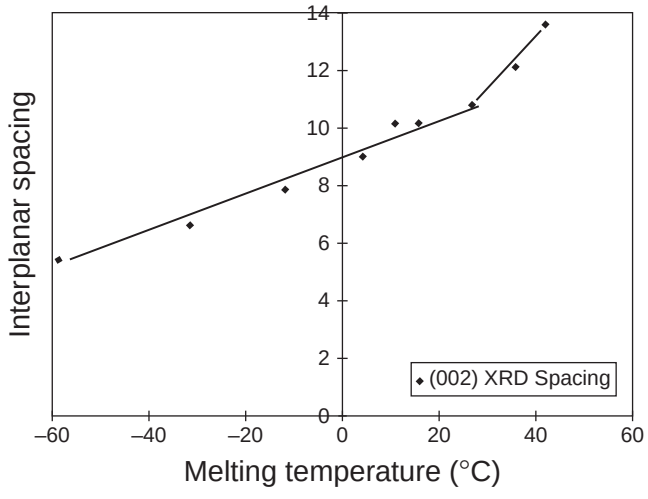

Figure 29 XRD interplanar distance (002) plotted versus DSC melt temperature.

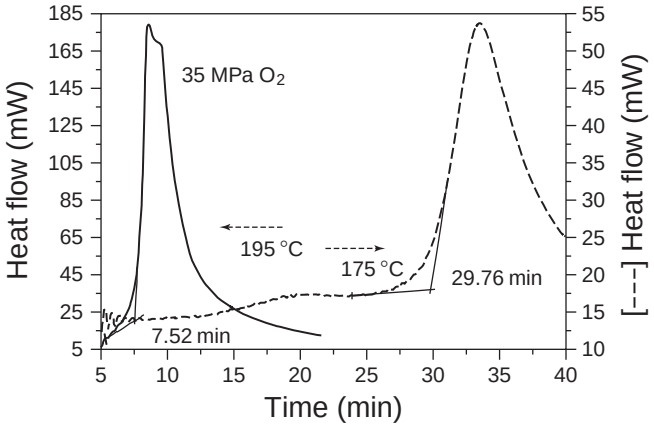

Figure 30 OIT at constant pressure and variable temperature.

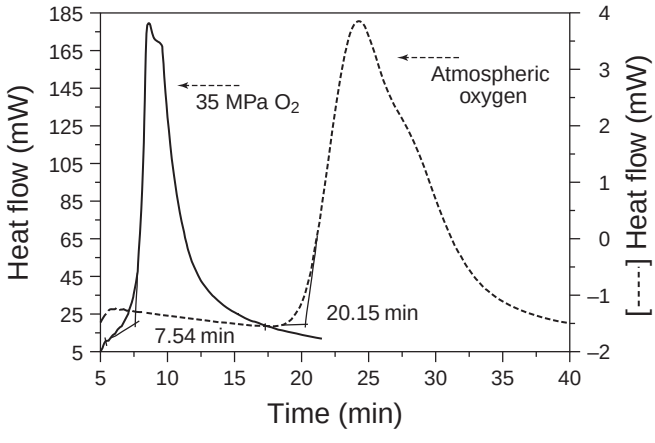

Figure 31 OIT at constant temperature and variable pressure.

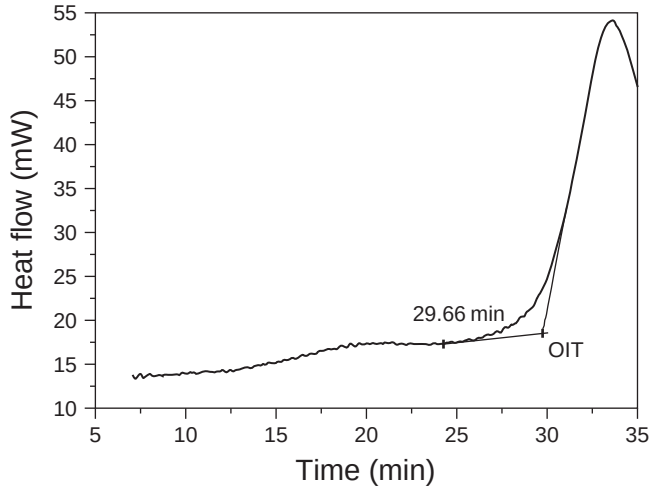

Figure 32 OIT by PDSC.

ASTM E-1858, by PDSC and DSC, see Figure 4 and Figures 30–32.

The variables used to develop this protocol were temperature, pressure, heating rate, sample mass, gas flow

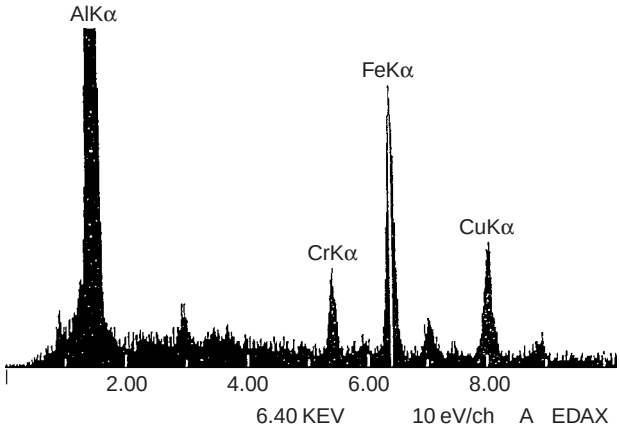

Figure 33 SEM–EDS (Energy Dispersive Spectroscopy) of aluminum pan with Fe.

rate, and gas type, air or oxygen. The DSC pan metallurgy played a statistically significant role in the measured OIT. Quality control charting of the DSC oxidation process discovered an out-of-control problem that was related to the impurities in the aluminum pans, see Figure 33 and Table 8. Iron impurities caused a decrease in the OIT of the reference fluid, while the presence of chromium stabilized the oxidation process. A costly engine test can be predicted when a new fully formulated oil blend is examined by PDSC using the OIT protocol, see Figure 34 and Table 9.

Oxidative behavior of commercial engineering plastics, polyolefins and elastomers were evaluated by DTA, TGA and PDSC.[61] There is a good correlation between the measured stability by PDSC in oxygen (ASTM E-1858, OIT) and DTA/TGA (ASTM E-2009, OOT) in air for olefin polymers, see Figures 35–39. The thermooxidative properties of the polymers studied were reliable and precise.

3.3 Inorganic Reactor Deposits by Differential Thermal Analysis, Thermogravimetric Analysis and X-ray Diffraction Analysis

Ash deposits on heat exchange tubes were characterized by DTA, TGA and X-ray diffraction methods.[78] The

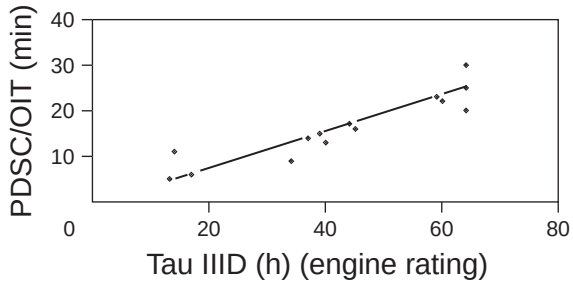

Figure 34 Engine test sequence IIID Tau plotted against PDSC/OIT at 185 °C.

Table 9 Comparison of PDSC/OIT engine ratings and performance: effect of base oils with the same lubricant additive package

Base oil	PDSC/OIT (min)	OIT Predicted[a] viscosity @ 64 h	Observed sequence IIID viscosity @ engine test time
A	42	Pass	Pass @ 64 h
B	35	Pass	Pass @ 64 h
C	35	Pass	Pass @ 64 h
D	36	Pass	Pass @ 56 h Fail @ 64 h
E	22	Borderline fail	Fail @ 48 h
F	14	Fail	Fail @ 48 h
G	8.2	Fail	Fail @ 16 h

[a] OIT predicted = PDSC at 175 °C and 35 MPa oxygen. OIT predicted viscosity = <20 min, fail; OIT predicted viscosity = 20–25 min, borderline fail; OIT predicted viscosity > 25 min, pass.

ash samples were deposited at different locations in the atmospheric fluidized bed combustor system at the TVA Shawnee Steam Plant, Paducha, KY (Figure 40). DSC aided in the identification of the endothermic and exothermic behavior of the deposits and helped develop the mechanism that promotes deposits. The major compound in the deposits from the convention pass inlet, superheater and multiclone inlet was calcium sulfate, $CaSO_4$. The predominant minor component was calcium oxide, CaO. The combined thermal and X-ray techniques are useful for profiling and identifying an unknown ash sample.

Table 8 Effect of pan surface metallurgy on PDSC/OIT and SEM/EDS elemental analysis

PDSC[a]	Elemental analysis					
Results	OIT (min)	Iron	Copper	Chromium	Ratios Fe/Cr	Fe + Cu/Cr
Poor	12	414	173	135	3.1	4.3
Fair–Poor	17	328	136	103	3.2	4.5
Good	26	315	157	235	1.3	2.1

[a] PDSC = ASTM E 1858, interlab study in aluminum pans.

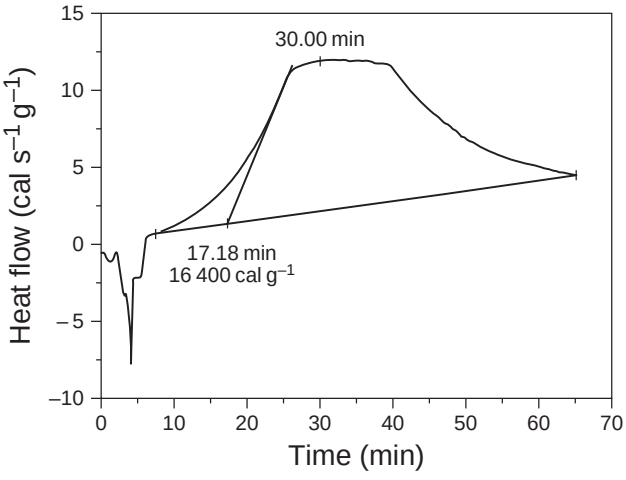

Figure 35 PDSC of low-density polyethylene (LDPE) in oxygen.

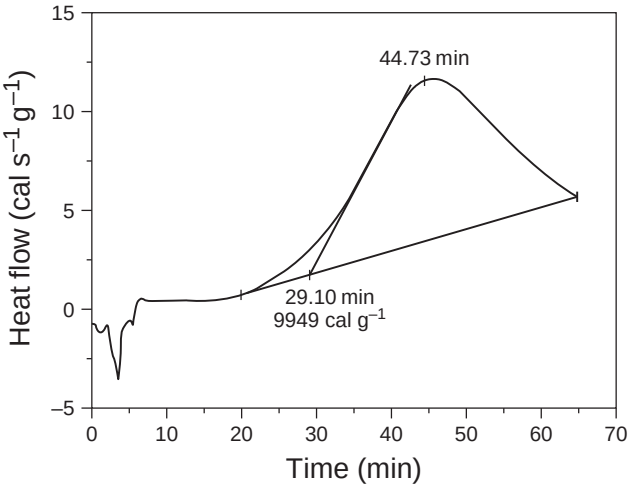

Figure 36 PDSC of polyethylene-covinyl acetate (EVA) in oxygen.

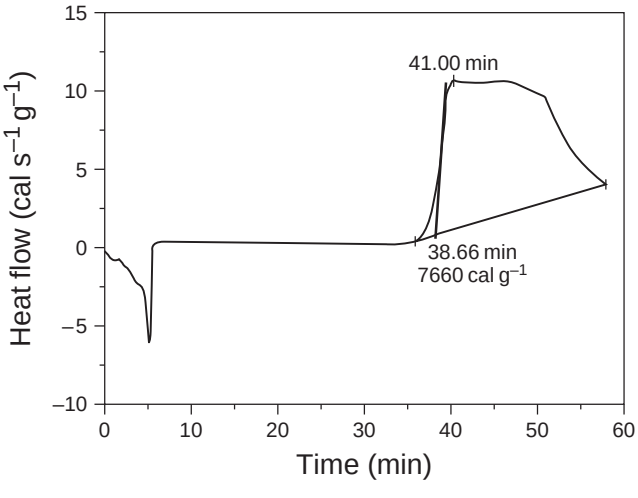

Figure 37 PDSC of polypropylene (PP) in oxygen.

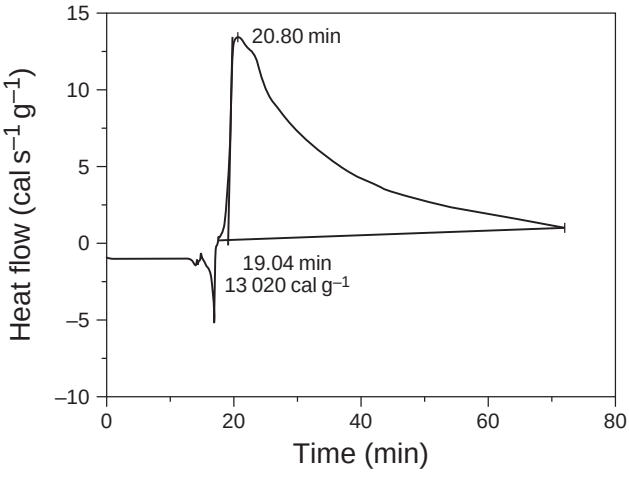

Figure 38 PDSC of PP plus $CaCO_3$ in oxygen.

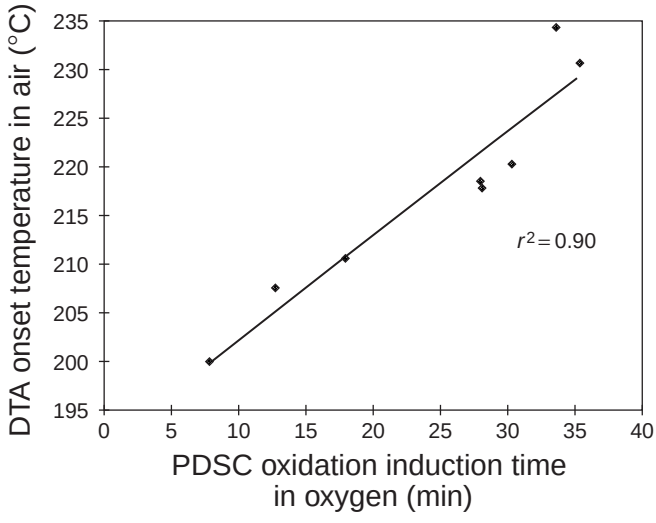

Figure 39 OOT by DTA plotted against OIT by PDSC.

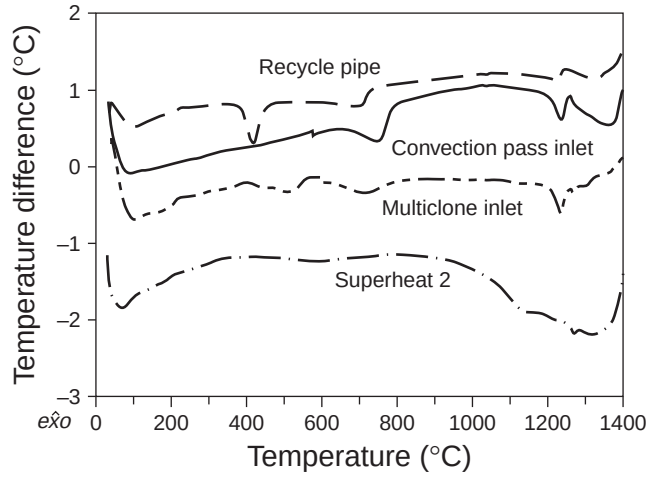

Figure 40 Comparison of DTA curves of four deposits heated to 1490 °C.

Table 10 Characterization of polymers by DSC, DTA and TMA[a]

Polymer	RK#	Transition temperature (°C)					
		DSC Set A	DTA	TMA			
				Set B			
		T_m, T_{oc}	T_m, T_p	T_g, T_{oc}	T_t, T_{oc}	T_m, T_{oc}	COE
1. HDPE	25	121	123			127	120
2. MDPE	49	na	121			126	115
3. LDPE	24	91	93			99	136
4. EVA	34	84	88			87	93
5. PP	27	145	147			154	
6. PP copolymer	26	133	137			147	
7. PP + flame retardant	38	na	151			153	
8. PP + talc	44	na	152			154	
9. PP + calcite	45	na	153			154	
10. PP + mica	46	na	151			153	
11. PS-general purpose	1			90	136		54
12. PS-medium impact	2			90	140		56
13. PS-high impact	3			84	140		57
14. ABS-medium impact	6			96	142		55
15. ABS-high impact	7			92	141		52
16. SAN-copolymer	4			98	144		53
17. SB-copolymer	8			68	111		103
18. Acetal	31			42			91
19. Acetal-copolymer	32			65			80
20. Acetal-copolymer	32			64			78
21. Nylon 6	16			51			34
22. Nylon 66	15			44			30
23. Nylon 6 + glass	47			40			13

[a] All data collected in 1996.
RK#, resin kit sponsored by the Society of Plastics Engineers, Inc.; Set A, resin kit, 1979; Set B, resin kit, 1994; T_g, glass transition temperature, °C; T_m, melting temperature, °C; T_t, transition temperature, marked change in COE, °C; COE, coefficient of linear expansion (mm mm^{-1} °C^{-1}); T_{oc}, extrapolated onset temperature, °C; T_p, peak temperature, °C; na, not available. MDPE, medium-density polyethylene.

3.4 Polymer Characterization by Thermogravimetric Analysis, Differential Thermal Analysis/Differential Scanning Calorimetry, Thermomechanical Analysis, Fourier Transform Infrared and X-ray Diffraction Analysis

A diverse set of commercial polymers were characterized by TGA/DTA, DSC, thermomechanical analysis (TMA), Fourier transform infrared (FTIR) spectroscopy and X-ray diffraction analysis.[79] The DSC and DTA were used to document the polymer melting temperature.

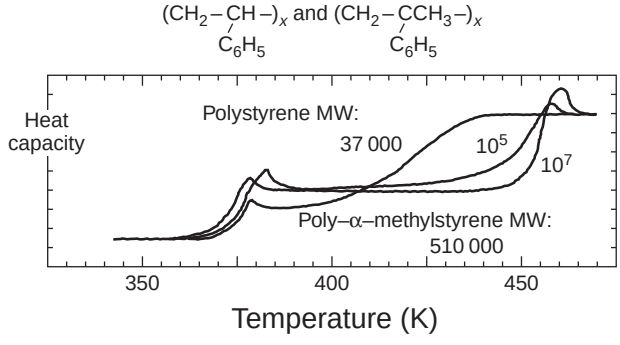

Figure 41 Change of breadth of T_g in copolymers (Wunderlich).

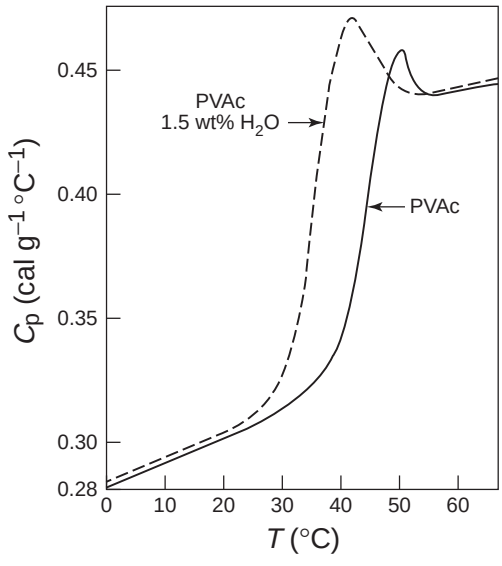

Figure 42 C_p versus T showing plasticizing effect on PVAcs T_g (Bair).

Polyolefin melt temperatures by DSC, DTA and TMA were very comparable, see Table 10.

3.5 Assignment of the Glass Transition Temperature

Assignment of the glass transition temperature was the focus of an ASTM conference and special technical publication.[80] R. Seyler, who was the editor and conference chairperson, lead the discussion on the T_g. His observations on DSC and DTA are summarized here: a consistently assigned T_g will vary because of differences in heating rate; composition, both chemical and physical, will cause the T_g to vary; diluent and fillers in a specimen usually affect the T_g, moisture is particularly troublesome in T_g measurements. There were many calorimetric contributions to the assignment of the T_g of polymers.

3.6 Copolymer T_g and Molecular Weight

Wunderlich[81] pointed out that the breadth of the T_g in copolymers varies with molecular mass, see Figure 41.

3.7 Polymer T_g and Moisture

Bair[82] demonstrated the effect of water on the T_g of polyvinyl acetate (PVAc); the T_g was lower with the addition of 1.5 wt% water, Figure 42.

3.8 Effect of High-pressure Gases on the Polymer T_g

Some representative calorimetric curves for PS and ethylene, C_2H_4, under various pressures can be seen in Figure 43.[83]

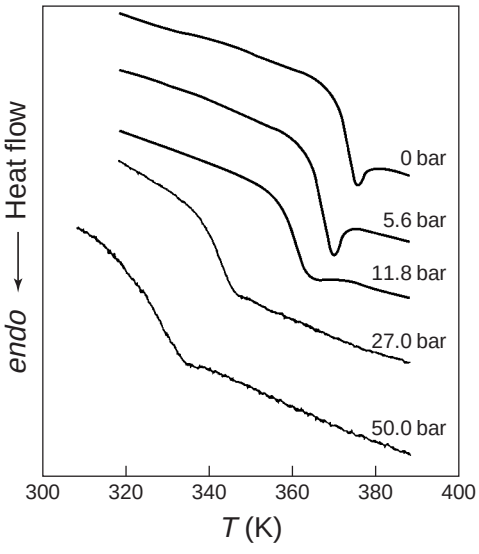

Figure 43 Some representative calorimeter outputs for PS–C_2H_4 scanned under various pressures (O'Neill).

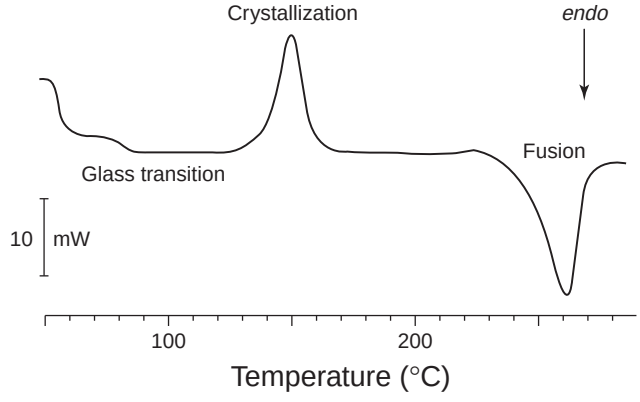

Figure 44 DSC curve of the T_g, crystallization and fusion of a PET sample quenched in liquid nitrogen (Wiedemann).

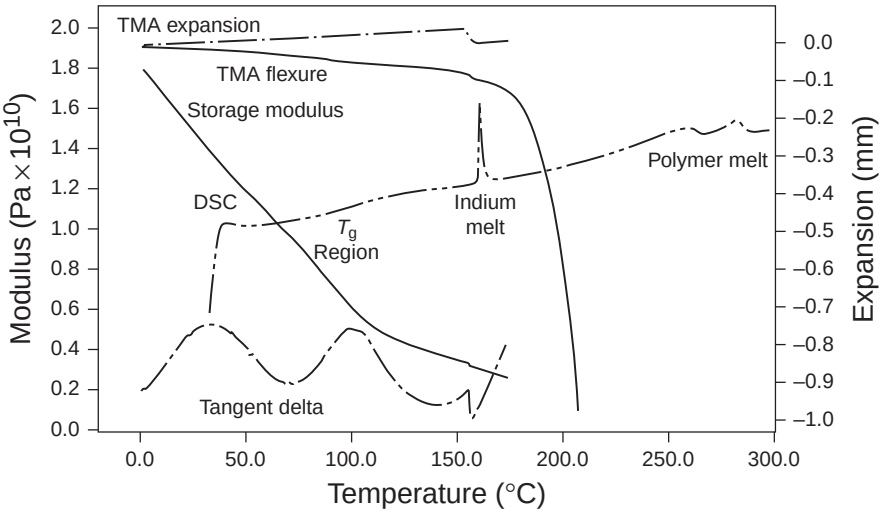

Figure 45 Thermal curves of a copolyester liquid crystal polymer, showing the use of indium as an internal standard (Cassel and Riga).

3.9 Polymer T_g, Crystallization and Fusion

Wiedemann[84] presented various thermal techniques in measuring the T_g in polymers, including a DSC curve of the T_g, crystallization and fusion of PET quenched in liquid nitrogen, Figure 44.

3.10 Liquid Crystal Polymer T_g and Internal Reference

Cassel and Riga[85] contributed a thermal curve of a copolyester liquid crystal polymer, showing the use of indium as an internal reference standard, Figure 45.

3.11 Polymer Film T_g and Thermal History

Moscato and Seyler[86] studied PET films oriented uniaxially without constraint: see Figure 46, first thermal cycle and second thermal cycle.

3.12 Effect of Inhibitors on Vinyl Monomer Polymerization

DSC[87,88] has been used to study the polymerization of vinyl monomers, acrylamides and acrylates and the effect of free radical inhibitors. Selection of the optimum inhibitor and concentration was based on the exothermic autopolymerization of the monomers. The autopolymerization temperature was proportional to the concentration of a given inhibitor, for example, *para*-methoxyphenol or *t*-butyl pyrocatechol, Figure 47.

3.13 Differential Scanning Calorimetry "Fingerprint" Curve Identifies Unknown

If you have an unknown material, a polymer or an excipient and drug mixture, DSC can give you vital physical and chemical information about it. In general, a DSC curve has many clues that will lead you on

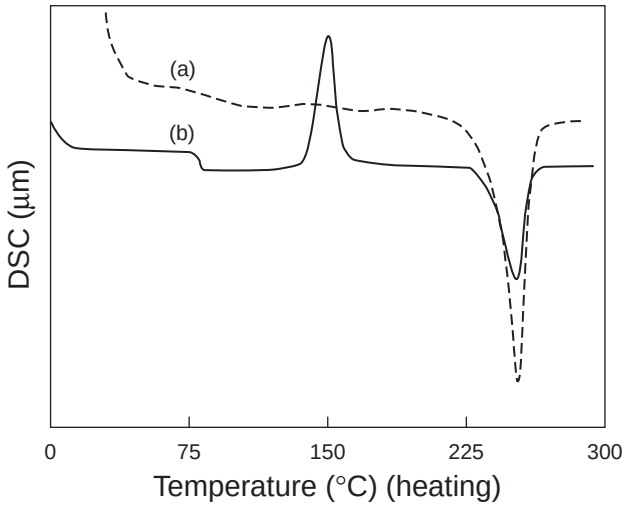

Figure 46 DSC curve of PET film oriented uniaxially with constraint (a) first thermal cycle and (b) second thermal cycle (Moscato and Seyler).

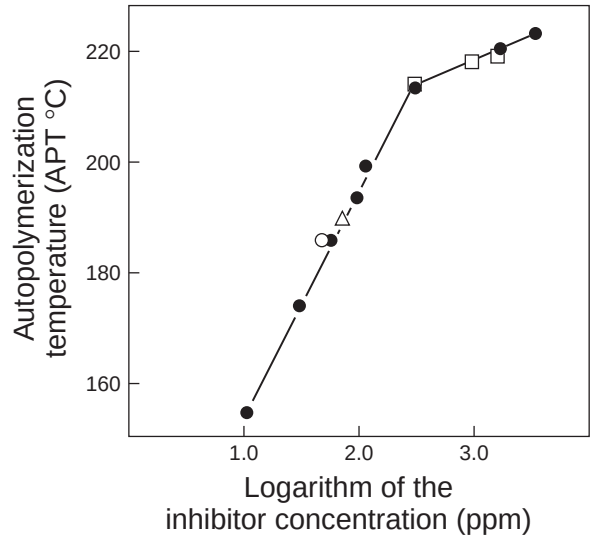

Figure 47 Effect of inhibitor concentration on the autopolymerization temperature of esters of acrylic acid. ● = isodecyl acrylate + *p*-methoxyphenol, = isodecyl acrylate + 2,6-di-*t*-butyl-*p*-cresol, △ = *n*-butylacrylate + *p*-methoxyphenol, ○ = 2-ethylhexylacrylate + *p*-methyoxyphenol.

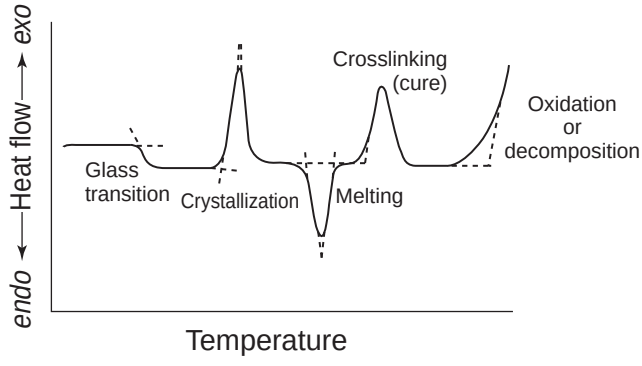

Figure 48 Schematic DSC thermal curve.

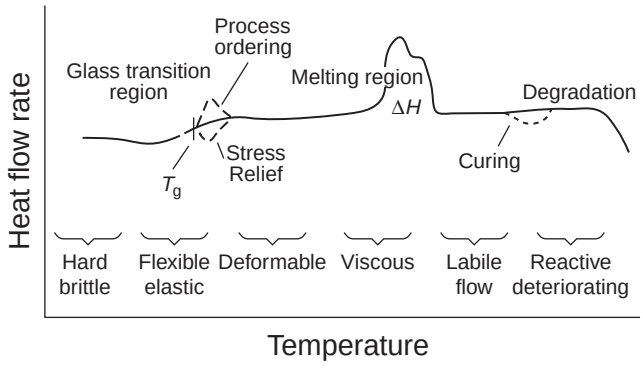

Figure 49 Polymer characterization by DSC.

a path to identification or a better understanding of your material. The DSC, in summary, can supply the T_g, crystallization temperature, heat of fusion and/or crystallization, crosslinking (cure) and oxidation or decomposition, Figure 48.

If you have a failed polymer, a DSC curve can reveal the T_g and melt temperatures, but can also suggest prior history of the part, Figure 49.

Endothermic process ordering or exothermic stress relief can be obtained from data surrounding the T_g. From low temperatures, below the T_g, a material is hard and brittle and above the T_g flexible and elastic. Upon further heating, the polymer is deformable, viscous and finally it reactively deteriorates or degrades. The T_g of thermoplastics, PS and PC are distinguishable from the elastomeric T_g of PBD and SBR, Figure 50. Two immiscible polymers, with a T_g difference between the two homopolymers or copolymers of greater than 10 °C, have two T_g values, Figure 51.

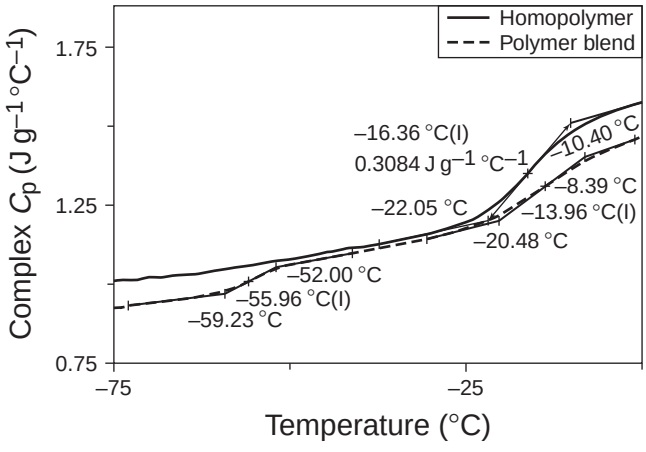

Figure 51 Immiscible polymer blend versus homopolymer by MTDSC.

3.14 Curing and Polymerization of Thermoset Polymers

Incomplete polymerization of an epoxy resin, thermoset polymer, can yield a lower than expected T_g, for example, 77.1 °C, Figure 52. However, continued heating of the polymer reveals an exothermic polymerization at 149 °C with an exothermic heat of 20.4 J g^{-1}. Upon cooling and reheating the epoxy had a higher T_g at 110.5 °C.[89] The glass transition temperature of another thermoset polymer varied with cure time and temperature, Figure 53. It is obvious from the plot that the polymer is cured in 20–40 min (plateau in the T_g versus cure time curve) at 175 °C, and does not reach cure even at 70 min at 140 °C.

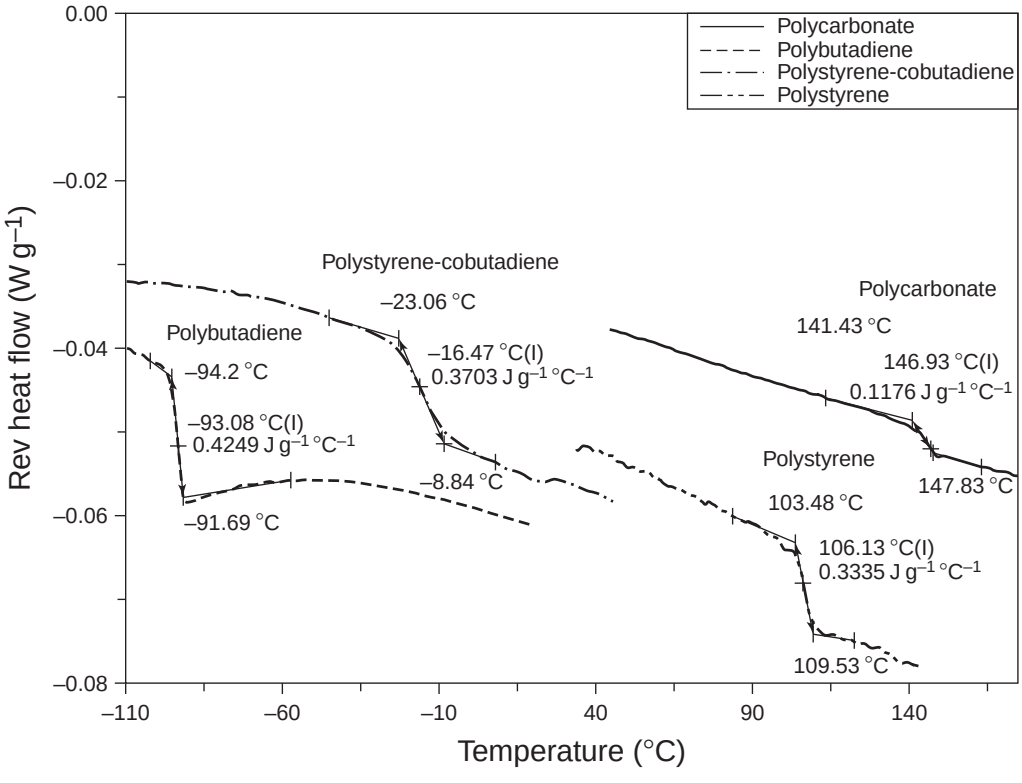

Figure 50 Polymer glass transition temperature by MTDSC.

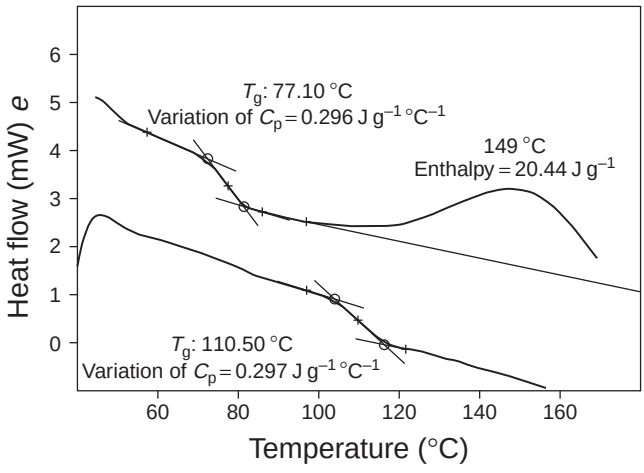

Figure 52 T_g and polymerization of an epoxy resin.

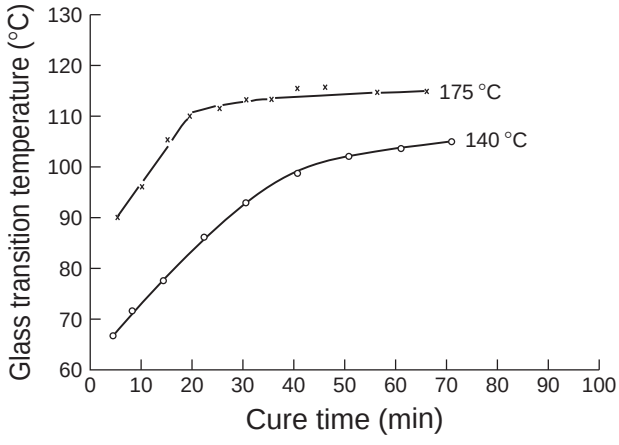

Figure 53 Variation of glass transition temperature with cure time and temperature.

3.15 Polymer Fabrics Identified by Differential Scanning Calorimetry

Hall and Cassel[90] identified fabric material by DSC. Rayon, cotton and wool underwent dehydration and degradation reactions at different temperatures. The DSC "fingerprint" curves clearly distinguished a polyester, polyester–cotton blend and cotton, Figure 54.

3.16 Pharmaceutical Applications: Purity, Polymorphism, Excipients and Lubrication

The DSC is also a problem-solving tool in pharmaceutical applications.[89] First, the purity of $C_9H_9BrO_3$ is determined from the melting peak and software using the Van't Hoft melting depression equation. This chemical was found to be 99.43% pure, heat of fusion was 90.9 J g^{-1} and the melting temperature was 114.8 °C, Figure 55. ASTM E-0928 is the standard test method for mole percent impurity by DSC.[91] This protocol which is based

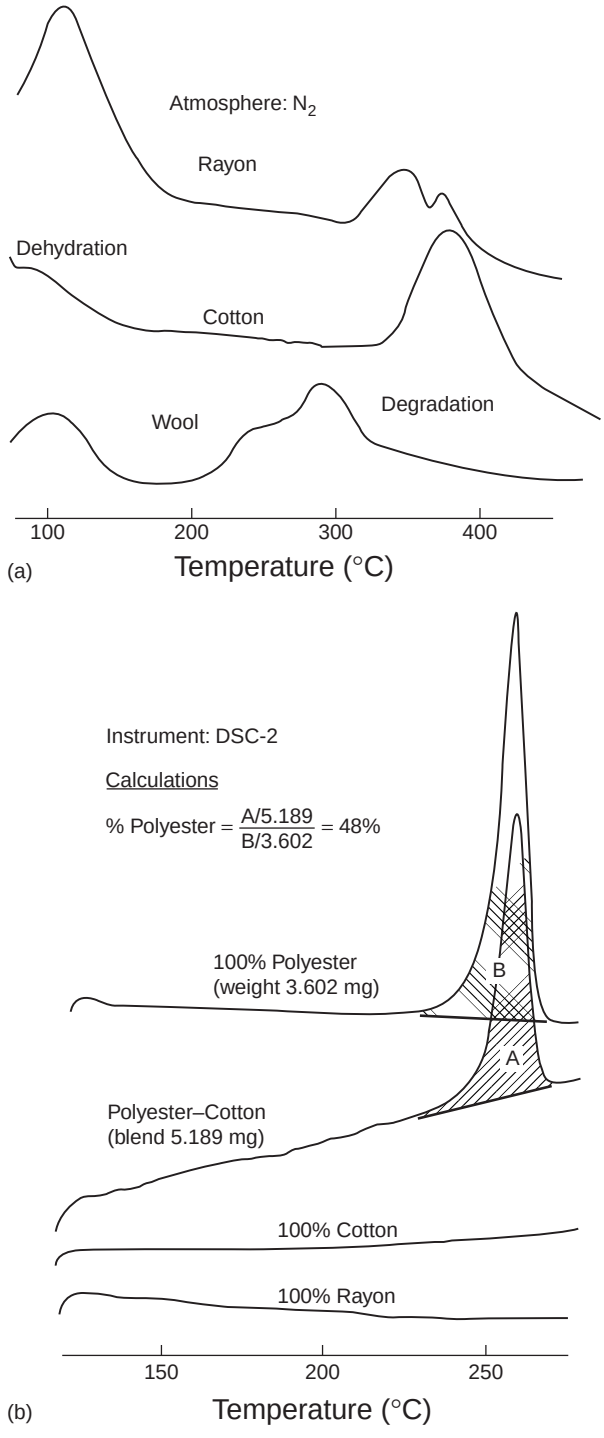

Figure 54 Identification of (a) fabric materials by DSC and (b) blend samples. (After Cassell and Hall.[90])

on the melting temperature range of a compound, broadens as the impurity level rises. Results obtained from this method include sample purity (in mol%), enthalpy of fusion (in J mol^{-1}) and the melting temperature (in Kelvin) of the pure form of the major component. The

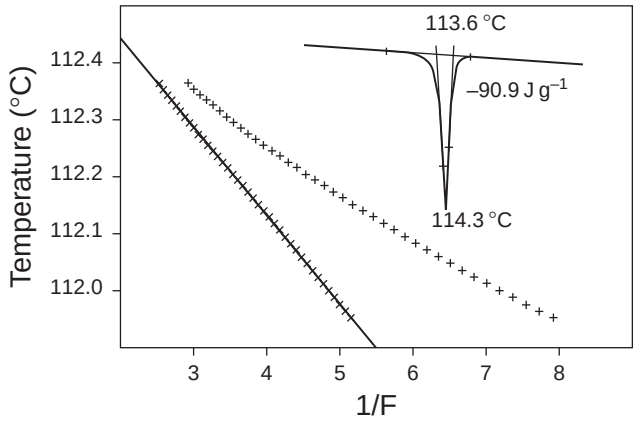

Figure 55 Determination of purity of $C_9H_9BrO_3$.

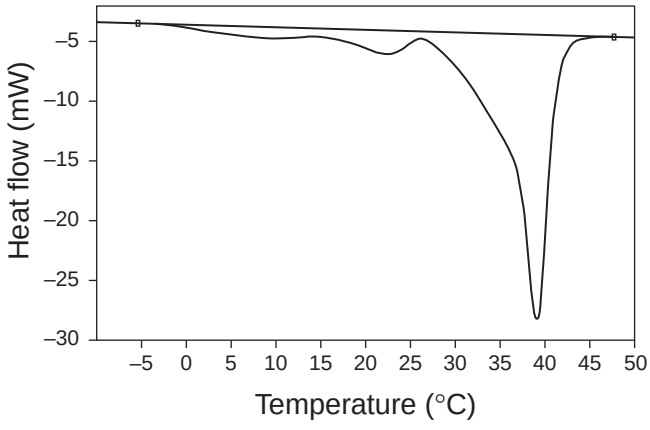

Figure 57 Melting of suppository excipient.

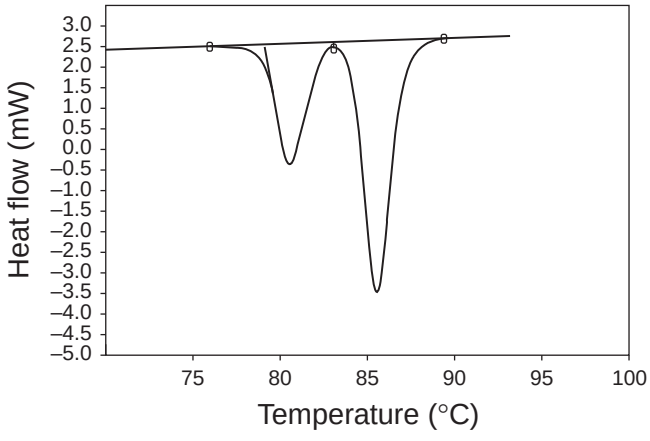

Figure 56 DSC separation of pharmaceutical polymorphs.

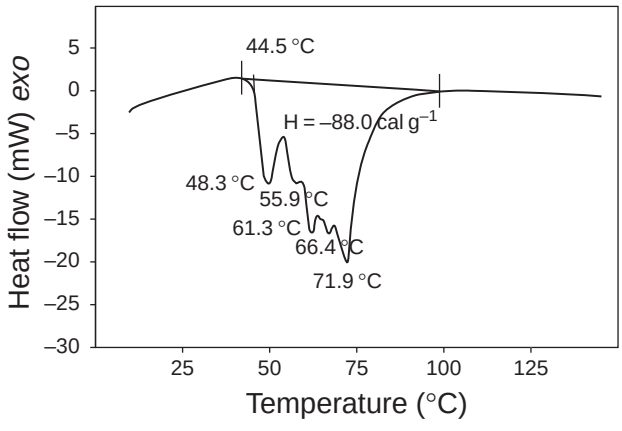

Figure 58 DSC analysis of pharmaceutical lubricants.

single instrument repeatability of purity at a 95% confidence level is that two results should be considered suspect if they differ by more than 0.19 mol%. The multiple instrument reproducibility at a 95% confidence level is that two averages should be considered suspect if they differ by more than 0.72 mol%.

Another pharmaceutical DSC application is the separation of polymorphs.[89] A 4.8-mg sample was heated at 2 °C min^{-1} from room temperature to 100 °C. Two endothermic peaks are seen at 80 °C and 85 °C corresponding to the transition of each polymorph. The enthalpy of polymorph 1 is 141.0 J g^{-1} and polymorph 2 is 78.2 J g^{-1}, Figure 56. The approximate composition of the mixture is 64% polymorph 1 and 36% polymorph 2. The actual composition was 52% polymorph 1.

The melting of a pharmaceutical suppository excipient (42.3 mg) was measured from −20 to 55 °C at 2 °C min^{-1}.[89] The thermal curve shows a very broad endothermic peak of melting from 0 to 45 °C. The multiple peaks are due to the melting of the different fatty compounds in the composition of the suppository excipient, see Figure 57.

DSC analysis of a pharmaceutical lubricant is seen in Figure 58. The sample, Sterotex, with a mass of 7.40 mg was heated from ambient to 150 °C at 10 °C min^{-1}. A broad endothermic peak occurred from 44 to 100 °C with an enthalpy of fusion of 88 cal g^{-1}. Superimposed on the broad peak were five additional peaks corresponding to at least five melting phenomena. The melting peaks at 48, 56, 61, 66, and 72 °C correspond to melting of different fatty compounds in the lubricant. If the components were paraffin in nature then the melt temperatures correspond to compounds with greater than 30 carbon atoms or a molecular mass of about 400 g mol^{-1}.

3.17 Food Application: Cocoa Butter

A food application of DSC is the melting analysis of cocoa butter, Figure 59.[89] An 11.1-mg sample was first heated to 50 °C at 10 °C min^{-1}, then cooled to −10 °C at 5 °C min^{-1} and finally heated to 60 °C at 5 °C min^{-1}. This DSC scheme of heating–cooling and reheating

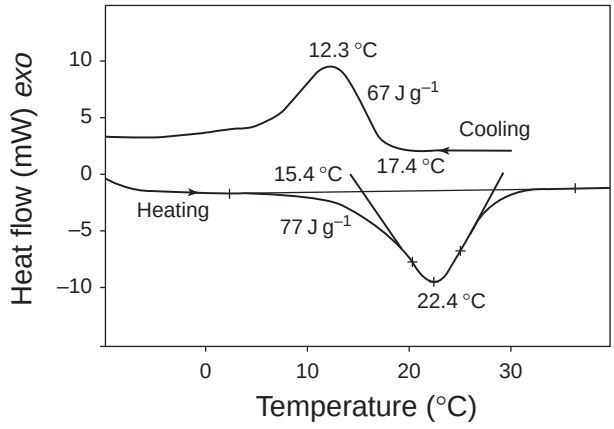

Figure 59 Analysis of cocoa butter by DSC.

was to impart a known thermal history to the cocoa butter. The cooling experiment shows an exotherm of crystallization, 67 J g^{-1} at 17.4–12.3 °C. The final heating cycle had a melt range of 15.4–22.4 °C with a 77 J g^{-1} heat of fusion. The coca butter slightly supercooled based on different melt and crystallization temperatures. The melting crystallization is almost reversible with the heats of crystallization and fusion being 67 and 77 J g^{-1}, respectively. Melting and crystallization heats and temperatures have been observed to be the same, approaching equilibrium values, when the ramp rate is ≅1.0–2.0 °C min^{-1}.[77]

3.18 Biochemical Application: Ribonuclease

A biochemical use of DSC is the denaturation of ribonuclease, Figure 60.[8] The sample is ribonuclease A from bovine pancreas. A solution of 2% of the ribonuclease is prepared in 0.2 mol L^{-1} of glycine and 0.2 mol L^{-1} of NaCl. The sample mass is 17.7 mg and scanned at 0.3 °C min^{-1} from 25 to 95 °C. The endothermic denaturation starts at 49.2 °C with a peak maximum at 56.1 °C and an enthalpy of reaction of 389 kJ mol^{-1}. The specific heat capacity variation measured before and after the denaturation is 3.3 kJ mol^{-1} °C^{-1}.

3.19 Hazardous Material: Aromatic Nitro Compound

Decomposition of an aromatic nitro command is easily detected by a DSC curve, Figure 61.[89] A stainless steel container was used with a 5.4-mg sample. The ramp rate was 2.0 °C min^{-1} from ambient to 330 °C. The nitro compound melts below 50 °C and exothermically decomposes from 200 to 273 °C. The kinetics of decomposition can be estimated when the ramp rate is varied. An ASTM test method, E-0698 can be used to determine Arrhenius kinetic constants for thermally unstable materials by DSC or DTA.[92] The Arrhenius parameters combined with the general rate law and the exothermic heat of reaction can be used for the determination of thermal explosion hazards. The accuracy of the kinetic constants is checked by an isothermal aging procedure. A difference of less than 10% relative between predicted and observed results, can be considered a reasonable test of accuracy.

3.20 Characterization of Inorganic Compounds

The dehydration of calcium sulfate hydrate, $CaSO_4 \cdot 2H_2O$ can be resolved by DSC, Figure 62.[89] A 34.7-mg sample is heated in a closed pan and lid system. The sample is heated at 5 °C min^{-1} from ambient to 250 °C. The dehydration occurs in two steps. In the first step $CaSO_4 \cdot 1/2H_2O$, hemihydrate is formed with the loss of 1.5 mol of H_2O. In the second step, $CaSO_4$, anhydrite is formed with the loss of 1/2 mol of H_2O. TGA of mass loss as a function of temperature or time, can further study this application.

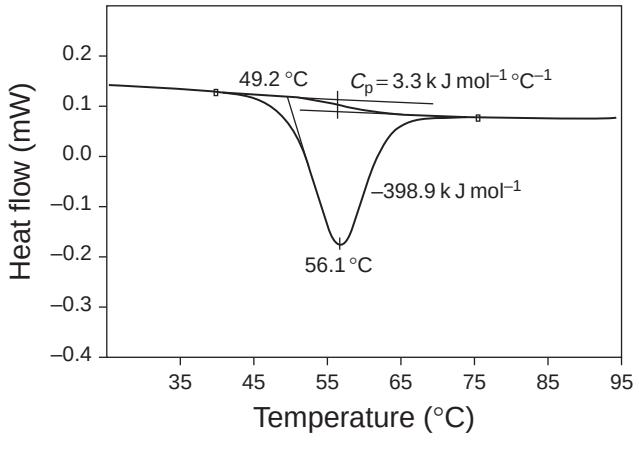

Figure 60 Denaturation of ribonuclease.

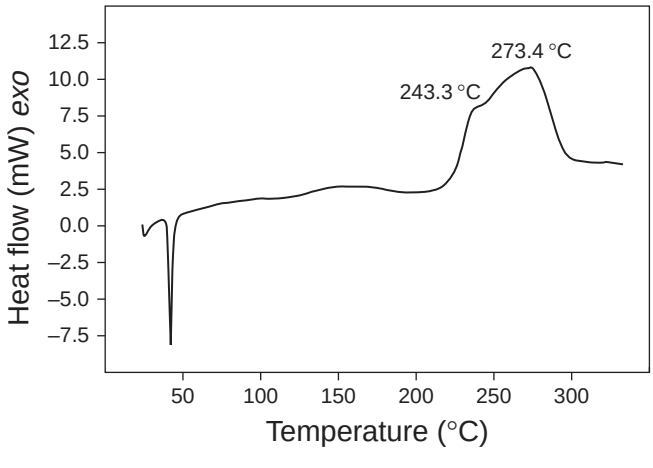

Figure 61 Decomposition of an aromatic nitro compound.

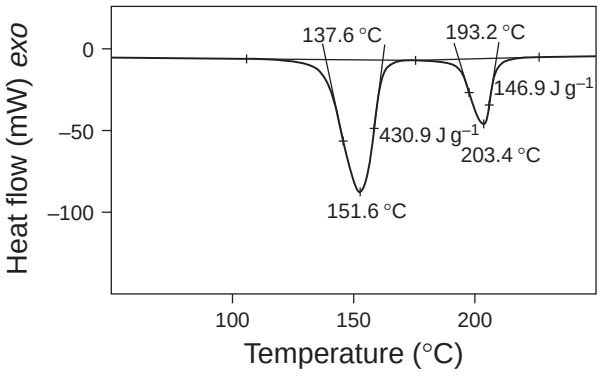

Figure 62 Dehydration of calcium sulfate dihydrate by DSC.

4 SUMMARY

DSC and DTA are invaluable research, quality control and performance measuring tools. It is easy to run a DSC or DTA experiment and interpretation is usually straightforward. There is an increase in the use of thermal analytical instruments in undergraduate and graduate college and university laboratories. This will result, eventually, in an increased use of DSC, MTDSC and DTA in industrial applications.

5 RESOURCES

The North American Thermal Analysis Society and the International Congress on Thermal Analysis and Calorimetry are resources for scientists seeking a high level of understanding and use of the many thermal analysis methods. ASTM Committee E-37 on Thermal Measurements is the source of standardized thermal analysis procedures, as well as new method development.

6 AMERICAN SOCIETY FOR TESTING AND MATERIALS METHODS USED IN DIFFERENTIAL SCANNING CALORIMETRY AND DIFFERENTIAL THERMAL ANALYSIS

- **D-3417**, 08.02, Standard Test Method for Enthalpies of Fusion and Crystallization of Polymers by DSC – This test method covers the determination of heat of fusion and heat of crystallization of polymers by DSC. It is applicable to polymers in granular form or to any fabricated shape from which appropriate specimens can be cut. The normal operating temperature range is from the cryogenic region to 600 °C.
- **D-3418**, 08.02, Standard Test Method for Transition Temperatures of Polymers by Thermal Analysis – This test method covers determination of transition temperatures of polymers by DTA and DSC. First-order transition temperatures or melting points of semicrystalline polymers may also be determined or approximated by the procedures found in D-2117 and D-789. The normal operating temperature range is from cryogenic region to 600 °C.
- **D-4591**, 08.03, Test for Temperatures and Heats of Transitions by DSC (fluoropolymers) – This test method defines conditions for the use of DSC with fluoropolymers: polytetrafluoroethylene, polyvinylidene difluoride, PCTFE and PVF and their copolymers. This test method is applicable to the analysis of powders as well as samples taken from semifinished or finished products.
- **D-5028**, 08.03, Test for Curing Properties of Pultrusion Resins by DSC – This test method covers determinations of curing parameters of pultrusion resins by DSC. The normal operating temperature range is from 0 to 200 °C.
- **D-3895**, 08.02, Test for OIT of Polymeric Materials by DSC – This test method outlines a procedure for the determination of the OIT of polymeric materials by DSC. It is applicable to polyolefin resins that are in a fully stabilized/compounded form.
- **D-4816**, 05.03, Test for Specific Heat Capacity of Aircraft Turbine Lubricants by DSC.
- **D-5483**, 05.03, Test for OIT of Lubricating Greases by PDSC.
- **D-4419**, 05.02, Test for Transition Temperatures of Petroleum Waxes by DSC.
- **E-0472**, 14.02, Standard Practice of Reporting Thermoanalytical – This practice is for general use in reporting experimental information from DTA, TGA, evolved gas analysis or detection and TMA studies. It incorporates laboratory practice with some of the specific needs in thermal analysis. These specific needs account for possible variations of the observed curve with the several parameters reported. Without changing the observed data, reporting in full the conditions under which the data were taken will enable another worker to reconcile differences that may be apparent in another study.

  Accompanying each TMA record should be: identification of substances, a statement of source of all substances, a statement of the temperature environment of the sample, identification of the sample atmosphere, statement of geometry, dimensions and materials of sample holder and identification of the abscissa scale in terms of time or temperature.

- **E-0473-94**, 14.02, Standard Definitions of Terms Relating to Thermal Analysis – This standard is a compilation of definitions used in thermal analysis. A definition is a single sentence with additional information included in the notes. It is reviewed every five years.
- **E-0537**, 14.02, Thermal Stability by DTA and DSC.
- **E-0698**, 14.02, Test for Arrhenius Kinetic Constants of Thermally Unstable Materials using DTA and DSC.
- **E-0793-95**, 14.02, Standard Test Method for Heats of Fusion and Crystallization by DSC – This test method covers the determination of the enthalpy (heat) of fusion (melting) and crystallization by DSC. This method is applicable to solid samples in granular form or in any fabricated shape from which an appropriate specimen can be cut, or to a liquid sample that crystallizes within the range of the instrument. The normal operating range is from −120 to 600 °C. This method is applicable to thermally stable materials with well-defined endothermic or exothermic behavior.
- **E-0794-95**, 14.02, Standard Test Method for Melting and Crystallization Temperature by Thermal Analysis (DSC and DTA) – This test method covers the determination of melting and crystallization temperatures of pure materials by DSC and DTA. It is applicable to thermally stable materials with well-defined melting temperatures.
- **E-0928**, 14.02, Standard Test Method for Mole Percent Impurity by DSC – This method covers the determination of purity by use of DSC and evaluation of the results using the van't Hoff equation. Determination of purity by this method is only applicable when the impurity dissolves in the melt and is insoluble in the crystal.
- **E-0967**, 14.02, Standard Practice for Temperature Calibration of DSC and DTA analyzers – This practice covers the temperature calibration of DSC and DTA analyzers over the temperature range −40 to +2500 °C.
- **E-0968**, 14.02, Standard Practice for Heat Flow Calibration of DSC – This practice covers the heat flow calibration of differential scanning calorimeters over the temperature range from −130 to +800 °C.
- **E-1142**, 14.02, Terminology Relating to Thermophysical Properties – This is a compilation of terms and corresponding definitions commonly used in the study of thermophysical properties. Terms that are generally understood or defined adequately in other readily available sources are either not included or their sources are identified.
- **E-1269-95**, 14.02, Test Method for Determining Specific Heat Capacity by DSC – This method covers the determination of specific heat capacity by DSC. It is applicable to thermally stable solids and liquids. The normal operating range of the test is from −100 to 600 °C.
- **E-1356-95**, 14.02, Test Method for Glass Transition Temperatures by DSC or DTA – This method covers the determination of the glass transition temperature, T_g, of materials using DSC or DTA. It is applicable to amorphous materials or to partially crystalline materials containing amorphous regions that are stable and do not undergo decomposition or sublimation in the T_g region.
- **E-1782-96**, 14.02, Test Method for Determining Vapor Pressure by Thermal Analysis – This method covers a procedure for the determination of the vapor pressure of pure liquids or melts from boiling measurements made using DTA or DSC instrumentation operated at different applied pressures. It may be used in the temperature range 273–773 K (0–500 °C) and for pressures between 5 and 2 MPa.
- **E-1858-97**, 14.02, Test Method for Determining OIT of Hydrocarbons by DSC (PDSC) – This test method covers the procedure for determining the oxidative properties of hydrocarbons by DSC or PDSC and is applicable to hydrocarbons that oxidize exothermically in their analyzed form.

  This standard reviews two test methods: A, a DSC is used at ambient pressure, for example, about 100 kPa of oxygen and B, a PDSC is used at high pressure, for example, 3.5 MPa (500 psig) oxygen.

  The test specimen in an aluminum pan and the reference aluminum pan are heated to a specified constant test temperature in an oxygen environment. Heat from the specimen is monitored at an isothermal temperature until the oxidative reaction is manifested by heat evolution on the thermal curve. The OIT, a relative measure of the oxidative stability at the test temperature, is determined from data recorded during the isothermal test. The OIT measurement is initiated upon reaching the isothermal test temperature.
- **E-1860-97**, 14.02, Test Method for Elapsed Time Calibration of Thermal Analyzers – This method covers the calibration or performance confirmation of the elapsed time signal from thermal analyzers.
- **E-1952-98**, 14.02, Test Method for Thermal Conductivity and Deriving Thermal Diffusivity by Modulated Temperature DSC.
- **E-2009-99**, 14.02, Test OOT of Hydrocarbons by DSC – This method covers the procedure for

determining the oxidative properties of hydrocarbons by DSC or PDSC and is applicable to hydrocarbons which oxidize exothermically in their analyzed form. In test method A, a DSC is used at ambient pressure, for example, 0.1 MPa or 1 atm of oxygen. In test method B, a PDSC is used at high pressure, for example, 3.5 MPa (500 psig) oxygen.

ABBREVIATIONS AND ACRONYMS

ABS	Acrylonitrilebutadiene Terpolymer
ASTM	American Society for Testing and Materials
DDC	Dynamic Differential Calorimetry
DDSC	Dynamic Differential Scanning Calorimetry
DEA	Dynamic Enthalpic Analysis
DETA	Dielectric Thermal Analysis
DMA	Dynamic Mechanical Analysis
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
EVA	Ethylene Vinyl Acetate Copolymer
FTIR	Fourier Transform Infrared
HDPE	High-density Polyethylene
ISO	International Standardization Organization
LDPE	Low-density Polyethylene
MTDSC	Modulated Temperature Differential Scanning Calorimetry
NIST	National Institute of Standards and Technology
OIT	Oxidation Induction Time
OOT	Oxidation Onset Temperature
PC	Polycarbonate
PBD	Polybutadiene
PDSC	Pressure Differential Scanning Calorimetry
PET	Poly(ethylene terephthalate)
PP	Polypropylene
PRT	Platinum Resistance Thermocouple
PS	Polystyrene
PVAc	Polyvinyl Acetate
QDTA	Quantitative Differential Thermal Analysis
SBR	Polystyrene-cobutadiene
SEM	Scanning Electron Microscopy
TGA	Thermogravimetric Analyzer
TMA	Thermomechanical Analysis
XDS	X-ray Diffraction Structure

RELATED ARTICLE

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction

REFERENCES

1. ASTM E-0473, Definitions of Terms Relating to Thermal Analysis, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1998.
2. B. Wunderlich, 'The Basis of Thermal Analysis', in *Thermal Characterization of Polymeric Materials*, ed. E. Turi, Academic Press, New York, Vol. 1, 1997.
3. B. Prime, 'Thermosets', in *Thermal Characterization of Polymeric Materials*, ed. E. Turi, Academic Press, New York, Vol. 2, 1997.
4. R. Chartoff, 'Thermoplastic Polymers', in *Thermal Characterization of Polymeric Materials*, ed. E. Turi, Academic Press, New York, Vol. 2, 1997.
5. P. Gallagher, 'Thermoanalytical Instrumentation', in *Thermal Characterization of Polymeric Materials*, ed. E. Turi, Academic Press, New York, Vol. 1, 1997.
6. H. Bair, 'Thermal Analysis of Additives in Polymers', in *Thermal Characterization of Polymeric Materials*, ed. E. Turi, Academic Press, New York, Vol. 2, 1997.
7. E. Turi, *Thermal Characterization of Polymers*, Academic Press, CA, 1981.
8. P. Slade, L. Jenkins (eds.), *Thermal Characterization Techniques*, Marcel Dekker, New York, 1970.
9. W. Wendlandt, *Thermal Analysis*, 3rd edition, John Wiley and Sons, New York, 1986.
10. R. Blaine, C. Schoff (eds.), *Purity Determination by Thermal Methods*, ASTM STP 838, West Conshohocken, PA, 1984.
11. M.J. O'Neill, J. Justin, N. Brenner, *Anal. Chem.*, **47**, 4, 630–637 (1964).
12. A. Gray, *American. Lab.*, **3**, 43 (1971).
13. H. Bair, *Polym. Eng. Sci.*, **10**, 247 (1970).
14. M. Casseino, D. Blevins, R. Sanders, *Thermochim. Acta*, **284**, 145 (1996).
15. S. Dyszel, *Thermochim. Acta*, **282**, 103 (1996).
16. S.St.J. Warne, *Thermochim. Acta*, **272**, 1 (1996).
17. D. Feldman, D. Banu, *Thermochim. Acta*, **272**, 243 (1996).
18. P. Navard, J. Haudin, *J. Therm. Anal.*, **30**(1), 61 (1985).
19. I. Kotula, A. Rabczu, *J. Therm. Anal.*, **30**(1), 95 (1985).
20. J. Sestak, B. Stepanek, *J. Therm. Anal.*, **43**(2), 371–376 (1995).
21. W. Hemminger, G. Hohne, *Calorimetry*, Verlag-Chemie, Basel, 1994.
22. R. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, 1994.
23. Perkin Elmer, Thermal Analysis Newsletter No. 9, 1970 and Perkin Elmer DSC-2 Manual, Wilton, CT.
24. B. Wunderlich, *Thermal Analysis*, Academic Press, MA, 1990.
25. R. Blaine, *TA-081, The Case for a Generic Definition of DSC*, TA Instruments, New Castle, DE, 1997.
26. 11th International Congress on Thermal Analysis and Calorimetry, 1996, Akademia Kiado RT Publishers, Budapest.
27. S.L. Boersma, *J. Am. Ceram. Soc.*, **38**, 281 (1955).

28. B. Cassel, Perkin Elmer Seminar CD, Materials Analysis in the QC and Analytical Laboratories, 1999.
29. B. Wunderlich, A. Boller, I. Okazaki, S. Kreitmeier, *J. Therm. Anal.*, **47**, 1013–1026 (1996).
30. M. Song, D. Hourston, H. Pollock, F. Schafer, A. Himmiche, *Polymer*, **38**(3), 503–507 (1997).
31. Y. Jin, J. Bonilla, G. Lin, J. Morgan, L. McCracken, J. Carnahan, *J. Therm. Anal.*, **38**(12), 3025–3034 (1997).
32. G. vanEkenstein, G. ten Brinke, T. Ellis, *Polym. Mater. Sci. Eng.*, **76**, 219–220 (1997).
33. D. Hourston, M. Song, A. Hammiche, H. Pollock, M. Reading, *Polymer*, **38**(1), 1–7 (1997).
34. X. Hu, C. Breach, R. Young, *Polymer*, **38**(4), 981–983 (1997).
35. S. Sauerbrunn, L. Thomas, *Am. Lab.*, **27**, 19–22 (1995).
36. L. Thomas, J. Cattiau, M. Chahinian, *Spectra. Anal.*, **26**(195), 31–33 (1997).
37. A. Luget, R. Wilson, *Thermochim. Acta*, **238**, 295–307 (1994).
38. M. Reading, *Trends Polym. Sci.*, **1**, 248–253 (1993).
39. ASTM E-0697, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1997.
40. ASTM E-0968, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1997.
41. ASTM E-1860, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1997.
42. 'Enthalpy of Melting for Standards', TA Instrument Co., Thermal Analysis and Rheology Applications Library, Version 2.5, 1998.
43. W. Eysel, K.H. Breuer, *Thermochim. Acta*, **57**, 317 (1982).
44. R. Hultgren, et al., *Selected Values of Thermodynamic Properties of the Elements*, American Society for Metals, Metals Park, OH, 1973.
45. G.T. Furukawa, R.E. McCoskey, G.J. King, *J. Res. Natl. Bur. Stand. (US)*, **47**, 256 (1951).
46. G.T. Furukawa, D.C. Ginnings, et al., *J. Res. Natl. Bur. Stand. (US)*, **46**, 195 (1951).
47. D.R. Stull, et al., *The Chemical Thermodynamics of Organic Compounds*, Wiley, New York, 367, 1969.
48. *Pure Appl. Chem.*, **48**, 506 (1976).
49. J.L. Margrave, *France Centre National de la Recherche Scientifique, Colloques Int.*, **205**, 71–77 (1972).
50. K.A. Gshneider, Jr, *Solid State Physics*, Academic Press, New York, Vol. 16, 1964.
51. O. Kubaschewski, E.L. Evans, *Metallurgical Thermochemistry*, Pergamon, London, 1974.
52. J.G. Aston, et al., *J. Am. Chem. Soc.*, **65**, 341 (1943).
53. J.G. Aston, et al., *J. Am. Chem. Soc.*, **65**, 1035 (1943).
54. H.L. Finke, et al., *J. Am. Chem. Soc.*, **76**, 33 (1954).
55. A.A. Schaerer, et al., *J. Am. Chem. Soc.*, **33**, 2017 (1955).
56. ASTM E-1858-97, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1998.
57. ASTM E-2009-99, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1999.
58. A. Riga, G. Patterson, 'Controlling Composition and Oxidation Tests of Hydrocarbon and Synthetic Fluids in a Thermal Analysis Laboratory', *Thermochim. Acta*, **243**, 277–288 (1994).
59. A. Riga, 'Oxidative Behavior of Materials by Calorimetric Methods', Plenary Lecture, Calorimetry Conference, Tallahassee, FL, 1999.
60. A. Riga, G. Patterson, *Thermochim. Acta*, **226**, 201–210 (1993).
61. A. Riga, et al., *Thermochim. Acta*, **324**, 135–149 (1998).
62. ASTM E-1356, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1998.
63. ASTM E-0794, 14.02 American Society for Testing and Materials, West Conshohocken, PA, 1997.
64. S. Aubuchon, Second Annual NATAS Follow Up Conference, MTDSC Fundamentals, 1998.
65. A. Riga, D. Brentgardner, ASTM Technical Professional Training, 'Instrumental Analysis of Polymers', West Conshohocken, PA, May, 1999.
66. Thermal Applications Note TN-45B, 'Choosing Conditions in MTDSC', TA Instruments, New Caste, DE, CD Ram, 2.5, 1997.
67. JMP 3.2, Professional Edition, SAS Institute, Cary, NC, 1997.
68. G. Box, W. Hunter, J. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.
69. ASTM E-1356, 14.02, 'Standard Test Method of Glass Transition Temperatures by DSC or DTA', American Society for Testing and Materials, West Conshohocken, PA, 1998.
70. B. Cassel, Perkin Elmer Corporation, Norwalk, CT, Private Communication, 1999.
71. J. Schawe, G. Hohne, *J. Therm. Anal.*, **46**, 893–904 (1996).
72. J. Foreman, Mettler-Toledo Corporation, Worthington, OH, Private Communication, 1999.
73. A. Riga, ASTM Technical Professional Training, 'Instrumental Analysis of Polymers', West Conshohocken, PA, May, 1999.
74. ASTM E-1269, 14.02, 'Standard Test Method of Specific Heat Capacity by DSC', American Society for Testing and Materials, West Conshohocken, PA, 1998.
75. B. Wunderlich, A. Boller, I. Okazaki, *Thermochim. Acta*, **284**, 1 (1996).
76. I. Hatta, S. Muramatsu, *Jpn. J. Appl. Phys.*, **35**, L858–L860 (1996).
77. A. Riga, *Thermochimica Acta*, **324**, 151–163 (1998).
78. Y. Xie, W. Xie, W.-P. Pan, A. Riga, K. Anderson, *Thermochim. Acta*, **324**, 123–133 (1998).
79. A. Riga, D. Young, G. Mlachak, P. Kovach, 'Thermoanalytical Evaluation of Readily Available Reference Polymers', *J. Thermal. Anal.*, **49**, 425–435 (1997).
80. R. Seyler, *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 13–15, 1994.
81. B. Wunderlich, 'The Nature of the Glass Transition and its Determination by Thermal Analysis', *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 17–31, 1994.

82. H. Bair, 'Glass Transition Measurements by DSC', *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 50–74, 1994.
83. M. O'Neill, Y. Handa, 'Plasticization of Polystyrene by High Pressure Gases: a Calorimetric Study', *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 165–173, 1994.
84. H. Wiedemann, G. Widmann, G. Bayer, 'Glass Transition in Polymers: DSC', *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 174–181, 1994.
85. B. Cassel, A. Riga, 'Glass Transition of Liquid Crystal Polymers', *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 202–213, 1994.
86. M. Moscato, R. Seyler, 'Assigning the Glass Transition Temperatures in Oriented PET', *Assignment of the Glass Transition*, ASTM STP 1249, ed. R.J. Seyler, ASTM, West Conshohocken, 239–252, 1994.
87. A. Riga, 'Inhibitor Selection for Vinyl Monomers by DSC', *Polym. Eng. Sci.*, **18**, 836 (1976).
88. A. Riga, 'Thermal Analysis as an Aid to Monomer Plant Design', *Polym. Eng. Sci.*, **15**, 349 (1975).
89. Setaram, Caluire, Cedex, France, Private Communication, 1999.
90. B. Cassel, J. Hall, 'Forensic Applications of DSC', in *Forensic Science*, ed. G. Davies, ACS Symposium Series, 114–133, Vol. 13, 1975.
91. ASTM E-0928, 14.02, American Society for Testing and Materials, West Conshohocken, PA, 1998.
92. D.A. Ditmars, et al., *J. Res. Natl. Bur. Stand.*, **87**(2), 159 (1982).

Inorganic Systems, Thermal Analysis Applications to

Patrick K. Gallagher*
The Ohio State University, Columbus, USA

1 Introduction **1**
2 Properties of Materials **2**
2.1 Thermal Properties 2
2.2 Electrical, Magnetic, and Optical Properties 2
2.3 Dimensional Properties 3
3 Phase Equilibria **4**
3.1 Metals 5
3.2 Ceramics 6
3.3 Others 8
4 Chemical Analysis **8**
4.1 Lithium Niobate 8
4.2 Lead–Tin Solder 8
4.3 Nonstoichiometry 8
4.4 Alloys and Ferrites by Thermomagnetometry 9
4.5 Thin Films 10
5 Reactivity **11**
5.1 Solid-state Reactions 11
5.2 Heterogeneous Kinetics 12
5.3 Influences on Reactivity 13
5.4 Heterogeneous Catalysis 13
5.5 Sintering 15
5.6 Synthesis 16
6 Concluding Remarks **16**
Abbreviations and Acronyms **16**
Related Articles **17**
References **17**

Thermoanalytical techniques can be used to characterize a wide range of inorganic materials and processes concerning them. The topics covered involve the measurement of thermal properties, phase equilibria, chemical analysis, and the study of solid-state reactivity. Examples are selected to include applications to metals, ceramics, minerals and a variety of inorganic compounds. Applications to processes such as catalysis, crystal growth, and synthesis are described.

The principle techniques chosen for these purposes are associated with the measurement of different properties as a function of temperature and time. Thermogravimetry (TG), thermomagnetometry (TM), and evolved gas analysis (EGA) are based on changes in weight and analysis of the materials evolved or consumed. Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) are concerned with changes in temperature and heat capacity. Thermodilatometry (TD) measures changes in length or volume. Examples illustrating the use of each of these techniques are described.

* Present address: Clemson University, Clemson, SC, USA

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

Analysis is taken in its broadest sense to include the characterization of processes and properties as well as the traditional chemical analytical context. In this regard thermoanalytical methods are an exceptionally versatile collection of techniques suitable for a very wide range of applications and materials. Other contributions to this section deal with the instrumentation and methodology associated with the major techniques. Consequently, those aspects will only be considered further when additional information is desired for the specific application under discussion. A substantial number of thermoanalytical applications are in the fields of polymers, pharmaceuticals, and biochemistry. These applications are discussed elsewhere.

Thermal methods encompass a very wide range of techniques and properties, but only the major techniques are considered here. In keeping with this constraint, some modern nontraditional time–temperature programs are not discussed. Approaches such as dynamic temperature[1] and controlled rate[2] techniques are not described in detail, however the reader should be aware of their advantages. Dynamic temperature methods use a nonlinear heating or cooling program that provides periods of both heating and cooling in a prescribed manner. This occasionally enables kinetically reversible events to be resolved from those that are irreversible in either a kinetic or thermodynamic sense. Controlled rate methods utilize feedback from the measurement of the property as a function of temperature to control the temperature program. These methods can offer better resolution of overlapping events and/or a saving in the time required for the experiment.

The range of inorganic materials is large and examples have been selected to include metals, ceramics, organometallics, and traditional inorganic chemicals. Similarly, applications to many of the major technological and scientific processes or equilibria will be considered, for example, catalysis, crystal growth, phase equilibria, thin

films, and so on. Finally, examples have also been chosen to illustrate some of the less common thermoanalytical methods. The organization of subsequent sections is based on the area of application rather than the thermoanalytical technique. This has been done since it is more in keeping with the purpose of this chapter and also because the specific application frequently needs more than one thermoanalytical method to provide the maximum information and benefits.

2 PROPERTIES OF MATERIALS

The discussion is divided into sections on thermal properties, electrical, magnetic and optical properties, and properties associated with surface area and changes in volume that can be investigated through thermal analysis.

2.1 Thermal Properties

DTA and DSC are concerned with the difference in temperature and the exchange of thermal energy between a sample and an inert reference material, as described in detail by Riga in **Differential Scanning Calorimetry and Differential Thermal Analysis**. Hence, DSC is ideally suited for measurements of such thermodynamic properties as heat capacity, thermal conductivity (diffusivity), and the changes in enthalpy associated with chemical reactions and physical transformations. DTA can be used to observe these phenomena in a more qualitative manner.

Riga has described in detail the methodology for the determination of specific heat at constant pressure, C_p, by DSC or modulated temperature differential scanning calorimetry (MTDSC). Values of C_p are of great interest for compilations of thermodynamic data of elements and compounds. Changes in C_p as a function of temperature reveal phase transformations. During first-order phase transformations a discontinuity in C_p occurs and hence a change in enthalpy, ΔH, is attributable to the transformation. Second-order transformations, however, lead only to a change in C_p and are distinguished in this manner. Most ferroelectric, magnetic, and glass transformations fall into this category. Figure 1 illustrates examples of these three second-order transformations.

Repeated scans of the congruent composition of lithium niobate are shown in Figure 1(a).[3] The Curie temperature, T_C, of the transition is indicated. It shows no indication of temperature hysteresis, even at 20 °C min^{-1}. The reproducibility is excellent. First-order and second-order transformations are contrasted in Figure 1(b).[4] The melting of lead is represented by a complete peak and the area under the endothermic peak is proportional to ΔH. The ferromagnetic transformation for nickel

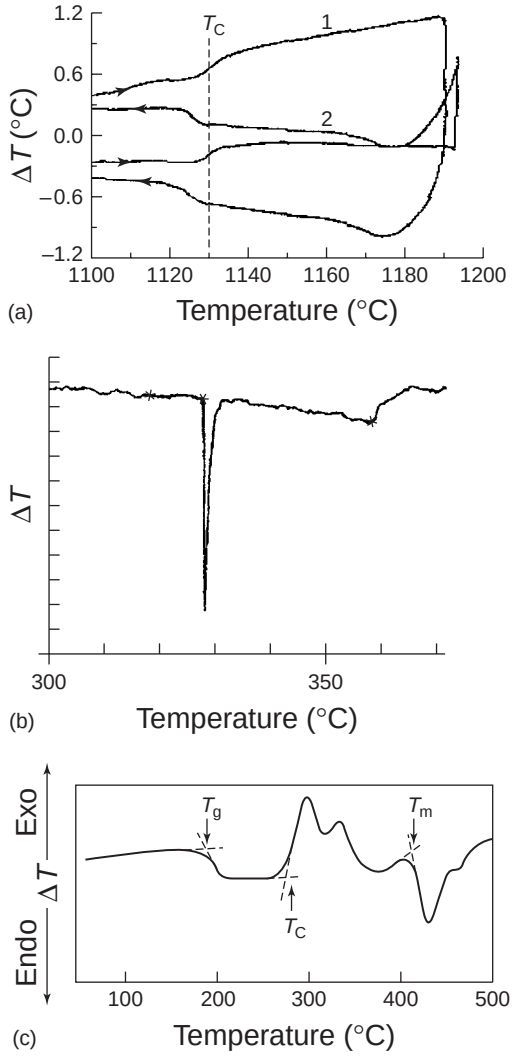

Figure 1 Examples of some second-order transitions: (a) ferroelectric transition in $Li_{1-x}NbO_3$;[3] (b) melting of lead and ferromagnetic transition of nickel;[4] (c) glass transition, crystallization, and melting of a fluoride glass.[5]

is indicated by a step in the curve similar to that in Figure 1(a). Figure 1(c) also presents a DTA scan of multiple transitions.[5] The curve is for a fluoride glass and shows in sequence a second-order glass transition at T_g, multiple exothermic crystallizations starting at T_C, and multiple endothermic meltings commencing at T_m.

Several efforts have been made, with moderate success, to adapt DSC instruments to measure thermal conductivity.[6,7] The newer modulated techniques hold greater promise.[1] Other thermal methods based on laser flash techniques are more common.[8]

2.2 Electrical, Magnetic, and Optical Properties

Thermal analysis is defined as the measurement of a property as it undergoes a programmed temperature

regime. Clearly, such properties include electrical conductivity, dielectric constant, or Seebeck coefficient; magnetization; and optical transmission, reflection, or emission. Instruments have been designed or modified to investigate all of these properties as a function of temperature.[9,10] Of course, the latter classes of spectroscopy can be expanded to include a much broader energy spectrum than the conventional optical range, that is, from γ-rays (Mössbauer spectroscopy) to acoustic waves (thermosonimetry). Wendlandt has pioneered in the field of dynamic as opposed to isothermal measurements in the areas of electrical conductivity,[11] thermovoltametry,[12] and reflection spectroscopy.[13] His textbook[10] can be consulted for numerous examples.

TM has been the topic of a tutorial review.[14] In TM a generally weak magnetic field gradient is superimposed on the sample position in conventional TG. Consequently a ferro- or ferrimagnetic substance, below its magnetic transition temperature (T_C), will react by apparently losing or gaining weight depending upon the direction of the field gradient. As the temperature passes through T_C a dramatic change in the apparent weight will occur.

The purpose of TM is not to determine the magnetic moment or saturation of the material, but simply the qualitative aspect of whether it is magnetic and what its T_C is. Simultaneous DTA and TM can be used to determine the value of T_C very accurately.[14] Figure 2 shows the simultaneous TM and DTA curves for a sample containing very pure specimens of lead, nickel, and zinc that have been sufficiently separated in the sample container to prevent their interaction. The onsets of melting for the lead and zinc can be compared with their known values[15] and used to make a correction line of measured versus true temperatures. The correction for the observed T_C, the final loss of magnetic attraction, can be interpolated from the linear correction plot. This revised value of T_C has now been directly traced to the primary standards which determine the International Temperature Scale of 1990 (ITS90).[15] Consequently materials measured in this manner are proposed as standards for the calibration of the temperature axis in TG. This simultaneous TG/DTA technique is also proposed as the ultimately accurate method of measuring T_C because of its traceability to ITS90.

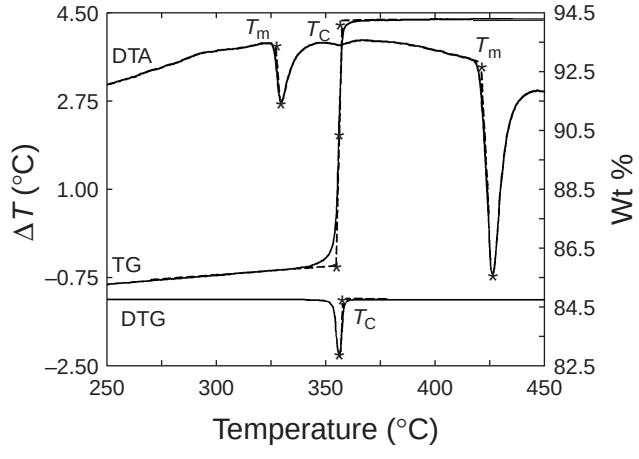

Figure 2 Simultaneous TM and DTA curves for lead, nickel, and zinc metals at 20 °C min^{-1} in Ar.[4] DTG, differential thermal gravimetry.

2.3 Dimensional Properties

The Brunauer–Emmett–Teller (BET) approach of measuring the physisorption of a monolayer of a gas, generally nitrogen, is widely used to determine the surface area of a powder.[16] One of the techniques available to measure this physisorption is TG. The change in mass associated with this process is measured and converted to the number of gaseous molecules adsorbed.[17] A complete adsorption isotherm as a function of partial pressure of the adsorbate can be used to describe the porosity of the material as well.[16]

TD is conventionally used to measure the change in volume, or more often length, of a material as a function of temperature. If there is an effect of atmosphere, such as for grossly nonstoichiometric materials, then it is necessary to control and specify the atmosphere as well as the temperature. The high-temperature superconductors are a good example of this dependence, in this case on the partial pressure of oxygen.[18] Descriptions of dilatometers may be found in most texts on thermal analysis.[8,9,19,20] Conventional instrumentation associated with thermomechanical analysis (TMA), is suitable. A flat-surfaced probe with a small load is used in order not to measure any penetration or softening of the sample. The most common sensor is a linear voltage transformer that provides a signal directly proportional to the change in length, having a polarity indicative of the direction, that is, expansion or contraction. Single-probe measurements measure the change in length relative to the material used to construct the probe and sample support. This is usually fused quartz below 1000 °C and alumina at higher temperatures. There are higher precision models that use two probes and measure the difference in expansion of the sample relative to some standard (having the same length as the sample), such as aluminum, sapphire, tungsten, and so on.

The coefficient of thermal expansion is an important property for many applications, for example, bonding of composite materials, matching thermal expansions of thin films and substrates, maintaining alignment of optical devices, and so on. Results of linear thermal expansion measurements are frequently reported as $\Delta L/L_{298}$, where ΔL corresponds to the change in length from 298 K to the

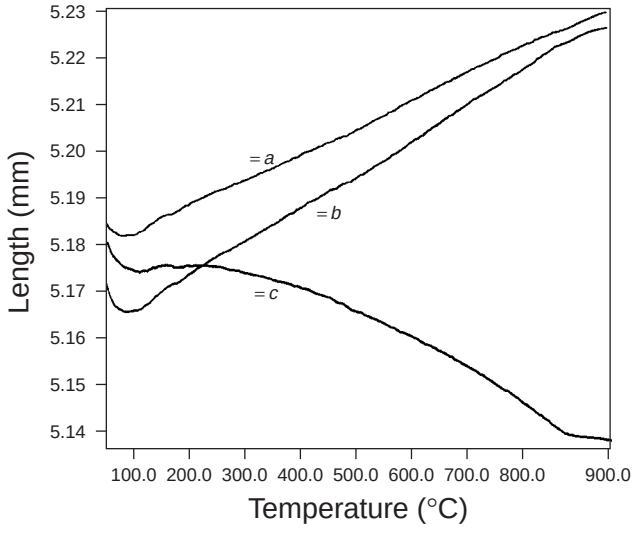

Figure 3 Thermal expansion of $KTiAsO_4$ along various crystallographic axes.[21]

current temperature and L_{298} is the length of the sample at 298 K. This is a dimensionless quantity that is often expressed in ppm (parts per million) or as a percentage. The coefficient of thermal expansion (α) is the derivative of that curve at the temperature of interest and is reported as ppm or percent per degree.

Most composite materials and single crystals have unequal expansion along different directions, that is, they exhibit anisotropy. When measuring polycrystalline solids it is frequently assumed that the crystallographic axes are randomly oriented, so that some general average or volume expansion is measured. Figure 3 shows thermal expansion curves for a highly anisotropic material, $KTiOAsO_4$.[21] The coefficient of thermal expansion is actually negative along the *c*-axis. An interesting aspect of this material is its second-order phase transformation near 850 °C. This is clearly evident as a change in slope of the thermal expansion, particularly along the *c*-axis.

To disseminate the information more accurately and conveniently, data on thermal expansion are generally fitted to a simple third-order polynomial in temperature. The fitting should not be done through a phase transformation because of the concomitant discontinuity or abrupt change in slope. The coefficient of thermal expansion can then be obtained by simply differentiating the polynomial and substituting the desired temperature.

3 PHASE EQUILIBRIA

The phase diagram for a system is the proverbial picture worth ten thousand words. It is the starting point for the characterization of a system, defining the specific phases present at equilibrium as a function of composition, temperature, and pressure. The pressure considered includes not only the total pressure, but also the partial pressure of an active component in the gas phase. Simple transformations such as melting and $solid_1 - solid_2$ transformations were discussed earlier in section 2.1 (Figure 1). In this section, however, the big picture is assembled over the full range of composition encompassing the specific system. Both DTA and DSC are appropriate for investigating phase transformations. If the primary concern is the temperature at which the events occur, then the simpler DTA approach is entirely suitable. If the C_p or ΔH of the events are desired as well, however, then DSC is the preferred method.

Pope and Judd[22] provide a nice description of the application of DTA to study generic phase diagrams. Figure 4 illustrates how the DTA or DSC curve evolves for a hypothetical binary phase diagram.[22] The hypothetical phase diagram is illustrated in Figure 4(a). A line

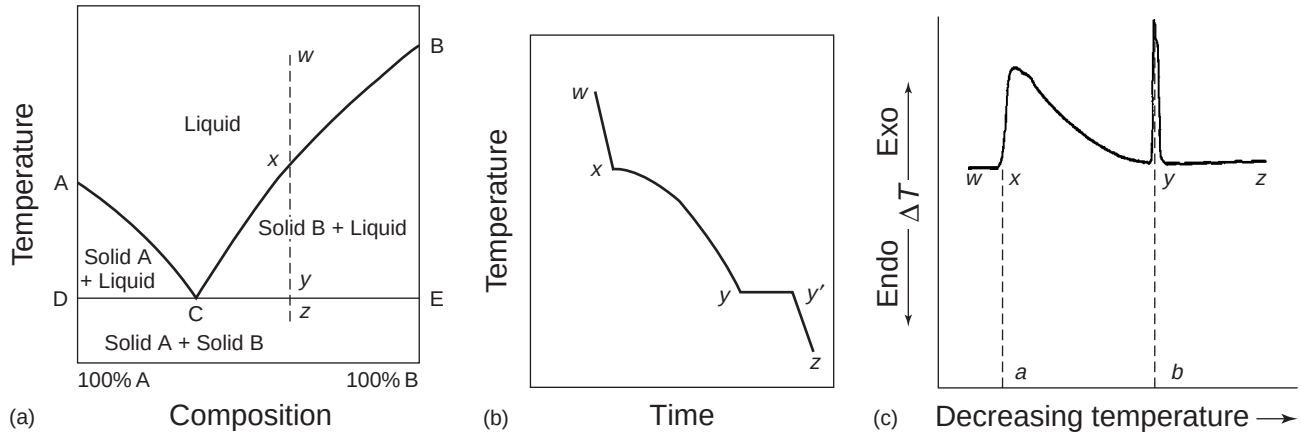

Figure 4 Evolution of a DTA curve for a simple hypothetical binary phase diagram: (a) the phase diagram with a line at constant composition; (b) the simple cooling curve along that line; (c) the DTA curve along that line.[22]

is drawn at constant composition representing a range of temperature from points *w* to *z*. That line crosses the liquidus at temperature *x* and the solidus at temperature *y*.

The simple cooling curve on going from point *w* to *z* is shown in Figure 4(b). While the material is above the liquidus, i.e. single phase, it cools at a rate depending on the exchange of heat with the surroundings. At point *x* solid phase B begins to crystallize from the liquid and the heat liberated by this exothermic process slows the rate of cooling. Material B is continuously solidifying as the two-phase mixture cools to the eutectic temperature *y*. At that point, three phases must coexist, solids A and B plus the eutectic liquid composition. The phase rule allows no degrees of freedom and the temperature must remain constant until one of the phases disappears at *y′*. On cooling this will be the liquid phase. Further cooling then follows the simple curve associated with the heat exchange of the two-phase mixture with its surroundings.

The DTA or DSC curve associated with this process is illustrated in Figure 4(c). The constant rate of cooling between temperatures *w* and *x* provides a constant derivative and hence a horizontal line. At point *x* the rate of cooling changes dramatically and there is an exothermic excursion. Since the amount of B solidifying decreases with time, that exothermic deflection decreases with time and temperature. At the eutectic temperature, *y*, the remaining liquid solidifies isothermally, which should give rise to an infinite derivative. This is tempered, however, by the finite kinetics of the process and is reflected as a sharp exotherm in the curve. Finally, the material cools at a nearly constant rate again below temperature *y* setting up another horizontal section of the curve.

This is extended to a more complex phase diagram in Figure 5 that includes congruently and incongruently melting compounds.[22] A series of cuts are made at key compositions in the system. The corresponding DTA curves, upon cooling, are illustrated directly above their respective compositions. These compositions were selected to indicate the variety of potential curves showing the peritectic point in curves 2 and 3, the simple eutectic in curve 4, and the congruently melting compound in curve 5. Curves 6 and 7 are similar to those in Figure 4. The liquidus curve is traced as the dashed line across the series of DTA curves.

Naturally, a simple collection of cooling curves is insufficient to establish the phase diagram. Supercooling and metastability are frequent problems so that heating curves should be determined and considered as well. Evidence of such kinetic behavior can be obtained by varying the heating and cooling rates. X-ray diffraction (XRD) is needed to establish the structures and, along with the electron microprobe analyses, verify the identities of the phases present in each region. Microscopy of polished sections is helpful for determining small quantities of phases and predicting whether any phases had been liquid at some time during the thermal treatment.

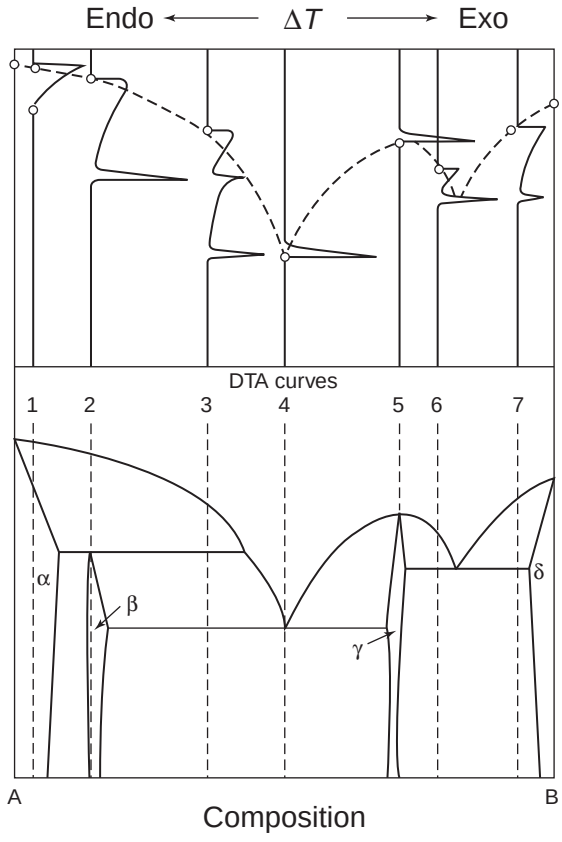

Figure 5 A complex hypothetical binary phase diagram.[22] The isocompositional lines are extended upwards showing the cooling DTA curves associated with the specific compositions.

3.1 Metals

Shull[23] discusses applications of DTA to phase transitions and phase equilibria in metals. Particular attention is given to the high-titanium portion of the complex Ti–Al diagram. In this paper, and references therein, detailed DTA studies showing the effects of subtle order ↔ disorder transitions are outlined. To explain these observations completely, however, as indicated earlier, it was necessary to invoke other measurements, such as high-temperature XRD and various microscopies.

Thermoanalytical techniques are ideally suited to investigate nonequilibrium phenomena as well as equilibrium diagrams. Metastable-to-stable transformations are quite exothermic and frequently occur abruptly once nuclei form. Altounian and Strom-Olsen[24] have reviewed some applications of amorphous to crystalline transformation in alloys. Figure 6(a) shows a DSC scan of initially amorphous $Pd_{40}Ni_{40}P_{20}$, while Figure 6(b) illustrates the effects of successive annealing periods of 1 h on the crystallization

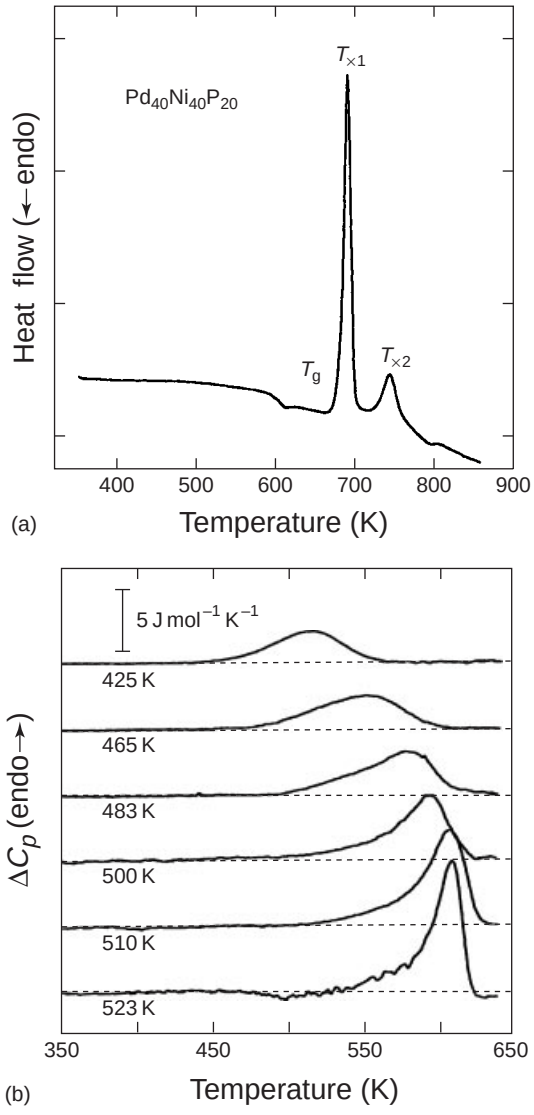

Figure 6 DSC scans of amorphous $Pd_{40}Ni_{40}P_{20}$: (a) as formed, at 40 K min^{-1}; (b) after 1 h anneals at the indicated temperatures.[24]

peaks. The glass transition of the disordered metal is clearly evident near 600 K. There are two sharp exotherms associated with crystallization of two phases at higher temperatures. Annealing at temperatures below T_g, however, will lower the crystallization temperature dramatically and sharpen the peak as well. This is because the values and range of activation energies for the process both decrease as the annealing temperature approaches T_g. Other examples include NiTi alloys[25,26] and AlLi alloys.[27]

TM can play a role in determining phase equilibria for magnetic materials. The value of T_C for instance can be used to identify the phases precipitated from metallic glasses, for example FeB and NdFe systems, as well as to follow the kinetics of crystallization.[24] Chromindur alloys (FeCrCo system) offer an interesting example illustrating magnetic effects during spinodal decomposition.[28] The very close continuous proximity of the two phases induced by spinodal decomposition allows the magnetic field of the phase having the higher T_C to clamp and maintain the magnetic alignment of both phases to the higher temperature.

3.2 Ceramics

Thermal analysis, particularly DTA, has played an active role in most determinations of phase equilibria. Establishing the liquidus and solidus portions of the diagram as described in Figure 5 is common. Similarly, determinations of subsolidus phase transformations, as illustrated in Figure 1, are also commonplace. TG, however, offers the added ability to determine the regions of stability as a function of the atmosphere. Variations in phase equilibria are measured as a function of the partial pressure of an active component in the gas phase, such as oxygen, sulfur, halogens, volatile compounds, and so on.

Vapor pressures can be measured using Knudsen cell techniques. The stepwise reduction and volatilities of gallium sulfides were determined through the homologous series from $BaGa_2As_4$ to $Ba_5Ga_2As_8$.[29] The vapor pressure is constant, but different, in each of the two-phase regions. It changes abruptly to a lower equilibrium pressure of Ga_2S and S_2 as each member in the series $BaS \cdot xGa_2S_3$ becomes exhausted at constant temperature. These isothermal, isobaric steps define the two-phase regions in the phase diagram.

The degree of oxygen nonstoichiometry controls the electrical, magnetic, and optical properties of materials. For example, the properties of a ferrite depend strongly on the Fe^{2+} to Fe^{3+} ratio. Normally this is set during the high-temperature sintering step and then quenched in, as much as possible, to yield a metastable single-phase spinel at room temperature. If the exact oxygen content is known at some point, then the changes in oxygen content can be followed by very careful TG measurements as the temperature and partial pressure of oxygen are varied. Careful work is needed to account for small changes in mass and losses other than oxygen.

Figure 7 shows both raw and processed data for a $Ni_{0.685}Zn_{0.177}Fe_{2.138}O_{4+\gamma}$ spinel.[30] Figure 7(a) shows the raw data achieved after long equilibration times in various partial pressures of oxygen at the indicated temperatures. The changes in mass indicated on the right ordinate are converted to γ in the chemical formula and shown on the left ordinate. The dependency of the results on the direction of crossing the phase boundary illustrates the metastability achieved on passing from the single-phase regime, I, into the two-phase region,

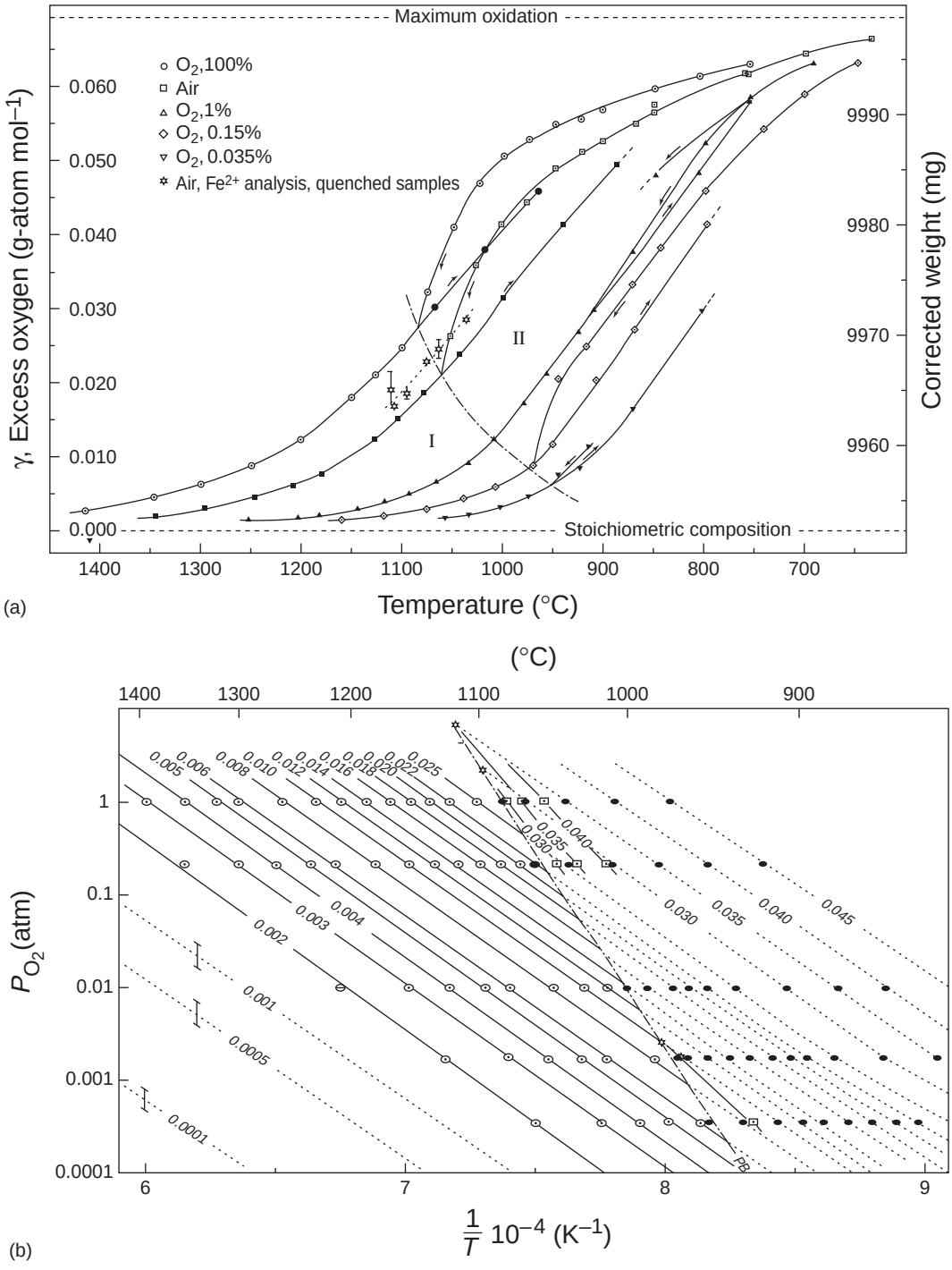

Figure 7 Raw and processed mass loss data for $Ni_{0.685}Zn_{0.177}Fe_{2.138}O_{4+\gamma}$ spinel: (a) corrected data; (b) processed data showing isocompositional lines and the phase boundary.[30]

II, when nucleation of a second solid phase must take place. Nucleation is not needed when passing from the two-phase into the single-phase region. The mass axis has been corrected for buoyancy and for the mass losses of ZnO and PtO_2 occurring at the higher temperatures.

Figure 7(b) presents the van't Hoff plots based on the data in Figure 7(a). The various lines are for the indicated values of γ and reflect how the partial pressure of oxygen and the temperature must adjust to preserve a constant composition. The dashed lines represent the metastable situation. Similar work has

been done for several high-temperature superconductors where the desired property depends closely on the oxygen content.[31]

3.3 Others

Besides the wide array of metallic and ceramic systems, there are many studies of fused salt systems, for example, nitrates, halides, sulfates, hydroxides, and so on. The results of such studies have been compiled, along with oxide-based systems, in *Phase Diagrams for Ceramists*.[32] Typical examples of the use of DTA to determine such diagrams are those of Touboul et al.[33] for systems of Li_2SO_4 with several divalent sulfates, Thiel and Seifert[34] for alkali halides, and Paulik et al.[35] for the many hydrates of $Ni(NO_3)_2$. Coupling DSC or DTA with optical microscopy can prove particularly informative. Such techniques have been described by Wiedemann and Felder-Casagrande[36] and the $NaNO_3-KNO_3$ system used as an example. Similarly the work of Arii and co-workers,[37,38] utilizing simultaneous XRD with DSC or TG, provides a valuable tool for studying phase equilibria and chemical reactivity.

4 CHEMICAL ANALYSIS

As indicated in the introduction, the general term ''thermal characterization'' is more appropriate than ''thermal analysis''. Nevertheless, there are many instances where thermoanalytical techniques can be used to perform chemical analysis in a convenient, relatively rapid, and accurate manner. Five examples are chosen to illustrate the use of TG, TM, DTA, DSC, and EGA for this purpose.

4.1 Lithium Niobate

The ferroelectric transformation discussed earlier and depicted in Figure 1(a) for $Li_{1-x}NbO_{3-0.5x}$ forms a very convenient and accurate way to determine the value of x. Large, virtually perfect single crystals of this material are desired for electro-optical devices. The compound melts incongruently and it is necessary to grow the homogeneous crystals from a nonstoichiometric liquid. The exact composition of the congruently melting composition was determined by DTA curves for the second-order transformation, such as those shown in Figure 1(a).

A carefully prepared series of standards was made and measured.[3] Figure 8(a) shows the resulting calibration curve for the determination of the Li content based on T_C. During crystal growth from these standard compositions by the Czochralski technique, samples were taken from the initial and final melts and the top and bottom of the resulting crystal. From the Li content in these samples it was possible to determine the distribution of Li between the solid and liquid phases. The distribution coefficients are plotted in Figure 8(b).[3] The congruent composition is that having a distribution coefficient of unity, 44.85 mol% Li_2O. When the crystal is pulled from a melt of that composition, the DTA curves for those four samples, that is, beginning and ending melt and solid, will have the same composition as indicated in Figure 8(c). Not only did this provide the necessary analytical data, it now provides an excellent method for quality control of the incoming cullet and outgoing crystal. A correlation was also made between the value of T_C and various aliovalent substitutions which produce defects similar to that accompanying Li deficiency.[3]

4.2 Lead–Tin Solder

A solid-state transition was used for analytical purposes in section 4.1. The liquidus curve can serve that function as well. Kuck[39] clearly demonstrated that the onset of a solid phase on cooling, or the onset of liquid phase on heating, could be compared with the phase diagram to establish the composition. Since the binary system forms a simple eutectic similar to that depicted in Figure 4(a), it is necessary to know which side of the eutectic the composition one is on, in order to establish the composition uniquely. Such analyses were used successfully to determine the solder composition of electroplated samples as a function of the various plating conditions.

4.3 Nonstoichiometry

In section 3.2, the use of TG was described to determine variations in the composition for various oxides as a function of the partial pressure of oxygen (see Figure 7). The success of this in terms of absolute composition, however, requires that the exact composition be known at some point in order to convert the relative concentrations to absolute values. Frequently, total or partial reduction of the starting material may be achieved to provide a known composition. Examples of this are for materials such as $BaPtO_{3-x}$[40] or $Ba_2YCu_3O_{7-\delta}$.[41]

Figure 9(a) presents TG curves for heating samples of $Ba_2YCu_3O_{7-\delta}$, a high temperature superconductor, in a nonexplosive forming gas (15% H_2, 85% N_2).[41] The thermodynamics of these oxides is such that only copper is reduced and the final elemental composition is $Ba_2YCu_3O_{3.50}$. Consequently, the overall mass loss at that point can be converted to the value of δ in the starting material. Once the starting composition has been determined, the changes in mass as a function of the partial pressure of oxygen and temperature indicate the changing composition.

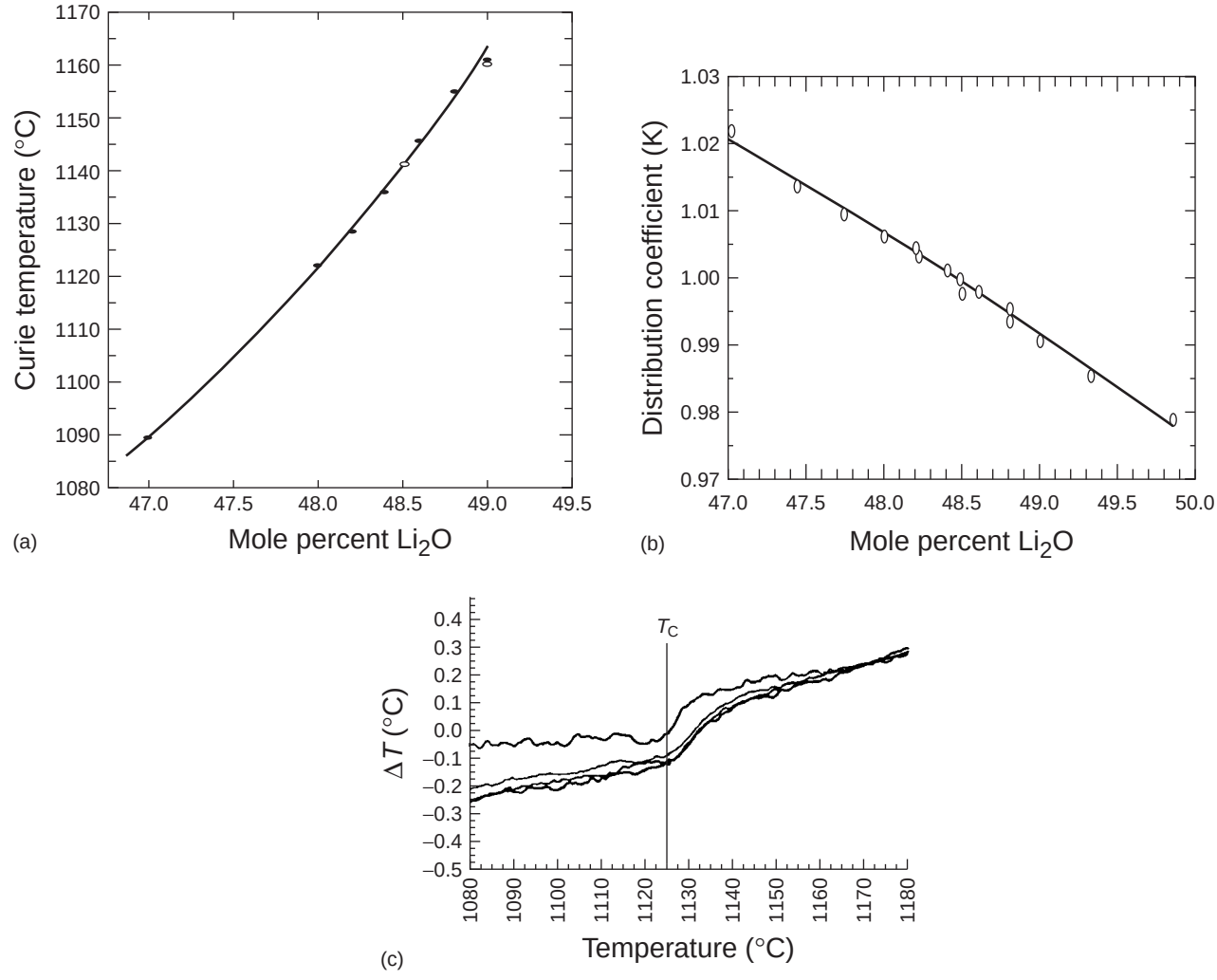

Figure 8 Use of DTA to establish the congruently melting composition of $Li_{1-x}NbO_{3-0.5x}$.[3] (a) T_C as a function of Li content; (b) distribution curve for lithium between the liquid and solid phase; (c) DTA curves from the starting and ending melts during Czochralski crystal growth and the top and bottom portions of the resulting crystal.

In this instance the changes are large and easily measured. The kinetics of the process are favorable so that measurements can be done in a dynamic manner, as contrasted with the isothermal approach used for ferrites (see Figure 7). The fact that equilibrium is established at the chosen heating or cooling rate can be established by comparing the results with those obtained at isothermal sections. Agreement between the values attained during the heating and cooling stages indicates that equilibrium was established in Figure 9(b). This approach is less viable for other superconducting or catalytic compositions containing, for example, Bi, Pb, Hg, and so on because of the volatility of the metal or its oxide.

4.4 Alloys and Ferrites by Thermomagnetometry

In section 2.2, the use of well-defined magnetic transition temperatures to calibrate the temperature axis during TG was described. Conversely, measured values of T_C can be used to establish the composition in many instances. The use of TM for qualitative analysis during studies of phase equilibria was briefly described in section 3.1. Other examples are covered in section 5.

Haglund[42] has reviewed some quantitative applications in metallurgy. An excellent example is the use of TM to analyze cemented carbide tool steels. Figure 10(a) indicates that T_C is a good indication of the WC content dissolved in the binder phase, Co.[42] Figure 10(b) shows how both the dissolved Ti and C affect the T_C for Ni-based binder systems.[42] The facts that Ni and Co form a complete solid solution and that the end members differ in their values of T_C by about 770 °C form a good basis for analysis of Ni–Co alloys.[43] Ferrites having the spinel structure and based on substitutions into Fe_3O_4 are very difficult to distinguish by XRD. However, the

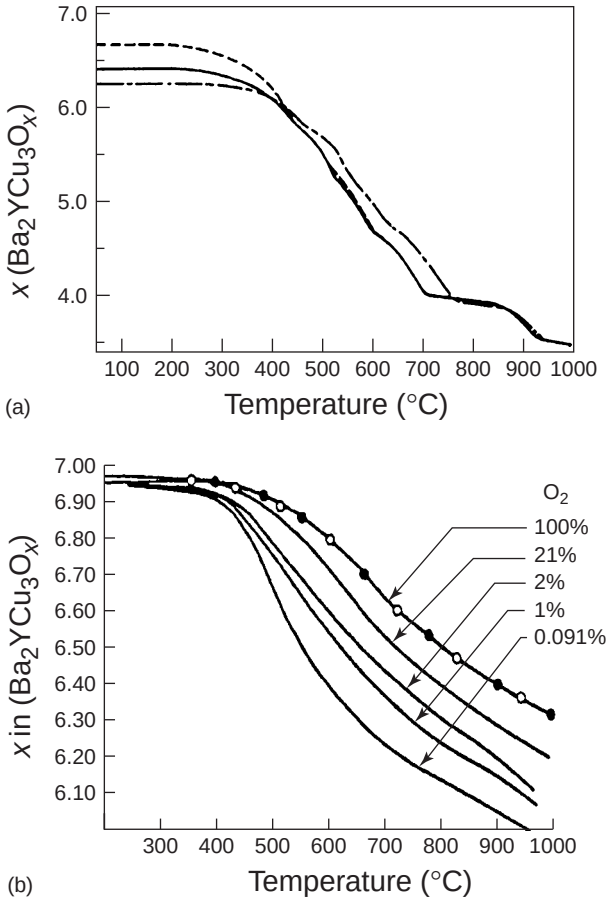

Figure 9 TG traces for high-T_C superconductors $Ba_2YCu_3O_{7-\delta}$.[31] (a) Several samples heated at 10 °C min^{-1} in 15% H_2, 85% N_2; (b) a sample of $\delta = 0.02$ ($x = 6.98$), heated in the indicated atmospheres at 1 °C min^{-1}. The circles are isothermal points after achieving constant mass on heating (closed) and cooling (open).

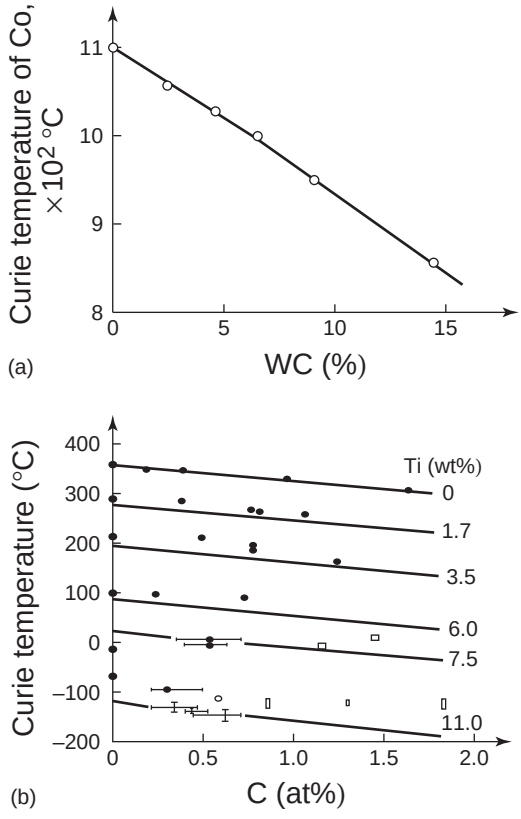

Figure 10 Examples of the use of TM for quantitative analysis: (a) variations of T_C with WC content in Co-based tool steels; (b) variations of T_C with C content in Ni–Ti alloys.[42]

value of T_C generally has a strong dependence on the extent and nature of the substitution and offers a method of analysis.[14]

4.5 Thin Films

The small amount of available sample in combination with the large amount of inert material, the substrate, presents a major challenge when analyzing thin films using thermoanalytical methods.[44,45] Gallagher[44] and Leskelä et al.[45] indicate that careful application of the traditional methods can be valuable. However, EGA, particularly by mass spectrometry (MS), is an especially viable technique because of its great sensitivity. Figure 11 summarizes the use of mass spectrometry/evolved gas analysis (MS/EGA) to analyze for key impurities in two thin films of importance in the manufacturing of semiconductor devices. Thin films of $TaSi_2$ are occasionally used as conductors during device fabrication. These films were prepared by sputtering in a reduced pressure of Ar and adhesion to the substrate was a problem initially. Extensive MS/EGA studies revealed that Ar was incorporated into the films and was evolved at an untimely stage during the subsequent processing (Figure 11a).[46] Studies of the processing conditions, in combination with characterization of the resulting films by MS/EGA, provided a means of reducing the problem to acceptable levels.

The opaque mask material frequently used in X-ray lithography of silicon wafers is BN. It is prepared by a CVD process involving various hydrogen-containing species, for example, diborane and ammonia. Frequently, H_2 is trapped in the thin film of BN and induces substantial mechanical stress therein. The resulting strain distorts the wafer and destroys the ability to align the mask. The broad overlapping peaks in the EGA of such a wafer and film, as shown in Figure 11(b), indicate that the H_2 is bound at a variety of sites. Annealing samples at different temperatures for various times to remove H_2, followed by subsequent EGA experiments, provided the opportunity to develop a suitable process to relieve the stresses.[47] Only the less strongly bonded H_2 needs to be removed, so a modest anneal at 550 °C is sufficient.

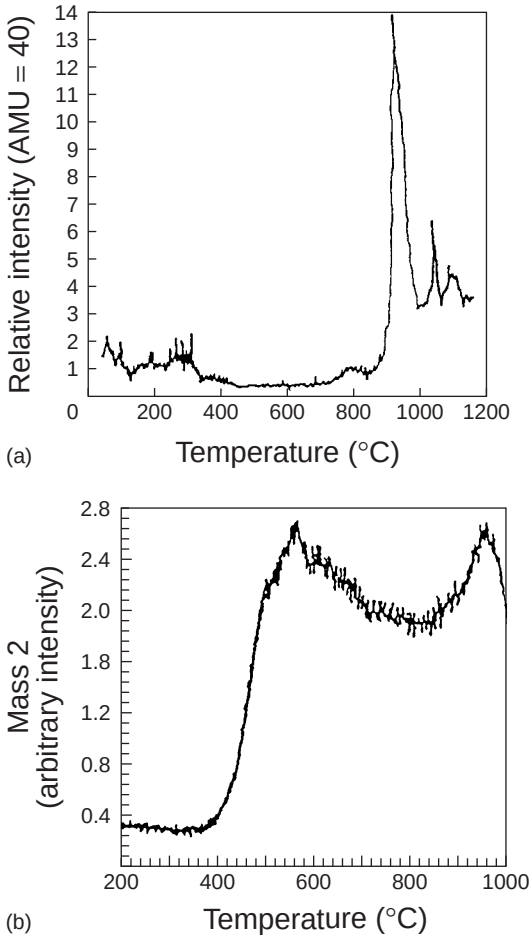

Figure 11 MS/EGA plots of electronic materials: (a) $TaSi_2$ films prepared by sputtering in Ar;[46] (b) BN films prepared by chemical vapor deposition (CVD) using B_2H_6 and NH_3.[47]

5 REACTIVITY

This is a very extensive and complex topic having great technological importance.[48,49] Thermal analysis plays an important role in many ways.[50,51] The first three subsections divide the topic into qualitative, quantitative, and overall considerations. The next two subsections discuss specific areas of application that have particular industrial importance. All these aspects of reactivity are then focused in the final subsection on the synthesis of materials.

5.1 Solid-state Reactions

Reactions involving solid phases are frequently divided into (1) solid–solid, (2) solid–liquid, and (3) solid–gas reactions. Solid-state decompositions may be considered as a separate class. All four categories lend themselves to thermoanalytical detection. Many solid–solid reactions are examples of acid–base reactions. Consider the equilibria shown in Equations (1) and (2), where the alkaline earth compound is basic, the silica is acidic, and the resulting compound is a salt:

$$CaO_{(s)} + SiO_{2(s)} = CaSiO_{3(s)} \quad (1)$$

$$CaCO_{3(s)} + SiO_{2(s)} = CaSiO_{3(s)} + CO_{2(g)} \quad (2)$$

Often reactions such as that shown in Equation (2) take place essentially in a single step without the appearance of a separate stage of carbonate decomposition forming an intermediate, CaO in this case. This is justification for writing the reaction shown in Equation (2) as a single equation.

Equation (1) occurs without a change in mass or the evolution of a gaseous product. Consequently, it is not detected by TG or EGA. No magnetic phases are present so TM is not suitable. The most appropriate thermoanalytical technique involves detection of the heat absorbed or evolved by the reaction, that is, DTA or DSC. XRD is also a viable means of observing the onset of such a reaction as the temperature is raised. Both methods are often used, although, DSC or DTA are far simpler and generally more sensitive. The reaction could also be detected by changes during TD experiments. However, this would be relatively indirect and less easily interpreted. The reaction shown in Equation (2) is amenable to the same methods of detection as that in Equation (1) but, because a gaseous product is formed, it is easily identified by both TG and EGA.

When any of the reactants, intermediates, or products are ferro- or ferrimagnetic it is possible to use TM as a tool to follow the reaction. During the thermal decomposition of many compounds of iron(II) in an oxidizing atmosphere, for example, carbonates[52] or oxalates,[43,53] the final product is usually an oxide of iron(III). Does an intermediate of mixed valence, for example, Fe_3O_4 or a related compound form? These spinels have very similar XRD patterns and it is difficult to determine the composition by that method.

Since the spinels containing iron are frequently magnetic, TM curves may be used to detect and identify the particular spinel formed. Figure 12(a) presents the TM curves for the decomposition of a very pure mineral, siderite ($FeCO_3$), in oxygen and nitrogen.[52] The value of T_C for Fe_3O_4 is about 585 °C. The decomposition in oxygen yields Fe_2O_3 directly without any evidence for the formation of Fe_3O_4 as an intermediate. However, the TM curve in nitrogen clearly shows a very substantial apparent gain in weight at the onset of the decomposition corresponding to the formation of Fe_3O_4. The curve returns to the normal weight as the magnetic attraction disappears. The agreement of the final weights in both atmospheres indicates that the oxygen content in ''inert''

gases is frequently well above that necessary to preserve the reduced species for long times.

Similar studies on common samples of siderite do not show this magnetic effect during decomposition in an inert atmosphere. It is tempting to conclude that impurities prevent the formation of the intermediate spinel. TM curves during cooling after the decomposition, however, indicate that the magnetic spinel intermediate is formed. Typical metallic impurities in siderite are other divalent cations, such as magnesium and manganese. Figure 12(b) presents TM curves for such partially substituted spinels or ferrites. The values of T_C for such substituted spinels are below the onset temperature of the decomposition and consequently there is no magnetic attraction observed during heating.

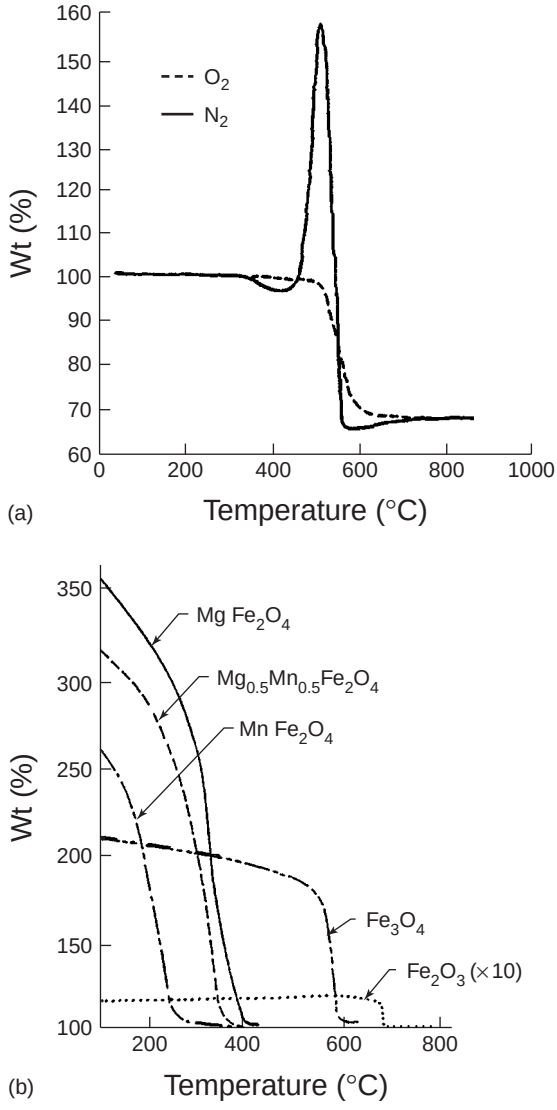

Figure 12 TM curves obtained at $10\,^{\circ}C\,min^{-1}$: (a) very pure siderite in O_2 and N_2; (b) various substituted spinels and hematite in O_2.[52]

The XRD patterns for nickel and cobalt, which have the same structure and nearly identical atomic radii, are very similar. Consequently, it is impossible to determine whether solid solutions of the co-precipitated oxalates decompose in an inert atmosphere to form initially the individual metals or an alloy. A comparison of TM curves for physically mixed and coprecipitated oxalates reveals that the physical mixtures decompose to individual metals while the co-precipitates are more reactive and form the alloy directly.[43]

The two examples so far have been for solid–solid and decomposition reactions. The field of corrosion provides many examples of solid–gas reactions.[54] Melting and crystallization reactions (solid–liquid) have been described in conjunction with phase equilibria. Clever corrosion studies were performed by Charles[55] using TM of sealed ampoules containing a ferromagnetic metal and corrosive liquid media. The apparent loss of weight was followed as the metal dissolved.

5.2 Heterogeneous Kinetics

Section 5.1 describes following reactions in a qualitative manner. Kinetic studies, however, require a quantitative description of the concentrations as a function of time. Several excellent treatises are available on the heterogeneous aspects of this topic.[56,57] Heterogeneous kinetics is a far more complex and controversial topic than homogeneous kinetics. Aspects of geometry, sample size, contact among reactants, thermal and mass transfer, structural order, defects, and so on, control the rate more than simple traditional terms related to concentrations.[58] The conventional expressions for concentration are not useful and the term "fraction reacted" (0 to 1) is commonly used instead. Determining the mechanism of the reaction under these conditions is a demanding task.

The fraction reacted, α, can be readily derived from various thermoanalytical measurements provided that certain assumptions are valid. While evolved or consumed, gas analysis is a direct measurement of the progress, and changes in mass or volume and the evolution or absorption of heat are indirect measurements of α predicated on an assumed general mechanism, stoichiometry and a known starting or ending point. In the simplest case, the overall integrated change in the measurement is normally assigned a range of α from zero to unity. The fraction of that total integral is then converted into α as a function of time and/or temperature assuming that the mechanism does not change in the region of interest.

Detailed discussion of kinetic analysis is beyond the scope of this chapter. Figure 13, however, indicates how thermoanalytical data, TG in this case, are used for this purpose. The simple thermal decomposition of $NaHCO_3$

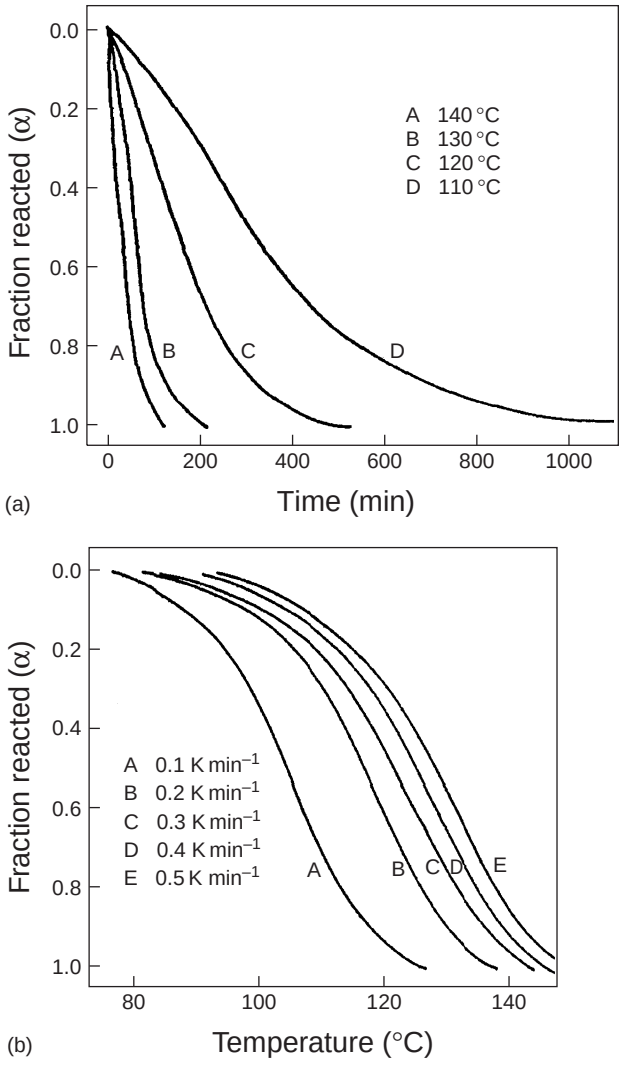

Figure 13 TG curves for determining the kinetics of the thermal decomposition of $NaHCO_3$: (a) isothermal; (b) dynamic at the indicated heating rates.[59]

is the example, shown in Equation (3)

$$2NaHCO_{3(s)} = Na_2CO_{3(s)} + H_2O_{(g)} + CO_{2(g)} \quad (3)$$

Isothermal data are presented in Figure 13(a) and examples of dynamic temperature results are given in Figure 13(b).[59]

The change of α or $d\alpha/dT$ with time or temperature is then fitted to various rate laws to find the best mathematical description. This process is fraught with uncertainties and caution must be exercised.[58,60,61] Decomposition of carbonates and hydrates, redox reactions, or other readily reversible reactions are particularly subject to variability because the removal of the products becomes a factor. Highly exothermic or endothermic reactions are subject to limitations due to thermal transport. It is advisable to vary the experimental parameters, for example, sample size, sample packing or shape, and gaseous flow rate, in a meaningful manner in addition to making other supporting measurements, such as XRD, microscopy, and so on, before making conclusions based on the mathematical analysis regarding the mechanism and significance of the resulting Arrhenius parameters.

5.3 Influences on Reactivity

The previous two subsections described the qualitative and quantitative aspects of following chemical reactions and the effects of temperature upon them through use of thermoanalytical methods. Consider now evaluating the effects of factors other than temperature which influence the nature of solid-state reactions. Anything that introduces defects or structural flaws into the reactants will generally increase their reactivity. Hedvall, in his famous review article,[62] listed these as (1) mechanical deformation; (2) impurities, vacancies, interstitials, and so on; (3) radiation; and (4) changes in crystal structure accompanying phase transformations or decompositions. The last aspect was a new concept and is still controversial. It has been referred to as the ''Hedvall Effect''. Radiation is generally considered as optical, X-ray, γ-ray, and so on however, magnetic or electrical fields may also influence reactivity.

Comparison of the thermoanalytical results for samples that have had carefully specified and controlled pretreatments will frequently reveal factors that influence reactivity. An alternative approach is to measure essentially identical samples under different experimental conditions involving exposure to fields or radiation, oxidizing or reducing atmospheres, or during phase transformations. Figure 14 shows examples of each of these three conditions based on thermoanalytical measurements.

In Figure 14(a) UV radiation is turned on at different stages during the isothermal decomposition of AlH_3. Clearly the rate of the reaction increases dramatically upon radiation.[63] Figure 14(b) indicates how lowering the partial pressure of oxygen enhances the reaction of $BaCO_3$ with Y_2O_3 and CuO to form the high temperature superconductor $Ba_2YCu_3O_{7-x}$.[64] The lower oxygen pressure leads to a larger value of x, which in turn raises the diffusion coefficients of the reactants through the product layer separating them. The ''Hedvall Effect'' is illustrated in Figure 14(c) where the rate for the oxidation of magnetite is altered substantially in the region of the magnetic phase transition of the reactant, Fe_3O_4.[61]

5.4 Heterogeneous Catalysis

Many commercially important processes are only viable because of the successful development of heterogeneous

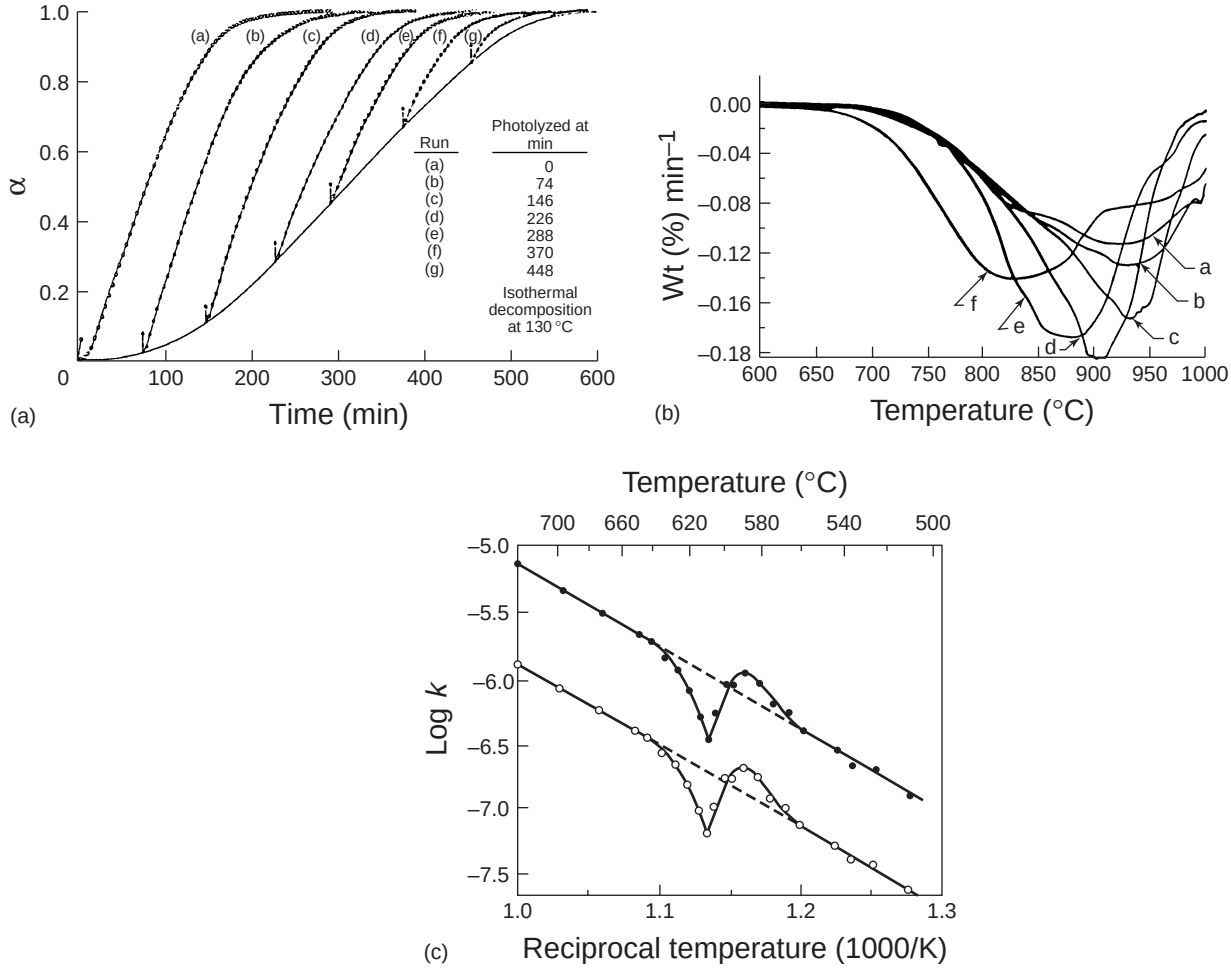

Figure 14 Thermoanalytical results related to solid-state reactivity: (a) isothermal TG curves at 130 °C for the thermal decomposition of AlH_3, showing the increase in rate when the ultraviolet (UV) exposure is initiated at different times,[63] (b) DTG curves for the reaction of $BaCO_3$ with stoichiometric amounts of Y_2O_3 and CuO to form $Ba_2YCu_3O_{7-x}$, showing the increased reactivity from a to f as the partial pressure of O_2 is decreased from 1 to 10^{-6} atm,[64] and (c) Arrhenius plots based on TG results for the oxidation of Fe_3O_4 showing the "Hedvall Effect" in the vicinity of the T_C of Fe_3O_4.[61]

catalysts. The catalytic action is based on surface active sites. These sites are generally chemically specific in nature such as acidic or basic, oxidizing or reducing, etc. Frequently the number and nature of such sites are determined by adsorption or desorption of specific gases. Acidic and basic sites can be determined from TG curves using suitable gaseous adsorbents, e.g. NH_3 for acidic sites or HCl for basic sites. Thermal desorption experiments are a common form of EGA. The temperatures at which various chemisorbed species are released or adsorbed from the catalyst is highly informative.

Most reactions have a substantial change in enthalpy associated with them. This evolution or adsorption of heat can be detected calorimetrically or by DTA. The oxidation of CO is very exothermic and can be rapidly and conveniently studied by DSC. These measurements require little sample and are simpler than conventional catalytic reactors for surveying potential catalysts and poisons. The DSC signal exhibits excellent correlation with analytical results for the progress of the reaction.

Some examples are depicted in Figure 15 for the oxidation of CO using supported oxide catalysts.[65] Figure 15(a) shows DSC curves at 10 °C min^{-1} for a flow of 2% O_2, 2% CO, and 96% N_2 over various catalysts. The earlier the temperature rise the more efficient the catalyst is. The unfortunate poisoning of the catalyst by as little as 150 ppm of SO_2 is indicated in Figure 15(b). If Pt is used in place of the oxide as a catalyst, the reaction is oscillatory in nature and the resulting temperature excursions can be followed by DSC under the right circumstances.[66] Another example is the work of Kosak[67] utilizing DSC at high pressure to follow the hydrogenation of unsaturated organic material using supported precious metal catalysts.

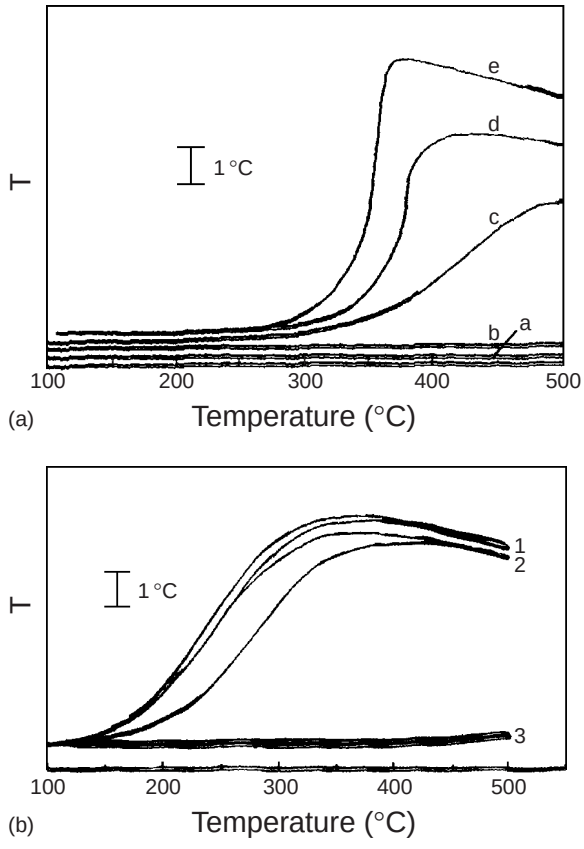

Figure 15 DSC curves of various catalysts heated at 10 °C min^{-1} in 2% CO, 2% O_2, and 96% N_2: (a) $La_{0.7}Pb_{0.3}MnO_3$ with increasing amounts of Pt from a to e, (b) same conditions using $La_{0.7}Sr_{0.3}MnO_3$, except that 150 ppm of SO_2 was added to the gas stream after the first scan.[66]

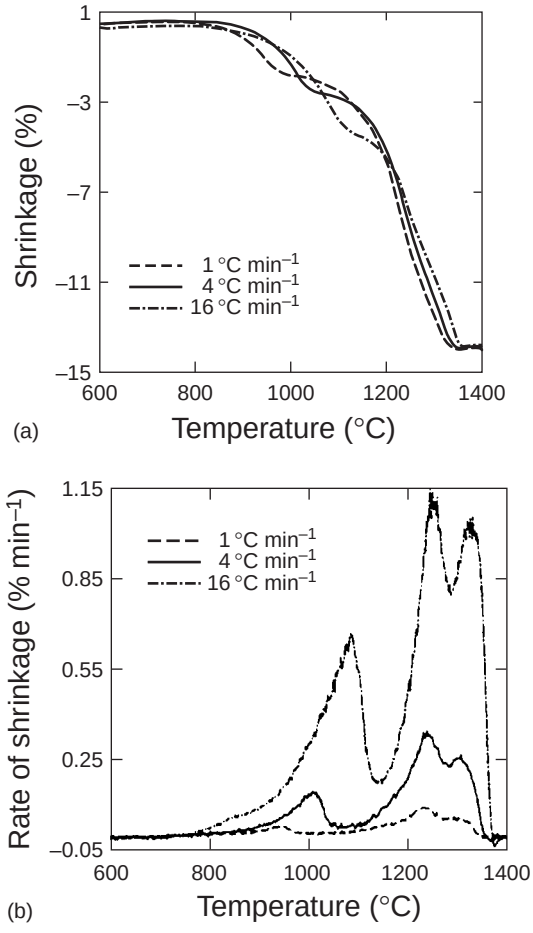

Figure 16 TD curves for the sintering of $BaTiO_3$ powders in air at the indicated heating rates: (a) shrinkage, (b) rate of shrinkage.[68]

5.5 Sintering

Most ceramic materials require a sintering step in order to achieve the desired mechanical, electrical, or magnetic properties. There is no change in mass associated with sintering and the process is so slow that the small evolution of heat accompanying the decrease in surface energy is not readily detected by DSC or DTA. The consolidation, however, leads to an increase in density and a decrease in volume. The latter is generally measured along a direction of interest by TD.

The change in length will exhibit discontinuities at first-order phase transformations and changes in slope at higher order transformations. Thermal decompositions may lead to expansion when gaseous products evolve. Sintering or densification is detected as a shrinkage. All these effects, plus the intrinsic thermal expansion of the sample, occur simultaneously and the observed TD trace is a composite of overall process. Figure 16 shows TD traces for the sintering of $BaTiO_3$.[68] The shrinkage curves are shown directly in Figure 16(a). The derivatives of these curves, that is, the rates of shrinkage, are depicted in Figure 16(b) and more clearly indicate the multiple stages that occur in these samples. These steps may arise from many causes, for example, density gradients in the powder compact, heterogeneous particle-size distribution, the presence of impurities, reactions with the atmosphere or support, variations in the reactivity of the powder, and so on.

This is particularly an auspicious point to mention briefly applications of unconventional temperature programs. Controlled-rate TD shows promise for achieving improved sintering while also conserving energy. The pioneering works of Palmour and Johnson[69] on the sintering of alumina and the recent work by Agarwal and Speyer[70,71] on other oxides are notable. Dynamic temperature TD may offer the opportunity to resolve some of the overlapping events accompanying sintering that were mentioned in the preceding paragraph.[68]

5.6 Synthesis

All the factors discussed under phase equilibria and reactivity are relevant when considering the synthesis of solid-state materials. The fundamental goals are to control, for each product, the

- chemical and phase composition
- macroscopic and microscopic morphology
- subsequent reactivity and stability.

It may also be necessary to control carefully the extent of the reaction, either to completion or to a limited extent when the desired product is an intermediate.

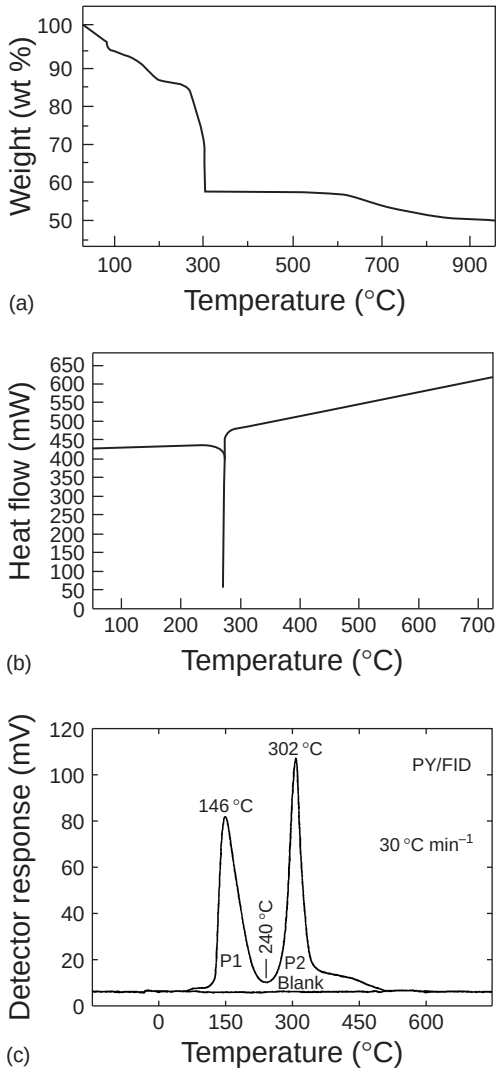

Figure 17 Thermoanalytical curves establishing processing conditions: (a) TG of $BaTiO(C_2O_4)_2 \cdot 4H_2O$,[73] (b) DSC of spray-dried superconducting salts,[74] and (c) EGA using a flame ionization detector of a tape-cast superconducting composition.[9]

Traditional methods of synthesis involving $solid_1$ plus $solid_2$ reactivity have been discussed. However, many modern methods of solid-state synthesis for ceramics, electronic materials, alloys, catalysts, and so on involve the thermal conversion of precursors. These precursors are generated from co-precipitates, spray-dried or freeze-dried powders, sol-gels, and so on.[72] Thermoanalytical methods are the primary tools for the characterization of the conversion process. Figure 17(a) shows an example of the use of TG to follow the thermal conversion of $BaTiO(C_2O_4)_2 \cdot 4H_2O$ into $BaTiO_3$,[73] Figure 17(b) uses DSC to follow the ignition temperature of a spray-dried mixture leading to a high temperature superconductor,[74] and Figure 17(c) uses EGA to follow the burnout of the plastic binder in a tape-cast superconducting composition.[9] Information from such experiments provides the basis for processing such precursors. TD provides the basis for subsequent sintering of the powders to dense, useful products.

6 CONCLUDING REMARKS

It was established in the introduction that space limitations would constrain the discussion to but a handful of the potential applications utilizing the major techniques. It is hoped that the selection herein will have served to point the way to further consideration and investigation by the reader. General references have been provided that are intended to encourage the reader to delve into the less common methods, such as those involving various forms of spectroscopy, acoustic emission or absorption, dielectric properties, and so on. Similarly the more sophisticated use of feedback and nonlinear programming is markedly improving the interpretation of the results and extending information achievable. The limits to applications of thermal methods are those imposed by our imaginations and ingenuity.

ABBREVIATIONS AND ACRONYMS

BET	Brunauer–Emmett–Teller
CVD	Chemical Vapor Deposition
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
DTG	Differential Thermal Gravimetry
EGA	Evolved Gas Analysis
MS	Mass Spectrometry
MS/EGA	Mass Spectrometry/Evolved Gas Analysis
MTDSC	Modulated Temperature Differential Scanning Calorimetry

ppm	parts per million
TD	Thermodilatometry
TG	Thermogravimetry
TM	Thermomagnetometry
TMA	Thermomechanical Analysis
UV	Ultraviolet
XRD	X-ray Diffraction

RELATED ARTICLES

Coatings **(Volume 2)**
Thermal Analysis of Coatings

Environment: Water and Waste **(Volume 3)**
Asbestos Analysis

Food **(Volume 5)**
Vitamins Analysis in Food

Forensic Science **(Volume 5)**
Pyrolysis Gas Chromatography in Forensic Science

Particle Size Analysis **(Volume 6)**
Surface Area and Pore Size Distributions

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Oil Shale and Shale Oil Analysis • Petroleum Residues, Characterization of

Polymers and Rubbers **(Volume 9)**
Dynamic Mechanical Analysis of Polymers and Rubbers • Pyrolysis Techniques in the Analysis of Polymers and Rubbers • Thermogravimetry of Polymers

Steel and Related Materials **(Volume 10)**
Atomic Absorption and Emission Spectrometry, Solution-based in Iron and Steel Analysis • Metal Analysis, Sampling and Sample Preparation in • Nickel Ore and Metals Analysis • Noble Metals, Analytical Chemistry of • Thermal Evolution Methods for Carbon, Sulfur, Oxygen, Nitrogen and Hydrogen in Iron and Steel Analysis

Surfaces **(Volume 10)**
Differential Reflectance Spectroscopy in Analysis of Surfaces

Chemometrics **(Volume 11)**
Signal Processing in Analytical Chemistry

Electroanalytical Methods **(Volume 11)**
Microbalance, Electrochemical Quartz Crystal

Kinetic Determinations **(Volume 13)**
Instrumentation for Kinetics

Mass Spectrometry **(Volume 13)**
Inorganic Substances, Mass Spectrometric in the Analysis of

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction • Differential Scanning Calorimetry and Differential Thermal Analysis • Simultaneous Techniques in Thermal Analysis • Thermogravimetry

General Articles **(Volume 15)**
Analytical Problem Solving: Selection of Analytical Methods • Quality Assurance in Analytical Chemistry

REFERENCES

1. P.J. Haines, M. Reading, F.W. Wilburn, 'Differential Thermal Analysis and Differential Scanning Calorimetry', in *Handbook of Thermal Analysis and Calorimetry*, ed. M.E. Brown, Elsevier, Amsterdam, Vol. 1, Chapter 5, 1998.
2. M. Reading, 'Controlled Rate Thermal Analysis and Related Techniques', in *Handbook of Thermal Analysis and Calorimetry*, ed. M.E. Brown, Elsevier, Amsterdam, Vol. 1, Chapter 8, 1998.
3. H.M. O'Bryan, P.K. Gallagher, C.D. Brandle, 'Congruent Composition and Li-rich Phase Boundary of $LiNbO_3$', *J. Am. Ceram Soc.*, **68**, 493–496 (1985).
4. P.K. Gallagher, Z. Zhong, E.L. Charsley, S.A. Mikhail, M. Todoki, K. Taniguahi, R.L. Blaine, 'A Study of the Effect of Purity on the Use of Nickel as a Temperature Standard for Thermogravimetry', *J. Therm. Anal.*, **40**, 1423–1430 (1993).
5. E.M. Vogel, P.K. Gallagher, 'Thermoanalytical Study of Some Fluoroberyllate Glasses', *Mater. Lett.*, **4**, 5–9 (1985).
6. J. Chiu, P.G. Fair, 'Determination of Thermal Conductivity by Differential Scanning Calorimetry', *Thermochim. Acta*, **34**, 267–273 (1979).
7. G. Hakvoort, L.L. Van Reijnen, 'Measurement of the Thermal Conductivity of Solid Substances by DSC', *Thermochim. Acta*, **93**, 317–320 (1985).
8. Y. Wang, M. Freeman, R. Nottenburg, K. Rajeshwar, J. Dubow, 'Automation of Laser-flash Thermal Diffusivity Measurements', *Thermochim. Acta*, **37**, 287–290 (1980).
9. P.K. Gallagher, 'Instrumentation, Techniques, and Methodology', in *Thermal Characterization of Polymeric Materials*, ed. E.A. Turi, Academic Press, New York, Vol. I, Chapter 1, 1997.
10. W.W. Wendlandt, *Thermal Analysis*, 3rd edition, John Wiley & Sons, New York, 1986.

11. Z. Halmos, W.W. Wendlandt, 'A Simultaneous Electrical Conductivity–Differential Thermal Analysis Apparatus', *Thermochim. Acta*, **7**, 95–101 (1973).
12. W.W. Wendlandt, 'A Thermovoltaic Detector (TVD) for Thermal Decomposition Reactions', *Thermochim. Acta*, **37**, 121–124 (1980).
13. W.W. Wendlandt, H.G. Hecht, *Reflectance Spectroscopy*, John Wiley & Sons, New York, 1966.
14. P.K. Gallagher, 'Thermomagnetometry', *J. Therm. Anal.*, **49**, 33–44 (1997).
15. T.J. Quinn, *Temperature*, Academic Press, New York, 1990.
16. S.J. Gregg, K.S.W. Sing, *Adsorption, Surface Area, and Porosity*, Academic Press, New York, 1967.
17. C.J. Keattch, D. Dollimore, *An Introduction to Thermogravimetry*, 2nd edition, Heyden, London, 1975.
18. H.M. O'Bryan, P.K. Gallagher, 'Characterization of $Ba_2YCu_3O_x$ as a Function of Oxygen Partial Pressure. Part II: Dependence of the O–T Transition on Oxygen Content', *Adv. Ceram. Mater.*, **2**, 642–648 (1987).
19. M.E. Brown, *Introduction to Thermal Analysis: Techniques and Applications*, Chapman and Hall, London, 1988.
20. R.F. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, New York, 1994.
21. Z. Zhong, P.K. Gallagher, D.L. Loiacono, G.M. Loiacono, 'The Thermal Expansion and Stability of $KTiOAsO_4$ and Related Compounds', *Thermochim. Acta*, **234**, 255–261 (1994).
22. M.I. Pope, M.D. Judd, *Differential Thermal Analysis*, Heyden, London, 1977.
23. R.D. Shull, 'Applications of Thermal Analysis Techniques to Phase Diagram Determinations', in *Thermal Analysis in Metallurgy*, eds. R.D. Shull, A. Joshi, TMS, Warrendale, 1992.
24. Z. Altounian, J.O. Strom-Olsen, 'Studies of Metastable Metallic Phase Transformations by Thermal Analysis', in *Thermal Analysis in Metallurgy*, eds. R.D. Shull, A. Joshi, TMS, Warrendale, 1992.
25. L.V. Meisel, P.J. Cote, 'Transient Phenomena Following Precipitation and Re-solution Anneals in a NiTi Alloy', in *Thermal Analysis in Metallurgy*, eds. R.D. Shull, A. Joshi, TMS, Warrendale, 1992.
26. W. Yu, 'The Application of Thermal Analysis in the Study of Ni–Ti Shape Memory Alloys', in *Thermal Analysis in Metallurgy*, eds. R.D. Shull, A. Joshi, TMS, Warrendale, 1992.
27. J.M. Papazian, 'Differential Scanning Calorimetry for Precipitate Characterization in Aluminum – Lithium Alloys', in *Thermal Analysis in Metallurgy*, eds. R.D. Shull, A. Joshi, TMS, Warrendale, 1992.
28. P.K. Gallagher, E. Coleman, S. Jin, R.C. Sherwood, 'Some Applications of Thermomagnetometry to the Study of Chromindur Alloys', *Thermochim. Acta*, **37**, 291–300 (1980).
29. P. Mukdeeprom-Burckel, J.G. Edwards, 'Chemistry and Thermodynamics of Solid and Vapor Phases in the Barium-sulfide, Gallium-sulfide System', *Mater. Res. Bull.*, **25**, 163–172 (1990).
30. P. Bracconi, P.K. Gallagher, 'Phase Diagram of a Nickel–Zinc Ferrite of Composition: $Ni_{0.685}Zn_{0.177}Fe_{2.138}O_{4+x}$', *J. Am. Ceram. Soc.*, **62**, 171–176 (1979).
31. P.K. Gallagher, 'Interactions of High-temperature Superconductors with the Atmosphere', *Thermochim. Acta*, **148**, 229–234 (1989).
32. M.K. Reser (ed.), *Phase Diagrams for Ceramists*, American Ceramic Society, Worthington, 1933-present.
33. M. Touboul, E. Le Samedi, N. Sephar, F. Broszniowski, P. Edern, E. Bétourné, 'Binary Systems with Li_2SO_4 as One of the Components', *J. Thermal Anal.*, **40**, 1151–1156 (1993).
34. G. Thiel, H.J. Seifert, 'Properties of Double Chlorides in the Systems $ACl/SmCl_3$ (A=Na–Cs)', *Thermochim. Acta*, **133**, 275–282 (1988).
35. F. Paulik, J. Paulik, M. Arnold, 'Investigation of the Phase Diagram for the System $Ni(NO_3)_2–H_2O$ and Examination of the Decomposition of $Ni(NO_3)_2•6H_2O$', *Thermochim. Acta*, **121**, 137–149 (1987).
36. H.G. Wiedemann, S. Felder-Casagrande, 'Thermomicroscopy', in *Handbook of Thermal Analysis and Calorimetry*, ed. M.E. Brown, Elsevier, Amsterdam, Vol. 1, 1998.
37. T. Arii, A. Kishi, Y. Kobayashi, 'A New Simultaneous Apparatus for X-ray Diffractometry and Differential Scanning Calorimetry (XRD–DSC)', *Thermochim. Acta*, **325**, 151–156 (1999).
38. T. Arii, A. Kishi, 'Thermal Dehydration of Cobalt and Zinc Formate Dihydrates by Controlled Rate Thermogravimetry (CRTG) and Simultaneous X-ray Diffractometry–Differential Scanning Calorimetry(XRD–DSC)', *Thermochim. Acta*, **325**, 157–166 (1999).
39. V.J. Kuck, 'Determination of the Liquidus Temperature and Composition of Tin/Lead Solders Using Differential Thermal Analysis', *Thermochim. Acta*, **99**, 233–242 (1986).
40. P.K. Gallagher, D.W. Johnson, Jr, E.M. Vogel, G.K. Wertheim, F.J. Schnettler, 'Synthesis and Structure of $BaPtO_3$', *J. Solid State Chem.*, **21**, 277–282 (1977).
41. P.K. Gallagher, H.M. O'Bryan, G.S. Grader, 'Reduction of $Ba_2YCu_3O_7$ and $Y_2Cu_2O_5$ by H_2', *Thermochim. Acta*, **137**, 373–381 (1989).
42. B.O. Haglund, 'Curie Temperature of Alloys, Its Measurement and Technical Importance', *J. Therm. Anal.*, **25**, 21–43 (1982).
43. P.K. Gallagher, P.T. Chao, L. Zhong, J. Subramanian, 'The Thermal Decomposition of Coprecipitates and Physical Mixtures of Nickel–Cobalt Oxalate Dihydrates', *J. Therm. Anal.*, **39**, 975–986 (1993).
44. P.K. Gallagher, 'Applications of Thermoanalytical Methods to the Study of Thin Films', *J. Therm. Anal.*, **38**, 17–26 (1992).

45. M. Leskelä, T. Leskelä, L. Niinistö, 'Thermoanalytical Methods in the Study of Inorganic Films', *J. Therm. Anal.*, **40**, 1077–1088 (1993).
46. R.A. Levy, P.K. Gallagher, 'Argon Entrapment and Evolution in Sputtered $TaSi_2$ Films', *J. Electrochem. Soc.*, **132**, 1986–1991 (1986).
47. T.M. Duncan, R.A. Levy, P.K. Gallagher, M.W. Walsh, Jr, 'Structural Characterization of Boron Nitride Films', *J. Appl. Phys.*, **64**, 2990–2994 (1988).
48. H. Schmalzried, *Solid State Reactions*, Verlag Chemie, Weinheim, 1981.
49. V.V. Boldyrev (ed.), *Reactivity of Solids: Past, Present and Future*, IUPAC, Oxford, 1996.
50. P.K. Gallagher, 'Ceramic Reactions and Phase Behavior', in *Characterization of Ceramics*, ed. R.E. Loehman, Butterworth-Heinemann, Boston, Chapter 8, 1993.
51. F. Paulik, *Special Trends in Thermal Analysis*, J. Wiley & Sons, Chichester, 1995.
52. P.K. Gallagher, S.St.J. Warne, 'Thermomagnetometry and the Thermal Decomposition of Siderite', *Thermochim. Acta*, **43**, 253–267 (1981).
53. R.G. Rupard, P.K. Gallagher, 'The Thermal Decomposition of Coprecipitates and Physical Mixtures of Magnesium–Iron Oxalates', *Thermochim. Acta*, **272**, 11–26 (1996).
54. U.R. Evans, *The Corrosion and Oxidation of Metals*, St. Martin's Press, New York, 1960.
55. R.G. Charles, 'Thermomagnetometry as a Means for Following Corrosion and Other Liquid-solid Chemical Reactions, a Review', in *Thermal Analysis in Metallurgy*, eds. R.D. Shull, A. Joshi, TMS, Warrendale, 1992.
56. M.E. Brown, D. Dollimore, A.K. Galwey, *Reactions in the Solid State*, Elsevier, Amsterdam, 1980.
57. A.K. Galwey, M.E. Brown, *Thermal Decomposition of Ionic Solids*, Elsevier, Amsterdam, 1999.
58. M. Maciejewski, 'Somewhere Between Fiction and Reality: The Usefulness of Kinetic Data of Solid-state Reactions', *J. Therm. Anal.*, **38**, 51–70 (1992).
59. H. Tanaka, H. Takemoto, 'Significance of the Kinetics of Thermal Decomposition of $NaHCO_3$ Evaluated by Thermal Analysis', *J. Therm. Anal.*, **38**, 429–435 (1992).
60. P.K. Gallagher, D.W. Johnson, Jr, 'The Effects of Sample Size and Heating Rate of the Kinetics of the Thermal Decomposition of $CaCO_3$', *Thermochim. Acta*, **6**, 67–83 (1973).
61. P.K. Gallagher, E.M. Gyorgy, H.E. Bair, 'An Anomaly in the Oxidation Rate of Magnetite at Its Curie Temperature', *J. Chem. Phys.*, **71**, 830–835 (1979).
62. J.A. Hedvall, 'Changes in Crystal Structure and their Effect on the Reactivity and Catalytic Effect of Solids', *Chem. Rev.*, **15**, 139–168 (1934).
63. P.J. Herley, O. Christofferson, 'Decomposition of Aluminum Hydride Powder', *J. Phys. Chem.*, **85**, 1882–1892 (1982).
64. P.K. Gallagher, D.A. Fleming, 'Influence of Oxygen Partial Pressure on the Synthesis of $Ba_2YCu_3O_7$ from a Novel Oxalate Precursor', *Chem. Mater.*, **1**, 659–664 (1989).
65. P.K. Gallagher, D.W. Johnson, Jr, E.M. Vogel, 'Examples of the Use of DTA for the Study of Catalytic Activity and Related Phenomena', in *Catalysis in Organic Syntheses–1976*, eds. P.A. Rylander, H. Greenstreet, Academic Press, New York, 113–136, 1976.
66. P.K. Gallagher, D.W. Johnson, Jr, 'Detection of Catalytic Oscillations by Differential Thermal Analysis', *Thermochim. Acta*, **15**, 238–240 (1976).
67. J. Kosak, 'High Pressure DTA of Hydrogenation Catalysts', in *Catalysis in Organic Syntheses–1976*, eds. P.A. Rylander, H. Greenstreet, Academic Press, New York, 137–148, 1976.
68. Z. Zhong, P.K. Gallagher, 'Combustion Synthesis and Characterization of $BaTiO_3$', *J. Mater. Res.*, **10**, 945–952 (1995).
69. H. Palmour, D.R. Johnson, 'Controlled Rate Sintering of Alumina', in *Sintering and Related Phenomena*, eds. G.C. Kuczynski, N.A. Hooten, C.F. Gibbon, Gordon & Breach, New York, 779–791, 1976.
70. G. Agarwal, R.F. Speyer, 'A Novel Programmed Shrinkage Dilatometer for Optimized Sintering of Powdered Ceramics', *J. Therm. Anal.*, **49**, 1297–1304 (1997).
71. G. Agarwal, R.F. Speyer, 'Effect of Rate-controlled Sintering on the Microstructure and Electrical Properties of ZnO with Bismuth and Antimony Oxide Additions', *J. Mater. Res.*, **12**, 2447–2454 (1997).
72. P.K. Gallagher, 'Chemical Synthesis', in *Engineering Materials Handbook*, ASM, Cleveland, 52–64, Vol. 4, 1991.
73. P.K. Gallagher, J. Thomson, Jr, 'Thermal Analysis of Some Barium and Strontium Titanyl Oxalates', *J. Am. Ceram. Soc.*, **48**, 644–647 (1965).
74. K. Kourtakis, M. Robbins, P.K. Gallagher, 'Synthesis of $Ba_2YCu_3O_7$ Powder by Anionic Oxidation (NO_3^-)–Reduction ($RCOO^-$, where R is H, CH_3, and CH_3CH_2)', *J. Solid State Chem.*, **83**, 230–236 (1989).

Simultaneous Techniques in Thermal Analysis

David Dollimore and Pauline Phang
The University of Toledo, Toledo, USA

1 Introduction 1
2 Theory and Operating Principles of a Simultaneous Thermogravimetry and Differential Thermal Analysis Unit 1
2.1 History of Simultaneous Thermogravimetry and Differential Thermal Analysis 2
2.2 Instrumentation of Simultaneous Thermogravimetry and Differential Thermal Analysis 2
2.3 Limitations of Simultaneous Thermogravimetry and Differential Thermal Analysis 3
2.4 Advantages of Simultaneous Thermogravimetry and Differential Thermal Analysis 3
3 Linking Computer Systems with Simultaneous Thermogravimetry and Differential Thermal Analysis 3
4 Evolved Gas Analysis and Associated Analytical Techniques 4
4.1 Mass Spectrometry 4
4.2 Fourier Transform Infrared Spectroscopy 6
4.3 Gas Chromatography 6
5 Applications of Simultaneous Techniques 6
5.1 Biological and Pharmaceutical Materials 6
5.2 Inorganic Compounds 6
5.3 Organic and Polymeric Compounds 7
Abbreviations and Acronyms 7
Related Articles 7
References 7

1 INTRODUCTION

While single techniques in thermal analysis are often used to analyze the thermal characteristics of a sample, there is an increasing preference for the application of coupled techniques. The operation of coupled techniques provides an analyst with a more comprehensive picture of the sample which is being investigated. For example, from a differential thermal analysis (DTA) trace, an analyst will be able to identify the various endothermic and exothermic reactions that are occurring as a sample undergoes the prescribed thermal treatment. However, it would be extremely useful if the analyst could also determine if the reactions corresponded to either physical or chemical changes. A thermogravimetry (TG) trace would show the physical and chemical changes for a sample as it undergoes heating. Hence, the coupling of these two thermal techniques would lead to a more complete profile of the material that is being analyzed.

When two analytical techniques are coupled and employed on any given sample within a single run, the term 'simultaneous' is used to refer to this procedure. Very often, the coupled thermal techniques, such as TG/DTA (simultaneous thermogravimetry and differential thermal analysis), are also referred to as 'hyphenated techniques'. It would be prudent to mention that the term simultaneous is not to be confused with the terms 'combined' or 'consecutive'. Both of these terms refer to either the investigation of a material by running independent samples on different instruments or on different arrangements of the same instrument.[1]

In this article the theory, operation and applications of a TG/DTA system will be explored. It is important to understand the advantages and disadvantages associated with the use of simultaneous techniques. Since the introduction of microcomputers, the inclusion of a computer workstation as an integral part of a thermal analysis operating system has evolved over time. Hence, some consideration will be given to the vital contributions of a computer workstation to a TG/DTA unit. A concise discussion on the various analysis techniques associated with TG/DTA will also be presented. Evolved gas analysis (EGA) is often performed on the gaseous products evolved from a sample that was analyzed on a TG/DTA unit. The methods used to detect and identify these gases include mass spectrometry (MS), Fourier transform infrared (FTIR) spectroscopy and gas chromatography (GC). The ultimate purpose of developing simultaneous techniques is to devise advanced systems which are capable of incorporating various conventional and thermal analytical techniques within a single system.

2 THEORY AND OPERATING PRINCIPLES OF A SIMULTANEOUS THERMOGRAVIMETRY AND DIFFERENTIAL THERMAL ANALYSIS UNIT

The main focus of this section is to provide some background information on the historical development,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the instrumentation and the working capabilities of a TG/DTA unit. As with any analytical technique, there are certain strengths and limitations linked with the use of simultaneous techniques. It is essential to know the boundaries of a chosen technique as this will help in the design of the experiments to be conducted.

2.1 History of Simultaneous Thermogravimetry and Differential Thermal Analysis

Due to the rising demand for coupled thermal techniques, simultaneous systems which afforded the collection of TG, derivative thermogravimetry (DTG) and DTA or differential scanning calorimetry (DSC) signals were developed. These units typically comprise thermobalances that are capable of monitoring the mass changes within a sample, and the temperature differences between a sample and reference cell as the sample is being subjected to a controlled temperature regime.

The pioneering TG/DTA system (dubbed the 'Derivatograph') involved the use of a derivatograph to allow for the simultaneous collection of TG/DTA data.[2–4] A schematic diagram of the Derivatograph is presented in Figure 1. The success of the first simultaneous unit spurred the emergence of improved TG/DTA (or DSC) systems such as the Stanton-Redcroft TG/DTA and the Setaram simultaneous TG/DSC.[5,6] Some latter day manufactured simultaneous systems include the TA Instruments SDT 2960 (TG/DTA) and the Netzsch STA 409 (TG/DSC).[7,8]

2.2 Instrumentation of Simultaneous Thermogravimetry and Differential Thermal Analysis

As mentioned in the previous section, TG/DTA unit designs are generally based on a thermobalance that is equipped with thermocouple leads. There are two basic arrangements for TG/DTA equipment. The first involves placing the sample and reference cells on one common vertical platform (refer to Figure 2). For this design, there is a single thermocouple that is responsible for the system temperature. The second arrangement entails separate sample and reference cells (see Figure 3). In this set-up, there are two thermocouple leads, and each of them is attached to the base of either the sample cell or the reference cell.

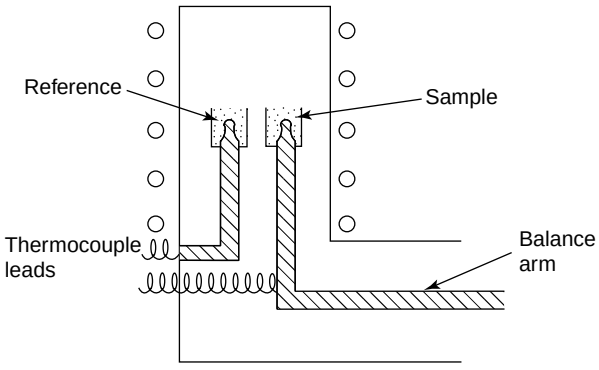

Figure 2 Schematic diagram of a TG/DTA equipment with separate sample and reference holders.

The modern TG/DTA system should consist of the following components. In addition to the microbalance with attached thermocouple leads, there should also be a computer workstation linked to the system. A discussion of the various roles of a computer workstation as linked to a TG/DTA unit, will be presented in the next part of this chapter. Generally, a direct current (DC) amplifier is also included in the system to enhance the DTA signal. Lastly, there should be a recorder to collect the various TG and DTA signals.

In newer models of simultaneous instruments, there are a few variations in the basic design. These changes were incorporated in an attempt to build better simultaneous thermal analysis units. For example, the TA Instruments

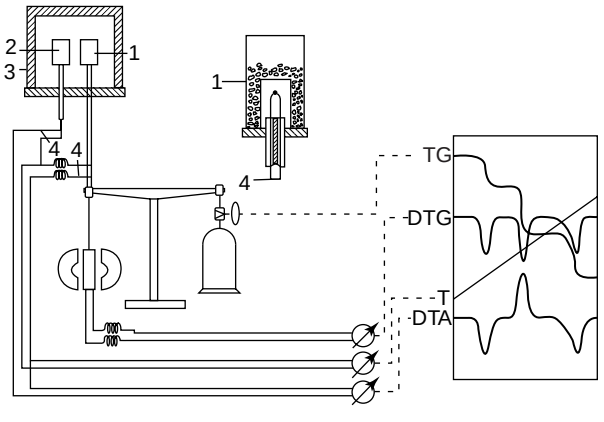

Figure 1 Diagram of Derivatograph for collecting TG/DTA data. 1, sample; 2, reference; 3, furnace; 4, thermocouples.

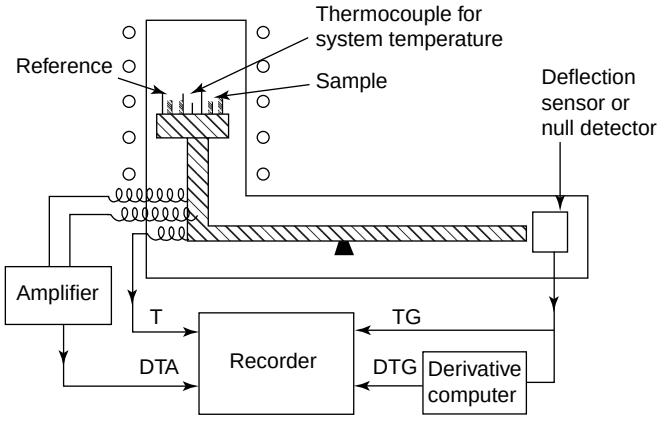

Figure 3 Diagram of the set-up for a TG/DTA unit with the sample and reference holders on a common platform.

2960 (TG/DTA) unit features a horizontal double beam balance set-up. The use of a horizontal system was introduced to counteract the effects of purge gas flow into the chamber unit. Yet another variation, the Setaram TG/DSC 111 involves the use of symmetrical dual furnaces linked with Calvet microcalorimeters, while the Netzsch STA 409 (TG/DSC) incorporates the use of a heat-flux panel within the unit.

2.3 Limitations of Simultaneous Thermogravimetry and Differential Thermal Analysis

A fundamental design limitation of the TG/DTA technique is that the sensitivity of both the TG and DTA methods are somewhat reduced when coupled. In practice, small sample sizes and fast heating rates are the recommended experimental parameters if one desires to achieve well-resolved DTA curves. This is necessary to ensure the rapid development and loss of temperature gradients as the sample is subjected to heating. Conversely, in order to enhance the output from TG experiments, large sample sizes and slower heating rates are advocated. Thus, a compromise has to be enforced if an analyst wants to optimize both signals from the TG/DTA system.

Another disadvantage of the TG/DTA equipment is the tendency for sample temperatures to deviate from the programmed temperature over the course of thermal events. This deviation may have a noticeable influence on measurements involving the rate of weight loss for the sample under investigation.[9]

It should further be noted that the DTA signal as obtained from a TG/DTA equipment is slightly different from the signal that is recorded from a conventional DTA unit. In a conventional DTA unit, differential temperature measurements are taken for a sample enclosed in a crimped crucible. For TG/DTA units, the DTA data obtained are generated from an open crucible. The event of the sample undergoing the process of sublimation is inevitable and this will therefore have a marked effect on the DTA signal. However, it is not possible to use a crimped crucible to collect TG data because the principle of thermogravimetric analysis is based on observations of mass changes while the sample is being heated.

2.4 Advantages of Simultaneous Thermogravimetry and Differential Thermal Analysis

There are several advantages that present an excellent case for advocating the use of simultaneous techniques. The obvious incentive of the time and cost efficiency of operating a simultaneous unit is perhaps the most important factor from an administrative point of view. However, there exist other technical aspects to the advantages derived from using simultaneous equipment.

If separate single techniques were used to analyze non-representative sample materials, a direct comparison of the TG and DTA curves would prove to be difficult, due to the nonuniform nature of the sample. By running a single sample of this sort of material with a simultaneous technique, the problem of analyzing nonrepresentative samples is eliminated. Some typical examples of nonuniform materials include coal, limestone and cement.

Another advantage of simultaneous techniques is that the analyst may ensure that the thermal environment and the sample temperatures are identical for the collected data sets. Hence, this technique yields data that are obtained under the same experimental conditions and the relationship between the TG and DTA traces are actual. This also obliterates the hassle of attempting to match the experimental conditions of two separate analytical units, in order to obtain comparable data.

In addition, the use of simultaneous techniques also allows for the regulation of the ambient temperature and the flow rate conditions of an experimental set-up. It is often assumed that the atmospheric environment immediately in contact with the sample for two different single techniques remains unchanged. In practice, this may differ by virtue of the discrepancies in the geometry of the sample container and the furnace enclosing the sample.

3 LINKING COMPUTER SYSTEMS WITH SIMULTANEOUS THERMOGRAVIMETRY AND DIFFERENTIAL THERMAL ANALYSIS

The tremendous influence of microcomputers on the evolution of thermal analytical techniques and applications is indisputable.[10–12] Consequently, the computer workstation is a vital part of any thermal analysis set-up as it is responsible for a multitude of tasks. These responsibilities range from the definition and execution of the various experimental conditions for a run, data capture, the storage and retrieval of collected data, the display of acquired data sets in the right format and the provision of a suitable arena for the application of sophisticated analysis programs to the acquired data.[13,14]

These days, commercial simultaneous instruments are often equipped with computer interfacing for thermal analysis units. A schematic representation of a thermal analysis system linked with a computer workstation is shown in Figure 4. Currently, there are numerous data analysis programs made available to an analyst. For example, there are programs that will calculate the kinetic parameters for isothermal kinetic runs at various

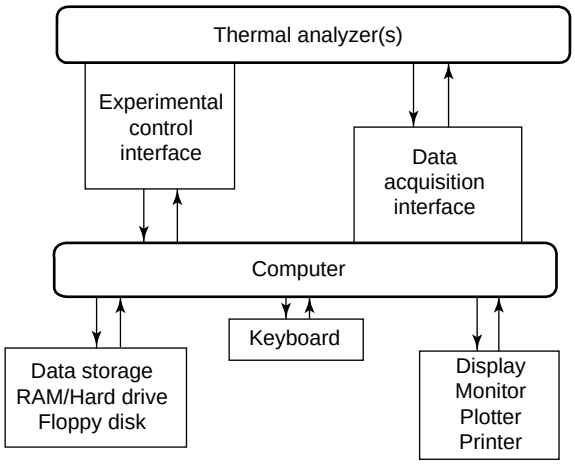

Figure 4 A model diagram illustrating the link of a computer workstation to a simultaneous thermal analyzer.

temperatures using raw data collected from a sample run. Obviously it is still feasible to analyze your data with the aid of a calculator or the use of spreadsheets. However it is now possible to do the same calculations in a matter of seconds.

While it is apparent that a computer workstation is a great tool to have for use with a simultaneous thermal analysis system, it would be judicious to note that the computer data, which are acquired from a unit's temperature sensors, are not always linear. This means that the computer acquisition system should be able to accurately convert the data signal to a temperature. In most cases, an analyst will require a 'change of scale' potential from the equipment and a computer program which is able to record the percentage mass change against temperature. Often, a derivative plot of the collected TG data is also desirable. Most of the newer simultaneous systems have interfaced computers which are able to process the collected data in many more ways, as required by the objectives of the particular experiment.

4 EVOLVED GAS ANALYSIS AND ASSOCIATED ANALYTICAL TECHNIQUES

The success of simultaneous techniques has led to the coupling of other conventional analytical methods with a TG/DTA system. Analytical methods such as MS, FTIR spectroscopy, GC, condensation of volatiles and chemical analysis have proven to be suitable for analyzing the gaseous output of a sample as it undergoes thermal treatment in a TG/DTA system. The main goals of coupling these techniques are to enhance the specificity of a single system and to eliminate the problems that

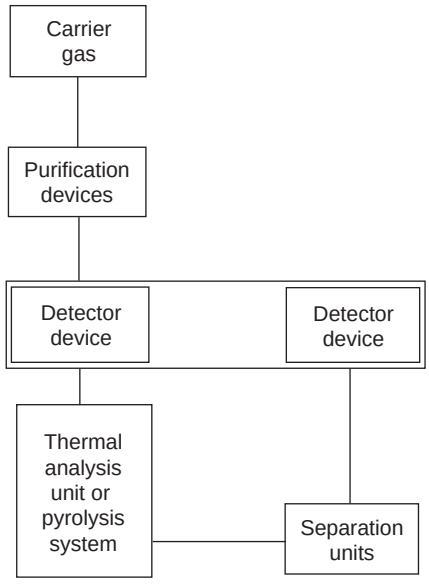

Figure 5 A schematic diagram detailing the various components that are found within an EGA system.

are associated with trying to compare data sets that were collected under different experimental conditions.

EGA is a technique that is rarely used alone. In most cases, it is added to an existing thermal analysis system, to enable a detailed study of the gaseous species that evolve from a sample as it undergoes thermal treatment.[15] The basic components of an EGA system are presented in Figure 5. In order to analyze gases from solid or liquid samples, a pyrolysis unit is needed to induce the thermal breakdown of the samples into its gaseous species. A TG/DTA system is often used for this purpose. A purified carrier gas (such as helium) is then utilized to direct the evolved gas species from the sample to a detector. There are many detection methods available to analyze the gaseous samples. The analytical methods of MS, FTIR and GC will be discussed in the following sections. A brief mention of other less common detection techniques will also be presented.

It should be noted that the method of EGA is not to be confused with evolved gas detection (EGD). EGA refers to a method that involves the quantitative detection and the determination of the identity of the various gases that are given off as the sample undergoes physical and chemical changes upon heating, while EGD merely entails the detection of gases that are evolved from a sample as it is being heated.

4.1 Mass Spectrometry

MS is an analytical technique that is widely used and often associated with EGA.[16,17] It is an excellent method for the detection and identification of gaseous species.

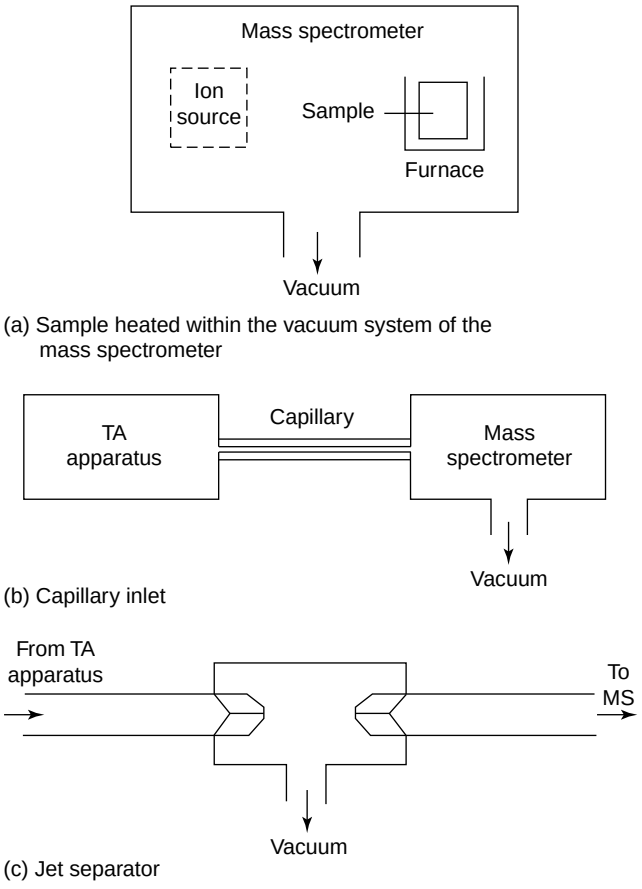

Figure 6 A schematic model of a coupled thermal analyzer and mass spectrometer used for EGA.

A prototype of a basic set-up of a TG/DTA/MS unit used for EGA is shown in Figure 6. The idea of the direct linkage of a thermal analyzer to the mass spectrometer has merit, as a simultaneous technique would overcome interfacing problems. However, the conditions of thermal decomposition under a controlled temperature program are rather unique and therefore the results should be viewed cautiously.

One of the earliest reports of the coupling of MS with EGA was published by Gohlke and Langer.[18] Over the years, several designs for linking a mass spectrometer to a TG/DTA system have been established.[19–22] Today, most of the modern commercial systems follow two basic designs and these are illustrated in Figures 7 and 8.

Though it is possible to purchase commercial units, it is also feasible to interface an existing thermal analyzer unit with a mass spectrometer. In the 1960s, Zitomer managed to interface a Du Pont Model 950 TG unit with a mass spectrometer in his laboratory.[23] A schematic diagram of his apparatus is provided in Figure 9.

Before an attempt is made to couple a thermal analysis unit with a mass spectrometer, one has to consider the

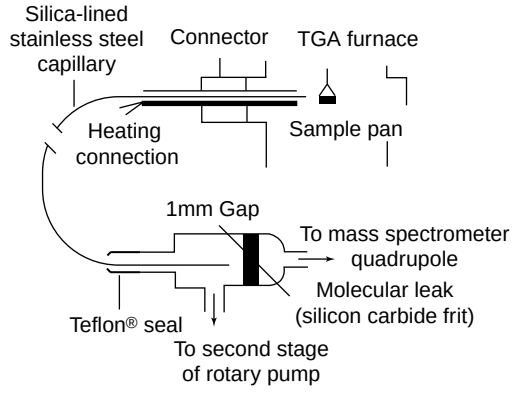

Figure 7 Diagram showing the basic design for a modern TG/MS unit.

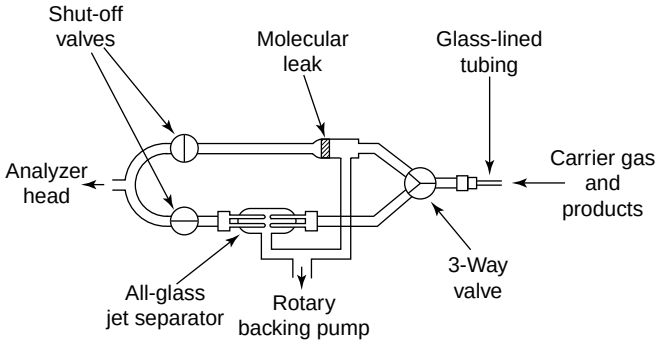

Figure 8 Model of a dual inlet system for a thermal analyzer coupled with MS.

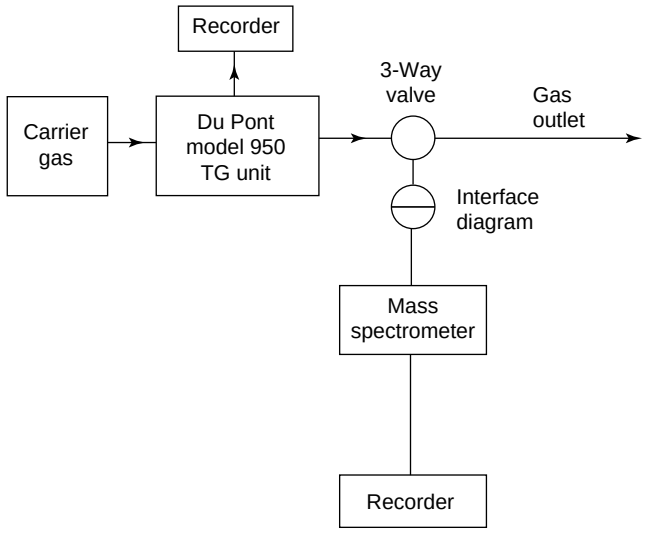

Figure 9 A schematic diagram of the apparatus assembled by Zitomer.

requirements of the equipment set-up. There should be a means of regulating the pressure of the evolved gas species from the thermal analysis unit as it is being channeled

into the mass spectrometer. A thermal analyzer is usually operated under atmospheric pressure conditions while the gas species that are introduced to a mass spectrometer should typically be of low pressure.

Generally, the use of a capillary tube or a diaphragm is sufficient to act as an interface to cope with the primary stage of reducing the differences in pressures between both of the analytical units. In a basic system, the evolved gas is extracted from the sample chamber through a flexible heated silica-lined steel capillary. Other methods of pressure reduction include the use of jet or orifice separators.[24,25] If required, a secondary pressure reduction can be achieved with the use of a molecular leak such as a silicon carbide frit.[26] The purpose of these manipulations is to minimize the chance of secondary reactions and to collect the EGA data as expeditiously as possible.

4.2 Fourier Transform Infrared Spectroscopy

The coupling of FTIR spectroscopy with a thermal analysis unit to perform EGA is another well-established method for the analysis of gaseous species.[27–30] The main feature of this system is the utilization of an infrared cell (set at a fixed wavelength) that is used as the detection chamber. The experimental design of this system involves channeling the gaseous products at atmospheric pressure, through a pathway of inert carrier gas and into the detection chamber. An ideal choice for the inert carrier gas would be dry nitrogen. A schematic representation of a typical system is shown in Figure 10.

There are several advantages to this method of EGA, including high specificity and the relatively fast data acquisition times. In addition, this is also an ideal technique for the analysis of samples which have proven to be difficult to study on a mass spectrometer. Some typical examples of such materials are carbon monoxide, carbon dioxide and water.

4.3 Gas Chromatography

The analytical technique of GC is not used as a detection method in an EGA system. Rather, the function of a gas chromatographic column is to separate the gaseous products evolved from a thermal analyzer before passing it on to a detector.[31–33] The main purpose of using a TG/DTA/GC/MS set-up is to induce a partial separation of the evolved gaseous species before they are further analyzed. Therefore, the scheme of equipment used in such an EGA system proceeds from a thermal analyzer, followed by a GC separation column and into a detector of choice. Frequently, the detector selected for use with this system is a mass spectrometer.

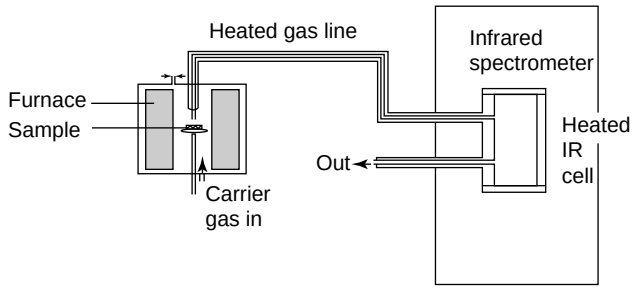

Figure 10 A schematic diagram illustrating the basic design for a thermal analyzer–FTIR system.

5 APPLICATIONS OF SIMULTANEOUS TECHNIQUES

In 1969, an interesting application of simultaneous techniques involved the use of a system that was specifically developed to analyze the characteristics of soil found on Mars.[34] The continued development and improvement of simultaneous thermal analyzers has brought forth a whole new range of applications. These applications have been used to study a vast selection of materials. In the following sections, the applications of simultaneous techniques to biological and pharmaceutical materials, organic and polymeric compounds and inorganic compounds will be stated briefly. While the various applications of simultaneous techniques mentioned here are not a comprehensive survey of all the applications in this area of thermal analysis, it is hoped that the information will challenge new and advanced analysts in exploring the uncharted territories of new applications to come.

5.1 Biological and Pharmaceutical Materials

Simultaneous techniques as applied to biological and pharmaceutical materials, have become indispensable tools to scientists in research and development laboratories. In the field of pharmaceutical studies, there are numerous applications. For example, the behavior of ibuprofen has been extensively studied with the aid of TG/DTA experiments.[35,36] The melt and evaporation processes of ibuprofen are easily observed from the data acquired from the simultaneous equipment. Other experiments which have been conducted on simultaneous thermal analyzers, include the evaluation of lactose crystalline forms with TG and DSC,[37] the compatibility of ascorbic acid,[38] nefazodene[39] and cephalexin monohydrate[40] with excipients.

5.2 Inorganic Compounds

Some recent investigations of inorganic compounds with the aid of simultaneous techniques include the growth of

MgO-doped $LiNbO_3$ single-crystal fibers,[41] the thermal behavior catalysts with superacid properties,[42–44] the thermal decomposition of ceric ammonium nitrate[45] and the thermal degradation processes of $Zr(SO_4)_2 \cdot 4H_2O$.[46] As with the field of pharmaceutical studies, simultaneous techniques are frequently used to perform routine analyses for novel inorganic compounds.

5.3 Organic and Polymeric Compounds

The use of simultaneous thermal analysis methods in the investigation of organic and polymeric compounds has evolved over the years. For example, a comprehensive analysis of starch has been conducted with the use of a TG/DTA unit.[47–59] Another example of recent applications in this area is the characterization of the resistance of diamond layers to oxidation on metal surfaces.[60]

ABBREVIATIONS AND ACRONYMS

DC	Direct Current
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
DTG	Derivative Thermogravimetry
EGA	Evolved Gas Analysis
EGD	Evolved Gas Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
MS	Mass Spectrometry
TG	Thermogravimetry
TG/DTA	Simultaneous Thermogravimetry and Differential Thermal Analysis

RELATED ARTICLES

Coatings **(Volume 2)**
Thermal Analysis of Coatings

Environment: Trace Gas Monitoring **(Volume 3)**
Fourier Transform Infrared Spectrometry in Atmospheric and Trace Gas Analysis

Environment: Water and Waste **(Volume 3)**
Gas Chromatography with Selective Detectors for Amines

Polymers and Rubbers **(Volume 9)**
Pyrolysis Techniques in the Analysis of Polymers and Rubbers • Thermogravimetry of Polymers

Gas Chromatography **(Volume 12)**
Gas Chromatography: Introduction

Mass Spectrometry **(Volume 13)**
Mass Spectrometry: Overview and History

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction • Differential Scanning Calorimetry and Differential Thermal Analysis • Inorganic Systems, Thermal Analysis Applications to • Thermogravimetry

REFERENCES

1. G. Lombardi, *For Better Thermal Analysis*, ICTA, Rome, 1977.
2. F. Paulik, J. Paulik, L. Erdey, *Z. Anal. Chem.*, **160**, 241 (1958).
3. J. Paulik, F. Paulik, *Simultaneous Thermoanalytical Examinations by Means of the Derivatograph*, Wilson and Wilson's Comprehensive Analytical Chemistry, Vol. XIIA, Elsevier, Amsterdam, 1981.
4. F. Paulik, *Special Trends in Thermal Analysis*, John Wiley & Sons, Chichester, 1995.
5. E.L. Charsley, J. Joannou, A.C.F. Kamp, M.R. Ottaway, J.P. Redfern, *Proc. of 6th ICTA*, Birkhaeuserm, Basel, 237, Vol. 1, 1980.
6. P. Le Parlouer, *Proc. of 8th ICTA*, Thermochim. Acta, **92**, 371 (1985).
7. P. Le Parlouer, *Thermochim. Acta*, **121**, 307 (1987).
8. C. Marcozzi, K. Reed, *Int. Lab.*, **23**(6), 10 (1993).
9. T. Daniels, *Thermal Analysis*, Kogan Page Limited, London, 1973.
10. W.W. Wendlandt, *Thermochim. Acta*, **5**, 225 (1973).
11. B. Wunderlich, *Int. Lab.*, **32**, October (1982).
12. E.B. James, *Chem. Br.*, **18**, 620 (1982).
13. M.E. Brown, *Introduction to Thermal Analysis: Techniques and Applications*, Chapman and Hall, London, 1988.
14. P.J. Haines, *Thermal Methods of Analysis: Principles, Applications and Problems*, Blackie Academic and Professional, London, 1995.
15. A. Kambe, K. Horie, T. Suski, *J. Therm. Anal.*, **4**, 461 (1972).
16. D. Dollimore, G.A. Gamlen, T.J. Taylor, *Thermochim. Acta*, **75**, 59 (1984).
17. M.R. Holdiness, *Thermochim. Acta*, **75**, 361 (1984).
18. R.S. Gohlke, H.C. Langer, *Anal. Chim. Acta*, **36**, 530 (1966).
19. R.M. Lum, *Thermochim. Acta*, **18**, 73 (1977).
20. P.K. Gallagher, *Thermochim. Acta*, **26**, 175 (1978).
21. F. Kinsbrow, P.K. Gallagher, A.T. English, *Solid State Electron.*, **22**, 517 (1979).
22. D. Price, D. Dollimore, N.J. Fatemi, R. Whitehead, *Thermochim. Acta*, **42**, 323 (1980).

23. F. Zitomer, *Anal. Chem.*, **40**, 1091 (1968).
24. E. Clarke, *Thermochim. Acta*, **51**, 7 (1981).
25. E. Kaisersberger, W.-D. Emmerich, *Proc. 7th ICTA*, Wiley, Chichester, 1982.
26. E.L. Charsley, C. Walker, S.B. Warrington, *J. Therm. Anal.*, **40**, 983 (1993).
27. D.J. Morgan, *J. Thermal Anal.*, **12**, 245 (1977).
28. J.O. Lephardt, *Appl. Spectros. Rev.*, **18**, 265 (1982–1983).
29. D.J. Morgan, S.B. Warrington, S.St.J. Warne, *Thermochim. Acta*, **135**, 207 (1988).
30. S.A. Liebman, D.H. Ahlstrom, P.R. Griffiths, *Appl. Spectros.*, **30**, 255 (1976).
31. P.D. Garn, *Talanta*, **11**, 1417 (1964).
32. P.D. Garn, G.D. Anthony, *Anal. Chem.*, **39**, 1445 (1967)
33. G. Blandenet, *Chromatographia*, **2**, 184 (1969).
34. E.M. Bolin, 'Thermal Analysis', *Proc. of 2nd ICTA*, Worcester, Massachusetts, USA, 1969.
35. S. Lerdkanchanaporn, D. Dollimore, *J. Therm. Anal.*, **49**, 879 (1997).
36. S. Lerdkanchanaporn, D. Dollimore, K.S. Alexander, *J. Therm. Anal.*, **49**, 887 (1997).
37. N. Drapier-Beche, J. Fanni, M. Parmentier, M. Vilasi, *J. Dairy Sci.*, **80**, 457 (1997).
38. S. Lerdkanchanaporn, D. Dollimore, K.S. Alexander, *Thermochim. Acta*, **284**, 115 (1996).
39. F. Balestrieri, A.D. Magri, A.L. Magri, D. Marini, A. Sacchini, *Thermochim. Acta*, **285**, 337 (1996).
40. M.A. Hassan, J. Kaloustian, H. Ramsis, K.A. Khaled, T.H. El-Faham, S.S. Tous, L. Maury, J. Joachim, *S.T.P. Pharma. Sci.*, **4**, 435 (1995).
41. A. Dakki, M. Ferriol, M.T. Cohen-Adad, *Eur. J. Solid State Inorg. Chem.*, **33**, 19 (1996).
42. A. Clearfield, A.I. Borton, L.N. Borton, *Spec. Publ – R. Soc. Chem.*, **182**, 338 (1996).
43. A.I. Borton, L.N. Borton, A. Clearfield, *Chem. Mater.*, **9**, 1854 (1997).
44. R.B. Borade, A. Clearfield, *Microporous Mater.*, **5**, 289 (1996).
45. N. Guillou, J.-P. Auffredic, D. Louer, *J. Solid State Chem.*, **122**, 59 (1996).
46. F. Li, J. He, B. Zhang, P. Sun, X. Duan, Z. Wang, *Beijing Huagong Daxue Xuebao, Ziran Kexueban*, **23**, 58 (1996).
47. S. Xu, X. Wan., C. Feng, Y. Wang, Q. Zeng, Y. Fu, X. Guo, L. Chen, *Theory Pract. Energ. Mater.*, **484** (1996).
48. P. Aggarwal, D. Dollimore, *Talanta*, **43**, 1527 (1996).
49. P. Aggarwal, D. Dollimore, *Thermochim. Acta*, **291**, 65 (1997).
50. F.F. Yamin, L. Svendsen, P. White, *J. Cereal Chem.*, **74**, 407 (1997).
51. R. Acquistucci, R. Bussi, A.D. Magri, A.L. Magri, *J. Anal. Chem.*, **357**, 97 (1997).
52. A. Kawabata, N. Takase, S. Akuzawa, S. Sawayama, *Oyo Toshitsu Kagaku*, **43**, 471 (1996).
53. M.J. Gidley, D. Cooke, A.H. Darke, R.A. Hoffmann, A.L. Russell, P. Greenwell, *Carbohydr. Polym.*, **28**, 23 (1995).
54. C.M. Bryant, B.R. Hamaker, *Cereal Chem.*, **74**, 171 (1997).
55. L. Slade, H. Levine, M. Wang, J. Ievolella, *J. Therm. Anal.*, **47**, 1299 (1996).
56. R.-M. Huang, C.-Y. Yeh, C.-Y. Li, *Zhongguo Nongye Huaxue Huizhi*, **35**, 86 (1997).
57. P.J. Stenhouse, J.A. Ratto, N.S. Schneider, *J. Appl. Polym. Sci.*, **64**, 2613 (1997).
58. S. Jacobsen, H.G. Fritz, *Polym. Eng. Sci.*, **36**, 2799 (1996).
59. T.M. Stein, R.V. Greene, *Starch/Staerke*, **49**, 245 (1997).
60. M. Wang, Y. Wang, C. Guan, J. Zang, *Fuhe Cailiao Xuebao*, **13**, 48 (1996).

Thermogravimetry

Jeffrey G. Dunn[1]
Curtin University, Perth, Western Australia

1 Introduction 1
2 Components of Thermogravimetry Apparatus 3
2.1 The Balance and Methods of Calibration 3
2.2 The Sample Chamber and Atmosphere Control 4
2.3 The Furnace 4
2.4 Measurement of Temperature and Methods of Calibration 4
2.5 The Temperature Programmer 5
2.6 Data Acquisition and Manipulation 5
3 Effect of Experimental Variables 5
3.1 Sample Preparation 6
3.2 Sample Mass 6
3.3 Sample Containers 7
3.4 Heating Rate 7
3.5 The Atmosphere 8
3.6 Optimizing Thermogravimetry Results 8
4 Reporting Thermal Analysis Results 9
4.1 Properties of Substances 9
4.2 Experimental Conditions 9
4.3 Data Acquisition and Manipulation Methods 9
5 Applications 10
5.1 Adsorption and Desorption Reactions 10
5.2 Sublimation and Vaporization 11
5.3 Thermal Decomposition and Thermal Stability 12
5.4 Gas–Solid Reactions 13
5.5 Solid–Solid Reactions 15
5.6 Analytical Applications 15
5.7 Kinetic Measurements 17
Abbreviations and Acronyms 19
Related Articles 19
References 19

Thermogravimetry (TG or TGA) is a technique in which the mass of a substance is measured as a function of temperature whilst the substance is subjected to a controlled temperature program.[1,2] Commonly materials can be heated or cooled in the temperature range from −160 to 1600 °C at rates of fractions of a degree per minute up to 100 °C min^{-1}. Sample sizes are from 10–100 mg, although apparatus is available which can take larger sample masses. Different atmospheres can be used, from inert gases such as nitrogen, to reactive gases such as oxygen, hydrogen and vapors. The sensitivity of the apparatus is usually in the microgram range, with an accuracy of ±1% and a similar level of precision. The results obtained from TG experiments are subject to a variety of experimental variables, and an understanding of the effects of these is essential to a correct interpretation of the results. TG can be used to study any physical or chemical reaction which gives a mass change, and includes absorption and desorption of gases, sublimation and vaporization, the thermal decomposition of a wide range of materials from inorganic compounds to polymers, the reaction between a gas and a solid, and solid–solid reactions. TG has also been widely used to carry out kinetic studies on solid materials, with the aim of determining the reaction model and the kinetic parameters of activation energy (E) and frequency factor (A). However, considerable care is required in the interpretation of the results obtained.

1 INTRODUCTION

The definition of TG as given by the International Confederation for Thermal Analysis and Calorimetry (ICTAC) is "a technique in which the mass of a substance is measured as a function of temperature whilst the substance is subjected to a controlled temperature programme".[1,2]

A controlled temperature program can mean:

- Heating or cooling the sample at some defined constant rate, and monitoring the mass change as a function of temperature. This is the most common mode of operation.
- Holding the sample at some constant temperature and monitoring the mass change as a function of time, called isothermal TG.
- Varying the heating rate in such a manner that the rate of mass change is constant. This mode of operation is called controlled rate thermal analysis (CRTA).[3,4]
- Varying the heating rate continuously in order to maximize the resolution of mass changes. When no mass change is taking place, the heating rate is high, but then slows down as a mass change begins. Once

[1] J. Dunn is now at the Department of Chemistry, University of Toledo, Toledo, OH 43606, USA.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the mass loss is complete the heating rate increases again until the next mass loss. This method is called high resolution TG by some manufacturers.

- Superimposing a sinusoidal temperature program on the underlying linear heating rate. This technique is called modulated thermogrametry (MTG or MTGA), and has been proposed as a new method for the determination of kinetic information.[5]

Thermogravimetric data is usually presented as a plot of the mass against time or temperature (the TG curve), with the mass loss on the ordinate plotted downwards and mass gains plotted upwards (Figure 1). The TG curve can be defined in several ways. The temperature at which a change in mass is first detected is called the initial temperature T_i, or onset temperature or procedural decomposition temperature. The initial temperature is not sufficiently well defined to use as a satisfactory reference point, since its detection is dependent on factors such as the sensitivity of the TG apparatus, and the rate at which the initial reaction occurs. If the initial rate of mass loss is very slow, then the determination of T_i can be uncertain.

A more satisfactory approach is to use the extrapolated onset temperature, T_e, which provides consistent values (Figure 1). If, however, the decomposition extends over a wide temperature range and only becomes rapid in its final stages, the extrapolated onset temperature will differ considerably from the onset temperature. For this kind of reaction, it is more satisfactory to measure the temperature at which a fractional weight loss, α, has occurred (T_α) (Figure 1). Clearly the temperature $T_{0.05}$ is close to that at the start of the reaction and $T_{0.50}$ is the mid-point of the reaction. In order to define the complete range of reaction, two further temperatures T_o and T_f may be identified as shown in Figure 1.

Alternatively the mass change can be presented as the derivative thermogravimetric curve which is a plot of the rate of change of mass with respect to time or temperature against time or temperature (Figure 2). Differential thermogravimetry (DTG) mass losses should also be plotted downwards and gains upwards. When a horizontal plateau is obtained in a TG curve, a minimum at which $dm/dt = 0$ is observed in the DTG curve. A maximum in the DTG curve corresponds to a point of inflection in the TG curve at which mass is lost (or gained) most rapidly. Minima can also occur in a DTG curve at which $dm/dt > 0$, corresponding also to a point of inflection in the TG curve, rather than to a horizontal plateau.

The DTG curve is particularly useful in displaying subtle changes in mass which are not easily discernible on the TG curve. The DTG curve also allows the ready determination of the temperature at which the rate of mass loss is a maximum (T_p), and this provides an additional characteristic temperature to those already described in Figure 1.

The DTG curve can assist the operator to be aware of overlapping reactions. A single reaction which occurs over a small temperature range gives rise to a single well-defined peak (Figure 3a). Two reactions which are partially overlapping result in two peaks in which the minimum does not return to the baseline (Figure 3b). If a slow reaction occurs over a wide temperature range, followed by a fast reaction, then the former gives a small broad initial DTG peak followed by a well-defined one (Figure 3c). Minor reactions which occur during or near a major reaction can frequently be identified by the appearance of a shoulder or separate small peak in the DTG curve (Figure 3d). Finally, slow reactions which occur over wide temperature ranges, in which other reactions happen, produce gradient changes in the DTG curve. Because the subtleties of the TG curve are visually emphasized in the DTG curve, the latter are frequently recorded as a 'fingerprint' of a new, unknown or standard material.

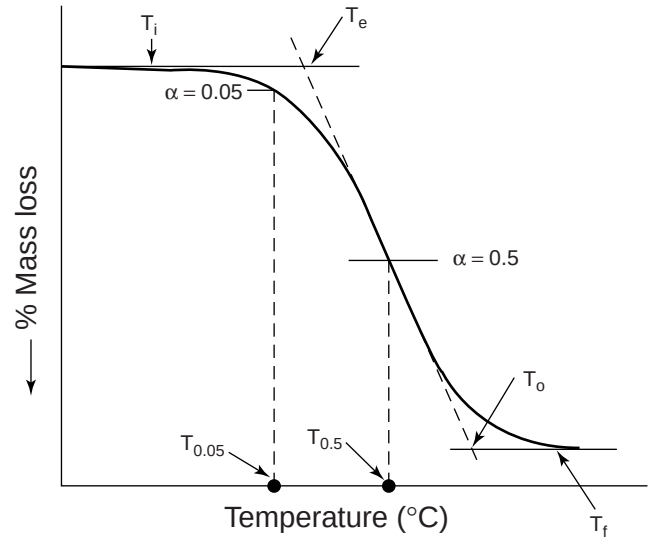

Figure 1 Defining temperatures of a TG curve.

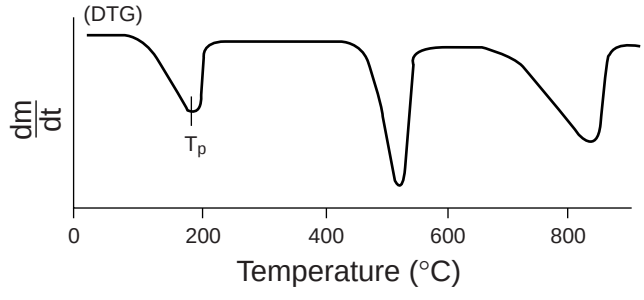

Figure 2 Differential thermogravimetric DTG curve.

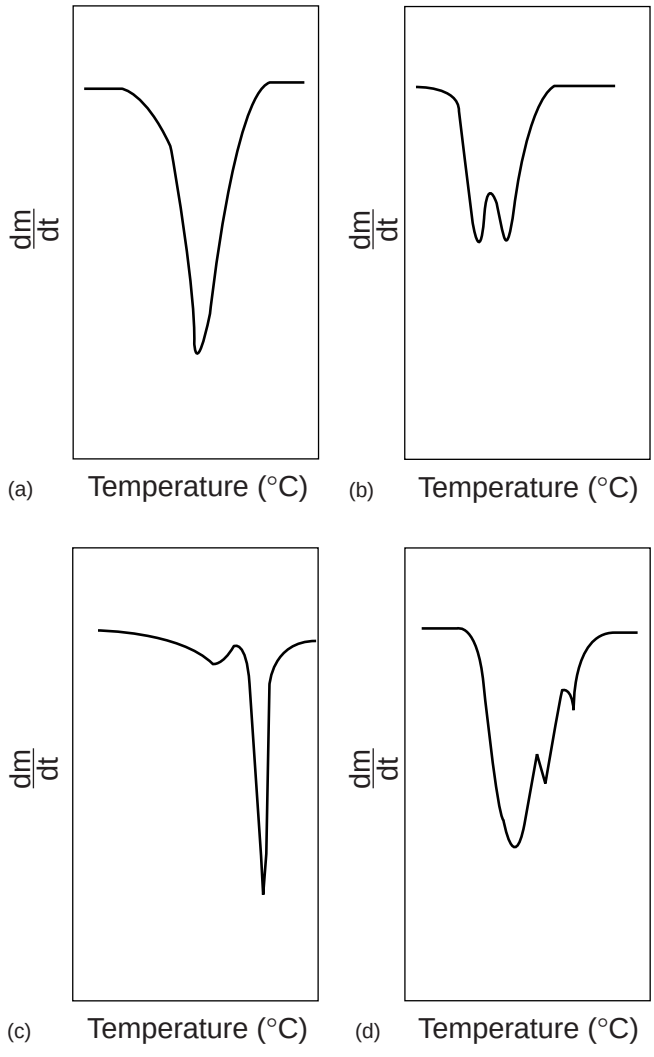

Figure 3 Use of DTG curves to detect multiple reactions. (a) single reaction, (b) two overlapping reactions, (c) slow reaction followed by a fast reaction, (d) minor reactions occurring in the same temperature range as a major reaction. (Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11])

Since the area under the DTG peak is directly proportional to the mass change, the area or even the peak height can be measured for quantitative purposes. The advantage of DTG in this application is that when no weight change is taking place, the baseline returns to the zero position and so only a simple baseline construction is required. The value of TG is considerably enhanced when the product at the end of each mass change is characterized. A range of qualitative and quantitative analytical methods are available which can be performed on a few milligrams of the sample. This kind of information can be used to write the chemical reaction associated with the mass loss. Information on reaction mechanisms can be obtained from techniques such as scanning electron microscopy (SEM).

There is an extensive literature associated with thermal methods of analysis, which includes two journals dedicated to the techniques, the *Journal of Thermal Analysis* and *Thermochimica Acta*. There are also several standard texts available which contain sections on TG, as well as review articles in texts.[6–13]

2 COMPONENTS OF THERMOGRAVIMETRY APPARATUS

All TG apparatus consist of the essential components of a balance, a sample chamber, a furnace with associated temperature programmer, a method of measuring sample temperature, and a means of acquiring and outputting the data collected. TG systems are available from a number of commercial manufacturers.

2.1 The Balance and Methods of Calibration

The most commonly used balance to follow the mass change in TG apparatus is the null deflection microbalance, in which the position of the beam is monitored. When a mass change occurs, a servo motor maintains the beam in its original position, and the restoring force is proportional to the mass change. The null deflection balance has advantages in that the sample stays in a constant position in the furnace. Sample masses of typically 10–20 mg are sufficient to be able to detect mass changes accurately and reliably. TG apparatus capable of taking larger samples, in the gram range, are also available, although tend to be used for specialist applications.

There are a number of factors which can affect the accuracy and precision of mass measurement and these are discussed in detail in Dunn and Sharp.[11] For most usage, that is where the mass loss is above 1% of the sample mass, the errors that arise in microbalance TG apparatus can be ignored. The extent of typical errors can be assessed by following a simple protocol:

- The microbalance is calibrated with standard weights to check that the balance is operating within specification (this procedure should be done at regular intervals, say once a month). Standardized masses are often provided as part of the accessories when a commercial thermobalance is purchased, or sets of such mass calibrants can be obtained from organizations such as ICTAC, the National Physics Laboratory (NPL), or NIST (National Institute for Science and Technology).
- A blank experiment is carried out with empty sample pans, which will enable the amount of noise and

base line stability (including balance drift) to be determined.

- A further experiment is carried out with a typical sample size of an inert material such as calcined alumina, which will enable the buoyancy effect to be established. The buoyancy effect produces an apparent mass gain which increases with increasing temperature and with increasing sample mass.
- Experiments are carried out with a standard material which decomposes to a known product, and the observed mass loss is compared with that expected theoretically. Multiple experiments on the same sample will give statistical information on the accuracy and reproducibility of the results. If the difference is within the acceptable error, experiments with unknown samples can then be carried out under the same experimental conditions.

2.2 The Sample Chamber and Atmosphere Control

The sample chamber usually consists of a glass envelope which encloses the balance, partly to protect the balance against corrosion or damage, but also to enable experiments to be carried out in a controlled atmosphere. One simple way to protect the balance mechanism from corrosion is to flow an inert gas over the balance first and then over the sample. A reactive gas, such as oxygen, can be introduced as a second gas stream nearer to the sample, but again arranged in such a way that gaseous products are swept away from the balance mechanism. A typical example is the reaction between chlorine and lead sulfate.[14] Other configurations are illustrated in Dunn and Sharp.[11] Gas flow rates are measured with a rotameter.

Pahlke and Gast[15] have overcome the problem of corrosive gases attacking the microbalance by designing a system in which the balance and sample chambers are completely separated. The mass change is monitored by a magnetic link. The sample chamber is often sufficiently sealed to permit work under vacuum down to 1 torr (1 torr = 133 Pa), and many thermobalances can operate down to 10^{-3} torr (= 0.133 Pa), although it is not usual to find instruments fitted with a means of monitoring the vacuum level. TG apparatus capable of operating at high pressures is available.

2.3 The Furnace

A furnace provides the thermal environment for the sample. The temperature ranges for commercially available instruments given in the manufacturers literature tend to be for the basic model, which operates typically in the temperature range ambient to 1000 °C, but other models may operate from −196 °C to 500 °C, or ambient to 2400 °C, with other ranges between.

One of the requirements of a furnace is that there should be a region, called the hot zone, within which the temperature is constant over a finite distance. The importance of this is that the sample should always be in the hot zone, so that all of the sample is at the same temperature. This is one of the reasons why it is advantageous to use a null deflection balance, as the sample is always in the same position in the furnace.

There are three main arrangements of the furnace relative to the weighing arm of the balance. Lateral loading is a horizontal arrangement, in which the balance arm extends horizontally into a furnace aligned horizontally. Two vertical arrangements are possible, one with the balance above the furnace with the sample suspended from the balance arm (bottom loading); and the other with the balance below the furnace, and the sample resting on a flat platform supported by a solid ceramic rod rising from the balance arm (top loading).

2.4 Measurement of Temperature and Methods of Calibration

The sample temperature is measured by a thermocouple ideally in contact with the sample, but in practice this is difficult and the best compromise situation is one in which the thermocouple is placed at a fixed distance from the sample. Again this is an advantage of the null deflection balance since the position of the sample is relatively fixed. Even so, the accurate measurement of temperature in TG experiments is difficult. Some typical configurations of thermocouples with respect to the sample are shown in Figure 4. The thermocouples are often made of platinum/rhodium and their alloys.

One of the better calibration methods involves the use of the Curie point transition in metal samples. In this procedure, a ferromagnetic material is placed in the sample crucible as the test specimen and a magnet is located above or below the crucible which creates a magnetic flux aligned to the gravitational field. When the material is heated at a particular temperature

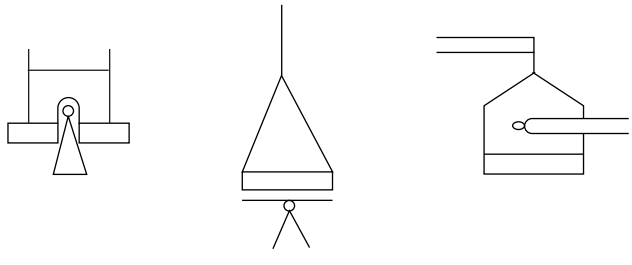

Figure 4 Possible arrangements of thermocouples to measure sample temperature in a TG apparatus. (Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11])

the ferromagnetic material becomes paramagnetic. The temperature at which the magnetic properties change is defined as the Curie temperature. At the Curie point a significant apparent mass loss in the specimen occurs.[16] Five metals have been certified as ICTAC certified reference materials for TG, GM761, as shown in Table 1.[17] The materials cover the temperature range 200–800 °C.

More reproducible and precise results for determining transition temperature can be achieved using simultaneous thermomagnetometry/differential thermomagnetometry/differential thermal analysis (TM/DTM/DTA).[18] The results from five different laboratories, based on DTM, showed that a very high degree of reproducibility between the laboratories was obtained with a standard deviation 2σ value of ± 0.2 °C. The potential of nickel–cobalt and nickel–copper alloys, which give magnetic transitions from 1100 °C to below room temperature, have also been investigated.[19]

The obvious advantage of the Curie point method is that the ferromagnetic material calibrates temperature directly at the sample position and the thermobalance can be calibrated under the same conditions as used with a sample. The magnetic transitions are reversible, so that the calibrant can be encapsulated to protect it against oxidation and used repeatedly for calibration purposes.

If the sample is visible when placed in the TG sample pan, then the melting of a pure compound can be used for calibration purposes. The method is to heat the calibrant to just below its melting point and then increase the temperature in small increments of say one degree until the calibrant just melts. Sets of melting calibration standards are available from organizations such as NIST, covering the temperature range ambient to 1000 °C. The obvious disadvantage of this method is that the calibrant needs to be visible, and the calibration is carried out under near isothermal conditions instead of the rising temperature program most commonly used for experimental purposes.

Table 1 Measured values of T_e for ICTAC certified reference materials for TG, GM761[17a]

Material	Temperature range of transition (°C)	Mean (°C)	Standard deviation
Permanorm 3	242–263	253.3	5.3
Nickel	343–360	351.4	4.8
Mumetal	363–392	377.4	6.3
Permanorm 5	435–463	451.1	6.7
Trafoperm	728–767	749.5	10.9

[a] Reproduced for Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11]

2.5 The Temperature Programmer

The temperature programmer is the device which enables the sample to be heated at a linear and reproducible set of heating or cooling rates. In the isothermal mode of operation, it should be able to hold a fixed temperature to within ±1 °C. The thermocouple output is usually fed to a microprocessor containing tabulated data on thermocouple readings, so that the conversion to temperature is made by direct comparison.

2.6 Data Acquisition and Manipulation

Data acquisition and manipulation is carried out by conversion of the millivolt output from both the balance and the sample thermocouple via an interface, and the digitized data stored on a microprocessor. Once the data has been acquired it can be manipulated in various ways, and some common examples are given as follows:

- rescaling of either the mass or temperature scales, so that the data can be displayed to best advantage;
- differentiation of the mass change to give the DTG curve;
- produce a report quality output of fully labeled TG/DTG curves with the various onset and offset temperatures identified, and mass losses expressed in various units including percentage values. Labels identifying the sample, operator, and experimental conditions used can be added;
- carry out various calculations including a kinetic analysis of the data. This aspect will be discussed in more detail in section 5.7.

The other role of the computer is as a control system, which enables a particular protocol to be established to carry out a repetitive test procedure. Heating and cooling rates can be varied at will, and over a range from fractions of a degree per minute to the maximum that the system will allow. This control function can be extended to switching on or off valves so that different atmospheres can be admitted to the TG equipment.

The availability of automatic sample loaders which allow multiple samples to be presented to the TG apparatus has been the most recent development.

3 EFFECT OF EXPERIMENTAL VARIABLES

The shape of the TG curve is dependent on the properties of the sample and the experimental variables chosen. These parameters include sample preparation, sample mass, sample containers, heating rate, and the atmosphere surrounding the sample. Most of the effects observed are

related to mass and heat transfer between the sample and the TG apparatus, which result in shifts in the temperature range over which the mass change occurs. In some instances the change in conditions will produce a change in the reaction mechanism and produce a TG curve which is quite different to that obtained under other conditions. An obvious example is when the atmosphere is changed from an inert gas to a reactive one, although sometimes a change in heating rate can bring about a change in the reactions that the material undergoes. If reactive materials (such as organic compounds, coals, and sulfides) are subjected to vigorous oxidizing conditions, the material may ignite and the excess heat generated in the reaction may heat the sample above that defined by the temperature programmer. This is evident as a distortion on the heating rate record. Hence it is necessary to know about the effects of these variables in order to interpret the results correctly and also to compare results with those published in the literature.

3.1 Sample Preparation

Most solid samples require some form of sample preparation prior to performing the experiment. Since TG apparatus based on microbalances require sample sizes in the milligram range, it is essential to grind the sample to a small enough particle size, typically less than 100 μm, to enable a representative sample to be taken. Particle size affects the TG record in two ways. First, smaller particles will heat more rapidly than larger ones, and second the rate of reaction becomes greater as the particle size decreases. In general, therefore, the smaller the particle size of the sample, the lower the temperature at the start and the completion of the reaction, which is manifest as a shift in the mass change along the temperature axis but with the general shape of the TG being retained. Figure 5 shows the TG curves for four different sample preparations. A thin film of a fine powder gives the lowest commencing and finishing temperature, followed by a thick film of a fine powder, then a coarse powder and finally at the highest temperatures a pellet of the material.

Some materials, when heated, can be ejected from the sample pan and give a mass loss which may be incorrectly attributed to a chemical reaction rather than a physical loss. Samples which are in this category are those which decrepitate, or which dehydrate with subsequent dissolution of the solid in the water which boils with frothing. Organic samples, if oxidized vigorously, may ignite and the solid material be carried away as a fine ash. Such problems can be solved by use of a cover on the sample pan.

3.2 Sample Mass

The sample mass is important for a number of reasons. For decomposition reactions, the sample mass affects the rate of diffusion of reaction products away from the sample. In general a larger sample mass will not significantly affect the onset temperature of reaction, but T_p and final temperatures will tend to be at higher values. Figure 6 shows the DTG curves for different masses of calcium carbonate.[20] The $T_{0.1}$ values change by 59 °C, the T_p

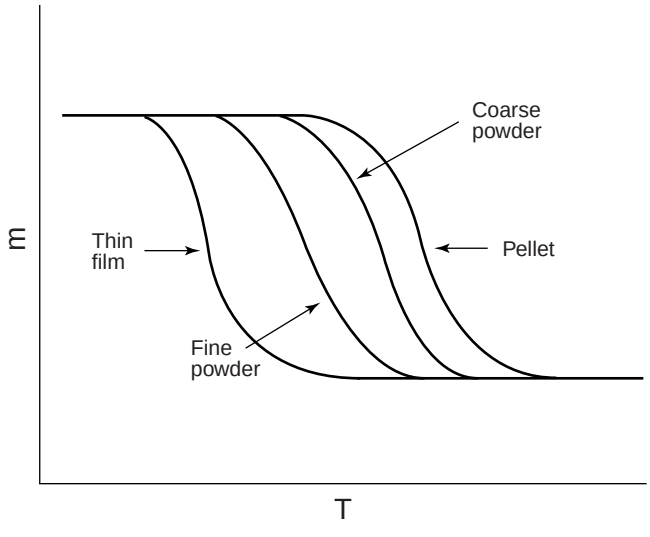

Figure 5 Effect of particle size on the TG curve. (Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11])

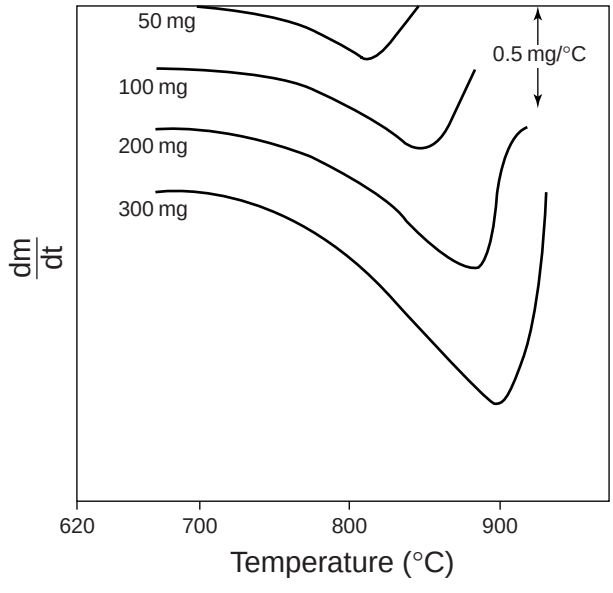

Figure 6 Effect of sample mass on the DTG curve. (Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11])

values by about 80 °C, and the $T_{0.90}$ values by 97 °C, as the sample mass changes from 50–300 mg. This is similar in the reaction between a sample and an introduced reactive gas. The larger the sample the larger will be the temperature interval over which the reaction occurs.

Another effect that needs consideration is the temperature gradients which exist within the sample. Temperature gradients are found in any procedure in which a dynamic heating or cooling rate is used. There is a temperature gradient between the furnace and the sample, as it takes a finite time for heat transfer to take place across the air gap between the furnace and the sample. Only under isothermal conditions will the sample and furnace temperatures be the same. At a constant heating rate this temperature gradient or, as it is more commonly known, the thermal lag, is approximately constant. The thermal lag increases as the heating or cooling rate increases, and also increases as the mass of the sample increases. The various heats of reaction that occur will contribute to the magnitude of the effect as well. Exothermic reactions will tend to heat the sample and reduce the thermal lag, whereas endothermic reactions will tend to cool the sample and increase the thermal lag.

There are also temperature gradients within the sample. Heat diffusion will be dependent on the thermal conductivity of the sample, so that heat gradients in metals will tend to be lower than minerals or polymers. If a large sample of a poorly conducting material is heated then large temperature gradients can be expected and so the rate of reaction will be faster at the exterior of the sample because it is at a higher temperature than the interior. This will cause the reaction to occur over a greater temperature range relative to a smaller sample mass or one with a better thermal conductivity.

These problems are in general minimized in the microbalance-based TG apparatus, since sample sizes are small and heat and mass transfer tend to be rapid. Nevertheless, it is sometimes necessary or advantageous to use large samples (300–1000 mg range) and the associated effects need to be considered.

To reduce heat and mass transport effects, small samples spread as a fine layer are preferred. This ensures that gaseous products are dispersed quickly, thermal gradients in the sample are small, and the reaction between solid and introduced reactive gas is rapid.

3.3 Sample Containers

Both the geometry of the sample container and the material from which it is made can influence the shape of a TG curve. The geometric effect is again related to gas diffusion to and from the sample. Deep narrow necked crucibles will reduce the rate of gas diffusion and tend to make the reaction take place over a larger temperature range relative to a shallow open pan.[21] Sometimes these effects are quite marked and it has been demonstrated that the rate of oxidation of pyrite decreases by about one half as the wall height of the sample pan changes from 2–4 mm.[22]

In certain situations it is advantageous to inhibit gas diffusion deliberately, by use of a lid with either a partial or complete seal, or of special design, for example the ''labyrinth'' crucible of Paulik et al.[23] The partial pressure of the gaseous product reaches atmospheric pressure and remains constant until the end of the process, so leading to the technique being named quasi-isobaric. Lids are of value in preventing sample loss, as well as holding the sample firmly in place.

Crucibles are usually fabricated from either metal (often platinum) or ceramic (porcelain, alumina or quartz) and their thermal conductivity varies accordingly. Care must be taken to avoid the possibility of reaction between the sample and its container, which may result in the corrosion of the crucible and/or a change in the rate of reaction if some catalytic effect is present.

3.4 Heating Rate

The heating rate will affect the temperature range over which a TG curve is observed. As the heating rate increases, the mass loss is usually displaced to a higher temperature range and occurs over a larger temperature range. Faster heating rates may also cause a loss of resolution when two reactions occur in similar temperature ranges and the two mass losses merge. Figure 7 shows the DTG curves for 20 mg of a 50:50 mixture of inorganic bicarbonates heated at 1 and 20 °C min^{-1}, respectively, in nitrogen.[24] The

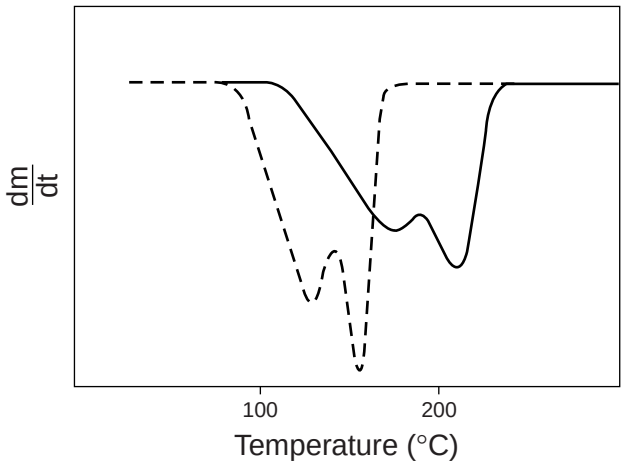

Figure 7 Effect of heating rate on the DTG curve. Solid line, sample heated at 20 °C min^{-1}; dotted line, sample heated at 1 °C min^{-1}. (Reproduced from Buckman and Hawkins, *Chem. in Austral.*[24] with permission.)

bicarbonates decompose on heating with the evolution of carbon dioxide. At the slower heating rate, the reaction commences at a T_e value of 88 °C, the T_p values for the first and second reactions are 127 °C and 154 °C, respectively, and the T_o value is 172 °C. At the faster heating rate the four values shift to 115, 172, 209 and 233 °C.

Changes in heating rate sometimes leads to a change in mechanism which produces a significantly different TG curve.

3.5 The Atmosphere

The atmosphere causes two major effects. First, the presence in the atmosphere of an appreciable partial pressure of a volatile product will suppress a reversible reaction and shift the decomposition temperature to a higher value. This effect can be achieved by having a large sample with a lid, or introducing the volatile component into the inlet of the gas stream. When the atmosphere within the crucible is the result of the decomposition of the reactant, the atmosphere is described as self-generated.[25] The second major effect is the interaction of a reactive gas with the sample, which will change the course of the reaction. Under an inert atmosphere, organic compounds will degrade by pyrolytic decomposition, whereas in an oxidizing atmosphere oxidative decomposition obviously takes place. Hydrogen can be used to study reductive processes, although great care has to be taken to remove all oxygen from the system before carrying out the reaction. Other gases frequently studied are SO_2 and halogens.

A more minor effect is the due to the differences in the thermal conductivity of gases. If two inert gases with differing thermal conductivities are used to conduct two experiments then there will be a difference in the TG curves. This is because the gas with higher conductivity will in general conduct heat to and from the sample at a greater rate than the one with lower thermal conductivity. This will not much affect the onset temperature of reaction, but the T_p and T_f values will be lower for the gas with higher conductivity. This is illustrated in Figure 8, which shows the DTG curves for the same bicarbonate mixture referred to in the previous example.[24] In both experiments the sample mass is 10 mg, and the heating rate is 10 °C min^{-1}, but the purge gases are helium and nitrogen. Helium, with a higher conductivity than nitrogen, causes the mass losses to occur at lower temperatures relative to nitrogen. The effect is not large, with the T_p values for the second reaction differing by 24 °C between the two different gases.

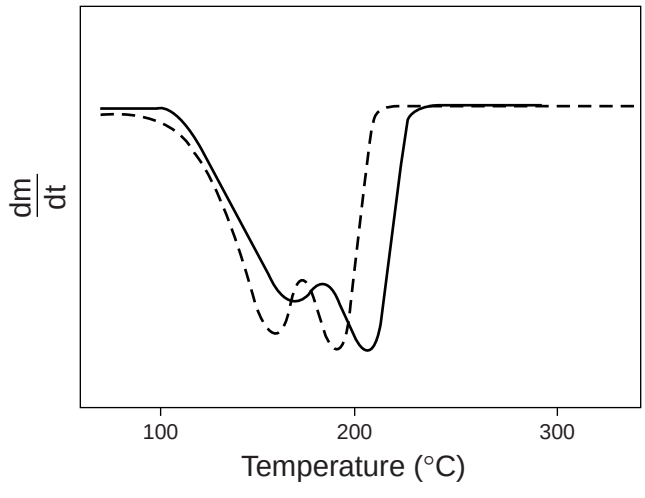

Figure 8 Effect of purge gas of differing thermal conductivities on the DTG curve. Dotted line, helium purge gas; solid line, nitrogen purge gas. (Reproduced from Buckman and Hawkins, *Chem. in Austral.*[24] with permission.)

3.6 Optimizing Thermogravimetry Results

It should be evident from the foregoing comments and examples that the conditions required to minimize the effects of heat and mass transfer as well as decreasing the possibility of change of reaction sequence are to have:

- a small sample (10–20 mg)
- finely ground material (<100 μm)
- a slow heating rate (10–20 °C min^{-1})
- a thin film spread evenly in a flat open pan made of an inert material.

Such conditions will allow maximum resolution of close reactions and give well-defined mass losses over relatively narrow temperature ranges. However, it should be emphasized that when dealing with a sample of unknown properties that it is advisable to use a range of different conditions to ensure that the sample behaves in a similar manner. Thus if the only effect noticed on increasing the heating rate is a modest shift to higher temperatures of the mass losses, then it can be assumed that the reaction sequence is the same and the increased temperature is just the result of an increase in the thermal lag of the sample. Any marked difference in the temperature at which mass losses occur, or the number of mass losses, may indicate a change in reaction sequence.

In some circumstances it is not possible to use the above conditions, for example when measuring small mass changes which may require large samples to give accurate mass values, or when doing combined TG–evolved gas analysis experiments where the evolved gas concentration needs to be increased in order to be detected.

4 REPORTING THERMAL ANALYSIS RESULTS

It has been stressed that TG results are very dependent on the sample properties and the experimental conditions used. In order to facilitate a comparison of results obtained by different workers, it follows that these various parameters, such as particle size, heating rate and so on, should be reported.

The existing recommendations for reporting thermal methods of analysis were developed by the Standardization Committee of ICTAC in the late 1960s and early 1970s,[26–28] and are also in various standards such as American Standard Test Methods, ASTM E472-86. The recommendations are presently in revision to take into account current practice.[29] The draft document is reproduced below, although it has not at the time of writing been officially ratified by the Council of the ICTAC. There has been some change of emphasis in the new draft and more emphasis has been placed on accurate characterization of the sample, as well as in giving details on the way in which the computer acquires and processes data. The section relating to the experimental conditions is largely unchanged. This paper does not address safety problems associated with the use of the related materials or equipment. It is the responsibility of the user to establish the appropriate safety standards and work within the validity of the given safety laws.

Accompanying each thermal analysis record should be information about the properties of substances, experimental conditions and data acquisition and manipulation methods.

4.1 Properties of Substances

1. All substances (sample, reference, diluent) should be identified by a definitive name, an empirical formula, or equivalent compositional information.
2. A statement of the source of all substances, details of their histories, pretreatments, physical properties, and chemical purities, as far as they are known should be given.
   (a) ''Histories'' includes method of acquisition or manufacture of the sample (how it was isolated or made, manufacturing conditions, e.g. grinding, sizing, etc.), previous thermal and mechanical treatments (e.g. repeated grinding, sintering, all experimental conditions), surface modification, and any other physical or chemical treatment.
   (b) ''Pretreatments'' includes preparation of the sample prior to the thermal analysis experiment.
   (c) ''Physical properties'' includes particle size, surface area, porosity.
   (d) ''Chemical purities'' includes elemental analysis, phase composition, and chemical homogeneity of sample.

4.2 Experimental Conditions

1. A clear statement should be given of the temperature environment of the sample during measurement or reaction, including initial temperature, final temperature, rate of temperature change if linear, or details if not linear.
2. The sample atmosphere should be identified by pressure, composition, and purity: whether the atmosphere is static, self generated or dynamic through or over the sample. Where applicable, the ambient atmospheric pressure and humidity should be specified. If the pressure is other than atmospheric, full details of the method of control should be given.
3. A statement of the dimensions, geometry and materials of the sample holder, and the method of loading the sample where applicable should be given.
4. The abscissa scale should be identified in terms of time or temperature at a specified location. Time or temperature should be plotted to increase from right to left.
5. A statement of the methods used to identify intermediate or final products should be provided.
6. Faithful reproductions of all original records should be provided.
7. Wherever possible, each thermal event should be identified and supplementary supporting evidence stated.
8. Sample mass and dilution of sample should be stated.
9. The apparatus should be identified by the provision of the manufacturer's name and model number. More detail should be provided if significant modifications to the standard commercial model have been made, or if the apparatus is new or novel, or if the apparatus is not commercially available.

4.3 Data Acquisition and Manipulation Methods

1. The instrument manufacturers software version should be identified, or details of any self-developed versions given.
2. Equations used to process data, or reference to suitable literature should be given.
3. The frequency of sampling, filtering and averaging of the signal should be given.

4. An indication should be provided of the smoothing and signal conditioning used to convert analog to digital signals, or reference to suitable literature. Similar information for subsequent processing of the digital data is also needed.

The following information is required specifically for TG.

- The ordinate scale should indicate the mass scale. Mass loss should be plotted as a downward trend and deviations from this practice should be clearly marked. Additional scales (for example, fractional decomposition, molecular composition) may be used for the ordinate where desired.
- The method used for the calibration of the temperature and mass should be indicated.
- If DTG is employed, the method of obtaining the derivative should be indicated and the units of the ordinate specified.

5 APPLICATIONS

TG can be used in any application in which there is a change in mass. Table 2 gives the range of physical and chemical reactions that have been studied. The information gained from TG experiments can be considerably enhanced by the characterization of intermediate products, so that the proposed reaction scheme is a confirmed one rather than speculation. If the TG experiment is stopped after a mass change and the sample cooled down to ambient temperature in an inert atmosphere, the product can be isolated and characterized. An X-ray diffraction (XRD) pattern, for example, may permit the determination of phases present. Fourier transform infrared spectroscopy (FTIR), electron probe microanalysis (EPMA), and SEM are other techniques that can assist in the elucidation of a reaction scheme. What is more, these methods can be carried out on a few milligrams of sample and can be used to supply both qualitative and quantitative information. Examples of the application of these techniques to TG studies will be given in the following sections.

5.1 Adsorption and Desorption Reactions

Any solid material which adsorbs gases, such as catalytic materials[30] and air cleaners, can have their sorptive and desorptive properties studied by TG. The gain in mass at any given temperature is related to the number of occupied sites. Usually a steady mass gain is observed up to a certain temperature and then a limiting diffusion rate gives rise to a plateau in the TG curve.

Table 2 Processes which can be studied by TG[a]

Process	Mass gain	Mass loss
Adsorption and absorption	*	
Desorption		*
Dehydration or desolvation		*
Vaporization		*
Sublimation		*
Decomposition		*
Oxidation	*	*
Reduction		*
Solid–gas reactions	*	*
Solid–solid reactions		*

[a] Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11]

During a heterogeneous catalytic reaction, gaseous molecules are adsorbed onto the surface of the active catalytic sites. Since the activity of the catalyst is dependent on the number of available adsorption sites, which in turn is related to the surface area of the catalyst material, it is of some importance to be able to measure the latter. TG has the major advantage of continuously monitoring the weight gain which results from gas adsorption and this gives information on the rate of gas uptake as well as the attainment of equilibrium conditions. Providing there is monolayer coverage and the number of molecules of gas adsorbed per unit of catalyst is known, from the gain in mass at a specific temperature it is possible to calculate the specific surface area of the catalyst material.

The total area of the catalyst material, i.e. catalyst plus support medium, is usually measured by the physical adsorption of an inert gas typically at subambient temperatures using a vacuum microbalance. This topic has been discussed in more detail by Keattch and Dollimore.[6] Measurement of the surface area of the active catalyst alone requires gases which are chemisorbed by the catalyst and hydrogen and carbon monoxide are frequently used in this role. Some knowledge of the stoichiometry of the surface reaction is required.

It is sometimes useful to monitor the adsorption behavior of the catalyst in the temperature range at which reaction occurs. To do this, the reaction gas mixture can be passed over the catalyst at specific temperatures and the weight changes followed until equilibrium is achieved. A rapid weight loss can result if the surface area of the metal are reduced by sintering, or if the decomposition temperatures of catalyst materials are exceeded.

Desorption isotherms can be monitored by heating up the solid with adsorbed material and following the mass loss. The vacuum desorption behavior of methanol adsorbed onto a zeolite catalyst between 47–227 °C[31] is

a typical example. Approximately 40% of the methanol remained adsorbed, which was explained by pore filling. TG has been used to investigate the regeneration of materials which are used as adsorbing agents, which may involve adsorption from the gas or liquid phases.[32–34] In the gold processing industry, for example, activated carbon is used to adsorb a gold cyanide complex. Unfortunately, the carbon is not very specific in its adsorption characteristics and becomes considerably fouled by the adsorption of inorganic and organic species, which reduces its gold adsorption activity. The fouled carbon is acid washed to remove inorganic materials and then thermally regenerated by heating to 650 °C for up to 30 min in steam. During this process, any organic compounds are likely to be decomposed. TG, especially when coupled with evolved gas analysis equipment such as FTIR and mass spectrometry (MS), provides a powerful tool for assessing the temperatures of decomposition of the adsorbed organic compounds, their identification, and any possible hazards associated with the evolved gases.

One reagent which is known to adsorb strongly onto activated carbon is sodium ethyl xanthate (SEX), which is used as a flotation reagent in the gold processing circuit. The thermal decomposition of SEX in a nitrogen atmosphere was studied by coupled thermogravimetry/Fourier transform infrared spectroscopy (TG/FTIR), and by pyrolysis/gas chromatography/mass spectrometry (py/GC/MS). The TG curve exhibited two discrete mass losses of 41% and 19%, respectively, in the temperature ranges 190–220 °C and 220–320 °C (Figure 9a).[33,34] The major evolved gases resulting from the first mass loss were carbonyl sulfide, ethanol, ethanethiol, carbon disulfide, diethyl sulfide, diethyl carbonate, and diethyl disulfide. The gases identified as a result of the second mass loss were carbonyl sulfide, ethanethiol, and carbon disulfide. Hydrogen sulfide was detected in both mass losses by py/GC/MS, but not detected by FTIR. The solid residue was sodium hydrogen sulfide.

SEX was adsorbed onto activated carbon and heated in nitrogen. Two discrete mass losses were still observed, but in the temperature ranges 100–230 °C and 200–340 °C (Figure 9b). Carbonyl sulfide and carbon disulfide were now the dominant gases evolved in each of the mass losses and the other gaseous products were relatively minor. Hence adsorbing the SEX onto carbon changed the temperature of decomposition as well as affecting the reaction sequence.[33,34] Samples of activated carbon were obtained from operating gold plants and heated in the TG/FTIR system. Carbonyl sulfide and carbonyl disulfide were detected in some of the samples, indicating the presence of adsorbed SEX.

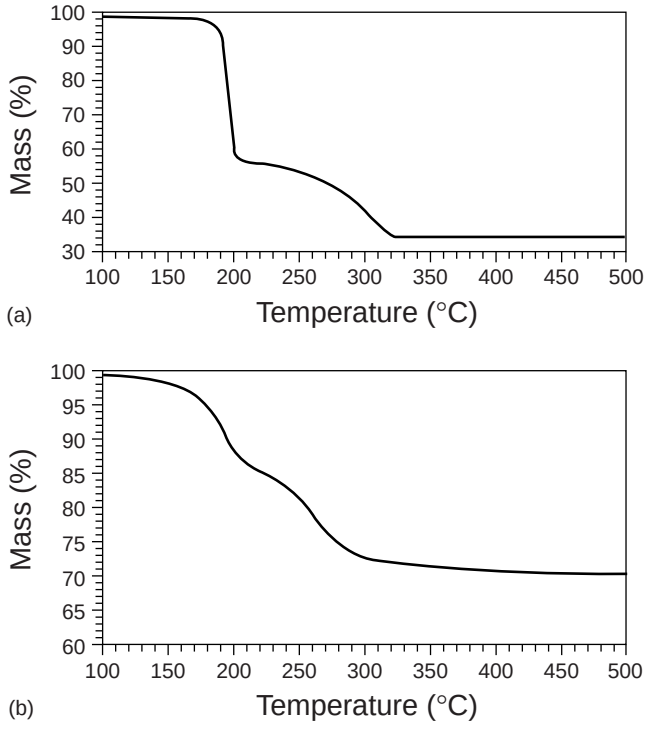

Figure 9 TG curves of (a) SEX alone and (b) when adsorbed onto activated carbon.[34]

5.2 Sublimation and Vaporization

TG can be used to provide a quantitative measure of the rate of sublimation or vaporization under a wide range of experimental conditions and can be used to obtain heats of sublimation or volatilization for materials which do not decompose in the temperature range being investigated. The vaporization characteristics of fuels, crude oil fractions, greases, aliphatic and aromatic organic compounds,[35] essential oils associated with perfumes,[36] and organic pollutants have been reported.[37]

TG can be used for the determination of vapor pressures by the Knudsen effusion method.[38] The technique involves measurement of the rate of loss of molecules of the evaporating material in an effusion cell through a calibrated orifice under reduced pressure. The equilibrium vapor pressures can be calculated using the ideal Knudsen equation, Equation (1):

$$p = \frac{\Delta m}{\Delta t q} \frac{(2\pi RT)^{1/2}}{M^{1/2}} \tag{1}$$

where p = vapor pressure in dyne cm^2, q is the orifice area in cm^2, $\Delta m/\Delta t$ the rate of weight loss in g s^{-1}, T is the temperature in K, and M is the molecular weight of the specimen. If the data fits the Clausius–Clapeyron relationship then the heat of sublimation (or if above the melting point, the heat of vaporization) can be determined.

Another method[39] has used the Langmuir method based on the rate of escape of molecules from a surface under vacuum conditions to determine heats of sublimation. According to the Langmuir method, the rate of sublimation per unit area of substance is related to the vapor pressure by Equation (2):

$$\frac{dm}{dt} = \frac{M^{1/2}p}{[2\pi RT]^{1/2}} \quad (2)$$

By choosing temperatures where the rate of weight loss was low and the overall loss was kept under 2%, good straight line Clausius–Clapeyron plots were given with slopes corresponding to the heats of sublimation reproducible to approximately 5%. Reasonable agreement was obtained with literature values for a number of the compounds studied. In addition, the value obtained for benzoic acid agreed with that obtained by Wiedemann[38] by the Knudsen method, in a similar temperature range, to better than 3%.

TG can be a useful technique to check on spurious results in DTA or differential scanning calorimetry (DSC). The measurement of the enthalpy of melting of volatile materials requires the sample to be hermetically sealed into a sample pan especially if repeat determinations are to be made. TG provides the ability to check if significant volatilization occurs in the vicinity of the melting point using an open sample crucible and to investigate the intergrity of the seal in an enclosed pan.

5.3 Thermal Decomposition and Thermal Stability

TG can be used to define various properties of a thermal decomposition reaction, as illustrated in Figure 1, and includes the extrapolated onset and offset temperatures. There is nothing absolute about these values, however, and they are all dependent on the experimental conditions used. The general application of thermal decomposition reactions is to determine the products of the reaction and so be able to write a reaction scheme. This task may be aided by an evolved gas analysis. Many compounds undergo a series of sequential decomposition reactions with increasing temperature, so that the appearance of a plateau after the first mass loss does not mean that the product at that temperature will not undergo further reaction at higher temperatures. The thermal decomposition of a wide range of materials has been studied and includes inorganic compounds, coordination compounds, polymers, natural products such as woods and wools, superconductors, cements and minerals.

An example of a thermal decomposition reaction for a coordination compound illustrates some of the points made above. $[Ni(\beta\text{-picoline})_4(SCN)_2]$[40,41] is a coordination compound which decomposes by loss of the relatively weakly bound picoline ligand. As the conditions are varied, the number of ligands lost vary, as does the temperature at which the decomposition occurs (see Table 3). In nitrogen at a moderate heating rate two ligands are lost commencing at 125 °C, followed by two further losses at 190 °C and 200 °C, respectively. If the ligand atmosphere is inhibited from diffusion away from the sample by putting a lid on the sample crucible, the ligands are lost in a 1:2:1 ratio and the decomposition temperatures shift to higher values. In a vacuum all four ligands are lost in a single step and the decomposition temperature becomes 30 °C.

TG is a technique which is particularly applicable as a rapid means of comparing the relative thermal stability of a range of materials. After the loss of volatiles such as solvents and other small molecules, the first significant mass loss is taken as the end of the thermal stability region. Heating the materials under the same experimental conditions will allow the decomposition temperature (T_e) or the temperature at which a chosen fraction has reacted (T_α) to be determined and used to place the materials in order of rank.

However, it must be emphasized that the decomposition temperature (or T_α) does not represent the upper limit of usefulness, as considerable decomposition can occur at lower temperatures albeit at a slower rate. Isothermal TG tests, although requiring more time, give more useful information on thermal stability. The material is heated rapidly to a preselected temperature and the mass loss followed as a function of time. Several temperatures are examined in separate experiments, thus generating a family of curves. A typical set of curves is shown in Figure 10. The material could be used quite successfully below 300 °C, but at 400 °C massive degradation had

Table 3 Effects of experimental conditions on the decomposition of $[Ni(\beta\text{-picoline})_4(SCN)_2]$[a]

Reference	Sample mass (mg)	Heating rate (°C min^{-1})	Atmosphere	Stoichiometry of reactions	Decomposition temperature (°C)
40	100	3	nitrogen	−2, −1, −1	125
41	500	6	ligand	−1, −2, −1	185
41	10.6	2.5	vacuum	−4	30

[a] Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & Sons.[11]

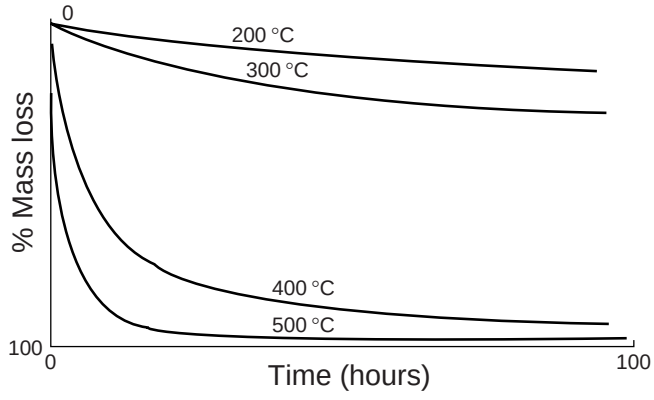

Figure 10 Isothermal TG experiments on a material which determines the effect of time at specific temperatures on thermal stability.

occurred and the integrity of the material compromised. This method is usually more successful for distinguishing between similar materials than the dynamic TG method.

An extension of the isothermal method is a test devised to measure the oxidative stability of materials, which involves heating the material in an inert atmosphere to a given temperature, establishing a stable baseline, and changing the atmosphere to oxygen or an oxidizing atmosphere. The time taken to produce the first major mass loss, called the oxidative induction period, is used as a measure of the oxidative stability. Polymers[42] and oils[43,44] have been investigated by this method. The effect of various additives, such as oxidative stabilizers and flame retardants, can be rapidly assessed.

The technique of TG has advantages over other dynamic techniques in evaluating thermal stability in that quantitative results are given directly and instrumental factors such as baseline drift are minimized. The quantitative nature of TG can be further exploited to enable stoichiometric equations for the decomposition mechanism to be postulated. The mechanism may then be confirmed by analysis of the solid and gaseous decomposition products.

5.4 Gas–Solid Reactions

One of the major uses of TG has been to study the reaction between a sample and a reactive gas. The most widely used atmospheres have been those of air and oxygen to study oxidation reactions, but if gas flows are arranged so that the balance mechanism is protected (see section 2.2) then corrosive gases e.g. chlorine,[14] or reductive gases, e.g. hydrogen[45] can be used. Vapors can also be introduced into the sample chamber, both under high or low pressures. In some cases, TG experiments have been carried out as the precursor to pilot-scale studies. The advantages of using TG in this role is that many experiments can be carried out rapidly and cheaply, experimental parameters can be carefully controlled, and the major factors affecting the reaction be identified. This then enables the pilot-scale study, which is a much more expensive operation to run, to focus on the important parameters and spend less time on the less significant ones. The oxidation of metal sulfides provides some typical examples of a range of reactions which cause mass and energy changes, and which can therefore be studied by thermal methods of analysis.

5.4.1 Direct Formation of Oxide

The formation of oxide can be expressed in general as Equation (3):

$$2MS_{(s)} + 3O_{2(g)} \longrightarrow 2MO_{(s)} + 2SO_{2(g)} \qquad (3)$$

The conversion of sulfide to oxide produces a mass loss coincident with an exothermic event.

5.4.2 Formation of Sulfate

The formation of sulfate usually takes place in the same temperature range as the formation of oxide and suggests that the sulfate is formed from the reaction scheme depicted by Equations (4–7):

$$MS_{(s)} + 1.5O_{2(g)} \longrightarrow MO_{(s)} + SO_{2(g)} \qquad (4)$$

$$SO_{2(g)} + 0.5O_{2(g)} \longrightarrow SO_{3(g)} \qquad (5)$$

$$MO_{(s)} + SO_{3(g)} \longrightarrow MSO_{4(s)} \qquad (6)$$

$$MS_{(s)} + 4SO_{3(g)} \longrightarrow MSO_{4(s)} + 4SO_{2(g)} \qquad (7)$$

The observed mass change is dependent on the relative rates of formation of the sulfate and oxide. Usually the first sign of reaction is a mass gain, as observed for FeS[46] and NiS.[47]

The formation of sulfates can be affected by several factors:

- Sulfate formation will tend to increase as particle size decreases.[48]
- Heating rate: sulfate formation is encouraged by slow heating rates.[49]
- Sample holder geometry: sulfate formation is enhanced in a shallow open sample holder, which allows good interchange between the incoming oxidizing gas and the outgoing sulfur dioxide product.[21]
- Sample holder material: the formation of SO_3 is encouraged by the presence of catalytic material and hence the extent of sulfation increases. The catalytic influence of a platinum sample holder on sulfate formation is well known.

- Mass size: increasing the mass size inhibits gas diffusion, and so inhibits formation of sulfate.[50]

The formation of sulfate and oxide layers can inhibit oxygen diffusion into the unreacted core of the particle. During the oxidation of chalcocite (Cu_2S), the first major mass gain is due to the simultaneous oxidation of the sulfide and formation of sulfate, but then the reaction starts to slow down and the TG curve starts to flatten off.[51] At 510 °C a second major mass gain commences. It could be interpreted that these were two separate reactions. Investigation using techniques such as SEM and hot stage microscopy revealed that a thick oxide/sulfate layer had formed around the particle and arrested further oxidation. At 510 °C, however, melting occurred through reaction between Cu_2S and $CuSO_4$, the protective film was destroyed and so further oxidation and sulfation could occur.

The determination of the above mechanisms could not have been made without the use of complementary techniques XRD, SEM, EPMA, chemical analysis, and FTIR spectroscopy to characterize the intermediate products. If the technique is quantitative then the appearance and disappearance of a particular phase can be followed as a function of temperature. Quantitative XRD can be carried out with small sample masses in the milligram range, which makes it applicable to thermal analysis studies. By using ratios of several interference-free peaks in the XRD record of the partially reacted mineral against some internal standard, it is possible to produce several results from one experiment which can be averaged to produce a reasonably reliable value. The XRD analysis of the Cu_2O content of partially oxidized chalcocite as a function of temperature is given in Table 4.[52] It is evident that wide variation can occur between individual values, but the average value has an acceptable standard deviation.

The combination of Fourier transform methods and infrared spectroscopy has caused a major resurgence of interest in this technique. The quality of the spectrum has been vastly improved even for very small samples in the milligram range, and so again this is a useful method for thermal analysis experiments. One significant advantage is that compounds with the cation in different oxidation states can often be differentiated. This technique has been used to follow the formation of iron(II) and iron(III) sulfates during the oxidation of pyrite.[53] Although sulfate is known to be formed in significant quantities, it is often difficult to detect even by XRD perhaps because it is amorphous. The FTIR technique is very sensitive to the detection of sulfate vibrations, and so the formation and decomposition of the iron(II) and iron(III) sulfates in quenched samples of partially oxidized pyrite can be followed as a function of temperature. It was found that $FeSO_4$ was the major species formed and that very little conversion of iron(II) to iron(III) took place under the experimental conditions used.

The technique was similarly used to determine $CuO \cdot CuSO_4$, which is one of the products formed during the oxidation of chalcocite. A peak at 439 cm^{-1} exhibited no interference by any of the other compounds present in the partially oxidized sample, and the measurement of its peak height for a set of standard samples was used to construct a calibration curve.[52] Excellent quantitative results were achieved and the method had better detection limits than XRD. The results from the compositional analysis by XRD (Table 4) and the FTIR are given for partially oxidized samples of chalcocite against temperature in Table 5. Summation of the data for those samples in which all the phases were determined by these methods, that is at 570, 670 and 775 °C, gave values that were better than ±10% relative error. Hence these analytical methods provide a reasonable estimation of the phase composition of the sample at any particular temperature.[52]

5.4.3 Decomposition of Sulfates

The decomposition of sulfates takes place in two stages as expressed by the Equations (8) and (9):

$$2MSO_{4(s)} \longrightarrow MO \cdot MSO_{4(s)} + SO_{3(g)} \quad (8)$$

$$MO \cdot MSO_{4(s)} \longrightarrow 2MO_{(s)} + SO_{3(g)} \quad (9)$$

These reactions are detectable by mass losses in the TG curve.

Table 4 Determination of Cu_2O by quantitative XRD in partially oxidized samples of chalcocite[52]

Temperature (°C)	Cu_2O content (% by weight)	Mean	Std. deviation
435	0.7, 0.9, 0.7, 0.8, 1.5, 2.0, 1.5, 1.7	1.2	0.5
470	2.0, 2.5, 1.9, 3.5, 4.4, 3.3	2.9	0.9
570	42.5, 40.5, 52.6, 50.7, 48.3, 45.9	46.8	4.7
670	38.1, 37.6, 40.6, 34.3, 33.3, 36.4, 40.6, 39.5, 43.1	38.2	3.2

Table 5 Phase composition of partially oxidized samples of chalcocite.[52] All values are in weight % in the products

Temperature (°C)	Cu_2S	Cu_2O	$CuSO_4$	$CuO \cdot CuSO_4$	CuO	Total
435	97.1[a]	1.2 ± 0.3	1.7 ± 0.4			100
470	83.5[a]	3.7 ± 0.6	12.8 ± 0.9			100
570		46.8 ± 5.3	55.6 ± 2.8	1.9 ± 0.1		104 ± 6.0
670		38.2 ± 3.8	9.8 ± 1.0	41.2 ± 1.4	11.9 ± 2.8	101 ± 5.0
775				23.8 ± 0.8	71.9 ± 9.4	95.7 ± 9.4
840					100	100

[a] Calculated by difference.

5.5 Solid–Solid Reactions

A thermobalance can be used to follow reactions of the type represented by Equation (10):

$$A(s) + B(s) \longrightarrow C(s) + D(g) \qquad (10)$$

as for example in the curing of rubbers by elimination or condensation reactions. The change in mass can be used to follow the rate of cure and in fact this method has been adopted by industry for routine quality control purposes.[54]

5.6 Analytical Applications

One of the major disadvantages of methods of mass loss analysis, such as loss on ignition measurements, is that the result is given as an undifferentiated mass loss from ambient temperature to 1000 °C. In contrast, TG methods permit the quantitative measurement of mass change as a continuous function of temperature. Hence in addition to being able to determine the purity of a high purity compound by measurement of the single mass loss associated with decomposition, in some circumstances several phases can be determined in a complex mixture. The required criteria are that each reaction is not interfered with by another reaction and that the stoichiometry of the product is known. Even when the reactions overlap to some extent, variation in the experimental conditions chosen can sometimes resolve the reactions sufficiently to permit individual quantification. Many examples of quantitative analysis by TG have been compiled by Earnest,[7] in areas such as drugs, polymers, inorganic compounds, minerals, fuels, and waste products.

One of the most common measurements on a sample is for water, both adsorbed and water of crystallization, as well as other solvating agents. Usually these can be differentiated by control of the experimental variables, such as an isothermal hold so that the adsorbed water is completely removed before returning to the dynamic heating rate with subsequent determination of water of crystallization. For some salts, the water of hydration occurs in discrete steps, for example with $CuSO_4 \cdot 5H_2O$, which loses water in the ratio 2:2:1. Mixtures can be analyzed easily if the decomposition reactions are resolved. Plaster consists of a mixture of gypsum ($CaSO_4 \cdot 2H_2O$) and lime ($Ca(OH)_2$). Gypsum looses its water of crystallization between 100–150 °C, and the lime dehydroxylates between 400–500 °C, as demonstrated by Equations (11) and (12):

$$CaSO_4 \cdot 2H_2O \longrightarrow CaSO_4 + 2H_2O \qquad (11)$$

$$Ca(OH)_2 \longrightarrow CaO + H_2O \qquad (12)$$

Hence in a single experiment at a heating rate of 20 °C min^{-1}, which would take 25 min, both phases can be determined with a greater accuracy than by any other method available.

Although the dehydration of gypsum appears as a single reaction in the above equation, it is a two step process as indicated by Equation (13):

$$\begin{aligned} CaSO_4 \cdot 2H_2O &\longrightarrow CaSO_4 \cdot 0.5H_2O + 1.5H_2O \\ &\longrightarrow CaSO_4 + 0.5H_2O \end{aligned} \qquad (13)$$

These two reactions can be resolved by alteration of the partial pressure of the water vapor above the sample. If the gypsum sample is sealed in a sample pan with a small leak, the water vapor trapped in the pan from the first reaction delays the commencement of the second step of the reaction by some 40–50 °C.[55] This method has practical application in the cement industry. Gypsum is added to cement clinker in order to slow down the rate of setting. The two components are ground together, but heat generated in the process may cause decomposition of gypsum to the hemihydrate, which is undesirable. Hence the TG method can be used rapidly to analyze the gypsum/clinker mix for gypsum and hemihydrate, a task which again cannot be performed as quickly or accurately by any other method.

If two overlapping decomposition reactions are taking place and one of the reactions is reversible to its product gas, then introduction of that gas into the system will delay that decomposition and it will take

place at a higher temperature. The decomposition of $CaCO_3$ is a reversible reaction, so the presence of CO_2 in the atmosphere will shift the decomposition temperature to higher values. It is possible, if the various thermodynamic constants are known, to calculate the effect of change in partial pressure of the evolved gas by use of Equation (14):

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = RT \ln P \tag{14}$$

where P is the partial pressure for this case of CO_2. Increasing the partial pressure from 0.1 to 1 atmosphere (1.01×10^4 to 1.01×10^5 Pa) of CO_2 produces a 138 °C shift in the decomposition temperature. Any other phase present in the system which was not reversible to CO_2 would be unaffected by the change in atmosphere.

One important application of TG/DTG has been in the study of high alumina cement (HAC). HAC is based on $CaAl_2O_4$, which can undergo a series of conversion reactions producing a change in volume which gives increased porosity with a concomitant loss in strength. It has been suspected that loss of strength was the cause of accidents which occurred in buildings constructed with HAC concrete beams. In order to assess the degree of conversion, and therefore to assess the degree of risk for the building, methods had to be developed to determine the phases present. DTA but also DTG was used and found to be acceptable.[56] A typical DTG curve is shown in Figure 11.

The amount of $Al(OH)_3$ (written as AH_3) formed, which is one of the conversion products from CAH_{10}, is given by the peak height at just below 300 °C. The amount of unconverted CAH_{10} is given by the peak height at 100 °C. Therefore the degree of conversion (DC) is given by a simple ratio, Equation (15).

$$DC = \frac{[AH_3]}{[AH_3] + [CAH_{10}]} \tag{15}$$

$$\%DC = \frac{100[113]}{[113 + 15]} = 88\%$$

The degree of conversion is, therefore, an empirical scale which can be used to estimate the likelihood of further conversion which might weaken the concrete or make it more permeable and, therefore, more vulnerable to chemical attack.

Changing the atmosphere from an inert one to a reactive one is a useful process. In an inert atmosphere many organic compounds can be decomposed and volatilized, and if the atmosphere is changed to air then some of the remaining material is oxidized with the formation of more volatile compounds. The residue is the ash. This analytical process has been used to carry out quantitative analysis on a variety of organic materials ranging from rubbers and plastics to coals. In the last case, the proximate analysis of a coal sample can be carried out by first heating in an inert atmosphere to 110 °C and holding until a constant mass is achieved.[57] This mass loss enables the moisture content of the coal to be determined. The sample is heated at constant rate to 900 °C and held until constant mass is again achieved. This second mass loss enables the volatile matter to be determined. The furnace temperature is then lowered to 815 °C and the atmosphere changed to air, which burns off the fixed carbon and leaves the ash as a residue (Figure 12). Fourteen coals of widely differing composition were analyzed by the British Standard Specification 1016, Part 3 (1973), and by the TG method. In general the analysis values were similar. The TG analysis could be carried out in about 35–40 min per sample.

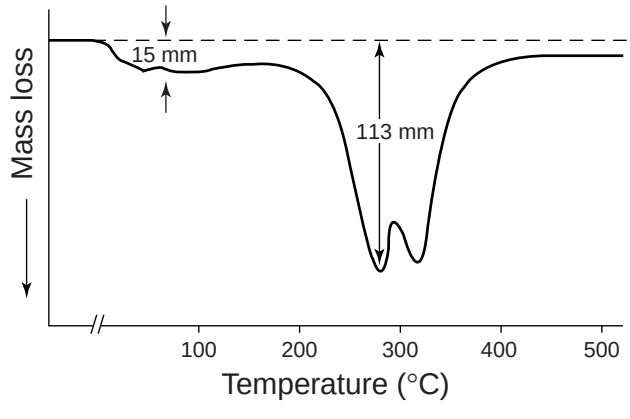

Figure 11 DTG curve for HAC. (Reproduced from Dunn and Sharp, *Treatise on Analytical Chemistry*, ed. J.D. Winefordner, Copyright 1993, by permission of John Wiley & sons.[11])

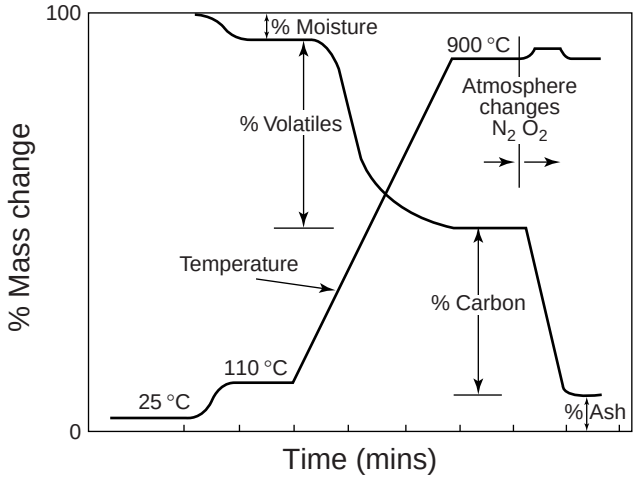

Figure 12 Proximate analysis of coal by TG.

An additional magnetic step has been included in the proximate analysis method which permits the determination of pyrite.[58] After the proximate analysis has been carried out, a magnetic field is switched on and any apparent increases in weight attributed to paramagnetic species such as Fe_2O_3. The ferric oxide is reduced to iron by the introduction of hydrogen at 400 °C. This produces a large apparent weight gain caused by the ferromagnetic effect. After cooling to room temperature the field is switched off and the apparent weight loss used to calculate the pyrite content. Good agreement was found between the TG magnetic method and the ASTM.

The analysis of rubber compounds can be carried out by a similar TG method.[59,60] Rubbers typically contain polymers, extending oils and other organic compounds, carbon black and inorganic compounds. Heating the rubber in nitrogen volatilizes or decomposes the organic compounds, and decomposes the polymer, so that the first mass loss can be related to the total content of these compounds. Changing the atmosphere to air oxidizes the carbon black to give a second mass loss, and the residue is the inorganic components converted to their oxides. This analysis can be done in less than 20 min, and provides a rapid routine quality control method in the elastomer industry. The method can be largely automated with the computer system controlling the gas changes, heating rates and so on. The decomposition of the pure polymer should be checked to ensure that it does decompose completely and does not form a char. In some circumstances, TG can be used to quantify the components of polymer blends and block copolymers.[61]

Sometimes, total phase analysis can be achieved by TG methods in conjunction with other analytical methods. Thus an iron ore containing a clay, goethite (FeOOH) and hematite (Fe_2O_3) was analyzed for the first two components by TG, and the total iron determined by dissolution and titrimetric analysis.[62] In order to make good standards, the clay and goethite were isolated from the original deposit. Making standards from the materials obtained from the usual chemical suppliers did not give such accurate results, as the thermal decomposition characteristics of the materials were different to those of the minerals.

5.7 Kinetic Measurements

One of the major applications of TG has been for the determination of the kinetic properties of a solid compound as it decomposes or reacts. The use of TG for this purpose, however, has been the cause of much controversy, mainly through its uncritical usage resulting in considerable variation in the kinetic parameters published for a given reaction. The experimentalist would be well advised to read some of the critical reviews of the use of TG for kinetic studies, for example Brown[63] and Flynn,[64] before commencing work.

Two equations are used to describe the kinetics of a reaction. The first one is the Arrhenius equation, which gives the relationship between the variation of the rate constant k with temperature (Equation 16):

$$k = A\mathrm{e}^{-E/RT} \tag{16}$$

where T is the absolute temperature, A is the pre-exponential factor, E is the activation energy and R is the gas constant.

The second one is the rate equation, which describes the relationship between the fraction of material reacted, the rate of the reaction and time (Equation 17):

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t} = kf(\alpha)\mathrm{e}^{-E/RT} \tag{17}$$

where α is the fraction reacted in time t, and $f(\alpha)$ is some function of α which is related to the kinetic or reaction mechanism, for example diffusion control or shrinking sphere model. The term ''reaction mechanism'' does not always have this definition, however, and may also mean the determination of the chemical steps by which a reactant proceeds to product. The various forms adopted by the function $f(\alpha)$ have been discussed elsewhere,[6,65–68] and are given in Table 6. The $f(\alpha)$ function defines the shape of the TG/DTG curve. Changes in the properties of the sample and the experimental variables will cause changes in the rate of reaction and shifts in the TG/DTG curves, but unless the reaction mechanism changes, the shape of the curves should remain the same.

Combination of these equations leads to an Equation (18) of the form:

$$\frac{\alpha}{\mathrm{d}t} = Af(\alpha)\mathrm{e}^{-E/RT} \tag{18}$$

At a constant heating rate, $T = T_0 + \beta t$, where T_0 is the initial temperature and β is the heating rate. Substitution for dt gives Equation (19):

$$\frac{\mathrm{d}\alpha}{\mathrm{d}T} = \frac{A}{\beta}f(\alpha)\mathrm{e}^{-E/RT} \tag{19}$$

Rearrangement gives Equation (20):

$$\frac{\mathrm{d}\alpha}{f(\alpha)} = \frac{A}{\beta}\mathrm{e}^{-E/RT}\,\mathrm{d}T \tag{20}$$

which is the basic equation of the DTG curve, which when integrated is the equation of the TG curve. If the form of the function $f(\alpha)$ is known, integration of the left hand side of the equation is straightforward, and gives the associated function, $g(\alpha)$. If E is constant, the right

Table 6 Broad classification of solid-state rate expressions

Classification		$g(\alpha) = kt$	$f(\alpha) = 1/k(\mathrm{d}\alpha/\mathrm{d}t)$
Acceleratory α–time curves			
P1	Power law	$\alpha^{1/n}$	$N(\alpha)^{(n-1)/n}$
E1	Exponential law	$\ln\alpha$	α
Sigmoidal α–time curves			
A2	Avrami–Erofeev	$[-\ln(1-\alpha)]^{1/2}$	$2(1-\alpha)(-\ln(1-\alpha))^{1/2}$
A3	Avrami–Erofeev	$[-\ln(1-\alpha)]^{1/3}$	$3(1-\alpha)(-\ln(1-\alpha))^{2/3}$
A4	Avrami–Erofeev	$[-\ln(1-\alpha)]^{1/4}$	$4(1-\alpha)(-\ln(1-\alpha))^{3/4}$
B1	Prout–Tompkins	$\ln[\alpha/(1-\alpha)] + c$	$\alpha(1-\alpha)$
Deceleratory α–time curves			
1. Based on geometrical models			
R2	Contracting area	$1-(1-\alpha)^{1/2}$	$2(1-\alpha)^{1/2}$
R3	Contracting volume	$1-(1-\alpha)^{1/3}$	$3(1-\alpha)^{2/3}$
2. Based on diffusion mechanisms			
D1	One-dimensional	α^2	$1/2\alpha$
D2	Two-dimensional	$(1-\alpha)\ln(1-\alpha)+\alpha$	$(-\ln(1-\alpha))^{-1}$
D3	Three-dimensional	$[1-(1-\alpha)^{1/3}]^2$	$3/2(1-\alpha)^{2/3}(1-(1-\alpha)^{1/3})^{-1}$
D4	Ginstling–Brounshtein	$(1-2\alpha/3)-(1-\alpha)^{2/3}$	$3/2((1-\alpha)^{-1/3}-1)^{-1}$
3. Based on "order" of reaction			
F1	First order	$-\ln(1-\alpha)$	$1-\alpha$
F2	Second order	$1/(1-\alpha)$	$(1-\alpha)^2$
F3	Third order	$[1/(1-\alpha)]^2$	$0.5(1-\alpha)^3$

hand side of the equation can be integrated leading to Equation (21):

$$g(\alpha) = \frac{A}{\beta} \quad \mathrm{e}^{-ER/T}\,\mathrm{d}T \tag{21}$$

The values of E and A affect the temperature range over which the TG curve is observed.[65]

One of the difficulties of dealing with kinetic work is that there are many choices available, for both the experimental approach and for ways of processing the data. The two main experimental approaches are:

- The isothermal approach involves rapidly heating the sample to some specific temperature and monitoring the mass change. The experiment is repeated with a new sample at another temperature, so generating a family of curves. The fraction reacted, α, (the ratio of the mass loss at time t and total mass loss for the reaction) is found for each temperature.
- The rising temperature or nonisothermal method[66,68] requires that the sample is heated at a constant rate over the reaction temperature range. The value of α is given by the ratio of the mass loss up to temperature T to the total mass loss of the reaction. Hence again α can be determined at several temperatures.

There are two main ways in which the data can be analyzed, the integral and differential methods. At first glance the former, as expressed by Equation (21), may seem the more attractive, as the measured mass loss is integral and the results can be used directly. Unfortunately, the right hand side of Equation (21) cannot be integrated analytically. Many attempts have been made to find appropriate approximations for the right hand side of the equation, but none of them seem to be entirely satisfactory. Differentiation of the data can be achieved, but this method also has its problems. The differentiation of the signal produces noise, which in turn affects the precision of the fit of the processed data with the various models. Both methods have their advocates, but so far no clear advantage has been shown for either.

The usual process for evaluation of the data is to fit the α values to the reaction models given in Table 6, and find the best fit. Calculation of k at several temperatures allows the values of A and E to be determined. Most of the instruments available provide the necessary algorithms to calculate kinetic data from the TG curves, although the user does not always know the algorithms being used, and it may not be the most appropriate choice for the system being investigated. It is also apparent that treatment of TG data by different mathematical analysis produces different values of A and E.[63]

There are several other methods used, including CRTA, which controls one of the experimental variables such as the temperature, evolved gas or the mass change to keep the rate of reaction constant,[4,69] constant

acceleration thermal analysis,[70] temperature jump,[71] and modulated TG.[5] Each of these methods has its own method of calculating the values of A and E.

The general consensus amongst kinetic scientists appears to be that there is no one method which has general applicability to all systems under investigation. Each system needs to be considered for both choice of method and data processing technique, and the ability to make the right choice comes with experience. However, these techniques are very useful if the purpose of the study is to obtain kinetic values to provide a quantitative comparison of materials for a specific use, or to assess the properties of similar samples prepared by different methods.

ABBREVIATIONS AND ACRONYMS

ASTM	American Standard Test Methods
CRTA	Controlled Rate Thermal Analysis
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
DTG	Differential Thermogravimetry
DTM	Differential Thermomagnetometry
EPMA	Electron Probe Microanalysis
FTIR	Fourier Transform Infrared Spectroscopy
HAC	High Alumina Cement
ICTAC	International Confederation for Thermal Analysis and Calorimetry
MS	Mass Spectrometry
MTG or MTGA	Modulated Thermogrametry
NIST	National Institute for Science and Technology
NPL	National Physics Laboratory
py/GC/MS	Pyrolysis/Gas Chromatography/ Mass Spectrometry
SEM	Scanning Electron Microscopy
SEX	Sodium Ethyl Xanthate
TG or TGA	Thermogravimetry
TG/FTIR	Thermogravimetry/Fourier Transform Infrared Spectroscopy
TM	Thermomagnetometry
XRD	X-ray Diffraction

RELATED ARTICLES

Coatings **(Volume 2)**
Thermal Analysis of Coatings

Polymers and Rubbers **(Volume 9)**
Thermogravimetry of Polymers

Thermal Analysis **(Volume 15)**
Thermal Analysis: Introduction • Simultaneous Techniques in Thermal Analysis

REFERENCES

1. R.C. Mackenzie, 'Nomenclature in Thermal Analysis, Part IV', *Thermochim. Acta*, **28**, 1–6 (1979).
2. R.C. Mackenzie, 'Nomenclature in Thermal Analysis', in *Treatise on Analytical Chemistry*, Part 1, eds. P.J. Elving, I.M. Kolthoff, John Wiley & Sons, New York, 1–16, Vol. 12, 1983.
3. J. Paulik, F. Paulik, 'Quasi-isothermal Thermogravimetry', *Anal. Chim. Acta*, **56**, 328–331 (1971).
4. J. Rouquerol, 'Controlled Transformation Rate Thermal Analysis: The Hidden Face of Thermal Analysis', *Thermochim. Acta*, **144**, 209–224 (1989).
5. N. Buckmann, R. Blaine, G. Dallas, 'Modulated Thermogravimetry', *Chem. in Austral., March*, 22–23 (1998).
6. C.J. Keattch, D. Dollimore, *An Introduction to Thermogravimetry*, 2nd edition, Heyden, London, 1975.
7. C.M. Earnest, *Compositional Analysis by Thermogravimetry*, ASTM STP 997, Philadelphia, USA, 1988.
8. W.W. Wendlandt, *Thermal Methods of Analysis*, 3rd edition, John Wiley & Sons, New York, 1986.
9. M.E. Brown, *Introduction to Thermal Analysis*, Chapman and Hall, London, 1988.
10. D. Dollimore, 'Thermogravimetry', in *Thermal Analysis – Techniques and Applications*, eds. E.L. Charsley, S.B. Warrington, Royal Society of Chemistry Special Publication No. 117, London, UK, 31–58, 1992.
11. J.G. Dunn, J.H. Sharp, 'Thermogravimetry', in *Treatise on Analytical Chemistry*, Part 1, ed. J.D. Winefordner, John Wiley & Sons, New York, 127–266, Vol. 13, 1993.
12. P.J. Haines, *Thermal Methods of Analysis*, Blackie Academic and Professional, Glasgow, UK, 1995.
13. E.A. Turi (ed.), *Thermal Characterization of Polymeric Materials*, Academic Press, New York, Vol. 2, 1997.
14. N. Menad, N. Kanari, I. Gaballah, 'Kinetics of chlorination of lead sulfate', *Thermochim. Acta*, **306**, 61–67 (1997).
15. W. Pahlke, Th. Gast, 'A New Magnetic Suspension Coupling for Microbalances', *Thermochim. Acta*, **233**, 127–139 (1994).
16. S.D. Norem, M.J. O'Neill, A.P. Gray, 'The Use of Magnetic Transitions in Temperature Calibration and Performance Evaluation of Thermogravimetric Systems', *Thermochim. Acta*, **1**, 29–38 (1970).
17. P.D. Garn, O. Menis, H.G. Wiedemann, 'Reference Materials for Thermogravimetry', *J. Therm. Anal.*, **20**, 185–204 (1981).
18. P.K. Gallagher, Z. Zhong, E.L. Charsley, S.A. Mikhail, M. Todoki, K. Taniguahi, R.L. Blaine, 'A Study of the Effect of Purity on the Use of Nickel as a Temperature Standard for Thermomagnetometry', *J. Therm. Anal.*, **40**, 1423–1430 (1993).

19. E.M. Gundlach, P.K. Gallagher, 'Synthesis of Nickel based Alloys for Use as Magnetic Standards', *J. Therm. Anal.*, **49**, 1013–1016 (1997).
20. F.W. Wilburn, J.H. Sharp, D.M. Tinsley, R.M. McIntosh, 'The Effect of Procedural Variables on TG, DTG, and DTA Curves of Calcium Carbonate', *J. Therm. Anal.*, **37**, 2003–2019 (1991).
21. G. Hakvoort, 'TG Measurement of Gas–Solid Reactions. The Effect of the Shape of the Crucible on the Measured Rate', *Thermochim. Acta*, **233**, 63–73 (1994).
22. R.A. Jorgensen, F.J. Moyle, 'Gas Diffusion during the Thermal Analysis of Pyrite', *J. Therm. Anal.*, **31**, 145–156 (1986).
23. F. Paulik, J. Paulik, M. Arnold, 'Kinetics and Mechanism of the Decomposition of Pyrite under Conventional and Quasi-isothermal – Quasi-isobaric Thermoanalytical Conditions', *J. Therm. Anal.*, **25**, 313–325 (1982).
24. N. Buckman, N. Hawkins, 'Improving Resolution in Thermogravimetric Anaysis', *Chem. in Austral.*, October, 566–567 (1994).
25. A.E. Newkirk, 'Thermogravimetry in Self-generated Atmospheres. A Decade of Practice and New Results', *Thermochim. Acta*, **2**, 1–23 (1971).
26. H.G. McAdie, 'Recommendations for Reporting Thermal Analysis Data', *Anal. Chem.*, **39**(4), 543 (1967).
27. H.G. McAdie, 'Recommendations for Reporting Thermal Analysis Data. Evolved Gas Analysis', *Anal. Chem.*, **44**(4), 640 (1972).
28. H.G. McAdie, 'Recommendations for Reporting Thermal Analysis Data. Thermomechanical Techniques', *Anal. Chem.*, **46**(8), 1146 (1974).
29. J.G. Dunn, 'Recommendations for Reporting Thermal Analysis Data', *J. Therm. Anal.*, **40**, 1431–1436 (1993).
30. D. Dollimore, 'The Use of Thermal Analysis in Studying Catalysts and the Chemisorption Process', *Thermochim. Acta*, **50**, 123–146 (1981).
31. P. Salvador, M.L.G. Gonzalez, 'Macroscopic Diffusion of Methanol in a Sodium and a Decationated Y-zeolite', *J. Colloid Interface Sci.*, **56**(3), 577–586 (1976).
32. M. Suzuki, D.M. Misic, O. Koyama, K. Kawazoe, 'Study of Thermal Regeneration of Spent Activated Carbons: Thermogravimetric Measurement of Various Single Component Organics Loaded on Activated Carbons', *Chem. Eng. Sci.*, **33**, 271–279 (1978).
33. J.G. Dunn, J. Avraamides, A.C. Chamberlain, N. Fisher, 'The Thermal Decomposition of Sodium Ethyl Xanthate Adsorbed onto Activated Carbon', The AusIMM Annual Conference, Australasian Institute for Mining and Metallurgy, Melbourne, Australia, 163–166, 1996.
34. J.G. Dunn, A.C. Chamberlain, N. Fisher, J. Avraamides, 'The Influence of Activated Carbon on the Thermal Decomposition of Sodium Ethyl Xanthate', *J. Therm. Anal.*, **49**, 1399–1408 (1997).
35. F. Noel, G.E. Cranton, 'Application of Thermal Analysis to Petroleum Research', *Am. Lab.*, **11**(6), 27–50 (1979).
36. P. Aggarwal, D. Dollimore, K. Alexander, 'The Use of Thermogravimetry to Follow the Rate of Evaporation of an Ingredient used in Perfumes', *J. Therm. Anal.*, **49**(2), 595–599 (1997).
37. C. Pichon, V. Risoul, G. Trouve, W.A. Peters, P. Gilot, G. Prado, 'Study of Evaporation of Organic Pollutants by Thermogravimetric Analysis: Experiments and Modelling', *Thermochim. Acta*, **306**, 143–151 (1977).
38. H.G. Wiedemann, 'Applications of Thermogravimetry for Vapor Pressure Determination', *Thermochim. Acta*, **3**, 355–366 (1972).
39. S.J. Ashcroft, 'The Measurement of Enthalphies of Sublimation by Thermogravimetry', *Thermochim. Acta*, **2**, 512–514 (1971).
40. E. Jona, T. Sramko, J. Gazo, 'Heterogeneous Reactions of Solid Nickel(II) Complexes, VIII. Study of the Stoichiometry of Thermal Decomposition of Nickel(II) Complexes with Ammonia', *J. Therm. Anal.*, **7**, 551–560 (1975).
41. T. Sramko, G. Liptay, E. Jona, 'Heterogeneous Reactions of Solid Nickel(II) Complexes, XI. The Influence of different Experimental Conditions on the Stoichiometry of Thermal Decomposition of the Complex $Ni(NCS)_2(\beta\text{-picoline})_4$', *J. Therm. Anal.*, **12**, 217–222 (1977).
42. H.E. Bair, 'Thermal Analysis of Additives in Polymers', in *Thermal Characterization of Polymeric Materials*, 2nd edition, ed. E.A. Turi, Academic Press, New York, 2279–2282, 1997.
43. H.J. Nieschlag, J.W. Hagemann, J.A. Rothfus, 'Rapid Thermogravimetric Estimation of Oil Stability', *Anal. Chem.*, **46**(14), 2215–2217 (1974).
44. R.L. Hassell, 'Thermal Analysis: An Alternative Method of Measuring Oil Stability', *J. Am. Oil Chem. Soc.*, **53**, 179–181 (1976).
45. L. Christel, A. Pierre, D.A.R. Abel, 'Temperature Programmed Reduction of Nickel Manganite Spinels', *Thermochim. Acta*, **306**, 51–59 (1997).
46. T. Kennedy, B.T. Sturman, 'The Oxidation of Iron(II) Sulfide', *J. Therm. Anal.*, **8**, 329–337 (1975).
47. J.G. Dunn, C.E. Kelly, 'A TG/DTA/MS Study of the Oxidation of Nickel Sulfides', *J. Therm. Anal.*, **12**, 43–52 (1977).
48. J.G. Dunn, G.C. De, B.H. O'Connor, 'The Effect of Experimental Variables on the Mechanism of the Oxidation of Pyrite. Part 1. Oxidation of Particles less than 45 μm in Size', *J. Therm. Anal.*, **145**, 115–130 (1989).
49. J.G. Dunn, S.A.A. Jayaweera, S.G. Davies, 'Development of Techniques for the Study of Flash Smelting Reactions using DTA and TG', *Proc. Australas Inst. Min. Metall.*, No 290, 75–82 (1985).
50. G.C. De, 'A Study of the Oxidation of Pyrite using Thermal Analysis Techniques', Master Thesis, Curtin University of Technology, 1987.
51. J.G. Dunn, A.R. Ginting, B. O'Connor, 'A Thermoanalytical Study of the Oxidation of Chalcocite', *J. Therm. Anal.*, **41**, 671–686 (1994).

52. J.G. Dunn, A. Ginting, B.H. O'Connor, 'Quantitative Determination of Phases Present in Oxidized Chalcocite', *J. Therm. Anal.*, **50**(1–2), 51–62 (1997).
53. J.G. Dunn, W. Gong, D. Shi, 'A Fourier Transform Infrared Study of the Oxidation of Pyrite', *Thermochim. Acta*, **208**, 293–303 (1992).
54. J.M. Barton, W.A. Lee, W.W. Wright, 'Some Applications of Thermal Methods of Analysis to Polymers', *J. Therm. Anal.*, **13**, 85–89 (1978).
55. J.G. Dunn, K. Oliver, G. Nguyen, I. Sills, 'The Quantitative Determination of Hydrated Calcium Sulfates in Cement by DSC', *Thermochim. Acta*, **121**, 181–191 (1987).
56. P.A. Barnes, J.H. Baxter, 'A Critical Analysis of the Application of Derivative Thermogravimetry to the Determination of the Degree of Conversion of High Alumina Content', *Thermochim. Acta*, **24**, 427–431 (1978).
57. J.W. Cumming, J. McLaughlin, 'The Thermogravimetric Behavior of Coal', *Thermochim. Acta*, **57**, 253–272 (1982).
58. D.M. Aylmer, M.W. Rowe, 'A New Method for the Simultaneous Determination of Pyrite Content an Proximate Analysis in Coal', *Thermochim. Acta*, **78**, 81–92 (1984).
59. D.W. Brazier, G.H. Nickel, 'Applications of Thermal Analytical Procedures in the Study of Elastomers and Elastomer Systems', *Rubber Chem. Technol.*, **53**, 437–511 (1980).
60. A.K. Sircar, 'Elastomers', in *Thermal Characterization of Polymeric Materials*, 2nd edition, ed. E.A. Turi, Academic Press, New York, 1219–1250, 1997.
61. A. Hale, H.E. Bair, 'Polymer Blends and Block Copolymers', in *Thermal Characterization of Polymeric Materials*, 2nd edition, ed. E.A. Turi, Academic Press, New York, 864–870, 1997.
62. P.K. Weissenborn, J.G. Dunn, L.J. Warren, 'Quantitative Thermogravimetric Analysis of Hematite, Goethite, and Kaolinite in WA Iron Ores', *Thermochim. Acta*, **239**, 147–156 (1994).
63. M.E. Brown, 'Steps in a Minefield. Some Kinetic Aspects of Thermal Analysis', *J. Therm. Anal.*, **49**, 17–32 (1997).
64. J.H. Flynn, 'A Critique of Lifetime Prediction of Polymers by Thermal Analysis', *J. Therm. Anal.*, **44**, 499–512 (1995).
65. V. Satava, F. Skvara, 'Mechanism and Kinetics of the Decomposition of Solids by a Thermogravimetric Method', *J. Am. Ceram. Soc.*, **52**, 591–599 (1969).
66. J. Sestak, 'Non-isothermal Kinetics', in *Thermal Analysis*, ed. H.G. Wiedemann, Birkhauser, Basel, 3–30, Vol. 2, 1972.
67. J.H. Sharp, 'Reaction Kinetics', in *Differential Thermal Analysis*, ed. R.C. Mackenzie, Academic Press, London, 47–77, Vol. 2, 1972.
68. J.H. Flynn, 'The Historical Development of Applied Non-isothermal Kinetics', in *Thermal Analysis*, eds. R.F. Schwenker, P.D. Garn, Academic Press, New York, 1111–1123, Vol. 2, 1969.
69. M. Reading, 'Controlled Rate Thermal Analysis and Beyond', in *Thermal Analysis – Techniques and Applications*, eds. E.L. Charsley, S.B. Warrington, Royal Society of Chemistry Special Publication No. 117, London, UK, 126–155, 1992.
70. A. Ortega, L.A. Perez-Maqueda, J.M. Criado, 'Shape Analysis of the $\alpha - T$ Curves obtained by CRTA with Constant Acceleration of the Transformation', *Thermochim. Acta*, **239**, 171–180 (1994).
71. O. Toft Sørensen, 'Quasi-Isothermal Methods in Thermal Analysis', *Thermochim. Acta*, **50**, 163–175 (1981).

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction

K. Siegbahn
Uppsala University, Uppsala, Sweden

1 Historical Background 1
2 Some Basic Features of Electron Spectroscopy 2
Abbreviations and Acronyms 3

1 HISTORICAL BACKGROUND

Röntgen's discovery of X-radiation in 1895 in Würzburg resulted in an immediate breakthrough, not just in physics but also in society, the latter mainly because of the sensational radiological applications of X-rays. Within a short time it also led indirectly to the discovery of radioactivity by Henri Becquerel. The discovery of X-radiation opened the gate to modern atomic physics, while the discovery of radioactivity did the same for nuclear physics. Later on, the discovery of X-ray diffraction made fundamental contributions to the understanding of atomic and crystal structures. Secondary electrons were soon observed in the scattered radiation when X-rays hit a sample. Development of the corresponding electron spectroscopy took much longer to reach maturity.

Röntgen carefully studied and described most of the properties of the radiation which he himself named X-radiation. With great experimental skill Röntgen attained order in a field that had previously been completely outside human experience. The phenomenon of diffraction of X-rays in crystals opened up two fields of fundamental importance in physics and chemistry, and later on in biology. First, very accurate X-ray wavelengths could be measured giving new and precise insights into the interior of atomic structures. Second, crystal structures, and distances and directions between lattice points of various crystal and molecular structures, could be calculated. All this has had a tremendous impact on atomic and molecular physics and, during the last decades, even on the understanding of life and reproduction, etc.

To go back to the time following the discovery of X-rays, Barkla and others had, by simple means, produced proof of some properties of X-radiation which had significance for the further development of atomic physics. After penetrating aluminium foils the ''soft'' part of the radiation was quickly absorbed. Left behind was a rather ''hard'' radiation. Allowing this harder radiation to strike an arbitrary plate, one could find three different components (Figure 1) in the scattered radiation. The first component consisted of scattered radiation of a similar kind to the impinging radiation. The second component was secondary X-radiation which was softer than the impinging radiation and characteristic of the irradiated material. From the study of this characteristic (fluorescence) radiation Barkla could distinguish between two series of radiation, which he called K and L fluorescence radiation. For a certain element the penetration of the K-radiation could be 300 times higher than that of the L-radiation. From these early observations X-ray spectroscopy was later developed by means of more and more refined X-ray diffraction methods. X-ray spectroscopy made possible atomic analysis of different materials and led to the discovery of new elements such as hafnium and rhenium. The shell structure of the atoms could be investigated by X-ray spectroscopy.

The third component of the scattered radiation was found to consist of electrons, which were expelled from the irradiated material. What happened in this field of spectroscopy? For a very long time it was overshadowed by the successful study of X-ray spectroscopy. In the 1920s spectroscopic investigations of the secondary electrons, the ''photoelectrons'', were begun, for example by H. Robinson in England and M. de Broglie in France. They showed, by magnetic deflection of the electrons, that this component also contained information on the shell structure of the atom and that one could obtain approximate values for the binding energies of the electrons in the different shells. This field turned out to be much more difficult to master, and the accuracy was low.

In our laboratory in Uppsala, interest in this third component began early in the 1950s. For some time previously we had been concerned with techniques to investigate electrons emitted in radioactive β-decay. A suitable way to investigate the γ-radiation was to cover the radioactive sample with a thin lead foil and to record the photoelectrons expelled from the foil using an electron spectrometer. We were finally able to observe and measure even the inherent widths of internal conversion-electron lines due to the finite widths of the atomic levels involved in the conversion process during the nuclear decay. To observe such phenomena a resolution of 1×10^4 at these energies, around 100 keV, was required. The investigations were performed using a magnetic double-focusing iron-free spectrometer with a large dispersion ($R = 30$ cm). It was tempting to go

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

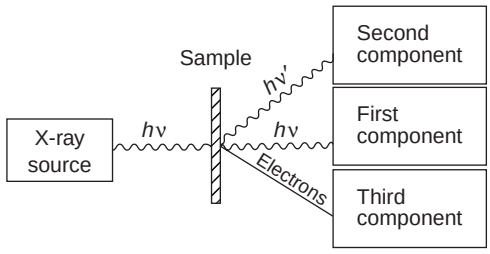

Figure 1 The interaction of X-radiation with matter.

back to Röntgen's original discovery and more precisely to the third component mentioned previously, to see if new methods could be applied to atoms, molecules and solids using an X-ray tube instead of nuclear γ-radiation. For this purpose an iron-free magnetic double-focusing instrument was constructed and used to investigate electron energies some orders of magnitude less than in nuclear β-decay. For X-ray energies double-focusing electrostatic instruments were later designed.

During the 1950s and the 1960s our work finally led to the development of an electron line spectroscopy technique which was characterized by narrow electron lines from the materials excited by X-radiation and ultraviolet radiation. The recorded electron lines were well defined by widths set by the inherent atomic widths themselves. Details of the spectroscopy technique, which had applications in many fields, were published in 1967 in the book "*ESCA: Atomic, Molecular and Solid State Structure Studied by Means of Electron Spectroscopy*", followed in 1971 by "*ESCA, Applied to Free Molecules*", both authored by our group.

Electron spectroscopy is nowadays an independent alternative to X-ray and other spectroscopies and contributes much new information about the properties of matter. Surface science and technology in particular makes use of electron spectroscopy as a convenient tool to characterize surface conditions and reactions. Applications are found in many other areas on the borders between physics and chemistry. To emphasize these latter aspects the acronym ESCA (electron spectroscopy for chemical analysis) was introduced. The acronym XPS (X-ray photoelectron spectroscopy) is also used.

2 SOME BASIC FEATURES OF ELECTRON SPECTROSCOPY

In electron spectroscopy the sample is introduced into a vacuum and irradiated by soft X-rays, ultraviolet light, laser light, or synchrotron radiation. In addition, an electron beam can also be used to cause Auger electron scattering spectroscopy. A certain number of the electrons which are expelled from their atomic bonds may leave the surface layer without energy loss, the scanning depth being set by the energy-dependent mean free paths. An electrostatic lens system collects and focuses the electrons onto the slit of the electron analyzer. An extended detector is situated in the focal plane where the different electrons are simultaneously being recorded. These events appear on the screen of a computer as a set of separate electron lines (see Figure 2). The lines characterize the examined substance, the elemental composition and the chemical bonds.

Solid materials with elements over the whole periodic system can be examined, as well as gases and molecular beams down to low pressures of $\sim 10^{-5}$ torr. The technique has also been applied to liquids and solutions. Applications of electron spectroscopy using synchrotron radiation and laser radiation for excitation have been increasing during more recent years.

Figure 3 illustrates an atom which is bound to other atoms in molecular valence bondings, adsorbed or chemically bound to a surface. To eject electrons from the inner part of an atom (the "core" region) harder radiation is required than for the external "valence" electron region. When an electron is emitted a hole is created in the electron shell. This is immediately followed by a relaxation of the other electrons which thereby adjust themselves to the new potential by means of "shrinkage" and "flow" in the surrounding electron cloud. The hole has a particular, very short, lifetime ($<10^{-14}$ s) before it is filled by an electron from some external shell. The atom can dispose of the released energy either in the form of an X-ray photon or by emitting another (Auger) electron. In the latter case three levels are involved (e.g. LMM, etc.). Such electrons are recorded together with the primary emitted photoelectron spectrum. One advantage of exciting Auger electrons by X-rays instead of electrons is the great improvement in signal-to-noise ratio. However, using an electron beam for excitation of Auger electron enables a high spatial resolution since the primary electron beam can be focused to a very small spot (<10 nm) on the sample, suitable for scanning and high resolution imaging.

When a chemical compound is formed between two or more atoms it causes a redistribution of the external valence electrons of the atoms which are responsible for the chemical bonds. The electric potential around a certain atom in the molecule will then be changed to an extent which is dependent on the specific nature of the chemical bond in the different cases (e.g. different "states of oxidation"). This new electric potential influences the bond strength of the inner electrons which surround the atomic nucleus and the consequence of this is a "chemical shift" in the recorded electron spectrum. This information is extremely important in electron

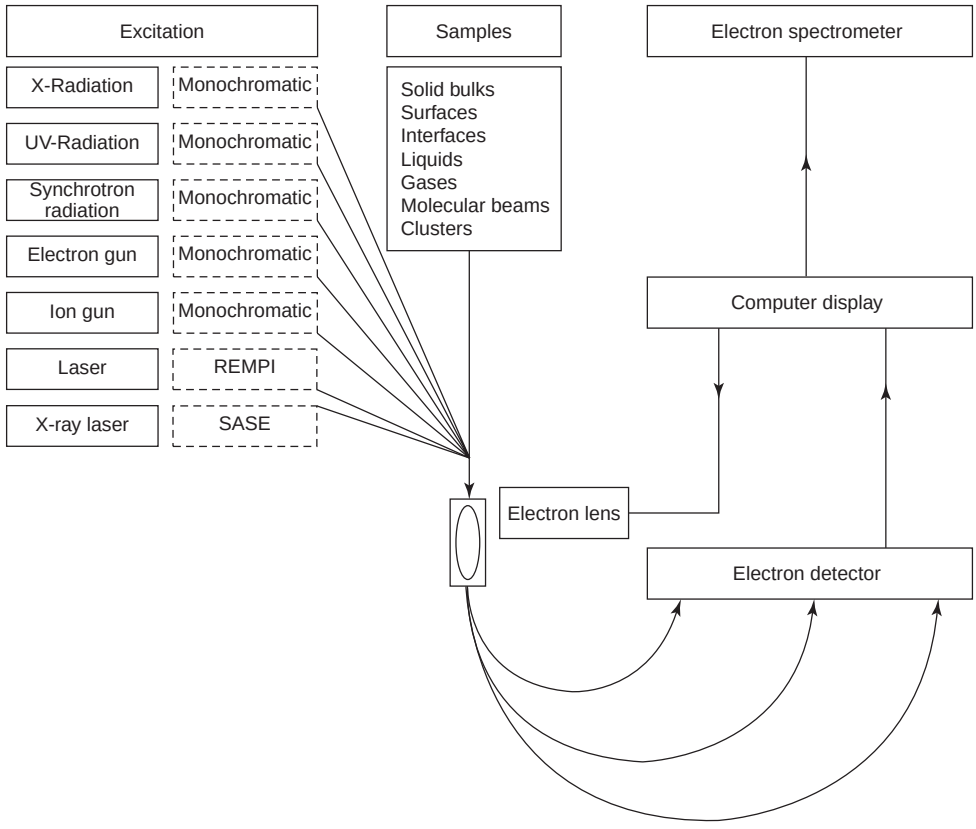

Figure 2 Scope of electron spectroscopy. REMPI, resonance-enhanced multiphoton ionization; SASE, self-amplification of synchrotron emission.

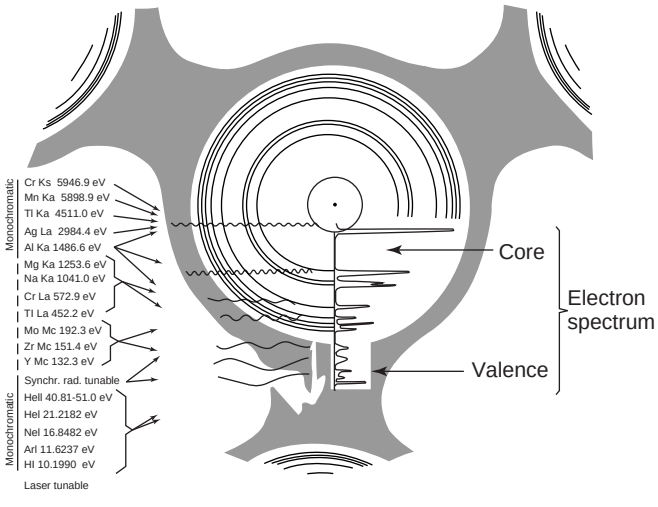

Figure 3 Excitation of core and valence electron spectra.

spectroscopy. One can for example easily distinguish, even at fairly modest resolutions, between metallic atoms and those which have been oxidized at the surface layer.

A great number of ESCA instruments are now in operation and the applications are increasing. Some are provided with X-ray monochromators for high-resolution work and some are combined with water-cooled rotating X-ray anodes. It is this combination which offers the best advantages for high resolution at high intensities in ordinary laboratories. Here and also in synchrotron laboratories the recorded electron line widths are ultimately set by the inherent level widths. An X-ray monochromator, using the (1010) plane of toroidally bent quartz wafers mounted in an assembly for the Al Kα radiation, significantly improves the conditions for accurate deconvolutions at the analysis of the spectra, since the lines get narrower and the signal-to-noise ratio is enhanced. The electron lines due to X-ray satellites situated close to the real line structures are eliminated and so is the background noise due to Bremsstrahlung. Ultrasoft X-ray lines are furthermore available in the Mζ X-radiation from yttrium, zirconium, etc. in the energy region of $\sim$150 eV. These can be focused by multilayer X-ray mirrors at high reflectivity.

ABBREVIATIONS AND ACRONYMS

ESCA	Electron Spectroscopy for Chemical Analysis
XPS	X-ray Photoelectron Spectroscopy

X-ray Photoelectron and Auger Electron Spectroscopy

Noel H. Turner
Springfield, USA

1 Introduction **1**
2 Basic Principles **2**
2.1 Surface Sensitivity 2
2.2 X-ray Photoelectron Spectroscopy 3
2.3 Auger Electron Spectroscopy 6
3 Instrumentation **8**
3.1 Vacuum Systems 8
3.2 Sample Introduction and Handling 11
3.3 Excitation Sources 13
3.4 Electron Energy Analyzers 16
3.5 Small Area Analysis 18
3.6 Data Collection and Handling 20
4 Accessories **25**
4.1 Ion Guns 25
4.2 Flood Gun 26
4.3 Fracture Stage 26
4.4 Special Auxiliary Chambers 27
4.5 Energy Dispersive X-ray Analysis 27
4.6 Residual Gas Analyzer 27
5 Conclusions **27**
Acknowledgments **28**
Abbreviations and Acronyms **28**
Related Articles **28**
References **28**

The use of X-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES) to analyze the surface composition and chemical environment of solids has been well established since about the early 1970s. These approaches find many applications in both basic and applied scientific investigations. The basic principles, i.e. surface sensitivity, the experimental observations employing these procedures, and the working methods, are described. The instrumentation utilized with the commonly available equipment for these approaches is summarized also.

1 INTRODUCTION

The ability to analyze the outermost few atomic layers of solid surfaces has expanded greatly since the early 1970s. Two analytical techniques for studying the surface composition have been in the forefront of the methods for such determinations: XPS and AES. XPS also is called electron spectroscopy for chemical analysis (ESCA). Overall, these procedures have been judged by Seah and Briggs as being the most useful of the modern surface analysis techniques, with a detection limit for most elements of about 0.5%.[1] XPS and AES have been used in numerous areas of important modern technologies; examples have included catalysis, corrosion, new materials, tribology, electrodes, electronic devices, cleaning procedures, magnetic storage media, adhesion, adsorbents, biological surfaces, and polymers. In addition, much basic information has been supplied by XPS and AES to help understand chemical reactions at solid interfaces, the atomic environment of solids at and near the surface (such as high-temperature superconductors), and compositional differences between the surface region and the bulk. Every year a few thousand scientific articles are published in which one or both of these techniques have been used. It does not appear that there will be any decrease in the interest and use of these techniques in the foreseeable future. There was very rapid advancement in both equipment and technique when these methods were developed in the 1970s. However, many more subtle improvements continue to be made in these approaches, and advances should continue for many years. Much of this current effort has been directed to making these methods better understood and more reliable in routine laboratory environments. The ability to analyze smaller areas has progressed greatly with XPS and AES since the techniques were developed initially.

In 1984 the Versailles project on advanced materials and standards (VAMAS) program in surface chemical analysis was initiated; it has 31 different projects that address many topics in surface analysis.[2] The main thrust has been directed toward a better understanding of the physical processes of these methods, the instruments that are used with these methods, standardized data output for communication to other laboratories, validation of data reduction programs, and reference materials and data. The American Standards for Testing and Materials (ASTM) has a continuing program of providing definitions and procedures using these techniques.[3] In 1992 the International Standards Organization (ISO) formed a committee on surface analysis.[4] It was expected that the ISO would base much of its work on the earlier VAMAS and ASTM efforts. The effects of these efforts have been explored.[5]

Many XPS and AES analysis systems are purchased as complete analysis systems. However, it is possible to build systems from components, such as energy analyzers, excitation sources, etc. Often these components have unique designs that possess special capabilities,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

reduced cost, or are targeted for the replacement of older components. Some parts of XPS and AES systems, notably vacuum components, can have useful lifetimes of decades, so that component replacement can be cost effective in certain situations. Components are available from a wider variety of sources than just system suppliers. Some lower-cost XPS and AES analysis systems are available that are designed for many routine applications, but will still have many fairly advanced features. Software for data analysis that has special capabilities is being offered from many sources besides that supplied with complete analysis systems.

2 BASIC PRINCIPLES

An extended discussion of the basic principles involved with AES and XPS is beyond the scope of this article; other sources give more detailed explanations.[6–10] However, many of the important factors are discussed as they relate to the available instrumentation and experimental procedures.

2.1 Surface Sensitivity

The most basic question that has to be asked about XPS and AES is: Why are these techniques sensitive to the surface region of solids? Both XPS and AES make use of the fact that electrons with kinetic energies from about 30–2500 eV (this is the range of almost all of the electrons typically analyzed by either XPS or AES that are ejected from an atom in a solid) will have an inelastic mean free path (IMFP or λ) of about 0.5–3 nm. The IMFP is defined as the average distance that an electron will travel before it undergoes an inelastic collision with another electron or nucleus of the matrix constituent atoms.[11] Thus, very few electrons created at depths greater than two to three times the IMFP will leave the bulk with the characteristic photoelectron or Auger electron kinetic energy with which they were created. Using data from experiments with very thin overlayers of one element deposited on another, it has been ascertained that the intensities of the substrate signal vary approximately as $1/e$ with film thickness.[12] In some cases it has been suggested that the IMFP for organic compounds and films is greater than IMFPs for metals and inorganic layers, perhaps being in the 8–12 nm range.[13,14] These proposed values have been questioned,[15] and somewhat smaller values have been suggested from the results in other investigations. Based upon the findings from many studies, IMFPs have been subdivided by Seah and Dench into three material-dependent subclasses: elemental, inorganic, and organic.[16] Even with these subdivisions, there still is considerable scatter in the data. More recent individual reports are usually in general agreement with the earlier work. It also has been proposed by Seah and Dench that the IMFP follows the relationship (Equation 1)

$$\lambda = aE^b + \frac{c}{E^2} \tag{1}$$

where a is a material-dependent constant, b is an exponent with a value of 0.5, c is a constant that depends on the class of material, and E is the kinetic energy of the ejected electron. The second term predominates at lower kinetic energies, i.e. less than about 100 eV. An idealized representation of the IMFP of electrons as a function of kinetic energy is given in Figure 1, based on Equation (1) for three elements.[16] Note that there is a minimum value for the IMFP of about 0.5 nm for kinetic energies in the 100 eV region. Then the IMFP increases slowly as the kinetic energy becomes larger. Others have suggested that b has a value near 0.75.[17–21] Regardless of the various suggested functional dependencies for the IMFP, the experimentally determined values all are approximately within the region shown above. Jablonski has questioned the functional relationship in Equation (1) and has proposed the Equation (2)

$$\lambda = \frac{aE}{\ln(\gamma E)} \tag{2}$$

with a and γ being material constants.[22] (The constant a in Equation 2 differs from that in Equation 1). Equation (2) follows much of the reported experimental data at kinetic energies of 500 eV and greater, but there are deviations for electrons with lower kinetic energies.

Other parameters have been proposed to account for the surface sensitivity of XPS and AES. These have included the attenuation length (AL), defined as the average distance that an electron travels between successive inelastic collisions as given by a particular

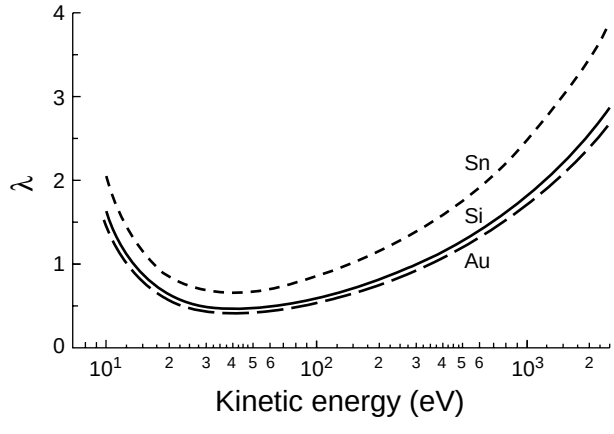

Figure 1 A representation of the IMFP versus the kinetic energy of an ejected electron computed for Sn (short dashed line), Si (solid line), and Au (long dashed line).

model.[11] More recently, a quantity called the depth distribution factor (DDF) has been defined as the emission probability from a given distance below the surface for the escape of an Auger or photoelectron.[23] The magnitude of these other parameters does differ from the IMFP, and sometimes there are variations of about 30% or more. However, these differences still do not change the basic fact that electrons in the kinetic energy region detected in almost all conventional XPS or AES experiments originate from only the first few atomic layers of a solid. For convenience the term IMFP will be used in this article. Different approaches have been used to predict IMFPs on a fundamental basis of material properties.[24,25] This entire topic has been reviewed extensively by Powell.[12]

Many Auger transitions, especially for most metals, are in the 50–100 eV region, but often there are overlapping peaks if other elements are present. With XPS the use of grazing take-off angles combined with angular resolution detection increase surface sensitivity; this topic will be treated in detail later. Either approach often gives enhanced surface sensitivity compared to more typical XPS or AES analyses.

2.2 X-ray Photoelectron Spectroscopy

The photoelectric effect in which a valence or core electron is ejected when irradiated with soft X-rays forms the physical basis of XPS. The Einstein relationship (Equation 3) governs the process

$$E_{KE}^{F} = h\nu - E_{BE}^{F} - \phi_{sp} \quad (3)$$

where E_{KE}^{F} is the measured kinetic energy (peak maximum) of the ejected electron with respect to the Fermi level, $h\nu$ is the photon energy (for XPS this is usually in excess of 1 keV), E_{BE}^{F} is the binding energy of the ejected electron with respect to the Fermi level, and ϕ_{sp} is the work function of the spectrometer. This last factor is constant for a particular instrument and should vary little during its lifetime. A diagram of the photoejection process is shown in Figure 2. A simple rearrangement of Equation (3) gives the binding energy. This value is usually reported with XPS data, as it is independent of the photon energy. All of the elements except hydrogen and helium can be detected by XPS, but there are some cases where the presence of hydrogen has been inferred from XPS spectra.[26–28] Each element has its own unique set of binding energies, so that identification can be made of the element or elements that are observed in an XPS spectrum. In a few cases there is an overlap of lines from different elements. However, there are other lines that will be observed that permit differentiation of the elements present in those overlapping situations, unless the signal is due to a very small amount of an individual element that is overwhelmed by that of the another element.

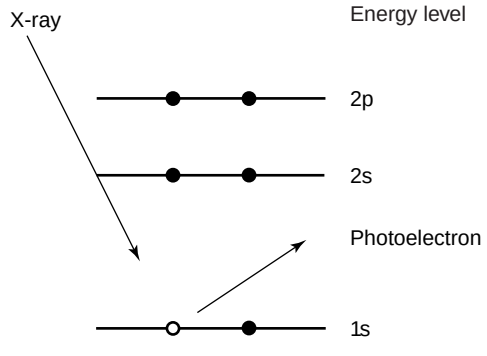

Figure 2 A diagram showing the ejection of a photoelectron from a carbon atom.

The energy level of the ejected photoelectron is indicated by atomic notation. For example, in a typical XPS spectrum of Ag, as shown in Figure 3, peaks due to the 3s, 3p, 3d, 4s, 4p, and 4d levels of Ag are observed. The spin–orbit split of the 3p and 3d levels can also be seen. The lower multiplicity levels, i.e. the $3p_{1/2}$ and $3d_{3/2}$ lines, are on the left side of these respective doublets. The higher intensity lines are the $3p_{3/2}$ and $3d_{5/2}$ peaks. The ratio for the intensities of the doublets is calculated from the total angular momentum, j, for the level being considered using j–j coupling, i.e. $j = l + s$ where l is the orbital angular momentum and s is the spin quantum number. For p levels, the $p_{3/2}$ peak has twice the relative intensity of the $p_{1/2}$ line. The relative ratios for d levels is 1.5 ($d_{5/2}$ to $d_{3/2}$) and for f levels is 1.33 ($f_{7/2}$ to $f_{5/2}$).

The deeper core levels for Ag could not be observed, as their binding energies are greater than the characteristic photon energies of the most common X-ray sources employed in XPS (1253.6 and 1486.6 eV for Mg and Al, respectively). In addition to the photoelectron peaks, several Auger transitions are found, as indicated in the spectrum. These lines are due to the holes created in the 3d levels in the photoemission process; details about the Auger mechanism are given later. Binding energy values for all elements have been published together with spectra from most elements.[29] A database of binding energies for use with a personal computer has been compiled under the direction of the National Institute of Standards.[30] This database has thousands of entries with numerous compounds; it also lists many Auger transitions. A review of the available databases has been made.[31]

Besides elemental identification in the surface region, information about the chemical environment of these elements usually can be deduced from XPS spectra. For example, there is a difference of about 5 eV between the $2p_{3/2}$ peak of metallic iron (707 eV) and Fe_2O_3 (712 eV).[32] For sulfur, there is about an 8 eV separation

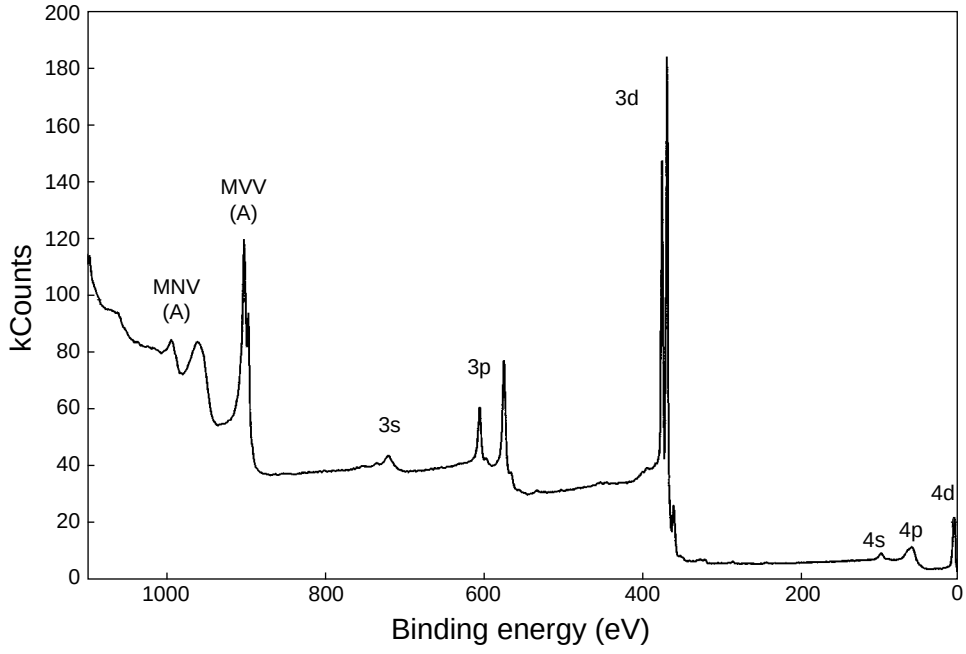

Figure 3 An XPS survey scan of sputter-cleaned silver with the energy of the photoelectron transitions in terms of binding energy using a Mg Kα X-rays. The X-ray excited Auger peaks are indicated by (A). The split of the 3d and 3p levels can easily be seen with the $3p_{1/2}$ and $3d_{3/2}$ peaks on the left side of the respective doublets. The shoulder on the right-hand side of the $3d_{5/2}$ peak is due to $K\alpha_{3,4}$ X-rays.

in the 2p level between S^{2-} and SO_4^{2-} (approximately 161 and 169 eV, respectively).[33] For some elements, such as Cu, these shifts are smaller, and other spectral features, namely the use of Auger lines and 'shake-up' peaks (considered below), have to be employed to gain chemical information.[34] In addition, there can sometimes be variations in binding energy with atomic environment, for instance aluminum bound either tetrahedrally or octahedrally.[35]

The accurate determination of binding energies can be made by comparison to standard values that have been measured with metals.[36] Most of these determinations have error limits of ±0.03 eV. A number of instrumental and experimental factors have been evaluated that affect these measurements.[37] In addition, the zero point of the binding energy scale is set by the Fermi edge of a metal. Ni has been used for this purpose, but Ag has been suggested to have a more distinct Fermi edge.[38] Surveys with instruments at many laboratories from different manufacturers have been conducted, and it has been found that most users usually agree to about 1 eV with oxide samples.[39,40] The effort required to obtain this level of agreement usually involved following manufacturers' procedures and some sample preparation that usually is not very time-consuming. With an individual instrument in a typical laboratory environment, the position of some metal peaks can usually be determined to within 0.05 eV or less.[41] Sometimes special methods are employed to determine accurately the peak position.[42,43] Unfortunately, under typical laboratory conditions, errors an order of magnitude larger often occur. In some very careful studies, small binding energy differences between the surface and bulk atoms in metals are observed; this topic has been reviewed elsewhere.[44] With alloys there can be small shifts in the Fermi level that will affect measured binding energies.[45] Similar findings have been made with bimetallic clusters.[46] Smaller clusters have higher binding energies than the bulk values, an effect suggested to be due to nonmetallicity increasing as size decreases.[47]

Many procedures have been utilized to calculate binding energies and binding energy shifts. Empirical methods employing the charge on an atom (computed by molecular orbital methods and compared to a standard material) have been used. Other procedures, such as Hartree–Fock calculations and electrostatic models, have also been employed. The equivalent core approximation, set up in a manner similar to the Born–Haber cycle, has also been applied. Further information on these computational procedures is available elsewhere.[45,48,49]

When insulators are being analyzed there is the complication of charging. This effect is observed when the surface of the specimen becomes positively charged, because more electrons leave the surface region than are replaced by those from the bulk. The peak maximum will

shift to lower kinetic energies (higher binding energies), sometimes by tens of electron volts. Usually a charge equilibrium is established after a reasonably short time. Sometimes a spectrum will not be observed, or the peak shape will be badly distorted. It has been suggested in some cases that surface layer atoms will have a different charging shift from those atoms below the surface.[50,51] Also, lateral charging across the surface has been observed due to a nonuniform X-ray flux and sample mounting procedures.[52,53] Even conductors deposited as clusters on insulators can experience charging. In addition, the cluster size and their distribution can affect observed binding energies.[54] The effects of substrate conductivity and photoelectron cross-section on differential charging have been explored for some polymer systems.[55] With uniform polymer layers on NaCl the effects due to the photoelectron flux from both layers and the electron energy from a flood gun (considered later) were found to be important factors in charging behavior.[56]

Several approaches can be tried to overcome, or at least minimize, charging effects in XPS. The Auger parameter (discussed below) can be used, because it is assumed that both the photoelectron and Auger peaks will be shifted by the same amount. Then a comparison is made with previously reported data. Another method is use of the adventitious carbon peak present on almost all sample surfaces that have been exposed to the atmosphere.[57] Then, all other peaks are referenced to the position of this peak, that is usually is set to a value of 284.6 or 285 eV. This procedure has been criticized because it assumes that the binding energy position of this peak will always be the same.[58] Measuring the binding energy of ion-implanted Ar (discussed below) has been suggested as a reference point to correct for the charging effect.[59,60] However, this technique might suffer from shifts in the Ar peak position due to extraatomic effects being different with various materials. Another approach is the use of a flood gun. A high transmission screen just above the sample has been used with a flood gun in some systems.[61] The use of a magnetic lens (discussed in detail below) with low-energy electrons can considerably reduce charging effects.[62] Low-energy positive ions with energies of a few electron volts also have been shown to greatly overcome charging effects. More details about charging in XPS are available elsewhere.[63]

Occasionally degradation of specimens occurs because of X-ray irradiation. This can often be observed as a diminution of a peak of an element in one oxidation state with the increase in intensity of another, such as for perchlorate to chloride.[64] This effect has also been noted with some polymers. In some instances it has been suggested that sample heating from the X-ray source may cause specimen degradation. In many situations this problem can be minimized by using a monochromatic X-ray source (discussed below). Also, many newer instruments can collect data much faster than was possible with older equipment, so that the time a sample is exposed to X-rays, and therefore the possible damage, is greatly reduced.

The use of X-ray-excited Auger peaks is usually based on the Auger parameter α, which is defined (Equation 4) as

$$\alpha = E_{KE}^{F}(j, k, l) - E_{KE}^{F} + h\nu \tag{4a}$$

or

$$\alpha = E_{BE}^{F} - E_{BE}^{F}(A) + h\nu \tag{4b}$$

Here $E_{KE}^{F}(j, k, l)$ is the kinetic energy of the Auger transition referenced to the Fermi level (where j, k, and l denote the energy levels involved in the Auger transition, discussed in more detail below) and $E_{BE}^{F}(A)$ is the energy position of the Auger peak on the *binding energy scale*. This definition usually is referred to as the modified Auger parameter, as an earlier definition sometimes gave negative values. Note that the Auger parameter is independent of the X-ray source employed to obtain a spectrum. The Auger parameter can also be used to determine the chemical environment of the detected elements. Often the differences in the Auger parameter are larger than the changes observed in the photoelectron peak energy alone. In many systems the peak shifts found for Auger transitions are larger than those for photoemission lines.[65] Peak positions will be shifted by charging: about the same amount for both photoelectron and Auger transitions. This can help in making assignments for the chemical state information in cases where a specimen is undergoing charging. Further details on the use of Auger parameters for many materials have been reviewed.[66]

Shake-up peaks result from a discrete energy loss of a core electron when a valence-band electron is ejected into an unoccupied conduction band during the photoemission event. Very strong shake-up peaks are observed with some transition metals with changes in the oxidation state, for example oxides versus the pure metals.[67] With some metals, the shake-up peaks may be approximately 10 eV higher in energy than the main line. Shake-up peaks are observed also in polymers where there are conjugated π systems.[68] In these cases, the shake-up peaks are about 5 eV above the main photoelectron line. The intensity of the shake-up peaks in these cases is much less than the main photoelectron line.

The valence band (VB) region of XPS spectra is not usually used extensively with X-ray excitation. In Figure 3 this area is denoted as the 4d peak. There are two reasons for this spectral region being neglected. First, the intensity

of the peaks is weak compared to the core level lines with the normal X-ray sources. Second, all of the elements in the sample under study will contribute to the spectrum in the VB region. Thus, making use of this portion of the spectrum is usually more difficult for obtaining elemental and chemical information for a specific element. Better VB spectra can be obtained with special sources, such as a He lamp that has a photon energy source of 21.2 eV. Such sources can be an attachment to a spectrometer.

However, the situation has changed with more efficient instruments using regular X-ray sources. First, as noted earlier, newer spectrometers can obtain spectra with much greater intensity than with earlier equipment. This allows data to be acquired in this region much more quickly. Secondly, the use of Xα calculations (a procedure used to compute the energies and intensities of atomic levels for chemical species) to model the VB spectra can be done on a routine basis with a personal computer. With this approach Sherwood and co-workers have been able to discover differences in surface species that could not be found by using just core level peaks alone.[69,70] This technique has opened new avenues for analyses in many systems.

The surface sensitivity of XPS can be enhanced by analyzing electrons that leave the surface at a near glancing angle; this technique is called angular resolved X-ray photoelectron spectroscopy (ARXPS). A simple model of a thin uniform layer of material B at a depth d on a substrate A illustrates this effect. The intensity of the substrate signal has been approximated by Equation (5)

$$I_a = I_a^o \exp\left(\frac{-d}{\lambda \sin\theta}\right) \tag{5}$$

where I_a is the measured intensity of the substrate signal, I_a^o is the signal of A without B, and θ is the angle at which the detected electrons leave the sample relative to the surface normal. The relationship for the signal intensity of the overlayer is (Equation 6)

$$I_b = I_b^o \;\; 1 - \exp\left(\frac{-d}{\lambda \sin\theta}\right. \tag{6}$$

where I_b is the intensity of the overlayer signal and I_b^o is the intensity for an overlayer of several λ. Nondestructive analysis of thin films, i.e. without using ion beams, can be performed by this approach, but added factors need to be considered.[71–73] Angular resolved procedures can also be used in AES.[74] The experimental procedures for using ARXPS are given later.

2.3 Auger Electron Spectroscopy

There are two pathways for the atom to relax when a core hole is created. One way is for an X-ray to be emitted when an electron from a higher orbital falls into the hole

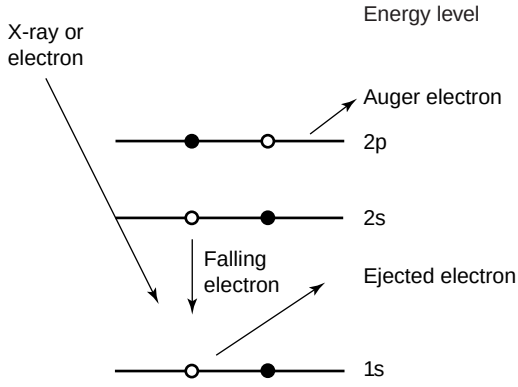

Figure 4 Diagram showing the ejection of an Auger electron from a carbon atom.

(X-ray emission). The creation of an Auger electron is the alternative mechanism. In this case, an electron from a higher orbital falls into the empty hole while another electron is ejected from the same or even higher energy level of the atom. This process is illustrated in Figure 4. The energy relationship for the Auger process with respect to the Fermi level is given by Equation (7)

$$E_{KE}^F(j,k,l) = E_{KE}^F(j) - E_{KE}^F(k) - E_{KE}^F(l) - U \tag{7}$$

where $E_{KE}^F(j)$ is the binding energy of the electron ejected during the initial ionization event, $E_{KE}^F(k)$ is the binding energy of the electron that falls into the empty level, $E_{KE}^F(l)$ is the binding energy of the ejected electron, and U is the hole–hole repulsion energy. This last term takes into account the fact there are two empty electron states in the excited atom; U is of the order of 1–10 eV. It should be noted that Auger energies often are given with respect to the vacuum level. The difference between the Fermi level and vacuum level is about 5 eV. There are more complex decay mechanisms found with some materials that contribute to the overall Auger transition beyond the description given above. However, these effects usually do not change the energy of the observed Auger transition to the extent that elemental identification cannot be made. Details about these more complex decay pathways have been reviewed.[75–77] AES is similar to XPS in that all of the elements except hydrogen and helium can be detected, but the presence of hydrogen has been observed by line shape changes.[78]

Auger electrons can be created in several ways. However, for surface analysis the two predominant methods are from an electron beam with a kinetic energy of about 2–10 kV, or from an X-ray source. The electron beam needs to have at least 2–3 times the kinetic energy to observe the Auger transition of interest in order to observe the transition with reasonable signal strength. The X-ray source has to have an energy greater than the binding energy of the initial core level that participates

in the observed Auger transition. Beam currents usually range from tens of microamperes to about a nanoampere. In certain Auger spectrometers, the beam diameter can be adjusted to be less than 20 nm. Compilations of Auger spectra have been made for routine elemental identification.[79,80] A high-resolution AES database has been constructed for many elements with an absolute intensity scale.[81] Instrumental contributions have been removed from the data so comparisons to any instrument can be made.

AES lines are identified with X-ray notation. AES spectra in surface analysis are displayed in two modes, derivative and integral. Details about the presentation methods for AES spectra are given later. For example, if an electron is ejected from the 1s level of Si, and a $2p_{3/2}$ electron falls into the empty hole and another $2p_{3/2}$ is ejected from the atom, the transition would have the notation KL_3L_3. The energy difference between the $2p_{3/2}$ and the $2p_{1/2}$ levels in Si is small (about 0.6 eV) in comparison to the observable width of this Auger transition. (The $2p_{3/2}$ and the $2p_{1/2}$ levels in Si levels can be resolved with many XPS instruments.) Thus, the overall observed peak is a combination of these levels, and its notation for this example is $KL_{2,3}L_{2,3}$. If valence electrons are involved in the Auger transition of interest, they are often denoted by V. Examples of this are illustrated in Figures 5 and 6, which show the AES spectrum of Li_2SO_4 in both integral and derivative display modes. In this case a hole in either the $2p_{3/2}$ or $2p_{1/2}$ level for S is filled with an electron falling from a valence level and ejecting another valence electron. For this overall Auger peak the notation would be $L_{2,3}VV$.

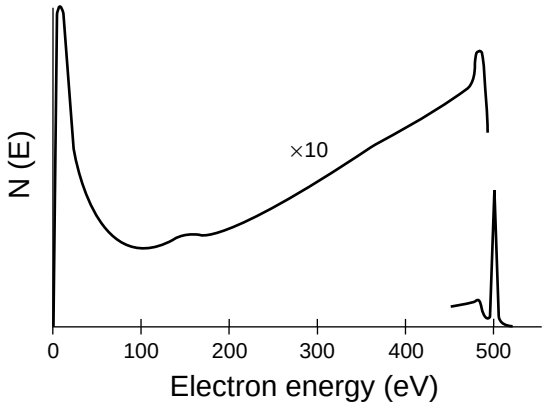

Figure 5 The electron-excited AES spectrum of Li_2SO_4 observed with an electron beam excitation of 500 V in the nonderivative display mode is shown. The small bump at about 140 eV is due to the sulfur LVV transition. The strong peak at 500 eV is due to the reflected primary electron beam. (Reproduced from D.E. Ramaker, J.S. Murday, N.H. Turner, 'Extracting Auger Line Shapes from Experimental Data', *J. Electron Spectrosc. Relat. Phenom.*, **17**, 45–65 (1979).)

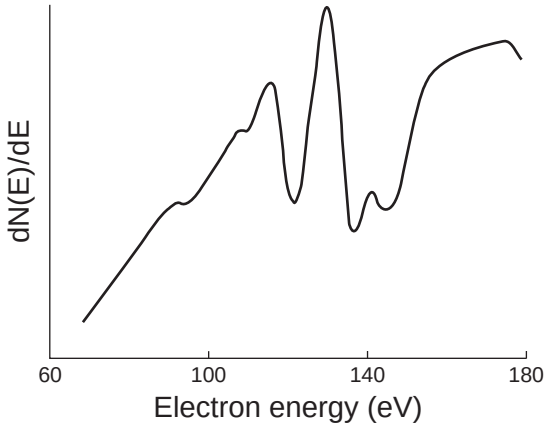

Figure 6 The derivative AES spectrum from Figure 5.

Chemical effects often are observed in AES. There are shifts in the position of the observed peaks for some elements depending on the chemical state; for example, the $KL_{2,3}L_{2,3}$ peak for Si is observed at 1616 eV whereas this line is at 1608 eV for SiO_2.[82] In other instances there is both an energy shift and a change in the peak shape. For example, with the S $L_{2,3}VV$ level, peaks of sulfides and sulfates are approximately 150 eV with a doublet 126 and 140 eV, respectively.[33] The peak at 150 eV in Figure 6 is due to the reduction of a sulfate type species by the electron beam. Most of the observable peak shape differences involve transitions with valence electrons. This method of identifying chemical species can be employed sometimes and often is referred to as fingerprinting.

When AES was first utilized for elemental identification in surface analysis, the derivative mode of spectral display was used. The signal levels of the integral peaks were very low compared to the overall intensity of the spectrum and this usually made observation of Auger peaks difficult. However, by taking the derivative of the spectrum, the Auger peaks were much easier to observe visually. In addition, the sloping background was reduced or eliminated. The compilations of Auger spectra usually are of derivative spectra, and the peak position is defined as the minimum of the peak. With improvements in analyzers and electron detectors, many spectra are now reported with integral data. Some compilations of Auger data contain spectra displayed in this format also. The choice of display method is made by the author or authors of a report. There is a difference in the position of the peaks between the methods. The peak maximum in the integral display is at the zero cross-point in the derivative mode representation. The energy difference usually is only a few electron volts between these methods unless there is a badly distorted peak.

Accurately determined values for AES transitions have been made in the same fashion as those in XPS. Values for some metals with errors of ± 0.03 eV have been determined.[36] The transitions used in these instances are unusually sharp for Auger peaks. The accuracy of these determinations is much greater than almost all reported values in general analysis.

Charging can be a much more serious problem when using electron beam AES compared to XPS with insulating samples. When charging happens, peaks will be shifted to higher kinetic energies or huge peaks will appear in the spectrum that are much larger than any observable Auger transition. When the latter phenomenon occurs, there is little if any possibility that useful information will be obtained. Also, the extent of charging can change from sweep to sweep. Several approaches can be taken in an attempt to overcome this problem. These include changing the beam voltage and current to try to balance the incident beam current with the secondary electron emission. Unfortunately, in some cases the beam voltage has to be lowered to a potential where some transitions will not have a strong signal. In a few instances the transition energy will be above the beam voltage that reduces charging. Tilting the specimen to a more grazing angle with respect to the electron beam has sometimes been successful. Rastering the beam over a large area of a specimen is occasionally effective. Wrapping the sample in metal foil with a small hole that exposes only a small part of the specimen is another approach that has sometimes been effective. No single procedure is successful in all cases, and the analyst has to be prepared to try several methods. Further information about charging in AES is available.[83–85]

Another effect that sometimes has to be considered in AES is sample degradation, which occurs more frequently than in XPS. An example of this effect has been observed with α-Al_2O_3[86] and some sulfates.[33] Some of the same approaches tried to overcome charging can be used to reduce or eliminate degradation, namely changes in the beam current and voltage, rastering, etc. The local temperature may rise where the beam strikes the specimen and may also play a role in specimen degradation. Sample degradation in AES has been reviewed elsewhere.[87]

3 INSTRUMENTATION

There are many components that comprise an XPS or AES spectrometer, and some commercially available components or those built from component systems are capable of performing both XPS and electron-excited AES. Often there are compromises with such multi-technique systems. The various items and experimental procedures used with XPS and AES spectrometers discussed here are the vacuum system, sample handling and introduction, excitation sources for the production of photoelectrons or Auger electrons, the detection of these electrons, and the data collection and handling procedures. Only passing references are made to the associated electronics. More details about other experimental procedures are available elsewhere.[88,89]

3.1 Vacuum Systems

Both XPS and AES require vacuums in the 10^{-3} Pa region (1 Pa is 1 N m^{-2} and is approximately 7.5×10^{-3} torr). With pressures above this value, a significant number of the electrons that are to be analyzed will undergo collisions with the residual gas in the system and will not be detected. Now, almost all commercially available or component-built systems are capable of achieving pressures in the ultra high vacuum (UHV) regime (about 10^{-8}–10^{-9} Pa). It should be noted that it will take several hours for a surface to be covered statistically (assuming that every molecule sticks to the surface) at these pressures from background gases in the analysis chamber. This means that once a clean surface has been produced, it should be possible under normal operating conditions to characterize completely most materials by either or both XPS and AES, without significant contamination from background gases in the vacuum system. Not all combinations of clean surfaces and gases will have a sticking coefficient of unity (i.e. upon each collision with a surface a gas molecule or atom will be adsorbed); this reduces the sticking coefficient as a potential problem in many analysis situations. A handy rule-of-thumb is that a surface will be covered in one second at a pressure of 1×10^{-6} torr if each molecule or atom has a sticking coefficient of one; this quantity is called a Langmuir. The time for a surface to be statistically covered will depend on the molecular or atomic weight of the species present in the vacuum chamber.

The achievement of UHV pressures can be accomplished in many ways. Consideration has to be given to the materials and techniques used in both the construction of a vacuum chamber and the components that are employed for analytical purposes. Fortunately, the technology to produce UHV pressures has been available for decades. Thus, only routine consideration now has to be given to such systems. Detailed information about vacuum systems is available elsewhere.[91]

Surface analytical instruments are almost entirely metal in construction. The metals used must be compatible with the requirements of obtaining UHV pressures, such as a low outgassing rate per unit area, the ability to be heated above room temperature (and in some uses for certain metals to 1000 °C or more), and the capability of

withstanding temperatures near that of liquid nitrogen (78 K), and ease of construction. Stainless steel is used for the main body of most vacuum chambers, as it fulfills these requirements. Some metals, such as brass, have an appreciable vapor pressure in the UHV regime, and therefore cannot be employed for such uses. Mumetal™ can be used also for the main body of a UHV chamber. This is used to reduce magnetic fields in the chamber. With most stainless steel systems some parts are shielded with Mumetal™. This will increase the internal surface area of the system. In addition, the flanges on the various ports of a UHV system are constructed of stainless steel. On these ports are attached the various devices required for a surface analysis system. The flanges, which are almost always circular, have a variety of trade names such as Con-Flat™, Vac-U-Flange™, etc. A circular recess with a knife edge is used to seal flanges with a gasket. Usually the gasket is made of copper, but other soft metals can be used. However, copper is the usual material of choice, as it is relatively inexpensive compared to other soft metals such as gold, and can withstand higher temperatures than indium, for example. For certain special flanges, a soft copper wire is used as the gasket. These special flanges are either rectangular or have diameters larger than the standard sizes. The flanges are held in position with a series of bolts. Almost all of the circular flanges are made in standard sizes and designs. In the USA, the flange size is referred to by the outside diameter; the range of sizes is from 1.33 to 10 in. In areas where the metric system is employed, the size is also determined by the outside diameter. However, in practice the size of the flanges is such that there is almost complete compatibility between flanges manufactured in the USA and those produced using the metric system. This usually permits attaching flanges from different manufacturers to each other. Thus, easy modification to existing systems usually can be made. However, there is a tendency for some manufacturers of surface analysis systems to have only a minimum of extra flanges on their chambers. This reduces the capability of making needed alterations or adding accessories to existing systems when experimental circumstances demand change.

Certain nonmetallic materials are also found in a UHV surface analysis chamber. Electrical feedthroughs (usually on flanges from the outside of the chamber to the interior of the system) require ceramic insulators. Ceramic materials are also used as insulators for wires inside the vacuum chamber, especially when there is the chance of shorting electrical leads. The ability to have liquid cooling for certain components requires special feedthroughs and careful construction. Windows are almost always present on surface analysis systems, so that the specimens can be observed visually and for ease in sample introduction and manipulation. Often small amounts of polymeric material are present in UHV systems. Their most common use is for gaskets on certain valves (e.g. Viton™). This material can be heated to about 150 °C without significant outgassing, if it is not under compression. The surface area that these gaskets expose to the vacuum system usually is relatively small compared to the total area in such systems. In addition, polymeric materials such as poly(tetrafluoroethylene) (PTFE) are sometimes used to insulate wires inside a UHV system. Such usage should be kept to a minimum, because these materials will have a high outgassing rate during bakeout (discussed below) of the system. Ceramic tubing or beads are also employed for insulating wires in vacuum systems.

Several different types of vacuum pumps are capable of producing the necessary vacuum for XPS and/or AES spectrometers. Each pumping system has advantages and disadvantages. In many systems combinations of pumps are used either when operating in the UHV environment or in pumping the system down from atmospheric pressure. When it is necessary to bring a vacuum chamber up to atmospheric pressure, a nonreactive, dry gas such as nitrogen is admitted to the analysis system. (If available, the boil-off from liquid nitrogen is a very good source.) This gas should also be free of contaminants such as oil, particulates, etc. For this reason and the possibility of having water present, compressed air should not be used.

Ion pumps are used quite extensively for UHV systems with surface analysis capabilities. Electrons produced from cold cathodes (usually titanium) ionize gas molecules. The cathodes and anodes are in a strong magnetic field that increases the electron path length. The positive ions then strike the cathode; this releases neutral Ti atoms that coat the surrounding surfaces. These fresh Ti atoms react with all gases except the inert gases, thus producing the pumping mechanism. Inert gases are buried under succeeding layers of Ti, and the efficiency for pumping these gases is less than with more reactive species. In normal operation the potential between the cathode and anodes ranges from 5000 to 7000 V and currents of a few microamperes. These pumps have a long lifetime, especially when operated in the UHV regime, as the amount of titanium consumed is proportional to the pressure. Common modes of failure with ion pumps are shorting between the cathode and anode assemblies by small flakes of Ti or the build-up of a metal film on internal insulators. Correcting these problems is usually not expensive in terms of replacement parts, but can be difficult because getting to the pump elements usually requires an almost complete dismantling of the spectrometer. If the exposure to inert gases is large, then small amounts of these gases are observed for extended periods even after the source has been removed. In addition, if the pressure in the system is raised, as for sputtering (discussed later) or to react a surface with a

gas, ion pumps will take a long time to recover to the base pressure. The pumping speed for surface analysis systems that use ion pumps is usually about $200\,l\,s^{-1}$, but systems with higher pumping speeds are available. Smaller ion pumps are sometimes used to evacuate special regions where the conduction to the main pump of a system is poor. With well-designed ion pumps, the magnetic field does not affect the paths of the ejected photoelectrons or Auger electrons.

Often built into the main ion pump is a separate Ti sublimation pump. This type of pump operates by the deposition of active layers of Ti onto special metal pieces within the vacuum system. These layers are produced by heating Ti filaments for a short time at a low voltage and high current. Until the surface is covered by gas molecules, this type of pump is very effective for removing active gases. In many systems a Ti sublimation filament is set to run with a timer for unattended operation. Titanium sublimation, often at short time intervals, is used during the pump-down of a system when the system pressure is in the 10^{-3} Pa region after it has been taken up to atmospheric pressure. Also, this type of pump is used when sputtering with an inert gas if a turbomolecular pump (discussed below) is not available. In this case, the ion pump usually is turned off. The effective pumping speed of the Ti sublimation pumps can be increased if the surface on which the Ti is deposited is cooled with liquid nitrogen.

An ion pump cannot be operated if the pressure is greater than about 10^{-2} Pa. The time for which these pumps can be used at this pressure should be brief, as extended operation at these pressures will drastically shorten the pump's lifetime. Evacuation of a chamber from atmospheric pressure to a point at which the ion pump can be turned on has to be done with a roughing pump. One method is the use of a sorption pump, a relatively small, valved vessel that contains molecular sieves. When the sieves are cooled with liquid nitrogen, they can adsorb a large amount of gas relative to their volume due to their large specific area. Sorption pumps can achieve pressures in the 10^{-3} Pa region, but in normal usage the pressure is one to two orders of magnitude higher. If the volume of the system that needs to be evacuated is relatively large, then more than one sorption pump may be required to bring the pressure down to a level at which an ion pump may be started. When the liquid nitrogen is removed from the sorption pump container, the trapped gas will vent to the atmosphere with a simple pressure relief valve or stopper assembly.

The use of a rotary mechanical pump for the initial pump-down of a system is not recommended because the possibility of oil contamination to a UHV system is too large. If such an event happens, repairs to the surface analysis system would be very time-consuming and expensive. An oil trap in the vacuum foreline will reduce this risk, but there are other pumping methods that can achieve the results (i.e. sorption pumps) at about the same cost.

Turbomolecular pumps can be used for surface analysis chambers. These pumps are somewhat similar to turbine engines. A rotating set of blades compresses gas molecules against a set of stationary blades. The gas is then forced to a following set of rotating and stationary blades. This process is repeated several times until the gas is expelled from the turbine area. When operating at full speed the turbines rotate above 20 000 rpm. In order to achieve UHV pressures, the turbomolecular pump is backed with a mechanical pump. For protection of the vacuum system from the oil of the mechanical pump, an oil trap should be installed between the mechanical and turbomolecular pumps. This is especially important in case of a power failure. Some turbomolecular pumps are capable of evacuating a chamber without the use of a separate roughing pump.

Turbomolecular pumps are capable of pumping all gases. However, the pumping speed of lighter gases, such as H_2 and He, is reduced compared to that for higher-molecular-weight species. This type of pump is capable of handling gas loads that are much greater than those of ion pumps. This is especially useful when sputtering or exposing a surface to a gas or gases in a controlled fashion. Cooling of turbomolecular pumps is achieved either with water or air. Care must be taken with these pumps to eliminate vibrations in the analysis chamber. This is especially important when analyzing small areas on a sample. Also, with some pumps, measures may be needed, such as baffles or enclosing the pump, to reduce the high-pitched noise that some of these pumps produce. Turbomolecular pumps require service more often than ion pumps, given equal operation time, and they are available in many sizes. Some systems will have both ion pumps and turbomolecular pumps, so as to take advantage of the strengths of each pump design.

Cryopumps have been employed occasionally in surface analysis chambers, but their use has not been as great as either ion pumps or turbomolecular pumps. These pumps utilize a high-specific-area adsorbent material, usually activated carbon, that has been cooled to a temperature near that of liquid helium. These pumps are vibration free, except for a compressor needed for the helium. This compressor can be away from the main vacuum chamber. Cryopumps have the capacity to adsorb large quantities of gas, e.g. 1000 liters at STP (standard temperature and pressure), and can have very high pumping speeds for most gases. However, when the gas load capacity has been reached, the pump must be warmed to release the adsorbed gas. Provision for this is made so that the evacuated chamber does not have to be exposed to the

released gases via proper valving. By careful operation of the system, and the use of an appropriate roughing pump, the time between warming cycles for a cryopump can be quite long. The ultimate pressure that can be achieved with a cryopumped system is in the 10^{-7}–10^{-8} Pa region.

The venerable oil diffusion pump can achieve the vacuum required for a surface analysis system. However, there are potential problems in the event of failure in the cooling or traps with this type of pump that makes their use questionable, such as pump oil diffusing throughout the vacuum system. Thus, oil diffusion pumps are rarely, if ever, used with surface analysis systems.

Each of these pumps has its own set of strengths and weaknesses. Some suppliers offer a choice of a vacuum pumps for their instruments, whereas others have a fixed design that does not allow any alternatives. For these latter systems, it is usually possible to have modifications made (most easily at the time of construction) to allow for an additional pump. With certain types of applications this can be very important. The cost between the various types of pumping systems is not large. This is especially true when considering the overall cost of a surface analysis system.

Most surface analysis systems provide for heating the various chambers to temperatures of about 150–200 °C. This baking-out of the system results in attaining the base pressure of the system much more quickly than if the system is not heated after it has been brought up to atmospheric pressure. Water vapor is the most usual contaminant in a vacuum system that impedes the achievement of the base pressure. This is true even when a dry gas has been used to vent the system, because there will be a very small residual amount of water present due to some atmospheric gases diffusing into the chamber. Sometimes a vacuum chamber will never reach the lowest pressure that it can attain without being baked out. Most systems are provided with either ovens or insulated blankets that enclose the system. These covers sometimes contain heating elements. In addition, there may be heating elements in the base of the instruments. When baking out a system, certain heat-sensitive components have to be removed, such as micrometer drives, certain electrical connections, etc. Heating tapes that are wrapped around the vacuum chambers can be used. However, as the thermal conductivity of stainless steel is relatively poor for a metal, this method may result in nonuniform heating of the vacuum chamber. This is because it is difficult to wrap a system uniformly with heating tapes. Another approach to baking out a system is the use of a quartz heater placed permanently in the vacuum chamber. This method of heating is not as effective as the use of an oven or heating blankets, but it reduces the need for the removal or protection of components that are heat sensitive and maybe easier and quicker.

3.2 Sample Introduction and Handling

Early XPS and AES spectrometers required that most of the vacuum system had to be brought up to atmospheric pressure in order to change samples. Usually the specimens were placed on a carousel or a probe device. With modifications to the carousel it was often possible to handle an irregularly shaped and/or bulky sample. This arrangement also allowed the introduction of several specimens at one time (provided that they were not very bulky). However, much time was lost in re-establishing adequate vacuum conditions for surface analysis on a routine basis. This method is sometimes still used when samples require heating and/or cooling beyond the capability of specimen holders designed for routine use. With some low-cost instruments designed for very routine studies, this approach of venting the main analysis chamber is still followed. However, the design of such instruments should allow only a relatively small volume to be evacuated, and thus the pumpdown time will not be large. Also, the routine operational pressure levels are higher than in most other instruments, and this poses problems with some specimens.

In place of this approach, many manufacturers use a prechamber for the placement of samples onto a holder or holders. This prechamber, often called a loadlock, generally has a small volume that can be evacuated quickly from atmospheric pressure to about 10^{-4} Pa. A small turbomolecular pump is usually appropriate for such systems. Then a gate valve is opened, and the sample or samples are transferred into the main chamber. The pressure in the main chamber usually should not rise by more than 1–2 orders of magnitude, and it will then fall to a value close to the original pressure in a matter of minutes. (Note that this presumes that the sample or samples do not outgas excessively.) Also, it is possible to design a glove-box arrangement around the prechamber such that a sample does not have to be exposed to the atmosphere. Other types of sample treatment are possible, such as plasma cleaning and fracturing. To use these types of approaches for specimen treatment successfully, the pressure level in the prechamber should be lower than in a typical introduction system, with relatively short sample introduction times. Sometimes extended pumpdown times will be sufficient to lower the pressure in the loadlock. Another way to overcome this problem is to have a second UHV chamber between the surface analysis system and the atmospheric pressure introduction assembly. Sometimes extremely surface sensitive materials are prepared in this second UHV chamber. A further approach is to have a

chamber and valve assembly that can be attached to the prechamber after sample preparation under controlled conditions elsewhere, such as a glove box. However, these methods increase the complexity and cost of the overall system, but are required for certain types of investigations. A sample cooling attachment is useful for some samples, for instance those that outgas excessively or may change markedly even at room temperature. This requires a sample holder in the main chamber that can also be cooled.

There are several different mechanisms available for the transfer of a sample holder assembly into a surface analysis chamber. One approach is a rack and pinion drive or a friction drive in the evacuated prechamber. The sample holder is placed on the end of the drive where it can be positioned in the main chamber specimen holder via simple rotation of an exterior mechanical feedthrough. Some systems use a magnet-controlled rod inside the introduction system, that in turn is attached to the sample holder. Another design for the transfer assembly involves a rod sealed with elastomeric gaskets. The sample holder is placed on the end of the rod. The transfer is accomplished by manually moving the rod into the main analysis chamber and then attaching the sample holder to the specimen mount. These gaskets require replacement on a regular basis, as this design requires a very tight fit between the rod and the gaskets (otherwise, the pressure in the main chamber will rise to unacceptable levels over time). Other specialized methods, including pneumatically controlled sample holders, have been utilized on occasions. Each of these designs has its own set of advantages. Many transfer systems are highly automated and interlocked. This reduces or eliminates the possibility of operator error, such as an unintentional exposure of the main chamber to atmospheric pressure. This feature is particularly useful if a system has many operators that do not use the equipment regularly.

Many of the regularly available sample holders only allow specimens of limited size, i.e. 1–2 cm on a side and less than 1 cm thick. In addition, the sample entry port of the analysis chamber may not allow the passage of anything much larger than the standard sample holder. Unfortunately, many real-world samples cannot easily be placed on such holders, due to their size or shape. Sometimes the specimens cannot be cut to dimensions that would allow them to be placed on such holders. However, sample holders can be designed for bigger samples, and larger sample entry ports can be placed on most surface analysis systems at the time of system construction. Some systems have been designed so that larger samples can be studied, thus silicon wafers 15–20 cm in size can be introduced into certain surface analysis systems. In such systems a television-type monitor is used to locate regions of interest. These areas are stored in a computer, and then used to place the wafer in the same spots in the analysis chamber. Unless only a certain size of specimen will always be analyzed, it is prudent to have as much flexibility as possible in the sample introduction chamber and holders.

Samples are placed on a stage so as to allow optimal positioning for the energy analyzer and should also permit visual observation. This is especially important for irregularly shaped samples and/or for analyses where high spatial resolution is required. Positioning should be relatively straightforward and allow movement in all directions. Sometimes this is achieved with motorized micrometers under computer control. For high spatial resolutions XPS and scanning Auger microscopy (SAM; discussed later), a completely eucentric specimen stage (i.e. as the settings for the specimen holder in the apparatus are changed, e.g. the angle, the analysis area is not altered) should be employed, and the ability to rotate and tilt the sample is required.

In addition to having the ability to accommodate samples of irregular geometry and size, sample manipulators should possess several other features if any nonstandard analyses are required. For example, it is sometimes useful to be able to heat or cool samples relative to the ambient temperature of the analysis chamber. Heating can be accomplished by direct or indirect electron bombardment, resistive heating (if the sample is conductive), or with special heating elements on the specimen holder. Heating is useful in certain cases to allow minor contaminants to diffuse into the bulk of the sample under investigation or to come to the surface where sputtering is used for their removal. Temperatures of over 1000 °C can be achieved with certain heatable sample manipulators. Cooling is useful if a sample has a relatively high vapor pressure, and nonambient temperature processes can also be investigated. Cooling also is needed to bring the temperature down quickly from an elevated level for a heated specimen; this is needed for certain cleaning operations or after exposure to various gas-phase species followed by surface analysis. Cooling has also permitted the analysis of materials that are liquids at room temperature.[91] Cooling is usually done with liquid nitrogen, either through flexible tubing connected to a small reservoir, or with flexible braids in the vacuum chamber. With certain designs for a sample holder it is possible to achieve temperatures close to 20 K (using liquid helium). With liquid-nitrogen-cooled systems the temperatures are somewhat above 77 K (the boiling point of liquid nitrogen). Obviously, provisions have to be made to measure the temperature of the sample with a thermocouple. Thus, as wide a temperature range above and below ambient as possible for a sample holder is useful for many analysis situations. The design and construction of sample holders that allow for temperature ranges above and below

ambient has to be done carefully, otherwise vacuum problems will arise. This usually limits the size and shape of a sample. A manipulator should also have the capability of electrically isolating a sample. This is used to measure the current from an ion gun (discussed later) or an electron beam.

It is essential that the samples to be analyzed are not touched with bare hands or fingers. Disposable plastic gloves (powder free) or lint-free cloth gloves should be always used when handling anything that goes into a UHV system. If the samples have been handled with bare hands and fingers, the specimens need to be cleaned with a solvent or series of solvents. High-purity acetone, isopropyl alcohol and methanol often are used, and an ultrasonic cleaner will speed the cleaning process. Other special sample cleaning methods can also be used. In some instances particular solvents may not be used due to incompatibilities with components in the analysis system. Sample holders and tools used with the samples and sample holders, such as small screwdrivers, tweezers, etc., also have to be cleaned in a similar manner. A laminar flow hood is often used as a workstation for the placement of samples onto holders, and for other operations for equipment that will be placed into a UHV system. Some surface analysis systems are placed in clean rooms to reduce exposure to particulate contaminants. The basic rule to keep in mind is everything that can come into contact with anything that goes in the vacuum system should be as clean as possible. This will help to keep the base pressure in the analytical system as low as possible, and reduce or eliminate analyzing unintended changes to the surface region.

3.3 Excitation Sources

As was noted above, the electrons analyzed in XPS are produced by an X-ray source. For most samples Auger electrons are also observed when an X-ray source is used. However, with AES a beam of electrons is most often employed.

3.3.1 X-ray Sources

The most common X-rays used in XPS are the Kα lines from Mg or Al anodes. The anode material is usually a thin coating of the metal on a copper substrate. X-rays are produced when the anode is bombarded by electrons having a potential of about 10–15 kV with respect to the anode. In order to produce a sufficient X-ray flux, the emission current from the electron source filament usually is in the tens of milliamperes range. The filament usually will have a current of several amps. With most X-ray sources a thin window of Al or Be is used to prevent stray electrons from striking the sample and being analyzed. This will reduce the X-ray flux. Also, adequate cooling of the anode is required, because up to several hundred watts have to be dissipated. Usually a closed-loop pressurized water system is used for cooling purposes. Some systems have filters and ion-exchange beds in the water cooling lines. Inadequate cooling reduces the lifetime of the anode. This can lead to the appearance of ghost peaks due to copper (the anode material, i.e. the Mg or Al becomes too thin). Also, anodes can become oxidized and this sometimes yields oxygen X-rays.[92] The photon energy of the ghost lines is approximately 925 eV and 530 eV for copper and oxygen, respectively. Ghost peaks are usually found for only the most intense lines observed with the regular source. Ghost peaks will also be observed for those lines with binding energies less than the photon energy of the foreign material. Designs where the anode is at ground potential or where the electron filament is at ground potential have been produced. When the latter arrangement is used, adequate precautions have to be made so there is no electrical leakage from the cooling fluid to the rest of the system. Proper safety precautions in both the electronics and cooling systems have to be taken with any X-ray source.

There are several reasons why Mg and Al are the most common X-ray sources used in XPS. The X-ray line width for the combined K$\alpha_{1,2}$ lines of Mg and Al is about 0.7 and 0.8 eV, respectively. Therefore, the sources using these elements will not excessively broaden or distort most XPS peaks. The K$\alpha_{1,2}$ lines are actually two closely spaced transitions that are separated by approximately 0.25 and 0.4 eV for Mg and Al, respectively.[93] The Kα_1 line has twice the intensity of the Kα_2 transition. All of the elements that can be observed by XPS have photoelectron peaks that are produced with X-rays when either of these two elements is used as the anode. The kinetic energy of the XPS peaks from either of these sources is such that the IMFP is only about 1–2.5 nm for most elements.

There is not a clear-cut advantage between these anodes for XPS analyses. Whereas Mg has a somewhat narrower X-ray line, a somewhat higher power level can be used with Al. Also, Al sources usually have a longer lifetime. However, there are some systems where there is an overlap of an Auger transition with an XPS line when employing a particular anode. As the kinetic energy of an Auger transition is constant and the kinetic energy of an XPS peak is dependent on the energy of the X-ray source (see Equations (3) and (7)), changing anode materials will resolve the overlap that might be present and make identifying peaks easier in some instances. In order to overcome this potential problem without having a changeover that requires venting the vacuum system, anode designs are produced that have both Mg and Al sources in one assembly. (X-ray sources with four different anodes have been made, but they are not

used widely.) Each anode requires its own filament. In some systems both filaments can be used independently or together with the other anode.[94] The design has to be made so there will not be any cross-talk between the two anodes when only one filament is being employed.[95] For some assemblies, modifications have to be made to existing products to eliminate this problem. Sometimes both anodes have been made with the same material. This allows a higher X-ray power to be used for an analysis, but the ability to shift photoelectron peaks from Auger transitions is obviously lost.

Although Mg and Al have been the most widely used X-ray sources, other elements have been employed. Silicon (1739 eV), silver (2984 eV), and Ti (4510 eV) are available. In addition, there are other possible sources.[96] Some of these materials are difficult to produce as anodes, and others have heat-conduction problems. In addition, the X-ray line widths are somewhat larger with these elements than those for Mg or Al. As the kinetic energy of the electrons excited by these anodes is higher than those from Mg or Al for a given core level peak or the VB region, it is possible to probe more deeply in the bulk of the samples being analyzed with these anodes. This has proved useful in certain studies of polymers.[97,98] Using these other anodes also offers the possibility of exciting higher-energy Auger transitions than would be possible with Mg or Al anodes. This can extend the range of the use for the Auger parameter. It is often possible to use one of these other anode materials with the usual Mg or Al anodes in a dual-anode assembly.

Certain Auger lines that are above the primary X-ray energy have been observed, such as the Si $KL_{2,3}L_{2,3}$ Auger line at about 1610 eV kinetic energy, with nonmonochromatic Al and Mg sources from Bremsstrahlung radiation.[99] This approach can be useful in some situations.[100]

3.3.2 Monochromatic X-ray Sources

It was noted earlier that the usual Mg or Al X-ray sources have two major X-ray lines, i.e. $K\alpha_{1,2}$. In addition, the $K\alpha_{3,4}$, satellite peaks appear with intensities of about 10% of the main lines with photon energies about 10 eV higher than the $K\alpha_{1,2}$ excited lines. Thus, these small additional X-ray lines will produce their own X-ray induced photoelectron peaks with normal X-ray sources. There are other, less intense, satellite peaks that have been observed with sensitive spectrometers.[101]

In order to overcome this problem, X-ray monochromators for use in XPS spectrometers have been developed. The X-rays produced from the anode irradiate a bent Si crystal as shown in Figure 7. Due to the crystal lattice spacing, X-rays of only one given energy will be reflected, as given by the Bragg equation. Mg cannot

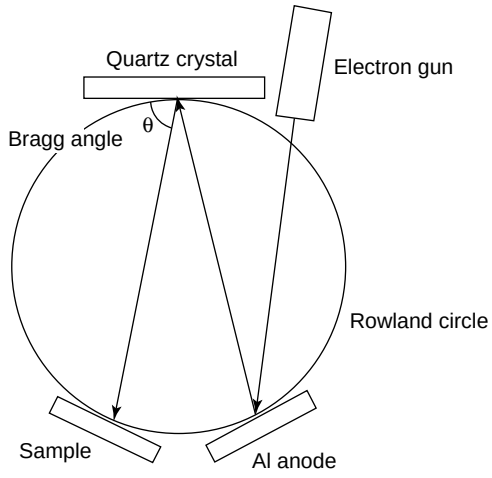

Figure 7 Diagram of the bent crystal arrangement for the production of monochromatic X-rays. (Reproduced with courtesy of Physical Electronics.)

be used with a monochromator, as there is no material available that has the proper crystal spacing. However, Al and elements that have a primary X-ray line at an energy greater than Al can be used. It also has been shown that a monochromator used for Al can be utilized also with only slight modifications for Ag as an anode material, because this latter element has a major X-ray line that is almost exactly double the energy of the Al $K\alpha_1$ line.[102]

There are both advantages and disadvantages with monochromatic X-ray sources. With only one component to the main X-ray line, namely the $K\alpha_1$ line for Al, it is possible to obtain sharper XPS peaks in many cases. This allows the resolution of the XPS line shape of an element in more than one elemental state to be obtained more easily than with nonmonochromatic X-ray sources. Vibrational structures have occasionally been found with polymers when an advanced monochromatic Al source has been used.[103–105] Also, damage to samples that are sensitive to X-ray radiation is reduced greatly with a monochromatic source, but cannot be eliminated for some materials, such as self-assembled monolayers on some metals.[106] Bremsstrahlung radiation that causes a higher background with the normal X-ray sources is absent with monochromatic sources. With certain monochromator designs, with a focused electron beam on the anode, it is possible to irradiate small areas of about 0.15 mm;[107] later improvements have reduced this size to about 0.01 mm longest dimension. Other designs have somewhat larger irradiation areas. Some advanced monochromators also employ several crystals to increase the X-ray intensity on a specimen.

The X-ray intensity that a sample receives with a monochromatic source usually is much less than that from a normal source. This requires longer data acquisition

times or advanced electron detection systems to obtain a spectrum in a reasonable amount of time with an adequate signal-to-noise (*S/N*) ratio. One highly advanced X-ray source uses a rotating anode.[108] When combined with a focused electron source, very high X-ray fluxes, and therefore large count rates, are produced. The rotating anode is used to reduce the effect of the electron beam being focused continuously on one spot of the anode. This is done to increase anode life and reduce cooling problems. However, this approach requires a very sophisticated mechanical assembly and, and as a result, is very costly.

Charging is observed much more often with monochromatic X-ray sources compared to nonmonochromatic sources. With normal X-ray sources the thin film in the assembly between the anode and samples releases secondary electrons because of the X-ray irradiation. This is often sufficient to compensate for the electrons leaving the surface.[109] However, it has been suggested that the electrons from the thin-film window could cause damage to sensitive samples.[107] The effect of the $K\alpha_{3,4}$ lines can be removed with fairly simple computer data-handling routines. These methods have to be used with analyses of the VB region line shapes.[69,70] In addition, deconvolution procedures have been suggested to remove the $K\alpha_2$ contribution to photopeaks.[110–112] Some of the results with these methods have been fairly good, but their usefulness in all cases is still questionable. With a focused X-ray beam it has been reported that some peaks shift locally in a non-uniform manner with certain insulators over the irradiated area.[113]

3.3.3 Electron Sources

Auger transitions are observed with X-ray excitation in many XPS spectra (and also by ion beams,[114]) but most AES spectra in surface analysis are produced by electron excitation. The source of the electrons is a heated filament that usually is at 2–10 kV below the sample, which is at ground potential. AES spectra can be obtained if the beam voltage is less than 2 keV, but many of the transitions used for routine analysis will be observed weakly or not at all. There are three types of electron sources available. Many Auger electron spectrometers are equipped with a tungsten filament source that is enclosed in a holder held at a potential of around 30–100 V less than the filament. Then electrons are drawn off through a small hole in the filament assembly. The electron beam is then focused with an Einsel lens arrangement. A simplified diagram of an Einsel lens for a charged particle gun is shown in Figure 8. This design can produce electron beams with a diameter of a few micrometers, and some general purpose systems use electron sources with this design.

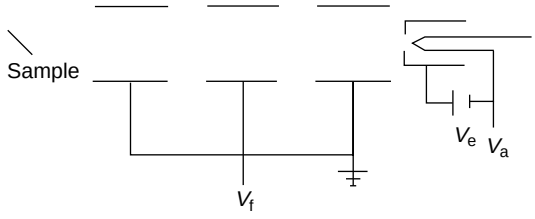

Figure 8 Cross-section of an electron gun with an Einsel lens focusing arrangement: V_f is the focus voltage, V_e is the emission voltage, and V_a is the acceleration voltage.

The electron beam can be deflected after it passes through the Einsel lens assembly by opposing electric fields; this is done with either plates or segmented cylindrical pieces. As electron beams are easily deflected, spatial resolution of the surface composition of materials by AES was developed soon after the first commercial AES spectrometers became available. Thus, the technique of SAM was developed. The addition of a secondary electron detector permitted images to be obtained using the same principles as those employed in scanning electron microscopy (SEM). Absorbed current images can also be acquired. Borrowing from SEM technology, electron sources with smaller beam diameters became available and this allowed the analysis of even smaller areas.[115] Eventually, LaB_6 electron sources were developed for use in SAM, offering beam diameters of less than 40 nm with beam voltages of 10 kV. This source has a higher brightness than tungsten filament-based systems. Even smaller beam diameters can be achieved when the beam voltage is raised beyond this value. However, the Auger yield decreases somewhat as the beam voltage increases. A rough rule is that the Auger electron yield is at its maximum when the beam voltage is about three times the transition energy. Thus, even using the highest-energy Auger transitions employed in surface analysis (around 2000–2500 eV), the signal intensity will be decreasing when the beam voltage is raised above 10 kV. As the range of Auger transitions used in surface analysis is from 25 to 2500 eV, there always is a compromise in choosing the beam voltage. Also, it is possible to obtain AES spectra even if the beam voltage is raised to 100 kV.[116] In addition to raising the beam voltage, electron beam diameters are decreased by lowering the beam current. The beam diameter is proportional approximately to the square of the beam current, i.e. if the beam current is lowered by a factor of four the beam diameter is halved.[115]

In the early 1990s Shockety barrier filaments were introduced for use with SAM. These filaments consist of a ZrO coating on a single crystal face of tungsten. These sources have two advantages over the LaB_6 filaments. First, useful data can be obtained with beam diameters of around 20 nm. Second, the beam current density is

higher compared to LaB_6. Spectra have been obtained with a beam diameter of less than 100 nm even with a beam voltage of 3 kV[117] This is not possible with a LaB_6 filament. The disadvantage with this source is the need to operate in the 10^{-8} Pa region. Even with a UHV system a special ion pump is placed near the Shockety source. Problems with vacuum systems are less with Auger systems than the typical SEM. Even smaller electron beam diameters can be achieved with field emission sources. However, with these sources there is flicker, i.e. the current varies with time. This obviously will give relative variations in the peak intensities beyond those from the inherent Auger process being investigated. Making estimates of the relative atomic amounts of the detected elements will have greater errors with a field emission source compared to other electron sources. Consequently, the usefulness of this type of electron beam source for AES is reduced greatly.

3.4 Electron Energy Analyzers

After a photoelectron or an Auger electron has been ejected from an atom and has left the solid surface, its characteristic kinetic energy has to be determined for surface analysis. There are only three main designs currently available, with some variations available. Each analyzer has its own set of strong and weak points, especially if both XPS and AES capabilities are present in the same instrument. This is because the requirements for XPS and AES differ when each technique is being performed at its maximum capability. The two major analyzer designs are the concentric hemispherical analyzer (CHA), sometimes called the spherical hemispherical analyzer (SHA), and the cylindrical mirror analyzer (CMA). A third design, the retarding field analyzer (RFA), is a four-grid low-energy electron diffraction (LEED) spectrometer equipped so that AES spectra can be obtained. (LEED is a technique that determines the atomic structure in the near-surface region and is used to study single crystals.) The performance of the RFA for AES is not as good as with either the CHA or the CMA. Many investigators who study single crystals use this type of analyzer for obtaining AES spectra, to check the surface cleaning procedures of their specimens. Also, the additional expense of adding this capability to a LEED system is small. RFA is not used for routine surface analysis, and is not discussed further. Detailed information on this analyzer design is available.[118]

The CHA has become the analyzer usually used in XPS; however, it also can be employed for AES. The three main parts to the CHA are a retarding and focusing input lens assembly, the hemispheres, and the electron detectors (Figure 9). A few analyzers do not have full hemispheres. Some of the electrons that leave the surface enter the

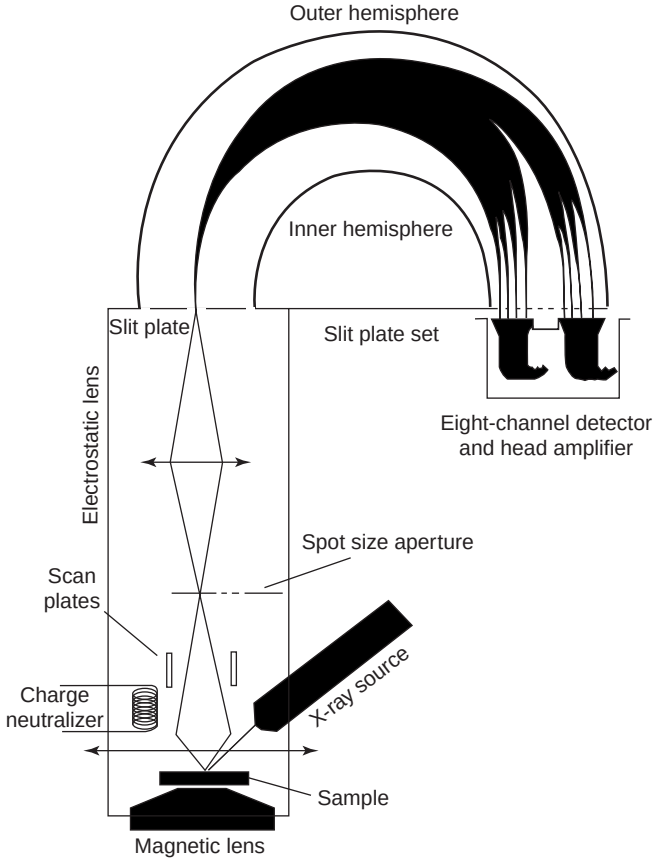

Figure 9 A schematic cross-section of a CHA is shown with eight-channel electron multipliers. After leaving the sample surface the electrons go through the input electrostatic lens assembly to the main part of the CHA and are detected at the exit plane of the analyzer. Also indicated are scan plates, an aperture, and a charge neutralizer. (This figure (modified slightly) is reproduced with courtesy of Kratos Analytical Instruments.)

input lens (some systems have a set of lenses) of the analyzer. The lens first focuses the electrons and retards their energy to a relatively low value. This permits much better energy resolution using a reasonably sized analyzer. The input lens also enables the acceptance angle of the analyzer to be changed with variable apertures. With some designs the input angle can be varied to less than 5°. The electrons then pass through an entrance slit into the main part of the analyzer, i.e. the hemispheres. The inner and outer hemispheres have a potential difference applied such that electrons with only small energy differences will reach the other end of the hemispheres. This potential difference then is ramped so that an energy spectrum of the ejected electrons can be recorded. At the end of the analyzer is the electron detection system. Electrons of slightly different energies fall across the exit plane of the CHA. With some electron detection systems this permits a small energy spread of electrons to be analyzed

simultaneously. This allows for much more rapid data collection, and permits the use of low-intensity excitation sources. Some manufacturers have placed a magnetic lens below the sample holder in XPS systems. The magnetic force lines steer more of the electrons emitted from the surface into the input lens assembly. This increases the signal level, but the magnetic lens cannot be used when making angular resolved measurements, except under certain special operating conditions. A discussion of the electron detection systems is given after CHA and CMA designs have been considered.

Two possible modes of operation with the CHA are possible. One method is known as fixed-analyzer transmission (FAT) or constant-analyzer transmission (CAT). The absolute resolution is independent of the kinetic energy of the incoming electrons in this operational mode, which permits easier quantification of electron spectroscopy data, especially with XPS. The other operational mode involves a fixed retardation ratio (FRR) or constant relative ratio (CRR). Peaks with low kinetic energy (usually found more in AES than XPS) are easier to detect with this method. This is due to the better S/N transmission characteristics in this energy region (up to about 200 eV).

A diagram of a CMA set-up is shown in Figure 10. This design is currently used only with new AES instruments, because the energy resolution is not as good as a CHA for XPS. With some CMAs the energy resolution can be varied, with typical values between 0.3% and 2%. This type of CMA only operates in the FRR mode. As an electron enters the region between the cylinders, it is deflected by a negative potential on the outer cylinder. The inner cylinder is held to ground potential, and the outer cylinder voltage is proportional to the kinetic energy of the electron being detected. Only those electrons that are in a solid-angle cone, usually $42 \pm 3^{\circ}$, with respect to the center axis of the cylinders, will be analyzed. Between the inner and outer cylinders are installed resistively coated ceramic pieces that have a high resistance (usually about 1 MΩ). These ceramic pieces are used so that there will be a uniform potential between the inner and outer cylinders. The effects of fringe fields at the ends of the cylinders are drastically reduced by these ceramic pieces. The sample surface has to be at the focal point of the analyzer, typically about 1 cm from the end of the CMA. This position is usually determined by analyzing the reflected peak at the energy of the incident electron beam. The largest peak intensity should be at this point. If the sample position is off by even a millimeter the signal intensity is reduced.[119] If the beam voltage being used for analysis purposes is higher than the analyzer's upper energy limit (usually around 3 keV), the beam voltage has to be lowered temporarily to determine the proper sample position.

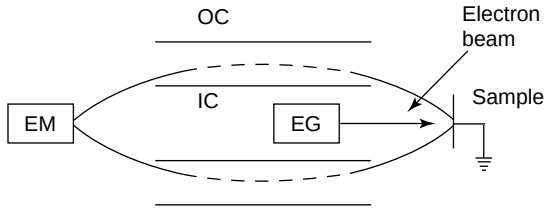

Figure 10 Cross-sectional diagram of a CMA with the electron gun coaxial; note that some CMAs have the electron gun independent of the analyzer. IC is the inner cylinder, OC is the outer cylinder, EG is the electron gun, and ED is the electron detector.

A variation of the CMA design was used for XPS and AES, starting in the mid-1970s.[120] This analyzer is known as the double-pass CMA and is still used in many laboratories. However, complete analysis systems using this analyzer are not available, but it can be obtained as a component. Although the transmission of the double-pass CMA was good, the energy resolution was poorer than with the CHA for XPS. With a modification to the double-pass CMA it is possible to obtain angular resolved spectra, but not to the small angles usually available with the CHA.[121] Also, the signal strength is greatly reduced in this operational mode. Other analyzer designs are available, but none of them is a part of a complete system. Instead, they are intended as replacements to existing analyzers or are used in custom built, usually research, UHV chambers. One analyzer is an adaptation of a design used for high-resolution electron energy loss spectroscopy (HREELS).[122] Another design is based on a large modification of the CMA, with the electron entrance and exit areas on the side of the main part of the analyzer.

A third analyzer has been designed to be added to a SEM for SAM. This analyzer is a modification of the CHA with less than a 180° deflection in a partial spherical portion design. Usually a SEM is not designed with the vacuum requirements needed for AES. However, modifications often can be made to many existing SEMs that will allow AES capability to be added, but the vacuum level will not be in the UHV region. Sometimes this approach may be less costly than using a regular SAM spectrometer, as the price competition in the SEM market is greater than that for SAMs.

After the electrons are energy analyzed they have to be detected and then the current has to be amplified. This is done with an electron multiplier. The two different types widely used in XPS and AES are the channel-electron multiplier and the position-sensitive detector. The channel-electron multiplier uses metal-coated pieces that have a high potential between them. As electrons strike the cone shaped entrance piece, other electrons are released and directed down the detector. The amplified

current is then measured and recorded after passing through the detector. These electron multipliers have a finite lifetime which can be shortened drastically if proper precautions are not taken. The usual failure mode is when too high a signal current is employed for a given multiplier operational voltage. In some systems several of these detectors are placed across the exit region of a CHA. This decreases the time required to obtain a spectrum. However, special procedures and equipment must be used so that gain is uniform for all the detectors.

The position-sensitive detector has the energy-filtered electrons striking a microchannel plate that consists of a series of holes, each having an individual electron multiplier. Then, the amplified signal strikes a resistive anode. At the corners of the anode, connectors are attached to a circuit that measures the amount of charge as a function of position on the anode. This information is then recorded. This type of detector is similar to that used with night vision devices and astronomical detectors.[123] The dynamic range of the position-sensitive detector is not as large as that of the channel-electron multiplier. However, with a CHA the energy range that can be analyzed simultaneously can be greater with a position-sensitive detector. Note that the pass energy (this is related to the amount of retardation incoming electrons experience in the input lens) also plays a role in the energy range of the detected electrons.

The choice of analyzer type is dictated by the expected use. For XPS the CHA, possibly with slight variations, is the analyzer that is employed in most complete commercial systems. This analyzer design is used also by some manufacturers for AES. The CMA is currently available only for dedicated AES systems. Some guidance for the choice of analyzer design can be gained from reported investigations.

Round-robin tests with many different surface analysis systems from various laboratories have been undertaken under the direction of the National Physical Laboratory in the UK as part of the VAMAS program.[124] These studies have centered on the observed signal intensity as a function of kinetic energy with standard samples. The spectra were ratioed with standard data in which instrumental effects were removed. It was found that the instrument function was dependent on the operational conditions, such as pass energy. In addition, the transmission factors sometimes varied markedly with different instruments from the same supplier. Sometimes the shape of the instrument function was similar with analyzers of the same type. It has to be noted that in these studies only instruments with either CHAs or CMAs were used. However, the procedures employed in the standardization studies could be applied to the new analyzer designs. A method to calibrate analyzers has been proposed.[125]

One very important factor is the electron multiplier response. In a study with several instruments it was observed that in some instances the channel-electron multiplier performance changed over time.[126] Any such changes in detector response would make comparisons with spectra taken at different times difficult unless the data were obtained with standard samples at the same time. It has been found that spectral distortion of low kinetic energy peaks can be reduced markedly with a biased screen placed before the electron multiplier.[127] A systematic study on any long-term behavioral differences has not been reported with position-sensitive detectors.

There have been some differences noted in the behavior of CMAs and CHAs when used for AES. The corrected intensity at low kinetic energies, i.e. for about 10–50 eV, decreases somewhat more for the CMA compared to the CHA. Internal scattering appears to be greater with some CMAs compared to CHAs.[128] This phenomenon is thought to be due to higher-energy electrons than those being energy analyzed hitting the outer cylinder or sphere and then being directed to the electron multiplier assembly. This results in a higher baseline signal, and therefore more noise in a spectrum. Field trimmers and the exit apertures have been identified as other sources for scattering with a CMA.[130] With an individual CMA this effect was very pronounced; it was suggested that extended sputtering might have been the cause of this observation.

3.5 Small Area Analysis

The analysis of small areas by XPS and AES has already been mentioned. The approaches used in XPS and AES for this type of analysis are very different. Spatial resolution is defined as the distance between two signal levels over a sharp feature, such as a knife-edge. The electron microscopy community uses the distance between 80% and 20% of the observed signal, but no firm definition is used by the electron spectroscopy community for elemental spatial resolution. However, when profiling (discussed later), depth resolution is usually taken as the distance between 84% and 16% of the signal change (this is two standard deviations).[11] The difference in spatial resolution between the two definitions is about 50%. The electron microscopy definition (based on a secondary electron image) is much less dependent on the material being analyzed. Several methods are used to analyze small areas by XPS, whereas with AES the strategy employed is to decrease the size of the electron beam. Also, it is necessary to distinguish between point analysis and imaging. Point analysis is investigating a single area of a small dimension with a suitable measurement approach or with a specially designed electron analyzer. These analyses can be either a survey type spectrum or

detecting the presence of selected elements. Imaging is analyzing the presence of designated elements or different chemical states of the same element over a larger area than an individual probe measurement. This can be a parallel measurement (effectively an elemental snapshot) or a sequential set of point analyses.

The analysis of areas of less than a few square millimeters commenced in the early 1980s with XPS, using a small-diameter electron beam to excite X-rays.[107] Also, the X-ray shape can be made to be nearly rectangular to various dimensions. The area of irradiation from the X-ray source is found by a phosphor or an X-ray-sensitive film that changes color and is observed with a microscope or a television monitor display. The photoelectron signal strength is very sensitive to the distance between the specimen and the input lens of the electron energy analyzer similar to that when using a CMA, as indicated above. This requires a method to measure quickly a photoelectron signal from an element thought to be present on the sample surface. In addition, that element should give a strong signal. In usual practice the focal point of the microscope or television camera can be used routinely, with only occasional checks being required by measuring a photoelectron peak intensity.

Somewhat later another method was developed for small area analysis by XPS with the placement of apertures in the input lens of CHAs. This technique initially permitted the analysis of areas of about 0.25 mm in diameter. Further improvements in X-ray sources and detectors have reduced the analysis size to about 20 μm in diameter. The area of analysis is identified by different strategies. In one approach the signal from a metal dot (often gold) is maximized, and this location is calibrated with a television monitor. Note that the metal dot has to have dimensions that are close to that of the analyzer spatial resolution. The metals used have a high relative XPS sensitivity. With systems that have an electron gun for AES, the secondary electron image is employed to locate the analysis area.

XPS imaging has been developed and several different approaches have been used. Before elemental images are recorded, the usual experimental procedure is to record a survey spectrum that determines the elements and possibly different chemical states to be imaged. One method has been to use an X-ray line source to obtain a spectrum over several electron volts and then move the sample.[108] This approach requires the use of a position-sensitive electron multiplier. The spatial resolution obtained with this technique is limited to that of the X-ray source and possible errors in moving the sample stage. Also, it is a time-consuming procedure. Having a spectrum over an energy range can detect changes in the chemical state of the element being analyzed on a routine basis.

A more widely used technique involves refinements in the input lens of CHAs. At the entrance to the input lens there are deflection plates (see Figure 9) that only allow electrons emitted from a small area to enter the assembly.[131,132] To use this approach an area larger than the imaged area has to be irradiated with X-rays. The electrons to be imaged are then magnified and passed through an aperture where deceleration occurs before entrance into the main part of the CHA. The electrons are then detected with either a series of channel electron multipliers or a position-sensitive detector. The signal intensity is then measured as the peak energy maximum minus a baseline value a few electron volts away from the peak. An image is produced by rastering the deflection plates over the field of view of the analyzer. With this approach the spatial resolution is of the order of 30 μm. It is also possible to obtain an entire survey spectrum or detailed elemental spectra at one setting of the deflection plates with the same spatial resolution as the image.

A modification to this approach has been developed and is illustrated in Figure 11 in a change to the design of a CHA. In this approach, imaged electrons enter the spheres of the CHA. However, after travelling a short distance in the main part of the hemispheres, they enter another analyzer called a spherical mirror analyzer that is on top of the main CHA. The electrons are then directed back to a point near the exit area of the main CHA. Finally, the imaged electrons are focused on to a position-sensitive electron multiplier that produces an image that is usually recorded with a video monitor. Then the image data is digitized. Spatial resolution of less than 5 μm have been achieved with this approach. However, it must be remembered that this is an image and other methods have to be used for analyzing a small region. With this analyzer it is possible to perform regular XPS analyses, as a series of channel-electron multipliers is placed alongside the position-sensitive detector.

An entirely different technique is used by one manufacturer.[133,134] Again, an area larger than that of interest is irradiated by X-rays. The electrons from the entire acceptance area of the analyzer pass into the input lens of a CHA. In two stages the electron image is magnified by about a factor of 16. Then the image is focused onto the entrance area of the main part of the CHA. This method makes use of the fact that if the distance from the image plane is the same as focal length of the lens, all of the electrons have a common intersection point on the center line of the lens. This process has been described in terms of a Fourier transform, in this case converting spatial information into angular information. The electrons of one energy then are focused at the exit area of the CHA. The electrons then pass through a hole where they enter another lens assembly where the image is reconstituted in the reverse fashion, i.e. angular information is

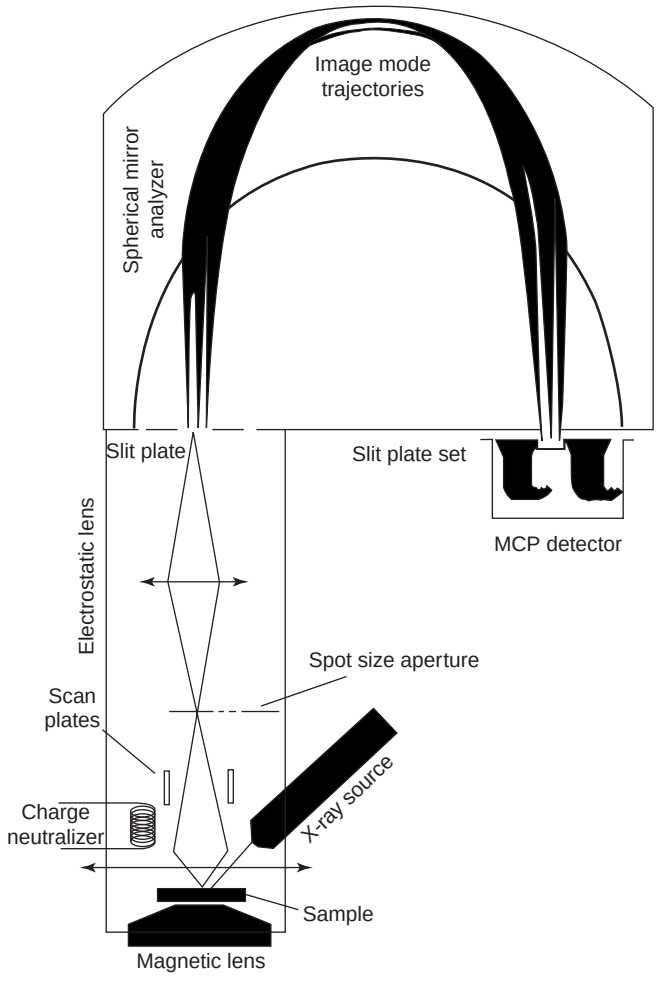

Figure 11 The cross-sectional diagram for an imaging CHA used for XPS. The analyzer uses a position-sensitive detector for images and channel-electron multipliers for regular XPS analyses. MCP, multichannel plate. (This figure (modified slightly) is reproduced with courtesy of Kratos Analytical Instruments.)

transformed into spatial information. The electrons are detected with a position-sensitive detector in a similar fashion to that mentioned above. Spatial resolutions of less than 5 μm have been achieved with this approach. With this analyzer it is possible to perform regular XPS analyses, as a series of channel-electron multipliers is placed at the exit plane of the CHA that is alongside the final electron lens assembly.

With the last three approaches discussed it is possible to obtain line scan data after the image has been recorded. With digital recording of the data a line can be connected between any two pixels. Those systems that use an X-ray line source, of course, produce a line scan directly.

Another method has been developed to obtain both high spatial resolution and images in XPS. The focused X-ray source using a monochromator has been designed so that the electron beam is rastered across the anode. This allows an image to be produced, and the spatial resolution is claimed to be about 10 μm. Also, the input lens region of the CHA is designed specifically to accept the spatially generated electrons. The area of analysis can be identified in two ways. First, a video display is used to identify the locations of interest while the sample is in the introduction system. Then, the coordinates of the locations of interest are stored in the system computer. The specimen then is transferred to the main vacuum chamber, and the sample is positioned for analysis. A second method can be employed also. A secondary electron image from the X-ray source is used to locate desired locations on the sample. The instrument also has a very unique configuration, but how adaptable the design is for a wide range of specimens is questionable, as it was designed primarily for large silicon wafers. Also, the ability to modify the instrument appears to be very limited. A review of this topic is available.[135]

With AES the spatial resolution is primarily a function of the electron beam diameter. However, there are other factors that have to be considered. When an electron beam bombards a solid surface, some of the electrons will be deflected from the point where they strike the surface. In addition, other electrons that have penetrated into the bulk can be scattered back toward the surface region; these electrons are called backscattered electrons. If these electrons have sufficient energy they can create Auger electrons that will be detected. Also, X-rays created by the electron bombardment will produce some detected Auger electrons. This effect becomes more significant at higher electron beam voltages, such as 10 keV or above. These effects will cause a broad, fairly weak background away from the point where the electron beam strikes the surface. Calculations have shown that this is not a large effect, but it is more important with higher-Z elements.[136]

Topography also can degrade spatial resolution in SAM, and this effect is illustrated in Figure 12. Here, as the electron beam strikes the edge of a feature, some electrons will exit the material and further bombard the surface away from the point of intersection. Again, if these electrons have sufficient energy, Auger electrons will be produced, and this will degrade the spatial resolution.[137] This has also been observed with small Al and Al_2O_3 particles on Si.[138]

3.6 Data Collection and Handling

The electron multiplier signal must be further processed before useful information concerning the surface composition can be obtained. With the earlier spectrometers used for XPS and AES the signals were detected either by pulse counting with a rate meter or by analog methods, depending on the current levels. The output then

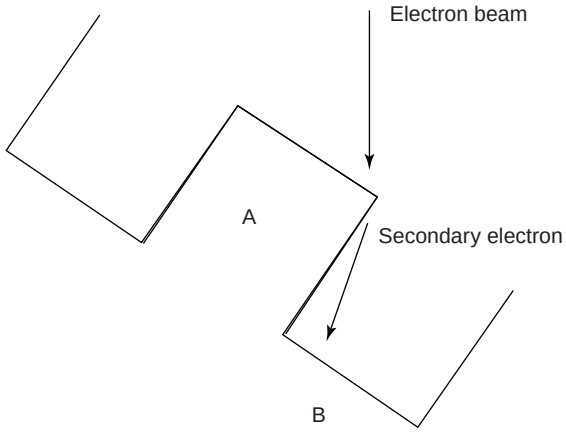

Figure 12 A representation of the effect that topography can have to reduce spatial resolution in AES. The A region is elevated slightly with respect to the B region.

was displayed with an ink on paper recorder. However, in most systems these approaches gave way to small computers for data collection and processing. As with practically all current advanced scientific equipment, a computer is now used almost exclusively for system operation in AES and XPS. With many advanced instruments the computer hardware and software systems are very sophisticated.

The primary computer use is in the operation of the basic electronics, i.e. the electron energy analyzer control and simple data collection. The computer provides a voltage via a digital to analog (D/A) converter to operate the ramp electronics for the electron energy analyzer. The voltage produced by the D/A usually will have a small range, such as 0–10 V. This voltage is then amplified by the analyzer control electronics to the level required by the energy analyzer. The D/A has to have sufficient resolution so that the energy resolution of the electron analyzer is not compromised. For example, a 16-bit D/A can produce a resolution of one part in 65 563 over its output range. If it is desired to analyze the kinetic energy of electrons over the range 0–3200 eV, then the smallest energy difference that could be sampled would be just under 0.05 eV. This, of course, assumes that there are no other sources of error in the electronics (e.g. drift) and analyzer systems. It has been determined that an error in just one D/A channel can affect a peak's energy resolution and shape.[139] The kinetic energy range 0–3200 eV is typical of many analyzers, as it covers the span for most AES and XPS transitions used in surface analysis. Note also that an exact difference between points of 0.10 eV separation, for example, is not possible. The situation is much more complex if the photoelectrons or Auger electrons are retarded before entering the main part of the energy analyzer. However, this does permit even increments, such as 0.1 eV. This provides some analyzers with the capability of measuring electron kinetic energies to values beyond 4000 eV. The computer will also provide the voltage dwell time at each step.

The signal from the electron multiplier then is stored in computer memory or a disk drive. With systems that employ a lock-in amplifier (used to display derivative AES data) or a rate meter, the output voltage has to be fed into an analog to digital (A/D) converter. This device converts an analog signal, such as a time-varying continuous voltage, into digital information that is stored by the computer. At each energy interval, a separate electron intensity will be recorded. The computer can repeat energy sweeps and sum them to improve the S/N of the spectrum. It is also possible to record spectra over narrow energy regions sequentially or serially. The former approach can be used if it is suspected that sample degradation due to the X-ray source or electron beam is damaging the specimen under investigation. The low levels encountered when pulse-counting techniques are used can be treated in a more direct fashion. Each individual pulse can be detected by counting circuitry, which then transfers the total count after a preset time to the computer. Such an approach must be used with position-sensitive detectors or analyzers that have several channel-electron multipliers because multiple signals are being detected and processed simultaneously. With signal levels up to several million counts per second, pulse counting can be employed. With count rates above this value the counting circuits are not able to function properly, as one pulse cannot be processed before the next one arrives. When this happens the signal has to be converted into the frequency domain for signal processing.[140]

Many of the operational settings on AES and XPS spectrometers can also be placed under computer control, such as resolution control of the analyzer and the various potentials needed for the input lenses on a CHA, the X-ray power in XPS, the electron gun settings in AES, and the voltages needed for the electron multipliers. The control of accessories, such as ion gun operation, can be placed under computer command. This allows for either continuous (if not prohibited due to other experimental conditions) or interrupted depth profiles. With certain systems the sample manipulator can be under computer control so that different samples or locations on the same specimen can be analyzed in succession without operator intervention. Angularly resolved determinations also could be made in this manner. System checks can be monitored with a computer. In the event of a failure in some part of the instrument, system diagnostics may be incorporated in the operational software to help pinpoint the problem.

Sputtering can be controlled with a system computer. This is very useful for profiles and a good profile routine

will permit the ion gun timing to be varied as the experiment proceeds. Selected elemental regions can then be analyzed, displayed, and stored during a profile. This is especially useful when the composition is changing rapidly as a result of small sputtering times. This may then followed by regions where the composition varies little. In this situation there is no need to have small time intervals between sputter cycles, as can occur with a thin oxide layer on a metal or at the interface of layered materials.

Another approach to find elemental depth distributions is to create a slight gradient across a specimen. Then line scans using SAM are made for the elements of interest. Ball cratering is one of the methods used to produce the gradient.[141]

Computers can do much more than just collect a spectrum and then display and store it. The storage of necessary experimental information, such as system pressure, electron multiplier settings, operator, date, and some sample identification should be included in the recorded data. The ability to scan only those regions of interest and to enhance low-level signals or take spectra at high resolution compared to survey experiments are additional advantages. Also, there is a reduction in the time required for an operator to be involved directly with all of the data collection procedures. Extended or even overnight data collection can be made routine. With the continuing rapid increase in computer capability and reduced costs, even more power will be available in the future. Real time displays of the experimental results, even with some data reduction (e.g. approximate intensities from profiles or images), are usually available with most systems.

After the spectrum has been recorded and stored by the computer it has to be displayed in order to make use of the collected information. Spectra initially are shown on a monitor for visual inspection by an instrument operator. Permanent plots can be obtained on output devices which often have color capabilities to enhance the display and clarify the observations. Spectra can also be archived on magnetic tapes or magnetic disks. High-speed computer systems having adequate memory offer the ability to collect data while processing previously acquired results.

Advanced data reduction routines are available with most current XPS and AES instruments. In addition, there are several independent data handling programs that can be procured for XPS and AES, providing capabilities that the instrument programs lack.

The display of only a small region of a spectrum often is useful and this feature is found in most programs. Occasionally a point will be quite far removed from its neighbors due to a noise spike or some other temporary instrument malfunction, and its repositioning can be easily accomplished without jeopardizing the value of the experiment. The ability to overlay two or more spectra, or to subtract one spectrum from another for comparative purposes, is usually available. Adding a fixed quantity to each point in a spectrum has advantages for certain display situations. Mathematical smoothing of a spectrum is readily accomplished.[142–148] The derivative of a spectrum is often computed, usually for rapid visual examination of AES spectra, even if the data has been collected in the undifferentiated mode. It has also been suggested that the second derivative of XPS spectra can indicate the presence of more than one component in what appears to be the envelope of a single peak.[149] As noted earlier, procedures utilizing deconvolution methods have been suggested as an alternative to the use of a monochromatic X-ray source.[110–112] Many find the ability to annotate a spectrum before making a hard copy to be useful, especially for reports or presentations.

Most of the above procedures are displayed on the computer monitor. However, some data-handling procedures do have drawbacks. For example, if smoothing is done over too wide a region (i.e. the number of points chosen for the smoothing algorithm is about the same as the total number of points defining the peak), some minor features will probably be lost. Also, smoothing is for visual purposes only, as no additional information is inherently available after smoothing. There are several different ways in which a derivative can be obtained mathematically, and each method has its own set of advantages and disadvantages.[150] The exact correction of a data point due to a spike is arbitrary.

Semiquantitative analysis from XPS or AES spectra is a data reduction procedure available with most software provided with commercially available instruments. However, the use of these programs requires much understanding and care. The accuracy of some methods has been questioned, and it has been suggested that the results could sometimes have an accuracy of no better than 50%. As a simple example of the potential pitfalls with both XPS and AES, it has to be recalled that the IMFP of the ejected electrons depends in part on the kinetic energy of the transition being analyzed. Therefore, a different volume will be analyzed for each element in the surface region for many materials. However, in many instances even a relatively poor estimate of the surface composition is of use for many problems.

In XPS the usual method for quantitative analysis is to determine the area of each photoelectron peak. A much less used approach is to measure the peak height of selected transitions.[151] This latter method usually is done manually, but it can be used as a rough indication of the composition of the surface region. The database used by Wagner et al.[151] was built with instruments that are no longer available commercially. In order to use this approach with most spectrometers, a restandardization would be required. Also, this method would yield poor

results if there is more than one chemical state present for an element in the analysis volume. Usually the most intense peak of each element being analyzed is used in either procedure, unless there is a strong reason not to do so. Before the area of a peak can be determined, a baseline has to be established. There are several methods can be used. The background intensity is due to photoelectrons and Auger electrons from the bulk that have lost some energy due to inelastic collisions before going into the vacuum phase. As such they will have a broad energy distribution. The simplest procedure is to select a point on either side of the peak being analyzed and assume that a straight line (defined by these two points) will represent the baseline for the peak. It also is possible to use a few points on either side of the peak to determine the baseline. This method will work, at least on a level that is visually satisfying for many peaks. However, this process cannot be employed for many of the major peaks with the transition metals, as they have significant losses in the immediate energy region of the photoelectrons. Another popular baseline selection procedure is the so-called Shirley method.[152] This approach subtracts a proportional quantity of the signal at higher binding energies times the intensity at a point below the peak from the experimental data. This yields a baseline-removed spectrum with the same intensity at both end-points. Empirical methods such as parabolic functions on the high binding energy side of a line or a tail function are sometimes employed. Several other theoretical techniques have been used. Tougaard has investigated this problem theoretically and has developed relationships to model the background. He has also found an empirical equation that closely follows the more complex theoretical expressions and which is relatively easy to implement. Many commercial data analysis programs now include this empirical expression. However, proper use of this empirical method requires that the data have to be acquired at least 40–50 eV below the peak of interest. This topic has been reviewed in detail elsewhere.[153,154]

Of the several ways to determine the area of a peak, the simplest is to simply integrate the peaks under the curve after baseline subtraction. However, most approaches involve fitting the curve to a function using a minimization procedure, such as nonlinear least-squares or the Simplex approach. One of the simplest curves used is the Gaussian, which is defined by three separate parameters: peak height, width, and position of maximum intensity. This is a useful approximation with peaks that are fairly broad. However, if the spectrum has been taken with good energy resolution and the peak is inherently narrow, the Gaussian representation is not an adequate fit of the experimental curve. This is because the natural line shape has Lorentz components due to the X-ray source and the electron emission processes. The Lorentz line shape has the same three parameters as the Gaussian function, but it has broader wings. This is also found with some relatively broad peaks, which are better described with the Lorentz line shape. Some programs allow for the calculated curve to be represented by a combined Gaussian–Lorentzian function. The fraction for the Lorentzian or Gaussian contribution is fixed or can be a variable, depending on the program. These curve-fitting programs can take into account the presence of more than one component in a complex spectrum envelope. In addition, often it is possible to fix or limit the range of the energy separation between two peaks if this quantity is known, as with doublets. The same can be done for area ratios of doublets; for example, $p_{3/2}$ peaks have twice the intensity of $p_{1/2}$ peaks, but the baseline subtraction routines may introduce some errors that would make fitting the exact ratio difficult. Care should be used when using such constraints, because the fits should usually come close without restraints. The computed areas then are normalized with the use of sensitivity factors computed by Scofield.[155] These theoretical cross-sections have agreed well with experimental results in comparative studies.[156] The Scofield cross-sections also have coincided reasonably well with other calculations.[157] With many instruments the transmission factor and electron detector response as a function of kinetic energy for the analyzed electrons must also be taken into account in computing the relative areas under the photoelectron peaks. When the relative areas of all the curves for the detected elements have been determined, an elemental composition of the surface region can be calculated. From the results of curve fitting the full-width at half-maxima and peak heights are computed. These results can be used for comparative purposes with other studies.

Some shortcomings in the above methods can be encountered in commercial software. First, the choice of baseline model is often done with one of the simple approaches, and this has been suggested as a source for large errors in some cases.[158,159] Second, even a combined Gaussian–Lorentzian curve often cannot correctly depict some narrow XPS peaks. The Doniach–Sunjic line shape can better describe the asymmetry experienced with many metals, arising from electron–hole interactions.[160] However, this line shape is not usually included in curve-fitting programs, and empirical asymmetry equations often are employed to take this effect into account. With some high-resolution studies of polymers, peak asymmetry has been detected due to vibrational effects.[103–105] Third, angular effects have to be considered when employing the theoretical cross-sections. When s-level peaks are being analyzed, a correction due to the electron ejection angle is not

needed. However, when other energy levels are being analyzed (i.e. p, d, and f), appropriate corrections should be made.[161] This assumes that the surface is flat, which is rarely true. By the proper choice of angle between the sample surface and the energy analyzer, a correction is not required. Many instruments are constructed with this 'magic angle' at about 56°. Also, for most analyzers there is a finite angular width so that any correction will be an approximation. Quite often this effect is neglected with typical specimens. Also, consideration has to be taken of shake-up effects when trying to measure the relative intensity of an XPS peak. As can be seen from this discussion, the correct determination of the area of an XPS peak is complex, and often many of these factors are not included in fitting programs. Other factors, such as detector response, X-ray intensity, and elemental density within the specimen must also be considered. Some of these values will tend to cancel out when making relative measurements, which is the most common method used.

Unfortunately, much the same can be said regarding the use of AES spectra for quantitative analysis purposes. There are many different approaches that can be used for quantitative AES. Many of the currently available procedures included with most commercial software usually employ only the simplest approaches. These methods include the use of peak-to-peak heights or the negative excursion in a derivative spectrum, and the peak height of an N(E) (number of electrons that are detected at a given energy, i.e. on the horizontal axes of Figures 5 and 6) AES spectrum. Comparisons between the spectrum of interest and published data made under standard conditions (the most important experimental variable is the electron beam voltage using this method) are then made. This is sometimes done with analyzers, such as a CMA of the same basic design, but not with the specific model. It has to be remembered that these methods use only a few points in a spectrum. All of these simple methods suffer from the fact that changes in the chemical state between the standard used and the material under investigation may significantly change the response between the two different elemental environments. There are numerous examples where the peaks will be at different energies between two chemical states for the same element (see Figure 6). Also, such approaches cannot be used if the signal results from two elements that overlap, as with Ti and N.

More advanced procedures have been used to obtain quantitative AES information and some of them are available in commercial software. These include the use of linear least-squares,[162] principal component analysis,[163] and factor analysis.[164] This latter approach has uncovered unexpected components or sputter effects in depth profiles in some systems. Some of these more sophisticated methods have also been used in XPS.[165,166] With these procedures a much wider spectral range is used. Thus, because the statistics are better, results based on using these approaches should be better than those employing the more simple methods. However, greater effort is required in the use of these methods in terms of analyst time and experience. Also, the data quality has to be relatively good, or the results may not have much meaning.[167] In some cases, improved results are obtained when the effects due to primary and secondary electrons are considered.

Matrix effects, backscattering, Auger transition probability, and IMFP should be considered also.[168] In some instances these effects are not important. The ability to have user-supplied standards that are more nearly representative for the problem at hand reduces some of the most glaring problems in the use of any analytical procedure with AES. The use of AES and XPS for quantitative purposes has been reviewed thoroughly elsewhere.[169,170]

Most image data collection methods involve using two data points, i.e. a point at the top of an elemental peak and another on the baseline at least several electron volts away on the higher-kinetic-energy side. The difference is then calculated to give a peak intensity. Three points can also be used, with the third point being at a lower energy than the peak maximum. Then a line is calculated between the first and third points to give a baseline, and the intensity is computed from the peak maximum to the baseline at that kinetic energy. The reason for employing so few points is that images require a large amount of computer memory and calculations. For example, with a 256 by 256 point image, the number of peak maxima points alone will be over 65 000.

A wide variety of procedures is available to process images from either XPS or AES digital elemental maps and secondary electron micrographs. Many of these routines are included in system software, but specialized programs can be purchased that can extend this approach. As computer capability increases, even more sophisticated image analysis will be possible. Various filters are used to enhance the quality of the images. The choice of the particular filter depends on the type of features in a given image and operator choice. The use of color images, with each element having a different color, gives an easily visualized distribution picture. For example, this is useful to picture the distribution of a catalyst on a support, a contaminant on a silicon wafer, or the wear area on a lubricated surface. It is also possible to use color to see gradations in the intensity of an element over the imaged area. Occasionally, these processed images can mislead an analyst and other factors that contribute to the signal must be considered. This was shown in the work of Prutton and co-workers in a study using Auger images of overlayers on various substrates.[171] Different image analysis approaches can

yield inconsistent findings.[172] There have some attempts to quantify data from images, but the task is difficult, as reviewed elsewhere.[173] Data from line scans also can be treated in the same fashion as images. With specimens that have pattern samples with regular features, line scans will produce the same information in a much shorter time. This is especially true with SAM. With either approach weak signals will often be missed. Much of the original effort in image analysis came from work on the pictures obtained with the deep-space probes.

The semiquantitative data analysis procedures used in AES and XPS depth profiles usually employ the simpler methods, such as simple peak areas or peak heights. Some data collection routines, especially with AES, will collect only two or three points per element per sputter cycle. However, it is better to collect the data over an energy range that encompasses the elements being analyzed. This will allow more points to be used, even if a simple area determination is made. In addition, this permits some of the more detailed procedures, e.g. factor analysis, to be employed if the need arises.

There are a number of other factors that are important in the software that is available for the collection and further treatment of AES and XPS data. The most advanced systems have programs that are easier to use than their predecessors. This relieves the operator from needing a detailed knowledge about the program, and is useful if there are a number of occasional users. However, although this frees the operator from an extensive learning process, just to get started, if there are some subtle operational or computational effects that influence the data and subsequent analysis, they could be easily missed.

The original program source code is usually not provided. This prevents an individual from introducing errors into the operational program for an instrument, which could have disastrous effects. On the downside, improved methods of data collection and treatment cannot be introduced by an advanced analyst and require waiting for updates from the instrument supplier. Also, errors in the data-treatment programs cannot be uncovered and corrected.

The ability to transport the data away from the original program usually is provided. Thus, in some instances the ability to use procedures other than those supplied with an instrument is possible. However this usually requires either some programming effort by an operator, or software from a third-party supplier. A VAMAS standard format is now available for data transfer.[174] This allows further treatment with other programs in a specific manner suitable for a given laboratory, and could include spreadsheets, word processor programs, special plotting devices, databases, and third party data reduction programs.

4 ACCESSORIES

4.1 Ion Guns

Most surface analysis systems are purchased with an ion gun as an accessory, and it has two main uses. It operates by striking the surface with energetic ions (usually from an inert gas with energies of a few hundred to a few thousand electronvolts) in order to sputter, i.e. remove atoms from the surface. The first use is to clean a surface of the almost inevitable contamination layer present on a sample. With many analysis systems this can be done within a few seconds for most specimens. The other use is to obtain a depth profile of a sample by either XPS or AES. This is useful in determining the thickness and composition of a layer that has been formed by some process, such as an oxide layer formed on a corrosion specimen or a sample from a deposition procedure. Another example is to study a material that has been implanted with highly energetic ions to change a material's physical or chemical properties.

It is possible to analyze the species that are removed from the surface. The most common method of analysis of the sputtered species is by secondary ionization mass spectrometry (SIMS). This capability is available on many XPS or AES systems, but much of the work with SIMS is done on instruments designed solely for this method of analysis.[175–177]

There are two main types of ion guns. The more common design is founded primarily in the Einsel lens approach. The main components are an inlet valve, ionization, acceleration, focusing, and deflection regions. The gas is usually admitted from a small reservoir with a variable leak valve. With small compressed-gas cylinders care has to be taken so that impurities are not introduced. There is a maximum pressure that can be employed with variable leak valves. With a variable leak valve the amount of gas being admitted into the ion gun can be finely adjusted. Some systems have variable leak valves that are thermally controlled, which eliminates operator control for fine adjustments in order to maintain a constant pressure of the bombarding gas. The pressure is measured by the system ion gauge or by the ion current in the ionization assembly of the ion gun.

A heated filament boils off electrons that are attracted by a positive potential, usually about 70 V, in an ionization chamber containing the sputtering gas. The newly created positive ions are accelerated to the desired sputtering energy by a voltage placed on the ionization chamber with respect to earth. These ions then are focused in the Einsel lens part of the ion gun. Most ion guns have deflection plates that steer the beam after the ions have passed through the Einsel lens. It has been suggested that the sputtered area should be at least 10 times the area being

analyzed.[178] This is because the ion beam, which is about 0.1–1.0 mm wide, with a well-designed ion gun, has a Gaussian distribution which gives rise to edge effects. The crater produced by the ion beam therefore has a depth gradient, whereas ideally it would be flat bottomed. If only a small area is being analyzed, this would not present much of a problem (typically in single-spot AES the analyses are a few square micrometers in area). This assumes that the ion gun is focused on the same area that is being analyzed. However, as the area of analysis in XPS is almost always much larger than with AES, the sputtered area must be correspondingly much greater. It is better to raster the ion beam, rather than defocus it, because with a defocused beam the edge effects of the crater will be larger. The same reasons apply if an AES analysis is performed by rastering the electron beam or if different locations on a sample are being studied simultaneously. It is also possible to use a defocused ion beam and then analyze the broad crater edge with SAM to study compositional differences as a function of depth.

Most ion guns operate in a dynamic or differentially pumped mode, i.e. the source gas is continuously being admitted into the ionization source. With this approach the ionization chamber of the ion gun is usually evacuated with a separate pump, most often a small turbomolecular pump. This permits the ion source to be at a few orders of magnitude higher pressure than the main analysis chamber, that also is being continuously pumped. There will be a pressure rise in the analysis chamber, but this is usually only about 1–2 orders of magnitude over the base pressure. With a vacuum system that uses ion pumps, residual sputtering gas will remain, even after the gas source is turned off. This is due to the memory effect with ion pumps, and is much more pronounced with inert gases.

Ion guns usually operate with a current density of a few tens of $\mu A\,cm^{-2}$ and an acceleration voltage of 2–5 kV. For many materials this current density results in a removal rate of 5–10 nm min^{-1} or more, which in many instances is too large. The sputter rate can be reduced in several ways. The pressure in the ionization chamber can be reduced, but this requires adjustments with the variable leak valve. Another way is to lower the ionization filament current. Also, the sputtering rate can be reduced by rastering the ion beam over a larger area than is needed to produce a crater consistent with the technique that is being used. The sputter rate also can be lessened by using ions with lower kinetic energy, such as a few hundred electronvolts. Relative sputtering rates can be measured with standard materials, such as a known thickness of an oxide layer on Si or Ta,[179] or with a standard sandwich material of Ni and Cr.[180]

Another ion gun design, called the dualplasmatron, has two ionization regions for the ion production, and can operate over a wide range of ion currents. Thus, very low to very high sputter rates can be used as required by a particular analysis. However, this type of ion gun is much more expensive than conventional designs. When sputtering to depths of more than about 1 μm, AES or XPS are not the usual methods used; dynamic SIMS would probably be a better choice of technique in these instances.

Sputtering is a destructive process. Even with the removal of surface contaminants the surface will be altered by disturbing surface crystallinity or changing the chemistry of the surface species. For example, the reduction of many metal oxides has been observed with sputtering.[181,182] In addition, each element has a different sputtering rate, and occasionally the same element in different oxidation states has a distinct sputtering rate.[183] These effects, along with such factors as topography and radiation-enhanced diffusion can affect the depth resolution of profiles. Several approaches can be taken to reduce these difficulties, such as lowering the ion beam voltage and using a heavier sputtering gas (Xe instead of Ar). Other methods require changes in the sample manipulator or analysis chamber design, for instance a glancing ion beam angle, the use of two ion guns striking the sample at different angles, or sample rotation. With this last procedure the specimen is slowly rotated (about 1 rpm) during sputtering. With some samples under ideal conditions, depth resolution of the order of 2–3 atomic layers may be possible.[184,185]

4.2 Flood Gun

Earlier, the problem of charging in XPS was discussed. An electron flood gun often is employed to help overcome charging. For use in charge compensation the voltage and current of the beam should be adjustable from above zero for both parameters. The production of electrons requires a potential of tens of volts for extraction purposes. This means that a deceleration assembly is required for the low voltages needed for a flood gun, and the beam is diffuse. There will be added complications with systems that have magnetic lenses. In these instances a heated filament will be used, but some acceleration potential still will be required.

4.3 Fracture Stage

For certain materials-science studies a fracture stage is a useful accessory. This permits a fresh surface to be analyzed without sputter-induced changes. The bulk composition of a sample can be determined by surface analysis methods in this fashion. With SAM for example, the composition of grain boundaries can be compared to that of the main crystallite. Also, such a fresh surface can be

exposed to study the initial stages of reactions with a reactive gas, such as oxygen or water. The material to be fractured is held in a special sample holder, and most specimens are cooled well below room temperature with liquid nitrogen. The sample, often notched, then is fractured with an internal hammer assembly. The specimen holder then is transferred to the sample stage for subsequent analysis. Most samples require machining before they can be placed into the special fracture specimen holders. The fracture stage usually has to be on its own vacuum flange and should be in the analysis chamber.

4.4 Special Auxiliary Chambers

As noted above many surface analysis systems use a loadlock for sample entry into the analysis chamber. More specialized systems can be used, but these are usually custom-designed features. Glove boxes can be constructed around the loadlock for atmosphere-sensitive materials. A temporary and inexpensive alternative, but with some obvious performance degradation, is a portable glove bag. Another possible method is to store sensitive materials in a portable transfer assembly that has UHV flanges and appropriate transfer capability. Such an assembly then can be attached to a loadlock for subsequent evacuation. For certain materials preparation applications that require UHV, such as molecular beam epitaxy, a surface analysis system is attached to the other chamber. Another example of such an auxiliary chamber is a high-pressure reaction chamber for catalysis studies. Sample transport with these systems is accomplished using transfer arms. Some of these complete systems can be very costly.

4.5 Energy Dispersive X-ray Analysis

Energy dispersive X-ray analysis (EDX) is a widely used technique with which to analyze materials with a SEM. EDX instrumentation can be added to most SAMs. This method detects X-rays produced as the result of electron beam bombardment of a solid, namely the other de-excitation process besides Auger electron production. The X-rays are produced to a depth of about 1–3 μm and are observed with a lithium-doped silicon detector that is cooled with liquid nitrogen. Depending on the design of the detector assembly, the elements starting with boron or sodium throughout the periodic chart are observable. The minimum volume that can be analyzed is about $1\,\mu m^3$, even though the electron beam diameter may be much smaller. This is because scattering of the primary beam and subsequent redistributed electrons can create detectable X-rays beyond the initial electron beam impact area. In order to obtain the best use of this technique, the placement of an EDX in a SAM requires careful consideration with relation to the sample position and the incident electron beam.[186]

4.6 Residual Gas Analyzer

A residual gas analyzer is a very useful addition to a surface analysis system that is relatively inexpensive. Most of these analyzers are small quadrapole mass spectrometers. They can be employed to test for an occasional leak in the vacuum system. The usual indication of a leak is a peak at 32 Da that indicates oxygen. In order to locate the source of the leak a very small stream of helium (from a glass or plastic capillary) is directed to potential problem sites. One of the most common problems is that a flange has not been sealed properly; less likely sources are the various feedthroughs, windows, and mechanical transfer assemblies. Very rarely does a weld fail to give a measurable leak.

The mass spectrometer also is useful to check the purity of sputter gases and to check the background composition in the analysis chamber. The identity of the gas phase species can be deduced if there is a relatively large amount of outgassing from a specimen. This can provide additional information about a specimen.

The mass spectrometer should have a range of 1–100 Da. Some small mass spectrometers can analyze up to a few hundred daltons. The two main types of detectors employed are a Faraday cup and an electron multiplier. The Faraday cup detector has a detection limit in the low 10^{-9} torr region for most gases. The electron multiplier can detect species with a relative partial pressure about three to four orders of magnitude lower than the Faraday cup. Many of the small quadrupole mass spectrometers have both detector types. The mass spectrometers can be computer controlled or have a separate electronic package and recording system.

5 CONCLUSIONS

XPS and AES are widely used techniques for the analysis of the surface region of solids. The elemental and often the chemical environment of surfaces can be identified using either technique. The spatial distribution of the detected species can be determined to less than 5 μm with XPS and under 20 nm with AES on a routine basis. Both can be employed in a wide range of basic and applied investigations. Variations in the depth distribution of elements can be obtained by sputter profiles. However, as with any sophisticated analysis method, care has to be used in evaluating any but the relatively routine analytical problems. The capabilities of these techniques have been expanded greatly since they were first introduced, and their development continues.

ACKNOWLEDGMENTS

The author would like to thank Dr John Hammond of Physical Electronics and Dr David Surman of Kratos Analytical for providing some of the figures.

ABBREVIATIONS AND ACRONYMS

A/D	Analog to Digital
AES	Auger Electron Spectroscopy
AL	Attenuation Length
ARXPS	Angular Resolved X-ray Photoelectron Spectroscopy
ASTM	American Standards for Testing and Materials
CAT	Constant-analyzer Transmission
CHA	Concentric Hemispherical Analyzer
CMA	Cylindrical Mirror Analyzer
CRR	Constant Relative Ratio
D/A	Digital to Analog
DDF	Depth Distribution Factor
EDX	Energy Dispersive X-ray Analysis
ESCA	Electron Spectroscopy for Chemical Analysis
FAT	Fixed-analyzer Transmission
FRR	Fixed Retardation Ratio
HREELS	High-resolution Electron Energy Loss Spectroscopy
IMFP	Inelastic Mean Free Path
ISO	International Standards Organization
LEED	Low-energy Electron Diffraction
MCP	Multichannel Plate
PTFE	Poly(tetrafluoroethylene)
RFA	Retarding Field Analyzer
SAM	Scanning Auger Microscopy
SEM	Scanning Electron Microscopy
SHA	Spherical Hemispherical Analyzer
SIMS	Secondary Ionization Mass Spectrometry
S/N	Signal-to-Noise
STP	Standard Temperature and Pressure
UHV	Ultra High Vacuum
VAMAS	Versailles Project on Advanced Materials and Standards
VB	Valence Band
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Pulp and Paper **(Volume 10)**
X-ray Photoelectron Spectroscopy, Paper Surface Analysis by

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Electron Energy Loss Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • Scanning Electron Microscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
Surface Analysis for Electrochemistry: Ultrahigh Vacuum Techniques

Mass Spectrometry **(Volume 13)**
Secondary Ion Mass Spectrometry as Related to Surface Analysis

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction

REFERENCES

1. M.P. Seah, D. Briggs, 'Auger and Photoelectron Spectroscopy', in *Practical Surface Analysis*, 2nd edition, eds. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 15, Vol. 1, 1990.
2. M.P. Seah, 'VAMAS Surface Chemical Analysis – A Unifying Approach for the 1990s', *Surf. Interface Anal.*, **19**, 247–252 (1992).
3. C.J. Powell, 'Activities of the ASTM Committee E-42 on Surface Analysis', *Surf. Interface Anal.*, **19**, 237–240 (1992).
4. C.J. Powell, R. Shimizu, 'Formation of Technical Committee 201 on Surface Chemical Analysis by the International Organization for Standardization', *Surf. Interface Anal.*, **20**, 322–325 (1993).
5. M.P. Seah, 'Standards for Surface Analysis: A Clash Between Bureaucracy and Science?', *Phil. Trans. R. Soc. Lond. A*, **354**, 2765–2780 (1996).
6. D. Briggs, M.P. Seah (eds.), *Practical Surface Analysis: Auger and Photoelectron Spectroscopy*, 2nd edition, John Wiley & Sons, New York, Vol. 1, 1990.
7. J.C. Rivière, *Surface Analytical Techniques*, Oxford University Press, New York, 1990.
8. J.M. Walls (ed.), *Methods of Surface Analysis*, Cambridge University Press, New Rochelle, NY, 1989.
9. I.F. Ferguson, *Auger Microprobe Analysis*, Adam Hilger, New York, 1989.
10. T.L. Barr, *Modern ESCA: The Principles and Practice of X-ray Photoelectron Spectroscopy*, CRC Press, Boca Rotan, FL, 1994.
11. American Society for Testing and Materials, 'Standard Terminology Relating to Surface Analysis (E673–90)', *Surf. Interface Anal.*, **17**, 951–958 (1991).

12. C.J. Powell, 'Inelastic Interactions of Electrons with Surfaces – Application to Auger Electron Spectroscopy and X-ray Photoelectron Spectroscopy', *Surf. Sci.*, **299/300**, 34–48 (1994).
13. B. Hupfer, H. Schupp, J.D. Andrade, H. Ringsdorf, 'Photoelectron Mean Free Paths in Poly(diacetylene) Mono- and Multilayers', *J. Electron Spectrosc. Relat. Phenom.*, **23**, 103–107 (1981).
14. C.D. Bain, G.M. Whitesides, 'Attenuation Lengths of Photoelectrons in Hydrocarbons', *J. Phys. Chem.*, **93**, 1670–1673 (1989).
15. D.T. Clark, M.M. Abu-Shbak, W.J. Brennan, 'Electron Mean Free Paths as a Function of Kinetic Energy: A Substrate Overlayer Investigation of Polyparaxylene Films on Gold Using a Ti Kα X-ray Source', *J. Electron Spectrosc. Relat. Phenom.*, **28**, 11–21 (1982).
16. M.P. Seah, W. Dench, 'Quantitative Electron Spectroscopy of Surfaces: A Standard Data Base for Electron Inelastic Mean Free Paths in Solids', *Surf. Interface Anal.*, **1**, 2–11 (1979).
17. K. Berresheim, 'Experimentelle Bestimmung oberer Schranken für Matrix-Effekte in XPS', *Fresenius' Z. Anal. Chem.*, **319**, 731 (1984).
18. J.C. Ashley, C.J. Tung, 'Electron Inelastic Mean Free Paths in Several Solids for 200 eV $\leq$ E $\leq$ 10 keV', *Surf. Interface Anal.*, **4**, 52–55 (1982).
19. C.D. Wagner, L.E. Davis, W.M. Riggs, 'The Energy Dependence of the Electron Mean Free Path', *Surf. Interface Anal.*, **2**, 53–55 (1980).
20. C.J. Powell, 'The Energy Dependence of Electron Attenuation Lengths', *Surf. Interface Anal.*, **7**, 256–262 (1985).
21. C.J. Powell, 'Calculations of Electron Inelastic Mean Free Paths from Experimental Optical Data', *Surf. Interface Anal.*, **7**, 263–274 (1985).
22. A. Jablonski, 'Universal Energy Dependence of the Inelastic Mean Free Path', *Surf. Interface Anal.*, **20**, 317–321 (1993).
23. W.S.M. Werner, 'The Role of the Attenuation Parameter in Electron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **59**, 275–291 (1992).
24. W.H. Gries, 'A Universal Predictive Equation for the Inelastic Mean Free Pathlengths of X-ray Photoelectrons and Auger Electrons', *Surf. Interface Anal.*, **24**, 38–50 (1996).
25. S. Tanuma, C.J. Powell, D.R. Penn, 'Calculations of Electron Inelastic Mean Free Paths (IMFPs)', *Surf. Interface Anal.*, **25**, 25–35 (1997).
26. T.A. Sasaki, Y. Baba, 'Chemical State Studies of Zr and Nb Surfaces Exposed to Hydrogen Ions', *Phys. Rev. B*, **31**, 791–797 (1985).
27. A. Fujimori, L. Schlapbach, 'Electronic Structure of Yttrium Hydride Studied by X-ray Photoemission Spectroscopy', *J. Phys. C*, **17**, 341–351 (1984).
28. S.J. Kerber, J.J. Bruckner, K. Wozniak, S. Seal, S. Hardcastle, T.L. Barr, 'The Nature of Hydrogen in X-ray Photoelectron Spectroscopy: General Patterns from Hydroxides to Hydrogen Bonding', *J. Vac. Sci. Technol. A*, **14**, 13134–1320 (1996).
29. J.F. Moulder, W.F. Stickle, P.E. Sobol, K.D. Bomben, 'Handbook of X-ray Photoelectron Spectroscopy', *Physical Electronics*, Eden Prairie, MN, 1992.
30. J.R. Rumble, D.M. Bickham, C.J. Powell, 'The NIST X-ray Photoelectron Spectroscopy Database', *Surf. Interface Anal.*, **19**, 241–246 (1992).
31. B.V. Christ, 'A Critical Review of Popular XPS Databanks', *J. Surf. Anal.*, **2**, 21–58 (1996).
32. N.S. McIntyre, D.G. Zetaruk, 'X-ray Photoelectron Spectroscopic Studies of Iron Oxides', *Anal. Chem.*, **49**, 1521–1529 (1977).
33. N.H. Turner, J.S. Murday, D.E. Ramaker, 'Quantitative Determination of Surface Composition of Sulfur Bearing Anion Mixtures by Auger Electron Spectroscopy', *Anal. Chem.*, **52**, 84–92 (1980).
34. S. Larrson, 'Shake-up and Multiplet Structure of ESCA Satellites of Cu Compounds', *Chem. Phys. Lett.*, **40**, 362–366 (1976).
35. T.L. Barr, S. Seal, K. Wozniak, J. Klinowski, 'ESCA Studies of the Coordination of Aluminum in Oxide Environments', *J. Chem. Soc., Faraday Trans.*, **93**, 181–186 (1997).
36. M.P. Seah, I.S. Gilmore, G. Beamson, 'XPS: Binding Energy Calibration of Electron Spectrometers 5 – Re-evaluation of the Reference Energies', *Surf. Interface Anal.*, **26**, 642–649 (1998).
37. M.P. Seah, I.S. Gilmore, S.J. Spencer, 'XPS: Binding Energy Calibration of Electron Spectrometers 4 – Assessment of Effects for Different X-ray Sources, Analyzer Resolutions, Angles of Emission and Overall Uncertainities', *Surf. Interface Anal.*, **26**, 617–641 (1998).
38. A.C. Miller, C.J. Powell, U. Geius, C.R. Anderson, 'Energy Calibration of X-ray Photoelectron Spectrometers. Part III: Location of the Zero Point on the Binding Energy Scale', *Surf. Interface Anal.*, **26**, 606–614 (1998).
39. T. Ohtsubo, K. Hirokawa, Y. Fukuda, Y. Oh-Hashi, N. Usuk, N. Gennai, H. Ishida, S. Yoshida, Y. Sekimoto, K. Suzuki, 'Surface Analysis of Metal 1: Activity Report of Surface Analysis Subcommittee of the Iron and Steel Analysis Committee, The Joint Research Society of ISIJ', *Trans. Iron Steel Instit. Jpn.*, **27**, 60–76 (1987).
40. T. Ohtsubo, K. Hirokawa, Y. Fukuda, Y. Oh-Hashi, N. Usuk, N. Gennai, H. Ishida, S. Yoshida, Y. Sekimoto, K. Suzuki, 'Surface Analysis of Metal 2: Activity Report of Surface Analysis Subcommittee of the Iron and Steel Analysis Committee, The Joint Research Society of ISIJ', *Trans. Iron Steel Instit. Jpn.*, **27**, 145–161 (1987).
41. M.P. Seah, 'Measurement – AES and XPS', *J. Vac. Sci. Technol. A*, **3**, 1330–1337 (1985).
42. C.J. Powell, 'Energy Calibration of X-ray Photoelectron Spectrometers: Results of an Interlaboratory Comparison to Evaluate a Proposed Calibration Procedure', *Surf. Interface Anal.*, **23**, 121–132 (1995).

43. P.J. Cumpson, M.P. Seah, S.J. Spencer, 'Simple Procedure for Precise Peak Maximum Estimation for Energy Calibration in AES and XPS', *Surf. Interface Anal.*, **24**, 687–694 (1996).
44. M.C. Desjonqueres, D. Sebilleau, G. Treglia, D. Spanjaard, C. Guillot, D. Chauveau, J. Lecante, 'Analysis of Clean Transition Metal Surfaces by Core Level Spectroscopy', *Scanning Microsc.*, **1**(4), 1557–1570 (1987).
45. W. Egelhoff, Jr, 'Core-level Binding Energy Shifts at Surfaces and in Solids', *Surf. Sci. Rev.*, **6**, 257–415 (1987).
46. K.R. Harikumar, S. Ghosh, C.N.R. Rao, 'X-ray Photoelectron Spectroscopic Investigations of Cu–Ni, Au–Ag, Ni–Pd, and Cu–Pd Bimetallic Clusters', *J. Phys. Chem. A*, **101**, 536–540 (1997).
47. C.N.R. Rao, V. Vijayakrishnan, H.N. Alyer, G.U. Kulkarni, G.N. Subbanna, 'An Investigation of Well Characterized Small Gold Clusters by Photoelectron Spectroscopy, Tunneling Spectroscopy, and Cognate Techniques', *J. Phys. Chem.*, **97**, 11 157–11 160 (1993).
48. C.S. Fadley, 'Basic Concepts of X-ray Photoelectron Spectroscopy', in *Electron Spectroscopy: Theory, Techniques and Applications*, eds. C.R. Brundle, A.D. Baker, Academic Press, New York, Chapter 1, 1978.
49. N.H. Turner, in *Investigations of Surface and Interfaces – Part B*, 2nd edition, eds. B.W. Rossiter, R.C. Batzold, Physical Methods of Chemistry Series, John Wiley & Sons, New York, 159–166, Vol. IXB, 1993.
50. T.L. Barr, 'Studies in Differential Charging', *J. Vac. Sci. Technol. A*, **7**, 1677–1683 (1989).
51. G. Beamson, A. Bunn, D. Briggs, 'High Resolution Monochromated XPS of Poly(methyl methacrylate) Thin Films on a Conducting Substrate', *Surf. Interface Anal.*, **17**, 105–115 (1991).
52. B.J. Tielsch, J.E. Fulghum, 'Differential Charging in XPS 1: Demonstration of Lateral Charging in a Bulk Insulator Using Imaging XPS', *Surf. Interface Anal.*, **24**, 422–447 (1996).
53. B.J. Tielsch, J.E. Fulghum, D.J. Surman, 'Differential Charging in XPS 2: Sample Mounting and X-ray Flux Effects on Heterogeneous Samples', *Surf. Interface Anal.*, **24**, 459–468 (1996).
54. S. Kohiki, 'Photoemission from Small Pd Clusters on Al_2O_3 and SiO_2 Substrates', *Appl. Surf. Sci.*, **25**, 81–94 (1986).
55. B.J. Tielsch, J.E. Fulghum, 'Differential Charging in XPS 3: A Comparison of Charging in Thin Polymer Overlayers on Conducting and Nonconducting Substrates', *Surf. Interface Anal.*, **25**, 904–912 (1997).
56. G. Vereecke, P.G. Rouxhet, 'Surface Charging of Insulating Samples in X-ray Photoelectron Spectroscopy', *Surf. Interface Anal.*, **26**, 490–497 (1998).
57. T.L. Barr, S. Seal, 'Nature of the Use of Adventitious Carbon as a Binding Energy Standard', *J. Vac. Sci. Technol. A*, **13**, 1239–1246 (1995).
58. P. Swift, 'Adventitious Carbon – The Panacea for Energy Referencing?', *Surf. Interface Anal.*, **4**, 47–52 (1982).
59. S. Kohiki, T. Ohmura, K. Kusoa, 'A New Charge-correction Method in X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **28**, 229–237 (1983).
60. S. Kohiki, T. Ohmura, K. Kusoa, 'Appraisal of a New Charge Correction Method in X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **31**, 85–90 (1983).
61. C.E. Bryson, 'Surface-potential Control in XPS', *Surf. Sci.*, **189/190**, 50–58 (1987).
62. K.S. Robinson, H. Schmiedel, 'XPS Analysis of Biomaterials Using a New Method of Integrated Charge Compensation', *Fizika A*, **5**, 49–56 (1996).
63. A. Cros, 'Charging Effects in X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **59**, 1–14 (1992).
64. R.G. Copperthwaite, 'The Study of Radiation-induced Chemical Damage at Solid Surfaces Using Photoelectron Spectroscopy', *Surf. Interface Anal.*, **2**, 17–25 (1980).
65. T. Sekine, N. Ikeo, Y. Nagasawa, 'Comparison of AES Chemical Shifts with XPS Chemical Shifts', *Appl. Surf. Sci.*, **100/101**, 30–35 (1996).
66. C.D. Wagner, A. Joshi, 'The Auger Parameter, its Utility and Advantages – A Review', *J. Electron Spectrosc. Relat. Phenom.*, **47**, 283–313 (1988).
67. D.E. Ramaker, N.H. Turner, J.S. Murday, L.E. Toth, M. Osofsky, F.L. Hutson, 'Observed Trends in the X-ray Photoelectron and Auger Spectra of High-temperature Superconductors', *Phys. Rev. B*, **36**, 5672–5675 (1987).
68. J.A. Gardella, Jr, S.A. Ferguson, R.L. Chin, 'π^* and π Shake-up Satellites for the Analysis of Structure and Bonding in Aromatic Polymers by X-ray Photoelectron Spectroscopy', *Appl. Spectrosc.*, **40**, 224–232 (1986).
69. S. Thomas, P.M.A. Sherwood, 'Valence Band Spectra of Aluminum Oxides, Hydroxides, and Oxyhydroxides Interpreted by Xα Calculations', *Anal. Chem.*, **64**, 2488–2495 (1992).
70. P.M.A. Sherwood, 'Surface Analysis for the Investigation of Electrochemical and Corrosion Systems', *Anal. Chim. Acta*, **283**, 52–61 (1993).
71. P.A. Dowben, 'Core Level Spectroscopy and the Characterization of Surface Segregation', in *Surface Segregation Phenomena*, eds. P.A. Dowben, A.A. Miller, CRC Press, Boca Raton, FL, 145–168, 1990.
72. H. Fischer, R. Svagera, H. Ebel, M.F. Ebel, B. Schossmann, 'Depth Profiling by ARXPS in Surface Analysis', *Fresenius J. Anal. Chem.*, **353**, 473–477 (1995).
73. J.E. Fulgham, R.W. Linton, 'Quantitation of Coverages on Rough Surfaces by XPS', *Surf. Interface Anal.*, **13**, 186–192 (1988).
74. W.H. Gries, 'A General Procedure for Extracting Quantitative Depth Information from Take-off Angle Resolved XPS and AES', *Appl. Surf. Sci.*, **100/101**, 41–46 (1996).

75. D.E. Ramaker, 'The Past, Present, and Future of Auger Line Shape Analysis', *Crit. Rev. Solid State Mater. Sci.*, **17**, 211–276 (1991).
76. D.E. Ramaker, 'Auger Line Shapes as a Probe of Electronic Structure in Covalent Systems', *Phys. Scr.*, **T41**, 77–87 (1992).
77. M. Cini, 'Correlation Effects in CVV Spectra from Partially Filled Bands', *Phys. Scr.*, **T41**, 59–66 (1992).
78. A.M. Prandini, A. Ruocco, S. Nannarone, 'Evolution of Auger Peaks in GaP(110) with Hydrogen Chemisorption', *J. Electron Spectrosc. Relat. Phenom.*, **59**, 393–397 (1992).
79. L.E. Davis, N.C. MacDonald, P.W. Palmberg, G.E. Raich, R.E. Weber, *Handbook of Auger Electron Spectroscopy*, 2nd edition, Perkin-Elmer Corp., Eden Prairie, MN, 1976.
80. T. Sekine, Y. Nagasawa, M. Kudoh, Y. Sakai, A.S. Parks, J.D. Geller, A. Mogami, K. Hirata, *Handbook of Auger Electron Spectroscopy*, JEOL Ltd., Tokyo, 1982.
81. M.P. Seah, I.S. Gilmore, 'High Resolution Digital Auger Data Base of True Spectra for Auger Electron Spectroscopy Intensities', *J. Vac. Sci. Technol. A*, **14**, 1401–1407 (1996).
82. C.D. Wagner, H.A. Six, W.T. Jansen, J.A. Taylor, 'Improving the Accuracy of Determination of Line Energies by ESCA: Chemical State Plots for Silicon–Aluminum Compounds', *Appl. Surf. Sci.*, **9**, 203–213 (1981).
83. I.F. Ferguson, *Auger Microprobe Analysis*, Adam Hilger, New York, 52–54, 1989.
84. I.F. Ferguson, *Auger Microprobe Analysis*, Adam Hilger, New York, 318–319, 1989.
85. S. Hofmann, 'Charging Effects in X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **59**, 15–32 (1992).
86. A. Hoffman, P.J.K. Paterson, 'Surface and Subsurface 1 keV Electron Stimulation Reduction of Sapphire Studied by Electron Spectroscopy', *Appl. Surf. Sci.*, **5**, 301–308 (1996).
87. C.G. Pantano, T.E. Madey, 'Electron Beam Damage in Auger Electron Spectroscopy', *Appl. Surf. Sci.*, **7**, 115–141 (1981).
88. J.T. Yates, Jr, *Experimental Innovations in Surface Science: A Guide to Practical Laboratory Methods and Instruments*, AIP Press (Springer-Verlag), New York, 1998.
89. *Annual Book of ASTM Standards*, American Society for Testing and Materials, West Conshohocen, PA, 905–1003, Vol. 3.06, 1998.
90. A. Roth, *Vacuum Technology*, 3rd edition, North Holland, New York, 1990.
91. D.H. Fairbrother, H. Johnston, G. Somorjai, 'Electron Spectroscopy Studies of the Surface Composition in the H_2SO_4/H_2O Binary System', *J. Phys. Chem.*, **100**, 13 696–13 700 (1996).
92. B.E. Koel, J.M. White, 'Interference of O Kα X-ray Excited Auger Spectra', *J. Electron Spectrosc. Relat. Phenom.*, **22**, 237–245 (1981).
93. J. Schweppe, R.D. Deslattes, T. Mooney, C.J. Powell, 'Accurate Measurement of Mg and Al $K\alpha_{1,2}$ X-ray Energy Profiles', *J. Electron Spectrosc. Relat. Phenom.*, **67**, 463–478 (1994).
94. O. Ganschow, P. Steffens, 'A New X-ray Generator for XPS Applications', *J. Vac. Sci. Technol.*, **21**, 845–852 (1982).
95. D.D. Hawn, 'Collimation of Dual Anode X-ray Source Used for X-ray Photoelectron Spectroscopy (XPS)', *Rev. Sci. Instrum.*, **54**, 767–768 (1983).
96. J.C. Rivière, 'Instrumentation', in *Practical Surface Analysis, Auger and Photoelectron Spectroscopy*, 2nd edition, eds. D. Briggs, M.P. Seah, John Wiley & Sons, New York, 50, Vol 1, 1990.
97. D.T. Clark, 'The Modification, Degradation, and Synthesis of Polymer Surfaces by ESCA', *ACS Symp. Ser.*, **162**, 255–291 (1981).
98. H. Fitzpatrick, P.F. Luckham, S. Eriksen, K. Hammond, 'Use of X-ray Photoelectron Spectroscopy to Study Protein Adsorption to Mica Surfaces', *J. Colloid Interface Sci.*, **149**, 1–9 (1992).
99. J.E. Castle, R.H. West, 'Bremsstrahlung Induced Auger Peaks', *J. Electron Spectrosc. Relat. Phenom.*, **18**, 355–358 (1980).
100. H.M. Liao, R.N.S. Sodri, T.W. Coyle, 'Surface Composition of AlN Powders by X-ray Photoelectron Spectroscopy and Bremsstrahlung Excited Auger Electron Spectroscopy', *J. Vac. Sci. Technol. A*, **11**, 2681–2686 (1993).
101. C. Klauber, 'Refinement of Magnesium and Aluminum Kα X-ray Source Functions', *Surf. Interface Anal.*, **20**, 703–715 (1993).
102. K. Yates, R.H. West, 'Monochromatized Ag Lα X-rays as a Source for Higher Energy XPS', *Surf. Interface Anal.*, **5**, 133–138 (1983).
103. G. Beamson, D.T. Clark, J. Kendrick, D. Briggs, 'Observation of Vibrational Asymmetry in the High Resolution Monochromatized XPS of Hydrocarbon Polymers', *J. Electron Spectrosc. Relat. Phenom.*, **57**, 79–90 (1991).
104. 'High Resolution XPS Spectra of Polymers', Kratos Analytical, Manchester, UK, 1992.
105. G. Beamson, D. Briggs, *High Resolution XPS of Organic Polymers*, John Wiley & Sons, New York, 1992.
106. R.L. Graham, C.D. Bain, H.A. Biebuyck, P.E. Laibinis, G.M. Whitesides, 'Damage to CF_3CONH-terminated Organic Self Assembled Monolayers (SAM) on Al, Ti, Cu, and Au by Al Kα X-rays is Due Principally to Electrons', *J. Phys. Chem.*, **97**, 9456–9465 (1993).
107. R. Chaney, 'Recent Developments in Spatially Resolved ESCA', *Surf. Interface Anal.*, **10**, 36–47 (1987).
108. U. Gelius, B. Wannber, F.H. Baltzer, G. Carlsson, C.-G. Johansson, J. Larsson, P. Munger, G. Vegerfors, 'A New ESCA Instrument with Improved Surface Sensitivity,

Fast Imaging Properties and Excellent Energy Resolution', *J. Electron Spectrosc. Relat. Phenom.*, **52**, 747–785 (1990).
109. C.D. Wagner, 'Studies of the Charging of Insulators in ESCA', *J. Electron Spectrosc. Relat. Phenom.*, **18**, 345–349 (1980).
110. M.F. Koning, J.T. Grant, 'Deconvolution in X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, 1984, **33**, 9–22 (1984).
111. M.F. Koning, J.T. Grant, 'Monochromator Versus Deconvolution for XPS Studies Using a Kratos ES300 System', *J. Electron Spectrosc. Relat. Phenom.*, **36**, 213–225 (1985).
112. G.K. Wertheim, 'Deconvolution and Smoothing: Applications in ESCA', *J. Electron Spectrosc. Relat. Phenom.*, **6**, 239–251 (1975).
113. M. Dutoit, S. Contarini, G. Margaritondo, 'Spatially Localized Energy Shifts in the Photoemission Spectroscopy on Insulators', *J. Appl. Phys.*, **76**, 3710–3713 (1994).
114. S. Valeri, 'Auger Electron Emission by Ion Impact on Solid Surfaces', *Surf. Sci. Rep.*, **17**, 85–150 (1993).
115. J.I. Goldstein, A.D. Roming, Jr, D.E. Newbury, C.E. Lyman, P. Echlin, C. Fiori, D.C. Joy, E. Lifshin, *Scanning Electron Microscopy and X-ray Analysis*, 2nd edition, Plenum Press, New York, Chapter 2, 1992.
116. G.G. Hembree, J.A. Venables, 'Nanometer Resolution Scanning Auger Electron Microscopy', *Ultramicroscopy*, **47**, 109–120 (1992).
117. R.R. Olson, L.A. LaVanier, D.H. Narum, 'Backscattering Limitations to Spatial Resolution in the Auger Microprobe', *Appl. Surf. Sci.*, **70/71**, L266–272 (1993).
118. E.N. Sickafus, H.P. Bonzel, 'LEED and AES in Surface Analysis', in *Progress in Surface and Membrane Science*, eds. J.F. Danielli, M.D. Rosenberg, D.A. Cadenhead, Academic Press, New York, 164–167, Vol. 4, 1971.
119. E.N. Sickafus, P.H. Holloway, 'Specimen Position Effects on Energy Shifts and Signal Intensity in a Single Stage Cylindrical Mirror Analyzer', *Surf. Sci.*, **51**, 131–139 (1975).
120. P.W. Palmberg, 'A Combined ESCA and Auger Spectrometer', *J. Vac. Sci. Technol.*, **12**, 379–384 (1975).
121. J.A. Knapp, G.J. Lapeyre, N.V. Smith, M. Traum, 'Modification of a Cylindrical Mirror Analyzer for Angle Resolved Electron Spectroscopy', *Rev. Sci. Instrum.*, **53**, 781–784 (1982).
122. J.C. Rivière, 'Instrumentation', in *Practical Surface Analysis: Auger and Photoelectron Spectroscopy*, 2nd edition, eds. D. Briggs, M.P. Seah, John Wiley, New York, 29–31, Vol. 1, 1990.
123. M. Lampton, C.W. Carison, 'Low Distortion Resistive Anodes for Two Dimensional Position Sensitive MCP Systems', *Rev. Sci. Instrum.*, **50**, 1093–1097 (1979).
124. M.P. Seah, G.C. Smith, 'AES-accurate Intensity Calibration of Electron Spectrometers – Results of a BCR Interlaboratory Comparison Co-sponsored by the VAMAS SCA TWP', *Surf. Interface Anal.*, **17**, 855–874 (1991).
125. M.P. Seah, 'A System for the Intensity Calibration of Electron Spectrometers', *J. Electron Spectrosc. Relat. Phenom.*, **71**, 191–204 (1995).
126. M.P. Seah, 'Versailles Project on Advanced Materials and Standards Study of Intensity Stability of Cylindrical Mirror Analyzer Based Auger Electron Spectrometers', *J. Electron Spectrosc. Relat. Phenom.*, **58**, 345–357 (1992).
127. M.P. Seah, G.C. Smith, 'Quantitative AES – Reducing Errors in Measured Analog Spectral Intensities Through Control of the Electron Detector', *Surf. Interface Anal.*, **15**, 701–704 (1990).
128. M.P. Seah, 'Scattering in Electron Spectrometers, Diagnosis and Avoidance 1-Concentric Hemispherical Analyzers', *Surf. Interface Anal.*, 1993, **20**, 865–875 (1993).
129. M.P. Seah, 'Scattering in Electron Spectrometers, Diagnosis and Avoidance 2-Cylindrical Mirror Analyzers', *Surf. Interface Anal.*, **20**, 876–890 (1993).
130. M.M. El Gomati, T.A. El Bakush, 'Sources of Internal Scattering of Electrons in a Cylindrical Mirror Analyzer (CMA)', *Surf. Interface Anal.*, **24**, 152–162 (1996).
131. G.C. Smith, M.P. Seah, 'ESCA Microscope – A New Approach for Imaging in XPS', *J. Electron Spectrosc. Relat. Phenom.*, **42**, 359–363 (1987).
132. M.P. Seah, G.C. Smith, 'Concept of an Imaging XPS System', *Surf. Interface Anal.*, **11**, 69–79 (1988).
133. P. Coxon, J. Krizek, M. Humpherson, I.R.M. Wardell, 'ESCASCOPE – A New Imaging Photoelectron Spectrometer', *J. Electron Spectrosc. Relat. Phenom.*, **52**, 821–836 (1990).
134. N.M. Forsyth, P. Coxon, 'Use of Parallel Imaging XPS To Perform Rapid Analysis of Polymer Surfaces with Spatial Resolution Less than 5 μmeters', *Surf. Interface Anal.*, **21**, 431–435 (1994).
135. I.W. Drummond, 'Spatial Resolution in X-ray Photoelectron Spectroscopy', *Philos. Trans. Royal Soc. Lond. A*, **354**, 2667–2682 (1996).
136. N.M. Glezos, A.G. Nassiopoulos, 'Lateral Resolution of Auger Electron Spectroscopy in the Energy Range 5–100 keV – Thin Overlayers on a High Z Material Substrate', *Surf. Sci.*, **254**, 309–319 (1991).
137. M.M. El Gomati, M. Prutton, B. Lamb, C.G. Tuppen, 'Edge Effects and Image Contrast in Scanning Auger Microscopy – A Theory Experimental Comparison', *Surf. Interface Anal.*, **11**, 251–265 (1988).
138. K.D. Childs, D. Narum, L.A. LaVanier, P.M. Lindley, B.W. Schueler, G. Mulholland, A.C. Diebold, 'Comparison of Submicron Particle Analysis by Auger Electron Spectroscopy, Time-of-flight Secondary Ion Mass Spectrometry, and Secondary Electron Microscopy with Energy Dispersive X-ray Spectroscopy', *J. Vac. Sci. Technol. A*, **14**, 2392–2404 (1996).

139. C.J. Powell, N.E. Erikson, T.E. Madey, 'Results of a Joint Auger/ESCA Round Robin Sponsored by ASTM Committee E-42 on Surface Analysis', *J. Electron Spectrosc. Relat. Phenom.*, 1979, **17**, 361–403 (1979).
140. H.V. Malmsted, C.G. Enke, S.R. Crouch, *Electronics and Instrumentation for Scientists and Engineers*, Benjamin/Cummings, Menlo Park, CA, 433, 1981.
141. J.M. Walls, I.K. Brown, D.D. Hall, 'The Application of Taper-sectioning Techniques for Depth Profiling Using Auger Electron Spectroscopy', *Appl. Surf. Sci.*, **15**, 93–107 (1983).
142. A. Savitzky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Anal. Chem.*, **36**, 1627–1639 (1964).
143. J. Steiner, Y. Termonia, J. Deltour, 'Comments on Smoothing and Differentiation of Data by Simplified Least Square Procedure', *Anal. Chem.*, **44**, 1906–1909 (1972).
144. G. Horkick, 'Digital Data Handling of Spectra Utilizing Fourier Transformations', *Anal. Chem.*, **44**, 943–947 (1972).
145. M.P. Seah, W.A. Dench, 'Smoothing and the Signal-to-noise Ratio of Peaks in Electron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **48**, 43–54 (1989).
146. K. Plyakis, E. Sacher, 'Spectral Noise Removal by Digital Filtering and Its Application to Surface Analysis', *Appl. Surf. Sci.*, **55**, 159–164 (1992).
147. K. Plyakis, E. Sacher, 'Reply to Comment of M.P. Seah, P.J. Cumpson on ''Spectral Noise Removal by Digital Filtering and Its Application to Surface Analysis'' ', *Appl. Surf. Sci.*, **63**, 361–364 (1993).
148. M.P. Seah, P. Cumpson, 'Comment on ''Spectral Noise Removal by Digital Filtering and its Application to Surface Analysis'' by K. Plyakis and E. Sacher', *Appl. Surf. Sci.*, **62**, 195–198 (1992).
149. A. Procter, P.M.A. Sherwood, 'Data Analysis Techniques in X-ray Photoelectron Spectroscopy', *Anal. Chem.*, **54**, 13–19 (1982).
150. M.P. Seah, M.T. Anthony, W.A. Dench, 'Characterization of Computer Differentiation of Spectra in AES and Its Relation to Differentiation by the Modulation Technique', *J. Phys. E.*, **16**, 848–857 (1983).
151. C.D. Wagner, L.E. Davis, M.V. Zeller, J.A. Taylor, R.H. Raymond, L.E. Gale, 'Empirical Atomic Sensitivity Factors for Quantitative Analysis by Electron Spectroscopy for Chemical Analysis', *Surf. Interface Anal.*, **3**, 211–225 (1981).
152. D.A. Shirley, 'High Resolution X-ray Photoemission Spectrum of the Valence Bands of Gold', *Phys. Rev. B*, **5**, 4709–4714 (1972).
153. S. Tougaard, 'Inelastic Background Correction and Quantitative Surface Analysis', *J. Electron Spectrosc. Relat. Phenom.*, **52**, 243–271 (1990).
154. S. Tougaard, 'Universality Classes of Inelastic Electron Scattering Cross-sections', *Surf. Interface Anal.*, **25**, 137–154 (1997).
155. J.H. Scofield, 'Hartree–Slater Subshell Photoionization Cross-sections', *J. Electron Spectrosc. Relat. Phenom.*, **8**, 129–137 (1976).
156. J. Szajman, J.C. Jenkin, R.C.G. Leckey, J. Liesengang, 'Subshell Photoionization Cross-sections, Electron Mean Free Paths and Quantitative X-ray Photoelectron Spectroscopy', *J. Electron Spectrosc. Relat. Phenom.*, **19**, 393–408 (1981).
157. J.J. Yeh, I. Lindau, 'Atomic Subshell Photoionization Cross-sections and Asymmetry Parameters-$1 \leq Z \leq 103$', *Atom. Data Nucl. Data Tables*, **32**, 1–155 (1985).
158. S. Tougaard, C. Jansson, 'Background Correction in XPS – Comparison of Validity of Different Methods', *Surf. Interface Anal.*, **19**, 171–174 (1992).
159. C. Jansson, S. Tougaard, G. Beamson, D. Briggs, S.F. Davies, A. Rossi, R. Haubert, G. Hobi, N.M.D. Brown, B.J. Meenan, C.A. Anderson, M. Repoux, C. Malitesta, L. Sabbatini, 'Intercomparison of Algorithms for Background Correction in XPS', *Surf. Interface Anal.*, **23**, 484–494 (1995).
160. S. Doniach, M. Šunjic, 'Many-electron Singularity in X-ray Photoemission and X-ray Line Spectra from Metals', *J. Phys.*, **C3**, 284–291 (1970).
161. R.F. Reilman, A. Msezane, S.T. Manson, 'Relative Intensities in Photoelectron Spectroscopy of Atoms and Molecules', *J. Electron Spectrosc. Relat. Phenom.*, **8**, 389–394 (1976).
162. W.F. Stickle, P.E. Sobol, 'Enhancing the Interpretation of X-ray Photoelectron Spectra Using Numerical Methods', *Surf. Interface Anal.*, **19**, 165–170 (1992).
163. N.H. Turner, J.H. Wandass, F.L. Hutson, 'Principal Component Analysis with Integral Auger Electron Spectroscopy', *J. Vac. Sci. Technol. A*, **8**, 4033–4038 (1990).
164. S.W. Gaarenstroom, 'Principal Component Analysis of Auger Line Shapes at Solid–Solid Interfaces', *Appl. Surf. Anal.*, **7**, 7–18 (1981).
165. M.A. Eberhardt, A. Proctor, M. Houalla, D.M. Hercules, 'Investigation of V Oxidation States in Reduced V/Al_2O_3 Catalysts by XPS', *J. Catal.*, **162**, 368–376 (1996).
166. G. Bubert, N. Korte, R.P.H. Garten, E. Grallath, M. Wielunski, 'Application of Factor Analysis in Electron Spectroscopic Depth Profiling on Copper Oxide', *Anal. Chim. Acta*, **297**, 187–195 (1994).
167. A. Arranz, C. Palacio, 'Factor Analysis, a Useful Tool for Solving Analytical Problems in AES and XPS: A Study of the Performances and Limitations of the Indicator Function', *Surf. Interface Anal.*, **22**, 93–97 (1994).
168. S. Mroz, 'Physical Foundation of Quantitative Auger Analysis', *Prog. Surf. Sci.*, **46**, 377–437 (1994).
169. C.J. Powell, M.P. Seah, 'Precision, Accuracy and Uncertainty in Quantitative Surface Analyses by Auger Electron Spectroscopy and X-ray Photoelectron Spectroscopy', *J. Vac. Sci. Technol. A*, **8**, 735–763 (1990).

170. I.S. Tilnin, A. Jablonski, W.S.M. Werner, 'Quantitative Surface Analysis by Auger and X-ray Photoelectron Spectroscopy', *Prog. Surf. Sci.*, **52**, 193–355 (1996).
171. L.R. Barkshire, J.M. Walton, M. Prutton, 'Testing the Success of Analytical Image Correction Routines – Surface Images', *Surf. Interface Anal.*, **20**, 583–588 (1993).
172. S.D. Böhmig, H. Beilschmidt, B.M. Reichi, 'Segmentation and Scatter Diagram Analysis of Scanning Auger Images – A Critical Comparison of Results', *Fresenius' J. Anal. Chem.*, **346**, 223–226 (1993).
173. M. Prutton, I.R. Barkshire, P.G. Kenny, R.H. Roberts, M. Wenham, 'Multi-imaging and Multivariate Statistics Used for 3D Characterization at Surfaces', *Philos. Trans. R Soc. Lond. A*, **354**, 2683–2695 (1996).
174. W.A. Dench, L.B. Hazell, M.P. Seah, and the VAMAS Community, 'VAMAS Surface Chemical Analysis Standard Data Transfer Format with Skeleton Decoding Programs', *Surf. Interface. Anal.*, **13**, 63–122 (1988).
175. J.C. Vickerman, 'Static Secondary Ion Mass Spectroscopy', in *Methods of Surface Analysis*, ed. J.M. Walls, Cambridge University Press, New Rochelle, NY, Chapter 6, 1989.
176. D.E. Sikes, 'Dynamic Secondary Ion Mass Spectrometry', in *Methods of Surface Analysis*, ed. J.M. Walls, Cambridge University Press, New Rochelle, NY, Chapter 7, 1989.
177. D. Briggs, M.P. Seah (eds.), *Practical Surface Analysis*, 2nd edition, John Wiley & Sons, New York, Vol. 2, 1992.
178. H.J. Mathieu, D. Landolt, 'On the Influence of Crater Geometry on Depth Resolution of AES and XPS Profiles of Tantalum Oxide Films', *Surf. Interface Anal.*, **5**, 77–82 (1983).
179. C.P. Hunt, M.T. Anthony, M.P. Seah, 'AES and XPS Depth Profiling Certified Reference Material', *Surf. Interface Anal.*, **6**, 92–93 (1984).
180. J. Fine, B. Navinsek, 'An NBS Standard Reference Material for Depth Profile Analysis', *Surf. Interface Anal.*, **11**, 542–543 (1988).
181. K.S. Kim, W.E. Baitinger, J.W. Amy, N. Winograd, 'ESCA Studies of Metal–Oxygen Surfaces Using Argon and Oxygen Ion Bombardment', *J. Electron Spectrosc. Relat. Phenom.*, **5**, 351–367 (1974).
182. D.F. Mitchell, G.I. Sproule, M.J. Graham, 'Sputter Reduction of Oxides by Ion Bombardment During Auger Depth Profile Analysis', *Surf. Interface Anal.*, **15**, 487–497 (1990).
183. H.H. Anderson, H.L. Bay, 'Sputtering Yield Measurements', in *Sputtering by Particle Bombardment I*, ed. R. Behrisch, Springer-Verlag, New York, 72, 1981.
184. S. Hofmann, 'Approaching the Limits of High Resolution Depth Profiling', *Appl. Surf. Sci.*, **70/71**, 9–19 (1993).
185. A. Zalar, S. Hofmann, 'Comparison of Rotational Depth Profiling with AES and XPS', *Appl. Surf. Sci.*, **68**, 361–367 (1993).
186. J.I. Goldstein, A.D. Roming, Jr, D.E. Newbury, C.E. Lyman, P. Echlin, C. Fiori, D.C. Joy, E. Lifshin, *Scanning Electron Microscopy and X-ray Analysis*, 2nd edition, Plenum Press, New York, Chapters 5–9, 1992.

X-ray Techniques: Overview

Ron Jenkins
International Centre for Diffraction Data, Newtown Square, USA

1 Introduction **2**
1.1 Brief History of the Development of X-ray Fluorescence and Powder Diffraction Methods 2
1.2 Role of X-ray Methods in the Modern Analytical Laboratory 2
2 Properties of X-radiation **3**
2.1 Continuous and Characteristic Radiation 3
2.2 X-ray Spectra 3
2.3 Absorption 4
2.4 Scattering 5
2.5 Role of Crystal Structure in X-ray Scattering and Diffraction 5
2.6 Interference and Diffraction Effects 6
3 Instrumentation for X-ray Fluorescence Analysis **6**
3.1 Basis of the Method 7
3.2 X-ray Sources 7
3.3 X-ray Detectors 7
3.4 The Wavelength-dispersive Spectrometer 8
3.5 The Energy-dispersive Spectrometer 8
4 Elemental Analysis by X-ray Fluorescence **8**
4.1 Specimen Preparation Techniques for X-ray Fluorescence 8
4.2 Qualitative Analysis with the X-ray Spectrometer 9
4.3 Development of Intensity/ Concentration Algorithms 9
4.4 Quantitative Methods 10
4.5 Trace Analysis 11
5 Instrumentation for Powder Diffraction **11**
5.1 Basis of the Method 11
5.2 Production of Monochromatic Radiation 12
5.3 Use of Diffractometers in the Powder Method 13
5.4 Powder Cameras 13
6 Phase Characterization by X-ray Diffraction **14**
6.1 Qualitative Analysis of Polycrystalline Material 14
6.2 Search Procedures 14
6.3 Modern Computer-based Methods 15
6.4 Limitations Due to the Specimen 16
6.5 Quantitative Methods 16
6.6 Limitations of the Diffraction Method 17
7 Future Trends for X-ray Diffraction and Fluorescence **17**
7.1 Analysis of Small Specimens and Thin Films 17
7.2 Impact of the Synchrotron 18
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

The use of X-ray methods in the field of materials analysis is now entering its seventh decade. While the broad definition of X-ray methods covers many techniques based on the scatter, emission and absorption properties of X-radiation, the two most common are X-ray fluorescence (XRF) spectrometry and X-ray powder diffractometry (XRD). When a sample of material is bombarded with energetic radiation (X-rays, γ-rays, electrons, protons, etc.) vacancies may arise from the removal of inner orbital electrons. One of the processes by which the atom regains stability is by transference of electrons from outer to inner electron shells. Each of these transitions is accompanied by the emission of an X-ray photon having an energy equal to the energy difference between the two states. The X-ray emission wavelengths are characteristic of the atom in question and there is a simple relationship (Moseley's law) between the wavelength of the emission line and the atomic number of the atom. Thus when a sample is made up of many different types of atoms, each atom will produce a series of wavelengths, and all of the contributions add up to become the total X-ray emission from the sample. XRF is a technique which utilizes the diffracting power of a single crystal, or the proportional characteristics of a photon detector, to separate the polychromatic beam of radiation from the sample into separate wavelengths, thus allowing qualitative and quantitative elemental measurements to be made.

A beam of monochromatic radiation may also be scattered when X-ray photons collide with atomic electrons. Where the scattered wavelengths interfere with one another diffraction of X-rays occurs. All substances are built up of individual atoms and nearly all substances have some degree of order of periodicity in the arrangement of these atoms. It is the scattering from these periodic arrays that leads to the diffraction effect, and there is a simple relationship (Bragg's law) between the scattering

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

angle, the wavelength of the radiation and the spacings between the planes of atoms. Since the distances between the atomic planes are dependent on the size and distribution of atoms – i.e. the structure of the material, XRD can be used for qualitative and quantitative phase identification.

1 INTRODUCTION

1.1 Brief History of the Development of X-ray Fluorescence and Powder Diffraction Methods

X-ray photons are a form of electromagnetic radiation produced following the ejection of an inner orbital electron and subsequent transition of atomic orbital electrons from states of high to low energy. When a monochromatic beam of X-ray photons falls onto a given specimen three basic phenomena may result, namely absorption, scatter or fluorescence. The coherently scattered photons may undergo subsequent interference leading in turn to the generation of diffraction maxima. These three basic phenomena form the bases of three important X-ray methods: the absorption technique, which is the basis of radiographic analysis; the scattering effect, which is the basis of X-ray diffraction; and the fluorescence effect, which is the basis of XRF spectrometry.

X-rays were discovered by Wilhelm Roentgen in 1895[1] and the property of the atomic number dependence of the absorption of X-ray photons was quickly established and applied for medical diagnostic purposes. Following the discovery of the diffraction of X-rays by Max von Laue in 1913, two major fields of materials analysis have developed. XRF spectrometry uses either the diffracting power of a single crystal to isolate narrow wavelength bands, or a proportional detector to isolate narrow energy bands, from the polychromatic beam characteristic radiation excited in the sample. The first of these methods is called wavelength-dispersive spectrometry and the second, energy-dispersive spectrometry. Because of the known relationship between emission wavelength and atomic number, isolation of individual characteristic lines allows the unique identification of an element to be made and elemental concentrations can be estimated from characteristic line intensities. Thus this technique is a means of materials characterization in terms of chemical composition. X-ray spectrometric techniques provided important information for the theoretical physicist in the first half of this century and since the early 1950s they have found an increasing use in the field of materials characterization. While most of the early work in X-ray spectrometry was carried out using electron excitation,[2] today most stand-alone X-ray spectrometers use X-ray excitation sources rather than electron excitation. XRF spectrometry typically uses a polychromatic beam of short wavelength X-radiation to excite longer wavelength characteristic lines from the sample to be analyzed. Modern X-ray spectrometers use either the diffracting power of a single crystal to isolate narrow wavelength bands, or a proportional detector to isolate narrow energy bands, from the polychromatic radiation (including characteristic radiation) excited in the sample.[3]

The second field of materials analysis involves characterization by means of atomic arrangement in the crystal lattice. XRD uses single or multiphase specimens comprising a random orientation of small crystallites, each of the order of 1–50 μm in diameter. Each crystallite in turn is made up of a regular, ordered array of atoms. An ordered arrangement of atoms (the crystal lattice) contains planes of high atomic density which in turn means planes of high electron density. A monochromatic beam of X-ray photons will be scattered by these atomic electrons and if the scattered photons interfere with each other, diffraction maxima may occur. In general, one diffracted line will occur for each unique set of planes in the lattice. A diffraction pattern is typically in the form of a graph of diffraction angle (or interplanar spacing) against diffracted line intensity. The pattern is thus made up of a series of superimposed diffractograms, one for each unique phase in the specimen. Each of these unique patterns can act as an empirical ''fingerprint'' for the identification of the various phases, using pattern recognition techniques based on a file of standard single-phase patterns. Quantitative phase analysis is also possible, albeit with some difficulty because of various experimental and other problems, not the least of which is of the large number of diffraction lines occurring from multiphase materials.

1.2 Role of X-ray Methods in the Modern Analytical Laboratory

Wavelength-dispersive spectrometers have been commercially available since the early 1950s and there are probably about 20 000 units in use in the world today. Energy-dispersive spectrometer systems became available in the early 1970s, and there are probably about 8000 stand-alone spectrometers in use, with perhaps slightly more than this number attached to scanning electron microscopes. XRF analysis finds a wide range of application since it allows the quantitation of all elements in the periodic table from F (atomic number 9) upwards.[4] Newer developments allow the determination of the ultralow atomic number elements including B, C, O and N. Accuracies of a few tenths of one percent are possible for most of the atomic number range, and elements are detectable in many cases to the low parts per million level. Excellent data treatment software is available allowing the rapid application of quantitative and semi-quantitative procedures.

X-ray powder diffractometers have been available in their modern form since the late 1940s, although camera systems were available back to the 1920s. There are probably of the order of 30 000 powder diffractometers in use – about half of these being automated to some degree. XRD is applicable to any ordered (crystalline) material and although much less accurate or sensitive than the fluorescence method, is almost unique in its ability to differentiate phases. Quantitative phase analysis is possible, but the accuracy of a determination is often compromised by preferred orientation of the specimen.

The fluorescence and diffraction techniques are to a large extent complementary, since one allows accurate quantitation of elements to be made and the other allows qualitative and semiquantitative estimations to be made of the way in which the matrix elements are combined to make up the phases in the specimen. Thus a combination of the two techniques will often allow the accurate determination of the material balance of a sample.

2 PROPERTIES OF X-RADIATION

2.1 Continuous and Characteristic Radiation

When a high-energy electron beam is incident upon a specimen, one of the products of the interaction is an emission of a broad-wavelength band of radiation called continuum, also referred to as white radiation or bremsstrahlung. This white radiation is produced as the impinging high-energy electrons are decelerated by the atomic electrons of the elements making up the specimen. The intensity/wavelength distribution of this radiation is typified by a minimum wavelength λ_{min} which is inversely proportional to the maximum accelerating potential V of the electrons. The intensity distribution of the continuum reaches a maximum at a wavelength 1.5 to 2 times greater than λ_{min}. Increasing the accelerating potential causes the intensity distribution of the continuum to shift towards shorter wavelengths. Most commercially available spectrometers utilize a sealed X-ray tube as an excitation source, and these tubes typically employ a heated tungsten filament as a source of electrons, and a layer of pure metal, such as chromium, rhodium or tungsten, as the anode. The broad band of white radiation produced by this type of tube is ideal for the excitation of the characteristic lines from a wide range of atomic numbers. In general, the higher the atomic number of the anode material, the more intense the beam of radiation produced by the tube. X-ray diffractometers typically employ a more focused filament assembly (12×0.05 mm) and run at a somewhat lower loading than the XRF tube. A key factor in the design of an X-ray tube for powder diffraction is the *specific loading* (watts per mm^2) of the tube.

2.2 X-ray Spectra

In addition to electron interactions leading to the production of white radiation, there are also electron interactions which produce characteristic radiation. If a high-energy particle, such as an electron, strikes a bound atomic electron, and the energy of the particle is greater than the binding energy of the atomic electron, it is possible that the atomic electron will be ejected from its atomic position, departing from the atom with a kinetic energy $(E - \phi)$ equivalent to the difference between the energy E of the initial particle and the binding energy ϕ of the atomic electron. Where the exciting particles are X-ray photons, the ejected electron is called a photoelectron and the interaction between primary X-ray photons and atomic electrons is called the photoelectric effect. As long as the vacancy in the shell exists, the atom is in an unstable state and can regain stability by transference of an electron from one of the outer orbitals to fill the vacancy. The energy difference between the initial and final states of the transferred electron may be given off in the form of an X-ray photon. Since all emitted X-ray photons have energies proportional to the differences in the energy states of atomic electrons, the lines from a given element will be characteristic of that element. The relationship (Equation 1) between the wavelength of a characteristic X-ray photon and the atomic number Z of the excited element was first established by Moseley.[5]

$$\frac{1}{\lambda} = K(Z - \sigma)^2 \tag{1}$$

in which K is a constant that takes on different values for each spectral series. σ is the shielding constant that has a value of just less than unity. The wavelength of the X-ray photon is inversely related to the energy E of the photon according to Equation (2):

$$\lambda(\text{Å}) = \frac{12.4}{E(\text{keV})} \tag{2}$$

Not all vacancies result in the production of characteristic X-ray photons since there is a competing internal rearrangement process known as the Auger effect.[6] The ratio of the number of vacancies resulting in the production of characteristic X-ray photons to the total number of vacancies created in the excitation process is called the fluorescent yield. Fluorescent yield values are several orders of magnitude less for the very low atomic numbers. In practice this means that if, for example, one were to compare the intensities obtained from pure barium ($Z = 56$) and pure aluminum ($Z = 13$), all other things being equal, pure barium would give about 50 times more counts than would pure aluminum. The L fluorescent yield for a given atomic number is always less by about a factor of three than the corresponding K

fluorescent yield, where K and L refer to the first and second electron shells. An excited atom can revert to its original ground state by transferring an electron from an outer atomic level to fill the vacancy in the inner shell. An X-ray photon is emitted from the atom as part of this de-excitation step, the emitted photon having an energy equal to the energy difference between the initial and final states of the transferred electron. The selection rules for the production of normal (diagram) lines require that the principal quantum number n must change by at least one, the angular quantum number ℓ must change by ± 1, and the J quantum number (the total momentum 'J' of an electron is given by the vector sum of $\ell + s$ where s is the spin quantum number) must change by 0 or 1. In effect this means that for the K series only p → s transitions are allowed, yielding two lines for each principal level change. Vacancies in the L level follow similar rules and give rise to L series lines. There are more of the L lines since p → s, s → p and d → p transitions are all allowed within the selection rules. Transition groups may now be constructed, based on the appropriate number of transition levels.

Figure 1 shows plots of the reciprocal of the square root of the wavelength, as a function of atomic number, for the K, L and M series. As indicated by Moseley's law such plots should be linear. A scale directly in wavelength is also shown, to indicate the range of wavelengths over which a given series occurs. In practice, the number of lines observed from a given element will depend upon the atomic number of the element, the excitation conditions and the wavelength range of the spectrometer employed.

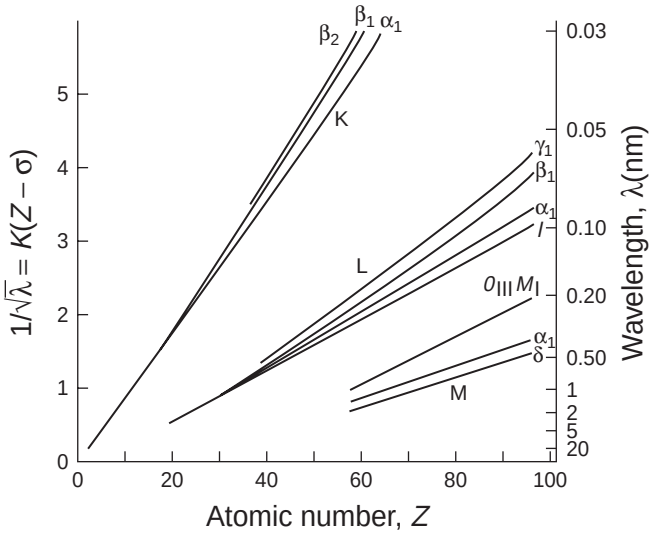

Figure 1 Moseley diagrams for the K, L and M series. (Reproduced from Jenkins, *X-ray Fluorescence Spectrometry*, 2nd edition, Wiley-Interscience, 1999. With permission from John Wiley & Sons.)

Generally, commercial spectrometers cover the K series, the L series and the M series, corresponding to transitions to K, L and M levels respectively. There are many more lines in the higher series and for a detailed list of all of the reported wavelengths the reader is referred to the work of Bearden.[7] In X-ray spectrometry most of the analytical work is carried out using either the K or the L series wavelengths.

While most of the observed fluorescent lines are normal, certain lines may also occur in X-ray spectra that do not at first sight fit the basic selection rules. These lines are called forbidden lines and are shown in the center portion of Figure 1. Forbidden lines typically arise from outer orbital levels where there is no sharp energy distinction between orbitals. A third type of line may also occur – called satellite lines – which arise from dual ionizations. Neither forbidden transitions nor satellite lines have great analytical significance; they may cause some confusion in qualitative interpretation of spectra and may even be misinterpreted as coming from trace elements. Most commercially available X-ray spectrometers have a range from about 0.2 to 20 Å (60–0.6 keV, $1\ \text{Å} = 10^{-10}\ \text{m} = 10^{-1}\ \text{nm}$) which will allow measurement of the K series from F ($Z = 9$) to Lu ($Z = 71$), and for the L series from Mn ($Z = 25$) to U ($Z = 92$). Other line series can occur from the M and N levels but these have little use in analytical X-ray spectrometry.

2.3 Absorption

When a beam of X-ray photons falls onto an absorber, a number of different processes may occur. The more important of these are illustrated in Figure 2. In this example, a monochromatic beam of radiation of wavelength λ_0 and intensity I_0, is incident on an absorber of thickness x_a and density ρ_a. The fate of each individual X-ray photon is governed by the following processes. Absorption occurs where only a certain fraction (I/I_0) of the radiation may pass through the absorber. Where

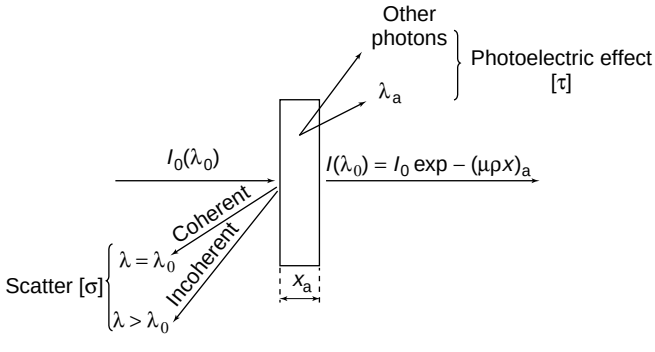

Figure 2 Interaction of X-ray photons with matter.

this happens the wavelength of the transmitted beam is unchanged and the intensity of the transmitted beam $I(\lambda_0)$ is given by Equation (3):

$$I(\lambda_0) = I_0 \exp -(\mu_a \rho_a x_a) \quad (3)$$

where μ_a is the mass attenuation coefficient of absorber a for the wavelength λ_0.

It will be apparent from the above that a number of photons equal to $(I_0 - I)$ have been lost in the absorption process, most of this loss being due to the photoelectric effect. Photoelectric absorption, usually designated τ, will occur at each of the energy levels of the atom. Thus the total photoelectric absorption will be determined by the sum of each individual absorption within a specific shell. The value of the mass attenuation μ referred to in Equation (3) is a function of both the photoelectric absorption and the scatter, Equation (4):

$$\mu = f(\tau) + f(\sigma) \quad (4)$$

However, $f(\tau)$ is usually large in comparison with $f(\sigma)$. Because the photoelectric absorption is made up of absorption in the various atomic levels, it is an atomic-number-dependent function. A plot of μ against λ contains a number of discontinuities called absorption edges, at wavelengths corresponding to the binding energies of the electrons in the various subshells. The absorption discontinuities are a major source of non-linearity between X-ray intensity and composition in both XRF and XRD.

2.4 Scattering

Scatter, σ, will occur when an X-ray photon collides with one of the electrons of the absorbing element. Where this collision is elastic, i.e. when no energy is lost in the collision process, the scatter is said to be coherent (Rayleigh) scatter. Since no energy change is involved, the coherently scattered radiation will retain exactly the same wavelength as the incident beam. As will be shown later, X-ray diffraction is a special case of coherent scatter, where the scattered photons interfere with each other. It can also happen that the scattered photon gives up a small part of its energy during the collision, especially where the electron with which the photon collides is only loosely bound. In this instance the scatter is said to be incoherent (Compton scatter) and the wavelength of the incoherently scattered photon will be greater than λ_0.

2.5 Role of Crystal Structure in X-ray Scattering and Diffraction

All substances are built up of individual atoms and nearly all substances have some degree of order of periodicity in the arrangement of these atoms. A crystal can be defined as a homogeneous, anisotropic body having the natural shape of a polyhedron. In practical terms, whether a substance is homogeneous or not can only be defined by the means that are available for measuring the crystallinity. In general, the shorter the wavelength, the smaller the crystalline region that is able to be recognized. Even noncrystalline materials have a degree of order and each will give some sort of a diffraction pattern. For example, glassy materials and liquids will give diffraction patterns of sorts, generally in the form of one or more broad diffuse peaks or halos. In XRD one is generally dealing exclusively with crystalline materials. The diffraction pattern can, moreover, be used to determine the degree of crystallinity, that is the dimensions of the crystalline regions in otherwise amorphous substances. A crystalline substance has a definite form which is retained no matter what the physical size of the crystal. A certain type of crystal can thus be defined in terms of specific physical characteristics which determine its shape. For example, a sodium chloride crystal has cubic symmetry and by definition the angles between all of the principal faces should be 90°. In fact there are other crystal classes which also have angles between faces of 90°, but the cube is unique in that the lengths of the sides are also equal. If our sodium chloride crystal were reduced in size until the smallest repeat unit was found (the so-called *unit cell*) it would indeed be found that all sides were equal. Hence it is apparent that a crystal type can be defined in terms of the lengths of the sides of its unit cell and the angles between the faces. Since every ordered material is made up of a unique arrangement and number of atoms, every ordered material will give a diffraction pattern which is, to all intents and purposes, also unique.

Figure 3 illustrates in a simplified manner the diffraction patterns that would be obtained from (a) a single

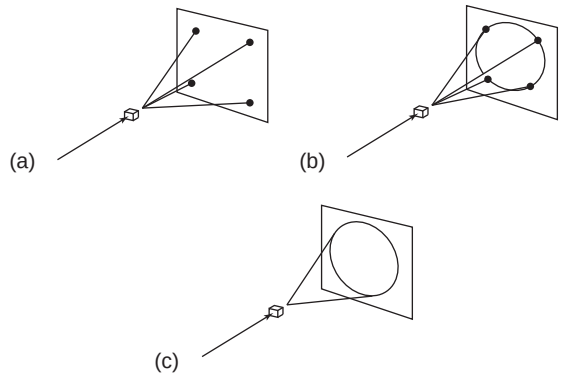

Figure 3 Diffraction patterns from (a) single crystal; (b) a random powder; and (c) an oriented powder. (Reproduced from Jenkins, Snyder, *Introduction to Powder Diffractometry*, Wiley-Interscience, 1996. With permission from John Wiley & Sons.)

crystal, (b) an oriented powder and (c) a random powder. In each instance the specimen is placed between a monochromatic beam of radiation and a piece of photographic film. In the case of the single crystal, because only certain crystal planes are in the position to diffract radiation onto the film, the pattern will appear as a series of spots on the film, the position of the spots being dependent upon the structure and orientation of the crystal. Changing the position of the crystal will bring other planes into diffracting position and the pattern of spots will change. In the case of a random powder, whatever the orientation of the specimen there is always a sufficient number of crystallites to diffract from the appropriate *d*-spacing(s) (see section 2.6). In the case of an oriented powder, however, there are more crystallites in certain orientations and less in others. Thus the pattern obtained is somewhere between the single-crystal and the random-powder patterns.

2.6 Interference and Diffraction Effects

As illustrated in Figure 4(a), a crystal lattice consists of a regular arrangement of atoms, with layers of high atomic density existing throughout the crystal structure. Planes of high atomic density means planes of high electron density. Since scattering occurs between impinging X-ray photons and the loosely bound outer orbital atomic electrons, when a monochromatic beam of radiation falls onto the high atomic density layers scattering will occur. In order to satisfy the requirement for constructive interference, it is necessary that the scattered waves originating from the individual atoms, the scattering points, be in phase with one another. The geometric conditions for this condition to occur are illustrated in Figure 4(b). Here, a series of parallel rays strike a set of crystal planes at an angle θ and are scattered as previously described. Reinforcement will occur when the difference in the path lengths of the two interfering waves is equal to a whole number of wavelengths. This path length difference is $2d\sin\theta$ where d is the inter-planar spacing; hence the overall condition (Equation 5) for reinforcement is that:

$$n\lambda = 2d\sin\theta \tag{5}$$

where n is an integer. Equation (5) is a statement of Bragg's law. Diffraction experiments are generally made at a fixed wavelength, thus a measure of the diffraction angles will allow the associated *d*-spacings to be calculated. Bragg's law is also important in wavelength-dispersive spectrometry since by using a crystal of fixed $2d$, each unique wavelength will be diffracted at a unique diffraction angle. Thus, measuring the diffraction angle 2θ allows the determination of the wavelength (inter-planar spacing) if the *d*-spacing of the crystal is known. Since there is a simple relationship between wavelength and atomic number, one can establish the atomic number(s) of the element(s) from which the wavelengths were emitted.

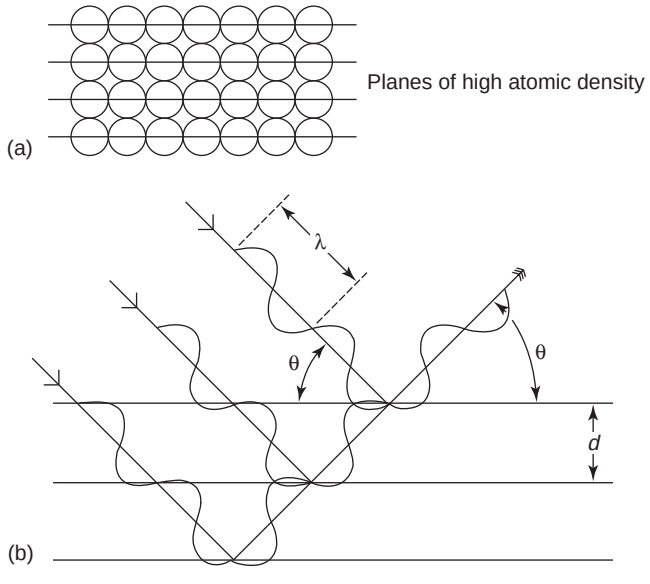

Figure 4 Scatter and diffraction from an ordered arrangement of atoms. (Reproduced from Jenkins, *X-ray Fluorescence Spectrometry*, 2nd edition, Wiley-Interscience, 1999. With permission from John Wiley & Sons.)

3 INSTRUMENTATION FOR X-RAY FLUORESCENCE ANALYSIS

X-ray spectrometers fall roughly into two categories: wavelength-dispersive instruments and energy-dispersive instruments. The wavelength-dispersive system was introduced commercially in the early 1950s and since the mid-1970s has developed into a widely accepted analytical tool. Analytical chemists have available to them a wide range of instruments for the qualitative and quantitative analysis of multielement samples, and in the choice of technique they will generally consider such factors as sensitivity, speed, accuracy, cost, range of applicability, and so on. Within the two major categories of X-ray spectrometers specified, there is of course a wide diversity of instruments available but most of these will fit the four basic types:

- simultaneous wavelength-dispersive
- sequential wavelength-dispersive
- bremsstrahlung source energy-dispersive
- secondary target energy-dispersive.

Although these different types of instrument may bear little physical resemblance to one another, for the purpose of comparison they differ only in the type of source used

for excitation, the number of elements which they are able to measure at one time, the speed at which they collect data and finally, their price range. All of the instruments listed above are, in principle at least, capable of measuring all elements in the periodic classification from $Z = 9\,^{\circ}$(F) and upwards. All can be fitted with multisample handling facilities and all can be automated by use of minicomputers. All are capable of precision of the order of a few tenths of one percent and all have sensitivities down to the low parts per million level. As far as the analyst is concerned, they differ only in their speed, cost and number of elements measurable at the same time.

Single-channel wavelength-dispersive spectrometers are typically employed for both routine and nonroutine analysis of a wide range of products, including ferrous and nonferrous alloys, oils, slags and sinters, ores and minerals, thin films, and so on. These systems are very flexible but relative to multichannel spectrometers are somewhat slow. The multichannel wavelength-dispersive instruments are used solely for routine, high throughput, analyses where the great need is for fast accurate analysis, but where flexibility is of no importance. Energy-dispersive spectrometers have the great advantage of being able to display information on all elements at the same time. They lack somewhat in resolution over the wavelength-dispersive systems but also find great application in quality control, troubleshooting problems, and so on. They have been particularly effective in the fields of scrap alloy sorting, in forensic science and in the provision of elemental data to supplement X-ray powder diffraction data.

3.1 Basis of the Method

All conventional X-ray spectrometers comprise three basic parts: the primary source unit, the spectrometer itself and the measuring electronics. The primary source unit consists of a very stable high-voltage generator, capable of providing up to around 3 kW of power at a potential of typically 60–80 kV, plus a sealed X-ray tube. The sealed X-ray tube has an anode of Cr, Rh, W, Ag, Au or Mo and delivers an intense source of continuous radiation which then impinges on the analyzed specimen where characteristic radiation is generated. A portion of the characteristic 'fluorescence' radiation is then collected by the actual spectrometer where the beam is passed via a collimator or slit onto the surface of an analyzing crystal, where individual wavelengths are diffracted in accordance with the Bragg law. A photon detector, typically a gas flow or a scintillation counter, is then used to convert the diffracted characteristic photons into voltage pulses which are integrated and displayed as a measure of the characteristic line intensity.

3.2 X-ray Sources

Several different types of source have been employed for the excitation of characteristic X-radiation, including those based on electrons, X-rays, γ-rays, protons and synchrotron radiation. Sometimes a bremsstrahlung X-ray source is used to generate specific X-radiation from an intermediate pure element sample called a secondary fluorescer. By far the most common source today is the X-ray photon source. This source is used in primary mode in the wavelength- and primary energy-dispersive systems, and in secondary fluorescer mode in secondary target energy-dispersive spectrometers. A γ-source is typically a radioisotope that is used either directly, or in a mode equivalent to the secondary fluorescer mode in energy-dispersive spectrometry. Most conventional wavelength-dispersive X-ray spectrometers use a high power (2–4 kW) X-ray bremsstrahlung source. Energy-dispersive spectrometers use either a high-power or low-power (0.5–1.0 kW) primary source, depending on whether the spectrometer is used in the secondary or primary mode. In all cases, the primary source unit consists of a very stable high-voltage generator, capable of providing a potential of typically 40–100 kV. The current from the generator is fed to the filament of the X-ray tube, which is typically a coil of tungsten wire. The applied current causes the filament to glow, emitting electrons. A portion of this electron cloud is accelerated to the anode of the X-ray tube, which is typically a water-cooled block of copper with the required anode material plated or cemented to its surface. The impinging electrons produce X-radiation, a significant portion of which passes through a thin beryllium window to the specimen.

3.3 X-ray Detectors

An X-ray detector is a transducer for converting X-ray photon energy into voltage pulses. Detectors work through a process of photoionization in which interaction between the entering X-ray photon and the active detector material produces a number of electrons. The current produced by these electrons is converted to a voltage pulse, by a capacitor and resistor, such that one digital voltage pulse is produced for each entering X-ray photon. In addition to being sensitive to the appropriate photon energies, which means being applicable to a given range of wavelengths or energies, there are two other important properties that an ideal detector should possess. These properties are proportionality and linearity. Each X-ray photon entering the detector produces a voltage pulse, and where the size of the voltage pulse is proportional to the photon energy, the detector is said to be proportional. Proportionality is needed where the technique of pulse height selection is to be used. Pulse height selection is a means of electronically rejecting

pulses of voltage levels other than those corresponding to the characteristic line being measured. X-ray photons enter the detector at a certain rate and where the output pulses are produced at this same rate the detector is said to be linear. Linearity is important where the various count rates produced by the detector are to be used as measures of the photon intensity for each measured line.

In the case of wavelength-dispersive spectrometers, the gas flow proportional counter is generally employed for the measurement of longer wavelengths and the scintillation counter is used for the measurement of shorter wavelengths. Neither of these detectors has sufficient resolution to separate multiple wavelengths on its own, and so they have to be employed along with an analyzing crystal. However, in the case of energy-dispersive spectrometry, where no dispersing crystal is employed, a detector of higher resolution must be used.

3.4 The Wavelength-dispersive Spectrometer

In order to maintain the required geometric conditions a goniometer is used to ensure that the angle between source and crystal, and crystal and detector, is kept the same. The output from a wavelength-dispersive spectrometer may be either analog or digital. For qualitative work an analog output is traditionally used, and in this instance a rate meter is used to integrate the pulses over short time intervals, typically of the order of a second or so. The output from the rate meter is fed to an x/t recorder which scans at a speed which is conveniently coupled with the goniometer scan speed. The recorder thus displays an intensity/time diagram, which becomes an intensity/2θ diagram. Tables are then used to interpret the wavelengths. For quantitative work it is more convenient to employ digital counting, and a timer/scaler combination is provided which will allow pulses to be integrated over a period of several tens of seconds and then displayed as count or count rate.

Most modern wavelength-dispersive spectrometers are controlled in some way by a minicomputer or microprocessor and by use of specimen changers are capable of very high specimen throughput. Once they are set up the spectrometers will run virtually unattended for several hours.

3.5 The Energy-dispersive Spectrometer

Like the wavelength-dispersive spectrometer, the energy-dispersive spectrometer also consists of the three basic units – excitation source, spectrometer and detection system. In this case, however, the detector itself acts as the dispersion agent. The detector generally employed in this context is the Si(Li) detector. The Si(Li) detector consists of a small cylinder (about 1 cm diameter and 3 mm thick) of p-type silicon that has been compensated by lithium to increase its electrical resistivity. A Schottky barrier contact on the front of the silicon disk produces a p-i-n type diode (one with p-doped, isolating and n-doped layers). In order to inhibit the mobility of the lithium ions and to reduce electronic noise, the diode and its pre-amplifier are cooled to the temperature of liquid nitrogen. Incident X-ray photons interact to produce a specific number of electron hole pairs. The charge produced is swept from the diode by the bias voltage to a charge-sensitive pre-amplifier. A charge loop integrates the charge on a capacitor to produce an output pulse. A pre-amplifier is responsible for collecting this charge on a feedback capacitor to produce a voltage pulse proportional to the original X-ray photon energy. Thus when a range of photon energies are incident upon the detector, an equivalent range of voltage pulses is produced as a detector output. A multichannel analyzer is used to sort the arriving pulses at its input in the same fashion as to produce a histogram representation of the X-ray energy spectrum.

The output from an energy-dispersive spectrometer is generally displayed on some sort of visual display unit. The operator is able to dynamically display the contents of the various channels as an energy spectrum and provision is generally made to allow zooming in on portions of the spectrum of special interest, to overlay spectra, to subtract background, and so on, in a rather interactive manner. As in the case of the modern wavelength-dispersive systems, nearly all energy-dispersive spectrometers will incorporate some form of minicomputer which is available for spectral stripping, peak identification, quantitative analysis and a host of other useful functions.

4 ELEMENTAL ANALYSIS BY X-RAY FLUORESCENCE

4.1 Specimen Preparation Techniques for X-ray Fluorescence

Because X-ray spectrometry is essentially a comparative method of analysis, it is vital that all standards and unknowns be presented to the spectrometer in a reproducible and identical manner. Any method of specimen preparation must give specimens that are reproducible and which, for a certain calibration range, have similar physical properties including mass attenuation coefficient, density, particle size and particle homogeneity. In addition the specimen preparation method must be rapid and cheap and must not introduce extra significant systematic errors, for example the introduction of trace elements from contaminants in a diluent. Specimen preparation is an all important factor in the ultimate accuracy of any X-ray determination, and many papers have been published

describing a multitude of methods and recipes for sample handling.

In general samples fit into three main categories:

- Samples which can be handled directly following some simple pretreatment such as pelletizing or surfacing, for example homogeneous samples of powders, bulk metals or liquids.
- Samples which require significant pretreatment, for example heterogeneous samples, samples requiring matrix dilution to overcome inter-element effects and samples exhibiting particle size effects.
- Samples which require special handling treatment, for example samples of limited size, samples requiring concentration or prior separation and radioactive samples.

The ideal specimen for XRF analysis is one in which the analyzed volume of specimen is representative of the total specimen, which is, itself, representative of the sample submitted for analysis. There are many forms of specimen suitable for XRF analysis, and the form of the sample as received will generally determine the method of pretreatment. It is convenient to refer to the material received for analysis as the sample, and that which is actually analyzed in the spectrometer as the specimen. While the direct analysis of certain materials is certainly possible, more often than not some pretreatment is required to convert the sample to the specimen. This step is referred to as specimen preparation. In general, the analyst would prefer to analyze the sample directly, because if it is taken as received any problems arising from sample contamination that might occur during pretreatment are avoided. In practice, however, there are three major constraints that may prevent this ideal circumstance from being achieved: sample size, sample size homogeneity and sample composition heterogeneity. Problems of sample size are frequently severe in the case of bulk materials such as metals, large pieces of rock, etc.

Problems of sample composition heterogeneity will generally occur under these circumstances as well, and in the analysis of powdered materials, heterogeneity must almost always be considered. The sample as received may be either homogeneous or heterogeneous; in the latter case, it may be necessary to render the sample homogeneous before an analysis can be made. Heterogeneous bulk solids are generally the most difficult kind of sample to handle, and it may be necessary to dissolve or chemically react the material in some way to give a homogeneous preparation. Heterogeneous powders are either ground to a fine particle size and then pelletized, or fused with a glass-forming material such as borax. Solid material in liquids or gases must be filtered out and the filter analyzed as a solid. Where analyte concentrations in liquids or solutions are too high or too low, dilution or preconcentration techniques may be employed to bring the analyte concentration within an acceptable range.

4.2 Qualitative Analysis with the X-ray Spectrometer

Both the simultaneous wavelength-dispersive spectrometer and the energy-dispersive spectrometers lend themselves admirably to the qualitative analysis of materials. As was shown in Equation (1) there is a simple relationship between the wavelength or energy of a characteristic X-ray photon, and the atomic number of the element from which the characteristic emission line occurs. Thus by measuring the wavelengths, or energies, of a given series of lines from an unknown material the atomic numbers of the excited elements can be established. Because the characteristic X-ray spectra are so simple the actual process of allocating atomic numbers to the emission lines is a relatively simple process and the chance of making a gross error is rather small. In comparison, the procedures for the qualitative analysis of multiphase materials with the X-ray powder diffractometer is a much more complex business. There are after all only 100 or so elements, and within the range of the conventional spectrometer each element gives, on an average, only half a dozen lines. In diffraction on the other hand, there are perhaps as many as several million possible compounds, each of which can give on average 50 or so lines. Similarly if one compares the X-ray emission spectrum with the ultraviolet (UV) emission spectrum, the X-ray spectrum is much easier to interpret than the UV emission spectrum because the X-ray spectrum arises from inner orbital transitions, the number of which is of course rather limited, and the UV spectrum arises from transitions to unfilled levels, of which there can be many. A further benefit of the X-ray emission spectrum for qualitative analysis is that because transitions do arise from inner orbitals, the effect of chemical combination, i.e. valence, is almost negligible.

4.3 Development of Intensity/Concentration Algorithms

The great flexibility and range of the various types of XRF spectrometer, coupled with their high sensitivity and good inherent precision make them ideal for quantitative analysis. Like all instrumental methods of analysis the high precision can be translated into high accuracy only if the various systematic errors in the analysis process are taken care of. The precision of a well-designed X-ray spectrometer is typically of the order of one tenth of a percent or so, the major source of this random error being the X-ray source, the high voltage generator plus the X-ray tube. In addition there is an error arising from the statistics of the actual counting process.

Systematic errors in quantitative X-ray spectrometry arise mainly from absorption- and specimen-related phenomena, rather like in X-ray powder diffraction except that in the spectrometry case they are much more complicated. This is because in spectrometry many wavelengths are involved, whereas in diffraction one is dealing with a single wavelength, i.e. the diffracted, monochromatic tube line. This means, for example, that the equivalent absorption-effect expression becomes rather complex. Although these so-called matrix effects are somewhat complicated, many excellent methods have been developed over the past 30 years or so for handling them. The advent of the personal computer controlled spectrometer has done much to enhance the application of these correction procedures and today, in most cases, one is able to quantify most elements in the periodic table of atomic number 9 (F) and upwards to an accuracy of a few tenths of a percent. The areas of application of the XRF technique now cover almost all areas of inorganic analysis.

In the conversion of net line intensity to analyte concentration it may be necessary to correct for any absorption and/or enhancement effects which occur. Absorption effects include both primary and secondary absorption. Primary absorption occurs because all atoms of the specimen matrix will absorb photons from the primary source. Since there is a competition for these primary photons by the atoms making up the specimen, the intensity/wavelength distribution of these photons available for the excitation of a given analyte element may be modified by other matrix elements. Secondary absorption refers to the effect of the absorption of characteristic analyte radiation by the specimen matrix. As characteristic radiation passes out from the specimen in which it was generated, it will be absorbed by all matrix elements, by amounts relative to the mass attenuation coefficients of these elements. The total absorption of a specimen is dependent on both primary and secondary absorption. Enhancement effects occur when a non-analyte matrix element A emits a characteristic line that has an energy just in excess of the absorption edge of the analyte element. This means that the nonanalyte element in question is able to excite the analyte, giving characteristic photons over and above those produced by the primary continuum. This gives an increased, or enhanced, signal from the analyte.

Since the early work of Sherman[8] there has been a growing interest in the provision of an intensity/concentration algorithm which would allow the calculation of the concentration values without recourse to the use of standards. Sherman's work was improved upon first by the Japanese team of Shiraiwa and Fujino[9] and later, by Criss and Birks[10,11] with their Naval Research Laboratory X-ray Fluorescence program (NRLXRF). The same group also solved the problem of describing the intensity distribution from the X-ray tube.[12] The problem for the average analyst in the late 1960s and early 1970s, however, remained that of finding sufficient computational power to apply these methods. In the early 1970s, de Jongh suggested an elegant solution[13] in which he proposed the use of a large mainframe computer for the calculation of the influence coefficients, then use of a minicomputer for their actual application using a concentration correction influence model.

4.4 Quantitative Methods

The simplest quantitative analysis situation to handle is the determination of a single element in a known matrix. In this instance, a simple calibration curve of analyte concentration vs line intensity is sufficient for quantitative determination. A slightly more difficult case might be the determination of a single element where the matrix is unknown. Three basic methods are commonly employed in this situation; use of internal standards, use of standard addition or use of a scattered line from the X-ray source. The most complex case is the analysis of all, or most, of the elements in a sample, about which little or nothing is known. In this case a full qualitative analysis would be required before any attempt is made to quantitate the matrix elements. Once the qualitative composition of the sample is known, again, one of three general techniques is typically applied; use of type standardization, use of an influence coefficient method or use of a fundamental parameter technique. Both the influence coefficient method and the fundamental parameter technique require a computer for their application. In principle, an empirical correction procedure can be described as the correction of an analyte element intensity for the influence of an interfering element(s) using the product of the intensity from the interfering element line and a constant factor as the correction term.[14] This constant factor is generally referred to as an influence coefficient, since it is assumed to represent the influence of the interfering element on the analyte. Commonly employed influence coefficient methods may use either the intensity or the concentration of the interfering element as the correction term. These methods are referred to as intensity correction and concentration correction methods, respectively. Intensity correction models give a series of linear equations which do not require much computation, but they are generally not applicable to wide ranges of analyte concentration. Various versions of the intensity correction models found initial application in the analysis of nonferrous metals where correction constants were applied as look-up tables. Later versions[15] were supplied on commercially available computer controlled spectrometers and were used for a wider range of application. A concentration

model[16] requires the solving of a series of simultaneous equations, by regression analysis or matrix inversion techniques. This approach is more rigorous than the intensity models, and so they became popular in the early 1970s as suitable low-cost minicomputers became available.

Influence coefficient correction procedures can be divided into three basic types: fundamental, derived and regression. Fundamental models are those which require starting with concentrations, then calculating the intensities. Derived models are those which are based on some simplification of a fundamental method but which still allow concentrations to be calculated from intensities. Regression models are those which are semi-empirical in nature, and which allow the determination of influence coefficients by regression analysis of data sets obtained from standards. All regression models have essentially the same form and consist of a weight fraction term, W (or concentration, C); an intensity (or intensity ratio) term, I; an instrument-dependent term which essentially defines the sensitivity of the spectrometer for the analyte in question, and a correction term which corrects the instrument sensitivity for the effect of the matrix. The general form is given by Equation (6):

$$W = I\langle\text{instrument}\rangle[1 + \{\text{model}\}] \tag{6}$$

Equation (6) shows that the weight fraction W of an analyte is proportional to the product of measured intensity I from the analyte, corrected for instrumental effects; and a matrix correction term. Different models vary only in the form of this correction term.

The major advantage to be gained by use of influence coefficient methods is that a wide range of concentration ranges can be covered using a relatively inexpensive computer for the calculations. A major disadvantage is that a large number of well-analyzed standards may be required for the initial determination of the coefficients. However, where adequate precautions have been taken to ensure correct separation of instrument- and matrix-dependent terms, the correction constants are transportable from one spectrometer to another and, in principle, need only be determined once.

4.5 Trace Analysis

The XRF method is particularly applicable to the qualitative and quantitative analysis of low concentrations of elements in a wide range of samples, as well as allowing the analysis of elements at higher concentrations in limited quantities of materials. A measurement of a line at peak position gives a counting rate which, in those cases where the background is insignificant, can be used as a measure of the analyte concentration. However, where the background is significant the measured value of the analyte line at the peak position now includes a count rate contribution from the background. The analyte concentration in this case is related to the net counting rate. Since both peak and background count rates are subject to statistical counting errors, the question now arises as to the point at which the net peak signal is statistically significant. The generally accepted definition for the lower limit of detection (LLD) is that concentration equivalent to two standard deviations of the background counting rate. A formula (Equation 7) for the LLD can now be derived:[17]

$$\text{LLD} = \frac{3/m}{R_\text{b}/t_\text{b}} \tag{7}$$

In Equation (7) R_b is the count rate at the background, t_b is the time spent counting the background and m is the sensitivity of the spectrometer in c/s/%. Even though we are using two standard deviations, two measurements, peak and background, must be made. This increases the error by $\sqrt{2}$. It is the practice to assume that $2\sqrt{2} \sim 3$, hence the '3' in Equation (7). It is important to note that not only does the sensitivity of the spectrometer vary significantly over the wavelength range of the spectrometer, but so too does the background counting rate. In general, the background varies by about two orders of magnitude over the range of the spectrometer. It will be seen that the detection limit will be best when the sensitivity is high and the background is low. Both the spectrometer sensitivity and the measured background vary with the average atomic number of the sample. While detection limits over most of the atomic number range lie in the low parts per million range, the sensitivity of the X-ray spectrometer falls off quite dramatically towards the long wavelength limit of the spectrometer owing mainly to low fluorescence yields and the increased influence of absorption. As a result, poorer detection limits are found at the long wavelength extreme of the spectrometer, which corresponds to the lower atomic numbers. Thus the detection limits for elements such as fluorine and sodium are at the levels of hundredths of one percent rather than parts per million. The detection limits for the very low atomic number elements such as carbon ($Z = 6$) and oxygen ($Z = 7$) are very poor, typically of the order of 3 to 5%.

5 INSTRUMENTATION FOR POWDER DIFFRACTION

5.1 Basis of the Method

The powder method derives its name from the fact that the specimen is typically in the form of a microcrystalline powder, although as has been indicated, any material

which is made up of an ordered array of atoms will give a diffraction pattern. The possibility of using a diffraction pattern as a means of phase identification was recognized over 50 years ago but it was not until the late 1930s that a systematic means of unscrambling the superimposed diffraction patterns was proposed.[18] The technique was based on the use of a file of single phase patterns, characterized in the first stage by their three strongest reflections and a search technique based on matching strong lines in the unknown pattern with these standard pattern lines. A potential match was then confirmed by a check using the full pattern in question. The identified pattern was then subtracted from the experimental pattern and the procedure repeated on the residue pattern until all lines were identified. Techniques for this ''search/matching'' process have changed little over the years and, although in the hands of experts manual search/matching is an extremely powerful tool, for the less experienced user it can be rather time-consuming. Round robin tests have indicated that 2–4 hours are typically required for the complete identification of a four-phase mixture.[19] A growing complication is that the file of standard patterns increases by about 1500 each year and currently stands at about 118 000 entries. The responsibility for the maintenance of the Powder Data File (PDF) lies with the International Centre for Diffraction Data (ICDD) which is a not-for-profit organization located in Newtown Square, Pennsylvania. This group is made up of a staff of permanent officers along with a number of academic and industrial scientists who are active in the field of XRD. The PDF is a unique assembly of good-quality single-phase patterns and is used by thousands of chemists, geologists, materials scientists, etc. all over the world.

5.2 Production of Monochromatic Radiation

With the growing need for high-quality X-ray powder diffraction data, one especially important parameter is the method by which the analytical wavelength is treated. Problems can occur in data treatment because of the polychromatic nature of the diffracted beam and the variability in the angular dispersion of the diffractometer. A combination of these two facts can lead to difficulty in manually assessing where the maximum of a peak occurs, especially in the range of angles from 30° to 60° 2θ. Of the many methods commonly employed to render the radiation monochromatic (or bichromatic), most will fit into two broad categories – instrumental methods and computer methods. The more common instrumental methods of monochromatization include the following:

- use of a β-filter
- use of a proportional detector and pulse height selection
- use of a Si(Li) solid-state detector
- use of a diffracted-beam monochromator
- use of a primary beam monochromator.

The basic purpose of the monochromatization of the diffracted radiation is to obtain an experimental pattern from a single, unique wavelength. Inspection of Bragg's law (Equation 5) reveals that each unique d-spacing will diffract different wavelengths at their own unique diffraction angles. Thus, if a pattern were measured using an X-ray beam containing two wavelengths, the observed pattern would, in fact, be two patterns (one for each wavelength) superimposed on one another, and clearly will be more difficult to interpret than a diffractogram from a single wavelength. While the problems associated with the interpretation of a diffractogram from two known wavelengths can be significant, much more of a problem occurs when additional wavelengths are diffracted. The characteristic K-radiation emission from typical X-ray tube anode materials is much more complex than the simple α-doublet and β-doublet model generally employed in classical powder diffractometry. For most practical purposes, however, the copper K spectrum is generally considered to consist simply of two pairs of lines, the α_1/α_2-doublet, occurring from a $2p \rightarrow 1s$ transition; and the β_1/β_3-doublet from a $3p \rightarrow 1s$ transition. The following table lists the values of the Kα and Kβ wavelengths for the more common X-ray tube target elements.

Anode	Kα_1(100)	Kα_2(50)	Kβ(15)
	Wavelength (Å)		
Cu	1.54060	1.54439	1.39222
Cr	2.28970	2.29361	2.08487
Fe	1.93604	1.93998	1.75661
Co	1.78897	1.79285	1.62079
Mo	0.70930	0.71359	0.63229

In most experimental work, the β-doublet intensity is typically reduced to less than a few percent of the α-doublet intensity by use of filtration, or it is removed by use of a diffracted beam monochromator or a high-resolution energy-resolving detector. In each case, what remains is essentially bichromatic X-radiation. The exception to this occurs with the primary beam monochromator where only the α_1 is directed onto the specimen. In applying the usual bichromatic radiation to the measurement of interplanar spacings, the main problems are the angle-dependent dispersion of the α_1/α_2-doublet, and differences in peak asymmetry between the two. It follows from Bragg's law that an error in d-spacing is linearly related to an error in wavelength. For manual qualitative phase identification, an accuracy of about 5 parts in 1000 for $\Delta d/d$ is sufficient;

and about twice this accuracy is required for computer search/matching. Provided that the true α_1-emission line is being used, i.e. either the diffracted beam is really monochromatic, or the contribution from the α_2 is effectively removed from the diffraction profile, then the wavelength need also be known only to about 2 parts per 1000. For accurate lattice parameter determination, it is common practice to use an internal standard and, in these cases, the wavelength value is, to all intents and purposes, redundant.

5.3 Use of Diffractometers in the Powder Method

A powder diffractometer is essentially a device which allows a range of θ values to be scanned, rotating the photon detector at twice the angular speed of the specimen thus maintaining the required geometrical condition. The specimen consists of a random distribution of crystallites so that the appropriate planes will be in the correct orientation to diffract the wavelength λ each time the Bragg condition is satisfied. Thus in X-ray diffraction each peak angle value corresponds to a certain *d*-spacing. In the wavelength-dispersive spectrometer a single crystal of known *d*-spacing is used to disperse the polychromatic beam of characteristic wavelengths coming from the sample, such that each wavelength will diffract at a discrete angle.

Figure 5 shows a schematic drawing of a typical vertical powder diffractometer system and illustrates the geometry of the system. This geometric arrangement is known as the Bragg–Brentano parafocusing system and is typified by a diverging beam from a line source F, falling onto the specimen S, being diffracted and passing through a receiving slit RS to the detector. Distances FS and SRS are equal. The amount of divergence is determined by the effective focal width of the source and the aperture of the divergence slit DS. Axial divergence is controlled by two sets of parallel plate collimators (Soller slits) SS1 and SS2 placed between focus and specimen, and specimen and scatter slit respectively. Use of the narrower divergence slit will give a smaller specimen coverage at a given diffraction angle, thus allowing the attainment of lower diffraction angles where the specimen has a larger apparent surface (thus larger values of *d* are attainable). This is achieved, however, only at the expense of intensity loss. Choice of the divergence slit, plus its matched scatter slit, is thus governed by the angular range to be covered. The decision as to whether or not the slit size should be increased at a given angle will be determined by the available intensity. A scintillation detector is typically placed behind the scatter slit and this converts the diffracted X-ray photons into voltage pulses. These pulses may be integrated in a rate meter to give an analog signal on an *x*/*t* recorder. By synchronizing the scanning speed of the goniometer with the recorder, a plot is obtained of degrees 2θ against intensity, called the diffractogram. A timer/scaler is also provided for quantitative work and this is used to obtain a measure of the integrated peak intensity of a selected line(s) from each analyte phase in the specimen. A diffracted beam monochromator may also be used in order to improve signal-to-noise characteristics.

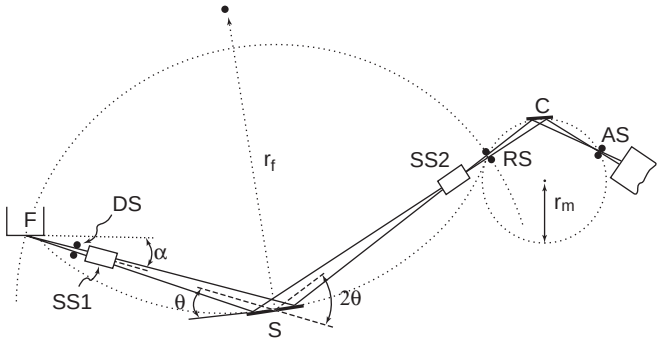

Figure 5 Bragg–Brentano parafocusing geometry. (Reproduced from Jenkins, Snyder, *Introduction to Powder Diffractometry*, Wiley-Interscience, 1996. With permission from John Wiley & Sons.)

5.4 Powder Cameras

The simplest device for the measurement of a powder diffraction diagram is the Debye–Scherrer powder camera. This consists of a cylindrical camera body carrying an entrance pinhole collimator and an exit beam collimator lying along the diameter of the camera. The specimen is mounted as a thin cylinder at the central axis of the camera. A piece of film is placed inside the cylindrical wall of the camera and small holes are cut in the film for the entrance and exit collimators. A beam of X-rays is directed onto the specimen via the entrance collimator and the diffracted X-rays fall onto the film. The film is then developed and laid flat for the interpretation. In order to make estimation of 2θ values from the film easy, the diameter of the camera is made such that one or two mm of film corresponds to exactly 1° 2θ. The powder camera finds great application in the analysis of very small specimens and good patterns can be obtained from only a few milligrams of specimen. The camera technique has the disadvantage that it may be rather slow and exposure times are typically in the range of 1–6 hours. In addition to the Debye–Scherrer powder camera there are also available more specialized cameras for various applications. As an example, the Guinier camera is a powder camera which incorporates a focusing monochromator and works in a transmission mode. Although the intensities are rather unreliable and it is more difficult to align than the Debye–Scherrer camera, the Guinier camera can give

very high quality patterns in a relatively short period of time. Another useful camera is the Gandolphi camera which is essentially a Debye–Scherrer camera with an eccentric specimen rotation movement added, allowing the generation of a 'powder' pattern from a small single crystal.

6 PHASE CHARACTERIZATION BY X-RAY DIFFRACTION

6.1 Qualitative Analysis of Polycrystalline Material

Of all of the methods available to the analytical chemist only X-ray diffraction is capable of providing general purpose qualitative and quantitative information on the presence of phases (e.g. compounds) in an unknown mixture. While it is true that techniques such as differential thermal analysis will provide some information on specific phase systems, under certain circumstances such methods could not be classified as general purpose. A diffraction pattern contains a good deal of information of which three parameters are of special interest:

- the position of the diffraction maxima
- the peak intensities
- the intensity distribution as a function of diffraction angle.

The three pieces of information can, in principle, be used to identify and quantify the contents of the sample, as well as to calculate the material's crystallite size and distribution, crystallinity, stress and strain. The ideal specimen preparation for a given experiment depends largely on the information desired. A sample which is used only for the identification of its constituents may be quite different from a sample used to measure strain, which in turn may be different from a sample used in quantitative analysis.

A diffraction pattern is characteristic of the atomic arrangement within a given phase and to this extent it acts as a *fingerprint* of that particular phase. Thus by use of the ICDD PDF a series of potential matches can be obtained. A complication in the application of this method for the analysis of multiphase materials is that the patterns are superimposed and consequently there may be uncertainties as to which lines belong to which phases. As shown in Figure 6, it is common practice to ''reduce'' the *experimental* pattern to a *reduced* pattern. This process was first carried out because of problems with storage and replication of powder patterns, and the procedure persists today.

In practice it is rather unusual to analyze specimens in which nothing at all is known. Generally one has access to

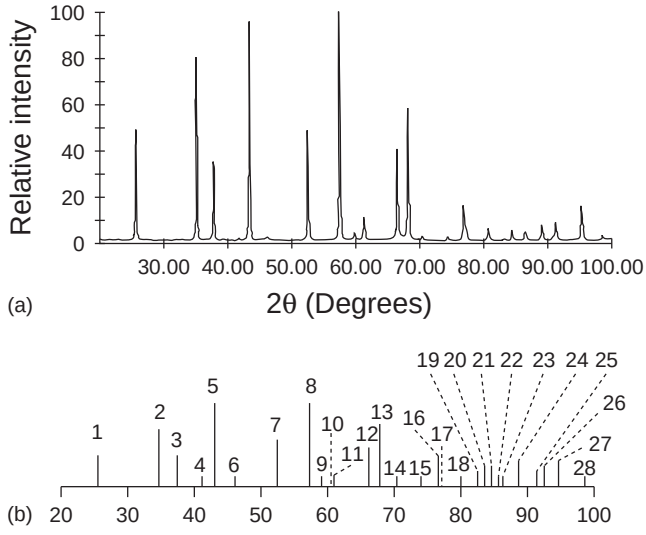

Figure 6 (a) The experimental pattern and (b) the reduced pattern of Al_2O_3. (Reproduced from Jenkins, Snyder, *Introduction to Powder Diffractometry*, Wiley-Interscience, 1996. With permission from John Wiley & Sons.)

information about sample history, environment, etc. and additional analytical information can be employed where available. As an example, in the routine analytical laboratory it is common to find the X-ray diffractometer and the X-ray spectrometer being used to give complementary information. Thus a more common means of identifying phases in an ''unknown'' mixture is to use the technique of preconceived. In this case educated guesses are made as to what phases may be present and these phases are sought first. As an example, if one were analyzing a specimen obtained by aspirating air through a filter paper, and the X-ray spectrometer indicated significant concentrations of silicon and iron, likely candidates to search for in the diffraction pattern would be SiO_2, Fe_2O_3 and $FeSiO_3$. In order to simplify the search procedure further the ICDD has subdivided the file of standard patterns into Mineral, Inorganic, Organic, Metals and Alloys, Common Phases, etc. XRD is an invaluable technique for qualitative phase identification but it is neither a sensitive nor a rapid means of analysis. The minimum detectable limit found by routine qualitative procedures would be of the order of several percent, compared, for example, with a few parts per million in XRF. A complete analysis would take several hours to complete, although this will of course depend to a very large extent on the experience of the analyst and the complexity of the problem in hand. The method is applicable to almost any crystalline material whether inorganic or organic.

6.2 Search Procedures

The PDF is described in section 5. Several different approaches are used to identify phases in an unknown

mixture. The first method is an analytical approach in which no basic assumptions are made about the sample being analyzed. The three strongest lines in the pattern are used to locate potential matches in the PDF index. Each time a potential candidate is found, a match is made with the complete pattern. If all lines agree, a phase confirmation is assumed and the lines for the match subtracted from the original pattern. This process is repeated until all significant lines in the pattern are identified. The second method is based on a series of guesses based in turn on preconceived ideas of what phases might be present. The two basic parameters being used in this search/match process are the "*d*" values which in turn have been calculated from the measured 2-θ values in the diffractogram, and the relative intensities of the lines in the pattern. Whereas the "*d*" value can be accurately measured – perhaps with an accuracy of better than 0.5% in routine analysis, the intensities are by comparison rather unreliable and can be subject to errors, sometimes running into tens of percent.

6.3 Modern Computer-based Methods

More than 90% of all new powder diffractometers now sold in the world are automated. Many excellent computer programs are provided with these instruments to aid the diffractionist in the reduction of powder data. Unfortunately, there can be pitfalls in the use of some of these programs, especially for the newcomer to the diffraction field. A popular misconception is that automated diffractometers are easier to use than manual systems. What, in fact, is true is that automated powder diffractometers (APD) can give many hours of unattended and reliable data collection and can greatly assist in the tedious and routine tasks of data analysis, provided one is careful to apply the programs in the correct manner.

During the 1960s, as mainframe computers evolved (e.g. IBM 7094, CDC 6600, etc.) crystallographic computing was performed in a noninteractive environment, where cards were input into a program operating in a batch stream. Owing to the very high cost of computers, laboratory automation was nearly nonexistent, so these early developments can be considered as the zeroth generation of laboratory computational software. One of the first APDs was developed by Rex[20] in the mid-1960s. However, it took the microelectronics revolution of the 1970s to initiate the general techniques which have led to today's automated instrumentation. Although the principal thrust in the early 1970s was to develop the hardware interfaces needed to allow a computer to control a diffractometer, this work rapidly gave way to the much more serious problems of devising algorithms for the control of the instrument and the processing of the digital data. The 1970s saw the evolution of the early laboratory computer (e.g. the Digital Equipment Corporation PDP-8, and PDP-11), and the first generation of process control software. Input to the program was first by paper tape and later by magnetic tape cassette, with programs written in low-level assembler language. Systems of this type were developed at Alfred University, Corning Glass Works and General Electric. The first commercial APD system in the USA was the Philips APD-3500[21] which developed from involvement in the NASA project for a Lunar Receiving Laboratory. This instrument utilized a 4000-word, 18-bit program logic controller (a forerunner of what was to become the microprocessor). Within the 4 K of core, routines were available for hardware control, data collection, data manipulation and sophisticated math – all without hardware arithmetic capability!

The second generation of laboratory software evolved on the next generation of laboratory computers (e.g. the PDP 11/34), with the introduction of floppy disks which permitted convenient use of high-level language compilers. The first of these systems was written in FORTRAN and allowed the user to conduct an interactive dialogue with the computer to set up and execute an automated experiment and analyze the data.[22] The development of the 5 and 10 Mbyte Phoenix and Winchester disk drives, allowed the development of the third generation of software involving extensive file structures with help screens on video terminals and default files containing the users' typical settings, rather than interactive dialogue.[23,24] A fourth generation of software is currently evolving where all interactions with the user are done through a "point and click" video interface with a mouse. The trend is to eliminate all keyboard activity and convert crystallographic methods, which used to rely on the user evaluating numbers, to visual examination of graphical data.

There are essentially four elements to the automation of a diffractometer:

- the replacement of the synchronous $\theta : 2\theta$ motor with a stepping motor and its associated electronics
- the replacement of a conventional scaler/timer with one which can be remotely set and read
- the conversion of the various alarms, limit switches and shutter controls to computer-readable signals
- the creation of an interface which will allow a computer to control 1 through 3.

These four items are easily obtained today by direct purchase of modules which often plug directly onto the bus of a modern minicomputer.

Today, nearly all laboratory APD systems are controlled by a personal computer. The major impact of computer automation on improving the accuracy of the measurement of diffraction angles has come from the algorithms which bring much more intelligence to the

process than has been conventionally used in manual measurements. There are two areas here which need to be considered; the first is the algorithms which control the collection of data, and the second, those which reduce the data to d values and intensities. The first generation of control algorithms were principally nonoptimizing, move and count methods. Although various attempts have been made to use the computer's decision-making capability to bring more ''intelligence'' to the data-collection process,[25] the algorithms in use today continue to be nonoptimizing. The fundamental problem is the determination of the intensity threshold above background, which limits the ability to detect small peaks that could be due to trace phases in the specimen. Thus, in the limit, an intelligent algorithm would spend all of its time counting where the peaks are not present. The key hindrance to the development of intelligent data-collection software remains our inability to predict or even accurately describe the diffractometer's background function.

6.4 Limitations Due to the Specimen

The aim of any diffraction experiment is to obtain the best possible data, within the appropriate constraints of the relevant circumstances, so that the data can be correctly interpreted and analyzed. One of the major problems in achieving this goal is the preparation of the specimen. Various methods of specimen preparation have been devised[26] and the success of a given diffraction experiment will invariably depend on the correct choice of preparation method for the sample being analyzed and for the instrument conditions being used for the analysis. Because of the many problems associated with the form of the analyzed material, it is useful to differentiate between the sample and the specimen. The sample is generally considered to be the material submitted for analysis. It is generally assumed that the sample has been correctly taken (sampled) by the submitter to give a representative measure of the problem at hand. It is now up to the diffractionist to take an aliquot of the sample and prepare a specimen for analysis. This specimen preparation may involve a number of steps such as drying, grinding, sieving, dilution, mounting, etc. It is tacitly assumed that the analysis of the specimen represents the analysis of the sample. While this assumption is generally true, it may not always be so, and the analyst must be on his guard to ensure that all reasonable precautions are taken to avoid contamination or phase changes during the process of specimen preparation.

The biggest overall problem in specimen preparation is that due to preferred orientation. Particle and crystallite size, sample size, sample position, crystallite orientation, and absorption will affect the quality and appearance of any diffraction pattern. As a rule of thumb, the best specimen preparation methods are those that allow the analyst to obtain the desired information with the least amount of sample treatment. This is especially important in the analysis of unknown materials, pharmaceuticals and organics where the materials being analyzed may be sensitive to grinding, humidity or atmospheric environment. Extensive specimen preparation may change the sample so that the analyst is no longer analyzing the original problem. A good example is the case of hydrates and polymorphs in pharmaceutical formulations. The presence of hydrates and/or polymorphs may change the melting, solubility and blending characteristics of the formulation. In such a case the analyst must ensure that the sample preparation method does not dehydrate the sample or change the polymorphic composition. This may eliminate specimen preparation methods involving grinding, spray drying and vacuum drying and may necessitate that the sample be sealed in a controlled environment.

6.5 Quantitative Methods

Once the presence of a phase has been established in a given specimen one can, at least in principle, determine how much of that phase is present by use of the intensities of one or more diffraction lines from the phase. However, as has been pointed out previously, it may be difficult to obtain an accurate value for these intensities. The intensities of the diffraction peaks are subject to a variety of random and systematic errors and the errors fall roughly into three categories:

- Structure-dependent: that is a function of atomic size and atomic arrangement, plus some dependence on the scattering angle and temperature.
- Instrument-dependent: that is a function of diffractometer conditions, source power, slit widths, detector efficiency, etc.
- Specimen-dependent: that is a function of phase composition, specimen absorption, particle size, distribution and orientation.

For a given phase, or selection of phases, all structure-dependent terms are fixed, and in this instance have no influence on the quantitative procedure. Provided that the diffractometer terms are constant, this effect can also be ignored. Thus if one calibrates the diffractometer with a sample of the pure phase of interest and then uses the same conditions for the analysis of the same phase in an unknown mixture, only the random errors associated with a given observation of intensity have to be considered. The biggest problem in the quantitative analysis of multiphase mixtures then remains the specimen-dependent terms, and specifically those dependent upon particle size and distribution, plus effects of absorption.

The absorption effect has already been mentioned and clearly, in a multiphase mixture, different phases will absorb the diffracted photons by different amounts. As an example, the mass attenuation coefficient for Cu Kα radiation is 308 ($cm^2\,g^{-1}$) for iron, but only 61 ($cm^2\,g^{-1}$) for silicon. Thus iron atoms are five times more efficient than silicon atoms for absorbing Cu Kα photons. There are a variety of standard procedures for correcting for the absorption problem of which by far the most common is the use of the internal standard. In this method a standard phase is chosen which has about the same mass absorption coefficient as the analyte phase, and a weighed amount of this material is added to the unknown sample. The intensities of a line from the analyte phase and the internal-standard phase are then used to estimate the relative concentrations of internal-standard and analyte phases. The relative sensitivity of the diffractometer for these two phases is determined by a separate experiment.

The handling of particle problems is unfortunately much more complex. As has been previously stated the powder method requires a specimen which is randomly oriented since the geometry of the system requires that an equal number of crystallites be in the correct position (i.e. orientation) to diffract at any diffraction angle where the goniometer happens to be. Where particles lie in a preferred orientation there will be more particles available to diffract at the angle corresponding to this orientation, and, what is equally important, less particles available to diffract at other diffraction angles. The overall effect is to enhance some intensities and to diminish others. In other words, the intensities are now dependent upon particle distribution and orientation. Some materials, just by virtue of their crystal habit, may become preferentially oriented during sample preparation. As an example, mica, being a rather ''platy'' material, will prefer to stack one plate on top of another rather than take up a random orientation. The overall effect of preferred orientation can vary from insignificance to the extent where errors of the order of tens of percent can accrue in cases where they are not taken care of. Careful specimen preparation is always called for and this may include grinding, sieving, spray-drying, and a host of other techniques. The areas of application of quantitative XRD are many and varied, and hundreds of analysts are using this technique on a daily basis. Some of the more common applications would include ore and mineral analysis, quality control of rutile/anatase mixtures, retained austenite in steels, determination of phases in airborne particulates, various thin-film applications, study of catalysts and analysis of cements. The current state of the art in the quantitative analysis of multiphase materials is that accuracy of the order of a percent or so can be obtained in those cases where the particle orientation effect is either nonexistent, or has been adequately compensated for.

6.6 Limitations of the Diffraction Method

There is little doubt that the diffraction method is without parallel as a general-purpose tool for phase identification but like all analytical techniques it is not without some shortcomings. Two of these are of particular importance. The first of these is related to the chance of misinterpretation during the course of a qualitative analysis procedure. Although, in principle, a diffraction pattern is ''unique'', in practice there are sufficient similarities between patterns as to cause confusion. This is particularly so in the case of multiphase specimens. In addition to this there are probably in excess of two million possible unique ''phases'', of which only 120 000 or so are on the ICDD file as single-phase patterns. There is as a result a certain chance of a given phase not being in the file. The second problem relates to the sensitivity of the powder diffraction method. As such, the diffraction method is not comparable in sensitivity (i.e. rate of change of measured signal per change in analyte-phase concentration) to the other X-ray-based techniques. Whereas in X-ray spectrometry one can obtain detection limits in the low parts per million region, the powder method has difficulty in identifying several tenths of one percent. To this extent it is less sensitive than the fluorescence method by about three orders of magnitude.

7 FUTURE TRENDS FOR X-RAY DIFFRACTION AND FLUORESCENCE

7.1 Analysis of Small Specimens and Thin Films

Conventional XRF spectrometers are generally designed to handle rather large specimens with surface areas of the order of several square centimeters, and problems occur where the sample to be analyzed is limited in size. The modern wavelength-dispersive system is especially inflexible in the area of sample handling, mainly because of geometric constraints arising from the need for close coupling of the sample to X-ray tube distance, and the need to use an airlock of some kind to bring the sample into the working vacuum. The sample to be analyzed is typically placed inside a cup of fixed external dimensions that is, in turn, placed in the carousel. This presentation system places constraints not only on the maximum dimensions of the sample cup, but also on the size and shape of samples that can be placed into the cup itself. Primary source energy-dispersive systems do not require the same high degree of focusing, and to this extent are more easily applicable to any sample shape or size,

provided that the specimen will fit into the radiation protected chamber. In some instances the spectrometer can even be brought to the object to be analyzed. Because of this flexibility, the analysis of oddly shaped specimens have been almost exclusively the purview of the energy-dispersive systems. In the case of secondary target energy-dispersive systems, while the geometric constraints are still less severe than the wavelength system, they are much more critical than in the case of primary systems. Where practicable, the best solution for the handling of limited amounts of material is invariable found in one of the specialized spectrometer systems.

One of the major problems that inhibits the obtaining of good detection limits in small samples is the high background due to scatter from the sample substrate support material. The suggestion to overcome this problem by using total reflection of the primary beam was made as long ago as 1971,[27] the technique being referred to as total reflection X-ray fluorescence (TRXRF) spectrometry. Unfortunately the absence of suitable instrumentation prevented real progress being made until the late 1970s.[28,29] The TRXRF method is essentially an energy-dispersive technique in which the Si(Li) detector is placed close to (about 5 mm), and directly above, the sample. Primary radiation enters the sample at a glancing angle of a few seconds of arc. The sample itself is typically presented as a thin film on the optically flat surface of a quartz plate. Here, a beam of radiation from a sealed X-ray tube passes through a fixed aperture onto a pair of reflectors that are placed very close to each other. Scattered radiation passes through the first aperture to impinge on the sample at a very low glancing angle. Because the primary radiation enters the sample at an angle barely less than the critical angle for total reflection, this radiation barely penetrates the substrate media; thus scatter and fluorescence from the substrate are minimal. Because the background is so low, picogram amounts can be measured or concentrations in the range of a few tenths of a ppb can be obtained without recourse to preconcentration.[30] Similar parallel-beam thin-film methods are finding increasing use in powder diffraction.

7.2 Impact of the Synchrotron

The availability of intense, linearly polarized synchrotron radiation beams[31] has prompted workers in the fields of XRF[32] to explore what the source has to offer over more conventional excitation media. An exciting new field developed in the early 1980s called synchrotron source X-ray fluorescence (SSXRF). In the synchrotron, electrons with kinetic energies of the order of several billion electron volts (typically 3 GeV at this time), orbit in a high-vacuum tube between the poles of a strong (about 10 000 Gauss) magnets. A vertical field accelerates the electrons horizontally, causing the emission of synchrotron radiation. Thus synchrotron source radiation can be considered as magnetic bremsstrahlung in contrast to normal electronic bremsstrahlung produced when electrons are decelerated by the electrons of an atom. It has been found that, because the primary source of radiation is so intense, it is possible to use a high degree of monochromatization between source and specimen, giving a source that is wavelength (and, therefore, energy) tunable, as well as being highly monochromatic. There are several different excitation modes that can be used using SSXRF including: direct excitation with continuum, excitation with absorber-modified continuum, excitation with source crystal monochromatized continuum, excitation with source radiation scattered by a mirror, and reflection and transmission modes. The intensity of the synchrotron beam is probably four to five orders of magnitude greater than the conventional bremsstrahlung source sealed X-ray tubes. This, in combination with its energy tunability and polarization in the plane of the synchrotron ring, allows very rapid bulk analyses to be obtained on small areas. Because the synchrotron beam has such a high intensity and small divergence, it is possible to use it as a microprobe of high spatial resolution (about 10 μm). Absolute limits of detection around 10^{-14} pg have been reported using such an arrangement.[33] SSXRF has also been used in combination with TRXRF. Very high signal/background ratios have been obtained employing this arrangement for the analysis of small quantities of aqueous solutions dried on the reflector, with detection limits of <1 ppb or 1 pg. Additional advantages accrue because synchrotron radiation is highly polarized and background due to scatter can be greatly reduced by placing the detector at 90° to the path of the incident beam and in the plane of polarization. A disadvantage of the SSXRF technique is that the source intensity decreases with time, but this can be overcome by bracketing analytical measurements between source standards and/or by continuously monitoring the primary beam.

ABBREVIATIONS AND ACRONYMS

APD	Automated Powder Diffractometer
ICDD	International Centre for Diffraction Data
LLD	Lower Limit of Detection
NRLXRF	Naval Research Laboratory X-ray Fluorescence
PDF	Powder Data File
SSXRF	Synchrotron Source X-ray Fluorescence
TRXRF	Total Reflection X-ray Fluorescence
UV	Ultraviolet
XRD	X-ray Powder Diffractometry
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Polymers and Rubbers **(Volume 9)**
X-ray Scattering in Analysis of Polymers

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Structure Determination, X-ray Diffraction for • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. W.C. Roentgen, 'On a New Kind of Rays', *Ann. Phys. Chem.*, **64**, 1–11 (1898).
2. G. Von Hevesey, *Chemical Analysis by X-rays and its Application*, McGraw-Hill, New York, 1932.
3. R. Jenkins, *An Introduction to X-ray Spectrometry*, Wiley/Heyden, London, Chapter 4, 1974.
4. L.S. Birks, 'History of X-ray Spectrochemical Analysis', American Chemical Society Centennial Volume, ACS, Washington DC, 1976.
5. H.G.J. Moseley, 'High Frequency Spectra of the Elements', *Phil. Mag.*, **26**, 1024–1034 (1912); **27**, 703–713 (1913).
6. P. Auger, 'Compound Photoelectric Effect', *Compt. Rend.*, **180**, 65 (1925); *Journal de Physique*, **6**, 205 (1925).
7. J.A. Bearden, 'X-ray Wavelengths', US Atomic Energy Commission Report NYO-10586, Oak Ridge, TN, 533, 1964.
8. J. Sherman, 'Theoretical Derivation of Fluorescent X-ray Intensities from Mixtures', *Spectrochim. Acta*, **7**, 283–306 (1955).
9. T. Shiraiwa, N. Fujino, 'Theoretical Calculation of Fluorescent X-ray Intensities of Nickel, Iron and Chromium Ternary Alloys', *Bull. Soc. Chem. Jpn.*, **40**, 2289–2296 (1967).
10. J.W. Criss, L.S. Birks, 'Calculation Methods for Fluorescent X-ray Spectrometry – Empirical Coefficients vs. Fundamental Parameters', *Anal. Chem.*, **40**, 1080–1086 (1968).
11. J.W. Criss, 'Fundamental Parameters Calculations on a Laboratory Microcomputer', *Adv. X-Ray Anal.*, **23**, 93–97 (1980).
12. J.V. Gilfrich, L.S. Birks, 'Spectral Distribution of X-ray Tubes for Quantitative X-ray Fluorescence Analysis', *Anal. Chem.*, **40**, 1077–1080 (1968).
13. W.K. de Jongh, 'X-ray Fluorescence Analysis Applying Theoretical Matrix Correction. Stainless Steel', *X-Ray Spectrom.*, **2**, 151 (1973).
14. M.J. Beattie, R.M. Brissey, 'Calibration Method for X-ray Fluorescence Spectrometry', *Anal. Chem.*, **26**, 980–983 (1954).
15. H.J. Lucas-Tooth, C. Pyne, 'The Accurate Determination of Major Constituents by X-ray Fluorescence Analysis in the Presence of Large Interelement Effects', *Adv. X-Ray Anal.*, **7**, 523–541 (1964).
16. G.R. Lachance, R.J. Traill, 'A Practical Solution to the Matrix Problem in X-ray Analysis – I. Method', *Can. Spectrosc.*, **11**, 43–48 (1966).
17. R. Jenkins, J.L. de Vries, *Practical X-ray Spectrometry*, 2nd edition, Springer, New York, 1970.
18. J.D. Hanawalt, H.W. Rinn, 'Identification of Crystalline Materials', *Ind. Eng. Chem.*, **8**, 244–247 (1936).
19. R. Jenkins, C.R. Hubbard, 'A Preliminary Report on the Design and Results of the Second Round Robin to Evaluate Search/Match Methods for Qualitative Powder Diffractometry', *Adv. X-Ray Anal.*, **22**, 133–142 (1978).
20. R.W. Rex, 'Numerical Control X-ray Powder Diffractometry', *Adv. X-Ray Anal.*, **10**, 366–373 (1966).
21. R. Jenkins, D.J. Haas, F.R. Paolini, 'A New Concept in Automated X-ray Powder Diffractometry', *Norelco Reporter*, **18**, 1–16 (1971).
22. C.L. Mallory, R.L. Snyder, 'The Control and Processing of Data from an Automated Powder Diffractometer', *Adv. X-Ray Anal.*, **22**, 121–132 (1979).
23. R.L. Snyder, National Bureau of Standards Technical Report, NBSIR-81-2229, The National Institute of Science and Technology, Gaithasburg, MD, 1981.
24. R.L. Snyder, C.R. Hubbard, N.C. Panagiotopoulus, 'A Second-generation Automated Powder Diffractometer Control System', *Adv. X-Ray Anal.*, **25**, 245–260 (1982).
25. R.L. Snyder, 'Accuracy in Angle and Intensity Measurements in X-ray Powder Diffraction', *Adv. X-Ray Anal.*, **26**, 1–11 (1983).
26. V.E. Buhrke, R. Jenkins, D.K. Smith, *Preparation of Specimens for XRF and XRD Analysis*, Wiley/VCH, New York, 1998.
27. Y. Yoneda, T. Horiuchi, 'Optical Flats for Use in X-ray Spectrochemical Microanalysis', *Rev. Sci. Instrum.*, **42**, 1069–1070 (1971).
28. J. Knoth, H. Schwenke, 'An X-ray Fluorescence Spectrometer with Totally Reflecting Sample Support for Trace Analysis at the p.p.b Level', *Fresenius' J. Anal. Chem.*, **301**, 200–204 (1978).
29. H. Schwenke, J. Knoth, 'A Highly Sensitive Energy Dispersive X-ray Spectrometer with Multiple Total Reflection of the Exciting Beam', *Nucl. Instrum. Methods*, **193**, 239 (1982).

30. H. Aiginger, P. Wobrauschek, 'Total Reflectance X-ray Spectrometry', *Adv. X-Ray Anal.*, **28**, 1–10 (1985).
31. C.J. Sparks, Jr, 'Synchrotron Radiation Research', in *Synchrotron Radiation Research*, eds. H. Winnick, S. Doniach, Plenum Press, New York, 459, 1980.
32. J.V. Gilfrich et al., 'Synchrotron Radiation X-ray Fluorescence Analysis', *Anal. Chem.*, **55**, 187–192 (1983).
33. W. Petersen, P. Ketelsen, A. Knoechel, R. Pausch, 'New Developments of X-ray Fluorescence Analysis with Synchrotron', *Nucl. Instrum. Methods Phys., Sect. A*, **A246**(1–3), 731 (1986).

Absorption Techniques in X-ray Spectrometry

Jun Kawai
Kyoto University, Kyoto, Japan

1 Introduction **1**
2 Acronyms and a Brief History **3**
3 X-ray Absorption Near-edge Structure **5**
3.1 Chemical Shift and Line Shape 5
3.2 Calculation Method for X-ray Absorption Near-edge Structure Spectra 8
4 Theory of Extended X-ray Absorption Fine Structure **8**
4.1 Single Scattering Theory 8
4.2 Relation to Other Techniques (X-ray Photoelectron Diffraction, Low-energy Electron Diffraction, X-ray Fluorescence Holography) 10
5 Data Analysis and Software Packages for X-ray Absorption Fine Structure **10**
6 Instrumentation **13**
6.1 Laboratory Extended X-ray Absorption Fine Structure 13
6.2 Synchrotron Radiation Extended X-ray Absorption Fine Structure 13
6.3 Secondary Yield Techniques and Applications 15
7 Sources and Databases **21**
8 Alternative Methods **21**
8.1 Electron Energy Loss Spectroscopy 21
8.2 Self-absorption 22
8.3 Extended X-ray Emission Fine Structure 22
8.4 X-ray Raman Scattering 22
8.5 Diffraction Anomalous Fine Structure 23
8.6 β-Environment Fine Structure 23
8.7 Inverse Photoemission Spectroscopy 23
9 Conclusion **23**
Abbreviations and Acronyms **23**
Related Articles **24**
References **24**

X-ray absorbance depends on the wavelength of the X-rays, atomic number, chemical environment, and concentration of analyte. X-ray absorption spectrometry is a technique for analyzing the chemical environment of an element in an unknown material. This method is closely related to photoelectron spectroscopy, Auger electron spectroscopy, and X-ray fluorescence spectroscopy.

Chemical information in the chemical shift and line shape of XANES (X-ray absorption near-edge structure) spectra is described. The history and theory of EXAFS (extended X-ray absorption fine structure) are discussed in relation to other experimental techniques. Data analysis methods, databases, software packages, instrumentation, and synchrotron radiation facilities for X-ray absorption analysis are overviewed. Alternative methods such as electron energy loss spectroscopy (EELS), self-absorption effect, extended X-ray emission fine structure (EXEFS), X-ray Raman scattering, diffraction anomalous fine structure (DAFS), β-environment fine structure (BEFS), and inverse photoemission spectroscopy (IPES) are also described.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION

X-rays are absorbed in matter and the energy of the X-rays is converted into the kinetic energy of photoelectrons, Auger electrons, secondary electrons, or fluorescent X-rays. The incident X-ray energy finally becomes the thermal energy of the absorber.

The amount of energy absorbed by a matter is usually estimated by a transmission method, but can also be estimated by measuring these secondary phenomena, such as photoelectrons, Auger electrons, secondary electrons, fluorescent X-rays, thermal radiation, and drain electric currents. The X-ray intensity of wavelength λ before (I_0) and after (I) the transmission of a thin film of thickness d is expressed by $I(\lambda) = I_0(\lambda)\exp[-\mu_i(\lambda)\rho_i d]$, where $\mu_i(\lambda)$ and ρ_i are the mass absorption coefficient and mass density, respectively, of the ith element in the thin film and their dimensions are [$cm^2\,g^{-1}$] and [$g\,cm^{-3}$], respectively.[1–7] The mass absorption coefficient μ of a specimen which contains n kinds of elements is expressed by $\mu = \mu_1(\lambda)W_1 + \mu_2(\lambda)W_2 + \cdots + \mu_n(\lambda)W_n$, where $W_1, W_2, \ldots, W_n$ are the weight fractions of element $1, 2, \ldots, n$ in the specimen. The wavelength dependence of the absorption coefficient $\mu(\lambda)$ is clarified when $\log\mu(\lambda)$ is plotted against $\log\lambda$ as shown in Figure 1; $\mu(\lambda)$ values are taken from Henke et al.[8] in this plot. Henke et al.[8] tabulated $\mu(\lambda)$ from $Z = 1$ to 92 at energy from 50 eV to 30 keV.

The plot of the mass absorption coefficients of matter against the incident X-ray energy or wavelength is called an X-ray absorption spectrum (XAS), where we find some jumps at particular X-ray energy, corresponding to K, L_I, L_{II}, L_{III}, ... electron shell binding energies as shown in

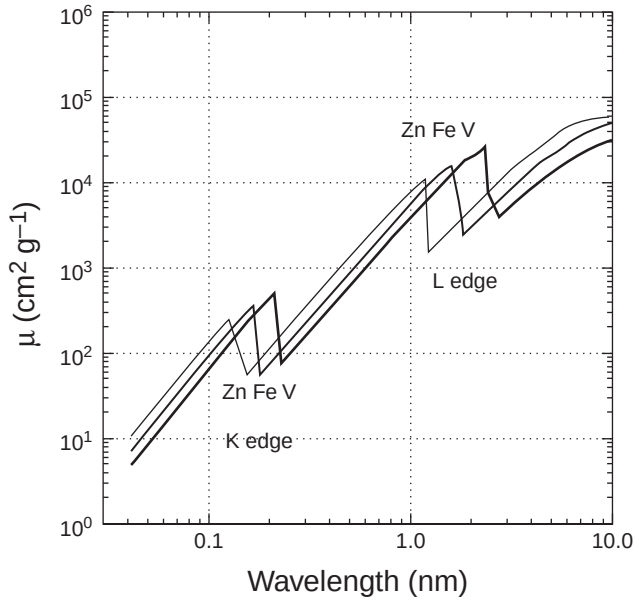

Figure 1 Mass absorption coefficients of V, Fe, and Zn plotted against wavelength. Both axes are on a logarithmic scale.

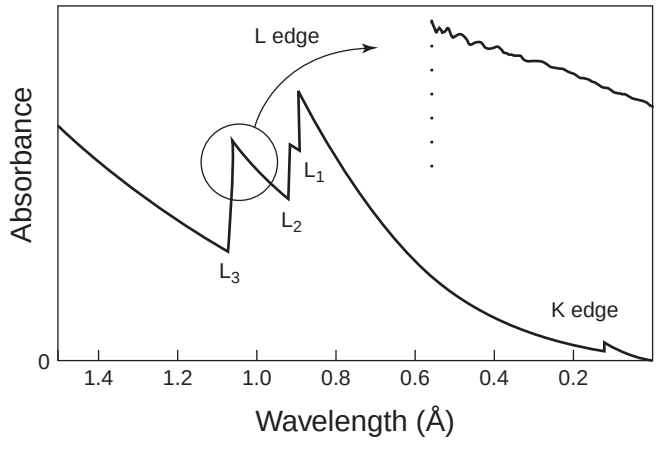

Figure 2 Platinum powder XAS. (Reproduced by permission from Udagawa.[9])

Figure 2.[9] K, L_I, L_{II}, and L_{III} denote electron deficiency states from 1s, 2s, $2p_{1/2}$, and $2p_{3/2}$ orbitals, respectively. Arabic numerals 1, 2, 3, ... have more recently been used in the subscript rather than Roman numerals I, II, III, ... The electron configuration of one electron deficiency from the 2p orbital is expressed as $1s^2 2s^2 2p^5 3s^2 3p^6$ for an Ar atom. This state has two energy levels corresponding to $j = \frac{1}{2}$ and $\frac{3}{2}$ states, where j is the eigenvalue of vector sum $\boldsymbol{s} + \boldsymbol{l}$, where $\boldsymbol{s}$ and $\boldsymbol{l}$ are called spin and angular momentum vectors, respectively, and the vector $\boldsymbol{j}$ is called the total angular momentum. These two states are written as $2p_{1/2}^{-1}$ and $2p_{3/2}^{-1}$, or $[2p_{1/2}]$ and $[2p_{3/2}]$.

The relation between the electron deficient state and the electron configuration is listed in Table 1. The s^{-1} hole state has a total angular momentum $\frac{1}{2}$, and the subscript $\frac{1}{2}$ is usually omitted. The multiplicity of the state, which is crudely proportional to the spectral intensity, is $2j + 1$.

Table 1 Relation between the hole state and the electron configuration

Hole state	Electron configuration
K	[1s]
L_1	[2s]
L_2	$[2p_{1/2}]$
L_3	$[2p_{3/2}]$
M_1	[3s]
M_2	$[3p_{1/2}]$
M_3	$[3p_{3/2}]$
M_4	$[3d_{3/2}]$
M_5	$[3d_{5/2}]$

The jump is called the absorption edge, and the wavelength is highly correlated with the atomic number similarly to Moseley's law[1] in X-ray emission spectra. Moseley's law in emission spectra is expressed as Equation (1):

$$\frac{1}{\sqrt{\lambda}} = K(Z - s) \tag{1}$$

where λ is the X-ray wavelength, Z is the atomic number, and K and s are constants for a spectral series. The absorption coefficient is crudely proportional to $Z^4\lambda^3$ except for the edge jumps. The energy at which the jump is observed is called the threshold energy, but the definition of the threshold is not exact, because it corresponds to the transition from a core orbital to the lowest unoccupied orbital. The ionization limit is a few or a few tens of electron-volts higher than the edge energy.

The mass per unit area is given by ρd, where ρ is the mass density. The linear absorption coefficient μ_l is defined by $\mu_l = \mu\rho$, and its dimension is [cm^{-1}]. The X-ray attenuation length $1/\mu_l$ is the length at which the X-ray intensity becomes 1/e after traveling in matter. The attenuation length of Cu Kα_1 X-rays (8047.8 eV) is 79 μm in aluminum, 4.2 μm in iron, 24 μm in copper, and 3.9 μm in lead. The attenuation length of Al Kα X-rays (1486.7 eV) is 9.2 μm in aluminum and 0.4 μm in iron. The intensity of X-rays emitted by a copper target X-ray tube is, however, attenuated by only half after transmission through 2-mm-thick aluminum, but depends on the applied power on the X-ray tube, because the X-rays emitted from an X-ray tube are not monochromatic. Thus it should be noted that the X-ray shielding thickness for safety cannot be determined only from the monochromatic X-ray attenuation length.

The linear absorption coefficient can otherwise be expressed as $\mu_l = 4\pi\beta/\lambda$, where λ is the X-ray wavelength and β the imaginary part of the complex refractive index

$(n = 1 - \delta - i\beta)$.[7] The atomic form factor, $f = f_1 + if_2$, which is used in the analysis of X-ray diffraction, is related to the refractive index (Equations 2 and 3):

$$\delta = \frac{Nr_0\lambda^2 f_1}{2\pi} \tag{2}$$

$$\beta = \frac{Nr_0\lambda^2 f_2}{2\pi} \tag{3}$$

where N is the number of atoms in unit volume and $r_0 = e^2/(mc^2) = 2.818 \times 10^{-13}$ cm is the classical electron radius (e the electron charge, m the mass, and c the speed of light). The real part f_1 is the Fourier transform of the electron density distribution in an atom. The relation between the absorption coefficient and atomic form factor is used in DAFS described below.

The mass absorption coefficient is the sum of two effects: photoelectric absorption and scattering of X-rays. The photoelectric absorption is the ionization of an inner-shell electron. Therefore, the absorption coefficient due to the photoelectric part can be calculated by the photoionization cross-section.[8] The scattering part is due to the Rayleigh (coherent) and Compton (inelastic) scattering of X-rays, but X-ray absorption spectra are often taken as if they represent only the photoelectric absorption effect, although the experimental spectra contain both effects.

The mass absorption coefficients or physically equivalent parameters[8,9–13] and the absorption edge energy or wavelength[14,15] can be found in the literature. The value of the absorption edge energy is close to the electron binding energy, which is used in electron spectroscopy, ESCA (electron spectroscopy for chemical analysis) or XPS (X-ray photoelectron (photoemission) spectroscopy).[16,17] The absorption edge jump is not exactly the same as the electron binding energy, because the absorption edge energy corresponds to the excitation of core electrons into the lowest unoccupied molecular orbital (LUMO) in the molecular orbital picture, or Rydberg state in the atomic orbital picture. The Rydberg state and continuum state threshold are clearly seen in rare gas X-ray absorption spectra but are not clear for condensed matter. The difference between the vacuum level and Fermi energy, which defines the highest energy of electrons in a conduction band, is called the work function, Φ. This is another source of the difference between the electron binding energy observed in XPS and the absorption edge. The photoionization cross-sections[18,19] and the electron binding energies[20,21] can be found in the literature. The relation between the electron photoionization cross-section (barns) and mass absorption coefficient is simple when the angular dependence is averaged.[11]

Absorption techniques in X-ray spectrometry are used to measure the X-ray absorption spectra using various methods described below, and to analyze the obtained spectral line shapes to obtain information on the element, oxidation state, concentration, atomic distance, coordination number, surface geometry, and reaction on solid surfaces, catalysts, or electrodes.

2 ACRONYMS AND A BRIEF HISTORY

The mass absorption coefficient plotted against the X-ray energy is called the XAS. The X-ray absorption spectra of condensed matter near the threshold energy have fine structures as shown in Figure 2. Fine structures are sometimes observable at energies less than the threshold energy, and are called the pre-edge structure (Figure 3).[22] These fine structures are called the XANES, usually pronounced as ''zaenz''. The absorption fine structure will extend up to 1000 eV above the threshold energy, and thus it is called the EXAFS,[23–26] pronounced ''eksafs''. XANES is restricted from the threshold to ca. 50 eV above (this energy approximately corresponds to $kR = 2\pi$, where k is the ejected photoelectron momentum and R the nearest-neighbor atomic

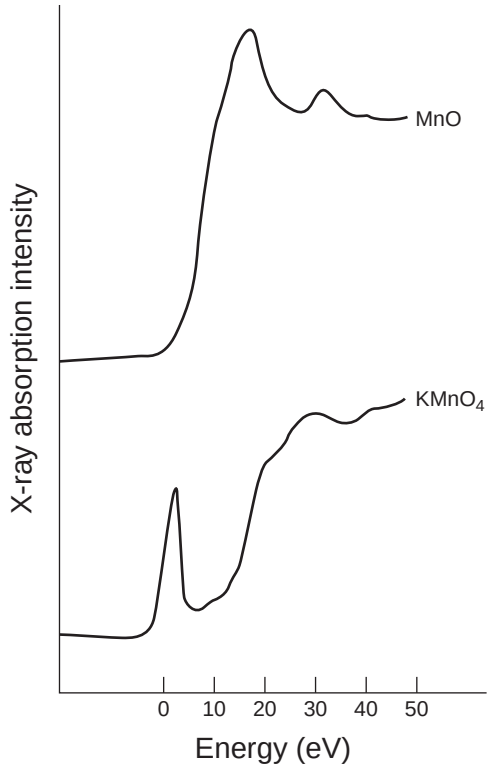

Figure 3 Mn K edge XANES spectra of MnO (octahedral, O_h, symmetry) and $KMnO_4$ (tetrahedral, T_d, symmetry). Pre-edge peak is found in the $KMnO_4$ spectrum. Chemical shift of the edge is found; the edge of $KMnO_4$ is higher than that of MnO. (Reproduced by permission from Pandya et al.[22])

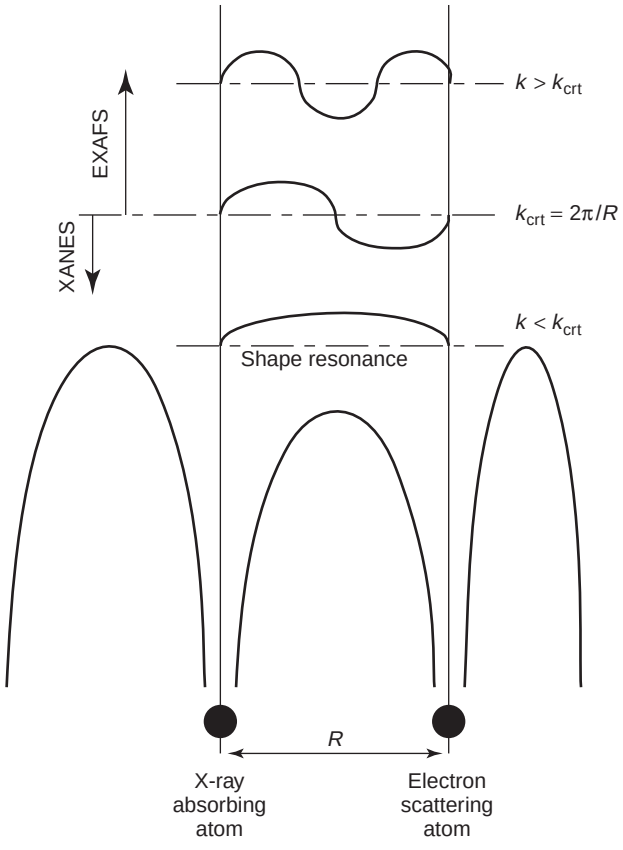

Figure 4 Illustration of electron standing wave between the X-ray absorption atom and its neighboring atom. (Reproduced by permission from Udagawa[9] and Ishii.[26])

distance as shown in Figure 4[9,26]). The momentum of the photoelectron is $k = [2m(E - E_0)]^{1/2}/\hbar$, where E is the photoelectron kinetic energy, E_0 the threshold energy, m the electron mass, and $\hbar$ Planck's constant. The photoelectron matter wave in a condensed system propagates as a spherical wave and forms a standing wave as shown in Figure 5.[9] Recently, XANES has come to be called near-edge X-ray absorption fine structure (NEXAFS),[27] pronounced ''neksafs''. All the fine structures including NEXAFS and EXAFS are grouped into the term X-ray absorption fine structure (XAFS), pronounced ''zafs''.

The history of the development of the understanding and application of XAFS has an interesting feature, as stated by Lytle et al.,[28] Shiraiwa,[29] Stern,[30] and Lytle.[31] Barkla (after Stern[30]) or de Broglie (after Lytle[31]) firstly found the X-ray absorption edge. Although XANES was found for both solids and gases, EXAFS was found only for condensed matter such as molecules, solids and liquids. EXAFS was first reported by Fricke in 1920[32] and was theoretically interpreted by Kossel.[33] He explained that the fine structure was due to the excitation of inner-shell electrons to an unoccupied level. This theory was valid for XANES,

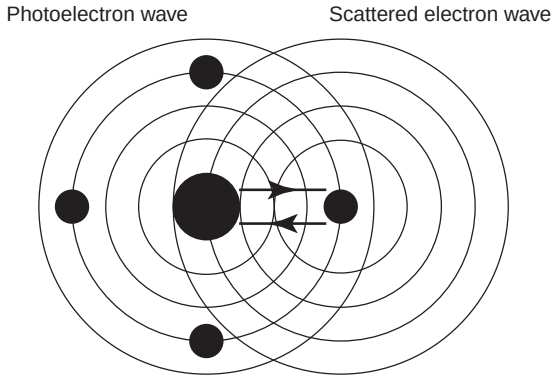

Figure 5 Schematic illustration of electron wave propagating and scattering in a solid. (Reproduced by permission from Udagawa.[9])

and thus XANES was called the Kossel structure. The Kossel theory was called short-range order (SRO) theory, because the electronic structures of unoccupied levels are mostly determined by orbital hybridization between the center atom and the nearest-neighbor atoms. On the other hand, Kronig[34] explained that fine structure was the result of the diffraction of photoelectrons as a matter wave when moving in a conduction band of a solid. The electron matter wave travels in a solid when the wavelength of an electron λ_e does not satisfy the Bragg condition, $2d \sin\theta = n\lambda_e$. When the Bragg condition is satisfied, then the electrons are scattered and leave the solid. His theory explained EXAFS and thus EXAFS was called the Kronig structure. His theory was called the long-range order (LRO) theory because the band structure is determined by the long-range periodic boundary conditions. Hayasi[35] considered that the electron waves that satisfied the Bragg condition form a standing wave in a solid, and thus the electron transition from an inner orbital to a standing wave state yields a maximum of X-ray absorption. Shiraiwa et al.[36] and Kozlenkov[37] improved the SRO theory to explain the EXAFS, but their method needed to solve a Schrödinger equation to obtain the EXAFS. Sayers et al.[38] proposed a Fourier transform method to obtain local structural information on condensed systems. Owing to their Fourier analysis, we do not need to solve the Schrödinger equation directly to obtain the local structure of matter. EXAFS had at that time great potential to be developed as a powerful method of analyzing the local structure of matter. The inelastic mean free path (IMFP) of a photoelectron is usually 2 nm. When the photoelectron is scattered inelastically, the coherence is forgotten. The coherent length, i.e. the length within which the electron matter waves emitted from a single source can interfere with each other, is an important length to apply in the EXAFS method to analyze a condensed system. When the IMFP is included in the LRO theory, it is equivalent to the SRO theory.

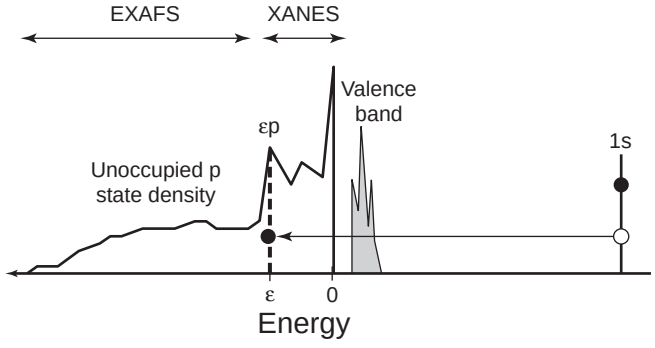

Figure 6 Schematic illustration of X-ray absorption and photoelectron excitation from the 1s to the unoccupied p state.

The XAFS represents the unoccupied electron density of states for atoms, molecules, solids, or liquids. One of the inner shell electrons, say a 1s electron, is excited into a discrete or continuum unoccupied state by the incident X-ray photon. The transition probability from the 1s to the unoccupied state equals the X-ray absorption intensity (only the photoelectric part is considered here), and thus the plot of the intensity against the incident X-ray energy is the XAS of a specimen. XANES is chiefly due to the transition from the inner shell to the unoccupied discrete level (Figure 6), and EXAFS is to the unoccupied continuum level.

3 X-RAY ABSORPTION NEAR-EDGE STRUCTURE

3.1 Chemical Shift and Line Shape

The XANES spectra show both the line shape modification and chemical shift[39] of the absorption edge or peak. Figure 7[40] shows typical examples for the S K edge for Na_2SO_4, Na_2SO_3, and $Na_2S_2O_3$. The sharp and prominent absorption peak shown in Figure 7[40] is called the ''white line''. This is because in the early days of X-ray experiments a white line developed on the X-ray film was observed. The white line for insulators is usually sharper than that for metals, because it corresponds to a $1s \rightarrow \pi^*$ electron transition, where the asterisk denotes an unoccupied antibonding orbital. The π^* state is usually a sharply localized state. The metal has a broad conduction band, and thus the absorption spectra show an edge jump but not a white line.

The white line energy plotted against the oxidation number of sulfur is shown in Figure 8.[41] The source of the chemical shift is both the unoccupied level shift and core level shift. The range of the unoccupied level shift ranges from the Fermi level (=0 eV) to the band gap energy (=a few electron-volts). The core level shift is due to the screening of core electrons by valence electrons;

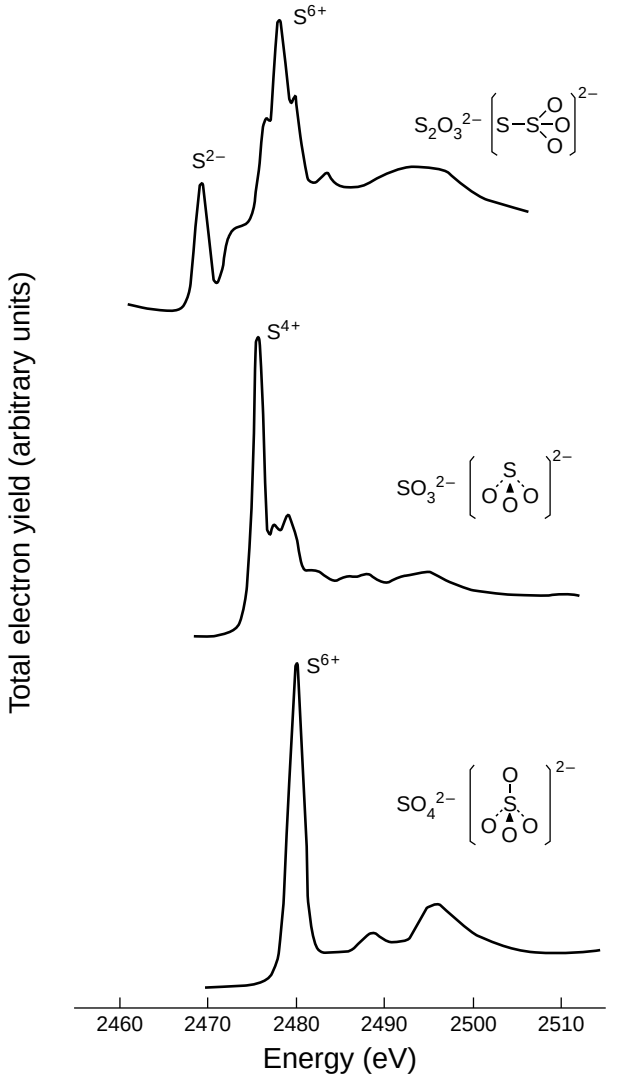

Figure 7 Sulfur K edge absorption spectra of Na_2SO_4, Na_2SO_3, and $Na_2S_2O_3$. (Reproduced by permission from Sekiyama et al.[40])

if the atom is negatively charged then the core level is shifted to a shallower binding energy, and if an atom is positively charged then it is shifted to a deeper energy. The source of the core level shift is the same as that of an ESCA chemical shift.

In Figure 9[42] is shown another example of a chemical shift of the absorption edge for Al compounds: Al metal, AlN, and four- and six-fold coordinated oxides. The Al−O distance of four-fold coordinated aluminum oxide (0.17 nm) is shorter than that of six-fold coordinated oxide (0.19 nm), because the oxygen ions interfere with each other and cannot be close to the Al atom for six-fold coordinated oxide. Thus the orbital hybridization of four-fold coordinated oxide is stronger than that of six-fold coordinated oxide, and consequently the six-fold form is

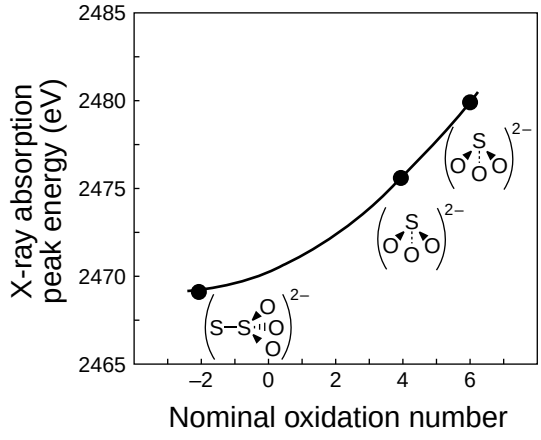

Figure 8 Relation between X-ray absorption peak and nominal sulfur oxidation number. (Reproduced by permission from Kawai et al.[41])

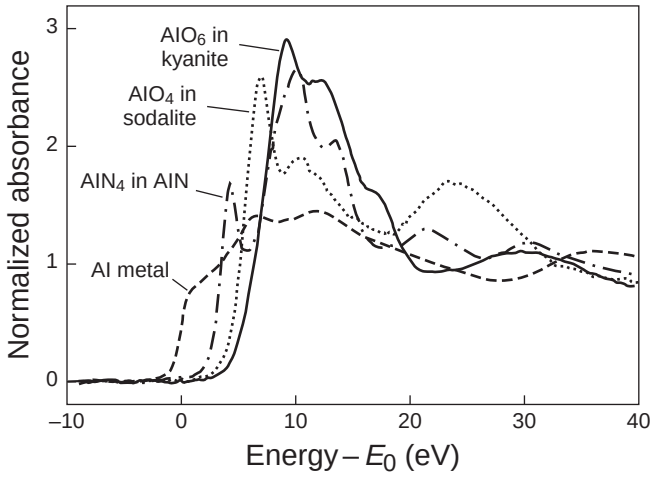

Figure 9 Al K edge XANES spectra of Al metal, AlN, sodalite, and kyanite. AlX_n denotes the first shell coordination of Al in each material. (Reproduced from J. Wong et al., 'New Opportunities in XAFS Investigation in the 1–2 keV Region', *Solid State Commun.*, **92**, 559–562, © 1994, with permission from Elsevier Science.)

ionic and the four-fold form is covalent. The effective positive charge of six-fold coordinated Al^{3+} is larger than that of four-fold coordinated oxide. The chemical shift of six-fold coordinated oxide is larger than that of four-fold coordinated oxide. The shift is strongly correlated to Pauling's electronegativity[43] of the neighboring atom, because the electronegativity determines the effective charge of the ion.

The unoccupied discrete level is composed of Rydberg states in the atomic picture, π^* and σ^* orbitals in the molecular-orbital picture (the asterisk denotes an antibonding molecular orbital), or conduction bands in crystals. The σ^* transition, which is formed in a potential well of neighboring atomic potentials, is called the shape resonance (Figure 4).

While the $1s \rightarrow \pi^*$ transition is a sharp white line, the $1s \rightarrow \sigma^*$ transition usually results in a broad and weak hump at higher energy,[44] which is called the shape resonance. The term shape resonance is used in the field of atomic spectra. The excited state or ionized state is bound in a potential wall, because of the centrifugal force potential of a high angular momentum orbital such as an f orbital, or surrounding potential such as F in SF_6. However, as shown schematically in Figure 4, such a surrounding potential does not have sufficient height to enclose the electron, but a weak resonance is observable. This is the origin of the term shape resonance.

The pre-edge structure shown in Figure 3 above is observed for the K edge of transition metal compounds whose local symmetry around the X-ray absorbing atom is T_d (tetrahedral). On the other hand, it is not observable for locally O_h (octahedral) symmetry solids. This pre-edge is sometimes said to be an electric quadrupole transition from 1s to 3d, whereas ordinary optical absorption is the electric dipole transition ($1s \rightarrow 2p$ or $2p \rightarrow 3s$, 3d). The quadrupole transition probability is, however, very weak, as shown in Table 2, where the probability is calculated by the Dirac–Fock method.[45] The origin of such a strong absorption as shown in Figure 3 is due to the electric dipole transition. The unoccupied p orbitals strongly hybridize with the d band for tetrahedral symmetry compounds based on the group theory as shown in Table 3,[46] where both p and d orbitals belong to the t_2 orbital. Thus the electric dipole transition is strongly observed at the energy of an empty d band. On the other

Table 2 Calculated transition probability for Cu

$K-L_1$	0.00000038
$K-L_2$	0.19
$K-L_3$	0.37
$K-M_1$	0.000000072
$K-M_2$	0.022
$K-M_3$	0.043
$K-M_4$	0.000025
$K-M_5$	0.000036

Table 3 Part of the character table of T_d

T_d	p	d
a_1		$x^2 + y^2 + z^2$
a_2		
e		$(2z^2 - x^2 - y^2, x^2 - y^2)$
t_1		
t_2	(x, y, z)	(xy, xz, yz)

Table 4 Part of the character table of O_h

O_h	p	d
a_{1g}		$x^2 + y^2 + z^2$
a_{2g}		
e_g		$(2z^2 - x^2 - y^2, x^2 - y^2)$
t_{1g}		
t_{2g}		(xy, xz, yz)
a_{1u}		
a_{2u}		
e_u		
t_{1u}		
t_{2u}	(x, y, z)	

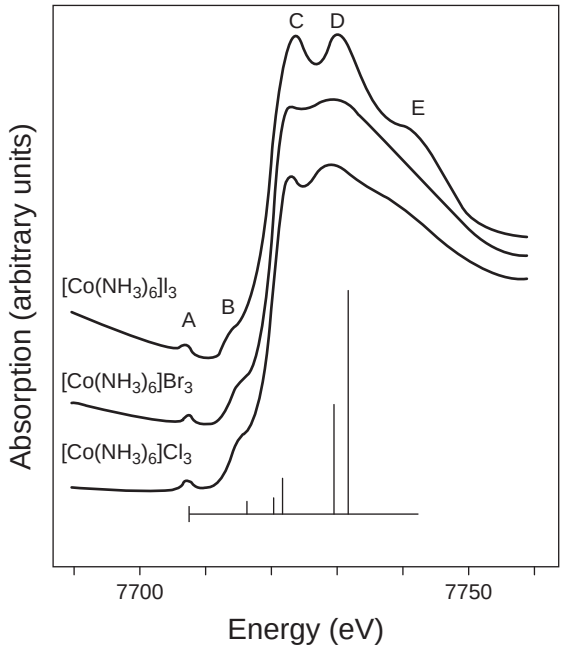

Figure 10 Co K edge spectra of $[Co(NH_3)_6]X_3$ (X = Cl, Br, and I). The vertical bars show calculated spectra for $[Co(NH_3)_6]^{3+}$. Peak A is due to the electric quadrupole transition but still mixed with the p orbital owing to the skewing of the molecular structure from exact octahedral symmetry. (Reproduced by permission from Sano.[47] © 1988 The American Chemical Society.)

hand, p and d orbitals never mix with each other for O_h symmetry, as shown in Table 4,[46] where the p orbital belongs to t_{2u} and d belongs to e_g and t_{2g}. Therefore, the p transition emerges at a different energy from the empty d band for octahedral symmetry compounds. The true electric quadrupole transition is very weak, as shown in Figure 10.[47] However, even in this case, the observable strength is due to the hybridization of the p character into the empty d band. The intensity and energy shift of the pre-edge peak are good indices for fingerprinting the chemical environment in the compounds, especially for biological samples.[48]

The chemical shifts of reference samples are measured and plotted against the electronegativity, and then the neighboring atom type is estimated for an unknown material from the chemical shift of the absorption edge.

After the discovery of high-temperature superconductors,[49] the understanding of the electron correlation effect of transition metal compounds[50] and rare earth compounds has been greatly improved by the study of XPS. Consequently, the understanding of the correlation effect, i.e. how the hole left in the final state of photoionization interacts with d holes in transition metal compounds, has developed substantially. Many reports have been published concerning the electron correlation effect on the XANES line shape of complicated materials.[51]

Mixed-valence rare earth compounds are clearly observed by the measurement of XANES, as shown in Figure 11,[52] but the intensity ratio sometimes does not directly represent the mixed-valence components because of a dynamic electron transfer, i.e. correlation effect, due to the core hole screening.[53] The peak decomposition of XANES spectra into Eu^{2+} and Eu^{3+}, as shown in Figure 11, yields a rough estimate of the mixed-valence state. However, the core hole created by the X-ray absorption rearranges the valence electrons and thus the peak intensity does not always represent the exact

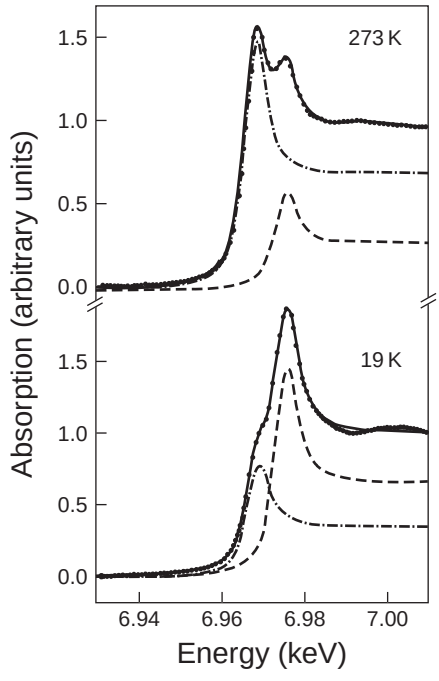

Figure 11 XANES spectra of $EuNi_2Si_{0.5}Ge_{1.5}$ at 19 and 273 K. The dashed-dotted and the dashed spectra indicate the Eu^{2+} and Eu^{3+} final state components, respectively. (Reproduced by permission from Wortmann et al.[52] © 1991 The American Physical Society.)

portion of the Eu^{2+} and Eu^{3+} states before the X-ray is absorbed.

3.2 Calculation Method for X-ray Absorption Near-edge Structure Spectra

The electronic states of photoelectrons whose kinetic energy is from a few electron-volts to a few tens of electron-volts are treated as conduction electrons in a conduction band. Thus a multiple scattering (MS) method or Green's function method, which has been used to calculate the electronic structure of conduction electrons near the Fermi energy in metals, is applicable to calculate the XANES of materials. The line shape of a XANES spectrum represents the partial and local electron density of states of the X-ray absorbing atom.[54] Hence any kinds of electronic structure calculations other than the MS theory, such as the LCAO-MO (molecular orbital derived from a linear combination of atomic orbitals) method or the APW (augmented plane wave) method, are also applicable to interpret the near-edge fine structure. One of the most popular methods for calculating XANES spectra is the MS theory.

In the MS theory, a sphere of radius r_i centered at the ith atom is considered, and the solid is divided by spheres. A spherically symmetric atomic potential $V(r)$ is put inside each sphere and the potential equals zero or constant outside the spheres. This is called the muffin-tin (MT) potential. The wave function in the solid is expressed as the overlap of spherical Bessel functions (radial part of the wave function) multiplied by the spherical harmonic functions (angular part of the wave function).

The wave function $\psi(r)$ of a photoionized electron is scattered by an atomic potential $V(r)$ near the ionized atom, and finally it becomes itself after being scattered many times (Equation 4)

$$\psi(r) = -\frac{1}{4\pi} \quad \frac{\exp(ik|r-r'|)}{|r-r'|} V(r')\psi(r')\,\mathrm{d}r' \qquad (4)$$

where $k^2 = \varepsilon$ is the kinetic energy of a photoelectron and k is real for $\varepsilon > 0$ (photoionized electron), $\exp(ik|r-r'|)/(|r-r'|)$ represents a spherically expanding wave, and $V(r)$ is the MT potential. This method is called the MS method, Green's function method, or Korringa–Kohn–Rostker (KKR) method.[55] The KKR method is only exact for solids that have translational symmetry, or periodic boundary conditions. Small clusters, molecules, amorphous or surface adsorbates have a lower symmetry, and it is difficult to apply directly the KKR method. Thus the cluster calculation method was proposed by Johnson[56] and was called the multiple scattering Xα (MS-Xα) method, because Slater's Xα exchange potential[57] is used in place of the Hartree–Fock (HF) exchange integral. The Xα method is also called the Hartree–Fock–Slater (HFS) method, and recently it has been developed as a local density approximation (LDA) theory. The calculation method for XANES spectra is a modified MS-Xα method.[58] In another way, the MS method is the expansion of the wave function of positive energy by an infinite sum of the spherically outgoing and incoming scattering waves. The electrons excited into the continuum level have a wave function of a standing wave formed by the infinite number of incoming and outgoing spherically traveling waves. That is to say, a wave whose intensity is $V(r')\psi(r')$ coming from every point r' in space is synthesized and forms a wave $\psi(r)$ at point r. This method produces a wave function similar to the APW method, which is an appropriate method to calculate a metallic band structure, but APW requires a greater number of basis functions than the MS method. The wave function at point r is the sum of all the scattered waves multiplied by the phase factor. The LCAO-MO method is another choice for calculating the electronic structures of solid or molecules, and is thus applicable to the calculation of XANES spectra.[59]

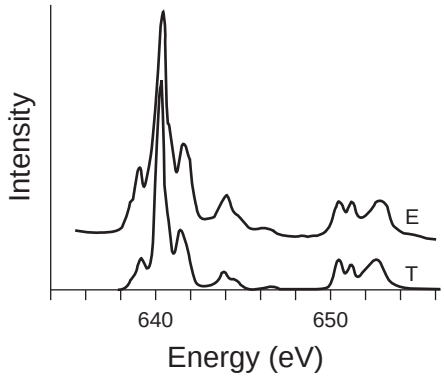

Figure 12 Mn $L_{2,3}$ XAS (E) of MnF_2 compared with atomic $3d^5$ multiplet calculation including the crystal field splitting (T). (Reproduced by permission from de Groot et al.[60] © 1990 The American Physical Society.)

An atomic calculation yields a satisfactory agreement between experiment and theory, as shown in Figure 12,[60] after the inclusion of the perturbation of crystal field splittings. Bragg reflection of electrons in a crystal reproduces a rough XANES spectrum.[61]

4 THEORY OF EXTENDED X-RAY ABSORPTION FINE STRUCTURE

4.1 Single Scattering Theory

Whereas MS of photoelectrons in a solid is a good approximation to treat XANES, because the electron kinetic energy of the EXAFS region is very high, single scattering is a good approximation to EXAFS except for

special cases. The wave function $\phi_k(r)$ of a photoelectron scattered by a single atom is asymptotically expressed by Equation (5):

$$\phi_k(r) \longrightarrow \exp(ikz) + \frac{f(\vartheta)}{r}\exp(ikr) \tag{5}$$

where $f(\vartheta)$ is the scattering amplitude and ϑ is the scattering angle ($\vartheta = 0^\circ$ for forward scattering and $\vartheta = 180^\circ$ for backscattering). The scattering amplitude of an electron of velocity v scattered by an atom of atomic number Z is expressed by the first Born approximation (Equation 6):

$$f(\vartheta) = \frac{e^2}{2mv^2}[Z - A(\vartheta)]\frac{1}{\sin^2(\vartheta/2)} \tag{6}$$

where e and m are the charge and the mass of an electron, respectively, and $A(\vartheta)$ is the atomic structure factor for X-rays, given by Equation (7):

$$A(\vartheta) = 4\pi \int_0^\infty \frac{\sin kr}{kr}\rho(r)r^2\,\mathrm{d}r \tag{7}$$

where $k = (4\pi mv/h)\sin(\vartheta/2)$ is the change in electron momentum before and after the scattering and $\rho(r)$ is the charge distribution in an atom. The forward scattering amplitude crudely depends on the atomic number in a way such that (Equation 8)

$$f(0^\circ) = \frac{1}{3}\int_0^\infty 4\pi\rho(r)r^4\,\mathrm{d}r = \frac{1}{3}Z\langle r^2\rangle \tag{8}$$

in atomic units, because (Equation 9)[62]

$$Z = \int_0^\infty 4\pi r^2\rho(r)\,\mathrm{d}r \tag{9}$$

where $\langle\ \rangle$ denotes an average. The calculated scattering amplitude is shown in Figure 13.[63]

The EXAFS is expressed by Equation (10):[64]

$$\chi(k) = -\sum_j \frac{N_j}{kR_j^2}|f_j(k,\pi)|\exp(-2\sigma_j^2k^2)\sin[2kR_j + \phi_j(k)] \tag{10}$$

where $k = \sqrt{2m(h\nu - E_0)}/\hbar$ is the photoelectron wave vector, N_j is the number of nearest neighbors, $|f(k,\pi)|$ is the backscattering amplitude, and R_j is the distance from the center atom. The exponential term contains the Debye–Waller-like vibrational effect and dumping. The dumping due to the finite coherent length of the photoelectron, $\exp[-2R_j/\lambda(k)]$, is multiplied for a more exact expression. The Debye–Waller factor contains both effects of thermal vibration and geometric randomness. The oscillating part of the EXAFS equation, $-\sin 2kR/(kR)^2$, if plotted as a function of kR, is the EXAFS oscillation.

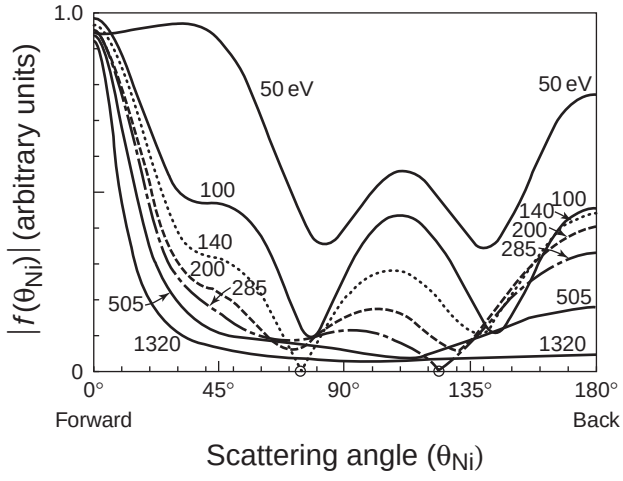

Figure 13 Calculated plane-wave scattering factor amplitude $|f|$ of nickel as a function of both the scattering angle θ_{Ni} and the photoelectron kinetic energy. (Reproduced from M. Sagurton et al., 'Derivation of Surface Structures from Fourier Transforms of Photoelectron Diffraction Data', *Phys. Rev. B*, **30**, 7332–7335, © 1984, with permission from Elsevier Science.)

The EXAFS oscillation amplitude is larger when the atomic number of neighboring elements is higher. For example, the Si K edge EXAFS oscillation amplitude of Si is stronger than that of SiO_2, because the atomic number of Si is higher than that of O. The white line of SiO_2 is sharper and stronger than that of Si. Hence the EXAFS oscillation and the white line intensity do not directly indicate the concentration of the atom in the analyte. However, the edge jump is a good measure of concentration, and the measurement of edge jump could determine the concentration without a working curve, as shown in Table 5.[65]

The effect of thermal vibration on the line shape of X-ray absorption spectra is shown schematically in Figure 14.[44] This is the line in the XANES region. Similarily, the EXAFS oscillation becomes unclear owing to thermal vibration. As the atomic number becomes

Table 5 Results of copper–zinc solution to test the trace element analysis

Zn ($\mu g\,mg^{-1}$ Cu solution)	Edge jump
0.092	0.0059
0.049	0.0035
0.020	0.0021
0.000	0.0010

Reproduced by permission from Nomura.[65] © 1992 The American Chemical Society.

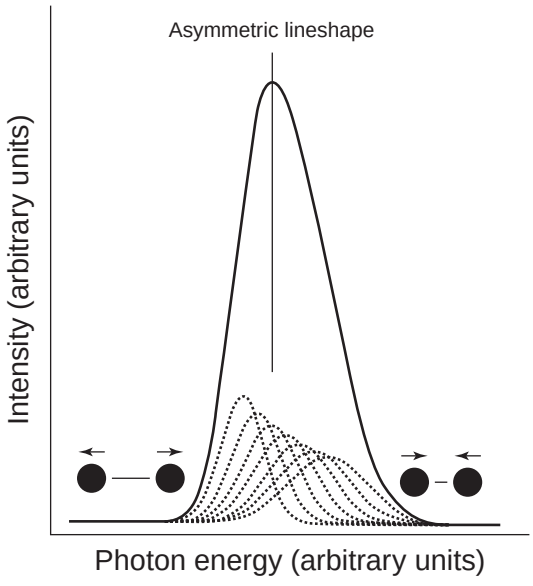

Figure 14 Asymmetric line shape of X-ray absorption spectra caused by the vibration of a diatomic molecule. (Reproduced by permission from D.A. Outka, J. Stöhr, 'Curve Fitting Analysis of Near-edge Core Excitation Spectra of Free, Adsorbed and Polymeric Molecules', *J. Chem. Phys.*, **88**, 3539–3554 (1988). © 1998 American Institute of Physics.)

higher, the core hole lifetime becomes shorter. A shorter lifetime of the inner shell level indicates that the energy of the inner shell becomes vague because of Heisenberg's uncertainty principle. Consequently, the EXAFS oscillation of the K spectrum for higher atomic number elements is not clear compared with that of the L edge spectrum of the same element.

The IMFP of photoelectrons is a function of kinetic energy for a particular material, as shown in Figure 15.[66] It is 1–2 nm for the usual EXAFS experiments. Hence the photoelectron is only coherent within a few nanometers,

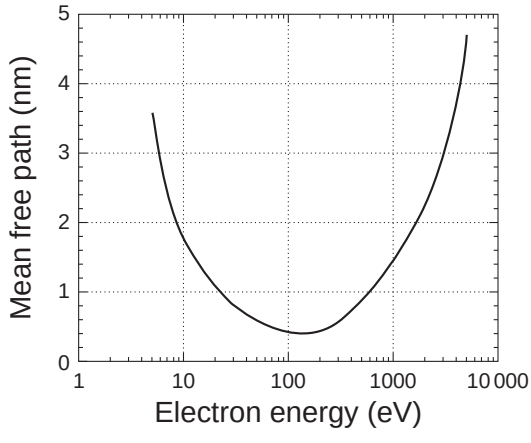

Figure 15 Calculated IMFP of photoelectrons in gold. (Reproduced by permission from Kuzmin.[66])

and it probes the local structure within the IMFP. Thus 1–2 nm regularity, usually up to the next-nearest neighbors, in the structure is sufficient for the EXAFS oscillation to emerge.

4.2 Relation to Other Techniques (X-ray Photoelectron Diffraction, Low-energy Electron Diffraction, X-ray Fluorescence Holography)

X-ray photoelectron diffraction (XPD)[62] is used to study the local structure of surfaces. The photoelectron intensity as a function of detected polar and azimuthal angles is measured in this technique. The photoelectron intensity is anisotropic in its detection angle. This effect is due to photoelectron diffraction, but roughly speaking it is due to the photoelectron's forward scattering by the nearest-neighbor atoms around the photoelectron-emitting atom. XPD uses forward scattering of photoelectrons; EXAFS uses backscattering of photoelectrons.

Recently XPD has been treated as photoelectron holography (PEH).[67] The intensity distribution of the photoelectrons of a single crystal is measured and Fourier transformed, and then a local atomic structure of the single crystal near the photoelectron-emitting atom can be constructed. The phase shift of electron waves scattered by a neighboring atom makes the analysis of the Fourier transform for PEH complicated. The angular distribution of the X-ray fluorescence intensity is measured and the Fourier transform of the angular distribution reconstructs the crystal image. This is called X-ray fluorescence holography (XFH).[67] XFH is free from the phase shift problem, but this method is bulk sensitive, whereas XPD and PEH are surface sensitive.

Low-energy electron diffraction (LEED)[68] is a surface crystallography experimental method where electrons of a few hundred electron-volts impinge on a single crystal and a diffracted electron pattern is observed. The penetration depth of these energy electrons is a few nanometers, hence this method is surface sensitive. Electron diffraction requires a periodic structure of at least 10 nm on the surface, hence the LEED method cannot probe the structure clusters of a few nanometers on a surface.

5 DATA ANALYSIS AND SOFTWARE PACKAGES FOR X-RAY ABSORPTION FINE STRUCTURE

The measured (Figure 16a) X-ray absorption spectral energy is converted into photoelectron momentum. The smooth background $\mu_0(E)$, which means an isolated atom absorption coefficient, is subtracted from the observed

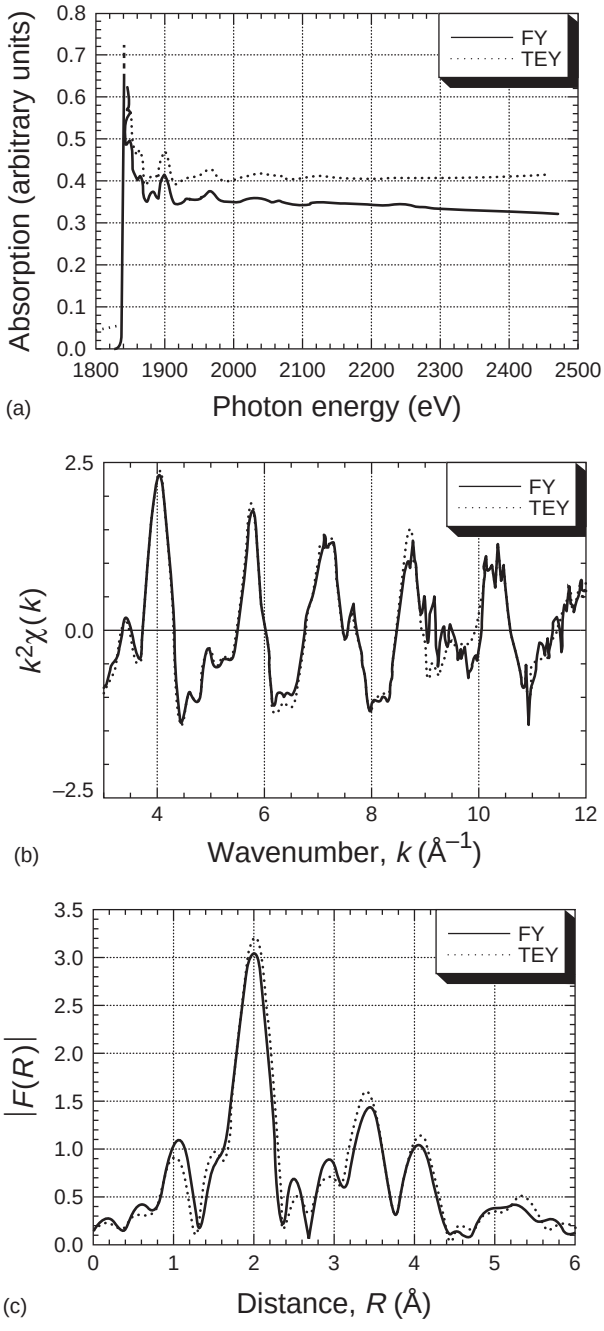

Figure 16 (a) Si K edge EXAFS spectra of an Si(001) wafer measured using the X-ray fluorescence yield (XFY) method and total electron yield (TEY) method. (b) Oscillation function of (a). (c) Fourier transforms of (b). (Reproduced by permission from Y. Kitajima, 'Fluorescence Yield X-ray Absorption Fine Structure Measurements in the Soft X-ray Region', *Rev. Sci. Instrum.*, **66**, 1413–1415 (1995). © 1995 American Institute of Physics.)

value $\mu(E)$ and normalized according to Equation (11):

$$\chi(k) = \frac{\mu(E) - \mu_0(E)}{\mu_0(E)} \tag{11}$$

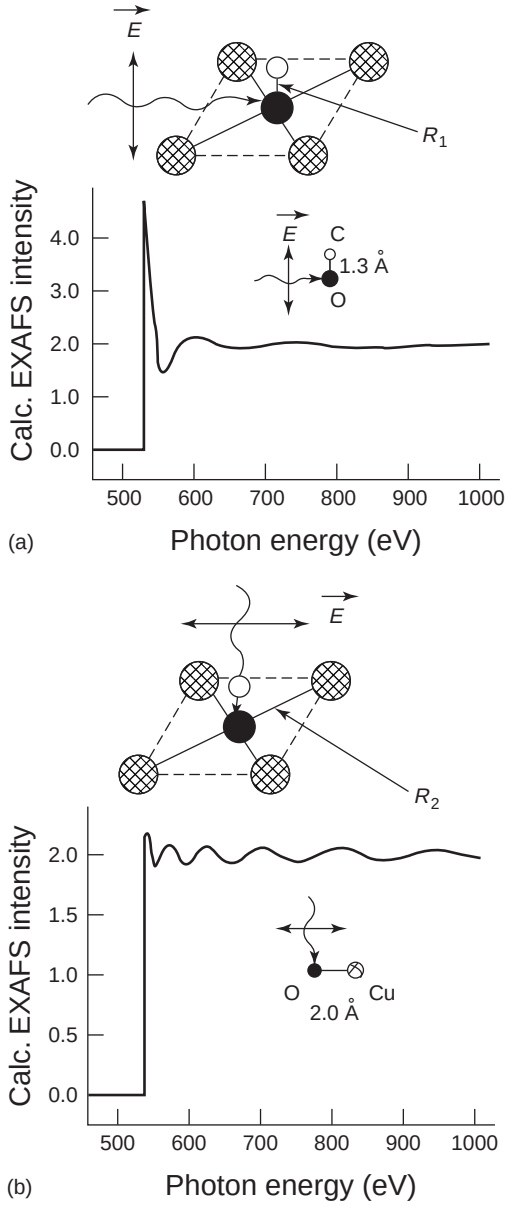

Figure 17 Schematic illustration of diatomic molecule (full and open circles) adsorbed on a four-fold coordinated site of metal substrate (hatched circles). When the incident X-ray beam electric vector is parallel to the diatomic molecular axis, the absorption spectrum of the molecule is obtained (a strong white line due to an insulating compound is observed). When the incident X-ray beam electric vector is perpendicular to the molecular axis, information on the metallic bond between the adsorbed atom and the substrate metal is obtained. (Reproduced by permission from J. Stöhr, 'Geometry and Bond Lengths of Chemisorbed Atoms and Molecules: NEXAFS and SEXAFS', *Z. Phys. B: Condens. Matter*, **61**, 439–445 (1985). © 1985 Springer-Verlag.)

This is the EXAFS oscillation. If we want to enlarge the oscillation for a larger k region, we sometimes plot $k^3\chi(k)$ or $k^2\chi(k)$ in place of $\chi(k)$ as shown in Figure 16(b).[69]

Then this is Fourier transformed as shown in Figure 16(c), and a radial distribution function is obtained. To obtain the coordination geometry, measurement of the polarization-dependent EXAFS oscillation is important, as shown in Figure 17(a) and (b).[70] An example of S K edge spectra of CS_2 on a Cu(111) surface is shown in Figure 18.[71]

A spline function or higher order polynomial determines the smooth background. Theoretically, the smooth background has a shape of $\tan^{-1}\theta$ at the threshold, because the discrete absorption line shape is a Lorentzian function and its sum in a Rydberg series becomes $\tan^{-1}\theta$ as shown in Figure 18.

An incident X-ray beam forms a standing wave in a large-sized single crystal. In this case, additional fine structure, the 1985-eV structure in Figure 19,[72] for example, depending on the incident angle of X-rays, is observable. The standing wave profile is sensitive to the location of impurity atoms in a crystal, i.e. which site in the lattice. The use of standing waves is one of the surface analysis methods, but sometimes interferes with obtaining $\chi(k)$, as shown in Figure 19.

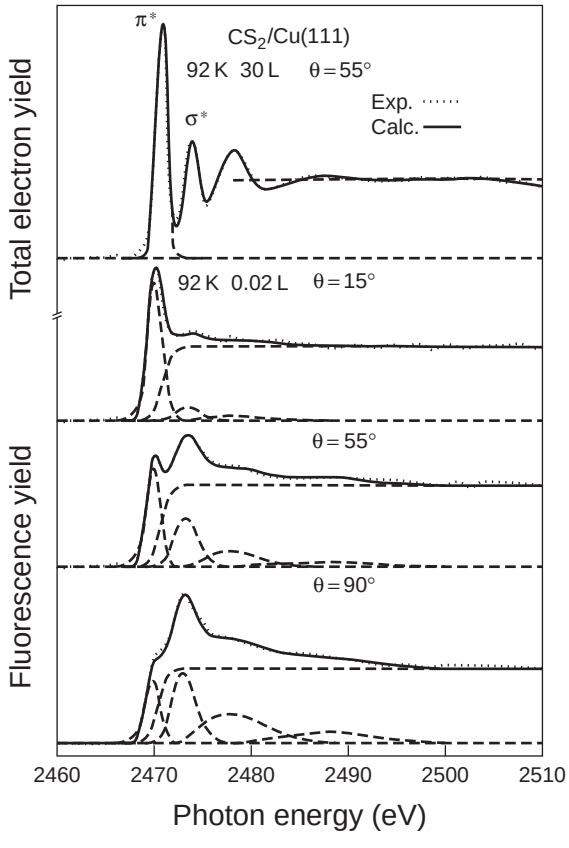

Figure 18 Experimentally obtained S K-edge XANES spectra (dots) of a CS_2 multilayer (30 L) at an X-ray incident angle of 55°, and submonolayer (0.02 L) at 15°, 55°, and 90°. (Reproduced from S. Yagi et al., 'Structural and Electronic Properties of Molecularly Adsorbed CS_2 on Cu(III) Studied by X-ray Absorption and Photoelectron Spectroscopies', *Surf. Sci.*, **311**, 172–180, © 1994, with permission from Elsevier Science.)

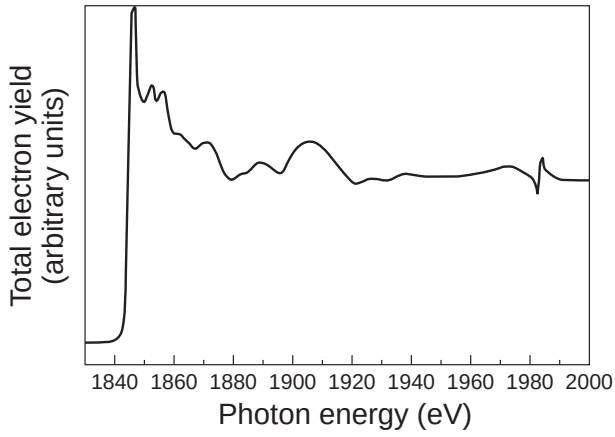

Figure 19 TEY spectrum of partially oxidized Si(111) wafer. Additional structure at 1985 eV is due to the incident X-ray Bragg diffraction (standing wave). (Reproduced from T. Ohta et al., 'A Possible Use of the Soft X-ray Standing Wave Method for Surface and Interface Structure Analysis', *Nucl. Instrum. Methods Phys. Res. A*, **246**, 760–762, © 1986, with permission from Elsevier Science.)

Another effect interfering with the observation of $\chi(k)$ is the multiple ionization effect. The effect of an additional one or two electrons ionized from outer shell(s) is not negligibly small.[73] The double ionization probability is sometimes more than 30% of single K shell ionization. This is a source of error in EXAFS analysis.

EXAFS Fourier analysis is sometimes not easy when additional peaks such as multiple ionization, standing wave structure, and impurity peaks originating from the analyte, X-ray source, or X-ray optics emerge.

Data analysis methods have been developed and several standard computer programs are now available.

Table 6 XAFS analysis computer programs

ATOMS	FUSE
AUTOBK	G4XANES
AUTOFIT	GNXAS
BAN	LASE
CDXAS	MacXAFS
CERIUS2	MURATA
EDA	REDUCE
EX.TR.As	REX
EXAFIT	REX2
EXAFS and FITEX	SEDEM
EXAFS (for Mac)	TT-MULTIPLETS
EXAFSPAK	UWXAFS
EXBACK	WinXAS
EXBROOK	XAFS
EXCALIB	XAID
EXCURVE98	XANADU
FEFF	XDAP
FEFFIT	XFIT

FEFF is the most popular program that is used by EXAFS users. The EXAFS analysis program is not difficult to code, and a laboratory that studies EXAFS may have its own program, but not always published. Some of the programs listed in Table 6 can be down-loaded from Web sites.

6 INSTRUMENTATION

6.1 Laboratory Extended X-ray Absorption Fine Structure

To measure the X-ray absorption spectra, a strong X-ray source of continuous energy is required, such as white radiation from an X-ray tube or synchrotron radiation (SR). A metallic wheel is rotated in vacuum and a high electric potential is applied between the wheel and a filament. Thermal electrons are emitted from the filament and bombard the wheel target. Water flows inside the wheel to cool it against heating by the electron bombardment. To eliminate the heating, the wheel is rotated. Therefore, this type of rotating anode X-ray tube produces one order of magnitude stronger X-rays than the ordinary sealed X-ray tubes. The electron deceleration at the metal target converts the electron kinetic energy into X-ray energy. The X-rays thus produced are continuum X-rays in addition to characteristic X-rays and the maximum energy is the acceleration electric potential applied. The X-rays from the X-ray tube are then monochromated by a crystal monochromator using the Bragg diffraction condition. Then the monochromatic X-rays are incident on the specimen as shown in Figure 20.[74] Sometimes, to compensate for the weak source intensity, a position-sensitive proportional counter is used as shown in Figure 21.[75] However, the very simple experimental set-up shown in Figure 22[76] is sometimes used.

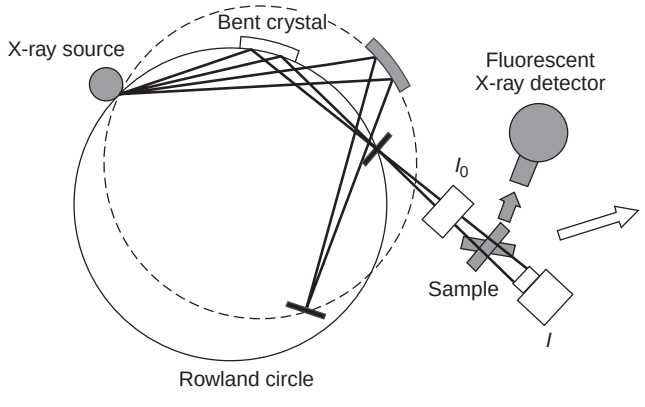

Figure 20 Example of EXAFS spectrometer using a rotating anode X-ray tube. (Reproduced by permission from Rigaku.[74])

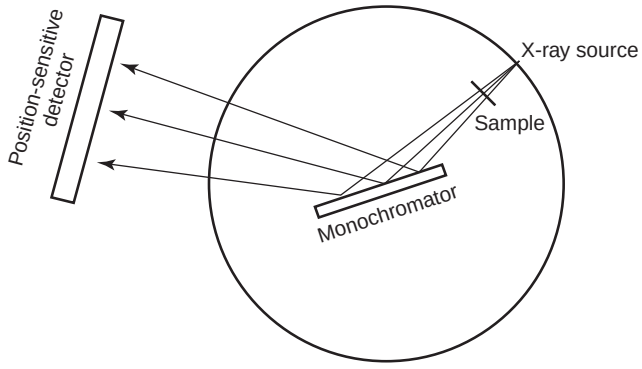

Figure 21 Example of EXAFS spectrometer using a position-sensitive proportional counter. (Reproduced by permission from Maeda et al.[75])

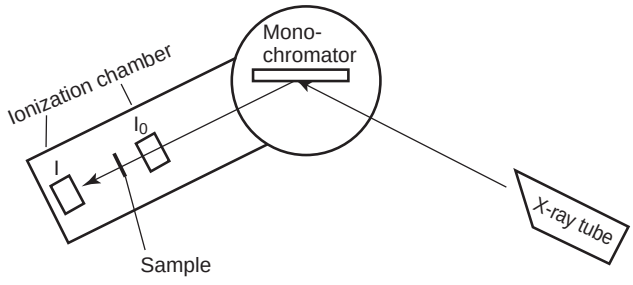

Figure 22 Laboratory EXAFS using a $\theta-2\theta$ goniometer. (Reproduced by permission from K. Sakurai, 'High-intensity X-ray Line Focal Spot for Laboratory Extended X-ray Absorption Fine-structure Experiments', *Rev. Sci. Instrum.*, **64**, 267–268 (1993). © 1993 American Institute of Physics.)

6.2 Synchrotron Radiation Extended X-ray Absorption Fine Structure

Recently, SR has frequently been used as an X-ray source. Synchrotron is the name of an electron (or positron) accelerator made of an ultrahigh vacuum (UHV) ring and electrons rotating inside the ring. High voltage is applied to the rotating electrons using variable-frequency radio waves and the frequency is synchronized with the electron rotation during the acceleration of the electrons in the ring. When the electron speed reaches close to the speed of light, a very sharp X-ray beam is emitted in a tangential direction because of the relativistic effect. This X-ray beam is called the SR. Usually SR from a storage ring is used. A ring in which electrons (or positrons) are rotated at a certain constant speed is called the storage ring. Usually electrons accelerated to a sufficient speed by a synchrotron or a linear accelerator are injected into the storage ring, and the tangential radiation when a magnet bends the electron beam is used as an X-ray source.[77] The SR thus produced from a bending magnet is continuous over a wide range (more than 1000 eV) of X-ray energy and a few orders of magnitude stronger than the rotating

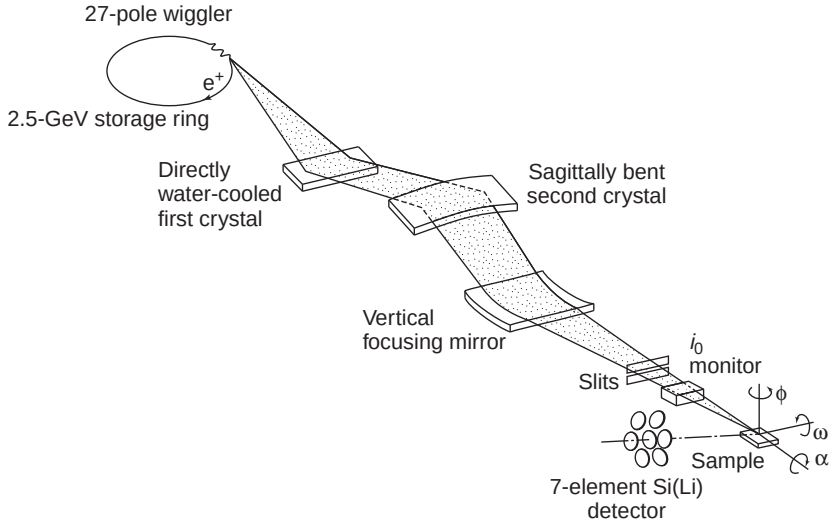

Figure 23 Schematic view of X-ray wiggler beam line. (Reproduced by permission from Oyanagi et al.[78])

anode X-ray tube. The smaller emittance indicates the electron orbit stored in a storage ring being sharpened, but practically it indicates a smaller X-ray beam size and higher photon density. The emittance is measured in meter · radians (m · rad). The emittance is a measure of beam quality. The smaller the emittance, the more the beam becomes parallel. The SR is monochromated by a single crystal.

Continuous X-rays from a bending magnet are most convenient for X-ray absorption spectroscopic experiments. The rotating axis of monochromator crystals is parallel to the electric vector of the SR X-rays to make good use of the X-rays. To obtain stronger X-rays, SR from an undulator or wiggler beam line is used. The undulator and wiggler are insertion devices in the storage ring, and made of many strong permanent magnets. The electron beam in the storage ring is undulating when it goes through an undulator, and a coherent quasi-monochromatic X-ray beam is produced. It consists of many harmonics and each harmonic has a narrow (say 100 eV) bandwidth. By changing the magnet gap width, the peak energy of the X-rays from the undulator is controlled; the wider the gap, the lower is the energy. Thus, to scan the energy over 1000 eV continuously using an undulator, both the undulator gap and monochromator crystal rotation should be controlled simultaneously. This is a difficult task but now routinely done in some SR undulator beam lines. An undulator can produce a few orders of magnitude stronger X-rays than a bending magnet beam line. If the X-rays produced by undulating by magnets are not coherently interfered, such an insertion device is called a wiggler. X-rays from a wiggler are not strong compared with those from an undulator but are continuous in energy, and thus much easier to use in X-ray absorption experiments than those from an undulator. An example of the experimental set-up of a wiggler is shown in Figure 23.[78]

SR facilities are classified into first-, second-, and third-generation sources. The first-generation synchrotrons were particle accelerators, and spectroscopists parasitically used the SR. Such SR was unstable. The second-generation synchrotrons use a storage ring to obtain SR but the emittance is still not small enough, i.e. 10^{-7} mrad. The emittance of the third-generation synchrotrons have emittance as small as 10^{-9} m · rad. Such a small-emittance storage ring requires a large-radius electron orbit because a large number of magnets are required to keep the electron beam cross-section small. The fourth generation of synchrotron has not yet been constructed and not defined, but may be a free electron laser facility using an electron accelerator. The third-generation SR facilities are tabulated in Table 7, and the second-generation SR facilities can be reached via the links of the Web pages listed.

The first crystal in the monochromator used in an SR beam line experiences a high heat load, and the lattice constant is slightly different from the second crystal. The crystal optics should be cooled by flowing water. To adjust the difference in the lattice constant between the first and the second crystals, usually one of the two crystals is independently finely moved by a piezoelectric

Table 7 Third-generation SR facilities and Web sites

ALS	http://www-als.lbl.gov/als/
APS	http://epics.aps.anl.gov/welcome.html
ESRF	http://www.esrf.fr/
SPring-8	http://www.spring8.or.jp/

mechanism. The adjustment of the two crystals should always be monitored for the measurement of EXAFS for the 1000-eV range. The monitoring of the adjustment is done by measuring the incident X-ray intensity, which is maximized at every energy point scanned.

6.3 Secondary Yield Techniques and Applications

The intensity of the monochromated X-rays is monitored by an ionization chamber. X-rays pass through the ionization chamber (I_0) and are then incident on a specimen, and the transmitted X-ray intensity is measured by another ionization chamber. The XAS is the plot of $-\log(I/I_0)$ against the X-ray energy. The absorption spectra are usually measured by this transmission method as shown in Figure 24(a).[79] When the X-rays are absorbed strongly by the sample, the secondary responses such as photoelectron intensity, sample electric current intensity, secondary electron intensity (electrons whose kinetic energy < 50 eV is mostly the secondary electrons), Auger electron intensity, and X-ray fluorescence intensity become strong. Therefore, equivalent spectra to the absorption spectra are measurable by these secondary phenomena such as photoelectron intensity, secondary electron intensity, Auger electron intensity (Figure 24b), sample drain current (Figure 24c and d), X-ray fluorescence intensity (Figure 24e), ion intensity due to photostimulated desorption, and other secondary techniques.

The photoelectron intensity for a single crystal has an anisotropy with respect to the observed direction because of the photoelectron diffraction. The angular average of the photoelectron intensity measured as the change in incident X-ray energy is the XAS. This method is called the photoelectron yield method. If Auger electrons are detected, this is the Auger electron yield method. The intensity of Auger electrons from a single crystal also has an angular anisotropy. The detection angle is therefore important for interpreting the observed data for these electron yield methods. Many kinds of electrons are detected, such as secondary electrons, core photoelectrons, valence photoelectrons, and Auger electrons, as shown in Figure 25(a–c).[80] The different kinetic energy electrons correspond to the different probing depths. Thus the electron yield spectra are a mixture of various depth spectra as well as a mixture of various processes of electron production. The relation between the X-ray absorption process and the electron emission process is neither direct nor clear.

The electric drain current is of the order of 10^{-9} A when using a bending magnet SR beam line. The electric drain current is a measure of X-ray absorption, and this method is called the TEY method. This is because the drain current represents the sum of all the electrons

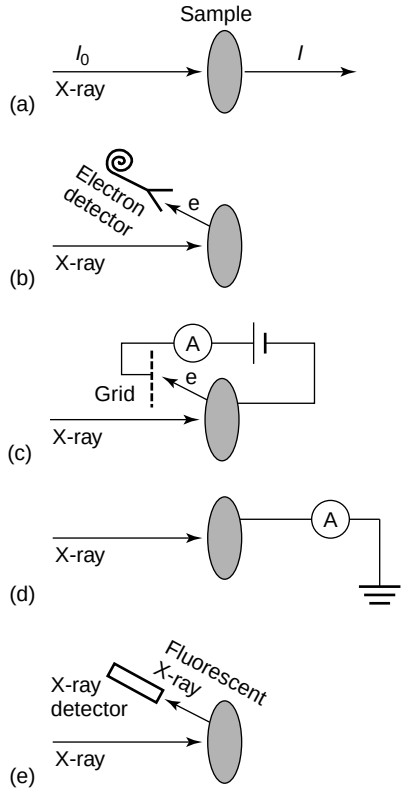

Figure 24 Various methods of measuring X-ray absorption spectra. (a) Transmission method; (b) partial electron yield method; (c, d) TEY method; and (e) XFY method. (Reproduced from J. Kawai et al., 'Depth Selective X-ray Absorption Fine Structure Spectroscopy', *Spectrochim. Acta Part B*, **49**, 739–743, © 1994, with permission from Elsevier Science.)

emitted from the sample. The electrons produced in a solid are scattered in the solid as shown in Figure 26,[81] and only electrons produced near the surface contribute to the electric drain current. When we detect ions desorbed from the surface due to X-ray absorption, the ion intensity represents the surface top layer.

These methods of detecting electrons (or electric current) or ions are surface sensitive and therefore are called surface extended X-ray absorption fine structure (SEXAFS). If the sample current is measured in an air or helium atmosphere, then the ejected electrons are converted into O^{2-} or other ions and a positive or negative current is observable depending on the sample electric potential with respect to the ground. This method is called the conversion electron yield (CEY) method.

The XFY method is not surface sensitive, because the fluorescent X-rays originate from as deep a location as the X-ray attenuation length. However, if XFY is combined with the grazing-incidence X-ray technique (Figure 27),[82] where the total reflection X-ray technique is combined with the TEY and CEY methods, it becomes more surface sensitive than normal-incidence TEY or

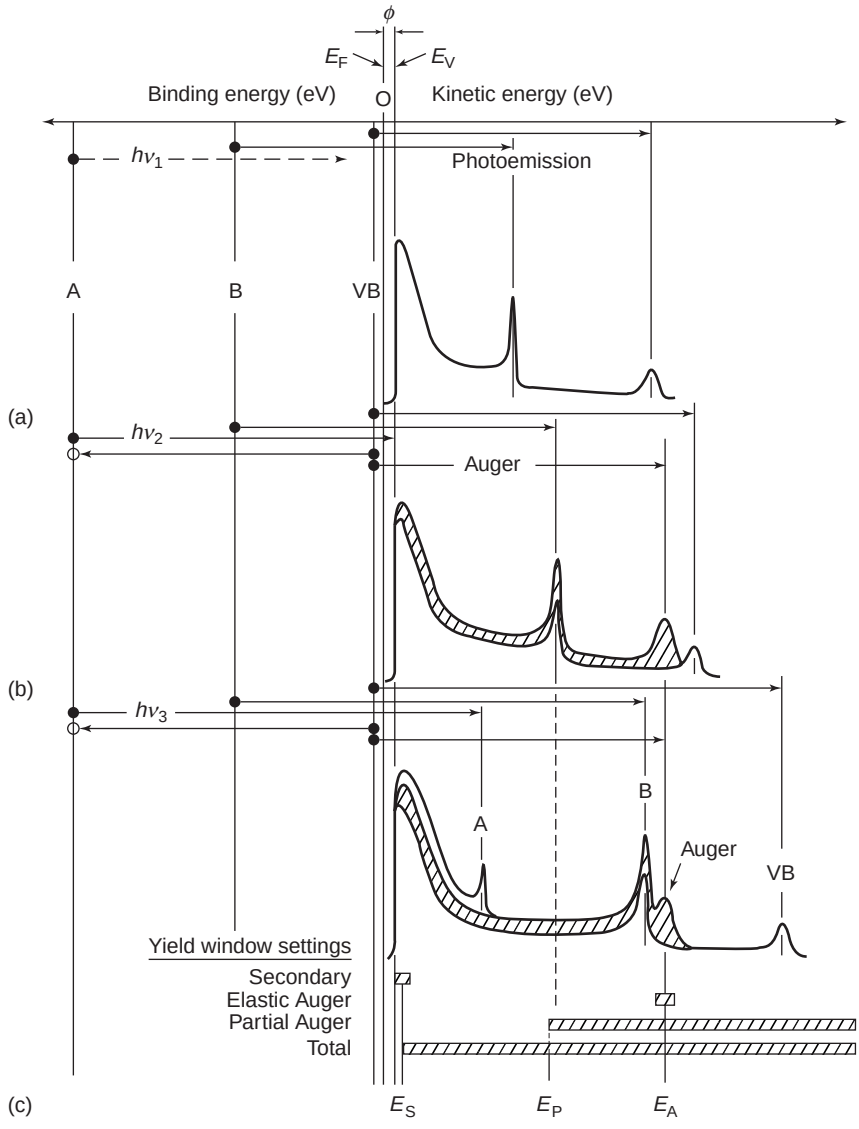

Figure 25 Schematic photoelectron spectra for (a) the excitation energy below the excitation threshold of core level A, (b) just above the absorption threshold but below the photoionization threshold, and (c) far above the threshold. The contributions to the photoelectron intensity from different level photoelectrons and Auger electrons are indicated. At the bottom the window settings for various electron yield X-ray absorption techniques are shown. (Reproduced by permission from Stöhr et al.[80] © 1984 The American Physical Society.)

CEY. The X-rays are totally reflected on the surface and the evanescent X-ray wave penetrates only a few nanometers from the surface. X-ray reflectivity is also a measure of XAFS.

The grazing exit angle method is also possible, as shown in Figure 28,[83] because of the reciprocal theorem of optical beams. The XFY method has atomic selectivity because an X-ray detector usually has energy resolution, hence the signal-to-noise ratio is better than in the TEY method, and minor elements adsorbed on the surface can be detected by the XFY method. X-ray absorption spectra of very dilute analytes can be detected using the XFY method; the detection limit is $<10^{12}$ atoms cm^{-2}.

The X-rays emitted from a deep location in a specimen suffer from the self-absorption effect and the spectral shape is different from that of an absorption-free spectrum.[84] The XFY method combined with the grazing incidence method does not suffer heavily from the self-absorption effect compared with the normal-incidence method.

Some materials emit luminescence in the visible wavelength range when irradiated with X-rays. This optical luminescence signal intensity corresponds to the

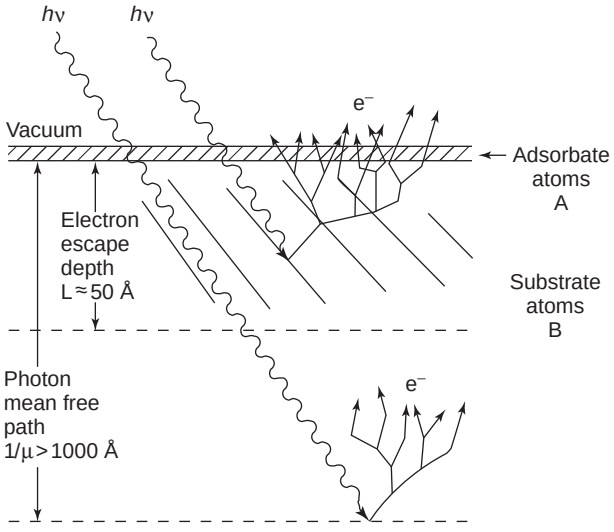

Figure 26 Photoabsorption and electron production in a solid consisting of substrate atoms B and an adsorbed layer A. Only electrons originating within a depth L from the surface will contribute to the TEY signal. (Reproduced from J. Stöhr et al., 'Surface Crystallography by Means of Electron and Ion Yield SEXAFS', *Surf. Sci.*, **117**, 503–524, © 1982, with permission from Elsevier Science.)

amount of X-rays absorbed by the specimen. Various processes of optical luminescence de-excitation are shown in Figure 29.[85] Optical luminescence is usually strong for rare earth compounds, but some crystals which have defects in their crystal structure emit stronger luminescence, although a perfect single crystal of the same material does not emit optical luminescence.

The photoacoustic (PA) effect produces sound on irradiating a sample surface by a chopped photon beam. This effect was discovered by Alexander Graham Bell.

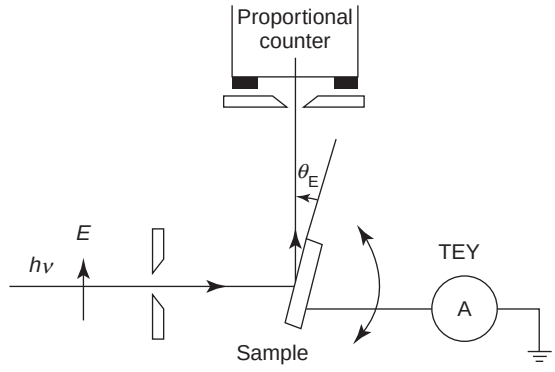

Figure 28 Schematic experimental set-up of grazing exit angle experiment for XFY and TEY. (Reproduced by permission from Kitajima.[83])

Heat is produced while the sample is irradiated by an optical beam and it diffuses during the beam chopping. The chopping frequency usually ranges from a few to a few hundred hertz. The sound wave is detected by a microphone or a piezoelectric device. When the chopping frequency is low, the heat diffuses into deeper location in the sample compared with when it is high. Thus the probing depth is variable by changing the chopping frequency. All the incident photon energy is finally converted into thermal energy through nonradiative transition processes in the solid. The PA process in the visible wavelength range is used for the very sensitive absorption spectrometry of thick bulk samples, which are not transparent to an optical beam. This method has been applied to the measurement of the X-ray absorption,[86] where the incident X-ray beam should be chopped to produce acoustic waves in the sample. This method is a thermal yield method.

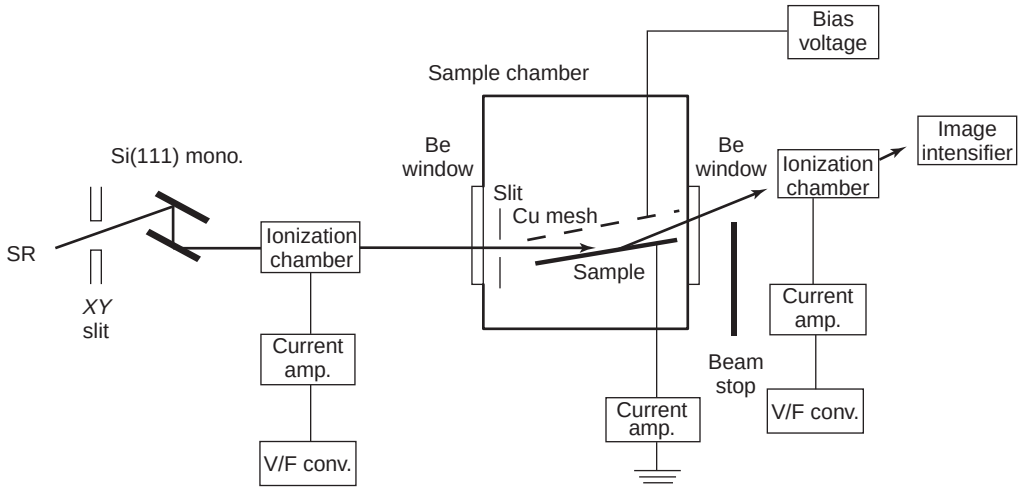

Figure 27 Schematic experimental set-up for TEY and CEY at grazing angles. (Reproduced by permission from Zheng and Gohshi.[82])

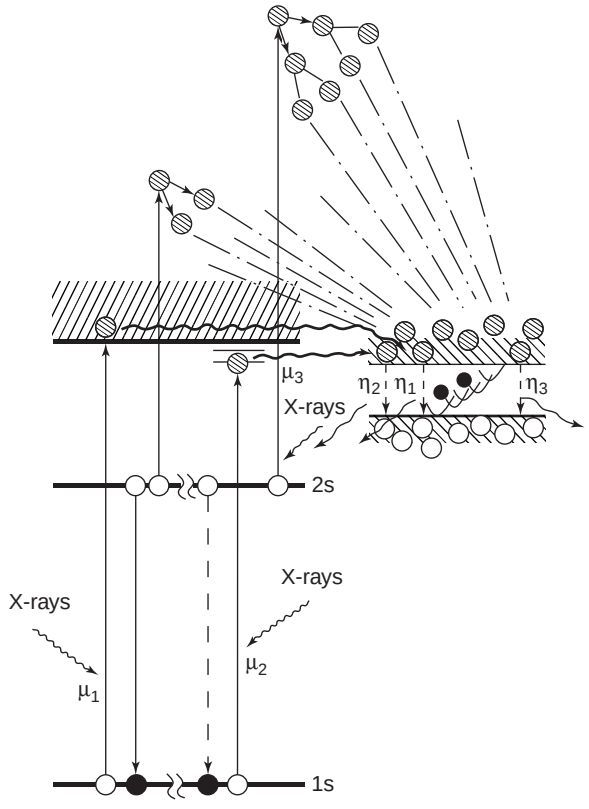

Figure 29 Schematic diagram of X-ray absorption and optical luminescence processes. Three different excitation processes, from the 1s orbital to a continuum state (absorption coefficient μ_1) and to a bound state (μ_2), and from the 2s orbital to a continuum state (μ_3), give rise to a single luminescence with luminescence yields η_1, η_2, and η_3, respectively. X-ray fluorescence, KLL Auger, electron multiscattering, nonradiative decay, and radiative decay processes are schematically shown. (Reproduced by permission from Emura et al.[85] © 1993 The American Physical Society.)

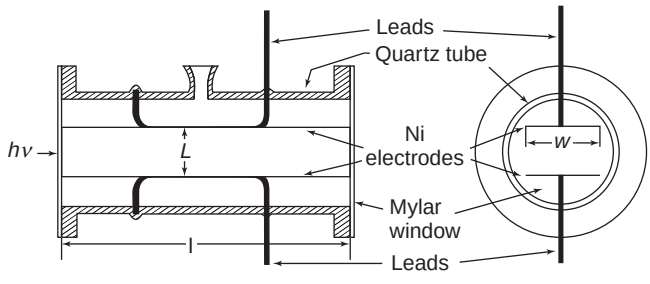

Figure 30 Schematic illustration of a liquid cell. (Reproduced by permission from T.K. Sham, R.A. Holroyd, 'Photoconductivity Measurements of X-ray Absorption of Liquids: Fe K-edge Spectrum of Ferrocene in 2,2,4-Trimethylpentane', *J. Chem. Phys.*, **80**, 1026–1029 (1983). © 1983 American Institute of Physics.)

Liquid samples are inserted into a cell shown in Figure 30,[87] and an electrode is used to measure the photoconductive spectra; electrical conductivity is induced by the incident X-rays. Spectra thus measured are sometimes the inverse of the transmission spectra and sometimes similar to the transmission spectra, depending on the concentration.[88] Electrode surface reaction processes can be measured using X-ray absorption in combination with the total reflection X-ray technique.

X-ray submicrometer beams are now available in major SR facilities, and using these beam lines micro- or nanobeam techniques are now applicable, as shown in Figure 31,[89] where the XFY is measured by a solid-state detector (SSD). The energy resolution of an SSD is of the order of 100 eV, while the energy shift of an absorption edge is a few electron-volts. If the incident X-ray energy is between the edge energies of two chemical states (say FeO and Fe_2O_3), then only one kind of chemical state (FeO) can emit the X-ray fluorescence. Using this technique, chemical state mapping is possible.

Using a bent crystal monochromator as shown in Figure 32,[90] multiple-energy X-rays can be focused on a sample and the transmitted X-rays are detected by a position-sensitive detector, which is made of a photodiode array (PDA). Using such a kind of energy-dispersive optics, one spectrum can be measured within a few milliseconds.[91] This method is called quick X-ray absorption fine structure (Q-XAFS).

Circularly polarized X-rays can be produced by linearly polarized X-rays transmitted through a diamond single crystal at a special angle depending on its wavelength, which is called the phase retarder, as shown in Figure 33.[92] The vertical and horizontal components, the ratio of which is called helicity, of the X-ray electric vector can be controlled by the small rotation and tilting of the diamond crystal. Absorption spectra of a magnetic sample are measurable. This is X-ray magnetic circular dichroism (XMCD) X-ray absorption. The magnetic S and N poles applying the magnetic field to the sample are inverted

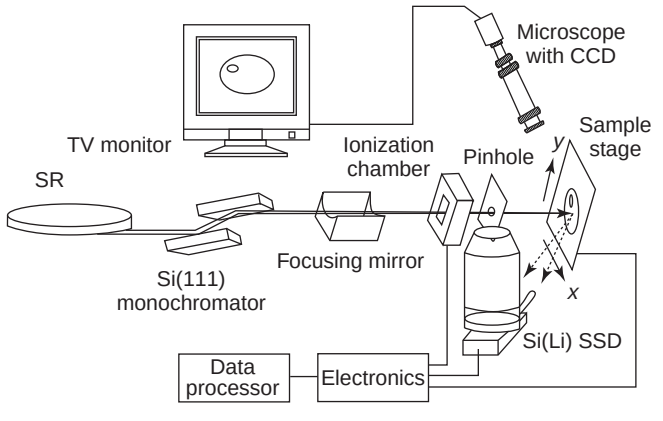

Figure 31 Schematic illustration of an X-ray fluorescence microprobe at the Photon Factory. (Reproduced from Nakai et al.,[89] courtesy of Marcel Dekker, Inc.)

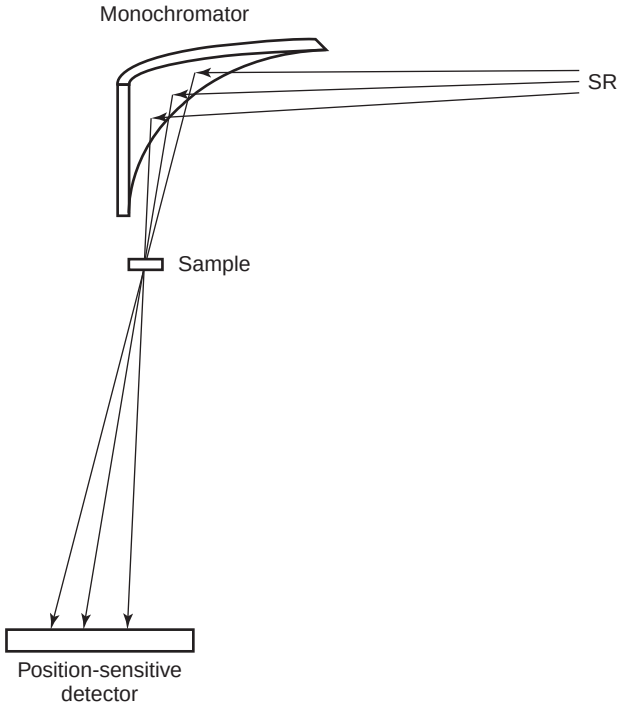

Figure 32 Energy-dispersive system at a synchrotron facility. (Reproduced by permission from Derbyshire et al.[90])

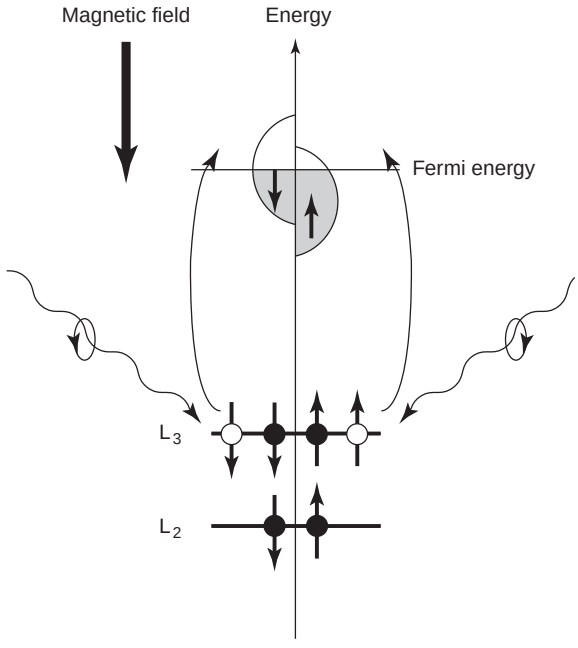

Figure 34 One-electron model used to explain the X-ray absorption dichroism process and intensities. The d band is split into spin-up and spin-down bands. The absorption of circularly polarized X-ray photons by the spin–orbit split 2p shell creates a spin-polarized core hole. (Reproduced by permission from Duda et al.[93] © 1994 The American Physical Society.)

or the circularly rotating direction of the X-ray electric field is inverted and the difference in the absorbance is measured. The former is usually used to measure the dichroism. Magnetic thin multilayers have recently become important for information mass storage devices, and these materials are characterized with microbeam XMCD X-rays. The difference in the absorption coefficients for the left and right circularly polarized X-rays is illustrated schematically in Figure 34.[93] The details of MCD are described in several books.[94–97]

High-resolution X-ray fluorescence spectra of transition metal compounds show multiplet splitting due to the exchange interaction between the unoccupied 3d level and the core X-ray hole. Thus the XFY absorption spectrum of each multiplet line provides spin-selective absorption spectra, as shown in Figure 35.[98] This method can measure spin-selective X-ray absorption spectra without applying a magnetic field to the sample. This method is useful for characterizing mixed-valence protein compounds.

The phase transition due to the temperature change is observable, as shown in Figure 36.[99] The phase transition is a small change in bond distance and bond angle, and consequently the electronic structure of the sample changes. Thus both XANES and EXAFS region spectra change their line shapes.

The surface of water, where a liquid monolayer is present, could be analyzed by grazing incidence X-ray reflection XAFS.[100] When the monolayer absorbs metal

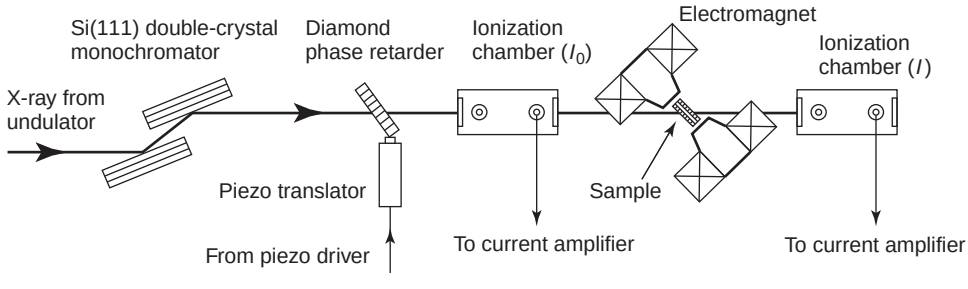

Figure 33 Experimental set-up for XMCD measurements with the helicity-modulation technique. (Reproduced by permission from Suzuki et al.[92])

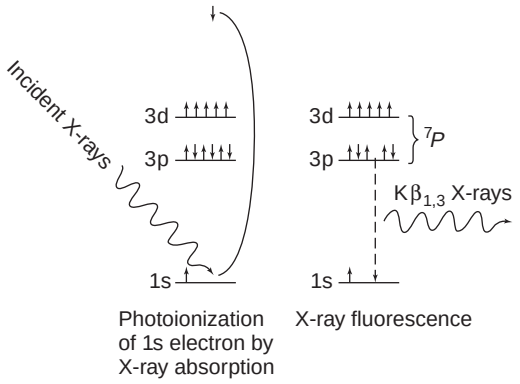

Figure 35 Schematic illustration of absorption and emission ($K\beta_{1,3}$ lines) of X-rays due to 1s down-electron photoionization of Mn^{2+} ion; 1s up-electron photoionization produces the $K\beta'$ X-ray emission spectral line. (Reproduced by permission from Grush et al.[98])

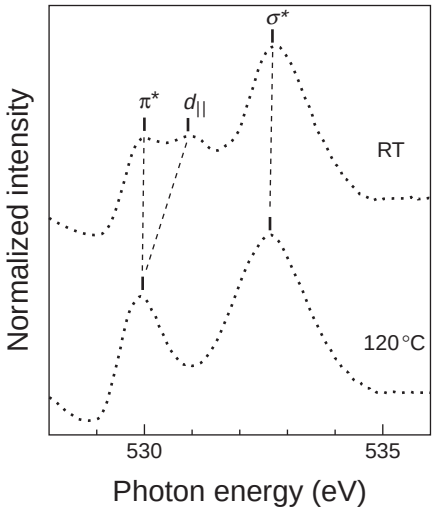

Figure 36 O 1s absorption spectra of VO_2 measured at room temperature and at 120 °C. (Reproduced by permission from Abbate et al.[99] © 1991 The American Physical Society.)

ions from the water solution, the concentration of the metal ion on the surface is slightly higher than that in the water. The coordination structure around the metal ion is analyzed by the EXAFS method.

When a powder is measured by the XFY and TEY methods on a substrate, then the depth-selective chemical state analysis of a fine particle can be performed. Fly ash is a powder of micrometer-sized particles, which are a source of acid rain when they are dispersed in the air. The particles are put on an aluminum foil and irradiated by monochromated X-rays, and XANES spectra are measured by TEY and XFY methods.[101] The TEY method is sensitive to the surface chemical state of the powder particle, and XFY is sensitive to deep parts of the particle (micrometers). The chemical shift of the white line in the absorption spectra is a measure of the oxidation state of an element in the particle.

The surface catalyst process could be elucidated by the analysis of EXAFS spectra. The incident X-ray polarization dependence is an important parameter for the determination of the geometry of a reactant and the surface.[102]

X-ray detectors in the XFY method have been developed for X-ray absorption experiments, such as a 19-element Ge detector array; a schematic illustration of a seven-element detector is shown in Figure 23; a 100-element detector array can be used for more efficient detection of X-rays.

Transition metals are usually a target of XAFS analysis, the energy range of which is from 5 to 20 keV. The beamline for X-rays of this energy range uses Be and polymer films as windows to separate the vacuum system from the atmosphere. Both lower and higher energy XAFS experiments require different techniques.

Soft X-ray XAFS experiments, ranging from 0.1 to 5 keV, require UHV techniques. This is because any windows between the SR storage ring and the sample heavily absorb X-rays, hence a windowless beamline is required. Consequently, the sample chamber is made of UHV components and must be baked out up to 200 °C. The contaminants in X-ray windows and other X-ray optics are carbon and oxygen, which are in the soft X-ray region (250–600 eV). Hence a windowless system is appropriate for the analysis of these elements. The shortcoming of the soft X-ray region experiment is the UHV system, because samples requiring X-ray absorption analysis could not always be inserted into the UHV system because they would emit gas into the clean system. A vacuum of 10^{-2} Pa is sufficient to avoid the absorption of X-rays in the X-ray path. The XAFS spectra of this soft X-ray region are surface sensitive and the profile changes of the spectra due to chemical state are very large. The spectra measured can be used for the same purpose (chemical state analysis) as XPS or ESCA. The detection limit of XAFS is usually lower than that of ESCA, hence XAFS has an advantage over ESCA if SR is available.

XAFS experiments higher than 20 keV require different experimental techniques. The number of application examples is not large, mainly because high-energy X-ray sources are limited without using third-generation SR facilities. In place of measuring this energy range K-edge XAFS, the L-edge XAFS, the energy of which usually falls in the range 5–20 keV, is measured. The L XAFS spectra are composed of L_1, L_2 and L_3 edge jumps, which interfere with each other. To avoid this, the K edge is useful for the analysis. The lifetime of the K hole state of high-energy K-edge XAFS of higher atomic number elements is, however, shorter than the long-wavelength region. Because of the Heisenberg uncertainty principle,

the line widths of the absorption spectral components are as large as 100 eV. Both the EXAFS oscillation and XANES lose fine structure owing to this lifetime broadening. High-energy XAFS has recently been measured with the development of third-generation SR facilities, because the numerical analysis overcomes the short lifetime effect.

7 SOURCES AND DATABASES

Academic societies and E-mail lists discuss the standardization of X-ray absorption spectrometry. As XAFS is used in many areas of research, standardization has been required. To achieve standardization, the International XAFS Society (IXS) was established in the 1990s. The purpose of the IXS is stated as follows:

> The International XAFS Society represents all those working on the fine structure associated with inner-shell excitation (near-edge and extended) by various probes (e.g. X-rays and electrons), and related techniques for which the data are interpreted on the same physical basis. The purpose of the IXS is to oversee activities which benefit the community as a whole, to establish operational committees, to provide for education in the field, to disseminate relevant information, to work with other related regional, national and professional organizations in promoting and developing XAS and related disciplines, and to act as representative for the community to other professional organizations.

This society has a close relation to the International Union of Crystallography (IUCr). The WWW home page of the IXS is at Illinois Institute of Technology, http://ixs.iit.edu/, where a large number of XAFS databases are presented. This WWW page links to other related WWW home pages. National society and working groups are active in many countries. All the information concerning these activities is obtainable at the international conferences on XAFS. The first XAFS international conference was held at Daresbury, UK, in 1981, and subsequent conferences are listed in Table 8.

Many kinds of activity reports published by SR facilities are useful sources of experimental methods and standard spectra. Journal sources are listed in Table 9. Fundamental reviews in *Analytical Chemistry* published in even years relate to X-ray absorption spectrometry.[113–120] In a book by Meisel et al.,[121] references are classified by atomic number and spectral series. Recently several books which treat newer X-ray techniques[122–126] and concerning X-ray absorption[127,128] have been published. The Materials Research Society has held symposia on applications of SR in materials science.[129–131]

The Denver X-ray analysis conference and international conferences on electron spectroscopy, on X-ray and inner-shell processes and on vacuum ultraviolet physics are sources of X-ray absorption spectroscopy and spectrometry.

Table 8 List of international XAFS conferences and proceedings

Conf. no.	Date	Location	Ref.
1st	March 1981	Daresbury	103
2nd	September 1982	Frascati	104
3rd	July 1984	Stanford	105
4th	July 1986	Fontevraud	106
5th	August 1988	Seattle	107
6th	August 1990	York	108
7th	August 1992	Kobe	109
8th	August 1994	Berlin	110
9th	August 1996	Grenoble	111
10th	August 1998	Chicago	112
11th	August 2000	Ako	–

Table 9 Source of information (journals)

Advances in X-ray Analysis
Analytical Chemistry, Fundamental Review, even years[113–120]
Journal of X-ray Science and Technology
Journal of Electron Spectroscopy and Related Phenomena
Journal of Synchrotron Radiation
Physical Review B
Surface Science
Synchrotron Radiation News
X-ray Spectrometry

8 ALTERNATIVE METHODS

8.1 Electron Energy Loss Spectroscopy

An electron beam, with an energy from a few hundred electron-volts to a few hundred kiloelectron-volts, impinges on a sample and loses its kinetic energy. When the sample is a bulk material, reflected electron energy is measured. Usually the loss of transmitted electron energy is measured for thin-film samples less than a few micrometers or a few tens of nanometers thick. This is called EELS.[132] EELS is usually combined with transmission electron microscopy (TEM). The electron energy loss structure is similar to the XAFS. The EXAFS region in EELS is called the extended electron energy loss fine structure (EXELFS). Forward-scattered (0°) electron energy loss spectra, formed when electrons are transmitted in a thin film, are approximately equivalent to the optical spectra; the selection rule is the electric dipole. Energy loss spectra of electrons scattered at a large angle are not treated by the electric dipole transition, and sometimes include optically forbidden transitions. The transmission method used in TEM has a very high spatial

resolution, hence chemical state imaging by the chemical shift of the absorption edge is possible.[133,134] High energy resolution and high spatial resolution are not always achieved by a single instrument. The EELS spectra are sensitive for low atomic number elements such as boron and carbon. It is not easy to measure the XAFS spectra of these long wavelengths using an SR facility.

8.2 Self-absorption

Although the characteristic X-ray wavelength of an element is usually separated from the absorption edge wavelength of the same element for hard X-rays, they are very close to each other for the soft X-ray region. These close lines are, for example, transition metal Lα,β X-ray emission lines and $L_{2,3}$ absorption edges. The Lα and Lβ X-ray emission lines emitted in a deep location in a solid are absorbed during the travel in the solid. Hence the X-ray emission spectra have dips due to the X-ray absorption spectra. The profiles of the Lα,β X-ray emission lines of transition metals excited by different electron energies (3 and 16 keV) change because of the self-absorption effect.[135] If the electron energy is high (16 keV), then the penetration depth of the electron is deeper. Hence the X-ray emission spectrum suffers heavily from the self-absorption effect. In contrast, if the electron energy is low (3 keV), then the penetration depth is shallow, and the X-ray emission spectrum is free from the self-absorption effect. The comparison of these two spectra yields an X-ray absorption. Similarly, one set of X-ray emission spectra is obtainable by tilting the sample to the X-ray detector or to the incident electron beam, when the electron energy is fixed.

8.3 Extended X-ray Emission Fine Structure

The radiative Auger effect (RAE) is always associated with the X-ray characteristic lines and this effect is an energy loss structure in characteristic X-ray emission, as shown in Figure 37.[136] The second electron shaken up into an unoccupied orbital has similar information to the XAFS. This is called EXEFS.[137] This method is used to measure low atomic number elements such as Na, Mg, Al, and Si, because the RAE satellite intensity is strong for these elements. If wavelength-dispersive electron probe microanalysis (EPMA) is available, XAFS spectra of 1 μm diameter area are measurable using this method.

8.4 X-ray Raman Scattering

X-ray Raman scattering is the effect of energy loss on X-ray scattering. Raman scattering is a similar physical process to Compton scattering. The difference is that Raman scattering involves scattering by core electrons whereas Compton scattering involves scattering by conduction band electrons. X-ray Raman scattering is a method for measuring soft X-ray absorption spectra (say of carbon) with a hard X-ray spectrometer (a few kiloelectron-volts). Hard X-rays can be measured in air; soft X-ray absorption spectroscopy, which usually requires UHV, is possible in air by this method. Hard X-rays (8265 eV) impinge on a carbon-containing sample, and the X-rays lose energy by the carbon K edge due to the Raman scattering (ca. 300 eV), as shown in Figure 38.[138]

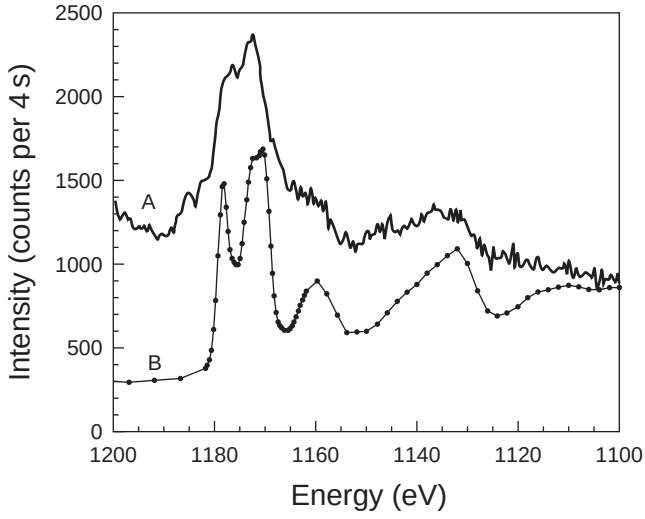

Figure 37 (A) Low-energy satellites (the radiative Auger satellites) of the Kα X-ray fluorescence spectrum of MgO compared with (B) XANES measured at an SR facility. (Reproduced by permission from Kawai and Takahashi.[136])

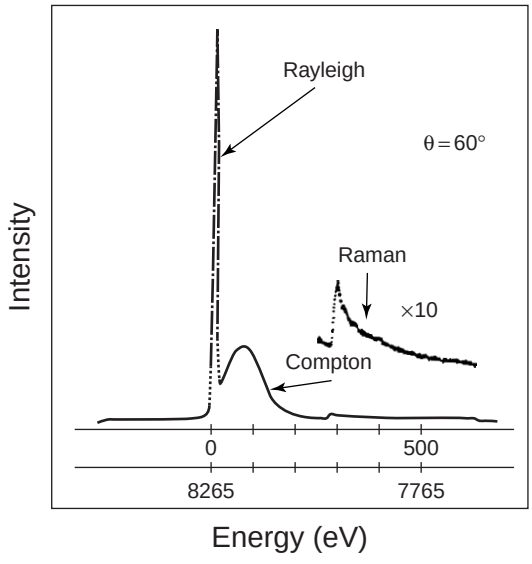

Figure 38 Rayleigh, Compton, and Raman scattering spectra of 8265-eV incident X-rays by graphite observed at $\theta = 60°$. (Reproduced by permission from Tohji and Udagawa.[138] © 1987 The American Physical Society.)

The X-ray that should be detected is at ca. 8 keV, which still falls in the hard X-ray region. The XAFS study of catalysts during reaction with gases is possible using the Raman effect.

8.5 Diffraction Anomalous Fine Structure

DAFS was proposed by Stragier et al.[139] XAFS usually measures the wavelength dependence of f_2, the imaginary part of the atomic structure factor; DAFS measures f_1. The wavelength dependence of f_1 and f_2 has a close relation through the Kramers–Kronig transformation. In the DAFS experiment, the intensity of a diffraction peak of a specimen is measured by the change in the incident X-ray energy. The sample and detector angles $(\theta - 2\theta)$ are measured by the change in the incident X-ray energy, or powder X-ray diffraction patterns are measured by the change in incident X-ray energy. This method can measure site-selective XAFS of the same element, because the diffraction peak corresponds to a different site in the crystal. If the diffraction peaks which originate from the surface and bulk phases are separated, space-selective EXAFS-like spectra are obtainable by this method.

8.6 β-Environment Fine Structure

The β-electron emission process in a nuclear conversion process suffers interference by the crystal structure for the same reason as EXAFS. This method is called BEFS.[140]

8.7 Inverse Photoemission Spectroscopy

IPES is an alternative method to measure the unoccupied electronic structure by irradiating electrons and detecting photons.[141] This method is otherwise called bremsstrahlung isochromat spectroscopy (BIS). The extended structure like EXAFS is also observable in BIS and this is called the extended X-ray bremsstrahlung isochromat fine structure (EXBIFS).[142] The BIS is usually combined with an ESCA instrument, and thus occupied and unoccupied electronic structures (similar to XANES) are measurable.[143]

9 CONCLUSION

X-ray absorption spectroscopy is chiefly used in the area of electronic structure study and structural analysis for the study of new materials, surfaces, and catalysts. The spectra measured are surface sensitive or bulk sensitive depending on the detection method. Chemical shift and profile changes are observable. Thus the spectral analysis is useful for materials characterization. This method is also powerful for analyzing mixed chemical states in industrial, environmental, and biological analytes. The development of SR facilities will make it possible to measure nanometer-sized samples in less than a few milliseconds.

ABBREVIATIONS AND ACRONYMS

APW	Augmented Plane Wave
BEFS	β-Environment Fine Structure
BIS	Bremsstrahlung Isochromat Spectroscopy
CEY	Conversion Electron Yield
DAFS	Diffraction Anomalous Fine Structure
EELS	Electron Energy Loss Spectroscopy
EPMA	Electron Probe Microanalysis
ESCA	Electron Spectroscopy for Chemical Analysis
EXAFS	Extended X-ray Absorption Fine Structure
EXBIFS	Extended X-ray Bremsstrahlung Isochromat Fine Structure
EXEFS	Extended X-ray Emission Fine Structure
EXELFS	Extended Electron Energy Loss Fine Structure
HF	Hartree–Fock
HFS	Hartree–Fock–Slater
IMFP	Inelastic Mean Free Path
IPES	Inverse Photoemission Spectroscopy
IUCr	International Union of Crystallography
IXS	International XAFS Society
KKR	Korringa–Kohn–Rostker
LCAO-MO	Linear Combination of Atomic Orbitals-Molecular Orbital
LDA	Local Density Approximation
LEED	Low-energy Electron Diffraction
LRO	Long-range Order
LUMO	Lowest Unoccupied Molecular Orbital
MS	Multiple Scattering
MS-Xα	Multiple Scattering Xα
MT	Muffin-tin
NEXAFS	Near-edge X-ray Absorption Fine Structure
PA	Photoacoustic
PDA	Photodiode Array
PEH	Photoelectron Holography
Q-XAFS	Quick X-ray Absorption Fine Structure
RAE	Radiative Auger Effect
SEXAFS	Surface Extended X-ray Absorption Fine Structure
SR	Synchrotron Radiation

SRO	Short-range Order
SSD	Solid-state Detector
TEM	Transmission Electron Microscopy
TEY	Total Electron Yield
UHV	Ultrahigh Vacuum
XAFS	X-ray Absorption Fine Structure
XANES	X-ray Absorption Near-edge Structure
XAS	X-ray Absorption Spectrum
XFH	X-ray Fluorescence Holography
XFY	X-ray Fluorescence Yield
XMCD	X-ray Magnetic Circular Dichroism
XPD	X-ray Photoelectron Diffraction
XPS	X-ray Photoelectron Spectroscopy

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Surfaces **(Volume 10)**
Auger Electron Spectroscopy in Analysis of Surfaces • Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces • X-ray Photoelectron Spectroscopy in Analysis of Surfaces

Electroanalytical Methods **(Volume 11)**
X-ray Methods for the Study of Electrode Interaction

X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy **(Volume 15)**
X-ray Photoelectron and Auger Electron Spectroscopy • X-ray Photoelectron Spectroscopy and Auger Electron Spectroscopy: Introduction

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Structure Determination, X-ray Diffraction for • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis • X-ray Techniques: Overview

REFERENCES

1. R. Jenkins, 'X-ray Techniques – Overview', *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Chichester.
2. M. Siegbahn, *The Spectroscopy of X-rays*, Translated by G.A. Lindsay, Oxford University Press, London, 1925.
3. A.H. Compton, S.K. Allison, *X-rays in Theory and Experiment*, Van Nostrand, Princeton, NJ, 1935.
4. A.E. Sandström, 'Experimental Methods of X-ray Spectroscopy: Ordinary Wavelengths', in *Handbuch der Physik, Vol. 30, Röntgenstrahlen*, ed. S. Flügge, Springer, Berlin, 78–245, 1957.
5. R.W. James, *The Optical Principles of the Diffraction of X-rays*, Ox Bow Press, Woodbridge, 1982.
6. B.K. Agarwal, *X-ray Spectroscopy*, 2nd edition, Springer, Berlin, 1991.
7. A.A. Markowicz, 'X-ray Physics', in *Handbook of X-ray Spectrometry*, eds. R.E. Van Grieken, A.A. Markowicz, Marcel Dekker, New York, 1–73, 1993.
8. B.L. Henke, E.M. Gullikson, J.C. Davis, 'X-ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at $E = 50–30\,000$ eV, $Z = 1–92$', *At. Data Nucl. Data Tables*, **54**, 181–342 (1993).
9. Y. Udagawa (ed.), *X-ray Absorption Fine Structure*, Gakkai-Shuppan-Center, Tokyo, 1993 (in Japanese).
10. C.H. Macgillavry, G.D. Rieck, *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham, Vol. III, 1968.
11. B.L. Henke, P. Lee, T.J. Tanaka, R.L. Shimabukuro, B.K. Fujikawa, 'Low Energy X-ray Interaction Coefficients: Photoabsorption, Scattering, and Reflection. $E = 100–2000$ eV, $Z = 1–94$', *At. Data Nucl. Data Tables*, **27**, 1–144 (1982).
12. B.L. Henke, J.C. Davis, E.M. Gullikson, R.C.C. Perera, 'A Preliminary Report on X-ray Photoabsorption Coefficients and Atomic Scattering Factors for 92 Elements in the 10–10 000 eV Region', LBL-26259, UC-411, Lawrence Berkeley Laboratory, University of California, Berkeley, CA, 1988.
13. J. Kirz, D.T. Attwood, B.L. Henke, M.R. Howells, K.D. Kennedy, K.-J. Kim, J.B. Kortright, R.C. Perera, P. Pianetta, J.C. Riordan, J.H. Scofield, G.L. Stradling, A.C. Thompson, J.H. Underwood, D. Vaughan, G.P. Williams, H. Winick, 'X-ray Data Booklet', PUB-490 Rev., Lawrence Berkeley Laboratory, University of California, Berkeley, CA, 1986 (this reference has many printing errors).
14. J.A. Bearden, 'X-ray Wavelengths', *Rev. Mod. Phys.*, **39**, 78–124 (1967).
15. E.W. White, G.G. Johnson, Jr, *X-ray Emission and Absorption Wavelengths and Two-theta Tables*, ASTM Data Series DS 37A, American Society for Testing and Materials, Philadelphia, PA, 1970.
16. K. Siegbahn, C. Nordling, A. Fahlman, R. Nordberg, K. Hamrin, J. Hedman, G. Johansson, T. Bergmark, S.-E. Karlsson, L. Lindgren, B. Lindberg, 'ESCA – Atomic, Molecular and Solid State Structure Studied by Means of Electron Spectroscopy', *Nova Acta Regiae Soc. Sci. Ups., Ser. IV*, **20**, 1–282 (1967). An extension of this reference has been published as K. Siegbahn, C. Nordling, G. Johansson, J. Hedman, P.F. Heden, K. Hamlin, U. Gelius, T. Bergmark, L.O. Werme,

R. Manne, Y. Baer, *ESCA Applied to Free Molecules*, North-Holland, Amsterdam, 1969.
17. H. Siegbahn, L. Karlsson, 'Photoelectron Spectroscopy', in *Encyclopedia of Physics, Vol. 31, Corpuscles and Radiation in Matter I*, ed. W. Mehlhorn, Springer, Berlin, 215–467, 1982.
18. J.H. Scofield, 'Theoretical Photoionization Cross-sections from 1 to 1500 keV', UCRL-51326, Lawrence Livermore Laboratory, University of California, Livermore, CA, 1973.
19. J.J. Yeh, I. Lindau, 'Atomic Subshell Photoionization Cross-sections and Asymmetry Parameters: $1 \leq Z \leq 103$', *At. Data Nucl. Data Tables*, **32**, 1–155 (1985).
20. J.A. Bearden, A.F. Burr, 'Reevaluation of X-ray Atomic Energy Levels', *Rev. Mod. Phys.*, **39**, 125–142 (1967).
21. J.F. Moulder, W.F. Stickle, P.E. Sobol, K.D. Bomben, *Handbook of X-ray Photoelectron Spectroscopy*, Perkin-Elmer, Eden Prairie, MN, 1992.
22. K.I. Pandya, D.C. Koningsberger, 'X-ray Absorption Fine Structure Spectroscopy', in *Analysis of Microelectronic Materials and Devices*, eds. M. Grasserbauer, H.W. Werner, Wiley, Chichester, 637–655, 1991.
23. B.K. Teo, D.C. Joy (eds.), *EAXFS Spectroscopy*, Plenum Press, New York, 1981.
24. P.A. Lee, P.H. Citrin, P. Eisenberger, B.M. Kincaid, 'Extended X-ray Absorption Fine Structure – its Strengths and Limitations as a Structural Tool', *Rev. Mod. Phys.*, **53**, 769–806 (1981).
25. D.C. Koningsberger, R. Prins (eds.), *X-ray Absorption, Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*, Wiley, New York, 1988.
26. T. Ishii, *Principles of the Theory of EXAFS*, Shokabo, Tokyo, 1994 (in Japanese).
27. J. Stöhr, *NEXAFS Spectroscopy*, Springer, Berlin, 1992.
28. F.W. Lytle, D.E. Sayers, E.A. Stern, 'The History and Modern Practice of EXAFS Spectroscopy', in *Advances in X-ray Spectroscopy*, eds. C. Bonnelle, C. Mande, Pergamon Press, Oxford, 267–286, 1982.
29. T. Shiraiwa, 'Progress of the Theory of Fine Structure of the X-ray Absorption Spectrum – from Kronig to Stern', *Oyo-Buturi*, **56**, 851–859 (1987) (in Japanese).
30. E.A. Stern, 'Development of XAFS into a Structure Determination Technique', in *Röntgen Centennial, X-rays in Natural and Life Sciences*, eds. A. Haas, G. Landwehr, E. Umbach, World Scientific, Singapore, 323–340, 1997.
31. F. Lytle, 'The EXAFS Family Tree: a Personal History of the Development of Extended X-ray Absorption Fine Structure', *J. Synchrotron Radiat.*, **6**, 123–134 (1999).
32. H. Fricke, 'The K-characteristic Absorption Frequencies for the Chemical Elements Magnesium to Chromium', *Phys. Rev.*, **16**, 202–215 (1920).
33. W. Kossel, 'Zum Bau der Röntgenspektren', *Z. Phys.*, **1**, 119–134 (1920).
34. R. de, L. Kronig, 'Zur Theorie der Feinstruktur in den Röntgenabsorptionsspektren', *Z. Phys.*, **70**, 317–323 (1931).
35. T. Hayasi, 'Zur Theorie der Feinstruktur des Röntgen K-Absorptionsspektrums II. Der Einfluss der thermischen-Schwingungen des Kristallgitters auf die Feinstrukture', *Sci. Rep. Tohoku Univ., Ser. 1*, **33**, 183–194 (1949).
36. T. Shiraiwa, T. Ishimura, M. Sawada, 'The Theory of the Fine Structure of the X-ray Absorption Spectrum', *J. Phys. Soc. Jpn.*, **13**, 847–859 (1958).
37. A.I. Kozlenkov, 'Teoriya Tonkoi Strukturi Rentogenovskich Spektrov Poglosheniya', *Izv. Akad. Nauk SSSR*, **25**, 957–976 (1961).
38. D.E. Sayers, E.A. Stern, F.W. Litle, 'New Technique for Investigating Noncrystalline Structures: Fourier Analysis of the Extended X-ray-absorption Fine Structure', *Phys. Rev. Lett.*, **27**, 1204–1207 (1971).
39. C. Mande, V.B. Sapre, 'Chemical Shifts in X-ray Absorption Spectra', in *Advances in X-ray Spectroscopy*, eds. C. Bonnelle, C. Mande, Pergamon Press, Oxford, 287–301, 1982.
40. H. Sekiyama, N. Kosugi, H. Kuroda, T. Ohta, 'Sulfur K-edge Absorption Spectra of Na_2SO_4, Na_2SO_3, $Na_2S_2O_3$, and $Na_2S_2O_x$ $(x = 5–8)$', *Bull. Chem. Soc. Jpn.*, **59**, 575–579 (1986).
41. J. Kawai, Y. Kitajima, Z. Asaki, 'Chemical State of Sulfur in Calcine of Zinc Fluidized Bed Roaster', *Adv. X-ray Chem. Anal. Jpn.*, **27**, 235–244 (1996) (in Japanese).
42. J. Wong, G.N. George, I.J. Pickering, Z.U. Rek, M. Rowen, T. Tanaka, G.H. Via, B. De Vries, D.E.W. Vaughan, G.E. Brown, Jr, 'New Opportunities in XAFS Investigation in the 1–2 keV Region', *Solid State Commun.*, **92**, 559–562 (1994).
43. F.A. Cotton, G. Wilkinson, *Advanced Inorganic Chemistry*, 2nd edition, Wiley, New York, 103, 1972.
44. D.A. Outka, J. Stöhr, 'Curve Fitting Analysis of Near-edge Core Excitation Spectra of Free, Adsorbed, and Polymeric Molecules', *J. Chem. Phys.*, **88**, 3539–3554 (1988).
45. T. Mukoyama, unpublished work.
46. P.W. Atkins, *Physical Chemistry*, 6th edition, Oxford University Press, Oxford, 1998.
47. M. Sano, 'XANES Study at the Co K Absorption Edge in a Series of Cobalt(III) Complexes', *Inorg. Chem.*, **27**, 4249–4253 (1988).
48. A.L. Roe, D.J. Schneider, R.J. Mayer, J.W. Pyrz, J. Widom, L. Que, Jr, 'X-ray Absorption Spectroscopy of Iron–Tyrosinate Proteins', *J. Am. Chem. Soc.*, **106**, 1676–1681 (1984).
49. J.G. Bednorz, K.A. Müller, 'Possible High T_c Superconductivity in the Ba–La–Cu–O System', *Z. Phys. B: Condens. Matter*, **64**, 189–193 (1986).
50. J. Zaanen, G.A. Sawatzky, J.W. Allen, 'The Electronic Structure and Band Gaps in Transition Metal Compounds', *J. Magn. Magn. Mater.*, 54–57, 607–611 (1986).

51. F.M.F. de Groot, 'X-ray Absorption of Transition Metal Oxides: an Overview of the Theoretical Approaches', *J. Electron Spectrosc. Relat. Phenom.*, **62**, 111–130 (1993).
52. G. Wortmann, I. Nowik, B. Perscheid, G. Kaindl, I. Felner, 'Critical Evaluation of Eu Valencies from L_{III}-edge X-ray-absorption and Mössbauer Spectroscopy of $EuNi_2Si_{2-x}Ge_x$', *Phys. Rev. B*, **43**, 5261–5268 (1991).
53. B.T. Thole, G. van der Laan, J.C. Fuggle, G.A. Sawatzky, R.C. Karnatak, J.-M. Estava, '3d X-ray-absorption Lines and the $3d^9 4f^{n+1}$ Multiplets of the Lanthanides', *Phys. Rev. B*, **32**, 5107–5118 (1985).
54. L.G. Parratt, 'Electronic Band Structure of Solids by X-ray Spectroscopy', in *Solid State Physics*, eds. K. Lark-Horovitz, V.A. Johnson, Academic Press, New York, Part B, 281–292, Vol. 6, 1959.
55. J.M. Ziman, *Principles of the Theory of Solids*, 2nd edition, Cambridge University Press, Cambridge, Chapter 3, 1972.
56. K.H. Johnson, 'Scattered-wave Theory of the Chemical Bond', *Adv. Quantum Chem.*, **7**, 143–185 (1973).
57. J.C. Slater, *The Calculation of Molecular Orbitals*, Wiley, New York, 1979.
58. D.D. Vvedensky, D.K. Saldin, J.B. Pendry, 'An Update of DLXANES, the Calculation of X-ray Absorption Near-edge Structure', *Comput. Phys. Commun.*, **40**, 421–440 (1986).
59. I. Tanaka, T. Nakajima, J. Kawai, H. Adachi, H. Gu, M. Ruhle, 'Dopant-modified Local Chemical Bonding at a Grain Boundary in $SrTiO_3$', *Philos. Mag. Lett.*, **75**, 21–27 (1997).
60. F.M.F. de Groot, J.C. Fuggle, B.T. Thole, G.A. Sawatzky, '2p X-ray Absorption of 3d Transition-metal Compounds: an Atomic Multiplet Description Including the Crystal Field', *Phys. Rev. B*, **42**, 5459–5468 (1990).
61. T. Tiedje, J.R. Dahn, Y. Gao, K.M. Colbow, E.D. Crozier, D.T. Jiang, W. Eberhardt, 'Bragg Reflection Model for X-ray Absorption Near Edge Structure in Crystalline Solids', *Solid State Commun.*, **85**, 161–165 (1993).
62. J. Kawai, K. Tamura, M. Owari, Y. Nihei, 'X-ray Photoelectron Diffraction of $SrTiO_3$', *J. Electron Spectrosc. Relat. Phenom.*, **61**, 103–122 (1992).
63. M. Sagurton, E.L. Bullock, C.S. Fadley, 'The Analysis of Photoelectron Diffraction Data Obtained with Fixed Geometry and Scanned Photon Energy', *Surf. Sci.*, **182**, 287–361 (1987).
64. M.J. Fay, A. Proctor, D.P. Hoffmann, D.M. Hercules, 'Unraveling EXAFS Spectroscopy', *Anal. Chem.*, **60**, 1225A–1243A (1988).
65. M. Nomura, 'Trace Element Analysis Using X-ray Absorption Edge Spectrometry', *Anal. Chem.*, **64**, 2711–2714 (1992).
66. A. Kuzmin, 'Anisotropy of the Photoelectron Mean Free Path and its Influence on the EXAFS Amplitude', *J. Phys.: Condens. Matter*, **6**, 5761–5770 (1994).
67. P.M. Len, C.S. Fadley, G. Materlik, 'Atomic Holography with Electrons and X-rays', in *X-ray and Inner-shell Processes, 17th International Conference*, eds. R.L. Johnson, H. Schmidt-Bocking, B.F. Sonntag, AIP Conference Proceeding 389, American Institute of Physics, Woodbury, NY, 295–319, 1997.
68. R.F. Egerton, *Electron Energy-loss Spectroscopy in the Electron Microscope*, Plenum Press, New York, 1986.
69. Y. Kitajima, 'Fluorescence Yield X-ray Absorption Fine Structure Measurements in the Soft X-ray Region', *Rev. Sci. Instrum.*, **66**, 1413–1415 (1995).
70. J. Stöhr, 'Geometry and Bond Lengths of Chemisorbed Atoms and Molecules: NEXAFS and SEXAFS', *Z. Phys. B: Condens. Matter*, **61**, 439–445 (1985).
71. S. Yagi, T. Yokoyama, Y. Kitajima, Y. Takata, T. Kanazawa, A. Imanishi, T. Ohta, 'Structural and Electronic Properties of Molecularly Adsorbed CS_2 on Cu(111) Studied by X-ray Absorption and Photoelectron Spectroscopies', *Surf. Sci.*, **311**, 172–180 (1994).
72. T. Ohta, Y. Kitajima, H. Kuroda, T. Takahashi, S. Kikuta, 'A Possible Use of the Soft X-ray Standing Wave Method for Surface and Interface Structure Analysis', *Nucl. Instrum. Methods Phys. Res. A*, **246**, 760–762 (1986).
73. R.D. Deslattes, R.E. LaVilla, P.L. Cowan, A. Henins, 'Threshold Studies of a Multivacancy Process in the Kβ Region of Argon', *Phys. Rev. A*, **27**, 923–933 (1983).
74. Rigaku, *R-EXAFS Series, Rigaku EXAFS and XANES Frame System Catalogue*, CJD427A/960820NS, Rigaku, Tokyo, 1996.
75. H. Maeda, H. Terauchi, K. Tanabe, N. Kamijo, M. Hida, H. Kawamura, 'Development of a Laboratory EXAFS Facility and its Application to Amorphous GeSe Semiconductors', *Jpn. J. Appl. Phys.*, **21**, 1342–1346 (1982).
76. K. Sakurai, 'High-intensity X-ray Line Focal Spot for Laboratory Extended X-ray Absorption Fine-structure Experiments', *Rev. Sci. Instrum.*, **64**, 267–268 (1993).
77. H. Saisho, Y. Gohshi (eds.), *Applications of Synchrotron Radiation to Materials Analysis*, Elsevier, Amsterdam, 1996.
78. H. Oyanagi, R. Shioda, Y. Kuwahara, K. Haga, 'Surface-sensitive XAFS in the Hard X-ray Region with Submonolayer Sensitivity', *J. Synchrotron Radiat.*, **2**, 99–105 (1995).
79. J. Kawai, H. Adachi, S. Hayakawa, S.Y. Zhen, K. Kobayashi, Y. Gohshi, K. Maeda, Y. Kitajima, 'Depth Selective X-ray Absorption Fine Structure Spectroscopy', *Spectrochim. Acta Part B*, **49**, 739–743 (1994).
80. J. Stöhr, C. Noguera, T. Kendelewicz, 'Auger and Photoelectron Contribution to the Electron-yield Surface Extended X-ray-absorption Fine-structure Signal', *Phys. Rev. B*, **30**, 5571–5579 (1984).
81. J. Stöhr, R. Jaeger, S. Brennan, 'Surface Crystallography by Means of Electron and Ion Yield SEXAFS', *Surf. Sci.*, **117**, 503–524 (1982).

82. S. Zheng, Y. Gohshi, 'A Simulation Study of Signal to Background Ratio of XANES by Total Electron Yield at Grazing Angle', *Anal. Sci.*, **13**, 997–1001 (1997).
83. Y. Kitajima, 'Soft X-ray Absorption Spectroscopy with Variable Surface Sensitivity Using Fluorescence Yield Detection', *J. Phys. IV*, **7**(C2), 705–706 (1997).
84. L. Tröger, D. Arvanitis, K. Baberschke, H. Michaelis, U. Grimm, E. Zchech, 'Full Correction of the Self-absorption in Soft-fluorescence Extended X-ray-absorption Fine Structure', *Phys. Rev. B*, **46**, 3283–3289 (1992).
85. S. Emura, T. Moriga, J. Takizawa, M. Nomura, K.R. Bauchspiess, T. Murata, K. Harada, H. Maeda, 'Optical-luminescence Yield Spectra Produced by X-ray Excitation', *Phys. Rev. B*, **47**, 6918–6930 (1993).
86. M.E. Garcia, Ch. Brouder, K.H. Bennemann, 'Theory for the Photoacoustic Response to Circularly Polarized X-rays: How to Detect Magnetism Using a Microphone', *Solid State Commun.*, **103**, 331–335 (1997).
87. T.K. Sham, R.A. Holroyd, 'Photoconductivity Measurements of X-ray Absorption of Liquids: Fe K-edge Spectrum of Ferrocene in 2,2,4-Trimethylpentane', *J. Chem. Phys.*, **80**, 1026–1029 (1983).
88. T.K. Sham, R.A. Holroyd, 'Photoconductivity Measurements of X-ray-absorption Spectra of Tetramethyl Tin in 2,2,4-Trimethylpentane Near the Sn K Edge', *Phys. Rev. B*, **39**, 8257–8262 (1989).
89. I. Nakai, C. Numako, S. Hayakawa, A. Tsuchiyama, 'Chemical Speciation of Geological Samples by Micro-XANES Techniques', *J. Trace Microprobe Tech.*, **16**, 87–98 (1998).
90. G.E. Derbyshire, W.I. Helsby, A.J. Dent, S.A. Wright, R.C. Farrow, G.N. Greaves, C. Morrell, G.J. Baker, 'Current and Future Energy Dispersive EXAFS Detector Systems', *Adv. X-Ray Anal.*, **34**, 177–185 (1990).
91. T. Matsushita, H. Oyanagi, S. Saigo, U. Kaminaga, H. Hashimoto, H. Kihara, N. Yoshida, M. Fujimoto, 'Twenty-five Millisecond Resolution Time-resolved X-ray Absorption Spectroscopy in Dispersive Mode', *Jpn. J. Appl. Phys.*, **25**, L523–L525 (1986).
92. M. Suzuki, N. Kawamura, M. Mizumaki, A. Urata, H. Maruyama, S. Goto, T. Ishikawa, 'Helicity-modulation Technique Using Diffractive Phase Retarder for Measurements of X-ray Magnetic Circular Dichroism', *Jpn. J. Appl. Phys.*, **37**, L1488–L1490 (1988).
93. L.-C. Duda, J. Stohr, D.C. Mancini, A. Nilsson, N. Wassdahl, J. Nordgren, M.G. Samant, 'Magnetic Dichroism in $L_{2,3}$ Emission of Fe, Co, and Ni Following Energy-dependent Excitation with Circularly Polarized X-rays', *Phys. Rev. B*, **50**, 16 758–16 761 (1994).
94. G. Schutz, P. Fischer, K. Attenkofer, D. Ahlers, 'X-ray Magnetic Circular Dichroism', in *Röntgen Centennial, X-rays in Natural and Life Sciences*, eds. A. Haas, G. Landwehr, E. Umbach, World Scientific, Singapore, 341–363, 1997.
95. L.H. Bennet, R.E. Watson (eds.), *Magnetic Multilayers*, World Scientific, Singapore, 1994.
96. S.W. Lovesey, S.P. Collins, *X-ray Scattering and Absorption by Magnetic Materials*, Oxford, University Press, New York, 1996.
97. H. Ebert, G. Schutz (eds.), *Spin–Orbit-influenced Spectroscopies of Magnetic Solids*, Springer, Berlin, 1996.
98. M.M. Grush, J. Chen, T.L. Stemmler, S.J. George, C.Y. Ralston, R.T. Stibrany, A. Gelasco, G. Christou, S.M. Gorun, J.E. Penner-Hahn, S.P. Cramer, 'Manganese L-edge X-ray Absorption Spectroscopy of Manganese Catarase from *Lactobacillus plantarum* and Mixed Valence Manganese Complexes', *J. Am. Chem. Soc.*, **118**, 65–69 (1996).
99. M. Abbate, F.M.F. de Groot, J.C. Fuggle, Y.J. Ma, C.T. Chen, F. Sette, A. Fujimori, Y. Ueda, K. Kosuge, 'Soft-X-ray-absorption Studies of the Electronic-structure Changes Through the VO_2 Phase Transition', *Phys. Rev. B*, **43**, 7263–7266 (1991).
100. I. Watanabe, H. Tanida, S. Kawauchi, 'Coordination Structure of Zinc(II) Ions on a Langmuir Monolayer, Observed by Total-reflection X-ray Absorption Fine Structure', *J. Am. Chem. Soc.*, **119**, 12 018–12 019 (1997).
101. J. Kawai, S. Hayakawa, F. Esaka, S. Zheng, Y. Kitajima, K. Maeda, H. Adachi, Y. Gohshi, K. Furuya, 'Depth-selective Chemical State Analysis of Fine Particles Using X-ray Absorption', *Anal. Chem.*, **34**, 1526–1529 (1995).
102. K.K. Bando, K. Asakura, H. Arakawa, K. Isobe, Y. Iwasawa, 'Surface Structure and Catalytic Hydroformylation Activities of Rh Dimers Attached on Various Inorganic Oxide Supports', *J. Phys. Chem.*, **100**, 13 636–13 645 (1996).
103. 'Proceedings of the First EXAFS Conference', Report DL/SCI/R17, Daresbury Laboratory, Warrington, 1981.
104. A. Bianconi, L. Incoccia, S. Stipcich (eds.), *EXAFS and Near Edge Structure*, Springer, Berlin, 1983.
105. K.O. Hodgson, B. Hedman, J.E. Penner-Hahn (eds.), *EXAFS and Near Edge Structure III*, Springer, Berlin, 1984.
106. P. Lagarde, D. Raoux, J. Petiau (eds.), 'EXAFS and Near Edge Structure IV', *J. Phys.*, **47**(Colloq.C8) (1986).
107. J.M. deLeón, E.A. Stern, D.E. Sayers, Y. Ma, J.J. Rehr (eds.), 'XAFS V', *Phys. B*, **158**, (1989).
108. S.S. Hasnain (ed.), *X-ray Absorption Fine Structure*, Ellis Horwood, London, 1991.
109. H. Kuroda, T. Ohta, T. Murata, Y. Udagawa, M. Nomura (eds.), 'Proceedings of 7th International Conference on X-ray Absorption Fine Structure, Kobe, Japan, 1992', *Jpn. J. Appl. Phys.*, **32**(Suppl 32-2) (1993).
110. K. Baberschke, D. Arvanitis (eds.), 'Proceedings of 8th International Conference on X-ray Absorption Fine Structure, Freie Universität Berlin, Germany, 28 August–2 September 1994', *Phys. B*, 208–209 (1995).
111. J. Goulon, C. Goulon-Ginet, N.B. Brooks (eds.), 'Proceedings of 9th International Conference on X-ray Absorption Fine Structure, Grenoble, France, 26–30 August 1996', *J. Phys. IV*, **7**(Colloq. C2) (1997).

112. S.S. Hasnain, J.R. Helliwell, H. Kamitsubo, 'Proceedings of 10th International Conference on X-ray Absorption Fine Structure, Chicago, USA, 10–14 August 1998', *J. Synchrotron Radiat.*, **6**(Part 3), 121–807 (1999).
113. W.J. Campbell, 'X-ray Absorption and Emission', *Anal. Chem.*, **36**, 312R–328R (1964).
114. L.S. Birks, 'X-ray Absorption and Emission', *Anal. Chem.*, **44**, 557R–562R (1972).
115. L.S. Birks, J.V. Gilfrich, 'X-ray Absorption and Emission', *Anal. Chem.*, **46**, 360R–366R (1972).
116. G.L. Macdonald, 'X-ray Spectrometry', *Anal. Chem.*, **52**, 100R–106R (1980).
117. A.A. Markowicz, R.E. Van Grieken, 'X-ray Spectrometry', *Anal. Chem.*, **56**, 241R–250R (1984).
118. S.B. Török, R.E. Van Grieken, 'X-ray Spectrometry', *Anal. Chem.*, **66**, 186R–206R (1994).
119. S.B. Török, J. Lábár, J. Injuk, R.E. Van Grieken, 'X-ray Spectrometry', *Anal. Chem.*, **68**, 467R–485R (1996).
120. S.B. Török, J. Lábár, M. Schmeling, R.E. Van Grieken, 'X-ray Spectrometry', *Anal. Chem.*, **70**, 495R–517R (1998).
121. A. Meisel, G. Leonhardt, R. Szargan, *X-ray Spectra and Chemical Binding*, Springer, Berlin, 1989.
122. J. Thieme, G. Schmahl, D. Rudolph, E. Umbach (eds.), *X-ray Microscopy and Spectromicroscopy*, Springer, Berlin, 1998.
123. A.G. Michette, C.J. Buckley, *X-ray Science and Technology*, Institute of Physics, Bristol, 1993.
124. K.L. D'Amico, L.J. Terminello, D.K. Shuh (eds.), *Synchrotron Radiation Techniques in Industrial, Chemical, and Materials Science*, Plenum Press, New York, 1996.
125. W. Eberhardt (ed.), *Applications of Synchrotron Radiation*, Springer, Berlin, 1995.
126. A.S. Schlachter, F.J. Wuilleumier, *New Directions in Research with Third-generation Soft X-ray Synchrotron Radiation Sources*, NATO ASI Series E, Kluwer, Dordrecht, Vol. 254, 1994.
127. K.B. Garg, E.A. Stern, D. Norman (eds.), *X-ray Absorption in Bulk and Surfaces*, World Scientific, Singapore, 1994.
128. Y. Iwasawa, *X-ray Absorption Fine Structure for Catalysts and Surfaces*, World Scientific, Singapore, 1996.
129. D.L. Perry, N.D. Shinn, R.L. Stockbauer, K.L. D'Amico, L.J. Terminello (eds.), *Applications of Synchrotron Radiation Techniques to Materials Science*, Materials Research Society Symposium Proceedings, Materials Research Society, Pittsburgh, PA, Vol. 307, 1993.
130. L.J. Terminello, N.D. Shinn, G.E. Ice, K.L. D'Amico, D.L. Perry (eds.), *Applications of Synchrotron Radiation Techniques to Materials Science II*, Materials Research Society Symposium Proceedings, Materials Research Society, Pittsburgh, PA, Vol. 375, 1995.
131. L.J. Terminello, S.M. Mini, H. Ade, D.L. Perry (eds.), *Applications of Synchrotron Radiation Techniques to Materials Science III*, Materials Research Society Symposium Proceedings, Materials Research Society, Pittsburgh, PA, Vol. 437, 1996.
132. J.P. Eberhart, *Structural and Chemical Analysis of Materials*, Wiley, Chichester, 350–371, 1991.
133. D.A. Muller, Y. Tzou, R. Raj, J. Silcox, 'Mapping sp^2 and sp^3 States of Carbon at Sub-nanometre Spatial Resolution', *Nature (London)*, **366**, 725–727 (1993).
134. P.E. Batson, 'Simultaneous STEM Imaging and Electron Energy-loss Spectroscopy with Atomic-column Sensitivity', *Nature (London)*, **366**, 727–728 (1993).
135. D. Chopra, 'Ni L Self-absorption Spectrum', *Phys. Rev. A*, **1**, 230–235 (1970).
136. J. Kawai, H. Takahashi, 'Easy, Quick, and Small Area Measurement of X-ray Absorption Spectra Using an Electron Probe X-ray Microanalyzer', *Bunseki Kagaku*, **48**, 793–795 (1999).
137. J. Kawai, K. Hayashi, Y. Awakura, 'Extended X-ray Absorption Fine Structure (EXAFS) in X-ray Fluorescence Spectra', *J. Phys. Soc. Jpn.*, **66**, 3337–3340 (1997).
138. K. Tohji, Y. Udagawa, 'Novel Approach for Structure Analysis by X-ray Raman Scattering', *Phys. Rev. B*, **36**, 9410–9412 (1987).
139. H. Stragier, J.O. Cross, J.J. Rehr, L. Sorensen, C.E. Bouldin, J.C. Woicik, 'Diffraction Anomalous Fine Structure: a New X-ray Structural Technique', *Phys. Rev. Lett.*, **69**, 3064–3067 (1992).
140. F. Gatti, F. Fontanelli, M. Galeazzi, A.M. Swift, S. Vitale, 'Detection of Environmental Fine Structure in the Low-energy β-Decay Spectrometry of ^{187}Re', *Nature (London)*, **397**, 137–139 (1999).
141. K. Ulmer, 'Isochromat Spectroscopy', in *Advances in X-ray Spectroscopy*, eds. C. Bonnelle, C. Mande, Pergamon Press, Oxford, 451–466, 1982.
142. E. Sobczak, 'Extended X-ray Bremsstrahlung Isochromat Fine Structure', *Acta Physiol. Pol.*, **83**, 135–155 (1993).
143. Y. Tezuka, S. Shin, T. Ishii, T. Ejima, S. Suzuki, S. Sato, 'Photoemission and Bremsstrahlung Isochromat Spectroscopy Studies of TiO_2 (Rutile) and $SrTiO_3$', *J. Phys. Soc. Jpn.*, **63**, 347–357 (1994).

Energy Dispersive, X-ray Fluorescence Analysis

E. Selin Lindgren
Chalmers University of Technology and Göteborg University, Göteborg, Sweden

1 Introduction 1
2 Physical Principles 2
2.1 Interactions of X-rays with Matter 2
2.2 Principles in Quantitative Evaluation of Element Concentration 3
2.3 Interelement Effects 3
2.4 Dependence of Detection Limits and Penetration Depth on X-ray Energy and Source Characteristics 4
3 Energy-dispersive X-ray Fluorescence Instrumentation 5
3.1 General Features 5
3.2 X-ray Sources 6
3.3 Energy-dispersive X-ray Fluorescence-spectrometer Designs 7
3.4 Detector Characteristics 8
4 Sample Handling and Quantification in Practice 8
4.1 Sample Preparation 8
4.2 Quantitative Analysis of Major or/and Trace Constituents 8
4.3 Surface Analysis and Microanalysis 10
5 Conclusions: Needs for Future Developments 10
Acknowledgments 11
Abbreviations and Acronyms 11
Related Articles 11
Appendix 1 Approximations for "Thin" and "Infinitely Thick" Sample Thickness 11
References 12

Energy-dispersive X-ray fluorescence (EDXRF) is an analytical method for qualitative as well as quantitative determination of elements in a sample, independent of their chemical form. It is built on the fact that elements that are irradiated with energetic X-rays have a certain probability of emitting characteristic X-rays, the energies of which are unique for each element. In the energy-dispersive systems, the emitted X-rays are detected by their energies. Use of the EDXRF technique has accelerated since the 1960s as a result of the development of solid-state detectors, nuclear electronics and small computers. EDXRF is multielemental and nondestructive and can be applied to large as well as small samples of different composition and character. If conditions are optimized, minimum detection limits (MDLs) can be below the nanogram level for small laboratory instruments and into the femtogram region for more advanced instrumentation [total reflection X-ray fluorescence (TXRF), synchrotron radiation].

EDXRF spectrometers exist at many degrees of sophistication, ranging from advanced laboratory instruments to small portable instruments for field observations. They can be designed for analysis of bulk material or for scanning and elemental mapping of small areas.

Typical applications for EDXRF are analysis of agricultural material, medical samples, archaeological and historical objects and environmental samples such as soil, ores, water and aerosol particles.

1 INTRODUCTION

EDXRF is an analytical method for qualitative as well as quantitative determination of elements in a sample, independent of their chemical form. It is built on the fact that elements which are irradiated with energetic X-rays have a certain probability of emitting characteristic X-rays. The energy, E, of this radiation is uniquely dependent on the element, which emits the characteristic X-rays by Moseley's law, $E = K(Z - \sigma)^2$, where K and σ are constants that vary with the spectral series and Z is the atomic number of the element. In energy-dispersive systems, the emitted X-rays are detected by their energies and not by their wavelengths as in wavelength-dispersive X-ray fluorescence (WDXRF).

Typical applications for EDXRF are analysis of agricultural material, medical samples, archaeological and historical objects and environmental samples such as soil, water and aerosol particles. EDXRF is multielemental and nondestructive and can be applied to small samples (about 0.1 mg). For scanning objects and mapping small areas in a specimen, EDXRF instruments with narrow beams of X-rays can be used.

Use of the EDXRF technique has accelerated since the 1960s as a result of the development of solid-state detectors, nuclear electronics and small computers. Today about 3000 EDXRF spectrometers are in use, many of which belong to research laboratories. A vast number of scientific publications exist on the subject of EDXRF and its applications. Authoritative textbooks on the principles and instrumentation for EDXRF are those of Bertin,[1] Jenkins et al.[2] and Van Grieken and Markowics.[3]

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

2 PHYSICAL PRINCIPLES

2.1 Interactions of X-rays with Matter

When a beam of X-rays impinges on a sample it will react with the atoms in the sample by three processes, namely the photoelectric effect and coherent (Rayleigh) and incoherent (Compton) scattering. For the photoelectric effect to occur, the energy of the impinging photons has to be large enough to create a vacancy in one of the shells of the studied elements. The photoelectric effect may result in emission of characteristic X-rays, but once a vacancy has been created in an inner shell the atom can also get rid of the excitation energy by emission of an Auger electron. The probability that characteristic X-rays will be emitted – and not an Auger electron – varies from one element to another and is described as the fluorescence yield. For elements of low atomic number, emission of Auger electrons dominates, whereas emission of characteristic X-rays is more likely for elements of high atomic number. A sketch of the X-ray fluorescence (XRF) process leading to emission of a characteristic X-ray is shown in Figure 1.

The scattering processes are of two types, incoherent and coherent. The probabilities of both kinds of scattering vary with photon energy and the composition of the sample. The scattered X-rays are an important part of the background radiation on which the characteristic peaks are superimposed. Figure 2 shows a typical X-ray spectrum with characteristic and scattered radiation.

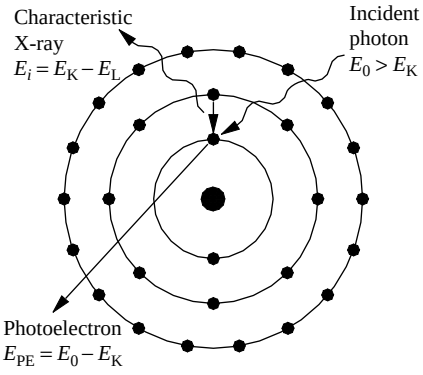

Figure 1 Schematic drawing of the XRF process in an atom: An incident photon of energy E_0 transfers its energy to an electron in the K-shell. The electron is expelled from the atom and leaves a vacancy. An electron from the L-shell makes a transition to the K-shell to fill the vacant site. The difference in binding energy between the two shells, $E_K - E_L$, is used by the atom to emit a characteristic X-ray of energy E_i. Note that the distances between the atomic shells are not drawn to scale.

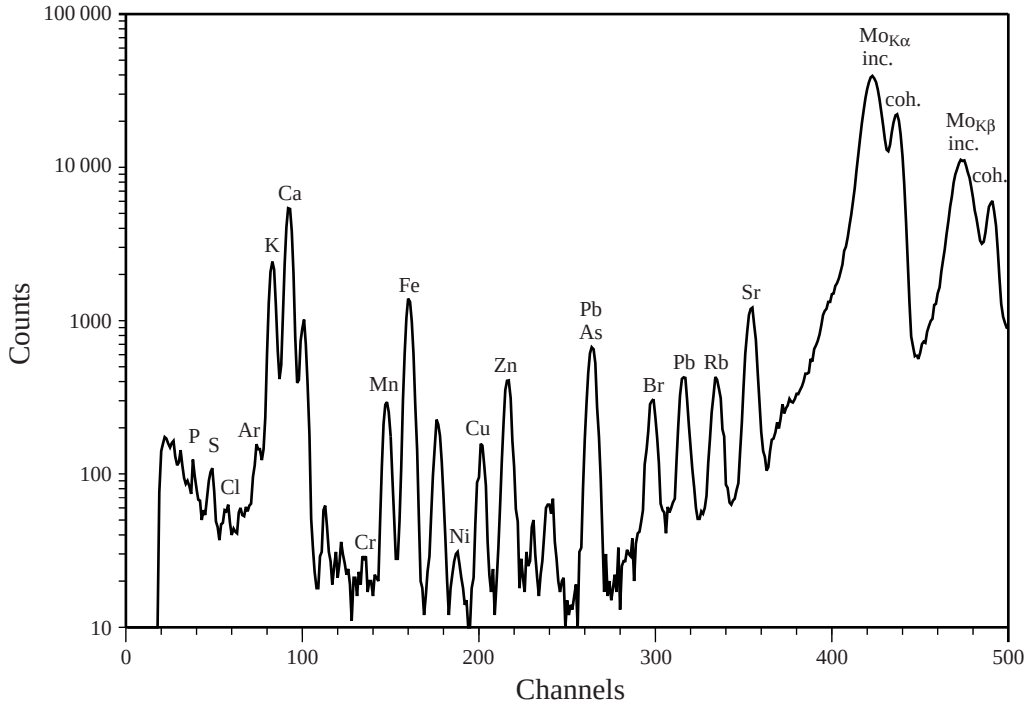

Figure 2 X-ray spectrum of a National Bureau of Standards (NBS) standard biological sample, 1571, ''orchard leaves'', measured in a secondary target EDXRF spectrometer of the geometry of Figure 3. The sample thickness is 30 mg cm^{-2}. The excitation conditions were as follows: tube voltage 55 kV, current 10 mA. The characteristic Kα lines are denoted by chemical symbols, except for Pb, for which both Lα and Lβ are marked. Note the high intensities of the coherently and incoherently scattered Kα and β radiation of the secondary target, which in this spectrum is Mo. The X-ray spectrum has been recorded with a Si(Li) detector.

2.2 Principles in Quantitative Evaluation of Element Concentration

The quantitative evaluation of element concentrations in EDXRF depends on whether the source is monochromatic or contains a range of different energies. Polychromatic X-ray sources can, however, be treated as a sum (or an integral) of monochromatic sources, and the discussion below is therefore limited to the case of monochromatic excitation. For this case, and with further limitations to homogeneous samples and in the absence of interelement effects, a simple relationship can be obtained between the mass thickness (or areal density), m_i, of a specific element and the measured intensity I_i of its characteristic X-ray peak:

$$I_i = I_0 K_i m_i D_i \tag{1}$$

In Equation (1) the following notation is used:

$I_0 =$ the intensity of the primary beam of energy E_0, which, in the case that the source is a secondary-target X-ray tube arrangement, is influenced by anode material and operating voltage and current.
$I_i =$ the intensity of characteristic X-rays (K- or L-radiation) of element "i" of energy E_i.
$D_i =$ a correction factor for attenuation of both the incident X-rays and the characteristic X-rays of element "i" in the sample.
$K_i =$ a factor that depends on the set-up of the instrument and basic physical processes.

The amount of analyte is often written as:

$$m_i = W_i m \tag{2}$$

In Equation (2) W_i is the weight fraction and m_i is the mass thickness of element "i" in the sample.

K_i thus depends on instrumental factors and on fundamental physical parameters related to the energies of the incident and characteristic X-rays. Important parameters to consider are:

- the geometry of the XRF spectrometer;
- the attenuation of the exciting and characteristic radiation along the beam paths and in the detector window;
- the detector efficiency for radiation of energy E_i;
- the probability of producing a vacancy with incident radiation of energy E_0 in the relevant shell of element "i";
- the fluorescence yield of element "i".

Note, however, that the essential feature of Equation (1) is that – under the assumptions of homogeneity and absence of interelement effects in the sample – the factor K_i is independent of the mass thickness, m_i, of the analyzed element and of the attenuation correction, D_i. Thus, it can be determined through calibration of the instrument if it is operated under well-defined conditions.

The attenuation correction, D_i, in Equation (1) has the form given in Equation (3):

$$D_i = \frac{1 - \exp -[(\mu(E_0) \csc \Phi_1 + \mu(E_i) \csc \Phi_2)m]}{[\mu(E_0) \csc \Phi_1 + \mu(E_i) \csc \Phi_2]m} \tag{3}$$

In this expression $\mu(E_0)$ and $\mu(E_i)$ are the mass attenuation coefficients for the exciting and characteristic radiation of energies E_0 and E_i respectively. These can be expressed as weighted sums of the individual mass attenuation coefficients for all elements in the sample. Φ_1 is the glancing angle between the exciting radiation and the sample plane and Φ_2 the glancing angle between the characteristic radiation and the sample plane. There are two cases for which the attenuation correction and thus Equation (1) takes on an especially simple form, namely for "thin" and "infinitely thick" samples (see Appendix 1).

In some cases it is also necessary to take into account the fact that the characteristic radiation produced in one element as well as the scattered radiation can cause fluorescence in another element (interelement and related effects, as discussed below). This may lead to enhancement of one spectral line and further attenuation of another. Furthermore, to perform quantitative evaluation one needs to know whether impurities are present in the spectrometer material and the properties of the detector.[1–3]

From Equations (1) and (2) it is evident that the net areas of the characteristic peaks in the X-ray spectrum will have to be evaluated in order to determine element concentrations in a sample. This can be done automatically by means of computer programs. In evaluation of net peak areas in a spectrum, the problem of interference between spectral lines will, however, have to be addressed. In a number of cases, the K-lines from one element will overlap or coincide with the L-lines of another element, as is illustrated in Figure 2 by the overlap of As Kα with Pb Lα. In other cases, the Kα from one element will coincide with the Kβ from another. By knowing relative ratios, Kα/Kβ, Lα/Lβ, it is possible to resolve the overlapping peaks. However, this procedure will reduce the accuracy of the measurement to an extent determined by the statistics of the measurement.

2.3 Interelement Effects

In a matrix composed of many elements, each element will interact with the radiation from all other elements. If the matrix elements have characteristic radiation that is on the short wavelength side of the absorption edge

of the element of interest, element "*i*", conditions may be favourable for enhancement of the lines of element "*i*". The combined effects of the matrix elements on each other are often called *interelement* or *absorption-enhancement* effects. Enhancement effects are especially important for cases in which the analyte is present at trace level and the enhancing elements at percentage levels. In order to illustrate this, consider the determination of one element, "*i*", by measuring its Kα radiation in a matrix consisting of lighter as well as heavier elements. In this case, the lighter elements will not be able to enhance the characteristic Kα radiation of the element "*i*" because their characteristic radiation will not be energetic enough to create the necessary vacancy in the K-shell of element "*i*" (see Figure 1). Thus, it is necessary to look at the effects of the heavy elements in the sample. If their Kα energies are lower than that of the absorption edge for the K-shell in element "*i*" they will not cause enhancement. If, on the other hand, their Kα lines have higher energies than that of the K-shell absorption edge they will enhance the Kα radiation of element "*i*" – in fact they are well matched for doing so. However, the general problem of determining whether or not enhancement will occur is not so difficult for homogeneous samples, since absorption edge energies for different shells and characteristic X-ray energies for all elements are long known and readily available in tabulated form in textbooks.

A situation similar to the interelement effect occurs when the backscattered exciting and/or fluorescent radiation excites the elements in a thick sample. These effects are of relatively large importance for analysis of trace elements embedded in organic matrices, since the medium-weight and heavy elements are usually present in low concentrations and the incoherent scattering in these cases is dominating.[4] (Compare the spectrum from the organic specimen as given in Figure 2.)

2.4 Dependence of Detection Limits and Penetration Depth on X-ray Energy and Source Characteristics

The intensity of a beam of X-rays passing through a sample is attenuated by the above-mentioned photoelectric and scattering processes according to the exponential law:

$$I(E) = I_0(E) \exp -[\mu(E)\rho x] \quad (4)$$

In Equation (4) the following notation is used:

$I_0(E)$ = the original intensity of the beam.
$I(E)$ = the intensity after the beam has travelled a distance x in the sample.
ρ = the density of the sample.
$\mu(E)$ = the mass attenuation coefficient of the sample material.

Since the three attenuation processes that contribute to the total attenuation of the beam are independent of each other, the mass attenuation coefficient can be written as a sum of coherent, incoherent and photoelectric mass attenuation coefficients. Each process has a functional dependence of the photon energy, E, and this is shown explicitly in Equation (4).

2.4.1 Penetration Depth

A relevant question in XRF analysis is how deeply the X-rays will penetrate into the material. A quantitative measure of this can be given in terms of the *penetration depth*, which is usually defined as the path length needed for the intensity of the X-rays to be reduced by a given factor. Since the mass attenuation coefficients, $\mu(E)$, in Equation (4) vary widely with X-ray energy and material composition, penetration depths are indeed very different for hard and soft X-rays and vary appreciably with sample composition. Table 1 shows some typical values for penetration depths for characteristic X-rays of some elements embedded in an iron and in a carbon matrix.[5] In this example, the penetration depth varies by 3–4 orders of magnitude. Generally, X-rays of low energy do not penetrate very far in a matrix of heavy material, whereas high-energy X-rays in a light material have high penetration. This fact has a large influence on whether corrections for attenuation have to be made when quantifying element concentrations. Concepts such as "thin" or "thick" samples are always seen in relation to the energy of the X-rays of interest.

2.4.2 Minimum Detection Limit

A widely used definition for the MDL in EDXRF is the concentration (or, alternatively, the amount of mass)

Table 1 Depth of penetration in micrometers of characteristic radiation from sulfur, chromium and yttrium in a matrix of iron (density 7.87 g cm^{-3}) and graphite (density 2.25 g cm^{-3}). The penetration depth for reduction of X-ray intensities to two-thirds of the original has been calculated according to the formula $I = I_0 \exp(-\mu\rho x_{2/3})$. The elements that generate characteristic radiation are assumed to be present at trace levels, which is why the density of the matrix is not affected

Characteristic radiation	Matrix	Mass attenuation[a] (μ, cm^2 g^{-1})	$x_{2/3}$ (μm)
S Kα	Fe	1170	0.44
Cr Kα	Fe	119	4.3
Y Kα	Fe	58.5	8.8
S Kα	C	208	8.7
Cr Kα	C	15.3	120
Y Kα	C	0.813	2200

[a] J.W. Robinson (ed.), *Handbook of Spectroscopy*, CRC Press, Boca Raton, Vol. 1, 1974.

needed for the number of counts in the characteristic peak of an element to be equal to or larger than $3\sqrt{B}$, where B is the number of counts in the background below the peak of element i[1,2] ($\sqrt{B}$ corresponds approximately to one standard deviation). Specifically:

$$\mathrm{MDL} = \frac{3\sqrt{B}}{k_i} \tag{5}$$

where k_i is the corresponding sensitivity constant measured as the net number of counts in the characteristic peak per unit concentration.

Since the attenuation of X-rays is stronger for soft X-rays than for hard X-rays, and the background in many designs of EDXRF spectrometers is rather flat for energies below the scattered radiation, optimum values for MDL are usually obtained for medium-weight elements. From Equation (5) it can be seen that the general problem in lowering detection limits can be achieved by finding ways of increasing the intensity of the characteristic peak while decreasing the intensity of the background radiation. A substantial portion of the background radiation is due to incoherent and coherent scattering of the primary radiation in the sample itself. These scattered peaks will, by means of secondary processes in the sample, give rise to a general increase in the background also at lower energies. Furthermore, the scattered radiation will also contribute to an increase in the background level owing to incomplete charge collection and other processes in the solid-state detectors. Therefore, sample properties that influence scattering processes, for example thickness and average atomic number, have an essential influence on the detection limits in an EDXRF spectrometer.

The counts in both peak and background are time dependent, but in different ways: the peak increases linearly with time, whereas $\sqrt{B}$ increases with the square root of time. This means that the detection limit will be reduced as analysis time is increased, and the analyst can use this fact if elements present in low concentrations are of special interest. The price for improvement in MDL is, of course, a longer running time for the spectrometer. For many laboratory instruments, collection times in the range 200–1000 s for each X-ray spectrum are chosen.

In practice, for bulk samples of organic material, MDLs are of the order of 0.1 ppm for laboratory instruments for monochromatic sources with energy matching the binding energy for the K-shell of the element of interest.[6] In bulk geological samples, MDLs are generally somewhat larger. Measured as absolute amounts (for example in mass units), the lowest MDLs can reach the picogram region for thin samples (especially with TXRF) and the femtogram region for micro beams that utilize a focused synchrotron beam as source.[7] MDLs in the 10^{-10} g region are reached for many elements in thin and thick samples in laboratory EDXRF spectrometers.

2.4.3 Sensitivity

A concept that is often used to describe the performance of an EDXRF method is the *sensitivity*, which is defined as the change in instrument response (net number of counts in a characteristic peak) per unit change in the concentration of a measured element. Thus, sensitivity in X-ray instruments can be given as count rate per ppm. The sensitivity concept is useful for comparing analyses of different elements in a given spectrometer. It varies to a high degree with the difference in energy between the incident and the characteristic radiation and is highest if the energy of the incident radiation is only slightly greater than the energy of the K-edge for the element in question.

However, sensitivity in XRF instruments[8] (expressed as count rate per ppm) is a measure of the excitation capability for each element with a specific XRF instrument. Usually, sensitivity is not used for comparing different instruments. For reasons of comparison, a more appropriate parameter is the MDL achieved with each instrument.

3 ENERGY-DISPERSIVE X-RAY FLUORESCENCE INSTRUMENTATION

3.1 General Features

A typical feature in element analysis in practice is that samples may contain major as well as trace concentrations of different elements. Samples such as alloys, mineral ores and ceramics usually contain a number of elements, the concentrations of which may vary by several orders of magnitude.

In an energy-dispersive X-ray spectrometer, the important parts to consider are the X-ray source, the beam and spectrometer geometry and the detector. An obvious goal for the designer of XRF spectrometers is to maximize the intensity of the characteristic X-rays from the elements while keeping the background radiation as low as possible. Most energy-dispersive (ED) spectrometers have an elemental range from Na to U and, if special detectors with ultra-thin windows are used, the range can be extended to still lighter elements such as C. There are many different designs of ED spectrometers for laboratory and field work.[9] Combinations of WDXRF and EDXRF systems are also used, in which the better resolution of wavelength-dispersive (WD) instruments is utilized for the lightest elements and the simplicity and higher sensitivity of ED systems is utilized for elements above Ca The cost of an ED system is about US$ 20 000 and upwards.

3.2 X-ray Sources

In photon-induced XRF, the most common sources are X-ray tubes, which can be used directly for broadband excitation or in combination with filters or secondary fluorescers to obtain semimonochromatic radiation. In some spectrometers, especially portable instruments for field work, and for problems in which only one or a few elements are of interest (for example in routine process control), radioisotopes are useful.

3.2.1 X-ray Tubes

In X-ray tubes, the radiation spectrum is influenced by two major processes that occur when electrons are accelerated by the high voltage from the cathode to the anode. One process is the creation of characteristic X-rays in the anode material itself, which occurs when the kinetic energy of the impinging electron is larger than the binding energy of a particular shell in the anode material. The other process is the creation of continuous bremsstrahlung, which is created when the electrons are slowed down in the anode. The bremsstrahlung spectrum contains energies from zero up to the kinetic energy of the electrons, which is equal to the electron charge times the tube voltage. Accordingly, the spectrum from the X-ray tube contains a continuous distribution overlaid with the characteristic radiation from the anode material. By varying parameters such as anode material and thickness, tube voltage and current settings, the broadband spectrum from the X-ray tube can be optimized for the elements of interest. Some X-ray tubes are equipped with dual or multiple anodes.

As the MDL for a given element is dependent on the intensity of the characteristic as well as the background radiation (Equation 5), there is often a desire to use monochromatic or semimonochromatic radiation. This can be achieved by use of filters and/or secondary target arrangements. Filters inserted between the X-ray tube and the sample will allow transmission of the characteristic line of the filter but will attenuate the low-energy part of the tube radiation by a substantial factor. Thus, filtering may change the shape of the tube spectrum and improve MDLs for elements lighter than the filter material. By varying filter material and filter thickness, this method can give a high flexibility in the use of tube excitation.

In secondary-target arrangements, the radiation from the X-ray tube is allowed to excite a target of a material whose characteristic radiation (almost monochromatic) is used as a source to excite the sample. Figure 3 shows an arrangement of this kind. The material in the secondary target can be chosen so that its characteristic radiation is just above the K-edge of the elements of most interest in the sample. The anode material in the X-ray tube is generally chosen to have a high output over a wide range of energies in order to allow different secondary target materials to be used. Although the intensity of the primary beam is reduced considerably by the secondary target arrangement, the technique usually gives low detection limits for many elements. This is mainly because of the low background in the X-ray spectra (see Figure 2). Other techniques to modify the output from an X-ray tube utilize scatterers to polarize the X-rays or multilayers, as demonstrated for TXRF spectrometers.[10]

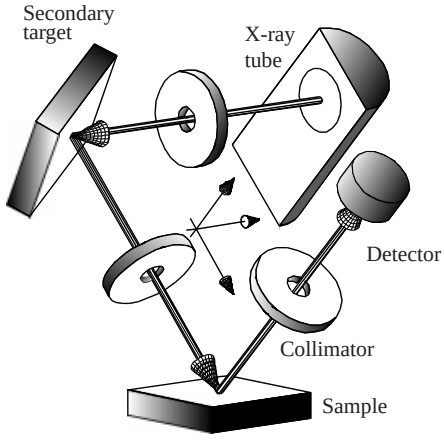

Figure 3 Preferred geometry of a secondary target EDXRF spectrometer. The broadband radiation from an X-ray tube excites a secondary target of a pure specimen, most often a metal. The characteristic Kα and β radiation from the secondary target is used as semimonochromatic X-ray source to excite the different elements in the sample. A detector registers the radiation from the sample. Note that the three X-ray beams are mutually orthogonal and that the sample plane is horizontal.

For microbeam EDXRF, microfocus low-power X-ray tubes with anode spots of the order of 100 μm^2 are now available, as are high-intensity rotating anode sources (18 kW) with various focal spot sizes.[11]

3.2.2 Radioactive Sources

Radioisotopes as X-ray sources utilize the γ-rays from the decay of the radioisotope or the characteristic X-rays emitted from the decay product. The output intensity is lower than that which can be achieved by an X-ray tube, but the advantages are that the intensity of the source is stable and no electric power supply is needed. Radioisotopes are therefore practicable in portable field instruments. The drawback is the limited half-lives for some of the most used isotopes (for example 2.7 years for ^{55}Fe with 5.9- and 6.4-keV X-rays and 1.3 years for ^{109}Cd with 22.2- and 25.5-keV X-rays plus 88.2 keV γ-rays). The most common geometry of the radioisotope sources is either that of a point source or that of an annular source. The sensitivity and MDLs are for some instruments

almost identical to those for X-ray tube secondary-target arrangements.[12]

3.2.3 Synchrotron Radiation

Synchrotron radiation is ideal as a source for EDXRF, since it can be obtained at the desired energy with high brilliance. Furthermore, it is polarized, which is useful for decreasing scattered radiation and thus reducing the background. Synchrotron radiation at the National Synchrotron Light Source (NSLS, Brookhaven National Laboratory) with collimated and focused X-rays of high energy has been used in an X-ray microscope mapping trace elements with a spatial resolution of $10 \times 10\,\mu m^2$ and with MDLs into the femtogram region for thin specimens.[7] However, synchrotron radiation is expensive and is not generally available in research laboratories.

3.3 Energy-dispersive X-ray Fluorescence-spectrometer Designs

As pointed out before, there are three main points to consider for achieving good performance in an EDXRF system, namely the X-ray source, the detector and the design of the spectrometer. A few things of importance for the design will be mentioned here, namely beam geometry, choice of materials and compactness of the design.

The problems will be discussed from the perspective of a secondary-target set-up but are easy to generalize to other arrangements.

3.3.1 Beam Geometry

A large part of the background signals in X-ray spectra depends on primary and secondary scattered radiation in the sample. This can be illustrated in the secondary target XRF arrangement shown in Figure 3. In such a system, contributions to the background in the final spectrum are obtained from: (a) the primary radiation from the tube, (b) the semimonochromatic beam from the secondary target and (c) the radiation from the sample.

The primary radiation from the X-ray tube is very intense and, although it will be scattered twice (in the secondary target and the sample), it is likely to contribute to the background unless special care is taken. The radiation from the secondary target will be scattered when reaching the sample and constitutes the two large scattered peaks shown in Figure 2. These peaks have been observed to be responsible for part of the flat background in the low-energy region by multiple scattering in the sample and by incomplete charge collection effects in the detector.

From basic physical principles it is well known that scattering is at a minimum when the scattering angle is 90°. Thus, it is important that the impinging and emitted radiation form an angle of 90° with respect to each other. This will be achieved if the secondary target and the sample are both oriented with angles of incidence of 45° with respect to the beam directions. Because the radiation from the X-ray tube is scattered twice – in the secondary target and in the sample – and since the reason for scattering at 90° is to utilize both minimum scattering cross-section and the production of a polarized beam, it is also of importance that the geometry of the three beam paths is a three-axial (*xyz*) geometry as shown in Figure 3 and not a *xyx* planar geometry.[13]

3.3.2 Choice of Materials – Compact Designs

In an EDXRF spectrometer the major part of the primary radiation emitted by the tube and secondary target will not be directed towards the sample but will interact with the material in the spectrometer. This interaction will take place by means of the same processes as occur in the sample itself. Different materials have different abilities to scatter and absorb radiation, and the purity of the material will have an influence on whether characteristic lines from impurities in the material close to the beam paths will pass into the detector. Thus, it is important that the materials in objects such as filters and collimators are of the highest possible purity.

Another important aspect to consider in an XRF set-up is the actual geometric dimensions of the various parts of the spectrometer. Since the general goal is to have efficient and high sensitivity in the instrument, long beam paths are of disadvantage. In air, a long beam path for the radiation from sample to detector will contribute to the attenuation of characteristic radiation from the sample, and thus significantly reduce detection efficiency for light elements in particular. Thus, a geometry as compact as possible with short beam paths is desirable. However, the material in the housing of the spectrometer has to be chosen so that unwanted radiation can be stopped efficiently. The combination of compact geometry and high stopping efficiency generally means that light materials such as aluminum should be avoided.

Another aspect to consider in relation to Figure 3 is the position of the sample. In many spectrometers, the sample position is vertical, which means that samples have to be self-supporting. By choosing an arrangement in which the sample plane is horizontal and the detector looks down at the sample, a more versatile instrument is achieved, in which the sample can also be in non-self-supporting forms, for example as liquids and powders. With a vertical sample position, such specimens are difficult to irradiate directly by the beam.

3.4 Detector Characteristics

For ED spectrometers, liquid-nitrogen-cooled Si(Li) detectors are the most widely used, having their optimum performance for X-rays of an energy ranging from a few to about 25 keV. For K-lines from heavier elements (30–140 keV) Ge(Li) and high-purity germanium detectors are preferable owing to better detection efficiency in this region. For benchtop and portable spectrometers, scintillation counters and gas-filled proportional counters are sometimes used, often in combination with filters to compensate for the lack of energy resolution. Other types of detectors have been developed, for example mercury iodide detectors at room temperature for high- and medium-Z-element analysis.[14] High-purity silicon detectors are now able to separate K fluorescence lines down into the 1 keV region and compare well with compound semiconductor detectors such as CdZnTe and HgI_2.[15] An overview of the present status of several semiconductor detectors can be found in McGregor and Hermon.[16]

4 SAMPLE HANDLING AND QUANTIFICATION IN PRACTICE

4.1 Sample Preparation

A typical feature in element analysis in practice is that samples may contain high as well as trace concentrations of different elements. Samples such as alloys, mineral ores and ceramics usually contain a number of elements whose concentrations may vary by several orders of magnitude.

EDXRF analysis can be performed on a vast number of samples of very different composition and status. Examples of common samples are:

- geological – for example minerals, rock, soil and sediments;
- biological – roots, leaves, stems, moss, lichens, algae, fish, animals;
- medical – blood, hair, skin, muscle, brain, teeth, bone, enzymes;
- industrial – metals, ores, paper, pulp;
- cultural – pottery, paintings, statues, coins, books, clothes;
- environmental – water, air pollutants, waste, pollen.

There is no strict division between the different categories, but the above list gives an idea of the diversity of the samples concerning range of elemental concentrations – from traces in biological material to major components in minerals – as well as the range of densities involved. The material may exist in different forms, such as airborne particles, liquids and solids.

4.1.1 Solid Samples

Homogeneous samples for which the concentrations of the elements of interest are significantly larger than the detection limits (usually 0.1–1 ppm) are conveniently analyzed without any sample preparation. This is a great advantage, since a minimum of sample treatment will minimize the risk of contamination. Many samples belong to this category.

Some samples may have sufficiently high element concentrations but be inhomogeneous and contain grains of different sizes. In view of the poor penetration depths for X-rays from light elements, the grain size may affect the quantitative evaluation. For such samples either grinding into a fine powder or digestion and dilution may be necessary.

4.1.2 Liquid Samples

Trace elements in water samples can be present in different forms, for example as insoluble particulates or as colloidal or as soluble species. Suspended material can be separated by filtration, and the deposits on the filter can be directly analyzed in the X-ray spectrometer. For water, many element concentrations are below MDLs, especially for heavy metals. The colloidal and soluble fractions may therefore need preconcentration, which may be achieved by evaporation due to heating or by freeze-drying. Care must be taken, however, to avoid loss of volatile elements in these processes. A freeze-dried residue may be acid digested. A number of other chemical preconcentration procedures have been used on natural waters, snow and rain, with acceptable precision and recovery rates.[17] Many of these techniques have the advantage that they increase element concentrations while at the same time giving rise to thin film samples that are easy to evaluate quantitatively. For many fluids – and in particular if TXRF technique can be used – the fluid can be deposited directly onto a carrier and dried at low temperature.

4.2 Quantitative Analysis of Major or/and Trace Constituents

As seen from section 3.2, quantification involves knowledge of some factors that are different for each spectrometer (geometrical properties, detector and source characteristics, etc.) and some parameters that are dependent on physical processes. In a practical case, both these aspects have to be taken into account.

4.2.1 Homogeneous Samples: General Approach

Samples analyzed in spectrometers with monochromatic or nearly monochromatic X-ray sources are usually rather

easy to work with, since Equations (1–3) and (6–8) can be used for quantitative evaluation. A convenient approach is to calibrate the spectrometer with known standards, which are either thin or of known mass thickness. Calibration gives information on the factors K_i for different characteristic X-ray energies. These factors can be stored for a special configuration of a given spectrometer and used for element quantification of the unknown samples. If an unknown sample is considered to be thin for some elements ($D_i = 1$), the amounts of these elements can be readily evaluated by use of Equation (6). If the sample is of intermediate or infinite thickness, the attenuation factors D_i will have to be evaluated either from measurements of mass thickness or from information contained in the coherently and incoherently scattered radiation, as discussed below.

For polychromatic X-ray sources, for example broad-band excitation from X-ray tubes, Equation (1) is not valid. In this case, the intensity, I_i, can be described by an integral over the distribution of excitation energies. In practice, the approach often used is to calibrate the instrument with standards as similar as possible to the investigated samples.

A special case of homogeneous samples is liquids. For most liquid samples, quantification is facilitated by the ability to spike the sample with internal standards before digestion or further treatment. By choosing two internal standards with different attenuation properties, the attenuation for all elements in the sample can be obtained together with information on the sample mass thickness.

4.2.2 Heterogeneous Samples

For heterogeneous samples, quantitative evaluation is more difficult, and this is particularly the case for X-rays from light elements, for which the penetration depths of the characteristic X-rays are small. Another problem with heterogeneous samples is the profile of the X-ray beam itself. In the ideal case, the beam intensity should be constant over the beam area and drop abruptly to zero at the edges of the beam. In reality, the beam intensity does not behave so well, at least not for the radiation from X-ray tubes. Owing to the geometrical effects and different path lengths for different parts of the radiation from the tube to the sample, the beam may not have a homogeneous intensity distribution. This has also been shown to be the case for secondary-target arrangements.[18] Thus, for laboratory instruments, knowledge of intensity variations within the beam is essential for heterogeneous samples. Alternatively, one can achieve good accuracy for heterogeneous samples and beams by scanning the sample, so that effects due to sample and beam inhomogeneity can be leveled out.

4.2.3 Methods for Quantification in Practice

Many samples can be regarded as "thin" for medium weight and heavy elements, but in reality few samples are thin with respect to light elements (see Table 1). Thus, one will often have to evaluate the attenuation properties of the sample, at least for some elements. Many methods have been developed to assist quantification. Some of these use an experimental approach, whereas others build on the fundamental parameters of the physical processes involved, often in combination with knowledge after calibration of the instrumental performance. A few methods to illustrate the principles in quantitative evaluations for nonliquid samples will be described.

The first method, which was one of the earliest, is the emission–transmission method.[1,2] A precondition is that the sample is of "intermediate thickness", as defined in Appendix 1. The method is based on an independent measurement of the factor $\exp -[(\mu(E_0) \csc \Phi_1 + \mu(E_i) \csc \Phi_2)m]$, which is contained in the attenuation correction, D_i. This factor, which measures the sample attenuation, can be directly evaluated by three measurements:

1. First the characteristic radiation from an element "*i*" in the sample is measured.
2. Under identical conditions, the same characteristic radiation is measured from the combination of the sample and a target of element "*i*" positioned behind the sample.
3. In the third measurement, the characteristic radiation is measured from the target of element "*i*" alone, without sample.

It is not difficult to show that these measurements will give knowledge of the factor $\exp -[(\mu(E_0) \csc \Phi_1 + \mu(E_i) \csc \Phi_2)m]$, which is needed in Equation (3). If this factor is known, it is straightforward to calculate m_i of element "*i*" from measurement of I_i and knowledge of the instrumental factors K_i and I_0 according to Equation (1).

Another well-known method is to make use of the information contained in the coherent and incoherent scattering of the incident (exciting) radiation. Under certain conditions, the intensity of the scattering peaks or combinations of these can be used as internal standards.[1,2,19–21] From Figure 2, which is an X-ray spectrum from a biological sample, in which molybdenum has been used as secondary target, it is seen that the incoherently scattered radiation is very pronounced. In a biological or other kind of light element sample in which medium and heavy elements are present at trace levels, the incoherently scattered radiation is essentially a measure of the mass seen by the impinging beam. The argument for this is that for the light elements (except hydrogen) e.g. carbon, nitrogen and oxygen, which are

the main constituents in most biological samples, the mass of an atom is closely proportional to its atomic number (Z).

Furthermore, the binding energies of the electrons in the light elements are small compared with the energy of the exciting radiation. Thus, the exciting radiation essentially sees a cloud of free electrons, and the probability of incoherent (Compton) scattering will be proportional to the number of electrons, which in turn is approximately proportional to the sample mass. This approach is also valid for some elements of importance in geochemical samples, such as Mg, Si and Ca (ratios of relative atomic mass to atomic number: 2.03, 2.01, 2.00) but less so for Na, Al and K (ratios of relative atomic mass to atomic number are in the range 2.09–2.06). For samples composed of heavy elements, the ratios of atomic mass to atomic number will be changed owing to increased neutron excess in the atomic nuclei. Furthermore, the binding energy of the inner electrons increases and the free electron incoherent scattering cross-section is no longer such a good approximation. Secondary scattering effects in the material may also appear. Several practical methods, however, have been developed in which use is made of the information contained in both coherently and incoherently scattered radiation for thick samples of geochemical origin. Giauque et al.[19] and Giauque[20] have used the ratios of the incoherently scattered radiation from Ag Kα and Kβ radiation together with the coherently scattered Ag Kα to ascertain sample mass thickness and photoelectric cross-sections for thick geochemical samples. Also, in the case of WDXRF, the scattered Compton lines from the X-ray tube have proved to be useful as internal standards in light element samples.[21]

The general aim of quantification in EDXRF is to arrive at the sample composition from the information contained in the spectrum. In an XRF spectrum the medium-heavy and heavy elements in the sample will reveal their presence by their characteristic radiation as well as by their contribution to the coherently and incoherently scattered radiation. Thus, a first estimate of the relative abundance of these elements can be made directly from the spectrum. As discussed earlier, however, this first estimate will be significantly modified by considering attenuation and interelement effects, etc. To evaluate the mass thickness of the light elements – which do not show up by their characteristic X-rays – their contribution to the scattering peaks can be determined from the spectrum by subtraction of the scattering due to the medium-heavy and heavy elements. From this evaluation, the total mass thickness from light as well as medium and heavy elements can be calculated.

In an iterative process, in which all the information from the elements through their characteristic as well as scattered X-rays is used, the concentrations can be recalculated until the concentration values obtained agree with the experimental spectrum.

In the practical situation, the analyst is dependent on computer programs in which an X-ray library is stored together with information on the actual spectrometer (obtained by calibration samples). Spectrum evaluation programs are available in commercial instruments and from research groups. A more or less continuous development of these programs takes place, in which peak shapes and background profiles are described by model functions.[22]

4.3 Surface Analysis and Microanalysis

Quantification of thin surface layers or microstructures essentially makes use of the same fundamental methods as discussed above, namely adding of an internal standard for liquids, use of external reference material and use of scattered radiation. For thin self-supporting homogeneous solid surface layers, quantification can be achieved by use of scattered radiation, calibration constants (Equation 6 in Appendix 1) or fundamental parameters.

In the case of heterogeneous microstructures smaller than the X-ray beam diameter, quantification by means of external references is difficult in spectrometers based on X-ray tubes owing to the inhomogeneous beam profile as discussed above and the difficulty in positioning the samples within the beam. In these cases internal standards such as the scattering peaks from the sample will give reasonably good average values, although structures with resolution below 10 μm are presently difficult to study with laboratory equipment.[23]

5 CONCLUSIONS: NEEDS FOR FUTURE DEVELOPMENTS

Although EDXRF has many advantages, in that it is multielemental, nondestructive and needs a minimum amount of sample treatment, it has two distinct disadvantages:

- For bulk samples it is not very sensitive. MDLs for laboratory equipment are in the order of 0.1 ppm. Measured in mass units, it has MDLs of the order of a fraction of a nanogram for the best cases in realistic samples. For synchrotron radiation sources and special designs [TXRF, grazing emission X-ray fluorescence (GEXRF)] MDLs may be reduced by several orders of magnitude, but not for all kinds of samples. A substantial part of the lack of sensitivity

for bulk samples is due to effects caused by the presence of the large scattered peaks entering the solid state detectors.

- In comparison with microprobes with charged particles – electrons or protons – microprobes of X-rays with high lateral resolution are difficult to obtain. This is because charged particles can be steered by electric and magnetic fields, whereas X-rays will have to be focused with mirrors or other kinds of optical devices. The advantage of X-rays over charged particles is that they deposit less energy into the sample and thus will cause less heating and loss of volatile material.

Future development which would imply breakthrough on the above items would be:

- improvements in detector technology to minimize effects that cause increased background, for instance trapping and polarization in the detector material;
- development of laboratory X-ray lasers in order to obtain parallel, polarized and intense beams.

For both these developments it is easy to observe that not only EDXRF but a vast field of other analytical methods would benefit from them.

ACKNOWLEDGMENTS

The author is grateful to Ing. Paulis Standzenieks for assistance in preparing the manuscript.

ABBREVIATIONS AND ACRONYMS

ED	Energy-dispersive
EDXRF	Energy-dispersive X-ray Fluorescence
GEXRF	Grazing Emission X-ray Fluorescence
MDL	Minimum Detection Limit
NBS	National Bureau of Standards
NSLS	National Synchrotron Light Source
TXRF	Total Reflection X-ray Fluorescence
WD	Wavelength-dispersive
WDXRF	Wavelength-dispersive X-ray Fluorescence
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Sample Preparation for Elemental Analysis of Biological Samples in the Environment • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Peptides and Proteins **(Volume 7)**
X-ray Crystallography of Biological Macromolecules

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Nuclear Methods **(Volume 14)**
PIXE (Particle-induced X-ray Emission)

Radiochemical Methods **(Volume 14)**
Nuclear Detection Methods and Instrumentation

X-ray Spectrometry **(Volume 15)**
Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis • X-ray Techniques: Overview

APPENDIX 1 APPROXIMATIONS FOR "THIN" AND "INFINITELY THICK" SAMPLE THICKNESS

The sample thickness for an element can be divided into three categories as follows:

- "Thin" sample: the intensity of the characteristic X-rays from a certain element increases linearly with the amount of analyte in the absence of interelement effects.
- "Infinitely thick" sample: an increase in sample thickness does not give rise to any increase in the intensity of the characteristic X-rays.
- Samples of "intermediate thickness": the intensity of the characteristic X-rays grows with sample thickness but it does not grow linearly.

Using the same notation as used in Equations (1–3), these equations take on simple expressions. Thus, for "thin" samples, in which $m \rightarrow$ zero, $D_i = 1$ and, as shown in Equation (6)

$$I_i = I_0 K_i m_i = I_0 K_i W_i m \tag{6}$$

For "infinitely thick" samples ($m \rightarrow$ infinity), Equations (7 and 8) apply,

$$D_i = \frac{1}{m[\mu(E_0) \csc \Phi_1 + \mu(E_i) \csc \Phi_2]} \tag{7}$$

and

$$I_i = \frac{I_0 K_i W_i}{[\mu(E_0) \csc \Phi_1 + \mu(E_i) \csc \Phi_2]} \quad (8)$$

REFERENCES

1. E.P. Bertin, 'Principles and Practise of X-ray Spectrometric Analysis', Plenum Press, New York, 1975.
2. R. Jenkins, R.W. Gould, D. Gedcke, 'Quantitative X-ray Spectrometry', Marcel Dekker, New York, 1981.
3. R.E. Van Grieken, A.A. Markowics (eds.), 'Handbook of X-ray Spectrometry', Marcel Dekker, New York, 1993.
4. A.G. Karydas, T. Paradellis, 'Calculation of the Enhancement of Characteristic Radiation Due to Scattering in Radioisotope XRF Using Exact Differential Cross-sections', *X-ray Spectrom.*, **22**(4), 208–215 (1993).
5. R.P. Pettersson, E. Selin Lindgren, 'Energy-dispersive X-ray Fluorescence Analysis', in *Surface Characterization, A User's Sourcebook*, eds. D. Brune, R. Hellborg, H.J. Whitlow, O. Hunderi, Wiley-VCH, Weinheim, 136–153, 1997.
6. A. Viksna, E. Selin, 'Intercomparison Between Energy Dispersive and Total Reflection X-ray Fluorescence in Whole Blood Analysis', *J. Trace Microprobe Techniques*, **14**(4), 763–774 (1996).
7. K.W. Jones, W.M. Kwiatek, B.M. Gordon, A.L. Hanson, J.G. Pounds, M.L. Rivers, S.R. Sutton, A.C. Thompson, J.H. Underwood, R.D. Giauque, Y. Wu, 'X-ray Microscopy Using Collimated and Focussed Synchrotron Radiation', *Adv. X-ray Anal.*, **31**, 59–68 (1987).
8. W. Ladisich, R. Rieder, P. Wobrauschek, 'Quantitative Total-reflection X-ray Fluorescence analysis with Monoenergetic Excitation', *X-ray Spectrom.*, **23**(4), 173–177 (1994).
9. A. Newman, 'X-ray Fluorescence Makes it Elementary', *Analytical Chemistry News and Features*, August 1, 493–500 A (1997).
10. J. Knoth, H. Schneider, H. Schwenke, 'Tunable Exciting Energies for Total Reflection X-ray Fluorescence Spectrometry Using a Tungsten Anode and Bandpass Filtering', *X-ray Spectrom.*, **23**(6), 261–266 (1994).
11. R.W. Ryon, H.E. Martz, J.M. Hernandez, J.J. Haskins, R.A. Day, J.M. Brase, 'X-ray Imaging: Status and Trends', *Adv. X-ray Anal.*, **31**, 35–52 (1987).
12. V. Valkovic, A. Markowics, N. Haselberger, 'Review of Recent Applications of Radioisotope Excited X-ray Fluorescence', *X-ray Spectrom.*, **22**(4), 199–207 (1993).
13. P. Standzenieks, E. Selin, 'Background Reduction of X-ray Fluorescence Spectra in a Secondary Target Energy Dispersive Spectrometer', *Nucl. Instrum. Methods*, **165**, 63–65 (1979).
14. J. Nissenbaum, A. Holzer, M. Roth, M. Schieber, 'X-ray Fluorescence Analysis of High Z Materials with Mercuric Iodide Room Temperature Detectors', *Adv. X-ray Anal.*, **24**, 303–309 (1980).
15. P. Jalas, H. Sipilä, S. Eränen, K. Leinonen, 'Near Room Temperature Energy-dispersive Detectors Processed on High-purity Silicon', *X-ray Spectrom.*, **24**(2), 63–69 (1995).
16. D.S. McGregor, H. Hermon, 'Room-temperature Compound Semiconductor Radiation Detectors', *Nucl. Instrum. Methods Phys. Res.*, **A 295**, 101–124 (1997).
17. B. Holynska, 'Sampling and Sample Preparation in EDXRS', *X-ray Spectrom.*, **22**(4), 192–198 (1993).
18. J. Boman, P. Standzenieks, E. Selin, 'Beam Profile Studies in an X-ray Fluorescence Spectrometer', *X-ray Spectrom.*, **20**(6), 337–341 (1991).
19. R.D. Giauque, F. Asaro, F.H. Stross, T.R. Hester, 'High Precision Non-destructive X-ray Fluorescence Method Applicable to Establishing the Provenance of Obsidian Artifacts', *X-ray Spectrom.*, **22**(1), 44–53 (1993).
20. R.D. Giauque, 'A Novel Method to Ascertain Sample Mass Thickness and Matrix Effects for X-ray Fluorescence Element Determinations', *X-ray Spectrom.*, **23**(4), 160–168 (1994).
21. H.A. van Sprang, M.H.J. Bekkers, 'Determination of Light Elements Using X-ray Spectrometry. Part 1 – Analytical Implications of Using Scattered Tube Lines', *X-ray Spectrom.*, **27**(1), 31–36 (1997).
22. B. Vekemans, K. Janssens, L. Vincze, F. Adams, P. Van Espen, 'Analysis of X-ray Spectra by Iterative Least Squares (AXIL): New Developments', *X-ray Spectrom.*, **23**(6), 278–285 (1994).
23. E.J. Suoninen, 'Selection of the Method', in *Surface Characterization, A User's Sourcebook*, eds. D. Brune, R. Hellborg, H.J. Whitlow, O. Hunderi, Wiley-VCH, Weinheim, 14–31, 1997.

Portable Systems for Energy-dispersive X-ray Fluorescence

Roberto Cesareo
Università di Sassari, Sassari, Italy

Giovanni E. Gigante
Università di Roma ''La Sapienza'', Rome, Italy

Alfredo Castellano
Università di Lecce, Lecce, Italy

Jan S. Iwanczyk
Photon Imaging Inc., Northridge, USA

1 Introduction 1
2 X-ray Sources 2
2.1 X-ray Tubes 2
2.2 Radioactive Sources 3
3 X-ray Detectors 4
4 Portable Energy-dispersive X-ray Fluorescence Systems 4
4.1 Portable Systems Based on Radioisotopic X-ray Sources and Scintillation or Gas Proportional Detectors 5
4.2 Portable Systems Based on Radioisotopic X-ray Sources and Thermoelectrically Cooled Semiconductor Detectors 6
4.3 Portable Systems Based on X-ray Tubes and Scintillation or Gas Proportional Counters 7
4.4 Portable Systems Based on X-ray Tubes and Thermoelectrically Cooled Semiconductor Detectors 7
5 Examples of Applications of Portable Energy-dispersive X-ray Fluorescence Systems 7
5.1 Analysis of Sulfur and Chlorine in Environmental Samples and in Works of Art 7
5.2 Analysis of Works of Art 7
5.3 Analysis of Alloys 8
5.4 Analysis of Lead in Paint 8
5.5 Analysis of Environmental Samples 9
5.6 Analysis of Soil 9
5.7 Underground Determination of Heavy Metal Contaminant Concentrations 9
5.8 The Mars Mission 10
5.9 General Applications 10
Abbreviations and Acronyms 11
Related Articles 11
References 11

Portable energy-dispersive X-ray fluorescence (EDXRF) spectrometers are becoming very popular in many fields for the on-site analysis of elements. This is mainly because EDXRF is a nondestructive, multielemental technique that is extremely well suited for the analysis of any material.

An EDXRF spectrometer mainly consists of an X- or γ-ray excitation source, an X-ray detector with electronics, and a pulse height analyzer. Recent technological developments have resulted in small, low-power, dedicated X-ray tubes, thermoelectrically cooled semiconductor detectors, and small pulse height analyzers. Therefore, completely portable EDXRF spectrometers are available that can be assembled on-site, having the size of a book and a weight ranging from as light as 500 g (using a radioactive source) to a few kilograms (using an X-ray tube). These spectrometers can be employed for on-site analysis in various fields, such as works of art, alloys, soil, environmental samples, forensic medicine, paper, waster materials, mineral ores and their products, or anywhere a portable apparatus would be required.

This paper reviews the present status of the development and application of EDXRF portable systems. The various components of a portable system are described: the radiation source, i.e. small, low-power, dedicated X-ray tubes or alternatively radioactive sources that emit X-rays or low-energy γ-rays; and X-ray detectors, i.e. scintillators, proportional gas counters and semiconductor detectors, with special emphasis on the recent thermoelectrically cooled X-ray detectors: Si-PIN (silicon positive–intrinsic–negative), Si-strip and CZT (cadmium–zinc–telluride).

Commercial systems are considered, and finally the most common and significant applications are described.

1 INTRODUCTION

EDXRF is a nondestructive technique that is extremely well suited for the analysis of any material. Its ease of use and adaptability have made EDXRF the most widely used method of nondestructive elemental analysis. Whether the sample is a solid, powder, liquid, thin film or coating, elements with concentrations ranging approximately from ppm to percentage levels can be determined

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

simultaneously with little or no sample preparation. EDXRF analysis simply consists of irradiating a sample with X- or γ-rays and of detecting the secondary X-ray spectrum emitted by the sample itself.[1] Each element emits a typical set of X-ray lines and is characterized by their energy. The X-ray spectrum is therefore composed (at least theoretically) of as many lines as the elements composing the sample.[1] The intensity of each line is proportional in some way to the concentration of the elements.

The sample also scatters (both coherently and incoherently) the incident radiation, generating lines due mainly to Rayleigh (coherent) and Compton (inelastic) scatterings.[1]

An apparatus for EDXRF analysis consists of an X- or γ-ray excitation source, an X-ray detector with electronics, and a multichannel analyzer. Portable ''laboratory'' systems can be assembled, and can be employed for on-site analysis in various fields such as analysis of works of art, alloys, environmental samples, forensic medicine samples, paper, waste materials, ferrous metal mineral ores and their products, etc.

Recent technological developments have resulted in both small-sized, low-power X-ray sources and small, thermoelectrically cooled semiconductor detectors.

This paper reviews the present state of development and application of EDXRF analysis, using miniature X-ray tubes and radioactive X- or γ-ray sources for excitation, and scintillation, gas and semiconductor detectors for detection. Special emphasis is given to thermoelectrically cooled X-ray detectors.

2 X-RAY SOURCES

For portable EDXRF equipment, X-ray tubes or radioactive sources may be employed.

An X-ray tube typically emits bremsstrahlung radiation with energy from zero (theoretically) to the value (in keV) of the maximum tube voltage, plus characteristic X-lines due to the anode material (see Table 1).

In this section the characteristics of X-ray tubes and radioisotopes emitting X- or γ-rays used for EDXRF analysis are described along with the detectors that can be used for portable systems.

2.1 X-ray Tubes

It should be noted that a 35–50 kV X-ray tube is adequate for the excitation of almost all elements of the periodic table, being able to excite K-lines of elements up to $Z = 60$ and L-lines of heavier elements.[1]

Table 1 Characteristics of small-size, portable X-ray tubes according to the element or elements to be analyzed

Element or groups of elements to be analyzed	Anode material	Anode voltage (kV)	Current[a] (mA)
Phosphorus, sulfur, chlorine	Calcium (K-lines)	6–8	0.1–1
	Palladium (L-lines) or silver (L-lines)	4–5	
Chlorine, potassium, calcium	Titanium (K-lines)	10	0.1–1
From potassium to yttrium (K-lines) and from cadmium to uranium (L-lines)	Molybdenum (K-lines)	30	0.1–1
From potassium to tin (K-lines) and from cadmium to uranium (L-lines)	Palladium (K-lines)	35	0.1–1
From potassium to tin (K-lines) and from terbium to uranium (L-lines)	Tungsten	35	0.1–1
From zinc to barium (K-lines) and from terbium to uranium (L-lines)	Tungsten	50	0.1–1

[a] The X-ray tube current determines the minimum detection limit (MDL). This parameter is also dependent on the geometry, the sample, the elements to be analyzed, the collimation, and so on. (For further details, see Cesareo.[1])

In the last few years, small-sized and low-powered X-ray tubes of various anode materials have been produced expressly for EDXRF analysis. Manufacturers include Oxford Analytical Systems Division, Scotts Valley, CA, USA; Hamamatsu Photonics KK, Hamamatsu City, Japan; and EIS-XRS, Rome, Italy. The low cost of these tubes allows the use of various dedicated X-ray tubes for the different problems.

Table 1 lists a series of different tubes, according to their characteristics (voltage, current, and anode material) that are optimized for various elements. Typical sizes of X-ray tubes for EDXRF analysis are:

- a length of about 10–15 cm;
- a diameter of about 3–5 cm;
- a window (of beryllium or glass) of 1.5–2 cm diameter;
- a weight of about 200–500 g.

Recently, a micro X-ray tube was designed and constructed explicitly to replace small, radioactive sources in portable gauging, fluorescence and imaging devices (X-Ray & Specialty Instruments Inc., Ann Arbor, MI, USA). This tube has a maximum voltage of 35 kV, a

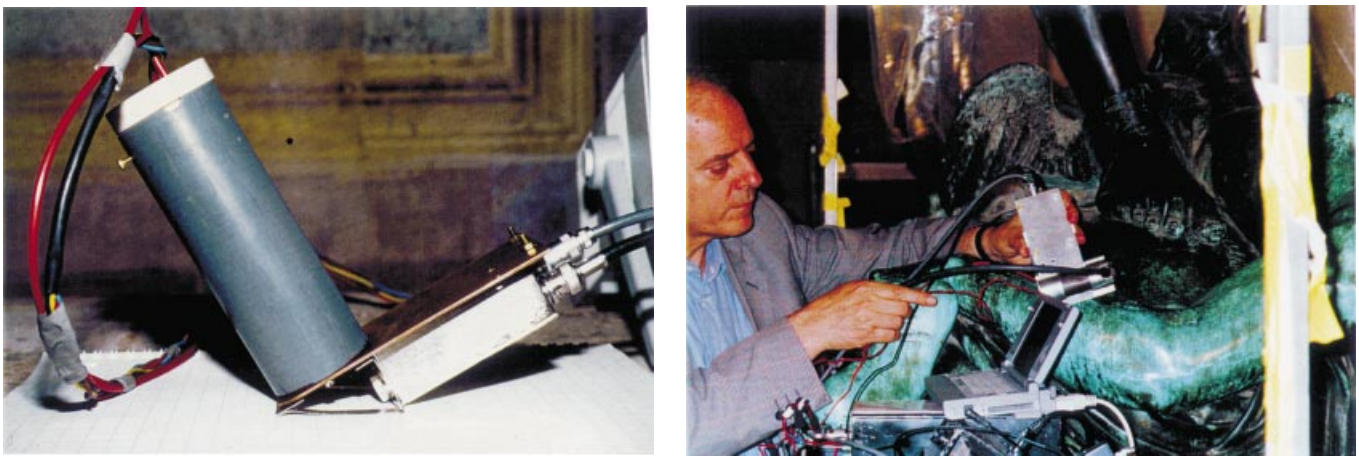

(a) (b)

Figure 1 EDXRF systems. (a) The Ca-anode Hamamatsu X-ray tube (on the left) and the thermoelectrically cooled AMPTEK Si-PIN (silicon positive–intrinsic–negative) detector (on the right), at approximately 30% life size. The tube is irradiating downwards, and the detector's window is visible on the left. (b) The complete EDXRF system used during the analysis of the bronze statue *Perseo*, by Benvenuto Cellini. The operator holds the Oxford W-anode X-ray tube (on the right) and the Si-PIN detector. The multichannel analyzer is also visible.

maximum power of 3 W, anode materials of Mo, Cu, W and a 0.25-mm diameter window of Be. It is a cylinder of length 8 cm and diameter 8 mm.

Photon output from an X-ray tube of the type described in Table 1 depends on many parameters, such as voltage, current and anode material. Typical outputs are 3–5 orders of magnitude larger than those of the radioactive sources. A typical small-size, low-cost X-ray tube for portable EDXRF systems is shown in Figure 1.

2.2 Radioactive Sources

α-, β-, X- and γ-ray sources may be employed for EDXRF analysis. Generally, they are characterized by their small size and low photon output compared with X-ray tubes and may be usefully employed for portable EDXRF systems.

α-ray sources are suited to the analysis of low atomic number elements. Those most used are ^{244}Cm, with a half-life of 17.8 years and emission of 5.76 and 5.81 MeV α particles, and ^{210}Po, with a half-life of 138 days and emission of 5.3 MeV α particles. Both are available from, for example, The Radiochemical Centre, Amersham, UK.

Table 2 Radioactive sources used for EDXRF analysis. (Available from The Radiochemical Centre, Amersham, UK.)

Radio-isotope	Half-life (years)	X- or γ-ray energy (keV)	Typical output (photons/s sr)
^{55}Fe	2.7	5.9–6.5 (Mn X-rays)	7×10^6
^{244}Cm	88	14.6–22 (U L X-rays)	
^{109}Cd	1.3	22–25 (Ag X-rays)	8×10^6
^{125}I	0.16	27–32 (Te K X-rays)	
^{241}Am	433	59.6 (γ-rays)	6×10^7
^{153}Gd	0.66	41, 48 (Eu K X-rays)	4×10^8
^{57}Co	0.74	122, 136 (γ-rays)	4×10^6

β-ray sources can also be employed both for direct EDXRF excitation of a sample and for producing bremsstrahlung radiation in a target to successively excite the sample. Typical sources of the first type are ^{22}Na, ^{85}Kr and ^{63}Ni, with half-lives of 2.6, 10.7 and 100 years, respectively, and emissions of 0.55, 0.67 and 0.066 MeV β^- particles. Typical sources of the second type are ^{147}Pm (with a zirconium target) and ^{3}H (with a titanium target), with half-lives of 2.6 and 12.4 years, respectively, and emitting photons up to 225 keV and 19 keV, respectively. Few radioactive sources emitting X- or γ-rays are used for EDXRF analysis; some are listed in Table 2, and typical radioactive sources and excitation geometries are shown in Figure 2. The sources detailed here are available from The Radiochemical Centre, Amersham, UK.

A range of elements that can be usefully analyzed with the various radioactive sources and X-ray tubes is shown in Figure 3.

Radioactive sources for EDXRF analysis are much smaller than X-ray tubes. However, the energy of these sources cannot be changed. Moreover, their output is often not adequate for an efficient excitation. Additionally, X-ray tube output can be monochromatized at various energies using proper filters. Finally, the costs of an EDXRF tube are currently comparable to those of a radioactive source.

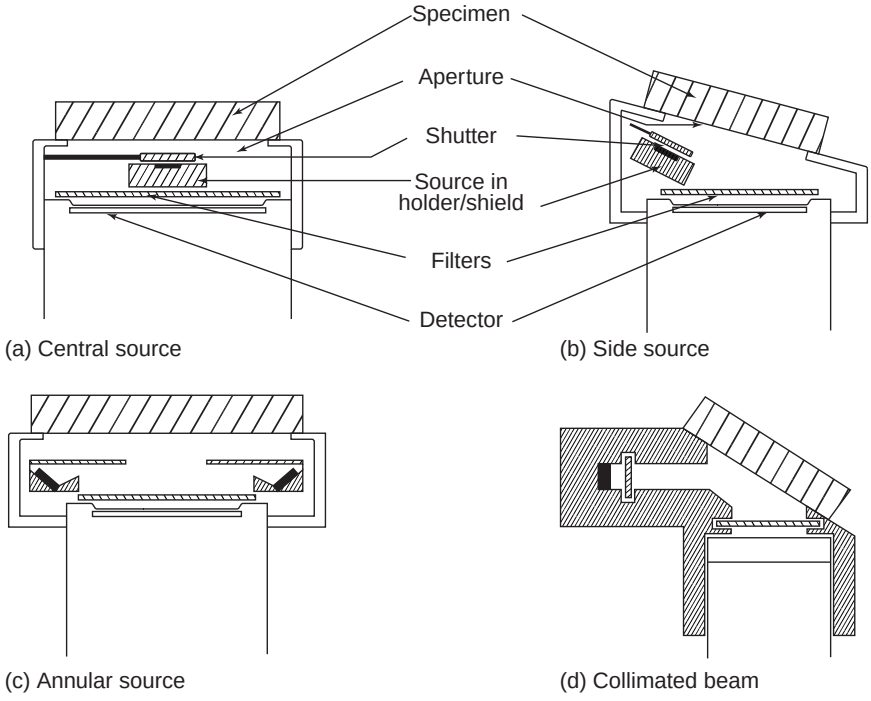

Figure 2 Typical source–specimen–detector arrangements, showing (a) the central source, (b) the side source and (c) the annular source disposition for radioactive sources. In the case of X-ray tubes (d), the collimated and in many cases monochromatized beam directly irradiates the sample (based on Rhodes[2]).

3 X-RAY DETECTORS

In the past, portable EDXRF analyzers traditionally involved the use of scintillation or proportional gas detectors, while laboratory systems typically used nitrogen-cooled semiconductor detectors, such as Si(Li) and HPGe (hyperpure germanium) (available from EG & G Ortec, Oak Ridge, TN, USA). X-ray scintillation detectors, such as NaI(Tl), CsI(Tl), are characterized by large areas but have very poor energy resolution, while proportional gas counters filled with neon, argon, krypton or xenon have better energy resolution[3] as well as a large useful area. Both are relatively cheap (approximately US $1000–2000).

In the last few years there has been a true revolution in the field of X-ray detectors. In fact, thermoelectrically cooled Si-PIN, $Cd_{1-x}Zn_xTe$ (CZT) and HgI_2 detectors[3] (available from AMPTEK Inc., Bedford, MA, USA) have been substituted for many applications, especially for portable apparatus, scintillation or proportional gas counters, or nitrogen-cooled Si(Li) and HPGe detectors.

The thermoelectrically cooled Si-PIN detector has a thickness typically of about 300 μm which makes this detector useful up to X-ray energies of 30 keV, and an energy resolution of about 200 eV at 5.9 keV,[3] which is only a little worse than that of a nitrogen-cooled Si(Li) detector. The CZT detector has a thickness of about 2 mm, and therefore an efficiency of 100% up to 150 keV, with an energy resolution of about 250 eV at 5.9 keV and about 1 keV at 60 keV (AMPTEK Inc., Bedford, MA, USA). The HgI_2 detector has a thickness of a few millimeters, sufficient for an efficiency of about 100% in the whole range of X-rays. It typically has an energy resolution of about 200 eV at 5.9 keV (XSIRIUS Inc., Camarillo, CA, USA).

Figure 4 shows the energy resolution versus energy of scintillation detectors, proportional gas counters, Si-PIN, CZT, Si(Li) and HPGe detectors in the X-ray energy range.

The only present limitation of the new thermoelectrically cooled Si-PIN and CZT detectors in their useful energy range is the reduced area, which limits the efficiency of the system. If high efficiency is required, or only one element is to be analyzed, larger-area detectors can be used, such as proportional gas counters or HPGe.

4 PORTABLE ENERGY-DISPERSIVE X-RAY FLUORESCENCE SYSTEMS

Portable systems are widely used in many circumstances to monitor one (or several) element(s), or to look at the

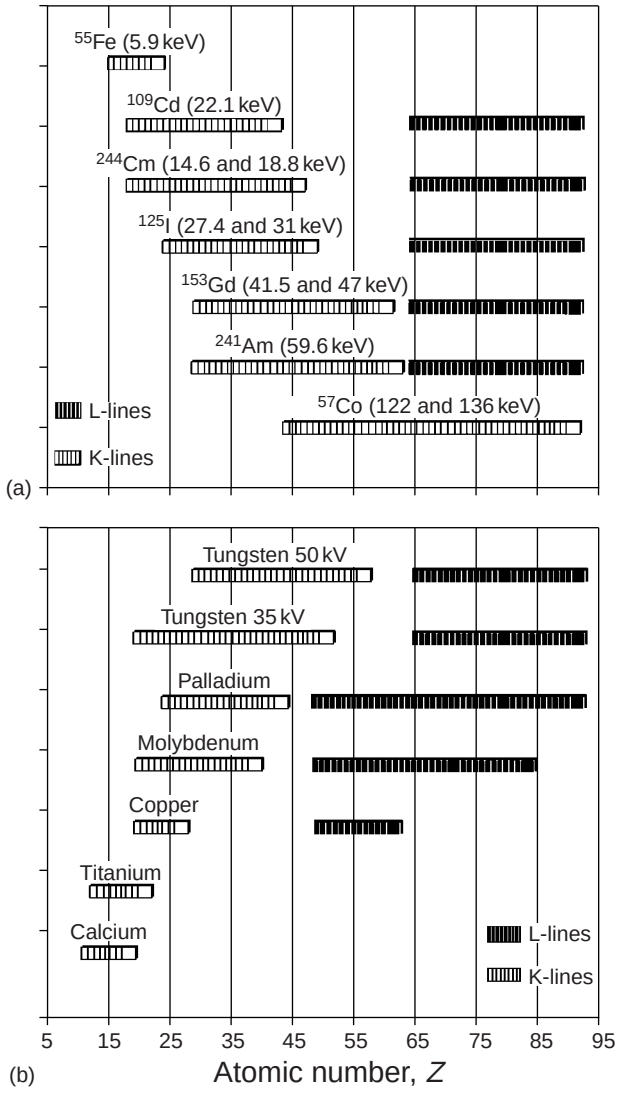

Figure 3 Range of elements that can be analyzed using (a) radioactive sources and (b) X-ray tubes with different anodes, showing excitation of K- and L-lines.

concentration of an element to indicate the quality of a particular technological process.

Several different combinations of X-ray sources and detectors can be constructed and are generally available on the market:

1. radioisotopic sources coupled to scintillation or gas proportional detectors;
2. radioisotopic sources coupled to thermoelectrically cooled semiconductor detectors;
3. X-ray tubes coupled to scintillation or gas proportional detectors;
4. X-ray tubes coupled to thermoelectrically cooled semiconductor detectors.

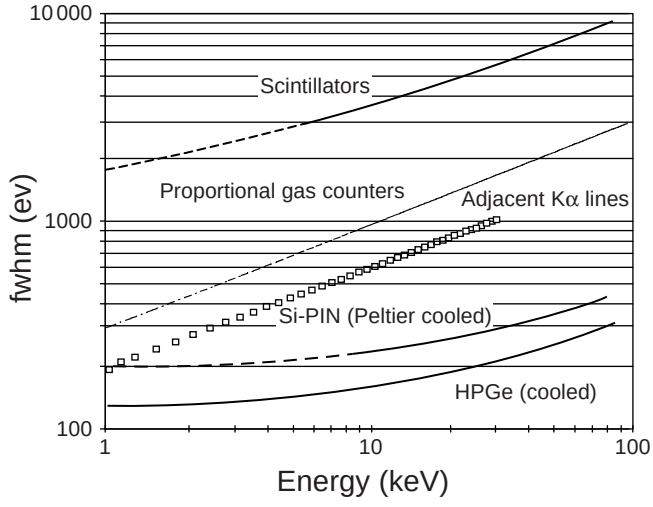

Figure 4 Energy resolution (in keV), as the full width at half-maximum (fwhm) of the X-ray peak, versus energy for different detectors. From top to bottom: a NaI(Tl) scintillator, a proportional gas counter, a thermoelectrically cooled semiconductor detector and a nitrogen-cooled semiconductor detector. The Kα energy difference of contiguous elements is also shown.

4.1 Portable Systems Based on Radioisotopic X-ray Sources and Scintillation or Gas Proportional Detectors

Scintillation detectors (typically NaI(Tl)) have a high efficiency, due to their large sensitive areas. However, the energy resolution is so poor that this detector can only be used with filters. Balanced filters placed between the sample and detector selectively transmit the X-lines of a given element, increasing the sensitivity of the system. A complete description of the method of balanced filters is given by Rhodes[4] and Watt.[5] For example consider filters of nickel and copper, having their K-shell absorption edges at 8.33 and 8.98 keV respectively, just below and above the energy of Kα X-rays of zinc (8.63 keV). Calibrating the masses per unit area of the filters to have the same transmission outside the zinc Kα X-rays energy, a ''pass band'' filtering system is produced, and the difference in the count rates using the two filters is proportional to the intensity of fluorescent zinc X-rays. Filter transmission is therefore different only in the energy window enclosing the zinc Kα X-rays. The technique is thus highly selective to the zinc Kα X-rays, but the sensitivity is poor compared with that obtained with detectors with sufficient energy resolution to isolate the X-rays of the element to be analyzed.

The balanced filters method is also useful when coupled to a gas proportional detector. These have the advantage over semiconductor detectors in that they have a larger sensitive area. Their energy resolution is poor,

although better than that of scintillation detectors, but not sufficient to separate Kα X-rays of contiguous elements (see Figure 4). The higher count rates obtainable with these detectors lead to shorter analysis times but an EDXRF system based on these detectors can only be employed in applications in which few elements are to be analyzed, the poor energy resolution is not a limitation, and possible overlaps from adjacent Z elements are not involved. There are various types of gas proportional detectors where the selection of the filling gas (neon, argon, krypton or xenon) is especially important. This selection should be made according to the excitation source and to the element or elements to be analyzed.[3]

There are many portable analyzers based on the use of radioactive excitation sources and scintillation or gas proportional detectors. These include the Outokumpu portable elemental analyzer, which is based on a gas proportional detector (Outokumpu Oy, Espoo, Finland); the Amdel system, which is based on the use of a scintillation detector (Australian Mineral Development Laboratories, Eastwood, Australia); and the Columbia Scientific (Austin, TX, USA) and Oxford Instruments (Abingdon, UK) systems, which use a gas proportional detector.

On-line analysis systems based on the use of radioisotope X-ray sources and scintillation or proportional gas counters have been listed by Watt,[15] who also described their applications, such as in determining the metal content of mineral slurries and mineral powder, sulfur in oil, calcium in raw cement mix, iron and chromium in ore, and tin in galvanizing solutions.

4.2 Portable Systems Based on Radioisotopic X-ray Sources and Thermoelectrically Cooled Semiconductor Detectors

Thermoelectrically cooled semiconductor detectors (Si-PIN, CZT, and HgI_2) generally have sufficient energy resolution to resolve the fluorescent X-rays of adjacent X-rays (see Figure 4). There are only a few cases of overlap where resolution of X-ray peaks is difficult (Cu Kβ and Zn Kα, As Kα and Pb Lα, and a few others). The small sensitive area of these detectors is their main limitation, but it may be compensated for by the large solid angle between the sample and detector.

As observed in section 3, the Si-PIN detector is the best one currently available up to about 25–30 keV. At higher energies the intrinsic efficiency is poor, due to the limited thickness of these detectors (about 300 μm). In this region the CZT detector can be employed, or the HgI_2 detector. The first has a poorer energy resolution, and both have the problem of ''escape peaks'', due to Cd and Zn in the first case and to Hg and I in the second.

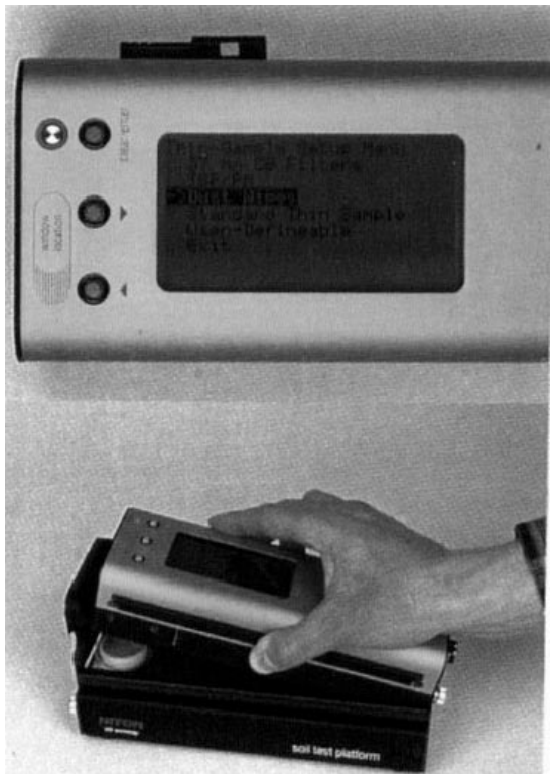

Figure 5 The NITON XRF multi-element XRF spectrum analyzer.

Figure 6 The Metallurgist XRF mobile metal analyzer from TN Technologies, Inc.

Portable systems using radioisotopic sources coupled to thermoelectrically cooled semiconductor detectors are manufactured and supplied by, among others, Niton Co. (Bedford, MA, USA) and TN Technologies (Round Rock, TX, USA) The NITON X-ray fluorescence (XRF) spectrum analyzers are equipped with a ^{109}Cd source (10–40 mCi) and a Si-PIN detector (Figure 5), while the TN Technologies systems use ^{55}Fe, ^{109}Cd or ^{241}Am sources with a HgI_2 thermoelectrically cooled detector (Figure 6).

4.3 Portable Systems Based on X-ray Tubes and Scintillation or Gas Proportional Counters

This combination has never been used, because combination of a high-intensity source with a high-efficiency detector would be illogical.

4.4 Portable Systems Based on X-ray Tubes and Thermoelectrically Cooled Semiconductor Detectors

The coupling of a high-intensity flexible source as the X-ray tube with a high-resolution detector (for example Si-PIN) in our opinion is the best possible current system for in situ XRF analysis. The combination of the appropriate X-ray tube (in terms of anode, high voltage, current, collimation) coupled to the corresponding X-ray detector (material, thickness of the material and of the Be-window) and sample geometry gives the best results.

Portable systems based on the use of an X-ray tube coupled to a thermoelectrically cooled semiconductor detector are manufactured and supplied by EIS-XRS (Rome, Italy), as one example. Various combinations of tubes (high voltage, anode material, current) and detectors (Si-PIN, CZT) are available, and should be selected according to the planned application.

A portable EDXRF system based on the use of an X-ray tube and a thermoelectrically or nitrogen-cooled Si detector is produced by Kevex Spectrace (model Spectrace 9000) and specifically dedicated to soil analysis.

5 EXAMPLES OF APPLICATIONS OF PORTABLE ENERGY-DISPERSIVE X-RAY FLUORESCENCE SYSTEMS

5.1 Analysis of Sulfur and Chlorine in Environmental Samples and in Works of Art

Sulfur in coal, oil and gasoline is a source of pollution, and strict environmental controls are now in place to limit sulfur release into the atmosphere. In the past, sulfur was freely released and formed deposits on monuments, frescoes and so on, causing damage. Chlorine is present as NaCl in the air coming from the sea and its deposits cause corrosion on monuments.

A portable system has been constructed by EIS-XRS, for the analysis of sulfur and chlorine in solids and liquids.[6] This system is based on a Ca-anode X-ray tube, working at 6–8 kV, coupled to a thin Be-window Si-PIN detector. The Ca-tube emits 3.7 keV X-rays that excite S and Cl X-rays with high efficiency without directly exciting X-rays of Ca, which is by far the most common element in lapideous monuments and frescoes (Figure 7). A Pd-anode Oxford X-ray tube can also be employed for the same purpose. In this case, the tube works at 4–5 kV, emitting Pd L-lines at 2.84 and 3 keV.

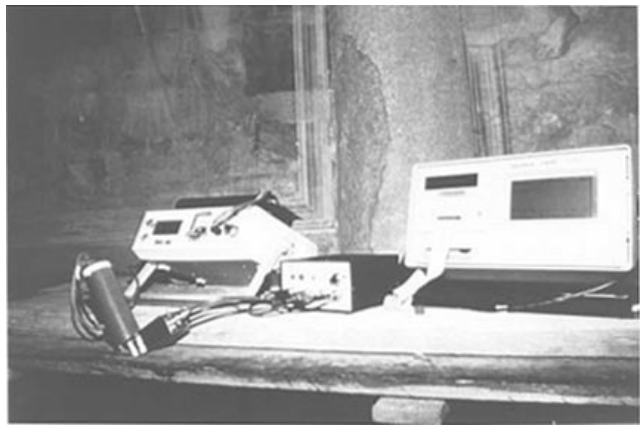

Figure 7 EDXRF system employed by the authors for the analysis of sulfur and chlorine in frescoes. From left to right: the measuring head (shown in Figure 1), the X-ray tube bias supply, the detector bias supply and amplifier, and the multichannel analyzer.

MDLs of about 0.02% and 0.1% are obtained in the analysis of S and Cl in calcareous surfaces respectively, while much better results are obtained for oil and gasoline.

5.2 Analysis of Works of Art

In the analysis of works of art (monuments, statues, frescoes, paintings, alloys, etc.) the portability of the system is a mandatory requirement, because in nearly all cases it is impossible to take the work of art to the laboratory. However, with a portable EDXRF apparatus, it is generally possible to take it to the work of art, to the museum, to the excavation, and so on.

Examining the analytical problems related to the study of works of art, one can deduce that in the great majority of the cases the elements to be analyzed are in the range between $Z = 16$ (sulfur) and $Z = 50$ (tin). This range corresponds to an excitation energy between 3 and 35 keV. High Z elements (gold and lead) are included in the same energy range, because the L-lines are more usefully analyzed.

Portable systems have been assembled by us since 1972, starting with the analysis of paintings and ancient alloys (gold, bronzes, and brasses). A ^{147}Pm radioactive source was employed, which is characterized by a useful bremsstrahlung spectrum ranging from 15 to 45 keV approximately, plus a xenon gas proportional detector and a single-channel analyzer.[7]

Recently, a new portable apparatus was manufactured to our design by EIS-XRF (Rome, Italy). It includes a 40-kV X-ray tube and a Si-PIN or CZT detector. The

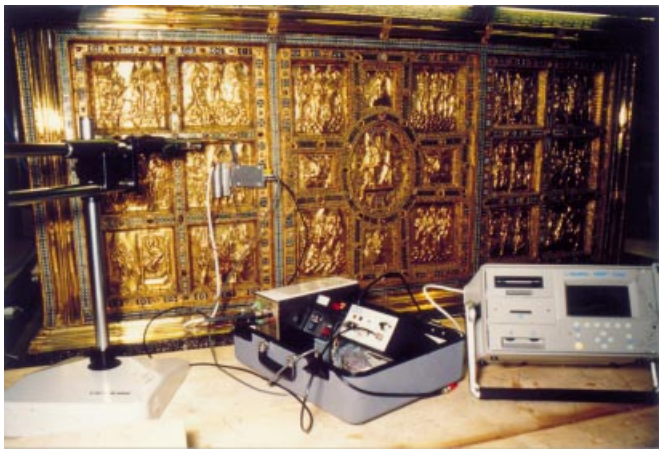

Figure 8 EDXRF system employed by the authors for the analysis of the golden altar of the basilica of St. Ambrogio, Milan. From left to right: the measuring head, comprising an Oxford W-anode X-ray tube and a Si-PIN detector, the bias supplies for both tube and detector, and the multichannel analyzer.

first detector may be employed up to about 20 keV, the second at higher energies.

The system described above has been used for the analysis of gold samples[8,9] (Figure 8), paintings[10] and bronzes[11] (Figure 9).

5.3 Analysis of Alloys

As observed in section 5.2, bronzes, brasses, gold and silver alloys have been analyzed using an EDXRF apparatus. Rapid sorting of alloys is required in many areas of the metal industry, such as fabrication, inventory control and the sorting of scrap. Some common alloy groups include nickel and copper alloys, stainless steels, special steels, precious metals, and so on. The main requirements of analysis equipment for alloy identification are portability, speed and reliability of identification. For the alloys, an accurate quantitative determination is generally required, and therefore sufficient standard samples are needed to calibrate the apparatus. Piorek and Rhodes[12] have shown that a 111 MBq ^{109}Cd source and a high-resolution proportional counter are adequate to identify many alloys.

Currently, better results can be obtained by using a portable system of the type described in section 5.2, which is adequate for the analysis of Ni, Cu, Fe, Cr/Mo, Au, Ag and other alloys.

TN Technologies (Round Rock, TX, USA) has developed portable EDXRF systems (Metallurgist XR) for the analysis of alloys. It uses 45 mCi ^{55}Fe or 5 mCi ^{109}Cd radioactive sources and a high-resolution thermoelectrically cooled HgI_2 detector, and may be employed for the analysis of low-alloy steel, Fe, Cr, Ni, Co alloys, brasses and bronzes, and Al alloys. The weight of the measuring head (source plus detector) is 1.1 kg (see Figure 6).

5.4 Analysis of Lead in Paint

One of the more serious public health hazards, which particularly affects children, is related to lead-based poisoning from the paint found in many old houses.

Figure 9 EDXRF system employed by the authors for the analysis of the famous bronze statue *Perseo* by Benvenuto Cellini. From top to bottom: the measuring head (X-ray tube horizontal, Si-PIN detector vertical), the multichannel analyzer (PC) and the bias supplies.

The estimated number of apartments in New York City alone that may be affected by this hazard is greater than 300 000,[13] and they are mainly occupied by low-income families. Recent legislation in the US provides specific requirements for new inspection procedures in federally funded housing programs as well as for the disclosure of information and inspections during the sale and transfer of all private residential houses and apartment units constructed prior to 1978. (Allowable lead content in paint was significantly reduced after 1978.) Inspection companies and state and federal agencies, which are charged with the responsibility of supervising compliance with this legislation, employ EDXRF hand-held instruments for on-site detection of lead concentration levels in paint. An abatement plan to eliminate paint poisoning hazards can be required for concentrations of lead in paint equal to or exceeding 1 mg cm^{-2}. Some states have adopted even more stringent standards.

This application[13] demonstrates the necessity of using K X-rays in the determination of lead concentration. The lead L X-rays can only serve as supplementary information, due to their strong attenuation by paint overlays. The high energy of the Pb K X-ray lines (75, 85 keV) imposes additional restrictions on the choice of a detector and an excitation source. High atomic number compound semiconductor detectors such as CdTe, CZT or HgI_2 are favored in this application because of their good detection efficiency, high-energy resolution, and operation near room temperature.[14,15] Also, excitation radioactive sources such as ^{57}Co or ^{109}Cd are more practical for hand-held instruments than X-ray generators, which are difficult to miniaturize for these relatively high X-ray energies. Table 3 lists portable instruments used to measure lead in paint and outlines their main characteristics.

Table 3 Portable instruments used to measure lead concentrations in paint

Instrument	Detector type	Excitation source
LPA-1 RMD[a]	CdTe (conduction counter type)	^{57}Co (10 mCi)
MAP-4 SCITEC	CdZnTe (conduction counter type)	^{57}Co
LeadStar Advanced Detectors	CdTe (PIN type)	^{57}Co (15 mCi)
700 series NITON	Si-PIN	^{109}Cd (10 mCi)
XL 309 NITON	Si-PIN/CdZnTe (2 detectors)	^{109}Cd (10 mCi)
TN Technologies	HgI_2	^{109}Cd (40 mCi)
Warrington Corp.	Scintillation detectors	^{57}Co (10 mCi)
PGT/Outokumpu[b]	Proportional counters	$^{109}Cd/^{57}Co$
HNU	Si[Li] (cryogenic cooling)	^{57}Co (10 mCi)

[a] RMD, radiation monitoring device.
[b] PGT, Princeton Gamma-Tech.

5.5 Analysis of Environmental Samples

The concentration of air particulates is determined by collecting a large volume of air through a filter paper, which is then submitted to EDXRF analysis. Rhodes[16] showed that a laboratory system combining an appropriate radioactive source and a nitrogen-cooled semiconductor detector allows the determination of elemental concentrations of many trace elements. A sensitivity of about 0.03 g cm^{-2} on the filter, corresponding to 6 ng m^{-3} of air over a 10-min period, was obtained. A portable system composed of an X-ray tube and a Si-PIN detector will give approximately the same results, because the smaller area of the detector is compensated for by a higher-intensity source.

On-line and in situ analysis of industrial waste is also important for environmental purposes. A dedicated portable system can be easily constructed, according to the element or elements to be analyzed and their concentration. In the case of only one element, electronics composed of a single-channel analyzer and a digital ratemeter can be constructed with a continuous output of data.

5.6 Analysis of Soil

In situ analysis of metals in soil, mainly S, K, Ca, Ti, Mn, Fe, Zn and others, can also be carried out with a portable system composed of a Mo-anode low-voltage X-ray tube and a thin Be-window Si-PIN detector.

Additionally the analysis of elements contained in fertilizers, and in substances introduced into the soil (P, S, Cl, K and others) can be carried out with the same portable apparatus.

Portable EDXRF systems for soil analysis are described in sections 4.2 (NITON XL-700) and 4.4 (Kevex Spectrace 9000).

5.7 Underground Determination of Heavy Metal Contaminant Concentrations

The site characterization and analysis penetrometer (SCAPS) XRF sensor provides XRF spectra on soil in situ down to depths of over 45 m. Its purpose is to measure heavy metal contaminant concentrations directly in the soil. Cone penetrometry, which was developed almost 100 years ago to explore soil strata for railroad bed construction, is an excellent method for bringing a variety of sensors into contact with subsurface

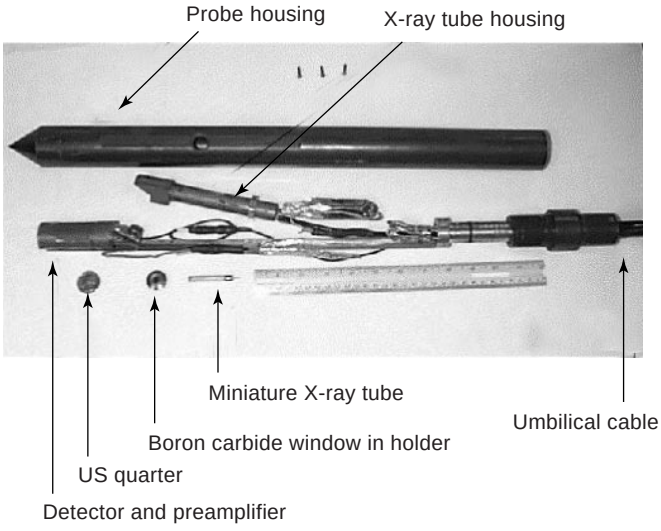

Figure 10 System for underground XRF analysis of heavy metals.

soils. Unlike other field-portable XRF instruments, the sensor uses an electrically excited X-ray tube as a source and is protected by a rugged but X-ray-transparent window. These features provide an instrument with excellent capabilities and which will withstand direct contact with underground soils. The probe and methods were developed by W.T. Elam of the US Naval Research Laboratory in conjunction with J.V. Gilfrich and K.R. Hudson of SFA, Inc., and B. McDonald of the University of Toledo.[17]

The XRF metal sensors consist of three subsystems: the actual below-ground probe (Figure 10), the umbilical cable and the above-ground electronics package. The probe contains the X-ray source (up to 30 kV), detector (Si-PIN) and preamp, appropriate X-ray optics, the mounting system and the rugged X-ray window. The above-ground electronics package contains the X-ray tube power supply with safety interlocks, the driver electronics for the detector and the multichannel analyzer.

5.8 The Mars Mission

On July 4, 1997, a portable EDXRF apparatus constructed for analyzing rock landed on Mars. It employs a ^{244}Cm α radioactive source and an AMPTEK XR-100 X-ray detector. The apparatus is shown in Figure 11, and a typical X-ray spectrum is shown in Figure 12.

5.9 General Applications

A portable apparatus (Spectrace 9000 Field Portable XRF Analyzer) developed by TN Technologies (Round Rock, TX, USA) may be applied to analysis of elements from sulfur to uranium, with concentrations from parts per million to percentage levels. This apparatus is characterized by various radioactive sources (^{55}Fe, ^{109}Cd and ^{241}Am) according to the elements to be analyzed, by a high-resolution HgI_2 detector, and software. The analyzer is a 2000-multichannel analyzer with software that automatically executes a selected menu. Each data acquisition procedure includes selection of source,

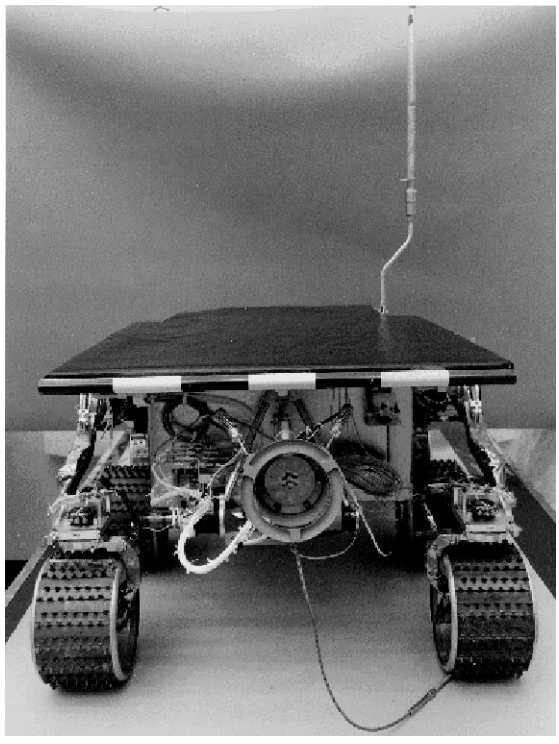

Figure 11 Mars *Pathfinder* rover with, at the right, the EDXRF-apparatus for the analysis of Martian rock, including the α-source and the AMPTEK detector and electronics.

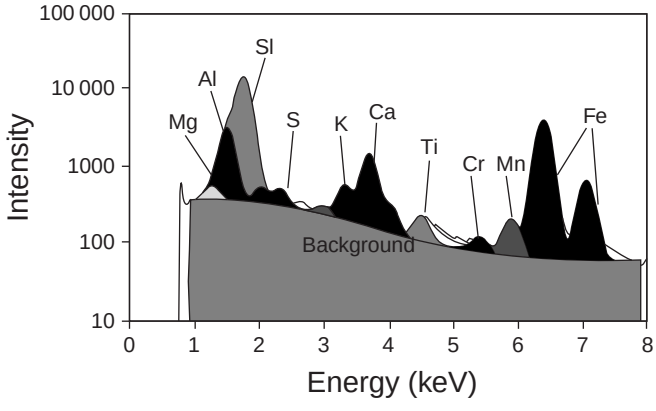

Figure 12 Typical XRF spectrum of "Barnacle Bill" Martian rock obtained with the equipment shown in Figure 11.

acquisition time, computation method and display of analysis results.

ABBREVIATIONS AND ACRONYMS

CZT	Cadmium–Zinc–Telluride
EDXRF	Energy-dispersive X-ray Fluorescence
fwhm	Full Width at Half-maximum
HPGe	Hyperpure Germanium
MDL	Minimum Detection Limit
PGT	Princeton Gamma-Tech
RMD	Radiation Monitoring Device
SCAPS	Site Characterization and Analysis Penetrometer
Si-PIN	Silicon Positive–Intrinsic–Negative
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Heavy Metals Analysis in Seawater and Brines

Environment: Water and Waste cont'd **(Volume 4)**
Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sampling Considerations for Biomonitoring • Soil Instrumental Methods • Soil Sampling for the Characterization of Hazardous Waste Sites • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in • Nickel Ore and Metals Analysis • X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Nuclear Methods **(Volume 14)**
PIXE (Particle-induced X-ray Emission)

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Sample Preparation for X-ray Fluorescence Analysis • X-ray Techniques: Overview

General Articles **(Volume 15)**
Archaeological Chemical Analysis • Quantitative Spectroscopic Calibration

REFERENCES

1. R. Cesareo, 'Photon Induced X-ray Emission' in *Nuclear Analytical Techniques in Medicine*, ed. R. Cesareo, Elsevier, Amsterdam, 19–121, 1988.
2. J.R. Rhodes, 'Design and Application of X-ray Emission Analyzers Using Radioisotope X-ray or Gamma Ray Sources', Special Technical Publication 485, American Society for Testing and Materials, Philadelphia, 243, 1971.
3. R. Cesareo, 'X-rays: Interaction with Matter, Production, Detection', *La Rivista del Nuovo Cimento* (a Review Section of *Il Nuovo Cimento*), Bologna, August 2000.
4. J.R. Rhodes, in *Non-destructive Testing*, ed. H.B. Egerton, Oxford University Press, London, Chapter 14, 1969.
5. J.S. Watt, in *Radioisotope X-ray Analysis*, eds. A. Markovicz, R. van Grieken, M. Dekker, New York, Chapter 7, 1992.
6. A. Castellano, R. Cesareo, 'A Portable Instrument for Energy-dispersive X-ray Fluorescence Analysis of Sulfur', *Nucl. Instrum. Methods Phys. Res., Sect. B*, **129**, 281–283 (1997).
7. R. Cesareo, F.W. von Hase, 'Non-destructive Radioisotope XRF Analysis of Early Etruscan Gold Objects', *Kerntechnik*, **15**, 565–569 (1973).
8. R. Cesareo, A. Castellano, C. Fiorini, G.E. Gigante, J.S. Iwanczyk, A. Longoni, J.A. Pantazis, J.L. Pena Chapa, M.A. Rosales, 'Thermoelectrically Cooled Semiconductor Detectors for Portable Energy Dispersive X-ray Fluorescence Equipments', Proc. SPIE, San Diego, CA July 31–August 1, 1997, Vol. 3115, 274–283, 1997.
9. R. Cesareo, A. Castellano, G.E. Gigante, M. Marabelli, S. Bandera, A. Longoni, C. Fiorini, 'EDXRF Analysis of the Golden Altar of S. Ambrogio in Milan', *J. Cultural Heritage*, in press.
10. G.E. Gigante, C. Maltese, S.Rinaldi, S. Sciuti, *Archaeometry '90*, Proceedings of the 7th International Symposium on Archaeometry, Heidelberg, April 2–6, 1990, eds. E. Pernicka, G.A. Wagner, Birkhäuser Verlag, Basel, 255–264, 1991.
11. R. Cesareo, G.E. Gigante, A. Castellano, M.A. Rosales, M. Aliphat, F. De la Fuente, J.J. Meitin, A. Mendoza, J.S. Iwanczyk, J.A. Pantazis, 'Portable Systems for Energy Dispersive X-ray Fluorescence Analysis of Works of Art', *J. Trace Microprobe Techn.*, **14**(4), 711–725 (1996).
12. S. Piorek, J.R. Rhodes, 'Application of a Microprocessor Based Portable Analyzer to Rapid Non-destructive Alloy Identification', Proc. ISA-1986, Houston, 1355–1368, 1986.
13. G.R. Laurer, T. Keneip, R.E. Albert, F.S. Kent, 'In-situ Determination of Lead on Painted Surfaces for the Prevention of Childhood Lead Poisoning', in *Application of Low Energy X- and Gamma Rays*, ed. C.A. Ziegler, Gordon and Breach, New York, 289–301, 1971.

14. J.L. Campbell, L.A. Cross, 'Portable Instrumentation for the Determination of Lead in Painted Surfaces', *Bull. Environ. Contam. Toxicol.*, **16**(4), 469–473 (1976).
15. Y.J. Wang, J.S. Iwanczyk, W.R. Graham, 'Evaluation of HgI_2 Detectors for Lead Detection in Paint', *IEEE Trans. Nucl. Sci.*, **40**(4), 846–850 (1993).
16. J.R. Rhodes, 'Energy-dispersive X-ray Spectrometry for Multielement Pollution Analysis', *IEEE Trans. Nucl. Sci.*, **NS-21**, 608–617 (1974).
17. W.T. Elam, J.W. Adams, K.R. Hudson, B. McDonald, J.V. Gilfrich, *Field Anal. Chem. Technol.*, **2**, 97–102 (1998).

Sample Preparation for X-ray Fluorescence Analysis

Jasna Injuk and René Van Grieken
University of Antwerp, Antwerp, Belgium

1 Introduction 1
2 Some Considerations that the X-ray Fluorescence Analyst Should be Aware of 1
3 Probabilistic Sampling Procedures 2
4 Solid Samples 2
4.1 Metallic Specimens 2
4.2 Powder Specimens 4
4.3 Fused Specimens 5
5 Liquid Specimens 5
5.1 Direct Methods 5
5.2 Preconcentration Methods 5
6 Biological Samples 7
7 Atmospheric Particles 8
8 Standards 8
Abbreviations and Acronyms 8
Related Articles 8
References 9

This article is a comprehensive review of sample preparation methods for various kinds of samples prior to X-ray measurement. Not only are commonly used methods presented but also included are some specific techniques for certain kinds of materials, with emphasis on more recent applications and developments.

1 INTRODUCTION

Sample preparation generally involves taking an aliquot from a submitted sample and then preparing a specimen for analysis from this aliquot, in such a way as to ensure that the relatively small volume of sample actually analyzed by the X-ray beam is representative of the total specimen. Only if the analyzed volume of the prepared specimen and the sample are of the same composition will the analysis be considered a meaningful one.

For the vast majority of sampling problems in the X-ray laboratory, only a few statistical principles are needed: representative sample, randomization, mean, variance and *t*-distribution. For this reason, a statistical reference book should be available always in the laboratory. It is important to replicate analyses because this serves as a check on instrumental reproducibility. Replication means drawing several samples from the bulk and preparing them separately. It is only by replication that the analyst can know that the drawn sample is representative. The number of replications necessary is always a question to which the answer depends on the importance of the analytical results and the required accuracy of analysis. When new procedures are being tested, much more replication is necessary than with well-established procedures. There are statistical methods (involving routine statistical testing, such as Student's *t*-test) for determining the number of samples required for a specific degree of accuracy.

Almost any sort of sample, liquid or solid, that will not damage the spectrometer can be analyzed by X-ray fluorescence (XRF). Samples are usually brought to the instrument, whereas field-portable XRF spectrometers are commonly used for various environmental evaluations.

The literature on the above topic is rather extensive and a more complete description of the sample preparation procedures can be found in some earlier books.[1,2] In this article, only the most common procedures are summarized. Specialized techniques for total reflection X-ray fluorescence (TXRF) and energy-dispersive X-ray fluorescence (EDXRF) are covered more extensively in other articles in this encyclopedia.

2 SOME CONSIDERATIONS THAT THE X-RAY FLUORESCENCE ANALYST SHOULD BE AWARE OF

In XRF the sample preparation procedure is at least as important as an analytical technique and it strongly influences the final quantitative result. Any method of specimen preparation must give specimens that are reproducible and, for a certain calibration range, have similar physical properties (e.g. comparable mass attenuation coefficient, density, particle size and particle homogeneity). Improper selection of a sample from the bulk or improper sample preparation can cause large systematic errors to be introduced; therefore the following principles should be considered as a rule:

- before any sample preparation is selected, one should think of the purpose of analysis;
- specimen preparation should be avoided as much as possible;
- a method should be selected that is cheap and rapid and does not introduce additional errors or contamination;

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

- if possible, preference should be given to simple physical sample preparation methods such as drying, freeze-drying, homogenizing, pulverizing, cutting;
- an adequately prepared specimen must be representative of the material and homogeneous and have a flat, smooth surface;
- all standards and unknown samples must be presented for analysis in a reproducible and identical way;
- standards must have similar physical properties, such as mass attenuation coefficient, density, particle size, homogeneity, etc.;
- contamination control and blank specimen preparation should be readily implemented;
- loss of analyte elements (particularly halogens) should be considered;
- careful and accurate records must be kept of samples received for analysis;
- routine statistical testing should be a part of all laboratory procedures and goes under the category of quality assurance;
- accuracy should be controlled by using Standard Reference Materials.

3 PROBABILISTIC SAMPLING PROCEDURES

It is important to emphasize that good sampling is absolutely crucial to the meaningfulness of the ultimate analysis. Because every step is linked materially to all other steps, it is essential to plan in detail the complete sampling program, starting from field sampling and through to the specimen proceedings in a balance room, before the first test portion from the bulk sample is taken.

A crucial early step in a sampling procedure is splitting of a sample; in its simplest form the sample is divided into two equal (in every sense) parts, one of which is archived in case it is needed later. Splitting a granular material is a dynamic process that usually involves pouring. Pouring always causes segregation, so it is vital that the segregated particles always appear in each split in their original proportions, otherwise an extraction error occurs. The oldest method of splitting is coning and quartering. This is attractive for samples ranging from tens of kilograms to a few grams. Firstly, a granular sample is carefully poured vertically onto a clean surface in order to form a cone. The dynamics of correct pouring will ensure that coarser fragments are uniformly concentrated around the periphery of the cone and that the finer ones are uniformly distributed in the center. Then, the cone is quartered by precisely dividing it into four equal sectors vertically. This quartering must be done without generating a delimitation error.

For large samples containing tens of kilograms of material, alternate shoveling is preferred to coning and quartering. In this method, the primary sample is poured in a symmetrical pile onto a clean surface. The original pile is made into two nonoverlapping piles by shoveling from the original one. It is important that the shovelfuls all be the same size and that each new pile contains the same number of transfers. In practice, it is difficult to come out exactly even; therefore the size of the shovel must be small relative to the size of the pile. For further reduction in the size of the primary sample, one of the two new piles is chosen at random and the process is repeated. In this way, the sample is reduced to a size that can be reduced further prior to final grinding.

Another shoveling or scooping method that allows probabilistic sampling is the linear Japan cake. The sample from a previous stage is spread in a linear pile from which a single scoop can be extracted randomly in one motion. The pile is constructed by pouring back and forth along its design length. Several samples can be extracted in this way. The pile should be at least 25 times longer than the width of the scoop. Any material that the scoop pushes out of the pile has to be recovered to the sample. Pitard[3] suggests that this method can be used even with small samples in the laboratory balance room.

The riffle splitter (or Jones splitter) is one of the best and simplest devices for splitting. It is a box containing an even number of adjacent chutes alternately inclined in opposite directions. When the primary sample is poured properly into the top of the splitter box, half of the sample passes through one set of chutes and half through the other. If the sample passes equally through all the chutes, no significant delimitation or extraction errors will be generated.

Sectorial splitters allow the sample to be split initially into more than two parts. In such a device normally several funnels are assembled together, with common sides forming the radii of a circle and their third sides forming the perimeter of the same circle.

4 SOLID SAMPLES

4.1 Metallic Specimens

XRF is one of the most common techniques for the analysis of samples in metal production and therefore the preparation of such a specimen should be fast, simple and reproducible. The most universally employed sample preparation procedures are cutting, milling and casting.

XRF analysis of metal samples mostly involves a relatively thin surface layer of the investigated specimen; therefore it must be representative of the bulk sample

structure. The major surface preparation methods are machining (milling, turning), mechanical grinding, and polishing and etching. In any employed surface preparation technique, special care must be taken to ensure that no additional surface contamination is introduced. This care is particularly important in the case of a relatively soft metal such as aluminum, where particles of the grinding agent may penetrate to the sample surface during surface finishing. Selection of a suitable abrasive is sometimes very difficult, particularly when elements like Si, Al or Fe need to be determined. For instance, SiC and Al_2O_3 are very effective abrasives, but both contain elements of interest. In such cases electrolytic polishing or etching is recommended. After each grinding or polishing step it is important to ensure that the surfaces are kept clean; hence all traces of lubricant, cutting fluid or finger marks must be removed by cleaning with isopropyl alcohol prior to examination. Occasionally the surface of several metals is affected by corrosion, which progresses with time; therefore the best results are always obtained when analyzing a specimen without much delay after surface preparation.

When the surface of the specimen is too rough and common methods like polishing and milling are not effective enough, the samples can be prepared by pressing them into pellets with a hydraulic press at several hundreds of megapascals. With such a procedure most of the turnings are compacted into satisfactory smooth surfaces.

Selection of the most suitable surface preparation technique depends on the individual characteristics of an alloy or group of alloys. For example, hard alloys that are resistant to cutting and abrasion, such as cast iron, can be ground and pelletized prior to analysis. On the other hand, if the sample is composed of soft, malleable, multiphase alloys, the smearing effects of softer components like Pb and Al can cause problems. The soft elements will be enriched onto the surface and the fluorescent intensities of the softer phase increase, whereas those of the harder ones decrease. In such cases, when the sample surface is not representative of the bulk composition, electrolytic polishing or etching should be employed. For some archeological materials, the etching procedure might be too excessive because many inclusions are dissolved during etching. For bronze alloys the recommended etching reagent is a solution of alcoholic ferric chloride (120 mL of C_2H_5OH, 30 mL of HCl and 10 g of $FeCl_3$), whereas for brass alloys a solution of aqueous ammonium persulfate (100 mL of H_2O and 10 g of $(NH_4)_2S_2O_8$) is preferred. Ancient metals etch very fast; only a 5–10 s etching time is required.

A number of metals and alloys that are otherwise very difficult to deal with can be brought into solution in aqua regia (a mixture of one part concentrated HNO_3 with three parts concentrated HCl) with gentle heating to 60 °C. Aqua regia is used to attack a number of metals (e.g. Pd, Pt, Sn, Ge) and alloys, including steels, alloys of Mo, Rh, Ir and Ru, and high-temperature alloys. The strength of this combination belongs to the formation of chloro complexes during the reaction, as well as to the catalytic effect of Cl_2 and NOCl. The effectiveness additionally can be improved if the solution is allowed to stand for 10–20 min before heating. Iron and vanadium ores and crude phosphates can be dissolved by this mixture, as well as some sulfides (e.g. pyrites and copper sulfide ores), although a part of the sulfur is lost as H_2S. Both HNO_3 and HCl are often used together in other proportions. So-called inverted aqua regia (mixture 3 : 1) is used to oxidize sulfur and pyrites. Another aggressive decomposition reagent is the combination HNO_3–HF. Here, the complexing effect of the fluoride ion is utilized and with this combination a number of metals (e.g. Si, Nb, Ta, Zr, Hf, W, Ti) and alloys (e.g. NbSn, AlCr, CuSi, CaSi) can be dissolved. However, some elements may be lost during the digestion step: elements like Se, Hg and Sn form volatile components with various kinds of acids and their combinations, samples treated with hot mixtures of HNO_3–HF–$HClO_4$ lose Se and Cr completely; and Hg, As, Ge, Te, Re, Os and Ru show particular losses. Another, rather new, technique is remelting of the sample. The original ferroalloy is ground to a grain size of 2 mm with a steel disk mill, followed by a W-mill, after previous crushing with a jaw breaker. Subsequently, the grained particles are mixed fairly with a portion of Fe (15 g of FeMo alloy plus 25 g of pure Fe) and this final mixture is melted inside an induction furnace. Afterwards, the melted metal is centrifuged and cast in an appropriate mold prior to analysis. Calibration samples can be obtained easily in the same way. This method is particularly useful for the determination of Cr in MoFe and NbFe alloys because, due to their inhomogeneity and brittleness, direct analysis by XRF is inappropriate.[4,5]

The analysis of semiconductors is often carried out by TXRF or synchrotron radiation X-ray fluorescence (SRXRF). However, there are also some interesting applications with XRF, such as the determination of Ni implantation in Si wafers, which can be measured directly with XRF simply by placing the wafer in the spectrometer. More difficult is the preparation of calibration standards for this application, which can be obtained in an unconventional way by spiking gelatin-containing Ni standard on parts of the Si wafer and drying prior to analysis.[6]

Irregularly shaped metallic specimens are prepared for analysis by embedding the piece in a special wax resin (e.g. acrylic resin and methyl acrylic resin). This block can be polished to appropriate surface smoothness by abrasives such as SiC and diamond paste, prior to analysis.

4.2 Powder Specimens

4.2.1 General

A common technique for geological, industrial and biological materials is the preparation of powders and pellets. Powdered specimens are prepared when the original sample is too heterogeneous for direct analysis or too brittle to form self-supporting disks or when a suitable surface finishing is not possible. The XRF analysis of powder samples requires particles in a uniform size range and with a small enough grain size to yield chemical homogeneity without absorption problems. Crushing, pulverizing and milling procedures may be used to achieve these goals. Additives (see section 4.2.2) are sometimes useful for proper grinding of powders; in general, 2–10% additive to aid in the grinding, blending and briquetting process is sufficient for most materials. Once the sample has been crushed, it is necessary to separate out the fraction that is too coarse for further treatment. Sieves are used for this step. The powder specimen has to fit into the specimen chamber of the X-ray spectrometer. It is desirable to have the specimen in the shape of a round, flat disk. For this purpose, presses and dies are applied. Another possibility is to present them directly to the spectrometer as loose powders, packed in cells or spread out on film materials (the so-called ''slurry'' technique). The slurry technique works for water-insoluble materials. A water slurry is prepared out of a few milligrams of powder and a few milliliters of water. A turbulent suspension is made, followed by rapid filtration through a Nucleopore filter for example. This method results in fairly uniform thin layers suitable for XRF analysis.

Processing large bulk samples poses the most problems. The challenge is to pulverize the sample, to produce fine particles without contamination or fractionation, to homogenize the product and to split it into representative aliquots for study.

4.2.2 Grinding

One of the principal considerations in sample preparation methods is a reduction of sample particle size. For a routine trace XRF analysis a particle size of less than 62 μm is commonly accepted. Reduction of the sample particle size is commonly accomplished by grinding. Various methods for grinding samples are available. For routine work, manual grinding using a pestle and mortar (agate, corundum, mullite) is traditionally used, whereas for finer (1–10 μm) and controlled particle size a variety of mills are commercially available. When dealing with extremely small amounts of sample, hand grinding is the best option, although it is tedious and time-consuming. Materials that are difficult to treat, such as rocks and ores, can be prepared much faster and more efficiently after the addition of sodium stearate. It is found that manual grinding of inorganic aluminum silicate, for example, by using a pestle and mortar reduces the particle size to about 38 μm, which is acceptable even for X-ray diffraction (XRD) methods. In the case of coal samples prepared as a powder (5 g of sample ground in a wet chemical rotary swing mill together with 1 g of boric acid binder and 100 mg of sodium stearate), a grinding time of about 6 min will reduce the particle size to 50 μm. For materials that may be damaged or undergo alteration during size reduction, grinding is usually carried out under alcohol using a pestle and mortar. Grinding in liquid nitrogen (in an SPEX freezer/mill for example) may also be effective for materials like polymers and plastics. Mechanical grinders can also be used under dry and wet conditions. When grinding samples that are mixtures, different physical properties may lead to significantly different results with the different phases. Experience has shown that all grinding tools present a source of contamination; therefore the material should be selected carefully for a particular application. Agate, for instance, introduces traces of SiO_2, Mg and Ca to the specimen. This might be of less interest for geological materials, but for biological materials these blank values represent a major source of error.

4.2.3 Pelletizing

Pelletizing of the powder samples is required to reduce surface effects and to yield a better precision than the loose powders. In general, provided that the powder particles are less than 50 μm in diameter (300 mesh), the sample should be pelletized at 6–8 tons cm^{-2}. If the self-bonding properties of the powder are good, low pressures of perhaps up to 1–4 tons cm^{-2} can be employed. High-pressure pelletizing in a die, or directly in a sample cup, often results in fracture of a pellet following removal of pressure from the die. It is sometimes necessary to add a binder before pelletizing to help form stable pellets and to prevent caking of the sample at the die surface. Binders are normally composed of light elements or organic materials and scattered X-rays increase the background of the spectrum. This effect can become important in the determination of trace elements in the low-atomic-number region, therefore the choice of the binding agent must be made with special care. It must be free from significant contamination and must have low absorption. As well as having good binder properties, it must not interfere with the elements being measured. The most useful binders are starch, cellulose, lucite, polyvinyl, urea, boric acid, graphite, etc. Liquid binders such as ethyl alcohol or

diethyl ether can be mixed with the sample manually, whereas powdered binders are recommended to be mixed mechanically to form a homogeneous mixture. A recommended procedure is the addition of 2–10% binder to the sample. Further stability can be achieved by spraying the pellet with a 1% solution of Formvar in chloroform.

4.3 Fused Specimens

Samples that do not go into solution easily or tend to remain heterogeneous after grinding and pelletizing are often treated by the technique of flux fusion. Probably the most effective way of preparing a homogeneous powder sample is the borax fusion method.[7] In principle this involves fusion of the sample with an excess of sodium or lithium tetraborate and casting into a solid bead. Chemical reaction in the melt converts the phases present in the sample into glass-like borates, giving a homogeneous bead of dimensions ideal for direct placement in the spectrometer. Manual application of the technique is rather time-consuming, but a number of automated and semi-automated borax-bead-making machines are commercially available. The critical stage in this method is to control the ratio of sample to fusion mixture, because this governs several factors: the speed and degree of the chemical reaction, the final mass absorption coefficient, and the actual dilution factor applied to the analyte element. By using fusion aids (such as iodides and peroxides) and high-atomic-number absorbers (such as barium or lanthanum salts) as part of the fusion mixture, these factors can be controlled. Claisse's original method proposed a sodium tetraborate/sample ratio of 100 : 1.[7] In 1962 Rose et al.[8] suggested the use of lithium tetraborate in a ratio of 4 : 1. A critical review on this subject was done by Bower and Valentine.[9] The actual fusion reaction can be performed at 800–1000 °C in a crucible made of Pt, Ni or SiO_2. All of these materials suffer from the disadvantage that the melt tends to wet the sides of the crucible and it is impossible to achieve complete recovery of the fused mixture. By using graphite crucibles this problem can be partially overcome, but the best characteristics are achieved in a crucible made of Pt and 3% Au.

The most severe disadvantages to the fusion techniques are the time and material costs involved and the dilution effect of the sample, which makes it often difficult or even impossible to determine trace elements in the specimen. For this reason the low-dilution fusion technique was developed, where the flux/sample ratio is 2 : 1. Mostly $LiBO_2$ is applied as a flux, because it is more reactive and forms fluxes with higher fluidity. Many useful recipes for fusions are given in a recent book.[2]

5 LIQUID SPECIMENS

5.1 Direct Methods

In most spectrometers, liquids can be analyzed as such, even under a pressure that is slightly below atmospheric. Special care should be taken that the analyzed surface remains reproducible and that no gas bubbles are formed under the influence of the X-rays, e.g. in acid solutions. For the analysis of liquid specimens, special cells can be made of stainless steel, polyethylene or polytetrafluorethylene (Teflon®). These cells are covered by Mylar or Kapton foils with a thickness of 3–6 μm. Sometimes, microporous films are applied. Because the most conventional spectrometers irradiate the sample from below, the support film attenuates the signal from the elements of lower atomic number, as well as introducing a significant blank. Absorption by air also poses a problem for measurements in the low-atomic-number region. Liquids are well suited for reducing matrix effects and for adding various amounts of internal standards. However, in contrast to standards prepared by fusion or pelletizing, liquid standards deteriorate with time due to adsorption and precipitation effects.

However, the major drawback of liquid analysis is the high background level due to X-ray scattering. This makes it difficult, or sometimes even impossible, to determine light elements at low concentrations. In general, direct analysis of aqueous samples leads to limits of detection in the part-per-million range, which is not satisfactory for most natural water analysis applications. Consequently, it is necessary to remove the matrix to improve detection limits.

5.2 Preconcentration Methods

In many cases (e.g. environmental aqueous samples), some preconcentration technique is to be applied to the water samples before analysis. The sample resulting from preconcentration is usually in the form of a thin sample and, as such, meets the thin-film criterion. Consequently, the conversion of XRF data to element concentration is a rather simple procedure. Many preconcentration methods for the analysis of water by X-ray spectrometric techniques have been proposed so far. A comprehensive overview of this subject, dating back to 1982, was given by Van Grieken.[10]

For samples of low salinity or hardness, preconcentration can be accomplished easily by simple evaporation. To achieve detection limits at the part-per-billion level, evaporation of about 100 mL of water is necessary. A large water sample also can be freeze-dried, and the evaporation residue can be pelletized, possibly after mixing with organic binder to reduce matrix effect variations.

Freeze-drying of 250 mL of waste water on 100 mg of graphite followed by grinding and pelletizing of the residue can lead to detection limits of a few micrograms per liter. In the evaporation residue, all nonvolatile elements are collected quantitatively and the risk of contamination is minimal.[11]

Unfortunately, taking samples to dryness causes some experimental problems, such as fractional crystallization, splashing, etc. For these particular reasons, preconcentration by evaporation has not found a great deal of application except when combined with some special techniques like TXRF. However, sometimes simple evaporation of the solution straight onto confined-spot filter paper can be sufficient, as is employed successfully in the so-called microdroplet sample preparation method. The advantage of the method is a rapid sample treatment with a small amount of sample for X-ray measurement. On the other hand, the disadvantage of the method is that the elemental distribution through a filter paper is heavily affected by the drying condition of the filter paper, due to condensation of the analyte on the surface of the filter material. Such experimental errors, derived from the drying conditions of the filter paper, can be minimized by drying at 10 °C and by using an internal standard. In addition, a new thin-film sample support was designed to accumulate a small amount of sample solution (about 50 μL) into a small spot. By a special surface treatment, this newly developed film (AP2™, Process Analytics) creates approximately an 1.5-mm hydrophilic spot in the center of a hydrophobic field, resulting in a low scattering of X-rays from the matrix of the sample. The reported precision of the method is 2–9% for transition metals at the 50 ppb concentration level and with an analyzing time of less than 15 min.[12] In contrast to commonly used sample support films (Prolene, Kapton, Mylar), this film has a high degree of temperature resistance and is therefore suitable for analysis of high-boiling acids and organics.

A variety of methods can be used for the preconcentration of aqueous samples. These include extraction precipitation, co-precipitation and the use of ion-exchange membranes and ion-exchange resins.

Low concentrations in solution can be concentrated with ion-exchange resins and eluted and evaporated onto filter-paper disks. From suitable ion-exchange material, preconcentration factors of up to 4×10^4 can be achieved by using around 100 mg of resin.[13] Alternatively, the ion-exchange resin itself can be used as a sample support, either by pelletizing the exchange resin or by use of ion-exchange filter papers. The former method has proved to be extremely useful for the determination of metals in very dilute solutions. The procedure involves shaking the solution with a known amount of resin for several minutes, followed by filtering, drying and pelletizing. By this method, for example, 0.1 ppm of Au in solution can be determined within less than 15 min.

Another effective method involves concentration of a litre of water, for example, on a cation-exchange resin column and direct analysis of the resin by EDXRF. With this procedure, trace metal contamination of a few parts per billion is easily determined in drinking water. Anion-exchange filters may be used to concentrate or selectively remove a particular component. An example is the determination of Cr^{3+} and Cr^{6+}. In acid medium, an anion-exchange filter paper will collect Cr^{6+}, whereas Cr^{3+} passes through. Subsequently, the toxic Cr^{6+} can be analyzed directly on the filter paper.

Another useful method for multielement trace analysis of aqueous samples is co-precipitation. Various co-precipitation agents have been proposed in the literature for XRF. By far the most appropriate are sodium diethyldithiocarbamates and ammonium pyrrolidine dithiocarbamate (APDC), or a combination of both. The carbamates are particularly attractive because of the low solubility of their metal chelates. The recommended procedure involves 100–500 mL of the water sample (adjusted to pH 3.5) and the addition of 2 mL of a 1% aqueous solution of APDC. After 40–60 min of stirring, the precipitate is filtered onto a 0.22- or 0.4-μm pore-size Nucleopore membrane and the loaded filter is analyzed after drying. The detection limits can be below $1\ \mu g\ L^{-1}$ for a suite of cations.

It should be emphasized that chemical preconcentration inherently holds some risks for sample contamination and may be inaccurate because of the possibly different behavior of the different chemical species of an element. Naturally occurring organic material might interfere with the co-precipitation reaction. Bringing the sample to pH 3.5 constitues a compromise in this respect: much of the organic material will be destroyed already and adjusting the sample to a lower pH does imply more contamination by the acid. Furthermore, many elements are not collected in the proposed way, e.g. alkali and alkaline-earth ions and Mn^{2+} (except at very high pH). Also, some species of elements, such as Cr^{3+}, might escape precipitation. Additionally, preconcentration can be labor-intensive and time-consuming.

Solid materials in aqueous samples must be filtered out through, for example, Nucleopore or Millipore membranes with 0.4-μm pore size (this is the conventional limit between ''dissolved'' and ''particulate'' matter in environmental waters), dried, and the filter analyzed as a solid.

The XRF analysis of liquids increases in importance from year to year.[14] Numerous applications have been reported recently, such as on-line process control in industry using a continuous flow of sample through a sample chamber. It appears, also, to be a well-suited

technique for various kinds of oils: from highly raffinated greases down to petroleum products and used oils. Even pitches and asphalt samples can be analyzed when applying a special preparation technique – the so-called ''hot casting'' procedure – which completely eliminates the use of mixing and grinding equipment and awkward clean-ups. The pitch or asphalt is directly ''hot cast'' into a plastic sample cup fitted with a thick Mylar window. After cooling, the window is removed carefully from the surface of the sample to expose a mirror-shiny and perfectly flat surface ready for XRF analysis.[2] Furthermore, determination of U and Ce in nuclear fuel can be achieved easily if the sample is dissolved in HNO_3 solution and subsequently microdroplets of this solution are spotted onto filter paper.[15]

6 BIOLOGICAL SAMPLES

Solid biological materials are generally heterogeneous. Therefore, drying, powdering, homogenizing and homogeneity testing might be necessary before preparing the samples for measurement. Various subsequent treatments of biological materials include lyophilization, ashing and wet digestion. The digestion method should be optimized and the recovery of the elements and precision of the procedure should be thoroughly tested.

Specimen preparation for XRF of plants, leaves and vegetation is simple and fast. The representative samples are collected (approximately 5 g), dried at 85 °C, pulverized and pressed into pellets without any contamination. Usually 10–20 elements can be determined quantitatively in a matter of minutes.[16]

Most human and animal tissues are very soft and hence they must first be stabilized and strengthened before sections can be cut. One possibility is to freeze-dry the tissue and pulverize it at liquid nitrogen temperatures. A small amount of that powder can be fixed to a thin membrane with a solution of 1% polystyrene in benzene and measured. Frozen organic tissue can be cut with a microtome, and if the sections are not too thin they can be dried and irradiated as self-supporting targets. To quantify the results, it is sufficient to add an internal standard before the analysis and to determine the dry mass of the sections after analysis.

The interest in clinical applications increases from year to year. In vivo analysis, particularly for Pb, is a major issue for XRF. The direct determination of Pb in bone gives an indication of the ongoing accumulation, whereas conventional blood studies only reflect the recent exposures. Other heavy metals of interest are As, Cd, Hg and U. Further in vivo investigations dealt with the concentration of various elements in different organs, mainly in the field of kidney, liver and lung research. However, the sensitivity and detection limits of the in vivo applications are still rather poor.

Other applications of XRF analysis include hair, teeth, nails, biological liquids (e.g. blood and serum), drugs and medicines. Blood samples, for example, can be pipetted in 100-μL portions onto filter paper. Serum samples can be treated almost in the same way but instead of filter paper, polypropylene foil should be selected as a carrier. To avoid losses of volatile components, urine and blood samples need to be dried in a refrigerator at temperatures of about 2–4 °C. Detection limits of between 1200 $\mu g\,L^{-1}$ for Ca in urine and 50 $\mu g\,L^{-1}$ for Sr in blood are achieved with such treatment.[17,18]

As for the preconcentration of traces from water samples, various kinds of co-precipitation or ion-exchange methods have been proposed for biological samples. For example, a very low concentration of Cr (0.3 $ng\,mL^{-1}$) in plasma could be determined by complexation with APDC and extraction with methyl isobutyl ketone. After evaporation of organic solvent, the residue is dissolved in acid and deposited on a thin polycarbonate foil.[19] Hair samples (0.1 g) can be digested with a 1 : 5 mixture of $HClO_4$ and HNO_3 in a Teflon® bomb and the trace elements are co-precipitated using APDC at pH 3–4. In this way, elements such as Fe, Ni, Cu, Zn and Pb could be determined down to 0.4 $\mu g\,g^{-1}$. The same limits of detection could be achieved if, for example, 1 g of hair sample is pelletized with 0.1 g of CaO and 20 g of cellulose.[20,21]

The wet digestion methods, by microwave or high-pressure ashing, are very effective. With a mixture of HNO_3, H_2O_2 and HF most organic material can be attacked. Even more effective is a combination of HNO_3 and HF because silicates in organic materials are also dissolved. Solvents $HClO_4$ and H_2SO_4 should not be mixed due to the possibility of explosion when heating. Care should be taken to work, whenever possible, in closed conditions to avoid contamination and/or losses of elements. The combination of microwave and pressure digestion allows the quantitative recovery of elements that may be volatilized in open digests (like Ge, Se, Hg, Pt, Os and Ir), as well as the recovery of rare trace metals (Pt, Os and Ir) to the part-per-billion level in solid samples, even in the extremely difficult digestion of organic material. For a basic understanding of microwave acid digestion theory, including safety guidelines and dissolution methods for geological, metallurgical, botanical, biological, food and other samples, we recommend the professional reference book edited by Kingston and Jassie.[22]

7 ATMOSPHERIC PARTICLES

XRF is frequently invoked for trace analysis in air pollution studies. In the air, trace elements are almost exclusively in the particulate phase at typical concentrations of 50–500 ng m^{-3}. By simply drawing a large volume of air through a filter, large preconcentration factors are easily achieved. Adequately loaded filters with air particulate matter are presented directly to the XRF unit. Unless the filters are extremely loaded, no correction for X-ray absorption in the aerosol material is usually necessary.

The part of the filter with a sample actually analyzed by XRF has an area of only a few square centimeters and hence it must be representative for the entire sample. Therefore, a critical step is the selection of an appropriate filter material. Teflon® and polycarbonate (Nucleopore) filters are ideal because of their high purity and because they are surface collectors. Whatman-41 cellulose filters have been used widely because of their low cost, but they are rather thick (around 9 mg cm^{-2} versus 1.1 mg cm^{-2} for Nucleopore), which leads to more X-ray scatter and higher detection limits. They also partially collect particulate matter within their depth, so that X-ray absorption corrections become more complicated. Glass-fiber filters should be avoided in any case owing to their high inorganic impurity. High-purity quartz-fiber filters have recently been proposed. Before selecting a filter material for a particular application, the blank count of the filter or background level of the material to be analyzed must be determined because all filters contain various elements as major, minor and trace constituents.

Regarding aerosol analysis, several factors that affect X-ray intensity must be considered: attenuation of X-rays within the individual aerosol particle is only important for particles larger than a few micrometers, but accumulation of particles can have an effect (in the case of heavily loaded filters) and the filter itself can attenuate incident and emerging X-rays (if particles are not collected on the surface). An appropriate correction involves knowledge of the particle size distribution and selection of an adequate sampling time and an appropriate filter material. For example, if thin filters like Nucleopore or Millipore are selected for the collection, the absorption effects in the filter itself are small and often negligible for X-rays above 1 keV, because normally all particles should be retained at the filter surface. However, for light elements such as Si, P, S and Cl the absorption effect needs to be corrected by using a suitable correction procedure.

Besides conventional XRF, TXRF also nowadays is applied more frequently for aerosol analysis.[23] To take advantage of the low detection limits and to reduce the matrix effects originating from the collection material, filter samples need to be digested. Normally, aerosol-loaded filters are dissolved with suprapure concentrated (70%) HNO_3 and HF acids in a high-pressure digestion vessel. A standard is normally added prior to digestion. With this procedure, element concentrations of from 0.2 ng m^{-3} for Cu and 1 ng m^{-3} for Mn are easily determined.

8 STANDARDS

X-ray spectrometry is essentially a comparative method of analysis; therefore it is necessary that all standards and unknowns have a high degree of similarity (approximately the same matrix composition) and that they are presented to the spectrometer in a reproducible and identical way. Standards may be a Certified Reference Material or a sample itself, but analyzed by another analytical technique. Many specimens in use may also be prepared in the laboratory. They are usually composed of one or more elements at different levels of concentration within different matrices. For standards, it is essential to provide high count rates for the analyte element. Sometimes use is made of an internal standard. Standards can be liquid, gas or solid and can be pure, mixed powder, natural or synthetic materials.

Today, there are 3000 reference standards listed in catalogues and hundreds more in preparation. These standards include ores, metals and alloys, rocks, minerals, waste products and dusts collected by electrostatic precipitators. Environmental problems have increased the demand for standards for tailings, industrial wastes and isotopic materials.

ABBREVIATIONS AND ACRONYMS

APDC	Ammonium Pyrrolidine Dithiocarbamate
EDXRF	Energy-dispersive X-ray Fluorescence
SRXRF	Synchrotron Radiation X-ray Fluorescence
TXRF	Total Reflection X-ray Fluorescence
XRD	X-ray Diffraction
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 3)**
Biological Samples in Environmental Analysis: Preparation and Cleanup • Heavy Metals Analysis in Seawater and Brines • Industrial Waste Dumps, Sampling and Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Proton-induced X-ray Emission in Environmental Analysis • Quality Assurance in Environmental Analysis • Sample Preparation for Elemental Analysis of Biological Samples in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Sampling Considerations for Biomonitoring • Soil Sampling for the Characterization of Hazardous Waste Sites • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Food **(Volume 5)**
Fluorescence Spectroscopy in Food Analysis • Food Analysis Techniques: Introduction • Sample Preparation for Food Analysis, General

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Industrial Hygiene **(Volume 6)**
Metals in Blood and Urine: Biological Monitoring for Worker Exposure

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Nuclear Methods **(Volume 14)**
PIXE (Particle-induced X-ray Emission)

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis • X-ray Techniques: Overview

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. J. Injuk, R. Van Grieken, 'Sample Preparation for XRF', in *Handbook of X-ray Spectrometry*, eds. R. Van Grieken, A. Markowicz, Marcel Dekker, New York, 657–692, 1993.
2. V.E. Buhrke, R. Jenkins, D.K. Smith, *A Practical Guide for the Preparation of Specimens for X-ray Fluorescence and X-ray Diffraction Analysis*, John Wiley & Sons, New York, 1998.
3. F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*, CRC Press, Boca Raton, FL, 1993.
4. A.G. Coedo, M.T. Dorado, C.J. Rivero, I.G. Cobo, 'X-ray Fluorescence Analysis of Ferroalloys: Development of Methods for the Preparation of Test and Calibration Samples', *Analyst*, **119**, 421–426 (1994).
5. I.G. Cobo, M.T. Dorado, A.G. Coedo, 'Remelting Technique for X-ray Fluorescence Analaysis of Main Constituent in Ferromolybdenum and Ferroniobium', *J. Trace Microprobe Techn.*, **13**, 431–441 (1995).
6. H. Bubert, P. Burba, R. Klockenkämper, A. Schönborn, M. Wielunski, 'Dose Determination of Nickel Implantations in Silicon Wafers', *Fresenius' J. Anal. Chem.*, **341**, 245–247 (1991).
7. R. Tertian, F. Claisse, *Principles of Quantitative X-ray Fluorescence Analysis*, Wiley–Heyden, London, 1982.
8. H. Rose, I. Adler, F.J. Flanagan, *US Geol. Surv. Prof. Pap.*, **450B**, 1–80 (1960).
9. N.W. Bower, G. Valentine, 'Critical Comparison of Sample Preparation Methods for Major and Trace Determination Using XRF', *X-ray Spectrom.*, **15**, 73–78 (1986).
10. R. Van Grieken, 'Preconcentration Methods for the Analysis of Water by X-ray Spectrometric Techniques', *Anal. Chim. Acta*, **143**, 3–34 (1982).
11. J. Smits, J. Nelissen, R. Van Grieken, 'Comparison of Preconcentration Procedures for Trace Metals in Natural Waters', *Anal. Chim. Acta*, **111**, 215–226 (1979).
12. A.F. Wilson, D.C. Turner, A.A. Robbins, *Process Analytics*, Company brochure, Orem, UT, 1998.
13. D.E. Leyden, T.A. Patterson, J.J. Alberts, 'Preconcentration and XRF Determination of Cu, Ni and Zn in Seawater', *Anal. Chem.*, **47**, 733–735 (1975).
14. A.T. Ellis, M. Holmes, P. Kregsamer, P.J. Potts, C. Streli, M. West, P. Wobrauschek, 'Atomic Spectrometry Update – X-ray Fluorescence Spectrometry', *J. Anal. At. Spectrom.*, **13**, 209R–232R (1998).
15. I. Hanif, J. Hanif, S.M. Hasany, M.Z. Iqbal, 'Studies of Uranium–Cerium Admixture Solutions by EDXRF Spectroscopy', *X-ray Spectrom.*, **24**, 298–306 (1995).
16. V. Kocman, T.E. Peel, G.H. Tomlinson, 'Rapid Analysis of Macro- and Micronutrients in Leaves and Vegetation by Automated X-ray Fluorescence Spectrometry – a Case Study of Acid Rain Affected Forest', *Commun. Soil Sci. Plant Anal.*, **22**(19/20), 2063–2075 (1991).
17. N.T. Hong, H.V. Ha, 'Application of EDXRF to the Determination of Lead and Other Trace Elements in the Body Fluids of Industrial Workers in Vietnam', *X-ray Spectrom.*, **25**, 3–14 (1996).
18. A. Viksna, G. Mwiruki, D. Jagner, E. Selin, 'Intercomparison Between Energy-dispersive X-ray Fluorescence and Stripping Potentiometry for the Determination of Copper Levels in Human Serum', *X-ray Spectrom.*, **24**, 76–80 (1995).
19. M. Simonoff, Y. Labador, G.N. Simonoff, M.R. Boisseau, M.F. Lorient Rodaut, 'PIXE Determination of Calcium in Red Blood Cells', *J. Radioanal. Nucl. Chem. Lett.*, **94**, 297–310 (1985).

20. M.A. Eltayeb, R. Van Grieken, 'Preconcentration and XRF Determination of Heavy Metals in Hair from Sudanese Populations', *J. Radioanal. Nucl. Chem. Art.*, **131**, 329–340 (1989).
21. M.A. Eltayeb, R. Van Grieken, 'Iron, Copper, Zinc and Lead in Hair from Sudanese Populations of Different Age Groups', *Sci. Total Environ.*, **95**, 157–165 (1990).
22. H.M. Kingston, L.B. Jassie (eds.), *Introduction to Microwave Sample Preparation, Theory and Practice*, ACS Reference Book Series, American Chemical Society, Washington, DC, 1988.
23. J. Injuk, R. Van Grieken, 'Optimisation of Total-reflection X-ray Fluorescence Spectrometry for Aerosol Analysis', *Spectrochim. Acta B*, **50**, 1787–1803 (1995).

Structure Determination, X-ray Diffraction for

D.D.L. Chung
State University of New York at Buffalo, Buffalo, USA

1 Introduction to X-ray Diffraction 1
1.1 Principles of Diffraction 1
1.2 Instrumentation for X-ray Diffraction 2
1.3 Applications of X-ray Diffraction 3
2 Symmetry 4
2.1 The Crystal Lattice and Net Planes 4
2.2 Symmetry of Point Groups 5
2.3 Space Groups 8
3 Theory of Diffraction 13
3.1 The Bragg Equation 13
3.2 The Reciprocal Lattice 13
3.3 The Ewald Construction 15
3.4 Symmetry in Reciprocal Space 15
3.5 The Structure Factor 15
3.6 Intensity 21
3.7 Diffuse Background 24
4 Production of X-rays and Their Characteristics 24
4.1 The X-ray Tube 24
4.2 Synchrotron Radiation 25
5 Diffractometers and Cameras 26
5.1 The Resolution Function 26
5.2 Single-crystal Methods 26
5.3 Powder Diffraction 35
Abbreviations and Acronyms 37
Related Articles 37
References 37

X-ray diffraction is a widely used experimental method for materials analysis. The scope of materials analysis that can be provided by X-ray diffraction includes crystalline phase identification, lattice parameter measurement, grain size measurement, and texture analysis. In particular, crystalline phase identification is a basic form of materials analysis that is critically important for metals, ceramics, and other crystalline materials. In contrast, elemental analysis, as provided by X-ray spectroscopy, atomic absorption, etc., provides information on the composition but not the phases. Because a material frequently consists of more than one phase, the composition of the multiphase mixture is not the same as the composition of an individual phase. A true understanding of a crystalline material cannot be obtained until one knows what phases are present in the material. Even if two phases have the same composition, their difference in crystal structure can make them have very different electrical, magnetic, thermal, mechanical, or chemical properties.

This is an adaptation of Chapter 1 of *X-ray Diffraction at Elevated Temperatures*, by D.D.L. Chung, P.W. De Haven, H. Arnold, Debashis Ghosh, Copyright © VCH Publishers, 1993. (Reprinted by permission of Wiley-VCH Publishers, Inc., a division of John Wiley & Sons, Inc.)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

1 INTRODUCTION TO X-RAY DIFFRACTION

1.1 Principles of Diffraction

X-rays are electromagnetic radiation with a wavelength of the order of 1 Å. Because of its inherent electric field, it interacts with charged particles, such as electrons in a solid. This interaction causes the electrons to emit wavelets. When these wavelets interfere with one another constructively, diffraction takes place.

Two criteria must be satisfied in order for constructive interference to occur. One criterion for constructive interference is given by Bragg's law (Equation 1):

$$\lambda = 2d \sin\theta \tag{1}$$

where λ is the wavelength of the X-ray, d is the spacing between planes of atoms in the solid, and θ is the angle of incidence, as illustrated in Figure 1.

The second criterion for constructive interference is that the arrangement of atoms in the unit cell of the solid under examination allows constructive interference to occur for that particular set of planes with interplanar spacing d. The allowed planes for constructive interference are those for which the structure factor is not zero. For example, for a face-centered cubic (fcc) crystal structure, the structure factor is not zero when the Miller indices h, k, and l of the planes are either all even or all odd, so that the allowed planes for constructive interference have Miller indices 111, 200, 220, etc. In contrast, for a body-centered cubic (bcc) structure, the structure factor is not zero when $h + k + l =$ an even integer, so that the allowed planes for constructive interference have Miller indices 110, 200, 211, 220, etc. Thus, the systematic absence of diffraction from certain planes provides a signature for identifying the atom arrangement in the unit cell.

The θ angle at which diffraction occurs (together with Equation 1) provides a determination of the interplanar

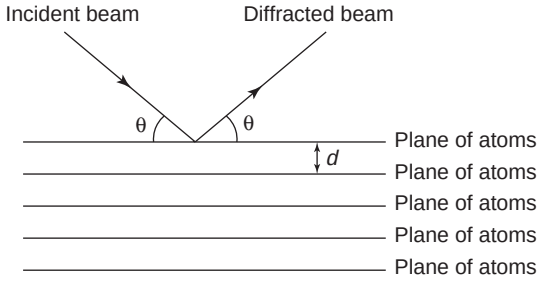

Figure 1 A set of atomic planes in a crystal, at an angle θ from the incident beam.

spacing d, which is related to the lattice parameters a, b and c (i.e. the unit cell dimensions) through Equations (2–5):

Cubic:

$$\frac{1}{d^2} = \frac{h^2 + k^2 + l^2}{a^2} \tag{2}$$

Tetragonal:

$$\frac{1}{d^2} = \frac{h^2 + k^2}{a^2} + \frac{l^2}{c^2} \tag{3}$$

Hexagonal:

$$\frac{1}{d^2} = \frac{4}{3}\frac{h^2 + hk + k^2}{a^2} + \frac{l^2}{c^2} \tag{4}$$

Orthorhombic:

$$\frac{1}{d^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \tag{5}$$

Information on the crystal structure helps in the identification of a phase. By determining the lattice parameters, one can measure the lattice strain due to residual stress, solid solution formation, or thermal expansion.

Most crystalline solids are polycrystalline, and the grains (or crystallites) in a solid are usually not exactly random in orientation. Preferred orientation (or texture) is quite common, especially in films, fibers, rolled products, wire-drawn products, etc. The texture affects the relative intensities of the diffraction peaks. However, for a quantitative analysis of the texture, pole figure measurement is necessary. In a pole figure measurement, the polycrystalline sample is rotated and tilted by known angles while the diffracted intensity at a fixed θ angle is measured as a function of the rotation and tilt angles. A sample may exhibit more than one preferred orientation (i.e. have more than one texture component). Different texture components may have different breadths in the distribution of orientations close to the corresponding ideal orientation (i.e. different texture components may have different strengths). Such detailed information on the texture can be obtained by pole figure measurement but cannot be obtained from the relative intensities of the diffraction peaks observed at various θ angles.

The grain size can be estimated from the width of a diffraction peak through Equations (6) and (7):

$$B = \frac{K\lambda}{L\cos\theta} \tag{6}$$

and

$$B^2 = B_M^2 - B_S^2 \tag{7}$$

where B_M is the measured width of the peak of the material under investigation, B_S is the measured width of a diffraction peak at a similar θ value from a standard having a grain size exceeding 1000 Å, L is the grain size, and λ is the wavelength of the X-ray used. The constant K is 0.9 when the full width at half-maximum (fwhm) is used and is 1.1 when the integral width is used. Equation (6) is known as the *Scherrer equation*. This method is limited to grain sizes above 1000 Å. It is also limited to samples that exhibit no nonuniform strain, because nonuniform strain also causes peak broadening.

1.2 Instrumentation for X-ray Diffraction

The instrumentation for X-ray diffraction involves an X-ray source, which provides the incident beam on the sample, and an X-ray detector, which measures the intensity of the diffracted beam emanating from the sample at a certain angle.

For a polycrystalline or powder sample, different crystallographic planes may be oriented simultaneously at their corresponding Bragg angle (θ) with respect to the incident beam, so that they may cause diffraction to occur simultaneously at different angles (2θ, the diffraction angle) from the incident beam. Thus, a number of diffracted beams may emanate from the same sample at the same time. All these diffracted beams need to be detected in order to obtain the complete diffraction pattern. For a single-crystal sample, the sample must be oriented appropriately in order to make a certain crystallographic plane be capable of causing diffraction. Thus, the single-crystal sample must be oriented differently to cause diffraction to occur from a differently oriented crystallographic plane.

The X-ray incident on the sample is ideally monochromatic. For that purpose, the beam to be monochromatized is incident on a monochromator. A *monochromator* is either a single crystal or a highly oriented polycrystalline material, which allows a diffracted beam at a well-defined monochromatic wavelength to emanate from it according to Bragg's law (Equation 1). For less stringent monochromatization, a filter is used; the filter is a material that preferentially absorbs radiation of wavelength below

a certain value, which is called the *absorption edge*, in order to remove the K_β radiation.

The X-ray source generates X-rays (called the *bremsstrahlung*) by the deceleration (upon hitting a target or being deflected by a magnet) of electrons traveling with a kinetic energy determined by the acceleration voltage encountered by the electrons. The higher the kinetic energy of the electrons, the higher the intensity of the bremsstrahlung. In addition, X-rays (called *characteristic X-rays*) are generated when the target of the electrons undergoes electronic transition due to excitation of the target atom. The bremsstrahlung has a continuous range of wavelengths (called the continuous spectrum); characteristic X-rays have wavelengths characteristic of the element (e.g. Cu) in the target (called the *characteristic spectrum*). In addition to generating X-rays, heat is generated in the bombarded target, which therefore must be cooled. If the electrons do not hit the same spot in the target all the time, the chance of melting the target is diminished and electrons of a higher current density can be tolerated. This is the idea behind the rotating-anode (target) X-ray source, which provides higher-intensity X-rays than the sealed-tube (stationary anode) X-ray source. Because electrons must travel in a vacuum, the target must be housed in a vacuum, which can be maintained in a sealed tube (possible for a stationary anode) or by continuous evacuation (necessary for a rotating anode). Electrons of extremely high kinetic energy are available in an accelerator such as a synchrotron (i.e. a storage ring). The bending of the electron beam in a synchrotron by using magnets can cause extremely intense X-rays (continuous spectrum) to be generated. Therefore, sealed tubes, rotating anodes, and synchrotron radiation are three kinds of X-ray sources, listed in order of increasing X-ray intensity.

The X-ray detector can be either a scintillation (or similar) counter or a position-sensitive detector, so that the detector must be moved in order to detect X-rays emanating from the sample at a different angle. This movement makes simultaneous detection of various diffracted beams impossible, thus making the collection of the full diffraction pattern time-consuming. In contrast, the latter detects and differentiates among the X-ray beams hitting different spots (along a line or in a two-dimensional array) in the detector, so that diffracted beams emanating from the sample at different angles can be detected simultaneously and the diffraction pattern can be collected quickly.

1.3 Applications of X-ray Diffraction

X-ray diffraction using a (more or less) monochromatic X-ray beam incident on a polycrystalline or powder sample is the most widely used configuration. The resulting diffraction pattern (i.e. diffracted intensity vs. 2θ) can be used for phase identification, although only crystalline phases can be identified. Both the 2θ positions and the relative peak intensities are used to help the phase identification. At least the strongest peaks of the suspected phase must be observed with about the expected intensity ratios in order for the phase identification to be certain. (A common pitfall is to claim that a certain phase has been identified on the basis of one or two peaks.) Coexisting phases in the same sample can be identified from the same diffraction pattern, provided that each phase is associated with a sufficient number of peaks at the expected intensity ratios. From the relative integrated intensities of the peaks due to different phases, the relative amounts of the phases can be estimated qualitatively. From the relative intensities of the peaks associated with different crystallographic planes of the same phase, the preferred orientation (texture) of the phase can be estimated qualitatively. From the fwhm of a peak, the grain size (for a polycrystalline sample) or the particle size (for a powder sample) can be estimated from the Scherrer equation for grain or particle sizes less than about 1000 Å.

For a quantitative determination of the concentration of a certain phase in a sample, the use of standards in the form of materials containing the same phases at known concentrations and with similar textures is necessary. Any difference in thickness between the sample and the standard can be corrected for by using $G_x = 1 - e^{-2\mu x \sin\theta}$ (for a diffractometer), where G_x is the fraction of the total integrated diffracted intensity contributed by a surface layer of depth x, μ is the linear absorption coefficient (product of the mass absorption coefficient and the density), and θ is the Bragg angle.

For a quantitative determination of the texture, pole figure determination is necessary. Figure 2 illustrates the Schulz reflection method, which is one of the most widely used methods. In this method, at a fixed 2θ angle, the sample is rotated about the AA' axis (angle $\alpha = 0$–90°) and also the BB' axis normal to the sample surface (angle $\delta = 0$–360°). The result of diffracted intensity (at a fixed 2θ angle) versus α and δ is conventionally displayed in a circle (called the *Wulff net*) with δ varying from 0° to 360° along the circumference and α varying from 0° to 90° along a radius from the circumference to the center of the circle (Figure 3). The intensity can be shown as contours in the circle, thereby indicating the sample orientations (the combinations of α and δ) that give a diffracted intensity of a certain level for the fixed 2θ angle. The smaller the width of a ''mountain'' depicted by the contours, the greater the degree of texturing for the texture component corresponding to that mountain. The higher the mountain, the greater the strength of that texture

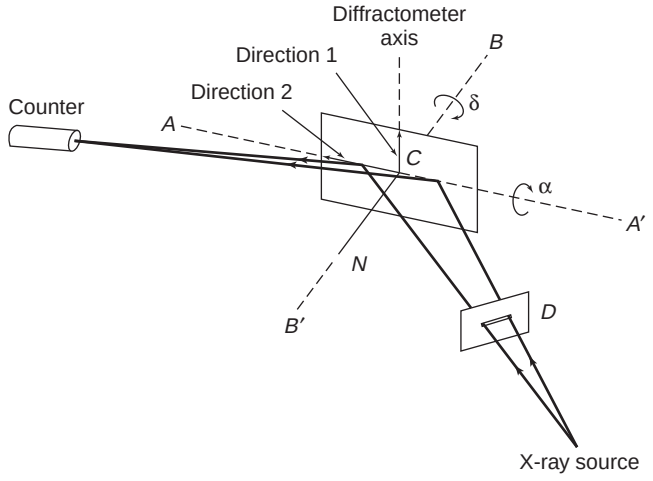

Figure 2 The Schulz reflection method for pole figure determination. The angle α is 90° when the sample sheet is in the vertical position drawn.

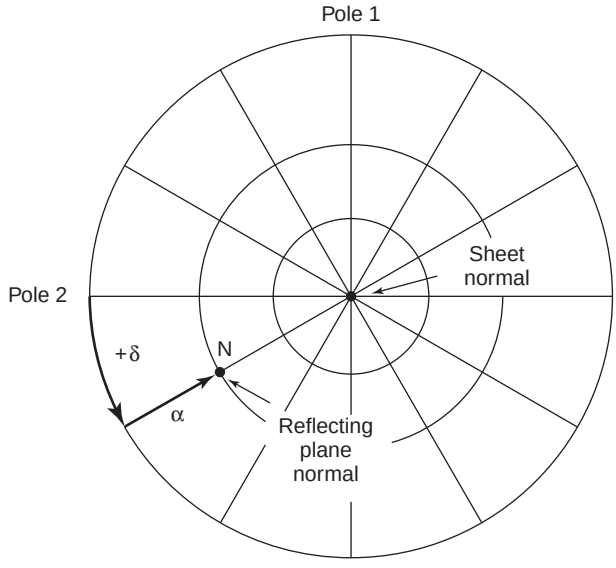

Figure 3 The Wulff net for plotting the pole figure. The angles α and δ correspond to those of Figure 2. Pole 1 represents direction 1 of Figure 2; pole 2 represents direction 2 of Figure 2. The center of the circle corresponds to $\alpha = 90°$.

component. Different mountains can belong to the same texture component (owing to symmetry) or to different texture components of the same phase. By changing 2θ to a different value of the same phase, additional clues can be obtained to ascertain the correctness of the texture determination. By changing 2θ to a value associated with a different phase, the texture of the other phase can be determined similarly. The preferred orientation relationship of coexisting phases can be determined by superimposing the pole figures of the different phases.

The lattice parameter (lattice constant) of a phase can be changed by solid solution formation or strain. For precise lattice parameter measurements, diffraction peaks with 2θ as near to 180° as possible should be used. This is because differentiation of Bragg's law with respect to θ gives $\Delta d/d = -\cot\theta\Delta\theta$.

In the case of a sample with a cubic unit cell of dimension a, Equation (2) is applicable and gives $\Delta a/a = \Delta d/d$. Therefore, $\Delta a/a = -\cot\theta\Delta\theta$. Because $\cot\theta$ approaches zero as θ approaches 90° (or 2θ approaches 180°), $\Delta a/a$ (the fractional error in a caused by a given error in θ) also approaches zero when θ approaches 90°.

2 SYMMETRY

2.1 The Crystal Lattice and Net Planes

A crystal with all its atoms is represented by its unit cell. The whole crystal can be constructed by placing one cell beside or beneath another. This procedure (i.e. shifting and copying the unit cell) is called *translation*. The translation vectors are shown in Equation (8):

$$\mathbf{t}_m = m_1\mathbf{a} + m_2\mathbf{b} + m_3\mathbf{c} \tag{8}$$

where **a**, **b** and **c** are the basis vectors of the unit cell and m_1, m_2, and m_3 are integers.

The vertices of all these cells are lattice points of the crystal. Each lattice point is symmetrically equivalent; from the point of view of symmetry, all lattice points have an identical environment. Through these points one can construct parallel planes, called the *net planes*. The different orientations of these planes with respect to the unit cell are indicated by the Miller indices h, k, and l. Net planes form the surface of a well-grown crystal. A lattice plane intersects the axes of the coordinate system. If the distance from the origin to the intersecting points is a_1 for the a-axis, a_2 for the b-axis, and a_3 for the c-axis, then the Miller indices are obtained according to Equation (9):

$$(h, k, l) = N\left(\frac{1}{a_1}, \frac{1}{a_2}, \frac{1}{a_3}\right. \tag{9}$$

where N is the smallest common factor such that the indices hkl become integers. Owing to the periodicity of the translation, the plane with the indices hkl consists of a set of equidistant parallel planes at the distance d. The multiples of hkl, say mh, mk, and ml, give a set of parallel planes at the distance d/m. The Miller indices of a plane or a set of planes are written in parentheses as (hkl); the set of symmetrically equivalent planes are written in braces as $\{h, k, l\}$. For example, the front face of a cube is denoted by (100); the six faces (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, and $(00\bar{1})$ are combined in one symbol {100}.

Lattice vectors are also used to indicate directions, which are written in brackets as [*uvw*]. For example, the positive directions parallel to the axes *a*, *b*, and *c* are [100], [010], and [001], respectively. The set of symmetrically equivalent directions to [*uvw*] is given by ⟨*uvw*⟩.

According to the directions of symmetry of the crystal, an appropriate unit cell is chosen. The unit cells of the seven crystal systems are listed in Table 1.

Table 1 Crystal systems

Crystal system		Relations
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$
Hexagonal	$a = b$; c	$\alpha = \beta = 90^\circ$; $\gamma = 120^\circ$
Rhombohedral	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$
Tetragonal	$a = b$; c	$\alpha = \beta = \gamma = 90^\circ$
Orthorhombic	a; b; c	$\alpha = \beta = \gamma = 90^\circ$
Monoclinic	a; b; c	β; $\alpha = \gamma = 90^\circ$; b-axis unique
	a; b; c	γ; $\alpha = \beta = 90^\circ$; c-axis unique
Triclinic or anorthic	a; b; c	α; β; γ

2.2 Symmetry of Point Groups

In this section, a short introduction to the symmetry of crystals is given. To gain a deeper insight, other textbooks should be considered.[1–13]

Let us look at a cube in a certain direction normal to one of its faces (Figure 4a). The origin of the coordinate system is placed at the body center of the cube. The *a*-axis is pointing at you, the *b*-axis to the right-hand side, and the *c*-axis upward. A rotation of 90° around the *a*-axis maps the cube onto itself. After four rotations of 90°, the cube is in its initial position. This symmetry

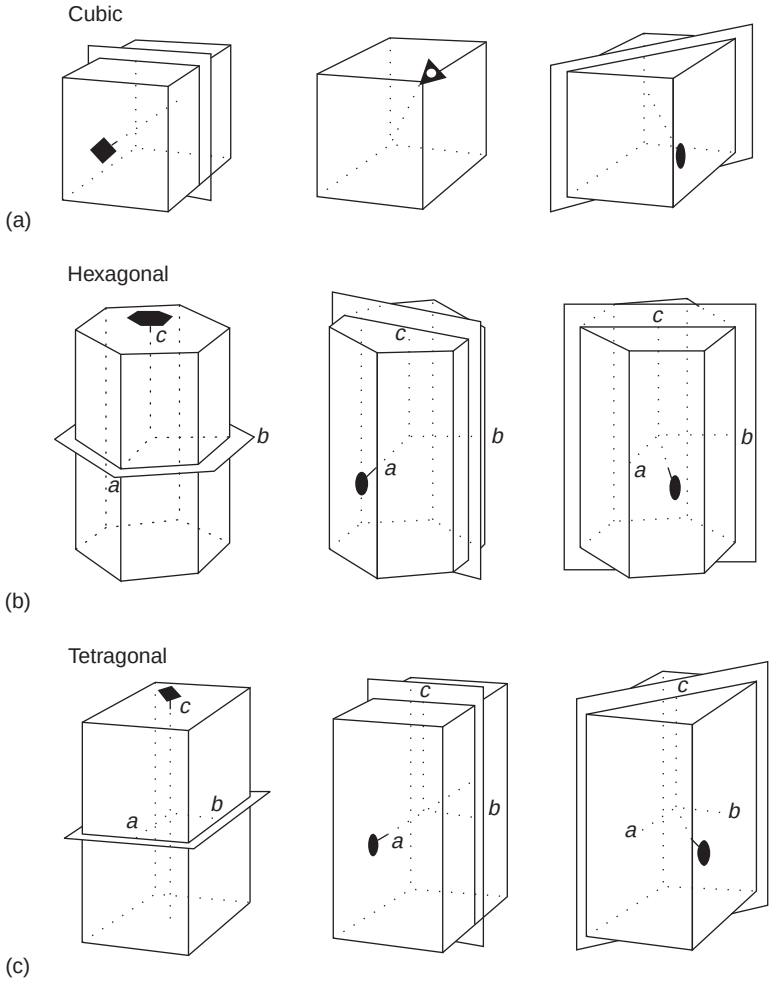

Figure 4 Symmetries in the directions of the point group symbol: (a) the cubic crystal system with the directions [100], [111], and [110]; (b) the hexagonal system with the directions [001], [100], and [210]; (c) the tetragonal system with the directions [001], [100], and [110]; (d) the orthorhombic system with the directions [100], [010], and [001]; (e) the monoclinic system with the direction [010]; (f) the rhombohedral system with the directions [111] and [110]. In the triclinic system, no direction is preferred by symmetry.

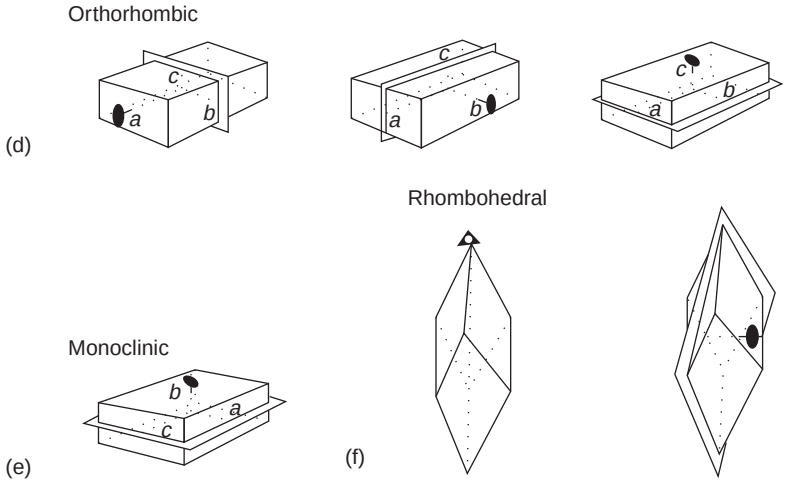

Figure 4 (*Continued*)

Table 2 Rotation axes

Rotation axis	Explanation	Symbol
Onefold	One rotation by 360°	1
Twofold	Two rotations by 180°	2
Threefold	Three rotations by 120°	3
Fourfold	Four rotations by 90°	4
Sixfold	Six rotations by 60°	6

Table 3 Rotoinversion axes

Rotoinversion axis	Explanation		Symbol
Onefold	Rotation by 360°	Followed by	$\bar{1}$
Twofold	Rotation by 180°	an inversion	$\bar{2} \equiv m$
Threefold	Rotation by 120°	through the	$\bar{3}$
Fourfold	Rotation by 90°	origin of the	$\bar{4}$
Sixfold	Rotation by 60°	coordinate system	$\bar{6}$

element is a fourfold rotation axis with the symbol 4. In crystals five rotation axes occur, as shown in Table 2. The rotation by 360° is identical to the operation identity.

Let us look along the body diagonal of the cube. This is a vector going through the origin in the body center of the cube with the vector components 1,1,1. This direction has the symbol [111]. Here, at first sight, we find a threefold rotation axis with the symbol 3. A rotation around the body diagonal by 120° maps the cube onto itself, but this is not the whole symmetry of this direction. This rotation, followed by an inversion through the origin, also maps the cube onto itself; this symmetry operation is called *rotoinversion*, with the symbol $\bar{3}$. In crystals five rotoinversion axes occur, as shown in Table 3 and Figure 5. The onefold rotoinversion is a pure center of inversion. The twofold inversion axis, $\bar{2}$, is identical to a mirror plane – ($\bar{2} \equiv m$). By a rotoinversion, a right-handed molecule is mapped onto a left-handed molecule. Thus, in a crystal with a rotoinversion axis a racemic mixture of right- and left-handed molecules exists, or the molecules, ions, etc. themselves have an appropriate symmetry.

For a well-grown crystal, only rotation axes and rotoinversion axes occur as symmetry elements. The variety of 32 crystal classes is given by the possible combination of rotation axes and rotoinversion axes. In this context, the following rules have to be mentioned:

- *The sixfold rotoinversion is a combination of a threefold rotation axis with a mirror plane normal to it*. Thus, $\bar{6} \equiv 3/m$, but only the symbol $\bar{6}$ is used.

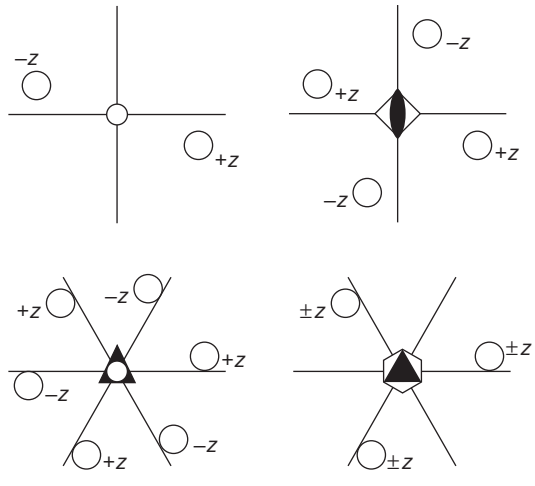

Figure 5 Rotoinversion: rotation axis along [001] and inversion center at the origin, i.e. $\bar{1}$, $\bar{3}$, $\bar{4}$, and $\bar{6}$.

- *The threefold inversion axis*, $\bar{3}$, *also contains a center of inversion*, $\bar{1}$. The following combinations of symmetry elements also contain a center of inversion: $2/m$, $4/m$, $6/m$. Consider a cube as a representation of a cubic crystal. First, look along the direction of a vector with the components 1, 0, 0. Such a direction has the symbol [100]. Normal to this direction, in the middle of the cube, is a mirror plane. This maps the front of the cube to the back, and vice versa. A mirror plane has the symbol m. The orientation of a plane is uniquely described by its normal direction. The symmetry of the cube with respect to the direction [100] is a fourfold rotation axis; normal to it is a mirror plane. The two symbols are combined to give $4/m$.

  Second, look along the direction of the body diagonal. As was mentioned earlier, this direction has the symbol [111]. It is a threefold rotoinversion axis, $\bar{3}$.

  Third, look along the direction [110], onto the edge. We can rotate the cube by 180°. This is a twofold rotation axis with the symbol 2. Normal to it is a mirror plane m. The combination of the two symbols 2 and m is written as $2/m$.

  With these three directions, [100], [111], and [110], all symmetry elements of the cube are described. In the direction [100], normal to a face, we have $4/m$; in the direction [111], a body diagonal, we have $\bar{3}$; in the direction [110], a plane diagonal, we have $2/m$. Thus, the symbol $4/m\,\bar{3}\,2/m$ describes the symmetry of the cube totally. The complete symmetry of the cube is exhibited by a *cubic holohedron*.

- *The symmetry of a crystal with a cube as unit cell can be lower*. The four threefold axes along the body diagonals make the three coordinate axes a, b, and c equivalent. Thus, these axes create the cubic coordinate system. In other words, a symbol of a crystal class with the symmetry element $\bar{3}$ or 3 in the second position of the symbol belongs to the cubic crystal system. The cubic crystal classes are: $4/m\,\bar{3}\,2/m$, $4\,3\,2$, $2/m\,\bar{3}$, $\bar{4}\,3\,m$, $2\,3$. The cubic holohedron can be reduced so that in [100] only a twofold rotation axis and in [111] only a threefold rotation axis persists, whereas in [110] no symmetry element is retained. The symmetry is still compatible with the cubic unit cell (i.e. $2\,3$ in the notation just described). The four threefold axes of the body diagonals make the axes, a, b, and c of the unit cell equivalent.

  Let us treat a tetragonal column with a square as the base plane (Figure 4c). This is a face form of a tetragonal crystal. We introduce the following coordinate system: a-axis normal to the rectangular face you are looking at, b-axis normal to the face at the right-hand side, and c-axis normal to the square face on top of the column. First, look along the direction [001], on top of the column. Here we have a fourfold axis and, normal to it, a mirror plane: $4/m$. Second, look along the direction [100], onto the rectangular face. Here we have a twofold axis and normal to it a mirror plane: $2/m$. The third direction is the bisector of the angle between the normals of the rectangular faces, i.e. the direction [110]. Here we find also the symmetry $2/m$. Thus, the holohedral tetragonal system is described by the symbol $4/m\,2/m\,2/m$.

  The symmetry of the tetragonal holohedron can be reduced and the following point groups occur: $4\,2\,2$, $4\,m\,m$, $\bar{4}\,2\,m$, $4/m$, $\bar{4}$, 4.

- *A crystal containing only one fourfold axis in [001] is tetragonal*. The same procedure is applied to a hexagonal column with a hexagon as the base plane (Figure 4b). We look in a direction normal to a rectangular face of the column. The coordinate system is chosen as follows: the a- and b-axes lie in the hexagonal plane passing through the midpoint of the column. The a-axis points from the midpoint to the left edge of the front face of the column. The b-axis points at an angle of 120° from the a-axis toward the right side. The c-axis points upward normal to the hexagonal base plane.

  To describe the symmetry, first look along the c-axis. This is the direction [001]. We find a sixfold rotation axis and a mirror plane normal to it: $6/m$. The second direction is along the a-axis, i.e. [100], and we find $2/m$. The third direction is normal to the rectangular front face, i.e. [210], and we find the symmetry $2/m$. Thus, the total symmetry of the hexagonal column is $6/m\,2/m\,2/m$, which is the symmetry of the hexagonal holoedrie. However, the smallest unit cell in the hexagonal system is not a hexagonal column. It is only one-third of the column, a section of 120° between the a-axis and the b-axis and parallel to the c-axis. The symmetry in the hexagonal system may be reduced to the following point groups: $6\,2\,2$, $6\,m\,m$, $\bar{6}\,2\,m$, $\bar{6}$, $6/m$, 6, $\bar{3}\,2/m$, $3\,2$, $3\,m$, $\bar{3}$, 3.

- *A crystal containing only one threefold axis in [001] belongs either to the hexagonal system or to the rhombohedral system*: An orthorhombic object is like a matchbox (Figure 4d). The coordinate axes are chosen normal to the faces, forming a right-handed, orthogonal coordinate system. Looking normal to the three faces in the directions [100], [010], and [001], we find in each case the symmetry $2/m\,2/m\,2/m$, the holoedrie of the orthorhombic system. There are also the point groups $2\,2\,2$ and $2\,m\,m$.

A monoclinic object is like a sheared matchbox (Figure 4e). Using an undeformed matchbox with the coordinate system defined previously, a shear force in the direction of the *b*-axis is applied to the face normal to the *a*-axis on the back side, and a shear force with the direction *b* is applied to the face at the front. Upon the shear deformation, the top and the bottom faces become parallelograms, whereas the other faces remain rectangular. Thus, the symmetry of the orthorhombic object in the direction [001] is retained as 2/*m*, whereas in the other two directions the symmetry is lost. If the symbol 1 denotes a direction in which a symmetry element is no longer found, the symbol of the orthorhombic holohedron is reduced to 1 1 2/*m*. Here the symmetry of the *c*-axis is retained; the *c*-axis is unique. It is customary to interchange the labels *a*, *b*, and *c* of the axes, so that the *b*-axis retains the symmetry 2/*m*. This procedure leads to the symbol 1 2/*m* 1 as the symbol for the monoclinic holohedron with the *b*-axis unique. The monoclinic system contains also the point groups 2, *m*. In the triclinic system the holohedron consists of the center of interversion, $\bar{1}$. Each face has a counterpart face related by the inversion. This symmetry can be reduced to the element 1, i.e. no symmetry along any direction.

A rhombohedron is a cube elongated or compressed in the direction [111] (Figure 4f). This deformation changes the angles of the cube uniformly. The lengths of the base vectors **a**, **b**, and **c** are unchanged. By the deformation, the symmetries of the cube in the direction [100] vanish, whereas those in the directions [110], [101], and [011] remain. From the symbol of the cubic holohedron $4/m\,\bar{3}\,2/m$, which represents the symmetry in the cubic directions [100], [111], and [110], the first position is canceled and the second and third positions of the cubic symbol are retained. As in the tetragonal and hexagonal cases, the main axis [111] (the threefold rotoinversion axis) is the first element in the symbol, followed by the symmetries normal to the main axis in the direction [110]. Thus, the symmetry of the rhombohedron is symbolized by $\bar{3}\,2/m$, the holohedral symmetry of the rhombohedral system. In the rhombohedral cell, the symmetry can be reduced to $3\,2$, $\bar{3}\,m$, $\bar{3}$, 3. We have already found these point groups among those of the hexagonal system. As is shown in the next section, the rhombohedral unit cell can be represented by a centered hexagonal unit cell. The hexagonal coordinate system is more comfortable than the rhombohedral coordinate system, so the hexagonal coordinate system is usually preferred for the representation of a rhombohedral crystal. However, there are several difficulties in choosing the proper transformation between the two coordinate systems. The *International Tables*[14] should be consulted carefully.

The 32 combinations of symmetry elements are crystallographic point groups. The midpoints of the bodies are left invariant by the symmetry. The 32 symmetry groups are also called *crystal classes*. Every crystal (or space group) belongs to only one crystal class (or point group). All crystal classes can be derived from the cubic or hexagonal holohedra as subgroups. The group–subgroup relations are shown in Figure 6. They are important to compare structures and to discuss phase transitions.

The symmetry of the macroscopic properties of a crystal obeys the symmetry of the point group of the crystal; the dielectric constants, the piezoelectric constants, the elastic constants, etc. are examples of macroscopic properties.

2.3 Space Groups

The point groups describe the symmetry of the crystal form and the physical properties of large well-grown crystals, but not their internal structure; this is performed by the 230 space groups. By the translation operation (Equation 8), space is filled by a replica of the unit cell. The macroscopic symmetry of a crystal is the result of the symmetry in the unit cell. The symmetry elements of a space group are isomorphic to those of the corresponding point group, i.e. in addition to mirror planes and rotation axes, glide planes and screw axes also occur.

2.3.1 Lattice Types

The directions of the symmetry elements of the point groups (i.e. rotation axes and rotoinversion axes) have been defined earlier. The conventional unit cell is chosen in accordance with these directions. As a consequence, lattice types (the Bravais lattices) were introduced. In addition to the primitive lattice (denoted by the symbol *P*), centered lattices are also used. The *P* lattice has lattice points at the eight corners of the unit cell, i.e. only one lattice point per cell. These points are defined by translation, as introduced by Equation (11). In a centered lattice the volume of the unit cell is enlarged and additional translations (or shift vectors) are needed. These are given in Table 4 and are shown in Figure 7. In the rhombohedral system the lattice is denoted by *R* in order to distinguish this primitive unit cell from that of the hexagonal system. However, the rhombohedral lattice can also be described by a centered hexagonal unit cell, R_{hex}. This cell is more

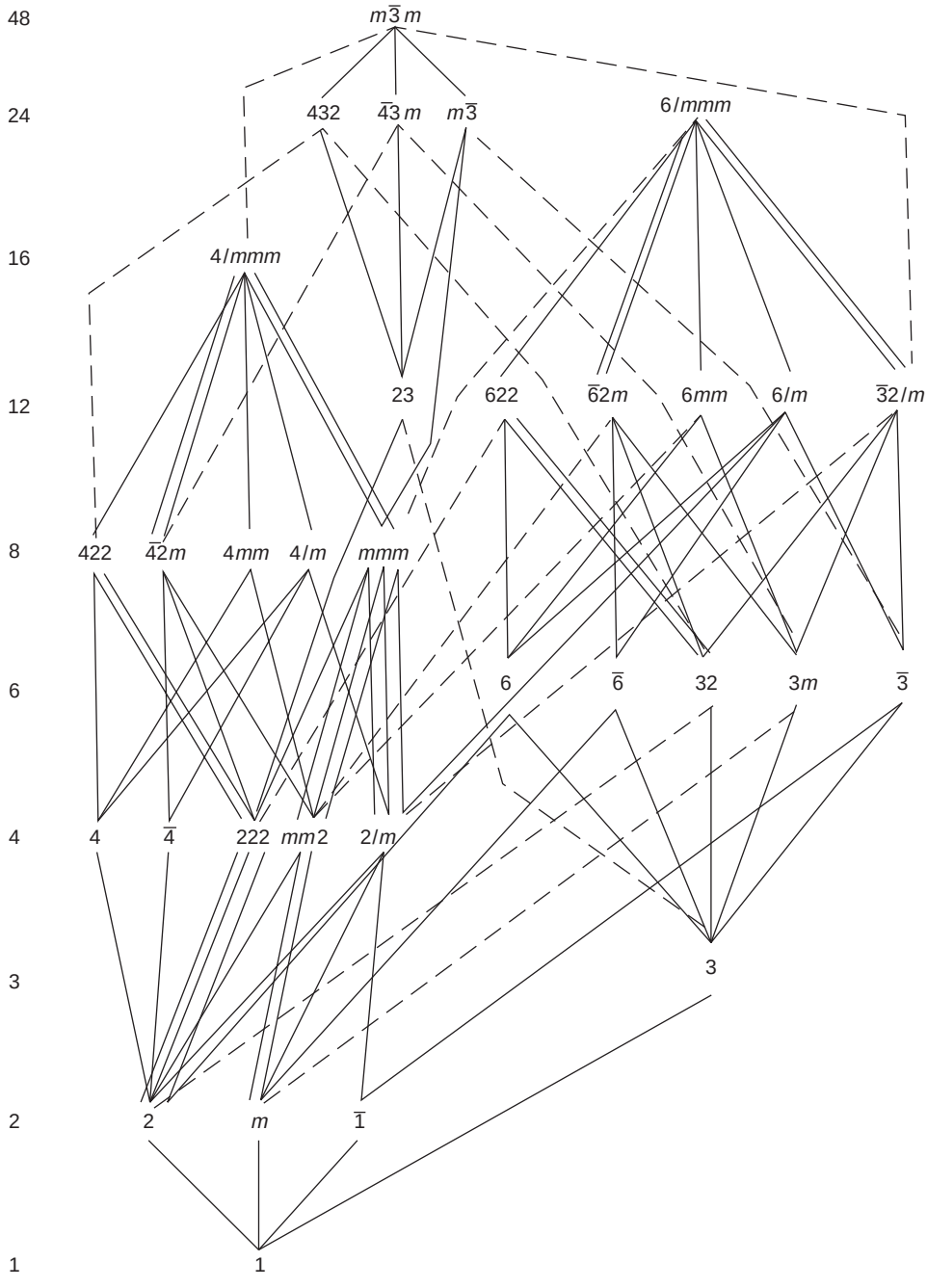

Figure 6 Subgroup relations between the 32 crystallographic point groups. In the left-hand part of the figure is listed the order of the group, i.e. the number of symmetry operations. This is also the number of symmetrically equivalent faces of the general face form with the indices *hkl*. Straight lines give the normal subgroup relations; dashed lines give the conjugated subgroup relations.

comfortable under many circumstances and is usually preferred.

2.3.2 *Position of an Atom*

The basis vectors of the unit cell, **a**, **b**, and **c**, are used to indicate the position of an atom in the unit cell. The coordinates of an atom are given by x, y, and z, so that the position vector emanating from the origin of the unit cell is given by Equation (10):

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} \tag{10}$$

2.3.3 *Symmetry Operations*

If we have a mirror plane normal to the basis vector **a** through the origin of the unit cell, corresponding to

Table 4 Lattices

Lattice type	Symbol	Additional translations	Crystal system[a]	Reflection conditions for *hkl*
Primitive	*P*	None	a,m,o,t,h,c	None
	R	None	r	None
Body-centered	*I*	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$	o,t,c	$h+k+l=2n$
Side-face-centered	*C*	$\frac{1}{2},\frac{1}{2},0$	m,o	$h+k=2n$
	or *A*	$0,\frac{1}{2},\frac{1}{2}$	m,o	$k+l=2n$
	or *B*	$\frac{1}{2},0,\frac{1}{2}$	m,o	$h+l=2n$
All-face-centered	*F*	$\frac{1}{2},\frac{1}{2},0$	o,c	*h*, *k*, *l* all even or all odd
		$\frac{1}{2},0,\frac{1}{2}$		
		$0,\frac{1}{2},\frac{1}{2}$		
Rhombohedral[b]	R_{hex}	$\frac{2}{3},\frac{1}{3},\frac{1}{3}$ $\frac{1}{3},\frac{2}{3},\frac{2}{3}$	h	$-h+k+l=3n$

[a] The crystal systems are denoted by: a, anorthic or triclinic; m, monoclinic; o, orthorhombic; t, tetragonal; r, rhombohedral; h, hexagonal; c, cubic.
[b] Centered lattice in hexagonal unit cell.

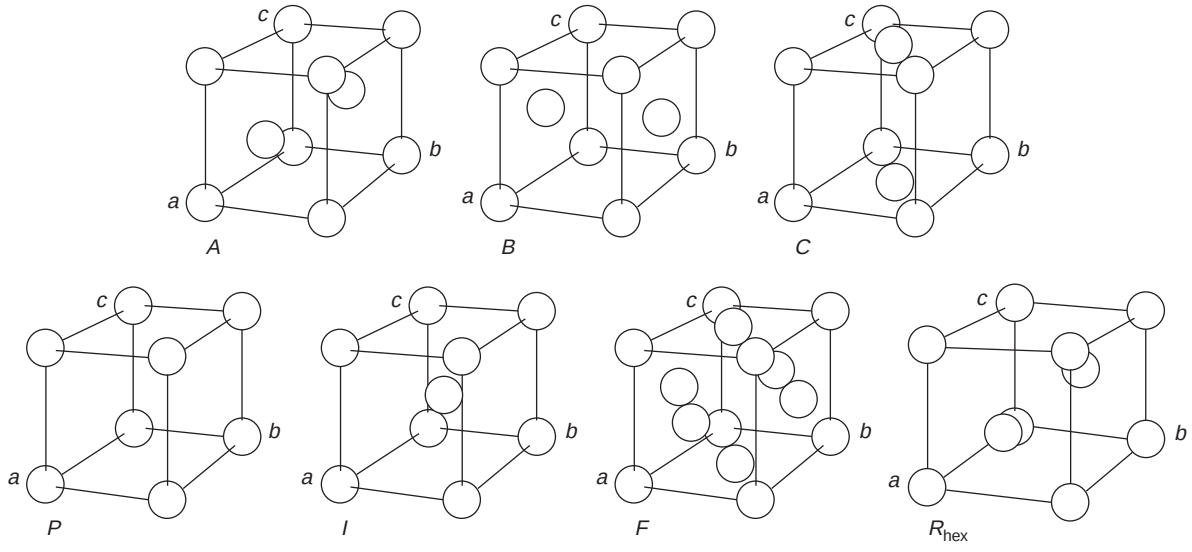

Figure 7 The lattice types. *First row*: the side-face-centered lattice (*S*; depending on the choice of the coordinate axes, the lattice type is called *A*, *B*, or *C*). *Second row*: the primitive lattice (*P*), the body-centered lattice (*I*), the all-face-centered lattice (*F*), and the centered hexagonal cell (R_{hex}). The primitive rhombohedral cell (*R*) is usually described by a centered hexagonal cell (R_{hex}).

each atom at *x*, *y*, and *z*, another atom exists with the coordinates $-x$, *y*, and *z*, i.e. the atom at a distance $x\mathbf{a}$ in front of the mirror plane has its counterpart at a distance $-x\mathbf{a}$ behind the mirror plane. A rotation by 90° around the rotation axis [001] transforms an atom at *x*, *y*, *z* to the coordinates $-y$, *x*, *z*. Thus, each symmetry operation is represented by a transformation of the coordinates *x*, *y*, *z* to x', y', z'.

The transformation of coordinates is easily performed by matrices. Equation (10) can thus be written in matrix notation according to Equation (11):

$$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} = (\mathbf{a},\mathbf{b},\mathbf{c}) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{11}$$

A symmetrically equivalent position x', y', z' is also given by Equation (11), leading to Equation (12):

$$\mathbf{r}' = x'\mathbf{a} + y'\mathbf{b} + z'\mathbf{c} = (\mathbf{a},\mathbf{b},\mathbf{c}) \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

$$= (\mathbf{a},\mathbf{b},\mathbf{c}) \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{12}$$

where **W** is the matrix of a point group symmetry operation, i.e. the matrix of a rotation or rotoinversion. As an example, the rotation by $+90^\circ$ around the *c*-axis, indicated by the symbol 4^+, is given by:

$$\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Insertion into Equation (12) gives the coordinates as described previously.

2.3.4 Symmetry of Space Groups

The symmetry operations of point groups return, after a number of operations, back to the starting position. In space groups, however, each lattice point is, with respect to symmetry, equivalent. Thus, after a number of operations, either the starting point or a point shifted by a translation from the starting point is reached. The symmetry operations in space groups contain a rotational part (or point group part), i.e. the matrix **W**, and a translational part **w**. Both together define a space group symmetry operation (**W, w**). The additional symmetry elements are screw axes and glide planes. An atomic position **r** is transformed to a symmetrically equivalent position **r**′ by Equation (13):

$$\mathbf{r}' = (\mathbf{a}, \mathbf{b}, \mathbf{c}) \left(\mathbf{W} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbf{w} \right) \tag{13}$$

For a screw axis, a rotation is accompanied by a shift parallel to the rotation axis. A rotation of an *n*-fold axis ($n = 2$, 3, 4, and 6) by $2\pi/n$ is associated with a shift by $m\mathbf{t}/n$, where **t** is the vector of the translation parallel to the rotation axis and $m = 1$ to $n-1$. The possible rotation and screw axes n_m are given in Table 5 and shown in Figure 8. As an example we consider the operation $4_1{}^+$. This is a rotation by $+90^\circ$, as was mentioned previously, followed by a shift of a quarter of the unit cell dimension along the *c*-axis. Thus, for this example, we get:

$$\mathbf{r}' = (\mathbf{a}, \mathbf{b}, \mathbf{c}) \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{1}{4} \end{pmatrix} = (\mathbf{a}, \mathbf{b}, \mathbf{c}) \begin{pmatrix} -y \\ x \\ z + \frac{1}{4} \end{pmatrix}$$

For a glide plane, the atom mapped behind the mirror plane is shifted in a direction parallel to the plane. For example, we consider the mirror plane normal to the *a*-axis, i.e. the direction [100], as was mentioned earlier. The possible shift vectors are $0,\frac{1}{2},0$ parallel to the *b*-axis, $0,0,\frac{1}{2}$ parallel to the *c*-axis, or a combination of both directions $0,\frac{1}{2},\frac{1}{2}$; these motions are denoted by the symbols *b*, *c*, and *n* respectively (Figure 9). In *F*-centered lattices, which have a lattice point at $0,\frac{1}{2},\frac{1}{2}$, *d* glide planes (Table 6) may occur, which have a glide vector $0,\frac{1}{4},\frac{1}{4}$ that is half of the translation $0,\frac{1}{2},\frac{1}{2}$.

In the cubic system normal to the direction [110], *c* glide planes exist with the components $0,0,\frac{1}{2}$, and *n* glides with $\frac{1}{2},\frac{1}{2},\frac{1}{2}$. In *I*-centered lattices with a lattice point at $\frac{1}{2},\frac{1}{2},\frac{1}{2}$, *d* glides may exist with the components $\frac{1}{4},\frac{1}{4},\frac{1}{4}$.

A symmetry operation may transform the atom at position *x*, *y*, *z* onto a point outside the unit cell. By a translation, this point can be transformed back into the unit cell. Mathematically, the positions, *x*′, *y*′, *z*′ are computed with mod 1, i.e. all coordinates *x*′, *y*′, *z*′ are reduced to the range $0 < x, y, z < 1$.

2.3.5 General and Special Positions

The set of all symmetry operations (**W, w**) of the space group transforms the position *x*, *y*, *z* of one atom to all symmetrically equivalent positions. This set of all positions is called the *general position*. However, if an atom lies on a mirror plane or a rotation axis, the symmetry operations of the symmetry elements map the position onto itself. The number of positions is reduced. This is called a *special position* and it has a point group

Table 5 Rotation and screw axes parallel to the *c*-axis

Symmetry element	Symbol	Components of the translation in direction of the rotation axis	Reflection conditions if the axis is parallel to the *c*-axis (i.e. 001)[a]
Twofold rotation axis	2	None	None
Twofold screw axis	2_1	$\frac{1}{2}$	$l = 2n$
Threefold rotation axis	3	None	None
Threefold screw axis	3_1	$\frac{1}{3}$	$l = 3n$
	3_2	$\frac{2}{3} = -\frac{1}{3}$	$l = 3n$
Fourfold rotation axis	4	None	None
Fourfold screw axis	4_1	$\frac{1}{4}$	$l = 4n$
	4_2	$\frac{2}{4} = \frac{1}{2}$	$l = 2n$
	4_3	$\frac{3}{4} = -\frac{1}{4}$	$l = 4n$
Sixfold rotation axis	6	None	None
Sixfold screw axis	6_1	$\frac{1}{6}$	$l = 6n$
	6_2	$\frac{2}{6} = \frac{1}{3}$	$l = 3n$
	6_3	$\frac{3}{6} = \frac{1}{2}$	$l = 2n$
	6_4	$\frac{4}{6} = -\frac{1}{3}$	$l = 3n$
	6_5	$\frac{5}{6} = -\frac{1}{6}$	$l = 6n$

[a] In the case of screw axes parallel to the *a*-axis, the extinction conditions must be replaced by $h00$ with $h = mn$; for a screw axis parallel to the *b*-axis, by $0k0$ with $k = mn$.

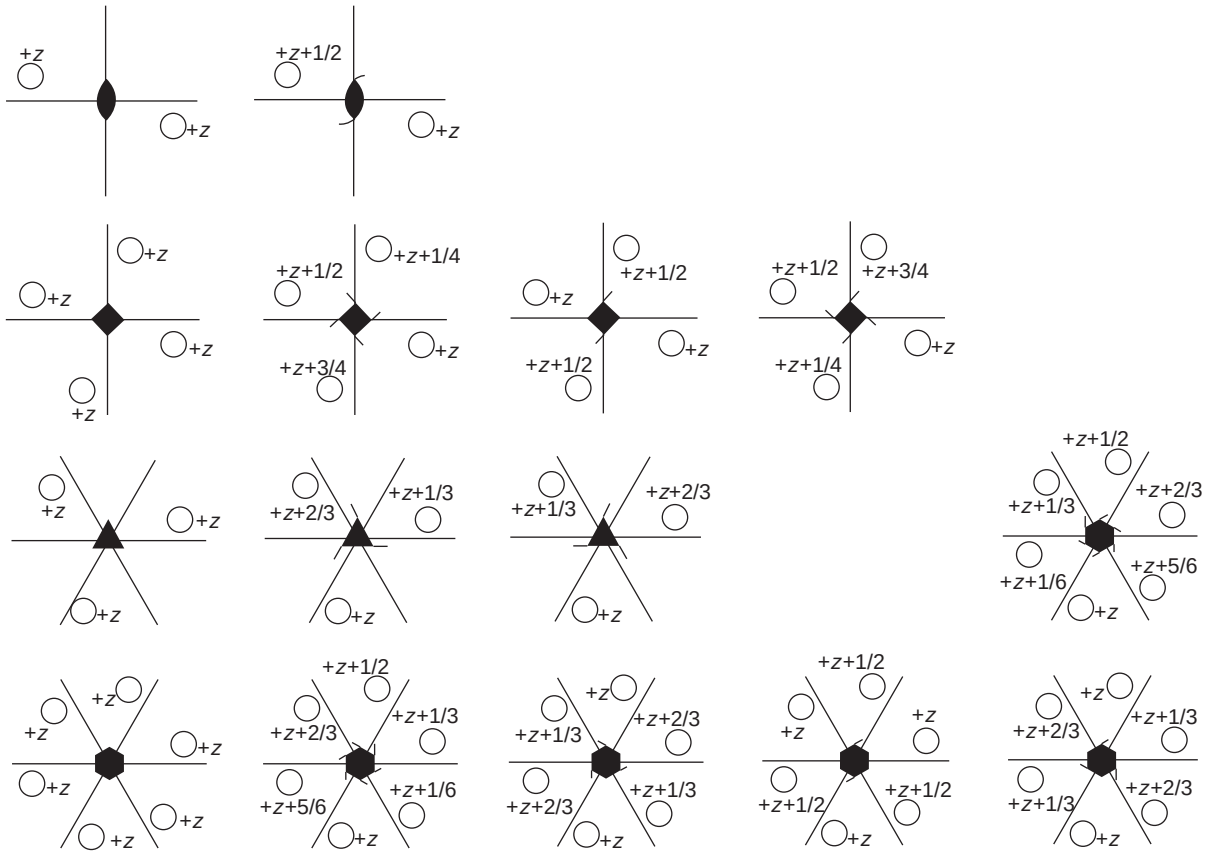

Figure 8 Rotation and screw axes parallel to *c*: first row, 2 and 2_1; second row, 4, 4_1, 4_2 and 4_3; third row, 3, 3_1, 3_2 and 6_5; last row, 6, 6_1, 6_2, 6_3 and 6_4.

symmetry defined by the symmetry elements passing through the point.

The special positions are important in inorganic structures; they enable the positioning of complex chemical formulae units in the unit cell. For example, quartz (SiO_2) crystallizes in the space group 152 ($P3_121$) with three formula units in the cell. The six oxygen atoms are in general positions, i.e.:

$$x, y, z;\ \bar{y}, x-y, z+\tfrac{1}{3};\ \bar{x}+y, \bar{x}, z+\tfrac{2}{3};\ y, x, \bar{z};$$
$$x-y, \bar{y}, \bar{z}+\tfrac{2}{3};\ \bar{x}, \bar{x}+y, \bar{z}+\tfrac{2}{3}$$

and the three silicon atoms are on twofold axes, i.e. $x,\ 0,\ \frac{1}{3};\ 0,\ x,\ \frac{2}{3};\ \bar{x},\ \bar{x},\ 0$. However, in structures with organic molecules, screw axes and glide planes are preferred. About 30% of these compounds crystallize in the monoclinic space group 14 ($P12_1/c1$).

2.3.6 Space Group Symbols

There are 230 space groups (i.e. the possible combination of symmetry elements to map the crystal onto itself). The translation is indicated by the symbol of the lattice type, followed by the symbol of the symmetry of the unit cell, i.e. a point group symbol, in which the symbols of the rotation axes and mirror planes can be replaced by those of screw axes and glide planes.

For example, the space group 227 ($F4_1/d\bar{3}2/m$) is the symmetry of the diamond structure. The lattice type is *F*. In [100] there is a screw axis 4_1 and normal to it a *d* glide; in [111] there is the rotoinversion axis $\bar{3}$; in [110] there is a twofold axis 2 with a mirror plane *m* normal to it.

The space group 62 ($P2_1/n2_1/m\,2_1/a$) is discussed as a second example. It has a primitive lattice *P*. In the direction [100] there are twofold screw axes normal to *n* glide planes; in [010] there are 2_1 screw axes normal to mirror planes *m*; in [001] there are 2_1 screw axes normal to *a* glides. An orthorhombic space group can be symbolized by several symbols. By exchanging the axes *a*, *b*, *c* by *b*, *c*, *a*, the space group symbol becomes $P2_1/m2_1/c2_1/n$, i.e. the direction of the glide vector of the former glide plane *a* must always point in the direction of the *n* glide, which is transformed to the direction *c*. Thus, a space group should be symbolized by its standard symbol, as given in the *International Tables*,[14] where the space groups are all listed extensively, including drawings.

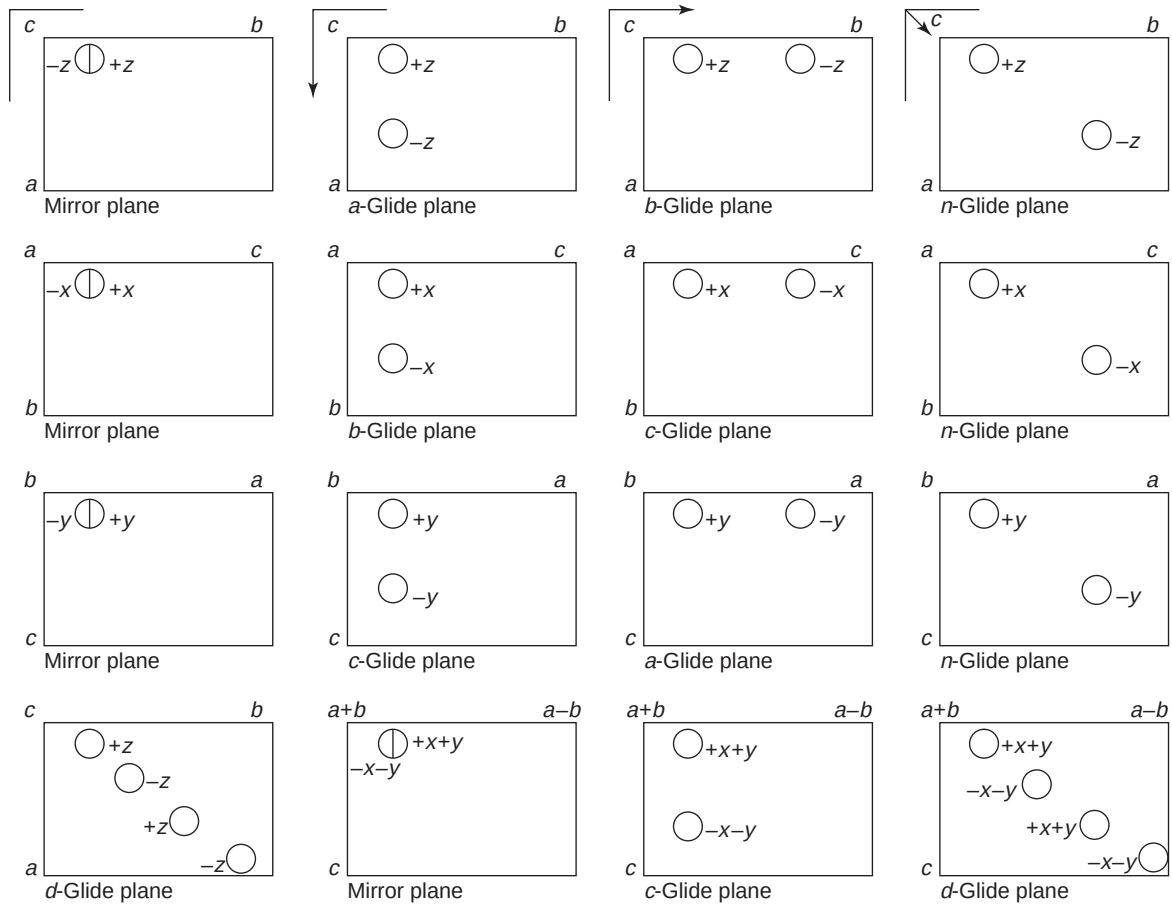

Figure 9 Mirror and glide planes. In the first row the projection is along [001] or the *c*-axis, in the second row the projection is along [100] or the *a*-axis, in the third row it is along [010] or the *b*-axis, and in the fourth row several examples are given: a *d*-glide plane normal to [100], which occurs only in *F*-centered lattices; in the tetragonal and cubic systems, normal to [110] mirror planes and *c*-glide planes; in *I*-centered lattices, also *d*-glide planes.

3 THEORY OF DIFFRACTION

X-rays interact with electrons. In the picture of classical electrodynamics, each electron oscillates, as it is agitated by the incoming wave, and emits a spherical wave. These waves interfere with the waves emitted by other electrons. In principle, the whole space is filled with emitted radiation. This can be seen in gases, liquids, and amorphous materials. Crystals, however, emit X-rays only in special directions. The reason for this effect is the crystal lattice and its condition of interference. Nonperiodic defects, inclusions of liquids, etc. in a crystal contribute to the diffuse background and can be analyzed if their concentration is high.

3.1 The Bragg Equation

The reflection of the X-rays by a crystal can be described as a reflection by its net planes (*hkl*). An incoming plane wave with the wavelength λ is reflected by a set of net planes when the path difference of the wave reflected by neighboring net planes is a period of the wave. From Figure 10 it can be derived that the diffraction angle θ_{hkl} is given by Equation (14):

$$2d_{hkl} \sin\theta_{hkl} = \lambda \tag{14}$$

where d_{hkl} is the distance between the reflecting net planes.

3.2 The Reciprocal Lattice

The diffraction of X-rays can be interpreted as the conservation of the momentum of the incoming wave $\mathbf{s}_i$ and that of the reflected wave $\mathbf{s}_f$ (the momentum of a wave is $h\mathbf{s}$, where h is Planck's constant; $h\mathbf{r}^*$ is the momentum transfer). Both vectors have the same

Table 6 Mirror and glide planes

Plane normal to	Symbol	Glide vector	Reflections	Conditions
[100]	m	None		None
	b	$0,\frac{1}{2},0$	$0kl$	With $k = 2n$
	c	$0,0,\frac{1}{2}$		With $l = 2n$
	n	$0,\frac{1}{2},\frac{1}{2}$		With $k + l = 2n$
	d	$0,\frac{1}{4},\frac{1}{4}$		With $k + l = 4n$
		$0,\frac{3}{4},\frac{3}{4}$		
[010]	m	None		None
	a	$\frac{1}{2},0,0$	$h0l$	With $h = 2n$
	c	$0,0,\frac{1}{2}$		With $l = 2n$
	n	$\frac{1}{2},0,\frac{1}{2}$		With $h + l = 2n$
	d	$\frac{1}{4},0,\frac{1}{4}$		With $h + l = 4n$
		$\frac{3}{4},0,\frac{3}{4}$		
[001]	m	None		None
	a	$\frac{1}{2},0,0$	$hk0$	With $h = 2n$
	b	$0,\frac{1}{2},0$		With $k = 2n$
	n	$\frac{1}{2},\frac{1}{2},0$		With $h + k = 2n$
	d	$\frac{1}{4},\frac{1}{4},0$		With $h + k = 4n$
		$\frac{3}{4},\frac{3}{4},0$		
[110] (in tetragonal and cubic systems only)	m	None		None
	c[a]	$0,0,\frac{1}{2}$	hhl	With $l = 2n$
	n[a]	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$,		
	d	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$		With $2h + l = 4n$
		$\frac{3}{4},\frac{3}{4},\frac{3}{4}$		

[a] In [110] there are always alternating parallel c glides and n glides.

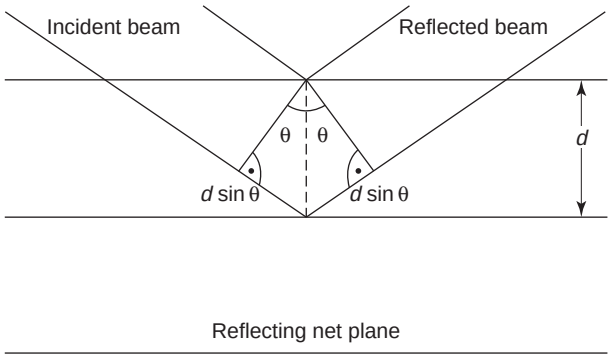

Figure 10 The Bragg equation. The difference of the path of a ray reflected by a net plane hkl with that reflected from the next neighboring net plane must be equal to the wavelength λ (i.e. $2d_{hkl}\sin\theta_{hkl} = \lambda$).

magnitude (modulus) 1/λ. The difference of the vectors is shown by Equation (15):

$$\mathbf{s}_f - \mathbf{s}_i = \mathbf{r}^*_{hkl} \tag{15}$$

and, as shown by Equation (16):

$$|\mathbf{r}^*_{hkl}| = \frac{1}{d_{hkl}} \tag{16}$$

The vector $\mathbf{r}^*_{hkl}$ is normal to the set of net planes (hkl) (Figure 10). Equation (14) can be written as shown by Equation (17):

$$\frac{2\sin\theta_{hkl}}{\lambda} = |\mathbf{r}^*_{hkl}| \tag{17}$$

For each set of net planes with the Miller indices (hkl), a normal vector $\mathbf{r}^*_{hkl}$ is defined. The end-points of such vectors are reciprocal lattice points. The vectors $\mathbf{r}^*_{hkl}$ emanate from the origin and their physical dimension is the reciprocal length.

The reciprocal lattice is given by Equation (18):

$$\mathbf{r}^*_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{18}$$

where $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ are the basis vectors of the reciprocal space. The collection of three integers h, k, and l (the Miller indices) define a lattice point.

If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are the basis vectors of the unit cell of a crystal, the basis vectors in reciprocal space $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ are obtained by Equations (19) and (20):

$$\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\\ \mathbf{c}\end{pmatrix}(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*) = \begin{pmatrix}\mathbf{a}\cdot\mathbf{a}^* & \mathbf{a}\cdot\mathbf{b}^* & \mathbf{a}\cdot\mathbf{c}^*\\ \mathbf{b}\cdot\mathbf{a}^* & \mathbf{b}\cdot\mathbf{b}^* & \mathbf{b}\cdot\mathbf{c}^*\\ \mathbf{c}\cdot\mathbf{a}^* & \mathbf{c}\cdot\mathbf{b}^* & \mathbf{c}\cdot\mathbf{c}^*\end{pmatrix} \tag{19}$$

$$= \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} \tag{20}$$

The vector $\mathbf{a}^*$ is normal to the $\mathbf{b}$–$\mathbf{c}$ plane, the vector $\mathbf{a}$ is normal to the $\mathbf{b}^*$–$\mathbf{c}^*$ plane, etc. (Figure 11). The lattice in direct space defines uniquely the lattice in reciprocal space, and vice versa. Thus, normal to a lattice translation $\mathbf{t}_m$ in direct space is a set of net planes in reciprocal space at a distance d^*_m. The choice of the unit cell in one space defines the unit cell in the other space. The volume of the reciprocal unit cell V^* is reciprocal to the volume of the unit cell in direct space, i.e. $VV^* = 1$. The rules for selection of the standard unit cells depend on the crystal system. If a choice does not agree with the rules, the unit cells in direct and reciprocal spaces should be transformed to the standard unit cells. This procedure is treated intensively in the *International Tables*.[14]

The modulus of the vector $\mathbf{r}^*$ is given by Equations (21) and (22):

$$|\mathbf{r}^*_{hkl}|^2 = (h, k, l)\begin{pmatrix}\mathbf{a}^*\cdot\mathbf{a}^* & \mathbf{a}^*\cdot\mathbf{b}^* & \mathbf{a}^*\cdot\mathbf{c}^*\\ \mathbf{b}^*\cdot\mathbf{a}^* & \mathbf{b}^*\cdot\mathbf{b}^* & \mathbf{b}^*\cdot\mathbf{c}^*\\ \mathbf{c}^*\cdot\mathbf{a}^* & \mathbf{c}^*\cdot\mathbf{b}^* & \mathbf{c}^*\cdot\mathbf{c}^*\end{pmatrix}\begin{pmatrix}h\\ k\\ l\end{pmatrix} \tag{21}$$

$$= \frac{1}{d^2_{hkl}} \tag{22}$$

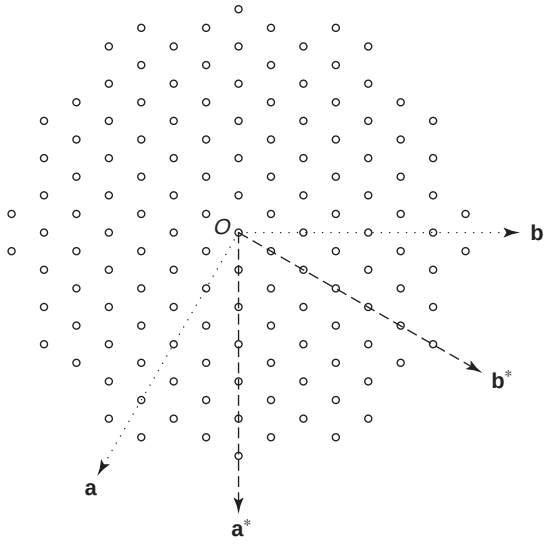

Figure 11 The $hk0$ plane of a hexagonal lattice with the directions of the basis vectors $\mathbf{a}^*$ and $\mathbf{b}^*$ of reciprocal space. The origin is denoted by O. The directions of the basis vectors of direct space **a** and **b** are schematically indicated. The base vectors **c** and $\mathbf{c}^*$ are normal to the $hk0$ plane.

and the metric tensor of reciprocal space is given by Equation (23):

$$\mathbf{g}^* = \begin{pmatrix} \mathbf{a}^* \cdot \mathbf{a}^* & \mathbf{a}^* \cdot \mathbf{b}^* & \mathbf{a}^* \cdot \mathbf{c}^* \\ \mathbf{b}^* \cdot \mathbf{a}^* & \mathbf{b}^* \cdot \mathbf{b}^* & \mathbf{b}^* \cdot \mathbf{c}^* \\ \mathbf{c}^* \cdot \mathbf{a}^* & \mathbf{c}^* \cdot \mathbf{b}^* & \mathbf{c}^* \cdot \mathbf{c}^* \end{pmatrix} \tag{23}$$

The formula is specified for the crystal systems in Table 7. The modulus of a vector in direct space is given by Equation (24):

$$|\mathbf{r}|^2 = (x, y, z) \begin{pmatrix} \mathbf{a} \cdot \mathbf{a} & \mathbf{a} \cdot \mathbf{b} & \mathbf{a} \cdot \mathbf{c} \\ \mathbf{b} \cdot \mathbf{a} & \mathbf{b} \cdot \mathbf{b} & \mathbf{b} \cdot \mathbf{c} \\ \mathbf{c} \cdot \mathbf{a} & \mathbf{c} \cdot \mathbf{b} & \mathbf{c} \cdot \mathbf{c} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{24}$$

and the metric tensor of direct space is given by Equation (25):

$$\mathbf{g} = \begin{pmatrix} \mathbf{a} \cdot \mathbf{a} & \mathbf{a} \cdot \mathbf{b} & \mathbf{a} \cdot \mathbf{c} \\ \mathbf{b} \cdot \mathbf{a} & \mathbf{b} \cdot \mathbf{b} & \mathbf{b} \cdot \mathbf{c} \\ \mathbf{c} \cdot \mathbf{a} & \mathbf{c} \cdot \mathbf{b} & \mathbf{c} \cdot \mathbf{c} \end{pmatrix} \tag{25}$$

The relations between the moduli of the basis vectors in direct space and reciprocal space are given in Table 8.

3.3 The Ewald Construction

The moduli of the incident and reflected waves $\mathbf{s}_i$ and $\mathbf{s}_f$ are both $1/\lambda$. Ewald made a geometric construction of Equation (15) in the reciprocal lattice. The reflection condition for a net plane (hkl) is that the sphere of reflection with radius $1/\lambda$ passes the lattice point hkl with the reciprocal lattice vector $\mathbf{r}^*_{hkl}$ (Figure 12). By the Ewald construction, reflection conditions for cameras and diffractometer are visualized. The reflection of a net plane can be shown by another type of drawing of the momenta (Figure 13). This representation is used to visualize the transfer of the divergence and the bandwidth of the incident beam to the reflected beam.

3.4 Symmetry in Reciprocal Space

In powder diffraction all reflections with equal d_{hkl} (i.e. equal $|\mathbf{r}^*_{hkl}|$) superimpose. The crystal system and the lattice constants determine the values of d_{hkl} and the positions of the powder reflections.

In the diffraction pattern of single crystals anomalous dispersion is not usually observed and Friedel's law is valid, i.e. by the diffraction experiment a center of inversion is introduced into the diffraction pattern; the intensity $I(hkl)$ of the reflection hkl is equal to $I(\overline{hkl})$. Noncentrosymmetric point groups have an additional center of inversion. Thus, only centrosymmetric classes, the 11 Laue classes, can be distinguished by inspecting the diffraction diagram of the single crystal (Table 9). In the case of anomalous dispersion (see section 3.5), Friedel's law is no longer valid and the 32 point groups can be detected by their single-crystal diffraction pattern.

To a primitive lattice in direct space belongs a primitive lattice in reciprocal space. If, however, a centered unit cell is chosen, the cell in direct space is larger than a primitive unit cell, and in reciprocal space it is smaller than the corresponding primitive unit cell. Thus, centering in direct space produces systematic absences in reciprocal space. The reflection conditions are given also in Table 4. The condition $h + k + l = 2n$ for a body-centered lattice I indicates that only reflections with the sum $h + k + l$ equal to an even number are present; the others are absent.

Glide planes give extinctions in reciprocal space. They can be seen in planes in reciprocal space. A glide plane normal to the a-axis, b-axis, or c-axis in direct space has extinctions in reciprocal space in the $0kl$ plane, $h0l$ plane, or $hk0$ plane respectively. For example, in the case of an a glide normal to the c-axis there are only reflections with $h = 2n$ in the $hk0$ plane present. The reflection conditions can be found in Table 6.

Screw axes of direct space give extinctions on the axes in reciprocal space: a 4_1 axis parallel to the c-axis produces $00l$ reflections, of which the reflections $l = 4n$ can be observed; the others with $l \neq 4n$ are extinct (see Table 5).

3.5 The Structure Factor

3.5.1 The Atomic Scattering Factor

The electrons of a single atom form a cloud around the nucleus. The incoming wave excites the electrons to

Table 7 Quadratic formulae

Crystal system	$\frac{1}{d^2}$
Cubic	$\frac{h^2+k^2+l^2}{a^2}$
Hexagonal	$\frac{4}{3}\frac{h^2+hk+k^2}{a^2}+\frac{l^2}{c^2}$
Tetragonal	$\frac{h^2+k^2}{a^2}+\frac{l^2}{c^2}$
Orthorhombic	$\frac{h^2}{a^2}+\frac{k^2}{b^2}+\frac{l^2}{c^2}$
Monoclinic *b*-axis unique	$\frac{h^2}{a^2\sin^2\beta}+\frac{k^2}{b^2}+\frac{l^2}{c^2\sin^2\beta}-\frac{2hl\cos\beta}{ac\sin^2\beta}$
Monoclinic *c*-axis unique	$\frac{h^2}{a^2\sin^2\gamma}+\frac{k^2}{b^2\sin^2\gamma}+\frac{l^2}{c^2}-\frac{2hk\cos\gamma}{ab\sin^2\gamma}$
Triclinic or anorthic	$\frac{1}{D}\left(\frac{h^2}{a^2}\sin^2\alpha+\frac{k^2}{b^2}\sin^2\beta+\frac{l^2}{c^2}\sin^2\gamma+\frac{2hk}{ab}(\cos\alpha\cos\beta-\cos\gamma)\right.$ $\left.+\frac{2kl}{bc}(\cos\beta\cos\gamma-\cos\alpha)\frac{2lh}{ac}(\cos\alpha\cos\gamma-\cos\beta)\right)$ $D=1-\cos^2\alpha-\cos^2\beta-\cos^2\gamma+2\cos\alpha\cos\beta\cos\gamma$
Rhombohedral[a]	$\frac{1}{a^2}\frac{(h^2+k^2+l^2)\sin^2\alpha+2(hk+kl+lh)(\cos^2\alpha-\cos\alpha)}{1-3\cos^2\alpha+2\cos^3\alpha}$

[a] Usually a hexagonal unit cell (R_{hex}) is used.

Table 8 The relations between the moduli of direct space and reciprocal space basis vectors

System	Relation
Cubic	$a^*=\frac{1}{a}$
Hexagonal	$a^*=\frac{2\sqrt{3}}{3a};\ c^*=\frac{1}{c}$
Rhombohedral[a]	$a^*=\frac{\sin\alpha}{aD}$ $\cos\alpha^*=\frac{\cos^2\alpha-\cos\alpha}{\sin^2\alpha}$ where $D^2=1-3\cos 2\alpha+2\cos^3\alpha$
Tetragonal	$a^*=\frac{1}{a};\ c^*=\frac{1}{c}$
Orthorhombic	$a^*=\frac{1}{a};\ b^*=\frac{1}{b};\ c^*=\frac{1}{c}$
Monoclinic *b*-axis unique	$a^*=\frac{1}{\alpha\sin\beta};\ b^*=\frac{1}{b};\ c^*=\frac{1}{c\sin\beta};\ \cos\beta^*=-\cos\beta$
Triclinic or anorthic	$a^*=\frac{\sin\alpha}{aD};\ b^*=\frac{\sin\beta}{bD};\ c^*=\frac{\sin\gamma}{cD}$ $\cos\alpha^*=\frac{\cos\beta\cos\gamma-\cos\alpha}{\sin\beta\sin\gamma}$ $\cos\beta^*=\frac{\cos\alpha\cos\gamma-\cos\beta}{\sin\alpha\sin\gamma}$ $\cos\gamma^*=\frac{\cos\alpha\cos\beta-\cos\gamma}{\sin\alpha\sin\beta}$ where $D^2=1+2\cos\alpha\cos\beta\cos\gamma-\cos 2\alpha-\cos 2\beta-\cos 2\gamma$

[a] It is convenient to use the centered hexagonal unit cell.

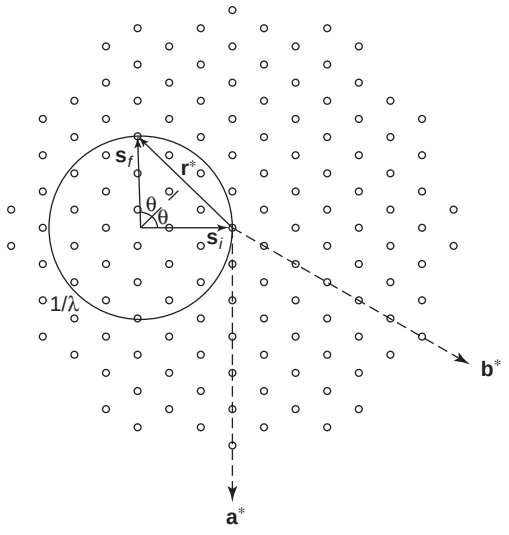

Figure 12 The Ewald construction in a reciprocal lattice plane $hk0$: $\mathbf{s}_i$ is the momentum of the incident beam, $\mathbf{s}_f$ is the momentum of the reflected beam, and $\mathbf{r}^*_{hkl}$ is the vector of the reciprocal lattice. A reflection is only possible if a lattice point hkl with the lattice vector $\mathbf{r}^*_{hkl}$ coincides with the sphere of reflection with the radius $1/\lambda$.

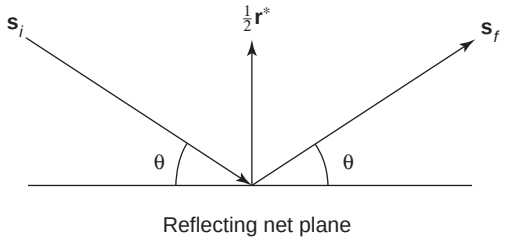

Figure 13 The momentum vectors of the incident beam $\mathbf{s}_i$, the reflected beam $\mathbf{s}_f$ and the normal vector $\mathbf{r}^*_{hkl}$ of the net plane hkl in a different representation.

Table 9 Laue classes

Laue class	Acentric members of the Laue class
$4/m\ \bar{3}\ 2/m$	$4\ 3\ 2, \bar{4}\ 3\ m$
$2/m\ \bar{3}$	$2\ 3$
$6/m\ 2/m\ 2/m$	$6\ 2\ 2, \bar{6}\ 2\ m, 6\ m\ m$
$6/m$	$6, \bar{6}$
$\bar{3}\ m$	$3\ 2, 3m$
$\bar{3}$	3
$4/m\ 2/m\ 2/m$	$4\ 2\ 2, \bar{4}\ 2\ m, 4\ m\ m$
$4/m$	$4, \bar{4}$
$2/m\ 2/m\ 2/m$	$2\ 2\ 2, 2\ m\ m$
$2/m$	$2, m$
$\bar{1}$	1

oscillations. In the direction of the incoming wave, all electrons oscillate in phase. However, with increasing scattering angle θ, i.e. with larger distance $|\mathbf{r}^*|$ from the origin of the reciprocal space, differences in the pathlength occur and the amplitude of the scattered wave is reduced. The atomic scattering factor $f_0(\sin\theta/\lambda) = f_0(\frac{1}{2}|\mathbf{r}^*|)$. For $|\mathbf{r}^*| = 0$, $f_0(0)$ is equal to the number of electrons in the atom. The function decreases rapidly and becomes almost constant at high values of $|\mathbf{r}^*|$ angles (Figure 14).[15]

If the frequency of the incident wave agitates an eigenfrequency of the atom, this simple picture must be modified. The scattering factor changes drastically at the wavelength near an absorption edge, anomalous dispersion occurs, and the scattering factor becomes a complex function. The real part must be corrected by $f'(\lambda)$ and an additional imaginary part $f''(\lambda)$ occurs owing to absorption. Thus, the atomic scattering factor is given by Equation (26):

$$f = f_0\ \ \tfrac{1}{2}|\mathbf{r}^*| + f'(\lambda) + \mathrm{i}f''(\lambda) \tag{26}$$

The atomic scattering factors are calculated by using quantum mechanics, assuming a free spherical atom. It should be noted that $f'(\lambda)$ can assume large negative values near the absorption edge. At a larger distance from the absorption edge, also small positive values of $f'(\lambda)$ occur. In many cases the distance from the absorption edge is so large that the contributions of f' and f'' are neglected. From a synchrotron, radiation with each wavelength can be selected. The anomalous dispersion is now applied in many experiments.

Slight deviations from the model of free spherical atoms occur in solids. Bonding electrons and polarization of the electron clouds in the crystal field deform the spherical symmetry of the clouds of electrons. Thus, the function f is also deformed. This was also detected for f' and f'' in the case of anomalous dispersion by Templeton and Templeton[16] (Figure 15).

The absorption of the X-rays at the absorption edge produces fluorescent radiation. The incident photon pushes out an electron and produces a photo effect, i.e. a hole in the K shell of the anomalously scattering atom. The hole is filled by an electron of an outer shell and fluorescent radiation is produced. This is similar to the production of the characteristic radiation in an X-ray tube. The incident photon must have a higher energy (or smaller wavelength) than that of the fluorescent radiation. This radiation can give a high contribution to the background, which can be avoided by the proper choice of anode material for the X-ray tube (e.g. an iron tube instead of a copper tube when a sample with iron or manganese is studied) or by placing an excellent monochromator in front of the detector.

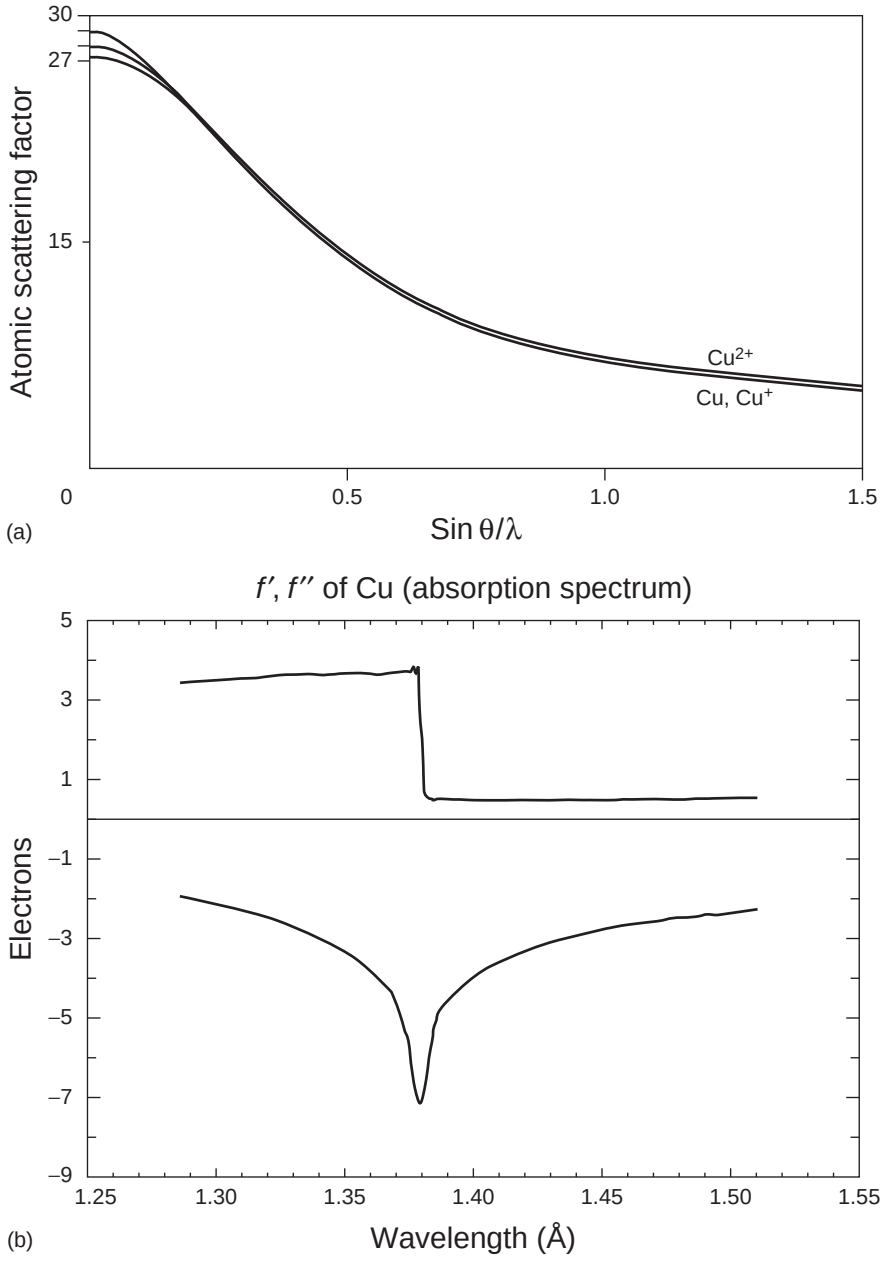

Figure 14 (a) The atomic scattering factor of Cu as a function of $(\sin\theta)/\lambda = \frac{1}{2}|\mathbf{r}^*|$ (from Coppens et al.[34]) (b) Anomalous dispersion: f' and f'' of Cu as a function of wavelength (from Kirfel and Eichorn.[15])

The absorption edge has a fine structure that depends on the electronic state of the atom and the environment of the scattering atom. The analysis of this effect has become a new field of investigation, called extended X-ray absorption fine structure (EXAFS). The structure of the neighborhood of the anomalously scattering atom can be determined. This procedure is also applied in the case of amorphous materials or in samples in which the anomalous scatterer is an interstitial atom.

3.5.2 The Structure Factor

An incoming plane wave excites the electrons of the atoms, and each of the N atoms in the unit cell is a source of a spherical wave. The sum of these waves is given by the structure factor as shown by Equation (27):

$$F(hkl) = \sum_{j=1}^{N} f_j \exp[2\pi\mathrm{i}(hx_j + ky_j + lz_j)] \qquad (27)$$

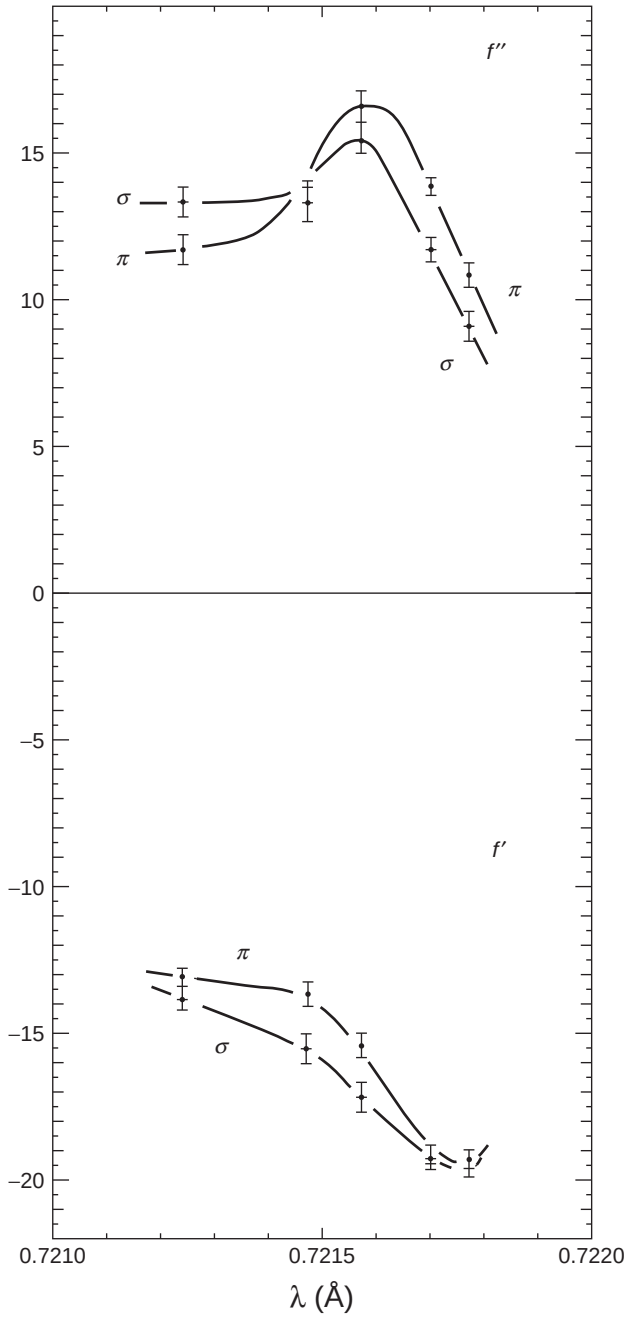

Figure 15 Polarized anomalous scattering of a uranyl ion. The ion is situated on a threefold rotation axis. The index π indicates a direction normal to the threefold axis and σ the direction parallel to the symmetry axis. (Reproduced from Hahn[14] by permission of The International Union of Crystallography.)

where hkl are the Miller indices of the reflecting net plane, x_j, y_j, z_j are the positions, and f_j is the scattering factor (Equation 26) of the jth atom in the unit cell. The exponential function gives the phase differences of the scattered waves with respect to a wave scattered at the origin of the unit cell. The structure factor describes the fact that the reflections hkl differ in amplitude and phase. If both are known, the crystal structure can be calculated by a Fourier transform. From the usual diffraction experiments, only the moduli can be determined. The procedure of structure determination from the moduli is difficult. Complicated methods have been developed that are in most cases successful. There are many textbooks for further reading.[17–30]

For the moment, only one type of atom in one general position in the unit cell is considered. If no anomalous dispersion occurs, the structure factor is given by Equations (28) and (29):

$$F(hkl) = f_0 \sum_{j=1}^{N} \cos[2\pi(hx_j + ky_j + lz_j)] + \mathrm{i}f_0 \sum_{j=1}^{N} \sin[2\pi(hx_j + ky_j + lz_j)] \quad (28)$$

$$= f_0(A + \mathrm{i}B) \quad (29)$$

and A and B are given by Equations (30) and (31):

$$A(hkl) = \sum_{j=1}^{N} \cos[2\pi(hx_j + ky_j + lz_j)] \quad (30)$$

$$B(hkl) = \sum_{j=1}^{N} \sin[2\pi(hx_j + ky_j + lz_j)] \quad (31)$$

The square of the modulus $F(hkl)$ is given by Equation (32):

$$FF^* = f_0^2 A^2 + f_0^2 B^2 \quad (32)$$

For a crystal, all different atomic positions in the unit cell have to be included in the structure factor.

In the general case, $F(hkl)$ is a complex number and its phase angle can assume any value. In the case of a centrosymmetric structure, when a center of inversion is situated at the origin, the formula is reduced. To each atom at x, y, z, an equivalent one exists at $\bar{x}$, $\bar{y}$, $\bar{z}$. Thus, $B(hkl) = 0$, and $F(hkl)$ becomes a real function that has either a positive or a negative sign. In order to solve an acentric structure, the phases of the reflections $F(hkl)$ have to be determined. In the centrosymmetric case the problem is less difficult; only the signs of the reflections have to be assigned to $F(hkl)$.

3.5.3 The Temperature Factor

The atoms in a crystal are not at rest; they oscillate because they are excited by phonons. This is taken into account by the temperature factor. During the time of an experiment, the positions of the atoms are smeared out

in first approximation by a spherical Gaussian function, which is written according to Equations (33) and (34):

$$q(hkl) = \exp\left[-B_j\left(\frac{\sin\theta}{\lambda}\right)^2\right] \tag{33}$$

$$= \exp[-B_j|\tfrac{1}{2}\mathbf{r}^*|^2] \tag{34}$$

where B_j is the isotropic temperature factor of an atom with index j. The physical interpretation of B_j is the mean square amplitude $\langle u_j^2\rangle$ of the jth atom in any direction, which is obtained by Equation (35):

$$\langle u_j^2\rangle = \frac{B_j}{8\pi^2} \tag{35}$$

According to Equation (34), the motion is spherical. This is valid in cubic crystals for atoms in positions with cubic point symmetry only. In general the motion of the atoms is not isotropic. For each atom an ellipsoidal smearing function is introduced. The anisotropic motion is written according to Equation (36):[31]

$$q(hkl) = \exp(-b_{11}h^2 + b_{22}k^2 + b_{33}l^2 + 2b_{23}kl + 2b_{13}hl + 2b_{12}hk) \tag{36}$$

In matrix notation the exponent becomes that shown in Equation (37):

$$-(h,k,l)\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}\begin{pmatrix} h \\ k \\ l \end{pmatrix} = -(khl)\mathbf{b}\begin{pmatrix} h \\ k \\ l \end{pmatrix} \tag{37}$$

where $\mathbf{b}$ is a symmetric tensor describing the anisotropic motion.

The mean square displacement (msd) of the atom in the direction of the vector $\mathbf{m}$ in direct space is obtained by Equations (38) and (39):

$$\mathrm{msd}(\mathbf{m}) = \frac{1}{2\pi^2}\frac{\mathbf{m}'\mathbf{g}'\mathbf{b}\mathbf{g}\mathbf{m}}{\mathbf{m}'\mathbf{g}\mathbf{m}} \tag{38}$$

$$= \frac{1}{2\pi^2}\frac{(m_1\mathbf{a}, m_2\mathbf{b}, m_3\mathbf{c})\begin{pmatrix} b_{11}\mathbf{a}\cdot\mathbf{a} & b_{12}\mathbf{a}\cdot\mathbf{b} & b_{13}\mathbf{a}\cdot\mathbf{c} \\ b_{12}\mathbf{a}\cdot\mathbf{b} & b_{22}\mathbf{b}\cdot\mathbf{b} & b_{23}\mathbf{b}\cdot\mathbf{c} \\ b_{13}\mathbf{a}\cdot\mathbf{c} & b_{23}\mathbf{b}\cdot\mathbf{c} & b_{33}\mathbf{c}\cdot\mathbf{c} \end{pmatrix}\begin{pmatrix} m_1\mathbf{a} \\ m_2\mathbf{b} \\ m_3\mathbf{c} \end{pmatrix}}{(m_1\mathbf{a}, m_2\mathbf{b}, m_3\mathbf{c})\begin{pmatrix} m_1\mathbf{a} \\ m_2\mathbf{b} \\ m_3\mathbf{c} \end{pmatrix}} \tag{39}$$

where $\mathbf{g}$ is the metric tensor (Equation 26).

To interpret the results, the tensor has to be transformed to the main axes in a Cartesian coordinate system. There are excellent computer programs to take care of this and to draw the structure, including the ellipsoids of the anisotropic motion.

Like the positions of the atoms, the temperature factors also have to obey the symmetry of the crystal. If a symmetry operation ($\mathbf{W}$, $\mathbf{w}$) transforms the position of the atom at x, y, z to x', y', z' (Equation 13), the tensor of the anisotropic motion also has to be transformed accordingly. The transformation is performed only by the rotational part $\mathbf{W}$ of the symmetry operation (see section 3.3). The tensor $\mathbf{b}$ is transformed according to Equation (40):

$$\mathbf{b}' = \mathbf{W}\mathbf{b}\mathbf{W}' \tag{40}$$

where $\mathbf{W}'$ is the transposed matrix of $\mathbf{W}$, i.e. in components according to Equation (41):

$$b'_{mn} = \sum_{m=1}^{3}\sum_{n=1}^{3} W_{mk}b_{kl}W_{nl} \tag{41}$$

The anisotropic motion is described by an ellipsoid. It has to fulfill the point symmetry of the position of the atom. On a position with cubic symmetry, an ellipsoid degenerates to a sphere. If the atom is located on a three-, four-, or sixfold rotation axis, the ellipsoid must be rotationally symmetric with respect to this axis. An orthorhombic symmetry requires that the main axes of the ellipsoid coincide with the orthorhombic symmetry axes. On a point with monoclinic point symmetry, the ellipsoid has to be oriented parallel to the mirror plane, or with one axis parallel to the twofold rotation axis. On a general position or on a center of symmetry, there is no restriction to the orientation of the ellipsoid.

Thus, the complete structure factor is given by Equation (42):

$$F(hkl) = \sum_{i=1}^{N}(f_0 + f' + if'')_j q_j(hkl) \times \exp[2\pi\mathrm{i}(hx_j + ky_j + lz_j)] \tag{42}$$

3.5.4 The Electron Density

The Fourier transform of all structure factors gives the electron density, $\rho(X, Y, Z)$, of the unit cell. However, not only the moduli $|F(hkl)|$ have to be known, but in an acentric structure their phases also; in a structure with an inversion center at the origin of the unit cell, the signs of $F(hkl)$ have to be known too. Thus, according to Equation (43):

$$\rho(X,Y,Z) = \frac{1}{V}\sum_{h=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} F(hkl) \times \exp[-2\pi\mathrm{i}(hX + kY + lZ)] \tag{43}$$

where X, Y, Z are the components of a vector $\mathbf{r}$ in the unit cell. By the inverse transformation, the structure factors can be computed from the electron density, as shown in

Equation (44):

$$F(hkl) = \rho(X, Y, Z) \times \exp[2\pi\mathrm{i}(hX + kY + lZ)]\,\mathrm{d}X\,\mathrm{d}Y\,\mathrm{d}Z \quad (44)$$

Because $\rho(X, Y, Z)$ is a continuous function in the unit cell, in the reverse transformation the sums are replaced by integrals. In the map of electron density, each atom is shown with its cloud of electrons smeared out by the temperature motion. Even bonding electrons between the atoms can be seen.

From the experiment, $|F|^2$ is obtained. Its Fourier transform, called the Patterson function, can be calculated as shown by Equation (45):

$$P(X, Y, Z) = \frac{1}{V}\sum_{h=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|F(hkl)|^2 \times \exp[-2\pi\mathrm{i}(hX + kY + lZ)] \quad (45)$$

The Patterson function is the convolution of the electron density functions $\rho(X, Y, Z)$ and $\rho(-X, -Y, -Z)$, as shown in Equations (46) and (47):

$$P(X, Y, Z) = \rho(X, Y, Z) \times \rho(-X, -Y, -Z) \quad (46)$$

$$= \rho(u, v, w) \times \rho(u + X, v + Y, w + Z)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w \quad (47)$$

The Patterson function has interesting properties. It is a periodic function that has the same unit cell as the crystal. Let us consider the atoms j and k in the unit cell, at the positions $\mathbf{r}_j$ and $\mathbf{r}_k$, which have the number of electrons n_j and n_k, respectively; the difference of the position vectors is given by Equation (48):

$$\Delta\mathbf{r}_{jk} = \mathbf{r}_j - \mathbf{r}_k \quad (48)$$

At the position $\Delta\mathbf{r}_{jk} = (\Delta X_{jk}, \Delta Y_{jk}, \Delta Z_{jk})$ with $j \neq k$, the Patterson function has a peak proportional to the product of the numbers of electrons, i.e. the convolution of the electron clouds of atoms j and k, as shown by Equation (49):

$$P(\Delta\mathbf{r}_{jk}) \sim n_j n_k \quad (49)$$

An example is shown in Figure 16.[32] From this function the pair correlation function can be derived, i.e. the number of atoms of kind k at the distance $\Delta\mathbf{r}_{jk}$ from atom j.

At the origin of the unit cell of the Patterson function, each electron cloud of an atom is convoluted with itself, and thus a peak is obtained that is the sum of the squares of the electrons of all atoms in the unit cell, as shown by Equation (50):

$$P(0) \sim \sum_{j=1}^{N} n_j^2 \quad (50)$$

If only one kind of atom is in the unit cell, the self-correlation function can be obtained.

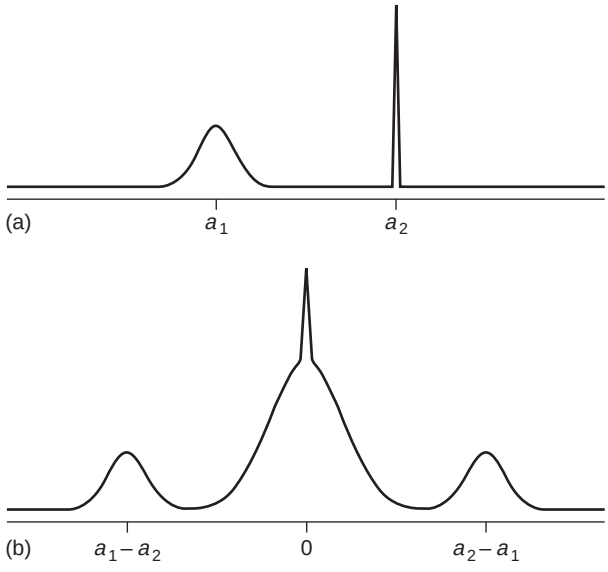

Figure 16 (a) Two Gaussian functions:

$$\rho(x) = \frac{c_1}{\sqrt{\pi b_1^2}}\exp\left(-\left(\frac{x - a_1}{b_1}\right)^2\right) + \frac{c_2}{\sqrt{\pi b_2^2}}\exp\left(-\left(\frac{x - a_2}{b_2}\right)^2\right)$$

(b) Convolution of the function $\rho(x)$ with $\rho(-x)$, according to Equation (47) (from Hosemann and Bagchi[32]):

$$\rho(x) \times \rho(-x) = \frac{c_1^2}{\sqrt{\pi 2b_1^2}}\exp\left(-\frac{x^2}{2b_1^2}\right) + \frac{c_2^2}{\sqrt{\pi 2b_2^2}}\exp\left(-\frac{x^2}{2b_2^2}\right) + \frac{c_1 c_2}{\sqrt{\pi(b_1^2 + b_2^2)}}\exp\left\{-\frac{[x \pm (a_1 - a_2)]^2}{b_1^2 + b_2^2}\right\}$$

3.6 Intensity

The structure factors $F(hkl)$ contain the information of the crystal structure. In order to obtain $|F(hkl)|$, the measured intensities must be corrected for systematic errors owing to the physics of the diffraction. This section begins with a discussion of the correction functions.

3.6.1 The Lorentz Factor

In order to measure the integrated intensity, the reciprocal lattice point must be swept through the sphere of reflection. The rotation axis is usually normal to the reflection plane. All reciprocal lattice points have the

same angular velocity. However, the lattice points are not real points. They show the width of the mosaic distribution. The time in which the reflection sweeps through the Ewald sphere depends on the angle at which the reciprocal lattice point passes the surface of the sphere. This is corrected by the Lorentz factor, as in Equation (51):

$$L = \frac{1}{\sin 2\theta} \tag{51}$$

If the reflection plane is not normal to the rotation axis, the Lorentz factor must be modified.

3.6.2 The Polarization Factor

Neutrons have no polarization and their polarization factor is given by Equation (52):

$$P_n = 1 \tag{52}$$

X-rays, however, are electromagnetic waves that can be polarized. The plane of diffraction is defined by the vectors of propagation of the incoming beam $\mathbf{s}_i$ and of the diffracted beam $\mathbf{s}_f$. The component of the polarization vector of the incoming beam, which is normal to the diffraction plane $\mathbf{s}_i$, $\mathbf{s}_f$, transfers unchanged to the diffracted beam. The component of the polarization vector of the incoming beam, which is parallel to the plane $\mathbf{s}_i$, $\mathbf{s}_f$, contributes to the diffracted beam only by its projection, i.e. with the factor $\cos 2\theta$. If the diffracting crystal is of perfect quality and reflects dynamically, the polarization factor is given by Equation (53):

$$P_{\mathrm{D}} = \tfrac{1}{2}(1 + |\cos 2\theta|) \tag{53}$$

However, Equation (54) gives the polarization factor in the case of a mosaic crystal, when intensities are added.

$$P_{\mathrm{M}} = \tfrac{1}{2}(1 + \cos^2 2\theta) \tag{54}$$

The X-ray tube produces an unpolarized radiation. Using a monochromator, the radiation becomes elliptically polarized. If the monochromator reflects as a mosaic crystal, the polarization factor becomes that shown in Equation (55):

$$P_{\mathrm{MK}} = \frac{1 + \cos^2 2\alpha_{\mathrm{M}} \cos^2 2\theta}{1 + \cos^2 2\alpha_{\mathrm{M}}} \tag{55}$$

according to kinematic theory. Otherwise, from dynamic theory, the polarization factor becomes that shown in Equation (56):

$$P_{\mathrm{MD}} = \frac{1 + |\cos 2\alpha_{\mathrm{M}}| \cos^2 2\theta}{1 + |\cos 2\alpha_{\mathrm{M}}|} \tag{56}$$

where α_{M} is the Bragg angle of the monochromator crystal.

A synchrotron produces almost perfect polarized radiation and its polarization factor depends on the take-off angle from the bending magnet or insertion device. The position of the electron beam is not stable, therefore the polarization is monitored by two counters positioned perpendicular to the incoming beam: one counter in the diffraction plane of the experiment, and the other normal to that plane. In this case, the polarization factor becomes that shown in Equation (57):

$$P_{\mathrm{s}} = \frac{c_{\mathrm{n}} + c_{\mathrm{p}} \cos^2 2\theta}{c_{\mathrm{n}} + c_{\mathrm{p}}} \tag{57}$$

where c_{n} is the counting rate of the counter normal to the diffraction plane and c_{p} is that within the plane. The equation assumes that the sample diffracts at the Bragg angle θ as a mosaic crystal.

3.6.3 The Geometric Factor

There are different cameras and diffractometers in use. Usually, the length of the beam path does not depend on the Bragg angle. However, when the length between crystal and counter (or film) varies with the Bragg angle, it has to be taken into account by the geometric factor G. In the case of single crystals measured by a four-circle diffractometer, the beam path is independent of θ and G is given by Equation (58):

$$G = 1 \tag{58}$$

The function is more complicated in the case of the Guinier powder diffraction, equi-inclination, or precession method.

In the case of powder diffraction the powder should have no preferred orientation (if the texture is not to be investigated). The vectors normal to the reflecting net plane *hkl* are distributed over a sphere in reciprocal space with the radius $1/d_{hkl} = |r^*|$. Only the intersection of this sphere with the sphere of reflection fulfills the reflection condition (Ewald construction). The reflected intensity of the total Debye–Scherrer cone varies with $\frac{1}{2}\cos\theta$, but only a small section of the cone is recorded, proportional to $1/2\pi r \sin 2\theta$. Thus, the geometric factor becomes that shown in Equation (59):

$$G_{\mathrm{DS}} = \frac{\cos\theta}{4\pi r \sin 2\theta} \tag{59}$$

where r is the distance between the sample and the detector.

3.6.4 Absorption Corrections

The intensity I_0 of an X-ray beam passing through a material is reduced by absorption, mainly by the photo effect. After a ray has passed through a sheet of thickness

t (more precisely, a path with length t), the intensity is reduced to that shown by Equation (60):

$$I = I_0 A_L \tag{60}$$

and the linear transmission factor A_L is given by Equation (61):

$$A_L = \exp(-\mu t) \tag{61}$$

where μ is the linear absorption coefficient. In the case of a Bragg–Brentano diffractometer, the transmission factor becomes that expressed in Equation (62):

$$A_{BB} = \frac{1}{2\mu} \tag{62}$$

This correction is important in the case of quantitative analysis of the composition of powders.[33]

The correction of absorption becomes tedious for strongly absorbing single crystals or cylindrical samples in the Debye–Scherrer case. There are sophisticated computer programs that calculate the absorption factor numerically by integrating the possible beam paths.[34] Usually small samples are used, to avoid an absorption correction.

The linear absorption coefficient μ can be calculated from the chemical composition of the sample by using the mass absorption coefficients (μ_i/ρ_i) of the chemical element i, according to Equation (63):

$$\frac{\mu}{\rho} = \sum_i \alpha_i \frac{\mu_i}{\rho_i} \tag{63}$$

where ρ is the density of the sample and α_i is the weight fraction of element i in the sample.

3.6.5 The Lattice Factor

The structure factor takes only the contents of a unit cell into account. The diffracted beams of all unit cells in the crystal add with equal phase. Let us assume a crystal with M_a unit cells in the direction of the lattice constant **a**. The unit cells are labeled by m, such that $0 < m < M_a - 1$. Each unit cell contributes to the diffracted beam with a phase factor $\exp(2\pi i m h)$. The sum over all unit cells is a geometric progression, as shown by Equation (64):

$$\sum_{m=0}^{M-1} \exp(2\pi i m h) = \frac{1 - \exp(2\pi i M_a h)}{1 - \exp(2\pi i h)} = \frac{\sin \pi M_a h}{\sin \pi h} \exp[\pi i (M_a - 1) h] \tag{64}$$

The lattice factor (Equation 65) is the square of the modulus of this function in three dimensions, i.e.:

$$|H|^2 = \frac{\sin^2 \pi M_a h}{\sin^2 \pi h} \frac{\sin^2 \pi M_b k}{\sin^2 \pi k} \frac{\sin^2 \pi M_c l}{\sin^2 \pi l} \tag{65}$$

where h, k, and l are the Miller indices (i.e. the components of a lattice vector in reciprocal space) and M_a, M_b, and M_c are the numbers of unit cells in the direction of the lattice constants a, b, and c. These numbers are usually very large. Thus, $|H|^2$ is essentially zero in reciprocal space and $N^2 = M_a M_b M_c$ at the position of a lattice point hkl; N is the number of unit cells in the crystal.

With small M values, the function H becomes a broad maximum. This can be used to determine the particle size by using the Scherrer equation (Equation 6).

3.6.6 The Multiplicity Factor

In powder diffraction all net planes with an equal value of $1/d_{hkl}$ contribute to the same powder line. The number of contributing reflections is determined by the crystal system and the Miller indices hkl of the reflecting net planes. The factor M is given by the number of symmetrically equivalent net planes hkl contributing to the reflection hkl. For example, $M = 6$ reflections contribute to a cubic powder line $h00$ according to the six faces {100} of the cube; $M = 48$ reflections contribute to a reflection hkl of the cubic holoedrie.

3.6.7 The Formula for the Intensity

The intensity is measured by sweeping the reciprocal lattice point hkl through the sphere of reflection and integrating the reflected radiation during the motion. This value, corrected for the overall background, is the integrated intensity, or, in short, the intensity. It is denoted by $I(hkl)$.

The interference of waves is responsible for diffraction. The amplitudes of the waves add in accordance with the phases. Thus, the diffracted intensity is proportional to $|F(hkl)|$, the amplitude of the waves. This is valid for perfectly grown crystals, such as silicon used in electronics. This is called *dynamic theory*, because it takes into account all the interactions of the waves in the perfect crystal.

The conditions for this effect are very sensitive. Crystals that were grown fast contain a great number of slight deformations of the crystal lattice, produced by defects such as dislocations and faults. They disturb the interference of the waves. Small regions in the crystal diffract independently of their environment. The small regions act like the stones of a mosaic. In this case, it is assumed that each individual mosaic block is slightly misoriented and interacts with an unweakened ray of the source. This is the *kinematic theory*. The intensity is proportional to the square of the modulus of $F(hkl)$, i.e. FF^*.

Unfortunately there are crystals that are neither perfect nor mosaic. In this case kinematic theory is applied and (strong) reflections, which have a reduced intensity, are corrected in a heuristic manner for ''primary extinction''. Sometimes, even powders have to be corrected. Thus,

the following formulae are extreme cases. In practice, kinematic theory is applied except in the case of really perfect crystals. The secondary extinction can be reduced by quenching a crystal in liquid nitrogen.

The measured intensity $I(hkl)$ in the case of kinematic theory is given by Equation (66):

$$I(hkl) = K^2\lambda^3 P_M LGMA|F|^2|H|^2 \tag{66}$$

and K is given by Equation (67):

$$K = \frac{e^2}{mc^2} \tag{67}$$

which involves the charge e and mass m of an electron and the velocity of light c.

In the case of dynamic theory, $I(hkl)$ is given by Equation (68):

$$I(hkl) = \frac{8}{3\pi} K\lambda^2 P_D LGMA|F||H| \tag{68}$$

Usually the proportionality factor, the product of K and λ, is fitted as an independent parameter by the least-squares method. It is difficult to measure the intensity absolutely.

3.7 Diffuse Background

The diffuse background is produced by phonons and by defects in the crystal lattice. The intensity found in the background reduces the intensities of the Bragg reflections. The temperature factor assumes a statistically independent motion of the atoms owing to thermal motion. However, the phonons are waves traveling through the crystal with a broad spectrum of wavelengths that are multiples of the unit cell dimensions. They form superlattices in direct space that can be seen as a spectrum of very small peaks in reciprocal space. In other words, the reciprocal space is filled with intensity obtained by the diffraction of phonons. In general, only the phonons with low frequencies (i.e. the acoustic phonons) contribute to the diffuse background. Their intensity is concentrated around the Bragg reflections.

The defects of a crystal (e.g. stacking faults, interstitial atoms, etc.) also contribute to the diffuse background. In a crystal with defects, a mean unit cell is defined that is the average of the unit cells of the whole crystal. This is the periodic contribution to the electron density function of the whole crystal. The aperiodic contribution to the electron density is made by the defects.

By the mean unit cell, a mean structure factor $\langle F(hkl)\rangle$ is defined. The three-dimensional periodic mean unit cell gives a sharp Bragg reflection with an intensity $I_B(hkl)$ according to Equation (69):

$$I_B(hkl) \sim \langle F(hkl)\rangle^2 |H|^2 \tag{69}$$

where $|H|^2$ is the lattice function (Equation 65). The aperiodic contribution of the electron density creates a diffuse background $I_D(r^*)$ in the entire reciprocal space, according to Equation (70):

$$I_D(r^*) \sim \langle |F(r^*)|^2\rangle - \langle F(r^*)\rangle^2 \tag{70}$$

i.e. the intensity of the diffuse background is given by the fluctuation of the structure factor.

Gases, liquids, and amorphous materials give a diffuse and modulated diffraction pattern. Sharp reflections cannot be detected and a unit cell does not exist. The intensity is given by Equation (71):

$$I(\theta) = \sum_u \sum_v f_u f_v \frac{\sin k\Delta r_{uv}}{k\Delta r_{uv}} \tag{71}$$

where $v = (4\pi \sin\theta)/\lambda$, and $\Delta r_{uv} = |r_u - r_v|$ is the magnitude of the interatomic distance between atoms u and v. Similarly to the Patterson function (Equation 45), a pair correlation function can be determined and in simple cases can be interpreted.[35]

4 PRODUCTION OF X-RAYS AND THEIR CHARACTERISTICS

4.1 The X-ray Tube

An X-ray tube is shown schematically in Figure 17. The X-rays are usually produced in an evacuated tube. The cathode is a tungsten filament, heated to emit

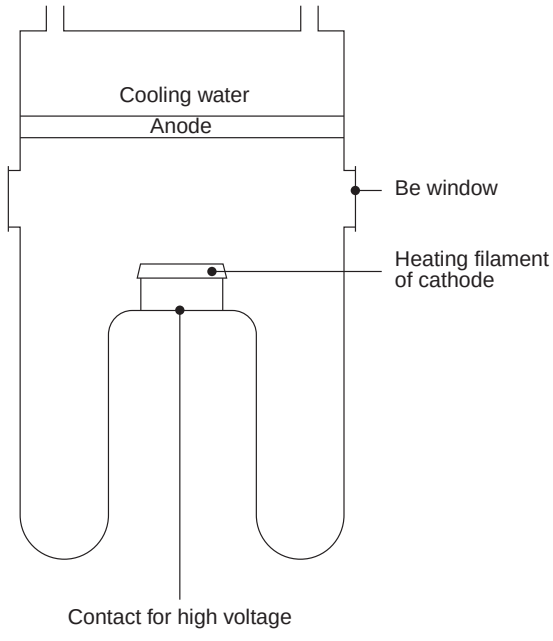

Figure 17 Schematic drawing of an X-ray tube.

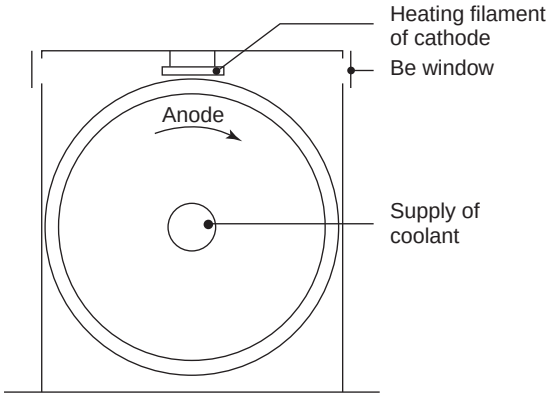

Figure 18 Schematic drawing of a rotating anode.

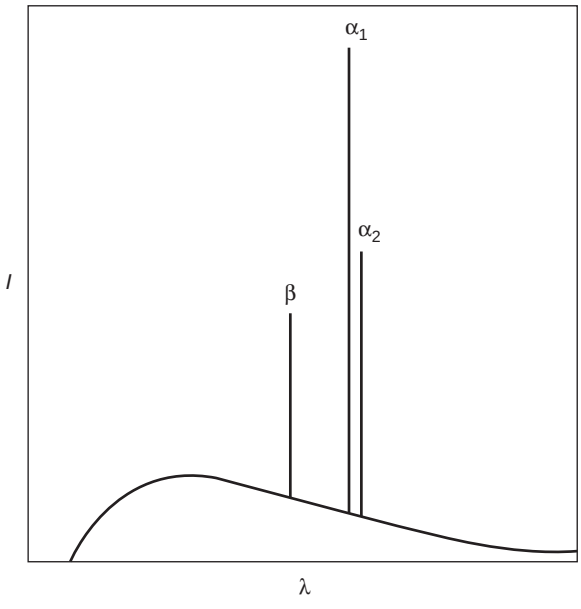

Figure 19 Schematic drawing of the X-ray spectrum emitted by an X-ray tube. On top of the continuous spectrum is the Kα, Kβ characteristic spectrum.

electrons. The cathode is at a negative potential in the range −5 to −60 kV. The anode is grounded at ground potential. Under this potential difference, the electrons are accelerated and hit the anode. Most of the energy heats the anode, which must be cooled by water. To increase the heat load on the anode without melting, the rotating anode was developed. The X-rays are produced in the anode (Figure 18). There are mainly two effects that produce X-rays in the anode with high intensity (Figure 19).

4.1.1 The Characteristic Radiation

The production of X-rays is the consequence of a collision between an accelerated electron and an electron bonded in the atom. The bonded electron is pushed out of the shell and expelled into the conduction band, leaving a hole in the shell. After this, the hole is refilled by an electron from an outer shell. The gain in potential energy is emitted as radiation. Because the electrons occupy precise potentials in the atom, a discrete spectrum is produced. In the case of X-rays, the hole must be produced in the innermost shell, the K-shell. The α-radiation results from the transition of an electron from the next shell: the L-shell. The β-radiation results from the transition of an electron from the next highest shell: from the M-shell to the K-shell. The Kα radiation is a doublet. The two lines indicate that the transition can be associated by a spin flip. The radiation is characteristic of the material of the anode.

4.1.2 Bremsstrahlung

The bremsstrahlung can be treated by classical electrodynamics. Each electric charge accelerated in a field emits radiation. The radiation is proportional to sin 2δ, where δ is the angle between the velocity vector of the charge and the vector of acceleration, i.e. the radiation is maximal normal to the direction of the acceleration. The electrons entering the anode are accelerated and deflected by the electric field of the atoms. This is a statistical process. When the electrons decelerate, they lose energy. The lost energy is emitted in the form of a continuous X-ray spectrum in all directions. The spectrum ends at low wavelengths when the total kinetic energy of an electron of the X-ray tube is converted into one phonon. The limiting wavelength λ_l (in Å) is given by Equation (72):

$$\lambda_l = \frac{12.398}{U} \tag{72}$$

where U is the voltage of the X-ray tube (in kV). This relation is important if one wants to obtain a spectrum free of $\lambda/2$ or $\lambda/3$ of the characteristic radiation.

4.2 Synchrotron Radiation

A synchrotron is an accelerator for electrons. As a source for synchrotron radiation, storage rings are used, i.e. a ring in which electrons (or positrons) are kept moving. The energy of the electrons is in the range 1–5 GeV, which is a factor of 10^6 larger than in the X-ray tube. A bending magnet is needed to keep the beam of electrons on the path of the ring. Each magnet accelerates the beam radially, and thus the bremsstrahlung is created. However, owing to the high kinetic energy, relativistic electrodynamics must be applied. The characteristics of synchrotron radiation are very different from those of the bremsstrahlung of the X-ray tube.

5 DIFFRACTOMETERS AND CAMERAS

5.1 The Resolution Function

The fwhm of a reflection or a powder line is a function of the angle θ (Figure 20). It depends on the diffractometer, i.e. the geometry of the sample, the divergence of the primary beam, and the bandwidth $\Delta\lambda/\lambda$ of the X-rays. The function in Equation (73) was found experimentally[36,37]

$$\text{fwhm} = U\tan^2\theta + V\tan\theta + W \qquad (73)$$

where U, V, and W are parameters that can be fitted by a least-squares procedure. The parameter W is essentially defined by the cross-section of the incident beam or the diameter of the sample. The parameter V, which is a negative number, depends on the divergence of the primary beam together with the width of the mosaic spread of the sample, and U depends mainly on the wavelength resolution $\Delta\lambda/\lambda$. The resolution function is valid for almost all diffractometers but the coefficients vary even from sample to sample.

5.2 Single-crystal Methods

In order to obtain the full information contained in an X-ray pattern, the intensities of all reflections *hkl* have to be determined. This can be done only by the single-crystal methods. In a powder diffraction pattern or a rotation photograph, reflections usually superimpose. The photographic methods were developed to resolve superimposed reflections and facilitate the indexing of reflections. From the measurements, i.e. the size of the unit cell, the symmetry elements determining the Laue class, and the reflection conditions, the reciprocal lattice is obtained. In rare cases the space group can be determined uniquely by this information. The measurement is completed when each reflection is labeled by the indices *hkl*; its intensity profile is measured and integrated. If the symmetry is known, one may measure only the intensities in the asymmetric unit of reciprocal space and save the time of measurement of symmetrically equivalent reflections. However, the assumption of the symmetry may be biased or the sample may be twinned.

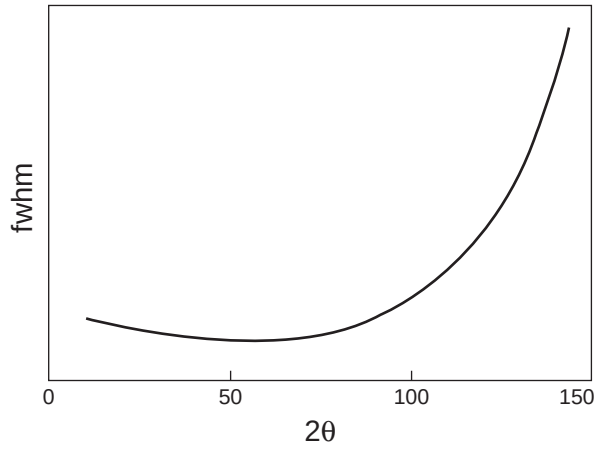

Figure 20 The shape of the resolution function (Equation 73).

In the next section a short introduction is given to the different experimental methods by using the Ewald construction (section 3.3) and the reciprocal lattice. It should be noted that in the drawings the reciprocal lattice is exaggerated (large) in comparison with the dimensions of the screens, film cylinders, etc. found in the same figure. It is the angles that should be visualized, not the distances.

5.2.1 The Laue Method

The Laue method uses the continuous X-ray spectrum. The region of wavelengths λ is limited at the short end of the spectrum by λ_{min}, which is governed by the high voltage applied to the X-ray tube (Equation 72). At the long end, the spectrum is limited by λ_{max}, owing to the absorption of the X-rays on the beam path by the window of the tube, by the air, or within the sample.

In Figure 21 the Ewald construction of the Laue method is shown schematically. The crystal and its reciprocal lattice are stationary. The direction of the incident beam $\mathbf{s}_i/|\mathbf{s}_i|$ within the reciprocal lattice is defined by the orientation of the crystal with respect to the incident beams. The end-points of the vectors $\mathbf{s}_i$ are fixed at the origin of the reciprocal lattice, that is, at the lattice point of the reflection $h, k, l = 0, 0, 0$. The moduli of the vectors $\mathbf{s}_i$ are $1/\lambda$. According to the range of wavelengths, $1/\lambda_{min}$ to $1/\lambda_{max}$, the vectors $\mathbf{s}_i$ start on the line that is enclosed in Figure 21 by a rectangle. Each point on the line is the center of a sphere of reflection with radius $1/\lambda$. Thus, all lattice points between the spheres with radii $1/\lambda_{min}$ and $1/\lambda_{max}$ reflect simultaneously.

A special case is the reflections that originate from lattice points lying on a line passing through the origin. In Figure 21 two vectors, $\mathbf{s}_{f1}$ and $\mathbf{s}_{f2}$, are drawn that fulfill this condition. The two lattice points are situated in the region between the limiting spheres with the radii $1/\lambda_{min}$ and $1/\lambda_{max}$. The vectors $\mathbf{s}_{f1}$ and $\mathbf{s}_{f2}$ are parallel. The reflection angle 2θ is equal in both cases. Thus, both lattice points reflect simultaneously in the same direction. The reflections, created by different wavelengths, superimpose on an X-ray film. By using an energy-dispersive detector the two reflections can be separated.

The Laue method is used to adjust crystals. The crystal, glued on a glass fiber, is fixed on a goniometer head that has two sliding carriages perpendicular to each other to center the crystal; the goniometer head also has two arcs perpendicular to each other, to turn the crystal into the

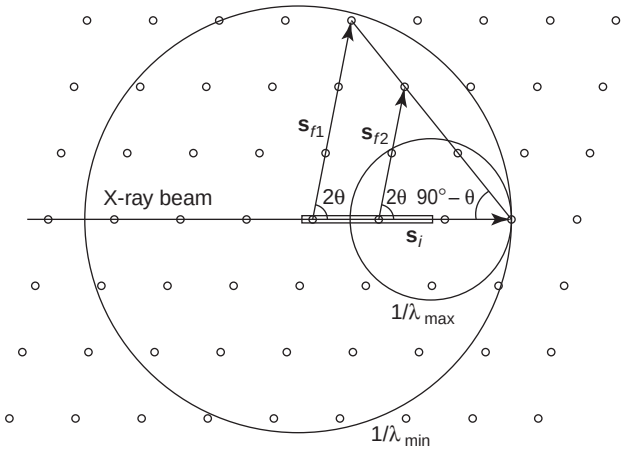

Figure 21 The Ewald construction of a Laue diagram. The range of wavelength is limited by two spheres of reflection with radii $1/\lambda_{min}$ and $1/\lambda_{max}$. All lattice points between the two spheres fulfill the reflection condition. The vectors $\mathbf{s}_{f1}$ and $\mathbf{s}_{f2}$ indicate the momenta of two reflected waves represented by two lattice points lying on a line passing through the origin of reciprocal space. The vectors $\mathbf{s}_{f1}$ and $\mathbf{s}_{f2}$ are parallel and reflect under the same diffraction angle 2θ. Thus, the reflections superimpose.

desired orientation. In a Laue pattern, mirror planes and rotation axes can be detected. The crystal is adjusted so that a symmetry plane is normal to the axis of the goniometer head and a rotation axis coincides with the incident beam. From Laue photographs, taken from the symmetry directions of the crystal, the Laue class can be determined. The adjusted crystal can be used in further investigations.

Laue diagrams usually are not used to determine the intensities of reflections. In the spectrum of an X-ray tube, the intensities vary too much with the wavelength. This is different in the case of synchrotron radiation, where the intensity is almost constant over a large range of wavelengths. By using two-dimensional position-sensitive detectors, the intensities of many reflections are determined simultaneously. The Laue method is of great advantage in protein crystallography, where the samples of biological material have only a short lifetime. The intensities of many reflections can be measured simultaneously in a short time and only a few reflections superimpose.

5.2.2 Monochromatic Methods

The indexing of the reciprocal lattice is simplified if one uses monochromatic radiation. Using an X-ray tube, it is often sufficient to remove the Kβ radiation by a filter. When filtered radiation is recorded by a film, one can observe the reflections of the Kα radiation on top of the streaks of the continuous spectrum. By using a monochromator (i.e. reflection by a net plane of a single crystal), monochromatic radiation is obtained. This radiation can be spoiled by higher harmonics of the wavelength (i.e. $\lambda/2$ and $\lambda/3$), which give additional reflections on a film. This radiation can be removed by reducing the high tension of the X-ray tube (Equation 72) or by using a double-crystal monochromator.

5.2.2.1 The Four-circle Diffractometer The four-circle diffractometer is the most important instrument to use for measuring the intensities from single crystals. One circle is used to position the detector at the Bragg angle $2\theta_{hkl}$; usually this circle is horizontal. Three additional circles are needed to turn a reciprocal lattice point into the diffraction plane, the plane of the 2θ-circle, and to sweep the lattice point through the reflection circle.

There are two types of instruments on the market:

1. The four circles of the Eulerian cradle are 2θ, w, χ, and φ. The cradle is shown in Figure 22(a). The 2θ-axis, supporting the detector, coincides with the w-circle. The χ-circle is turned around the w-axis. The φ-axis, mounted on a carriage, can slide on the χ-circle. The Eulerian cradle is mechanically very stable. Heavy furnaces or cryostats can be attached to the χ-circle opposite the position of the φ-circle. A disadvantage of the Eulerian cradle is that at certain positions the beam path is interrupted by the shadow of the χ-circle.
2. The kappa diffractometer is shown in Figure 22(b). The χ-circle of the Eulerian cradle is replaced by the κ-circle. Its axis is inclined by 50° to the w-axis. A motion of the χ-circle of the Eulerian cradle is simulated by combined rotations of the w-, κ-, and φ-circles. The advantage of the kappa diffractometer is that almost no shadows occur.

Intensities are measured in two ways. In the first method, the 2θ-circle and the w-circle scan simultaneously, the latter with half-speed (θ, 2θ scan). In reciprocal space, this corresponds to a motion along a line passing through the origin. The scan follows the trace of the reflections produced by the white radiation of the X-ray tube. The contribution of the white radiation is subtracted as background from the intensity of the Kα reflection. In the second method, the detector (with a sufficiently opened entrance slit) is kept stationary and the crystal is turned by the w-circle. In reciprocal space, this corresponds to a motion of the reciprocal lattice point on a circle with radius $r^* = 1/d_{hkl}$. This method is applied in the case of the background radiation produced by the material around the sample, e.g. a capillary. In this case the intensity of the material is subtracted as background from the measured intensity.

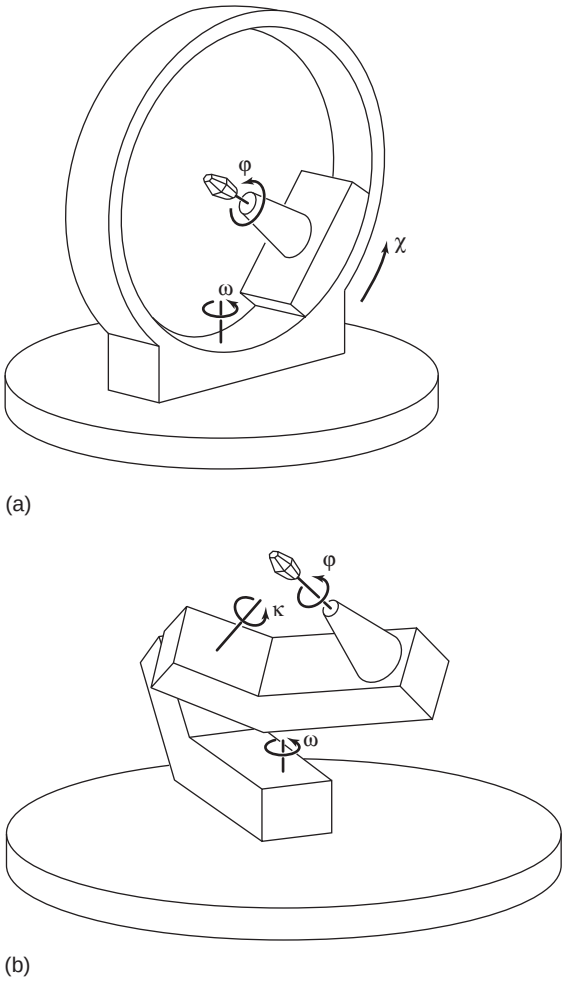

Figure 22 Schematic drawings of two species of four-circle diffractometers. The 2θ-circle, which supports the detector, is omitted. The 2θ-axis coincides with the w-axis. The crystal is rotated: (a) in the Eulerian cradle by the three angles w, χ and φ; (b) in the kappa diffractometer by the angles w, κ and φ.

The four-circle diffractometer is a universal instrument. With an x,y sliding carriage on the φ-circle, the texture on the surface of large platelet samples is also investigated.

The adjustment of a crystal on the diffractometer has been described for an instrument with a Eulerian cradle. In the case of a kappa diffractometer the procedure is similar. It is advantageous to know the lattice constants, the indices of strong reflections, and the approximate orientation of the sample at the start of the adjustment. The procedure is easy in the case of a sample with an orthogonal unit cell. The angle χ is set to 90°. Let the c-axis of the crystal be approximately parallel to the φ-axis of the diffractometer. A reflection 001 of a lattice plane at known 2θ, which is expected to be normal to the φ-axis, can be found by scanning the w-circle. The offset of the crystal from the w-axis is stored as an offset of the χ-circle in the computer. In a second step, the φ-circle is turned by 90° and the procedure is repeated. The offset is stored as an offset of the w-axis. Now the orientation of the c-axis with respect to the φ-circle is known. The χ-axis is turned by 90° and reflections are searched in the plane perpendicular to the φ-axis. This defines the zero point of the φ-circle. The adjustment is correct when addition reflections are found. In a special procedure, the positions of a set of reflections are measured automatically to refine the orientation matrix by least squares, i.e. the orientation and the measured lattice constants. The adjustment of a crystal with an anorthic unit cell is more difficult.

There are computer programs for the automatic search of reflections. However, reflections with low intensity are easily overlooked. This may result in too large a unit cell in reciprocal space, i.e. too small a unit cell in direct space. Also, twinning is difficult to detect by the use of a four-circle instrument, therefore each sample should be investigated in advance by photographic methods to determine the lattice constants and, as far as possible, the space group. The crystal, already oriented on a goniometer head, can be transferred easily from the photographic camera to the diffractometer and it is easy to adjust the small deviations that may have occurred during the transfer.

5.2.2.2 Photographic Methods Photographic methods permit the planes of reciprocal space to be investigated and indexed. For the rotation or Weissenberg method, a lattice plane is adjusted normal to the rotation axis, i.e. the axis of the goniometer head. Planes normal to the rotation axis are investigated. However, the precession method investigates planes parallel to the rotation axis. Usually the Weissenberg photograph is made with Cu Kα radiation. The precession pattern is obtained by Mo Kα radiation so that the radii of reciprocal space, explored by the two methods, are approximately equal and the two methods are complementary. A crystal with an orthogonal unit cell, adjusted in a Weissenberg camera, can be transferred immediately to a precession camera, and vice versa. It is easier to adjust a crystal in a Weissenberg camera than in a precession camera. The great advantage of the precession camera is to give an undistorted picture of reciprocal lattice planes.

5.2.2.3 The Rotation Method The cross-section of a rotation camera is given in Figure 23. Usually, the radius of the cylindrical film is $r_F = 90°/\pi = 28.649\,\text{mm}$. One millimeter on the circumference of the film corresponds to an angle of 2°. A single crystal mounted on a thin glass rod on the goniometer head is placed at the middle of the film cylinder. The crystal is rotated and the rotation axis is perpendicular to the incident beam. A rotation photograph is shown schematically in Figure 24.

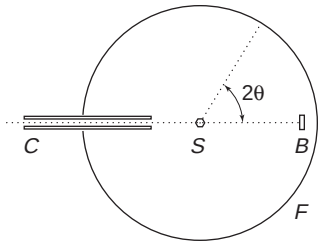

Figure 23 Cross-section of a rotation camera: *C*, collimator; *S*, sample; *B*, beam stop; *F*, cylindrical film or position-sensitive detector or circle on which the entrance slit of a detector scans.

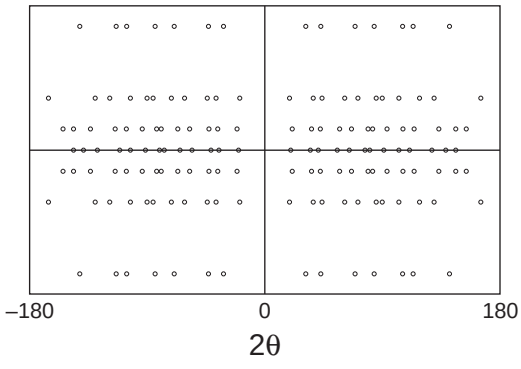

Figure 24 Schematic drawing of a rotation photograph of a hexagonal crystal: *c*-axis vertical.

The crystal is adjusted if the normal of a plane of lattice points is parallel to the rotation axis, i.e. the incident beam passes this plane of lattice points independently of the rotation angle w. For example, if the c-axis of the crystal is adjusted so that it coincides with the rotation axis, the plane $hk0$ is normal to the cylinder axis of the camera and coincides with the incident beam. On a photograph with a well-adjusted crystal, the reflections appear on layers ordered according to the aligned axis. In the example with the c-axis parallel to the rotation axis, the layers are $hk0$, $hk1$, $hk2$, etc.

The reflection geometry can be seen in Figure 25. On the X-ray film, the distance of the nth layer line from the zero layer is y_n. The angle ν_n is obtained by Equation (74):

$$\tan \nu_n = \frac{y_n}{r_F} \tag{74}$$

where r_F is the radius of the film cylinder. The distance d_n^* of the nth net plane to the zero plane in reciprocal space normal to the rotation axis is given by Equation (75):

$$d_n^* = \frac{1}{nt} \tag{75}$$

where t is the translation of the crystal in the direction of the adjusted axis. From Figure 25 Equations (76) and (77) can be derived:

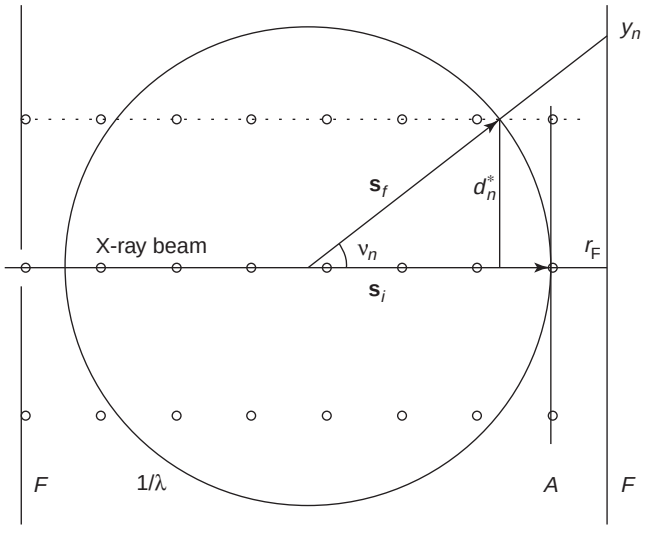

Figure 25 Schematic drawing of the geometry of a rotation photograph of an nth layer in reciprocal space (dotted line): the recorded net planes are perpendicular to the drawing plane; the sphere of reflection is stationary and has a radius $1/\lambda$, A, the rotation axis by which the reciprocal lattice is turned through the sphere of reflection; F, cylindrical film with the radius r_F; y_n, the distance of the nth layer line from the zero layer line on the film; d_n^*, distance of the nth layer from the zero layer in reciprocal space; and ν_n, the angle between the circle obtained by intersection of the nth layer line with the sphere of reflection, and the zero layer line.

$$d_n^* = \frac{1}{\lambda} \sin \nu_n \tag{76}$$

$$= \frac{1}{\lambda} \frac{y_n}{\sqrt{r_F^2 + y_n^2}} \tag{77}$$

Thus, the lattice constant of the adjusted axis can be determined from the rotation photograph directly by Equation (78):

$$t = \frac{\lambda}{n y_n} \sqrt{r_F^2 + y_n^2} \tag{78}$$

As in a powder diffraction pattern, reflections superimpose within the layer lines. Owing to the rotation, there are two mirror planes in the photograph; the vertical plane is produced by the rotation, and the horizontal plane relates reflections hkl with reflections $hk\bar{l}$. Owing to this effect, the symmetry information is obscured.

It is difficult to index a set of reflections if the crystal is not properly adjusted. The procedure for aligning a crystal is described in the legend of Figure 26.

5.2.2.4 The Weissenberg Method The Weissenberg method was developed to separate the superimposed

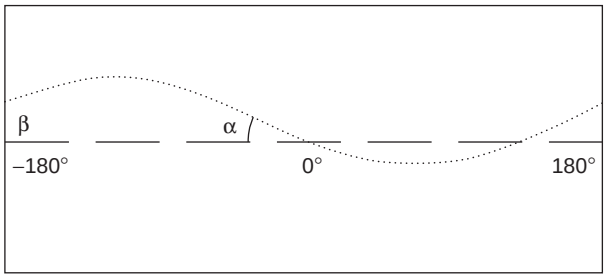

Figure 26 A zero layer line misoriented by the angles α and β. One arc of the goniometer head is placed parallel and the other arc is placed perpendicular to the incident beam. From the trace of lattice plane obtained by a Laue photograph in combination with a small oscillation Δw of the crystal, the two correction angles α and β can be determined. The angle α is read out from the photograph directly. The angle β is obtained by the deviation d of the lattice plane at $2\theta = 180°$, i.e. $\tan\beta = d/2r_F$, where r_F is the radius of the film cylinder; usually $\tan\beta = d/57.296$.

reflections within the layers of a rotation photograph. For each layer a Weissenberg picture has to be taken. The Weissenberg camera is a rotation camera (as shown in Figure 23) in which a cylindrical layer-line screen is inserted between the crystal and the film, which is shifted while the crystal is rotated. For nonzero layer photographs, the camera is inclined with respect to the incident beam by the angle μ_e.

5.2.2.5 The Zero Layer At first the technique is explained for the zero layer, which is seen on the equator of a rotation photograph. The incident beam is normal to the rotation axis. A cylindrical screen permits only the rays of the zero layer line to pass to the cylindrical film (Figure 27). Simultaneously with the rotation of the crystal by the angle w, the film is shifted parallel to the rotation axis. After a rotation of more than 180°, the rotation of the crystal and the motion of the film are reversed. On the film, the position of a reflection is recorded as a function of w and 2θ (Figure 28). The example shown in Figure 28 is the picture of a hexagonal $hk0$ plane, which is shown in Figure 11.

The shift of the film is coupled with the rotation angle by $2°\,\mathrm{mm}^{-1}$. A straight line passing through the origin of reciprocal space is recorded on the film also as a straight line. In a typical camera with the film radius $r_F = 28.649\,\mathrm{mm}$, $\theta = 2°\,\mathrm{mm}^{-1}$ and the straight line on the film is inclined by 63.43°. A line passing through the origin at a distance d^* gives a line on the film defined by Equation (79):

$$\theta = w_0 \pm \arccos\left(\cos w + \frac{d^*}{\lambda}\right. \tag{79}$$

A chart is drawn with parallel lattice lines appearing on a Weissenberg photograph. The use of the chart is shown

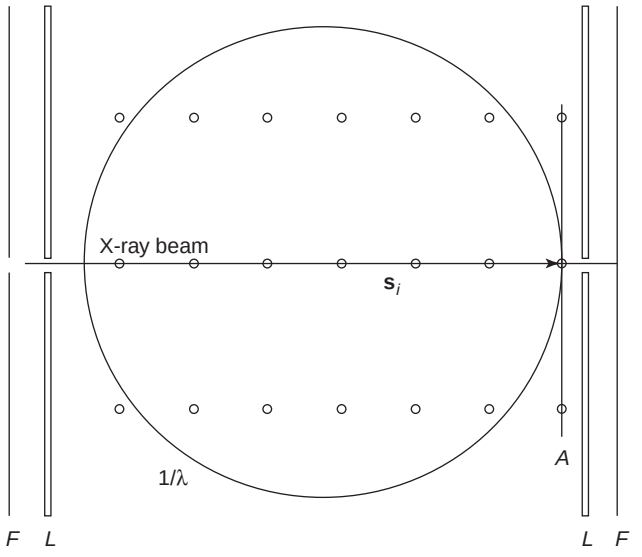

Figure 27 Schematic drawing of a zero layer Weissenberg photograph. From a rotation photograph (Figure 25) only the reflections of the zero layer can pass the cylindrical layer screen (L). The zero layer is normal to the drawing plane. The vectors $\mathbf{s}_f$ of the reflected beams have components perpendicular to the drawing plane. When a reciprocal lattice point sweeps through the sphere of reflection, a reflection is recorded on the cylindrical film (F). The reciprocal lattice is turned around the rotation axis (A). With the rotation of the crystal, the film is shifted parallel to A.

in Figure 28. Each reflection lies at the intersection of two lines relating two indices, say the indices h and k in an $hk0$ reciprocal lattice plane.

5.2.2.6 The Equi-inclination Method With the Weissenberg camera, the nonzero-layer lines are also investigated. The most comfortable way is the equi-inclination method. From the rotation photograph the angle ν_n is obtained, the angle between the zero layer line on the film and the nth layer line. For the equi-inclination method, the Weissenberg camera is inclined by μ_e, given by Equation (80):

$$\mu_e = \arcsin\left(\tfrac{1}{2}\sin\nu_n\right) \tag{80}$$

and the layer-line screen is shifted according to Equations (81) and (82):

$$s_L = r_L \tan\mu_e \tag{81}$$

$$= r_L \frac{\sin\nu_n}{\sqrt{4 - \sin^2\nu_n}} \tag{82}$$

where r_L is the radius of the screen. Thus, the angle between the incident beam $\mathbf{s}_i$ and the zero layer is μ_e, i.e. the same as the angles between the reflected beams and the zero layer ν_e (equi-inclination; Figure 29).

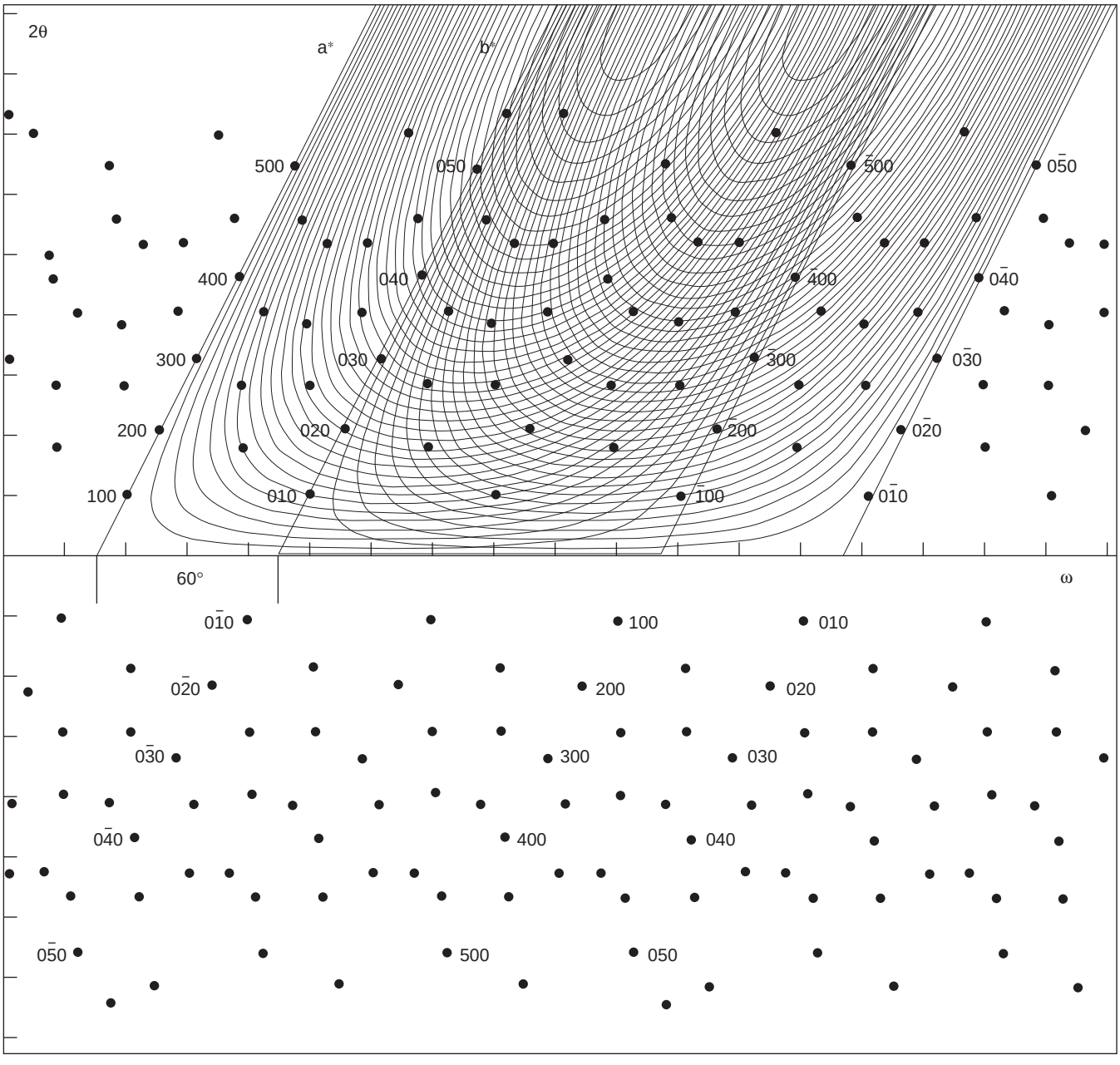

Figure 28 Schematic Weissenberg pattern of the $hk0$ plane of a hexagonal lattice. The equatorial line is the position of the reflection $h, k, l = 0, 0, 0$, which is removed by the beam stop of the incident beam. The reflections of the straight lines of the a^*- and b^*-axes are indexed as $h00$ and $0k0$. Each lattice point lies at the intersection of two lines of the charts. The line passing $h00$ defines the index h; the line passing $0k0$ defines the index k of a reflection $hk0$.

The advantage of this method is that the nth layer line is recorded on the film in almost the same way as the zero layer line. The same chart can be used to index the reflections. However, the scale is changed. In the zero layer line the angle 2θ was recorded as a function of the rotation angle w of the crystal; in the nth layer line the angle 2σ is recorded as a function of w. The relation between these angles is given by Equation (83):

$$\cos 2\theta_{hkl} = \cos 2\sigma \cos \mu_e \tag{83}$$

The relation is easily derived by spherical trigonometry. The equation of a line at a distance d^* from the intersection of the rotation axis with the nth layer in reciprocal space must be modified. These

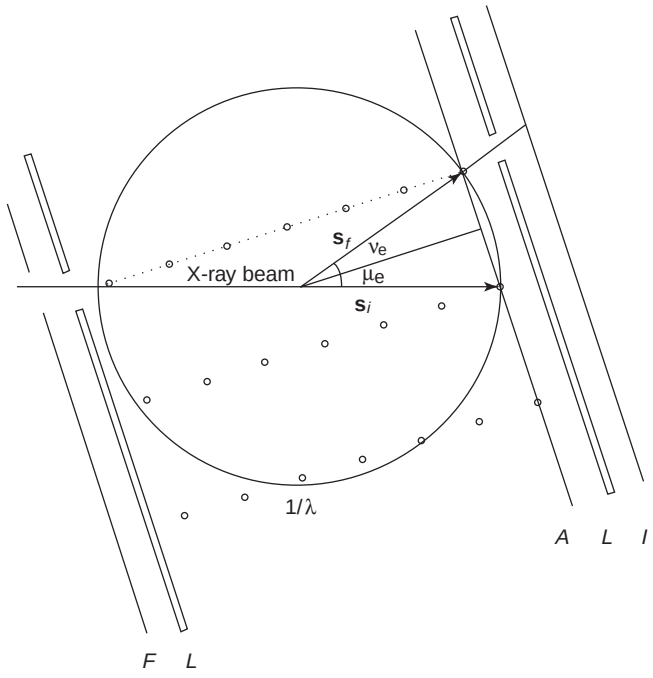

Figure 29 Schematic drawing of an equi-inclination Weissenberg photograph. The angle of the incident beam with the equatorial plane μ and the angle of the reflected beam with the equatorial plane ν_e are equal, where $\mu = \nu_e$. The sine of the angle ν_e is half of the sine of the angle ν_n, which was obtained from the rotation photograph. Thus only the reflections $\mathbf{s}_f$ of the upper lattice plane can pass the slit of the layer line screen (L), the gap between two cylinders. With the rotation of the crystal by the angle w, the cylindrical film F is shifted parallel to the rotation axis (A).

lines are mapped on the film and are given by Equation (84):

$$2\sigma = w_0 \pm \arccos\left(\cos w + \frac{d^*}{\lambda}\cos\mu_e\right) \qquad (84)$$

5.2.3 The Precession Method

5.2.3.1 The Zero Layer The precession method uses monochromatic radiation. A plane of the reciprocal lattice is investigated that is normal to the direction of the incident beam $\mathbf{s}_i/|\mathbf{s}_i|$. For example, the a-axis is adjusted parallel to the incident beam, i.e. the $0kl$ plane is perpendicular to it. In this position the $0kl$ plane is a tangential plane and touches the sphere of reflection at the origin O^*, i.e. at $h, k, l = 0, 0, 0$. In order to obtain reflections from the lattice plane $0kl$, the plane is inclined by the angle μ (Figure 30). In this position the plane intersects the sphere of reflection in a circle with the radius given by Equation (85):

$$r_c^* = \frac{\sin\mu}{\lambda} \qquad (85)$$

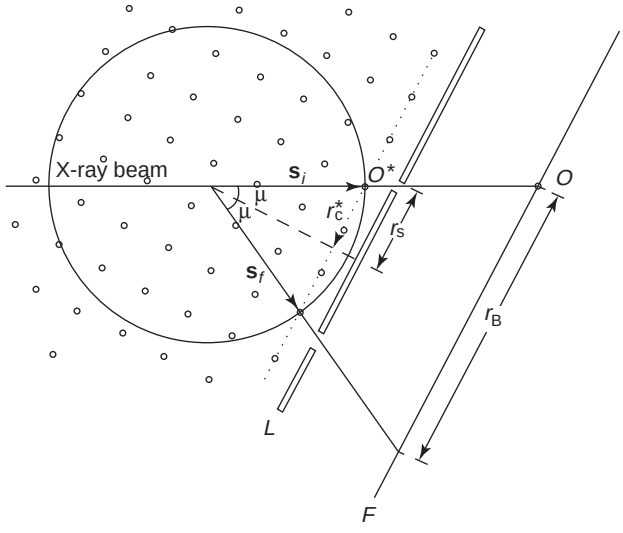

Figure 30 Schematic drawing for the reflection conditions of a zero layer line of a precession photograph. L is the layer-line screen; F is the film; O^* is the origin of reciprocal space, which is projected onto the film at O; μ is the angle of inclination of the camera; and r_c^* is the radius of the circle by which the reflecting net plane (dotted line) intersects the sphere of reflection.

In order to obtain a total map of the $0kl$ plane, all possible circles that can be obtained by an inclination μ have to be scanned through the sphere of reflection. This is obtained by a motion by which the normal of the lattice plane $0kl$, fixed at the origin O^*, moves on a cone with an opening angle of 2μ.

The cone of reflected beams emanates at the center of the sphere of reflection. From this point a circle of the reciprocal lattice plane is projected onto the film. In order to obtain an unperturbed map, the film must be kept parallel to the reflecting net plane $0kl$, i.e. the film has to perform the same motion with O as a fixed point as the lattice plane does with its origin O^* as a fixed point. In Figure 31 are shown projections of cones obtained by the intersection of the net plane with the sphere of reflection. The radii r_c of the circles on the film are as shown in Equation (86):

$$r_c = M\sin\mu \qquad (86)$$

where M is the magnification factor, i.e. the distance from the center O of the film to the center of the sphere of reflection, which is usually 60 mm. A reciprocal lattice point on the net plane at a distance r_{hkl}^* from the origin O^* is projected onto the film at a point that is at the distance s_{hkl} from O, the center of the film (Figure 31). The relation between r_{khl}^* and s_{hkl} is given by Equation (87):

$$2M\sin\theta_{hkl} = \lambda M r_{hkl}^* = s_{hkl} \qquad (87)$$

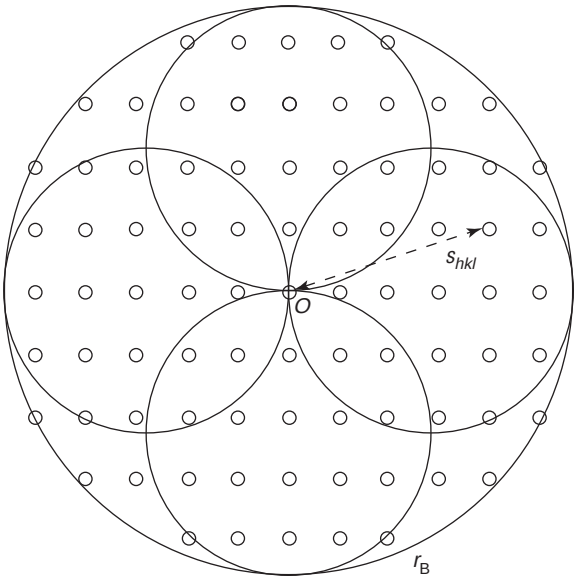

Figure 31 The picture of the layer-line screen at four moments of the motion, permitting a cone of reflected intensity to pass to the film. The incident beam could pass the screen as reflection $hkl = 0, 0, 0$ at O, which is prevented by the beam stop. By the motion of the camera the cone of reflected intensity moves and covers the film. The envelope of all reflected cones has the radius r_B; s_{hkl} is the distance of a reflection hkl from the origin $O = 0, 0, 0$.

In order to protect the film from reflections of other layers, a screen is inserted. The screen has a transparent circle with a given radius r_s, usually 15, 20, 25, or 30 mm. The screen is attached to the axis of the crystal at a distance s_L. When the inclination μ and the radius r_L are chosen, the screen has to be set at a distance given by Equation (88):

$$S_L = \frac{r_L}{\tan \mu} \tag{88}$$

The lattice plane is mapped onto the film within the radius given by Equation (89):

$$r_B = 2M \sin \mu \tag{89}$$

Outside r_B is a blind region where no reflections are found.

Before a precession photograph can be taken, the angle μ of the instrument must be set to zero and the normal of the lattice plane to be observed must be oriented parallel to the direction of the incident beam. There are three angles to adjust for the crystal: the two arcs of the goniometer head and the angle φ of the axis on which the goniometer head is mounted (Figure 32). A first overview can be obtained by a Laue picture without a screen. The parameters of the correction of small deviations are found in precession photographs

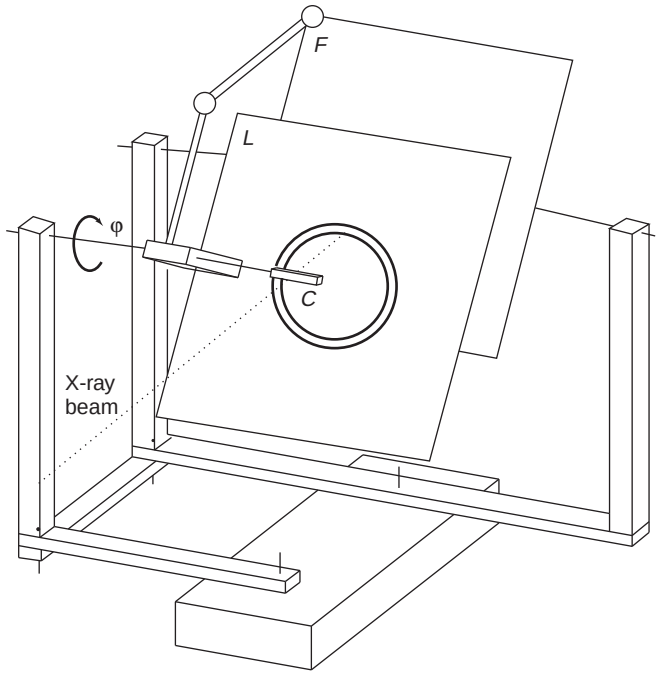

Figure 32 Schematic drawing of the front view of a precession camera. The crystal C is mounted in the ordinary precession suspension (i.e. the crystal is in the center of two intersecting rotation axes). The reflecting reciprocal lattice plane, the layer-line screen L, and the film F are kept always parallel. The screen L is fixed on the shaft of the axis on which the crystal is mounted. The motion of the film and the motion of the crystal are coupled by a rod. By the angle φ the reciprocal lattice plane of the crystal is adjusted parallel to the film.

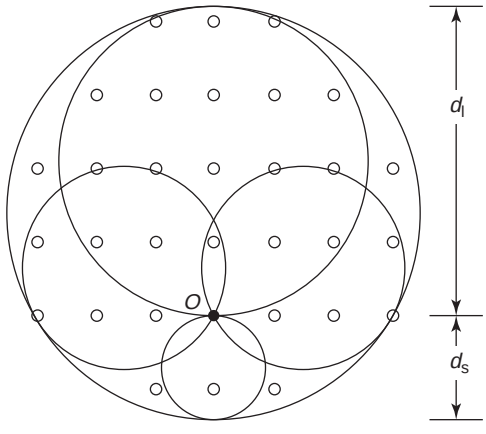

Figure 33 If the crystal is not adjusted, the intersections of the sphere of reflection with the different orientations of the reciprocal plane give circles with different radii r_c^* (cf. Figure 30). With this information the crystal can be adjusted (see text).

with a small precession angle, specifically $\mu = 10°$, and a layer-line screen with a circular hole of 10 mm. The effect is shown in Figure 33. In this picture the crystal has to be adjusted by a turn around the horizontal axis by the

angle φ (Figure 32). In Figure 33 four circles are drawn that touch the center of the photograph O, the point of the reflection $h, k, l = 0,\ 0,\ 0$; these circles have different radii and correspond to different positions in which the zero net plane intersects the sphere of reflection. The blind region – the envelope of all circles – is clearly seen on a photograph. This circle cuts off the tails of the white radiation of the reflections. In Figure 34 a vertical deviation is shown. The diameter of the large circle above O is $2r_l$; the diameter of the small circle below O is $2r_s$. Taking into account small angles, the opening angles of the lattice plane are given by Equation (86), i.e. $r_l = M(\mu + \Delta\varphi)$ and $r_s = M(\mu - \Delta\varphi)$. Thus the correction angle is given by Equation (90):

$$\Delta\varphi = \frac{d_l - d_s}{M} \tag{90}$$

A horizontal deviation is corrected correspondingly.

5.2.3.2 The Upper-level Layer The distance of the nth layer from the origin is $d_n^* = 1/nt$, where t is the translation of the crystal. The distance of the zero layer from the center of the sphere is $(\cos\mu)/\lambda$ (Figure 34);

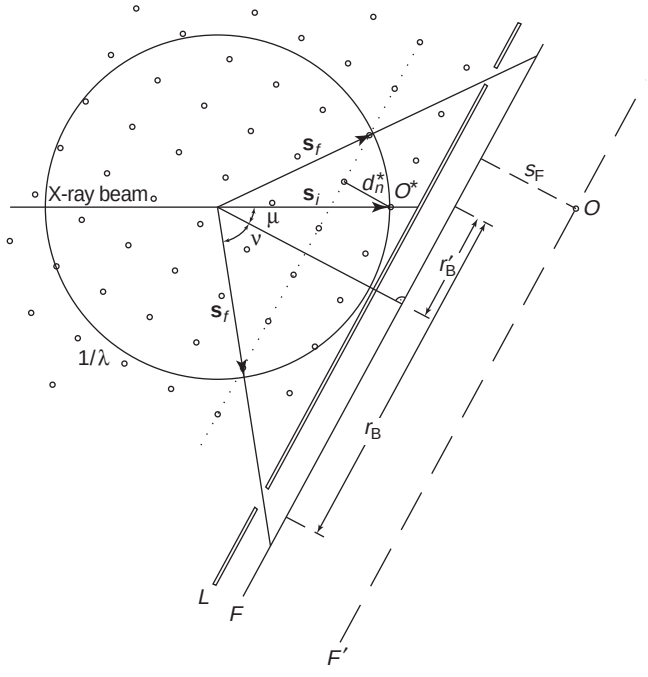

Figure 34 Schematic drawing of the reflection condition of an upper-layer precession photograph: the reflecting plane of reciprocal space is indicated by a dotted line; L is the layer-line screen; F is the correct position of the film; F' is the position of the film in the case of a zero layer photograph; O^* is the origin of reciprocal space; O is the center of the film; μ is the inclination of the camera; and ν is the aperture of the cone that can pass the screen. The distance of the reflecting plane from O^* is d_n^*, which is proportional to the distance of the film F from O.

the distance of the nth layer is thus $(\cos\mu)/\lambda - d_n^*$, as in Equations (91) and (92):

$$\frac{\cos\nu}{\lambda} = \frac{\cos\mu}{\lambda} - d_n^* \tag{91}$$

$$\sin\nu = \sqrt{1 - (\cos\mu - \lambda d_n^*)^2} \tag{92}$$

The radius r_{cn}^* of the circle obtained by the intersection of the nth layer with the sphere of reflection is given by Equation (93):

$$r_{cn}^* = \frac{\sin\nu}{\lambda} \tag{93}$$

Thus, the screen with radius r has to be placed at a distance s_s from the crystal, as shown by Equations (94) and (95):

$$s_s = \frac{r_s}{\tan\nu} \tag{94}$$

$$= r_s \frac{\cos\mu - \lambda d_n^*}{\sqrt{1 - (\cos\mu - \lambda d_n^*)^2}} \tag{95}$$

The nth lattice plane is at the distance d_n^* from O^*. In order to obtain an unperturbed projection of the lattice plane, the film must also be shifted from O toward the crystal as given by Equation (96):

$$s_F = d_n^* M\lambda \tag{96}$$

The motion of the film relative to the origin O must be similar, in the mathematical sense, to the motion of the nth lattice plane relative to the center of reciprocal space O^*. The distance s_{hkl} of a reflection hkl from the origin of the film is related to $r_{hkl}^* = 2(\sin\theta_{hkl})/\lambda$ as shown by Equations (97) and (98):

$$r_{hkl}^* = \sqrt{(d_n^*)^2 + \left(\frac{s_{hkl}}{\lambda M}\right)^2} \tag{97}$$

$$= \frac{1}{\lambda M}\sqrt{s_F^2 + s_{hkl}^2} \tag{98}$$

The lattice plane is mapped on the film within the radius as shown by Equation (99):

$$r_B = M(\sin\nu + \sin\mu) \tag{99}$$

which is the radius of the blind region (Figure 35). However, there is another blind region in the center of the precession photograph with the radius as shown by Equation (100):

$$r'_B = M(\sin\nu - \sin\mu) \tag{100}$$

The mechanics of a precession camera are shown in Figures 32 and 36. The crystal (C) is positioned in the center of two perpendicular rotation axes to enable a

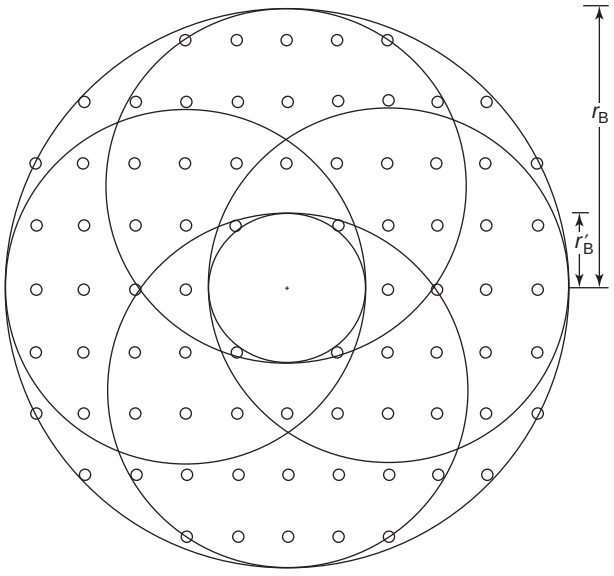

Figure 35 The picture of the layer-line screen at four moments of the motion of an upper-level photograph. The four circles are produced by the cones of reflected intensity that can pass to the film. In the center of the picture a blind region occurs: the inner envelope of the motion of the circles.

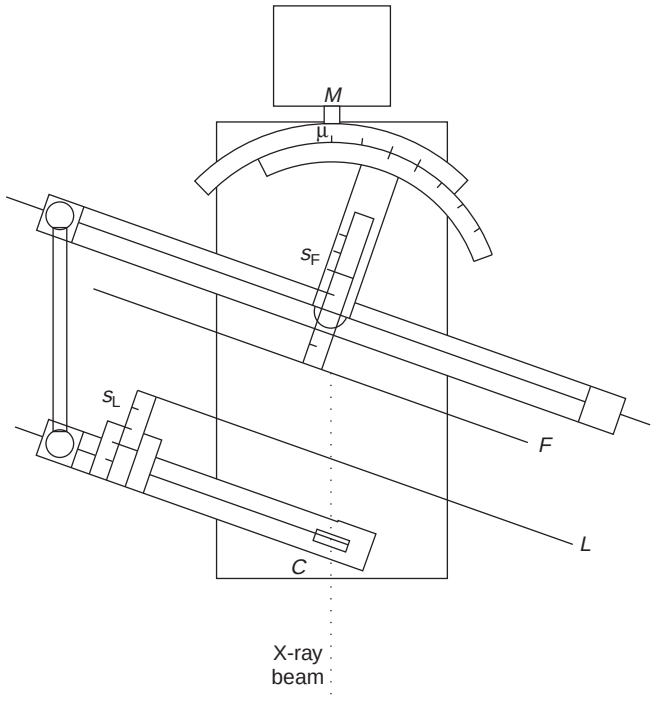

Figure 36 Schematic drawing of the top view of a precession camera. The motor (M) drives the precession motion, i.e. a cone with the opening angle μ (in the drawing 20°). The film (F), the layer-line screen (L), and the lattice plane of the crystal (C) are always parallel, s_F is the scale for the distance of the screen from the crystal; s_L is the scale for the shift of the film from the zero point, and the path of the incident beam is indicated by a dotted line.

precession motion. Two other rotation axes intersect in the zero point of the film (F). By using a rod, the motion of the crystal is coupled to the motion of the film. With $\mu = 0$, the reciprocal lattice of the crystal and the plane of the film are normal to the incident X-ray beam, the direction of which coincides with the motor axis. With $\mu = 0$, the crystal and film do not move when the motor rotates. With $\mu \neq 0$, the crystal and film perform a precession motion, i.e. the normals of the crystal in O^* and the film in O yield a cone with an opening angle 2μ. On varying φ, the holder for the layer screen (L) is fixed. Its distance from the crystal can be set by using the scale s_L. For photographs of higher levels, the film must be shifted from the zero point (O) toward the crystal. The distance can be read out on the scale s_F. (The precession camera was invented and described by Buerger.[38])

5.3 Powder Diffraction

The diffraction of a single crystal was described by the reciprocal lattice, which consists of mathematical points. However, all reflections have a profile, even those of a perfect crystal. In a mosaic crystal the domains are slightly misoriented, which can be measured by the broadening of the reflections: the mosaic spread. There are various possible transitions between a perfect crystal and a randomly oriented sample such as a powder. A statistically preferred orientation is called a *texture*, which can be measured by using a four-circle diffractometer. The detector is set at a value of 2θ for a Bragg reflection and the

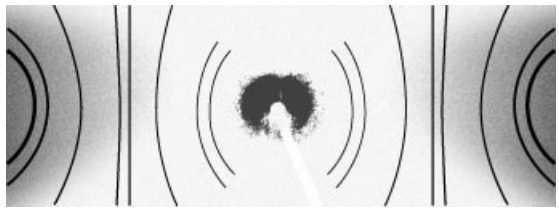

Figure 37 Powder diffraction pattern of a copper wire showing texture.

distribution of the reflected intensity in reciprocal space is measured. There are textures of rotational symmetry in fibers or wires, and textures on the surface of rolled material. Even powders may have a texture. For example, clay minerals prepared on a flat surface lie parallel. This is used to identify the different modifications according to their c lattice constant normal to the plane of preparation. It is sometimes very difficult to measure the correct intensities of a powder pattern owing to texture. Texture is immediately observed in a Debye–Scherrer photograph (Figure 37). With a powder diffractometer one is often not aware of the effect. The effect of texture can be reduced by spinning the sample.

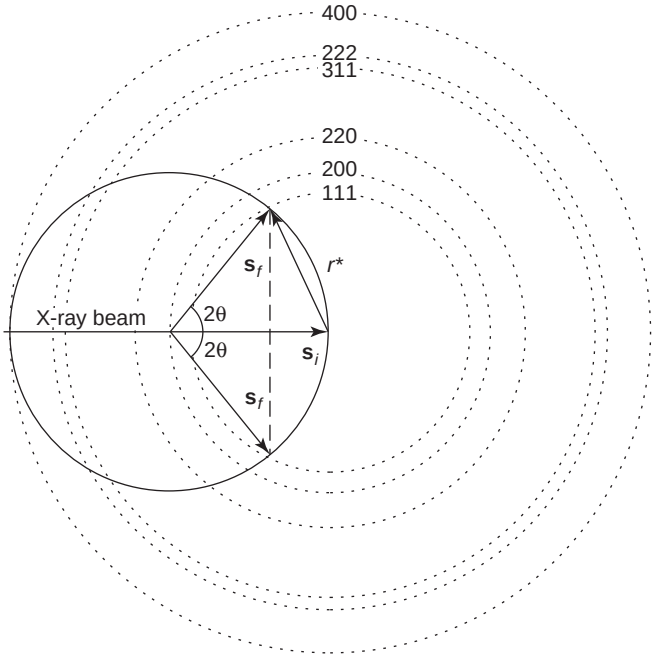

Figure 38 Reciprocal space of a powder with an fcc structure. The reciprocal lattice points of the powder are randomly oriented on spheres as indicated by dotted lines. The intersection of the sphere $hkl = 111$ with the sphere of reflection results in a cone, emanating from the center of the circle of reflection, with an opening angle of $4\theta_{111}$. This angle is observed in a Debye–Scherrer camera.

Powder diffraction is mainly used to identify crystalline phases. Structure determination by X-ray powder diffraction has become an important field. There is usually enough material for a powder sample and a powder diagram is easily obtained experimentally.

In Figure 38 the Ewald construction for powder diffraction is shown. The isotropic distribution of the crystals in the sample produces spheres of reflecting net planes with the radius $r^* = 1/|d_{hkl}|$ in reciprocal space, i.e. the lattice points hkl are smeared out to spheres. The sphere of one powder line hkl with the Bragg angle $2\theta_{hkl}$ intersects the sphere of reflection in a circle. From the center of the sphere of reflection opens a cone of diffraction with an opening angle of $4\theta_{hkl}$, which is recorded on a film or measured by using a diffractometer. For further reading, several textbooks are available.[39–42]

5.3.1 The Debye–Scherrer Method

A schematic drawing of the Debye–Scherrer giometry is given in Figure 39. The advantage of the method is that all powder lines can be recorded simultaneously. The resolution $\Delta\theta/\theta$ is determined by the divergence of the primary beam and the radius r between the sample and the detector. However, if one improves the resolution by

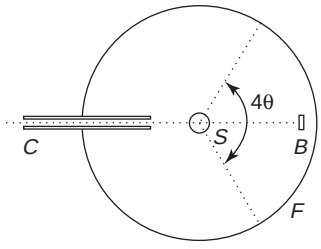

Figure 39 Schematic drawing of a Debye–Scherrer camera or diffractometer: C, entrance collimator; S, powder sample; B, beam stop; F, film or position-sensitive detector of the circle on which a detector scans.

enlarging this radius r, the observed intensity is reduced, being proportional to r^{-2}. Well-ground powder is filled into a capillary, which is centered in the middle of the device. In order to improve the statistics (i.e. the homogeneity of the diffraction pattern), the sample is spun. In the case of a photographic camera, a cylinder containing the film is placed around the sample. The radius r of the camera is chosen to be $n \times r = 28.648$ mm, so that 1 mm on the film is equal to $2^\circ/n$. Today the film is replaced by a position-sensitive detector. In the case of a Debye–Scherrer diffractometer, a counter or a counter bank sweeps at a constant distance around the sample.

5.3.2 The Bragg–Brentano Diffractometer

The resolution of the Debye–Scherrer method was improved by the invention of the Bragg–Brentano diffractometer. This is a parafocusing method. The resolution is improved in the region of the focus, but elsewhere the resolution is reduced. Thus, only one powder line and its environment can be detected simultaneously. The diffractometer is shown schematically in Figure 40. The flat sample is rotated by an angle θ while the detector is rotated by an angle 2θ around the axis common to both circles. The distance from the X-ray source to the sample is equal to the distance from the sample to the entrance slit of the detector. The focusing circle is defined by the following three points: the source, the center of the sample, and the entrance slit of the detector. The flat sample is positioned tangential to the focusing circle. The focusing condition is fulfilled only for the points on the focusing circle. Thus, the sample obeys only a parafocusing condition. The resolution is mainly given by the divergence of the X-ray beam and the source–sample distance or the sample–detector distance. The divergence normal to the diffraction plane is improved by Soller slits: a set of parallel foils of strongly absorbing material at an equal distance from one another. Owing to the lines of the Kα spectrum, the reflections are broadened in the range of small 2θ and split at high 2θ. One can suppress the Kα_2 radiation by placing a

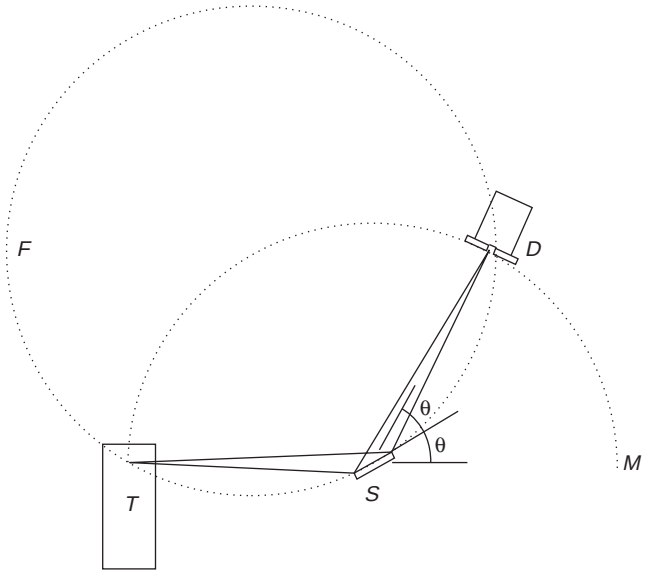

Figure 40 Schematic drawing of the Bragg–Brentano diffractometer: *T*, X-ray tube; *S*, sample; *D*, detector; *M*, circle on which the entrance slit of the detector scans with angle 2θ while the sample is turned by the angle θ; *F*, focal circle.

monochromator after the source. The fluorescent radiation produced in the sample by the primary beam can be removed by inserting a monochromator in front of the detector.

ABBREVIATIONS AND ACRONYMS

bcc	Body-centered Cubic
EXAFS	Extended X-ray Absorption Fine Structure
fcc	Face-centered Cubic
fwhm	Full Width at Half-maximum

RELATED ARTICLES

Forensic Science **(Volume 5)**
Scanning Electron Microscopy in Forensic Science • X-ray Fluorescence in Forensic Science

Particle Size Analysis **(Volume 6)**
Diffraction in Particle Size Analysis

Peptides and Proteins **(Volume 7)**
X-ray Crystallography of Biological Macromolecules

Steel and Related Materials **(Volume 10)**
Metal Analysis, Sampling and Sample Preparation in • X-ray Fluorescence Spectrometry in the Iron and Steel Industry

X-ray Spectrometry **(Volume 15)**
Energy Dispersive, X-ray Fluorescence Analysis • Sample Preparation for X-ray Fluorescence Analysis • Total Reflection X-ray Fluorescence • Wavelength-dispersive X-ray Fluorescence Analysis

REFERENCES

1. I. Bernal, W.C. Hamilton, J.S. Ricci, *Symmetry: A Stereoscopic Guide for Chemists*, Freeman, San Francisco, 1972.
2. M.B. Boison, G.V. Gibbs, 'Mathematical Crystallography', *Rev. Mineral.*, **15** (1990).
3. M.J. Buerger, *Elementary Crystallography*, Wiley, New York, 1956.
4. G. Burns, A.M. Glazer, *Space Groups for Solid State Scientists*, Academic Press, New York, 1978.
5. F.A. Cotton, *Chemical Application of Group Theory*, 3rd edition, Wiley, New York, 1990.
6. H.H. Jaffé, M. Orchin, *Symmetry in Chemistry*, Wiley, New York, 1965.
7. E.H. Lockwood, R.H. Macmillan, *Geometric Symmetry*, Cambridge University Press, Cambridge, 1978.
8. G.E. Martin, *Transformation Geometry: An Introduction to Symmetry*, Springer, New York, 1982.
9. H. Megaw, *Crystal Structures: A Working Approach*, Saunders, Philadelphia, 1973.
10. A.V. Shubnikov, V.A. Koptsik, *Symmetry in Science and Art*, Plenum Press, New York, 1974.
11. J.V. Smith, *Geometrical and Structural Crystallography*, Wiley, New York, 1982.
12. B.K. Vainstein, *Modern Crystallography. Symmetry of Crystals*, Springer, Berlin, Vol. 1, 1981.
13. H. Weyl, *Symmetry*, Princeton University Press, Princeton, 1952.
14. T. Hahn (ed.), *International Tables for Crystallography*, 2nd revised edition, D. Reidel, Dordrecht, Vol. A, 1989.
15. A.H. Kirfel, K.-D. Eichorn, private communication, 1990.
16. A.H. Templeton, L.K. Templeton, *Acta Crystallogr.*, **A38**, 62–67 (1982).
17. L.V. Azároff, *Elements of X-ray Crystallography*, McGraw-Hill, New York, 1968.
18. C.S. Barret, T.B. Massalski, *Structure of Metals – Crystallographic Methods, Principles and Data*, McGraw-Hill, New York, 1966.
19. M.J. Buerger, *X-ray Crystallography*, Wiley, New York, 1962.
20. M.J. Buerger, *The Precession Method in X-ray Crystallography*, Wiley, New York, 1964.
21. M.J. Buerger, *Contemporary Crystallography*, McGraw-Hill, New York, 1970.
22. G.B. Carpenter, *Principles of Crystal Structure Determination*, Benjamin, New York, 1969.
23. J.W. Jeffrey, *Methods in X-ray Crystallography*, Academic Press, London, 1971.

24. M.F.C. Ladd, R.A. Palmer, *Structure Determination by X-ray Crystallography*, Plenum Press, New York, 1977.
25. P. Luger, *Modern X-ray Analysis on Single Crystals*, De Gruyter, Berlin, 1980.
26. D. McKie, C. McKie, *Essentials of Crystallography*, Blackwell, Oxford, 1986.
27. E.W. Nuffield, *X-ray Diffraction Methods*, Wiley, New York, 1974.
28. L.H. Schwartz, J.B. Cohen, *Diffraction from Materials*, 2nd edition, Springer-Verlag, Berlin, 1987.
29. G.H. Stout, L.H. Jensen, *X-ray Structure Determination – A Practical Guide*, Wiley, New York, 1989.
30. M.M. Woolfson, *An Introduction to X-ray Crystallography*, Cambridge University Press, Cambridge, 1970.
31. C.K. Johnson, H.A. Levy, *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham, 311–336, Vol. 4, 1974.
32. R. Hosemann, S.N. Bagchi, *Direct Analysis of Diffraction by Matter*, North-Holland, Amsterdam, 1962.
33. H.P. Klug, L.E. Alexander, *X-ray Diffraction Procedures for Polycrystalline and Amorphous Materials*, 2nd edition, Wiley, New York, 1974.
34. P. Coppens, L. Leiserowitz, D. Rabinovich, *Acta Crystallogr.*, **18**, 1035–1038 (1965).
35. B.E. Warren, *X-ray Diffraction*, Addison-Wesley, Reading, MA, 1969.
36. G. Caglioti, A. Paoletti, F.P. Ricci, *Nucl. Instrum.*, **3**, 223 (1958).
37. H.M. Rietveld, *J. Appl. Crystallogr.*, **2**, 65–71 (1969).
38. M.J. Buerger, *The Precession Method*, Wiley, New York, 1964.
39. H. Lipson, H. Steeple, *Interpretation of X-ray Powder Diffraction Patterns*, Macmillan, London, 1970.
40. D.L. Bish, J.E. Post (eds), 'Modern Powder Diffraction', *Rev. Miner.*, **20** (1989).
41. A.J.C. Wilson, *Mathematical Theory of X-ray Powder Diffractometry*, Philips Technical Library, Eindhoven, 1963.
42. A.J.C. Wilson, *Elements of X-ray Crystallography*, Addison-Wesley, Reading, MA, 1970.

Total Reflection X-ray Fluorescence

Peter Wobrauschek and Christina Streli
Atominstitut der Österreichischen Universitäten, Vienna, Austria

1 Introduction **1**
1.1 Physical Principles 1
1.2 X-ray Fluorescence Analysis 2
1.3 Features of X-ray Fluorescence Analysis 2
2 History **3**
2.1 First Considerations and Applications of Total Reflection of X-rays 4
2.2 External Total Reflection 5
3 Total Reflection X-ray Fluorescence Analysis **6**
3.1 Historical Development 6
3.2 Total Reflection X-ray Fluorescence Analysis Experimental Set-up 6
4 Sources for Total Reflection X-ray Fluorescence Analysis and Spectral Modifying Devices **7**
5 Reflector Specifications: Surface Quality, Material, Shapes **8**
6 Technical Instrumentation for Total Reflection X-ray Fluorescence Analysis **8**
7 Detectors and Electronics **10**
8 Quantification in Total Reflection X-ray Fluorescence Analysis **12**
9 Sample Preparation **14**
10 Applications **16**
10.1 Environmental Samples 16
10.2 Medical Applications 17
10.3 Industrial Applications 17
10.4 Forensic and Art Historical Applications 19
11 Characterization of Near Surface Layers **19**
12 Synchrotron Radiation Excited Total Reflection X-ray Fluorescence Analysis **21**
13 Total Reflection X-ray Fluorescence Analysis of Low-Z Elements **24**
14 Conclusions and Outlook **26**
15 Manufacturers of Total Reflection X-ray Fluorescence Analysis Spectrometers **26**
Acknowledgments **26**
Abbreviations and Acronyms **26**
Related Articles **27**
References **27**

Total reflection X-ray fluorescence (TXRF) analysis is an energy-dispersive (ED) X-ray fluorescence (XRF) technique using a special excitation geometry. The narrow collimated beam impinges at grazing incidence on a special sample carrier. This is a flat polished quartz plate on which the sample is deposited. The most significant advantages of TXRF over conventional XRF analysis are the improved detection limits achieved by inherent signal increase and reduced background. Detection limits are in the picogram range for X-ray tube excitation and even down to femtograms for excitation using synchrotron radiation (SR). The main application is the analysis of some microliters of a liquid sample deposited and dried on the quartz reflector or of some microliters of slurry or microgram grains of solid samples. The most widely used application is the quality control of contaminations on Si wafer surfaces. The detection limits demanded by the semiconductor industry can be reached today using X-ray tube excitation, but metrology shows that in the future SR is required or special preconcentration techniques. Straight TXRF is the only analytical technique offering nondestructive and spatial resolved analysis. A special ED detector with an ultrathin window is suitable for the detection of low atomic number, Z elements (C upwards) using TXRF. Another promising application is the characterization of thin layers on a reflecting surface or implanted atoms in Si by variation of the incidence angle and measurement of the fluorescence signal. It is possible to determine the elements forming the layer, the layer thickness and its density. For thin layered samples as well as implantations, thickness, depth and depth profile for implanted atoms can be evaluated by comparing experimental results with theoretical models.

1 INTRODUCTION

1.1 Physical Principles

XRF analysis is a powerful analytical tool for the spectrochemical determination of almost all the elements

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

present in a sample. XRF radiation is induced if photons with sufficiently high energies emitted from an X-ray source impinge on the sample. These primary X-rays undergo interaction processes with the atoms of the analyte. Taking Bohr's model of the atom as a basis to explain the mechanism of the excitation and creation of fluorescence radiation, this process can be described as follows. High-energy photons induce ionization of inner shell electrons by the photoelectric effect and thus electron holes in the innermost K shell or the L or M shells are created. After the interaction the photoelectron leaves the atom carrying the difference energy of incident photon energy minus the binding energy of the specific shell electron which was expelled. The prompt transition of outer shell electrons into the vacancies within some 100 fs causes emission of characteristic fluorescence radiation to take place. Not all transitions from outer shells or subshells are allowed, only those obeying the selection rules for electric dipole radiation. All the excited atoms of each element in the sample emit their characteristic X-rays. An electron transition following the electric dipole selection rules, from e.g. L_{III} to K, fills up the K shell but leaves the L shell with an electron hole. This L shell vacancy is again filled by another transition (e.g. $M_V - L_{III}$) and so on until the last outer shell vacancy is filled from the pool of free electrons always present. The result is a cascade of electron transitions all correlated with the emission of photons with a well-defined energy corresponding to the difference in energy between the atomic shells involved. The X-ray family of characteristic X-rays from each element including all transitions allows the identification of the element. In fact XRF is not the only possible process following the electron transitions from higher energy levels to the lower energy states. There is a competing process associated, the emission of Auger electrons. Both processes have certain Z-dependent probabilities which are complementary: Auger yield is high for light elements and fluorescence yield is high for heavy elements. The determination of Auger electron energy and intensity also allows spectroscopy to identify the elements but is not considered here.

1.2 X-ray Fluorescence Analysis

The working principle of XRF analysis is the measurement of wavelength or energy and intensity of the characteristic photons emitted from the sample. This allows the determination of the elements present in the analyte and their mass or concentration. All the information for the analysis is stored in the measured spectrum, which is a line spectrum with all the characteristic lines superimposed above a certain fluctuating background. The background is induced by other interaction processes, mainly the elastic and inelastic scattering of the primary radiation on sample and substrate. Detailed information about XRF analysis is given in the literature.[1–3]

Measurement of the spectrum of the emitted characteristic fluorescence radiation is performed using wavelength-dispersive (WD) and ED spectrometers. In wavelength-dispersive X-ray fluorescence (WDXRF) analysis the result is a spectrum intensity of the characteristic lines versus wavelength measured with a Bragg single crystal as dispersion medium and counting the photons with a Geiger-Mueller, proportional or scintillation counter. In energy-dispersive X-ray fluorescence (EDXRF) analysis the photon energy is directly measured by a suitable detector which selects the photons due to energy and counts them directly and stores the result in a multichannel memory. The result is a spectrum with intensity on the y-axis and energy on the x-axis. The range of detectable elements is limited. It ranges from Be ($Z = 4$) for the light elements and goes up to U ($Z = 92$) on the high atomic number Z side. The concentrations that can be determined with standard spectrometers of both types WD or ED lie between micrograms per gram and weight percent, thus having a wide dynamic range. In terms of mass the nanogram range is reached with spectrometers that have excitation geometry typical of incident primary radiation (45° to the flat sample surface).

In the course of the developments and improvements which are discussed in this chapter, by introducing special excitation geometry, optimized sources and detectors, picogram and even femtogram ranges for the detection limits were reached and in terms of concentration the same improvement factor was attained from microgram per gram to picogram per gram under the best conditions. In principle, XRF analysis is a multielement analytical technique and in particular, the simultaneous determination of all the detectable elements present in the sample is inherently possible with EDXRF. In WDXRF both the sequential and the simultaneous detection modes are possible.

1.3 Features of X-ray Fluorescence Analysis

What are the defining features of XRF analysis? It can be said that this analytical technique allows the qualitative and quantitative analysis of almost all the elements (Be–U) in an unknown sample. The analysis is in principle nondestructive, has high precision and accuracy, has simultaneous multielement capacity, requires only a short time, a high sample throughput is possible, on-line analysis is possible, running costs are low, is extremely versatile for applications in many fields of science, research and quality control, has low detection limits, and a large dynamic range of concentrations covering up to 10^9 orders of magnitude. The physical size of an XRF spectrometer ranges from handheld, battery-operated field units to

high-power laboratory units with compact table top units and larger ones requiring several cubic meters of space including a 10–20 kW electrical power supply and efficient cooling units with high pressure water and a heat sink. In contrast to all these attractive properties there are some disadvantages. The absorption effects of primary radiation and the fluorescence radiation created in the analyte result in a narrow layer a few tenths of a millimeter wide which provides information from the source of the signal. This requires a perfectly homogeneous sample which often occurs naturally but must sometimes be produced by dissolution into liquids or by grinding and pressing pellets. In both examples the feature of nondestructiveness is lost. Most ideally thin films or small amounts of microcrystalline structure on any substrate are the ideal analyte where even the quantification process is simple because there is linearity between fluorescence intensity and concentration. In thick samples there is only a proportionality between intensity and concentration or mass of the element which requires calculations to correct for absorption and enhancement effects.

The interactions of primary X-rays create the fluorescence signal from the analyte. Considering the interactions only, the photoelectric effect is responsible for the creation of characteristic lines. However, this is not the only effect observed in this energy range of X-rays mainly used for XRF, namely to 60 keV or 100 keV. Elastic and inelastic scattering of the primary radiation are observed simultaneously on the sample itself and on the substrate. This results in a background in the measured spectrum upon which the characteristic lines are superimposed. The primary radiation displays a spectral distribution according to the way it was produced. An X-ray tube is most often in use where the direct beam shows a continuum (bremsstrahlung) and characteristic lines of the anode material. Figure 1 shows a spectrum emitted from an X-ray tube with a Mo anode and operated 45 kV and 3 mA. Owing to the scattering processes those photons scattered towards the detector will be registered in the same way as the fluorescence radiation and appear in the measured spectrum as background. This background is the limiting factor as the signal from lower and lower concentrations or masses will decrease proportionally, whereas the scatter background from the sample matrix or the substrate will remain the same. When a particular lowest mass or concentration value is reached the signal cannot be distinguished from the statistical fluctuations of the background, in other words the detection limits are reached. Ways of improving the detection limits are discussed in section 3.1.

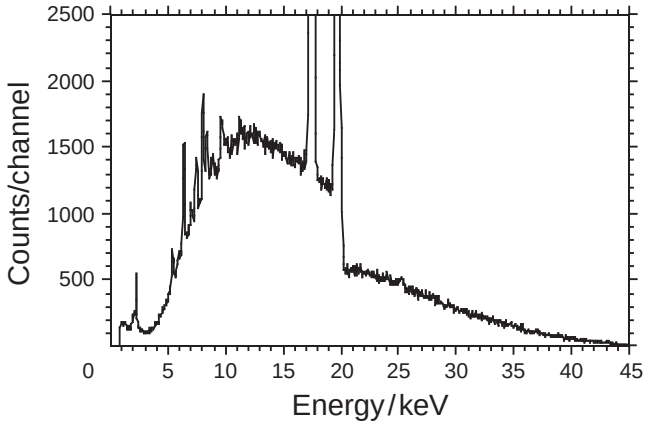

Figure 1 Primary spectrum from a fine focus Mo tube at 45 kV, 5 mA, 1000 s and 6° take-off angle.

A renaissance in XRF took place as the importance of trace element analysis in samples from environment, medicine and technology was clearly shown and because XRF was such a perfect candidate for a spectrochemical analytical technique.

The burden on the environment from natural or artificial sources requires the use of analytical techniques so that careful control of the emitters and what part is absorbed or accumulated in all forms of life can be maintained. Knowledge about the importance of trace amounts in the uptake from food, air, and water by humans requires reliable critical measurements at extremely low levels of concentration. Changes in the production and preparation of food has changed the uptake of essential trace elements by human beings, thus changing the levels of both biopositive and negative trace elements in body fluids, tissues, bones and organs. Medical diagnosis is supported by knowledge of the expected average levels of trace elements in the body and changes in the low microgram per gram levels can point to serious diseases. Medical treatment like chemotherapy can lead to overexposure of carrier elements in some organs which should be avoided to keep them working properly. The wide field of nanotechnology and microelectronics requires new levels of detectability in ultratrace amounts where femtogram levels must be considered because of the relevant damage they cause in the final device in reducing the production yield. The capability of detecting lower and lower masses or concentrations by the introduction of improved analytical techniques and methods opens up new applications and can have a new impact on science and technology.

2 HISTORY

On November 8th, 1895 Conrad Wilhelm Roentgen, at age 50, discovered X-rays. Within a few weeks he had performed in his laboratory at the University of Würzburg, Germany, a great number of experiments

with this ''new kind of radiation'', which is the literal translation of his first publication in December 1895, *Über eine neue Art von Strahlung*.[4] He found that X-rays had many properties, like ionization, transmission and absorption, blackening of photographic plates, as well as inducing fluorescence in the visible range of certain materials. Some other properties of the new kind of radiation like polarizability, refraction and reflection were tested by W.C. Roentgen in experiments but were not observed and thus declared to be nonexistent. Later C.G. Barkla,[5] in a series of experiments which he performed in 1906–1907, found that he could polarize X-rays by scattering on a piece of carbon through an angle of 90°. This constituted an important step on the way to explaining the true nature of X-rays by showing evidence that this property which is well known for visible light and electromagnetic transversal waves could be applied to X-rays. The true nature of X-rays as electromagnetic waves was certified 17 years after their discovery by the Munich research group of M. Laue, W. Friedrich and P. Knipping[6] and later by the University of Cambridge scientist W.L. Bragg.[7]

In all the early publications no experimental observation of refraction or reflection of X-rays was made, meaning that an important property which was expected to exist, as it did for visible light, was missing. In 1919 W. Stenström[8] described in his thesis the observation of slight deviations from the calculated Bragg angles, in particular at diffracted peaks of higher orders, which he interpreted correctly as refraction phenomenon. Finally he concluded theoretically from this fact, that if reflection occurs then refraction must also be occurring and must be observable. The first experimental evidence of the effect of the ''total reflection of X-rays'' was confirmed and published by A.H. Compton[9] in 1922.

The reason for the relatively late experimental verification is because the critical angles for the total reflection of X-rays are small, in the range of milliradians or tenths of a degree, and in contrast to light optical experiments the effect is observed for X-rays while they are passing from air or vacuum to matter, that is to say in the light optical sense into denser material. Experimentally the set-up was such that a narrow collimated X-ray beam was directed towards the top edge of a glass prism, thus giving a minimum of absorption, and the refraction–reflection phenomenon was observed at the tip, resulting in three split X-ray beams, one primary beam which passes on in the same direction and one passing above (the reflected beam) and one passing in between (the refracted beam), all registered on a photographic plate. Figure 2 shows a sketch of the experimental set-up.

On the photographic plate the three split beams were observed well. This experiment was performed by A. Larsson, M. Siegbahn and J. Waller[10] in 1924. This

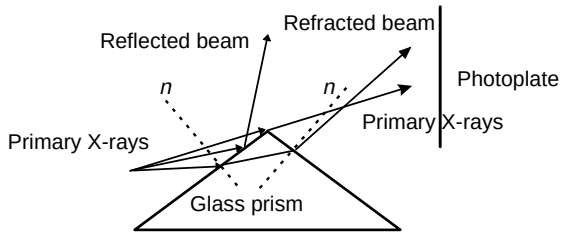

Figure 2 Sketch of the experimental set-up to show X-ray refraction and reflection; *n* denotes the normal to the prism planes.

experiment also shows the effect of refraction from the normal to the plane, which means that $n < 1$ (where $n =$ index of refraction) and that there is a phase velocity inside the medium larger than c, the speed of light in vacuum or air.

2.1 First Considerations and Applications of Total Reflection of X-rays

Some applications of the X-ray total reflection effect were suggested and performed immediately after the total reflection effect interpreted in terms of classical electrodynamics and Fresnel formalism had been used to calculate reflectivity and transmission intensities. Following the suggestions of E. Nähring,[11] an idea was realized for use of the total reflection effect as a possible means of modifying the spectral distribution of primary radiation. The instruments so created were falsely named ''total reflection monochromators'' but did indeed suppress the high-energy part of the spectrum, thus filtering the short wavelengths of the continuum emitted by the X-ray tube. M. Montel[12] (1956) reported and described this effect again. G. Damaschun et al.[13] (1967) used this phenomenon to improve the monochromaticity of the primary beam, performing small angle X-ray scattering measurements with a Kratky camera.

Optical components for X-ray astronomy were designed using the total reflection effect and spectacular X-ray optical components were made to focus the X-ray beam into micrometer spots. This is achieved by arranging bent mirrors in one, two or more planes to produce a point focus from a source that originally emitted in a large solid angle. Another interesting application of the total reflection effect was direct measurement of the energy-dependent fluorescence signal, in particular characteristic X-rays from light elements in the range 180 (B) to 1024 eV (Na). This was performed using the reflectance of a mirror and the energy dependence of the critical angle of total reflection as dispersive element, in the early 1960s by H.K. Herglotz[14] using a ''paraffin mirror'' for Be, B, C and a ''LiF mirror'' for N, O, F.

In 1954 L.G. Parrat[15] published an excellent theoretical paper about the external total reflection of X-rays.

2.2 External Total Reflection

This phenomenon is observed if a fine narrow collimated X-ray beam impinges on the plane-polished smooth surface of any material. Normally, solid materials are used as mirrors but total reflection from liquid surfaces was also reported. The surface quality of a reflector is described well by mean roughness, which should be in the range of a few nanometers or less and an optical flatness of $\lambda/20$, where λ is typically 589 nm in the visible light range.

Experimental observation of total reflection is best described by starting at large angles of several degrees of incidence, measured between the beam direction and the reflector surface. The X-ray beam having typically 5–20-keV energy (wavelength λ from 0.24–0.06 nm) shows the normal behavior of penetration and absorption in the reflector. Looking in more detail, the beam inside the medium after penetration through the interface shows a slight angular deviation from its original direction to a smaller angle measured between the new direction of the beam and the plane of the interface. This refraction of the beam in the medium is the basis of the following conclusions. In decreasing the incidence angle a certain so-called critical angle is achieved where the direction of the propagation of the refracted beam is parallel to the surface of the reflector. Further reduction of the incidence angle leads to an intensive reflected beam having almost the same intensity as the incoming one. This is the totally reflected beam, where total refers to the almost identical number of photons in the reflected beam as in the primary beam. Figure 3 shows three different cases.

From theoretical considerations of Fresnel's formalism it follows that even in the case of total reflection there is a certain amount of penetration of the primary X-rays into the reflector. This represents an energy transfer into the medium. The correlation between critical angle and the material constants as well as the energy of the impinging photons on the reflector is given by the physical parameters described by classical electrodynamics.

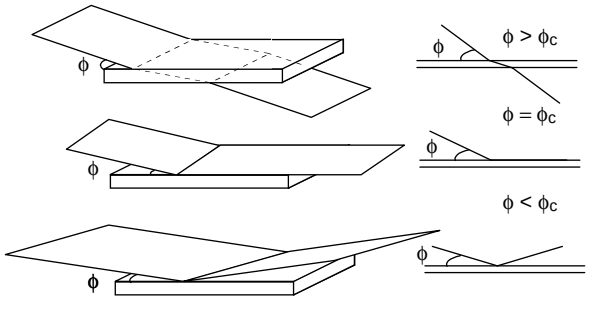

Figure 3 Principle of total reflection, for three cases: $\phi > \phi_c$, $\phi = \phi_c, \phi < \phi_c$.

Starting with the complex index of refraction, n which is given by its real and imaginary part, Equation (1)

$$n = 1 - \delta - i\beta \tag{1}$$

where the two terms β and δ are described by the results of the classical electrodynamics of a dipole in the electric field of an incoming wave. The δ is often named the decrement of n with small values in the 10^{-5} range; β corresponds to the absorption term of the X-rays.

These values are correlated to the scattering factors f_1 and f_2 from Kramers Kronig dispersion relations which are tabulated by Henke et al.,[16] Equations (2) and (3)

$$\delta = Kf_1(E) = 4.15 \times 10^{-4} \frac{\rho}{A} \frac{1}{E^2} f_1(E) \tag{2}$$

$$\beta = Kf_2(E) = 4.15 \times 10^{-4} \frac{\rho}{A} \frac{1}{E^2} f_2(E) = 9.86 \times 10^{-9} \frac{\mu(E)}{E} \tag{3}$$

where ρ is the material density ($g\,cm^{-3}$), A is the atomic weight, E is the photon energy (keV) and μ is the mass absorption coefficient ($cm^2\,g^{-1}$). K is given by Equation (4)

$$K = \frac{r_0 \lambda^2}{2\pi} \frac{N_A}{A} \rho \tag{4}$$

with λ as the wavelength of the incoming radiation, r_0 as the classical electron radius and N_A as Avogadro's number. f_1 approaches Z (number of electrons) and shows only a discontinuity at the absorption edges and f_2 is proportional to the absorption coefficient.

The correlation between the critical angle (ϕ_c) and the index of refraction (n) is derived assuming no absorption, i.e. $\beta = 0$ by Equations (5) and (6)

$$n = 1 - \delta = \frac{\cos\phi_c}{\cos 0} = \cos\phi_c = \sqrt{1 - \sin^2\phi_c} \tag{5}$$

$$(1 - \delta)^2 = 1 - 2\delta + \delta^2 = 1 - \sin^2\phi_c \tag{6}$$

neglecting δ^2 and approximating $\sin\phi_c = \phi_c$ which is valid for small angles, Equation (7) follows

$$\phi_c = \sqrt{2\delta} \tag{7}$$

Measuring the reflectivity R which is defined as the ratio between the intensity of the incident beam versus the intensity of the reflected beam, curve R versus angle shows a typical shape. The sharp decrease of R at a certain angle defines the critical angle of total reflection. A remarkable observation is also the high value of R below the critical angle which is close to unity (see Figure 4). The influence of parameter β/δ can also be seen. Higher absorption leads to a ''smearing'' of the step function at ϕ_c.

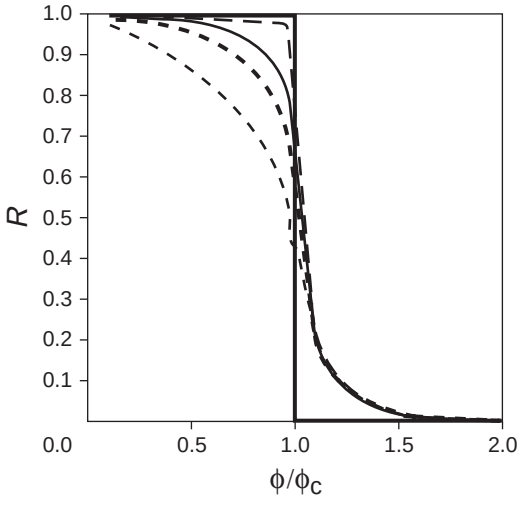

Figure 4 Calculated reflectivity versus incident angle/critical angle of Si for various values of β/δ. ———, $\beta/\delta = 0$; — —, $\beta/\delta = 0.0054$ (17.5 keV); ———, $\beta/\delta = 0.22$ (8 keV); - - - -, $\beta/\delta = 0.049$ (5.4 keV); - - - -, $\beta/\delta = 0.14$ (3 keV).

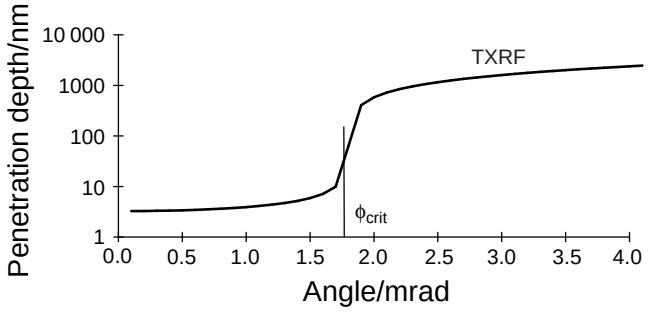

Figure 5 Penetration depth as a function of the ratio of incident angle to critical angle for Cu Kα radiation using a quartz reflector; 17.5 keV in Si.

Relating the behavior of the electromagnetic wave inside the medium to the associated damping effect, a term "penetration depth" can be defined which is the depth z_p normal to the surface where the intensity of the incoming beam at the surface $z = 0$ is reduced to 1/e; a classical definition whenever absorption phenomena are described. Figure 5 shows the penetration depth as a function of the relative incidence angle for Mo Kα radiation.

3 TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS

3.1 Historical Development

In 1971 Yoneda and Horiuchi[17] published a paper with the first experimental data from an EDXRF measurement using a Ge detector and excitation geometry in total reflection geometry from a sample of a dried spot of 100-ng Ni from an aqueous solution of $NiCl_2$ on an "optical flat". In 1974 Aiginger and Wobrauschek[18] published results using a Si(Li) detector and a Suprasil reflector, where 5 μL of Cr salts in aqueous solution were dried and analyzed with a Cu-anode X-ray tube. A more detailed publication by Wobrauschek and Aiginger[19] appeared in 1975 showing theoretical considerations and details of the set-up, calibration curves, steps for the quantification and detection limits of 4 ng of Cr. In 1977 the GKSS group Knoth, Schwenke, Marten and Glauer[20] published the first prototype of a TXRF system and the analytical results of blood serum analysis. From that time on many more teams worldwide have been working on the development of this technique. In particular from 1986 there have been regular biannual meetings, TXRF workshops and conferences for the users of this young analytical technique.[21–25] In 1997 the first book dealing exclusively with TXRF, written by Klockenkämper, became available and gives an excellent summary of the principle, advantages and applications of TXRF.[26]

3.2 Total Reflection X-ray Fluorescence Analysis Experimental Set-up

TXRF is basically an ED analysis with a special excitation geometry. In the standard case of EDXRF the angle between the primary incident radiation and the sample is 45°. In addition the detector is placed normal to the incident beam so that the angle between sample and detector is also 45°. The principle set-up is shown in Figure 6. A narrow almost parallel beam impinges at angles below the critical angle on the surface of the reflector which carries the sample as randomly distributed microcrystals in the center part of its surface. The X-rays scarcely penetrate the reflector and thus the contribution from scattered primary radiation from the substrate is minimized. The sample itself is excited by both the primary and the reflected beam so that the fluorescent signal is practically twice as intense as in the standard EDXRF excitation mode.

One of the differences between EDXRF and TXRF is the excitation geometry which is changed in such a way

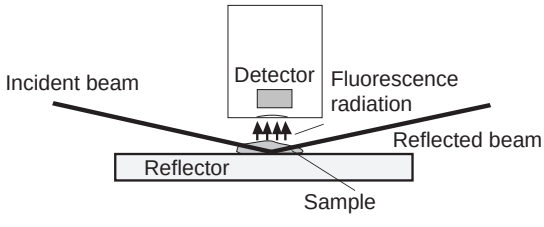

Figure 6 Principle of TXRF; geometrical arrangement.

that the primary radiation impinges at an angle below the critical angle of total reflection on the surface of the reflector carrying the sample. The critical angles are in the range of a few milliradians for typical reflector materials like quartz or Si and primary radiation of 9.4 keV from a W-L or 17.5 keV from a Mo-anode X-ray tube.

With higher energies in the exciting spectrum, adjustments must be made for the proper incident angle below the critical angle which is given, corresponding to Equation (7), with the assumption that $f_1 \approx Z$, and $Z/A \approx 1/2$ by Equation (8):

$$\phi_{\text{crit}} = \frac{20.7}{E}\sqrt{\rho} \tag{8}$$

where ϕ_{crit} is in mrad, E in keV and ρ in g cm^{-3}.

The main advantages of TXRF are:

1. The background caused by scattering of the primary radiation on the substrate is reduced.
2. The fluorescence intensity is doubled as the primary and reflected beams pass through the sample giving efficient excitation.
3. The distance between the sample on the reflector surface and the detector can be made small, thus the solid angle for detection is large.
4. All these advantages lead to lower limits of detection (LLD) compared to the standard EDXRF mode.

Depending on the X-ray source and the spectral modification devices, the LLD are in the ''picogram'' range for 2–3-kW standing anode X-ray tubes and in the ''femtogram'' range with SR.

4 SOURCES FOR TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS AND SPECTRAL MODIFYING DEVICES

The excitation source used most in TXRF is a high-power diffraction X-ray tube with 2–3 kW applied where the line focus is taken. This type of X-ray tube is available with Cr, Cu, Mo, Ag and W anodes. For better excitation of light elements, Al or other special anode materials are chosen. The X-ray tubes need a power supply from an X-ray generator with high voltage heating for the filament and efficient water cooling. The line focus of the anode has to be used so that the brilliance emitted from that tube correlates with the slit-like collimator which is necessary to produce a narrow almost parallel beam with divergence less than the critical angles involved. A larger photon flux on the sample can be achieved technically by using rotating anodes made from Mo or W which can stand up to 18 kW. In all cases the focal size of the electron beam on the anode is a line of dimensions $0.4 \times (8 \text{ or } 12)$ mm. Emission of X-rays is observed at less than an angle of 6° to the anode surface, such that the optical focus is reduced by projection with $\sin 6^\circ$ equal to 0.1–0.04-mm width.

The emitted spectrum is usually a continuum (white) with the highly intensive characteristic lines of the anode superimposed. In modern spectrometers the X-ray tube is usually operated with an applied voltage up to 40 or 60 kV. Photons with an energy of 40–60 keV are also available but not essential for excitation. An excellent tool for modifying the spectral distribution, in a way that suppresses the high-energy photons whereas lower energies including the intensive characteristic lines will be recorded, is the insertion of another reflector in the beam path. The energy dependence of the critical angle is used to achieve this effect. In adjusting for a fixed value for the angle, only photons up to a certain value of E with respect to the angle are reflected. Those having more energy do not fulfill the reflection condition and are absorbed in the reflector material. This effect is comparable to a low-pass filter known from electronics. The reflected beam is of the same intensity as the primary beam but the high-energy photons are missing from its spectrum. These would cause an increase in the low-energy background due to Compton backscattering at the front side of the detector. If the higher energy photon is backscattered, the electron carrying the energy difference is scattered into the detector leading to an increased low-energy background. Thus complete suppression of the unwanted high-energy photons is achieved using a cut-off filter.[13,14,27]

Another more effective way of modifying the exciting spectrum is by using a monochromator for the intensive characteristic primary radiation, e.g. Cr Kα, Cu Kα, W Lβ, Mo Kα. The most common are crystal monochromators, but they have the disadvantage of transmitting a very narrow energy band, in an electronvolt range smaller than the natural linewidth. The more efficient monochromators are layered synthetic microstructures.[28–30] This multilayer (ML) is a perfectly polished flat substrate on the back of which a number of layer pairs with varying high Z and low Z material are sputtered. A few nanometers distance separates the layers and the system acts like a Bragg single crystal selecting a certain energy at a certain angle but with a larger $\Delta E/E$ and high reflectivity. Using such a device, the spectrum measured shows some extraordinary features. The background is practically zero in all regions except for the scattered primary energy. Only the fluorescence lines of interest are observed and detection limits with such an excitation mode are improved.

The excitation of elements is best if the photon energy of the exciting radiation is close to the individual

absorption edge. The search for anode materials for special applications has led to the development of an Al anode or a Si anode for analyzing all the elements below Si, such as F, Na, Mg and Al. The use of the L-lines of Rh or Mo for the light element excitation is also well known.

As there is a steady call for greater and greater improvements in detection limits, an approach has been applied using the most intensive source available at the moment, synchrotron radiation (SR), which has almost ideal features for TXRF. SR is produced by high-energy (GeV) relativistic electrons or positrons circulating in a storage ring. It is several orders of magnitude greater in brightness compared to X-ray tubes, has a natural collimation in the vertical plane and is linearly polarized in the plane of the orbit. The spectral distribution is continuous, so by proper monochromatization the performance of selective excitation under the best conditions is possible. Many applications can only be performed with SR and its combination with TXRF will be mentioned in the section on SR/TXRF. Experiments have been performed by several groups worldwide. The properties of SR are so challenging that much more effort should be made to develop accessibility to SR and in particular to beamlines in which a TXRF system can be installed.

The use of MLs or cut-off reflectors is not helpful if the task is to analyze high Z elements by their K shell emissions, which requires high-energy photons up to more than 120-keV energy. In this case a slit system of high Z elements will be taken to define the geometry of the beam.

All the possibilities for modifying the exciting beam are shown schematically in Figure 7. For further improvements in the spectral background, the scattering and absorption of primary photons in air should be avoided. Therefore the use of a vacuum chamber is recommended for optimized excitation conditions. The Ar K-line which is also excited in air will disappear from the spectrum and absorption of both primary and fluorescence photons is avoided. A background of low intensity is observed, and scattered primary X-rays still remain after interacting with the sample and the substrate, as do the characteristic lines from the reflector. If for any reason related to sample type or due to technological problems measurements in vacuum are not possible, improvements compared to measurements in air are achieved by flushing the chamber with He.

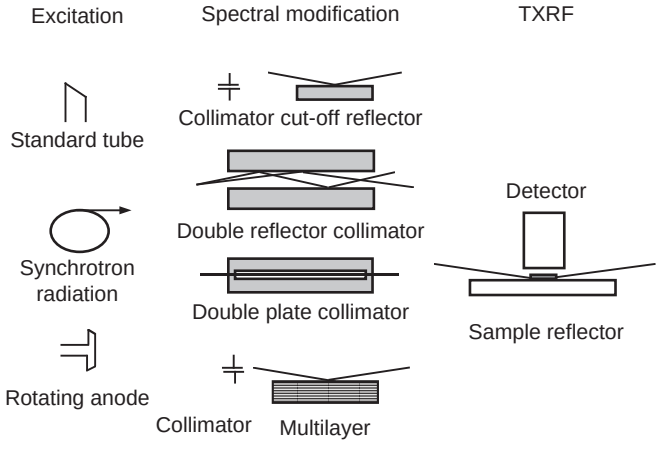

Figure 7 Schematic presentation of possible excitation sources and spectral modification tools.

5 REFLECTOR SPECIFICATIONS: SURFACE QUALITY, MATERIAL, SHAPES

The presence of characteristic lines from the reflector material are sometimes disturbing especially if the analyte has the same elements. In this case the use of other materials is suggested and several materials have been tested. Quartz and synthetics like Synsil, Suprasil, fused silica and elemental Si are generally in use. The heavier elements like Ti, Nb, Ta have been tested and show excellent performance. Boron nitride, glassy carbon and the most simple types of perspex are light element reflectors that show no interfering characteristic lines. All geometric shapes are possible for reflectors, and typically a disk 3–4 mm thick, with a 30-mm diameter, is standard. The only important requirement is the surface quality, the mean roughness should be in the nanometer range. The overall flatness is given in terms of optical wavelength (589 nm), so typically $\lambda/20$ is sufficient for a good surface.

Requirements for the reflectors include:

- mechanical stability
- good machinability for grinding and polishing
- hard surface
- hyperpure material
- chemically resistive against acidic or chemically aggressive samples
- easy to clean for multiple use
- XRF lines of the reflector not in the energy range of interest
- an economic price; the perspex reflector is a few cents, whereas the polished Synsil reflector is about 40 US$.

6 TECHNICAL INSTRUMENTATION FOR TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS

The complete TXRF experimental set-up with the various components is seen in Figure 8. Technically all

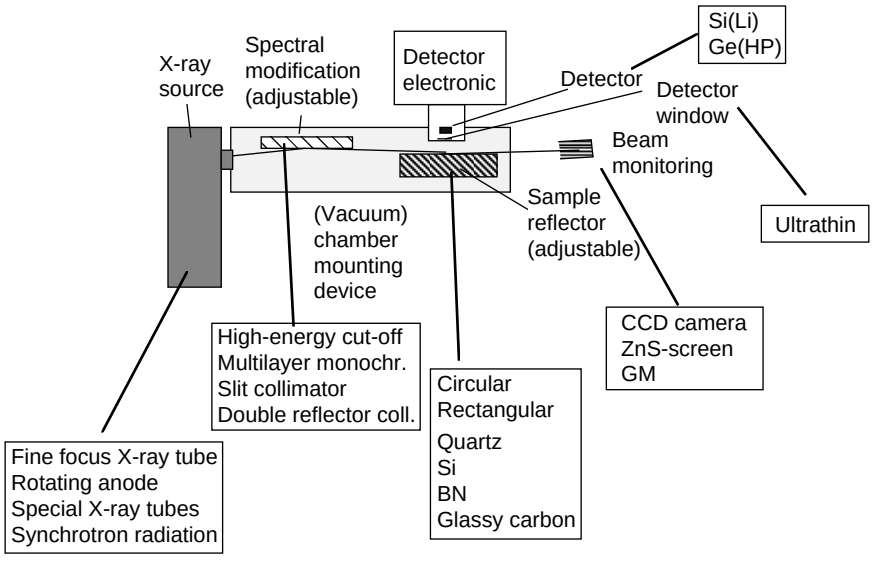

Figure 8 Set-up of a TXRF unit with possible components.

components need to be adjusted by translation and rotation stages and tilters to obtain perfect alignment with the X-ray beam and the total reflection condition. These motions can be mechanically adjusted by hand or remotely by controllers. Another possibility is to use a firm prealigned module with all components fixed, where the only movable part is the sample changer. To observe the effect of total reflection several units can be used as specified in Figure 8. The modern way is the charge coupled device (CCD) camera where the X-rays fall directly on the chip and induce an image where the pixels are hit. This is made visible by a monitor in the normal way. The ZnS screen can be used to watch the X-rays either by eye or again via a CCD camera and lens focused on the screen. Because part of the direct beam passing above the surface of the reflector can be observed, the image for proper alignment is always a double beam – the primary and the reflected one separated according to the adjusted angle. Knowing the distance between the reflector and the screen, one can then calculate the angle adjusted.

To demonstrate the advantages of TXRF in a realistic comparison between the same sample measured in standard EDXRF geometry and TXRF set-up, the spectrum of a reference standard water 1634a from NIST (National Institute of Standards and Technology) is shown in Figure 9. The spectrum in Figure 9(a) was measured by putting a droplet of 10 µL on a Kapton foil (8 µm), drying the sample and inserting it with a spectrocup into a standard EDXRF spectrometer with a Rh tube operated at 40 kV and 0.35 mA, leading to 65% deadtime under vacuum conditions, which were found to be the optimal conditions. Also 10 µL of the sample was dried on a quartz reflector and inserted in the TXRF spectrometer equipped with a Mo tube, a ML monochromator and a vacuum chamber (40 kV, 50 mA) to obtain the spectrum shown in Figure 9(b). Optimized conditions were applied in both cases for a fair comparison. The lower detection limits and the determination of elements which were not seen in standard EDXRF but clearly in TXRF, demonstrate the analytical power of the method.

To demonstrate the development of the method over time and the improvements in the detection limits achieved with TXRF, the following data for the detection limits are given. In 1975 LLDs were in the nanogram range, but by 1985 improvements in spectral modification, excitation geometry and detection systems led to LLDs in the picogram range. Using SR, in 1995 LLDs at femtogram levels were achieved.

1975	⟶	1985	⟶	1995
nanograms		picograms		femtograms

These values are extrapolated values, obtained from interference-free single-element standard solutions. All the figures were calculated from the accepted definition of detection limits for spectrometric methods recommended by IUPAC (International Union of Pure and Applied Chemistry), expressed by the simple relation, Equation (9), between background and sensitivity

$$\mathrm{LLD} = \frac{3}{S}\sqrt{\frac{I_\mathrm{B}}{t}} \tag{9}$$

where S is the sensitivity in counts per second of the respective net signal divided by the mass or concentration

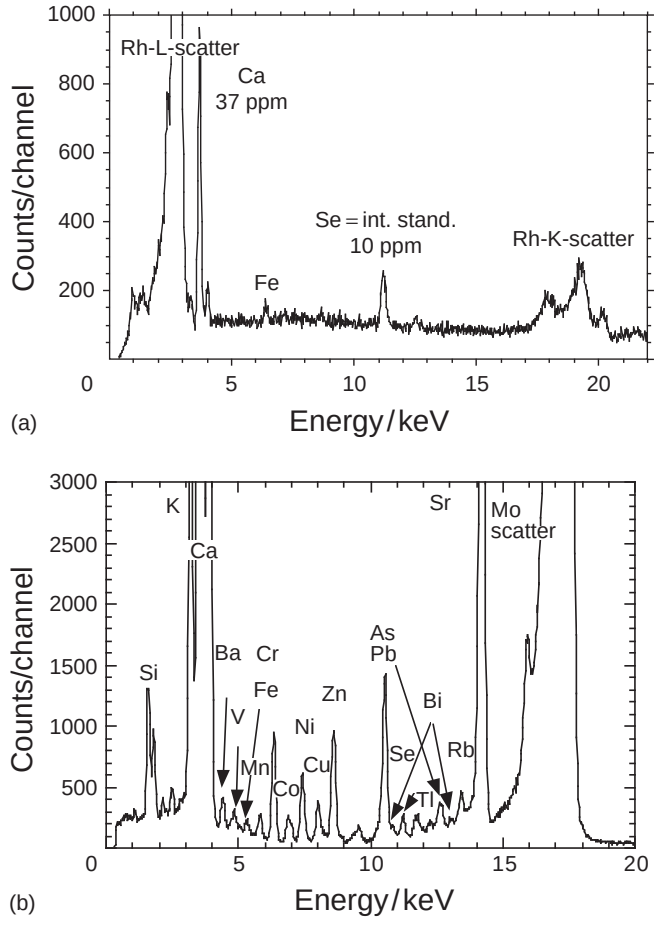

Figure 9 Comparison of a sample (10 µL of NIST water SRM 1634) analyzed by standard EDXRF (Rh low-power tube, dried on Kapton 8 µm) (a) and TXRF (ML monochromatized Mo tube, vacuum chamber) (b).

of the element present, I_B is the background intensity and t the measuring time in seconds.

7 DETECTORS AND ELECTRONICS

The detectors used in TXRF are semiconductor detectors. The lithium drifted silicon (Si(Li)) detector, the hyperpure germanium (GeHP) and from older days the lithium drifted Ge (Ge(Li)) detectors are used. Their main advantages are compact size, nonmoving system components, and excellent energy resolution, where the best have 120 eV at 5.9 keV. Because of their operation principles, these detectors have an inherent simultaneous multielement capacity, which leads to a short measuring time for all elements as the detectors select the energy and collect counts at the same time. Some disadvantages include the need for liquid nitrogen (LN) cooling during operation, the necessity of having a very thin 8–25 µm sensitive Be window and the fact that the maximum processable number of counts is limited to about 40 000 cps. This figure can be increased to 100 000 cps, but with loss of the excellent performance data. The operation principle is as follows. The detector is a disk of very pure Si or Ge with dimensions of 4–10-mm diameter and 3–5-mm thickness. Even careful production of the Si ingot where this disk is cut will have some trace impurities. To compensate and bind all free electrons, lithium ions drift in the silicon crystal to neutralize the Si crystal defects in a particular zone, the so-called intrinsic zone. Au contacts are evaporated and a reverse voltage applied. To keep the leakage current as low as possible the crystal is cooled with LN, so that the detector has to be operated in vacuum and the radiation to be measured has to pass a thin entrance window, usually Be.

Applying a reverse voltage in the charge carrier free intrinsic zone, an absorbed X-ray photon is converted into charge by ionization. In the Si crystal, electron–hole pairs are created and their number is the ratio between the energy of the absorbed photon and the energy required to create one electron–hole pair. The electrons and holes are quickly swept to the contact layers by the electric field created by the applied reverse bias on the crystal.

Figure 10 shows the operation principle schematically. The charge induces a signal at the gate of the cooled field effect transistor (FET) which is the input stage of the charge sensitive preamplifier. The output signal is fed to a pulse processor which shapes the pulse and amplifies it further. This signal is in the range up to 10 V and is proportional to the energy of the absorbed photon. The signal is stored digitally, so the analog-to-digital converter (ADC) prepares the digitized signal. Finally all signals are stored in a multichannel analyzer (MCA) where the event is stored according to the digitized number in that channel.

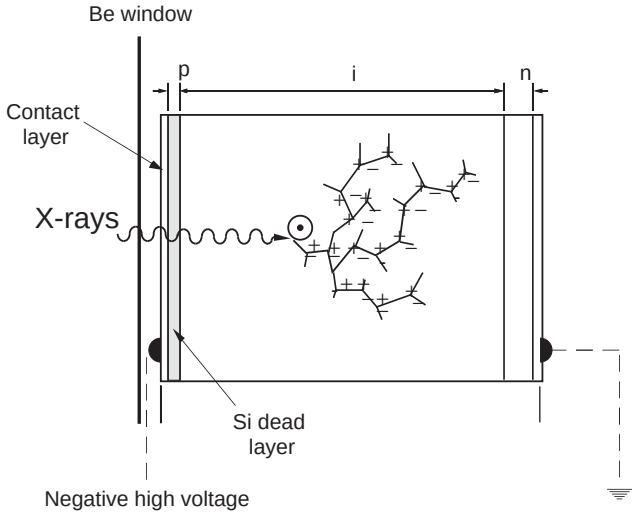

Figure 10 Scheme of the working principle of a Si(Li) detector.

All events created from photons with the same energy are stored in one channel and the result will be a line in the spectrum where the number of events is the *y*-axis and the energy is the *x*-axis. Even trying hard to reduce all noise contributions in the electronics, there are no ideal lines. The natural linewidth of the characteristic radiation, the statistical fluctuations in the collection process, and the noise from resistors and FET, all cause a broadening of the measured X-ray lines. The energy resolution, which is given at the full width at half-maximum (fwhm) of the Gaussian distributed X-ray line, demonstrates the quality of a detector. The fwhm is proportional to the square root of the energy, therefore it is necessary to define a reference energy, the Mn Kα energy at 5890 eV. Another attribute demonstrating the merit of an X-ray detector is its efficiency: the ratio of the number of counted photons to the number of incoming photons.

A typical calculated efficiency curve versus energy is shown in Figure 11. The reason for the decrease in efficiency for low-energy photons is the absorption of the photons in all the layers between the incoming photon and the sensitive volume. The layers are: Be window, gold contact layer and dead layer of the p-type Si material. On the high-energy side the absorption of the high-energy photons in the detector material itself is reduced (lower mass absorption coefficient) and fewer pulses result. From the efficiency it can be seen that for the low- and the high-energy side some precautions have to be made if the elements of interest are emitting low- or high-energy photons.

The light elements emit fluorescence radiation in the range 100–1000 eV. Be windows would not allow them to pass, so new window materials have to be used offering high transmission for low-energy radiation. For K shell radiation from high *Z* elements a better absorbing material is needed, like germanium. A high-quality detector offers high energy resolution (low fwhm),

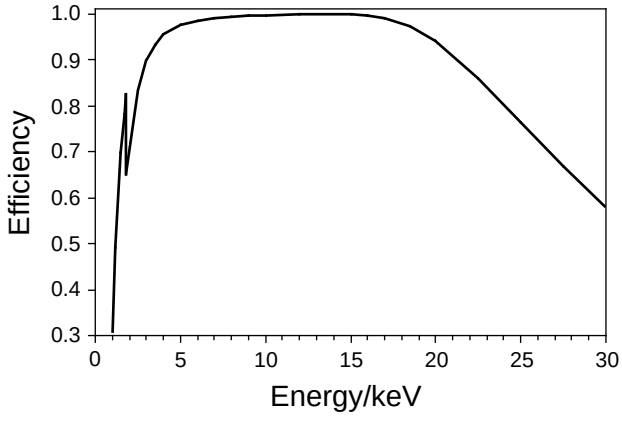

Figure 11 Typical calculated efficiency curve of a Si(Li) detector, 3 mm thick.

a high efficiency, symmetric peak shape, and a small detector contribution to peak and background, leading to a high peak to background ratio. The background is influenced by the detector quality as well as by the source sample interactions. Incomplete charge collection at the electrodes leads to low-energy tailing of the peak and a higher remaining background. Additionally the escape effect creates escape peaks and thus an increased background. This escape effect occurs in the outer zones of the crystal where the incoming photon is absorbed close to the surface of the crystal. If the energy is high enough the crystal material itself is excited to emit its own fluorescence Si K or Ge K or L lines. From the near surface these fluorescence photons can escape and do not contribute to the charge carrier production. This causes an additional line in the spectrum, called an escape line, which is shifted in energy just by the value of the energy of the characteristic Si or Ge radiation. Overlap with fluorescence lines can occur and has to be kept in mind. Nevertheless the software for the spectral deconvolution of the measured peaks is so sophisticated that safe interpretation is possible. Another inherent advantage of semiconductor detectors is the possibility of bringing the detector crystal very close to the sample. Assuming that the crystal is 3–4 mm behind the Be window and a 1-mm distance from sample to entrance window was chosen, this leads to a total distance of 5 mm which gives rise to a large solid angle. The physically largest component of the detector is the storage tank for LN. This Dewar has a storage capacity from 5 to 30 L. The system inside must be under vacuum, so a housing "snout" is kept around the crystal, shielding it from the outside environment. The modern detector can be mounted in all positions, vertical, horizontal and inclined because the Dewar has a special filling and venting system.

Two recent developments show an interesting approach to improving two problematic features of the ED detector. LN filling can be a problem if there is no delivery or production. The detector, with the exception of Ge(Li) types, can warm up. The use of Peltier-cooled detectors is a step in the direction of independence of the LN supply. A new generation of ED detectors with a Si PIN diode and Peltier cooling are resulting in lightweight hand-sized units with acceptable resolution of 160–200 eV.[31] Si drift chambers can also be used. Another development is in cryogenic detector technology, a superconducting tunnel junction operating at a few millikelvin to achieve energy resolution of 10 eV at 5900 eV.[32] Another group uses the idea of a microcalorimeter where the temperature change created by the absorbed photon is measured.[33] These developments are in progress and can provide a big step forward for specific applications.

The signal from the preamplifier has to be pulse shaped and amplified. This is done by electronic units

named pulse processors. Gaussian amplifiers or triangular shapers are state of the art, but over the last few years time variant filter techniques and digital pulse processing have been introduced. The basic problem is the fact that an increased throughput results in a decrease of resolution. A reduced shaping time (e.g. from 12 to 2 μs) increases the throughput of pulses but a loss in energy resolution is observed in the case of a Gaussian amplifier. The digital pulse processor (DPP), processing the same input rate, improves the resolution compared to the Gaussian. Thus the count rate may be increased in the case of the DPP, improving the throughput of pulses and maintaining the energy resolution as for the Gaussian. Many more approaches have been suggested by electronic experts and companies to obtain an optimized electronic chain which allows a high throughput at best resolution. In particular at high count rates the number of counts arriving may exceed the processable count rate. The system is not ready to accept counts as it is busy with the processing. This dead time must be added to the preset time required to collect data. In conclusion, the components required for EDXRF are:

- detector
- preamplifier
- detector high-voltage bias supply
- shaping amplifier
- ADC
- MCA
- a PC plug-in board
- software for data deconvolution
- data storage
- data display
- output and printing unit.

8 QUANTIFICATION IN TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS

In XRF the intensity of the fluorescence radiation depends on the following parameters, Equation (10)

$$I(E^i_{K\alpha}) = \int_{E=E^i_{abs}}^{E=E_{max}} \int_{x=0}^{x=d} I_0(E) G_1 \frac{\rho}{\sin\varphi} \frac{\tau^i_K(E)}{\rho} w^i_K p^i_\alpha c^i V^i(E) \times \exp\left(-\left(\frac{\mu(E)}{\rho \sin\varphi} + \frac{\mu(E^i_{K\alpha})}{\rho \sin\psi}\right) \rho x\right) \times G_2 f(E^i_{K\alpha}) \varepsilon(E^i_{K\alpha}) \, dx \, dE \quad (10)$$

where i is the index of the respective element, E_{abs} is the energy of the absorption edge of element i, E_{max} is the maximum energy of the excitation spectrum, d is the thickness of the sample perpendicular to the sample surface, $E_{K\alpha}$ is the energy of the Kα-radiation of element i, $I_0(E)dE$ is the spectral distribution of the exciting radiation, x is the depth of the inspected layer perpendicular to the sample surface, dx is the thickness of the inspected layer, G_1, G_2 are geometry factors, ρ is the density of the sample, φ is the incidence angle of the exciting radiation, ψ is the take-off angle of the fluorescence radiation, $\tau_K(E)/\rho$ is the photoelectric mass absorption coefficient of the K shell of element i, $\mu(E)/\rho$ is the mass absorption coefficient of the sample, w_K is the fluorescence yield of the K shell, p_α is the emission probability of the Kα-line, c is the concentration, $V(E)$ is a factor, taking secondary excitation into account, $f(K\alpha)$ is the absorption of the fluorescence radiation between sample and detector and $\varepsilon(K\alpha)$ is the detector efficiency.

After integration of Equation (10) for x the absorption term A can be defined by Equation (11):

$$A(E) = \frac{1 - \exp\left(-\left(\frac{\mu(E)}{\rho \sin\varphi} + \frac{\mu(E^i_{K\alpha})}{\rho \sin\psi}\right) \rho d\right)}{\left(\frac{\mu(E)}{\rho \sin\varphi} + \frac{\mu(E^i_{K\alpha})}{\rho \sin\psi}\right) \rho d} \quad (11)$$

Terms depending only on physical properties of the atoms are called *fundamental parameters*. σ is the fluorescence cross-section and is tabulated for various excitation energies and elements, Equation (12)

$$\sigma^i_{K\alpha}(E) = \frac{\tau^i_K(E)}{\rho} w_K p_\alpha \quad (12)$$

Therefore, dealing with low concentrations where the factor $V(E) = 1$, Equation (10) can be written by Equation (13)

$$I(E^i_{K\alpha}) = \int_{E=E^i_{abs}}^{E=E_{max}} I_0(E) G_1 \frac{\rho d}{\sin\varphi} \sigma^i_{K\alpha}(E) c^i A(E) \times G_2 f(E^i_{K\alpha}) \varepsilon(E^i_{K\alpha}) \, dE \quad (13)$$

Theoretically it is possible to perform a quantitative analysis with Equation (13) without standard samples, but several problems have to be considered. The mass absorption coefficient of the sample $\mu(E)/\rho$ is determined by the composition of the sample, a summation about all elements in the sample is necessary. Thus an iterative procedure has to be applied for a solution. A further problem is knowledge of the spectral distribution of the exciting radiation.

Assuming a monochromatic beam, the integral of Equation (13) can be removed and all parameters depending on the energy of the exciting radiation, like $\sigma^i_{K\alpha}(E) \equiv \sigma^i_{K\alpha}(E_0)$, $A(E) \equiv A(E_0, E^i_{K\alpha})$ become more

simple, if E is substituted by the monochromatic excitation energy E_0.

$$I(E^i_{K\alpha}) = I_0(E_0)G_1\frac{\rho}{\sin\varphi}\sigma^i_{K\alpha}(E_0)c^i A(E_0)G_2 f(E^i_{K\alpha})\varepsilon(E^i_{K\alpha}) \tag{14}$$

If the sample is very thin, $(x \to 0)$, the integration in Equation (14) becomes $\lim_{d\to 0} \frac{1-e^{-ad}}{ad} = 1$. Absorption can be neglected, as well as secondary enhancement. The intensity of the fluorescence radiation is then given by Equation (15)

$$I(E^i_{K\alpha}) = \int_{E=E^i_{abs}}^{E=E_{max}} I_0(E)G_1\frac{\rho d}{\sin\varphi}\sigma^i_{K\alpha}(E)c^i G_2 f(E^i_{K\alpha})\varepsilon(E^i_{K\alpha})\,dE \tag{15}$$

with (ρd) as the areal mass density of the sample $(\rho d = m/F)$.

In this special case the relation between concentration and fluorescence intensity is linear. The so-called *sensitivity* (cps/sample mass) can be defined, Equation (16):

$$S^i := \frac{I^i}{c^i m} \tag{16}$$

S depends only on fundamental parameters and the measuring conditions, which can be assumed to be constant in a spectrometer or an experimental set-up. Thus the *relative sensitivities* (*sensitivity factors*) of various elements i referring to the intensity of a standard element I^s (*internal standard*) can be determined, Equation (17):

$$S^i_{rel} = \frac{I^i}{I^s} = \frac{\int_{E=E^i_{abs}}^{E=E_{max}} I_0(E)\sigma^i_{K\alpha}(E)f(E^i_{K\alpha})\varepsilon(E^i_{K\alpha})\,dE}{\int_{E=E^s_{abs}}^{E=E_{max}} I_0(E)\sigma^s_{K\alpha}(E)f(E^s_{K\alpha})\varepsilon(E^s_{K\alpha})\,dE} \tag{17}$$

with s being the index for the internal standard element.

This definition is valid for the same concentration of the element i and the internal standard element. Relative sensitivities can be determined experimentally with standard samples or calculated theoretically if all parameters are known (thin film approximation, monoenergetic excitation, detection efficiency = 1). Owing to the ratio, all geometry factors cancel.

The determination of the concentration c^i of an unknown element in a sample spiked with an internal standard s is simple, if relative sensitivities S^i_{rel} are determined, Equation (18)

$$c^i = \frac{I^i}{I^s}\frac{1}{S^i_{rel}}c^s \tag{18}$$

One of the inherent advantages in TXRF is that the sample can be assumed to be thin, so the simple conversion of fluorescent intensities I into concentration data c is valid, as there is a linear correlation between intensity I and concentration c. After establishing a calibration curve either from known multielement standards or by using the fundamental parameters to calculate the calibration curve theoretically, the conversion of I into c can be performed immediately. The addition of one element as internal standard of known concentration to the sample improves the accuracy of the results, because geometric and volumetric errors will cancel out and the inhomogeneous excitation caused by the standing wave field above the reflecting surface vanishes.

Figure 12 shows the linear calibration curves with respect to Se as internal standard for Fe, Zn and Sr, obtained from standard solutions, from which the relative sensitivity can be determined. Figure 13 shows

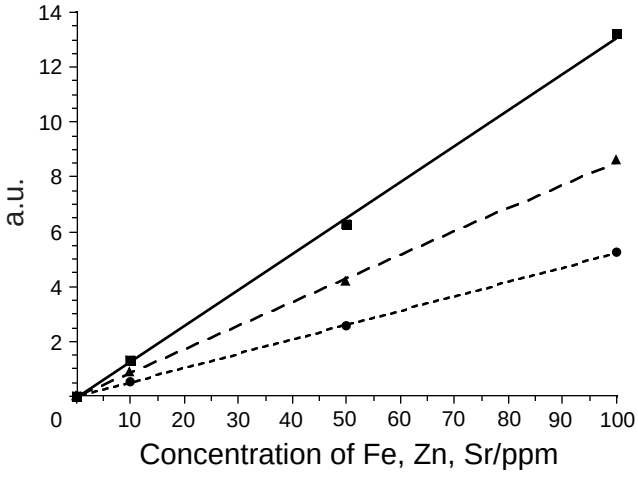

Figure 12 Calibration curve of Fe, Zn and Sr relative to Se as internal standard. • I(Fe Kα)/I(Se Kα); I(Zn Kα)/I(Se Kα); ■ I(Sr Kα)/I(Se Kα).

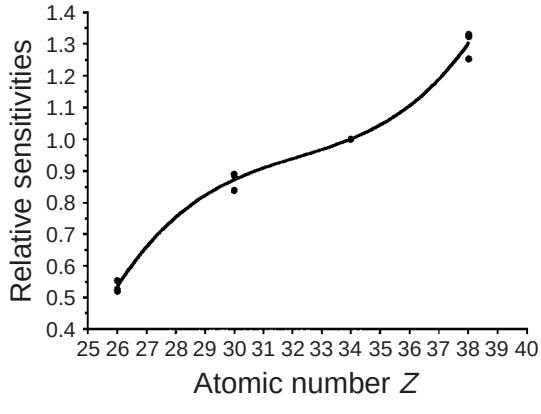

Figure 13 Measured and extrapolated relative sensitivities versus atomic number.

the relative sensitivities versus atomic number and a fitted curve through the data points allowing interpolation of measured data for elements not measured. To calibrate the spectrometer it is necessary to determine the relative sensitivities once. Then any unknown sample can be easily quantified after adding the respective internal standard.

A sample is ''thin'' if the thickness does not exceed the critical thickness (d_{max}) calculated by Equation (19)

$$d_{max} = \frac{0.05/\rho_{matrix}}{(\mu/\rho)_{sample}} \quad (19)$$

This is about 4 μm for organic tissue, 0.7 μm for mineral powder and 0.01 μm for metallic smears, following Klockenkämper.[26] However, on the assumption that the matrix absorption for the analyte differs only slightly from that of the internal standard, these critical thicknesses are much greater.

9 SAMPLE PREPARATION

The preparation of the sample is a very important factor influencing the achievement of satisfying detection limits. Generally Figure 14 shows the various common and applied methods of sample preparation in TXRF depending on the kind of sample to be analyzed. One of the big advantages of TXRF is the small sample volume required for analysis, only a few nanograms of sample mass or a few microliters of sample volume are required. This is in contrast to other sensitive analytical methods for trace analysis and in some cases is a great advantage.

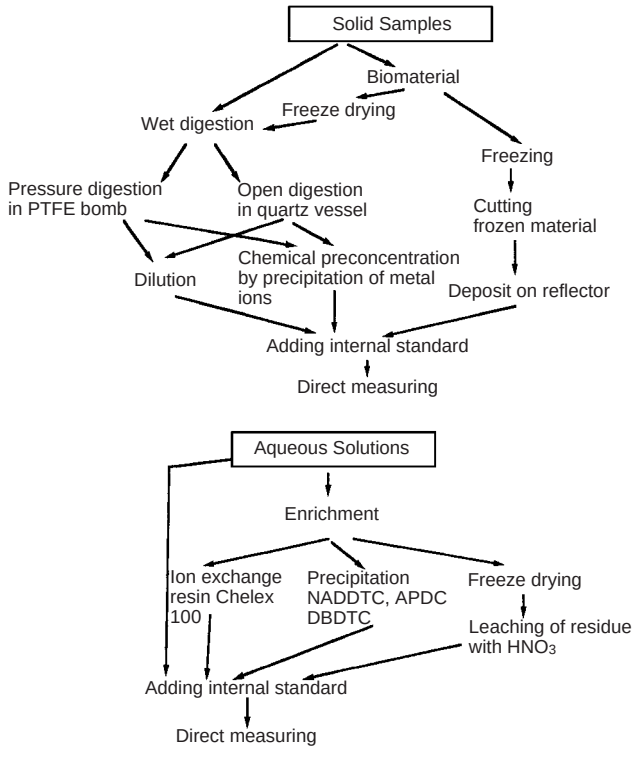

Figure 14 Sample preparation methods for TXRF. PTFE, polytetrafluoroethylene; NADDTC, Na-diethyldithiocarbamate; APDC, ammonium pyrrolidinedithiocarbamate; DBDTC, dibenzyldithiocarbamate.

First of all the sample must be representative of the whole specimen, therefore homogenization is required. The easiest way is the direct analysis of liquid samples. A few microliters (of the sample solution mixed with the internal standard) are pipetted on to the sample reflector and dried, either in a desiccator or with infrared light or on a hot plate. Depending on the matrix of the sample, this matrix should be digested to eliminate the low Z elements like C and O. As an alternative, solid samples can be crushed and then ground to a fine powder of micrometer grain size. This powder can be mixed with a liquid to produce a suspension which can be pipetted, after adding an internal standard, on to the sample reflector. The pulverized sample can also be dissolved in a suitable solvent and after adding the internal standard an aliquot is pipetted on to the sample reflector and dried.

Various methods for decomposing biological and environmental materials are used. Biomaterial can be decomposed using a low-temperature oxygen plasma asher. The ash can then be dissolved in acid. The most popular method for decomposing biological and environmental samples like plants, tissues, and sediments is wet digestion in Teflon™ vessels (Teflon bombs). It is possible to use the open or closed version with high pressure or without. Strong acids like HNO_3, HF (needed to decompose sediments and rocks), $HNO_3 + HCl$ or $HNO_3 + H_2O_2$ in different proportions are mostly used. HF can be used only at low concentrations if quartz glass reflectors are used, because this acid attacks them. The use of microwaves often reduces the time of digestion to less than 1 h. Extraction of traces of organic phases or reverse phase chromatography can also be used for preconcentration of the traces.

Freeze drying is a method of reducing the volume of a solution or any sample containing water. The sample is frozen and water is evaporated under vacuum conditions. The dried residue can be dissolved in small volume of acid or wet digested. It is also possible to extract traces of certain elements by phase separation. An organic solvent is added to a given volume of sample in an aqueous solution, which has an appropriate pH and is spiked with an internal standard, and mixed thoroughly. Then the two phases are separated. Traces of the metal ions stay in the organic phase, while the matrix elements are left in inorganic solutions. The organic liquid can be directly pipetted on to the reflector. The separation of traces by

Table 1 Influence of sample preparation on detection limits (following Klockenkämper[26])

Sample	Drying	Freeze-drying	Chemical matrix separation	Open digestion	Ashing	Suspension	Solution	Pressure digestion	Freeze cutting
Rain, river water	0.1–3 ng mL^{-1}	20–100 pg mL^{-1}	3–20 pg mL^{-1}	1–3 ng mL^{-1}					
Blood, serum			Digest: 2–30 ng mL^{-1}	20–80 ng mL^{-1}	40–220 ng mL^{-1}				
Air dust, ash, aerosols				5–200 µg g^{-1}		10–100 µg g^{-1}	0.1–3 µg g^{-1}		
Air dust on filter					0.6–20 ng cm^{-2}			0.2–6 ng cm^{-2}	
Suspended matter				3–25 µg g^{-1}		10–100 µg g^{-1}			
Sediment						10–100 µg g^{-1}		15–300 µg g^{-1}	
Powdered biomaterial						1–10 µg g^{-1}		0.2–2 µg g^{-1}	
Fine roots				1–10 µg g^{-1}	Digest: 0.1–1 µg g^{-1}				
High-purity acids	5–50 pg mL^{-1}								
Tissue, foodstuffs, biomaterials									0.5–5 µg g^{-1}
Mineral oil							1–15 µg g^{-1}		
Mussels, fish								0.1–1 µg g^{-1}	
High-purity water		1 pg mL^{-1}							

adding a chelating agent and precipitating the metal ions is another commonly used technique. The metal complexes are filtered through a membrane filter and dissolved in a suitable organic solvent.

Generally loss of elements and contamination by other elements can occur during sample preparation.

Pigments from oil paints can be sampled by using a cotton bud. A few nanograms of material are rubbed from the painting and deposited via the cotton bud on to the sample reflector. Aerosols can be directly collected on sample carriers, used as an impaction stage. Individual particles can simply be put directly on the sample carrier. Solid samples can also be rubbed on the reflector; a small amount of sample material is smeared on the carrier. Biomaterial can be freeze-cut by a microtome and the frozen section placed on the reflector.

For surface and thin layer analysis, the sample has to offer a smooth polished Si surface and can be analyzed directly. A very common preconcentration technique used for the analysis of Si-wafer surfaces is the VPD (vapor phase decomposition) technique. The wafer is exposed to HF vapor, leading to dissolution of the SiO_2 surface layer of the wafer.[34] The impurities can then be collected by scanning the surface systematically with a droplet of a special liquid or water. This integrates the impurities of the complete surface in the droplet, which can then be dried and measured by TXRF.

Generally the sample preparation method has a tremendous influence on the detection limits that can be obtained, documented in Table 1.

Details of sample preparation procedures can be found in Prange and Schwenke,[35] Holynska et al.,[36] and Dargie et al.[37]

10 APPLICATIONS

TXRF allows the simultaneous determination of trace elements in samples of small volume. The three advantages, namely multielement capability, LLD and small sample volume required characterize the analytical capacity of TXRF. The additional advantages of insensitivity to matrix effects, easy calibration, fast analysis and low costs lead to almost unlimited applications. Table 2 gives an overview of various kinds of samples already analyzed with TXRF.

Generally all kinds of aqueous or acidic liquids (where the liquid matrix is evaporated) and also solids, can be analyzed, a small amount being deposited on a sample reflector. Oils, alcohols, whole blood and blood serum can be analyzed after special treatment.

10.1 Environmental Samples

The almost optimal sample for TXRF is the water sample. Pure water like rain or tap water can be analyzed directly.[38,39] River water or seawater, as well as wastewater, require sample preparation to remove suspended matter or the salt content. Figure 15 from Klockenkämper[26] shows the spectrum of a rainwater sample with Ga added as internal standard. 10–100 µL

Table 2 Applications of TXRF

Application area	Substance/research tool	Sample
Environment	Water	rain, river, sea, drinking water, wastewater.
	Air	aerosols, airborne particles, dust, fly ash.
	Soil	sediments, sewage sludge.
	Plant material	algae, hay, leaves, lichen, moss, needles, roots, wood.
	Foodstuff	fish, flour, fruits, crabs, mussels, mushrooms, nuts, vegetables, wine, tea.
	Various	coal, peat.
Medicine/biology/ pharmacology	Body fluids	blood, serum, urine, amniotic fluid.
	Tissue	hair, kidney, liver, lung, nails, stomach, colon.
	Various	enzymes, polysaccharides, glucose, proteins, cosmetics, bio-films.
Industrial/technical applications	Surface analysis Implanted ions Thin films	wafer surfaces.
	Oil	crude oil, fuel oil, grease.
	Chemicals	acids, bases, salts, solvents.
	Fusion/fission research	transmutational elements in Al + Cu, iodine in water
Mineralogy		ores, rocks, minerals, rare earth elements.
Fine arts/ archaeology/ forensic		pigments, paintings, varnish. bronzes, pottery, jewelry. textile fibers, glass, cognac, dollar bills, gunshot residue, drugs, tapes, sperm, fingerprints.

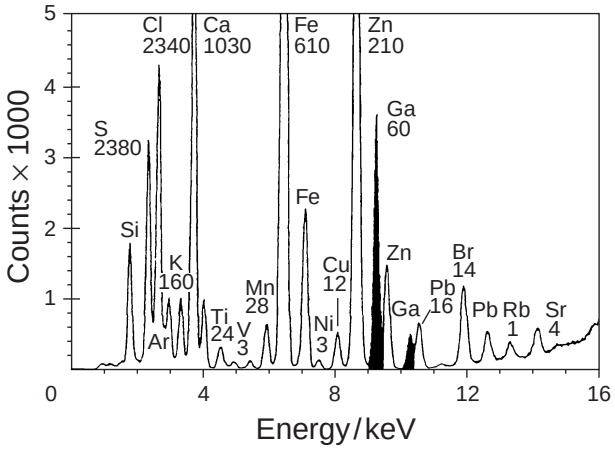

Figure 15 Typical TXRF spectrum of rainwater sample with 60 ppb Ga added as internal standard. (Reproduced by permission of John Wiley & Sons from Klockenkämper.[26])

are required and dried on a clean sample reflector. Detection limits are in the low nanogram per milliliter range. To reduce the detection limits down to the 10 pg mL^{-1} level, freeze-drying with additional leaching of the residue with HNO_3 have to be applied. To remove the salt matrix from seawater, complexation, chromatographic adsorption and subsequent elution have to be performed. This method was developed by Prange et al.[40] Suspended matter separated from filtrates can be analyzed by digesting the filter with HNO_3.[41] For the analysis of sediments microwave digestion is recommended.[42]

Aerosols can be analyzed by collecting on a filter and digesting the loaded filter material, or more conveniently, by direct sampling on the sample reflector, deposited in an impactor stage. Plexiglas™ (Perspex) carriers can be used. If the collected air is wet, the aerosols are reliably deposited on the carrier; if the air is dry, they can be bounced off. This can be prevented by coating the reflectors with a thin film of suitable grease.[43] The reflectors loaded with collected aerosols can be directly analyzed after adding an internal standard. Detection limits are in the range of 0.1 ng for 1 h collection time and a sampling volume of 0.5 m^3. Owing to direct analysis systematic errors due to digestion are avoided.[44]

Environmental monitoring can be done by analyzing appropriate plants like moss or lichen (biomonitoring). All individual components of plants can be analyzed after sample preparation. After cleaning, shredding, freeze-drying and finally pulverizing, the powder is mixed with HNO_3 and digested in a Teflon™ bomb.

It is recommended that vegetable oil be diluted with toluene and an oil-based standard should be added as internal standard.[45] After pipetting an aliquot on the sample reflector, it is heated up to evaporate the volatile parts of the matrix. Detection limits are in the range of 3–20 ng mL^{-1}. Low-temperature plasma ashing is also suitable; the residue has to be dissolved by HNO_3.

10.2 Medical Applications

There is an extensive interest in trace element determination in the medical field, because either depletion or accumulation of trace elements can effect the biological functions of human beings. Whole blood and blood serum, amniotic fluid, organ tissue, hair and dental plaque were analyzed. Detection limits down to 20 ng mL^{-1} for body fluids after microwave digestion[46] and 10 ng g^{-1} for freeze cutting were reported.[47] A small piece of tissue is frozen and then cut by a microtome in thin sections of about 15-μm thickness. The section is then placed on a sample reflector and spiked with an internal standard. This method was applied to tissue of kidney, liver and lung.

10.3 Industrial Applications

Quality control is performed by analyzing ultrapure reagents, like acids, bases and solvents.[48] High-purity metals like aluminum,[49] iron or silicon[50] can also be measured after appropriate matrix separation.

Another field is the analysis of mineral and synthetic oils. Crude oils,[51] lubricating oils,[52] motor oils[53] and diesel fuel[54] have been analyzed. Light oils can be diluted with chloroform or toluene, the volatile matrix then removed by evaporation, and detection limits of micrograms per milliliter are achieved. If detection limits of nanograms per milliliter are required, low-temperature oxygen plasma ashing is necessary.

10.3.1 Analysis of Wafer Surfaces

The real breakthrough in industrial TXRF applications came, when it was discovered that optically flat, polished wafer disks, made from pure Si are suitable for investigation with TXRF. In order to remain competitive, the semiconductor industry is being constantly forced to increase the performance and reduce the cost of integrated circuits (ICs) by shrinking device dimensions and increasing the number of devices per unit area on IC chips. Strong correlations have been found between the presence of metal contamination on the wafer surface and process yields. Even very low concentrations of atomically dispersed metals can create atomic scale defects which lead to leakage of current, gate insulator breakdown, or poor threshold voltage control. All of these can result in device failure or reliability problems. Examples of metals which can cause degradation include Na, Al, Ca, Fe, Ni, Cu and Zn. In decreasing the size of devices the acceptable level of metal contamination decreases as

well. The ability to measure low levels of contamination is crucial for developing techniques to remove these contaminations. The Semiconductor Industry Association (SIA) 1997 National Technology Roadmap for Semiconductors (NTRS) projects that the required sensitivities for transition metals will be at 3×10^8 atoms cm^{-2} in the year 2001.[55] At the moment (1999) 10^9 atoms cm^{-2} level is reached by conventional TXRF using standard tube excitation (Atomika TXRF8030W, http://www.atomika.com) or rotating anode tube excitation (Philips TREX 630 T/S, http://www.analytical.philips.com; Rigaku 3750 TXRF).

The advantage of measuring the surface contamination with TXRF is that the wafer can be mapped by moving the wafer in the plane below the detector to obtain contamination information from each point of the wafer. The inspected area is determined by the detector aperture and detector crystal size. Spatially resolved analysis is completely nondestructive and can be performed ''in-fab'', close to the point of fabrication. Figure 16 from Berneike[56] shows a wafer mapping for S, Br, Fe and Ca. It can be seen, that the spatial distribution of contamination is different for different elements.

If the wafer is treated with VPD (see section 9 on Sample Preparation), all impurities of the oxide layer of the whole wafer surface are collected in one drop, leading to a detection capability of 10^6 atoms cm^{-2}, but losing the advantage of local information as well as nondestructiveness.[57]

To measure surface contamination on wafers a special set-up is required, allowing measurement of the wafer without any surface contact and with the possibility of an angle scan. Measuring the angle dependence of the fluorescence signal allows the form of the contamination, film-type or particulate-type to be distinguished (see section 11 on Characterization of Near Surface Layers).[58] From this kind of measurement it is also possible to check whether the signal comes from surface contamination or whether the atoms are below the surface.

10.3.2 Analysis of Thin Layers

Measuring the angle dependence of the fluorescence signal allows analysis of single or MLs on smooth substrates like wafers. Composition, density and thickness can be determined. Figure 17 from Weisbrod et al.[59] shows an example of a bilayer system, the first layer consists of a Cr–Fe–Ni alloy, the second layer of Pd being followed by a Si wafer substrate. From the measured angle profile the thickness of the alloy layer was determined

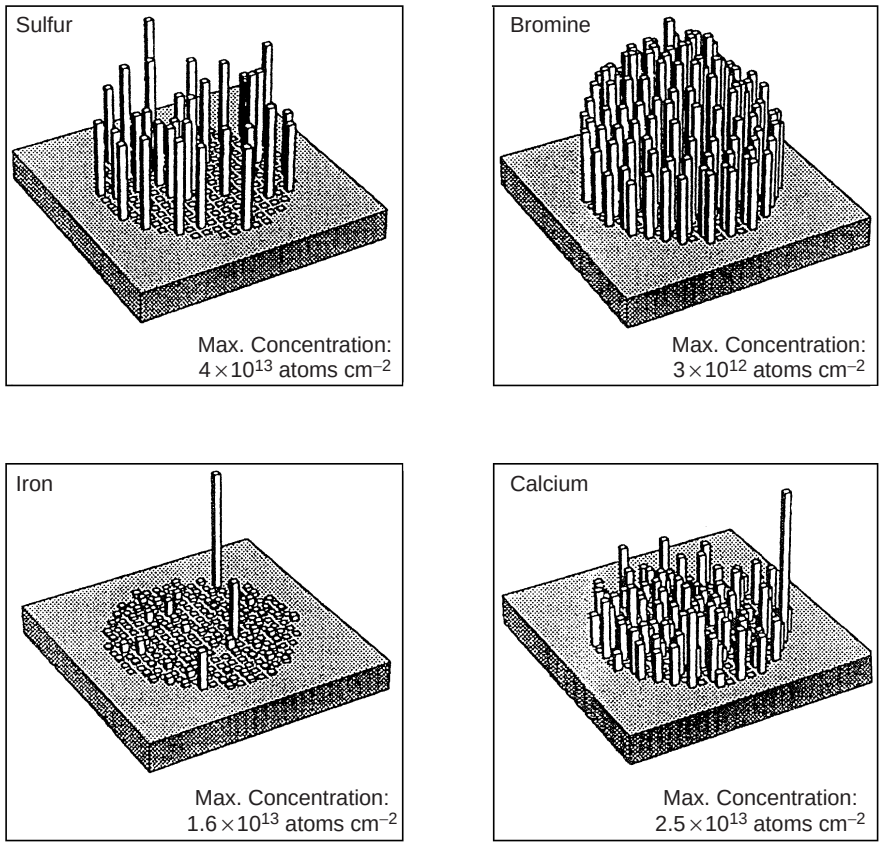

Figure 16 Results of a wafer mapping. (Reprinted from Berneike,[56] Copyright (1992) with permission from Elsevier Science.)

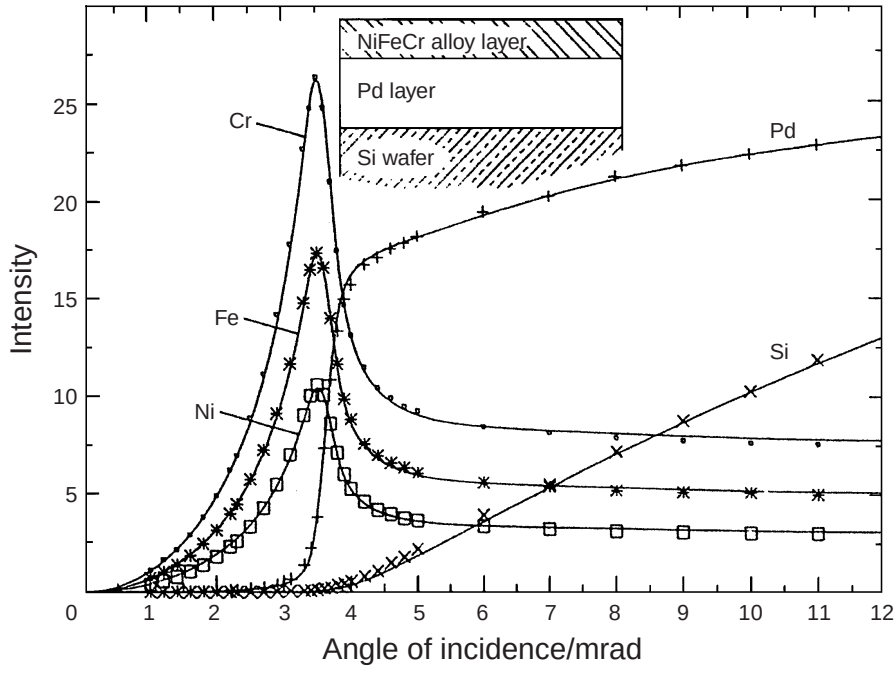

Figure 17 Fluorescence signal versus angle of incidence for a two-layer system on a Si wafer; the best fit was found for: Ni: 23.8%, Fe: 29.5%, Cr: 46.7%, $\rho = 7.0\,\mathrm{g\,cm^{-3}}$, $d = 5.9\,\mathrm{nm}$ for the alloy layer and $\rho = 11.2\,\mathrm{g\,cm^{-3}}$, $d = 257\,\mathrm{nm}$ for the Pd layer. (Reprinted from Weisbrod et al.,[59] Copyright (1991) with permission from Springer-Verlag.)

to be 5.9 nm (C(Cr) = 46.7%, C(Fe) = 29.5%, C(Ni) = 23.8%) with $\rho = 7\,\mathrm{g\,cm^{-3}}$, a thickness of 257 nm and a density of $11.2\,\mathrm{g\,cm^{-3}}$ were determined for the Pd layer.

10.4 Forensic and Art Historical Applications

For forensic applications the microanalytical capability of TXRF is the outstanding feature. In this field, microsamples are very common. Specific problems have been investigated, like hair samples, glass particles, tape fragments, drug powder, semen traces, gun-shot residues and textile fibers.[60,61]

Another application relating to the microanalytical capabilities of TXRF is the investigation of art objects. Favorable results include dating of objects and identification of fakes. Oil paintings have been investigated by analyzing the pigments, which are characterized by a limited number of major elements. Cotton buds can be used to remove an amount of about 1 μg from the paint before depositing it on a sample reflector.[62]

11 CHARACTERIZATION OF NEAR SURFACE LAYERS

An application of TXRF that is gaining importance is the nondestructive element analysis of near surface layers: layers on top of a reflecting surface as well as so-called "buried" layers below a reflecting surface. This is a consequence of the variation in the primary intensity with the angle of incidence above and below the surface. The variation results from interference caused by the superposition of incoming and reflected beams, as can be seen in Figure 18 from Klockenkämper, p. 55[26] for Si and Mo Kα radiation. A standing wave is formed above the surface. At a critical angle 0.1° (1 mrad), nodes and antinodes follow at an interval of about $d = 18\,\mathrm{nm}$ and the first antinode coincides with the surface. Assuming a reflectivity of 80%, the nodes have a 3.6-fold intensity of the primary beam. Below the surface the intensity is damped exponentially within a depth of some 10 nm (penetration depth). For angles smaller than the critical angle, the distance d is stretched and the first antinode moves away from the surface. Inside the substrate, the intensity is damped within a few nanometers. For angles larger than the critical angle, d is compressed, the oscillations vanish and the intensity approaches unity; also inside the substrate, the penetration depth increases to a few micrometers.

As the primary field varies with the angle of incidence, so the intensity of the fluorescence signal also shows this variation. From the shape obtained it is possible to distinguish between film-type samples, residue samples, thin layers and buried layers. The theory is discussed in detail by Knoth et al.[63–69] and De Boer et al.[70–73] as

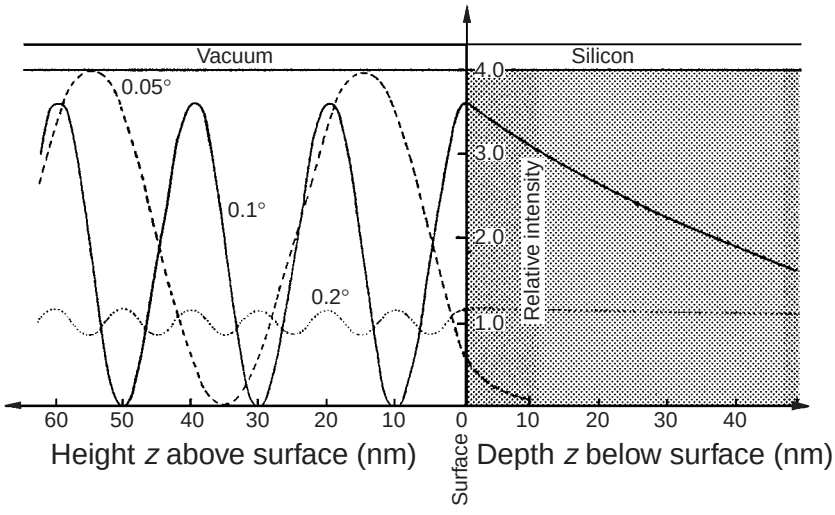

Figure 18 Primary intensity above and below a thick Si substrate, Mo Kα energy, various angles of incidence: 0.05° (- - - -), 0.1° = critical angle (———) and 0.2° (·····). (Reproduced by permission of John Wiley & Sons from Klockenkämper.)

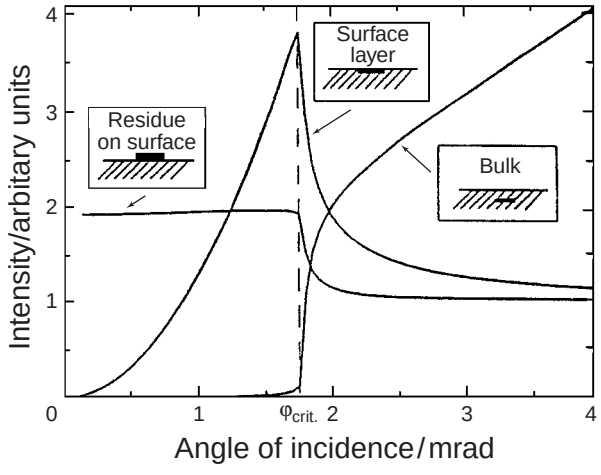

Figure 19 Characteristic shape of the angular dependence of the fluorescence signal versus angle of incidence for three cases of analyte locations: residue on the surface, surface layer (analyte near surface) and homogeneous bulk. (Reprinted from Weisbrod et al.,[64] Copyright (1991) with permission from Springer-Verlag.)

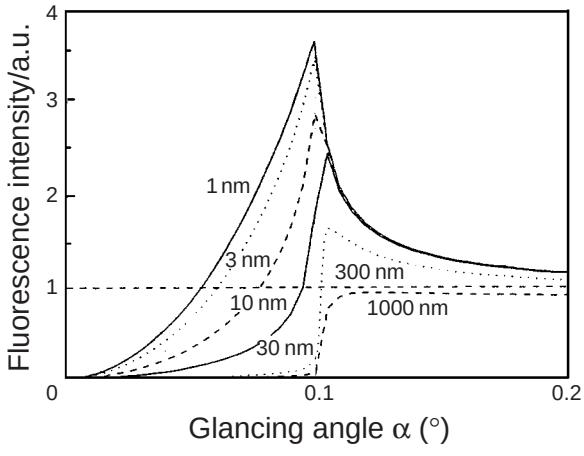

Figure 20 Angle dependent fluorescence signal from analyte atoms buried in a thick Si substrate. Mo Kα was the energy of the exciting radiation, the depth was varied but a constant area density was assumed. (Reproduced by permission of John Wiley & Sons from Klockenkämper.[26])

well as by Klockenkämper.[26] Figure 19 from Weisbrod et al.[64] shows the differences. In the case of residue on a surface (droplet pipetted on reflector, liquid removed), the signal is proportional to the factor $1 + R$, where R is the reflectivity. This is only valid if the grain size of the residue is in the range of 100 nm, because the oscillations vanish. Therefore the addition of a homogeneously distributed internal standard is necessary for calibration. If the reflectivity is assumed to be nearly unity, the fluorescence signal is doubled for angles less than the critical angle. There is angle dependence of the bulk signal for atoms exceeding a few 1000 nm depth; the signal is low for small angles and increases rapidly above the critical angle due to an increase in the penetration depth. For the near surface impurities the angle dependence leads to a strong peaking at the critical angle. Looking at an analyte below the surface, so-called ''buried'' layers, the fluorescence signal demonstrates a behavior like that shown in Figure 20, taken from Klockenkämper.[26] A rectangular profile of Co atoms in Si is assumed for depths varying from 1 to 1000 nm. Reflecting layers on a substrate induce some oscillations due to the standing waves. Figure 21, taken from Klockenkämper,[26] shows various Co layers on Si excited by Mo Kα radiation. The critical angle changes from that of Si to that of Co

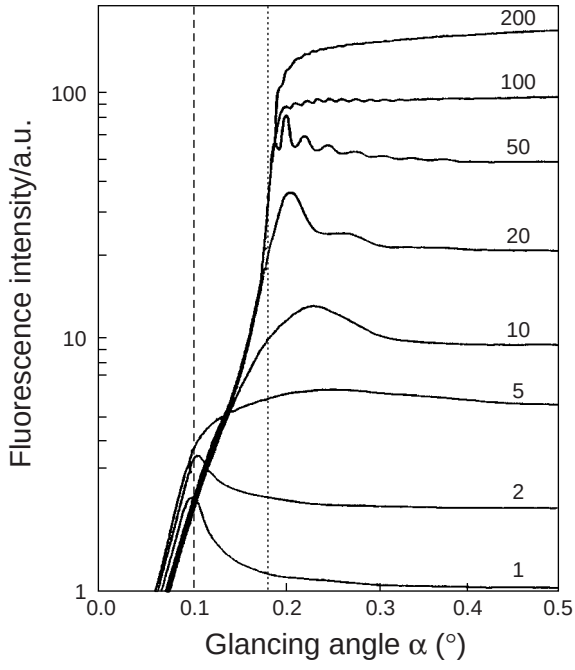

Figure 21 Angle dependent fluorescence signal emitted from Co-layers on a Si substrate with various layer thickness (in nm) and Mo Kα as exciting radiation. The maximum for the ultrathin layers is located at the critical angle of Si (dashed vertical line), and is shifted to the critical angle of Co (dotted vertical line) if the layer is more then 10 nm thick. (Reproduced by permission of John Wiley & Sons from Klockenkämper.[26])

with increasing Co thickness. Comparing the shape of the angle dependence signal for near surface layers below and above, no difference caused by the continuity condition is apparent, which is also seen in Figure 19: the primary field is nearly the same just above and just below the surface.

Thus it is possible to determine the location of analytes, the composition, the density and the layer thickness or the implantation depth. Multiple layered samples can also be analyzed.

12 SYNCHROTRON RADIATION EXCITED TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS

It can easily be seen from Equation (9) that there are different ways of improving the detection limits, namely increasing the sensitivity S, reducing the background and increasing the measuring time, which, however, is limited for practical reasons.

One way to reduce the spectral background is to use total reflection geometry. A further possibility for reducing the background is the use of polarized primary radiation.[74,75] A scheme is shown in Figure 22.

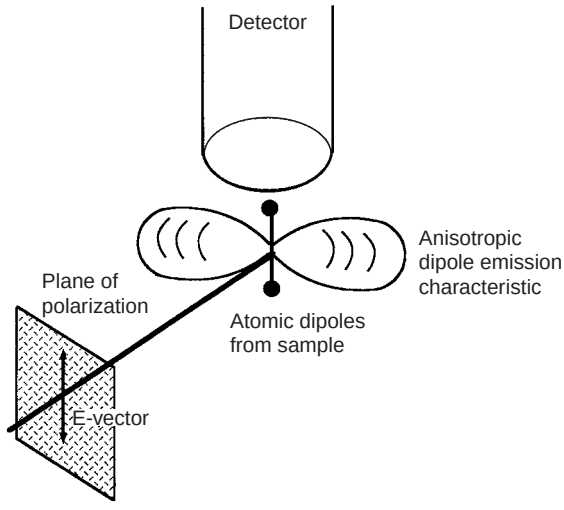

Figure 22 Anisotropy of the scattering effect for atomic dipoles and its application to XRF.

Owing to the anisotropic emission characteristics of the scattered radiation based on classical dipole radiation, it is advantageous to place a detector in such a position that only the isotropic emission of the fluorescence signal is detected. Thus the combination of TXRF with polarized radiation leads to a lower background. Moreover, the use of monochromatic primary radiation improves the background conditions because only photons with one energy can be scattered. An increase in sensitivity can be attained by using a tunable intense excitation source, enabling the exciting energy to be adjusted to just above the absorption edge of the element of interest.

SR with its outstanding properties offers new possibilities for improving the power of TXRF. The intense beam with a continuous spectral distribution from infrared to high-energy photons, as well as the linear polarization in the orbit plane and its natural collimation, are features best suited for excitation in total reflection geometry. Figure 23 shows the spectral brightness of Beamline L at HASYLAB, allowing excitation by photons between a few electronvolts and a few hundred kiloelectronvolts. For optimal excitation conditions the spectral distribution can be modified by elements like cut-off mirrors, monochromators and filters. Details can be found in Wobrauschek and Streli.[76]

ML monochromators are best suited for combination with SR excited XRF.[77,78] In comparison with crystal monochromators they offer a larger bandwidth ($\Delta E/E \approx 0.01$) which leads to a much larger photon flux on the sample. Another advantage is the possibility of selecting the excitation energy just below a matrix element with high concentration and just above the absorption edge of the element of interest (selective excitation), with the

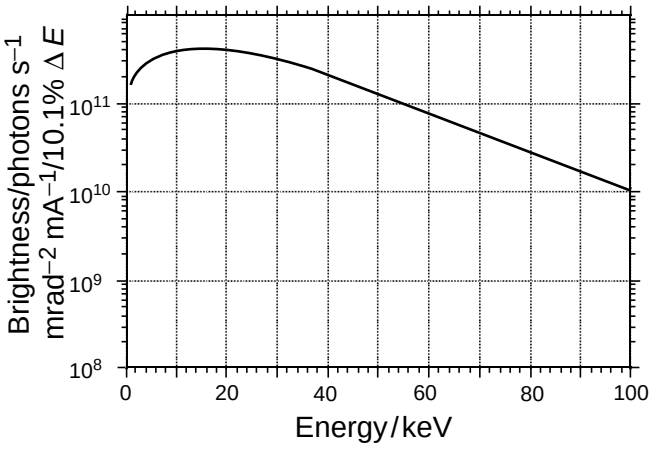

Figure 23 Calculated spectral brightness of Hamburg Synchrotronstrahlungslabor (HASYLAB), Beamline L.

drawback, however, of an increased background caused by Raman scattering.

The combination of TXRF with SR can be achieved with various geometrical arrangements for reflector and detector. Figure 24 shows three possibilities. For geometry in Figure 24(a) the polarization effect is fully utilized by positioning the detector axis in the plane of the orbit. Scattered radiation is not emitted in this direction. The sample is excited efficiently and there is full homogeneous illumination of the sample by the width of the beam in the horizontal plane. There are hardly any losses due to the collimators because the beam is naturally collimated in the vertical plane 0.1–0.2 mrad depending on the energy. The detection of the fluorescence signal is not optimal because the detector must be positioned sideways in order to make use of the polarization effect. The fluorescent radiation follows a long path in the sample before reaching the detector.

The excitation conditions for the arrangement as displayed in Figure 24(b) are poor. Most of the photons in the horizontal plane are absorbed in the collimation system. The intensity distribution in the vertical plane drops drastically and therefore the fluorescence intensity also drops with the deviation of sample regions from the plane of reference. Thus restriction to a 2–4-mm sample diameter is advisable because of the intensity and polarization distribution in the vertical plane. However, the detection efficiency is perfect because of the large solid angle. Excellent excitation and detection will be achieved with the arrangement shown in Figure 24(c). This combination of sample–detector position results in a complete loss of the polarization effect. If the sample is small, which is the case in ultratrace analysis, the scattering contribution from the sample itself is negligible. Scattering from the substrate is reduced by total reflection.

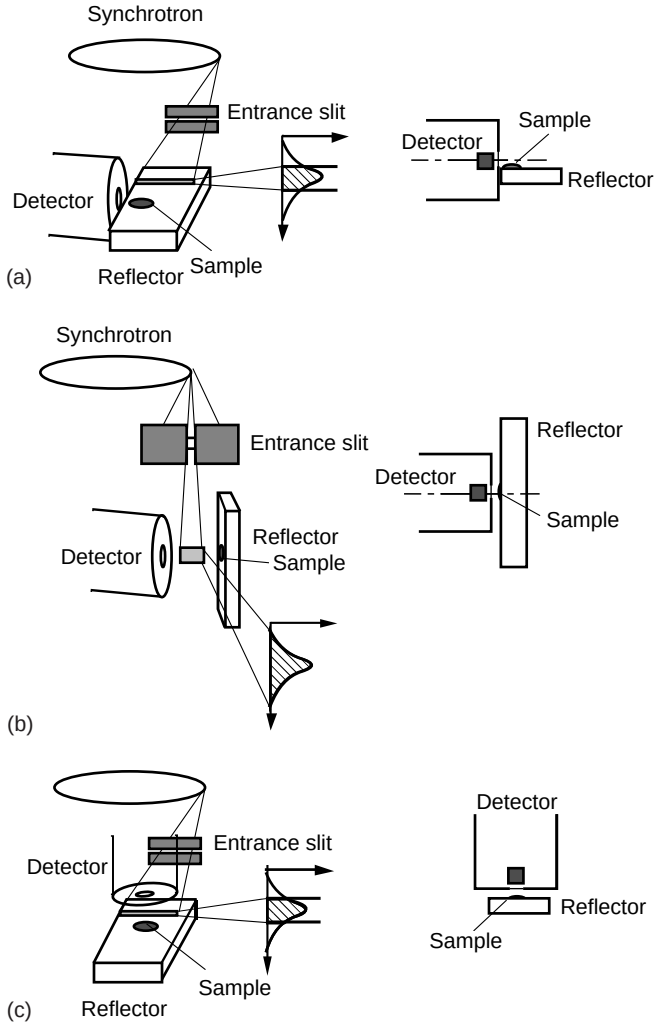

Figure 24 Various possible arrangements of sample reflector and detector in SR/TXRF.

If the experiments are performed in air, scattering of the exciting radiation contributes to the background. Therefore, all measurements have to be done in a vacuum chamber and give excellent results.

A group from Atominstitute, Vienna, have performed experiments at HASYLAB, DESY Hamburg, Beamline L, which is a beamline bending magnet. Details are described by Wobrauschek and Streli.[76] The arrangement can be seen in Figure 25 following geometry shown in Figure 24(b) with a vertical reflector and sideways detector. The beam is collimated by a primary slit system and then impinges on the ML monochromator, after the monochromator the primary beam is absorbed by a beamstopper. Shielding the vacuum chamber is extremely important because there are always multiple scattered high-energy photons which would penetrate into the measuring chamber through the Al wall. The beam is totally reflected on a rectangular Suprasil reflector which

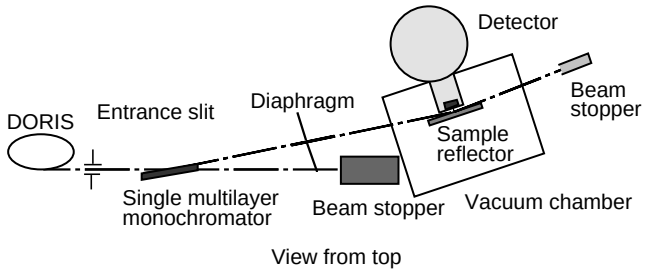

Figure 25 Experimental set-up for TXRF at HASYLAB.

has a Ta plate at the front edge to prevent scattering on the front edge. A typical multielement spectrum is shown in Figure 26, excited with 17.5 keV simulating a Mo tube. Figures 27 and 28 show the extension to higher Z elements, setting the monochromator at 31 keV and at 70 keV. Detection limits can be found in Table 3. Detailed information concerning the experiments is given by Wobrauschek et al.[79–81]

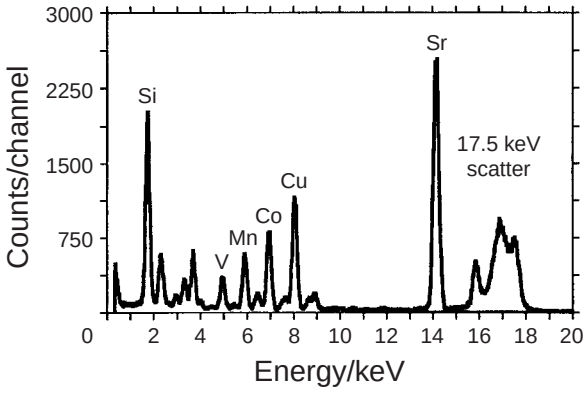

Figure 26 Spectrum of a multielement sample measured at HASYLAB, Beamline L, excited with 17.5 keV.

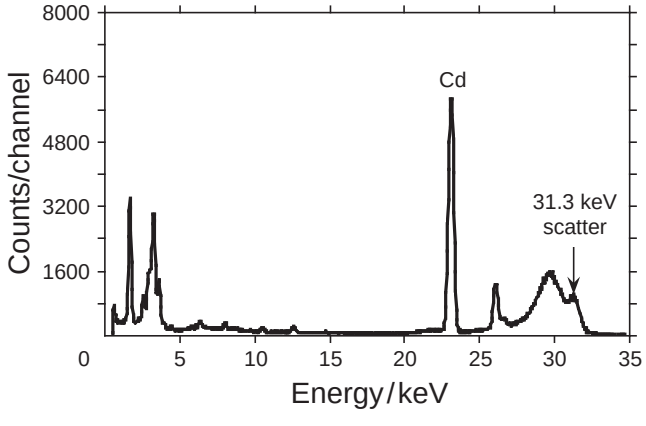

Figure 27 Spectrum of a Cd sample measured at HASYLAB, Beamline L, excited with 31.3 keV.

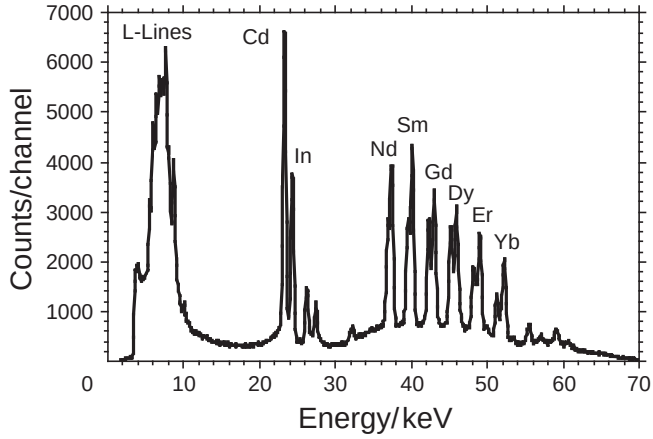

Figure 28 Spectrum of rare earth elements measured at HASYLAB, Beamline L, excited with 70 keV.

Table 3 Comparison of detection limits obtained with monochromatic excitation at HASYLAB Beamline L

E(keV)	I(mA)	Mn	Ni	Sr	Cd	Nd	Pr
10	103	15	12				
17	70			20			
31	70				150		
50	57					1500	600

Also some other groups made SR/TXRF experiments, in Japan[82–85] in USA (NSLS, National Synchrotron Light Source,[86] and SSRL, Stanford Synchrotron Radiation Laboratory),[87] and Italy (Frascati).[88]

As pointed out in the section on Applications, the semiconductor industry requires very low detection limits in the femtogram range or 10^8 atoms cm^{-2} level for the analysis of wafer surface contaminations at ultratrace levels.

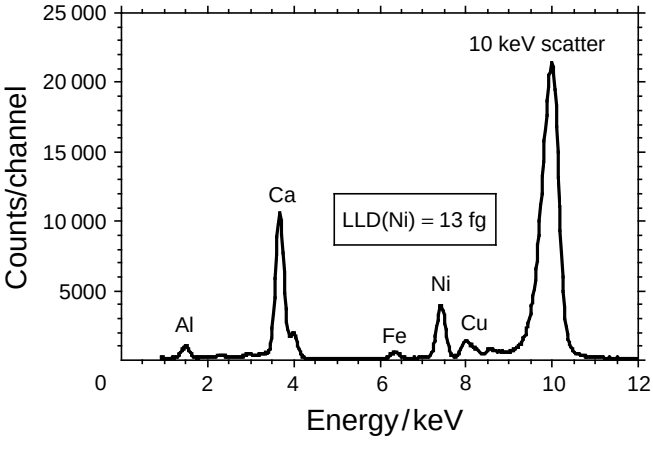

Figure 29 Spectrum of 10-pg Ni on a Si wafer measured at HASYLAB, Beamline L.

SR/TXRF is an interesting tool for this task. Experiments were performed at HASYLAB in cooperation with Wacker Siltronic, leading to detection limits of 13 fg or 1.3×10^8 atoms cm^{-2} for Ni, assuming an inspected area of 1 cm^2.[89,90] Figure 29 shows the relevant spectrum. The group at SSRL (Pianetta, Brennan, Fischer-Colbry, Ladderman[91]) established an instrument for routine wafer analysis at SSRL Beamline 6-2.[92–94] With its offer of a clean environment, automatic wafer handling (150 mm as well as 200-mm wafers), wafer mapping with detection limits of 3×10^8 atoms cm^{-2} for transition metals meets the requirements of the wafer industry perfectly. A focused wiggler beamline with a double ML monochromator is used. The wafer is held by a combined vacuum–electrostatic chuck in the vertical position, the detector is sideways and measurements are performed under vacuum conditions. At ESRF (European Synchrotron Radiation Facility) there is a project being realized on an Undulator beamline.[95] A Si(111) double monochromator will be used as well as a ML monochromator. The wafer will be held by a chuck in a horizontal position and two detector arrays consisting of seven elements in a sideways geometry will be established, measuring the fluorescence radiation. It is also planned to measure low-Z elements down to Na. Detection limits below 10^8 atoms cm^{-2} are proposed for 200 and 300-mm wafers.

13 TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS OF LOW-Z ELEMENTS

There is a lack of analytical methods for performing nondestructive and rapid multielement analyses of light elements at trace levels. TXRF can meet these requirements,[96,97] if a special spectrometer adapted to the specific problems of the ED detection of low-energy radiation is used. The detection limits achievable are mainly influenced by the type of excitation source, which should provide a large number of photons with an energy near the K-absorption edge of these elements (from 200 eV upwards). The determining factor is the integral over the source spectrum of the intensity multiplied by the photoelectric absorption coefficient ($I(E)\tau(E)\,dE$). Since the absorption coefficient $\tau(E)$ drops steadily as the energy E increases above an element's absorption edge, X-ray tubes with standard anode materials (e.g. Sc, Cr, Cu) are poor exciters for light elements as their characteristic emissions are far above the absorption edges. To improve the sensitivity for light elements it is necessary to use either an X-ray tube which emits intensive characteristic radiation as close as possible to the absorption edges of the interesting element, or SR with its continuous spectral range down to low energies.

The spectrometer used for the experiments described in the following section is characterized in detail by Streli et al.[98] A special ED Ge(HP) detector is used that meets all the requirements for low-energy detection. It has a thin entrance window (diamond window, 0.4-μm thick, transmission for oxygen Kα = 85%), an ion implanted contact layer, a thin deadlayer and low-electronic noise contributions (fwhm (5.9 keV) = 125 eV). Measurements are performed under vacuum conditions to reduce absorption. A Cr 1300 W tube is used under standard conditions and a W/C ML is used for monochromatization. Details are described by Streli et al.[99]

To improve the excitation conditions a windowless X-ray tube was used. Various anode materials have been tested after the take-off angle of the X-ray tube had been optimized to obtain a high photon flux on the sample. Optimization was effected with respect to the brilliance required and the self-attenuation of the emitted photons in the anode. Al, Si and Mo were tested as anode materials. Si offers the advantage of a high τ value for Al, but a low one for Si, which is mainly used as reflector material. For the analysis of impurities on Si-wafer surfaces this is especially an advantage, because the intensity of Si is reduced and cannot saturate the detector. Figure 30 shows the beam path of the spectrometer. Details are described in Streli et al.[100] Detection limits obtained differ greatly with the excitation conditions. The standard Cr tube provides about 90 pg for Mg, and with the windowless Si-anode tube 7 pg were obtained. Figures 31 and 32 show spectra.

Using SR as an excitation source offers the advantage of an intense naturally collimated X-ray beam with a wide spectral range, which is also in the low-energy region. Therefore, SR is the ideal source for exciting

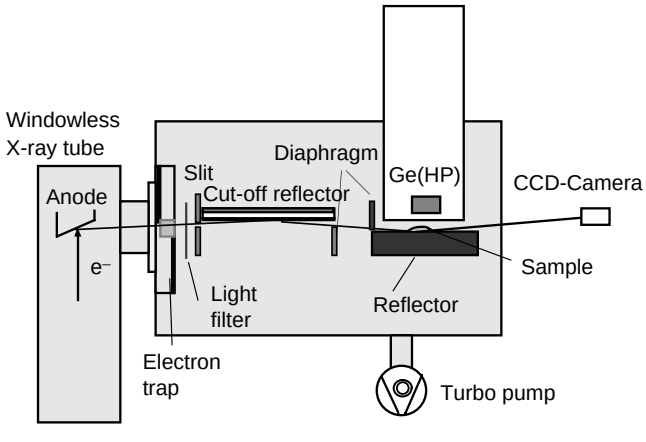

Figure 30 Scheme of the low Z-TXRF spectrometer equipped with a windowless Si-anode X-ray tube.

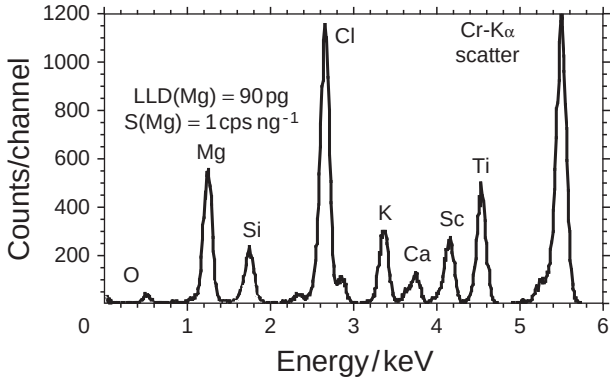

Figure 31 Spectrum measured with the low *Z*-TXRF spectrometer using a standard Cr fine focus tube.

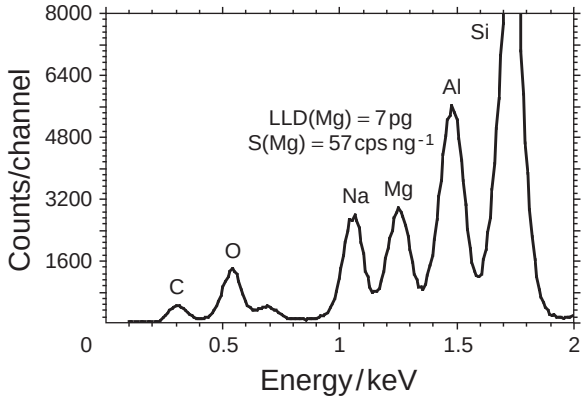

Figure 32 Spectrum measured with the low *Z*-TXRF spectrometer using a Si-anode windowless tube.

light elements efficiently. Experiments were performed at SSRL, at Beamline III-4. This beamline is equipped with a horizontally deflecting toroidal Au-mirror to cut off photons with an energy higher than 3 keV. A differential pumping system offers the possibility of connecting the measuring chamber directly without any window. Different kinds of filters can be inserted. A double ML monochromator can be inserted in the beam and removed easily. Figure 33 shows the schematic arrangement. The TXRF vacuum chamber was connected to the monochromator chamber directly with a flexible coupling and an interlock valve which closes the main valve upstream if the pressure exceeds 10^{-3} mbar. Because of the downlooking detector, the advantage of linear polarization of SR could not be used. Details are described by Streli et al.[101,102] To adapt the spectral distribution for analysis on Si wafer surfaces a 12-μm Si filter was inserted, leading to ''quasimonochromatic'' radiation with a bandwidth of about 400 eV, below the absorption edge of Si.

An interesting and promising application of low *Z*-TXRF is the quality control of Si wafer surfaces. Al

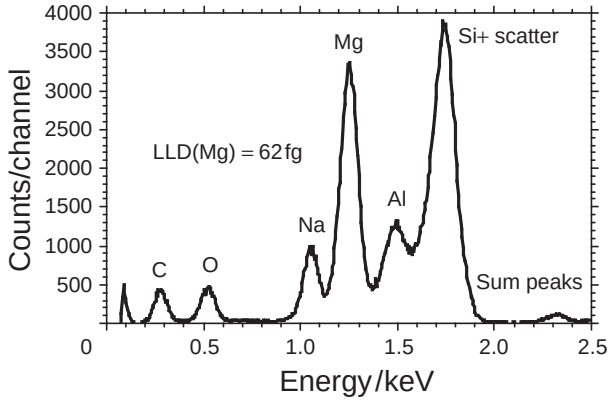

Figure 34 Spectrum of a Mg sample on a Si wafer measured at SSRL, Beamline 3–4.

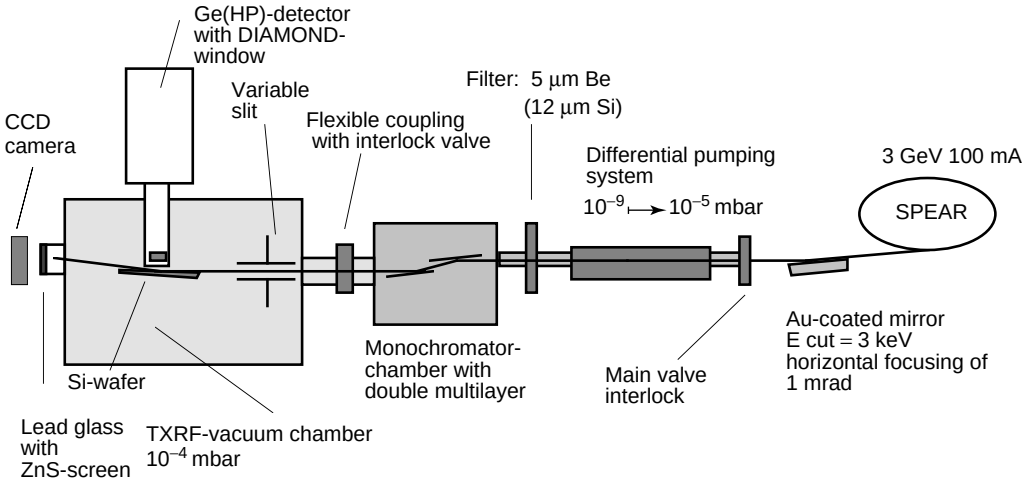

Figure 33 Scheme of the low *Z*-TXRF set-up at SSRL, Beamline 3–4.

and Na are elements of interest for the semiconductor industry, influencing the production yield of ICs negatively. Therefore it is very important to have an analytical method sensitive enough for the analysis of ultratraces of these elements. Droplet samples of 100 pg of Mg were prepared on Si wafer surfaces and analyzed with the set-up described using SR. The best results are found with ''quasimonochromatic'' SR, 62 fg for Mg (see Figure 34). Details are described by Purba et al.[49]

In analyzing low-Z elements with an ED detector, the problem of overlapping L and M-lines from higher Z elements is critical and has to be considered, as well as the fact that low-energy radiation is more easily absorbed and has to be controlled, if the thin film criterion is valid for the matrix under investigation.

14 CONCLUSIONS AND OUTLOOK

TXRF has become accepted worldwide and is the applied method for trace element analysis in various kinds of samples. Owing to the LLDs in the picogram range (for standard X-ray tube excitation) and the requirement for only small sample volume/mass and simultaneous multielement capability, the range of application is wide. At the moment the application for quality control for Si wafer surfaces at ultratrace levels is the most prevalent use, because of the nondestructiveness and spatial resolution of the method. The possibility of surface analysis due to the angle dependence of the fluorescence signal indicates promising applications like thin film analysis and nondestructive depth profiling. Using SR as an excitation source reduces the detection limits down to the femtogram level. The main application at present is in the field of semiconductor surface analysis.

Generally TXRF has become a relevant competitor with other analytical methods in the ultratrace range due to the manyfold interdisciplinary applications, where the analytical power of TXRF in the sense of detectable elemental range, detection limits, stability, accuracy and precision, almost background-free spectra and simple quantification can be usefully applied.

15 MANUFACTURERS OF TOTAL REFLECTION X-RAY FLUORESCENCE ANALYSIS SPECTROMETERS

Atomika Instruments (Germany), Bruckmannring 40, 85764 Oberschleissheim/Munich; chemical and wafers: web page: http://www.atomika.com, email: atomika @atomika.com

Atominstitut (Austria), Stadionallee 2, 1020 Vienna, chemical; email: wobi@ati.ac.at

Diffraction Technology (Australia), PO Box 444, Mitchwll Act 2911, Canberra, chemical; www.difftech.com.au, email: difftech@difftech.com.au

Italstructures (Italy), Via Monte Misone 11/d, 38066 Riva del Grada, chemical; email: isinfo@italstructures.com

IUT (Germany), Rudower Chaussee 5, D-12489 Berlin, chemical; http://www.iut-berlin.com, email: info@iut-berlin.com

Philips (TREX 630 T/S), web page: http://www.analytical.philips.com

Rigaku (Japan), 3-9-12 Matsubara-cho, Akishima-shi, Tokyo 196-8666, wafers; http://www.rigaku.com, email: rinttyo@rigaku.co.jp

Technos (Japan) 1-32-1 Nagao-Tanimachi, Hirakata-shi, Osaka 573-0164, represented by Philips, Lelyweg 1, 7602 EA Almelo, The Netherlands, wafers; http://www.philips.com/AXR

ACKNOWLEDGMENTS

The authors would like to thank P. Kregsamer for helpful contributions to the preparation of the manuscript and S. Zamini for preparing some figures.

ABBREVIATIONS AND ACRONYMS

ADC	Analog-to-digital Converter
APDC	Ammonium Pyrrolidinedithiocarbamate
CCD	Charge Coupled Device
DBDTC	Dibenzyldithiocarbamate
DPP	Digital Pulse Processor
ED	Energy-dispersive
EDXRF	Energy-dispersive X-ray Fluorescence
ESRF	European Synchrotron Radiation Facility
FET	Field Effect Transistor
fwhm	Full Width at Half-maximum
HASYLAB	Hamburg Synchrotronstrahlungslabor
IC	Integrated Circuit
IUPAC	International Union of Pure and Applied Chemistry
LLD	Lower Limits of Detection
LN	Liquid Nitrogen
MCA	Multichannel Analyzer
ML	Multilayer
NADDTC	Na-Diethyldithiocarbamate

NIST	National Institute of Standards and Technology
NSLS	National Synchrotron Light Source
NTRS	National Technology Roadmap for Semiconductors
SIA	Semiconductor Industry Association
SR	Synchrotron Radiation
SSRL	Stanford Synchrotron Radiation Laboratory
TXRF	Total Reflection X-ray Fluorescence
VPD	Vapor Phase Decomposition
WD	Wavelength-dispersive
WDXRF	Wavelength-dispersive X-ray Fluorescence
XRF	X-ray Fluorescence

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Surfaces **(Volume 10)**
Soft X-ray Photoelectron Spectroscopy in Analysis of Surfaces

X-ray Spectrometry **(Volume 15)**
X-ray Techniques: Overview • Energy Dispersive, X-ray Fluorescence Analysis • Sample Preparation for X-ray Fluorescence Analysis

REFERENCES

1. E. Bertin (ed.), *Principle and Practice of X-ray Spectrometric Analysis*, Plenum Press, New York, 1975.
2. R. VanGrieken, A.A. Markowicz (eds.), *Handbook of X-ray Spectrometry*, Marcel Dekker, 1993.
3. R. Jenkins, R.W. Gould, D. Gedcke (eds.), *Quantitative X-ray Spectrometry*, Marcel Dekker, 1981.
4. W.C. Röntgen, 'Über eine neue Art von Strahlung', *Sitzungsbericht der Würzburger Physikalisch medizinischen Gesellschaft* (1895).
5. C.G. Barkla, 'Polarisation in Secondary Röntgen Radiation', *Proc. R. Soc. London, Sec. A*, **77**, 247–255 (1906).
6. M. Laue, W. Friedrich, P. Knipping,'Interferenzererscheinungen bei Röntgenstrahlen', Sitzung. der Bayrischen Akademie der Wissenschaften, presented by A. Sommerfeld (June 1912).
7. W.L. Bragg, *Camb. Philos. Soc. Proc.*, **17**, 43–57 (Dec. 1912).
8. W. Stenström, 'Doctoral Thesis', University of Lund, Sweden, 1919.
9. A.H. Compton, *Phys. Rev.*, **20**(2), 84 (1922).
10. A. Larsson, M. Siegbahn, J. Waller, *Naturwissenschaften*, **52**, 1212 (1924).
11. E. Nähring, *Phys. Z.*, **31**, 401 (1930).
12. M. Montel, 'Utilisation de la Reflexion Totale des Rayons X pour la Suppression du Rayonnement de Courte Longueur Dónde dans un Faisceau Polychromatique', *C. R. Acad. Sci. Paris*, **7**, 2335–2337 (1956).
13. G. Damaschun, G. Kley, J.J. Müller, 'Über eine Röntgen–Kleinwinkelkamera nach Kratky mit Totalreflexionsmonochromator', *Acta Phys. Austr.*, **28**, 233–244 (1967).
14. H.K. Herglotz, 'Paraffin Mirrors for Ultrasoft X-rays', *Nature*, **214**, 263–265 (1967).
15. L.G. Parrat, 'Surface Studies of Solids by Total Reflection of X-rays', *Phys. Rev.*, **95**, 359–369 (1954).
16. B.L. Henke, P. Lee, T.J. Tanaka, R.L. Shimabukuro, B.K. Fujikawa, 'Low-energy X-ray Interaction Coefficients: Photoabsorption, Scattering, and Reflection', *At. Data Nucl. Data Tables*, **27**, 1–144 (1982).
17. Y. Yoneda, T. Horiuchi, 'Optical Flats for Use in X-ray Spectrochemical Microanalysis', *Rev. Sci. Instrum.*, **42**(7), 1069–1670 (1971).
18. H. Aiginger, P. Wobrauschek, 'A Method for Quantitative XRF Analysis in the Nanogram Region', *Nucl. Instrum. Methods*, **114**, 157–158 (1974).
19. P. Wobrauschek, H. Aiginger, 'Total-reflection X-ray Fluorescence Spectrometric Determination of Elements in Nanogram Amounts', *Anal. Chem.*, **47**(6), 852–854 (1975).
20. J. Knoth, H. Schwenke, R. Marten, J. Glauer, 'Determination of Cu and Fe in Human Blood Serum by EDXRF', *J. Clin. Chem. Clin. Biochem.*, **15**, 537–560 (1977).
21. Proceedings of the TXRF Conference 1988 in Dortmund, *Spectrochim. Acta, Part B*, **44**, 1989.
22. Proceedings of the TXRF Conference 1990 in Vienna, *Spectrochim. Acta, Part B*, **46**, 1991.
23. Proceedings of the TXRF Conference 1992 in Gheesthacht, *Spectrochim. Acta, Part B*, **48**, 1993.
24. Proceedings of the TXRF Conference 1994 in Tsukuba, *Adv. X-ray Anal., Jpn*, **26s**, 1995.
25. Proceedings of the TXRF Conference 1996 in Dortmund and Eindhoven, *Spectrochim. Acta, Part B*, **52**, 1997.
26. R. Klockenkämper, *Total Reflection X-ray Fluorescence Analysis*, John Wiley & Sons, 1997.
27. H. Schwenke, J. Knoth, 'A Highly Sensitive Energy-dispersive X-ray Spectrometer with Multiple Total

Reflection of the Exciting Beam', *Nucl. Instrum. Methods*, **193**, 239–243 (1982).
28. T.W. Barbee, N.S.F. 26 Ann. Rep., US Gov. 038-000-00313-5, 019, 1976.
29. R. Day, J. Grosso, R. Bartlett, 'Layered Synthetic Microstructures: Measurements and Applications', *Nucl. Instrum. Methods*, **208**, 245–249 (1983).
30. J. Dumond, J.P. Youtz, *J. Appl. Phys.*, **11** (1940).
31. Moxtek Inc., 452 West 1260, North Oren 84057; Ketek GmbH, Am Isarbach 30, D-85764 Oberschleissheim; Amtek, 6 De Angelo Drive, Bedford, MA 10730-2204, USA.
32. S.E. Labov, C.A. Mears, M. Frank, L.J. Hiller, H. Netel, M.A. Lindeman, 'Assessment of Low Temperature X-ray Detectors', *Nucl. Instrum. Methods, Section A*, **370**, 65–68 (1996).
33. J. Hohne, et al. Oral presentation at the TXRF98 Conference in Austin, Texas, October, 1998.
34. C. Neumann, P. Eichinger, 'Ultra-trace Analysis of Metallic Contaminations on Silicon Wafer Surfaces by Vapor Phase Decomposition/Total Reflection X-ray Fluorescence Analysis (VPD/TXRF)', *Spectrochim. Acta, Part B*, **46**, 1369–1377 (1991).
35. A. Prange, H. Schwenke, 'Sample Treatments for TXRF–Requirements and Prospects', *Adv. X-Ray Anal.*, **32**, 211–220 (1989).
36. B. Holynska, B. Ostachowicz, D. Wegrzynek, 'Simple Method of Determination of Copper, Mercury and Lead in Potable Water with Preliminary Pre-concentration by TXRF', *Spectrochim. Acta, Part B*, **51**, 769–773 (1996).
37. M. Dargie, A. Markowicz, A. Tajani, V. Valkovic, 'Optimized Sample Preparation Procedure for the Analysis of Solid Materials by Total Reflection XRF', *Fresenius' Z. Anal. Chem.*, **357**, 589–593 (1997).
38. B. Holynska, M. Olko, B. Ostachowicz, J. Ostachowicz, D. Wegrzynek, M. Claes, R. Van Grieken, P. de Bokx, P. Kump, M. Necemer, 'Performance of Total Reflection and Grazing Emission X-ray Fluorescence Spectrometry for the Determination of Trace Metals in Drinking Water in Relation to Other Analytical Techniques', *Fresenius' Z. Anal. Chem.*, **362**(3), 294–298 (1998).
39. M.A. Barreiros, M.L. Carvalho, M.M. Costa, M.I. Marques, M.T. Ramos, 'Application of TXRF to Elemental Studies of Drinking Water', *X-Ray Spectrom.*, **26**, 165–168 (1997).
40. A. Prange, A. Knöchel, W. Michaelis, 'Multielement Determination of Dissolved Heavy Metal Traces in Sea Water by Total Reflection X-ray Fluorescence Analysis', *Anal. Chim. Acta*, **172**, 79–100 (1985).
41. A. Prange, H. Böddeker, K. Kramer, 'Determination of Trace Elements in River-water Using Total-reflection X-ray Fluorescence', *Spectrochim. Acta, Part B*, **48**, 207–215 (1993).
42. Ch. Koopmann, A. Prange, 'Multielement Determination in Sediments from the German Wadden Sea–Investigations on Sample Preparation Techniques', *Spectrochim. Acta, Part B*, **46**, 1395–1402 (1991).
43. R. Klockenkämper, H. Bayer, A. VonBohlen, 'Total Reflection X-ray Fluorescence Analysis of Airborne Particulate Matter', *Adv. X-Ray Chem. Anal.*, **26s**, 41–46 (1995).
44. J. Injuk, R. Van Grieken, 'Optimisation of Total-reflection X-ray Fluorescence for Aerosol Analysis', *Spectrochim. Acta, Part B*, **50**, 1787–1803 (1995).
45. U. Reus, 'Determination of Trace Elements in Oils and Greases with Total Reflection X-ray Fluorescence: Sample Preparation Methods', *Spectrochim. Acta, Part B*, **46**, 1403–1411 (1991).
46. A. Prange, H. Böddeker, W. Michaelis, 'Multi-element Determination of Trace Elements in Whole Blood Serum with TXRF', *Fresenius' Z. Anal. Chem.*, **335**, 914–918 (1989).
47. A. VonBohlen, R. Klockenkämper, G. Tölg, B. Wiecken, 'Microtome Sections of Biomaterials for Trace Analyses by TXRF', *Fresenius' Z. Anal. Chem.*, **331**, 454–458 (1988).
48. A. Prange, K. Kramer, U. Reus, 'Determination of Trace Element Impurities in Ultrapure Reagents by Total Reflection X-ray Spectrometry', *Spectrochim. Acta, Part B*, **46**, 1385–1393 (1991).
49. P. Burba, P. Willmer, M. Becker, R. Klockenkämper, 'Determination of Trace Elements in High-purity Aluminium by Total Reflection X-ray Fluorescence After their Separation on Cellulose Loaded with Hexamethylenedithiocarbamates', *Spectrochim. Acta, Part B*, **44**(5), 525–530 (1989).
50. U. Reus, 'Total Reflection X-ray Fluorescence Spectrometry: Matrix Removal Procedures for Trace Analysis of High-purity Silicon, Quartz and Sulphuric Acid', *Spectrochim. Acta, Part B*, **44**(5), 533–537 (1989).
51. N. Ojeda, E.D. Greaves, J. Alvarado, L. Sajo-Bohus, 'Determination of V, Fe, Ni and S in Petroleum Crude Oil by Total-reflection X-ray Fluorescence', *Spectrochim. Acta, Part B*, **48**, 247–253 (1993).
52. D.B. Bilbrey, D.J. Leland, D.E. Leyden, P. Wobrauschek, H. Aiginger, 'Analysis of Aerosols Using Total Reflection X-ray Spectrometry', *X-Ray Spectrom.*, **16**, 161–165 (1987).
53. K. Freitag, U. Reus, J. Fleischhauer, 'Extension of the Analytical Range of Total Reflection X-ray Fluorescence Spectrometry to Lighter Elements ($11 = < Z < 16$) and Increase in Sensitivity by Excitation with Tungsten LA Radiation', *Spectrochim. Acta, Part B*, **44**(5), 499–503 (1989).
54. C.T. Yap, R.E. Ayala, P. Wobrauschek, 'Quantitative Trace Element Determination in Thin Samples by Total Reflection X-ray Fluorescence Using the Scattered Radiation Method', *X-Ray Spectrom.*, **17**, 171–175 (1988).

55. *National Technology Roadmap for Semiconductors: Technology Needs* is the Semiconductor Industry Association (SIA) 1997 edition report published by SEMATECH, Inc. http://notes.sematech.org/1997pub.htm.
56. W. Berneike, 'Basic Features of Total-reflection X-ray Fluorescence Analysis on Si-wafers', *Spectrochim. Acta, Part B*, **48**, 269–275 (1993).
57. L. Fabry, S. Pahlke, L. Kotz, G. Tölg, 'Trace-analytical Methods for Monitoring Contaminations in Semiconductor-grade Si Manufacturing', *Fresenius' Z. Anal. Chem.*, **349**, 260–265 (1994).
58. A. Prange, H. Schwenke, 'Trace Element Analysis Using Total-reflection X-ray Fluorescence Spectroscopy', *Adv. X-Ray Anal.*, **35**, 899–923 (1992).
59. U. Weisbrod, R. Gutschke, J. Knoth, H. Schwenke, 'Total Reflection X-ray Fluorescence Spectrometry for Quantitative Surface and Layer Analysis', *Appl. Phys. A*, **53**, 449–456 (1991).
60. T. Ninomiya, S. Nomura, K. Taniguchi, S. Ikeda, 'Application of GIXE to Forensic Samples', *Adv. X-Ray Chem. Anal.*, **26s**, 9–18 (1995).
61. A. Prange, U. Reus, H. Böddekker, R. Fischer, F.P. Adolf, S. Ikeda, 'Microanalysis in Forensic Science: Characterization of Singel Textile Fibers with TXRF', *Adv. X-Ray Chem. Anal.*, **26s**, 1–8 (1995).
62. R. Klockenkämper, A. VonBohlen, L. Moens, W. Devos, 'Analytical Characterization of Artists' Pigments Used in Old and Modern Paintings by Total-reflection X-ray Fluorescence', *Spectrochim. Acta, Part B*, **48**, 239–246 (1993).
63. H. Schwenke, J. Knoth, 'Total Reflection X-ray Fluorescence Analysis', *Handbook on X-Ray Spectrometry*, eds. R. Van Grieken, A. Markowicz, Marcel Dekker, New York, 453–488, 1993.
64. U. Weisbrod, R. Gutschke, J. Knoth, H. Schwenke, 'X-ray Induced Fluorescence Spectrometry at Grazing Incidence for Quantitative Surface and Layer Analysis', *Fresenius' Z. Anal. Chem.*, **341**, 83–86 (1991).
65. U. Weisbrod, R. Gutschke, J. Knoth, H. Schwenke, 'Total Reflection X-ray Fluorescence Spectrometry for Quantitative Surface and Layer Analysis', *Appl. Phys. A*, **53**, 449–456 (1991).
66. J. Knoth, H. Schwenke, U. Weisbrod, 'Total Reflection X-ray Fluorescence Spectrometry for Surface Analysis', *Spectrochim. Acta, Part B*, **44**(5), 477–482 (1989).
67. H. Schwenke, R. Gutschke, J. Knoth, 'Characterization of Near Surface Layers by Means of Total Reflection X-ray Fluorescence Spectrometry', *Adv. X-Ray Anal.*, **35**, 941–946 (1992).
68. J. Knoth, R. Bormann, R. Gutschke, C. Michaelsen, H. Schwenke, 'Examination of Layered Structures by Total-reflection X-ray Fluorescence Analysis', *Spectrochim. Acta, Part B*, **48**, 285–290 (1993).
69. H. Schwenke, R. Bormann, J. Knoth, A. Prange, 'Some Potential Developments for Trace Element and Surface Analysis Using a Grazing Incident X-ray Beam', *Spectrochim. Acta, Part B*, **48**, 293–297 (1993).
70. D.K.G. De Boer, W.W. Van Den Hoogenhof, 'TXRF of Thin Layers on and in Solids', *Adv. X-Ray Anal.*, **34**, 35–40 (1991).
71. D.K.G. De Boer, W.W. Van Den Hoogenhof, 'TXRF of Single and Multiple Thin-layer Samples', *Spectrochim. Acta, Part B*, **46**, 1323–1328 (1991).
72. D.K.G. DeBoer, A.J.G. Leenaers, W.W. Van Den Hoogenhof, 'The Profile of Layered Materials Reflected by Glancing-incidence X-ray Analysis', *Appl. Phys A*, **58**, 169–173 (1994).
73. D.K.G. DeBoer, A.J.G. Leenaers, W.W. Van Den Hoogenhof, 'Glancing-incidence X-ray Analysis of Thin-layered Materials: a Review', *X-Ray Spectrom.*, **24**, 91–95 (1995).
74. R.W. Ryon, J.D. Zahrt, 'Polarized Beam X-ray Fluorescence', *Handbook of X-ray Spectrometry*, Marcel Dekker, New York, 491, 1993.
75. H. Aiginger, P. Wobrauschek, C. Brauner, 'Energy-dispersive Fluorescence Analysis Using Bragg-reflected Polarized X-rays', *Nucl. Instrum. Methods*, **120**, 541–546 (1974).
76. P. Wobrauschek, C. Streli, 'TXRF with Synchrotron Radiation and Other Sources for Trace Element Determination', *X-ray and Inner-shell Processes*, eds. R.L. Johnson, H. Schmidt-Böacking, B. Sonntag, AIP Conference Proceedings, **389**, 233–247, 1997.
77. D.H. Bilderback, B.M. Lairson, T.W. Barbee, Jr, G.E. Ice, C.J. Sparks, Jr, 'Design of Doubly Focusing, Tunable (5–30 keV), Wide Bandpass Optics Made from Layered Synthetic Microstructures', *Nucl. Instrum. Methods*, **208**, 251–256 (1983).
78. P. Pianetta, T. Barbee, 'Applications of Multilayers to Synchrotron Radiation', *Nucl. Instrum. Methods, Sect. A*, **266**, 441–445 (1988).
79. R. Rieder, P. Wobrauschek, W. Ladisich, C. Streli, H. Aiginger, S. Garbe, G. Gaul, A. Knöchel, F. Lechtenberg, 'Total-reflection X-ray Fluorescence Analysis with Synchrotron Radiation Monochromatized by Multilayer Structures', *Nucl. Instrum. Methods, Sect. A*, **355**(2,3), 648–653 (1995).
80. P. Wobrauschek, P. Kregsamer, W. Ladisich, R. Rieder, C. Streli, 'TXRF–Sources–Samples and Detectors', *Adv. X-Ray Anal*, **39**, 755–765 (1997).
81. R. Görgl, P. Wobrauschek, P. Kregsamer, C. Streli, M. Haller, A. Knöchel, M. Radtke, 'Total Reflection X-ray Fluorescence Analysis (TXRF) Excited by Synchrotron Radiation (SR–TXRF)–Variation of Excitation Conditions and Sample Geometries', *X-Ray Spectrom.*, **26**(4), 189–193 (1997).
82. A. Iida, A. Yoshinaga, K. Sakurai, Y. Gohshi, 'Synchrotron Radiation Excited XRF Analysis Using Total Reflection of X-rays', *Anal. Chem.*, **58**, 394–400 (1986).

83. A. Iida, 'Semiconductor Surface Characterization by Synchrotron XRF', *Adv. X-Ray Anal.*, **34**, 23–30 (1991).
84. A. Iida, 'Grazing Incidence X-ray Fluorescence Analysis Using Synchrotron Radiation', *Adv. X-Ray Anal.*, **35**, 795–802 (1992).
85. N. Kondo, N. Ota, I. Nakai, A. Sasaki, 'Synchrotron Radiation-induced TXRF of Cu and Zn in Carcinoma Tissues Obtained by a Biopsy', *Bunseki Kagaku*, **46**(12), 957–963 (1997).
86. P.A. Pella, R.C. Dobbyn, 'Total Reflection Energy-dispersive XRF Spectrometry Using Monochromatic SR: Application to Se in Blood Serum', *Anal. Chem.*, **60**, 684–690 (1988).
87. S. Brennan, W. Tompkins, N. Takaura, P. Pianetta, S. Laderman, A. Fischer-Colbrie, J.B. Kortright, M. Madden, D. Wherry, 'Wide Band-pass Approaches to Total-reflection X-ray Fluorescence Using Synchrotron Radiation', *Nucl. Instrum. Methods, Sect. A*, **347**, 417–423 (1994).
88. H. Sanchez, E. Burattini, M. Rubio, A. Riveros, 'XRF Analysis of Volcanic Ash Using Synchrotron Radiation', *Nucl. Instrum. Methods, Sect. B*, **84**, 408–412 (1994).
89. P. Wobrauschek, P. Kregsamer, W. Ladisich, C. Streli, S. Pahlke, L. Fabry, S. Garbe, M. Haller, A. Knöchel, M. Radtke, 'TXRF with Synchrotron Radiation – Analysis of Ni on Si-wafer Surfaces', *Nucl. Instrum. Methods, Sect. A*, **363**, 619–621 (1995).
90. P. Wobrauschek, R. Görgl, P. Kregsamer, C. Streli, S. Pahlke, L. Fabry, M. Haller, A. Knöchel, M. Radtke, 'Analysis of Ni on Si-wafer Surfaces Using Synchrotron Radiation Excited TXRF', *Spectrochim. Acta, Part B*, **52**(7), 901–908 (1997).
91. S. Laderman, A. Fishcer-Cobrie, A. Waldauer, N. Takaura, S. Ghosh, S. Brennan, P. Pianetta, E. Adem, 'TXRF R&D Studies', SSRL Annual Report 1995 P200M, 1995.
92. S. Brennan, W. Tompkins, N. Takaura, P. Pianetta, S. Laderman, A. Fischer-Colbrie, J.B. Kortright, M. Madden, D. Wherry, 'Wide Band-pass Approaches to Total-reflection X-ray Fluorescence Using Synchrotron Radiation', *Nucl. Instrum. Methods, Sect. A*, **347**, 417–421 (1994).
93. P. Pianetta, N.Takaura, S. Brennan, W. Tompkins, S. Laderman, A. Fischer-Colbrie, A. Shimazaki, K. Miyazaki, M. Madden, D. Wherry, J.B. Kortright, 'Total Reflection X-ray Fluorescence Spectroscopy Using Synchrotron Radiation for Wafer Surface Trace Impurity Analysis', *Rev. Sci. Instrum.*, **66**(2), 1293–1298 (1995).
94. A. Fischer-Colbrie, S. Laderman, S. Brennan, S. Ghosh, N. Takaura, P. Pianetta, A. Shimazaki, A. Waldauer, D. Wherry, A. Barkan, 'State-of-the-art Evaluation of Ultra-clean ULSI Processes', *Mater. Res. Soc. Symp. Proc.*, **477**, 403–408 (1997).
95. F. Comin, M. Navizet, P. Mangiagalli, G. Apostolo, 'ID27: an Industrial TXRF Facility', *ESRF Newsl.*, September, 28–29 (1998).
96. C. Streli, P. Wobrauschek, H. Aiginger, 'A New X-ray Tube for Efficient Excitation of Low-*Z*-elements with Total Reflection X-ray Fluorescence Analysis', *Spectrochim. Acta, Part B*, **46**, 1351–1358 (1991).
97. C. Streli, H. Aiginger, P. Wobrauschek, 'Light Element Analysis with a New Spectrometer for Total-reflection X-ray Fluorescence', *Spectrochim. Acta, Part B*, **48**, 163–168 (1993).
98. C. Streli, P. Wobrauschek, E. Unfried, H. Aiginger, 'A New Spectrometer for Total Reflection X-ray Fluorescence Analysis of Light Elements', *Nucl. Instrum. Methods, Sect. A*, **334**, 425–430 (1993).
99. C. Streli, P. Wobrauschek, W. Ladisich, R. Rieder, H. Aiginger, 'Total Reflection X-ray Fluorescence Analysis of Light Elements Under Various Excitation Conditions', *X-Ray Spectrom.*, **24**, 137–142 (1995).
100. C. Streli, V. Bauer, P. Wobrauschek, 'Recent Developments in TXRF of Light Elements', *Adv. X-Ray Anal.*, **39**, 771–779 (1997).
101. C. Streli, P. Wobrauschek, W. Ladisich, R. Rieder, H. Aiginger, R. Ryon, P. Pianetta, 'Total Reflection X-ray Fluorescence Analysis of Light Elements Using Synchrotron Radiation', *Nucl. Instrum. Methods, Sect. A*, **345**, 399–405 (1994).
102. C. Streli, P. Wobrauschek, V. Bauer, P. Kregsamer, R. Görgl, P. Pianetta, R. Ryon, S. Pahlke, L. Fabry, 'Total Reflection X-ray Fluorescence Analysis of Light Elements with Synchrotron Radiation and Special X-ray Tubes', *Spectrochim. Acta, Part B*, **52**(7), 861–867 (1997).

Ultrafast Diffraction Techniques

A. Rousse
LOA, ENSTA, CNRS, Ecole Polytechnique, Palaiseau, France

1 Introduction 1
2 Ultrafast Time Resolution 2
2.1 High-speed Detectors 2
2.2 Femtosecond Sources 3
3 Ultrafast X-ray Diffraction 3
3.1 Femtosecond X-ray Sources 3
3.2 Operating Principles 4
3.3 Experiments 4
4 Ultrafast Electron Diffraction 6
4.1 Electron Femtosecond Sources 6
4.2 Operating Principles 6
4.3 Experiments 6
5 Conclusions 7
Acknowledgments 7
Abbreviations and Acronyms 7
Related Articles 7
References 8

The need for femtosecond time resolution in the analysis of processes governed by atomic motion is well established. In biology, chemistry, and physics different times are of interest, but the shortest relate to the vibration of matter, being typically about 10^{-13} s (100 femtoseconds). Time-resolved experiments are used to monitor time-dependent processes. Depending on the spectral range of observation, different properties can be studied. Usually, visible or near-visible radiation provides information on the electronic properties, but atomic motion can be directly monitored only through very short wavelength radiation, such as X-rays. The electronic changes associated with vibrations in single molecules, liquids or crystal lattices, isomerization and the breaking or formation of chemical bonds, can be monitored by ultrafast optical spectroscopy techniques, but the accompanying ultrafast structural rearrangements cannot be directly observed. In the detailed characterization of structures at the atomic level, X-ray and electron diffraction is an unparalleled tool, allowing reaction intermediates to be followed.

Synchrotrons have become essential to the study of structures. Nevertheless, the shortest durations attainable with the latest generation of synchrotrons, hold the time resolution of experiments around 50 ps. This article shows how subpicosecond time resolution can be achieved in the analysis of very short-lived transient structures. The two emerging techniques in subpicosecond time-resolved X-ray (electron) diffraction are (1) ultrashort X-ray (electron) probe pulses, produced through a femtosecond laser system, and (2) fast X-ray detection during quasicontinuous synchrotron-produced X-ray probe illumination.

1 INTRODUCTION

Atoms share their outer-shell electrons to establish the bonds that build the structures of solids, molecules, and macromolecules. Strict energetic laws govern the movement of electrons, atoms, and molecules. At room temperature, atoms and molecules vibrate at a few hundred femtoseconds. When excited, these atoms may redistribute themselves inside their environment. Molecules may twist, to establish and destroy bonds. These basic ultrafast atomic displacements are at the origin of many processes, even though the expected event is reached sometimes a few seconds later, following transitions through multiple intermediate states. The ideal would be to follow a reaction – at the atomic level – en route from reactant to product. However, bond dissociation and atomic displacement following charge transfer or isomerization have been shown to occur in the femtosecond timescale.[1–3] The subpicosecond transient structures involved are critical to the understanding of the first steps of elementary processes such as vision, oxygen transport in hemoglobin, or phase transitions such as the melting of a solid.

The scattering properties of fully ordered (crystaline) or poorly ordered (fluid) matter on electrons and X-rays can directly reveal its structure through atomic spatial resolution.[4,5] KeV X-radiation ($\lambda \approx 1$ Å) scatters mainly with the core electrons. The interferences that occur between electromagnetic waves produced from all the scattering centers – periodically positioned within the crystal – lead to a diffraction pattern which is very sensitive to the atomic structure of the material. In comparison to X-rays, electrons interact directly with the nuclei of the atoms by elastic scattering. Being several orders of magnitude stronger than X-ray scattering, this process is efficiently used in gas-phase structure determination, instead of X-ray diffraction. Since the mid-1990s significant advances have occurred towards picosecond and subpicosecond time-resolution in diffraction experiments.[6] Using 100 fs time-resolved X-ray diffraction, the picosecond transient behavior of atomic disordering in a Langmuir–Blodgett (LB) organometallic film has been resolved, following ultrashort laser

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

pulse excitation.[7] At UCSD (University of California, San Diego), picosecond time resolution with very high spatial resolution (mÅ) has been demonstrated in the ultrafast nonthermal melting of GaAs crystal.[8,9] The laser energy transport (disordering, optical phonon, acoustic phonon) in an InSb crystal was analyzed with a resolution of 300 fs by X-ray diffraction.[10] In less than 20 years, the time resolution of X-ray diffraction experiments has increased by five orders of magnitude, owing to the maturation of laser-produced plasma X-ray sources (Figure 1). At present only the few most intense Bragg reflections of a crystal can be studied at such temporal resolutions, excluding the analysis of complex systems. The analysis of intermediate states in time-resolved protein crystallography requires thousands of Bragg reflections. Such high numbers of reflections can be obtained with synchrotrons using Laüe geometry, but the time resolution is limited to the 50 ps shortest X-ray burst achievable with the present third-generation devices.[6]

As with crystal X-ray diffraction, gas phase electron diffraction was soon recognized for the analysis of molecular structures in their ground state.[6,11] The first investigations of transient excited molecular species were with a pulsed 1 µs electron beam (≈10 keV), electromagnetically chopped and synchronized with the exciting laser pulse. High temporal resolution resulted from the subsequent production of short laser-driven electron sources (Figure 2). Picosecond time-resolution has been demonstrated in the studied photochemical processes.[12,13] Electron diffraction is also extensively used to characterize solid surfaces. The escape depth of electrons at energies of a few hundreds of electron volts is limited to 10 Å, whereas electrons at higher energies (few tens of keV) may probe several atomic layers. Experiments have recently been performed at a few picoseconds resolution.[6]

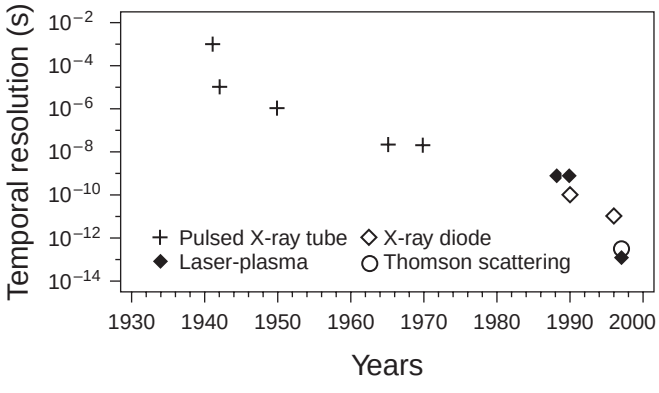

Figure 1 Experimental time resolution in X-ray diffraction experiments.

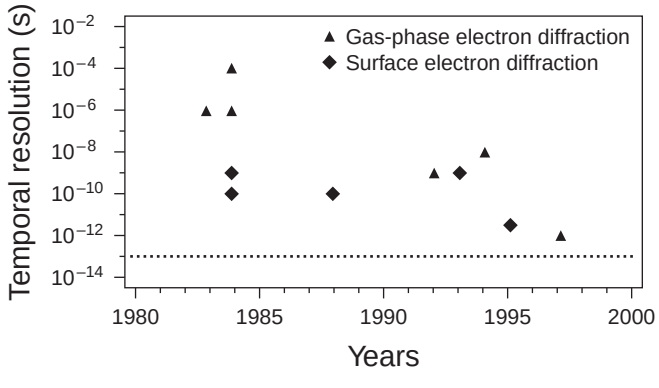

Figure 2 Experimental time resolution in electron diffraction experiments.

2 ULTRAFAST TIME RESOLUTION

There are two methods with which to reach subpicosecond time resolution for the analysis of ultrafast atomic conformation changes. Both require an optical pump pulse to excite the sample for less time than the ultrafast process under study.

2.1 High-speed Detectors

One method is to use long X-ray or electron beams associated with a fast detector, such as a streak camera (Figure 3).[6,14,15] The principle of the fast detector is displayed in Figure 4. At the photocathode, the radiation is converted into an electron beam which is accelerated to few tens of keV and streaked by pulsed electrostatic fields. The time resolution is mainly limited by the initial electron energy dispersion at the photocathode

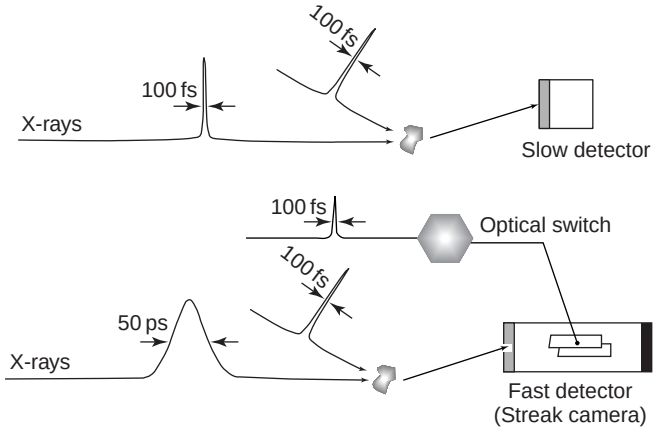

Figure 3 The two schemes to reach subpicosecond time resolution with X-rays and electrons.

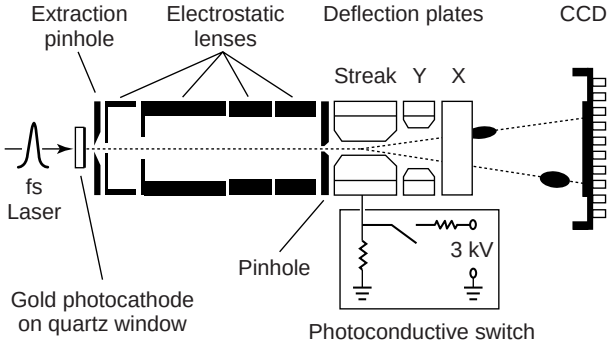

Figure 4 Streak camera principle.

(Δt_P), by the temporal broadening of the electron packet affected by Coulombic repulsion during its transport from the photocathode to the sample, by the transit time dispersion in the electron focusing optics region, and by the geometrical arrangement. As is shown in Table 1, Δt_P may be kept below 100 fs, thereby providing a very high extraction field in the range of 27 kV mm^{-1}. Nevertheless, solving the equations of motion for an electron packet propagating under Coulombic repulsion clearly shows that subpicosecond time resolution may be reached by only few electrons in the packet. A high-efficiency detector such as a CCD must be used in that case. An overview of streak camera limits is reported in Table 1 for two different photoelectron energy widths produced by the X-ray beam at the photocathode. Time resolutions lower than 300 fs may be reached in the future for single-shot experiments. However, time-resolved X-ray diffraction experiments in the subpicosecond timescale provide a low signal-to-noise ratio behavior. The main drawback of this device is undoubtedly its very poor overall efficiency (0.1%). Accumulation over many shots is necessary to obtain a usable signal. Its synchronization with both the pump and probe pulses then leads to a shot-to-shot temporal jitter, which takes the minimum resolution into the picosecond domain.

2.2 Femtosecond Sources

In the other method (Figure 3), both the optical pump and X-ray or electron pulses are short and come from the same laser system, thus preventing any temporal jitter. For a long time, short optical laser pulses have been the only way to probe ultrafast phenomena in biochemistry and physics. Using this method, the output energy has now become high enough to produce these intense electron or X-ray bursts directly for transient states analysis. The detector can be large and very efficient and single-shot experiments are achievable.

3 ULTRAFAST X-RAY DIFFRACTION

3.1 Femtosecond X-ray Sources

Laser-produced plasmas from 100 fs intense laser systems are good candidates to produce subpicosecond X-ray sources.[16–21] The energy deposited at the surface of a solid-state target produces a high temperature (500 eV) and a dense plasma which remains at near-solid densities during the interaction because it has no time to expand. In such plasmas, high ionization stages of ions are reached and strong nonequilibrium states are produced. X-ray emission produced from this thermal plasma lasts longer than the laser pulse duration because it takes few picoseconds to be quenched by cooling. However, at laser peak power densities around 10^{17} W cm^{-2}, strong nonlinear processes accelerate some plasma electrons to energies of 10–60 keV. As shown in Figure 5, these fast electrons penetrate inside the solid target deeper than the plasma depth and lose their energy mainly by K-shell ionization of the atoms producing monochromatic Kα X-radiation. At LOA (Laboratoire d'Optique Appliquée), the efficiency of X-ray sources[20,21] has now reached $7–10^{-5}$ and 10^9 photons per steradian can be produced at 7.12 Å. To date, these laser-produced plasma X-ray bursts are the most intense that can be produced in the 100 fs timescale. Third-generation synchrotrons cannot reduce

Table 1 Current limits of the streak camera technology for two photoelectron energy widths. The total temporal resolution is defined as $((\Delta t_P)^2 + (\Delta t_{CB})^2 + (\Delta t_D)^2 + (\Delta t_S)^2)^{1/2}$

Problem	Current limits	Possible solutions
Electron energy dispersion (Δt_P)	5 eV → 260 fs 0.5 eV → 55 fs	More voltage at the photocathode; low energy-dispersion photocathode
Space charge effect (Δt_{CB})	0 fs	Internal CCD
Drift region before plates (Δt_D)	5 eV → 140 fs 0.5 eV → 30 fs	Move plates closer to the photocathode; high average field before plates
Spatial extend of the focus (Δt_S)	263 fs at 1.9 c sweep	Smaller slit at the photocathode; faster sweep
Total temporal resolution	5 eV → 478 fs 0.5 eV → 276 fs	

CCD, charge-coupled device.

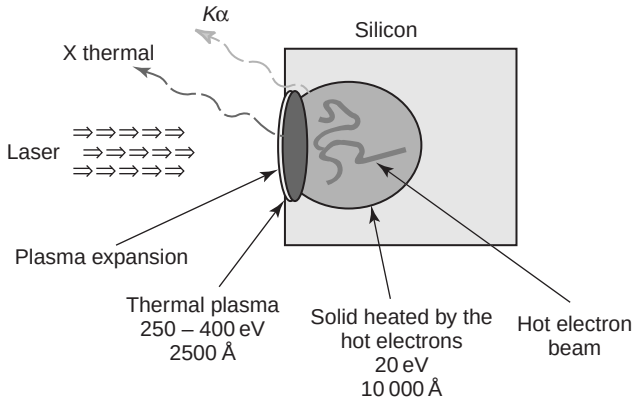

Figure 5 Kiloelectron-volt and 100 fs X-ray source from laser-produced plasma experiments.

X-ray pulse durations below a few tens of picoseconds.[6] At ESRF (European Synchrotron Radiation Facility), 50 ps X-ray flashs can contain 10^7 collimated photons at 8 keV and 10^{11} collimated photons within the range 7–11 keV. Thomson scattering of 100 fs intense laser pulses of such electron bunches may significantly reduce the available pulse duration from such systems. At the ALS (Advanced Light Source) synchrotron in Berkeley, 10^5 collimated X-rays at 30 keV and with a pulse duration of 300 fs have been produced.[22] At UCI (University of California, Irvine), a bunch of photoelectrons is produced on a photocathode by a 100 fs laser beam. These electrons are then accelerated to few 10 keV and focused onto the anode, producing $10^4\,\mathrm{sr}^{-1}$ monochromatic $K\alpha$ photons.[6] The X-ray pulse duration is limited to few picoseconds due to temporal broadening of the electron beam from the coulombic repulsion between electrons during their transport from the photocathode to the anode.

3.2 Operating Principles

The first experiments with subpicosecond time resolutions have recently been realized on simple systems using ultrafast X-ray sources.[7,9,10,23] In these experiments, first-order X-ray Bragg reflections are used to analyze the lattice strain and the disordering of atoms inside the surface of simple crystals following its excitation by an optical 100 fs laser pulse. Being produced by the same laser system, the exciting pulse and the probe X-ray flash are perfectly synchronized. The X-ray flash duration gives the resolution at which the atomic dynamics is probed and the delay lines on both laser arms allow X-ray probing of the sample before, during, and after its excitation. A spectral shift of the Bragg reflection provides information on the transient atomic plane spacing being compressed or expanded. The width of the transient rocking curve is the signature of the lattice correlation length over which each atomic plane gives a similar contribution to the scattering amplitude. It is inversely proportional to the number of atomic planes (n) participating coherently in the diffraction, i.e. $W_\mathrm{B} = (2\pi n/tg\theta_\mathrm{B})^{-1}$. Thus, high spectral resolution of the width of the transient rocking curves may allow propagation of the perturbation inside the lattice to be observed over the stationary correlation length. It may also give information on the homogeneity of the sample excitation by the pumping laser. Finally, the peak intensity is a signature of the remaining order/disorder of the diffracting atoms inside the correlation length.

3.3 Experiments

Table 2 summarizes experiments with the available ultrafast X-ray sources. The LOA experiment demonstrates that laser-produced plasma X-ray sources can follow simple atomic processes lasting less than 1 ps with a time resolution of 100 fs. The spatial resolution was kept low enough (0.3 Å) to maintain high signal-to-noise ratio data acquisition. However, laser-produced plasma X-ray sources can provide high spatial resolution, as shown in the UCSD experiment in which mÅ precision was reached.

As an example of such experiments, the LOA set-up is shown in Figure 6. The 7.12 Å $K\alpha$ X-ray radiation emitted from a silicon target is collected over a solid angle of 7.4×10^{-3} sr by a toroidally bent quartz crystal (100) and focused onto the sample. The X-ray focal spot contained 5.8×10^4 photons per shot for efficient laser–solid coupling. The first-order Bragg reflection

Table 2 Summary of ultrafast X-ray sources

Parameter	UCSD	UCI	ALS	LOA
X-ray source	Plasma–laser	X-ray diode	Thomson scattering	Plasma–laser
Time resolution	100 fs	Few ps	300 fs	100 fs
Sample	GaAs (111)	Au crystal	InSb (111)	LB
Observed process	Phonon lattice dilatation	Phonon lattice dilatation	Phonon lattice dilatation (disordering?)	Coulombic repulsion disordering
Time response of the observed process	10 ps	50 ps	40 ps (<few ps?)	0.6 ps

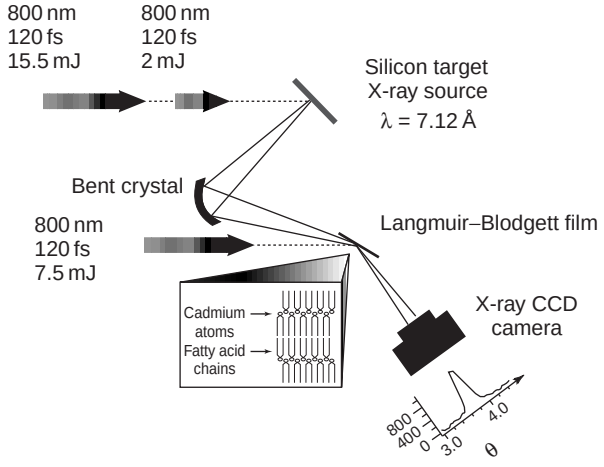

Figure 6 Experimental set-up for ultrafast X-ray diffraction experiments.

(3.6° Bragg angle), of a cadmium arachidate LB film was analyzed.[24] The structure of the 3500 Å thick LB film is shown in Figure 6. This fatty acid salt has a polar head group, where the cadmium atom is located, and a long hydrophobic tail. When deposited on a glass substrate by the LB technique, the Cd arachidate molecules form a multilayer film, with alternating orientation of the layers: head group against head group and hydrophobic tail against hydrophobic tail. This places the cadmium atoms in well-defined planes with a separation of 55 Å. The $300 \times 75\,\mu m^2$ X-ray focal spot was stretched onto the sample's surface to 1.2 mm. The Bragg reflection was detected by a cooled X-ray CCD camera, allowing the measurement of the low amplitude signals which do not exceed 40 diffracted photons in one laser shot.

The transient diffracted signal is studied as a function of the delay between the X-ray pulse and a pumping laser pulse used to perturb the sample at fluxes ranging from $1.8\,J\,cm^{-2}$ to $26.7\,J\,cm^{-2}$. One experimental run consisted of five reference shots without the pumping beam on a fresh surface, one shot with the pumping beam on and five shots without the pumping beam in order to probe the damaged area.

For all fluxes, a 75% decrease of the diffracted X-ray intensity was observed in a timescale from few hundreds of femtoseconds to few tens of picoseconds, depending on the energy of the exciting pulse. Results are shown in Figure 7 for four different pumping fluxes. The time response of the film is 15 ps at $1.8\,J\,cm^{-2}$. It drops down to 1 ps at $11.8\,J\,cm^{-2}$ and is lower than 600 fs at $27.6\,J\,cm^{-2}$. This loss in diffracted intensity can be understood as a disordering of the diffracting atoms produced inside each cadmium layer before any thermal expansion of the lattice has time to occur.

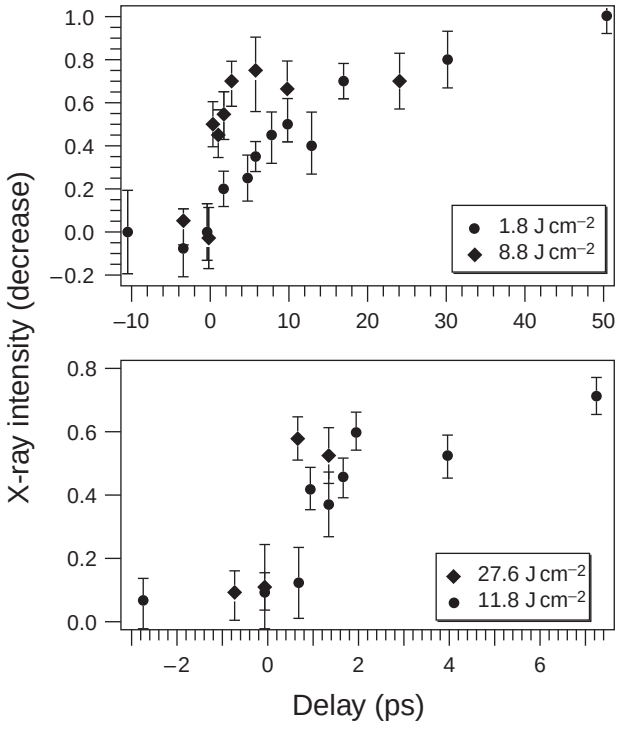

Figure 7 Decrease in diffracted X-ray intensity by LB film as a function of the delay between the pump pulse and the X-ray probe pulse. The sample is excited at fluxes ranging from $1.8\,J\,cm^{-2}$ to $26.7\,J\,cm^{-2}$.

Using the Debye–Waller theory,[25] a broadening of 8 Å of the electronic density in each layer of the film must be induced to provide the 75% loss of diffracted X-ray intensity. At higher than $27\,J\,cm^{-2}$, the signal-to-noise ratio on the detector becomes very poor due to X-ray emission produced from the sample itself, which hides the transient diffracted X-ray signal. Such large atomic displacements occurring in less than 1 ps probably come from the very specific film structure. In fact, each film layer contains two layers of cadmium atoms separated by 3 Å and the distance between two atoms is about 5 Å. At high pumping intensities ($10^{14}\,W\,cm^{-2}$), bonds are broken and atoms are ionized; ions then experience strong electrostatic repulsions. Rough numerical simulations, not taking into account the influence of the organic chains, are consistent with the observed subpicosecond film time response. In the case of the LB film, cadmium atoms experience weak ionic bonding with the carboxyl group of the organic chain. Experiments were conducted on thallium acid phthalate crystals in which the thallium atoms are strongly linked to the crystal structure by four covalent bonds. At high pumping fluxes ($21\,J\,cm^{-2}$), the time response is found to be 3 ps, much slower than for the film at the same flux. This confirms the ability of the LB film to experience high disordering.

4 ULTRAFAST ELECTRON DIFFRACTION

4.1 Electron Femtosecond Sources

The production of femtosecond electron bunches takes advantage of developments in streak camera technology.[12,13,26–28] Femtosecond electron bunches are produced from femtosecond laser pulses focused on negatively biased photocathodes. Ejected photoelectrons are accelerated to energies around a few tens of kilo electron volts and are focused onto the sample instead of being streaked in the case of streak camera operations. Figure 8 reports the experimental data obtained at CALTECH (California Institute of Technology) using this technology. A temporal broadening of 3 ps is already produced for 1000 electrons in the packet due to coulombic repulsion inside the beam.[12] Subpicosecond time resolution may be reached with very few electrons in the packet and a highly efficient detector.

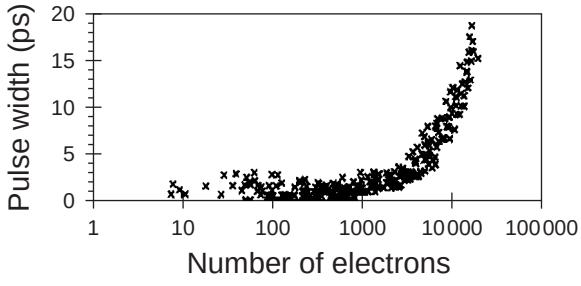

Figure 8 Influence of the space–charge effect on the electron pulse duration using the streak camera technology.

4.2 Operating Principles

In contrast to X-rays, electrons are scattered by the electronic density cloud and by the nuclei of the atoms. The intensity of scattering is then orders of magnitude larger than for X-rays. Electron scattering is well suited to probe less-dense matter, such as gas-phase samples, and is also used to probe solid-state surfaces up to few 100 Å thicknesses.[29–31] As shown in Figure 9, information on the internuclear distances is projected as a macroscopic diffraction diagram made of concentric rings of radius equal to the amplitude of the momentum transfer vector **s**. Fourier inversion of the signals gives a direct snapshot of the interatomic distance for each delay following the excitation. The calculations for the dissociation of IBr, also shown in this figure, demonstrate that time-dependent phenomena can be resolved with a shorter resolution than the width of the electron probe pulse.[6]

4.3 Experiments

Two kinds of experiments have been carried out. The smallest time resolution to date is 4 ps, obtained at

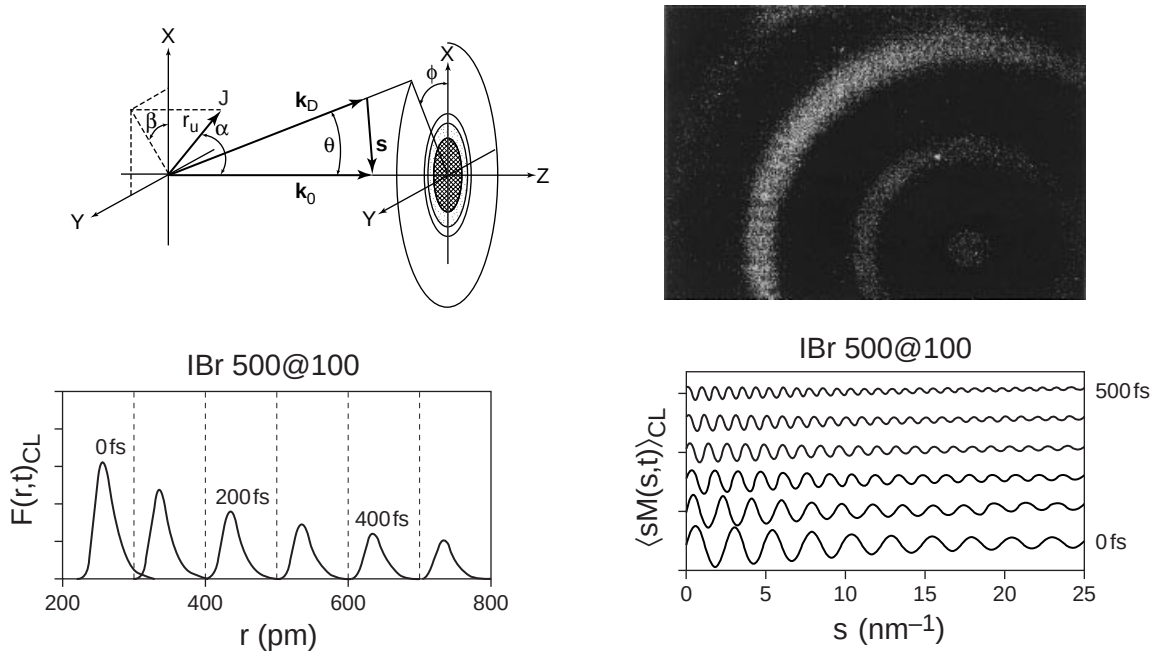

Figure 9 Elementary principles of electron diffraction. The incident electron beam travels along the wave vector $\mathbf{k}_0$. Electrons scatter at angle θ with the wave vector $\mathbf{k}_D$. The quantity $s = (4\pi/\lambda_e)\sin(\theta/2)$ is the scattering parameter defined by $\mathbf{s} = \mathbf{k}_0 - \mathbf{k}_D$. The scattering intensities are in units of s. Simulations are shown for IBr excitation from its ground state to the dissociative state of IBr probed by a 500 fs electron beam.[6]

CALTECH by using back-illuminated streak camera technology.[12] The C–I and I–I internuclear distances were followed in the photodissociation of CH_2I_2. Two other groups, at Arkansas[13] and Norfolk,[6] have demonstrated time a resolution of few tens of picoseconds from a front-illuminated streak camera.

5 CONCLUSIONS

The important challenge in electron time-resolved diffraction is now to demonstrate subpicosecond resolution in gas-phase experiments. For ultrafast time-resolved X-ray diffraction experiments, the available X-ray photon flux prevents any subpicosecond or picosecond time-resolved experiments on low-Z materials, even in the solid phase. Furthermore, the study of more complex structures requires the analysis of thousands of Bragg reflections and the potential use of laser-produced plasma X-ray sources strongly depends on an ability to demonstrate new X-ray diffraction geometries. Synchrotrons provide polychromatic and collimated X-ray beams. A static Laüe diffraction pattern of Photo Yellow Protein, with an X-ray flash lasting 50 ps, has been obtained.[32] Dynamic studies might follow this result. However, the available third-generation technology will not permit shorter time resolutions. The laser-produced plasma X-ray source is highly monochromatic and divergent. According to the Bragg law, this is another way to pick up many Bragg reflections inside the sample,[33] as shown in Figure 10.

For a triclinic crystal of lysozyme illuminated with $20 \times 100\,\text{mrad}^2$ 1.54 Å X-rays, we have estimated a 42-min accumulation time at 10 Hz to be necessary to reach a minimum of 10^2–10^3 photons in each spot with the available X-ray flux (2×10^9 photons/rad^2/shot). This new geometry will really open up new capabilities in ultrafast time-resolved X-ray diffraction experiments.

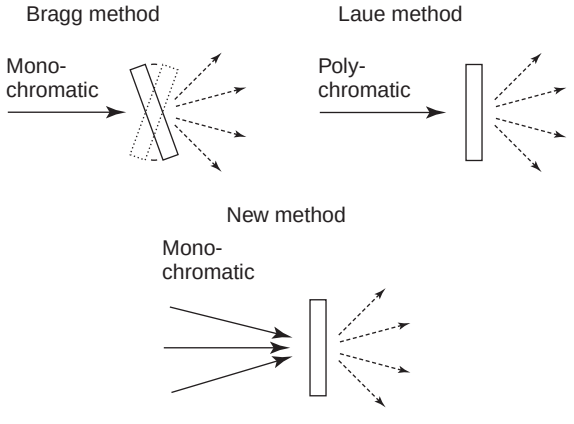

Figure 10 X-ray diffraction geometry with a divergent source for probing multiple Bragg planes.

ACKNOWLEDGMENTS

The author would like to thank Christian Rischel (Royal Veterinary and Agricultural University of Copenhagen), Ingo Uschmann and Eckhart Foerster (Jena University), Jean-Paul Geindre, P. Audebert, and J.C. Gauthier (LULI laboratory, France) for their collaboration to this research project. This work is supported by the Centre National de la Recheche Scientifique, Ecole Nationale Superieure des Techniques Avancées, Ecole Polytechnique and by the European Community under the Large Facility Programme.

ABBREVIATIONS AND ACRONYMS

ALS	Advanced Light Source
CALTECH	California Institute of Technology
CCD	Charge-coupled Device
ESRF	European Synchrotron Radiation Facility
LB	Langmuir–Blodgett
LOA	Laboratoire d'Optique Appliquée
UCI	University of California, Irvine
UCSD	University of California, San Diego

RELATED ARTICLES

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Nucleic Acids Structure and Mapping **(Volume 6)**
X-ray Structures of Nucleic Acids

Peptides and Proteins **(Volume 7)**
X-ray Crystallography of Biological Macromolecules

Surfaces **(Volume 10)**
Electron Microscopy and Scanning Microanalysis • Ellipsometry in Analysis of Surfaces and Thin Films

Nuclear Methods **(Volume 14)**
PIXE (Particle-induced X-ray Emission)

X-ray Spectrometry **(Volume 15)**
Structure Determination, X-ray Diffraction for • X-ray Techniques: Overview

General Articles **(Volume 15)**
Ultrafast Laser Technology and Spectroscopy

REFERENCES

1. J.L. Martin, A. Migus, G.A. Mourou, A.H. Zewail (eds.), *Ultrafast Phenomena VIII*, Springer-Verlag, 1992.
2. P.F. Barbara, W.H. Knox, G.A. Mourou, A.H. Zewail (eds.), *Ultrafast Phenomena IX*, Springer-Verlag, 1994.
3. P.F Barbara, J.G. Fujimoto, W.H. Knox, W. Zinth (eds.), *Ultrafast Phenomena X*, Springer-Verlag, 1996.
4. A.H. Compton, S.K. Allison, *X-rays in Theory and Experiment*, ed. D. van Nostrand, 1954.
5. R.W. James, *The Optical Properties of the Diffraction of X-rays*, Ox Bow Press, Woodbridge, 1947.
6. J.R. Helliwell, P.M. Rentzepis, in *Time-resolved Diffraction*, eds. J. Chikawa, J.R. Helliwell, S.W. Lovesey, Oxford University Press, 1997.
7. C. Rischel, A. Rousse, I. Uschmann, P.A. Albouy, J.P. Geindre, P. Audebert, J.C. Gauthier, E. Foerster, J.L. Martin, A. Antonetti, 'Femtosecond Time-resolved X-ray Diffraction from Laser-heated Organic Films', *Nature*, **390**, 490 (1997).
8. C. Rose-Petruck, T. Guo, R. Jimenez, F. Ràski, J.A. Squier, B.C. Walker, K. Wilson, C.P.J. Barty, 'Measurements of Ultrafast Dynamics in GaAs Crystals Using Time-resolved X-Ray Diffraction', Presented at the Application of High Field & Short Wavelength Sources Topical Meeting, Santa Fe, 1997.
9. T. Guo, C. Rose-Petruck, R. Jimenez, F. Ràski, J.A. Squier, B.C. Walker, K. Wilson, C.P.J. Barty, *Time-resolved X-ray Diffraction of GaAs with a 20 fs, Laser-driven Plasma Source*, eds. A. Kyrala, J.C. Gauthier, SPIE, San Diego, 1997.
10. A.H. Chin, R.W. Schoenlein, T.E. Glover, P. Balling, W.P. Leemans, C.V. Shank, 'Ultrafast Structural Dynamics in InSb Probed by Time-resolved X-ray Diffraction', Presented at the Ultrafast Phenomena XI, 1998.
11. E.D. Potter, J.L. Herek, S. Pederson, Q. Liu, A.H. Zewail, 'Femtosecond Laser-control of a Chemical Reaction', *Nature*, **355**, 66–68 (1992).
12. J.C. Williamson, J. Cao, H. Ihee, H. Zewail Frey, 'Clocking Transient Chemical Changes by Ultrafast Electron Diffraction', *Nature*, **386**, 159–162 (1997).
13. V.A. Lobastov, J.D. Ewbank, L. Schäfer, A.A. Ischenko, 'Instrumentation for Time-resolved Electron Diffraction Spanning the Time Domain from Microseconds to Picoseconds', *Rev. Sci. Instrum.*, **69**(7), 2633–2643 (1998).
14. J. Larsson, Z. Chang, E. Judd, P.J. Schuck, R.W. Falcone, P.A. Heimann, H.A. Padmore, H.C. Kapteyn, P.H. Bucksbaum, M.M. Murnane, R.W. Lee, A. Machacek, J.S. Wark, X. Liu, B. Shan, 'Ultrafast X-ray Diffraction using a Streak-camera Detector in Averaging Mode', *Optics Letters*, **22**(13), 1012 (1997).
15. J. Larsson, P.A. Heimann, A.M. Lindenberg, P.J. Schuck, P.H. Bucksbaum, R.W. Lee, H.A. Padmore, J.S. Wark, R.W. Falcone, 'Ultrafast Structural Changes Measured by Time-resolved X-ray Diffraction', *Applied Physics A*, **66**, 587 (1998).
16. J.C. Gauthier, S. Bastiani, P. Audebert, J.P. Geindre, A. Rousse, C. Quoix, G. Grillon, A. Mysyrowicz, A. Antonetti, R. Mancini, A. Shlyaptseva, *Characterization of a Femtosecond Laser-produced Plasma X-ray Source by Electronic, Optical and X-ray Diagnostic Techniques*, SPIE, San Diego, 52, 1997.
17. J.P. Geindre, P. Audebert, A. Rousse, F. Falliès, J.C. Gauthier, A. Mysyrowicz, G. Grillon, J.P. Chombaret, A. Antonetti, A. Mens, R. Verrichia, R. Scuneuf, P. Schirman, 'X-rays Generated by Femtosecond Laser-produced Plasmas', Presented at the Ultrafast Phenomena VIII, Antibes – France, 1993, ed. H.K.V. Lotsch, Series in Chemical Physics, Springer-Verlag, 272, Vol. 55, 1993.
18. J.C. Kieffer, M. Chaker, J.P. Matte, H. Pépin, C.Y. Côté, Y. Beaudoin, T.W. Johnston, C.Y. Chien, S. Coe, G. Mouron, O. Peyrusse, 'Ultrafast X-ray Sources', *Phys. Fluids B*, **5**, 2676 (1993).
19. M.M. Murnane, H.C. Kapteyn, M.D. Rosen, R.W. Falcone, 'Ultrafast X-ray Pulses from Laser-Produced Plasmas', *Science*, **251**, 531 (1991).
20. A. Rousse, P. Audebert, J.P. Geindre, F. Falliès, J.C. Gauthier, A. Mysyrowicz, G. Grillon, A. Antonetti, 'An Efficient $K\alpha$ X-ray Source from Femtosecond Laser-Produced Plasmas', *Phys. Rev. E*, **50**(3), 2200 (1994).
21. S. Bastiani, A. Rousse, J.P. Geindre, P. Audebert, C. Quoix, G. Hamoniaux, A. Antonetti, J.C. Gauthier, 'Experimental Study of the Interaction of Subpicosecond Laser Pulses with Solid Targets of Varying Initial Scale Lengths', *Phys. Rev. E*, **56**, 7179 (1997).
22. R.W. Schoenlein, 'Femtosecond X-ray Pulses at 0.4 Å Generated by 90° Thomson Scattering: A Tool for Probing the Structural Dynamics of Materials', *Science*, **274**, 236–238 (1996).
23. P. Chen, V. Tomov, P.M. Rentzepis, 'Time-resolved Heat Propagation in a Gold Crystal by Means of Picosecond X-ray Diffraction', *J. Chem. Phys.*, **104**, 1001–1007 (1996).
24. D.K. Schwartz, 'Langmuir–Blodgett Film Structure', *Surf. Sci. Rep.*, **27**, 334 (1997).
25. H.P. Klug, L.E. Alexander, *X-Ray Diffraction Procedure*, John Wiley and Sons, New York, 1959.
26. H. Niu, W. Sibbett, 'Theoretical Analysis of Space-charge Effects in Photochron Streak Cameras', *Rev. Sci. Instrum.*, **52**(12), 1830–1836 (1981).
27. M. Dantus, S.B. Kim, C. Williamson, A.H. Zewail, 'Ultrafast Electron Diffraction. 5. Experimental Time-resolution and Application', *J. Phys. Chem.*, **98**, 2782–2796 (1994).
28. W.L. Faust, D.L. Monts, L. Schäfer, 'Molecular Electron Diffraction From Space-charge Limited Beam', *Rev. Sci. Instrum.*, **59**(4), 550–556 (1988).

29. J.C. Williamson, A.H. Zewail, 'Ultrafast Electron Diffraction. 4. Molecular Structures and Coherent Dynamics', *J. Phys. Chem.*, **98**, 2766–2781 (1994).
30. J.C. Williamson, M. Dantus, S.B. Kim, A.H. Zewail, 'Ultrafast Diffraction and Molecular Structure', *Chem. Phys. Lett.*, **196**(6), 529–534 (1992).
31. A.A. Ischenko, V.P. Spiridonov, L. Schäfer, J.D. Ewbank, 'The Stroboscopic Gas Electron Diffraction Method for Investigation of Time-resolved Structural Kinetics in Photoexcitation Processes', *J. Mol. Struct.*, **300**, 115–140 (1993).
32. M. Wulff, D. Bourgeois, F. Schotte, 'Life Sciences', in *Highlights ESRF 1996/1997*, 1997.
33. A. Rousse, C. Rischel, J.C. Gauthier, 'Ultrafast X-ray Diffraction', *Rev. Mod. Phys.*

Wavelength-dispersive X-ray Fluorescence Analysis

Ron Jenkins
International Centre for Diffraction Data, Newtown Square, USA

1 Introduction **2**
1.1 Properties of X-radiation 2
1.2 General Applications of X-rays 2
1.3 Historical Development of X-ray Spectrometry 2
2 Characteristic X-ray Wavelengths **4**
3 Analyzing Crystals and Dispersion Devices **5**
4 Detectors for X-ray Fluorescence **6**
4.1 Desirable Properties of an X-ray Detector 6
4.2 Types of X-ray Detector 6
5 X-ray Spectrometers **8**
5.1 X-ray Excitation Sources 8
5.2 X-ray Spectrometer Configurations 8
5.3 Resolution of the Spectrometer 9
5.4 Spectrometers for X-ray Fluorescence 10
6 Wavelength-dispersive Spectrometer **11**
6.1 Sequential Spectrometers 11
6.2 Simultaneous Spectrometers 12
6.3 Other Types of X-ray Spectrometer 12
7 Qualitative Analysis with the X-ray Spectrometer **14**
8 Accuracy of X-ray Fluorescence **14**
8.1 Counting Statistical Errors 15
8.2 Matrix Effects 15
8.3 Specimen Preparation for X-ray Fluorescence 16
9 Quantitative Analysis **17**
9.1 Single-element Techniques 18
9.2 Use of Internal Standards 18
9.3 Type Standardization 18
9.4 Influence Correction Methods 19
9.5 Fundamental Methods 19
10 Trace Analysis **20**
10.1 Analysis of Low Concentrations 20
10.2 Analysis of Small Amounts of Sample 20
Abbreviations and Acronyms **21**
Related Articles **21**
References **21**

When an element is bombarded by high-energy particles, orbital electrons may be ejected creating inner orbital atomic vacancies. These vacancies may be filled by transition of outer level electrons giving rise to characteristic X-radiation. X-ray fluorescence spectrometry provides the means of the identification of an element by measurement of its characteristic X-ray emission wavelength of energy. The method allows the quantization of a given element by first measuring the emitted characteristic line intensity and then relating this intensity of element concentration. While the roots of the method go back to the early part of this century, where electron excitation systems were employed, it is only during the last 30 years or so that the technique has gained major significance as a routine means of elemental analysis. Wavelength-dispersive spectrometers employ diffraction by a single crystal to separate characteristic wavelengths emitted by the sample. Today, nearly all commercially available X-ray spectrometers use the fluorescence excitation method and employ a sealed X-ray tube as the primary excitation source. The first commercial X-ray spectrometer became available in the early 1950s and although these earlier spectrometers operated only with an air path, they were able to provide qualitative and quantitative information on all elements above atomic number 22 (titanium). Later versions allowed use of helium or vacuum paths that extended the lower atomic number cut-off to around atomic number 9 (fluorine). X-ray detectors used include the flow counter, the scintillation counter and the Si(Li) detector.

The X-ray method has good overall performance characteristics. In particular, the speed, accuracy and versatility of X-ray fluorescence are the most important features among the many that have made it the method of choice in over 30 000 laboratories all over the world. Most wavelength-dispersive spectrometers fall into two broad categories – single channel and multichannel. Single channel spectrometers are typically employed for both routine and nonroutine analysis of a wide range of products, including ferrous and nonferrous alloys, oils, slags and sinters, ores and minerals, thin films, and so on. These systems are very flexible but relative to multichannel spectrometers are somewhat slow. The multichannel wavelength-dispersive instruments are used almost exclusively for routine, high-throughout analysis where the great need is for fast accurate analysis, but where flexibility is of no importance.

Interelement (matrix) effects often complicate quantitative analysis by X-ray fluorescence. However, a wide selection of methods is now available for minimizing these

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

effects, allowing excellent accuracy to be obtained in many cases. Detection limits are achievable down to the low parts per million (ppm) range and it is possible to obtain reasonable responses from as little as a few milligrams of material.

1 INTRODUCTION

Much of the basic physics of X-ray production, absorption, scatter and diffraction has been covered in **X-ray Techniques: Overview**. The following section briefly reviews this material.

1.1 Properties of X-radiation

In the analysis of a specimen, both the incoming primary X-ray beam and the secondary fluoresced beam will be attenuated by an amount which can be theoretically described or modeled from practical experiments. The effect of absorption is typically to produce nonlinearity in a calibration curve of elemental concentration versus X-ray response. Scattering occurs when an X-ray photon interacts with the electrons of the target element. Where this interaction is elastic, that is no energy is lost in the collision process, the scattering is referred to as coherent (Rayleigh) scattering. Since no energy change is involved, the coherently scattered radiation will retain exactly the same wavelength as that of the incident beam. It can also happen that the scattered photon gives up a small part of its energy during the collision, especially where the electron with which the photon collides is only loosely bound. In this instance, the scatter is referred to as incoherent (Compton scattering). X-ray diffraction is a combination of two phenomena – coherent scatter and interference. At any point where two or more waves cross one another, they are said to interfere. The principal of superposition is that the resulting displacement, at any point and at any instant, may be formed by adding the instantaneous displacements that would be produced at the same point by independent wave trains, if each were present alone. Under certain geometric conditions, wavelengths that are exactly in phase may add to one another, and those that are exactly out of phase may cancel each other out. Under such conditions coherently scattered photons may constructively interfere with each other giving diffraction maxima (Equation 1):

$$n\lambda = 2d\sin\theta \tag{1}$$

where n is an integer. Equation (1) is a statement of Bragg's law. Bragg's law is important in wavelength-dispersive spectrometry since by using a crystal of fixed $2d$, each unique wavelength λ will be diffracted at a unique diffraction angle (Bragg angle). Thus, by measuring the diffraction angle θ, knowledge of the d-spacing of the analyzing crystal allows the determination of the wavelength. Since there is a simple relationship between wavelength and atomic number, as given by Moseley's law, one can establish the atomic number(s) of the element(s) from which the wavelengths were emitted.

1.2 General Applications of X-rays

X-ray photons are produced following the ejection of an inner orbital electron from an excited atom, and subsequent transition of atomic orbital electrons from states of high to low energy. Each element present in the specimen will produce a series of characteristic lines making up a polychromatic beam of characteristic and scattered radiation coming from the specimen. This polychromatic beam is then directed to the spectrometer where individual wavelengths (or energies) can be measured. X-ray spectrometric techniques provided important information for the theoretical physicist in the first half of this century and since the early 1950s they have found an increasing use in the field of materials characterization. While most of the early work in X-ray spectrometry was carried out using electron excitation,[1] today, use of electron-excited X-radiation is restricted mainly to X-ray spectrometric attachments to electron microscopes. Most modern stand-alone X-ray spectrometers use X-ray excitation sources rather than electron excitation. X-ray fluorescence spectrometry typically uses a polychromatic beam of short wavelength X-radiation to excite longer wavelength characteristic lines from the sample to be analyzed. Modern X-ray spectrometers use either the diffracting power of a single crystal to isolate narrow wavelength bands, or a proportional detector to isolate narrow energy bands, from the polychromatic radiation (including characteristic radiation) excited in the sample. The first of these methods is called wavelength-dispersive spectrometry and the second, energy-dispersive spectrometry (see **Portable Systems for Energy-dispersive X-ray Fluorescence**). Because the relationship between emission wavelength and atomic number is known, isolation of individual characteristic lines allows the unique identification of an element to be made and elemental concentrations can be estimated from characteristic line intensities. Thus this technique is a means of materials characterization in terms of chemical composition.

1.3 Historical Development of X-ray Spectrometry

X-ray fluorescence spectrometry provides the means of identification of an element by measurement of its characteristic X-ray emission wavelength or energy. The method allows the quantization of a given element by first

measuring the emitted characteristic line intensity and then relating this intensity to elemental concentration. While the roots of the method go back to the early part of this century, it is only during the last 30 years or so that the technique has gained major significance as a routine means of elemental analysis.[2] The first use of the X-ray spectrometric method dates back to the classic work of Henry Moseley.[3] In Moseley's original X-ray spectrometer, the source of primary radiation was a cold cathode tube in which the source of electrons was residual air in the tube itself, the specimen for analysis forming the target of the tube. Radiation produced from the specimen then passed through a thin gold window onto an analyzing crystal whence it was diffracted to the detector. One of the major problems in the use of electrons for the excitation of characteristic X-radiation is that the process of conversion of electron energy into X-rays is relatively inefficient: about 99% of the electron energy is converted to heat energy. This means in turn that it may be difficult to analyze specimens which are volatile or tend to melt. Nevertheless, the technique seemed to hold some promise as an analytical tool and one of the first published papers on the use of X-ray spectroscopy for real chemical analysis appeared as long ago as 1922, when Hadding[4] described the use of the technique for the analysis of minerals.

Some time in the mid-1920s, it was suggested that one could use primary X-ray photons for the excitation of secondary characteristic X-ray spectra. The use of X-rays, rather than electrons, to excite characteristic X-radiation avoids the problem of heating the specimen. It is possible to produce the primary X-ray photons inside a sealed X-ray tube under high vacuum and efficient cooling conditions, which means that the specimen itself need not be subject to heat dissipation problems or the high vacuum requirements of the electron beam system. Use of X-rays rather than electrons represented the beginnings of the technique of X-ray fluorescence as we know it today. The fluorescence method was first employed on a practical basis in 1928 by Glocker and Schreiber.[5] Unfortunately, data obtained at that time were rather poor because X-ray excitation is even less efficient than electron excitation. Also the detectors and crystals available at that time were rather primitive and thus the fluorescence technique did not seem to hold too much promise. In any event, widespread use of the technique had to wait until the mid-1940s when X-ray fluorescence was rediscovered (see Birks[2]). The basis of their spectrometer was a diffractometer that had been originally designed for the orientation of quartz oscillator plates. A Geiger counter was used as a means of measuring the intensities of the diffracted characteristic lines and quite reasonable sensitivity was obtained for a very large part of the atomic number range.

The first commercial X-ray spectrometer became available in the early 1950s and although these earlier spectrometers operated only with an air path, they were able to provide qualitative and quantitative information on all elements above atomic number 22 (titanium). Later versions allowed use of helium or vacuum paths that extended the lower atomic number cut-off. Most modern commercially available X-ray spectrometers have a range from about 0.4 to 20 Å (40 to 0.6 keV) and this range will allow measurement of the K series from fluorine ($Z = 9$) to lutetium ($Z = 71$), and the L series from manganese ($Z = 25$) to uranium ($Z = 92$). Other line series arise from the M and N levels but these have little use in analytical X-ray spectrometry. In practice, the number of vacancies in electron levels resulting in the production of characteristic X-ray photons is less than the total number of vacancies created in the excitation process, because the atom can also regain its initial state by reorganization of atomic electrons without the emission of X-ray photons (the Auger process). This is an important factor in determining the absolute number of counts that an element will give under a certain set of experimental conditions. It is mainly for this reason that the sensitivity of the X-ray spectrometric technique is rather poor for the very low atomic number elements, since fluorescent yields for these low atomic numbers are very small.

Today, nearly all commercially available X-ray spectrometers use the fluorescence excitation method and employ a sealed X-ray tube as the primary excitation source. Some of the simpler systems may use a radioisotope source, because of considerations of cost and/or portability. While electron excitation is generally not used in stand-alone X-ray spectrometers, it is the basis of X-ray spectrometry carried out on electron column instruments. The ability to focus the primary electron beam allows analysis of extremely small areas down to a micrometer or so in diameter, or even less in specialized instruments. This, in combination with imaging and electron diffraction, offers an extremely powerful method for the examination of small specimens, inclusions, grain boundary phenomena, etc. The instruments used for this type of work may be in the form of a specially designed electron microprobe analyzer, or simply an energy or wavelength-dispersive attachment to a scanning electron microscope.

Over the past 30 years or so, the X-ray fluorescence method has become one of the most valuable methods for the qualitative and quantitative analysis of materials. Many methods of instrumental elemental analysis are available today and among the factors that will generally be taken into consideration in the selection of one of these methods are accuracy, range of application, speed, cost, sensitivity and reliability. While it is certainly true that no one technique can ever be expected to offer all of

the features that a given analyst might desire, the X-ray method has good overall performance characteristics. In particular, the speed, accuracy and versatility of X-ray fluorescence are the most important features among the many that have made it the method of choice in over 20 000 laboratories all over the world. Both the simultaneous wavelength-dispersive spectrometer and the energy-dispersive spectrometers lend themselves admirably to the qualitative and quantitative analysis of solid materials and solutions.

2 CHARACTERISTIC X-RAY WAVELENGTHS

X-ray Techniques: Overview showed that there are two basic types of X-radiation – white radiation (produced as the impinging high energy electrons are decelerated by the atomic electrons of the elements making up the specimen) and characteristic radiation. The relationship between the wavelength of a characteristic X-ray photon and the atomic number Z of the excited element was first established by Moseley. Moseley's law is written as shown in Equation (2)

$$\frac{1}{\lambda} = K(Z - \sigma)^2 \qquad (2)$$

where Z is atomic number and K and σ are constants. Since there are two competing effects by which an atom may return to its initial state, and since only one of these processes will give rise to the production of a characteristic X-ray photon, the intensity of an emitted characteristic X-ray beam will be dependent upon the relative effectiveness of the two processes within a given atom. As an example, the number of quanta of K series radiations emitted per ionized atom is a fixed ratio for a given atomic number, this ratio being called the fluorescent yield.

An excited atom can revert to its original ground state by transferring an electron from an outer atomic level to fill the vacancy in the inner shell. An X-ray photon is emitted from the atom as part of this de-excitation step, the emitted photon having an energy equal to the energy difference between the initial and final states of the transferred electron. Each unique atom has a number of available electrons which can take part in the transfer and, since millions of atoms are typically involved in the excitation of a given specimen, all possible de-excitation routes are taken. These de-excitation routes can be defined by a simple set of selection rules that account for the majority of the observed wavelengths. Each electron in an atom can be defined by four quantum numbers. The first of these quantum numbers is the principal quantum number n, which can take all integral values. When n is equal to 1, the level is referred to as the K level; when n is 2, the L level, and so on. ℓ is the angular quantum number and this can take all values from $(n-1)$ to zero. m is the magnetic quantum number and can take values of $+\ell$, zero and $-\ell$. s is the spin quantum number with a value of $\pm\frac{1}{2}$. The total momentum J of an electron is given by the vector sum of $(\ell + s)$. Since no two electrons within a given atom can have the same set of quantum numbers, a series of levels or shells can be constructed. Table 1 gives the atomic structures of the first three principal shells. The first shell, the K shell, has a maximum of two electrons and these are both in the 1s level (orbital). Since the value of J must be positive in this instance the only allowed value is $\frac{1}{2}$. In the second shell, the L shell, there are eight electrons: two in the 2s level and six in the 2p levels. In this instance J has a value of $\frac{1}{2}$ for the 1s level and $\frac{3}{2}$ or $\frac{1}{2}$ for the 2p level, thus giving a total of three possible L transition levels. These levels are referred to as L_I, L_{II} and L_{III} respectively. In the M level, there are a maximum of 18 electrons: 2 in the 3s level, 8 in the 3p level and 10 in the 3d level. Again, with the values of $\frac{3}{2}$ or $\frac{1}{2}$ for J in the 3p level and $\frac{5}{2}$ and $\frac{3}{2}$ in the 3d level, a total of five M transition levels are possible. Similar rules can be used to build up additional levels: N, O, etc.

As was shown in a previous chapter (see **X-ray Techniques: Overview**), the selection rules for the production of normal (diagram) lines require that the principal quantum number must change by at least one, the angular quantum number must change by only one, and the J quantum number must change by zero or one. Transition groups may now be constructed, based on the appropriate number of transition levels. Application of

Table 1 Atomic structures of the first three principal shells

Shell(electrons)	n	m	ℓ	s	Orbitals	J
K(2)	1	0	0	$\pm\frac{1}{2}$	1s	$\frac{1}{2}$
L(8)	2	0	0	$\pm\frac{1}{2}$	2s	$\frac{1}{2}$
	2	1	1	$\pm\frac{1}{2}$	2p	$\frac{1}{2}$, $\frac{3}{2}$
	2	0	1	$\pm\frac{1}{2}$	2p	$\frac{1}{2}$, $\frac{3}{2}$
	2	−1	1	$\pm\frac{1}{2}$	2p	$\frac{1}{2}$, $\frac{3}{2}$
M(18)	3	0	0	$\pm\frac{1}{2}$	3s	$\frac{1}{2}$
	3	1	1	$\pm\frac{1}{2}$	3p	$\frac{1}{2}$, $\frac{3}{2}$
	3	0	1	$\pm\frac{1}{2}$	3p	$\frac{1}{2}$, $\frac{3}{2}$
	3	−1	1	$\pm\frac{1}{2}$	3p	$\frac{1}{2}$, $\frac{3}{2}$
	3	2	2	$\pm\frac{1}{2}$	3d	$\frac{3}{2}$, $\frac{5}{2}$
	3	−1	2	$\pm\frac{1}{2}$	3d	$\frac{3}{2}$, $\frac{5}{2}$
	3	0	2	$\pm\frac{1}{2}$	3d	$\frac{3}{2}$, $\frac{5}{2}$
	3	−1	2	$\pm\frac{1}{2}$	3d	$\frac{3}{2}$, $\frac{5}{2}$
	3	−2	2	$\pm\frac{1}{2}$	3d	$\frac{3}{2}$, $\frac{5}{2}$

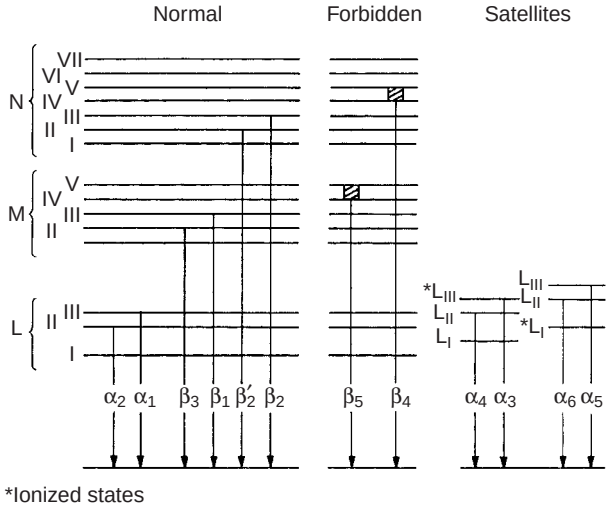

Figure 1 Observed lines in the K series. (Reproduced from R. Jenkins, *X-ray Fluorescence Spectrometry*, 2nd edition, John Wiley & Sons, New York, 1999, by permission of John Wiley & Sons, Ltd.)

the selection rules indicates that in, for example, the K series, only $L_{II} \rightarrow K$ and $L_{III} \rightarrow K$ transitions are allowed for a change in the principal quantum number of one. There are equivalent pairs of transitions for $n = 2$, $n = 3$, $n = 4$, etc. Figure 1 shows the lines that are observed in the K series. Three groups of lines are indicated. The normal lines are shown on the left-hand side of the figure and these consist of three pairs of lines from the L_{II}/L_{III}, M_{II}/M_{III} and N_{II}/N_{III} sub-shells respectively. While most of the observed fluorescent lines are normal, certain lines may also occur in X-ray spectra that do not at first sight fit the basic selection rules. These lines are called forbidden lines and are shown in the center portion of the figure. Forbidden lines typically arise from outer orbital levels where there is no sharp energy distinction between orbitals. As an example, in the transition elements, where the 3d level is only partially filled and is energetically similar to the 3p levels, a weak forbidden transition (the β_5) is observed. A third type of line may also occur: satellite lines, which arise from dual ionizations. Following the ejection of the initial electron in the photoelectric process, a short, but finite, period of time elapses before the vacancy is filled. This time period is called the lifetime of the excited state. For the lower atomic number elements, this lifetime increases to such an extent that there is a significant probability that a second electron can be ejected from the atom before the first vacancy is filled. The loss of the second electron modifies the energies of the electrons in the surrounding sub-shells, and other pairs of X-ray emission lines are produced, corresponding to the $K\alpha_1/K\alpha_2$. In the K series the most common of these satellite lines are the $K\alpha_3/K\alpha_4$ and the $K\alpha_5/\alpha_6$ doublets. These lines are shown at the right-hand side of the figure. Although, because they are relatively weak, neither forbidden transitions nor satellite lines have great analytical significance, they may cause some confusion in qualitative interpretation of spectra and may even be misinterpreted as coming from trace elements.

3 ANALYZING CRYSTALS AND DISPERSION DEVICES

The angular dispersion $d\theta/d\lambda$ of a crystal of spacing $2d$ is given by Equation (3)

$$\frac{d\theta}{d\lambda} = \frac{n}{2d\cos\theta} \tag{3}$$

It will be seen from Equation (3) that the angular dispersion will be high when the $2d$-spacing is small. This is unfortunate as far as the range of the spectrometer is concerned, because a small value of $2d$ means in turn a small range of wavelengths coverable. Thus, as with the resolution and peak intensities, high dispersion can only be obtained at the expense of cutting the wavelength range covered by a particular crystal. In order to circumvent this problem, it is likely that several analyzing crystals will be employed in the coverage of a number of analyte elements. Many different analyzing crystals are available, each having its own special characteristics, but three or four crystals will generally suffice for most applications. Table 2 lists the more commonly used crystals. While the maximum wavelength covered by traditional spectrometer designs is about 20 Å, recent developments now allow the extension of the wavelength range significantly beyond this value.

Classically, large single crystals have been used as diffracting structures in the wavelength-dispersive spectrometer. The three-dimensional lattice of atoms is oriented and fabricated such that Bragg planes form the interatomic $2d$-spacing for the wavelength in question. The selection of crystals for the longer (>8 Å) wavelength region is difficult, however, mainly because there are not

Table 2 Analyzing crystals used in wavelength-dispersive X-ray spectrometry

Crystal	Planes	$2d$ (Å)	Range K lines	L lines
Lithium fluoride	(220)	2.848	>Ti(22)	>La(57)
Lithium fluoride	(200)	4.028	>K(19)	>Cd(48)
Pentaerythritol	(002)	8.742	Al(13)–K(19)	–
TAP	(001)	26.4	F(9)–Na(11)	–
LSMs	–	50–120	Be(4)–F(9)	–

LSM, layered synthetic microstructure; TAP, thallium acid phthalate.

too many crystals available for work in this region. The most commonly employed crystal is probably thallium acid phthalate (TAP; $2d = 26.3$ Å), and this allows measurement of the K lines of elements including magnesium, sodium, fluorine and oxygen. Several alternatives to single crystals as diffracting structures have been sought over the years including the use of complex organic materials with large $2d$-spacings, gratings, and specular reflectors, metal disulfides, organic intercalation complexes such as graphite, molybdenum disulfide, mica and clays. For a comprehensive list of analyzing crystals see Bertin.[6]

In addition to crystals and multilayer films, a third alternative became available in the late 1970s, this alternative being the layered synthetic microstructure (LSM). LSMs are constructed by applying successive layers of atoms or molecules on a suitably smooth substrate. In this manner, both the $2d$-spacing and composition of each layer are selected for optimum diffraction characteristics, thus, to a certain extent, they can be designed and fabricated to give optimum performance for special applications. Typical experimental data show factors of about four to six times improvement in peak intensities compared to TAP, for the range of elements measured.

4 DETECTORS FOR X-RAY FLUORESCENCE

4.1 Desirable Properties of an X-ray Detector

As was stated in **X-ray Techniques: Overview**, an X-ray detector is simply a transducer for converting X-ray photon energy into voltage pulses. In order to adequately define the properties of X-ray detectors it is necessary to define certain terms, which, while in common use, are often misused. Intensity (I) is a flux of radiation, measured in number of photons per steradian. It is, in effect, a measure of what comes out of the X-ray tube. The photons are converted to pulses by the detector generating a count rate (R), measured in pulses (counts) per second. The detector pulses are integrated for time t giving a number of counts (N). The ideal detector should possess certain important properties. These properties are quantum counting efficiency, linearity, proportionality and resolution, and are illustrated in Table 3. The table shows that a photon flux I_o is incident upon the detector, and a fraction I of this incident flux passes through the detector without producing ions within the detector. The quantum counting efficiency Q is expressed as I/I_o. The property of linearity is also illustrated. Here, a number of X-ray photons are entering the detector at a rate of I photons per second, producing a rate of R pulses per second. Where R is proportional to I the detector is said to be linear. Linearity is important where the various count rates produced by the detector are to be used as measures of the photon intensity for each measured line. Table 3 also shows that when an X-ray photon of energy E enters the detector a pulse of V volts is produced. Where V is proportional to E the detector is said to be proportional. Proportionality is needed where the technique of pulse height selection is to be used. The last important detector property is associated with the use of pulse height selection and this property is the detector resolution. While each incident photon of energy E gives an average value of V, there is a distribution in the value of V equal to δV. This distribution is a measure of the detector resolution. Pulse height selection is a means of electronically rejecting pulses of voltage levels other than those corresponding to the characteristic line being measured. This technique is a very powerful tool in reducing background levels and the influence of overlapping lines from elements other than the analyte. The properties of proportionality and linearity are, to a certain extent, controllable by the electronics associated with the actual detector. To this extent, while it is common practice to refer to the characteristics of detectors, the properties of the associated pulse processing chain should always be included.

Table 3 Desirable properties of an X-ray detector

Property	Input	Output	Notes
Quantum efficiency (Q)	Photon flux I_o	I transmitted	$Q = \frac{I}{I_o}$
Linearity	Intensity (I)	Count rate (R)	$R \propto I$
Proportionality	Photon energy (E)	Voltage (V)	$V \propto I$
Resolution	Photon energy (E)	$E \pm \delta V$	$R \propto \delta V$

4.2 Types of X-ray Detector

4.2.1 The Gas Flow Proportional Detector

A gas flow proportional counter consists of a cylindrical tube about 2 cm in diameter, carrying a thin (25–50 μm) wire along its radial axis. The tube is filled with a mixture of inert gas and quench gas – typically 90% argon/10% methane (P-10). The cylindrical tube is grounded and around 1400–1800 V is applied to the central wire. The wire is connected to a resistor shunted by a capacitor. An X-ray photon entering the detector produces a number n of ion pairs, each comprising one electron and one argon positive ion. The first ionization potential for argon is about 16 eV, but competing processes during the conversion of photon energy to ionization cause the average energy required to produce an ion pair to be greater than this amount. The fraction relating the average energy to produce one ion pair, to the first

ionization potential, is called the Fano factor F. For argon, F is between $\frac{1}{2}$ and $\frac{1}{3}$ and the average energy e_i required to produce one primary ion pair is equal to 26.4 eV. Equation (4) shows that the number of ion pairs produced by a photon of energy E will equal:

$$n = \frac{E}{e_i} \tag{4}$$

Following ionization, the charges separate with the electrons moving towards the (anode) wire and the argon ions to the grounded cylinder. As the electrons approach the high field region close to the anode wire they are accelerated sufficiently to produce further ionization of argon atoms. Thus a much larger number N of electrons will actually reach the anode wire. This effect is called gas gain, or gas multiplication, and its magnitude is given by M, where M is equal to N/n. For gas flow proportional counters used in X-ray spectrometry M typically has a value of around 10^5. Provided that the gas gain is constant the size of the voltage pulse V produced is directly proportional to the energy E of the incident X-ray photon. In practice not all photons arising from photon energy E will be exactly equal to V. There is a random process associated with the production of the voltage pulses and the resolution of a counter is related to the variance in the average number of ion pairs produced per incident X-ray photon. The resolution is generally expressed in terms of the full width at half-maximum of the pulse amplitude distribution. The theoretical resolution R_t of a flow counter can be described by Equation (5):

$$R_t\% = \frac{38.3}{\sqrt{E}} \tag{5}$$

4.2.2 The Scintillation Detector

While the gas flow proportional counter is ideal for the measurement of longer wavelengths it is rather insensitive to wavelengths shorter than about 1.5 Å. For this shorter wavelength region it is common to use the scintillation counter. The scintillation counter consists of two parts, the phosphor (scintillator) and the photomultiplier. The phosphor is typically a large single crystal of sodium iodide that has been doped with thallium. When X-ray photons fall onto the phosphor, blue light photons are produced, where the number of blue light photons is related to the energy of the incident X-ray photon. These blue light photons produce electrons by interaction with a photosurface in the photomultiplier, and the number of electrons is linearly increased by a series of secondary surfaces, called dynodes, in the photomultiplier. The current produced by the photomultiplier is then converted to a voltage pulse, as in the case of the gas flow proportional counter. Since the number of electrons is proportional to the energy of the incident X-ray photon, the scintillation counter is a proportional counter.

Because of inefficiencies in the X-ray/blue-light/electron conversion processes, the average energy to produce a single event with the scintillation counter is more than a magnitude greater than the equivalent process in the flow counter. For this reason, the resolution of the scintillation counter is much worse than that of the flow counter. The theoretical resolution R_t of the scintillation counter for photons of energy E is given by Equation (6):

$$R_t\% = \frac{128}{\sqrt{E}} \tag{6}$$

4.2.3 The Si(Li) Detector (see ***Portable Systems for Energy-dispersive X-ray Fluorescence***)

The Si(Li) detector consists of a small cylinder (about 1 cm diameter and 3 mm thick) of p-type silicon that has been compensated by lithium to increase its electrical resistivity. A Schottky barrier contact on the front of the silicon disk produces a p-i-n type diode. In order to inhibit the mobility of the lithium ions and to reduce electronic noise, the diode and its preamplifier are cooled to the temperature of liquid nitrogen. By applying a reverse bias of around 1000 V, most of the remaining charge carriers in the silicon are removed. Incident X-ray photons interact to produce a number n of electron–hole pairs, as given in Equation (4), where, in this instance, e_i is equal to about 3.8 eV for cooled silicon. This charge is swept from the diode by the bias voltage to a charge-sensitive preamplifier. A charge loop integrates the charge on a capacitor C to produce an output pulse as in the case of the flow proportional counter, although in this case the M term is equal to unity since the Si(Li) detector does not have an equivalent property to gas gain.

4.2.4 Other Detector Systems

Even though the Si(Li) detector is the most common detector used in energy-dispersive X-ray spectrometry, it is certainly not the only one. The higher absorbing power of germanium makes it an alternative for the measurement of high-energy spectra, and both cadmium telluride, CdTe, and mercuric iodide, HgI_2, show some promise as detectors capable of operating satisfactorily at room temperature. There are, however, many practical problems in the manufacture of these devices. As an example, in the preparation of CdTe, a Br–methanol etchant is used and the side product $CdBr_2$ has been found to poison the surface causing high leakage current with increase in background noise, but newer technology is reducing the problem and much lower backgrounds are now being reported. The major problem with HgI_2 is the growth of crystals of suitable size. The resolution

of the detector is, however, promising, making the rather tedious research worthwhile.

5 X-RAY SPECTROMETERS

5.1 X-ray Excitation Sources

As was discussed in **X-ray Techniques: Overview**, several different types of source have been employed for the excitation of characteristic X-radiation, including those based on electrons, X-rays, γ-rays, protons and synchrotron radiation. Most conventional wavelength-dispersive X-ray spectrometers use a high power (2 to 4 kW) X-ray bremsstrahlung source. In all cases, the primary source unit consists of a very stable high voltage generator, capable of providing a potential of typically 40 to 100 kV. The current from the generator is fed to the filament of the X-ray tube, which is typically a coil of tungsten wire. The applied current causes the filament to glow, emitting electrons in all directions. A portion of this electron cloud is accelerated to the anode of the X-ray tube, which is typically a water-cooled block of copper with the required anode material plated or cemented to its surface. The impinging electrons produce X-radiation, a significant portion of which passes through a thin beryllium window to the specimen. In order to excite a given characteristic line the source must be run at a voltage V_o well in excess of the critical excitation potential V_c of the element in question. The relationship between the measured intensity of the characteristic line I, the tube current i, and the operating and critical excitation potentials is shown in Equation (7):

$$I = Ki(V_o - V_c)^{1.6} \quad (7)$$

The product of i and V_o represents the maximum output of the source in kilowatts. The optimum value for V_o/V_c is 3 to 5. This optimum value occurs because at very high operating potentials the electrons striking the target in the X-ray tube penetrate so deeply into the target that self absorption of target radiation becomes significant. Since it is the intention to eventually equate the value of I for a given wavelength or energy to the concentration of the corresponding analyte element, it is vital that, over the short term (1–2 h), both the tube current and voltage be stabilized to better than 0.1%. The current from the generator is fed to the tungsten filament of the X-ray tube. A sealed X-ray tube has an anode of Cr, Rh, W, Ag, Au or Mo and delivers an intense source of continuous and characteristic radiation, which then impinges onto the analyzed specimen, where characteristic radiation is generated. In general, most of the long wavelength excitation comes from the longer wavelength characteristic lines from the tube, and most of the short wavelength excitation from the continuous radiation from the tube.

Table 4 Stability of X-ray generators

Type of drift	Source	Magnitude (%)	Time duration
Short	Stabilization circuits	<1	30–120 min
Long	Thermal changes	0.2–0.5	Days
Ultra-short	Transients	0.2–10	50–500 ms
Ultra-long	Aging of X-ray tube	1–20	Months/years

Since the accuracy of any X-ray measurement will depend on the stability of the source it is important to understand exactly where any instability comes from. Table 4 shows that source drift is divided into four separate time-related zones: short, long, ultra-short and ultra-long. Most high voltage generators are designed to give a very low (0.05%) short-term drift. The effect of long-term drift can be calibrated out by use of internal standards. Ultra-long-term drift is mainly to do with X-ray tube aging and should be monitored on a monthly basis. The biggest problem is often the ultra-short-term drift. The stabilization circuits of the high voltage generator typically have a response time of a few hundred milliseconds. Thus if a short duration pulse (a transient) should enter the input of the generator, this may pass through giving a voltage spike with a resulting X-ray intensity spike. The solution here is to ensure that the quality of the mains supply is adequate.

5.2 X-ray Spectrometer Configurations

There are many different types of X-ray fluorescence instruments available, based on several different source and dispersion configurations. The most commonly employed instruments are based either on the wavelength-dispersive method or the energy-dispersive method (see **Portable Systems for Energy-dispersive X-ray Fluorescence**). Further subdivisions in the energy-dispersive instruments include primary or secondary source excitation mode; and in the case of wavelength-dispersive instruments, single channel (sequential) or multichannel (simultaneous). In all cases this may or may not include microprocessor control, and may or may not include a data processing computer. While each of these configurations has clear advantages over its competitors, each also has disadvantages – other than the obvious ones of cost and flexibility. Most instrument manufacturers continually strive to develop instruments that offer a good price/performance ratio, and that minimize some of the inherent limitations of a given procedure.

The function of the spectrometer is to separate the polychromatic beam of radiation coming from the specimen so that the intensities of each individual

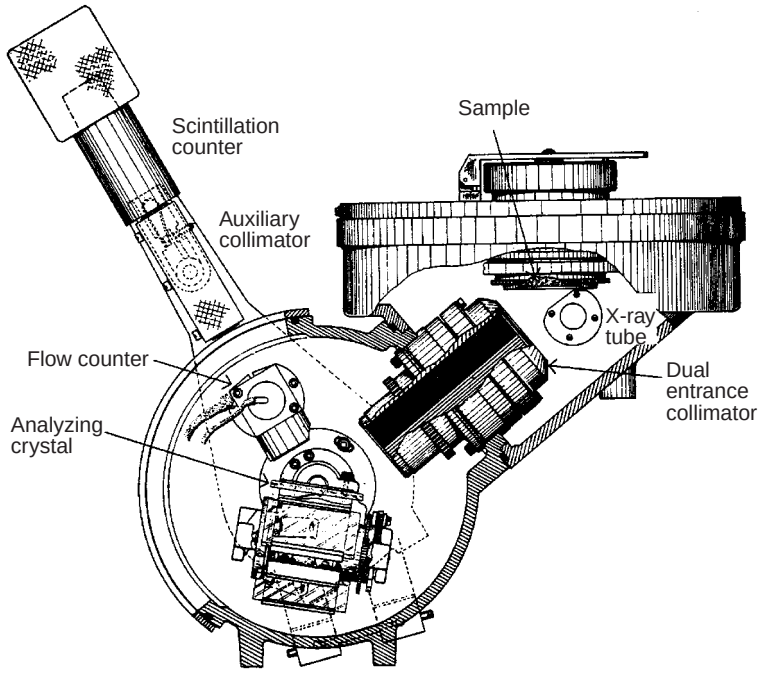

Figure 2 The wavelength-dispersive spectrometer. (Reproduced from Jenkins[7] by permission of John Wiley & Sons, Ltd.)

characteristic line can be measured. A spectrometer should provide sufficient resolution of lines to allow such data to be collected, at the same time providing a sufficiently large response above background to make the measurements statistically significant, especially at low analyte concentration levels. It is also necessary that the spectrometer allow measurements over the wavelength range to be covered. Thus, in the selection of a set of spectrometer operating variables, four factors are important: resolution, response, background level and range. Due to many factors, optimum selection of some of these characteristics may be mutually exclusive as, for example, attempts to improve resolution invariably cause lowering of absolute peak intensities.

The wavelength-dispersive spectrometer may be a single-channel instrument in which a single crystal and a single detector are used for the sequential measurement of a series of wavelengths, or a multichannel spectrometer in which many crystal/detector sets are used to measure many elements simultaneously. In the typical wavelength-dispersive spectrometer geometry, a single crystal of known interplanar spacing is used to disperse the collimated polychromatic beam of characteristic wavelengths coming from the sample, such that each wavelength will diffract at a discrete angle. As was shown in **X-ray Techniques: Overview**, there is a simple relationship between the interplanar spacing of the crystal, the diffracted wavelength and the diffraction angle. Since the maximum achievable angle on a typical wavelength-dispersive spectrometer is around 73 °θ, the maximum wavelength that can be diffracted by a crystal of spacing $2d$ is equal to about $1.9d$.

5.3 Resolution of the Spectrometer

Many factors are significant in the design of an X-ray spectrometer, and not the least of these is the ability to disperse and separate the polychromatic beam of radiation coming from the excited specimen. In terms of the separating power of a crystal spectrometer this will be dependent upon two factors – the divergence allowed by the collimators (that to a first approximation determines the width of the diffracted lines) and the rocking angle of the crystal. Figure 2 shows a diagrammatic representation of the optical path of the spectrometer and shows radiation paths coming from the specimen, passing through a primary collimator, onto the surface of a crystal which diffracts over a range of $\pm\delta\theta$. The radiation then passes through a secondary collimator to the detector. The reason why the crystal diffracts over a range of angles, as opposed to one discrete Bragg angle, is due to the mosaic nature of the crystal. In order to avoid problems such as secondary extinction, the surface of the single crystal analyzer is abraded. This process results in some broadening of the half-height width B_{cr} of the diffracted beam. The divergence of the collimator B_{cc} is determined by the length of the collimator blades and the spacing between the blades. The experimental breadth (B_e) of a diffracted line in a crystal spectrometer can be

expressed by Equation (8):

$$B_e = \sqrt{B_{cc}^2 + B_{cr}^2} \quad (8)$$

$\Delta\lambda/\lambda$ is a function of B_e and the angle at which the line is diffracted, as shown in Equation (9):

$$\frac{\Delta\lambda}{\lambda} = \frac{B_e}{\tan\theta} \quad (9)$$

Typical values for B_e range from 0.1° for 2θ at low angles to 0.5° at high angles, giving $\Delta\lambda/\lambda$ values in the range 0.002–0.02. In terms of energy, this corresponds to approximately 10–100 eV.

From the above, it will be clear that there is about one order of magnitude variation in the resolution of a wavelength-dispersive spectrometer over the usual wavelength range, depending upon the selection of crystals and collimators. Since mechanical limitations prevent wide selectability of line shape just by selection of collimator divergence, in practice, the resolution of the spectrometer will mainly be a function of the angular dispersion of the analyzing crystal, albeit, with some influence of the breadth of the diffracted line profile. It will be seen from Equation (3) that the angular dispersion will be high when the d-spacing is small. This is unfortunate as far as the range of the spectrometer is concerned because a small value of $2d$ means in turn a small range of wavelengths coverable. Thus, as with the resolution and peak intensities, high dispersion can only be obtained at the expense of cutting down the wavelength range covered by a particular crystal. In order to circumvent this problem, it is likely that several analyzing crystals will be employed in the coverage of a number of analyte elements. Many different analyzing crystals are available,[6] each having its own special characteristics, but three or four crystals will generally suffice for most applications.

The output from a wavelength-dispersive spectrometer may be either analog or digital. Pulses from the detector are amplified and passed via the pulse height selector on to one of two circuits. For qualitative work an analog output is traditionally used and in this instance a rate meter is used to integrate the pulses over short time intervals, typically of the order of a second or so. The output from the rate meter is fed to an x/t recorder that scans at a speed that is synchronously coupled with the goniometer scan speed. The recorder thus displays an intensity/time diagram, which becomes an intensity/2θ diagram. Tables are then used to interpret the wavelengths. For quantitative work it is more convenient to employ digital counting and a timer/scaler combination is provided that will allow pulses to be integrated over a period of several tens of seconds and then displayed as count or count rate. In more modern spectrometers a scaler/timer may also take the place of the rate meter, using a process known as step-scanning. In this instance the contents of the scaler are displayed on the x-axis of the x/t recorder as a voltage level. The scaler/timer is then reset and then started for a selected time interval. At the end of this time the timer sends a stop pulse to the scaler that now holds a number of counts equal to the product of the counting rate and the count time. The contents of the scaler are then displayed as before, the goniometer stepped to its next position and the whole cycle repeated. Generally the process is completely controlled by a microprocessor or a personal computer.

5.4 Spectrometers for X-ray Fluorescence

Simultaneous wavelength-dispersive spectrometers were introduced in the early 1950s, and sequential systems about a decade later. At this time, about 30 000 or so wavelength-dispersive instruments have been supplied commercially, roughly one-third of these in the USA. Energy-dispersive spectrometers became commercially available in the early 1970s and today there are of the order of 20 000 units in use. The two major categories of wavelength-dispersive X-ray spectrometers differ mainly in the type of source used for excitation, the number of elements that they are able to measure at one time, the speed at which they collect data and their price range. For high specimen throughput quantitative analysis where speed is of the essence, and where high initial cost can be justified, simultaneous wavelength-dispersive spectrometers are probably the best. For more flexibility, where speed is important but not critical and where moderately high initial cost can be justified, sequential wavelength-dispersive spectrometers are probably the best. Both of the instruments are, in principle at least, capable of measuring all elements in the periodic classification from $Z = 9$ (F) and upwards, and most modern wavelength-dispersive spectrometers can do some useful measurements down to $Z = 6$ (C). Both can be fitted with multisample handling facilities and both can be automated by use of personal computers. Both are capable of precision of the order of a few tenths of 1% and both have sensitivities down to the ppm level. Single-channel wavelength-dispersive spectrometers are typically employed for both routine and nonroutine analysis of a wide range of products, including ferrous and nonferrous alloys, oils, slags and sinters, ores and minerals and thin films. These systems are very flexible but, relative to multichannel spectrometers, are somewhat slow. The multichannel wavelength-dispersive instruments are used almost exclusively for routine, high throughput analyses where there is need of fast accurate analysis, but where flexibility is of no importance.

While the various types of specialized X-ray spectrometers are not generally available to the average

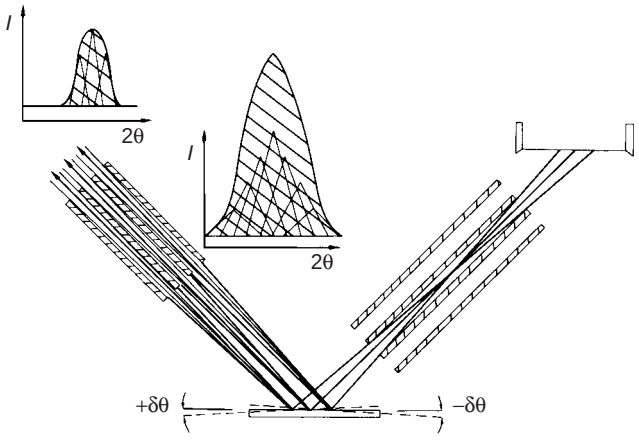

Figure 3 The collimation system in a scanning wavelength-dispersive spectrometer. (Reproduced from Jenkins[7] by permission of John Wiley & Sons, Ltd.)

user, they do have important roles to play in special areas of application. Included within this category are total reflection X-ray fluorescence (TRXRF) spectrometers and synchrotron source X-ray fluorescence (SSXRF) spectrometers. Two things that these systems have in common are a very high sensitivity and ability to work with extremely low concentrations and/or small specimens. The TRXRF system makes use of the fact that at very low glancing angles, primary X-ray photons are almost completely absorbed within thin specimens and the high background that would generally occur due to scatter from the sample support is absent. The recent development of high intensity synchrotron radiation beams has led to interest in their application for X-ray fluorescence analysis. The high intensity available in synchrotron beams allows use of very narrow band path monochromators between source and specimen, giving, in turn, a high degree of selective excitation. A compromise that must always be made in the design and set-up of any spectrometer is that between intensity and resolution, resolution being defined as the ability of the spectrometer to separate lines. In the flat-crystal wavelength-dispersive system this resolution is dependent upon the angular dispersion of the analyzing crystal and the divergence allowed by the collimators (Figure 3).[7]

6 WAVELENGTH-DISPERSIVE SPECTROMETER

Wavelength-dispersive spectrometers employ diffraction by a single crystal to separate characteristic wavelengths emitted by the sample. Energy-dispersive spectrometers use the proportional characteristics of a photon detector, typically lithium-drifted silicon, to separate the characteristic photons in terms of their energies. Since there is a simple relationship between wavelength and energy, these techniques each provide the same basic type of information, and the characteristics of the two methods differ mainly in their relative sensitivities and the way in which data are collected and presented. Generally speaking the wavelength-dispersive system is roughly one to two orders of magnitude more sensitive than the energy-dispersive system. Against this, however, the energy-dispersive spectrometer measures all elements within its range at essentially the same time, whereas the wavelength-dispersive system identifies only those elements for which it is programmed. To this extent, the energy-dispersive system is more useful in recognizing unexpected elements. Both wavelength- and energy-dispersive spectrometers typically employ a primary X-ray photon source operating at 0.5 to 3 kW. A disadvantage with this type of source is that the specimen scatters the white radiation from the source leading to significant background levels that tend to become one of the major limitations in the determination of low concentration levels. Typical analysis times vary from about 10 s to 3 min per element. The minimum sample size required is of the order of a few milligrams, although typical sample sizes are probably around several grams. Good accuracy is obtainable and in favorable cases standard deviations of the order of a few tenths of 1% are possible. This is because the matrix effects in X-ray spectrometry are well understood and relatively easy to overcome. The sensitivity is fair and determinations down to the low ppm level are possible for most elements.

6.1 Sequential Spectrometers

A sequential spectrometer, often referred to as a scanning spectrometer, uses a single crystal and a single detector. A simultaneous spectrometer contains many crystal/detector sets which are used to measure elements simultaneously. Of these two basic types, the sequential systems are the most common. A typical sequential spectrometer system consists of the X-ray tube, a specimen holder support, a primary collimator, an analyzing crystal and a tandem detector. The gas flow counter is ideal for the measurement of the longer wavelengths and the scintillation counter is best for the short wavelengths. A portion of the characteristic fluorescence radiation from the specimen is passed via a collimator or slit onto the surface of an analyzing crystal, where individual wavelengths are diffracted to the detector in accordance with the Bragg law. A goniometer is used to maintain the required θ to 2θ relationship between crystal and detector. Typically, six or so different analyzing crystals and two different collimators are provided in this type of spectrometer,

giving the operator a wide range of choice of dispersion conditions. In general, the smaller the $2d$-spacing of the crystal, the better the separation of the lines but the smaller the wavelength range that can be covered. Since the maximum achievable angle θ on a wavelength-dispersive spectrometer is generally around 73°, the maximum wavelength that can be diffracted by a crystal of spacing $2d$ is equal to about $1.9d$. The separating power of a crystal spectrometer is dependent upon the divergence allowed by the collimators (which to a first approximation determines the width of the diffracted lines) and the angular dispersion of the crystal. Since mechanical limitations prevent wide selectability of line shape just by selection of collimator divergence, in practice, the resolution of the spectrometer is typically determined by the angular dispersion of the analyzing crystal, albeit with some influence of the breadth of the diffracted line profile.

6.2 Simultaneous Spectrometers

The simultaneous spectrometer is generally nothing more than a group (6–30) of fixed-channel spectrometers arranged round the irradiated specimen. This arrangement allows one to tailor-make each channel, just for the element for which it is being set, so, in principal, if one had a 30-channel simultaneous spectrometer, every crystal/collimator/slit/detector combination could be different for each channel. It is common to use a curved crystal arrangement, as shown in Figure 4, since this is more efficient and requires only a very small detector window. The curved crystals may be simply bent, or bent and ground, or even may be formed as a section of a logarithmic spiral. The simultaneous spectrometer offers great speed and high specimen throughput – a sample containing 24 elements can be analyzed in less than 1 min. Against this, however, there is a loss in flexibility in comparison with the sequential spectrometer, and the equipment cost is higher – perhaps by a factor of two. Some flexibility can be regained by use of one or more scanning channels. These scanning channels are, in essence, the same as a conventional optical system in the sequential spectrometer.

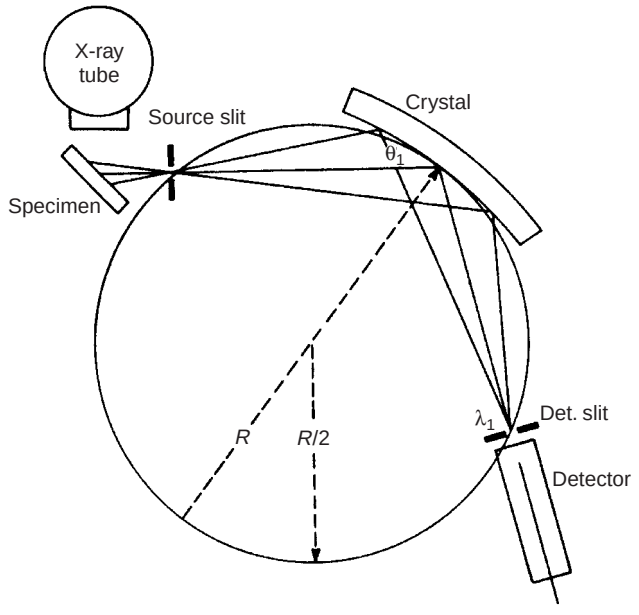

Figure 4 A curved crystal spectrometer. (Reproduced by permission of Plenum Press.)

6.3 Other Types of X-ray Spectrometer

*6.3.1 Energy-dispersive Systems (see **Portable Systems for Energy-dispersive X-ray Fluorescence**)*

The energy-dispersive spectrometer consists of the excitation source and the spectrometer/detection system. The spectrometer/detector is typically an Si(Li) detector, which is a proportional detector of high intrinsic resolution. A multichannel analyzer is used to collect, integrate and display the resolved pulses. While similar properties are sought from the energy-dispersive system as with the wavelength-dispersive system, the means of selecting these optimum conditions are very different. Since the resolution of the energy-dispersive system is equated directly to the resolution of the detector, this feature is of paramount importance. The output from an energy-dispersive spectrometer is generally displayed on a cathode ray tube and the operator is able to dynamically display the contents of the various channels as an energy spectrum and provision is generally made to allow zooming in on portions of the spectrum of special interest, to overlay spectra, to subtract background, and so on. Even though the Si(Li) detector is the most common detector used in energy-dispersive X-ray spectrometry, it is certainly not the only one. As an example, the higher absorbing power of germanium makes it an alternative for the measurement of high-energy spectra. Both cadmium telluride (CdTe) and mercuric iodide (HgI_2) show some promise as detectors capable of operating satisfactorily at room temperature.

6.3.2 On-stream Analyzers

The use of conventional and nonconventional wavelength-dispersive X-ray spectrometers for on-line process control has long been recognized, and currently there are several hundreds of these units in use. The major use is found in mineral beneficiation processes, although some units are used for the control of cement kilns. Essentially three major areas of application exist:

1. Analysis of flowing slurries found in most ore dressing and mineral beneficiation plants.
2. Analysis of flowing dry solids, for example, crushed ores on conveyor belts, or dry raw cement mix en route to the kiln.
3. Analysis of drill cores and insides of bore holes.

The first category is typically covered by a conventional wavelength-dispersive spectrometer, generally in a multichannel configuration, and using a specially designed flow cell in place of the regular fixed sample holder. A complex plumbing system is then required to effectively sample the flowing slurry, and inject it continuously into the flow cell. Because of the effect of absorption by the (thick) flow cell windows, and the necessary air path in the system, it is almost impossible to use this technique to determine elements below atomic number 22 (Ti). The second category can be handled by either continuous analysis or by successive sampling. Special care must be made in continuous analysis because of variations in particle size of the feed. An advantage of the successive sampling technique is that automated samplers, pellet presses and borax bead makers can all be employed to provide a rapid and reliable analysis for most of the atomic number range. The third category is covered mainly by the use of radioisotope-based systems. The major disadvantage of any on-stream analyzer based on a source/multichannel, crystal dispersion system is its high inherent cost and physical size, not to mention the amount of specimen to be analyzed resting in the plumbing required to bring the specimen to the analysis chamber. Many of these problems can be resolved by use of a radioisotope-based system, which can be small, compact, and relatively inexpensive.[8] Am^{241}, Cd^{109}, and Gd^{153} have all been successfully employed as sources. The great disadvantage of the radioisotope source is that its total photon yield is small in comparison to conventional sealed-source X-ray tubes. In order to recover some of this inefficiency, it is common to use energy discrimination, rather than the classic dispersive crystal arrangement, in radioisotope-based systems.

6.3.3 Total Reflection Spectrometers

One of the major problems that inhibits obtaining good detection limits in small samples is the high background due to scatter from the sample substrate support material. The suggestion to overcome this problem by using total reflection of the primary beam was made long ago, but the absence of suitable instrumentation prevented real progress being made until the late 1970s. Good sample preparation and presentation procedures are now available, making TRXRF a valuable technique for trace analysis in which the Si(Li) detector is placed close to (about 5 mm), and directly above, the sample.[9] Primary radiation enters the sample at a glancing angle of a few seconds of arc. The sample itself is typically presented as a thin film on the optical flat surface of a quartz plate. A typical instrument contains a series of reflectors to aid in the reduction of background. A beam of radiation from a sealed X-ray tube passes through a fixed aperture onto a pair of reflectors that are placed very close to each other. Scattered radiation passes through the first aperture to impinge on the sample at a very low glancing angle. Because the primary radiation enters the sample at an angle barely less than the critical angle for total reflection, this radiation barely penetrates the substrate media; thus scatter and fluorescence from the substrate are minimal. Because the background is so low, picogram amounts can be measured or concentrations in the range of a few tenths of a part per billion can be obtained without recourse to preconcentration.

6.3.4 Use of Synchrotron Source Radiation

The availability of intense, linearly polarized synchrotron radiation beams[10] has prompted workers in the fields of X-ray fluorescence[11] to explore what the source has to offer over more conventional excitation media. In the synchrotron, electrons with kinetic energies of the order of several billion electron-volts (typically 3 GeV), orbit in a high vacuum tube between the poles of a strong (about 10^4 G) magnet. A vertical field accelerates the electrons horizontally, causing the emission of synchrotron radiation. Thus synchrotron source radiation can be considered as magnetic bremsstrahlung in contrast to normal electronic bremsstrahlung produced when electrons are decelerated by the electrons of an atom. It has been found that because the primary source of radiation is so intense, it is possible to use a high degree of monochromatization between source and specimen, giving a source that is wavelength (and, therefore, energy) tunable, as well as being highly monochromatic. The intensity of the synchrotron beam is probably four to five orders of magnitude greater than the conventional bremsstrahlung source-sealed X-ray tubes. This, in combination with its energy tunability and polarization in the plane of the synchrotron ring, allows very rapid bulk analysis to be obtained on small areas. Because the synchrotron beam has such a high intensity and small divergence, it is possible to use it as a microprobe of high spatial resolution (about 10 μm). Absolute limits of detection around 10–14 μg have been reported using such an arrangement.

A good example of the power of synchrotron source radiation is found in the study of the thin film superconductors which are required for applications of high-temperature superconductors (HTSCs) in microelectronics technology. An example of such work[12]

involves the analysis of HTSC films produced by laser evaporation of elements in the Y–Ba–Cu–O system. Ten films were simultaneously deposited with various target-to-substrate distances allowing study to be made of the laser plasma expansion in vacuo. A very important advantage of this method is that it allows extremely low detection limits to be achieved. As an example, in the referenced study the authors claim a detection limit of 5×10^{13} atoms cm^{-2}. Additional advantages accrue because synchrotron radiation is highly polarized and background due to scatter can be greatly reduced by placing the detector at 90° to the path of the incident beam and in the plane of polarization.

7 QUALITATIVE ANALYSIS WITH THE X-RAY SPECTROMETER

Both the simultaneous (scanning) wavelength-dispersive spectrometer and the energy-dispersive spectrometers (see **Portable Systems for Energy-dispersive X-ray Fluorescence**) lend themselves admirably to the qualitative analysis of materials, since there is a simple relationship between the wavelength or energy of a characteristic X-ray photon and the atomic number of the element from which the characteristic emission line occurs. Thus by measuring the wavelengths, or energies, of a given series of lines from an unknown material, the atomic numbers of the excited elements can be established. The inherent simplicity of characteristic X-ray spectra makes the process of allocating atomic numbers to the emission lines relatively easy, and the chance of making a gross error is rather small. There are only 100 or so elements, and within the range of the conventional spectrometer each element gives, on average, only half a dozen lines. If one compares the X-ray emission spectrum with the ultraviolet (UV) emission spectrum, since the X-ray spectrum arises from a limited number of inner orbital transitions, the number of X-ray lines is similarly rather few. UV spectra, however, arise from transitions to empty levels, of which there may be many, leading to a significant number of lines in the UV emission spectrum. A further benefit of the X-ray emission spectrum for qualitative analysis is that because transitions do arise from inner orbitals, the effect of chemical combination or valence state is almost negligible.

Qualitative analysis has traditionally been performed by scanning the goniometer synchronously at a fixed angular speed. However, most scanning spectrometers are slow in this sequential angular/intensity data collection mode, not only because the data are taken sequentially, but also because in order to cover the full range of elements a series of scans must be made with different conditions. In addition to this, scanning at a fixed speed is somewhat inefficient because in the crystal dispersive system the atomic number varies as a function of one over the square root of the angle. In effect this means that in the low atomic number region much scanning time is wasted in scanning angular space that contains no characteristic line data. Although some unique designs have brought about some reduction in data acquisition times, this remains a major limitation of wavelength-dispersive spectrometers. Data interpretation involves identifying each line in the measured spectrogram. However, since the relationship between atomic number and Bragg angle is rather complicated, it is common practice to use sets of tables for interpretation relating wavelength and atomic number with diffraction angle for specific analyzing crystals. Some automated wavelength-dispersive spectrometers provide the user with software programs for the interpretation and labeling of peaks.

8 ACCURACY OF X-RAY FLUORESCENCE

Table 5 lists the four main categories of random and systematic error encountered in X-ray fluorescence analysis. The first category includes the selection and preparation of the sample to be analyzed. Two stages are generally involved before the actual prepared specimen is presented to the spectrometer, these being sampling and specimen preparation. The actual sampling is rarely under the control of the spectroscopist and it generally has to be assumed that the container containing the material for analysis does, in fact, contain a representative sample. It will be seen from the table that, in addition to a relatively large random error, inadequate sample preparation and residual sample heterogeneity can lead to very large systematic errors. For accurate analysis these errors must be reduced by use of a suitable specimen preparation method. The second category includes errors arising from the X-ray source previously discussed. Source errors can be reduced to less than 0.1% by use of the

Table 5 Sources of error in X-ray fluorescence analysis

	Source	Random (%)	Systematic (%)
1	Sample preparation	0–1	0–5
	Sample inhomogeneity	–	0–50
2	Excitation source	0.05–0.2	0.05–0.5
	Spectrometer	0.05–0.1	0.05–0.1
3	Counting statistics	Time dependent	–
	Dead time	–	0–25
4	Primary absorption	–	0–50
	Secondary absorption	–	0–25
	Enhancement	–	0–15
	Third element	–	0–2

ratio counting technique, provided that high frequency transients are absent. The third category involves the actual counting process and these errors can be both random and systematic. System errors due to detector dead time can be corrected either by use of electronic dead time correctors or by some mathematical approach. The fourth category includes all errors arising from inter-element effects. Each of the effects listed can give large systematic errors that must be controlled by the calibration and correction scheme.

8.1 Counting Statistical Errors

The production of X-rays is a random process that can be described by a Gaussian distribution. Since the number of photons counted is nearly always large, typically thousands of hundreds of thousands, rather than a few hundred, the properties of the Gaussian distribution can be used to predict the probable error for a given count measurement. There will be a random error $\sigma(N)$ associated with a measured value of N, this being equal to $\sqrt{N}$. As an example, if 10^6 counts are taken, the 1σ standard deviation will be $\sqrt{10^6} = 10^3$, or 0.1%. The measured parameter in wavelength-dispersive X-ray spectrometry is generally the counting rate R and, based on what has been already stated, the magnitude of the random counting error associated with a given datum can be expressed as Equation (10):

$$\sigma(\%) = \frac{100}{\sqrt{N}} = \frac{100}{\sqrt{Rt}} \tag{10}$$

Care must be exercised in relating the counting error (or indeed any intensity related error) with an estimate of the error in terms of concentration. Provided that the sensitivity of the spectrometer in counts per second per percent, is linear, a count error can be directly related to a concentration error. However, where the sensitivity of the spectrometer changes over the range of measured response, a given fractional count error may be much greater when expressed in terms of concentration.

In many analytical situations the peak lies above a significant background and this adds a further complication to the counting statistical error. An additional factor that must also be considered is that whereas with the scanning wavelength-dispersive spectrometer the peaks and background are measured sequentially, in the case of the energy-dispersive and the multichannel wavelength-dispersive spectrometers, a single counting time is selected for the complete experiment, thus all peaks and all backgrounds are counted for the same time. To estimate the net counting error in the case of a sequential wavelength-dispersive spectrometer it is necessary to consider the counting error of the net response of peak counting rate R_p, and background counting rate R_b, since the analyte element is only responsible for $(R_p - R_b)$. Equation (10) must then be expanded to include the background count rate term to give Equation (11):

$$\sigma(R_p - R_b) = \frac{100}{\sqrt{t}} \frac{1}{\sqrt{R_p} - \sqrt{R_p}} \tag{11}$$

One of the conditions for Equation (11) is that the total counting time t must be correctly proportioned between time spent counting on the peak t_p, and time spent counting on the background t_b, as shown in Equation (12):

$$\frac{t_p}{t_b} = \frac{\overline{R_p}}{R_b} \tag{12}$$

8.2 Matrix Effects

In the conversion of net line intensity to analyte concentration it is necessary to correct for any absorption and/or enhancement effects which occur. Absorption effects include both primary and secondary absorption, and enhancement effects include direct enhancement, involving the analyte element and one enhancing element, plus third element effects that involve additional element(s) beyond the analyte and enhancer. Primary absorption occurs because all atoms of the specimen matrix will absorb photons from the primary source. Since there is competition for these primary photons by the atoms making up the specimen, the intensity/wavelength distribution of these photons available for the excitation of a given analyte element may be modified by other matrix elements. Secondary absorption refers to the effect of the absorption of characteristic analyte radiation by the specimen matrix. As characteristic radiation passes out from the specimen in which it was generated, it will be absorbed by all matrix elements, by amounts relative to the mass absorption coefficients of these elements. The mass absorption coefficient is a parameter which defines the magnitude of the absorption of a certain element for a specific X-ray wavelength. The total absorption of a specimen is dependent on both primary and secondary absorption. The total absorption by element i for an analyte wavelength λ_j is given by Equation (13):

$$\alpha_i = \mu_i(\lambda) + A[\mu_i(\lambda_j)] \tag{13}$$

The factor A is a geometric constant equal to the ratio of the sines of the incident and take-off angles of the spectrometer. This factor is needed to correct for the fact that the incident and emergent rays from the sample have different pathlengths. The term λ in the equation refers to the primary radiation. Since most conventional X-ray spectrometers use a bremsstrahlung source, in

practice λ is a range of wavelengths, although in simple calculations it may be acceptable to use a single equivalent wavelength value where the equivalent wavelength is defined as a single wavelength having the same excitation characteristics as the full continuum.

There are a number of routes by which the analyte element can be excited or enhanced and these are illustrated in Figure 5(a) which shows the direct excitation of an analyte element *i* by the primary continuum P_1. There may also be excitation of the analyte by characteristic lines from the source (P_2 in Figure 5b). Both the continuous and characteristic radiation from the source may be somewhat modified by Compton scatter and the excitation by this modified source radiation is indicated in Figure 5(c) by P_3. Enhancement effects (Figure 5d) occur when a nonanalyte matrix element A emits a characteristic line that has an energy just in excess of the absorption edge of the analyte element. This means that the nonanalyte element in question is able to excite the analyte, giving characteristic photons over and above those produced by the primary continuum. This gives an increased, or enhanced, signal from the analyte. The third element effect is also shown. Here, a third element B is also excited by the source. Not only can B directly enhance *i*, but it can also enhance A, thus increasing the enhancing effect of A on *i*. This last effect is called the third element effect.

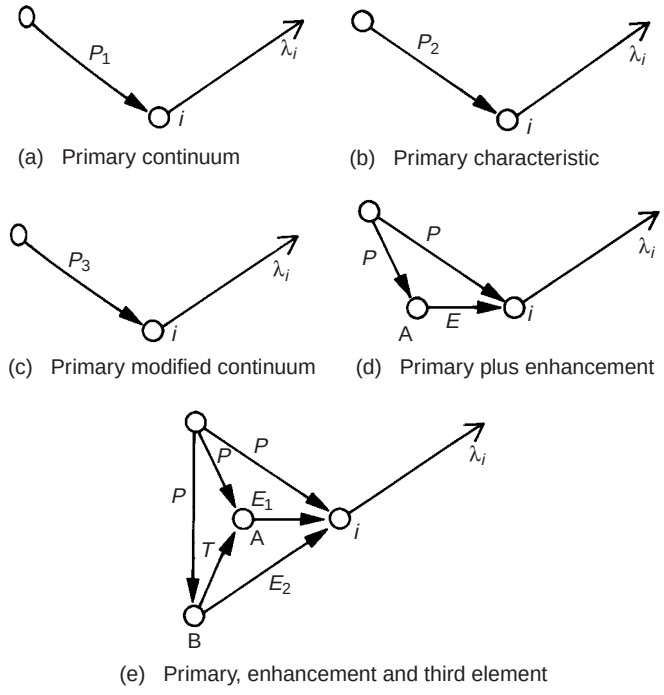

Figure 5 Various routes by which characteristic analyte radiation can be excited. (Reproduced from R. Jenkins, *X-ray Fluorescence Spectrometry*, 2nd edition, John Wiley & Sons, New York, 1999, by permission of John Wiley & Sons, Ltd.)

8.3 Specimen Preparation for X-ray Fluorescence

Because X-ray spectrometry is essentially a comparative method of analysis, it is vital that all standards and unknowns be presented to the spectrometer in a reproducible and identical manner. Any method of specimen preparation must give specimens which are reproducible and which, for a certain calibration range, have similar physical properties including mass attenuation coefficient, density, particle size, and particle homogeneity. In addition the specimen preparation method must be rapid and cheap and must not introduce extra significant systematic errors, for example, the introduction of trace elements from contaminants in a diluent. Specimen preparation is an all important factor in the ultimate accuracy of any X-ray determination, and many papers have been published describing a multitude of methods and recipes for sample handling.[13]

In general samples fit into three main categories:

1. Samples which can be handled directly following some simple pretreatment such as pelletizing or surfacing. For example, homogeneous samples of powders, bulk metals or liquids.
2. Samples which require significant pretreatment. For example, heterogeneous samples, samples requiring matrix dilution to overcome inter-element effects and samples exhibiting particle size effects.
3. Samples which require special handling treatment. For example, samples of limited size, samples requiring concentration or prior separation and radioactive samples.

The ideal specimen for X-ray fluorescence analysis is one in which the analyzed volume of specimen is representative of the total specimen, which is, itself, representative of the sample submitted for analysis. There are many forms of specimen suitable for X-ray fluorescence analysis, and the form of the sample as received will generally determine the method of pretreatment. It is convenient to refer to the material received for analysis as the sample, and that which is actually analyzed in the spectrometer as the specimen. While the direct analysis of certain materials is certainly possible, more often than not some pretreatment is required to convert the sample to the specimen. This step is referred to as specimen preparation. In general, the analyst would prefer to analyze the sample directly, because if it is taken as received, any problems arising from sample contamination that might occur during pretreatment are avoided. In practice, however, there are three major constraints that may prevent this ideal circumstance from being achieved: sample

size, sample size homogeneity and sample composition heterogeneity.

Problems of sample size are frequently severe in the case of bulk materials such as metals, large pieces of rock, etc. Problems of sample composition heterogeneity will generally occur under these circumstances as well, and in the analysis of powdered materials heterogeneity must almost always be considered. The sample as received may be either homogeneous or heterogeneous; in the latter case, it may be necessary to render the sample homogeneous before an analysis can be made. Heterogeneous bulk solids are generally the most difficult kind of sample to handle, and it may be necessary to dissolve or chemically react the material in some way to give a homogeneous preparation. Heterogeneous powders are either ground to a fine particle size and then pelletized, or fused with a glass-forming material such as borax. Solid material in liquids or gases must be filtered out and the filter analyzed as a solid. Where analyte concentrations in liquids or solutions are too high or too low, dilution or preconcentration techniques may be employed to bring the analyte concentration within an acceptable range.

9 QUANTITATIVE ANALYSIS

In the X-ray analytical laboratory the quantitative method of analysis employed will be typically predicated by a number of circumstances of which probably the four most common are: the complexity of the analytical problem; the time allowable; the computational facilities at the disposal of the analyst; and the number of standards available. It is convenient to break quantitative analytical methods down into two major categories: single-element methods and multiple-element methods, as shown in Table 6. The simplest quantitative analysis situation to handle is the determination of a single element in a known matrix. A slightly more difficult case might be the determination of a single element where the matrix is unknown. As shown in the table, three basic methods are commonly employed in this situation: use of internal standards, use of standard addition, or use of a scattered line from the X-ray source. The most complex case is the analysis of all, or most, of the elements in a sample about which little or nothing is known. In this case a full qualitative analysis would be required before any attempt is made to quantitate the matrix elements. Once the qualitative composition of the sample is known, again, one of three general techniques is typically applied: use of type standardization, use of an influence coefficient method, or use of a fundamental parameter technique. Both the influence coefficient and fundamental parameter technique require a computer for their application.

Table 6 Quantitative procedures employed in X-ray fluorescence analysis

Single-element methods	Internal standardization
	Standard addition
	Use of scattered source radiation
Multiple-element methods	Type standardization
	Use of influence coefficients
	Fundamental parameter techniques

One of the problems with any X-ray spectrometer system is that the absolute sensitivity (i.e. the measured counts per second per percent) decreases dramatically as the lower atomic number region is approached (Figure 6). The three main reasons for this decrease are the reduction in the fluorescence yield with decrease in atomic number, the decrease in the absolute number of useful long wavelength X-ray photons from a bremsstrahlung source with increase of wavelength and, thirdly, the increase of absorption effects, generally with increase of the wavelength of the analyte line. The first two of these problems are inherent to the X-ray excitation process and to constraints in the basic design of conventional X-ray tubes. The third, however, is a factor that depends very much on the instrument design, and, in particular, upon the characteristics of the detector. The detector that is used in long wavelength spectrometers is typically a gas flow proportional counter, in which an extremely thin, high-transmission window is employed.

The great flexibility, sensitivity and range of the various types of X-ray fluorescence spectrometer make them ideal for quantitative analysis. In common with

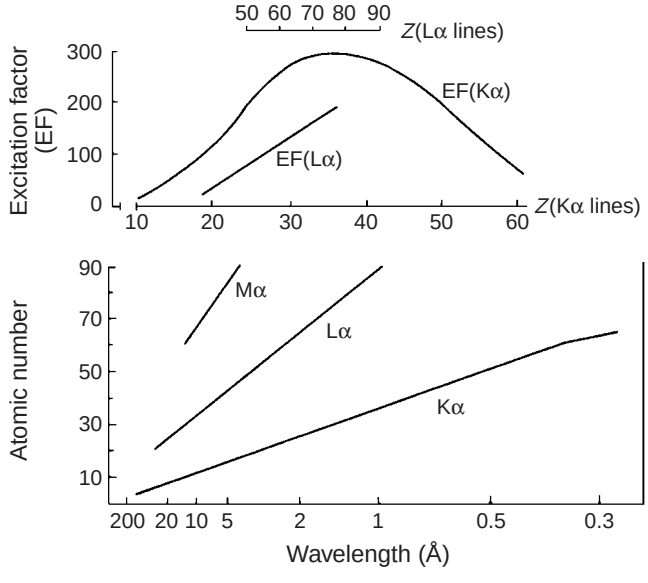

Figure 6 Sensitivity of the X-ray spectrometer. (Reproduced by permission of Marcel Dekker Inc.)

all analytical methods, quantitative X-ray fluorescence analysis is subject to a number of random and systematic errors that contribute to the final accuracy of the analytical result. Like all instrumental methods of analysis, the potential high precision of X-ray spectrometry can only be translated into high accuracy if the various systematic errors in the analysis process are taken care of. The precision of a wavelength-dispersive system for the measurement of a single, well separated line is typically of the order of 0.1%, and about 0.25% for the energy-dispersive system. A good rule-of-thumb which can be used in X-ray fluorescence analysis to estimate the expected standard deviation σ at an analyte concentration level C is given by Equation (14)

$$\sigma = K\sqrt{C + 0.1} \tag{14}$$

where K varies between 0.005 and 0.05. For example, at a concentration level $C = 25\%$, the expected value of σ would be between about 0.025% and 0.25%. A K value of 0.005 would be considered very high quality analysis and a value of 0.05 rather poor quality. The value of K actually obtained under routine laboratory conditions depends upon many factors but with reasonably careful measurements a K value of around 0.02 to 0.03 can be obtained. The correlation between the characteristic line intensity of an analyte element and the concentration of that element is typically nonlinear over wide ranges of concentration, due to inter-element effects between the analyte element and other elements making up the specimen matrix. However, the situation can be greatly simplified in the case of homogeneous specimens, where severe enhancement effects are absent, and here, the slope of a calibration curve is inversely proportional to the total absorption μ of the specimen for the analyte wavelength. In this instance the slope K of the calibration curve is taken as W/I, where I is the line intensity and W the weight fraction of the analyte element. Thus, Equation (15) holds:

$$W = \frac{I}{K\mu} \tag{15}$$

9.1 Single-element Techniques

Single-element techniques reduce the influence of the absorption term μ in Equation (15), generally by referring the intensity of the analyte wavelength to a similar wavelength, arising either from an added standard or from a scattered line from the X-ray tube. In certain cases, limiting the concentration range of the analyte may allow the assumption to be made that the absorption value does not significantly change over the concentration range and the calibration curve is essentially linear. This assumption is applied in the traditional type standardization technique.

9.2 Use of Internal Standards

One of the most useful techniques for the determination of a single analyte element in a known or unknown matrix is to use an internal standard. The technique is one of the oldest methods of quantitative analysis and is based on the addition of a known concentration of an element which gives a wavelength close to that of the analyte wavelength. The assumption is made that the effect of the matrix on the internal standard is essentially the same as the effect of the matrix on the analyte element. Internal standards are best suited to the measurements of analyte concentrations below about 10%. The reason for this limit arises because it is generally advisable to add the internal standard element at about the same concentration level as that of the analyte. When more than 10% of the internal standard is added, it may significantly change the specimen matrix and introduce errors into the determination. Care must also be taken to ensure that the particle sizes of specimen and internal standard are about the same, and that the two components are adequately mixed. Where an appropriate internal standard cannot be found it may be possible to use the analyte itself as an internal standard. This method is a special case of standard addition, and it is generally referred to as spiking.

9.3 Type Standardization

Provided that the total specimen absorption does not vary significantly over a range of analyte concentrations, and provided that enhancement effects are absent and that the specimen is homogeneous, a linear relationship will be obtained between analyte concentration and measured characteristic line intensity. Where these provisos are met, type standardization techniques can be employed. It will also be clear from previous discussion that by limiting the range of analyte concentration to be covered in a given calibration procedure, the range in absorption can also be reduced. Type standardization is probably the oldest of the quantitative analytical methods employed, and the method is usually evaluated by taking data from a well characterized set of standards, and, by inspection, establishing whether a linear relationship is indeed observed. Where this is not the case, the analyte concentration range may be further restricted. The analyst of today is fortunate in that many hundreds of good reference standards are commercially available. While the type standardization method is not without its pitfalls, it is nevertheless extremely useful and is especially useful for quality control type applications where a finished product is being compared with a desired product.

Special reference standards may be made up for particular purposes, and these may serve the dual purpose of instrument calibration as well as establishing working curves for analysis. As an example, two thin glass film standard reference materials (SRMs) specially designed for calibration of X-ray spectrometers are available from the National Institute of Standards and Technology in Washington as SRMs 1832 and 1833. They consist of a silica-based film deposited by focused ion-beam coating onto a polycarbonate substrate. SRM 1832 contains aluminum, silicon, calcium, vanadium, manganese, cobalt and copper, and SRM 1833 contains silicon, potassium, titanium, iron, zinc and rhodium. The standards are especially useful for the analysis of particulate matter.

9.4 Influence Correction Methods

It is useful to divide influence coefficient correction procedures into three basic types: fundamental, derived and regression. Fundamental models are those which require starting with concentrations, then calculating the intensities. Derived models are those which are based on some simplification of a fundamental method but which still allow concentrations to be calculated from intensities. Regression models are those which are semi-empirical in nature, and which allow the determination of influence coefficients by regression analysis of data sets obtained from standards. All regression models have essentially the same form and consist of a weight fraction term, W (or concentration C), an intensity (or intensity ratio) term, I, an instrument-dependent term which essentially defines the sensitivity of the spectrometer for the analyte in question, and a correction term which corrects the instrument sensitivity for the effect of the matrix. The general form is shown in Equation (16):

$$\frac{W}{R} = \text{constant} + [\text{model-dependent term}] \qquad (16)$$

where W is the weight fraction of the analyte and R is the ratio of the analyte intensity measured from a pure elemental standard to that of the analyte line intensity measured from the specimen, with each intensity corrected for instrumental effects.

The different models vary only in the form of this correction term. Table 7 shows several of the more important of the commonly employed influence coefficient methods. All of these models are concentration correction models in which the product of the influence coefficient and the concentration of the interfering element are used to correct the slope of the analyte calibration curve. The Lachance–Traill model[15] was the first of the concentration correction models to be published. Some years after the Lachance–Traill paper appeared, Heinrich et al. suggested an extension to the Lachance–Traill approach[16] in which absorbing and enhancing elements are separated as α- and β-terms. These authors suggested that the enhancing effect cannot be adequately described by the same hyperbolic function as the absorbing effect. A thorough study of Lachance–Traill coefficients based on theoretically calculated fluorescence intensities shows that all binary coefficients vary systematically with composition. Both the Claisse–Quintin[17] and Lachance–Claisse[18] models use higher order terms to correct for so-called crossed effects, which includes enhancement and third element effects. These models are generally more suited for very wide concentration range analysis.

Table 7 Influence correction models

Linear model:
$$\frac{W_i}{R_i} = K_i$$

Lachance–Traill (1966):
$$\frac{W_i}{R_i} = K_i + \sum_j a_{ij} W_j$$

Claisse–Quintin (1967):
$$\frac{W_i}{R_i} = K_i + \sum_j a_{ij} W_j + \sum_j \gamma_{ij} W_j^2$$

Rasberry–Heinrich (1974):
$$\frac{W_i}{R_i} = K_i + \sum_j a_{ij} W_j + \sum_{k \neq j} \beta_{ik} \left(\frac{W_k}{1 + W_i}$$

Lachance–Claisse (1980):
$$\frac{W_i}{R_i} = 1 + \sum_j a_{ij} W_j + \sum_j \sum_{k>j} a_{ijk} W_j W_k$$

9.5 Fundamental Methods

Since the early work of Sherman there has been a growing interest in the provision of an intensity/concentration algorithm which would allow the calculation of the concentration values without recourse to the use of standards. Sherman's work was improved upon first by the Japanese team of Shiraiwa and Fujino[19] and later by the Americans Criss and Birks[20] with their program NRLXRF. The same group also solved the problem of describing the intensity distribution from the X-ray tube. The problem for the average analyst in the late 1960s and early 1970s, however, remained that of finding sufficient computational power to apply these methods. In the early 1970s, de Jongh[21] suggested an elegant solution in which he proposed the use of a large main-frame computer for the calculation of the influence coefficients, then use of a small minicomputer for their actual application using a concentration correction influence model. One of the problem areas remains that of adequately describing the

intensity distribution from the X-ray tube. While software packages are available for fundamental type calculations using data obtained with the energy-dispersive system, one major drawback remains in their application system which uses a modified primary excitation spectrum. Most fundamental quantitative approaches in use today employ measured or calculated continuous radiation functions in the calculation of the primary absorption effect. Where sharp discontinuities or ''breaks'' in this primary spectrum occur, as in the case of the energy-dispersive system, the calculation becomes very complicated.

10 TRACE ANALYSIS

10.1 Analysis of Low Concentrations

The X-ray fluorescence method is particularly applicable to the qualitative and quantitative analysis of low concentrations of elements in a wide range of samples, as well as allowing the analysis of elements at higher concentrations in limited quantities of materials. The measured signal in X-ray analysis is a distribution of counting rate R as a function of either 2θ angle (wavelength-dispersive spectrometers) or as counts per channel as a function of energy (energy-dispersive spectrometers). A measurement of a line at peak position gives a counting rate which, in those cases where the background is insignificant, can be used as a measure of the analyte concentration. However, where the background is significant the measured value of the analyte line at the peak position now includes a count rate contribution from the background. The analyte concentration in this case is related to the net counting rate. Since both peak and background count rates are subject to statistical counting errors, the question now arises as to the point at which the net peak signal is statistically significant. The generally accepted definition for the lower limit of detection (LLD) is the concentration equivalent to two standard deviations of the background counting rate. A formula for the LLD can now be derived as shown in Equation (17):[22]

$$\mathrm{LLD} = \frac{3}{m}\sqrt{\frac{R_\mathrm{b}}{t_\mathrm{b}}} \tag{17}$$

Note that in Equation (17) t_b represents one half of the total counting time. The sensitivity m of the X-ray fluorescence method is expressed in terms of the intensity of the measured wavelength per unit concentration, expressed in counts per second per percent. Figure 6 showed the sensitivity (excitation factor) of a wavelength-dispersive spectrometer and indicates that the sensitivity varies by about four orders of magnitude over the measurable element range when expressed in terms of rate of change in response per rate of change in concentration. For a fixed analysis time the detection limit is proportional to $m/\sqrt{R_\mathrm{b}}$ and this is taken as a figure of merit for trace analysis. The value of m is determined mainly by the power loading of the source, the efficiency of the spectrometer for the appropriate wavelength and the fluorescent yield of the excited wavelength. The value of R_b is determined mainly by the scattering characteristics of the sample matrix and the intensity/wavelength distribution of the excitation source. From inspection of Equation (17) it will be seen that the detection limit will be best when the sensitivity is high and the background is low. Both the spectrometer sensitivity and the measured background vary with the average atomic number of the sample. While detection limits over most of the atomic number range lie in the low ppm range, the sensitivity of the X-ray spectrometer falls off quite dramatically towards the long wavelength limit of the spectrometer due mainly to low fluorescence yields and the increased influence of absorption. As a result, poorer detection limits are found at the long wavelength extreme of the spectrometer that corresponds to the lower atomic numbers. Thus the detection limits for elements such as fluorine and sodium are at the levels of hundredths of 1% rather than ppm. The detection limits for the very low atomic number elements such as carbon ($Z = 6$) and oxygen ($Z = 8$) are, however, very poor and are typically of the order of 3–5%.

10.2 Analysis of Small Amounts of Sample

Conventional wavelength-dispersive X-ray fluorescence spectrometers are generally designed to handle large specimens with surface areas of the order of several square centimeters, and problems occur where the sample to be analyzed is limited in size. The modern wavelength-dispersive system is especially inflexible in the area of sample handling, mainly because of geometric constraints arising from the need for close coupling of the sample to X-ray tube distance, and the need to use an airlock of some kind to bring the sample into the working vacuum. The sample to be analyzed is typically placed inside a cup of fixed external dimensions that is, in turn, placed in the carousel. This presentation system places constraints not only on the maximum dimensions of the sample cup, but also on the size and shape of samples that can be placed into the cup itself. Primary source energy-dispersive systems do not require the same high degree of focusing, and to this extent are more easily applicable to any sample shape or size, provided that the specimen will fit into the radiation-protected chamber. In some instances the spectrometer can even be brought to the object to be analyzed. Because of this flexibility,

the analysis of odd-shaped specimens has been almost exclusively the purview of the energy-dispersive systems. Where practicable, the best solution for the handling of limited amounts of material is invariably found in one of the specialized spectrometer systems. This is also generally true for the analysis of low concentrations. The TRXRF system is ideally suited for the analysis of small samples in those cases where the specimen can be dispersed as a thin film onto the surface of the reflector substrate. The high specific intensity synchrotron source offers sensitivities many times greater than the sealed X-ray tube source and because of the small beam divergence it is possible to obtain good signals from very small specimens.

The detection limits directly achievable by the X-ray fluorescence method generally lie in the low ppm range or, in absolute terms, a nominal detection limit of 0.1 to 0.01 μg on a thin film sample. In favorable cases, such as the transition elements, this limit may go to about 0.1 ppm, and in the most unfavorable case typified by the lower atomic number elements, to about 50 ppm. One of the special problems encountered in the direct analysis of water samples stems from the need to support the specimen under examination. Most conventional spectrometers irradiate the sample from below and the support film both attenuates the signal from the longer wavelength characteristic lines as well as introducing a significant ''blank''. Absorption by air also becomes an important factor for the measurement of wavelengths longer than about 2 Å and the need to work in a helium atmosphere introduces further attenuation of the longer wavelength signals.

ABBREVIATIONS AND ACRONYMS

HTSC	High-temperature Superconductor
LLD	Lower Limit of Detection
LSM	Layered Synthetic Microstructure
ppm	Parts per Million
SRM	Standard Reference Material
SSXRF	Synchrotron Source X-ray Fluorescence
TAP	Thallium Acid Phthalate
TRXRF	Total Reflection X-ray Fluorescence
UV	Ultraviolet

RELATED ARTICLES

Environment: Water and Waste **(Volume 4)**
Polynuclear Aromatic Hydrocarbons Analysis in Environmental Samples • X-ray Fluorescence Spectroscopic Analysis of Liquid Environmental Samples

Forensic Science **(Volume 5)**
X-ray Fluorescence in Forensic Science

Steel and Related Materials **(Volume 10)**
X-ray Fluorescence Spectrometry in the Iron and Steel Industry

Nuclear Methods **(Volume 14)**
PIXE (Particle-induced X-ray Emission)

X-ray Spectrometry **(Volume 15)**
Absorption Techniques in X-ray Spectrometry • Portable Systems for Energy-dispersive X-ray Fluorescence • Sample Preparation for X-ray Fluorescence Analysis • Structure Determination, X-ray Diffraction for • Total Reflection X-ray Fluorescence • X-ray Techniques: Overview

REFERENCES

1. G. Von Hevesey, *Chemical Analysis by X-rays and its Application*, McGraw-Hill, New York, 1932.
2. L.S. Birks, *History of X-ray Spectrochemical Analysis*, American Chemical Society Centennial Volume, Washington, DC, 1976.
3. H.G.J. Moseley, 'High Frequency Spectra of the Elements', *Philos. Mag.*, **26**, 1024–1034 (1912); **27**, 703–713 (1913).
4. A. Hadding, 'Mineral Analysis by X-ray Spectroscopic Methods', *Z. Anorg. Allg. Chem.*, **122**, 195–200 (1922).
5. R. Glocker, H. Schreiber, 'Quantitative Roentgen Spectrum Analysis by Means of Cold Excitation of the Spectrum', *Ann. Phys.*, **85**, 1089–1102 (1928).
6. E.P. Bertin, *Principles and Practice of X-ray Spectrometric Analysis*, 2nd edition, Plenum Press, New York, 1975.
7. R. Jenkins, *An Introduction to X-ray Spectrometry*, Wiley/Heyden, London, Chapter 4, 1974.
8. R. Jenkins, J.L. de Vries, 'Isotope Excitation in On-stream Analysis', *Can. Spectrosc.*, **16**, 3–8 (1971).
9. H. Aiginger, P. Wobrauschek, 'Total Reflection X-ray Spectrometry', *Adv. X-ray Anal.*, **28**, 1–10 (1985).
10. C.J. Sparks, Jr, *Synchrotron Radiation Research*, eds. H. Winnick, S. Doniach, Plenum Press, New York, 459, 1980.
11. J.V. Gilfrich, E.F. Skelton, S.B. Quadri, J.P. Kirkland, D.J. Nagel, 'Synchrotron Radiation X-ray Fluorescence', *Anal. Chem.*, **55**, 187–190 (1983).
12. V.A. Trounova, V.S. Danilovich, K.V. Zolotarev, M.R. Predtechensii, O.F. Bobrinok, 'Applications of Synchrotron Radiation XRF Analysis in High Temperature Super-conducting Thin Film Technology', *X-ray Spectrom.*, **23**, 187–189 (1994).

13. V.E. Buhrke, R. Jenkins, D.K. Smith, *Preparation of Specimens for XRF and XRD Analyses*, Wiley/VCH, New York, 1998.
14. J. Sherman, 'Theoretical Derivation of Fluorescent X-ray Intensities from Mixtures', *Spectrochim. Acta*, **7**, 283–306 (1955).
15. G.R. Lachance, R.J. Traill, 'A Practical Solution to the Matrix Problem in X-ray Analysis', *Can. Spectrosc.*, **11**, 43–48 (1966).
16. S.D. Rasberry, K.F.J. Heinrich, 'Calibration for Interelement Effects in X-ray Fluorescence Analysis', *Anal. Chem.*, **46**, 81–89 (1974).
17. F. Claisse, M. Quintin, 'Generalization of the Lachance–Traill Model for the Correction of the Matrix Effect in X-ray Fluorescence Analysis', *Can. Spectrosc.*, **12**, 129–134 (1967).
18. G.R. Lachance, F. Claisse, 'A Comprehensive Alpha Coefficient Algorithm', *Adv. X-ray Anal.*, **23**, 87–92 (1980).
19. T. Shiraiwa, N. Fujino, 'Theoretical Calculation of Fluorescence X-ray Intensities of Nickel, Iron and Chromium Ternary Alloys', *Bull. Chem. Soc. Jpn.*, **40**, 2289–2296 (1967).
20. J.W. Criss, L.S. Birks, 'Calculation Methods for Fluorescence X-ray Spectrometry – Empirical Coefficients vs. Fundamental Parameters', *Anal. Chem.*, **40**, 1080–1086 (1968).
21. W.K. de Jongh, 'X-ray Fluorescence Analysis Applying Theoretical Matrix Correction. Stainless Steel', *X-ray Spectrom.*, **2**, 151–158 (1973).
22. R. Jenkins, J.L. de Vries, *Practical X-ray Spectrometry*, 2nd edition, Macmillan, London, Chapter 9, 1970.

Analytical Problem Solving: Selection of Analytical Methods

Alan H. Ullman, Thomas M. Thorpe, and Glenn D. Boutilier
The Procter and Gamble Company, Cincinnati, USA

1 Introduction 1
2 Definition of the Problem 1
3 Information Gathering 3
3.1 Problem Originator 3
3.2 Company Sources 4
3.3 External Sources 4
4 Considerations in Selecting an Analytical Method 4
4.1 Technical Considerations 5
4.2 Nontechnical Considerations 6
5 Implementation of the Analysis 7
5.1 Planning 7
5.2 Preparation, Sampling, and Pretreatment 7
5.3 Calibration, Standardization, and Analysis 7
5.4 Data Reduction 7
6 Interpretation of Results 8
7 Communication of Results and their Meaning 8
Acknowledgments 8
Abbreviations and Acronyms 9
Related Articles 9
References 9

This article of the Encyclopedia explains that analytical chemistry is problem solving and provides a systematic procedure to follow that will increase the likelihood of successfully providing a solution to the problem. The problem solving process is shown in a simple flow chart diagram with amplification in the text.

1 INTRODUCTION

Analytical chemistry is about solving problems. As analytical chemists, that is what we do. When we can, we try to prevent the problem from occurring. Otherwise, we provide a solution. Whether we work in academia or industry, government or private consulting, we solve problems. In our experience, the novice analytical chemist, when given a sample or assigned a problem to solve, has a tendency to run immediately to the laboratory to use whatever expensive instrument is at hand. **That is exactly the wrong thing to do!** This article will explain why that is wrong and provide a systematic procedure to follow that will increase the likelihood of successfully providing a solution to the problem.

The guiding philosophy to this procedure is summarized in Figure 1, the Problem Solving Flow Chart. As you can see, it is a flow chart or decision tree. At each point, the option to recycle back is offered. These ''loops'' provide the opportunity to refine continuously the problem definition and the problem solving approach as new information is obtained. This is crucial to achieving success: never lose sight of the problem!

2 DEFINITION OF THE PROBLEM

Lewis and Clark, the famous early American explorers, could have been analytical chemists. The approaches they used during their expedition to find a Northwest Passage (1804–1806) were not too different from those that a modern analytical scientist might use to solve a challenging technical problem.[a]

Lewis and Clark realized that a poor definition of their ''Northwest Passage Problem'' would at least prevent them from finding this long sought route to Western North America. Errors in setting their course, misreading the weather, choosing their method of transport, or misinterpreting the intentions of the people they encountered during their journey could also lead to tragedy for them and the members of their expedition. If they made a faux pas in any of these areas they would expend substantial effort, extra time, and perhaps lose their lives before they reached their destination and returned safely to St. Louis. The results of their quest might also be lost to other explorers who would help ''solve the problem'' of opening easily traveled routes through the Louisiana Purchase for further development

[a] In 1803, the fifty-year interest in North America, west of the Mississippi River, by United States President Thomas Jefferson culminated in two events. The first was the authorization of an expedition to the Pacific Ocean via the Missouri River and the second was the purchase of the Louisiana Territory from France. The expedition was led by Meriwether Lewis and William Clark. They returned in 1806 with the first map and scientific observations of the flora and fauna of the route up the Missouri, over the Rocky Mountains, and down the Columbia River to the Pacific. Only one member of the expedition was lost, Sargeant Charles Floyd. He died, most likely, from an infected appendix that had ruptured.(1)

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

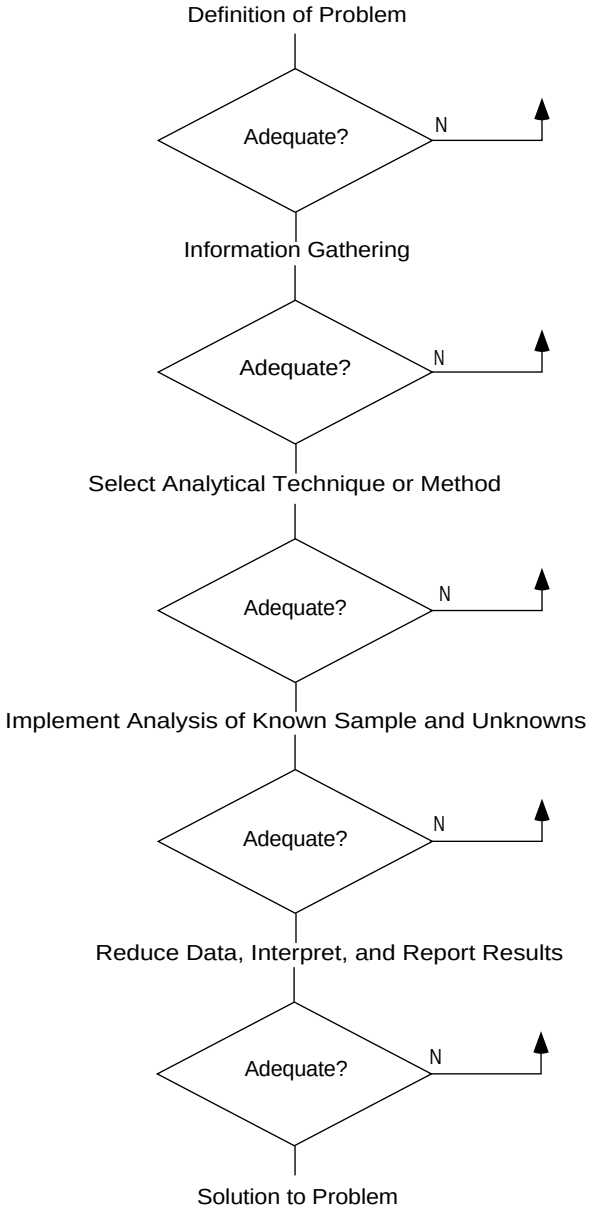

Figure 1 Analytical problem solving flow chart.

of the United States. The extra care taken by Lewis and Clark as they planned their expedition certainly helped assure the success of their two-year adventure.

Analytical problem solving is no different. Defining the problem effectively when it is first presented assures that it will be solved more efficiently. The analytical chemist must be certain that she is solving the correct problem and not merely treating a symptom of the problem. For example, if we were faced with a number of drums containing ethoxylated alcohol that have bulging tops and bottoms we could solve the symptom of the problem (i.e. the bulging) by simply releasing the gas from the drums. While this solution eliminates the symptom, it does not deal with the fundamental cause of the bulging – the chemical reactions producing the gas. Without examining the problem more thoroughly for its underlying causes the analytical chemist has no way of developing a solution that can be used to prevent recurrence of the ''Bulging Drums''.[2]

Careful definition of a problem's scope and limiting conditions – time available, expense, safety, etc. – is an essential first step to effective problem solving. In the definition phase, the analytical scientist needs to describe the chemical and physical systems from which information must be obtained in order to meet technical and organizational objectives that require a solution to the problem.

The descriptions of the ''chemical and physical systems'' are much like the information that Lewis and Clark sought as they planned their journey. These descriptions answer the questions: Who? What? Where? When? and How? In the examples, the Bulging Drum and the Lewis and Clark Expedition, we might begin defining the scope of the problem by asking the following questions:

Bulging Drum	Northwest Passage Expedition
Who discovered the drums?	Who has explored any territory west of St Louis?
What did this person observe when he found the drums?	What did these explorers see as they traveled west on the Missouri River?
Where were the drums found? What were the conditions?	What are the conditions on the river?
When were they found? Was it hot or cold?	When were portages encountered? Were they long or short?
How do these drums differ from ''normal'' drums of this chemical?	How does the terrain of these portages differ?
How does the chemical inside the drums differ from ''normal'' material?	How many rapids are there? Can keel boats navigate them or do we need canoes?

(The answers to these questions are the starting points for information gathering – the second stage of analytical problem solving that will be discussed shortly.)

It is also important when defining a problem to understand the ''technical and organizational objectives'' associated with it. These objectives also set some of the boundary conditions and a framework within

which the problem must be solved. The technical and organizational objectives are often intertwined and hard to distinguish as separate. When Lewis and Clark started their journey, they set off with a ''technical objective'' of discovery and exploration: to catalogue plant and animal life in the Northwest Territory. President Thomas Jefferson, however, had an ''organizational objective'' for Lewis and Clark: to find an all water route to the western edge of the Louisiana Purchase and to establish relations with the Native American inhabitants. Clearly the solution to the problem of exploring the American Northwest had to accommodate both objectives if it were to be considered a success.

In analytical chemistry, technical objectives are typically intended to produce purely technical outcomes – develop a better compositional determination, lower the detection limit of a method, understand the chemistry responsible for the performance of a product, or find novel chemical leads that may be fruitful avenues for new products. These objectives appear somewhat removed from organizational objectives, yet the technical objectives are shaped by organizational goals. In many respects, the technical objectives are a subset of the organizational objectives.

Organizational objectives usually guide the direction of problem solving in industry. These objectives focus on meeting the needs of customers for a product that is novel, can be produced efficiently, performs consistently and safely, is legal, and can be purchased at an acceptable price.

3 INFORMATION GATHERING

This is the second step in analytical problem solving. It was also an important component of the Lewis and Clark journey – they had to plan an itinerary based on whatever information they could find about the land and water courses along their intended route west from St. Louis.

Sometimes, even with careful definition of a problem, preliminary exploration may be necessary. When Lewis and Clark reached the junction where the Marias River joins the Missouri River, they had to decide which of these was the Missouri. When scouting reports were inconclusive, Lewis set out with a small party up the north fork and Clark took his party up the south fork. Upon returning, both Lewis and Clark believed the south fork was the Missouri while every other member of the expedition believed the north fork to be the correct choice. Lewis and Clark were correct in choosing the south fork as the Missouri, but Lewis went ahead with a small expedition by land to look for the expected falls or the mountains. Due to the lateness of the season, this was almost an irrevocable decision.

Similarly, an analytical scientist may have to reconsider her approach to the problem as she gathers new information. The skilled analytical problem solver collects enough information to narrow the many possible approaches to solve the problem, so as to use the most effective few. This may require her to reset her course to reach the true solution – an explanation for production of the gas in the bulging drums and a plan to prevent future occurrences of the chemical reactions that were responsible.

The questions – Who? What? Where? When? How? and Why? – posed as the problem was defined are now revisited to tease out essential details of information that will guide the eventual selection of the analytical methods used to solve the problem.

A few of the plentiful information sources available to the analytical chemist are presented in Table 1. These sources are classified into three categories.

3.1 Problem Originator

The first, problem originator(s) – either the person(s) who found the problem or the person(s) who caused it – are rich with details about the problem. These people often have first hand, specific, information that no one else has seen or knows about. The challenge for the analytical problem solver, like a good criminal detective, is to extract all of the pertinent details from the problem originator. This may best be accomplished by actually inspecting the problem first hand with the originator, or by asking the originator to explain the chemical process and

Table 1 Sources of information for analytical problem solving

Problem Originator(s)
Internal Company Sources
Company experts
Company reports
Memoranda
Standard analytical methods
Laboratory notebooks
External Sources
Published technical literature
Primary literature: journals reporting original research
Secondary literature: sources that review the literature, such as: handbooks, encyclopedias and treatises, monographs and reviews, textbooks (general and specialized)
Published standard analytical methods
Patents
Consultants: academic or industrial experts
Contract research laboratories: Laboratories that do research or perform analytical determinations for a fee, such as Battelle Laboratories, Kovance Laboratories America, and Southwest Research Institute
Government agencies
Instrument manufacturers and suppliers

equipment that produced the problem: ''Did operators do anything differently when they filled the drums of ethoxylated alcohol that would cause them to bulge?'' If the ''originator'' only discovered the problem and is not knowledgeable about all of the technical aspects of it, he may at least be able to refer us to a person who is more familiar with the situation. Also, don't forget that the originator may be well aware of the organizational ramifications of the problem – the manufacturing plant must shut down, the gas in the bulged drums may pose a safety hazard, or the ethoxylated alcohol may be contaminated and should not be used to make detergent products. The problem originator is frequently the best source of a broad set of background information. That information becomes a foundation from which to build a better understanding of the problem using internal company sources or by consulting external sources of information. This background information is also useful in creating early hypotheses about the cause of the problem that can be tested as more details are collected.

3.2 Company Sources

Internal company sources are obviously the next place to search in order to add information about the problem – its chemistry, previous occurrences of the problem under similar conditions, analytical procedures used to obtain critical data about the substances involved, and possible experts in the technology who continue to work with it or may have moved on to other assignments, yet are still accessible. The utility of the other internal company sources listed in Table 1 generally decreases in the order they are presented. Company reports are usually easily obtained, although the amount of detail they contain and frequent lack of direct applicability to a problem may limit their value. Memoranda often deal with very specific situations that may be difficult to link to the problem at hand, and they may be difficult to locate except in the most far-reaching and robust corporate information management systems. Analytical methods are much like company reports; relatively easily available, often with an archive of previous versions, and fairly specific in their application. Finally, there are laboratory notebooks. While these are frequently archived for many years, and references to them come from reports and analytical methods, their content is often difficult to decipher and put to direct use in solving a problem, except in the best of cases.

3.3 External Sources

External sources of information and the means to access them, either in hardcopy or through the Internet, are generally well known to the practicing analytical chemist. This is certainly the case with primary and secondary published literature, standard analytical methods, and patents (Table 1). Identifying useful consultants to address a very specific problem is usually a longer term proposition since the consulting scientist must be located, their credentials verified, consulting agreements drawn up and consulting time arranged. All of this may take more time than is available to solve a problem. On the other hand, industrial experts may be readily identified and easily accessed, especially if they are in some way connected with the problem. In the case of the bulging drums, the manufacturer of the ethoxylated alcohol is knowledgeable about its chemistry and is probably interested in helping the customer solve the problem in order to keep the customer's business. Contract research laboratories can also be rich sources of information. When a problem relates to health, safety of chemicals, occupational safety, the environment, and regulated substances like food and drugs, government agencies may be a preferred source of information. Finally, manufacturers of analytical instruments or suppliers of chemicals often have an applications department whose scientists may have worked on analytical problems similar to the one you need to solve. They may be willing to assist – it could lead you to purchase one of their instruments. However, that help may be provided in a circumspect manner so as not to divulge information that the instrument manufacturer has obtained during their work with another company.

Ideally, we would like to perform a very thorough search for information so that we turn up all possible leads that can point the way to a solution. However, we must always remember the boundary conditions that we uncovered when we defined the problem. They may necessitate very rapid information gathering so that we can move forward quickly to solve the problem. Said differently, we may need to identify two or three of the most promising leads very quickly so that they can be evaluated before we try to exhaustively identify all of the potential causes for a problem. We can always recycle back to look for new explanations for the problem if the first ones do not work out. Lewis and Clark certainly charted and recorded their course regularly, and, when necessary, adjusted it to travel westward as directly as they could.

The result of information gathering should be several hypotheses about the cause of the problem that can be tested experimentally. With these hypotheses in hand we can proceed to the next stage of the problem solving process – selection of appropriate analytical techniques that we apply either to confirm or refute our hypotheses.

4 CONSIDERATIONS IN SELECTING AN ANALYTICAL METHOD

We have now defined the problem, gathered and studied the available information, and perhaps redefined the

problem, so now it is time to begin thinking about the analytical method(s) we will use when we go to the laboratory. What are the issues that we must consider? More than likely, the first thing that came to mind was one of the ''technical issues'' such as lower detection limits of the method, so let's address these issues first.

4.1 Technical Considerations

4.1.1 Limit of Detection

One of the first factors you are likely to consider is the limit of detection (LOD) of a candidate method. The LOD is the lowest concentration or amount of analyte which produces a signal that can reliably be differentiated from the response obtained in the absence of analyte or from a blank. Clearly, this is a critical issue – a method with a LOD of 1% Zn in water won't be very useful if your problem involves trace level Zn contamination in a lake. The LOD is generally defined as the amount of analyte producing a signal which is three times the noise level of an analyte-free blank (S/N = 3). At a S/N of 3, there is still significant uncertainty that data are valid, so we generally try to work at a S/N of at least 10 (the *limit of determination*). Inherent to the LOD, and often confused with it, is *sensitivity*, the analytical response per unit concentration of analyte (i.e. the slope of the *calibration curve* for the technique). Remember that a method with high sensitivity does not necessarily have a low LOD; LOD also includes the noise of the measurement. Figure 2 shows these terms graphically.

4.1.2 Accuracy and Precision

These two terms are confused as often as sensitivity and LOD, but actually are easy to understand. *Accuracy* is just the difference between your measured result and the true or accepted value. In actual practice, the average (mean) of multiple determinations is compared to the accepted value; but what is the accepted value? This may come from another older, proven method, an industry standard method, or an ''official'' standard. Accuracy of final results will be no better than the accuracy of the standards used in calibrating the analytical method. There are two kinds of standards, primary and secondary, which can be used to calibrate a method. In addition, the accuracy of methods can be validated using a variety of reference materials including those from the US National Institute of Standards and Technology (NIST, formerly the National Bureau of Standards, NBS), specialty standards from organizations such as the Environmental Protection Agency (EPA), United States Geological Survey (USGS) and their non-US equivalents, as well as, industry organizations.

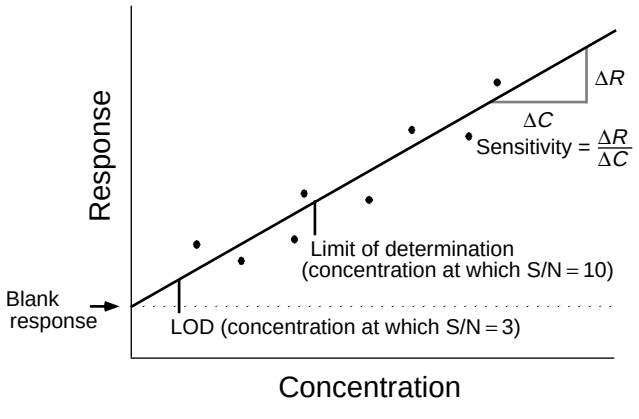

Figure 2 An analytical calibration curve with annotations defining several analytical figures of merit.

4.1.3 Precision

Precision is the repeatability or reproducibility of a measured result for a set of replicate analyses of the same sample. Figure 3 shows the difference between accuracy and precision in a simple, graphical way and introduces additional terms – systematic and random error. Systematic errors give results that are skewed in

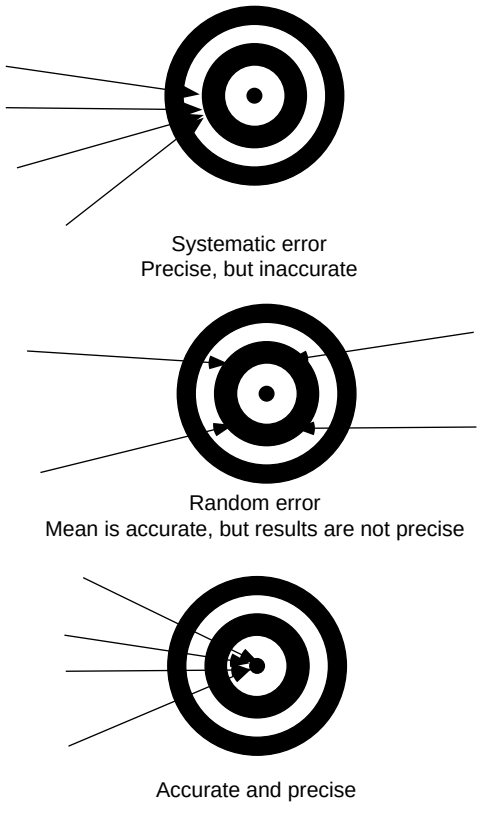

Figure 3 The difference between accuracy and precision.

one direction; random errors cause results to be high or low, randomly.

4.1.4 Selectivity

The complexity of the *matrix* of a sample is a critical factor as you consider the appropriate analytical method to use. The selectivity of a method is its ability to respond to the analyte of interest in preference to other substances in the matrix. In general, we prefer using methods with higher selectivity to increase the reliability of the analytical results. *Interferences* are these other substances in the sample to which the method responds.

The sample matrix is the material with which the analyte is physically associated and/or chemically bound. In addition to the number of different chemical species, the matrix can be complex in its physical state. As you see in Table 2, this is not a trivial concern.

Quantitatively determining an unknown analyte in an unfamiliar matrix is a more daunting task than determining a known analyte in a well understood or simple matrix. Note that the more complex the matrix of the sample, the more important the *sampling* step becomes. Unfortunately, sampling is not always in the direct control of the analytical chemist. It is therefore, very important that the sampling procedure used by the ''customer'' (the individual who first presented the problem) be understood and, perhaps, questioned.

Selectivity is a continuum from nonselective methods to those that are specific for an analyte. As you will see in the later articles of this encyclopedia, there are many ways to achieve the selectivity needed for a given measurement. Some of these include chemical reaction-based methods, including inorganic qualitative analysis and organic functional group analysis, electrochemistry, radioimmunoassay, and enzymatic analysis. Physical separation techniques such as solvent extraction, chromatography, volatilization, and precipitation can all be used to separate the analyte from interfering compounds in the sample. Additional approaches such as spectroscopy, coupled techniques [e.g. gas chromatography/mass spectroscopy (GC/MS)], and modern data reduction and manipulation (e.g. chemometric techniques including factor analysis, pattern recognition, partial least squares, etc.) may also provide selectivity.

Table 2 Different physical states of the sample matrix

State of matrix	Location of analyte
Solid	Uniformly dispersed
	Segregated in or with particles of particular size, shape, density or composition
	Segregated at surface of solid
	Localized at a single point within the solid
Liquid	Uniform solution
	Colloidal suspension
	Micellar systems
Gas	Uniform mixture
	Suspended particles
	Aerosols

4.2 Nontechnical Considerations

In addition to considering technical factors such as accuracy, LOD, etc. your choice of analytical method will depend on many nontechnical factors. Does your laboratory have access to the needed equipment? Are personnel available to run the method? How many samples will be analyzed and how often? Is this a one of a kind problem, requiring only a small number of determinations or will the method be installed in a quality control (QC) lab and used hourly? This is particularly important to consider if the method of choice is expensive to perform because of high reagent cost or waste disposal costs. How long does it take to perform the determination? If it takes a long time, but you only have to do it a few times, that is quite different from use of the method in a QC laboratory. If you only need to analyze a few samples, perhaps using a contract lab or borrowing/renting an instrument is more appropriate. Don't forget that any method you choose must be safe to perform in your laboratory. Do you have the appropriate hood for a perchloric acid digestion, for example? Each of these must be considered before you are ready to move onto the implementation phase of the problem solving flow chart. Tables 3 and 4 provide more details on the *cost* and *safety* issues.

4.2.1 Equipment Availability

Hopefully your own lab has the needed capability, but if not it may be available within another laboratory of your company. If not, it can be acquired by borrowing, leasing or renting, or purchase. It may be faster and more cost effective to use a contract laboratory or arrange to use equipment at a nearby university, regional research

Table 3 Cost issues in selecting an analytical method

Tangible expenses	Intangible expenses
Chemicals	Analyst time
Supplies	Development effort
Instrumentation	Lost production due to time needed to conduct determination
	Safety
	Overhead costs:
	Employee benefits
	Electricity, water, heat, etc.
	Space
	Quality assurance

Table 4 Safety issues in selecting an analytical method

Chemical hazards	Physical hazards
Flammables	High voltage/current
Corrosives	Intense light sources
Explosives	Radioactive materials
Toxics	
Carcinogens	
Teratogens	
Chemical fumes	

center, equipment supplier, or a government laboratory. There may be additional resources, such as a glass blower, machine shop, or photo lab, needed to solve the problem. These may be available in your company or university, or contracted.

5 IMPLEMENTATION OF THE ANALYSIS

5.1 Planning

After an analytical method has been selected, it is necessary to plan for the implementation of the analysis. If any hazardous samples or reagents are involved, the appropriate safety procedures should be reviewed with the affected laboratory personnel. This is especially important if the hazard level and corresponding safety procedure fall outside the norm for the laboratory. In the example of the Bulging Drum, the laboratory did not routinely handle explosive gases. If the samples are being analyzed to comply with government regulations, those regulations should also be reviewed. If, for example, the samples are the result of a designed experiment, then not all samples may be available at the same time. The customer and the analytical laboratory may need to agree on whether the samples will be analyzed as they become available, or only when all samples are available. When limited amounts of sample are available for multiple analyses, it may be necessary to plan the order of analyses to maximize the available sample.

5.2 Preparation, Sampling, and Pretreatment

Preparation and sampling may be concurrent or sequential. Frequently they are concurrent when the sampling is done by someone other than the laboratory analyst. Sampling is a topic where communication between the analytical laboratory, the customer, and the person performing the sampling is crucial. The type of sampling is influenced by both the physical state of the material (solid, liquid, or gas) and the nature of the question being addressed. Random sampling might be appropriate to characterize the variability in a sample, while systematic sampling will be necessary to characterize properties as a function of time. If a representative sample from a large bulk quantity of material is required, all parties must agree on whom is responsible for reducing the sample size to the appropriate level. In the example of the Bulging Drum, quickly learning to sample gases was the first order of business for a laboratory that normally received liquids and solids. For the laboratory to provide the most useful interpretation of the results, a complete understanding of the origin of the samples is necessary. Solid samples frequently require dissolution, digestion, extraction, etc. to be in a form suitable for analysis. Each member of the problem solving team needs to remember that the goal is a solution to the problem and the result of the sampling process will define the scope of the information to be applied to a solution. Next, the required reagents, apparatus, and instruments must be assembled; then sample pretreatment can begin. This is also usually the first stage where chemical selectivity is applied. This can range from extraction from water at a controlled pH into an organic phase, to liquid chromatography as a precursor to gas chromatography. The type of pretreatment is also strongly dependent on whether the goal is preconcentration (increase the analyte to matrix ratio) or cleanup (removal of items that will interfere with the analysis).

5.3 Calibration, Standardization, and Analysis

Calibration and standardization first come into play in quantitative analysis. Any spectroscopic method of identification relies on a calibrated frequency or wavelength scale. Verification of this calibration should be a routine part of analytical laboratory operation. If an internal standard is used, it should be added as early as practical in sample preparation. Quantitative analysis to determine one or more components of the sample requires standards to be procured or synthesized. Deciding on the range of calibration can be done in an iterative method. This is consistent with the recycle option illustrated in Figure 1. Additional standards can be added if results for the samples fall outside the range of the initial calibration. For complicated and varying sample matrices, the method of standard addition may be preferable to additional pretreatment or separation during sample preparation. When samples are finally analyzed, it is often a very satisfying point for the problem solver. A more challenging point may be convincing the customer on the need for a blank and agreeing on the appropriate blank.

5.4 Data Reduction

Data reduction is usually a combination of the application of statistics to analytical chemistry and the calculation of concentration, mass, volume, etc. from

a measured response. Integrated microprocessors on analytical instruments and desktop computers have greatly reduced the tedium involved in the calculations for large amounts of data. It is still possible, however, to use the wrong formula in a spreadsheet and turn the most carefully obtained analytical information into wrong answers. When any analytical figure of merit, such as precision, LOD, etc., is reported, the method used to calculate the figure of merit should be included. This is especially important if a result is ''None Detected'' and the value of the LOD is to be used in further calculations.

6 INTERPRETATION OF RESULTS

One of the reasons for doing careful problem definition, information gathering, and method selection is to make interpretation of results straightforward. The best experiments have results that are easily understood in terms of a solution to the problem. In the example of the bulging drum, results obtained prior to the identification of hydrogen as the explosive gas are not easily understood. The detection of hydrogen gas is easily understood as an explosive gas, but not in terms of a solution to the problem. Only when identification of hydrogen gas is combined with the presence of sodium borohydride are the results easily understood from the standpoint of problem solution. The focus here has shifted from samples to problem solution. The complete interpretation of results should also have the total problem as its focus rather than the analyst or analytical laboratory. Remember, the problem solution is not the analytical result; the analytical result is the key to understanding and solving the problem.

7 COMMUNICATION OF RESULTS AND THEIR MEANING

We have now reached the bottom of the Problem Solving Flow Chart (Figure 1) and the solution to the problem, assuming the data collected has, in fact, led to results which make sense. If not, we must go back to an earlier step and start again. If we are satisfied with the results obtained by our analytical problem solving skills, we must communicate the results and their meaning to everyone involved in the problem. **Without sharing the knowledge, the problem has not been solved!** Unless the results are documented and circulated, filed, etc., your efforts have been wasted. The complete problem has not really been solved; the situation cannot be rectified. In the Bulging Drum example used earlier, the technical solution of the problem was determined to be hydrogen gas derived from sodium borohydride decomposition in the drums of ethoxylated alcohol that had slightly elevated water content and lower than expected pH. With this knowledge in hand, the analytical chemist can meet with the process engineers and supplier of the ethoxylate to discuss options concerning the borohydride. Perhaps it isn't needed and the supplier won't add it in the future. If that isn't possible, perhaps the ethoxylate specifications can be modified to require lower water content or higher pH. Plant procedures might also be changed so that the ethoxylate is used sooner and stored at lower temperatures. If these changes aren't made, why shouldn't the problem recur? In addition, the complete story, from problem to analytical solution to process resolution needs to be documented in a form that can be located in the future.

If Lewis and Clark hadn't documented their journey, how would successive explorers and settlers have gained anything from it? Unfortunately, because of Lewis' suicide before the journals were prepared for publication, few did gain much from the expedition's scientific observations for the next hundred years. It was not until 1904 that a complete edition of the journals was published.

A report in the company library, appropriately indexed for searching, is the logical answer. The company intranet might also be a repository for the information, and circulation of a brief summary to appropriate people is also a useful tool. Presentation of good analytical detective work (problem solving!) at company technical symposia is another great way to prevent similar problems from occurring in the future.

The bottom line: good laboratory work is wasted if the results are not documented and translated into information that can be used to deliver an organizational solution to the problem, as well as the technical solution.

ACKNOWLEDGMENTS

The material in this article is derived, in part, from a one day short course we have taught which has had as its aim, explaining industrial analytical chemistry, careers in said field, and analytical problem solving to undergraduate students. The authors acknowledge other present and past Procter and Gamble Company analytical chemists who contributed to the ideas expressed in this article through their participation as course instructors: Val Adams, Bonnie Charpentier, Margi Conditt, Linda Cooper, Dick DePalma, Matt Doyle, Chris Frank, Mark Greenberg, Anita Guy, Susan Lunte, Sandy

Murawski, Diane Parry, Larry Perkins, David Pilosof, Doug Raynie, Pamela Schofield, and Suzanne Tanguay. We also acknowledge Ted Logan, Joel Shulman, Grover Owens, Bob Wade, and Whit Wharton for their input and support.

ABBREVIATIONS AND ACRONYMS

EPA	Environmental Protection Agency
GC/MS	Gas Chromatography/Mass Spectroscopy
LOD	Limit of Detection
NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
QC	Quality Control
USGS	United States Geological Survey

RELATED ARTICLES

Chemometrics **(Volume 11)**
Chemometrics

General Articles **(Volume 15)**
Literature Searching Methodology • Quantitative Spectroscopic Calibration

REFERENCES

1. S.E. Ambrose, *Undaunted Courage: Meriwether Lewis, Thomas Jefferson, and the Opening of the American West*, Simon and Schuster, New York, 1996.
2. T.M. Thorpe, 'What Caused the Drums to Bulge', *Anal. Chem.*, **56**(4), 603A–605A (1984).

Archaeological Chemical Analysis

A. Mark Pollard and Carl Heron
University of Bradford, Bradford, UK

1 Introduction 1
2 The History of Analytical Chemistry in Archaeology 2
3 The Role of Chemical Analysis in Archaeology 4
3.1 Questions of Process 4
3.2 Archaeological Questions 5
3.3 Special Considerations in the Analysis of Archaeological Material 11
4 Spectrometric Methods Using Optical Wavelengths 12
5 X-ray and Electron Techniques 13
6 Nuclear Techniques 14
7 Mass Spectrometry 15
8 Chromatography and Other Analytical Techniques 16
9 Summary 16
Abbreviations and Acronyms 17
References 17

In recent years, archaeology has interacted more and more with the physical sciences. In particular, chemical analysis has been established as a significant contributor to science-based archaeology. The principal applications include compositional analysis of natural and synthetic materials and residues to ascertain artifact manufacturing processes and use; chemical and isotopic studies of biological remains and identification of plant and animal residues to investigate ancient diet, nutrition, and resource use; the determination of geographical sources of procurement or production of materials to establish long-distance contact or trade; and the processes governing preservation and decay of materials and scientific investigation in the context of conservation and restoration.

This article explores the role of chemical analysis in archaeology and highlights the importance of the findings in contributing to our understanding of the past. In particular, taking examples from the last 200 years, it demonstrates that chemical analysis of archaeological materials is not a routine application but one that must be informed by a thorough knowledge of the archaeological context, degradative processes, and material – environment interactions. Case studies span a wide range of elemental, isotopic, and molecular investigations with a bibliography comprising nearly 200 publications.

1 INTRODUCTION

Archaeology is about people. In one of the leading introductory texts on the subject, Renfrew and Bahn[1] state that archaeology is concerned with ''the full range of past human experience – how people organized themselves into social groups and exploited their surroundings; what they ate, made, and believed; how they communicated and why their societies changed''. Archaeology has been called ''the past tense of cultural anthropology'', but it differs crucially from anthropology in one respect – it is impossible in archaeology to interview the subjects, or to observe them directly in their everyday life. Inferences about past societies are made from the evidence recovered by archaeological excavation – sometimes material evidence in the form of surviving artifacts or structures, sometimes fortuitous evidence in the form of preserved associated biological material such as insect remains. Sometimes it is the soils and sediments themselves – their nature and stratigraphy – which provide the evidence, or add additional information by providing a context. Hence the occasionally acrimonious debates in archaeology about the effects of looting or the undisciplined use of metal detectors, where objects are removed from their contexts without proper recording and informational value is lost.

Although archaeology is a historical discipline, in that it reconstructs events in the past, it is not the same as history. If history is reconstructing the past from documentary sources, then 99.9% of humanities' five million years of evolution is beyond the reach of history. Even in historic times, there is still a distinctive role for archaeology. Documentary evidence can provide evidence on the ''big events'' – famous people, battles and invasions, and the history of states, but such written sources are often biased. History is written by the literate, and usually by the victorious. It can obscure the past as well as illuminate it. In contrast, archaeology is generally the story of the common people – about how they lived and died.

It is the focus on material evidence in archaeology which creates the need for scientific approaches to the past. There is a necessity to extract the maximum possible information from the generally mundane collection of bones, stone tools, shards of broken pots, corroded metalwork, and biological assemblages which make up the vast bulk of archaeological finds. Trade routes are

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

inferred from fragments of broken glass or pottery manufactured in one place but found in another. The economies of ancient cities are reconstructed from a study of the animal bones found on midden tips. In this respect, archaeology has much in common with modern forensic science – events, chronologies, relationships, and motives are reconstructed from a wide range of material remains. To set the scene for fresh recruits to the study of the science of archaeology, it is instructive to challenge new students to name a scientific discipline which has no relevance to modern-day archaeology. One can easily go through the scientific alphabet, from astronomy to zoology, and find many obvious applications. It is possible, of course, to carry out the same exercise in the social sciences, and also in engineering. Few disciplines exist which have no relevance or application in archaeology, and archaeology is therefore truly an interdisciplinary subject.

A comprehensive history of scientific analysis applied to the study of past people and materials has not been written. Caley[2–4] gives summaries of the early applications of chemistry to archaeology, and a review paper by Trigger[5] gives an overview of the relationship between archaeology and the physical and biological sciences. A collection of recent scientific studies, largely relating to museum objects, including dating, authenticity, metalwork, ceramics, and glass, can be found in the volume edited by Bowman.[6] Many conference proceedings (especially those entitled *Archaeological Chemistry*, produced by the American Chemical Society[7–11]) contain a very wide range of chemical studies in archaeology. Of the several books covering the chemical aspects of archaeological science, Goffer[12] gives a very broad introduction to archaeological chemistry, covering basic analytical chemistry, the materials used in antiquity, and the decay and restoration of archaeological materials. More recent publications include that of Pollard and Heron,[13] which gives a basic introduction to instrumental chemical analysis followed by seven chapters of case studies, and the book by Lambert,[14] which has eight chapters, each based on the study of a particular archaeological material.

This review examines the historical origins of the applications of chemical analysis within archaeology. It then discusses the types of problem that can be answered by such analyses, dividing them somewhat arbitrarily into those designed to elucidate processes acting on archaeological materials, and those designed to answer specific archaeological questions. Consideration is given to the particular problems caused to the analyst by archaeological materials. This is followed by a brief review of the applications of analytical chemistry to archaeology, classified by broad analytical technique.

2 THE HISTORY OF ANALYTICAL CHEMISTRY IN ARCHAEOLOGY

Chemistry can claim to have a longer relationship with archaeology than most other sciences, going back well into the eighteenth century. The range of materials studied has varied enormously, ranging from the more obvious archaeological finds such as obsidian, flint, glass, pottery, and metals, to those which survive only occasionally and under exceptional conditions, such as skin, hair, and (possibly) faint traces of blood on stone tools.

Many eminent chemists feature in the early history of the scientific analysis of antiquities, such as Klaproth (1743–1817), Davy (1778–1829), Berzelius (1779–1848), Faraday (1791–1867), Berthelot (1827–1907), and Kekulé (1829–1896). Klaproth determined the approximate chemical composition of some Greek and Roman coins, a number of other metal objects, and a few pieces of Roman glass.[15] In producing compositional data on ancient materials, Klaproth had first to devise workable quantitative gravimetric schemes for the analysis of copper alloys and glass. He became Berlin's first Professor of Chemistry, and it is fitting, therefore, that he can also be credited with the inception of archaeological chemistry.

The discovery of the ruins of the Roman cities of Herculaneum and Pompeii in 1594, and their subsequent "excavation" starting in 1709, ignited a passion for Neo-Classicism in Europe.[16] By the mid-eighteenth century, Pompeii was an important stop on the European "Grand Tour", and several of the Royal Houses of Europe were sponsoring systematic recovery of objects from the sites. In 1815, Humphry Davy published a paper on the examination of ancient pigments collected at Rome and Pompeii.[17] In addition to evidence for natural pigments, he was also able to identify a synthetic pigment later to be called *Egyptian Blue*, formed by fusing copper, silica, and naturally occurring natron (sodium carbonate). A report by Diamond in 1847 includes a short contribution on a Roman pottery glaze studied by Michael Faraday, which provided the first indications of lead glaze in antiquity.[18] Similarly, Kekulé carried out the analysis of an ancient wood tar composed, in part, of compounds with aromatic rings, the structure of which he subsequently proposed in 1865.

These early forays were largely devoted to satisfying curiosity and to the acquisition of basic knowledge about the nature of materials used in antiquity. As early as the mid-nineteenth century, however, Wocel suggested that correlations in chemical composition could be used to *provenance* or identify the source of archaeological materials, and even to provide relative dates for manufacture and use.[19,20] During the 1840s, Göbel, at the University of Dorpat in Estonia, studied large numbers of copper alloy artifacts from the Baltic region, comparing them

with known artifacts of prehistoric, Greek, and Roman date.[21] He concluded that the Baltic artifacts were probably Roman in origin. This marked a progression from the study of individual objects to a consideration of assemblages. The French mineralogist Damour proposed that the geographical source of stone axes could be located by considering the density and chemical composition of a number of rock types, including jade and obsidian.[22] He also exhorted archaeologists to work closely with specialists from other disciplines such as geology, zoology, and paleontology – a plea which is still heard today.

The appearance of the first appendixes of chemical analysis in major excavation reports represents the earliest systematic collaboration between archaeologists and chemists. Examples include the analysis of four Assyrian bronzes and a sample of glass in Layard's *Discoveries in the Ruins of Nineveh and Babylon* published in 1853[23] and Schliemann's *Mycenae* first published in 1878.[24] So influential was this latter work that William Gladstone, the British Prime Minister, wrote the preface. The reports in both these works were overseen by John Percy, a metallurgist at the Royal School of Mines in London. Between 1861 and 1875, Percy himself wrote four major works on metallurgy that include significant sections on the early production and use of metals, and which remain key sources for those interested in ''pre-Industrial'' metal processing.[25–28] An important aspect of this work was the realization that the benefit of using chemical analysis to elucidate the composition of archaeological material was not just one-way. Percy himself wrote that ''Some of the results are, I think, both novel and important, in a metallurgical as well as archaeological point of view''.

Schliemann, referring to more than 2000 amber beads recovered from Mycenae, wrote ''It will, of course, for ever remain a secret to us whether this amber is derived from the coast of the Baltic or from Italy.'' Towards the end of the 19th century, Otto Helm attempted to source amber in one of the earliest systematic applications of chemistry to a specific problem in archaeology.[29] Helm based his approach, in part, on the succinic (butanedioic) acid content of Baltic amber, but did not undertake a study of other sources of fossil resins in Europe. A full account of these investigations has been compiled by Beck, who published the results of an intensive study using infrared (IR) spectroscopy which successfully discriminated between Baltic and non-Baltic European fossil resins.[30] It appears that the vast majority of amber finds throughout prehistoric Europe do indeed derive from the Baltic coastal region, implying the presence of long exchange routes at an early date.

As more archaeological objects were recovered, there was an increasing need for restoration and conservation, which demanded a better understanding of the decay mechanisms for a wide range of materials. The pioneer in this area was Friedrich Rathgen, who in 1888 established a laboratory at the State Museum in Berlin and later published the first book dealing with practical procedures for the conservation of antiquities, including electrolytic removal of corrosion products and the use of natural consolidants (pine resin and gelatin) in conservation.[31] Thus was born archaeological conservation science, in which substantive chemical understanding of degradation processes, and how to stabilize or reverse them, is applied to an ever-increasing range of museum materials.

By the end of the nineteenth century, new ideas began to appear. In 1892, Carnot suggested that fluorine uptake in buried bone might be used to provide an indication of the age of the bone,[32] although the method was not tested until the 1940s, when such studies were central to the detection of the famous Piltdown hoax.[33] The chemical study of archaeological bone, especially human, and more recently teeth, has continued to be a principal interest of archaeological chemists. The chemical composition of the mineral fraction is related to factors such as diet, health status, and exposure to environmental toxins – information important to an understanding of the quality of life in past societies. The isotopic composition of the organic component is similarly related to diet and health, and differences between various components of the skeleton with different turnover rates have been linked to mobility.[34] Obviously, the analysis of bone presents different challenges to the analysis of inorganic artifacts. In life, it is a living tissue which does not respond simply to changes in the chemistry of dietary intake – assimilation of essential and trace elements into bone mineral is governed by physiological factors, which therefore require a detailed knowledge of human biochemistry. After death, biochemical control ceases, and another regime with different laws takes over – that of geochemistry. The composition of buried bone is controlled by a great many factors which, again, are not simple. Thus, in order to study one of the most important and abundant materials in the archaeological record – bone – archaeological chemistry has had to become familiar with the vast and complex subject of biogeochemistry.[35]

Occasionally – perhaps once or twice a century – an archaeological find is made which radically influences life, not least in terms of fashion, art, and architecture. The revelation of Pompeii in the latter part of the eighteenth century was one such event. The discovery of the tomb of Tutankhamun by Howard Carter in 1922 was another.[36] Perhaps less globally influential, but certainly of great public interest, was the discovery in September 1991 of the well-preserved body of a man high up in the Ötztaler Alps on the Austrian–Italian border.[37] Subsequent investigations recovered not just the body,

but also a large collection of his equipment, some of which was made from materials such as birch bark which rarely survive in normal archaeological contexts. A large program of scientific investigations was initiated, ranging from obvious questions such as the date of the remains, through to the identification of the parasites associated with the body, and the manufacturing technology of his clothing and equipment. Radiocarbon dating of bone and tissue from the body, as well as fibers from his clothing, revealed a date of between 3350 and 3120 BC, placing him in the Neolithic Period of northern Italy. Chemical analysis of the surface of his copper alloy ax revealed it to be almost pure copper (better than 99% copper, with traces of arsenic and silver). This composition is one which is thought to be characteristic of the Neolithic/Bronze Age transition, traditionally dated to about 2500 BC. The shape of the ax, however, is one which would conventionally have been described as Bronze Age, and dated to about 2000 BC, more than 1000 years later than the actual date of the body. He was also carrying with him a small flint dagger, classically dated to the Neolithic. Such a find, unique in Europe, has shown us that we need to know more about the dating of the transition from the Neolithic to the Bronze Age. Above all, this study underlines the forensic nature of the scientific investigation of ancient material, involving not only archaeologists and forensic pathologists, but many other specialists.

3 THE ROLE OF CHEMICAL ANALYSIS IN ARCHAEOLOGY

The historical introduction in section 2 has identified the origins of many of the issues addressed by the application of analytical chemistry to archaeology. They can be divided into those projects which attempt to understand the processes acting upon archaeological material before, during, and after burial, and those which use chemical methods to derive information to answer specific questions of interest to archaeology.

3.1 Questions of Process

3.1.1 Degradative Processes

Most material which enters the archaeological record degrades until it ceases to be a recognizable entity. Molecular evidence may remain, but to all intents and purposes the object has disappeared. Exceptions to this rule in the main constitute the material evidence upon which archaeological inference is based. Some materials, such as stone, almost always survive degradative processes (although they may succumb to other processes such as translocation). Others, such as skin, hair, and organic fabrics, only survive in exceptional circumstances such as extremes of cold or dryness. Many materials, such as metals, glass, and some of the more resistant organic materials such as amber, will undergo some degradation, but are likely to survive in a recognizable and recoverable form. Biological hard tissue (e.g. bone, teeth, horn, shell) undergoes particularly complex patterns of degradation because of its composite nature.

Chemical and biological degradation processes are part of a wider phenomenon which is termed *taphonomy*, originally defined as the process of transition of an organism from the biosphere to the lithosphere.[38] It includes all natural and anthropogenic processes which create death assemblages before deposition, as well as those chemical, physical, and biological processes which act on the assemblage after deposition. It is also possible to conceive of the postdepositional ''taphonomy'' of nonbiotic material, e.g. metal and ceramic artifacts, since they too experience change as a result of environmental interaction, although this goes well beyond the original definition.

Analytical chemistry has a fundamental role to play in helping to understand some of the major aspects of taphonomic change. Some processes are likely to be primarily chemical in nature, such as the electrochemical corrosion of a metal object in an aqueous environment, although even here microbiological mediation is likely to be important.[39] Some processes are structural, such as mineralogical changes taking place in ceramics as a result of interaction with groundwater. Others, such as the degradation of organic materials, may be largely biological, although chemical hydrolysis may also have an important role. Whatever the driving force, analytical chemistry is essential as a means of measuring, monitoring, modeling, and verifying these processes.[40]

3.1.2 Material – Environment Interactions

The objective of understanding degradative processes is to improve our knowledge of the factors which control the preservation of archaeological evidence in the burial environment. The potential for survival is governed largely by the interaction of the material with its depositional environment. This process forms part of a continuum between the degradation of very recent material (perhaps a few months to years), knowledge of which may have forensic interest, the behavior of archaeological deposits (few tens of years up to tens of thousands), and ultimately to material of geological and paleobiological interest (hundreds of thousands to millions of years). There is little systematic understanding of the factors which control preservation for the wide range of materials encountered archaeologically, and

even less in the way of predictive models. Soil pH and Eh are often referred to as the ''master variables'' in the consideration of soil chemistry.[41] However, their measurement in the field is not always easy or even possible because of fluctuating conditions. Moreover, the chemical composition of the soil water is a complex interaction of the mineralogical, organic, and atmospheric composition of the soil, further complicated by speciation, redox, and solubility factors within the soil solution.[42] Again, direct measurement in the field is often impossible. Nevertheless, knowledge of such factors is vital for understanding the chemical environment of buried archaeological objects. In response to this, a whole family of groundwater geochemical modeling programs has been developed over the last 30 years.[43] These programs allow speciation to be calculated for given total ion concentrations under specified conditions and the behavior of particular mineral species in contact with defined waters to be modeled, enabling the stabilities of such systems and their response to environmental change to be predicted. It is clear that this is the necessary point of departure for modeling the behavior of archaeological objects in contact with soil water and groundwater, but little has yet been done.

3.1.3 Conservation Science

The term ''conservation science'' includes characterization of the constituent materials and production techniques of objects produced by humans, the study and understanding of decay processes, and the study and evaluation of conservation products and techniques.[44] It is generally taken to include all materials which might be put into a museum, such as ethnographic material and objects of industrial and military interest, as well as more conventional museum exhibits. It also includes issues surrounding the environmental monitoring of display conditions, impact of visitor numbers, and the like.[45] Chemistry is generally at the heart of the conservation process, since the first step in conservation is to stabilize the object by preventing any further degradation. This requires an understanding of the composition of the object itself, and also the mechanisms by which such objects degrade, which usually requires microstructural analysis, or the analysis of corrosion products. In the museum context, nondestructive (or quasi-nondestructive techniques) such as X-ray fluorescence (XRF) are often preferred for the analysis of objects, although microanalysis by inductively coupled plasma mass spectrometry (ICPMS) or neutron activation analysis (NAA) is of importance, with Raman spectrometry (RS) and IR spectrometry being used for structural information and the identification of corrosion products.

3.2 Archaeological Questions

3.2.1 Identification

Perhaps the simplest question that can be answered chemically is ''what is this object made from?'' The chemical identity of many archaeological artifacts may be obscured for a number of reasons. Simply, it may be too small, corroded, or dirty to be identified by eye. Furthermore, it may be made of a material which cannot be identified visually, or by simple (nondestructive) field tests. An example might be a metal object made of a silvery colored metal. It may be silver or a silver-rich alloy (in which case the precise composition may well carry information on debasement, which relates to economic history), but it may be an alloy designed to deceive, such as ''nickel silver'' (cupronickel alloys, such as are used in modern ''silver'' coinage). Conceivably, it might consist of some more exotic silvery metal, such as platinum, but this would excite great interest if identified in a European context prior to the mid-eighteenth century AD since this metal was supposedly unknown in Europe before then.

Thus, even the simple identification of a material may have important ramifications. In general, the required levels of analysis are relatively simple, subject to the usual constraints of archaeological materials. One preferred technique for many years has been XRF, because of its nondestructive nature (providing the sample can fit into a sample chamber), its restricted sample preparation requirements, and its multielement capability. During the 1960s an air path machine was developed specifically to allow the nondestructive analysis of larger museum objects,[46] and since then there have been attempts to devise a portable XRF system to use in museum displays or at an archaeological excavation, as well as for geological purposes.[47]

Identification of organic materials in archaeological contexts can pose more problems. Instances of identification of amorphous organic residues (either visible or occluded in another matrix) are addressed next. An example of a situation where the identification of the organically derived raw material used to manufacture artifacts is important is the discrimination between jet, shale and various forms of coal. Until 25 years ago, the classification of small pieces of jewelry made from various black materials was carried out by eye using a number of simple criteria such as color and structure.[48] Although there is little difficulty when applying these techniques to geological hand specimens, the small size of most archaeological finds and the nature of the destructive sampling such as thin sectioning or even streak testing often renders such judgments difficult to make. Such identifications are, however, rather important because of the restricted number of geological sources of jet. In the British Bronze Age, for example, if a piece of jet is identified in a Wessex burial

context in southern England, then it is taken as evidence of trading links with Whitby on the northeastern coast (approx. 400 km distant), since this is the only significant source of jet in England. Other similar materials, such as shales and the various workable types of coal, are more widely distributed. Analytical work, initially by NAA and then using XRF, showed that inorganic composition could be used to partially discriminate between these sources, and also showed that many of the original attributions were likely to be incorrect.[49] Subsequent work has refined the procedures,[50] and most recently further progress in characterizing such material[51] has been made using pyrolysis/gas chromatography/mass spectrometry (Py/GC/MS). Hindsight suggests that, given the nature of such materials, the use of organic analysis techniques may have yielded an earlier solution to the problem, but the approach taken reflects the trajectory of analytical work in archaeology, starting from studies of inorganic materials.

3.2.2 Provenance

As noted in section 2, the idea that the chemical composition of an archaeological artifact might be related to the geological source of the raw materials originated in the nineteenth century. Interestingly, this appears to be many years in advance of the development of the same idea in geochemistry. Quantitative study of the partitioning behavior of the elements between iron-rich and silicate-rich phases in the Earth's crust was carried out in the first half of the twentieth century and resulted in the geochemical classification of the elements as lithophile and siderophile. Much of this early work was summarized by Goldschmidt in his seminal work on geochemistry.[52] It was really not until this theoretical basis had been established that the concept of chemical provenance using trace elements acquired currency in geochemistry.

Perhaps an explanation of this can be found in the fact that the idea of provenance (based on stylistic or other visual characteristics) has a long history in archaeology, going back to at least the eighteenth century.[53] In the absence of any scientific means of dating artifacts in museum and private collections, a great deal of attention was paid to the observation of stylistic development, and the search for ''parallels'' in other collections. These effectively gave a relative chronology for a particular set of objects, and allowed proposals to be made about where certain objects might have originated, if they were deemed to be ''exotic'', or ''imports''. It is not unexpected, therefore, that with the advent of instrumental methods of analysis (primarily optical emission spectroscopy (OES), in the 1920s), the composition of an object was added to the list of characteristics which might be used to indicate either the provenance of the object (taken here to mean the place of manufacture), or the position of an object in some evolutionary sequence of form or decoration. Thus were born the great ambitious programs of analytical studies of ancient artifacts, perhaps typified by the Stuttgart program for the analysis of European Bronze Age metalwork during the 1950s.[54,55] Subsequently the analyses of other materials such as faience beads and ceramics were incorporated into analytical programs. During the second millennium BC, faience (sintered and glazed quartz) beads were distributed across prehistoric Europe, and they were thought to represent the contact of the more ''barbarian'' areas of Europe with the more advanced technologies of the Near East. In 1956 Stone and Thomas[56] reported on the use of OES to ''find some trace element, existent only in minute quantities, which might serve to distinguish between the quartz or sand and the alkalis used in the manufacture of faience and glassy faience in Egypt and in specimens found elsewhere in Europe''. Initial results suggested that OES data could not unequivocally distinguish between British and Near Eastern sources, but statistical re-evaluation[57] suggested a British origin for the beads found in England and Scotland on the basis of the quantities of tin, aluminum, and magnesium present. This was significant, because it suggested independent evolution of manufacturing technology, rather than long-distance trade.

The ''New Archaeology'' of the 1960s, emphasizing as it did explicit relationships between material culture and human behavior,[58] rejuvenated interest in scientific research into prehistoric trade and exchange. Invasion or diffusion of peoples was no longer viewed as the principal instigator of cultural change, but rather internal social processes. Evidence for social contact arising from exchange of artifacts and natural materials (as well as the transmission of ideas) was seen as important information in this debate, and one which scientific analysis might help evaluate. This initiated a golden era in archaeological chemistry, in which a wide range of scientific techniques were deployed in the hope of chemically characterizing materials such as obsidian, marble, ceramics, metals, glass, and natural organic materials including amber.[14] These characterization studies were aimed at ''the documentation of culture contact on the basis of hard evidence, rather than on supposed similarities of form''.[59] Perhaps most successful over the years has been the chemical characterization of ceramics, which has been mainly carried out by NAA.[60] As an indication of the geographical and temporal spread of such research, recent studies have included Islamic pottery from medieval Cairo,[61] Classic period pottery from the US Southwest,[62] Iron Age pottery from South Africa,[63] Chinese and European red stonewares,[64] and Roman glazed ceramics[65] and Greek and Cypriot pottery[66] from the Classical world.

Scientific characterization studies remain an important research area in archaeology, utilizing a range of trace element compositions as determined by increasingly sensitive analytical instrumentation, but now also including biomarker compositions and isotopic measurements. Despite the increasing sophistication of the analytical techniques, the fundamental limitations of the process must, however, be remembered. In order to be successful, the project requires carefully chosen samples to answer a well-constructed archaeological question, which in turn must be securely based on an appropriate archaeological model of the situation. Even if the archaeological side of the problem is well defined, there remain limitations as to what can be achieved. In his review of chemical characterization, Harbottle[67] reminded practitioners that:

> ''with a very few exceptions, you cannot unequivocally source anything. What you can do is characterize the object, or better, groups of similar objects ..., and also characterize the equivalent source materials, if they are available, and look for similarities to generate attributions. A careful job of chemical characterization, plus a little numerical taxonomy and some auxiliary archaeological and/or stylistic information, will often do something almost as useful: it will produce groupings of artifacts that make archaeological sense. This, rather than absolute proof of origin, will often necessarily be the goal.''

3.2.3 Manufacturing Technology, Date, and Authenticity

Another subset of questions which can be meaningfully addressed via chemical analysis relates to the determination of the technology used to produce an object. Often manufacturing technology can be adequately determined by careful visual and microscopic examination of the object, although experience has shown that laboratory or field simulations of ancient technologies are essential in order to fully understand ancient technologies, and can reveal some unexpected results.[68] Occasionally, however, chemical analyses are required, either of the object itself, or sometimes of the waste material from the process, such as the vast quantities of vitreous slag produced during iron manufacture. In this case a knowledge of the purity of the iron produced, the composition of the waste slag, and the composition of any residual slag included in the metal can be combined to give an understanding of the general nature of the technology involved (e.g. bloomery or blast furnace), as well as a more detailed knowledge of the operating conditions of the process.[69]

Given the increasing interest in our industrial heritage, and the resulting pressures to extend the legal protection and public explanation of its monuments, it is becoming more important to improve our understanding of the technological processes involved, including those from our recent past, some of which are now all but extinct. Experience has shown that even contemporary literary and patent evidence cannot always be taken as reliable, as has been shown by studies of the post-Medieval European brass industry. The traditional method for the manufacture of brass is known the ''calamine process'', introduced into Europe by the Romans. This is a small-scale procedure carried out in a sealed crucible, in which small lumps of copper metal are mixed with calamine (taken to be zinc carbonate or the roasted form, zinc oxide) and heated with charcoal. The zinc vapor is absorbed by the copper before it melts, therefore producing brass by a solid–vapor reaction. The more modern process is called the direct process, and involves mixing metallic zinc with molten copper. Because of thermodynamic restrictions in the calamine process, the maximum uptake of zinc into the brass alloy is limited to around 28–30%, whereas the direct process can be used to give any desired alloy of copper and zinc. Thus the chemical analysis of a brass object can be used to give an indication of the process by which it was made, and also some idea of date – European brass with more than 30% zinc is taken to be a product of the direct process, and therefore implicitly to date from some time after the introduction of that process into Europe. Extensive analyses of well-dated objects including scientific instruments and coinage has shown, however, that the British patent to manufacture brass by the direct process, dated 1738, was taken out some time after the actual introduction of the process into western Europe, which was probably around 1650.[13]

This rather crude analytical test to distinguish between manufacturing processes for brass is somewhat unsatisfactory, and cannot distinguish between calamine brass and brass made by the direct process but containing less than 30% Zn. There has been considerable interest in recent years in the possibility that certain high-temperature anthropogenic metal-producing processes might introduce measurable isotopic fractionation into the product.[70] Early interest concentrated on lead, and more recently on copper,[71] but theoretical studies and experimental observations on zinc have demonstrated for the first time that anthropogenic processes in brass manufacture might introduce sufficient differential isotopic fractionation of the zinc to allow the processing methodology to be distinguished.[72] If verified by higher-precision measurements, this observation has not only archaeological significance, but also wider implications for environmental geochemical monitoring.

The example of brass illustrates how the determination of manufacturing technology (by chemical or perhaps isotopic analysis) can also give an indication of the date of manufacture. More specifically, it gives an indication of a date before which a particular object could not have been manufactured, providing our understanding of the

appropriate ancient technology is accurate. This leads directly into the complex and sometimes controversial field of authentication of ancient objects, in which chemical analysis plays a large role. Thus any European brass object shown by analysis to contain more than 30% Zn must be dated to some time after the introduction of the direct process into Europe (remembering the uncertainty in the actual dates involved). This might be an extremely important consideration when judging the authenticity of a potentially valuable brass object. Perhaps the most famous example of authentication of a brass object is that of the ''Drake Plate'', so called because it was said to have been left by Sir Francis Drake to claim the San Francisco Bay area in the name of Queen Elizabeth I of England, dated to June 17, 1579. Analysis of the plate[73] by XRF showed it to have a very high zinc content (around 35%), with very few impurities above 0.05%. This was quite unlike any other brass analyzed from the Elizabethan period, which typically had around 20% zinc and between 0.5% and 1% each of tin and lead. It was therefore adjudged unlikely to be of Elizabethan manufacture (a view supported by that fact that it had a thickness consistent with the No. 8 American Wire Gage standard used in the 1930s, when the plate first appeared). In fact, European brass was imported into North America during the first half of the seventeenth century, and there have been several very successful analytical studies using the composition of such objects to map relationships between native North Americans and the early European traders.[74]

A wide range of archaeological materials has been subjected to authenticity studies.[75] Where possible, this takes the form of a direct determination of the date of the object, such as by radiocarbon dating for organic materials or thermoluminescence analysis for ceramics and the casting cores of cast objects. For metal objects in particular, it has of necessity taken the form of chemical analysis and comparison with reliably dated objects from the same period. Coins have been particularly subjected to such studies, since the variations in fineness for precious metal coinage can give a reasonably reliable calibration curve by which to date or authenticate other coins, and also because the fineness of the precious metals in circulation can give a great deal of information about the economic conditions prevalent at the time.[76] Authenticity has been a particular concern for all the major museums in the world, and most have facilities for carrying out a number of tests similar to those described here before making any acquisition. Considerably more questionable is the situation with respect to the commercial trade in antiquities, where access to scientific laboratories willing to carry out authentication on objects of undefined provenance has been partially blamed for encouraging the uncontrolled looting of some of the richest archaeological sites in the world.[77]

3.2.4 Chemical Analysis of Human Remains

The closest we can ever come to the physical presence of our ancestors is to excavate and study human bone, and it is therefore natural that the scientific study of human remains should be accorded special status in modern archaeological research. The well-known saying ''you are what you eat'' has been taken almost literally in archaeology for the last 20 years, and reconstruction of long-term dietary averages has been attempted using trace element levels in bone mineral and stable isotope studies on collagen and mineral in both bone and teeth.[78] Early on in this work it became apparent that inorganic trace element studies in bone were potentially bedeviled by post mortem diagenetic effects, the magnitude and significance of which continue to be debated.[79–83] Isotopic studies have been analytically far less controversial and, for Holocene material at least, appear to avoid most of the diagenetic problems encountered with trace elements.[84] Most authors conclude that if some collagen survives in a molecularly recognizable form, then the isotopic signal measured on this collagen is unchanged from that which would have been measured in vivo. There are several reviews of dietary reconstruction using isotopic measurements on bone collagen,[85–88] bone lipid[89] and bone and dental carbonate.[90] The chemical study of the protein and mineral fractions of archaeological bone and teeth can therefore reveal information on diet and health, providing that our knowledge of living bone metabolism is adequate, and that we can account for the changes which may occur during burial. Both of these factors provide significant scientific challenges to archaeological chemists.

The isotopic method of dietary reconstruction is based on the observation that carbon and nitrogen isotope ratios ($^{13}C/^{12}C$ and $^{15}N/^{14}N$) in bone collagen (and carbon in bone mineral carbonate) can reflect the corresponding isotopic ratios in the diet.[34] In principle these ratios can be used to distinguish between a reliance on terrestrial and on marine food resources in the diet of prehistoric humans. Carbon isotope ratios may further be used to differentiate between the consumption of terrestrial plants which photosynthesize using the C_3 pathway and of those using the C_4 pathway. Virtually all land plants photosynthesize using the C_3 pathway, but a few plants, mainly tropical grasses, have evolved the C_4 pathway as an adaptation to a hot dry environment. Archaeologically the most important C_4 plant is maize, which has its origins in central America, and carbon isotope measurements on human bone have been used with great success to

plot the spread of maize agriculture in prehistoric North American cultures.[86] The nitrogen isotope ratios in bone predominantly reflect the importance of marine inputs into the diet, but can also reflect the differential utilization of nitrogen-fixing terrestrial plants (legumes) and non-nitrogen-fixing plants. These models of isotopic ecology have also been applied to modern terrestrial and marine ecosystems, and have been shown to have great value as ecological markers, as well as having significant archaeological implications.

One significant recent development is the observation that, because different parts of the skeleton are resorbed and remodeled at different rates during life, the isotopic (and potentially the trace element) composition of different skeletal elements might reflect changes in diet with time. More specifically, permanent dental tissue in humans is formed during early life (up to about 18 years of age) and then is chemically unchanged in adulthood. Bone, on the other hand, is constantly remodeled during life, and although there are differences of opinion about the rate at which different skeletal components remodel, it is true to say that for a mature individual the rib isotopic composition (for example) should reflect the average diet over the last few years of life. If the location of a particular individual changes, such that the dietary balance is altered significantly, then this might be apparent from a chemical analysis of the skeleton. Such approaches have therefore been used to address the particularly significant question of human mobility in antiquity. Successful studies have used stable carbon isotopes in dental and skeletal tissue from prehistoric and colonial period individuals found in Cape Town,[91] or strontium isotopes in similar tissue from prehistoric Beaker period burials in southern Germany.[92,93]

The use of trace element data, and also strontium isotope data, from archaeological human tissue demands that we are able to discriminate between the in vivo signal and any subsequent post mortem changes. In an attempt to detect such post mortem changes, some studies have concentrated on the mineralogy of excavated bone. In view of the poor crystallinity and large surface area of living bone, it is not surprising that mineralogical changes take place post mortem, which might affect the trace element composition. For example, Sillen[94] has suggested that an increase in the Ca/P stoichiometric ratio of bone mineral to more than 2 may be due to recrystallization of the hydroxyapatite to other calcium phosphates, and the deposition of calcite crystals into the bone. It is not immediately clear what effect these processes will have on the supposed dietary indicator elements, but it is certain that they need to be better understood before convincing archaeological interpretations can be made. In order to address these processes, it has become common to analyze the soil material from the burial context in an attempt to assess the elemental exchange between sediment and bone. This is clearly not of itself adequate and some workers have emphasized the need to consider the availability of exchangeable ions rather than the total soil composition.[95] This requires an understanding of the physics and chemistry of the soil, including factors such as soil pH and temperature, the presence or absence of organic matter, as well as the geochemistry of the species present and the local groundwater movement, as discussed elsewhere in this review.

Many physical and chemical parameters have been measured on archaeological bone to identify diagenetically altered material (C/N ratio, organic content, amino acid profile of collagen, Ca/P ratio, porosity, mechanical strength, histology, density, ''crystallinity'').[96] Some progress has been made on the physical nature of the interaction between groundwater and buried bone,[97] but this needs to be augmented by better geochemical modeling, using information derived from laboratory experimentation on both the organic and inorganic components of the system. On the assumption that most archaeological bone is mineralogically altered in some way, a highly significant question is ''can this alteration be satisfactorily removed in order to allow trace element or isotopic analysis of the in vivo mineral phase?''. A common approach to this problem (at least in the context of Sr and Sr isotope measurements on fossil and archaeological bone) is a process known as ''sequential washing'',[98] which has been demonstrated to give a residual strontium level that is chemically and isotopically plausible for an in vivo signal. This raises a number of questions focusing around the nature of such recrystallization processes (which are presumably driven by Ostwald processes during the burial period), and the differential solubility of various substituted calcium phosphates, which might be susceptible to a series of controlled laboratory experiments.[99]

3.2.5 Organic Analysis in Archaeology

Until recently, archaeological chemists have generally paid more attention to the analysis of inorganic artifacts – encompassing both natural (e.g. stone) and synthetic (e.g. ceramics, metals, glass and glazes) materials. These represent the most obviously durable artifacts in the archaeological record. However, in recent years, increasing attention has been directed at biological materials, including natural products such as waxes and resins, accidental survivals such as food residues, and, above all, human remains, including bone, protein, lipids, and DNA. Some of the methodology for this work has been imported not only from chemistry, biochemistry, and molecular biology, but also from organic geochemistry,

which has grown from a discipline interested in elucidating the chemical origins of oil and coal into one which studies the short-term alteration and long-term survival of a very wide range of biomolecules.[100]

Another related discipline in this quest for ancient biomolecular information is molecular paleontology. Collectively, these disciplines and archaeology have much to offer each other, particularly in the recovery of genetic information from animals and plants. In terms of specific archaeological interest, the ability to extract DNA has considerable significance. Hitherto, extraction of nucleic acids from bone some 40 000 years old has been claimed.[101] Preserved soft tissue and seed remains also yield extractable DNA. Specific DNA sequences can be targeted, amplified using the polymerase chain reaction (PCR), and compared with sequences in other individuals and modern specimens. However, ancient DNA is severely damaged and fragmented. Contamination of aged samples and extracts with modern DNA is a serious problem and whilst the study of DNA in archaeological samples will constitute a major area of future activity in the discipline,[102] current research will continue to focus on the authentication of samples of ancient DNA[103] and understanding the extent of DNA damage.[104]

Preservation of a wider range of biomolecules has been demonstrated in a number of archaeological contexts. In particular, proteins preserved in human bone have been subject to immunological investigation.[105,106] The survival of recognizable protein residues on stone tool surfaces[107,108] hints at the possibility of characterizing artifact use and identifying utilization of particular animal resources and dietary items, although the specificity of the results obtained remains contentious.[109]

Organic residues represent amorphous biological debris in the archaeological record. They differ in one crucial respect from the main body of organic remains which archaeologists have studied for many years in that they lack the macroscopic cellular structure present in seeds, wood, leather, pollen, and so on. Typical organic residues include food deposits surviving in pottery containers used for cooking, storing, and serving solid and liquid organic products, the balms in the wrappings of mummified bodies, and traces of coloring dyes impregnated in ancient textiles.

The most effective approach to the identification of organic residues is molecular analysis. Ideally, the presence of a specific compound or distribution of compounds in an unknown sample is matched with its presence in a contemporary natural substance. The use of such molecular markers is not without problems since many compounds are widely distributed in a range of natural materials, and the composition of an ancient residue may have changed significantly during use and burial. In general, molecular markers belong to the compound class defined as lipids, a heterogeneous group of molecules which includes fats and oils and often includes compound classes with common solubilities, such as the constituents of resins and waxes.

Combined gas chromatography/mass spectrometry (GC/MS) has been used widely in the identification of ancient lipid residues contributing to studies of artifact use patterns and food consumption.[110] The potential for preservation is relatively high and a wide range of fatty acids, sterols, acylglycerols, and wax esters has been identified in visible surface debris on pottery fragments or as residues absorbed into the permeable ceramic matrix. Isolation of lipids from these matrixes is achieved by solvent extraction of powdered samples. Despite their stability relative to other classes of biomolecules, lipids often undergo alteration, with the reactive sites undergoing oxidation (leading to the loss of double bonds) or hydrolysis (cleavage of ester linkages). Lipid degradation combined with other factors, such as the mixing of fats from diverse sources, has made precise identifications problematic. Often it is only possible to conclude that animal or plant lipid, or indeed a mixture of the two, is present. In certain circumstances, organic products consistent with specific sources can be identified due to the presence of long-chain waxy compounds found in epicuticular waxes of plants.[111] The recent introduction of gas chromatography/combustion/isotope ratio mass spectrometry (GC/C/IRMS) has shown great promise in differentiating ancient lipid residues (e.g. discriminating ruminant from nonruminant animal fats).[112]

Lipids can be used to inform studies of decay sequences and decay products associated with human remains, notably preserved soft tissue from peat-buried bog bodies and soft tissue remains in permafrost. Lipid analysis of the skin tissue from the Tyrolean Ice man has been undertaken. Although many of the ester bonds of the triacylglycerols had been hydrolyzed releasing the fatty acids, some preservation of acyl lipids was evident. However, all triacylglycerols with more than one double bond were completely degraded. Combined histological (showing loss of epidermis) and chemical investigations (revealing transformation of the fat into adipocere) indicated submersion of the body in water for at least several months followed by freeze-drying.[113]

Not all the lipids associated with human remains are endogenous. Recent work[114] suggests that mycolic acid lipid biomarkers for tuberculosis (TB) (from *Mycobacterium tuberculosis*) are preserved in archaeological bone from the burial ground (18th–19th century) at Newcastle Infirmary (UK). Mycolic acids are molecules with 60–90 carbon atoms. These acids are mainly esterified to a polysaccharide, which is in turn linked to the cell wall. The authors report the chemical identification of species-specific mycolic acids in 5 out of 21 individuals, which

correlates with the documented frequency of 27.1% TB for the Infirmary. None of the rib samples in the Newcastle study had lesions that might be linked to TB, indicating that TB would not have been diagnosed without the molecular study. The mycolic acids have been ''fingerprinted'' using high-performance liquid chromatography (HPLC) and this approach is highly complementary to studies based on genetic indicators of ancient disease.

The versatility of lipid molecules as indicators of specific human activities has been demonstrated in the persistence in soils and sediments of biomarkers of fecal material. Remarkably, ratios of certain biomarkers (α- and β-stanols) and the relative abundance of others (bile acids) suggest that it is possible to provide an indication of the animal donor to the archaeological record![115] The occurrence and persistence of cholesterol in long-buried bone and the ability to determine compound-specific $\delta^{13}C$ values is being used as a source of paleodietary information complementary to those obtained on collagen or bone mineral with their longer turnover times. Furthermore the unaltered carbon skeleton of ancient cholesterol ensures that chemical modification has not occurred.[116] Food lipids are by no means the only source of amorphous organic debris. Higher plant resins and their heated derivatives (wood tar and pitch) served as sealants and adhesives, perfumes, caulking materials, and embalming substances. The protective functions of resinous substances in nature reflect their widespread survival in waterlogged contexts. The use of a tar derived from heating birch bark (*Betula* sp.) has been demonstrated in prehistoric Europe from the early Holocene onwards.[117] This tar served as a ubiquitous hafting adhesive for attaching stone tools to handles of wood, bone, or antler. Birch bark tar is also the source of chewing ''gums'' excavated from bog sites of Mesolithic date in southern Scandinavia. Recent historical evidence suggests that chewing the tar may have played a role in dental hygiene and in treating throat disorders. Molecules consistent with beeswax (*Apis mellifera*) have been identified on a pottery vessel dating to the fourth millennium BC in Europe and provide among the earliest evidence in antiquity for the collection of wax and, by assumption, honey.[118] Resins consistent with those of *Pistacia* sp. have been identified as a part of the bulk cargo of a Late Bronze Age shipwreck.[119] Finally, a find of frankincense from a sixth century AD site in Egypt has been identified by organic analysis.[120]

Although food lipids, plant resins, and wax comprise the largest categories of organic debris identified by GC/MS, there are other striking identifications. Biomarkers identified from plant extracts with psychoactive properties have been reported. By way of example, lactones congruent with the intoxicating drink kava (prepared from the roots of *Piper methysticum*) have been identified in residues adhering to pottery fragments from Fiji.[121] More systematic investigations have been undertaken on bituminous substances.[122] Bitumen represents the fraction of sedimentary organic matter that is soluble in organic solvents (the insoluble fraction is known as kerogen). The liquid or semisolid varieties of bitumen were widely used in the Near East and Middle East in antiquity, serving as a multipurpose glue and waterproofing material, a building mortar, medicinal agent and as one of the constituents of the organic preparations applied to mummified bodies in Ancient Egypt dating from 1000 BC to 400 AD. Compounds consistent with a bituminous substance include saturated hydrocarbons which have linear (alkylated alkanes) or cyclic (steranes, terpanes) carbon skeletons. These molecules derive largely from microscopic plants deposited in the sediments as well as from bacterial inputs. It has proved possible to identify molecular and isotopic characteristics of the bitumen which enable archaeological finds to be assigned to a particular source of bitumen.[123]

Other methodologies have contributed to identification of residues. In particular, traces of wine have been deduced from chemical ''spot tests'' for tartaric acid, supported by IR spectroscopy, ultraviolet/visible (UV/VIS) spectroscopy and HPLC. Positive results have been reported with a thin yellowish deposit on a shard from a Neolithic jar (5400–5000 BC) from the site of Hajji Firuz Tepe in the Zagros mountains, Iran.[124]

A range of spectroscopic techniques have been useful in archaeological investigations. Although IR spectroscopy can be limited for complex mixtures comprising many degraded components, the technique offers some possibilities for screening organic residues and ''fingerprinting'' organic substances. In particular, IR spectroscopy has been very successful in characterizing Baltic amber found at prehistoric sites in Europe and distinguishing it from other European fossil resins.[30] A related technique, Raman spectroscopy, has been used to study archaeological materials nondestructively.[125] UV/VIS absorption spectroscopy has been useful in confirming the identity of dyestuffs from the New and Old Worlds. Nuclear magnetic resonance (NMR) has provided valuable information on the structure of highly cross-linked fossil resin samples from around the world.[126] This work builds upon the classic IR spectroscopic program of analysis that began in the 1960s, and identified the source of origin of amber from archaeological sites throughout Europe.[127]

3.3 Special Considerations in the Analysis of Archaeological Material

For the most part, archaeological materials are now analyzed on factory-standard equipment – sometimes on a machine dedicated to archaeological research,

but often on a multiproject instrument. In this sense, archaeological materials are no different from any other environmental or geological samples which require analysis. In other ways, however, archaeological material can pose special problems to the analyst. Restrictions are often placed on sampling either by physical limitations or by consideration of aesthetic value. Typically the resulting samples can be far from ideal from the analytical point of view – small, fragmentary, and, particularly in the case of biological samples, often considerably degraded. They are likely to be contaminated in some way during burial, and after recovery (postexcavation) may suffer further contamination due to storage media, handling, or airborne particles. Samples taken from museum material which was collected some time ago may have been subjected to unknown conservation, restoration, or fumigation procedures. These ubiquitous problems of degradation and contamination, although not insoluble, make archaeological chemistry a challenging field, and usually not one which can be regarded as just another routine analytical application.

In many ways, there is a close relationship between forensic chemistry and archaeological chemistry. Both derive evidence from samples obtained from a controlled recovery situation, and both attempt to reconstruct patterns of human behavior from material evidence. In both cases, too, the samples are often unique, and often far from ideal from an analytical point of view. Attempts have been made to draw together the two disciplines of archaeology and forensic science,[128] but an examination of the analytical literature on both subjects reveals as yet little evidence of cross-fertilization.

4 SPECTROMETRIC METHODS USING OPTICAL WAVELENGTHS

Analytical chemistry began to make a sustained impact on the subject of archaeology following the introduction of OES as a method of analysis in the 1930s. The principal interest at the time was understanding the level of technology represented by finds of ancient metalwork, especially in terms of alloying, and systematic programs of analysis were initiated in Britain and Germany leading to substantial compilations of analytical data.[129] Although now outmoded as an analytical technique, it provided the bulk of analyses (in Europe at least) of archaeological metals and pottery until the 1970s, and was particularly prominent in the large-scale study of the composition of Greek pottery carried out by the Fitch Laboratory of the British School at Athens and the Research Laboratory for Archaeology and the History of Art, Oxford University.[66] Using OES, it was possible to measure virtually any element present in a 10 mg sample, in the concentration range from 0.001% to 10%. In practice, because of the complexity of the emission spectra, the maximum number of elements measurable from a single exposure was about 20, with a coefficient of variation of between 5% and 25% for major and minor elements.

During the 1970s and 1980s most OES laboratories switched to atomic absorption spectrometry (AAS), and this became one of the standard methods for the analysis of archaeological metals and pottery,[130] with a growing application to bone in the 1980s. In comparison with OES or NAA (the principal competitor techniques until the advent of inductively coupled plasma (ICP) techniques in the 1980s), AAS had the disadvantage that it required a liquid sample (causing dissolution problems for silicate materials), and that it was sequential in operation, thus limiting the range of elements which could be determined. Nevertheless, a substantial database of archaeological analyses has been compiled using AAS, and it has only been replaced by inductively coupled techniques in archaeology during the 1990s.

Temperatures up to 10 000 °C are achievable using an ICP torch in an inductively coupled plasma atomic emission spectrometer (ICPAES). At these temperatures, all compounds are completely dissociated and excited, so that they strongly emit characteristic lines, which are subsequently dispersed and detected using a single detector system similar to that used in atomic absorption. Although strictly sequential in operation, the combination of software-controlled analysis programs and automated sampling makes the machine quasi-simultaneous. Further developments have enhanced its performance considerably, such as the use of a mass spectrometer as the detection system, and the recent development of laser ablation (LA) systems that allow a solid sample to be precisely vaporized by a high-power laser, which clearly has a great deal to offer archaeological analysis.

A major benefit of the ICP torch which has been exploited over the past few years is that it can be connected up to a mass spectrometer to give the very powerful technique of ICPMS. The temperature of the torch (up to 10 000 °C) is sufficient to ionize approximately 50% of the atoms in the sample, making the plasma an ideal source of ions which, given the appropriate interface to deal with the high temperatures involved, can be extracted from the plasma and injected directly into a mass spectrometer. Not only is this more sensitive than the measurement of emission intensities (thus lowering even further the detection limits for certain elements), but it also allows the analyst to monitor the concentration of individual isotopes of a particular element. It also enables the direct measurement of the ratios of specific isotopes for a particular element, such as lead.

There have been several comparative reviews of the performances of AAS and the ICP technique, including a recent one by Slavin.[131] Although ICPAES usually has a limit of detection comparable with flame AAS, it can be seen that an element-by-element comparison often reveals significant differences. In general, the refractory elements (Al, Ti, etc.) are better detected by ICPAES, whereas the heavy (nonrefractory) elements are better detected by flame AAS. For example, with ICPAES Ba has a detection limit of around 0.5 ppb (parts per billion), compared with about 8 ppb by flame AAS, whereas Pb is better detected by AAS (10 ppb compared with about 50 ppb). The use of an electrothermal furnace with AAS generally gives an order of magnitude improvement in limits of detection, and the coupling of a mass spectrometer to an ICP torch gives a further order of magnitude improvement. Equally (if not more) important are the analytical precisions of the various techniques – a measure of the repeatability of the experimental method. In routine applications, the precisions are usually quoted as around 0.5% for flame AAS, 1.5% for ICPAES, 3% for furnace AAS and 2–3% for ICPMS, although individual laboratories may be able to improve substantially on these figures for particular applications. Recently, a review has been published concerning the intercomparability of silicate analyses carried out by ICPAES and AAS for the purposes of comparing data on archaeological ceramics.[132] It concluded that the results from both methods were sufficiently close that common databanks could be established, providing adequate care had been taken to include certificated standards within each run to monitor performance on the more difficult elements.

LA is potentially extremely attractive for archaeological applications, since it offers the possibility of spatially resolved microanalysis of solid samples, in a manner similar to electron microscopy, but with greater sensitivity and the potential for isotopic analysis. In this technique, a high-energy pulsed laser is directed onto a solid sample, with a beam diameter of less than 25 μm. The pulse vaporizes about 1 μg of material, to leave a crater 50 μm deep. The vaporized sample is swept into the ICP torch via a carrier gas. Current equipment allows solid samples, approximately flat, of 35 mm diameter to be accommodated in the LA unit. The analytical sensitivity in terms of minimum detection limits is poorer for LA than for solution analysis, and the detection limit for a particular element (or isotope) depends on the solid matrix being analyzed. Nevertheless, the potential advantages of such a system far outweigh these problems, and laser ablation inductively coupled plasma mass spectrometry (LA/ICPMS) is likely to become a powerful technique in archaeology, as in all other materials sciences. The potential of the new generation of high-resolution ICPMS instruments using multicollector detection is discussed in section 5.

5 X-RAY AND ELECTRON TECHNIQUES

XRF has been a popular technique for the chemical analysis of archaeological artifacts since the early 1960s, largely because it was perceived as nondestructive providing a reasonably flat and clean part of the object could be presented to the spectrometer. Early instruments for archaeological research used wavelength-dispersive detection, and were soon adapted to use a curved crystal to reduce the spot size of the analyzed area from a circle of ca. 5 mm diameter down to a couple of millimeters (the so-called "Milli-probe"[133]). The use of energy-dispersive X-ray fluorescence (EDXRF) instruments was pioneered during the 1970s,[134] largely using home-made mounting equipment so that a large sample could be orientated toward the X-ray tube and detector to give maximum flexibility of sample size. Some machines were designed to use a large vacuum chamber enclosing the sample (in which cases elements down to sodium could be detected, but not with great sensitivity). Others operated in a nonevacuated environment ("air path"), which limited the analyses to elements heavier than potassium in the periodic table, but did allow the majority of ancient metal alloys to be analyzed in a rapid fashion. Yet others experimented for a while with an enclosing bag containing helium, which allowed analysis down to about silicon. Several large analytical programs on materials such as metals[135] (especially coinage[136]), ceramics,[137] porcelain[138] and glass,[139] obsidian,[140] and even jet[141] were carried out using XRF during the 1980s and later. XRF continues to be a standard tool for major and minor elements in geological whole-rock analysis where the sample can be pelletized or standardized in some way,[142] and this method has been used to some extent for the routine analysis of ceramics and lithic raw materials in archaeology. XRF is still a valued tool in many museum and archaeological laboratories since it can provide an almost instantaneous semiquantitative identification of a wide range of inorganic (and occasionally organic) objects. As a research tool, however, electron microscopy is often the preferred technique, for the following reasons.

One major consideration in the archaeological use of EDXRF on essentially unprepared archaeological samples such as coins and glass has been the problems of surface sensitivity. Although by no means the most surface-sensitive of analytical techniques, EDXRF only gives an analysis of the top fraction of a millimeter of the sample. In the case of metals, where the phenomenon of surface enrichment of the more noble metals (either deliberately during manufacture[143] or naturally as a

result of interaction with the burial environment[144]) has long been known, this can be a critical restriction. Similar problems, due to the selective leaching of the alkali elements, have been noted in glass.[145] Although careful sample preparation can minimize these difficulties and provide good quantitative data, it is usual to regard EDXRF analyses of unprepared archaeological materials as qualitative, or at best semiquantitative. This does not compromise its usefulness in areas such as conservation, where a rapid identification of the material may be all that is required, or in other areas as a preliminary analytical screening technique.

The analytical uses of the electron microscope in archeology were developed at about the same time as the development of wavelength-dispersive X-ray fluorescence (WDXRF), again using specially developed equipment.[146] More recently, however, standard instrumentation has been used, especially since the introduction of large environmental chambers, which ease the sample size restrictions found with conventional equipment. Electron microscopes can be operated with either a wavelength-dispersive (WD) X-ray detector or an energy-dispersive (ED) detector. More powerful instruments are fitted with both. Comparative studies have been made of the performance of WD and ED detectors in the analysis of silicates[147] and vitreous material.[148] These studies show that the accuracy and precision of both techniques are comparable over the normal concentration ranges found in geological and archaeological material, but ED has considerably poorer limits of detection, typically by one or two orders of magnitude – representative figures for the limit of detection by ED detection are 0.05–0.26 wt% of the element.

The principal advantage conferred by electron beam stimulation is the ''steerability'' and ''focusability'' of the primary beam. Not only does this allow analysis of small regions of the sample, but it also allows line scans and area scans to be carried out, giving spatially resolved chemical information from the surface of the sample. Line scans across sections through the object are particularly valuable in the study of the changes caused by corrosion of metal artifacts. Area scans give rise to ''elemental maps'' of the sample surface. Thus, for example, the distribution of uranium can be displayed across the surface of a section through an archaeological bone. This facility is particularly valuable in archaeology, where the samples tend to be inhomogeneous and corroded.

Transmission electron microscopy (TEM) has had as yet limited application in archaeology – an exception is the work of Barber and Freestone,[149] who used TEM to identify the nature of the tiny particles involved in the phenomenon of dichroism in the Roman glass known as the Lycurgus Cup. Small dispersed particles of silver–gold alloy in the glass make this object appear red in transmitted light and green in reflected light, and made it one of then technological marvels of the ancient world.

During the 1980s an alternative approach to the X-ray analyses of inorganic materials was developed, utilizing high-intensity, highly focused beams of protons. These beams can be ''tapped off'' from a particle accelerator and focused onto a sample outside the accelerator (i.e. not in a chamber under high vacuum). This is valuable for archaeological material, since it again removes the need for sampling. The first generation of proton-induced X-ray emission (PIXE) machines had beam diameters of the order of half a millimeter,[150] but more recent machines have micrometer-diameter beams (μ-PIXE), giving spatial resolution similar to that obtainable by electron microscopy. The major advantage offered by PIXE is that the primary beam of protons does not generate a high X-ray background, giving improved analytical sensitivity. Detection levels in PIXE may be as low as 0.5–5 ppm (parts per million) for a wide range of elements in thin organic specimens such as biological tissue. Limits of detection better than 100 ppm have been reported for elements above calcium in the periodic table in inorganic archaeological material. Analyses of archaeological and historical art samples ranging from ceramics, metals, and paintings to postage stamps have been reviewed by Johansson and Campbell.[151]

One other analytical use of X-rays is in the technique of X-ray diffraction (XRD), used to identify crystalline phases in a range of archaeological objects. It has been used extensively (in conjunction with IR spectroscopy) to monitor changes to the crystallinity of archaeological bone as a monitor of diagenetic processes.[152] It is also used (in conjunction with IR and Raman spectroscopy) to identify the various mineral phases in corrosion products on metal surfaces[153,154] and also to identify the high-temperature phases present in archaeological ceramics as an indication of firing temperatures and techniques.[155]

6 NUCLEAR TECHNIQUES

Until the advent of ICP and PIXE techniques during the 1980s, the standard analytical method for producing multielement analyses with detection limits at the ppm level or better was NAA, and it has consequently been widely applied in all disciplines from geology to medicine. It has been popular because it applies to a wide range of materials, giving analytical sensitivities down to the ppb level for some elements on small solid samples. One review showed that 50% of the trace element data reported in the major geochemical journals between 1990 and 1992 were obtained by NAA.[156] NAA has been used on archaeological material since the inception of

the technique in the 1950s, particularly for coinage[157] and ceramics.[158] It has been particularly prominent in the USA, where very few inorganic analyses of any sort have been published using a technique other than NAA. It is still a major technique in archaeological chemistry but increasing difficulties associated with obtaining irradiation facilities, and increasing competition from ICPMS, are eroding this position.

The procedure in use at the British Museum for ceramic analysis is typical of the methods used on a range of materials.[159] Powdered samples (40–80 mg) removed from the ceramic are sealed into silica tubes and packed into an irradiation canister, which can contain up to 70 samples. The canisters are irradiated at a nuclear reactor, and returned to the laboratory four days after irradiation, for a program of γ counting using a germanium solid-state detector. The British Museum program allows 23 elements to be determined from the same irradiation, making the whole analytical procedure extremely economical in terms of information obtained for the cost and time involved. Detection levels can be as low as 1.5×10^{-5} ppb for very sensitive elements in a suitable matrix[160] but more typical figures would be around 10 ppb, up to perhaps 10 ppm, for trace elements in geological or biological material. As with all analytical techniques, there is a variation in the sensitivity and detection levels from element to element, but additionally with NAA there are some elements which cannot be ''seen'' at all, either because neutron irradiation does not produce suitable radioactive nuclei, or possibly because the spectrum has severe spectral interference. Most important archaeologically here are elements such as lead and silicon, which means that NAA often cannot produce ''total analyses'' for metals, ceramics, and glasses. Some of these problems can be overcome using variations on NAA, such as prompt gamma NAA, but often other methods have to be used for some elements. Despite this, NAA remains the method of choice for many analytical provenance studies, especially for ceramics.

7 MASS SPECTROMETRY

Mass spectrometry (MS) is becoming increasingly important in archaeology, as in other scientific disciplines, both as a technique in its own right, and also as a sensitive detection system in the so-called ''hyphenated techniques'' such as ICPMS, GC/MS, etc. It is has been of great importance in archaeology for the determination of heavy stable isotope ratios such as lead,[161] or for light stable isotopes (carbon, nitrogen) as used in dietary reconstruction[34] (see section 3.2.4 for the latter). For the heavy elements, a solid source is needed, and this is the basis of thermal ionization mass spectrometry (TIMS), which has become of prime importance in isotope geochemistry. Modern instrumentation uses highly sophisticated systems, such as triple filaments for increased beam stability, and multiple collector devices to monitor several isotope ratios at the same time, which allows samples as small as 10^{-9} g to be measured satisfactorily.[162] Lighter elements (up to sulfur) require a gaseous sample, which is prepared by combustion of the sample off-line and injection directly into the spectrometer. Again, multiple collectors are used for increased stability in isotope ratio monitoring. A significant development in the measurement of carbon and nitrogen stable isotope ratios has been the coupling of an on-line combustion system and gas chromatography (GC) interface onto the mass spectrometer, which allows the direct input of solid samples. Not only does this considerably simplify the sample preparation, but it greatly speeds up the throughput of measurements with little loss of precision.[163] In the past five years, as discussed above, continuous-flow isotope ratio mass spectrometry (CFIRMS) has almost revolutionized the isotopic study of human remains, as well as allowing more rapid analysis of other organic materials such as cellulose in tree rings for the purposes of climatic reconstruction.[164]

MS has played an expanding role as an alternative detection system offering much greater sensitivity (as well as isotopically resolved measurements, if required) in a wide range of analytical applications. ICPMS instrumentation using relatively low-resolution quadrupole detection has been used in environmental applications to measure isotope ratios of heavy elements such as Pb in body fluids, plant material, and dust samples to determine the source of metal contamination.[165] It is generally accepted that quadrupole ICPMS measurements of lead isotope ratios are sufficiently precise to detect differences between sources of pollution, but interlaboratory studies[166] of ICPMS measurements on a range of environmental samples suggest that realistic precisions are 0.3% for $^{206}Pb/^{207}Pb$, 0.8% for $^{206}Pb/^{204}Pb$ and 1.4% for $^{208}Pb/^{204}Pb$, which are roughly an order of magnitude poorer than can be achieved for lead isotope measurements by TIMS. This is, however, changing rapidly, as a new generation of ICPMS machines becomes available, which have an ICP source linked to a double-focusing multiple collector ''conventional'' mass spectrometer, giving precisions expected to be even better than those achievable by TIMS.[167,168] The first commercial versions of these machines have recently been commissioned, and are certain to have a tremendous impact in archaeology, as elsewhere, over the next few years.[169]

Another important ''hyphenated technique'' is GC/MS, achieved by coupling a mass spectrometer to the effluent of a GC set-up, which combines the powerful analytical capability of MS with the high degree of separation

possible with GC. Combined GC/MS has become the workhorse of organic analysis in archaeology, with more than 100 papers published using GC/MS applications during the 1990s. The majority of these investigations have made use of electron ionization mass spectra, although more sophisticated ionization and spectrometric methods of detection are being applied. Analytical pyrolysis is increasingly being used in conjunction with GC/MS. Pyrolysis is the preferred method for breaking up the insoluble or polymeric fractions of organic residues that are not themselves volatile enough for conventional analysis, and thereby allowing separation and identification of the fragments. Py/GC/MS has been applied successfully to the study of fossil and recent higher plant resins and to the macromolecular debris remaining from the burning of food in archaeological pottery vessels.[170] GC/C/IRMS provides an extremely versatile approach to studying isotopic patterns in individual molecules.[171]

8 CHROMATOGRAPHY AND OTHER ANALYTICAL TECHNIQUES

Both preparative and instrumental methods of chromatography have been widely applied in archaeological analysis. Some methods such as thin-layer chromatography (TLC) can be used to separate and isolate molecular compound classes in complex mixtures. The method is less satisfactory for actual identification purposes. GC was employed as early as 1965 to identify wood tars from Neolithic potsherds.[172] The technique remains useful in screening solvent extracts and offers some value in ''fingerprinting'' certain substances (such as beeswax). High-temperature GC employing capillary columns enables a wide range of lipid compound classes to be separated. Most often, flame ionization detection (FID) is employed to monitor the column effluent. HPLC is more versatile than GC because the sample components need not be in the vapor phase. HPLC has been used in archaeology for the separation of amino acids and peptides (for the purposes of dating and amino acid racemization studies), the analysis of ancient dyes and even the identification of alkaloids (e.g. caffeine and theobromine) characteristic of cacao preserved in Mayan archaeological ceramics.

RS has been used increasingly in archaeology as in other disciplines to characterize a wide range of organic and inorganic substances.[173] Materials studied include pigments from Medieval English wall paintings[174] and Roman tombs,[175] human teeth,[176] hair[177] and skin,[178] waxes and resins,[179] linen textiles,[180] and ceremonial polished greenstone ax-heads from Mesoamerica.[181] Similarly, IR spectroscopy (especially Fourier transform infrared, FTIR) has been used to characterize and monitor the degradation of materials such as bone,[182,183] but also materials of conservation interest such as adhesives.[184]

Electron spin resonance (ESR) is probably best known in archaeology as a dating technique, where it provides crucial evidence about human evolution in the time period before radiocarbon dating becomes feasible.[185] It has, however, been used as an analytical tool to provenance lithic raw materials such as marble[186] and jasper.[187] It has also been used to detect and measure the degree of heating of ancient materials[188] such as ceramics,[189] bone,[190] soil,[191] and even cereal grains.[192] NMR has been sparingly applied to archaeological material, but solid-state ^{13}C-NMR has been used to look at organic materials encountered during conservation.[193] Perhaps inevitably, NMR has also been applied to the study of the degradation of archaeological bone.[194]

9 SUMMARY

Although chemical analysis of archaeological artifacts is recognized as a fundamental component of the inquiry into past human behavior and development, the demand for relevant data has been echoed on many occasions. Scientific analysis should be much more than a descriptive exercise which simply documents the date, morphology, or composition of ancient materials. As DeAtley and Bishop[195] have pointed out, no analytical technique has ''built-in interpretative value for archaeological investigations; the links between physical properties of objects and human behavior producing the variations in physical states of artifacts must always be evaluated.'' The demand for meaningful scientific data also needs to be viewed against the evolution of the discipline of archaeology. Trigger[5] states that ''archaeologists have asked different questions at different periods. Some of these questions have encouraged close relations with the biological and physical sciences, while other equally important ones have discouraged them.'' Against this backdrop, the contribution of scientific analysis to the study of the past has come under increasing scrutiny. Although the majority of archaeologists acknowledge the contribution of scientific dating and analytical techniques to increasing the information potential of the past, the central concern prevails that scientific studies sometimes proceed in a context substantially devoid of a specific archaeological problem.[196] It would, however, be equally misleading to suggest that chronological, compositional, or locational data generated by scientific techniques have no role to play in providing foundations for interpretations of past human behavior. Perhaps success in archaeological chemistry is difficult to measure. For

some, it may be the implementation of a robust or elegant scientific methodology; for others it may be the degree of integration within archaeological problems. Ideally, it should display characteristics of both. The promotion of genuine interdisciplinary studies rather than multidisciplinary investigation lies at the heart of all archaeological endeavor.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
CFIRMS	Continuous-flow Isotope Ratio Mass Spectrometry
ED	Energy-dispersive
EDXRF	Energy-dispersive X-ray Fluorescence
ESR	Electron Spin Resonance
FID	Flame Ionization Detection
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
GC/C/IRMS	Gas Chromatography/Combustion/ Isotope Ratio Mass Spectrometry
GC/MS	Gas Chromatography/Mass Spectrometry
HPLC	High-performance Liquid Chromatography
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectrometer
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
LA	Laser Ablation
LA/ICPMS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
MS	Mass Spectrometry
NAA	Neutron Activation Analysis
NMR	Nuclear Magnetic Resonance
OES	Optical Emission Spectroscopy
PCR	Polymerase Chain Reaction
PIXE	Proton-induced X-ray Emission
ppb	Parts Per Billion
ppm	Parts Per Million
Py/GC/MS	Pyrolysis/Gas Chromatography/ Mass Spectrometry
RS	Raman Spectrometry
TB	Tuberculosis
TEM	Transmission Electron Microscopy
TIMS	Thermal Ionization Mass Spectrometry
TLC	Thin-layer Chromatography
UV/VIS	Ultraviolet/Visible
WD	Wavelength-dispersive
WDXRF	Wavelength-dispersive X-ray Fluorescence
XRD	X-ray Diffraction
XRF	X-ray Fluorescence

REFERENCES

1. C. Renfrew, P. Bahn, *Archaeology: Theories, Methods, Practice*, 2nd edition, Thames and Hudson, London, 1996.
2. E.R. Caley, 'Klaproth as a Pioneer in the Chemical Investigation of Antiquities', *J. Chem. Ed.*, **26**, 242–247, 268 (1949).
3. E.R. Caley, 'Early History and Literature of Archaeological Chemistry', *J. Chem. Ed.*, **28**, 64–66 (1951).
4. E.R. Caley, 'The Early History of Chemistry in the Service of Archaeology', *J. Chem. Ed.*, **44**, 120–123 (1967).
5. B.G. Trigger, 'Archaeology's Relations with the Physical and Biological Sciences: A Historical Review', in *Proceedings of the 26th International Archaeometry Symposium*, eds. R.M. Farquhar, R.G.V. Hancock, L.A. Pavlish, University of Toronto Press, Toronto, 1–9, 1988.
6. S. Bowman (ed.), *Science and the Past*, British Museum Press, London, 1991.
7. C.W. Beck (ed.), *Archaeological Chemistry*, Advances in Chemistry Series 138, American Chemical Society, Washington, 1974.
8. G.F. Carter (ed.), *Archaeological Chemistry II*, Advances in Chemistry Series 171, American Chemical Society, Washington, 1978.
9. J.B. Lambert (ed.), *Archaeological Chemistry III*, Advances in Chemistry Series 205, American Chemical Society, Washington, 1984.
10. R.O. Allen (ed.), *Archaeological Chemistry IV*, Advances in Chemistry Series 220, American Chemical Society, Washington, 1989.
11. M.V. Orna (ed.), *Archaeological Chemistry V*, Advances in Chemistry Series, American Chemical Society, Washington, 1996.
12. Z. Goffer, *Archaeological Chemistry*, Wiley-Interscience, New York, 1980.
13. A.M. Pollard, C. Heron, *Archaeological Chemistry*, Royal Society of Chemistry, Cambridge, 1996.
14. J.B. Lambert, *Traces of the Past: Unravelling the Secrets of Archaeology Through Chemistry*, Addison-Wesley, Reading, MA, 1997.
15. M.H. Klaproth, 'Memoires de l'Academie Royale des Sciences et Belles-Lettres, Berlin, Classe de Philosophie Experimentale', 97–113, 1798.
16. S. Nappo, *Pompeii*, Weidenfeld and Nicolson, London, 1998.

17. H. Davy, 'Some Experimental Observations on the Colours Used in Painting by the Ancients', *Phil. Trans.*, **CV**, 97–124 (1815).
18. H.W. Diamond, 'Account of Wells or Pits, Containing Roman Remains, Discovered at Ewell in Surrey', *Archaeologia*, **32**, 451–455 (1847).
19. J. Wocel, 'Archäologische Parallelen', Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften (Wien), Philosophisch-Historische Classe 11, 716 (1853).
20. J. Wocel, 'Archäologische Parallelen', Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften (Wien), Philosophisch-Historische Classe 16, 169 (1855).
21. C.C.T.C. Göbel Fr, *Ueber den Einfluss der Chemie auf die Ermittelung der Völker der Vorzeit oder Resultate der Chemischen Untersuchung Metallischer Alterthümer Insbesondere der in den Ostseegouvernements Vorkommenden Behufs der Ermittelung der Völker von Welchen sie Abstammen*, Erlangen, 1842.
22. A. Damour, 'Sur la Composition des Haches en Pierre Trovees dans les Monuments Celtiques et chez les Tribus Sauvages', *Comptes Rendus*, **61**, 313–357 (1864).
23. A.H. Layard, *Discoveries in the Ruins of Nineveh and Babylon, with Travels in Armenia, Kurdistan, and the Desert*, J. Murray, London, 1853.
24. H. Schliemann, *Mycenae: A Narrative of Researches and Discoveries at Mycenae and Tiryns*, J. Murray, London, 1878.
25. J. Percy, *Metallurgy. Volume I: Fuel; Fire-clays; Copper, Zinc; Brass*, J. Murray, London, 1861.
26. J. Percy, *Metallurgy. Volume II: Iron; Steel*, J. Murray, London, 1864.
27. J. Percy, *Metallurgy. Volume III: Lead*, J. Murray, London, 1870.
28. J. Percy, *Metallurgy. Volume IV: Silver; Gold*, J. Murray, London, 1875.
29. O. Helm, 'Über die Herkunft des in den Alten Königs-Gräbern von Mykenae Gefundenen Bernsteins', *Schriften der Naturforschenden Gesellschaft in Danzig*, **6**(2), 234–239 (1885).
30. C.W. Beck, 'Spectroscopic Studies of Amber', *Appl. Spectrosc. Rev.*, **22**, 57–110 (1986).
31. F. Rathgen, *The Preservation of Antiquities: A Handbook for Curators*, Cambridge University Press, Cambridge, 1905.
32. A. Carnot, 'Sur la Composition des Ossements Fossiles et la Variation de leur Teneur en Fluor dans les Diffèrents Étages Géologiques', *Comptes Rendus*, **115**, 243–246 (1892).
33. K.P. Oakley, 'The Composition of the Piltdown Hominoid Remains', *Bull. Brit. Mus. Nat. Hist.*, **2**, 254–261 (1955).
34. A.M. Pollard, 'Archaeological Reconstruction Using Stable Isotopes', in *Stable Isotopes. Integration of Biological, Ecological and Geochemical Processes*, ed. H. Griffiths, Bios Scientific Publishers, Oxford, 285–301, 1998.
35. M.K. Sandford (ed.), *Investigations of Ancient Human Tissue. Chemical Analyses in Anthropology*, Gordon and Breach, Langhorne, PA, 1993.
36. N. Reeves, *The Complete Tutankhamun: The King, the Tomb, the Royal Treasure*, Thames & Hudson, London, 1990.
37. K. Spindler, *The Man in the Ice: The Preserved Body of a Neolithic Man Reveals the Secrets of the Stone Age*, Phoenix, London, 1995.
38. I.A. Efremov, 'Taphonomy: New Branch of Palaeontology', *Pan-Am. Geol.*, **74**, 81–94 (1940).
39. B. Little, P. Wagner, F. Mansfeld, 'Microbiologically Influenced Corrosion of Metals and Alloys', *Int. Mater. Rev.*, **36**, 253–272 (1991).
40. D.R. Brothwell, A.M. Pollard (eds.), *Handbook of Archaeological Sciences*, John Wiley and Sons, Chichester, in press, 2000.
41. A.M. Pollard, 'The Chemical Nature of the Burial Environment', in *Preserving Archaeological Remains In Situ*, eds. M. Corfield, P. Hinton, T. Nixon, A.M. Pollard, Museum of London Archaeology Service, London, 1998.
42. W.L. Lindsay, *Chemical Equilibria in Soils*, Wiley, New York, 1979.
43. E.A. Jenne (ed.), *Chemical Modeling in Aqueous Systems*, American Chemical Society, Washington, 1979.
44. N.H. Tennent (ed.), *Conservation Science in the U.K.*, James & James, London, 1993.
45. J.M. Cronyn, *The Elements of Archaeological Conservation*, Routledge, London, 1990.
46. E.T. Hall, 'X-ray Fluorescent Analysis Applied to Archaeology', *Archaeometry*, **3**, 29–35 (1960).
47. O. Williams-Thorpe, P.J. Potts, P.C. Webb, 'Field-portable Non-destructive Analysis of Lithic Archaeological Samples by X-ray Fluorescence Instrumentation Using a Mercury Iodide Detector: Comparison with Wavelength-dispersive XRF and a Case Study in British Stone Axe Provenancing', *J. Arch. Sci.*, **26**, 215–237 (1999).
48. A.M. Pollard, G.D. Bussell, D.C. Baird, 'The Analytical Investigation of Early Bronze Age Jet and Jet-like Material from the Devizes Museum', *Archaeometry*, **23**, 139–167 (1981).
49. G.D. Bussell, A.M. Pollard, D.C. Baird, 'The Characterization of Early Bronze Age Jet and Jet-like Material by X-ray Fluorescence', *Wilts. Arch. Mag.*, **76**, 27–32 (1981).
50. F.J. Hunter, J.G. McDonnell, A.M. Pollard, C.R. Morris, C.C. Rowlands, 'The Scientific Identification of Archaeological Jet-like Artefacts', *Archaeometry*, **35**, 69–89 (1993).
51. S. Watts, A.M. Pollard, G.A. Wolff, 'The Organic Geochemistry of Jet: Py-GCMS Applied to Identifying Jet and Similar Black Lithic Materials', *J. Arch. Sci.*, **26**, 923–933 (1999).
52. V.M. Goldschmidt, *Geochemistry*, Oxford University Press, London, 1954.

53. B.G. Trigger, *A History of Archaeological Thought*, Cambridge University Press, Cambridge, 1989.
54. S. Junghans, E. Sangmeister, M. Schroeder, *Kupfer und Bronze in der Fruehen Metallzeit Europas: Die Materialgruppen Beim Stand von 12 000 Analysen*, Gebr. Mann, Berlin, 1968.
55. S. Junghans, E. Sangmeister, M. Schroeder, *Metallanalysen Kupferzeitlicher und Fruehbronzezeitlicher Bodenfunde aus Europa*, Mann, Berlin, 1960.
56. J.F.S. Stone, L.C. Thomas, 'The Use and Distribution of Faience in the Ancient Near East and Prehistoric Europe', *Proc. Prehis. Soc.*, **22**, 37–84 (1956).
57. R.G. Newton, A.C. Renfrew, 'British Faience Beads Reconsidered', *Antiquity*, **44**, 199–201 (1970).
58. D.L. Clarke, *Analytical Archaeology*, Methuen, London, 1968.
59. C. Renfrew, *Problems in European Prehistory*, Edinburgh University Press, Edinburgh, 1979.
60. H. Neff (ed.), *Chemical Characterization of Ceramic Pastes in Archaeology*, Monographs in World Archaeology, Prehistory Press, Madison, WI, 1992.
61. J.W. Beal, I. Olmez, 'Provenance Studies of Pottery Fragments from Medieval Cairo, Egypt', *J. Radioanal. Nucl. Chem.*, **221**, 9–17 (1997).
62. A.W. Simon, J.H. Burton, 'Anthropological Interpretations from Archaeological Ceramic Studies: An Introduction', *J. Anthropol. Res.*, **54**, 435–446 (1998).
63. C. Punyadeera, A.E. Pillay, L. Jacobsen, G. Whitelaw, 'Application of XRF and Correspondence Analysis to Provenance Studies of Coastal and Inland Archaeological Pottery from the Mgeni River Area, South Africa', *X-ray Spectrom.*, **26**, 249–256 (1997).
64. G.J.P.A. Anders, C.J.A. Jorg, W.B. Stern, N. Andersbucher, 'On Some Physical Characteristics of Chinese and European Red Wares', *Archaeometry*, **34**, 43–52 (1992).
65. H. Hatcher, A. Kaczmarczyk, A. Scherer, R.P. Symonds, 'Chemical Classification of Some Roman Glazed Ceramics', *Am. J. Archaeol.*, **98**, 431–456 (1994).
66. R.E. Jones, *Greek and Cypriot Pottery: A Review of Scientific Studies*, British School at Athens, Athens, 1986.
67. G. Harbottle, 'Chemical Characterization in Archaeology', in *Contexts of Prehistoric Exchange*, eds. J.E. Ericson, T.K. Earle, Academic Press, New York, 13–51, 1982.
68. J.M. Coles, *Experimental Archaeology*, Academic Press, London, 1979.
69. G.R. Thomas, T.P. Young, 'The Determination of Bloomery Furnace Mass Balance and Efficiency', in *Geoarchaeology: Exploration, Environments, Resources*, ed. A.M. Pollard, Geological Society Special Publication, London, 155–164, 1999.
70. P. Budd, A.M. Pollard, B. Scaife, R.G. Thomas, 'The Possible Fractionation of Lead Isotopes in Ancient Metallurgical Processes', *Archaeometry*, **37**, 143–150 (1995).
71. N.H. Gale, A.P. Woodhead, Z.A. Stos-Gale, A. Walder, I. Bowen, 'Natural Variations Detected in the Isotopic Composition of Copper: Possible Applications to Archaeology and Geochemistry', *Int. J. Mass Spectrom.*, **184**, 1–9 (1999).
72. P.D. Budd, P. Lythgoe, R.A.R. McGill, A.M. Pollard, B. Scaife, 'Zinc Isotope Fractionation in Liquid Brass (Cu–Zn) Alloy: Potential Environmental and Archaeological Applications', in *Geoarchaeology: Exploration, Environments, Resources*, ed. A.M. Pollard, Geological Society Special Publication, London, 147–153, 1999.
73. R.E.M. Hedges, 'Analysis of the Drake Plate: Comparison with the Composition of Elizabethan Brass', *Archaeometry*, **21**, 21–26 (1979).
74. R.G.V. Hancock, L.A. Pavlish, R.M. Farquhar, D. Knight, 'The Analysis of Brass Samples from the Ball and Warminster Sites in Southern Ontario, Canada', in *Metals in Antiquity*, eds. S.M.M. Young, R.A. Ixer, P. Budd, A.M. Pollard, Archaeopress, Oxford, 341–347, 2000.
75. S.J. Fleming, *Authenticity in Art: The Scientific Detection of Forgery*, Institute of Physics, London, 1975.
76. D.M. Metcalf, F. Schweizer, 'The Metal Contents of the Silver Pennies of William II and Henry I (1087–1135)', *Archaeometry*, **13**, 177–190 (1971).
77. C. Chippindale, 'Editorial', *Antiquity*, **65**, 6–9 (1991).
78. F.D. Pate, 'Bone Chemistry and Paleodiet', *J. Arch. Meth. Theory*, **1**, 161–209 (1994).
79. R.G.V. Hancock, M.D. Grynpas, K.P.H. Pritzker, 'The Abuse of Bone Analyses for Archaeological Dietary Studies', *Archaeometry*, **31**, 169–179 (1989).
80. T.D. Price, 'Multi-element Studies of Diagenesis in Prehistoric Bone', in *The Chemistry of Prehistoric Human Bone*, ed. T.D. Price, Cambridge University Press, Cambridge, 126–154, 1989.
81. M.K. Sandford, 'Understanding the Biogenic–Diagenetic Continuum: Interpreting the Elemental Concentrations of Archaeological Bone', in *Investigations of Ancient Human Tissue*, ed. M.K. Sandford, Gordon and Breach, Langhorne, PA, 3–57, 1993.
82. S.C. Radosevich, 'The Six Deadly Sins of Trace Element Analysis: A Case of Wishful Thinking in Science', in *Investigations of Ancient Human Tissue*, ed. M.K. Sandford, Gordon and Breach, Langhorne, PA, 269–332, 1993.
83. J.H. Burton, T.D. Price, W.D. Middleton, 'Correlation of Bone Ba/Ca and Sr/Ca due to Biological Purification of Calcium', *J. Arch. Sci.*, **26**, 609–616 (1999).
84. B.K. Nelson, M.J. DeNiro, M.J. Schoeninger, D.J. De Paulo, P.E. Hare, 'Effects of Diagenesis on Strontium, Carbon, Nitrogen and Oxygen Concentrations and Isotopic Composition of Bone', *Geochim. Cosmochim. Acta*, **50**, 1941–1949 (1986).
85. S.H. Ambrose, 'Isotopic Analysis of Paleodiets: Methodological and Interpretative Considerations', in *Investigations of Ancient Human Tissue*, ed.

M.K. Sandford, Gordon and Breach, Langhorne, PA, 59–130, 1993.
86. N.J. van der Merwe, 'Light Stable Isotopes and the Reconstruction of Prehistoric Diets', in *New Developments in Archaeological Science*, ed. A.M. Pollard, Proceedings of the British Academy 77, Oxford University Press, Oxford, 247–264, 1992.
87. H.P. Schwarcz, M.J. Schoeninger, 'Stable Isotope Analysis in Human Nutritional Ecology', *Yearbook Phys. Anthropol.*, **34**, 283–321 (1991).
88. M.J. DeNiro, 'Stable Isotopy and Archaeology', *Am. Sci.*, **75**, 182–191 (1987).
89. A.W. Stott, R.P. Evershed, S. Jim, V. Jones, J.M. Rogers, N. Tuross, S. Ambrose, 'Cholesterol as a New Source of Palaeodietary Information: Experimental Approaches and Archaeological Applications', *J. Arch. Sci.*, **26**, 705–716 (1999).
90. S.H. Ambrose, L. Norr, 'Experimental Evidence for the Relationship of the Carbon Isotope Ratios of Whole Diet and Dietary Protein to those of Bone Collagen and Carbonate', in *Prehistoric Human Bone: Archaeology at the Molecular Level*, eds. J.B. Lambert, G. Grupe, Springer Verlag, Berlin, 1–37, 1993.
91. J. Sealy, R. Armstrong, C. Schrire, 'Beyond Lifetime Averages: Tracing Life Histories Through Isotopic Analysis of Different Calcified Tissues from Human Archaeological Skeletons', *Antiquity*, **69**, 290–300 (1995).
92. T.D. Price, G. Grupe, P. Schröter, 'Reconstruction of Migration Patterns in the Bell Beaker Period by Stable Strontium Isotope Analysis', *Appl. Geochem.*, **9**, 413–417 (1994).
93. G. Grupe, T.D. Price, P. Schröter, F. Sölner, C.M. Johnson, B.L. Beard, 'Mobility of Bell Beaker People Revealed by Strontium Isotope Ratios of Tooth and Bone: A Study of Southern Bavarian Skeletal Remains', *Appl. Geochem.*, **12**, 517–525 (1997).
94. A. Sillen, 'Diagenesis of the Inorganic Phase of Cortical Bone', in *The Chemistry of Prehistoric Human Bone*, ed. T.D. Price, Cambridge University Press, Cambridge, 211–219, 1989.
95. C.T. Williams, 'Alteration of Chemical Composition of Fossil Bones by Soil Processes and Groundwater', in *Trace Elements in Environmental History*, eds. G. Grupe, B. Herrmann, Springer-Verlag, Berlin, 27–40, 1988.
96. R.E.M. Hedges, A.R. Millard, A.W.G. Pike, 'Measurements and Relationships of Diagenetic Alteration of Bone from 3 Archaeological Sites', *J. Arch. Sci.*, **22**, 201–209 (1995).
97. R.E.M. Hedges, A.R. Millard, 'Bones and Groundwater–Towards the Modeling of Diagenetic Processes', *J. Arch. Sci.*, **22**, 155–164 (1995).
98. A. Sillen, R. LeGeros, 'Solubility Profiles of Synthetic Apatite and of Modern and Fossil Bone', *J. Arch. Sci.*, **18**, 385–397 (1991).
99. P.L. Koch, N. Tuross, M.L. Fogel, 'The Effects of Sample Treatment and Diagenesis on the Isotopic Integrity of Carbonate in Biogenic Hydroxylapatite', *J. Arch. Sci.*, **24**, 417–429 (1997).
100. M.H. Engel, S.A. Macko, *Organic Geochemistry: Principles and Applications*, Plenum Press, New York, 1993.
101. M. Krings, A. Stone, R.W. Schmitz, H. Krainitzki, M. Stoneking, S. Pääbo, 'Neandertal DNA Sequences and the Origin of Modern Humans', *Cell*, **90**, 19–30 (1997).
102. K.D. Thomas, 'Molecular Biology and Archaeology: A Prospectus for Interdisciplinary Research', *World Archaeol.*, **25**, 1–17 (1993).
103. M.B. Richards, B.C. Sykes, R.E.M. Hedges, 'Authenticating DNA Extracted from Ancient Skeletal Remains', *J. Arch. Sci.*, **22**, 291–300 (1995).
104. M. Hoss, P. Jaruga, T.H. Zastawny, M. Dizdaroglu, S. Pääbo, 'DNA Damage and DNA Sequence Retrieval from Ancient Tissues', *Nucleic Acids Res.*, **24**, 1304–1307 (1996).
105. A.M. Child, A.M. Pollard, 'A Review of the Applications of Immunochemistry to Archaeological Bone', *J. Arch. Sci.*, **19**, 39–47 (1992).
106. J.A. Eisele, D.D. Fowler, G. Haynes, R.A. Lewis, 'Survival and Detection of Blood Proteins on Stone Tools', *Antiquity*, **69**, 36–46 (1995).
107. T.H. Loy, 'Prehistoric Blood Residues: Detection on Stone Tool Surfaces and Identification of Species of Interest', *Science*, **220**, 1269–1271 (1983).
108. T.H. Loy, E.J. Dixon, 'Blood Residues on Fluted Points from Eastern Beringia', *Am. Antiq.*, **63**, 21–46 (1998).
109. N. Tuross, T. Dillehay, 'The Mechanism of Organic Preservation at Monte Verde, Chile, and One Use of Biomolecules in Archaeological Interpretation', *J. Field Archaeol.*, **22**, 97–110 (1995).
110. C. Heron, R.P. Evershed, 'The Analysis of Organic Residues and the Study of Pottery Function', in *Archaeological Method and Theory V*, ed. M.B. Schiffer, University of Arizona Press, Arizona, 247–286, 1993.
111. R.P. Evershed, C. Heron, L.J. Goad, 'Epicuticular Wax Components Preserved in Pot Sherds as Chemical Indicators of Leafy Vegetables in Ancient Diets', *Antiquity*, **65**, 540–544 (1991).
112. R.P. Evershed, H.R. Mottram, S.N. Dudd, S. Charters, A.W. Stott, A.M. Gibson, A. Conner, P.W. Blinkhorn, V. Reeves, 'New Criteria for the Identification of Animal Fats Preserved in Archaeological Pottery', *Naturwissenschaften*, **84**, 402–406 (1997).
113. B.X. Mayer, C. Reiter, T.L. Bereuter, 'Investigation of the Triacylglycerol Composition of Iceman's Mummified Tissue by High-temperature Gas Chromatography', *J. Chromatogr.*, **B692**, 1–6 (1997).
114. A.M. Gernaey, D.E. Minnikin, M.S. Copley, J.J. Power, A.M.S. Ahmed, R.A. Dixon, C.A. Roberts, D.J. Robertson, J. Nolan, A. Chamberlain, 'Detecting Ancient

Tuberculosis', *Internet Archaeology*, **5** (1998) (http://intarch.ac.uk/journal/issue5/gernaey_index.html).
115. I. Bull, I.A. Simpson, P.F. van Bergen, R.P. Evershed, 'Muck 'n' Molecules: Organic Geochemical Methods for Detecting Ancient Manuring', *Antiquity*, **73**, 86–96 (1999).
116. R.P. Evershed, A.W. Stott, '$\delta^{13}C$ Analysis of Cholesterol Preserved in Archaeological Bones and Teeth', *Anal. Chem.*, **68**, 4402–4408 (1996).
117. E.M. Aveling, C. Heron, 'The Chemistry of Birch Bark Tars at Mesolithic Star Carr', *Ancient Biomol.*, **2**, 69–80 (1998).
118. C. Heron, N. Nemcek, K.M. Bonfield, J. Dixon, B.S. Ottaway, 'The Chemistry of Neolithic Beeswax', *Naturwissenschaften*, **81**, 266–299 (1994).
119. J.S. Mills, R. White, 'The Identity of the Resins from the Late Bronze Age Shipwreck at Ulu Burun (Kas)', *Archaeometry*, **31**, 37–44 (1989).
120. R.P. Evershed, P.F. van Bergen, T.M. Peakman, E.C. Leigh-Firbank, M.C. Horton, D. Edwards, M. Biddle, B. Kjølbye-Biddle, P.A. Rowley-Conwy, 'Archaeological Frankincense', *Nature*, **390**, 667–668 (1997).
121. C.H. Hocart, B. Fankhauser, D.W. Buckle, 'Chemical Archaeology of Kava, a Potent Brew', *Rapid Commun. Mass Spectrom.*, **7**, 219–224 (1993).
122. J. Connan, 'Use and Trade of Bitumen in Antiquity and Prehistory: Molecular Archaeology Reveals the Secrets of Past Civilizations', *Philos. Trans. R. Soc. London*, **B354**, 33–50 (1999).
123. J. Connan, A. Nissenbaum, D. Dessort, 'Molecular Archaeology: Export of Dead Sea Asphalt to Canaan and Egypt in the Chalcolithic–Early Bronze Age (4th–3rd Millennium BC)', *Geochim. Cosmochim. Acta*, **56**, 2743–2759 (1992).
124. P.E. McGovern, D.L. Glusker, L.J. Exner, M.M. Voigt, 'Neolithic Resinated Wine', *Nature*, **381**, 480–481 (1996).
125. H.G.M. Edwards, D. Farwell, C. Heron, H. Croft, A.R. David, ''Cats' Eyes in a New Light: Fourier Transform Raman Spectroscopic and Gas Chromatographic–Mass Spectrometric Study of Egyptian Mummies', *J. Raman Spectrosc.*, **30**, 139–146 (1999).
126. J.B. Lambert, C.W. Beck, J.S. Frye, 'Analysis of European Amber by Carbon-13 Nuclear Magnetic Resonance Spectroscopy', *Archaeometry*, **30**, 248–263 (1988).
127. C.W. Beck, E. Wilbur, S. Meret, 'Infrared Spectra and the Origins of Amber', *Nature*, **201**, 256–257 (1964).
128. J.R. Hunter, C.A. Roberts, A. Martin, *Studies in Crime: An Introduction to Forensic Archaeology*, Batsford, London, 1995.
129. H. Otto, W. Witter, *Handbuch der Altesten Vorgeschichtlichen Metallurgie in Mitteleuropa*, Leipzig, 1952.
130. M.J. Hughes, M.R. Cowell, P.T. Craddock, 'Atomic Absorption Techniques in Archaeology', *Archaeometry*, **18**, 19–37 (1976).
131. W. Slavin, 'A Comparison of Atomic Spectroscopic Analytical Techniques', *Spectrosc. Int.*, **4**, 22–27 (1992).
132. H. Hatcher, M.S. Tite, J.N. Walsh, 'A Comparison of Inductively-coupled Plasma Emission Spectrometry and Atomic Absorption Spectrometry Analysis on Standard Reference Silicate Materials and Ceramics', *Archaeometry*, **37**, 83–94 (1995).
133. M. Banks, E.T. Hall, 'X-ray Fluorescent Analysis in Archaeology: The 'Milli-probe'', *Archaeometry*, **6**, 31–36 (1963).
134. E.T. Hall, F. Schweizer, P.A. Toller, 'X-ray Fluorescence Analysis of Museum Objects: A New Instrument', *Archaeometry*, **15**, 53–78 (1973).
135. M.F. Guerra, 'Analysis of Archaeological Metals. The Place of XRF and PIXE in the Determination of Technology and Provenance', *X-ray Spectrom.*, **27**, 73–80 (1998).
136. M.F. Araujo, L.C. Alves, J.M.P. Cabral, 'Comparison of EDXRF and PIXE in the Analysis of Ancient Gold Coins', *Nucl. Instrum. Methods Phys. Res.*, **B75**, 450–453 (1993).
137. C. Punyadeera, A.E. Pillay, L. Jacobson, G. Whitelaw, 'Application of XRF and Correspondence Analysis to Provenance Studies of Coastal and Inland Archaeological Pottery from the Mngeni River Area, South Africa', *X-ray Spectrom.*, **26**, 249–256 (1997).
138. C.T. Yap, Y.N. Hua, 'Raw-materials for Making Jingdezhen Porcelain from the 5 Dynasties to the Qing Dynasty', *Appl. Spectrosc.*, **46**, 1488–1494 (1992).
139. G.A. Cox, O.S. Heavens, R.G. Newton, A.M. Pollard, 'A Study of the Weathering Behaviour of Mediaeval Glass from York Minster', *J. Glass Stud.*, **21**, 54–75 (1979).
140. D.R. Mitchell, M.S. Shackley, 'Classic Period Hohokam Obsidian Studies in Southern Arizona', *J. Field Archaeol.*, **22**, 291–304 (1995).
141. F.J. Hunter, J.G. McDonnell, A.M. Pollard, C.R. Morris, C.C. Rowlands, 'The Scientific Identification of Archaeological Jet-like Artifacts', *Archaeometry*, **35**, 69–89 (1993).
142. P.J. Potts, 'Geoanalysis: Past, Present and Future', *Analyst*, **122**, 1179–1186 (1997).
143. M. Cowell, S. La Niece, 'Metalwork: Artifice and Artistry', in *Science and the Past*, ed. S. Bowman, British Museum Publications, London, 74–98, 1991.
144. E.T. Hall, 'Surface Enrichment of Buried Metals', *Archaeometry*, **4**, 62–66 (1961).
145. G.A. Cox, A.M. Pollard, 'X-ray Fluorescence of Ancient Glass: The Importance of Sample Preparation', *Archaeometry*, **19**, 45–54 (1977).
146. A.P. Hornblower, 'Archaeological Applications of the Electron Probe Microanalyser', *Archaeometry*, **5**, 108–112 (1962).
147. A.C. Dunham, F.C.F. Wilkinson, 'Accuracy, Precision and Detection Limits of Energy-dispersive Electron-microprobe Analyses of Silicates', *X-ray Spectrom.*, **7**, 50–56 (1978).

148. M. Verità, R. Basso, M.T. Wypyski, R.J. Koestler, 'X-ray Microanalysis of Ancient Glassy Materials: A Comparative Study of Wavelength Dispersive and Energy Dispersive Techniques', *Archaeometry*, **36**, 241–251 (1994).
149. D.J. Barber, I.C. Freestone, 'An Investigation of the Origin of the Colour of the Lycurgus Cup by Analytical Transmission Electron Microscopy', *Archaeometry*, **32**, 33–45 (1990).
150. S.J. Fleming, C.P. Swann, 'PIXE Spectrometry as an Archaeometric Tool', *Nucl. Instrum. Methods Phys. Res.*, **A242**, 626–631 (1986).
151. S.A.E. Johansson, J.L. Campbell, *PIXE: A Novel Technique for Elemental Analysis*, John Wiley, Chichester, 1988.
152. A. Person, H. Bocherens, J.F. Saliège, F. Paris, V. Zeitoun, M. Gérard, 'Early Diagenetic Evolution of Bone Phosphate: An X-ray Diffractometry Analysis', *J. Arch. Sci.*, **22**, 211–221 (1995).
153. K. Tronner, A.G. Nord, G.C. Borg, 'Corrosion of Archaeological Bronze Artefacts in Acidic Soil', *Water, Air, Soil Pollut.*, **85**, 2725–2730 (1995).
154. L. Robbiola, J.M. Blengino, C. Fiaud, 'Morphology and Mechanism of Formation of Natural Patinas on Archaeological Cu–Sn Alloys', *Corros. Sci.*, **40**, 2083–2111 (1998).
155. J.L. Ruvalcaba-Sil, M.A.O. Salamanca, L. Manzanilla, J. Miranda, J.C. Ortega, C. Lopez, 'Characterization of Pre-Hispanic Pottery from Teotihuacan, Mexico, by a Combined PIXE-RBS and XRD Analysis', *Nucl. Instrum. Methods Phys. Res.*, **B150**, 591–596 (1999).
156. L.L. Jackson, P.A. Baedecker, T.L. Fries, P.J. Lamothe, 'Geological and Inorganic Materials', *Anal. Chem.*, **65**, 13R–28R (1993).
157. C.M. Kraay, 'Gold and Copper Traces in Early Greek Silver', *Archaeometry*, **1**, 1–5 (1958).
158. E.V. Sayre, R.W. Dodson, 'Neutron Activation Study of Mediterranean Potsherds', *Am. J. Archaeol.*, **61**, 35–41 (1957).
159. M.J. Hughes, M.R. Cowell, D.R. Hook, 'Neutron Activation and Plasma Emission Spectrometric Analysis in Archaeology', British Museum Occasional Paper 82, London, 1991.
160. M.D. Glascock, 'Nuclear Reaction Chemical Analysis: Prompt and Delayed Measurements', in *Chemical Analysis by Nuclear Methods*, ed. Z.B. Alfassi, John Wiley, Chichester, 75–99, 1994.
161. N.H. Gale, 'Lead Isotope Studies in the Aegean', in *New Directions in Archaeological Science*, ed. A.M. Pollard, Proceedings of the British Academy 77, Oxford University Press, Oxford, 63–108, 1992.
162. H.E. Duckworth, R.C. Barber, V.S. Venkatsubramanian, *Mass Spectrometry*, 2nd edition, Cambridge University Press, Cambridge, 1986.
163. S.J. Prosser, S.T. Brookes, A. Linton, T. Preston, 'Rapid, Automated-analysis of C-13 and O-18 of CO_2 in Gas Samples by Continuous-flow, Isotope Ratio Mass-spectrometry', *Biol. Mass Spectrom.*, **20**, 724–730 (1991).
164. I. Robertson, V.R. Switsur, A.H.C. Carter, A.C. Barker, J.S. Waterhouse, K.R. Briffa, P.D. Jones, 'Signal Strength and Climate Relationships in C-13/C-12 Ratios of Tree Ring Cellulose from Oak in East England', *J. Geophys. Res.–Atmospheres*, **102**, 19 507–19 516 (1997).
165. M. Hamester, H. Stechmann, M. Steiger, W. Dannecker, 'The Origin of Lead in Urban Aerosols–A Lead Isotopic Study', *Sci. Total Environ.*, **146/7**, 321–323 (1994).
166. N. Furuta, 'Interlaboratory Comparison Study on Lead Isotope Ratios Determined by Inductively Coupled Plasma Mass Spectrometry', *Anal. Sci.*, **7**, 823–826 (1991).
167. A.J. Walder, P.A. Freedman, 'Isotopic Ratio Measurement Using a Double Focusing Magnetic-sector Mass Analyzer with an Inductively Coupled Plasma as an Ion-source', *J. Anal. At. Spectrom.*, **7**, 571–575 (1992).
168. A.J. Walder, N. Furuta, 'High-precision Lead-isotope Ratio Measurement by Inductively-coupled Plasma Multiple Collector Mass-spectrometry', *Anal. Sci.*, **9**, 675–680 (1993).
169. A.N. Halliday, D.C. Lee, J.N. Christensen, A.J. Walder, P.A. Freedman, C.E. Jones, C.M. Hall, W. Yi, D. Teagle, 'Recent Developments in Inductively-coupled Plasma Magnetic Sector Multiple Collector Mass-spectrometry', *Int. J. Mass Spectrom. Ion Processes*, **146**, 21–33 (1995).
170. T.F.M. Oudemans, J.J. Boon, 'Molecular Archaeology: Analysis of Charred Food Remains from Prehistoric Pottery by Pyrolysis – Gas Chromatography/Mass Spectrometry', *J. Anal. Appl. Pyrol.*, **20**, 197–227 (1991).
171. P.E. Hare, M.L. Fogel, T.W. Stafford, Jr, A.D. Mitchell, T.C. Hoering, 'The Isotopic Composition of Carbon and Nitrogen in Individual Amino Acids Isolated from Modern and Fossil Proteins', *J. Arch. Sci.*, **18**, 277–292 (1991).
172. W. Sandermann, 'Untersuchungen Vorgeschichthlicher 'Graberhärze' und Kitte', *Technische Beiträge zur Archäeologie*, **2**, 58–73 (1965).
173. H.G.M. Edwards, 'Raman Spectroscopy in Art, Medicine and Archaeology', *J.Raman Spectrosc.*, **28**, 77 (1997).
174. H.G.M. Edwards, C.J. Brook, J.K.F. Tait, 'Fourier Transform Raman Spectroscopic Study of Pigments from English Mediaeval Wall Paintings', *J. Raman Spectrosc.*, **28**, 95 (1997).
175. D.C. Smith, A. Barbet, 'A Preliminary Raman Microscopic Exploration of Pigments in Wall Paintings in the Roman Tomb Discovered at Kertch, Ukraine, in 1891', *J. Raman Spectrosc.*, **30**, 319–324 (1999).
176. M.T. Kirchner, H.G.M. Edwards, D. Lucy, A.M. Pollard, 'Ancient and Modern Specimens of Human Teeth: A Fourier Transform Raman Spectroscopic Study', *J. Raman Spectrosc.*, **28**, 171–178 (1997).
177. A.S. Wilson, H.G.M. Edwards, D.W. Farwell, R.C. Janaway, 'Fourier Transform Raman Spectroscopy: Evaluation as a Non-destructive Technique for Studying

the Degradation of Human Hair from Archaeological and Forensic Environments', *J. Raman Spectrosc.*, **30**, 367–373 (1999).
178. M. Gniadecka, H.G.M. Edwards, J.P.H. Hansen, O.F. Nielsen, D.H. Christensen, S.E. Guillen, H.C. Wulf, 'Near-infrared Fourier Transform Raman Spectroscopy of the Mummified Skin of the Alpine Iceman, Qilakitsoq Greenland Mummies and Chiribaya Mummies from Peru', *J. Raman Spectrosc.*, **30**, 147–153 (1999).
179. H.G.M. Edwards, M.J. Falk, 'Fourier Transform Raman Spectroscopic Study of Ancient Resins: A Feasibility Study of Application to Archaeological Artefacts', *J. Raman Spectrosc.*, **28**, 211–218 (1997).
180. H.G.M. Edwards, M.J. Falk, 'Investigation of the Degradation Products of Archaeological Linens by Raman Spectroscopy', *Appl. Spectrosc.*, **51**, 1134–1138 (1997).
181. D.C. Smith, F. Gendron, 'Archaeometric Application of the Raman Microprobe to the Non-destructive Identification of Two Pre-columbian Ceremonial Polished 'Greenstone' Axe-heads from Mesoamerica', *J. Raman Spectrosc.*, **28**, 731–738 (1997).
182. S. Weiner, S. Schiegl, P. Goldberg, O. Bar-Yosef, 'Mineral Assemblages in Kebara and Hayonim Caves, Israel–Excavation Strategies, Bone Preservation, and Wood Ash Remnants', *Isr. J. Chem.*, **35**, 143–154 (1995).
183. A. Sillen, A. Morris, 'Diagenesis of Bone from Border Cave: Implications for the Age of the Border Cave Hominids', *J. Human Evol.*, **31**, 499–506 (1996).
184. E. Lopez-Ballester, M.T. Domenech-Carbo, J.V. Gimeno-Adelantado, F. Bosch-Reig, 'Study by FT-IR Spectroscopy of Ageing of Adhesives Used in Restoration of Archaeological Glass Objects', *J. Mol. Struct.*, **483**, 525–531 (1999).
185. H.P. Schwarcz, R. Grün, 'Electron-spin-resonance (esr) Dating of the Origins of Modern Man', *Philos. Trans. R. Soc. London*, **B337**, 145–148 (1992).
186. D. Cordischi, D. Monna, L. Segre, 'Electron-spin-resonance Analysis of Marble Samples from Mediterranean Quarries of Archaeological Interest', *Archaeometry*, **25**, 68–76 (1983).
187. T. Warashina, 'Allocation of Jasper Archaeological Implements by Means of ESR and XRF', *J. Arch. Sci.*, **19**, 357–373 (1992).
188. J. Bartoll, A. Tani, 'Thermal History of Archaeological Objects, Studied by Electron Spin Resonance', *Naturwissenschaften*, **85**, 474–481 (1998).
189. Y. Bensimon, B. Deroide, S. Clavel, J.V. Zanchetta, 'Electron Spin Resonance and Dilatometric Studies of Ancient Ceramics Applied to the Determination of Firing Temperature', *Jpn. J. Appl. Phys.*, **37**, 4367–4372 (1998).
190. V. Michel, C. Falgueres, J.M. Dolo, 'ESR Signal Behavior Study at *G* Similar to 2.002 of Modern and Fossil Bones for Heating Palaeotemperature Assessment', *Radiat. Meas.*, **29**, 95–103 (1998).
191. J. Bartoll, A. Tani, M. Ikeya, T. Inada, 'ESR Investigations of Burnt Soil', *Appl. Magn. Reson.*, **11**, 577–586 (1996).
192. G.C. Hillman, G.V. Robins, D. Oduwole, K.D. Sales, D.A.C. McNeil, 'Determination of Thermal Histories of Archaeological Cereal-grains with Electron-spin Resonance Spectroscopy', *Science*, **222**, 1235–1236 (1983).
193. E.L. Ghisalberti, I.M. Godfrey, 'Application of Nuclear Magnetic Resonance Spectroscopy to the Analysis of Organic Archaeological Materials', *Stud. Conservation*, **43**, 215–230 (1998).
194. A.P. Lee, J. Klinowski, E.A. Marseglia, 'Application of Nuclear-magnetic-resonance Spectroscopy to Bone Diagenesis', *J. Arch. Sci.*, **22**, 257–262 (1995).
195. S.P. DeAtley, R.L. Bishop, 'Toward an Integrated Interface for Archaeology and Archaeometry', in *The Ceramic Legacy of Anna O. Shepard*, eds. R.L. Bishop, F.W. Lange, University Press of Colorado, Boulder, 358–380, 1991.
196. N. Yoffee, A. Sherratt (eds.), *Archaeological Theory: Who Sets the Agenda?*, Cambridge University Press, Cambridge, 1993.

Gravimetry

M. Widmer*
Ciba-Geigy, Basel, Switzerland

1 Introduction 1
2 Gravimetric Techniques 1
3 Evaluation of Gravimetric Analyses 2
4 Electrogravimetry 3
4.1 Principle of Electrolysis with Controlled Current 3
4.2 Electrogravimetric Methods 5
4.3 Electroseparation 5
Acknowledgments 5
Related Articles 6
Further Reading 6

Gravimetry is one of the oldest techniques of quantitative analysis. In gravimetry, the analyte is precipitated as a stoichiometrically defined compound. After collecting and drying, the precipitate is weighed on an analytical balance and the analyte is determined from the mass and the known stoichiometry of the analytes in the precipitated compound. Although many instrumental methods have superseded gravimetry it is a very important method for standardization processes. Since calibration is inherent in the relative molecular masses and the stoichiometry of the chemical reaction, it represents one of the few primary methods.

As a closely related method, electrogravimetry is also considered. In electrogravimetry the analyte is deposited on a metal, for example a platinum electrode, by electrolysis with a controlled current and its mass is evaluated from the weight difference of the electrode before and after electrolysis. Electroseparation of two analytes is feasible by judicious choice of the electrode potential.

1 INTRODUCTION

In gravimetry, an analyte (in most cases ionic) is precipitated in the form of an insoluble compound of defined stoichiometry. After collection and drying, the product is weighed on an analytical balance and, from the mass and known stoichiometry, the original analyte is quantitatively determined.

Gravimetric methods are widely used in standardization processes, although volumetric and instrumental techniques have superseded gravimetric methods in most routine and general analytical investigations. Generally speaking, gravimetric methods are extremely accurate, owing to the fact that it is possible to weigh substances to great accuracy with analytical balances. It is common practice to determine a weight to five digits; however, even the most sensitive volumetric and instrumental techniques are rarely more precise than about 1–0.1%.

The accuracy (trueness and precision) of a gravimetric method depends on the precipitation technique and also on the properties of the precipitate. If high accuracy is demanded, the following requirements must be met:

- definite and reproducible stoichiometric composition of the precipitate;
- low solubility in the mother liquor and in the wash solvent (the analyzed substance must be quantitatively incorporated into the final precipitate);
- minimal interference from other elements and components of the system;
- low surface area of the precipitate (crystalline) so that adsorption of impurities is minimal;
- properties that enable convenient separation of the solid from the mother liquid and efficient washing with a suitable solvent;
- thermal stability, so that the precipitate may be dried conveniently without changing its composition;
- stability of the dried product (hygroscopic properties of the substance are troublesome).

2 GRAVIMETRIC TECHNIQUES

It is of the utmost importance that the precipitation leads to a product of highest purity that can be filtered. Therefore some skill and experience is required when gravimetric methods are to be applied.

Some of the above-mentioned properties can be influenced by the precipitation technique. Crystalline substances generally have a definite stoichiometric composition, they are easy to separate from the original solution, either by filtration or centrifugation, and they may be efficiently washed with a minimum volume of the rinsing solvent. In addition, crystalline substances are characterized by a relatively small specific surface area. Therefore, the surface activity is relatively small (particularly when compared with amorphous or colloidal precipitates) and the amount of adsorbed or occluded material is minimal.

Crystal habit and size can be controlled to a certain extent by the precipitation technique. Rapid mixing of two concentrated reagent solutions results in the formation of

* Since this article was written the author has died.

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

undesirable microcrystalline products. If the solubility of a substance is extremely temperature dependent, small crystals are obtained by sudden temperature changes, resulting in a lower solubility. If a hot solution is cooled very slowly, the crystal seeds have an opportunity to grow before new crystallization centers are formed. Supersaturated solutions are subject to sudden crystallization processes that produce a heavily occluded material; mother liquor with all its constituents is occluded in the precipitate. These impurities cannot be removed by washing.

As a general rule, hot solutions should be used to form precipitates with the appropriate features necessary for the filtration. Add the reagents dropwise and choose as small a concentration as possible. Stir during the addition of any reagents so that you achieve a homogeneous distribution of all solutes.

A number of volatile solutes and electrolytes are known. Solutions of such chemicals should not be heated, otherwise substantial losses may occur. For this reason, the heating of a solution containing a volatile acid (H_2S, HCl, etc.) or base (NH_3, etc.) is not recommended. For instance, in the gravimetric determination of hydrochloric acid with silver nitrate, the acid solution should not be heated before the addition of $AgNO_3$. However, after addition of $AgNO_3$, a temperature close to the boiling point (80 °C) is recommended in order to obtain a precipitate which can be conveniently filtered. Therefore, chloride solutions used in gravimetric determinations should not be heated before the addition of $AgNO_3$, since the precipitation is carried out in an acidic medium (addition of HNO_3) and therefore contains volatile HCl. Before filtration, the precipitation should have settled to the bottom of the vessel, so that the most of the liquid can be removed by decantation.

In the early days the precipitation was collected on a filter paper, which was then burned completely in a platinum crucible to leave the final product only. Later, special ceramic crucibles were made available as filtering devices. In both applications the washing process is essential for the accuracy of final results. To obtain a dry product, the crucible containing the precipitation is dried in an oven. From the weight difference between the empty crucible and crucible plus precipitate, the weight of the precipitate is determined, and from that the weight of the analyte is calculated.

3 EVALUATION OF GRAVIMETRIC ANALYSES

The aim of a gravimetric analysis is the determination of the concentration w_A of an analyte in a given sample using the laws of stoichiometry. The result is expressed in Equation (1) and is based on two precise weighings:

$$w_A = \frac{m_A}{e}100 = \frac{af}{e}100 \quad (1)$$

where w_A is the content of the component sought (%), m_A is the mass of the component sought (mg), e is the mass of the sample before gravimetric treatment (mg), a is the mass of the substance weighed after precipitation, drying, etc. (mg), and f is the stoichiometric factor.

The stoichiometric factor f is the ratio of the molar masses of component A and its weighing form as shown in Equation (2):

$$f = \frac{\text{formula weight of the substance sought (g mol}^{-1})}{\text{formula weight of the substance weighed (g mol}^{-1})} \quad (2)$$

$f \times 100$ is the percentage of the substance sought in the substance weighed.

Table 1 gives some examples of the calculation of stoichiometric factors.

The stoichiometric factor f should be as small as possible, in order to provide for a high sensitivity of

Table 1 Example for calculation of stoichiometric factors

Species sought	Species weighed	Stoichiometric factor	
MgO	$Mg_2P_2O_7$	$\frac{2M(MgO)}{M(Mg_2P_2O_7)}$	0.3622
Fe_3O_4	Fe_2O_3	$\frac{2M(Fe_3O_4)}{3M(Fe_3O_4)}$	0.9666

Table 2 Examples of stoichimetric factors for precipitations with different precipitation reagents

Species sought	Precipitation reagent	Species precipitated	Species weighed	Stoichiometric factor
Fe^{3+}	OH^-	$Fe(OH)_3$	Fe_2O_3	0.6994
Al^{3+}	8-Hydroxyquinoline	$Al(HQ)_3$	Al_2O_3	0.5293
Al^{3+}	8-Hydroxyquinoline	$Al(HQ)_3$	$Al(HQ)_3$	0.05873
Ni^{2+}	Dimethylglyoxime (sodium salt)	$Ni(DMG)_2$	$Ni(DMG)_2$	0.2032

the method, and to keep the weighing error small. This can be obtained by using organic precipitation reagents with high molar mass (Table 2).

A practical example is given for the gravimetric determination of Ba^{2+} after precipitation with H_2SO_4 as $BaSO_4$, filtration, and thorough drying in a quartz crucible:

Sample weight: 0.6537 g; Weight of $BaSO_4$: 0.4288 g

$$f = \frac{M(\text{Ba})}{M(\text{BaSO}_4)} = \frac{137.34}{233.40} = 0.5884$$

$$w_{\text{Ba}} = \frac{af}{e} \times 100 = 38.6\%$$

4 ELECTROGRAVIMETRY

In contrast to voltammetric experiments (see section 3), electrolysis can be performed under conditions where the composition of the solution is altered, owing to the quantitative electrooxidation or electroreduction of solution components. These bulk electrolysis techniques are characterized by a large ratio of working electrode area A to solution volume V and convective diffusion-controlled mass transport conditions. The use of large electrodes permits the flow of relatively high current and makes possible the fast conversion of material into product in electrooxidation or electroreduction. The basic principles governing electrode reactions in bulk electrolysis are the same as discussed in connection with microelectrolysis (voltammetric) techniques.

Bulk electrolysis techniques can be classified according to the quantities measured. In electrogravimetric experiments, the weight of the material deposited at one of the electrodes (the working electrode) in a pure form of known stoichiometry is determined. Electrogravimetry belongs to the group of microdetermination gravimetry using the current as reagent.

Electrogravimetry is among the oldest electroanalytical techniques and has become a useful analytical tool in the analysis of metal ions. It is now used primarily for the electroseparation of metals from complex mixtures.

In coulometric methods, however, the total amount of electricity Q consumed in the electrolysis serves as the basis for the quantitative determination of the amount of analyte present or reagent produced (see section 3).

Both electrogravimetric and coulometric methods are highly accurate (true and precise) analytical methods. The evaluation of the analyte concentration requires no preliminary calibration with standard solutions.

4.1 Principle of Electrolysis with Controlled Current

In the electrolytic process a chemical reaction is driven in the nonspontaneous direction by an external source of electricity, by applying a DC voltage to the cell consisting of two electrodes in contact with the solution. The total voltage applied across the cell is distributed as shown in Equation (3) if $E_{\text{appl}} < E_{\text{equ}}$:

$$E_{\text{appl}} = E_{\text{a}} - E_{\text{c}} + iR \tag{3}$$

where E_{a} is the anode potential, E_{c} is the cathode potential, i is the current, and R is the resistance.

Each electrode potential is considered as the sum of a reversible potential and the overpotential. The reversible potential depends on the standard potential of the electrochemical half-reaction taking place at the electrode and the activities of the relevant substances in the vicinity of the electrode surface.

Let us consider the overall reversible reaction shown in Equation (4):

$$\text{O} + ne^- \rightleftharpoons \text{R} \tag{4}$$

where both O and R are soluble.

The rate of the electrode reaction depends on the applied potential, and the concentrations of the substances O and R are expressed by the Nernst equation as shown in Equation (5):

$$E_{\text{appl}} = E^0 + \frac{\text{R}T}{nF} \ln \frac{c_{\text{O}}}{c_{\text{R}}} \tag{5}$$

As the electrolysis proceeds, that is, as the metal deposition proceeds, the electrode potential decreases with decrease of the metal cation concentration c_{O} in the solution. The degree of completion of the bulk electrolysis process can be predicted from the applied potential and the Nernst equation. Supposing that, in the beginning, the concentration of O is $c_{\text{o,i}}$, E^0 is the standard electrode potential, V_{s} is the volume of solution and x is the fraction of O reduced to R at the electrode potential E, then the number of moles of O at equilibrium is $V_{\text{s}}c_{\text{i}}(1-x)$, while the number of moles of R at equilibrium is $V_{\text{s}}c_{\text{i}}x$; and Equation (6) shows:

$$E_{\text{appl}} = E^0 + \frac{\text{R}T}{nF} \ln \frac{V_{\text{s}}c_{\text{i}}(1-x)}{V_{\text{s}}c_{\text{i}}x} = E^0 + \frac{\text{R}T}{nF} \ln \frac{1-x}{x} \tag{6}$$

For example at 99.99% completeness of reaction of O to R (i.e. $x = 0.9999$) the working electrode potential is given by Equation (7):

$$E_{\text{appl}} = E^0 + \frac{0.059}{n} \log \frac{0.0001}{0.9999} \approx E^0 + \frac{0.059}{n} \log 10^{-4} \tag{7}$$

or $236/n$ mV more negative than E^0 at 25 °C.

Similarly, Equation (8) shows, for electrogravimetric conditions, when metal ions are deposited, for example, on an inert electrode such as platinum:

$$O + ne^- \rightleftharpoons R \text{ (solid)} \quad (8)$$

The Nernst equation yields Equation (9):

$$E_{appl} = E^0 + RT \ln[c_i(1 - x)] \quad (9)$$

where c_i is the initial concentrations of oxidant, and x is the fraction of the oxidized form reduced at electrode potential E.

Considering the extent of electroreduction, the potential shift due to the given degree of completion of electrolysis can be calculated (Equation 6).

From this it follows that an alteration for the electrode potential due to electrolysis must be kept in mind when designing a bulk electrolysis experiment for the selective determination of metals.

Let c_A and c_B be the initial concentrations of components A and B, respectively, and E^0_A and E^0_B the redox potential values of the relevant half-reactions. If $E^0_A > E^0_B$ then the condition for the selective separation at $x = 0.9999$ is given in Equations (10) and (11):

$$E^0_A + \frac{0.0592 \times 4}{n} E^0_B + \frac{0.059}{n} \log c_B \quad (10)$$

$$\Delta E = \frac{0.059}{n}(4 + \log c_B) \quad (11)$$

In the foregoing it was assumed that the electrochemical reduction of both metal cations is reversible.

For the case shown, the extent of the electrode reaction will be governed by equilibrium conditions, but the rate of electrolysis will be small at potentials predicted by the Nernst equation. The rate of electrolysis is naturally increased with increase of the electrode potential applied, and reaches a limiting value at a potential corresponding to the plateau region of the current–voltage curves. As a consequence the actual potential for the electrolysis is usually selected on the basis of experimental current–voltage curves of microelectrodes (see section 3) recorded under conditions close to those of the intended bulk electrolysis.

The selectivity of electrolysis techniques (electroseparation) can also be affected by the composition of the supporting electrolyte chosen for electrolysis. For the separation of two oxidation–reduction systems with similar E^0 values (e.g. $\Delta E^0 < 236$ mV at 25 °C), the addition of a complexing agent, which complexes one of the metals, may improve the separation. The addition of an agent that forms a complex with a metal ion usually shifts the E^0 value of the metal ion/metal redox system to more negative values. This approach can be employed for the separation of copper(II) from bismuth(III) ($E^0_{Cu} = 0.345$; $E^0_{Bi} = 0.320$). When the deposition is carried out in a tartrate-ion-containing buffer of pH 6 the quantitative electroseparation of the two metal ions can be accomplished. The metal deposition is often disturbed by H_2 formation, which can be avoided simply by making the solution more alkaline.

On the basis of the foregoing calculation, one can determine the standard potential difference theoretically needed for determining one ion without interference from another. This potential difference is 240 mV at 25 °C for singly charged ions, and 120 mV at 25 °C for doubly charged ones.

The working electrode potential can be altered only by variation of the potential applied to the electrolysis cell. However, as Equation (6) predicts, E_{appl} affects all potential values developed in the cell, all of which vary as the electrolysis proceeds. Thus, the only practical way of keeping the electrode potential at a preselected value is to measure the cathode potential continuously with respect to a reference electrode. The cell potential can then be adjusted within the desired range. This type of analysis technique is called controlled cathode potential electrolysis or controlled anode potential electrolysis.

The working electrode potential with respect to a third reference electrode is controlled automatically with a potentiostat.

4.1.1 Current Efficiency

The current efficiency for an electrode reaction (*i*th reaction) is equal to the fraction of total current rendered for the *i*th reaction as shown in Equation (12):

$$\text{current efficiency for the } i\text{th reaction} = \frac{i_i}{i_{total}} v \quad (12)$$

where i_{total} is the total current passing through the electrolysis cell.

A current efficiency of 100% means that no side reaction takes place at an electrode, that is, there is only one working electrode process.

It is beneficial for bulk electrolysis to be carried out with high current efficiency in order to reduce analysis time. In electrogravimetric experiments, 100% current efficiency is not required unless the side reactions produce insoluble reaction products. However, for coulometric titrations, 100% current efficiency is a prerequisite.

4.1.2 Electrolysis Cells

Electrode designs used for controlled potential bulk electrolysis experiments are shown in Figure 1. As a cathode, Pt gauze or foil cylinders of diameter 2–3 cm are

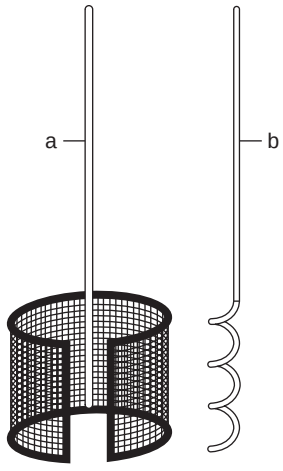

Figure 1 Electrode designs for controlled potential bulk electrolysis: (a) platinum gage working electrode; (b) platinum auxiliary electrode.

commonly used. Platinum electrodes have the advantage of being relatively nonreactive, and can be used for the deposition of a number of different metal ions. However, a protective coating is required before the electrolysis of certain metal ions, notably tin and zinc, to avoid alloy formation.

Amalgam-forming metals can be deposited on Hg cathodes. This procedure is especially used for removing easily reducible elements as a preliminary to the analysis.

The anode can also be made of Pt. Proper positioning of the auxiliary electrode is essential in order to provide uniform current density across the working electrode surface. The location of the reference electrode is also important for the long-term stability of the reference electrode and the *iR* drop.

The working and the auxiliary electrodes are often placed in separate compartments in order to avoid mixing of the electrolytes. The two compartments are usually separated by a sintered glass disk or an ion-exchange membrane of relatively low resistance that does not contribute appreciably to the overall cell resistance. The use of separators can be avoided by the judicious choice of the auxiliary reaction, in which solid products such as AgCl or inert gaseous products (e.g. N_2) are formed.

The cell resistance and the proper positioning of the reference electrode in the cell are important when the electrolysis is carried out in nonaqueous solvents with lower dielectric constant (e.g. acetonitrile, dimethylformamide, ammonia).

4.2 Electrogravimetric Methods

Electrogravimetric determinations require adherent, smooth deposits. The quality of the deposit depends, among other things, on the form of the metal ion in the solution. In general, depositions from solutions of complexed ions are smoother and more adherent than those obtained from solution containing the aquo form of the ion.

The time required for electrogravimetric analysis depends on the current density. However, too high current values may result in deposits of poorer quality. Hydrogen formation during deposition also results in a rougher deposit.

The sensitivity of an electrogravimetric method is limited by the difficulty in determining the small difference in weight between the electrode and electrode plus deposit. In general, the deposit of 0.2–5 mmol material is especially advantageous. Electrogravimetric determinations have been supplemented by coulometric methods, except if the 100% current efficiency requirement cannot be fulfilled.

4.3 Electroseparation

In the previous section, the conditions for selective quantitative deposition of one metal M_1 on a solid electrode were discussed. The selectivity in this case means that no appreciable deposition of the other metal M_2 occurs under the experimental conditions selected. Thus, for complete deposition of M_1 ($\leq$99.9%) as Hg amalgam, the electrode potential E must be selected as shown in Equation (13):

$$E \leq E_1^0 - \frac{0.18}{n_1}\, V \text{ at } 25\,^{\circ}\text{C} \tag{13}$$

where E_1^0 is the standard potential and n_1 is the number of electrons taking part in the electrode reaction, whereas for 0.1% deposition of M_2, Equation (14) shows:

$$E \leq E_2^0 - \frac{0.18}{n_2}\, V \text{ at } 25\,^{\circ}\text{C} \tag{14}$$

Thus, for 99.9% separation of M_1, Equation (15) shows:

$$\Delta E = E_1^0 - E_2^0 = 0.18(n_1 + n_2) \tag{15}$$

that is, for ions of valency one, it is 0.36 V.

For electroseparation the Hg pool cathode is the most frequently used.

ACKNOWLEDGMENTS

The Publishers are grateful to Professors Mermet and Otto for allowing reproduction of this chapter from *Analytical Chemistry*, edited by Professors Kellner, Mermet, Widmer and Otto and published by Wiley-VCH in 1998.

RELATED ARTICLES

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

FURTHER READING

D.C. Harries, *Quantitative Chemical Analysis*, 4th edition, Freeman, New York, Chapter 25, 1995.

'Gravimetric Determination of the Elements', in *Comprehensive Analytical Chemistry*, eds. C.L. Wilson, D.W. Wilson, Elsevier, New York, Vol. 1c, 1960.

I.M. Kolthoff, E.B. Sandell, E.J. Meehan, S. Bruckenstein, *Quantitative Chemical Analysis*, 4th edition, Macmillan, London, 1969.

Karl Fischer Moisture Determination

Katrin Schöffski
Sigma–Aldrich Laborchemikalien, Seelze, Germany

Dieter Strohm
Metrohm Ltd, Herisau, Switzerland

1 Introduction 1
2 History 2
3 Definitions 3
4 Instrumentation 4
4.1 The Titration Stand 4
4.2 Volumetric Karl Fischer Titration 4
4.3 Coulometric Karl Fischer Titration 5
4.4 Thermostatic Titration Cells 5
4.5 Karl Fischer Oven 5
4.6 Automation of the Volumetric Karl Fischer Titration 5
4.7 Oven Sample Processor 6
5 Reagents 6
5.1 One-component Reagent 6
5.2 Two-component Reagents 6
5.3 Auxiliary Products 7
6 Standardization, Calculation and Control of the Results 7
6.1 Titer Determination 7
6.2 Sample Size and Calculation of the Water Content 7
6.3 Control of the Result 7
6.4 Interferences 8
7 Sample Preparation 8
7.1 Solids 8
7.2 Liquids 9
7.3 Gases 9
7.4 Oven Method 9
8 Important Applications 9
8.1 Chemicals 9
8.2 Pharmaceutical Products 10
8.3 Petroleum Products 11
8.4 Plastics 11
8.5 Foodstuffs 12
9 International Standard Procedures 12
10 Comparison with Loss-on-drying 12
Abbreviations and Acronyms 13
Related Articles 13
References 13

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

The Karl Fischer (KF) titration is a method of determining the water content of solid, liquid and gaseous samples. It is the technique preferred for use in industrial quality control. In principle it involves the oxidation of sulfur dioxide by iodine, in the presence of water, in a buffered solution. An alcohol is used as the preferred solvent. The water is converted stoichiometrically and therefore its quantity is determined indirectly. The end-point (EP) is reached when there is an excess of iodine. It can be indicated visually, photometrically or electrochemically. According to state-of-the-art technology a double platinum electrode determines a voltammetric indication. The KF titration can be carried out either volumetrically or by coulometry. For a volumetric titration the iodine is added by volume to the titration cell containing the sample. In a coulometric titration iodide is oxidized at a platinum electrode and the iodine formed reacts with the water. The amount of current necessary in the generation of the iodine is directly related to the quantity of iodine generated, according to Faraday's first law. In practice, special titrators for the KF titration are available. Preformulated reagents are similarly available ready for use. Water in an extensive number of materials can be determined by KF titration. Chemicals, pharmaceuticals, oils, plastics and foodstuffs are examples of typical sample types. The measurement range spans a few ppm to 100% water. The KF titration is described in ISO 760 and as part of many other international standard procedures. Other methods for the determination of water content include loss-on-drying, IR spectroscopy or azeotropic distillation.

1 INTRODUCTION

The water content of various products is an important quality consideration in industrial processes, and, for a raw material, is often taken as a correction factor in the measure of its worth. For example, for oils and cereals, the proportion of water in the gross weight of a delivery is taken into consideration in the price. For intermediate products, such as plastic granulates or tablet powders, the water content influences the mechanical characteristics, e.g. the flow capability in a press. In the final product of a foodstuff the water content must sometimes be determined by law and can also influence the storage stability. For these reasons an exact determination of the water component of the material is necessary.

It is necessary to distinguish between moisture and water determination. To be exact, moisture is referred to

as bound water. It can also be used to describe a bound solvent or an evaporated substance. In general, moisture is determined by using an oven. Water, on the other hand, is determined chemically by the KF titration. Both surface water and water bound in crystals can be determined in this way.

Water determination according to the KF method is an analytical technique for determining the water content in various matrixes. Chemically, it involves the oxidation of sulfur dioxide by iodine, with the consumption of water, in a buffered solution. Organic solvents are used as the solvent for the sample as well as for the working medium in the titration cell. Methanol, ethanol and higher propylene glycol mixtures are most often used. The KF titration can be carried out both volumetrically and coulometrically.

The volumetric titration is carried out with a one- or two-component reagent. When a one-component reagent is used, the buret of the titrator contains all the necessary reagents dissolved in an inert solvent. Methanol or a mixture of alcohols serves as the working medium in the titration cell. This is then pretitrated by the KF reagent in order to remove any traces of water from the cell and working medium. Then the sample is added and its water content determined automatically.

In a two-component reagent the reactive components are separated. The so-called ''titrant'' contains a preset amount of iodine in a solvent, which can be either methanol or ethanol. The buret of the titrator is then filled with this solution. A base and sulfur dioxide are dissolved in the ''KF solvent'', which is added to the titration cell and pretitrated to dryness. The sample is then added and its water content determined.

The one-component reagent has the advantage that it can be used with a wide variety of different working media. The two-component reagent is distinguished by its faster titration rate and greater polarity. Typical water contents for volumetric KF determinations are between 0.5 and 50 mg water. Figure 1 shows a typical KF titration curve of time vs reagent added.

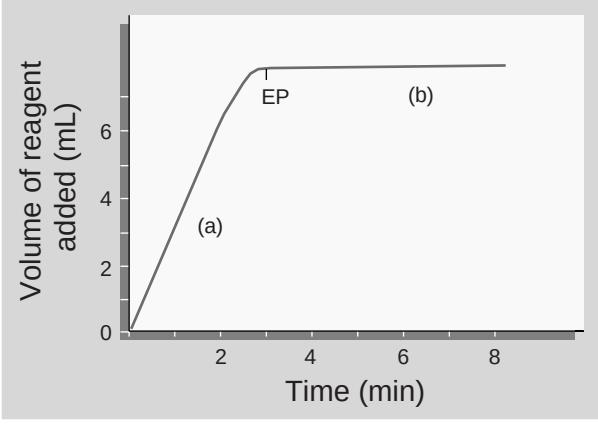

Figure 1 Typical Karl Fischer titration curve. (a) Reagent adding mode; (b) conditioning mode.

A different composition is used for KF coulometry. The reagents contain a solution of iodide rather than iodine, which is then generated from the iodide at a platinum electrode. The iodine then reacts with sulfur dioxide according to the KF chemistry, with the loss of water. The current necessary to generate the iodine is measured. The amount of iodine is determined using Faraday's law and the water content consequently calculated. To carry out a coulometric determination the reagents are added to the titration cell. These solutions must first be pretitrated as in the volumetric method. Any water in the titration cell or dissolved by the reagent from the air is thereby removed. Since coulometry is more sensitive, this pretitration process takes longer than a volumetric titration. The sample is injected, via a septum, into the reagent. Modern titrators calculate the amount of water automatically after injection of the sample. Typical absolute values of water analyzed by coulometry are between 50 and 2000 μg.

2 HISTORY

In 1935 the German chemist Karl Fischer was required to determine the water content of liquid sulfur dioxide. Neither loss-on-drying nor distillation were suitable for liquid samples, and he had to find an alternative method. During his investigations he stumbled across the Bunsen reaction, Equation (1), which describes the iodometric determination of sulfate in a buffered, aqueous solution.

$$SO_2 + I_2 + 2H_2O \Longleftrightarrow H_2SO_4 + 2HI \qquad (1)$$

According to Equation (1), a determination of the water content should also be possible if the sulfur dioxide is present in excess and the protons that are consequently produced are taken up by a base. Fischer put together his reagent using sulfur dioxide, iodine and pyridine, with methanol as solvent. He formulated Equation (2),[1]

$$SO_2 + I_2 + 2H_2O^*Py \Longleftrightarrow H_2SO_4 + 2HI^*Py \qquad (2)$$

where Py represents pyridine. The EP was indicated by the color change from yellow to brown. Using this one-component reagent, he carried out the first volumetric KF titrations and determined the water content of sulfur dioxide and various different solvents.

In 1939 the reaction mechanism was re-examined by an American research group. Smith et al. formulated the two-step reaction, Equation (3), with a different molar relationship of water to iodine of 1 : 1,[2]

$$H_2O + SO_2{}^*Py + Py^*I_2 + Py + MeOH \Longleftrightarrow 2HI^*Py + Py^*SO_3 \quad (3a)$$

$$Py^*SO_3 + MeOH \Longleftrightarrow Py^*MeSO_3H \quad (3b)$$

where MeOH represents methanol. By 1943 Wernimont and Hopkinson had introduced the electrochemical dead-stop EP detection for the KF titration.[3]

During the years that followed, there were further improvements in KF reagents,[4] but the next big step came in 1959 when Meyer and Boyd published the coulometric KF titration.[5] Further investigations were published in the 1970s concerning the kinetics of the reaction. In 1974 Cedergren discovered that the reaction rate is dependent on the concentration of iodine, sulfur dioxide and water.[6] Verhoef and Barendrecht found that pyridine functioned only as a base and was not part of the reaction itself.[7] From these observations Scholz formulated Equation (4), the version of the reaction equation that has been accepted since 1984,[8]

$$SO_2 + HO\text{-}R + B \Longleftrightarrow R\text{-}SO_3{}^- + BH^+ \quad (4a)$$

$$R\text{-}SO_3{}^- + I_2 + H_2O + 2B \Longleftrightarrow R\text{-}SO_4{}^- + 2I^- + 2BH^+ \quad (4b)$$

where B is a base. An alkyl sulfite is formed within the reagent: from the alcohol used as solvent and the sulfur dioxide (Equation 4a). Wuensch and Seubert isolated this reaction product in 1988.[9] The alkyl sulfite is oxidized by iodine in the presence of water in the second step (Equation 4b). The equations illustrate two essential points. Firstly, in order to guarantee exact stoichiometry, an alcohol must be present in the KF titration. In practice, it is found that approximately 50% methanol or ethanol is necessary in the working medium. Secondly, since the pyridine does not take part directly in the reaction, it should be possible to replace it by other bases. Pyridine has been unwelcome in laboratories since the end of the 1970s due to its unpleasant odor and the fact that it is harmful. Since the 1980s pyridine-free reagents have been the standard.

Comprehensive application support and information on the titration of many different samples is available in the form of the monograph the *HYDRANAL Manual*.[10] Imidazole is used most often as the base in pyridine-free reagents. The reagents manufactured with imidazole as base have a higher pH because imidazole is a stronger base than pyridine. Since the reaction rate, K, of the KF reaction is dependent on the pH (Figure 2), a faster titration and more stable EP can be achieved.[11] By 1998 further advances in the field of the KF water determination included the development of a coulometric cell without a diaphragm, and reagents free of harmful materials.[12,13]

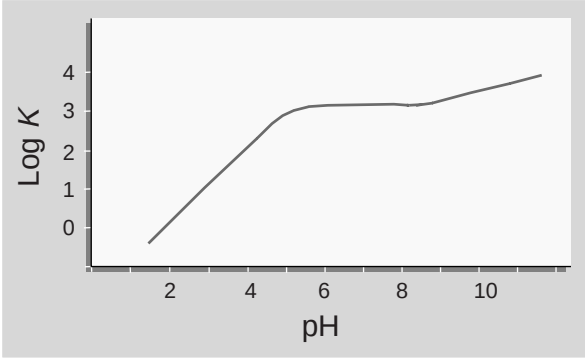

Figure 2 The relationship between reaction rate, K, and pH.

3 DEFINITIONS

One-component reagent. A volumetric reagent containing sulfur dioxide, base and iodine in an organic solvent. It is used with working media such as alcohols, or mixtures of organic solvents and alcohols.
Two-component reagent. A volumetric reagent combination consisting of a titrant and solvent. The solutions are not mixed prior to use but are used separately, in contrast to the former pyridine-containing KF reagents.
Titrant. An alcoholic iodine solution.
Solvent. The solution containing the base and sulfur dioxide and which is added to the titration cell.
Anodic reagent = anolyte. A reagent consisting of iodide, a base and sulfur dioxide in an alcohol. It is the reagent in which the coulometric KF reaction takes place.
Cathodic reagent = catholyte. A reagent consisting of organic salts in an alcohol. It is the reagent in which the coulometric cathode reaction takes place, i.e. the reduction reaction. It counteracts the iodide oxidation and normally involves the reduction of protons to hydrogen.
Titer. The water equivalent of the measurement reagent given in mg water per ml reagent.
Drift. The blank value in the titration cell prior to the addition of water. Drift is caused by moisture ingress into the cell and by side-reactions which consume iodine. Modern titrators determine and display the drift automatically. The drift is given in different units depending on the instrument supplier: $\mu g\,min^{-1}$, $mg\,min^{-1}$, $\mu L\,min^{-1}$ or $\mu g\,s^{-1}$ are in use.

4 INSTRUMENTATION

Due to the special qualities of the water content determination by KF titration it is perhaps understandable that specific analytical instruments for this titration have been on offer for many years. As a result of the enormous developments in the field of microprocessor technology, many of these analytical instruments are now driven by microprocessors. In practice, there are two different instruments accepted for the KF water determination: the volumetric KF titrator and the coulometric KF titrator. Both techniques use the EP titration method, and the titration stand consisting of the titration cell and stirrer plays a particularly critical role.

4.1 The Titration Stand

The atmospheric moisture present in every laboratory is a common source of error in KF titration. This moisture can penetrate the sample, titration reagent and titration cell, and thereby cause incorrect results. Modern analytical instruments for the KF water determination (such as that in Figure 3) reduce the influence of atmospheric moisture as far as possible by ensuring that the following measures are incorporated into the titration stand.

The stand is particularly airtight and has special input ports for the addition of solid, liquid or gas samples. In order to minimize the ingress of atmospheric moisture when changing the reagents, a pump is often integrated into the stand. This pump can suck out used reagent and pump in fresh without having to open the titration cell. During the pumping, it is impossible to avoid air being sucked into the titration cell. Moisture ingress in this way is normally avoided by passing the air over a molecular sieve. No titration stand is absolutely closed to the atmosphere and very small quantities of water can always seep in. Modern KF titrators continually condition the titration cell and determine the quantity of water that has seeped into the cell over a particular period of time. This value is displayed as drift, e.g. as μg H_2O or μL of titration reagent required per minute. The drift is continually on display and the user therefore has a constant measure of the condition of the titration stand. The drift value is also measured at the beginning of the titration and it can therefore be taken into consideration in the final result.

Figure 3 A typical, automated volumetric Karl Fischer titrator.

As with all EP titrations, good mixing of the reagents is an important prerequisite for a fast and exact KF titration. Therefore a magnetic stirrer is built into the titration stand of a modern analytical instrument for KF water determination.

4.2 Volumetric Karl Fischer Titration

KF titrators of today are, without exception, fully automatic. The water determination is carried out independently by the press of a button, with the water content calculated and displayed or printed out, according to good laboratory practice (GLP), on an attached printer. To ensure that an automatic titration is possible the volumetric KF titrator has the following essential components.

4.2.1 Motorized Volumetric Buret with a Titration Stand

This buret enables a precise dosing of the titration reagent. To ensure that the hub can accurately add reagent in the microliter range, it is set up with 10 000 steps.

4.2.2 Exchange Unit

The exchange unit was developed to dose the titration reagent. It consists of a cylinder, buret, tap, buret tip and the stock bottle containing the titration reagent. In order to protect the reagent from ingress of atmospheric moisture, the stock bottle is fitted with a drying tube.

4.2.3 Indication System

Normally a bi-voltammetric indication is used as the indication system for the EP of a volumetric KF titration. For this a double platinum electrode with a constant current (e.g. 50 μA) is used. Whilst iodine is reacting with water in the titration cell, there is no free iodine in the titration solution. A high voltage is necessary to maintain the preset polarization current of the electrode. As soon as all the iodine has reacted with the water, free iodine will remain in the titration solution. This free iodine considerably reduces the voltage, which is necessary to maintain the preset polarization current.

4.2.4 Controls for the End-point Titration

In a KF titration the titration reagent must be added as fast as possible from the buret and exchange unit, and the addition should be stopped exactly at the EP. The control of modern KF titrators adapts itself to the course of the titration curve and thus leads to short titration times and very accurate results.

4.2.5 Integrated Communications

A display and keypad are available to the user in order to input control parameters, sample data for the calculation and the documentation and to be able to read the results. Top-of-the-range KF titrators display a live titration curve in the display, which can give the user useful information about the course of the titration.

4.2.6 Method of Storage

A storage capability is necessary to save preset KF titration methods. In each laboratory at least two methods are necessary: the titer determination and the water content determination. If various different samples are to be analyzed the optimum titration conditions can be saved for each different sample type. Then all that has to be done is to call up the right sample, add the sample and start the determination.

4.2.7 Interfaces

Interfaces for a balance and data system (laboratory information management system (LIMS)) enable the KF titrator to be integrated into the modern analytical laboratory.

4.3 Coulometric Karl Fischer Titration

Coulometric KF titrators are also exclusively automatic. The water determination is carried out independently at the press of a button, with the water content calculated and displayed and documented on a printer connected to the titrator. They are differentiated from volumetric titrators by the following aspects. Instead of the motorized buret to dose the iodine solution, coulometers have a generator current circuit such that iodine can be generated electrochemically. For the iodine production there is a generator in the cell rather than the buret tip. In practice there are two types of generator electrode: those with and those without a diaphragm. Wherever possible a generator without a diaphragm is preferred because handling is then easier.

Owing to the smaller amounts of water being determined, a more sensitive indication set-up is preferred. The indication principle is, in effect, identical to that in the volumetric KF titration, only here an alternating current is used for the polarization of the double platinum electrode. Coulometric KF titration is a micro method and particularly suitable for the determination of small quantities of water. Owing to the high sensitivity, particular consideration has to be paid to contamination by moisture from external sources.

The water content of a wide range of samples, from diverse industrial processes, can be determined by using the KF technique. For this reason, special instruments are available for the preparation of samples that are analyzed regularly. Additionally, automated instruments are available.

4.4 Thermostatic Titration Cells

Substances which dissolve slowly in methanol or the KF working medium, cause drifting EPs. The same is true for solids, for example foodstuffs, which give up their water only very slowly. In such cases, titration at 50–60 °C can accelerate the dissolution of the sample or the release of its water. Slow side-reactions sometimes involve iodine and therefore a stable EP is impossible to achieve. It can then be advantageous to carry out the titration at low temperatures, which can reduce the effect of the side-reaction.

Special titration cells are available for these techniques. They are constructed with an outer glass jacket through which a liquid can be pumped. When attached to a thermostat or cryostat the titration temperature can be adjusted.

4.5 Karl Fischer Oven

Many substances only release their water at high temperatures and are therefore unsuitable for the KF titration. These samples can be heated in an oven to between 100 and 300 °C and their water evaporated. The evaporated water is then carried over into the titration cell via a heated, inert stream of gas and titrated either coulometrically or volumetrically. The KF oven can be used for insoluble solids (e.g. plastics and salts), where the water can be released at temperatures above 60 °C fairly rapidly, and also for solids and liquids that react with the KF titration reagents (e.g. ascorbic acid, mineral oil).

4.6 Automation of the Volumetric Karl Fischer Titration

The aim of many laboratories is to increase the throughput of samples and so improve business efficiency and relieve the personnel of routine work. For this reason, sample exchangers were developed. Samples are weighed into sample beakers, which are then closed with aluminum foil in order to protect them from the ingress of atmospheric moisture. When they are at the titration position they are

Figure 4 A typical, automated sample processor.

raised and the foil is pierced. A defined volume of solvent is added. Then the desired dissolving time or extraction time elapses before the KF titration is started. Using such equipment, large numbers of samples can be processed fully automatically. Sample exchangers will also work overnight or during the weekend.

4.7 Oven Sample Processor

Water determination in plastics, oils, emulsions or salts often cause problems. They can cause contamination of the oven and titration cell, side-reactions, reduced lifetime of the reagent, long analysis times and high cost due to the fact that the possibility of automation is limited. An oven sample processor (Figure 4) has been developed to avoid direct contact between the sample and titration cell and thereby avoid the above-mentioned problems. In addition, the full capacity of the reagent is utilized to the optimum. Both the KF volumetric and coulometric titrator can be set up in coordination with the oven sample processor. The samples are weighed in the small (glass) sample boat, which is then sealed with a septum and placed on the exchanger. When the small sample boat reaches the workstation, the oven sample processor is lowered and the septum is pierced with a needle. At the same time the sample boat is pushed into the aluminum heating block and the heater is switched on. The needle has two openings: an inert gas enters the sample boat via the first one, and leaves through the second. The inert gas then passes into the titration cell along with the evaporated water of the sample. The water is then titrated.

5 REAGENTS

The reagent Karl Fischer described has a number of disadvantages. Apart from the strong smell of pyridine, it displays a very fast fall in the titer value. One-component KF reagents are generally unstable. A slow-side reaction, which involves iodine, takes place within the reagent, and so diminishes the titer.[14] Pyridine-containing reagents often decompose completely within a very few weeks but a modern pyridine-free reagent has a titer decrease of less than 10% per year if stored correctly. Additionally, the ingress of atmospheric water into the reagent can cause the titer value to decrease. Air humidity has a great influence on the stability. A liter of air contains up to 15 mg water and therefore it must be dried before being pumped into the titration cell or bottle of reagent.

5.1 One-component Reagent

A pyridine-containing one-component reagent has been described.[15] The titer of these reagents is not given, but must be determined daily since it drops rapidly. Commercially available, pyridine-free reagents are offered with various different concentrations. The water equivalence is often contained in the name of the reagent. Reagents of titer 5, 2 and 1 are the norm. They are used in conjunction with methanol or other special working media. The working media can be varied in order to improve the solubility of the sample or to minimize side-reactions. A chloroform/methanol mixture would be used, for example, to aid the solubility of fats. An ethanolic medium will reduce a side-reaction of the sample with methanol.

Aldehydes and ketones react with the components of a usual KF reagent, and therefore special reagents are available for such samples. They have a K in the name, referring to ketone. The water determination is carried out with a one-component reagent of titer 5 in a special, methanol-free working medium. The main components of this working medium are chlorinated or methoxylated alcohols.

5.2 Two-component Reagents

Two-component reagents consist of a titrant and a KF solvent. The titrant is available with a titer of 5 or 2. For general use both methanol- and ethanol-based KF solvents are used. Solvents such as chloroform or formamide can be added to the KF solvent to aid sample solubility. There are also special products for nonpolar samples available.

5.2.1 Coulometric Reagents

Reagents for coulometry are in the form of an anolyte and catholyte. The KF reaction takes place in the anode reagent. This contains sulfur dioxide, an organic base and a soluble iodide in an alcoholic solvent. Anode reagents

have an A in the name. For general purposes an anolyte that does not contain a chlorinated hydrocarbon is used. For oils, a chloroform-containing anolyte is used; for ketones, a methanol-free anolyte. The reduction takes place in the associated catholyte. The sample does not come into direct contact with the catholyte, and therefore an alcoholic catholyte usually based on methanol can be used for almost all applications. The methanol-free reagent is necessary only for a ketonic sample. Catholytes can be recognized from the C in their name.

5.3 Auxiliary Products

In addition to the actual KF reagents, a whole array of auxiliary products is available. Buffers and acids allow the setting of an optimal pH. Karl Fischer standards with a predetermined water content are used to carry out a titer determination and to check the functioning of the instrumentation.

6 STANDARDIZATION, CALCULATION AND CONTROL OF THE RESULTS

6.1 Titer Determination

Although a titer is given for volumetric reagents, this must be checked at regular intervals. The titer of a one-component reagent falls due to a side-reaction and there is always the possibility of ingress of moisture from the atmosphere, which also reduces its value. Since the titer is actually a measurement of volume, there is also a temperature dependency.

The titer can be determined using pure water, salts with a constant water content and liquid water standards. Pure water is injected into the cell using a microliter syringe. The titer is calculated according to Equation (5):

$$\text{titer} = \text{water equivalent} = \frac{\text{amount of water}}{\text{volume of reagent}} \qquad (5)$$

where the amount of water is in mg, and the volume of reagent is in mL. Sodium tartrate-2-hydrate is used as the solid standard. This salt is not hygroscopic, it dissolves in methanol and has a constant water content of 15.66%. The titer is calculated according to Equation (6):

$$\text{titer} = \text{water equivalent} = \frac{\text{sample size} \times 0.1566}{\text{volume of reagent}} \qquad (6)$$

where the sample size is in mg, and the volume of reagent is in mL. Liquid water standards consist of a non-hygroscopic solvent mixture that contains 0.01–1% water. The standard (1–2 g) is weighed into the cell by difference. The calculation for the 1% standard would be as in Equation (7):

$$\text{titer} = \text{water equivalent} = \frac{\text{sample size} \times 0.01}{\text{volume of reagent}} \qquad (7)$$

where the sample size is in mg, and the volume of reagent is in mL.

The titer determination is not only an exact measure of the water equivalent but is also a control of the working conditions. If the titer falls dramatically, this can indicate moisture in the reagent. If the values vary considerably from one another the titration cell may not be firmly closed.

A titer determination or any other calibration procedure is unnecessary in the case of coulometry. It is an absolute method in which the amount of iodine is calculated by using Faraday's constant.

6.2 Sample Size and Calculation of the Water Content

Sample size depends on the method, water content of the sample and the desired accuracy. Table 1 gives some idea of appropriate sample sizes.

6.3 Control of the Result

When validating the KF titration, various sources of error need to be considered. The result can be incorrect due to an error in the titer value, or a deficient reagent or instrument. Additional errors due to a reaction between the sample and reagent are also possible. The whole titration system, i.e. the instrumentation together with the reagents, can be checked using a certified water standard. The water content of the standard is determined during its manufacture by obtaining numerous results at various sample weights. In order to guarantee the KF system, the result obtained with the standard must lie within a predetermined standard deviation limit. Table 2 shows a typical overview of results.

This procedure checks both the KF apparatus and reagents together. If the results do not fall within the accepted limits of error then the source of the error must

Table 1 Recommended sample sizes

Water content of sample (%)	Sample weight (g) for volumetric titration	Sample weight (g) for coulometric titration
100	0.05	–
10	0.5	0.05
1	5	0.2
0.1	10	2
0.01	10	5
0.001	20	10
0.0001	–	10

Table 2 Typical results from a reliability test of a KF coulometer

Sample no.	Weight of water standard (g)	Water (calculated) (mg)	Water (found) (mg)	Water content ($mg\,g^{-1}$)
1	0.5355	0.5301	0.5366	1.0021
2	0.5001	0.4951	0.5089	1.0176
3	0.4785	0.4737	0.4794	1.0019
4	1.0268	1.0165	1.0405	1.0133
5	0.9911	0.9812	0.9866	0.9955
6	0.9994	0.9894	1.0046	1.0052
7	1.9515	1.9320	1.9575	1.0031
8	1.9800	1.9602	1.9964	1.0083
9	2.0169	1.9967	2.0060	0.9946

Water content:
Mean of nine determinations, 1.0046
Relative standard deviation, 0.75%
Standard deviation control limit, 2%
Control requirements fulfilled? Yes.

be found and eradicated. Then the procedure is repeated with fresh reagents. The drift and titration curve should be noted. The balance being used must also be checked. If the error persists then the buret of the volumetric titrator can be checked. In order to do this, definite volumes of water are dosed at 20 °C from the buret and these volumes are then compared with their exact weight. Further details can be obtained from the manufacturers of the instrumentation.[15] If the source of error remains, the manufacturer should check the instrumentation.

Manufacturers of the instruments are able to certify them, and also support the user in many ways in terms of quality management in the analytical laboratory. Practically all manufacturers are ISO 9001 accredited and consider the requirements of the international standards of quality management during the development of their instrumentation. This support involves all phases of instrument procurement and use. Advice involves the set-up of the equipment, the necessary solutions to problems, delivery of the instrument, instructions of use, operator training, instrument servicing, and solving customer problems. On request, the manufacturer will help the operator with installation qualification (IQ), the operation qualification (OQ) and performance qualification (PQ). With the integration of GLP tools, for example the monitoring of validation intervals, reagent life and result limits, the operator is efficiently supported by the instrument itself during daily determination of water content by KF titration.

6.4 Interferences

If it is to be determined whether or not a sample is interfering with the KF chemistry, the method of ''subsequent standard addition'' is recommended. The working medium of the volumetric titration or the anolyte in coulometry is titrated to dryness. Then the sample is added and its water content found. A known amount of water is then added to the titration cell still containing the titrated sample. If the amount of water titrated is 100%, within a predetermined standard deviation control limit, interference with the sample can be eliminated.

7 SAMPLE PREPARATION

Due to the wide area of application for the KF titration, there are countless different samples where the water content needs to be determined. It is only possible here to cover the sample preparation and methods for the various different classes of substance rather than individual materials.

7.1 Solids

Solids usually have water bound in two different states. It can be adsorbed on the surface of the solid or be in the form of crystallized water or trapped water in the solid. In order to determine the whole water content it is preferable that the sample is fully dissolved. Since many materials do not dissolve in methanol, additional solvents are often added to aid solubility. For lipophilic substances, e.g. fats, a working medium is used which consists of 50% methanol or KF solvent and the remainder is chloroform or a suitable alcohol. To dissolve polar substances the working medium is made up of 50% formamide.

Mechanical sample preparation is also possible. For better solubility it is always recommended to reduce the size of large pieces. The sample can be crushed using a pestle and mortar, ground in a laboratory mill, or cut into small pieces. If there is no working medium that dissolves the sample, an extraction titration is sometimes also possible. The sample is ground as fine as possible and suspended in the working medium. The water is then titrated. The hygroscopic solvent extracts the water from the solid. After the EP has been reached it is recommended that a subsequent standard addition is carried out, as described in section 6.4, to ensure that the water is completely extracted. To improve the distribution of the sample, a high-speed stirrer or homogenizer (Ultra-Turrax®, IKA Werke, Stauffen, Germany) can be installed inside the titration cell. The sample is dispersed during the titration, which optimizes the conditions for such samples.

A further possibility for determining the water content of insoluble samples is external extraction. Here, a known amount of a hygroscopic solvent, typically dried methanol, is added to the sample in a volumetric flask. After a period of stirring, an aliquot is taken and the

water content titrated. The result must take into account the blank value of the methanol. When small quantities of water are analyzed, not only the blank value of the solvent but also a blank of the procedure should be determined. For this an identical volumetric flask is taken, filled with the same volume of methanol, stirred under the same conditions for the same length of time and an aliquot taken and titrated to the same volume. Already titrated KF solvent can also be used to extract water from samples. Pretitrated anolyte from coulometry is best suited here due to the very small amounts of water involved. A syringe is rinsed numerous times with the pretitrated anolyte until the syringe is completely free of water. A distinct volume of water-free anolyte is then used as extracting solvent. In this case the blank is always zero.

7.2 Liquids

Liquids are ideal samples for the KF titration. They are injected into the titration cell, via a septum, by using a syringe. The cell remains in an ideal water-free condition because ingress of atmospheric water from outside can be minimized. The coulometric cell should, in general, never be opened for sample addition. Therefore, a sample in a liquid or gaseous state is a prerequisite for the KF titration by coulometry. The mass of sample titrated is measured by weighing by difference. Dosing by volume is generally less exact than by weight.

Liquids can also present a solubility problem or, put another way, a problem with mixing, if the sample and working medium have very different polarities. In principle, a titration in two phases is possible with sufficient stirring, but, where possible, only one phase should be present. To improve the solubility of fats, oils and other nonpolar samples, chloroform, propanol or other higher alcohols, e.g. hexanol, can be used. Polar samples can be dissolved if formamide is added.

7.3 Gases

The water content of gases can also be determined by the KF titration. The gas is directed into the titration cell. The water content of the gas is then transferred to the KF reagent. Sometimes the gas dissolves in the reagent and sometimes it bubbles out. If the bubbles are not too large this does not affect the KF titration. In practice, a fine stream of bubbles is achieved by using a sieve-type tip to the insert tube and therefore good dispersion of the gas. Gases normally contain very small amounts of water and therefore coulometry is the preferred method of analysis. Volumetric titration can also be carried out, although a reagent with a low titer must be used. The amount of gas passed through the titration cell is measured in volume and the water content result obtained is consequently in $mg\,L^{-1}$. It must be noted that the water content of gases is not always distributed homogeneously. If the water is adsorbed on the inside of the gas cylinder, higher values are obtained at the start of the determination than when equilibrium has been reached. In order to ensure that there is a homogeneous distribution of water in the gas, a constant flow of the gas should be allowed to bubble through the titration cell for an hour and the drift observed. A titration should only then be carried out when the drift has reached a stable value.

7.4 Oven Method

For the water content of samples that react with the KF reagents or that are completely insoluble, the oven method is recommended. This method is described in section 4.5.

8 IMPORTANT APPLICATIONS

The KF reaction is a redox reaction that takes place in an organic solvent. Oxidizing and reducing agents affect the reaction just like substances that react with the working medium. Other interferences take place when a sample is not fully dissolved. Here the water is only released slowly and the titration displays a drifting EP. In order to ensure that the sample fully dissolves, different solvents or physical techniques can be used. The following section gives an overview of the most frequently titrated sample types. A more complete coverage of sample applications is given in the *HYDRANAL Guide*.[16]

8.1 Chemicals

8.1.1 Inorganic Compounds

Inorganic salts rarely dissolve in methanol. In order to guarantee the determination of the entire water content, the salt must first be fully dissolved. A mixture of 50% formamide and 50% methanol dissolves some alkali and alkali earth salts. Prior grinding of the sample and warming of the titration cell accelerates solubility. Most inorganic salts do not dissolve in organic solvents. The oven method is then used. The water is evaporated from the sample, carried into the titration cell via an inert gas, and titrated according to the KF method. Many oxides and carbonates react with the KF reagent and a direct titration is not possible.

Inorganic salts and bases can also be titrated by KF titration, in principle, although the pH in the titration cell must be considered. The reaction proceeds too slowly if the environment is too acid, and the titrator consequently finds no EP. In an alkaline environment there is a side-reaction that consumes iodine and therefore causes a high

result or a vanishing EP. When titrating acids, a suitable buffer, e.g. one based on imidazole and sulfur dioxide, is recommended and, for basic samples, salicylic or benzoic acid is added to the working medium.

8.1.2 Organic Compounds

Organic compounds are very diverse and often characterized by many different functional groups. Whether the compound causes a side-reaction or not can often be judged prior to titration, by taking into consideration the functional groups present. If the substance is sensitive to oxidation or reduction it could react with iodine or iodide. Active carbonyl groups can react with methanol; iodine and sulfur dioxide can add across double bonds. Figure 5 shows titration curves with and without a side-reaction.

Hydrocarbons, halogenated hydrocarbons, alcohols, ethers, esters and other similar substances normally pose no difficulties for KF titration. Propanol or chloroform must sometimes be added to aid the solubility of long-chain compounds. Hydrocarbons that are mainly unsaturated behave inertly in the KF reagent. However, if a multiple bond is present, an addition reaction can take place which consumes iodine. Most carboxylic acids do not have to be neutralized in the same way as mineral acids. However, very strong acids, such as dichloroacetic acid or bromoacetic acid, require the addition of a buffer based on imidazole. Some carboxylic acids form esters with alcohols and therefore need to be titrated very rapidly or require a modified working medium. Working media based on ethanol esterify slower than those based on methanol. Salts of carboxylic acids can be dissolved, often with addition of formamide. Basic organic compounds such as amines require the addition of benzoic or salicylic acid to maintain the optimum pH.

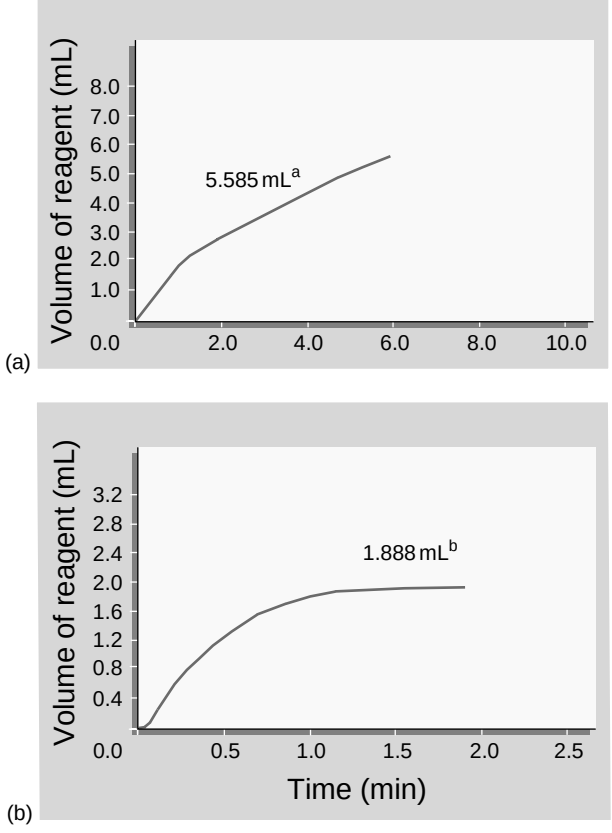

Figure 5 Titration curves with and without side-reaction. (a) Titration of benzylamine in an unbuffered solution. (b) Titration of benzylamine in a buffered solution. [a] This is the volume of reagent consumed until the titration is stopped manually; [b] this is the volume of reagent consumed until the titration is stopped automatically.

Polar organic compounds such as sugars and proteins are titrated volumetrically because they are solids. These substances are dissolved by using a working medium of formamide–methanol 1 : 1. Warming the titration cell accelerates the dissolving process.

Aldehydes and ketones present particular difficulties for the KF titration. These substances form acetals and ketals with methanol and these reactions release a stoichiometric quantity of water. There are special reagents available for these sample types which contain alcohols that react according to the KF reaction shown in Equation (4) but which do not form acetals and ketals with the sample. Alcohols such as chloroethanol, or sterically hindered compounds such as 2-methoxypropanol, can be used as suitable alcohols. Aldehydes also react according to the bisulfite addition reaction with the KF reagent. Sulfur dioxide reacts with aldehyde groups and this reaction requires water. Therefore with conventional KF reagents a result is obtained which is too low. Special reagents for aldehydes have a reduced sulfur dioxide content to decelerate the bisulfite addition reaction and enable the KF titration to be carried out accurately. Only for very reactive aldehydes, e.g. acetaldehyde, does the side-reaction predominate.

An example of substances that are sensitive to oxidation and which react with iodine is the mercaptans. Peroxides are sensitive to reduction and react with iodide to generate iodine. There are many exceptions within each group of compounds and these can behave differently in the KF environment. It is therefore worthwhile carrying out a test titration in each case. A two-step validation is necessary with addition of standard after the titration has been carried out, as described in section 6.4, and the titration curve should be investigated.

8.2 Pharmaceutical Products

Different substance groups can be thought of under the term pharmaceuticals: raw materials, intermediates and finished products. Raw materials consist of the active substances, drugs derived from vegetable material, inorganic

salts, e.g. sodium phosphate, solvents, e.g. ethanol or polyethylene glycol, and non-active components such as saccharin and isomalt. For these materials there is often a valid, quality control method according to the relevant pharmacopeia. The methods quoted by the pharmacopeia for the determination of water content are KF titration and loss-on-drying.[17–19]

Interference with the KF reagents can occur and can be avoided as described in section 8.1. Normally, raw materials are pure substances whose behavior can be determined by looking at their chemical structure.

If a reagent being used has a different composition to that described by the relevant pharmacopeia, it has to be validated as described in section 6.4. This validation must be carried out for each individual product being tested. Sometimes loss-on-drying is quoted as the preferred method for water determination. A KF titration can be used instead if a validation is carried out. This validation is necessary since in some cases loss-on-drying and KF titration give results that are not identical. The water content given in the specification of the raw material must also be taken into consideration. If the water content value found by the KF titration is within the specification control limit then this method can be employed.

Intermediates are mixtures of active and non-active substances that are not yet in their final state. The water content is an important quality criterion since it affects the physical characteristics of the product, such as its flow ability or its adhesiveness. In comparison to the quality requirements of raw materials, where, as a rule, only a maximum value for water content is quoted, for intermediates there is normally both an upper and lower limit specified. Tablet powders and other similar solids often display solubility problems. External extraction (section 7.1) or the KF oven (section 4.5) must then be used.

Similarly, in the finished product, the water content is an important quality criterion because it influences physical properties and storage stability. The water in tablets and capsules can be determined using either the KF oven or by external extraction. Direct titration is rarely possible. Effervescent tablets that release water and carbon dioxide are particularly problematic. A KF titration is not possible in this situation. Liquid medicine can be added direct to the titration cell. Creams and suppositories can also be added to the titration cell and dissolved by using either propanol or chloroform, and warming the cell if necessary.

8.3 Petroleum Products

Oils, lubricants and related products generally do not dissolve in the methanolic or ethanolic working medium of KF titration. In addition, they contain very small amounts of water (with the exception of crude oil), and so the determination must be carried out very accurately. Coulometry is recommended when used in conjunction with an anolyte modified specifically for oils. These special reagents contain chloroform or a long-chained alcohol to aid solubility of the oils. Sometimes xylene is added.

Crude oils must be homogenized before a sample is taken, because the water is often unevenly dispersed. Homogenization can be carried out either by a high-speed mixer or in an ultrasonic bath. Crude oil contains tar, which can stick to the electrodes. Xylene or toluene should therefore be added to the working medium of a volumetric titration and likewise to the anolyte in coulometry. In this way the tar stays in solution and cannot stick to the electrodes. Refined products such as benzene, diesel and kerosene are added to the specially modified anolytes designed for oils and fats. Disturbances with such products are not expected.

Lubricants, insulating oil and motor oils contain additives to improve their performance and increase their active life. Many of these additives cause problems for the KF titration. Antioxidants and mercaptans react with iodine, ketone groups form ketals, and metal oxides create water. Generally, the water content determined for these products by direct titration is far too high. Therefore the oven method (section 7.4) is carried out. The oil is heated at a temperature between 120 and 140 °C. Coulometry is absolutely necessary because the water contents are usually in the ppm region. The water content determination of petroleum products is controlled by a large number of ASTM (American Society for Testing and Materials) and ISO standards.

8.4 Plastics

The water content of polymers, like pharmaceuticals, determines their physical characteristics and therefore it is a very important criterion. The samples to be analyzed tend to be in the form of granules, fibers or solutions. The water content is often very low and therefore coulometry is the preferred choice of instrumentation. Many of the polymers can be brought into solution by the addition of chloroform or methyl pyrrolidone and can then be titrated directly. Ketones, which are also often used as solvents for polymers, are not recommended for KF titration because they cause side-reactions (see section 8.1). Small amounts of ketone can be titrated when special reagents for ketones are used.

A volumetric titration can be tried if the polymer is in the form of a fine powder. The sample is added directly to the methanolic working medium. The methanol extracts the water from the very finely ground polymer without dissolving it.

The oven method is used for less fine powders, fibers and granules. The sample is heated to between 120 and 170 °C and its water carried over to the coulometric cell in a stream of dried gas. Since the water is released very slowly a minimum determination time of 10 min must be set on the titrator to avoid it switching off prematurely. Finding an optimum oven temperature is extremely difficult because the polymers can further polymerize and form extra water. Sometimes they can decompose at higher temperatures. It is extremely difficult to distinguish between free water and water formed during the heating process and so in practice the oven temperature is standardized for each particular type of polymer.

8.5 Foodstuffs

The water content of foodstuffs is an important quality parameter. It influences, amongst other things, taste, shelf life and appearance. An upper legal limit is set for many foods. At the same time the water content determination of food is very difficult because of the very complex material involved, and requires very careful sample preparation. Food is often a mixture of polar material, e.g. protein, and nonpolar material, e.g. fat. Additionally, ethereal oils and other chemicals that enhance the taste of a food can cause side-reactions. The water can also be distributed nonhomogeneously. Sometimes it is dissolved, sometimes suspended and sometimes, in the case of plant material, trapped in cellular structures. If the samples are heated they often easily degrade, releasing water. In such cases the results have to be carefully validated. A titration curve yields a great deal of information about possible side-reactions and the course of the titration.

Volumetric titrations are carried out almost exclusively for the titration of foodstuffs. Coulometry is only used as an exception, such as for the titration of a vegetable oil. The sample must be dissolved or very finely dispersed. A working medium mixture of methanol/formamide is used for polar ingredients such as proteins and sugars. To enable the ingredients to dissolve faster the cell can be warmed to 50 °C. Fats, and samples containing fats, are dissolved in a working medium of methanol and chloroform. A mixture of methanol, formamide and chloroform is used for samples containing both fat and sugar, e.g. milk powder.

Samples containing starch never dissolve completely and their water is therefore extracted from the fine particles in warm methanol. Generally, vegetable material is very difficult to dissolve or does not dissolve at all. In order to extract the water satisfactorily, the material must be chopped as finely as possible. Homogenizers or high-speed mixers are state-of-the-art technology. These mixers can be set up inside the titration cell through a special opening in the lid. The sample is added and the mixer switched on for a certain set time. Short titration times and sharp EPs are possible due to the very fine dispersion of the material. Heterogeneous samples such as chocolate bars, biscuits and noodles can be investigated using this technique. Extensive coverage of the titration methods for foodstuffs can be found in Isengard.[20,21]

9 INTERNATIONAL STANDARD PROCEDURES

The International Standard ISO 760 controls the volumetric water determination according to the KF method. Reagents, instruments and procedures are quoted. Sodium tartrate-2-hydrate or water is used for the titer determination of the reagents. Since ISO 760 was written in 1978 it does not take into consideration new methods, reagents and instrumentation. There is no international standard available for the coulometric KF titration.

The general KF titration is also quoted in the various national pharmacopeias. The *US Pharmacopeia* and the *Japanese Pharmacopeia* describe the volumetric and coulometric KF titration.[18,19] The *European Pharmacopeia* only covers the volumetric technique.[17]

In addition to the standards that quote the general methods for the KF titration, there are also national and international standards that describe the titration of individual matrixes. Such standards, given by the American Society for Testing and Materials (ASTM), are, for example, ASTM D 1533-79 ''Standard Test Method for Water in Insulation Liquids'', ASTM D 1744-83 ''Standard Test Method for Water in Liquid Petroleum'' and the DIN ISO 51869 ''Coulometric Water Determination in Gases''.

10 COMPARISON WITH LOSS-ON-DRYING

Loss-on-drying and the KF titration do not determine the same parameters of a sample, and therefore the analytical values found can not really be compared with one another. In the KF titration the water within a sample reacts chemically. During loss-on-drying the amount of material that escapes is measured. For some samples the water content found according to KF titration and loss-on-drying is identical. Such samples would be those that release water easily and do not alter on heating. Samples belonging to this category are the salts of carboxylic acids, some inorganic salts and mineral samples. Loss-on-drying is, however, still the preferred method quoted in the pharmaceutical and food industries, where samples can yield different results in loss-on-drying and the KF titration. Results found by KF titration are, in comparison,

mainly high because the water is fully extracted and reacted. Foodstuffs often yield higher results due to the degradation that often takes place under the conditions of loss-on-drying.

ABBREVIATIONS AND ACRONYMS

ASTM	American Society for Testing and Materials
EP	End-point
GLP	Good Laboratory Practice
IQ	Installation Qualification
KF	Karl Fischer
LIMS	Laboratory Information Management System
OQ	Operation Qualification
PQ	Performance Qualification

RELATED ARTICLES

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Sample Preparation for Food Analysis, General • Water Determination in Food

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction

Petroleum and Liquid Fossil Fuels Analysis **(Volume 8)**
Hydrocarbons Analysis: Introduction • Diesel Fuels Analysis • Fuels Analysis, Regulatory Specifications for • Full Range Crudes, Analytical Methodology of • Lubricant Base Oils: Analysis and Characterization of

Pharmaceuticals and Drugs **(Volume 8)**
Pharmaceuticals and Drugs: Introduction

Polymers and Rubbers **(Volume 8)**
Polymers and Rubbers: Introduction

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry • Traceability in Analytical Chemistry

REFERENCES

1. K. Fischer, 'Neues Verfahren zur massanalytischen Bestimmung des Wassergehaltes von Fluessigkeiten und festen Koerpern', *Angew. Chemie*, **48**, 394 (1935).
2. D.M. Smith, W.M.D. Bryant, J. Mitchell, 'Analytical Procedures Employing Karl Fischer Reagent I', *J. Am. Chem. Soc.*, **61**, 2407–2412 (1939).
3. G. Wernimot, F.J. Hopkinson, 'The Dead-stop End-point. As Applied to the Karl Fischer Method for Determining Moisture', *Ind. Eng. Chem. Anal. Ed.*, **15**, 272–274 (1943).
4. E.D. Peters, J.L. Jungnickel, 'Improvements in Karl Fischer Method for Determination of Water', *Anal. Chem.*, **27**, 450–453 (1955).
5. A.S. Meyer, C.M. Boyd, 'Determination of Water by Titration with Coulometric Generated Karl Fischer Reagent', *Anal. Chem.*, **31**, 215–219 (1959).
6. A. Cedergren, 'Reaction Rates Between Water and the Karl Fischer Reagents', *Talanta*, **21**, 265–271 (1974).
7. J.C. Verhoef, E. Barendrecht, 'Mechanism and Reaction Rate of the Karl Fischer Titration Reaction', *J. Electroanal. Chem. (Lausanne)*, **71**, 305–315 (1976).
8. E. Scholz, *Karl Fischer Titration*, Springer Verlag, Berlin, Heidelberg, New York, Tokyo, 1984.
9. G. Wünsch, A. Seubert, 'Stoechiometrie und Kinetik der Karl-Fischer-Reaktion in Methanol als Reaktionsmedium', *Fresenius' Z. Anal. Chem.*, **334**, 16–21 (1988).
10. 'HYDRANAL-Manual', RdH Laborchemikalien, Seelze, Germany, 1999. (http://www.netvertise.de/hydranal)
11. E. Scholz, 'Karl-Fischer-Reagenzien ohne Pyridin (3): Die Genauigkeit der Wasserbestimmung', *Fresenius' Z. Anal. Chem.*, **306**, 304–396 (1981).
12. W. Richter, W. Huerlimann, 'Diaphragmalose Messzelle fuer die Karl-Fischer-Titration', EP 0 390 727, CA 114(26):258756b, Metrohm Ltd, Herisau, Switzerland.
13. K. Schoeffski, 'Der lange Weg zur giftfreien Karl-Fischer-Titration', *GIT Lab. Fachz.*, **10/98**, 996–997 (1998).
14. G. Wuensch, K. Schoeffski, 'Die iodierende Nebenreaktion im Karl-Fischer-System', *Anal. Chim. Acta*, **239**, 283–290 (1990).
15. 'Application Bulletins', Metrohm Ltd, Herisau, Switzerland. (email: sales@metrohm.ch)
16. *HYDRANAL Guide PC*, RdH Laborchemikalien, Seelze, Germany, 1999.
17. 'Europaeisches Arzneibuch', 3. Ausgabe, Deutscher Apotheker Verlag, Stuttgart, 1997.
18. *United States Pharmacopeia*, 24th edition, The United States Pharmacopeia Convention Inc., Rockville, Maryland, USA, 1999.
19. *Japanese Pharmacopeia*, 13th edition, Hirokawa Shoten, Tokyo, 1996.
20. H.-D. Isengard, 'Bestimmung von Wasser in Lebensmitteln nach Karl Fischer', *Eur. Food Science, ZFL*, **42**, 1–6 (1991).
21. H.D. Isengard, 'Karl Fischer Titration in Boiling Methanol', *Fresenius' Z. Anal. Chem.*, **342**, 287–291 (1992).

Literature Searching Methodology

Dana L. Roth
California Institute of Technology, Pasadena, CA, USA

1 Introduction **1**
Further Reading **2**
References **2**
2 Cataloging and Indexing Specificity **2**
Further Reading **3**
References **3**
3 The Literature of Analytical Chemistry **3**
References **4**
4 Textbook and Official Analytical Methods **5**
Further Reading **5**
5 Encyclopedias, Reference Books, and Treatises **6**
Further Reading **6**
6 Review Journals, Research Monographs, AND Monographic Series **8**
6.1 Review Journals 8
6.2 Research Monographs 9
Further Reading **9**
6.3 Research Monographic Series 9
Further Reading **9**
7 Primary Literature (Journal Articles, Conference Papers, Technical Reports, Theses, and Patents) **10**
7.1 Journal Articles 10
References **11**
7.2 Conference Papers 11
References **11**
7.3 Technical Reports 11
Further Reading **11**
7.4 Theses 12
References **12**
7.5 Patents 12
Further Reading **12**
8 Abstracting and Indexing Services **12**
8.1 Chemical Abstracts 13
8.2 Web of Science 13
8.3 Analytical Abstracts 14
8.4 Ingenta Connect (Previously Uncover) 14
8.5 Institut de l'Information Scientifique et Technique 14
8.6 Google Scholar 14
9 Data Compilations **14**
10 Dictionaries, Catalogs, Directories, and so on **16**
Further Reading **16**
11 Internet Search Engines and Guides **16**
References **17**
11.1 Analytical Chemistry Guides 17
11.2 Internet Search Engines 17
12 Miscellaneous Tips **17**
Abbreviations and Acronyms **18**

Descriptions of analytical methods and instrumental techniques are found in an astonishing variety of resources. Because of this wide variety of sources and the extensive time period over which they have been published, developing an effective literature searching methodology is absolutely essential for students, research faculty, and practicing chemists. This article covers traditional methodologies and basic resources. In addition, recent widespread adoption of electronic journals and reference works, plus the virtually universal adoption of Google searching, are substantially transforming literature searching methodology.

1 INTRODUCTION

Descriptions of analytical methods and instrumental techniques are found in an astonishing variety of resources. Books, handbooks, journal articles, technical reports, theses, patents, standards, protocols, or encyclopedias may contain a specific method or describe a methodology appropriate for an analytical problem. Because of this wide variety of sources and the extensive time period over which they have been published, developing an effective literature searching methodology is absolutely essential for students, research faculty, and practicing chemists. Developing this methodology, fortunately, is largely a matter of guidance and extensive practice.

Guidance is often provided by colleagues, textbooks, and current awareness readings, but needs to be supplemented by interacting with knowledgeable librarians, making use of literature guides,[1–3] and extensive practice. Many technical libraries have former bench chemists on their staff and many academic libraries require an appropriate undergraduate degree for their subject-specialist librarians. Asking for help can be an excellent

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

technique both for getting started and for learning about new and unusual resources.

Practice is necessary because of the difficulty, often encountered, in answering the simple questions, ''where do I start?'' or ''where do I look next''? The *Encyclopedia of Analytical Chemistry* (EAC) is an obvious first answer to either of these questions, since this monumental work provides professional-level coverage of all aspects of analytical chemistry, from theory and instrumentation through all applications in all matrices. Availability of the EAC, however, does not obviate the need for a comprehensive understanding of the basic fundamentals of literature searching and regular practice using the full range of analytical chemistry resources.

The fundamentals of literature searching include a working knowledge of the cataloging and indexing practices in both library catalogs and indexing and abstracting services; a thorough familiarity with the structure and interrelationships of the literature of analytical chemistry; experience in using the various resources; a well-developed discipline of current awareness; and a knowledge of the Internet's strengths and weaknesses as a search tool. These factors are essential for the development of an effective literature searching methodology.

Recently, approaches to literature searching have undergone a dramatic transformation since the print publication of the EAC. The Internet and, specifically, Google and Google Scholar (GS) have been almost completely transformative. In fact, ''conservatively, ninety-five percent of all scholarly inquiries start at Google.''[4]

Given this situation, traditional sources of information and their electronic equivalents are increasingly vulnerable to being marginalized by freely available Internet resources. *Chemical Abstracts* (CA) and *SciFinder*/*SciFinder Scholar* continue to provide a level of excellence in the indexing of chemistry-related material.[5] In addition, efforts are currently under way in many libraries to develop ''federated searching'', which is a process whereby simultaneous searching of multiple electronic resources can be accomplished. This emerging feature is viewed as a portal, since the results can be merged, deduplicated, and sorted, in contrast to a web search engine, such as Google, which is limited by crawler technology and lack of access to deep web documents.[6]

FURTHER READING

Google Scholar, http://scholar.google.com/.

A. Grafton, *Future Reading: Digitization and its Discontents*, http://www.newyorker.com/reporting/2007/11/05/071105fa_fact_grafton?currentPage = all.

D.L. Roth, 'Literature Searching Methodology', *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, 2000.

REFERENCES

1. R.E. Maizell, *How to Find Chemical Information: A Guide for Practicing Chemists, Educators, and Students*, 3rd edition, John Wiley & Sons, New York, 1998.
2. G. Wiggins, *Chemical Information Sources*, McGraw-Hill, New York, 1991. The Analytical Chemistry Material is Updated at: http://www.indiana.edu/~cheminfo/400ac.html.
3. R.T. Bottle, *Information Sources in Chemistry*, 4th edition, Bowker-Saur, London, New York, 1993.
4. Google, http://www.google.com/.
5. Searching by Subject. *Chemical Information Sources Wiki*, Indiana University http://cheminfo.informatics.indiana.edu/cicc/cis/index.php/Searching_by_Subject.
6. Federated search, From *Wikipedia*, the free encyclopedia http://en.wikipedia.org/wiki/Federated_search.

2 CATALOGING AND INDEXING SPECIFICITY

One of the basic fundamentals of developing a methodology for searching the literature of any subject is an understanding of the concept of ''cataloging or indexing specificity''. In library card catalogs and their successor OPACs (Online Public Access Catalogs), books are cataloged as specifically as they are written.

An excellent example is Kuwana's three-volume set on *Physical Methods in Modern Chemical Analysis*,[1] which has chapters on transform techniques, electrochemical characterizations, high-performance liquid chromatography, X-ray diffraction methods, gas chromatography, and flame and plasma emission analysis. It is only retrievable (by subject) in library catalogs by searching the keywords in the title or ''chemistry, analytic'' as a subject heading. The inability to retrieve this book by searching on chapter-title words or subject terms related to the individual analytical techniques is obviously somewhat confusing. This is especially true when there are monographs on each of these techniques that are easily located by searching on keywords and subject terms in a library catalog. A further example of the specificity problem is exemplified by books dealing with ''chronoamperometry'', which may also be found cataloged or indexed only under ''amperometry'', ''voltammetry'', or ''analysis, electrochemistry''.

The practice of indexing specificity is carried to its extreme, however, in the cataloging of the encyclopedia,

treatise, or reference set. This concept explains the single subject headings: ''chemistry, analytic – encyclopedias'' for *Encyclopedia of Analytical Science*, ''chemistry, analytic'' for *Treatise on Analytical Chemistry*, and ''chemistry – manipulation'' for *Physical Methods of Chemistry*, in library catalogs. All three of these titles are comprehensive multivolume works that cover virtually all aspects of their corresponding subject fields. As an aid to searching library catalogs and OPACs at a more general level, the Library of Congress publishes *Library of Congress Subject Headings*, which is in its 30th edition.[2] This book provides broader terms (BT), narrower terms (NT), and related terms (RT) for a wide variety of subject concepts.

In Chemical Abstracts Service (CAS),[3] which indexes journal articles, patents, and so on, the specificity problem is not as severe, since many multivolume works are indexed as separate volumes or at the chapter level. For example, each of the chapters in volumes 1 and 2 of *Physical Methods in Modern Analytical Chemistry* is indexed separately. Coverage issues, however, complicate the situation. SciFinder/SciFinder Scholar[1] has replaced CAS Online, and combines CAplus, CAS Registry, CASREACT, CHEMCATS, and CHEMLIST.

While SciFinder/SciFinder Scholar can be searched with title and CAS abstract keywords, CAS index terms are selected from ''new'' information reported in the document. This is in contrast to the expanded keyword searching (title and author abstract keywords, author supplied keywords, and significant keywords from the titles of an article's references) that is possible in the Web of Science, the web version of the Science Citation Index (SCI),[4] another journal-article database. This is obviously a subtle distinction, and is pointed out to emphasize the need to search multiple databases in a comprehensive literature retrieval.

In either SciFinder/SciFinder Scholar or print, ''The fact that a substance or concept, reported in a document, is not in itself new does not mean that it will not be indexed, only that the indexing will focus on the novel aspects. However, when long lists of known substances and accompanying previously published experimental data are reported, but the author's main interest is in some property or source of the substances, the substances are not indexed specifically but receive General (Subject Index) entries only.''[5]

In addition, ''CA chooses the most specific term possible as an index heading. Thus, separate entries would be found at ''fluorometers'' and ''phosphorimeters'', when an original document discusses both of these processes, rather than a single entry at ''spectrometers''. However, a review article on spectrometers (that includes examples of fluorometers and phosphorimeters) may only be assigned the more general index term. As an aid to searching at a more general level, CA provides ''Hierarchies of General Subject Headings''.[6]

Today the importance of cataloging and indexing specificity is underscored by what is widely known as *search engine chaos*. Despite the ability of Google/Google Scholar to retrieve a few pertinent articles or web sites, library catalogs and professionally indexed databases (e.g. CA) continue to be essential for comprehensive research needs.

FURTHER READING

Chemical Abstracts Service, *SciFinder/SciFinder Scholar*, http://www.cas.org/expertise/cascontent/ataglance/index.html, 2008.

REFERENCES

1. T. Kuwana, ed, *Physical Methods in Modern Chemical Analysis*, Academic Press, New York, Vols. 1–3, 1978–1983.
2. Library of Congress, *Library of Congress Subject Headings*, 30th edition, United States Government Printing Office, Washington, DC, 2007.
3. Chemical Abstracts Service, *CAS Database*, http://www.cas.org/expertise/cascontent/index.html, 2008.
4. Thomson, *Web of Science*, http://scientific.thomson.com/products/wos/, 2008.
5. *Chemical Abstracts Index Guide Appendix II*, 100I, 1992–1996.
6. *Chemical Abstracts Index Guide Appendix I*, 1992–1996.

3 THE LITERATURE OF ANALYTICAL CHEMISTRY

One of the obvious difficulties in approaching the literature of analytical chemistry is its enormity. The 100-fold growth in the annual output of CA from about 10 K abstracts in 1907 to over 1 M abstracts in 2007 is an excellent indicator of the need for a carefully developed search methodology.

Before approaching the specialized literature of analytical chemistry, however, it is essential to use and become familiar with basic reference and general chemical resources (the importance of developing this skill cannot be overemphasized). The vast majority of questions in a public library, for example, can be answered by referring to an encyclopedia, for example, *Encyclopedia Britannica*[1]; a dictionary, for example, *Webster's Third New International Dictionary*[2]; an almanac, for example, *Information Please Almanac*[3];

or an atlas/gazetteer, for example, *The Times Atlas of the World*[4] or *Merriam-Webster's Geographical Dictionary*.[5]

Many basic chemistry questions can be answered by an encyclopedia, for example, *Encyclopedia of Physical Science and Technology*[6] or *Encyclopedia of Chemical Technology*[7]; a dictionary, for example, *The Condensed Chemical Dictionary*[8]; a data book, for example, *Handbook of Chemistry and Physics*[9] or *Merck Index*[10]; or a catalog/structure index, for example, *Aldrich Catalog*,[11] *Aldrich Structure Index*,[12] or *Dictionary of Organic Chemistry*.[13]

Extending this analogy to analytical chemistry requires familiarity with encyclopedias, for example, *Encyclopedia of Analytical Science*[14] or the *Instrumentation and Applications* section of *EAC* (this publication); dictionaries, for example, *Dictionary of Spectroscopy*[15] or *Dictionary of Chromatography*[16]; data books, for example, *Critical Stability Constants*[17] or *Tables of Standard Electrode Potentials*[18]; and compilations of analytical methods and instrumental techniques, for example, *Treatise on Analytical Chemistry*.[19]

Some information formerly found in general encyclopedias, yearbooks, almanacs, and dictionaries can be easily found through Google or Yahoo.[20] This partially obviates the need to be thoroughly familiar with the basic reference and general chemical resources. Many of the specialized chemical references continue to retain their importance, with many appearing in new editions and as associated electronic versions. Electronic versions are generally sold on a subscription basis that includes regular updating.

Wikipedia[21] is a unique attempt to provide encyclopedic information. It is reputed to be the ''biggest multilingual free-content encyclopedia on the Internet, with over 7 million articles in over 200 languages, and still growing.'' *Wikipedia* is written collaboratively by volunteers from around the world. *Citizendium*[22] is an attempt to offer an ''edited'' counterpart to *Wikipedia*. *Knovel*[23] is a unique service that aggregates reference works, including handbooks and databases, in 19 subject areas, including chemistry. There are currently 18 ''analytical chemistry'' titles, which are freely searchable; however, the electronic content is only available to subscribers.

There are also a variety of electronic ''news'' magazines and web sites, such as *SpectroscopyNow, SeparationsNow,*[24] *News Analytic*,[25] the ''analytical highlights'' section of Chemistry & Industry,[26] the Royal Society of Chemistry (RSC) *Analytical Gateway*,[27] the American Chemical Society (ACS) Division of Analytical Chemistry's Analytical Sciences web site,[28] and the *ChemLin Virtual Library*.[29]

Progressing to the full range of analytical chemistry literature requires familiarity with textbooks and sources of official analytical methods or protocols, encyclopedias, reference books and treatises, review articles and specialist monographs, journal articles, conference papers, technical reports, theses and patents, abstracting and indexing resources, comprehensive data books, and dictionaries, catalogs, and directories.

Thus, the crux of developing an efficient literature searching methodology is a thorough familiarity with the wide variety of information resources, through consultation and extensive practice, so that quickly locating an appropriate resource for a particular problem becomes routine.

Perhaps the most important concept the reader must learn is when to begin searching CA.[30] Because of this resource's comprehensive coverage, it is often difficult to determine when to progress beyond what can quickly be found in library catalogs and easily available reference sources and begin a more formal search of the literature. For this reason, it is absolutely essential to become familiar with the wealth of secondary literature resources.

The description of the analytical chemistry literature that follows progresses from the secondary literature (textbook and official analytical methods; encyclopedias, reference books and treatises; review journals, research monographs, and monographic series) to the primary literature (journal articles, conference papers, technical reports, theses and patents) and their indexing and abstracting services. This is followed by the tertiary and miscellaneous literature (data compilations; dictionaries, catalogs, directories; Internet search engines and guides; and miscellaneous tips). This approach allows readers to orient themselves quickly and makes the next step, for any given problem, much more obvious.

REFERENCES

1. *Encyclopaedia Britannica*, 2007 edition, Encyclopaedia Britannica, Chicago, 2007, http://www.eb.com/.
2. *Webster's Third New International Dictionary of the English Language* (unabridged), G. & C. Merriam, Springfield, 1981 and *12 000 Words: A Supplement to Webster's Third New International Dictionary*, G. & C. Merriam, Springfield, 1986.
3. Infoplease, http://www.infoplease.com/, 2008.
4. *The Times Atlas of the World*, 9th comprehensive edition, Times Books, New York, 1994.
5. *Merriam–Webster's Geographical Dictionary*, 3rd edition, Merriam–Webster, Springfield, 1997.
6. R.A. Meyers, *Encyclopedia of Physical Science and Technology*, 3rd edition, Elsevier Academic Press, Vols. 1–18, 2001, http://www.sciencedirect.com/science/referenceworks/0122274105 (listing of contents with links to related journal articles).

7. A. Seidel, *Kirk-Othmer Encyclopedia of Chemical Technology*, 5th edition, John Wiley & Sons, Vols. 1–27, 2004, http://www.mrw.interscience.wiley.com/emrw/0471238961/home/.
8. R.J. Lewis Sr, *Hawley's Condensed Chemical Dictionary*, 14th edition, John Wiley & Sons, 2001, http://mrw.interscience.wiley.com/emrw/9780470114735/home/.
9. D.R. Lide, *CRC Handbook of Chemistry and Physics*, 88th edition, CRC Press, 2007–2008, http://www.hbcpnetbase.com/ (freely searchable but content only available to subscribers).
10. *Merck Index*, 14th edition, Merck Manuals, Whitehouse Station, 2006 (Includes a companion text searchable CD-ROM, that also includes over 1000 monographs from the 12th and 13th editions that are not available in the print copies of the 14th edition).
11. *Aldrich Catalog*, Sigma-Aldrich, Milwaukee, 2007–2008, http://www.sigmaaldrich.com/Brands/Aldrich.html (electronic version includes access to both IR and NMR spectra and MSDS for the chemicals).
12. *Aldrich Structure Index*, Aldrich Chemical, Milwaukee, 1996–1997.
13. Combined Chemical Dictionary (CCD) Chapman & Hall/CRC (includes *Dictionary of Organic Compounds, Dictionary of Natural Products, Dictionary of Inorganic and Organometallic Compounds, Dictionary of Drugs, Dictionary of Analytical Reagents, Dictionary of Carbohydrates, and Dictionary of Marine Natural Products*.) http://www.chemnetbase.com/.
14. P. Worsfold, A. Townshend, C. Poole, *Encyclopedia of Analytical Science*, 2nd edition, Academic Press, Vols. 1–10, 2004, http://www.sciencedirect.com/science/referenceworks/9780123693976.
15. *Dictionary of Spectroscopy*, 2nd edition, John Wiley & Sons, New York, 1982.
16. H.-P. Angele, *Dictionary of Chromatography: English, German, French, Russian*, Huthig, New York, 1984.
17. A.E. Martell & R.M. Smith, *Critical Stability Constants*, Plenum Press, New York, Vols. 1–6, 1974–1989 and NIST, *Critically Selected Stability Constants of Metal Complexes* (computer file), National Institute of Standards and Technology, Gaithersburg, 1997.
18. G. Milazzo, S. Caroli, *Tables of Standard Electrode Potentials*, John Wiley & Sons, New York, 1978.
19. I.M. Kolthoff & P.J. Elving, *Treatise on Analytical Chemistry*, 2nd edition, 1959–1977, Interscience, New York, 1978–1986.
20. Google, http://www.google.com/; (a) Yahoo, http://www.yahoo.com/.
21. *Wikipedia,* http://wikipedia.org.
22. Citizendium, *The Citizen's Compendium*, http://en.citizendium.org/wiki/Main_Page, 2008.
23. Knovel Corporation, *Knovel Library*, New York, http://www.knovel.com/knovel2/library/default.jsp, 2008.
24. SpectoscopyNow, John Wiley & Sons, Chichester, http://www.spectroscopynow.com/coi/cda/home.cda; (b) SeparationsNow, John Wiley & Sons, Chichester, http://www.separationsnow.com/coi/cda/home.cda, 2008.
25. Analytic News, (English & German editions), Dr. Beyer Internet-Beratung, Ober-Ramstadt, http://www.analytic-news.com/, 2008.
26. Chemind, *Chemistry and Industry*, Ten Alps Publishing/SCI, London, http://www.chemind.org/CI/index.jsp, 2008.
27. Royal Society of Chemistry, *Royal Society of Chemistry online gateway: Analytical*, http://rsc.org/Gateway/Subject/Analytical/, 2008.
28. American Chemical Society/Analytical Sciences Digital Library (ASDL), *Analytical Sciences*, http://www.analyticalsciences.org/, 2008.
29. Chemlin Virtual Chemistry Library, ChemLin Digitalverlag GmbH, Verden, http://www.chemlin.net/, 2008.
30. Chemical Abstracts Service, *Chemical Abstracts*, Columbus, 1907, +http://www.cas.org/.

4 TEXTBOOK AND OFFICIAL ANALYTICAL METHODS

Analysis and testing are frequent operations in chemical laboratories. For many routine purposes, textbook methods may be adequate; however, in certain circumstances, an official or standard method (SM) is required. World Wide Web (WWW) sites are provided to allow readers to check for up-to-date information from these resources. Books and reference works that fit into this category are listed here. Current and archival official methods are now both generally available on the WWW.

FURTHER READING

1. D.C. Harris, ed, *Quantitative Chemical Analysis*, 7th edition, Freeman, 2006.
   This is the definitive modern textbook of quantitative analysis, which includes a final chapter of experiments. http://bcs.whfreeman.com/qca7e/
2. D.A. Skoog, ed, *Fundamentals of Analytical Chemistry*, 8th edition, Saunders, 2004.
   This is an introductory textbook designed primarily for a one or two semester course for chemistry majors, with selected methods of analysis in the final chapter.

3. G.H. Jeffery, J. Bassett, J. Mendham, R.C. Denney, eds, *Vogel's Textbook of Quantitative Chemical Analysis*, 5th edition, John Wiley & Sons, New York, 1989.
Vogel is intended as a comprehensive textbook for both students and practicing analytical chemists and contains an extensive listing of specific procedures, albeit without references to the original literature.
4. *Standard Methods of Chemical Analysis*, 6th edition, Van Nostrand, New York, 1962–1966.
This final edition, while dated, is in five parts, largely devoted to specific applications. Analysis of the chemical elements is covered in volume 1, classes of substances in volume 2, and instrumental methods in volume 3.
5. *Standard Methods for the Examination of Water and Wastewater*, 21st edition, American Public Health Association, 2005.
More than 340 methods are presented, each with step-by-step procedures for precise analysis of water and the chemical constituents of wastewater, sanitary quality, and physical and biological characteristics. http://www.standardmethods.org/store/l
6. *Annual Book of ASTM Standards*, American Society for Testing and Materials, 2007.
The ASTM standards continue publication in print, but have added electronic access to individual volumes, an online basic subscription to active standards, and an online plus subscription which also shows changes from a previous edition, previous editions, and withdrawn standards. http://www.astm.org/cgi-bin/SoftCart.exe/STORE/standardsearch.shtml?E + mystore
7. *Official Methods of Analysis of Association of Official Analytical Chemists (AOAC) International*, 18th edition, Revision 2, Association of Official Analytical Chemists, 2007.
Official methods have been developed by federal, state, local, and industrial scientists for products related both to agriculture and public health and for enforcing laws and regulations. http://eoma.aoac.org/
8. R.-K. Smith, ed, *Guide to Environmental Analytical Methods*, 5th edition, Genium Publishing Corporation, 2003.
This includes the latest and most important methods required by several major environmental statutes. A glossary of terms and abbreviations, as well as an EPA regulatory contact directory, is included.
9. *US EPA Environmental Test Methods and Guidelines*. Provides links to the *EPA Test Method Index*, revised edition, 2003 and *Sources for EPA Test Methods*, a listing of sources cited in the EPA Test Method Index. http://www.epa.gov/epahome/Standards.html.
10. *Compilation of EPA's Sampling and Analysis Methods*, 2nd edition, Lewis, Boca Raton, 1996.
This contains over 1700 analyte/method summaries with a double table of contents that lists chemical names/CAS registry numbers in alphabetical order and CAS registry numbers/names in numerical order. http://www.environetbase.com/ejournals/books/book_summary/summary.asp?id = 80
11. *Current Protocols in Field Analytical Chemistry (CPFAC)*, John Wiley & Sons, New York, 1998.
CPFAC covers four main analyte types: volatile organic compounds, semivolatile/nonvolatile organic compounds, metals, and radionuclides. It is available in either a loose-leaf or CD-ROM format and is updated quarterly.

5 ENCYCLOPEDIAS, REFERENCE BOOKS, AND TREATISES

For more specialized problems, when textbook or SM are either unavailable or inappropriate, encyclopedias, reference books, and treatises provide comprehensive coverage of analytical methods and instrumental techniques. The importance of becoming familiar with these publications cannot be overemphasized. These publications often refer to specialized monographs that may also be found in library catalogs by searching on title keywords or appropriate subject headings. Many encyclopedias and reference books are now also available, as annual subscriptions, on the WWW.

FURTHER READING

1. R.A. Meyers, ed, *Encyclopedia of Environmental Analysis and Remediation*, John Wiley & Sons, New York, 1998.
Composed of over 280 articles in eight volumes, it provides the background information necessary to prepare and evaluate environmental compliance documentation as well as laboratory protocols for pollutant analysis.
2. F.A. Settle, ed, *Handbook of Instrumental Techniques for Analytical Chemistry*, Prentice Hall, Upper Saddle River, 1997. Reviewed in *J. Am. Chem. Soc.*, **120**(26), 6633, (1998).
The 51 chapters are grouped in the following sections:

- Introduction (sample preparation, laboratory automation, etc.)

- Separation methods/Qualitative optical spectroscopicmethods
- Quantitative optical spectroscopic methods/ Mass spectrometry
- Electroanalytical techniques/Microscopic and surface analysis
- Polymer analysis
- CD-ROM technique selection software program.

3. A. Townshend, ed, *Encyclopedia of Analytical Science*, 2nd edition, Elsevier Academic Press, Vol. 1–10, 2004.

The articles cover three broad areas:

- Techniques (e.g. spot tests, mass spectrometry, chemometrics)
- Applications (e.g. asbestos, pharmaceuticals, pigments, forensics)
- Analytes (e.g. aluminum, cholesterol, nucleic acids).

http://www.sciencedirect.com/science/reference works/9780123693976

4. P. Patniak, ed, *Analytical Chemistry Handbook*, 2nd edition, McGraw-Hill, 2004.

There are sections covering, amongst others, the following areas:

- Sample preparation
- Separation, gravimetric, volumetric, and chromatographic methods
- Instrumental methods
- Organic, water, and geological/inorganic analysis
- Statistics.

5. *Ullmann's Encyclopedia of Industrial Chemistry*, 7th edition, Wiley-VCH, 2007, http://www.wiley-vch.de/contents/ullmann/index.html.

The printed version comprises 28 alphabetically arranged volumes, 8 volumes on general aspects such as chemical engineering, industrial safety, and analytical methods, and a cumulative index volume with more than 80 000 carefully selected keywords. It also contains a German–English dictionary of technical terms. An alphabetical listing of the contents is available on the web site.

- Vol. B5 *Analytical Methods – I*
- Vol. B6 *Analytical Methods – II and Process Control Engineering*.

The analytical methods sections discuss commonly used procedures in both process engineering and research activities.

http://mrw.interscience.wiley.com/emrw/97835273-06732/home/

6. *Kirk-Othmer Encyclopedia of Chemical Technology*, 5th edition, John Wiley & Sons, Vol. 1–27, 2004.

This encyclopedia focuses on the significant aspects of chemical substances, industrial processes, unit operations, and fundamental concepts. It includes charts, graphs, and tables and provides CAS registry numbers for substances.

http://www.mrw.interscience.wiley.com/emrw/0471-238961/home

7. *Encyclopedia of Analytical Instrumentation*, *Chemistry Hypermedia Project*, Virginia Polytechnic Institute and State University, Blacksburg, http://www.chem.vt.edu/chem-ed/ac-meths.html.
8. J. Cazes, ed, *Ewing's Analytical Instrumentation Handbook*, 3rd edition, CRC, New York, 2004.

This is a compilation of practical descriptions of the most frequently used instrumental analytical methods, with information on sensitivity and applicability, and advice on how to optimize the measurements. Contents include ''chapters'' on all the standard instrumental techniques as well as the following: laboratory use of computers; continuous-flow analyzers; photoacoustic spectroscopy; techniques of chiroptical spectroscopy; and capillary electrophoresis.

9. I.M. Kolthoff, P.J. Elving, eds, *Treatise on Analytical Chemistry*, Wiley-Interscience, New York, 1959–1989.

- *Part I – Theory and Practice*, Vols. 1–12, 1959–1976.
- *Part II – Systematic Analytical Chemistry of the Elements*, Vols. 1–10, 1961–1975.
- *Part II – Organic Analysis*, Vols. 11–16, 1965–1980.
- *Part III – Analytical Chemistry in Industry*, Vols. 1–4, 1967–1977.
- *Part I (2nd edition) – Theory and Practice*

–*Analytical Chemistry*, 1978.
–*Analytical Chemistry Methodology*, 1978.
–*General Chemistry and Analytical Chemistry*, 1978.
–*Solution Equilibria and Chemistry*, 1978–1979.
–*Principles of Instrumentation for Analysis*, 1984.
–*Sample Preparation*, 1982.
–*Separation Techniques*, 1982.
–*Optical Methods of Analysis*, 1981–1986.
–*Magnetic Field and Related Methods of Analysis*, 1983–1989.
–*Thermal Methods*, 1983–1993.
–*Nuclear Activation and Radioisotopic Methods of Analysis*, 1986.

10. A. Weissberger, ed, *Techniques of Chemistry*, Wiley-Interscience, New York, Vols. 1–20, 1971–1990.

- Vol. 21 *Solubility Behavior of Organic Compounds*, 1990

- Vol. 22 *Molecular Design of Electrode Surfaces*, 1992
- Vol. 23 *Laser Techniques in Chemistry*, 1995.

11. A. Weissberger, B.W. Rossiter, eds, *Physical Methods of Chemistry*, 2nd edition, Wiley-Interscience, New York, 1986–1993.

This is a sub-series of *Techniques of Chemistry*.

- Vol. 1 *Scientific Instruments . . . Computers*, 1986
- Vol. 2 *Electrochemical Methods*, 1986
- Vol. 3 *Chemical Composition and Molecular Structure*, 1987
- Vol. 4 *Microscopy*, 1991
- Vol. 5 *Crystal and Amorphous Solids*, 1990
- Vol. 6 *Thermodynamic Properties*, 1992
- Vol. 7 *Elastic and Mechanical Properties*, 1991
- Vol. 8 *Electronic and Optical Properties*, 1993
- Vol. 9 *Surfaces and Interfaces*, 1992–1993
- Vol. 10 *Index*.

12. S. Srinivasan, ed, *Comprehensive Treatise of Electrochemistry*, Plenum Press, New York, Vols. 1–10, 1980–1985.

- Vol. 1 *The Double Layer*
- Vol. 2 *Electrochemical Processing*
- Vol. 3 *Electrochemical Energy Conversion and Storage*
- Vol. 4 *Electrochemical Materials Science*
- Vol. 5 *Thermodynamic and Transport Properties of Aqueous and Molten Electrolytes*
- Vol. 6 *Electrodics: Transport*
- Vol. 7 *Kinetics and Mechanisms of Electrode Processes*
- Vol. 8 *Experimental Methods of Electrochemistry*
- Vol. 9 *Electrodics: Experimental Techniques*
- Vol. 10 *Bioelectrochemistry*.

6 REVIEW JOURNALS, RESEARCH MONOGRAPHS, AND MONOGRAPHIC SERIES

The primary role of both review articles (including plenary conference lectures) and chapters in research monographs is to summarize new research on a particular topic that updates information found in textbooks, encyclopedias, or treatises. The review is expected both to summarize and evaluate this information and to provide a bibliography of the original sources.

Locating review articles is fairly straightforward, because both CA and the Web of Science use ''review'' as a subject term. In the web of Science, a review article is defined as either having 100 or more references, or appearing in a review publication or a review section of a journal, or if review or overview appears in the title, or the abstract states that it is a review or survey. In CA, trained editorial staff categorize review articles on the basis of an assessment of the content. A review may be an article, a bibliography, a discussion, a portion of a book, and so on, and it may be critical or noncritical.

Locating specialized monographs, chapters in research monographs and plenary conference lectures, however, can be difficult owing to the absence of suitable descriptions of monographic works and conference proceedings in library catalogs. As mentioned previously, multivolume works may receive only a single subject entry in library catalogs, books in series may only be cataloged under the series title (without title or subject entries for individual volumes), and plenary conference lectures are only cataloged under the conference title. Finding these publications requires imaginative searching of library catalogs, followed by a search of CA and limiting the search to ''book'' as a document type or ''review'' as a keyword. Google is an alternative route for locating such material.

Maintaining a current awareness of newly published review material requires familiarity with a wide variety of bibliographic tools. These include the following electronic resources:

- Amazon http://www.amazon.com/
- Baker-Taylor's monthly listing of new academic books http://www.baker-taylor.com/Academia/Academia.html
- Index to Chemical Education Resources (textbooks) from the *Journal of Chemical Education* http://www.umsl.edu/~chemist/books/texts.html
- Comprehensive library catalogs such as the University of California's Melvyl http://www.melvyl.ucop.edu/ and Linda Hall's Leonardo http://opac.lhl.lib.mo.us/.

6.1 Review Journals

1. American Chemical Society, *Analytical Chemistry*, http://pubs.acs.org/journals/ancham/index.html. Fundamental Reviews and Application Reviews are published in alternate years as the June 15 issue. Starting with the first issue of 2006, Analytical Chemistry's News, Features, and Departments (previously called the *A-Pages*) are integrated with the research papers into one page-numbering scheme.
2. *European Conference on Analytical Chemistry, Euroanalysis 13*, 2004, http://www.euroanalysis13.com/.
3. *Trends in Analytical Chemistry* (*TrAC*), Elsevier, http://www.sciencedirect.com/science/journal/01659936.
4. American Chemical Society, *Chemical Reviews*, http://pubs.acs.org/journals/chreay/index.html.

5. The Royal Society of Chemistry, *Chemical Society Reviews*, http://www.rsc.org/Publishing/Journals/CS/Index.asp.
6. Russian Academy of Sciences/Turpion, *Russian Chemical Reviews*, http://www.uspkhim.ru/rcr.html, http://www.turpion.org/php/homes/pa.phtml?jrnid = rc.
7. Informa, *Critical Reviews in Analytical Chemistry*, http://www.informaworld.com/smpp/title~content = *t*713400837.
8. Informa, *Current Opinion in Chemical Biology*, Each issue deals with one or more special topics. Analytical Techniques/Mechanisms appeared in **11**(5), October 2007, http://www.informaworld.com/smpp/title~content = t713400837.
9. Bentham Science Publishers, *Current Organic Chemistry*, http://www.ingentaconnect.com/content/ben/coc.

6.2 Research Monographs

Lists of specialized monographs are generally included in the appendices of analytical chemistry textbooks and reference books. Recent publications are listed in the Book Buyer's Guide supplement to the *Journal of Chemical Education*, and the *Subject Guide to Books in Print* (published annually by R.R. Bowker). Subject searching for books with the subject ''analytical chemistry'' is also available on the WWW at (http://www.amazon.com). Amazon.com also provides a list of ''50 best sellers'' in a wide variety of categories. The titles of the top five analytical chemistry books, in terms of sales, as of December 2007 are:

FURTHER READING

1. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd edition, Springer, 2006.
2. A.J. Bard, L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, 2nd edition, John Wiley & Sons, 2000.
3. D.A. Skoog, F.J. Holler, S.R. Crouch, *Principles of Instrumental Analysis*, 6th edition, Thomson Brooks/Cole, 2007.
4. R. de Levie, *Advanced Excel for Scientific Data Analysis*, Oxford University Press, 2004.
5. J.M. Hollas, *Modern Spectroscopy*, 4th edition, John Wiley & Sons, 2004.

6.3 Research Monographic Series

In addition to individually published research monographs, there are a number of continuing research monographic series that routinely publish books on analytical topics. As mentioned previously, libraries may only catalog these books in series under their series title and not provide keywords or subject headings for each individual volume. Thus, searching under the series title is very strongly recommended, if the individual volume is not found.

FURTHER READING

1. *Analytical Spectroscopy Library*, Elsevier, http://www.sciencedirect.com/science/bookseries/09264345.
   - Vol. 9 *Flow Analysis with Atomic Spectrometric Detectors*
   - Vol. 10 *Separation, Preconcentration and Spectrophotometry in Inorganic Analysis*
2. *Chemical Analysis*, John Wiley & Sons.
   - Vol. 165 *Quadrupole Ion Trap Mass Spectrometry*, 2005
   - Vol. 167 *Introduction to Soil Chemistry: Analysis and Instrumentation*, 2005
   - Vol. 169 *Identification of Microorganisms by Mass Spectrometry*, 2006
   - Vol. 171 *Fourier Transform Infrared Spectrometry*, 2007.
   - Vol. 172 *New Frontiers in Ultrasensitive Bioanalysis: Advanced Analytical Chemistry Applications in Nanobiotechnology, Single Molecule Detection, and Single Cell Analysis*, 2007.
3. *Comprehensive Analytical Chemistry*, Elsevier, http://www.sciencedirect.com/science/handbooks/0166526X.
   - Vol. 46 *Proteomics and Peptidomics: New Technology Platforms Elucidating Biology*, 2005.
   - Vol. 47 *Modern Instrumental Analysis*, 2006.
   - Vol. 48 *Passive Sampling Techniques in Environmental Monitoring*, 2007.
   - Vol. 49 *Electrochemical Sensor Analysis*, 2007.
   - Vol. 50 *Analysis, Fate and Removal of Pharmaceuticals in the Water Cycle*, 2007.
4. *Journal Of Chromatography Library*, Elsevier, http://www.elsevier.com/wps/find/bookdescription.cws_home/BS_JCL/description#description.
   - Vol. 67 *Monolithic Materials; Preparation, Properties and Applications*, 2003
   - Vol. 68 *Emerging Technologies in Protein and Genomic Material Analysis*, 2003
   - *Chromatography; Fundamentals and Applications of Chromatography and Related Differential*

Migration Methods – Part A: Fundamentals and Techniques, Part B: Applications, 2004.
- *Quantitation of Amino Acids and Amines by Chromatography; Methods and Protocols*, 2005.
- *Liquid Chromatography of Natural Pigments and Synthetic Dyes*, 2006.

5. *Methods of Biochemical Analysis*, John Wiley & Sons.
   - Vol. 46 *Evaluation of Enzyme Inhibitors in Drug Discovery : A Guide for Medicinal Chemists and Pharmacologists*, 2005.
   - Vol. 47 *Green Fluorescent Protein: Properties, Applications, and Protocols*, 2006.
   - Vol. 48 *Biochemical Thermodynamics: Applications of Mathematica*, 2006.
   - Vol. 49 *Microbial Proteomics: Functional Biology of Whole Organisms*, 2006.
   - Vol. 50 *Physical Chemistry for the Biological Sciences*, 2007.
6. *Methods in Enzymology*, Academic Press, http://www.sciencedirect.com/science/bookseries/00766879.
   - Vol. 431 *Translation Initiation: Cell Biology, High-Throughput Methods, and Chemical-Based Approaches*, 2007.
   - Vol. 432 *Lipidomics and Bioactive Lipids: Mass-Spectrometry-Based Lipid Analysis*, 2007.
   - Vol. 433 *Lipidomics and Bioactive Lipids: Specialized Analytical Methods and Lipids in Disease*, 2007.
   - Vol. 434 *Lipidomics and Bioactive Lipids: Lipids and Cell Signaling*, 2007.
   - Vol. 435 *Oxygen Biology and Hypoxia*, 2007.
7. *Techniques and Instrumentation in Analytical Chemistry*, Elsevier, http://www.sciencedirect.com/science/bookseries/01679244.
   - Vol. 23 *Molecularly Imprinted Polymers – Man-made Mimics of Antibodies and their Applications in Analytical Chemistry*, 2000.
   - Vol. 24 *Acceleration and Automation of Solid Sample Treatment*, 2002.
   - Vol. 25 *Analytical Pyrolysis of Synthetic Organic Polymers*, 2005.
   - Vol. 26 *Analytical Applications of Ultrasound*, 2007.
   - Vol. 27 *UV–Visible Spectrophotometry of Water and Wastewater*, 2007.
8. *Techniques in Analytical Chemistry*, John Wiley & Sons.
   - *Introduction to Bioanalytical Sensors*, 1998.
   - *Introduction to Fluorescence Spectroscopy*, 1999.
   - *Chemical Measurements in Biological Systems*, 2000.

7 PRIMARY LITERATURE (JOURNAL ARTICLES, CONFERENCE PAPERS, TECHNICAL REPORTS, THESES, AND PATENTS)

These seemingly disparate types of publication are included in the broadest definition of the primary literature. The formal primary literature consists of peer-reviewed journal articles, which can be reliably expected to present data and procedures upon which further research can be conducted. The informal primary literature includes conference papers, technical reports, theses and patents, which are not peer-reviewed and, in many cases, should be considered the raw material for the formal primary literature. Searching for information in these resources is greatly facilitated by the indexing and abstracting tools described in the next section.

7.1 Journal Articles

Analytical chemistry research journals, as defined by Institute for Scientific Information (ISI), are listed below with their journal ranking data and sorted by ''impact factor'' (Table 1). Analytical articles, however, appear in virtually all chemistry journals.

A consistent habit of reading the current analytical literature is greatly facilitated by perusing the high-impact titles (*Analytical Chemistry, TrAC*, etc.) and using current awareness services such as Current Contents http://www.isinet.com/prodserv/cc/cchp.html (which is available both in print and on line) and CA Selects http://www.cas.org/PRINTED/caselects.html. In its ''Highlights Section'', *Chemistry and Industry* http://ci.mond.org/ reviews the most significant new papers published in the scientific journals (new materials and analytical chemistry in the odd-numbered issues).

Journal articles have undergone a major transformation as they are now widely available on the WWW (albeit generally requiring a subscription). Open access journals[4] are growing in number, although the concept

Table 1 General analytical chemistry journals ranked by impact factor[a]

1. Analytical Chemistry	5.646
2. Trends in Analytical Chemistry	5.068
3. Biosensors & Bioelectronics	4.132
4. Electrophoresis	4.101
5. Journal of Analytical Atomic Spectrometry	3.630
6. Journal of Chrotomatography – A	3.554
7. Journal of American Society for Mass Spectrometry	3.307
8. ANALYST	3.198
9. Analytical Biochemistry	2.948
10. Analytica Chimica Acta	2.894

[a] ISI Web of Knowledge. Journal Citation Reports. JCR Science Edition, 2006.

of author payments is not widely popular. A wide variety of updating options is now available on each journal's web site, along with articles that have been accepted but not assigned to a specific journal issue. An extensive listing of ''analytical chemistry'' journals is given at CiteSeer.[5]

REFERENCES

1. *Current Contents Connect,* http://scientific.thomson.com/products/ccc/.
2. *CA Selects,* http://www.cas.org/products/print/selects/index.html.
3. *Chemistry & Industry,* http://www.chemind.org/CI/index.jsp.
4. *Directory of Open Access Journals*, http://www.doaj.org/.
5. *CiteSeer (Analytical Chemistry Journals)*, http://www.chemseer.com/journals/journals_analytical.shtml.

7.2 Conference Papers

Collections of conference papers are considered in this context, since they are very similar to a focused journal issue. In fact, many conference papers are collected and published as special issues in a wide variety of research journals. In this case, papers are indexed and abstracted as if they were journal articles. Individually published conference proceedings, however, are a special problem. Literature citations to conference proceedings rarely correspond to the way in which they are listed in library catalogs. The availability of the British Library's Document Supply Centre Conference Index on the WWW[1] and CISTI Catalog[2], however, greatly assists in locating the necessary bibliographic information. Once the official conference name or conference title is known, libraries holding the volume can be found in CASSI (Chemical Abstracts Service Source Index) or WorldCat (http://www.worldcat.org/). In many cases, recent conferences also have their own web page, which may include abstracts of the presentations. These can be found by searching for keywords and the conference location on an Internet search engine such as Google. Both the Eastern Analytical Symposium (http://www.eas.org) and the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (http://www.pittcon.org) have web sites.

REFERENCES

1. The British Library, http://blpc.bl.uk/.
2. Canada Institute for Scientific and Technical Information, *CISTI Catalog*, http://cat.cisti-icist.nrc-cnrc.gc.ca/search.

7.3 Technical Reports

In addition to peer-reviewed journal articles, technical reports are an accepted method of publishing research results in fields such as environmental engineering, aerospace engineering, and energy research. Technical reports are also a necessary requirement for reporting the results of government-sponsored research. These submissions to the National Technical Information Service (NTIS) database are often progress reports and may simply be references to a published journal article. Searching for technical reports is usually based on a citation to an author or report number. The Stanford University Engineering Library has posted a 'How do I find?' web page for technical reports: http://www-sul.stanford.edu/collect/techrpt.html.

FURTHER READING

1. NTIS Technical Reports Database, http://www.ntis.gov/search/index.aspx.
   This collection covers government-sponsored US and worldwide scientific, technical engineering, and business-related publications. It includes references to the 3M publications and technical reports that NTIS has received from Government agencies since 1964 with a few exceptions.
2. Defense Technical Information Center (DTIC) Scientific and Technical Information Network (STINET), http://stinet.dtic.mil/info/s-stinet.html.
   The DTIC Technical Report collection includes all unclassified and unlimited technical report citations to full-text reports since 1965. It is freely searchable.
3. US Department of Energy (DOE) Reports Bibliographic Database, http://www.osti.gov/bridge/.
   The DOE database contains references to all DOE-sponsored scientific and technical reports sent to Government Printing Office (GPO) Depository Libraries, beginning with January 1, 1994. It is supplemented by the access to full-text DOE research and development reports in physics, chemistry, materials, biology, environmental sciences, energy technologies, etc. from 1996 onward.
4. NASA Technical Reports Server (NTRS), http://ntrs.nasa.gov/search.jsp.
   The NTRS provides access to NASA's (National Aeronautics and Space Administration) current and historical technical literature – approximately 500 K aerospace-related citations, 90 K full-text electronic documents, and 111 K images and videos. Information includes conference papers, images, journal articles, photos, meeting papers, movies, patents, research

reports, and technical videos. NTRS integrates three separate information collections through a common search interface:

- NACA Collection: Citations and reports from NACA period (1915–1958).
- NASA Collection: Citations and documents from 1958 to date.
- NIX Collection: Citations and images, photos, movies and videos downloaded from the NASA Image eXchange.

5. National Service Center for Environmental Publications NSCEP/NEPIS, http://www.epa.gov/ncepihom/. EPA's Gateway to 7000 in stock and 26 000 digital titles available free of charge to search and retrieve, download, print and/or order.
6. Yahoo-Technical Reports, http://dir.yahoo.com/Science/computer_science/technical_reports.

7.4 Theses

Virtually, all theses from US and Canadian universities (as well as an increasing number of European institutions) are deposited with ProQuest Digital Dissertations (formerly *University Microfilms*), which publishes *Dissertation Abstracts International*. Access to the complete file from 1871 is available on the WWW[(1)] (a fee-based service, which is available in most academic libraries).

In addition, CA indexes (but does not abstract) chemical theses from records provided by ProQuest.[(2)] NDLDT (Networked Digital Library of Theses and Dissertations) is a project that facilitates the development of institutional databases of electronic theses,[(3)] whose records can be harvested by Google.

REFERENCES

1. *Dissertation Abstracts International*, ProQuest Digital Dissertations, http://wwwlib.umi.com/dissertations/.
2. *ProQuest*, http://proquest.umi.com/i-std/en/dis/about.htm.
3. NDLDT (Networked Digital Library of Theses and Dissertations), http://www.ndltd.org/.

7.5 Patents

The patent literature is a very important albeit often underutilized resource, especially in an academic setting. Many of the patent documents referenced in CA are European or Japanese patent applications because these applications are publically available prior to the final granted patent. This results in a family of patent documents from different countries. Locating a granted US patent from, for example, a European patent (EP) application number is facilitated by the CA Patent Index (cumulated in SciFinder/SciFinder Scholar), as well as other commercial vendors. Over 126 K patent documents are currently referenced in CA each year and of these, over 1300 were classified in the Inorganic Analytical Chemistry or Organic Analytical Chemistry sections in 1997. A sampling of the titles of these patents is given in Table 2.

While patents are extensively covered in CA, the World Patent Index, and Claims (all fee-based services on STN International, as well as other commercial vendors), there are also several US patent databases available, at no cost, on the WWW. Each of these free services provides something a little different, in searching or display, so it is worth checking each of them out.

FURTHER READING

1. *Esp@cenet database,* http://ep.espacenet.com/. Provides access to worldwide patents and offers full-text PDF downloads for patents of less than 50 pages.
2. *MicroPatent,* http://www.micropat.com/static/index.htm.
3. *QPat US,* http://www.qpat.com/index.htm.
4. American Chemical Society Committee on Patents and Related Matters, *What Every Chemist Should Know About Patents*, 3rd edition, http://portal.acs.org/portal/fileFetch/C/WPCP_006903/pdf/WPCP_006903.pdf, *What Every Chemist Should Know About Patents* (2006 Supplement), http://membership.acs.org/C/CPRM/2006Supplement.pdf, (2002).

8 ABSTRACTING AND INDEXING SERVICES

An abstract is a brief summary or digest of the most significant results reported in a primary source. The services described below are clearly indispensable for

Table 2 Sample titles of analytical chemistry patents in *Chemical Abstracts (CA)* in 1997

Carbon content analyzer
Fiber-optic gas analyzer
Gas detection method and apparatus therefor
Heating-type oxygen sensor for determination of oxygen in automobile exhaust gas
Impedance-type humidity sensor with proton-conducting electrolyte
Impurity analyzer
Mass spectrometric analysis method and device
Methods and apparatus for detection of gases in oils
Quantitative analytical method of heavy metal ion in semiconductor grinding slurry
Semiconductor chemical sensor device and method of forming

collecting materials to write a review of a topic, to locate a specific fact or technique not easily found in secondary sources, and to follow current research advances.

The author abstracts are being increasingly used by the major services. CA (see Section 8.1), while depending on volunteer chemists as abstractors in the past (especially for articles published in Cyrillic and oriental languages), now employs staff abstractors and is increasing its use of author abstracts. This is especially true for ACS publications that provide electronic versions that are added to the CAPlus database ASAP (i.e. as soon as publishable). CAPlus is an enhanced version of the CA online database. ASAP is a new feature of the ACS journal's web editions (http://pubs.acs.org/) that allows the posting of articles prior to formal publication.

ISI's Web of Science, (see Section 8.2) which includes abstracts for about 65% of the articles indexed since 1991, uses the author's abstract exclusively.

Analytical Abstracts (see Section 8.3), however, because of its specialized perspective, does not use author abstracts. All of their abstracts are written in-house, from an analytical perspective, giving as much analytical detail as is practical.

The currency of these three services varies widely. The CAPlus online file is updated daily, and includes articles from ACS journals before they are published in print. Web of Science generally depends on the print journal and is usually about 3–4 weeks behind the current journal issue, with CA generally about 3–4 months behind and Analytical Abstracts generally about 7–8 months behind.

UnCover (see Section 8.4) and Institut de l'Information Scientifique et Technique (INIST) (see Section 8.5) are journal-article databases that are freely searchable on the WWW. They both are limited to brief bibliographic details and do not provide abstracts.

8.1 Chemical Abstracts

CA (http://www.cas.org/expertise/cascontent/caplus/index.html) currently abstracts over 1 million items (journal articles, patents, dissertations, reports, book chapters and reviews), of which only about 10 K are specifically classified each year as either inorganic or organic analytical chemistry. This, however, grossly underestimates their coverage since analysis articles are also classified in many other subject sections (e.g., drugs, enzymes, fossil fuels, and rocks).

CA, in print, is issued weekly with semiannual indexes, author, general subject (molecular) formula, chemical substance (names), and patent (numbers) that are cumulated every 5 years. The companion fee-based online service covers the chemical literature from 1967 to date and currently includes over 75% of all the abstracts published since 1907.

The key to effective use of CA is a thorough familiarity with the Index Guides. Each edition (e.g. 1992–1996) of the Index Guide is intended for use with the corresponding General Subject Index and Chemical Substance Index for that time period. The CA Index Guide is composed of a main body that contains cross-references, indexing notes, valid general subject index headings, and structural diagrams for cyclic and acyclic stereoparent compounds, as well as four appendices that describe the organization and interrelationship of subject terms and the rules for naming chemical compounds.

The entry under Analysis, in the 1997 CA Index Guide, lists both NT (chromatography, ellipsometry, kinetic analysis, polarimetry, titration) and RT (analytical numbers, gas sensors, sampling) and is supplemented by the entry in the CA General Subject Index, shown below.

8.1.1 Analysis

Studies of chemical analysis in only a broad sense are indexed at this heading.

More specific aspects are indexed when possible under hierarchically narrower headings; see Hierarchy 2, Appendix I of the Index Guide. For studies on the analysis of specific substances, see headings for those substances in the Chemical Substance Index. For classes of substances, such as amines, proteins, or minerals, see headings for those classes in the General Subject Index.

The importance of using the Index Guide, specifically the one for the time period being searched, cannot be overemphasized. For example, this is where you will find that diethyl ether is indexed as Ethane, 1,1′-oxybis- and that prior to 1997, the various types of chromatography, e.g. gas chromatography and paper chromatography, were assigned inverted General Subject Index headings, e.g. Chromatography, gas and Chromatography, paper.

The main difference between searching the print and online versions of CA is the use of CA Registry Numbers in the online CA file and the use of inverted CA Index Names in the print Chemical Substance Index. The online Registry File provides the links between the registry numbers and index names.

A compendium of instructional materials for searching CA (both print and on line) is available at http://www.indiana.edu/~cheminfo/cciim31.html.

8.2 Web of Science

Web of Science (http://scientific.thomson.com/products/wos/) is ISI's web product, now a component of ISI's Web of Knowledge. It includes the SCI, the Social Science Citation Index (SSCI), and the Arts and Humanities

Citation Index (AHCI). *Web of Science* is exclusively a journal-article database, but this definition includes some serial and review publications (e.g. *Materials Science Forum, Advances in Chemical Physics, Annual Review of Biochemistry*). With the Century of Science initiative, approximately 850 000 fully indexed journal articles have been added from 262 scientific journals published in the first half of the twentieth century.

The SCI continues to be available as a print publication. It serves as a very useful journal-article index (the Source Index is the author index and the Permuterm Index is the subject index), with the Citation Index providing an author index to articles that have been cited in more recent articles.

The *Web of Science* provides the same search capabilities, but expands the subject searching to include author keywords, words from the abstract (if present), and significant words from the titles of related articles (keywords plus). Web of Science also allows searching on journal titles and significant elements from an author's address (e.g. zipcodes) as well as providing links to related articles (i.e. articles that cite articles in common).

8.3 Analytical Abstracts

Analytical Abstracts (http://www.rsc.org/Publishing/CurrentAwareness/AA/index.asp) has nearly 400 K records. It is available in a monthly print version with a subject index that is cumulated annually along with an author index. An electronic fee-based version is also available covering 1980+ with abstracts since 1984. *Analytical Abstracts* covers the major analytical journals (and some of the more specialized journals) as well as British and Australian Standards, books, and some manufacturers' application notes.

One of the major features of *Analytical Abstracts* is the indexing. Index terms are qualified to denote the analyte (the substance being determined, detected, etc.), the matrix (the sample being analyzed) and the concept (the technique used, properties such as pH, humidity, chemometric technique, etc.).

The electronic version of Analytical Abstracts is a relatively small database (about 250 K records) compared to CA (15.6 M records) and the Web of Science (9.7 M records) and is limited to journal articles (98%), books (1%), and reports (1%). By comparison, CA and Web of Science, contain over 900 K and 180 K records that mention either analysis or analytical chemistry, respectively.

Because of the careful indexing and abstracting, Analytical Abstracts is not as up-to-date as either CA or Web of Science.

8.4 Ingenta Connect (Previously Uncover)

Ingenta Connect (http://www.ingentaconnect.com/) is a database of current article information taken from well over 17 000 multidisciplinary journals (including over 350 of the major chemistry journals) and contains brief descriptive information for over 23 000 000 articles that have appeared since the autumn of 1988. Ingenta has an automated alerting service that delivers the table of contents of your favorite periodicals directly to your e-mail box.

8.5 Institut de l'Information Scientifique et Technique

INIST (http://articlesciences.inist.fr/) is a journal-article database, sponsored by the Centre National De La Recherche Scientifique (CNRS) in France. It includes records for over 4.5 M articles since 1990. It has both an English and French version. While explicit journal coverage is not readily available, the fact that INIST indexes 49 journals with ''analytical'' in the title suggests that its coverage in the sciences is comprehensive.

8.6 Google Scholar

GS (http://scholar.google.com/) is a freely accessible web search engine that indexes the full text of scholarly literature across a variety of formats and disciplines. It is similar, albeit far more comprehensive, than Scirus, CiteSeer, and getCITED. GS is also similar to Scopus and Web of Science, but lacks both a structured presentation of results and any source information. GS also provides a citation index analogous to Scopus and Web of Science.

9 DATA COMPILATIONS

Data compilations have evolved from the primarily descriptive individual efforts of Beilstein (*Handbook of Organic Chemistry*), Gmelin (*Handbook of Inorganic Chemistry*), and Landolt–Bornstein (*Numerical Data and Functions in Physics, Chemistry, Astronomy, Geophysics and Technology*) to the International Critical Tables (ICT) (*Numerical Data of Physics, Chemistry, and Technology*), which was the first comprehensive data compilation subjected to rigorous evaluation by cooperating experts. Following publication of the ICT in 1933, an increase in the publication of specialized data compilations complemented completion of the sixth edition of Landolt–Bornstein, which appeared in 28 volumes between 1950 and 1980, and the Landolt–Bornstein New Series (Numerical Data and Functional Relationships in Science and Technology), which is a work in progress.

The important distinction between the named compilations (which are described in depth by Wiggins in

Chemical Information Sources, McGraw-Hill, New York, 1991) is that Beilstein (organic compounds) and Gmelin (inorganic/organometallic compounds) are organized by chemical compounds and Landolt–Bronstein is organized by physical property.

The *CRC Handbook of Chemistry and Physics*, under the editorship of David Lide, is an obvious first choice for basic property data. It also provides information on sources of additional data for each property. It is available in print, on CD-ROM, and on the WWW (13). *Knovel Critical Tables* (14), which are freely available to academic institutions, complement the physical-property sections of the HCP. Beilstein and Gmelin are now also searchable in conjunction with a wide variety of specialized databases through DiscoveryGate (15). Landolt–Bornstein has added a freely available electronic index (16). *Chemical Abstracts* is adding a wide variety of both experimental (including spectra) and predicted physical properties (17). The US National Institute for Standards and Technology (NIST) offers the freely available *NIST Chemistry WebBook*, which provides a wide variety of spectra (18). The Mineralogy Database (19) offers data on over 4.4 K mineral species.

1. Electrochemical Science and Technology Information Resource (ESTIR): http://electrochem.cwru.edu/estir/
   This web site includes information on the following topics:
   - Electrochemistry and related subjects on the Internet WWW Sites
   - Newsgroups/mailing lists/frequently asked questions
   - Public domain information on the Internet (bibliographies/software/physical and chemical data)
   - Popular science information
   - Review chapters (index of 2000 items)
   - Books (index of 750 items)
   - Proceedings volumes (index of 450 items)
   - Graduate schools (300 listings from over 45 countries)
   - Scientific and technical societies
   - Scientific and technical journals
   - Handbooks, bibliographies
   - Nomenclature, standards, etc.
   - Meetings.
2. *Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*, 14th edition, Merck Manuals, Whitehouse Station, 2006.
   Useful for finding CA Index Names.
3. *Aldrich Catalog: Handbook of Fine Chemicals*, Aldrich Chemical Co., Milwaukee, 2007/2008. http://www.sigmaaldrich.com/Brands/Aldrich.html
   Provides spectra and Beilstein references.
4. *Combined Chemical Dictionary (CCD)* includes the *Dictionary of Organic Compounds, Dictionary of Natural Products, Dictionary of Inorganic and Organometallic Compounds, Dictionary of Drugs, Dictionary of Analytical Reagents, Dictionary of Carbohydrates,* and *Dictionary of Marine Natural Products.* The CCD is a subset of ChemNetBase (below)
5. *Hazardous Substances Data Bank*, Toxicology Data Network, (National Library of Medicine), http://toxnet.nlm.nih.gov.
   Chemical/physical properties, safety, toxicity, environmental fate, etc. http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html.
6. *Handbook of Analytical Chemistry*, McGraw-Hill, New York, 1963. http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html
   Includes both procedures and numerical data.
7. *National Institute of Standards and Technology (NIST) Critically Selected Stability Constants of Metal Complexes* (computer file) Version 8.0, NIST Standard Reference Database 46, NIST, Gaithersburg, MD.
8. *Reagent Chemicals: ACS Specifications,* 10th edition, ACS, Washington, DC, 2005.
   Includes general physical properties and analytical uses for all 500 reagent chemicals.
9. *NIST Online Reference Databases*, http://www.nist.gov/srd/online.htm
   Spectra, physical constants, crystal structures, sequence databases.
10. *ChemFinder WebServer*, http://chemfinder.camsoft.com/
    Physical property data and two-dimensional chemical structures for common chemicals. Indexes over 75K unique substances from over 350 chemical information sites.
11. *Mineralogy Database*, David Barthelmy http://www.webmineral.com
    Database of over 4.4K individual mineral species; crystallography, chemical composition, physical, and optical properties.
12. *WebElements,* The University of Sheffield/WebElements Ltd, UK http://www.webelements.com/
    Using a periodic table arrangement, extensive data is provided for each chemical element.
13. *CRC Handbook of Chemistry and Physics*. http://www.chemnetbase.com/
14. *Knovel Critical Tables* http://www.knovel.com/knovel2/Toc.jsp?BookID = 761

15. *DiscoveryGate* http://www.mdl.com/solutions/solutions_for/academics/dg_academics.jsp
16. *Landolt–Bornstein Substance/Property Index* http://lb.chemie.uni-hamburg.de/index.html
17. *Chemical Abstracts Registry File*. http://www.cas.org/expertise/cascontent/registry/index.html
18. *NIST Chemistry WebBook* http://webbook.nist.gov/chemistry/
19. *Mineralogy Database* http://webmineral.com/

10 DICTIONARIES, CATALOGS, DIRECTORIES, AND SO ON

There are a number of specialized dictionaries, catalogs, and directories related to analytical chemistry, which can be very useful. For example, since literature references in journal articles only include journal abbreviations, and full titles are necessary for searching library catalogs, the CASSI (http://www.cas.org/PRINTED/cassi.html) provides a letter-by-letter alphabetical journal abbreviation index to full journal titles.

Of late, many of these dictionaries have not been updated with new editions, perhaps as a testament to the ubiquity of Google's define option. Some new electronic resources include IATE (Inter-Active Terminology for Europe), which is used for translating a term from one European language to another. BioABACUS (Biotechnology Abbreviation and Acronym Uncovering Service) is a searchable database of abbreviations and acronyms in Biotechnology.

FURTHER READING

1. Huthig, *Dictionary of Chromatography: English, German, French, Russian*, revised edition, Huthig, New York, 1984.
   Contains about 3500 terms.
2. *A Dictionary of Spectroscopy*, 2nd edition, John Wiley & Sons, New York, 1982.
3. *Academic Press Dictionary of Science and Technology*, Academic Press, San Diego, 1992.
4. *Comprehensive Biological Catalysis*, Academic Press, San Diego, Vol. 4, *The Lexicon of Terms and Concepts in Mechanistic Enzymology*, 1998.
5. *German Dictionary of Analytical Chemistry. German-English/English-German*, Routledge, London, 1998.
6. *ACS LabGuide*, American Chemical Society, Washington, DC.
   Instrument and laboratory supplies, laboratory reagents and standards, services, original equipment manufacturers supplies and suppliers, company directory and product index. The 2006 volume was issued as a supplement to *C&E News*.
7. *ACS Directory of Graduate Research (DGR)*, American Chemical Society, Washington, DC, 1997 http://dgr.rints.com/.
8. The Chemical Industry Home Page. http://www.neis.com/.
   A directory of associations, resources, and web sites.
9. WWW-CHE Info Virtual Library, *Analytical Methods, Products and Services*, University of Florida, http://www.che.ufl.edu/WWW-CHE/topics/analytical.html.
   Listings relevant to chemical and process engineering.
10. National Institute of Standards and Technology (NIST), *Analytical Chemistry Databases and Publications*, http://www.nist.gov/srd/analy.htm.
    A listing of databases (mass spectra, infrared, surface, diffraction, spectrum analysis) available from NIST.
11. IATE (Inter-Active Terminology for Europe), http://iate.europa.eu/iatediff/SearchByQueryLoad.do.
12. *BioABACUS*, *BioABACUS: Biotechnology Abbreviations and Acronyms Uncovering Service*, New Mexico State University, http://www.nmsu.edu/~molbio/bioABACUShome.htm.
13. D.M., Glick, ed. *Glossary of Biochemistry and Molecular Biology*, Portland Press, London, http://www.portlandpress.com/pp/books/online/glick/default.htm.
14. *Water Quality Association Glossary of Terms*, http://www.wqa.org/glossary.cfm, 2008.
    Click 'tools and resources', then click 'Glossary of Terms'

11 INTERNET SEARCH ENGINES AND GUIDES

The transition from hand-set type and print indexes (in the 1950s and 1960s), through computer composition and online searching (in the 1970s and 1980s), to web-based publication and Internet search engines (in the 1990s) has been truly revolutionary. Current estimates suggest that there are ~30B pages on the WWW[1] and Google is the preeminent Web search engine for most WWW users.[2]

If the Internet were a book, Search Engines (e.g. Google) would be the index and Internet Guides (e.g. Yahoo) would be the contents pages. To maintain currency, given the dynamic nature of the WWW, search engines are constantly running software called "robots" or "crawlers" that read entire web sites and

update their index entries. Search engines save their data alphabetically, just like a book's index. Owing to the enormous amount of information, a search engine's listings are not browsable. They are searched by entering words (or a phrase) that specifies what you are looking for, and the results are presented as a list in descending order of relevance (which is an indication of the positional frequency of the search term/s). Internet Guides, on the other hand, are designed to be browsed and offer a much more structured and selective approach to information gathering.

Unfortunately, of the estimated ~30B pages on the WWW, coverage by the various search engines varies widely.[1] This situation is further complicated[2] by the possible exclusion of major search engines (e.g. HotBot and Northern Light) from the multifile search engines (e.g. Dogpile), in certain circumstances.

In addition, the popularity of Internet searching also magnifies the ''principle of least effort'', which is exemplified by the obviously uninformed proclamations about the unique benefits of searching the WWW. A recent article[3] has exposed some obvious deficiencies of depending on WWW searches for research articles (e.g. only about 10% of what one could retrieve from CAS Online is found on the WWW.)

This is not to diminish the importance of many important web pages and Internet-accessible databases. The WWW is extremely useful for finding material (specialized databases, personal and company-related information, and conference proceedings, etc.) that is not indexed in traditional sources, such as CA, Analytical Abstracts, or the Web of Science.

In addition to searching the WWW, there are numerous news groups that offer an opportunity to connect with others sharing similar interests. Some groups related to analytical chemistry include sci.chem; sci.chem.electrochem.battery; sci.chem.analytical; and sci.chem.electrochem.[3]

REFERENCES

1. Boutell.com, http://www.boutell.com/newfaq/misc/sizeofweb.html.
2. Search Engine Showdown, http://searchengineshowdown.com/.
3. Chemistry Forum, http://www.docendi.org/chemistry-f173.html.

11.1 Analytical Chemistry Guides

- Links to Internet Sources for Analytical Chemistry http://cheminfo.informatics.indiana.edu/cicc/cis/index.php/Link_to_Internet_Sources_for_Analytical_Chemistry is an extensive listing covering Spectral & Crystallographic Sources; Structural Databases: Biological Molecules; Amino Acids, Peptides, and Proteins: Structural Databases; Nucleic Acids: Sequence and Structural Databases; Miscellaneous Sources for Analytical Chemistry; and Printed Sources for Analytical Chemistry Information
- Analytical Sciences Digital Library http://www.asdlib.org/ is an NSF-funded digital library responsible for collecting, cataloging, linking, and publishing peer-reviewed web-based materials on innovations in curricular development and supporting technical resources.
- UMEA University: Analytical Chemistry Springboard http://www.anachem.umu.se/jumpstation.html. This site provides a comprehensive list of analytical chemistry resources on the Internet, covering both descriptions of analytical techniques (e.g. chemical sensors, mass spectrometry, sample preparation) and information resources (departmental listings, meetings and short courses, software).
- Links for Chemists – Analytical Chemistry http://www.liv.ac.uk/chemistry/links/refanal.html
- University of Florida. Analytical Methods, Products and Services. http://www.che.ufl.edu/WWW-CHE/topics/analytical.html
- Chemical Information Sources Wiki http://cheminfo.informatics.indiana.edu/cicc/cis/index.php/main_page

Sections on How and Where to Start; How and Where to Search: General; How and Where to Search: Specialized (Analytical Chemistry), Communicating in Chemistry; Miscellaneous; and Supplemental Resources.

11.2 Internet Search Engines

- DOGPILE http://www.dogpile.com is a Metasearch engine that provides combined results from: Google, Yahoo! Search, Live Search, Ask.com, About, MIVA, LookSmart and more).
- GS http://scholar.google.com/schhp? tab = ws provides access to a broad range of disciplines and multiple formats from a variety of sources: peer-reviewed papers, theses, books, abstracts, and articles, from academic publishers, professional societies, preprint repositories, universities, and other scholarly organizations.

12 MISCELLANEOUS TIPS

1. There are two slightly different transliteration schemes used by the Library of Congress and scientific publishers. For example,

Library of Congress – Ioffe, Izvestiia
Chemical Abstracts – Joffe, Izvestiya
Web of Science – Joffe, Izvestiya

2. There are also variations in American and British (Commonwealth) spellings: fetus–foetus, aging–ageing.
3. Accent marks, umlauts and compound names have not been treated consistently by various publishers

Published as:	WoS(<1998):
Brown-Wensley	BrownWensley
Van der Waals	Vanderwaals
WoS(>1997):	CAS Print:
Brown-Wensley	BrownWensley
Van der Waals	VanderWaals

An author search in SciFinder/SciFinder Scholar provides a listing of alternative spellings of author's last names.

4. Alphabetical arrangement also varies between library catalogs (word-by-word) and dictionaries (letter-by-letter).

Library Catalogs (word-by-word)	Dictionaries (letter-by-letter)
Chemic green	Chemically pure
Chemical warfare	Chemical warfare
Chemically pure	Chemic green

5. The CASSI has a unique letter-by-letter arrangement of the abbreviations (which are bold faced). This results in the following alphabetical listing.

Z. Chem.	Zeitschrift fuer Chemie
Zeiss	Zeiss Information
Zh. Obshch. Khim.	Zhurnal Obshchei Khimii
Z. Kristallogr	Zeitschrift fuer Kristallographie
Zoon	Zoon
Z. Phys	Zeitschrift fuer Physik

6. The CA Author Index (in print) is unique in that the alphabetical arrangement of the author's forenames are by the initials and not the full forenames. This results in the following alphabetical order:

Brown, D.;
Brown, Douglas D.;
Brown, Dennis Judson;
Brown, D. Kyle;
Brown, David Ward.

ABBREVIATIONS AND ACRONYMS

ACS	American Chemical Society
AHCI	Arts and Humanities Citation Index
AOAC	Association of Official Analytical Chemists
ASAP	As Soon as Publishable
ASTM	American Society for Testing and Materials
BioABACUS	Biotechnology Abbreviation and Acronym Uncovering Service
BT	Broader Terms
CA	Chemical Abstracts
CAS	Chemical Abstracts Service
CASSI	Chemical Abstracts Service Source Index
CNRS	Centre National De La Recherche Scientifique
CPFAC	Current Protocols in Field Analytical Chemistry
DOE	Department of Energy
DTIC	Defense Technical Information Center
EAC	Encyclopedia of Analytical Chemistry
EP	European Patent
EPA	Environmental Protection Agency
ESTIR	Electrochemical Science and Technology Information Resource
GPO	Government Printing Office
GS	Google Scholar
IATE	Inter-Active Terminology for Europe
ICT	International Critical Tables
INIST	Institut de l'Information Scientifique et Technique
ISI	Institute for Scientific Information
NASA	National Aeronautics and Space Administration
NDLDT	Networked Digital Library of Theses and Dissertations
NEPIS	National Environmental Publications Information System
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
NT	Narrower Terms
NTIS	National Technical Information Service
NTRS	NASA Technical Report Server
OPAC	Online Public Access Catalog
RSC	Royal Society of Chemistry
RT	Related Terms
SCI	Science Citation Index

SM	Standard Method	USEPA	United States Environmental Protection Agency
SSCI	Social Science Citation Index		
TrAC	Trends in Analytical Chemistry	WWW	World Wide Web
TRS	Technical Report Server		

Microwave Techniques

Frank Smith
Laurentian University, Sudbury, Canada

Guohua Xiong
Jiujiang Teachers College, Jiujiang, China

1 Introduction **1**
2 Theoretical Considerations **2**
2.1 Interactions of Microwaves with Matter 2
2.2 The Magnetron and Instrument Components 4
2.3 Microwave Heating and Superheating 5
3 Instrumentation and Equipment **5**
3.1 Ovens and Complete Microwave Systems 5
3.2 Vessels 7
4 Safety Considerations **7**
5 Applications **8**
5.1 Introduction and Review Articles 8
5.2 Environmental Samples 9
5.3 Biological and Clinical Samples 14
5.4 Food and Pharmaceutical Samples 17
5.5 Mineral and Ceramic Samples 19
5.6 Industrial Samples 20
5.7 Microwave-assisted Sample Drying 21
5.8 Microwave Fusions 21
6 Concluding Remarks **22**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **24**

A concise yet comprehensive review of the use of microwave-assisted techniques for sample preparation in analytical chemistry is presented. A historical review of the development of the field is followed by summaries of the most significant theoretical principles, available instrumentation and safety considerations. The major part of the review concerns recent applications of the techniques to the preparation of a wide range of sample types. These include environmental, soil, biological, clinical, food, pharmaceutical, mineral, ceramic and industrial samples. Examples of microwave-assisted drying and fusion of analytical samples are also given. Throughout, reference is made to the most important review articles. Where appropriate, original papers are summarized and cited.

1 INTRODUCTION

Over the past decade, manufacturing industry generally has come under increasing pressure on two main fronts – economic and environmental. On the economic front, increasing competition due largely to globalization has forced companies to improve quality, efficiency and productivity. At the same time, society has become increasingly concerned with environmental effects, leading to more stringent controls on emissions and effluents. These influences have in turn made efficient management more and more dependent on reliable chemical analysis at all stages of the manufacturing process. Analytical chemistry has never had a more important role to play and the signs are that this role will become even more crucial in the future. These pressures have favored the trend for analytical chemistry instrumentation to become more and more sophisticated. Improved instruments with the capability to detect ever-smaller quantities of more and more elements and compounds, more rapidly than before, are constantly being developed and usually incorporate powerful computational capabilities. However, many of these advanced techniques require extensive sample pretreatment. Yet 10 years ago, few innovations had been made to sample dissolution or pretreatment methods for a century or more. Thus sample preparation had become the slow step in the analytical process, providing the impetus for the development of fast and reliable alternative methods to those traditionally employed. The innovative techniques introduced to speed up sample pretreatment include supercritical fluid extraction (SFE),[1] solid-phase extraction (SPE),[2] solid-phase microextraction (SPME),[3] the supported liquid membrane (SLM) technique for extraction,[4] accelerated solvent extraction (ASE)[5] and high-performance capillary electrophoresis (HPCE).[6] However, some of the most significant advances in this area have been based on the use of microwave radiation to accelerate the process of sample preparation.

This article will focus on these developments and is an attempt to identify the most significant advances described in the literature up to the middle of 1999, along with sufficient background information to provide an overall perspective of the potentials and limitations of microwave-assisted techniques.

The preparation of biological samples by wet ashing at atmospheric pressure was the first chemical decomposition process to exploit the microwave oven for analytical purposes. The first microwave-assisted high-pressure dissolutions were reported about a decade later.[7] Conventional wet ashing dissolution techniques are very slow,

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

and involve heating samples in various acids or mixtures of acids in an oven or on a hot-plate for extended periods of time. Often the sample – acid mixture must be brought down to dryness, cooled and then reheated with more acid in order to ensure dissolution. Also, these techniques are susceptible to loss of volatile analytes and cross-contamination. Other problems often associated with conventional sample preparation techniques include tedium, danger due to the use of strong oxidizers, incomplete digestion and the need for large amounts of reagents, constant supervision and special fume hoods. For those samples containing intractable matrices such as refractory minerals, closed vessels such as steel-jacketed Teflon™ polytetrafluoroethylene (PTFE) bombs or glass Carius tubes heated in a conventional oven have traditionally been used – again, lengthy heating times are required.

In 1975, Abu-Samra et al. reported the use of microwave ovens for the purpose of wet ashing biological samples prior to elemental analysis.[8,9] These authors used a domestic microwave oven equipped with a spray fume scrubber to wet ash a range of biological samples. The only previously recorded laboratory uses of microwave ovens were for the drying of wet cake and paste in an inorganic chemical manufacturing plant[10] and for the sintering of silica and alumina mixtures.[11] The inspiration for pressure-dissolution studies came from a US Bureau of Mines report,[12] which described how rapid dissolution of some mineral samples had been achieved by using a microwave oven to heat samples and an acid mixture contained in polycarbonate bottles. When this experiment was repeated,[7] the method showed great potential, but the vessels proved unsuitable for more rigorous conditions. A great deal of effort went into the search for better reaction vessels – the vessels needed to be transparent to microwave radiation, chemically inert and resistant to acid attack, able to withstand high temperatures and pressures and yet be reasonably cheap. Vessels should also be small enough to allow up to a dozen inside the oven at the same time, yet be readily sealed and opened to enable routine use by laboratory staff. A range of commercially available Teflon™ bottles was tried, but these deformed or burst during the heating process. Some test vessels were machined out of solid Teflon™ blocks, but it was found that the screw threads in the cap deformed during heating, allowing the cap to explode off. Finally, a vessel was constructed from 4-cm diameter Teflon™-PFA (polyfluoroalkoxy) tubing and end caps and this proved successful. Arrangements were then made with the manufacturer[13] to make vessels from the tubing. These were marketed and eventually became the vessel of choice for microwave experimentation. The identification and development of this more suitable vessel were a very important step forward and led to the development of a variety of digestion protocols,[7,14–20] and also to a range of organic and organometallic syntheses,[21–26] all extremely fast compared with the traditional methods that they were intended to replace. Since then, both oven and vessel design have improved markedly and include the development of a continuous-flow synthesis system,[27] a continuous-flow analysis system and very high-pressure and-temperature (including fusion) capabilities.

Dramatic reductions in sample preparation times, containment of volatiles and sample type versatility are evident advantages of closed vessel dissolutions over classical wet digestion methods. Closed vessels result in the development of high temperatures and pressures which make dissolutions much more vigorous and rapid, hence the greater speed of the technique. The increased acceptance of the microwave pressure-dissolution technique is reflected by the dramatic rise in the number of publications appearing over the past few years. These report dissolution methods involving geological, botanical, zoological, environmental, food, sludge, coal and ash, metallic materials and synthetic material samples. Other papers focus on drying, on-line/flow-injection/continuous-flow systems, extractions/desorptions, distillations, total nitrogen determinations (Kjeldahl), reviews, general reports and databases, automated/robotic systems, oven modifications, special dissolution techniques and equipment, safety and moisture content determinations. These topics are all discussed in some detail in the sections that follow.

In many analytical laboratories, the microwave pressure-dissolution technique is already the method of choice for sample preparation. The new protocols reported have been, in most instances, verified exhaustively against standard reference materials. Microwave pressure-dissolution must now be considered an accepted technique. Other microwave-assisted processes for sample preparation are being constantly developed, put into practice and gaining general acceptance. This is an extraordinarily dynamic and rapidly changing area of chemistry.

2 THEORETICAL CONSIDERATIONS

These have been covered in depth in a number of excellent reviews,[28–33] that of Mingos and Baghurst being particularly recommended.[28]

2.1 Interactions of Microwaves with Matter

Microwaves form part of the spectrum of electromagnetic radiation, and fall below the infrared (IR) band, with frequencies of from 30 GHz to 300 MHz (corresponding to wavelengths of from 1 cm to 1 m). This is nonionizing radiation, with insufficient energy to rupture chemical bonds, but enough energy to cause molecular rotation

or movement of ions. These effects will, in turn, cause heating. This region of the electromagnetic spectrum is also used for radio detection and ranging (RADAR) and telecommunications purposes. In order to cause as few disruptions as possible to these long-range uses, certain fixed frequencies have been identified by international agreement (International Radio Regulations, Geneva, 1959) for industrial, scientific and medical (ISM) use. The approved ISM frequencies are 13.56, 27.12, 40.68, 915 and 2450 MHz and 5.80 and 22.125 GHz.

Some materials (metals, for example) are completely impervious to microwaves, reflecting them and absorbing no radiation. Other materials are completely transparent to these waves, freely allowing their passage, again with no energy absorption. However, there are a whole range of substances which do absorb microwave radiation, and become heated as a result. The extent of the observed heating effect is dependent on a variety of factors.

Both conduction and dielectric polarization are sources of microwave heating. Dielectric polarization comprises several individual components, namely electronic polarization, atomic polarization, dipolar polarization and interfacial polarization (or the Maxwell–Wagner effect). With an oscillating electric field, such as that associated with electromagnetic radiation, the response of a material depends on the timescales of the orientation and disorientation phenomena relative to the frequency of the radiation. Since the timescales for electronic and atomic polarization/depolarization are much faster than the microwave frequencies, these two kinds of polarization can usually be ignored. In practice, only dipolar polarization and interfacial polarization make any significant contributions to the total polarization and hence to microwave heating because the timescales associated with them are comparable to microwave frequencies.

Polarizability may be represented by the dielectric constant (ε'). The greater the ε' value, the more electromagnetic field energy can be stored in the material. Moreover, the dielectric loss (ε'') measures the efficiency with which the energy of the electromagnetic radiation can be converted into heat. It is known that ε' and ε'' vary with frequency in different manners. Thus the (dielectric) loss tangent ($\varepsilon''/\varepsilon' = \tan\delta$) is defined to describe the ability of a material to convert electromagnetic energy into heat energy at a given frequency and temperature.

Dielectric relaxation occurs when the electric field that induces the polarization is removed. The time required to reduce the order to 1/e of its original value is referred to as the relaxation time (τ). It can be calculated by the Debye equation[30] [Equation 1]:

$$\tau = \frac{\xi}{2kT} \tag{1}$$

where ξ is a coefficient associated with the size of the molecule and its environmental effects. The theoretical calculation of ε' and ε'' may be carried out by the use of the following Debye equations[30,31] [Equations 2 and 3]:

$$\varepsilon' = \varepsilon'_\infty + \frac{\varepsilon'_0 - \varepsilon'_\infty}{1 + w^2\tau^2} \tag{2}$$

$$\varepsilon'' = \frac{(\varepsilon'_0 - \varepsilon'_\infty)w\tau}{1 + w^2\tau^2} \tag{3}$$

where w is the frequency and ε'_∞ and ε'_0 are referred to as the high-frequency ($w \gg 2\pi/\tau$) and static dielectric constants, respectively. The loss factor (ε'') has a maximum value when $w = 1/\pi$. Generally, because of the interaction of the molecules with their neighbors, the real relaxation spectra of liquids and solids do not coincide well with those predicted by the Debye equations. In these cases, a modification to the calculation of relaxation time (τ) is necessary[32] [Equation 4]:

$$\tau = \tau_0 e^{U/kT} \tag{4}$$

where τ_0 is a calculated relaxation time from the Debye equation and U is the value of the potential barrier that a molecule has to cross when rotation occurs.

The dielectric constant of an inhomogeneous material, typically a suspension of conducting particles in a nonconducting medium, is also frequency dependent. Owing to the build-up of charges between the interfaces, dielectric loss is induced. This is known as the Maxwell–Wagner effect. The related absorption centred around a frequency of 10^7 Hz may lie in the microwave region.

As one of the two sources of microwave heating, conduction loss can be viewed as an extended Maxwell–Wagner effect, which contributes to the total loss. This conductive part depends on the direct current (DC) conductivity itself. For highly conductive liquids and solids, conduction losses may be on the same scale as dipolar relaxation loss, and there comes a point where the conductive loss effects are larger than dipolar relaxation effects.

Finally, for microwave heating of liquids and solids, the rate of rise of temperature due to the microwave electric field is governed by Equation (5):[33]

$$\frac{\delta T}{\delta t} = \frac{\text{constant} \times \varepsilon'' f E^2}{\rho C_p} \tag{5}$$

where E is the electric field intensity, ρ the density of the material, C_p the specific heat capacity and f the frequency of the electromagnetic field in hertz. Since these physical properties and also the dielectric loss for liquid and solid samples, are all temperature dependent, carrying out a full theoretical analysis of dielectric heating is an extraordinarily complex task. Nevertheless, it is possible

to make useful generalizations which assist in designing protocols for specific experiments.

In the case of liquids, polar compounds (e.g. water, ethanol) are readily heated in a microwave field, whereas nonpolar compounds (e.g. aliphatic or aromatic hydrocarbons) or those with molecules that have no net dipole moment (e.g. carbon tetrachloride) tend to be more transparent to microwave radiation and thus absorb very little energy and heat up far less readily. The electric field component of the microwave radiation interacts with the electric dipoles on the polar molecules leading to heating via dipole relaxation. For solutions containing dissolved ions, these can also interact with the oscillating electric field associated with microwave radiation, augmenting the dipole heating effect with a conduction effect.

Those solid materials that absorb microwave radiation efficiently (e.g. CuO and NiO) usually do so through a mechanism involving conduction by electrons, ions or holes. In addition, there may be microwave absorption by polar molecules weakly bound within the solid structure or lattice. For example, solid ice is transparent to microwave radiation, but heats rapidly owing to absorption by the liquid water molecules present within the structure. Other dipole interactions may be present in solids owing to irregularities in lattice structure and across grain boundaries.

The extent of the heating effect may also be frequency dependent. Hence the penetration of microwave radiation into water varies roughly with the wavelength of radiation used, so for deep penetration (to ensure the thorough cooking or warming of foods, for example) a longer wavelength (lower frequency) than that giving optimal absorption of microwave energy and leading to excessive surface heating is preferable. In practice, domestic microwave ovens operate at a frequency of 2.45 GHz, much lower than the 20 GHz that would give the optimal heating effect.[28] Some restaurant models do have an additional and higher frequency (9.8 GHz) available for the browning of foods – the heating effect is augmented but confined to the surface region.

Hence the major mechanisms leading to microwave absorption and consequent heating in solid, liquid and gaseous materials are understood. However, the detailed mechanisms involved for any particular real-world system are complex and extremely difficult to model, since they include dipole moments, numbers and sizes of ions involved, concentration effects, temperature effects and frequency effects, in addition to changes in the microwave field which occur as the temperature and conditions within the substrate change.

2.2 The Magnetron and Instrument Components

The magnetron generates the microwave radiation. The external configuration of different magnetrons varies, but the basic internal structures are the same.[34] These include the anode, the cathode or filament, the antenna and the magnets. The anode is a hollow cylinder of iron with an even number of cylindrical (or trapezoidal) cavities running on the inside parallel to the axis. The metal segments separating these cavities are called the anode vanes. When the magnetron is operating, the matched cavities behave as tuned circuits and determine the output frequency of the tube. Alternate segments must be connected, or strapped, so that each segment is opposite in polarity to the segment on either side. The cathode is a wire running through the center of the hollow anode, supported by the filament leads, which are carefully sealed into the tube and shielded. The antenna is a probe or loop that is connected to the anode and extends into one of the cavities. It is coupled to the waveguide, a hollow sheet metal conduit, into which it transmits the microwave radiation, which subsequently follows the waveguide into the oven cavity. Strong permanent magnets are mounted around the magnetron so that the magnetic field is parallel with the axis of the cathode.

Magnetron operation starts when a low voltage is applied to the wire cathode and, at the same time, a high potential is applied between the cathode and anode. Electrons stream off the heated cathode and are attracted strongly to the anode, but this motion is opposed by the magnetic field and so the electrons spiral around the cathode towards the anode. The whirling cloud of electrons, influenced by the high voltage and the strong magnetic field, forms a rotating pattern that interacts with the configuration of the interior surface of the anode to produce oscillating charges in the resonant cavities of the anode. The charge oscillation leads to the release of electromagnetic radiation at a specific frequency determined by the geometry of the cavities. Once the oscillations have increased sufficiently in amplitude, the radiation flows via the antenna into the waveguide and then into the oven cavity.

The cavity walls are metallic and reflect microwave radiation so that it may pass through the sample many times and yet be contained within the cavity. Oven doors are usually of transparent material covered with metallic gauze, which reflects the radiation. Particular care is taken with door design to ensure no radiation leakage. Standing waves within the cavity cause an uneven distribution of energy leading to 'hot spots' – this problem may be reduced by the installation of a mode stirrer, which looks rather like a small ceiling fan. Alternatively, a rotating carousel on the floor of the oven, on which samples are placed, smoothes out the field experienced by the target materials.

In a domestic microwave oven, the magnetron is either on, and operating at full power, or switched off. Hence the variable power settings available are made possible

by pulsing the microwave power – that is, by alternately switching the magnetron on and off according to a duty cycle. Thus a 600-W oven with a 30-s duty cycle can deliver an average of 300 W by switching the magnetron on and off every 15 s. For chemical applications and the minimization of oscillating temperature conditions in the sample, very short pulse times are an advantage and give greater control. Some specialized laboratory ovens have programmable, continuously variable, unpulsed microwave power sources.

For poorly absorbing samples, particularly where only small quantities are available, it may be worth considering an alternative to the multimode oven described above. A single-mode resonant cavity may be tuned to the characteristics of the particular sample. Such an experimental setup has been described for use in the processing of poorly absorbing polymers.[35]

2.3 Microwave Heating and Superheating

During the early experiments involving microwave digestion and synthesis, it was clearly understood that when reactions were carried out in closed containers under high pressures, solvent boiling points became elevated, allowing the reaction mixtures to attain far higher temperatures than would be present under atmospheric pressure, where the normal boiling point of the solvent was the limiting temperature. Thus accelerated reaction rates and dramatically reduced reaction times were readily explicable. However, some observers noticed increases in reaction rates or dissolution times when reactions were carried out in open vessels under atmospheric pressure. This was one of the reasons why the idea that there was some special 'microwave effect' in operation arose. This was of particular concern to the food industry, since it was feared that microwave heating might lead to a different series of reactions in foods subjected to microwave heating than those heated in conventional ovens. It was also of great interest to synthetic chemists seeking to alter the products of various reactions. A simple series of experiments led to some very significant results, which went a long way towards explaining this particular anomaly.[36] The rapid heating associated with microwaves can lead to superheating in strongly absorbing solutions, e.g. water can be heated to about 10 °C above its normal boiling point under atmospheric pressure. An increase in temperature of this order can lead to a two- or threefold increase in reaction rate. Thus the more rapid dissolution of analytical samples in open vessels observed with microwave heating[37,38] may be explained without recourse to a special effect. Repeated experiments have failed to confirm that microwave heating is in any way different from conventional heating in its chemical or physical effects on matter.[28]

3 INSTRUMENTATION AND EQUIPMENT

The intention of this section is to provide an overview of the main types of instrumentation and equipment available commercially. Since this is such a rapidly expanding field, the list given is not exhaustive. In most instances it is possible to use relatively cheap domestic ovens to perform microwave-assisted chemical procedures. However, the more expensive ovens specifically designed for laboratory operation have many desirable features which reduce the risk of accidents permit a far higher degree of control over the progress of the reaction taking place and allow higher temperatures and pressures to be safely employed.

3.1 Ovens and Complete Microwave Systems

A number of manufacturers now supply microwave ovens specially designed for laboratory use. In general, these

Table 1 Ovens and systems available

Manufacturer [US distributor]	Model and details
Anton Paar GmbH Kärntner Strasse 322 A-8054 Graz Austria 43 316 2570 www.anton-paar.com/ap [www.perkin-elmer.com]	**Multiwave** Power 1000 W, dim. (w × d × h) 64 × 67 × 49 cm, wt 45 kg. P_{max} 75 bar, T_{max} 300 °C, 12-position carousel. Vessels: 50/100 mL TFM, 25/50 mL quartz. *P* and *T* monitored in each vessel, full system control with built-in PC and methods library. Unpulsed microwave power. **PMD** Power 750 W, dim. (w × d × h) 47 × 47 × 53 cm, wt 20 kg. P_{max} 80 bar, T_{max} 300 °C, holds 2 vessels only. Vessels: 35/50/120 mL Exhaust and cooling unit (ECU) available.

(*continued overleaf*)

Table 1 (*continued*)

Manufacturer [US distributor]	Model and details
Berghof Laborprodukte GmbH Harretstrasse 1 72800 Eningen Germany 49 7121 8940 www.berghof.com [www.berghofusa.com]	**MWS-1** Power 850 W, dim. (w × d × h) 55 × 42 × 37 cm, wt 29 kg. P_{max} 75 bar, T_{max} 240 °C, 10- or 6-position carousel. Vessels: 15, 30, 60 and 80 mL TFM, quartz inserts available. *T* monitored in each vessel, system monitoring (temp. and power) optional with RS-232 PC interface. Vessels have pressure release membrane, central gas collector.
CEM Corporation P.O. Box 200 Matthews, NC 28106 USA 704 821 7015 www.cemx.com	**MARS 5** Power 1500 W, dim. (w × d × h) 51 × 64 × 58 cm, wt 54 kg. P_{max} 100 bar, T_{max} 300 °C, 12-position carousel. Vessels: 100 mL, TFM or quartz. Optional *P* and *T* control. PC interface and software for data collection and applications library. Many accessories available.
Milestone Via Fatebenefratelli, 1/5 24010 Sorisole (BG) Italy 0039 3557 3857 [Milestone Inc., 160B Shelton Road, Monroe, CT 06468, USA 203 261 6175 www.milestonesci.com]	**Ethos 1600** Power 1000 W, dim. (w × d × h) 55 × 55 × 70 cm, wt 90 kg. P_{max} 100 bar, T_{max} 300 °C, 12-position carousel. Vessels: 120 mL TFM. Monobloc or segmented rotors hold vessels secure during whole process. *P* and *T* monitored, full system control with built-in PC and methods library and programming. Many unique features. Dual magnetron/premixing/rotating diffuser. Modular design allows flexible specification from models Ethos 900 to 1600. **UltraCLAVE** Power 1000 W (unpulsed), dim. (w × d × h) 60 × 105 × 190 cm, wt 250 kg. P_{max} 200 bar, T_{max} 350 °C. Large (3.5 L) chamber may be used as single sample vessel or to hold multiple vessels. Operates under high nitrogen pre-pressurization. Many unique features.
OI Analytical P.O. Box 9010 151 Graham Road College Station TX 77842-9010 USA 409 690 1711 www.oico.com	**Model 7295** Power 950 W, dim. (w × d × h) 57 × 46 × 34 cm, wt 27 kg. P_{max} 41 bar, T_{max} 200 °C, 12-position carousel. Vessels: 90 mL. *P* control, optional *T* control. Winwave software and applications library.
Plazamatronica Ul. Osbowicka 70 51-008 Wrocław Poland 48 7172 6666 parosa@manager.ae.wrocl.pl	**UniClever** Power 300 W, dim. (w × d × h) 20 × 40 × 60 cm, wt 20 kg. P_{max} 110 bar, T_{max} 270 °C. Vessels: 12 × 110 mL, TFM. *P* and *T* control. Triple safeguard pressure system; water-cooled vessels. Open digestion adapter.
Prolabo 54 Rue Roger Salengro F-94126 Fontenay-sous-Bois France 33 1 4514 8680 www.prolabo.fr	**MCS 950** Power 950 W, P_{max} 40 or 15 bar, T_{max} 300 °C, 12-position reversing carousel, variety of vessels available. *P* and *T* control. PC interface and software for data collection and applications library. Many accessories available, including remote control, exhaust module and central collector for released acid fumes.
Questron Technologies Corporation 6741 Columbus Road, Unit 12 Mississauga, Ontario L5T 2G9 Canada 905 564 8417 www.qtechcorp.com	**QLAB 6000** Power 1000 W, dim. (w × d × h) 58 × 46 × 48 cm, wt 56 kg. P_{max} 42 bar, T_{max} 230 °C, 12-position carousel. Vessels: 100 mL. *P* and *T* control, full system control with built-in PC and methods library. May be set for closed or open vessel operation or flow-through. Sophisticated exhaust system. Many unique features.

are one or two orders of magnitude more expensive than domestic microwave ovens but have many features which make them safer and more effective for use in chemical analysis and research. Laboratory ovens have electronic components that are protected from the corrosive gases that will often occupy the oven cavity and are designed to eliminate the risk of sparking by switches, etc., which could ignite flammable vapors in the cavity. The overall construction is more robust and extensive use is made of stainless-steel components where appropriate. Ventilation from the cavity is often a feature. The elimination or reduction of hotspots in the cavity is usually one of the design targets – some instruments are fitted with rotating diffusers in the roof of the cavity, most make use of a carousel or rotating platform to house the sample vessels. Most ovens use a pulsed system in which power adjustment is controlled by switching the magnetron on and off successively, with the more advanced instruments having a shorter pulse time. Other instruments do have an unpulsed and variable source of microwave power. Some manufacturers sell the ovens as stand-alone instruments whereas others sell a complete microwave sample treatment system, including oven, vessels and temperature and pressure monitors. Available accessories may include distillation equipment and facilities for handling open vessels where an efficient ventilation capability is essential. Some manufacturers either equip their ovens for carrying out Kjeldhal analyses, or may sell specialized ovens designed specifically for this purpose.

Computer control is available on most higher end instruments – this can include a library of stored protocols for a variety of sample types, full temperature and pressure predetermination with a printout of the whole experiment and even automatic sample and reagent dispensing.

One system now on the market (the QLAB 6000) enables users to carry out closed vessel, open vessel or discrete flow digestions all on the one instrument. In the discrete flow mode, each sample is pneumatically isolated and transported through a complete digestion cycle. The operator weighs and loads the samples and the instrument then completes the task – it adds reagents, agitates, transports the samples, digests, cools, rinses, dispenses, dilutes and cleans up ready for the next cycle, automatically. Other manufacturers will have similar systems available soon. A short review of this topic appeared recently.[39] A summary of some of the ovens and oven systems currently available is presented in Table 1.

3.2 Vessels

Different vessels are used for high-pressure dissolutions or reactions and those carried out under atmospheric pressure. In addition, specialized vessels are available for fusions, distillations or Kjeldhal analyses.

Vessels used for high-pressure work are usually made with a high-purity Teflon™ or PFA liner and an outer jacket of a stronger microwave transparent material (e.g. ULTEM, made by GE). Some kind of pressure-relief valve or membrane is usually fitted. These vessels normally have a capacity of 150 mL and can withstand pressures up to 40 bar (600 psi). The temperature is limited by the 260 °C softening point of Teflon™. Some manufacturers claim that these vessels can withstand much higher pressures, but at the high temperatures generated, it is probably wise to adopt a conservative approach. Closed vessel digestion is ideal for those samples which are being dissolved in HNO_3 or HCl, since the boiling point of the acids is raised under the high-pressure conditions and the digestion reactions are speeded up. In a closed vessel digestion, there is no loss of volatile elements and no reagent loss, hence smaller reagent quantities are needed. However, for those digestions where H_2SO_4 is required, such as Kjeldahl or digestions of petroleum products, there is little advantage in using the regular closed vessel approach. This is because the boiling point of H_2SO_4 is 330 °C, way above the temperature available in a Teflon™-lined vessel. Consequently, a higher temperature and faster reaction time is attainable in a glass or quartz vessel with the reaction run at atmospheric pressure.

Very high operating temperatures and pressures (200 bar/350 °C or 2900 psi/350 °C) are available with the ultraCLAVE by Milestone. This has an unusual design in which the reaction vessels are completely contained inside a high-pressure enclosure.

4 SAFETY CONSIDERATIONS

The main hazards associated with the laboratory use of microwave-assisted techniques can be divided into two areas:

- accidental exposure to microwave radiation
- explosions/fires/chemical spills or leakages.

Exposure to microwave radiation in small doses is not necessarily dangerous – in fact, it has been used for many years in the treatment of a variety of medical conditions.[40] The danger arises when the exposure is not detected. Human skin can readily detect heat radiation and if one accidentally comes into contact with a hot object, there is an automatic reaction to withdraw from the heat source. Any injury is immediately obvious and treatment may be applied. However, microwave radiation tends to penetrate the first layers of skin and tissue,

passing through them imperceptibly. The heating occurs at points deeper within the body or limb, in places where nerve ends are absent. Hence tissue can be heated to temperatures where serious damage may take place yet the victim remains unaware of the injuries that are being inflicted. In cases of high exposure, such injuries could lead to death, so avoidance of accidental exposure to microwave radiation is a crucial concern in all work associated with microwaves.

The dangers of chemical spills, fires and explosions are always present in the chemical laboratory when vigorous reactions take place, either by accident or design. So in principle, the risks involved in microwave-assisted chemical processes are not very different from those encountered in more traditional chemistry. The major difference is that fairly often, users are unfamiliar with the apparatus and protocols that they should be using in the microwave-assisted procedures which they are attempting. This can lead to the development of excessively high temperatures or pressures and the rupture or melting of reaction vessels, leading to accidents.

Microwave-assisted preparation of solid samples can occasionally cause electrical arcing inside the sample vessels when the sample is not submerged in the solvent.[40] This carries the potential danger of igniting a flammable solvent and causing a fire.

Another significant safety aspect is the maintenance of the microwave unit itself. Since microwave-assisted pretreatment of samples often involves the use of strong acids or organic solvents, most laboratory microwave units exhaust poisonous gases from the sample cavity. This must be kept separate from the air that cools the circuitry, or it may lead to deterioration of the insulating materials leading to electronic failure. Such an occurrence has been reported[41] and led to accidental exposure to high powered microwave radiation by an individual, when the safety power cut-off failed to function when the oven door was opened. Therefore, it is important not to store the entire microwave system under a fume hood, where corrosion of the electronics is almost certain to take place.

5 APPLICATIONS

5.1 Introduction and Review Articles

Microwave energy has a number of different applications in analytical chemistry:

- In electron spin resonance spectroscopy (ESR) [electron paramagnetic resonance spectroscopy (EPR)]. This technique, which has had a long history of development, uses microwave energy as the irradiation source and is widely used for structural analysis and in the elucidation of a variety of problems, particularly biochemical ones, involving species with unpaired electrons.
- In a microwave plasma (MP). These have been used as sources for atomic emission spectroscopy (AES) since the 1970s. Several common forms of this plasma source exist, including the microwave-induced plasma (MIP), the capacitatively coupled microwave plasma (CMP), the surface-wave or surfatron plasma, the microwave plasma torch (MPT), and some other unique designs. It is suggested that these MPs offer certain cost benefits in comparison with inductively coupled plasma (ICP) sources.[42]
- In microwave-assisted sample preparation techniques such as microwave-assisted digestion (MAD), microwave-assisted extraction (MAE), microwave-assisted fusion, microwave-assisted derivatization and microwave drying. It is the applications of these sample preparation techniques that form the focus of this study.

A number of books, book chapters and review articles have appeared recently. These provide a general coverage of the topic. Their contents are summarized below.

- *Microwave-enhanced Chemistry – Fundamentals, Sample Preparation and Applications*.[20] A wide-ranging survey of microwave-assisted chemistry. Sixteen chapters covering fundamental theory right through to applications in chemical synthesis and including sample preparation techniques.
- 'SamplePrep Web: Analytical Sample-preparation and Microwave-chemistry Resource Center'.[43] The authors have developed a web site [http://www.sampleprep.duq.edu/sampleprep], which is a resource for a wide range of protocols and ancillary information.
- 'Standardization of Sample Preparation for Trace Element Determination Through Microwave-enhanced Chemistry'.[44] The development of standard sample protocols using microwave digestion is discussed and a brief overview presented of some of the key steps in EPA method 3052. Advantages and disadvantages of the technique are outlined.
- 'Microwave Digestion Procedures for Environmental Matrices'.[45] A review of the applications of microwave energy to the digestion of environmental samples. Both open and closed microwave digestion methods are discussed. Trends in research are highlighted and universal digestion procedures for particular matrices identified.

- 'Sample Preparation Using MAD or Extraction Techniques'.[46] A review on microwave digestion or extraction as a sample preparation technique. Describes how the technique can be applied in both organic and inorganic determinations for various sample matrices.
- 'Microwave Analysis Systems for Rapid Process Analysis'.[47] The determination of moisture, solids, ash and fat contents are summarized and the speed of determinations using microwave systems is discussed, with particular reference to the CEM LabWave 9000 drying apparatus and the CEM MAS 7000 and AirWave 7000 microwave ovens. The replacement of wet ashing by microwave ashing for the determination of sulfate ash is recommended. Ashing times for a range of samples with a conventional and a microwave muffle oven are tabulated.
- 'Microwave Processing of Ceramics'.[48] This review is focused on the microwave processing of ceramics, which is fast emerging as a new field in material synthesis. Recent times have witnessed significant progress in the aspect of commercialization and application of the technology to new areas. The most significant developments have been the use of microwaves in the sintering of nonoxides, such as tungsten carbide-based components and powder metals, fabrication of transparent ceramics and the design of continuous microwave systems. While not covering analytical applications, this article will be of interest to those planning to use microwave fusion in sample preparation.
- 'Microwave-assisted Sample Preparation in Analytical Chemistry'.[49] A comprehensive review giving a broad overview of the topic and details of protocols used in the preparation of a wide variety of sample types.

5.2 Environmental Samples

This section is fairly broad and includes examples of treating soil samples and sewage sludge samples and related materials. There is some overlap between this and the next section, which deals with biological and clinical samples.

- 'A Study of the Analytical Parameters Important for the Sequential Extraction Procedure Using Microwave Heating for Pb, Zn and Cu in Calcareous Soils'.[50] This is a report of a four-step microwave extraction procedure for Pb, Zn and Cu fractionation for calcareous soils. For 1 g of soil with grain size $< 50\,\mu m$, 20 mL of 1 M NaOAc at pH 5.0 was used for extraction of the fraction bound to the carbonate in the first step; 20 mL of 0.072 M $NH_2OH \cdot HCl$ in HOAc for the fraction bound to iron and manganese oxides in the second step; and HNO_3 (0.02 M) + NH_4OAc (0.8 M) for the fraction bound to organic matter in the third step. Finally, the residue was dissolved with $HClO_4$ + HF, then diluted to 25 mL with HCl (1 M) for analysis of the residual fraction in the fourth step. Metals were determined by flame atomic absorption spectroscopy (FAAS). The validity of the method was verified by a mass balance using calcareous soils and two certified reference materials. The results obtained were appropriate for the assessment of the environmental risks.
- 'Application of Microwave Extraction for Partitioning of Heavy Metals in Sewage Sludge'.[51] Describes the application of microwave extraction for partitioning of heavy metals (Cu, Cr, Ni, Pb and Zn) in sewage sludge. The procedure used follows the four-stage Tessier sequential extraction method but with conventional heating replaced by microwave heating and elimination of the shaking operation. A significant time saving (more than 10 h) was obtained by this microwave-assisted Tessier extraction with comparable efficiencies to the conventional Tessier method.
- 'Elemental Speciation and Coupled Techniques – Towards Faster and Reliable Analyses'.[52] Procedures for simplifying and shortening sample preparation protocols in environmental speciation analysis are discussed. Advances in gas chromatography (GC) sample introduction for plasma source atomic spectrometry and sample preparation methods for multicapillary GC, including microwave-assisted solvent extraction (MASE) and microwave-assisted purge-and-trap techniques are considered. Examples given of the speciation of organometallic compounds in biological tissues and road dust by GC/MIP/AES and of cobalamins by high-performance liquid chromatography (HPLC)/inductively coupled plasma mass spectrometry (ICPMS). Integration of sample preparation and chromatographic separation into a single step is evaluated. The importance of multidimensional separation and detection techniques for the unambiguous determination of metallocompounds in biological species is discussed.
- 'Microwave-assisted Distillation in Tritium Monitoring (in Japanese)'.[53] Describes a simple microwave-heating distillation device for the rapid preparation of natural samples for environmental tritium (T) monitoring. The device consists of a hard-glass test-tube, an N_2-gas flow piping system, a cold trap and a microwave oscillator (2.45 GHz, 250 W max.). The T concentration in H_2O in pine needles was determined by distillation of a 13-g sample in the

test-tube at 90 °C (200 W) for 3 min under an N_2-gas flow-rate of 3 L min^{-1}. The H_2O was collected by a conventional low-background scintillation apparatus. T concentrations of the needles collected in the vicinity of an organic IR waste disposal facility were 5.0 Bq L^{-1} before burning and 7.9 Bq L^{-1} after burning operation.

- 'Continuous Microwave Assisted Pervaporation/ Atomic Fluorescence Detection: an Approach for Speciation in Solid Samples'.[54] This report describes a method for the speciation of mercury in solid samples. It is based on microwave-assisted pervaporation with atomic fluorescence detection in a continuous flow system. The sample was placed in a laboratory-made cell, and either $SnCl_2$ [for $Hg(NO_3)_2$] or an oxidizing solution followed by $SnCl_2$ (for PhHgOAc) were injected into the sample via a syringe. The cell was then placed inside a microwave device so that the microwaves were focused only on the sample chamber. After application of the microwaves, the flow of gas or liquid in the upper acceptor chamber was started, and the Hg driven to the detector via a gas–liquid separator. Acceptable results were obtained for the speciation of mercury in a certified reference material.
- 'Stability of Organic Pollutants During MAE from Solid Matrices'.[55] Dry and wet soil samples (5 g) were dispersed in solvent (30 mL) and exposed to MASE at 100% power for 5 or 20 min at 50 or 145 °C. The solvents were hexane–acetone (1 : 1), methylene chloride–acetone (1 : 1), toluene–methanol (10 : 1) and methyl *tert*-butyl ether. The samples were spiked with a solution containing 95 semi-volatile organic compounds (VOCs) including neutral (34), polycyclic aromatic hydrocarbon (PAH) (22), basic (19) and phenolic (19) compounds and benzoic acid, prior to extraction. The extracts were concentrated under N_2 and analyzed by gas chromatography/mass spectrometry (GC/MS). Recoveries and correction for losses during concentration were obtained by comparing the results with those for extracts of spiked samples not subjected to MASE. Mean recoveries and 95% confidence intervals are reported. Hexane–acetone gave recoveries of >80% except for basic compounds and benzoic acid in dry soil. An extraction time of 20 min did not significantly increase recoveries, and an extraction temperature of 145 °C decreased the recoveries of basic compounds. Recoveries of basic compounds, benzoic acid and some phenolic compounds were higher for wet soil than dry.
- 'Open Focused-microwave-assisted Sample Preparation Procedures: Fundamentals and Application to the Speciation of Tin and Mercury in Environmental Samples'.[56] An attempt was made to investigate some of the fundamental principles underlying microwave digestion by measuring the permittivity changes of solutions during digestion. The experiments revealed a decrease in permittivity with increasing microwave heating time. The microwave-assisted leaching of organotin compounds and Hg species from sediments and the microwave-assisted solubilization of organotin compounds and Hg species from biological materials were also studied. Results of work recently carried out by the authors are compared with those obtained using extraction/determination procedures reported in the literature since 1989. The development and potential applications of open microwave sample preparation procedures are also discussed.
- 'Analysis of Trace Elements in Coal Fly Ash and Their Leachates: Results of the ENEL–EDF Round-robin Test'.[57] Eight Italian laboratories and four French laboratories participated in a round-robin test of the determination of trace elements (As, Cd, Cr, Cu, Hg, Ni, Pb, Sb, Se, V and Zn) in coal fly ash and its leachates. The analysis procedure consisted of MAD of fly ash samples, followed by FAAS, electrothermal atomic absorption spectroscopy (ETAAS) or inductively coupled plasma atomic emission spectroscopy (ICPAES) of the digests. Acetic acid and water leachates of fly ash were also analyzed. The accuracy and precision of the results were acceptable, except for Hg and Sb. The greatest source of variability in the fly ash analysis was the digestion step, because of the possibility of incomplete digestion of the matrix. For the leaching tests, the greatest source of variability was the pH of the leaching solution, as the metals' solubilities varied with pH, and therefore variations in pH could affect the final concentrations of the metals in the leaching solutions.
- 'A Rapid Microwave-digestion Method for Colorimetric Measurement of Soil Organic Carbon'.[58] Ground (<0.2 mm), air-dried and oven-dried soil (0.2 g) and 0.5-mL portions of standard solutions containing 0–10 mg of sucrose-C were treated with 5 mL each of 0.17 M $K_2Cr_2O_7$ and H_2SO_4, and the mixtures were heated in a microwave oven to an applied energy of 500 J mL^{-1}. The cooled mixtures were diluted to 30 mL with H_2O, centrifuged at 5000 rpm for 5 min and the absorbances measured at 590 nm. The recovery of organic C relative to the result obtained by the dry-combustion method with use of a LECO analyzer was 91.7%. The proposed method is more rapid and precise, involves the use of smaller volumes of reagents and results

in less waste than the conventional wet-digestion method.

- 'Voltammetric Procedure for Determination of Mercury in Soil Wastes After Wet Microwave Digestion'.[59] Describes how the determination of mercury in solutions after microwave digestion of solid waste sample can be performed by using differential pulse anodic stripping voltammetry (DPASV) in a flow-through system with a gold disk as the working electrode.
- 'Automatic Flow-injection System for the Determination of Heavy Metals in Sewage Sludge by Microwave Digestion and Detection by Inductively Coupled Plasma Atomic Emission Spectroscopy (MW–ICPAES)'.[60] Results from the optimization of an automatic flow injection system combining MAD with atomic spectrometric detection (FAAS, ICPAES) to determine heavy metals in sewage sludge are presented. Digestion was performed by preparing a suspension of the sample in 1.5 M HNO_3 and making it flow through a PTFE capillary tube placed inside a conventional microwave oven. The effects of capillary tube length and inner diameter and pumping rate were studied to establish the experimental conditions that allow a quantitative elemental recovery in the shortest possible time. The elements determined were Zn, Cu, Pb, Cd, Ni, and Cr. The optimized procedure was applied to one ordinary and one certified sewage sludge sample. In comparison with conventional methods of sewage sludge analysis, the proposed method was less time consuming and equally precise and accurate.
- 'Microwave-assisted Pre-treatment of Environmental Samples for the Determination of ^{90}Sr'.[61] In this paper, microwave-assisted pretreatment of environmental samples for the determination of ^{90}Sr are presented. Six matrices were studied: organic and mineral soils, sediment, grass and ashes from grass and milk samples. Total attack (using HNO_3, $HClO_4$, HF and H_2O_2) and leaching digestion methods, which avoided the use of HF, were optimized for each matrix. Since no statistically significant differences were found between the results obtained using the two procedures, leaching digestion was preferred, as long as the digestion time and the number and amount of reagent were lower. The influence of sample type on ^{90}Sr activity concentration determination was studied and no significant influence was observed.
- '^{210}Po and ^{210}Pb Analysis in Sediments and Soils by Microwave Acid Digestion'.[62] A microwave acid digestion method prior to the determination of ^{210}Pb and ^{210}Po in sediments and soils is described. It involves an acid (HNO_3, HCl, HF and HBO_3 mixture) digestion with microwave heating in closed vessels at high pressures. Analyses carried out for various reference materials showed that the results were statistically the same as the certified values and the precision was good. The advantage of the microwave technique compared with the traditional leaching procedures is that the solid materials are completely dissolved and, therefore, almost 100% efficiency is achieved in the extraction of ^{210}Po and ^{210}Pb, even though a fraction is tightly bound to the silica matrix. Moreover, the time of analysis is drastically reduced, as are the risks associated with vapor inhalation and material corrosion.
- 'An Optimized Microwave Digestion Procedure for Cadmium Analysis of Mussel Samples'.[63] Mussel (0.2 g dried sample) was treated with concentrated HNO_3 in a closed vessel by microwave heating. After cooling, the resulting solution was adjusted to pH 5 with NH_3 and concentrated on-line with a chelating resin. The cadmium was determined by AAS (atomic absorption spectroscopy). The analytical results agreed well with certified values for a mussel standard reference material.
- 'Microwave-assisted Soil and Waste Dissolution for Estimation of Total Phosphorus'.[64] For the determination of P by ICPAES, the United States Environmental Protection Agency (USEPA) 3051(SW-486) microwave-digestion method was more effective and convenient, less labor intensive and safer than digestion with HNO_3–$HClO_4$ in an aluminum block for the decomposition of soil, sediments, waste solids, sewage sludge and compost. The recovery and precision can be improved further for samples containing large amounts of organic matter by dry ashing before the microwave digestion. Recoveries of >93.5% were obtained.
- 'Microwave Digestion for Analysis of Metals in Soil'.[65] The effects of H_2O_2 addition to HNO_3 for the microwave digestion of soil on the analyses of 10 heavy metals by ICP were investigated. For HNO_3-only digestion (method A), based on USEPA method 3051, organic soil (200 mg, or 500 mg for mineral soil) was digested with 10 mL of concentrated HNO_3 in a closed vessel (heating at 180 °C for 10 min, taking 20 min to reach 180 °C). On cooling, the filtration, dilution and analytical procedures were followed. For method B, 2 mL of H_2O_2 were added to the digestion vessel after the HNO_3 digestion cycle and the microwave digestion repeated with the same program as for A, and with identical filtration, dilution and analytical procedures. The results showed that metal concentrations were greater in organic than mineral soils and addition of H_2O_2

to HNO_3 did not enhance metal recovery from soil compared with HNO_3 alone except in the case of Mn.

- 'Sample Preparation for Trace Metals. Part II: Microwave Digestion'.[66] Microwave digestion as a sample preparation tool for trace metal environmental analysis is discussed with a focus on contamination issues.
- 'Comparison of Microwave-assisted Leaching Techniques for the Determination of Heavy Metals in Sediments, Soils, and Sludges'.[67] Microwave-assisted USEPA method 3051 for the HNO_3 leaching of environmentally key elements from sediments, soils and sludges was tested, and the influence of leaching temperature and time on element recovery for an estuarine sediment was investigated. The leaching efficiencies for four certified reference materials applying USEPA method 3051, an optimized HNO_3 procedure and an aqua regia [$HCl–HNO_3$ (3 : 1)] procedure were compared. Digestions were carried out in a high-pressure microwave system offering simultaneous temperature and pressure control for all digestion vessels employed. Eight elements (Cd, Co, Cr, Cu, Hg, Ni, Pb and Zn) were determined by ICPMS. The extraction efficiency strongly depended on the applied leaching parameters and varied for certain elements among different materials when an HNO_3 procedure was applied. In general, element recoveries obtained from the aqua regia procedure were superior to those obtained from HNO_3 procedures and showed good agreement (95% confidence interval) with the certified value for most of the elements investigated.
- 'Use of Microwave Digestion for Estimation of Heavy Metal Content of Soils in a Geochemical Survey'.[68] Soil (0.5 g) was microwave digested for 30 min in a mixture of 10 mL of H_2O, 5 mL of HNO_3, 4 mL of HF and 1 mL of HCl, and the resultant digest was treated with 2 g of boric acid to complex excess fluoride as tetrafluoroborate. Cu, Ni, Zn, Cd, Cr and Pb were determined with FAAS or ETAAS.
- 'Microwave Digestion of Ancient Peat and Determination of Pb by Voltammetry'.[69] Different acid compositions (HNO_3, H_2O_2, HF and $HClO_4$) and low-pressure microwave digestion were applied to the microwave-assisted dissolution of ancient peats. The digests were evaluated with respect to the decomposition of the inorganic and organic fraction of the peats and to an optimized determination of Pb by DPASV. Addition of HF was necessary to achieve a complete dissolution of the resistant inorganic phase of the peat and to obtain total Pb recovery. Addition of $HClO_4$ decreased the concentration of dissolved organic C significantly, allowing reliable voltammetric signals for Pb determinations with appropriate peak shape, baseline, sensitivity and precision.
- 'Optimization of the Reduction of Selenium(VI) to Selenium(IV) in a Microwave Oven'.[70] Water samples were made 10 mM in HCl and agitated ultrasonically to remove CO_2. Portions were analyzed directly to determine Se(IV). Further portions (25 mL) were transferred to PTFE vessels, made 4 mM in HCl and the vessels were capped and heated in a microwave oven at 600 W for 3 min, giving 100% reduction of Se(VI) to Se(IV). The total Se(IV) was then determined. The samples were pumped at 7 mL min^{-1} and mixed with 1% $NaBH_4$ in 1% NaOH and with 8 M HCl (both at 1 mL min^{-1}). The H_2Se was determined by AAS. Selenomethionine and selenocystine gave partial and variable yields of H_2Se under these conditions and needed to be separated, e.g. by cation-exchange treatment. The method was applied to the analysis of six samples of mineral water, for which the prior removal of high CO_2 concentrations, as described, was essential.
- 'Levels of Organochlorines in Penguin and Skua Eggs from the Antarctic. Determination after Application of Focused Open-vessel MAE in Combination with Gel Permeation Chromatography (GPC)'.[71] Organochlorines in penguin and skua eggs from the Antarctic were determined after application of focused open-vessel MAE in combination with GPC. Extraction of the entire, partly lyophilized eggs was performed with ethyl acetate–cyclohexane. It was shown to be an excellent method for the quantitative determination of organochlorines in the parts per billion to trillion range because it allows the use of large quantities of sample.
- 'Combined MAE and GPC for the Determination of Chlorinated Hydrocarbons in Seal Blubber and Cod Livers'.[72] A fast and effective sample clean-up procedure for the quantification of chlorinated hydrocarbons in seal blubber and cod livers is reported. Lipophilic sample ingredients were extracted by application of microwave energy. MAE was performed with ethyl acetate–cyclohexane as solvent. Without exchange of the solvent, the organochlorine compounds were separated from matrix co-extractives by GPC. Traces of matrix remainders were separated on deactivated silica prior to gas chromatography/electron capture detection (GC/ECD). The polychlorinated biphenyl (PCB) compounds 153, PCB 138 and PCB 180, hexachlorobenzene (HCB) and *p*, *p*′-dichloro-di(chloro)phenyl-ethylene were targets for quantification. In seal blubber, the recovery rates for these organochlorines were >90% following the complete sample clean-up procedure. MAE of cod

livers required milder conditions. After optimization of the MAE conditions, the organochlorine levels in cod livers were almost identical with those determined with other independent clean-up methods.

- 'Fast Extraction of Dioxins from Fly Ash. Comparison of Soxhlet, SFE and Microwave-assisted Soxhlet Extraction'.[73] MASE was tested for the extraction of dioxins from waste incinerator fly ash, for which SFE is sometimes not effective. SFE can be considered as an alternative to Soxhlet extraction for low C-level fly ash, but when the percentage of C is high (about 15%), MASE may be preferable.
- 'Extraction of Hexaconazole from Weathered Soils: a Comparison Between Soxhlet Extraction, MAE, SFE and ASE'.[74] Extraction of hexaconazole residues, weathered over periods of time varying between 0 and 52 weeks, was undertaken from two characterized soils. The various extraction techniques considered were Soxhlet extraction, MAE, SFE and ASE. In this case, the best precision was obtained for the automated ASE system.
- 'Pesticide Analysis without Sample Transfer'.[75] Problems in Soxhlet extraction are discussed, and the advantages of microwave extraction outlined, e.g. extraction of 12 samples takes only 60 min including sample preparation and samples from a few milligrams up to 500 g can be handled. The easy WAVE software for controlling the heating program and the filtEX glass filtration and drying unit are briefly described. For the extraction of 2,4-dichlorophenoxyacetic acid, 2,4,5-trichlorophenoxyacetic acid and 2-(2,4,5-trichlorophenoxy)propionic acid from soil, a 2-g sample was extracted with 10 mL of methanol at 120 °C and 300 W for 15 min in a filtEX unit, which allowed subsequent filtration and concentration of the extract without transfer before analysis by HPLC. The determination of hexachlorocyclohexane and other polychlorinated compounds was carried out by GC/ECD after extraction in a filtEX unit.
- 'Extraction of Organic Substances with Solvents'.[76] MAE was carried out under pressure in an MLS-ETHOS apparatus equipped with Teflon™ vessels, which allowed nonpolar solvents to be heated by microwaves. The vessels contained magnetic stirring bars and were purged with an inert gas. The system was used for the extraction of 5-g soil samples with hexane–acetone (1 : 1); up to 12 sample holders could be accommodated in the pressurized rotor and extraction at 500 W and 115 °C with stirring took 10 min. The extracts could be concentrated by evaporation without transfer. Reproducibility was better and extraction was 20–30% faster with stirred than with unstirred samples. Recoveries of, e.g. acetophenone, aniline derivatives, phenol derivatives and lindane by microwave (with and without stirring), Soxhlet and ultrasonic extraction are tabulated.
- 'Microwave vs. Soxhlet for the Extraction of PCDD (polychlorinated dibenzo-*p*-dioxin) and PCDF (polychlorinated dibenzofuran) from Sewage Sludge Samples'.[77] Dried, ground, sewage sludge (10 g) was added to 0.3 g of H_2O and equilibrated for 10 min. A standard solution containing 15 ^{13}C-labeled 2,3,7,8-substituted isomers of PCDD and PCDF was added, followed by 30 mL of toluene. Microwave extraction was carried out at 15, 20, 25 and 30% using a Soxwave 100 (Prolabo). Soxhlet extraction was carried out for comparison, for 48 h with 300 mL of toluene in the presence of 5 g of copper to remove sulfur interferences. Quantitative determination was carried out by isotope dilution. Recoveries ranged from 30 to 70% for the 15 analytes and two recovery standards. The advantages of MAE include reduced extraction time and reduced solvent use. The method is proposed as a viable alternative to Soxhlet extraction.
- 'Focused MAE (FMAE) of Polynuclear Aromatic Hydrocarbons from Contaminated Soil: Role of Acetone and Water Content Impact on Microwave Efficiency'.[78] Portions (5 g) of dried soil were mixed with 1-mL aliquots of H_2O, transferred to a quartz container and microwave extracted with 40 mL of CH_2Cl_2–acetone (1 : 1) at 30 W for 10 min. The recovery of 15 PAHs (0.07–2.3 ppm), in comparison with 8-h extraction with 220 mL of anhydrous CH_2Cl_2 in a Soxhlet apparatus, taken as 100%, was 70.8–128.1% (lowest for naphthalene) and the RSDs were 9.6–17.2% (mean 13.1%) compared with 8.1–36% (mean 22.7%) for the Soxhlet method ($n = 5$).
- 'A Comparison of MAE and Soxhlet Extraction for Phenols in Soil Samples Using an Experimental Design'.[79] A comparative study of MAE and Soxhlet extraction for the separation of five phenols from soil is reported. Factorial and central composite designs together with the simplex method were used to optimize the extraction conditions for both techniques. For phenol, 2-chlorophenol and 2,4-dichlorophenol, comparable recoveries were obtained by both methods. For 2-ethylphenol and 2-nitrophenol, MAE gave better recoveries than Soxhlet extraction. The RSDs were 7.5–31% and 2.5–30% for MAE and Soxhlet extraction, respectively.
- 'The Evaluation of Rapid Solid Sample Extraction Using the Microwave-assisted Process (MAP) under Closed-vessel Conditions'.[80] Samples of soil or

sediment (1–5) were microwave extracted in a closed vessel with 30 mL of hexane–acetone (1 : 1) at 115 °C for 10 min. After cooling, the extract was dried by filtration through a bed of Na_2SO_4. The filtrate was mixed with 1 mL of toluene, evaporated to 1–2 mL under vacuum and then diluted. The extract was subjected to clean-up if required. The prepared sample (1 μL) was then analyzed by GC/MS, GC/ECD or gas chromatography/flame ionization detection (GC/FID). Recoveries of PAHs, PCBs, base-neutrals, chlorinated pesticides and substituted phenols were >80% and RSDs were generally of the order of 10%. The method was assessed by the use of standard reference materials, and by comparing results with those obtained by a contract laboratory.

- 'Abbreviated Microwave Extraction of Pesticides and PCBs in Soil'.[81] Samples of soil material were microwave extracted with 8 mL of hexane–acetone (1 : 1) in a closed vessel. The microwave oven was operated at full power and samples maintained at 115 °C for 15 min. After cooling, the extracts were analysed by GC/ECD. Recoveries of pesticides and PCBs were comparable to those obtained by sonication or Soxhlet extraction but the microwave experiments took less than half the time required by other methods and consumed 12–30 times less solvent.
- 'GC–Atomic-emission Determination of Organotin Compounds in Marine Biomaterials After Microwave Extraction'.[82] Marine biological samples (0.2–1 g) were microwave extracted with 15 mL of tetramethylammonium hydroxide in a closed vessel. The mixture was irradiated with microwaves until the temperature reached 120 °C and maintained at this temperature for 3 min. After cooling, the mixture was adjusted to pH 5 with 3.8 mL of glacial acetic acid and 5 mL of buffer, and 1 mL of isooctane and 2 mL of 1% sodium tetraethylborate were added. The mixture was again irradiated with microwaves and the temperature was held at 90 °C for 2 min. The isooctane phase was then purified on alumina and analyzed on a column (30 m × 0.25 mm ID) coated with DB-5 (0.25 μm) and operated with temperature programming from 80 °C (held for 1 min) to 260 °C (held for 10 min) at 20 °C min^{-1}, He as carrier gas (20 psi; 9 mL min^{-1}) followed by microwave-induced plasma atomic emission spectroscopy (MIPAES) detection of organotin compounds at 303.419 nm. Recoveries were >80% for 200–500-mg samples; quantification limits using 500-mg samples were ~10 ng g^{-1} (of tin) for monobutyltin trichloride, dibutyltin dichloride, tributyltin chloride and triphenyltin chloride.
- 'Focused MAE of PAHs in Open Cells from Reference Materials (Sediment, Soil, Air Particulates)'.[83] Samples (0.1–1 g) of freeze-dried sediment, soil or airborne particulates were microwave extracted with 30 mL of dichloromethane using an open-vessel focused-microwave system. The microwave oven was operated at 30 W and the extraction time was 10 min. The extract was desulfurized on activated Cu, then purified on an alumina microcolumn (elution with dichloromethane) and fractionated on a silica microcolumn. The analytes were determined by GC/MS. Recoveries were generally >80% and RSDs were generally <10%. The recoveries were compared with those obtained by Soxhlet extraction; where reference materials were involved, the results were also compared with the certified values.

5.3 Biological and Clinical Samples

There is some overlap between this and the previous section, since biological samples are often analyzed to assess environmental pollution levels, and some clinical samples are used for monitoring exposure to environmental pollutants.

- 'A Method for Analysis of Exhaled Air by Microwave Energy Desorption Coupled with GC–Flame Ionization Detection–Mass Spectrometry'.[84] A method for the chemical analysis of volatile constituents of the exhaled air of mechanically ventilated patients is described. Exhaled substances were adsorbed and concentrated on to activated charcoal, desorbed by microwave energy and transferred into a gas chromatograph for separation without prior cryofocusing. Substances were identified by flame ionization detection and mass spectrometry. This method gave reproducible results and appears well suited for clinical studies.
- 'Microwave-assisted Vapour-phase Nitric Acid Digestion of Small Biological Samples for ICP Spectrometry'.[85] The sample digestion container developed earlier for use with a commercially available high-temperature–high-pressure asher was modified to fit into the digestion vessel of either a PMD, QWave 3000 or a Floyd-RMS150 microwave decomposition system. The modified container was constructed from quartz, and had an effective volume of 1.8 mL. Typically 50–100 mg of powdered sample were placed in the container and this was moistened with 250 μL of H_2O. Then the container was placed in the microwave digestion vessel containing 2 mL (for the PMD) system) or 10 mL (for other microwave digestion system) of sub-boiling HNO_3, and digestion was carried out at 450 W and 1200 psi, 75 W and

400–480 psi or 595 W and 140 psi with one or other of the systems. Digestion blanks were lower than those obtained by conventional microwave digestion as a result of the concurrent distillation of the HNO_3 during the digestion process. Results for Ca, Cd, Cu, Fe, Mn, Mo, Pb, Zn, As, Al and Cr in various standard reference materials, e.g. National Institute of Standards and Technology (NIST) bovine liver and apple leaves, International Atomic Energy Agency (IAEA) horse kidney, National Research Council of Canada (NRCC) lobster hepatopancreas and National Institute for Environmental Studies (Japan) (NIES) mussel, were in good agreement with the certified values.

- 'The Importance of Vessel Materials for Ultratrace Analysis (Part 2)'.[86] Gallstones were subjected to microwave digestion at 220 °C and ≤80 bar in HNO_3 (1:1)–30% H_2O_2 (2:1) in a vessel with a Hostaflon TFM PTFE liner. Waste water and soil were decomposed similarly in 65% HNO_3 at 170 °C and ≤8 bar and 180 °C and ≤4.5 bar, respectively, in a vessel with a polyfluoroalkoxy (PFA) liner. Raster electron microscopy shows that the PFA surfaces were virtually unaltered after 100 digestion cycles whereas the Hostaflon TFM surface was damaged and made porous after 45 digestion cycles. Hence PFA but not Hostaflon TFM liners should be used in the microwave digestion of samples before ultratrace analysis.
- 'Voltammetric Analysis of Biological Samples and Soil After Wet Digestion Using Microwave Irradiation'.[87] The use of low-pressure CEM vessels for determining Cd, Pb, Cu and Zn in trees and soil by DPASV and testing the effectiveness of mineralization for determining the Al content in trees by absorptive cathodic voltammetry using Alizarin S as complexing agent are described.
- 'Microwave-accelerated Sample Preparation for Speciation Analysis of Organotin Compounds in Biomaterials'.[88] One-stage (1S) and two-stage (2S) procedures are described for the preparation of biological materials for speciation analysis of organotin compounds by GC. For the 1S procedure, 0.1–0.2 g of lyophilized tissue (1–2 g of wet tissue) was microwave digested at 40 W power for 3 min at a temperature of 130 °C with 5 mL of acetic acid, 1 mL of nonane and 3 mL of 2% sodium tetraethylborate ($NaBEt_4$). The organic supernatant was analysed by gas chromatography/atomic emission spectroscopy (GC/AES) or gas chromatography/flame photometric detection (GC/FPD) after purification on alumina with diethyl ether as eluent. For the 2S procedure, 0.1–0.2 g of lyophilized tissue was microwave digested at 50–60 W power for 3 min with 5 mL of aqueous 25% tetramethylammonium hydroxide. The solution was diluted to 15 mL with H_2O, adjusted to pH 5 with acetic acid and extracted with 1 mL of 2% $NaBEt_4$ and 1 mL of nonane for 5 min. The organic phase was separated and analyzed for tin species as described. The 1S procedure was applied to the analysis of mussel and sea urchin eggs.
- 'Determination of Aluminium in Serum by Electrothermal Atomic Absorption Spectrometry with Prior Sample Pretreatment by Microwave Digestion and SPE'.[89] Serum (0.5 mL) was mixed with 0.5 mL of concentrated HNO_3 and 0.2 mL of concentrated $HClO_4$ (both purified by sub-boiling distillation) in 7-mL PTFE/PFA phials. Three of the sealed phials were mounted on top of each other in twelve 120-mL digestion vessels and the 36 phials were simultaneously subjected to microwave digestion. The solutions were evaporated under an IR lamp. The residues were dissolved in 1 mL of 0.1 M HCl, treated with 0.1 mL of 1% 8-quinolinol in dilute HCl and adjusted to pH 9 with ammonia. Each solution was applied to a short column of 0.5 g of XAD-4 resin (20–50 mesh), the resin bed was washed with 5 mL of H_2O and the Al(III) was eluted with 1 mL of 0.1 M HCl at 0.25 mL min^{-1}. The eluates were analyzed by graphite furnace AAS.
- 'Focused MAD of Biological Reference Materials for the Determination of Trace Metals by ICPMS'.[90] A rapid method for the preparation of biological materials based on their digestion in a low-power focused microwave system prior to ICPMS was developed. The samples were decomposed using a three-step attack with 30% nitric acid at (a) 30 W for 5 min, (b) 50 W for 6 min and (c) 80 W for 6 min. This was followed by cooling the mixture and subsequently reacting it with a 15% solution of hydrogen peroxide at 30 W for 5 min. The resultant solution was diluted with 2% nitric acid and aspirated to the nebulizer. Eight elements (Mg, Mn, Ni, Cu, Zn, As, Sr and Cd) were determined simultaneously using the peak hopping mode and external calibration of the ICPMS system. The method was validated by the analysis of four reference materials.
- 'Low Volume Microwave Digestion for the Determination of Selenium in Marine Biological Tissue by Graphite Furnace AAS'.[91] An MAD procedure for determining Se in marine biological tissues was developed using three certified reference materials containing 1.62–6.88 μg g^{-1} Se as model samples. Quantitative recoveries were obtained with a power/time microwave digestion program consisting of 600 W for 2 min, 0 W for 2 min and 450 W

for 45 min. The digestions were performed in sealed PTFE vials (7-mL capacity) loaded with 0.025–0.1 g of freeze dried tissue and 1 mL of HNO_3. Recovery tests were carried out by spiking oyster, dogfish and lobster samples with six species of selenium (selenite, selenate, selenomethionine, selenocysteine, selenocystamine and trimethylselenonium). These gave quantitative recoveries (94–105%) for all the species apart from trimethylselenonium (90–101%).

- 'Microwave Radiation as a Factor Intensifying Sample Preparation: Analysis of Samples with an Organic Matrix'.[92] The necessity of using a controlled-pressure microwave system (e.g. the MDS-81D, CEM) for the safe and rapid decomposition of organic matrices with the simplest reagent mixtures is emphasized. Conditions were established for the decomposition of vegetable materials (leaves, grass, cereals and foods), vegetable oils, cosmetics (shampoo, cosmetic milk and cream, lipstick and hair gel) and biological fluids (blood and serum) for the determination of trace elements by AAS and ICPAES.
- 'Online Microwave Slurry Sample Digestion Using Flow Systems for the Spectrophotometric Determination of Iron in Seafood'.[93] Lyophilized, powdered samples (250 mg) were ultrasonicated with HNO_3–H_2O_2 (oyster) or aqua regia–H_2O_2 (mussel, fish). A portion (200 µL) of the resulting slurry was injected into a flow system and carried by an air flow (8 mL min^{-1}) to a 2-m PTFE coil for microwave irradiation at 700 W for 10–20 min. After digestion, the slurry was eluted from the flow system with 14 mM HNO_3 (oyster) or 0.15 M HNO_3–0.35 M HCl (mussel, fish) and collected in a final volume of 10 mL (5 mL for fish). A portion (200 µL for oyster and fish; 130 µL for mussel) of the resulting solution was injected into a carrier stream (2 mL min^{-1}) of 14 mM HNO_3 (oyster) or 0.15 M HNO_3–0.35 M HCl (mussel, fish) and mixed successively with streams (all 1 mL min^{-1}) of 1% ascorbic acid, 0.25% 1,10-phenanthroline and acetate buffer to give a final pH of 3.7. Finally, the absorbance was measured at 512 nm.
- 'The Optimization of Microwave Digestion for Determination of Selenium in Human Urine by Flow Injection–Hydride Generation–Atomic Absorption Spectrometry'.[94] A microwave digestion program designed to decompose and oxidize completely selenium compounds in urine to selenate is reported. This involves monitoring the pressure and the temperature during microwave digestion. The complete decomposition and quantitative recovery of trimethylselenonium iodide spiked into urine was achieved in 18 min using an optimized microwave program reaching 200 °C and 8 bar. The selenate in the digest was reduced to selenite by hydrochloric acid with the aid of microwave energy. Selenomethionine and selenoethionine were found to be unstable during the microwave heating used to reduce selenate to selenite. It is suggested that such a microwave reduction procedure could be used to distinguish selenate from selenite in those matrices which might contain organic selenium compounds.
- 'Spectrophotometric Determination of Pb in Urine with Microwave Digestion and a Dithizone–Water Phase'.[95] A method for the spectrophotometric determination of Pb in urine with microwave digestion and a dithizone–H_2O phase is described. A complex is formed by Pb^{2+} and dithizone in the pH 8.5–9.0 H_2O phase with Tween-20 as solubilizing agent.
- 'Determination of Five Selenium Compounds in Urine by Liquid Chromatography with Focused MAD and Hydride Generation–Atomic Absorption Spectrometric Detection'.[96] A method is proposed for the determination of trimethylselenonium ion, selenomethionine, selenocystine, selenite and selenate in urine, following purification with C_{18} cartridges. The on-line system consists of an anion-exchange chromatographic column for species separation, a focused microwave oven for the oxidation of organic Se species in the presence of 3% $K_2S_2O_8$ and 3% NaOH solution and reduction of selenate to selenite in 10 M HCl.
- 'Determination of Mercury Species in Fish Reference Materials by Isothermal Multicapillary GC with Atomic Emission Detection after Microwave-assisted Solubilization and MASE'.[97] Samples of freeze-dried tissue (0.1–0.2 g) were mixed with tetramethylammonium hydroxide (5 mL) and exposed to a 45-W microwave field for 2.5 min. After digestion, the samples were diluted with H_2O (15 mL) and the pH was adjusted to 4 with 1.5 mL of concentrated acetic acid (1.5 mL) and acetate buffer (5 mL). Sodium tetrathylborate (1 mL of 1%) and hexane (2 mL) were added and the mixture shaken. The extraction was repeated twice with fresh tetrathylborate (hexane not added), the sample centrifuged and an aliquot of the supernatant analyzed by GC. The method was validated by the use of certified reference materials.
- 'The Molecular Stability of Genomic DNA of Phytoplasma in the Witches'-broom-affected Paulownia Tissues after Microwave Heat Treatment'.[98] A comparative study was made on the use of microwave heat, air drying of tissue samples or fresh tissues from the midribs, petioles and tissue

cultures of paulownia that were infected or uninfected by phytoplasma prior to extraction of total genomic DNA. The DNAs extracted after these treatments were subjected to polymerase chain reaction (PCR) amplification using oligonucleotide primers targeting 16S rDNA, followed by restriction endonuclease treatments to establish restriction fragment length polymorphism (RFLP) profiles for comparison of treatment conditions and for comparison of witches'-broom-diseased tissues with normal paulownia tissues. DNAs extracted from microwave heat treated and from untreated phytoplasma showed identical nucleotide sequences, suggesting that the brief microwave heat treatment of 0.5 g of midribs and petioles for up to 14 min and 0.5-g tissue cultures for up to 20 min rendered phytoplasma noninfectious, but facilitated the preservation of phytoplasma DNA in situ without compromising the molecular stability. This simple method provides a means for long-term preservation and international exchange of phytoplasma specimens for molecular genetic studies between different laboratories.

- 'Microwave Extraction of Aerial Parts of *Zygophyllum Gaetulum*'.[99] MAE of saponins from the aerial parts of *Zygophyllum gaetulum* was investigated. Samples were ground and mixed with an appropriate solvent [MeOH, EtOH, MeOH–H_2O (8:2) and MeOH–H_2O (1:1)]. Microwave irradiation was performed for 2 min. After cooling to room temperature, the irradiation was repeated several times. The MAE procedure proved to be better than Soxhlet extraction with regard to the time and effort employed.
- 'A Rapid Microwave Method to Extract Plasmid DNA from *Saccharomyces cerevisiae* Suitable for the Transformation of *Escherichia coli*'.[100] A rapid microwave method to extract plasmid DNA from *Saccharomyces cerevisiae* suitable for the transformation of *Escherichia coli* is reported. The method was developed using different yeast strains and different microwave powers.
- 'Optimization and Comparison of Two MAE Procedures of Terpenic Compounds in *Vitis vinifera* Must Samples'.[101] In the first experiment, a sample (150 mL) of *Vitis vinifera* must was shaken with 2.5 g of Amberlite XAD-2 resin for 15 min, the mixture was filtered and the residue was microwave extracted with 30 mL of acetone–dichloromethane (1:1) in a sealed vessel using a 950-W microwave oven operated at 50% of its maximun power level. The mixture was heated to 0.5 mL and analyzed by GC/MS (details given). In the control experiment, 6 mL of must were microwave extracted directly with 10 mL of dichloromethane at 100 °C for 10 min. The extract was then analyzed as described above. The calibration graphs were linear from 10 to 30 and from 1 to 16 mg L^{-1} of terpenic compounds using the two extraction procedures, respectively. Recoveries were 88.7–94.9 and 80.1–86.8% for the two procedures, respectively. The methods were optimized by factorial design. Full details of the factor levels and extraction parameters used during the optimization processes are given.

5.4 Food and Pharmaceutical Samples

- 'Detection of Residues of the Epoxy Adhesive Component Bisphenol: a Diglycidyl Ether (BADGE) in Microwave Susceptors and its Migration into Food'.[102] Susceptors are an example of one of the many new products being introduced into food packaging. They are compounds which strongly absorb microwave radiation, creating very high temperatures on the surface of specific foods. This has the effect of browning the food during microwave cooking.
- 'Determination of Major Minerals in Dairy Products Digested in Closed Vessels Using Microwave Heating'.[103] An improved digestion method for dairy products (cheeses, caseinates and milk powders) for subsequent determination of Ca, Mg, K and Na is described. Comminuted samples (100 mL) were digested with three acid mixtures: (a) 2 mL of HNO_3 + 2 mL of H_2O_2, (b) 2 mL of $HClO_4$ + 2 mL of H_2O_2 and (c) 2 mL of $HClO_4$ + 1 mL of H_2O_2, in a microwave reactor for 5 min at 650 W and 10 min at 380 W. The reaction vessels were cooled in an ice bath.
- 'A Streamlined Flame Atomic Absorption Method for Animal Feed Analysis'.[104] A faster digestion protocol using nitric acid and microwave digestion is reported for the preparation of animal feed samples. FAAS using on-line dilution and the automated addition of lanthanum solution provided a much easier and quicker analysis of the feed samples.
- 'Accelerated Microwave Pre-treatment of Meat Samples Before Automated Determination of L-hydroxyproline'.[105] The method described is one in which 3 g of meat were microwave digested with 50 mL of 50% HCl solution in a closed vessel for 15 min at 800 W, or in an open vessel for 2 h at 200 W. The digest was neutralized with NaOH, diluted to 250 mL with water and a portion (100 µL) was analyzed by injection into an H_2O carrier stream. This was merged with a stream containing chloramine-T, and then merged with a stream containing *p*-dimethylaminobenzaldehyde. The resulting

colored product was monitored at 560 nm. Both methods (open and closed vessel) gave results comparable to those obtained by digestion in a sand bath, but required only 2 h or 15 min, respectively, compared with the 8 h needed for sand bath digestion. The closed vessel method was the most efficient as it required the shortest digestion time, and because 12 samples could be treated simultaneously, compared with one sample at a time for the open-vessel method.

- 'Use of a Microwave System for Polyphosphate Determination in Meat Products'.[106] Polyphosphate in meat was determined on the basis of the difference between total phosphorus (determined spectrophotometrically after microwave-based ashing) and organic phosphorus (estimated by using a conversion factor applied to the total protein determined by the Kjeldahl procedure). For total phosphorus determination, samples were digested in a microwave oven using nitric acid and hydrogen peroxide. Data obtained by using sausage and ham indicate that recoveries obtained after microwave digestion were higher than with traditional methods.
- 'Microwave Mineralization of Confectionery Samples and Determination of Trace Quantities of Heavy Metal Ions in Mineralizates'.[107] A wet mineralization method using a microwave irradiation energy source was optimized for the decomposition of confectionery samples for analysis. HNO_3 and H_2O_2 were used as oxidizing reagents. Optimal reagent quantities, mineralization time and irradiation energy were determined. The amounts of selected heavy metals (Pb, Cu, Zn, As and Sn) were determined in various confectionery samples. It was shown that the microwave mineralization was very fast and suitable for analytical use.
- 'Application of Atmospheric-pressure Microwave Digestion to Total Kjedahl Nitrogen Determination in Pharmaceutical, Agricultural and Food Products'.[108] Samples (200 mg) were digested with H_2SO_4 (20 mL) and 2 g of catalyst in a borosilicate reactor and then heated for 27 min in a microwave oven. Next, the mixture was cooled, treated with 33% H_2O_2 (7 mL) and heated for a further 3 min. Total nitrogen was determined according to the conventional method, by distilling the reaction mixture after rendering it alkaline and titrating the distillate. The results were in good agreement with those obtained by the official classical method.
- 'Determination of Dissolved Combined Amino Acids Using Microwave-assisted Hydrolysis and HPLC Precolumn Derivatization for Labeling of Primary and Secondary Amines'.[109] Microwave-assisted, vapor-phase acid hydrolysis of natural water samples for measurements of individual dissolved combined amino acids (DCAA) was examined and related to standard liquid-phase hydrolysis. The microwave technique allowed the labor-saving, simultaneous hydrolysis of 20 samples in 20 min. Relative to liquid-phase hydrolysis, a higher reproducibility and lower blanks were obtained with this procedure. In some samples, the microwave vapor-phase hydrolysis produced $\leq$39% higher DCAA concentrations than did the liquid-phase hydrolysis.
- 'Flow-injection Spectrophotometric Determination of Acetylsalicylic Acid in Tablets After Online Microwave-assisted Alkaline Hydrolysis'.[110] A known mass of powdered tablets, corresponding to 500 mg of acetylsalicylic acid (aspirin), was dissolved in and diluted with H_2O. A portion (250 µL) of the resulting solution and 250 µL of 0.2 M NaOH were simultaneously injected into two separate aqueous carrier streams (both 3.4 mL min^{-1}). The two streams were merged synchronously before being passed through a 200-cm reactor coil housed in a microwave oven operated at 600 W. After irradiation, the solution was passed through a 100-cm coil immersed in a water-bath and was then mixed with a reagent stream (1.3 mL min^{-1}) of 0.5% iron(III) nitrate in 0.4 M HNO_3 and the absorbance was measured at 525 nm. The results obtained agreed with those obtained by titrimetry.
- 'Hydrolysis of Phytic Acid by Microwave Treatment: Application to Phytic Acid Analysis in Pharmaceutical Preparations'.[111] An investigation of the use of a microwave-assisted technique for sample preparation in the analysis of phytic acid in pharmaceuticals, based on its hydrolysis to phosphate and subsequent analysis by spectrophotometry, is described. Ground tablets (15–20 mg) were suspended in 50 mL of H_2O with magnetic stirring for a few minutes. Portions (0.5–2.2 mL) of the resulting suspension were transferred to a PTFE vessel, 0.2 mL of 5.3 M HCl was added and the mixture was diluted to 2.4 mL with H_2O. The reactor was transferred to a domestic microwave oven, then the oven was sealed and heated at 650 W for six 2-min periods. The reactor was allowed to cool between each heating period for alternating 20- and 40-min intervals and finally to room temperature.
- 'Microwave-induced Rapid Transmethylation of Fatty Acids for Analysis of Food Oil'.[112] A 3-mL sample of vegetable oil in toluene (0.01 g mL^{-1}) and methyl heptadecanoate (3 mg mL^{-1}) internal standard and appropriate quantities of 0.3 or 1%

sodium methoxide in methanol in a PTFE vial were irradiated in a microwave oven or placed in a heating block. Reaction was stopped with 0.1 mL of acetic acid. Transesterification was completed in 40 s in methanol–toluene (1 : 10) and in 15 s in methanol–toluene (1 : 3) and the procedure gave a yield of fatty acid methyl esters comparable to or higher than that obtained by the conventional method and was sufficiently mild for the analysis of oils rich in polyunsaturated fatty acids.

- 'Extraction of Selected Drugs from Serum Using Microwave Irradiation'.[113] MAE of selected drugs from human blood/serum was performed in an 'atmospheric pressure' system. Before irradiation with microwaves, an appropriate solvent mixture was added to the buffered specimen. Lidocaine, methadone, diazepam, nordiazepam, propoxyphene and norpropoxyphene were tested as model substances. The final analysis was performed by gas chromatography/nitrogen–phosphorus detection (GC/NPD). The procedure had been applied successfully in a number of forensic cases.
- 'MAE of Free Amino-acids from Foods'.[114] MASE was used to separate free amino acids from foods. The MLS Mega microwave sample preparation system permitted concurrent extractions from ten 5-g samples into aqueous 7% $HClO_4$. Foods extracted were salami, cheese, broccoli and cauliflower. Optimum extraction temperatures were 40 °C for vegetables and 50 °C for higher fat foods. Comparative results were tabulated for extraction of up to 17 amino acids by microwave extraction (20-min analysis time) and traditional extraction by shaking. The microwave procedure yielded 5–10% more amino acids from each product.
- 'MAE of Taxanes from *Taxus* Biomass'.[115] Taxanes, including paclitaxel, were extracted from *Taxus* needles by MAE with an MES-1000 extraction system. Conditions were optimized with 95% ethanol as solvent, 85 °C and time 9–10 min. Advantages over the conventional overnight soaking extraction method were reported. Recoveries were similar to those obtained with the traditional method.
- 'Extraction of Incurred Sulphamethazine in Swine Tissue by MAE and Quantification without Cleanup by HPLC Following Derivatization with Dimethylaminbenzaldehyde'.[116] Samples (freeze-dried or wet tissues) were irradiated in methanol for 25 s in a household microwave oven. After filtration, sulfamethazine (sulfadimidine) was determined by HPLC. The results agreed well with those obtained by a homogenization technique for freeze-dried samples.
- 'Development and Validation of MAE of Fortified and Incurred Chloramphenicol Residues in Egg Albumin and Yolk'.[117] A sample of egg equivalent to 1 g of nonlyophilized egg (containing 0.58 g of yolk per 0.3 g of albumin) was microwave extracted with 15 mL of acetonitrile and 2 mL of 2-propanol. The residue was re-suspended in acetonitrile–2-propanol (15 : 2) and the supernatant was decanted. The combined extracts were centrifuged, the new supernatant was evaporated to dryness and the residue was dissolved in ethanol. Recoveries from conventional and MAE are compared.

5.5 Mineral and Ceramic Samples

Geological samples are included in this section.

- 'Determination of the Platinum-group Elements in Geological Materials by ICPMS Using Microwave Digestion, Alkali Fusion and Cation-exchange Chromatography'.[118] A new method for the determination of platinum group elements (Ru, Rh, Pd, Ir, Pt) in geological samples, using microwave-assisted acid digestion and alkali fusion, separation of the elements by cation-exchange chromatography and analysis by ICPMS, is described.
- 'Microwave Treatment of Minerals – A Review'.[119] This is a short review with 25 references of the advances in the microwave treatment of minerals from the early stages of development to possibilities for future utilization. It is included since it gives some ideas which suggest a number of analytical applications. Many different applications are considered, including fundamental heating rate studies, microwave grinding, possible exploitations in the area of extractive metallurgy and also microwave treatment of coal. Further indications of the potential for the common exploitation of microwaves within the mineral processing and extractive metallurgical industries are included.
- 'Potential of Water for Continuous Automated Sample Leaching'.[120] This summary describes the use of water as a leaching agent in discontinuous and continuous solid–liquid extraction procedures. The ways in which new methods have improved earlier ones are discussed. The use of ultrasound and microwaves is considered, with particular reference to the use of high pressure and temperature for accelerating leaching with water under sub- and supercritical conditions (45 references).
- 'Microwave-assisted Acid Dissolution of Sintered Advanced Ceramics for ICP Atomic Emission

Spectrometry'.[121] A systematic study is described of the conditions required for the microwave-assisted dissolution of sintered samples of 28 structural and electronic advanced ceramic materials. These include zirconia-, alumina-, ceria-, titania-, lead titanate-, lead zirconate titanate- and niobate-based ceramics, nonoxide-based ceramics such as AlN, BN, Si_3N_4 and SiC, and oxide and nonoxide ceramics such as β'-sialon (silicon aluminum oxynitride). Dissolutions were carried out using an RMS-150 microwave-digestion system and 15 acids or mixtures of acids were investigated. Optimized methods were developed for the complete dissolution of all the materials studied except for β'-sialon and SiC. The resulting solutions were analyzed by ICPAES and detection limits for 16 elements are tabulated together with those obtained by decomposition methods based on alkali fusion in a graphite or platinum crucible. Results are also presented for the determination of major and trace elements in a gadolinium lead titanate by microwave digestion and by alkali fusion in graphite or platinum crucibles. The main advantage of microwave digestion over fusion methods is that it involves the use of small amounts of high-purity acids which generally contain less impurities, resulting in improved precision and freedom from matrix effects.

- 'Microwave-assisted Sample Preparation and Preconcentration for ETAAS'.[122] A report of a broad study of the microwave-assisted dissolution of a wide variety of materials, e.g. sulfide ores, carbonaceous shales, coals, high-purity oxides of Nb and Ta, soils and dusts, botanical samples and blood, before analysis by ETAAS. Application of microwave-assisted dissolution procedures is shown to reduce the sample preparation time and offer possibilities for the automation of procedures.

5.6 Industrial Samples

- 'Decomposition of Coal Under Microwave Irradiation'.[123] This article describes a proposed industrial process for the decomposition of coal in a CH_4 atmosphere under microwave irradiation. It is included to spur analytical adaptations of this interesting process.
- 'A Method for Decreasing Metal Ion Content in Chemical Compounds with a Dipole Moment or a Mixture of these Compounds by Microwave Distillation'.[124] The frequency of the microwave radiation used in the distillation is 1 MHz–50 GHz, and the metal ion content in the title compound(s) was decreased to $\leq$1 ppb. The compounds studied include HCl, HF, H_2SO_4, HNO_3, HOAc, aqueous NH_3, acetone, MeOH, EtOH, PrOH, *N*-methylpyrrolidone, propylene glycol monoethyl acetate and propylene glycol monoethyl ether. The distillation vessel has a casing of electrically conductive material such as steel, Cu, Al, Ni, Cr, Ti, Ag or Au. The method is used in the semiconductor industry and is included here for potential analytical applications.
- 'Open-vessel Microwave Digestion of Wear Metals in Lubricating Oil'.[125] The Star 2 microwave decomposition system (CEM, Kamp-Lintfort, Germany) is described. With use of this apparatus, oil samples were decomposed with sulfuric and nitric acids and subsequently with HCl and hydrogen peroxide by programmed heating to a maximum temperature of 280 °C. Decomposition took 30 min. Cu, Fe, Ti, Ni and Pb in the resulting solution were determined by ICPAES. The results obtained on a 'wear metals in oil' reference material agreed with the certified values.
- 'Chemical Characterization of Municipal Solid Waste Incineration Residue: Dissolution of Elements with a Microwave–Dilute Acid Digestion Technique Compared to Conventional Methods'.[126] The focused microwave technique in an open system for the rapid dissolution of elements in fly ash and filter cakes was evaluated in comparison with conventional wet and dry methods. The method involved a 15-min heating time in 2 M nitric or hydrochloric acid. It was tested on different types of samples and led to the dissolution of Ca, Na, K, Zn, Pb and As with 2 M HNO_3 or HCl, and Hg with 2 M HNO_3 for subsequent determination by AES or AAS. The results justify the adoption of the new method.
- 'Dissolution of Polymers Bearing Acid Groups Using Microwave Radiation'.[127] The polymers studied can be dissolved in H_2O using microwave radiation much more readily than by conventional means. Placing 2.5 g of sulfonated (51%), milled polyether ether ketone (PEEK) and 47.5 g of H_2O in a microwave autoclave, flushing with N_2, subjecting the apparatus to microwave radiation for 5 min, cooling and centrifuging resulted in 98.3% dissolution, the molecular weight of the polymer being unchanged.
- 'A Rapid Method for the Determination of Trace Element Impurities in Silicone Oils by ICPMS After MAD'.[128] A MAD procedure has been developed for the treatment of silicone oil samples. ICPMS was used to determine the concentration level of 40 trace element impurities.
- 'The Analysis of Coal Tar Pitch by ICP Optical Emission Spectrometry After Digestion in a Microwave Oven System'.[129] A controlled microwave heating and digestion method was developed for the dissolution and preparation of samples of coal-tar pitch prior

to analysis by ICP optical emission spectrometry. A clear solution indicated minimal residual carbon levels, whereas partially digested solutions tended to give yellow solutions.

5.7 Microwave-assisted Sample Drying

- 'Microwave Vacuum Drying of Marine Sediment Prior to Analysis for Carbon and Metals'.[130] Errors that arise during an intermediate step in the analytical procedure of sediment drying are considered. Microwave drying in vacuum was used to reduce the boiling temperature of water and prevent loss of volatile species. The key issues in the study relate to the accuracy and precision of this process, reflected by comparison with the established oven drying procedure at 105 °C.
- 'Sample Preservation for Determination of Organic Compounds: Microwave Versus Freeze-drying'.[131] In a search of a reliable drying method, which might be used even under field conditions, microwave drying was compared with freeze-drying of plant material. Certain leaves, buds and phloem were used and checked for one or more of the following substances: sugars, sugar alcohols, organic and amino acids, total nitrogen and glycinebetaine. With most samples good agreement was achieved between the two drying methods. Only in the case of *Ananas comosus* leaves, which exhibited low pH and high water content, did appreciable differences occur in organic and amino acids. For *Ananas comosus* leaf samples it was shown that microwave drying could also be used prior to extraction of tissue sap.
- 'Determining the Water Content of Aggregate and Freshly-mixed Concrete by the Microwave Method'.[132] A new standard method for determining the water content of concrete was developed and compared with the conventional spirit-drying method. The effects of the water/cement (w/c) ratio, plasticizers, concrete consistency, cement strength and further influencing factors on the quality of the analytical results were studied. Using the microwave method, the water content can be reliably determined up to 90 min after water addition, if the concrete is thoroughly homogenized, and the thickness of the sample is $\leq$15 cm. For concrete containing high-strength cement, 1000 W of microwave power are required to remove water completely (90 min after mixing). In contrast to the spirit-drying method, the core moisture and the water content of the additives are also determined. The reproducibility of the analytical results obtained from the microwave method was comparable to or better than that of the conventional technique.

5.8 Microwave Fusions

There has been far less use of microwave energy to bring about the fusion or sintering of analytical samples than the potential of this technique warrants. Therefore, several examples illustrating the versatility and effectiveness of microwave fusions or sintering processes are given in this section in the hope that readers may be encouraged to develop new fusion protocols for difficult samples.

- 'Microwave Melt Texturing of Bulk YBCO Superconductors'.[133] The potential for the use of microwave irradiation as an alternative energy source for the melt texturing of bulk, ceramic high-critical temperature YBCO (Y–Ba–Cu–O) superconductors has been investigated with a view to taking advantage of the opportunities presented by this novel heating technique. T-shaped samples of 83% dense, sintered YBCO were melt-processed, both conventionally and using microwave energy, while suspended within two different casketing arrangements. Samples consisting of densely packed, 150-μm long grains all aligned with the axis of the sample could be obtained under the best conditions. In addition, completely uniform and full oxygen lattice occupancy ($X = 7$ in $YBa_2Cu_3O_x$) could be obtained with all the microwave-processed samples when the melt processing temperature was kept in the range 910–940 °C. However, on the negative side, this appeared to limit the maximum density to 94% of theory unless a slight drop in x to 6.98 could be tolerated. The higher processing temperatures, which yielded the best microstructures, also resulted in 5–8% of the 211 phase being present.
- 'Accelerated Sintering and Phase Transformation of TiO_2 in Microwave Radiation'.[134] The sintering process and phase transformation of submicrometer powder compacts of TiO_2 were investigated using 2.45-GHz microwave radiation. The microwave-sintered samples were densified more rapidly and in a much shorter time and at a lower temperature than that of the conventionally sintered samples. Also, an accelerated phase transformation from anatase to rutile was observed in microwave processing when the dwelling time was reduced from 6 h for conventional heating to 30 min for microwave heating at 1000 °C. The samples sintered in microwave radiation show a refined microstructure and smaller grain sizes at high density near 98–99% of the theoretical value. The reasons for rapid heating behavior and enhanced sintering of TiO_2 under microwave radiation are discussed.
- 'Analysis of Heat Generation in a Conductive Polymer in Single Mode Microwave Welding'.[135]

Microwave welding of plastics can be achieved by placing a conductive polymer heating element at a joint interface. This heating element absorbs the microwave energy and converts it into heat, which determines the heat generation and peak temperature during welding. Hence it controls the welding cycle and final joint strength.

6 CONCLUDING REMARKS

Although only relatively recently developed, the microwave-assisted dissolution technique is rapidly becoming the method of choice for the routine handling of large numbers of analytical samples. Although the approach does not so far provide a universal sample preparation technique, it does offer certain advantages over conventional dissolution techniques. For example, the rate-determining step for sample throughput is usually sample preparation. Microwave-assisted dissolution techniques offer a reasonably economical solution to this problem. As the result of saving time, these techniques are also generally more economical in the long run, through savings due to reduced labour costs. Also, these techniques often require smaller amounts of sometimes expensive reagents, can contain corrosive fumes, thereby preventing contamination and loss of volatile elements, require less supervision, are simpler and more efficient and have proven to be reliable.

Another major advantage of the speed with which microwave-assisted dissolutions occur is that this can lead to rapid availability of results. Modern process control often requires constant monitoring of analytical parameters – these are sometimes available via on-line X-ray fluorescence (XRF) analyses, but often AAS or ICP results are needed, so that a rapid sample dissolution method becomes a necessity for efficient plant operation.

Techniques involving microwaves are not without faults, albeit they can be minimized. One disadvantage when compared with the use of open digestion vessels is that dissolution variables such as sample weight and type, acid volume, microwave power, microwave distribution homogeneity and duration of exposure must be collectively acknowledged and controlled with all due respect to the possible explosive nature of the method. However, such control itself is advantageous since dissolution conditions are accurately reproduced without surpassing vessel limitations. In addition, analyte accuracy and precision limits within batches of samples should be reduced owing to this reproducibility.

Microwave-assisted dissolutions, although relatively rapid, do themselves have one disconcertingly slow step. This is the necessary delay in opening closed vessels with high internal pressures. These must be cooled to room temperature in order for this pressure to be reduced to a safe level. Several methods of shortening cooling times have been investigated, such as internal vessel cooling systems, refrigeration or use of a freezer or cooling in ice, water or some other cold material such as liquid nitrogen. However, the time required for cooling of vessels by whatever method is only a minor inconvenience when considering how much time can be saved in performing dissolutions. For example, it takes 24 h for the complete dissolution of wool using classical methods whereas only 8 min are required using microwave digestion; this is 180 times faster!

The usefulness of microwaves in analytical chemistry is increasingly evident. From a meagre start in the 1970s represented by only seven references, publication had increased to 132 references in the 1980s and to 166 references in only the first four years of the 1990s! This review is a small selection of some of the most significant papers in the last 2–3 years, and yet covers ~90 such papers. The growth rate is increasing ever more rapidly, reflecting a wide acceptance of a variety of microwave-assisted sample preparation techniques. Microwave energy has proven to be useful in shortening drying, extraction, Kjeldahl determination, ashing and fusion times and also in increasing efficiency. Microwave techniques are also exceptionally applicable to continuous flow and on-line systems owing to their simplicity and ease of use. One or two such systems are now commercially available and have passed the prototype stage. Within the next few years, flow-through protocols will most likely replace many of the closed-vessel methods described above in high-throughput situations.

ACKNOWLEDGMENTS

One of the authors (F.S.) acknowledges the contributions of those who have worked with him in this field during the past 15 years, particularly John Bozic, Richard Gedye and Ken Westaway. The other author (G.X.) acknowledges the guidance provided by Professor Zhanxia Zhang during his doctoral study at Zhongshan (Sun Yat-Sen) University, Guangzhou, China. His research related to this article was supported by the National Natural Science Fund of China (No. 29967001).

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectroscopy
AES	Atomic Emission Spectroscopy

ASE	Accelerated Solvent Extraction
CMP	Capacitatively Coupled Microwave Plasma
DC	Direct Current
DPASV	Differential Pulse Anodic Stripping Voltammetry
EPR	Electron Paramagnetic Resonance Spectroscopy
ESR	Electron Spin Resonance Spectroscopy
ETAAS	Electrothermal Atomic Absorption Spectroscopy
FAAS	Flame Atomic Absorption Spectroscopy
GC	Gas Chromatography
GC/AES	Gas Chromatography/Atomic Emission Spectroscopy
GC/ECD	Gas Chromatography/Electron Capture Detection
GC/FID	Gas Chromatography/Flame Ionization Detection
GC/FPD	Gas Chromatography/Flame Photometric Detection
GC/MS	Gas Chromatography/Mass Spectrometry
GC/NPD	Gas Chromatography/Nitrogen–Phosphorus Detection
GPC	Gel Permeation Chromatography
HCB	Hexachlorobenzene
HPCE	High-performance Capillary Electrophoresis
HPLC	High-performance Liquid Chromatography
IAEA	International Atomic Energy Agency
ICP	Inductively Coupled Plasma
ICPAES	Inductively Coupled Plasma Atomic Emission Spectroscopy
ICPMS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
ISM	Industrial, Scientific and Medical
MAD	Microwave-assisted Digestion
MAE	Microwave-assisted Extraction
MASE	Microwave-assisted Solvent Extraction
MIP	Microwave-induced Plasma
MIPAES	Microwave-induced Plasma Atomic Emission Spectroscopy
MP	Microwave Plasma
MPT	Microwave Plasma Torch
NIST	National Institute of Standards and Technology
PAH	Polycyclic Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PCDD	Polychlorinated Dibenzo-*p*-dioxin
PCDF	Polychlorinated Dibenzofuran
PCR	Polymerase Chain Reaction
PEEK	Polyether Ether Ketone
PFA	Polyfluoroalkoxy
PTFE	Polytetrafluoroethylene
RADAR	Radio Detection and Ranging
RFLP	Restriction Fragment Length Polymorphism
SFE	Supercritical Fluid Extraction
SLM	Supported Liquid Membrane
SPE	Solid-phase Extraction
SPME	Solid-phase Microextraction
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
XRF	X-ray Fluorescence

RELATED ARTICLES

Biomolecules Analysis **(Volume 1)**
High-performance Liquid Chromatography of Biological Macromolecules

Carbohydrate Analysis **(Volume 1)**
Carbohydrate Analysis: Introduction • Disaccharide, Oligosaccharide and Polysaccharide Analysis

Chemical Weapons Chemicals Analysis **(Volume 2)**
Fourier Transform Infrared in On-site and Off-site Analysis of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Drugs of Abuse, Analysis of

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction

Environment: Water and Waste **(Volume 3)**
Environmental Analysis of Water and Waste: Introduction • Detection and Quantification of Environmental Pollutants • Flow-injection Techniques in Environmental Analysis • Industrial Waste Dumps, Sampling and Analysis • Inorganic Analysis in Environmental Samples by Capillary Electrophoresis • Inorganic Environmental Analysis by Electrochemical Methods • Ion Chromatography in Environmental Analysis

Environment: Water and Waste cont'd **(Volume 4)**
Ion-selective Electrodes in Environmental Analysis • Microwave-assisted Techniques for Sample Preparation in Organic Environmental Analysis • Microwave-enhanced Solvent Extraction of Organics in Environmental Analysis • Pervaporation, Analytical • Sample Preparation for Elemental Analysis of Biological Samples

in the Environment • Sample Preparation for Environmental Analysis in Solids (Soils, Sediments, and Sludges) • Sample Preparation Techniques for Elemental Analysis in Aqueous Matrices • Soxhlet and Ultrasonic Extraction of Organics in Solids • Waste Extraction Procedures

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Sample Preparation Analytical Techniques for Food • Sample Preparation for Food Analysis, General

Pesticides **(Volume 7)**
Pesticide Analysis: Introduction

Pesticides cont'd **(Volume 8)**
Soil and Sediments: Pesticides Content Sampling, Sample Preparation, and Preservation

Pulp and Paper **(Volume 9)**
Pulp and Paper Matrices Analysis: Introduction

Atomic Spectroscopy **(Volume 11)**
Microwave-induced Plasma Systems in Atomic Spectroscopy

REFERENCES

1. H. Engelhalt, A. Gross, 'Supercritical Fluid Extraction and Chromatography: Potential and Limitations', *Trends Anal. Chem.*, **10**, 64–70 (1991).
2. B.A. Bidlingmeyer, 'Guidelines for Proper Usage of Solid-phase Extraction', *LC Liq. Chromatogr. HPLC Mag.*, **2**, 578–580 (1984).
3. C.L. Arthur, J. Pawliszyn, 'Solid Phase Microextraction with Thermal Desorption Using Fused Silica Optical Fibers', *Anal. Chem.*, **62**, 2145–2148 (1990).
4. G. Audunsson, 'Aqueous/aqueous Extraction by Means of a Liquid Membrane for Sample Cleanup and Preconcentration of Amines in a Flow System', *Anal. Chem.*, **58**, 2714–2723 (1986).
5. B.E. Richter, B.A. Jones, J.L. Ezzell, N.L. Porter, N. Avdalovic, C. Pohl, 'Accelerated Solvent Extraction: a Technique for Sample Preparation', *Anal. Chem.*, **68**, 1033–1039 (1996).
6. R. Kuhn, F. Stocklin, F. Erni, 'Chiral Separations by Host–Guest Complexation with Cyclodextrin and Crown Ether in Capillary Zone Electrophoresis', *Chromatogr.*, **33**, 32–36 (1992).
7. F. Smith, B. Cousins, J. Bozic, W. Flora, *Anal. Chim. Acta*, **177**, 243 (1985).
8. A. Abu-Samra, J.S. Morris, S.R. Koirtyohann, 'Wet Ashing of some Biological Samples in a Microwave Oven', *Anal. Chem.*, **47**(8), 1475–1477 (1975).
9. A. Abu-Samra, J.S. Morris, S.R. Koirtyohann, *Trace Subs. Environ. Health*, **9**, 297 (1975).
10. J.A. Hesek, R.C. Wilson, 'Use of Microwave Oven in In-process Control', *Anal. Chem.*, **46**(8), 1160 (1974).
11. A.J. Berteaud, J.C. Badot, *J. Microwave Power*, **11**, 315 (1976).
12. S.A. Matthes, R.F. Farrell, A.J. Mackie, *Tech. Prog. Rep. U.S. Bur. Mines*, TPR **120**, 9 (1983).
13. The Savillex Corporation, 6133 Baker Rd., Minnetonka, MN 55345, USA.
14. F.E. Smith, B.G. Cousins, J.Y. Maillet, *Educ. Chem.*, **24**(1), 13 (1987).
15. J. Huang, D. Goltz, F. Smith, *Talanta*, **35**(11), 907 (1988).
16. H.M. Kingston, L.B. Jassie (eds.), *Introduction to Microwave Sample Preparation*, ACS Professional Reference Book, American Chemical Society, Washington, DC, 1988.
17. F.E. Smith, B. Cousins, J. Bozic, W. Flora, *Proceedings of the IUPAC Symposium on Analytical Chemistry in the Exploration, Mining and Processing of Materials*, Pretoria, South Africa, 1985.
18. J. Bozic, W. Flora, F. Smith, B. Cousins, *Proceedings of the 68th Canadian Chemical Conference*, Kingston, Ontario, Canada, 1985.
19. F.E. Smith, B. Cousins, J. Bozic, W. Flora, *Proceedings of the 17th Annual Canadian Mineral Analysts Conference*, Val d'Or, Québec, Canada, September, 1985.
20. H.M. Kingston, S.J. Haswell (eds.), *Microwave-enhanced Chemistry – Fundamentals, Sample Preparation and Applications*, American Chemical Society, Washington, DC, 1997.
21. F.E. Smith, 'Microwaves in the Laboratory – Recent Advances', Keynote Presentation, *Proceedings of the 2nd ANIAC International Conference*, Kuala Lumpur, Malaysia, November, 1993.
22. R.N. Gedye, F.E. Smith, K.C. Westaway, H. Ali, L. Baldisera, L. Laberge, J. Rousell, *Tetrahedron Lett.*, **27**, 279 (1986).
23. R.N. Gedye, F.E. Smith, K.C. Westaway, *Can. J. Chem.*, **66**, 17 (1988).
24. R.N. Gedye, F.E. Smith, K.C. Westaway, *Educ. Chem.*, **25**, 55 (1988).
25. R.J. Giguere, T.L. Bray, S.M. Duncan, G. Majetich, *Tetrahedron Lett.*, **27**, 4945 (1986).
26. R.J. Giguere, A.M. Naman, B.O. Lopez, A. Arepally, D.E. Ramos, *Tetrahedron Lett.*, **28**, 6553 (1987).
27. C.R. Strauss, R.B. Trainor, *Aust. J. Chem.*, **48**, 1665 (1995).
28. D.M.P. Mingos, D.R. Baghurst, *Chem. Soc. Rev.*, **20**, 1 (1991).
29. A. Zlotorzynski, *Crit. Rev. Anal. Chem.*, **25**(1), 43–76 (1995).
30. P. Debye, *Polar Molecules*, Chemical Catalog Co., New York, 1929, and Dover Publications, New York, 1945.
31. P. Debye, *Phys. Z.*, **36**, 100 (1935).

32. H. Fröhlich, *Theory of Dielectrics*, Oxford University Press, London, 1949.
33. T.T. Meek, *J. Mater. Sci. Lett.*, **6**, 638 (1987).
34. http://www.gallawa.com/microtech/magnetron.html.
35. M. Mingos, M. Teffal, A. Gourdenne, *Eur. Polym. J.*, **19**, 543 (1983).
36. C.R. Buffler, T. Lindstrom, *Microwave World*, **9**, 10 (1988).
37. K.I. Mahan, T.A. Foderaro, T.L. Garza, R.M. Martinez, G.A. Maroney, M.R. Trivisonno, E.M. Willging, *Anal. Chem.*, **59**, 938 (1987).
38. P. Aysola, P.D. Anderson, C.H. Langford, *Anal. Lett.*, **21**, 2003 (1988).
39. H.M. Assenheim, D.A. Hill, E. Preston, A.B. Cairnie, 'The Biological Effects of Radio-frequency and Microwave Radiation', Publication No. NRCC 16448, National Research Council of Canada, Ottawa, 1978.
40. B. Erickson, 'Standardizing the World with Microwaves', *Anal. Chem.*, **70**(13), 467A–471A (1998).
41. I.J. Barnabas, J.R. Dean, I.A. Fowlis, S.P. Owen, 'Extraction of Polycyclic Aromatic Hydrocarbons from Highly Contaminated Soils Using Microwave Energy', *Analyst*, **120**(7), 1897–1904 (1995).
42. A.E. Croslyn, B.W. Smith, J.D. Winefordner, 'A Review of Microwave Plasma Sources in Atomic Emission Spectrometry: Literature from 1985 to the Present', *Crit. Rev. Anal. Chem.*, **27**(3), 199–255 (1997).
43. S.J. Chalk, H.M. Kingston, P.J. Walter, K. McQuillin, J. Brown, 'SamplePrep Web: Analytical Sample-preparation and Microwave-chemistry Resource Center', in *Microwave-enhanced Chemistry – Fundamentals, Sample Preparation and Applications*, eds. H.M. Kingston, S.J. Haswell, American Chemical Society, Washington, DC, 667–693, 1997.
44. H.M. Kingston, 'Standardization of Sample Preparation for Trace Element Determination Through Microwave-enhanced Chemistry', *At. Spectrosc.*, **19**(2), 27–30 (1998).
45. K.J. Lamble, S.J. Hill, 'Microwave Digestion Procedures for Environmental Matrices', *Analyst*, **123**(7), 103R–133R (1998).
46. M.-K. Wong, W. Gu, T.-L. Ng, 'Sample Preparation Using Microwave-assisted Digestion or Extraction Techniques', *Anal. Sci.*, **13**(Suppl.), 97–102 (1997).
47. U. Sengutta, 'Microwave Analysis Systems for Rapid Process Analysis', *CLB Chem. Labor. Biotech.*, **49**(5), 166–168,170 (1998).
48. D.K. Agrawal, 'Microwave Processing of Ceramics', *Curr. Opin. Solid State Mater. Sci.*, **3**(5), 480–485 (1998).
49. F.E. Smith, E.A. Arsenault, 'Microwave-assisted Sample Preparation in Analytical Chemistry', *Talanta*, **43**, 1207–1268 (1996).
50. E. Campos, E. Barahona, M. Lachica, M.D. Mingorance, 'A Study of the Analytical Parameters Important for the Sequential Extraction Procedure Using Microwave Heating for Pb, Zn and Cu in Calcareous Soils', *Anal. Chim. Acta*, **369**, 235–243 (1998).
51. B. Perez-Cid, I. Lavilla, C. Bendicho, 'Application of Microwave Extraction for Partitioning of Heavy Metals in Sewage Sludge', *Anal. Chim. Acta*, **378**, 201–210 (1999).
52. R. Lobinski, I.R. Pereiro, H. Chassaigne, A. Wasik, J. Szpunar, 'Elemental Speciation and Coupled Techniques – Towards Faster and Reliable Analyses', *J. Anal. At. Spectrom.*, **13**(9), 859–867 (1998).
53. Y. Yamaguchi, S. Mano, B. Kyoh, T. Yamamoto, 'Microwave-assisted Distillation in Tritium Monitoring (in Japanese)', *Hoken Butsuri*, **33**(3), 337–340 (1998).
54. D.W. Bryce, A. Izquierdo, M.D. Luque de Castro, 'Continuous Microwave Assisted Pervaporation/Atomic Fluorescence Detection: an Approach for Speciation in Solid Samples', *Anal. Chim. Acta*, **324**(1), 69–75 (1996).
55. V. Lopez-Avila, R. Young, W.F. Beckert, 'Stability of Organic Pollutants During Microwave-assisted Extraction from Solid Matrices', *J. AOAC Int.*, **81**(2), 462–476 (1998).
56. V.O. Schmitt, A. De Diego, A. Cosnier, C.M. Tseng, J. Moreau, O.F.X. Donard, 'Open Focused-microwave-assisted Sample Preparation Procedures: Fundamentals and Application to the Speciation of Tin and Mercury in Environmental Samples', *Spectroscopy*, **13**(1), 99–111 (1997).
57. M. Bettinelli, S. Spezia, G. Quattroni, A. Giove, 'Analysis of Trace Elements in Coal Fly Ash and Their Leachates: Results of the ENEL–EDF Round-robin Test', *Ann. Chim.*, **88**(3–4), 269–280 (1998).
58. K.R. Islam, R.R. Weil, 'A Rapid Microwave-digestion Method for Colorimetric Measurement of Soil Organic Carbon', *Commun. Soil Sci. Plant Anal.*, **29**(15–16), 2269–2284 (1998).
59. J. Gorecki, J. Golas, 'Voltammetric Procedure for Determination of Mercury in Soil Wastes After Wet Microwave Digestion', *Chem. Anal. (Warsaw)*, **43**(6), 1027–1031 (1998).
60. L. Bordera, V. Hernandis, A. Canals, 'Automatic Flow-injection System for the Determination of Heavy Metals in Sewage Sludge by Microwave Digestion and Detection by Inductively Coupled Plasma Atomic Emission Spectrometry (MW–ICPAES)', *Fresenius' J. Anal. Chem.*, **355**(2), 112–119 (1996).
61. J.M. Torres, M. Llaurado, G. Rauret, 'Microwave-assisted Pre-treatment of Environmental Samples for the Determination of ^{90}Sr', *Anal. Chim. Acta*, **379**(1–2), 135–142 (1999).
62. J.A. Sanchez-Cabeza, P. Masque, I. Ani-Ragolta, '^{210}Po and ^{210}Pb Analysis in Sediments and Soils by Microwave Acid Digestion', *J. Radioanal. Nucl. Chem.*, **227**(1–2), 19–22 (1998).
63. M.C. Yebra, M.F. Enriquez, 'An Optimized Microwave Digestion Procedure for Cadmium Analysis of Mussel Samples', *Analusis*, **26**(7), 261–263 (1998).

64. W.S. Dancer, R. Eliason, S. Lekhakul, 'Microwave-assisted Soil and Waste Dissolution for Estimation of Total Phosphorus', *Commun. Soil Sci. Plant Anal.*, **29**(11–14), 1997–2006 (1998).
65. B.S. Xing, P.L.M. Veneman, 'Microwave Digestion for Analysis of Metals in Soil', *Commun. Soil Sci. Plant Anal.*, **29**(7–8), 923–930 (1998).
66. M.E. Tatro, 'Sample Preparation for Trace Metals. Part II: Microwave Digestion', *Environ. Test. Anal.*, **7**(5), 16–17, 28 (1998).
67. D. Florian, R.M. Barnes, G. Knapp, 'Comparison of Microwave-assisted Leaching Techniques for the Determination of Heavy Metals in Sediments, Soils, and Sludges', *Fresenius' J. Anal. Chem.*, **362**(7–8), 558–565 (1998).
68. D. McGrath, 'Use of Microwave Digestion for Estimation of Heavy Metal Content of Soils in a Geochemical Survey', *Talanta*, **46**(3), 439–448 (1998).
69. D. Weiss, W. Shotyk, H. Schaefer, U. Loyall, 'Microwave Digestion of Ancient Peat and Determination of Pb by Voltammetry', *Fresenius' J. Anal. Chem.*, **363**(3), 300–305 (1999).
70. C. Brunori, M.B. De la Calle-Guntinas, R. Morabito, 'Optimization of the Reduction of Selenium(VI) to Selenium(IV) in a Microwave Oven', *Fresenius' J. Anal. Chem.*, **360**(1), 26–30 (1998).
71. M. Weichbrodt, W. Vetter, E. Scholz, B. Luckas, 'Levels of Organochlorines in Penguin and Skua Eggs from the Antarctic. Determination after Application of Focused Open-vessel Microwave-assisted Extraction in Combination with Gel Permeation Chromatography', *Organohalogen Compd.*, **35**, 195–198 (1998).
72. W. Vetter, M. Weichbrodt, K. Hummert, D. Goltz, B. Luckas, 'Combined Microwave-assisted Extraction and Gel Permeation Chromatography for the Determination of Chlorinated Hydrocarbons in Seal Blubber and Cod Livers', *Chemosphere*, **37**(9/12), 2439–2449 (1998).
73. I. Windal, G. Eppe, A.-C. Gridelet, L.E. Garcia-Ayuso, M.D. Luque de Castro, E. de Pauw, 'Fast Extraction of Dioxins from Fly Ash. Comparison of Soxhlet, Supercritical Fluid and Microwave-assisted Soxhlet Extraction', *Organohalogen Compd.*, **35**, 199–203 (1998).
74. S.P. Frost, J.R. Dean, K.P. Evana, K. Harradine, C. Cary, M.H.I. Comber, 'Extraction of Hexaconazole from Weathered Soils: a Comparison Between Soxhlet Extraction, Microwave-assisted Extraction, Supercritical Fluid Extraction and Accelerated Solvent Extraction', *Analyst*, **122**(9), 895–898 (1997).
75. W. Lautenschlaeger, I. Floeter, G. Schwedt, 'Pesticide Analysis without Sample Transfer', *LaborPraxis*, **22**(7–8), 42–44 (1998).
76. 'Extraction of Organic Substances with Solvents', *LaborPraxis*, **22**(1), 48–49 (1998).
77. E. Eljarrat, J. Caixach, J. Rivera, 'Microwave vs. Soxhlet for the Extraction of PCDD and PCDF from Sewage Sludge Samples', *Chemosphere*, **36**(10), 2359–2366 (1998).
78. S. Dupeyron, P.M. Dudermel, D. Couturier, 'Focused Microwave-assisted Extraction (FMAE) of Polynuclear Aromatic Hydrocarbons from Contaminated Soil: Role of Acetone and Water Content Impact on Microwave Efficiency', *Analusis*, **25**(8), 286–292 (1997).
79. A. Egizabal, O. Zuloaga, N. Etxebarria, L.A. Fernandez, J.M. Madariaga, 'A Comparison of Microwave-assisted Extraction and Soxhlet Extraction for Phenols in Soil Samples Using an Experimental Design', *Analyst*, **123**(8), 1679–1684 (1998).
80. K. Li, J.M.R. Belanger, M.P. Llompart, R.D. Turpin, R. Singhvi, J.R.J. Pare, 'The Evaluation of Rapid Solid Sample Extraction Using the Microwave-assisted Process (MAP) under Closed-vessel Conditions', *Spectroscopy*, **13**(1), 1–13 (1997).
81. R. McMilllin, L.C. Miner, L. Hurst, 'Abbreviated Microwave Extraction of Pesticides and PCBs in Soil', *Spectroscopy*, **13**(1), 41–50 (1997).
82. I. Rodriguez, M. Santamarina, M.H. Bollain, M.C. Mejuto, R. Cela, 'Gas Chromatography–Atomic-emission Determination of Organotin Compounds in Marine Biomaterials After Microwave Extraction', *Spectroscopy*, **13**(1), 51–59 (1997).
83. M. Letellier, H. Budzinski, P. Garrigues, S. Wise, 'Focused Microwave-assisted Extraction of Polycyclic Aromatic Hydrocarbons in Open Cells from Reference Materials (Sediment, Soil, Air Particulates)', *Spectroscopy*, **13**(1), 71–80 (1997).
84. W. Mueller, J. Schubert, A. Benzing, K. Geiger, 'A Method for Analysis of Exhaled Air by Microwave Energy Desorption Coupled with Gas Chromatography–Flame Ionization Detection–Mass Spectrometry', *J. Chromatogr. B*, **716**(1–2), 27–38 (1998).
85. D. Amarasiriwardena, A. Krushevska, R.M. Barnes, 'Microwave-assisted Vapour-phase Nitric Acid Digestion of Small Biological Samples for Inductively Coupled Plasma Spectrometry', *Appl. Spectrosc.*, **52**(6), 900–907 (1998).
86. H.M. Ortner, C. Berger, O. Boese, P. Hoffmann, A. Dedik, K. Englert, G. Giebenhain, U. Sengutta, H.J. Staerk, 'The Importance of Vessel Materials for Ultratrace Analysis (Part 2)', *GIT Labor-Fachz.*, **42**(4), 390, 392–394 (1998).
87. J. Opydo, 'Voltammetric Analysis of Biological Samples and Soil After Wet Digestion Using Microwave Irradiation', *Chem. Anal. (Warsaw)*, **42**(1), 123–128 (1997).
88. V.O. Schmitt, J. Szpunar, O.F.X. Lobinski, 'Microwave-accelerated Sample Preparation for Speciation Analysis of Organotin Compounds in Biomaterials', *Can. J. Anal. Sci. Spectrosc.*, **42**(2), 41–44 (1997).
89. S.C. Chen, C.R. Lan, M.H. Yang, 'Determination of Aluminium in Serum by Electrothermal Atomic Absorption Spectrometry with Prior Sample Pretreatment by Microwave Digestion and Solid-phase Extraction', *Anal. Sci.*, **13**(3), 415–421 (1997).

90. H. Garraud, M. Robert, C.R. Quetel, J. Szpunar, O.F.X. Donard, 'Focused Microwave-assisted Digestion of Biological Reference Materials for the Determination of Trace Metals by ICPMS', *At. Spectrosc.*, **17**(5), 183–185 (1996).
91. M. Deaker, W. Maher, 'Low Volume Microwave Digestion for the Determination of Selenium in Marine Biological Tissue by Graphite Furnace Atomic Absorption Spectroscopy', *Anal. Chim. Acta*, **350**(3), 287–294 (1997).
92. I.V. Kubrakova, T.F. Kudinova, E.B. Stavnivenko, N.M. Kuzmin, 'Microwave Radiation as a Factor Intensifying Sample Preparation: Analysis of Samples with an Organic Matrix', *J. Anal. Chem. (Transl. Zh. Anal. Khim.)*, **52**(6), 522–527 (1997).
93. E.R. Filho Pereira, J.J.R. Rohwedder, M.A.Z. Arruda, 'Online Microwave Slurry Sample Digestion Using Flow Systems for the Spectrophotometric Determination of Iron in Seafood', *Analyst*, **123**(5), 1023–1028 (1998).
94. F. Li, W. Goessler, K.J. Irgolic, 'The Optimization of Microwave Digestion for Determination of Selenium in Human Urine by Flow Injection–Hydride Generation–Atomic Absorption Spectrometry', *Anal. Commun.*, **35**(11), 361–364 (1998).
95. H.-H. Wang, Z.-Y. Wang, 'Spectrophotometric Determination of Pb in Urine with Microwave Digestion and a Dithizone–Water Phase', *Guangdong Weiliang Yuansu Kexue*, **5**(6), 9–11 (1998) (in Chinese).
96. M.M. Gomez, T. Gasparic, M.A. Palacios, C. Camara, 'Determination of Five Selenium Compounds in Urine by Liquid Chromatography with Focused Microwave Assisted Digestion and Hydride Generation–Atomic Absorption Spectrometric Detection', *Anal. Chim. Acta*, **374**(2–3), 241–251 (1998).
97. I.R. Pereiro, A. Wasik, R. Lobinski, 'Determination of Mercury Species in Fish Reference Materials by Isothermal Multicapillary Gas Chromatography with Atomic Emission Detection after Microwave-assisted Solubilization and Solvent Extraction', *J. Anal. At. Spectrom.*, **13**(8), 743–747 (1998).
98. K. Wang, C. Hiruki, 'The Molecular Stability of Genomic DNA of Phytoplasma in the Witches'-broom-affected Paulownia Tissues after Microwave Heat Treatment', *J. Microbiol. Methods*, **33**(3), 263–268 (1998).
99. O. Safir, S. Fkih-Tetouani, M. Soufiaoui, K. Bougrin, R. Aquino, 'Microwave Extraction of Aerial Parts of *Zygophyllum Gaetulum*', *Ital. EPPOS*, **25**, 3–10 (1998).
100. S. Mendez-Alvarez, R.I.L. Eggen, 'A Rapid Microwave Method to Extract Plasmid DNA from Saccharomyces Cerevisiae Suitable for the Transformation of Escherichia Coli', *Biotechnol. Tech.*, **12**(8), 605–606 (1998).
101. N. Carro, C.M. Garcia, R. Cela, 'Optimization and Comparison of Two Microwave-assisted Extraction Procedures of Terpenic Compounds in *Vitis vinifera* Must Samples', *Spectroscopy*, **13**(1), 61–70 (1997).
102. M. Sharman, C.A. Honeybone, S.M. Jickells, L. Castle, 'Detection of Residues of the Epoxy Adhesive Component Bisphenol: a Diglycidyl Ether (BADGE) in Microwave Susceptors and Its Migration into Food', *Food Addit. Contam.*, **12**(6), 779–787 (1995).
103. M.A. De la Fuente, B. Carazo, M. Juarez, 'Determination of Major Minerals in Dairy Products Digested in Closed Vessels Using Microwave Heating', *J. Dairy Sci.*, **80**(5), 806–811 (1997).
104. C.P. Bosnak, K.W. Barnes, 'A Streamlined Flame Atomic Absorption Method for Animal Feed Analysis', *At. Spectrosc.*, **19**(2), 40–42 (1998).
105. A. Velasco-Arjona, J.A. Garcia-Garrido, R. Quiles-Zafra, M.D. Luque de Castro, 'Accelerated Microwave Pre-treatment of Meat Samples Before Automated Determination of L-hydroxyproline', *Spectroscopy*, **13**(1), 81–88 (1997).
106. R. Cozzani, D. Barchi, G. Catalli, A.L. Cinquina, 'Use of a Microwave System for Polyphosphate Determination in Meat Products', *Ind. Aliment.*, **35**(348), 515–516 (1996).
107. J. Maslowska, B. Swat, 'Microwave Mineralization of Confectionery Samples and Determination of Trace Quantities of Heavy Metal Ions in Mineralizates', *Zesz. Nauk. Politech. Lodz., Technol. Chem. Spozyw.*, **59**, 63–75 (1998).
108. Z. Chemat, D.E. Hadj-Boussaad, F. Chemat, 'Application of Atmospheric-pressure Microwave Digestion to Total Kjedahl Nitrogen Determination in Pharmaceutical, Agricultural and Food Products', *Analusis*, **26**(5), 205–209 (1998).
109. N.O.G. Jorgenson, R.E. Gensen, 'Determination of Dissolved Combined Amino Acids Using Microwave-assisted Hydrolysis and HPLC Pre-column Derivatization for Labeling of Primary and Secondary Amines', *Mar. Chem.*, **57**(3–4), 287–297 (1997).
110. A.V. Pereira, C. Aniceto, O. Fatibello-Filho, 'Flow-injection Spectrophotometric Determination of Acetylsalicylic Acid in Tablets After Online Microwave-assisted Alkaline Hydrolysis', *Analyst*, **123**(5), 1011–1015 (1998).
111. J.G. March, F. Grass, A. Salvador, 'Hydrolysis of Phytic Acid by Microwave Treatment: Application to Phytic Acid Analysis in Pharmaceutical Preparations', *Microchem. J.*, **59**(3), 413–416 (1998).
112. N. Jeyashoke, K. Krisnangkura, S.T. Chen, 'Microwave-induced Rapid Transmethylation of Fatty Acids for Analysis of Food Oil', *J. Chromatogr. A*, **818**(1), 133–137 (1998).
113. M. Franke, C.L. Winek, H.M. Kingston, 'Extraction of Selected Drugs from Serum Using Microwave Irradiation', *Forensic Sci. Int.*, **81**(1), 51–59 (1996).
114. A. Kovacs, K. Ganzler, L. Simon-Sarkadi, 'Microwave-assisted Extraction of Free Amino-acids from Foods', *Z. Lebensm.-Unters.-Forsch.*, **207**(1), 26–30 (1998).
115. M.J. Incorvia Mattina, W.A. Iannucci Berger, C.L. Denson, 'Microwave-assisted Extraction of Taxanes

from *Taxus* Biomass', *J. Agric. Food Chem.*, **45**(12), 4691–4696 (1997).
116. M.H. Akhtar, M. Wong, S.R.H. Crooks, A. Sauve, 'Extraction of Incurred Sulphamethazine in Swine Tissue by Microwave Assisted Extraction and Quantification without Cleanup by HPLC Following Derivatization with Dimethylaminbenzaldehyde', *Food Addit. Contam.*, **15**(5), 542–549 (1998).
117. M.H. Akhtar, L.G. Croteau, C. Dani, K.A. Elsooud, 'Development and Validation of Microwave-assisted Extraction of Fortified and Incurred Chloramphenicol Residues in Egg Albumin and Yolk', *Spectroscopy*, **13**(1), 33–40 (1997).
118. I. Jarvis, M.M. Totland, K.E. Jarvis, 'Determination of the Platinum-group Elements in Geological Materials by ICPMS Using Microwave Digestion, Alkali Fusion and Cation-exchange Chromatography', *Chem. Geol.*, **143**(1–2), 27–42 (1997).
119. S.W. Kingman, N.A. Rowson, 'Microwave Treatment of Minerals – A Review', *Miner. Eng.*, **11**(11), 1081–1087 (1998).
120. M.D. Luque de Castro, M.M. Jimenez-Carmona, 'Potential of Water for Continuous Automated Sample Leaching', *Trends Anal. Chem.*, **17**(7), 441–447 (1998).
121. M.T. Larrea, I. Gome-Pinilla, J.C. Farinas, 'Microwave-assisted Acid Dissolution of Sintered Advanced Ceramics for Inductively Coupled Plasma Atomic Emission Spectrometry', *J. Anal. At. Spectrom.*, **12**(11), 1323–1332 (1997).
122. I. Kubrakova, 'Microwave-assisted Sample Preparation and Pre-concentration for ETAAS', *Spectrochim. Acta, Part B*, **52**, 9–10, 1469–1481 (1997).
123. T. Yamaguchi, K. Onoe, 'Decomposition of Coal Under Microwave Irradiation', *Jpn. Kokai Tokkyo Koho*, JP 10 310,777 [98 310,777], 24 November 1998, Appl. 97/120,749, 12 May 1997.
124. W. Sievert, 'A Method for Decreasing Metal Ion Content in Chemical Compounds with a Dipole Moment or a Mixture of these Compounds by Microwave Distillation', *Ger. Offen.*, DE 19,719,720, 12 November 1998, Appl. 19,719,720, 9 May 1997.
125. M. Hippler, U. Sengutta, 'Open-vessel Microwave Digestion of Wear Metals in Lubricating Oil', *GIT Labor-Fachz.*, **41**(6), 645–646,648,651 (1997).
126. S. Dugenest, M. Olle, A. Ribes, M.F. Grenier-Loustalot, 'Chemical Characterization of Municipal Solid Waste Incineration Residue: Dissolution of Elements with a Microwave–Dilute Acid Digestion Technique Compared to Conventional Methods', *Analusis*, **26**(7), 256–260 (1998).
127. T. Soczka-Guth, H. Witteler, G. Deckers, G. Frank, K. Brehl, J. Lenze, H. Bonsel, R. Knauf, 'Dissolution of Polymers Bearing Acid Groups Using Microwave Radiation', *PCT Int. Appl.*, WO 98 55,534, 10 December 1998; *DE Appl.*, 19,723,854, 6 June 1997.
128. A.K. Das, R. Chakarborty, M.L. Cervera, M. De La Guardia, 'A Rapid Method for the Determination of Trace Element Impurities in Silicone Oils by ICPMS After Microwave-assisted Digestion', *Spectrosc. Lett.*, **31**(6), 1245–1253 (1998).
129. M. Thomsen, P. Kainrath, 'The Analysis of Coal Tar Pitch by ICP Optical Emission Spectrometry After Digestion in a Microwave Oven System', *At. Spectrosc.*, **19**(2), 60–61 (1998).
130. P.A. Tanner, L.S. Leong, 'Microwave Vacuum Drying of Marine Sediment Prior to Analysis for Carbon and Metals', *Am. Environ. Lab.*, **10**(6), 20–21 (1998).
131. M. Popp, W. Lied, A.J. Meyer, A. Richter, P. Schiller, H. Schwitte, 'Sample Preservation for Determination of Organic Compounds: Microwave Versus Freeze-drying', *J. Exp. Bot.*, **47**(303), 1469–1473 (1996).
132. P. Nischer, 'Determining the Water Content of Aggregate and Freshly-mixed Concrete by the Microwave Method', *Betonwerk Fertigteil-Tech.*, **64**(11), 60–62, 64–67 (1998).
133. J.G.P. Binner, I.A.H. Al-Dawery, 'Microwave Melt Texturing of Bulk YBCO Superconductors', *Supercond. Sci. Technol.*, **11**(11), 1230–1236 (1998).
134. Z.-P. Xie, X.-D. Fan, H. Yong, 'Accelerated Sintering and Phase Transformation of TiO_2 in Microwave Radiation', *J. Mater. Res.*, **13**(12), 3417–3422 (1998).
135. C.-Y. Wu, R. Lee, A. Benatar, 'Analysis of Heat Generation in a Conductive Polymer in Single Mode Microwave Welding', in *The 56th Annual Technology Conference of American Society of Plastic Engineers*, American Society of Plastic Engineers, 1034–1038, Vol. 1, 1998.

Multivariate and Hyperspectral Image Analysis

Paul Geladi
Swedish University of Agricultural Sciences, Umeå, Sweden

Hans F. Grahn
Karolinska Institute, Stockholm, Sweden

1 Introduction **1**
2 Images and Pictures in the Sciences **2**
3 Images and Computing **2**
3.1 The Digital Image 2
3.2 Image Storage Formats 3
4 Image-forming Analytical Equipment **4**
4.1 Plane Images 4
4.2 Volumes 5
5 Obtaining Multivariate/Hyperspectral Images **6**
5.1 Definitions 6
5.2 Variables 7
6 Statistical Analysis **10**
6.1 Pretreatments of the Data 10
6.2 Exploratory Study of Raw Images 11
6.3 Principal Component Analysis: Latent Variable Images 13
6.4 Multivariate Exploration and Classification 15
6.5 Regression and Calibration 16
6.6 Experimental Design for Imaging 16
7 Visualization of Results **16**
7.1 Latent Variable Images 16
7.2 Scatter Plots 16
7.3 Polygon Segmentation 17
7.4 Local Modeling 18
7.5 Hyperspectral Visualization through Score Images 19
8 Definitions **19**
Acknowledgments **22**
Abbreviations and Acronyms **22**
Related Articles **23**
References **23**

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)
Copyright © 2008 John Wiley & Sons Ltd

Multivariate image analysis (MIA) and hyperspectral image analysis (HIA) are methodologies for analyzing multivariate images, where the image coordinates are position (two or three dimensions) and variable number. Multivariate/hyperspectral images can have typical sizes 1024 × 1024, 512 × 512, 256 × 256, etc. and have between two and many hundreds of variables. The variables can be wavelength, electron energy, particle mass, and many others. Classical image analysis concentrates mainly on spatial relationships between pixels in a gray level image. MIA/HIA concentrates on the correlation of structure between the variables to provide extra information useful for exploring images and classifying regions in them. The many variables can be transformed into a few latent variable images containing condensed information. The sheer size of the data arrays necessitates visualization of raw data, intermediate data, model parameters, and residuals. When the images consist of continuous spectra having more than 100 variables, the name hyperspectral *is used and the emphasis of the analysis is often on the spectral interpretation.*

All physical techniques for measuring materials can be made into imaging techniques, describing not only a property but also its position in a plane or volume. All imaging techniques can be expanded to become multivariate/hyperspectral. Multivariate imaging is used in many fields of research, but for practical reasons a subdivision in the classes remote sensing, medical imaging, and microscopy (including macroscopy) can be made. In microscopy, MIA/HIA can be used to study materials and biological processes by optical, electron, and charged particle techniques.

1 INTRODUCTION

In analytical chemistry, titration and making solutions are examples of important operations that rely on homogeneous solutions containing a constant concentration of the analyte of interest. In nature and industry, gradients are important: all biological processes rely on heterogeneous environments, many manufacturing processes deal with heterogeneous mixtures and gradients, and geology deals with systems that are heterogeneous. The interpretation of data from homogeneous solutions of constant concentration or activity is rather simple. For a heterogeneous situation, a more complex representation is needed and some systems to be analyzed can only be understood by mapping in two, three, or even more dimensions. Early applications of this are found in geological and geochemical mapping[1] and early clinical X-ray photography, which is a chemical map of calcium phosphate concentrations in the body.[2]

The availability of electronics and computing has increased the use of mapping of heterogeneous systems, especially with the advent of satellite imaging, providing images as data files with intensities on which mathematical and statistical methods can be applied for error correction, contrast improvement, data extraction, and many other operations. Digital image analysis has evolved rapidly, but is mostly applied to one gray level image, where the spatial information in the images is explored.

It did not take long for scientists to realize that spectral information is available in imaging and to acquire data with many layers of gray level images. Satellite imaging soon began to use more than one wavelength band, electron microscopists used many X-ray wavelengths, and optical microscopists started using color and even more wavelengths outside the visual range such as the ultraviolet (UV) and infrared (IR).

This article is concerned with the use of multivariate spectral images combined with spatial exploration. Multivariate statistical methods and linear algebra are used to handle the huge amounts of data present in multivariate images. The large amounts of resulting data also are best displayed as images. Images need to be explored both in geometrical and spectral space. Classification and detection of gradients in images are important and calibration against external information is sometimes necessary.

Section 4 discusses image formation in two dimensions (planes) and three dimensions (volumes) and Section 5 deals with multivariate image formation. Four examples are used to show the vital aspects of MIA. Their explanation is spread out over Sections 5, 6, and 7. Section 6 concerns statistical analysis and Section 7 is about visualization, but both topics overlap.

2 IMAGES AND PICTURES IN THE SCIENCES

The human eye is an organ that can process huge amounts of complex information in parallel. Because of this, illustration has always played an important role in describing technological processes and scientific principles. In the ancient cultures, Egyptian frescoes and Greek pottery explain a lot about processes of agriculture, medicine, architecture, and manufacturing. Cave paintings are even older. They date from a time when written language did not exist and the only record of activities in those days is an interpretation of the drawings. During the renaissance, artists like Leonardo da Vinci and Albrecht Dürer perfected technical painting and drawing while scientists like Andreas Vesalius, Olof Rudbeck, and Carl von Linné[3–6] explained their work by technical drawings. Exploration of the seas and continents increased the accuracy and precision of map making. Microscopes were invented and the most obvious way of representing results of microscopic studies was visualizing them with geometrical accuracy and in color. Visualization using images has always been an integral part of scientific work and its use is constantly increasing.

Much of the early work in drawing and visualization was very tedious and was therefore soon replaced by photography, but the introduction of electronic imaging by television cameras with recording on video tape was necessary to make images practically useful for analysis. The possibilities were even more enhanced with the advent of digitization and storage of images as computer files, allowing automation in registration, storage, database management, and analysis. A good historical overview is given by Johnson.[7]

The color TV camera is nowadays replaced by multivariate and hyperspectral imaging devices and terabytes of data are collected daily in both airborne and medical imaging.

3 IMAGES AND COMPUTING

3.1 The Digital Image

Figure 1 gives one possible setup for collecting digital images. A video camera registers a scene through an objective and the video signal is then digitized to become data file. A digital image is an array of $i = 1, \ldots, I$ lines and $j = 1, \ldots, J$ columns where for each pair of coordinates (i, j) a gray value is given as seen in Figure 2. This is also called *raster data*. A similar definition for a 3-D (three-dimensional) image is given in Figure 3. This cube is also called the *hypercube*. The digital representation makes it possible to carry out complex and calculation-intensive transformations. Some transformations are meant to reduce noise or take away systematic errors and others are meant to improve the visual quality of the image. This is called *image processing*. It is also possible to reduce a digital image to fewer parameters. This is called *image analysis*. There is a large literature on these topics.[8–11] One example is particle sizing where images of bright particles on a dark background are collected. Thresholding on a certain gray level makes it possible to separate the bright particles from the dark background to create a binary image with white = particle and black = background (see Figure 4 for a simplified representation). In this binary image, the particles can be counted and measured, resulting in histograms of particle size, area, perimeter, etc. It is also possible to measure form factors (round, elongated, needle-like, etc.) and present these as histograms. Such operations can be

done visually, and the human eye is extremely good at recognizing products by their shape, color, texture, size, but accurate measurement requires precise, sensitive measurement systems, and huge amounts of material, and fast analysis demands require automation. Therefore, storage formats and database issues are an integral part of image analysis considerations. A multivariate image consists of more than one 2-D plane or image. These could be the same object imaged at different times or in different wavelengths as shown in Figure 5. A more formal definition follows later.

3.2 Image Storage Formats

There is still a flurry of different file formats for pixel-based images. Many are based on fulfilling specific demands and each has a good reason for existing. However, demands on industry standards for image file formats are becoming more important. Real multivariate image standards are still not well established, although one could consider the RGB (red–green–blue image) format for color images as a primitive multivariate format. There are also many formats for movies (image sequences), but much of advanced image analysis is done on still images and formats and some of these are mentioned here. Lossy compression is not considered appropriate when subsequent analysis is envisioned, but some future compression developments may be less harmful for calculations. A good overview is given in the literature.[12,13] A possible multivariate image format is shown in Figure 6.

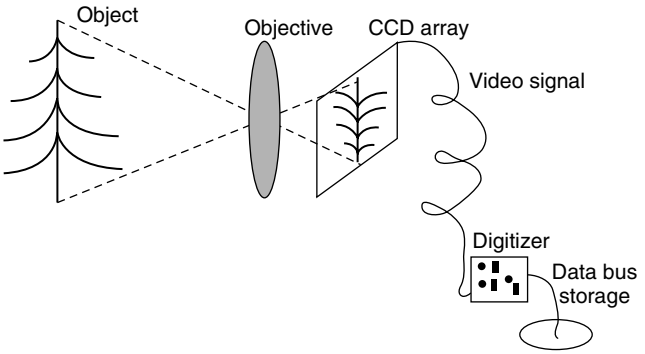

Figure 1 Classical imaging situation. An object is projected through an objective onto a CCD (charge coupled device) array. The video signal from the array is digitized and stored as a file.

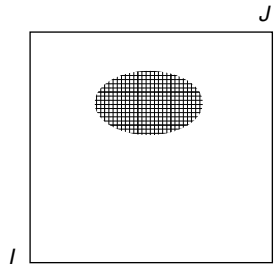

Figure 2 Digitized two-dimensional (2-D) image is an array of I lines (rows) and J columns. For each index pair (i, j) there is an intensity value called a *pixel*. Together the pixels form the gray value image.

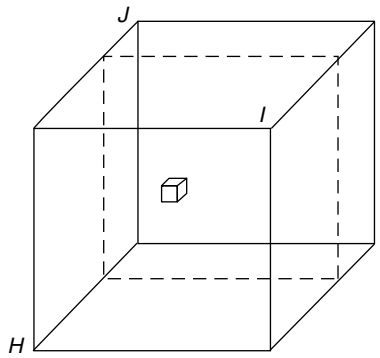

Figure 3 Digitized 3-D image is an array of H lines, I columns, and J planes. For each index triplet (h, i, j) there is an intensity value, the voxel.

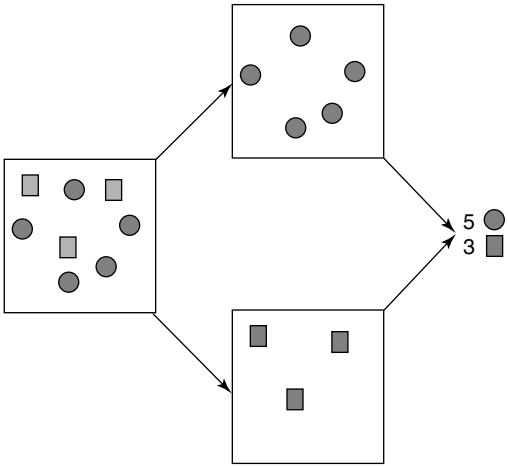

Figure 4 Image analysis consists of extracting classes of interesting objects (e.g. by thresholding or by size/shape recognition) and counting/measuring them.

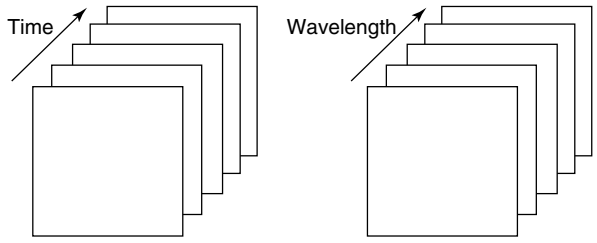

Figure 5 A multivariate image is a stack of images, each collected for a different variable. The same object can be imaged over time or in different optical wavelengths. For a large number of semicontinuous wavelengths, the term *hyperspectral* is used.

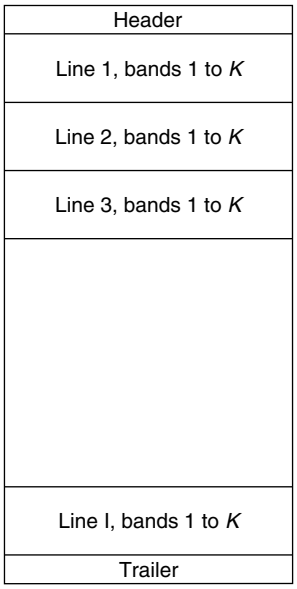

Figure 6 Schematic representation of a possible multivariate image file format.

The TIFF (tagged image file format) was created by Aldus, Microsoft, and NeXT and it can be used for raster images for black/white, grayscale, or color. There are four types of TIFF image: gray level TIFF images, RGB TIFF images, which contain colors based on the RGB color model, the CMYK (cyan magenta yellow and black image) TIFF images where separated image information is stored for printing color separations, and CIE (Commission Internationale de l'Éclairage) TIFF images, which contain dot-created image data in an independent unified color model. The TIFF format supports the LZW (Lemfel Zif Welsh) compression, which is a lossless compression technique. The TIFF format is an almost universal image format for photos and scanned graphics. TIFF contains the image and a relatively simple screen preview. The format is quite robust and is used extensively.

Joint photographic experts group (JPEG) is often used to show images for hypertext markup language (HTML) documents on the World Wide Web. All information about the image coding is kept in JPEG images. The lossy compression technique used for JPEG is very efficient in compressing the size of document, a process which removes ''unnecessary'' data in the image.

The raw format is a flexible format that is being used to transfer various documents between different programs and computer platforms. The format consists of some bytes that describe the color coding of the document. Each pixel is described in binary file format where 0 means black and 255 means white. The format also contains information that identifies the document and additional channels that can be included. The raw format needs no special byte for coding, and it can have more than 1 byte per pixel.

Recent developments in medical imaging make use of the possibilities that the large-scale image distribution offers through networking. The majority of medical imaging modalities such as magnetic resonance imaging (MRI), X-ray computed tomography (CT), and ultrasound all produce digital images. The images are stored in digital archives and can be accessed by various institutions, radiologists, and physicians.

In medical imaging, the two large organizations, American College of Radiology and National Electrical Manufacturers Association (ACR/NEMA), have defined a common file format and communication standard, the digital imaging and communications in medicine (DICOM) standard. The purpose of a common standard in medical imaging is to increase interoperability between biomedical imaging modalities such as ultrasound equipment, various modalities based on X rays, MRI with associated text, and measurements and waveforms in an electronic method of communication. Manufacturers are using the DICOM standards in the design of their imaging equipment.

Hyperspectral images are usually very large, usually up to several hundred megabytes. Since the size creates a practical complexity, the data are often stored in a compressed mode. A well-known raw data format is the environment for visualizing images (ENVI) format, where two files are stored for each dataset: one raw data file and a header file. This seems to be a general trend; the values are stored as long binary sequences with a header file to allow reconstruction of the image hypercube. A lot of software allows easy transformation of file formats, e.g. to the popular MATLAB format.

Compressed file formats such as compressed National Imagery Transmission Format (NITF), enhanced compression wavelet (ECW), and JPEG2000 are loaded into memory by a C++ library called *Geospatial Data Abstraction Library* (*GDAL*). Also the libecw2 library from ERMapper can be used for JPEG2000 images that have been embedded into other formats such as the NITF. There is a constant variation and addition that changes the file formats and these changes are usually driven by various applications.

4 IMAGE-FORMING ANALYTICAL EQUIPMENT

4.1 Plane Images

Plane images are formed by two major techniques: projection and scanning (see Figures 1 and 7). Hybrid techniques are also possible. A simple example of this is

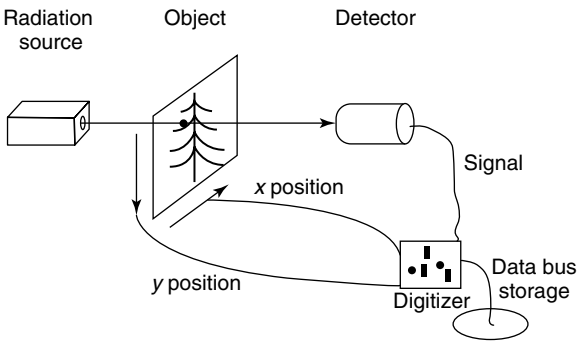

Figure 7 Scanning registration with one radiation source and one detector. The sample moves and the coordinates and signal from the detector are combined into a digital image.

the fax machine or document scanner, where the object is scanned and line after line of reflected intensities is measured to form the complete image. Satellite and airborne imaging mainly make plane images of curved surfaces. The slight curvature is usually ignored. Scanning is the major technique for satellite and airborne images, because it allows the use of one expensive, reliable, high-quality detector and because it is also very difficult to make objectives that work well for all remote sensing wavelength bands. In medical applications, X rays (direct radiography (DR) and computed radiography (CR)), ultrasound and MRI are the main 2-D techniques, but video imaging of wounds and skin disorders, microscopy imaging (pathology studies), and eye studies through charge coupled device (CCD) or complementary metal oxide semiconductor (CMOS) camera are also in use.

Many techniques allow the production of images of analytical samples. The easiest would be to simply use a color or black and white television camera to study inhomogeneous materials. Attempts have been made to extend the sensitivity of the television camera toward the UV and near infrared (NIR) regions of the spectrum. IR sensitive cameras are used to study temperatures and temperature gradients.[14] This is macroscopic imaging. In the laboratory, one might use macroscopic imaging in the visible (VIS), UV, or NIR regions on thin-layer chromatography (TLC) plates or electrophoresis plates. In this way automatic analysis can be done, avoiding observer fatigue and allowing a higher throughput.

Visualization in the sciences has usually been associated with microscopy. By putting a CCD camera on top of a microscope, it is possible to obtain digital microscopy images.[15,16] The increase in sensitivity of the cameras has made it possible to follow very weak but extremely specific fluorescence of certain materials.[17] Specific reagents have been developed to induce this fluorescence.

In electron microscopy, different techniques may lead to digital images.[18–21] In transmission electron microscopy, the electron transmission through a thin sample in vacuum is made into an image when the electrons are focused on a phosphorescent material by advanced electron optics. In scanning electron microscopy, thicker samples are scanned by a thin electron beam that follows a predetermined rectangular pattern. It is possible to detect reflected electrons, scattered electrons, energy loss in electrons, X rays, and cathodoluminescence (optical emission under electron excitation). The scanning pattern and the detector response can be combined into an image. A hybrid technique is scanning transmission microscopy. Other images can be made using scanning proton and ion beams in the techniques of proton (particle)-induced X-ray emission (PIXE) and secondary ion mass spectrometry (SIMS).[22–25] Sometimes with these techniques, the scanning is stopped and the beam is focused on a small area in the sample for a more accurate and precise analysis. This is the microprobe technique.[26,27] There are applications using beams of charged particles for image formation in biology, industry, materials science, geology, agriculture, forensic applications, etc. Most of these techniques can only take small samples, because they all need a vacuum to work properly and because the electron and ion optics have a small numerical aperture.

Acoustic signals can also be used for imaging materials.[28] A special imaging technique with very high spatial resolution is scanning force microscopy.[29] This technique uses piezoelectric scanning of a probe that can be used to measure force and field parameters of a surface.

In summary, one might conclude that wherever there is a technique that allows study of a material by physical principles (such as interaction with radiation, mechanical vibrations, etc.) it is possible to devise an imaging technique based on projection or scanning.

4.2 Volumes

In airplane photogrammetry, stereoscopic or depth viewing is made possible by stereo photography. Some satellite systems also allow stereo images and depth information. In medical imaging, slices through solid objects are made by the technique of X-ray tomography, also called *CT scan*. By taking different slices, a volume can be constructed. A similar technique is applied to MRI, giving slices and volumes.[30,31] By using radioactive markers, the techniques of PET (positron emission tomography)[32] and SPECT (single photon emission computed tomography)[33] can give information from inside a volume. The general theory of tomography is discussed by Hermann.[34]

Stereoscopic imaging and mechanical slicing with measuring of the slices are used both in optical and electron microscopy.[35] An important new technique is

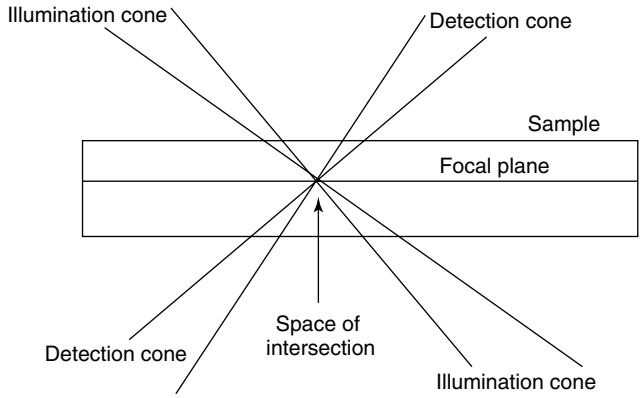

Figure 8 Schematic view of the principles of confocal microscopy. A restricted illumination cone is focused on a spot in the plane of focus. The restricted detector beam overlaps with it to give a small volume of intersection. By scanning in the plane and moving to different planes of focus, the whole sample volume can be imaged in three dimensions.

confocal imaging. Confocal microscopy is based on the fact that optical focusing is in a narrow layer and that unsharp images of the other layers interfere in a thick specimen (see Figure 8 for the principle). There are two major techniques of confocal microscopy: scanning and computing. The scanning technique is based on a narrow beam of radiation (illumination), often a laser and an aperture restricted detection as shown in Figure 8. The illumination and detector beam paths overlap by a very small volume that can be moved around in the sample under study. By scanning the whole volume, a 3-D image is created. Scanning can be done either by laser scanning or by moving the sample around. This works especially well for fluorescence. The computing confocal technique makes use of 3-D blurring functions and deconvolution to make planes of sharp images in the volume. See, for example, the books by Kriete,[35] Shotton,[16] Corle and Kino,[36] and Stevens et al.[37] for more details. The confocal microscopy technique relies heavily on not having too many opaque objects in the sample.

In the SIMS technique, the ion beam sputters away layer after layer of material and this can be used to do depth imaging and to construct 3-D volumes. To represent and analyze 3-D volumes of voxels, see, for example, the books by Kriete[35] and Lohmann.[38]

5 OBTAINING MULTIVARIATE/ HYPERSPECTRAL IMAGES

5.1 Definitions

A multivariate/hyperspectral image is an image where each pixel or voxel is represented not by one gray value, but by a vector of gray values. From now on,

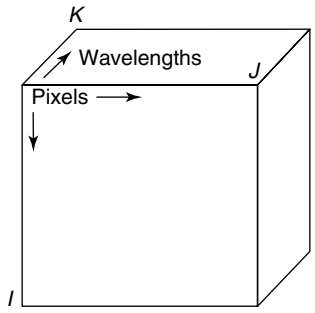

Figure 9 A multivariate 2-D image has pixel coordinates I and J and variable (wavelength, mass, electron energy, etc.) coordinates K.

"multivariate" and "hyperspectral" are referred to as *hyperspectral*. This simplifies the text without creating confusion. In many cases the vectors are spectra. With 2-D images, the hyperspectral image is a stack of gray level images, one for each variable (e.g. wavelength) (Figures 5 and 9). Congruence is very important for such a stack. This means that wavelengths for the same pixel represent the same position in the object studied. The easiest way of representing this is as a three-way or 3-D array where two of the ways are image coordinates and the third way contains variables. The indices are $i = 1, \ldots, I$ and $j = 1, \ldots, J$ for the pixel locations and $k = 1, \ldots, K$ for the variables. In satellite images, the variables are called *channels* or *bands*. For 3-D images this becomes more complicated because of the difficulty of showing four-way arrays (indices H, I, J for the voxel positions and K for the variables). Usually the parallelepipeds are shown under or beside each other as in Figure 10. Sometimes a system evolves over time and images are taken as a time series (just like a movie).[39,40] This may be necessary in order to follow a crystallization, an erosion process, or a biological process on a microscopic scale. Arrays as described here are generally called *N-way arrays*. A univariate 2-D image is a two-way array; a 3-D image is a three-way array, etc.

There are two classes of variables: homogeneous and heterogeneous. A stack of images taken at increasing wavelengths is a hyperspectral image with homogeneous variables. A stack consisting of a magnetic resonance image, an X-ray image, and an ultrasound image of the same object is a hyperspectral image with heterogeneous variables. For homogeneous variables, it is relatively easy to obtain congruent images in the stack. For heterogeneous variables, a transformation may be needed to achieve congruence of the images. Also, the intensity and spatial resolution may vary widely between techniques for heterogeneous variables.

A color image is multivariate since there are three color planes needed to construct it. For television these are red, green, and blue and for photography they are

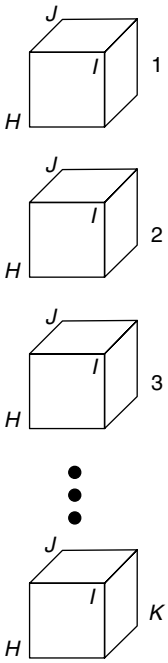

Figure 10 A multivariate 3-D image is a four-way array with indices H, I, J and K. It may be represented by writing K three-way arrays in a sequence.

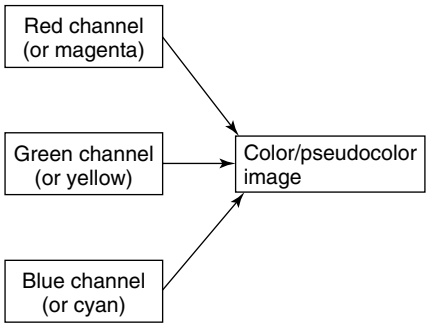

Figure 11 A color image on a screen is produced by mixing images in the red, green, and blue channels. If the data in the images are true red, green, and blue, almost true colors are produced. If any other combinations are used, a distorted color results.

cyan, magenta, and yellow. Color plays an important role in visualizing image data and analysis results. Figure 11 shows how a color image can be obtained.

5.1.1 Examples 1 and 2: Electron Microscopy and Secondary Ion Mass Spectrometry Data

Figure 12 shows congruent images obtained by electron microscopy. One image is for backscattered electrons (no energy loss) and the other one is for secondary electrons (energy loss by interaction with the sample). Figure 13 shows a multivariate image obtained by SIMS. There are nine 256×256 congruent images representing the

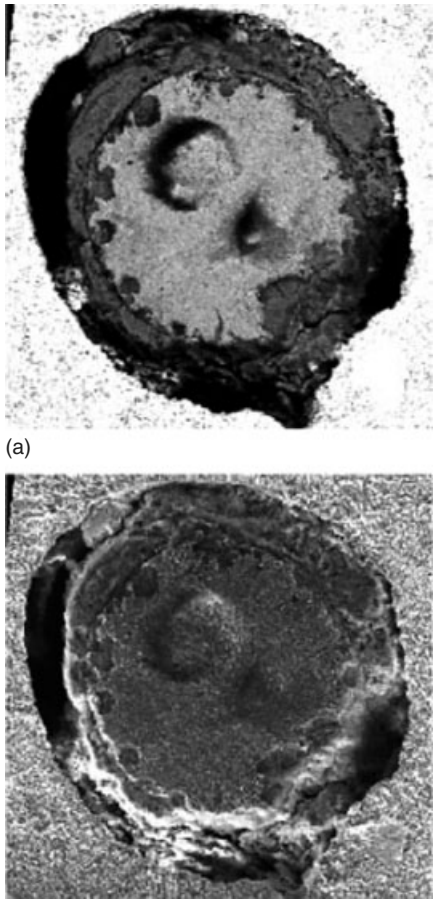

Figure 12 An image of a sample measured by scanning electron microscopy in two variables: backscattered (a) and secondary (b) electrons. The image size is 256×256.

intensity values for selected isotopes of the elements Na, Mg, Al, Si, Cr, Fe, Ni, Cu, and Zn. The sample is a corroded metal wire of 0.2-mm diameter. If the images of Figures 12 and 13 were made congruent by some transformation, the resulting multivariate $11 \times 256 \times 256$ image would have heterogeneous variables: two electron intensities and nine m/z values.

5.2 Variables

Many variable types can be used to produce 2-D and 3-D hyperspectral images. It is useful to consider three major fields of image analysis: remote sensing, medical applications, and analytical chemistry. The emphasis is on analytical chemistry, but for historical reasons some remote sensing and medical applications are also mentioned. Table 1 gives some variables used in imaging. The table is not complete and new developments appear

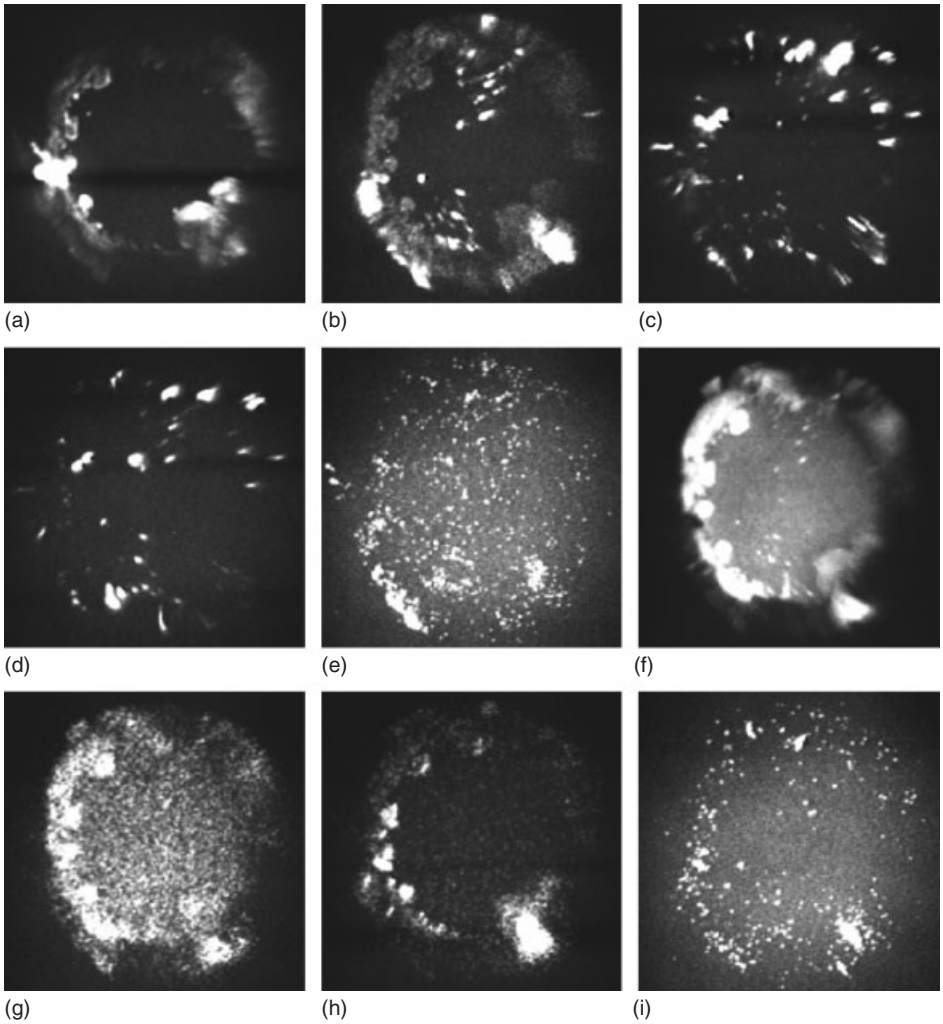

Figure 13 SIMS images for nine elements form a $9 \times 256 \times 256$ multivariate image stack. (a) Na, (b) Mg, (c) Al, (d) Si, (e) Cr, (f) Fe, (g) Ni, (h) Cu, and (i) Zn.

Table 1 Some variables used in imaging

Field	Technique	Variables
Remote sensing	Optical	Wavelengths 400–2500 nm
	Radar	Wavelength, polarization
Medical	X ray	Wavelength
	MRI	Pulse sequence, contrast chemical
	Acoustic	Frequency
	PET, SPECT	Decay
Analytical	Optical microscopy (diascopic)	Brightfield, dark field, phase contrast, etc.
		Polarization
		Wavelengths
	Fluorescence (episcopic)	Reagents, excitation/emission wavelength
	IR, Raman microscopy	
	Electron microscopy	Transmission/scanning/microprobe
		Electron energy/secondary/backscatter
		Brightfield/dark field
		X rays
		Cathodoluminescence
	Macroscopy	UV, VIS, NIR, IR, color
	Ion microscopy	m/z/X rays
	Scanning force microscopy	Force, magnetic field, conductivity, etc.

Table 2 Wavelength bands in some airborne/satellite imaging systems

Band/ channel	Wavelength (nm)			
	SPOT-XS	LANDSAT-TM	JERS-OPS	AVIRIS
1	500–590	450–520	520–600	Every 10 nm
2	610–680	520–600	630–690	390–2500 nm
3	790–890	630–690	760–860	224 bands
4	–	760–900	760–860[a]	–
5	–	1550–1750	1600–1710	–
6	–	10 400–12 500	2010–2120	–
7	–	2080–2350	2130–2250	–
8	–	–	2270–2400	–

[a] Bands 3 and 4 are a stereo pair. SPOT-XS = Système Probatoire d'Observation de la Terre. Image size 60 × 60 km; resolution 20 × 20 m. (Panchromatic = B/W (black and white, gray level or intensity image) resolution 10 × 10 m). LANDSAT-TM = Thematic Mapper. Image size 185 × 172 km; resolution 30 × 30 m. JERS-OPS = Japanese Earth Resources Satellite. Optical sensor 75 × 75 km; resolution 18.3 × 25.4 m. There is also 1.275 GHz radar. AVIRIS = Airborne VIS/IR Imaging Spectrometer. Typical 512 × 614 pixels and 20 × 20 m resolution, typical file size 70.5 MB of short integers.

continually, but it gives a good overview of the major concepts and their mode of use.

Airborne imaging from balloons and airplanes on black and white, color, and IR film is a very old technique and has given rise to the important field of photogrammetry. The first weather satellites were launched during the 1960s. Satellite and airborne imaging are used in agriculture, geology, hydrology, urban planning, climate studies, archeology, etc. Over the years different systems have been employed, and a good historical overview is given by Lillesand and Kiefer.[41] More details are also found in Sabins[42] and Horst.[43] Satellite imaging works continuously with automatic gathering of data for as long as the satellite works, which usually is a number of decades. Airborne imaging is based on scheduled flights of a shorter duration. VIS, NIR, IR (collectively called *optical*), and radar (or microwave) wavelengths are used. Table 2 gives the wavelength bands of some important systems capable of giving multivariate images. The optical systems are passive, depending on reflection of sunlight while the radar systems are mainly active, meaning that the exciting radiation comes from the satellite. Radar waves are insensitive to clouds and atmospheric phenomena and work during both day and night. Many optical satellites have orbit patterns that follow the sunlight. For the radar systems, wavelength and polarization can be varied to produce many imaging variables. Wavelength bands range from 7.5 mm (40 GHz) to 1000 mm (300 MHz) and a popular range is the K_u band: 12.5–18 GHz. More details on specific airborne and satellite systems can be found in Kramer.[44]

The multivariate information in satellite and airborne images has also been called *hyperspectral* since the late 1980s[45] and can be used to study chemical composition on the earth's surface and just a few examples are highlighted here. The combination of microwave and optical satellite date is described by Moran et al.[46] The detection of mineral content using optical data is found in Deroin et al.[47] The use of high resolution airborne images for studying vegetation with respect to minerals is described in Roberts et al.[48] and in Van der Meer and Bakker.[49] These applications are really chemical analysis from a distance.

Imaging in medicine began when Konrad Röntgen discovered the X rays. Almost immediately (1905) the technique was introduced into diagnostic work. The technique relies on the fact that various tissues in the body absorb the X-ray beam differently. If the beam that passes through the body is trapped by a film, various gray levels will be generated, each related to the specific part of the body that has been X-ray-exposed. The technique has been film-based since the beginning and the technical procedure has not changed much over the years. With the introduction of the CR technique in the 1970s, film was replaced by photostimulable phosphor radiographic systems, which digitized images in scanners. A new era in conventional medical imaging began[50]: images were now visualized with adjustable gray levels by a monitor and image processing tools became available. A prominent tool is edge detection, because edges characterize object boundaries. For diagnostic work, many image processing techniques received slow acceptance by radiologists because of problems with interpretation. Today, many processing tools are accepted and used frequently.[51]

In medical applications, images can be made multivariate/hyperspectral by using different X-ray wavelengths, different MRI pulse sequences, different contrast chemicals, different ultrasound wavelengths, color television cameras (extended by UV and NIR wavelengths) in wound and skin studies, and in endoscopic studies.

With the radiotracer materials used in PET and SPECT, time-dependent diffusion and decay of the tracers give a temporal multivariate sequence. Sometimes images from different studies may be combined to form a multivariate/hyperspectral image with heterogeneous variables, e.g. MRI images with X-ray images and acoustic images. The advantage of this is that some techniques give better spatial resolution, while others give better contrast between tissue types, but with a less spatial resolution.

In parallel with the CR development the CT technique evolved. This imaging technique is, in principle, an X-ray method based on the generation of cross sections of an object from several images of its transaxial projections. The projection is basically a shadowgram obtained by radiating an object. The CT technique is nowadays considered a standard diagnostic technique for examination of any part of the body.

The MRI technique uses an image reconstruction technique based on the Fourier transform. The images received reveal the proton density in tissue. It is increasingly being used in medical imaging.[52]

In optical microscopy, different techniques exist to make images of the same sample: brightfield (the usual technique), dark field, phase contrast, and polarized light. The illumination can be white or different wavelengths from the UV and up to the NIR. The sample can be studied in diascopic mode (transmission) or in episcopic mode (reflection or fluorescence). The latter mode makes the use of fluorescence possible. Many combinations of exciting and emitted wavelengths may be chosen. This is improved by the use of fluorescent reagents that specifically mark certain regions in a tissue.

The special microscopic techniques are IR and Raman microscopy, where a sample is studied in VIS light under the microscope and a useful region for study is found. Then, the selected region is studied by detailed laser Raman or IR spectroscopy. By scanning, image formation in IR and Raman wavelengths is possible.[53]

In electron microscopy, many modes of operation are available and each of these provides many modes of detection. Usually, a distinction is made between transmission, scanning (reflection), and microprobe modes. Also, combinations are possible such as scanning transmission. The phenomena detected are brightfield transmission, dark field transmission, backscattered electrons, secondary electrons, electron energy, X rays, and optical emission. All these phenomena are usually affected by the energy of the exciting electrons, giving extra variables to be considered.

In ion microscopy, the SIMS technique allows direct imaging of ions by quadrupole mass spectrometry. A better mass resolution is obtained with time-of-flight mass spectrometry (TOFMS), giving fractional masses and the possibility of studying organic materials and isotopic composition. In PIXE, the charged particles (usually protons) give X-ray emission in different X-ray wavelengths that are specific for certain elements.

More suggestions on how to make imaging methods hyperspectral are presented in Geladi and Grahn[54] and in Grahn and Geladi.[55] In principle, all imaging methods can be made hyperspectral. With this whole plethora of possible modes of measuring samples and selecting a detected variable, there is rarely one image that provides the solution to the problem at hand. Most of the images that can be produced contain correlated information on the sample studied and, therefore, it is necessary to treat them together to obtain the maximum information from the images.

5.2.1 Example 2: Secondary Ion Mass Spectrometry Data

The SIMS example of Figures 12 and 13 has images of the sizes $2 \times 256 \times 256$ and $9 \times 256 \times 256$. It has been described in part in Van Espen et al.[56] The technique used is described by Van Espen and Janssens.[57] The sample is a corroded string of an antique musical instrument. The original problem was a study of the metallurgical techniques used and the nature of the corrosion. For this purpose, a part of the string was fixed in a polymer medium and polished flat. The SIMS measurements were made in a Cameca 3f instrument. The sample is excited by a thin Ar^+ beam. The Ar ions produce secondary ions on impact and these secondary ions enter a mass spectrometer. The instrument has an electrostatic and a magnetic sector for mass to charge (m/z) filtering. The m/z filtered ions are projected in focus on a fluorescent target and the intensity image from this target is read by an analog video camera and digitized. Two variables of the $2 \times 256 \times 256$ image were secondary and backscattered ion intensity, measured by scanning electron microscopy. The nine SIMS variables of the $9 \times 256 \times 256$ image were positive ion intensities at different m/z as in Table 3. The images are present as 256×256 pixel image files with 256 gray levels for each pixel. This means that the ion and electron intensities had to be transformed logarithmically and stretched to fit the smaller range of 0–255. Figure 13 shows the SIMS images. Some of them have artefacts in the background due to low ion intensities.

6 STATISTICAL ANALYSIS

6.1 Pretreatments of the Data

Those who produce images try to obtain the best possible raw images by using the instrumentation properly, but

Table 3 Mass spectrometry variables for the SIMS experiment; only the most abundant isotope is used

Element	m/z (rounded)
Na	23
Mg	24
Al	27
Si	28
Cr	52
Fe	56
Ni	58
Cu	63
Zn	64

errors cannot always be avoided. Any small errors that obviously do not belong in the images are removed. Images to be used for MIA/HIA are usually corrected for errors, missing data, and excessive noise before the MIA/HIA takes place. The whole gamut of univariate image processing techniques is available for this purpose. In some kinds of images, parts may be removed completely, e.g. by filling in zeros, because it is known in advance that they are not needed. Making images congruent or of the same spatial resolution is also part of preprocessing before performing MIA/HIA. Registration is the process of making images congruent by rotation, translation, stretching, and shrinking to make fixed points (landmarks) match. This can be done in 2-D or 3-D. These techniques of correction belong to univariate image analysis and there is an abundant literature on this subject.[8–11]

6.2 Exploratory Study of Raw Images

The separate images in a hyperspectral image can be visualized as gray level images or as pseudocolor images. Usually, the image data are stretched in a linear or nonlinear fashion to obtain the best contrast and intensity conditions for the human observer. Using triplets of images to create pseudocolor is a useful way of studying hyperspectral images, but in some cases there are too many combinations to be checked. Colors can be ''natural'' when using a color television camera, or completely artificial, e.g. when combining electron microscopy images into color composites. A technique often used in remote sensing is ratio imaging, where the ratio between two wavelength bands is used to create new images.

Scatter plots are quite useful because they show correlations and groupings in images. A scatter plot between two images is obtained by taking intensities for matching pixel positions and using them as horizontal and vertical positions in a plot. Figure 14 shows how this is done. There are many combinations of two images selected from a hyperspectral image and looking at all combinations is not always possible. It is also possible to make 3-D scatter plots. These can be made visible in 3-D and rotated or projected to 2-D on three surfaces of a cube. Scatter plots of raw images sometimes show useful information about the individual images in a stack. Preprocessing for scatter plots is done by stretching or shrinking the intensities to the range of 0–255, or any other convenient range.

The scatter plot of the backscatter and secondary electron images in Figure 12 is shown as Figure 15. The $256 \times 256 = 65\,536$ pixels cannot be shown as individual points, so they are given as densities of points: dark = low density and bright = high density. It can be seen that there is a clustering of pixels and that there are *outlier* pixels. There is also an artefact from intensity stretching in the secondary electrons image. This artefact could not be observed by studying the images themselves.

The clusters can be used to define classes in the scatter plot and these pixel classes can be projected back to image

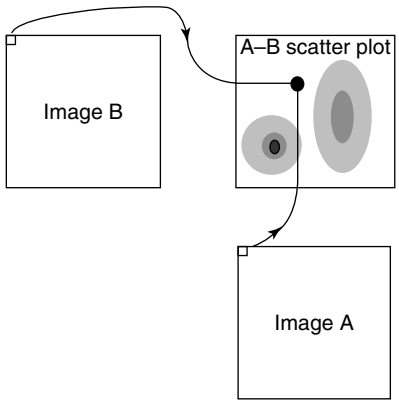

Figure 14 A scatter plot can be constructed using intensities of matching pixel positions in two different images. This is shown here for the first pixel pair. When the points have been drawn for all pixel pairs, the scatter plot often contains clusters of points.

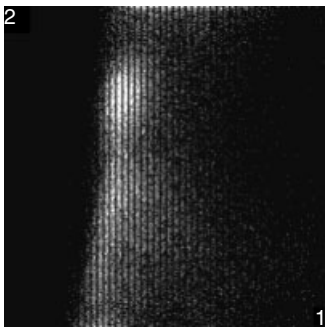

Figure 15 The scatter plot of the backscattered and secondary electrons images shows a clustering of points. Point densities within pixels are shown as gray levels. The vertical line pattern is a result of intensity stretching.

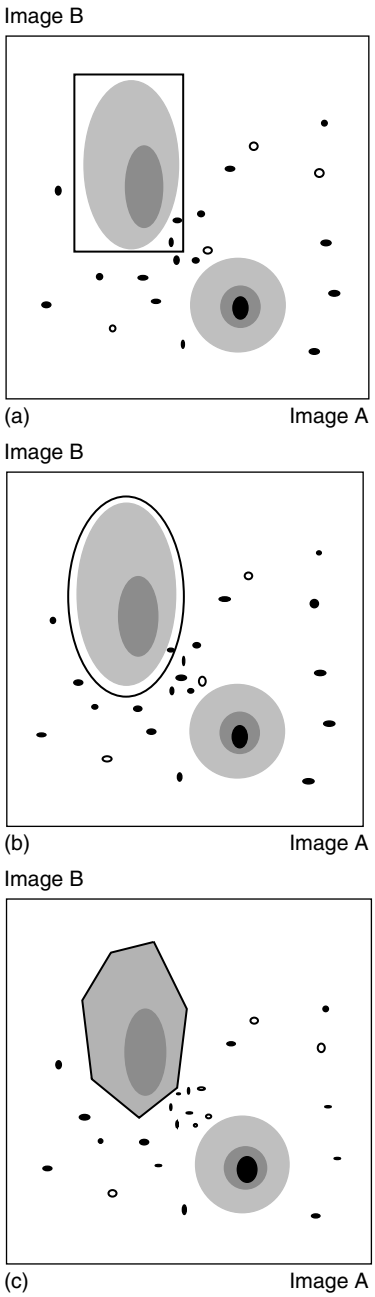

Figure 16 A number of classifiers can be used on pixel clusters in a scatter plot: the box classifier (a), the ellipsoid classifier (b), and the polygon classifier (c) are shown here.

space for inspection. Figure 16 shows some methods of selecting subsets of pixels in scatter plots. The first method is the box classifier using rectangles or parallelepipeds (for 3-D scatter plots). The box is easily defined, but may not always be of the correct form. The second method shown in Figure 16 is the ellipsoid classifier. It is based on the assumption of multinormal distribution of the pixels in a class. The ellipsoids are class membership probability ellipsoids. The most versatile is the polygon classifier. It can be objectively adapted to perceived shapes of the pixel classes and improved interactively to most shapes. For 3-D scatter plots, the box classifier is easily defined while the ellipsoid classifier is more difficult and the polygon method becomes quite awkward. When using polygons, they are filled to define a contiguous area, called *mask* or *class mask*.

Figure 17 shows some results of a rough classification. The classes are defined interactively by a polygon

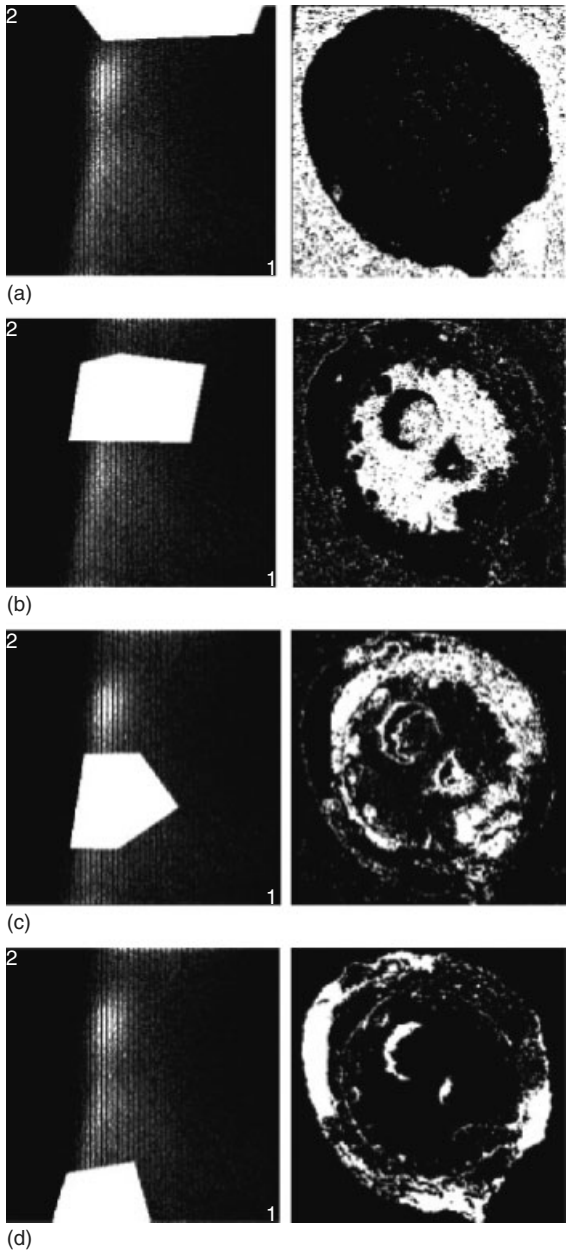

Figure 17 A rough partition of the scatter plot of Figure 15 in four classes and back projection of the classes to image space. The classes are (a) background, (b) core, and (c, d) two corrosion product classes.

classifier. Compared with box classifiers or ellipsoid classifiers, the polygon classifier gives a more realistic class boundary, since the true classes are rarely rectangular or ellipsoidal. It is also possible to define classes in the image and to find out where they fall in the scatter plot. The backscattered electrons (vertical axis in Figure 15) are the main contributors to the clustering, but there is also a small contribution from the secondary electrons (horizontal axis in Figure 15). This gives a small improvement over just using segmentation in one image. The major regions of the scatter plot are background, core, and two corroded regions. Some misclassification always occurs, but the main feature to look for is larger contiguous regions belonging to the same class. Single pixels and smaller regions may be misclassifications. With only two variables, this is all the information that can be extracted and there is only a little bit on chemical composition, where the backscattered electrons may reflect the atomic number. The classes are based on interaction of the sample with electrons. The SIMS images (Figure 13) may be able to give more information on chemical composition. It is always possible to refine the classification in Figure 17, but that is not shown here.

6.3 Principal Component Analysis: Latent Variable Images

The problem with hyperspectral images described above is that there are many variables to be checked and that there is usually a high correlation between them. Therefore, decorrelation is needed. This can be done easily by principal component analysis (PCA).[54,55,58–60] PCA calculates new variables, called *latent variables* or *scores*, thereby going from K to $A < K$ image variables. These are orthogonal and therefore they show no correlation. The latent variables also form images and therefore these are called *latent variable images* or *score images*. Another interesting property of scores is that

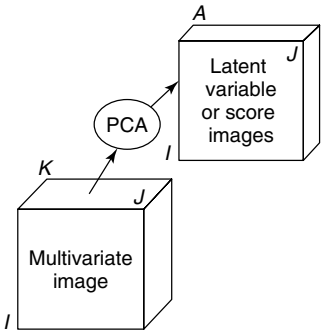

Figure 18 With PCA of an $I \times J \times K$ multivariate image, K image variables become $A < K$ latent variables or scores. The score or latent variables are orthogonal and explain much of the variability in the original multivariate image, leaving out some of the noise.

a few latent variables describe most of the variability in the hyperspectral image and that the remaining ones show little variability and can be considered as noise and discarded. This means that a data reduction is obtained by going, for example, from 10 variables to 3 latent variables explaining over 90% of the original variability. The three latent variables obtained in this way can easily be studied in pseudocolor images or in scatter plots called *score plots* (Figure 18). Alternatives to PCA may be based on curve resolution or factor analysis and can be found in the literature.[61,62] Also, hybrid algorithms using spatial information have been tried.[63]

6.3.1 *Example 2: Secondary Ion Mass Spectrometry Data*

The $9 \times 256 \times 256$ SIMS hyperspectral image was subjected to PCA analysis. The images were used as shown in Figure 13, stretched to fill the 0–255 intensity range. PCA was carried out without mean centering or rescaling of the nine variables. The results of the analysis are shown in Table 4. With four principal components, over 95% of the original sum of squares is explained. Figure 19 gives RGB composites of components 1, 2, 3 and 2, 3, 4 using the principle of Figure 11. These color images form a condensed representation of the main information in the nine original images of Figure 13. The colorings in the color composites give a good subjective overview of the different classes in the sample. The main constituents of these images are background, metal core, and corroded regions around the core. The background and metal core are homogeneous, while the corroded parts are very heterogeneous. Figure 20 gives the loading plots for the SIMS images. These explain the role of the variables in the different components. The first component has strong contributions from ⟨Zn,Cr⟩ and ⟨Fe,Ni⟩ while the second has strong contributions from ⟨Fe,Ni⟩. Fe and Ni are the major elements of the core of the sample. Zn and Cr seem to be a surface contamination, probably from mechanical processing. Other groupings are ⟨Al,Si⟩ and ⟨Na,Mg,Cu⟩. Al and Si are usual in the earth's crust and are an indication of contamination by sand.

6.3.2 *Example 3: The Visible/Near Infrared Data*

Images were collected with a PbS camera connected to a low-magnification microscope.[54,64] The wavelengths were in the VIS and NIR regions. The setup uses interference filters and a ring of optical fibers to give a monochromatic illumination of the object under study. An overview of wavelengths used is given in Table 5. The object is a piece of chinaware that has undergone painting and glazing in different layers. The total image is $12 \times 256 \times 256$. The original images were 512×512,

Table 4 Results of PCA analysis on the 9 × 256 × 256 SIMS image; percentages of the sum of squares (SS) and cumulative SS are shown

Component numbers	%SS	$\%SS_{cum}$
1	88.8	88.8
2	4.1	92.9
3	1.9	94.8
4	1.6	96.4
5	1.1	97.5
6	0.8	98.3
7	0.6	98.9
8	0.6	99.5
9	0.4	99.9[a]

[a] Because of rounding, this is the same as 100%.

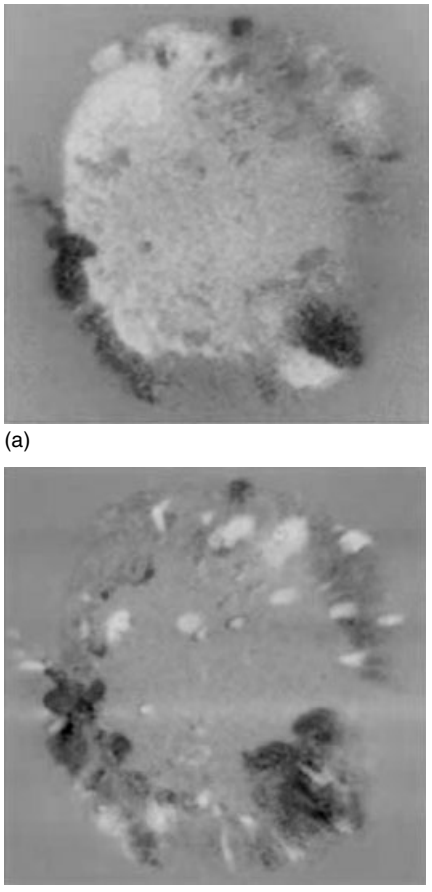

Figure 19 (a) A color composite of principal components 1 = red, 2 = green, and 3 = blue for the SIMS example. (b) The color composite of principal components 2 = red, 3 = green, and 4 = blue.

but because of border errors in the NIR region with the camera, only the central part was used. The images were averaged for noise reduction and some were amplified during this averaging. The image size is 9.5 mm × 9.5 mm.

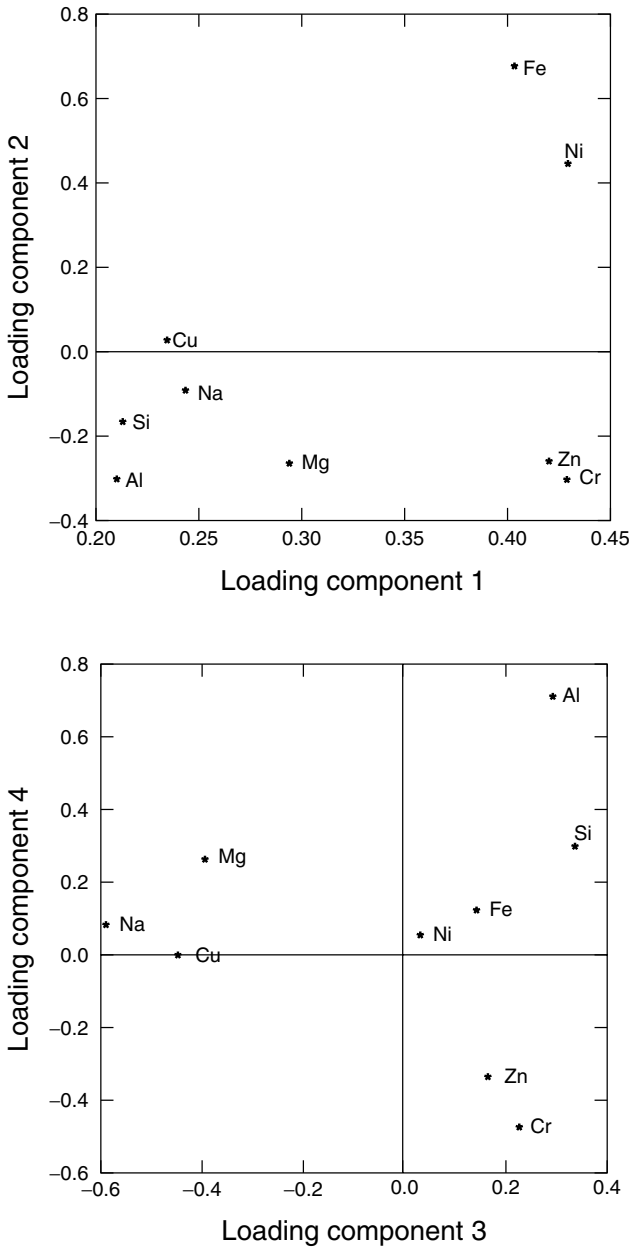

Figure 20 Loadings plots for the components 1–2 and 3–4 for the SIMS example.

Figure 21 is an RGB composite of wavelengths 460, 540, and 630 nm, which gives a reasonable impression of the colors. Table 5 also shows means and standard deviations. Two images have low standard deviation because of low transmission of the filter combined with low sensitivity of the camera. The VIS/NIR image was subjected to PCA with centering of each image, but without rescaling. PCA can give different results depending on the preprocessing of the data that is used. Image-wise (variable-wise) mean centering takes away differences in average intensity and compensates for the fact that some wavelengths may be

Table 5 Variables used in the macroscopic multivariate image

Number	Wavelength (nm)	Averaging[a]	Mean	Standard deviation
1	460	64/16	22	6
2	500	32/16	26	12
3	540	16/16	44	14
4	580	16/16	67	16
5	630	16/16	67	18
6	680	16/16	42	16
7	740	16/16	42	16
8	800	16/16	46	17
9	840	128/16	32	1
10	1010	128/16	32	4
11	1110	16/16	29	16
12	1200	16/16	44	17

[a] 16/16 means averaging over 16 images, 64/16 means averaging over 64 images, and increasing intensity by a factor of 4, etc.

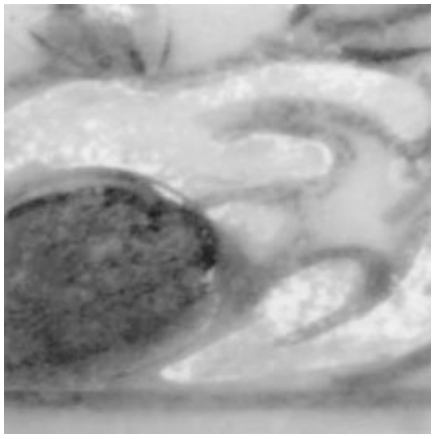

Figure 21 A color composite of the 460, 540, and 630 nm images of the VIS/NIR image.

darker or brighter on the average than others. Image-wise scaling gives more equal contrast over the images. It is also needed when different variables are measured in different units. The general use of these operations is to avoid that any image in a stack dominates the analysis because PCA is supposed to look for correlations or covariances. The PCA results can be seen in Table 6. Three components explain a large amount of the total sum of squares and five components explain almost all the sum of squares, giving an appreciable data reduction.

6.3.3 Example 4: Hyperspectral Near Infrared Data

The hypercube of eight pieces of polymer material is collected with an InGaAs camera and an LCTF (liquid crystal tunable filter) in the range 960–1662 nm with a band every 6 nm.[65] Illumination is with four quartz halogen lamps and the polymer pieces are positioned on a white ceramic tile in order to obtain reflectance/transflectance images.[66] This gives a hyperspectral image

Table 6 Percentage sum of squares results for PCA analysis on the variable-wise mean-centered VIS/NIR image $12 \times 256 \times 256$

Component numbers	%SS	$\%SS_{cum}$
1	84.7	84.7
2	13.1	97.8
3	1.4	99.2
4	0.4	99.6
5	0.2	99.8

(hypercube) of size $256 \times 320 \times 118$. A background and a reference image are also collected in order to correct for noisy background from the camera and nonhomogeneous illumination. The corrected image is subjected to PCA and three score images (99.63, 0.16, and 0.11% of the total sum of squares) are obtained. This is a huge data reduction coming from 118 wavelength bands. Of the eight pieces of polymer, four are PET (polyethylene terephthalate), three are PP (polypropylene), and one is PE (polyethylene). Three pieces are turbid (nontransparent): the PE piece and two PP pieces.

6.4 Multivariate Exploration and Classification

Exploration and classification in many wavelengths were first used in remote sensing by Ready and Wintz.[67] A good early overview is found in Lillesand and Kiefer.[41] In one of the chapters, scatter plotting of raw data, box classifiers, nonrectangular box classifiers, and Gaussian maximum likelihood classifiers are explained. There is also a description of how to rotate axes to obtain better discrimination. The concepts of region of interest and unsupervised classification are introduced.

An exploration of a hyperspectral image data set starts with obtaining histograms of the images and two-image scatter plots. After that PCA gives useful information. A table or plot of percentage sum of squares explained in each component (Tables 4 and 6) is used to select the number of components that is to be kept. The components are used for making score plots, but also loading plots are useful. They allow an interpretation of the role of the variables in the model. A vast number of methods exist for classifying in a hyperspectral image. This can be done automatically, or interactively. There are different degrees of interactivity. The results of the PCA for the VIS/NIR image are shown in Table 6. Since the images in the stack are correlated, four to five components suffice to explain most of the sum of squares.

Classification can be supervised or unsupervised. In the supervised case, one is looking for known classes. In the unsupervised case, one tries to find classes in the data without prior knowledge. Supervised classification

often starts from a region of interest. The multivariate properties of this region then define the class and all other pixels are tested for class membership.

6.5 Regression and Calibration

Regression between a multivariate image and external data is a very useful technique. In satellite and airborne imaging, images over an area of 1600 km^2 are available, e.g. a river delta. The ground truth for water quality must be studied on the ground and even 1 km^2 of ground truth data is quite difficult and expensive to collect. In this case, it would be convenient if a water quality parameter could be modeled from the spectral information in the hyperspectral image. The same situation occurs in medicine, where tissue quality or disturbances such as tumors and plaque have to be modeled from hyperspectral magnetic resonance images and the ''ground truth'' may be a small, expensive biopsy. Multivariate/hyperspectral image regression (MIR/HIR) is about relating desirable localized external information to the hyperspectral image. This information can be qualitative or quantitative. The regression techniques used are preferably based on latent variables. Examples are given in the literature.[64,68–75]

6.6 Experimental Design for Imaging

With many variables available for imaging, it may be useful to select the interesting ones more carefully. Leaving out really useless or noisy variables gives quicker image collection and a saving in data storage. In medical imaging, it means less strain on the patients. In a number of cases, screening by experimental design is possible in order to decide on useful imaging variables.[76] Alternative techniques can be used to find subsets of useful wavelengths.[77]

7 VISUALIZATION OF RESULTS

7.1 Latent Variable Images

Latent variable (score) images are constructed from the original images by methods like PCA, with different pretreatments. Latent variable images are constructed to have desirable properties. PCA gives orthogonality, allowing a condensation of the information in a few images. The use of score images also has noise reducing properties. Latent variable images can be studied as they are, or used as color composites (see Figures 19 and 22 for examples). The color contrasts in a color composite are subjective and they may be misleading or difficult to interpret in some cases. In other cases, they give a good impression of the different compositions in the image and their localization.

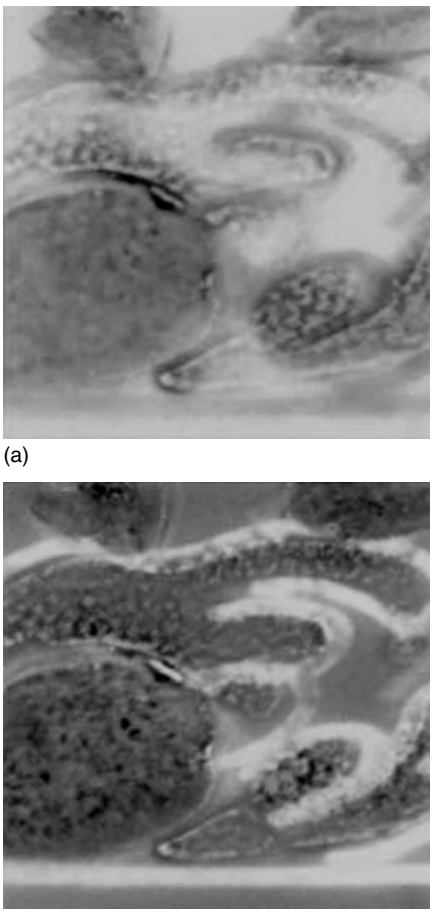

Figure 22 An RGB composite of (a) red = score 1, green = score 2, blue = score 3, and (b) red = score 2, green = score 3, and blue = score 4 for the PCA result of the VIS/NIR image.

7.2 Scatter Plots

Scatter plots of images are useful, as shown in Figure 17, but scatter plots of score images are even more useful. These are called *score plots*. The principle is the same as in Figure 14, but the data are more concentrated and the score images are orthogonal, giving better scatter plots. Figure 23 gives some score plots for the SIMS example. The plots are color coded for pixel density. The color coding is found in Geladi and Grahn.[54] The figure shows the score plots for principal components 1–4 in combinations 1–2, 1–3, 1–4, and 2–3. The score plots show two dense clusters corresponding to the background and core, and also some interesting smaller clusters corresponding to the corroded parts. An example of how to use such plots for classification is shown in Van Espen et al.[56] The loading plots allow a chemical interpretation of what is found in the score images and score plots.

An important loading plot for the VIS/NIR image is shown in Figure 24. It shows a good spread of the VIS

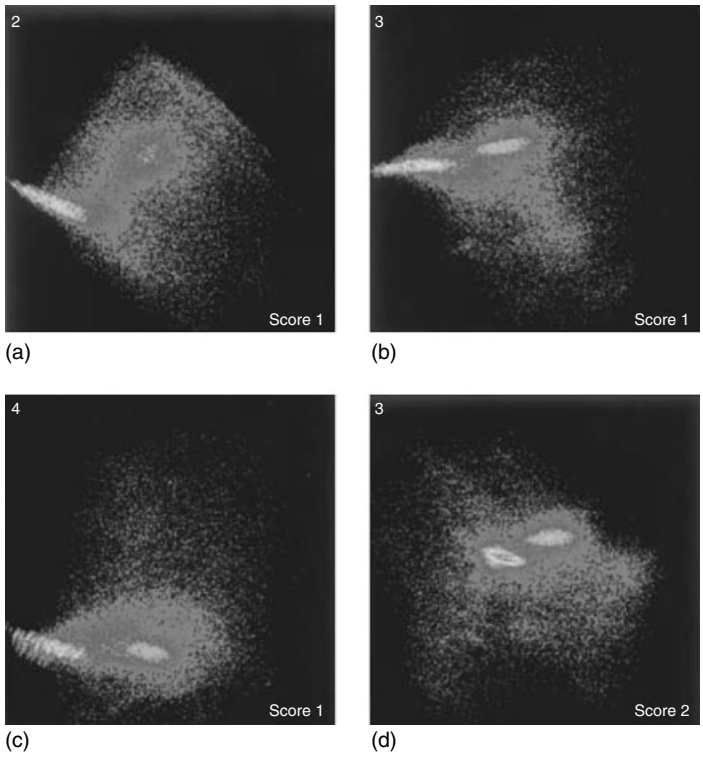

Figure 23 Score plots for the principal components 1–4 of the SIMS example. Combinations 1–2 (a), 1–3 (b), 1–4 (c), and 2–3 (d) are given. Dense clusters, smaller clusters and single pixels can be observed.

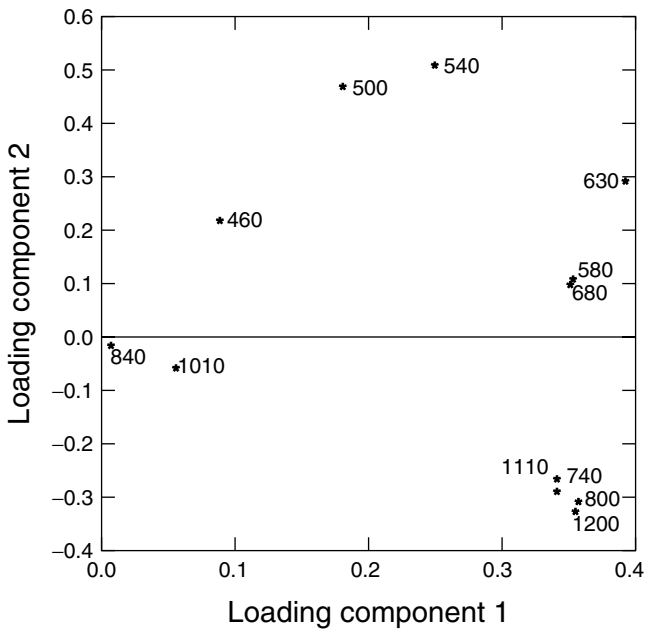

Figure 24 Loading plot of components 1 and 2 of the VIS/NIR image. Wavelengths in nanometer are indicated.

wavelengths. The NIR wavelengths are clumped together. The two NIR wavelengths with small standard deviation have almost zero loadings. The loading plots of the SIMS image allow a chemical interpretation, while the loading plot of the VIS/NIR image gives a spectral interpretation, which can lead to a chemical interpretation if sufficient spectroscopic evidence is given for the sample. Score images for components 1 and 2 for the VIS/NIR example are shown in Figure 25. The first score image is greatly influenced by the NIR wavelengths, while the second one gives a ''visual'' impression. The NIR radiation reacts most to the metallic ''gold'' coloring by giving total reflection. The score plots for the VIS/NIR example are seen in Figure 26. They show a useful number of pixel classes for components 1, 2, and 3. Component 4 seems to be influenced mainly by a gradient of outlier pixels.

7.3 Polygon Segmentation

Polygon segmentation was explained for scatter plots of raw images. The technique becomes even more powerful when latent variable images are used, because scatter plots of latent variable images show even better clustering of pixels and more separation of the clusters. A simple example is in order. Figure 27 shows the score plot of components 3 and 2 for the VIS/NIR image. A polygon classifier is used by putting a mask on the score plot to cover selected pixels. The pixels under this mask are then projected back to the image where they are highlighted in white (other colors are possible). The highlighted pixels

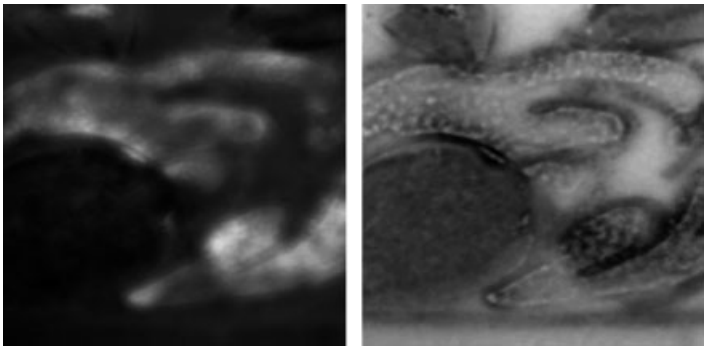

(a) (b)

Figure 25 The first (a) and second (b) component of the VIS/NIR image. These are score or latent variable images.

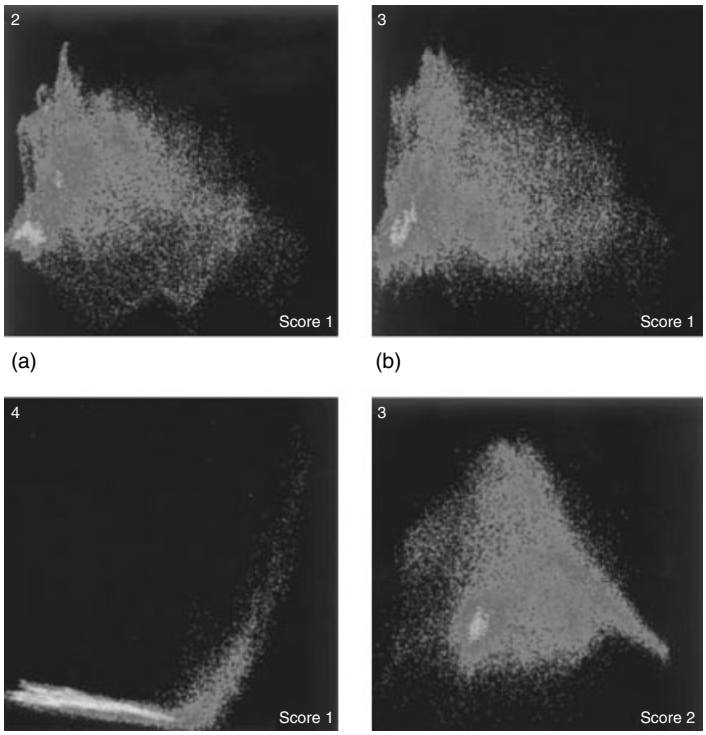

(a) (b) (c) (d)

Figure 26 Score plots for the principal components 1–4 of the VIS/NIR example. Combinations 1–2 (a), 1–3 (b), 1–4 (c), and 2–3 (d) are given. Dense clusters, smaller clusters, gradients, and single pixels can be observed.

are the *red* class, corresponding to the red dye used in painting the porcelain. The classification can be refined by modifying the polygon mask. In this way a number of classes can be defined, until most of the area in the image is classified. This classification corresponds to the different chemical compositions of the colored areas. In some regions, the coloring is homogeneous, while in other regions it is heterogeneous, indicating that many strokes of different colors were used to create a visual effect. The visual wavelengths are most important, since the dyes used were meant to give a visual impression. The NIR wavelengths also give a contribution by reacting with the metallic inclusions in the glazing. Miscellaneous references on classification in multivariate images are available.[78–88]

7.4 Local Modeling

Images can be quite large and one is not always interested in analyzing the whole image. There may be large regions that are of little or no interest or only small regions that are of real interest. In medical imaging, only the patient and not the air surrounding him/her is used and even inside the patient, only a lesser region of diseased

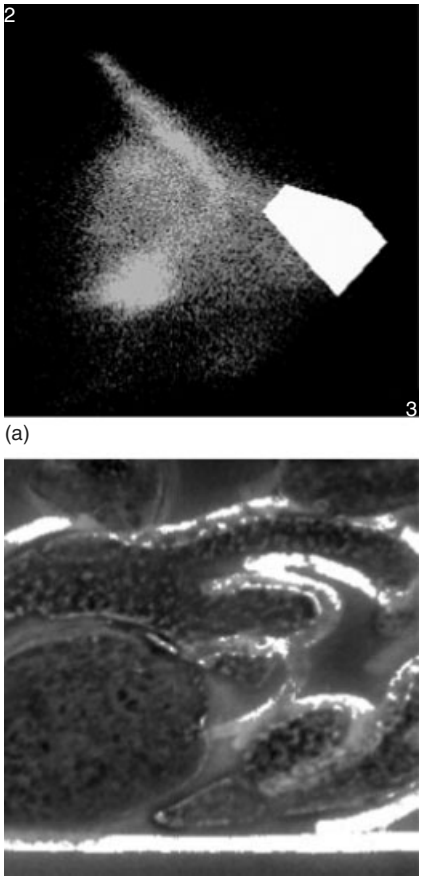

Figure 27 An example of polygon classification. (a) A polygon is drawn and filled to form a mask over a class of pixels in the score plot of components 3 (horizontal) and 2 (vertical). (b) The pixels under the class mask are highlighted in the score image of component 3. This is the class of red lines.

tissue may really be worth studying. In satellite imaging, different research groups focus on different surface-based phenomena: only wetlands, only water, only forest, and only clear-cuts.

Also in microscopic techniques, details may be more interesting than the whole image, e.g. a few cells in a transparent medium or only cell membranes. There exist different techniques of neutralizing uninteresting regions or selecting only smaller regions. All the data analysis models and visualization techniques mentioned earlier can also be used with these smaller local image regions.[89]

7.5 Hyperspectral Visualization through Score Images

Figures 28–30 are the first, second, and third score images of the corrected $256 \times 320 \times 118$ hypercube of the polymer pieces. In Figure 28, one of the pieces is slightly darker and this is a piece of PE. In Figure 29, three pieces are highlighted. The PE piece is very bright and two other pieces are also brighter. They are both PP pieces. All three pieces are made of nontransparent plastic. In Figure 30, four pieces are highlighted. These are all PE terephthalate. The information obtained is partly available by visual inspection of the pieces in white light, but the NIR spectra contribute unique spectral (= chemical) information. The classification is easily done without any prior need to know where to look for wavelength bands. The analysis shown here can be carried out in much more detail, but already the score images give an indication of the power of hyperspectral imaging in the NIR region.

8 DEFINITIONS

In image analysis, certain conventions are needed for working and communicating more easily. A proposed list is given below. More extensive and alternative definitions can be found in the literature.

- 2-D digital image: an array of intensities consisting of lines (rows) and columns, indexed by $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Each (i, j) pair or pixel has a number expressing intensity. The origin of the coordinate system is the upper left-hand corner. See also Figure 2. The image is also called *gray level, intensity,* or *black and white image*.
- 3-D digital image: an array of intensities consisting of lines (rows), columns and planes, indexed by $h = 1, \ldots, H$, $i = 1, \ldots, I$ and $j = 1, \ldots, J$. Each triplet (h, i, j) or voxel has a number expressing intensity. The origin of the coordinate system is the upper left-hand corner of the frontal plane. See also Figure 3.
- 2-D multivariate/hyperspectral image: an array of intensities consisting of lines (rows), columns, and variables, indexed by $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. See also Figures 5 and 9. Alternatively, see 2-D digital image, but each pixel is a vector (spectrum) of values instead of a scalar (intensity, gray value).
- 3-D multivariate/hyperspectral image: an array of intensities consisting of lines (rows), columns, planes, and variables, indexed by $h = 1, \ldots, H$, $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. See also Figure 10. Alternatively, see 3-D digital image, but each voxel is a vector (spectrum) of values instead of a scalar (intensity, gray value).
- 2-D multitemporal image: an array of intensities consisting of lines (rows), columns and times, indexed by $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $m = 1, \ldots, M$. The m-index forms a sequence in time. See Figure 5.
- Binary image: an image in which pixel (voxel) values are only black or white. The image is often used to represent classification results. See Figure 17.

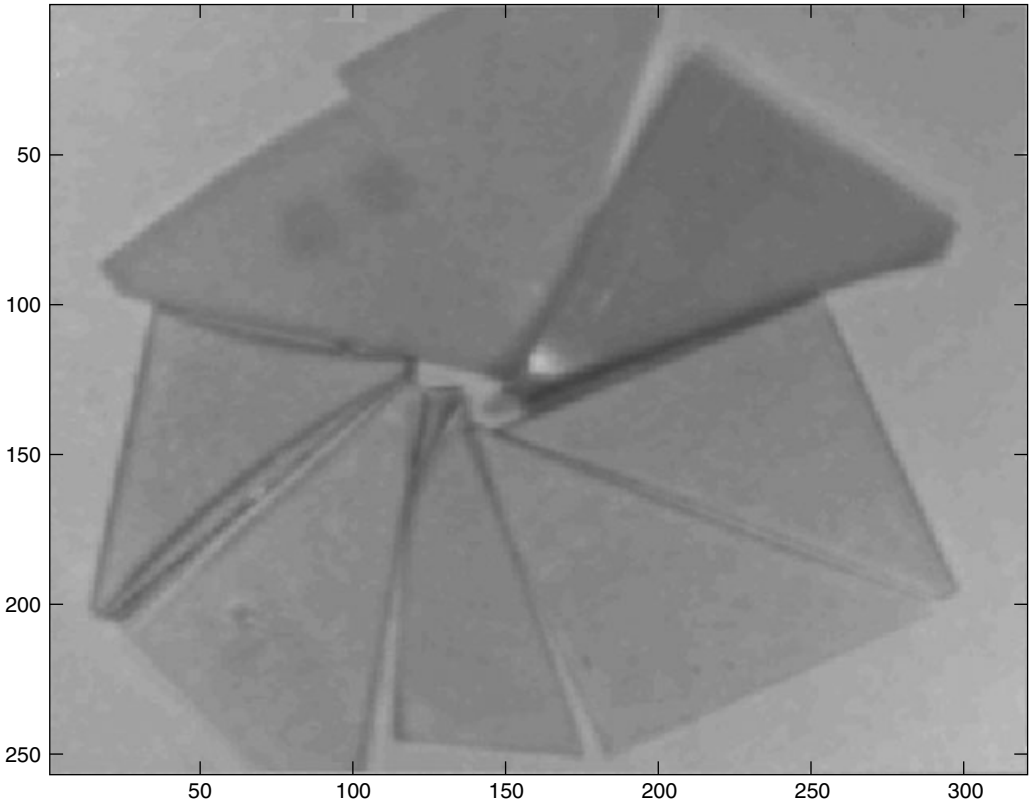

Figure 28 First score image of the NIR hyperspectral image. The piece of PE plastic is slightly darker.

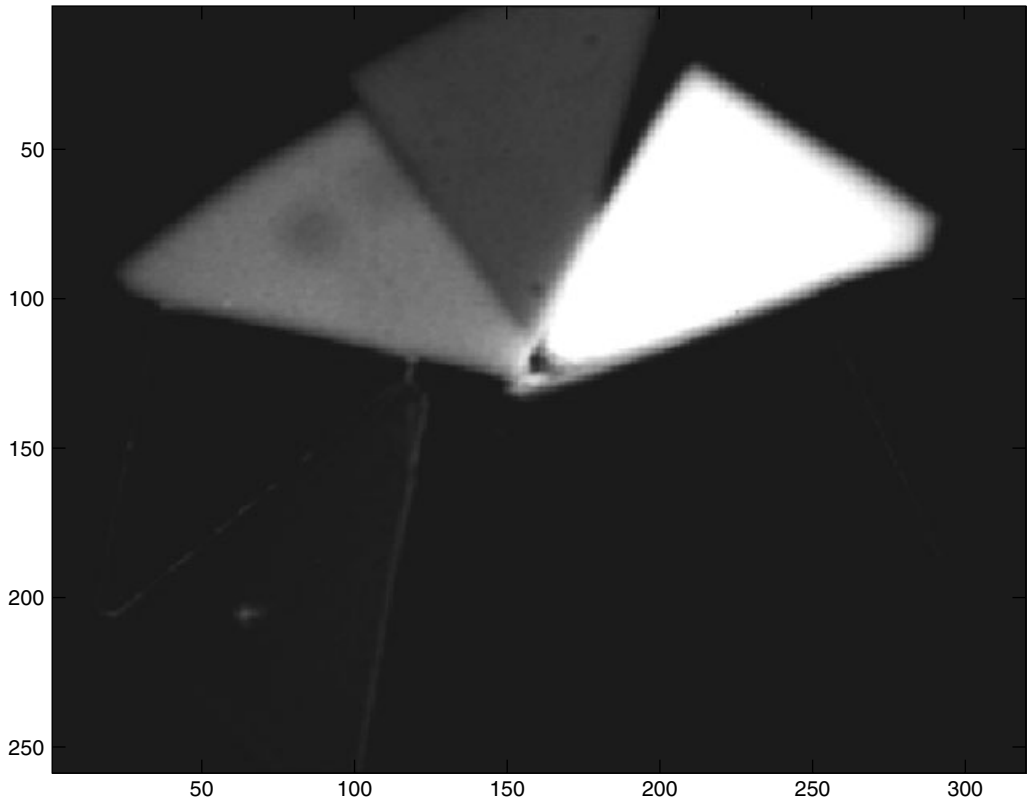

Figure 29 Second score image of the NIR hyperspectral image. The three nontransparent pieces of plastic are brighter.

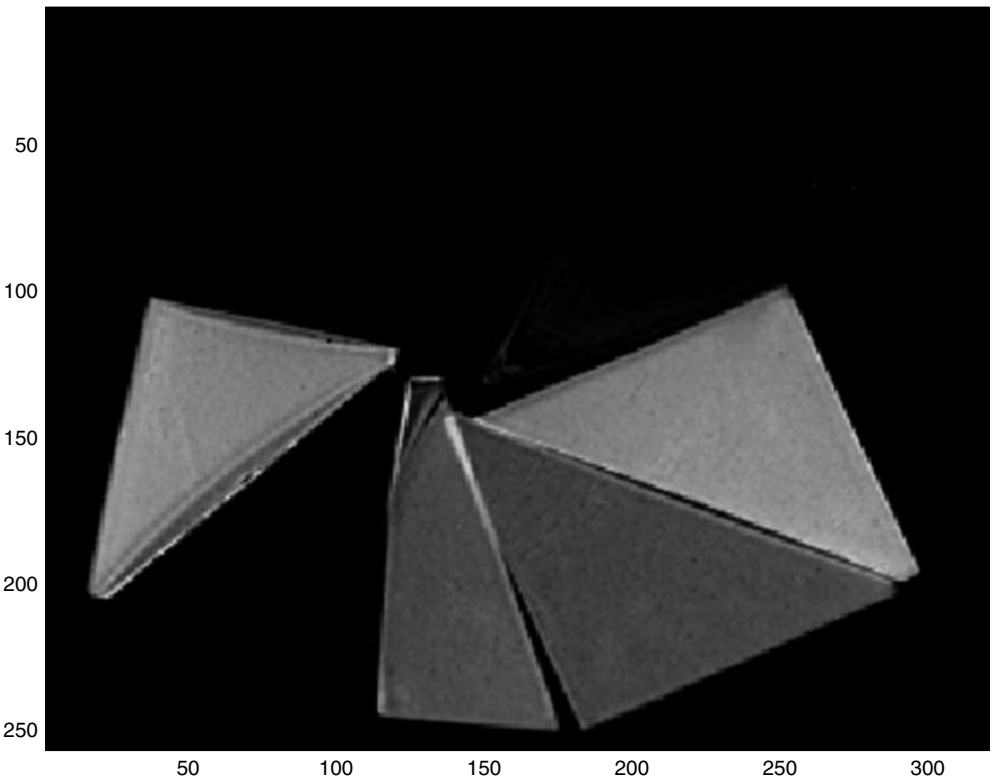

Figure 30 Third score image of the NIR hyperspectral image. The four PET pieces are brighter.

- Box classifier: a box is used to select a cluster of pixels in a scatter plot that form a class. See Figure 16.
- Confocal: a (dark field) technique of imaging in volumes by using a narrow illumination beam and a narrow return path to the detector. All slices in the volume are in focus. See Figure 8.
- Congruence: when measuring images using different variables or techniques congruence means a one-to-one position correspondence for the pixels in the different images. Congruence is an absolute condition for making MIA possible.
- Density coding: used to show pixel densities in scatter plots. See Figures 15, 23, and 26.
- Ellipsoid classifier: an ellipsoid is used to select a cluster of pixels in a scatter plot that form a class. See Figure 16.
- Focal plane scan: a system in hyperspectral imaging, when a complete image is taken at one wavelength. The hypercube is constructed by scanning over the wavelengths using electronically controlled filters.
- Hypercube: another name for the hyperspectral image with indices i, j, and k.
- Intensity resolution: many images have intensities between 0 and 255. Radiological standard is -1000 to 3095. A binary image has only 0 and 1 as possible intensities.
- Image analysis: a mathematical operation that transforms an image into desired information. Often a mathematical data reduction is involved.
- Image processing: a mathematical operation that transforms a digital image to a new digital image with (almost) the same spatial and intensity resolution, by taking away noise and/or errors.
- Latent variable image or score image: an image in which the values for the pixels are the results of a PCA calculation. Composites of latent variables images are shown in Figures 19 and 22. Figure 25 shows latent variable images.
- Line scan: a system in hyperspectral imaging, when a line of spectra is scanned. The hypercube is constructed by adding lines when moving the camera or object to be scanned; sometimes also called *push broom technique*.
- Loading plot: a scatter plot showing the importance of the original variables in the latent variable (PCA) model by the positions of the variables in the plane. See Figures 20 and 24.
- Mask: a filled polygon used to select pixels under it (also class mask). See Figures 17 and 27.
- MIA/HIA: as image analysis, but used on multivariate images and makes use of the correlation between the images.

- MIR/HIR: regression between multivariate images or between multivariate images and external information.
- *N*-way array: an array of numbers, where each number has *N* indices. A matrix is a two-way array. A multivariate image is a three-way array, or a four-way array if it describes volume.
- Object: the entity under study by image analysis.
- Point scan: a process in hyperspectral imaging, when a complete spectrum (often fourier transform FT) is taken in a point of an image and the hypercube is constructed by scanning the detector or object in x and y coordinates.
- Pixel: an element of a 2-D digital image. Pixels are square or rectangular. A pixel has four side neighbors and four corner neighbors.
- Polygon: a closed figure made of straight lines and corners, used to define a region of interest.
- Polygon classifier: a polygon is drawn around a cluster of pixels in a scatter plot to define a class. See Figure 16.
- Projection: the most often used imaging technique in photography, video cameras, transmission electron microscopes, etc. See Figure 1.
- Pseudocolor image: commonly used to denote a lookup table with assigned colors for gray level image.
- PCA: a technique of reducing many variables to a few meaningful latent variables. See Figure 18.
- RGB image: a color image in which three different congruent images are used to create color information in print or on a computer/video screen. The colors do not have to be true colors. See Figure 21 (almost natural colors), Figures 11 and 23 (artificial colors for latent variables).
- Raw images: images as they are collected by an imaging device. Before analysis these are usually pretreated or corrected.
- Region of interest: a region of pixels in an image or in a score plot.
- Registration: the process of making two images congruent by translation, rotation, stretching, shrinking, and warping with given fixed points (landmarks) as references.
- Sample: the entity under study by image analysis.
- Scanning: an imaging technique using a small radiation source and a small detector. The radiation source, the detector, or the sample can be moved. Many variations are possible. See Figure 7.
- Scatter plot: a plot of 2(3) vectors, matrices or other arrays in which the (horizontal and vertical) coordinates of the plotted points are given by the respective vectors, matrices, or arrays. For images, the intensities of corresponding positions in both dimensions are used as (horizontal and vertical) coordinates in the scatter plot. A scatter plot is a cloud of points in a plane (volume). With very many points, density coding is used instead of showing individual points. See Figure 15.
- Score plot: a scatter plot, where scores (latent variables) are used as input, see Figures 23 and 26.
- Scene: the entity under study by image analysis.
- Score image: latent variable image.
- Segmentation: a transformation of an image (univariate or multivariate) to a binary image containing objects for study and a background.
- Spatial resolution: has to do with the number of rows, columns (and planes) in an image.
- Tixel: a pixel of a multitemporal image.
- Tomography: a technique of reconstructing slices or volumes in 3-D objects from projections through the object.
- Variable: any physical method of extracting image information from an object. Each variable used creates a different image. The simplest example is different wavelengths.
- Voxel: an element of a 3-D image. Voxels are cubes or parallelepipeds. A voxel has 6 side neighbors, 12 edge neighbors, and 8 corner neighbors.

ACKNOWLEDGMENTS

The SIMS example was recorded by Helena Wouters of the University of Antwerp, Belgium. Piet Van Espen is acknowledged for providing the files and giving permission to use them.

ABBREVIATIONS AND ACRONYMS

ACR/NEMA	American College of Radiology and National Electrical Manufacturers Association
B/W	Black and White, Gray Level or Intensity Image
CCD	Charge Coupled Device
CIE	Commission Internationale de l'Éclairage
CMOS	Complementary Metal Oxide Semiconductor
CMYK	Cyan Magenta Yellow and Black Image
CR	Computerized Radiography
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DR	Direct Radiography

ECW	Enhanced Compression Wavelet
ENVI	Environment for Visualizing Images
GDAL	Geospatial Data Abstraction Library
HIA	Hyperspectral Image Analysis
HIR	Hyperspectral Image Regression
HTML	Hypertext Markup Language
IR	Infrared
JPEG	Joint Photographic Experts Group
LCTF	Liquid Crystal Tunable Filter
LZW	Lemfel Zif Welsh
MIA	Multivariate Image Analysis
MIR	Multivariate Image Regression
MRI	Magnetic Resonance Imaging
NIR	Near Infrared
NITF	National Imagery Transmission Format
PCA	Principal Component Analysis
PE	Polyethylene
PET	Positron Emission Tomography
PET	Polyethylene Terephthalate
PP	Polypropylene
PIXE	Proton (Particle)-induced X-ray Emission
RGB	Red–Green–Blue Image
SIMS	Secondary Ion Mass Spectrometry
SPECT	Single Photon Emission Computed Tomography
TIFF	Tagged Image File Format
TLC	Thin-layer Chromatography
TOFMS	Time-of-flight Mass Spectrometry
UV	Ultraviolet
VIS	Visible
2-D	Two-dimensional
3-D	Three-dimensional

RELATED ARTICLES

Forensic Science **(Volume 5)**
Scanning Electron Microscopy in Forensic Science

Polymers and Rubbers **(Volume 9)**
Thermogravimetry of Polymers

Remote Sensing **(Volume 10)**
Hyperspectral Remote Sensing: Data Collection and Exploitation • Imaging Spectrometry for Geological Applications • Land Cover Assessment and Monitoring • Polar Environments Assessment by Remote Sensing • Processing and Classification of Satellite Images • Satellite and Sensor Systems for Environmental Monitoring • Sea Ice Monitoring by Remote Sensing • Stellar SpectroscopyB • Temperate Forest Resource Assessment by Remote Sensing • Tropical Forest Resource Assessment by Remote Sensing

Surfaces **(Volume 10)**
Electron Energy Loss Spectroscopy in Analysis of Surfaces • Electron Microscopy and Scanning Microanalysis • Scanning Electron Microscopy in Analysis of Surfaces • Scanning Tunneling Microscopy/Spectroscopy in Analysis of Surfaces

Chemometrics **(Volume 11)**
Clustering and Classification of Analytical Data • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher

Electroanalytical Methods **(Volume 11)**
Scanning Tunneling Microscopy, In Situ, Electrochemical

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Signal Processing in Analytical Chemistry

Mass Spectrometry **(Volume 13)**
Secondary Ion Mass Spectrometry as Related to Surface Analysis

Raman Spectroscopy **(Volume 15)**
Raman Microscopy and Imaging

REFERENCES

1. C. Lyell, *Principles of Geology*, John Murray, London, 1868.
2. R. Grashey, *Atlas Typischer Röntgenbilder Vom Normalen Menschen*, Urban and Schwarzenberg, Berlin, 1905.
3. A. Vesalius, *De Humani Corporis Fabrica (1543)*, Culture et Civilisation, Brussels, 1964.
4. O. Rudbeck den Yngre, *Iter Lapponicum (1695)*, René Coeckelberghs Editioner, Stockholm, 1987.
5. C. Linnæi, *Iter Lapponicum (1732)*, Wahlström and Widstrand, Stockholm, 1965.
6. C. Linnæus, *Flora Lapponica (1737)*, Bokförlaget Rediviva, Stockholm, 1977.
7. S. Johnson, *On Digital Photography*, O'Reilly Media, Sebastopol, 2006.
8. R. Gonzalez, R. Woods, *Digital Image Processing*, Addison-Wesley, Reading, 1992.
9. A. Low, *Introductory Computer Vision and Image Processing*, McGraw-Hill, London, 1991.
10. W. Pratt, *Digital Image Processing*, John Wiley & Sons, New York, 1978.
11. A. Rosenfeld, A. Kak, *Digital Picture Processing*, Academic Press, New York, 2 Vols., 1979.

12. B. Furht, M. Milenkovic, eds. *A Guided Tour of Multimedia Systems and Applications*, IEEE Computer Society Press, Los Alamitos, 1995.

13. P. Andleigh, K. Thakrar, *Multimedia Systems Design*, Prentice Hall, Upper Saddle River, 1995.

14. S. Burnay, T. Williams, C. Jones, *Applications of Thermal Imaging*, Adam Hilger, Bristol, 1988.

15. S. Inoue, K. Spring, *Video Microscopy: the Fundamentals*, Plenum, New York, 1997.

16. D. Shotton, ed, 'Electronic Light Microscopy', *The Principles and Practice of Video-enhanced Contrast, Digital Intensified Fluorescence and Confocal Scanning Light Microscopy*, Wiley-Liss, New York, 1993.

17. X. Wang, B. Herman, *Fluorescence Imaging Spectroscopy and Microscopy*, John Wiley & Sons, New York, 1994.

18. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, eds, *Handbook of Microscopy*, VCH, Weinheim, 1997.

19. T. Rochow, P. Tucker, *Introduction to Microscopy by Means of Light, Electrons, X-rays and Acoustics*, 2nd edition, Plenum, New York, 1994.

20. D. Newbury, D. Joy, P. Echlein, C. Fiori, *Advanced Scanning Electron Microscopy and X-ray Microanalysis*, Plenum, New York, 1986.

21. K. Heinrich, D. Newbury, eds, *Electron Probe Quantitation*, Plenum, New York, 1991.

22. A. Benninghoven, ed, *Secondary Ion Mass Spectrometry*, Springer, Berlin, 1984.

23. A. Benninghoven, F. Rudenauer, H. Werner, *Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Applications and Trends*, John Wiley & Sons, Chichester, 1987.

24. D. Briggs, A. Brown, J. Vickerman, *Handbook of Static Secondary Mass Spectrometry*, John Wiley & Sons, Chichester, 1989.

25. J. Campbell, K. Malmqvist, S. Johansson, *Particle-induced X-ray Emission Spectrometry*, John Wiley & Sons, Chichester, 1995.

26. S. Reed, *Electron Microprobe Analysis*, Cambridge University Press, Cambridge, 1997.

27. L. Murr, *Electron and Ion Microscopy and Microanalysis*, 2nd edition, Marcel Dekker, 1991.

28. G. Wade, *Acoustic Imaging: Cameras, Microscopes, Phased Arrays and Holographic Systems*, Plenum, New York, 1976.

29. D. Sarid, *Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces*, Oxford University Press, 1994.

30. D. Dixon, L. Dugdale, *Introduction to Clinical Imaging*, Churchill Livingstone, Edinburgh, 1987.

31. R. Robb, *Three-dimensional Biomedical Imaging: Principles and Practice*, VCH, Weinheim, 1995.

32. P. Ell, B. Holman, eds, *Computed Emission Tomography*, Oxford University Press, Oxford, 1983.

33. R. English, SPECT, *Single-photon Emission Computed Tomography: A Primer*, Society of Nuclear Medicine, Reston, 1996.

34. G. Hermann, *Image Reconstruction from Projections*, Academic Press, New York, 1980.

35. A. Kriete, ed, *Visualization in Biomedical Microscopies. 3D Imaging and Computer Applications*, VCH, Weinheim, 1992.

36. T. Corle, G. Kino, eds, *Confocal Scanning Optical Microscopy and Related Imaging Systems*, Academic Press, London, 1996.

37. J. Stevens, L. Mills, J. Trogadis, *Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens*, Academic Press, London, 1994.

38. G. Lohmann, *Volumetric Image Analysis*, John Wiley & Sons, Chichester, 1998.

39. J. Van der Weerd, S. Kazarian, 'Validation of Macroscopic ATR-FTIR Imaging to Study Dissolution of Swelling Pharmaceutical Tablets', *Applied Spectroscopy*, **58**, 1413–1419 (2004).

40. J. Van der Weerd, K. Chan, S. Kazarian, 'An innovative Design of a Compaction Cell for *in-situ* FT-IR Imaging of Tablet Dissolution', *Vibrational Spectroscopy*, **35**, 9–13 (2004).

41. T. Lillesand, R. Kiefer, *Remote Sensing and Image Interpretation*, John Wiley & Sons, New York, 1979.

42. F. Sabins, *Remote Sensing Principles and Interpretation*, Freeman & Co, San Francisco, 1978.

43. M. Horst, *Digital Image Processing of Remotely Sensed Data*, Academic Press, New York, 1987.

44. H. Kramer, *Observation of the Earth and its Environment. Survey of Missions and Sensors*, 2nd edition, Springer, Berlin, 1994.

45. G. Vane, ed, *Proceedings of the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Performance Evaluation Workshop*, NASA Jet Propulsion Laboratory, Pasadena, 1988.

46. M. Moran, A. Vidal, D. Troufleau, J. Qi, T. Clarke, P. Pinter, T. Mitchell, Y. Inoue, C. Neale, 'Combining Multifrequency Microwave and Optical Data for Crop Management', *Remote Sensing of Environment*, **61**, 96–109 (1997).

47. J.-P. Deroin, Y. Marchand, J.-P. Auffret, 'Littoral Survey Using the JERS-OPS Multispectral Sensor: Example of the Mont-Saint-Michel Bay (Normandy, France)', *Remote Sensing of Environment*, **61**, 119–131 (1997).

48. D. Roberts, R. Green, J. Adams, 'Temporal and Spatial Patterns in Vegetation and Atmospheric Properties From

AVIRIS', *Remote Sensing of Environment*, **61**, 223–240 (1997).

49. F. van der Meer, W. Bakker, 'Cross Correlogram Spectral Matching: Application to Surface Mineralogical Mapping Using AVIRIS Data From Cuprite, Nevada', *Remote Sensing of Environment*, **61**, 371–382 (1997).
50. A. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall International Editions, Englewood Cliffs, 1989.
51. Y. Yateno, T. Linuma, M. Takano, eds, *Computer Radiography*, Springer-Verlag, Berlin, 1987.
52. Z. Cho, H. Kim, H. Song, J. Cumming, 'Fourier Transform Nuclear Magnetic Resonance Tomographic Imaging', *Proceedings of the IEEE*, **70**, 10 (1982).
53. G. Turrell, J. Corset, eds, *Raman Microscopy: Developments and Applications*, Academic Press, New York, 1996.
54. P. Geladi, H. Grahn, *Multivariate Image Analysis*, John Wiley & Sons, Chichester, 1996.
55. H. Grahn, P. Geladi, eds, *Techniques and Applications of Hyperspectral Image Analysis*, John Wiley & Sons, Chichester, 2007.
56. P. Van Espen, G. Janssens, W. Vanhoolst, P. Geladi, 'Imaging and Image Processing in Analytical Chemistry', *Analysis*, **20**, 81–90 (1992).
57. P. Van Espen, G. Janssens, 'Imaging Secondary Ion Mass Spectrometry', in *Handbook of Microscopy Methods, II*, eds. S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo, VCH, Weinheim, 691–718, 1997.
58. J. Joliffe, *Principal Component Analysis*, Springer, Berlin, 1986.
59. J. Jackson, *A User's Guide to Principal Components*, John Wiley & Sons, New York, 1991.
60. E. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, John Wiley & Sons, New York, 1991.
61. W. Windig, S. Markel, 'Simple-to-use Interactive Self-Modeling Mixture Analysis of FTIR Microscopy Data', *Journal of Molecular Structure*, **292**, 161–170 (1993).
62. M. Keenan, P. Kotula, 'Accounting for Poisson Noise in the Multivariate Analysis of ToF-SIMS Spectrum Images', *Applied Surface science*, **231–232**, 240–244 (2004).
63. J. Noordam, W. van den Broek, 'Multivariate Image Segmentation Based on Geometrically Guided Fuzzy c-Means Clustering', *Journal of Chemometrics*, **16**, 1–11 (2002).
64. P. Geladi, F. Lindgren, J. Swerts, 'Multiwavelength Microscopic Image Analysis of a Piece of Painted Chinaware: Classification and Regression', *Chemometrics and Intelligent Laboratory Systems*, **24**, 145–167 (1994).
65. P. Geladi, J. Burger, T. Lestander, 'Hyperspectral Imaging: Calibration Problems and Solutions', *Chemometrics and Intelligent Laboratory Systems*, **72**, 209–217 (2004).
66. J. Burger, P. Geladi, 'Hyperspectral NIR Image Regression Part I: Calibration and Correction', *Journal of Chemometrics*, **19**, 355–363 (2005).
67. P. Ready, P. Wintz, 'Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery', *IEEE Transactions on Communications*, **COM21**, 1123–1131 (1973).
68. H. Grahn, N. Szeverenyi, M. Roggenbuck, P. Geladi, 'Tissue Discrimination in Magnetic Resonance Imaging: a Predictive Multivariate Approach', *Chemometrics and Intelligent Laboratory Systems*, **14**, 357–374 (1992).
69. P. Geladi, K. Esbensen, 'Regression on Multivariate Images: Principal Component Regression for Modeling, Prediction and Visual Diagnostic Tools', *Journal of Chemometrics*, **5**, 97–111 (1991).
70. K. Esbensen, P. Geladi, H. Grahn, 'Strategies for Multivariate Image Regression (MIR)', *Chemometrics and Intelligent Laboratory Systems*, **7**, 67–86 (1989).
71. F. Lindgren, P. Geladi, S. Wold, 'Kernel-based PLS Regression: Kernel-based Cross Validation and Applications to Spectral Data', *Journal of Chemometrics*, **8**, 377–389 (1994).
72. H. Grahn, J. Sääf, 'Multivariate Image Regression and Analysis. Useful Techniques for the Evaluation of Clinical Magnetic Resonance Images', *Chemometrics and Intelligent Laboratory Systems*, **14**, 391–396 (1992).
73. W. van den Broek, E. Derks, E. van de Ven, D. Wienke, P. Geladi, L. Buydens, 'Plastic Identification by Remote Sensing Spectroscopic NIR Imaging Using Kernel Partial Least Squares (KPLS)', *Chemometrics and Intelligent Laboratory Systems*, **23**, 187–197 (1996).
74. N. Gallagher, B. Wise, D. Sheen, 'Estimation of Trace Vapor Concentration-Pathlength in Plumes for Remote Sensing Applications from Hyperspectral Images', *Analytica Chimica Acta*, **77**, 105–117 (2003).
75. J. Burger, P. Geladi, 'Hyperspectral NIR Imaging for Calibration and Prediction: a Comparison Between Image and Spectrometer Data for Studying Organic and Biological Samples', *The Analyst*, **10**, 1152–1160 (2006).
76. H. Grahn, P. Geladi, 'Application of Multivariate Data Analysis Techniques to NMR Imaging', in *Signal Treatment and Signal Analysis in NMR*, ed, N. Rutledge, Elsevier, Amsterdam, 1996.
77. W. van den Broek, 'Chemometrics in Spectroscopic Near Infrared Imaging for Plastic Material Recognition', PhD Thesis, Catholic University of Nijmegen, 1997.
78. D. Bright, S. Newbury, 'Concentration Histogram Imaging. A Scatter Diagram Technique for Viewing Two or Three Related Images', *Analytical Chemistry*, **63**, 243A–249A (1992).
79. O. Hagman, 'Multivariate Prediction of Wood Surface Features Using an Imaging Spectrograph', *Holz als Roh- und Werkstoff*, **55**, 377–382 (1997).

80. F. Pedersen, E. Bengtsson, B. Nordin, 'An Extended Strategy for Exploratory Multivariate Analysis Including Noise Considerations', *Journal of Chemometrics*, **9**, 389–409 (1995).

81. N. Bonnet, 'Preliminary Investigation of Two Methods for the Automatic Handling of Multivariate Maps in Microanalysis', *Ultramicroscopy*, **57**, 17–27 (1995).

82. P. Trebbia, N. Bonnet, 'EELS Elemental Mapping With Unconventional Methods. I. Theoretical Basis: Image Analysis With Multivariate Statistics and Entropy Concepts', *Ultramicroscopy*, **34**, 165–178 (1990).

83. Y. Chtioui, D. Bertrand, Y. Dattée, M.-F. Devaux, 'Identification of Seeds by Colour Imaging. Comparison of Discriminant Analysis and Artificial Neural Network', *Journal of the Science of Food and Agriculture*, **71**, 433–441 (1996).

84. P. Geladi, K. Esbensen, 'Multivariate Image Analysis in Chemistry: An Overview', in *Applied Multivariate Analysis in SAR and Environmental Studies*, eds. J. Devillers, W. Karcher, Kluwer Academic, Dordrecht, 415–445, 1991.

85. P. Geladi, H. Grahn, F. Lindgren, 'Chemical Multivariate Image Analysis: Some Case Studies', in *Applied Multivariate Analysis in SAR and Environmental Studies*, eds. J. Devillers, W. Karcher, Kluwer Academic, Dordrecht, 447–478, 1991.

86. K. Esbensen, P. Geladi, F. Krogh, K. Kvaal, 'Multivariable Bildeanalyse-Kjemisk Bildeanalyse', in *Anvendelse av Kjemometri Innen Forskning og Industri*, eds. R. Nortvedt, F. Brakstad, O. Kvalheim, T. Lundstedt, Tidsskriftforlaget Kjemi AS, Oslo, 305–321, 1996.

87. J. Richards, *Remote Sensing Digital Image Analysis. An Introduction*, 2nd edition, Springer, Berlin, 1993.

88. R. Schowengerdt, *Remote Sensing Models and Methods for Image Processing*, Academic Press, London, 1997.

89. P. Geladi, 'Sampling and Local Models for Multivariate Image Analysis', *Microchimica Acta*, **120**, 211–230 (1995).

Quality Assurance in Analytical Chemistry

Christoph Fischbacher
Analytik Jena AG, Jena, Germany

1 Introduction **1**
2 General Aspects and Goals **2**
2.1 Traceability 3
2.2 Measurement Uncertainty 3
2.3 Statistical Process Control and Control Samples 6
2.4 Regulatory Aspects and Standards 10
3 Quality Assurance in the Analytical Process **11**
3.1 Sources and Significance of Errors 11
3.2 Organization, Personnel and Premises 12
3.3 Sampling and Sample Preparation 13
3.4 Reagents and Standards 14
3.5 Instruments and Computers 15
3.6 Reporting Results 16
3.7 Documentation and Software 17
4 Method Development and Validation **18**
4.1 Calibration and Recovery Studies 19
4.2 Performance Characteristics 19
4.3 Reference Materials 21
5 External Quality Assurance **22**
5.1 Proficiency Testing 22
5.2 Accreditation 22
Acknowledgments **23**
Abbreviations and Acronyms **23**
Related Articles **23**
References **23**

Analytical methods have to meet certain quality requirements, so that the results are credible and can be used with confidence for regulatory, economic and research decisions. Analytical quality assurance (QA) contains preventive action and comprises planning steps, quality control (QC), quality assessment, documentation and quality improvement. All factors leading to the analytical result have to be controlled, clearly described and documented. Furthermore, the analytical result should be linked to the sample and related to reference materials (RMs) or measurement standards by an unbroken chain of traceability. Two elements are needed in order to achieve this goal: internal QA measures comprise elements performed by the laboratory, such as method validation and documentation and the use of RMs, external QA relates to performance evaluation by persons outside the laboratory, such as proficiency tests and accreditation.

1 INTRODUCTION

Results from analytical measurements form the basis for decisions which may have serious economic, environmental, safety and health related consequences. Confidence in these decisions therefore depends directly on the quality of the analytical results. The International Organization for Standardization (ISO) defines quality as 'the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs'.[1] Some of these demands are stated by the customer, who may be located inside a company or a public organization, others may be required by standards or laws, e.g. safety requirements. Because needs change with time, what was regarded as being of high quality in analytical chemistry some years ago may not satisfy today's requirements. On the other hand, qualitative or half-quantitative results, for example a pH value acquired by using litmus paper, may be fully adequate for a certain task. It follows that quality is a relative concept.

In order to ensure that products or services meet defined standards of quality with a stated level of confidence, appropriate QA measures have to be taken. While traditional product-centered QC is based on the principle of checking conformance with the specifications of the final product, process-oriented QA contains preventive action and comprises planning steps, QC, quality assessment, documentation and quality improvement. QA can thus be regarded as a feedback loop (Shewhart cycle)[2] and consists of the following steps: (1) development of a plan to improve a process; (2) testing of the plan; (3) checking of the test results; and (4) implementation or rejection of the plan (Figure 1).

Analytical chemistry is related to QA in two ways: first, analytical methods are used to test and characterize products such as pharmaceuticals, thus giving the customer the assurance that the product conforms with the stated requirements and that conformance is reliable and consistent. The second aspect is that analytical methods themselves have to meet certain quality requirements, so that the results are credible and can be used with confidence for regulatory, economic and research decisions. Since all steps of chemical analysis, from sampling and method selection to choice of equipment and the taking and reporting of measurements, affect the quality of analytical data, confidence in the results can only be achieved

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

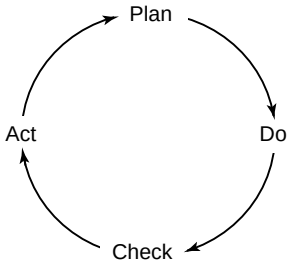

Figure 1 Shewhart cycle (plan-do-check-act cycle).

if the laboratory has the competence and capability to perform the analyses (Figure 2).

This means that all factors leading to the analytical result have to be controlled, clearly described and documented. Furthermore, the analytical result should be linked to the sample and related to RMs or measurement standards by an unbroken chain of traceability. This chain has to be transparent and well documented. Another aim of QA in analytical laboratories is to increase work safety by the use of standardized procedures, regular instrument maintenance, and qualified personnel. Non-technical aspects of QA cover integrity and independence of the laboratory, which assures that the results are not biased by political, economic or other reasons.

The long-term goal of analytical QA is that the methods and processes that lead to analytical results are globally consistent and accepted. This can be attained by (written) standards and technical specifications. These were traditionally developed for products but can also be used to describe processes, services and systems.

The quality concept has far-reaching implications for trade, e.g. the amount of retesting of products is reduced. With the development of the regional trading blocks such as the European Union (EU) and the North American Free Trade Agreement (NAFTA), efforts have been made to accept standards by mutual agreements and to harmonize standards between the different regions and countries.[3] Certain terms, however, especially those regarding analytical method validation and accreditation, are still used inconsistently in a number of national and international documents. Since most terms relating to analytical QA are defined in English, the possibility exists that ambiguities may be introduced by translation into other languages, a problem which has been dealt with by producing a glossary of analytical terms.[4]

Two elements of QA are necessary in order to assure consistent and reliable analytical results: Internal QA measures comprise elements performed by the laboratory, such as method validation and documentation and the use of RMs; external QA relates to performance evaluation by persons outside the laboratory. Examples of external QA are accreditation by independent bodies or proficiency testing aimed at producing comparable data generated in different laboratories. Accreditation is a system of voluntary peer assessment of laboratory competence conducted by a third party according to national or international standards or guidelines.

2 GENERAL ASPECTS AND GOALS

In this section some fundamental terms and concepts of analytical QA are discussed. Since most of the terms, standards and guidelines were developed for laboratories in general, ranging from materials testing to electrical laboratories, an interpretation concerning analytical tasks is necessary. While some aspects of analytical QA, especially regulatory matters, vary in different countries, there

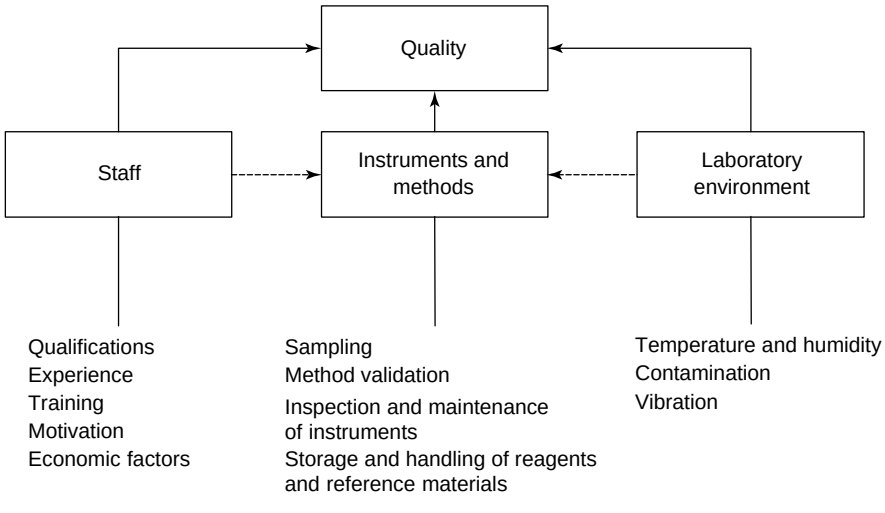

Figure 2 Some elements affecting quality in the analytical laboratory.

is a general agreement about the principles of analytical QA. A trend towards international harmonization of the relevant standards and guidelines can be observed. Several books deal with general aspects of analytical QA.[5–8]

2.1 Traceability

Traceability is a key concept in QA and is defined in the International Vocabulary of Basic and General Terms in Metrology (VIM)[9] as the ''property of the result of a measurement or the value of a standard whereby it can be related, with a stated uncertainty, to stated references, usually national or international standards, through an unbroken chain of comparisons''. Measurement traceability is hierarchical and means that an unbroken and demonstrable path of calibration exists which preferably begins with a national or international standard, for example a standard reference material (SRM) of the US National Institute of Standards and Technology (NIST) or a physical fundamental constant, and ends with the sample. The purpose of traceability is that measurements in different laboratories are accurate representations of the measured quantity within the uncertainty of the measurement, which is a prerequisite for the consistency and comparability of the measurements.

The term ''traceability'' is sometimes used in a slightly different sense and refers to the chain of documentation from the sample to the result, which is needed in order to repeat the analysis. For this audit trail, the term ''trackability''[10] has been proposed.

An example from physical metrology is the calibration chain which links a yardstick to the definition of the meter in terms of the path traveled by light during a time interval of 1/299 792 458 s. By analogy, quantitative chemical measurements should in principle be traceable to the mole, which is the basic quantity in the international measurement system of units [Système International (SI)][11] for the amount of a defined substance (Table 1).

In order to prove the accuracy of the measurements, laboratories obtain certified reference materials (CRMs). In an ideal case the certified value is directly traceable to the fundamental units of measurement (mass, length, time, etc.) through the use of a primary method such as coulometry, titrimetry, gravimetry or isotope dilution mass spectrometry (IDMS).[12,13] These methods are fully understood, give results with a small degree of uncertainty and do not rely on empirical determination of the measurement function by calibration. In practice, however, applicability of primary methods is very limited owing to the small number of these available.

Table 1 The SI base units

Quantity	Name	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

Buffer solutions made from or traceable to primary standards can be used for the determination of the pH value. A national metrology institute uses a primary standard measurement device to certify primary standards with a defined uncertainty. Secondary RMs can be derived from these with assigned pH values and an uncertainty which includes the uncertainty of the primary standard. The user of a pH meter who buys the certified calibration buffer solutions can therefore trace his or her pH value measurement to a national standard.[14]

In many cases, a rigorous application of the traceability concept in chemical analysis is not viable, since the measurement step is only one part of the analytical process. The uncertainties of sampling or preconcentration steps are mostly uncalibrated and therefore not directly traceable. Furthermore, for some analytical methods, e.g. sum parameters such as adsorbable organically bound halogens (AOX) or the protein content of a meat sample, accuracy can only be stated relative to a reference method.

The strictness of the implementation of the traceability concept depends on the field and on the required level of accuracy. For most analytical laboratories, a practicable interpretation of traceability is that the accuracy and overall measurement uncertainty of the result have been assessed in a transparent and demonstrable way.[15,16] The aim should be traceability to either matrix-matched RMs or to a generally accepted reference method. If both are lacking, then samples spiked with the analyte provide a restricted possibility to assess accuracy and uncertainty.

2.2 Measurement Uncertainty

In chemical analysis, significant sources of uncertainty may arise from poor sampling, incomplete extraction of the analyte, contamination, uncertainty of RMs and matrix effects during the measurement step. An analytical result is therefore never more than an approximate estimation of the true value of the measured quantity, and incomplete without a realistic quantitative statement of its uncertainty, which ''characterizes the dispersion of the values that could reasonably be attributed to the measurand'' VIM. The uncertainty is needed to determine whether a result can be used for its intended purpose and in order to compare a result with others, regardless of the analytical method used. Uncertainty

must be distinguished from error, which is defined as the difference between an individual result and the true value of the measurand.

An analytical result, e.g. concentration of an analyte, is not measured directly but is estimated using a function $y = f(x_1, x_2, x, \ldots, x_n)$ of intermediate measurement results x_i. The input estimates x_i are, for example, instrument responses, calibration parameters or correction factors, all associated with uncertainties which have an effect on the uncertainty of y.

Often the uncertainty of the result is evaluated by taking into account solely the standard deviation of the calibration function or from repeated determinations. In this way, the uncertainty is underestimated because only uncertainty components arising from random effects (repeatability) are considered. The overall analytical uncertainty is often much larger and contains uncertainty components arising from systematic effects, as can be demonstrated by the use of CRMs or inter-laboratory comparisons where between-laboratory reproducibility is estimated.

Various approaches in order to realistically quantify the overall measurement uncertainty can be distinguished:

- the step-by-step approach, which calculates the overall uncertainty of analytical result by taking into account the individual contributions of uncertainty (error propagation);
- the experimental approach, which systematically varies all parameters which show a significant influence on the uncertainty of the result;
- the top-down method, which estimates uncertainties directly using suitable reference samples, established reference methods or interlaboratory comparisons in order to determine reproducibility and laboratory bias;[17]
- other methods based on numerical evaluation of uncertainty components using the bootstrap method, cross-validation or other resampling statistical methods,[18] or sensitivity analysis using simulations.

In order to achieve consistent and realistic expressions of measurement uncertainty, the ISO *Guide to the Expression of Uncertainty in Measurement*[19] has been prepared by several organizations. The ISO Guide describes the step-by-step approach to measurement uncertainty by assessing the individual components of uncertainty in the various analytical steps and combining them to an uncertainty estimation using an error propagation model of the measurement process. A guidance document for the interpretation of the ISO guide in relation to analytical application which also contains worked examples has been prepared by a EURACHEM working group.[20]

The following steps must be performed in order to assess the overall uncertainty for analytical measurements:

1. identify and list the steps of the analytical procedure, e.g. sampling, sample preparation and calibration;
2. for each step, identify and list the sources of uncertainty which influence the performance;
3. decide which uncertainty contributions have to be further investigated and which can be neglected;
4. express the determination of the measurand as a functional relationship of input parameters;
5. for each significant contribution of uncertainty, evaluate the contribution to the measurand;
6. if relevant correlations between the individual uncertainty contributions exist, their corresponding covariance has to be taken into account;
7. calculate the combined uncertainties of the target quantity as the square root of the sum of squares of the uncertainties of the individual steps and their relevant covariances;
8. an expanded combined uncertainty can be calculated (for security) by multiplying the combined uncertainty by a coverage factor, mostly 2.

According to this enumeration, the main task is to estimate the individual uncertainty components. Two methods of evaluation of the individual uncertainty components can be distinguished according to the method used for their estimation: type A uncertainty components are evaluated by statistical analysis of a series of observations, while type B uncertainty components are evaluated by other means. Each uncertainty component, regardless of the method of evaluation, can be represented by an estimated standard deviation termed standard uncertainty u_i.

Uncertainty components evaluated by type A can be directly represented by the statistically estimated standard deviation s_i with the associated degrees of freedom: $u_i = s_i$. Examples are the standard deviation of a mean of observations or the standard deviation of calibration parameters fitted by the least-squares method.

In type B evaluation, the uncertainty component is represented by a quantity u_i which uses a hypothetical distribution based on assumptions and available information, such as previously determined data from method validation, data from literature or from specifications by the manufacturer.

The uncertainty evaluated according to type B is treated like a standard deviation and is equivalent to u_i. For example, if a maximum upper a_u and lower value a_l for the quantity in question is given, then it can be

approximately assumed that all values (exactly 99.73% for normal distribution) lie in the interval $a \pm 3$ standard deviations with $a = (a_u - a_l)/2$ if no other information is given. An approximation of the standard uncertainty would therefore be $u_i = a/3$. For simplicity, in the absence of other information, a uniform rectangular or triangular distributions may be assumed in order to compute the standard uncertainty, which leads to a standard uncertainty of $u_i = a/\sqrt{3}$ and $u_i = a/\sqrt{6}$ relating to 58 and 65%, respectively, of the area in the interval $a \pm u$.

The individual standard uncertainties u_i evaluated either by type A or B are then combined, $u_c(y)$, according to the law of propagation of uncertainty, resulting in the combined standard uncertainty of the result. The expression for uncorrelated uncertainty components is given by Equation (1):

$$u_c^2(y) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) \qquad (1)$$

The uncertainties of additive or subtractive corrections or correction factors have to be included in the assessment according to Equation (1). The combined standard uncertainty resulting from uncorrelated individual uncertainty components always exceeds the uncertainty of the individual contributions.

If some variables are not independent of each other, i.e. significant correlations between the variables exist, then the result may be underestimated for a positive correlation, since the contribution terms reinforce each other, or overestimated for a negative correlation, since the contribution terms partially cancel each other out. Therefore, for correlated uncertainty contributions an additional covariance term has to be included.[19,20]

The probability distribution of the measurand and its combined standard uncertainty is often approximately normal (due to the Central Limit Theorem, which states that even if the original population is not normally distributed, the sampling distribution of means will increasingly approximate a normal distribution as sample size increases). In this the level of confidence is approximately 68% that the measurand y lies in the interval $y - u_c(y) \leq y \leq y + u_c(y)$ given a negligible uncertainty of the determination of $u_c(y)$ itself.

For additional safety, e.g. in industrial applications, the combined uncertainty may be multiplied by a coverage factor k to give the expanded uncertainty $U = ku_c(y)$, where k is usually in the range 2–3, relating to a level of confidence from approximately 95% to approximately 99% for a normal distributed measurand. Generally $k = 2$ is used; the exact value for k has to be stated.

Figure 3 shows typical steps of a spectrometric, e.g. inductively coupled plasma optical emission spectrometry (ICPOES) or atomic absorption spectrometry (AAS), determination of a solid sample after sample digestion together with the uncertainty contributions.

If suitable RMs are available, the top-down approach which directly estimates the combined uncertainty is in

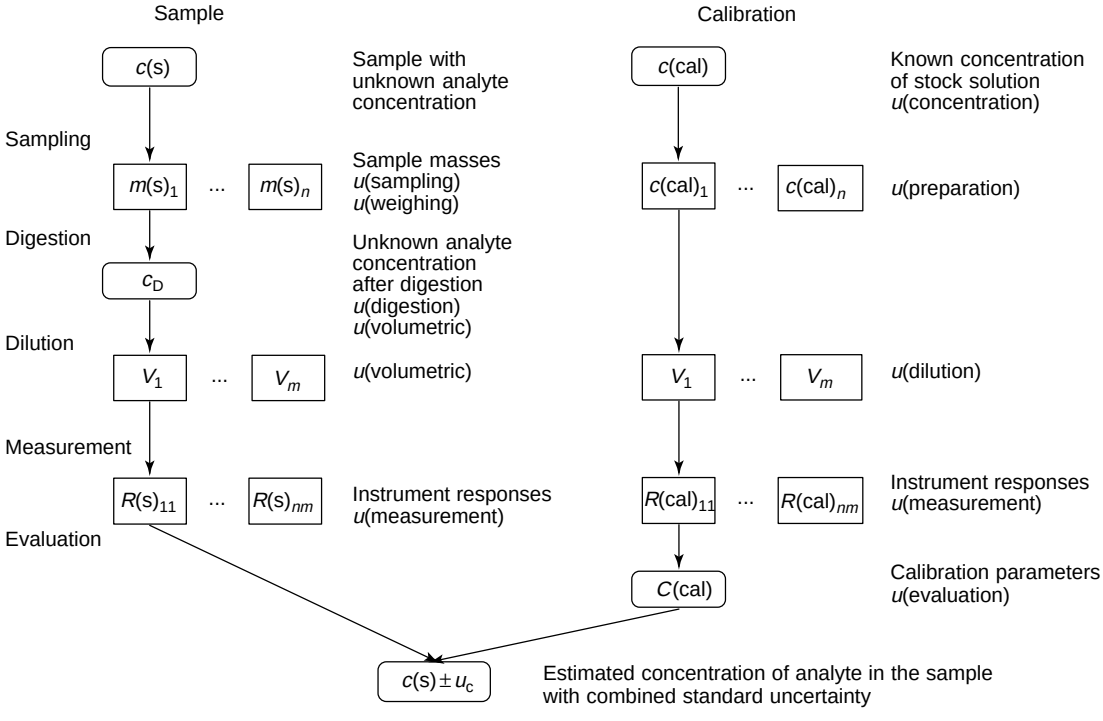

Figure 3 Typical steps and sources of uncertainty of a spectrometric determination of a solid sample after digestion.

many cases more suitable to assess the uncertainty of the result, since the combined effect of several sources of uncertainty is taken into account in one step. The remaining uncertainty components are the uncertainty of the assigned value of the RM, the reproducibility of the measurements made on the RM, matrix effects due to difference in the composition between sample and RM and steps that are not accounted for, e.g. sample inhomogeneity and sample storage.

In chemical analysis, expert judgement is often required to assess uncertainty components. For example, subjective assumptions on the type of matrix of a sample are often necessary. The EURACHEM guidance document states that ''The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement method and procedure.''[20]

2.3 Statistical Process Control and Control Samples

Statistical process control (SPC) is a means of detecting and controlling unaccounted sources of uncertainty in the analytical process. A process is a structured order of activities and events necessary to reach a certain goal, e.g. a production process or analytical process. A stable process is predictable and shows only small random variations from the average, which can be attributed to factors such as unavoidable random variations of temperature or contamination. The stability of the process is disturbed by assignable causes (also called special causes), which are unnatural variations of the process, resulting, for example, from problems with instruments and reagents or human mistakes. The aim is to achieve and maintain a stable process, since this leads to more reliable results and to less work with adjusting the process.

SPC monitors the process in order to differentiate systematic deviations of the process from the natural random variations and assists in the decision if it is necessary to intervene in order to reduce variability and to maintain the stability of a process. Statistical assumptions are used to help determine whether a variation is solely by chance or due to an assignable cause. Reference works dealing with SPC are those by Grant and Leavenworth[21] and Montgomery.[22]

A process can be visually represented by QC charts, which were at first introduced by Shewhart[23] for the SPC of production processes. A number of variables which characterize product quality are monitored by measuring samples of a certain size extracted from the production process. A graph of the individual variable or some sample statistics, e.g. mean or standard deviation, is then plotted over time or sampling sequence to reflect the course of the process. A control chart usually consists of a central line together with warning and control limits which are based on statistical parameters estimated from earlier samples. Depending on the type of control chart, significant deviations from the target specifications or excessive variability around target specifications can be detected. Control charts can be used to document the process history. Two process characteristics which are of special interest are the location of the mean of the process and the variation of the process.

In contrast to industrial product QC for analytical chemistry, special control samples have to be monitored in order to use control charts. Materials or solutions used for QC must be available in sufficient amounts, stable under defined storage and handling conditions for several months, and homogeneous. The following types of control samples are frequently used:

- standard samples which are solutions of a substance or materials with known purity and concentration without the sample matrix;
- blank samples which are free from analyte and are processed in the same way as the sample in order to evaluate and control contamination from reagents and environment;
- real samples which have the same matrix as the sample but contain an unknown amount of analyte;
- spiked real samples spiked with a known amount of the analyte (if possible of the same chemical species) and forming the basis of recovery experiments;
- matrix-matched synthetic samples which are calibration samples with additional substances contained in the sample matrix;
- CRMs which are real or synthetic samples for which a conventional true value with known uncertainty has been certified by a defined and reliable procedure, e.g. a round-robin test.

Table 2 shows which performance characteristics can be controlled with the different types of control samples. Additionally, instrumental or calibration parameters can be monitored by SPC, such as the temperature of a refrigerator or the slope of the calibration line.

The control samples have to be analyzed at regular intervals, which depend on the reliability of the method and the level of confidence required.

In most cases, analytical applications control charts are used which require quantitative variables. Generally a normal distribution is assumed. Control charts for qualitative attributes also exist but are rarely used. Qualitative attributes follow discrete distributions, e.g. binomial or Poisson distributions. The simplest chart is the individual chart (x-chart) where the individual observations are plotted in the sequence of observation. The course of the plot displays both variation and location

Table 2 Control of performance characteristics using control samples and charts

Type of control sample	Type of control chart	Control of	
		Precision	Accuracy
Calibration standard sample	$\bar{x}$-chart	+	–
Blank sample	Blank chart ($\bar{x}$-chart)	+	+ (control of reagents and contamination)
Real sample	s- or R-chart	+	–
Spiked real sample	Recovery chart ($\bar{x}$-chart)	–	+
Matrix-matched synthetic sample	$\bar{x}$-, R- and s-chart	+	+
CRM	$\bar{x}$-, R- and s- chart	+	+

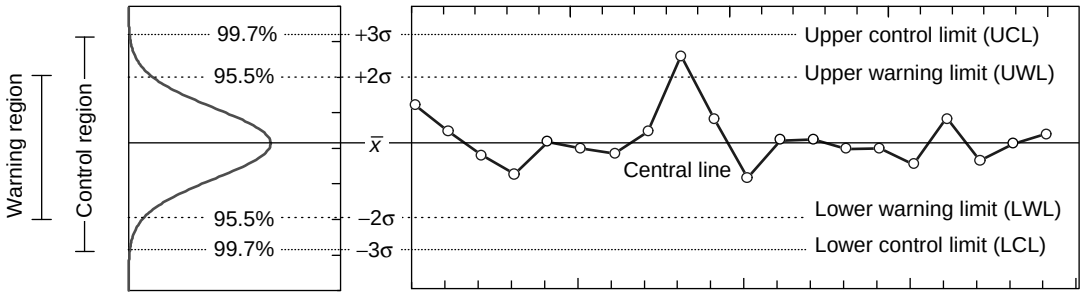

Figure 4 An $\bar{x}$-chart with control and warning limits.

of the process. Figure 4 shows an $\bar{x}$-chart (x-bar chart) which is used to track the location of mean values $\bar{x}_i$ of samples of size n. The central line is given by the overall mean $\bar{\bar{x}}$, usually calculated from stable periods of the process history. The lower control limit (LCL) and upper control limit (UCL) can be expressed by Equations (2) and (3):

$$\mathrm{UCL} = \mu + 3\sigma_x \tag{2}$$

$$\mathrm{LCL} = \mu - 3\sigma_x \tag{3}$$

where μ is estimated by the mean value $\bar{\bar{x}}$ and σ_x is the standard deviation of the mean, $\sigma_x = \sigma/\sqrt{n}$, which can be estimated according to Table 3 using the mean standard deviation $\bar{s}$. Given a stable process with only random variations and normal distributed data, approximately 99.7% of the sample means are expected to fall inside the $3\sigma_x$ interval. The probability of error is thus $p \approx 0.3\%$. With a decrease in the factor the sensitivity and the risk of giving false alarm (type I error) increases. For increased sensitivity of the $\bar{x}$ -chart warning lines are often added using $2\sigma_x$ ($p \approx 0.05$). Other multiples of σ_x can be calculated to obtain limits with specified p-values.[24]

As stated previously, applications of the $\bar{x}$-chart in analytical QC require samples with known concentrations, e.g. reference samples, in order to control accuracy. The blank chart often used in trace analysis is an $\bar{x}$-chart where the mean values of the blank determinations are controlled to detect changes and trends of blanks. Another application is the recovery chart used to control proportional changes in sensitivity due to matrix components. The basis is the percentage recovery rate [Equation 4]:

Table 3 Parameters for the construction of $\bar{x}$-, R- and s-charts

Type of chart	Quantity tracked	Estimator for μ	Estimator for σ
x-Chart	Individual observations x_i	$\bar{x}$	$\bar{R}/d_2$ with $n = 2$ and $R = \lvert x_i - x_{i+1}\rvert$
$\bar{x}$-Chart	Mean values $\bar{x}_i$ of a sample with size n	$\bar{\bar{x}}$	$\dfrac{\bar{s}}{c_4\sqrt{n}}$
R-Chart	Range R_i of a sample with size n	$\bar{R}$	$d_3\dfrac{R}{d_2}$
s-Chart	Standard deviation s_i of a sample with size n	$\bar{s}$	$\sqrt{1-c_4^2}\dfrac{\bar{s}}{c_4}$

$$RR = \frac{x_{\mathrm{spiked,meas}} - x_0}{x_{\mathrm{spiked}}} \times 100\% \tag{4}$$

where $x_{\mathrm{spiked,meas}}$ is the concentration of analyte measured in the spiked sample, x_0 is the concentration of analyte measured in the sample prior to the spiking procedure

and x_{spiked} is the concentration of analyte added (added value, not determined by method).

Homogeneity of the variances over time is a prerequisite for a reliable use of the $\bar{x}$-chart. Changes in precision can be controlled by tracking the range (R-chart) or the standard deviation (s-chart). Since the number of replicate determinations in the analytical laboratory is typically $n = 2–3$, the R-chart is often used to assess the spread of the measurement values. The range of a sample of measurements is defined according to Equation (5):

$$R = x_{max} - x_{min} \quad (5)$$

If the number of samples for each point is higher ($n > 5$), the standard deviation can be calculated to assess the spread of data. Range and standard deviation are always positive, a fact which must be considered for the construction of lower limits. These correspond to higher precision and can be used to find causes which lead to the improvement of precision or to detect fraud, e.g. by using identical values instead of replicates.

A combination of the $\bar{x}$-chart with the s- or r-chart using the same data is especially useful in order to monitor precision and trueness simultaneously. A decision can be made if the variability between the subsets is smaller or greater than that in the subset. The cause of out-of-control situations can be better detected. Out-of-control results differ in a statistically significant way from measurements which were taken in order to establish the control limits, but this does not necessarily mean that the results are useless for the customer. Therefore, in addition to control and warning limits, specification limits can be established, e.g. the measurement uncertainty of a RM.

The construction of the $\bar{x}$-, R- and s-charts is shown in Table 3. The values c_4, d_2 and d_3 are functions of the sample size n tabulated in Table 4 and characterize the distribution of the standard deviation and the range.[25]

When the laboratory works at a level of high quality with previously validated methods, the process is stable. This means that the process meets the performance criteria determined during the validation, shows only random deviations and independence between successive points (no trends) and only rarely exceeds the specified limits. On the other hand, a process which displays trends or exceeds the established limits because of an assignable cause, is defined as being out of control. Appropriate measures can then be taken in order to find the cause of the variation and bring the process back to stability. This can mean instrument maintenance, new reagents or even revalidation of the procedure.

A number of tests based on statistical assumptions exist to decide objectively if a process is out of control. For example, a systematic deviation of the mean value can be assumed if a certain number of points consecutively fall below or above the central line. The probability that this is solely by chance is $P = 0.5$ for each point, so that $P = 0.0312$ (3.12%) for five consecutive points lying on one side of the central line. Similar assumptions exist for identifying trends from a number of consecutively increasing or decreasing values. Various rules for chart interpretation exist. Some examples which indicate problems with the process are one value outside the 3σ control limits, two of three consecutive values outside of the 2σ warning limits, six values consecutively increasing or decreasing and eight values on one side of the central line and also every unusual pattern (see Figure 5).

Prior to using the charts two cycles have to be completed, which are usually part of the validation of an analytical method. In the first cycle, the statistical parameters of the process are assessed. It is important to investigate these data thoroughly, e.g. using histograms or $P–P$ plots. Realistic limits for the control chart can only be estimated if all possible variations have been included, e.g. temperature variations and different analysts. For example, establishing meaningful warning and control limits for the $\bar{x}$-chart requires that the mean scatters symmetrically around the central line and no trend can be observed. A validation cycle should be used to check these limits. If an unexpected number of out-of-control situations is observed, the limits have to be redefined.

Another chart often used in analytical QC is the Cusum chart, which gives information on small shifts and trends often observed due to ageing of instrument parts or stock solutions and changes of environmental conditions, e.g. temperature. The cumulative sum (Cusum) is calculated accordingly to Equation (6):

$$\text{Cusum} = \sum_{i=1}^{t} (x_i - \mu_0) \quad (6)$$

where t is the time sequence, x_i are the observations and μ_0 is the mean of the undisturbed process, usually estimated

Table 4 Factors for the estimation of σ

Factor	n								
	2	3	4	5	6	7	8	9	10
d_2	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078
d_3	0.8525	0.8884	0.8798	0.8641	0.8480	0.8332	0.8198	0.8078	0.7971
c_4	0.7979	0.8862	0.9213	0.9400	0.9515	0.9594	0.9650	0.9693	0.9727

$\bar{x}$-chart	Observation	Cause (example)
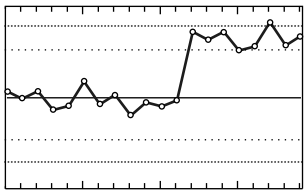	Mean exhibits trend	Changing instrumental condition, e.g. temperature, ageing and contamination of reagents, memory effect
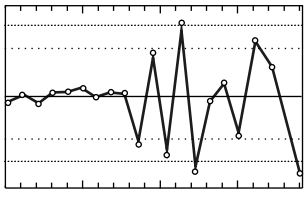	Sudden change of scattering of the process average	New personnel, method or instrument
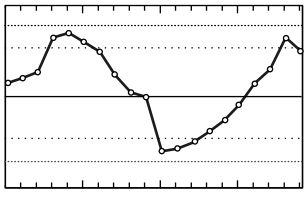	Cyclic variations of mean	Day–night temperature variation

Figure 5 Typical patterns and possible causes of out-of-control situations for the $\bar{x}$-chart.

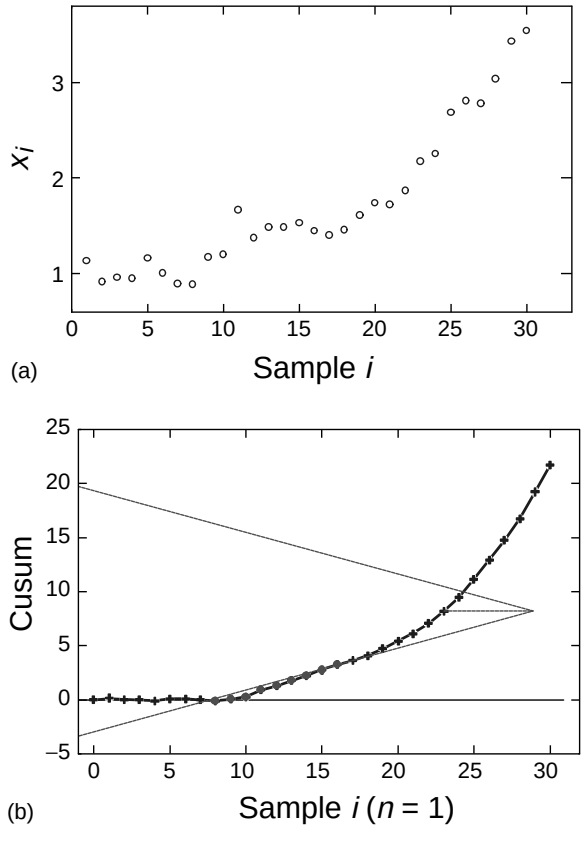

Figure 6 Cusum chart (a) course of measurement values and (b) Cusum chart from these data.

by calculating $\bar{x}$ from the process history. Usually individual observations x_i are used to attain increased sensitivity in comparison with the $\bar{x}$-chart. As long as the only random deviations affect the observations, the Cusum scatters around μ_0. If shifts or trends occur, then the course of the Cusum shows a certain slope (Figure 6a and b).

If the slope exceeds a certain value which is based on statistical assumptions then the process is declared out of control. The limits can be estimated either graphically by using a V-mask or by computation of the slope.

Various process capability indices and operating characteristic curves are described in the literature to assess the performance of control charts. An important characteristic is the average run length (ARL) The ARL is the average number of samples before a value falls falsely outside of the control limit (type I error) and should be high. Figure 7 shows an operation characteristic curve which shows the probability with which a certain shift of the process can be detected for various samples sizes n with the $\bar{x}$-chart using a 3σ interval.

Other charts which display time series are the moving average (MA) and the exponentially weighted moving average (EWMA) charts. When several process variables are observed which are correlated and thus cannot be controlled independently, there is an increased probability of finding false out of control situations (Type 1 error). In this case multivariate control charts

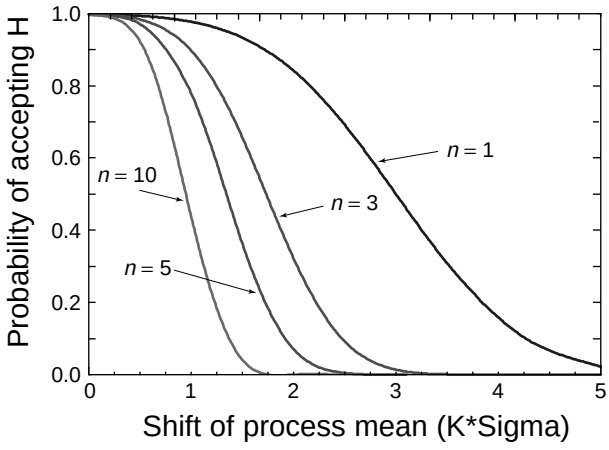

Figure 7 Operation characteristic curve for the $\bar{x}$-chart using a 3σ interval.

can be used which are based on principal component analysis, e.g. Hotelling's T^2 statistic.[26]

2.4 Regulatory Aspects and Standards

In order to assure minimal requirements for the quality of analytical results, a number of standards and legal regulations for the analytical sector have been developed. Rules, standards and legislation related to laboratory work form the basis for national and international acceptance of analytical results and for the accreditation of laboratories. The compliance with regulatory and legal requirements varies in different countries and may also depend on the scope of the laboratory, such as first party, representing the vendor or manufacturer, second party, representing the client, or an independent third party laboratory. Generally legal regulations exist for certain branches such as medical and health, pharmaceutical protection and safety, explosives, food and feed related and forensic analyses.

Some standards and guidelines relevant for analytical laboratories are the following: ISO/International Electrotechnical Commission (IEC) Guide 25[27] (General Requirements for the Competence of Calibration and Testing Laboratories, will soon be ISO 17025, General Requirements for the Competence of Testing and Calibration Laboratories), European Norm (EN) 45001[28] (General Criteria for the Operation of Testing Laboratories) for laboratories in the EU (will be replaced by the ISO 17025), the ISO 9000 series of standards[29] addressing quality systems (in the EU the identical EN 29000 series, in the USA ANSI/ISO/ASQC Q9000 series) and the good laboratory practice (GLP) standards and regulations [Organization for Economic Cooperation and Development (OECD) Principles of GLP,[30]] Environmental Protection Agency (EPA) GLPs[31,32] and Food and Drug Administration (FDA) GLPs.[33] These standards and guidelines are regularly reviewed and modified.

The technical competence of the laboratory is addressed by the ISO/IEC Guide 25 and the EN 45001, both of which focus on the training and experience of personnel, appropriate equipment, calibration and maintenance procedures, the laboratory organization and the environmental conditions in the laboratory. The new standard ISO 17025 (Table 5)[34] integrates EN 45001 and will form the basis for worldwide acceptance of analytical results and for the assessment of laboratories by accreditation bodies. Laboratories operating under ISO 17025 have to implement a quality system according to ISO 9001/2. ISO 17025 addresses all kinds of laboratories (except those working in legally regulated sectors such as clinical, safety or pharmaceutical laboratories) including testing, calibration and sampling laboratories.

Table 5 Contents of the ISO Draft International Standard 17025[a], 'General Requirements for the Competence of Testing and Calibration Laboratories'

1. Scope
2. Normative references
3. Definitions
4. Management system requirements
   - 4.1. Quality management system
   - 4.2. Organization and management
   - 4.3. Document and information control
   - 4.4. Request, tender and contract review
   - 4.5. Sub-contracting of tests and calibrations
   - 4.6. Procurement of services and supplies
   - 4.7. Service to and feedback from customers
   - 4.8. Control of nonconforming work
   - 4.9. Corrective action
   - 4.10. Preventive action
   - 4.11. Records
   - 4.12. Internal audits
   - 4.13. Management reviews
5. Technical requirements
   - 5.1. General
   - 5.2. Personnel
   - 5.3. Accommodation and environmental conditions
   - 5.4. Test and calibration methods
   - 5.5. Equipment
   - 5.6. Measurement traceability
   - 5.7. Sampling
   - 5.8. Handling of test and calibration items
   - 5.9. Assuring the quality of test and calibration results
   - 5.10. Reporting the results
   - 5.11. Records

Annexes

A Cross-references between ISO/IEC Guide 25 and ISO 9001/2

B Guidance for establishing specific requirements for laboratory competence

C Bibliography

[a] ISO/CASCO Working Group 10/40, Draft ISO 17025.

These guides and standards only address general requirements of QA for various types of laboratories. Therefore, detailed guidance documents have been prepared by various international or national organizations, e.g. by International Laboratory Accreditation Cooperation (ILAC), which interpret the various standards and guides with regard to the practical application in the analytical laboratory.

The ISO 9000 series of standards, especially ISO 9001 and ISO 9002, address the capability to implement and run a quality system. ISO 9000 defines a quality system as ''the organization, structure, responsibilities, procedures, processes and resources for implementing quality management.'' The goal is to improve the operational processes of the laboratory in order to minimize human errors. Operational processes relate only to the definition and documentation of operation procedures. The ISO 9000 series (Table 6) does not address the technical competence of the laboratory or the equipment and premises needed to perform analyses.

GLP is a concept covering the organization and conditions under which nonclinical laboratory studies are planned and conducted in order to assess chemicals relating to the safety and protection of humans and the environment. The aim is primarily to prevent fraud, and to assure transparency and clear responsibilities by extensive documentation of studies relating to research and marketing permits of products in fields such as pharmaceuticals, food and feed additives or pesticides. In contrast to the above-mentioned voluntary standards and guidelines, regulations according to GLP are in many countries mandatory for studies in the above mentioned fields such as pharmaceutical, toxicological studies or pesticide residue studies.

The primary aim of GLP documentation can be summarized as documenting 'who did what and when it was done'. The GLP standards requires a separate quality assurance unit (QAU) which supervises the conformance of the regulations. The requirements concerning facilities and documentation are more comprehensive than those found in EN 45001 or ISO/IEC Guide 25. Owing to the toxicological relevance of the acquired data, GLP also has also stricter rules for the storage and archiving of documents and test items. The development of the GLP Regulations was put forward by the FDA in 1978 following irregularities in toxicological studies. The GLP Regulations were later adopted by the US EPA for pesticide and chemicals metabolism and residue data in the OECD member countries.[30]

Table 6 The ISO 9000 series of standards

Number	Title	Last updated
ISO 9000	Quality Management and QA Standards:	
	Part 1: Guidelines for Selection and Use	1994
	Part 2: Generic Guidelines for the Application of ISO 9001, ISO 9002 and ISO 9003	1997
	Part 3: Guidelines for the Application of ISO 9001:1994 to the Development, Supply, Installation and Maintenance of Computer Software	1997
	Part 4: Guide to Dependability Program Management	1993
ISO 9001	Quality Systems – Model for QA in Design, Development, Production, Installation and Servicing	1994
ISO 9002	Quality Systems – Model for QA in Production, Installation and Servicing	1994
ISO 9003	Quality Systems – Model for QA in Final Inspection and Test	1994
ISO 9004	Quality Management and Quality System Elements:	
	Part 1: Guidelines	1994
	Part 2: Guidelines for Services	1991
	Part 3: Guidelines for Processed Materials	1993
	Part 4: Guidelines for Quality Improvement	1993

3 QUALITY ASSURANCE IN THE ANALYTICAL PROCESS

In order to provide confidence that the analytical result satisfies the stated quality requirements, the laboratory has to take appropriate measures in every step of the analytical process. Prerequisites for developing and using valid analytical methods are appropriate personnel and technical competence. These cover the organizational structure of the laboratory and also instrumentation, maintenance and reagents. With respect to the consistency of the quality measures taken, special attention has to be given to the documentation, which comprises recorded data and laboratory documents. The central document is the quality manual in which the policy of the laboratory and all elements of QA are addressed. This section is primarily orientated towards the requirements of ISO/IEC Guide 25 (ISO 17025).

3.1 Sources and Significance of Errors

Various sources of uncertainty in the analytical process, from sampling to reporting, lead to a certain amount of error (deviation from the true value) of the analytical result. Two components can be attributed to the observed error, namely random error and systematic error.

Random error arises from unpredictable variations of the quantities affecting the analytical result and manifests

itself in the precision of repeated measurements. The random component of error cannot be compensated by correction but can be reduced by computing the mean value of an increasing number of measurements. Often the amount of random error depends on the measured value.

Systematic error remains constant or varies in a predictable way in repeated analyses and leads to inaccurate results. Sources of constant systematic error are, for example, uncorrected reagent blanks, incomplete extraction and inaccurate calibration. Systematic errors may affect the analytical result in different ways (Table 7).

Figure 8 shows the influence of the different types of systematic error on the observed values x^*.

Systematic errors varying over time can often be attributed to variations of the experimental conditions such as temperature, induced instrument drift or ageing of parts. A reduction of systematic error components can often be achieved by correction by correction of a blank value or a correction factor, e.g. correction of a recovery rate from analyte extraction. Another type of error is spurious error due to human mistakes or instrument failures. Typical examples are confusing the samples, transposing digits in an instrument reading or extreme contaminations of the sample. These errors can be considered as an extreme manifestation of random or systematic error and should usually be eliminated from the data set. Random spurious errors can be detected by statistical outlier tests on replicate measurements. Table 8 shows some examples of possible sources of error in trace analysis, together with the contribution to the analytical result.

Problems with analytical work can, for example, be identified by SPC (section 2.2), staff observations, complaints by customers or internal audits which should be conducted regularly in order to assure that the laboratory work is consistent with the quality system.

Table 7 Types of systematic error

Type of systematic error	Observed value x^*	Example
Constant term	$x^* = x + a$	Unrecognized blank value Contamination
Multiplicative (proportional) term	$x^* = bx$	Inaccurate calibration factor Incomplete recovery
Nonlinear term	$x^* = f(x)$	Linear calibration model in case of a nonlinear concentration–signal dependence

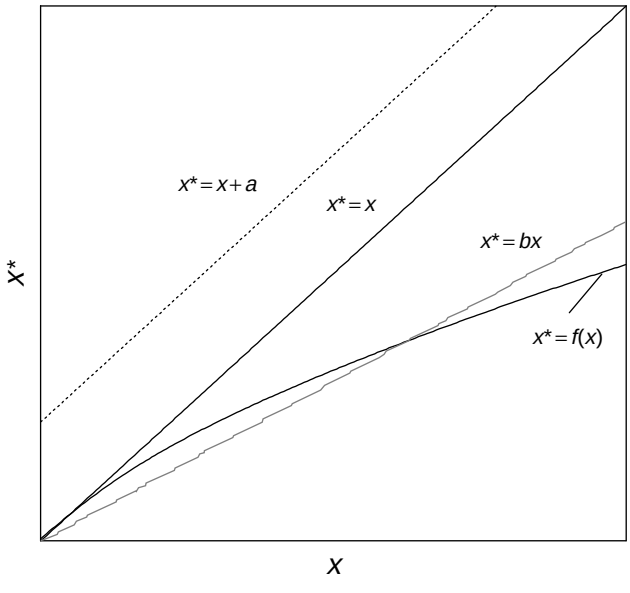

Figure 8 Influence of the different types of systematic error on the observed values x^*.

Table 8 Possible sources of error in trace analysis

Analytical step	Cause	Reduction
Sampling	Contamination of sample	Cleansing of containers
	Sample not representative	Homogenization, different sampling scheme, proper sample labeling
Preparation	Weighing	Calibrated balance, temperature control
	Pipetting	Calibrated glassware
	Moisture	Drying of sample
	Reagent contamination	Reagents of higher purity, blank measurements
	Contamination from tools	Acid washing or steaming, PTFE containers, blank measurements
	Adsorption	Acid-washed materials, PTFE containers
	Contamination from laboratory environment	Clean bench, clean room, glove-box, dust filter, reduction of air circulation
Measurement	Instrument stability	Warm up instrument, air conditioning
Data evaluation	Inadequate calibration model	Method validation

PTFE, polytetrafluoroethylene.

3.2 Organization, Personnel and Premises

Constant quality of the laboratory work requires a clear organizational structure (usually documented in an organizational diagram) with responsibilities and duties clearly assigned to its staff members. Responsibilities

include the technical operations and resources of the laboratory. One particular member of the staff should be responsible for ensuring the quality of laboratory operations (part-time or full-time, depending on the size and the resources of the laboratory). A prerequisite for quality in the laboratory work is that the staff members have appropriate competence, training and experience. They should attend external or internal training courses. For unfamiliar analytical tasks appropriate supervision should be available. Records have to be kept, which include information on qualification, additional training and methodological expertise of the staff members. ISO/IEC Guide 25 (ISO 17025) furthermore requires that the laboratory and its staff have to be free 'from any commercial, financial and other internal and external pressures that may adversely affect the quality of their work'.[27]

Procedures have to be established to implement corrective action in an acceptable period when problems with the quality management system or with the technical operations of the laboratory have been identified. This should include an investigation of the cause of the problem. A guideline for performing internal audits is published by the EAL.[35]

3.3 Sampling and Sample Preparation

The sampling procedure can strongly affect the analytical result and its uncertainty and depends on the purpose for which the analytical result is to be used.[36] Generally, for analysis, only a small amount, the sample, is drawn at a certain time from the total bulk of material of interest.

Two aspects have to be considered. First, the sample taken must show average properties, spatially and temporally, of the medium being sampled. Spatial resolution presents a problem with heterogeneous or grainy solid materials, temporal resolution is critical for example, in process analysis or environmental analyses of rivers. The second aspect of representativeness addresses the reliability of the determination of the uncertainty of sampling as a measure of the variability of analyte present in a heterogeneous sample volume.

The analytical result contains the uncertainties respectively of the standard deviations of all steps (Equation 7):

$$u_{\text{analysis}} = \sqrt{u^2_{\text{sampling}} + u^2_{\text{preparation}} + u^2_{\text{measurement}} + u^2_{\text{evaluation}}} \quad (7)$$

Since sampling uncertainty can be only assessed after preparing and analyzing the samples, analysis of variance in combination with statistical designed experiments or collaborative trials has to be used in order to estimate the sampling uncertainty.[37–39] The degree of spatial and temporal resolution depends on the problem definition, on the sampling strategy, on the sample mass and on the analytical method. For example, different strategies have to be used in order to determine local or average concentration of a trace contaminant in an alloy. While local contamination can be assessed by analysis of small sample masses, e.g. by laser ablation inductively coupled plasma mass spectrometry (ICPMS), determining average contamination requires higher sample masses, using ICPMS after sample digestion or electrothermal vaporization. In environmental and geological analysis, but also in food analysis, the uncertainty of the sampling step is often the strongest contribution to the overall uncertainty of the analytical result. An example comparing the consequences of different sampling schemes for soil analysis is given in the literature.[40]

Even for homogeneous materials, sampling, sample transport and sample storage can present problems related to sample contamination, loss of analyte or inappropriate sampling procedures. If the sample is not taken by the laboratory then the person who takes it has to be instructed in order to prevent ambiguities, e.g. appropriate sample mass.

The chain of traceability from the primary sample to the sample prepared for analysis has to be fully documented, especially if deviations from the usual protocol occur, in order to assess the representativity of the analytical result of the sample in relation to the medium being sampled. The documentation should include type and source of container, preparation techniques and cleaning procedures if the containers are reused. The interval between sampling and analysis should be as short as possible. Usually samples are stored in a cool, dark place. Prior to analysis the authenticity and the suitability of the sample have to be controlled. The sample has to conform to the sample description, e.g. whether the sample container is sealed.

Analyte loss primarily depends on the nature of the sample and is especially relevant in trace and ultratrace analysis. For example, degradation of organic samples by light or microbiological activity can be reduced by using dark bottles, cooling with ice or stabilizing reagents. Volatile analytes have to be sampled in closed and completely filled containers. Precautions against oxidation must be taken if the aim is chemical speciation. Special attention has to be given to the material of the sample container with respect to:

- short- and long-term resistance against sample, added reagents and, if necessary, pressure and temperature;
- leachability of trace contaminants, filler, lubricant, catalyst;
- sorption/desorption of analytes;
- permeability of organic liquids through the walls of the material;
- cleaning procedures.

It is well known that cations, including iron, copper and lead, are subject to loss due to adsorption and ion exchange, especially on glass surfaces, which is particularly relevant in the sub-micrograms per gram region. This can be minimized by acidification with nitric acid. Silica and sodium can also be leached from glass and contaminate the sample. Polyethylene (PE) or polypropylene (PP) containers contain traces from catalysts, e.g. aluminum, titanium and others, which can be leached. Containers made from fluorinated polymers such as PTFE are to be preferred. Liquid organic samples have to be compatible with plastic material. Volatile organic compounds may migrate into the walls of plastic containers or may leach substances such as phthalate esters used as plasticizers. Glass containers or containers made from fluorinated polymers are therefore preferred for samples such as volatile organic substances, oil and grease. If plastic caps have to be used, they have to be PTFE lined.

Every sample container has to be labeled in order to achieve unique identification. Sample records should contain the unique sample identification, date, time, analytes or methods, added preservatives, potential hazards, location and the name of the person who took the sample. To assure legible and complete information, preprinted labels should be used.

Contamination can be controlled by analyzing blank samples. The inclusion of a sample container blank may be adequate. If the sample blank exceeds the specified limit then appropriate measures have to be taken. For inorganic trace analysis these may comprise high-purity reagents, metal-free tools, and the use of glove-boxes, clean benches or clean rooms. Analyte loss can be controlled by using spiked samples at varying stages of the sample preparation and by calculation of the recovery rate (Equation 4). The quantity added should be based on the typical values found in the sample. A prerequisite for realistic assessment of recovery rates is that the spiked analyte is of the same chemical species (oxidation state, binding state) as that in the sample.

Sample preparation covers methods such as filtration, centrifugation, extraction methods, purging and trapping, microcolumns, digestion methods and grinding dependent on the type of sample. Problems mainly arise from contamination by reagents, especially acids and solvents, and the environment, e.g. dust, and loss of analyte.

3.4 Reagents and Standards

Various grades of analytical reagents exist, ranging from technical grade to ultrapure grade depending on the specifications of the supplier. Often the method specifies the required grade; if this is not the case, usually analytical-grade reagents are used. Special attention has to be given to the purity of the reagents and solvents, which is of utmost importance in high-performance liquid chromatography (HPLC), ICPOES and ICPMS and electrochemical methods. Depending on the purpose, water or acids may have to be further purified, e.g. by sub-boiling distillation. Solvents and solid reagents often have to be dried prior to use.

Supplied chemicals should be marked with the date of arrival and the date of opening in order to control the shelf-life. Information on potential hazards, restrictions of use and reagent disposal must be available. Storage conditions depend on the chemicals, acids and volatile solvents should be stored under hoods or in ventilated cabins. Organic reagents should generally be stored in cool, dark places. In order to reduce deterioration of the initial grade of the reagent by contamination, spatulas, pipets and other tools should never be used directly in reagent bottles. Instead, an approximate amount of the reagent is dispensed into a clean vessel. If reagent is left over it is never returned to the reagent bottle. For acids, solvents and other liquids, bottle-top dispensers made from fluorinated polymers are ideal for dispensing defined volumes directly from the reagent bottle without risk of contamination. Standard solutions and other laboratory-prepared reagents have to be stored in an appropriate container and labeled, dated and signed. Records should exist describing the preparation of the a standard, the source of the reagents and all weight or volume operations. To prevent cross-contamination, standards should not be prepared and stored in the sample preparation or storage areas.

Reagent-grade water is a prerequisite for most analytical work. Since feed water always contains impurities, such as dissolved salts and gases, organic compounds, suspended particles, microorganisms and their toxins, techniques such as distillation, ion exchange, carbon adsorption, reverse osmosis, microporous membrane filtration, and ultrafiltration have been used in order to produce water. As shown in Table 9, with each of these

Table 9 Comparison of water purification techniques for laboratory use

Method	Contaminants removed[a]			
	Ionized	Un-ionized	Pyrogens and endotoxins	Particulates
Distillation	++[b]	++[b]	++	++
Activated carbon	−	−	−−	−−
Ion exchange	++	−−	−−	−−
Reverse osmosis	+	−	++	++
Ultrafiltration	−	−	+	++

[a] Removal of contaminants: ++, nearly total; +, good; −, poor; −−, very poor.
[b] If vapor pressure of the contaminant differs from that of water.

methods only specific types of contaminants are removed. Therefore, for the production of ultrapure water several methods have to be combined, e.g. reverse osmosis and distillation.[41] Important parameters characterizing the suitability of water for analytical purposes are conductivity, total organic carbon (TOC), dry residue and ultraviolet (UV) absorbance. Cartridge-based systems in particular have to be constantly controlled for possible exhaustion or failure.

3.5 Instruments and Computers

Laboratory equipment can be categorized as consisting of measuring instruments, such as spectrometers, chromatographs, balances, thermometers and timers, volumetric equipment, such as pipets, burets and volumetric flasks, general equipment, such as refrigerators, ovens, stirrers and non-volumetric glassware, and computer equipment including printers, backup systems and network components. The level of maintenance necessary varies from cleaning to regular renewal of wearable parts. Measuring equipment has to be calibrated frequently. The calibration interval ranges from daily, or prior to each use, such as spectrometers or pH meters, to monthly or annual calibration intervals as in the cases of balances and timers. Other equipment which influences the analytical results, such as volumetric glassware, refrigerators and ovens, should be checked regularly.

For major pieces of equipment documentation should exist which includes manufacturer, type, serial number or other unique identification, year of manufacture, operating conditions and installation requirements (environment temperature, cooling water, electricity and gases), modifications (description, when, by whom) and application areas, cleaning, calibration and maintenance procedures and intervals.

Analytical instruments, e.g. spectrometer systems, usually consist of various subsystems. It is therefore necessary to evaluate the equipment consisting of electronics, analytical operations and samples as an integral system. For example, a spectrophotometer system comprises mechanical and optical components, such as light source, monochromator and detector, hardware and software for instrument control, data acquisition, data evaluation and printing or displaying the measurements. The system should therefore be tested using control samples with concentrations over the whole specified range. Test criteria for a spectrophotometer are accuracy of wavelengths, signal-to-noise ratio and linearity of the absorbance response. The recorded data, e.g. spectra, are then compared with previously recorded spectra, for example provided by the vendor, regarding peak positions, intensities and relative standard deviation (RSD) of the measured signal. This process is often referred to as operational qualification or system verification.[42] Table 10 shows some parameters for operational qualification or verification for different types of instrumentation. Computer hardware and software components of analytical instruments can also be tested by using stored test data files. These tests are employed to check compatibility after hardware components have been changed or instrument software or operating system updates have been performed. The results of these tests should be automatically documented.

Table 10 Parameters for operational qualification or verification of instrumentation

Equipment	Component	Parameter
Spectrometers (e.g. AAS, ICPOES, UV/VIS)	Wavelength scale	Accuracy, stability, precision
	Source	Intensity, stability
	Detector	Resolution, signal-to-noise ratio, linearity, stability, calibration
Chromatographs (e.g. GC, HPLC, IC)	Column	Resolution, capacity, retention
	Detector	Linearity, selectivity, noise, stability
	Autosampler	Precision, accuracy

UV/VIS, ultraviolet/visible; GC, gas chromatography; IC, ion chromatography.

In pharmaceutical analysis, the combination of method, software and hardware validation is referred to as system suitability testing. Appropriate methods are described in several pharmacopeias, e.g. the USP.[43] For example, in order to check if a chromatographic system is suitable to perform a particular analytical task, several parameters, including resolution, repeatability and capacity are determined using a test solution.

Today, laboratories use analytical instruments which rely on computers for control and data acquisition, storage and evaluation. Furthermore, data are stored in databases, e.g. using a laboratory information management system (LIMS), which in many cases are networked. Hardware and software components are the frequent cause of failures, data loss and erroneous results. Since the quality of software and hardware greatly influences the reliability of an analytical method, specific GLP regulations relating to computers and computerized instruments exist. These are the OECD GLP consensus document 'The Application of the Principles of GLP to Computerized Systems'[44] and the EPA Good Automated Laboratory Practices (GALPs). The GALPs are applicable for computerized instruments when it is possible for the user to modify the data, defined as LIMS by the EPA.[45] Validation of

computerized analytical instruments is also covered by Huber.[46]

Software comprises all programs and program systems necessary for controlling the hardware, e.g.

- application software which may be commercially available, such as word processors, databases and spreadsheets, or by the laboratory, such as customized data evaluation procedures;
- system software, such as operating systems, system utilities, compilers and diagnostic software;
- firmware, which is software that is embedded in a hardware device that allows the execution of a program or reading of data but does not allow modification by the user, e.g. system BIOS (basic input output system) or signal processing hardware;
- data files, such as raw and processed data and archives;
- documentation files, such as standard operating procedures (SOPs) and operating manuals.

Hardware and especially software components are subject to frequent modifications. Changes comprise new peripheral devices, new firmware, changes to documentation and modifications of programs and data, etc. Since all components interact, the installation of a new software version may lead to complete loss of data. Therefore, a configuration management system, often termed ''change control'', must exist to ensure that changes are only made by authorized staff and followed by adequate tests. The procedure requires documentation of all changes and results of the validation tests. Version numbers have to be accessible to the user. It should always be possible to restore the previous software versions, if problems with new versions occur. Because life cycles of software versions and hardware are limited, special attention has to be given to ensure compatibility with archived data. This can mean that outdated hardware components, e.g. drives or printers, may have to be maintained to access the archives.

Preventive measures have to be taken in order to assure the data safety of computer-based systems. Data loss and damage can occur by human mistakes, software and hardware failures and environmental influence. Therefore, regular backup procedures have to be scheduled. The write-protected backup copies have to be archived in a safe place. In addition, data safety provisions have to be taken against intentional manipulation of the data, a problem which has increased due to networking. While computer software is an immaterial product and not subject to ageing, the media, e.g. disks and tapes, on which software and data reside can deteriorate with time. Usually hard copies (printouts) are archived together with the electronic archives because the task of copying and verifying the electronic copies is time and cost intensive.

Software QA, verification and validation are a field on their own.[47,48] The ISO/IEC 9126[49] states six standard basic characteristics of software quality, namely reliability, functionality, usability, efficiency, maintainability and portability. The quality of software algorithms greatly influences analytical results. For example, poor peak integration algorithms may lead to unreliable results if the background noise is high or peaks are not well separated. One approach to software validation is by testing the program with reference data sets, which must be representative for the application. Another approach is to use computer-generated data with known solutions. Reference data sets for statistical software such as regression calculation are provided online, for example, by NIST.[50] In contrast to these ''black box'' approaches, i.e. no information about the structure of the program is necessary in order to perform the tests, 'glass box' tests assess the structure of the program, e.g. control structures. Often reference software which has been validated by use, due to the number of installations, exists for a problem. In this case the results achieved with the test data have to be compared with the reference software. Simple programs can be verified by investigating all possible input values together with the given output.

3.6 Reporting Results

The results of analytical determinations have to be clearly and unambiguously documented in a test report or certificate. This report has to include the address of the laboratory, identification and description of the samples, date of sample receipt, a short description of sampling and analytical procedures and analytical result with the accompanying uncertainty, and has to be dated and signed. It is important to make clear whether the analytical result relates to the whole sample or a portion, to an extract, to dry or wet sample mass, etc.

To avoid ambiguity when reporting quantitative results, special attention has to be given to the number of digits and to the units. Only significant figures are reported and include a single doubtful digit. The measurement uncertainty has to be considered, which means that the number of digits has to be justified. For example, if the result of an analytical determination is reported as being $20.4\,\mu g\,L^{-1}$ then it is expected that the last digit is uncertain. If, on the other hand, the expanded uncertainty is known to be $u_c = 1\,\mu g\,L^{-1}$ then the concentration value should be rounded to $20\,\mu g\,L^{-1}$. Preferably the uncertainty of the result should be included, especially in cases where there is doubt about the number of significant digits because of zeros in the last digits, leading to $20 \pm 1\,\mu g\,L^{-1}$. Results relating to a concentration or

content must never be reported as zero. An acceptable way is to report a result as being smaller than a certain limit, e.g. the quantitation limit.

In research and development and also in product conformance testing, the customer is often interested in the opinion of the laboratory in relation to the stated problem. This is usually termed ''professional judgment'' and comprises opinions, conclusions and interpretations based on the measurements and observations and also on experience, knowledge and other sources of information such as literature. Professional judgement is therefore based on facts and observations but includes a certain level of subjectivity. The judgement should not be biased by political opinions. It depends on the relevant standards and regulations whether professional judgement may be included in the report. A problem is that the ability of the laboratory to perform qualified professional judgment is difficult to assess.

3.7 Documentation and Software

The documentation of a laboratory can be subdivided into three categories: documents contain instructions and guidelines regarding quality measures and operating procedures; records comprise all data which demonstrate the work done; and organizational documents contain data which have to be changed frequently, such as a list of reagents.

The documents are organized hierarchically. The quality manual forms the top of the hierarchy and addresses all relevant topics which influence the quality of the laboratory work. The main purpose of the quality manual is to achieve internal transparency, e.g. documentation of responsibilities, and external transparency, e.g. for customers and accreditation bodies, of the work performed. An important point is the description of the scope of the analytical work performed by the laboratory such as analytical methods or fields of work, for which appropriate staff training, facilities and resources have to be accessible. This is needed in order to decide if a contract can be accepted by the laboratory and if the contract contains routine or new work. For accredited laboratories this includes the scope of accreditation.

The contents of a quality manual usually include a quality policy statement by the management, the definition of organizational structure of the laboratory, description of the scope of the analytical work, qualification and training of personnel, document control, sample handling, a description of premises, instruments and reagents with references to more detailed documents, internal QC schemes, aspects concerning safety, health and waste disposal, procedures for ensuring traceability of the results, procedures for reporting analytical results, procedures for corrective action following deviations from documented policies or customer complaints, procedures for dealing with subcontractors and procedures executing internal audits. Usually the quality manual contains references to external documents describing the relevant legal requirements, standards and guidelines and internal documents such as a SOP and sample forms of protocols, e.g. for reporting results.

SOPs are formalized documents which describe in detail all routine work repetitively performed in the laboratory relating to analytical methods and instruments but also to procedures such as reporting results, document handling and writing a new SOP. In this way consistency of the work, traceability of the results and aspects regarding work safety, environmental and economical aspects such as waste production and reagent use are documented and well regulated. Furthermore, SOPs help auditors and customers to assess the work done in the laboratory. SOPs usually contain the following information prior to the written procedure: title, author, approved by whom, released by whom, date until valid, revision date, version history.

Systematic document control has to be performed in order to assure that only current versions of the documents are used and that revised documents are properly archived for future access. This consists of measures for approval, release, modification, drawback, identification, indexing, storage, retrieval and distribution of documents. All relevant documents and records have to be accessible for a defined time in order to allow historical reconstruction of analytical work.

Reports and certificates containing the results have to be clearly written and laid out in order to prevent misuse or misunderstanding by the customer. If ambiguous terms are used, e.g. limit of detection, then reference to the definition should be included. Amendments should be not made in the report but as a separate or a new document which supersedes the old one.

For laboratories working under GLP, archiving is mandatory for all documents, raw data, which are data necessary to repeat the data evaluation, protocols, specimens and reports. For example, the German adaptation of the GLP[51] requires that raw data and other records have to be stored for a period of 30 years. For other laboratories it is advisable to keep records such as sampling records, raw data from calibrations and measurements, calculations for processing the data, calibration records, records from SPC, analytical results and certificates, staff records and records from audits for some years, so that the long-term performance of the laboratory can be assessed and possible problems can be identified retrospectively. Records should always be signed and dated, and possible confidentiality, especially of analytical results, has to be preserved. Archived records have to be protected from deterioration and loss.

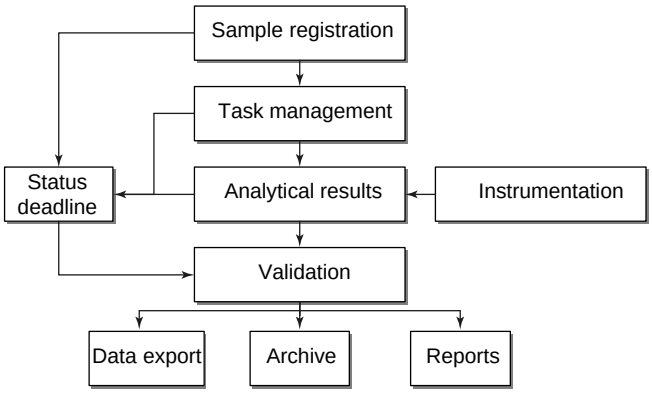

Figure 9 Information flow in a LIMS.

Paper documentation is increasingly being replaced by data processing. With the advent of laboratory automation and computer networks, specialized database systems were introduced in the laboratory. These LIMS model the sample flow of the laboratory (Figure 9)[52] and fully document the ''life cycle'' of a sample. Typical tasks of a LIMS comprise registration and labeling of samples, entering results, generating reports, management and version control of documents such as SOPs and management of instrument calibration and maintenance data. Furthermore, often, data such as customer addresses or accounting data are processed and stored, which may lead to additional requirements regarding the confidentiality of these data. Advantages of electronic over paper documentation are that consistency, completeness and access of the data can be controlled. Depending on the authorization of the user of the LIMS, different access levels are possible, while simultaneous access by different users is possible. In many cases an audit-trail function records all actions performed by the user. Today most LIMS use client–server architecture using an SQL (structured query language) database server which can be accessed simultaneously by several workstations in order to query, modify or complete the data sets.

4 METHOD DEVELOPMENT AND VALIDATION

In order to demonstrate that an analytical method is suitable for the required purpose, the method has to be validated. An important aim is to determine systematically certain performance characteristics of the analytical method, e.g. limit of detection or linearity, in order to demonstrate the technical capabilities of the method. The ISO definition states that validation is ''confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled'' (ISO 8402).[1] A prerequisite for method validation is that the equipment used has to be checked and clearly described.

The amount and type of validation data depend on the intended purpose of the method. Newly developed methods always have to be validated to some extent. If the method has been developed especially for a certain customer then the degree of validation will be lower than for methods which are to be used routinely. Even well-established standard and official methods need to be validated to a certain degree when performed with different laboratory equipment and reagents and by different analysts. The highest degree of validation is of course necessary in order to establish new standard or official methods, such as the Association of Official Analytical Chemists (AOAC),[53] EPA or Deutsche Institut für Normung (DIN) methods. In this case interlaboratory studies are mandatory. Furthermore, the degree of validation depends on the consequences of an erroneous result. As ISO Guide 25 states: ''Validation is always a balance between costs, risks and technical possibilities''.[27] Therefore, comprehensive method validation is mandatory in pharmaceutical analysis. Minimum criteria and requirements of the analytical method should be listed before the method is developed and validated.

Various validation strategies and procedures are described in the literature.[5,54–58] Terms are often not used consistently, but the procedures and parameters discussed are generally identical. Therefore, it should always be stated which convention has been used, e.g. for reporting the limit of detection. Certain validation procedures may be required by regulations, e.g. in the pharmaceutical sector. Performance characteristics typically assessed are precision, accuracy, linearity, working range, specificity/selectivity, limits of detection and quantitation and robustness/ruggedness of the method. The final step of the validation procedure is a written documentation of the method SOP, which contains the validation data. The validation data also form the basis for SPC measures.

Some possible approaches to method validation are as follows:

- Validation using RMs, which is the most common approach for intra-laboratory method validation. The reliability of the validation procedure depends on the quality of the RM. If no RM is available, spiked samples can be used, with limitations.
- Comparison with another analytical method. The results obtained using the analytical method to be validated are compared with a reliable, previously validated method.
- Comparison with other laboratories by a collaborative study.[59] Parameters which can be obtained include uncertainty, ruggedness and reproducibility,

while systematic deviations cannot be assessed. In many cases the use of collaborative studies for method validation is impractical because the intended scope of the method is narrow or, as is often the case in pharmaceutical analyses, the samples must not be made available to competitors.

4.1 Calibration and Recovery Studies

With a few exceptions, e.g. titrimetry, analytical methods require adequate calibration. In this case the calibration forms the basis of the method validation. The most common model is the linear model with the functional relationship in Equation (8):

$$\hat{y} = a + bx + \varepsilon \tag{8}$$

fitted by least-squares regression, where $\hat{y}$ is the estimated measurement value depending on the concentration or content x of the calibration sample, b is the slope of the calibration function and a equals the blank value, e.g from solvents or reagents. The model deviations (residuals) are $\varepsilon = \hat{y} - y$. If the blank value obtained from the calibration sample and from the sample which is to be analyzed differ significantly from each other, it has to be corrected for. Assumptions for using the model according to Equation (8) are (1) a linear relationship or one that can be linearized by transformation, (2) the concentration error of the calibration standards is negligible compared with the measurement error, (3) the measurement errors follow a normal distribution, (4) homogeneity of the variances, which means that the precision of repeated measurement remains constant over the working range. If deviations from assumption (1) are observed, either the working range has to be reduced or a nonlinear model has to be used. If assumption (2) is violated severely, e.g. $s_x \approx s_y$, then error-in-variables regression models may be used.[60] Violation of assumption (3) is usually unimportant except in severe cases where outliers are observed, e.g. problems with the instrument. Outliers have to be eliminated prior to estimation of the regression parameters, preferably by using statistical methodology. For asymmetric distributions transformations can be used. The assumption of homogeneity of variances often does not apply, and most methods, e.g. ICPOES, show an increase in absolute signal variance with increasing concentration while the RSD ($s/\bar{x}$) usually decreases or remains constant. Reduction of the working range or taking into account of the functional relationship between concentration and standard deviation of the observed signal by using weighted regression may be appropriate.

Recovery studies are performed in order to assess signal depression or increase which is caused, for example, by incomplete analyte preconcentration or extraction steps or matrix effects. Real samples are spiked with increasing amounts of analyte and analyzed before and after each spiking procedure. The sample must be homogenized and the added analyte should be of the same species as the analyte to be determined. The recovery can then be calculated according to Equation (4). The recovery function (Figure 9) can be determined by regression of the added versus found values. Ideally, the recovery function has a slope of 1 and an intercept at the origin. If the deviation from the ideal relationship is due to influences of the sample matrix significantly (relationship $x^* = bx$ in Figure 9) then either matrix-matched calibration samples or the standard addition method have to used.

4.2 Performance Characteristics

Table 11[61] shows which performance characteristics have to be assessed for different purposes. For qualitative parameters, e.g. substance identification, it has to be assured that other substances do not interfere. Also, the lower limiting concentration for positive identification has to be assessed. Quantitative methods are usually based on some kind of calibration curve. Relevant parameters are range, linearity, precision and accuracy of the method. Both qualitative and quantitative determinations are affected by small variations of the experimental conditions to a certain extent, which determine the robustness/ruggedness of the method.

The scope of the validation is determined by the range and the types of matrices covered by the procedure. The determination of a physico-chemical parameter, e.g. temperature or mass, is of a much broader scope than an analytical method combining several steps. If an analytical problem is outside the scope of the validation, e.g. because of a different sample matrix, then the method has to be revalidated in order to show that the reliability is not affected.

Table 11 Performance characteristics relevant for different purposes of the analytical method

Parameter	Qualitative analysis	Quantitative analysis	Trace analysis	Physico-chemical parameters
Accuracy		×	×	×
Precision		×	×	×
Working range/linearity		×	×	×
Selectivity	×	×	×	
Limit of detection	×		×	
Limit of quantitation			×	
Ruggedness	×	×	×	×

4.2.1 Precision

Precision of a method is determined by replicate analyses of a homogeneous material. Usually 10 replicates are measured. The replicates have to be independent to assess realistically the precision of the method, which means that the samples have been independently prepared. Precision is generally expressed as absolute or relative standard deviation of the measurements. If the calibration indicates nonhomogeneity of variances, then several replicates (about 3–5) have to be measured for different concentrations. The quantitative assessment of precision depends on the measuring conditions: the lowest variability is expected if the replicates are prepared and analyzed within a short period of time by the same analyst using the same equipment. The repeatability standard deviation, s_r, is an important parameter for SPC. Higher variability is observed if subsamples of the same homogeneous sample are analyzed by different analysts in different laboratories using different equipment and reagents. This so-called reproducibility standard deviation, s_R, is a realistic assessment of the long-term precision of the method. The term intermediate precision[57] describes the within-laboratory precision achieved by different analysts on different days using different equipment and reagents, and usually lies between repeatability and reproducibility. For qualitative measurements, precision can be interpreted as the percentage of false (positive or negative) identifications.

4.2.2 Accuracy

Accuracy is defined as the closeness of agreement between the means of measurement results and the accepted reference value for the sample (ISO 3534-1).[62] (Note: for a set of values, ISO defines accuracy as a qualitative concept combining bias and precision. Trueness relates to closeness of agreement of the average result of a large number of determinations.) For many analytical methods accuracy is strongly affected by the matrix composition and interferences. Accuracy can be determined (1) by analyzing samples with defined and known analyte concentration with identical or very similar matrix composition, as a matter of preference using CRMs, (2) by using a reference method of known accuracy and (3) by recovery experiments using samples or blank matrices spiked with the analyte. Because of the lack of suitable RMs and methods, the third method is often performed. In many cases recovery studies are problematic since the species of the added analyte may differ from that in the sample. This is especially true for solid samples such as food, geological samples or material where a realistic assessment of accuracy requires RMs or methods. Accuracy cannot be deduced from interlaboratory comparisons.

4.2.3 Working Range

A reliable application of quantitative methods is possible only in a certain range of analyte concentrations between the limit of quantitation and an upper limit which depends on the measuring principle and technical aspects. This working range relates to the concentration in the measuring samples following all sample preparation steps and not to the primary samples. In many cases a range of linear response exists where a proportional relationship between signal and e.g. analyte concentration can be observed. Often a linear response can be achieved by mathematical transformation of the signals, e.g. by Beer's law. Linearity can be assessed by visual inspection of the residual concentrations $c - \hat{c}$ between the measured concentration and the concentration calculated using the regression model. A prerequisite is that the concentrations, usually ≥ 5 independently prepared samples and a blank, cover the whole linear range. Statistical tests comparing the residuals between different regression models may also be used, while regression statistics such as the correlation coefficient or residual variance can be misleading.[63] Outliers can be identified by replicate analysis and removed by using appropriate statistical tests. Nonlinear relationships between signal and analyte concentration can also be evaluated, but depending on the function, more calibration samples are needed since more parameters have to be estimated.

4.2.4 Uncertainty

The combined effects of bias, long-term precision and uncertainty due to calibration can be quantified using the measurement uncertainty concept (section 2.2).

4.2.5 Limit of Detection and Quantification

The lower limit of the working range is of special importance in trace analysis. Since the signal-to-noise ratio of the analytical signal decreases with lower concentrations, a concentration range exists where the probability of distinguishing a trace concentration from the blank sample lies below a certain significance level P. Several definitions exist for determining this 'limit of detection' or 'detection limit'[64] and therefore the method used has to be stated. All approaches are based on assessing the repeatability of measurement near zero concentration either using blank samples or information acquired from the regression function. The most frequently used approach is based on the standard deviation of the sample blank s_B usually determined from 10 replicate measurements and expressed as Equation (9):

$$x_{DL} = \frac{k s_B}{b} \tag{9}$$

where b is the slope of the calibration function and k is a factor; usually $k = 3$ or 3.3. In chromatography the baseline noise is used to estimate the limit of detection. The limit of detection strongly depends on the sample matrix. Detection limits provided by instrument manufacturers are generally determined using simple matrices, e.g. water, and should therefore be interpreted with caution. More sophisticated definitions use the residual variance of the calibration curve.[65] Only values near the detection limit should be taken into account, since nonhomogeneity of the variances and even a low degree of nonlinearity can lead to a biased estimation. The lowest concentration of analyte in a specific matrix for which a certain RSD can be specified, is termed ''limit of quantification'', ''quantitation limit'' or ''limit of determination''. A value of $k = 10$ is often used as a factor which relates to an RSD $\approx 10\%$.

4.2.6 Selectivity and Specificity

Generally selectivity and specificity both address the fact that interfering substances lead to increased or decreased concentrations determined for the analyte. Both terms are sometimes used with different meanings. Specificity has been defined as ''the ability of a method to measure only what it is intended to measure''[53] and as ''the ability to assess unequivocally the analyte in the presence of components which may be expected to be present'',[57] while selectivity relates to ''the extent to which other substances interfere with the determination of a substance according to a given procedure''.[66] Sometimes selectivity is only used in connection with multicomponent analysis, e.g. in chromatography, while specificity relates to single-component methods, such as AAS.[24] Selectivity and specificity have to be investigated for both qualitative and quantitative methods. Except in special cases such as atomic spectroscopy, where all interferences may be known, it is almost impossible to state that a method is not affected by interferences or to identify all possible interferents. Since also combinations of noninterfering substances may lead to inference, not all potential interfering substances can be tested. In practice, interferents which are expected to occur are deliberately introduced in the sample, and favorable statistical experimental design is employed. Selectivity can be enhanced by isolation, preconcentration, or clean-up of the analyte in the pretreatment step, by optimizing instrumental parameters or by using chemometric methods such as peak separation or multivariate calibration in the data evaluation step.

4.2.7 Ruggedness

The reliability of an analytical method depends on various environmental parameters and operating conditions which change during normal usage. If the effect of these changes on the accuracy and precision of the method is small then the method behaves ruggedly. The term robustness is used synonymously.[57] The limits of the parameters where the reliability is not affected have to be assessed and documented. Ruggedness can be determined by two methods:

- by an interlaboratory study (method comparison study), where the influence of different instruments, analysts and reagents can be studied.
- by an intralaboratory study, where small deliberate changes have to be made in a systematic way on all parameters which significantly influence the analytical result.

Preferably, an experimental design based on factorial experiments should be used for varying the influence factors.[53] For quantitative methods the calibration has to be included in the study.

4.3 Reference Materials

An RM is a substance of which one or more properties are sufficiently established to be used for validation, calibration and method comparison. RMs are supplied with a certificate of analysis but must be distinguished from CRMs, ''one or more of whose property values are certified by a technically valid procedure, accompanied by or traceable to a certificate or other documentation which is issued by a certifying body'' (ISO Guide 30).[67] The term SRM is also used, e.g. by NIST (Table 12). The certificate usually covers aspects such as form of the material, e.g. powder, intended use of the material, e.g. analytical method, method used for analysis and analyte concentration or content together with an uncertainty statement.

Table 12 Categories found in the NIST SRM/RM catalog

NIST category	SRM/RM (example)
Ferrous metals	Valve steel
Nonferrous metals	Battery lead
Microanalysis	Gold–copper wire for microprobe analysis
High-purity materials	High-purity platinum wire
Inorganics	Trace elements in water
Food and agriculture	Nonfat milk powder
Health and industrial hygiene	Human serum
Primary gas mixtures	CO_2 in nitrogen
Engine wear materials	Wear metals in lubricating oil
Fossil fuels	Moisture in mineral oil
Organics	Polycyclic aromatic hydrocarbons test mixture
Geological materials and ores	Zinc concentrate
Ceramics and glasses	Soda-lime float glass
Cement	Portland cement clinker

The properties of RMs and CRMs usually have been determined through interlaboratory testing or definite methods. CRMs have to be certified by following the procedures described in ISO Guide 35.[68] In most cases working standards are derived from the CRM for calibration and SPC. Therefore, for the laboratory a CRM forms a useable end-point of the traceability chain. The general properties of RMs should be the same as for control samples, e.g. stability and homogeneity. CRMs and RMs used in the analytical laboratory relate to chemical composition and trace elements; examples are shown in Table 12.

CRMs and RMs are available from, among others, NIST (US Department of Commerce, NIST, Gaithersburg, MD, USA), BCR (Bureau Communautaire de Référence, Commission of the European Community, Brussels, Belgium), BAM (Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany) and LGC (Laboratory of the Government Chemist, Teddington, Middlesex, UK). Data bases on RMs/CRMs are the ISO Directory of CRMs and COMAR (International Data Bank on RMs, Central Secretariat of COMAR, Laboratoire National d'Essais, Paris).

5 EXTERNAL QUALITY ASSURANCE

5.1 Proficiency Testing

Various types of inter-laboratory comparisons exist, which are performed for investigation the properties of methods, to characterize a material or to compare laboratories. When laboratory performance is investigated by a third party, the comparison is termed a proficiency test. The purpose is then to demonstrate the capability of the laboratory to perform a certain type of analysis using an established method with known properties. Special attention has to be given to the homogeneity and stability of the sample which is distributed.

Procedures for proficiency testing are described in ISO Guide 43, 'Proficiency Testing by Laboratory Intercomparisons'. Criteria for evaluating the laboratory performance are accuracy and precision of the results. The most commonly chosen approach[69] is to calculate the standardized z-values according to Equation (10):

$$z_i = \frac{|x_i - \mu|}{s} \tag{10}$$

where x_i is the analytical result from the laboratory i, μ is the true or accepted reference value and s is the standard deviation calculated from the results from all laboratories. Prior to computing the standard deviation, outliers must be identified and eliminated. Another prerequisite is approximate normal distribution of the data.

The z-values are interpreted as follows:

$z \leq 2$ acceptable
$2 < z < 3$ questionable
$z \geq 3$ not acceptable

The interpretation of the z-values is based on the confidence regions of approximately 95% and 99%. Proficiency tests should be used for learning and correction rather than judgment to increase acceptance.[70]

5.2 Accreditation

ISO/IEC Guide 2 defines accreditation as a "procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks".[71] The laboratory can thus demonstrate its competence in certain fields by conforming to international or national standards. Typically following steps can be distinguished in an accreditation procedure: application, assessment, accreditation and surveillance. Laboratories applying for accreditation have to fulfill certain requirements according to national or international standards with the scope of accreditation ranging from a specific test method to an entire field of testing. The guides and standards most widely used for laboratory accreditation are ISO Guide 25 (soon ISO 17025) and in Europe EN 45001. In future the ISO 17025 will further improve internationally standardization of laboratory accreditation standards. Various programs for accreditation exist in different countries; an overview is given in an ILAC publication.[72] In Europe where an important aspect of laboratory accreditation is the reduction of trade barriers most accreditation bodies are governmental organizations whereas in the USA various programs are run by different levels of government and the private sector. These are in some cases state and industry specific and often use their own terminology, e.g. designated, inspected, accepted or nationally recognized laboratory.

Breitenberg[73] gives six reasons for a laboratory to participate in an accreditation program:

- outside check of its internal quality program;
- proof of competence to higher level management within the organization;
- competitive advantage over other unaccredited laboratories;
- a means of protection in liability proceedings;
- a means of established credibility within the public;
- contracts or procurement requirements.

Accreditation does not imply that the results of the laboratory are generally better than those from laboratories that have not taken part in an accreditation program, but it demonstrates the laboratory's capability to perform certain tasks.

ACKNOWLEDGMENTS

The author thanks Klaus Danzer, Gabi Thiel and Kay-Uwe Jagemann for inspiring discussions and constructive comments.

ABBREVIATIONS AND ACRONYMS

AAS	Atomic Absorption Spectrometry
AOAC	Association of Official Analytical Chemists
AOX	Adsorbable Organically Bound Halogens
ARL	Average Run Length
BAM	Bundesanstalt für Materialforschung und -prüfung
BCR	Bureau Communautaire de Référence
CRM	Certified Reference Material
DIN	Deutsche Institut für Normung
EN	European Norm
EPA	Environmental Protection Agency
EU	European Union
EWMA	Exponentially Weighted Moving Average
FDA	Food and Drug Administration
GALP	Good Automated Laboratory Practice
GC	Gas Chromatography
GLP	Good Laboratory Practice
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography
ICPMS	Inductively Coupled Plasma Mass Spectrometry
ICPOES	Inductively Coupled Plasma Optical Emission Spectrometry
IDMS	Isotope Dilution Mass Spectrometry
IEC	International Electrotechnical Commission
ILAC	International Laboratory Accreditation Cooperation
ISO	International Organization for Standardization
LCL	Lower Control Limit
LGC	Laboratory of the Government Chemist
LIMS	Laboratory Information Management System
MA	Moving Average
NAFTA	North American Free Trade Agreement
NIST	National Institute of Standards and Technology
OECD	Organization for Economic Cooperation and Development
PE	Polyethylene
PP	Polypropylene
PTFE	Polytetrafluoroethylene
QA	Quality Assurance
QAU	Quality Assurance Unit
QC	Quality Control
RM	Reference Material
RSD	Relative Standard Deviation
SI	Système International
SOP	Standard Operating Procedure
SPC	Statistical Process Control
SQL	Structured Query Language
SRM	Standard Reference Material
TOC	Total Organic Carbon
UCL	Upper Control Limit
UV	Ultraviolet
UV/VIS	Ultraviolet/Visible
VIM	International Vocabulary of Basic and General Terms in Metrology

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis

Chemometrics **(Volume 11)**
Chemometrics

General Articles **(Volume 15)**
Traceability in Analytical Chemistry

REFERENCES

1. 'Quality Management and Quality Assurance; Vocabulary', ISO 8402, International Organization for Standardization, Geneva, 1994.
2. W.E. Deming, *Out of the Crisis*, Massachusetts Institute of Technology, Cambridge, MA, 1982.
3. 'The Role of Testing and Laboratory Accreditation in International Trade', International Laboratory Accreditation Cooperation ILAC-I3, 1996.
4. E. Prichard, H. Albus, B. Neidhart, W. Wegscheider, 'Glossary of Analytical Terms (X)', *Accred. Qual. Assur.*, **3**, 171–173 (1998).
5. W. Funk, V. Dammann, G. Donnevert, *Quality Assurance in Analytical Chemistry*, VCH, Weinheim, 1995.
6. H. Günzler (ed.), *Accreditation and Quality Assurance in Analytical Chemistry*, Springer, Heidelberg, 1994.
7. G. Kateman, L. Buydens, *Quality Control in Analytical Chemistry*, Wiley, Chichester, 1991.
8. F.E. Prichard, *Quality in the Analytical Chemistry Laboratory (Analytical Chemistry by Open Learning)*, Wiley, Chichester, 1995.

9. *International Vocabulary of Basic and General Standard Terms in Metrology*, (VIM), 2nd edition, International Organization for Standardization, Geneva, 1993.
10. J. Fleming, B. Neidhart, C. Tausch, W. Wegscheider, 'Glossary of Analytical Terms', *Accred. Qual. Assur.*, **1**, 41–43 (1996).
11. 'SI Units and Recommendations for the Use of Their Multiples and of Certain Other Units', ISO 1000, International Organization for Standardization, Geneva, 1992.
12. J.R. Moody, M.S. Epstein, 'Definitive Measurement Methods', *Spectrochim. Acta, Part B*, **46**, 1571–1575 (1991).
13. W. Richter, 'Primary Methods of Measurement in Chemical Analysis', *Accred. Qual. Assur.*, **2**, 354–359 (1997).
14. P. Spitzer, R. Eberhardt, I. Schmid, U. Sudmeier, 'Improved Traceability of pH Measurements', *Fresenius' J. Anal. Chem.*, **356**, 178–181 (1996).
15. W. Hässelbarth, 'Die Forderung der Rückführbarkeit/ Rückverfolgbarkeit von Messergebnissen und ihre Umsetzung in der Chemischen Analytik', *DIN-Mitteil.*, **72**, 9–12 (1993).
16. F. Adams, 'Traceability and Analytical Chemistry', *Accred. Qual. Assur.*, **3**, 308–316 (1998).
17. W. Hässelbarth, 'Measurement Uncertainty Procedures Revisited: Direct Determination of Uncertainty and Bias Handling', *Accred. Qual. Assur.*, **3**, 418–422 (1998).
18. P. Good, *Resampling Methods: a Practical Guide to Data Analysis*, Birkhäuser Verlag, Basle, 1998.
19. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, *Guide to the Expression of Uncertainty in Measurement*, International Organization for Standardization, Geneva, 1993.
20. *Quantifying Uncertainty in Analytical Measurement*, EURACHEM Guide, LEC, Teddington, 1995.
21. E. Grant, R. Leavenworth, *Statistical Quality Control*, McGraw-Hill, New York, 1980.
22. D.C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley, New York, 1985.
23. W.A. Shewhart, *The Economic Control of Quality of Manufactured Products*, Van Nostrand, New York, 1931.
24. K. Danzer, 'The Significance of Statistics in Quality Assurance', in *Accreditation and Quality Assurance in Analytical Chemistry*, ed. H. Günzler, Springer, Heidelberg, 105, 1994.
25. American Society for Testing and Materials (ASTM), *Manual on Presentation of Data and Control Chart Analysis*, Special Technical Publication 15D, 6th edition, ASTM, Philadelphia, PA, 1990.
26. T.P. Ryan, *Statistical Methods for Quality Improvement*, John Wiley, New York, 1989.
27. 'General Requirements for the Competence of Calibration and Testing Laboratories', ISO/IEC Guide 25, International Organization for Standardization, Geneva, 1990.
28. 'General Criteria for the Operation of Testing Laboratories', CEN/CENELEC EN45001.
29. 'Quality Management and Quality Assurance Standards', ISO 9000, International Organization for Standardization, Geneva.
30. 'OECD Principles of Good Laboratory Practice', OECD Document ENV/MC/CHEM (98) 17, OECD Environment Directorate, Paris, 1998.
31. Environmental Protection Agency, FIFRA GLPs, 40 CFR 160, US Government Printing Office, 1999.
32. Environmental Protection Agency, TSCA GLPs, 40 CFR 792, US Government Printing Office, 1999.
33. Food and Drug Administration, 21 CFR Part 58, 'Nonclinical Laboratory Studies – Good Laboratory Practice Regulations', *Fed. Regist.*, **41**(225), 1976.
34. H. Kohl, 'The New ISO 17025 – Basic Idea', *Accred. Qual. Assur.*, **3**, 422–425 (1998).
35. 'Internal Quality Audits and Management Review for Laboratories', EA-4/04 European Cooperation for Accreditation, 1996.
36. B. Kratochvil, J.K. Taylor, 'Sampling for Chemical Analysis', *Anal. Chem.*, **53**, 925A–938A (1981).
37. G. Kateman, L. Buydens, *Quality Control in Analytical Chemistry*, 2nd edition, John Wiley, New York, 1993.
38. P. Quevauviller (ed.), *Quality Assurance in Environmental Monitoring: Sampling and Sample Pretreatment*, Wiley-VCH, Weinheim, 1995.
39. M. Thompson, 'Uncertainty of Sampling in Chemical Analysis', *Accred. Qual. Assur.*, **3**, 117–121 (1998).
40. A. Argyraki, M.H. Ramsey, M. Thompson, 'Estimation of Sampling Bias Between Different Sampling Protocols on Contaminated Land', *Analyst*, **120**, 2799 (1995).
41. T.H. Meltzer, *High Purity Water Preparation*, Tall Oaks Publications, Littleton, CO, 1993.
42. L. Huber, 'Qualification and Validation of Software and Computer Systems in Laboratories. Part 3. Installation and Operational Qualification', *Accred. Qual. Assur.*, **3**, 140–144 (1998).
43. *United States Pharmacopeia*, 23rd Revision, United States Pharmacopeial Convention, Rockville, MD, 1994.
44. OECD Series on the Principles of GLP and Compliance Monitoring, No. 10, GLP Consensus Document, 'The Application of the Principles of GLP to Computerised Systems', Environment Monograph No. 116, Environment Directorate, Organization for Economic Cooperation and Development, Paris, 1995.
45. US Environmental Protection Agency, 'Good Automated Principles and Guidance to Regulations for Ensuring Data Integrity in Automated Laboratory Operations with Implementation Guidance', No. 2185, Office of Information, Research Triangle Park, NC, 1995.
46. L. Huber, *Validation of Computerized Analytical Systems*, Interpharm Press, Englewood, CO, 1995.
47. R.F. Boisvert (ed.), *The Quality of Numerical Software: Assessment and Enhancement*, Chapman & Hall, London, 1997.

48. M.G. Cox, P.M. Harris, 'Design and Use of Reference Data Sets for Testing Scientific Software', *Anal. Chim. Acta*, **380**, 339–351 (1999).
49. 'Information Technology – Software Products Evaluation – Quality Characteristics and Guidelines for Use', ISO/IEC 9126, International Organization for Standardization, Geneva, 1991.
50. NIST Statistical Reference Data Sets, http://www.nist.gov/div898/strd/index.html.
51. 'Gesetz zum Schutz vor gefährlichen Stoffen', Verwaltungsvorschrift Bundesgesetzblatt 1, 29/10/90 (Chem. VWV – GLP), Bundesanzeiger No. 204, 31/10/90, Bonn, 1990.
52. S.P. Maj, *The Use of Computers for Laboratory Automation*, CRC Press, Boca Raton, FL, 1993.
53. *AOAC Peer Verified Methods Program – Manual on Policies and Procedures*, AOAC International, Arlington, VA, 1993.
54. J.M. Green, 'A Practical Guide to Analytical Method Validation', *Anal. Chem.*, **68**, 305A–309A (1996).
55. W. Wegscheider, 'Validation of Analytical Methods', in *Accreditation and Quality Assurance in Analytical Chemistry*, ed. H. Günzler, Springer, Heidelberg, 135, 1994.
56. 'EURACHEM Fitness for Purpose of Analytical Methods: A Laboratory Guide to Validation and Related Topics', LEC, Teddington, 1998.
57. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 'Q2A Validation of Analytical Methods – Definition and Terminology', ICH, Geneva, 1994.
58. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 'Q2B Validation of Analytical Methods – Methodology', ICH, Geneva, 1996.
59. W. Horwitz, 'Protocol for the Design, Conduct and Interpretation of Method-performance Studies', *Pure Appl. Chem.*, **67**(2), 331–343 (1995).
60. K. Danzer, M. Wagner, Ch. Fischbacher, 'Calibration by Orthogonal and Common Least Squares – Theoretical and Practical Aspects', *Fresenius' J. Anal. Chem.*, **352**, 407–411 (1995).
61. 'The Validation of Analytical Procedures Used in the Examination of Pharmaceutical Materials', WHO/PHARM/89.541, World Health Organization, Geneva, 1989.
62. 'Statistics – Vocabulary and Symbols – Part 1: Probability and General Statistical Terms', ISO 3534-1, International Organization for Standardization, Geneva, 1993.
63. F.J. Anscombe, 'Graphs in Statistical Analysis', *Am. Stat.*, **27**, 17–21 (1973).
64. 'Nomenclature in Evaluation of Analytical Methods, Including Detection and Quantification Capabilities (IUPAC Recommendations)', *Pure Appl. Chem.*, **67**, 1699–1723 (1995).
65. 'DIN 38402-51 Deutsche Einheitsverfahren zur Wasser-, Abwasser und Schlammuntersuchung; Kalibrierung von Analysenverfahren, Auswertung von Analysenenergebnissen und Lineare Kalibrierfunktionen für die Bestimmung von Verfahrenskenngröpen', Beuth Verlag, Berlin, 1994.
66. *Compendium of Chemical Technology (IPAC Recommendations)*, Blackwell Scientific, Oxford, 1987.
67. 'Terms and Definitions Used in Conjunction with Reference Materials', ISO Guide 30, International Organization for Standardization, Geneva, 1992.
68. 'Certification of Reference Materials – General and Statistical Principles, ISO Guide 35, International Organization for Standardization, Geneva, 1989.
69. 'Proficiency Testing of Chemical Analytical Laboratories', ISO/REMCO N280, International Organization for Standardization, Geneva, 1993.
70. H.O.F. Andersson, 'The Rational Use of Proficiency Tests and Intercomparisons', *Accred. Qual. Assur.*, **3**, 224–226 (1998).
71. 'General Terms and Their Definitions Concerning Standardization and Related Activities', ISO/IEC Guide 2, International Organization for Standardization, Geneva, 1996.
72. International Laboratory Accreditation Cooperation, 'Testing, Quality Assurance, Certification and Accredition', ILAC-I2:1994, 1996.
73. M.A. Breitenberg, 'The ABC's of the U.S. Conformity Assessment System', NIST IR 6014, Office of Standards Services, National Institute of Standards and Technology, Gaithersburg, MD, 1997.

Quantitative Spectroscopic Calibration

Howard Mark
Mark Electronics, Suffern, USA

1 Introduction **1**
1.1 Why Calibration? 2
1.2 Generic Procedures 3
2 Individual-wavelength Methods (also called Multiple Linear Regression, P-matrix, Inverse Least Squares, and Ordinary Least Squares) and Wavelength Search Techniques **7**
2.1 Classical Methods 8
2.2 Modern Methods 10
3 Full-spectrum Methods **12**
3.1 Generic Considerations 13
3.2 K-matrix 15
3.3 Principal Components 16
3.4 Partial Least Squares 17
3.5 Fourier Transform 17
3.6 Neural Nets 17
Abbreviations and Acronyms **18**
Related Articles **19**
References **19**

Spectroscopic calibration is the process of using mathematical calculations to relate the measured spectroscopic properties of a set of samples to the chemical composition (and sometimes other properties) of those samples. Any of several algorithms are employed to perform these calculations. Regardless of the algorithms used, however, successful execution of this process depends on properly performing several tasks before and after the main calibration algorithm is applied: selecting a suitable set of calibration samples to ''train'' the system, using the correct algorithm to match the characteristics of the measured data, and properly evaluating and validating the calibration model developed. In this article, these various aspects of calibrating spectroscopic instruments to perform quantitative analysis are discussed. The article ends with an overview of the various algorithms that are available to perform the calibration process.

1 INTRODUCTION

''Calibration'' means different things to different people. The definition in one widely accepted dictionary is ''to determine, rectify, or mark the graduations of''.[1] This definition is fairly specific: it is intended to be applied to devices such as thermometers. A more general, and perhaps more generic, meaning might be ''to make readings (presumably from an instrument or some other measurement device) conform to those of a standard''. This definition, while more generally applicable, has problems of its own (which readings?, what standard?, etc.) but, even if those problems were solved, it would not even begin to address the meaning of ''calibration'' as used in modern spectroscopic practice. Some time ago, I started to discuss the question of calibrating a spectrometer with a well-known, highly regarded spectroscopist. It took some time to realize that we were talking at cross-purposes, since his definition of calibration was essentially ''to ensure the accuracy of the wavelength scale of the spectrometer'', which had nothing to do with the points I was trying to make. Other meanings, which are certainly reasonable and justifiable in appropriate spectroscopic contexts, are: to ensure accuracy of the transmittance scale, or of radiant energy measurement, but these meanings also have nothing to do with ''calibration'' as used in the current terminology.

What, then, do we mean by the term ''spectroscopic calibration''? The definition given in one recent book about chemometrics is ''the process of constructing a model that is used to predict characteristics or properties of unknown samples'',[2] and this is pretty close. In this discussion, at least, the meaning ascribed to the term ''spectroscopic calibration'' is the use of mathematical algorithms to relate spectroscopically measured quantities to the chemical composition (and sometimes other properties) of samples; Beebe's definition can be expanded slightly since the models may or may not ever be used to make actual predictions. The most common use of the term is to express the percentage of various chemical constituents, although sometimes the interest is in the cumulative effect of the components (e.g. the octane rating of gasoline), or a physical property (e.g. viscosity), or an operational parameter (e.g. heat content of fuel). It is often desired, and usually possible, to calibrate a set of samples for more than one constituent or property. This is advantageous, because when performing routine analysis on the actual ''real'' samples of interest, it then becomes possible to analyze those samples for all of the constituents and properties for which

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

calibration models exist, from a single measurement of the spectrum.

Modern spectroscopic practice has progressed far beyond the simple use of Beer's law to relate the absorbance of an analyte at a characteristic wavelength to the concentration of that analyte. Although Beer's law has not failed us, and indeed is still the basis of the methods we use, as described in at least one book,[3] a simple measurement of absorbance is not sufficient to accommodate the variety of conditions encountered when using spectroscopic measurements to perform ''real-world'' analysis.

To allow accurate analysis in the face of electronic and other forms of noise, inhomogeneous samples, and a plethora of other phenomena that can affect spectroscopic readings, a variety of sophisticated mathematical techniques has been developed that attempt (with varying degrees of success) to extract the analytical information from the spectroscopic data. It is the application of these techniques that define the modern meaning of the verb ''to calibrate''.

Note, however, that the term ''spectroscopic calibration'' also has a related meaning as a noun: it is the collective set of mathematical parameters resulting from the application of the calibration process to a set of data. This is also sometimes called the ''mathematical model'', or the ''calibration model'', but the simpler term ''calibration'' is often used on its own.

1.1 Why Calibration?

As indicated above, Beer's law still works and can be applied in those measurement situations that are simple enough for it to continue to apply. If the samples to be measured are clear solutions, if the solvent has no absorbance at the wavelength of analytical importance, if the analyte of interest has absorbance bands, if those absorbance bands are not interfered with by bands of the solvent or of other components of the sample, if the analyte does not interact (chemically) with either the solvent, other components of the sample, or itself as the concentration changes, if the samples are stable with time, if the spectrometer is sufficiently noise free and linear in its response to the radiant energy, then we can perform the following experiment:

1. Measure the radiant energy through the system at an appropriate wavelength in the absence of a sample (call this I_o).
2. Measure the radiant energy through the system at the same wavelength with the sample present in known thickness (call this I).
3. Calculate T (transmittance) as I/I_o.
4. Calculate A (absorbance) as $\log_{10}(1/T)$.
5. Calculate ε (the absorptivity) from the formula (Equation 1):

$$A = \varepsilon bc \tag{1}$$

   where b is the thickness of the sample and c is the concentration of the analyte.
6. Measure the absorbance of other samples and calculate the concentration from the measured absorbance and the (now known) values of ε and b.

Steps 1–6 essentially constitute the classical, standard, Beer's law approach to analysis; indeed, this in itself is a limited, specific type of calibration. But it involves a lot of ''ifs''.

In the ''real world'', where inhomogeneous samples contain multiple components having overlapping absorbance bands, the analytes are subject to interactions with themselves, each other, and the matrix (or solvent) as well as the other components, are affected by extraneous physical phenomena, such as particle size variation, and extraneous conditions, such as temperature changes, and are measured by less than ideal instruments that are prone to noise, drift, non-linearity, and other defects, simply choosing ''a'' wavelength at which to perform the analysis is not an option: such a wavelength will not exist, probably not even in principle. Add to that brew the fact that very often the ''analyte'' of interest is not a well-defined chemical species but rather a physical property or operational parameter of some sort (examples of which were described above, where the chemistry and/or spectroscopy is ill defined or unknown), then the question of characterizing the situation by ''a'' wavelength becomes ludicrous. Consider further the fact that spectroscopic measurements are often taken via reflectance rather than transmittance, so that the measurements are affected by the physical nature of the sample; thus, the choice of a wavelength becomes even more difficult (what is the absorbance band for ''particle size effect'', for example?).

There are generically two ways in which all these difficulties can be addressed. The first is to apply known physical and chemical theory to characterize each individual phenomenon that affects the measurement and use that theory to make appropriate corrections. Historically, this is how science has progressed. In principle, this would result in an ''absolute'' measurement that would be independent of particular conditions existing in each laboratory and would consequently be highly advantageous. There are two difficulties with this approach: first, not all of these obstacles to accurate analysis have rigorous ab initio theories available to adequately describe them and, second, each and every routine analytical application of spectroscopy would then become a major research project (with the attendant time, money, and resource

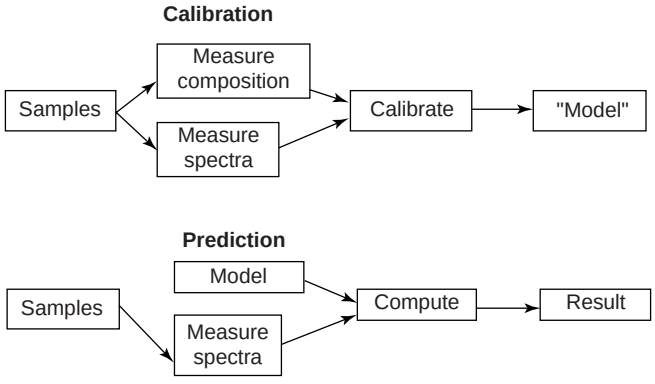

Figure 1 Block diagrams of generic procedures for calibration and prediction.

requirements), preventing spectroscopic analysis from becoming a routine tool, the state of affairs that we currently enjoy.

Therefore, in practice, the second way of addressing the difficulties is used: the situation is treated completely empirically, and following the dictum that ''math is cheaper than physics'', the relationship between the composition of a suitable set of samples and their optical properties is derived from the spectroscopic measurement of the samples and computation of the relationships using the sophisticated mathematical procedures mentioned above. Figure 1 presents a block diagram of how this is done.

These procedures have been developed to find empirical ways of accommodating variations in the data that are not describable by fundamental a priori knowledge, through the ability to use information at multiple wavelengths to compute corrections for the physical effects that would otherwise cause errors. Random errors are accommodated by using a large number of readings. This is accomplished by using data measured at multiple wavelengths from many samples, to allow the errors to ''average out'' so that the net result is close to the ''truth''. The procedures used are collectively described by the term ''chemometrics'' (literally, the measurement of chemistry).

1.2 Generic Procedures

Some considerations are common to all methods of spectroscopic calibration, regardless of the particular mathematical algorithm used. Use of standard good laboratory practice (GLP), such as proper sampling techniques, ensures that the sample measured is representative of the material whose composition is to be determined. Equally important for calibration purposes, it assures that the material measured optically by the instrument is substantially the same as that measured by the reference laboratory.

Good chemical laboratory practice prevents the sample from changing between the time it is taken and the time it is measured. In general, the application of common sense and standard chemical procedures will prevent many sources of error from affecting the measurement process.

In addition, there are some procedures that are unique to the application of the chemometric techniques that are used here. These procedures devolve from the nature of the calibration process, the need to ''relate spectroscopically measured quantities to the chemical composition of samples'', as defined above. Thus, to accomplish this, we need to do two things:

1. take the spectroscopic measurements; and
2. determine the chemical composition of the samples,

so that we can find the relationships between them. Both of these need to be done properly.

Thus, the concept of ''training'' the spectroscopic system (usually considered to be the instrument), which comprises the procedures used to find the necessary relationships, arose. To accomplish this, it is necessary to accumulate an appropriate set of samples, measure them using the spectrometer to be calibrated, and have the composition of those same samples measured by some other, independent method (often wet chemical analysis), in order to create the data that the algorithms will be applied to.

1.2.1 Sample Collection

As hinted at above, a single sample is useless for calibration purposes. The methodology of calibration requires that a set of samples be used. Furthermore, although samples are often collected at random out of necessity, not all sets of samples are equally suitable for calibration. Here we discuss criteria for collecting sample sets with optimum characteristics. Although it is usually not possible to achieve and optimize all these desirable characteristics simultaneously, attention to them, and attempts to come as close as reasonably possible, will result in ultimately achieving calibration results that are as accurate and robust as possible.

So, what are the desirable characteristics of a set of samples to be used for calibration? The first critical characteristic is that the samples used for creating the calibration model must be essentially the same as those to be analyzed in the future, using that model. This usually means that actual, real, samples of the type to be analyzed must be collected. Synthetic laboratory samples have been used to calibrate instruments but this is rare as such samples are usually not sufficiently representative of the variations of actual samples from a real chemical process for a calibration model based on them to give accurate results when the actual samples that the calibration is

intended for are measured. This is particularly true when the samples are a natural product, or an intermediate or final product of a chemical process stream.

A second key characteristic is that the samples must cover the range of values of the constituents contained in the samples. It is found that extrapolation beyond the range of constituent concentrations contained in these training samples generally results in very poor accuracy in the extended parts of the range. Furthermore, to include the range of values of the constituents for which calibrations are to be generated is not sufficient. It is also necessary that the calibration sample set includes the range of variation of those constituents that are not measured, but which may exist in the samples and vary from sample to sample. While this is a tall order, the reason for it is clear: it is necessary for the variations of effects of those other, interfering, materials to be present in the sample set so that the necessary corrections can be calculated and included in the calibration model.

It is also most desirable that the various constituents vary independently of each other. In other words, if a set of samples contains two constituents, A and B, say, then constituent B should vary throughout its concentration range in those samples containing low values of constituent A, intermediate values of constituent A, and high values of constituent A. Similarly, constituent A should vary over its range for all values of constituent B.

In those few cases where it is possible to use synthetic laboratory samples, they do have an advantage since they can be made up according to prescribed formulas that can assure the independence of the constituents. However, more commonly, when this method cannot be used, it is necessary to collect ''sufficiently many'' samples to maximize the chance that the range and independence conditions are met. ''Sufficiently many'' is obviously not a very concrete description, but the actual number will depend on how much variation can be introduced into the samples while they are being collected.

Some rules of thumb have been devised to help with this. One such rule is that the minimum number of samples should be 10 times the number of constituents, plus 10 more (''for the pot''). Another rule is that, since these calibrations are often used for analysis of samples from some sort of a process, the samples for calibration should be taken during start-up, shutdown, or any other time the process is in a condition other than ''normal operation''.

Another good piece of advice is to collect samples over long time periods. Often a more or less temporary calibration is created at the time an instrument is first installed, and then more calibration samples collected even while the instrument, with this temporary calibration, is used for routine analysis. In this case, the instrument can even help select which samples to collect: those would be the ones for which the readings are found to be inaccurate.

1.2.2 Reference Value Considerations

Few aspects of the calibration process have caused more confusion than the question of obtaining reference, or training, values for the constituents in the sample set. Part of this is attributable to the historical experience of new users of the instruments that need this type of calibration: they are used to depending on their current laboratory procedure for their analytical answers. Normally, this is the same procedure that will serve to provide the values for the composition of the training samples. Since this was always the only value available, it has always been regarded as being ''correct''. Most chemists know intellectually that every measurement has some error associated with it, but this does not easily translate into an intuitive appreciation for the effects of those errors on an instrument calibration.

With experience, and hopefully from the advice of experienced workers, there comes the realization that the errors need to be quantified, and that there are relatively standard procedures for doing this. If no other rationale can be found, anyone calibrating an instrument should know and appreciate the fact that the instrument cannot be expected to agree with the reference laboratory results any better than those results agree with themselves.

Hence, good practice in this regard indicates that every sample that is to be included in the calibration set (or in the validation set: see below) should have at least two aliquots taken and sent to the laboratory for independent analysis. The samples should be ''blinded'' so that whoever is doing these reference analyses should not know which of the samples being analyzed represent the aliquot pairs. It is preferable, in fact, that, if the procedure is not too time or resource consuming, the different aliquots be sent on different days, or at least be timed so that they will be analyzed on different shifts; the idea here is to try to ''capture'' as much as possible of the true laboratory variability in the readings, so as to obtain as good an idea as possible of the true value of the reference laboratory error.

These paired readings can be used for two purposes. The first is to assess the error of the reference laboratory, so that a comparison can later be made with the instrument's agreement to the reference laboratory values. Normally, a good calibration model will provide results that agree with the reference laboratory results about as well as the reference laboratory results agree with each other. If the agreement between the instrument readings using a given model and the reference laboratory results is too good, then that is evidence that the model has been ''overfit'' to the calibration data, a calibration problem that will be discussed further below.

The second use of the paired values is to average them together, and use the average of each pair of readings as the reference value for that sample. The process of averaging reduces the error of the analysis, so that the calibration training values are more accurate than single readings.

1.2.3 Data Pretreatments

Spectral data are often converted, prior to applying the main calibration algorithm, to a form different from the absorbance values they are collected in, for the purpose of calibration. Indeed, it could be argued that they are always converted, since spectral data are invariably collected as measured energy, and the absorbance itself is a calculated value. However, since absorbance is invariably the format stored, used for calibration based on Beer's law and often used for presentation of the spectra, it is considered the "standard" format for calibration work, particularly in the near-infrared (NIR) spectral region, and in others when quantitative analysis is the consideration.

Thus, data transformations almost invariably use the absorbance spectrum as their starting point. The intent of most of the transformations is to remove or reduce variations of the spectra that might cause errors in the calculated results, although some transformations are intended to improve the linearity of the relationship between constituent concentrations and the spectral measurements. There are a number of ways in which the various transformations can be classified; one convenient way is to consider whether they work on individual wavelengths or use the readings from two or more wavelengths to compute the final value corresponding to a given wavelength. Table 1 lists the most common transformations encountered.

Of the pretreatments listed in Table 1, all except the Kubelka–Munk transform are for the purpose of removing extraneous variations superimposed on the spectrum, although the various transformations are designed to remove different types of variations.

Smoothing is the averaging together of several readings in the neighborhood of a nominal wavelength, and is intended to reduce instrumental (or high-frequency) noise. A simple averaging together of the readings at some number of wavelengths on either side of the nominal wavelength with the reading at the nominal wavelength is most effective at reducing this noise, but has the disadvantage of increasing the apparent width of the absorbance bands in the spectrum. A compromise sometimes used is to perform a weighted averaging, whereby those wavelengths closest to the nominal wavelength contribute more to the final value than do those farther away. The classic paper by Savitsky and Golay[4] is often used as the basis for selecting the weighting factors for this transformation (as well as some of the other data transformations used). An alternate method that has been recommended for smoothing spectra is Fourier smoothing (see, for example, Chapter 8 in Burns and Ciurczak[5]), but this method is not widely used.

Table 1 Common data transformations applied to spectral data

Used with individual wavelengths	Used with multiple wavelengths
Normalization	Smoothing
Kubelka–Munk transform	Derivatives ($dA/d\lambda$)
Multiplicative scatter correction	Ratios (of absorbances, derivatives, etc.)

The remaining methods listed in Table 1 are intended to remove low-frequency noise from the spectra. "Low-frequency" in this case means that it affects all wavelengths uniformly, or at least consistently. Although in rare cases pathlength changes in a transmittance measurement would give rise to this type of variation, it is most commonly seen when measuring reflectance spectra from powdered solids. It is a characteristic of powdered solids that the reflectance depends on the details of the size, shape, and orientation of the surface layer of particles, and hence the same sample will give a different spectrum each time it is reinserted into the instrument and remeasured. The variations are systematic across the spectrum, and therefore mathematical manipulations can correct these variations (or at least bring all readings to a common basis). Some of these characteristics are equivalent to a baseline shift, some are equivalent to a sensitivity change, and others are not describable by such simple means. However, most of the variations can be interpreted as one of those two mechanisms, therefore the listed data treatments address those types of variation.

Normalization is performed on a spectrum-by-spectrum basis. The mean absorbance of each spectrum is computed and subtracted from the absorbance at each wavelength. This corrects the baseline of each in such a way that the average absorbance of each spectrum is zero. The standard deviation of each spectrum's absorbances is computed and the mean-corrected absorbances divided by that value; this adjusts the scales for the spectra, to minimize sensitivity differences.

Multiplicative scatter correction is similar, except that, instead of bringing each spectrum to an arbitrary baseline and sensitivity, the mean spectrum is computed, wavelength by wavelength, from all the spectra in the calibration set, and then each individual spectrum's mean and standard deviation adjusted to match that of this mean spectrum.

Various orders of the derivatives of spectra (with respect to wavelength) are used to correct baseline variations. The first derivative will correct a constant offset, similar to the subtraction of the mean absorbance from each spectrum. Higher-order derivatives extend this capability by correcting higher-order baseline shifts: a linear shift can be corrected by a second derivative, a quadratic (parabolic) shift by the third derivative, etc. Computing derivatives has no effect on differences due to sensitivity changes. Computing derivatives, however, has a disadvantage in that it emphasizes the high-frequency noise. It is usually necessary, therefore, to apply a smoothing transformation to the spectra as well.

Ratios are used to compensate for sensitivity (and pathlength) changes. Ratios of absorbance to that at a chosen wavelength, to a derivative, and other combinations have been tried, with varying success. A difficulty arises when the true ''zero'' for the spectra cannot be determined; this may be the cause of the lack of success in some cases.

Finally, the Kubelka–Munk correction has been used to obtain a more linear relationship between the spectral readings and the concentration composition of samples measured by reflectance. The reflectance (R) at each wavelength is back-calculated from the absorbance, and then the Kubelka–Munk function is computed from Equation (2):

$$f(R) = \frac{(1 - R)^2}{2R} \qquad (2)$$

Since the theory upon which this is based requires that R be known in absolute terms, this approach also encounters difficulties when the true zero and true unity reflectance cannot be determined.

1.2.4 Calibration

The process of applying one or more of the various algorithms used to relate the spectroscopic data to the compositions of the training samples constitutes the ''calibration'' process. The algorithms used will be discussed below. What is important to remember is that, regardless of the algorithm, certain considerations should be kept in mind. The mathematics behind all the algorithms are based on certain assumptions being met. Depending on the computer programs used to implement the calibration procedure, the software may or may not automatically test whether the data at hand conform to these assumptions. Good practice indicates that such tests should be made in any case. Information about the underlying assumptions and ways to test the data can be found in Draper and Smith.[6]

In addition to not meeting the fundamental assumptions, data can be defective in other ways, e.g. contain outliers or be nonlinear. Draper and Smith discuss ways to examine data for various defects; a discussion more directed to the effects found in spectroscopic data is given by Mark.[3]

An additional consideration involves the question of how many factors' worth of data to use in a given calibration model (or, alternatively, how many wavelengths, where that is the appropriate consideration) and which ones. Including too few factors results in a model that does not account for all the interferences in the samples, and thus is inaccurate.

Including too many factors (or wavelengths) results in a model that is ''overfit'', i.e. it tends to model some of the random noise present in the data rather than the systematic variations due to real physical phenomena, and therefore it is not robust. That is, it will not stand up to variations in the samples, their preparation, the instrument, or the environment, and thus the long-term accuracy is compromised.

Therefore, finding the optimum number of factors, and the correct ones, to include in the model is crucial to obtaining best performance. Some of the statistics produced by calibration software help in this search. Some experts recommend creating several calibration models, monitoring their performance over time, and weeding out those that are deficient in accuracy or robustness.

Another recommended practice, mentioned above, is to improve the robustness by continuing to collect samples over time and including them in the calibration calculations. This procedure will eventually include all the variations normally seen in the calibration samples.

1.2.5 Testing the Calibration Model

Whenever a calibration model is created, it is necessary to ascertain whether that model can accurately analyze the constituents in the samples it is intended for. To this end, a number of statistics have been devised, to test different aspects of the performance of the calibration model. An exhaustive list would be impossible to present, since all statisticians have their own favorite set of statistics to use for this purpose. However, some common ones are:

1. Standard error of estimate (SEE, sometimes called standard error of calibration SEC): the standard deviation of the errors found for the samples in the calibration set. This value should be compared with the accuracy of the reference laboratory in order to determine whether the calibration model can be improved. If the SEE is smaller than the error of the reference laboratory, the model is probably overfit.
2. Multiple correlation coefficient (R): a dimensionless number. The minimum value of R is zero; this happens if there is no agreement between the reference

laboratory values and the analytical values found using the instrument with the calibration model under consideration. The maximum value of R is unity; this is obtained when the instrumental measurements agree exactly with the reference laboratory measurements, with absolutely no error. Usually, neither of these extreme values is seen: in-between values are inevitably found. An average ''good'' calibration will achieve a value of R between 0.90 and 0.95, while a superior calibration may attain a value of 0.98–0.99, or even higher.

3. Standard error of prediction (SEP; sometimes the abbreviation SECV (standard error of cross-validation) is found in the literature (incorrectly) used in place of SEP): the standard deviation of the errors found for samples not in the calibration set (see section 1.2.6).

1.2.6 Validation

''Validation'' is the name of the process used to assess the performance of a calibration model. Although there is some disagreement over the exact meaning of the term, the majority opinion at this time is that it consists of ways to estimate the accuracy and robustness of any given calibration model at the time it is created.

There is widespread, even universal, agreement, however, that the best way to determine these qualities is to have a set of samples, similar to but separate from those used to create the calibration model, that also have known constituent compositions (normally from measurements made using the same reference laboratory procedure applied to the calibration samples themselves). When any calibration model is created, it is used to analyze these validation samples, and the result of that analysis is compared with the reference laboratory values. If the agreement between these two sets of analyses (as measured by the SEP) is approximately the same as the agreement for the calibration samples (as measured by the SEE), then, subject to the qualification that the results of any other statistical tests carried out are satisfactory, the model is probably a valid one, and will continue to perform at the accuracy that the SEE and SEP indicate.

1.2.7 Prediction (Analysis)

Whether the intended purpose of the instrument (along with the corresponding calibration model) is raw material inspection, quality control (QC) of product, or any other purpose, this purpose is realized in the actual ongoing routine analysis of the samples for which the calibration model was created. Good practice indicates that the performance of the calibration model used should be monitored on an ongoing basis, by comparing the results from the instrument with those from the reference laboratory at periodic intervals. A continuing QC process of this sort will guard against inaccurate analyses due to instrument malfunctions, as well as other error sources that can occur, such as carelessness in sample handling and preparation.

Indeed, a minority opinion extends the concept of ''validation'' to include this ongoing QC process. There is also a proposal that several calibration models be created and subjected to this extended validation process; then the best model will be chosen after evidence is available as to which model is in fact most accurate and robust.

2 INDIVIDUAL-WAVELENGTH METHODS (ALSO CALLED MULTIPLE LINEAR REGRESSION, P-MATRIX, INVERSE LEAST SQUARES, AND ORDINARY LEAST SQUARES) AND WAVELENGTH SEARCH TECHNIQUES

We now come to the point of dealing with the various mathematical methods that have been developed for actually relating the spectroscopic measurements to the samples' compositions. There are a number of ways in which these can be organized, each of them justifiable. I do not think that one way is so superior that it sticks out as the ''best'', but since there must be some sort of organization imposed on the many methods available, the one used is simply one that is convenient, and perhaps somewhat idiosyncratic in that it is attuned to the way I think about them.

Figure 2 illustrates the situation. Here there are three spatial dimensions; these dimensions can represent three

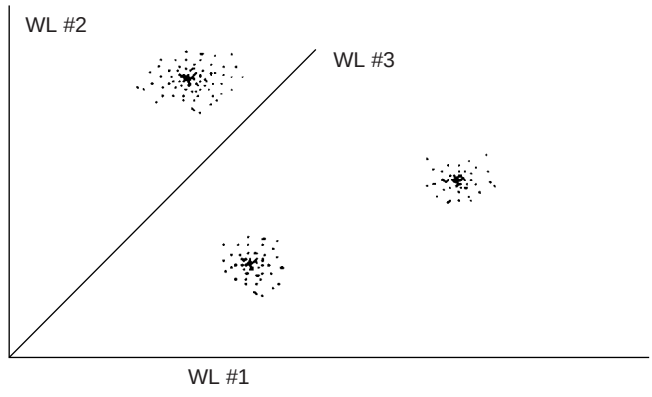

Figure 2 Regions of optimum performance are localized in different parts of the multidimensional (here, three-dimensional) space. It should be noted, of course, that, while the labels of the axes in this space indicate wavelengths, they may equally well be axes representing any of the various types of factors in the factor spaces used.

wavelengths (or three abstract ''factors'' as will be discussed presently). Each combination of the three wavelengths will have a calibration model associated with it, and that model will have some performance characteristics, which are characterized by the values of the auxiliary statistics for that model.

Certain combinations of the factors will have better performance than others. The density of points in Figure 2 represents the ''quality'' of the calibration, as evidenced by the values of the auxiliary statistics, which are thus seen to be localized in the vicinity of those optimum combinations of wavelengths.

The general problem is to find those regions of optimum performance representing the wavelength combinations at which the spectral data give the best analytical behavior. For models based on more than three wavelengths, the situation is the same. The main difference is that it is not possible to draw a picture to show four, five, or more dimensions. The problem of finding the best wavelength set also becomes more difficult, but there is no change in its basic nature.

The existence of more than one region of apparent optimum performance complicates the situation, because some algorithms will find one of these ''local maxima'' and stop searching before finding the global optimum. On the other hand, often the local maximum will behave almost as well as the global maximum, so that in practical terms it makes no difference. The main problem then is to compare models, developed at different times, resulting from these different wavelength choices. A similar situation exists for models based on full-spectrum algorithms. One solution to this dilemma will be discussed shortly.

Historically, individual-wavelength methods were the first to be developed and put into widespread use. While Malinowski may have developed K-matrix (see below) methods contemporaneously with, or possibly even a bit earlier than, Norris' work, this approach was pretty much restricted to academic research laboratories. It was the application of multiple-regression analysis (also now called P-matrix) methods by Karl Norris to NIR spectroscopy that brought calibration out of the academic laboratory and into regular commercial use. The P-matrix methods depend on the careful selection of a relatively small number of wavelengths from the spectrum, and the use of the spectroscopic data at those wavelengths to perform the analysis. Hence, a good deal of effort has been expended in investigating ways to find the ''best'' wavelengths.

To a large extent, the use of P-matrix methodology has been superseded by full-spectrum methods, although it is lately enjoying a slight resurgence, partially because of the new methods of wavelength selection that are being developed.

Therefore, it is worth spending some time discussing this approach because of its historical importance (to help understand the current and future status of this topic), because it is eminently suited to use with lower-cost instrumentation, and because new research reveals situations in which this approach can provide superior results, which may further promote its future applications.

In general, all of the wavelength-search methods are based on selecting some test set of wavelengths and performing a trial set of P-matrix calibration calculations, including the auxiliary statistics, using the data found at those wavelengths. These calculations are then repeated using a different set of test wavelengths, and the statistics from the two trial calibrations are compared. The set with the better statistics is retained. This process is then repeated for however many trial sets of wavelengths are predetermined by the wavelength-search process being used; when this phase is complete, the best wavelength set found, along with the results from the corresponding calibration, is reported.

This procedure presupposes the existence of a larger set of data from which the trial sets of wavelengths are drawn. Normally, this larger set is composed of optical readings at a continuum of closely spaced wavelengths, i.e. a spectrum. The question that immediately comes to mind is: ''Why not simply use all the data available?'' The answer is multifold, but in general it has to do with the mathematical properties of such large sets of data. The mathematics of some types of algorithms, for example of the multiple-regression (P-matrix) approach require that there be more samples' worth of data than there are variables, otherwise divide-by-zero errors will necessarily occur during the intermediate calculations (e.g. during the matrix inversion). In other cases, the nature of the data itself is such as to cause difficulties of that sort; even if the problem is not so extreme as to cause divide-by-zero error, a lesser degree of that characteristic of the data can, and will, cause the final model calculated to be useless (in the sense that it would be even more inaccurate than simply guessing), and any tests involving it will become meaningless.

2.1 Classical Methods

The wavelength-selection methods can further be divided into ''classical'' and ''modern''. The classical methods are loosely defined as those developed before the full-spectrum methods gained ascendancy, and the modern ones are those developed since then, and are helping the P-matrix approach make its comeback. Again, this division is purely idiosyncratic on my part, but it helps keep the various methods in historical perspective. There are also, as we will see, some significant differences in the nature of the search process between the two groups.

2.1.1 Searches at Constant Numbers of Wavelengths

The hierarchy of wavelength-search methods can further be broken down into those that maintain a constant number of wavelengths and those that change the number of wavelengths used in the trial models. Although in actuality these are often combined, it is pedagogically useful to describe them separately.

The search procedures that maintain the number of wavelengths constant are, relatively speaking, easier to compare and evaluate. This is because all the statistical descriptors will improve or degrade simultaneously as different trial models are calculated and compared. Therefore, using any of these descriptors will provide an unambiguous choice of which model to retain for future comparisons during the wavelength-search, a property that is not necessarily maintained when comparing models with different numbers of wavelengths.

2.1.1.1 All Possible Combinations One solution to the problem of finding the best wavelength combination is to try all possible combinations of wavelengths. This is really the only method guaranteed to find the global optimum. It will do this even if the data are discontinuous, i.e. if only separated segments of the spectrum are available, or even if, as in interference filter-based instruments, spectral data at only sporadic and arbitrary wavelengths exist.

There are actually two definitions of ''all possible combinations'', and users should make sure that they know which definition applies to any software package in use. By one definition, trial models for $n!/(m!(n-m)!)$ combinations of the n available wavelengths taken m at a time are calculated; all those models contain m wavelengths. By the other definition, the software calculates all 2^n trial models for the wavelengths taken $1, 2, 3, \ldots, n$ at a time. The second definition is more comprehensive but introduces the difficulty of comparing models containing different numbers of wavelengths.

By either definition, the main limitation of this approach is the amount of calculation required. Since the number of models that need to be tested increases factorially or exponentially as the number of wavelengths available, or the number of wavelengths in the model increases, the amount of computation can quickly overwhelm any physical computer, and require computation time many times that of the lifetime of the universe. Software exists to mitigate these limitations, in order to make the use of the all-possible-combinations approach feasible, but care must still be taken to avoid bogging down the computer this way.

2.1.1.2 Seek and Replace The general idea behind this method of seeking the next wavelength set is to sweep one trial wavelength at a time through the spectrum, and find the wavelength in the spectrum that gives the best model in conjunction with the other wavelengths, while those others are held fixed. Figure 3 illustrates this procedure, for a search for a three-wavelength model, compared with the spectrum, which is indicated by the horizontal line. Figure 3 indicates the starting configuration, which in this case has the three wavelengths all starting from one end of the spectrum. This is not necessarily the case in actual search software, where the initial configuration may be determined by applying extraneous chemical or spectroscopic knowledge, or even by choosing the wavelengths randomly. However, showing all the wavelengths as starting from one end of the spectrum is convenient pedagogically. At step 1, the ''first'' wavelength is swept through the spectrum and the best model noted, so that after step 1 we see that one of the wavelengths is in its optimum position. Then the ''second'' wavelength is swept through the spectrum so that after step 2 another one has been optimized, and after step 3 the third. There is a temptation to stop at this point and declare the search complete.

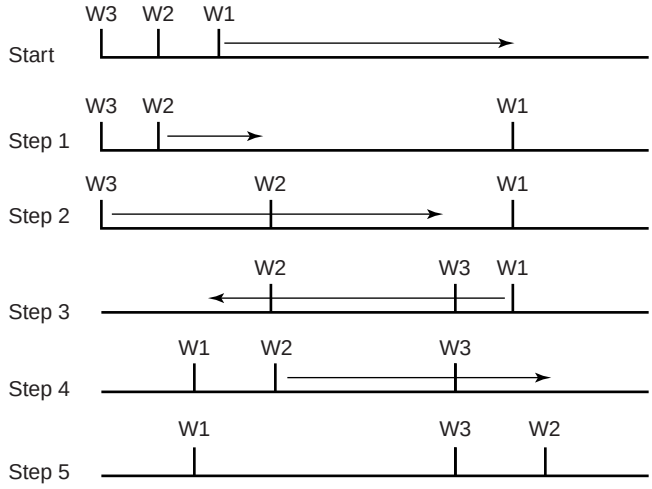

Figure 3 A common method of finding a region of optimum performance is to use the various wavelengths on a trial basis, trying to optimize each one, one at a time. This usually requires cyclically optimizing all of the available ones, until the selection becomes stable.

In practice, however, it is found that the optimum wavelength for any one of the trial wavelengths depends on the actual wavelengths of the others. Therefore, it is necessary to sweep the ''first'' wavelength through the spectrum again, and continue through the trial wavelengths, sometimes several times, before the best set is found. This is usually considered to be achieved when the set of wavelengths becomes stable, that is, sweeping any of the individual trial wavelengths through the spectrum does not find any different wavelength which provides an improved model compared with the current

actual wavelength, so that no new wavelength is chosen. In our example, shown in Figure 3, this occurs at step 5.

2.1.2 Step-up Searches

The main limitation of any search methods that restrict the number of wavelengths to a constant number is, in fact, just that: that the number may be too few or too many for an optimum model. As described in section 1.2.4, a model with too few wavelengths will not allow that model to correct for all the interferences present in the samples, and one with too many will overfit the data, with a consequent lack of robustness. One common way in which these difficulties are addressed is by using ''step-up'' (adding wavelengths) and ''step-down'' (deleting wavelengths) procedures. The step-up procedure is perhaps the easiest to describe, since in one implementation it builds on the seek-and-replace algorithm described in section 2.1.1.2.

Basically, from the best model found using n wavelengths (say, the three found in Figure 3), a new model is sought using $n + 1$ wavelengths, in which the ones already found determine the starting configuration for the search for $n + 1$st. This situation is shown in Figure 4, in which a fourth trial wavelength is added to the three found in the previous search. Normally, the new wavelength added would be the first one to try to improve by sweeping it through the spectrum, as shown, and continuing from that point as before.

In general, given a set of n wavelengths, it is possible to step up from that set as the starting point for finding $n + 1$. This, however, leaves open the question of how to start the search. In principle, there is no reason why one should not simply start from zero wavelengths and search for the best single one. However, in practice, it is often found that this does not lead to such good models as when a search for the best two or even three wavelengths, using a different search method (e.g. all-possible-combinations) is used to start the search going.

2.1.3 Step-down Searches

A step-down search, of course, is just the opposite of a step-up search. That is, having found a model using n wavelengths, one of these wavelengths is deleted, so that a model containing $n - 1$ wavelengths is calculated.

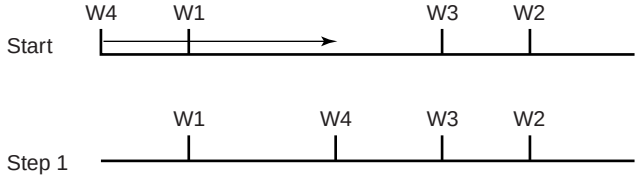

Figure 4 Adding a new wavelength to an already optimized set requires optimizing the value of the new wavelength, and may also require reoptimizing the ones already chosen. Here a fourth wavelength is added to the three finally selected as shown in Figure 3.

However, when starting a new calibration development effort, no wavelengths have yet been chosen; in order to ''step down'' it is desirable to then include all available wavelengths in the first trial model. In order to avoid the problems (described above) encountered when applying multiple linear regression (MLR) calculations to full spectra data, this procedure is sometimes used when the spectral array does not contain data from many wavelengths, for example when the data are from an instrument using a limited number of interference filters. One reason why this approach is limited to that type of data is that it is necessary (as it is always necessary when using an MLR-type of calibration methodology) to have data from many more samples than there are wavelengths in the calibration model. Since the maximum number of wavelengths can be no more than the number of filters in the instrument, this helps ensure that the condition is met, even though data from all the available wavelengths are included in the model. Including all wavelengths also insures that no potentially useful wavelength is excluded from the trial wavelengths.

This approach to calibration also has the advantage of being the least computation intensive of the methods so far discussed. The reason for this is that, when all the available wavelengths are already included in the calibration model, it is a relatively easy task to determine which are contributing the least to the model's performance, and to delete that one. It is therefore rare that it is necessary to sweep the various trial wavelengths through the spectrum multiple times, as was necessary in the step-up approach. In fact, if the software provides the appropriate statistical outputs along with the calibration model itself, no sweeping of any wavelengths is needed; the wavelength that can be deleted can be determined directly from these auxiliary statistics.

2.2 Modern Methods

It is possible to distinguish between the ''classical'' and the ''modern'' methods of wavelength-selection in two ways, which give the same classification of the methods but provide some different insights into the historical and technical developments of the techniques. One method of drawing the distinction is the one described in section 2.1: the temporal relationship between the methods and whether they were developed before or after the full-spectrum calibration methods became popular.

The other way of distinguishing them is as follows: if the various classical wavelength-selection methods described in section 2.1 are examined, it can be noted that, although the modeling calculations themselves are multivariate,

the wavelength-search patterns used to find the proper, or best, wavelengths are, by analogy, univariate. That is, only one wavelength at a time is changed and its effect on the calibration model examined. By contrast, the modern search algorithms are multivariate in the search patterns as well as in the model generation calculations. That is, the search algorithms vary the locations of two or more wavelengths within the spectrum simultaneously, and compare the various models generated by those more complicated sets of wavelengths.

The advantage of these algorithms is that they are computationally much less intensive than, say, the all-possible-combinations search, while having a greater likelihood of finding the best wavelength set at which to create the model than the ''classical'' search methods.

2.2.1 Simplex Optimization

The generic concepts of simplex optimization are nicely described by Walters et al.[7] The advantage of simplex optimization as a wavelength-search technique is efficiency: in general, a simplex will find an optimum set of wavelengths more surely and with much less computation than most other wavelength-search methods.

A simplex is a generalization of a triangle to higher-dimensional spaces. As we saw in Figure 2, for example, in the types of multidimensional spaces that are of interest to us, the axes of the coordinate system of that space represent wavelengths, and in Figure 2 the values of the quantity that varies in that space are represented by the density of the points. In Figure 2, the quantity that varies throughout that ''space'' of wavelengths is the ''goodness'' of calibration models, and so the density there is proportional to the ''goodness''.

To represent this situation in a small number of dimensions, a diagram can be created such as that in Figure 5(a), in which two axes represent wavelengths and the third axis represents the ''goodness'' quantity. In this case, the way in which ''goodness'' varies can be represented by a surface. This situation is shown in Figure 5(a). The simplex, in this case, is an actual triangle: the small triangle shown that represents the search plan. Each corner of the triangle represents two wavelengths and the ''goodness'' parameter of a model at those wavelengths. Being a triangle, the simplex lies in a plane, and this plane is embedded in the surface representing the responses. The triangle (simplex) moves around the surface, seeking the location of the optimum response. It moves by replacing the existing triangle with one that has two corners the same but a third corner that has moved to a location where the response has improved over the response at its original location on the surface. Successive moves of this sort quickly bring the simplex to the optimum position at the ''top of the mountain'', where the response is optimum. Each move thus causes both of the wavelengths corresponding to that corner of the simplex to change simultaneously.

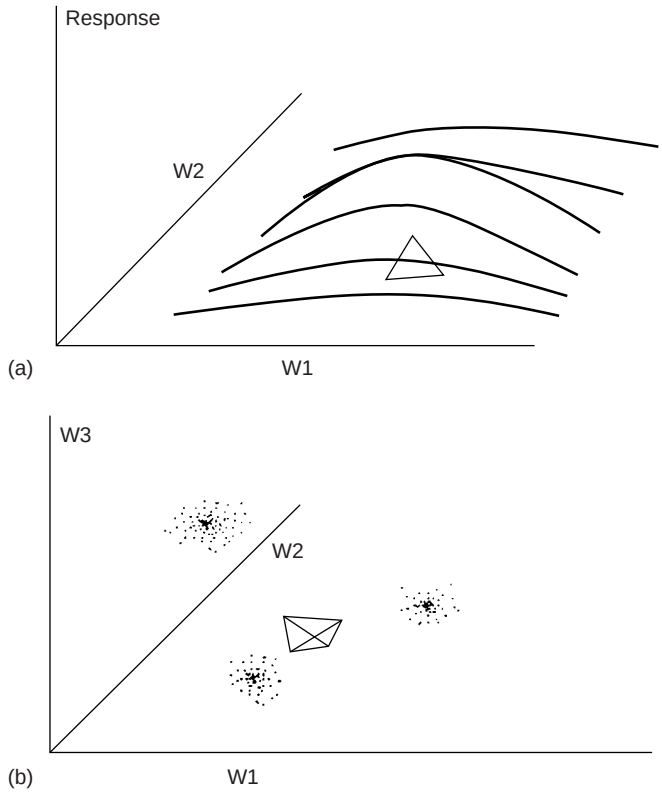

Figure 5 Simplex optimization can be done in any number of dimensions. (a) With two independent variables and one response variable, the situation can be conceptualized as a response surface; optimizing the response is equivalent to finding the ''mountaintop''. The simplex is a triangle, in this case, that ''crawls'' up to the top of the mountain. (b) The regions of optimum performance are the same as in Figure 2. In three dimensions, the simplex is a tetrahedron.

Figure 5(b) shows the similar situation, extended to the case of three wavelengths (plus one response axis, giving a four-dimensional situation); indeed, this figure is almost a duplicate of Figure 2. In this four-dimensional case the simplex is, as shown, a tetrahedron: a three-dimensional object. The tetrahedron moves in a manner similar to the triangle; here, however, three wavelengths change whenever one corner of the simplex moves.

In general, the dimensionality of the space is one more than the number of wavelengths, while the dimensionality of the simplex is one less than the dimensionality of the space. Diagrams for higher dimensions (i.e. more wavelengths) cannot be drawn, but expressing them mathematically, or in a computer program, is relatively straightforward.

2.2.2 Genetic Algorithms

This topic is discussed for the sake of completeness, since at this time (at least as far as we know), this method of wavelength-selection is used only for research purposes and no commercial implementations are available for routine analytical use. More extensive discussions than we can provide here are available in the literature (see Jouan-Rimbaud et al.[8]).

As will be seen, while the genetic algorithm approach to wavelength-selection is similar to the simplex optimization approach in that it provides for changing several wavelengths at once, it differs in one key respect: it provides for different *numbers* of wavelengths to be included in the various models, whereas the simplex can only generate models containing a predefined number of wavelengths.

Genetic algorithms are based on, and named after, analogous operations in biological evolution. The spectrum of a set of samples is considered analogous to a chromosome, and wavelengths selected from those spectra are thus analogous to genes. Just as biological systems (organisms) evolve by selection of the most successful genes from all the combinations that are created at random, so too is a calibration model allowed to ''evolve'' by selecting the most accurate and robust model(s) from randomly created sets of trial combinations, following rules that cause the process to behave in a manner analogous to phenomena found in nature.

Figure 6 shows the effects involved. A number of ''chromosomes'' (spectrum segments) and ''chromosome fragments'' are available for recombination. Some of these ''chromosomes'' are subjected to random effect analogous to ''mitosis'' and ''recombination'' of the genes (wavelengths). In Figure 6, for example, the ''chromosome'' containing ''genes'' w10, w11 and w12 is undergoing ''mitosis'' – splitting into fragments. Figure 5 also shows two fragments joining: the ''chromosome'' containing ''genes'' w4 and w5 is joined to the one containing ''genes'' w13 and w14 to form a new ''organism'' containing four ''genes'' in its ''chromosome''. Each time a new ''chromosome'' is formed, by either splitting or recombination, the ''fitness'' of the ''organism'' to survive is determined by computing the calibration model corresponding to those ''genes'' (wavelengths) and ''only the fit survive'' to ''breed'' a new ''generation'' of calibration models, i.e. only those calibration models meeting a minimum criterion for performance are retained to use for trying new combinations of the available wavelengths. In this manner, better and better combinations of wavelengths are found as succeeding ''generations'' of wavelengths are developed and tested.

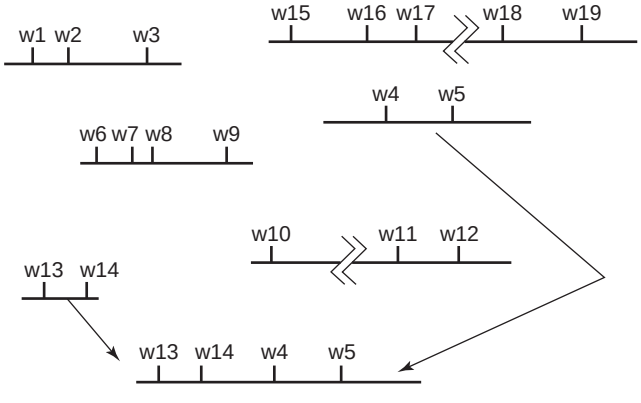

Figure 6 The genetic algorithm works by having the ''chromosomes'' (wavelength sets) randomly splitting and recombining, so that the ''genes'' (wavelengths) eventually get brought together in different combinations that can be tested and kept (or discarded) depending on their calibration performance, thus mimicking ''natural'' selection.

Normally, a ''mutation'' mechanism is also introduced into the system. The ''mutation'' is the random introduction into a ''chromosome'' of a new ''gene'' (wavelength). Through this mechanism, a wavelength that has not been used before can be tested for its effect on the ''evolution'' of the calibration model, thereby removing what would otherwise be a restriction on the possible wavelengths to only those wavelengths initially introduced. With ''mutations'' introduced, any wavelength present in the spectrum can be given a chance to contribute to the calibration model.

3 FULL-SPECTRUM METHODS

These were originally developed to overcome the difficulties encountered in the individual-wavelength methods, of selecting the proper wavelengths and the proper number of wavelengths to include in a calibration model. Historically, they were also developed simultaneously with the development of full-spectrum instruments (NIR interference filter-based instruments were commercially available before other types, particularly diffraction-grating-based instruments), and it seemed sensible to use the full-spectrum capability of the instrument by including the full spectrum's worth of data in the calibration model.

It is not mathematically possible to include the several hundred wavelengths' worth of data typically produced by a diffraction-grating-based instrument (or any of the other full-spectrum instruments) in an MLR calibration calculation. The data must somehow be modified in order to reduce or remove intercorrelations. Indeed, this is one of the major purposes of selecting wavelengths when performing MLR calibrations. If we prespecify, however, that wavelength-selection is not an allowed option, then other methods must be used to perform that task.

3.1 Generic Considerations

While these methods are classified as ''full-spectrum'' methods, that is merely a convenient label, used to distinguish this class of algorithms from the ''individual-wavelength'' algorithms: it is not necessary in fact to have a full spectrum's worth of data in order to apply these algorithms. Disconnected sections of spectra, individual wavelengths, and even mixtures of these are all satisfactory as input data for these approaches. The point, however, is that none of the data is left out of the calculations; whatever is available is included.

We have seen that we cannot simply carry out regression-type calculations using data with so many wavelengths that intercorrelation is severe. Even more importantly, it is not possible to perform regression-type calculations when there are more wavelengths (or variables) than samples, as there often are when the instrument produces several hundred or even several thousand wavelengths' worth of data from each sample that is measured.

The full-spectrum methods work (i.e. allow all the wavelengths to be used) by reducing the intercorrelations among the data used in the final regression step by means other than deleting the wavelengths.

Normally, this is accomplished by transforming the spectral data, by some means, into other variables. These ''other variables'' (sometimes called ''latent variables'') are defined and derived in such a way that they can be used to reconstruct, or recreate, the data spectra. There are two parts to this: the spectra are replaced by the sum of other ''spectra'' that have desirable properties; the main property of these other spectra is that each of the original data spectra can be expressed as a weighted sum of the replacements. The replacement spectra are called ''factors''.

In addition even though each one of the factors is formed from, and contains information from, the entire spectrum, they often have the desirable property that they are uncorrelated with each other; this is desirable because it means that the contribution of any one factor to the original data spectrum is unaffected by the presence or absence of any other factor. The mathematics behind the transformations that provide these properties is very complicated and sophisticated, and far beyond the level of this article. However, we can get a conceptual idea of what and how this is accomplished by inspecting Figure 7. The concept of a ''spectrum'' has been simplified in Figure 7 in order to concentrate on the salient points. The ''data spectrum'' of Figure 7 is marked at the top of the figure. The spectrum, and the factors, each have their own zero line for reference. The zero lines are the heavy horizontal lines associated with the four functions of interest (the spectrum + three factors). The ''factors'' shown in Figure 7 have the following two properties:

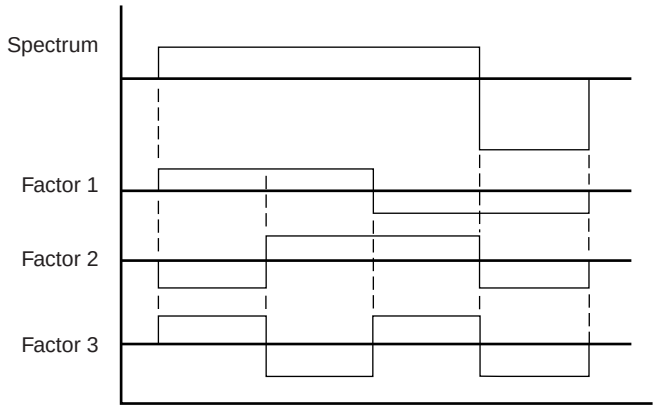

Figure 7 Factor orthogonality requires that the sum of the cross-products between any two factors be zero. To accomplish this, equal numbers of positive and negative cross-products must be present, as in the simplified case shown here, where this is achieved by having the negative and positive parts coincide, and in equal ranges.

1. A suitably weighted sum of these factors can be added up to recreate the spectrum.
2. Each factor has both a positive and negative section to correspond with a positive section and a negative section of the other two factors, i.e. one part of the positive section of factor 1 corresponds to a positive section of factor 2 and the other to a negative section of factor 2.

The important point about this idealized conceptual spectrum and the factors is that they share the properties of real spectra and factors. Property 1 is the crucial one; all decompositions of spectra into factors must provide factors that have this property. Property 2 is the one that causes the factors to be uncorrelated, as described above. This desirable property is called ''orthogonality''.

In principle, there is an infinite number of ways that a set of spectra can be expressed as the sum of other functions (i.e. other spectra), because there is an infinite set of possible mathematical functions that can be used for the reconstruction process. These sets of functions are also called ''basis functions'', ''basis factors'', and ''latent variables''. In practice, there are only a relatively few types of basis functions that are of interest to spectroscopists who wish to use this approach for calibration purposes.

Figure 8 shows how the calibration procedure must be modified, to use the ''latent variable'' or ''factor'' approach. We can compare Figure 8 with the calibration procedure shown in Figure 1 to see how the two approaches are related. The factors are represented by a set of values that can be put into a one-to-one correspondence with the values at each wavelength

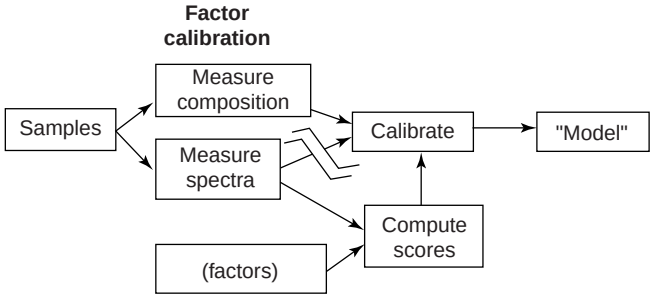

Figure 8 Generalized block diagram for a factor-based calibration. The cross-product between the spectra and factors are the scores, which replace absorbances in the regression computation.

in the data spectra. Multiplying these pairs of values together and adding the products gives a result that transforms the spectrum into a value called the ''score'' of that factor for that spectrum. Thus, although the basic calibration process is the same, the spectra are modified, or transformed, by this multiplication with whatever basis factors are involved, in order to put them into the form that is suitable for calibration. The result of this mathematical operation is thus the ''score'' that essentially describes the ''amount'' or the contribution of that particular basis vector (or basis function) to the actual data spectrum. By including enough basis functions of the type being used, the data spectrum can be recreated to as close an approximation as desired; the scores from the spectra after applying whatever factors are in use then replace the actual spectral wavelengths in the calibration process. The different type of calibration algorithms that are in common use are thus distinguished by the nature of the basis factors that are used to decompose the spectra.

There is an issue here that still needs to be discussed: this is the question of how many factors, and which factors, should be used? This question is entirely analogous to the similar question about wavelength-selection. There is an added complication here, but even with that, the question is somewhat easier to answer.

The complication is the presence of many types of basis functions that might be used, in addition to the multiplicity of functions of any given type. The multiplicity of functions within a type (e.g. the existence of sine and cosine waves of different frequencies when performing Fourier conversion) is analogous to the multiplicity of wavelengths within a spectrum. However, the existence of different sets of basis functions has no analog in the wavelength-selection domain.

As noted above, there is, in principle, an infinite number of these, but even limiting this discussion to those that are commonly used, there are at least four that are commonly encountered; the question is then how to choose among those?

The answer is that, in practice, a choice will rarely have to be made, since that decision will have been made when the software in use was purchased: except for some ''third-party'' software packages, calibration software packages will usually contain an implementation of only one or two of these algorithms, implementing whichever set of basis functions the manufacturer chose to include. Fortunately, in the vast majority of situations, it does not matter which set of basis functions is included, since any of them can achieve adequate calibration performance results. In certain cases, one or another specialized ability of one of the approaches is of special interest; it will then be necessary to make sure that the software available is capable of using the basis functions with that capability.

The calibration algorithms described below are thus the same underlying calibration procedure, using different sets of basis functions. The reader is warned, however, about an inconsistency in the usage of the term ''basis functions'' between mathematicians and chemometricians. The usage here conforms to that of mathematicians; chemometricians use the term ''basis functions'' to mean only those functions that are included in the calibration model, rather than the ones used to recreate the data spectra. When reading the chemometric literature, this difference in usage should be kept in mind.

There is another difference between these various procedures, however, that has to do with the nature of the basis functions themselves, that is the method used to create the various basis functions in any given set. Some types of basis functions are based on a priori mathematical considerations. These are not discussed here to any great extent, since the mathematical bases of these can be found elsewhere, in books dealing with calculus or other higher mathematics, with much more complete explanations than could be provided here. The rest of the calibration algorithms rely on basis functions that are created from the data themselves; those will be discussed in more detail, especially their relative advantages and disadvantages.

The last generic issue to be dealt with concerning calibration is the question of, having selected a set of basis functions as the foundation of the calibration algorithm, how to select the actual set of functions that will be included in any given calibration model. In fact, there is potential here to apply any of the selection algorithms that are used to select wavelengths, but this is rarely, if ever, done. There is good reason for this; since many of the basis functions are chosen so as to be mutually orthogonal, the effect of the inclusion or exclusion of any given function on the model is independent of the presence of any of the others, therefore the decision may be based much more confidently on the values of auxiliary calibration statistics, which greatly simplifies the selection procedure.

The calculation of the calibration models in these cases, however, has an important extra step: the calculation of

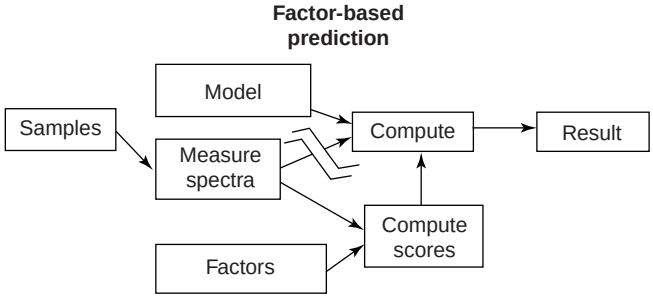

Figure 9 For a factor-based prediction, the scores replace absorbances just as in the calibration.

the basis functions from the data. Besides the mathematical properties of the basis functions themselves, this step creates the biggest differences between the different approaches to calibration modeling.

Figure 9 presents the generic requirement for performing predictions when using a full-spectrum (or factor-based) model. There is an extra step of computing the scores for each factor used from the measured data spectrum, which is computationally intensive compared with the direct calculation of a prediction from an individual-wavelength model. This can also be seen by comparing Figure 9 with the prediction model shown in Figure 1, where the change in flow of computation is emphasized in Figure 9, as was the calibration description shown in Figure 8.

3.2 K-matrix

The K-matrix approach to calibration was, perhaps, one of the first to be developed, although it never got far beyond the walls of academia. This algorithm is the direct application of the concepts behind Beer's law, and in modern times is still called ''the Beer's law method''. It is sometimes seen in other guises, for example spectral subtraction, which is essentially the same concept used in reverse.

The idea behind the K-matrix approach is that, assuming linearity of the instrument and non-interaction of the components, the absorbance spectrum of a mixture is composed of the combined absorbances of the components of the mixture, and can thus be recreated from the spectra of those components. In light of earlier discussion in this article, this causes the spectra of the components to be the factors. A mathematical description of this approach is available in a book by Beebe et al.,[2] (p. 188). The nomenclature used by Beebe et al. for this process is ''direct (classical) least squares (CLS)''.

Although theoretically sound, this concept has a major limitation: the assumption that the spectra of the components of the mixture are the same as their corresponding spectra in the pure state is often not valid. Thus, it does not suffice to measure the spectra of the pure materials and use them in an unmodified form. This difficulty can be circumvented by calculating the spectra of the components in the mixture from spectra of mixtures, which can be achieved if enough of the necessary information is available. By measuring the spectra of a suitable set of mixtures of known composition, it is indeed possible to calculate the spectra of the components as they exist in the mixture, which are the needed and desired spectra. Several authors also describe how these calculations are performed (Beebe et al.,[2] p. 218; Kramer,[9] p. 52). The terminology used is indirect CLS.

Another difficulty arises here, though: it is necessary to know the concentrations of all the components in the mixture, whether they are of analytical interest or not. In many cases, not only are there more constituents that are not of interest than that are of interest, but often the reference analyses for those constituents are more difficult, error prone, and resource intensive than for constituents that are actually of interest. This difficulty has prevented this approach from becoming of commercial interest and it has remained more of a curiosity in academic research laboratories.

In a simplified situation, we could imagine the basis functions being computed as the difference between the spectra of pairs of mixtures, each containing one component at a ''low'' concentration and one at a ''high'' concentration, with ''everything else the same''. Each difference spectrum would clearly represent the spectrum of that constituent, as it exists in the mixture. In the real world, with noise and other nonidealities in the data, the extraction of such approximations to the pure component spectra must be done by regression, in a manner similar to that in which regression was needed to compute the models in the individual-wavelength cases, instead of just solving the simpler case of multiple variables in multiple equations there.

Figure 10 shows a block diagram of the K-matrix approach. Compare Figure 10 with Figure 8, to see how not having the basis functions available a priori complicates the process. To compute the basis functions from the available data is normally the most difficult and computation-intensive portion of these types of calibration modeling algorithms. Note that both the spectra and the set of corresponding concentration values are used twice: once to compute the basis functions and again to compute the final model. This dual use is generally found in all the algorithms that require computing the basis functions from the data. Another distinction of the K-matrix approach is that during prediction, because the basis functions are not orthogonal, the computation of the scores must also be done using a regression approach rather than the simpler cross-product computation used in those cases where the basis functions are orthogonal.

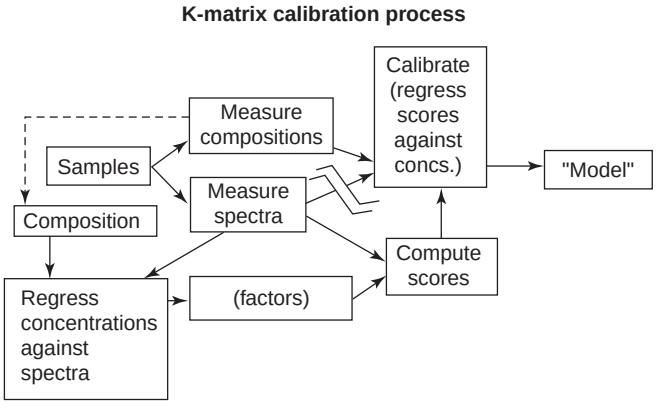

Figure 10 As shown in this block diagram of the generic procedure for a K-matrix calibration, one way to generate the factors is to regress the spectra against the concentrations (instead of the concentrations against the spectra, as is done in P-matrix calibrations).

Note the key difference between the regression used here to create the factors and the regression used earlier to create a calibration model directly from individual wavelengths, even though both use the spectral data and the constituent concentrations: in the earlier situation the spectral data were regressed against the constituent concentrations, whereas here the roles are reversed, and the constituent concentrations are regressed against the spectra. The details of this are also described by Beebe et al.[2] and Kramer.[9]

Finally, it must be reiterated that the defining characteristic of the basis functions determined by the K-matrix approach is that they are estimates of, or approximations to, the pure-component spectra.

3.3 Principal Components

The principal-component approach to calibration is another calibration method based as factors or basis functions which also uses factors that are defined by the set of spectral data from which they are computed.[2,10] (See also Kramer,[9] pp. 103–106). There are some notable differences between the factors produced by this method and the ones created by the K-matrix algorithm.

First, the factors produced are orthogonal. This is a mathematical property that will not be examined in any depth here. Note, however, that it provides the convenient property, mentioned previously, that the contribution of each factor is independent of the contribution of any of the other factors, in the behavior, i.e. the accuracy and robustness, of the calibration model.

The second difference is that the factors are abstract. Whereas the factors produced by the K-matrix algorithm (particularly the indirect least-squares method) recreate the spectra of the pure components, the basis functions

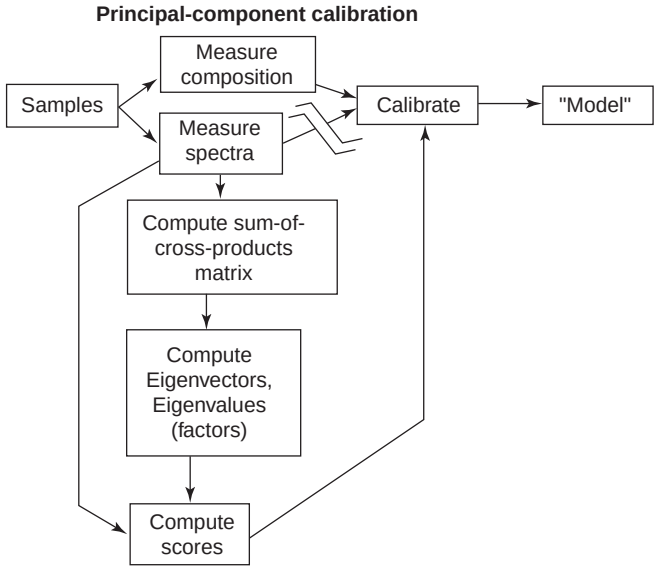

Figure 11 Block diagram showing the generic procedure for generating principal-component factors.

created by the principal-component algorithm do not necessarily have any relationship to any particular physical characteristic of the samples used in their creation. This can happen by coincidence, however, and it is not uncommon to see various signatures of the underlying component spectra in principal-component factors. In Figure 11 the flow of computation involved in the computation of a principal-component calibration model is presented.

The third difference between this algorithm and K-matrix calibration methodology is that, whereas computation of the factors comprising the K-matrix requires both the spectral data and the compositions corresponding to each sample, the computation of the principal-component factors requires only spectral data; the composition does not enter into the factor computation but only into the computation of the final model.

The fourth difference is also the difference between the principal-component algorithm and all other methods of calibration modeling procedures, since it is the key defining characteristic that distinguishes the principal-component algorithm from all other methods of creating basis functions. The key defining characteristic of principal components, which distinguishes them from all other sets of basis functions, is that, when they are used to reconstruct the original data spectra from which they were created, they can do so with less error than any other possible basis function or set of functions. That is, the first principal component can reconstruct the data spectra better than can any other single function, of any type. The first two principal components can reconstruct the data spectra better than any two functions can, whether based

on theory or data. And so forth. This capability is well described graphically.[9,11]

3.4 Partial Least Squares

The partial-least-squares (PLS) algorithm is an attempt to improve on the factors produced by the principal-component calculations and to create basis functions that can better relate to the constituent compositions, rather than simply being the ones that best relate to the spectra. PLS has been widely adopted, and several treatises are available that describe the algorithm in detail.[12,13] (See also Kramer,[9] pp. 139–142).

Figure 12 provides a simplified overview of the workings of the algorithm. Some key differences from the K-matrix algorithm, and some differences from the principal component algorithm are apparent.

A major difference of the PLS approach from the other methods can be seen from an appraisal of Figure 12: information from the constituent concentrations of the samples is included in the computation of the factors. Algorithms exist that do not require iteration and this is illustrated in Figure 12 by the cyclic flow of computations between the model, the factors, the scores, and back again. This is indicative of a key characteristic of the algorithm shown: rather than working solely with a set of factors that model only the principal variations of the spectral data in isolation, it also includes a contribution from the information about the consituents. This inclusion of the constituent information in the factor computation modifies the factors so that they are better able to model these constituents, rather than solely modeling the spectra.

While the PLS method is somewhat similar to the principal-component method in that the PLS factors are partially determined by their ability to reconstruct the spectral data, the inclusion of the constituent concentrations in the loop calculations causes those factors to be ''rotated'' (this is the statistical nomenclature used to describe this operation) in multivariate space to a direction that causes the factors to become more predictive of these concentrations; this rotation distinguishes the PLS factors from principal-component factors. They also differ from principal-component factors in that PLS factors are not necessarily orthogonal.

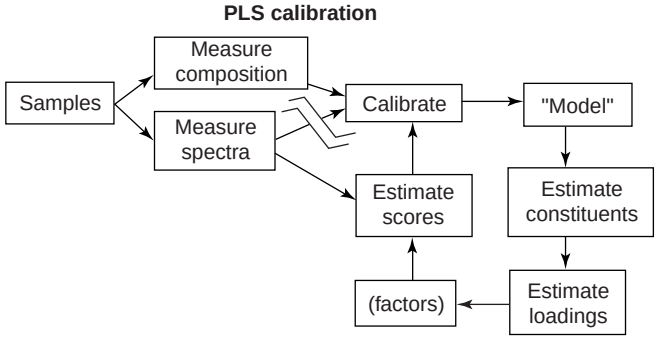

Figure 12 Block diagram showing the generic procedure for generating PLS factors. Note the cyclic computation resulting from the iterative nature of the calculation.

The PLS factors differ from K-matrix factors in that, like principal-component factors, they are abstract, as discussed above. Also, the PLS factors are optimized for their task of quantitative analysis. Neither K-matrix factors nor principal-component factors are necessarily optimized for this purpose, since they are defined and computed to meet other criteria.

3.5 Fourier Transform

The Fourier transform approach is, perhaps, the simplest of the algorithms discussed so far. This is because the basis functions used are defined by a priori mathematics: sine and cosine functions. This being the case, the calibration and prediction explanatory diagrams are exactly those presented in Figures 8 and 9, with the arbitrary (factors) designation replaced with the sine and cosine functions, the full set of functions consisting of the sines and cosines of various frequencies. The use of this approach was developed by McClure for use in connection with NIR analysis; an extensive description is available: see Chapter 8 in Burns and Ciurczak.[5]

An interesting use that McClure made of this calibration methodology was data compression. Since the Fourier components can be calculated from a priori mathematical considerations, it is not necessary to maintain a copy of the basis functions along with the model in order to do predictions, as is often required by other full-spectrum methods. In those cases, a copy of the exactly calculated basis functions must be associated with the rest of the model, since each set is unique to the data set from which they were calculated.

Thus, McClure found that a large saving in computer storage space could be achieved, since each spectrum could be represented by only a relatively few of the Fourier coefficients, and the full spectrum reconstructed from just those values, since the basis functions themselves could be calculated on the spot when needed, from their definition.

This calibration methodology never gained widespread acceptance, and has lately fallen into disuse.

3.6 Neural Nets

This is one of the newer calibration methodologies that has been developed, and currently enjoys the status of being used mainly for academic research, although that situation is likely to change, and is the reason it is discussed

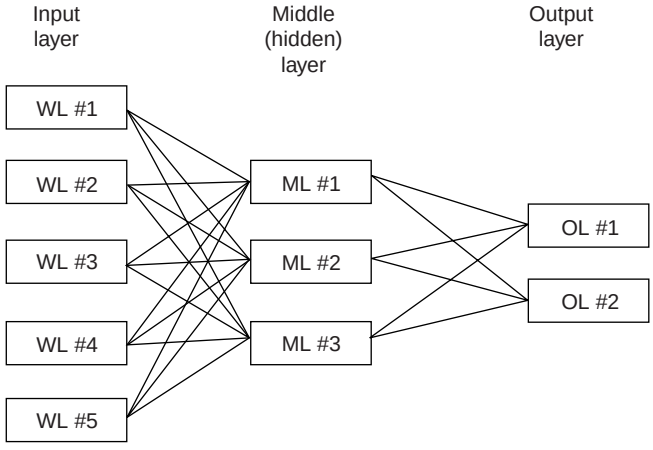

Figure 13 Block diagram showing the generic procedure for generating calibration models using a neural net procedure.

here. As with genetic algorithms, the topic has so recently become used for calibration purposes that one must often go to the primary literature for detailed information (see, for example, Naes et al.[14] and the references therein). An elementary description comparing artificial neural networks with natural (biological) neural networks is also available.[15]

An explanatory diagram is presented as Figure 13; this differs from the format of the previous diagrams because, by presenting it this way, it not only conforms to the standard presentation of this topic, but it can also be used to show how the neural net approach is a generalization of some of the other calibration modeling procedures.

Figure 13 shows some of the key characteristics of neural nets: an input layer, one or more middle (or hidden) layers and an output layer. The boxes represent the nodes of the neural net. For the sake of presentation, only a small number of nodes are shown in each layer. One of the variables in neural net construction is the number of nodes in each layer; actual neural nets may contain hundreds or even thousands of nodes in each layer. Another variable in the construction of neural nets is the number of hidden layers; one hidden layer is shown, and that is a common architecture for these nets, but more than one hidden layer may be employed.

The input layer represents the input data, or spectrum, so the number of input nodes would ordinarily equal the number of wavelengths in the spectrum. The output layer represents the results; but the nodes of the output layer also include computation capability, as do the nodes in the hidden layer. The hidden and output layers are where the computations are performed, to create the model that allows calibration results to be formed from the input data. Each node of the output layer can be seen to represent the results of models for a different constituent. The nature of the computation determines the output result, and the computation internal to a node is termed the transfer function of the node.

Each node in one layer is connected, as shown, to all the nodes in the next layer and, conversely, each node is connected to all the nodes in the previous layer. The connections, shown here simply as lines, in an actual neural net comprise weighting functions (w_i) and the result of the computation of any node at level L_i is given here in Equation (3):

$$L_i = T_j \sum_j w_j L_{j,i-1} \tag{3}$$

or, in words, the weighted sum w_j of the j outputs of the previous $(i-1)$ layer, multiplied by the transfer function of the layer. Thus, the calibration algorithms examined above can all be seen to be special cases of the neural net concept, in that if the hidden layer were removed, the neural net would degenerate to one or another of the various modeling methods discussed earlier, depending upon the calculations included in the output nodes. For example, if the transfer function of the output layer computes the square of the errors, the minimizing of the sum-squared difference and limiting the wavelength set becomes the recipe for the MLR algorithm. Commonly, the transfer function is a sigmoid, which thus causes the result of the node to vary between zero and unity, depending on the result of the computation of the weighted sum.

Clearly there is room for much flexibility here. Simply by changing the transfer function to something other than minimizing the sum-squared-error, completely different types of model generation algorithms can be specified, even without a hidden layer. The inclusion of a hidden layer (or layers), then, immensely increases the potential for creating models to characterize any sort of data. Research is ongoing to determine the effects that can be achieved, and also to characterize the pitfalls encountered using this methodology.

One pitfall which everyone seems to agree on is that it is very easy to overfit the data.

ABBREVIATIONS AND ACRONYMS

CLS	Classical Least Squares
GLP	Good Laboratory Practice
MLR	Multiple Linear Regression
NIR	Near-infrared
PLS	Partial-least-squares
QC	Quality Control
SEC	Standard Error of Calibration
SECV	Standard Error of Cross-validation
SEE	Standard Error of Estimate
SEP	Standard Error of Prediction

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Infrared Spectroscopy in Clinical and Diagnostic Analysis • Near-infrared Spectroscopy, In Vivo Tissue Analysis by

Polymers and Rubbers **(Volume 9)**
Mechanical Properties of Polymers and Rubbers

Process Instrumental Methods **(Volume 9)**
Chemometric Methods in Process Analysis

Chemometrics **(Volume 11)**
Chemometrics • Classical and Nonclassical Optimization Methods • Multivariate Calibration of Analytical Data • Second-order Calibration and Higher • Soft Modeling of Analytical Data

Infrared Spectroscopy **(Volume 12)**
Quantitative Analysis, Infrared • Spectral Data, Modern Classification Methods for

REFERENCES

1. *Webster's Seventh New Collegiate Dictionary*, G. & C. Merriam Co., Springfield, MA, 1970.
2. K.R. Beebe, R.J. Pell, M.B. Seascholtz, *Chemometrics: A Practical Guide*, John Wiley & Sons, New York, 1998.
3. H. Mark, *Principles and Practice of Spectroscopic Calibration*, John Wiley & Sons, New York, 1991.
4. A. Savitsky, M.J.E. Golay, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedure', *Anal. Chem.*, **36**, 1627–1639 (1964).
5. D.A. Burns, E.W. Ciurczak, *Handbook of Near-Infrared Analysis*, Marcel Dekker, New York, 1992.
6. N. Draper, H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, 1981.
7. F.H. Walters, L.R. Parker, S.L. Morgan, S.N. Demin, *Sequential Simplex Optimization*, CRC Press, Boca Raton, FL, 1991.
8. D. Jouan-Rimbaud, D. Massart, 'Genetic Algorithms as a Tool for Wavelength Selection in Multivariate Calibration', *Anal. Chem.*, **67**, 4295–4301 (1995).
9. R. Kramer, *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York, 1998.
10. E.R. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, John Wiley & Sons, New York, 1991.
11. H. Mark, 'Data Analysis: Multilinear Regression and Principal Components Analysis', in *Handbook of Near-Infrared Analysis*, eds. D. Burns, E. Ciurczak, Marcel Dekker, New York, 107–158, 1992.
12. H. Martens, T. Naes, *Multivariate Calibration*, John Wiley & Sons, New York, 1989.
13. M. Sharaf, D. Illman, B. Kowalski, *Chemometrics*, John Wiley & Sons, New York, 1986.
14. T. Naes, K. Kvaal, T. Isaksson, C. Miller, 'Artificial Neurol Networks in Multivariate Calibration', *J. Near Infrared Spectrosc.*, **1**, 1–12 (1993).
15. W.C. McClure, M. Hana, J. Sugiyama, 'Neural Networks in NIR Spectroscopy', in *Making Light Work: Advances in Near Infrared Spectroscopy*, eds. I. Murray, I. Cowe, VCH Publishers, New York, 652, 1992.

Spot Test Analysis

Ervin Jungreis
The Hebrew University of Jerusalem, Jerusalem, Israel

1 Introduction **1**
2 Techniques and Equipment **2**
2.1 Techniques 2
2.2 Equipment 2
3 Selected Tests in Inorganic Analysis **2**
3.1 Examples of Preliminary Tests 2
3.2 Tests for Cations 3
3.3 Selected Tests for Anions 5
4 Selected Tests in Organic Analysis **6**
4.1 Preliminary Tests 6
4.2 Selected Tests for Functional Groups 7
4.3 Detection of Individual Organic Compounds 9
5 Selected Examples of Applications of Spot Tests **11**
5.1 Applications in Clinical Analysis 11
5.2 Forensic Application of Spot-test Analysis 14
5.3 Selected Geochemical Applications of Spot Tests 15
5.4 Rapid Selected Screening Tests for Soil and Plant Tissues 15
5.5 Selected Examples of Applications of Spot Tests in Air and Water Quality Control 16
5.6 Food Quality Control Analysis by Spot Test 17
Abbreviations and Acronyms **18**
Related Articles **18**
References **18**

Sensitive and selective qualitative tests based on chemical reactions using small amounts of complex and reagents are referred to as ''spot tests''. The target is to reach the utmost sensitivity and selectivity with a minimum of physical and chemical operations.

Rapid advances in spectral, chromatographic, nuclear, mass spectroscopic and electroanalytical research have changed analytical chemistry spectacularly and developed highly sophisticated analytical tools. At the same time another development has taken place in certain distinct areas of application in exactly the opposite direction. This is the search for simple, compact and inexpensive analytical devices for semiquantitative evaluation of certain elements or compounds where such approximate evaluation has diagnostic value at least in the first stage of examination.

Major manufacturers of chemical reagents concentrated their efforts mainly in a number of applicative areas, namely clinical analysis, forensic tests, air and water quality control (QC) tests and soil and food QC tests.

It must be emphasized that the applicability of spot test methodology is limited and for the exact determination of chemical substances, complex analytical procedures are unavoidable. Although the use of spot and screening tests is marginal, the margin, however, is quite significant.

1 INTRODUCTION

Spot tests are defined as an analytical technique that enables the analyst to accomplish satisfactory semimicro and ultramicro tests with simple equipment in a minimum time. The target is to reach the utmost sensitivity and selectivity with a minimum of physical and chemical operations, and the procedure must be as simple as possible.

The modern systematic development of spot test methods occupied Fritz Feigl[1,2] for half a century in Vienna and Rio de Janeiro, but analytical chemists have long used single chemical tests on filter paper. These are the indicator papers for the detection of an excess of hydrogen or hydroxy ions, potassium iodide–starch paper for chlorine detection, silver carbonate-impregnated filter paper for uric acid detection in urine, and others. According to the writings of Pliny the Elder, the early Romans detected iron in vinegar by means of a reagent paper prepared by soaking papyrus in an extract obtained from gall nuts.

In parallel with the trend towards instrumental sophistication using the phenomenal development of microprocessors, a trend towards simplification was seen in some marginal selected areas in the form of simple, rapid and inexpensive spot and screening tests.[3] Commercial companies sell vast numbers of compact spot-test systems in which the goal is the rapid establishment of the presence or absence of certain materials. It would be wasteful to devote resources to attaining precision or accuracy several magnitudes beyond that necessary in clinical urine analysis, for example, when pathological changes in the composition of urine are evidenced by a simple test. A new look was taken at the classical tests; sometimes they were refined and sometimes new ones were elaborated. In most of the tests the human eye was used as a detector; others employed simple reflectometric instruments and UV

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

(ultraviolet) lamps. Most commercial companies nowadays concentrate their efforts in the following applicative areas: clinical, forensic, soil and plant, air, water and food QC tests, because of the vast number of screening tests needed in these areas.

2 TECHNIQUES AND EQUIPMENT

2.1 Techniques

Spot-test procedures are the ultimate in simplicity. Classical tests are ordinarily run by using one of the following techniques:

- by bringing together one drop each of the test solution and reagent on porous or nonporous supporting surfaces such as paper, glass or porcelain;
- by placing a drop of the test solution on a medium impregnated with appropriate reagents;
- by subjecting a drop of reagent or a strip of reagent paper to the action of liberated gases from a drop of the test solution or from a minute quantity of the solid specimen.

Commercial companies produce strips of absorbent cellulose, one end of which is impregnated with the dry reagent system. These strips are mainly used in clinical, chemical and water analysis systems. The impregnated reagent area is sometimes covered with a semipermeable membrane to prevent staining by a colored matrix such as blood. Another form of reagent system is a stable compact reagent tablet, which can contain several compatible agents such as buffers and sequestering agents. Test strips for body-fluid analyses are prepared in a modern technique by printing liquid-absorbing substances on a support, drying the treated support and applying reagents to the film formed on the support. Simple color comparing devices may be part of commercial test systems. Some are calibrated color charts printed on the container, others are colored cubes molded of acrylic plastic and matched to laboratory standards. A series of nonfading glass color standards mounted on a circular disc may be used in a comparator box or continuous color discs may be applied for visual comparison.

For the semiquantitative testing of airborne contaminants in a workroom atmosphere, the detector tube technique has gained ground. The reagent system placed into the tube consists of supporting material impregnated with the chromogen selective for the air contaminant.

The technique of separation of small water-soluble molecules from high-molecular-weight proteins by a semipermeable layer is utilized in the routine and simple screening test for glucose and urea in blood.

2.2 Equipment

The amount and type of equipment needed is relatively small. It is most important to have an ample supply of reagents immediately available. Balances, UV lamps, centrifuges, dryers, ovens, furnaces and a hood can be considered to be essential in the spot-test laboratory. Assorted sizes of beakers, volumetric flasks, Conway cells, crystallizing and evaporating dishes, filter sticks, separation funnels, extraction pipets, fritted glass crucibles, graduated cylinders, vials, test tubes, centrifuge tubes, cover glasses, white and black spot plates, porcelain crucibles and platinum dishes are standard for many spot-test operations. Tweezers and spatulas, microburners, blow torches, sand and water bath are essential, too.

3 SELECTED TESTS IN INORGANIC ANALYSIS

The sample should be subjected to preliminary tests before beginning a systematic search, although it is often possible to carry out conclusive tests with as little as one drop of dilute sample solution, even though considerable concentrations of other substances are present.

3.1 Examples of Preliminary Tests

A first test is carried out by heating the solid sample in contact with air in a microcrucible. If the sample's color and structure remain unaltered, it shows the absence of volatile compounds, e.g. absence of ammonium and mercury salts, carbonates, organic compounds, water of crystallization. If the sample melts, it is an indication of the possible presence of nitrates, nitrites, carbonates, formates, chlorates, silver chloride and bromide and fusible organic compounds. If the sample chars, it is a sign of the presence of organic materials, organometallic compounds or metal salts of organic acids. If the sample changes color without charring, it shows the presence of certain heavy metal salts, such as cupric salts which are oxidized to black cupric oxide, and cadmium salts which are oxidized to dark-brown cadmium oxide. Lightening of the color may indicate the decomposition of higher oxides, e.g. lead dioxide to lead oxide, or oxidation of sulfides, e.g. lead sulfide to lead sulfate.

A second test is carried out by heating the solid sample in an ignition tube. If the sample gives off water, the presence of salts containing water of crystallization, metal hydroxides and oxyhydrates and ammonium salts of organic materials are indicated. If litmus paper detects acidic vapors, it shows the presence of acid salts or heavy-metal sulfates or sulfites.

If alkaline vapors are detected by litmus paper, ammonium salts of heat-stable nonoxidizing acids or

complex metal amines might be present. If a white sublimate appears, it indicates the presence of ammonium salts of volatile materials, As_2O_3, Sb_2O_3, or mercury salts of volatile acids or oxalic acid. The appearance of a yellow sublimate is a sign of the presence of As_2O_3, As_2O_5, HgI_2 or free sulfur. If a gray sublimate shows up, it might be a sign of metallic mercury derived from mercury compounds or of free arsenic. Yellow or brown vapor might show bromine derived from bromide in the presence of oxidizing agents, or nitrogen oxide derived from nitrate or nitrite, or chromyl chloride provided by dichromate and chloride. Colorless vapors show up when CO_2 derives from carbonates, CO derives from oxalate and formate, HCN derives by thermal decomposition of silver, mercuric or palladium cyanides, H_2S derives from water-containing sulfide or thiosulfate or NH_3 derives from ammonium salts and so on.

3.2 Tests for Cations

A few selected examples are described here. The selection is based on the simplicity of execution, on the selectivity and on the sensitivity of the reagents. The short description here might be enough to allow the tests to be carried out, but Feigl and Anger[1] should be consulted for more detail and a deeper understanding of the reactions.

Aluminum ions form a lake stable to acetic acid with quinalizarin. The test solution is placed on the reagent paper (filter paper soaked in 10 mg reagent dissolved in 2 mL pyridine + 20 mL acetone) and held over concentrated NH_3 until the initial blue color disappears. Visible change: a remaining red-violet color. Sensitivity: 0.005 μg aluminum. Interference: Be^{2+}.

Ammonium ions liberate volatile NH_3 by adding alkali to them in gas-evolving test tubes. A strip of moist red litmus paper is hung on the hook of the tube, which is closed and warmed to about 40 °C for 5 min. Visible changes: blue color. Sensitivity: 0.01 μg ammonia. Interference: cyanides.

Antimony ions reduce phosphomolybdic acid to molybdenum blue. A drop of the test solution is placed on filter paper impregnated by 5% aqueous reagent and held over steam. Visible change: blue coloration. Sensitivity: 0.2 μg antimony. Interference: stannous ions.

Arsenic ions when oxidized to arsenate form an insoluble compound with silver ions. The test solution is warmed in a microcrucible with a few drops of ammonia and 10% H_2O_2, then acidified with acetic acid and 1% $AgNO_3$ added. Visible change: red-brown precipitate. Sensitivity: 6 μg arsenious acid. Interference: the test is specific in the absence of chromates and ferricyanides.

Barium ions form an insoluble salt with sodium rhodizonate solution. To a drop of the neutral or slightly acid test solution on filter paper is added 0.2% aqueous reagent (freshly prepared) dropwise. Visible change: red-brown stain. Interference: bivalent heavy-metal salts must be absent. Sensitivity: 0.25 μg barium.

Beryllium ions give red-violet color lake with chromeazurole. One drop of the test solution is treated with a drop of 2 N sodium acetate and one drop of the alcoholic reagent solution. Visible change: red-violet color. Sensitivity: 0.3 μg beryllium. Interference: Fe, Al, Zn and copper salts must be masked.

Bismuth ions form insoluble double iodide with univalent organic bases, among them cinchonine. Slightly acidic test solution is dropped onto filter paper impregnated with the reagent solution (1 g cinchonine dissolved in 100 mL water containing a little nitric acid, by warming; after cooling, 2 g potassium iodide is added) Visible change: orange-red stain. Interference: from lead, copper and mercury can be avoided by zone precipitation.

Cadmium ions are precipitated with ferrous dipyridyl iodide, a compound in which both the cation and the anion are large complexes. A drop of the test solution is treated with a drop of the reagent solution. Visible change: red precipitate. Sensitivity: 0.05 μg cadmium. Interference: mercury and nickel ions. The reagent solution is prepared as follows: 0.25 g α,α'-dipyridyl and 0.146 g $FeSO_4 \cdot 7H_2O$ are dissolved in 50 mL water, then 10 g KI are added, and after vigorous shaking any solids remaining are filtered off.

Calcium ions form colored inner complex salts with glyoxal-bis(2-hydroxyanil), which are stable against alkali carbonate and extractable in chloroform. A drop of the test solution is treated successively in a test tube with 3–4 drops of the reagent solution, one drop of 10% NaOH, one drop of 10% Na_2CO_3 and 3–4 drops of chloroform. Visible change: red color in the chloroform layer. The test is specific for calcium. Although barium and strontium ion form a red complex with the reagent, they are not extractable with chloroform. Sensitivity 0.05 μg calcium.

Chromium ions that are oxidized to chromate react with diphenyl carbazide and form an inner complex salt in which bivalent chromous ion participates. A drop of the test solution is mixed on a spot plate with a drop of saturated solution of potassium persulfate and a drop of 2% $AgNO_3$ solution and allowed to stand for 2–3 min. By adding a drop of 1% alcoholic diphenylcarbazide solution, a violet to red color is formed. Sensitivity: 0.8 μg chromium. Interference: Mo, Hg, V.

Cobalt ions form a series of cationic and anionic complexes with thiocyanates in the presence of ethyl alcohol or acetone. A drop of the test solution is mixed with a few drops of saturated acetone solution of ammonium thiocyanate on a spot plate. Visible change: green to blue color. Sensitivity: 0.5 μg cobalt. Interferences: Cu^{2+} and Fe^{3+}, which can be avoided by removal of copper through precipitation in the form

of cuprous thiocyanate by adding reducing agents, and complexing iron with fluoride, respectively.

Copper ions are precipitated with rubeanic acid (dithiooxamide). A drop of the neutral test solution is placed on filter paper, held over ammonia and treated with a drop of 1% alcoholic solution of rubeanic acid. Visible change; black or olive-green color. Sensitivity: 0.06 µg. copper Interferences: Ni^{2+}, Co^{2+}, Fe^{3+}, Ag^{+}, Bi^{3+}, Pt^{2+}, Pt^{2+}. All these interferences are avoided by complexing them with 20% malonic acid and 10% ethylenediamine.

Gold ions (tervalent) are easily reduced to metallic gold with strong reducing agents. Filter paper impregnated with stannous chloride is dried at not more than 40 °C for 2 h. On adding a drop of a solution containing gold, violet-colored metallic gold forms. Sensitivity: 1 µg gold.

Iron (bivalent) ions give a stable complex cation with α,α′-dipyridyl. A drop of the test solution is placed on filter paper which has been impregnated with a 2% solution of the reagent in thioglycolic acid. The latter reduces tervalent iron. Visible change: red color. Interferences: considerable quantities of cobalt and nickel ions. Sensitivity: 0.03 µg iron.

Lead ions are precipitated as chelates with dithizone. A drop of the test solution is shaken in a test tube with the reagent dissolved in carbon tetrachloride. Visual change: red color. Sensitivity: 0.04 µg lead. Interferences: heavy metal ions disturb, but complexing with alkali tartrate and cyanide [CARE], makes the identification of lead specific.

Magnesium ions give a blue lake with quinalizarin in alkaline solutions. A drop of the test solution and a drop of distilled water are placed in adjoining depressions in a spot plate and mixed with 2 drops of an alcoholic 0.02% solution of the reagent. 2 N NaOH solution is added drop by drop until a change to violet occurs. Visual change: a blue precipitate in comparison with the violet blank. Sensitivity: 0.25 µg magnesium. Interferences: La, Be, Nd, Pr, Ce, Zr and Th.

Manganese ions are oxidized catalytically to violet permanganates. A drop of the test solution is mixed with a drop of concentrated sulfuric acid in a microcrucible. A drop of 0.1% silver nitrate solution is stirred in, followed by a few milligrams of ammonium persulfate. The mixture is gently heated. Visual change: red-violet color. Sensitivity: 0.1 µg manganese. Interference: halides should be absent.

Mercuric ions form a red cuprous mercuric iodide complex with a white suspension of cuprous iodide. The acidic test solution (pH = 0) is treated on filter paper with a drop of potassium iodide (5%)–sodium sulfite solution (20%), followed by a drop of 5% $CuSO_4$ solution Visual change: red or orange color. Sensitivity: 0.003 µg mercury. Interferences: Ag^{+}, Hg_2^{2+}, Au^{3+}, Fe^{3+}, Ce^{4+}. All interferences might be avoided by a pretest treatment.

Molybdenum (tervalent) ions produce a red water-soluble $H_3[Mo(CNS)_6]$ complex. A drop of the test solution and a drop of 10% potassium thiocyanate solution are placed on filter paper previously moistened with 1:1 hydrochloric acid. In the presence of ferric iron, a red stain appears which disappears when adding 5% stannous chloride in 3 N HCl. In the presence of molybdate, a brick red Mo^{3+}–thiocyanate complex forms. Sensitivity: 0.1 µg molybdenum.

Nickel ions form a red chelate with dimethylglyoxime. A drop of the test solution and one drop of 1% alcoholic reagent solution are placed on filter paper and held over ammonia. Sensitivity: 0.16 µg Ni. Interferences: large amounts of oxidizing substances. Pd gives an analogous yellow chelate.

Palladium(II) ions form an acid-insoluble yellow chelate on the surface of red nickel dimethylglyoxime paper. This layer protects the underlying red compound against dissolution in acid. A drop of the test solution is placed on the reagent paper and almost dried by gentle warming. The paper is then bathed in dilute HCl. The red compound dissolves except for the palladium protected spot. Visual change: a red fleck remains. Reagent paper preparation: filter paper is bathed in cold saturated alcohol solution of dimethylglyoxime. After drying, the paper is placed in 2 N $Ni(NO_3)_2$ solution made ammoniacal. The paper is washed with alcohol. Sensitivity: 0.05 µg palladium.

Platinum(IV) salts, like those of gold, palladium and osmium in acid solution, are reduced to the elementary state by stannous chloride using the following procedure. The platinum can be detected even in the presence of these other noble metals. A drop of saturated thallium nitrate solution is placed on filter paper, followed by a drop of the test solution and another drop of $TlNO_3$ solution. The paper is washed with ammonia. $Tl_2[PtCl_6]$ remains on the paper only. A drop of stannous chloride solution in strong HCl is added. Visual change: orange-red stain. Sensitivity: 0.025 µg platinum.

Potassium ions are precipitated with sodium tetraphenylboron. The test is particularly valuable if potassium is to be detected in the presence of a lot of sodium. A drop of the test solution is placed on a black spot plate with a drop of a 2% water solution of the reagent. Visual change: white precipitate. Sensitivity: 1 µg potassium. Interferences: NH_4^{+}, Cs^{+}, Rb^{+}, heavy metal ions.

Silver ions are precipitated with manganese salts and alkali, as MnO_2 and metallic silver. Both are black. Since mercurous salts, noble metals and stannous salts react analogously, AgCl should be isolated beforehand in order to prevent such interferences. A drop of 0.1 N HCl is placed on a filter paper followed by a drop of the test solution and then a further drop of HCl. The fleck is

blackened at the addition of a drop of 0.1 N manganese nitrate and a drop of 0.1 N NaOH. Sensitivity: 2 μg silver.

Sodium ions impart a characteristic yellow color in the flame. This flame test is extremely sensitive, so even omnipresent sodium traces are detectable. A sodium concentration of as little as 0.01 ppm yields a very persistent and intense yellow coloration in the flame.

Strontium ions react in neutral solution with sodium rhodizonate and form a brown-red precipitate. To detect Sr in the presence of Ba, the latter must be converted first to insoluble sulfate. The solubility of $SrSO_4$ allows its detection with rhodizonate. A drop of the test solution is placed on filter paper impregnated with saturated sodium sulfate solution and the sulfates are formed. After a minute, a drop of freshly prepared 0.2% reagent solution is placed on the spot. Visible change: brown-red precipitate.

Tin(II) ions reduce ammonium phosphomolybdate (in contrast antimony salts reduce the free phosphomolybdic acid only) to molybdenum blue. Filter paper impregnated with an aqueous 5% phosphomolybdic acid solution is held over ammonia until an insoluble yellow material forms on the reagent paper. Quadrivalent tin must be reduced with a tiny piece of zinc and a few drops of concentrated hydrochloric acid. Visual change: deep blue color. Sensitivity: 0.03 μg tin.

Tungsten(VI) ions are precipitated as a white amorphous product with 2,4′-diaminodiphenyl hydrochloride. A drop of the test solution is mixed in a microtest tube with a 1% solution of the reagent in 2 N HCl. Sensitivity: 6 μg tungsten. Interference: molybdates disturb in concentrations larger than 10%.

Uranium(VI) ions fluoresce in crystalline form, but only slightly in solution. Sodium fluoride beads of uranyl salts are most striking and are used to detect uranium. Sodium fluoride is fused in a loop of platinum wire. A neutral test solution is placed on the bead and evaporated slowly. After fusing for a short time, the bead is cooled and examined in UV light. Visual change: deep yellow fluorescence. Sensitivity: 0.001 μg uranium. Interferences: only niobium and beryllium give a much weaker similar fluorescence.

Vanadium(V) as the vanadates forms a yellow precipitate in strongly acidic solution with α-benzoinoxime. Analogous white compounds are formed with the same reagent by molybdates and tungstates. Two drops of the saturated reagent solution in ethyl alcohol are added to a drop of the test solution followed by a drop of 3 N H_2SO_4. Sensitivity: 1 μg vanadium. It is a specific reaction, only the color of Fe^{3+} might interfere. This can be avoided by adding phosphate.

Zinc ions raise the oxidation potential of the ferricyanide by removing ferrocyanide ions from the redox equilibrium between ferricyanide and 3,3′-dimethylnaphthidine. A drop of the reagent solution (freshly prepared mixture of one part 5% solution of potassium ferricyanide in water and two parts saturated aqueous solution of 3,3′-dimethylnaphthidine chloride) and a drop of weakly acid test solution are brought together on a spot plate. Visual change: red-violet color. Sensitivity: 0.1 μg zinc.

Zirconium(IV) ions with morin form a pale-yellow, acid-resistant color in daylight and a yellow-green fluorescing lake in UV light. A drop of the test solution is mixed on the spot plate with a drop of 0.001% reagent solution in ethyl alcohol and some drops of concentrated HCl. Sensitivity: 0.1 μg zirconium. Interferences: it is a specific reaction but cupric, ferric, vanadate, chromate and noble-metal ions quench the fluorescence.

3.3 Selected Tests for Anions

If the original material does not consist solely of alkali salts, the sample is boiled with strong sodium carbonate solution or is fused with sodium potassium carbonate in a platinum spoon. Heavy-metal salts are thus converted in water-insoluble carbonates. The only soluble metal carbonate, except for the alkali carbonates, is Tl_2CO_3. Spot reactions to detect the various anions can be successfully made on drops of the sodium carbonate extract. Acetic acid is added to the carbonate extract until no further carbon dioxide evolves.

Borates can liberate free boric acid which gives a deep red compound when evaporated with tincture of curcuma (turmeric). The boric acid changes the yellow curcumin into the isomeric red-brown rosocyanine. A drop of the test solution acidified with HCl is placed on curcuma paper and dried at 100 °C. Visual change: a red-brown fleck, which turns blue to greenish-black on treatment with 1% NaOH. Sensitivity: 0.02 μg boron. Interferences: oxidizing substances (H_2O_2, CrO_4^{2-}, NO_2^-, ClO_3^- and iodides). They must be decomposed before carrying out the reaction.

Chloride ions are detected in the presence of other halide ions by selective oxidation of bromides and iodides and halogenating 8-hydroxyquinoline with the halogen. Chlorides are not oxidized and are precipitated with $AgNO_3$ solution. A drop of the test solution is warmed with a drop of 2% reagent in 1 : 4 acetic acid and a drop of H_2O_2 solution (2 parts of 6% H_2O_2 and 1 part dilute CH_3COOH). A drop of 1% $AgNO_3$ solution is added. Visual change: white precipitation.

Cyanide ions can be converted to thiocyanate, which can be identified by the sensitive iron reaction. A drop of the test solution is stirred with a drop of yellow ammonium sulfide on a watch glass and warmed until a rim of sulfur is formed around the liquid. After the addition of some

dilute HCl, 1–2 drops of 1% $FeCl_3$ are added. Visual change: red color. Sensitivity: 1 μg cyanide. Interference: in the presence of thiocyanates, the cyanide should be first isolated as $Zn(CN)_2$.

Ferrocyanide, thiocyanate and thiosulfate anions are detected simultaneously by capillary separation of their respective ferric compounds. Filter paper is impregnated with ferric chloride solution and dried. The paper is spotted with dilute HCl and then with the test solution. After a short time, individual colored areas are observed due to capillary separation. From the center outward they reveal: ferrocyanide, blue fleck; thiosulfate, white ring; thiocyanate, red ring. The sensitivities of the detections depend on the ratio of the three components but they are lower than 1 μg each.

Iodide ions react with palladous chloride to form a brown-black precipitate of PdI_2. The latter is insoluble in mineral acids but soluble in ammonia, NaCl and $MgCl_2$. A drop of the test solution is spotted on filter paper with a drop of 1% reagent Sensitivity: 1 μg iodide.

Nitrate ions oxidize diphenylamine to blue quinoid imonium dye. About 0.5 mL of the strongly acid diphenylamine solution is placed on a spot plate, and in the middle a drop of the test solution is added. Sensitivity: 0.5 μg nitrate. Interferences: alkali halogenates, perchlorates, periodates, permanganates, persulfates, peroxides and nitrites. The interferences of the six anions can be averted by drying the sample and heating the residue to 400°–500 °C. All but the nitrate and nitrite are decomposed. Sulfamic acid, on the other hand, destroys the nitrite in advance.

Nitrite ions form diazonium cations with primary aromatic amines in acid solution which couple with the same (or with another) primary aromatic compound and result in colored azo compounds. A drop of the test solution is mixed on a spot plate with a drop each of 1% sulfanilic acid in 30% acetic acid and a 0.03% solution of α-naphthylamine in 30% acetic acid solution Visual change: red color. Sensitivity: 0.01 μg nitrite.

Phosphate ions react with molybdates forming complex phosphomolybdic acid. The latter has an enhanced oxidizing power, oxidizing tetrabase to its blue quinoidal compound. A drop of the acid test solution is placed on the quantitative filter paper followed by a drop of ammonium molybdate solution (5 g salt dissolved in 100 mL water and poured into 35 mL HNO_3 specific gravity 1.2) and a drop of saturated tetrabase solution in 2 N CH_3COOH. Sensitivity: 0.05 μg phosphate. Interferences: germanic, arsenic and silicic acids behave analogously.

Silicate ions form the complex silicomolybdic acid H_2SiO_4, $12MoO_3$ when treated with molybdate. It is reduced by sodium stannite solution to molybdenum blue. Three drops of ammonium molybdate solution is placed on a watch glass followed by a drop of the test solution and a drop of alkali stannite solution Visual change: blue color. Sensitivity: 1 μg SiO_2. Interferences: phosphoric and molybdic acid.

Sulfate ions decompose the red-brown precipitate of barium rhodizonate by forming insoluble $BaSO_4$. A drop of $BaCl_2$ is placed on filter paper followed by a drop of freshly prepared 0.2% solution of sodium rhodizonate. The red fleck is treated with a drop of the test solution Visual change: the red fleck disappears. Sensitivity: 5 μg SO_4^{2-}.

Sulfide ions form methylene blue with *p*-aminodimethylaniline and $FeCl_3$. H_2S is liberated from the test solution and absorbed by glass fiber paper moistened with dilute NaOH. Add a mixture of a drop each of concentrated HCl, 0.1 N $FeCl_3$ and several granules of *p*-aminodimethylaniline. Visual change: blue. Sensitivity: 1 μg sulfide.

4 SELECTED TESTS IN ORGANIC ANALYSIS

The main field of qualitative organic analysis is in the detection of certain groups of compounds, the detection of characteristic types of compounds and, whenever possible, the identification of individual compounds. These goals are mainly attained through observation of the results of the chemical reactions into which the functional group enters. Such reactions include chelate formation, catalyzed and induced reactions, masking procedures, solid–solid reactions at elevated temperatures, reactions in the gas phase through contact with solid or dissolved reagents and reactions which yield fluorescent or colored products or those which fluoresce.

These qualitative tests[2,4] are instrumentally unsophisticated but their use has considerable practical importance. Analytical problems seldom involve totally unknown materials. The information available concerning the origin, method of preparation, color of the sample and so on, almost always gives valuable clues as to the direction of the spot-test examination. The analytical problem is not always to detect a particular organic compound, but rather to find out whether members of the class are present or absent.

4.1 Preliminary Tests

4.1.1 Sensory Tests (Color and Odor)

As most pure organic compounds are colorless and odorless, the appearance of color and odor has some limited diagnostic value. Yellow material can indicate nitro, nitrozo and azo compounds, and a shift to a longer wavelength might suggest the presence of conjugation, chelation or dyestuffs. Fluorescence is indicative also,

but only in material carefully purified by sublimation or recrystallization.

The detection of odoriferous materials such as menthol, phenol, pyridine, butyric acid and vanilin can be extremely sensitive. Nitrobenzene and benzaldehyde have an almond odor, ethyl sulfide has a garlic and long-chained fatty acids have characteristic rancid odors.

4.1.2 Burning Tests

Different visual phenomena appear by ignition of various classes of organic compounds under similar conditions. Burning aromatic compounds and halogen compounds furnish a smoky flame whereas the lower aliphatic compounds burn with a smokeless flame. Compounds containing a high percentage of halogen do not ignite easily and oxygen-containing compounds show a bluish flame. Proteins and carbohydrates burn with an odor resembling burnt hair.

4.1.3 Pyrolysis

Rapid external heating generally leads to the formation of lower-molecular-weight compounds, which can be detected in the gas phase. The latter are HCN, $(CN)_2$, hydrogen halides, acetaldehyde, CO, SO_2, nitrous oxide, phenols and so on. HCN appears in the gas phase by pyrolysis of nitrogenous organic material. The detection of $(CN)_2$ is almost specific for uric acid and purine derivatives by pyrolysis. Hydrogen halides are released by aromatic and aliphatic halogen compounds, and nitrogen–oxygen containing compounds always split off nitrous oxide. Aromatic compounds containing oxygen atoms release phenol on strong pyrolysis.

4.1.4 Acid–Base and Redox Behavior

The basic or acidic behavior of a material can afford a valuable clue in the detection of basic or acidic functional groups. The derivatives of ammonia, hydroxylamine and hydrazine are bases, whereas functional groups which split hydrogen ions, such as carboxylic, sulfonic, sulfinic, arsonic, nitroxylic, oximic, thioenolic, phenolic and thiophenolic are detectable by acid character. For water-soluble compounds, conventional indicators would be useful, whereas for water-insoluble compounds, basic groups are detected by shifting of the reaction equilibrium of nickel with dimethylglyoxime. After removal of the precipitate, the equilibrium solution which results may react with the basic unknown and the consumption of protons will lead to the appearance of the red precipitate. On the other hand, extremely weak organic acids heated with prefused potassium iodate (which contains iodide) leads to iodine liberation.

Organic reductants are indicated by a molybdenum blue reaction, whereas a preliminary test with tetrabase reagent paper might detect the few classes of oxidizing compounds, such as polyhalides of organic bases, quinones and peroxoacids.

4.1.5 Elemental Analysis

The organic material is pyrolyzed with carbonates, strong oxidizing acids, oxidants such as vanadium pentoxide or heated in aluminum foil under a stream of oxygen. For details, consult Feigl and Anger.[1]

4.2 Selected Tests for Functional Groups

An organic molecule may include several of such groups and their detection may largely characterize its chemical and physical properties. The most common functional groups are discussed here.

Aromatic compounds react with concentrated H_2SO_4 containing formaldehyde and form red, green and violet precipitates. The reaction occurs only when the *para* position is free. This enables the formation of methylene compounds, which are partly oxidized to *p*-quinoidal compounds. A drop of the nonaqueous test solution or a little of the solid is treated on a spot plate with a drop of formaldehyde–sulfuric acid reagent (0.2 mL of 37% formaldehyde in 10 mL concentrated H_2SO_4). Sensitivity: 2–25 μg aromatics.

Primary, secondary and tertiary alcohols convert $[Ce(NO_3)_6]^{2-}$ into $[Ce(OR)(NO_3)_5]^{2-}$ with a color change from yellow to red. The reagent solution is 40% $[(NH_4)_2Ce(NO_3)_6]$ in 2 N HNO_3. One milliliter of the reagent solution diluted with 2 mL of dioxane is treated with the test solution. Sensitivity: 400 μg alcohol. Interferences: aliphatic bases and oxidizable compounds.

Secondary alcohols fused with sulfur release H_2S on heating. In a microtest tube, the test solution is treated with a drop of 2% sulfur solution in carbon disulfide and dried. The mouth of the tube is covered with a disc of lead acetate paper and placed in a glycerol bath previously heated to 150 °C. Visual change: appearance of black lead sulfide on the paper. Sensitivity: 10–200 μg alcohol.

Volatile phenols condense with 2,6-dichloroquinone-4-chloroimine to give colored indophenols. A drop of the test solution is dried in a test tube. The mouth of the tube is covered with a disc of filter paper impregnated with a saturated benzene solution of the reagent. The tube is placed in a glycerol bath preheated to 150 °C. The reagent paper is held in NH_3 vapors. Visual change: blue stain. Sensitivity: 0.3–20 μg phenol.

Phenols react with nitrous acid containing mercuric nitrate to form nitro compounds reacting with the phenol.

A drop of the test solution is mixed in a microcrucible with a drop of the reagent solution (one part mercury is dissolved in one part of fuming HNO_3 and diluted with two parts of water). Visual change: red color. Sensitivity: 0.5–10 μg phenol. Note: di-*o*- and di-*m*-substituted phenols do not react.

Aromatic and aliphatic aldehydes condense with primary aromatic amine forming colored Schiff bases. Dianisidine is an especially suitable reagent for this. A drop of the sample is mixed with 3–4 drops of the saturated solution of the reagent in glacial acetic acid. The light color is intensified on heating. Sensitivity: 0.05–50 μg aldehyde. Interference: large amounts of certain ketones.

Aromatic aldehydes condense with *N*,*N*-dimethyl-*p*-phenylenediamine to give an intensely red protonated Schiff base. Several milligrams of the solid oxalate or chloride of the reagent are added to a drop of the test solution in a microtest tube and warmed in a preheated water bath. Aliphatic aldehydes and ketones do not interfere. Sensitivity: 0.3–0.5 μg aldehyde.

Aliphatic ketones react with *m*-dinitro compounds in alkaline solution to yield violet products. The reaction is selective for aliphatic monoketones. A drop of a 1% solution of *m*-dinitrobenzene in alcohol is mixed in a microtest tube with a saturated methanolic solution of KOH and a drop of the test solution. The tube is warmed to 70–80 °C in a water bath for 2–3 min. Visual change: violet-red color.

Carboxylic acids are converted via acid chloride into hydroxamic acids. The latter react with ferric ions to give a red or bluish-red chelate. A drop of the test solution is evaporated to dryness in a microcrucible and treated with two drops of thionyl chloride. The mixture is evaporated almost to dryness. Two drops of saturated alcoholic solution of hydroxylamine hydrochloride are then added and made basic with alcoholic alkali. The mixture is acidified with 0.5 N HCl and treated with 1% aqueous $FeCl_3$ solution. Visual change: dark-violet. Sensitivity: 10–100 μg carboxylic acids. Interferences: citric and thioacetic acid.

Alkyl esters of carboxylic acids react with metallic sodium to form acyloins. The latter is identified by its reaction with 1,2-dinitrobenzene to produce the red acid form of nitrosonitrobenzene. The test is conducted in the depression of a spot plate. A small piece of metallic sodium (CARE) is pressed into a disk with a glass rod. A drop of a benzene solution of the test material is added and then a drop of 2.5% benzene solution of 1,2-dinitrobenzene. The system is stirred with a glass rod and a drop of water is introduced. Visual change: deep-violet color.

Ethers are oxidized at higher temperatures (230 °C) to etherperoxides which oxidize, via cupric peroxide, tetrabase to tetrabase blue. A drop of the test solution is placed in a microtest tube. The tube is plunged into a glycerol bath preheated to 230 °C. The mouth of the test tube is covered with a disc of filter paper moistened with a mixture of 5% aqueous copper ethylacetoacetate and 5% chloroform solution of tetrabase. Sensitivity: 20 μg diethylether.

Interferences: the test fails in the presence of benzene and CS_2. However, it is also given by dioxan and tetrahydrofuran; it seems that the $-CH_2-O-CH_2$ group is responsible for the auto-oxidation.

Aromatic primary amines react with $Na_2[Fe(CN)_5 \cdot H_2O]$ and produce bluish-green $Na_2[Fe(CN)_5ArNH_2]$. A drop of the test solution is mixed with a drop of 1% reagent and a drop of 2% Na_2CO_3 solution on a spot plate. Sensitivity: 0.2–50 μg amines.

Primary and secondary aliphatic and aromatic amines condense with *p*-dimethylaminocinnamaldehyde to yield colored Schiff bases. A drop of saturated solution of the reagent dissolved in methanol and trichloroacetic acid is placed on filter paper and treated with a drop of the test solution The filter paper is kept in an oven at 100 °C for about 3 min. Visual change: blue for primary and purple color for secondary amines. Sensitivity: 0.02–2.2 μg amines.

Primary and secondary aliphatic amines form dithiocarbamates with carbon disulfide instantly at room temperatures. The dithiocarbamates can be detected by the catalytic enhancement of the sodium azide–iodine reaction. A drop of the alcoholic test solution is mixed in a microcrucible with a few drops of a 1 : 1 mixture of alcohol : CS_2. The excess of CS_2 is volatilized after about 5 min. A few drops of the reagent (3% NaN_3 in 0.1 N iodine solution) is added. Visual change: the color of the iodide disappears and evolution of nitrogen bubbles is observed. Sensitivity: 1–250 μg amines. Interferences: mercaptans, thioketones and thiol compounds. Tertiary aliphatic amines do not react.

Aliphatic secondary amines form blue-violet compounds reacting with sodium nitroprusside and acetaldehyde in alkaline solution. Primary and tertiary amines do not interfere. A drop of the test solution is mixed with a drop of the reagent (freshly prepared 1% solution of sodium nitroprusside to which 10% acetaldehyde is added.) The mixture is made basic with a 2% solution of sodium carbonate. Visual change: blue-violet color. Sensitivity: 1–100 μg amine. Interference: diisopropylamine does not react.

Nitro compounds added to molten tetrabase (melting point 91 °C) lead to yellow-orange-red melts, molecular compounds are obtained between the two components. One drop of the ethereal or benzene solution of the sample and one drop of 5% benzene solution of the reagent are mixed in a microconical tube and dipped into boiling water. Visual change: after evaporation of the solvent, a

yellow-orange melt appears. Sensitivity: 0.3–50 μg nitro compound. Interference: *p*-quinoid compounds.

Primary aliphatic nitro compounds are pyrolytically reduced to nitrous acid when heated to 160 °C with benzoin. The benzene solution of the test material is brought together with some solid benzoin in a microtest tube. After solvent volatilization, the test tube is immersed in a glycerol bath preheated to 130 °C. The mouth of the test tube is covered with a filter paper moistened with Griess reagent (a mixture of 1% solution of sulfanilic acid and 0.1 aqueous solution of *N*-1-naphthylethylenediamine dihydrochloride) and the temperature raised to 160 °C. Visual change: red fleck on the reagent paper. Sensitivity: 10–20 μg nitro compounds.

Aromatic nitro compounds can be reduced with metallic zinc in weakly acidic solution to arylhydroxylamines. The latter form violet complexes with pentacyanoamine ferroate. A few milligrams of the sample are dissolved in hot alcohol in a test tube and treated with a few drops of 10% $CaCl_2$. About 50 mg of zinc dust is added and the suspension is heated in a strongly boiling water bath. The filtrate is cooled and 1 drop of 1% reagent solution is added. Visual change: purple, blue or green color. Sensitivity: 0.8–15 μg nitro compound.

Nitriles (cyanides) are converted to oxamide by heating them with hydrated oxalic acids (pyrohydrolysis). The resulting oxamide is detected by its sintering with thiobarbituric acid. A drop of the test solution is dried in a test tube, a few milligrams of solid thiobarbituric acid are added and the mixture is heated to 140°–160 °C in a glycerol bath. Visual change: orange product. Interferences: ammonium salts, amides of acids, amino acids, ureides. Interference is avoided by selective dissolution of nitriles in benzene. Sensitivity: 2–20 μg nitriles.

Aliphatic nitriles release HCN pyrolytically, when heated to 250 °C (differentiation from aromatic nitriles). A little of the solid test material is mixed with a few centigrams of a 1 : 1 mixture of calcium carbonate and oxide. The tube is placed in a glycerol bath preheated to 250 °C and the mouth of the tube is covered with a paper disc moistened with a mixture of 5% Cu ethylacetoacetate and 5% tetrabase in chloroform. Visual change: blue stain on the paper after 3–4 min. Sensitivity: 2–50 μg nitriles.

Thioketones and thiols accelerate the iodine–azide reaction catalytically, in contrast to other organic sulfur derivatives. A drop of a solution or the test substance in water or organic solvent is mixed on a watch glass with a drop of 3% NaN_3 in 0.1 N iodine solution Visual change: the brown color of the iodine disappears and bubbles of nitrogen evolve. Sensitivity: 3×10^{-4}–0.6 μg.

Primary and secondary thiols when heated with concentrated NH_3 at boiling water temperature release H_2S. A small amount of the test material or a drop of the solution is treated in a microtest tube with a drop of concentrated NH_3. The mouth of the tube is covered with wet lead acetate paper and the tube is placed in a boiling water bath. Visual change: black stain on the paper. Sensitivity: 1–200 μg. Tertiary thiols and thioketones do not interfere.

Sulfonic acids can be converted pyrohydrolytically to sulfuric acid which is instantly reduced to sulfurous acid.

$(R, Ar)\ SO_3H + H_2O \longrightarrow (R, Ar)\ H + H_2SO_4$. The pyrohydrolysis is accomplished by phthalic acid at 230 °C. A little of the solid or a drop of its solution is mixed with a few centigrams of phthalic acid and the solvent is removed by evaporation. The test tube is placed in a glycerol bath preheated to 230 °C. A disc of filter paper moistened with ferric ferricyanide solution (0.08% anhydrous $FeCl_3$ and 0.1% of $K_3Fe(CN)_6$ in water) is placed over the mouth of the test tube. Visual change: blue stain on the paper. Sensitivity: 1–20 μg sulfonic acids.

4.3 Detection of Individual Organic Compounds

The measurement of certain physical properties provides a very reliable method for the identification of an organic material. Determination of melting and boiling points, density of compounds and optical qualities such as refractive index, optical activity, UV and infrared (IR) absorption are very useful, but the essential requirement of isolating the compound in a state of a high purity is very costly with regard to time and material. For this reason rapid chemical identifications that do not require expensive apparatus are important. The possibility of finding a direct specific test for an individual compound is highly improbable but the available information about the origin, method of preparation and use of the material in question provide valuable clues. In fact, the analytical goal is often well defined by the question of whether particular compounds are present or absent. The following section contains descriptions of selected spot tests for technically important individual compounds.

Acetone condenses with salicylaldehyde to form a red conjugated compound. One drop of 10% alcoholic solution of the reagent is mixed in a test tube with one drop of 40% NaOH solution and a drop of the test solution The tube is kept for 2–3 min in a water bath at 70–80 °C. Visual change: deep-red color. Sensitivity: 0.2 μg acetone.

Acetic acid forms a dark-blue adsorption compound of iodine and basic lanthanum acetate when mixed with iodine and ammonia. The interference of many anions is avoided by the microdistillation of acetic acid in the presence of concentrated H_3PO_4. In the depression of a spot plate a drop of the solution is mixed with a drop of 5% solution of lanthanum nitrate, 0.01 N iodine solution and a drop of 1 N NH_3. Visual change: blue color in a few minutes. Sensitivity: 50 μg acetic acid.

Monochloroacetic acid heated with ammonium thiocyanate forms rhodamine, which subsequently, owing

to its reactive $-CH_2-$ group, reacts with 1,2 naphthoquinone-4-sulfonate. A colored condensation product results. A drop of the test solution is mixed with some solid NH_4SCN and heated in boiling water for 2 min. The cooled mixture is treated with 1–2 drops of 0.5% reagent and made alkaline with 1% NaOH solution Visual change: blue color. Sensitivity: 5 μg monochloroacetic acid.

Acetylene forms an orange precipitate with ammoniacal Ag_2CrO_4. It is an acid-insoluble addition compound $Ag_2C_2 \cdot Ag_2CrO_4$ which serves as a protective layer around silver chromate. A drop of ammoniacal silver chromate (Ag_2CrO_4 dissolved in 1 : 5 NH_3 and filtered) is placed in each of two small porcelain crucibles. One drop of the solution being tested for acetylene is added to one of the suspensions, and a drop of water to the other. The mixtures are stirred and treated with the same number of drops of 1 : 10 HNO_3 until the Ag_2CrO_4 precipitate in the blank disappears. Visual change: in the presence of acetylene a residue of red-brown Ag_2CrO_4 remains. Sensitivity: 1 μg acetylene.

Anthracene evaporated with concentrated HNO_3 is oxidized to anthraquinone, which can be detected by reductive conversion with solid sodium hydrosulfite in alkaline solution to the red water soluble salt of anthrohydroquinone. Sensitivity: 3 μg anthracene.

Ascorbic acid reduces ammonium phosphomolybdate to molybdenum blue. A drop of the test solution is placed on the reagent paper (filter paper immersed in a saturated alcoholic solution of phosphomolybdic acid, dried in cold air, bathed in concentrated Ag_2CrO_4 solution and dried again). Visible change: blue color. Sensitivity: 0.1 μg ascorbic acid. Interference of uric acid or urate is avoided by acidifying the sample.

Barbituric acid reacts instantaneously with nitrous acid to give violuric acid. A drop of the test solution is mixed in the depression of a spot plate with one drop each of 2 N CH_3COOH and saturated $NaNO_2$ solution. Visual change: red violet. Sensitivity: 10 μg barbituric acid.

Carbon disulfide forms dithiocarbamates with secondary aliphatic amines, which in turn produce chelates with transition-metal ions. In a microtest tube a drop of the test solution is treated with a drop of reagent solution (9 mL 0.2% $CuSO_4$ solution mixed with 1 mL concentrated NH_3 and 0.5 g dimethylamine chloride.) The tube is shaken with 2 drops of benzene. Visual change: yellow-brown color in the benzene layer. Sensitivity: 3 μg carbon disulfide.

Carbon tetrachloride vapor releases chlorine when heated in the presence of quartz sand. It is most likely that a disproportionation occurs on the surface of the quartz. In a microtest tube a drop of the test solution is mixed with several centigrams of quartz dust. The tube is immersed in a glycerol bath preheated to 150 °C and the mouth of the tube is covered with a disc of Congo Red paper moistened with 5% H_2O_2 solution. Visual change: blue stain. Sensitivity: 5 μg CCl_4. Interference: *sym*-dichloro(bromo)ethane and *sym*-tetrachloro(bromo)ethane show the same behavior.

Citric acid is converted by alkali permanganate into acetone dicarboxylic acid, which in the presence of bromine precipitates as pentabromoacetone. A drop of the test solution is mixed in a microtest tube with one drop each of 0.01 N $KMnO_4$ and a drop of saturated bromine water. After cooling, the excess bromine and MnO_2 are destroyed with the addition of some solid sulfosalycilic acid. Visual change: white precipitate. Sensitivity: 6 μg citric acid.

Formaldehyde condenses with J-acid dissolved in concentrated H_2SO_4 to produce a yellow fluorescent xanthylium dyestuff. Add one drop of aqueous test solution to a drop of a 0.1% solution of J acid (6 amino-1-naphthol-3-sulfonic acid in concentrated H_2SO_4) on a glass filter paper. Visual change: yellow fluorescence under a UV lamp after addition of a drop of water, the yellow solution turns blue. Sensitivity: 0.01 μg formaldehyde.

Glycerol is dehydrated with concentrated H_2SO_4 to acrolein, which according to the classical Skraup synthesis forms 8-hydroxyquinoline with *o*-aminophenol. The magnesium salt of 8-hydroxyquinoline is highly fluorescent. Two drops of 2% alcoholic solution of *o*-aminophenol are evaporated in a microtest tube. A drop of the test solution and 4 drops 1% arsenic acid in concentrated H_2SO_4 are added. The tube is kept at 140 °C for about 15 min, and after cooling 5 drops of concentrated NaOH, 1 drop of 2 N $MgSO_4$ and three drops of concentrated NH_3 are added. Visual change: bluish-green fluorescence. Sensitivity: 0.5 μg glycerol. Interference: crotonaldehyde.

Malonic acid fused with urea forms barbituric acid. The latter is identified by its reaction with pyridylpyridinium dichloride to give a pentamethine dye. A drop of the test solution is treated with several drops of a saturated methanolic solution of urea in a microtest tube. The tube is heated for several minutes in a glycerol bath at 120 °C. A drop of 1% solution of pyridylpyridinium dichloride in dimethylformamide (DMF) is added. The mixture is heated to 120 °C for several minutes. Visual change: reddish-blue color develops on cooling, which shows a red fluorescence red under UV. Sensitivity: 50 μg malonic acid. The detection is specific.

Methanol is oxidized by acidic permanganate to formaldehyde. The latter is detected specifically by forming a bright violet *p*-quinoidal compound with chromotropic acid. A drop of the test solution is mixed with a drop of 5% H_3PO_3 and a drop of 5% $KMnO_4$. The mixture is decolorized by sodium bisulfate crystals and 4 mL of 12 N H_2SO_4. A little finely powdered

chromotropic acid is added and heated to 60 °C for 10 min. Visual change: violet color. Sensitivity: 3.5 μg methanol.

Naphthalene vapor reacts selectivity with chloranil to give a brick-red addition compound. The solid unknown or its ethereal solution is placed in a microtest tube and the mouth of the tube is covered with a disc of filter paper impregnated with a saturated ethereal solution of chloranil. The tube is immersed in a water bath at 60 °C. Visual change: brick-red color on the paper. Sensitivity: 25 μg of naphthalene.

Oxalic acid decomposes pyrolitically at 250 °C to give formic acid, which forms aniline blue with diphenylamine. In a microtest tube a drop of the sample is evaporated and melted with a little diphenylamine over a free microflame. After cooling, a drop of alcohol is added. Visual change: blue coloration. Sensitivity: 5 μg oxalic acid.

Piperidine reacts through its reactive NH_2 group with 1,2-naphthoquinone-4-sulfonate to produce intensely colored *p*-quinoidal condensation products. In a microtest tube a drop of the acidic test solution is mixed with several centigrams of Na_2CO_3. The tube is immersed in a water bath heated to 90 °C and the mouth of the tube is covered with a disc of filter paper moistened with a 2.5% aqueous solution of the reagent Visual change: red stain on the yellow paper. Under the conditions described piperazine does not disturb the detection. Sensitivity: 20 μg piperidine.

Pyridine reacts with trisodium pentacyanoamminoferrate, by replacement of the ammonia with pyridine in the coordination sphere, forming a colored compound. On a Whatman filter paper No.1, a drop of the test solution is sprayed with one drop of 0.2% solution of the reagent in buffer solution at pH 6.5. Visual change: yellow-orange coloration. Sensitivity: 10 μg pyridine.

Quinones give colored condensation products with rhodamine. A drop of the test solution is shaken with one drop of a saturated aqueous solution of the reagent and one drop of ammonia. Visual change: green or blue with *p*-quinones and violet-red with *o*-quinones. Sensitivity: 0.2–5 μg quinones.

Salicylic acid is converted easily to fluorescent alkali salicylate with dissolved alkali and alkaline-earth salts. This acid can be separated adequately by sublimation on heating the substance with concentrated H_2SO_4. A drop of the test solution is evaporated in a microtest tube. A drop of concentrated H_2SO_4 is then added and the mouth of the tube is covered with a glass-paper disc wetted with 3% KOH. The tube is heated in a glycerol bath at 130 °C. Visual change: violet fluorescence on the glass paper. Sensitivity: 5 μg salicylic acid.

Tartaric acid heated with concentrated H_2SO_4 containing β,β′-dinaphthol produces green fluorescence. The test solution is treated in a test tube with a few milliliters of 0.05% concentrated H_2SO_4 solution of the reagent and kept for 30 min in a water bath at 85 °C. Visual change: luminous green fluorescence. Sensitivity: 5 μg tartaric acid.

Thiophene condenses with ninhydrin in concentrated H_2SO_4 to give a violet dyestuff. In a depression of a spot plate one drop of freshly prepared 0.01% reagent in concentrated H_2SO_4 is covered with a drop of the test solution Visual change: deep violet to pink color. Sensitivity: 5 μg thiophene.

Urea is quickly hydrolyzed enzymatically to ammonia even at room temperature. A drop of the neutral or alkaline test solution is treated in the depression of a spot plate with several milligrams of urease and after 2–5 min one drop of Nessler solution (50 g KI dissolved in 35 mL water is treated with saturated $HgCl_2$ solution until a slight precipitate persists. Then 400 mL of 9 N caustic alkali is added. The solution is diluted to 1000 mL, allowed to settle and decanted). Visual change: brown precipitate. Sensitivity: 1 μg urea Interference: ammonium salts. These can be eliminated by drying the test solution with a drop of dilute alkali.

5 SELECTED EXAMPLES OF APPLICATIONS OF SPOT TESTS

The classical spot tests have been successfully applied to solving large-scale analytical problems in clinical analysis, in crime laboratories, in judicial chemical studies, in geochemical prospecting, in soil and plant tissue testing, and in air and water QC. Commercial companies are marketing a great variety of compact test systems. Their extreme simplicity, time and money-saving nature and ultimate reliability make these tests very useful. These tests systems, marketed generally by commercial companies in the form of reagent strip systems, prepared reagent pillows or tablets constitute a significant proportion of the tools used for solving the total spectrum of analytical problems.

5.1 Applications in Clinical Analysis

Substantial changes in the concentrations of glucose, protein, ketone bodies, bilirubin, urubilinogen nitrites, leucocytes and erythrocytes, and so on in urine may have significant diagnostic value.[5] A simple spot test can give valuable quick information about many health anomalies and provides the possibility of performing many rapid bedside methods for some key materials in the human blood. Selected examples only are presented here.

5.1.1 Tests in Urine

Glucose in urine is screened quickly by enzymatic tests based on the activity of the enzyme glucose oxidase,

which uses dry-reagent chemical technology (Ames Co., Boehringer-Mannheim, Hoechst AG). A firm plastic strip, to which a stiff absorbent cellulose area is affixed, is impregnated with a buffered mixture of glucose oxidase, peroxidase and *o*-tolidine. In the first stage of the reaction glucose is oxidized by atmospheric oxygen in the presence of glucose oxidase to gluconic acid and hydrogen peroxide, while in the second stage H_2O_2/peroxidase oxidizes *o*-tolidine to a blue quinoidal compound. The detection is either visual or by simple reflectance instruments, irradiating the dipstick with polychromatic light and measuring the reflected radiation.

Protein in urine test is based ''on the protein error''. The test exploits the fact that the aqueous solution of the potassium salt of tetrabromophenol-phthalein ethyl ester becomes yellow on addition of dilute acetic acid. If, however, a solution or suspension of native albumin is added to a dilute blue solution, a great deal of acetic acid can be introduced without causing the blue to change to yellow (Ames Co., Boehringer-Mannheim). These standardized tests for the estimation of protein in urine are composed of a strip of absorbent cellulose, one end of which is impregnated with the indicator and buffered at pH 3.

Ketone bodies in urine are quickly screened by a dipstick utilizing the reaction of ketones with nitroprusside. The NO group of the nitroprusside gives isonitrosoketone, which remains in the complex anion, with the ketone whereas the Fe^{3+} is reduced to Fe^{2+} (Ames Co., Boehringer-Mannheim, Hoechst AG). A microprocessor-controlled benchtop reflectance monitor (Kyoto) is designed to measure the change in the color of the strips.

Bilirubin in urine can be screened with a tablet-configuration reagent system involving a diazotization procedure, which couples with the bilirubin to give a characteristic purple color. It is a fiber mat test (Ames Co.) using a tablet composed of *p*-nitrobenzenediazonium *p*-toluenesulfonate, NH_4HCO_3, sulfosalycilic acid and boric acid. The sensitivity of the reaction coincides with the lower limit of accepted pathological significance.

Urubilinogen in urine is tested by a dip-and-read solid state reagent incorporated in a *p*-dimethylaminobenzaldehyde and an acid buffer-containing strip. Urubilinogen forms a red Schiff base (Ames Co.)

Nitrite in urine tests with the dipstick technique are based on the classical diazotization Griess method (Ames Co., Boehringer-Mannheim, Hoechst AG, Merck Co.) The tests reveal significant bacteriuria when the bacterial count reaches 1×10^7 per mL of urine.

Occult blood in urine is detected on the basis of peroxidase-like activity of hemoglobin, myoglobin and some of their degradation products. The pseudoperoxidase activity of hemoglobin transfers an oxygen atom from the peroxide to the chromogen (*o*-tolidine, *o*-dianisidine, 2,6-dichlorophenolindophenol.) Typical dip-and-read spot tests for occult blood (Ames Co., Boehringer-Mannheim, Helena Laboratories, Finnpipet, Smith Kline Diagnostics) consist of absorbent cellulose impregnated with an organic peroxide, *o*-tolidine and buffer system.

Hemoglobin in urine is detected with a test strip coated with an organic hydroperoxide. The test strip can be used for the detection of peroxidase-like substances.

Phenylketones in urine are detected by a stiff strip of cellulose impregnated with iron(III) and magnesium ions. The desired acidity (pH = 2.3) is provided by the presence of cyclohexylsulfuric acid. The ferric ion forms colored chelates with phenylpyruvic acid, whereas the magnesium ions are masking agents to minimize interference in the color development by urinary phosphate (Ames Co.)

Chloride in urine is estimated by discharging the red color of silver dichromate by chloride ion. A solid-reagent spot test for chloride is the Quantab chloride titrator (Ames Co.) It is a chemically inert plastic strip laminated within, which is a capillary column impregnated with $Ag_2Cr_2O_7$. The reaction of $Ag_2Cr_2O_7$ with chloride produces a white color in the capillary column.

Ethanol in urine can be estimated by test strips containing alcohol oxidase and horseradish peroxidase with colorimetric reagents (3-methylbenzothiazolinone hydrazone hydrochloride and 3-dimethylaminobenzoic acid). The urine sample is placed in a centrifuge tube and the test strip is suspended in the air space above the sample. The stoppered tube is heated at 65 °C for 5 min and the color development is estimated from a color chart.

Ascorbic acid in urine is screened by the reductive decolorization of 2,6-dichlorophenolindophenol. The reagent paper set is prepared by impregnating and drying filter paper in 0.1%, 0.05% and 0.0025% alcoholic reagent solutions. They are suitable for detecting 0.4–1 mg mL^{-1}, 0.2–0.44 mg mL^{-1} and 0.01–0.24 mg mL^{-1} ascorbic acid, respectively.

Cysteine in urine can be tested semiquantitatively based on the reaction between cysteine, sodium nitroprusside and K_2CO_3 to give a cherry-red complex. Limit of detection: 0.003–0.12 mM cysteine.

Ionic strength and specific gravity in urine can be tested using a strip prepared by soaking a filter paper in a solution of sodium heparin and 1 mM methylene blue and attaching the treated sample to a plastic holder. The density of the developed color correlates positively with the specific gravity and ionic strength of the urine sample.

Edetates in urine screening is based on the complexing action of edetate anion to remove ferric ion from ferrioxamine B immobilized on a Dowex HRC(H^+) resin.

Decomposition of the colored chelate leads to a gradual fading of the deep brown-violet resin beads.

5.1.2 Tests in Blood

Glucose in blood screening is based on the specific glucose–oxidase–peroxidase reaction. In a reaction catalyzed by glucose oxidase, D-glucose is oxidized by atmospheric oxygen to β-D-gluconolactone and H_2O_2. The latter oxidizes the chromogenic system (Ames Co., Boehringer-Mannheim, Biodynamic, Inc.) The test consists of a firm plastic strip to which an impregnated reagent area is affixed. A semipermeable membrane serves as a barrier that retains the high-molecular-weight and cellular components on its surface and can be wiped clean after exactly 60 s. Although significantly hypoglycemic results can be rapidly recognized with the visual test, small deviations from normal values should be confirmed by simple reflectance measurements.

Urea in blood screening is based on the principle of catalyzed hydrolysis of the urea to carbon dioxide and ammonium hydroxide, the concentration of which is measured by the color change of the indicator (Ames Co., Boehringer-Mannheim, E. Merck Co.). The test stick is coated with a membrane which is freely permeable to urea but prevents staining of the paper by blood pigments.

Lactate screening in whole blood is as follows. Apply 20 μL whole blood to a filter paper strip coated on its other side with 3.65 μmol min^{-1} enzyme activity Pediococcus lactase oxidase, 0.12 μmol 4-aminoantipyrine, 0.2 μmol 3-(*N*-ethyl-*m*-toluidine)-2-hydroxypropane-1′-sulfonic acid and 18.4 μmol min^{-1} enzyme activity horseradish peroxidase and attached at one end of a supporting strip. After 1 min the absorbance is measured at 560 nm by a reflectometer.

Cholesterol in blood can be quickly analyzed by use of the cassette strip system, which is based on enzyme action on cholesterol to produce H_2O_2 and determination of the latter (Diagnostics, Inc.). The enzyme is cholesterol oxidase and the strip contains horseradish peroxidase and leuco dye.

5.1.3 Other Tests

Occult blood tests in feces depend on the estimation of peroxidase activity as an indication of hemoglobin content. Reagents used include guaiac, *o*-tolidine, *o*-dianisidine and 2,6-dichlorophenolindophenol. The products of Smith Kline Diagnostics and Finnpipette are based on the gauiacum resin, whereas other self-contained systems contain *o*-tolidine which is oxidized by the peroxidase-like reaction. The reagent tablets (Ames Co. Stoke Pages) contain *o*-tolidine, strontium peroxide, tartaric acid and calcium acetate. In contact with water, tartaric acid, calcium acetate and strontium peroxide instantly produce H_2O_2, which in the presence of hemoglobin oxidizes *o*-tolidine to a blue compound.

Spot tests are performed for *urinary calculi*, the most common varieties of which are the calcium oxalate, uric acid, ammonium urate and magnesium ammoniumphosphate calculi. The test element for oxalate determination contains oxalate oxidase, peroxidase and optional riboflavin (reaction promoter) and quinolines (enhancer). A filter paper is soaked in the first solution containing oxalate oxidase, peroxidase, riboflavin, citrate buffer and sodium alginate and then in the second solution containing tetramethylbenzidine (color indicator) and 3-aminoquinoline. Uric acid and urates are identified by the classical murexide test, heating the sample with a number of drops of concentrated HNO_3. The resulting red murexide indicates the presence of urates. Phosphates in calculi are detected by heating the sample with concentrated HNO_3 and $(NH_4)_2$ MoO_4 to produce yellow ammonium phosphomolybdate.

Urinary lactic acid screening is based on the transformation of the acid to acetaldehyde, which reacts with sodium nitroprusside to produce a blue color. A sample is added to a tablet, which contains 30 mg $CaCl_2$ and 30 mg Ce_2SO_4 in a test tube. The test tube is covered with a piece of filter paper that has been impregnated with a 2.5% aqueous sodium nitroprusside solution and a drop of 10% aqueous morpholine solution. Lactic acid is indicated by the appearance of a blue color.

Sweat test for *cystic fibrosis* (CF) is based on the elevated concentration of chlorides characterizing this genetic disease. An indicator system for CF consists of three filter paper strips impregnated with 0.1 M K_2CrO_7. The dry strips are soaked in 0.05 N, 0.066 N and 0.1 N $AgNO_3$ solution, respectively. The indicator system is attached to a suitable part of the body of the child for 15 min. The indicator strip is removed and read. A positive result is signaled by the decoloration of the originally red-brown papers. Zero or one indicator showing a positive reaction means that the sweat chloride ion concentration is $\leq$35 equivalents per liter and the child is considered normal. Children showing chloride concentration of 35–55 equivalents per liter (two of three papers are reacting positively) need to be checked more thoroughly and concentrations of more than 55 equivalents per liter chloride (all three indicators reacting positively) indicate a high probability of CF.

Urine leucocyte reagent strip (Boehringer-Mannheim, Shiojiri) detects leucocytes in urine. The detection is based on the esterase activity of the leucocytes. Upon contact between the reagent matrix and leucocytes in urine, the indoxyl ester substrate is hydrolyzed by the esterase to indoxyl, which then couples with a diazonium salt to produce a purple azo dye.

5.2 Forensic Application of Spot-test Analysis

Inexpensive screening by spot-test analysis has considerable importance, among other disciplines, in a modern police laboratory,[3,6–9] mainly as quick first orientation. Only selected examples are described here.

5.2.1 Firearm Discharge Residue Tests

These search mainly for the presence of antimony, barium nitrates, nitrites and lead as residues after shooting. The three elements can be identified by direct test on the cloth used to remove the residues. The cotton cloth is dried and a drop of 10% alcoholic solution of triphenylmethylarsonium iodide is added. In the presence of antimony an orange ring develops within 2 min. After drying a drop of freshly prepared saturated aqueous solution of sodium rhodizonate is added. A red color within the orange ring indicates the presence of barium, lead or a combination of both. After repeated drying the red area is spotted with 1 : 20 HCl. The appearance of a blue spot indicates lead. A reddish-brown spot turning vivid red confirms barium. Nitrate and nitrite detection on the hand is achieved by removing traces from the fingers by hot filter paraffin. After solidification, diphenylamine reagent (10 mL concentrated H_2SO_4 + 2 mL H_2O + 0.05 g diphenylamine) is added to the cast. A positive reaction is shown by the appearance of dark-blue particles. Although the reaction is not specific, it is accepted as corroborative evidence for the firing of a gun.

Since the color of the blood and the red azo dye produced by the classical Griess reaction for nitrite are similar, nitrite is detected in this case by fluorescence under UV radiation. A Whatman No.1 filter paper is soaked in freshly prepared 5% solution of *o*-phenylenediamine dihydrochloride and partially dried. This paper is placed over the suspected area of the gunshot on the clothing. The paper is then placed in a Petri dish containing a 25% NH_3 solution The paper is illuminated with 254 nm radiation. The yellow fluorescence pattern may be photographed as permanent evidence.

5.2.2 Some Quick Field Tests for Drug Detection

As the problem of drug abuse reaches frightening proportions, the use of simple screening tests are gaining considerable importance beside the enormously useful and sophisticated instrumental methods. Only some representative examples are described here.

Morphine can be detected by an internally referenced test strip immunoassay. The test strips comprise two active surfaces, each of which contains co-immobilized glucose oxidase and antibody against morphine; the color that develops is inhibited by the presence of morphine in the sample. The reference surface contains antiperoxidase and it is used to set the assay detection limit. To determine morphine in urine, the test strip is immersed for 1 min in the sample, then for 10 min in a ''developer'' solution containing morphine-bound peroxidase, glucose and 4-chloro-1-naphthol as a chromogenic substrate for peroxidase.

Barbiturates in blood and urine samples may be screened by thin-layer chromatography (TLC) on Silica Gel G with $CHCl_3$–acetone (9 : 1) or benzene–acetone (17 : 3) as mobile phase. Visualization is effected by spraying with diphenylcarbazone–$HgCl_2$ reagent followed by 1% methanolic KOH.

LSD (lysergic acid diethylamide) is identified by its reaction with Ehrlich reagent (also referred to as Van Urk reagent) producing a purple to deep-blue color.

Amphetamines in urine are sensitively (>0.1 μg) detected by a strip prepared using the Simon reagent. A 2% sodium nitroprusside and 5% aqueous Na_2CO_3 solution is dried for 3 h at 110 °C. A strip is cut from the plate to use in the test. Five milliliters of alkaline urine is extracted with 2 mL of benzene and a portion of the benzene layer is spotted onto the strip. When exposed to acetaldehyde vapor an intense blue color results.

Marijuana and coca leaf in illicit cigarette samples can be revealed by shaking the suspected material with a chloroform solution of Fast Bordeaux Gp. salts (C.I. Azoic Diazo Component 1) and with 0.5 M NaOH solution. Cannabinoids are indicated by the appearance of red color in the chloroform layer. The test is applicable to urine samples too.

Salicylates in overdoses are commonly detected in urine by the classical violet iron(III)-salicylate chelate formation. 0.3 mL of blood is mixed with 0.6 mL of methanol and centrifuged. The supernatant solution is filtered and 50 μL of 0.5% $FeCl_3$ solution is added to 0.2 mL of the filtrate. A violet color at the diffuse interface indicates the presence of salicylate.

5.2.3 Chemical Testing for Blood

Chemical testing is needed to provide physical evidence in cases in which direct visual inspection is not a decisive proof. Chemical spot tests can verify the presence or absence of blood stains, when blood spots on a dark background are not visible or a stain may resemble a bloodstain. In most tests the blood is identified on the basis of the peroxidase-like activity of the heme group of hemoglobin. Free hydroxy ions liberated by this activity oxidize colorless phenolphthalein (reduced phenolphthalein with metallic zinc in basic solution). Addition of a drop of 3% H_2O_2 solution produces a pink color in the presence of blood traces.

5.2.4 Detection of Semen

Rapid identification of semen is necessitated by the high incidence of sex crimes. Semen is generally identified by its high acid phosphatase content, which hydrolyzes aromatic orthophosphoric acid esters; the phenol resulting from the enzyme-catalyzed reaction can be detected with the Folin–Ciocalteu reagent.

The suspected stain is cut and incubated for 30 min at 37 °C with two drops of an 0.1% aqueous solution of disodium phenylphosphate and two drops of Folin–Ciocalteu reagent is added. Phenol forms molybdenum blue on reaction with the molybdophosphate incorporated in the reagent.

5.2.5 Detection of Saliva

Saliva occurrence may be related to spittle, or may be detected on objects in contact with the mouth. Two-thirds of the solid content (0.6%) is salivary amylase. This enzyme, also called ptyalin or diastase hydrolyzes starch to sugar or insoluble cross-linked blue starch polymer to soluble blue fragments. A cutting from the suspected saliva stain is placed in a tube with 1 mL distilled water. Some drops of slurry are made by suspending a reagent tablet in 5 mL water which contains the cross-linked starch polymer (Pharmacia Laboratories, Inc.). After 30 min incubation at 37 °C the enzyme action is stopped by adding 1 mL of 0.5 M NaOH solution and the mixture is centrifuged. The appearance of a blue color in the supernatant fluid indicates the presence of saliva.

5.3 Selected Geochemical Applications of Spot Tests

The aim of geochemical prospecting is to find new deposits of metals and nonmetals and to detect crude oil and natural gas accumulation. Chemical tests may indicate the extension of existing deposits. Most of the analytical methods used in geochemistry are instrumental, but because a precision of 10% at the 95% confidence level is good enough for geochemical purposes, at least in the initial stages of the project, simple colorimetric screening tests are still applied, which can be carried out in the field by relatively unskilled operators.[10] Very low capital cost is associated with these methods and the operators can be trained in a relatively short time. Many tests applicable to geochemical exploration are described in Feigl and Anger[1] and Jungreis.[2]

In semiquantitative field tests calibrated sampling spoons and plastic standards, which are prepared by adding suitable dyes to a transparent polyester resin, are used. These standards are far more stable than the conventional liquid standards. No time-consuming preparation of a standard series in the field is necessary and the hardened plastic endures rough handling.

5.3.1 Some Examples in Geochemical Screening

Differentiation between calcite and dolomite is based on the different behaviors of the two carbonates towards weak acids. Calcite is attacked by dilute acid at room temperature. Some crystals of the sample material are placed in a depression of a spot plate. One drop of 0.1% alizarin S solution in saturated tartaric acid solution is added. Calcite is indicated by the appearance of a purple-red precipitate. Dolomite does not react.

A spot test for *Beryllium* in minerals and ores is based on the conversion of all beryllium compounds to the soluble alkali beryllium fluoride, which reacts with ammoniacal quinalizarine solution and forms a blue-violet complex. The test is specific. The sample is fused in a platinum spoon with an excess of KHF_2. The residue is put in some cold water and centrifuged. A drop of the supernatant fluid is treated with a drop of 0.05% quinalizarine solution in 10% NH_3. A few drops of bromine water destroys the violet color of the excess reagent.

A test for *chromium* in rocks is accomplished by grinding a tiny particle to a fine powder in an agate mortar and mixing with a four-fold excess of 1 : 1 mixture of K_2CO_3 and Na_2O_2. A red-hot platinum wire is dipped in the mixture to allow bead formation. The cold bead is dissolved in 1 : 1 H_2SO_4 in the depression of a spot plate and a drop of 1% alcoholic diphenylcarbazide solution is added. A violet color appears even when the rocks contain Cr in concentrations as low as 0.04%.

Lead in ores and minerals is directly estimated even in the most water-insoluble ores with the rhodizonate reaction. In the presence of Ba or Sr the sample must be fumed with H_2SO_4 beforehand. In a depression in a spot plate a drop of pH 2.8 solution and a drop of freshly prepared 0.2% reagent solution is added. A red-violet color appears on the surface of lead-containing ores. Galena, cerussite, anglesite, crocoite and all water-insoluble lead ores give positive results.

Molybdenum in ores is identified extremely sensitively by the specific color reaction with potassium xanthate. A few grains of the ore in a small crucible are digested with concentrated NaOH and acidified with H_3PO_4. Add a few crystals of the reagent. A dark-violet colored compound $MoO_3 \cdot 2[SC(SH)(OC_2H_5)]$ forms in the presence of Mo.

5.4 Rapid Selected Screening Tests for Soil and Plant Tissues[3]

The simplification of complex soil chemical analyses became necessary to make them performable by minimally trained personnel. The screening of soil quality is required in conjunction with testing of plant sap.

A test of *soil aeration* shows the oxygen concentration. The soil is sampled with an open-faced sampler and

a $CaCO_3$ suspension is added dropwise to its surface. The extent of penetration of the white chalk particles is proportional to the degree of aeration.

A test of *soil pH* is accomplished with a mixed indicator of bromocresol green, bromocresol purple and bromocresol red (0.05%, 0.1% and 0.02%, respectively.) covering a pH range of 4.0–7.5 (Hellige Inc.) Some dried soil grains are dispersed on a white spot plate and some drops of the indicator mixture are added. Purified barite mineral masks the color of the soil.

A test for *exchangeable potassium* is necessary to decide whether additional potassium fertilizer is needed. A portable soil test laboratory (Hach Co.) estimates K content using powder pillows containing tetraphenylborate, which precipitates the ion. Because heavy-metal ions interfere with the test a second powder pillow is included that contains the appropriate complex-binding agents.

A rapid test for *nitrate* is accomplished by adding a noncorrosive stable reagent tablet to the soil extract. The tablet contains sulfanilamide, *N*-(1-naphthyl)ethylenediaminedihydrochloride, zinc dust, sulfosalicylic acid, boric acid and magnesium stearate. Rapid reduction of nitrate to nitrite is accomplished by formation of the stable complex between zinc and sulfosalycilic acid. The color formed is compared with a standardized color chart.

Some tests of the sap of plant tissue are very useful as guides to the interpretation of the relative amounts of nutrients taken up by the plant.

A screening test for *phosphorus* in plant tissue sap is based upon production of the complex phosphomolybdic acid with sodium molybdate and the reduction of the heteropolyacid to molybdenum blue with a tin rod.

The plant tissue is sliced with a sharp knife and a filter paper is wetted with a plant juice. A drop of 0.1% NH_4MoO_4 solution in 0.1 N HCl is added, and a tin rod is pressed to the spot for ca. 10 s. If no color appears, the plant is very deficient in phosphorus, a blue color indicates a deficiency, a medium color indicates a slight deficiency in phosphorus, and the appearance of dark blue signals that the plant is adequately supplied with phosphorus.

A screening test for *potassium* in plant tissue is based on indirect testing for potassium deficiency which is feasible through detection of iron in the nodes of corn stalks. It has been shown that the iron concentration is inversely related to the total potassium in the leaf, nodal and internodal tissue. The stalk is cut with a sharp knife and some drops of KSCN are placed on the nodal area followed by a drop of 6 N HCl. K deficiency is indicated by the appearance of a red color.

5.5 Selected Examples of Applications of Spot Tests in Air and Water Quality Control

To protect the worker's well-being the concentrations of air pollutants in industrial areas must not exceed certain admissible levels. The simple estimation of *carbon monoxide* in ambient air is carried out by sucking the air sample through a glass tube containing a mixture of iodine pentoxide and fuming H_2SO_4. The extent of green coloration is the indication of CO concentration.

Glass-indicating tubes containing solid reagents are produced and marketed by Drägerwerke AG, Bacharach Industrial Instruments Co., Davis Engineering Co., Union Industrial Equipment Corp., Mine Safety Appliances Co, Acme Protection Equipment Corp. and so on. The simple operating procedure for the detecting tubes is as follows: the two sealed ends of the tube are broken and a recommended air volume is drawn through the tube with a calibrated bellows pump. The color change produced in the tube is compared visually against a set of standard colors. The length over which the color change has occurred in the tube is an alternative indication of the amount of airborne contaminant.[11]

Arsine has a maximum allowable concentration (MAC) of 0.05 ppm. A typical detector tube for it includes a precleansing layer of cupric ions, which precipitate the interfering reducing gases such as H_2S, H_2Se and mercaptans. The indicator layer is auric gold, which oxidizes the arsine to violet-gray metallic gold (Drägerwerke AG(DW); Mine Safety Appliances Co. (MSA); Union Industrial Equipment Corp. (UIE)).

Carbon monoxide detector tubes utilize the selenium dioxide-catalyzed reaction of I_2O_5 by CO in the presence of fuming H_2SO_4. The interfering materials are selectively oxidized by a precleansing layer ((DW), Bacharach Ind. Instruments (BII)), MSA, UIE).

Chlorine is estimated by an indicator tube, in which the chromogen is *o*-tolidine. It is oxidized by chlorine into a full-quinoidal yellow compound (DW, MSA, UIE).

Hydrocyanic acid is measured in the indicator tube by its complexation with Hg^{2+} ion, and liberation of HCl indicated by methyl red (DW, UIE).

Sulfur dioxide is estimated in the tubes by reaction with free iodine decolorizing the blue iodine–starch complex. The interference of H_2S is screened out with a cupric compound-containing precleansing layer (DW, MSA, UIE).

5.5.1 Water Quality Screening

Since the Safe Drinking Water Act passed by the US Congress in 1974 the Environmental Protection Agency (EPA) proposed a list of materials that present potential hazards when present in drinking water. Recent water

pollution controls rely heavily on advanced instrumental methods, but spot tests based on simple kits for quick and inexpensive water testing still find wide application and are approved by EPA (Hach Co., Merck Co., Gallard-Schlesinger Chem. Manufacturing, Motte Chem. Products Co., Taylor Chem. Co., Hellige Inc., etc.). In the present limited space only a few of them can be mentioned.[3,12,13]

The *chlorine* content in drinking water may be estimated by a stable tablet test. It contains equal parts by weight of KI, potassium bitartrate and soluble starch. Although Br_2, I_2, ozone and other oxidants react similarly to chlorine in this test, their presence is unlikely in water. The EPA accepts the use of the visual comparison method using color-step cubes (Hach Co.) or continuous color discs (Hach Co., Hellige Inc., Merck).

Chromium ions are suspected carcinogens. Their semiquantitative detection is based on their oxidation in alkaline solution to chromate and subsequent reaction of the latter with diphenylcarbazide. On-the-spot estimation kits use comparator blocks (Merck), calibrated continuous color discs and color cubes (Hack Co., Hellige Inc.).

Cyanide ions are estimated by the demasking of the red-violet mercury(II) dimethylaminobenzylidenerhodamine complex. An indicator stick based on this reaction is on the commercial market (MINTEK, Council for Mineral Technology, Randburg, South Africa).

Nitrate nitrogen in drinking water may be measured semiquantitatively on test strips (Merck). The test is based on diazotization with an aromatic amine of nitrous acid produced in situ in the reaction zone by a reducing agent and coupled with *N*(1-naphthyl) ethylenediamine.

Organophosphorus compounds are detected by a field test in water based on an enzyme reaction, namely on the inhibition of cholinesterase activity. Soman (GD) ($0.12\,\mu g\,L^{-1}$), sarin (GB) ($9.9\,\mu g\,L^{-1}$) and tabun (GA) ($26\,\mu g\,L^{-1}$) are detected, whereas the detection limit of the pesticide dichlorvos is $50\,\mu g\,L^{-1}$.

Total hardness of water is evaluated by test strips based on the complexing reaction between the Ca^{2+} and Mg^{2+} and ethylenediaminetetraacetic acid (EDTA) disodium salt. The test strip has four zones of gradually increasing reagent content. It is dipped into the water sample and the water hardness is given by the number of zones changing color from green to red-violet (E. Merck, Macherey-Nagel, Ames Co.).

5.6 Food Quality Control Analysis by Spot Test[3,14]

The partial substitution of one product for another in which the substituted material is inferior to the original product is defined as adulteration. Mineral oil is sometimes substituted for vegetable oil and saccharin is sometimes used instead of sugar. Starch is sometimes used to extend powdered cocoa, and butter can be replaced by margarine.

One of the most common food adulterants is water. Some of the proven toxic materials can be detected by simple screening tests, although most of the analytical procedures used are of a complex nature. A few examples of quick tests are described here.

Saccharin in food may be detected by conversion into phenolsulfonphthalein. The aqueous extract of the food is treated with a few drops of H_2SO_4 and a few grains of phenol and heated for 10 min at 180 °C. After cooling, it is neutralized with solid Na_2CO_3. The presence of saccharin is indicated by the appearance of a red color.

Artificial sweeteners (saccharin, cyclamate, dulcin and P-4000) can be separated and identified by TLC on silica gel using a solvent system of butanol, ethanol, ammonia and water. Saccharin is detected as a fluorescent spot under shortwave UV radiation. Successive spraying with solutions of 5% bromine in DMF–alcohol (1 : 1) and 2% *N*(1-naphthyl)ethylenediamine dihydrochloride in alcohol reveals cyclamate as a bright pink spot, dulcin as a brownish or blue spot and P-4000 as a brown-pink spot.

Objectionable food preservatives such as H_2O_2, salicylic acid, boric acid, borates, formaldehyde and sulfites may be screened using simple spot tests.

To test for *hydrogen peroxide*, equal volumes of milk sample and 1% KI solution dissolved in 2% starch paste are mixed. The presence of peroxides is indicated by the appearance of a blue color.

To test for *salicylic acid*, 50 mL of the sample or a small volume of the aqueous extract is acidified with 5 mL dilute (1 : 3) HCl solution in a separation funnel and extracted with diethyl ether. After evaporation of the ether, a drop of 0.5% neutral $FeCl_3$ solution is added. Violet ferric salicylate indicates the presence of salicylic acid.

Boric acid produces a deep-red compound when evaporated with tincture of curcuma (turmeric), changing yellow diferuloylmethane into the isomeric red-brown rosocyanin. It turns blue to greenish-black under alkaline conditions. The test is specific for boric acid.

Formaldehyde is extracted from food samples with water and after acidifying with phosphoric acid, the slurry is distilled. One milliliter of the distillate is treated with 5 mL of 0.5% chromotropic acid dissolved in 72% H_2SO_4. The tube is heated for 10 min in a boiling water bath. In the presence of formaldehyde, a bright-violet color appears.

The presence of *sulfur dioxide* or *sulfite* as preservatives is allowed in dehydrated fruits and vegetables, but their addition is prohibited to meats. A test strip for rapid identification of sulfite was prepared from a mixture of orange I (C.I. Orange 20), brilliant green (C.I. Basic Green 1), sodium bicarbonate and alumina. The black strip in contact with the moist sample of food will turn

red within 15 s in the presence of ≥5 μg sulfite and green in its absence.

5.6.1 Tests for Mammalian Feces and Urine Residues in Food

Food-processing plants and warehouses are subject to occasional invasion of rats and mice. The detection of rodent-contaminated lots, urine-contaminated grains, and mammalian feces in foods is a crucial task for law enforcement officials.

The spot test for *mammalian feces* is based on the fact that the alkaline phosphate isoenzyme present in the feces at certain pH splits phosphate radical from phenolphthalein diphosphate to produce the reddish free phenolphthalein. A pH 9.5 borate–carbonate reaction buffer optimizes the response of the enzyme.

Mammalian-urine contamination in food is detected by strips (Warner-Chilcott) in which the lowest part is impregnated with phosphate-buffered urease and the next band contains K_2CO_3. This area is bordered by a plastic barrier. Above the barrier is the indicator area containing bromocresol green.

Adulteration of *milk* by addition of $(NH_4)_2SO_4$ increases both the nonfat solid content and the nitrogen content as determined by the Kjeldahl procedure. One milliliter of milk is treated with 0.5 mL each of 2% NaOH and 2% NaOCl and an aqueous 5% phenol solution. The mixture is heated in a boiling water bath for 20 s. A rapidly deepening blue color appears in the presence of $(NH_4)_2SO_4$.

To test for differentiation of synthetic from natural vinegar, 10 mL of the sample is mixed with 1 mL concentrated H_3PO_4 and 1 mL 3% $KMnO_4$ solution. After 5 min the $KMnO_4$ color changes for natural but not for synthetic vinegar. The solution is then mixed with 1 mL saturated oxalic acid, 1 mL 10% H_2SO_4 and 1 mL Schiff's reagent (0.1% magenta, 1% anhydrous Na_2SO_3 in 0.1 N HCl). After 5 min natural vinegar gives a violet color reaction, but synthetic vinegar does not.

ABBREVIATIONS AND ACRONYMS

CF	Cystic Fibrosis
DMF	Dimethylformamide
EDTA	Ethylenediaminetetraacetic Acid
EPA	Environmental Protection Agency
GA	Tabun
GB	Sarin
GD	Soman
IR	Infrared
LSD	Lysergic Acid Diethylamide
MAC	Maximum Allowable Concentration
QC	Quality Control
TLC	Thin-layer Chromatography
UV	Ultraviolet

RELATED ARTICLES

Carbohydrate Analysis **(Volume 1)**
Disaccharide, Oligosaccharide and Polysaccharide Analysis • Monosaccharides and Sugar Alcohol Analysis

Chemical Weapons Chemicals Analysis **(Volume 1)**
Detection and Screening of Chemicals Related to the Chemical Weapons Convention

Clinical Chemistry **(Volume 2)**
Clinical Chemistry: Introduction • DNA Arrays: Preparation and Application • Drugs of Abuse, Analysis of • Glucose Measurement • Phosphorescence, Fluorescence, and Chemiluminescence in Clinical Chemistry • Urinalysis and Other Bodily Fluids

Environment: Water and Waste **(Volume 4)**
Soil Instrumental Methods

Food **(Volume 5)**
Food Analysis Techniques: Introduction • Atomic Spectroscopy in Food Analysis • Water Determination in Food

Forensic Science **(Volume 5)**
Forensic Science: Introduction

Pharmaceuticals and Drugs **(Volume 8)**
Alkaloids, Pharmaceutical Analysis of

REFERENCES

1. F. Feigl, V. Anger, *Spot Tests in Inorganic Analysis*, 6th edition, Elsevier, London, 1972.
2. F. Feigl, *Spot Tests in Organic Analysis*, 7th edition, Elsevier, Amsterdam, 1966.
3. E. Jungreis, *Spot Test Analysis, Clinical, Environmental, Forensic and Geochemical Applications*, 2nd edition, John Wiley & Sons, New York, 1997.
4. E. Jungreis, L. Ben-Dor: *Organic Spot Test Analysis, Comprehensive Analytical Chemistry*, ed. G. Svehla, Elsevier, Amsterdam, Vol. X, Chap. 1, 1980.
5. L. Gershenfeld, *Urine and Urinalysis*, 3rd edition, Saunders, Philadelphia, 1948.
6. E. Jungreis, in *Spot Test Analysis–Inorganic Ions, Encyclopedia of Analytical Science*, Academic Press, London, 4742–4748, 1995.

7. E. Jungreis, 'Organic Compounds', in *Encyclopedia of Analytical Science*, Academic Press, London, 4748–4755, 1995.
8. E. Jungreis, 'Applications', in *Encyclopedia of Analytical Science*, Academic Press, London, 4755–4760, 1995.
9. P.L Kirk, *Crime Investigation*, 2nd edition, Wiley, New York, 1974.
10. R.E. Stanton, *Rapid Methods of Trace Analysis for Geochemical Applications*, Edward Arnold, London, 1966.
11. *Detector Tube Handbook: Air Investigations and Technical Gas Analysis*, Dräger Tubes, 4th edition, A.G. Drägerwerk, Lübeck, Germany, 1979.
12. M.J. Suess, *Examination of Water for Pollution Control*, Pergamon Press, New York, Vol. 2, 1982.
13. *APHA Standard Methods for the Examination of Water and Wastewater*, 14th edition, American Public Health Association, Washington, DC, 1976.
14. F.L. Hart, H.J. Fisher, *Modern Food Analysis*, Springer Verlag, Berlin, 1971.

Titrimetry

M. Widmer (Deceased)
Ciba-Geigy, Basel, Switzerland

1 Acidimetric and Alkalimetric Titrations 1
1.1 Introduction 1
1.2 The Titration Curve 2
1.3 Titration of a Strong Acid (Hydrochloric Acid) with a Strong Base (Sodium Hydroxide) 2
1.4 Titration of a Weak Acid (Acetic Acid) with a Strong Base (Sodium Hydroxide) 3
1.5 Titration of a Weak Base (Ammonia) with a Strong Acid (Hydrochloric Acid) 4
1.6 Acid Enforcement Techniques 4
2 Complexometric Titrations 5
2.1 Introduction 5
2.2 Complexing Agents 6
2.3 Simultaneous Equilibria Influencing Complexometric Reactions 7
2.4 Indication in Complexometry 8
2.5 Applications of Complexometric Titrations 9
3 Electrochemical Titrations 9
3.1 Potentiometric Titrations 9
3.2 Amperometric Titrations 9
3.3 Conductometric Titrations 12
Acknowledgments 12
Abbreviations and Acronyms 12
Related Articles 12
Further Reading 13

In a titration, the analyte reacts with a reagent added as a solution of known concentration. The standard solution added from a buret is called a titrant. The volume of the titrant required to react completely with the analyte is measured as the volume at the equivalence point. The notions titrimetry and volumetry have the same meaning. Only well-defined stoichiometric reactions are applied. No additional calibration is necessary, i.e. titrimetric methods are primary methods. Indication of the equivalence point can be carried out by visual indicators or instrumental methods.

Acidimetric and alkalimetric titrations can be described by Brønsted's theory of strong and weak acids and bases. Theoretical titration curves for reactions of strong or weak acids (bases) with strong bases (acids) are demonstrated. Extremely weak electrolytes can be titrated in the presence of precipitating, oxidizing or complexing agents.

Determinations of metal ions are feasible by means of complexometric titrations. The theory of complexometry, including coupled chemical equilibria, can be derived from the fundamentals of complex-formation reactions in aqueous solutions. Applications are important for analyzing individual metal ions, mainly for their standardization, or for determining the hardness of water.

With regard to the instrumental indications of a titration, the potentiometric, amperometric and conductometric principles are outlined. Amperometric titrations of metal ions are performed with one polarizable electrode. The use of two polarizable electrodes (biamperometry) is especially important to determine the water content of organic solvents.

1 ACIDIMETRIC AND ALKALIMETRIC TITRATIONS

1.1 Introduction

Proton transfer reactions are usually extremely fast, although exceptions, such as the deprotonation of $CO_{2(aq)}$, are known, and therefore suitable for application in volumetric procedures. In addition, products of definite stoichiometry are formed and a number of chemical indicators and instrumental detectors are available for the end-point detection.

The instrumental devices are usually based on pH measurement, but other techniques, such as conductometric methods, are also in current use. The only limiting condition is the self-dissociation of the medium (water or any other solvent with acid and/or base properties) which may produce an acid (H_3O^+) or a base form (OH^-). It restricts the practical range for acidimetric or alkalimetric titration to a pH range between 0 and 14. For this reason, acidimetric or alkalimetric titrations are sometimes performed in nonaqueous solvents, where a different or larger practical range may apply.

Acidimetric and alkalimetric methods are usually based on pH measurements. With chemical indicators, estimates are usually accurate to within about one pH unit and more accurate determinations are better done with electrodes which respond to the hydrogen ion concentration, or preferably the hydrogen ion activity. Several such electrodes are known.

The most popular and widely used sensor is the glass electrode, although it may exhibit some nonideal behavior in solution with high pH values or high ionic strength. A typical glass electrode shows the

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

ideal pH response in the pH range from about 1 to 12. A constant potential difference for each pH unit, which is practically identical with the theoretical (minimum) slope, is then observed. The following millivolt difference is calculated per pH unit, from Equation (1):

$$E = E_2 - E_1 = \frac{2.3026RT}{F}(\log[H^+]_2 - \log[H^+]_1) \quad (1)$$

where R is the gas constant ($8.3143\,J\,K^{-1}\,mol^{-1}$), T is the absolute temperature (K), F is the Faraday constant ($96\,485.0\,C\,mol^{-1}$) and E is the potential difference observed between solutions 1 and 2 with hydrogen ion concentrations $[H^+]_1$ and $[H^+]_2$, respectively. The calculated values are as follows:

0 °C: 54.20 mV	30 °C: 60.14 mV
10 °C: 56.18 mV	40 °C: 62.13 mV
20 °C: 58.16 mV	50 °C: 64.12 mV
25 °C: 59.16 mV	60 °C: 66.11 mV

1.2 The Titration Curve

The results of a titration are best plotted in an appropriate diagram. However, before one can discuss the results of such a plot, a number of conventions must be introduced and defined. In the following discussion, an alkalimetric titration is outlined, but similar results are obtained in an acidimetric titration and the conclusions may be applied to both types of titration.

The following conventions are introduced:

1. In a theoretical treatment it is assumed that the titrant is sodium hydroxide (NaOH). One is dealing with an alkalimetric titration.
2. The concentration of the titrant (NaOH) is large ($>1.0\,mol\,L^{-1}$), so that no significant volume changes occur during the titration.
3. The concentration of the analyte (acid) is symbolized by C_s. Its acid form is HB and its base form is B^-. It is assumed that the acid is a neutral species (without charge) and that the base form is a univalent anion (Equation 2):

$$C_s = [HB] + [B^-] = [HB]_{total} = [B]_{total} \quad (2)$$

The quantity a_s is the degree of neutralization, defined according to Equation (3):

$$a_s = \frac{\text{number of moles of added base}}{\text{number of moles of original acid AB}} \quad (3)$$

a_s is a normalized quantity that replaces the volume (milliliters) in the titration curve.

The quantity p_s is the degree of protonation, defined according to Equation (4):

$$p_s = \frac{[\text{protonated base}]}{[\text{total base}]} = \frac{[HB]}{C_s} \quad (4)$$

and therefore (Equations 5 and 6):

$$[HB] = C_s p_s \quad (5)$$

$$[B^-] = C_s(1 - p_s) \quad (6)$$

One has to consider Equations (7–11), which enable us to calculate the magnitude of the five unknown quantities, $[Na^+]$, $[H^+]$, [HB], $[B^-]$ and $[OH^-]$:

$$[Na^+] = a_s C_s \quad (7)$$

$$C_s = [HB] + [B^-] \quad (8)$$

acid dissociation:

$$K_A = \frac{[H^+][B^-]}{[HB]} \quad (9)$$

self-dissociation of water:

$$K_w = [H^+][OH^-] = 10^{-14} \quad (10)$$

electroneutrality:

$$[Na^+] + [H^+] = [OH^-] + [B^-] \quad (11)$$

We now consider three different titrations.

1.3 Titration of a Strong Acid (Hydrochloric Acid) with a Strong Base (Sodium Hydroxide)

We assume $C_s = 0.01\,mol\,L^{-1}$. The values of the different quantities are given in Table 1.

Table 1 Hydrogen ion and hydroxide ion concentrations (mol/L) as a function of degree of neutralization

a_s	$[H^+]$	$[OH^-]$	pH
0.0	10^{-2}	10^{-12}	2.00
0.1	9.0×10^{-3}	1.1×10^{-12}	2.05
0.5	5.0×10^{-3}	2.0×10^{-12}	2.30
0.9	10^{-3}	10^{-11}	3.00
0.99	10^{-4}	10^{-10}	4.00
0.999	10^{-5}	10^{-9}	5.00
1.0	10^{-7}	10^{-7}	7.00
1.001	10^{-9}	10^{-5}	9.00
1.01	10^{-10}	10^{-4}	10.00
1.1	10^{-11}	10^{-3}	11.00
1.5	2.0×10^{-12}	5.0×10^{-3}	11.70
1.9	1.1×10^{-12}	9.0×10^{-3}	11.95
2.0	10^{-12}	10^{-2}	12.00

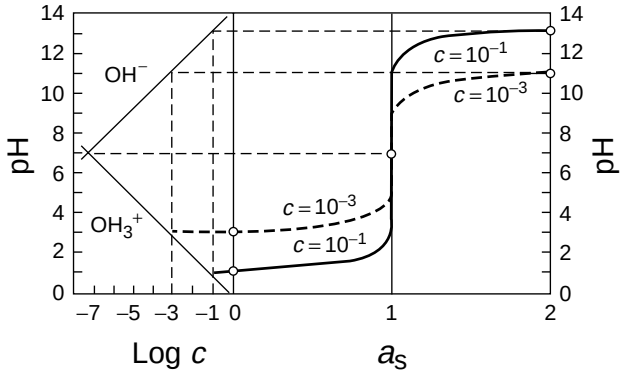

Figure 1 Titration curve of a strong acid (HCl) with a strong base (NaOH) in aqueous solution. The left part represents the logarithmic (Hägg) diagram.

The results of this calculation are shown in Figure 1, where the pH of the solution is plotted against the degree of neutralization. It represents the curve of any titration of a univalent strong acid with a strong base, and the abscissa may be replaced by the corresponding volume of added base.

At the point $a_s = 1.0$, the expression $\mathrm{dpH}/\mathrm{d}a_s$, representing the slope of the titration curve, reaches its maximum value. It is also the end-point of the titration.

It is easily detected by the pronounced pH jump or by any acid–base indicator with a pK_i value between 4 and 10. This means that the choice of the indicator is not critical. The expression $\mathrm{dpH} = \mathrm{d}a_s$ is at a minimum at the beginning and at the end of the titration curve. This indicates that strong acids and strong bases themselves have a pronounced buffer capacity.

1.4 Titration of a Weak Acid (Acetic Acid) with a Strong Base (Sodium Hydroxide)

We assume $C_s = 0.01\ \mathrm{mol\,L^{-1}}$. The values of the different calculated quantities are given in Table 2. Some simplified relationships have been introduced to reduce the effort of calculation. The acid dissociation constant is $K_A = 1.74 \times 10^{-5}$ or $pK_A = 4.76$.

The results are also represented in a titration curve (Figure 2). It can be seen that at the points $a_s = 0.5$ and $a_s = 1.0$ the term $\mathrm{dpH}/\mathrm{d}a_s$ assumes a minimum and maximum value, respectively. At $a_s = 0.5$ the system exhibits its largest buffer capacity (since $[H^+] = [B^-]$) and therefore shows the smallest change in pH per unit volume of added base. The point of minimum slope can therefore be used to determine the pK_A value of an unknown weak acid (the point where $pH = pK_A$). The end-point ($a_s = 1.0$) is again characterized by a pH jump, but it is less pronounced than in a titration of a strong

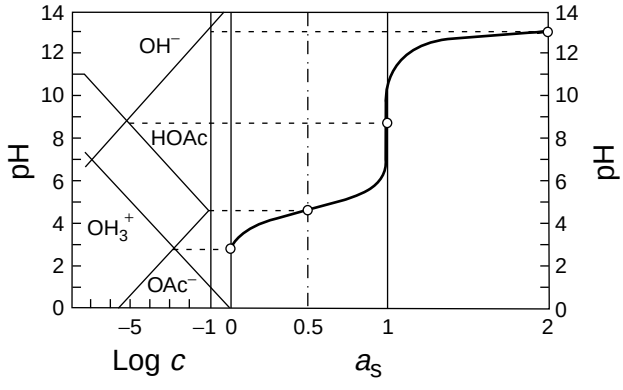

Figure 2 Titration curve of a weak acid (CH_3COOH) with a strong base (NaOH) in aqueous solution.

Table 2 Titration of a weak acid (acetic acid, $pK_A = 4.76$) with a strong base (NaOH)

a_s	$[Na^+]$ (mol/L)	$[H^+]$ (mol/L)	$[B^-]$ (mol/L)	[HB] (mol/L)	pH	Simplified relation
0.0	0.0	4.20×10^{-4}	4.20×10^{-4}	9.58×10^{-3}	3.38	$[H^+] = [B^-]$
0.1	1.0×10^{-3}	1.35×10^{-4}	1.14×10^{-3}	8.86×10^{-3}	3.87	$[Na^+] + [H^+] = [B^-]$
0.2	2.0×10^{-3}	6.76×10^{-5}	2.06×10^{-3}	7.94×10^{-3}	4.17	$[Na^+] + [H^+] = [B^-]$
0.3	3.0×10^{-3}	44.00×10^{-5}	3.04×10^{-3}	6.96×10^{-3}	4.40	$[Na^+] + [H^+] = [B^-]$
0.4	4.0×10^{-3}	2.58×10^{-5}	4.03×10^{-3}	5.97×10^{-3}	4.59	$[Na^+] + [H^+] = [B^-]$
0.5	5.0×10^{-3}	1.74×10^{-5}	5.02×10^{-3}	4.98×10^{-3}	4.76	$[Na^+] + [H^+] = [B^-]$
0.6	6.0×10^{-3}	1.15×10^{-5}	6.01×10^{-3}	3.99×10^{-3}	4.94	$[Na^+] + [H^+] = [B^-]$
0.7	7.0×10^{-3}	7.14×10^{-6}	7.01×10^{-3}	2.99×10^{-3}	5.13	$[Na^+] + [H^+] = [B^-]$
0.8	8.0×10^{-3}	5.40×10^{-6}	8.00×10^{-3}	2.00×10^{-3}	5.26	$[Na^+] = [B^-]$
0.9	9.0×10^{-3}	1.93×10^{-6}	9.00×10^{-3}	1.00×10^{-3}	5.72	$[Na^+] = [B^-]$
0.99	9.9×10^{-3}	1.76×10^{-7}	9.90×10^{-3}	1.00×10^{-4}	6.76	$[Na^+] = [B^-]$
1.0	1.0×10^{-2}	4.16×10^{-9}	1.00×10^{-2}	2.40×10^{-6}	8.38	$[Na^+] = [OH^-]$
1.1	1.1×10^{-2}	1.00×10^{-11}	1.00×10^{-2}	5.76×10^{-9}	11.00	$[Na^+] - [B^-] = [OH^-]$
1.5	1.5×10^{-2}	2.00×10^{-12}	1.00×10^{-2}	1.15×10^{-9}	11.70	$[Na^+] - [B^-] = [OH^-]$
1.9	1.9×10^{-2}	1.11×10^{-12}	1.00×10^{-2}	6.46×10^{-10}	11.95	$[Na^+] - [B^-] = [OH^-]$
2.0	2.0×10^{-2}	1.00×10^{-12}	1.00×10^{-2}	5.76×10^{-6}	12.00	$[Na^+] - [B^-] = [OH^-]$

acid with a strong base. Only a limited number of pH indicators can be used for the end-point detection. The pK_i must be between 7 and 10 for the choice of indicator.

1.5 Titration of a Weak Base (Ammonia) with a Strong Acid (Hydrochloric Acid)

The results of the titration of NH_3, $C = 0.1\,mol\,L^{-1}$, with HCl are presented in Figure 3. It can be seen that at the end-point of the titration the slope assumes a maximum value. However, the pH jump is smaller than in the titration curve of a strong acid with a strong base. It is comparable to the pH jump in the titration of a weak acid with a strong base, but it is shifted toward lower pH values.

As can be seen from the titration curve displayed in Figure 4, mixtures of strong and weak acids in aqueous solutions can be titrated in a single sample with a single titrant (a strong base in this example) if the dissociation constants of the acids are different enough to provide for a sufficient pH jump around the equivalence points. The example shown is a mixture of $0.1\,mol\,L^{-1}$ HCl and $0.1\,mol\,L^{-1}$ H_3PO_4. The first equivalence point

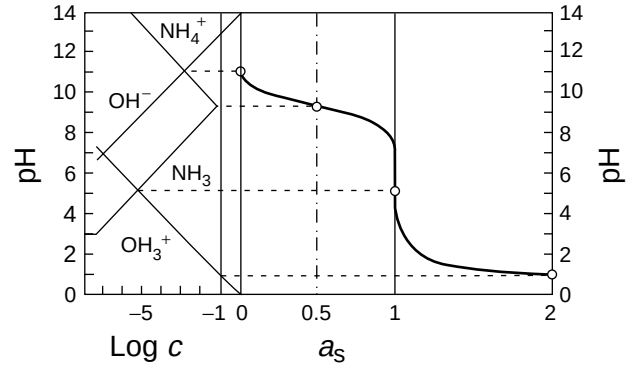

Figure 3 Titration curve of a weak base (NH_3) with a strong acid (HCl) in aqueous solution.

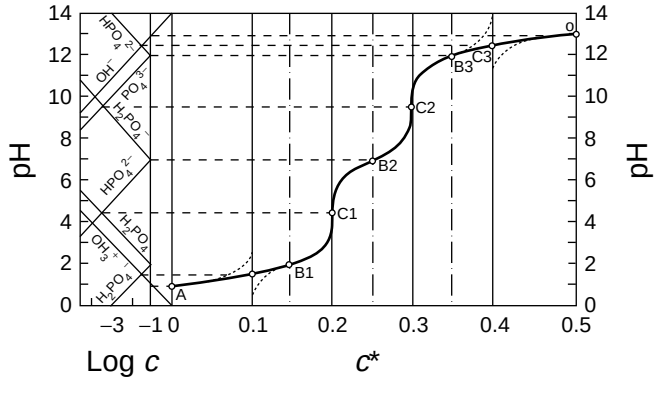

Figure 4 Titration curve of a mixture of HCl ($0.1\,mol\,L^{-1}$) and H_3PO_4 ($0.1\,mol\,L^{-1}$) with NaOH in aqueous solution.

Cl marks the sum of the concentrations of HCl and H_3PO_4, because the first proton of the phosphoric acid ($pK_{A_1} = 2$) and the HCl proton are indistinguishable in aqueous systems. The second dissociation of phosphoric acid ($pK_{A_2} = 7$), however, marks its own equivalence point C2, in such a way that the titrant volume C2 − C1 indicates the concentration of the phosphoric acid alone. From that result, the concentration of HCl can also be calculated. B1, B2 and B3 are buffer points.

It is even possible to titrate extremely weak acids with a strong base if the acidity of such an acid is enhanced through a chemical reaction. Several acid enforcement techniques are discussed later in this section.

The most common acids and bases may be arranged in the acidity order given in Table 3, according to the behavior in an alkalimetric or acidimetric titration.

1.6 Acid Enforcement Techniques

Extremely weak acids cannot be titrated with bases in pure aqueous solutions. However, through certain chemical reactions, weak acids may be transformed into strong acids, so that they can be titrated. These reactions involve one of the following.

1. Precipitation of the anion of the weak acid. Barium ions precipitate a number of anions and they may be used for this purpose (Equation 12):

$$2HB + Ba^{2+} \longrightarrow \{BaB_2\} + 2H^+ \qquad (12)$$

The braces { } denote the solid state of the barium precipitate, for example (Equation 13):

$$2HPO_4^{2-} + 3BaCl_2 \longrightarrow \{Ba_3(PO_4)_2\} + 2H^+ + 6Cl^- \qquad (13)$$

2. Oxidation of the weak acid to an oxygen-richer acid, e.g. Equation (14):

$$HNO_2 + \text{“O”} \longrightarrow HNO_3 \qquad (14)$$

or (Equation 15):

$$H_2SO_3 + \text{“O”} \longrightarrow H_2SO_4 \qquad (15)$$

3. Complexation: examples include the formation of an ester with a polyalcohol (Equation 16):

$$H_2BO_3 + 2C_3H_5(OH)_3 \longrightarrow BO_4(C_3H_5OH)_2^- + H_3O^+ + 2H_2O \qquad (16)$$

In a similar way it is possible to titrate the ammonium ion (weak acid) with a strong base after the addition of formaldehyde which converts it to

Table 3 Relative strengths of acids and bases

pK_A	Acid	Base	pK_B
Very strong:		*Extremely weak*:	
<-5	ClO_4H	$H^+ + ClO_4^-$	*ca.* 23
<-5	IH	$H^+ + I^-$	*ca.* 23
ca. −3.5	BrH	$H^+ + Br^-$	*ca.* 20
ca. −2	ClH	$H^+ + Cl^-$	*ca.* 17
ca. −2	SO_4H_2	$H^+ + SO_4H^-$	*ca.* 17
−1.74	OH_3^+	$H^+ + OH_2$	15.74
−1.32	NO_3H	$H^+ + NO_3^-$	15.32
0	ClO_3H	$H^+ + ClO_3^-$	14
0	BrO_3H	$H^+ + BrO_3^-$	14
Strong:		*Very weak*:	
1.42	$C_2O_4H_2$	$H^+ + C_2O_4H^-$	12.58
1.64	IO_3H	$H^+ + IO_3^-$	12.36
1.92	SO_3H_2	$H^+ + SO_3H^-$	12.03
1.92	SO_4H^-	$H^+ + SO_4^{2-}$	12.08
1.96	PO_4H_3	$H^+ + PO_4H_2^-$	12.04
2	$S_2O_3H^-$	$H^+ + S_2O_3^{2-}$	12
2.22	$Fe(OH_2)_4^{2+}$	$H^+ + Fe(OH)(OH_2)_3^{2+}$	11.78
2.32	AsO_4H_4	$H^+ + AsO_4H_3^-$	11.68
3.14	FH	$H^+ + F^+$	10.86
3.35	NO_3H	$H^+ + NO_3^-$	10.65
3.7	HCOOH	$H^+ + HCOO^-$	10.3
4	NCSH	$H^+ + NCS^-$	10
4.21	$C_2O_4H^-$	$H^+ + C_2O_4^{2-}$	9.79
Weak:		*Weak*:	
4.74	CH_3COOH	$H^+ + CH_3CO_2^-$	9.25
4.85	$Al(OH_2)_4^{3+}$	$H^+ + Al(OH)(OH_2)_3^{2+}$	9.15
6.52	$CO_2 + OH_2$	$H^+ + CO_3H^-$	7.48
6.92	SH_3	$H^+ + SH_2^-$	7.03
7	AsO_4H_3	$H^+ + AsO_4H_2^-$	7
7	SO_3H^-	$H^+ + SO_3^{2-}$	7
7.12	$PO_4H_2^-$	$H^+ + PO_4H^{2-}$	6.38
7.25	ClOH	$H^+ + ClO^-$	6.75
8.69	BrOH	$H^+ + BrO^-$	5.31
9.24	BO_3H_2	$H^+ + BO_3H^-$	4.76
9.22	AsO_3H_3	$H^+ + AsO_3H_2^-$	4.78
9.25	NH_4^+	$H^+ + NH_3$	4.75
Very weak:		*Strong*:	
9.40	NCH	$H^+ + NC^-$	4.60
9.61	$Zn(OH_2)_6^{2+}$	$H^+ + Zn(OH)(OH_2)_5^+$	4.39
10	SiO_4H_4	$H^+ + SiO_4H_3^-$	4
10.6	IOH	$H^+ + IO^-$	3.4
10.4	CO_3H^-	$H^+ + CO_3^{2-}$	3.6
11.62	O_2H_2	$H^+ + O_2H^-$	3.38
12	SiO_4H^{2-}	$H^+ + SiO_4H_2^{2-}$	2
12.32	PO_4H^{2-}	$H^+ + PO_4^{3-}$	1.68
13	$AsO_4H_2^-$	$H^+ + AsO_4^{3-}$	1
13	SH^-	$H^+ + S^{2-}$	1
13.52	$AsO_3H_2^-$	$H^+ + AsO_3H^{2-}$	0.48
Extremely weak:		*Very strong*:	
15.74	OH_2	$H^+ + OH^-$	−1.74
ca. 23	NH_3	$H^+ + NH_2^-$	*ca.* −9
ca. 24	OH^-	$H^+ + O^{2-}$	*ca.* −10
ca. 34	CH_4	$H^+ + CH_3^-$	*ca.* −20
38.6	H_2	$H^+ + H^-$	−24.6

hexamethylenetetrammine and a free hydronium ion (urotropine reaction) (Equation 17):

$$NH_4^+ + H_2O \rightleftharpoons NH_3 + H_3O^+ \longrightarrow C_6H_{12}N_3(\text{urotropine}) + H_3O^+ \quad (17)$$

2 COMPLEXOMETRIC TITRATIONS

2.1 Introduction

Complexometric titration methods are based on the stoichiometric reaction of metal ions in aqueous solutions with a class of suitable water-soluble bidentate or polydentate ligands, called 'complexones'.

To be of practical use, the reaction of the metal ions and the complexone must be fast, must result in a water-soluble product and should go nearly to completion when equivalent quantities of both substances are present. The first condition is a kinetic requirement, as in all titrations, and the last implies that the stability constant of the metal complex formed is large.

The most practical ligand systems are those involving chelating polyamines and aminopolycarboxylic acids (the Greek *chele* means 'scissors'). Most complexes with nonchelating monodentate ligands do not have sufficiently large stability constants to fulfill the stability requirement, as can be shown in the example of (monodentate) ammonia complexes with Cu^{2+} and the (tetradentate) triaminotriethylamino complex with the same metal ion.

2.1.1 Copper–Tetrammine Complexes (See Also Table 4)

A series of four relatively unstable monodentate complex species are formed consecutively (Equations 18–22):

$$Cu^{2+} + NH_3 \longrightarrow [Cu(NH_3)]^{2+} \quad \log K_1 = 4.1 \quad (18)$$

$$[Cu(NH_3)]^{2+} + NH_3 \longrightarrow [Cu(NH_3)_2]^{2+} \quad \log K_2 = 3.5 \quad (19)$$

$$[Cu(NH_3)_2]^{2+} + NH_3 \longrightarrow [Cu(NH_3)_3]^{2+} \quad \log K_3 = 2.9 \quad (20)$$

$$[Cu(NH_3)_3]^{2+} + NH_3 \longrightarrow [Cu(NH_3)_4]^{2+} \quad \log K_4 = 2.1 \quad (21)$$

$$Cu^{2+} + 4NH_3 \longrightarrow [Cu(NH_3)_4]^{2+} \quad \log K_{ST} = 12.6 \quad (22)$$

2.1.2 Copper–Triaminotriethylamine Complex (See Also Table 4)

The final complex with the tetradentate ligand is formed in a single step, resulting in a much higher stability

Table 4 Stability constants of some polyamine–metal complexes with copper and zinc (log K_{ST}). [Taken in part from G. Schwarzenbach, *Analyst*, **80**, 713 (1995)]

Complexing agent	Structure[a]	Zinc	Copper
Ammonia	NH_3	2.3, 2.3, 2.4, 2.1	4.1, 3.5, 2.9, 2.1
Ethylenediamine	$H_2NCH_2CH_2NH_2$	5.9, 5.2	10.7, 9.3
Diethylenetriamine	$H_2NCH_2CH_2NHCH_2CH_2NH_2$	8.9, 5.5	16.0, 5.3
1,2,3-Triaminopropane	$H_2NCH_2CHNH_2CH_2NH_2$	6.8, 4.3	11.0, 9.0
Triethylenetetramine (Triene)	$H_2NCH_2CH_2NHCH_2CH_2NHCH_2CH_2NH_2$	12.1	20.4
β,β',β''-Triaminotriethylamine	$N(CH_2CH_2NH_2)_3$	14.7	18.8

[a] Experimental conditions are 20 °C and ionic concentration 0.1 mol L^{-1}.

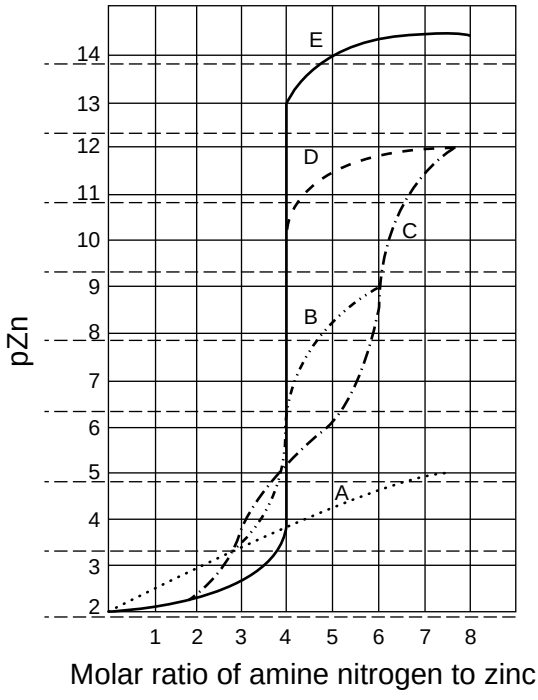

Figure 5 Titration curves for the titration of zinc with polyamines. Titrants: (A) ammonia; (B) ethylenediamine; (C) diethylenetriamine; (D) triethylenetetramine (Triene); (E) triaminotriethylamine.

constant of the complex and a steeper titration curve, owing to the chelate effect. The higher stability of a chelate complex is the result of an increase in the entropy of the system. One polydentate ligand, on reaction with a hydrated metal ion (e.g. Cu^{2+} binds four water molecules) liberates several water molecules, increasing the number of unbound particles and hence the entropy (Figure 5).

2.2 Complexing Agents

Several useful complexing agents with chelating properties, forming water-soluble chelate complexes, are commercially available. The major field of application is the quantitative determination of water hardness.

2.2.1 Nitrilotriacetic Acid (NTA) (Titriplex I) ***(1)***

$M = 191.14\,g\,mol^{-1}$, tetradentate, white powder, easily soluble in alkaline solutions.

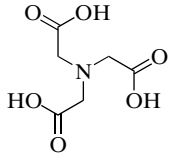

(1)

2.2.2 Ethylenediaminetetraacetic Acid (Titriplex II) ***(2)***

$M = 292.25\,g\,mol^{-1}$, hexadentate, white powder, not easily soluble in water, but soluble in alkaline solutions.

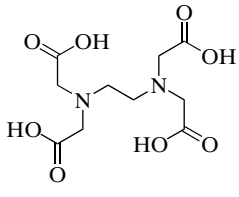

(2)

2.2.3 Ethylenediaminetetraacetic Acid Disodium Salt Dihydrate (Titriplex III) ***(3)***

$M = 372.24\,g\,mol^{-1}$, hexadentate, white powder, easily soluble in water, the aqueous solution is slightly acidic.

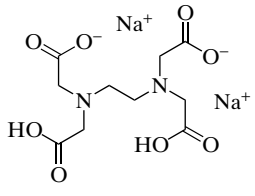

(3)

2.2.4 *Trans-1,2-Diaminocyclohexane-N,N,N′,N′-tetraacetic Acid Monohydrate (DCYTA)* ***(4)***

$M = 364.36\,\mathrm{g\,mol^{-1}}$, hexadentate, white powder, similar to Titriplex II, but with larger K_{ST} values (see Table 5).

Ethylenediaminetetraacetic acid (EDTA) (hexadentate) is the most frequently used complexometric agent. Structure **(5)** shows the three-dimensional octahedral structure of the chelate with a divalent cation, displaying the six coordination sites in the ligand, completely surrounding the central ion. The coordination sites are the two amino-nitrogen atoms and four oxygen atoms of the carboxylate groups.

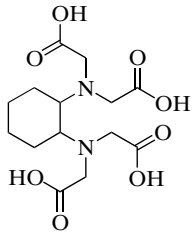

(4)

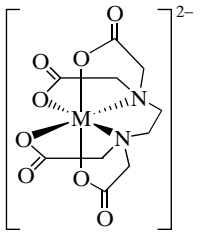

(5)

Table 5 Stability constants of some aminopolycarboxylic acid complexes

Metal ion	Log K_{ST}		
	NTA	EDTA	DCYTA
Ba^{2+}	4.8	7.8	8.0
Ca^{2+}	6.4	10.7	12.5
Fe^{3+}	15.9	25.1	29.3
La^{3+}	10.4	15.4	16.3
Mg^{2+}	5.4	8.7	10.3
Zn^{2+}	10.5	16.5	18.7

NTA, nitrilotriacetic acid; EDTA, ethylenediaminetetraacetic acid; DCYTA, *trans*-1,2-diaminocyclohexane-*N*,*N*,*N*′,*N*′-tetraacetic acid monohydrate.

2.3 Simultaneous Equilibria Influencing Complexometric Reactions

The formation of complexes may be accompanied by simultaneous parallel reactions in complex aqueous solutions, according to Scheme 1.

Without such side reactions, which are the ideal case, the total amount of ligand and metal ions would be available for the main titration reaction, as characterized by the equilibrium constant K_{ST} (Equation 23):

$$K_{ST} = \frac{C_{[MY]}}{C_M C_Y} \tag{23}$$

Prerequisite: $c_M = C_M$ and $c_Y = C_Y$. Owing to each of the side reactions, alone or in combination with each other, the available equilibrium concentration of metal ions and/or ligand may be (much) smaller than the total concentrations. A factor f is introduced to correct from the ideal case (f is between 1 and 0) and to give the apparent or conditional stability constant K'_{ST} (Equations 24–26):

$$f_{Me} = \frac{c_M}{C_M} \tag{24}$$

$$f_Y = \frac{c_Y}{C_Y} \tag{25}$$

$$K_{ST} = \frac{c_{[MY]}}{f_M C_M f_Y C_Y} \tag{26}$$

The conditional stability constant K'_{ST} describes the complex stability of the real system (Figure 6), including all side reactions (Equation 27):

$$K'_{ST} = K_{ST} f_M f_Y \tag{27}$$

An example is the titration of Zn^{2+} with $EDTAH_2Na_2$-(YH_2^{2-}) in NH_3 solution. The overall reaction of the titration is (Equation 28)

$$Zn_2^{+} + YH_2^{2-} \longrightarrow [ZnY]^{2-} + 2H^{+} \tag{28}$$

The complex reaction is (Equation 29)

$$Zn^{2+} + Y^{4-} \rightleftharpoons [ZnY]^{2-} \quad K_{ZnY}$$

Side reactions are as follows:

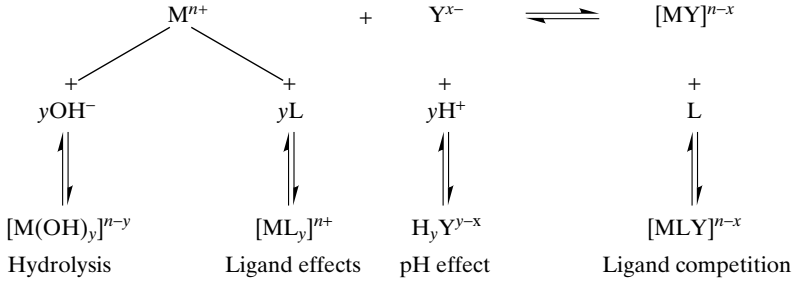

Scheme 1

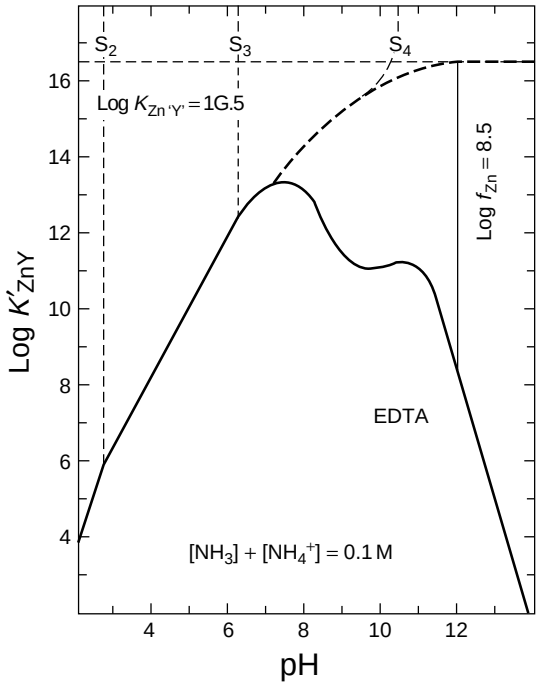

Figure 6 Plot of the conditional stability constant K'_{ZnY} of the Zn–EDTA complex as a function of pH at C_{NH_3}(total) = 0.1 mol L^{-1}.

1. EDTA reacts as an acid (Equations 29 and 30):

$$Y^{4-} + H^+ \rightleftharpoons YH^{3-} \quad pK_{S_4} = 10.3 \quad (29)$$

$$YH^{3-} + H^+ \rightleftharpoons YH_2{}^{2-} \quad pK_{S_3} = 6.2 \quad (30)$$

2. NH_3 reacts as a complexing agent (Equations 31 and 32):

$$Zn^{2+} + NH_3 \rightleftharpoons [Zn(NH_3)]^{2+} \quad (31)$$

$$[Zn(NH_3)]^{2+} + NH_3 \rightleftharpoons [Zn(NH_3)_2]^{2+} \quad (32)$$

3. OH^- reacts as a further complexing agent (pH > 10) (Equation 33):

$$Zn^{2+} + 4OH^- \rightleftharpoons [Zn(OH)_4]^{2-} \quad (33)$$

Values of selected stability constants can be found in Table 5 and in the literature.

2.4 Indication in Complexometry

The equivalence point of a complexometric titration is represented by the inflection point of a titration curve according to Figure 5. It can be found by the following methods:

1. Direct registration of the titration curve by electrochemical techniques. Major methods are amperometry, if the ''free'' metal ion and the metal chelate have different half-wave potentials, and potentiometry (ion-selective electrodes)
2. The use of metallochromic indicators. These are also complexing agents, but have the special property of color changes between the free ligand and the metal complex. In addition, the stability of the indicator complex must be significantly smaller than the stability of the titrant complex with the same metal ion.

At the beginning of the titration, the indicator reacts with the metal ion, forming a colored metal–indicator complex. During the titration, the stronger metal–titrant complex is formed, which causes the indicator complex to liberate its metal ions around the equivalence point, and to change into the free indicator form, also changing the color. At comparable color intensities of the free and complexed forms, the equivalence point is marked by the appearance of the mixed color, when 50% of the indicator has been liberated.

Metallochromic indicators are available for several metal ions. Some selected examples are as follows:

1. Eriochrome Black T (EBT) **(6)** (for Mg, Pb, Cd, Mn, Zn).

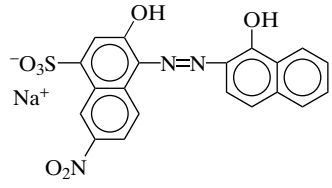

(6)

EBT is a trivalent acid, in practice used as the NaH_2E form. The indicator reaction is shown in Equation (34):

$$\underset{\text{blue}}{HEBT^{2-}} + Mg^{2+} \rightleftharpoons \underset{\text{red}}{[MgEBT]^-} + H^+ \quad (34)$$

EBT is sensitive to oxidation; the aqueous solution is stabilized with ascorbic acid. Fe, which would block the indicator, is masked with hydroxylamine. Commercially available ''indicator buffer tablets'' are a mixture of an indicator based on EBT and dyes (color change red–gray–green) with a buffer for pH 10.

2. Calgon carboxylic acid **(7)**, for Ca, and also for Cd, Mg, Mn, Zn.
3. Tiron **(8)** for Fe^{3+}.

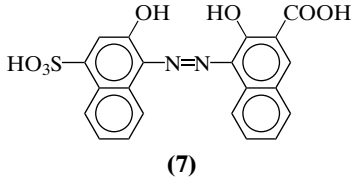

(7)

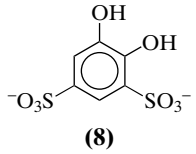

(8)

2.5 Applications of Complexometric Titrations

As a general rule, high selectivity for the metal ion to be determined by complexometric titrations of metal ion mixtures is required, and can be achieved by exploiting the principles of complex chemistry. The following variables have to be optimized for each individual case:

- selection of ligand
- selection of indicator
- adjustment of pH
- masking of interfering cations, forming stronger chelates with the titrant
- precipitation of interfering cations.

2.5.1 Titration Methods

2.5.1.1 Direct Titration EDTA forms chelate complexes with many cations, but with different stability. Fe^{3+} and Ca^{2+} can be titrated with EDTA in the same solution, Fe^{3+} at pH 3, and then Ca^{2+} after changing the pH to 10.

2.5.1.2 Back Titration For metal ions without an appropriate indicator, the sample solution is mixed with a known amount of EDTA in excess, and the excess is determined with Mg^{2+} and EBT.

2.5.1.3 Substitution Titration The sample solution (e.g. containing Fe^{3+}) is mixed with an excess of $[MgEDTA]^{2-}$, then the freed Mg^{2+} can be determined in the usual way.

2.5.1.4 Indirect Titration The determination of anions is possible by this technique, e.g. sulfate determination: precipitation of the sulfate with excess Ba^{2+}, followed by titration of the excess Ba^{2+} with EDTA.

2.5.2 Examples

2.5.2.1 Determination of Water Hardness Water hardness is represented by the sum of the concentrations of Ca^{2+} and Mg^{2+} in millimoles per liter, and is determined by the simultaneous titration of the two ions with EDTA (Titriplex III) and EBT at pH 10. Masking agents for the major interferents Fe^{3+} and Mn^{2+} (hydroxylamine), Al^{3+} (triethanolamine), and Cu^{2+} (cyanide) are described in the literature.

2.5.2.2 Individual Determination of Ca^{2+} and Mg^{2+} The solution is first adjusted to pH 12 with NaOH, and $Mg(OH)_2$ is then precipitated. Then Ca^{2+} is titrated alone with EDTA and Calgon carboxylic acid. To prevent errors due to coprecipitation of Ca^{2+}, the precipitate is redissolved in HCl and $Mg(OH)_2$ and precipitated a second time with NaOH. The exact Mg^{2+} content is calculated from the difference between the total hardness and the Ca^{2+} concentration.

3 ELECTROCHEMICAL TITRATIONS

Electrochemical titrations are classical titrimetric methods with electrochemical end-point detection. They are mainly potentiometric, amperometric and conductometric techniques, as described below.

3.1 Potentiometric Titrations

Potentiometric methods can be used to monitor the course of titration procedures, i.e. to determine the titration end-point if an indicator electrode is available for at least one of the chemical species taking part in the stoichiometric equivalence point of the chemical reactions (either the titrant or the titrand or both). The inflection point of the titration reaction [i.e. with the largest $\Delta E = \Delta V_R$ value (ΔE = potential difference, V_R = reagent volume)] coincides with the end-point of the reaction if the titration curve is symmetric. In any reaction where the stoichiometry is not 1:1, the titration curve is not symmetric. The titration error associated with the asymmetric titration curves can be partly overcome if the titrant concentration is determined under exactly the same experimental conditions as that of the analyte.

Typical titration curves recorded during the course of a precipitation titration are shown in Figure 7(a–c), together with different end-point determination approaches (direct reading, first-derivative and second-derivative). It must be kept in mind that the error of potentiometric titrations is also affected by the error of the titration end-point location.

The potentiometric method of end-point location can also be used in different automated titrations. However, in any titration procedure, a potentiometric indicator electrode serves only as an end-point indicator, and the analytical concentration of the analyte is determined on the basis of the stoichiometric chemical reaction.

3.2 Amperometric Titrations

Amperometry as a detection mode is widely used in different flow analytical techniques, including flow injection analysis and chromatographic separation methods, and also in conventional volumetric and coulometric titrations. In the latter, current measurement is used to locate the titration end-point.

The working electrode material is more critical in continuous-flow analytical techniques than in titrations,

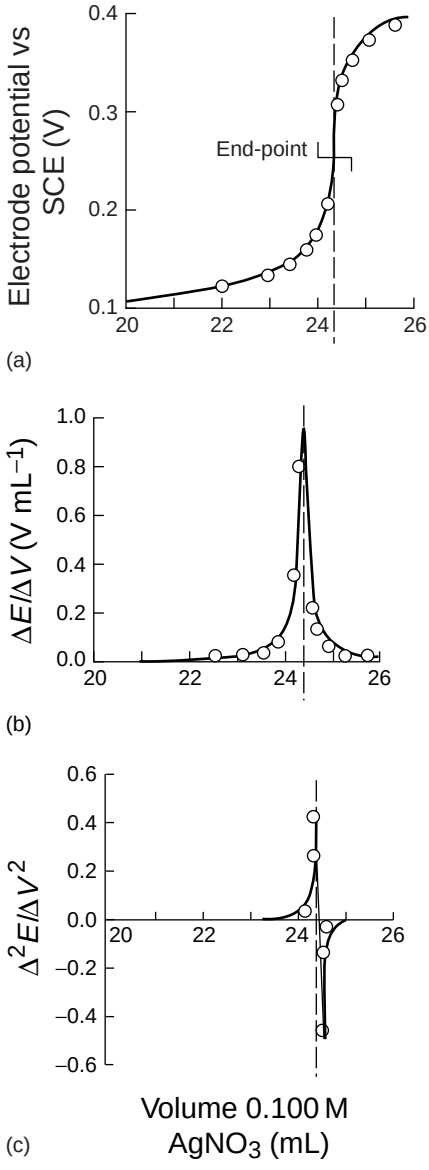

Figure 7 Typical potentiometric titration curves for titration of chloride with silver ions.

mainly because of mechanical and long-term operational stability. Different carbon electrodes, such as glassy carbon or carbon paste (a mixture of graphite powder and paraffin or silicone oil), are especially popular for high-performance chromatography and flow injection analysis. For conventional titrations, the indicator electrode is a polarizable electrode, e.g. a dropping mercury electrode (DME) or a platinum or a graphite microelectrode. The other electrode is either a reference electrode (i.e. a nonpolarizable electrode, for which the potential remains unchanged at relatively small current flow) such as a saturated calomel electrode (SCE), or another polarizable electrode. Thus, amperometric titration methods may be subdivided into

- titrations with one polarizable electrode
- titrations with two polarizable electrodes (biamperometric titrations).

3.2.1 Amperometric Titration with One Polarizable Electrode

This method involves measurement of the current during the titration in the presence of background electrolyte at a preselected working electrode potential in a stirred solution. The working electrode potential is selected so that the current is due to the electrochemical oxidation or reduction of the analyte or the titrant or both. The amperometric titration curve is constructed on the basis of the alteration of the limiting current versus titrant volume or titration percentage. The shape and characteristics of the titration curves depend on the current–voltage curve for the working electrode and the reversibility of the redox couples involved in the titration reaction.

As an example let us consider the titration reaction in Equation (35):

$$Pb^{2+} + Cr_2O_7^{2-} \longrightarrow PbCr_2O_7 \qquad (35)$$

The working electrode is a DME and the reference electrode is a SCE. The working electrode potential is set and kept at a constant value with respect to the reference electrode, so that the limiting current of the reducible species, Pb^{2+}, is measured as a function of volume of titrant added. If the potential is set to 0 V the current remains at zero level up to the end-point because Pb^{2+} is not reduced at that potential. After the equivalence point, the diffusion current rises, owing to the added excess of the reducible chromate ions. The titration end-point is the intersection point obtained by extrapolation of the two linear portions of the titration curve. The titration points close to the end-point are not considered, owing to the dissociation of complexes and the solubility of the precipitates formed during the titration process.

When both titrant and analyte ions give diffusion current at the selected operating voltage, a V-type titration curve is obtained. This is true for the lead–chromate couple if the potential is set to −1 V.

The rotating Pt electrode can also be used as a working electrode in amperometric titrations. The increased limiting current value due to convective diffusion and the lowered residual current at an electrode of constant surface area in the amperometric operating mode significantly lower the detection limit of the technique. Hence amperometric titration can be used for quantitative analyses in the concentration range 10^{-5}–10^{-6} mol L^{-1}. However, it must be kept in mind that the low overvoltage of hydrogen on the platinum electrode may be a serious problem when using the electrode as cathode in acidic solutions.

The relative error of amperometric titration is less than ±2–3% relative standard deviation. The selectivity of the direct amperometric technique can be improved by appropriate selection of the titration reaction. Thus, compounds of similar $E_{1/2}$ may be determined in the presence of each other by amperometric titration with the proper selection of the titrants.

3.2.2 Amperometric Titration with Two Polarizable Electrodes

Amperometric titrations using two identical polarizable electrodes of equal size are also called biamperometric titrations. The current is measured while a relatively small, fixed potential (10–100 mV) is applied across the electrodes as a function of the titrant volume. One electrode functions as the anode and the other as the cathode. The shape of the titration curve is strongly dependent on the reversibility of the electrode reaction of the titrant and the analyte system. Biamperometric titrations are primarily used for the end-point detection of redox titrations.

Let us consider, as an example, the titration of I_2 with $S_2O_3^{2-}$ at a low potential, ΔE, applied across the two platinum electrodes. The titration reaction is given by Equation (36):

$$I_2 + 2S_2O_3^{2-} \rightleftharpoons 2I^- + S_4O_6^{2-} \quad (36)$$

For the qualitative interpretation of the titration curves, the $I_2|I^-$ and $S_2O_3^{2-}|S_4O_6^{2-}$ redox couples must be considered. As is known, current will flow in the electrochemical cell if an electrode reaction proceeds at both electrodes. If the fixed potential applied is low, the current flows in the cell when both components of the reversible redox system are present in the solution. The magnitude of the current is determined by the component present in the solution at a lower concentration.

The $I_2|I^-$ system is reversible at platinum electrodes, whereas the $S_2O_3^{2-}|S_4O_6^{2-}$ system is irreversible. Therefore, at a low applied potential, the reaction shown in Equation (37) proceeds:

$$I_2 + 2e^- \longrightarrow 2I^- \quad (37)$$

and the reaction shown in Equation (38):

$$S_2O_3^{2-} \longrightarrow S_4O_6^{2-} + 2e^- \quad (38)$$

does not.

Accordingly, at the start of the titration there is no current flow, since no I^- is present in the solution to be titrated; only the residual current can be measured.

As the titration proceeds, I^- is generated by the redox reaction and faradaic current flows between the Pt electrodes according to the electrode reactions in Equations (39) and (40):

At the cathode:

$$I_2 + 2e^- \longrightarrow 2I^- \quad (39)$$

At the anode:

$$2I^- \longrightarrow I_2 + 2e^- \quad (40)$$

The current is determined by the concentration of the species of the redox couple present in the solution at lower concentration. Hence, the current has a maximum at 50% titration when the concentrations of both I_2 and I^- are equal, then it decreases owing to the consumption of the I_2 in the titration reaction. At 100% titration and beyond this point there is no I_2 present and only residual current is observed. The titration curves of this type of biamperometric titration are said to have 'dead stop' end-points since the current falls essentially to zero at the end-point and remains there.

Similar titration curves can be derived by considering the current–voltage curves of the titrant and the analyte systems and the potential applied.

3.2.3 Karl Fischer Titration of Water

The technique can be used for the determination of water in purified solvents and for water of crystallization in crystals. The Karl Fischer reagent, containing, for example, iodine, sulfur dioxide and pyridine in a 1 : 3 : 10 mole ratio in methanol, can be used for the direct titration of water in any solvent which does not react with sulfur dioxide and/or iodine. It must be kept in mind that solvents containing aldehydes and ketones cannot be titrated, since they bind sulfur dioxide.

The basis of the method is that sulfur dioxide reacts with iodine in the presence of water, according to the reaction in Equation (41), leading to equilibrium:

$$SO_2 + I_2 + 2H_2O \rightleftharpoons SO_4^{2-} + 2I^- + 4H^+ \quad (41)$$

The equilibrium can be shifted to the right-hand side owing to the presence of pyridine (Equations 42 and 43):

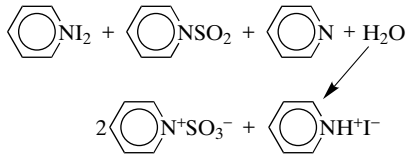

(42)

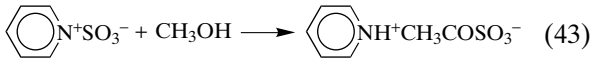

(43)

The water equivalent of the Karl Fischer standard is determined with a methanol solution containing a known amount of water, e.g. $CH_3COOH \cdot 3H_2O$.

In the course of the titration, the variation of the current as a function of a known volume of standardized Karl Fischer reagent, i.e. $i-V_{\text{Reagent}}$ (mL) is recorded. The titration end-point is determined by dead-stop end-point detection.

3.3 Conductometric Titrations

The specificity of conductometry is enhanced by the addition of a selective reagent. This fact is used in conductometric titrations in which the variation of the electrical conductivity of a solution as a function of the reagent addition V_R is monitored.

The quantitative determination is based on the determination of the reagent consumption until chemical equivalence is reached. The titration end-point is obtained by the intersection of two straight lines fitted to the linear parts of the titration curves (i.e. conductance–titration volume data pairs) recorded during the course of titration.

The main advantages of conductance measurements over classical and potentiometric end-point detection are as follows:

- they can be used for the titration of colored or turbid solutions;
- the titration curve can be constructed on the basis of a few data points, far from the chemical equivalence point.

The latter allows one to overcome problems such as the slow potential response in the neighborhood of the chemical equivalence point in potentiometric titrations. Thus the speed of titrations can be significantly enhanced.

Conductivity end-point detection is used especially in acid–base and precipitation titrations. As an illustration, let us consider the acid–base titration in Equation (44):

$$\underline{H}^+ + \underline{A}^- + M^+ + OH^- = H_2O + \underline{M}^+ + \underline{A}^- \quad (44)$$

The underlined ions are present in the solution before the titration end-point.

Thus, with the addition of titrant (MOH), H^+ ions are replaced by metal ions M^+, while the anion concentration is unchanged. Until the chemical equivalence point is reached, the number of ionic species is unchanged; only the hydrogen ions are exchanged for metal ions, and since the mobility of the hydrogen ions is much greater than that of metal cations, the conductance of the solutions decreases until the equivalence point is reached. After the equivalence point, the conductance corresponds to the excess of hydroxyl ions added, compared with that corresponding to the equivalence point. Thus the conductance increases after the equivalence point as the base concentration increases, and the corresponding titration curve (strong acid with strong base) is V-shaped.

In contrast, in the titration of a weak acid with a weak base, the conductivity of the starting solution is low, owing to limited dissociation. The salt produced in the course of the titration dissociates completely, resulting in a higher conductivity. However, after the equivalence point is reached, the concentration of the strong electrolyte remains constant, and only that of the weak base increases.

In general, a well-defined titration point is obtained only if a strong acid is titrated with strong base or a weak acid with weak base.

Hydrolysis, dissociation of the reaction product or solubility phenomena give rise to curvature of the titration curve in the vicinity of the end-point. At a distance away from the end-point, i.e. from 0 to 50% and between 150 and 200% of the equivalent volume of titrant, these effects are suppressed. By extrapolating these linear portions, the position of the end-point can be determined precisely.

As in all titration techniques, the titrant must be 10 times more concentrated than the solution being titrated, in order to keep the volume change small. If necessary, a correction may be applied; all conductance reagents are multiplied by the terms $(V + v)/V$, where V is the initial solution volume and v is the titrant volume added.

ACKNOWLEDGMENTS

The publishers are grateful to Professors J.-M. Mermet and M. Otto for allowing reproduction of this article from *Analytical Chemistry*, edited by Professors Kellner, Mermet, Widmer and Otto, published by Wiley–VCH in 1998.

ABBREVIATIONS AND ACRONYMS

DCYTA	*trans*-1,2-Diaminocyclohexane-*N*,*N*,*N*′,*N*′-tetraacetic acid monohydrate
DME	Dropping Mercury Electrode
EBT	Eriochrome Black T
EDTA	Ethylenediaminetetraacetic Acid
NTA	Nitrilotriacetic Acid
SCE	Saturated Calomel Electrode

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Fluorescence Spectroscopy In Vivo

Food **(Volume 5)**
Water Determination in Food

Process Instrumental Methods **(Volume 9)**
Titration Techniques for Process Analysis

Electroanalytical Methods **(Volume 11)**
Electroanalytical Methods: Introduction • Ion-selective Electrodes: Fundamentals

FURTHER READING

T.P. Hadhjioannou, G.D. Christian, C.E. Efstathiou, D. Nikolis, *Problem Solving in Analytical Chemistry*, Pergamon Press, Oxford, 1988.

I.M. Kolthoff, E.B. Sandell, E.J. Meehan, S. Bruckenstein, *Quantitative Chemical Analysis*, 4th edition, Macmillan, London, 1969.

Traceability in Analytical Chemistry

Freddy Adams
University of Antwerp (UIA), Antwerp, Belgium

1 Introduction 1
2 Analytical Chemistry and Chemical Analysis 2
3 Traceability and Analytical Chemistry 3
3.1 Primary Methods 3
3.2 Primary References 4
3.3 Standard or Reference Materials 5
3.4 Confidence Limits 6
4 Conclusions 7
Abbreviations and Acronyms 8
Related Articles 8
References 8

This article describes the concepts of traceability for chemical measurements (in chemical analysis) as they are defined at present, and their relation to analytical chemistry. Traceability concerns the direct or indirect linkage of analytical results to the international metric system of units [Système International (SI)], thereby showing a way to achieving comparability in chemical analysis. The metrological concepts to do this, as they have been developed up to the present, are critically evaluated. They concern the units in which analytical results should be expressed and the tools, such as primary reference methods and primary standards, certified reference materials (CRMs) and other reference materials (RMs), that can be used to increase the accuracy of results of analysis.

1 INTRODUCTION

In the global world of today, the validity of data, including those obtained by chemical analysis, must be properly guaranteed for their quality. There is, at present, growing pressure on improving the overall quality of analytical information and on making it consistent with general metrological principles. There is a widespread belief that, all too often, analytical results are far from being accurate and that these inaccuracies do not occur only on the discipline's borderline (in the application of methods under development), but also during the application of well established methods. Any methods that might help in increasing the quality of analysis are, hence, welcome. One of the principles selected towards proper quality control is the concept of traceability. The term traceability was gradually introduced into chemical analysis over the last 20 years.[1]

Metrology, the science of measurements, was traditionally used in connection with the measurement of a number of physical parameters (such as length, time, voltage, ...). Needs have been expressed during the last decade for connecting analytical measurements with the SI and for judging the quality of analytical measurements by the estimation of their uncertainties according to internationally agreed guidelines.[2] For instance, the criteria for the operation of testing laboratories of the EU (European Union) (EN 45001) specify that:

- measurement equipment is regularly recalibrated;
- a RM, traceable to a national or international standard reference material (SRM), is used whenever possible;
- whenever traceability to national/international standards of measurement is not applicable, the laboratories shall provide evidence of accuracy by indirect means, e.g. by participating in programs of interlaboratory comparison.

These guidelines impose concepts connected with traceability as one of the requirements of the results of chemical analysis. In general terms, traceability refers to the direct link between the producer of analytical data and the national and international measurement system.

In the vocabulary of the International Organization for Standardization (ISO) (Geneva, Switzerland), for quality management and quality assurance, traceability is rather broadly defined as "the ability to trace the history, application or location of an entity by means of recorded identifications" (ISO 8402, 1994). For calibration purposes there was added to this definition, "traceability relates measuring equipment to national or international standards, basic physical constants or RMs". In connection with chemical measurements, the basic concepts of traceability were more specifically defined by the CCQM (Consultative Committee on Quality of Measurements) as "the property of the result of a measurement whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties". In this definition, the uncertainty of a measurement is the parameter associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurement.[3] According to common metrological concepts, all this should be achieved by

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

the application of a so-called metrological pyramid with primary standards on the top and using international and national metrological laboratories as the priviledged intermediates to the SI system. Basically, the application of these principles should guarantee that the analytical results are accurate (true) and comparable both over time and between individual laboratories. Implied also is that the results are given to the user with sufficient information on their significance expressed as confidence levels or uncertainties.

The basic definition of traceability given by the CCQM was implemented in recent years, mostly in connection with the question of how this unbroken chain of comparisons with the SI system could be achieved in analytical practice. The sometimes very rigid aspects that have been put forward for doing this are not always well accepted in the analytical chemistry community.[4–7] Most of the differences that arise are connected with the differences between analytical chemistry on one side and metrology on the other, as distinct scientific disciplines (see further).

It is of interest to determine to what extent the concepts of traceability are applicable in analytical practice and how they might be put to use for increasing the overall quality of analytical results. As a concept, traceability is then reduced to an important and practical tool instead of a dogmatic undertaking. It can be useful to complement or assist any other available tools for achieving high quality and comparability of analytical data.

2 ANALYTICAL CHEMISTRY AND CHEMICAL ANALYSIS

In its diversity of approaches, analytical chemistry is a complex science that relies on a variety of different techniques and concepts. Chemical analysis is also an important factor in the economy and society as they exist today. Chemical analysis involves a considerable economic cost, but it results in considerable added value of the goods produced and helps in promoting international trade. Chemical analysis constitutes a considerable fraction (up to 6%) of the gross national product that advanced nations spend on measurements and related operations. It is important for advanced society in, for example, environmental protection, food, safety and health. It is, hence, essential that results are trustworthy and reliable for the many diverse purposes they serve.

Over the last 50 years, the advent of a number of rapid, highly sensitive and precise instrumental analytical techniques has contributed to a tremendous increase in the overall number of chemical (analytical) measurements. These instrumental techniques now cover nearly every imaginable analytical need. Most of these techniques are comparative in nature (requiring calibration of the analytical signal) and suffer from a number of matrix-dependent influences in the expression of the relation between the signal measured and the concentration of the measurand (the selected item to be quantified in a sample, the analyte). They therefore need careful calibration and the validation of the methods and procedures through the use of RMs, thus linking the measurements to the truth and achieving traceability in its broadest definition.

Chemical analysis is a particular measurement process, distinct from the physical measurement processes. The differences in attitude between analytical chemists and metrologists are conceptual and terminological. They are the result of the push to solve the problems that analysts face through the exploitation of the analytical sciences, as opposed to the dogmatism behind the metrological efforts for achieving utmost accuracy, sometimes at any cost that this might imply.

The analytical act is an indivisible process including the measurement equipment, the measurand in its relation to the particular environment (the sample subjected to analysis), and the very often complex overall procedure that leads to the result. The errors of the determination depend on random error sources in sampling, sample preparation for analysis during transfer, preparation and storage and chemical separations. The quality of the results depends also on systematic sources of error such as incomplete recovery and other losses of the measurand, interferences and contaminations. Analysis based exclusively on a measurement is extremely rare in practice (limited to some methods of nondestructive instrumental analysis).

In its application, analytical chemistry is complex as it draws on any available technique or information that is suitable for the specific purpose involved, namely deriving the concentration of the measurand in a given object subjected to analysis. In its applications, analysis is also very strongly dependent on the properties and the overall composition of the samples involved. Analytical procedures and the use of measurement standards are very often different when applied to different types of samples. Therefore, processes of method validation need to be related to the overall analytical act and to the specific sample involved in order to cover errors due to sampling, incomplete recovery of the analytes, contaminants, reagent blanks and interferences. This is commonly achieved by studies of all parameters affecting the analysis in interlaboratory comparisons, the use of reference samples, proficiency testing and other practical and operational validation tools.

Owing to the importance of matrix effects, the appropriate standards for calibration and reference

to the measurement system are, in general, lacking. Measurement standards for most physical measurements, on the other hand, are, in general, readily available. There is a widespread and growing need for more (and more reliable) reference samples. At the same time, it is realized that well-characterized RMs are extremely costly to produce, store and distribute.

3 TRACEABILITY AND ANALYTICAL CHEMISTRY

It is important to review the basic concepts behind traceability as they have been implemented in practice by the CCQM. This consultative Committee of the International Committee for Weights and Measures (CIPM) was established in 1993 to translate the traceability concepts to chemical analysis.[8] The major items addressed by the CCQM up to now are as follows:

- International comparability of chemical measurements is to be achieved by linking all measurements (analyses) to the SI system and expressed as the basic SI unit of amount of substance with the mole or derived entities as the unit. Since its introduction in 1971, the mole is the amount of substance containing as many elementary items (atoms, ions, ...) as there are atoms in 0.012 kg of carbon-12. Analytical chemistry, on the other hand, is used to express results of the measurement process in concentration units expressed in different units according to the specific problem being addressed. As long as these concentration units are properly defined (and traceable to the SI basic unit), there cannot be a problem in converting results. There is an extensive literature on the formal aspects of the measurement unit, e.g. on the way to express it (with the subtle difference between mass and amount of substance and other problems of nomenclature), how it relates to the different elements (e.g. for polyisotopic elements such as lead with a variable isotopic composition) and to the Avogadro constant and the uncertainty with which it is known at present (ca. 10^{-6}, or more than sufficient for nearly all analytical purposes).
- The CCQM defined the analytical methods for measurement of the amount of substance in mole units that are directly linked to SI units as so-called primary methods. The term definite method was used earlier. The definition of primary methods implies that these methods are completely described and understood and that a complete uncertainty statement can be expressed in terms of the SI units *without reference to a standard of the quantity being measured*. The basis of the choice of methods is that they are specific for the measurand and that all parameters or corrections dependent on the matrix or other constituents are known or can be calculated with appropriate uncertainty. The requirement that no comparative measurements are allowed with the pure or diluted entity subjected to analysis, excludes nearly all of the available instrumental analytical techniques.

3.1 Primary Methods

Only a very limited number of methods have been identified as having the potential to be a primary method. Methods selected so far are as follows:

1. Isotope dilution mass spectrometry (IDMS), a method whose result depends on the accurate mass spectrometric determination of isotope ratios of unknown sample, a spiked sample brought in full equilibrium with the spike. It should be taken into account that for elemental analysis the method is only applicable to part of the periodic system (the polyisotopic elements), and that it is slow and expensive when used with thermal ionization mass spectrometry. It might become more easily applicable for elemental analysis with inductively coupled plasma mass spectrometry. Its use can be extended, however, to the determination of an extensive list of organic compounds through the use of labeled compounds, e.g. through the addition to the sample of a ^{13}C-enriched homolog of the measurand.
2. Coulometry and differential scanning coulometry, where the analytical signal is directly related to the measure and through the Faraday constant.
3. Reaction chemistry using gravimetry and titrimetry. These methods are based on stoichiometric solution chemistry. Gravimetry derives the result from the measurement of the mass of a slightly soluble product, and thus relates directly to the measurement of mass. Volumetric analysis is based on the measurement of the volume of a primary standard solution of another chemical entity, with pure Ag as the ultimate standard.
4. Determinations based on lowering of the freezing point, change of the freezing point and other methods based on the theorem that in sufficiently dilute solutions the solvent behaves ideally.

Primary methods are carefully defined by the CCQM as methods for the determination of the amount of substance in pure or simple compound systems, i.e. in samples which do not contain any impurities which might potentially act as interferences (and, hence, become an

error source) in the primary method applied. The CCQM considers it as one of its future tasks to investigate the applicability and robustness of these methods for complex mixtures encountered in practical analytical chemistry. Many reports, however, tend to identify primary methods already as methods of analysis (to be used on complex samples of unknown overall composition). This is, of course, an overoptimistic (and unwarranted) enlargement of the definition that would, for example, imply that all titrimetric methods of analysis could be considered as primary methods, putting aside any of the many potential interferences that occur in their application in complex samples. Kolthoff's handbooks on titrimetry are full of such interferences!

The CCQM is at present evaluating other analytical methods for their potential of being promoted as a primary method, e.g. nuclear magnetic resonance (NMR) spectrometry. It also recently (1998) redefined and extended the definition of ''primary methods''. A distinction was made between a primary ''direct'' method (e.g. coulometry, gravimetry) that can be used to measure the value of an unknown without reference to a standard of the same quantity, and primary ''ratio'' methods where the value of the unknown is determined by means of the measurement of a ratio of the unknown to a standard of the same quantity (e.g. IDMS). The latter primary methods thus rely on reference of the unknown to a reference that is itself traceable to SI units. The distinction was made because three of the methods originally identified by the CCQM as ''primary'' are primary ratio methods, e.g. IDMS, which relies on a gravimetric assay (or an assay with another primary direct method) of the pure spike.

3.2 Primary References

According to the definition, these methods, and no others, can be used for the determination of the purity of materials used in the preparation of (primary) RMs. A primary reference material (PRM) is one made on the basis of primary methods, thus primary standards and primary methods are linked to each other, the latter being always the result of measurements with the first. PRMs can be used for calibration of instruments and, hence, are of help in the measurement step of the analytical procedures to evaluate uncertainties in the measurement process and, for example, to evaluate the extent of matrix-dependent effects.

Considering the limitations in available primary methods, emphasis must be placed on other measurement standards. For determinations of nonisolated substances, (those present in complex matrixes) it is considered that other RMs are needed which are firmly linked to SI units by using primary methods. In practice, this implies that such RMs must be artificially produced, e.g. by mixing them together starting from the PRMs. This distinguishes them from any real-life RMs (matrix RMs; see further).

The possible set of ways of realizing traceability to SI units using intermediate points and primary methods is shown in Figure 1.[9] In Figure 2, the structure is schematically shown of how a network of laboratories, ranging from international metrology institutes, over secondary (national) reference laboratories, could link working-level laboratories to the SI units.[10] In the physical measurements field, networks linking working-level laboratories to internationally recognized metrology institutes exist, but not at present in chemical analysis.

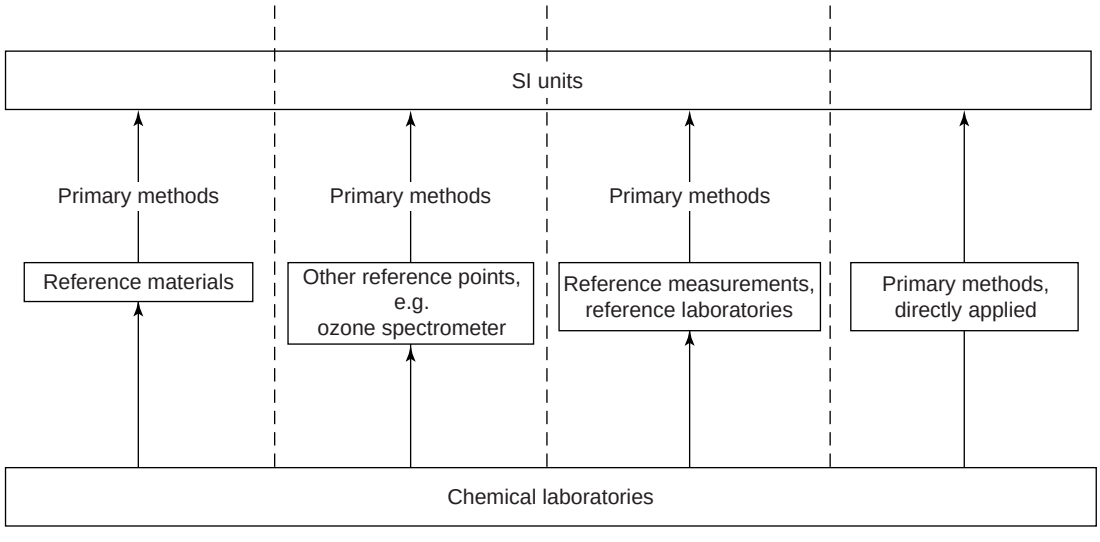

Figure 1 Ways of realizing traceability of chemical measurements to SI units using primary methods and intermediate reference points. (Reproduced by permission from Richter and Dube.[9])

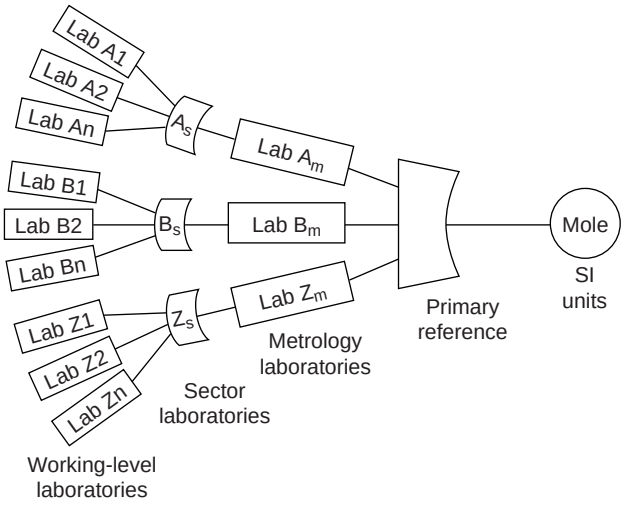

Figure 2 A network of laboratories in the traceability chain. (Reproduced by permission from King.[10])

3.3 Standard or Reference Materials

The term standard material or RM is used to describe a generic class of stable and well characterized materials with one or more properties or components experimentally determined within stated measurement uncertainties. There is an extensive terminology, even a hierarchy on the basis of traceability, on such less stringently defined RMs. As opposed to a PRM, as defined in the previous section, a secondary or composite RM is a standard whose value is assigned by comparison with a primary reference of the same quantity. Their certification requirements are less stringent than those of the primary references. They can be used for calibration and validation of analytical methods or measuring systems. Composite RMs can be made up artificially from synthetic RMs derived from pure materials or using PRMs. Such synthetic RMs, however, are usually not appropriate for most of the analytical problems in day-to-day practice.

A traceable CRM relying on primary methods is available only in exceptional circumstances, as the exclusive use of primary methods of analysis is impossible in practice.[11] For trace element determinations, the National Institute of Standards and Technology (NIST) employs, whenever possible, methods such as IDMS for its CRMs (NIST SRMs).

CRMs, accompanied by a certificate, are materials in which one or more components are certified by procedures which establish their traceability and whose certified values are accompanied by a stated level of confidence. The stability and homogeneity of such materials must be carefully assessed. They constitute the simplest and most appropriate way for a laboratory to ensure traceability and to verify the analytical procedure. It is important that these CRMs are ''matrix'' CRMs. Matrix CRMs are those of partly known matrix composition, based on natural (nonsynthetic) samples. Using these in analysis for validation purposes, values in disagreement with the certified values provide a warning of errors of which the sources must be identified. Matrix CRMs with a natural composition are preferred in chemical analysis. Their certification can only be done using complex procedures with limited precision compared with those based on primary methods.

The ISO (Guide 35) describes how the production and use of RMs should be carried out as ''the certified value should be an accurate estimate of the true value with a reliable estimate of the uncertainty compatible with the end use requirements''. Many existing CRMs, such as those issued by the EU, use a range of analytical methodologies regardless of whether they are directly linked to the SI system or not, provided that they are properly calibrated.

The majority of existing ''matrix'' CRMs are usually characterized by interlaboratory studies using different analytical techniques. According to the current state-of-the-art in RMs certification, a number of selected participants in a certification deliver a set of replicate measurement results. In the statistical evaluation a consensus value is obtained as the mean of means of the participants' replicate data, excluding outliers. The uncertainty is given as the standard deviation of the mean or a multiple of this (e.g. providing the 95% confidence interval of the true value).

The strict reliance on the SI system by primary methods for the determination of certified concentrations remains illusory in many cases, except for the occasional use of IDMS. Even this primary method of analysis is not fail-proof in this application for matrix CRMs, as chemical equilibrium between the spike and the measurand often cannot be ensured.

CRMs should be the major bridge between metrology laboratories and the broader community. Most CRMs are issued by organizations other than metrology laboratories and are produced through certification campaigns that rely on a plethora of different analytical methods. Hence strict traceability of the data produced according to the SI system is not always ensured.

Hence, such CRMs are consensus items for which it is assumed that the use of different methods of analysis will cancel out systematic variations if independent reference and validated methods are employed by expert analysts, working independently under carefully prescribed and controlled conditions. A typical certification scheme, as applied by the Community Bureau of Reference (BCR) of the Standards, Measurements and Testing (SMT), Program of the EU is shown in Figure 3.[12] A set of data

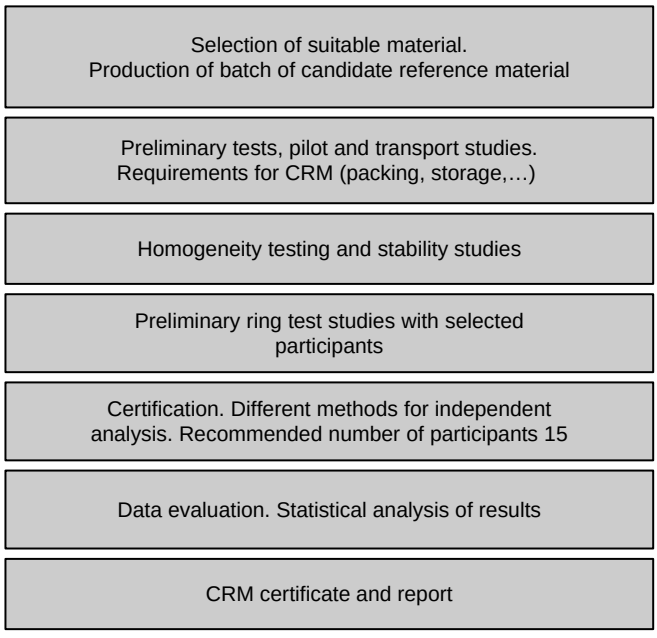

Figure 3 Flow diagram of the steps in the preparation for certification of matrix CRM. (Adapted from Boenke[12] and the 'Guidelines for the Production and Certification of BCR Reference Materials' of the EU, Doc. BCR/48/93, 1994.)

belonging to a specific certification exercise is illustrated in Figure 4.[13]

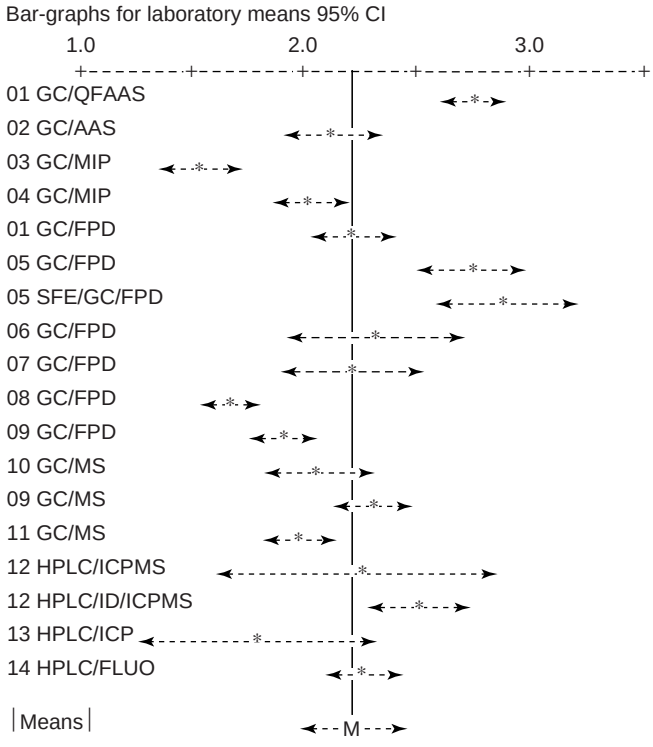

Figure 4 Result of certification of dibutyltin in CRM 477 (organotin compounds in mussel tissue) using results from a number of laboratories. The different methods used are identified by acronyms with hyphenated methods of analysis. (Reproduced by permission from Quevauviller.[13])

3.4 Confidence Limits

In addition of accuracy (absence of systematic errors), the implementation of traceability also concerns stated uncertainty (the coefficient of variation or confidence interval) of analytical measurements. The ''bottom-up'' approach of the ISO consists in the estimation of all the individual error components in a total uncertainty budget and combining the quantified individual uncertainty components.[14] Another approach (top-to-bottom) is based on the participation of a number of laboratories in method performance exercises.[15] Both systematic and random errors occurring between laboratories are considered as random errors and are used to estimate statistically the random measurement error. This method has the advantage over the former approach that unforeseen errors turn up in the uncertainty estimates. It is the method of choice for analytical chemists. Metrology advocates the former.

Analytical chemists are accustomed to report their measurement results together with the standard deviations of individual results (the repeatability). It became clear from many interlaboratory comparison studies that discrepancies (the reproducibility) are consistently much larger than these repeatability estimates. Hence it is necessary that more emphasis is placed on the determination of the uncertainty of analytical methods.[16] Approaches for evaluating total measurement uncertainty may rely on metrology laboratories and on carefully selected analytical laboratories using a standard operating procedure (SOP). CRMs and proficiency tests can be used for this purpose. The total analytical error being larger than the measurement alone, it is necessary that other sources of measurement uncertainty are accounted for, e.g. those resulting from sampling.

Many papers discussing traceability concepts are heavily biased towards analytical determinations that require extremely high accuracy and, hence, full metrological orthodoxy. It is obvious that in specific situations strictly defined concepts must be enforced and adhered to, and even that in some circumstances comparability of measurements is more important than the true analytical result as such. Chemical analysis being a problem-solving tool, its measurement results should provide answers to specific problems in science, society and industry. In defining the measurement process, it is necessary to define the degree of accuracy necessary for ''fitness for purpose''. Economic considerations come in here: it is not necessary that the accuracy of the determination is higher than required for the specific problem at hand,

provided that stated uncertainties include the true value and that measurements are comparable. In many discussions on traceability, measurement errors are insignificant compared with both fitness of purpose requirements and other error sources in analysis or sampling.

4 CONCLUSIONS

In the fields of physical measurements (length, voltage, etc.), straightforward calibration procedures ensuring direct traceability to national and international calibration systems are generally applicable. In chemical analysis, calibration and traceability to the SI system is considerably less obvious. The underlying concepts originating from metrology, i.e. measurement science, tend overly to reduce or simplify the analytical practice to the measurement. The measurement step is only one, often minor, part of the entire analytical process, among many other important parts that contribute to the overall bias of the results and to the confidence limits.

Traceability is not an end in itself, but is just one of the possible ways to foster quality assurance as a drive towards analytical reliability. The concepts as put forward by the CCQM are heavily centered on the measurement part of the overall analytical process. Their practical application is hampered by the adoption of a number of strict rules, so that truly traceable CRMs relying on primary methods are only available in rather exceptional circumstances.

Accuracy remains a central theme in analysis and its pursuit can no longer be the work of a single analytical chemist but requires comparative studies involving many methodologies and many, widely dispersed, experts. It is sensible to expect that pooling different kinds of data (methods and users) will lead to better estimates of the truth, even if some of the contributions are more reliable than others. There is a need for education, to increase the awareness of analysts about quality enforcement and control.

The SMT Program of the European Commission (Fourth Framework Program 1994–1998) included measures for the maintenance and development of metrological systems. It put emphasis on both (i) the accuracy and traceability of measurements to the SI system and (ii) the development of metrology in chemistry and routine analysis, including the development, improvement and validation of analytical steps (sampling, digestion, preconcentration, separation and calibration). The Fifth Framework Program (1999–2002) is less specific and avoids any direct reference to traceability. The Program description makes the following general comment on measurements and testing: "...by ensuring that European standards and testing laboratories provide consistent measurements and tests which are equivalent to similar measurements made by Europe's major trading partners; by making available technical tests, CRMs and measuring instruments for use in Member States in order to assure compliance with Community directives; by supporting standardization and certification...".

Quality control of the laboratories depends on the availability of CRMs, round-robin studies, intercomparisons and proficiency tests between methods and between laboratories. Of special importance is the full knowledge of the complex analytical process and the painstaking pursuit of the true value by the definition of all sources of errors, and the application of an adequate error source budget. Internal quality control (IQC) comprises all procedures used in a laboratory that enable one to assess the quality of each run of analysis produced and accept the results only if they fulfill a predetermined standard with a high probability. In general, IQC schemes operate by the inclusion of traceable control materials and by the duplication of measurements.[15] The pursuit of accuracy in chemical analysis is a matter of quality assurance, in the laboratory and in the system, less one of the concepts of traceability alone as a means for anchoring measurements in time or space. A number of analytical measurement practices as advocated by international organizations such as the European Collaboration in Measurement Standards (EUROMET), Cooperation on International Traceability in Analytical Chemistry (CITAC), International Union of Pure and Applied Chemistry (IUPAC), ISO and International Measurement Evaluation Program (IMEP) to increase accuracy of analysis are given in Table 1.[17]

Organizations requiring a high throughput of analyses are now undergoing a paradigm shift in which performance is specified, not the specific methodologies used. The United States Environmental Protection Agency (USEPA), for example, is at present shifting to a "performance-based measurement system" (PBMS) for the environmental analyses of hazardous wastes. In this system, performance needs (instead of specific measurement technologies) are specified for reliable, cost-effective analyses, thus avoiding costly measurement overkill. Feedback on successes and failures are then used to expand knowledge on new or modified approaches

Table 1 Valid analytical measurement practices (Reproduced by permission from King[17].)

• Work on agreed customer requirements
• Use validated methods/equipment and qualified staff
• Participate in independent assessment of technical performance (proficiency testing)
• Ensure comparability with measurements in other laboratories (traceability and measurement uncertainty)
• Use well-defined quantity control and quality assurance practices

in a flexible working environment and in ''real-world'' conditions. In such types of approaches the generator of the data becomes directly responsible for demonstrating regulatory compliance by defining a sampling/analysis plan and strict record-keeping of proofs of concepts and levels of validation.

ABBREVIATIONS AND ACRONYMS

BCR	Community Bureau of Reference
CCQM	Consultative Committee on Quality of Measurements
CIPM	International Committee for Weights and Measures
CITAC	Cooperation on International Traceability in Analytical Chemistry
CRM	Certified Reference Material
EU	European Union
EUROMET	European Collaboration in Measurement Standards
IDMS	Isotope Dilution Mass Spectrometry
IMEP	International Measurement Evaluation Program
IQC	Internal Quality Control
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
PBMS	Performance-based Measurement System
PRM	Primary Reference Material
RM	Reference Material
SI	Système International
SMT	Standards, Measurements and Testing
SOP	Standard Operating Procedure
SRM	Standard Reference Material
USEPA	United States Environmental Protection Agency

RELATED ARTICLES

Clinical Chemistry **(Volume 2)**
Point-of-care Testing • Statistical Quality Control in Clinical Laboratories

Environment: Water and Waste **(Volume 4)**
Quality Assurance in Environmental Analysis

General Articles **(Volume 15)**
Quality Assurance in Analytical Chemistry

REFERENCES

1. G. Uriago, C.C. Gravatt, 'The Role of Reference Materials and Reference Methods in Chemical Analysis', *CRC Crit. Rev. Anal. Chem.*, **6**, 361–411 (1977).
2. P. De Bievre, R. Kaarls, H.S. Peiser, S.D. Rasberry, W.P. Reed, 'Measurement Principles for Traceability in Chemical Analysis', *Accred. Qual. Assur.*, **1**, 3–13 (1996).
3. IUPAC, 'IUPAC Protocol for the Design, Conduct and Interpretation of Collaborative Studies', *Pure Appl. Chem.*, **60**, 856–864 (1988).
4. M. Thompson, 'Sense and Traceability', *Analyst*, **121**, 285–288 (1996).
5. Y.I. Alexandrov, 'Once More on the Account of Chemical Composition Analysis, Traceability and Calibration in Chemical Analysis', *Fresenius' J. Anal. Chem.*, **357**, 563–571 (1997).
6. W. Ackermann, W. Breitschu, H. Büchler, G. Küppers, C. Lührs, B.J. Schlotmann, G. Staats, R. Zens, 'Primary Reference Materials for Chemical Composition Analysis – Proposals for an Operational Definition', *Fresenius' J. Anal. Chem.*, **357**, 562–564 (1997).
7. F. Adams, 'Traceability and Analytical Chemistry', *Accred. Qual. Assur.*, **3**, 308–316 (1998).
8. R. Kaarls, T.J. Quinn, 'The Comité Consultatif pour la Quantité de Matière: A Brief Review of its Origins and Present Activities', *Metrologia*, **34**, 1–5 (1997).
9. W. Richter, G. Dube, 'Measurement Standards and the General Problem of Reference Points in Chemical Analysis', *Metrologia*, **34**, 13–18 (1997).
10. B. King, 'Traceability of Chemical Analysis', *Analyst*, **122**, 197–204 (1997).
11. R.F. Walker, 'The Work of the ISO–REMCO's Accreditation Task Group', *Fresenius' J. Anal. Chem.*, **360**, 325 (1998).
12. A. Boenke, 'Pure Substances and a New Generation of CRMs for the Chemical Industry', *Fresenius' J. Anal. Chem.*, **360**, 388–392 (1998).
13. P. Quevauviller, L. Ebdon, *Methods Performance Studies for Speciation Analysis*, Royal Society of Chemistry, Cambridge, 1998.
14. *ISO Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1995.
15. Analytical Methods Committee, 'Uncertainty of Measurement: Implications of its Use in Analytical Science', *Analyst*, **120**, 2303–2308 (1995).
16. IUPAC Interdivisional Working Party for Harmonization of Quality Assurance Schemes for Analytical Laboratories, *Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement*, Protocol 1997-02-14, IUPAC, Oxford, 1997.
17. B. King, 'Metrology and Analytical Chemistry: Bridging the Cultural Gap', *Metrologia*, **34**, 41–47 (1997).

Ultrafast Laser Technology and Spectroscopy

Gavin D. Reid
University of Leeds, Leeds, UK

Klaas Wynne
University of Strathclyde, Glasgow, UK

1 Introduction 1

2 Ultrafast Lasers and Amplifiers 2
2.1 Oscillators 2
2.2 Dispersion and Pulse Broadening 3
2.3 Chirped-pulse Amplification 6
2.4 Pulse Recompression 8
2.5 Saturation Effects 9

3 Wavelength Conversion 9
3.1 White-light Generation and the Optical Kerr Effect 9
3.2 Generation of Ultraviolet and X-rays 10
3.3 Optical Parametric Amplifier for Infrared Generation 11
3.4 Noncollinear Optical Parametric Amplification 12
3.5 Terahertz-pulse Generation and Detection 12
3.6 Femtosecond Electron Pulses 14

4 Time-resolved Experiments 14
4.1 Auto- and Cross-correlation 14
4.2 Pump–Probe Techniques 17

5 Applications 18
5.1 The Study of Fast Chemical Reactions 18
5.2 Imaging 20
5.3 Structure Determination: Electron Beams and X-rays 21

Acknowledgments 22

Abbreviations and Acronyms 22

Related Articles 22

References 23

Ultrafast laser technology and spectroscopy involves the use of femtosecond (10^{-15} s) laser and other (particle) sources to study the properties of matter. The extremely short pulse duration allows one to create, detect and study very short-lived transient chemical reaction intermediates and transition states. Ultrafast lasers can also be used to produce laser pulses with enormous peak powers and power densities. This leads to applications such as laser machining and ablation, generation of electromagnetic radiation at unusual wavelengths (such as millimeter waves and X-rays), and multiphoton imaging. The difficulty in applying femtosecond laser pulses is that the broad frequency spectrum can lead to temporal broadening of the pulse on propagation through the experimental set-up. In this article, we describe the generation and amplification of femtosecond laser pulses and the various techniques that have been developed to characterize and manipulate the pulses.

1 INTRODUCTION

Ultrafast spectroscopy has become one of the most active areas of physical chemistry. Rather than postulating mechanisms for chemical and biological reactions, ultrashort laser pulses can now be used to observe and even control the outcome of reactions in real time. Because of our improved understanding of reaction pathways, the ''arrows'' describing purported electronic motion in mechanistic organic chemistry are no longer sufficient.[1] A state-of-the-art laser system can generate 1-J ca. 20-fs pulses and the peak fluence at the focus of these lasers can exceed 10^{20} W cm^{-2}. In contrast, the total solar flux at the Earth is only 10^{17} W. An exciting new era is beginning, which will allow the possibility of using these intense ultrashort laser pulses as sources of short X-ray and electron pulses. These will reveal the positions of atoms as a function of time as reactants proceed to products through the transition states. More routinely, femtosecond lasers can be used to detect and monitor transient chemical species in solution or the gas phase, to image living cells with micrometer resolution, for laser ablation mass spectrometry (MS) and micromachining applications, which will all be of immediate interest to the analytical chemist.

With the invention of flash photolysis in 1950,[2] radical intermediates were observed by light absorption (rather than fluorescence) during the progress of a chemical reaction. When the pulsed laser followed, nanosecond experiments derived from the flash-photolysis technique were to reveal the chemistry of singlet states in solution. However, it was not until the mode-locked ruby[3] and Nd : glass[4] lasers were built in the mid-1960s that the picosecond timescale became accessible and the field of ''picosecond phenomena'' was born. When excited-state processes were studied in the picosecond domain, such as energy redistribution in molecules and proteins, proton and electron transfer reactions, photoisomerization and dissociation, and relaxation in semiconductors, many

Encyclopedia of Analytical Chemistry
R.A. Meyers (Ed.) Copyright © John Wiley & Sons Ltd

measured rate constants were found to be instrument limited. Nevertheless, important results such as the observation of the ''inverted region'' in electron transfer reactions[5–7] and Kramers' turnover in excited-state reactions in solution.[8,9]

However, ultrafast spectroscopy was revolutionized in the 1980s by the invention of the colliding-pulse mode-locked (CPM) dye laser, which generated 100-fs pulses in its early form[10] and 30-fs pulses as the technology was perfected.[11] This ring laser, operating at about 620 nm, coupled with improvements in dye–amplifier chains,[12] allowed the exciting field of femtochemistry to be developed,[13] which will be discussed briefly in section 5.1, and the generation of 6-fs probe pulses,[14] a record which stood until very recently. Self mode-locking in Ti : sapphire-based lasers was discovered[15] in 1990 and the early 1990s brought a new revolution – simplicity of use – and, with it, the commercialization of ultrashort pulse technology. Ti : sapphire oscillators now produce 10–20-fs pulses routinely and 4–5-fs pulses in optimized configurations using mirrors designed to reverse the ''chirp'' introduced by the Ti : sapphire rod. The limiting factor on the exceptional stability of these oscillators is the pump source and here diode-pumped solid-state laser sources are rapidly replacing large, expensive, low-efficiency ion lasers.

In parallel with improvements in oscillator technology, the technique of chirped pulse amplification[16] using solid-state gain media in regenerative or multi-pass schemes has replaced the dye–chain amplifiers of the past. A typical amplifier for routine chemistry applications produces 1 mJ per pulse at a 1-kHz repetition rate or 100 mJ at 10–20 Hz with a duration between 20 and 100 fs. Moreover, a Ti : sapphire laser oscillator and amplifier combination can be purchased in a single box less than 1 m square, operating at mains voltages with no external water-cooling requirements. Total hands-off operation is a reality and, in fact, the complete laser system can be computer controlled.

In the next section, we shall describe the technology behind Ti : sapphire lasers and amplifiers and discuss how light at almost any frequency from X-rays to terahertz can be generated and employed by the chemist. A history of the field can be found in the papers submitted to the biennial conference 'Ultrafast Phenomena', the proceedings of which are published in the *Springer Series in Chemical Physics*, which is now in its eleventh volume.[17]

2 ULTRAFAST LASERS AND AMPLIFIERS

2.1 Oscillators

Ultrashort pulses are generated by mode-locked lasers. By constructive interference, a short pulse is formed when many longitudinal modes are held in phase in a laser resonator. Various techniques have been employed, usually grouped under the terms ''active'' or ''passive'' mode locking and descriptions of these can be found in many standard texts[18,19] and review articles.[20] Active mode locking uses a modulator in the laser cavity whereas passive schemes use a saturable absorber, often a thin semiconductor film, to lock the relative phases. Modern solid-state mode-locked lasers use a different scheme called self-mode-locking and titanium-doped sapphire (Ti : sapphire) has become by far the most common laser material for the generation of ultrashort pulses. Developed in the mid-1980s,[21] Ti : sapphire has a gain bandwidth from 700 to 1100 nm peaking around 800 nm, the broadest of the solid-state materials yet discovered, high-gain cross-section and extremely good thermal conductivity. Mode locking is achieved through the action of an instantaneous nonlinear Kerr lens in the laser rod (see section 3.1). The peak fluence of the laser approaches 10^{11} W cm^{-2}, which is enough to focus the beam as it travels through the gain medium on each pass. This Kerr lens then couples the spatial and temporal modes and maintains phase locking.

A basic oscillator–cavity configuration[22] is shown schematically in Figure 1. The laser is pumped by about 5 W from a continuous wave (CW) laser source, usually now an intracavity-doubled diode-pumped neodymium laser. This light is focused into the Ti : sapphire rod, collinearly with the laser axis, through the back of one of the mirrors. The cavity consists of a Brewster-angle cut Ti : sapphire rod, 5 mm or less in length, doped to absorb about 90% of the incident pump radiation, two concave focusing mirrors placed around it, a high reflector and an output coupler. A pair of Brewster-cut fused-silica prisms are inserted to control the spectral dispersion (chirp) introduced in the laser rod. Dispersion arises from the variation of the refractive index of the material across the gain bandwidth of the laser, which can lead to a temporal separation of the resonant wavelengths and

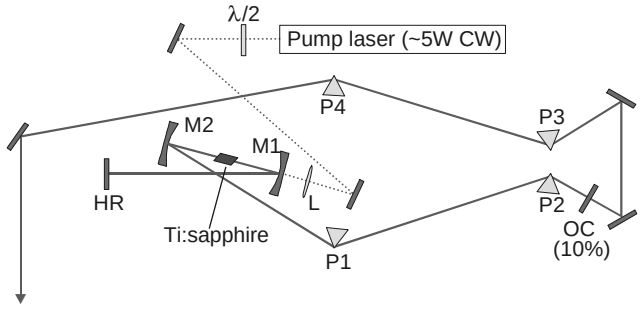

Figure 1 Diagram of a basic self-mode-locked Ti : sapphire oscillator showing the cavity layout. The pulse is coupled out from the dispersed end of the cavity, which requires a pair of matching extracavity prisms.

place a limit on the generated bandwidth (see section 2.2). The cavity dispersion, coupled to the Kerr lens effect, is an intrinsic part of the pulse-formation process. Ordinarily, the Kerr lens in the rod would contribute to the overall loss but this is overcome by a small adjustment to the resonator. Displacement of one of the curved mirrors by only ca. 0.5 mm pushes the cavity into pulsed mode. Here the cavity is corrected for the nonlinear lens effect and CW operation is restricted. Pulsing is established by perturbing the cavity to introduce a noise spike, literally by tapping a mirror mount. This configuration delivers 12-fs pulses centered at 800 nm with 5 nJ energy at an 80-MHz repetition rate. Using a shorter rod, shorter pulses have been obtained[23] as the dispersion is better compensated and space–time focusing effects are controlled. The laser repetition rate can be adjusted by the insertion of a cavity dumper in a second fold, without prejudice to the pulse duration.[24] The use of mirrors that have dispersion opposite to that of the rod[25] obviates the need for prisms entirely. Alternatively, a mixture of prisms and mirrors can be used to generate pulses as short as 5 fs.[26] At the time of writing, the use of mirrors with well-defined chirp characteristics is complicated by the demand for extreme tolerances in the manufacturing process. Many schemes have been proposed for self-starting oscillators, perhaps the best of which is the use of a broadband semiconductor saturable-absorber mirror in the cavity.[27] Advances in these areas will surely continue.

Other important solid-state materials include Cr : Li-SAF (chromium-doped lithium–strontium–aluminum fluoride), which can be pumped by red diode lasers and operates close to the peak Ti : sapphire wavelength, and Cr^{4+} : YAG, which lasers at around 1.5 μm, an important communication wavelength. Cr : forsterite lases, operating at about 1.2 μm, can be frequency doubled to the visible region and have been used for imaging applications (see section 5.2). Furthermore, passively mode-locked frequency-doubled erbium-doped fiber lasers have been developed commercially. These lasers operate between 1530 and 1610 nm and efficient frequency doubling to 765–805 nm is possible in periodically poled nonlinear crystals (i.e. these lasers can be used in place of Ti : sapphire for many applications).[28] They have the advantage of being cheap and extremely compact since they do not require complicated dispersion-compensation schemes, owing to the soliton nature of the pulses supported in the fiber. They can also be pumped using cheap large-area telecommunications-standard laser-diode sources. Unfortunately, the pulse duration is limited to a minimum of about 100 fs.

2.2 Dispersion and Pulse Broadening

A bandwidth-limited pulse has a spectral width given by the Fourier transform of its time-domain profile.

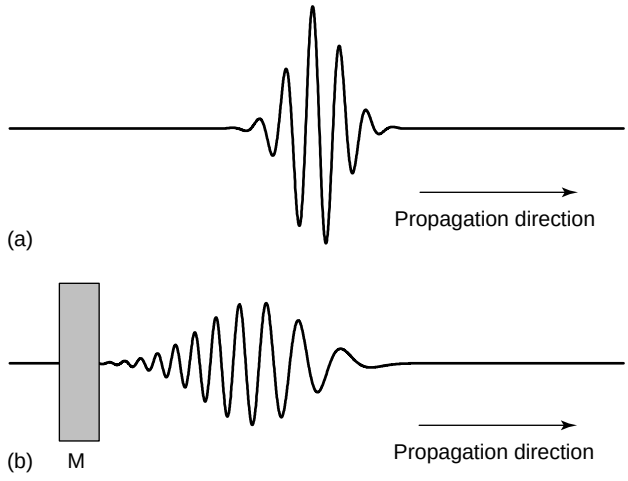

Figure 2 Schematic diagram of the electric field of (a) an undispersed Gaussian pulse and (b) the same pulse after traveling through a positively dispersing medium, M. A frequency sweep from low to high frequency (right to left) can be observed on the trace in (b).

Consequently, a 10-fs full width at half-maximum (fwhm) Gaussian pulse centered at 800 nm has a bandwidth of 94 nm (1466 cm^{-1}). When a short pulse travels through a dispersive medium, the component frequencies are separated in time. Figure 2(a) and (b) shows the effect of dispersion on a Gaussian pulse traveling through a piece of glass. There are two points to note. First, the center of the pulse is delayed with respect to a pulse traveling in air. This is usually called the group delay, which is not a broadening effect. Second, normally dispersive media such as glass impose a positive frequency sweep or "chirp" on the pulse, meaning that the blue components are delayed with respect to the red.

In order to get a physical feel for the effect of the chirp, it is common to consider the phase shift as a function of frequency w. The phase, $\varphi(w)$ can be developed as a power series about the central frequency w_0, assuming the phase varies only slowly with frequency according to Equation (1):

$$\varphi(w) = \varphi(w_0) + (w - w_0)\varphi'(w_0) + \tfrac{1}{2}(w - w_0)^2\varphi''(w_0) + \tfrac{1}{6}(w - w_0)^3\varphi'''(w_0) + \tfrac{1}{24}(w - w_0)^4\varphi''''(w_0) + \cdots \quad (1)$$

where (Equation 2):

$$\varphi'(w_0) = \left.\frac{\partial\varphi}{\partial\varphi}\right|_{w=w_0}, \varphi''(w_0) = \left.\frac{\partial^2\varphi}{\partial w^2}\right|_{w=w_0},$$

$$\varphi'''(w_0) = \left.\frac{\delta^3\varphi}{\delta w^3}\right|_{w=w_0}, \text{etc.} \quad (2)$$

φ' is the group delay, φ'' the group delay dispersion (GDD) [or group velocity dispersion (GVD)] and φ'''

and φ'''' are simply the third-order dispersion (TOD) and fourth-order dispersion (FOD).

For the sake of simplicity, consider a transform-limited Gaussian pulse with a central frequency w_0 and a pulse width (fwhm) τ_{in}. Then its electric field, E_{in}, takes the form (Equation 3)

$$E_{in}(t) = E_0 \exp\ -\left(\frac{2\ln 2t^2}{{\tau_{in}}^2}\right. \quad + iw_0t \tag{3}$$

The electric field after traveling through a dispersive medium can be found by transforming E_{in} to the frequency domain and adding the components from the phase expansion $\varphi(w)$ in Equation (1) before transforming back. If we assume only a contribution from the GDD term, the electric field E_{out} is given by Equation (4):[18]

$$E_{out}(t) = E_0 \exp\ \ i(w_0t - \varphi) - \Gamma(t - \varphi')^2 \tag{4}$$

where (Equation 5)

$$\Gamma = \left(\frac{\tau_{in}^2}{2\ln 2} + 2i\varphi''\right.^{-1} \tag{5}$$

The effects of this dispersion are twofold. First, by inspection of the Gaussian part of E_{out}, τ_{out} is analogous to τ_{in} from E_{in} and is broadened with respect to the input-pulse width by a factor (Equation 6)

$$\frac{\tau_{out}}{\tau_{in}} = \sqrt{1 + \frac{\varphi''^2}{\tau_{in}^4}16(\ln 2)^2} \tag{6}$$

Second, a frequency sweep is introduced in the output pulse (because the expression in Equation (5) is complex) with a sign that is opposite to that of φ''.

The GDD, φ''_m, due to material of length l_m, is related to the refractive index of the material, $n(\lambda)$, at the central wavelength, λ_0, through its second derivative with respect to wavelength (Equation 7):

$$\varphi''_m = \frac{\lambda_0^3 l_m}{2\pi c^2}\frac{d^2 n(\lambda)}{d\lambda^2} \tag{7}$$

Figure 3(a) shows the variation of refractive index with wavelength for some common materials and Figure 3(b) shown the GDD. The data were obtained from glass suppliers and the fits are to Sellmeier-type equations, which are then used to take the numerical derivatives for use in subsequent calculations.

Using Equations (6) and (7) together with refractive index data, we can calculate values for the GDD arising from a length of material. Figure 4 shows the effect of 10 mm of fused silica on a short pulse. Silica is one of the least dispersive materials available and 10 mm is chosen to represent one or two optical components, which might be part of an experimental arrangement. If we consider a pulse of around 100 fs in duration, the effect is minimal but visible. However, a 10-fs pulse is broadened by more than a factor of 10!

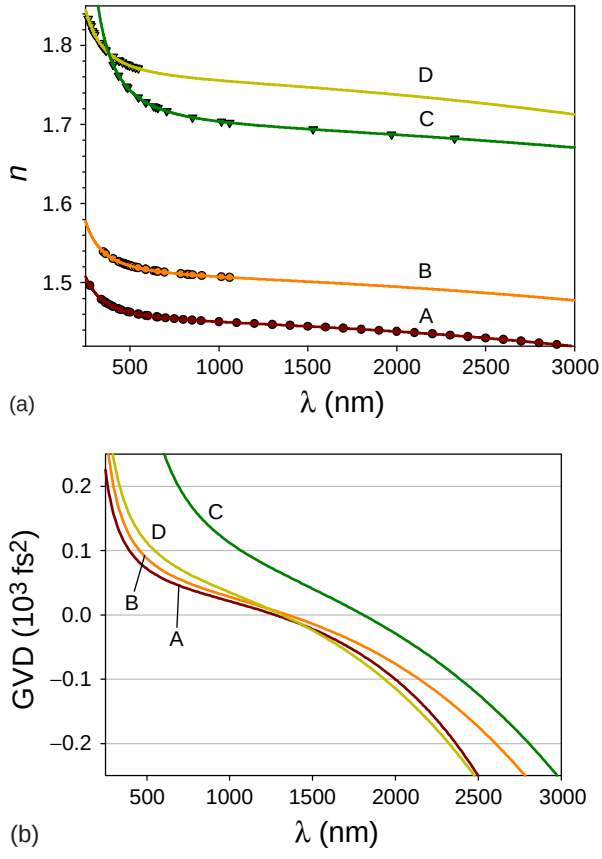

Figure 3 (a) Refractive index versus wavelength data for some common materials: (A) fused silica, (B) Schott BK7, (C) Schott SF10 and (D) sapphire. The points are measured values and the lines are fits to Sellmeier equations. (b) The GDD for 1 mm of material derived from the fits.

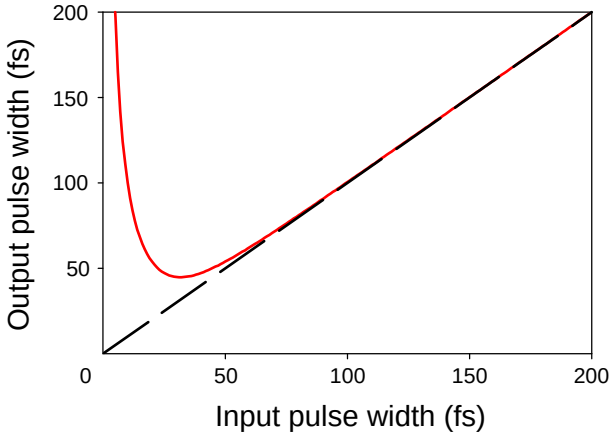

Figure 4 Gaussian pulse width before and after 10 mm of fused silica (solid line), corresponding to one or two optical components. The broadening is due to GDD.

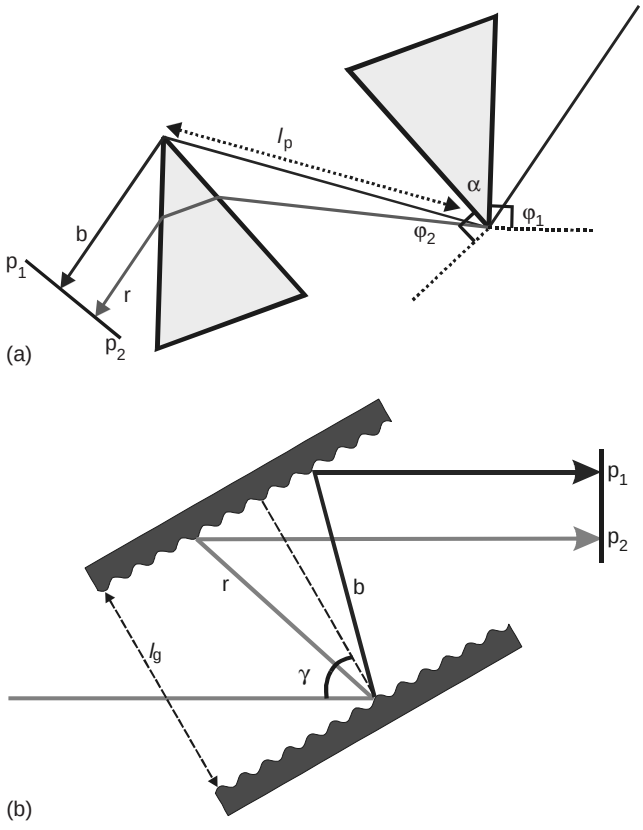

Figure 5 (a) Prism and (b) grating pairs used in the control of dispersion; r and b indicate the relative paths of arbitrary long- and short-wavelength rays. φ_1 is the (Brewster) angle of incidence at the prism face. The light is reflected in the plane $p_1 - p_2$ in order to remove the spatial dispersion shown.

A good understanding of dispersion is essential in order to deliver a short pulse to the sample and careful control of the phase shift is necessary. Fortunately, a number of designs using prism and grating pairs have been devised whereby this can be achieved.[29] The two most important schemes are shown in Figure 5(a) and (b). The Ti : sapphire oscillator we discussed above uses the prism pair.[30] This arrangement creates a longer path through the prism material for the red wavelengths compared with the blue, introducing a negative dispersion. Provided that the prism separation, l_p (defined tip to tip), is sufficiently large, the positive dispersion of the material can be balanced. The prism apex angle is cut such that at minimum deviation of the center wavelength, the angle of incidence is the Brewster angle. Here, the Fresnel reflection losses for the correct linear polarization are minimized and the system is essentially loss free. The second scheme is the parallel-grating pair[31] and, again, a longer path is created for the red over the blue. Grating pairs (Figure 6) introduce negative GDD at very modest separation leading to compact designs but suffer from losses of close to 50% in total. Both prism and grating

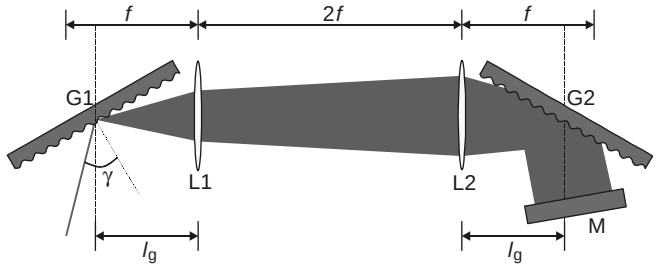

Figure 6 Schematic diagram of a grating-pair pulse stretcher showing the arrangement for positive dispersion. G1 and G2 are diffraction gratings, L1 and L2 are identical lenses separated by twice their focal length, f, and M is a mirror acting to double-pass the beam through the system. The distance $l_g - f$ determines the total dispersion.

pairs are used in a double-pass arrangement to remove the spatial dispersion shown in the diagram.

Expressions for calculating the dispersion are given in Table 1. The equations look a little daunting but the dispersion can be modeled easily on a personal computer. To illustrate this point, Figure 7 shows the total GDD and TOD arising from 4.75 mm of sapphire balanced against a silica prism pair separated by 60 cm. This is typical of a Ti : sapphire oscillator. The net GDD is nearly zero at 800 nm but the net TOD remains negative at the same wavelength. In fact, it is a general observation that prism compressors overcompensate the third-order term. The greater the dispersion of the glass, the less distance is required, but the contribution of the third-order term increases. Experimentally, there must usually be a compromise between prism separation and material.

Grating pairs are important in amplification and will be considered in more detail below. Significantly, however, the sign of the third-order contribution from the grating pair is opposite to that of the prisms, allowing a combined approach to dispersion compensation, which has been used to compress pulses in the 5-fs regime.[14] It can be more useful to think in terms of the total GDD versus wavelength with a view to keeping the curve as flat and as close to zero as possible across the full bandwidth of the pulse.

Mirror coatings have been developed which can provide second- and third-order compensation. This so-called ''chirped'' mirror reflects each wavelength from a different depth through the dielectric coating, which is made up of multiple stacks of varying thickness. In combination with a prism pair, this technique has been as successful as the grating/prism combination without the associated losses. This type of system will become more common as the mirrors, tailored to individual requirements, become available commercially.

Table 1 Expressions (Equations 8–17) for the dispersion for material, prism and grating pairs

The accumulated phase in a double-pass *prism compressor* (see Figure 5a) is

$$\varphi(w) = \frac{2wl_p}{c}\cos\left(\varphi_2^{\text{bluest}} - \varphi_2\right) \tag{8}$$

where φ_2 is the (frequency-dependent) exit angle, $\varphi_2^{\text{bluest}} = \varphi_2(w_{\text{bluest}})$ is the exit angle of the shortest-wavelength light transmitted by the prism pair:

$$\varphi_2(w) = \arcsin\{n\sin[\alpha - \arcsin(n^{-1}\sin\varphi_1)]\} \tag{9}$$

where n is the (frequency-dependent) refractive index of the prism material, φ_1 is the (frequency-independent) angle of incidence on the first prism and α is the prism top angle. If the prism compressor is designed for frequency w_{design}:

$$\alpha = 2\arcsin\frac{1}{\sqrt{1+n_{\text{design}}^2}} \tag{10}$$

$$\varphi_1 = \arcsin\ n_{\text{design}}\sin\frac{\alpha}{2} \tag{11}$$

where n_{design} is the refractive index of the prism at the design wavelength. Expressions for the group delay, GVD, TOD and FOD can be obtained by taking derivatives of Equation (8).

The group delay in a *grating compressor/stretcher* (see Figures 5b and 6) is

$$\frac{d\varphi}{dw} = \frac{P}{c} \tag{12}$$

where the optical pathlength is

$$P = L\frac{1+\cos\theta}{\cos(\gamma-\theta)} \tag{13}$$

$$\theta(w) = \gamma - \arcsin\left(\frac{2\pi c}{wd} - \sin\gamma\right. \tag{14}$$

with γ the angle of incidence on the first grating and d the groove frequency. In the case of a grating *compressor*, $L = l_g$. In the case of a grating *stretcher*, $L = 2(l_g - f)\cos\gamma$.

Expressions for the GVD, TOD and FOD can be obtained by taking derivatives of Equation (12). For a double-pass grating compressor/stretcher, the analytical expression

$$\frac{d^2\varphi}{dw^2} = \frac{-\lambda^3 L}{\pi c^2 d^2}\ 1 - \left(\frac{\lambda}{d} - \sin\gamma\right.^{\ 2\ -3/2} \tag{15}$$

is found, where λ is the wavelength.

The accumulated phase in *material* is

$$\varphi(w) = \frac{wnl_m}{c} \tag{16}$$

where l_m is the length of the material. Expressions for the group delay, GVD, TOD and FOD can be obtained by taking derivatives of Equation (16). An analytical expression for the GVD is

$$\frac{d^2\varphi}{dw^2} = \frac{\lambda^3 l_m}{2\pi c^2}\frac{d^2n}{d\lambda^2} \tag{17}$$

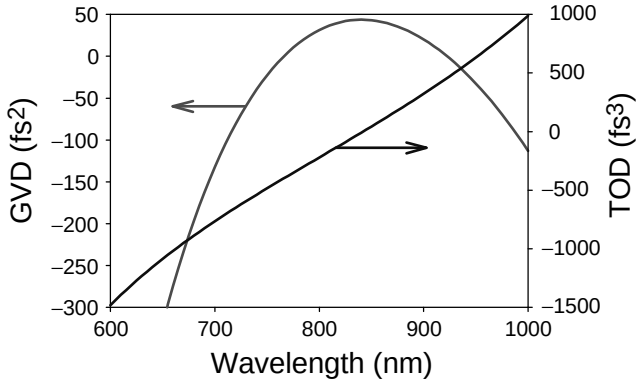

Figure 7 The residual GVD and TOD arising from the balance between the dispersion of a 4.75-mm Ti : sapphire laser rod and the intracavity silica prism pair separated by 60 cm.

2.3 Chirped-pulse Amplification

The amplification of nanojoule-level femtosecond pulses to the millijoule level and above is complicated by the extremely high peak powers involved. A 1-mJ 20-fs pulse focused to a 100-μm spot size has a peak fluence of $5 \times 10^{12}\ \text{W cm}^{-2}$. The damage threshold of most optical materials is only a few gigawatts per square centimeter, 1000 times lower. The problem is overcome by stretching the pulse in time using dispersion to advantage. This is followed by amplification and subsequent recompression to the original pulse duration. This technique also has the benefit of eliminating unwanted nonlinear effects in the amplifier materials. A diagram showing the principle of chirped-pulse amplification (CPA) is shown in Figure 8.

The pulse stretcher is a variation of the grating pair described above.[32] A unity-magnification telescope is placed between two gratings in an antiparallel rather than parallel geometry. This reverses the sign of the dispersion

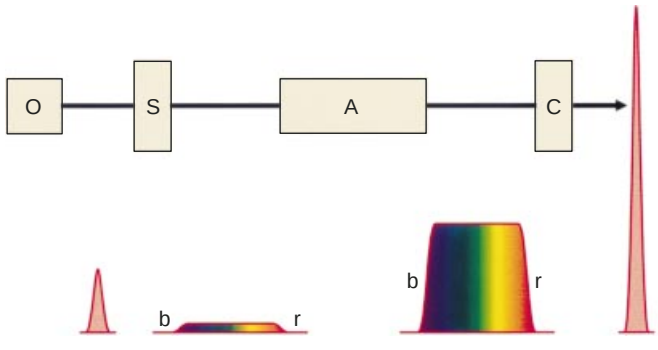

Figure 8 Diagram showing the principle of CPA. The oscillator output (O) is stretched in the grating stretcher (S) such that the red frequency components (r) travel ahead of the blue (b). The peak intensity is reduced in the process. The stretched pulse is then amplified in a regenerative or multipass amplifier (A) before recompression in a grating-pair compressor (C).

of the grating pair but otherwise the mathematical expressions given in Table 1 are identical.

The stretching factor is defined by the effective grating separation $L = 2(l_g - f)$, where f is the focal length of the lens and l_g the distance from the lens to the grating. Typically, the pulse duration is increased to 100 ps or more for efficient extraction of the stored energy. When l_g is equal to f, there is no dispersion, and when l_g becomes larger than f, the dispersion changes sign. In practice, the lenses are replaced by a single spherical or parabolic mirror in a folded geometry, which eliminates chromatic aberration and allows gold-coated holographic gratings to be used near their most efficient Littrow angle of incidence, $\arcsin(\lambda_c/2d)$, where λ_c is the central wavelength and d the line separation of the grating. Other more sophisticated stretcher designs, one for example based on an Offner triplet,[33] have been made but are outside the scope of this discussion. Ignoring the amplifier for now, the pulse is recompressed using an identical parallel grating pair separated by $2l_g$.

CPA technology developed rapidly during the 1990s. Solid-state materials usually have long upper-state radiative lifetimes compared with laser dyes. The large saturation fluence (1 J cm^{-2}) and long storage time (3 μs) of Ti : sapphire make it an ideal amplifier gain material. Here we shall consider two basic schemes, the regenerative and the multipass amplifier. These operate either at 10–20 Hz, pumped by a standard Q-switched Nd : YAG laser giving up to 100 mJ at 532 nm, or at 1–5 kHz pumped by an intracavity-doubled acousto-optically modulated CW Nd : YLF laser, which usually provide more than 10 W at 527 nm in a 200-ns pulse. These pump lasers are normally flash-lamp pumped but diode-pumped equivalents have appeared commercially in recent months. A good spatial mode quality is essential and a clean top-hat profile is ideal.

Figure 9(a) and (b) shows two arrangements for regenerative amplification. The arrangement in Figure 9(a) is often used at 1 kHz.[34] Briefly, a single vertically polarized pulse from the oscillator, stretched to ~100 ps, is injected into the amplifier using a fast-switching PC. This is performed by stepping the voltage in two stages, first by a quarter wave, in order to trap the pulse in the amplifier cavity, and then up to a half wave for ejection. Typically, the pulse makes around 12 round trips in the cavity before the gain is saturated. A FR is used to isolate the output pulse from the input. The arrangement in Figure 9(b)[35] differs in two respects. First, the focusing in the cavity is relaxed in order to remain near the saturation fluence for the more energetic pulses at 10 Hz. Second, the PC is used to switch the pulse in and out while it is traveling in opposite directions. This has two great advantages for short pulses: (1) there is only one pass made through the Faraday isolator, which has extremely large dispersion

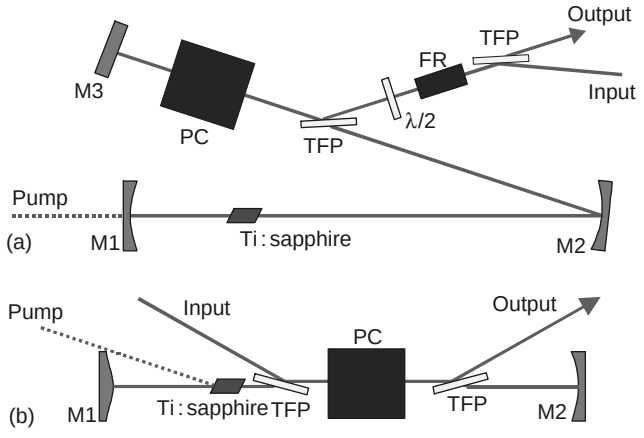

Figure 9 Two schemes for regenerative amplification. The design in (a) is often used for kilohertz repetition rate amplifiers and that in (b) at a 10–20-Hz repetition rate. The Ti : sapphire rod is usually ca. 20 mm long and doped for 90% absorption. TFP, thin-film polarizing beamsplitter; PC, Pockels cell; FR, Faraday rotator; λ/2, half-wave plate. In (a), M1 is 150 mm radius of curvature, M2 is 1 m and M3 is flat. In (b) M1 is −20 m and M2 is +10 m.

and only a limited spectral bandwidth, and (2) it is only necessary to apply a half-wave voltage to the PC at the moment the pulse is switched in or out of the cavity. Again, this is to avoid bandwidth-limiting effects.

An alternative and perhaps more straightforward design for amplification of femtosecond pulses is based on the multipass scheme that has been used in the past with dye amplifiers. One of the best arrangements for use at kilohertz repetition rates[36] is shown in Figure 10 and an example of a 10-Hz system can be found in a paper by Chambaret et al.[37] A PC is used to inject a single pulse from the 80-MHz pulse train into the amplifier, where it is allowed to make about eight passes with a slight offset at each cycle before being picked off and ejected. The

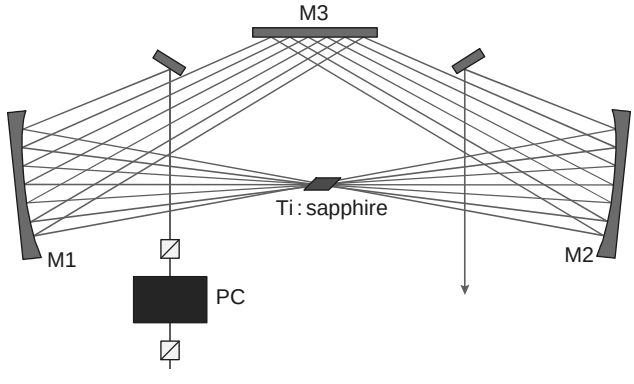

Figure 10 Multipass-amplifier arrangement. M1, M2 are 1-m radius of curvature mirrors and M3 is flat and up to 15 cm wide. A PC and a pair of polarizers are used to inject a single pulse into the amplifier.

pulse incurs significantly less loss in these arrangements, accrues much less chirp and only one pass is made through the PC. There is also no need for a Faraday isolator.

Additional power-amplification stages may be added to increase the pulse energy further. The ring configuration can be modified by arranging for the beams to cross away from the focus to achieve a different saturation fluence. In this configuration, 4 W after recompression at 1 kHz has been obtained.[38] At kilohertz repetition rates, the presence of a thermally induced refractive index change across the spatial profile of the beam must be avoided in order to achieve diffraction-limited output. This is achieved either by cooling to 120 K where the thermal properties of sapphire are much improved[39] or by using the lens to advantage.[40] At lower repetition rates, the medium has time to recover and a thermal lens is usually not established. In this regime, power amplifiers can be added to the limit of available pump energy. An additional complication requires that short pulses of more than a few millijoules in energy must be recompressed under vacuum to avoid nonlinear effects and ionization of the air.

2.4 Pulse Recompression

Clearly, the amplification process introduces extra dispersion and one of the major considerations in system design is the recompression process. The naive approach is to add additional separation between the gratings in the pulse compressor and this is how the first systems were built. Unfortunately, the stretcher and compressor combination is the most dispersive part of the system and a mismatch between the two introduces vast third- and fourth-order contributions to the phase expansion. The easiest way to correct the TOD is to adjust the angle of incidence between the stretcher and compressor, which changes the third-order contribution. Figure 11 (line A) shows the net group-delay dispersion versus wavelength for the kilohertz regenerative amplifier discussed above. The GDD curve is essentially flat over only a narrow wavelength range although the GDD and TOD are zero simultaneously at the center wavelength. Table 2 shows the bandwidth (fwhm) of Gaussian pulses of different durations. Ideally, the GDD curve should be flat over perhaps twice the bandwidth to avoid phase distortions.

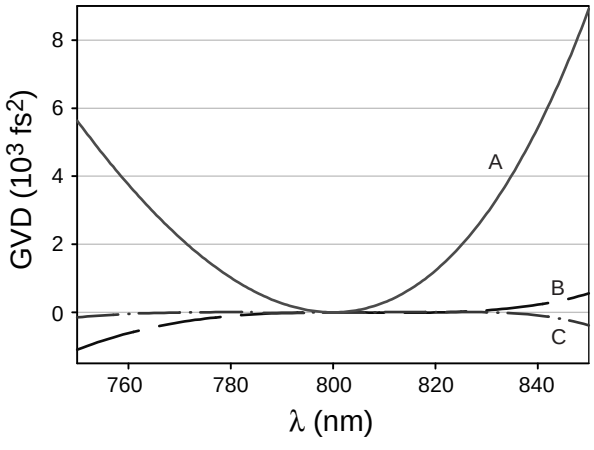

Figure 11 Solid line (A): the net GVD due to the stretcher–amplifier–compressor combination using 1200 lines mm^{-1} diffraction gratings. The difference between the angle of incidence in the stretcher compared to the compressor is 10.32° and the extra grating displacement is 69.4 mm. Dashed line (B): the same using 1200 lines mm^{-1} gratings in combination with an SF18-glass prism pair. The difference between the angle of incidence in the stretcher compared with the compressor is 3.4° and the extra grating displacement is 26.2 mm. The prism separation is 2.4 m. Dashed-dotted line (C): the same, using a 1200 lines mm^{-1} diffraction grating in the stretcher and an 1800 lines mm^{-1} grating in the compressor in combination with an SF10-glass prism pair. The angle of incidence in the stretcher is 13° and in the compressor 39°. Littrow is 41°. The prism separation is 1.65 m.

Recall that the third-order contribution from prism pairs has the opposite sign to gratings.[41] This fact can be used to null the GDD, TOD and FOD terms and this is illustrated in Figure 11 (line B). The difference is dramatic and this system will support much shorter pulses at the expense of 2.4 m of pathlength between the SF18 prism pair.

Another method in the recent literature[42] uses gratings of different groove density in the compressor compared with the stretcher balanced against additional round trips in the amplifier. This method, while better than the angular adjustment technique, is not as effective as the grating–prism combination approach, although it has found favor in commercial application owing to its inherent compactness. Typical parameters are a 1200 lines mm^{-1} grating in the stretcher at an angle of incidence of only 6° and 2000 lines mm^{-1} in the compressor at 57°, which is reasonably close to Littrow geometry (55°). Gratings with 2000 lines mm^{-1} are significantly more efficient than those with 1200 lines mm^{-1}, giving improved throughput in the compressor. However, a combination of mixed gratings and a prism pair as shown in Figure 11 (line C) is a little better than matched gratings and prisms with the benefit of extra throughput and reduced prism separation without the need to balance the

Table 2 Pulse duration versus bandwidth (fwhm) for Gaussian-shaped pulses centered at 800 nm

Gaussian pulse duration (fwhm) (fs)	Gaussian bandwidth (fwhm) (nm)
100	9.4
50	18.8
30	31.3
10	94

total dispersion with extra amplifier material.[43] This system will support 20-fs pulses with little phase distortion.

2.5 Saturation Effects

The gain cross-section, σ_g, is not constant as a function of wavelength. Since this appears as an exponent in calculating the total gain, successive passes through the amplification medium can lead to pulse spectral narrowing and a shift in the central wavelength. This limits the maximum bandwidth to 47 nm in Ti : sapphire.[38] As the gain approaches saturation, the leading red edge of the pulse extracts energy preferentially and the spectrum will red shift. The maximum gain bandwidth can be achieved by seeding to the blue of the peak, allowing the gain to shift towards the maximum on successive passes. Second, by altering the gain profile by discriminating against the peak wavelengths using an étalon or a birefringent filter, termed regenerative pulse shaping, the theoretical limit can be overcome.[44]

3 WAVELENGTH CONVERSION

Ultrafast lasers and amplifiers typically operate at a very limited range of wavelengths. For example, Ti : sapphire-based ultrafast lasers are tunable in the near-infrared (NIR) region from about 700 to 1000 nm but typically work best at about 800 nm. The high peak power of these lasers can be used, however, to convert the laser light to different wavelengths. In fact, in some cases ultrafast laser systems may be the ideal or only route to make radiation at certain wavelengths. Below, a series of techniques are described to convert femtosecond laser pulses at visible wavelengths to other wavelengths.

3.1 White-light Generation and the Optical Kerr Effect

At high intensities such as on the peak of an ultrashort laser pulse, the refractive index of any medium becomes a function of the incident intensity. This effect, which is often referred to as the optical Kerr effect (OKE),[45] can be described by Equation (18):

$$n(I) = n_0 + n_2 I + \cdots \tag{18}$$

where n_0 is the normal refractive index of the medium and n_2 is the nonlinear refractive index. The nonlinear refractive index is very small, for example, in fused silica $n_2 \approx 3 \times 10^{-16}\,\mathrm{cm^2\,W^{-1}}$. A laser pulse with center frequency w traveling through a medium of length L will acquire an optical phase wLn/c and therefore the effects of the nonlinear refractive index will become important when this phase factor becomes comparable to a wavelength. As an example, consider the propagation of femtosecond pulses through an optical fiber. With a 9-μm fiber-core diameter (typical for communications-grade fiber), 10-nJ pulse energy and 100-fs pulse width, the peak power is 100 kW, corresponding to a power density of $1.5 \times 10^{11}\,\mathrm{W\,cm^{-2}}$. If the length of the fiber is 1 cm, the optical pathlength changes by $Ln_2I = 472$ nm or half a wavelength. From this calculation, it can be seen that the OKE can have a significant effect on a femtosecond pulse traveling through a medium. For pulses with energies in the order of millijoules or higher, even the nonlinear refractive index of air becomes important.

To understand how the OKE can modify the spectral properties of an ultrashort pulse, one has to consider how the nonlinear refractive index modifies the optical phase of the pulse. The electric field of a laser pulse traveling in the x direction can be written as Equation (19):

$$E(t)\cos(wt - kx) \tag{19}$$

where $k = wn/c$ is the wavenumber. Since the wavenumber depends on the (nonlinear) refractive index of the medium, the pulse will acquire a time-dependent phase induced by the Kerr effect. Inserting Equation (18) into Equation (19) and performing a Taylor expansion around the peak of the ultrashort pulse, one finds (Equation 20)

$$E(t)\cos[(w + \xi t)t - kx] \tag{20}$$

where ξ depends on the nonlinear refractive index, the pulse width and peak power, and the distance traveled. Consequently, the nonlinear refractive index induces an approximately linear frequency sweep (or chirp, see section 2.2 and Figure 12) on the pulse. In other words, the spectrum of the pulse has broadened owing to the nonlinear interaction. If a single pulse modifies its own characteristics in this way, the effect is often referred to as self-phase modulation. If one pulse modifies the effective refractive index causing a second pulse to change its characteristics, this is referred to as cross-phase modulation.

The spectral broadening induced by the nonlinear refractive index is extremely useful in spectroscopic applications. For example, an 800-nm femtosecond pulse can be sent through a short length of fiber or through a few millimeters of glass or sapphire to produce a broadband output pulse. Often there will be significant power at wavelengths ranging from 400 nm to 1.6 μm. For this reason, such spectrally broadened pulses are referred to as white-light continuum pulses. The white-light continuum generated in a fiber has been used to generate some of the world's shortest pulses of around 5 fs.[46,47] A white-light continuum pulse is an ideal seed for an optical parametric amplifier (see section 3.3).

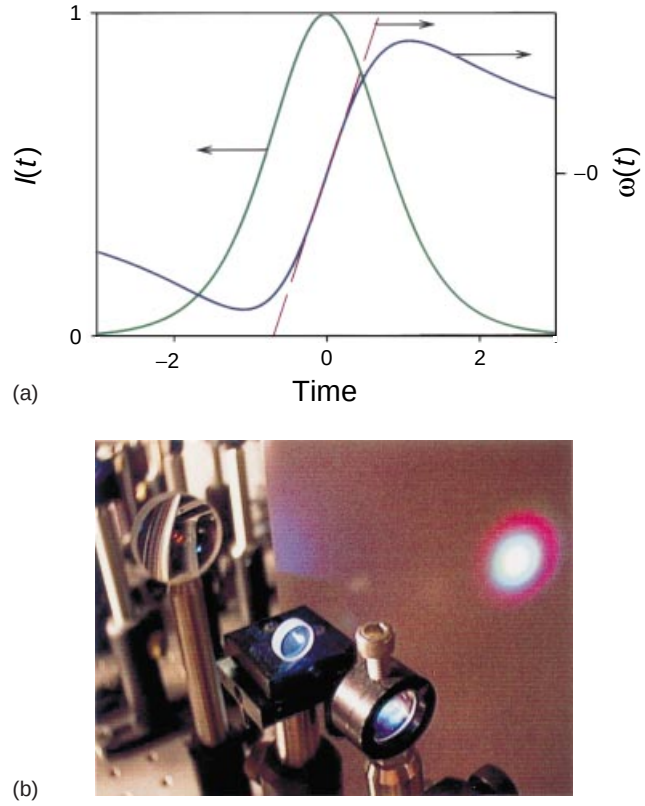

Figure 12 (a) Chirp on an ultrashort pulse induced by the nonlinear refractive index of a dielectric medium. The input pulse has a sech^2 envelope and the figure shows the instantaneous frequency within the pulse. A Taylor expansion around the peak of the pulse shows that the frequency sweep is approximately linear around time zero. (b) Photograph of a femtosecond white-light continuum beam generated in a piece of sapphire. (Picture courtesy of the Center for Ultrafast Optical Science, University of Michigan.)

An implication of the nonlinear refractive index is self-focusing or defocusing. As a laser beam is typically more intense in its center, the nonlinear change of the refractive index will be strongest in the center. As a result, the medium will act as an intensity-dependent lens. In a set-up in which the sample is translated through the focus of a beam, this effect can be used to measure the nonlinear refractive index of a sample quickly. Self-focusing is the basis of the Kerr lens mode locking (KLM) effect used in ultrashort lasers (see section 2.1). Self-focusing can become a run-away process leading to beam distortion and catastrophic damage to optical components. In white-light generation, self-focusing can result in the beam breaking up into multiple filaments that make the white-light output extremely unstable. It is therefore of the utmost importance to choose the incident power such that white light is generated without producing multiple filaments. For a 100-fs pulse, this usually means that the pulse energy should be limited to approximately 1 μJ.

3.2 Generation of Ultraviolet and X-rays

The very high peak power that can be achieved with femtosecond pulses means that in principle nonlinear frequency conversion should be very efficient. It should be straightforward to use second-harmonic generation (SHG), third-harmonic generation (THG) and fourth-harmonic generation (FHG) to produce femtosecond pulses in the near- to deep-ultraviolet (UV) region. However, the group velocity of the pulses (see section 2.2) depends on the center wavelength and changes significantly as the UV region is approached. For example, in a 1-mm β-barium borate (BBO) crystal (suitable for harmonic generation down to about 180 nm) the difference in group delay between the fundamental and the fourth harmonic is 3.3 ps (see Table 3). Depending on the set-up, this implies that the conversion efficiency is very low or the pulses produced are very long.[48] Therefore, when UV pulses are produced, it is extremely important to use the thinnest possible nonlinear crystals. As a rough guide, one should choose the thickness of the crystal such that the group-delay difference between the fundamental and the harmonic is about equal to or less than the width of the pulse. An efficient harmonic-generation set-up will also use multiple conversion steps and between steps readjust the relative time delay between the laser pulses at different wavelengths.[49]

Most nonlinear crystals used for harmonic generation [BBO, lithium triborate (LBO), potassium dihydrogen-phosphate (KDP), etc.] are opaque in the deep-UV region. Very high harmonic generation in low-pressure gases has been used successfully to generate deep-UV and soft X-ray pulses. It was recently shown[50] that a glass capillary could be used to modify the phasematching condition for coherent soft X-ray generation. Using this set-up, femtosecond pulses at 800 nm were converted to the 17–32-nm wavelength range (~30th harmonic) with about 0.2 nJ energy per harmonic order. A similar set-up was used[51] to mix 800- and 400-nm pulses to produce 8-fs pulses at 270 nm. Extremely short (5 fs) amplified laser pulses have been used[52] to generate X-rays (~4 nm) in the water window by harmonic generation in a gas jet.

Table 3 Group delays calculated from the refractive index of BBO, of femtosecond pulses at various wavelengths traveling through a 1-mm crystal (ordinary polarization)

Wavelength (nm)	τ_{group} (ps)	$\Delta\tau_{group}$ (ps)
800	5.62	–
400	5.94	0.32
267	6.74	1.12
200	8.97	3.34

The advantage of very high harmonic generation is that the X-rays are generated in a well-collimated beam. The disadvantage is that it has not yet been shown to be possible to generate hard X-rays with wavelengths smaller than a molecular bond length.

Electron impact sources could generate femtosecond hard X-rays if high-charge, high-energy femtosecond electron bunches were available. Even so, electron impact sources have the disadvantage that the X-rays are emitted in a 2π solid angle, which makes their brightness typically very low. Therefore, these sources are difficult to use in diffraction experiments although not strictly impossible. Very high-power laser pulses [peak powers of the order of 1 TW (10^{12} W)] can be used to generate laser-produced plasmas.[53] Typically, a high-power laser pulse is used to evaporate a (metal) target and produce a plasma by stripping electrons off the atoms with multiphoton ionization.[54] Recombination of these electrons with the ions results in the emission of hard X-ray pulses (wavelengths of the order of 1 Å). Very recently, it has been shown that reverse Thompson scattering of terawatt laser pulses off a high-energy (50 MeV) electron beam can be used to produce hard X-rays (0.4 Å).[55] However, only 10^4 X-ray photons were generated per laser shot, resulting in about one diffracted X-ray photon per shot in diffraction off Si ⟨111⟩.

3.3 Optical Parametric Amplifier for Infrared Generation

The infrared (IR) region of the spectrum is very important for the sensing of a great variety of (transient) molecular species. Femtosecond IR pulses can be used to determine which bonds in a molecule break or form. The mid-IR fingerprint region is ideally suited to determine the presence of specific molecules in a sample. Therefore, a great deal of effort has been invested over the last decade to produce femtosecond pulses tunable in the near- and mid-IR regions. Most techniques in use now are based on parametric difference-frequency processes (see Figure 13a and b).[56] In difference-frequency mixing, a strong femtosecond pulse at frequency $w_1 + w_2$ mixes in a nonlinear crystal with a (weaker) pulse at frequency w_1, to produce a new beam of femtosecond pulses at frequency w_2. If the incident power at frequencies w_1 and w_2 is zero, a nonlinear crystal can produce these frequencies spontaneously in a process referred to as optical parametric generation (OPG). If the incident power at frequencies w_1 and w_2 is small but nonzero, the pump pulse at frequency $w_1 + w_2$ can amplify the former frequencies in what is referred to as OPA. Since the peak power of a femtosecond pulse can be extremely high while the pulse energy is relatively low, one can produce enormous parametric gain without destroying the nonlinear crystal.

The frequencies or wavelengths that are produced in OPG or OPA depend on the phasematching condition, group velocity walk-off and the type of crystal used. Figure 13(b) shows materials that are commonly used for parametric generation in the IR region. BBO has a limited tuning range in the IR range (approximately 1.2–2.8 μm when pumped at 800 nm) but is very efficient owing to the small group-velocity walk-off and high damage threshold. KTP and its analogs rubidium titanyl phosphate and cesium titanyl phosphate (RTA and CTA) are not as efficient as BBO but allow the generation of femtosecond pulses at wavelengths as long as 3–4 μm. There is a variety of crystals suitable for generation of pulses in the mid-IR region such as $AgGaS_2$ and GaSe. However, when these crystals are used to convert directly from the visible to the mid-IR, they suffer from enormous group velocity walk-off, (two-photon) absorption of the pump and very poor efficiency. Therefore, generating mid-IR pulses is typically performed in a two-stage

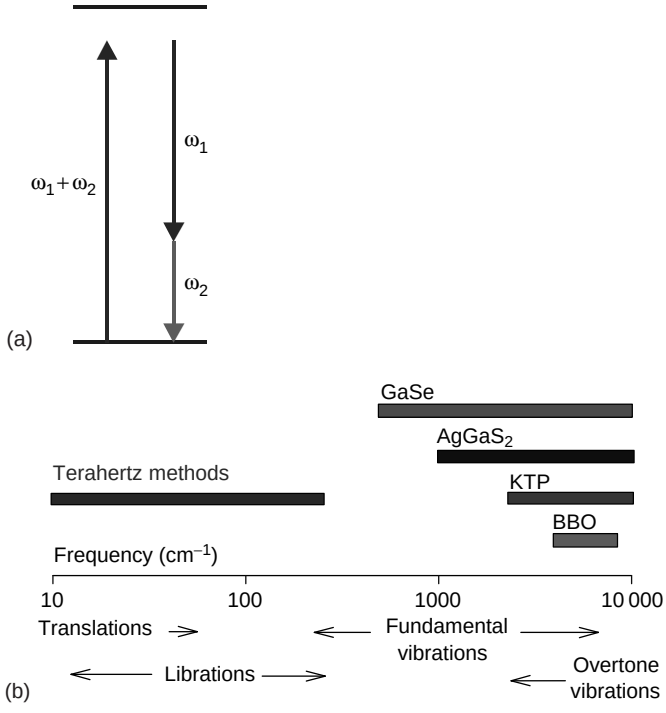

Figure 13 (a) Difference-frequency mixing. A strong femtosecond pulse at frequency $w_1 + w_2$ can mix in a nonlinear crystal with a (weaker) pulse at frequency w_1, to produce a new beam of femtosecond pulses at frequency w_2. (b) Diagram showing the various femtosecond IR tuning ranges that can be obtained using an optical parametric amplification (OPA). Using the common crystals BBO and potassium titanyl phosphate (KTP), tuning between 1 and 3 μm can be obtained. With the more unusual crystals $AgGaS_2$, $AgGaSe_2$ and GaSe, one can obtain radiation to wavelengths as long as 20 μm.

process:[57] an OPA generates two NIR frequencies (e.g. 800 nm → 1.5 μm + 1.7 μm, for example, in BBO) followed by a difference frequency mixing stage (e.g. 1.5–1.7 μm → 12 μm, for example, in $AgGaS_2$).

All known materials suitable as a nonlinear difference-frequency mixing crystal absorb strongly in the far-IR region. GaSe and $AgGaSe_2$ have the longest wavelength cutoff of about 18 μm. In order to produce longer wavelengths, different techniques have to be used (see section 3.5). It has recently been shown that periodically poled crystals such as periodically poled KTP[58] can be used with great success to generate femtosecond IR pulses efficiently.

When femtosecond IR pulses are generated, they inevitably have a very large bandwidth, e.g. about 100 cm^{-1} for a 100-fs pulse, which is clearly much larger than the line width of a typical vibrational transition. This does *not* mean that femtosecond IR pulses are useless for vibrational spectroscopy – quite the contrary! The IR pulse can be spectrally resolved after the sample. The recent availability of IR diode arrays and IR charge-coupled device (CCD) cameras means that an entire IR spectrum can be taken in a single laser shot.[57,59] However, in time-resolved experiments, the broadband IR pulse can excite multiple transitions at once and one has to take account of this carefully in the theoretical analysis of experiments.[60]

3.4 Noncollinear Optical Parametric Amplification

Noncollinear optical parametric amplification (NOPA) is a technique used to generate sub-20-fs tunable visible and NIR pulses.[61] In a collinear geometry, the temporal output of parametric generators and amplifiers is restricted by group-velocity mismatch between the pump and the generated signal and idler fields. In a noncollinear arrangement, this effect can be overcome since only the projection of the idler group velocity on to the seed is important. By arranging the pump and seed incidence angles with respect to the phase-matching angle correctly, this group-velocity mismatch can be zeroed in some nonlinear crystals. A seed pulse, which is a single filament of white-light continuum generated in a 1-mm thick piece of sapphire, is amplified by the frequency-doubled output of a Ti : sapphire amplifier, in type I BBO cut at 31°. The sapphire must be cut such that the optical axis runs perpendicular to the cut face. Type I LBO is used to generate the second harmonic of the Ti : sapphire around 400 nm. The pump beam focused on to the BBO generates a cone of parametric superfluorescence (inset of Figure 14). When the angle of incidence at the crystal is correct, there is no appreciable spatial divergence of the superfluorescence. By directing the continuum seed beam along the cone axis, a large spectral bandwidth from

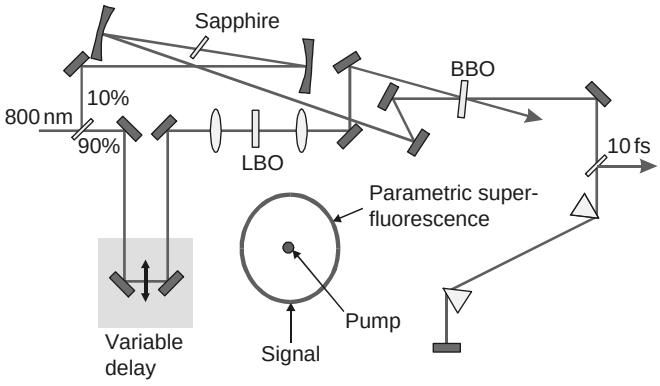

Figure 14 Schematic diagram of a noncollinear OPA. The inset shows the arrangement of the seed beam (the signal) relative to the pump and the generated superfluorescence.

the white light can be simultaneously phase matched. Adjustment of the relative delay between pump and seed and some control of the chirp on the continuum together change the center wavelength and bandwidth of the amplified light. One great advantage of this scheme is that relatively thick crystals can be employed (typically 2 mm), which results in high single-pass gain. The amplified output is then recompressed using a prism pair (BK7 separated by ~60 cm) yielding sub-20-fs visible pulses continuously tunable from 480 to ~700 nm. Pulses as short as 5 fs have been generated[62] using a prism–double chirped mirror recompression scheme, although the mirrors are unavailable commercially at the time of writing. Pumping with only 10 μJ of blue light, 2 μJ can be generated at the signal wavelength. If higher energy is available, a second amplification stage can be added and 10 μJ can be obtained for a 75-μJ pump. Using the same technique, extremely short NIR light can be generated in type II phase-matched BBO pumped at the Ti : sapphire fundamental.

3.5 Terahertz-pulse Generation and Detection

As described above, NIR and mid-IR pulses can be generated using parametric down conversion. However, down conversion cannot produce pulses with a wavelength longer than about 18 μm. Long-wavelength pulses can be generated (and detected) using so-called terahertz techniques. There are effectively two methods for generating subpicosecond terahertz (10^{12} Hz) pulses: photoconduction or optical rectification. In photoconduction, a laser pulse incident on an absorbing semiconductor creates (real) charge carriers in the conduction band. Acceleration of these carriers in an electrical bias field gives rise to a transient photocurrent that radiates electromagnetic waves. In the far field the radiated electric field is given by $\mathbf{E}_R(t) \propto d\mathbf{J}_S(t)/dt$, where $\mathbf{J}_S(t)$ is the time-dependent surface current. This method is typically used

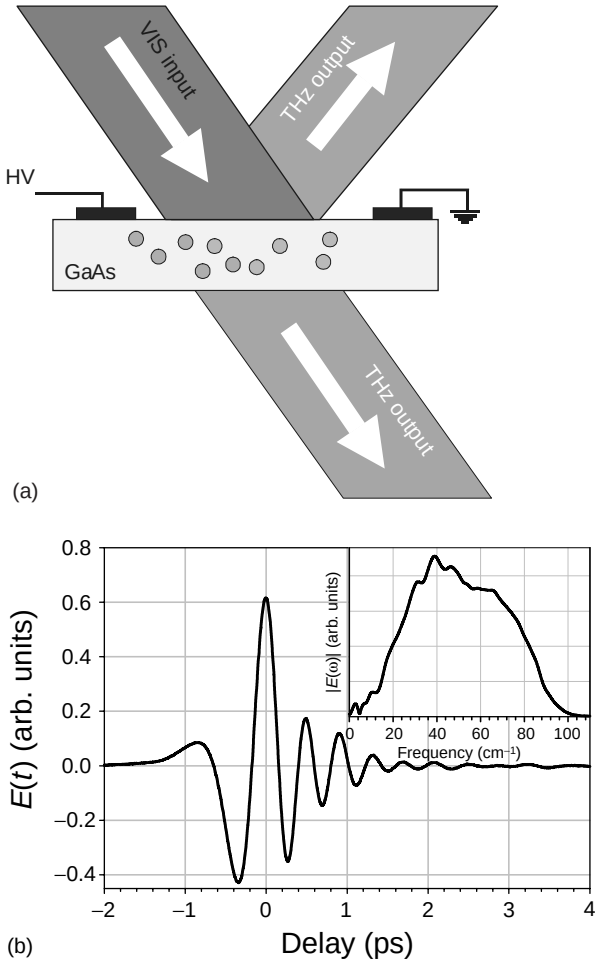

Figure 15 (a) Schematic diagram illustrating the generation of terahertz pulses. A visible femtosecond input pulse irradiates a semiconductor antenna and creates conduction electrons. The acceleration of these electrons in an external DC field results in the emission of two beams of terahertz pulses. (b) A terahertz pulse generated and detected in a 1-mm long ZnTe crystal cut perpendicular to the ⟨110⟩ akis by a 150-fs pulse at 800 nm. The inset shows the amplitude spectrum. The oscillations in the pulse are due to the finite crystal thickness. (1 THz ≈30 cm^{-1}).

in conjunction with an antenna structure (see Figure 15a and b), which allows an external bias field to be applied. An antenna structure ideally suited to be used with low-power mode-locked lasers was developed in the 1980s at Bell Laboratories[63] and IBM,[64] and is now the most common method for generating and detecting terahertz pulses. In such a set-up, two metal electrodes are laid down on a silicon or GaAs substrate, typically with a separation of ∼100 μm. A beam of femtosecond laser pulses is focused between the electrodes, in a spot with a diameter of a few micrometers. On the generation side, the metal electrodes are biased with a few tens of volts and the excitation by the pump laser triggers the emission of terahertz radiation. On the detection side, the incident terahertz beam accelerates carriers created by another visible beam, resulting in a measurable photocurrent. Since the visible beam has to be focused to a very tight spot in this method, only unamplified ultrafast lasers can be used. There is no overriding reason, however, why one should use such closely spaced electrodes. Large-aperture photoconducting antennas work very well for the generation of terahertz pulses when pumped by amplified pulses. The conversion efficiency is about 0.1%. With low repetition rate (10–1000 Hz) ultrafast laser systems, far-IR pulses with energies as high as 1 μJ have been generated.[65]

An external bias field is not strictly necessary for photoconductive generation of terahertz pulses, as real carriers generated by a visible laser pulse can be accelerated in the field of the depletion layer of the semiconductor. This surface field will accelerate the carriers perpendicular to the surface of the semiconductor and hence the terahertz oscillating dipole will be perpendicular to the surface. Therefore, terahertz radiation generated through this effect is only observed if the angle of incidence of the exciting visible laser beam is nonzero. Typically, the effect maximizes at the Brewster angle.

Optical rectification is distinct from photoconduction, in that the visible exciting beam creates virtual rather than real carriers. A more appropriate way to describe this is that the second-order susceptibility, $\chi^{(2)}$, of the crystal is used for difference-frequency mixing. Thus, the second-order polarization can be written in the time domain as Equation (21):

$$P^{(2)}_{\mathrm{THz}}(t) = \varepsilon_0 \chi^{(2)} E_{\mathrm{VIS}}(t) E^{*}_{\mathrm{VIS}}(t) \tag{21}$$

which shows that the electric field of the terahertz pulse has the same shape as the intensity envelope of the visible exciting pulse. In the last few years, it has been discovered that optical rectification is an efficient method for generating terahertz pulses if used appropriately. Since a subpicosecond terahertz pulse has a spatial length comparable to its center wavelength, it travels through a material at its phase velocity. Therefore, for optimum conversion from visible to far-IR wavelengths, one has to match the group velocity of the visible pulse with the phase velocity of the terahertz pulse.[66,67] This condition is met in some zinc blende, large band-gap semiconductors such as ZnTe and GaP when the exciting laser has a center wavelength of ca. 800 nm.

The inverse of optical rectification is electrooptic sampling: a terahertz pulse incident on an electrooptic crystal such as ZnTe will induce a birefringence through the Pockels effect. An ultrafast visible probe pulse with a variable delay copropagating through the same crystal will experience a retardation that can be retrieved with balanced detection. Scanning the relative time delay of

the probe pulse, one can record a time-domain trace of the electric field of the terahertz pulse. Using this method, signal-to-noise ratios, defined as the ratio of the terahertz pulse peak to the noise background, as high as 10^7 have been reported.[68] However, the signal-to-noise ratio with which one can measure the peak of the terahertz pulse is typically of the order of 10^2–10^3 in 100 ms. An exciting new variation on this technique is the use of a chirped probe pulse.[69] A femtosecond pulse at 800 nm can be stretched and chirped to tens of picoseconds using a grating pair. If this chirped pulse is used in the electrooptic sampling process, there will be a relationship between wavelength and relative time delay. Thus, detection of this probe pulse with a spectrometer and diode-array detector allows one to measure the entire terahertz pulse shape in a single laser shot.

As rectification and electrooptic sampling are nonresonant effects, the minimum duration of the terahertz pulses that can be generated or detected is limited only by the thickness of the crystal scaled with the difference in phase and group velocity. Thus, with ca. 10–15-fs exciting pulses at 800 nm, it was shown that terahertz pulses could be generated with detectable frequencies as high as 70 THz.[70] At these large bandwidths, it is unavoidable that the t-ray spectrum will overlap with a photon absorption band in the generating and detecting crystals, leading to large oscillations in the terahertz field trailing the main peak.

The current record highest detectable frequency for a terahertz pulse is ~70 THz, but there is no reason to believe that this could not be improved upon. Using the simple time–bandwidth relation, $\Delta\nu\Delta\tau = 0.32$, it follows that with the shortest visible pulses achievable, ca. 4–5 fs, usable power at frequencies from 0 to 160 THz ($\lambda = 1.8\,\mu m$) could be achieved. As femtosecond lasers continue to shrink in size, it may be expected that ultrafast terahertz devices may well take over from Fourier transform infrared (FTIR) spectrometers as general-purpose IR spectrometers. As an entire terahertz time-domain trace can be acquired in a single shot,[69] these devices would combine the reliability and accuracy of FTIR spectrometry with real-time speeds. The most significant aspect of ultrafast terahertz pulses, however, is that they are synchronized with visible or UV pulses, allowing time-domain spectroscopy.

3.6 Femtosecond Electron Pulses

Very short (picosecond to femtosecond) electron pulses are useful for a variety of applications. For example, picosecond electron pulses have been used as a probe pulse for determining the time-dependent structure of molecules undergoing chemical reactions (see section 5.3). However, such pulses might also be used as a seed for electron accelerators or for the generation of femtosecond X-ray pulses. Current technology for generating femtosecond electron pulses is based on experiments done in the past with picosecond and nanosecond laser pulses.[71–73] A UV laser beam is used to irradiate a metal target. If the photon energy is higher than the work function of the metal (typically 4–5 eV for common transition metals such as gold, silver or tungsten), electrons are ejected into the vacuum through the photoelectric effect. These emitted electrons are electrostatically extracted and accelerated into a narrow beam. In recent experiments (see section 5.3), this technique has been brought into the subpicosecond domain. The main problem in maintaining the time resolution is the space-charge effect: nonrelativistic electrons repel each other through Coulomb repulsion. It is therefore of the greatest importance to accelerate the electrons as quickly as possible and to keep the number of electrons per pulse as low as possible. Thus, it was seen in an experiment[74] that for a 100-μm laser spot size on the photocathode, the electron pulse would broaden to about 15 ps when there were 1000 electrons in the pulse. When the number of electrons per pulse was reduced to 100, the pulse width was less than 1 ps. There is no fundamental reason why one could not work with ten or even one electron per pulse and therefore it should be possible to achieve electron-pulse durations in the order of tens of femtoseconds. Of course, to achieve a reasonable signal-to-noise ratio in an experiment, the pulse repetition rate should be high. Such pulses will be of great value in time-dependent molecular structure determination.

4 TIME-RESOLVED EXPERIMENTS

Most electronic devices cannot measure transients much faster than about 1 ns. Although there are specialized electronic devices such as streak cameras that may be able to resolve picosecond or even hundreds of femtosecond transients in real time, in most cases it makes more sense to look for alternative detection techniques. The techniques that are used most frequently are based on auto- or cross-correlation of two beams of femtosecond pulses (see Figure 16a). If the sample is a nonlinear crystal used for sum-frequency generation, this technique can be used to determine the shape and relative arrival time of two short pulses. If the sample contains molecules or atoms that resonantly absorb the incident radiation, the experiment is a pump–probe experiment.

4.1 Auto- and Cross-correlation

In an autocorrelator, an incoming beam of pulses is split in two. One beam travels through an optical path with

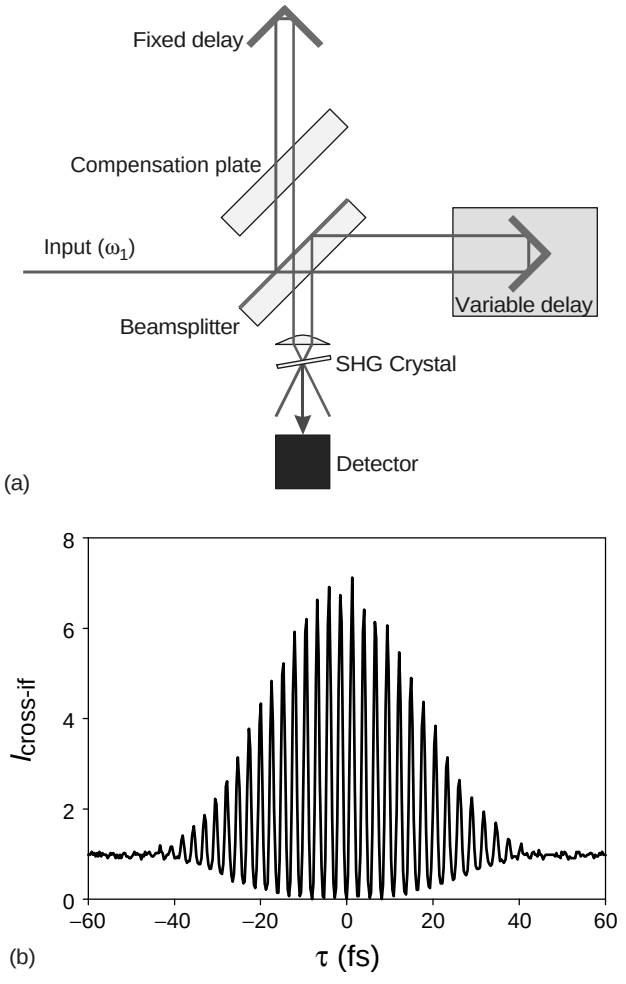

Figure 16 (a) Schematic diagram illustrating the layout of an autocorrelator. Two laser beams of femtosecond pulses at frequency w_1 are overlapped in a nonlinear crystal such as BBO. The sum-frequency signal at frequency $2w_1$ as a function of relative time delay is proportional to the shape of the pulse. (b) Interferometric autocorrelation of a 25-fs fwhm laser pulse at 800 nm.

a fixed length, the other through a path that includes an optical delay line. An optical delay line consists simply of a pair of mirrors or a retro-reflector mounted on a motorized translation stage or a sine-wave-driven loudspeaker. In such a set-up, the spatial position of the mirrors on the translation stage is related directly to the relative time delay between the two beams of pulses. For example, if the optical delay line is translated by $L = 1\,\mu\text{m}$, this corresponds to a change in relative time delay of $\tau = 2L/c \approx 6.6\,\text{fs}$. Both beams are focused in a nonlinear crystal such as BBO or KDP in order to produce second-harmonic radiation. Phase matching[56] in the crystal is used to produce a beam at the second-harmonic frequency in which one photon has been taken from one beam and one from the other. Alternatively, two-photon absorption in a photodiode may be used to provide the required second-order response.[75] The averaged measured auto- or cross-correlation signal then has the form (Equation 22)

$$I_{\text{cross}}(\tau) \propto \chi^{(2)\,2} \quad I_{w_1}(t) I_{w_2}(t-\tau)\,\mathrm{d}t \tag{22}$$

where $\chi^{(2)}$ is the second-order susceptibility of the nonlinear crystal and τ is the relative delay time $t_1 - t_2$ between the two pulses. Of course, the cross-correlation signal consists itself of a train of femtosecond pulses but since one is only interested in the *average* signal as a function of the relative time delay, no fast detector is required. Since the nonlinear crystal is chosen such that it does not exhibit any time-dependent processes of its own, the cross-correlation signal as a function of time delay is proportional to the shape of the incoming pulses. For example, in the case of an autocorrelation of a pulse with a Gaussian envelope, the autocorrelation signal as a function of delay has itself a Gaussian shape. In this example, one can measure the fwhm of the autocorrelation signal and find the width of the intensity envelope of the pulse by multiplying by 0.71. The value of the conversion factor depends on the pulse shape and a list of conversion factors can be found in Table 4. In practice, it may be wise to use nonlinear curve fitting to determine the pulse shape and corresponding conversion factor.

Clearly, the auto- or cross-correlation signal contains limited information about the pulse shape. For example, an asymmetric pulse shape will go undetected in an autocorrelation. More serious is that an autocorrelation will provide almost no information about possible chirp on the pulse (see section 2.2). If the pulse is chirped, the autocorrelation will be broader than might be expected from the pulse spectrum. Table 4 has a column with time–bandwidth products, that is, the product of the fwhm of the intensity envelope of the pulse in the time domain and the frequency domain. However, checking that the time–bandwidth product is close to that expected for a given pulse shape is *not* a very good method of making sure that the pulse is chirp free.

A better method for detecting chirp is the measurement of an interferometric autocorrelation and is taken by overlapping the two input beams. The correlation signal is now much more complicated[77] and oscillates at frequencies corresponding to the laser fundamental and second harmonic. Since the second-harmonic intensity is proportional to the square of the incident intensity, the detected second-harmonic intensity as a function of delay is (Equation 23)

$$I_{2w}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} [E(t) + E(t-\tau)]^{2\,2}\,\mathrm{d}t \tag{23}$$

Table 4 Second-order autocorrelation functions and time–bandwidth products for various pulse-shape models,[76] where τ_p = pulse width (fwhm), τ_G = autocorrelation width (fwhm) and $\Delta\nu$ = spectral width (fwhm)

$I(t)$ $(x \equiv t/T)$	$\Delta\nu\Delta\tau_p$	τ_p/T	$G_0^2(\tau)$ $(y \equiv \tau/T)$	τ_G/T	τ_p/τ_G
Gaussian $I(t) = e^{-x^2}$	0.4413	$2\sqrt{\ln 2}$	$e^{-y^2/2}$	$2\sqrt{2\ln 2}$	0.7071
Diffraction function $I(t) = \dfrac{\sin^2 x}{x^2}$	0.8859	2.7831	$\dfrac{3}{2y^2}\left(1 - \dfrac{\sin 2y}{2y}\right.$	3.7055	0.7511
Hyperbolic sech $I(t) = \operatorname{sech}^2 x$	0.3148	1.7627	$\dfrac{3(y\coth y - 1)}{\sinh^2 y}$	2.7196	0.6482
Lorentzian $I(t) = \dfrac{1}{1+x^2}$	0.2206	2	$\dfrac{1}{1+(y/2)^2}$	4	0.5000
One-sided exponential $I(t) = \begin{matrix} e^{-x} & \wedge t \geq 0 \\ 0 & \wedge t < 0 \end{matrix}$	0.1103	$\ln 2$	$e^{-\lvert y\rvert}$	$2\ln 2$	0.5000
Symmetric two-sided exponential $I(t) = e^{-2\lvert x\rvert}$	0.1420	$\ln 2$	$(1 + 2\lvert y\rvert)e^{-2\lvert y\rvert}$	1.6783	0.4130

where T is the measurement time, and $T \gg \tau$. Figure 16(b) shows an experimental interferometric autocorrelation trace of a 25-fs 800-nm pulse. If the pulse is bandwidth limited, the interferometric autocorrelation signal yields a ratio between peak and background of 8 : 1. Linear chirp (caused by GDD) is clearly visible as the peak-to-background ratio is much less than 8 : 1 in that case. Nonlinear chirp (caused, for example, by residual TOD) is clearly visible as a ''pedestal'' on the interferometric autocorrelation.

Another useful technique for characterizing femtosecond pulses is called frequency-resolved optical gating (FROG)[78] with which both the intensity and phase of a femtosecond pulse can be determined. The critical feature of FROG is that it measures the (second- or third-order) autocorrelation as a function of both time delay and frequency (see Figure 17). As before, a beam of femtosecond pulses is split in two and recombined on a nonlinear crystal. In this case, the two beams are focused on to the nonlinear crystal using a cylindrical lens. As a result, different positions on the crystal correspond to different relative delay times. If the crystal produces the second harmonic of the incident beams, the emergent beam will have the autocorrelation trace spatially encoded on to it. If the signal beam is sent through a spectrometer, a two-dimensional detector such as a CCD camera can be used to measure a signal that depends on delay in one direction and frequency in the other. Various nonlinearities can be used with the FROG technique, such as SHG and self-diffraction (effectively the OKE in a piece of glass; see section 3.1).

In the polarization-gate arrangement of self-diffraction,[79] the probe pulse $E(t)$ passes through crossed polarizers and is gated at the nonlinear medium by the gate pulse $E(t-\tau)$, which has a 45° relative polarization. A cylindrical lens focuses each beam to a line in the sample (e.g. a piece of glass) so that the delay varies spatially across the sample. Consequently, at the spectrometer,

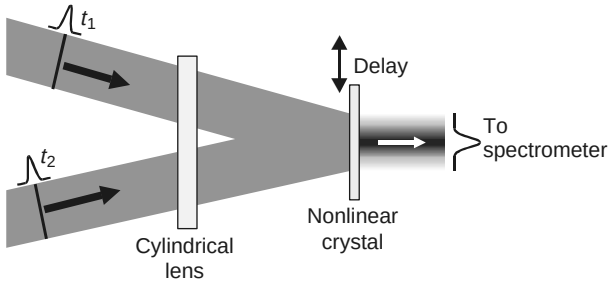

Figure 17 Experimental arrangement for single-shot FROG. Two beams of femtosecond pulses are focused with a cylindrical lens on to a nonlinear crystal, e.g. one that generates the second-harmonic of the input beam. Because of the focusing geometry, the position on the nonlinear crystal corresponds directly with time delay. This allows one to measure the correlation trace as a function of both time and frequency.

the delay τ varies along the slit, whereas the frequency w varies in a direction perpendicular to this. For such an arrangement, the resulting signal-pulse electric field is given by Equation (24):

$$E_{sig}(t,\tau) \propto \chi^{(3)} E(t)|E(t-\tau)|^2 \tag{24}$$

where $\chi^{(3)}$ is the third-order susceptibility of the sample (i.e. the strength of the OKE). The FROG trace can be considered a ''spectrogram'' of the field $E(t)$ as the signal intensity is given by Equation (25):

$$I_{FROG}(w,\tau) \approx \left| \int_{-\infty}^{\infty} E(t)g(t-\tau)e^{-iwt}\,dt \right|^2 \tag{25}$$

where the variable-delay gate function $g(t-\tau)$ is equal to $|E(t-\tau)|^2$. This quantity can be thought of as a gate that chooses a slice of the time-varying signal pulse.

At each delay, the signal consists of different frequency components, so the gate function builds up a spectrum of the pulse for every value of τ. For a transform-limited pulse, it is expected that only frequencies within the

bandwidth of that pulse would be resolved and that the instantaneous frequency remains the same at each point in time. However, for a chirped pulse, where lower frequencies lead higher ones or vice versa, the frequency-resolved trace will be considerably different, since the instantaneous frequency is a function of time. Since the phase of the signal pulse is contributed only by $E(t)$ [$E(t-\tau)$ appears as the square, so all phase information is lost], the phase can be retrieved from a FROG trace. This is achieved using an iterative Fourier-transform algorithm.

Ultrashort pulse characterization techniques generally require instantaneously responding media. The polarization gate geometry FROG technique uses the electronic Kerr effect, which is accompanied, however, by a Raman-ringing effect. The slower nuclear motion of the material must therefore be considered in the iterative algorithm used to retrieve the pulse phase and intensity. The pulse-retrieval algorithm can be modified[80] by including terms due to the slow response of the medium in the signal-field equation. Such problems with a slow response of the correlating medium can be avoided by using two-photon absorption in a diode.[75]

4.2 Pump–Probe Techniques

Most ultrafast spectroscopy experiments use the pump–probe technique, which is very similar to the cross-correlation technique described above. In pump–probe spectroscopy the pump beam is typically much stronger than the probe beam, the two beams have a different center wavelength (or the probe may not even be a laser beam), and the sample tends to be more interesting. As in the cross-correlation technique, the femtosecond time resolution is obtained by sending one of the beams through a motor-driven optical delay line. The (relatively) strong pump beam initiates some process of interest, e.g. a chemical reaction. The probe beam, entering the sample later, will be amplified, attenuated or refracted because of the changes taking place in the sample. There are so many variations of this scheme and literally thousands of experimental examples that it would be pointless to list them all. Therefore, two examples of the pump–probe technique will be described in some detail here to convey the general idea.

Figure 18 shows a very typical pump–probe spectroscopy set-up. A relatively strong pump beam is sent through a (motorized) optical delay line and focused into the sample where it initiates some chemical reaction. In many cases, the pump beam will be modified, e.g. by doubling or tripling its frequency through SHG or THG. A (small) fraction of the pump beam is split off by an optical beamsplitter and converted to another wavelength. In this example, a white-light continuum is generated as a probe but one could also use other conversion techniques to make probes at widely varying wavelengths. In any event, the probe pulse is sent through another optical delay line and focused into the sample. One may temporarily replace the sample by a nonlinear crystal to determine at which setting of the optical delay lines the pulses overlap in the sample. In Figure 18, the white-light probe is spectrally resolved by a spectrometer after the sample in order to measure the transient transmission spectrum of the sample. The signal measured as a function of the pump–probe delay time reflects the creation and destruction of chemical species in the sample. However, on short timescales ($\sim$<1 ps), there are numerous phenomena such as cross-phase modulation (see section 3.1) that may complicate the signal.[45,60] A variation on the pump–probe technique is the photon echo.[56] Photon echoes may be used to remove inhomogeneous broadening from an absorption spectrum and can therefore be very useful in analyzing complex samples.

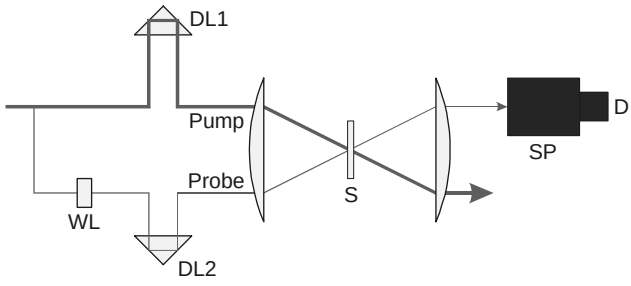

Figure 18 "Classical" pump–probe spectroscopy set-up. A relatively strong pump beam is sent through an optical delay line (DL1) and focused into the sample (S). A small fraction of the pump beam is split off and converted to another wavelength. In this example, a white-light continuum is generated as a probe, which is sent through another optical delay line (DL2) and into the sample. After the sample, the white-light probe is spectrally resolved by a spectrometer (SP) before being detected by a "slow" detector (D).

In the above example of the pump–probe technique, both the pump and probe were laser pulses. However, lasers can also be used to make pump and probe pulses of an entirely different kind. Figure 19 shows an example where femtosecond laser pulses are used to generate and gate voltage pulses.[81–84] A short laser pulse can be used to create carriers in the conduction band of a semiconductor and thereby create a femtosecond voltage pulse that can be sent down a transmission line deposited on to the sample. The voltage pulse can be picked up, for example, by a STM tip. If the current flowing through the STM tip is gated by a second laser pulse, the averaged tunnel current can be measured as a function of pump–probe delay in addition to the spatial position of the tip. Therefore, this type of pump–probe technique provides both very high temporal resolution

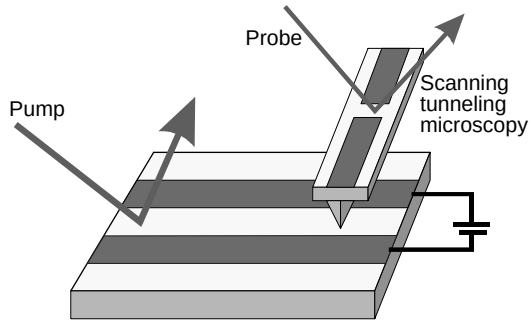

Figure 19 Unconventional pump–probe technique. In this example, the pump beam excites electrons in the conduction band of a semiconductor, thereby creating a femtosecond voltage pulse in the metal transmission line deposited on top of the semiconductor sample. The voltage pulse can be measured by an scanning tunneling microscopy (STM) tip. If the current flowing through the STM cantilever is itself gated by a probe pulse, time- and space-resolved measurements can be performed.

and spatial resolution. There are many other examples of unconventional pump–probe techniques. The probe may consist of X-ray pulses, electron pulses or any other kind of pulsed particle beam. The only critical requirement is that the pulses in the pump beam are synchronized to those in the probe beam.

5 APPLICATIONS

As ultrafast lasers are becoming smaller and easier to use, they will find increasingly down-to-earth applications. For example, IR generation using ultrafast lasers may become so simple and cheap that its advantages can be used: spatially coherent beams (resulting in better focusing and directionality), coherent detection (allowing one to measure refractive indices directly without resorting to the Kramers–Kronig relations), broadband (allowing single-shot detection of an entire spectrum) and more power at ''odd'' wavelengths (especially in the far-IR region). One of the main advantages of ultrafast lasers is, of course, that they allow time-resolved experiments with which one can observe transient species. Another advantage is that short pulses with modest energy can have huge peak power. This makes femtosecond pulses very suitable for laser ablation of materials, multiphoton absorption (for imaging of biological materials), fragmentation (e.g. of DNA into fragments that may be analyzed using MS), etc. The high peak power also permits the efficient conversion to other wavelengths using nonlinear crystals. A slightly unexpected recent observation is that when femtosecond lasers are used for machining the resulting cuts are much cleaner because the laser pulses turn the material into a plasma rather than melting it. The number of applications of ultrafast lasers is already large and will only grow. It would be impossible to describe all these applications in the limited space available here. Therefore, a few key applications will be described.

5.1 The Study of Fast Chemical Reactions

In principle, there is no difference in the information that can be obtained from a frequency-resolved or a time-resolved experiment. In a frequency-resolved spectrum, a vibronic transition may show up as a series of sharp absorption lines. In the equivalent time-resolved experiment, one will observe quantum beats whose periods correspond to the inverse of the spacings between the absorption lines. In practice, however, there may be differences. For example, even in the gas phase, the spacing between the absorption lines may be so small that it is hard to resolve them. More importantly, in the solution phase, the molecule under study will couple to the degrees of freedom of the surrounding bath. Strictly, the Schrödinger equation for the liquid could be solved, which would show that the dynamics in the liquid in fact correspond to about one-trillion-trillion (10^{24}) well-defined quantum states. Most of these bath states will be coupled to the vibronic states in the molecule, resulting in the broadening of the transition. In such a case, a spectrally resolved experiment will provide little information about the dynamics of the molecule or the dynamics of a chemical reaction that one would like to study. Even hole-burning spectroscopy will not provide much more information in most cases. Only time-resolved pump–probe techniques can elucidate the molecular dynamics.

In the 1980s and 1990s, a series of experiments were performed by a variety of groups which came to be known as ''femtochemistry'' or ''femtobiology''. These experiments exploited the property that if a molecule is excited with a laser pulse shorter than the oscillation period of relevant vibrations in the molecule, a coherent vibrational wavepacket is created in the excited state (see Figure 20a and b). Vibrations with a period shorter than the pulse width are simply not coherently excited and do not give rise to quantum beats.

The ''hydrogen atom'' of femtochemistry is the diatomic molecule NaI that was extensively studied in the gas phase.[13] When NaI is excited by a femtosecond UV pulse (see Figure 21a and b), a wavepacket is created on the repulsive excited state, which (at large interatomic distances) corresponds to the ion pair Na^+ and I^-. However, this repulsive potential forms an avoided crossing with the ground-state potential, which (at large interatomic distances) corresponds to the radical pair Na + I. If the system is probed in the excited-state potential well, it is seen that the wavepacket coherently oscillates

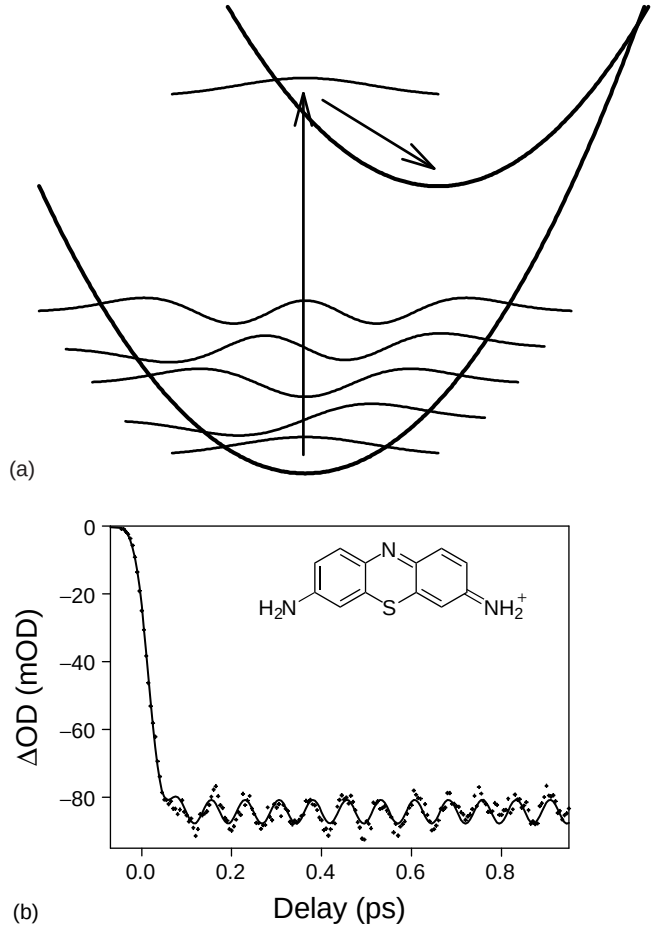

Figure 20 (a) Schematic diagram of a vibronic transition excited by a pulse whose width is smaller than the vibrational period. The pulse creates a coherent vibrational wavepacket in the excited electronic state that will oscillate in the excited-state potential well. (b) Time-resolved 600-nm pump, 700-nm probe signals in the dye molecule thionine studied with 42-fs time resolution. The signal is stimulated emission from the excited electronic state and the oscillations correspond to a vibrational mode at 443 cm^{-1} in the excited-state potential well.

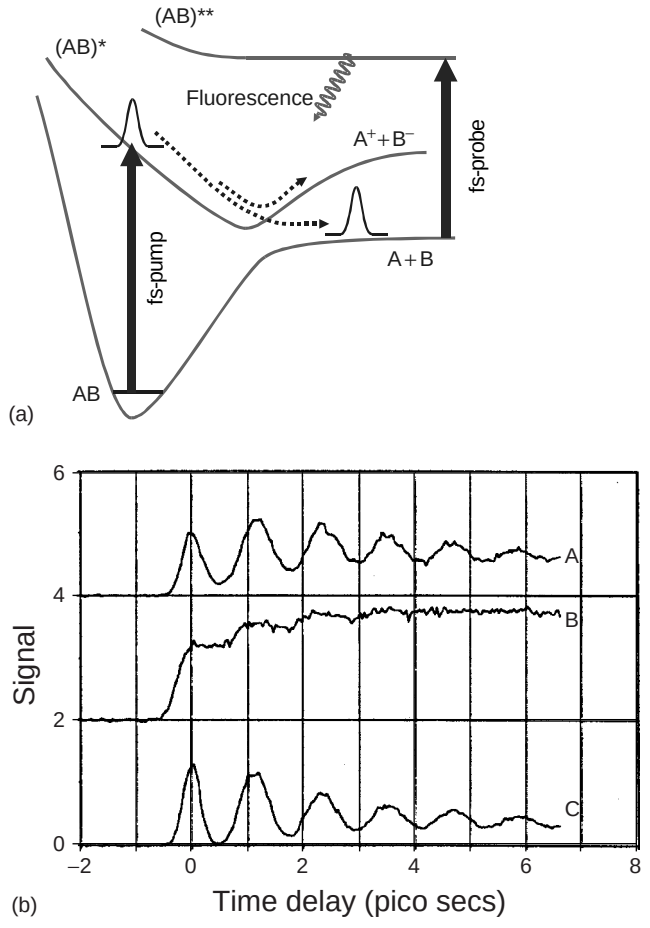

Figure 21 (a) Simple diagram depicting a typical coherent dissociation reaction. The pump pulse excites a small molecule to its first excited state. If the pump pulse is shorter than the vibrational period, it may create a coherent wavepacket that will oscillate in the excited-state potential well. In this diagram, the molecule splits into two fragments consisting of either two radicals or two ions. A probe pulse excites the molecule to a third state, from which it fluoresces back to the ground state. The total fluorescence is measured as a function of pump–probe delay. (b) Pump–probe results[86] from NaI with $\lambda_{pump} = 310$ nm. (A) $\lambda_{probe} = 580$ nm; (b) $\lambda_{probe} = 589$ nm (on resonance); and (C) $\lambda_{probe} = 612$ nm. (Reproduced from Rosker et al.[86] by permission of Elsevier Science. © Elsevier Science BV, 1988.)

in the bound state. However, if the sodium radical is probed by inducing fluorescence with the probe pulse, it is seen that the population of the radical increases in a stepwise fashion. Therefore, such an experiment shows that the chemical reaction (a dissociation in this case) is a quantum-mechanical process in which a fraction of the bound wavepacket tunnels to the state of dissociated radicals each time the wavepacket travels through the crossing region. Interestingly, it is possible to use quantum mechanical interference of wavepackets to steer a reaction in a desired direction and away from undesired directions. This is referred to as coherent control and has been applied on small molecules with varied success.[85]

A study, such as that on NaI described above, becomes interesting if one can compare the behavior in the gas phase with that in the condensed phase.[87] Various molecules that had been studied in the gas phase have now been studied in solution where again one could observe quantum beats due to vibrational wavepackets. Examples are HgI_2[88] and I_2.[89,90] Vibrational quantum beats have now been observed in a large variety of molecules in the condensed phase, including proteins.[91] In these large, complicated systems, quantum beats provide a unique opportunity to determine the potentials relevant for chemical and biological reactions. Not least, the femtochemistry technique may be used to detect

short-lived intermediates that would be invisible when ordinary spectrally resolved techniques were used. A recent experiment[1] has shown that femtochemistry can be combined with MS: pump and probe laser beams were crossed with a molecular beam and radical fragments were detected with MS as a function of pump–probe delay. The technique was used to study Diels–Alder reactions and the concept of concertedness and may well find more general use in analytical chemistry. Now that high-repetition rate, high-power lasers are available, ultrafast lasers may be used to make a variety of elusive molecules that normally occur only in the atmosphere or in space.

5.2 Imaging

Nonlinear ultrafast optics can be of immense value for (biological) imaging studies. Femtosecond lasers can be used to perform multiphoton absorption confocal-imaging studies. Using this technique, one can image the spatial distribution of one specific molecule inside a living cell.[92] Tunable femtosecond NIR pulses are ideal for this purpose as scattering and damage to the cell are reduced compared with UV radiation. A new alternative involving THG at the focus of a microscope has been used recently in imaging applications.[93] At an interface, giving rise to a slight change in refractive index or third-order nonlinear susceptibility, THG becomes phase matched.[94] The UV light generated can be detected using a photomultiplier tube or CCD camera. The conversion efficiency is low (10^{-8}), but at the tight focus of a high numerical-aperture objective lens, a fluence of 100–300 GW cm^{-2} (a few nanojoules per pulse) is enough to generate a detectable third-harmonic signal without damage to the sample. The technique benefits from being background free. NIR light at 1.2 μm has been employed,[95] giving a third-harmonic signal at 400 nm away from potentially harmful UV wavelengths, which would damage delicate (possibly living) specimens. The chirp induced by the lens is also much smaller in the NIR region. In distinct contrast to multiphoton absorption of samples tagged with fluorescent probes, no reduction in signal is observed over time since the sample is not repeatedly bleached. Scanning mirrors are used to provide traveling Lissajou patterns giving a nearly uniform sample illumination and the image is recorded on a CCD camera. An example image[95] is shown in Figure 22(a). Here, a rhizoid (a tubular single cell) tip from chara plants (green algae) is shown. The bright spots are stratoliths, vesicles containing $BaSO_4$ crystals, which flow within the cell and are necessary for the gravitational response of the alga. The CCD camera detects the motion of these stratoliths over time. This example shows that the technique is applicable to the same systems as more conventional

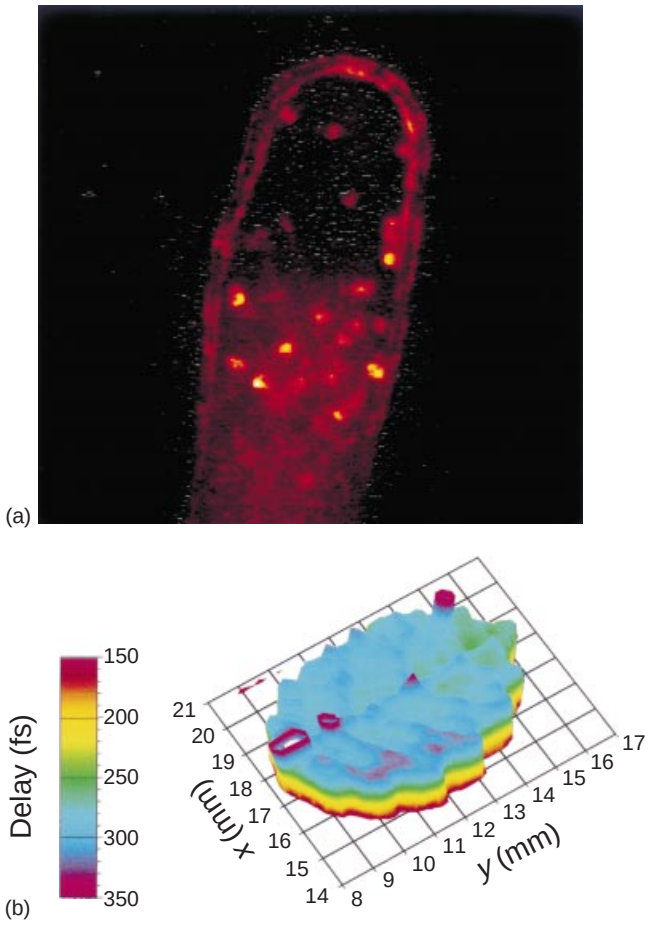

Figure 22 (a) THG image[95] at root tip showing motion of stratoliths. The image is approximately 75 × 75 μm^2. (Reproduced from J.A. Squier, M. Muller, G.J. Brakenhoff, K.R. Wilson, *Opt. Express*, **3**(9), 315–324 (1998), by permission of the Optical Society of America, 1998.) (b) THz image of a bonsai leaf taken by plotting the transmission time as a function of position.[99]

phase-contrast microscopy and will surely be developed further as techniques for optimizing the duration of the pulse at the objective are developed.[96–98]

The frequency range between ca. 100 GHz and a few terahertz has not been studied very well, as not many light sources are available. FTIR spectrometers typically peter out at frequencies below about 50 cm^{-1} (1.5 THz), even when equipped with bolometric detectors and far-IR optics. Microwave devices may go to frequencies as high as 1 THz but usually have poor tunability. However, this frequency range is of great interest as many low-frequency vibrational and rotational modes absorb here. The pseudo-band gaps of superconductors and inter-band gaps of semiconductor heterostructures are also in this range. An exiting new application of terahertz pulses is t-ray imaging. T-ray imaging was first performed at Bell Laboratories:[100,101] a beam of terahertz pulses was focused to a tight spot and a sample was scanned through

the beam. Owing to the long wavelengths at ca. 1 THz, spatial resolution was limited to about 1 mm. However, because the transit time of subpicosecond terahertz pulses can be measured with an accuracy of a few femtoseconds (see Figure 22b), it is possible to make tomographic images.[100] Thus, by measuring the internal reflections in a sample, for example, the various surfaces in a floppy disk, it is possible to detect hidden features with a depth resolution much higher than the wavelength. Very recently, several attempts have been made to improve the lateral resolution by applying near-field techniques such as using small apertures in front of the sample[102] or terahertz pulse generation in subwavelength regions.[99] It was shown recently[103] that electrooptic sampling can be performed in parallel using a large-area electrooptic crystal and an unfocused probe beam, thus allowing the acquisition of t-ray images in real time.

Using a technique analogous to ultrasound, optical-coherence tomography utilizing femtosecond lasers has permitted micrometer resolution in real-time imaging applications, particularly in biomedicine.[104] Cross-sectional images are made in a similar manner to the terahertz-imaging application by measuring the ''echo'' time of backscattered light using interferometric methods. Since the axial resolution is limited by the coherence length of the light, short-pulse Ti : sapphire and particularly NIR Cr : forsterite lasers are seen as attractive sources. An image is measured by building up the backscattered pattern at several transverse positions and these high average-power lasers are ideal for imaging at a few frames per second.

5.3 Structure Determination: Electron Beams and X-rays

Time-resolved pump–probe spectroscopy has been used for a long time to study chemical reactions. Often such spectroscopy relies on excitation and probing at visible wavelengths and transient visible spectra have to be carefully analyzed and interpreted to find out what the reaction products are. More recently, several groups have started to use femtosecond IR pulses in the hope that transient IR spectra will give more direct information about which chemical bonds are broken or formed. However, neither visible nor IR spectroscopy provides direct information about the spatial atomic structure of a given reaction intermediate.

X-ray diffraction has been used for close on a century to determine the atomic structure of crystals. Especially in the life sciences this has resulted in a much better understanding of the structure of proteins and other biomolecules and hence to a better understanding of the fundamentals of biochemistry. Because of the great absorption of electrons by materials, time-resolved electron-diffraction experiments have so far only been performed in the gas phase and on surfaces. However, as steady-state electron diffraction has also been used to determine the structure of proteins in the solid state,[105] femtosecond time-resolved electron diffraction in the solid state may also become a possibility. Electrons at present provide the most structural information for a given amount of radiation damage.[106] The atomic cross-section of carbon for (elastic or inelastic) scattering of electrons is more than six orders of magnitude larger than that of X-rays. Furthermore, radiation damage by 80–500-keV ($\lambda = 0.043$–0.017 Å) electrons is 400–1000 times less than that of $\lambda = 1.5$ Å X-rays.[106] Clearly, it would be a great advantage if one could apply the technique of diffraction on a femtosecond timescale. Modern synchrotron facilities such as the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, and the Advanced Light Source (ALS) in Berkeley, CA, USA, now make it possible to study structural dynamics using time-resolved X-ray crystallography on a nanosecond timescale.[107,108] However, many chemical and biological reactions take place on a picosecond or even femtosecond timescale. Many research groups around the world are now working towards the goal of femtosecond timescale diffraction. The technology of generating femtosecond X-ray or electron pulses is described in sections 3.2 and 3.6.

A series of (sub)picosecond time-resolved electron-diffraction experiments was recently performed on gas phase samples in a molecular beam (Figure 23a and b).[109,110] In these experiments, electron pulses were generated by exciting a back-illuminated photocathode with 400-fs 307-nm laser pulses at a 30-Hz repetition rate. The electron pulses (~6000 electrons) were then accelerated to 18.8 keV and focused with the 307-nm pump beam into a gas jet containing the sample. Two-dimensional electron diffraction patterns were recorded as a function of pump–probe delay using a CCD camera fiber-coupled to an image intensifier. The time resolution (determined by space–charge effects) was about 7 ps. In these experiments, dissociation reactions in, for example, CF_2I_2 and $Fe(CO)_5$ were studied. The method allowed the observation of rare reaction intermediates and, most importantly, the determination of their spatial structure. Picosecond time-resolved electron diffraction experiments have also been performed on surfaces. For example, the effect of melting a Pb (110) surface has been studied.[72]

Generating femtosecond hard X-ray pulses is much more difficult than generating electron pulses, which may explain the small number of unique experiments.[53] All ultrafast X-ray diffraction experiments to date[53,111] have studied the lattice expansion of semiconductors, metals

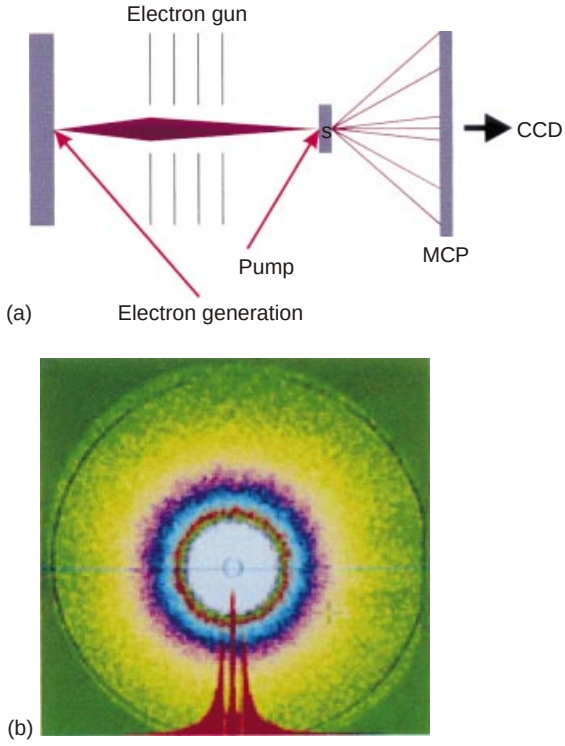

Figure 23 (a) Schematic diagram of a time-resolved electron-diffraction experiment. A strong femtosecond laser pulse irradiates a metal target and produces femtosecond electron pulses. The electrons are accelerated by an electron gun and focused into a sample. Diffracted electrons are detected by an imaging multichannel array plate (MCP) and CCD detector. If the sample is pumped by another beam of femtosecond pulses, one can detect the change in the diffraction pattern induced by the pump. (b) Femtosecond electron-diffraction pattern.[74,109,110] Picture courtesy of the California Institute of Technology.

or organic films under intense laser irradiation. In such an experiment, one typically selects one diffraction peak and measures the amount of diffracted X-rays as a function of the pump–probe delay or the shift or broadening of the diffraction peak. The changes in the diffraction peak reflect, for example, the expansion of the lattice and sound waves bouncing through the crystal.[112] It may be a while before the technique is mature enough to study the time dependence of complete diffraction patterns.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (EPSRC) and the Royal Society. In addition, we thank Professor Robin M. Hochstrasser for his support and guidance over the years.

ABBREVIATIONS AND ACRONYMS

BBO	β-Barium Borate
CCD	Charge-coupled Device
CPA	Chirped-pulse Amplification
CPM	Colliding-pulse Mode-locked
CTA	Cesium Titanyl Phosphate
CW	Continuous Wave
FHG	Fourth-harmonic Generation
FOD	Fourth-order Dispersion
FR	Faraday Rotator
FROG	Frequency-resolved Optical Gating
FTIR	Fourier Transform Infrared
fwhm	Full Width at Half-maximum
GDD	Group Delay Dispersion
GVD	Group Velocity Dispersion
IR	Infrared
KDP	Potassium Dihydrogenphosphate
KLM	Kerr Lens Mode Locking
KTP	Potassium Titanyl Phosphate
LBO	Lithium Triborate
MCP	Multichannel Array Plate
MS	Mass Spectrometry
NIR	Near-infrared
NOPA	Noncollinear Optical Parametric Amplification
OKE	Optical Kerr Effect
OPA	Optical Parametric Amplification
OPG	Optical Parametric Generation
PC	Pockels Cell
RTA	Rubidium Titanyl Phosphate
SHG	Second-harmonic Generation
STM	Scanning Tunneling Microscopy
TFP	Thin-film Polarizing Beamsplitter
THG	Third-harmonic Generation
TOD	Third-order Dispersion
UV	Ultraviolet

RELATED ARTICLES

Biomedical Spectroscopy **(Volume 1)**
Biomedical Spectroscopy: Introduction • Fluorescence Spectroscopy In Vivo • Infrared Spectroscopy in Microbiology • Optical Coherence Tomography

Biomolecules Analysis **(Volume 1)**
Infrared Spectroscopy of Biological Applications • Single Biomolecule Detection and Characterization

Clinical Chemistry **(Volume 2)**
Infrared Spectroscopy in Clinical Chemistry • Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry

Environment: Trace Gas Monitoring **(Volume 3)**
Environmental Trace Species Monitoring: Introduction • Laser Absorption Spectroscopy, Air Monitoring by Tunable Mid-infrared Diode • Laser-induced Breakdown Spectroscopy, Elemental Analysis

Environment: Water and Waste **(Volume 3)**
Infrared Spectroscopy in Environmental Analysis

Peptides and Proteins **(Volume 7)**
X-ray Crystallography of Biological Macromolecules

Atomic Spectroscopy **(Volume 11)**
Laser Ablation in Atomic Spectroscopy • Laser-induced Breakdown Spectroscopy

Electroanalytical Methods **(Volume 11)**
Ultrafast Electrochemical Techniques

Electronic Absorption and Luminescence **(Volume 12)**
Fluorescence Imaging Microscopy • Fluorescence Lifetime Measurements, Applications of

Infrared Spectroscopy **(Volume 12)**
Infrared Spectroscopy: Introduction

Kinetic Determinations **(Volume 12)**
Kinetic Determinations: Introduction

Kinetic Determinations cont'd **(Volume 13)**
Instrumentation for Kinetics

Mass Spectrometry **(Volume 13)**
Time-of-flight Mass Spectrometry

X-ray Spectrometry **(Volume 15)**
Structure Determination, X-ray Diffraction for • Ultrafast Diffraction Techniques

REFERENCES

1. E.W.G. Diau, S. DeFeyter, A.H. Zewail, 'Femtosecond Dynamics of Retro Diels–Alder Reactions: The Concept of Concertedness', *Chem. Phys. Lett.*, **304**(3–4), 134–144 (1999).
2. G. Porter, 'Flash Photolysis', *Proc. R. Soc. London, Ser. A*, **200**, 284–300 (1950).
3. H.W. Mocker, R.J. Collins, 'Mode locked Ruby Laser', *Appl. Phys. Lett.*, **7**, 270–273 (1965).
4. A.J. DeMaria, D.A. Stetser, H. Heynau, 'Mode locked Nd Glass', *Appl. Phys. Lett.*, **8**, 174–176 (1966).
5. M.P. Irvine, R.J. Harrison, G.S. Beddard, P. Leighton, J.K.M. Sanders, 'Detection of the Inverted Region in the Photoinduced Intramolecular Electron Transfer of Capped Porphyrins', *Chem. Phys.*, **104**(2), 315–324 (1986).
6. J.R. Miller, J.V. Beitz, R.K. Huddleston, 'Effect of Free Energy on Rates of Electron Transfer Between Molecules', *J. Am. Chem. Soc.*, **106**(18), 5057–5068 (1984).
7. R.A. Marcus, 'ET', *J. Chem. Phys.*, **24**, 966–978 (1956).
8. R.M. Bowman, K.B. Eisenthal, D.P. Millar, 'Frictional Effects on Barrier Crossing in Solution – Comparison with the Kramers Equation', *J. Chem. Phys.*, **89**(2), 762–769 (1988).
9. M.G. Hyde, G.S. Beddard, 'Picosecond Photodissociation and Geminate Radical Recombination in Alkane Solutions of Tetraphenylhydrazine', *Chem. Phys.*, **151**(2), 239–248 (1991).
10. R.L. Fork, B.I. Greene, C.V. Shank, 'First CPM', *Appl. Phys. Lett.*, **38**(9), 671–672 (1981).
11. J.A. Valdmanis, R.L. Fork, 'prism CPM', *IEEE J. Quantum Electron.*, **22**, 112–118 (1986).
12. R.L. Fork, C.V. Shank, R.T. Yen, '10 Hz Amp', *Appl. Phys. Lett.*, **41**(3), 223–225 (1982).
13. A. Zewail, *The Chemical Bond. Structure and Dynamics*, Academic Press, Boston, 1992.
14. R.L. Fork, C. Brito, C.H.P.C. Becker, C.V. Shank, 'Compression of Optical Pulses to Six Femtoseconds by Using Cubic Phase Compensation', *Opt. Lett.*, **12**, 483–485 (1987).
15. D.E. Spence, P.N. Kean, W. Sibbett, '60-fs Pulse Generation From a Self-mode-locked Ti–Sapphire Laser', *Opt. Lett.*, **16**(1), 42–44 (1991).
16. D. Strickland, G. Mourou, 'Compression of Amplified Chirped Optical Pulses', *Opt. Commun.*, **55**(6), 447–449 (1985).
17. T. Elsaesser, J.G. Fujimoto, D.A. Wiersma, W. Zinth (eds.), *Ultrafast Phenomena XI (formerly Picosecond Phenomena)*, Springer, Berlin, Vol. 63, 1998.
18. A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986.
19. O. Svelto, *Principles of Lasers*, 3rd edition, Plenum Press, New York, 1989.
20. J.D. Simon, 'Ultrashort Light Pulses', *Rev. Sci. Instrum.*, **60**(12), 3597–3624 (1989).
21. P.F. Moulton, 'Spectroscopic and Laser Characteristics of $TiAl_2O_3$', *J. Opt. Soc. Am. B*, **3**(1), 125–133 (1986).
22. M.T. Asaki, C.P. Huang, D. Garvey, J.P. Zhou, H.C. Kapteyn, M.M. Murnane, 'Generation of 11-fs Pulses from a Self-mode-locked Ti–Sapphire Laser', *Opt. Lett.*, **18**(12), 977–979 (1993).
23. I.P. Christov, V.D. Stoev, M.M. Murnane, H.C. Kapteyn, 'Sub-10-fs Operation of Kerr-lens Mode-locked Lasers', *Opt. Lett.*, **21**(18), 1493–1495 (1996).
24. M.S. Pshenichnikov, W.P. Deboeij, D.A. Wiersma, 'Generation of 13-fs, 5-MW Pulses from a Cavity-dumped Ti–Sapphire Laser', *Opt. Lett.*, **19**(8), 572–574 (1994).

25. A. Stingl, C. Spielmann, F. Krausz, R. Szipocs, 'Generation of 11-fs Pulses from a Ti–Sapphire Laser without the Use of Prisms', *Opt. Lett.*, **19**(3), 204–206 (1994).
26. U. Morgner, F.X. Kartner, S.H. Cho, Y. Chen, H.A. Haus, J.G. Fujimoto, E.P. Ippen, V. Scheuer, G. Angelow, T. Tschudi, 'Sub-two-cycle Pulses from a Kerr-lens Mode-locked Ti:Sapphire Laser', *Opt. Lett.*, **24**(6), 411–413 (1999).
27. I.D. Jung, F.X. Kartner, N. Matuschek, D.H. Sutter, F. Morier-Genoud, G. Zhang, U. Keller, V. Scheuer, M. Tilsch, T. Tschudi, 'Self-starting 6.5-fs Pulses from a Ti:Sapphire Laser', *Opt. Lett.*, **22**(13), 1009–1011 (1997).
28. M.A. Arbore, M.M. Fejer, M.E. Fermann, A. Hariharan, A. Galvanauskas, D. Harter, 'Frequency Doubling of Femtosecond Erbium-fiber Soliton Lasers in Periodically Poled Lithium Niobate', *Opt. Lett.*, **22**(1), 13–15 (1997).
29. S. Backus, C.G. Durfee, M.M. Murnane, H.C. Kapteyn, 'High Power Ultrafast Lasers', *Rev. Sci. Instrum.*, **69**(3), 1207–1223 (1998).
30. R.L. Fork, O.E. Martinez, J.P. Gordon, 'Prism Compression', *Opt. Lett.*, **9**, 150–152 (1984).
31. E.B. Treacy, 'Optical Pulse Compression with Diffraction Gratings', *IEEE J. Quantum Electron.*, **5**(9), 454–458 (1969).
32. O.E. Martinez, '3000 Times Grating Compressor with Positive Group-velocity Dispersion – Application to Fiber Compensation in 1.3–1.6 μm Region', *IEEE J. Quantum Electron.*, **23**(1), 59–64 (1987).
33. G. Cheriaux, P. Rousseau, F. Salin, J.P. Chambaret, B. Walker, L.F. Dimauro, 'Aberration-free Stretcher Design for Ultrashort-pulse Amplification', *Opt. Lett.*, **21**(6), 414–416 (1996).
34. J.V. Rudd, G. Korn, S. Kane, J. Squier, G. Mourou, P. Bado, 'Chirped-pulse Amplification of 55-fs Pulses at a 1-kHz Repetition Rate in a Ti–Al_2O_3 Regenerative Amplifier', *Opt. Lett.*, **18**(23), 2044–2046 (1993).
35. C.P.J. Barty, C.L. Gordon, B.E. Lemoff, 'Multiterawatt 30-fs Ti–Sapphire Laser System', *Opt. Lett.*, **19**(18), 1442–1444 (1994).
36. S. Backus, J. Peatross, C.P. Huang, M.M. Murnane, H.C. Kapteyn, 'Ti–Sapphire Amplifier Producing Millijoule-level, 21-fs Pulses at 1 kHz', *Opt. Lett.*, **20**(19), 2000–2002 (1995).
37. J.P. Chambaret, C. LeBlanc, G. Cheriaux, P. Curley, G. Darpentigny, P. Rousseau, G. Hamoniaux, A. Antonetti, F. Salin, 'Generation of 25-TW, 32-fs Pulses at 10 Hz', *Opt. Lett.*, **21**(23), 1921–1923 (1996).
38. S. Backus, C.G. Durfee, M.M. Murnane, H.C. Kapteyn, 'High Power Ultrafast Lasers', *Rev. Sci. Instrum.*, **69**(3), 1207–1223 (1998).
39. S. Backus, C.G. Durfee, G. Mourou, H.C. Kapteyn, M.M. Murnane, '0.2-TW Laser System at 1 kHz', *Opt. Lett.*, **22**(16), 1256–1258 (1997).
40. F. Salin, C. LeBlanc, J. Squier, C. Barty, 'Thermal Eigenmode Amplifiers for Diffraction-limited Amplification of Ultrashort Pulses', *Opt. Lett.*, **23**(9), 718–720 (1998).
41. K. Wynne, G.D. Reid, R.M. Hochstrasser, 'Regenerative Amplification of 30-fs Pulses in Ti–Sapphire at 5 kHz', *Opt. Lett.*, **19**(12), 895–897 (1994).
42. J. Squier, C.P.J. Barty, F. Salin, C. LeBlanc, S. Kane, 'Use of Mismatched Grating Pairs in Chirped-pulse Amplification Systems', *Appl. Opt.*, **37**(9), 1638–1641 (1998).
43. G.D. Reid, 'Use of Mismatched Gratings in Conjunction with High Dispersion Prism Pairs in Chirped-pulse Amplification Systems', in preparation.
44. C.P.J. Barty, T. Guo, C. LeBlanc, F. Raksi, C. RosePetruck, J. Squier, K.R. Wilson, V.V. Yakovlev, K. Yamakawa, 'Generation of 18-fs, Multiterawatt Pulses by Regenerative Pulse Shaping and Chirped-pulse Amplification', *Opt. Lett.*, **21**(9), 668–670 (1996).
45. L. Dhar, J.A. Rogers, K.A. Nelson, 'Time-resolved Vibrational Spectroscopy in the Impulsive Limit', *Chem. Rev.*, **94**(1), 157–193 (1994).
46. W.H. Knox, R.L. Fork, M.C. Downer, R.H. Stolen, C.V. Shank, J.A. Valdmanis, 'Optical Pulse Compression to 8 fs at a 5-kHz Repetition Rate', *Appl. Phys. Lett.*, **46**(2), 1120–1121 (1985).
47. M. Nisoli, S. Stagira, S. DeSilvestri, O. Svelto, S. Sartania, Z. Cheng, M. Lenzner, C. Spielmann, F. Krausz, 'A Novel High Energy Pulse Compression System: Generation of Multigigawatt Sub-5-fs Pulses', *Appl. Phys. B*, **65**(2), 189–196 (1997).
48. A.M. Weiner, 'Effect of Group Velocity Mismatch on the Measurement of Ultrashort Optical Pulses via Second Harmonic Generation', *IEEE J. Quantum Electron.*, **19**(8), 1276–1283 (1983).
49. J. Ringling, O. Kittelmann, F. Noack, G. Korn, J. Squier, 'Tunable Femtosecond Pulses in the Near Vacuum Ultraviolet Generated by Frequency Conversion of Amplified Ti–Sapphire Laser Pulses', *Opt. Lett.*, **18**(23), 2035–2037 (1993).
50. A. Rundquist, C.G. Durfee, Z.H. Chang, C. Herne, S. Backus, M.M. Murnane, H.C. Kapteyn, 'Phase-matched Generation of Coherent Soft X-rays', *Science*, **280**(5368), 1412–1415 (1998).
51. C.G. Durfee, S. Backus, H.C. Kapteyn, M.M. Murnane, 'Intense 8-fs Pulse Generation in the Deep Ultraviolet', *Opt. Lett.*, **24**(10), 697–699 (1999).
52. M. Lenzner, M. Schnuerer, C. Spielmann, F. Krausz, 'Extreme Nonlinear Optics with Few-cycle Laser Pulses', *IEICE Trans. Electron.*, **E81-C**(2), 112–122 (1998).
53. J. Wark, 'Time-resolved X-ray Diffraction', *Contemp. Phys.*, **37**(3), 205–218 (1996).
54. D. Giulieti, L.A. Gizzi, 'X-ray Emission from Laser-produced Plasmas', *Riv. Nuovo Cimento*, **21**(10), 1–93 (1998).
55. R.W. Schoenlein, W.P. Leemans, A.H. Chin, P. Volfbeyn, T.E. Glover, P. Balling, M. Zolotorev, K.J. Kim, S. Chattopadhyay, C.V. Shank, 'Femtosecond X-ray Pulses at 0.4 Ångstrom Generated by 90 Degrees Thomson Scattering: A Tool for Probing the Structural

Dynamics of Materials', *Science*, **274**(5285), 236–238 (1996).
56. Y.R. Shen, *The Principles of Nonlinear Optics*, Wiley, New York, 1986.
57. B. Akhremitchev, C.F. Wang, G.C. Walker, 'A Femtosecond Absorption Spectrometer Tunable from 50 000 to 800 cm^{-1}: Nonlinear Optics and Pump/Probe Geometries', *Rev. Sci. Instrum.*, **67**(11), 3799–3805 (1996).
58. D.T. Reid, G.T. Kennedy, A. Miller, M. Sibbett, M. Ebrahimzadeh, 'Widely Tunable, Near- to Mid-infrared Femtosecond and Picosecond Optical Parametric Oscillators Using Periodically Poled $LiNbO_3$ and $RbTiOAsO_4$', *IEEE J. Sel. Top. Quantum Electron.*, **4**(2), 238–248 (1998).
59. P. Hamm, S. Wiemann, M. Zurek, W. Zinth, 'Highly Sensitive Multichannel Spectrometer for Subpicosecond Spectroscopy in the Midinfrared', *Opt. Lett.*, **19**(20), 1642–1644 (1994).
60. K. Wynne, R.M. Hochstrasser, 'The Theory of Ultrafast Vibrational Spectroscopy', *Chem. Phys.*, **193**(3), 211–236 (1995).
61. T. Wilhelm, J. Piel, E. Riedle, 'Sub-20-fs Pulses Tunable Across the Visible from a Blue-pumped Single-pass Noncollinear Parametric Converter', *Opt. Lett.*, **22**(19), 1494–1496 (1997).
62. A. Shirakawa, I. Sakane, M. Takasaka, T. Kobayashi, 'Sub-5-fs Visible Pulse Generation by Pulse-front-matched Noncollinear Optical Parametric Amplification', *Appl. Phys. Lett.*, **74**(16), 2268–2270 (1999).
63. P.R. Smith, D.H. Auston, M.C. Nuss, 'Subpicosecond Photoconducting Dipole Antennas', *IEEE J. Quantum Electron.*, **24**(2), 255–260 (1988).
64. D. Grischkowsky, S. Keiding, M. van Exter, C. Fattinger, 'Far-infrared Time-domain Spectroscopy with Terahertz Beams of Dielectrics and Semiconductors', *J. Opt. Soc. Am. B*, **7**, 2006–2015 (1990).
65. D. You, R.R. Jones, P.H. Bucksbaum, D.R. Dykaar, 'Generation of High-power Sub-single-cycle 500-fs Electromagnetic Pulses', *Opt. Lett.*, **18**(4), 290–292 (1993).
66. A. Nahata, A.S. Weling, T.F. Heinz, 'A Wideband Coherent Terahertz Spectroscopy System Using Optical Rectification and Electro-optic Sampling', *Appl. Phys. Lett.*, **69**(16), 2321–2323 (1996).
67. Q. Wu, X.-C. Zhang, '7 Terahertz Broadband GaP Electro-optic Sensor', *Appl. Phys. Lett.*, **70**(14), 1784–1786 (1997).
68. Q. Wu, X.-C. Zhang, 'Ultrafast Electro-optic Field Sensors', *Appl. Phys. Lett.*, **68**(12), 1604–1606 (1996).
69. Z. Jiang, X.-C. Zhang, 'Electro-optic Measurement of THz Field Pulses with a Chirped Optical Beam', *Appl. Phys. Lett.*, **72**(16), 1945–1947 (1998).
70. A. Leitenstorfer, S. Hunsche, J. Shah, M.C. Nuss, W.H. Knox, 'Detectors and Sources for Ultrabroadband Electro-optic Sampling: Experiment and Theory', *Appl. Phys. Lett.*, **74**(11), 1516–1518 (1999).
71. G. Mourou, S. Williamson, 'Picosecond Electron Diffraction', *Appl. Phys. Lett.*, **41**(1), 44–45 (1982).
72. H.E. Elsayed-Ali, J.W. Herman, 'Ultrahigh Vacuum Picosecond Laser-driven Electron Diffraction System', *Rev. Sci. Instrum.*, **61**(6), 1636–1647 (1990).
73. V.A. Lobastov, J.D. Ewbank, L. Schaefer, A.A. Ischenko, 'Instrumentation for Time-resolved Electron Diffraction Spanning the Time Domain from Microseconds to Picoseconds', *Rev. Sci. Instrum.*, **69**(7), 2633–2643 (1998).
74. M. Dantus, S.B. Kim, J.C. Williamson, A.H. Zewail, 'Ultrafast Electron Diffraction. 5. Experimental Time Resolution and Applications', *J. Phys. Chem.*, **98**, 2782–2796 (1994).
75. D.T. Reid, W. Sibbett, J.M. Dudley, L.P. Barry, B. Thomsen, J.D. Harvey, 'Commercial Semiconductor Devices for Two Photon Absorption Autocorrelation of Ultrashort Light Pulses', *Appl. Opt.*, **37**(34), 8142–8144 (1998).
76. K.L. Sala, G.A. Kenney-Wallace, G.E. Hall, 'CW Autocorrelation Measurements of Picosecond Laser Pulses', *IEEE J. Quantum Electron.*, **16**(9), 990–996 (1980).
77. K. Naganuma, K. Mogi, H. Yamada, 'General Method for Ultrashort Light-pulse Chirp Measurement', *IEEE J. Quantum Electron.*, **25**(6), 1225–1233 (1989).
78. K.W. DeLong, D.N. Fittinghoff, R. Trebino, 'Practical Issues in Ultrashort-laser-pulse Measurement Using Frequency-resolved Optical Gating', *IEEE J. Quantum Electron.*, **32**(7), 1253–1264 (1996).
79. R. Trebino, D.J. Kane, 'Using Phase Retrieval to Measure the Intensity and Phase of Ultrashort Pulses – Frequency-resolved Optical Gating', *J. Opt. Soc. Am. A*, **10**(5), 1101–1111 (1993).
80. K.W. Delong, C.L. Ladera, R. Trebino, B. Kohler, K.R. Wilson, 'Ultrashort-pulse Measurement Using Noninstantaneous Nonlinearities – Raman Effects in Frequency-resolved Optical Gating', *Opt. Lett.*, **20**(5), 486–488 (1995).
81. G. Nunes, M.R. Freeman, 'Picosecond Resolution in Scanning-tunneling-microscopy', *Science*, **262**(5136), 1029–1032 (1993).
82. D. Botkin, J. Glass, D.S. Chemla, D.F. Ogletree, M. Salmeron, S. Weiss, 'Advances in Ultrafast Scanning Tunneling Microscopy', *Appl. Phys. Lett.*, **69**(9), 1321–1323 (1996).
83. G.M. Steeves, A.Y. Elezzabi, M.R. Freeman, 'Nanometer-scale Imaging with an Ultrafast Scanning Tunneling Microscope', *Appl. Phys. Lett.*, **72**(4), 504–506 (1998).
84. M.J. Feldstein, P. Vöhringer, W. Wang, N.F. Scherer, 'Femtosecond Optical Spectroscopy and Scanning Probe Microscopy', *J. Phys. Chem.*, **100**, 4739–4748 (1996).
85. N.F. Scherer, R.J. Carlson, A. Matro, M. Du, A.J. Ruggiero, V. Romero-Rochin, J.A. Cina, G.R. Fleming, S.A. Rice, 'Fluorescence-detected Wave Packet Interferometry: Time Resolved Molecular Spectroscopy with

Sequences of Femtosecond Phase-locked Pulses', *J. Chem. Phys.*, **95**(3), 1487–1511 (1991).
86. M.J. Rosker, T.S. Rose, A.H. Zewail, 'Femtosecond Real-time Dynamics of Photofragment-trapping Resonances on Dissociative Potential Energy Surfaces', *Chem. Phys. Lett.*, **146**(3,4), 175–179 (1988).
87. G.A. Voth, R.M. Hochstrasser, 'Transition State Dynamics and Relaxation Processes in Solutions: A Frontier of Physical Chemistry', *J. Phys. Chem.*, **100**(31), 13034–13049 (1996).
88. N. Pugliano, A.Z. Szarka, R.M. Hochstrasser, 'Relaxation of the Product State Coherence Generated Through the Photolysis of HgI_2 in Solution', *J. Chem. Phys.*, **104**(13), 5062–5079 (1996).
89. Y. Yan, R.M. Whitnell, K.R. Wilson, A.H. Zewail, 'Femtosecond Chemical Dynamics in Solution. Wavepacket Evolution and Caging of I_2', *Chem. Phys. Lett.*, **193**(5), 402–412 (1992).
90. N.F. Scherer, L.D. Ziegler, G.R. Fleming, 'Heterodyne-detected Time-domain Measurement of I_2 Predissociation and Vibrational Dynamics in Solution', *J. Chem. Phys.*, **96**, 5544–5547 (1992).
91. K. Wynne, R.M. Hochstrasser, 'Coherence and Adiabaticity in Ultrafast Electron Transfer', *Adv. Chem. Phys.*, **107**, 263–309 (1999).
92. R.M. Williams, J.B. Shear, W.R. Zipfel, S. Maiti, W.W. Webb, 'Mucosal Mast Cell Secretion Processes Imaged Using Three-photon Microscopy of 5-Hydroxytryptamine Autofluorescence', *Biophys. J.*, **76**(4), 1835–1846 (1999).
93. T.Y.F. Tsang, 'Optical 3rd-harmonic Generation at Interfaces', *Phys. Rev. A*, **52**(5), 4116–4125 (1995).
94. Y. Barad, H. Eisenberg, M. Horowitz, Y. Silberberg, 'Nonlinear Scanning Laser Microscopy by Third Harmonic Generation', *Appl. Phys. Lett.*, **70**(8), 922–924 (1997).
95. J.A. Squier, M. Muller, G.J. Brakenhoff, K.R. Wilson, 'Third Harmonic Generation Microscopy', *Opt. Express*, **3**(9), 315–324 (1998).
96. D.N. Fittinghoff, A.C. Millard, J.A. Squier, M. Muller, 'Frequency-resolved Optical Gating Measurement of Ultrashort Pulses Passing Through a High Numerical Aperture Objective', *IEEE J. Quantum Electron.*, **35**(4), 479–486 (1999).
97. A.C. Millard, D.N. Fittinghoff, J.A. Squier, M. Muller, A.L. Gaeta, 'Using GaAsP Photodiodes to Characterize Ultrashort Pulses Under High Numerical Aperture Focusing in Microscopy', *J. Microsc.*, **193**(Pt. 3), 179–181 (1999).
98. J.A. Squier, D.N. Fittinghoff, C.P.J. Barty, K.R. Wilson, M. Muller, G.J. Brakenhoff, 'Characterization of Femtosecond Pulses Focused with High Numerical Aperture Optics Using Interferometric Surface-third-harmonic Generation', *Opt. Commun.*, **147**(1–3), 153–156 (1998).
99. K. Wynne, D.A. Jaroszynski, 'Superluminal Terahertz Pulses', *Opt. Lett.*, **24**(1), 25–27 (1999).
100. D.M. Mittleman, R.H. Jacobsen, M.C. Nuss, 'T-ray Imaging', *IEEE J. Sel. Top. Quantum Electron.*, **2**(3), 679–692 (1996).
101. D.M. Mittleman, M. Gupta, R. Neelamani, R.G. Baraniuk, J.V. Rudd, M. Koch, 'Recent Advances in Terahertz Imaging', *Appl. Phys. B*, **68**(6), 1085–1094 (1999).
102. S. Hunsche, M. Koch, I. Brener, M.C. Nuss, 'THz Near-field Imaging', *Opt. Commun.*, **150**(1–6), 22–26 (1998).
103. Q. Wu, T.D. Hewitt, X.-C. Zhang, 'Two-dimensional Electro-optic Imaging of THz Beams', *Appl. Phys. Lett.*, **69**(8), 1026–1028 (1996).
104. G.J. Tearney, B.E. Bouma, S.A. Boppart, B. Golubovic, E.A. Swanson, J.G. Fujimoto, 'Rapid Acquisition of in Vivo Biological Images by Use of Optical Coherence Tomography', *Opt. Lett.*, **21**(17), 1408–1410 (1996).
105. N. Grigorieff, T.A. Ceska, K.H. Downing, J.M. Baldwin, R. Henderson, 'Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin', *J. Mol. Biol.*, **259**, 393–421 (1996).
106. R. Henderson, 'The Potential and Limitations of Neutrons, Electrons and X-rays for Atomic Resolution Microscopy of Unstained Biological Molecules', *Q. Rev. Biophys.*, **28**(2), 171–193 (1995).
107. B.L. Stoddard, 'New Results Using Laue Diffraction and Time-resolved Crystallography', *Curr. Opin. Struct. Biol.*, **8**, 612–618 (1998).
108. K. Moffat, 'Ultrafast Time-resolved Crystallography', *Nature Struct. Biol.*, **5**, 641–643 (1998).
109. H. Ihee, J. Cao, A.H. Zewail, 'Ultrafast Electron Diffraction: Structures in Dissociation Dynamics of $Fe(CO)_5$', *Chem. Phys. Lett.*, **281**, 10–19 (1997).
110. J. Cao, H. Ihee, A.H. Zewail, 'Ultrafast Electron Diffraction and Direct Observation of Transient Structures in a Chemical Reaction', *Proc. Natl. Acad. Sci. USA*, **96**, 338–342 (1999).
111. A.H. Chin, R.W. Schoenlein, T.E. Glover, P. Balling, W.P. Leemans, C.V. Shank, 'Ultrafast Structural Dynamics in InSb Probed by Time-resolved X-ray Diffraction', *Phys. Rev. Lett.*, **83**(2), 336–339 (1999).
112. C. Rose-Petruck, R. Jimenez, T. Guo, A. Cavalleri, C.W. Siders, F. Raksi, J.A. Squier, B.C. Walker, K.R. Wilson, C.P.J. Barty, 'Picosecond-milliangstrom Lattice Dynamics Measured by Ultrafast X-ray Diffraction', *Nature (London)*, **398**(6725), 310–312 (1999).